IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 March 2024, accepted 10 April 2024, date of publication 18 April 2024, date of current version 3 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390412

RESEARCH ARTICLE

FishIR: Identifying Pufferfish Individual Based on Deep Learning and Face Recognition

YUAN LIN[1], (Member, IEEE), SHAOMIN XIE[2], (Graduate Student Member, IEEE), DEBASISH GHOSE[1], (Senior Member, IEEE), XIANGRONG LIU[2], JUNYONG YOU[3], (Senior Member, IEEE), JARI KORHONEN[4], (Member, IEEE), JUAN LIU[5], AND SOUMYA P. DASH[6], (Member, IEEE)

[1]School of Economics, Innovation, and Technology, Kristiania University College, 0153 Bergen, Norway
[2]School of Information Science and Technology, Xiamen University, Xiamen 361005, China
[3]NORCE Norwegian Research Centre, 5008 Bergen, Norway
[4]King's College, University of Aberdeen, AB24 3SW Aberdeen, U.K.
[5]School of Aerospace Engineering, Xiamen University, Xiamen 361000, China
[6]School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar, Odisha 752050, India

Corresponding author: Yuan Lin (yuan.lin@kristiania.no)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Pufferfish, globally recognized for its distinctive delicacy, carries high culinary value. However, it is also notorious for the lethal toxicity, and there is a great demand for traceability measures in the commercial trade of pufferfish to assure safety and accountability. This research introduces a novel deep learning approach, utilizing facial recognition techniques, to identify pufferfish individuals. This method specifically leverages distinctive back skin texture patterns as key biological traits. Our initial step involved assembling a collection of annotated and augmented images of Takifugu bimaculatus, a species of pufferfish native to East China Sea, which is accessible upon request. We then extensively investigated fundamental components of Deep Face Recognition (deep FR) systems, focusing on segmentation and extraction models, and assessed their effectiveness in identifying pufferfish. Following this, we developed FishIR (Fish Individual Recognition), a framework to identify pufferfish individuals that consists of four deep FR stages while incorporating enhanced segmentation and feature extraction techniques. Experimental results show that this framework successfully captures unique representations of individual pufferfish, as verified by the high accuracy achieved in recognition tasks.

INDEX TERMS Fish recognition, deep face recognition, deep learning.

I. INTRODUCTION

Pufferfish, known for its exquisite delicacy, unique texture, and nutritional value, has enjoyed enduring popularity as a luxury food ingredient, particularly in the eastern hemisphere, for many centuries. In 2019, the Ministry of Agriculture and Rural Areas of China made an announcement stating that the National Pedigree and Fine Aquatic Breed Verification Commission would grant approval for 14 new aquatic varieties, including new varieties of pufferfish [1]. This development has spurred the growth of pufferfish farming as a sustainable alternative to wild harvesting.

However, despite its culinary appeal, it is widely known that pufferfish can be highly dangerous due to the presence of tetrodotoxin, a toxic and fatal substance found in its liver and ovaries, persists even after cooking. As a result, the US Food and Drug Administration (FDA) has stringent regulations regarding the import of pufferfish to ensure public health [2]. In China, regulations have been implemented since 2016 to allow certified companies to farm Takifugu rubripes and Takifugu obscurus, which can be sold after undergoing proper processing and packaging with traceability codes

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

indicating their origin. However, the sale of fresh pufferfish remains banned [3]. Despite these measures, there has been a significant increase in cases of intoxication caused by wild pufferfish caught and consumed by individuals, highlighting the challenges in ensuring safety. The complex nature of the food supply chain, involving multiple economic stakeholders, makes it difficult to obtain reliable information about the origin and safety of food products, leading to possible incidents of food fraud and safety concerns [4]. There is an increasing awareness among customers about the safety and security of the food distribution network.

To address the aforementioned problem, techniques for identifying and tracking pufferfish individuals can prove valuable for the aquaculture and food processing industries. Relying on humans for individual identification is notably labor-intensive and prone to errors [5]. Previous attempts such as using alphanumeric or barcode systems [6], lack direct physical interaction with the fish and easily raises concerns about reliability in the food supply chain. Alternatively, the emerging approach based on Radio Frequency Identification (RFID) utilizes wearable devices and offers significant operational improvements compared to traditional methods. However, implanting RFID chips in fish poses a potential risk of injury [6], [7]. Moreover, RFID systems are generally expensive and involve labor-intensive processes. In contrast, a non-invasive computer vision method for individual fish identification minimizes the need for invasive techniques. Traditional approaches rely on shape and texture feature extraction [8], [9], [10], [11], [12], [13], which however have certain limitations, including sensitivity to background noise, restricted capacity for generalization, and challenges in identifying distinctive features [14].

Recent years have witnessed remarkable progress in the use of deep learning (DL) techniques across various cutting-edge disciplines, including ecology, biology, marine science, nutrition, and food engineering. Convolutional neural networks (CNNs), a type of deep learning architecture, have become fundamental in various computer vision tasks, ranging from image classification to object detection and semantic segmentation [15]. Deep learning has demonstrated its effectiveness in identifying and counting animal or plant species [16], [17]. Furthermore, it has been extensively researched and proven highly successful in individual identification, particularly in human-related studies [18]. More recently, similar methodologies have been adapted for individual recognition in other animal species, leveraging visual biometrics to capture unique and stable biological characteristics. Notable examples include individual recognition in primates [19], [20], pigs [21], cattle [22], [23], and elephants [24], employing computer vision approaches.

One major challenge in applying deep learning methods for individual recognition lies in acquiring a substantial amount of training data. This process usually involves annotating images with specific identities or attributes of individual entities. The amount of data required for training a CNN is contingent upon the complexity of the classification challenge, especially in cases of subordinate class recognition or identifying species associated with certain regions or habitats. Therefore, developing a robust data collection strategy is essential for effective individual recognition tasks.

Deep learning techniques have been successfully implemented in marine environments, i.e., underwater video and acoustic surveillance systems for the monitoring of fish species [25], fish recognition competitions such as the Kaggle challenge [26], among other research activities. Despite these advances, there has been very limited work in the area of individual fish recognition, possibly due to the historical preference for statistical approaches in marine biodiversity and ecosystem studies. The remarkable biodiversity of deep ocean, in terms of species count and variety, further complicates the task of individual fish identification. Yet, the growing interest in learning individual characteristics calls for the development of more robust and flexible methods in both aquaculture and food traceability sectors.

