ICIC Express Letters
Part B: Applications
Volume **15**, Number **5**, May **2024**

ICIC International ©2024 ISSN 2185-2766
pp. **487–493**

FACE MASK IMAGERY CLASSIFICATION WITH DATA AUGMENTATION

Kritsana Kumphet and Wansuree Massagram*

Department of Computer Science and Information Technology
Naresuan University
99 Moo 9, Thapo Sub-district, Muang District, Phitsanulok 65000, Thailand
kritsanaku64@nu.ac.th; *Corresponding author: wansureem@nu.ac.th

Received September 2023; accepted November 2023

Abstract. *This study focuses on training image analysis models to classify face mask usage in public places and enhance their performance through data augmentation techniques. Publicly available imagery depicting people using face masks was used to refine the training weights of pre-trained networks – small, medium, and large models in the YOLOv5 family. These models can detect three key aspects in a single analysis: 1) if someone is wearing a mask, 2) if the mask is worn correctly, and 3) the type of mask being worn. Multiple mask variations (color, style, type), varied subject numbers, and subject-observer orientations are accounted for in the training data. Augmentation techniques were used to improve the detection performance of each model, with augmented data added to the training set data provided a mean mAP increase of 13.9%. The data set augmentation for training with drawn mask imagery helped classify masks, but did not improve face detection when used on live subjects.*

Keywords: Face mask detection, Object detection, COVID-19, Transfer learning, Training data, Data augmentation, YOLOv5

1. **Introduction.** Face masking has been demonstrated to be an effective prophylactic in place of pharmaceutical approaches for reducing and slowing the spread of COVID-19, for which people can become infectious before displaying obvious symptoms [1]. Public health authorities wishing to ascertain rates of mask usage and details impacting effectiveness must monitor people and mask usage. Automated observation and categorization would reduce human workload, provide more immediate reporting of data, and reduce risks (such as infection) to human observers. Many facial recognition and detection systems have shown disparate performance for intersectional racial groups [2]. To accurately assess population usage of face masks, the data set must include classes matching the diversity of the population and also the diversity of masks used by the population [3].

This proposed study covers an assessment of an existant face mask dataset [4], training YOLOv5 models, and using a software tool for drawing face masks on images [5] for augmenting the training data set.

The remainder of this article is split into background material on face mask identification, methodology for data augmentation, results and discussion, and concluding remarks.

2. **Background.** Mask face detection models based on deep transfer learning and with classical machine learning classifiers have been proposed by Loey et al. [6] and Yang et al. [7]. The comparison results identified the most suitable algorithm that achieved the highest accuracy and consumed the least time in the process of training and detection. The performance of four different models: Faster R-CNN, R-FCN, SSD, and YOLOv5 was compared. The results indicated that YOLOv5 [8] outperformed the other models with 97.9% accuracy.

DOI: 10.24507/icicelb.15.05.487

Another study by Ieamsaard et al. describes the procedure for training a face mask detection model using the YOLOv5 algorithm. The dataset used in their study contains 628 images and three classes: mask, no mask, and wearing incorrect. During the training process, the model was trained for different epochs, including 20, 50, 100, 300, and 500. Their conclusion reveals that the highest accuracy achieved was 96.5% after training the model for 300 epochs, which serves as a useful guide for implementing the YOLOv5 model to identify different types of masks in our proposed training techniques [9].

Thus, our main objective is to leverage YOLOv5 for training a face mask detection model. However, the detection of face masks has become a widely explored area of research, primarily focused on distinguishing individuals who are wearing masks from those who are not. The contribution of this work focuses on identifying four types of commonly used masks and also marking masks when worn incorrectly. This enables a more refined and precise classification of various types of face masks.

In addressing the issue of dataset limitations, we have employed a solution by means of data augmentation, utilizing an open-source software project, "MaskTheFace". This open-source tool enabled the creation of a diverse dataset with face masks of multiple types added to the set of original photographs for training. Our use of this technique aligns with the research conducted by Anwar and Raychowdhury [5], where the method was proposed to expand the dataset for mask insertion and facial recognition model training. The process involves facial detection in input images, identification of six key facial landmarks, and subsequent overlaying of mask images onto the facial region using these key points. Notably, Anwar and Raychowdhury's work demonstrated that models trained on this augmented dataset achieved a high level of face detection accuracy, reaching up to 38% performance. Salim and Surantha [10] further applied augmentation techniques to compiling a dataset for training a face recognition model. Their research primarily aimed to achieve facial recognition even when individuals were wearing face masks. One of the outcomes revealed in this study is an approximately 9% increase in accuracy after utilizing the augmented dataset for training.

