Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/54/20241137

Federated learning-based YOLOv8 for face detection

Ruijia Peng

Department of Computer Science, University of Birmingham, Birmingham, B15 2TT, United Kingdom

rxp303@student.bham.ac.uk

Abstract. Recognizing the paramount importance of face detection in the realm of computer vision, there is an urgent need to address the vital concern of protecting individuals' privacy. Face detection inherently involves the handling of extremely sensitive personal information. To tackle this challenge, this study puts forth a proposal to incorporate Federated Learning into the face detection model. The objective is to maintain data localization and enhance security throughout the experiments by harnessing the decentralized nature of collaborative learning. The experimental procedure for Federated learning in face recognition models encompasses several key steps: device selection, global model initialization, model distribution to devices, local training, local model updates, model aggregation, global model updates, and multiple iterations. This methodology enables the collective training of models by dispersed devices, hence enhancing recognition performance, all the while ensuring the preservation of user data privacy. In addition, it is imperative to integrate Federated learning with YOLOv8 in order to establish a distributed target detection system. This method entails numerous devices engaging in local YOLOv8 model training, hence safeguarding data privacy and minimising data transmission. The empirical findings indicate that the use of joint learning in the face detection model leads to successful identification of the face model. In the future, there will be a consideration of novel federated learning algorithms with the aim of enhancing privacy.

Keywords: Federated Learning, Face Detection, Yolov8.

1. Introduction

The utilization of artificial intelligence technology has become increasingly prevalent in various domains, encompassing functions such as facial recognition and face detection. The detection of faces is a fundamental objective within the realm of computer vision, encompassing the automated identification and spatial localization of facial features within an image or video [1]. In contrast to face recognition, face detection is primarily concerned with ascertaining the existence of a face inside an image and identifying its location if one is present. The identification of faces serves as an essential initial stage for numerous artificial intelligence and computer vision applications. These applications encompass face recognition, analysis of facial expressions, estimation of gender and age, crowd counting, and security monitoring.

In order to develop a face identification model with superior performance, the acquisition of a substantial volume of data is crucial. However, the face serves as the fundamental basis of an individual's identity. This particular piece of information pertaining to an individual is highly sensitive. In contrast to the utilisation of accounts and passwords on the Internet. Modifying facial features is a complex

DOI: 10.54254/2755-2721/54/20241137

endeavour [2]. Therefore, the collection of facial data inherently raises concerns regarding personal privacy. Obtaining real-world facial photos necessitates obtaining consent from the individuals involved. In the conventional approach to deep learning, it is of great necessity for the data to be centralised on the server side prior to the training process. The aforementioned scenario presents a potential concern regarding the potential for extensive disclosure of face data.

Due to this rationale, the integration of federated learning can be considered. Federated learning is an innovative machine learning methodology characterised by its decentralised nature, enabling the training of models across a multitude of devices or servers. This technique ensures that data remains localised and secure throughout the learning process. In the context of conventional machine learning, the customary practise involves the collection of data, which is subsequently transmitted to a centralised server for the purpose of model training. However, federated learning presents an alternative approach wherein models are trained directly on the device that generates the data, hence eliminating the need to transmit the data to a centralised server. Federated learning enables the training of models utilising private data while mitigating concerns over privacy, as the data is exclusively saved on individual devices. The integration of federated learning with additional tasks holds significant promise within the contemporary field of artificial intelligence. By incorporating federated learning into various applications such as face recognition, natural language processing, medical data analytics, image segmentation, autonomous driving, and the Internet of Things, it is possible to attain objectives such as safeguarding privacy, distributing data, providing instantaneous responses, and developing personalized models. Previous studies have demonstrated that federated learning can engage in collaborative efforts inside decentralized settings, such as smartphones, edge devices, and sensors [3, 4]. This collaborative approach has been found to enhance accuracy, safeguard data privacy, and promote the development of socially responsible artificial intelligence systems.