It has been noted that each pufferfish has a distinctive pattern on its back skin texture, similar to that of human fingerprints. This suggests that such unique texture patterns can be employed for pufferfish individual identification, analogically to the process of learning feature representations to distinguish human faces. Note that in this study, we specifically focus on the unique Chinese pufferfish species, Takifugu bimaculatus, which inhabits in East China sea. Unless stated otherwise, any reference to pufferfish in this paper will be concerning Takifugu bimaculatus.

This study introduces a well engineered DL framework to identify individual pufferfish, specifically leveraging the back skin texture patterns as key biological traits. By training an architecture similar to Deep Face Recognition (deep FR) on pufferfish images, we facilitate a cost-effective, non-invasive, user-friendly and robust approach for the identification of individual pufferfish. Previous attempts that incorporate face recognition techniques can be found in [27], where an architecture based on FaceNet [28] was proposed to identify salmon. And similarly, a novel face identification framework by integrating light-weight RetinaFace-mobilenet with Additive Angular Margin Loss (ArcFace), namely CattleFaceNet, was proposed in [23]. It should be emphasized that our approach is not merely a replication of existing deep FR architecture but rather a comprehensive examination and selection of the most effective methods among these fundamental components. Based on the evaluation, we proposed a unique DL framework tailored for pufferfish, named FishIR, structured similarly to the 4-stage deep FR, but augmented with enhanced segmentation and feature extraction techniques. This approach has achieved remarkable result and provides us with a deeper understanding of well engineered features.

The main contributions of this study are as follows:

- We have developed an annotated and augmented dataset for the Takifugu bimaculatus pufferfish, which is accessible upon request.

- We have proposed a DL framework for individual pufferfish recognition, introducing well-engineered face embedding techniques.

II. RELATED WORK

A. FACE RECOGNITION

Face recognition, a typical example of individual recognition, is a technology that can detect faces in video or images based on facial features and recognize their identities [29]. Traditional face recognition methodologies have predominantly relied on manually engineered facial features. The process often involves segmenting the face into distinct regions and subsequently applying conventional feature extraction techniques, such as Histogram of Oriented Gradients (HOG) [30], Local Binary Patterns (LBP) [31], Scale-Invariant Feature Transform (SIFT) [32], and Speeded Up Robust Features (SURF) [33], among others, to derive features. These extracted features are then aggregated to a composite representation of the overall facial features. The Eigenfaces method [34], for instance, is a notable method in this approach. Moreover, certain hybrid methods employ feature-based approaches for feature extraction, followed by dimensionality reduction techniques to achieve compact, low-dimensional features.

Inspired by the breakthrough work launched by DeepFace [35], DeepID [36], and FaceNet [28], research in face recognition focus has shifted to DL-based approaches and has reached a high level of maturity [37], [38]. The deep FR framework typically consists of four phases: face detection, face alignment, facial feature extraction, and face matching.

In our work, four similar steps have been used to identify pufferfish. The pufferfish image is scaled to a dimension of 1500×800, which ensures standardization across all images. Our approach utilizes the deep FR's generic segmentation component to separate pufferfish from the background. Following this, we perform mask alignment to address intra variations such as poses, illuminations, expressions, and occlusions both during training and testing. Following this, a feature extractor is employed to extract the back skin features during testing. The final stage involves the computation of a Euclidean space feature vector to assess the resemblance between the features extracted from the pufferfish image and those stored in the database, facilitating the feature matching process [39].

B. SEGMENTATION

Image segmentation separates the target object from the background in an image. Traditional approaches segment images into distinct regions based on grayscale, color, spatial texture, and geometric shape criteria. These include Threshold-Based Method (e.g., Otsu's thresholding algorithm [40]), Cluster-Based Method (e.g., K-means clustering [41]), Area-Based Method [42], and Edge-Based Method (e.g., Sobel and Canny edge detectors [43], [44]) approaches, which are still widely used.

Deep learning based segmentation methods, on the other hand, leverage neural networks to automatically learn and extract meaningful features for segmentation tasks. In this study, we will present several representative models for semantic segmentation and instance segmentation tasks, including Fully Convolutional Networks (FCNs), Encoder-Decoder Based Models, Pyramid-Based Models, R-CNN Based Models, and Dilated Convolution Based Models.

1) SEMANTIC SEGMENTATION

Semantic segmentation aims at classifying each pixel in an image to a specific class, without distinguishing between different instances of the same class.

a: FULLY CONVOLUTIONAL NETWORKS

Neural networks composed of only convolutional layers, known as Fully convolutional networks (FCNs), have been applied to a variety of segmentation tasks as demonstrated in [45] and [46]. FCN-8, a variant initially proposed by [47], is employed in this study for semantic image segmentation.

b: ENCODER-DECODER ARCHITECTURES

The encoder-decoder based architecture is usually composed of an encoder that employs convolutional layers derived from VGG16 and a deconvolution layer that that offers segmentation masks and pixel-wise class labels. Our study evaluates a recently developed encoder-decoder architecture known as SegNet [48], which has achieved notable success in the field.

c: PYRAMID NETWORK BASED MODELS

Feature Pyramid Network (FPN), as first introduced by [49], leverages the multi-scale, pyramidal structure intrinsic to deep convolutional networks for the creation of feature pyramids with marginal extra cost. The Pyramid Scene Parsing Network (PSPNet) [50] further enhances global context representation in scenes. We evaluate PSPNet as a benchmark method for performance comparison.

d: DILATED/ATROUS CONVOLUTIONAL MODELS

Dilated convolution employs sparse convolution kernels to enlarge the receptive field through dilation rates in to convolutional layers. The DeepLab family [51], [52], [53] uses dilated convolution to aggregate multi-scale contextual information while preserving resolution. In 2018, DeepLabV3+ was released in [54], achieving significant performance in the PASCAL VOC challenge after pretraining on diverse datasets.

2) INSTANCE SEGMENTATION

On the other hand, instance segmentation goes beyond categorizing each pixel in an image by its class, by identifying and delineating each distinct instance.

R-CNN Based Models: Among the models that are developed to handle instance segmentation tasks, the Region-based

Convolutional Neural Networks (R-CNN) and its successors represent a significant advancement in the field of object detection and segmentation tasks. Mask R-CNN, introduced by He et al. [55], can efficiently detect objects in images while simultaneously generating precise segmentation masks for each detected instance. Their subsequent work, Mask Scoring R-CNN [56], has further introduced a mask scoring mechanism that calibrates the misalignment between mask quality and mask score, leading to a significant performance enhancement. Both Mask R-CNN and Mask Scoring R-CNN are evaluated in this study.

