2020

On Body Mass Index Analysis from Human Visual Appearance

Min Jiang
mijiang@mix.wvu.edu

On Body Mass Index Analysis from Human Visual Appearance

Min Jiang

Dissertation submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Electrical Engineering

Guodong Guo, Ph.D., Committee Chairperson
Matthew C. Valenti, Ph.D.
Xin Li, Ph.D.
Donald Adjeroh, Ph.D.
Hong-Jian Lai, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2020

Keywords: Body Mass Index Analysis, 2D and 3D Visual Appearance, Feature Extraction, Two-stage Learning, Label Assignment Matching

Abstract

On Body Mass Index Analysis from Human Visual Appearance

Min Jiang

In the past few decades, overweight and obesity are spreading widely like an epidemic. Generally, a person is considered overweight by body mass index (BMI). In addition to a body fat measurement, BMI is also a risk factor for many diseases, such as cardiovascular diseases, cancers and diabetes, etc. Therefore, BMI is important for personal health monitoring and medical research. Currently, BMI is measured in person with special devices. It is an urgent demand to explore conveniently preventive tools. This work investigates the feasibility of analyzing BMI from human visual appearances, including 2-dimensional (2D)/3-dimensional (3D) body and face data.

Motivated by health science studies which have shown that anthropometric measures, such as waist-hip ratio, waist circumference, etc., are indicators for obesity, we analyze body weight from frontal view human body images. A framework is developed for body weight analysis from body images, along with the computation methods of five anthropometric features for body weight characterization. Then, we study BMI estimation from the 3D data by measuring the correlation between the estimated body volume and BMIs, and develop an efficient BMI computation method which consists of body weight and height estimation from normally dressed people in 3D space.

We also intensively study BMI estimation from frontal view face images via two key aspects: facial representation extracting and BMI estimator learning. First, we investigate the visual BMI estimation problem from the aspect of the characteristics and performance of different facial representation extracting methods by three designed experiments. Then we study visual BMI estimation from facial images by a two-stage learning framework. BMI related facial features are learned in the first stage. To address the ambiguity of BMI labels, a label distribution based BMI estimator is proposed for the second stage. The experimental results show that this framework improves the performance step by step. Finally, to address the challenges caused by BMI data and labels, we integrate feature learning and estimator learning in one convolutional neural network (CNN). A label assignment matching scheme is proposed which successfully achieves an improvement in BMI estimation from face images.

Acknowledgments

My deepest gratitude goes to my committee chair and advisor Dr. Guodong Guo, who consistently and generously supported my research, gave me invaluable advice, encouraged and inspired me. This dissertation would not be possible without his enormous work and help. He brings me a spirit of exploring science to real-world applications. I am deeply impressed by his passionate, rigorous and diligent attitude toward both teaching and research, as well as his kindness and caring. It was a great pleasure to learn from his classes, to carry out research work with him, and to know him as a mentor and friend.

I would like to thank Dr. Matthew C. Valenti, Dr. Xin Li, Dr. Donald Adjeroh and Dr. Hong-Jian Lai for serving as my committee members. They spent precious time and effort on giving suggestions to my research and helping to improve and refine this dissertation.

I would also like to thank Dr. Natalia Schmid and Dr. Maura McLaughlin for advising and supporting me during the first three years of my Ph.D. study. Their patience, innovations, enthusiasm, and prestigious contributions to radio astronomy realm greatly impact me, and encourage me to explore in this area.

I am highly grateful to my colleagues Bingyi Cui, Zhicheng Cao, Qiangchang Wang and Mohammad Iqbal Nouye for their kind help and assistance to complete my research projects during these years.

Last but not least, I would like to thank my family and friends, especially my parents who gave me life, nurtured me, and unconditionally love and support me.

Contents

List of Figures

List of Tables

List of Acronyms

BMI	Body Mass Index
RGB-D	Red-Green-Blue and Depth
3D	Three-Dimensional
2D	Two-Dimensional
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
APE	Absolute Percentage Error
Std	Standard Deviation
PCC	Pearson's Correlation Coefficient
WTR	Ratio of Waist Width to Thigh Width
WHpR	Ratio of Waist Width to Hip Width
WHdR	Ratio of Waist Width to Head Width
HpHdR	Ratio of Hip Width to Head Width
CI	Confidence Interval
SVM	Support Vector Machine
SVR	Support Vector Regression
GPR	Gaussian Process Regression
LD	Label Distribution
CCA	Canonical Correlation Analysis
PLS	Partial Least Square
PCA	Principal Component Analysis
PIGF	Psychology Inspired Geometric Feature
PF	Pointer Feature
SP	Sample Pose
FIW-BMI	Face In Wild for BMI Analysis
SDG	Stochastic Gradient Decent

CNN	Convolutional Neural Network
FC	Fully Connected
KLD	Kullback-Leibler Divergence
GAN	Generative Adversarial Network

Chapter 1

Introduction

Biometrics [1] refers to metrics of human physiological and behavioral characteristics, such as the face, fingerprint, iris, DNA, voice, gait, gender, height, weight and age, etc. Researchers working on biometrics put a spotlight on the so-called soft biometrics [2, 3], including height, weight, age and hair color, etc. Different from hard biometrics (such as the face, fingerprint, iris and DNA, etc) which have some peculiar characteristics such as robustness and distinctiveness [4, 5], soft biometrics do not exhibit such characteristics. However, recently it has been demonstrated that soft biometrics are useful to ameliorate the quality of identification and recognition [6–8].

Among the soft biometric measures, body mass index (BMI) is a good indicator of health condition and a risk factor for many diseases such as diabetes, cancer and renal disease, etc [9–13]. BMI is calculated by an individual's height and weight. It is an important visual characteristic to describe a person, which is widely used for measuring adiposity, especially for the overweight issue [14–16]. Generally, BMI is measured in person with special devices. Thereby, automatically accessing BMI values from human visual appearance is a great benefit to health condition monitoring and researchers who are interested in studying obesity in large populations.

This dissertation is devoted to the soft biometrics using human visual appearance as the trait known as visual BMI estimation. The main cases of interest discussed in this work are BMI estimation from various types of visual appearances, such as 2-dimensional (2D) body images, 3-dimensional (3D) body reconstruction, and 2D face

images. Computational methods are proposed in turn for these cases. The main focus of this dissertation is to develop new computational approaches for BMI estimation.

1.1 Background and Motivation

Extreme overweight and obesity are spreading widely like an epidemic. The United States spends more than 300 billion each year to treat obesity, diabetes, and cardiovascular diseases [17]. Overweight has been identified as one of the main factors that generate those diseases. Generally, a person is considered overweight or obese by the BMI value, which is used as a general measure for body fat. BMI is an attempt to quantify the amount of tissue mass (muscle, fat, and bone) in an individual. It has been widely used in public health and clinical practice. Given an individual's height and weight, the calculation of BMI is computed by [18]:

$$BMI = \frac{weight(lb) \times 703}{height(in)^2} \tag{1.1}$$

According to the values of BMI, people are categorized as underweight (BMI ≤ 18.5), normal weight ($18.5 <$ BMI ≤ 25), overweight ($25 <$ BMI ≤ 30) and obese (BMI > 30).

In addition to a body fat measurement, BMI is also a risk factor for many diseases. For example, several works [10, 11] demonstrated that increased BMI is associated with some cancers for both males and females, such as breast cancer, colon cancer, thyroid cancer, etc. Wolk et al. [13] presented that BMI is a risk factor for unstable angina and myocardial infarction in patients. Meigs et al. [9] studied the risk of type 2 diabetes or cardiovascular disease (CVD) stratified by BMI.

Taking into account the close connection between BMI and some diseases, BMI is important for personal health monitoring and medical research. Generally, BMI is measured in person with special devices. Populations feel the urge for conveniently preventive tools and methods to increase their self-awareness so as to achieve a better state of health. Computer vision, by now entered in our daily life could be a favored mean for providing such new techniques. This dissertation investigates the feasibility of analyzing BMI from visual appearance. In other words, we want to decode the BMI information from the aspect of visual appearance.

The motivation for this proposal comes from several aspects. First, from human vision, body weight and fat can be intuitively observed by humans from the 2D body and face images. Some examples with corresponding BMI values are shown in Fig 1.1. The increase in body and face adiposity can be observed by human vision without difficulty.

Figure 1.1: Some images with BMI values and corresponding categories. The increase in body adiposity is observed as the BMI value increases.

Second, many studies in health science [19–26] had shown that some anthropometric measures, such as waist-thigh ratio, waist-hip ratio, waist circumference, the cheekbone width to jaw width ratio, average size of eyes and facial shape, etc. are indicators for obesity and are correlated to BMI values. Based on the above intuitive observation and health science studies, we believe that it is worth analyzing body weight or BMI from human visual appearance.

1.2 Related Work

This section gives a detailed literature review on BMI related study, including health science studies and computer vision based computational methods. Existing works study BMI and body weight from various types of data, such as anthropometric measurements, 3D body data, face images, etc. According to the types of the data, first, we simply divide them into two parts: body related data and face related data. In the following section, we introduce the related work from these two types of data and present their limitations and existing challenges.

1.2.1 Study Body Weight and Obesity from Body Data

Anthropometric indicators for obesity or BMI: In the past few years, a lot of studies in the health science had shown that some anthropometric measures, such as waist-thigh ratio, waist-hip ratio, waist circumference, etc., are indicators for obesity or are correlated to BMI values. Ashwell et al. [20] proposed a simple method to classify

the female fat distribution from two factors: waist and thigh circumferences. Later on, Ashwell et al. [27] found that the correlations between body measurements and intra-abdominal/subcutaneous fat are related to the fat distribution of the female body. The body measurements they studied include the ratio of waist to hip circumference, and the ratio of waist to thigh circumference. In addition to analyzing the correlation between body measurements and intra-abdominal fat, Seidell et al. [22] showed that the amounts of intra-abdominal fat and subcutaneous abdominal fat could be accurately predicted from several circumferences, skinfold measurements, BMI and age. Considering skinfold thicknesses and body circumferences both are associated with body fat, Mueller et al. [28] addressed the question of whether body circumferences are inherently more reliable than skinfold thicknesses. According to their study, it is shown that the reliability of six body circumferences is significantly higher than that of skinfold thicknesses at five sites, suggesting that circumferences are a more reliable method. These six body circumferences include that of the forearm, mid-upper arm, calf, chest, waist, hip. Though a variety of anthropometric indicators for abdominal obesity have been suggested in many studies, the literature lacks a systematic evaluation of the proposed indicators taking into account possible differences between genders, age categories and ethnic groups and different diseases and mortality. To provide solid basis for the selection and use of anthropometric indicators for analyzing abdominal obesity, Molarius et al. [19] listed a number of different indicators suggested in the literature as best measures for abdominal obesity, such as waist-thigh ratio, waist-hip ratio, waist-height ratio and waist circumference, etc, and evaluated their performance taking into account possible differences.

Analyzing human weight from anthropometric data: Taking into account the abundant literature in health science have shown the close association between anthropometric measurements and obesity, Velardo et al. [29] studied the feasibility of estimating weight from given anthropometric. A polynomial regression model was employed to predict body weight from the anthropometric data. A large health database NHANES [30] was exploited for the model training, while its validation was performed both on ideal and realistic conditions. The experimental results showed that under noisy data conditions the method could provide accurate estimations, putting the basis for future work towards an automatic weight estimation. Later on, Cao et al. [31] investigated the method of predicting certain soft biometrics, such as gender and weight from a large set of true measurements of the body (provided by CAESAR 1D database). Detailed definitions, usage and performance of 43 anthropometric measurements were included in

their work.

Estimating human weight from body RGB-D data: There are a few works analyzing body weight from RGB-D data. Nahavandi et al. [32] presented a neural network based method to estimate BMI from simulated depth images of the human body. First, 3D manikins are generated using the MakeHuman open source software. During this stage, body weight is estimated as the ground-truth. Then depth images are generated from 3D manikins by Blender open source software. Pfitzner et al. [33] presents a method for estimating body weight from RGB-D body data and thermal data of lying people in clinical environment. First, anthropometric features are extracted from the frontal view of RGB-D data. Thermal data is used to ease the segmentation of a person from the background. Then the features are forwarded to an artificial neural network for weight estimation. Later on, Pfitzner et al. [34] extended their previous work [33] by adding two more scenarios: standing and walking people.

Estimating human weight from video frames: Besides the body weight estimation from RGB-D data, it is also possible to estimate it from a sequence of video frames. Labati et al. [35] developed a weight estimation approach from frame sequences representing a walking person. The method analyzes pairs of frame sequences captured by two cameras (frontal and side views) and extracts features related to the dimensional characteristics of the silhouette. A computational approach is then used to estimate weight from extracted features by evaluating the relations between the visual characteristics and the weight of the person. Experiments were performed with 20 subjects, walking in eight different directions. A maximum absolute mean error was recorded with less than 2.4 kg. In another work, Arigbabu et al. [36] presented an approach for estimating soft biometrics, e.g., body height and weight, from video frames which record the walk process of each participant (frontal and side views). First, they extracted the silhouette of people from each frame by image processing techniques like background subtraction. Then 13 pixel-density based features are extracted from the segmented body regions. The features are finally forwarded to an artificial neural network (ANN) to estimate body weight. In experiments with 80 subjects, they reached a mean average error of 4.66 kg the estimation of body weight.

Predicting human weight from body 3D Data: Taking the advantages of 3D reconstruction technology, such KinectFusion, researchers began to study body weight from 3D body data. In 2012, Velardo et al. [37] studied the weight estimation from 3D body data collected by Microsoft Kinect sensor. Several anthropometric measures (same

as those used in their previous work [29]) are extracted from the body 3D data. They utilized a neural network regressor instead of a polynomial regression model used in [29] to learn the map between the extracted features and body weight. Velardo et al. [38] presented an approach to estimate the weight of a person within 4% error using 2D and 3D data extracted from a low-cost Kinect RGB-D camera output. To obtain accurate 3D data, they smoothed the depth map by convolving its 2D projection with a Gaussian kernel, and then they removed the outliers at the edges by taking only the pixel belonging to the binary mask provided by the background separation algorithm. The body weight is estimated from the extracted anthropometric features by a regression model. Both [37] and [38] estimate BMI from the extracted anthropometric features, a new method which directly estimates BMI from people in 3D space is desired.

From the above reviewed literatures, we found that there are two limitations of estimating weight from body data. First, existing methods [32–34] depend on both color and depth images. Is it possible to analyze weight just from single-shot body images? Second, both [37] and [38] estimate BMI from the extracted anthropometric features, a new method which directly estimates BMI from people in 3D space is desired.

1.2.2 Study BMI from Facial Data

Relationship between facial appearance and health: The relationship between facial cues and perceptions of health have been studied by many researchers from various aspects for a long time. There are several studies explored the association between facial adiposity and various diseases. Coetzee et al. [23] demonstrated that facial adiposity (the perception of weight in the face) significantly improve the prediction of perceived health and attractiveness in a curvilinear relationship, and perceived facial adiposity is significantly associated with measures of cardiovascular health and reported infections. Tinlin et al. [39] showed that young adult women's facial adiposity is better predicted by their body weight than by their body shape, and are correlated with a composite measure of their physical and psychological condition (such as stress, anxiety, and depression). In [40], De Jager et al. reviewed that facial adiposity has also been linked to various health outcomes such as cardiovascular disease, respiratory disease, blood pressure, immune function, diabetes, arthritis, oxidative stress, hormones, and mental health. In addition to facial adiposity, facial skin coloration also affects perceived health. Stephen et al. [41] suggested that facial skin colors are associated with health and also play a role in the perception of health in human faces. Increased skin yellowness and lightness suggest a

role for high carotenoid and low melanin coloration in the healthy appearance of faces. Furthermore, the facial shape has been successfully used for predicting physiological health. Stephen et al. [42] applied the geometric morphometric methodology to facial shape data, producing models that successfully predict aspects of physiological health (including percentage body fat, body mass index and blood pressure) from 272 Asian, African, and Caucasian. The experimental results suggested that facial shape provides a valid cue to aspects of physiological health.

Association between facial appearance and BMIs/body fat: Over the past decade, researchers began to pay attention to the association between facial appearance and BMI. Coetzee et al. [24] studied the correlation between BMI and three facial geometric metrics−cheek-to-jaw-width ratio (CJWR), face width-to-height ratio (WHR) and face perimeter to area ratio (PAR). They recruited 95 Caucasian and 99 African participants to capture face images for this study. According to the experimental results, there is a significant correlation existing between the facial feature and BMI values. Later on, Pham et al. [25] further analyzed the correlation between BMI values and four other geometric metrics−eye size, lower face to face height ratio (LF/FH), face width to lower face height ratio (FW/LFH) and mean of eyebrow height among the group of young and elder in Korean. Facial images of 911 participants were analyzed. These data indicated that these four facial metrics are correlated with BMI. To estimate the strength of the relationship between perceived facial adiposity and BMI, De Jager et al. [40] conducted a meta-analysis to evaluate the quantified the relationship between perceived facial adiposity and BMI. A model weighted by sample size was performed. The analysis revealed a strong positive overall correlation between perceived facial adiposity ratings and BMI with $r = 0.71$. In addition to BMI, facial metrics are correlated to visceral obesity. To determine the best predictor of the normal waist and visceral obesity among these characteristics, Lee et al. [43] investigated the association of visceral obesity with facial characteristics. They extracted 15 facial characteristics from 2D images and identified the strongest predictor for each age-gender group. They also assessed the predictive power of different combinations of characteristics. Recently, researchers found facial skin color and texture are also associated with BMI prediction. Henderson et al. [26] investigated the effect of multiple facial cues on health judgments from both 2D and 3D face images. They found that except general face geometric features, global face shape and skin color are also associated with BMI prediction. Mayer et al. [44] assessed the association of BMI values with facial attributes−shape and texture (color pattern) in

the female group. The experiment was conducted on 49 standardized images of female participants. They found that the faces of women with high BMIs had wider and rounder facial outlines relative to the size of the eyes and lips, and relatively lower eyebrows. Furthermore, they showed that women with higher BMIs have brighter and more reddish facial skin color.

Estimating BMI from 3D face data: There are a few works estimating BMI from 3D volume reconstruction of the face. Pascali et al. [45] proposed a framework for estimating body weight from 3D facial data collected by low-cost depth scanners. Body weight is estimated by the geometric features extracted from the 3D model. A method for automatically computing geometric features is proposed. Given a 3D face scan labeled with a set of landmarks, Giorgi et al. [46] utilized persistent homology descriptors to get geometric and topological information of the face. By applying dimension reduction techniques to the dissimilarity matrix of descriptors, they got a space in which each face was a point and face shape variations are encoded as trajectories in that space. By analyzing the shape patterns of single individuals as trajectories, it is shown that this method helps to assess the weight gain or loss of individuals.

Estimating BMI from 2D face images by geometric feature: The computational method which predicts BMI values from 2D face images began with utilizing geometry based facial features. Wen et al. [47] proposed a novel geometry based computational method for automatically predicting BMI values from 2D face images. This is the first work on visual BMI estimation from facial images. The psychology inspired geometric features (PIGF) are computed for BMI prediction. Three regression methods: the support vector regression (SVR) [48], Gaussian process (GP) [49], and the least-squares estimation [50] are used for learning the map between facial features and BMI values. This method was evaluated on a large dataset: Morph II database [51]. Barr et al. [52] utilized the method proposed in [47] to identify whether BMI values can be correctly identified from participants' facial images in order to improve data capturing in dissemination and implementation research. According to the BMI values, there are mainly four BMI categories: underweight (BMI$\leq$18.5), normal (18.5<BMI$\leq$25), overweight (25<BMI$\leq$30), obese (BMI>30). Experimental analysis indicated estimated BMIs are more accurate in normal and overweight categories while they are less accurate in underweight and obese categories. [53] explored the utility of a data-driven approach for assessing BMI from face images. They employed a data-driven approach in which statistical models were built using principal components (PCs) derived from the objectively defined shape and

color characteristics in face images. The predictive power of these models is compared with models based on previously studied facial proportions (perimeter-to-area ratio, width-to-height ratio, and cheek-to-jaw width). Experimental results showed that 2D shape/color PCs based models perform better than others.

Estimating BMI from 2D face images by deep feature: Recently, deep learning based approaches have shown promising results in face recognition [54–56], and other visual tasks, such as image retrieval [57, 58] and pose estimation [59, 60]. To take advantages of deep learning based features, some works utilize pre-trained deep network for BMI estimation. Kocabey et al. [61] analyzed BMIs from face images collected from a social media website. The pre-trained VGG-Net and VGG-Face models [62] are used to extract features from facial images. They employed SVR models to predict BMIs from the extracted features. A comparison of BMI prediction between this method and human cognition was presented. It is shown that human performs better on small BMI differences predictions, and there is no performance difference for larger BMI difference predictions. Dantcheva et al. [63] explored the possibility of estimating height, weight and BMI from facial images by a regression method based on the 50-layers ResNet architecture. They evaluated their methods on a celebrity dataset of facial images with the annotation of weight, height and gender. They also analyzed the influence of gender on BMI estimation.

According to the above literatures review, we found that there are two challenges have not been well studied yet. First, so far there is no literature systematically evaluate and compare various face representations for visual BMI analysis, especially the two types of facial representations: the geometric features and deep learning based features. Second, the above existing methods for estimating BMI from 2D images all consider the estimation as a regression problem. They ignore the ambiguity of BMI labels.

1.3 Research Goals

With the above limitations and challenges for BMI estimation from human visual appearance, in this dissertation, we focus on addressing these problems. Specifically, we describe these research goals (RG1-RG2) in detail as follows:

- *RG1-Body weight analysis from 2D and 3D body data:* Existing methods [32–34] use both color and depth images to estimate weight. Our first research goal is to investigate the feasibility of analyzing body weight from single 2D frontal view human body images. To the best of our knowledge, there is no existing work that

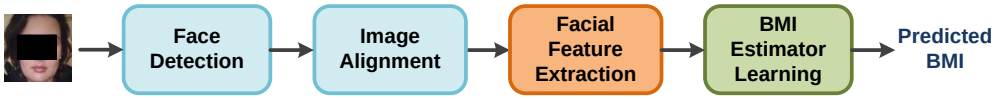


Figure 1.2: A typical framework for visual BMI estimation from 2D face images.

explores weight or BMI from 2D body images only. In addition, considering [37] and [38] which estimate BMI by anthropometric features extracted from 3D data, we aim to develop a computational approach that directly estimates body weight and height from dressed people in 3D space.

- *RG2-BMI analysis from facial images:* Fig. 1.2 shows a typical framework for BMI estimation from 2D facial images. It consists of four steps: face detection, image alignment, facial representation extraction, and BMI estimator learning. The third and fourth steps both are important which dominantly determines the performance of a BMI estimation method. For the third step, so far there is no literature systematically evaluate various facial feature extraction methods for BMI estimation, thereby we dedicate to this blank research area. For the fourth step, all existing methods consider the estimation as a regression problem and ignore the ambiguity of the BMI labels. We aim to explore the efficient method to address the ambiguity of BMI labels.

1.4 Summary of Contributions

In this dissertation, to address the aforementioned research problems, we conduct the feasibility study for weight estimation from body images, evaluate facial extraction methods for BMI estimation, propose a two-stage learning framework to address label ambiguity and develop an end-to-end convolutional neural network (CNN) for BMI estimation. The contributions can be summarized as follows:

- *Analyze body weight from human body images [64]:* Motivated by the recent health science studies [19, 21], we investigate the feasibility of analyzing body weight from 2D frontal view human body images. A framework is developed for analyzing body weight and BMI from 2D human body images. Computation of five anthropometric features is proposed for body weight characterization. Correlation is analyzed between the extracted anthropometric features and BMI values, which validates the usability of the selected features. A visual-body-to-BMI dataset is collected and

cleaned to facilitate the study, which contains 5900 images of 2950 subjects along with weight, height and gender information. Body weight analysis is studied at three levels of difficulties (from easy to difficult, based on human perception). The proposed method outperforms two state-of-art facial images based weight analysis approaches on most test sets. To the best of our knowledge, this is the first work to explore BMIs from 2D body images only.

- *Estimate BMI from dressed people in 3D Space [65]:* We study BMI estimation from the 3-dimensional (3D) visual data by measuring the correlation between the estimated body volume and BMIs and then develop an efficient BMI computation method. Our approach consists of body weight and height estimation from normally dressed people in 3D space. To address the influence of loose clothes on body volume estimation, two clothes models are developed to make the volume estimation more accurate. A new RGB-D video dataset is collected for this study, and the reconstructed 3D data are provided by the KinectFusion on depth data. Experimental results show the effectiveness of the approach to work on normal conditions of dressed people. The MAE of the estimated BMI can achieve 2.54 in our experiments.

- *Evaluate and analyze facial feature extraction methods for BMI estimation [66]:* We studied the visual BMI estimation problem based on the characteristics and performance of different facial representations. Various facial representations, including geometry based and deep learning based representations, are comprehensively evaluated and analyzed from three perspectives: the overall performance on visual BMI prediction, the redundancy in facial representations and the sensitivity to head pose changes. The experiments are conducted on two databases: a new dataset we collected, called the FIW-BMI and an existing large dataset Morph II. Our studies provide some deep insights into the facial representations for visual BMI analysis.

- *Propose a two-stage learning method for visual BMI estimation [67]:* We investigate the problem of visual BMI estimation from facial images by a two-stage learning framework. BMI related facial features are learned from the first stage. Then a label distribution based BMI estimator is proposed for the second stage. Two label assignment strategies are analyzed for modeling the single BMI value as a discrete probability distribution over the whole ranges of BMIs. Extensive experiments are conducted on FIW-BMI and Morph II databases. The experimental results

show that the two-stage learning framework improves the performance step by step. More importantly, the proposed estimator efficiently reduces the estimated error and outperforms other regression and label distribution methods (LDL-IIS and LDL-CPNN).

- *Propose a label assignment matching based neural network for BMI estimation:* To address the challenges caused by limited BMI data and ambiguity of BMI labels, we integrate feature learning and estimator learning in one neural network. A label assignment scheme is embedded into the deep network which models the scalar BMI label as a probability distribution. A triple-loss function is proposed for label assignment matching which minimizes the discrepancy between estimated labels and ground-truth labels. Extensive experiments are conducted on two datasets: Morph II and FIW-BMI. The experimental results show that the three loss functions all contribute to the improvement of the performance. Furthermore, the proposed method is more accurate than other state-of-the-art regression based and label distribution based methods.

1.5 Dissertation Outline

The rest of the dissertation is organized as follows.

Chapter 2 addresses the problem of analyzing body weight from 2D frontal view body images. It begins with describing the three problems to be studied in Section 2.1. A newly collected and cleaned visual-body-to-BMI dataset is introduced in Section 2.2. Then Sections 2.3 and 2.4 present the feature extraction and learning methods. Detailed experimental results and discussion are provided in Section 2.6. Conclusions are summarized in Section 2.7.

Chapter 3 studies BMI estimation from the 3D visual data. We present the methods developed for weight and height estimation and the two clothes models for correcting the estimated body weight in Section 3.2. Then the RGB-D video dataset newly collected for this work is introduced in Section 3.3. The correlation between the estimated body volume and BMIs, and experimental results are given in Section 3.5. Section 3.6 summaries the major conclusions of this chapter.

In Chapter 4, we investigate the visual BMI estimation from the aspect of the facial feature extracting methods. First, the problem to be studied is described in Section 4.1. The principles and existing facial feature extraction methods are systematically

presented and discussed in Section 4.2. Then, Section 4.3 presents two databases used for performance evaluation: the newly collected FIW-BMI and Morph II. Detailed experimental results and analysis are provided in Section 4.4. The conclusions are given in Section 4.5.

Chapter 5 proposed a two-stage learning framework for visual BMI estimation from facial images. First of all, the existing challenges of facial BMI estimation are described in Section 5.1. Then, Section 5.2 presents the proposed method for BMI estimation. The results of extensive experiments are reported in Section 5.4. Finally, a brief summary of this chapter is given in Section 5.5.

In Chapter 6, we propose a convolutional neural network (CNN) for visual BMI estimation which integrates feature learning and estimator learning in one network. This chapter begins with introducing the challenge of this topic in Section 6.1. Then regression based, ranking based and label distribution based methods are reviewed in Section 6.2. Section 6.3 describes the details about the label assignment matching based learning network and the triple-loss function. The experimental setting and results are presented in Section 6.5. Conclusions and future work are given in Section 6.6.

Chapter 7 summarizes the conclusions obtained in this dissertation and addresses the future work.

Chapter 2

Body Weight Analysis from Human Body Images

This chapter is dedicated to a new problem termed as body weight analysis from single-shot 2D human body images. Considering there is no existing work that has addressed this problem, the purpose of this chapter is to investigate the feasibility of analyzing the body weight from 2-dimensional (2D) frontal view human body images. To investigate the problems at different levels of difficulties, three feasibility problems, from easy to hard, are studied. More specifically, a framework is developed for analyzing body weight from human body images. Computation of five anthropometric features is proposed for body weight characterization. Correlation is analyzed between the extracted anthropometric features and the BMI values, which validates the usability of the selected features. The social networks provide abundant data. A visual-body-to-BMI dataset is collected and cleaned to facilitate this study.

The outline of this chapter is as the following. Section 2.1 introduces the three problems we study and presents the framework for body weight analysis. Section 2.2 describes the newly collected and cleaned visual-body-to-BMI dataset. Details about feature detection and computational method are presented in Section 2.3. Section 2.4 describes the employed machine learning models. The calculation of Pearson's correlation and performance measurements are presented in Section 2.5. In Section 2.6, we first calculate the correlation between the extracted features and the BMI values; then the detailed experimental results and discussion are provided. Finally, summery and discoveries of this chapter are given in Section 2.7.

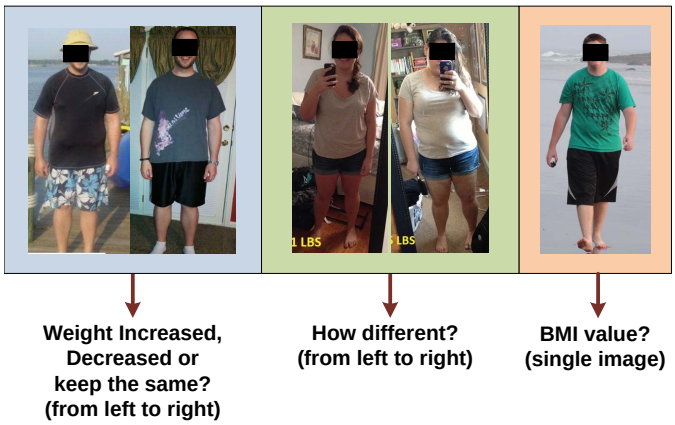


Figure 2.1: Three kinds of problems explored for body weight analysis. For the pairwise images, the change is from the left one to right one.

2.1 Problems to Study

The studies in health science [19–21] show evidence on the relation between some anthropometric measures and obesity. Considering that the BMI is a widely used body weight/fat indicator, we employ this index as the measure for the body weight. This study explores the relation between the BMI values and the visual appearance of the human body. The correlation between BMI values and the computed anthropometric features is studied first. Given the correlation, we analyze the body weight issue from 2D human body images at different levels of difficulties (from easy to difficult, based on human perception).

Fig. 2.1 shows the three problems studied for body weight analysis. First, we recognize the weight difference from a pair of frontal view body images. This is defined as a three-class classification problem. The output is a triple classification result which decides whether the weight of the subject is increased, decreased or keeping the same. In our dataset, the height of each subject remains the same height in the corresponding pairwise images, thereby the weight difference is equivalent to BMI difference. Then, we go further and estimate how big the weight or BMI difference between the pairwise images is. The above two problems are studied based on the pairwise images from the same individual. The key of these two problems is to measure whether the change between the two images can be computed or not. A more challenging task is to directly estimate the BMI value from a single body image.

Fig. 2.2 depicts the framework of the body weight analysis approach, which consists of three steps:

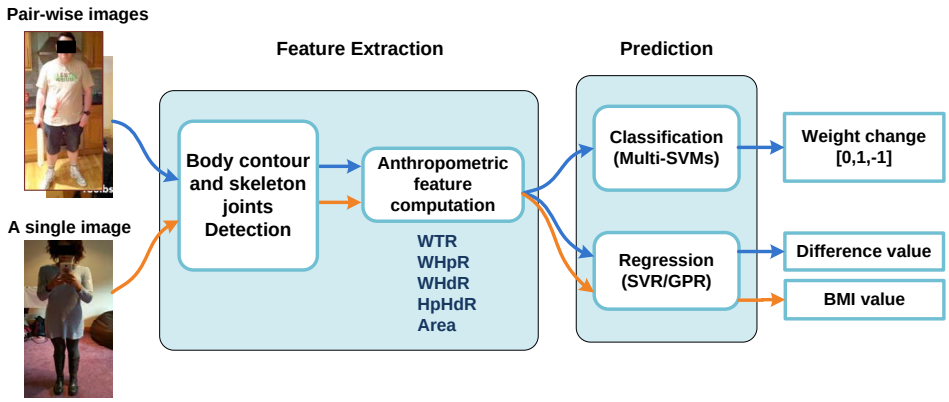


Figure 2.2: The framework of our proposed weight analysis approach. The approach can take either pairwise body images or a single image as input. It classifies and predicts the BMI difference from pairwise images, or estimate the exact BMI value from a single image.

1. Body contour and skeleton-joints detection.
2. Anthropometric feature computation from the body images.
3. Apply statistical models to map the features to the weight differences or the BMI values.

As shown in Fig. 2.2, the approach can classify the weight difference from the pairwise images. The classification is an output of three different results $\{0,1,-1\}$: *0* indicates no weight change, *1* indicates weight increased, and *-1* means weight decreased. The order of the images in the pair does matter, and the change is from 'previous" left to 'current" right, as indicated in Fig. 2.1. Note that the prediction of the BMI differences and the BMI values are solved by two different regression models. The details about feature extraction and mapping will be given in Section 2.3 and 2.4, respectively.

2.2 Dataset with Cleaning

The human body images are downloaded from the website Reddit posts [1]. In total there are $47,574$ images of $16,483$ individuals. Each individual has at least one "previous" and one "current" images (or a collage which was made by sticking several images). As shown in Fig. 2.3, all the images under the same individual folder have the same annotations (except the image number). The format of the original annotation

[1] Website: `http://www.reddit.com/r/progresspics`

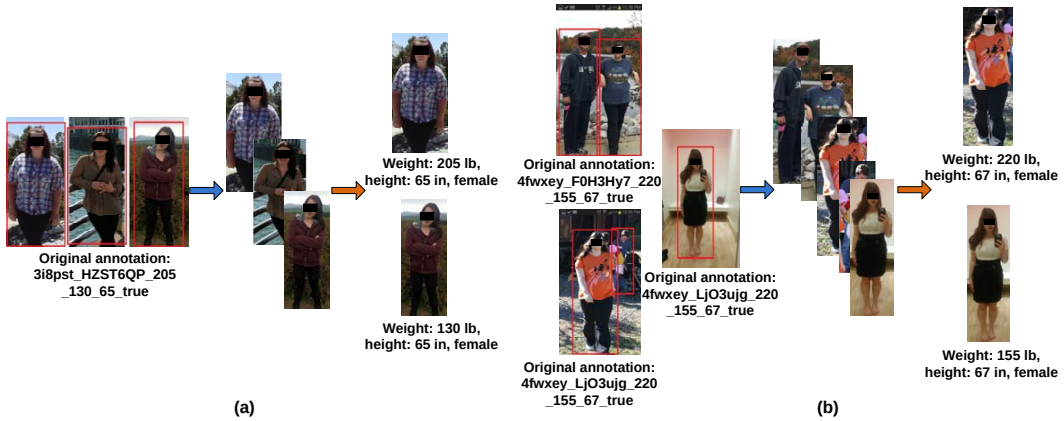


Figure 2.3: The illustration of cleaning images with automatic and manual steps. Two cases are given. The first case (left panel) shows the individual just contains one collage (an image made by sticking several images). The second case (right panel) shows the individual contains 3 images, among them there are 2 group photos (more than one person shown on the image). The blue arrow represents the process of cropping each single body from a composite image based on automated body detection. The orange arrow represents the manual process of correcting annotations. The annotations for the "previous" and "current" images are visually distinguished by body size and shape.

is "ID_image number_previous weight_current weight_height_gender". Thereby, all the images under the same individual folder share the same information about weights ("previous" and "current") and height, the weight for each image cannot be automatically distinguished by algorithms. It needs manual processing (visually check) to correct the weight for the individuals.