The experimental procedure for incorporating federated learning into face recognition models comprises several steps: device selection, global model initialization, model distribution to devices, local model training, local model updating, model aggregation, global model updating, and multiple iterations of training. This procedure enables the training of models on dispersed devices while preserving local privacy. The model parameters are then aggregated on a central server, leading to iterative enhancements in the performance and resilience of face recognition models. Throughout this process, utmost attention is given to safeguarding user data confidentiality and privacy.

2. Method

2.1. Dataset preparation

The dataset utilised in this study is the Human Faces (Object Detection) dataset obtained from Kaggle [5]. A dataset consisting of 2, 204 human face photographs has been provided, each of which is accompanied by a labelled bounding box. The selection of the micro version of YOLOv8 was made based on constraints related to hardware resources and data volume. The decision was made to utilise a pre-trained YOLOv8 model for the analysis of the collected dataset. More detailed information about the dataset is shown in Figure 1.

DOI: 10.54254/2755-2721/54/20241137

image_name Image file name	# width Image width	# height Image height	# x0 X0	# y0 Y0
2204 unique values	150 8192	115 6680	1 5212	1
00001722.jpg	1333	2000	490	320
00001044.jpg	2000	1333	791	119
00001050.jpg	667	1000	304	155
00001736.jpg	626	417	147	14
00003121.jpg	626	418	462	60
00003121.jpg	626	418	316	157
00003121.jpg	626	418	35	71
00003121.jpg	626	418	254	94
00003121.jpg	626	418	166	118
00000400.jpg	600	400	148	61

Figure 1. The detailed information of the collected dataset [5].

2.2. YOLOv8

The chosen baseline model for this study is the YOLOv8 model. YOLOv8 represents a significant advancement over its predecessor, YOLOv5, which was publicly released by ultralytics on January 10, 2023 [6]. This updated version now encompasses a broader range of functionalities, including image categorization, object detection, and instance segmentation activities [7]. The YOLO algorithm was first introduced by Joseph Redmon and his colleagues in 2016.In contrast to alternative object detection algorithms that utilise sliding window technique or region proposals, YOLO adopts a distinctive methodology by processing the entire image in a single forward pass over a deep convolutional neural network backbone. The YOLO algorithm divides an input image into an S × S grid [8]. Within this grid-based framework, each grid cell assumes the responsibility of making predictions for a predefined group of bounding boxes. Additionally, it computes probabilities for various object classes and confidence scores, which convey the likelihood of an object existing within its corresponding bounding box.

Non-maximum suppression (NMS) is a significant approach that was introduced in the YOLO framework.In the context of object detection, it is possible for the model to produce several bounding boxes corresponding to a single object. The Neural Matching System (NMS) is utilised to detect and eliminate superfluous or inaccurate bounding boxes, resulting in the generation of a solitary bounding box for each object present within the image [9].

