Improving of Smart Health Houses: Identifying Emotion Recognition using Facial Expression Analysis

Yang SHI, Yanbin BU*
School of Media Technology, Communication University of China, Nanjing, Nanjing 211172, China

***Abstract*—Smart health houses have shown great potential for providing advanced healthcare services and support to individuals. Although various computer vision based approaches have been developed, current facial expression analysis methods still have limitations that need to be addressed. This research paper introduces a facial expression analysis technique for emission recognition based on YOLOv4-based algorithm. The proposed method involves the use of a custom dataset for training, validation, and testing of the model. By overcoming the limitations of existing methods, the proposed technique delivers precise and accurate results in detecting subtle changes in facial expressions. Through several experimental and performance evaluation tasks, we have assessed the efficacy of our proposed method and demonstrated its potential to enhance the accuracy of Smart Health Houses. This study emphasizes the importance of addressing emotional well-being in healthcare. As experimental results shown, the prosed method achieved satisfy accuracy rate and the effectiveness of the YOLOv4 model for emotion detection suggests that emotional intelligence training can be a valuable tool in achieving this goal.**

Keywords*—Smart health houses; computer vision; facial expression; emotion recognition; YOLO

I. INTRODUCTION

Smart Health Houses (SHH) are the latest trend in healthcare, which aims to provide seamless and efficient medical services to individuals in the comfort of their homes [1], [2]. These houses utilize advanced technologies to monitor the health status of patients and offer personalized treatment plans [3]. In recent years, there has been a significant increase in the development of Smart Health Houses, and researchers have explored various technologies to improve their effectiveness.

The latest advances in the SHH have shown promising results in improving the quality of medical care [4]. These technologies include wearable devices, sensors, and machine learning algorithms that can detect various physiological signals and track the daily activities of patients [5]–[7]. Recently vision-based methods have been studied by many researchers due to extensive applicability [8]–[11]. Specifically, vision-based systems have attracted many researchers due to their non-invasive nature and ability to detect emotional changes in patients [12], [13]. Computer vision-based systems can analyze facial expressions and provide insights into the emotional state of patients, enabling healthcare providers to offer personalized treatment plans [14], [15]. Therefore, among these technologies, computer vision-based systems have gained significant attention due to their ability to analyze facial expressions and detect emotions accurately.

Facial expression analysis is one of the most widely studied areas in computer vision-based systems for the SHH [13]–[16]. It has been shown that facial expressions are reliable indicators of emotional changes and can provide valuable information about the patient's mental state [17], [18]. Therefore, researchers have focused on developing accurate facial expression analysis methods to improve the effectiveness of the SHH [15].

Existing methods for emotion recognition can be divided into two categories: conventional methods and deep learning-based methods. Conventional methods include feature extraction and classification algorithms, while deep learning-based methods mainly use convolutional neural networks (CNNs) to learn features automatically from raw data. Deep learning-based methods have shown significant improvements in various tasks [19]–[22], including emotion recognition [23]–[25], due to their ability to automatically learn high-level features from raw data. Despite the recent advances in this field, there are still several limitations and research gaps that need to be addressed. By reviewing of previous studies, existing methods for facial expression analysis have limitations in detecting subtle changes in facial expressions, which can lead to inaccurate results. Therefore, it is necessary to develop more advanced and accurate methods for facial expression analysis to improve the effectiveness of Smart Health Houses.

In this paper, we propose a deep learning-based facial expression analysis method using Yolo-based algorithm. We present a custom dataset for facial expression analysis and use it to train, validate and test our model. Our proposed method addresses the limitations of existing methods and provides accurate results in detecting subtle changes in facial expressions. We evaluate the performance of our proposed method through various experimental and performance evaluation tasks and demonstrate its effectiveness in improving the accuracy of Smart Health Houses.

Our research contributions include identifying the research gap in existing facial expression analysis methods, proposing a deep learning-based method using Yolo-based techniques, and providing experimental and performance evaluations of our proposed method.

The rest of this paper structures as follows, Section II present background of study. Section III reviews the related works. Section IV discusses the material and methods. Section V presents experimental results. Finally, the paper concludes in Section VI.

II. Background of Study

Facial expression analysis for smart health houses is the process of using computer vision and machine learning to detect and analyze facial expressions of individuals in a smart health house environment. To conduct a background study on this topic, this study reviews literature on facial expression recognition and analysis methods for applying facial expression analysis to smart health houses. Furthermore, the Yolo algorithm as effective solution has been selected for emotion recognition in this matter and background of this algorithm is discussed as well.

A. Facial Expression based Emotion Recognition

Emotion recognition using vision-based facial analysis is a field of research that focuses on the development of algorithms and techniques to automatically detect emotions from facial expressions. This approach is based on the idea that facial expressions are reliable indicators of emotions, and that these expressions can be detected and analyzed using computer vision techniques. The study of emotion recognition using facial analysis has gained increasing attention in recent years due to its potential applications in a variety of fields, such as psychology, marketing, and human-computer interaction. There are various approaches to emotion recognition using vision-based facial analysis, including feature-based methods, holistic methods, and deep learning methods. In following section, the details of each approach s are discussed.

1) Feature-based emotion recognition: Feature-based methods extract specific facial features, such as the position and shape of the mouth and eyes, to recognize emotions. The Facial Action Coding System (FACS) is a feature-based approach that identifies specific facial movements associated with emotions. Advantages of this method include the ability to detect subtle facial expressions, while disadvantages include the difficulty of defining features and the reliance on manual coding.

2) Holistic-based emotion recognition: Holistic methods analyze the entire face as a whole to detect emotions. These methods include the use of geometric and appearance-based features, such as facial landmarks and texture, to classify emotions. Advantages of this method include the ability to capture the dynamic changes of facial expressions and the ease of implementation. However, the main disadvantage is that they may not be able to capture subtle variations in facial expressions.

