Privacy-Preserving Face Detection: A Comprehensive Analysis of Face Anonymization Techniques

Ricardo Andrade
Mestrado em Ciência de Dados
Departamento de Ciência de Computadores
2023

Orientador
Prof. Dr. João Paulo Vilela, Professor Auxiliar,
Faculdade de Ciências da Universidade do Porto

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

" The pursuit of knowledge is the most noble of all human endeavors, for it enriches the mind and empowers the spirit. "

, written by ChatGPT

Acknowledgements

I would like to express my sincere gratitude to everyone who has supported me throughout this journey of completing my master's thesis.

First and foremost, I am deeply thankful to my thesis supervisor, Professor João Vilela, and advisor, João Pinto, for their unwavering guidance, expertise, and patience. Their valuable insights and constructive feedback have been instrumental in shaping this research.

I extend my heartfelt appreciation to my family for their constant encouragement and belief in my abilities - Lola, Tony, Cláudia, Júlio, Baba, Pedro, Clara, Pedro. Their love and support have been my pillars of strength.

I want to acknowledge my friends and peers for their camaraderie and for being a source of inspiration and motivation. Above all, I want to extend my heartfelt gratitude to my muse, Mafalda França.

Lastly, I thank all the participants and individuals who contributed to this study by sharing their time and insights particularly, to the 200 volunteers who willingly participated in the data collection process. Your contributions were immensely valuable to the success of this research.

This thesis would not have been possible without the collective support and encouragement of these wonderful people. Thank you for being part of this significant milestone in my academic journey.

This work was funded by European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) and Portugal 2020 [Project nº 047264 - Theia; Funding Reference: POCI-01-0247-FEDER-047264].

UNIVERSIDADE DO PORTO

Abstract

Faculdade de Ciências da Universidade do Porto

Departamento de Ciência de Computadores

MSc. Data Science

Privacy-Preserving Face Detection: A Comprehensive Analysis of Face Anonymization Techniques

by Ricardo ANDRADE

The advancing capability of facial recognition technology, alongside the pervasive collection and exchange of facial data, raises substantial privacy concerns for individuals. While prior research has extensively examined privacy-preserving solutions in the 2D space, there is still a noticeable void in the development of such solutions for the 3D space. This thesis addresses the gap in the field by analyzing and proposing novel 3D face anonymization techniques for point clouds, which are collections of data points in 3D space representing the external surface of objects. It also conducts a comprehensive assessment to measure the effectiveness of these techniques in providing privacy protection while preserving data utility.

The methodological framework encompasses three pivotal components. First, a custom-made 3D facial dataset is curated, featuring a roster of 201 distinct identities. This dataset represents an expansion compared to several existing datasets and serves as the foundation for conducting experiments related to anonymization solutions. Second, the research proposes and implements six anonymization techniques spanning a diverse spectrum, encompassing paradigms that embrace sampling, noise injection, warping, smoothing, morphing, and point-level operations. Ultimately, a comprehensive evaluation of the anonymization techniques is conducted using a proposed evaluation methodology that measures their effectiveness by assessing the interplay between the level of privacy protection offered and the preservation of data utility. The privacy assessment involves rigorous testing against a robust attacker face recognition model, encompassing both verification and closed-set recognition modes. Simultaneously, the utility evaluation assesses its impact on face detection and a range of image quality assessment metrics.

The results indicate that all the anonymization techniques effectively conceal the identity of individuals. However, the smoothing technique based on the nearest neighbors algorithm and the face-morphing technique achieved the best compromise between data utility and privacy. Also, the evaluation methodology has proven its reliability in selecting the parameter configurations for these techniques and has the potential for further extension to assess the specific requirements of other use cases.

The findings of this study have implications for the potential integration of the proposed anonymization techniques in the field of 3D imaging technology. Moreover, this study expands the current knowledge base of 3D face anonymization, thereby laying the foundation for the development of more sophisticated methods.

Keywords— Face anonymization, Face detection, Point cloud, Privacy, Utility, Tradeoff

UNIVERSIDADE DO PORTO

Resumo

Faculdade de Ciências da Universidade do Porto
Departamento de Ciência de Computadores

Mestrado em Ciência de Dados

Deteção Facial com Garantias de Privacidade: Uma Análise Abrangente de Técnicas de Anonimização Facial

por Ricardo ANDRADE

A crescente capacidade da tecnologia de reconhecimento facial, associada à coleta e troca generalizada de dados faciais, levanta preocupações significativas de privacidade para os indivíduos. Embora estudos anteriores tenham examinado extensivamente soluções de preservação de privacidade no espaço 2D, ainda existe uma lacuna percetível no desenvolvimento de tais soluções para o espaço 3D. A presente dissertação aborda esta lacuna existente no campo, com a análise e desenvolvimento de técnicas inovadoras em nuvens de pontos, que correspondem a coleções de pontos no espaço 3D que representam a superfície externa de objetos. Por outro lado, também realiza uma avaliação abrangente dessas técnicas para determinar a sua capacidade de equilibrar a proteção da privacidade com a manutenção da utilidade dos dados.

O quadro metodológico abrange três componentes principais. Em primeiro lugar, é criado um conjunto de dados faciais 3D que contabiliza uma total de 201 identidades distintas. Este conjunto de dados representa uma expansão em comparação com vários conjuntos de dados existentes e serve como base para o desenvolvimento e avaliação das soluções de anonimização. Em segundo lugar, a pesquisa incide na conceção e implementação de seis técnicas de anonimização distintas que englobam um espectro diversificado, incluindo paradigmas que abrangem amostragem, injeção de ruído, distorção, suavização, "morphing", e operações ao nível dos pontos. Por fim, é realizada uma avaliação extensiva das técnicas de anonimização utilizando uma metodologia de avaliação proposta que mede a sua eficácia ao avaliar a interação entre o nível de proteção de privacidade oferecido e a preservação da utilidade dos dados. A avaliação de privacidade envolve testes rigorosos contra um modelo robusto de reconhecimento facial enquanto atacante, abrangendo os modos de verificação e reconhecimento em conjunto fechado. Simultaneamente, a avaliação de utilidade avalia o seu impacto na detecção facial e incorpora uma variedade de métricas de avaliação da qualidade da imagem.

Os resultados indicam que todas as técnicas de anonimização possuem a capacidade de ocultar a identidade do indivíduo. No entanto, a técnica de suavização baseada no algoritmo dos vizinhos mais próximos e a técnica de morfismo facial alcançaram o melhor compromisso entre a utilidade dos dados e a vertente da privacidade. A metodologia de avaliação também se mostrou competente na seleção das configurações de parâmetros destas técnicas e poderá potencialmente ser estendida para avaliar os requisitos específicos de outros casos de uso.

As conclusões deste estudo evidenciam que as técnicas de anonimização propostas oferecem garantias para a sua integração no domínio da tecnologia de imagens 3D. Além disso, este estudo amplia o conhecimento sobre a anonimização facial em 3D e fortalece as bases para o desenvolvimento de outros métodos mais sofisticados.

Palavras-chave— Anonimização facial, Deteção facial, Nuvem de pontos, Privacidade, Utilidade, Compromisso

Contents

List of Figures

List of Tables

Glossary

2D	Two-Dimensional
3D	Three-Dimensional
AAM	Active Appearance Model
AFW	Annotated Faces in the Wild
AP	Average Precision
ARS	Angular Radial Signature
AUC	Area Under Curve
B-PET	Biometric Privacy-Enhancing Techniques
BU-3DFE	Binghamton University 3D Facial Expression
CNN	Convolutional Neural Network
CMC	Cumulative Match Characteristic
CS-KDA	Class-Specific Kernel Discriminant Analysis
DCN	Deformable Convolution Network
DPM	Deformable Part-Based Model
DSLR	Digital Single Lens Reflex
FAR	False Acceptance Rate
FA	False Acceptance
FDDB	Face Detection Dataset and Benchmark
FID	Fréchet Inception Distance
FN	False Negative
FMR	False Match Rate

FNMR	False Non-Match Rate
FPS	Frames Per Second
FP	False Positive
FPR	False Positive Rate
FPFH	Fast Point Feature Histograms
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HOG	Histogram of Oriented Gradients
ICA	Independent Component Analysis
ICNP	Iterative Closest Normal Point
ICP	Iterative Closest Point
IJB-A	IARPA Janus Benchmark-A
IJB-B	IARPA Janus Benchmark-B
IJB-C	IARPA Janus Benchmark-C
IoU	Intersection over Union
KITTI	Karlsruhe Institute of Technology and Toyota Technological Institute
LBP	Local Binary Pattern
LDA	Linear Discriminative Analysis
LFW	Labeled Faces in the Wild
LiDAR	Light Detection and Ranging
MAE	Mean Average Error
MSE	Mean Squared Error
MRI	Magnetic Resonance Imaging
NN	Neural Network
PCA	Principal Component Analysis
PASCAL VOC	PASCAL Visual Object Classes Challenge
PII	Personally Identifiable Information

PMP	Point-Mesh-Point
PNG	Portable Network Graphic
PSNR	Peak Signal-to-Noise Ratio
RANSAC	Random Sample Consensus
ROC	Receiver Operating Characteristic
RoI	Region of Interest
R-CNN	Region-based Convolutional Neural Network
RPN	Region Proposal Network
RGB	Red Green Blue
S^{3}FD	Scale-invariant Face Detector
SBSR	Sparse Bounding Sphere Representation
SFR	Spherical Face Representation
SIFT	Scale-Invariant Feature Transform
SSH	Single Stage Headless
SSD	Single Shot Detector
SSIM	Structural Similarity Index Measure
SURF	Speeded-Up Robust Features
TAR	True Acceptance Rate
TA	True Acceptance
TPIR	True Positive Identification Rate
TP	True Positive
TR	True Rejection
VRML	Virtual Reality Modeling Language
WIDER	Web Image Dataset for Event Recognition

Chapter 1

Introduction

The growing capabilities of facial recognition technology, coupled with the extensive collection and dissemination of facial data, have emerged as considerable concerns regarding individuals' privacy. Beyond infringing upon the fundamental right to privacy, the unauthorized gathering of facial data also introduces the dangers of profiling, discrimination, and unwarranted surveillance, thereby disregarding ethical principles.

In May 2018, the General Data Protection Regulation (GDPR) took effect as a European Union regulation focused on safeguarding information privacy, encompassing provisions and requirements that govern the collection, utilization, processing, and storage of facial data. In adherence to privacy policies and principles, while safeguarding individual rights, extensive research efforts have been committed to the exploration of anonymization and pseudonymization techniques applied to facial data within images and videos. These techniques aim to strike a balance between preserving data utility and safeguarding individuals' privacy. Conversely, the domain of 3D data has garnered limited attention, with minimal to no research focused on addressing similar privacy concerns. This limitation stems from the current use of three-dimensional (3D) mapping technology, which lags behind its two-dimensional (2D) counterpart. Factors such as sensor resolution, limited demand, and the increased complexity associated with working with 3D data contribute to this disparity.

Nevertheless, with the escalating adoption of 3D imaging technologies, the imperative for robust privacy-enhancing solutions becomes accentuated. For example, the realm of autonomous driving has spurred the utilization of Light Detection and Ranging (LiDAR) technology, a key component within a car's perception system. LiDAR is a 3D mapping technology that captures the physical geometric properties of a scene by emitting pulsed laser beams that are reflected from objects and recaptured. The light's reflection time, also called the time of flight, is stored, as well as its intensity, which enables the exact computation of an object's distance. These sensors generate 3D data in the form of point clouds, as illustrated in Figure 1.1. Despite the current limitations in sensor resolution, the ongoing advancements in these sensors over the years highlight the importance of addressing this privacy issue, positioning it as a paramount challenge for the times ahead.

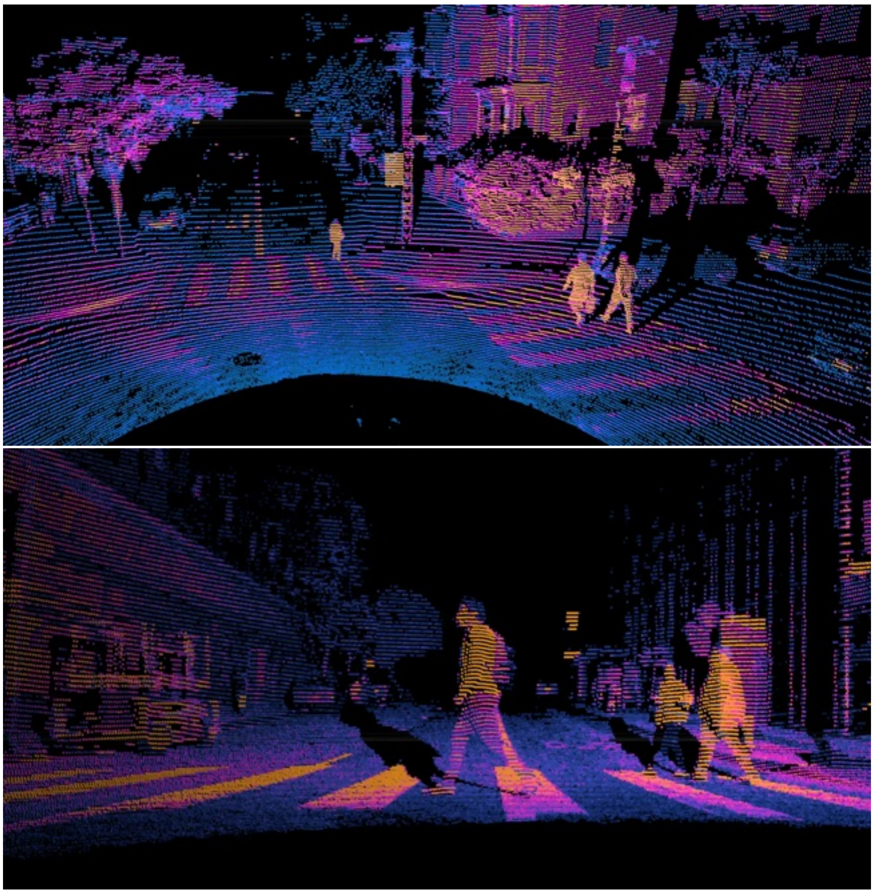


FIGURE 1.1: Point cloud captured by a LiDAR sensor in an autonomous driving setting. Extracted from OUSTER1.

Research Context This master's thesis is rooted in the innovative THEIA project, a collaborative effort between Bosch and the University of Porto. The core aim of THEIA is to strengthen the sensory capabilities of autonomous vehicles through the implementation and validation of perception algorithms. These algorithms leverage data from vehicle sensors, with a particular emphasis on LiDAR sensors, with the overarching goal of establishing a highly precise, resilient, and secure vision and perception system. This thesis aligns with SP5 (Subproject 5) within the project's scope, which focuses on Infrastructure and Security.

1.1 Objectives

In this context, this master's thesis delves into the constrained realm of 3D face anonymization, intending to propel advancements in this field. Specifically, the investigation is guided by the following objectives:

- Formulate novel face anonymization techniques designed for point cloud data;
- Conduct a comprehensive assessment of their inherent effectiveness, through implementation and evaluation of the privacy-utility trade-off of various anonymization techniques.

1.2 Contributions

This master's thesis bridges the acknowledged gap in the existing literature by addressing face anonymization for point cloud data. Consequently, the central contributions of this thesis encompass the following:

- A systematic literature review covering face detection, face recognition, and face anonymization across both 2D and 3D spaces;
- The development of a straightforward framework to curate a 3D facial dataset optimized for both 3D face analysis and anonymization;
- The implementation of innovative face anonymization techniques in point cloud data by integrating and combining pre-existing methods from other fields;
- A comprehensive and rigorous assessment of the effectiveness of the anonymizations, accompanied by a proposed methodology and metrics for effective privacy and utility analysis;
- An Artificial Intelligence-based solution for de-anonymization to test the effectiveness of the proposed anonymization solutions.

1.3 Document Structure

This thesis is organized into five additional chapters outlined as follows: Chapter 2 provides a comprehensive review of the existing literature related to face anonymization and the related topics of face detection and recognition in both 2D and 3D spaces. The key concepts, theories, and empirical findings are examined, providing the foundational context for this research; Chapter 3 outlines the research methodology employed in this study. The face anonymization techniques are proposed, and theoretical models that guide the analysis are outlined; Chapter 4 details the experimental setup, denoting the dataset and empirical configurations of the components under testing. The evaluation strategy is presented, providing a detailed explanation of how it was developed and executed, along with initial insights gleaned from the conducted experiments; Chapter 5 reports the results obtained regarding the anonymization techniques and provides a comprehensive analysis of them concerning the research questions; and at last, Chapter 6 draws conclusions about the work done, highlights the main findings, and discusses future work prospects.

Chapter 2

Related Work

The related work forming the foundation of this master thesis revolves around three pivotal components: face detection, face recognition, and face anonymization, with paramount emphasis on anonymization. The models for each component are designed to process facial information in various formats, including 2D data, such as still images, video frames, live video feeds, and any other digital format containing visual information or 3D data, encompassing depth images[1], point clouds, and meshes[2]. The reviewed literature covers both data types, 2D and 3D, particularly for digital images and point clouds. A digital image is a 2D array organized in rows and columns, represented as a matrix where individual elements correspond to pixels, which are the smallest units. A point cloud is a set of data points in a 3D coordinate system that represents the external surface of an object in the form of discrete points lying on the surface of the object. Each data point contains information about the position of a specific location in space, typically defined by its X, Y, and Z coordinates.

This chapter aims to present a comprehensive study of the interplay between face detection, face recognition, and face anonymization and highlight the essential aspects that characterize each field. The chapter starts by providing an overview of the relevant topics for each of the three areas concerning the algorithms, as well as datasets and evaluation

[1] A depth image is a 2D representation of a scene where each pixel encodes the distance from the camera to the corresponding point in the scene. Despite representing 3D information, they are also referred to as 2.5D due to their 2D nature.

[2] A 3D mesh is composed of vertices, edges, and faces, collectively defining a 3D object. Vertices represent points in 3D space, edges connect adjacent vertices, and faces (or polygons) enclose these edges, defining the object's surface.

metrics. Accordingly, the chapter analyzes face detection, recognition, and anonymization, particularly addressing those topics, and presents the current state-of-the-art results for reference.

2.1 Overview

Before delving into the specifics, an introduction of the interdisciplinary elements that will be addressed regarding each field is provided. The division is aligned with the reviewed literature, enabling a presentation of the most relevant information for characterization and paving the way for a deeper exploration. The following elements will be discussed in the remaining chapter:

Algorithms Given the rapid advances in face detection, recognition, and anonymization, it is virtually impossible to cover all methods entirely. Therefore, this topic highlights pioneering works and milestone models in the history of these fields while exploring recent trends. The milestones in the three areas can be categorized into two eras: the *Traditional Methods* (pre-Deep Learning) and the *Deep Learning-based Methods* (post-Deep Learning) [1]. While both eras are discussed, particular emphasis is placed on the latter as these works have demonstrated substantially better performance and have dominated the research field in recent years [2].

The *Traditional Methods* in computer vision rely on handcrafted feature extraction techniques to design and extract feature representations that may involve edge detection, corner detection, or threshold segmentation [2]. These features are defined as descriptive or informative patches in the data, providing fragments of information about its content. Until around 2015, these methods were considered the standard approaches [3]. However, a paradigm shift prompted the widespread adoption of Deep Learning solutions.

The *Deep Learning-based Methods*, on the other hand, leverage Deep Learning algorithms to automatically learn and generalize complex discriminative facial features without human supervision [4]. Deep Learning is a subset of a broader family of Machine Learning methods grounded in Artificial Neural Networks [5]. Within the Deep Learning community, Convolutional Neural Networks (CNNs) are some of the most successful and widely used architectures [6], serving as essential building blocks for various facial analysis approaches. While their widespread popularity began in 2014 due to the availability of large training datasets [7] and the accessibility of high-performance graphics processing units (GPUs) [8], they have been employed in face detection since as early as 1994 [9].

Evaluation and Performance Metrics This topic addresses commonly employed performance metrics in the research community within each field.

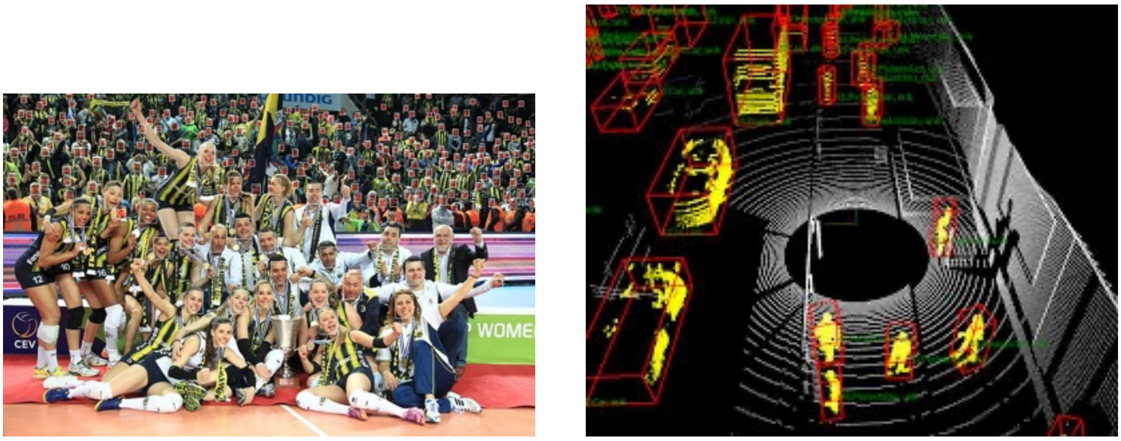
Datasets and Benchmarks The advent of *Deep Learning-based Methods* has emphasized the importance of large-scale data for model training, significantly impacting model performance. Wang and Deng [10], and Yi *et al.* [11] claim that developing large-scale face databases has become a leading factor in face recognition research progress. Similar trends can be observed in face detection and anonymization. Various datasets are publicly available for training and benchmarking algorithms in face detection and recognition, categorized as constrained/controlled and unconstrained/uncontrolled, according to their characteristics. Constrained datasets are created under controlled conditions, allowing for the study of specific model parameters [12]. However, they lack the complexity and variability of real-world scenarios. In contrast, uncontrolled or "in-the-wild" datasets reflect real-world complexities, containing more images and diverse factors, such as extreme poses, lighting variations, occlusions, low resolution, and facial expression variations. The unconstrained scenarios avoid metric saturation, enabling the exploration of model strengths and real-world application utility.

Benchmarks provide standardized evaluation protocols and metrics on datasets to assess the performance of these computer vision algorithms. They offer a common ground for assessing progress, comparing methods, identifying limitations, and advancing research boundaries.

This topic provides an overview of the primary datasets within each field, accompanied by benchmark results.

2.2 Face Detection

Object detection is a computer vision technique that predicts the location and class of objects in the input data, empowering a computer with the knowledge of *What objects are where?* [1]. Face detection is a specific subset of object detection focusing exclusively on identifying and locating an unknown number of human faces. As a result, the algorithms are designed to be more specialized in recognizing facial features and distinguishing them from other objects and backgrounds. Nevertheless, face detection algorithms mostly follow the approaches of generic object detection [13]. Face detectors determine the pixel-wise and point-wise coordinates of the faces with bounding boxes defined as the smallest 2D rectangle-shaped or 3D box-shaped structure surrounding the entire human face, irrespective of its shape and occlusion degree. The 2D boxes are axis-aligned, whereas the 3D boxes are oriented. Refer to Figure 2.1 for a visual representation of the face detection output.



(A) 2D face detection

(B) 3D object detection.

Figure 2.1: Output of a 2D face detector and 3D object detector. Extracted from RetinaFace and [14], respectively.

Face detection is the stepping stone to all facial analysis algorithms [15], including face anonymization and recognition. The accurate identification and isolation of the facial regions serve as the foundation for subsequent stages of the pipelines that focus specifically on analyzing and processing facial features and attributes. Thus, accurate detection of human faces significantly impacts the performance and overall effectiveness of the entire face analysis process.

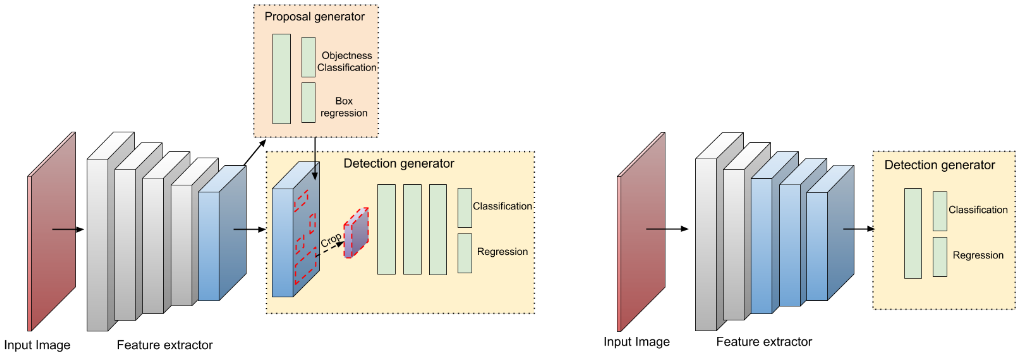
2.2.1 Algorithms

The academic community has extensively explored 2D face detection, with major surveys analyzing the evolution and trends in this research area. Hjelmås and Low [16] outlined the efforts within the face detection field dating back to the beginning of the 1970s until the early 2000s in their acknowledged survey. Zhang and Zhang [17] reviewed the advances in the field over the following ten years, covering essential feature extraction and learning algorithms. Minaee *et al.* [3] covered the rapid progress in this field from the beginning of the Deep Learning wave in 2015 until 2021, including more than fifty methods. Kumar *et* *al.* [18] presented a comprehensive survey of face detection techniques in digital images, focusing on the traditional approaches and their evolution.

In contrast, the 3D domain has been relatively underexplored, resulting, to the best of the research efforts, in no surveys on this topic.

2D Face Detection Regarding the *Traditional Methods*, the Viola-Jones face detector [19] is a milestone algorithm proposed in 2001 that revolutionized real-time face detection with comparable detection accuracy to contemporaneous algorithms. The detector uses a sliding window to search for Haar-like features on the input image, which are efficiently calculated with an integral image. The Adaboost algorithm finds the best features with an attentional cascade structure. In 2005, the Histogram of Oriented Gradients (HOG) [20] was proposed and led to considerable improvements compared to other algorithms of the time. The detector divides the image into grids and computes the gradients of pixel intensities and their orientations to generate feature representations. The HOG features are fed into a linear Support Vector Machine (SVM) classifier for detection. According to Zafeiriou *et al.* [21], this approach inspired the use of robust descriptors, such as Scale-Invariant Feature Transform (SIFT) [22] and Speeded-Up Robust Features (SURF) [23], with weak classifiers. The Deformable Part-based Model (DPM) [24] was initially proposed in 2008 as a detector that incorporates flexible parts and models their spatial relationships within an object. The approach decomposes objects into parts and learns their appearance and deformation properties. The model uses a hierarchical structure and combines deformable part templates to capture local and global contexts. Other DPM-based models proposed by Felzenszwalb *et al.* [25, 26], have achieved great results. Despite the significant improvements of handcrafted algorithms, these methods rely on the robustness of handcrafted features and have sub-optimal component optimization [27]. Furthermore, their performance became saturated [1], and a paradigm shift occurred with the rebirth of CNNs and advancements brought by Deep Learning [28].

Concerning *Deep Learning-based Detectors*, there are two groups of face detectors inspired by object detection: two-stage detectors and single-stage detectors [29]. The former adopts a two-stage process that sequentially finds an arbitrary number of object proposals and then classifies and localizes them. In contrast, the latter classifies and localizes the objects in a single shot - refer to Figure 2.2 for a visual representation of the architecture of both groups. Generally, one-stage detectors prioritize fast inference time, whereas two-stage detectors excel in accuracy [30]. Specifically, a comparison reveals the following advantages of two-stage detectors: 1) The region proposal in two-stage detection effectively addresses class imbalance by filtering out most of the negative proposals [31]; 2) the two-stage process focuses on a smaller number of proposals, allowing for larger detection heads and the extraction of richer features; and 3) two-stage detection regresses the object location twice, resulting in more precise box localization. For a more comprehensive comparison, refer to [32], while a detailed bibliometric analyses of both groups can be found in [33].



(a) Two-stage Faster R-CNN

(b) One-stage RetinaNet

FIGURE 2.2: Deep Learning architectures of one-stage and two-stage object detectors. Extracted from [34].

Among the two-stage models, the Region-based Convolutional Neural Network (R-CNN) [7] stands as a pioneering member of the R-CNN family, introduced in 2014. The model generates region proposals called Regions of Interest (RoI) with Selective Search [35], and a pre-trained CNN model extracts features from these regions. The detection leverages a linear SVM to predict the presence and location of the objects within each region. Despite the performance improvement at the time, it has drawbacks related to slow inference time, unnecessary feature computations on numerous overlapping proposals, and inefficient handling of objects of varying scales. In 2015, Fast R-CNN [36] addressed the limitations of its predecessor and introduced a unified framework that combines region proposal generation and feature extraction, significantly speeding up training and inference time. In the same year, Faster R-CNN [37] pushed closer to real-time inference by introducing a Region Proposal Network (RPN), making the region proposal generation more efficient. Although originally designed for object detection, the above models, especially Faster R-CNN, considered the most classical anchor-based generic object detection method [29], were inherited by face detection [27]. For instance, works like

[9, 38–40] consist of slight variations of Faster R-CNN adapted to the face detection problem.

In the scope of one-stage models, the Single Shot Detector (SSD) [41] emerged in 2016 as a pioneer single-shot object detector demonstrating comparable performance to two-stage detectors [42]. The main contribution was incorporating multiple feature maps with different resolutions from different layers to detect objects of various scales and the discretization of the output space with a set of reference bounding boxes that are adjusted during prediction time. However, it struggles with detecting small objects due to the default anchor reference designs, a common problem of anchor-based detectors whose performance significantly deteriorates as objects become smaller [43]. This model exerted some influence on subsequent works within the field of face detection. In 2017, Single Shot Scale-invariant Face Detector (S^{3}FD) [27] introduced a framework that can handle different scales of faces, improving the detection of small faces. The model has extra convolutional layers to generate more anchors for small faces and reduces the stride sizes to increase potential matches with more small-scale faces. Other anchor strategies, such as FaceBoxes[44] and ScaleFace [45], effectively address small objects and multi-scale face detection. Single Stage Headless (SSH) [46] was introduced in 2017 as a headless single-stage detector avoiding the computation of the parameters of the fully connected layers. The model is scale-invariant by design and detects faces from the early convolutional layers, making it fast and lightweight. More recently, in 2020, TinaNet [47] achieved state-of-the-art results in face detection by only using models and techniques constructed on pre-existing object detection modules. Zhu *et al.*, the authors, claim that there is no gap between face detection and generic object detection.

3D Face Detection While some papers have addressed the 3D face detection problem [48–50], they are limited to the facial data only comprising the region above the neck. On the other hand, 3D object detection has been extensively reviewed, especially in autonomous driving, and the models operate on data that captures various aspects of the driving environment, including other vehicles, vegetation, cars, pedestrians, and more. Given the considerable difference between current 3D object detection and 3D face detection regarding the complexities of the data they operate, they are not depicted here.

2.2.2 Evaluation and Performance Metrics

Numerous datasets commonly employ rectangular bounding boxes as ground truth, encoded by the pixel coordinates of their upper-left corner, height, and width [51–53]. However, there is no consensus regarding the most suitable shape [54]. In most benchmark evaluations, a correct detection is determined using the Intersection over Union (IoU) metric. This metric calculates the overlapping area between the predicted bounding box B_{pred} and the ground-truth bounding box B_{truth}, divided by the area of their union:

$$IoU = \frac{\text{area}(B_{pred} \cap B_{truth})}{\text{area}(B_{pred} \cup B_{truth})} \tag{2.1}$$

Comparing the IoU value to a given threshold $\theta \in \mathbb{R}$ allows for classifying detections as correct if IoU $\geq \theta$, or incorrect if IoU $< \theta$. The commonly adopted threshold value is $\theta = 0.5$ [51–54], although different values can be chosen based on the specific application and requirements. The analysis of correct detections allows for the evaluation of a detector's performance using various metrics, which are computed based on the fundamental concepts defined below:

- *True Positive (TP):* The system correctly detects a face;
- *False Positive (FP):* The system incorrectly detects a non-existent face or inaccurately detects an existing face;
- *False Negative (FN):* The system fails to detect a face, leaving it undetected.

The concept of a True Negative (TN) is impractical in face detection, as it would require correctly identifying all non-face regions within an image, a task practically infeasible due to their vast number. Hence, evaluation metrics that rely on TN, such as the FPR and the standard Receiver Operating Characteristic (ROC) curve of binary classifier systems, cannot be used.

However, benchmarks in face detection propose an adaptation of the ROC curve that does not rely on the TN computation. A standard ROC curve is created by plotting the True Positive Rate (TPR) on the y-axis against the False Positive Rate (FPR) on the x-axis. Therefore, benchmarks like [54] replace the x-axis with the total number of FP, and [55] use the number of FPs per image. The measurement of the Area Under Curve (AUC), obtained through the integral of the ROC curve, is also used to quantify and summarize the entirety of the ROC curve.

Also, to avoid the computation of the TN value, the benchmark's assessment is primarily grounded on Precision and Recall, which are mathematically defined as:

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.3}$$

Precision measures the quality of the identified faces (positive predictions) made by the model, whereas Recall measures the model's ability to detect faces (relevant instances). These two metrics are inversely related, and evaluating a precision-recall curve provides valuable insights into this trade-off. A well-performing detector will consistently exhibit high Precision and Recall, regardless of variations in the confidence threshold. Thus, the AUC effectively summarizes and evaluates this relationship. However, computing the AUC requires additional processing due to the irregular shape of the precision-recall curve. The Average Precision (AP) metric is used as an alternative way to calculate the AUC [56], addressing the issue. The AP is the mean Precision calculated across multiple Recall values ranging from 0 to 1. One common approach, as discussed in [57], involves using an 11-point interpolation, where the precision-recall curve is summarized by averaging the maximum precision values at 11 equally spaced recall levels $\{0, 0.1, ..., 0.9, 1\}$. Mathematically, it can be defined as follows:

$$\text{AP} = \frac{1}{11} \sum_{r \in \{0,0.1,...,0.9,1\}} p_{interp}(r) \tag{2.4}$$

where the Precision at each recall value is interpolated given by:

$$p_{interp}(r) = max_{\tilde{r} \geq r} p(\tilde{r}) \tag{2.5}$$

Another performance indicator is the Frames Per Second (FPS), which quantifies the runtime efficiency of detectors by measuring the number of frames processed per second. This metric holds significant importance in determining the suitability of detectors for real-time applications.

2.2.3 Datasets

2D Face Detection Over the past few years, numerous datasets have emerged as valuable resources for face detection research. These datasets have been instrumental in training and evaluating face detection algorithms, serving as benchmarks for performance assessment. Table 2.1 presents the specifications of some of the most prominent and widely used face detection datasets, including significant details such as the number of images, the count of annotated faces, the image sources, and the bounding box shapes. Additional information is provided below for each dataset. While these datasets have primarily been designed for face detection tasks, other facial analysis datasets may also be employed for the intent of face detection.

TABLE 2.1: Datasets used for training and testing 2D deep face detection.

Name	Year	Images	Faces	Source	Bounding box
FDDB [54]	2010	2 845	5 171	Yahoo's news articles	Elliptical
AFW [51]	2012	205	468	Flickr	Rectangular
PASCAL FACE [52]	2014	851	1 335	PASCAL VOC	Rectangular
WIDER FACE [53]	2016	32 303	393 703	WIDER	Rectangular

All the presented datasets are used for testing, except for the WIDER FACE [53], which serves both training and testing purposes.

AFW The Annotated Faces in the Wild (AFW) [51] dataset was introduced in 2012 to simultaneously address face detection, pose estimation, and landmark localization. It contains 205 images from Flickr, with 468 annotated faces. Each face is labelled with a rectangular bounding box and up to six landmarks, including the centre of the eyes, the tip of the nose, and the two corners and centre of the mouth. The dataset showcases diverse backgrounds and exhibits substantial facial viewpoints and appearance variations.

The benchmark used in the AFW dataset follows the same approach as the PASCAL Visual Object Classes Challenge (PASCAL VOC) dataset [57], a widely used dataset in computer vision research.

FDDB The Face Detection Dataset and Benchmark (FDDB) [54] dataset was introduced in 2010, offering a larger number of faces with more accurate face region annotations. It also proposes a standardized protocol for evaluating the performance of face detection algorithms. The dataset consists of 2 845 grayscale and colour images collected from news articles on the Yahoo website, featuring various resolutions and containing 5 171 annotated faces. Instead of using typical bounding boxes, the faces are manually annotated using elliptical regions, as the author claims that ellipses provide a more accurate specification of the face region without increasing the number of parameters. The dataset covers a range of challenging scenarios, including difficult pose angles, out-of-focus faces, and low resolution.

The benchmark employed in the FDDB dataset considers the correspondence between the set of detections and annotations as a maximum weighted matching in a bipartite graph, computing two distinct metrics accumulated in an ROC curve.

PASCAL FACE The PASCAL FACE [52] dataset was introduced in 2014 along with a face detection method, with limited information regarding its specifications. The dataset focuses on face detection and recognition and comprises 851 images with 1 341 annotated faces, which exhibit limited variations in appearance. These images were collected from the Pascal Person Layout test set, a subset of the larger-sized PASCAL VOC dataset [57] created in 2010. Similar to FDDB [54], PASCAL FACE is commonly used as a test set only. The evaluation metric used is the AP with an IoU threshold of 0.5.

The benchmark employed in the PASCAL FACE dataset follows the same approach as the PASCAL VOC dataset [57].

WIDER FACE The Web Image Dataset for Event Recognition (WIDER FACE) [53] was introduced in 2016 to bridge the gap between real-world requirements by introducing a high degree of variability in scale, pose, and occlusion. It contains 32 203 images with 393 703 labelled faces, which Yang *et al.*, the authors, claim was ten times larger than existing datasets. The images were selected from the publicly available Web Image Dataset for Event Recognition (WIDER) [58], created in 2015. It comprises three subsets: 40% allocated for training, 10% for validation, and the remaining 50% for testing. The images in the dataset are divided into three splits based on detection difficulty levels: Easy, Medium, and Hard, which are determined based on the detection rate achieved by the EdgeBox [59] method for object proposals. Detection difficulty is associated with varying degrees of face scale, occlusion, and pose.

The benchmark employed in the WIDER FACE dataset follows a similar approach to the PASCAL VOC dataset [57].

3D Face Detection While various 3D object detection datasets exist, such as KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) [60], nuScenes [61], and Waymo Open Dataset [62], they are not suitable for face detection due to their limited resolution, which hinders the capture of fine facial features. As a result, the development of sensors with sufficient resolution remains a significant challenge for advancing the 3D face detection field. Furthermore, there is a need for affordable 3D acquisition systems capable of generating large volumes of data required to handle the inherent complexity of representing and processing 3D data [63].

2.2.4 Benchmarks

Table 2.2 presents the performance of some of the leading models on the WIDER FACE benchmark, the most commonly used benchmark for evaluating face detection algorithms [64]. When the dataset was introduced, state-of-the-art face detectors achieved values lower than 32% on the hard difficulty split [53], a difference higher than 60% from current models, illustrating the significant development in this field over time. Currently, the model performances on the FDDB [54] and PASCAL FACE [52] datasets exceeds 99% [3], indicating that these datasets have reached their saturation point and are no longer able to effectively evaluate the quality of new models.

TABLE 2.2: State-of-the-art models for 2D face detection, on the WIDER FACE benchmark.

Method	Easy	Medium	Hard
FACE R-FCN [65]	94.3	93.1	87.6
SRN [66]	95.9	94.8	89.6
FDNet [67]	95.0	93.9	89.6
DSFD [68]	96.0	95.3	90.0
PyramidBox [69]	95.6	94.6	90.0
AInnoFace [70]	96.5	95.7	91.2
RetinaFace [29]	-	-	91.4
TinaFace [47]	-	-	92.4

2.3 Face Recognition

Biometrics encompasses biological measurements and distinctive physical attributes for individual identification [71]. Face recognition is a specific biometric technology that relies exclusively on analyzing facial characteristics for identity verification or identification purposes. Face recognition technology finds extensive applications across numerous fields, including security and surveillance, healthcare, banking, and retail.

Face recognition can be categorized into two modes according to its application. On one side, face verification (or 1:1 matching) consists of verifying an identity claim through a face image and comparing its facial attributes with the stored facial data associated with the claimed identity. The system performs a one-to-one comparison, either accepting or rejecting the claim. Conversely, face identification (or 1:N matching) involves determining an individual's identity by comparing their face image with a database of known identities. The system performs a one-to-many comparison, assigning the identity label associated with the closest match. Thus, recognition is a generic term that encompasses both verification and identification. The basic workflow of an end-to-end face recognition system involves three key components [8]:

1. *Face detection:* This component detects all faces in the input image and provides the corresponding coordinates of the bounding boxes.

2. *Face alignment:* After detecting the faces, this step normalizes them to a canonical view, accounting for pose, scale, and expression variations to facilitate subsequent tasks.

3. *Feature representation:* In this stage, various discriminative facial features are extracted from the aligned faces. These features are designed to map the aligned face images into a feature space where features of the same identity are closer together while those of different identities are far apart.

In the literature, the final step of face recognition is often known as *Feature Matching* [72], generating the ultimate predictions for verification and identification. Refer to Figure 2.3 for a visual representation of the general pipeline of a face recognition model.

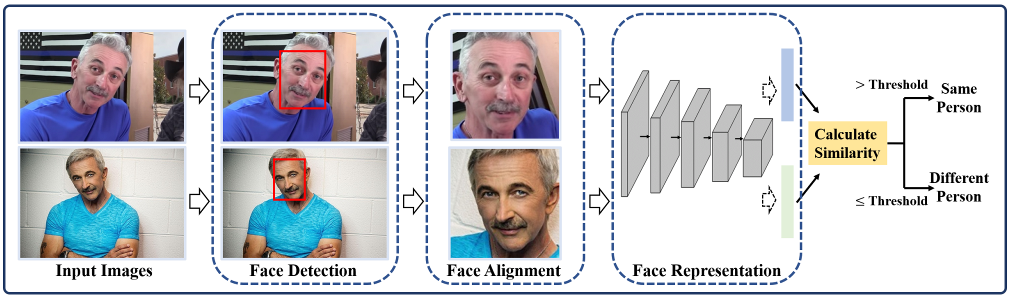


FIGURE 2.3: The standard pipeline of an end-to-end deep face recognition system. Extracted from [73].

2.3.1 Algorithms

Similar to 2D face detection, extensive research has been conducted by the academic community on 2D face recognition. Jafri *et al.* [74] and Zhao *et al.* [75] provide comprehensive surveys on the Traditional Methods employed in face recognition. Recent advancements in *Deep Learning-based Methods* have attracted significant attention, leading to studies focusing on specific pipeline elements. For instance, Wang *et al.* [76] and Jin *et al.* [77] present exhaustive surveys on face alignment techniques, while Masi *et al.* [15] focus on the representation of facial features. The preceding chapter has already outlined existing studies on face detection, which serves as the initial stage of the pipeline. Additionally, Du *et al.* [73] offer a comprehensive review of the end-to-end face recognition problem, highlighting recent advances across the entire pipeline. Adjabi *et al.* [78] present the evolution of face recognition technology, reviewing 180 publications from 1990 to 2020, with a particular focus on the current 2D face recognition state-of-the-art, namely *Deep Learning-based Methods*, although also reviewing 3D models.

In the 3D domain of facial recognition, research has developed, although the number of surveys remains somewhat limited when compared to their 2D counterparts. Conducted by Li et al. [79], the survey offers an extensive overview that spans both traditional and contemporary methods. It delves into the associated limitations and advantages of these techniques, specifically focusing on aspects like face processing, feature extraction, and the classification of recognition methods. The survey also delves into the challenges encountered in 3D facial recognition. Addressing some of these challenges, Zhou and Zhang [80] and Zhang et al. [81] have contributed to the survey landscape, placing particular emphasis on the complexities introduced by facial expressions, occlusions, and pose variations. For a broader perspective encompassing multi-modal approaches, Bowyer et al. [82] have provided a survey that covers the approaches and challenges in 3D face recognition and integrates the fusion of 3D and 2D information.

2D Face Recognition The Traditional face recognition methods can be classified into three distinct approaches: *Holistic Methods*, *Feature-based Methods*, and *Hybrid Methods*, as outlined in [83] and [10], with initial efforts in the field dating back to 1966 [84].