C. FEATURE EXTRACTION

The significance of feature extraction in face recognition lies in its ability to reduce dimension while preserving informative features. Classical feature extraction methods often involve the use of filters or descriptors to extract key information and highlight distinguishing characteristics. Representative examples include the implementation of Gabor filters [57], Local Binary Pattern (LBP) [31], and Gray-level Co-occurrence Matrix (GLCM) [58] into face extraction. Detailed information can be found in their respective original work for interested readers.

In recent years, CNNs have become increasingly prominent for feature extraction, primarily due to their ability of learning hierarchical representations and identifying complex patterns in data. Face features are considered discriminative when they demonstrate a high degree of intra-class similarity and inter-class dissimilarity. There is a growing research efforts on incorporating well crafted classification loss functions, as well as adopting efficient architectural designs to enhance discriminative power of the learned features. In this study, we will review several significant achievements in classification loss functions design.

1) SOFTMAX LOSS

In the context of face recognition, the development of margin-based softmax loss functions is crucial for acquiring discriminative features, as highlighted in the study by [59]. Examples include angular, additive, additive angular margins, etc.

Softmax loss consists of a series of components, including the last fully connected layer, the softmax function and the cross-entropy loss. In the following formulation, with d being the feature dimension and K being the number of classes, $w_k \in R^d$ represents the weights of k-th classier where $k \in 1, 2, \ldots, K$, while $x \in R^d$ is the feature vector associated with input data.

$$L_1 = -log \frac{e^{w_y^T x}}{e^{w_y^T x} + \sum_{k \neq y}^{K} e^{w_k^T x}}$$

In face recognition, it is a common practice to transform the original softmax loss into a cosine similarity-based formulation. Given an input feature vector x with its ground truth label y, the cosine similarity is calculated by $\cos\left(\theta_{w_k,x}\right) = w_k^T x$, where the angle between w_k and x is represented by $\theta_{w_k,x}$.

$$L_2 = -\log \frac{e^{s\cos\left(\theta_{wy,x}\right)}}{e^{s\cos\left(\theta_{wy,x}\right)} + \sum_{k \neq y}^{K} e^{s\cos\left(\theta_{w_k,x}\right)}}$$

2) LOSS FUNCTIONS AND ITS VARIANCE

In following we will present three different loss functions which are evaluated in our experiments. The interested readers can refer to [59], [60], and [61] for theoretical details.

a: ANGULAR MARGIN SOFTMAX

The angular softmax (A-softmax) [60], modifies the traditional softmax loss to enforce an angular margin (A-Softmax) between feature vectors and their corresponding class centers in the embedding space. It is designed to enhance the discriminative power of deep learning models, especially in face recognition related tasks.

b: ADDITIVE ANGULAR MARGIN LOSS

A-Softmax, while effective, can be sensitive to parameter settings. Deng et al. [59] developed an additive angular margin loss, which has a clear geometric interpretation to address the stability of A-Softmax loss.

c: LARGE MARGIN COSINE LOSS

Large Margin Cosine Loss, was proposed in [61] to maximize inter-class variance and minimize intra-class variance from a different perspective, in contrast to traditional softmax loss and angular softmax loss. It reformulates the softmax loss as a cosine loss by L2 normalizing both the feature vectors and weight vectors to remove radial variations. Additionally, a cosine margin term is introduced to further maximize the decision margin in the angular space.

III. MATERIALS AND METHODS

A. DATA COLLECTION AND PROCESSING

Recent advances in deep learning based face recognition have been driven by the availability of extensive, large annotated databases, that enable the extraction of comprehensive and concise facial representations. Several benchmark datasets are now available for researchers to assess their algorithms, such as PASCAL [62], MS COCO [63], and ILSVRC [64]. To the best of our knowledge, there is no existing publicly accessible dataset of labeled pufferfish. To bridge the gap and foster pufferfish identification, we have constructed a properly annotated pufferfish database in this study.

In order to construct our dataset for Takifugu bimaculatus, we followed a systematic approach. We initiated the dataset construction process by collecting labeled pictures of takifugu bimaculatus, a species of pufferfish found in the East China Sea. To obtain these labeled pictures, we utilized video clips and extracted frames from the recorded footage. This process involved acquiring the video clips and subsequently extracting individual frames, which were then labeled and

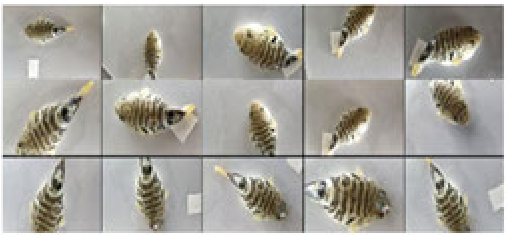


FIGURE 1. Takifugu bimaculatus images extracted and filtered from one video clip.

incorporated into our dataset. The specific implementation details of this process are outlined below:

- We carefully set up the photography environment and device for clear image acquisition. A D65 light source with a color temperature of 6500K was applied to ensure a light background with a distinct contrast to the fish's color.
- To document the various features of the fish, we recorded video clips from multiple angles (e.g., 30 or 45 degrees), allowing the camera to rotate for about 60 seconds. From each video segment recorded, we derived approximately 200 images.
- Finally, we categorized and annotated the images according to a specific naming convention, ensuring systematic organization and easy reference within the dataset.

By following these steps, we constructed a comprehensive dataset for Takifugu bimaculatus. In our data collection process, we acquired a total of 146 video clips, with each clip capturing an individual Takifugu bimaculatus. To ensure consistency, we scaled each image to the dimensions of 1500 × 800. From these video clips, we extracted and meticulously filtered a total of 20,793 high-quality pufferfish images. Fig. 1 showcases a selection of Takifugu Bimaculatus images that were extracted and filtered from one video clip.

To create a dataset for the segmentation task, we employed the MS COCO-style approach [63]. We utilized 30 different Takifugu bimaculatus fish, encompassing a total of 1,267 images, as the training dataset for this relatively straightforward task. We utilized the LabelImg software [65] to manually label and annotate the images. For feature extraction, we utilized the Labeled Faces in the Wild (LFW) method, which incorporates pair matching to organize the dataset. The dataset was divided into verification and identification subsets. The verification subset contained 126 fish, while the identification subset included 20 fish. Within the verification subset, the dataset was further split into training, verification, and recognition sets at a ratio of 8:1:1.