DOI: 10.54254/2755-2721/54/20241137

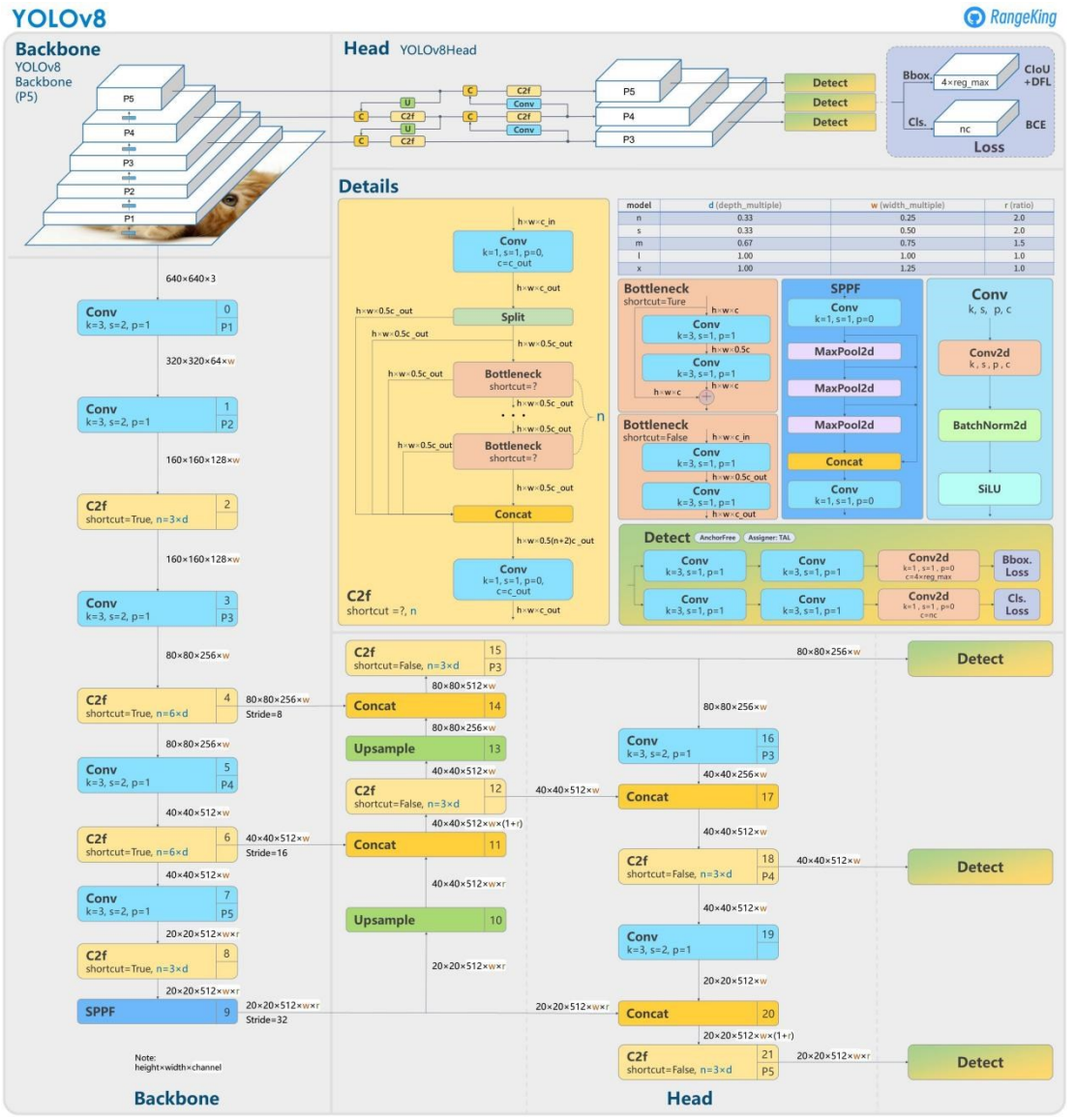

Figure 2. Model Architecture Design for yolov8 [6].

2.3. Federated Learning +YOLOv8

The subsequent phase in integrating federated learning with YOLOv8 shown in Figure 2 involves establishing a distributed target detection system wherein numerous devices or nodes cooperate to train a shared YOLOv8 model. This approach ensures that data is processed locally, eliminating the need for centralization in a single location. The primary objectives of this approach are to enhance privacy protection and minimize data transfer. Initially, it is vital to undertake the necessary steps to configure the YOLOv8 model and thereafter opt for a federated learning framework. It is imperative to divide the data among numerous local devices and incorporate mechanisms to safeguard privacy. The configuration of model update and communication protocols inside the federated learning framework involves the establishment of safe mechanisms to facilitate the sharing of model weights and the coordination of training rounds on local devices. This process enables the aggregation of weight updates, which in turn allows for the updating of the global model. In conclusion, the deployment of YOLOv8 models trained with federated learning is executed to achieve real-time inference for target detection

DOI: 10.54254/2755-2721/54/20241137

tasks. It is crucial to evaluate the performance of the deployed models and carry out essential fine-tuning and optimization procedures.