Holistic Methods utilize the entire face as input and consider global facial information for face recognition. These methods encode the global information by extracting a small set of features from the pixels, capturing the variations among different faces, and enabling the unique identification of individuals [85]. EigenFaces [86] was introduced in 1991 as a groundbreaking method using statistical Principal Component Analysis (PCA). This method aims to identify the principal components of the face image distribution by calculating the eigenvectors of the covariance matrix. These eigenvectors, known as eigenfaces, collectively describe the variations between face images, allowing for a precise representation of each face within a lower-dimensional feature space than the original data. In 1997, Fisherface [87] followed a similar principle of similarity as EigenFaces. However, instead of using PCA to reduce the high dimensionality of the image space, Fisherface employs Linear Discriminative Analysis (LDA). A version of Independent Component Analysis (ICA) was used in a study by Bartlett *et al.* [88] in 2002, which is a generalization of PCA. These methods extract a low-dimensional face representation based on certain distribution assumptions and employ linear techniques to represent the subspace. However, nonlinear techniques such as Kernel Principal Component Analysis (KPCA) [89], Locality Preserving Projections (LPP) [90], and Class-specific Kernel Discriminant Analysis (CS-KDA) [91] are also employed. Nonlinear methods often leverage kernel techniques, which involve mapping the data into a higher-dimensional space where linear algorithms are well-suited for efficiently modelling complex relationships.

In contrast to the previous approaches, *Feature-based/Local Methods* involve extracting and analyzing specific facial features (e.g., eyes, mouth, nose) at different locations in a face image to identify individuals [92]. These methods aim to discover distinctive features within facial regions and match them across the entire image. They demonstrate higher robustness than *Holistic Methods* when dealing with variations in facial expressions, illuminations, and occlusions [92]. One feature-based method is Local Binary Pattern (LBP) [93], which was introduced in 1996, significantly influencing the development of other methods in the field [94, 95]. LBP leverages the extraction of local texture patterns from facial images and represents them using histograms. The image is divided into small local regions, and binary codes are generated based on pixel intensity comparisons with the center pixel of each region. These codes are then used to construct concatenated local histograms, which characterize the distribution of local patterns within the face. The development of local feature descriptions in computer vision has also found applications in the recognition problem [96–98].

Hybrid Methods combine the strengths of both *Holistic Methods* and *Feature-based Methods*, either in a serial or parallel manner, to overcome the limitations of individual approaches [99]. These methods involve extracting local features such as LBP or SIFT and then projecting them onto a lower-dimensional and discriminative subspace using techniques like PCA or LDA [92].

Deep learning Methods for face recognition rely on two essential factors, apart from the training data: the backbone CNN architecture and the loss function [92]. They determine how well the network can extract discriminative features from facial images and encourage the network to learn meaningful feature representations for recognition by defining the objective that the network is trained to minimize. These two factors have enabled these methods to become state-of-the-art [100], surpassing the distinctiveness and compactness limitations of handcrafted methods [10]. The main deep face recognition methods benefit from the advancement of general architectures used in visual recognition tasks, which serve as the foundation for more specialized architectures for face recognition. Some of these architectures include:

- AlexNet [28]: Winner of the 2012 ImageNet image classification competition [101], it consists of five convolutional layers, some followed by a max-pooling layer, and three fully connected layers with a 1000-way softmax;
- VGGNet [102]: Introduces smaller convolutional kernels of size 3x3, increasing the depth of the network with configurations ranging from VGG-16 and VGG-19, with 16 and 19 layers;
- GoogleNet[1] [103]: Winner of the 2014 ImageNet, it is a deep and wide architecture with 22 layers that uses inception modules to aggregate information from different spatial scales efficiently;

[1] Also known as Inception-V1.

- ResNet (Residual Network) [104]: Winner of the 2015 ImageNet, it introduces a residual learning framework with "shortcut connections", enabling the training of deeper networks with hundreds of layers, addressing the vanishing gradient problem[2].

These architectures have been optimized and adapted for face recognition since 2014 when DeepFace [105], developed by Facebook researchers, achieved an approaching human-level performance. Inspired by Alexnet, the model was trained on four million facial images using 3D shape modelling to align the faces and a nine-layer deep neural network for face representation. Recent models tend to use ResNet backbones to achieve even higher performance levels.

The loss function is crucial in enhancing feature discrimination, intending to promote intra-class compactness and inter-class separability [106]. Here, the classes refer to the subjects' identities, such that faces belonging to the same identity have the same label and belong to the same class, whereas distinct identities do not. Researchers have focused on re-designing classical classification loss functions like Softmax to achieve this goal. Some well-known loss functions include Center Loss [107], Contrastive Loss [108], and Triplet Loss [109] that embed the images into the Euclidean space, or Large Margin Loss [110], A-Softmax [111], ArcFace [112], and ElasticFace [113] that leverages angular separability between different classes.

3D Face Recognition The 3D face recognition methods follow a similar categorization to the 2D approaches. They can also be divided into *Holistic Methods* and *Feature-based Methods* following the same principles.

In the realm of *Holistic Methods*, the Iterative Closest Point (ICP), initially introduced by Besl and McKay [114], stands as a foundational iterative algorithm employed for the alignment of point clouds. This alignment process allows the computation of matching errors between two point cloud sets, which can be subsequently leveraged for recognition purposes. Nevertheless, it exhibits limitations when confronted with facial expressions characterized by non-rigid deformations. The ICP algorithm has served as a spur for the development of subsequent methodologies in the field. For instance, Li *et al.* [115] incorporated the Hausdorff distance in a central profile alignment step preceding the application of ICP, developing an efficient rejection classifier. Similarly, Mohammadzade

[2] Phenomenon that occurs during DNN training in gradient-based optimization algorithms.

and Hatzinakos [116] introduced the Iterative Closest Normal Point (ICNP), leveraging the higher discriminative information encoded within surface normal vectors compared to the point coordinates.

In the domain of *Feature-based/Local Methods*, these approaches are primarily applied to 3D depth images or meshes. However, a number of approaches center on the identification and characterization of 3D keypoints, as exemplified by Emambakhsh and Evans [117], who focused on the nasal region and identified seven distinct keypoints. From these keypoints, they extract nasal surface normals using Gabor-wavelet filtered depth maps. These surface normals serve as the foundation for generating a set of curves and spherical patches, which, in turn, are employed as descriptors for the facial representation. Beyond keypoints, other methods within this category introduce surface curves as a means of representing the facial structure. Drira *et al.* [118] employ radial curves as a feature representation, while Lei *et al.* [119] leverage Angular Radial Signatures (ARS).

The *Deep Learning* methods have seen advancements, driven by innovative architectures capable of directly processing point clouds. One example is PointNet [120] which was initially designed for object classification and part segmentation. The model learns point-wise features independently using Multilayer Perceptron (MLP) layers and extracting global shape features through max-pooling operations, functioning as symmetric functions. What sets PointNet apart is its ability to learn the original geometry of unordered point clouds, achieved through permutation-invariant operators. However, it does not account for local spatial relationships within the data. In response to this limitation, PointNet++ [121] emerged as a successor. PointNet++ enhances feature quality by enabling the network to learn local structures in point clouds at various scales. These architectural innovations have not only influenced 3D face recognition but have also made significant contributions to 3D object detection, for example. Bhople *et al.* [122] introduced PointNet-CNN, a deep CNN based on PointNet that employs a PointNet-based module for feature extraction and incorporates a Siamese network to compare extracted features and perform classification. Another work is FR3DNet [123], a deep CNN model inspired by VGG-Face [124] that embeds facial data into feature vectors of length 1 024. Unlike traditional 2D face recognition kernels with small filter sizes, FR3DNet introduces filters with larger kernel sizes. To train this model, the authors proposed novel data augmentation techniques involving the generation of synthetic identities through non-linear transformations in facial expressions and interpolations between identities.

2.3.2 Evaluation and Performance Metrics

Before delving into the evaluation, it is important to introduce some basic terminology and notation:

- *Gallery set:* The gallery consists of a collection of enrolled reference faces with a known identity in the system;
- *Probe set:* The probe set denotes a set of faces used to perform the recognition tasks.

The two face recognition applications, verification and identification, influence the choice of performance evaluation methodologies, performance statistics, and visualization charts.

Verification A face recognition system determines whether to accept or reject a probe face by comparing its feature representation with the feature representation of a gallery face using a distance or similarity measurement, given a predefined threshold. The decision of the system allows the computation of the following values:

- *True Acceptance (TA):* The system identifies a genuine match between a gallery face and a probe face, leading to the correct acceptance;
- *False Rejection (FR):* The system fails to identify a genuine match between a gallery face and a probe face, leading to an incorrect rejection;
- *True Rejection (TR):* The system accurately identifies no genuine match between a gallery face and a probe face, leading to a correct rejection decision;
- *False Acceptance (FA):* The system identifies a match between a gallery face and a non-matching probe face, leading to erroneous acceptance.

These concepts share similarities with those introduced in the evaluation of face detection but are specifically applied in the context of face recognition using the appropriate terminology. Refer to [125] for more details and in-depth concepts. Consequently, confirming a user's claimed identity involves two types of errors: false acceptance and false rejection. As such, the False Acceptance Rate (FAR) or False Match Rate (FMR) denotes the probability of mistakenly accepting as a match a non-matching face, i.e., an imposter. On the other hand, the False Rejection Rate (FRR) or False Non-Match Rate (FNMR) represents the likelihood of the system inaccurately rejecting a genuine identity match. Mathematically, these two metrics are defined as follows:

$$FAR = \frac{FA}{FA + TR} \tag{2.6}$$

$$FRR = \frac{FR}{FR + TA} \tag{2.7}$$

Another common metric is the True Acceptance Rate (TAR), which represents the probability of correctly accepting a genuine match. Mathematically is defined as:

$$TAR = \frac{TA}{TA + FR} = 1 - FRR \tag{2.8}$$

In the biometrics context, the ROC is a graphical representation that plots the FRR on the y-axis against the False Acceptance Rate (FAR) on the x-axis achieved by varying the threshold used for verification. Alternatively, the ROC curve can also depict the TAR against the FAR. The AUC of the ROC curve is a widely used metric to evaluate the overall performance of face verification tasks.

Identification Identification can be categorized into two evaluation protocols: closed-set and open-set. In the closed-set protocol, the system assumes that all identities in the probe set are included in the gallery set. On the other hand, the open-set protocol is designed to handle probe identities not present in the gallery set, representing a more realistic and complex scenario. This protocol must address the challenge of rejecting unknown identities [126]. As a result, the closed-set protocol is a more straightforward and less realistic protocol when compared to the open-set.

In a closed-set, the performance evaluation involves computing the rank-k performance. This evaluation involves calculating the distance or similarity scores between a specific probe and the gallery elements. The scores are then sorted in descending order[1], and the rank of the true match in the sorted list is determined. The identification rate for a given rank-k, denoted as IR(k), represents the proportion of probes at rank-k or lower. Mathematically, it can be expressed as follows:

$$IR(k) = \frac{|\{b \mid rank(b) \leq k, \forall b \in \mathcal{P}\}|}{|\mathcal{U}_{\mathcal{P}}|} \tag{2.9}$$

[1] or increasing order depending on the use of distance or similarity measures.

where e $\mathcal{U}$ represents the set of unique identities, and $\mathcal{P}$ is the probe set. The Cumulative Match Characteristic (CMC) curve plots the Identification Rate (IR) at rank-k on the y-axis against the value of k. Rank-1 performance is commonly used to summarize the closed-set identification performance [127].

In an open-set scenario, the commonly used metrics are the True Positive Identification Rate (TPIR), and False Positive Identification Rate (FPIR) [73]. For more in-depth information, refer to [128], which provides a comprehensive survey of the open-recognition problem, focusing as well on the facial standpoint.

2.3.3 Datasets

2D Face Recognition The advancement of deep face recognition heavily depends on the existence of large-scale training datasets, which are essential for learning deep facial features and representations. Additionally, high-complexity testing datasets are essential to avoid performance saturation. Table 2.3 provides specifications for some of the most prominent and widely used face recognition datasets, offering important details such as the number of subject images, total image count, and the number of images per subject.

TABLE 2.3: Datasets used for training and testing 2D deep face recognition.

Name	Year	# Subjects	# Images	# Images per subject	Annotations
CASIAWebFace [11]	2014	10 575	494 414	47	–
MS-Celeb-1M [129]	2016	20*K*	100*K*	5	BB
VGGFace2 [130]	2017	9 131	3.31*M*	363	pose, age
LFW [12]	2008	5 749	13 233	2.3	–
IJB-B [131]	2017	1 845	11 754	11.4	BB

These datasets can be utilized jointly to achieve optimal models, as demonstrated in a systematic study conducted by Cao *et al.* [130]. In their study, the authors trained a ResNet-50 model on VGGFace2, on MS- Celeb-1M, and their union, achieving an improved recognition performance on the latter.

CASIAWebFace The CASIAWebFace [11] dataset, introduced in 2014, was the first publicly available large-scale training face dataset. This dataset comprises 494 414 images of 10 575 subjects, which were collected semi-automatically from the Internet without annotations. The subjects featured in the dataset are celebrities, resulting in a long-tail distribution of images per subject [132]. CASIAWebFace has gained significant popularity as a training dataset owing to its moderate size and user-friendly nature.

MS-Celeb-1M The MS-Celeb-1M [129] dataset was introduced in 2016 and praised as the largest publicly available training dataset at its release. The dataset was created using Bing Search and contains approximately $10M$ images of 10 000 celebrities, with an average of 100 images per subject. Each subject in the dataset is annotated with a bounding box. Although the dataset has some limitations, such as the need for more quality annotations and the presence of duplicate and non-face images, it remains one of the primary and most extensive training datasets used by the community.

VGGFace2 The VGGFace2 [130] dataset was introduced in 2018 as a large-scale training dataset. It consists of $3.31M$ images of 9 131 subjects, including celebrities with significant pose, age, illumination, and ethnicity variations. Each identity in the dataset is represented by an average of 363 images, which is considered a substantial number for training purposes. The images were collected from Google Image Search and annotated through a manual and automated procedure.

LFW The LFW (Labeled Faces in the Wild) [12] is a classic benchmark dataset introduced in 2008 to address the problem of unconstrained face recognition. Over the years, it has become the most widely used benchmark for evaluating face recognition systems. The dataset comprises 13 233 images featuring 5 749 different identities collected from the web. For evaluation purposes, the standard LFW protocol employs 6 000 face pairs, consisting of 3 000 genuine pairs and 3 000 impostor pairs, to evaluate the mean verification accuracy. Each image in the dataset is 250 × 250 pixels in size. However, despite its significance, the state-of-the-art performance on LFW has reached a saturation point [15].

IJB-B The IARPA Janus Benchmark-B (IJB-B) [131] dataset was created in 2017 as an extension of the 2015 IJB-A dataset [133]. Its purpose was to address the limitations of the previous dataset, which included limitations of unconstrained traits, a relatively low number of impostors, and a more uniform geographic distribution. The images in IJB-B, totalling 11 754, were collected from the Internet and feature 1 845 different subjects, including video frames from various sources. IJB-B serves as a testing dataset for face verification and identification protocols, applicable to both images and videos. Subsequently, the 2018 IJB-C dataset [134] was introduced, representing the latest addition to this series, further improving dataset size and variability.

3D Face Recognition Within the realm of 3D data, the emphasis is placed on point clouds and meshes rather than range images. Table 2.4 presents the specifications of some of the most prominent and widely used 3D face recognition datasets, including essential details such as the number of subjects, the number of faces, and the data format type. Additional information is provided below for each dataset. The iPhonePLYv3 dataset and its data collection process will be discussed in detail in Section 4.1, and it is presented here for the purpose of comparison within that section. It is important to note that this dataset was not specifically designed for 3D face recognition; rather, its creation is primarily related to 3D face anonymization.

TABLE 2.4: Datasets used for 3D face recognition.

Name	Year	#Subjects	#3D Models	Data Type
GavabDB [135]	2004	61	427	Mesh
BU-3DFE [136]	2006	100	2 500	Mesh
Bosphorus [137]	2008	105	4 666	Point Cloud
BJUT-3D [138]	2009	500	500	Mesh
FRAV3D [139]	2013	106	1 696	Mesh
FaceScape [140]	2020	938	18 760	Mesh
iPhonePLYv3	2023	201	201	Point Cloud

Compared to 2D face datasets, 3D face datasets are less common and smaller in scale [80]. A significant challenge in this domain lies in developing affordable 3D acquisition systems capable of providing the vast amount of data required by these techniques to handle the inherent complexity of 3D data representation and processing [63].

GavabDB The GavabDB [135] is a 3D face analysis dataset established in 2004. It comprises 427 meshes of facial surfaces without texture from 61 subjects, including 45 males and 16 females, with seven images per subject. All subjects in the dataset belong to the white ethnicity and are aged between 18 and 40 years old. The database offers systematic variations in pose and facial expressions, making it suitable for diverse research purposes. The 3D facial data was captured using a Minolta Vi-700 laser range scanner.

BU-3DFE The Binghamton University 3D Facial Expression (BU-3DFE) [136] dataset was released in 2006 to develop expression-invariant face recognition. This dataset comprises 100 subjects, with 56% female and 44% male, totalling 2 500 facial models. It includes six expressions: anger, happiness, sadness, surprise, disgust, and fear. The subjects' ages range from 18 to 70, with diverse ethnic ancestries represented.

Bosphorus The Bosphorus [137] dataset, released in 2008, is a comprehensive 2D and 3D human face dataset designed to simulate adverse conditions for facial analysis. It incorporates diverse expressions, including the six basic emotions, variations in head pose (13 yaw and pitch rotations), and different types of occlusions (beard, mustache, hair, hand, and eyeglasses). The dataset includes 105 subjects and 4 666 faces, with some subjects being actors, contributing to a more realistic representation of emotions. The 3D facial data was acquired using an Inspeck Mega Capturor II 3D scanner.

BJUT-3D The BJUT-3D [138] dataset, released in 2009, is a 3D database designed for face analysis tasks. The dataset comprises 500 Chinese individuals, including 250 males and 250 females, all depicted with neutral expressions and without accessories. The 3D facial data was captured using a CyberWare 3030 RGB/PS laser scanner.

FRAV3D The FRAV3D [139] dataset, introduced in 2013, is a multimodal dataset customized for 2D, 2.5D, and 3D controlled facial analysis. The dataset consists of 106 subjects, roughly one-third being women, all captured under controlled lighting conditions and without accessories. Each subject in the dataset has 16 captures, each featuring different poses or lighting conditions, although not simultaneous. The data was acquired using a Minolta VIVID 700 scanner, which provides texture information in the form of a 2D image and a Virtual Reality Modeling Language (VRML) file for the 3D image representation.

FaceScape The FaceScape [140] dataset was released in 2020, presenting a large-scale high-quality 3D face dataset. It comprises 18 760 textured 3D faces from 938 subjects, each captured with 20 specific controlled expressions. The age range of the subjects is between 16 and 70 years old, and most of the subjects are of Asian ethnicity. The facial data was acquired using a dense multi-view system comprising 68 Digital Single Lens

Reflex (DSLR) cameras to create accurate 3D face models with detailed textures for each subject.

2.3.4 Benchmarks

Table 2.5 presents the performance of some of the leading models on the LFW benchmark, which is the most commonly used benchmark for evaluating face detection algorithms [64]. The best algorithms exhibit only marginal differences, with higher than 99.5% accuracy. This level of performance indicates that, despite using more powerful models, capturing a significant quantitative gain over the previous models and accurately measuring its true strength becomes impossible [3]. Consequently, the performance on the LFW dataset is now saturated, and it may no longer be sufficient to accurately evaluate the quality of the latest models.

TABLE 2.5: State-of-the-art models on 2D face verification, on the LFW benchmark.

Method	Accuracy (%)
FaceNet [109]	99.63
CosFace [141]	99.73
PRN [142]	99.76
Deep Embedding [143]	99.77
L_2-Softmax [144]	99.78
ArcFace [145]	99.83

Regarding the 3D, Table 2.6 reports the results on the Bosphorus dataset, which is one of the main 3D facial point cloud datasets. The benchmarks exhibit that the deep-learning techniques tend to have a higher performance compared to the traditional methods. Although the recognition rate is high, it closely relates to the size of the testing set, which is considerably lower than the 2D counterpart.

TABLE 2.6: State-of-the-art models on 3D face verification, on the Bosphorus dataset.

Method	Recognition Rate (%)
Li *et al.* [146]	95.40
Berretti *et al.* [147]	95.70
WESC [148]	97.75
FR3DNet [123]	98.60
PointNet-CNN [122]	98.91

2.4 Face Anonymization

The ISO/IEC 29100:2011, published in 2011 and reviewed and confirmed in 2017, defines anonymization as:

> *Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal*[1] *can no longer be identified directly or indirectly, either by the PII controller*[2] *alone or in collaboration with any other party.*

Within the domain of face anonymization, the focus is solely on protecting the privacy of an individual's identity captured in facial data, achieved through the irreversible alteration of PII associated with the face. Face anonymization is a crucial tool employed in various applications to address ethical concerns by mitigating the privacy risks of biometric recognition systems. It plays an essential role in balancing the benefits of biometric systems and the need to uphold ethical principles, fostering trust and accountability in using facial data.

An effective face anonymization system should possess several key properties to ensure the privacy and protection of individuals. These properties may include [149]:

- *Preservation of Anonymity:* The result must conceal the original identity;
- *Realistic:* The result must look authentic and preserve the performance of state-of-the-art detection and recognition systems;
- *Controllable:* The anonymization process should be controllable through a control parameter that determines the fake identity of the anonymized image;
- *New Identities:* The anonymized identity must not belong to the training set.

2.4.1 Algorithms

Much research has been conducted regarding 2D face anonymization, although not to the same extent as face detection and anonymization. Meden *et al.* [150] present a comprehensive introduction to privacy-related research, reviewing existing works on Biometric Privacy-Enhancing Techniques (B-PETs) applied to face biometrics. Rakhmawati *et al.* [151] provide a comparison of several existing Traditional methods, underlying the advantages and disadvantages of each method. Ribaric *et al.* [152] offer a broader overview

[1] The natural person to whom the PII relates.

[2] The privacy stakeholder that determines the purposes for processing PII.

of de-identification approaches for both biometric and non-biometric identifiers, including behavioural and soft biometric data. In the context of privacy and security, Cai *et al.* [153] comprehensively survey recent trending approaches that leverage Generative Adversarial Networks (GANs), which are still in an early stage of development, depicting both their advantages and drawbacks.

However, the 3D domain has been relatively underexplored, resulting, to the best of the research efforts, in no surveys on this topic.

2D Face Anonymization In their work, Meden *et al.* [150] present a comprehensive taxonomy of face anonymization techniques. The following outline provides a simplified version by presenting a selection of illustrative global methods. As such, the *Traditional Methods* are divided into two groups: *Obfuscating Techniques* and *Synthesis Techniques*.

Obfuscating Techniques consist of simple techniques[1] that degrade the image quality at the expense of reduced biometric data utility. This group encompasses the following approaches:

- *Masking:* Conceals the entire facial region or parts using masks or shapes, including rectangles, ellipses, or circles with solid colours, typically black [154, 155]. These works relate to video surveillance, which demands a module to first detect and track the faces in the video frames and then create the obscuring masks using the detected face locations and scales;
- *Blurring:* Reduces the detail level of images with a smoothing technique such as Gaussian filters [156]. A Gaussian filter, commonly used in image processing, serves as a low-pass filter. This filter is implemented as a symmetric kernel that is convolved with either the input image or a selected region of interest. For instance, when considering a standard deviation σ of 1, the Gaussian kernel approximations for both the 3×3 and 5×5 cases take the following form:

$$\frac{1}{16} \times \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array} \quad \frac{1}{273} \times \begin{array}{|c|c|c|c|c|}\hline 1 & 4 & 7 & 4 & 1 \\ \hline 4 & 16 & 26 & 16 & 4 \\ \hline 7 & 26 & 41 & 26 & 7 \\ \hline 4 & 16 & 26 & 16 & 4 \\ \hline 1 & 4 & 7 & 4 & 1 \\ \hline \end{array} \tag{2.10}$$

[1] Also known as *naïve* or *ad hoc* methods.

These values are a discrete representation of the Gaussian Function defined as:

$$G(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.11}$$

A Gaussian filter assigns a higher weight to pixels near the center of the kernel and gradually reduces the weight as it moves away from the center. Another smoothing technique is the average filter, which averages the pixel values of all neighboring pixels instead of weighting the pixels like the previous;

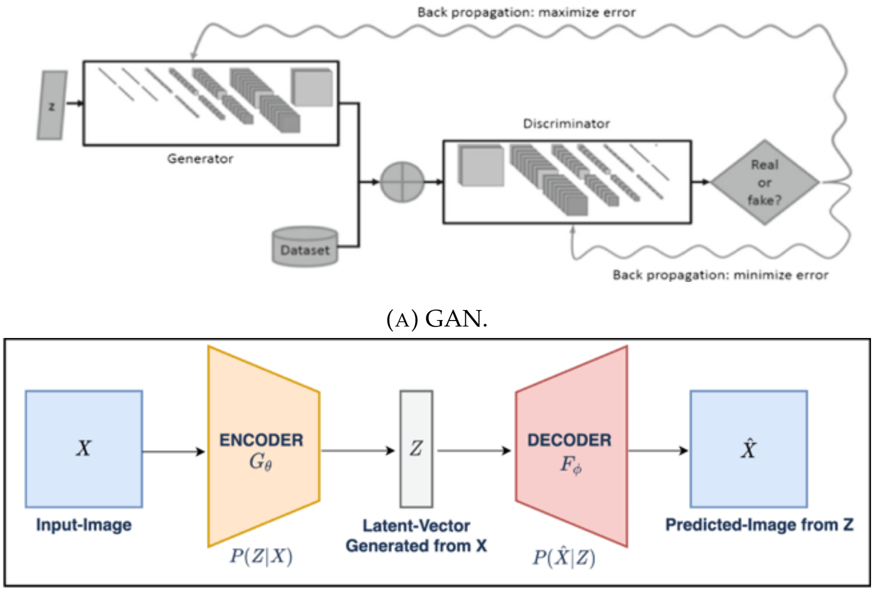
- *Pixelization:* Reduces the resolution of an image by aggregating the pixels into groups of uniform squares or rectangles, stretching them to a point beyond their original size [157];

- *Warping:* A geometrical transformation that destroys neighbouring pixel relationships by shifting their positions and interpolating their intensities [158]. Korshunov and Ebrahimi, the authors, select a set of key points in the image and shifts their coordinates based on random values and warping strength. A transformation matrix is computed according to the destination coordinates and applied to each pixel using cubic interpolation.

Other studies introduce artifacts in the image, such as noise [159], or apply other simple digital image processing transformations. However, these approaches have proven to highly distort the integrity of the original face and be vulnerable to comparatively simple attacks [160].

The *Synthesis Techniques* generate synthetic facial data with predefined attributes and are more sophisticated than the previous group. One of the most famous is the K-Same family, which implements the k-anonymity [161] strategy on facial images. The k-anonymity property ensures that the information of each person cannot be distinguished from at least $k-1$ individuals, with a possible success rate of recognition of $1/k$. For example, Newton *et al.* [162] presented the k-Same-Pixel and k-Same-Eigen. While both approaches consider the k-closest faces to the de-identification input image, the former is based on the pixel-wise average of the original face images, whereas the latter leverages PCA to average the projected images. Despite attaining adequate levels of privacy, they can lead to undesirable artifacts, the so-called “ghosting” artifacts, caused by alignment errors [160]. Other research efforts aimed at improving results utility, such as *k*-Same-Select [163] and *k*-Same-M [164], utilize Active Appearance Models (AAMs) to statistically represent and model facial appearance. Unlike traditional methods like PCA, AAMs combine facial shape and texture, generating more visually convincing faces with fewer artifacts.

The *Deep learning Methods* follow the *Synthetic Technique* approach of generating new facial data with higher complexity. These methods utilize generative modelling, a Machine Learning approach that aims to approximate complex and high-dimensional probability distributions of sample data using Neural Networks (NN) to generate new statistically similar data [165]. Within the context of face anonymization, GANs and Variational Autoencoders (VAE) are the most popular generative approaches. Refer to Figure 2.4 for a visual representation of the architecture of both deep learning models.



(A) GAN.

(B) VAE.

FIGURE 2.4: General Deep Learning architectures. Extracted from [166] and LearnOpenCV, respectively

GANs employ an adversarial process between a generator and a discriminator. The generator network learns to generate new data samples that resemble the original data, while the discriminator network learns to distinguish between real and generated data. Through an iterative training process, the generator and discriminator networks compete against each other, improving the quality of the generated samples over time. The process of replacing sensitive facial data with synthetic data ensures data privacy preservation while retaining the statistical properties that contribute to the photorealism of the generated faces. For instance, DeepPrivacy [167] employs a U-Net architecture for its generator, incorporating background and pose information to enhance the realism of generated faces. The discriminator, mirroring the generator's filter count, includes the background information as a conditional input and concatenates the pose information at each resolution. In the case of CIAGAN [149], it introduces a module for controlling the generator's characteristics, allowing target identity injection. The input representation includes facial landmarks (silhouette, mouth, nose bridge) to ensure pose preservation. A masked background image encompassing the forehead region is fed as input to the generator, improving overall visual quality. FPGAN [168] introduces a pixel loss function to guide the privacy de-identification process. Its generator employs an enhanced U-Net architecture, while the discriminator module comprises two custom-designed discriminators.

On the other hand, VAEs combine the capabilities of autoencoders with the generative power of probabilistic modelling. VAEs consist of an encoder network that maps the input data to a latent space representation and a decoder network that reconstructs the input data from the latent space. By learning the underlying distribution of the input data in the latent space, VAEs can generate new data samples by sampling from this distribution. Similar to GANs, the resulting faces appear realistic and exhibit the desired properties in terms of facial characteristics. For instance, studies such as [169, 170] have investigated the use of VAEs in face anonymization. These studies train the VAE's encoder to shift faces toward other faces with desirable attributes. They explore methods for acquiring new encoding targets in both supervised and unsupervised settings or even incorporate multiple privacy protection modes with VAEs to achieve privacy-preserving faces.

3D Face Anonymization Despite extensive research efforts, the existing literature addressing 3D face anonymization on point clouds was quite limited. In its bachelor thesis, Rustici [171] proposes an approach that involves globally registering the source point cloud to an oriented template and refining the alignment using the ICP technique. Subsequently, a point cloud template is utilized to eliminate unnecessary facial features by computing their nearest neighbours on the source face model, resulting in a face model that retains only essential facial traits. Figure 2.5 depicts the anonymization outcomes for two subjects following the procedure above.

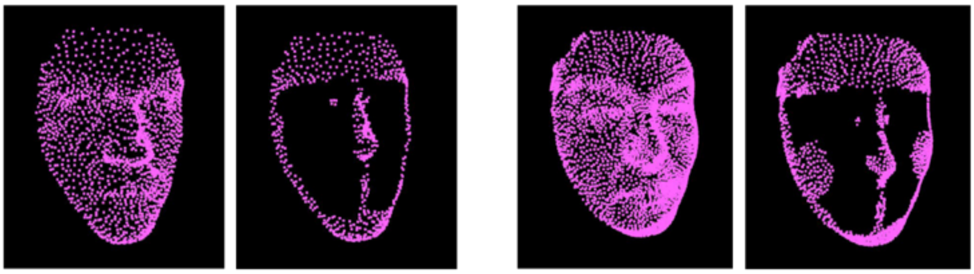


FIGURE 2.5: Anonymization results of two subjects following Rustici procedure. Extracted from [171].

Singh and Ramachandra [172] present a 3D face morphing technique that combines the 3D face point clouds of two individuals. The process begins by transforming the initial 3D facial point clouds into depth-maps and 2D color images in a canonical view. Morphing operations are then performed on the color images using facial landmarks to detect key points and perform Delaunay Triangulation to estimate the affine warping. These operations are extended to the depth-maps. Subsequently, they project the 2D morphed color-map and the depth-map back to the point cloud and refine the results by filling in any gaps, resulting in a 3D face morphing model. In their work, they target the vulnerability of a morphing model to face recognition systems by morphing two individuals in a way that both subjects can be identified as having the same identity in the morphed result. This contrasts with the anonymization goal, where the identity being anonymized should not be associated with its anonymized result. However, their work does provide insights into the potential application of morphing techniques in anonymization, albeit with a different objective.

Other studies in this domain primarily focus on medical image input data, particularly high-resolution Magnetic Resonance Imaging (MRI). These MRIs enable the generation of 3D images using volume rendering software that may inadvertently expose the identity of the subjects, necessitating anonymization [173]. For instance, [174] identifies five landmarks on the MR image corresponding to the ear, eyes, and nose and proceeds to either distort or remove these features for anonymization purposes within the 3D MRI representation.

2.4.2 Evaluation and Performance Metrics

Evaluating privacy protection techniques involves two key aspects: privacy and utility. Privacy metrics are used to assess the level of privacy protection provided by an anonymization technique. On the other hand, utility metrics focus on evaluating how well the visual appearance of the data is preserved and whether relevant information necessary for downstream tasks is retained, closely relating to the intended data usage. The main objective of face anonymization algorithms is to strike a balance between maximizing data utility and ensuring adequate privacy protection. However, achieving this balance proves to be challenging as privacy and utility often have an inverse relationship, referred to as the privacy-utility trade-off. Balancing these competing goals is crucial, as illustrated in Figure 2.6. The acceptable trade-off is closely intertwined with the intended purpose for the subsequent use of the data.

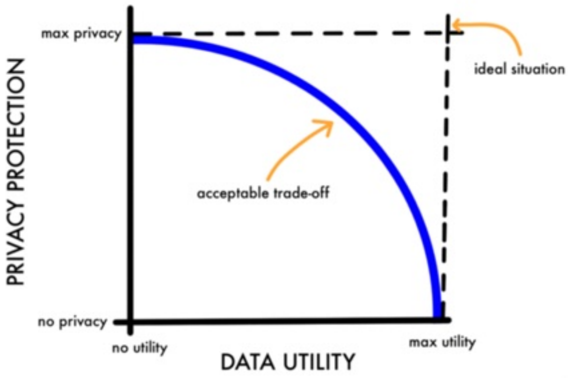


FIGURE 2.6: The trade-off between data privacy and utility. Extracted from [175].

The privacy-utility trade-off analysis involves two distinct evaluation approaches: subjective and objective. The subjective evaluation is a qualitative approach that relies on human evaluators and their subjective judgments, which may be influenced by biases. It includes user feedback collection through surveys or interviews, aiming to assess the visual quality of the results [176, 177], as well as privacy preservation. In contrast, the objective evaluation is a quantitative approach that employs measurable and quantifiable metrics, drawing from related domains like face recognition and digital image processing.

Privacy Metrics Anonymization algorithms are commonly evaluated by assessing their ability to protect the identity of anonymized faces. This evaluation is often conducted by subjecting them to state-of-the-art facial recognition models acting as attackers and using standard face recognition evaluation metrics. These metrics include ROC curves, AUC, or CMCs, as discussed in Section 2.3.2. The evaluation involves measuring these metrics both with and without applying the anonymization techniques, resulting in two sets of results. These results are then compared to measure the privacy enhancement provided by the anonymization method. This comparison does not quantify the anonymization strength in the form of a scalar measure, although some efforts have been made towards that goal by Pavel Korshunov *et al.* [178], and Terhörst Philipp *et al.* [179].

To obtain the two sets of results, various attacker paradigms exist in the literature, such as the *naïve recognition* that compares the recognition models' performance on the original images (the gallery) to their performance on the altered images (the probes). Given this setting, the attacker takes no action to account for the de-identification effect. In addition to *naïve recognition*, there are two other types of attackers related to the arrangement of the gallery and probe sets, as denoted in [162]. The second type is *reverse recognition*, where the altered images (the gallery) are matched to the original images (the probes). The third type is *parrot recognition*, where the altered images are matched to altered images, functioning as the gallery and probe sets. Furthermore, another attacker methodology [180] aims to revert the anonymization data before applying *naïve recognition*. The assumption is that the de-anonymized data is closer to the original data than the anonymized data. This reversibility is achieved with an Autoencoder model, demonstrating a high versatility due to its ability to generalize the reversibility to adapt to multiple anonymizations. Despite the authors not explicitly labelling the approach, henceforth, it shall be referred to as "reversibility recognition" in this work. Refer to Figure 2.7 for a visual representation of the four attacker configurations.

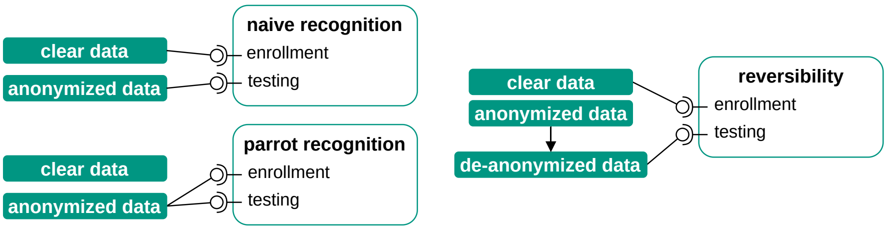


FIGURE 2.7: Types of attackers to reidentify anonymized data. Extracted from [180].

Utility Metrics Most research studies utilize image quality evaluation strategies to measure the data utility. Among the well-known metrics are the Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Fréchet Inception Distance (FID).

The *Structural Similarity Index Measure* (SSIM) is a widely used similarity metric for comparing two images. It was proposed by Wang *et al.* [181] as a more robust solution to the problem of image quality assessment. The SSIM metric comprises three terms that estimate the impact of image luminance, contrast, and structural changes. Mathematically, the SSIM between images x and y is defined as:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{2.12}$$

where $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ control the relative significance of the correspondent three terms of the index. The luminance, contrast, and structural components of the index can be defined individually as:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2.13}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2.14}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{2.15}$$

where μ_x and μ_y, σ_x^2 and σ_y^2 represent the means and variance of the pixels in the original and anonymized images, respectively, while σ_{xy} denotes their covariance. Additionally, $C_i, i = 1,2,3$ are stabilization constants. The metric is bounded by the interval $[-1,1]$ but is often presented on the interval $[0,1]$. A higher SSIM value indicates a stronger similarity in terms of perceptual quality between the two images compared.

The PSNR is commonly used in image compression to quantify the reconstruction quality. It is a modified version of the Mean Squared Error (MSE). The PSNR is mathematically defined as:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{2.16}$$

where MAX represents the maximal variation in the original image I. For an 8-bit image RGB image, $MAX = 255$. The MSE is defined as:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I_{ij} - K_{ij})^2 \tag{2.17}$$

where M and N represent the number of rows and columns of the image pixels, while I and K denote the original and anonymized images, respectively. The PSNR is typically expressed as a logarithmic quantity using the decibel scale (dB), where a higher value implies a more substantial similarity in perceptual quality.

The FID was proposed by Heusel *et al.* [182] as a means of calculating the similarity of generated images to real images, originally created to evaluate the performance of GANs. The metric uses the model Inception-v3 [183] up to the last layer before the output classification to compute the features of input images from a collection of real and generated images. The collection is summarized with a multivariate Gaussian by computing their mean and covariance, and the distance between the two distributions is calculated using the Fréchet distance [184]. The FID is mathematically defined as:

$$FID(P, Q) = \|\mu_P - \mu_Q\|^2 + \mathrm{Tr}(C_P + C_Q - 2(C_P C_Q)^{1/2}) \tag{2.18}$$

where *Tr* refers to the linear algebra trace function, and, P and Q, represent two multidimensional Gaussian distributions with mean and covariance matrices denoted as μ_P, C_P, and μ_Q, C_Q, respectively.

Additionally, several other studies investigate the effects of anonymization on face detectors and employ conventional evaluation metrics, such as AP, to quantify the impact [167].

2.4.3 Datasets

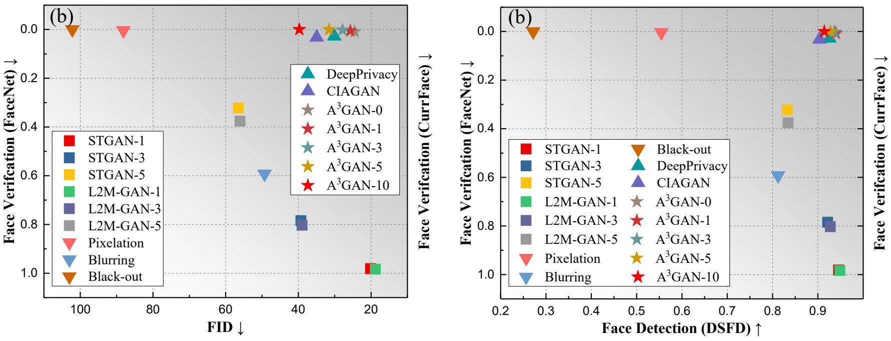
In the 2D domain, experiments often rely on standard face datasets commonly used in face recognition and related tasks, including those mentioned in Sections 2.2.3 and 2.3.3. Additionally, custom-made datasets, not publicly available, are occasionally utilized [150].

In the 3D domain, due to the literature absence and to the best of the research efforts, no dataset trend selection can be observed. Nevertheless, the only identified work regarding the 3D point cloud anonymization [171] employs a custom-made dataset captured using an Artec Leo scanner, each scan containing between 20 000 and 30 000 points. However, specific details regarding the dataset's size, identity numbers, and additional information are not provided.

2.4.4 Benchmarks

The absence of a standardized benchmark for evaluating face anonymization techniques, coupled with variations in summarizing state-of-the-art results, poses challenges when comparing models. Different authors employ a range of evaluation strategies, taking into account diverse privacy and utility metrics and utilizing various face recognition attacker models across multiple datasets.

Nevertheless, in Figure 2.8, a visual representation of the privacy-utility trade-off of various approaches is presented. These encompass three basic techniques: 8 × 8 pixelization, 9 × 9 Gaussian blur, and black-box masking. Additionally, two state-of-the-art deep generative face de-identification methods are included, DeepPrivacy [167] and CIAGAN [149], along with variations of the method proposed by Zhai *et al.* [185], A^{3}GAN. Three variants of two facial attribute editing methods, STGAN, and L2M-GAN, are also considered. The results highlighted in the figure underscore the superior performance of deep learning-based methods.



(A) Face Verification vs. Image Quality.

(B) Face Verification vs. Face Detection.

FIGURE 2.8: Privacy-utility trade-off (top-right corner is the best). Extracted from [185].

In the context of the 3D domain, the sole identified study evaluated anonymization effectiveness with an online questionnaire, gathering responses from 100 participants as detailed in [171]. The findings revealed that less than 2% of respondents correctly identified the identity of the anonymized 3D face model among six images, with only one image being the actual match. The evaluation did not include any additional metrics or quantifiable measures.

2.5 Discussion

2.5.1 Challenges

While face detection and recognition have achieved notable success, specific difficulties persist that hinder their performance in real-world scenarios. Kumar *et al.* [18] enumerates nine challenges in the field of face detection, which are shared across face recognition [186]. These challenges hinder the model performance and demand robust algorithms to handle them. Some of those challenges include:

- *Background complexity:* Denotes the presence of many background objects that function as distractor elements, making it challenging to distinguish them from the face;
- *Illumination:* Refers to changes in lighting conditions, such as varying brightness, shadows, or uneven illumination, which can affect the visibility and appearance of faces;
- *Image resolution:* The number of pixels determines image resolution in an image, which in turn affects the level of detail and clarity of faces. Lower-resolution images may make it impossible to capture well-defined facial features adequately, resulting in a loss of essential details and reducing the amount of information available for face analysis;
- *Occlusion:* Refers to the partial or complete obstruction of facial features caused by the subject, objects, or other individuals. This obstruction can alter the natural structure and appearance of the face;
- *Odd expressions:* Can introduce significant variations in the appearance of faces and cause changes in the position and shape of facial features;
- *Orientation variations:* Reflect variations in one or more of the head's degrees of freedom, i.e., pitch, roll, and yaw of the subject, which introduce changes in the overall geometry of the facial features.

These challenges mainly concern the facial attributes and variations depicted in the input images, promoting ongoing research efforts to mitigate them and develop problem-oriented methods designed to address them. Additionally, Du *et al.* [73] highlight other challenges concerning data and label distribution, computational efficiency, and the explainability of the overall process, which is currently a major topic in the community [187].

As an alternative solution to the challenges mentioned above, researchers have resorted to 3D approaches. While 3D data can be sensitive to facial expressions, its advantage lies in its invariance to pose and lighting conditions, leading to improved efficiency in recognition systems [78]. According to Kusuma *et al.* [188], 2D+3D algorithms can be considered complementary, as the additional dimension compensates for the absence of depth information and addresses issues related to pose and illumination variations. Regarding point clouds, which is the type of data being approached, other challenges arise due to their inherent nature:

- *Sparsity:* Point clouds may exhibit a spread distribution of points over an object's surface. In some regions, the density of points collected by the sensor might be reduced.
- *Unstructured data:* The cloud points have no structure or organization, making it impossible to index them in a way that their neighbours are related. This lack of structure poses challenges in analysis and processing.

2.5.2 Critical Reflection

Considerable research has been dedicated to face detection, face recognition, and face anonymization in the 2D space. This research has yielded impressive results, driven by advancements in Deep Learning algorithms and the availability of large-scale face datasets. Simultaneously, 2D+3D detection and recognition research has emerged as a parallel field to overcome the challenges faced in 2D, unlike anonymization, which presents minimal studies within this paradigm frame.

However, in the 3D domain, face detection and anonymization remain significantly underdeveloped, with the reviewed literature revealing a near-absence of scientific papers on these subjects. One possible explanation for this scarcity is the limited availability of large-scale 3D facial datasets, the increased complexity of working with 3D data, and the low resolution of sensors used in current applications that would demand this research, such as autonomous driving. Nevertheless, investigating these fields is paramount, given the increasing resolution of sensors and the widespread use of LiDAR technology.