We processed and cleaned the dataset with automatic and manual steps which are described below. First, we went through the original images by a body detector, using a method similar to [59]. Then, given the detected bodies, each single body image was cropped from the original images. During the process, we kept the cropped body images containing both head and frontal body (with required joints detected). If there are greater than or equal to 2 cropped images kept for the individual, the algorithm kept the individual folder. Now the left (cropped) images under the same individual folder still share the same annotation ("ID_image number_crop number_previous weight_current weight_height_gender"). The next step was to visually distinguish which image has the "previous" weight and which has the "current" weight. Since the annotations only have the "previous" and "current" body weights for each individual, just one "previous" image and one "current" image was kept for each individual. Finally, we manually corrected

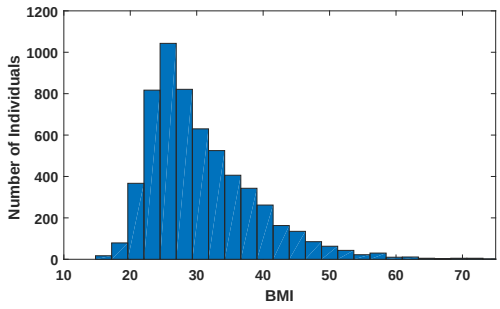


Figure 2.4: The BMI distributions of the body-to-BMI dataset. The BMI distribution is in a wide range from 15 to 75.

the annotations for these images.

Fig. 2.3 shows the procedure of processing the images with automatic and manual steps. Two cases are introduced: the first case shows the individual just contains one collage (an image made by sticking several images); the second case shows the individual contains 3 images, among them there are 2 group photos (more than one person shown in the image). The blue arrow represents the automatic process of cropping every single body from the images. Each cropped body in the image is labeled by a red boundary box. The orange arrow represents the manual process of distinguishing and correcting annotations. The annotations for the "previous" and "current" body images are visually distinguished by body size and shape. A pair of images mentioned throughout this work is one "previous" and one "current" body images from the same individual.

After these procedures, there are 2950 subjects (individuals) left, each contains two images: one "previous" and one "current". This leads to a total of 5900 images with corresponding labels of gender, height, and weight. The set of images is noted as visual-body-to-BMI dataset containing 966 females and 1984 males. The ground truth of BMI can be calculated. The BMIs distribution of the body-to-BMI dataset is shown in Fig. 2.4. The BMIs distribution is in a wide range from 15 to 75. Specifically, 46 body images are in the underweight range (BMI ≤ 18.5), 1416 are normal ($18.5 <$ BMI ≤ 25), 1863 are overweight ($25 <$ BMI ≤ 30) and 2575 are obese (BMI > 30). By comparing the weight of the "previous" and "current" images, we conclude that 1246 subjects show the increase in weight, 1233 subjects show the decrease in weight, and the rest 481 subjects have the same weight in both images. The height of each subject remains the same in a pair of images. The subjects are natural with various clothing styles.

Figure 2.5: The body contour and skeleton-joints detected by the CSJ detector. The brick red area represent the detected body part. The asterisks represents the skeleton joints.

2.3 Feature Extraction

In this section, we present the details about feature extraction for the proposed approach. Body contour and skeleton joints (CSJ) detection is the first step for feature extraction. The output of the detection is used for anthropometric feature computation.

2.3.1 Contour and Skeleton-joints Detection

Body contour and skeleton joints (CSJ) detection are based on deep networks for contour and skeleton joints detection. Fig. 2.5 shows the body contour and skeleton joints detected by the CSJ detector. The brick-red area represents the detected body part. The asterisks represent the detected skeleton joints.

To detect the body contour from an image, pixel-level image segmentation is applied to it. We use the conditional random fields as recurrent neural networks (CRF-RNN) method [68] for body detection. The mean-field CRF inference is reformulated as a RNN, then the CRF-RNN layer (iterative mean-field layer) is plugged into a fully convolutional neural network (FCN). By applying the CRF-RNN method to the image, the body regions are labeled out, while all other regions in the image are labeled as the background. This leads to a set B contains all pixels which are labeled as the human body, and a set G contains all pixels which are labeled as the background. $l_{x,y}$ represents the label assigned to the pixel locates at (x, y), where (x, y) denotes the horizontal and vertical coordinates on the image. $l_{x,y}$ is from a pre-defined set of labels $L = \{b, g\}$. Here b is the label for the human body and g is for the background. Then $B = \{(x, y) : l_{x,y} = b\}$ and $G = \{(x, y) : l_{x,y} = g\}$. We will use this in Section 2.3.2 for computing anthropometric

features.

With the locations of skeleton joints in an image, the key parts (waist, hip, etc.) are located for extracting the anthropometric features. In this work, the convolutional pose machine (CPM) [69] is employed to detect the skeleton joints from body images. CPM consists of a series of convolutional neural networks (CNN) that repeatedly produce 2D belief maps for the location of each body part. The belief maps produced by the previous CNN are used as the input of the next CNN. By using the CPM, a list of coordinates of the key skeleton joints can be obtained, such as left hip, right hip, left shoulder, and right shoulder, etc. The coordinates of skeleton joints will be used for computing anthropometric features.

2.3.2 Anthropometric Feature Computation

Several anthropometric indicators suggested in health science [19–22] are used as measures for obesity. Some listed indicators include waist-thigh ratio, waist-hip ratio, abdominal sagittal diameter, waist circumference, and hip circumference. Taking into account these indicators, we have five anthropometric features automatically detected and computed from the body images, including waist width to thigh width ratio (WTR), waist width to hip width ratio ($WHpR$), waist width to head width ratio ($WHdR$), hip width to head width ratio ($HpHdR$), and body area between waist and hip ($Area$). Among these features, $Area$ is inspired by our human perception.

The measurement of the waist circumference and the hip circumference cannot be directly obtained from 2D images. We consider the particular body part as a cylinder. Then we use the width of the body part (on a 2D image) to approximately represent the circumference of a particular body part. A similar approximation has been utilized and verified in [29]. They used the width of the upper arm, leg, waist, and calf to test a polynomial regression model, which was trained by the real circumferences of the body parts. Since the absolute measures of the waist width and hip width cannot be obtained from 2D images without metric/scale information, thereby we compute the ratio to characterize the relative measures.

Fig. 2.6 illustrates the anthropometric features visually. There are 18 detected skeleton joints shown in the figure labeled with asterisks. In the following, we use the coordinates of 8 detected skeleton joints for computing anthropometric features. These 8 skeleton joints are the nose, left ear, right ear, center shoulder, left hip, right hip, left knee, and right knee. The abbreviations of skeleton joints or boundaries involved for

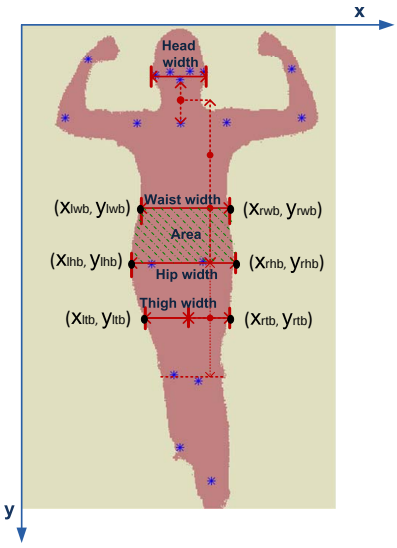


Figure 2.6: The anthropometric features computed for body weight analysis. The 18 skeleton joints (labeled by asterisks) are nose, left eye, right eye, left ear, right ear, center shoulder, left shoulder, right shoulder, left elbow, left hand, right elbow, right hand, left hip, right hip, left knee, right knee, left ankle and right ankle. The area filled with green dash dots denotes the feature *Area*.

Table 2.1: Abbreviations of body parts for feature computation.

Body part	Abbrev.	Body part	Abbrev.
Nose	n	Hip	h
Left ear	le	Left hip	lh
Right ear	re	Right hip	rh
Center shoulder	cs	Left hip boundary	lhb
Waist	w	Right hip boundary	rhb
Left waist	lw	Thigh	t
Right waist	rw	Left thigh boundary	ltb
Left waist boundary	lwb	Right thigh boundary	rtb
Right waist boundary	rwb	Knee	k
Left knee	lk	Right knee	rk

feature computation are given in Table 2.1. The abbreviation of a body part is used as an index that denotes the location of the pixel. For example, the pixel of left hip is denoted as p_{lh}, and its coordinate is denoted as $(x_{\text{lh}}, y_{\text{lh}})$. The size of the input image is $M \times N$. The methods for computing the five anthropometric features are described below:

1) *WTR* the ratio of waist width to thigh width. A general knowledge about human body proportions [70] is used to initially estimate the location of waist and thigh based

on the detected locations of hip and head. As shown in Fig. 2.6, the vertical location of the waist is computed by: $y_{\mathrm{w}} = \frac{2}{3}y_{\mathrm{h}} + \frac{1}{6}(y_{\mathrm{n}} + y_{\mathrm{cs}})$, where $y_{\mathrm{h}} = \frac{1}{2}(y_{\mathrm{lh}} + y_{\mathrm{rh}})$. Similarly, the vertical location of the thigh $y_{\mathrm{t}} = \frac{1}{2}(y_{\mathrm{k}} + y_{\mathrm{h}})$, where $y_{\mathrm{k}} = \frac{1}{2}(y_{\mathrm{lk}} + y_{\mathrm{rk}})$. With the vertical locations of waist and thigh, the next step is to estimate the waist width and thigh width. Taking waist width as an example, this calculation can be considered as fixing $y = y_{\mathrm{w}}$, and searching for the x-axis coordinates of the left and right waist boundaries x_{lwb} and x_{rwb} from the contour image. The x-axis coordinate of left waist boundary x_{lwb} can be computed by:

$$\begin{aligned} x_{\mathrm{lwb}} &= \underset{x}{\operatorname{argmin}} \left| x - x_{\mathrm{cw}} \right|, \\ s.t.\ & x \in [0, x_{\mathrm{cw}}],\ (x, y_{\mathrm{lw}}) \in G. \end{aligned} \tag{2.1}$$

Here x_{cw} is x-axis coordinate of the center waist, which can be approximated by x-axis coordinate of the center shoulder x_{cs}. y_{lw} and y_{rw} both are equal to y_{w}. G is a set contains all pixels labeled as the background. Similarly, x_{rwb} is given by:

$$\begin{aligned} x_{\mathrm{rwb}} &= \underset{x}{\operatorname{argmin}} \left| x - x_{\mathrm{cw}} \right|, \\ s.t.\ & x \in [x_{cw}, M],\ (x, y_{rw}) \in G. \end{aligned} \tag{2.2}$$

Here M is the width of the image. y_{lwb} and y_{rwb} both are equal to y_{w}. The thigh boundary along the x-axis is determined by x_{ltb} and x_{rtb}, which can be calculated in the same way as Eqn. (2.1) and (2.2). With the coordinates of these boundaries, the waist width is the Euclidean distance between p_{lwb} and p_{rwb}. Thigh width is half of the Euclidean distance between p_{ltb} and p_{rtb}. So WTR is computed by:

$$WTR = \frac{d\left(p_{\mathrm{lwb}}, p_{\mathrm{rwb}}\right)}{0.5 \cdot d\left(p_{\mathrm{ltb}}, p_{\mathrm{rtb}}\right))}, \tag{2.3}$$

where $d(\cdot)$ denotes the Euclidean distance between the two pixels.

2) $WHpR$ the ratio of waist width to hip width. Given the left and right hip skeleton-joints p_{lh} and p_{rh}, the left hip boundary p_{lhb} and right hip boundary p_{rhb} are calculated following the rules in Eqns. (2.1) and (2.2). Then hip width is the Euclidean distance between p_{lhb} and p_{rhb}. The $WHpR$ is computed by:

$$WHpR = \frac{d\left(p_{\mathrm{lwb}}, p_{\mathrm{rwb}}\right)}{d\left(p_{\mathrm{lhb}}, p_{\mathrm{rhb}}\right)}. \tag{2.4}$$

3) $WHdR$ the ratio of waist width to head width. Since the images may have different scales, the waist widths computed from the images cannot directly represent the measured

waist width. According to the anthropometry study on adult head circumferences [71], there are tiny differences on the width of adult heads. Thereby, WHdR is computed to represent the waist width. Here head width is the Euclidean distance between left ear p_{le} and right ear p_{re}. Then $WHdR$ is given by:

$$WHdR = \frac{d\left(p_{\text{lwb}}, p_{\text{rwb}}\right)}{d\left(p_{\text{le}}, p_{\text{re}}\right)}. \tag{2.5}$$

4) *HpHdR* the ratio of hip width to head width. As described above, we use this ratio to represent the hip width in each body image. The $HpHdR$ is computed by:

$$HpHdR = \frac{d\left(p_{\text{lhb}}, p_{\text{rhb}}\right)}{d\left(p_{\text{le}}, p_{\text{re}}\right)}. \tag{2.6}$$

5) *Area* the area between waist and hip. Because of the unknown scale information for each image, *Area* is expressed as the number of pixels per unit area between waist and hip. The number of pixels between waist and hip is given by:

$$\#pixels = \sum_{\substack{x\in[0,M]\\ y\in[y_{\text{w}},y_{\text{h}}]}} \mathbb{1}\left[l_{x,y} = b\right], \tag{2.7}$$

here $\mathbb{1}\left[\cdot\right]$ is a indicator function. $l_{x,y}$ represents the label (obtained from CSJ detection) assigned to the pixel locates at (x, y). Then the *Area* is calculated by:

$$Area = \frac{\#pixels}{(y_{\text{h}} - y_{\text{w}}) \cdot 0.5 \cdot [d(p_{\text{lwb}}, p_{\text{rwb}}) + d(p_{\text{lhb}}, p_{\text{rhb}})]} \tag{2.8}$$

As shown in Fig. 2.4, the prediction approach can take the different input: either a pair of images or a single image. The BMI difference can be classified and estimated from a pair of input images. On the other hand, the BMI value can be estimated from a single body image. Five anthropometric features are extracted from each body image, resulting in a feature vector $\mathbf{f} = [WTR, WHpR, WHdR, HpHdR, Area]^T$. For a single image, $\mathbf{f}$ is the feature used for estimation. For pairwise images, the following transformation is applied to the features $\mathbf{f_1}$ and $\mathbf{f_2}$ for generating the transformed feature:

$$\mathbf{f_t} = log\mathbf{f_1} - log\mathbf{f_2}, \tag{2.9}$$

where $\mathbf{f_1}$ and $\mathbf{f_2}$ are features extracted from the "previous" and "current" images, respectively. $log(\cdot)$ denotes applying logarithmic operation to each element in the vector.

After extracting the features from a pair of images or single images, we apply a normalization to the features by:

$$\mathbf{m}' = \frac{\mathbf{m} - \mu}{\sigma}, \tag{2.10}$$

where $\mathbf{m}$ is the extracted feature (denoted as $\mathbf{f}$ or $\mathbf{f_t}$ above). μ is the mean value and σ is the standard deviation, both are calculated from the training data along each feature dimension (there are 5 feature dimensions). Normalization is essential in order to obtain a robust estimation.

2.4 Learning Method

Weight/BMI analysis is to map the anthropometric features to BMI values. The training process is to learn the mapping function. In the estimation, the learned function is used to estimate the BMI values from extracted features. We study the problem with different settings. Since the problem is relatively new and challenging, we explore how well we can achieve at different levels of difficulties:

- Recognize the weight difference (increase, decrease or the same) $\hat{t}_c$, given a pair of images.
- Predict how big is the weight or BMI difference $\hat{t}_d$ between a given pair of images.
- Estimate the BMI value $\hat{t}_v$ from a single body image.

Weight difference recognition is a three-class classification problem. The pairwise features $\mathbf{f_t}$ used for training and testing in this problem are obtained from Eqn. (2.9). The ground-truth label t_c is generated based on the weight change on the pairwise images (suppose each subject has the same height in the image pair). $t_c \in [0, 1, -1]$: *0* denotes keeping the same weight, *1* denotes weight increased, and *-1* denotes weight decreased.

The level of BMI differences is considered as a regression problem. The pairwise-features $\mathbf{f_t}$ are also used for training and testing in this problem. The ground-truth label t_d is the BMI difference of the pair image which may be positive or negative.

BMI value estimation is also defined as a regression problem. The features extracted from each single image $\mathbf{f} = [WTR, WHpR, WHdR, HpHdR, Area]^T$ are used for this problem. The ground-truth label t_v is the BMI value.

We employ the multi-class support vector machines (multi-SVMs) [72] for classification, and the support vector regression (SVR) [48] and Gaussian process regression (GPR) [73] for weight or BMI differences mapping and BMI estimation.

1) Support vector machine (SVMs) are supervised learning algorithms that analyze data for classification or regression. There are two main categories for SVMs: support

vector classification (SVC) and support vector regression (SVR). It has been widely utilized in many problems [74, 75]. SVM can do nonlinear classification using kernel functions. Gaussian radial basis Function (RFB) kernel is one of the most popular kernels. The RBF achieves a better performance in classification or regression than some other kernels.

The SVC is a binary classifier. To get multi-class classification, a set of binary classifiers can be constructed with each trained to separate one class from another. For n classes, this results in $\frac{(n-1)n}{2}$ binary classifiers. Since our classification on BMI difference has three classes $\{0,1,-1\}$ for the a pair of images, 3 binary classifiers are trained accordingly. The SVR uses the same principle is similar to the SVC, but with differences in the optimization.

2)Gaussian processing regression: A Gaussian process (GP) is a collection of random variables, and a finite number of variables that have a joint Gaussian distribution [76]. GPR means Gaussian process regression. The prior mean and covariance of the GP need to be specified. The prior mean is assigned constant and zero, or the mean of the training data. The prior covariance is specified by passing a kernel object. The hyperparameters of the kernel are optimized by maximizing the log-marginal-likelihood. A rational quadratic kernel is employed for GPR. Given a set of training examples $(a_1, b_1), ...(a_n, b_n)$, the rational quadratic kernel is defined as:

$$k(a_i, a_j) = \left(1 + \frac{D(a_i, a_j)^2}{2\alpha\iota^2}\right)^{-\alpha}, \tag{2.11}$$

here ι is a length-scale parameter, α is a scale mixture parameter, and $D(\cdot)$ denotes the distance between two sample points.

2.5 Performance Measures

It is critical to measure the correlation between the extracted anthropometric features and body weight or BMI. Pearson's correlation coefficient (PCC) is employed for measuring the correlation. It is a measure of the linear correlation between two variables. It was developed by Karl Pearson in 1895 from a related idea introduced by Francis Galton [77]. Given two sets of data $\{a_1, ..., a_n\}$ and $\{b_1, ..., b_n\}$, the formula for PCC is:

$$PCC = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2}\sqrt{\sum_{i=1}^{n}(b_i - \bar{b})^2}}, \tag{2.12}$$

here PCC is a scalar value between -1 and 1. If $PCC < 0$, it shows a negative correlation between the two sets. If $PCC > 0$, it shows a positive correlation. When $PCC = 0$, it indicates that there is no correlation between the two sets. When PCC is close to -1 or 1, there is a very strong correlation.

We apply a hypothesis testing with a statistical significance measure [78]. A p-value is utilized to decide whether a significant correlation exists between the two sets of data. We can make a decision by:

- If the p-value is smaller than the significance level α, it can reject the null hypothesis (there is no correlation between the two sets).
- If the p-value is larger than the significance level α, it fails to reject the null hypothesis.

The significance level α can be set to, e.g., 0.001, 0.01 or 0.05. If the p-value is equal or smaller than the thresholds, it indicates a significant correlation between the two data sets.

In addition to the correlation, we measure the performance of the proposed approach for weight or BMI estimation. The recall is used to evaluate the classification. And mean absolute error (MAE), mean absolute percentage error (MAPE) and absolute percentage error (APE) are used to measure the regression results:

- Recall: It is a performance measure that quantifies the ability of the classifier to correctly classify the positive training instances (also true positive rate, sensitivity). It is computed as the number of corrected classification divided by the number of samples that should have been classified as this class.
- MAE: It is defined as the average of absolute error between the estimated values and the ground truth:
$$MAE = \frac{1}{N}\sum_{j=1}^{N}|\hat{r_j} - r_j|\,, \tag{2.13}$$
here $\hat{r_i}$ is the estimated value for $j-th$ sample, r_j is the ground truth for $j-th$ sample, and N is the total number of test samples.
- MAPE: It is the mean absolute percentage error, computed as:
$$MAPE = \frac{100}{N}\sum_{j=1}^{N}\left|\frac{\hat{r_j} - r_j}{r_j}\right|, \tag{2.14}$$

Table 2.2: Pearson's correlation between the extracted features and the BMI in different gender groups.

	Male		Female		Overall	
	n = 1334		n = 666		n = 2000	
	p−value	correlation	p−value	correlation	p−value	correlation
WTR	0.0000	0.1774	0.0078	0.1033	0.0000	0.1320
WHpR	0.0000	0.1771	0.0018	0.1301	0.0000	0.1371
WHdR	0.0000	0.3317	0.0000	0.2992	0.0000	0.3038
HpHdR	0.0000	0.2791	0.0000	0.2769	0.0000	0.2785
Area	0.0000	0.4082	0.0000	0.3219	0.0000	0.3873

where all variables in Eqn. (2.14) have the same meaning as in Eqn. (2.13). Considering the large BMI range (15 to 75) of the visual-body-to-BMI dataset, the absolute percentage error can be another useful measure for the performance of BMI prediction from single images. For example, two individuals with the same height, one's BMI is 20 and another is 40. If they both have their BMI increase by 2, such a change is more obvious on the individual with BMI = 20. MAPE measures the error by taking the BMI as the base. APE is calculated by a single estimated value and ground-truth. It is a relative error.

2.6 Experiments

In this section, we explore the feasibility of analyzing body weight from 2D body images. We first examine the correlation between the extracted anthropometric features and the BMI values and then perform three estimation experiments using the extracted features.

The visual-body-to-BMI dataset is randomly split into training and test sets. The training set contains 2000 subjects (4000 images) of 1334 males and 666 females. The test set contains 950 subjects (1900 images): 650 males and 300 females. There is no overlap of subjects between the training and test sets.

2.6.1 Correlations between Body Features and BMI Values

According to the hypothesis test, we can measure whether the extracted features and BMI values are correlated. Here we assume the correlation with p−value < 0.01 is a significant correlation, and vice versa.

We measure the correlation between the extracted anthropometric features and BMI values on the training set. The results are shown in Table 2.2. $p-value = 0.0000$ indicates that the value which is smaller than 0.0001. From Table 2.1 we can see that the feature *Area* shows a higher correlation with BMI than other features. The correlation is a little lower in the female group than that in the male group, which may be caused by the different dress styles or body fat distribution between females and males. The dress and other loose clothes bring negative influence on the extracted features. The correlation coefficients of WTR and WHpR are a little lower than the other three features. Velardo et. al [22] reported an average correlation coefficient of 0.27 for BMI and waist to thigh ratio and Vazquez et. al [21] reported a correlation coefficient of 0.34 for BMI and waist to hip ratio. These correlation coefficients reported in health studies were calculated from the precise body size measurements (in person). And considering the various clothes styles that exist in this dataset may bring a negative influence to calculation, the correlation coefficients of these two features obtained in Table 2.1 are acceptable. Given the significant correlation (most $p-values < 0.0001$) between the anthropometric features and BMI, we draw the conclusion that the extracted features are correlated with the BMI values. Thereby, it is reasonable to estimate BMI values using the extracted features.

2.6.2 Recognize Weight Difference from A Pair of Images

The proposed approach takes either a pair of body images or a single body image as the input. For the pairwise images, the approach performs a three-class classification which decides the subject in the pairwise images as weight increased, decreased or keeping the same. Furthermore, we estimate how much the BMI difference between the pairwise images is.

1) Three-class classification: The approach can process a three-class classification $\{0,1,-1\}$ for a pair of images from the same subjects. We use the features calculated by Eqn. (2.9) in Section 2.3.2 to train a multi-SVMs which contains 3 binary classifiers for this task. Gaussian Radial Basis Function (RBF) is utilized for the SVM kernel. RBF achieves a better performance in classification than other kernels.

The recall of weight difference classification is given in Table 2.3. Taking into account the different body fat rate between males and females, the recall is measured for each gender group. It is seen that the accuracy for class *0* (keep the same weight) is much lower than the other two classes: *1* (weight increase) and *-1* (weight decrease). The

Table 2.3: Recall of triple classification from the pair-wise images.

	Recall (%)		
Class	Male	Female	Overall
0	63.6	40.0	56.3
1	81.0	89.2	83.6
-1	77.3	88.0	81.1

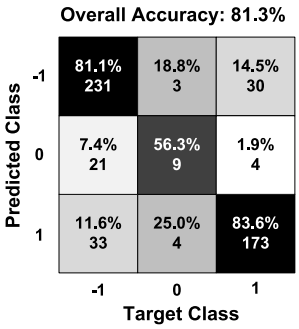


Figure 2.7: Confusion matrix of weight difference classification results, the diagonal cells show the number and percentage of correct classifications by the method.

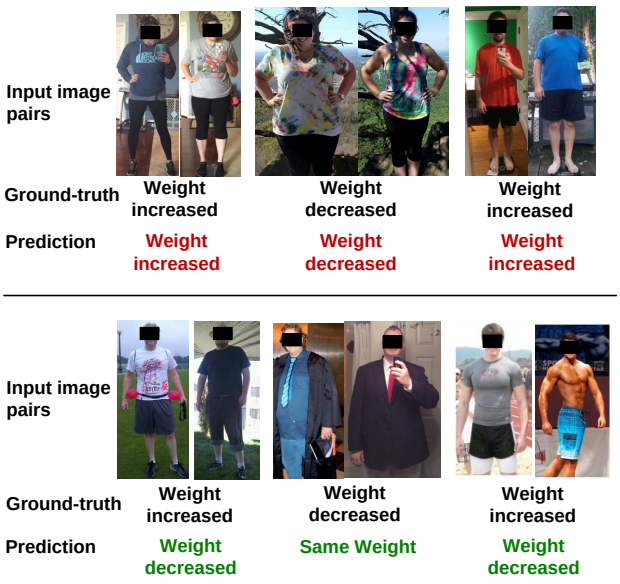


Figure 2.8: Some results of the weight difference classification. The upper panel shows good cases, and the lower panel shows failure cases. The BMI difference is from the left one to the right one.

Table 2.4: The MAEs and standard deviations of the estimated BMI differences using SVR and GPR models.

	MAE			Std		
Model	Male	Female	Overall	Male	Female	Overall
SVR	3.6	4.1	3.8	3.6	3.5	3.6
GPR	3.6	4.0	3.7	3.4	3.5	3.5

reason may be that the number of subjects in class *0* (481) is much less than the other two classes ($1246 + 1223$). There is an uneven distribution among the three classes. Fig. 2.7 shows the confusion matrix of weight change classification. The accuracy of weight increased pairs is 83.6%, and the accuracy of weight decreased pairs is 81.1%. They are both within the acceptable range. Fig. 2.8 shows some examples of the classification. The upper panel shows some good cases, while the lower panel shows some failure cases. Failure cases are observed due to the interference, occlusion of the body, or large body pose, etc.

2) How big is the weight difference? Further exploration is to discover how big the weight or BMI change between pairwise images is. The features computed by Eqn. (2.9) are used to train the regression model. Here we employ the SVR (with the RBF kernel) and GPR models. Table 2.4 shows the MAEs and standard deviations of the estimated BMI differences by the two regression models. We can see that the GPR model performs slightly better than the SVR model. Fig. 2.9 depicts the comparison of MAEs between the SVR and the GPR broken down by the absolute BMI differences. The differences between SVR and GPR for all ranges are less than 1 except for the range of $0.5 - 5.5$ (approximately 1.16). The MAEs in the absolute BMI difference range > 15.5 are relatively higher than other ranges. This may be caused by the small number of subjects (about 7.90%) with BMI differences larger than 15.5. The distribution of BMI differences in the visual-body-to-BMI dataset is given as: 492 subjects are in the range of BMI difference < 0.5, 921 are between 0.5 and 5.5, 866 are between 5.5 and 10.5, 438 are between 10.5 and 15.5 and 233 are in the range of BMI difference > 15.5. The proposed approach shows effectiveness in predicting how big the weight or BMI change is from a pair of body images.

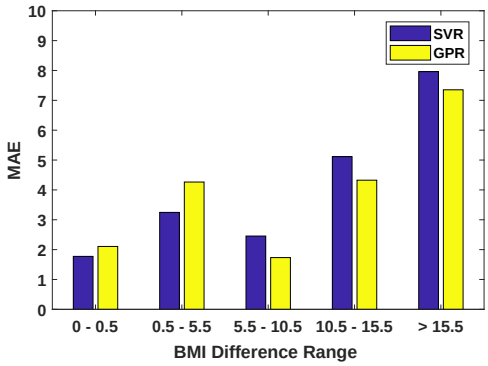


Figure 2.9: Comparison of MAEs between SVR and GPR broken down by the absolute BMI differences.

2.6.3 Estimate BMI from A Single Image

Now we study the BMI estimation from single images, by the SVR and GPR models for regression. Different from the previous, we use the anthropometric features $\mathbf{f}$ (calculated by Eqn. 2.9) extracted from every single image for BMI estimation.

The MAEs and MAPEs of the estimated BMI values by two regression models are given in Tables 2.5 and 2.6, respectively. The overall MAEs of the predicted BMI is between 3 and 4, the range of BMI values in the dataset is among 15 to 75, as shown in Fig. 2.2. The error of BMI estimation is relatively small compared to the large range of BMI values in the dataset. Fig. 2.10 shows the MAEs and MAPEs between SVR and GPR in different BMI categories: underweight, normal, overweight and obese. We can see that the two regression methods perform better in the normal category. Though the MAEs in obese category is between 5 and 6.5, taking into account the large range of BMI distribution in the obese category (30 to 75), the MAPEs of this category are acceptable. To compare the ground-truth BMIs with the estimations, a scatter plot based on the SVR results is shown in Fig. 2.11. The red dash-dot line shows where the two values are the same. The two green lines show where the absolute differences between the two values are 5. It is shown that points mainly distribute around the red line. Most outliers have the ground truth BMI values larger than 55. It can be seen that the proposed method tends to have a bias with an overestimation for low BMIs (BMI values between 20 and 30) and have an underestimation of high BMIs (BMI values larger than 35).

Figure 2.12 shows some examples of prediction. The absolute percentage error ($APE = \left|\frac{\hat{r_j} - r_j}{r_j}\right|$) is calculated for each case. Some failure cases are caused by ambiguous boundaries between the foreground and background, image blur, or large body pose. A

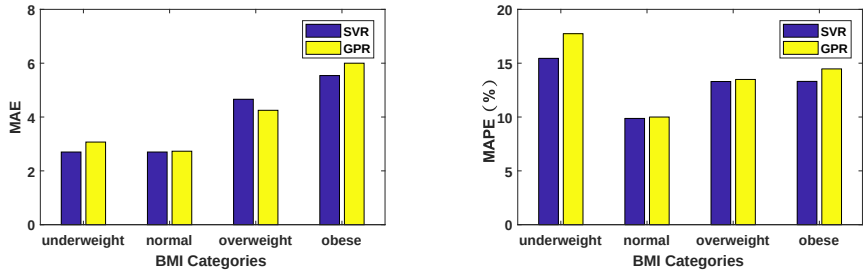


Figure 2.10: Comparison of MAEs and MAPEs between SVR and GPR in different BMI categories: underweight (BMI $\leq$ 18.5), normal (18.5 $<$ BMI $\leq$ 25), overweight (25 $<$ BMI $\leq$ 30) and obese (BMI $>$ 30).

Table 2.5: The MAEs and standard deviations of predicted BMI in different gender groups using SVR and GPR models.

	MAE			Std		
Model	Male	Female	Overall	Male	Female	Overall
SVR	3.4	4.5	3.8	3.3	4.8	3.6
GPR	3.5	4.4	3.9	3.5	4.0	3.7

Table 2.6: MAPEs of predicted BMI in different gender groups using SVR and GPR models.

Model	Male	Female	Overall
SVR	11.3%	15.0%	12.5%
GPR	12.1%	15.2%	13.1%

detailed discussion about estimation errors and failure cases will be given in Section 2.6.5.

2.6.4 Comparison with Other Methods

To the best of our knowledge, there is no previous approach that can estimate the BMI values from 2D body images only. Thereby, we compare with two methods that predict BMI values from face images. One is a geometric feature based method [47] and another is a VGG-face feature based method [61]. They are denoted as PIGF (psychology inspired geometric feature) and VGG features, respectively. These two methods both require clear frontal face images as the input, while some images in visual-body-to-BMI dataset do not meet this requirement. For a fair comparison, we select 2000 images which contain the clear frontal view face and then crop the face images. The 2000 images are

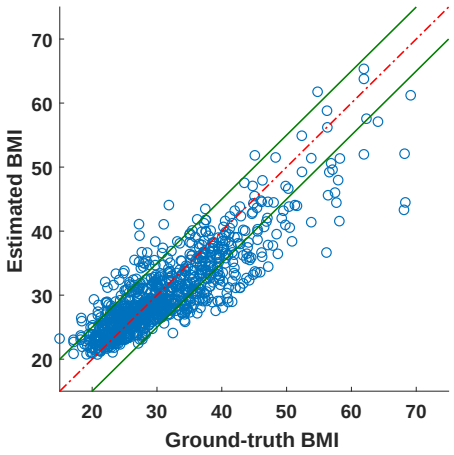


Figure 2.11: Scatter plot of the ground-truth BMIs over the estimated BMIs based on SVR model. The red dash-dot line shows where the two values are same. The two green lines show where the absolute differences of the two values are 5.

Figure 2.12: Examples of estimating BMI from a single body image.

split into training and test sets, which contains 1500 and 500 images, respectively. The input of our approach is a single body image, and the input of the other two methods is a face image cropped from the same body image. The comparison of the results is shown in Table 2.7. It can be seen that the proposed method outperforms the PIGF and VGG-face feature based methods in most cases, except on the male set. Moreover, the proposed method does not require a clear frontal view face image as input, which is useful for more general applications.

Furthermore, considering the features learned in deep neural networks (DNN) are demonstrated to be transferable and effective when used in other visual recognition tasks [79], we compare our anthropometric features with that the deep features. In this experiment, we employ the VGG-Net [80] model which is pre-trained on ImageNet

Table 2.7: Comparison of BMI estimation between our method and other methods.

	MAE			MAPE (%)		
Method	Male	Female	Overall	Male	Female	Overall
PIGF	4.61	4.58	4.60	15.50	15.64	15.54
VGG feature	3.72	4.48	3.94	12.69	16.04	13.66
Ours	3.74	4.16	3.86	12.68	14.15	13.11

Table 2.8: Results of BMI estimation from our anthropometric features and the VGG-Net feature.

	MAE			MAPE (%)		
Feature	Male	Female	Overall	Male	Female	Overall
VGG-Net	4.65	5.55	4.94	15.6	17.8	16.3
Ours	3.41	4.52	3.76	11.3	15.0	12.5

database [81] to extract the deep feature. Then an SVR model is trained based on the extracted deep feature. The feature from the $fc6$ layer is extracted for each body image in the training and test sets. VGG-Net takes an image of size 224 $\times$ 224 with the average image subtracted as the input. To normalize the images in visual-body-to-BMI dataset to a common size, we apply zero-padding to the images, and then resize them to 224 $\times$ 224. The training and test sets in this experiment are the same as the experiment in Section VII-C. Table 2.8 presents the results obtained based on the two features. It can be seen that our anthropometric features outperform the VGG-Net feature significantly.

2.6.5 Discussion

We analyze the errors in feature extraction and regression. The statistical analysis will be given, discussing whether the errors are acceptable for the application of BMI estimation from a single image. Finally, some failure cases are shown, and we analyze the influencing factors for the proposed method and possible reasons for the failure cases.