TABLE 1. Dataset before and after data augmentation.

Task	Before Augmentation	After Augmentation
Segmentation	1267	7620
Feature Extraction	20793	55837

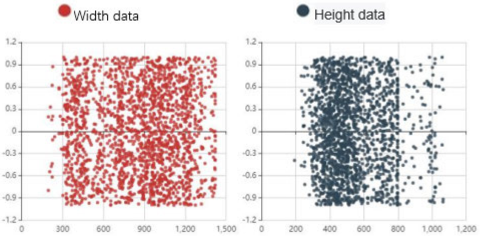


FIGURE 2. Scatter diagram of bounding box width and height from training dataset.

B. DATA AUGMENTATION

Data augmentation, achieved through various transformations such as translation, reflection, rotation, warping, scaling, color space shifting, and cropping are usually applied to enhance the training sample diversity. This further leads to faster convergence, reduced overfitting, and improved generalization, especially when training neural networks with limited datasets. We employed several augmentation methods in our study. The number of images employed for segmentation task has been expanded from 1,267 to 7,620, while for feature extraction, it has increased from 20,793 to 55,837, as shown in Table 1.

C. EXPERIMENTAL SETUP

We conducted experiments using NVIDIA GeForce GTX 2080 Ti GPUs on a workstation running the Linux Ubuntu 16.04 LTS operating system. To create the software environment for DL, we utilized Python and PyTorch [66]. Stochastic gradient descent (SGD) served as the optimization algorithm across all experiments, ensuring efficient training and convergence of the models.

IV. EXPERIMENTS

In this section, we present a comprehensive examination of the essential components of deep FR, focusing particularly on segmentation and feature extraction models, and exploring their applicability for the identification of pufferfish.

A. SEGMENTATION EXPERIMENTS

For the segmentation experiments, we used a training dataset consisting of 30 distinct Takifugu bimaculatus fish, with a total number of 1,267 images. For instance segmentation, we assessed the performance of Mask R-CNN [55] and Mask Scoring R-CNN [56]. For semantic segmentation, we explored the effectiveness of DeepLabV3+ [54], PSPNet [50], FCN-8 [47], and SegNet [48].A learning rate of 0.0025 was applied to Mask R-CNN and Mask Scoring R-CNN, while the remaining models were configured with a learning rate of 0.01. These settings were carefully chosen to optimize the training process and achieve optimal segmentation results.

1) PERFORMANCE METRICS

We use the standard COCO metrics to assess the model performance in terms of segmentation accuracy, which include Mean Pixel Accuracy (MPA) and Mean Intersection

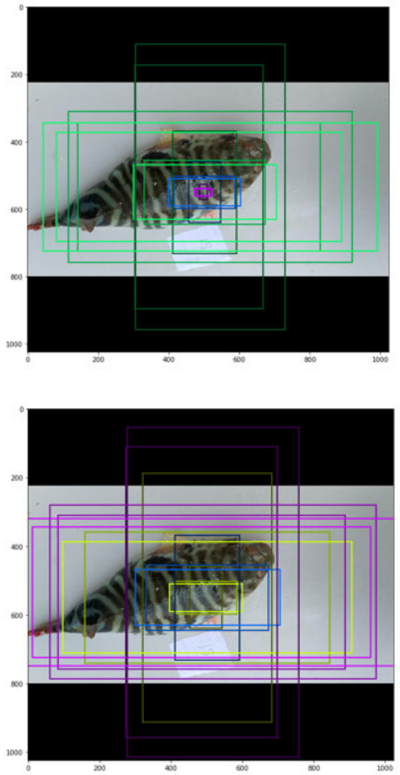


FIGURE 3. Two different sets of anchor boxes.

TABLE 2. MPA and MIoU of two different sets of anchor boxes.

Anchor box set size	MPA	MIoU
(32,128,256,512,600) #1	0.9859	0.9733
(128,256,512,600,680) #2	0.9858	0.9732

TABLE 3. Quantitative performance analysis of image segmentation models with different backbone networks.

Model	Network	MPA	MIoU	Time (s)	Size
Mask R-CNN	ResNet_50	0.9884	0.9775	0.0659	335M
	ResNet_101	0.9894	0.9772	0.0869	480M
	MobileNet	-	-	-	-
	ShuffleNet	-	-	-	-
Mask Scoring R-CNN	ResNet_50	0.9883	0.9773	0.0886	459M
	ResNet_101	0.9889	0.9774	0.0662	604M
	MobileNet	0.7518	0.5276	0.0442	295M
	ShuffleNet	0.9840	0.9699	0.0662	309M
DeepLabV3+	ResNet_101	0.9949	0.9856	0.0344	453M
	MobileNet	0.9827	0.9761	0.0196	45M
	DRN	0.9886	0.9838	0.0211	311M
PSPNet	-	0.9927	0.9850	19.8835	393M
FCN-8	-	0.9915	0.9822	11.9410	1.1G
SegNet	-	0.9872	0.9783	15.6391	125M

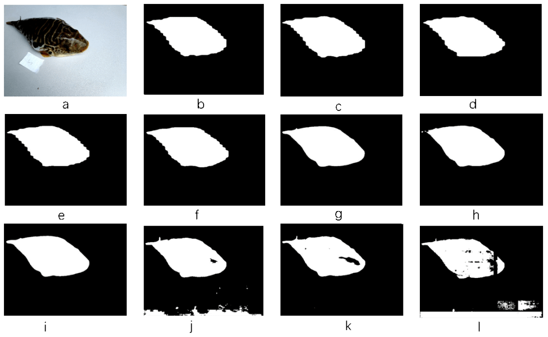


FIGURE 4. Visual comparison of image segmentation models using different backbone networks.

over Union (MIoU) [67]. MPA measures the proportion of correctly classified pixels across all classes, and MIoU calculates the ratio between the intersection and union of the ground truth and predicted segmentation. The equations for MPA and MIoU are as follows:

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}}$$

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}$$

2) ANCHOR SIZE

Anchor box selection is a process of choosing predefined bounding boxes with different sizes and ratios, serving as a reference to object detection in CNN models such as Mask R-CNN and Mask Scoring R-CNN. By selecting an appropriate set of anchors, we can improve both the speed and accuracy of our models. For performance optimization of Mask R-CNN and Mask Scoring R-CNN, we conducted an evaluation of choosing the optimal anchor size, following a K-means clustering approach to cluster the bounding boxes of objects, inspired by YOLO [68]. The clustering process helped to identify suitable anchor box sizes and propose potential regions, which in turn guided a more accurate object localization.