Consequently, there undoubtedly exists a clear gap in academic research regarding 3D face anonymization on point clouds. This gap motivates this thesis to contribute advancements and insights into this unexplored field. This work serves as an initial stepping stone toward more complex research endeavours.

Chapter 3

Methodology

This chapter is devoted to the methodology involved in this research, outlining the techniques employed regarding face detection, recognition, and anonymization to achieve the study's objectives. It commences with a succinct high-level description, offering a broad understanding of the methodology by outlining the main components addressed in subsequent sections. Then, the chapter thoroughly explores the methods of face anonymization, recognition, and detection, substantiating their selection and presenting their operational settings. Additionally, insights are presented regarding the evaluation process, which includes a description of an attacker model used for assessing the anonymization. At last, acknowledging the importance of transparency, it openly addresses the study's limitations, offering an assessment of encountered constraints during the research.

3.1 High-Level Description

The primary aim of this research is to develop novel 3D face anonymization techniques designed for point clouds and to conduct a thorough evaluation of their effectiveness in balancing privacy protection and data utility preservation.

In an initial step, a range of 3D face anonymization solutions is proposed, drawing from various domains, as there is limited prior research in this field. While these techniques are built upon pre-existing algorithms from diverse domains, they have never been adapted for anonymization purposes. Hence, this thesis represents an innovative step in extending these techniques to the anonymization context.

The subsequent step involves evaluating these proposed solutions, following the established procedures in existing literature that assess the interplay between privacy and utility from both qualitative and quantitative perspectives.

In the privacy evaluation, a critical aspect involves the utilization of a face recognition model as an attacker[1] regardless of the attacker paradigm considered from the ones presented in Section 2.4.2. In this work, both the *naïve recognition* and *reversibility recognition* paradigms are explored under the verification and identification (with closed-set protocol) modes of face recognition, discussed in Section 2.3. For *reversibility recognition*, an autoencoder is designed as the de-anonymization model with the goal of reversing the anonymization process. On the utility front, a comprehensive evaluation was conducted without immediate plans for specific applications, aiming to uncover the potential benefits and limitations of these techniques across various use cases. This evaluation includes the utilization of a face detection model for computing specific metrics. Ultimately, a proposed evaluation methodology proceeds with the analysis of the relationship between privacy and utility, highlighting their trade-off and emphasizing the overall effectiveness of the anonymization techniques. This comprehensive methodology assures the achievement of the thesis objectives by proposing and subjecting multiple 3D face anonymization solutions to a well-rounded evaluation process.

However, despite the proposed facial anonymization techniques operating in point clouds within the 3D space, the evaluations related to privacy, utility, and their trade-off were conducted within the 2D space. This decision was influenced by constraints related to the availability of 3D evaluation resources, particularly open-source 3D face recognition models. More details about the evaluation strategy will be presented in the next chapter.

As outlined in the high-level description of the framework pipeline above, the domains of face detection, face recognition, and face anonymization, discussed in the previous Chapter 2, are pivotal components of this research. To visualize their integration within the operational pipeline and gain insights into the sequence of procedures mentioned above, refer to Figure 3.1.

The blue arrows represent the dataset used for conducting experiments, which includes facial data in the form of point clouds (Dataset 3D) and images (Dataset 2D). Initially, face anonymization techniques are applied to the 3D data of the dataset, resulting in a collection of anonymized point clouds (Anon 3D). Since their evaluation takes

[1] This also implies the use of a face detection model, which forms the initial stage of the recognition pipeline.

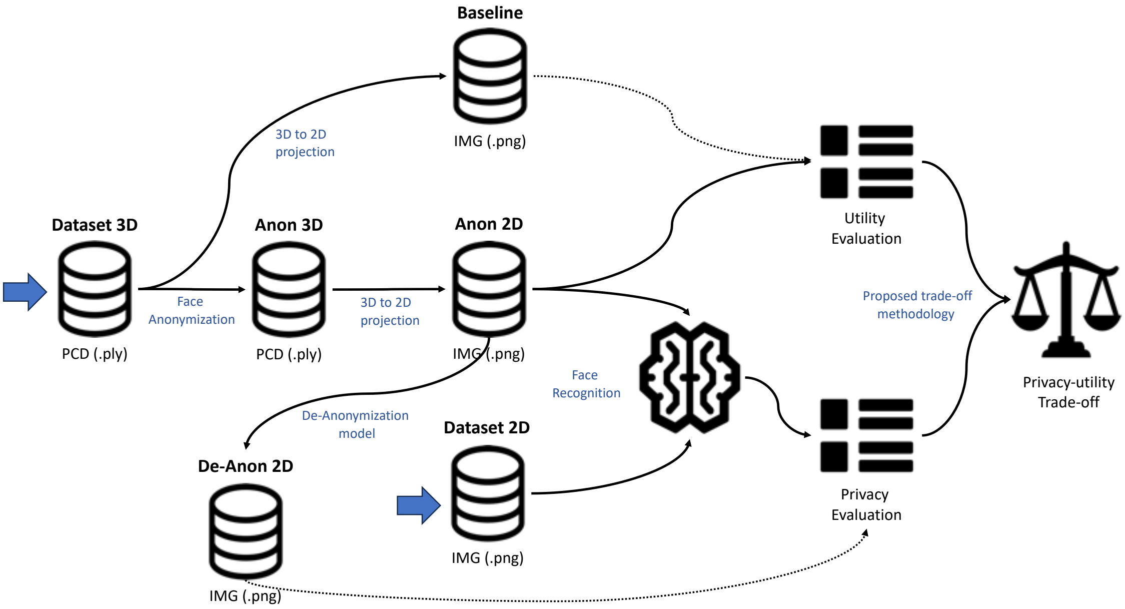


FIGURE 3.1: The overarching pipeline of the conducted experiments.

place within the 2D space, both the anonymized and non-anonymized point clouds are projected into the lower-dimensional space, producing a set of anonymized and non-anonymized images (Baseline, and Anon 2D, respectively). The non-anonymized images serve as a baseline for assessing privacy[1]. Their results are compared to the results obtained with a face recognition model, using the original dataset images as the gallery set and the anonymized images as the probe set. To further test privacy protection, the anonymized images are then subjected to de-anonymization using an autoencoder model (De-Anon 2D). The non-anonymized images are also used to calculate specific utility metrics within the employed utility evaluation.

In the remainder of the chapter, the rationale behind the selection of the methods used in face anonymization, face recognition, face detection, and de-anonymization steps will be expounded, along with detailed explanations for each.

3.2 Face Anonymization Techniques

In Section 2.4, attention was drawn to the near-absence of scientific papers dedicated to point cloud face anonymization. To the best of the research efforts, the existing literature addressing 3D face anonymization on point clouds is quite limited. Consequently, one of

[1] They are also enrolled as the probe set in the face recognition step, separately.

the primary contributions of this thesis involves proposing multiple techniques to tackle the challenge of face anonymization in point clouds.

3.2.1 Algorithms Selection

In the literature review conducted in Section 2.4.1 regarding face anonymization techniques in point clouds, only a singular work addressing this problem has been identified. This highlights the necessity for the development of new solutions and the exploration of alternative techniques. However, the lack of existing research poses challenges in formulating solutions, as the restricted knowledge within the field hinders insights into potential approaches that could serve as foundational principles for other variants. As a result, the rationale behind proposing these solutions was based on extending the principles from the well-established 2D face anonymization field into the 3D domain. This process also entailed the utilization of pre-existing methods from other domains, which may require some degree of adaptation to meet the specific data requirements.

In accordance with this conceptual shift, the categorization of the diverse 2D techniques has been reconfigured for the 3D context. The solutions are then classified based on their inherent methodologies into six categories: Sampling-based, Noise-based, Warping-based, Smoothing-based, Morphing-based, and Point Operations-based techniques. While each category has the potential to include a variety of methods, only one per category will be outlined, as others could be derived from the same underlying principles. Here, the emphasis was placed on testing a wide range of different methodologies rather than limiting the scope to various techniques from just a few categories. This approach was adopted to develop a broader understanding of how various distinctive strategies can be integrated into the framework of 3D anonymization. It is significant to note that these techniques have not been previously applied in this context by any prior work. This research marks their pioneering implementation and evaluation within the 3D face anonymization realm. Hence, their name was deliberately coined for this research, even though the anonymization techniques integrate pre-existing methods from other domains.

3.2.1.1 Sampling-Based

Point cloud sampling is a process used to reduce the number of points in a point cloud while striving to preserve its original shape and characteristics. This process is commonly employed as an essential pre-processing strategy before implementing 3D deep learning models to enhance memory and computational efficiency [189]. However, enlarging the sample size results in severe information loss, boosting the potential for anonymization. The various sampling techniques have adjustable parameters that regulate the process and control the information loss magnitude. These attributes exhibit promising potential when it comes to anonymization. This category is similar to pixelization in 2D images, as it reduces the amount of information in the point cloud data through removal or aggregation. In this category, a voxel-based sampling approach was selected, referred to as *CentroidVoxel*. This selection was made because it has the capability to generate a 3D uniform grid-like result similar to pixelization in the lower space. Other sampling approaches that could have also been considered include Poisson Disk Sampling or Farthest Point Sampling (FPS).

3.2.1.2 Noise-Based

Point cloud noise addition entails the introduction of random or structured variations to the position, or color of the data points. Noise, a naturally occurring phenomenon during data management, can undermine data quality. This intrinsic nature has driven the exploration of innovative solutions to mitigate its impact to the forefront of research [190]. Despite the research efforts, noise can disrupt data to such an extent that it erodes its inherent structure, rendering it a promising avenue for anonymization techniques. This approach inherently bears a resemblance to the addition of noise in 2D images. In this category, uniform noise was arbitrarily selected, and the approach is designated as *UniformNoise*. Alternatively, other noise types, such as Gaussian noise and Poisson noise, could have been taken into consideration.

3.2.1.3 Warping-Based

Point cloud warping involves altering a point cloud's structure using geometric transformations. These transformations can be selectively applied to specific points or uniformly across the entire cloud. In this process, individual points within the point cloud are adjusted, leading to changes in both the overall shape and visual characteristics. Furthermore, the possibility exists to amalgamate multiple transformations, allowing more complex and customized modifications. This approach inherently bears a resemblance to the warping techniques in 2D images. In this category, the selection was made for the tapering transformation as the method for shape deformation. It will be denoted as *Tapering*. There are also other options for transformations, including bending and twisting, that could have been considered.

3.2.1.4 Morphing-Based

Point cloud face morphing refers to the process of smoothly transforming one facial point cloud into another while achieving a gradual and realistic change in facial appearance. This transformation involves adjusting the spatial coordinates of the points within the point clouds to blend facial features between the two states, ensuring a seamless transition. This method highlights the potential for anonymization and privacy protection, particularly when changing facial states entirely. Furthermore, the analysis of the work [172], as outlined in the preceding section, appears to support its potential. Although face morphing may include warping to transform one face into another, warping alone does not inherently produce a morphing effect, as these are separate and distinct concepts. This approach shares similarities with face morphing techniques used in 2D space. In this category, a new pipeline was designed to operate in point clouds and named *Merge2Faces*.

3.2.1.5 Smoothing-Based

Point cloud smoothing is a computational technique commonly employed to reduce noise and irregularities, resulting in a more regular and visually pleasing point cloud representation. However, increasing the intensity of the smoothing effect may lead to significant information loss, potentially causing facial features to vanish, thus enhancing the anonymization potential. This approach shares similarities with smoothing techniques used in images, such as blurring and smoothing filters. In this category, a method called *SmoothkNN* was proposed, which is based on the k-nearest neighbor framework. Another potential smoothing-based approach that could have been considered involves using an average voxel filter as a means of recreating the operation of a standard smoothing filter in the 2D space.

3.2.1.6 Point Operations-Based

This category was established to encompass techniques that defy classification within any specific class. These methods involve manipulations of points that alter the structure and shape of the point cloud in various ways. The implemented technique within this group is referred to as *Point-Mesh-Point* (PMP).

3.2.2 Algorithms Description

According to the selected anonymization techniques outlined earlier, this section will provide an explanation of their functioning. To facilitate their understanding, some preliminary concepts and mathematical notation are introduced to provide additional context.

Let $P \in \mathbb{R}^3$ represent a finite point set of a subject, and let $p \in P$ denote an arbitrary point. Each point p is defined by two key components: its 3D coordinates, p_{coord}, and its color, p_{color}. While specific techniques may alter both components, others may target only one, leaving the other unaffected. The RGB color model characterizes the color component p_{color}, utilizing an 8-bit representation per channel. Actions performed on the p_{color} component, such as computing the mean color of a point set, are equivalent to calculating each RGB component's mean value.

3.2.2.1 Sampling-Based

CentroidVoxel The sampling technique *CentroidVoxel* groups the facial points in a voxel grid[1]. A voxel grid is a 3D depiction of space that partitions it into a systematic arrangement of compact volumetric entities known as voxels. Then, each voxel is represented by its centroid point[2], and its color is determined by calculating the average colors of all the points contained within the voxel. The pseudo-code outlining this technique can be found in Algorithm 1.

The parameter governing the sampling intensity is *ratio*, representing the size of each voxel and ranging from $]0, \infty]$. As this value increases, more data is compressed onto the centroid point of each voxel. In an extreme scenario where the entirety of the subject's head is confined within a single voxel, the resultant point cloud would consist of an isolated point positioned at the centroid of the voxel, displaying the average color derived from the complete set of facial points.

[1] The process is known as voxelization, consisting of converting the continuous geometric representation of a point cloud into a discrete representation using voxels.

[2] The centroid corresponds to the geometric center of the voxel, irrespective of the points it contains.

Algorithm 1 *CentroidVoxel*

Input: P: target point cloud, *ratio*: voxel size

Output P_{anon}: sampled point cloud with *CentroidVoxel*

```
$voxel\_grid$ ← generate a voxel grid from point cloud $P$ with voxel size equal to $ratio$
for voxel $v \in voxel\_grid$ do
    $centroid$ ← compute the centroid of voxel $v$
    $mean\_RGB$ ← compute the average color value of the points inside voxel $v$
    $p'$ ← create point with coordinates $centroid$ and color $mean\_RGB$
    $P_{anon} \leftarrow P_{anon} \cup p'$
end for
return $P_{anon}$
```

3.2.2.2 Noise-Based

UniformNoise The *UniformNoise* technique involves adding random values drawn from a uniform distribution to the coordinates of each point. As a result, the 3D noise's individual coordinates adhere to a uniform distribution, denoted as $\mathcal{U}(a, b), a, b \in \mathbb{R} \wedge a < b$. To align the point cloud with its original position, an essential step is taken: the coordinates of each point are translated by subtracting the mean of the uniform distribution, calculated as $(a + b)/2$. This adjustment prevents any horizontal or vertical shifts in the point cloud's placement, especially when utilizing parameter values a and b that generate a non-centered uniform distribution interval around the origin 0, given by $|a| \neq |b|$.

Algorithm 2 *GaussianNoise*

Input: P: target point cloud, a: uniform distribution parameter, b: uniform distribution parameter

Output P_{anon}: noisy point cloud with *UniformNoise*

```
for point $p \in$ point cloud $P$ do
    $unif\_noise$ ← generate a random 3D vector with each coordinate following a uniform distribution, $\mathcal{U}(a, b)$
    $p$ ← add the noise $unif\_noise$ to the point $p$ coordinates
    $P_{anon} \leftarrow P_{anon} \cup p'$
end for
$P_{anon}$ ← subtract $(a + b)/2$ from all the point coordinates of $P_{anon}$
return $P_{anon}$
```

The regulatory parameters responsible for controlling the noise intensity are denoted as a and b, representing the lower and upper bounds of the uniform distribution, respectively. Both parameters extend across the interval $]\infty, \infty[$, with the condition $a < b$ duly observed.

3.2.2.3 Warping-Based

Tapering In mathematics, tapering represents a form of shape deformation characterized by a nonconstant scaling according to a specified tapering function. This transformation differentially changes the length of two global components while keeping the length of the third unchanged [191]. It constitutes a higher-order deformation, capable of yielding nonlinear deformations. The introduction of nonlinearity stems from the utilization of a nonconstant transformation matrix denoted as T. The subsequent equation represents its application to a point $p \in P$, yielding a new point p'.

$$p' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = T \cdot p = \begin{bmatrix} r & 0 & 0 \\ 0 & r & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} rx \\ ry \\ z \end{bmatrix} \tag{3.1}$$

where $r = f(z)$ symbolizes the scaling factor with f representing the tapering function. If $f(z) = 1$, the deformed affected portion of the object remains unaltered; for $f'(z) > 0$, the object experiences enlargement, while $f'(z) < 0$ leads to a reduction in size. The algorithmic representation of this technique is outlined in Algorithm 3, incorporating a minor modification by calculating the scaling factor within a specific interval rather than solely based on the z-coordinates. This refinement enhances control over the process.

Algorithm 3 *Tapering*

```
Input: $P$: target point cloud,$f$: tapering function, $val_m, val_M$: minimum and maximum interval extremes
Output $P_{anon}$: warped point cloud with Tapering
$x\_values \leftarrow$ generate $\|P\|$ equally spaced points over the interval$[-val_m, val_M]$
for point $p \in$ z-axis sorted point cloud $P$ do
    $M \leftarrow$ compute the transformation matrix with $f(x\_values[\text{index of } p])$
    $p' \leftarrow M \cdot p$, matrix multiplication
    $P_{anon} \leftarrow P_{anon} \cup p'$
end for
return $P_{anon}$
```

The regulating parameters consist of the tapering function f and the function's restricted domain, $[-val_m, val_M]$, signifying a subset of the original domain of definition within which the values of f are computed. These two parameters jointly govern the transformation, with the function closely linked to the appearance of deformation. This connection is also influenced by the interval values that regulate the function's outputs. The tapering function $f(x) = 1$ results in no alteration of the data.

However, the inverse transformation is described by the subsequent equation, implying the potential to reverse the tapering effect. In the context of anonymization, this attribute could expose a vulnerability to malicious attackers.

$$\begin{aligned} r(Z) &= f(Z), \\ x &= X/r, \\ y &= Y/r, \\ z &= Z \end{aligned} \tag{3.2}$$

3.2.2.4 Morphing-Based

Merge2Faces The *Merge2Faces* technique combines the attributes of two facial point clouds by averaging their points' coordinates and colors. Within this framework, the point cloud designated for anonymization is called the *source* face, while the other is defined as the *target* face. The *Merge2Faces* algorithm is composed of two distinct stages.

The initial stage is dedicated to registering the *source* and *target* faces. The alignment process commences with a coarse global registration executed using the Random Sample Consensus (RANSAC) algorithm [192]. During each iteration of RANSAC, a subset of randomly chosen points from the *source* face is selected. By employing the nearest neighbor query in the Fast Point Feature Histograms (FPFH) feature space, points exhibiting similar local geometric structures are identified within the *source* face. FPFH features, as described by Ruse *et al.* [193], encompass robust multi-dimensional descriptors that encapsulate the local geometry surrounding a point. After a pruning step, the derived matches contribute to the computation of a uniformly applied transformation across all points. This transformation is systematically applied to the *source* face in successive iterations, attempting to attain its alignment with the *source* face. Subsequently, the outcome of global registration is further refined through local registration. This refinement employs a variant of the ICP algorithm termed Point-to-Plane ICP [194], which Rusinkiewicz and Levoy [195] demonstrated to possess a faster convergence rate when contrasted with other ICP variants [114, 194].

The Point-to-Plane ICP algorithm iterates through the following steps until a predetermined stoppage criterion is satisfied:

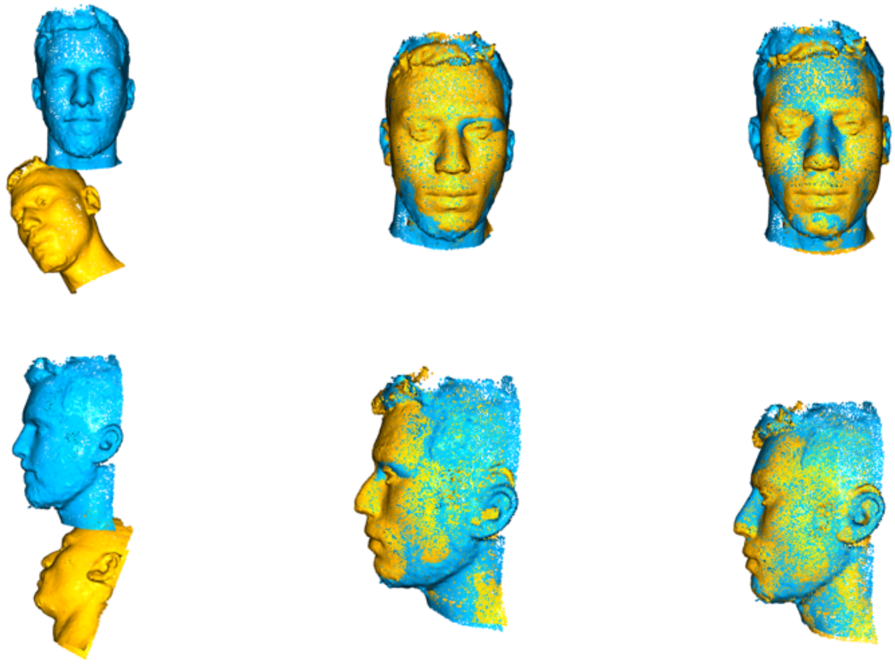
1. Select $\mathcal{K} = \{(p, q) | p \in source, q \in target\}$, comprising the closest matched points between the *source* and *target* point clouds.

2. Apply the current transformation matrix T to the *target* point cloud.

3. Update the transformation matrix T by minimizing an objective function $E(T)$, as defined in Equation 3.3, over the selected points.

$$E(T) = \sum_{(p,q)\in\mathcal{K}} \left((p - Tq) \cdot n_p\right)^2 \tag{3.3}$$

where n_p represents the normal vector of point p. Figure 3.2 illustrates the initial stage of the *Merge2Faces* technique using a toy example featuring two identities of the iPhone-PLYv3 dataset, introduced in a later chapter.



(A) Initial state of the point clouds.

(B) Global registration using RANSAC.

(C) Local refinement using ICP.

FIGURE 3.2: First stage of the *Merge2Faces* algorithm for front and profile views - the blue and yellow point clouds represent the *source* and *target* faces, respectively.

The second stage assumes the alignment of both faces and executes a weighted average for the coordinates and color of each point on the *source* face, along with the corresponding closest point on the *target* face.

The pseudo-code outlining this technique is provided in Algorithm 4.

Algorithm 4 *Merge2Faces*

```
Input: $P$: source point cloud, $w$: weighted average weight, $target$: facial point cloud
Output $P_{anon}$: morphed point cloud with Merge2Faces
target_global ← global registration using RANSAC between $P$ and $target$
target_local ← local registration using Point-to-Plane ICP between $P$ and target_global
for point $p \in$ point cloud $P$ do
    $p_s$ ← select the nearest neighbor of $p$ from target_local
    mean_coord ← weighted mean between the coordinates of $p$ and $p_s$, with weight $w$
    mean_RGB ← weighted mean between the RGB colors of $p$ and $p_s$, with weight $w$
    $p'$ ← update point with coordinates mean_coord and mean_RGB color
    $P_{anon} \leftarrow P_{anon} \cup p'$
end for
return $P_{anon}$
```

The regulatory parameters responsible for controlling the attributes change of the *source* face encompass *target*, referring to the point cloud of the target face to be merged, and the weight w assigned to the *target* face when calculating the weighted average with the *source* face. The weight w lies within the range of $[0, 1]$. As w augments, the distinctive facial characteristics of the *target* face gradually integrate into the *source* face, altering its appearance towards a more distant version. A $w = 0$ value yields an identity transformation to the point cloud, while $w = 1$ results in a complete substitution of the *source* face with the *target* face, commonly referred to as face swapping.

3.2.2.5 Smoothing-Based

SmoothKNN The *SmoothKNN* algorithm embodies a smoothing approach akin to an average filter. This technique involves substituting each point within the facial point cloud with the mean coordinate and color attributes of its k-nearest neighbors (the point is contained within the set of k points). The pseudo-code delineating this technique can be located in Algorithm 5.

Algorithm 5 *SmoothKNN*

```
Input: $P$: target point cloud, $k$: number of neighbours
Output $P_{anon}$: smoothed point cloud with SmoothKNN
for point $p \in$ point cloud $P$ do
    $S_k$ ← select the $k$ nearest neighbors of $p$ from $P$ (including point p)
    mean_coord ← compute the average point's coordinates of set $S_k$
    mean_RGB ← compute the average RGB color value of the points of set $S_k$
    $p'$ ← update point with coordinates mean_coord and mean_RGB color
    $P_{anon} \leftarrow P_{anon} \cup p'$
end for
return $P_{anon}$
```

The regulating parameter that manages the degree of smoothing is designated as k, signifying the count of neighbors taken into account for the averaging computations. The parameter k spans from 1 to $|\mathbb{P}|$, where $k = 1$ represents the identity transformation[1], and $k = |\mathbb{P}|$ yields a singular point with coordinates and color equivalent to the mean value among all the facial points. A larger k corresponds to a heightened number of adjacent points, resulting in a more pronounced smoothing effect. Conversely, a smaller value retains more intricacies and offers less pronounced smoothing.

3.2.2.6 Point Operations-Based

PMP The *PMP* algorithm undergoes a sequential transformation of the facial point cloud. It begins by converting the point cloud into an α-shape mesh and subsequently reverting this α-shape mesh back into a point cloud.

The α-shape of a finite set of points is a polytope, a geometric object characterized by its flat sides, which is uniquely determined by the point set and the parameter α [196]. Consider a finite point set $S \in \mathbb{R}^3$ and an $\alpha \in \mathbb{R}$ satisfying the constraint $0 \leq \alpha \leq \infty$. When $\alpha = \infty$, the α-shape coincides with the convex hull of S, diminishing in size as α decreases, thus giving rise to cavities. Edelsbrunner and Mücke [197] provide an intuitive definition of α-shapes as a broader conception of the convex hull for a point set, wherein elements vanish as α decreases sufficiently for a sphere with radius α to encompass its space without encapsulating any of the points in S.

To gain visual insight into the α-shape of a point set for varying α values, refer to Figure 3.3. The erasing sphere is shown to the right of the shape. The transformation of the original point cloud into an α-shape induces a loss of information regarding the geometric contours of the face when converting back to a point cloud, thereby facilitating the anonymization of the subject.

The pseudo-code outlining this technique is found in Algorithm 6.

Algorithm 6 *PMP*

```
Input: $P$: target point cloud, $alpha$: trade-off parameter, $n$: final number of points
Output $P_{anon}$: altered point cloud with PMP
$A \leftarrow$ convert the point cloud into an alpha shape with $alpha$
$P_{anon} \leftarrow$ convert the $alpha$-shape $A$ back to a point cloud with $n$ points
return $P_{anon}$
```

[1] The closest point to a point is the point itself.

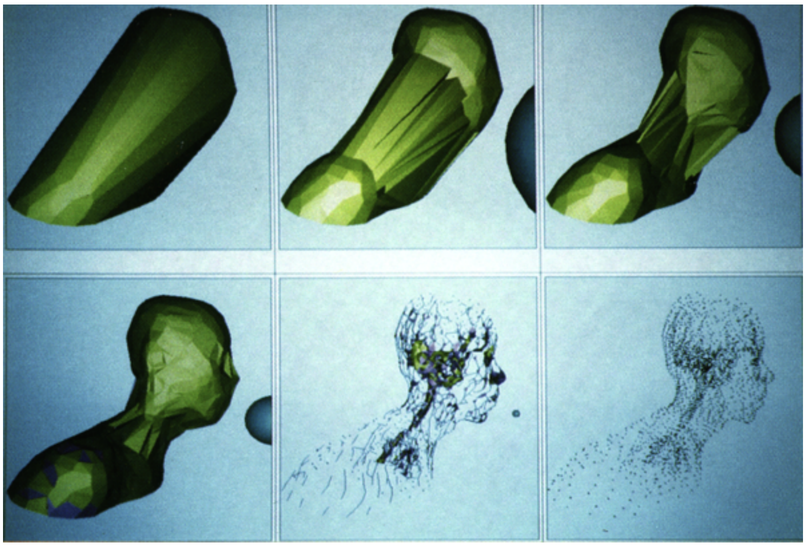


FIGURE 3.3: The effect of the α on the α-shape of a set of points. Extracted from [197].

The regulatory parameters governing the information loss are denoted as α, a real positive number that dictates the trade-off level of fineness for the α-shape, and an integer n, signifying the number of points resulting from the conversion of the α-shape back into a point cloud.

3.3 Face Recognition Model

As highlighted in Section 2, the privacy assessment of face anonymization techniques requires using a face recognition model, which assumes the role of an attacker, regardless of the specific approach paradigm employed. Whether it involves *naïve recognition*, *reverse recognition*, *parrot recognition*, or the *reversibility recognition*, a face recognition model is integral to the evaluation process. As such, the cornerstone of a robust privacy evaluation hinges on selecting a strong face recognition attacker model, striving to re-identify individuals from anonymized facial data. For instance, using a weaker attacker model may undermine the simulation of real-world threats, potentially resulting in misleading conclusions about the security provided by the anonymization technique.

3.3.1 Algorithm Selection

The selection process of the face recognition model was guided by a predefined set of criteria, established *a priori*. These criteria have been organized in descending order of significance. Each criterion is elaborated upon below, accompanied by an explanation for its inclusion

1. *State-of-the-art model:* These models serve as robust potential adversaries, being at the forefront of performance and advancement within the field. Their utilization mirrors real-world scenarios where sophisticated recognition systems are deployed, unveiling potential vulnerabilities and limitations that may not surface when employing outdated or less accurate models.
2. *Open-source code model:* Adopting open-source code provides an existing framework that can be readily utilized, streamlining the evaluation process. This approach fosters a focused evaluation of the face anonymization technique, relieving the burden of creating the model from scratch. Moreover, open-source code promotes research transparency, providing a transparent implementation of the recognition model and enabling reproducibility by fellow researchers.
3. *Pre-trained model:* Pre-trained models come equipped with learned parameters and weights, making them readily applicable without requiring extensive training. This characteristic conserves valuable time and computational resources. Pre-trained models are typically trained on extensive datasets, and their rigorous training, evaluation, and validation processes also inspire confidence in their reliability.

Since the evaluated anonymization techniques operate on 3D point clouds, opting for a 3D face recognition model seems most logical. However, while adhering to the above-mentioned criteria, an obstacle emerged involving criterion 2. It was impossible to locate any open-source implementations of 3D face recognition models, let alone state-of-the-art ones. In light of this challenge, the proposed solution is to convert the 3D point clouds into 2D images using a 3D to 2D projection and proceed with the evaluation within the lower-dimensional space. The prime performance of 2D face recognition (even in the most demanding scenarios) is attributed to the creation of large-scale training datasets containing millions of images, which empower these models and render them suitable for the task. The extensive body of research in this realm, as expounded upon in Section 2.3, corroborates the effectiveness of 2D methods. The proposed solution leverages the strengths of state-of-the-art face recognition models designed for the 2D space, allowing the application of advanced face recognition methods to evaluate the proposed 3D anonymization techniques.

The selection of the state-of-the-art 2D face recognition model was founded upon the outcomes of the previously described 2D face recognition benchmarks detailed in Section 2.3.3. The LFW dataset is a classic and extensively used benchmark, presenting the most exhaustive and up-to-date list of algorithms. Consequently, it was chosen as the benchmark for this study. In Section 2.3.4, the state-of-the-art results for the selected benchmark have already been presented, rendering them the prime candidates for selection. A compilation of potential algorithms was gathered and summarized concisely in Table 3.1, taking into account their performance on LFW, the availability of open-source code (official or unofficial), and the existence of pre-trained versions.

TABLE 3.1: Candidate of 2D face recognition attacker models.

Method	Code
FaceNet [109]	Unofficial
DeepID [100]	Unofficial
DeepFace [105]	Unofficial
ArcFace [145]	Official

From the candidate models, Additive Angular Margin Loss (ArcFace) [145] was chosen due to its widespread usage in anonymization-related studies [112, 198, 199] and excellent performance. It is reasonable to assume that selecting alternative models would likely yield comparable outcomes. Nevertheless, exploring alternative options for the attacker model presents an intriguing avenue for future research.

3.3.2 Algorithm Description

Deng *et al.* [145] introduced ArcFace in 2019, an advanced face recognition model that achieved state-of-the-art results. The model's superior performance has been demonstrated through extensive experiments on different benchmarks, such as LFW and MegaFace. The researchers' primary contributions to the field of face recognition center around the introduction of a novel loss function known as Additive Angular Margin Loss (ArcFace). This loss function plays a crucial role in acquiring highly discriminative features for improved face recognition. The framework for ArcFace is visually depicted in Figure 3.4.

The ArcFace loss further improves the face recognition model's discriminative power by enforcing intra-class compactness and inter-class difference of embeddings on the hypersphere surface and stabilizes the training process. The function is categorized as a type

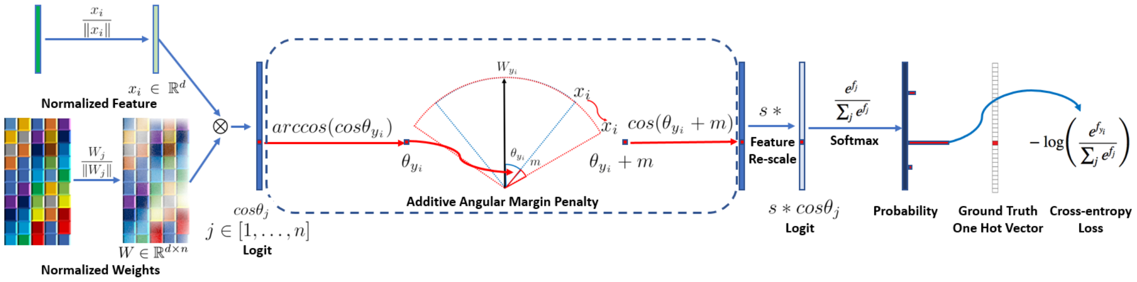


FIGURE 3.4: Framework of the ArcFace loss function. Extracted from [145].

of angular separability loss from the ones presented in Section 2.3.1, revolving around the use of an angular margin penalty based on the cosine similarity measure. Arcface is a modification of the most widely used classification loss, the Cross-entropy loss function, which is mathematically defined as:

$$\sigma(z_{y_i}) = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{z_{y_i}}}{\sum_{j=1}^{n} e^{z_j}} \tag{3.4}$$

where the z_{y_i} is the logit[1] of the y_i-th class that is divided by the sum over all the logits in the final layer. In an effort to enhance both intraclass similarity and inter-class diversity within the previously mentioned underoptimized loss, the authors have introduced several modifications to the function. Accordingly, the loss undergoes a normalization step on features and weights. A logit can be described with the embeddings, weights, and biases of the laster neural network layer as:

$$z_j = W_j^T x_i + b_j \tag{3.5}$$

where $x_i \in \mathbb{R}^d$ denotes the embeddings of the i-th sample, $W_j \in \mathbb{R}^d$ denotes the j-th column of the weight $W \in \mathbb{R}^{d \times n}$ and $b_j \in \mathbb{R}^n$ the bias term, both responsible for the j-th logit. Fixing the bias to $b_j = 0$, the dot product of the W_j and x_i is geometrically given by:

$$W_j^T x_i = \|W_j\| \|x_i\| cos\theta_j \tag{3.6}$$

After performing L2-normalization on both $\|W_j\|$ and $\|x_i\|$, followed by re-scaling to s, the predictions are now solely dependent on the angle θ between the feature and the weight, causing the learned embedding to be distributed on a hypersphere with a radius of s. Thus, obtaining:

[1] In a neural network, the logits (or raw scores) are computed for each class as the output of the final layer, representing the model's unnormalized confidence levels for each class based on a given input sample.

$$W_j^T x_i = scos\theta_j \tag{3.7}$$

Finally, introducing an additive angular margin penalty m between x_i and W_{y_i} to improve the intra-class compactness and inter-class discrepancy, the mathematical formulation of ArcFace is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{3.8}$$

3.4 Face Detection Model

As outlined in Section 2.3, the initial phase of a conventional end-to-end face recognition system encompasses the detection of all faces within the input image. These identified faces delineate a set of sub-regions in the image, demarcated by bounding box coordinates supplied by a facial detection module. These are the only regions undertaken by the subsequent pipeline stages of the face recognition system. Hence, the model plays a crucial role in the evaluation strategy, serving as an integral component of the recognition model pipeline. In addition, the face detection model integrates the utility evaluation as two of the metrics rely on its output.

3.4.1 Algorithm Selection

After methodically selecting the face recognition model and evaluating its impact on the methodology design, the next step involves the utilization of a 2D-based face detection framework. Similar to the criteria employed for the face recognition algorithm, the selection of the face detection model follows a comparable principle, prioritizing state-of-the-art models with available open-source code and, ideally, pre-trained weights. In Section 2.2.4, a selection of state-of-the-art models from the widely recognized WIDER FACE benchmark was introduced to narrow down potential candidate models. Subsequently, after cross-referencing these models with those included in the same project as the ArcFace model implementation detailed in the following chapter, the RetinaFace [29] model was chosen. In particular, the selection of RetinaFace was based on its benchmark, outperforming the seven other models within the implementation project and establishing itself as the most promising candidate.

3.4.2 Algorithm Description

In 2019, Deng *et al.* [29] introduced RetinaFace, a single-shot detector designed to address multiple face-related tasks within a unified framework. The model encompasses bounding box prediction, 2D facial landmark localization, and 3D vertices regression, offering a solution to these interconnected tasks by taking advantage of joint extra-supervised and self-supervised multi-task learning. For example, the semantic points inclusion in the facial landmark localization process significantly enhances the accuracy of box prediction during face detection. For training purposes, Deng *et al.* leveraged the WIDER FACE dataset, which required additional processing steps involving the manual annotation of five 2D facial landmarks on each image due to their absence. The approach resulted in superior performance compared to the prevailing state-of-the-art models on the WIDER FACE benchmark. The model's architecture, illustrated in Figure 3.5, comprises three pivotal components: the feature pyramid network, the context head model, and the cascade multi-task loss.

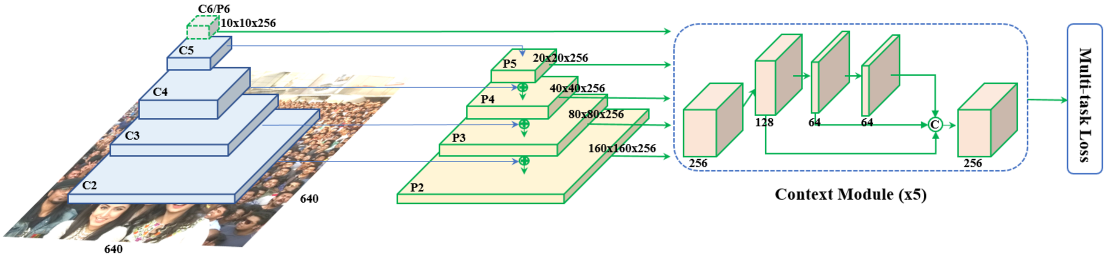


FIGURE 3.5: RetinaFace face localization framework. Extracted from [29].

The pyramid network takes the image as its input and employs top-down and lateral connections derived from a ResNet backbone to generate feature maps at five different scales. The bottom-up pathway in the ResNet backbone follows the standard network forward pass, extracting features from the input image and producing feature maps with decreasing spatial resolution but escalating semantic information. Conversely, the top-down pathway complements this process. It begins from the highest layer of the backbone, possessing the greatest spatial resolution and highest semantically rich features, whose feature maps are aligned with the resolution of corresponding lower-level feature maps obtained from the bottom-up pathway. This iterative procedure establishes lateral connections, merging the upsampled feature maps with their lower-level counterparts. This fusion integrates high-level semantic content with precise spatial details, generating a feature pyramid with multi-scale feature maps.

Within the lateral connection and context modules, a Deformable Convolution Network (DCN) is employed. Unlike conventional fixed-grid convolutions, DCN introduces learnable offsets to convolutional filters, enabling them to adaptively modify their receptive fields in response to input data variations. The deformable convolution operation empowers the network to adjust its receptive field flexibly, facilitating more flexible and precise feature learning and, consequently, leading to more accurate face detection.

3.5 De-Anonymization Model

As previously noted, *reversibility recognition* is a paradigmatic framework for evaluating anonymization methods, as introduced by Todt *et al.* [180]. In contrast to the conventional *naïve recognition* approach, the *reversibility recognition* paradigm shifts its focus towards evaluating the extent to which the anonymizations are reversible by performing the de-anonymization of the images using a powerful de-anonymization model. These de-anonymized images resemble the original images more closely and serve as the probe set for assessment, contrasting from the direct comparison between anonymized images and the gallery set of original images of the other paradigm.

The *reversibility recognition* ensures a more accurate assessment of anonymization techniques, providing a deeper insight into their privacy protection effectiveness and potential vulnerabilities when confronted with sophisticated de-anonymization attempts. In their study, Todt *et al.* [180] adopted an autoencoder as the de-anonymization model due to its inherent versatility and robust generalization capabilities. The researchers chose a unified autoencoder model over creating individualized de-anonymization models in scenarios involving multiple sets of diverse image groups originating from distinct anonymization techniques. This approach ensures temporal efficiency and delivers commendable results, showcasing the autoencoder's effectiveness in seamlessly managing diverse de-anonymization tasks, as observed within this work.

Regarding Autoencoders, the problem is formally defined by Bank *et al.* [200] as learning the functions $A : \mathbb{R}^n \to \mathbb{R}^p$, and $B : \mathbb{R}^p \to \mathbb{R}^n$, representing the encoder and decoder, respectively, such that they satisfy:

$$\text{argmin}_{A,B} E\left[\triangle(x, B \circ A(x))\right] \tag{3.9}$$

where E denotes the expectation taken over the distribution of x, while $\triangle$ represents the reconstruction loss function, quantifying the dissimilarity between the input to the encoder and the output from the decoder.

Todt *et al.* employed an under-complete autoencoder as the de-anonymization function. Under-complete autoencoder is a type of autoencoder architecture where the dimensionality of the latent space, also known as the bottleneck layer, is smaller than the dimensionality of the input data. The main characteristic of an under-complete autoencoder is that it learns to represent the input data in a compressed form, as the number of neurons in the bottleneck layer is less than the number of neurons in the input and output layers. This compression representation fosters the capture of only the most essential features that benefit the learning of the dependencies in the data and aid data reconstruction. There are several other Autoencoder variants, as denoted in [200], one of them being VAEs which have already been addressed in this work with the opposite intent of anonymizing images.

3.5.1 Algorithm Description

The encoder component contains two convolutional layers followed by an activation function and a max pooling layer that halves the spatial dimensions of the input image. On the other hand, the decoder is designed symmetrically, containing two transposed convolutions that quadruple the spatial dimensions of the input data followed by an activation function, matching the pixel-wise input resolution. Refer to Figure 3.6 for a visual representation of the architecture of the de-anonymization Autoencoder model.

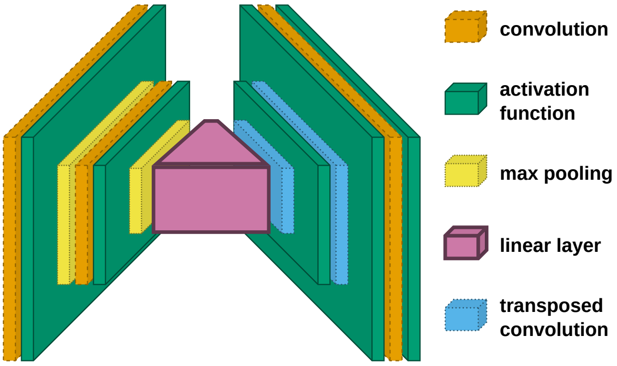


FIGURE 3.6: Architecture of the de-anonymization model employed in [180]. Extracted from [180].

Although not explicitly mentioning which activation functions and loss functions were incorporated into the model, the authors refer they tested the Sigmoid, Tanh, ReLu, and LeakyReLu for the former, and MSE, Mean Average Error (MAE), and SSIM for the latter.

3.6 Limitations

This study has made strides in exploring and implementing 3D face anonymization techniques. However, it is imperative to acknowledge and address the limitations encountered throughout the research journey.

Regarding the proposed 3D face anonymization methods, the initial aim was to extend state-of-the-art 2D anonymization approaches into the higher-dimensional space. Consequently, attention was directed towards generative deep learning-based models, given their superior performance and widespread usage. Nevertheless, challenges emerged due to time constraints and limited data availability, as generative models often require extensive datasets for effective training. Therefore, emphasis was placed on the pursuit of simpler 3D face anonymization solutions akin to traditional 2D methods. Also, given the field's early stage, commencing with elementary procedures, assessing their effectiveness, and subsequently advancing to more intricate methods is a mindful approach.

In addition to the previously mentioned limitation, another constraint that arose during the course of the study pertains to the search for suitable 3D face recognition models for the evaluation phase. The scarcity of models compelled a shift towards a 3D to 2D approach, where evaluations were conducted on projected anonymization images. This decision bears significant implications, as evaluation outcomes are deeply tied to projection quality, potentially introducing errors and bias. Nonetheless, precautions were undertaken to mitigate these effects and enhance the accuracy of the result. Recent research has predominantly favored 2D face recognition over its 3D counterpart. With increased attention and performance that surpasses human capabilities, 2D face recognition serves as a robust evaluator for attackers, ensuring the integrity of the results.