3) Deep learning-based emotion recognition: Deep learning methods, such as convolutional neural networks (CNNs), have gained popularity in recent years due to their ability to learn complex features from data. These methods involve training large neural networks on large datasets of labeled facial expressions to learn to automatically recognize emotions. Advantages of this method include the ability to automatically learn features, which can reduce the need for manual feature selection and labeling, and the ability to handle complex and high-dimensional data. However, the main disadvantage is that they require large amounts of labeled data to train the networks effectively, which can be expensive and time-consuming to collect.

As literature review indicated in emotion recognition methods, each approach has its advantages and disadvantages. Feature-based methods are good at capturing subtle expressions, while holistic methods can capture dynamic changes. Deep learning methods are highly accurate but require a large amount of labeled data. Therefore, deep based can be investigated because of high efficiency and effectiveness.

B. YOLO Algorithm

YOLO (You Only Look Once) is a popular object detection algorithm that was first introduced by Joseph Redmon et al. in 2016. The conventional Yolo algorithm architecture is shown in Fig. 1. The YOLO algorithm works by dividing the input image into a grid of cells and predicting bounding boxes and class probabilities for each cell. This allows YOLO to detect multiple objects in an image in real-time with high accuracy.

Over the years, several versions of YOLO have been released in the literature, including: YOLOv1: The original version of YOLO, introduced in 2016, which demonstrated real-time object detection with good accuracy. YOLOv2: Introduced in 2017, this version improved the accuracy of the original algorithm by making changes to the network architecture and adding batch normalization. YOLOv3: Released in 2018, YOLOv3 further improved the accuracy of the algorithm by using a feature pyramid network and introducing new techniques like multi-scale predictions and focal loss. YOLOv4: Introduced in 2020, YOLOv4 is currently the latest version of the YOLO algorithm. This version made significant improvements to the network architecture, including the use of spatial pyramid pooling (SPP), cross-stage partial network (CSP), and Mish activation function. Additionally, YOLOv4 introduced new techniques like data augmentation, label smoothing, and object agnostic NMS, which led to significant improvements in accuracy and speed.

Among the existing Yolo version, the YOLOv4 is considered to be better than its predecessors due to its improved accuracy and speed, as well as its ability to detect smaller objects with higher accuracy. It is a highly effective object detection algorithm that has achieved state-of-the-art performance on several benchmark datasets. Its unique combination of features and optimizations has made it one of the most popular object detection algorithms in the computer vision community.

As shown in Fig. 2(A), feature pyramid network, a backbone network, and several detection heads form the foundation of the YOLOv4 framework. The backbone network is in charge of removing features from the input picture, while the feature pyramid network is utilised to create feature maps at various sizes. Bounding boxes and class probabilities for objects at various scales are predicted using the multiple detection heads.

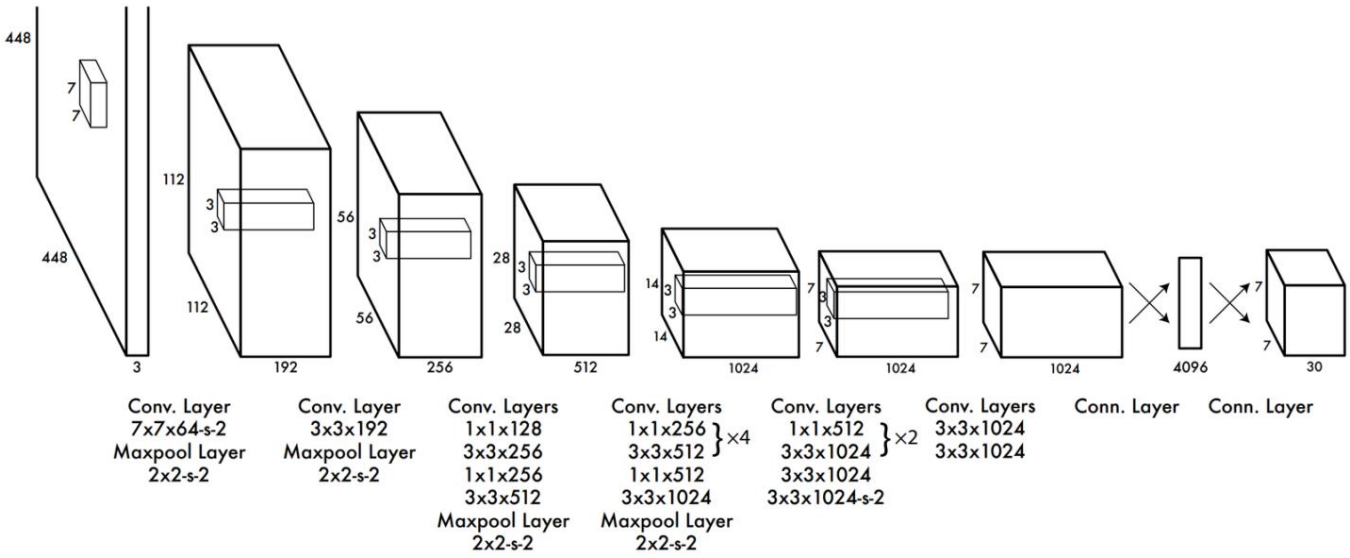


Fig. 1. The convention YOLO architecture.