We randomly selected 2000 (width, height) data points and depicted the scatter diagram in Fig. 2. It can be seen that the height data is mainly grouped at 400, while width data is clustered at 300-400 and 900-1200. Based on the clustering results, we derived two sets of anchor boxes: one with the sizes (32,128,256,512,600) and the other with sizes (128,256,512,600,680), and both with aspect ratio set to (0.5,1.8,2.5), as shown in Fig. 3. After 10, 000 iterations of training, we observed that the first anchor box set delivered marginally better results in terms of MPA and MIoU, as detailed in Table 2. This suggests us to apply the first anchor box set configuration with size (32,128,256,512,600) and aspect ratio (0.5,1.8,2.5) in both Mask R-CNN and Mask Scoring R-CNN models.

3) NETWORK ARCHITECTURE

Several ablation experiments were conducted to assess the impact of backbone networks. We tested performance of ResNet_50, ResNet_101, MobileNet, ShuffleNet on Mask Scoring R-CNN, ResNet_50, ResNet_101 on Mask R-CNN, and ResNet_101, MobileNet, DRN on DeepLabV3+ respectively. The quantitative data and visual representations are presented in Table 3 and Fig. 4 respectively. Note that the order of pictures in Fig. 4 is aligned with the sequence in Table 3, arranged from left to right and top to bottom. All backbone networks were pre-trained in advance. Furthermore, Table 3 also includes results from PSPNet, FCN-8 and SegNet, which are optimized without any modifications of underlying backbone network.

For instance segmentation tasks, Mask Scoring R-CNN demonstrates only a marginal performance gain over Mask R-CNN across all backbone networks, likely due to the fact that segmentation of single object has limited sensitivity to a more precise scoring function. Additionally, Mask Scoring R-CNN has yielded a larger number of model parameters due to its additional prediction branch. For semantic segmentation, DeepLabV3+ based on a ResNet_101 backbone network achieved the highest MPA and MIoU at 99.49% and 98.56%, respectively. It's worth noting that DeepLabV3+ maintains an efficient average processing time of 0.034 seconds. Conversely, MobileNet, while offering the fastest processing speed and smallest model size, sacrifices performance in terms of MPA and MIoU. PSPNet and FCN-8 are not efficient in terms of both speed and model size, despite their high MPA and MIoU. As a result, DeepLabV3+ has been selected as the segmentation model for our application.

B. FEATURE EXTRACTION EXPERIMENTS

1) TRADITIONAL FEATURE EXTRACTION

We first applied traditional methods such as LBP [31], GLCM [58], and Gabor [57] for feature extraction. These classical feature extraction methods require no training procedures, and the parameter configuration during experimental setup can be directly derived from empirical studies. Among these approaches, Gabor performs best in comparison to other methods. However, its accuracy in feature extraction stands merely at 50.31%, and the method demands considerable computational time. The experimental results are depicted in Table 4.

TABLE 4. Experimental results of 3 traditional methods: LBP, GLCM, and Gabor.

Algorithm	ACC	time(s)	threshold
LBP	0.0120	0.0840	0.0725
GLCM	0.0278	0.0553	0.0225
Gabor	0.0.5031	7.9931	0.0225

2) LOSS FUNCTION IN CNNs

During the feature extraction process using CNNs, training the network is an initial step. Subsequently, the validation step is performed on a verification set. In our experiments, we utilized the LFW-style dataset [69], which consists of 4, 158 pairs of sample images. We employed the ten-fold cross-validation approach on the test dataset. In each trial, we randomly selected nine folds for training and allocating the remaining fold for testing. The final accuracy results were obtained by averaging the accuracy across ten distinct trials.

To assess the resemblance between pairs of images, we measured the Euclidean distance value for each pair.

TABLE 5. Various backbone networks and their performance across different loss functions.

Model	Loss Function	ACC	Threshold	Time (s)	Size
ResNet_50	AAML	0.9997	1.1124	0.2111	250M
	LMCL	0.9985	1.2450	0.2243	250M
	A-Softmax	0.9987	1.0900	0.2207	250M
ResNet_101	AAML	0.9997	1.2974	0.3883	323M
	LMCL	0.9990	1.2025	0.4152	323M
IR_50	AAML	0.9992	1.1475	0.2726	251M
	LMCL	0.9963	1.4049	0.3066	251M
IRSE_50	AAML	0.9987	1.2450	0.3592	252M
	LMCL	0.9997	1.2050	0.3596	252M

TABLE 6. The ResNet_50_*rc* structure is composed of building blocks, each detailed within brackets alongside the respective stack counts. Downsampling is performed at four stages with a stride of 2.

layer name	output size	ResNet_50_rc
stage_0	150 x 350 73 x 175	7 x 7, 64, stride 2 7 x 3 max pool, stride 2
stage_1	73 x 175	$\begin{bmatrix} 1\times 1, 64 \\ 7\times 3, 64 \\ 1\times 1, 256 \end{bmatrix} \times 3$
stage_2	37 x 88	$\begin{bmatrix} 1\times 1, 128 \\ 7\times 3, 128 \\ 1\times 1, 512 \end{bmatrix} \times 4$
stage_3	19 x 44	$\begin{bmatrix} 1\times 1, 256 \\ 7\times 3, 256 \\ 1\times 1, 1024 \end{bmatrix} \times 6$
stage_4	10 x 22	$\begin{bmatrix} 1\times 1, 512 \\ 7\times 3, 512 \\ 1\times 1, 2048 \end{bmatrix} \times 3$
	5 x 11	3 x 5 average pool, stride 2

TABLE 7. Results using 7 × 3 rectangular kernels in stages 1 to 4 of ResNet$_{50}$.

model	stage_1	stage_2	stage_3	stage_4	ACC
ResNet_50_$rc1$	x	x	x	x	0.9992
ResNet_50_$rc2$		x	x	x	0.9995
ResNet_50_$rc3$			x	x	0.9997
ResNet_50_$rc4$				x	0.9997

We categorized each image pair as either similar or dissimilar, based on a predetermined threshold value s. For each image pair, this process yields one of four outcomes: a True Positive (TP) if the pair is correctly identified as similar, a True Negative (TN) if the pair is correctly identified as dissimilar, a False Positive (FP) if the pair is misclassified as similar, and a False Negative (FN) if the pair is misclassified as dissimilar. The accuracy is calculated by the ratio of the sum of TP and TN to the overall number of pairs, represented by the equation:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

In our feature extraction experiments, we focused on assessing the impact of different loss functions while keeping the backbone network architecture consistent. The backbone networks selected for these experiments were ResNet_50, ResNet_101, IR_50 and IRSE_50, where IR_50 and IRSE_50 are minor modifications of ResNet_50. We tested the following loss functions: Angular Softmax Loss [60], Large Margin Cosine Loss [61], and Additive

TABLE 8. A summary of recognition results on the pufferfish dataset.