Moreover, this methodology is unsuitable for point clouds that lack a color component in their data points. Specifically, employing this methodology to assess an anonymization technique that exclusively introduces drastic color changes to the points (such as converting all points to a uniform color) without altering the geometric attributes of the point cloud would compromise the integrity of the recognition model when applied to the projected images. Nonetheless, a facial recognition model operating exclusively within the

3D space would remain unrestrained by the anonymization process, countering the misleading outcomes arising from the overly optimistic anonymization requirement in the projected evaluation. This anticipated scenario underscores the importance of exercising prudence in color manipulations and emphasizes the significance of minimizing their potential contribution to misleading results.

Chapter 4

Experimental Development and Implementation

This chapter examines the critical components of implementing and testing the 3D point cloud anonymization techniques conducted during the experiments. It begins with a complete characterization of the custom-made dataset used in this investigation. The dataset's data sources, preprocessing steps, rationale for selection, and encountered limitations are discussed. Then, the evaluation framework is detailed, outlining the metrics and procedures used for assessing the anonymization techniques. The technical aspects related to the parameters and configurations of the anonymization techniques and the models described in the last chapter are also presented. Lastly, some limitations of the experiments are addressed, acknowledging constraints and potential challenges encountered during the research and discussing their impact on the findings.

4.1 Dataset

This master thesis introduces a new 3D facial dataset named iPhonePLYv3[1]. Curated explicitly for this research purpose, the iPhonePLYv3 dataset comprises diverse subjects captured under favorable conditions, rendering it an ideal resource for comprehensive evaluations of anonymization algorithms and potentially suitable for other 3D face analysis tasks

[1] The iPhonePLYv3 dataset, the third iteration of its kind, was collected using an iPhone and comprises PLY-format point cloud data, in line with its name iPhone + PLY + v3.

4.1.1 Motivation

Due to the lack of datasets designed explicitly for anonymizing 3D point cloud facial data, the candidate datasets primarily focus on 3D face recognition model development. In Section 2.3.3, several 3D face recognition datasets have been discussed along with their specifications. Specifically, Table 2.4 reveals that although these datasets include multiple scans per subject capturing various facial expressions, poses, or occlusions to create robust face recognition models, the majority feature around 100 or fewer identities, which is a relatively small number. However, using unconstrained datasets to evaluate face anonymization techniques introduces complexities in privacy evaluation that restrict the performance of the face recognition attacker model, making it challenging to assess the actual effectiveness of the anonymization techniques accurately. Hence, evaluating anonymization techniques on controlled datasets is more suitable to address these challenges.

While the created dataset exhibits limited variability in facial expressions, poses, and occlusions, it serves as an evaluation tool for assessing anonymization methods under controlled conditions. These limitations enhance the effectiveness of an attacker's face recognition model, making the anonymization process more challenging. This implies that the attacker operates in optimal conditions, and if the anonymizations prove to be reliable in this scenario, their effectiveness can be extrapolated to real-world conditions with higher variability in facial expressions, poses, and occlusions. In addition, during the initial stages of this emerging field, focus on a controlled scenario to gain better insights into the implications of model parameters is preferable. Certain constraints related to dataset accessibility, cost, and unresponsiveness to requests further support the need for a new dataset for this research.

The new dataset, iPhonePLYv3, addresses identified issues providing favorable conditions for evaluating face anonymization techniques with an increased number of identities. The dataset includes color information in the point clouds, which is not universal in traditional 3D facial databases. However, the number of identities in the iPhonePLYv3 dataset remains relatively low, suggesting room for future advancements. The acquisition methodology of the iPhonePLYv3 is denoted by its simplicity which serves as another contribution to the broader research community. Refer to Figure 4.1 for an illustration of some existing 3D face datasets.

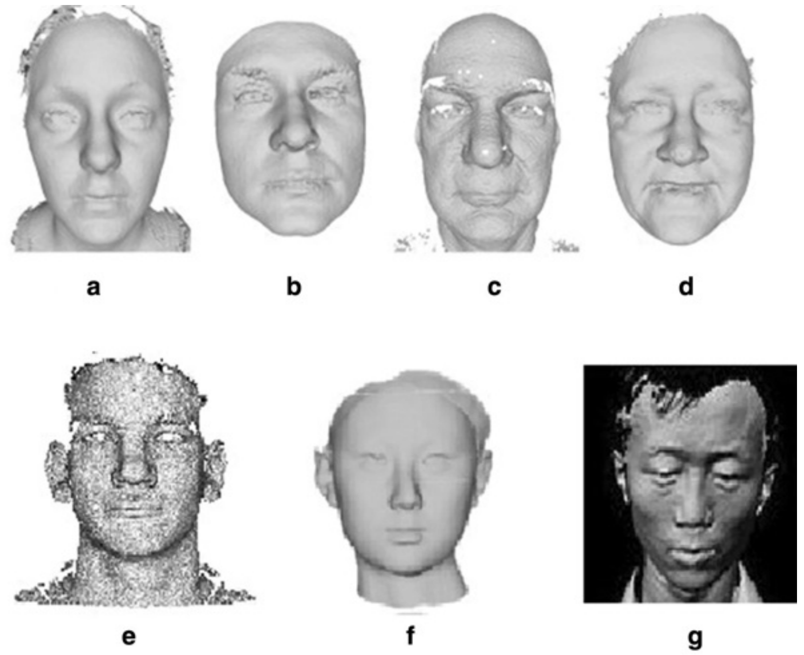


FIGURE 4.1: Comparison of several 3D face datasets. Extracted from [80].

4.1.2 Dataset Content

The iPhonePLYv3 consists of 201 subjects captured in a consistent pose, featuring a neutral facial expression free from occlusion. The dataset contains two distinct data types: 2D and 3D data, in the format of point clouds and high-resolution digital images with a pixel resolution of 3024 × 4032, both containing the facial information of the subjects. The point clouds depict a 180-degree region of the frontal part of the face, extending from one ear to the other[1]. The images, on the other hand, provide a frontal view of each face. Each subject has a corresponding point cloud and image pair collected consecutively.

The dataset's gender distribution is unbalanced, with 119 men and 82 women predominantly of Caucasian ethnicity. The subjects' ages range from 12 to 83 years, with a prominent concentration (172 subjects) in the 18 to 25 age group. The selection process of individuals was random, resulting in a high representativeness of facial characteristics in the dataset. This diversity is evident in subjects with beards, no facial hairs, dark skin color, light skin, baldness, and other unique features. For a visual representation, refer to Figure 4.2, which showcases a dataset sample featuring three subjects. The top row represents the high-resolution images, while the bottom is a screenshot of the point clouds, on the MeshLab [201] processing tool. For a more detailed description of the dataset structure and detailed specifications, refer to Appendix A.1.

[1] The back of the head was excluded due to its low compromise of a subject's identity.

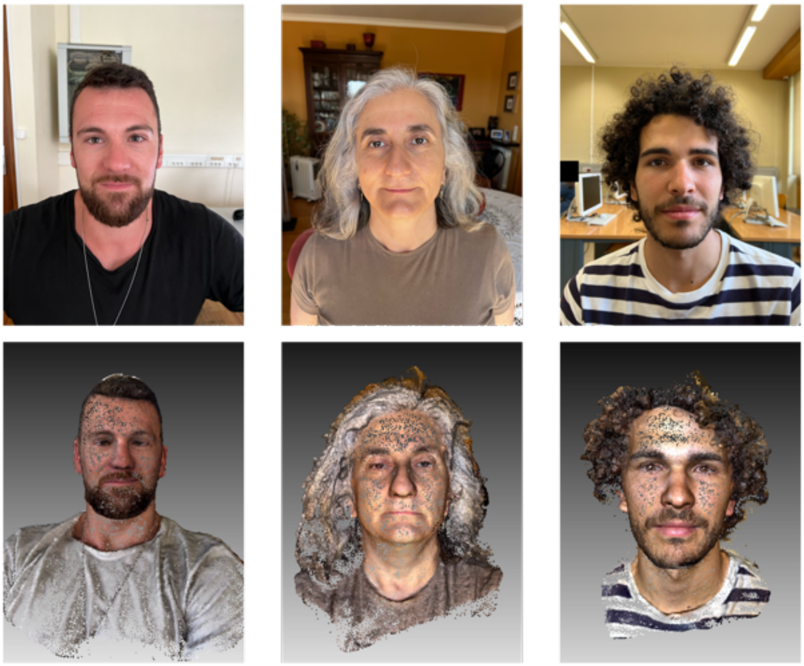


Figure 4.2: Illustration of three point cloud and image pairs for three subjects from the iPhonePLYv3 dataset.

4.1.3 Data Acquisition

The data acquisition process for the iPhonePLYv3 dataset introduces a straightforward methodology for collecting 3D facial data without the need for expensive equipment or extensive expertise. The 3D facial data was acquired using the *Scaniverse - 3D Scanner* mobile application on an iPhone 13 Pro device, while the 2D data was captured using the primary camera. In particular, from the *Scaniverse - 3D Scanner*, three files are obtained containing the information of the 3D facial model of each subject in a mesh format. The mesh undergoes multiple processing stages executed manually using the MeshLab [201] processing tool until it is converted into a point cloud, the final data format. The point cloud is registered onto a specific position and presents a uniform pitch, roll, and yaw for all the subjects. For a more detailed description of the conducted process, refer to Appendix A.2, and for the dataset evolution stages, refer to Appendix A.3.

The dataset was gathered with the explicit consent of all 201 subjects. Prior to collecting their facial data, a concise explanation of how it would be used within the scope of this thesis was provided, and each subject willingly accepted these terms. Upon collection, subjects were scanned in indoor and outdoor environments under favorable lighting conditions. Although occlusions were not explicitly accounted for, some subjects' hair may partially cover the forehead and a significant portion of the face profile, particularly for female subjects. Each scan took approximately 20 to 30 seconds, with subjects instructed to remain stationary and refrain from any movement about 30 centimeters away from the 3D scanner device. Other preventive measures forming the standardized protocol obeyed during the data acquisition process included:

- Standardizing the pose of all subjects;
- Instructing all subjects to exhibit a neutral facial expression[1];
- Restricting the use of any props that could obstruct facial features, such as glasses, masks, or other facial garments.

Despite the effort to maximize data quality, some records may still present defects related to movement, hair obstruction, or the 3D model generation by the application. Nevertheless, by adhering to these guidelines, the resulting data is of satisfactory quality.

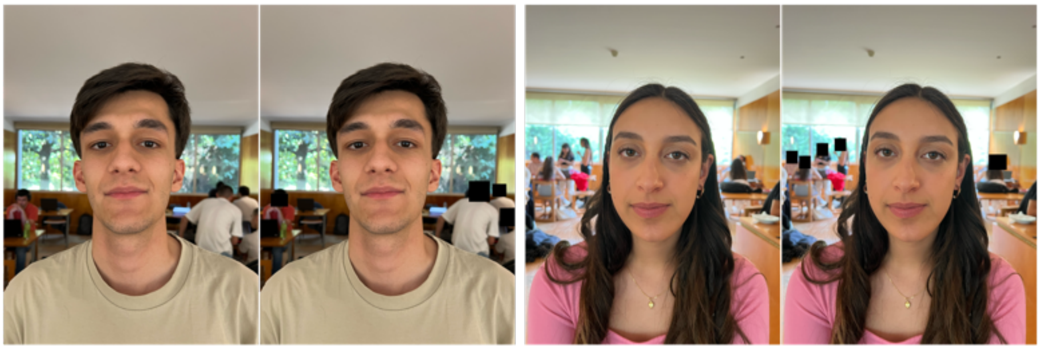
4.1.4 Data Preprocessing

To ensure the suitability of the data for the experiments, both the point cloud and image data underwent a data preprocessing process to facilitate the experimental design and enhance the dataset's usability for research endeavours.

4.1.4.1 2D Images

Regarding the 2D data corresponding to high-resolution frontal images of the subjects, the primary concern revolves around handling background information. While the capture of the 3D facial model is robust to background information, effectively capturing the relevant details only up to a certain distance from the capturing device, the same does not hold for the 2D images. Despite the efforts to optimize the data-capturing process, some subjects were acquired in crowded environments, resulting in unwanted facial information from other individuals in the image's background. To address this issue, these undesirable background elements were covered with black patches, as illustrated in Figure 4.3 for two subjects. This approach ensures that the evaluation process is facilitated by guaranteeing the presence of only one face per image.

[1] Some subjects may show a slight smile, but nothing exaggerated.



(A) id_058.

(B) id_082.

FIGURE 4.3: Preprocessing results of 2D data from two subjects in the iPhonePLYv3 dataset.

4.1.4.2 3D Point Clouds

The iPhonePLYv3 3D information comprises a collection of faces displayed in a 3D coordinate system as a set of points, such that their nose tips are located at the origin with coordinates $(0, 0, 0)$. These faces are oriented towards the x-axis orientation, and the y-axis serves as the reflection axis, dividing the facial region into two identical vertical halves. The above characteristics were crafted during the dataset creation and initial processing, acknowledging their significance in the subsequent stages of implementing and evaluating face anonymization techniques. However, this data may be affected by outliers and may also encompass irrelevant regions beyond the face, such as the neck and torso. Consequently, there is a variation in the number of points within these scans. Preprocessing is undertaken to address these issues with the aim of obtaining a point cloud exclusively containing the facial region. This facial region is the sole area requiring anonymization, as opposed to the other irrelevant regions such as the neck and hair. Additionally, it ensures that the final data is free from outliers that could potentially harm the anonymization techniques and promotes uniformity in the number of points across all point clouds, fostering consistent results. The preprocessing involves four sequential stages: Face segmentation $>$ Standardization $>$ Outlier removal $>$ Standardization, which are described as follows:

1. Face segmentation

   (a) *Bounding box cropping:* A bounding box with predefined coordinates is placed over the face point cloud, and all points outside the box are removed (first substage);

(b) *Sphere cropping:* A fixed-radius sphere centered at $(0, 0, 0)$ (nose tip location) is placed over the face point cloud, filtering the points outside the sphere - the filtering process results in two point sets, the inner facial set and the outer facial set (second sub-stage).

2. Standardization
   - *Sampling:* The Farthest Point Sampling (FPS) technique is used to reduce and standardize the number of points in the cloud.
3. Outlier removal
   - *Outlier removal:* Points that exhibit a greater distance from their neighbors than the average distance for the point cloud are removed.
4. Standardization
   - *Sampling:* The Farthest Point Sampling (FPS) technique is used to reduce and standardize the number of points in the cloud.

The anonymization process could be extended to cover the entire facial region, eliminating the need for the *Sphere cropping* step during the *Face Segmentation* stage. However, for superior results, the point cloud region was exclusively cropped to include the facial area requiring anonymization (the inner facial set), leaving other non-sensitive information untouched. The *Outlier removal* was executed subtly to avoid excessively reducing the scan density, only removing the most extreme cases. Similarly, both *Standardization* stages aimed to introduce minimal changes in the scan density by employing the FPS technique.

After completing the four sequential preprocessing steps, point clouds are obtained, containing only the segmented face region, free of outliers, and with a standardized number of points. For an illustration of the preprocessing stages employed for a subject in the dataset, refer to Figure 4.4. The point cloud of subject *id*_008 initially contained 44 726 points and underwent a reduction to 10 352 points after the three stages of preprocessing. At each stage, Face segmentation, standardization (the first), and outlier removal, the point cloud had 36 063, 23 756, and 10 500 points, respectively.

To gain further insights into the point cloud transformation results, Figure 4.5 illustrates the evolution of the number of points in the clouds at the following three stages: raw point clouds, *Face Segmentation*, *Outlier Removal*. While direct comparisons between

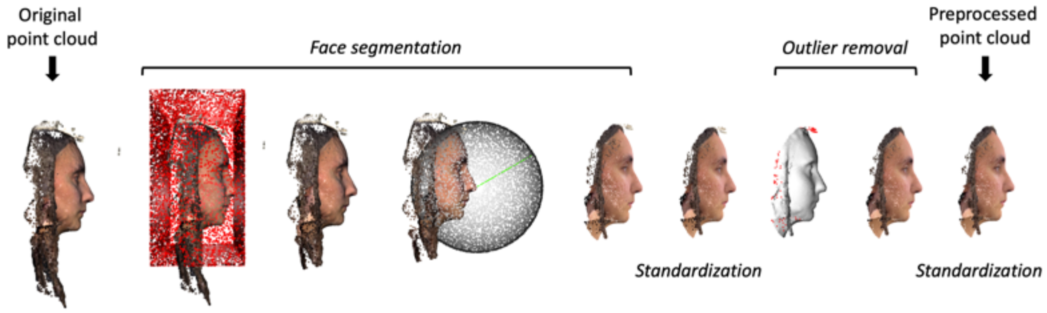
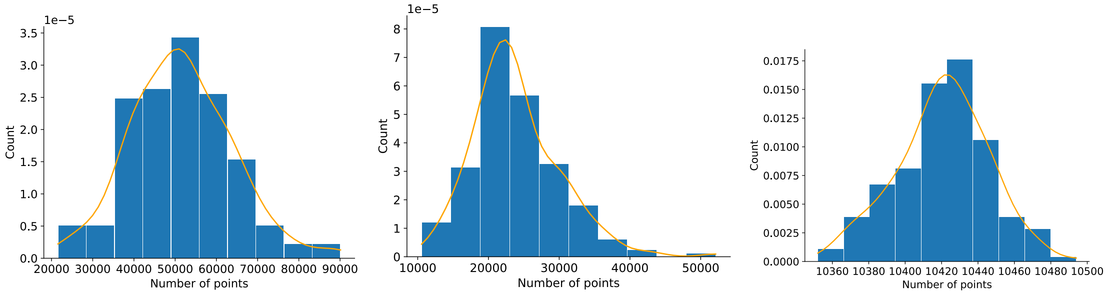


FIGURE 4.4: Point Cloud data preprocessing stages for subject *id*_008.

the plots are hindered by non-standardized axis limits, it is relevant to observe the overall variations in the number of points at each stage. The two *Standardization* stages are not depicted because all the clouds have a standardized number of points, corresponding to 10 500 and 10 352. The number of points of the raw point clouds does not directly represent the quality of the facial information in a scan. Scans with higher point counts may be attributed to the presence of additional hair or regions extending below the face, possibly encompassing the upper part of the torso, rather than necessarily indicating the higher facial resolution of the subjects.



(A) Raw point clouds. (B) *Face Segmentation.* (C) *Outlier Removal.*

FIGURE 4.5: Evolution of the number of points of the clouds through various stages of the preprocessing pipeline.

The subsequent information provides a detailed explanation of each stage, delving into their specific processes and the implications they have on the workflow.

Face Segmentation The face segmentation process consists of two steps, whose objective is segmenting the facial area that requires anonymization from non-sensitive regions and undesirable artifacts[1].

[1] Artifacts are undesirable and erroneous point clusters located at a distance from the subject's face.

The first step, called *Bounding box cropping*, involves retaining only a subset of the point cloud that closely relates to the subject's face while removing some of the rough artifacts. The dimensions of the bounding box employed to enclose the region of interest are of 30 centimeters in height, 24 centimeters in width, and 16 centimeters in depth. These values remain constant for all subjects. Despite variations in head sizes, they offer a simplified approximation of the overall dimensions, effectively narrowing the 3D space to focus more closely on the region of interest, which is the face. Mathematically, it is represented as follows:

$$B = \{(x, y, z) \mid -0.16 \leq x \leq 0.00, -0.13 \leq y \leq -0.17, -0.12 \leq z \leq 0.12\} \tag{4.1}$$

The second step, termed as *Sphere cropping*, is a simplified version of the face segmentation technique employed by Nair *et al.* [50]. While the original procedure considers a sphere centered on the nose tip with a radius determined by the nose length, the chosen approach simplifies this by adopting a fixed radius of 0.14 (14 centimeters), centered on the nose tip $(0, 0, 0)$. This simplification allows for more straightforward implementation while still achieving reasonable results in capturing the facial region of interest.

Outlier Removal The outlier removal process enhances the overall quality of the facial data by eliminating points that do not align with the facial structure due to sparseness reasons. Before its implementation, the point clouds are sampled using the FPS technique to 10 500 points, corresponding to the minimum number of points from all the segmented faces.

In essence, the employed *Outlier Removal* procedure is referred to as the *Statistical Outlier Removal*, eliminating points that exhibit a greater distance from their neighbors compared to the average distance within the entire point cloud, as depicted in Algorithm 7. The implementation requires two passes over the whole set of points in the cloud, and is regulated by two parameters that control the selectiveness of the points. The first parameter, *nb_neighbors*, determines the number of neighboring points considered when calculating the average distance for each point in the point cloud. The second parameter, *std_ratio*, sets the threshold level based on the standard deviation of the average distances across the entire point cloud. A lower value of *std_ratio* makes the point removal filter more aggressive, while a higher value makes it more lenient.

Algorithm 7 Statistical Outlier Removal

```
Input: $P$: point cloud, $K$: number of neighbors, $ratio$: standard deviation multiplier
Output: $P$: processed point cloud without outliers
for point $p \in P$ do
    $K_{nn} \leftarrow$ Find the $K$ nearest neighbors to point $p$
    $d \leftarrow$ compute the average distance from point $p$ to the points in $K_{nn}$
    $D \leftarrow D \cup \{d\}$, the set with all average distances $d$ for every point $p$
end for
$\mu_D \leftarrow$ compute the mean distance of the set $D$
$\sigma_D \leftarrow$ compute the standard deviation of the set $D$
$T \leftarrow \mu_D + \text{ratio} \times \sigma_D$, the threshold computation
for point $p \in P$ do
    $K_{nn} \leftarrow$ Find the $K$ nearest neighbors to point $p$
    $d \leftarrow$ compute the average distance from point $p$ to the points in $K_{nn}$
    if $d > T$ then
        Remove point $p$ from $P$
    else
        Keep the point $p$
    end if
end for
    return $P$
```

In this research, both parameters were empirically set to $nb\_neighbors = 50$ and $std\_ratio = 4$. This configuration sets the threshold value $T = \mu + 4\sigma$, where μ represents the mean of the mean distances of every point to its 50 nearest neighbors, and σ stands for the standard deviation. Thus, for each point with a mean distance to its 50 nearest neighbors of d, if $d > T$, the point is removed. Otherwise, it is kept.

Standardization The standardization stages ensure uniformity in the number of points across all point clouds, promoting algorithmic efficiency without compromising facial details. They leverage FPS, which is a greedy algorithm that iteratively selects points without repetition from point cloud data. This algorithm's objective is to maximize the distance between selected points, thus ensuring a maximum and well-distributed spatial coverage across the entirety of the original point set. It was selected as the sampling strategy due to its widespread use and ability to describe structural characteristics [202].

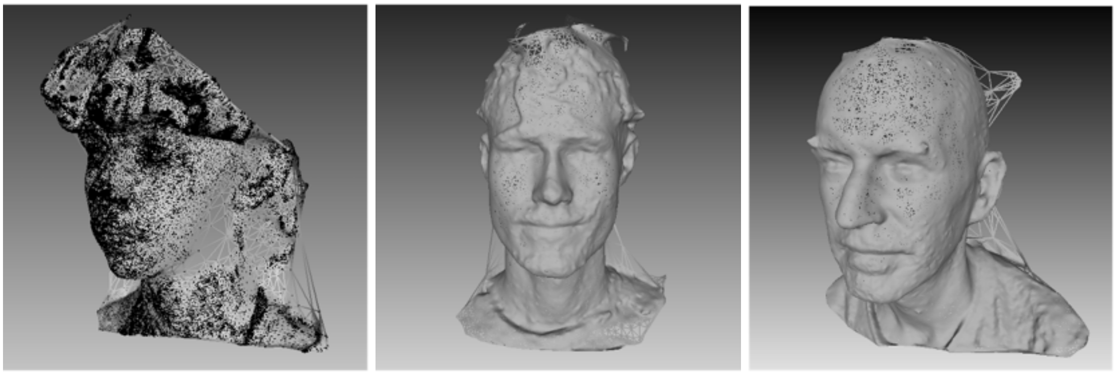
As previously mentioned, the first standardization reduces the number of points to 10 500, while the latter reduces it to 10,352. These values were chosen to ensure that the reduction in the number of points is minimized, setting the number of points equal to the point cloud with the lowest number after the previous implemented stages. Furthermore, this indicates that the outlier removal stage resulted in the removal of a maximum of 140 outliers. While further point reduction is feasible, preserving as much detail as possible is essential, especially for the 2D projections.

4.1.5 Data Limitations

Despite implementing various precautions to ensure high-quality data, certain errors persisted in the dataset, primarily attributed to the 3D digitizing system and setup conditions. The following are some examples of these errors:

- *Low-density:* Regions with a low density of points are mainly found in the chin area, but they may also appear on the forehead and cheeks;
- *Deformations:* Imperfections in the scans may be introduced by the presence of hair due to the inherent nature of their texture. Additionally, the eye region may present a concave shape instead of a convex one;
- *Movements:* Some scans within the dataset may exhibit perturbations caused by the movements of the subjects. Although the scans have a short duration, it is unrealistic to expect the subjects to remain completely motionless throughout the entire process.

Refer to Figure 4.6 for a visual representation of some of the mentioned occurring problems during the acquisition process. The images depict the point cloud uniformly colored or display the corresponding mesh from which the point cloud was derived. This visualization approach aims to enhance the perception of the issues at hand. These constraints are present in only 15% of the subjects, which does not compromise the data quality requirements for this research, as will be demonstrated later.



(A) Low-density region.

(B) Hair imperfections.

(C) Eyes deformation.

FIGURE 4.6: Limitations of the iPhonePLYv3 dataset.

4.2 Evaluation Strategy

The evaluation procedure of the anonymization techniques is conducted on the 2D space leveraging the 3D to 2D projections of the anonymized results provided by the anonymization techniques. As mentioned earlier, its design addresses the limitations of 3D face recognition, specifically addressing the lack of state-of-the-art open-source implementations and taking advantage of the more mature stage of development in the 2D space.

4.2.1 Preliminary Considerations

The experimental results will pertain to the six distinct anonymization techniques presented in the last chapter, each encompassing 35 configurations that correspond to the use of different regulating parameter values. These specific values were chosen based on the AUC metric to standardize results across all anonymization methods, enabling meaningful comparisons, as explained later. Altogether, these arrangements contribute to a total of 210 unique configurations devoted to testing.

Within each configuration, the *gallery set* is composed of high-resolution frontal images, while the *probe set* encompasses point cloud projections with and without anonymization applied. According to this aspect, the probe set is further divided into two subsets: the *baseline subset* and the *anonymization subset*. A visual representation of these different sets and their interrelationships can be seen in Figure 4.7. The smaller sets within the *anonymization subset* symbolize the diverse configurations, each involving multiple parameters linked to a specific technique.

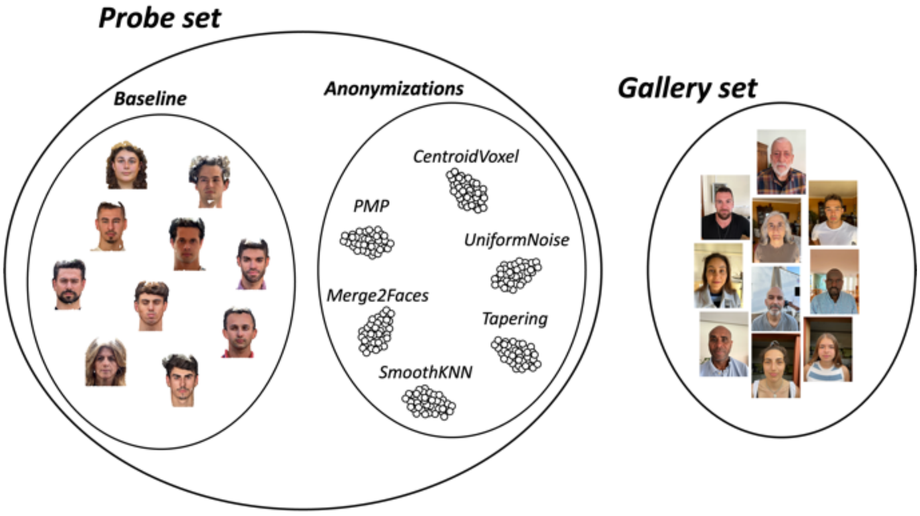


FIGURE 4.7: Data structure used for the anonymization techniques evaluation.

All sets contain facial data from 200 subjects, identified as *id_000* to *id_200*. However, *id_001* is an exception as it is excluded due to its integral role in the *Merge2Faces* anonymization, rendering it unsuitable for other experiments. Both the *gallery* and the *baseline subset* remain consistent across all configurations. The *baseline subset* serves a dual purpose: it acts as a benchmark for evaluating both privacy and utility metrics, measuring how the anonymization affects the data. Furthermore, it validates the overall experimental design.

In essence, every anonymization configuration is tested under specific metrics presented in the following section, generating a comprehensive overview of the technique's performance.

4.2.2 Description

The characterization of the effectiveness of an anonymization technique revolves around the balance between the privacy and utility it provides, involving the evaluation of the privacy-utility trade-off. Consequently, it is mandatory to design the assessment strategy for the privacy and utility components separately and then combine the two to obtain the trade-off.

Privacy Assessment Throughout the testing process, both face recognition modes, namely identification and verification, will be considered. This approach offers a more comprehensive and exhaustive understanding of the protective capacity of each technique.

The identification mode will be restricted to the closed-set protocol, simplifying its implementation and performance evaluation. Although it may not represent real-world scenarios, it avoids introducing additional complexity to the face recognition attacker model that may hinder the effectiveness of the primary objective of assessing the privacy strength of the anonymization technique. In Section 2.4.2, four distinct evaluation methodologies were presented for privacy assessment, each representing different paradigm attacker models. Gross *et al.* [160] demonstrated the effectiveness of *parrot recognition* in defeating *naïve anonymization* techniques, thereby reducing the privacy protection offered by such methods. In this research, most anonymization techniques are classified as *naïve anonymizations* due to their simplicity, making *parrot recognition* an appealing testing alternative. However, due to the methodology adopted in this research, the requirement of this attacker paradigm of applying the same distortion of the probe set in the gallery set is infeasible - the probe set results from a 3D distortion followed by a 3D to 2D projection, and the gallery set is inherently composed of 2D images. This contrast makes it impossible to recreate the same distortion effect in both sets, potentially compromising the effectiveness of the *parrot recognition* and, consequently, the quality of the evaluation results. As an alternative, the *reversibility recognition* has shown promising results in attacking face anonymization techniques [180], comparable to or even better than other approaches. Hence, it is considered a suitable choice for evaluation and it was selected for anonymization assessment. In addition, the *naïve anonymization* will also be considered as it is the widely used paradigm in the reviewed literature. By employing the two attacker paradigm frameworks, namely *naïve recognition* and *reversibility recognition*, a comprehensive evaluation of the privacy potential of the techniques is achieved.

Utility Assessment The utility assessment is closely related to the subsequent application of the data. Since there is no specific intent, the evaluation is expected to consider a broader aspect that enables the characterization of the overall utility of the data. It serves a general purpose without narrowing the focus to a particular application. Accordingly, the utility assessment revolves around four distinct metrics that complement each other. These four metrics pertain to the ability of anonymization to maintain a human-like facial appearance, preserve the facial structure of the subject, and ensure image quality and visual perception.

Within the scope of this work, as an integral part of the Bosch innovation project for autonomous driving, assessing the impact of anonymization on other perception algorithms relevant to self-driving cars, such as pedestrian detection, lane estimation, tracking, and motion estimation, is of utmost importance. However, this evaluation is infeasible due to the current low resolution of LiDAR sensors, which results in the unavailability of suitable data. This limitation prompted the use of datasets outside the scope of autonomous driving in the first place. Nevertheless, it is worth noting this consideration for future research.

Privacy-Utility Trade-off The privacy-utility trade-off represents the ultimate objective that algorithms aspire to achieve by striking a positive balance between privacy and utility. However, the reviewed literature does not present a unanimous methodology for evaluating this trade-off. For instance, Abbasi *et al.* [203] formally define a privacy-utility trade-off optimization criterion as:

$$T = [Acc_{Det} + Acc_{Agg} + PG]/3 \tag{4.2}$$

where Acc_{Det} is the face detection accuracy, Acc_{Agg} is the average between the matched and mismatched face verification accuracies, and PG is a privacy gain component, pivoting the maximum trade-off T score as a linear optimization problem. Inspired by this formulation, a practical solution is proposed to visualize the privacy-utility trade-off by attempting to condense the overall privacy and utility of each anonymization into a single value, despite its potential difficulty in quantification.

4.2.3 Evaluation Metrics

In Section 2.4.2, various standard metrics used in the literature for evaluating privacy and utility were presented. For this work, a subset of those metrics has been selected and is summarized in Table 4.1, along with a brief description of their aim. In addition to the quantitative metrics mentioned earlier, a qualitative assessment will be conducted to evaluate the anonymization results from both privacy and utility perspectives.

Privacy Metrics The verification mode entails analyzing ROC curves and their corresponding AUC values. Figure 4.8 visually represents the ROC curve computation framework for the selected dataset, which encompasses two primary stages. In the initial step, each anonymized image is subjected to comparison with every identity in the *gallery set*,

TABLE 4.1: Summarization of the evaluation metrics employed for assessing the face anonymization techniques.

Name	Description
Privacy Metrics	
CMC and Rank-1 Identification Rate	Face identification under the closed-set protocol
ROC curve and *AUC*	Face verification
Utility Metrics	
Delta Detection	The impact on the perception of faces by face detectors
Landmark Distance	Measures the impact on the facial structure
SSIM	Image quality metric related to the human visual perception
FID	Visual perception by distribution distances
Inference Time	Measures the feasibility of real-world applications

resulting in a distance value calculated through the cosine similarity metric[1]. This process yields a total of $40,000$ computed distances: 200 representing valid matches and the remaining $39,800$ denoting non-matching instances. Such an arrangement introduces a substantial class imbalance. Gu *et al.* [204] highlight that the ROC curve might exhibit limitations when assessing precision within these imbalanced contexts. Hence, embracing precision-recall curves could unveil disparities between algorithms that might remain concealed in the ROC space. In forthcoming investigations, this evaluation metric holds promise for providing deeper insights into the efficacy of the anonymizations. In the second phase, the distance threshold is systematically adjusted, and for each setting, the False Acceptance Rate (FAR) and the True Acceptance Rate (TAR) are calculated. In the second step, the ROC curve is generated by adjusting the classification threshold value for all the computed distances. In Figure 4.8, various thresholds are depicted as black lines overlaid on the distributions of the distances computed between genuine and impostor matches. Different thresholds result in different outcomes.

The closed-set recognition mode capitalizes on the insights the Cumulative Matching Characteristic (CMC) plot reveals. The computational framework of the CMC curve for the selected dataset is visually depicted in Figure 4.9, encompassing three distinct phases. In the initial phase, a distance-based similarity metric is calculated for an anonymized image in relation to every identity within the *gallery set*, generating 200 distance values. Subsequently, the second step entails arranging the identities of the subjects in descending order based on their similarity, effectively sorting them from the most closely aligned

[1] Contrary to the cosine metric, this can yield values exceeding 1.

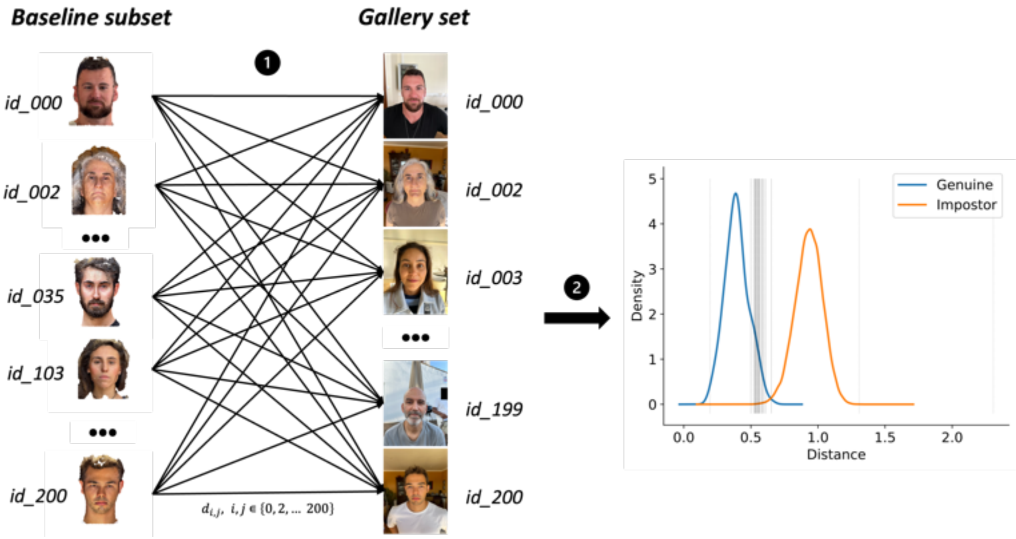


FIGURE 4.8: Face verification framework.

identity to the least similar counterpart according to the distance-based metric. Finally, the rank of the anonymized identity is established, yielding a value denoted as k, which spans the range of 1 to 200. This value signifies that the genuine match is positioned among the top k candidates. This procedure is iterated for all the anonymized images. In the example illustrated in Figure 4.9, the anonymized subject id_003 is ranked at position 2. Finally, the CMC corresponds to the percentage of subjects with a given rank valued between 1 and 200.

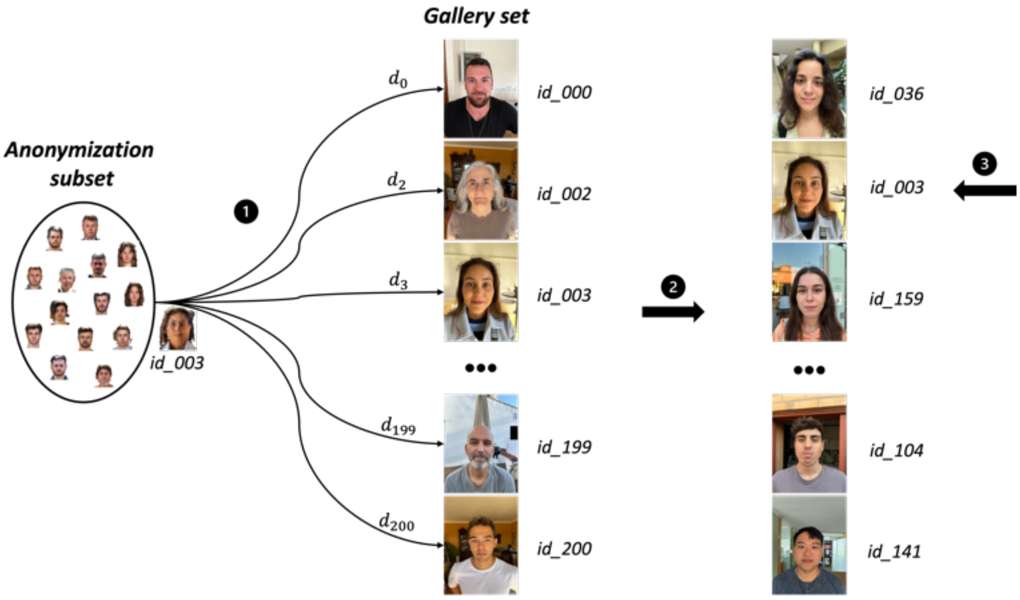


FIGURE 4.9: Face identification under closed-set framework.

Utility Metrics The *Delta Detection*, a term coined for this research as it lacks a clear name in the literature, quantifies the impact of the anonymization on the performance of a face detector. RetinaFace is chosen as the detector due to its use as the initial stage in the pipeline of the face recognition attacker model. Figure 4.10 illustrates the face detector results, encompassing both a bounding box and five landmarks, across various anonymization levels.

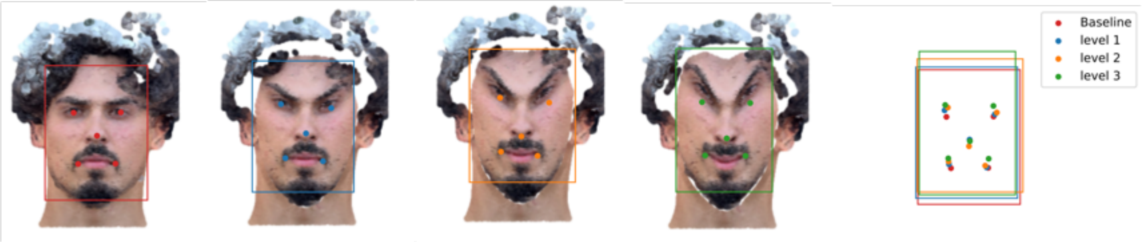


FIGURE 4.10: Data extracted by a face detector for *Delta Detection* and *Landmark Distance* computation.

The impact is measured by comparing the face detector's accuracy with and without the anonymization techniques for a model's prediction confidence of 0.9 and considering an Intersection over Union (IoU) of 0.5 as a correct detection.

Among the three presented image quality and visual perception assessment metrics, only two will be considered: the SSIM and the FID. The reason for this selection is that SSIM has been shown to outperform PSNR in measuring the quality of natural images across a wide variety of distortions [205]. However, this metric is computed on a subset region rather than considering the entire image. This subset region is delineated by the bounding box of the detected face on the baseline image, which matches the position in the anonymized image, as the facial location in the anonymized image remains unchanged. Refer to Figure 4.11, where each image corresponds to the bounding box region depicted in Figure 4.10 for each anonymization level used for computing the SSIM.

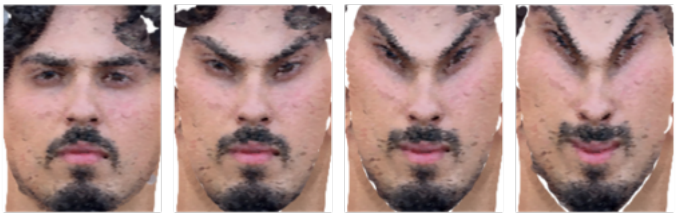


FIGURE 4.11: Subset region for the *SSIM* computation.

The reason for this approach is that a substantial amount of white space is present in the images, and by considering only the facial region, it is possible to observe the variations introduced by the anonymization more clearly. This way, the metric is not restricted to a small interval between 0.9 and 1, but it varies across a broader range of values, enhancing the interpretability of the anonymization effect. On the other hand, the FID does not suffer from this problem. Furthermore, FID provides additional insights into visual perception by measuring the difference between distributions of feature vectors for the two image sets corresponding to the anonymized and non-anonymized data.

The *Landmark Distance* is an additional metric not commonly encountered in the literature, offering further insights into the effects of the anonymization techniques on the subject's facial structure. This metric calculates the Euclidean distance between a set of keypoints denoted by five facial landmarks - the center of the right and left eye, nose tip, and right and left corner of the mouth - computed by the RetinaFace model on the anonymized result (Figure 4.10). If no detection is produced, the metric is penalized by considering a high distance value. The metric ranges between 0, indicating an unaltered facial structure, and the predefined penalty (a high value), indicating a face that is no longer identifiable as such due to excessive anonymization.

Although not directly a utility metric for evaluating anonymization results, the execution time is crucial for specific applications that demand real-time inference or have limited computational resources. The execution time can be considered a utility metric for the anonymization algorithm (not its results) and is presented as well.

Privacy-Utility Trade-off Metrics The privacy-utility trade-off is hard to fully capture by a single metric, as it would oversimplify its complexity and nuanced nature. Instead, three visual representations will illustrate the relationship, using a pair plot, a correlation matrix, and a mean aggregation of the privacy and utility metrics. Although the latter may suffer from oversimplification, it may still uphold valuable information. Further information is provided in the next Chapter.

The pair plot is a powerful visualization for exploring interactions among multiple anonymization and utility variable pairs, revealing connections between privacy and utility metrics. It uses scatter plots for metric relationships and diagonal plots for metric distributions. The correlation matrix is essential for quantifying variable relationships through numeric coefficients. It condenses privacy metric interactions, with coefficients near 1 indicating strong positive correlations and those near -1 indicating strong negative correlations. It helps identify prominent associations between metric pairs. On the other hand, the mean aggregation will summarize the overall strength of the metrics in terms of privacy and utility, enabling a straightforward comparison of the effectiveness of different anonymization techniques.

4.3 Implementation Details

The implementation details of the conducted experiments offer readers a comprehension of the experimental setup and procedures. The devoted time dedicated to exploring the tools, frameworks, and configurations utilized in the models of the three major research fields has been instrumental in ensuring reliable results for this study.

4.3.1 Software and Tools

This master's thesis was developed entirely using the high-level programming language Python. The decision to use Python as the primary programming language was driven by its extensive collection of libraries and frameworks ideally suited for machine learning, computer vision, and image processing tasks. Python 3.8.12 was chosen for this project without any specific reason for the version selection. The research primarily focused on face detection, face recognition, and face anonymization in 2D and 3D spaces, necessitating the adaptation of Python libraries and frameworks.

In addition to conventional Python libraries for data analysis, scientific and mathematical computation, machine learning, image processing, and data visualization, such as pandas, NumPy, scikit-learn, scikit-image, and Matplotlib, specific libraries crucial for this research and not typically utilized in standard projects are highlighted as follows.