For feature extraction and regression, the widths of head, waist, hip and thigh are estimated from the 2D body images, and used to calculate the four anthropometric features (WTR, $WHpR$, $WHdR$, $HpHdR$). To analyze the error, we randomly selected 300 images from the dataset and manually labeled the widths of the head, waist, hip, and thigh for each image. Then the labeled widths are used as the ground truth values (v) to calculate the relative error (ε) of the estimated values ($\hat{v}$) by: $\varepsilon = \frac{|v-\hat{v}|}{v}$. The mean

Table 2.9: The mean relative errors of the extracted features.

	Head	Waist	Hip	Thigh
Error	2.1%	5.4%	4.7%	9.7%

Table 2.10: The accuracy of predicted category.

	Underweight	Normal	Overweight	Obese
Accuracy	11.1%	78.3%	64.2%	81.0%

relative errors of the extracted widths are shown in Table 2.9. The four errors are within a relatively low range. Since it is hard to label the area between waist and hip, where the relative error of estimated *Area* is not given.

To demonstrate whether the errors are acceptable for BMI estimation from a single body image, we further calculate the accuracy of the predicted category. According to the estimated BMI values, we can classify the body belong to which BMI category (underweight, normal, overweight and obese). The accuracy of the predicted category is the proportion of the total number of predictions that are correct. This measurement is helpful to decide if the errors are acceptable. For example, given a body image with a ground truth BMI value of 24.5, the estimated value is 20. Though the absolute error is 4.5 which is larger than the MAE (3.8), the predicted category (normal) is correct. On the other hand, this measurement has a limitation. For example, if the ground truth BMI is 25 and the estimated value is 25.5, though the absolute error is 0.5, the predicted category (overweight) is not correct. Considering the advantage and limitation of this measure, we combine the accuracies of the predicted category (as shown in Table 2.9) with the MAEs of predicted BMIs (shown in Fig. 2.10) to evaluate the performance. All predicted results shown in Table 2.10 are based on the SVR method. As it can be observed from Table 2.9 and Fig. 2.10, the prediction accuracy and MAE of the obese category are 81% and 5.5, respectively. Taking into account the large range of BMI on the obese category (30 to 75), the error of the obese category is reasonable. The prediction accuracy and MAE of the overweight category are 64.2% and 4.6, respectively. The performance in the overweight category is a little lower than the obese. The prediction accuracy of the underweight category is the lowest since there are only 46 body images in the database belong to the underweight category, among them, 9 are in the test set and 37 are in training set. The lack of enough underweight body images in the training set could be the reason for this lower performance.

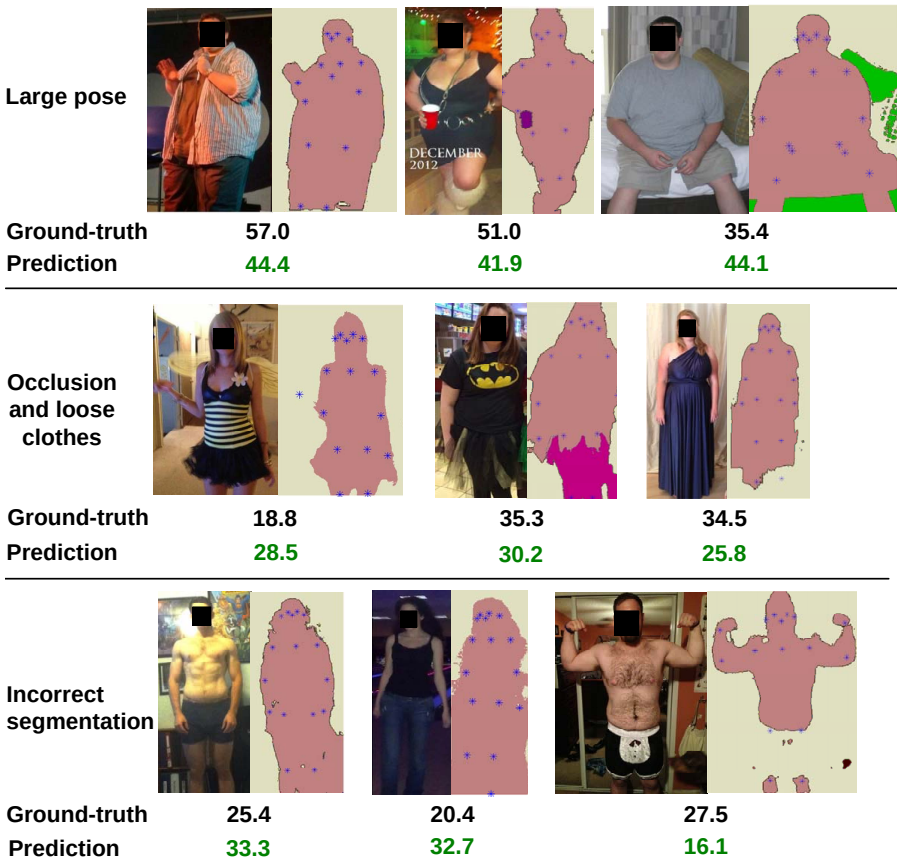


Figure 2.13: Examples of failure cases with the corresponding body contour and skeleton-joints detection. The upper panel shows the cases with the large pose. The middle panel shows the cases with body occlusion or loose clothes. The lower panels show the incorrect segmentation cases.

To analyze influencing factors (such as pose, occlusion, loose clothes, and scale) for the proposed method, and the reasons (such as incorrect body contour) of failure, Fig. 2.13 shows some failure cases with the detected body contour and skeleton joints. Most images in the dataset are frontal view body images with limited pose changes. Since there is no annotation about body pose, it is difficult to conduct an experiment to evaluate the performance with regard to pose changes. Theoretically, the extracted anthropometric features can tolerate small pose changes. The estimation may be significantly influenced if the input is a profile view image or with a large pose. The upper panel of Fig. 2.13 shows three failure cases with different poses. Though the detected body contour and skeleton joints are correct, the absolute errors are large. The occlusion always brings

negative influences to the method, decreasing the accuracy of body contour detection. Loose clothes are another negative factor to influence the real body shape. Three cases with large body occlusions and loose clothes are shown in the middle panel of Fig. 2.13. Because all the extracted anthropometric features are relative values (see details in Section 2.3.2), the scale changes in the image will not impact the method. The lower panel of Fig. 2.13 shows three failure cases caused by inaccurate contour. The incorrect or inaccurate body contour direct influences the accuracy of the extracted features. The failure of contour detection could be caused by image blurs, ambiguous boundaries between the foreground and background, etc. The proposed method can be further improved by employing more accurate body contour detection algorithms.

2.7 Summary

This chapter raises a new topic of study: estimating the BMI values from 2D body images. We investigate the relation between body weight and visual body appearance and estimate the BMI values from 2D body images. Correlation is analyzed between the extracted anthropometric features and BMI values, which validates the usability of the selected features. More specifically, body weight analysis is studied at three different levels of difficulties: the weight change classification is first investigated from a pair of body images of the same subjects; further investigation is conducted to estimate how big the weight change between the pairwise images is; the last is to predict the BMI value from a single body image. To address the visual body weight analysis problem, the computational method of five anthropometric features is developed. And a new visual-body-to-BMI image dataset has been collected and cleaned to facilitate this study. The errors of the three estimation tasks evaluated by several measurements are within acceptable ranges. Comparing with the facial images analysis approaches, the proposed method performs better in most cases. Furthermore, our anthropometric features significantly outperform the VGG-Net feature on BMI estimation. Based on all experimental results, it is promising to analyze body weight or BMI from the 2D body images visually. In the future, we will combine body images with face images to improve the BMI prediction, and will explore the DNN-based method to address this visual body weight analysis problem.

Chapter 3

Body Mass Index Estimation from Dressed People in 3D Space

This chapter is dedicated to studying BMI estimation from the 3D visual data. The proposed BMI computation approach consists of body weight and height estimation from normally dressed people in 3D space. To address the influence of loose clothes on body volume estimation, two clothes models are developed to make the volume estimation more accurate. A new RGB-D video dataset is collected for this study, and the reconstructed 3D data are provided by the KinectFusion on the depth data. Experimental results show the effectiveness of the approach to work on normal conditions of dressed people.

The outline of this chapter is as follows. Section 3.2 describes the method developed for weight and height estimation, and the two clothes models for correcting the estimated body weight. Section 3.3 describes the RGB-D video dataset newly collected for this work. The measure metrics used to evaluate the performance of the estimated BMIs are introduced in Section 3.4. The experimental results and analysis are presented in Section 3.5. Finally, some conclusions are drawn in Section 3.6.

3.1 Problem Definition

Body mass index (BMI) is an important soft biometric measure that is related to people's daily lives. Given an individual's height and weight, BMI= $\frac{weight(lb)}{height(in)^2} \times 703$. BMI is an important visual characteristic to describe a person. It is widely used for measuring the adiposity, especially for the overweight issue [14]. In medical science, both BMI and body weight can be used to estimate the risk for some diseases, such as breast and endometrial cancers [11, 82]. Currently, computer vision has been a favored approach

for providing new techniques to automatically detect various diseases [83,84]. Considering the inconvenience of measuring BMI with special devices, exploring an automatic BMI estimation method from visual images/3D data could make it efficient to monitor the health conditions in a large scale setting.

The standard approach to collect 3D data of the human body involves laser scanning and RBG-D sensor fusion [85]. The commercial laser scanning systems cost approximately from $35,000 to $500,000 dollars. A Kinect sensor [86] contains a depth sensor, a color camera, and a four-microphone array. With the depth data streamed from the sensor, a global surface model of the observed scene can be generated in real-time by the Kinect fusion technique [87]. The cost of collecting 3D data by Kinect sensors is much lower than the traditional laser systems. And Kinect 3D data can be generated in real-time by the KinectFusion algorithm, the time cost is also much lower. Thereby, the Kinect 3D data have more general applications in the real world than laser scanning data.

Considering the advantages of the Kinect 3D data, this work analyzes BMI from 3D data of the human body collected by the Kinect sensor. We propose to estimate the body weight and height of dressed people from the Kinect sensor, and compute the BMI based on the two estimations. The proposed weight estimation approach is based on the volume estimation. RGB-D videos are collected to facilitate this study. The condition for collecting the data is simple. Though the noise may be complicated with the scanning and fusion process, the proposed weight estimation approach includes clustering and fitting stages to suppress such noise.

The significance of this study comes from several aspects. First, this work provides a non-contact way by using affordable Kinect devices for accessing BMI and body weight. It can be used as a convenient self-monitoring tool or telemedical equipment for users rather than asking them to find scales and metric tapes to measure their body weight and height. Second, this approach, as well as data acquisition, may give more opportunities in real applications. As the first step, we focus on BMI estimation from 3D data, which is a frequently used index parameter in reality. However, BMI estimation is not the only goal. Given the 3D data capture, more properties can be examined in addition to BMI. For instance, in smart health, one can assess the body volume, body shape, etc. to get a more accurate estimation of a person's health condition; In E-commence, one can assess the 3D body in whole for clothes selection with different fashions. Third, BMI and body weight are soft biometric traits that can be utilized as auxiliary information for recognition or tracking of persons. Currently, it is still a challenge to visually extract BMI and weight

from visual 2D or 3D data. Thereby, it is of great importance to estimate weight or BMI from 3D data.

In the past few years, 3D related techniques rapidly developed on various applications [88, 89]. There are some studies estimating the body shape from clothing 3D body data. Bualan et al. [90] proposed a method to estimate the detailed 3D shape of a person from images of that person wearing clothing by a shape model that is independent of the body pose. Hasler et al. [91] proposed a method to estimate the detailed 3D body shape of a person even if heavy or loose clothing is worn by fitting the statistical model. Zhang et al. [92] presented an approach to recover a personalized shape of the person under clothing from a sequence of 3D scans.

There are a few works analyzed body weight from 3D body data. Velardo et al. [37] studied the weight estimation from 3D human body data by analyzing the anthropometric features. Pfitzner et al. [33] presents a method for estimating body weight from RGB-D body data and thermal data of lying people in the clinical environment. First, anthropometric features are extracted from the frontal view RGB-D data. Thermal data is used to ease the segmentation of a person from the background. Then the features are forwarded to an artificial neural network for weight estimation. Later on, Pfitzner et al. [34] extended their previous work [33] by adding two more scenarios: standing and walking people. There are two limitations of the above three work. First, [37] estimate BMI from the extracted anthropometric features. Second, [33] and [34] reply on thermal data to accomplish body segmentation. Considering these limitations, a new method which directly estimates BMI from people in 3D space is desired.

Different from all the above studies, we explore a new application for BMI estimation from 3D body volume. Though in [91, 92] some simple biometric features, such as height, arm length, leg length, etc., can be computed directly on the estimated 3D body shape model, the body weight cannot be computed simply. They employed another independent step (statistical model fitting) to compute the body weight. While our method is to measure from real 3D data.

The main contributions include:

- A new RGB-D video dataset is collected which comprises human body 3D data, color video and depth video, using the Kinect sensor for 163 subjects, along with the corresponding gender, age, weight and height labels.
- An efficient weight estimation approach is developed to work on normally dressed people in 3D space.

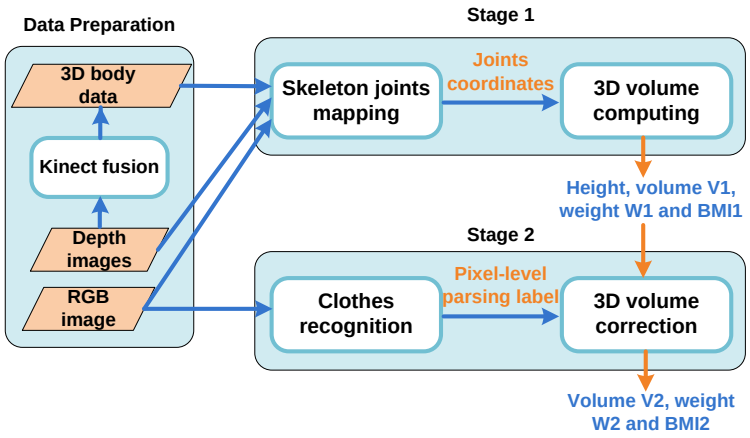


Figure 3.1: The framework of our BMI estimation approach.

- Two clothes models are proposed to reduce the negative influence of loose clothes on body volume and weight estimation.

3.2 Method

The framework of our proposed BMI estimation approach is shown in Fig. 3.1. The input data includes the depth and color image streams, and the reconstructed 3D data of the human body provided by the KinectFusion. The BMI estimation approach consists of four main steps: skeleton joints mapping (Section 3.2.2), 3D volume computing (Section 3.2.3), clothes recognition (Section 3.2.4), and 3D volume correction by clothes models (Section 3.2.5). There are two stages in this approach. Both height and weight are estimated in stage 1, then the estimated weight (volume) is corrected in stage 2. BMI is calculated from the estimated weight and height. As shown in Fig. 3.1, weight W_1 is estimated directly from the 3D volume of the body. We multiply the volume of the body by the body density to obtain the body weight. Weight W_2 is the outcome of applying 3D volume correction to V_1. The clothes model is selected based on the output of clothes recognition.

3.2.1 KinectFusion

3D body volume reconstruction is provided by the KinectFusion, which reconstructs a single dense surface model smoothly by integrating the depth data from multiple viewpoints continuously [87]. In this work, we do not focus on 3D volume reconstruction but rather utilizing the KinectFusion. The output of the fusion is a 3D body volume reconstruction in the format of a voxel grid.

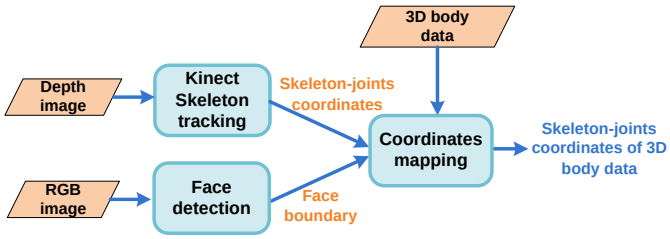


Figure 3.2: The pipeline of skeleton joints coordinates mapping. The input is depth image, color image and 3D body data. The output is the skeleton-joints coordinates located in the 3D body data.

3.2.2 Skeleton joints mapping

Skeleton joints mapping is a preprocessing step for 3D volume estimation. Since two different fitting methods (see details in Section 3.2.3) are applied to various body parts of the 3D data, the skeleton-joints coordinates of the body are used to mark different body parts. The skeleton-joints coordinates computed from the depth images need to be mapped to 3D data results.

First, skeleton tracking is applied to the depth video, and then the skeleton-joints coordinates are computed for each frame of the depth video. One frame that contains a frontal face image in depth video and its corresponding frame in color video is selected as the input depth and color images. From the skeleton-joints coordinates of the chosen depth image, the approximated coordinates of a person's feet can be obtained. Then a face detector is applied to the chosen color image and the bounding box of the head is obtained. The coordinates of the head are estimated from the location of the head bounding box. Since the color image and depth image are aligned, the coordinates of them can be mapped simply. Now the approximated coordinates of head and feet in color and depth images are obtained. Then we automatically locate the head and feet on 3D data and obtain the head and feet coordinates accordingly. The height of the body is computed from the coordinates of the head and feet in 3D data. Note that such height estimation method works within scenario where people are standing straight up. Using the coordinates of head and feet obtained on color/depth images and 3D data, we can find a linear mapping from color/depth image to 3D data. Thereby the skeleton-joints coordinates in the 3D data of the full body can be obtained. Fig. 3.2 shows the processing for this mapping.

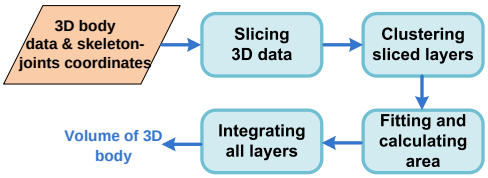


Figure 3.3: Block diagram of the 3D volume calculation process applied to the 3D data after the KinectFusion.

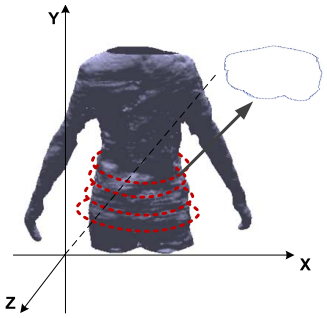


Figure 3.4: The planar view of a single slice from the 3D volume.

3.2.3 3D volume computing

To estimate the body weight, the 3D volume is estimated first and then the body weight is calculated by multiplying the volume by the body density. The volume is calculated by integrating the volume of sliced layers that are parallel to the x-z planer, as shown in Fig. 3.4. The 3D data is sliced into many thin layers (the thickness of each layer is set to 0.5cm), and each layer is considered as a cylinder. Then the area of each sliced layer is calculated, and the volume of the cylinder is estimated by multiplying the area by the height of the layer. Finally, the volumes of all these layers (cylinders) are added together to get the whole volume of the 3D body. Fig. 3.3 depicts the volume calculation process applied to the 3D data from the KinectFusion. And Fig. 3.4 shows the planar view of a single slice of the 3D body. In the following part, two methods (clustering and fitting) employed for volume calculation are described.

Clustering

After slicing the 3D data horizontally, the numbers of independent sections on the sliced layers depend on the specific part of the body. For instance, when slicing around the knee, there are two independent sections (two knees) shown on the sliced layer; when slicing around the waist, there are three sections (waist and two arms) on the layer as shown in Fig. 3.5. The skeleton-joints coordinates of 3D data (the output in Fig. 3.2)

Figure 3.5: The left panel is a sliced layer around the waist, there are three independent sections (waist and two arms) on this layer. The right panel is the outcome after being applied DBSCAN. The noisy data which lays inside and outside of the biggest circle (waist) are all removed after clustering. The data are clustered into 3 groups (drawn by blue, red and green, respectively).

are used to identify different body parts.

Meanwhile, the reconstructed 3D data by KinectFusion contains some noise. As shown in the left panel of Fig. 3.5, there are some noisy points which are inside and outside of the biggest circle (waist section). To do denoising in the data and identify the number of independent sections on the layer, we apply clustering to the 3D points in each sliced layer. Density-based spatial clustering for applications with noise (DBSCAN) [93] is employed for clustering. It is a density-based clustering algorithm, designed to discover clusters of arbitrary shape as well as to distinguish noise in the spatial data. There are two main parameters for this algorithm, the maximum radius of the neighborhood from the core point (ϵ) and the minimum number of points required to form a dense region ($MinPts$). In this work, we set $\epsilon = 0.02$ and $MinPts$=10. Fig. 3.5 shows the result of applying DBSCAN to the sliced layer which is around the waist. We can see the noisy data inside and outside of the largest section are all removed, and the points are clustered into 3 groups.

Fitting

After applying the clustering to the data of the sliced layers, fitting is applied next. With the parameters obtained from fitting, the area of the sections on the sliced layer can be calculated. Then using this area to calculate the volume of a sliced layer (cylinder). The shape of the sections on the sliced layers can vary if they are from different body parts. In this work, two fitting methods are utilized. The shape of the sliced section of head, neck, arm, leg and hip can be approximated as the ellipses. For the other parts, since people wear casual clothes during the data collection, the shape of these sliced sections can be very different from an ellipse, as shown in Fig. 3.7. To address this

Table 3.1: Fitting methods applied to the corresponding body parts.

Fitting method	Body parts
Ellipse	Head, neck, arm, leg, hip
B-spline	Shoulder, chest, waist

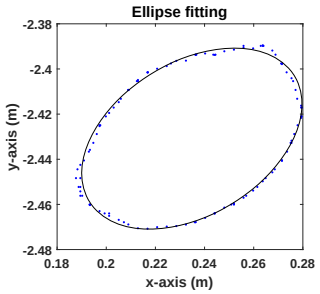


Figure 3.6: Applying ellipse fitting to the data of a sliced section.

issue, he B-spline fitting [94, 95] is applied to these data. Table 3.1 shows the fitting methods applied to the corresponding body parts. During the data collection stage, all participants keep a similar specific pose (see details in Section 3.3), thereby different parts of the reconstructed 3D body data can be divided by the skeleton joints obtained from the previous stage: skeleton joints mapping (Section 3.2.2).

The ellipse fitting [96] is done by minimizing the least squares error: $E = \sum_{i=1}^{n}(ax_i^2 + bx_iy_i + cy_i^2 - \rho^2)^2$. Fig. 3.6 shows the result of a sliced section of elbow being processed by an ellipse fitting. After the fitting, the parameters of the fitted functions are used to calculate the area of the ellipse, which is the estimated area of the section.

For B-spline fitting, given the unorganized data points, representing a target shape $X_k, k = 1, 2, ..., n$, the control points $P_i, i = 1, 2, ..., m$ are estimated to minimize the following objective function:

$$f = \frac{1}{2}\sum_{k=1}^{n} \|P(t), X_k\| + \lambda f_s, \tag{3.1}$$

f_s is a regularization term to ensure a smooth solution curve, and λ is a positive constant to modulate the weight of f_s.

These unorganized data points X_k are approximated by a closed or open planar B-spline curve:

$$P(t) = \sum_{i=1}^{m} B_i(t) \cdot P_i, \tag{3.2}$$

where $B_i(t)$ are the B-spline basis functions of a fixed order and knots, and P_i are the control points. Since in Eqn. (3.1), f is a nonlinear objective function, an iterative minimization is performed. Suppose that given a specific B-spline curve $P_c(t) = \sum_{i=1}^{m} B_i(t) \cdot P_{c,i}$

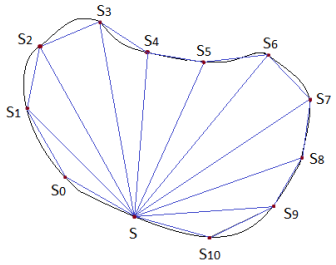


Figure 3.7: The area of a sliced section is divided into a collection of triangles.

with control points $P = (P_{c,1}, P_{c,2}, ..., P_{c,m})$, which can be an initial fitting curve, or the current fitting generated from the last iteration. Let $D = (D_1, D_2, ..., D_m)$ be the variable updates to P_c to give the new control points $P_+ = P_c + D$, then $P_+(t) = \sum_{i=1}^{m} B_i(t)(P_{c,i} + D)$ is the B-spline curve with updated control points P_+.

After applying the B-spline curve fitting, the following steps are processed to estimate the area of the section, as shown in Fig. 3.7. First, a number of samples $(s_0, s_1, ..., s_n)$ are selected on the curve. Then, choosing a random point s on the plane. As shown in Fig. 3.7, this collection of points define a sequence of triangles of the form (s, s_i, s_{i-1}), with $i = 1, 2, ..., n$. The sum of all these triangles is equal to the area inside the curve, subtracting the sum E of the small residuals e_i as $E = \sum_{i=1}^{n} e_i$, here e_i is defined by the line segment between points s_i and s_{i-1}. With the increasing number of samples s_i, there are more triangles and smaller residuals e_i, so the sum of the areas of these triangles gets closer to the true area within the curve.

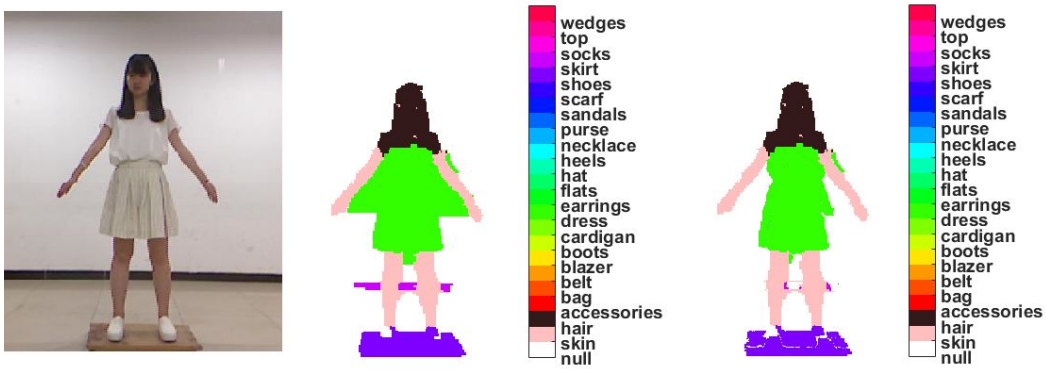


Figure 3.8: The left panel is the input color image. The middle and right panels show a comparison of the pixel-level labeling based on the clothes parsing before and after the mask correction computed from the depth image.

3.2.4 Clothes recognition

When the weight W_1 is calculated directly from the estimated 3D volume of the body, loose clothes that people wear are the main reason for the error in estimating the true body volume. In order to eliminate or reduce such errors as much as possible, clothes recognition is applied to the color images and then the volume of the 3D data could be corrected by different clothes models. The clothes recognition method employed in this work is similar to the paper doll parsing work. The outcome is a pixel-level labeling result. Given an input image, a labeling tag such as null, skin, pant and skirt, etc. (null denotes the background of the image) is given for each pixel in the image.

However, the clothes parsing method [97,98] sometimes cannot perform well, especially when the color of clothes is similar to the background, as shown in Fig. 3.8. To address this issue, we combine the depth image with the result of clothes parsing to solve the confusion between background and foreground. A mask M is generated from the depth image to segment the background and foreground. For a depth image I, let $i \in I$ be a pixel, then the mask is given as:

$$M_i = 1[I_i > T], \tag{3.3}$$

where M_i denotes the mask for pixel i, I_i denotes the value of pixel i in the depth image (the higher the value is, the closer the object to the camera) and T is a threshold. $1[\cdot]$ is an indicator function. Then the correct results of clothes parsing can be obtained from: $parsing_{correct} = parsing \cdot M_i$. Fig. 3.8 shows a comparison of the pixel-level parsing result of clothes recognition before and after the correction by using the Mask M.

3.2.5 3D volume correction

In this stage, the corrected pixel-level clothes parsing result (Section 3.2.4) are mapped to the 3D data. The mapping method is similar to that in Section 3.2.2. For different clothes parsing results, different clothes models are developed to correct the estimated volumes. This work mainly focuses on correcting the volume estimation when wearing the dress/skirt (female) and shorts (male). The two proposed clothes models are presented: the dress (skirt) model and the shorts model.

Dress (skirt) model

We assume that the skirt/dress in 3D data can be considered as an elliptic truncated cone, as shown in Fig. 3.9. The following equation is used to calculate the volume of the

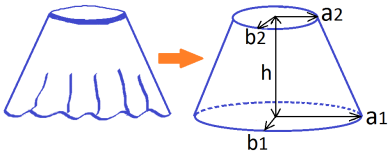


Figure 3.9: Dress model: the skirt/dress in 3D can be approximated by an elliptic truncated cone.

skirt/dress:

$$V_{dress} = \frac{\pi}{3} \cdot \frac{a_1^2 b_1 h - a_2^2 b_2 h}{a_1 - a_2}, \tag{3.4}$$

where a_i denotes the major radius and b_i denotes minor radius of an ellipse. Subscript 1 and 2 denote the upper base and lower base, respectively, as shown in Fig. 3.9. h is the height of the elliptic truncated cone (skirt). a_i and b_i can be estimated by ellipse fitting. h can be estimated from mapping pixel-level parsing result to the 3D data.

The next step is calculating the volume of legs covered by the dress/skirt. We also assume that this part of the leg (covered by the dress/skirt) can be approximately considered as an elliptic truncated cone. The volume $V_{dress}{}'$ need to be subtracted from V_1 as $V_{dress} - V_{legs}$, which can be obtained from the following equation:

$$V_{dress}{}' = \frac{\pi h}{3} \cdot \left[\frac{a_1^2 b_1 - a_2^2 b_2}{a_1 - a_2} - 2 \cdot \frac{a_4^2 b_4 - a_3^2 b_3}{a_4 - a_3} \right], \tag{3.5}$$

Here a_4 and b_4 are the radius of upper base of leg elliptic truncated cone, which are approximately equal to $0.5a_2$ and $0.5b_2$. Then a_3 and b_3 can be obtained by applying an ellipse fitting to the sliced layer around the bottom of the skirt in 3D. Then the corrected volume is: $V_2 = V_1 - V_{dress}{}'$, where V_1 is the volume estimated in stage 1 (Section 3.2.3).

Shorts model

The shorts worn by males are usually very loose, so we try to eliminate such a negative influence on weight estimation. A shorts model for the male is proposed to correct the estimated volume, as shown in Fig. 3.10. The solid lines represent the shape of shorts and the dash lines are the shape of legs. S_1 is the area of the section at the short's bottom, S_2 is the area of the leg section at the short's bottom. We assume that:

$$\frac{V_{legs}}{V_{shorts}} = \frac{S_2}{S_1}, \tag{3.6}$$

where V_{legs} is the volume of legs covered by shorts and V_{shorts} is the volume of shorts. Then the volume ($V_{short}{}'$) to be subtracted from V_1 can be derived from Eqn. (3.6):

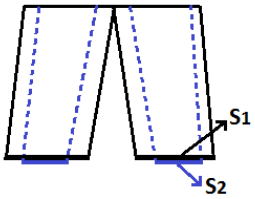


Figure 3.10: Short model for male: the solid lines represent the shape of shorts and dash lines are the shape of legs.

$V_{short}{}' = V_{short} \cdot (1 - \frac{S_2}{S_1})$. The following five steps summarize the processing to correct the volume of the 3D body based on the proposed shorts model:

- Map the pixel-level clothes parsing result to the location of the shorts (the coordinates of shorts in the color image), from color image to 3D data.
- Using the coordinates of shorts, calculate the volume of shorts V_{short} by integrating the sliced layers, and calculate the area S_1 and S_2 by applying ellipse fittings to the corresponding sliced layers.
- Obtain coefficient from: $f = 1 - \frac{S_2}{S_1} \cdot \alpha$. According to different ranges of $\frac{S_2}{S_1}$, α is empirically set to 1.5 in our implementation.
- The volume need to be subtracted: $V_{short}{}' = V_{short} \cdot f$.
- The correct volume is $V_2 = V_1 - V_{short}{}'$.

3.2.6 Weight estimation

The weight of the 3D body is estimated by multiplying the volume of the body by body density. The body density varies with ages, gender and anthropometric measurements. Several studies [99–101] employed multiple regression based methods to analyze body density. The regression models consist of different variables, such as skinfolds, body diameters and circumferences, etc. For instance, [99] used the following regression model to analyze body density:

$$D = 1.11847 - 0.00078V_5 - 0.00048V_{27}, \tag{3.7}$$

here V_5 is abdominal skinfold, V_{27} is abdomen circumference. Based on the regression results from a large amount of data, [101] summarized a query table based on age and gender for body density estimation, which is given in Table 3.2. In this work, the RGB-D

Table 3.2: The base body density varies with different age and gender.

	Density (kg/m^3)	
Age	Male	Female
17-19	1.066×10^3	1.040×10^3
20-29	1.064×10^3	1.034×10^3
30-39	1.046×10^3	1.025×10^3
40-49	1.043×10^3	1.020×10^3
50-59	1.036×10^3	1.013×10^3

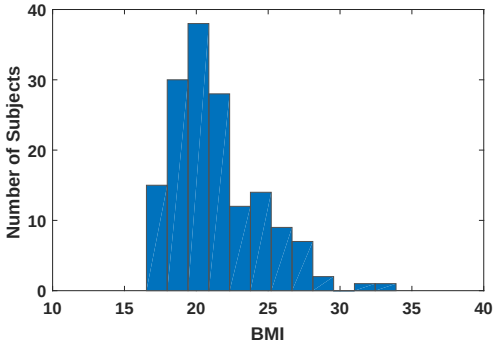


Figure 3.11: Distribution of BMI values on the dataset. The BMI values mainly distribute between 18 to 30.

dataset has the annotation of age and gender for each subject. Thereby, we choose the proper body density for each subject according to Table 3.2.

3.3 Dataset

We collected a new dataset, since there is no previous dataset appropriate for this study. The dataset consists of 163 subjects captured by the Kinect sensor, with color videos, depth videos and 3D fusion. Each subject has gender, age, weight and height information recorded. There are 70 males and 93 females. The ages of people are in the range of 16 to 52 years old. Fig. 3.11 shows the distribution of BMI values on the dataset. The mean BMI value of the database is 21.4, the standard deviation of the BMI values is 3.1.

The Kinect sensor features a 640×480 pixels RGB camera, and 640×480 pixels depth sensor. Each frame of the depth video is made up of pixels that contain the distance from the camera plane to the nearest object [86]. The reconstructed 3D data of each subject is obtained from the depth video by the KinectFusion [87]. The location of the Kinect

Table 3.3: Pearson's correlation cofficient *PCC* between weight/BMI and estimated volume in clothes groups.

	Dress/skirt		Short		Pant		T-shirt		Jacket/long sleeves	
	PCC	p-value	PCC	p-value	PCC	p-value	PCC	p-value	PCC	p-value
weight	0.6051	0.0324	0.7978	0.0000	0.8301	0.0000	0.8476	0.0000	0.7280	0.0262
BMI	0.5444	0.0954	0.7024	0.0000	0.7731	0.0000	0.7895	0.0000	0.6269	0.0708

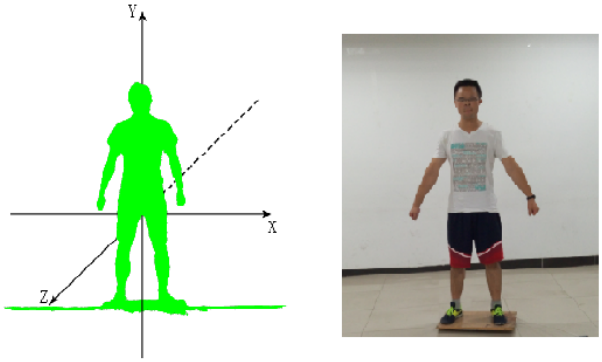


Figure 3.12: The left panel shows the measurement coordinate systems and the riht panel shows the guesture during data collection.

sensor is set as the origin of coordinates. The x-axis is along the horizontal, the y-axis is along the vertical and the z-axis represents the distance between the subject and the Kinect sensor. Fig. 3.12 shows the measurement coordinate system and the specific pose during data collection. The subject stands in front of the sensor and turns around for 360 degrees with a constant speed. Some samples of the 3D volume by the KinectFusion are shown in Fig. 3.13.