Segmentation model	Feature Extraction	Loss function	TPR	FPR	Time(s)
MobileNet	ResNet_50	AAML	0.9432	0.0010	0.4644
DRN	ResNet_50	AAML	0.9492	0.0010	0.6689
ResNet_101	ResNet_50	AAML	0.9485	0.0011	0.7439
MobileNet	ResNet_50	LMCL	0.9423	0.0067	0.4993
DRN	LMCL	ResNet_50	0.9412	0.0065	0.8865
ResNet_101	ResNet_50	LMCL	0.9429	0.0066	0.7438
MobileNet	ResNet_50	A-Softmax	0.9312	0.0053	0.4994
DRN	ResNet_50	A-Softmax	0.9374	0.0049	0.8859
ResNet_101	ResNet_50	A-Softmax	0.9361	0.0044	0.4436

Angular Margin Loss [59] to compare their performance. The ablation study results presented in Table 5 indicate that most state-of-the-art CNNs are robust to perform feature extraction task with the loss functions assessed in this study. Note that the threshold value referenced in Table 5 corresponds to the aforementioned Euclidean distance value.

The experimental results lead to several important observations: Firstly, deeper and larger neural networks in general yield better performance, i.e., ResNet_101 consistently achieved a high ACC value that exceeds 99.9%. Secondly, ResNet_50 backbone using AAML is the fastest method to learn representative features. In contrast, ResNet_101 using LMCL requires the longest time. Finally, AAML is outperformed than LMCL for most backbone networks, with IRSE_50 being the exception.

Observing from a time perspective, ResNet_50 with the AAML loss function was the most efficient, requiring only 0.21 seconds to extract the features of a single image. On the contrary, ResNet_101 with the LMCL loss function was the most time-intensive, demanding a total of 0.41 seconds to perform feature extraction for a single image.

When considering the model size, ResNet_50, IR_50, and IRSE_50 exhibit similar dimensions, all approximately around 250M. However, ResNet_101 exhibits a slightly larger size of 323M.

3) CONVOLUTIONAL KERNELS

Inspired by the observations of rectangular patterns on pufferfish back skin, we hypothesized that employing rectangular convolutional kernels, as opposed to the conventional square-shaped kernels, might lead to more efficient and precise feature representation. Our study integrated 7 × 3 rectangular convolutional kernel with padding dimensions 3 × 1 and a stride of 2, into different stages of ResNet_50, as shown in Table 6. Surprisingly, the application of rectangular convolutional kernels resulted in only marginal performance improvements or, in some cases, even a decrease in accuracy, as demonstrated in Table 7.

C. FishIR: A DL BASED PUFFERFISH RECOGNITION ARCHITECTURE

In this Section, we present a DL framework named FishIR (Fish Individual Recognition) based on the extensive study of essential building blocks of deep FR. This framework is composed of four deep FR stages, while tailored by incorporating enhanced segmentation and feature extraction methods for pufferfish identification. Fig. 5 shows the architecture of the proposed FishIR framework.

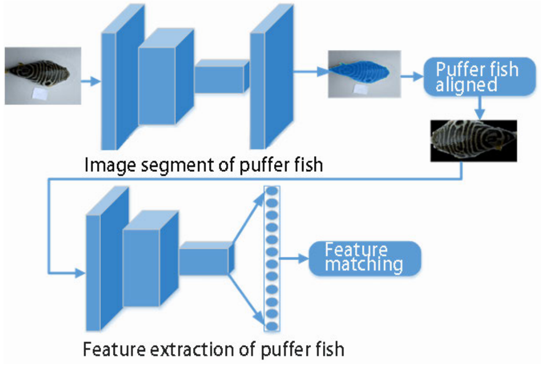

Puffer fish aligned
Image segment of puffer fish
Feature matching
Feature extraction of puffer fish

FIGURE 5. Architecture of FishIR: an individual pufferfish recognition system.

We have selected DeepLabV3+ coupled with ResNet_50 as the backbone network for feature extraction in our segmentation model. The pre-trained model parameters were utilized to initialize weights. The performance evaluation of this model was conducted using three distinct loss functions. To evaluate the model, we use TPR (true positive rate) and FPR (false positive rate) metrics from the confusion matrix. The experimental results are summarized in Table 8.

$$FPR = \frac{FP}{TN + FP}; \quad TPR = \frac{TP}{TP + FN}$$

We can draw several conclusions from the experimental results. First, network architectures affect speed more than accuracy. Secondly, the efficiency in learning representative features can vary depending on the choice of loss functions, and AAML performs the best for individual pufferfish recognition tasks among all tested loss functions. Last while not least, our experiments demonstrate that FishIR is highly effective in the precise recognition of individual pufferfish based on the analysis of their unique back skin patterns.

V. CONCLUSION

This study presents FishIR, a novel DL framework tailored to recognize individual pufferfish of the species Takifugu bimaculatus, which are native to the East China Sea. This system incorporates principles from face recognition techniques by leveraging unique back skin texture patterns as key biological traits and achieved significant performance proved by experimental results. To facilitate identity recognition, we constructed a collection of annotated and augmented images of Takifugu bimaculatus, which is accessible for scholarly use upon request. Our methodology entailed a comprehensive evaluation of fundamental components of deep Face Recognition (deep FR), with a particular focus on segmentation and extraction models, to gauge their effectiveness in pufferfish identification. As a result, we introduced FishIR (Fish Individual Recognition), a novel system that integrates the four fundamental stages of deep FR technology while incorporating advanced segmentation and feature extraction techniques tailored for pufferfish identification.