Open3D [206] is an open-source library that supports the development of software dealing with 3D data. It provides diverse modules, including the *Geometry* module, which supports fundamental 3D processing algorithms and offers three geometrical representations: point clouds, triangle meshes, and images. The *I/O* module functionalities enable the reading and writing of 3D data files for each representation. Moreover, Open3D's *Visualization* module provides functionalities for 3D visualization, while the *Registration* module implements multiple methods for surface registration, including local and global registrations. Despite Open3D having other modules, the ones mentioned above were pivotal in achieving the research objectives.

DeepFace [207] is a lightweight Python framework for face recognition and facial attribute analysis. Developed using popular deep learning frameworks such as Keras and

TensorFlow, it ensures efficiency and accessibility for researchers. The library provides over five state-of-the-art deep learning models, including ArcFace for face recognition tasks, supporting verification and recognition.

The RetinaFace [207] library is a TensorFlow-based re-implementation of the state-of-the-art face detector RetinaFace, drawing inspiration from the InsightFace project. This library offers face detection and landmark localization functionalities, providing coordinates of the facial area (bounding box corners) and landmark coordinates (eyes, nose, and mouth extremities) along with a confidence score.

PyTorch [208] is a widely-used open-source deep learning framework developed by Facebook's AI Research Lab. Renowned for its effectiveness in building and training deep neural networks, PyTorch has gained popularity in the fields of machine learning and artificial intelligence. PyTorch provides extensive support for GPU acceleration, enabling faster training and inference on compatible hardware.

Below is an overview of the hardware setup employed for this research:

- *Machine 1:* MacBook Air 14" Apple M2 chip 8-Core 8GB;
- *Machine 2:* MacBook Pro 13" 2.3 GHz Dual-Core Intel Core i5 8GB;
- *Machine 3:* One Tesla V100-SXM2 32GB GPU with Dual-Core 8GB CPU.

The first two machines were utilized for most experiments. The need for two machines arose due to incompatibilities between TensorFlow and the M2 chip, which could not be resolved, limiting the use of the Deepface and RetinaFace libraries. As a result, all TensorFlow computations were conducted on Machine 2, while the remaining computations were performed on Machine 1. Adopting Machine 3 with GPUs was crucial for training the de-anonymization autoencoder model. GPUs play a vital role in deep learning algorithms, significantly accelerating model training and inference and enhancing efficiency and cost-effectiveness for deep learning tasks.

4.3.2 3D to 2D Projection

The term 3D projection, also known as graphical projection, refers to converting 3D objects or scenes into a 2D representation on a surface. In the context of 3D point cloud projection, the points of the object within the 3D space are mapped onto a 2D plane while preserving specific spatial relationships, some of which may create the illusion of depth and perspective. Numerous techniques are available for performing 3D projection, each possessing unique characteristics and applications. Refer to Figure 4.12 for an illustrative categorization of some 3D projections.

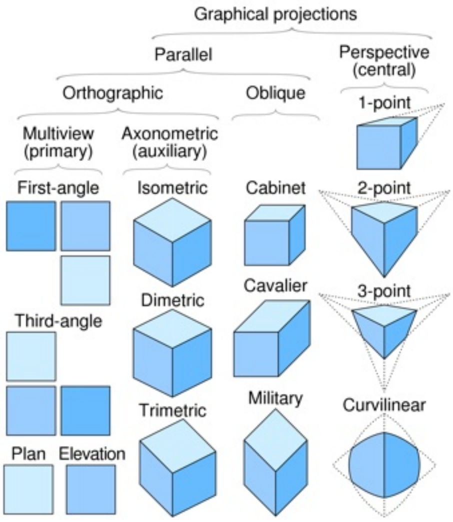


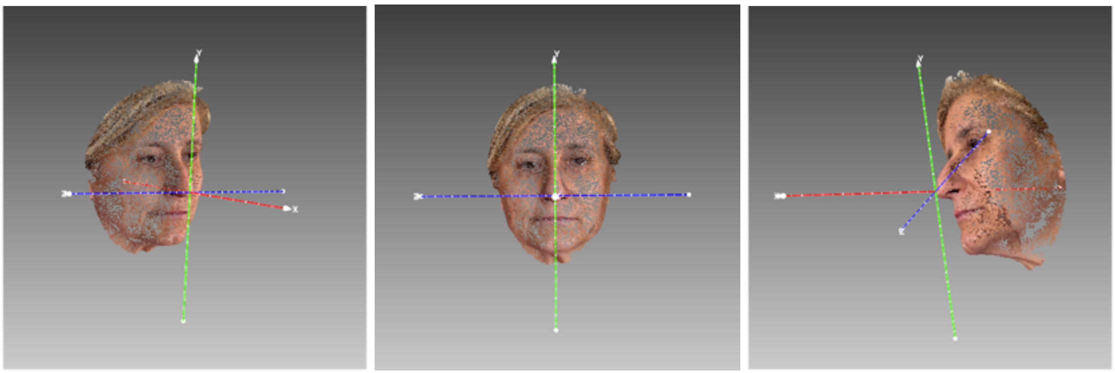
FIGURE 4.12: Classification of some 3D projections. Extracted from 3D Projection.

As face detection and face recognition algorithms will be applied to the 2D data, the obtained projection results should emulate the appearance of faces in standard digital images, which serve as the training input for these pre-trained models. Among the existing projections, the Multiview Orthographic projection primarily captures the facial structure and closely resembles digital images. This projection was also employed by Jovančević *et al.* [209] to detect and characterize defects on an airplane's exterior surface from 3D point could data. Despite not simulating perspective or depth perception, resulting in a flat and uniformly scaled representation unlike real-world human perception, the projection offers an accurate depiction of object sizes and shapes, and its implementation is straightforward.

Let $p = (p_x, p_y, p_z)$ denote an arbitrary 3D point in a point cloud. The resulting point p' from an orthographic projection onto the plane $z = 0$ of the point p is defined by:

$$p' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \\ 0 \end{bmatrix} \tag{4.3}$$

In this research, as explained in the previous Section 4.1, the design of the iPhone-PLYv3 dataset involved registering all point clouds to a precise location. This registration ensures that the nose tip is at the origin, the front view of the face is oriented towards the x-axis, and the yaw, pitch, and roll angles are standardized. For a visual representation of the point clouds' orientation compared to the 3D axis, refer to Figure 4.13. The images are taken from the MeshLab processing tool. The x, y, and z axis are represented by the colors red, green, and blue, respectively.



(A) Predefined viewpoint.

(B) x-axis is perpendicular to the screen.

(C) Predefined viewpoint.

FIGURE 4.13: Visual comparison of subject id_004's face with the reference axis.

Given these settings and the insight obtained from Figure 4.13, the orthographic projection onto a plane $x = k$, where k is a constant, captures the front view of the subject's face. In the implementation, a scatter plot was generated in Python using the 3D coordinates of the point clouds. The X coordinates were excluded, and the Y values (x-axis of the scatter) were plotted against the Z values (y-axis of the scatter) of the point clouds. The color component of each point remained unchanged. However, this approach involves two parameters that require special attention: the size of the scatter points and the ordering of the points based on their coordinates[1].

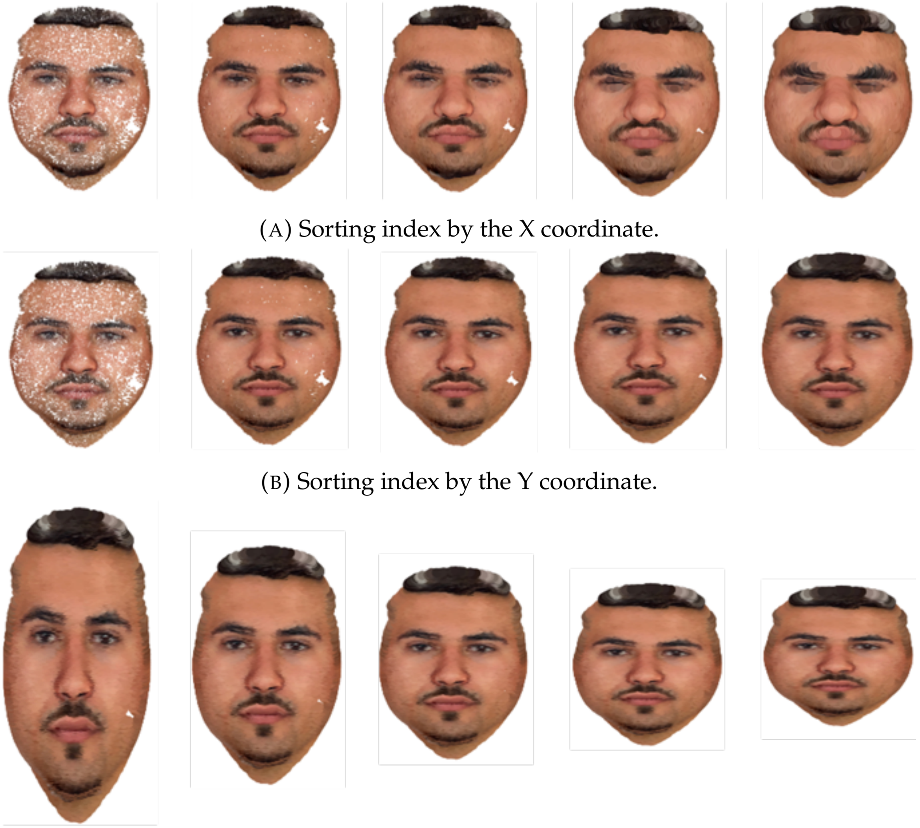
The first parameter (Figure 4.14a), scatter point size, is crucial in controlling the appearance of white patches induced by low-density regions on the 3D face models. Setting a low size value prevents overlapping points, causing the white patches to be more prominent in the projection. On the other hand, using a size value that is too high introduces some priors, particularly in the surrounding region of the face, or deforms the facial features according to the scatter points order.

For the second parameter (4.14b), scatter points order, the intuitive choice would be to display the points in an increasing order of the X-coordinate to avoid superimposition of further away points on closer ones (recap Figure 4.13 for the axis orientation). However, the empirical analysis highlighted issues with this approach. The eye region, for

[1] The organization of the scatter points based on the X, Y, or Z coordinates in ascending or descending order.

instance, exhibits concavity instead of convexity, as already mentioned, which leads to severe deformations when points in the eye's vicinity superimpose the points of the eyes. Similar deformations are also observed in other regions, such as the nose and mouth. As an alternative, a tested solution was to select the index order based on the Y-coordinate in increasing order, which yielded satisfactory results and was chosen for the final implementation.

Another constraint involved the axis limits of the scatter plot (Figure 4.14c), which were initially set on the same scale for both X and Y coordinates. However, this led to an undesirable effect of overly wide faces. Consequently, the X scatter coordinate limits were reduced to flatten the faces, resulting in a more accurate representation.



(A) Sorting index by the X coordinate.

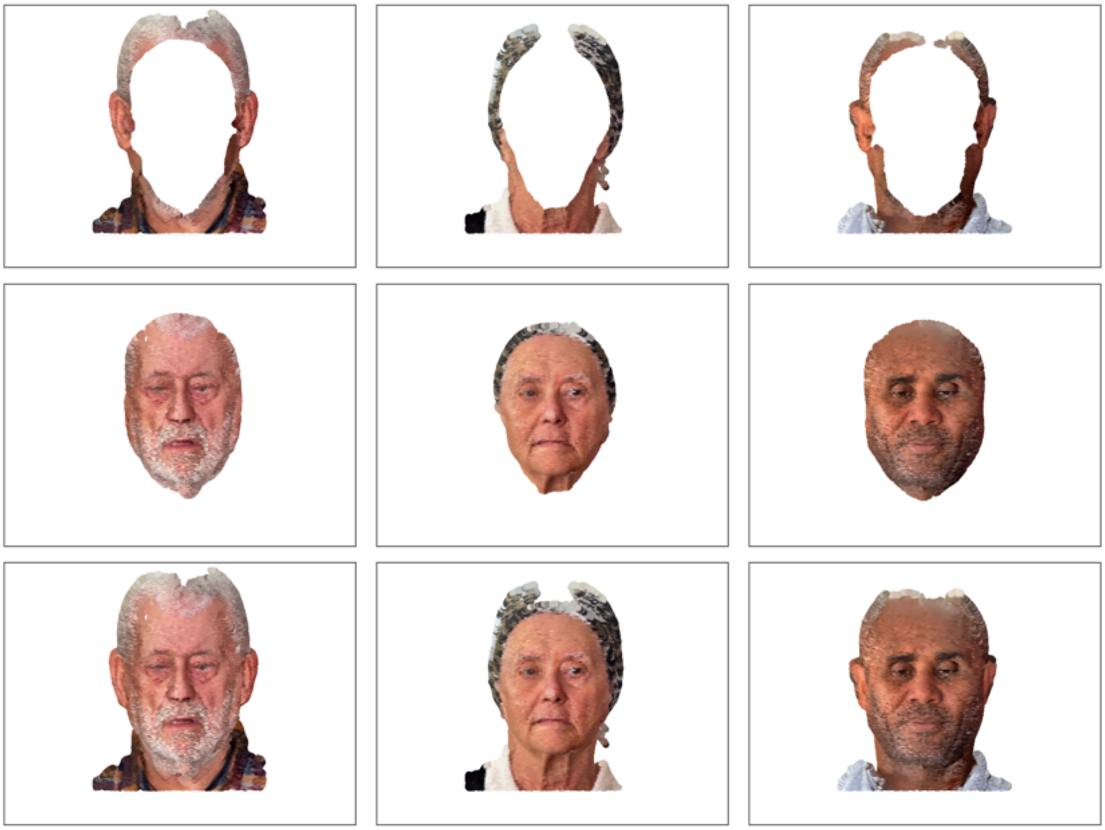
(B) Sorting index by the Y coordinate.

(C) Y axis scale variation.

FIGURE 4.14: Impact of parameter adjustments on 3D to 2D projection results.

After conducting experiments, it was evident that choosing a point size that avoids both deformation and white holes while sorting the index by the X coordinate is impossible. Consequently, the points were sorted according to the Y-coordinate in increasing order with a size of 25, preventing major deformations and white holes on the faces. As for the scatter axis limits, the x-axis ranges in the interval $[-0.25, 0.25]$, whereas the y-axis ranges between $[-0.18, 0.18]$.

The resulting images are 389 × 515 pixels in size, each depicting a single identity with their face centered on a white background. Regarding the anonymization techniques, they are exclusively applied to the inner part of the face, as detailed in Chapter 3. After applying them, the anonymized 3D information is merged with the 3D non-anonymized outer region of the head, which was excluded through the sphere cropping in the preprocessing pipeline. This merging provides additional contextual information, and the resulting combined points are then projected using the 3D to 2D projection procedure explained earlier. Figure 4.15 illustrates the further contextual information that is provided by this procedure. The top row represents the outer face projection, the middle row depicts the inner face projection, and the bottom row shows the stitching result. A black border has been added to enhance the perception of the true dimensions of the images.



(A) id_005. (B) id_006. (C) id_007.

FIGURE 4.15: Final 3D to 2D projections of point clouds for three subjects.

4.3.3 Anonymization Techniques

The proposed anonymization techniques include a set of regulator parameters that control the degree of magnitude of the data modification, providing varying levels of control over the point cloud anonymization process. Some techniques may have single or multiple regulator parameters. However, for those with multiple, the regularization was employed to one parameter keeping the remaining fixed at a specific value. As a result, the evaluation procedure is simplified, avoiding the need to test an extensive number of configurations, which would be impractical. Besides, this approach enables a more explicit assessment of the effect of each anonymization technique parameter and its impact on the evaluative metrics. However, future research can focus on exploring the full potential of the proposed techniques by considering a more comprehensive range of tests, delving into the effects of different parameter combinations.

For each anonymization technique, the process of choosing the regularization parameter led to the creation of 35 distinct configurations, known as anonymization degrees. These degrees span from 1 to 35, with 1 indicating the least impact and 35 producing the most significant effect. The regularization parameters were empirically tested, consisting of equally spaced values ranging from no modification to a level of modification that entirely hinders face recognition, measured by an AUC[1] value from approximately 1 to 0.5. This approach enables the illustration of the evolution of privacy and utility metrics and the trade-off for various standardized degrees of anonymization. It effectively depicts the strengths and weaknesses of the anonymization techniques, allowing for their comparison on a common basis. Table 4.2 presents the parameters used for all tested anonymization algorithms.

For each regulator parameter, specific adjustments had to be made to the size and index order parameters of the 3D to 2D projection procedure to cope with the effects of the modifications. Thus, thorough testing was conducted to achieve visually satisfactory projections results, minimizing biased errors that could hinder the recognition performance.

4.3.3.1 Sampling-Based

CentroidVoxel The *CentroidVoxel* anonymization technique is a straightforward approach centered around a single regulating parameter, dictating the resolution of the voxel

[1] The AUC is an evaluation metric for face identification that also captures the attributes of the anonymization technique in the verification scenario.

TABLE 4.2: Regulatory parameters for all anonymization algorithms in the conducted experiments, originating the anonymization degrees.

Anonymization	**Parameter**	**Values**
CentroidVoxel	*size*	$size = \{x_i \mid x_i = 0.00025 \cdot i,\ i = 1,2,\ldots,35\}$
Merge2Faces	*weight*	$weight = \{x_i \mid x_i = 0.0285 \cdot i,\ i = 1,2,\ldots,35\}$
PMP	*alpha*	$alpha = \{x_i \mid x_i = 0.00215 \cdot i,\ i = 1,2,\ldots,35\}$
SmoothKNN	k-NN	$k = \{x_i \mid x_i = 8 \cdot i,\ i = 1,2,\ldots,35\}$
Tapering	$[r_m, r_M]$	$\lvert r_m \rvert = \lvert r_M \rvert = \{x_i \mid x_i = 0.087 \cdot i,\ i = 1,2,\ldots,35\}$
UniformNoise	$\sim \mathcal{U}(a,b)$	$a = 0, b = \{x_i \mid x_i = 0.00064 \cdot i,\ i = 1,2,\ldots,35\}$

grid. For the voxel grid creation, the Open3D's *Geometry* module was leveraged, which features a method for generating a voxel grid directly from a point cloud. This process designates a voxel as occupied if it encompasses at least one point from the point cloud, while the color of a voxel is computed as the average color of all the points within it. The voxel grid resolution, highlighted by the voxel size, is depicted in Figure 4.16 for the subject *id_001*. With an increase in voxel size, more points are condensed into the voxels, resulting in higher information loss. Consequently, the number of voxels decreases until it ultimately converges into a single voxel that aggregates the information of the entire point cloud.

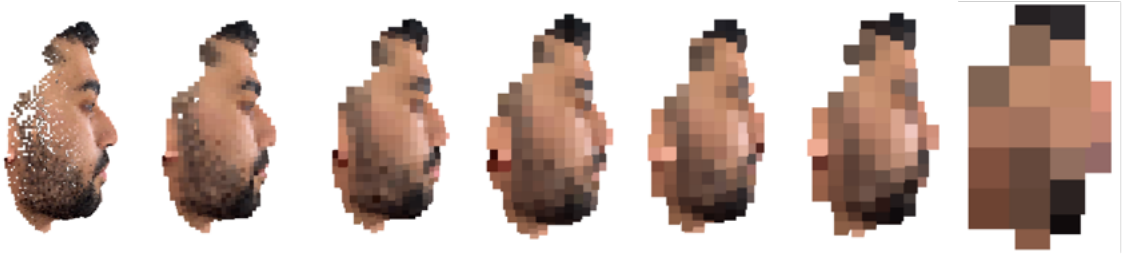


FIGURE 4.16: Voxel grid generation with voxel sizes ranging from 0.0025 to 0.150 with 0.0025 increments, spanning from 7812 voxels down to 49 voxels.

As depicted, the increments in voxel size lead to voxels occupying larger portions of space. Consequently, the points originating from each voxel – where the coordinates and color correspond to the centroid coordinate and average color of all points within the voxel – become more distant from each other.

As the voxel resolution decreases, the positions of the points shift, ultimately moving beyond the confines of the initial input facial region. Furthermore, these points accumulate more information from other points in the point cloud, all aggregated within the same voxel. This aggregation results in information loss and confers an inherent sampling nature to this technique. The result of the anonymized point clouds is a notable decrease in point density, which becomes especially pronounced in scenarios involving large voxel sizes, leading to significant reductions through extreme sampling.

Consequently, these characteristics introduce uniform patterns into the projected anonymization if no action is taken to adjust the point size on the scatter plot. These white patches were addressed by methodically selecting the appropriate point size for each anonymization parameter, ranging from 25 in moderate cases to 55 in the most extreme scenarios. For a visual representation of the impact of point size on resulting anonymized images across various modification scenarios, refer to Figure 4.17.

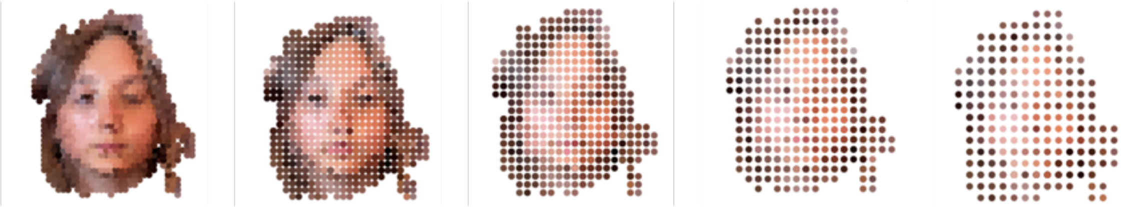


FIGURE 4.17: *Centroid Voxel* technique with uniform point size equal to 25 of the inner face of subject id_197. From left to right, the parameter *size* starts with 0.006 with an increment of 0.002 until 0.014, containing between 1 850 and 395 points.

4.3.3.2 Noise-Based

UniformNoise The *UniformNoise* technique involves two regulator parameters related to the noise distribution, which follows a uniform distribution. The lower limit of the uniform distribution is fixed at the origin by setting $a = 0$, while the upper limit b is set to vary. Consequently, only positive perturbations are introduced to the coordinates of the points, shifting their location toward the positive direction of the X, Y, and Z axes. However, this shift is compensated by subtracting a constant corresponding to the mean of the uniform distribution from the coordinates, as already explained. Figure 4.18 depicts a visual representation of a uniform distribution with a sample size of 10 352, the same as the number of facial points of each point cloud, generated in Python.

If a noise value has coordinates $x = 0.020, y = 0, z = 0$, the facial point shifts its position 2 centimeters in the positive direction of the x-axis (the units are in meters).

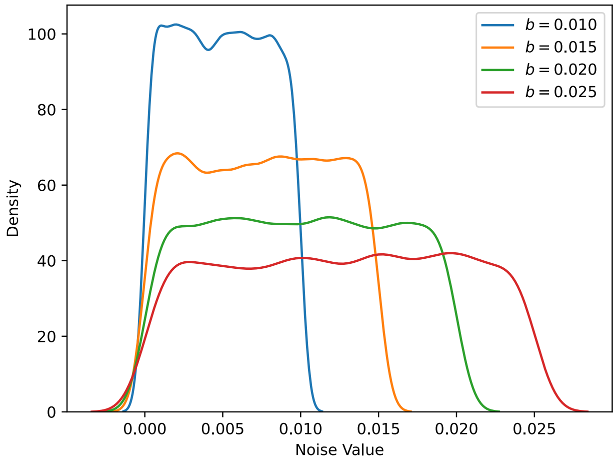
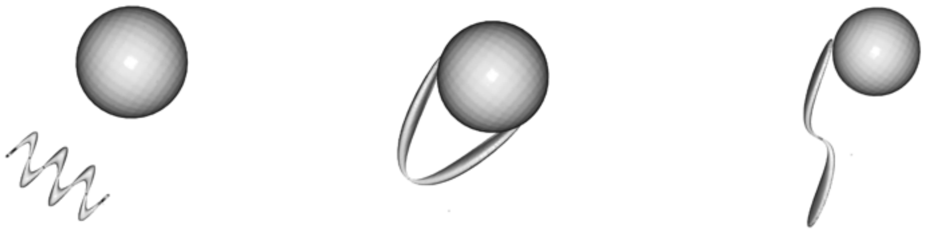


FIGURE 4.18: A uniform noise distribution $\mathcal{U}(a, b)$, with $a = 0$.

4.3.3.3 Warping-Based

Tapering The *Tapering* technique is governed by two regulator parameters, one of which is the tapering function that guides the deformation result, shaping the final appearance. To observe the impact of the tapering function on the anonymization result, refer to Figure 4.19. In this toy example, three distinct tapering functions are applied to a sphere centered at the origin with a radius of 0.1, containing 50 000 points.



(A) $F(x) = sin(x)/6$.

(B) $F(x) = x^2/100$.

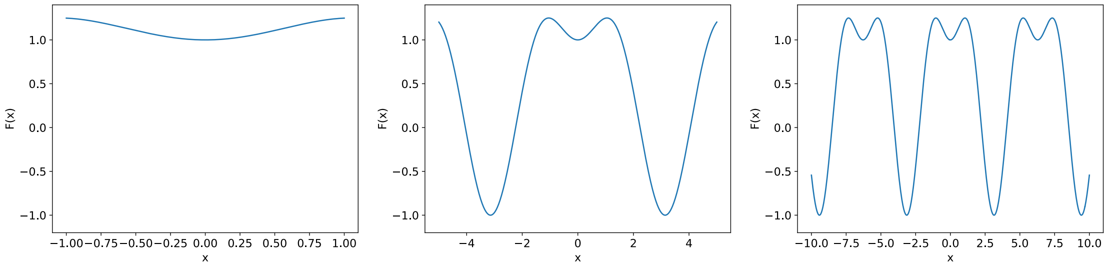
(C) $F(x) = x^3/1000$.

FIGURE 4.19: Transformation of a 3D sphere using three distinct tapering functions varying across the interval $[-10, 10]$.

The anonymization shape closely resembles the tapering functions' shapes. Therefore, the tapering function is designed to be computationally efficient, avoiding complex computations while achieving a desired unique shape. The tapering function was selected arbitrarily from a set of functions with desirable characteristics, which were determined through empirical observation. The definition of the selected tapering function is as follows:

$$F_{anon}(x) = sin(x)^2 + cos(x) \tag{4.4}$$

The other parameter to consider is the restricted domain of the tapering function. Even with the same tapering function, this parameter can significantly impact the final shape of the anonymization. Figure 4.20 visually represents three different restricted domains of the function F_{anon}. This graph illustrates that the same function can produce highly distinct results, considering that the shape shown in the graph will resemble the final shape of the anonymization.



(A) $r = 1$.

(B) $r = 5$.

(C) $r = 10$.

FIGURE 4.20: Different restricted domains of the function defined in Equation 4.4.

The impact of the *Tapering* technique is assessed by altering the function F_{anon} restricted domain, defined in the form $[-r, r]$ for simplicity, where different real values of r are taken into account.

4.3.3.4 Morphing-Based

Merge2Faces The *Merge2Faces* anonymization technique stands out as one of the most complex methods proposed. The anonymization consists of incorporating the facial traits of a target face, one of the regulator parameters, into a source with varying degrees of influence. In this study, an invariant target face from subject *id*_001 was employed to anonymize all identities.

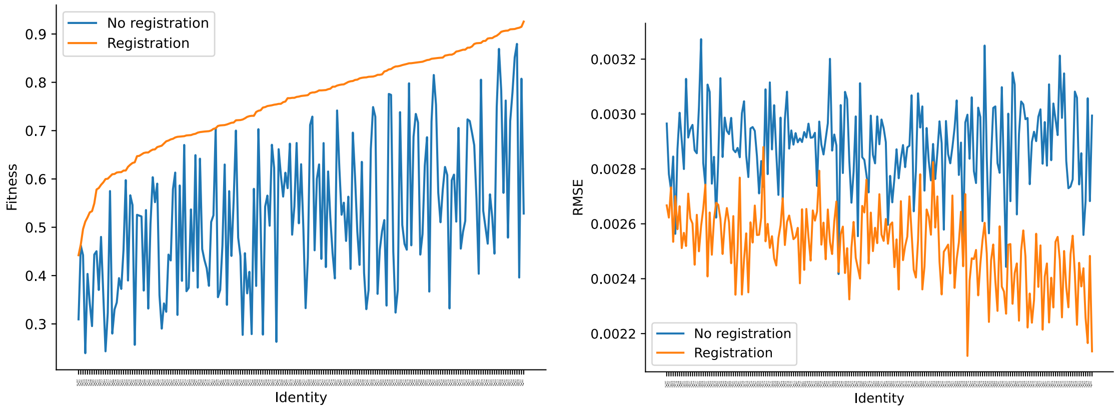
This algorithm unfolds in two distinct stages. The initial stage concerns the alignment of a source face and the face of the subject *id*_001. However, due to the preliminary manual alignment performed on the point clouds during the dataset's acquisition and preprocessing phases, the need for an initial application of the global registration RANSAC algorithm was obviated. The initial alignment of all the faces served as an adequate starting point for the subsequent local registration. For this purpose, the Point-to-Plane ICP algorithm variant was executed using Open3D's *Registration* module. The convergence criterion corresponding to the Root Mean Square Error (RMSE) difference of the two-point clouds between algorithm iterations was left at its default value. The second stage was executed with direct simplicity leveraging the Open3D's *Geometry* module for a fast nearest-neighbor search across the two point clouds, following the completion of the initial alignment process.

The quality of the outcomes produced by this implementation is closely interconnected with the effectiveness of the initial registration phase in achieving a robust alignment between the *source* and *target* point clouds. The assessment of alignment quality between these two point clouds can be estimated through the utilization of two defined metrics:

- *Fitness*: quantifies the overlapping area by evaluating the ratio of inliers[1] to the total number of points in the target point cloud. A higher value indicates better alignment;
- *Inlier RMSE*: computes the RMSE of all inlier correspondences. A lower value signifies more accurate alignment.

These metrics were calculated using Open3D's *Pipelines* module, with a distance threshold of 0.005, equivalent to 5 millimeters, for identifying inlier points. The calculations were performed before and after applying the ICP algorithm for local registration, contrasting the initial rough alignment with the refined alignment utilized in the subsequent algorithmic steps. Figure 4.21 visually presents the outcomes for both metrics. The x-axis displaying subject identities has been arranged in increasing order of the Fitness metric after ICP registration for both plots, promoting better visualization and standardization. The results indicate a significant enhancement in the Fitness metric across all subjects, with improvements spanning from 9.7×10^{-5} to 0.525. Figure 4.22a illustrates the alignment that achieved the most substantial Fitness improvement, reaching 0.525, for subject *id_129*. A Fitness value of over 0.5 suggests that the number of inliers more than doubled compared to the initial alignment, with over half of the points in the *target* point cloud located within 5 millimeters of a point in the *source* point cloud. However, the RMSE of seven out of 199 point clouds increased. The most significant rise in RMSE before and after alignment was equivalent to 10^{-4}, with absolute improvements diminishing to 8.9×10^{-4}.

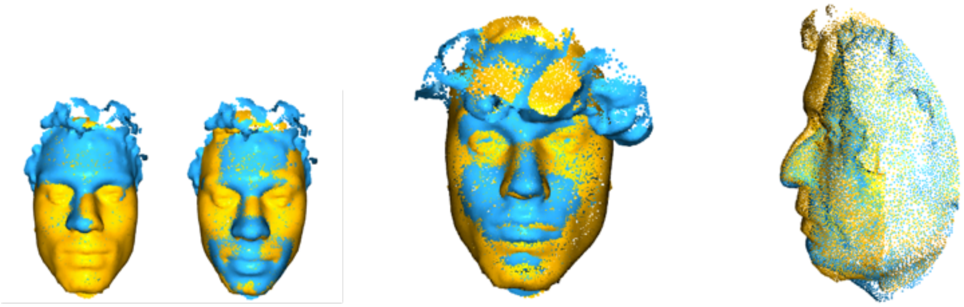
[1] Inliers refer to source points whose distance to a target point falls below a correspondence distance threshold.



(A) Fitness.

(B) Inlier RMSE.

FIGURE 4.21: Evaluation metrics for point cloud registration.



(A) Best fitness improvement.

(B) Underestimation fitness illusion.

(C) Overestimation fitness illusion.

FIGURE 4.22: Visual results of local registration.

Nonetheless, the fitness measure may underestimate results in cases where the facial point cloud segmentation is suboptimal and includes a portion of the hair. For instance, Figure 4.22b showcases the ICP alignment outcome for subject *id_197*. While visually satisfactory, the Fitness metric scores a low value of 0.44, marking the poorest result among all subjects. This discrepancy is attributed to the hair of the subject negatively affecting the metric, even though the facial alignment is well-executed. Conversely, Figure 4.22c[1] illustrates the ICP alignment output for subject *id_002*. Despite a Fitness score of 0.78, the visual results are unsatisfactory as the eye regions remain significantly apart in both faces. This instance underscores that specific Fitness values might overestimate alignment success. Hence, it's advisable to exercise caution when considering both metrics.

In addition, Figure 4.22c highlights that the registration process can prove highly challenging for certain subjects, owing to substantial variability in facial structure and characteristics between the two individuals. In these samples with suboptimal alignment,

[1] The variance in points dispersion across the three images is due to differing zoom factors.

the 2D projection is denoted by the appearance of white patches. The reason for this occurrence is that when the neighbors taken into account for the weighted average are not well-aligned, the corresponding mean point will be pushed farther away.

4.3.3.5 Smoothing-Based

SmoothKNN The *SmoothKNN* technique is governed by a single parameter that controls the extent of facial smoothing and color blurring in the anonymized point cloud. The initial step of this approach involves computing the neighboring points for each point within the point cloud. Similar to the *Merge2Faces* method, neighboring points were determined using Open3D's *Geometry* module, enabling rapid nearest-neighbor searches. Refer to Figure 4.23, which illustrates the facial region of interest used to calculate the average coordinates and colors. This region is defined by the (k-1) nearest neighbors in a light tonality relative to a predefined point marked in red.

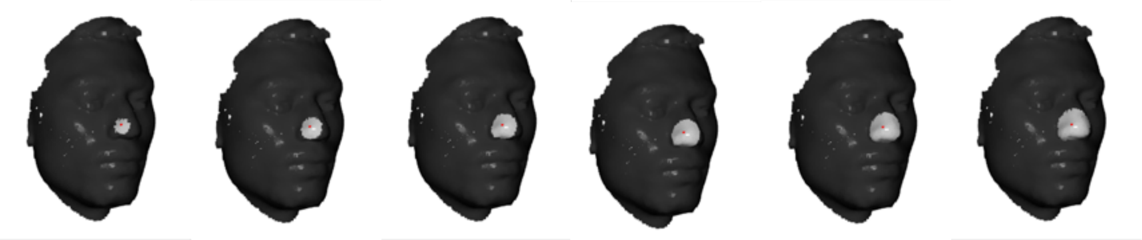
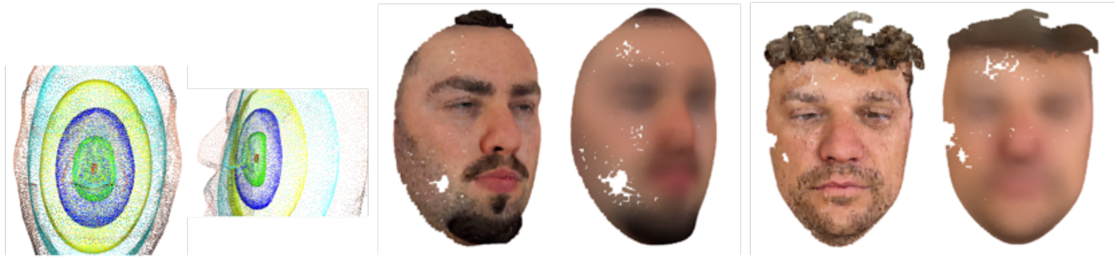


FIGURE 4.23: Number of neighbors to consider when computing the average color and coordinates for the *SmoothKNN* for subject id_001. From left to right, the parameter *k*-NN starts with 50 with an increment of 50 until 300.

As the number of nearest neighbors considered for computing averages progressively rises, approaching the total count of facial points at its maximum, the point cloud gradually condenses, eventually converging into a singular point. This phenomenon is illustrated in Figure 4.24a, where the generated point clouds result from diverse anonymization parameters, spanning from 2 000 to 10 000 with increments of 2 000. These point clouds are presented in different colors to improve visual clarity and offer two viewing perspectives. For reference, the original point cloud is also showcased.

Furthermore, Figure 4.24 highlights two issues associated with the number of neighbors, considering 200 neighbors. In instances where the segmentation of the facial point clouds is suboptimal, encompassing a portion of the hair, the hair's influence leads to the displacement of points toward it, resulting in certain regions being left blank on the forehead 4.24c (*id*_090). Additionally, Figure 4.24b (*id*_068) illustrates a potential concern arising from sparsity in specific areas, which can become more pronounced for a relatively low number of neighbors as surrounding points exert influence on these regions, propelling the points in their direction. As the number of neighbors increases, this issue gradually diminishes. This leads to the surface of white patches in the projections of the anonymized data, which is addressed by augmenting the size of the points in the scatter plot to 60.



(A) Shrinkage.

(B) Sparsity.

(C) Poor segmenting.

FIGURE 4.24: Challenges associated with *SmoothKNN*.

4.3.3.6 Point Operations-Based

PMP The *PMP* anonymization technique is a straightforward method governed by two regulatory parameters. These parameters oversee the detail of the α-shape obtained from the initial point cloud transformation and the count of points following the α-shape transformation back into a point cloud. For this specific technique, the count was fixed at 10 352 points, mirroring the number of points in the original point clouds.

As the detail of the α-shape diminishes due to the anonymization parameter, the shape progressively approximates the convex hull of the face. This evolution is depicted in Figure 4.25, which illustrates the gradual transformation of the α-shape into its convex hull. Owing to their unique characteristics, prominent facial attributes like the nose, eyes, and mouth undergo substantial alterations while the overall facial structure is still preserved. This trait proves advantageous for effective anonymization. As the α-shape approaches the convex hull, the concave regions surrounding the eyes are progressively concealed (akin to an eye patch), and the nose region is transformed into a pyramid-like structure. This transformation unfolds as facial regions gradually connect to the elevated sections of the nose, ultimately culminating at its tip. A comparable pattern of alteration is evident in the mouth region.

Subsequently, the α-shape is transformed into a point cloud, accompanied by color alterations that are influenced by the surrounding regions. Although not explicitly depicted,

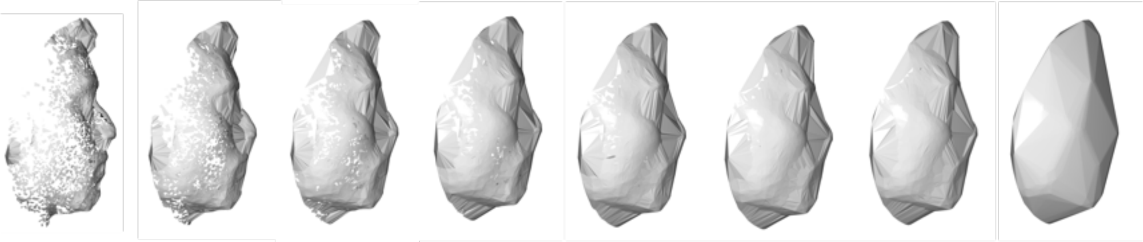


FIGURE 4.25: *PMP* initial point cloud to mesh conversion for subject id_001. From left to right, the parameter $alpha$ starts with 0.02 with an increment of 0.01 until 0.07. The furthest image to the right is the convex hull.

the rear section of the face undergoes the transformation, becoming fully interconnected, resulting in an unusual effect.

During the 3D to 2D projection, it was discovered that for an $alpha$ value greater than 0.030, the projection outcome appears more consistent with the point cloud's anonymization result when the scatter plot indexing is based on the x-axis order as opposed to the y-axis order. This indexing adjustment deviates from other anonymization techniques, which tend to use the y-axis order. Furthermore, this alteration necessitated resizing the point clouds to mitigate the swelling effect, as elucidated in Section 4.3.2.

4.3.4 Face Detection and Recognition Models

On the Deepface framework, the face recognition model considered was ArcFace, which comes with a pre-trained model. This pre-trained model is stored in a file containing the learned parameters (weights and biases) after extensive training on a large dataset.

The face detector backend was set to the RetinaFace model, and its default settings remained unchanged. This backend detector is responsible for locating facial regions within the image, leaving Deepface to exclusively manage the alignment and normalization steps for these identified regions. As outlined in Section 2.3, this process aligns with the conventional face recognition pipeline, where subsequent stages are centered on representing the pre-processed facial regions as feature vectors and proceeding with the feature matching. However, in cases where the backend detector fails to identify any facial region within the image, the library treats the entire input image as a face and proceeds to compute its embedding. This behavior is enforced by setting the *enforce_detection* command to *True*.

The similarity metric used in the feature matching stage is the cosine similarity. It calculates the cosine similarity between the vector representations from the feature representation stage of the two identity images given as input. The similarity score and a threshold value θ are then used to determine whether the two identities are regarded as matching or non-matching.

As part of this research, a slight code adjustment was made to a Deepface function, allowing users to consider a custom threshold for the ArcFace model instead of being restricted to the default value $\theta = 0.68$. This modification aimed to enhance downstream tasks by improving the speed and practicality of computing evaluation metrics.

On the *retinaface* library, all the default configurations were used. The five landmarks outputted by the model and the facial region bounding box are used as integral parts of the utility metrics.

4.3.5 De-Anonymization Model

The de-anonymization model is specifically designed to reverse the 3D to 2D projections, following the *reversibility recognition* attack configuration. In order to achieve optimal results, certain adjustments were made to the dataset to accommodate the training and testing of the model. The model's design was determined through empirical testing, and the best configuration was chosen.

4.3.5.1 Dataset

Regarding the model implementation, the dataset was divided into three sets, each containing disjoint identities: the training, validation, and testing sets. The division follows standard practices, with the data splits consisting of 80%, 10%, and 10% of the total 200 identities, respectively. Consequently, the training set comprises 160 identities not present in the other two sets, while the validation set includes 20 identities out of the 40, leaving the remaining 20 for the testing set. The three sets were randomly chosen. This data partitioning strategy ensures that the model is trained, validated, and tested on different identity samples, enhancing the reliability and generalizability of the results. Nonetheless, the results may exhibit bias as a result of the single random instance employed in this strategy.

To improve the model training, the training set contains the subject images corresponding to different view angles from the 160 subjects, ranging from -10 degrees until 10 degrees with increments of 2 degrees for the six anonymization techniques. In total, it comprises 52 800. The model may improve learning by getting exposure to more data by considering different viewing angles instead of only the frontal image of the subjects. Refer to Figure 4.26 for a visual representation of different viewpoints, considering various degree angles.

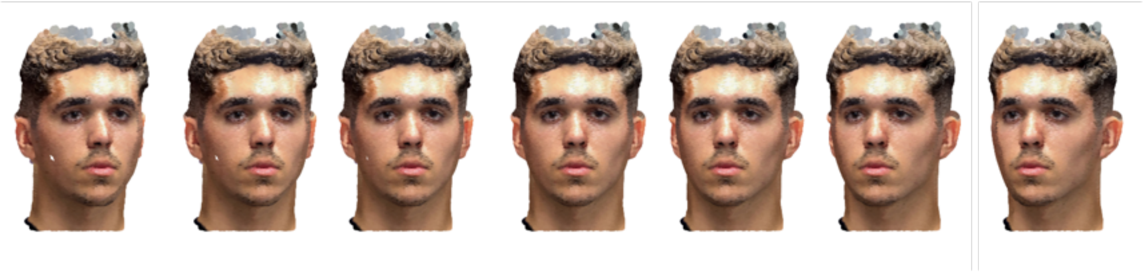


FIGURE 4.26: Autoencoder training sample images of *id*_016, ranging from -10 degrees to 10.

The testing set contains 20 identities, each originating a de-anonymized frontal image that further contributes to the privacy evaluation.

4.3.5.2 Architecture

The architecture of the Autoencoder comprises an encoder and a decoder such that:

1. Encoder
   - Receives as input an RGB image with three channels;
   - The encoder part consists of three convolutional layers with ReLU activation functions and max-pooling layers;
   - Outputs a compressed representation with 8 channels.
2. Decoder
   - Receives as input the compressed representation from the encoder;
   - The decoder part consists of three transposed convolutional layers, also known as deconvolution or upsampling layers, with ReLU activation functions;
   - The final layer uses the sigmoid activation function to scale the output values between 0 and 1, which is appropriate for reconstructing and outputting the reconstructed RGB images.

The decoder aims to reconstruct the original input from the compressed representation possessing a symmetrical design compared to the encoder.

4.3.5.3 Training Procedure

The training loop implements the Autoencoder using MSE as the loss function to measure the difference between reconstructed and clean images, and the Adam optimizer, with a learning rate of 1×10^{-3}, updates the model's parameters during training. Due to memory limitations, a batch size of 8 is used. The model is trained for 250 epochs, and early stopping is applied if the validation loss does not improve for 20 consecutive epochs. The training loop records the training and validation losses, computed by averaging the losses over their respective datasets. The model is saved whenever an improvement in the validation loss occurs, preventing overfitting and identifying the best model configuration. As a result, the training reconstructs anonymous images by minimizing the MSE loss between the reconstructed and clean images.