3.4 Performance Measure

It is critical to measure the correlation between the estimated body volume and body weights or BMIs. The Pearson's correlation coefficient (PCC) is employed for measuring the correlation. A hypothesis testing is used to test the significance of the obtained correlations. Based on the hypothesis testing, we can understand whether the estimated volumes and weights/BMIs (ground truth) are correlated. Pearson's correlation coefficient was developed by Karl Pearson in 1895 from a related idea introduced by Francis Galton [102]. Given two variables $X = x_1, ..., x_n$ and $Y = y_1, ..., y_n$, PCC is

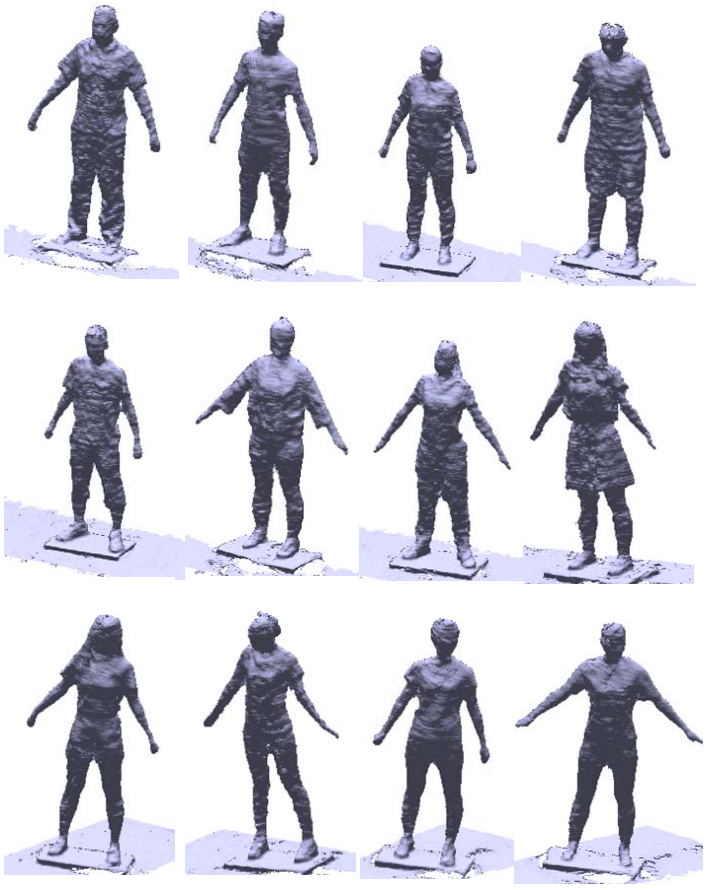


Figure 3.13: Samples of the under clothing 3D data reconstructed by KinectFusion.

given by:

$$PCC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{3.8}$$

where PCC is between -1 and 1. If $PCC < 0$, it indicates a negative correlation between X and Y. If $PCC > 0$, it indicates a positive correlation. If $PCC = 0$, it indicates no correlation between X and Y. When PCC is close to -1 or 1, it indicates a very strong correlation.

The above Pearson's correlation coefficient is computed from the observed/tested samples. To extend the correlation measure to a population, we need to do hypothesis testing with a statistical significance measure. The p-value is used to indicate whether a significant correlation exists between the estimated volume and weight (ground truth). The p-value uses the t-distribution. First, t is calculated by:

$$t = \frac{r(n-2)}{\sqrt{1-r^2}}, \tag{3.9}$$

where r is the correlation coefficient and n is the number of observations. Then computing

p-value from t uses the t-distribution function. The number of degrees of freedom is equal to $(n-2)$, and the tails is set to 2.

We can make a decision from the p-value by:

- If the p-value is smaller than the significance level α, it can reject the null hypothesis (there is no correlation between the two sets).
- If the p-value is larger than the significance level α, it fails to reject the null hypothesis.

Some significance level α can be set, e.g., 0.001, 0.01 or 0.05. If the p-value is equal to or smaller than the thresholds, it indicates a significant correlation between the estimated volume and weight.

In addition to correlation, the mean absolute error (MAE), percent error (ε) and are used to evaluate the performance of the proposed approach for BMI estimation. MAE is defined as the average of absolute error between the estimated value and the ground truth: $MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{b_i} - b_i\right|$, where $\hat{b_i}$ is the estimated BMI for $i-th$ sample, b_i is the ground truth for $i-th$ sample and N is the number of samples. This measure is motivated by its usage in age estimation, e.g. [103]. Percent error is defined as: $\varepsilon_i(\%) = \frac{\hat{w_i}-w_i}{w_i} \times 100$, where $\hat{w_i}$ is the estimated weight for $i-th$ sample, w_i is the ground truth for $i-th$ sample.

3.5 Experiments

In this section, experiments are conducted to validate the proposed approach to BMI estimation. We first examine the correlation between the estimated body volume and body weights or BMIs, and then perform the estimation experiments.

3.5.1 Correlation between the estimated volumes and true weights

BMI is calculated from the estimated weight and height. The weight is calculated directly from the volume of 3D data. The correlation between the estimated volumes and weights/BMIs (ground truth) is analyzed. PCC is used to measure the correlations. Based on the hypothesis testing, we can tell whether the estimated volumes and weights/BMIs are correlated or not. Therefore the performance of the BMI estimation approach can be verified.

Table 3.3 shows the correlations between the estimated volume and weight/BMI (ground truth) in five clothes groups. In this table, $p-value = 0.0000$ indicates a

Table 3.4: MAE of estimated height, weight and BMI in the two stages.

	Stage 1	Stage 2
Height (cm)	3.72	–
Weight (kg)	7.28	5.15
BMI	3.40	2.54

very small value, which is less than 0.0001. Here we consider the correlation with $p - value < 0.05$ as a significant correlation. From Table 3.3, we can see the correlation between the estimated volumes and the weights/BMIs are significant in short, pant and T-shirt groups. While the correlations are much lower in dress and jacket groups than in others, which means the dress and jacket brings a large negative influence on the volume estimation of the 3D body.

3.5.2 BMI estimation in two stages

As shown in Fig. 3.1, there are two stages for the BMI estimation. In stage 1, the height and weight are directly estimated from the 3D data. In stage 2, clothes recognition and clothes models (skirt and short models) are applied to correct the estimated weight obtained in stage 1. Table 3.4 shows MAEs of the estimated weight and BMI in two stages, respectively (height is only estimated in stage 1). We can see that the MAEs of the estimated weight and BMI reduced significantly in stage 2.

Table 3.5 shows MAEs of weight estimated from the data wearing dress/skirt (female) and shorts (male) in two stages. The first column is the category of data for testing. The second column is the number of subjects. The third and fourth columns are the MAEs of estimated weight in stage 1 and stage 2, respectively. Both MAEs are decreased significantly in stage 2. It shows the necessity and effectiveness of the two clothes models (dress/skirt and shorts) on volume correction and body weight estimation.

Table 3.5: MAE of estimated weight before and after applying clothes models to specified data. Stage 1 is before applying the clothes models while stage 2 is after applying the clothes models.

Data	Number of subjects	Stage 1 (W_1)	Stage 2 (W_2)
wearing skirt or dress	18	24.45	5.04
wearing shorts (male)	15	5.89	2.05

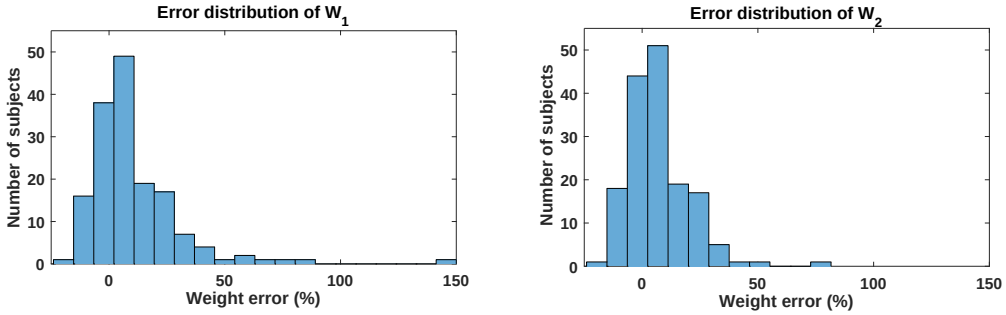


Figure 3.14: The upper panel is the error distribution of estimated weight of all data in stage 1. The lower panel is the error distribution of estimated weight in stage 2.

Table 3.6: Mean and standard deviation of weight percentage error in two stages.

	Stage 1	Stage 2
Mean of percentage errors	10.15	3.95
Standard deviation of percentage errors	2.45	1.28

Fig. 3.14 shows the error distribution of estimated weight from all subjects. The x-axis is the percent error of the estimated weight. The y-axis is the number of subjects whose percent errors are within the corresponding percent error interval. The upper panel in Fig. 3.14 is the error distribution of estimated weight over all subjects in stage 1. The lower panel is the error distribution of estimated weight in stage 2. The mean and standard deviation of percent errors in both stages are given in Table 3.6. We can see both the mean and standard deviation of percent errors become smaller in stage 2, which means there are more subjects with reduced errors. This again shows the effectiveness of the clothes models on weight estimation.

3.5.3 BMI estimation in separate clothes groups and gender groups

Based on the parsing results of clothes recognition, there are several groups of clothing styles. The performance of the BMI estimation approach for different clothes groups is evaluated. As shown in Table 3.7, there are five main clothes groups: dress/skirt, shorts, pants, T-shirt and jackets/long-sleeves. Since the weight estimation approach is based on estimating the volume of the 3D body, the errors of estimated weight vary with different styles of clothes. In Table 3.7, the first column is the clothes group, the second column is the number of subjects in each clothes group, the third and fourth columns are MAEs of weight and BMI estimation in each clothes group. We can see the MAEs of weight and BMI in shorts group are the smallest which means shorts have the lowest negative

Table 3.7: MAE in different clothes groups

Clothes group	# of subjects	MAE of weight	MAE of BMI
Dress/skirt	18	5.05	3.05
Shorts	43	4.01	2.09
Pants	107	5.55	2.70
T-shirt	135	4.84	2.32
Jacket/ long sleeves	9	10.7	5.55

Table 3.8: MAE calculated in gender groups

	Number of subjects	MAE of weight	MAE of BMI
Female	93	5.92	3.02
Male	70	4.13	1.92

influence on our weight estimation approach. While the highest MAE of BMI 5.55 is from jackets/long-sleeves group, which is much higher than MAEs over all subjects 2.54 (see details in Table 3.4). Since there are only 9 subjects wearing jackets or long-sleeves in this dataset, we did not design a jackets/long-sleeves model for volume correction. The MAEs vary in different clothes groups show that the performance of the BMI estimation approach performs unevenly for different clothes style.

Then we study the performance of BMI estimation in separated gender groups. The results are shown in Table 3.8. The third column is MAEs of estimated weight, the fourth column is MAEs of estimated BMI. We can see that the MAEs are smaller for the males, while is a little larger for females. This result suggests that our current BMI estimation approach performs slightly better for males. The higher errors for females may be caused by various clothing styles.

3.5.4 Compare with other volume calculating methods

In this work, the proposed volume calculating method for 3D reconstructed data is based on slicing and integration. We compare it with another method for 3D data volume calculation. It is a triangular projecting based method [104]. Given a projecting plane, the volume of the pentahedron, which consists of the triangle and its projection, is computed. The volume of the 3D data is the sum of the volume of all pentahedrons. We first apply method [104] to the dataset and obtain the volume V_1. Then the clothes models are applied to correct V_1. Finally the corrected volume V_2 is used to estimate the body weight. Table 3.9 gives the MAEs of body weight estimation based on these two volume calculating methods. One can see that our method gives smaller errors for

Table 3.9: Comparison of volume calculating between our method and other method.

Method	Male	Female	Overall
[104]	4.93	7.02	6.12
Ours	4.13	5.92	5.3

different groups.

3.5.5 Compare with other visual BMI estimation methods

As shown in Table 3.4, the MAE of BMIs of the proposed approach is 2.54. Wen and Guo [47] proposed a BMI computation approach from facial images with a reported MAE of 3.13. Dantchev et. al [63] explored the possibility of estimating height, weight, and BMI from facial images by a regression based deep network. The reported MAEs related to BMI is in the range of $2.3 + 0.06$. Jiang and Guo [64] developed a body weight analysis method from frontal view body images with a reported MAE of 3.8. [66] evaluated the performances of several facial representations for BMI estimation, the best performance is from Arcface 3.15 ± 0.06. Comparing with the above existing visual BMI estimation methods, the performance of the proposed method is quite acceptable, although the datasets and extracted features are different. It shows that the use of 3D data could be promising for BMI analysis.

3.6 Summary

This chapter presents an effective computational approach to BMI estimation from the normally dressed people in 3D space. Two clothes models have been proposed to obtain a more accurate estimation of the body volumes. Though the Kinect 3D fusions contain some noise, the proposed BMI estimation includes clustering and fitting components to suppress such noise. A new RGB-D dataset is collected for this study. Experimental results have shown the effectiveness of the proposed approach to people with different styles of clothes, for both females and males. Comparing to another 3D volume estimation method, our method achieves a significantly lower error. In the future, we would like to explore more precious clothes models (such as jackets and pants models) to make the estimated weight and BMI become more accurate.

Chapter 4

On Visual BMI Analysis from Facial Images

In this chapter, we study an interesting and challenging problem in computer vision-automatically assessing body mass index (BMI) from facial images. Facial feature extraction is an important step for visual BMI estimation. This work studies the visual BMI estimation problem based on the characteristics and performance of different facial representations, which has not been well studied yet. Various facial representations, including geometry based representations and deep learning based, are comprehensively evaluated and analyzed from three perspectives: the overall performance on visual BMI prediction, the redundancy in facial representations and the sensitivity to head pose changes. The experiments are conducted on two databases: a new dataset we collected, called the FIW-BMI and an existing large dataset Morph II. Our studies provide some deep insights into the facial representations for visual BMI analysis.

The organization of this chapter is as follows. Section 4.1 describes the problem studied in this chapter. The principles and related methods are systematically presented and discussed in Section 4.2. Section 4.3 characterizes two databases used for performance evaluation: a newly collected FIW-BMI dataset and Morph II. In Section 4.4, we conduct three experiments: the overall performance on visual BMI prediction, the redundancy in representations and the sensitivity to variant head pose, and provide detailed analysis and discussion. Section 4.5 summarizes the work and discoveries of this chapter.

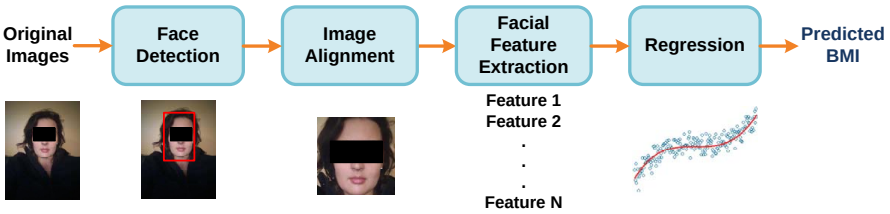


Figure 4.1: A typical framework for visual BMI estimation from two-dimensional (2D) facial images.

4.1 Problem Definition

Fig. 4.1 shows a typical framework for BMI estimation from two-dimensional (2D) facial images. It consists of four steps: face detection, image alignment, facial representation extraction, and regression. The first and second steps are the preparation for feature extraction. The third step is the most important which dominantly determines the performance for BMI estimation. Thereby, in this work, we study the visual BMI estimation problem from this key aspect: facial feature extraction methods, and explores methods to improve their performance.

We study the visual BMI estimation problem by analyzing two types of facial representation methods. The psychology inspired geometric features (PIGF) is used in [47,105]. Considering the whole facial shape may not be exactly defined by the PIGF, we explore another method for extracting geometric facial representation-pointer feature (PF), which defines the face shape by a series of facial landmarks. In addition, to take advantage of the above two geometric representations, a fusion method is utilized to extract a richer geometric representation, denoted as PIGF+PF. In terms of deep learning, the VGG-Face model has been utilized for BMI prediction in [61]. Considering the very high dimension of the VGG-Face feature, we also explore other deep models to extract the deep representations, e.g. the LightCNN [106], Centerloss models [107] and Arcface [108]. Thus we can get a deep insight into deep learning based facial representations for visual BMI analysis.

4.2 A deep insight into the visual BMI representations

As mentioned above, we consider that there are two representative types of facial representations for visual BMI analysis from facial images. We examine the principles of the facial representations systematically and discuss some related issues.

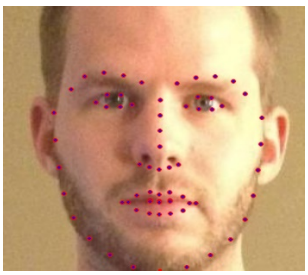


Figure 4.2: Illustration of pointer feature (PF), which consists of a series of facial landmarks.

4.2.1 Geometry based representations

The principle of a geometric model is to mathematically describe the facial shapes related to body fat. It shows that facial geometry measures are correlated to body fat or BMI. Inspired by these psychology studies [24, 25], the first computational method PIGF was developed by Wen and Guo [47], using geometric features to estimate seven facial metrics: cheek-to-jaw-width ratio (CJWR), face width-to-height ratio (WHR), face perimeter to area ratio (PAR), eye size (ES), lower face to face height ratio (LF/FH), face width to lower face height ratio (FW/LFH) and the mean of eyebrow height (MEH). Given the geometric features, some statistical methods can be used to map the features to BMI values. The facial landmarks need to be extracted prior to the geometric feature extraction. So the performance of BMI estimation is related to the accuracy of facial landmark detection. The experimental results reported in [47] show that the PIGF performs quite well on BMI estimation.

On the other hand, Mayer et al. [44] analyzed the association of facial landmarks with body fat. They found that comparing with WHR, the whole facial shape is also good at reflecting the total fat proportion of the body. The facial shape based points were used to study the correlation between facial shape and body fat. Inspired by this, we want to explore another method for extracting the geometric facial representation, called the pointer feature (PF). It defines the facial shape and features by a series of facial landmarks as shown in Fig. 4.2. The PF consists of coordinates of M facial landmarks, denoted as $(x_i, y_i), i = 1, ..., M$, which can be directly concatenated into a vector. Then the PF representation is a 2M-dimensional vector: $[x_1, y_1, ..., x_n, y_n, ..., x_M, y_M]^T$. It is obvious that PF relies on the accurate detection of facial landmarks.

In addition, a fusion of these two kinds of geometric features could be considered. A simple way is to concatenate the PIGF and PF, denoted as PIGF+PF. Hopefully, it can

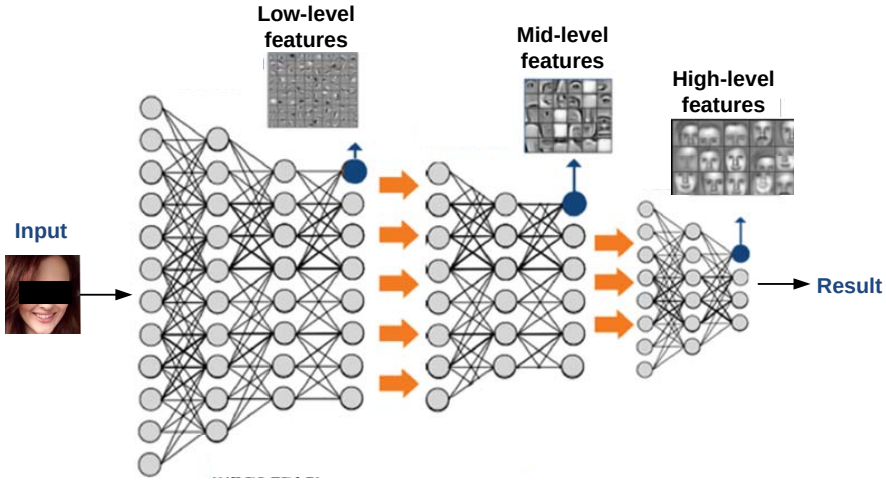


Figure 4.3: The pipeline of deep learning approach.

take advantages of the two geometric representations.

As a result, we will investigate three different geometric features, in order to get a deep understanding. The computation of geometric features is very fast, and the geometric representations are with very low dimensions.

4.2.2 Deep learning based representations

Recently, deep neural networks have been successfully applied to various applications. Fig. 4.3 shows a general pipeline of the deep learning approach. The VGG-Face is one of the deep convolutional networks originally proposed for face recognition, which learns a face embedding using a triplet loss function [62]. The network contains 13 convolutional layers, 5 max-pooling layers, 3 fully-connected (fc) layers and a final layer with the soft-max function. VGG-Face model is trained on 2.6 million face images from the web. It takes a face image of size 224×224 with a constant image with all pixels equal to (94,105,129) subtracted as the input. Having about 144 million parameters indicates that the VGG-Face is a complex model. Kocabey et al. [61] employed the pre-trained VGG-Face models to extract facial features for BMI analysis. The extracted features from layer fc6 of VGG-Face is utilized. The size of the feature vector in VGG-Face is 4096. Thus the dimension of the VGG-Face representation is quite high.

Considering the high computational complexity for VGG-Face model, it is interesting to investigate other deep models with a lower computational cost for visual BMI analysis. Here we explore the LightCNN, Centerloss and Arcface models.

LightCNN is a network with a low computational complexity which learns a compact

face embedding on a large-scale dataset with noisy labels [106]. It proposed a Max-Feature-Max (MFM) activation function to suppress a small number of neurons and to make CNN models light and robust. The model is trained on 493,456 face images from CASIA WebFace dataset. The input of the network is a gray-scale face image of size 128×128. Considering the significant performance achieved by this model on the face recognition task, the layer fc1 (with size 256×1) of LightCNN is used to extract deep facial features for BMI estimation.

The features learned by the deep networks trained under the supervision of softmax loss [109] may not be discriminative enough. In order to improve the discriminative power of the learned features, Wen et al. [107] proposed a center loss function to minimize the intra-class variations while keeping features of different classes separable. The center loss based network takes a face image of size 96×112 as the input. This model was developed for face recognition, which we evaluate its use for visual BMI estimation. We extract features for each image and its horizontally flipped one, and concatenate them as a 1024 dimensional feature vector.

Additive Angular Margin Loss (ArcFace) [108] can extract highly discriminative features for face recognition by directly optimizing the geodesic distance margin through the correspondence between the angle and arc in the normalized hypersphere. It outperforms many other deep models for face recognition. The input of Arcface model is face images of size 112×112. The final 512-dimension embedding feature of the network is utilized for visual BMI analysis.

4.3 Databases

Given various face representations for BMI analysis, we conduct experiments on two databases: a newly collected face database by us and the Morph II. They are different in size and characteristics. The details about the two databases are given below.

4.3.1 FIW-BMI database

We collect a new dataset, called face in the wild for BMI analysis (FIW-BMI) [1]. The facial images were collected from a social website−Reddit posts [2]. We went through the original images by a deep cascaded multi-task based face detector [110]. Given the detected face landmarks (two eyes, nose and mouth), each face image is cropped and

[1] Please contact the authors for the dataset.

[2] Website: `http://www.reddit.com/r/progresspics`

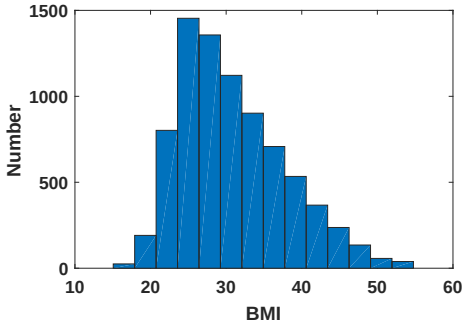


Figure 4.4: Distribution of BMI values on FIW-BMI database. The BMI values span a wide range with most of the values distribute between 20 to 50.

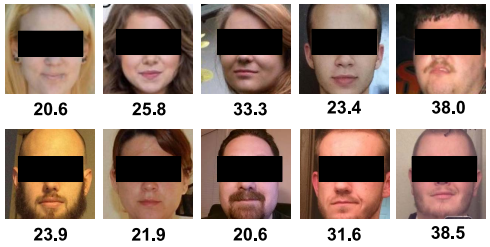


Figure 4.5: Samples of the cleaned images in FIW-BMI database.

normalized to the size of 256×256. Considering all the images are from social networks, they are not strictly frontal view face images with a clean background. We visually checked all images and discarded the images which are not appropriate for visual BMI analysis, such as large head pose changes or exaggerated facial expressions. Finally, the annotation for each image is manually checked to generate the correct labels.

After all the above procedures, 7930 images from 4881 individuals were kept, along with the corresponding gender, height, and weight labels. Among these individuals, there are 3192 males (5197 images) and 1689 females (2733 images). Fig. 4.4 shows the BMI distribution of the whole database. Because the Reddit posts is a social network displaying people's progress of weight loss, weight gain, or essentially any type of body changes, the BMI values of these images distribute over a very wide range: 15 to 60. The mean BMI value of the database is 30.8, the standard deviation of the BMI values is 6.97. Among these images, 43 are underweight (BMI$\leq$18.5), 1662 are normal (18.5<BMI$\leq$25), 2455 are overweight (25<BMI$\leq$30), 3770 are in obese (BMI>30). Fig. 4.5 shows a few examples of the facial images from our FIW-BMI database.

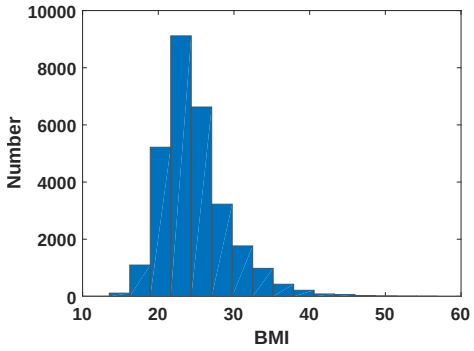


Figure 4.6: Distribution of BMI values on Morph II. The BMI values mainly distribute between 15 to 35.

Table 4.1: Details about the selected Morph II database.

	Black male	Black female	White male	White female
#Subject	6497	1096	1565	535
#Images	19290	2824	4862	2057

4.3.2 Morph II database

Morph II database [51] contains 55,608 mugshot-style frontal view face images along with the age, gender, and ethnicity labels. Most of them are with height and weight values. The BMI values can be computed from the weight and height. There is an uneven distribution of the ethnicity in the database, eg. about 96% identities are Black and White, while 4% are Hispanic, Asian, Indian, and others. Only images from African Americans and White are used for this study. Totally 29,033 images of 9693 identities were selected. Details about the selected Morph II dataset are given in Table 4.1. The images are separated into four groups by gender and ethnicity. The distribution of BMI values on the selected database (includes training and test sets) is shown in Fig. 4.6. Comparing to Fig. 4.4, the BMI values of Morph II mainly distribute on a relatively small range: 15 to 35. The mean BMI value of selected Morph II is 24.8, the standard deviation of the BMI values is 4.61. Among these, 893 are underweight, 16,582 are normal, 8,237 are overweight and 3,321 are obese.

4.4 Experiments and analysis

Experimentally, we evaluate and analyze the two major types of facial representations for BMI estimation. The experiment settings and performance metrics are briefly described

Table 4.2: Splitting FIW-BMI by gender.

	Training set		Test set	
	#Subject	#Images	#Subject	#Images
Male	2551	4136	641	1061
Female	1329	2164	360	569

first. Then the overall performance of the facial representations on two databases is presented. We further analyze the facial representations from two perspectives: the redundancy of the facial representations, and the sensitivity of facial representations to various head pose changes. Finally, two influence factors for BMI analysis are discussed.

4.4.1 Experiment setting

1) Database setting: The BMI values are estimated based on separated training in each gender (and ethnicity) group using the SVR model in this work without any other specification. To evaluate the overall performance of the seven facial representations, FIW-BMI is split into 10 subsets for each gender group (10 subsets for the male group and 10 subsets for female). A similar process is applied to Morph II. It is also split into 10 subsets for each gender and ethnicity group. We use cross-validation for performance measure. 8 subsets are used as the training set and the remaining 2 are the testing set in each round for each gender-ethnicity group. There is no overlap of individuals between the training and test sets in each round. Such a process is repeated 30 rounds for each group (the training and test sets are different for each round). Then 95% confidence intervals are calculated based on the results of these 30 repeated experiments.

To analyze the redundancy of the facial representations and the sensitivity to head pose variations, the training and test sets of the two databases are given in Tables 4.2 and 4.3, respectively. The same individual can only appear in either training or test set, but not both.

2) Images preprocessing: The face image alignment is applied prior to the extraction of geometric features. The alignment is based on the detected eye coordinates. It basically performs translation, rotation, and scaling of the faces so as to align all face images into the common eye coordinates. The output is a cropped 256×256 image. The Openface toolkit [111] is employed for detecting 68 face landmarks. The output of the PIGF is a 7 dimensional representation: $[CJWR, WHR, PAR, ES, LF/FH, FW/LFH, MEH]^T$. The PF consists of the coordinates of 68 facial landmarks, denoted as $(x_i, y_i), i = 1, ..., 68$, resulting in a 136

Table 4.3: Splitting selected Morph II by gender and ethnicity.

	Training set		Test set	
	#Subject	#Images	#Subject	#Images
Black male	4568	13574	1929	5716
Black female	873	2218	223	606
White male	1245	3856	320	1006
White female	428	1615	107	442

dimensional representation: $[x_1, y_1, ..., x_n, y_n, ..., x_{68}, y_{68}]^T$.

The implemented and pre-trained models of VGG-Face, LightCNN-29 and Centerloss are used from the Caffe deep learning framework [112]. All weights of the fully-connected layer of each deep network are used for feature extraction. These layers are noted as fc6 in VGG-Face, fc1 in LightCNN and fc5 in Centerloss. The VGG-Face model takes a 224×224 color image with the mean subtracted and outputs a 4096 dimensional feature vector. The LightCNN model provides a 256 dimensional representation extracted from a 128×128 gray-scale image. The Centerloss model outputs a 1024 dimensional representation with the input 96×112 color images. The Arcface model takes 112×112 color image and outputs a 512 dimensional feature vector. The image alignment is required before extracting deep representations. It is done by following the alignment protocol provided by each deep model.

3) Implementation details for machine learning: As shown in Fig. 4.1, the extracted facial representations are then used to train a regression model. We employ the support vector regression (SVR) [48] model to learn the mapping from the extracted representations to BMI values. The SVR is selected due to its robust generalization behavior. The Gaussian Radial Basis Function (RBF) is utilized as the SVR kernel.

4.4.2 Performance metrics

Mean absolute error (MAE) is employed to measure the performance on BMI estimation. It is defined as the average of the absolute error between the estimated BMI values and the ground truth BMI values, which is computed by: $MAE = \frac{1}{N}\sum_{k=1}^{N}|\hat{p}_k - p_k|$, here p_k is the ground truth BMI value for image k, $\hat{p}_k$ is the corresponding estimated BMI value, N is the number of test images. This measure is motivated by its use in age estimation [103].

The second measurement is the accuracy of the predicted BMI category. According to

the estimated BMI values, we can compute the corresponding BMI category (underweight, normal, overweight and obese). The accuracy of the predicted category is the proportion of the total number of predictions that are correct. This measurement is helpful to decide if the errors are acceptable. For example, given an image with a ground-truth BMI value 24, the estimated value is 19. Though the absolute error is 5, the predicted category (normal) is correct. On the other hand, the category has a limitation. For example, if the ground-truth BMI of an image is 30 and the estimated value is 30.5, though the absolute error is 0.5, the predicted category (obese) is incorrect.

Mean absolute percentage error (MAPE) is proposed as the third measure. It is a relative error computed as:

$$MAPE = \frac{100}{N}\sum_{k=1}^{N}\left|\frac{\hat{p_k} - p_k}{p_k}\right|. \tag{4.1}$$

Considering the advantages and limitations of the above three measurements, we use all of them to evaluate the performance.

A 95% confidence interval (CI) is a range of values that it can be 95% certain contains the true mean of the population. We calculate the 95% confidence intervals of the above three metrics based on the results of 30 repeated experiments. It is computed by:

$$CI = \bar{X} \pm Z\frac{s}{\sqrt{n}}, \tag{4.2}$$

here n is the number of observations, $\bar{X}$ is the mean of observations, and s is the standard deviation. For 95% confidence interval, the Z value is 1.96.

4.4.3 Overall performance comparison

The 95% confidence interval of MAEs for the seven representations for BMI estimation on the two databases are given in Tables 4.4 and 4.5. The MAEs are calculated from the whole test set. To better present the details about the performance, we further calculated the MAEs from each BMI category. In addition to the MAEs, the 95% confidence intervals of the accuracy for the predicted BMI category are given in Tables 4.6 and 4.7. The 95% confidence interval of MAPEs are given in Tables 4.8 and 4.9. Combining MAEs (Tables 4.4 and 4.5), the accuracy for category classification (Tables 4.6 and 4.7) and MAPEs (Tables 4.8 and 4.9) to evaluate and analyze the performances with more specific information, some interesting observations can be obtained.

Performance of the two types of facial representations: The performances of the seven facial representations are different from each other. Overall, the experimental results

Table 4.4: 95% confidence interval of MAEs for the seven facial representations for BMI prediction on FIW-BMI.

	Male				
	All	Underweight	Normal	Overweight	Obese
PIGF	3.78±0.07	7.90±0.25	3.79±0.08	2.52±0.50	4.76±0.14
PF	3.76±0.07	8.96±0.37	3.81±0.10	2.51±0.04	4.56±0.12
PIGF+PF	3.70±0.07	8.16±0.27	3.79±0.09	2.50±0.04	4.49±0.10
VGG-Face	3.26±0.06	5.61±0.34	2.99±0.10	2.50±0.05	4.24±0.10
LightCNN	3.44±0.06	5.76±0.36	3.17±0.09	2.50±0.05	4.30±0.09
Centerloss	3.40±0.05	8.02±0.35	3.19±0.08	2.54±0.05	4.30±0.11
ArcFace	**3.15±0.07**	5.52±0.21	3.18±0.04	2.25±0.05	4.07±0.14
	Female				
	All	Underweight	Normal	Overweight	Obese
PIGF	4.26±0.08	10.37±1.10	5.30±0.07	2.68±0.05	4.68±0.13
PF	4.15±0.08	9.43±0.90	5.08±0.09	2.91±0.07	4.60±0.10
PIGF+PF	4.10±0.07	9.97±0.87	5.02±0.08	2.71±0.07	4.53±0.12
VGG-Face	3.66±0.08	9.79±0.95	4.42±0.11	2.67±0.09	3.81±0.12
LightCNN	3.90±0.03	9.94±1.00	4.62±0.09	2.86±0.07	4.12±0.06
Centerloss	3.82±0.11	8.56±0.50	5.00±0.21	2.70±0.11	3.97±0.11
ArcFace	**3.51±0.09**	9.76±0.85	4.47±0.08	2.53±0.08	3.62±0.17

show that these two types of facial representations both are effective for addressing BMI estimation. And the deep model based methods (VGG-Face, LightCNN, Centerloss and Arcface) perform better than the geometry based methods (PIGF, PF, and PIGF+PF). Among them, measuring with MAEs, the VGG-Face and Arcface show more robustness than the others in most cases.

For the white female group, the deep learning based representations do not show clear advantages over the geometric representations as in other groups. From Table 4.1, we can see this group has the least number of images for training and testing. Since the training time of SVR models for deep representations is much longer than the geometric representations. The geometric representations are more suitable for small datasets. The deep representations perform better on a large dataset with much more time cost.

From Tables 4.4 and 4.5, it can be seen that the confidence intervals of PIGF+PF are smaller than both PIGF and PF for most groups. To decide whether a significant performance difference exists between the fused geometric feature (PIGF+PF) and the individual feature (PIGF, PF), we apply a hypothesis testing with a statistical significance measure. The null hypothesis is: there is no performance difference between the two features. We can make a decision by:

Table 4.5: 95% confidence interval of MAEs for the seven facial representations for BMI prediction on Morph II database.