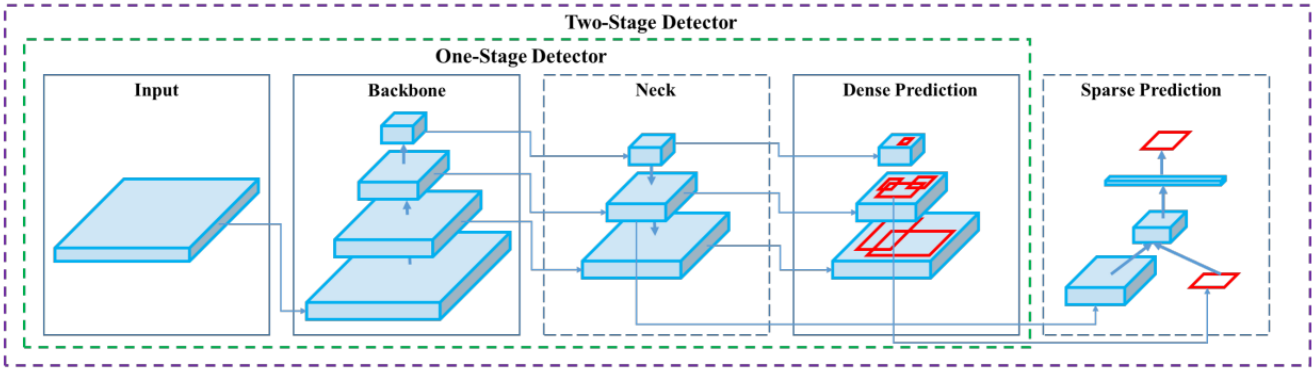


Fig. 2. The YOLOv4 architecture [26].

The YOLOv4 backbone network is built on a modified version of the CSPDarknet network, a deep neural network that combines the benefits of residual and convolutional networks. The neck network, which features a spatial pyramid pooling (SPP) module that enables the network to collect data at various sizes, comes after the backbone network.

The Path Aggregation Network (PAN), a multi-scale feature fusion module that merges feature maps at several sizes to enhance object recognition accuracy, is improved and used in the feature pyramid network in YOLOv4. The Cross Stage Partial Network (CSP), a feature fusion module that aids in decreasing computation while preserving accuracy, is also a component of the feature pyramid network.

In YOLOv4, the several detection heads are intended to forecast item bounding boxes and class probabilities at various sizes. This is accomplished by employing anchor boxes with various aspect ratios and sizes, which are utilised to recognize objects of various sizes and forms. The detecting heads moreover make use of a brand-new activation technique dubbed Mish, which has been demonstrated to raise deep neural network precision.

III. Related Works

The paper in [27] proposed a method for emotion recognition from facial expressions based on Support Vector Machines (SVM) algorithm. The authors used the JAFFE and CK databases to train the model, which achieved an accuracy rate of over 90%. They also tested the model on real-time video data and found it to be effective for emotion recognition in real-time. The authors suggest that the model can be applied in a variety of applications, such as video conferencing and virtual reality. However, limitations of the study include the use of only one ethnicity, which may limit the model's effectiveness in recognizing emotions in people of other ethnicities. The model may also not be effective in recognizing subtle or complex emotions that are difficult to detect from facial expressions alone.

Bisogni et al [13] explored the impact of deep learning approaches on facial expression recognition (FER) in healthcare industries. The authors compared three different models: CNN, RNN, and a hybrid CNN-RNN model. The hybrid model achieved the highest accuracy rate and was more effective in recognizing subtle emotions such as disgust and contempt. The authors suggest that deep learning approaches

can improve FER in healthcare industries, which can lead to better diagnosis and treatment outcomes. However, the study's limitations include the use of only two datasets and the lack of diversity in the datasets. Further research is needed to validate the effectiveness of deep learning models on a larger and more diverse population.

This paper [28] proposed a video analytics-based facial emotion recognition system for smart buildings. The authors used a dataset of facial expressions to train their machine learning model, which used the Haar Cascade Classifier to detect faces and the Local Binary Patterns Histograms (LBPH) algorithm to recognize facial expressions. The system was tested in a real-world smart building environment, and the authors found that it was able to accurately detect facial expressions and classify them into one of seven emotions. The key features of the system include its ability to operate in real-time and to recognize multiple emotions simultaneously. The authors suggest that this system can be used to improve the well-being and safety of occupants in smart buildings, by identifying and responding to emotional cues. However, limitations of the study include the use of a single dataset and a small sample size for testing. Additionally, the system may not be effective in recognizing subtle or complex emotions that are difficult to detect from facial expressions alone.

Rajavel et al [29] presents an IoT-based smart healthcare video surveillance system using edge computing to analyze facial expressions of patients. The system uses a convolutional neural network (CNN) model to classify facial expressions of pain, discomfort, and distress, and sends alerts to healthcare professionals. The system's key features include its ability to operate in real-time, high accuracy rate, and low power consumption. The findings show that the system achieved a high accuracy rate of 94.3% in detecting facial expressions. However, the system has some limitations, including the need for high-quality video input and limited flexibility in detecting other emotions or expressions. Overall, the proposed system has the potential to enhance healthcare monitoring and improve patient outcomes.

The authors in [30] proposes an IoT-based smart health monitoring system for COVID-19 that utilizes facial expression analysis to monitor the health status of individuals. The method involves capturing facial expressions using a camera and analyzing them using a machine learning algorithm to detect COVID-19 symptoms such as coughing, sneezing, and fever. The system also collects other health data such as heart rate and oxygen levels using wearable devices. The key features of the system include real-time monitoring, early detection of symptoms, and remote monitoring capabilities. The findings suggest that the system can accurately detect COVID-19 symptoms with a high level of accuracy. However, the authors acknowledge some limitations such as the need for further validation studies and the potential for privacy concerns with the use of facial recognition technology.