The training losses of the model are illustrated in Figure 4.27. While the model demonstrates the ability to learn how to de-anonymize the anonymizations, the limited reduction in validation loss in comparison to the training loss suggests that the model may struggle to generalize effectively beyond the training dataset.

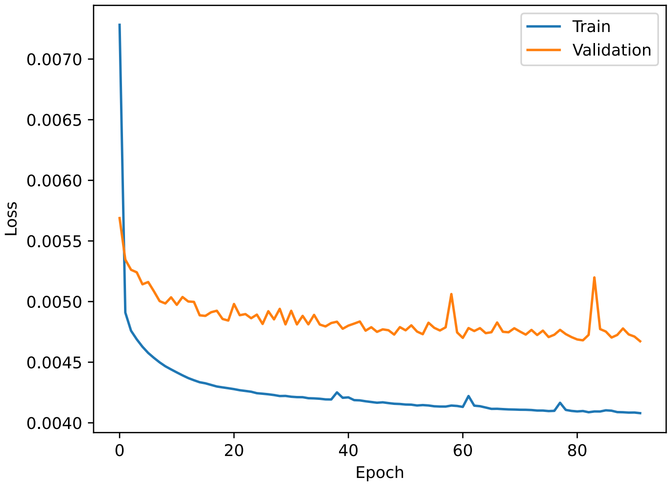


FIGURE 4.27: Train and validation losses of the autoencoder.

4.4 Limitations

One limitation of this study is the size of the relatively small dataset. Although the dataset included 200 identities, larger than many existing datasets in the literature, it may still be insufficient for robust generalization. Strengthening the validity and generalizability of the findings requires considering a larger dataset with a substantially greater number of identities in future research. Within the academic community of Faculdade de Ciências da Universidade do Porto (FCUP) and even at the University of Porto, dedicated efforts could be undertaken to collect and integrate a more extensive dataset complementing the current one. This endeavor would enable a more comprehensive evaluation of the face anonymization approach, producing more meaningful insights and results.

Another limitation was the evaluation procedure for privacy and utility, which was a time-consuming process that required over two days of code running time to compute all the metrics for all the anonymizations, only considering 200 identities. This duration becomes prohibitive when aiming for a higher number of identities, such as in large-scale datasets. Therefore, a more efficient and optimized evaluative framework design becomes essential. Nevertheless, a total of 8 400 000 identity comparisons between two images were performed within this work, and over 40 000 images were subjected to face detection. These numbers are not insignificant and reflect the extensive analysis carried out in this work. In addition, many parameters required fine-tuning and adjustments during the experiments, which was time-consuming and may introduce challenges in achieving optimal results. In particular, regarding the 3D to 2D projections, there is room for improvement in streamlining the approach.

Finally, the de-anonymization autoencoder model could be further improved by considering a more sophisticated loss function, increasing the training data, or enhancing the overall architecture. Along the same lines, some anonymization techniques can be further optimized concerning efficiency and overall design.

Chapter 5

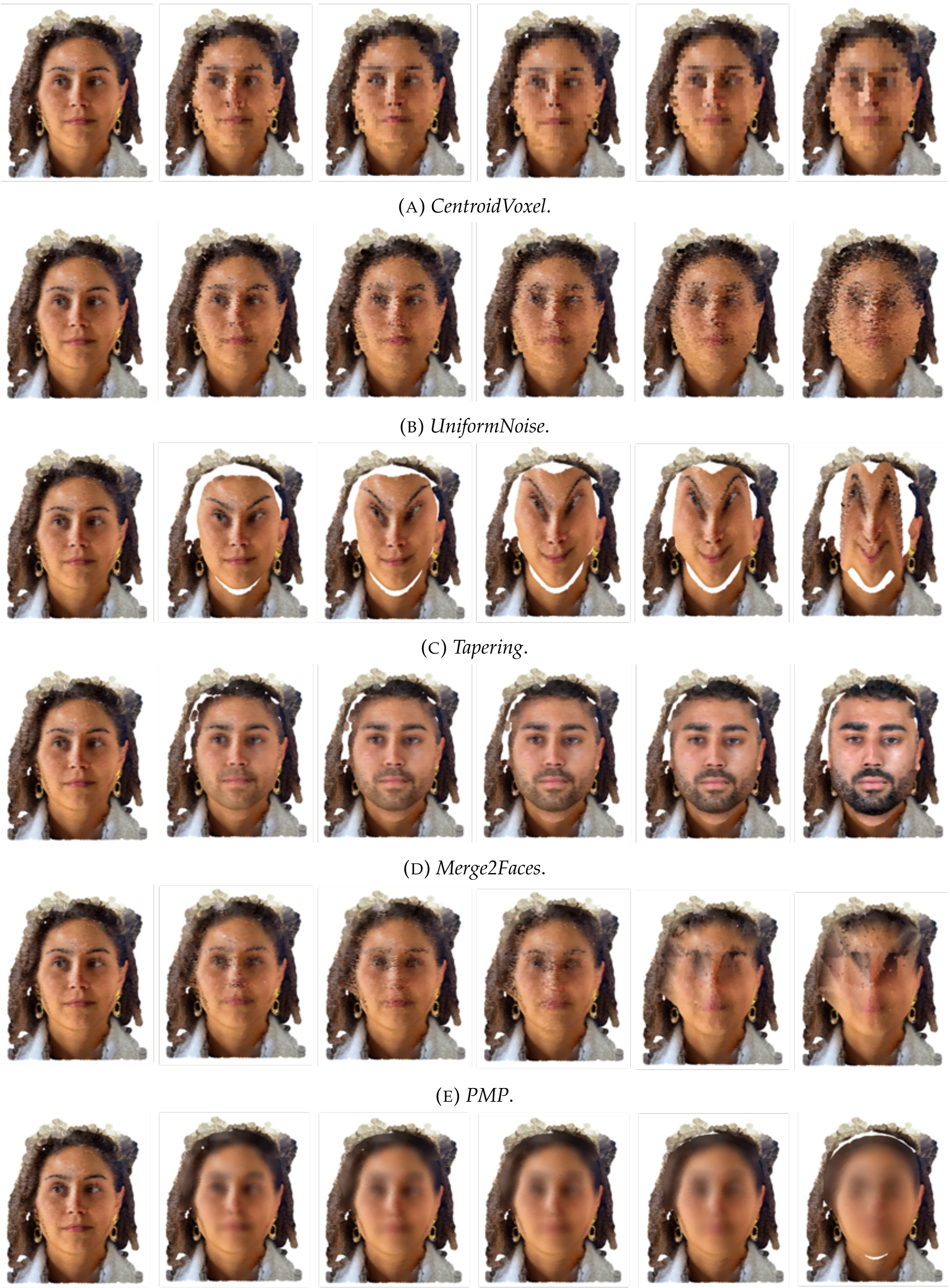
Evaluation and Results

This chapter signifies the culmination of the investigation, unveiling a thorough evaluation of the proposed anonymization techniques. The evaluation commences with a qualitative assessment of the visual outcomes for both the anonymization technique and the de-anonymization model. Subsequently, a quantitative assessment is conducted from both privacy and utility perspectives, separately. Finally, the overall effectiveness of each anonymization is measured, taking into account the privacy-utility trade-off.

5.1 Visual Results

The visual depictions presented in Figures 5.1 and 5.2 offer a representation of how distinct regulatory parameters influence privacy protection levels in the context of each anonymization approach applied to two subjects, a man and a woman. The choice of these parameters aligns with the AUC metric of the face recognition verification mode (presented in Section 2.3.2), with each image conveying anonymization outcomes across five equidistant intervals ranging from 0.5 to 1.0, with an interval size of 0.1. Starting from level 1, which corresponds to transformations producing an AUC within the range $[0.9, 1.0]$, and culminating in level 5, representing an AUC value within $[0.5, 0.6]$, these images provide a gradual insight into the impact of varying parameters. Note that these five levels constitute a subset of the 35 anonymization degrees detailed in Table 4.2 and were chosen for illustrative purposes.

In both figures, a gradual decrease in the AUC values is observed as moving from left to right, signifying an increase in privacy protection. The baseline image is included in the first column as a reference for comparative analysis, representing the 3D to 2D projection



(A) *CentroidVoxel*.

(B) *UniformNoise*.

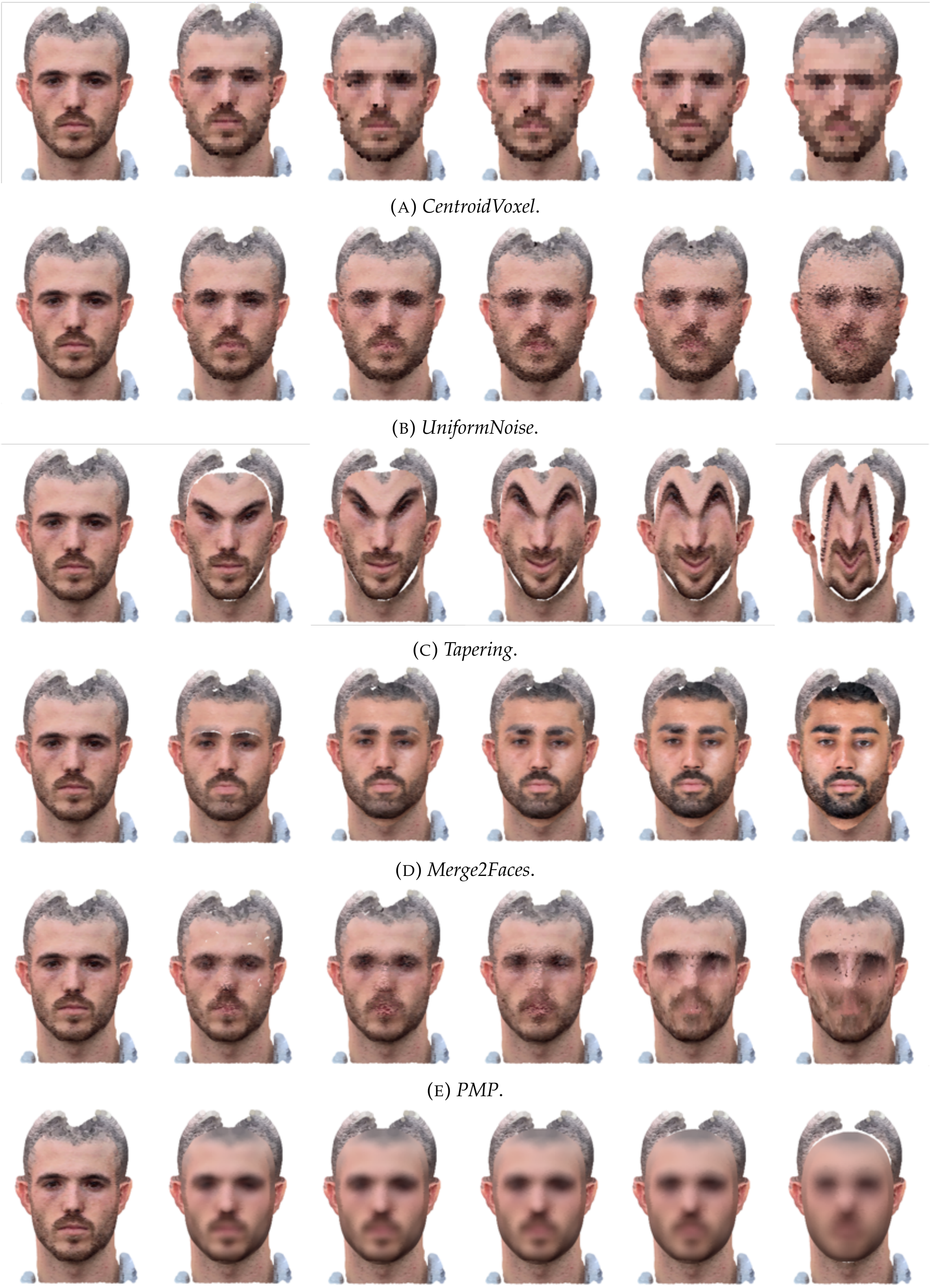
(C) *Tapering*.

(D) *Merge2Faces*.

(E) *PMP*.

(F) *SmoothKNN*.

FIGURE 5.1: Visual results of the anonymization techniques across five levels of privacy protection for subject *id_003*.



(A) *CentroidVoxel*.

(B) *UniformNoise*.

(C) *Tapering*.

(D) *Merge2Faces*.

(E) *PMP*.

(F) *SmoothKNN*.

FIGURE 5.2: Visual results of the anonymization techniques across five levels of privacy protection for subject *id_192*.

of the non-anonymized point cloud data for each subject. From the second to the last column, the images illustrate the projections of anonymized inner facial structures overlaid onto the outer facial features of the subject. This integration enhances the contextual comprehension of the anonymization process, as already explained. Additionally, the incorporation of additional techniques aimed at refining blending was deliberately avoided, ensuring the preservation of raw outcomes. This decision facilitates a deeper grasp of the authentic effects of anonymization on the original data, thereby offering better insights into the mechanisms at play.

5.2 Qualitative Evaluation

5.2.1 Anonymization Techniques

CentroidVoxel The *CentroidVoxel* anonymization method shares similarities with the 2D pixelization technique, resulting in a mosaic-like pattern of large, circular pixels. Once characterized by distinct and expressive features, facial details are transformed into a composition of coarse and oversized pixel clusters. These clusters replace the finer nuances of the face, simplifying features like eyes, nose, and mouth into arrangements of circular blocks. The edges defining facial contours exhibit noticeable irregularity, and the transitions between different facial regions lack the smoothness in the original depiction. While still recognizable, the colors display a diminished gradient and less nuanced shading, further contributing to the image's anonymized appearance.

UniformNoise The *UniformNoise* anonymization technique materializes as a scattered arrangement of random pixels that merge and overlay with the underlying facial features. The introduction of noise has resulted in a granular texture that partially obscures the subtleties of the face. In terms of color, the noise-infused image retains the original color palette while introducing subtle variations in pixel values. These variations are most evident in areas with gradual color transitions, where the noise generates a delicate dotted effect. Although the overall color accuracy is maintained, the tonal gradients experience a gentle disruption, contributing further to the image's anonymized appearance. However, some visual issues emerge as the noise increases, causing the points to extend beyond the facial region, as depicted in Figure 5.3.

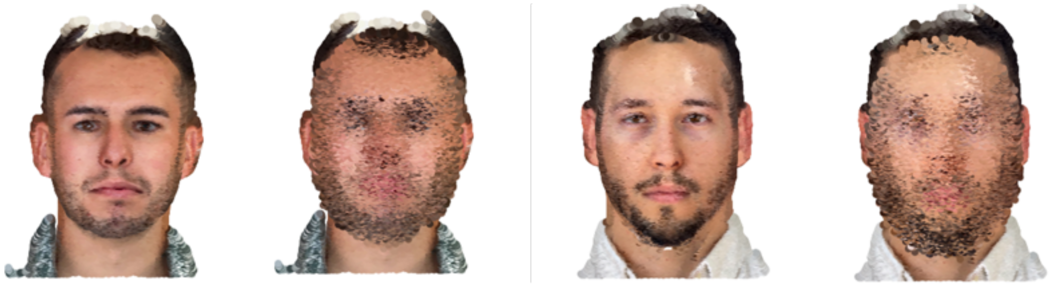


FIGURE 5.3: *UniformNoise* artifacts.

Tapering The *Tapering* technique, owing to its manipulation of facial geometry, yields a modified layout of facial features that diverges from their original recognizable form. The warping process introduces fluidity to the contours of the face, causing edges to curve and features to adopt new positions. This curvilinear influence gently softens facial boundaries, veering away from the structured appearance of the unaltered image. The extent of warping varies across distinct facial regions, with specific areas undergoing more pronounced adjustments than others (the eye region, for example). Regarding color and texture, the warped image retains the original color palette and the reshaping of facial geometry leads to changes in the arrangement of skin textures and tones. While not radical, these shifts contribute to altered visual perception, enhancing the overall anonymization impact. Nonetheless, the technique introduces artifacts linked to facial size, progressively compressing the face and resulting in significant visual alterations, as depicted in Figure 5.4.

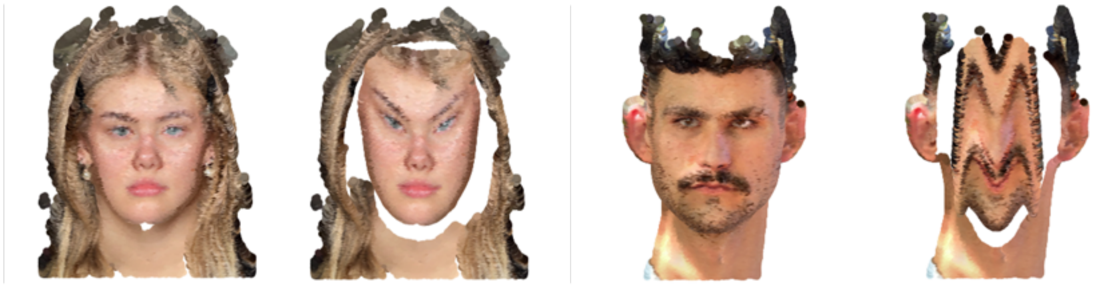


FIGURE 5.4: *Tapering* artifacts.

Merge2Faces The *Merge2Faces* technique amalgamates elements from two distinct facial sources. This method involves a deliberate merging of features and attributes, culminating in a composite appearance that challenges easy recognition and identification. The eyes, nose, mouth, and other distinguishing elements undergo a purposeful fusion, yielding an entirely novel facial configuration that gradually loses any immediate resemblance to the original visage. The resulting facial structure blends traits evenly, yet certain areas highlight more distinct characteristics from one source than the other. In terms of color and texture, the images often retain a unified color palette that slowly diverges from the original.

However, there may be instances of ghosting effects due to less than optimal alignment, particularly noticeable around the eye regions. This misalignment might also introduce white patches. Furthermore, the resulting anonymized image could experience a decrease in size compared to the original due to the prevalence of hair during face segmentation, causing it to diminish. All these artifacts are represented in Figure 5.5.

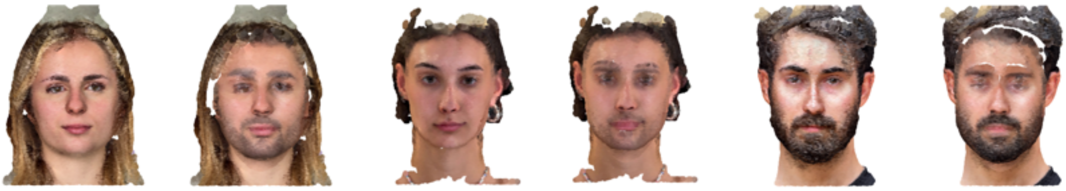


FIGURE 5.5: *Merge2Faces* artifacts.

PMP The *PMP* anonymization method shares similarities with the noise approach when introducing slight alterations to the data. However, it exhibits a distinct appearance when these values are higher. Facial features such as the nose, mouth, and eyes gradually become obscured and lose finer details. The resultant effect resembles a coarse sculpture of the face with diminishing levels of detail. In terms of color, the technique modifies the palette, leading to regions with distinct colors that might not harmonize well with the facial region. For lower regulating parameter values, the anonymization introduces noticeable artifacts. Only a few regions contain points, particularly those that mark the transition between facial features. Moreover, with higher parameter values, the technique can yield peculiar colors or regions with sparse point density, as depicted in Figure 5.6.

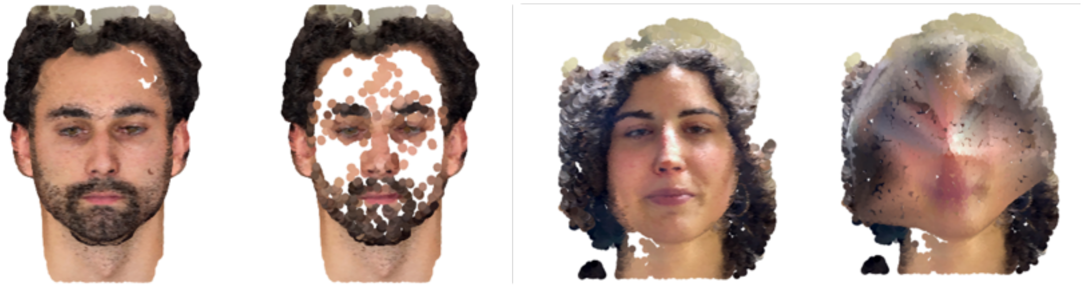


FIGURE 5.6: *PMP* artifacts.

SmoothKNN The *SmoothKNN* technique bears a resemblance to the blurring 2D anonymization approach, producing a comparable effect of gentle pixel diffusion across the facial contours. By applying the blurring effect, pixel values across facial contours gradually diffuse, resulting in a softened and slightly obscured depiction that eradicates features like the eyes, nose, mouth, and others. Particularly around facial features, the blurring is most prominent, subtly tempering fine textures and edges. Consequently, these edges become less sharp, creating a more diffused boundary between distinct facial areas. Concerning color and tonal qualities, the blurred image adheres to the original color palette but presents a unified and harmonious appearance due to the softened edges. This aspect fosters cohesiveness among facial regions while reducing distinct color patterns or tonal variations that could facilitate recognition. However, as modification levels increase, the technique leads to a reduction in the size of the face. Additionally, it might introduce an augmentation in the visibility of pre-existing white patches, as illustrated in Figure 5.7.

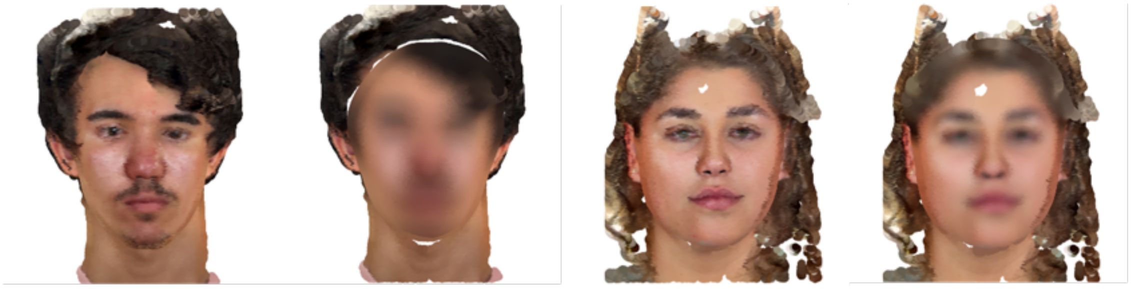


FIGURE 5.7: *SmoothKNN* artifacts.

5.2.1.1 Discussion

The visual outcomes of the anonymization techniques underscore their effectiveness in safeguarding the individual's identity, particularly at higher privacy levels, as evidenced in Figures 5.1 and 5.2. However, the degree of identity protection varies among the techniques, which also influences their realism.

For example, the *Merge2Faces* approach begins to conceal the subject's identity at lower privacy levels. At this stage, it becomes challenging for humans to discern the facial attributes of the anonymized subject due to the blending of facial features. Furthermore, this merging process yields convincing results that nearly give the impression of an absence of any anonymization.

Conversely, the *Tapering* technique exhibits the least authenticity. The pronounced deformations it introduces severely alter the facial shape, proportions, and attributes, leading to a less genuine appearance. Nonetheless, these deformations significantly hinder the human ability to identify the subject, even at lower privacy levels.

The *CentroidVoxel* method appears to have less impact, with subject *id_192* showing a relative resemblance to the original image even at high levels, especially for the male subject in Figures 5.1 and 5.2. A similar but less intense effect is observed with *SmoothKNN*. However, the latter approach offers a smoother transition between the facial attributes, resulting in a more aesthetically pleasing appearance, even though it may not achieve a high degree of realism.

In contrast, the *UniformNoise* and *PMP* techniques yield similar outcomes at lower anonymization levels, but their distinctions become more pronounced as the intensity increases. Notably, the *PMP* approach appears to better conceal the subject's identity while imparting a more visually pleasing aesthetic.

5.2.2 De-Anonymization

The de-anonymization is an integral component of the *reversibility recognition*, for which an autoencoder has been designed and trained. This attacker paradigm operates by first reversing the anonymization results. The obtained de-anonymized set will form the probe set, containing the faces used to perform the recognition task against the gallery set, containing the 200 subjects with a known identity for both verification and identification face recognition modes.

The autoencoder results regarding the de-anonymization of the six techniques indicate that the model's performance is unsatisfactory, as the generated images closely resemble the anonymized ones, showing minimal reversibility for all the algorithms. Figure 5.8 showcases the outcomes of the trained model for subject *id_047* during testing. The top row displays the original image, the middle row displays the anonymized version, and the bottom row presents the de-anonymization results achieved by the Autoencoder. These outcomes serve as indicative results for the remaining 19 identities that constitute the entire testing set.

While the model's inability to reverse anonymization techniques may be partially attributed to the strengths of the anonymization methods, it is plausible to assume that

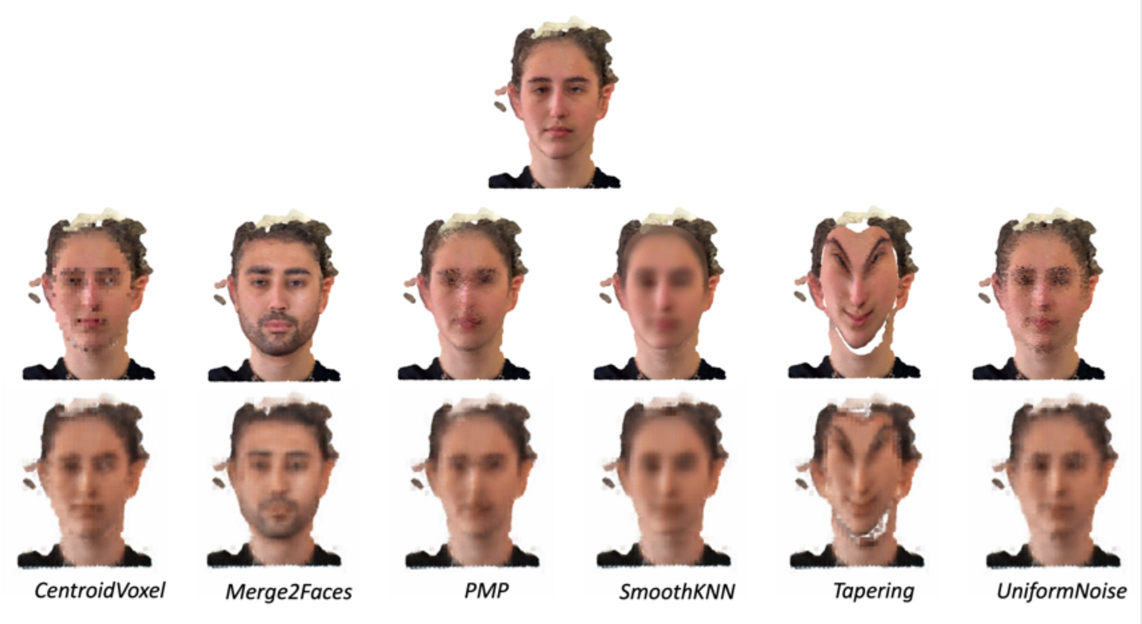


FIGURE 5.8: Results of de-anonymization achieved by the trained Autoencoder.

the primary issue lies in the insufficient training data. Hence, due to the substantial similarity between the anonymized and de-anonymized images, it is unnecessary to proceed with further quantitative evaluation, as the results would likely yield similar metrics for both sets of images.

Consequently, despite the *reversibility recognition* potential, this approach could not be seamlessly integrated into the evaluation framework of this study. As an alternative to addressing data scarcity, one could consider employing specialized approaches tailored for each anonymization technique, such as deblurring or denoising. These methods have shown success in reversing simple anonymizations and might offer more effective solutions in this particular scenario. Therefore, the subsequent quantitative evaluation relies solely on the *naïve recognition* paradigm.

5.3 Quantitative Evaluation

5.3.1 Privacy Metrics

The privacy metrics considered encompass a total of two metrics regarding the two face recognition modes of verification and identification under closed-set. In the context of verification assessment, the employed metric was the AUC value, whereas for identification under closed-set analysis, the metric used was the Rank-1 Identification Rate, represented as a vertical cross-section within the CMC plot. As a reminder, both metrics are elaborated upon in Section 4.2.3, encompassing both their description and the rationale behind their computation.

To ensure reader clarity and avoid overwhelming numerous results, the six anonymizations along their anonymization degrees are synthesized into a unified graph for each mode. Nonetheless, the comprehensive evaluation of each anonymization technique is presented in Appendix B.1, encompassing ROC curves (used to compute the AUC) and CMC curves (used to compute the Rank-1 Identification Rate).

The AUC lies within the range of $[0, 1]$. An AUC score of 1 indicates minimal privacy protection, while a score of 0.5 signifies maximum protection. An AUC score of 1 emphasizes that when confronted with an identity claim of an anonymized subject, the face recognition model can consistently and accurately accept or reject that claim by establishing a one-to-one comparison between the anonymized subject and the claimed identity. Conversely, an AUC of 0.5 indicates that the face recognition model performs no better than a random classifier when accepting or rejecting the claim, indicating a loss of its discernibility ability. Values below 0.5 are often indicative of incorrect model configurations or instances where the model mistakenly identifies the negative class as positive. Therefore, attention is primarily directed toward the interval $[0.5, 1]$.

The Rank-1 Identification Rate lies within the range of $[0, 1]$. A Rank-1 Identification Rate score of 1 indicates minimal privacy protection, while a score of 0 indicates maximum protection. A Rank-1 value of 1 signifies that when determining the identity of an anonymized subject, the face recognition model correctly identifies the subject by returning their identity label as the best match from the pool of 200 identities, achieved through a one-to-many comparison. In simpler terms, the model always ranks the real identity of the anonymized subject as the top match, resulting in an identification rate of 100%. Conversely, a Rank-1 score of 0 means that the face recognition model consistently fails to identify the true identity of the anonymized subject as the best match from the pool of 200 identities. In simpler terms, the model never ranks the real identity of the anonymized subject as the top match, resulting in an identification rate of 0%.

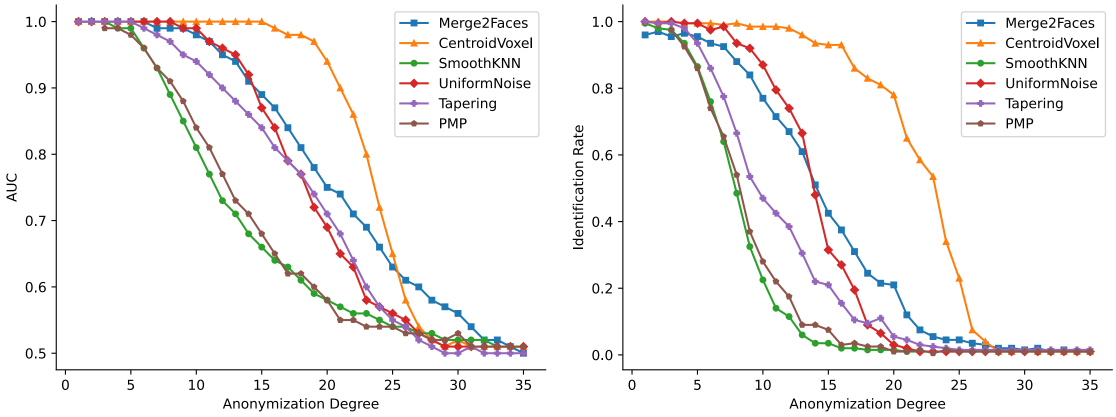
5.3.1.1 Comparative Analysis

The privacy metrics reveal that the baseline set, represented by the projection of the subjects' 3D point clouds into 2D space without any anonymization, achieves a Rank-1 Identification Rate of 0.995 and an AUC of 1 (dotted lines in Appendix B.1). These values indicate that a face recognition model exhibits nearly perfect recognition ability in both verification and identification tasks. Thus, it underscores that the quality of the iPhone-PLYv3 dataset does not compromise the research, despite the constraints outlined in Section 4.1.5. It also highlights the effectiveness of the 3D to 2D projection in preserving the facial traits of the subjects.

In line with the baseline set results, the reference values for both metrics used for comparison with the anonymization outcomes are approximately 1. Figure 5.9 illustrates the two privacy metrics for the anonymization techniques across the considered 35 privacy levels, ranging from 1 to 35, representing varying levels of anonymization intensity. This figure confirms that all the techniques are capable of providing a broad spectrum of privacy protection levels, spanning from minimal to complete. This is evident from the range of values across the maximum variation amplitude of both the AUC and Rank-1 Identification Rate, which span from their lower limits of 0.5 and 0, respectively, to 1. Such a continuous spectrum of privacy protection is achieved by adjusting the corresponding control parameter to achieve the desired level of protection. However, the control parameter does not uniformly affect the protection levels, as most of the curves exhibit non-linearity in both face recognition modes. An exception is the *Merge2Faces* technique, which demonstrates some linearity within specific ranges of anonymization levels, for AUC between levels 12 to 32 and for Rank-1 between levels 12 to 18. Besides, the steepness of the curves can impact how easily parameter fine-tuning achieves the desired level of privacy. Particularly for methods with steeper gradients, slight variations in the parameter can lead to significant differences in privacy protection levels. For instance, the control parameter of the *SmoothKNN* and *PMP* techniques has a substantial impact on privacy protection levels, as evidenced by the steep slopes of the curves between anonymization degrees ranging from 5 to 15 in both plots.

Comparing both plots of Figure 5.9 reveals that the Rank-1 Identification Rate is more responsive to changes in the anonymization degree than the AUC. This is confirmed by the fact that, at a given anonymization degree, the high level of privacy protection achieved in face recognition verification (as represented by Rank-1) does not correspond to the same high level of identification (as represented by the AUC). For example, while the *SmoothKNN* and *PMP* techniques reach a maximum Rank-1 protection value at an anonymization degree near 20, at the same degree, the AUC still has a score of approximately 0.6. This pattern is consistent among the other techniques as well. This conclusion suggests that ensuring protection in verification also implies protection in identification.

The *CentroidVoxel* technique presents a unique profile, indicating minimal impact over a broader range of lower anonymization degrees up to 15. Additionally, the proximity of the *SmoothKNN* and *PMP* techniques in both plots confirms that these two techniques have a similar influence on privacy protection levels across anonymization degrees. For the *PMP* algorithm, the first two anonymization degrees are considered outliers and have been excluded from the evaluation, as they do not provide relevant information. At both of these levels, the privacy protection outliers are characterized by an AUC and Rank-1 close to 0.5 and 0, respectively. This behavior can be attributed to the presence of a significant number of holes caused by low α values, an intrinsic characteristic of α-shapes explained in the qualitative evaluation and illustrated in Figure 5.6. The remaining three anonymization techniques, namely *Merge2Faces*, *UniformNoise*, and *Tapering*, have a similar effect on the AUC, while in the Rank-1, *Tapering* offers slightly more protection at lower anonymization degrees, as its curve is below the other two.



(A) AUC.

(B) Rank-1 Identification Rate.

FIGURE 5.9: Privacy metrics comparison for the anonymization techniques.

5.3.2 Utility Metrics

The utility metrics considered encompass a total of four metrics, each targeting different aspects of utility. These metrics complement each other in assessing the overall anonymized data utility, as explained in Section 4.2.3. Specifically, the four metrics and their respective evaluative utility components are as follows:

- *Delta Detection:* This metric evaluates the anonymization's ability to maintain a human-like facial appearance. It quantifies this ability by measuring the impact on the accuracy of a face detection algorithm in identifying faces;
- *Landmark Distance:* This metric assesses the anonymization's capability to preserve the overall facial structure of the original face. It does so by calculating the average distance between five landmarks on both the anonymized and non-anonymized instances;
- *SSIM:* It measures the image quality of the anonymized face in comparison to the original image. It provides a correlation with human perception, indicating how closely the anonymized image resembles the original;
- *FID:* It also evaluates the image quality of the anonymized face compared to the original image. However, it focuses on the distinction between the feature vector distributions for the anonymized and non-anonymized image sets.

The *Delta Detection* score falls within the range of $[0, 1]$. A Delta Detection score of 1 signifies that the anonymization effectively preserves the human-like appearance of the face, allowing a face detector to correctly identify all anonymized faces among the 200 subjects. This value is attainable because the face detector achieves an accuracy of 1 on the baseline set, which corresponds to the non-anonymized 3D to 2D projected images[1]. Conversely, a Delta Detection score of 0 indicates that the anonymization entirely disrupts the human-like appearance of the face, rendering the face detector incapable of detecting any faces.

The *Landmark Distance* metric ranges from 0 to 40. A Landmark Distance score of 0 signifies that the facial structure of the face is fully preserved, indicating that the locations of the eyes, nose, and mouth remain exactly the same as in the original. Conversely, higher values indicate that the configuration of the facial structure for the anonymized subjects deviates from the original, which has a detrimental effect on the face detector's ability to accurately determine the locations of these five landmarks. The value 40 represents the penalty imposed whenever the face detector fails to identify any of the five landmarks.

The *SSIM* metric ranges from 0 to 1. An SSIM score of 1 suggests that the image quality of the anonymized image closely resembles that of the non-anonymized image,

[1] This accuracy level serves as the upper quality bound that an anonymization cannot surpass.

indicating a perfect similarity, while an SSIM score of 0 implies no similarity between the two images.

The *FID* metric falls within the range of $[0, \infty]$. Lower FID scores have been demonstrated to be associated with higher-quality images. An FID score of 0 signifies that the sets of anonymized and non-anonymized images are identical in terms of their feature distributions, indicating the highest level of image quality.

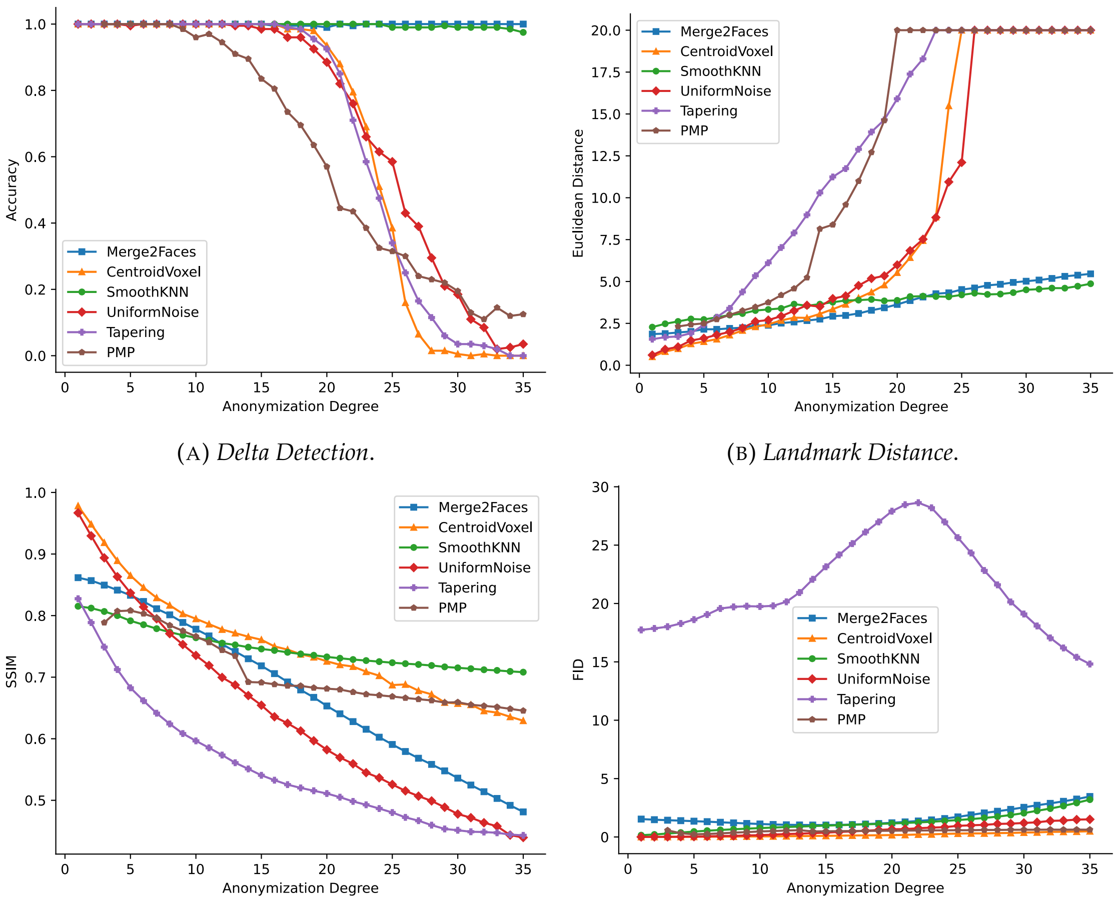
Following the same approach as with the privacy metrics, all anonymization variants have been synthesized into four distinct graphs representing the four selected utility metrics. These graphs offer a simplified view of the comprehensive results presented in Appendix B.2, which includes additional details. In the Appendix graphs, for each level of anonymization, mean and standard deviation values of the Intersection over Union (IoU) corresponding to detection accuracy for Delta Detection are provided, violin plots illustrate SSIM and Landmark Distance, and the plain FID scores are presented. In contrast, the four synthesized graphs that follow focus solely on depicting detection accuracy, median Landmark Distance, and average SSIM, while the FID scores remain consistent.

In addition, the inference time will also be discussed, although it is not a utility metric for evaluating the anonymization results; rather, it pertains to the performance of the anonymization process itself.

5.3.2.1 Comparative Analysis

The utility metrics rely on the baseline set as a reference for computing utility values. This baseline set has been demonstrated to possess a high quality that the anonymization techniques strive to maintain. Figure 5.10 presents the outcomes of the four utility metrics across 35 privacy levels, spanning from 1 to 35, which correspond to different degrees of anonymization impact. Similar to the privacy metrics, the first two anonymization outlier levels of the *PMP* technique are excluded from consideration. The subsequent analysis addresses each utility metric individually and is supported by the data presented in this figure.

Delta Detection The *Merge2Faces* and *SmoothKNN* techniques consistently demonstrate high face detection accuracy across all levels of anonymization. The former maintains an accuracy of no less than 0.995, while the latter stays above 0.975. This indicates that even



(A) *Delta Detection.*

(B) *Landmark Distance.*

(C) *SSIM.*

(D) *FID.*

FIGURE 5.10: Utility metrics comparison for the anonymization techniques.

though these techniques significantly alter the facial features of the subjects in the original image, the anonymized results still closely resemble human faces, allowing them to be recognized as such. In contrast, the *CentroidVoxel*, *UniformNoise*, and *Tapering* techniques start to introduce noticeable changes in the appearance of the results, making them less recognizable as faces, starting from an anonymization level of 20. This continues until they reach a point where no faces can be identified among the 200 subjects. However, starting from level 25, the *UniformNoise* technique has a less severe impact on facial detection, as indicated by the milder slope of its curve. For example, while the other two techniques fail to detect any faces at level 30, the noise-based approach can identify 18.5% of the faces (37 subjects) at level 25. The *PMP* technique exhibits an earlier decline in accuracy, starting at level 8 and decreasing at a slower rate than the three previous techniques. By level 25, it performs better in facial detection than the sampling-based and deformation-based techniques, and it even surpasses the noise-based technique at level 29. Ultimately, at level 35, the *PMP* achieves a relatively higher face detection accuracy of 12.5% (25 subjects) compared to the other three techniques. However, the difference between the other two top-performing methods is still significant.

Landmark Distance The *Merge2Faces* and *SmoothKNN* techniques result in a minimal alteration to the facial structure of the subjects, with a median distance of the five landmarks not exceeding 5. Both techniques exhibit an approximately linear trend with slight variations in the values. In contrast, the *Tapering* technique consistently shows the highest median distance between the facial landmarks in comparison to all the techniques, likely due to its inherent nature of deforming the facial structure. The *PMP* technique introduces the second-largest difference. While it doesn't directly modify the facial dimensions as the warping-based method does, the loss of detail and the more uniform appearance of facial features compromise the detector's ability to place the landmarks accurately, resulting in unusual positions such as eyes on the forehead and exaggerated mouth inclinations. The remaining two anonymization techniques, *CentroidVoxel* and *UniformNoise*, exhibit a similar exponential increase, comparable to *Merge2Faces* and *SmoothKNN*, until an anonymization degree of 10.

SSIM The *SmoothKNN* technique exhibits a relatively narrow range of variation, starting with a high SSIM value of 0.82 at its maximum and gradually decreasing to 0.71 at its lowest point. In contrast, the *UniformNoise* technique displays the widest amplitude, initiating with an SSIM value close to one and progressively decreasing to less than 0.5 at the most extreme anonymization levels. Both the *CentroidVoxel* and *PMP* techniques follow a similar trajectory, with their SSIM values reaching around 0.65 for anonymization degrees greater than 6. The dip registered in the SSIM values of the *PMP* technique between levels 13 and 14 are attributed to changes in sorting indices as explained in Section 4.3.2. This pattern of fluctuations is consistently observed in all the graphs. As for the *Merge2Faces* and *Tapering* techniques, they exhibit a comparable range of image quality variation. However, the *Tapering* technique consistently yields lower SSIM values, primarily due to the shrinkage of facial dimensions it introduces.