	Black male				
	All	Underweight	Normal	Overweight	Obese
PIGF	2.70±0.04	6.61±0.22	1.82±0.02	2.36±0.01	7.38±0.10
PF	2.67±0.04	6.69±0.24	1.84±0.02	2.29±0.02	7.22±0.11
PIGF+PF	2.63±0.04	6.61±0.24	1.84±0.03	2.28±0.02	7.14±0.11
VGG-Face	2.45±0.05	6.01±0.21	1.87±0.03	2.10±0.03	5.73±0.11
LightCNN	2.42±0.10	5.81±0.16	1.78±0.02	2.16±0.02	5.84±0.10
Centerloss	2.50±0.04	6.23±0.28	1.84±0.03	2.12±0.04	6.37±0.12
ArcFace	**2.40±0.03**	6.24±0.28	1.85±0.03	2.01±0.03	5.65±0.08
	Black female				
	All	Underweight	Normal	Overweight	Obese
PIGF	3.77±0.08	6.25±0.16	2.26±0.05	2.95±0.09	8.38±0.19
PF	3.76±0.08	6.53±0.15	2.35±0.05	2.43±0.09	8.33±0.19
PIGF+PF	3.68±0.07	6.37±0.14	2.39±0.06	2.60±0.08	8.11±0.15
VGG-Face	3.48±0.04	5.21±0.14	2.25±0.05	2.61±0.08	7.25±0.16
LightCNN	3.55±0.05	5.40±0.17	2.38±0.05	2.53±0.05	7.82±0.23
Centerloss	3.63±0.06	5.41±0.16	2.42±0.06	2.71±0.10	7.94±0.16
ArcFace	**3.51±0.07**	5.12±0.13	2.43±0.05	2.57±0.07	7.26±0.17
	White male				
	All	Underweight	Normal	Overweight	Obese
PIGF	2.67±0.03	5.73±0.36	1.94±0.03	2.35±0.06	7.30±0.15
PF	2.57±0.03	5.85±0.43	1.86±0.03	2.28±0.04	7.17±0.15
PIGF+PF	2.49±0.03	5.69±0.37	1.84±0.03	2.22±0.05	7.15±0.17
VGG-Face	**2.30±0.03**	4.83±0.31	1.82±0.03	2.08±0.04	5.35±0.21
LightCNN	2.35±0.04	4.73±0.25	1.82±0.04	1.98±0.03	6.15±0.15
Centerloss	2.41±0.03	5.21±0.47	1.87±0.07	2.09±0.05	6.03±0.25
ArcFace	2.32±0.02	5.45±0.30	1.77±0.03	1.96±0.04	6.27±0.18
	White female				
	All	Underweight	Normal	Overweight	Obese
PIGF	2.96±0.04	4.27±0.21	1.73±0.03	4.74±0.09	8.82±0.38
PF	3.13±0.07	4.68±0.18	1.88±0.03	4.41±0.10	8.88±0.38
PIGF+PF	3.01±0.05	4.42±0.17	1.78±0.04	4.35±0.09	8.80±0.41
VGG-Face	2.96±0.06	4.11±0.15	1.77±0.08	3.99±0.09	8.72±0.32
LightCNN	**2.87±0.07**	4.08±0.17	1.72±0.05	3.89±0.07	7.80±0.61
Centerloss	2.94±0.09	4.22±0.22	1.79±0.07	3.89±0.15	8.94±0.52
ArcFace	2.90±0.05	4.02±0.16	1.76±0.03	3.42±0.10	8.63±0.36

Table 4.6: 95% confidence interval of BMI category prediction accuracy (%) for the seven facial representations on FIW-BMI.

	Male				
	All	Underweight	Normal	Overweight	Obese
PIGF	72.1±0.6	3.8±2.1	45.8±1.0	79.0±0.7	80.6±0.7
PF	73.9±0.6	4.0±2.7	46.0±1.3	76.7±0.8	81.7±0.7
PIGF+PF	74.1±0.5	4.1±2.5	46.2±1.1	78.5±0.7	82.3±0.6
VGG-face	78.0±0.6	15.5±5.8	62.5±1.3	79.8±1.2	85.5±0.8
LightCNN	76.3±0.5	3.3±2.2	60.0±2.6	74.3±0.8	86.7±0.5
Centerloss	75.4±0.5	4.1±3.0	60.3±1.6	72.9±1.0	85.8±0.6
ArcFace	**79.1±0.3**	5.7±2.7	62.4±0.9	77.5±0.9	90.3±0.6
	Female				
	All	Underweight	Normal	Overweight	Obese
PIGF	68.6±0.7	3.5±2.1	18.0±1.5	78.9±1.0	82.5±1.2
PF	69.4±0.7	3.0±1.9	29.8±1.1	70.1±4.4	83.2±1.0
PIGF+PF	70.1±0.6	3.2±1.8	28.8±1.3	75.4±2.3	83.5±1.1
VGG-face	74.5±0.7	4.9±1.9	35.9±2.3	73.5±1.6	89.9±0.8
LightCNN	71.9±0.6	4.5±2.0	36.8±1.6	68.2±1.3	87.8±0.7
Centerloss	73.3±1.3	4.7±2.1	33.1±3.7	73.8±1.7	88.2±0.8
ArcFace	**75.7±0.8**	5.2±2.3	39.5±1.8	72.0±1.2	91.4±1.0

- If the p-value is smaller than the significance level α, it can reject the null hypothesis;
- If the p-value is larger than the significance level α, it fails to reject the null hypothesis.

Here the significance level α is set to 0.01. The p-value is computed from the MAEs of the two features obtained from the repeated (30 times) experiments on each group of the two databases. According to the calculation, the range of p-value is from 2.7e-3 to 1.3e-06. This result reveals the significant differences between the fused geometric feature (PIGF+PF) and the individual feature (PIGF, PF).

Performance on the four BMI categories: All seven representations have different performances in the four categories. As shown in Tables 4.4, 4.5, 4.8 and 4.9, the MAEs and MAPEs are higher for underweight category on FIW-BMI dataset; MAEs and MAPEs are high for underweight and obese categories on selected Morph II dataset. This is caused by the BMI distributions of the datasets. From Figs. 4.4 and 4.6, it can be seen that most images in the FIW-BMI database are in the categories of overweight and obese, while most images in Morph II are in the normal and overweight categories. The performance of the facial representations is influenced by the number of training

Table 4.7: 95% of BMI category prediction accuracy (%) confidence interval for the seven facial representations on Morph II.

	Black male				
	All	Underweight	Normal	Overweight	Obese
PIGF	71.7±0.7	4.0±0.3	94.7±0.2	51.0±0.9	19.5±0.9
PF	73.0±0.6	3.8±0.3	94.5±0.3	55.2±1.0	24.7±1.2
PIGF+PF	73.0±0.6	3.7±0.5	94.7±0.2	55.9±1.0	21.8±1.2
VGG-Face	73.5±0.3	14.1±2.5	89.0±0.7	57.1±2.1	35.4±1.7
LightCNN	77.1±0.2	17.3±3.5	90.7±0.2	62.7±0.3	48.3±1.1
Centerloss	75.6±0.4	16.9±4.1	93.1±0.6	61.7±1.9	33.5±1.7
ArcFace	**78.4±0.4**	18.7±3.5	93.0±0.7	66.4±1.6	47.3±1.3
	Black female				
	All	Underweight	Normal	Overweight	Obese
PIGF	65.0±1.0	2.5±0.8	93.9±0.6	48.5±1.9	29.3±3.1
PF	67.7±1.2	0.6±0.5	91.4±1.0	58.5±1.9	33.2±2.5
PIGF+PF	67.9±1.1	0.9±0.3	92.4±1.3	59.7±1.0	31.2±2.6
VGG-Face	65.9±1.0	10.1±1.7	87.1±1.4	60.7±1.3	30.0±2.3
LightCNN	70.3±0.8	14.2±5.5	88.9±0.7	64.6±2.0	48.7±2.1
Centerloss	66.4±1.2	9.3±1.6	89.1±1.1	59.4±1.7	30.7±3.6
ArcFace	**69.2±1.0**	7.0±1.2	88.2±1.0	64.7±1.1	45.0±3.2
	White male				
	All	Underweight	Normal	Overweight	Obese
PIGF	71.9±1.0	5.5±1.1	95.9±0.4	53.2±1.9	4.6±0.6
PF	72.5±0.9	4.5±0.9	95.7±0.3	54.4±1.6	7.7±1.4
PIGF+PF	74.5±1.0	4.7±1.0	91.7±0.4	61.5±2.0	28.9±1.5
VGG-Face	75.2±0.7	11.1±2.4	90.5±0.9	60.5±1.1	33.0±1.7
LightCNN	**76.7±0.6**	9.2±1.9	91.5±0.4	64.1±1.4	36.7±2.5
Centerloss	75.8±0.7	10.7±2.5	91.6±0.9	62.8±1.0	30.6±2.6
ArcFace	75.8±0.5	14.5±3.7	94.0±1.5	63.8±1.1	17.8±3.7
	White female				
	All	Underweight	Normal	Overweight	Obese
PIGF	70.2±1.0	5.5±1.0	99.1±0.5	19.2±2.2	25.3±3.9
PF	70.6±1.2	13.9±3.6	99.6±0.2	22.5±1.9	24.7±1.4
PIGF+PF	69.1±1.1	9.5±1.5	99.7±0.2	14.1±3.4	15.1±4.9
VGG-Face	**73.2±0.8**	30.4±1.9	98.7±0.2	27.9±2.1	15.9±3.1
LightCNN	72.9±0.6	27.4±2.5	97.6±0.2	30.2±2.3	26.5±3.6
Centerloss	70.1±1.0	17.0±1.8	99.4±0.2	27.8±2.7	19.4±2.7
ArcFace	70.6±0.7	17.5±1.4	98.7±0.3	34.5±1.5	14.6±2.6

Table 4.8: 95% confidence interval of MAPEs (%) for the seven facial representations for BMI prediction on FIW-BMI.

	Male				
	All	Underweight	Normal	Overweight	Obese
PIGF	13.7±0.3	31.6±0.7	14.0±0.3	8.7±0.2	17.5±0.5
PF	13.4±0.3	33.0±1.0	13.9±0.4	8.9±0.2	16.7±0.4
PIGF+PF	13.3±0.3	32.1±0.8	14.1±0.3	8.8±0.2	16.2±0.4
VGG-face	11.4±0.2	23.9±1.5	12.0±0.3	8.5±0.1	13.0±0.3
LightCNN	11.6±0.2	24.2±1.2	11.9±0.3	8.5±0.2	13.5±0.2
Centerloss	11.9±0.1	30.0±1.0	12.3±0.2	9.1±0.2	13.7±0.3
ArcFace	**11.0±0.2**	24.0±0.7	11.8±0.1	9.1±0.1	12.3±0.3
	Female				
	All	Underweight	Normal	Overweight	Obese
PIGF	15.5±0.3	38.3±2.7	19.4±0.2	8.7±0.1	17.6±0.5
PF	15.0±0.5	34.4±2.3	18.8±0.3	9.5±0.2	16.3±0.6
PIGF+PF	14.9±0.3	36.7±2.5	18.2±0.2	8.8±0.1	16.3±0.5
VGG-face	12.3±0.2	35.5±3.5	17.1±0.4	9.0±0.2	12.5±0.3
LightCNN	12.7±0.1	34.8±2.7	16.2±0.3	9.1±0.2	13.4±0.3
Centerloss	12.6±0.3	35.6±1.9	17.5±0.7	8.8±0.3	12.7±0.4
ArcFace	**12.0±0.2**	35.9±2.2	16.7±0.3	9.3±0.2	11.8±0.4

images. Less training images in the specific category leads to worse performance for the corresponding category.

Performance on the two databases: Comparing to Morph II, all facial representations show less robustness on FIW-BMI database. This is caused by the wild data collection of FIW-BMI. Slight head pose changes exist on this database, and the BMI values distribute in a larger range on FIW-BMI (20-55) than the Morph II (15-35). To further analyze the performances of these facial representations on the two databases, we do another experiment that uses Morph II for training and FIW-BMI for testing, and vice versa. The experimental results are given in Tables 4.10 and 4.11. Comparing with the results in Tables 4.4 and 4.5, one can see that the performances of all seven facial representations drop significantly. This may be caused by the quite different BMI distributions of the two databases and the different "domains" of the images (Morph II has mugshot face images, while FIW-BMI is with daily life face images).

4.4.4 Redundancy in facial representations

According to the overall performance of these facial representations on BMI estimation, it is shown that the deep representations perform better on a large dataset. Since the

Table 4.9: 95% confidence interval of MAPEs (%) for the seven facial representations for BMI prediction on Morph II.

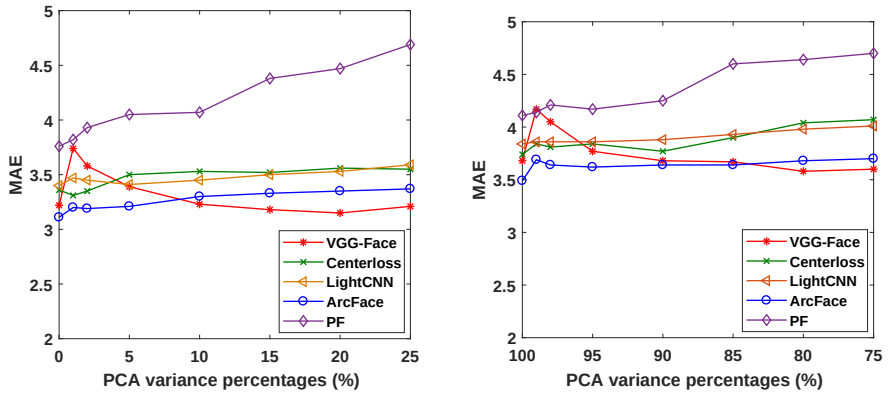
	Black male				
	All	Underweight	Normal	Overweight	Obese
PIGF	10.9±0.2	27.9±0.9	7.5±0.1	9.7±0.1	28.5±0.4
PF	10.7±0.1	28.2±0.9	7.4±0.1	9.4±0.1	27.7±0.4
PIGF+PF	10.7±0.1	27.9±0.9	7.5±0.1	9.3±0.1	27.7±0.4
VGG-Face	9.5±0.1	25.9±0.8	7.8±0.1	8.7±0.3	21.3±0.4
LightCNN	9.3±0.1	24.7±0.6	7.0±0.1	8.3±0.1	20.9±0.2
Centerloss	10.0±0.1	26.4±1.0	7.5±0.1	8.7±0.2	23.7±0.5
ArcFace	**9.1±0.1**	25.5±1.1	7.4±0.1	8.1±0.1	20.2±0.3
	Black female				
	All	Underweight	Normal	Overweight	Obese
PIGF	15.3±0.3	26.2±0.6	9.3±0.2	12.6±0.4	32.4±1.0
PF	15.1±0.3	26.0±0.5	9.6±0.2	11.8±0.4	31.7±0.9
PIGF+PF	15.1±0.2	26.1±0.5	9.0±.2	12.0±0.3	33.5±0.9
VGG-Face	**12.4±0.2**	21.9±0.5	9.2±0.1	9.7±0.4	24.3±0.6
LightCNN	12.9±0.2	21.1±0.7	9.5±0.2	9.8±0.3	24.5±0.8
Centerloss	14.5±0.2	23.0±0.6	10.0±0.2	11.2±0.3	29.8±0.7
ArcFace	12.8±0.3	22.1±0.5	10.0±0.2	10.5±0.3	24.3±0.7
	White male				
	All	Underweight	Normal	Overweight	Obese
PIGF	10.5±0.1	24.6±1.5	7.6±0.1	9.5±0.2	28.5±0.6
PF	10.5±0.1	24.9±1.6	7.7±0.3	9.5±0.2	28.1±0.7
PIGF+PF	10.1±0.1	24.7±1.3	7.3±0.3	9.6±0.2	28.0±0.5
VGG-Face	8.6±0.1	22.1±1.1	7.0±0.2	7.5±0.1	19.3±0.7
LightCNN	**8.5±0.1**	20.3±0.9	6.8±0.2	7.7±0.1	18.9±0.6
Centerloss	9.7±0.1	24.1±1.8	7.7±0.2	8.5±0.2	22.6±1.0
ArcFace	9.6±0.1	24.0±1.2	7.7±0.1	8.4±0.1	23.0±0.6
	White female				
	All	Underweight	Normal	Overweight	Obese
PIGF	13.8±0.2	19.6±0.8	7.9±0.1	22.0±0.4	36.8±1.9
PF	10.5±0.1	24.9±1.6	7.7±0.1	12.5±0.2	28.1±0.7
PIGF+PF	11.9±0.2	22.5±1.1	7.7±0.1	16.5±0.3	32.1±1.1
VGG-Face	13.9±0.5	17.3±0.9	8.7±0.2	19.9±0.5	37.1±2.3
LightCNN	**11.3±0.3**	16.6±0.6	7.4±0.2	13.6±0.3	28.7±3.1
Centerloss	13.4±0.4	19.2±0.8	8.7±0.2	17.4±0.7	36.0±2.6
ArcFace	12.8±0.2	18.7±0.6	8.2±0.1	15.3±0.1	34.4±1.8

Table 4.10: Performance of the seven facial representations where using Morph II for training and FIW-BMI for testing.

	Male			Female		
	MAE	Accuracy (%)	MAPE(%)	MAE	Accuracy(%)	MAPE(%)
PIGF	5.41	53.5	20.6	6.73	39.4	25.2
PF	5.35	54.3	21.9	6.86	40.7	26.8
PIGF+PF	5.30	54.2	21.2	6.88	40.9	26.9
VGG-Face	4.32	64.5	15.3	5.79	51.9	20.5
LightCNN	4.45	63.2	15.6	5.91	51.3	20.6
Centerloss	4.61	60.8	16.5	6.22	49.5	22.6
ArcFace	4.21	64.1	15.1	5.73	52.2	20.1

Table 4.11: Performance of the seven facial representations where using FIW-BMI for training and Morph II for testing.

	Black male			Black female		
	MAE	Accuracy(%)	MAPE(%)	MAE	Accuracy(%)	MAPE(%)
PIGF	4.82	50.1	15.3	5.80	44.7	18.4
PF	4.95	46.7	16.7	5.95	42.6	20.0
PIGF+PF	4.72	48.5	16.1	5.86	43.6	19.8
VGG-Face	4.00	56.4	14.1	5.52	44.5	18.8
LightCNN	3.59	63.5	12.8	5.52	44.0	18.7
Centerloss	3.92	59.8	13.8	6.08	40.5	20.2
ArcFace	3.59	62.7	13.2	4.90	53.0	17.4
	White male			White female		
	MAE	Accuracy(%)	MAPE(%)	MAE	Accuracy(%)	MAPE(%)
PIGF	5.58	40.2	18.3	6.43	37.7	22.4
PF	5.82	36.8	19.0	6.58	37.0	24.5
PIGF+PF	5.78	37.2	18.9	6.40	38.6	23.9
VGG-Face	3.69	63.6	13.1	5.00	46.7	18.3
LightCNN	3.51	64.9	12.5	4.99	49.1	18.0
Centerloss	4.08	60.0	14.1	5.23	48.5	18.5
ArcFace	3.32	67.4	11.1	5.03	48.6	18.1



(a) (b)

Figure 4.7: BMI estimation error (measured by MAEs) of applying PCA to facial representations by different percentages of explained variance on FIW-BMI database. (a) is the results of the male group, and (b) is the female group.

Table 4.12: Performance (MAEs) of applying PCA to the five facial representations for BMI prediction. A downward arrow (↓) denotes the MAE decreases, comparing with the method without PCA. And an upward arrow (↑) denotes the MAE increases.

Method	FIW-BMI		Morph II			
	Male	Female	Black male	Black female	White male	White female
PF + PCA	3.82 ↑	4.14 ↑	2.67 ↑	3.75 ↑	2.73 ↑	3.14 ↑
VGGFace + PCA	**3.15** ↓	**3.57** ↓	**2.41** ↓	**3.56** ↓	**2.41** ↓	2.91 ↓
LightCNN + PCA	3.41 ↑	3.86 ↑	2.45 ↑	3.71 ↑	2.49 ↑	3.08 ↑
Centerloss + PCA	3.31 ↓	3.77 ↑	2.50 ↑	3.73 ↑	2.50 ↑	**2.86** ↓
ArcFace + PCA	3.19 ↑	3.62 ↑	2.38 ↑	3.51 ↑	2.57 ↑	2.94 ↓

number of training samples is limited, we try to eliminate the negative influence caused by the small number of training samples. Thereby it is essential to analyze the redundancy in facial representations and explore efficient methods to improve their performance.

One of the problems with high-dimensional features is that, in many cases, not all the measured features are relevant or important for understanding the underlying phenomena of interest. It is, therefore, interesting to analyze the redundancy in the representations. To figure out the issue, we first apply dimension reduction to the five facial representations (VGG-Face, LightCNN, Centerloss, Arcface and PF), then evaluate the performance of the reduced dimensions. As one of the typical dimension reduction methods, Principal Component Analysis (PCA) is selected. Note that the PCA projection is only learned

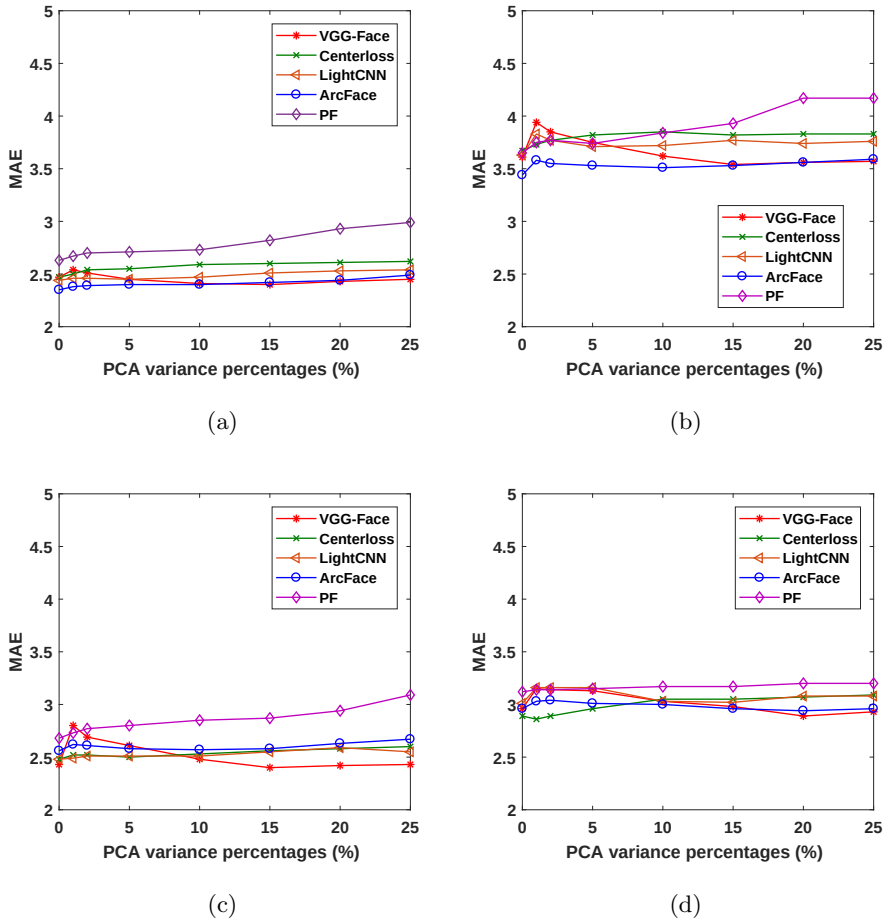


Figure 4.8: BMI estimation error (measured by MAEs) of applying PCA to facial representations by different percentages of explained variance on Morph II. Each sub-figure shows the result of the different gender-ethnicity group: (a) black male, (b) black female, (c) white male, and (d) white female.

with the training set. The dimensions of the four analyzed facial representations are as follows: PF is 128-dimension, VGG-Face is 4096-dimension, LightCNN is 256-dimension, Centerloss is 1024-dimension and Arcface is 512-dimension. Because the dimension of PIGF is seven and each dimension has its physical meaning (as mentioned in Section 4.2), PIGF and PIGF+PF are not involved in this investigation.

The percentage of explained variance is an index of the goodness of fit when applying PCA. It can be easily computed as the eigenvalues of corresponding components divided by the total variance. Here the total variance is the sum of all eigenvalues. Because the percentage of explained variance is a key factor to influence the performance of dimension reduced representations, different percentages (99%, 98%, 95%, 90%, 85%, 80%, and

Table 4.13: The number of kept dimensions corresponding to different percentages of explained variance on FIW-BMI database.

	Male					Female				
	PF	VGG	LightCNN	Centerloss	ArcFace	PF	VGG	LightCNN	Centerloss	ArcFace
99 %	19	1763	124	203	228	18	1326	121	199	218
98 %	15	1458	101	178	215	14	1101	99	174	203
95 %	10	988	67	138	189	10	755	67	130	174
90 %	7	616	50	104	160	7	479	47	94	145
85 %	6	414	41	85	138	6	325	38	74	124
80 %	5	285	34	71	121	5	224	32	61	107
75 %	4	196	29	60	106	4	153	27	51	93

Table 4.14: The number of kept dimensions corresponding to different percentages of explained variance on Morph II.

	Black male					Black female				
	PF	VGG	LightCNN	Centerloss	ArcFace	PF	VGG	LightCNN	Centerloss	ArcFace
99 %	18	1675	118	177	216	16	1140	115	181	208
98 %	14	1352	93	144	199	13	905	90	154	188
95 %	9	825	59	97	167	9	568	58	112	153
90 %	6	441	41	66	135	6	325	40	78	119
85 %	5	265	32	50	114	5	204	31	59	98
80 %	4	166	26	41	97	4	131	25	46	82
75 %	3	104	22	34	84	4	84	20	38	69
	White male					White female				
	PF	VGG	LightCNN	Centerloss	ArcFace	PF	VGG	LightCNN	Centerloss	ArcFace
99 %	18	1625	115	184	226	18	1076	115	191	213
98 %	13	1311	91	158	213	14	885	91	167	196
95 %	9	845	59	120	186	8	591	60	127	165
90 %	6	502	44	91	156	6	363	44	94	134
85 %	5	328	36	74	134	5	242	35	74	112
80 %	4	222	30	61	116	4	166	30	61	95
75 %	3	151	26	52	101	4	114	25	51	82

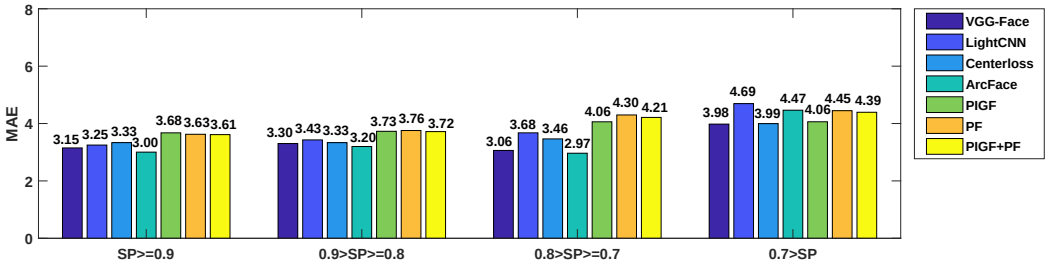
75%) of explained variance for PCA are analyzed.

This experiment is conducted on FIW-BMI and selected Morph II dataset, respectively. And the details about the training and test sets are given in Tables 4.2 and 4.3. Table 4.12 presents the MAEs of applying PCA to the five facial representations for BMI estimation. Here the reported MAE is the best performance of each representation among the different percentages of explained variance. Comparing the MAEs of facial representations without applying PCA as given in Table 4.4 and 4.5, we mark each MAE with a sign indicating the positive or negative effect of applying PCA to the representation. More specifically, a downward arrow (↓) denotes the MAE decreases (positive effect), and an upward arrow (↑) denotes the MAE increases (negative effect). It can be seen that VGG-Face+PCA performs better than VGG-Face in all groups on both databases. Centerloss+PCA achieves lower MAEs than Centerloss in the male group (BMI analysis database) and the white female group (Morph II). Arcface+PCA achieves lower MAE only in the white female group (Morph II). While applying PCA to LightCNN and PF representations does not bring any positive effect. Such different changes observed in the five facial representations caused by the different feature redundancy. Thereby, it is concluded that removing the redundancy in VGG-Face representation can increase the accuracy and efficiency in BMI estimation.

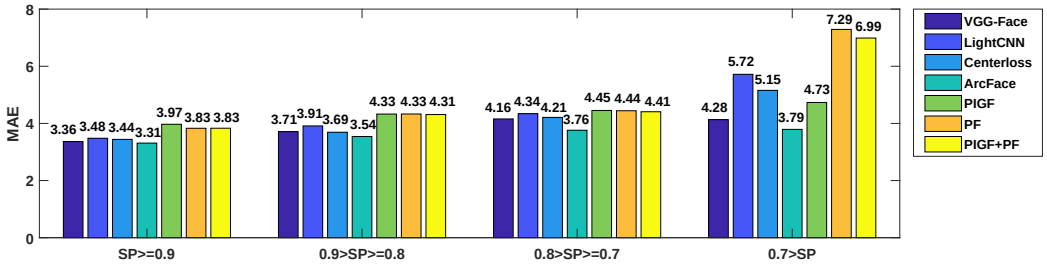
More details about BMI estimation performance (MAEs) obtained by applying different percentages of explained variance are shown in Fig. 4.7 (FIW-BMI database) and Fig. 4.8 (Morph II). The horizontal axis denotes the percentages of explained variance. We conduct the experiment by seven different percentages: 99%, 98%, 95%, 90%, 85%, 80% and 75%, respectively. Here 100% denotes the facial representation without applying PCA. It can be seen that the curve denoted VGG-Face+PCA drops obviously after a short rise, while most of the other curves are in a gradual uptrend. The best performance of VGG-Face is obtained at 80%−85% of the explained variance. Tables 4.13 and 4.14 report the kept dimensions of the five facial representations after applying PCA based on different percentages of explained variance on the two databases, respectively.

4.4.5 Sensitivity to head pose variations

The performance of face recognition is related to head pose changes. However, the influence of head pose on visual BMI estimation has not been studied yet. BMI estimation influenced by various head poses is conducted on the FIW-BMI dataset. As mentioned in Section 4.3.1, head pose variations exist in this database. To benchmark the robustness



(a)



(b)

Figure 4.9: The sensitivity of facial representations to invariant head pose. (a) shows the performance on the male group, and (b) shows the performance on the female group.

of the seven facial representations against pose variations, we group the face images by head pose angles.

Face pose distortion based sample pose (SP) index proposed by Marsico et al. [113] is utilized for measuring the head pose angles. SP index is given by the linear combination of three components, which are inversely proportional to $roll$, yaw, and $pitch$, respectively:

$$SP = \alpha(1 - roll) + \beta(1 - yaw) + \gamma(1 - pitch), \tag{4.3}$$

with $\alpha = 0.1$, $\beta = 0.6$ and $\gamma = 0.3$. See details about the calculation for $roll$, yaw, and $pitch$ in [113], whose ranges are from 0 to 1, where 0 means almost no distortion and 1 means the worst distortion. Thereby, large SP represents small head pose and vice versa.

This experiment is conducted on FIW-BMI database. The dataset is divided as shown in Table 4.2. The number of images of the test set for each range of SP values is given in Table 4.15. The obtained MAEs of the seven facial representations for BMI estimation with various head poses on FIW-BMI database are shown in Fig. 4.9. The values of SP index are divided into four intervals: $SP >= 0.9$, $0.9 > SP >= 0.8$, $0.8 > SP >= 0.7$ and $SP < 0.7$. It can be seen that when the SP decreases (head pose increases), the MAEs of the seven facial representation all increase, except the VGG-Face and Arcface

Table 4.15: The number of images for each range of SP values in the test set of FIW-BMI database.

	Male	Female
$SP >= 0.9$	464	307
$0.9 > SP >= 0.8$	468	57
$0.8 > SP >= 0.7$	109	202
$SP < 0.7$	20	3

representations on the male group in the interval $0.8 > SP >= 0.7$. This experimental result demonstrates that large head pose changes lead to low performance for both geometric based and deep learning based representations. Thus the visual BMI estimation can be further improved by employing efficient pose normalization approaches.

It is interesting to observe that the VGG-Face and Arcface perform better on the range from 0.7 to 0.8 than on higher SP ranges in the male group. While such a phenomenon does not exist in the performance of the other two deep features (Centerloss and LightCNN). This may be caused by the different architectures of the four deep models and the different properties of the training sets. The VGG-Face and Arcface were trained on larger datasets that contain more pose conditions. In addition, the VGG-Face and Arcface have more sophisticated architectures which may lead to richer representations.

Among the seven facial representations, the Arcface, VGG-Face and PIGF show greater robustness than other representations w.r.t head pose variations, since the MAEs increase much less than the others. LightCNN, PF and PIGF+PF show lower robustness to head pose variations, since their performances drop significantly with the decrease of SP value, especially when the SP values are smaller than 0.7.

4.4.6 Discussion

We discuss the two influence factors on the performance of facial representations. One is the BMI distribution on the dataset. Another is the accuracy of landmark detection.

Influence of BMI distribution on the estimation: As shown in Figs. 4.4 and 4.6, there is unbalanced BMI distribution over the two datasets. Very few samples distribute on the underweight category (BMI$\leq$18.5), while most samples distribute on the normal and overweight categories. This phenomenon also exists in real life. Most people are in normal and overweight ranges. To analyze the influence of unbalanced data on the estimated BMIs, we conduct an experiment on a balanced dataset. 5556 images were

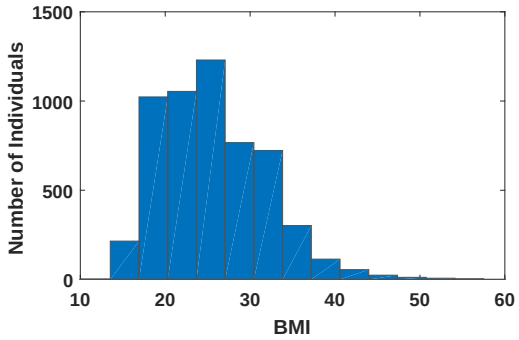


Figure 4.10: The BMI distribution of the balanced dataset.

Table 4.16: MAE of estimated BMIs on the balanced dataset of selected Morph II.

	Black male	Black female	White male	White female
PIGF	3.91	4.87	3.53	3.92
PF	3.87	4.81	3.47	4.03
PIGF+PF	3.85	4.75	3.47	3.95
VGG-Face	3.43	4.22	2.90	3.54
LightCNN	3.50	4.38	3.43	3.99
Centerloss	3.43	4.46	2.91	3.43
Arcface	3.44	4.13	2.93	3.51

selected from the Morph II database. Among the selected images, 893 are underweight, 1788 are normal, 1505 are overweight and 1370 are obese. Fig. 4.10 shows the BMI distribution over the selected dataset. Comparing with the BMI distribution in Fig. 4.6, Fig. 4.10 shows a relatively balanced distribution (with a relatively higher portion for underweight and obese). Then the images are randomly split into training and test sets. The training set contains 4319 images, and the test set contains 1237 images. There is non-overlap of individuals between the training and test sets. Considering the size of the training set is small, we use mixed training without separating the four gender and ethnicity groups. The experimental results on this balanced dataset are given in Table 4.16. Comparing with the results shown in Table 4.5, the performance on the balanced test set becomes worse. The experimental results indicate that the performance of BMI estimation depends on the prior distribution of the training set and the specific properties of the test set.

Influence on accuracy by landmark detection: The three geometric facial representations are computed from the detected landmarks. Though the recently proposed facial landmark detection methods [111, 114] achieve quite a high accuracy with good resists

Table 4.17: MAE of estimated BMIs from 119 landmarks and 68 landmarks.

	Black male	Black female	White male	White female
119 landmarks	2.85	3.97	2.82	3.15
68 landmarks	2.63	3.65	2.68	3.12

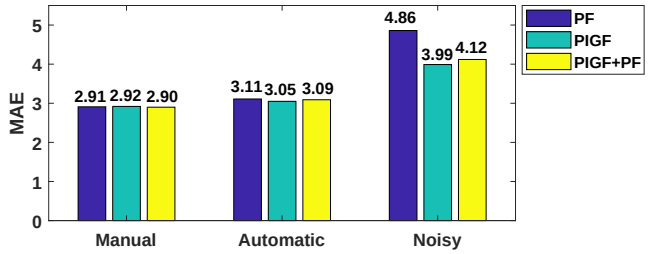


Figure 4.11: Influence of accuracy of landmark detection to geometric facial representations.

for low resolution, blur and noisy images, an evaluation of the influence on the accuracy of BMI analysis is still necessary. To report a fair evaluation, we generate three sets of data. First, we randomly select 100 images from the Morph II dataset (there is no head pose variations in this dataset), and manually label all the needed landmarks (68 landmarks) for each image. These manually labeled landmarks are used as the ground truth. Then we apply an automatic landmark detection by Openface toolkit [111] to the selected 100 images, with 68 landmarks detected for each image. Finally, we generated noisy landmarks by adding white Gaussian noise to the ground truth landmarks with the mean set to 3 pixels, variance set to 2 pixels. The BMIs are estimated from these three sets of data by the three geometric representations. The experimental results are presented in Fig. 4.11. One can see that the difference between the MAEs from manually labeled landmarks and the automatically detected are very small. While the performance degrades significantly on the noisy set. Among the three representations, PIGF shows relatively more robust to inaccurate and noisy landmarks. These results justify that the accuracy of landmark detection methods has the limited influence on geometric facial representations.