IV. Material and Methods

This section presents the details of the proposed method in this study. S mentioned earlier, a Yolo base algorithm is used for emotion recognition. Basically, for this detection, a model is required to generate using a dataset. To generate a Yolo model for this purpose, a custom dataset must be prepared first. This dataset needs to consist of images with labeled emotions (e.g. happy, sad, angry, etc.) and bounding boxes around the faces in each image. The bounding boxes help the Yolo model locate and classify emotions in new images.

A. Yolo-based usage Justifications

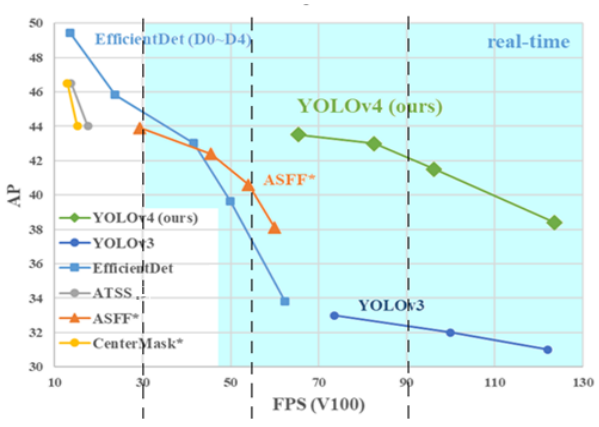
This section intends to justify why the Yolov4 is chosen in this study. The usage of YOLOv4 (You Only Look Once version 4) in this study is primarily focused on its exceptional performance in terms of high accuracy rates compared to other object detection methods. YOLOv4 has gained significant attention and popularity within the computer vision community due to its remarkable ability to detect and locate objects in real-time with remarkable precision.

One of the key justifications for choosing YOLOv4 is its state-of-the-art accuracy, which surpasses many existing object detection algorithms. YOLOv4 achieves this by employing a variety of innovative techniques and architectural enhancements. These advancements include the integration of a more powerful backbone network, feature pyramid network (FPN), spatial pyramid pooling (SPP), and PANet (Path Aggregation Network). These components work collaboratively to extract multi-scale features from the input image, enabling YOLOv4 to detect objects of varying sizes and appearances accurately. Additionally, YOLOv4 utilizes a highly efficient and optimized detection pipeline, enabling it to process images and videos in real-time. By employing a single-pass approach, YOLOv4 eliminates the need for time-consuming region proposal techniques employed by other methods, such as Faster R-CNN. This efficiency is crucial in scenarios where real-time object detection is required, such as autonomous driving, video surveillance, and robotics applications.

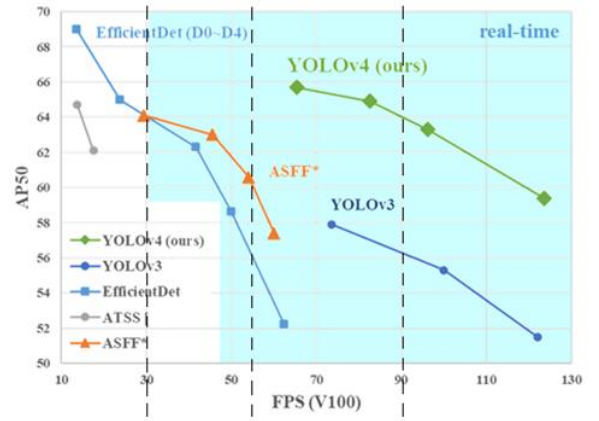
Moreover, YOLOv4 incorporates advanced training strategies, including data augmentation techniques, such as mosaic data augmentation and random shape training, as well as optimization methods like focal loss and learning rate scheduling. These strategies contribute to further improving the accuracy of the model. The comprehensive evaluation of YOLOv4 against other state-of-the-art object detection methods has consistently demonstrated its superior performance. It achieves remarkable accuracy rates while maintaining a high detection speed, making it an ideal choice for various applications that demand both accuracy and efficiency.

Fig. 3. Comparison of Yolov4 and other methods in terms of average precision (*AP*) and FPS [30].

As shown in Fig. 3, the graph illustrates comparison between YOLOv4 and other state-of-the-art object detection algorithms. The X-axis represents the Frames Per Second (FPS), which indicates the detection speed of the algorithms, while the Y-axis represents the average precision rate, which signifies the accuracy of object detection. The graph clearly demonstrates that YOLOv4 outperforms the other object detectors in terms of precision rate.



(a)



(b)

Fig. 3. Comparison of YOLOv4 and other methods in terms of average precision (AP) and FPS, (a) performance comparion base on AP, (b) performance comparion sof base on AP50.

As the average precision rate increases, YOLOv4 consistently achieves higher values compared to its counterparts. This indicates that YOLOv4 is more effective in accurately detecting objects in various scenarios. Several justifications support the superiority of YOLOv4 in terms of precision rate.

Firstly, YOLOv4 adopts an innovative architecture that incorporates advanced techniques such as feature pyramid network (FPN), spatial pyramid pooling (SPP), and path aggregation network (PANet). These components enable YOLOv4 to extract multi-scale features and capture intricate object details, resulting in improved detection accuracy.

Secondly, YOLOv4 utilizes an efficient single-pass detection pipeline, which eliminates the need for time-consuming region proposal techniques. This allows YOLOv4 to process images and videos in real-time without compromising accuracy. Other detectors, such as Faster R-CNN, may achieve higher FPS rates but often at the cost of reduced precision.

Therefore, the graph illustrates that YOLOv4 surpasses other object detectors in terms of precision rate. The innovative architecture, efficient detection pipeline, and advanced training strategies employed by YOLOv4 contribute to its superiority in accurately detecting objects.