3. **Methodology.** The steps taken for preparation and analysis of our proposed YOLOv5-based mask type and use classification are illustrated in Figure 1. The procedure

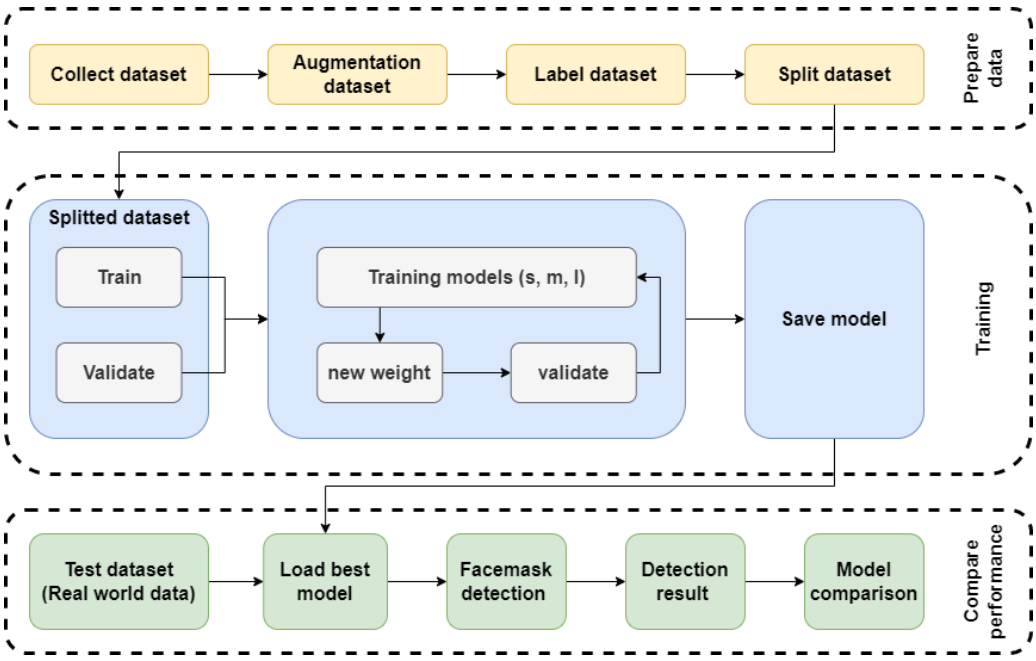

FIGURE 1. The framework for model training with data augmentation for face mask imagery analysis

can be divided into three parts: data preparation, model training, and evaluation. Data preparation involves data selection, data augmentation, data labeling, and a training/validation/test data split. Model training combines the training portion of the data and the model to generate weights for detection and recognition of face masks and types. Evaluation uses the data unseen during training to assess performance for each model.

3.1. **Data collection.** The dataset was assembled from publically available imagery, including a face mask dataset available through Kaggle [11]. A total of 1108 images were labeled with five categories: surgical mask, N95 mask, homemade mask, incorrectly worn, and no mask. This task was performed using the Makesense.ai service [12] and a sample of this dataset can be seen in Figure 2. The this dataset includes 1400, 400, 400, 150, and 500 labeled objects for training, classified as surgical, N95, homemade, incorrect, and no mask, respectfully.