Finally, we study another interesting problem. Whether more landmarks could bring an improvement to BMI estimation? The performance of the PF feature with 119 landmarks [44] is analyzed on Morph II dataset, and compared with 68 landmarks. Fig. 4.12 shows an example of the detected 119 landmarks on a face image. The extended landmarks are around the neck, ears, forehead and around the vertex to the ears. The

Figure 4.12: An example of the detected 119 landmarks on a face image.

training and test sets are the same as shown in Table 4.3. Table 4.17 shows a comparison of the performance between 119 landmarks and 68 landmarks. As it can be observed, PF with 68 landmarks providing more promising results than 119 landmarks on each set. This reveals that the facial points around the neck, ears, forehead and the vertex to the ears are not as important as those around the face for estimating BMI values.

4.5 Summary

This chapter studies the visual BMI estimation problem systematically based on facial representation or feature extraction. According to the inherent properties of representations, they are grouped into two types: geometric based and deep learning based. In addition to the two existing approaches (VGG-Face and PIGF), five other facial approaches: PF, PIGF+PF, LightCNN, Centerloss and Arcface are explored for the first time for BMI analysis. The performance and characteristics of the two types of facial representations have been comprehensively evaluated and analyzed from three perspectives: the overall performance on visual BMI prediction, the redundancy in representations and the sensitivity to head pose changes. The experiments are conducted on two databases: FIW-BMI and Morph II, exploring the capability of these approaches, which are summarized below.

Experimentally we have found that the geometric representations are more suitable for the small dataset while the deep representations could perform better on large datasets with a much higher computation time cost. Among the seven representations, the VGG-Face and Arcface perform better than the others in most cases. For geometric features, more advantages can be achieved by the fused representation, PIGF+PF. The performance of the representations could be influenced by the training images and the BMI distribution.

Considering the limited number of training samples and high dimensions of some facial representations, we explored the efficient methods to improve the performance. We have analyzed the redundancy of the five facial representations (VGG-Face, LightCNN, Center-loss, Arcface and PF) by investigating the effect of applying PCA to the representations. Experimental results have shown that applying PCA to VGG-Face representation leads to better performance on BMI prediction with $80\%-85\%$ explained variance. Removing the redundancy in VGG-Face representation can increase the accuracy and efficiency in BMI estimation.

The sensitivity of facial representations to head pose variations for BMI estimation has been investigated as well. Experimental results have shown that large head pose changes lead to low performance. Among the seven representations, The Arcface, VGG-Face and PIGF show better robustness than the others to head pose variations. The performance of LightCNN, PF and PIGF+PF drop significantly with the increase of head pose angles.

Chapter 5

Visual BMI Estimation using Label Distribution based Method

In this chapter, we investigate the problem of visual BMI estimation from facial images by a two-stage learning framework. BMI related facial features are learned from the first stage. Then a label distribution based BMI estimator is learned by an optimization procedure that is implemented by projecting the features and assigned labels to a new domain which maximizing the correlation between them. Two label assignment strategies are analyzed for modeling the single BMI value as a discrete probability distribution over the whole ranges of BMIs. Extensive experiments are conducted on FIW-BMI and Morph II datasets. The experimental results show that the two-stage learning framework improves the performance step by step. More importantly, the proposed estimator efficiently reduces the estimated error and outperforms other regression and label distribution methods.

The organization of this chapter is as the following. Section 5.1 describes the challenge existing in BMI estimation from facial images. The details of the proposed method for BMI estimation are presented in Section 5.2. Section 5.3 describes the two databases used in this work: Morph II and FIW-BMI. Extensive experiments, detailed analysis and discussion are reported in Section 5.4. Finally, the conclusion is summarized in Section 5.5.

5.1 Introduction

Recent research shows that facial adiposity is associated with perceived health and is important for body mass index (BMI) prediction [23, 53]. As a body fat indicator, BMI

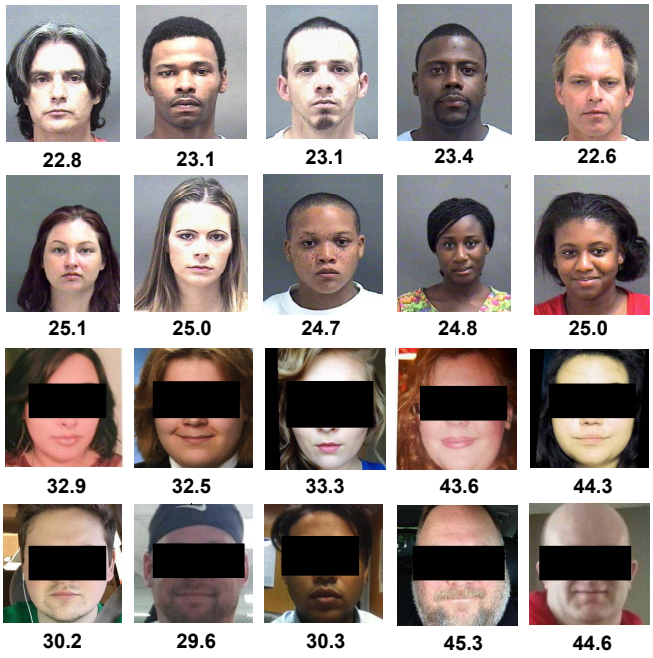


Figure 5.1: Samples from Morph II and FIW-BMI dataset with corresponding BMI values.

is widely used in health monitoring and health research. There are close connections between BMI and some diseases, such as cancers, unstable angina and type 2 diabetes and cardiovascular disease (CVD), etc [9, 10, 13]. Generally, BMI is measured in person with special devices. Therefore, automatically estimating BMI from facial images is a great benefit to health monitoring and researchers who are interested in studying obesity in large populations.

BMI estimation from facial images is a challenging problem in computer vision and pattern recognition. First, different from other human visual tasks, such as face recognition [115, 116], motion capture [117, 118] which have sufficient data for training and testing, it is difficult to collect a database covering images with all BMI values. Second, the distribution of BMIs on the database is uneven. According to the BMI values, there are mainly four BMI categories: underweight (BMI$\leq$18.5), normal (18.5<BMI$\leq$25), overweight (25<BMI$\leq$30), obese (BMI>30). Very few BMIs distribute on underweight and severe obese categories. Therefore, it is hard to ensure each category have enough associated images. Currently, the number of public databases for visual BMI study is limited. This work uses two databases: Morph II [51] and FIW-BMI [66]. Finally, the BMI label is an ambiguous label. e.g., one person looks like with BMI around 25 which means that some neighbor values (24.5 or 25.5) can also be used to describe this person;

and some people may look lower than their real BMI, while others may look higher than their real BMI. Fig. 5.1 shows some facial images from Morph II and FIW-BMI dataset with corresponding BMI values. We can see some samples with the same gender and the adjacent BMI values but have different facial appearances.

Single label estimation assumes one image has one label. Regression based method directly predicts the label from the images which ignores the label ambiguity existing in images. In order to describe the label ambiguity associated with the images, a label distribution scheme is proposed by Geng et al. [119] to describe such ambiguity. Later on, several distribution learning based approaches have been proposed for age estimation and other tasks. These methods utilized label correlation or entropy model to solve the problem. [119] proposed two label distribution based algorithms named IIS-LLD and CPNN. Comparing with other single label methods, their methods showed good performances. A multivariate label distribution (MLD) based method was also proposed by Geng et al. [120] for further improving the performance on head pose estimation. In addition, Xing et al. [121] used Logistic Boosting Regression (LogitBoost) to learn a general label distribution model family which can avoid the potential influence of the specific model.

Some work explored regression based methods for visual BMI estimation. Wen and Guo [47] proposed a computational method for automatically predicting BMI from 2D face images. This is the first work on visual BMI estimation from facial images. Kocabey et al. [61] employed the pre-trained VGG-Face model [62] to extract facial representation for BMI estimation. Then a support vector regression model is learned to map the facial representation to predicted BMIs. The above two works treated BMI prediction as a regression problem. The performance of such methods may be influenced by outliers. Recently, convolution neural networks (CNN) have shown promising performance in many applications [55, 59, 122]. A recent work using CNN for BMI estimation is proposed by Dantcheva wt al. [63], where estimating height, weight, and BMI from single-shot facial images by a regression method based on the 50-layers ResNet-architecture.

Different from the above work, this work addresses the visual BMI estimation problem by a label distribution based method. Particularly, a two-stage learning framework is shown in Fig. 5.2. First, the BMI related facial representation is learned by fine-tuning the pre-trained deep face model. This step is expected to obtain sufficient visual BMI characteristics and reinforce the learning process using the limited number of BMI data. More importantly, the label distribution method models the single BMI value as a discrete

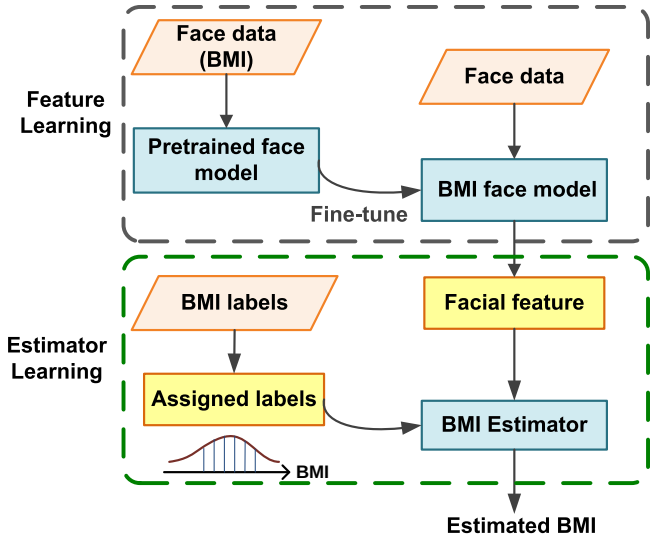


Figure 5.2: The pipeline of two-stage learning framework: BMI related feature learning and the BMI estimator learning.

probability distribution over the whole ranges of BMIs. Given the extracted facial features from the first stage, a BMI estimator is trained by an optimization procedure which is applied to the features and the assigned distribution labels. The main contributions of this work include:

1. A two-stage learning framework is presented to address the visual BMI estimation problem from face images.

2. A label distribution based learning method which regards each BMI label as a discrete probability distribution is proposed to learn the BMI estimator.

3. Two distribution strategies are analyzed to model BMI labels. The output can either be a discrete probability distribution or a single value.

5.2 Method

Fig. 5.2 depicts the two-stage learning framework, which contains BMI related facial features learning and BMI estimator learning. The BMI related face model is learned based on a pre-trained deep face model. Then two different strategies are analyzed for assigning the distributed BMI labels. And a projection optimization is obtained by maximizing the correlation between the facial features and the assigned labels. Finally,

the BMI estimator is learned from the projected features and assigned labels. Below the detailed procedure and derivation are presented.

5.2.1 Deep model for BMI related facial feature

1) Face model: Representing face structure using a pre-trained face model. We utilize the feature extracted from publicly released Centerloss face model [107]. This network improves the discriminative power of the learned features by using a new Centerloss function to minimize the intra-class variations while keeping features of different classes separable. This model performs impressively in face recognition tasks which achieved face verification accuracy of 98.28% on LFW, 94.9% on YTF. fc5 layer of Center loss model C is used to extract facial features.

2) Fine-tuning: Adapting from general facial structure model to BMI related face model. Our goal is to estimate BMI values from face images. We tune the pre-trained Centerloss face model to the BMI face model before feature extraction. FIW-BMI dataset is used to fine-tuning the deep model. The aim of this step is to learn sufficient BMI related facial structure. We replace the original softmax loss and centerloss functions in Centerloss network with Euclidean loss function during the fine-tuning process. Euclidean loss function E computes the sum of squares of differences between two inputs, which can be written as:

$$E = \frac{1}{2N} \sum_{i=1}^{N} \|\hat{y}_i - y_i\|^2 , \tag{5.1}$$

where N is the number of samples, $\hat{y}_i$ is the output from the network and y_i is the true BMI value. After the above steps, the fine-tuned face model is expected to have the capability to capture more BMI related facial structures.

5.2.2 Modelling BMI values with label distribution

BMI value is labeled by a single real number. BMI estimation from facial images is different from other traditional regression tasks because there is ambiguous information among BMI labels. Based on this observation, given an image labeled with the BMI value b, the BMI value is transformed to discrete probabilities distribution $\mathbf{P} = [p_1, p_2, ..., p_k]^T \in \mathbb{R}^k$ over the whole range of BMIs. Inside of the BMI range, every BMI value could be a possible label to describe true BMI with different confidence. A similar definition is proposed by [119].

We assume that the BMI range is from 15 to 60 in this work according to the BMI distribution of the two databases. Two strategies are investigated for modeling the single

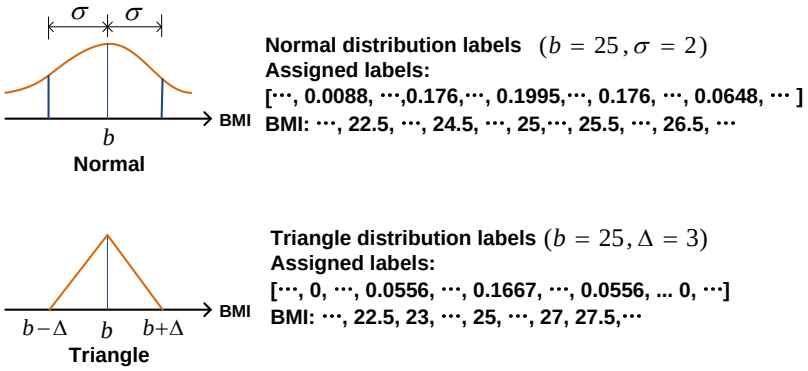


Figure 5.3: Two strategies for BMI label distribution.

BMI value, namely normal distribution, and triangle distribution, which are shown in Fig. 5.3.

Specifically, denoting the BMI value as b. For a normal distribution, the BMI labels are modeled as a Gaussian distribution centered at b. It can be computed by:

$$p(z_i) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(z_i - b)^2}{2\sigma^2}), \quad 15 \leq z \leq 60, \tag{5.2}$$

where σ is the standard deviation of the Gaussian distribution, and $\mathbf{Z} = [z_1, z_2, ..., z_k]^T \in \mathbb{R}^k$ is a set of discrete values from 13 to 60 with interval of 0.1. For example, if the range of BMI is from 15 to 60, then $\mathbf{Z} = [15, 15.2, ..., 59.8, 60]$. Note that k can be adjusted with different BMI ranges and intervals. Thereby, $p(z_i)$ denotes the probability that is corresponding to the BMI value z_i.

For a triangle distribution, the neighbor BMIs are considered with a length of Δ on each side of the BMI value, here $\mathbf{Z} = [z_1, z_2, ..., z_k]^T \in \mathbb{R}^k$ is a set of discrete values from 15 to 60 with interval of 0.1. The probability function $p(z_i)$ is defined as:

$$p(z_i) = \begin{cases} \frac{\Delta - b + z_i}{\Delta^2}, & \text{if } b - \Delta \leq z_i \leq b \\ \frac{\Delta + b - z_i}{\Delta^2}, & \text{if } b \leq z_i \leq b + \Delta \\ 0, & \text{otherwise} \end{cases} \tag{5.3}$$

A normalization process is applied to the assigned labels, which defined as: $y_i = \frac{p(z_i)}{\sum_{i=1}^{k} p(z_i)}$. This leads to a discrete range of BMIs with different levels of "probabilities".

5.2.3 Learning with assigned label

The BMI estimator is learned by a label optimization procedure based on the correlation. The optimization procedure is implemented by projecting the features and assigned

labels to a new domain which maximizes the correlation between them. Then the BMI estimator is learned from the projected features and labels as the least square problem.

Given N training samples, denote $\mathbf{x_i} = \left[x_i^1, x_i^2, ..., x_i^d\right]^T \in \mathbb{R}^d$, $\mathbf{y_i} = \left[y_i^1, y_i^2, ..., y_i^k\right]^T \in \mathbb{R}^k$ as the feature vector and assigned distribution label of i-th training sample, respectively. Here d is the dimension of the feature vector $\mathbf{x_i}$, and k is the length of the assigned label $\mathbf{y_i}$. We assume that both $\mathbf{x_i}$ and $\mathbf{y_i}$ are centered, i.e., $\sum_{i=1}^N \mathbf{x_i} = 0$ and $\sum_{i=1}^N \mathbf{y_i} = 0$. Denote $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}] \in \mathbb{R}^{d\times N}$, $\mathbf{Y} = [\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_N}] \in \mathbb{R}^{k\times N}$, the main idea of canonical correlation analysis (CCA) [123] is to project the two sets of variables into latent variables (a new domain), such that the correlation ρ between them is maximized, which can be written as:

$$\rho = \max_{\mathbf{w_X}, \mathbf{w_Y}} \frac{\mathbf{w_X}^T C_{\mathbf{XY}} \mathbf{w_Y}}{\sqrt{\mathbf{w_X}^T C_{\mathbf{XX}} \mathbf{w_X} \mathbf{w_Y}^T C_{\mathbf{YY}} \mathbf{w_Y}}}, \tag{5.4}$$

here $\mathbf{w_X}$ and $\mathbf{w_Y}$ are projection vectors. Observe that the solution of Eqn. (5.4) is invariant to re-scaling $\mathbf{w_X}$ or $\mathbf{w_Y}$ either together or independently:

$$\frac{\alpha \mathbf{w_X}^T C_{\mathbf{XY}} \mathbf{w_Y}}{\sqrt{\alpha^2 \mathbf{w_X}^T C_{\mathbf{XX}} \mathbf{w_X} \mathbf{w_Y}^T C_{\mathbf{YY}} \mathbf{w_Y}}} = \frac{\mathbf{w_X}^T C_{\mathbf{XY}} \mathbf{w_Y}}{\sqrt{\mathbf{w_X}^T C_{\mathbf{XX}} \mathbf{w_X} \mathbf{w_Y}^T C_{\mathbf{YY}} \mathbf{w_Y}}}. \tag{5.5}$$

The solution of Eqn. (5.4) is only related to the direction of the two projection vectors $\mathbf{w_X}$ and $\mathbf{w_Y}$. To obtain a unique solution, the constraints are added. Thereby the CCA is equivalent to maximizing the numerator with the constraints:

$$\begin{aligned} &\max_{\mathbf{w_X}, \mathbf{w_Y}} \ \mathbf{w_X}^T \mathbf{X} \mathbf{Y}^T \mathbf{w_Y}, \\ &s.t. \ \mathbf{w_X}^T \mathbf{X} \mathbf{X}^T \mathbf{w_X} = 1, \ \mathbf{w_Y}^T \mathbf{Y} \mathbf{Y}^T \mathbf{w_Y} = 1. \end{aligned} \tag{5.6}$$

The corresponding Lagrangian is:

$$\begin{aligned} L(\lambda, \mathbf{w_X}, \mathbf{w_Y}) =& \mathbf{w_X}^T C_{\mathbf{XY}} \mathbf{w_Y} - \frac{\lambda_X}{2}(\mathbf{w_X}^T C_{\mathbf{XX}} \mathbf{w_X} - 1) \\ &- \frac{\lambda_Y}{2}(\mathbf{w_Y}^T C_{\mathbf{YY}} \mathbf{w_Y} - 1). \end{aligned}$$

Taking derivatives of $\mathbf{w_X}$ and $\mathbf{w_Y}$, respectively:

$$\frac{\partial L}{\partial \mathbf{w_X}} = C_{\mathbf{XY}} \mathbf{w_Y} - \lambda_X C_{\mathbf{XX}} \mathbf{w_X} = \mathbf{0}, \tag{5.7}$$

$$\frac{\partial L}{\partial \mathbf{w_Y}} = C_{\mathbf{YX}} \mathbf{w_X} - \lambda_Y C_{\mathbf{YY}} \mathbf{w_Y} = \mathbf{0}. \tag{5.8}$$

Then subtracting $\mathbf{w_Y}^T$ multiplies Eqn. (5.8) from $\mathbf{w_X}^T$ multiplies Eqn. (5.7):

$$\begin{aligned} \mathbf{0} &= \mathbf{w_X}^T \left(C_{\mathbf{XY}} \mathbf{w_Y} - \lambda_X C_{\mathbf{XX}} \mathbf{w_X}\right) - \mathbf{w_Y}^T \left(C_{\mathbf{YX}} \mathbf{w_X} - \lambda_Y C_{\mathbf{YY}} \mathbf{w_Y}\right) \\ &= \lambda_Y \mathbf{w_Y}^T C_{\mathbf{YY}} \mathbf{w_Y} - \lambda_X \mathbf{w_X}^T C_{\mathbf{XX}} \mathbf{w_X}. \end{aligned}$$

Taking into account the constraints $\mathbf{w_X}^T\mathbf{X}\mathbf{X}^T\mathbf{w_X} = 1$ and $\mathbf{w_Y}^T\mathbf{Y}\mathbf{Y}^T\mathbf{w_Y} = 1$, we can obtain that $\lambda_Y - \lambda_X = 0$. Let $\lambda = \lambda_Y = \lambda_X$ and assuming $C_{\mathbf{YY}}$ is invertible, we can obtain:

$$\mathbf{w_Y} = \frac{C_{\mathbf{YY}}^{-1} C_{\mathbf{YX}} \mathbf{w_X}}{\lambda}. \tag{5.9}$$

substituting Eqn. (5.9) into Eqn. (5.7):

$$C_{\mathbf{XY}} C_{\mathbf{YY}}^{-1} C_{\mathbf{YX}} \mathbf{w_X} = \lambda^2 C_{\mathbf{XX}} \mathbf{w_X}. \tag{5.10}$$

Now Eqn. (5.10) is a generalized eigenvalue problem of the form $Ax = \lambda Bx$. $\mathbf{w_X}$ can be obtained via solving the following generalized eigenvalue problem. To avoid the singularity problem of $\mathbf{YY^T}$ and $\mathbf{XX^T}$, we adopt regularized CCA to get $\mathbf{w_X}$ by the following form:

$$C_{\mathbf{XY}}(C_{\mathbf{YY}} + \eta_y \mathbf{I})^{-1} C_{\mathbf{YX}} \mathbf{w_X} = \lambda^2 (C_{\mathbf{XX}} + \eta_x \mathbf{I}) \mathbf{w_X}. \tag{5.11}$$

Let $\mathbf{W} = [\mathbf{w_1}, \mathbf{w_2}, ..., \mathbf{w_q}]$ denotes the matrix of top q eigenvectors of the generalized eigenvalue problem. Here $\mathbf{W}$ is the projection vector which is used to project $\mathbf{X}$ into a new domain, such that the correction ρ is maximized. For each original feature vector $\mathbf{x} \in \mathbb{R}^d$, we obtain the new representation $\mathbf{x}^{CCA} = \mathbf{W}^T\mathbf{x}$.

After obtaining the new representation of all N training samples $\mathbf{x_i}^{CCA} = \mathbf{W}^T\mathbf{x_i}$, we can obtain the BMI distribution by solve the following least square problem:

$$\min_B \sum_{i=1}^{N} \left\| \mathbf{x_i}^T \mathbf{W}\mathbf{B} - \mathbf{y_i}^T \right\|, \tag{5.12}$$

where $\mathbf{B} \in R^{q \times k}$ is a coefficient matrix, which can be shown that the solution to Eqn. (5.12) is:

$$\mathbf{B} = \left(\mathbf{X}^T\mathbf{W}\right)^+ \mathbf{Y}^T, \tag{5.13}$$

where $\left(\mathbf{X}^T\mathbf{W}\right)^+$ denotes the pseudoinverse of $\mathbf{X}^T\mathbf{W}$. Given a test sample (feature) $\mathbf{x_t}$, the corresponding estimated assigned label distribution can be obtained by:

$$\hat{\mathbf{y}} = \mathbf{B}^T\mathbf{W}^T\mathbf{x_t}, \tag{5.14}$$

here $\hat{\mathbf{y}} = \left[\hat{y^1}, \hat{y^2}, ..., \hat{y^k}\right]$ is a vector denotes the predicted probabilities distribution. $\hat{y^i}$ is a factor in vector $\hat{\mathbf{y}}$ which denotes the predicted probability belongs to the BMI label l^i. Then the estimated BMI value $\hat{b}$ is computed by:

$$\hat{b} = \sum_{i}^{k} \hat{y^i} l^i. \tag{5.15}$$

$\mathbf{l} = \left[l^1, l^2, ..., l^k\right]$ is a set of discrete values from the whole range of BMIs $(13 - 60)$ with the interval of 0.1.

Table 5.1: Characteristic of FIW-BMI dataset. Mean and standard deviations pertained to BMI for male and female.

	Male	Female
#Subjects	3192	1689
#Images	5197	2733
Mean	30.7	31.2
Std	7	6.9

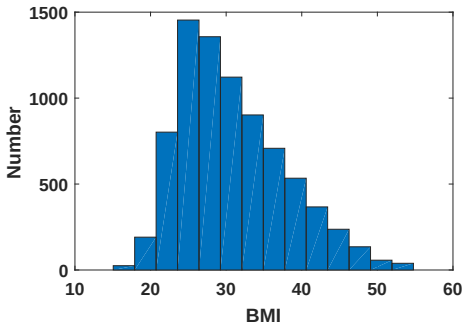


Figure 5.4: Distribution of BMI values on BMI-analysis face database. The BMI values span a wide range with most of the values distribute between 20 to 50.

5.3 Dataset

We conduct extensive experiments on two databases. First, FIW-BMI dataset [66] is used to fine-tune the deep face model. Then, Morph II database [51] is utilized to evaluate the effectiveness of the proposed method.

Face in the wild for BMI analysis (FIW-BMI) dataset: It contains 7930 images from 4881 individuals, along with the corresponding gender, height and weight information. Among these individuals, there are 3192 males and 1689 females. Each individual has 1 to 4 images. Details about the dataset are described in Table 5.1. It is separated into two groups by gender. The same individual does not exist in both training and test sets. Most images in this dataset are collected from a social network−Reddit posts [1]. Because this is a social network displaying people's progress of weight loss, weight gain, or essentially any type of body transformation, the BMI values of these images distribute in a very wide range from 15 to 60 as shown in Fig. 5.4. Comparing with the distribution of BMI values on Morph II database (as shown in Fig. 5.5), this database has a much wider BMI distributed range. Thereby we select it to tune the deep face model.

Morph II database: It contains 55,608 passport-style frontal face images along with

[1]Website: `http://www.reddit.com/r/progresspics`

Table 5.2: Characteristic of selected data from Morph II. Mean and standard deviations pertained to BMI for four gender and ethnicity groups.

	Black male	Black female	White male	White female
#Subjects	6497	1096	1565	535
#Images	19290	2824	4862	2057
Mean	25.0	25.2	24.6	22.8
Std	4.4	6.0	4.0	5.5

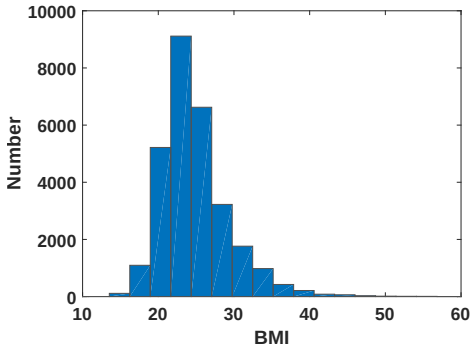


Figure 5.5: Distribution of BMI values on Morph II. The BMI values mainly distribute between 15 to 35.

age, gender and ethnicity information. Moreover, there are 40,330 images have height and weight information. Considering the uneven distribution of the ethnicity in the database, only images from Black and White are used for this work. There are 29033 images kept. Details about the selected data are described in Table 5.2. The same individual does not exist in both the training and test set. Most images from the same individual have different BMI values. The BMI values of Morph II mainly distribute in the range of 15 to 35. Among these, 893 are underweight, 16,582 are normal, 8,237 are overweight and 3,321 are obese. In this work, this dataset is used for evaluating the methods.

5.4 Experiments

The performance metrics are introduced in this section. Then the experimental setting and results are presented in detail.

5.4.1 Performance metrics

Two measure metrics are utilized for evaluating the performance of BMI estimators. The first one is mean absolute error (MAE) which is motivated by that used in age

Table 5.3: Performance of the five estimation methods on Morph II database based on separated training in each of the gender and ethnicity groups.

		Black male		Black female		White male		White female	
Feature	Method	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy
Face model	SVR	2.47	75.7%	3.67	66.8%	2.48	73.8%	2.89	74.4%
	PCA-SVR	2.50	75.7%	3.73	66.8%	2.50	73.8%	2.86	73.5%
	GPR	2.54	75.8%	3.73	66.3%	2.52	75.4%	3.09	75.1%
	PLS	2.52	76.5%	4.76	54.1%	2.82	70.4%	4.64	51.8%
	CCA	2.51	76.3%	4.42	56.1%	2.78	70.9%	4.05	60.7%
Fine-tuned	SVR	**2.45**	76.7%	**3.40**	**68.6**%	2.37	76.2%	**2.78**	75.8%
	PCA-SVR	2.57	75.7%	3.56	67.8%	**2.34**	**78.1**%	3.47	69.0%
	GPR	2.53	**76.8**%	3.62	68.3%	2.38	76.7%	2.79	**78.1**%
	PLS	2.47	76.5%	4.54	56.2%	2.71	73.0%	4.27	58.4%
	CCA	2.46	76.6%	4.36	55.7%	2.67	74.2%	3.84	62.9%

Table 5.4: BMI estimation results using label distribution based method on Morph II database.

		Black male		Black female		White male		White female	
Feature	Method	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy
Face model	LD-PLS	2.49	77.0%	3.49	66.0%	2.48	74.7%	2.96	71.5%
	LD-CCA	2.42	76.5%	3.50	67.1%	2.38	**77.1**%	2.86	**75.9**%
Fine-tuned	LD-PLS	2.41	76.6%	3.69	59.4%	2.54	74.3%	2.98	72.0%
	LD-CCA	**2.35**	**77.0**%	**3.40**	**67.3**%	**2.25**	75.6%	**2.72**	73.8%

estimation, e.g., [103]. It is defined as the average of the absolute error between the estimated BMIs and the ground truth BMIs: $MAE = \frac{1}{N}\sum_{j=1}^{N}\left|\hat{b}_j - b_j\right|$, here b_j is the ground truth BMI for image j, $\hat{b}_j$ is the estimated BMI, N is the number of test images. Another metric is the accuracy of the predicted category. According to the estimated BMI, one can predict the image belong to which category (underweight, normal, overweight or obese). The accuracy of the predicted category is the proportion of the total number of predictions that are correct.

It should be noted that the two metrics have advantages and limitations. e.g. Given an image with true BMI value is 24, the estimated value is 19. Though the absolute error is 5, the predicted category (normal) is correct. While if the true BMI of an image is 30 and the estimated value is 30.5, though the absolute error is 0.5, the predicted category (obese) is incorrect. Thereby we combine them together to evaluate the performance of each method.

5.4.2 Experimental settings

1) Data preprocessing: Given the images, we first applied face detection and landmark localization using Openface toolkit [111]. Then the images are aligned by the eye locations and cropped with the size of a 96×112. In addition, two geometric facial BMI models-PIGF and PF are utilized to further evaluate the effectiveness of the proposed method. For these two geometric models, the face images are cropped with the size of 256×256.

1) Implementation details for BMI related feature learning: The fine-tuning of Centerloss network is implemented by the Caffe platform [112]. We fine-tune the network parameters using face images with BMI labels in FIW-BMI dataset. In this fine-tuning step, we used mini-batch stochastic gradient descent (SGD) with momentum settings. The mini-batch size is set to 64 and momentum is set to 0.9. We initialize the learning rate to 0.00001. The learning rate decreases in polynomial decay with a power of 0.1. The training procedure stops after 20000 iterations. A feature vector of 512 dimensions is extracted from layer fc5 in Centerloss. As mentioned in [107] which extracts the features for each image and its horizontally flipped one, and concatenates them as the representation with the size of 1024×1.

2) Implementation details for evaluating different BMI estimators: After facial features being extracted, five estimators are learned: Support Vector Regression (SVR) [48], PCA-SVR, Gaussian Process Regression (GPR) [73], Canonical Correlation Analysis (CCA) [123], and Partial Least Square analysis (PLS). Considering the dimension of

Table 5.5: BMI estimation results using label distribution based method on Morph II database by geometric features.

		Black male		Black female		White male		White male	
Feature	Method	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy
PIGF	SVR	2.66	72.7%	3.73	65.8%	2.71	70.8%	2.96	70.7%
	GPR	2.72	**74.2**%	3.74	66.5%	2.74	**72.1**%	2.99	70.7%
	PLS	2.76	71.9%	3.81	67.1%	2.74	71.2%	3.15	**74.5**%
	CCA	2.77	71.8%	3.77	67.1%	2.74	71.3%	3.14	74.5%
	LD-CCA	**2.62**	72.4%	**3.56**	66.6%	**2.62**	71.6%	2.96	73.4%
	LD-PLS	2.64	72.5%	3.61	**67.7**%	2.70	71.1%	**2.87**	72.0%
PF	SVR	2.63	73.6%	3.65	68.3%	2.68	70.8%	3.12	72.5%
	GPR	2.68	**75.0**%	3.79	**68.4**%	2.79	**71.0**%	3.09	73.8%
	PLS	2.73	73.1%	3.69	66.5%	2.73	70.9%	3.37	72.0%
	CCA	2.71	73.4%	3.68	66.8%	2.71	70.9%	3.38	72.0%
	LD-CCA	**2.57**	73.7%	3.52	67.2%	**2.58**	69.5%	**2.91**	**76.9**%
	LD-PLS	2.61	73.5%	**3.50**	65.9%	2.64	69.4%	3.02	75.2%

the deep facial feature, principal component analysis (PCA) is applied to the features before training the SVR, which denoted as PCA-SVR. The PCA percentage of explained variance for different SVR is various, but all selected based on the best performance. In our implementation, SVR is trained with RBF kernel, and GPR is trained with the rational quadratic kernel. The parameters for each SVR and GPR lead to the best predicted result are utilized.

3) Implementation details for label distribution based estimator: For label distribution based method, the first step is to convert the BMI labels to distribution labels as described in Section 5.2.2. Particularly in this implementation, the corresponding BMI range is from 13 to 60 with the interval of 0.1. The assigned labels are calculated according to the true BMI value. With the label distribution based method, we train the BMI estimator which are named LD-CCA and LD-PLS.

5.4.3 Experimental results

1) Evaluation of BMI estimation based on deeply learned BMI features: First, we conduct an evaluation based on features extracted from the pre-trained Centerloss face model and fine-tuned BMI related face model. In order to explore the performance based on different BMI estimators using the deep features, we conduct experiments using five different estimators: SVR, PCA-SVR, GPR, PLS and CCA. Facial features extracted

from the deep network are fed into the estimators for training thus they are more suitable for the BMI estimation. Tables 5.3 presents the experimental results. All the results are obtained based on separated training and testing in each of the gender and ethnicity groups. Face model means the feature directly extracted from the pre-trained Centerloss model. Fine-tuned means the feature extracted from the fine-tuned facial BMI model. As described before, FIW-BMI dataset is used to fine-tuning the deep model. From the results, one can see that the performance (MAEs and the accuracy of predicted category) based on fine-tuned model are all better than face model, which shows that fine-tuning the deep model using facial BMI data derives more robust representations for BMI estimation. No matter which estimator is used, the errors are all reduced on the test set. This demonstrates that after fine-tuning the deep face model is more capable of capturing BMI related facial features.

Since all the results presented in Table 5.3 are obtained from the estimators trained by original BMI labels (single BMI value for each face image), they will be compared with the experimental results from label distribution based methods.