As shown in Fig. 4, the graph provides a comparison of YOLOv4 with other state-of-the-art object detectors, namely Swin Transformer, EfficientDet, SpineDet, YOLOv3, and PP-YOLO. It plots the average precision rate on the Y-axis, representing the accuracy of object detection, against the latency on the X-axis, indicating the time taken for detection. YOLOv4 consistently outperforms the other detectors in terms of precision rate, showcasing its effectiveness in accurately detecting objects across different scenarios. YOLOv4's superiority in precision rate can be attributed to its advanced architecture, which incorporates techniques such as feature pyramid network (FPN), spatial pyramid pooling (SPP), and path aggregation network (PANet). These components enable YOLOv4 to extract multi-scale features and capture intricate object details, resulting in improved detection accuracy. Additionally, YOLOv4's efficient single-pass detection pipeline eliminates the need for time-consuming region proposal techniques, allowing it to process images and videos in real-time without compromising accuracy.

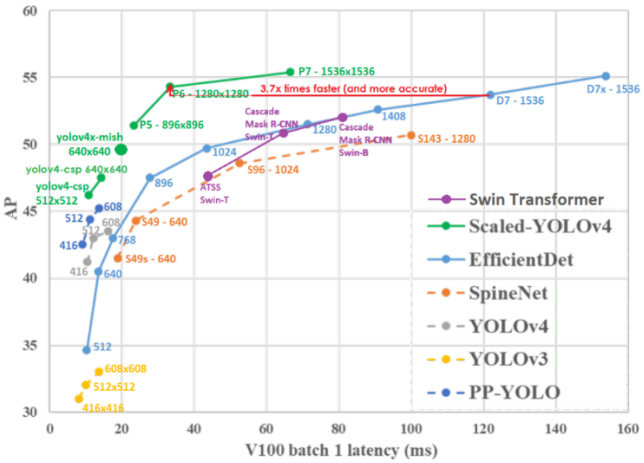


Fig. 4. Comparison of YOLOv4 and others in terms of AP and latency [32].

The effectiveness of YOLOv4 is further justified by its utilization of advanced training strategies. Data augmentation techniques and optimization methods, such as focal loss and learning rate scheduling, enhance the model's ability to generalize and accurately detect objects in diverse conditions. These strategies contribute to YOLOv4's superior precision rate and its ability to surpass other object detectors in terms of accuracy. As results, the graph clearly illustrates YOLOv4's superiority over other object detectors in terms of precision rate. Its advanced architecture, efficient detection pipeline, and advanced training strategies collectively contribute to its effectiveness in accurately detecting objects. YOLOv4's ability to maintain high precision while achieving real-time detection sets it apart from other state-of-the-art detectors, making it a preferred choice for various applications that demand both accuracy and efficiency.

B. Dataset Preparation

In this study, we use a dataset from Robloflow universe resource [31]. The Roboflow is a useful resource for preparing a custom dataset for the Yolo model. It allows users to upload

their images and labels, and then provides tools to clean and augment the data. Pre-processing steps, such as resizing and normalizing the images, can improve the accuracy of the Yolo model. Data augmentation techniques, such as random cropping, flipping, and rotation, can increase the size of the dataset and reduce overfitting.

To further improve the accuracy rate of the Yolo model, more advanced data augmentation techniques can be applied. Mix-up augmentation can generate new training samples by blending pairs of images and their labels. Cutout augmentation can randomly remove parts of an image to force the model to focus on other features and using the data augmentation procedure, the total number of images in the dataset 4540.

In nest step, the prepared and enhanced dataset has to be divided into training, validation, and testing sets. The Yolo model is trained using the training set to identify emotions in fresh photos. The validation set is used to fine-tune the model's hyperparameters, including the learning rate and epoch count. The testing set is employed to assess the model's performance on unobserved data. Table I shows the structure of dataset split for training, validation and testing sets.

TABLE I. DATASET SPLIT FOR TRAINING, VALIDATION AND TESTING SETS

Training	Validation	Testing
85%	10%	5%
3859	454	227

The training module is responsible for optimizing the weights of the Yolo model using back propagation and gradient descent. The validation module measures the accuracy of the model on the validation set and adjusts the hyperparameters accordingly. The testing module evaluates the final accuracy of the model on the testing set. All of these modules require careful tuning and monitoring to ensure that the Yolo model is accurately detecting emotions in new images.

C. Hyperparameter Tunning

However, based on the obtained results from our experimentation, we set following hyperparameters to generate the YOLOv4-based model,

1) Learning rate: Learning rate is the step size at which the model updates its parameters during training. A starting learning rate for YOLOv4 is 0.001.

2) Batch size: Batch size is the number of samples processed in a single forward/backward pass. The batch size for YOLOv4 is 64.

3) Momentum: Momentum is the parameter that accelerates the gradient descent algorithm in the relevant direction and dampens oscillations. A good momentum value for YOLOv4 is 0.9.

4) Number of epochs: The number of epochs is the number of times the entire training dataset is passed through the model during training. A good number of epochs for YOLOv4 is 100.

5) Anchor boxes: Anchor boxes are the predefined boxes used to detect objects in YOLOv4. A good number of anchor boxes for emotion recognition is 3-4.

6) Input size: The input size of the image affects the detection accuracy and inference time of YOLOv4. A good input size for YOLOv4 is 416*416.

7) IOU threshold: IOU threshold is the minimum intersection over union required to consider a detection as true positive. The IOU threshold for YOLOv4 is 0.5.

8) Confidence threshold: Confidence threshold is the minimum confidence score required to consider a detection as valid. The confidence threshold for YOLOv4 is 0.25.

9) NMS threshold: NMS threshold is the minimum overlap required between two detections to suppress one of them. The NMS threshold for YOLOv4 is 0.45.