Our experiments show that training a facial recognition model on pufferfish images enables accurate individual pufferfish identification without physical intervention. The experimental results could inspire further research efforts in applying deep learning and facial recognition mechanisms in animal ecology, identify recognition and marine science. It could open up new research areas requiring long-term monitoring of individual animals, such as studying feeding behavior, disease detection, and social interactions.

In future work, we aim to test the model's ability to recognize individuals in the wild environment. We also plan to examine potential improvements in performance by incorporating other critical variants. Finally, we would like to test the generalizability of the concept to other species beyond pufferfish.

REFERENCES

[1] *The Fourth Edition of the Chinese Fish Price Report*, Fisheries Aquaculture Dept., Food Agricult. Organizations United Nations (FAO), FAO, Rome, Italy, 2019.

[2] US FDA. (2022). *Exchange of Letters Between Japanese Ministry of Health and Welfare and the Us Food and Drug Administration*. [Online]. Available: https://www.fda.gov/international-programs/cooperative-arrangements/fdajapan-exchange-letters-regarding-puffer-fish

[3] X. Yin, R. Xing, Z. Li, B. Hu, L. Yang, R. Deng, J. Cao, and Y. Chen, "Real-time qPCR for the detection of puffer fish components from lagocephalus in food: *L. inermis, L. lagocephalus, L. gloveri, L. lunaris, and L. spadiceus*," *Frontiers Nutrition*, vol. 9, Dec. 2022, Art. no. 1068767.

[4] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of deep learning in food: A review," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 18, no. 6, pp. 1793–1811, Nov. 2019.

[5] B. G. Weinstein, "A computer vision for animal ecology," *J. Animal Ecol.*, vol. 87, no. 3, pp. 533–545, May 2018.

[6] A. Regattieri, M. Gamberi, and R. Manzini, "Traceability of food products: General framework and experimental evidence," *J. Food Eng.*, vol. 81, no. 2, pp. 347–356, Jul. 2007.

[7] A. B. Eilertsen, J. S. Dyrstad, and M. S. Bondø, "Identifikasjon AV lakseindivider—Biometri fase 1 (salmID)," Tech. Rep., 2017.

[8] H. S. Yi, "Fish observation, detection, recognition and verification in the real world," in *Proc. Int. Conf. Image Process., Comput. Vis., Pattern Recognit.*, 2012, pp. 1–6.

[9] K. Blanc, D. Lingrand, and F. Precioso, "Fish species recognition from video using SVM classifier," in *Proc. 3rd ACM Int. Workshop Multimedia Anal. Ecological Data*, Nov. 2014, pp. 1–6.

[10] H. Yu, Z. Wang, H. Qin, and Y. Chen, "An automatic detection and counting method for fish lateral line scales of underwater fish based on improved YOLOv5," *IEEE Access*, vol. 11, pp. 143616–143627, 2023.

[11] A. Rova, G. Mori, and L. Dill, "One fish, two fish, butter fish, trumpeter: Recognizing fish in underwater video," in *Proc. Mach. Vis. Appl.*, 2007, pp. 404–407.

[12] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1491–1498.

[13] F. Wu, Z. Cai, S. Fan, R. Song, L. Wang, and W. Cai, "Fish target detection in underwater blurred scenes based on improved YOLOv5," *IEEE Access*, vol. 11, pp. 122911–122925, 2023.

[14] M. Ravanbakhsh, M. R. Shortis, F. Shafait, A. Mian, E. S. Harvey, and J. W. Seager, "Automated fish detection in underwater images using shape-based level sets," *Photogrammetric Rec.*, vol. 30, no. 149, pp. 46–62, Mar. 2015.

[15] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," in *Computational Intelligence and Neuroscience*, 2018.

[16] M. A. Tabak, M. S. Norouzzadeh, D. W. Wolfson, S. J. Sweeney, K. C. Vercauteren, N. P. Snow, and R. S. Miller, "Machine learning to classify animal species in camera trap images: Applications in ecology," *Methods Ecol. Evol.*, vol. 10, no. 4, pp. 585–590, Apr. 2019.

[17] M. Rzanny, M. Seeland, J. Wäldchen, and P. Mäder, "Acquiring and preprocessing leaf images for automated plant identification: Understanding the tradeoff between effort and information gain," *Plant Methods*, vol. 13, no. 1, pp. 1–11, Dec. 2017.

[18] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.-C. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, "Deep learning for understanding faces: Machines may be just as good, or better, than humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, Jan. 2018.

[19] D. Deb, S. Wiper, S. Gong, Y. Shi, C. Tymoszek, A. Fletcher, and A. K. Jain, "Face recognition: Primates in the wild," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–10.

[20] D. Schofield, A. Nagrani, A. Zisserman, M. Hayashi, T. Matsuzawa, D. Biro, and S. Carvalho, "Chimpanzee face recognition from videos in the wild using deep learning," *Sci. Adv.*, vol. 5, no. 9, Sep. 2019, Art. no. eaaw0736.

[21] M. F. Hansen, M. L. Smith, L. N. Smith, M. G. Salter, E. M. Baxter, M. Farish, and B. Grieve, "Towards on-farm pig face recognition using convolutional neural networks," *Comput. Ind.*, vol. 98, pp. 145–152, Jun. 2018.

[22] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark, "Individual cattle identification using a deep learning based framework," in *Proc. 6th IFAC Conf. Sens., Control Autom. Technol. Agricult.*, vol. 52, 2019, pp. 318–323.

[23] B. Xu, W. Wang, L. Guo, G. Chen, Y. Li, Z. Cao, and S. Wu, "CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106675.

[24] M. Körschens, B. Barz, and J. Denzler, "Towards automatic identification of elephants in the wild," 2018, *arXiv:1812.04418*.

[25] M. Moniruzzaman, "Deep learning on underwater marine object detection: A survey," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2017, pp. 150–160.

[26] Kaggle. (2017). *The Nature Conservancy Fisheries Monitor*. Kaggle Competition. [Online]. Available: https://www.kaggle.com/c/the-nature-conservancy-fisheries-monitoring

[27] B. M. Mathisen, "FishNet: A unified embedding for salmon recognition," in *Proc. 9th Int. Conf. Prestigious Appl. Intell. Syst.*, 2020, pp. 1–8.

[28] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.