2) Experimental results on label distribution based methods: Now we conduct the experiment using our proposed estimator for BMI estimation. The experiment uses the features extracted from Centerloss face model and fine-tuned facial BMI model. The results of applying label distribution based methods to the facial feature are presented in Table 5.4. The estimated BMI values are obtained based on separated training and testing in each of the gender and ethnicity groups. Note that the method is based on the normal label distribution model as mentioned in Section 5.2.2. The comparison of the performance between the two different label distribution strategies will be analyzed later.

Comparing with the results given in Table 5.3, LD-CCA and LD-PLS outperform the previous five estimators-SVR, PCA-SVR, GPR, PLS and CCA. More specifically, with the features extracted from the face model, MAEs of CCA and PLS are 2.51 and 2.52, respectively; while MAEs of LD-CCA and LD-PLS are 2.42 and 2.49, respectively. With the features extracted from the fine-tuned model, MAEs of CCA and PLS are 2.46 and 2.47, respectively; while MAEs of LD-CCA and LD-PLS are 2.35 and 2.42, respectively. This result shows the advantages of the proposed estimator when utilizing the label distribution schemes. Given one facial image, its corresponding label distribution consists of a series of probabilities. Each probability represents the confidence that the corresponding label describes the image. The label distribution scheme well defines the increase of BMI as a continuous process. With this scheme, one image not only

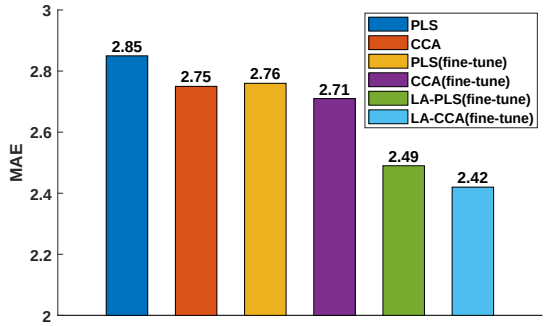


Figure 5.6: Comparison of overall MAE on Morph II database in each step of our proposed method.

contributes to the learning of its BMI but also provides auxiliary information to the learning of its adjacent BMIs.

To further evaluate the performance, Fig. 5.6 shows the overall MAE of the estimated BMI values in each step of the two-stage learning method. The overall MAE is calculated from the whole test set. From this figure, one can see that the BMI estimation error is reduced step by step using our proposed method on deeply learned representations. The MAE of applying CCA to Centerloss face feature is 2.75. After BMI feature learning by fine-tuning, it drops from 2.75 to 2.71. Then by applying LD-CCA to deeply learned BMI features, the MAE significantly drops to 2.42. This further demonstrates the effectiveness of the proposed two-stage learning method.

3) Experimental results using different label assignment strategies: As mentioned in Section 5.2.2, there are two schemes for modeling BMI values with label distributions: normal distribution and triangle distribution. We compare the performance of the two different modeling strategies. Three facial features are used in this experiment, they are Centerloss, PIGF and PF. Note that Centerloss feature is extracted from the fine-tuned BMI face model. The results are shown in Fig. 5.7. One can see that the normal distribution performs better than the triangle distribution in the three cases. This result indicates that the normal distribution is more appropriate for defining the BMI labels than triangle distribution. This may because the increase and decrease of facial BMI (adiposity) is not a uniform change. The correlation between facial appearance and BMI is related to age and gender [25]. This means with different age or gender the facial appearance variance caused by BMI are different. However, the triangle distribution describes the increase and decrease of facial BMI (adiposity) as a uniform change. In

Table 5.6: Comparison of BMI estimation using our method to other methods.

Feature	Method	Black male		Black female		White male		White female	
		MAE	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy
Deep feature	LDL-IIS	2.48	75.6%	**3.33**	**69.0%**	2.39	**76.5%**	2.89	73.4%
	LDL-CPNN	7.60	30.4%	4.20	60.1%	3.73	59.8%	4.66	54.0%
	LD-PLS (ours)	2.41	76.6%	3.69	59.4%	2.54	74.3%	2.98	72.0%
	LD-CCA (ours)	**2.35**	**77.0%**	3.42	67.3%	**2.25**	75.6%	**2.72**	**73.8%**
PIGF	LDL-IIS	2.99	**85.6%**	4.10	74.4%	3.08	77.2%	3.34	66.8%
	LDL-CPNN	2.98	80.0%	4.7	50.9%	3.19	**78.4%**	3.26	66.4%
	LD-PLS (ours)	2.64	72.5%	3.61	**67.7%**	2.70	71.1%	2.87	72.0%
	LD-CCA (ours)	**2.62**	72.4%	**3.56**	66.6%	**2.62**	71.6%	**2.69**	**73.4%**
PF	LDL-IIS	3.40	60.7%	4.50	72.4%	3.45	59.5%	3.75	64.1%
	LDL-CPNN	3.70	54.7%	5.24	24.5%	4.47	30.5%	6.45	13.4%
	LD-PLS (ours)	2.61	73.5%	**3.50**	**65.9%**	2.64	69.4%	3.02	75.2%
	LD-CCA (ours)	**2.57**	**73.7%**	3.52	67.2%	**2.58**	**69.5%**	**2.91**	**76.9%**

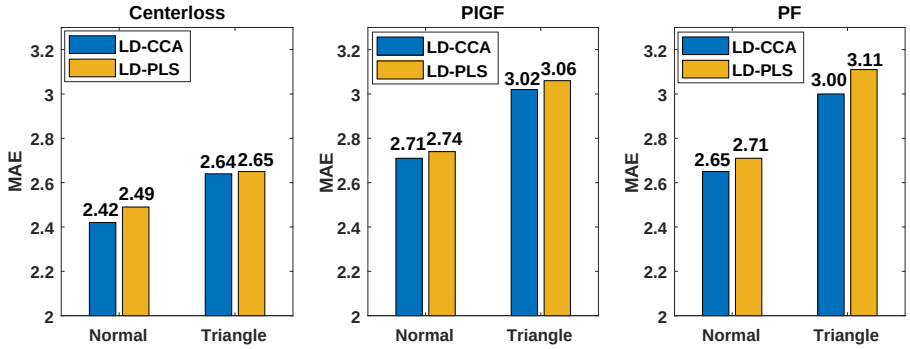


Figure 5.7: Comparison of BMI estimation results on Morph II dataset using different label distribution strategies.

addition, it can be observed that in most cases LD-CCA method shows better performance than LD-PLS by both two label assignment strategies.

To further analyze the parameters sensitivity of the two label assignment strategies, we evaluate their performances with various parameter settings. Fig. 5.8 shows the performance with the corresponding parameters. It can be seen that the best performance is achieved with setting σ to 4, Δ to 3 for LD-CCA, and setting σ to 5, Δ to 2 for LD-PLS.

4) Evaluate label distribution based estimators on geometric features: To further evaluate the effectiveness of the proposed label distribution based estimator, we apply it to two geometric features: psychology inspired geometric feature (PIGF) [47], pointer

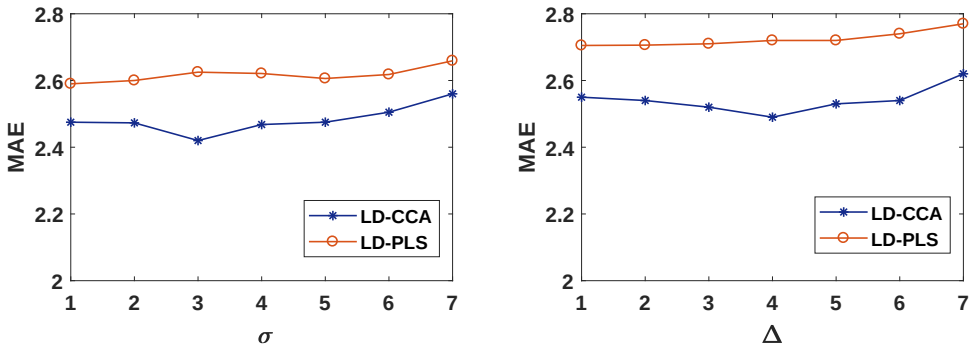


Figure 5.8: Parameters sensitivity in estimation results.

Table 5.7: The computing time (sec) taken for training the proposed methods and the other methods.

	LDL-IIS	LDL-CPNN	LD-PLS	LD-CCA
Black male	21457.4	1206.0	423.1	154.8
Black female	30710.8	152.5	605.5	221.6
White male	17019.6	828.3	335.6	122.8
White female	20156.9	298.2	397.4	145.4

feature(PF) [66]. PF is the geometric facial representation which can well define the face shape. It consists of coordinates of 68 facial landmarks extracted by Openface toolkit [111]. Here the coordinates of detected landmarks are simply concatenated as: $[x_1, y_1, ..., x_n, y_n, ..., x_{68}, y_{68}]^T$. The dimension of the feature is 136. The results are given in Table 5.5. It can be seen that LD-CCA and LD-PLS perform better than the other four methods in most cases. LD-CCA shows more robustness than LD-PLS.

5) Comparing with other methods: Finally, we compare our methods with two label distribution learning methods, namely LDL-IIS and LDL-CPNN [119] on Morph II dataset. The two methods are first proposed for facial age estimation. One assumption made in the IIS-LLD algorithm is the derivation of conditional probability $p(y|\mathbf{x})$ as the maximum entropy model [124]. A strategy similar to improved iterative scaling (IIS) [125] is used to optimize the cost function. Alternatively, LDL-CPNN uses a three-layer network to approximate $p(y|\mathbf{x})$ to remove the above assumption. The results are given in Table 5.6. Three features are utilized for comparison. Note that the deep feature is extracted from the fine-tuned BMI face model. From the result, one can see that our method outperforms LDL-IIS and LDL-CPNN in most cases. Table 5.7 shows the computing time taken for training the proposed methods and the compared methods [119]. Note that the computing time is based on applying deep features. One can see that our methods take significant

Table 5.8: BMI estimation results by applying label distribution method twice on Morph II dataset.

Feature	Method	Black male		Black female		White male		White female	
		MAE	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy
Fine-tuned	LD-PLS	2.41	76.6%	3.69	59.4%	2.54	74.3%	2.98	72.0%
	LD-CCA	**2.35**	**77.0**%	3.40	**67.3**%	**2.25**	**75.6**%	**2.72**	**73.8**%
Fine-tuned	LD-PLS(twice)	2.45	76.1%	3.60	62.4%	2.55	75.1%	2.96	71.8%
	LD-CCA(twice)	2.40	75.3%	**3.35**	67.1%	2.31	75.2%	2.77	74.1%

less time for training than the LDL-IIS and LDL-CPNN methods. Though LDL-IIS method performs competitively with the proposed methods, it takes much longer time for training. This further demonstrates the effectiveness and efficiency of the proposed methods.

5.4.4 Discussion

Considering many iterative methods have been proposed for optimization [126–128], it is important to figure out if applying label distribution method twice will lead to better performance on BMI estimation. To aim this, we conduct an experience on Morph II dataset. The label distribution methods (LD-CCA and LD-PLS) are applied twice to the BMI-related facial features. The experiment setting is described in Section 5.4.2. The experimental results are compared with that of applying label distribution methods once. As shown in Table 5.8, we can see that applying label distribution methods twice can not lead to better performance on most sets except on black female by LD-CCA (twice). The MAE is decreased from 3.40 to 3.35.

5.5 Summary

In this chapter, we study the problem of BMI estimation from facial images by a two-stage learning framework. More specifically, first, a BMI related face model is fine-tuned to learn more BMI related facial features. Then the facial features are extracted from the fine-tuned face model, and the BMI labels are modeled into discrete probability distributions. Finally given the extracted BMI related facial features and the probability distributions, a BMI estimator is learned by maximizing the correlation between them. Two different label assignment strategies are presented and compared in this work. The results show that the two-stage framework reduces the estimated errors step by step.

The proposed label distribution based estimator shows more robustness than regression based methods and methods without label distribution schemes. We further evaluated the effectiveness of the estimator on two geometric features. In addition, our method outperforms the two label distribution based methods: LDL-IIS and LDL-CPNN.

Chapter 6

Label Assignment Matching based Network for BMI Estimation

Take into account the existing challenges in visual BMI estimation, we propose a label assignment matching based convolutional neural network for BMI estimation from facial images. First, considering the limited BMI training data, the network can simultaneously learn the BMI related facial feature and BMI estimator. Second, the label assignment scheme well defines the ambiguity of BMI labels. Third, the triple-loss function takes advantage of both relative entropy loss and distribution shape matching. Extensive experiments are conducted on two datasets to evaluate the proposed method. Comparing with state-of-the-art methods, the proposed method successfully achieves an improvement in BMI estimation from facial images.

The remainder of this paper begins with introducing the existing challenges in visual BMI estimation in Section 6.1. Related works on BMI estimating methods, ranking based and label distribution based methods are summarized in Section 6.2. Details about the label assignment matching based learning network and the triple-loss function are presented in Section 6.3. Section 6.4 presents two databases used for performance evaluation: an extended version of FIW-BMI2 and Morph II. In Section 6.5, first, we describe the evaluation metrics and experimental setting; and then we provide the detailed experimental results and discussion. Finally, conclusions are given in Section 6.6.

6.1 Challenge

BMI estimation from facial images is a challenging problem in computer vision and pattern recognition. The first challenge is caused by the BMI data. Different from the

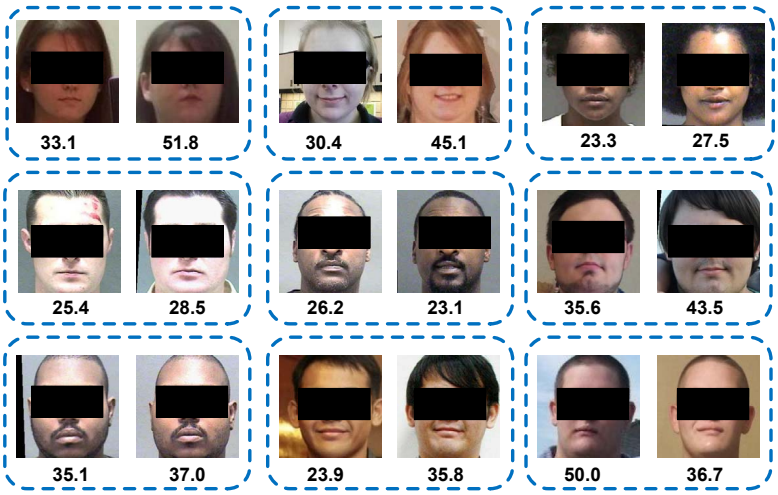


Figure 6.1: Some frontal face images with corresponding BMI values. The increase in facial adiposity is a continuous process.

dataset used for age estimation and face recognition, it is difficult to collect a dataset contains images that cover all BMI values. According to the BMI values, there are four BMI categories: underweight (BMIs$\leq$18.5), normal (18.5<BMIs$\leq$25), overweight (25<BMIs$\leq$30), obese (BMIs>30). Most of BMI data distribute on normal and overweight categories. The distribution of BMIs on the dataset is uneven. In addition, there are very few public datasets for visual BMI estimation. Exploiting efficient learning method for BMI estimation with limited training data is an urgent demand. The second challenge is the ambiguity of BMI labels. As shown in Fig. 6.1, the increase in facial adiposity is a continuous process. A person with ground-truth BMI 32 means the probability of 32 is higher than other adjacent BMI values (e.g. 31, 32.5 and 33, etc.). Though the ground-truth is 32, other adjacent BMI values also have the probabilities to describe this person. In addition, the correlation between facial appearance and BMI is related to age and gender [25]. With different ages or gender, variances of facial appearance caused by BMI are different. Both traditional regression based methods [47] and Euclidean loss based neural networks [63] ignored such ambiguity. Several works have studied the ambiguity problem in age estimation, such as ranking based methods [129], label distribution based methods [130, 131]. They describe the ambiguity by different deep network architectures, but the label assignment matching solution has not been fully exploited in the existing works.

In this chapter, we propose an end-to-end convolutional neural network (CNN) for visual BMI estimation which integrates feature learning and estimator learning in one

network. A label assignment scheme is embedded into the deep network which models the scalar BMI label as discrete probabilities distribution. A triple-loss function is proposed for label assignment matching which minimizes the discrepancy between estimated labels and ground-truth labels. We use a pre-trained face model as the based structure of the network. Then the network is followed by the label assignment matching module. There are three advantages of this proposed network. First, considering the limited BMI training data, the network can simultaneously learn the BMI related facial feature and BMI estimator. Second, the label assignment scheme well defines the ambiguity of BMI labels. Third, the triple-loss function takes both advantages of relative entropy loss and distribution shape matching. The main contributions of this work are summarized as follows:

- A label assignment matching based learning network is designed for visual BMI estimation from facial images. The structure of the network is evaluated by extensive experiments.
- A triple-loss function is proposed for label matching which consists of relative entropy loss, absolute value loss and variance loss.

6.2 Related Work

BMI Estimation Approaches: Computationally the BMI values can be estimated from 2D facial images by geometric features and deep features. Wen et al. [47] first proposed geometric features based computational method for BMI prediction from face images. The psychology inspired geometric features (PIGF) are computed from facial images. Three regression methods: the support vector regression (SVR) [48], Gaussian process regression (GPR) [49], and the least-squares estimation [50] are used for learning the map between facial features and BMI values. The approach was evaluated on the selected Morph II dataset [51]. Pascali et al. [45] proposed a method for automatically extracting geometric features which are related to weight parameters, from 3D facial data collected by low-cost depth scanners. Kocabey et al. [61] analyzed BMIs from face images collected from a social media website. The pre-trained VGG-Net and VGG-Face model [62] are used to extract features from facial images. Then they employed SVR models to predict BMIs from the extracted features. Dantcheva et al. [63] explored the possibility of estimating height, weight, and BMI from facial images by using a regression method based on a 50-layer ResNet architecture. They evaluated their methods on a celebrity

dataset. Jiang et al. [66] comprehensively studied the visual BMI estimation problem based on the characteristics and performance of several geometric facial features (PIGF, PF and PIGF+PF) and deep features (VGG-Face, Arcface, Centerloss and LightCNN). SVR models are used for predicting BMI values. All the above work considered BMI prediction as regression problems and ignored the ambiguity of BMI labels. In addition to the above regression based methods, recently Jiang et al. [67] proposed a two-stage learning framework for BMI estimation from facial images. A label distribution based BMI estimator is learned by an optimization procedure that is implemented by projecting the features and assigned labels to a new domain which maximizing the correlation between them.

Ranking based Methods: Some ranking based methods are proposed for age estimation. Considering age-related ordinal information, these methods transform the ordinal regression problem to a series of binary classification. Li et al. [132] presented a reduction framework from ordinal regression to binary classification based on extended examples. Chen et al. [129] proposed a CNN based framework for age estimation which contains a series of basic CNNs. Different from age labels (e.g. 18, 19, 20, etc.), BMI labels have a much finer interval (e.g. 22.3, 22.7, 23, etc.) between each other. Thereby ranking based methods can not well define the BMI labels.

Label Distribution Methods: Label distribution scheme was first proposed by Geng et al. [119] for age estimation to describe the ambiguity of age labels. In [119], two optimization methods−IIS-LLD and CPNN were presented. Later on, several distribution learning based approaches have been proposed for age estimation and other tasks. A multivariate label distribution (MLD) based method was also proposed by Geng et al. [120] for further improving the performance. Xin et al. [121] used Logistic Boosting Regression (LogitBoost) to learn a general label distribution model family which can avoid the potential influence of the specific model. Gao et al. [130] proposed a deep label distribution learning (DLDL) method by minimizing a Kullback–Leibler divergence between the predicted and ground-truth label distributions. These methods utilized label correlation or a single entropy model to optimize the divergence between the estimation and ground-truth. The label matching problem has not been fully exploited in these works.

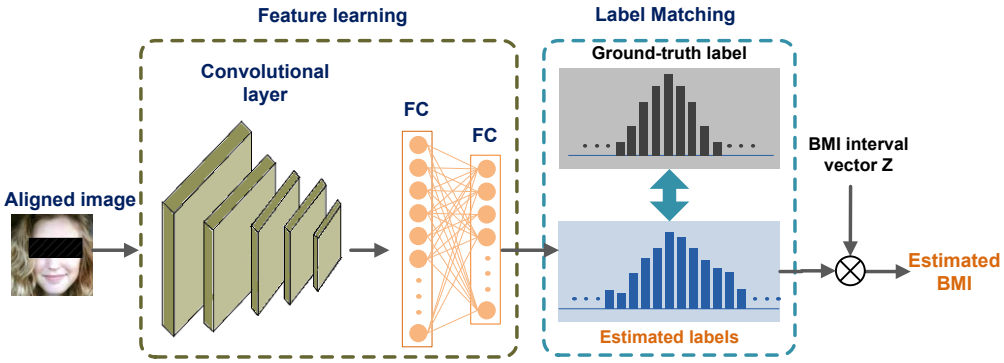


Figure 6.2: The pipeline of the proposed BMI estimation method. It consists of two main steps: feature learning and label matching. A convolutional neural network is utilized to extract features from the aligned images. Then extracted features are normalized by the softmax function. The estimated BMI value is the dot product of the estimated labels and the corresponding BMI range vector Z. The whole network is optimized by the triple-loss function.

6.3 Method

Fig. 6.2 shows the pipeline of the proposed method which takes the aligned face images as the input. The images are passed through convolutional layers and fully connected (FC) layers. The output of the last fully connected layer is normalized by the softmax function. The estimated BMI value is the dot product of the estimated probabilities from the label matching module and the corresponding BMI range vector Z. In this section, first we describe the definition of BMI estimation problem with the label assignment scheme. Then we present the structure of the network and give the details about the proposed label matching solution.

6.3.1 Modeling BMI values with label assignment

The ground-truth BMI value for each image is a real number usually rounding to one decimal place. Considering the ambiguity of the BMI label discussed in Section 6.1, we use the label assignment scheme to model BMI labels.

As shown in Fig. 6.3, given an image labeled with the BMI value m, the BMI value is transformed to discrete probabilities distribution $\mathbf{p} = [p_1, p_2, ..., p_k]^T \in \mathbb{R}^k$ over the whole range of BMIs which follows a Gaussian distribution centered at m:

$$p(z_i) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(z_i - m)^2}{2\sigma^2}) \tag{6.1}$$

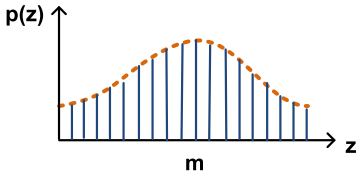


Figure 6.3: Probability density function of Gaussian distribution $\mathcal{N}\left(m, \sigma^2\right)$.

where σ is the standard deviation of the Gaussian distribution. And $\mathbf{z} = [z_1, z_2, ..., z_k]^T \in \mathbb{R}^k$ is a set of discrete values from the whole range of BMIs with interval of 0.5, e.g. if the range of BMI is from 15 to 60, then $\mathbf{z} = [15, 15.5, ..., 59.5, 60]$. Note that k can be adjusted with different BMI ranges and intervals. The final assigned labels $\mathbf{y} = [y_1, y_2, ..., y_k]^T \in \mathbb{R}^k$ should be normalized by:

$$y_i = \frac{p(z_i)}{\sum_{i=1}^{k} p(z_i)}. \tag{6.2}$$

One advantage of the label assignment is that it covers a discrete range of BMIs with different levels of "probabilities". It is more suitable to represent BMI value, because inside of this range, every BMI value could be a possible label to describe the true BMI with different confidence. In addition, by applying label assignment, BMI estimation can obtain a "continuous" value, which cannot be achieved by the classification approaches.

6.3.2 Label Matching solution

As shown in Fig. 6.2, softmax function is applied to the output of FC layer with the purpose of normalizing the estimated labels. The parameters of the whole network are optimized by matching the estimated labels with the ground-truth. A triple-loss function is proposed to optimize the network which contains relative entropy loss, MAE loss and variance loss. The details of this approach are given below.

Relative entropy loss

Given an image with BMI labeled as a scalar b, the assigned labels (ground-truth) $\mathbf{y} = [y_1, y_2, ..., y_k]^T$ are computed according to Eqns. (6.1) and (6.2). The estimated labels from the network is denoted as $\hat{\mathbf{y}} = [\hat{y_1}, \hat{y_2}, ..., \hat{y_k}]^T$. To maximize the similarity between $\mathbf{y}$ and $\hat{\mathbf{y}}$, a loss function is used to penalize their discrepancy. This work utilizes Kullback-Leibler (KL) divergence to measure the discrepancy between the ground-truth

and estimated labels:

$$\begin{aligned} D_{KL}(\mathbf{y}, \hat{\mathbf{y}}) &= \sum_{i=1}^{k} y_i \cdot log\frac{y_i}{\hat{y}_i} \\ &= \sum_{i=1}^{k} y_i \cdot (logy_i - log\hat{y}_i)\,. \end{aligned} \tag{6.3}$$

Because $y_i \cdot logy_i$ is a constant, the KL divergence based loss function is defined as the following:

$$L_{kld}(\mathbf{w}) = -\sum_{i=1}^{k} y_i \cdot log\hat{y}_i, \tag{6.4}$$

here $\mathbf{w}$ denotes the parameters to be optimized.

MAE loss

As shown in Fig. 6.2, the estimated BMI value is computed by:

$$\hat{m} = \sum_{i=1}^{k} \hat{y}_i \cdot z_i, \tag{6.5}$$

here $\hat{y}_i$ denotes the estimated probability that belongs to BMI value z_i. $\hat{m}$ is also the mean of estimated labels $\hat{\mathbf{y}}$. Euclidean loss function is utilized to match the ground-truth BMI and estimated BMI:

$$L_{mae}(\mathbf{w}) = |m - \hat{m}|\,, \tag{6.6}$$

As mentioned in Section 6.3.1, the assigned ground-truth labels follow the Gaussian distribution $P \sim \mathcal{N}\left(m, \sigma^2\right)$. Among the two parameters, m defines the location of the Gaussian distribution, σ^2 determines the shape of the distribution. Eqn. (6.6) not only penalizes the distance between the estimated value and ground-truth, but also penalizes the location difference between estimated distribution and true distribution.

Variance loss

The variance of estimated labels is expressed as:

$$\hat{Var} = \left(\sum_{i=1}^{k} \hat{y}_i \cdot z_i^2\right) - \hat{m}^2. \tag{6.7}$$

Variance determines the shape of Gaussian distribution. The purpose of variance loss is to match the shape of distributions. Variance loss penalizes the decentralized labels by:

$$L_{var}(\mathbf{w}) = max\left(0, \hat{Var} - \sigma^2\right), \tag{6.8}$$

where σ^2 is the variance used for generating the ground-truth labels in Eqn. (6.1), which is predefined as a hyperparameter of the network. This loss function leads to concentrated distributions by penalizing violated variance which is lager than σ^2. Eq. (6.8) has the same format as the hinge loss function. Though it is not differentiable, a sub-gradient of the loss function can be computed as:

$$\frac{\partial L}{\partial w} = \begin{cases} \frac{\partial \hat{Var}}{\partial w}, & \hat{Var} > \sigma^2 \\ 0, & otherwise \end{cases} \tag{6.9}$$

6.3.3 Full Objective

The proposed label matching solution is embedded in the convolutional network. Such a framework takes both advantages of CNN and label matching scheme. The BMI related features are learned with CNN, and the ambiguity of BMI label is addressed by label matching scheme. Given a training set $\mathcal{D}$, the goal of the network is to optimize parameters $\mathbf{w}$ by the loss function. The final loss function is defined as a combination of the three above loss functions:

$$L = L_{kld}(\mathbf{w}) + \lambda_1 L_{mae}(\mathbf{w}) + \lambda_2 L_{var}(\mathbf{w}), \tag{6.10}$$

where λ_1 and λ_2 are the tradeoff parameters which balance the importance between three losses. Substituting Eqns. (6.4), (6.6) and (6.8) into Eqn. (6.10):

$$L = -\sum_{i=1}^{k} y_i \cdot log\hat{y}_i + \lambda_1 \left| m - \hat{m} \right| + \lambda_2 max\left(0, \hat{Var} - \sigma^2\right). \tag{6.11}$$

The proposed label assignment matching solution can be easily implemented by deep learning libraries, such as Tensorflow [133], Theano [134], etc. Furthermore, it can be embedded into any CNNs. In this work, we embed it into a modified VGG13 [80] network. Original VGG13 consists of 13 convolutional layers, 5 max-pooling layers and 3 fully connected layers. We replace the last fully connected layer with the label assignment matching module.

6.4 Dataset

Extensive experiments are conducted on three datasets: CASIA-WebFace, Morph II and FIW-BMI. CASIA-WebFace is used to train the general face model by a face recognition task. The proposed label assignment based network is finetuned and evaluated on the other two datasets.

Table 6.1: The number of images in the training and test set of Morph II.

Training		Test	
Male	17416	Black male	5713
		White male	1006
Female	3828	Black female	605
		White female	440

6.4.1 CASIA-WebFace

CASIA-WebFace [135] is a large-scale database including 494,414 face images from 10,575 subjects. Face images in this database are crawled from the Internet (i.e., IMDb website) and annotated in a semiautomatic manner. First, the names of interested celebrities are crawled from the website, then photos on their webpages are downloaded. As most photos contain more than one face, a simple and fast clustering method was used to annotate the identity of faces in the photos. Finally, the authors checked the whole database manually and corrected false annotations. It was originally collected to train a deep convolutional neural network for face recognition and obtained state-of-the-art accuracy. This database is used to train the VGG13 network by a face recognition task. We randomly split the dataset into two parts, 80% of the whole dataset for training and the remaining 20% for validation. There is no overlap of subjects between the training and validation sets.

6.4.2 Morph II

Morph II dataset [51] contains 55,608 passport-style frontal face images along with age, gender and ethnicity information. Moreover, there are 40,330 images have height and weight information. Considering the uneven distribution of the ethnicity in the database, only images from Black and White are used for this work. There are 29,033 images kept. Details about the selected data are described in Table. The same individual does not exist in both the training and test set. Most images from the same individual have different BMI values. The BMI values of Morph II mainly distribute in the range of 15 to 35. Among these, 893 are underweight, 16,582 are normal, 8,237 are overweight and 3,321 are obese. Table 1 shows the details about the training and test sets. We train the BMI estimation network on the training set and report the results on each gender-ethnicity group.

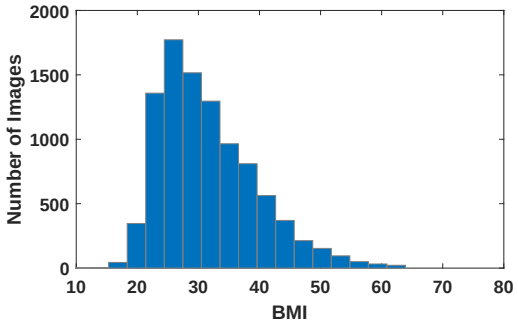


Figure 6.4: Distribution of BMI values on FIW-BMI.

6.4.3 FIW-BMI

We expand FIW-BMI [66] by adding 1685 images from 577 individuals to it. The updated dataset contains 9615 images from 5458 individuals with the annotation of gender, weight and height. Each individual contributes 1 to 4 images. We use the same image alignment and normalization protocols as used for FIW-BMI. The BMI values of the dataset distribute from 14 to 64, as shown in Fig. 6.4. Because FIW-BMI has a much wider BMI range than Morph II, it is more challenging to estimate BMI on this dataset. The dataset is split to the training set (7726 images) and the test set (1889 images). Among the test set, there are 1225 images from the male, and 664 images from the female. We train the BMI estimation network on the training set and report the results on two gender groups separately.

6.5 Experiments

Experimentally, we evaluate the performance of the proposed BMI estimation network on two datasets. First, the performance metrics and experiment settings are briefly described. Then the experimental results and analysis are presented.

6.5.1 Evaluation Metrics

Mean absolute error (MAE) is utilized to measure the performance on BMI estimation. It is defined as the average of the absolute error between the estimated BMI values and the ground truth BMI values, which is computed by:

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|\hat{p}_k - p_k|\,, \tag{6.12}$$

here p_k is the ground truth BMI value for image k, $\hat{p}_k$ is the corresponding estimated BMI value, N is the number of test images. This measure is motivated by its use in age estimation [103].

6.5.2 Experiment Setting

As mentioned in [66], image alignment is a preprocessing step for all BMI estimation methods. The alignment is based on the five detected face landmarks (two eyes, nose and two corners of the mouth). It performs translation, rotation and scaling of the faces so as to align all face images into the common coordinates. The output is a cropped 256×256 image. The MTCNN toolkit [110] is employed for detecting the required face landmarks.

Considering the number of images for the BMI estimation method is limited, we apply data augmentation to the training set to avoid overfitting. The training images are random cropped and rotated.

All experiments are implemented using the Tensorflow framework on an NVIDIA GP102 GPU. First, we train VGG13 network with softmax loss for face recognition on the WebFace dataset. With the pre-trained VGG-face network, we remove the last fully connected layer and resize the dimension of the second fully connected layer. The dimension of the second fully connected layer should be the same as the dimension of BMIs range vector. In this work, the BMIs range vector is defined as: $[15 : 0.5 : 80]$ (Matlab notation). Then the label assignment matching module is added to the end of the network. Finally, we fine-tune this modified network on BMI datasets.

When training the VGG13 network, we use mini-batch stochastic gradient descent (SDG) with momentum settings. The mini-batch size is set to 32 and momentum is set to 0.9. We initialize the learning rate to 0.01. The learning rate decreases in exponential decay with 0.1. The training procedure stops after 30 epochs. When fine-tuning BMI estimation network, we use mini-batch SDG without momentum. The mini-batch size is set to 64. We initialize the learning rate to 0.0001. The learning rate decreases in exponential decay with 0.1. The fine-tuning procedure stops after 60 epochs. In this work, the better performance is achieved by setting the two hyperparameters λ_1 and λ_2 to 0.5 and 0.1, respectively.

6.5.3 BMI Estimation Results

We evaluate the performance of the proposed method on each ethnicity and gender group of the test set. First, the proposed method is compared with regression based and

Table 6.2: The number of images in $t1$ and $t2$ sets used for LD-CCA method.

	$t1$	$t1$
Morph male	8023	9393
Morph female	1836	1992
FIW-BMI	3327	4423

labels distribution based methods. Then we validate the efficiency of the label matching solution by ablation study. Finally, we explore the sensitivity of hyperparameters λ_1 and λ_2.

Comparisons with the State-of-the-art

We compare the proposed network with several regression based and label distribution based methods for BMI estimation on Morph II and FIW-BMI. Theses methods include support vector regression (SVR) [48], principal component analysis (PCA) + SVR [66], Gaussian processing regression [73], LDL-IIS and LDL-CPNN [119], LD-CCA and DLDL [130]. For SVR, SVR+PCA, GPR, LDL-IIS and LDL-CPNN methods, all the features are extracted by the VGG-Face network which is pre-trained on the Web-face dataset (described in Section 6.5.2). Then the BMI estimators are learned with the training set. Note that when we implement the PCA+SVR method, the PCA projection is only learned with the training set. LD-CCA, DLDL and the proposed methods all use the same pre-trained VGG-Face network as the backbone. Because LD-CCA is a two-stage learning method, the training set is divided into two parts ($t1$ and $t2$), $t1$ is used for BMI-feature learning and $t2$ is used for LD-CCA estimator learning. The details about $t1$ and $t2$ are given in Table 6.2. DLDL and our method both use the training set to fine-tune the BMI estimation network.

Table 6.3 shows the comparisons of the BMI estimation MAEs by the proposed method and regression based methods on Morph II and FIW-BMI datasets. The experimental results are reported based on each of the gender and ethnicity groups. It is shown that the proposed method outperforms the three regression based methods with a clear margin on most sets, except on the female test set of FIW-BMI. As described in Section 6.4.3, the percentage of females is about 35% in the training set of FIW-BMI. The poor performance on this female set may be caused by the fewer training data of females. Our method achieves the lowest MAE of 2.25, 3.37, 2.27, 2.69, 3.20 and 3.85 on the six gender and ethnicity groups, respectively. This result demonstrates the advantages of the label

Table 6.3: Comparisons of the BMI estimation (MAEs) by the proposed method and regression based methods on Morph II and FIW-BMI dataset.