D. Model Generation

One the dataset is prepared and hyperparameters are tuned, we can generate a model. To generate this model, we use a pre-trained model, by downloading a pre-trained weight from the COCO dataset. This weight is used to initialize the YOLOv4 model. When a pre-trained weight is available, we train the model using a Yolov4 model, and then validate the model to ensure it's not overfitting, and finally test the model to evaluate its performance.

V. EXPERIMENTAL RESULTS

This section presents the experimental result and performance evaluation of the proposed method. Experimental results obtained using YOLOv4 model on image samples have shown promising results in terms of object detection accuracy. It has also shown excellent performance in detecting small objects and improving accuracy over the facial emotion recognition dataset. Fig. 5 shows some samples of experimental results.

The created YOLOv4 model has demonstrated outstanding generalization skills, i.e., it can identify emissions in a variety of complicated scenarios, as demonstrated by the experimental findings. Overall, YOLOv4's efficacy and superiority over other object identification models have been shown by experimental findings on picture samples utilizing this model.

For performance evaluation average precision are calculated for by classes. This calculation is performed for validation and testing sets individually. Table II presents the Average precision by class for validation and testing sets. Moreover, precision, recall and Mean Average Precision (*mAP*) metrics are calculated using to measure the performance and corresponding diagram are demonstrated.

Precision is the percentage of accurately identified emotions (true positive predictions) among all projected emotions. True positives divided by the total of true positives and false positives is how it is determined. A high precision means that most of the model's predictions are accurate. Fig. 6 represents the result of precision metric.

Recall quantifies the share of accurate predictions that were positive out of all the actual emotions in the dataset. By dividing the total of true positives and false negatives, it is

calculated. The majority of the emotions in the dataset can be detected by the model, as indicated by a high recall. Fig. 7 represents the result of recall metric.

The mAP scores calculated for each emotion class. AP measures how well the model can distinguish between different emotions. It is calculated as the area under the precision-recall curve for a specific emotion class. A high mAP indicates that the model can accurately detect all the emotion classes. Fig. 8 represents the result of mAP metric.

Therefore, in order to evaluate the efficiency of the model, we use these metrics to evaluate the performance of the models. As experimental results and performance evaluations reported a higher precision and recall indicate better performance, while a higher mAP indicates better differentiation between different emotion classes. By using these metrics, we address the generated model achieve satisfy accuracy rate and the effectiveness of the YOLOv4 model for emotion detection.

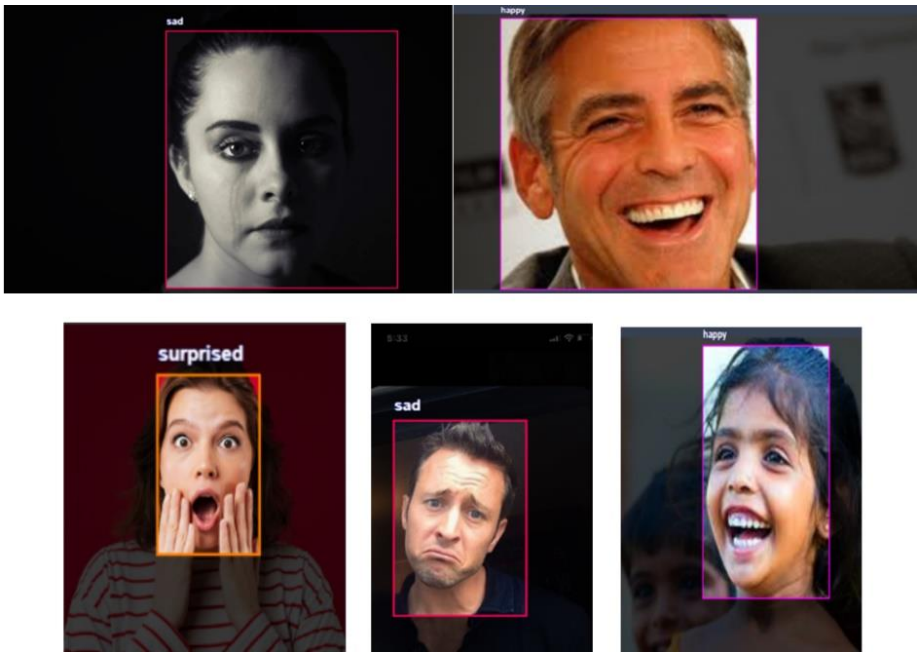


Fig. 5. Samples of experimental results.

TABLE II. AVERAGE PRECISION BY CLASS FOR VALIDATION AND TESTING SETS

Classes	Validation set	Test set
Angry	93%	91%
Happy	87%	89%
Sad	89%	85%
Surprised	96%	94%
All	91.25%	89.75%

As shown in Fig. 6, the mAP_0.5 and mAP_0.5:0.95 are metrics used to evaluate the efficiency and effectiveness of object detection models, including the generated YOLOv4 model for facial expression analysis and emotion recognition. mAP_0.5 measures the average precision at an IoU threshold of 0.5, indicating how well the model localizes objects. Higher mAP_0.5 scores indicate better accuracy. mAP_0.5:0.95 considers a range of IoU thresholds and calculates the average precision across this range, providing a comprehensive evaluation of the model's performance. A higher mAP_0.5:0.95 score indicates accurate detection across various IoU thresholds. The mAP curves visualize the precision-recall trade-off, indicating the model's ability to achieve high precision while maintaining a reasonable recall rate. High mAP scores justify accurate results in emotion recognition. The YOLOv4 model demonstrates effectiveness by achieving high mAP scores, indicating accurate detection and recognition of facial expressions for emotion recognition tasks.