FID The *Tapering* technique consistently exhibits higher values than all other methods across various privacy levels. Unexpectedly, this anonymization does not peak at higher levels; instead, it demonstrates a decrement from levels beyond 20, ultimately reaching values lower than those at lower anonymization degrees. This distinct behavior lacks a clear explanation. As the curves of the remaining anonymizations might be challenging to discern because of the value discrepancy, referring to Appendix B.2 can enhance insights by providing the values of the metric individually. Concerning the *CentroidVoxel* and *UniformNoise*, they exhibit an exponential increase, with disparities in the FID amplitude, ranging respectively between $[0.0016, 0.5]$ and $[0.002497, 1.5]$. Unlike expected, *Merge2Faces* demonstrates a decline from 1.53 to 1.02 in the mid-range anonymization levels, followed by an increase up to a value of 3.47. This metric's unexpected behavior lacks a clear rationale. In fact, the presence of white patches in the middle and facial contours, which result from poor alignment and segmentation, could theoretically increase the metric for intermediate anonymization values, the opposite occurs. *PMP* experiences a swift incline in FID from 0.23 to 3.20. This incline is marked by a slight slope increase, which is preceded by a disruption attributed to ordering indices. Following this, it maintains an almost constant behavior until reaching 0.62. Lastly, *SmoothKNN* displays an increase from 0.15 to 3.20, characterized by a cubic trend that gradually intensifies as parameter values increase.

Another crucial aspect to consider when evaluating anonymization techniques is their execution time. Although not classified as a utility metric for the anonymization outcomes, execution time offers a more comprehensive gauge of anonymization's overall practicality. This metric serves as a criterion for distinguishing techniques suited for real-time demands from those requiring offline processing. In Figure 5.11, a comparison shows the mean execution times for processing one face from the 200, considering all parameters. The range of mean execution times spans significantly across all anonymization methods, ranging from 0.00004 to 0.6751 seconds. However, it's worth noting that relying solely on the mean value can be misleading, as some anonymization techniques exhibit substantial standard deviations, such as *SmoothKNN* and *CentroidVoxel*. This highlights the fact that the chosen regulating parameter profoundly influences the latency period.

To provide more insightful findings, Figure 5.12 comprehensively illustrates the execution times across all anonymization techniques in relation to the anonymization degree, related to the regulating parameter. It's important to note that the y-scale, corresponding to the execution time, differs for each anonymization. The execution time of *CentroidVoxel* exhibits a descending trend as the regulating parameter increases. This technique initiates at approximately 0.045 seconds and progressively diminishes to under 0.005 second. Consequently, as the level of anonymization intensifies, the execution time experiences

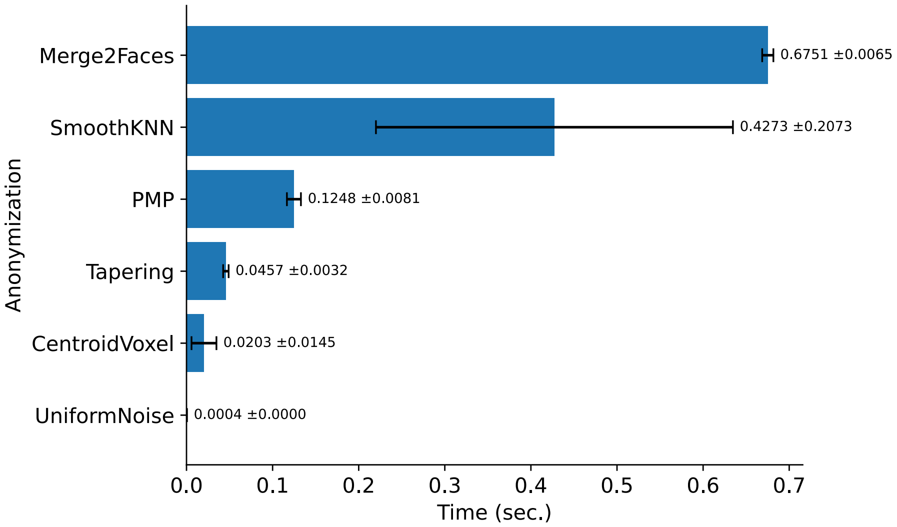
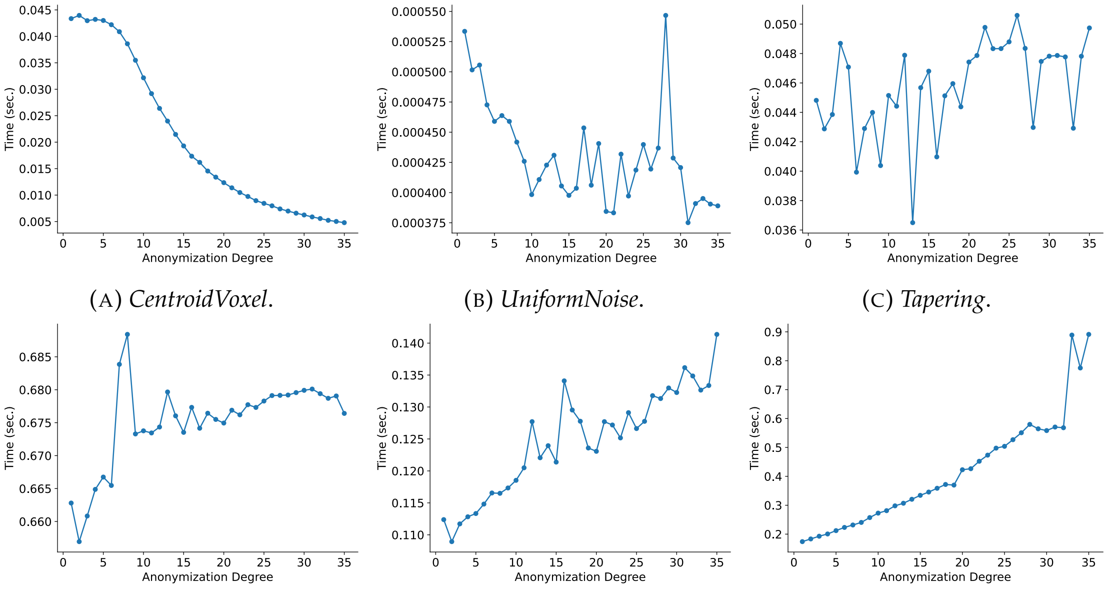


FIGURE 5.11: Mean and standard deviation execution time for all the anonymization techniques.

a notable exponential reduction. This feature favors the technique for scenarios where rapid inference precedes utility considerations. Conversely, the latency of *SmoothKNN* linearly increases, with a pronounced surge at an anonymization degree of 33, corresponding to nearly 300 neighbors. Starting at around 0.18 seconds, it escalates to over 0.85 seconds. In contrast, the *PMP* technique displays a relatively linear progression in execution time, spanning from 0.11 seconds to over 0.14 seconds. Its amplitude of variation is smaller, merely 0.03 seconds, compared to the preceding techniques at 0.04 and 0.7 seconds. Meanwhile, *Merge2Faces* showcases a marginally ascending execution time from 0.66 until reaching 0.69, after which it remains relatively stable around 0.67, causing an amplitude shift of merely 0.026 seconds in total. On the other hand, both *UniformNoise* and *Tapering* exhibit execution times largely unaffected by the regulating parameter, both maintaining a relatively stable duration with only occasional random fluctuations.

5.3.3 Privacy-Utility Trade-Off

The privacy protection and data utility achieved by an anonymization technique are delineated by an adversarial relationship. This analysis of the relationship, commonly referred to as the privacy-utility trade-off, is paramount as it is the primary criterion for measuring the overall anonymization effectiveness. The trade-off between privacy and utility is explored through three visualizations: a pair plot, a correlation matrix, and a



(A) *CentroidVoxel*. (B) *UniformNoise*. (C) *Tapering*.

(D) *Merge2Faces*. (E) *PMP*. (F) *SmoothKNN*.

FIGURE 5.12: Execution time of each anonymization technique.

scatter plot that illustrates the connection between mean privacy and utility, as elaborated upon later.

5.3.3.1 Comparative Analysis

Figures 5.13 and 5.14 present a pair plot encompassing all privacy and utility variables, along with the corresponding correlation matrix. These visual representations complement each other and are jointly analyzed to extract valuable insights. Before delving into the analysis of the privacy-utility trade-off, it is beneficial to examine the relationships between privacy and utility metrics independently.

Privacy Metrics Both privacy metrics, AUC and Rank-1 Identification Rate, display a strong positive correlation of 0.93 (as shown in Figure 5.14), suggesting a robust alignment between privacy protection levels in both verification and closed-set identification scenarios. Additionally, Figure 5.13 highlights that the *CentroidVoxel* exhibits a unique characteristic compared to the others, demonstrating a linear relationship between both metrics, while the remaining techniques share a similar trait, nearly all falling within a curve with negative concavity.

Utility Metrics The *FID* metric demonstrates a moderate negative correlation of -0.51 with *SSIM*, which is somewhat lower than expected. Thus, while both metrics evaluate the quality of images, they shed light on different elements associated with "quality", which in essence is a subjective concept (despite the metrics being objective numerical measurements). This emphasizes the significance of the careful selection of the evaluation metrics based on the specific requirements of the use cases. Moreover, *FID* shows a weak correlation of 0.28 with *Landmark Distance* and virtually no correlation of -0.095 with *Delta Detection*, indicating the absence of a linear relationship between these metrics. In contrast, *SSIM* exhibits a moderate correlation of -0.67 with *Landmark Distance* and a positive correlation of 0.54 with *Delta Detection*. This suggests that decreases in image quality correspond to disruptions in facial structure and reduced facial detection accuracy, as anticipated. The strong negative correlation of -0.93 between *Delta Detection* and *Landmark Distance* is logical, as the inability to detect faces and landmarks is inherently interconnected. Consequently, these variables tend to approach their extreme values jointly.

Privacy-Utility Metrics Both the AUC and Rank-1 privacy metrics exhibit a similar relationship with the four utility metrics, as illustrated in Figure 5.13. The overall arrangement of the curves remains consistent, with the same order and relative positions maintained throughout. However, Rank-1 tends to show a more pronounced decline in comparison to AUC when evaluating the same values of utility metrics. For example, when examining the *UniformNoise* technique against Landmark Distance values in the range of 5 to 10, Rank-1 values hover around 0, while AUC reaches values as high as 0.6 [1]. Additionally, Rank-1 curves often feature steeper inclines, indicating regions of abrupt changes across utility metrics and anonymization degrees. Another conclusion pertains to the adversarial relationship between privacy and utility metrics, as demonstrated in Figure 5.13 where data utility diminishes as privacy protection is enhanced. This relationship is further substantiated by the presence of moderate to strong correlations between the two metric categories, with absolute correlation values falling within the range of 0.51 to 0.72, as illustrated in Figure 5.9. The *FID* metric deviates from this trend, as it does not exhibit a significant correlation with the two privacy metrics.

Acknowledging the intrinsic adversarial nature of privacy and utility metrics, the central goal and challenge of any anonymization technique lies in attaining an equilibrium between these elements. To capture and convey information about the effectiveness of

[1] Still, do note the different range variation between the metrics of $[0, 1]$, $[0.5, 1]$.

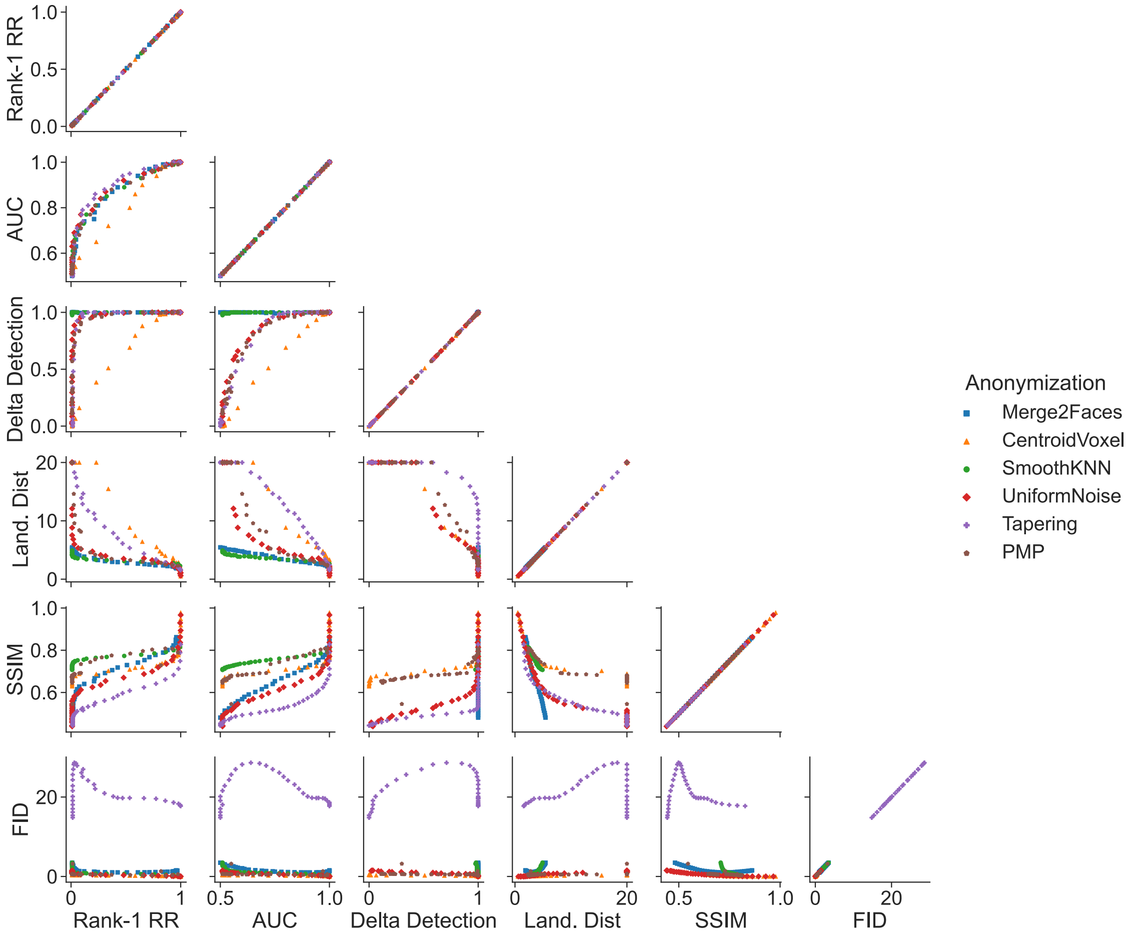


FIGURE 5.13: Pair plot for the privacy-utility trade-off.

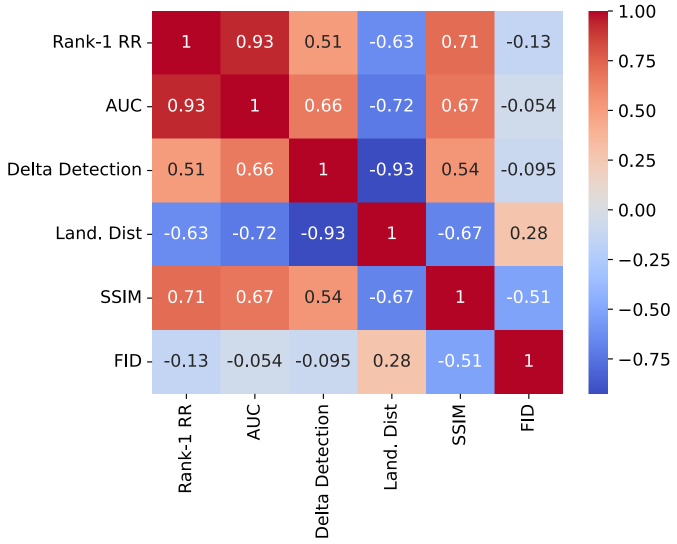


FIGURE 5.14: Correlation matrix for the privacy-utility trade-off.

anonymization while emphasizing the trade-off, a novel visualization approach is proposed. This approach amalgamates the mean privacy and utility into a unified plot by consolidating the two privacy metrics through the use of the mean. As the AUC metric range was considered between 0.5 and 1, it was rescaled to fit within the $[0, 1]$ interval. The *Rank-1* identification rate already resides within this interval. The perceptibility of the visualization was optimized by inverting the values and setting a mean privacy score of 1, signifying maximum protection, while a score of 0 denoting minimum protection. Shifting the focus to the utility metrics, the *Delta Detection* accuracy and SSIM were employed. Both of these metrics range from 0 to 1, wherein higher values correspond to heightened utility. In this context, the *Landmark Distance* metric was omitted due to its high correlation with *Delta Detection*, and due to the absence of a straightforward conversion mechanism that would enable its comparison on a uniform scale, the same goes for the *FID*. The conceptual underpinning of mean privacy and mean utility is defined as follows:

$$\mu_{privacy} = 1 - \frac{norm(AUC) + Rank1}{2} \tag{5.1}$$

$$\mu_{utility} = \frac{DeltaDetection + SSIM}{2} \tag{5.2}$$

Figure 5.15 provides a visual representation of the overall performance demonstrated by the anonymization techniques concerning privacy and utility standpoints. The x-axis denotes mean privacy, while the y-axis signifies mean utility, with both variables being calculated according to Equations 5.1 and 5.2. Anonymization configurations that achieve a favorable balance between privacy and utility are positioned in the upper-right corner, representing the most satisfactory privacy-utility trade-off. Conversely, points in the lower-left region indicate instances with the least desirable trade-offs, resulting in both low privacy and utility. However, due to the inverse relationship between the two variables, observations lying below the diagonal line $y = 1 - x$ are considered atypical. Additionally, instances located in the upper region of the plot signify anonymization configurations that are more effective in providing privacy protection, while those on the right are better at preserving data utility.

The least effective technique is the *CentroidVoxel*, which exhibits a severe compromise in data utility at the expense of privacy protection. For instance, when targeting a mean privacy level surpassing 0.8, the technique results in a mean utility hovering around 0.4.

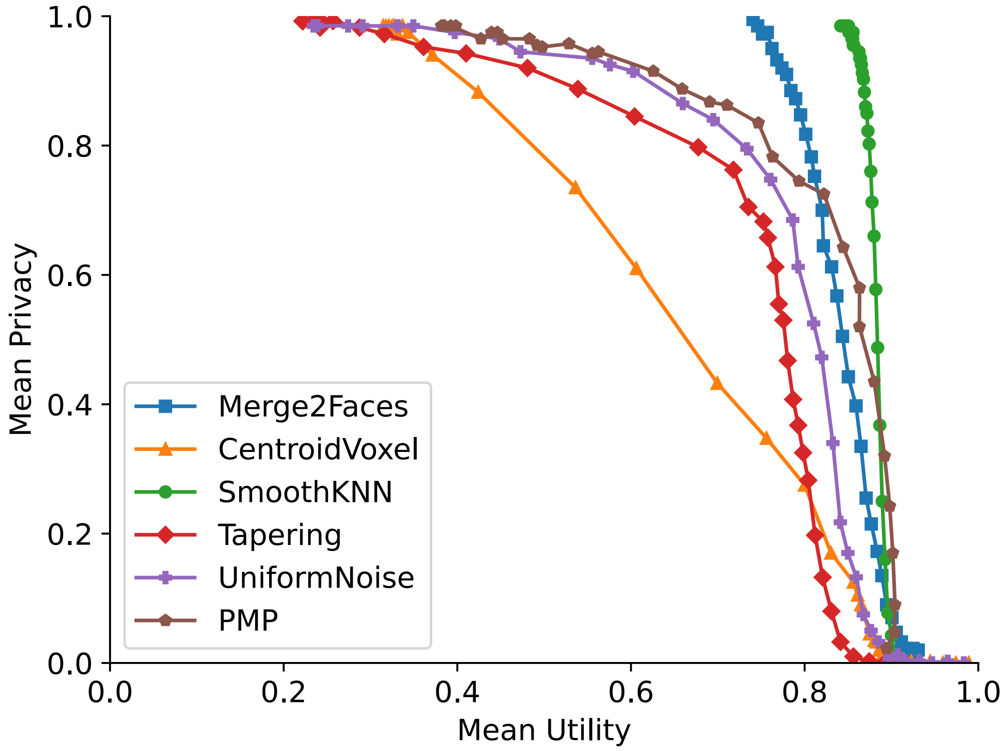


FIGURE 5.15: Privacy-utility trade-off for the anonymization techniques, with the proposed mean aggregation solution.

The deficiency in privacy protection is further supported by the qualitative evaluation, where it appears that a human observer may be able to visually identify the identity of the anonymized subject even for the higher anonymization degrees.

The trade-off curves of *Tapering*, *UniformNoise*, and *PMP* exhibit a characteristic shape characterized by a non-linear, concave-down curve. Despite their similar shapes, *PMP* yields more robust results compared to the other two, as it is positioned closer to the upper-right region. Consequently, *PMP* ranks as the third most suitable option among the evaluated anonymization techniques. Nonetheless, when considering a mean privacy level exceeding 0.8, this technique exhibits nearly identical values compared to *UniformNoise*, with both curves almost superimposed. Specifically, the *PMP* technique demonstrates the best ability to preserve the most utility when the privacy protection falls within the range of approximately 0.1 to 0.4. However, it's worth noting that typically, more emphasis is placed on the heightened privacy component, rendering this status less relevant. The *Tapering* is positioned as the worst between the three and it is the second least satisfactory anonymization technique in terms of trade-off results, surpassing the *CentroidVoxel*.

Finally, Figure 5.15 also demonstrates that both the *Merge2Faces* and *SmoothKNN* techniques excel in achieving a favorable balance between privacy and utility, with the latter having a slight edge. This is supported by their nearly vertical trade-off lines positioned on the right side of the plot, where the mean utility exceeds 0.74. This indicates that increasing the intensity of anonymization has minimal impact on the utility of the data. In particular, by adjusting the regulating parameter configuration for both techniques, *Merge2Faces* reaches a coordinate of $(0.95, 0.76)$, while *SmoothKNN* achieves $(0.95, 0.86)$. However, the qualitative evaluation suggests that *Merge2Faces* may offer greater reliability in terms of privacy, as it can significantly attenuate or even replace facial traits in an extreme case. Additionally, the visual appearance of the anonymization is more realistic, and the anonymization is less perceptible, almost unnoticed, in comparison with the *SmoothKNN*.

5.3.3.2 Summary

The trade-off results are closely tied to the metrics used to assess both privacy and utility. In this regard, the pair plot and correlation matrix in Figures 5.13 and 5.14 provided initial insights into the privacy-utility trade-off for the anonymization techniques and facilitated the selection of the most relevant metrics for characterization. As a result, the proposed mean performance solution was implemented to consolidate all the key evaluation components relating to both privacy and utility within a single graph. This approach enabled not only the ranking of anonymization techniques based on their effectiveness and the highlighting of their strengths and weaknesses but also the determination of the appropriate control parameter for an effective privacy-utility trade-off. When combined with the qualitative evaluation, this evaluation procedure identified *Merge2Faces* and *SmoothKNN* as the most effective techniques, potentially suitable for a wide variety of use cases, owing to the comprehensive nature of the proposed evaluation.

Chapter 6

Conclusion

The goal of this thesis was to create 3D face anonymization methods for point clouds and perform a thorough assessment of their effectiveness. Considering the scarcity of existing research in the domain of 3D face anonymization in point clouds, this study holds significant relevance, particularly in response to increasing privacy concerns and the growing collection of 3D facial data, driven in part by developments in autonomous vehicles.

In response to this challenge, this research demonstrates the feasibility of extending concepts from the well-established realm of 2D face anonymization into the 3D domain. This conceptual shift led to the creation of six novel anonymization techniques. These techniques are rooted in 2D solutions but have been adapted to the higher-dimensional space by leveraging algorithms from various fields. In addition, these techniques have never been extended to the context of anonymization, and this thesis innovates in this regard. The development of such solutions benefited from the iPhonePLYv3, a custom dataset created specifically for this research. This dataset includes nearly twice the number of identities compared to some existing datasets. The data collection process, conducted without the need for specialized expertise or expensive equipment, yields a relevant methodology that can be replicated by the scientific community.

The effectiveness of the anonymization techniques was assessed using a set of privacy and utility metrics on the iPhonePLYv3 dataset. Both types of metrics were integrated into a unified evaluation methodology that was proposed and implemented to address the challenge of moderating the adversarial relationship between privacy and utility. This methodology can be adapted for evaluating the specific requirements of other use cases and can offer guidance on configuring anonymization parameters to achieve an effective privacy-utility trade-off. In conjunction with the qualitative evaluation, the evaluation methodology identified *Merge2Faces* and *SmoothKNN* as the most effective techniques displaying a positive compromise between privacy protection and data utility. Furthermore, this status demonstrates the potential for developing other solutions based on the same principles employed by these algorithms in morphing and smoothing.

6.1 Practical Implications

The findings derived from the experiments conducted in this thesis on 3D face anonymization suggest that several anonymization techniques hold promise for application within practical contexts that demand robust privacy protection mechanisms. In the realm of autonomous driving applications, the integration of the proposed anonymization techniques is likely to become indispensable with the advent of higher-resolution LiDAR sensor technology. This integration would ensure the preservation of individuals' privacy. However, additional investigation is warranted to determine the extent to which it maintains the integrity of perception tasks executed on the collected data. Moreover, various other fields that leverage 3D mapping technology, including aerial inspection and general robotics, stand to benefit from these solutions. Particularly in the latter, where LiDAR plays a pivotal role as a perception component, these techniques could play a crucial role in enhancing privacy without compromising functionality.

6.2 Limitations and Future Research

While the present study has provided valuable insights into the realm of 3D face anonymization within point clouds, it is imperative to acknowledge certain limitations that have influenced the scope and findings of this research.

The efficacy of the proposed anonymization techniques was evaluated on a custom-made 3D facial dataset. Consequently, the results are limited to the dataset size and characteristics. For a more robust assessment, the extent to which these techniques generalize across diverse datasets with varying demographics, facial expressions, and environmental conditions, warrants further investigation. Besides, the evaluation methodology exerts significant influence over the outcomes. Eliminating the need for 2D projection and directly measuring the privacy and utility of anonymization within the 3D space holds promise, mitigating biases and errors originating from the projection. Comparing the results of such evaluation to those presented in this work is crucial for a comprehensive assessment of the efficacy of anonymization techniques.

These two primary limitations can serve as a starting point for future research. However, other significant aspects can also be explored. For instance, the utilization of deep learning and generative models may open up avenues for more sophisticated anonymization techniques. Moreover, exploring the potential adaptation of the techniques initially designed for anonymizing 3D facial data within point clouds to anonymize various other objects or scenes has the prospect of broadening the practical applicability of these methods.

Appendix A

iPhonePLYv3

A.1 Folder Structure

The directory tree representation of the dataset is depicted in Figure A.1. This directory includes the information of 201 subjects, identified by an id code from *id_000* to *id_200*, based on the scanning order of each.

The iPhonePLYv3 dataset directory encompasses two main branches: *PointClouds* and *Images*, containing 3D and 2D related information, respectively. The *PointClouds* branch has two sub-branches: *Raw*, containing raw 3D face models of each subject with separate files for geometry data, material color, textures, and surface properties, and *MeshLab*, containing point clouds obtained from the processed 3D facial models. On the other hand, the *Images* branch has a single sub-branch called *Gallery*, which contains high-resolution frontal images of each subject. Each file in both branches corresponds to a specific subject and is named according to its id number. The dataset comprises 1 005 samples, with five data files per subject. Specifically, three files store the raw 3D model, one file contains the point cloud data, and the other file stores the corresponding 2D image. The entire dataset requires approximately 2.1 GB of storage space.

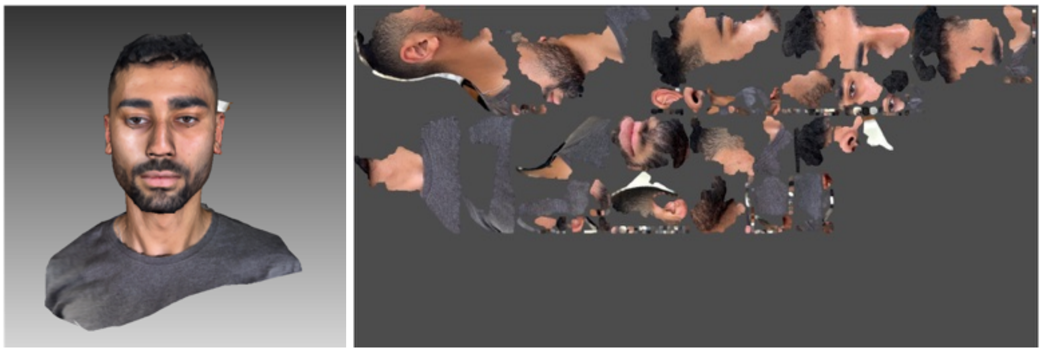
The dataset employs different file formats for storing the 3D face models and point clouds. The 3D face models are saved in the *OBJ*, *MTL*, and *JPG* formats. The *OBJ* format represents the surface geometry of the 3D model using a polygon mesh, while the *MTL* file type contains descriptions of surface appearance properties to be applied to the facets. Additionally, the textures used in the model are stored in *JPG* files. An example of the texture file for a subject within the dataset can be seen in Figure A.2.

```
iPhonePLYv3
│
├── PointClouds/ ............................3D information of the
│   │                                         subjects.
│   ├── Raw/
│   │   ├── id_000/
│   │   │   ├── id_000.jpg
│   │   │   ├── id_000.mtl
│   │   │   └── id_000.obj
│   │   ├── id_001/
│   │   │   └── ⋮
│   │   └── ⋮
│   └── MeshLab/
│       ├── id_000.ply
│       ├── id_001.ply
│       └── ⋮
└── Images/
    └── Gallery/ ............................2D information of the
        │                                     subjects.
        ├── id_000.png
        ├── id_001.png
        └── ⋮
```

FIGURE A.1: iPhonePLYv3 directory tree.

The point clouds are stored in the Polygon File Format (PLY) format, a versatile file format used for describing objects as polygonal models. It supports various properties for storage, such as color, transparency, surface normals, texture coordinates, and more.

The 2D face images are stored in the Portable Network Graphic (PNG) file format, capable of displaying high-quality digital images.



(A) 3D model. (B) Texture image.

FIGURE A.2: The *JPG* file contains texture information for loading the raw 3D model from the *OBJ* file, along with material data stored in the *MTL* file.

A.2 Construction Details

The iPhonePLYv3 utilizes the Scaniverse mobile application, incorporating photogrammetry and leveraging the iPhone 13 Pro LiDAR sensor for enhanced precision. Photogrammetry is a method that reconstructs detailed 3D models of objects, scenes, or landscapes using multiple photographs from different angles. By analyzing overlapping images, the technique applies triangulation[1] to derive precise measurements and depth information, ultimately reconstructing the 3D structure of the subject.

Figure A.3 presents the main steps in creating the iPhonePLYv3, which are further elaborated in detail later. The image furthest to the left corresponds to a raw 3D face scan, which is a mesh. The mesh is then registered to a specific location depicted in the subsequent image. In the following image, mesh components are illustrated, with points represented in yellow, edges depicted in gray, and the facets with the corresponding color. Subsequently, the resulting mapping of texture colors to the mesh's vertices and edges is depicted in the next image with the facets removed. Finally, the vertices are filtered, originating the final point cloud.

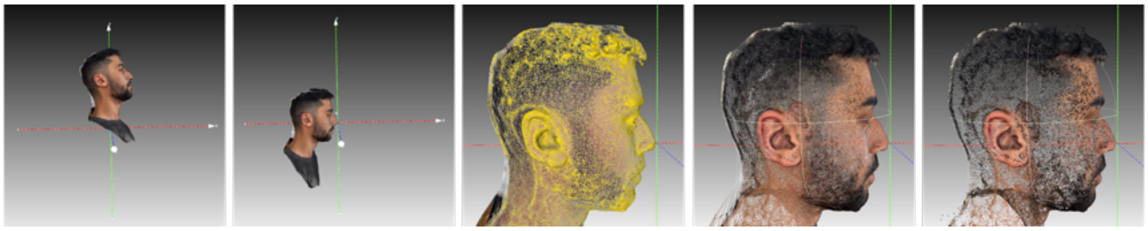


FIGURE A.3: Illustration of some steps of the employed processing for the dataset construction.

[1] A fundamental principle of the technique that identifies common points on the overlap images and calculates their relative positions through intersecting converging lines in space.

The iPhonePLYv3 construction methodology was initiated with the Scaniverse mobile application. First, the *NEW SCAN* button was pressed, followed by selecting *Small Object* and setting the range to *0.3 M*. The iPhone was moved slowly and smoothly around the subject's face, covering different angles from the right to the left side. The processing mode was set to *Detail* to capture surface textures accurately. The obtained 3D model was exported by selecting *Share > Export Model* and choosing the *OBJ* format, which resulted in a zip file containing *JPG*, *MTL*, and *OBJ* files with mesh information. The files were stored within the *Raw* sub-branch.

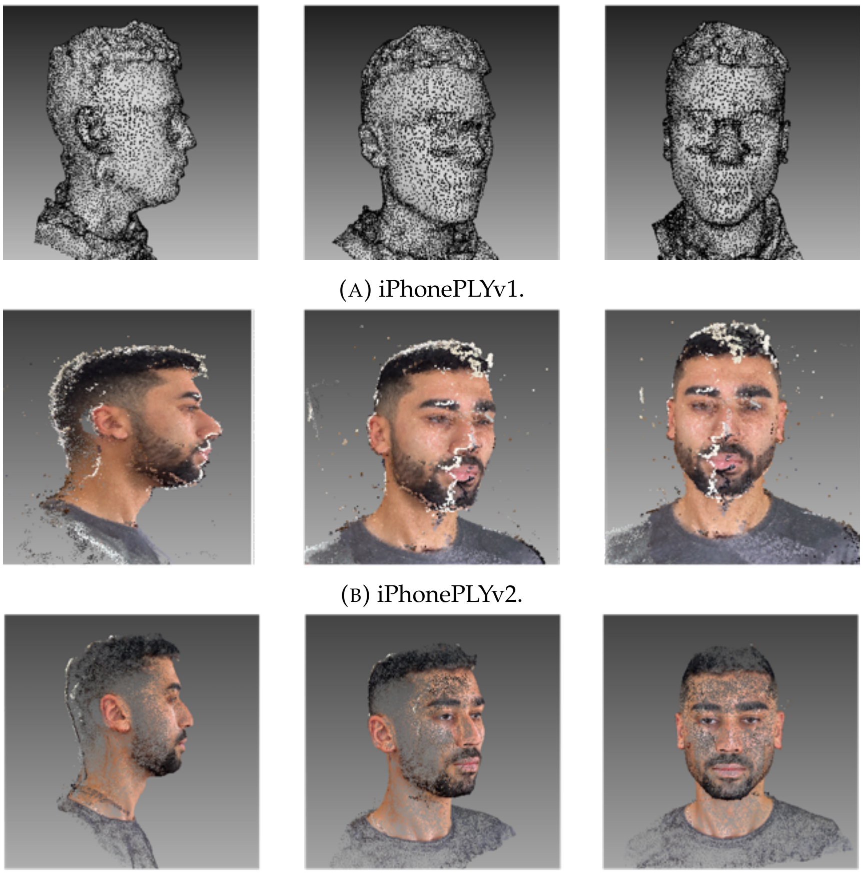
In MeshLab, the 3D face model was opened as a new project. The *Draw XYZ axis in world coordinates* and *Points* utilities were used to display the XYZ axis coordinates and the mesh's vertices. The mesh was manually aligned using the *Manipulator*, and transformations were applied and saved using *Filters > Normals, Curvatures, and Orientation > Matrix: Freeze Current Matrix*. RGB color was assigned to each vertex with *Filters > Color Creation and Processing > Transfer Color: Texture to Vertex*. All faces and edges were deleted using *Filters > Selection > Delete ALL Faces*, and the altered mesh was saved as a *PLY* file through *File > Export Mesh As...*, selecting the file name and directory for saving. The result was a point cloud obtained from the original mesh, stored within the *MeshLab* sub-branch.

A.3 iPhonePLYv1 and iPhonePLYv2

The iPhonePLYv3 dataset represents the third iteration of its kind, resulting from a sequence of iterative developments guided by strategic decisions. Figure A.4 illustrates the raw point clouds obtained from the three dataset iterations.

The initial version, named iPhonePLYv1, was captured using the Polycam application. However, due to export limitations regarding the payment of fees, the 3D model dataset was restricted to the *GLTF* file format, resulting in low point density, no color information, and overall low data quality.

Seeking higher-quality and budget-friendly options led to the discovery of Scaniverse for the second iteration, iPhonePLYv2. For its creation, the 3D face models were directly exported as *PLY* files, bypassing the MeshLab stage of the current version. Although a more straightforward approach, many subjects' point clouds exhibited noise and visual blurring, despite preprocessing attempts to alleviate these effects. Thus, while efforts



(A) iPhonePLYv1.

(B) iPhonePLYv2.

(C) iPhonePLYv3.

FIGURE A.4: Raw point clouds taken from the three dataset iterations.

were made to preserve color and enhance point density, the results still lacked the desired quality.

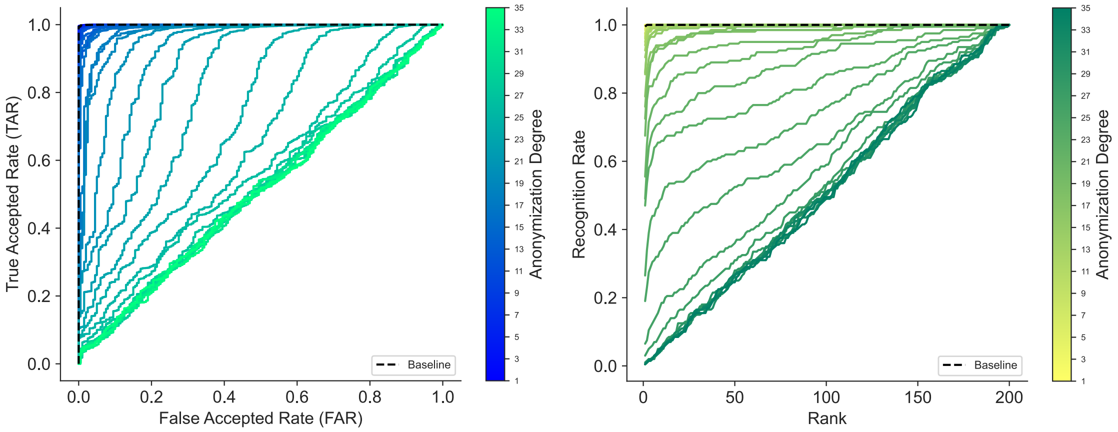
In the latest version, iPhonePLYv3, these limitations were addressed by incorporating color, achieving commendable point density, albeit lower than version two, and significantly enhancing overall quality. The *OBJ* file exportation, and the additional processing steps were crucial to the satisfactory outcome.

This iterative development process demonstrates progressive dataset refinement, with each version improving critical aspects such as point density, color representation, and overall data quality.

Appendix B

Anonymization Results

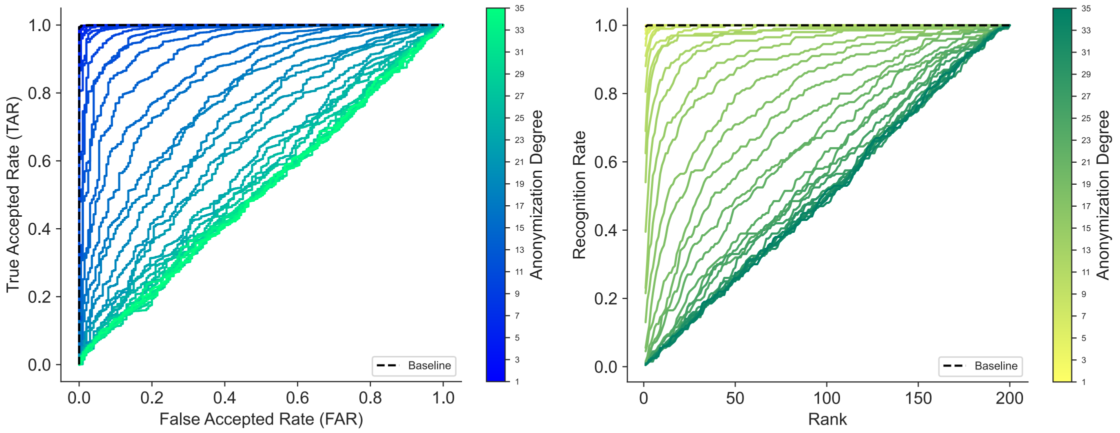
B.1 Privacy Metrics



(A) ROC.

(B) CMC.

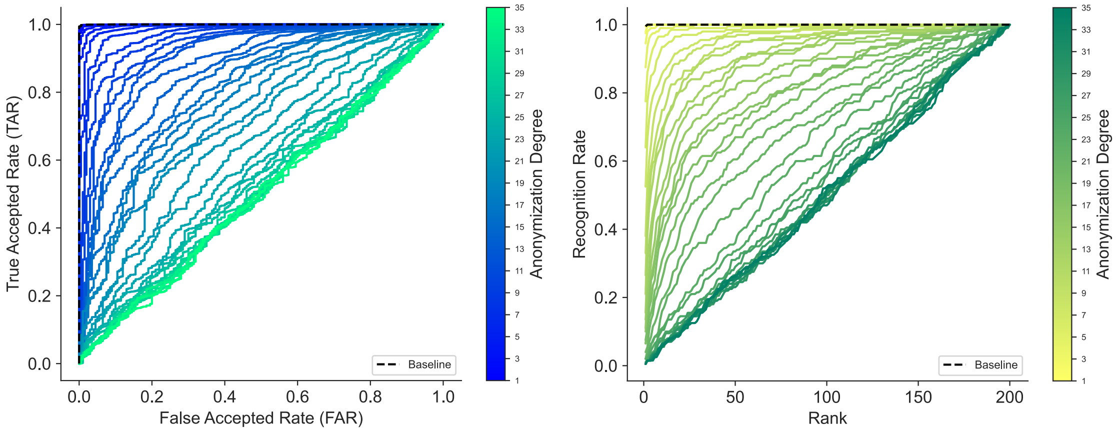
FIGURE B.1: Privacy metrics for *CentroidVoxel*.



(A) ROC.

(B) CMC.

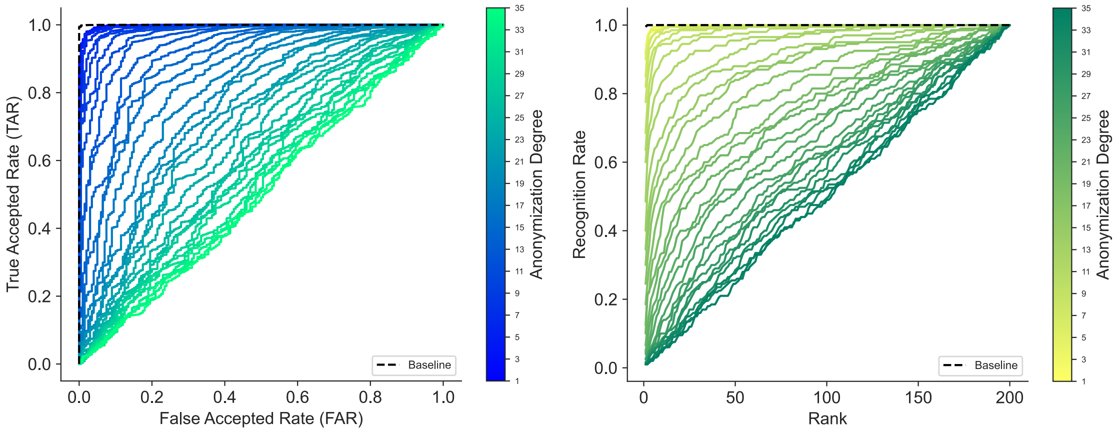
FIGURE B.2: Privacy metrics for *UniformNoise*.



(A) ROC.

(B) CMC.

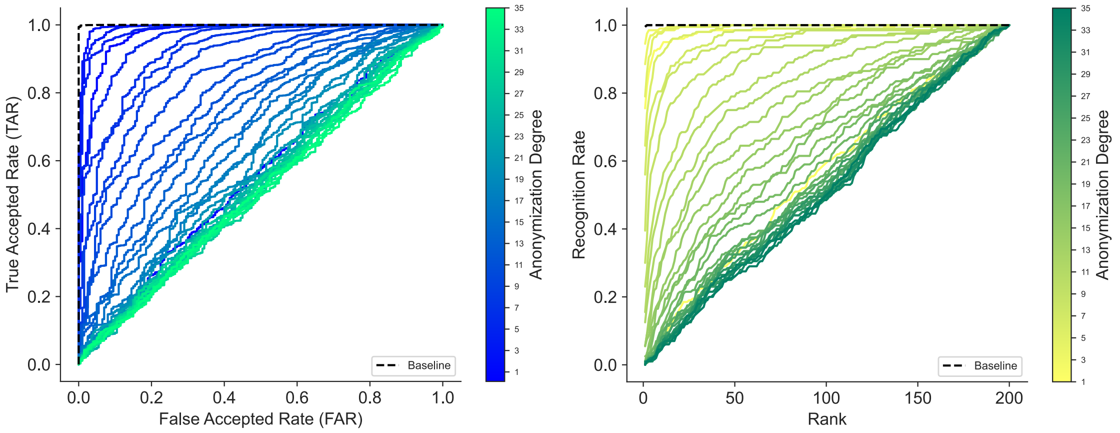
FIGURE B.3: Privacy metrics for *Tapering*.