Method	Morph II				FIW-BMI	
	Black male	Black female	White male	White female	Male	Female
SVR [48]	2.55	3.65	2.43	3.01	3.46	3.85
PCA+SVR [66]	2.51	3.60	2.42	2.96	3.40	3.87
GPR [73]	2.60	3.67	2.45	3.10	3.45	3.97
Ours	**2.14**	**3.37**	**2.27**	**2.61**	**3.20**	3.85

Table 6.4: Comparisons of the BMI estimation (MAEs) by the proposed method and other label distribution based methods on Morph II and FIW-BMI dataset.

Method	Morph II				FIW-BMI	
	Black male	Black female	White male	White female	Male	Female
LDL-IIS [119]	2.99	4.49	3.45	4.03	3.64	4.34
LDL-CPNN [119]	7.66	8.34	5.43	6.37	5.51	8.87
LD-CCA	2.41	3.50	2.38	2.86	3.30	**3.73**
DLDL [130]	2.36	3.55	2.30	2.80	3.43	3.93
Ours	**2.14**	**3.37**	**2.21**	**2.61**	**3.20**	3.85

assignment matching based network.

Table 6.4 shows the comparisons of the BMI estimation MAEs by the proposed method and four label distribution based methods on Morph II and FIW-BMI datasets. Among the four compared methods, LD-CCA is a two-stage learning method and DLDL is an end-to-end CNN method. Though all five methods in Table 6.4 are based on a similar label distribution scheme, we can see that our method still outperforms the other four methods on most test sets, except on the female test set of FIW-BMI. This demonstrates that the proposed label matching scheme (triple-loss function) is more capable of optimizing the BMI estimation network.

Fig. 6.5 shows some examples of BMI estimation along with estimated label distributions by our approach on Morph II and FIW-BMI datasets. The upper panel shows some good cases, and the lower panel shows some failure cases. The red curves are ground-truth probabilities distributions, and the blue curves are estimated probabilities distributions. We can see that most of estimated distributions follow Gaussian distribution. In good cases, most parts of the estimated distributions overlap with the ground-truth distributions. While in failure cases, the estimated distributions have larger variances, and are

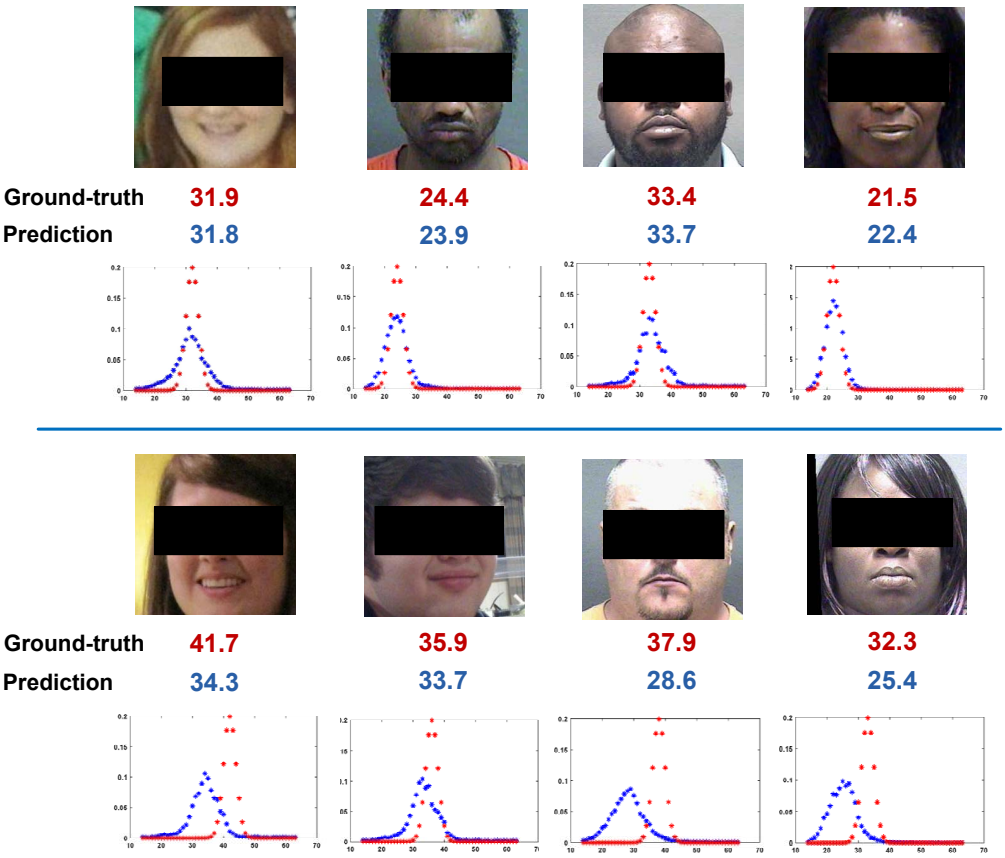


Figure 6.5: Examples of BMI estimation by the proposed method. The upper panel shows good cases, and the lower panel shows failure cases. The red curve is ground-truth probabilities distribution, and the blue curve is estimated probabilities distribution.

separated from the ground-truth. Some failure cases are observed due to the large pose, occlusion of the face, etc.

Ablation Study

To further evaluate the contribution of each loss function, we design this ablation study by comparing the performance of different combinations of loss functions. The experiments are conducted on Morph II dataset and Table 6.5 shows the experimental results. λ_1 and λ_2 denote the coefficients applied to the loss function. They are set to 0.5 and 0.1, respectively. First, we evaluate the performance of three single-loss functions: *kld* (calculated by Eqn. (6.4)) and *mae* (calculated by Eqn. (6.6)) and *var* (Eqn. (6.8)). It is shown that both *kld* and *mae* perform well on optimizing the network which leads to much lower MAEs than *var*. Among them, *kld* loss is relatively more efficient than *mae*

Table 6.5: Performance (MAEs) of the BMI estimation network optimized by different combinations of loss functions. λ_1 and λ_2 are set to 0.5 and 0.1, respectively.

Loss function	Morph II			
	Black male	Black female	White male	White female
kld	2.36	3.55	2.30	2.80
mae	2.50	3.68	2.49	2.98
var	12.11	15.47	15.64	16.55
$kld + \lambda_1 mae$	2.27	3.46	2.22	2.74
$kld + \lambda_1 var$	2.32	3.56	2.30	2.84
$mae + \lambda_1 var$	2.42	3.58	2.43	2.83
$kld + \lambda_1 mae + \lambda_2 var$	**2.14**	**3.37**	**2.21**	**2.61**

loss, while var performs worst. Because var is a loss function that is used to penalize the decentralized distributed labels, we think it can be used as an auxiliary loss function rather than the main loss function.

Next, we evaluate the performance of three combinations: $kld + \lambda_1 mae$, $kld + \lambda_1 var$ and $mae + \lambda_1 var$. From Table 6.5, it is shown that these three combinations perform better than the above three single-loss functions on the four test sets. $kld + \lambda_1 mae$ achieves a little lower MAE than $mae + \lambda_1 var$ and $kld + \lambda_1 var$. Comparing with different combinations of loss functions, we can see the triple-loss function is the most robust. We can draw the conclusion that three loss functions all contribute to improving performance. Among them, kld loss is relatively more efficient than the other two.

Hyperparameters Sensitivity Analysis

Hyperparameters λ_1 and λ_2 balance the contribution of three functions during the training stage. We take the values of these two parameters by changing $0.1, 0.3, 0.5, 0.7, 0.9, 1$, and conduct the experiments on Morph II dataset. The matching MAEs of the proposed method are shown in Table 6.6. Note that the experiment is based on separately training on each gender set, and the MAEs are computed from the whole test set of Morph II. According to the experimental results, we can see that the best performance is achieved by $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$.

Table 6.6: Performance (MAEs) of the BMI estimation network with different hyperparameters.

λ_2 \ λ_1	0.1	0.3	0.5	0.7	0.9	1
0.1	2.44	2.33	**2.27**	2.30	2.34	2.39
0.3	2.46	2.41	2.32	2.35	2.41	2.45
0.5	2.46	2.44	2.42	2.42	2.47	2.44
0.7	2.47	2.44	2.42	2.44	2.54	2.58
0.9	2.58	2.56	2.49	2.49	2.56	2.62
1	2.62	2.60	2.55	2.59	2.64	2.65

6.6 Summary

To address the challenges caused by BMI data and ambiguity of labels, this chapter proposes a label assignment matching based convolutional neural network for BMI estimation from facial images. There are three advantages of this network. First, the network can simultaneously learn the BMI related facial feature and BMI estimator. Second, the label assignment scheme well defines the ambiguity of BMI labels. Third, the triple-loss function takes the both advantages of relative entropy loss and distribution shape matching. Extensive experiments are conducted on two datasets. Comparing with three regression methods and four label distribution based methods, the proposed method performs better in most cases. Additionally, to fully evaluate the proposed triple-loss function, we compare the performance of different combinations of loss functions. According to the experimental results, it is shown that the three loss functions all contribute to the improvement of performance. Among them, *kld* loss is relatively more efficient than *mae* loss. Both *kld* and *mae* have counted more with the improvement than *var*.

Chapter 7

Conclusion and Future Work

The final chapter summarizes the work and contributions made in the dissertation as well as envisioning possible research problems that can be further explored.

7.1 Conclusion

BMI and Weight Analysis from Visual Body Data The studies in health science [19–21] show evidence on the relation between anthropometric measures and obesity. Recently, researchers utilized machine learning based methods to analyze weight from various types of body data and have achieved a few success. However, there are still some limitations that exist in these body weight analysis methods. Considering that existing methods [32–34] use both color and depth images to estimate weight, we investigate the feasibility of analyzing body weight from single 2D frontal view human body images. We also study a computational approach that directly estimates body weight and height from dressed people in 3D space. The conclusions for BMI and weight analysis from visual body data can be summarized as follows:

- We propose an approach to analyze body weight just from 2D body images [64]. Neither depth images nor clear face images are required for this approach. To the best of our knowledge, this is the first work to explore weight/BMI related information from 2D body images only. A computational framework is developed for body weight and BMI analysis from 2D human body images, which can process either a single image or a pair of images. Five anthropometric features are proposed for body weight analysis from 2D body images. Correlation is analyzed between the extracted anthropometric features and BMI values, which validates the usability of

the selected features. More specifically, body weight analysis is studied at three different levels of difficulties: the weight change classification is first investigated from a pair of body images of the same subjects; further investigation is conducted to estimate how big the weight change between the pairwise images is; the last is to predict the BMI value from a single body image. A new visual-body-to-BMI image dataset has been collected and cleaned to facilitate this study. The errors of the three estimation tasks evaluated by several measurements are within acceptable ranges. Comparing with the facial images analysis approaches, the proposed method performs better in most cases. Furthermore, our anthropometric features significantly outperform the VGG-Net feature on BMI estimation. Based on all experimental results, it is promising to analyze body weight or BMI from the 2D body images visually.

- An efficient weight estimation framework is developed to work on normally dressed people in 3D space [65]. Two clothes models are proposed to reduce the negative influence of loose clothes on body volume and weight estimation. Though the Kinect 3D fusions contain some noise, the proposed BMI estimation includes clustering and fitting components to suppress such noise. A new RGB-D dataset is collected for this study. Experimental results have shown the effectiveness of the proposed approach to people with different styles of clothes, for both females and males. Comparing to another 3D volume estimation method, our method achieves a significantly lower error.

BMI Estimation from Facial Images In the past few years, several computational methods are proposed for BMI estimation from facial images. A typical framework for BMI estimation consists of four steps: face detection, image alignment, facial representation extraction, and BMI estimator learning. The third and fourth steps both are important which dominantly determines the performance of a BMI estimation method. In this dissertation, we take a further step to delve deeper into the third step-characteristics and performance of different facial representations, and investigate how we can improve the performance of the fourth step by developing effective learning frameworks.

Conclusions for BMI estimation from facial images are summarized as follows:

- We study the visual BMI estimation problem systematically based on the facial representations [66]. According to the inherent properties of representations, they are grouped into two types: geometric based and deep learning based. In addition

to the two existing approaches (VGG-Face and PIGF), five other facial approaches: PF, PIGF+PF, LightCNN, Centerloss and Arcface are explored for the first time for BMI analysis. The performance and characteristics of facial representations have been comprehensively evaluated and analyzed from three perspectives: the overall performance on visual BMI prediction, the redundancy in representations and the sensitivity to head pose changes. The experiments are conducted on two databases: FIW-BMI and Morph II. Our studies provide some deep insights into the facial representations for visual BMI analysis: 1) The deep model based methods perform better than geometry based methods. Among them, the VGG-Face and Arcface show more robustness than others in most cases; 2) Removing the redundancy in VGG-Face representation can increase the accuracy and efficiency in BMI estimation; 3) Large head poses lead to low performance for BMI estimation. Among the seven representations, the Arcface, VGG-Face and PIGF are more robust than the others to head pose variations.

- We study the problem of BMI estimation from facial images by a two-stage learning framework [67]. First, the BMI related facial representation is learned by fine-tuning the pre-trained deep face model. This step is expected to obtain sufficient visual BMI characteristics and reinforce the learning process using the limited number of BMI data. More importantly, the label distribution method models the single BMI value as a discrete probability distribution over the whole ranges of BMIs. Given the extracted facial features from the first stage, a label distribution based BMI estimator is learned by an optimization procedure by projecting the features and assigned labels to a new domain which maximizing the correlation between them. Two different label assignment strategies are presented and compared in this work. The results show that the two-stage framework reduces the estimated errors step by step. The proposed label distribution based estimator shows more robustness than regression based methods and methods without label distribution schemes. We further evaluated the effectiveness of the estimator on two geometric features. Additionally, our method outperforms the two label distribution based methods: LDL-IIS and LDL-CPNN.

- A convolutional neural network (CNN) is developed for visual BMI estimation which integrates feature learning and estimator learning in one network. A label assignment scheme is embedded into the deep network work which models the scalar BMI label as a probability distribution. A triple-loss function is proposed for label

assignment matching which minimizes the discrepancy between estimated labels and ground-truth labels. The experiments are conducted on three databases: CASIA-WebFace, FIW-BMI and Morph II. The results demonstrate that the proposed method outperforms state-of-the-art regression based methods and label distribution based methods.

7.2 Future Work

In this section, several future research topics are summarized as follows.

Deep learning based BMI estimation from body images: In this dissertation, we study BMI estimation from frontal view body images by anthropometric features in Chapter 2. To achieve improvement, we would like to explore innovative lightweight deep network to learn the latent feature representations for BMI estimation from body images. Furthermore, we also would like to investigate how to use the profile view of body images as auxiliary information to improve the estimation accuracy.

Generating facial images by varying BMI values: We have explored several facial representations for BMI estimation by analyzing their characteristics and performances in Chapter 4. One extension of this topic is generating different realistic versions of an input facial image by varying the BMI values. This would lead to algorithms or applications which allow users to modify the facial image using sliding knobs, like faders, to change the appearance of a facial image. Currently, many encoder-decoder based [136, 137] and generative adversarial network (GAN) based [138–140] architectures have been developed for the purpose of generating fake images. We would like to explore a feasible approach based on these two architectures.

Cross-BMI face verification: We have evaluated the performance of several deep face models on cross-BMI face verification. The experimental results demonstrate that large BMI-differences lead to low performance for face verification. To fully address this topic, we would like to investigate how to use the deep network to learn the mapping kernel that can reduce the variance between intra-subjects. The network aims to map the images with different BMIs to the common subspace, and to construct new feature representation which is robust to BMI variations and discriminative to different subjects.

List of Publications

1. M. Jiang, B. Cui, N. Schmid, M. McLaughlin, and Z. Cao, "Wavelet denoising of radio observations of rotating radio transients (RRATs): Improved timing parameters for eight RRATs," The Astrophysical Journal, vol. 847, no. 1, p. 75, 2017.

2. M. Jiang and G. Guo, "Body weight analysis from human body images," IEEE Transactions on Information Forensics and Security, vol. 14, pp. 2676-2688, Oct. 2019.

3. M. Jiang, Y. Shang, and G. Guo, "On visual BMI analysis from facial images," Image and Vision Computing, vol. 89, pp. 183-196, 2019.

4. M. Jiang, B. Cui, Y.-F. Yu, and Z. Cao, "DM-free curvelet based denoising for astronomical single pulse detection," IEEE Access, vol. 7, pp. 107389-107399, 2019.

5. M. Jiang, Y. Shang, and G. Guo, "A computational approach to body mass index from 3D reconstruction of dressed people," IET Image Pprocessing, 2020.

6. M. Jiang, G. Guo, and G. Mu, "Visual BMI estimation from face images using label distribution based method," Computer Vision and Image Understanding, under major revision, 2020.

7. M. Güunther, P. Hu, C. Herrmann, C.-H. Chan, M. Jiang, S. Yang, A. R. Dhamija, D. Ramanan, J. Beyerer, J. Kittler, et al., "Unconstrained face detection and open-set face recognition challenge," in Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), pp. 697-706, IEEE, 2017.

8. Y.-F. Yu, G.-X. Xu, M. Jiang, H. Zhu, D.-Q. Dai, and H. Yan, "Joint transformation learning via l2,1-norm metric for robust graph matching," IEEE Transactions on Cybernetics, 2019.

9. Y.-F. Yu, C.-X. Ren, M. Jiang, M.-Y. Sun, D.-Q. Dai, and G. Guo, "Sparse approximation to discriminant projection learning and application to image classification," Pattern Recognition, vol. 96, p. 106963, 2019.

Bibliography

[1] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. Springer Science & Business Media, 2007.

[2] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer Science and Business Media, 2006.

[3] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," *Pattern Recognition Letters*, vol. 68, pp. 218–230, 2015.

[4] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, 2006.

[5] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[6] A. K. Jain, S. C. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?" in *Biometric Technology for Human Identification*, vol. 5404. International Society for Optics and Photonics, 2004, pp. 561–572.

[7] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 406–415, 2010.

[8] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739–777, 2011.

[9] J. B. Meigs, P. W. Wilson, C. S. Fox, R. S. Vasan, D. M. Nathan, L. M. Sullivan, and R. B. D'agostino, "Body mass index, metabolic syndrome, and risk of type 2 diabetes or cardiovascular disease," *The Journal of Clinical Endocrinology & Metabolism*, vol. 91, no. 8, pp. 2906–2912, 2006.

[10] M. Arnold, M. Leitzmann, H. Freisling, F. Bray, I. Romieu, A. Renehan, and I. Soerjomataram, "Obesity and cancer: an update of the global impact," *Cancer Epidemiology*, vol. 41, pp. 8–15, 2016.

[11] A. G. Renehan, M. Tyson, M. Egger, R. F. Heller, and M. Zwahlen, "Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies," *The Lancet*, vol. 371, no. 9612, pp. 569–578, 2008.

[12] C. Y. Hsu, C. E. McCulloch, C. Iribarren, J. Darbinian, and A. S. Go, "Body mass index and risk for end-stage renal disease," *Annals of Internal Medicine*, vol. 144, no. 1, pp. 21–28, 2006.

[13] R. Wolk, P. Berger, R. J. Lennon, E. S. Brilakis, and V. K. Somers, "Body mass index," *Circulation*, vol. 108, no. 18, pp. 2206–2211, 2003.

[14] D. Gallagher, M. Visser, D. Sepulveda, R. N. Pierson, T. Harris, and S. B. Heymsfield, "How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?" *American Journal of Epidemiology*, vol. 143, no. 3, pp. 228–239, 1996.

[15] A. Pietrobelli, M. S. Faith, D. B. Allison, D. Gallagher, G. Chiumello, and S. B. Heymsfield, "Body mass index as a measure of adiposity among children and adolescents: a validation study," *The Journal of Pediatrics*, vol. 132, no. 2, pp. 204–210, 1998.

[16] N. R. Shah and E. R. Braverman, "Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin," *PloS One*, vol. 7, no. 4, p. e33308, 2012.

[17] M. Pollan, "Big food vs. big insurance," *The New York Times*, 2009.

[18] A. Keys, F. Fidanza, M. J. Karvonen, N. Kimura, and H. L. Taylor, "Indices of relative weight and obesity," *Journal of Chronic Diseases*, vol. 25, no. 6-7, pp. 329–343, 1972.

[19] A. Molarius and J. Seidell, "Selection of anthropometric indicators for classification of abdominal fatness—a critical review." *International Journal of Obesity and Related Metabolic Disorders*, vol. 22, no. 8, p. 719, 1998.

[20] M. Ashwell, S. Chinn, S. Stalley, and J. Garrow, "Female fat distribution-a simple classification based on two circumference measurements." *International Journal of Obesity*, vol. 6, no. 2, pp. 143–152, 1982.

[21] G. Vazquez, S. Duval, D. R. Jacobs Jr, and K. Silventoinen, "Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis." *Epidemiologic Reviews*, vol. 29, no. 1, pp. 115–128, 2007.

[22] J. C. Seidell, A. Oosterlee, M. Thijssen, J. Burema, P. Deurenberg, J. Hautvast, and J. Ruijs, "Assessment of intra-abdominal and subcutaneous abdominal fat: relation between anthropometry and computed tomography," *The American Journal of Clinical Nutrition*, vol. 45, no. 1, pp. 7–13, 1987.

[23] V. Coetzee, D. I. Perrett, and I. D. Stephen, "Facial adiposity: a cue to health?" *Perception*, vol. 38, no. 11, pp. 1700–1711, 2009.

[24] V. Coetzee, J. Chen, D. I. Perrett, and I. D. Stephen, "Deciphering faces: quantifiable visual cues to weight," *Perception*, vol. 39, no. 1, pp. 51–61, 2010.

[25] D. D. Pham, J.-H. Do, B. Ku, H. J. Lee, H. Kim, and J. Y. Kim, "Body mass index and facial cues in sasang typology for young and elderly persons," *Evidence-Based Complementary and Alternative Medicine*, vol. 2011, 2011.

[26] A. J. Henderson, I. J. Holzleitner, S. N. Talamas, and D. I. Perrett, "Perception of health from facial cues," *Phil. Trans. R. Soc. B*, vol. 371, no. 1693, p. 20150380, 2016.

[27] M. Ashwell, T. J. Cole, and A. K. Dixon, "Obesity: new insight into the anthropometric classification of fat distribution shown by computed tomography." *Br Med J (Clin Res Ed)*, vol. 290, no. 6483, pp. 1692–1694, 1985.

[28] W. H. Mueller and R. M. Malina, "Relative reliability of circumferences and skinfolds as measures of body fat distribution," *American Journal of Physical Anthropology*, vol. 72, no. 4, pp. 437–439, 1987.

[29] C. Velardo and J.-L. Dugelay, "Weight estimation from visual body appearance," in *Proceedings of the IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*. IEEE, 2010, pp. 1–6.

[30] C. L. Johnson, R. Paulose-Ram, C. L. Ogden, M. D. Carroll, D. Kruszan-Moran, S. M. Dohrmann, and L. R. Curtin, "National health and nutrition examination survey. analytic guidelines, 1999-2010," 2013.

[31] D. Cao, C. Chen, D. Adjeroh, and A. Ross, "Predicting gender and weight from human metrology using a copula model," in *Proceedings of the IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*. IEEE, 2012, pp. 162–169.

[32] D. Nahavandi, A. Abobakr, H. Haggag, M. Hossny, S. Nahavandi, and D. Filippidis, "A skeleton-free kinect system for body mass index assessment using deep neural networks," in *Proceedings of the IEEE International Systems Engineering Symposium (ISSE)*. IEEE, 2017, pp. 1–6.

[33] C. Pfitzner, S. May, and A. Nüchter, "Evaluation of features from RGB-D data for duman body weight estimation," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10 148–10 153, 2017.

[34] C. Pfitzner, S. May, and A. N, "Body weight estimation for dose-finding and health monitoring of lying, standing and walking patients based on RGB-D data," *Sensors (Basel, Switzerland)*, vol. 18, no. 5, 2018.

[35] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Weight estimation from frame sequences using computational intelligence techniques," in *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA)*. IEEE, 2012, pp. 29–34.

[36] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, S. Yussof, V. Iranmanesh, and F. L. Malallah, "Estimating body related soft biometric traits in video frames," *The Scientific World Journal*, vol. 2014, 2014.

[37] C. Velardo and J.-L. Dugelay, "What can computer vision tell you about your weight?" in *Proceedings of the IEEE European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 1980–1984.

[38] C. Velardo, J.-L. Dugelay, M. Paleari, and P. Ariano, "Building the space scale or how to weigh a person with no gravity," in *Proceedings of the IEEE International Conference on Emerging Signal Processing Applications*. IEEE, 2012, pp. 67–70.

[39] R. M. Tinlin, C. D. Watkins, L. L. Welling, L. M. DeBruine, E. A. Al-Dujaili, and B. C. Jones, "Perceived facial adiposity conveys information about women's health," *British Journal of Psychology*, vol. 104, no. 2, pp. 235–248, 2013.

[40] S. De Jager, N. Coetzee, and V. Coetzee, "Facial adiposity, attractiveness and health: a review," *Frontiers in Psychology*, vol. 9, p. 2562, 2018.

[41] I. D. Stephen, M. J. L. Smith, M. R. Stirrat, and D. I. Perrett, "Facial skin coloration affects perceived health of human faces," *International Journal of Primatology*, vol. 30, no. 6, pp. 845–857, 2009.

[42] I. D. Stephen, V. Hiew, V. Coetzee, B. P. Tiddeman, and D. I. Perrett, "Facial shape analysis identifies valid cues to aspects of physiological health in caucasian, asian, and african populations," *Frontiers in Psychology*, vol. 8, p. 1883, 2017.

[43] B. J. Lee and J. Y. Kim, "Predicting visceral obesity based on facial characteristics," *BMC Complementary and Alternative Medicine*, vol. 14, no. 1, p. 248, 2014.

[44] C. Mayer, S. Windhager, K. Schaefer, and P. Mitteroecker, "BMI and WHR are reflected in female facial shape and texture: a geometric morphometric image analysis," *PloS One*, vol. 12, no. 1, p. e0169336, 2017.

[45] M. A. Pascali, D. Giorgi, L. Bastiani, E. Buzzigoli, P. Henríquez, B. J. Matuszewski, M.-A. Morales, and S. Colantonio, "Face morphology: can it tell us something about body weight and fat?" *Computers in Biology and Medicine*, vol. 76, pp. 238–249, 2016.

[46] D. Giorgi, M. A. Pascali, P. Henriquez, B. J. Matuszewski, S. Colantonio, and O. Salvetti, "Persistent homology to analyse 3D faces and assess body weight gain," *The Visual Computer*, vol. 33, no. 5, pp. 549–563, 2017.

[47] L. Wen and G. Guo, "A computational approach to body mass index prediction from face images," *Image and Vision Computing*, vol. 31, no. 5, pp. 392–400, 2013.

[48] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, 1997, pp. 155–161.

[49] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Springer, 2003, pp. 63–71.

[50] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression: I," in *Breakthroughs in Statistics*. Springer, 1992, pp. 237–259.

[51] K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2006, pp. 341–345.

[52] M. Barr, G. Guo, S. Colby, and M. Olfert, "Detecting body mass index from a facial photograph in lifestyle intervention," *Technologies*, vol. 6, no. 3, p. 83, 2018.

[53] K. Wolffhechel, A. C. Hahn, H. Jarmer, C. I. Fisher, B. C. Jones, and L. M. DeBruine, "Testing the utility of a data-driven approach for assessing BMI from face images," *PloS one*, vol. 10, no. 10, p. e0140347, 2015.

[54] 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*. IEEE, 2014, pp. 1701–1708.

[55] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 815–823.

[56] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: deep hypersphere embedding for face recognition," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 212–220.

[57] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *European Conference on Computer Vision*. Springer, 2014, pp. 584–599.

[58] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2016.

[59] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 7291–7299.

[60] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European Conference on Computer Vision*. Springer, 2016, pp. 483–499.

[61] E. Kocabey, M. Camurcu, F. Ofli, Y. Aytar, J. Marin, A. Torralba, and I. Weber, "Face-to-BMI: using computer vision to infer body mass index on social media," in *11th International AAAI Conference on Web and Social Media*, 2017.

[62] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, "Deep face recognition," in *British Machine Vision Conference*, vol. 1, no. 3, 2015, p. 6.

[63] A. Dantcheva, F. Bremond, and P. Bilinski, "Show me your face and I will tell you your height, weight and body mass index," in *24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3555–3560.

[64] M. Jiang and G. Guo, "Body weight analysis from human body images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2676–2688, Oct 2019.

[65] M. Jiang, G. Guo, and Y. Shang, "A computational approach to body mass index estimation from dressed people in 3D space," *IET Image Processing*, 2020.

[66] M. Jiang, Y. Shang, and G. Guo, "On visual BMI analysis from facial images," *Image and Vision Computing*, vol. 89, pp. 183–196, 2019.

[67] M. Jiang, G. Guo, and G. Mu, "Visual BMI estimation from face images using label distribution based method," *Computer Vision and Image Understanding*, Under Major Revision.

[68] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2015, pp. 1529–1537.

[69] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 4724–4732.

[70] R. Drillis, R. Contini, and M. Bluestein, "Body segment parameters: a survey of measurement techniques," *Artificial Limbs*, vol. 8, pp. 44–66, 1963.

[71] K. Bushby, T. Cole, J. Matthews, and J. Goodship, "Centiles for adult head circumference." *Archives of Disease in Childhood*, vol. 67, no. 10, pp. 1286–1287, 1992.

[72] J. Weston and C. Watkins, "Multi-class support vector machines," Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, Tech. Rep., 1998.

[73] C. K. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, 1996, pp. 514–520.

[74] G. Guo, S. Z. Li, and K. L. Chan, "Support vector machines for face recognition." *Image and Vision Computing*, vol. 19, no. 9, pp. 631–638, 2001.

[75] A. Statnikov, D. Hardin, and C. Aliferis, "Using SVM weight-based methods to identify causally relevant and non-causally relevant variables." *Sign*, vol. 1, no. 4, 2006.

[76] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. MIT Press Cambridge, 2006.

[77] K. Pearson, "Note on regression and inheritance in the case of two parents." *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.

[78] W. C. Navidi, *Statistics for Engineers and Scientists*. McGraw-Hill New York, 2006.

[79] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems*, 2014, pp. 3320–3328.

[80] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[81] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.

[82] R. Ballard-Barbash and C. A. Swanson, "Body weight: estimation of risk for breast and endometrial cancers," *The American Journal of Clinical Nutrition*, vol. 63, no. 3, pp. 437S–441S, 1996.

[83] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin, "Multiscale rotation-invariant convolutional neural networks for lung texture classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 184–195, 2017.

[84] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers *et al.*, "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, 2018.

[85] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using kinects," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 643–650, 2012.

[86] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[87] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2011, pp. 127–136.

[88] A. Girdhar and V. Kumar, "Comprehensive survey of 3D image steganography techniques," *IET Image Processing*, vol. 12, no. 1, pp. 1–10, 2017.

[89] A. Moeini, K. Faez, and H. Moeini, "Real-world gender classification via local gabor binary pattern and three-dimensional face reconstruction by generic elastic model," *IET Image Processing*, vol. 9, no. 8, pp. 690–698, 2015.

[90] A. O. Bălan and M. J. Black, "The naked truth: estimating body shape under clothing," in *European Conference on Computer Vision*. Springer, 2008, pp. 15–29.

[91] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, and H.-P. Seidel, "Estimating body shape of dressed humans," *Computers & Graphics*, vol. 33, no. 3, pp. 211–216, 2009.

[92] C. Zhang, S. Pujades, M. J. Black, and G. Pons-Moll, "Detailed, accurate, human shape estimation from clothed 3D scan sequences." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2017, p. 3.

[93] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD*, vol. 96, no. 34, 1996, pp. 226–231.

[94] C. De Boor, C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor, *A Practical Guide to Splines*. Springer-Verlag New York, 1978, vol. 27.

[95] T. Tjahjowidodo, V. Dung, and M. Han, "A fast non-uniform knots placement method for b-spline fitting," in *Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics*. IEEE, July 2015, pp. 1490–1495.

[96] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, 1999.

[97] K. Yamaguchi, M. Hadi Kiapour, and T. L. Berg, "Paper doll parsing: retrieving similar styles to parse clothing items," in *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2013, pp. 3519–3526.

[98] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg, "Retrieving similar styles to parse clothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 5, pp. 1028–1040, 2015.

[99] J. H. Wilmore and A. R. Behnke, "An anthropometric estimation of body density and lean body weight in young men," *Journal of Applied Physiology*, vol. 27, no. 1, pp. 25–31, 1969.

[100] J. Wilmore and A. Behnke, "An anthropometric estimation of body density and lean body weight in young women," *The American journal of clinical nutrition*, vol. 23, no. 3, pp. 267–274, 1970.

[101] J. V. Durnin and J. Womersley, "Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years," *British Journal of Nutrition*, vol. 32, no. 1, pp. 77–97, 1974.

[102] F. Galton, "Regression towards mediocrity in hereditary stature." *The Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 15, pp. 246–263, 1886.

[103] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 112–119.

[104] D. Wang, "Fast and accurate volume calculation method for arbitary triangular meshes," *Computer Engineering and Application*, vol. 45, no. 18, p. 32, 2009.

[105] A. Z. Guo and M. Jiang, "Artificial intelligence techniques as potential tools for large scale surveillance and interventions for obesity," *Crimson Publishers*, vol. 3, 2020.

[106] X. Wu, R. He, Z. Sun, and T. Tieniu, "A light CNN for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[107] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 499–515.

[108] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2019, pp. 4690–4699.

[109] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of Electronic Imaging*, vol. 16, no. 4, 2007.

[110] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[111] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: a general-purpose face recognition library with mobile applications," *CMU School of Computer Science*, 2016.

[112] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*. ACM, 2014, pp. 675–678.

[113] M. De Marsico, M. Nappi, and D. Riccio, "Measuring measures for face sample quality," in *Proceedings of the 3rd International ACM Workshop on Multimedia in Forensics and Intelligence*. ACM, 2011, pp. 7–12.

[114] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D and 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)," in *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2017, pp. 1021–1030.

[115] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-Celeb-1M: a dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 87–102.

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

[117] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[118] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[119] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2401–2412, 2013.

[120] X. Geng and Y. Xia, "Head pose estimation based on multivariate label distribution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 1837–1842.

[121] C. Xing, X. Geng, and H. Xue, "Logistic boosting regression for label distribution learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4489–4497.

[122] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[123] B. Thompson, "Canonical correlation analysis," *Encyclopedia of Statistics in Behavioral Science*, 2005.

[124] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[125] S. Della Pietra, V. Della Pietra, and J. Lafferty, "Inducing features of random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 380–393, 1997.

[126] M. Diehl, H. G. Bock, and J. P. Schlöder, "A real-time iteration scheme for nonlinear optimization in optimal feedback control," *SIAM Journal on Control and Optimization*, vol. 43, no. 5, pp. 1714–1736, 2005.

[127] P.-W. Wang and C.-J. Lin, "Iteration complexity of feasible descent methods for convex optimization," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1523–1548, 2014.

[128] A. Safari and H. Shayeghi, "Iteration particle swarm optimization procedure for economic load dispatch with generator constraints," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6043–6048, 2011.

[129] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-CNN for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 5183–5192.

[130] B.-B. Gao, C. Xing, C.-W. Xie, J. Wu, and X. Geng, "Deep label distribution learning with label ambiguity," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2825–2838, 2017.

[131] W. Li, J. Lu, J. Feng, C. Xu, J. Zhou, and Q. Tian, "Bridgenet: a continuity-aware probabilistic network for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2019, pp. 1145–1154.

[132] L. Li and H.-T. Lin, "Ordinal regression by extended binary classification," in *Advances in Neural Information Processing Systems*, 2007, pp. 865–872.

[133] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *ArXiv Preprint ArXiv:1603.04467*, 2016.

[134] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, "Theano: a CPU and GPU math expression compiler," in *Proceedings of the Python for Scientific Computing Conference (SciPy)*, vol. 4, no. 3, 2010.

[135] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *ArXiv Preprint ArXiv:1411.7923*, 2014.

[136] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato, "Fader networks: manipulating images by sliding attributes," in *Advances in Neural Information Processing Systems*, 2017, pp. 5967–5976.

[137] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "Beta-vae: learning basic visual concepts with a constrained variational framework." *ICLR*, vol. 2, no. 5, p. 6, 2017.

[138] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: interpretable representation learning by information maximizing generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2016, pp. 2172–2180.

[139] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2019, pp. 4401–4410.

[140] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," *ArXiv Preprint ArXiv:1710.10196*, 2017.