KZYJC
ISSN:1001-0920

ISSN: 1001-0920
Volume xx, Issue xx, January, 2023

Revolutionizing Face Recognition: An Improved MobileNetV2 System

Chi Jing[1], Zhang Haopeng[2], and Chin Kim On[3], Chai Soo See[4]

Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kinabalu, Sabah 88400, Malaysia [1,3.]
Hebei University of Engineering, Handan, Hebei 056038, China [2.]
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak, Sarawak 94300, Malaysia[4.]

Abstract— In this paper, we present an impressive face recognition model, which represents a robust improvement over the original MobileNetv2. Our model introduces the Receptive Field Block (RFB) to prevent any loss of facial feature information, expands the perceptual field, and implementing multi-scale feature fusion to enhance the model's feature extraction capability. Moreover, we have incorporated Coordinate Attention (CA) into the RFB to enhance recognition accuracy within the lightweight network. The proposed model is named CA_RFB_MobileNetv2. Our experimental results from eight public datasets demonstrate that the recognition accuracy rate of the proposed CA_RFB_MobileNetv2 model is either greater than or equal to that of MobileNetv2. In one of the eight datasets, the recognition accuracy of CA_RFB_MobileNetv2 was slightly reduced by 0.18% compared to FaceNet. However, it offers a significant advantage, a 2.3 times reduction in processing time per image and an 8.8 times decrease in the number of parameters used. Finally, our proposed model was used in a face recognition system, achieving an impressive accuracy of 97.5% with a low false acceptance rate of 2% when tested on 200 randomly selected face images from the Labeled Faces in the Wild dataset.

Keywords— face recognition, lightweight convolutional neural network, mobilenetv2, attention mechanism, multi-scale receptive field.

1. Introduction

Face recognition is a crucial biometric-based authentication technology widely used in various fields such as military, finance, public security, and everyday life [1]. The non-contact nature of face recognition, coupled with high stability, accuracy, and difficulty of replication, has made it an essential tool in the context of current epidemic prevention and control, which aims to reduce the risk of contact transmission. In recent years, significant progress has been made in face recognition research with Convolutional Neural Networks (CNNs) [2]. To improve the accuracy of face recognition, deep learning models have been developed with an increased number of operations and layers. This has resulted in higher hardware requirements. Therefore, research focusing on lowering the hardware requirements while enhancing recognition accuracy is critical.

The emergence of lightweight CNN has successfully addressed the impact of hardware requirements on face recognition in practical applications. Although increasing the number of network layers can significantly enhance the recognition accuracy of network models, it also increases the number of model parameters, which is impractical for many applications. To address this issue, researchers have developed lightweight CNN such as SENet [3], MobileNet [4], ShuffleNet [5], and GhostNet [6]. These lightweight models have significantly reduced the number of parameters and computational costs with only a slight loss of accuracy [7]. For instance, MobileNet, when compared to VGG16, it achieved similar accuracy performance in ImageNet classification but used only 1/32 of the parameters and 1/27 of the computational cost [8].

Feature fusion is another method that can enhance recognition accuracy by effectively addressing the problem of partial feature loss and incomplete semantic information in deep networks [9]. Combining the output of different scales of convolutional kernels can enhance the diversity of semantic information, and merging

global features with local features can increase the output of multi-scale feature maps, further improving recognition accuracy.

The attention mechanism is a resource allocation mechanism that allows the network model to selectively amplify valuable feature channels and suppress useless ones during the training process, enhancing the model's learning and understanding of the key information in the feature map [10]. However, in deep CNN, important face features can be lost due to extensive convolution and pooling operations, leading to a reduced generalization ability of the network model [11]. Adding an attention module to the network model can effectively improve the recognition accuracy of the network model [12].

In this paper, we propose the Coordinate Attention (CA) into the Receptive Field Block (RFB) MobileNetv2, namely CA_RFB_MobileNetv2 face recognition model, aiming to reduce computational time and produce higher recognition accuracy than MobileNetv2. The model uses MobileNetv2 as the base network model and introduces a Multi-scale RFB to enhance the feature extraction capability of the lightweight network. The CA attention mechanism is utilized to autonomously enhance and suppress the features in the Multi-scale RFB to improve the algorithm's robustness. The experiments demonstrate that the proposed CA_RFB_MobileNetv2 model achieves both real-time recognition computational time reduction and higher recognition accuracy than MobileNetv2.

The remainder of this article is structured as follows: Section 2 presents the theory related to face recognition systems, Section 3 discusses the CA_RFB_MobileNetv2 model, Section 4 presents and analyzes the experimental results, and Section 5 builds and tests a face recognition system on Labeled Faces in the Wild (LFW). The last section concludes and discusses future works.

2. Related Works

The face recognition system is comprised of three main components: face detection, face alignment, and face recognition. Face detection accurately locates the position of the face, while face alignment involves calibrating key facial features such as the eyes, nose tip, mouth corner points, eyebrows, and contour points of face parts. Face recognition involves several steps, including face image pre-processing, feature extraction, and face matching. First, the detected face is normalized and geometrically corrected to reduce noise interference. Then, the processed image undergoes feature extraction, and the feature data of the extracted face image is matched with the feature template stored in the database. Finally, the output displays the result with the highest matching degree. In the following subsections, we will discuss some of the existing work on face recognition systems.

2.1 Retina Face

In 2019, [13] proposed RetinaFace, a one-stage face detection algorithm that uses multi-task joint additional supervised learning and self-supervised learning for pixel-level localization of faces at different scales [14]. As illustrated in Fig. 1, RetinaFace's feature extraction network utilizes a feature pyramid comprising of five levels ranging from P2 to P6. P2 to P5 are obtained by horizontally stitching the residual connected output feature maps from C2 to C6, while P6 is obtained by performing a 3×3 convolution of C5 using a stride of 2. RetinaFace also incorporates five independent contextual modules, corresponding to the five levels of the feature pyramid, to expand the perceptual field of the feature map and enhance the semantic segmentation of the context.