3. Results and discussion

3.1. Results

In the conducted trials, FL is employed as the methodology for training the YOLOv8 model. Given the limited availability of extensive datasets in real-world Internet of Things (IoT) devices, the dataset utilised in this experiment is characterised by its small size yet specificity. The models employed in this study are trained using pre-existing yolov8n models. In this experiment, there are four clients engaged in training their separate local models. Consequently, the entire dataset is partitioned into four distinct segments. Within the server, a function exists that facilitates the updating of the global model. This updating process is contingent upon the reception of the local model from the clients, which adheres to the Federated Averaging methodology. The objective of this experiment is to integrate the Federated Averaging algorithm into the face recognition model and evaluate the efficacy of the resulting global model. To prevent premature convergence of the model and to effectively showcase the experimental outcomes, a significantly low learning rate of 0.0000001 is employed. The local step size is configured as 1, whereas the batch size is configured as 32. The outcomes of executing the model on the photos within the validation dataset are illustrated in Figure 3. Additionally, the presented Table 1 provides an overview of the global model's accuracy at various stages of training. According to the findings presented in Figure 4, it can be observed that the face model is successfully detected by the model after undergoing a training process consisting of 30 rounds.

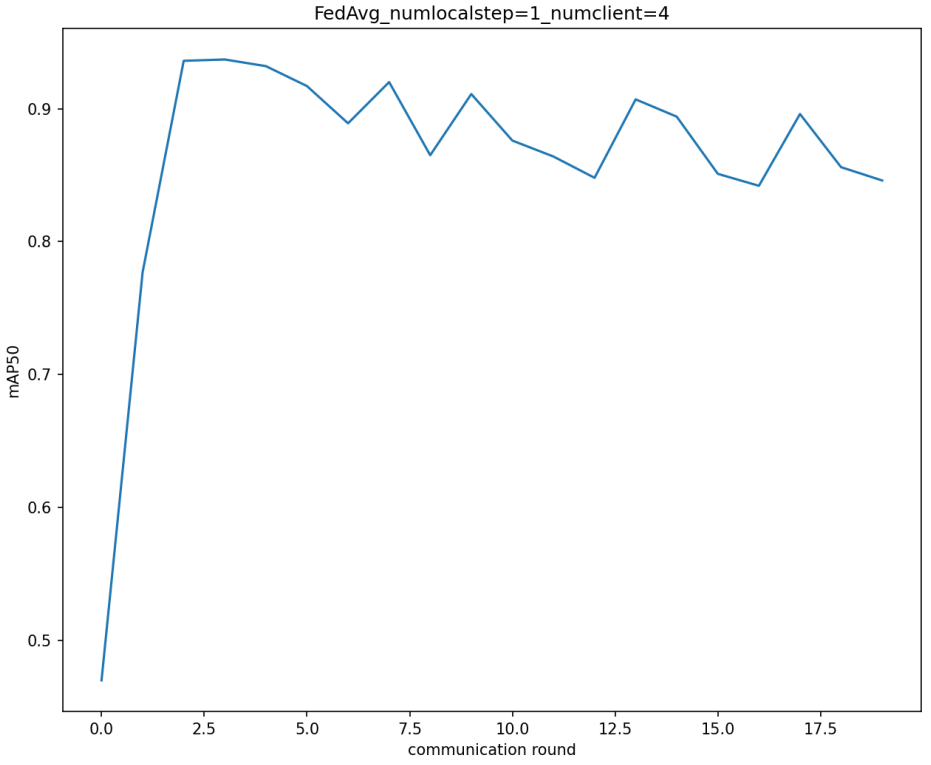

Figure 3. The outcomes of executing the model on the photos within the validation dataset (Photo/Picture credit: Original).

DOI: 10.54254/2755-2721/54/20241137

Table 1. The global model's accuracy at various stages of training.

Round	Method	mAP50
5	FedAvg	0.615
10	FedAvg	0.83
30	FedAvg	0.849

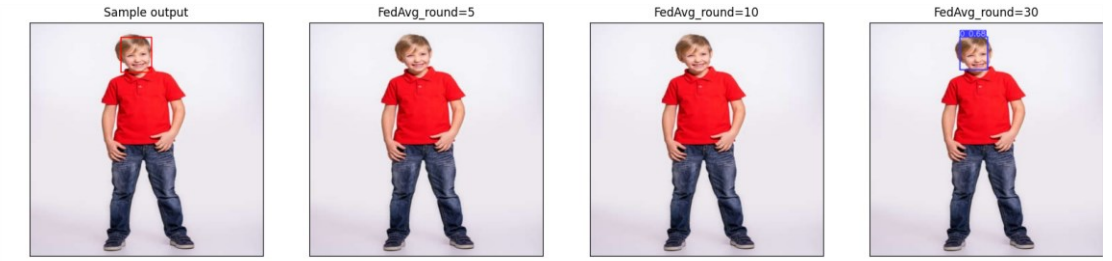

Figure 4. Model training results based on the sample image (Photo/Picture credit : Original).