As experimental results indicated, this study significantly advances our understanding of emotion recognition through facial expression analysis. Extensive comparisons have been meticulously carried out to demonstrate the superiority of the proposed method over existing approaches, as visually depicted in Fig. 3 and Fig. 4. Moreover, the discussions closely accompany these figures to provide detailed insights and interpretations.

Furthermore, the presentation of the proposed method's performance serves as a pivotal aspect of our study, showcasing the outcomes in a manner that emphasizes their significance. As obtained results and illustrated in Fig. 3 and Fig. 4, the developed method not only exhibits effectiveness but also outperforms other existing methods, underscoring its importance in the field of emotion recognition.

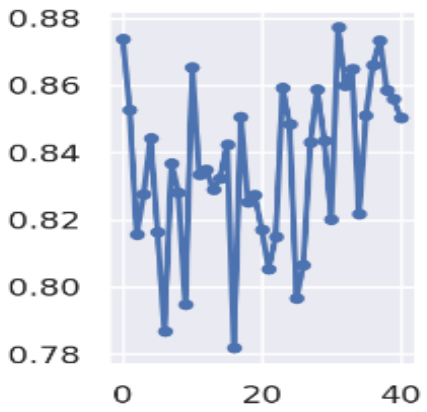


Fig. 6. Performance evaluation based on precision metric.

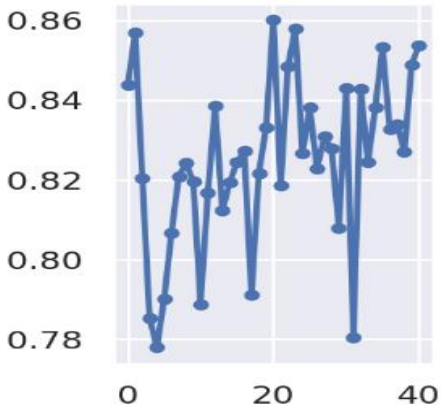


Fig. 7. Performance evaluation based on recall metric.

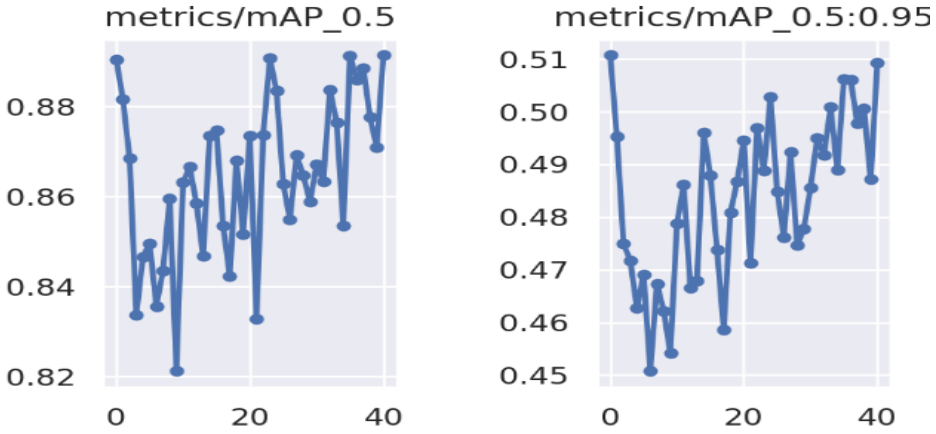


Fig. 8. Performance evaluation based on mAP metrics.

VI. Conclusion

Smart Health Houses aim to provide efficient and personalized medical services to individuals in their homes using advanced technologies, such as wearable devices, sensors, and machine learning algorithms. Among these technologies, computer vision-based systems that can analyze facial expressions and detect emotions accurately have gained significant attention due to their non-invasive nature. However, existing methods for facial expression analysis have limitations in detecting subtle changes in facial expressions. To address this issue, this paper proposes a deep learning-based facial expression analysis method using a YOLOv4-based algorithm. The method utilizes a custom dataset for training, validation, and testing and overcomes the limitations of existing methods, providing accurate results. The study concludes that the proposed method has the potential to enhance the accuracy of Smart Health Houses. For directions and future studies, investigating the potential of combining multiple technologies, including computer vision-based systems, wearable devices, and sensors, to enhance the accuracy and effectiveness of Smart Health Houses. This study holds considerable significance as it not only advances our understanding of emotion recognition through facial expression analysis but also offers a superior method for achieving accurate results. The outcomes presented in this research provide a valuable foundation for future studies in the realm of emotion recognition, paving the way for more sophisticated and precise techniques. Researchers can build upon these findings to develop enhanced algorithms and applications, ultimately contributing to a deeper comprehension of human emotions and their applications in various fields, such as psychology, human-computer interaction, and affective computing. Further exploring the limitations of existing facial expression analysis methods and developing more advanced techniques is to improve their accuracy. This could involve investigating the potential of using different deep learning architectures, such as convolutional neural networks, to enhance the performance of facial expression analysis models. Additionally, research could focus on developing methods to address issues such as variations in lighting conditions and occlusions of facial features.

Acknowledgments

This work was supported by Special Project of Philosophy and Social Sciences Research Ideological and Political Work of Jiangsu Province Higher Education Institutions (2022SJSZ0219), and Special Project of Jiangsu Higher Education Association (2021JDKT065), (2022JDKT128), and Project of the 2022 National Association for Basic Computer Education in Higher Education Institutions: (2022-AFCEC-410).

REFERENCES

[1] T. M. Ghazal et al., "IoT for smart cities: Machine learning approaches in smart healthcare—A review," Future Internet, vol. 13, no. 8, p. 218, 2021.

[2] M. M. Islam, A. Rahaman, and M. R. Islam, "Development of smart healthcare monitoring system in IoT environment," SN Comput Sci, vol. 1, pp. 1–11, 2020.