(A) ROC.

(B) CMC.

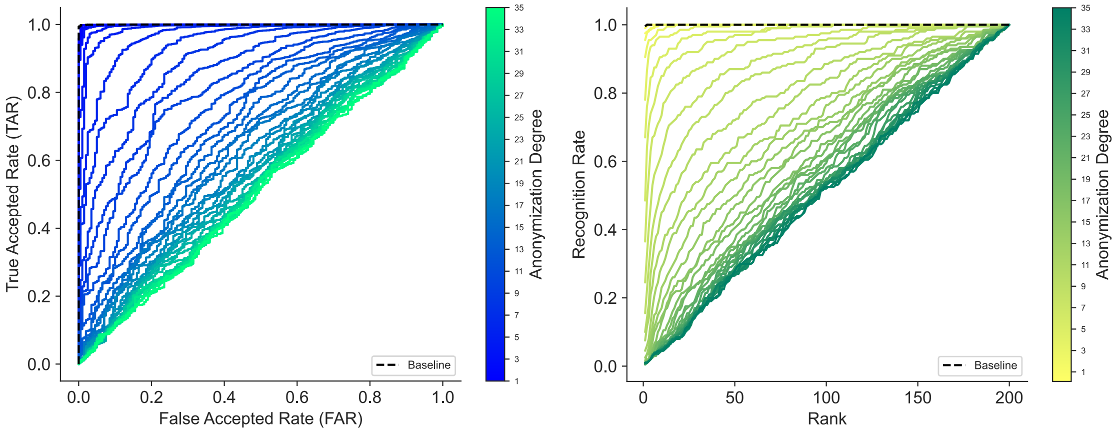
FIGURE B.4: Privacy metrics for *Merge2Faces*.



(A) ROC.

(B) CMC.

FIGURE B.5: Privacy metrics for *PMP*.

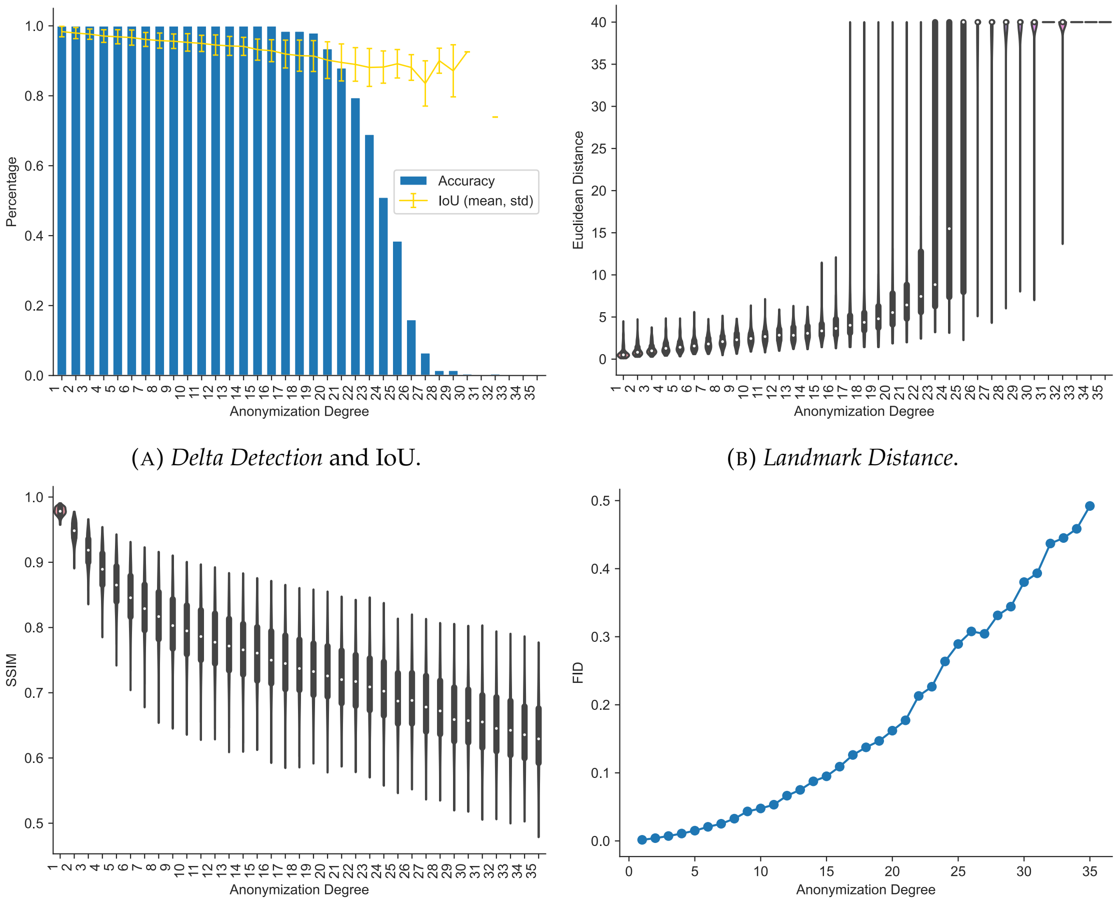


(A) ROC.

(B) CMC.

FIGURE B.6: Privacy metrics for *SmoothKNN*.

B.2 Utility Metrics



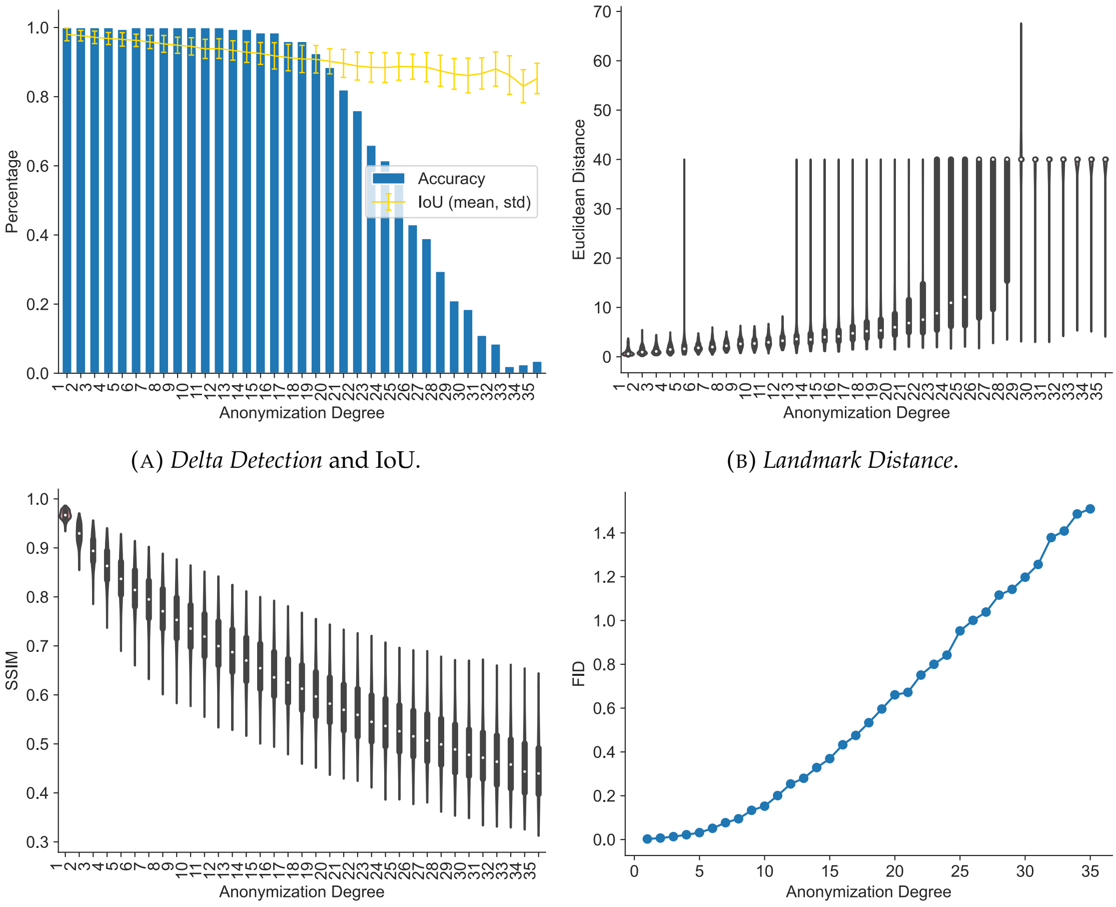
(A) *Delta Detection* and IoU.

(B) *Landmark Distance*.

(C) *SSIM*.

(D) *FID*.

Figure B.7: Utility metrics for *CentroidVoxel* anonymization.



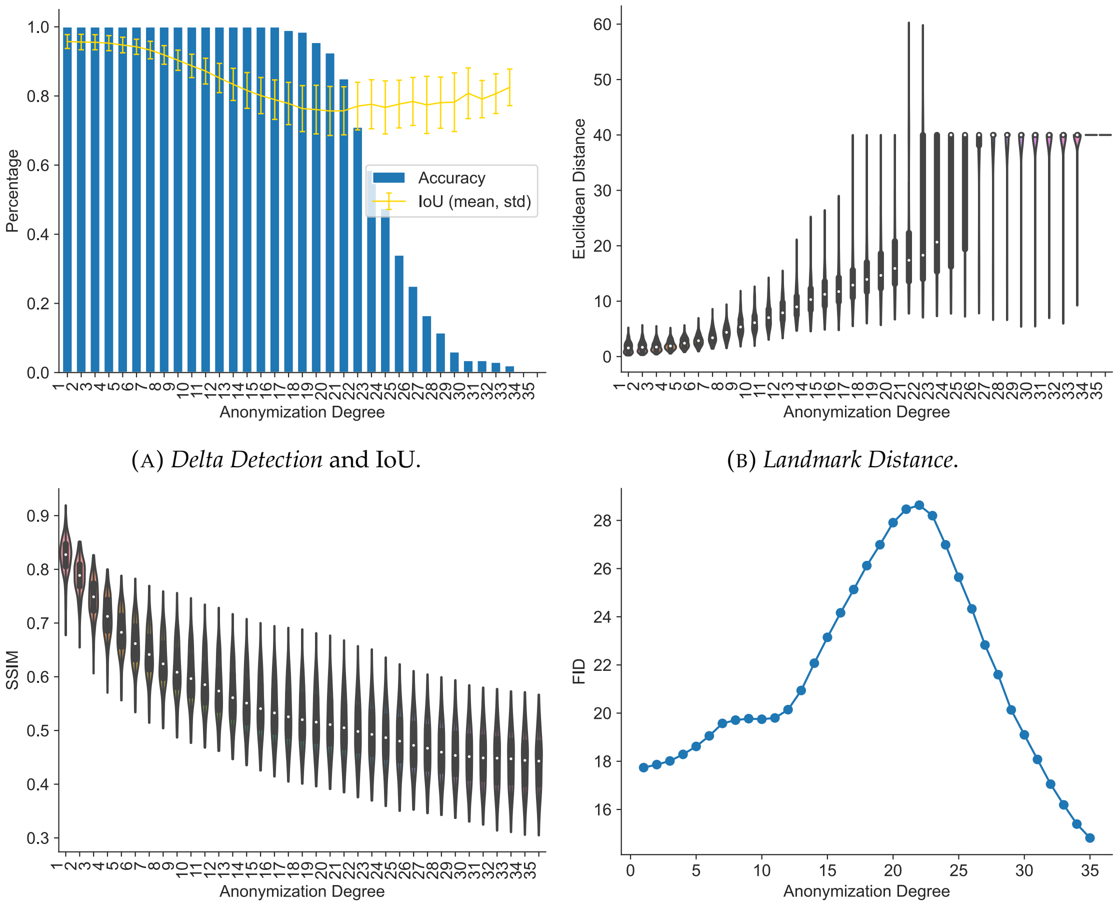
(A) *Delta Detection* and IoU.

(B) *Landmark Distance*.

(C) *SSIM*.

(D) *FID*.

FIGURE B.8: Utility metrics for *UniformNoise* anonymization.



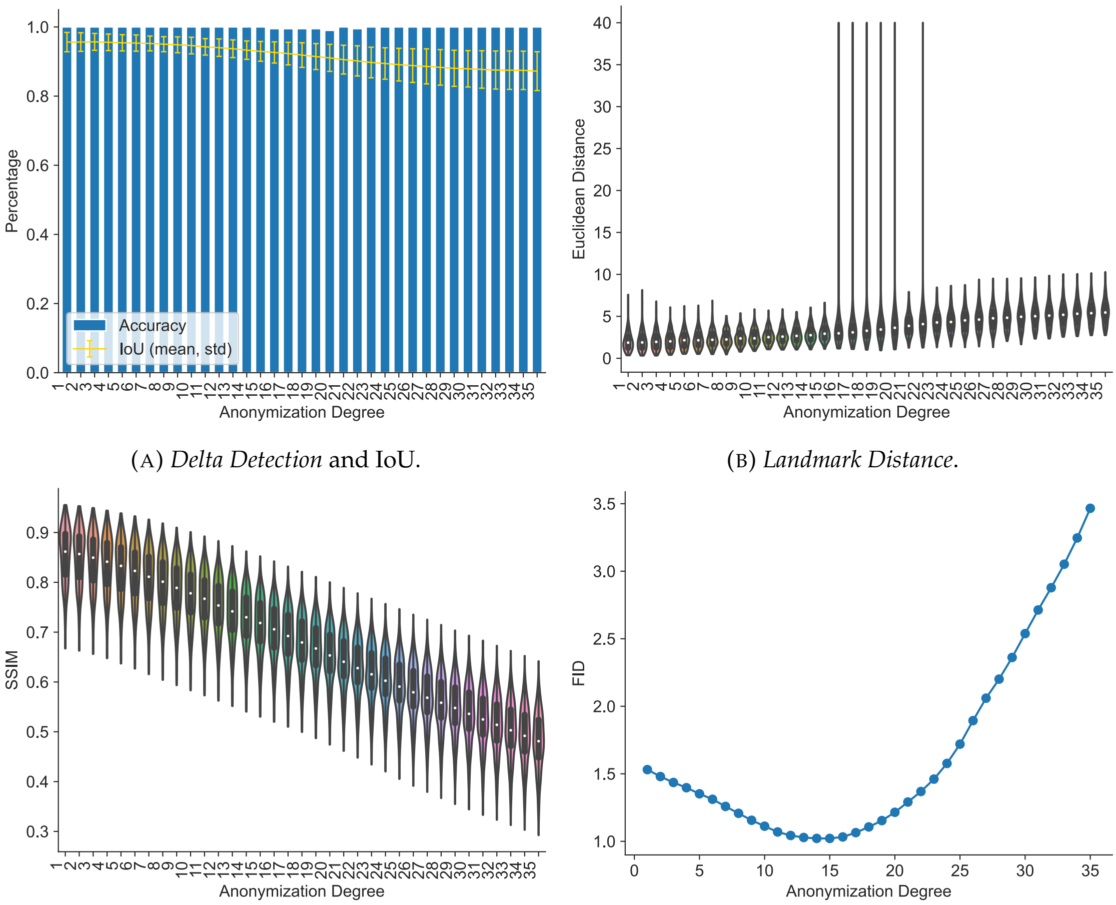
(A) *Delta Detection* and IoU.

(B) *Landmark Distance*.

(C) *SSIM*.

(D) *FID*.

FIGURE B.9: Utility metrics for *Tapering* anonymization.



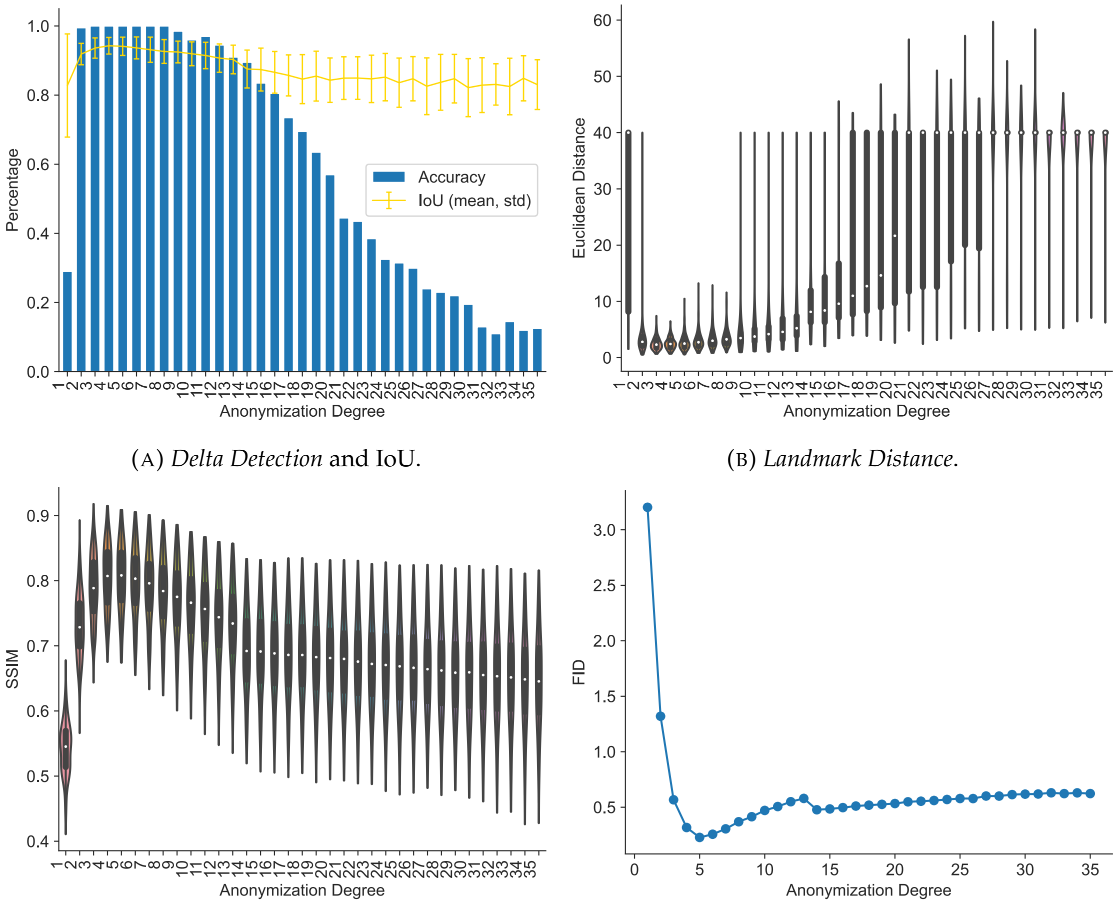
(A) *Delta Detection* and IoU.

(B) *Landmark Distance*.

(C) *SSIM*.

(D) *FID*.

FIGURE B.10: Utility metrics for *Merge2Faces* anonymization.



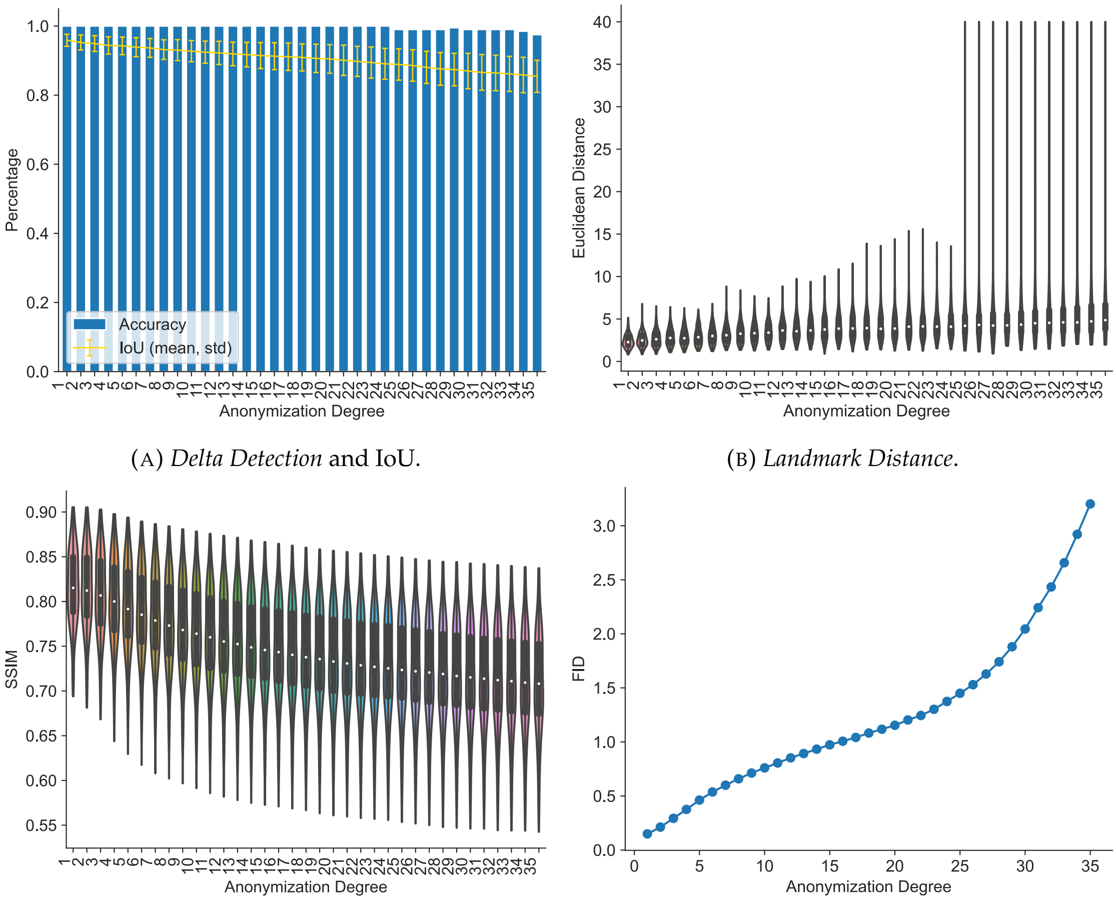
(A) *Delta Detection* and IoU.

(B) *Landmark Distance*.

(C) *SSIM*.

(D) *FID*.

FIGURE B.11: Utility metrics for *PMP* anonymization.



(A) *Delta Detection* and IoU.

(B) *Landmark Distance*.

(C) *SSIM*.

(D) *FID*.

FIGURE B.12: Utility metrics for *SmoothKNN* anonymization.

Bibliography

[1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[2] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1*. Springer, 2020, pp. 128–144.

[3] S. Minaee, P. Luo, Z. Lin, and K. Bowyer, "Going deeper into face detection: A survey," 2021.

[4] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.

[5] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, pp. 685–695, 2021.

[6] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[8] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.-C. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, "Deep learning for understanding faces: Machines may

be just as good, or better, than humans," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 66–83, 2018.

[9] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2017, pp. 650–657.

[10] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[11] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014.

[12] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, Oct. 2008.

[13] Y. Zhou, D. Liu, and T. Huang, "Survey of face detection on low-quality images," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 769–773.

[14] T. Shimizu, K. Koide, S. Oishi, M. Yokozuka, A. Banno, and M. Shino, "Sensor-independent pedestrian detection for personal mobility vehicles in walking space using dataset generated by simulation," in *25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 1788–1795.

[15] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2018, pp. 471–478.

[16] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[17] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," 2010.

[18] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," *Artificial Intelligence Review*, vol. 52, pp. 927–948, 2019.

[19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2001, p. 511–518.

[20] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, 2005, pp. 886–893.

[21] S. Zafeiriou, C. Zhang, and Z. Zhang, “A survey on face detection in the wild: Past, present and future,” *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, 2015.

[22] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer, 2006, pp. 404–417.

[24] P. Felzenszwalb, D. McAllester, and D. Ramanan, “A discriminatively trained, multiscale, deformable part model,” in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[25] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, “Cascade object detection with deformable part models,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2241–2248.

[26] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[27] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Li, “S^3FD: Single shot scale-invariant face detector,” *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 192–201, 2017.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[29] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “RetinaFace: Single-shot multi-level face localisation in the wild,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5202–5211.

[30] P. Soviany and R. T. Ionescu, "Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction," in *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 2018, pp. 209–214.

[31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[32] X. Lu, Q. Li, B. Li, and J. Yan, "MimicDet: Bridging the gap between one-stage and two-stage object detection," in *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV*. Berlin, Heidelberg: Springer-Verlag, 2020, p. 541–557.

[33] A. Lohia, K. D. Kadam, R. R. Joshi, and A. M. Bongale, "Bibliometric analysis of one-stage and two-stage object detection," *Libr. Philos. Pract*, vol. 4910, 2021.

[34] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data," *Remote Sensing*, vol. 13, no. 1, p. 89, 2020.

[35] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[36] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[37] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[38] Y. Li, B. Sun, T. Wu, and Y. Wang, "Face detection with end-to-end integration of a convnet and a 3D model," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 420–436.

[39] S. Wan, Z. Chen, T. Zhang, B. Zhang, and K.-k. Wong, "Bootstrapping face detection with hard negative examples," *arXiv:1608.02236*, 2016.

[40] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "CMS-RCNN: contextual multi-scale region-based cnn for unconstrained face detection," *Deep learning for biometrics*, pp. 57–79, 2017.

[41] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[42] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, p. 103514, 2022.

[43] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3296–3297.

[44] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "FaceBoxes: A CPU real-time face detector with high accuracy," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 1–9.

[45] S. Yang, Y. Xiong, C. C. Loy, and X. Tang, "Face detection through scale-friendly deep convolutional networks," *arXiv:1706.02863*, 2017.

[46] M. Najibi, P. Samangouei, R. Chellappa, and L. Davis, "SSH: Single stage headless face detector," *arXiv:1708.03979*, 2017.

[47] Y. Zhu, H. Cai, S. Zhang, C. Wang, and Y. Xiong, "TinaFace: Strong but simple baseline for face detection," *arXiv:2011.13183*, 2021.

[48] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern recognition*, vol. 39, no. 3, pp. 444–455, 2006.

[49] A. Mian, M. Bennamoun, and R. Owens, "Automatic 3D face detection, normalization and recognition," in *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*. IEEE, 2006, pp. 735–742.

[50] P. Nair and A. Cavallaro, "3-D face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on multimedia*, vol. 11, no. 4, pp. 611–623, 2009.

[51] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2879–2886.

[52] J. Yan, X. Zhang, Z. Lei, and S. Z. Li, "Face detection by structural models," *Image and Vision Computing*, vol. 32, no. 10, pp. 790–799, 2014.

[53] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5525–5533.

[54] V. Jain and E. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, Tech. Rep. UM-CS-2010-009, 2010.

[55] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Fine-grained evaluation on face detection in the wild," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, 2015, pp. 1–7.

[56] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.

[57] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, pp. 303–338, 6 2010.

[58] Y. Xiong, K. Zhu, D. Lin, and X. Tang, "Recognize complex events from static images by fusing deep channels," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1600–1609.

[59] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 391–405.

[60] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[61] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 618–11 628.

[62] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo Open Dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.

[63] Z. Guo, Y.-N. Zhang, Y. Xia, Z.-G. Lin, Y.-Y. Fan, and D. D. Feng, "Multi-pose 3D face recognition based on 2D sparse representation," *Journal of Visual Communication and Image Representation*, vol. 24, no. 2, pp. 117–126, 2013, sparse Representations for Image and Video Analysis.

[64] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.

[65] Y. Wang, X. Ji, Z. Zhou, H. Wang, and Z. Li, "Detecting faces using region-based fully convolutional networks," *arXiv:1709.05256*, 2017.

[66] C. Chi, S. Zhang, J. Xing, Z. Lei, S. Z. Li, and X. Zou, "Selective refinement network for high performance face detection," *arXiv:1809.02693*, 2018.

[67] C. Zhang, X. Xu, and D. Tu, "Face detection using improved Faster-RCNN," *arXiv:1802.02142*, 2018.

[68] J. Li, Y. Wang, C. Wang, Y. Tai, J. Qian, J. Yang, C. Wang, J. Li, and F. Huang, "DSFD: Dual shot face detector," *arXiv:1810.10220*, 2019.

[69] X. Tang, D. K. Du, Z. He, and J. Liu, "PyramidBox: A context-assisted single shot face detector," *arXiv:1803.07737*, 2018.

[70] F. Zhang, X. Fan, G. Ai, J. Song, Y. Qin, and J. Wu, "Accurate face detection for high performance," *arXiv:1905.01585*, 2019.

[71] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Springer, 2011.

[72] T. Agagu and B. Akinnuwesi, "Automated students' attendance taking in tertiary institution using hybridized facial recognition algorithm," *Journal of Computer Science and Its Application*, vol. 19, no. 2, pp. 1–13, 2012.

[73] H. Du, H. Shi, D. Zeng, X.-P. Zhang, and T. Mei, "The elements of end-to-end deep face recognition: A survey of recent advances," *arXiv:2009.13290*, 2021.

[74] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques," *journal of information processing systems*, vol. 5, no. 2, pp. 41–68, 2009.

[75] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[76] N. Wang, X. Gao, D. Tao, H. Yang, and X. Li, "Facial feature point detection: A comprehensive survey," *Neurocomputing*, vol. 275, pp. 50–65, 2018.

[77] X. Jin and X. Tan, "Face alignment in-the-wild: A survey," *Computer Vision and Image Understanding*, vol. 162, pp. 1–22, 2017.

[78] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electronics*, vol. 9, no. 8, p. 1188, 2020.

[79] M. Li, B. Huang, and G. Tian, "A comprehensive survey on 3D face recognition methods," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104669, 2022.

[80] S. Zhou and S. Xiao, "3D face recognition: a survey," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, pp. 1–27, 2018.

[81] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.

[82] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[83] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, 2020.

[84] W. Bledsoe, "Man-machine facial recognition: Report on a large-scale experiment. panoramic research (1966)," *Inc., Palo Alto, CA*, 1966.

[85] S. Karamizadeh, S. Abdullah, and M. Zamani, "An overview of holistic face recognition," *International Journal of Research in Computer and Communication Technology*, vol. 2, pp. 738–741, 09 2013.

[86] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[87] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[88] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.

[89] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 40–42, 2002.

[90] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[91] S. R. Arashloo and J. Kittler, "Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2100–2109, 2014.

[92] D. S. Trigueros, L. Meng, and M. Hartnett, "Face recognition: From traditional to deep learning methods," *arXiv:1811.00116*, 2018.

[93] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[94] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram fourier features," in *Image Analysis: 16th Scandinavian Conference, SCIA 2009, Oslo, Norway, June 15-18, 2009. Proceedings 16*. Springer, 2009, pp. 61–70.

[95] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE transactions on image processing*, vol. 19, no. 2, pp. 533–544, 2009.

[96] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2. IEEE, 1999, pp. 1150–1157.

[97] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG-EBGM," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.

[98] P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, and G. Aachen, "SURF-Face: Face recognition under viewpoint consistency constraints." in *BMVC*, 2009, pp. 1–11.

[99] T. M. Kodinariya, "Hybrid approach to face recognition system using principle component and independent component with score based fusion process," *arXiv:1401.0395*, 2014.

[100] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891–1898.

[101] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 248–255.

[102] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2015.

[103] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[104] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv:1512.03385*, 2015.

[105] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[106] X. Wang, S. Wang, C. Chi, S. Zhang, and T. Mei, "Loss function search for face recognition," in *International Conference on Machine Learning*. PMLR, 2020, pp. 10 029–10 038.

[107] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*. Springer, 2016, pp. 499–515.

[108] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1735–1742.

[109] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[110] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," *arXiv:1612.02295*, 2016.

[111] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6738–6746.

[112] J. Deng and S. Zafeririou, "ArcFace for disguised face recognition," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 485–493.

[113] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "ElasticFace: Elastic margin loss for deep face recognition," *arXiv:2109.09416*, 2022.

[114] P. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[115] X.-L. LI and F.-P. DA, "A rapid method for 3D face recognition based on rejection algorithm," *Acta Automatica Sinica*, vol. 36, pp. 153–158, 01 2010.

[116] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 381–397, 2013.

[117] M. Emambakhsh and A. Evans, "Nasal patches and curves for expression-robust 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 995–1007, 2017.

[118] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava, "Pose and expression-invariant 3D face recognition using elastic radial curves," in *British machine vision conference*, 2010, pp. 1–11.

[119] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," *Pattern Recognition*, vol. 47, no. 2, pp. 509–524, 2014.

[120] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3d classification and segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 77–85.

[121] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," *arXiv:1706.02413*, 2017.

[122] A. R. Bhople, A. M. Shrivastava, and S. Prakash, "Point cloud based deep convolutional neural network for 3D face recognition," *Multimedia Tools Appl.*, vol. 80, no. 20, p. 30237–30259, aug 2021.

[123] S. Zulqarnain Gilani and A. Mian, "Learning from millions of 3D scans for large-scale 3d face recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1896–1905.

[124] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association, 2015.

[125] P. Grother and M. Ngan, "Face recognition vendor test (FRVT): Performance of face identification algorithms. NIST interagency report 8009," *National Institute of Standards and Technology, Tech. Rep*, 2013.

[126] M. Günther, P. Hu, C. Herrmann, C. H. Chan, M. Jiang, S. Yang, A. R. Dhamija, D. Ramanan, J. Beyerer, J. Kittler, M. A. Jazaery, M. I. Nouyed, G. Guo, C. Stankiewicz, and T. E. Boult, "Unconstrained face detection and open-set face

recognition challenge," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 697–706.

[127] S. Marcel, "Beat–biometrics evaluation and testing," *Biometric technology today*, vol. 2013, no. 1, pp. 5–7, 2013.

[128] C. Geng, S.-J. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3614–3631, 2021.

[129] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 87–102.

[130] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 67–74.

[131] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother, "IARPA janus benchmark-B face dataset," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 592–600.

[132] E. Zhou, Z. Cao, and Q. Yin, "Naive-deep face recognition: Touching the limit of lfw benchmark or not?" *arXiv preprint arXiv:1501.04690*, 2015.

[133] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA janus benchmark A," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1931–1939.

[134] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, and P. Grother, "IARPA janus benchmark - C: Face dataset and protocol," in *2018 International Conference on Biometrics (ICB)*, 2018, pp. 158–165.

[135] A. Moreno, "GavabDB: A 3D face database," in *Proc. 2nd COST275 Workshop on Biometrics on the Internet, 2004*, 2004, pp. 75–80.

[136] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *7th international conference on automatic face and gesture recognition (FGR06)*. IEEE, 2006, pp. 211–216.

[137] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *Biometrics and Identity Management*, B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 47–56.

[138] Y. Baocai, S. Yanfeng, W. Chengzhang, and G. Yun, "BJUT-3D large scale 3D face database and information processing," *Journal of Computer Research and Development*, vol. 6, no. 020, p. 4, 2009.

[139] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "FaceWarehouse: A 3D facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2013.

[140] H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, and X. Cao, "FaceScape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, 2020, pp. 601–610.

[141] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," *arXiv:1801.09414*, 2018.

[142] B.-N. Kang, Y. Kim, and D. Kim, "Pairwise relational networks for face recognition," *arXiv:1808.04976*, 2018.

[143] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting ultimate accuracy: Face recognition via deep embedding," *arXiv:1506.07310*, 2015.

[144] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," *arXiv:1703.09507*, 2017.

[145] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4685–4694.

[146] H. Li, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 128–142, 2015.

[147] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala, "Matching 3D face scans using interest points and local histogram descriptors," *Computers & Graphics*, vol. 37, no. 5, pp. 509–525, 2013.

[148] S. Soltanpour and Q. M. J. Wu, "Weighted extreme sparse classifier and local derivative pattern for 3D face recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3020–3033, 2019.

[149] M. Maximov, I. Elezi, and L. Leal-Taixé, "CIAGAN: Conditional identity anonymization generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5447–5456.

[150] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, and V. Štruc, "Privacy–enhancing face biometrics: A comprehensive survey," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4147–4183, 2021.

[151] L. Rakhmawati *et al.*, "Image privacy protection techniques: A survey," in *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018, pp. 0076–0080.

[152] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, vol. 47, pp. 131–151, 2016.

[153] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, "Generative adversarial networks: A survey toward private and secure applications," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–38, 2021.

[154] K. Chinomi, N. Nitta, Y. Ito, and N. Babaguchi, "PriSurv: Privacy protected video surveillance system using adaptive visual abstraction," in *Advances in Multimedia Modeling: 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008. Proceedings 14*. Springer, 2008, pp. 144–154.

[155] D. Chen, Y. Chang, R. Yan, and J. Yang, "Protecting personal identification in video," *Protecting Privacy in Video Surveillance*, pp. 115–128, 2009.

[156] O. Sarwar, A. Cavallaro, and B. Rinner, "Temporally smooth privacy-protected airborne videos," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 6728–6733.

[157] L. Fan, "Image pixelization with differential privacy," in *Data and Applications Security and Privacy XXXII: 32nd Annual IFIP WG 11.3 Conference, DBSec 2018, Bergamo, Italy, July 16–18, 2018, Proceedings 32*. Springer, 2018, pp. 148–162.

[158] P. Korshunov and T. Ebrahimi, "Using warping for privacy protection in video surveillance," in *2013 18th International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1–6.

[159] P. Chriskos, J. Munro, V. Mygdalis, and I. Pitas, "Face detection hindering," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 403–407.

[160] R. Gross, L. Sweeney, J. Cohn, F. de la Torre, and S. Baker, *Face De-identification*. London: Springer London, 2009, pp. 129–146.

[161] L. Sweeney, "K-anonymity: A model for protecting privacy," *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 10, no. 05, pp. 557–570, 2002.

[162] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[163] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *Privacy Enhancing Technologies*, G. Danezis and D. Martin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 227–242.

[164] R. Gross, L. Sweeney, F. De la Torre, and S. Baker, "Model-based face de-identification," in *2006 Conference on computer vision and pattern recognition workshop (CVPRW'06)*. IEEE, 2006, pp. 161–161.

[165] L. Ruthotto and E. Haber, "An introduction to deep generative modeling," *GAMM-Mitteilungen*, vol. 44, no. 2, p. e202100008, 2021.

[166] C. Yinka-Banjo and O.-A. Ugot, "A review of generative adversarial networks and its application in cybersecurity," *Artificial Intelligence Review*, vol. 53, pp. 1721–1736, 2020.

[167] H. Hukkelås, R. Mester, and F. Lindseth, "DeepPrivacy: A generative adversarial network for face anonymization," in *International symposium on visual computing*. Springer, 2019, pp. 565–578.

[168] J. Lin, Y. Li, and G. Yang, "FPGAN: Face de-identification method with generative adversarial networks for social robots," *Neural Networks*, vol. 133, pp. 132–147, 2021.

[169] P. Nousi, S. Papadopoulos, A. Tefas, and I. Pitas, "Deep autoencoders for attribute preserving face de-identification," *Signal Processing: Image Communication*, vol. 81, p. 115699, 2020.

[170] Y. Qiu, Z. Niu, Q. Tian, and B. Song, "Privacy preserving facial image processing method using variational autoencoder," in *International Conference on Big Data and Security*. Springer, 2021, pp. 3–17.

[171] P. Rustici, "Anonymization of 3D face models for GDPR compliant outsourcing to 3rd party companies," 2020, Delft University of Technology.

[172] J. M. Singh and R. Ramachandra, "3D face morphing attacks: Generation, vulnerability and detection," *arXiv:2201.03454*, 2022.

[173] N. Schimke, M. Kuehler, and J. Hale, "Preserving privacy in structural neuroimages," in *Data and Applications Security and Privacy XXV: 25th Annual IFIP WG 11.3 Conference, DBSec 2011, Richmond, VA, USA, July 11-13, 2011. Proceedings 25*. Springer, 2011, pp. 301–308.

[174] Y. U. Jeong, S. Yoo, Y.-H. Kim, and W. H. Shim, "De-identification of facial features in magnetic resonance images: Software development using deep learning technology," *J Med Internet Res*, vol. 22, no. 12, p. e22739, Dec 2020.

[175] P. Churi, D. A. Pawar, and A. Moreno Guerrero, "A comprehensive survey on data utility and privacy: Taking indian healthcare system as a potential case study," *Inventions*, vol. 6, pp. 1–30, 06 2021.

[176] L. Rakhmawati, Wirawan, and Suwadi, "Image privacy protection techniques: A survey," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 0076–0080.

[177] Z. Kuang, H. Liu, J. Yu, A. Tian, L. Wang, J. Fan, and N. Babaguchi, "Effective de-identification generative adversarial network for face anonymization," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3182–3191.

[178] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi, "Framework for objective evaluation of privacy filters," in *Applications of Digital Image Processing XXXVI*, vol. 8856. SPIE, 2013, pp. 265–276.

[179] P. Terhörst, M. Huber, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised enhancement of soft-biometric privacy with negative face recognition," *arXiv:2002.09181*, 2020.

[180] J. Todt, S. Hanisch, and T. Strufe, "Fantômas: Evaluating reversibility of face anonymizations using a general deep learning attacker," *arXiv preprint arXiv:2210.10651*, 2022.

[181] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[182] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," *arXiv:1706.08500*, 2018.

[183] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[184] M. Fréchet, "Sur la distance de deux lois de probabilité," in *Annales de l'ISUP*, vol. 6, no. 3, 1957, pp. 183–198.

[185] L. Zhai, Q. Guo, X. Xie, L. Ma, Y. E. Wang, and Y. Liu, "A3GAN: Attribute-aware anonymization networks for face de-identification," in *Proceedings of the 30th ACM International Conference on Multimedia*, ser. MM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 5303–5313.

[186] M. Sharif, F. Naz, Y. Mussarat, M. Shahid, and A. Rehman, "Face recognition: A survey," *Journal of Engineering Science and Technology Review*, vol. 10, pp. 166–177, 06 2017.

[187] W. T. Hrinivich, T. Wang, and C. Wang, "Interpretable and explainable machine learning models in oncology," *Frontiers in Oncology*, vol. 13, p. 1184428, 2023.

[188] G. P. Kusuma and C.-S. Chua, "Pca-based image recombination for multimodal 2D+ 3D face recognition," *Image and Vision Computing*, vol. 29, no. 5, pp. 306–316, 2011.

[189] J. Li, J. Zhou, Y. Xiong, X. Chen, and C. Chakrabarti, "An adjustable farthest point sampling method for approximately-sorted point cloud data," in *2022 IEEE Workshop on Signal Processing Systems (SiPS)*, 2022, pp. 1–6.

[190] X.-F. Han, J. S. Jin, M.-J. Wang, W. Jiang, L. Gao, and L. Xiao, "A review of algorithms for filtering the 3d point cloud," *Signal Processing: Image Communication*, vol. 57, pp. 103–112, 2017.

[191] A. H. Barr, "Global and local deformations of solid primitives," in *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '84. New York, NY, USA: Association for Computing Machinery, 1984, p. 21–30.

[192] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in Computer Vision*, M. A. Fischler and O. Firschein, Eds. San Francisco (CA): Morgan Kaufmann, 1987, pp. 726–740.

[193] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 3212–3217.

[194] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.

[195] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proceedings third international conference on 3-D digital imaging and modeling*. IEEE, 2001, pp. 145–152.

[196] N. Akkiraju, H. Edelsbrunner, M. Facello, P. Fu, E. Mucke, and C. Varela, "Alpha shapes: Definition and software," in *Proceedings of the 1st international computational geometry software workshop*, vol. 63, no. 66, 1995.

[197] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Transactions On Graphics (TOG)*, vol. 13, no. 1, pp. 43–72, 1994.

[198] M.-H. Le and N. Carlsson, "StyleID: Identity disentanglement for anonymizing faces," *arXiv:2212.13791*, 2022.

[199] J. Li, L. Han, R. Chen, H. Zhang, B. Han, L. Wang, and X. Cao, "Identity-preserving face anonymization via adaptively facial attributes obfuscation," in *Proceedings of the 29th ACM International Conference on Multimedia*, ser. MM '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 3891–3899.

[200] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *arXiv:2003.05991*, 2020.

[201] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, V. Scarano, R. D. Chiara, and U. Erra, Eds. The Eurographics Association, 2008.

[202] D. Rethage, J. Wald, J. Sturm, N. Navab, and F. Tombari, "Fully-convolutional point networks for large-scale point clouds," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 596–611.

[203] W. Abbasi, P. Mori, A. Saracino, and V. Frascolla, "Privacy vs accuracy trade-off in privacy aware face recognition in smart systems," in *2022 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2022, pp. 1–8.

[204] Q. Gu, L. Zhu, and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in *Computational Intelligence and Intelligent Systems: 4th International Symposium, ISICA 2009, Huangshi, China, October 23-25, 2009. Proceedings 4*. Springer, 2009, pp. 461–471.

[205] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "SSIM-optimal linear image restoration," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 765–768.

[206] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.

[207] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2020, pp. 23–27.

[208] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," *arXiv:1912.01703*, 2019.

[209] I. Jovančević, H.-H. Pham, J.-J. Orteu, R. Gilblas, J. Harvent, X. Maurice, and L. Brèthes, "3D point cloud analysis for detection and characterization of defects on airplane exterior surface," *Journal of Nondestructive Evaluation*, vol. 36, pp. 1–17, 2017.