3.2. Discussion

Several limitations were taken into account throughout the implementation of this technique. The application of FL to object classification tasks follows a similar configuration, wherein a shallow network serves as the underlying framework. The final classification is achieved by utilising a shared fully connected layer, which is accessible to all clients. However, this method can give rise to potential privacy issues as it opens the door to information leakage. This issue stems from the retrieval of a client's confidential class embedding, which allows other clients to efficiently create a highly accurate facial image of that client by optimizing random noise [10]. Furthermore, current FL methodologies predominantly employ shallow networks on the client side, hence increasing the susceptibility to network failures when applied to more complex and deeper network architectures. Hence, it is imperative to reevaluate the FL algorithm in order to guarantee the provision of an individual private fully-connected layer to each client. This not only enhances privacy but also facilitates the convergence of the neural network. In the future, other effective modules e.g. the advanced attention mechanism that are considered in many studies [11, 12] may be also employed for further improving the performance of the developed method.

4. Conclusion

The aim of this research is to incorporate joint learning into face detection models using the yolov8 framework training dataset. In order to evaluate the effectiveness of the proposed method, the project conducted a series of comprehensive experiments. These experiments were designed to train the accuracy of the model under various hyper-parameter configurations. After 30 iterations of training, the proposed model showed the ability to correctly detect face models. However, the experiments show that the use of joint learning in item classification tasks may lead to privacy leakage due to the non-IIID issue. Therefore, the FL algorithm will be reevaluated in subsequent studies to ensure that privacy measures are enhanced.

References

[1] Kumari Sirivarshitha A Sravani K Priya K S and Bhavani V 2023 An approach for Face Detection and Face Recognition using OpenCV and Face Recognition Libraries in Python (India: Coimbatore) pp1274-1278.

[2] Kim J Park T Kim H and Kim S 2021 Federated Learning for Face Recognition (USA:NV) pp. 1-2.

DOI: 10.54254/2755-2721/54/20241137

[3] Ren J Zhou J Lyu Y and Liu L 2022 Accelerated Federated Learning with Decoupled Adaptive Optimization.
[4] Lai W and Yan Q 2022 Federated Learning for Detecting COVID-19 in Chest CT Images: A Lightweight Federated Learning Approach (China:Qingdao) pp. 146-149.
[5] Kaggle 2023 Human faces object detection https://www.kaggle.com/datasets/sbaghbidi/human-faces-object-detection.
[6] Zhou Y Zhu W He Y Li Y 2023 YOLOv8-based Spatial Target Part Recognition (China : Chongqing2023) pp. 1684-1687.
[7] Li Q Ge Z Wang F Luo X Yang Y Yu T 2022 Small Target Repair Parts Detection Algorithm Based on Improved YOLOv5 (China :Guilin) pp 420-433.
[8] Mahendru M Dubey S K 2021 Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3(India: Noida) pp. 734-740,
[9] Liu Y Huang A and Luo Y 2020 Fedvision: An online visual object detection platform powered by federated learning.
[10] Niu Y and Deng W 2022 Federated Learning for Face Recognition with Gradient Correction pp. 1999-07.
[11] Qiu Y Wang J Jin Z Chen H Zhang M and Guo L 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training Biomedical Signal Processing and Control 72 103323.
[12] Shao H Li W Cai B Wan J Xiao Y and Yan S 2023 Dual-threshold attention-guided GAN and limited infrared thermal images for rotating machinery fault diagnosis under speed fluctuation IEEE Transactions on Industrial Informatics.