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

[30] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *Proc. Int. Workshop Autom. Face Gesture Recognit.*, vol. 12, 1995, pp. 296–301.

[31] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[32] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Oct. 1999, pp. 1150–1157.

[33] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 3951, 2006, pp. 404–417.

[34] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, May 1994, pp. 704–708.

[35] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[36] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.

[37] S.-H. Lin, "An introduction to face recognition technology," *Informing Sci., Int. J. Emerg. Transdiscipline*, vol. 3, pp. 1–7, Jan. 2000.

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

[39] D. Lucas and N. Helen, "Facial recognition technology," A Surv. Policy Implement. Issues Lancaster Univ., U.K. Centre Study Technol. Org., Tech. Rep., 2009.

[40] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[41] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, 1967, pp. 281–297.

[42] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 641–647, Jun. 1994.

[43] I. Sobel and G. Feldman, "A 3×3 isotropic gradient operator for image processing," *Pattern Classification Scene Anal.*, pp. 271–272, 1973.

[44] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[45] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proc. Int. MICCAI Brainlesion Workshop*, Sep. 2017, pp. 178–190.

[46] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.

[47] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[48] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[49] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[50] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2881–2890.

[51] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[52] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[53] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[54] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[55] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[56] Z. J. Huang, L. C. Huang, Y. C. Gong, C. Huang, and X. G. Wang, "Mask scoring R-CNN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 6409–6418.

[57] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[58] P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar, "Image texture feature extraction using GLCM approach," *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, pp. 1–5, 2013.

[59] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5962–5979, Oct. 2022.

[60] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6738–6746.

[61] W. Hao, W. Yitong, Z. Zheng, J. Xing, G. Dihong, C. Z. Jing, F. L. Zhi, and L. Wei, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5265–5274.

[62] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[63] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. ECCV*, vol. 14, 2014, pp. 740–755.

[64] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. H. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F. F. Li, "ImageNet large scale visual recognition challenge," in *Proc. Eur. Conf. Comput. Vis.*, vol. 115, 2015, pp. 211–252.

[65] *LabelImg*. [Online]. Available: https://www.v7labs.com/blog/labelimg-guide

[66] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. NIPS Workshop*, 2017, pp. 1–4.

[67] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*.

[68] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[69] M. O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2015, pp. 1–12.

YUAN LIN (Member, IEEE) received the bachelor's and master's degrees in information science and electronic engineering from Zhejiang University, China, in 2000 and 2003, respectively, and the Ph.D. degree from the Center for Quantifiable Quality of Service, Norwegian University of Science and Technology, Trondheim, Norway, in 2009.

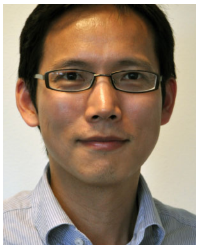
From 2009 to 2021, she was a System Developer with Nera Networks, DNB Bank ASA, and Sparebanken Vest. She is currently an Associate Professor with the School of Economics, Innovation, and Technology, Kristiania University College, Norway. Her current research interests include AI in health informatics, marine technology, computer vision, and quality assessment.

SHAOMIN XIE (Graduate Student Member, IEEE) is currently pursuing the master's degree with the School of Information Science and Engineering, Xiamen University, China. His research interests include computational intelligence and pattern recognition.

DEBASISH GHOSE (Senior Member, IEEE) received the Ph.D. degree in information and communication technology from the University of Agder, Grimstad, Norway, in 2019.

From 2020 to 2021, he was a System Developer with Confirmit, Grimstad, where he honed his practical skills in system development. From 2021 to 2022, he was a Postdoctoral Fellow with the University of Agder. He is currently an Associate Professor with the School of Economics, Innovation, and Technology, Kristiania University College, Bergen, Norway. His academic endeavors focus on protocol design, modeling, and performance evaluation within the Internet of Things (IoT) domain. His research interests include realms of data analytics, cybersecurity, and machine learning.

XIANGRONG LIU is currently a Professor with the Department of Computer Science and Technology, School of Informatics, Xiamen University. His main research interests include computational intelligence, computational theory, data mining, biological information processing, mobile, and micro-sensing technology.



JUNYONG YOU (Senior Member, IEEE) received the bachelor's and master's degrees in computational mathematics and the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, in 1998, 2001, and 2006, respectively. He was with Tampere University, Nokia Research Center, and Norwegian University of Science and Technology. He is currently a Senior Researcher with the NORCE Norwegian Research Center. He is the author or coauthor of more than 70 scientific articles. His research interests include AI, machine learning, computer vision, and their applications in different industries, such as energy, transportation, and aquaculture.

JARI KORHONEN (Member, IEEE) received the M.Sc. (Eng.) degree in information engineering from the University of Oulu, Finland, in 2001, and the Ph.D. degree in telecommunications from Tampere University of Technology, Finland, in 2006. Since 2022, he has been a Senior Lecturer with the University of Aberdeen, U.K. His research experience covers both telecommunications and signal processing aspects in multimedia communications. His current research interests include quality of experience (QoE) and deep learning for image and video quality assessment.

JUAN LIU received the Ph.D. degree in electronics science and technology from Huazhong University of Science and Technology, China. She is currently an Associate Professor with the School of Aerospace Engineering, Xiamen University, China. She has published more than 20 papers in journals and conferences. Her research interests include data mining and embedded systems.

SOUMYA P. DASH (Member, IEEE) received the B.Tech. degree (Hons.) in electrical engineering from Indian Institute of Technology, Bhubaneswar, in 2014, and the Ph.D. degree in electrical engineering from Indian Institute of Technology, Delhi, in 2019.

From January 2019 to March 2019, he was an Early-Doctoral Research Fellow with the Department of Electrical Engineering, Indian Institute of Technology, Delhi. Since March 2019, he has been with the School of Electrical Sciences, Indian Institute of Technology, Bhubaneswar, where he is currently an Assistant Professor. His research interests include communication theory for hybrid communication systems, power line communications, smart grid communications, next-generation wireless communication systems, reconfigurable intelligent surfaces, quantum communications, visible light communications, and diversity combining. He is a member of the IEEE Communications Society and the IEEE Vehicular Technology Society. He is also a Young Associate of Indian National Academy of Engineering (INAE). He was a recipient of Odisha Young Scientist Award, the VDGOOD Young Scientist Award, and the President of India Gold Medal for the academic year 2013–2014 for obtaining the highest CGPA amongst the students graduating with B.Tech. degree.