[3] V. Bhardwaj, R. Joshi, and A. M. Gaur, "IoT-based smart health monitoring system for COVID-19," SN Comput Sci, vol. 3, no. 2, p. 137, 2022.

[4] A. Das Gupta, S. M. Rafi, B. R. Rajagopal, T. Milton, and S. G. Hymlin, "Comparative analysis of internet of things (IoT) in supporting the health care professionals towards smart health research using correlation analysis," Bull. Env. Pharmacol. Life Sci., Spl, no. 1, pp. 701–708, 2022.

[5] J. Gao et al., "Ultra‐robust and extensible fibrous mechanical sensors for wearable smart healthcare," Advanced Materials, vol. 34, no. 20, p. 2107511, 2022.

[6] A. Sujith, G. S. Sajja, V. Mahalakshmi, S. Nuhmani, and B. Prasanalakshmi, "Systematic review of smart health monitoring using deep learning and Artificial intelligence," Neuroscience Informatics, vol. 2, no. 3, p. 100028, 2022.

[7] A. A. Nancy, D. Ravindran, P. M. D. Raj Vincent, K. Srinivasan, and D. Gutierrez Reina, "Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning," Electronics (Basel), vol. 11, no. 15, p. 2292, 2022.

[8] A. Aghamohammadi, M. C. Ang, E. A. Sundararajan, K. W. Ng, M. Mogharrebi, and S. Y. Banihashem, "Correction: A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos," PLoS One, vol. 13, no. 3, p. e0195418, 2018.

[9] V. Nandini, R. D. Vishal, C. A. Prakash, and S. Aishwarya, "A review on applications of machine vision systems in industries," Indian J Sci Technol, vol. 9, no. 48, pp. 1–5, 2016.

[10] C.-Z. Dong and F. N. Catbas, "A review of computer vision–based structural health monitoring at local and global levels," Struct Health Monit, vol. 20, no. 2, pp. 692–743, 2021.

[11] Y. Jiang, W. Wang, and C. Zhao, "A machine vision-based realtime anomaly detection method for industrial products using deep learning," in 2019 Chinese Automation Congress (CAC), IEEE, 2019, pp. 4842–4847.

[12] S. P. Yadav, S. Zaidi, A. Mishra, and V. Yadav, "Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (RNN)," Archives of Computational Methods in Engineering, vol. 29, no. 3, pp. 1753–1770, 2022.

[13] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, "Impact of deep learning approaches on facial expression recognition in healthcare industries," IEEE Trans Industr Inform, vol. 18, no. 8, pp. 5619–5627, 2022.

[14] P. Silapasuphakornwong and K. Uehira, "Smart mirror for elderly emotion monitoring," in 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), IEEE, 2021, pp. 356–359.

[15] F. A. Pujol, H. Mora, and A. Martínez, "Emotion recognition to improve e-healthcare systems in smart cities," in Research & Innovation Forum 2019: Technology, Innovation, Education, and their Social Impact 1, Springer, 2019, pp. 245–254.

[16] J. K. Pandey, V. Veeraiah, S. B. Talukdar, V. B. Talukdar, V. M. Rathod, and D. Dhabliya, "Smart City Approaches Using Machine Learning and the IoT," in Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, IGI Global, 2023, pp. 345–362.

[17] A. P. Plageras and K. E. Psannis, "IOT-based health and emotion care system," ICT Express, vol. 9, no. 1, pp. 112–115, 2023.

[18] R. Jahangir, Y. W. Teh, F. Hanif, and G. Mujtaba, "Deep learning approaches for speech emotion recognition: State of the art and research challenges," Multimed Tools Appl, pp. 1–68, 2021.

[19] A. Aghamohammadi, R. Ranjbarzadeh, F. Naiemi, M. Mogharrebi, S. Dorosti, and M. Bendechache, "TPCNN: two-path convolutional neural network for tumor and liver segmentation in CT images using a novel encoding approach," Expert Syst Appl, vol. 183, p. 115406, 2021.

[20] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," Procedia Comput Sci, vol. 175, pp. 689–694, 2020.

[21] A. M. Abu Nada, E. Alajrami, A. A. Al-Saqqa, and S. S. Abu-Naser, "Age and gender prediction and validation through single user images using cnn," 2020.

[22] R. Ranjbarzadeh et al., "Lung infection segmentation for COVID-19 pneumonia based on a cascade convolutional network from CT images," Biomed Res Int, vol. 2021, pp. 1–16, 2021.

[23] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Appl Sci, vol. 2, no. 3, p. 446, 2020.

[24] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," Pattern Recognit Lett, vol. 120, pp. 69–74, 2019.

[25] H. Ge, Z. Zhu, Y. Dai, B. Wang, and X. Wu, "Facial expression recognition based on deep learning," Comput Methods Programs Biomed, vol. 215, p. 106621, 2022.

[26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[27] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," Procedia Comput Sci, vol. 108, pp. 1175–1184, 2017.

[28] K. S. Gautam and S. K. Thangavel, "Video analytics-based facial emotion recognition system for smart buildings," International Journal of Computers and Applications, vol. 43, no. 9, pp. 858–867, 2021.

[29] R. Rajavel, S. K. Ravichandran, K. Harimoorthy, P. Nagappan, and K. R. Gobichettipalayam, "IoT-based smart healthcare video surveillance system using edge computing," J Ambient Intell Humaniz Comput, pp. 1–13, 2022.

[30] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," in Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, 2021, pp. 13029–13038.

[31] "Facial Emotion Rrecognition Dataset," Roboflow Universe, 2023.

[32] M. Murugavel, "YOLO V4.," https://manivannan-ai.medium.com/yolo-v4-750cd627064f.