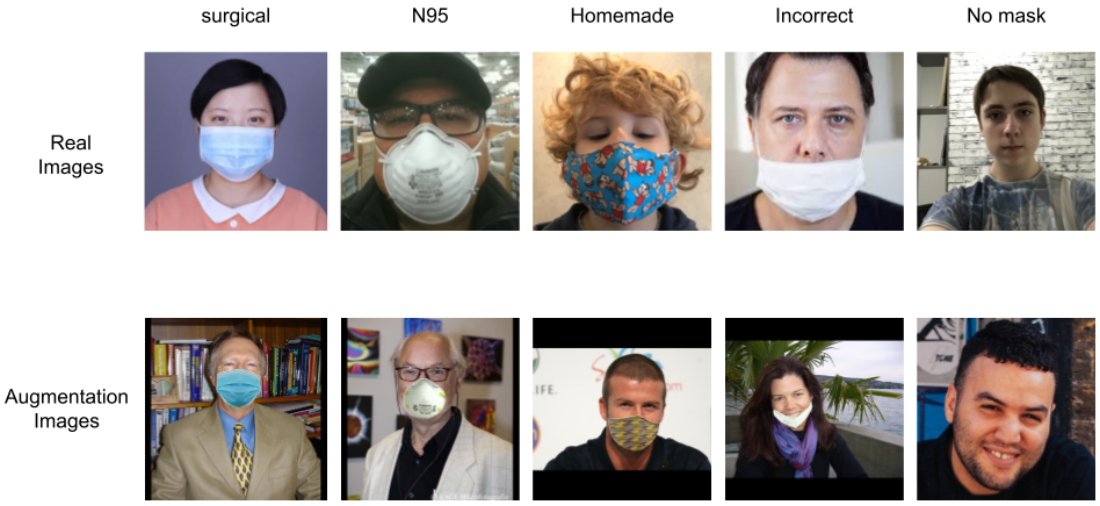

FIGURE 2. Example of actual images and augmentation images from our collected dataset with various numbers of objects for each class identified

3.2. **Augmentation, synthesis, and image generation.** Starting with the unaugmented image dataset, MaskTheFace [13] and Dlib [14] were collectively used for shape prediction, facial landmark localization, overlaying images, and object labeling. As a preparatory step, we readied templates for the types of face masks (adjusting to fit facial angles, resized the images, and modified the shapes for incorrect fit). In the mask application step, we used Dlib [14] for face detection and landmark localization and then MaskTheFace [13] to apply randomized masks by combining the appropriate mask image with the base facial image. For use in training or validation, images need to have identifying annotation (location and type of mask) in each image. We developed new functionality for MaskTheFace to support automatic labeling, writing the class and bounding box as part of the mask application process. The resultant dataset had 2500 images, of which 2023 were of usable quality, examples of which are depicted in Figure 2. Variation in mask types, colors, and patterns found in real-world conditions are included in this dataset, blue, white, green, black, and patterned masks, with people wearing some correctly, some incorrectly.

3.3. **Model training.** We have selected three YOLOv5 models for comparison, small, medium, and large. Our training started with pre-generated COCO dataset based weights, using Ultralytics-provided high-augmentation set of hyperparameters to delay overfitting. We modified the configuration, adding object classes for face mask usage and then trained the networks on two different datasets. We used training durations of 100 and 300 epochs

and with a batch size of 10. After training, we evaluated model performance with precision, recall, accuracy, and mean average precision (mAP).

3.4. **Evaluation.** We compared model performance to determine the benefits of training set augmentation with generated imagery. For this model evaluation, we used data collected from offline scenarios to analyze real-world effects. We used standard metrics (Precision, Recall, mAP) for figure-of-merit outputs from classification quantification.

4. **Result and Discussion.** Results of model performance evaluation can be found in Table 1, for models trained using only real-world data, or real-world data augmented with additional generated imagery. Additional model variations are model size and number of training iterations. The larger models offer modest performance advantages over the small model, systematically, while the number of training iterations has little effect, only sometimes positive. The largest improvement is increasing the amount of data in the training set, which provides a significant boost in performance irrespective of the other variations.

Table 1. The result from model training

Dataset	YOLOv5 model	Epoch	mAP@0.5
Real	s	100	78.4%
Real	s	300	78.1%
Real	m	100	79.9%
Real	m	300	79.6%
Real	l	100	79.2%
Real	l	300	78.8%
Real&Augment	s	100	92.4%
Real&Augment	s	300	92.7%
Real&Augment	m	100	**93.6**%
Real&Augment	m	300	93.2%
Real&Augment	l	100	92.2%
Real&Augment	l	300	**93.4**%

As part of our performance evaluation, we utilized two sets of real-world data, each comprising 100 images. Each image shows a single person exhibiting different behaviors, such as correctly worn masks, incorrectly worn masks, and masks not being worn. The dataset also contains images with individuals positioned with straight faces and faces turning at approximately 45 degrees. Figure 3 displays the results of our image detection in real time with different styles of masks and wear. The orange boxes represent homemade masks, the red boxes represent surgical masks, the yellow boxes represent incorrect masks, the pink boxes represent N95 masks, and the green boxes represent individuals not wearing masks.

As can be seen in the results summarized in Table 2, there were no clear advantages for any of the configurations. The highest Accuracy scoring configuration was a medium sized network trained on Real data for 300 epochs, closely followed by a large network trained with Real&Augmented dataset for 300 epochs, and two medium networks trained for 100 epochs.

Figure 4 shows the confusion matrix from a medium sized model trained on real data for 300 epochs. The model showed the highest accuracy in identifying people without masks and those wearing surgical masks, detecting fewer faces when people wore N95 respirators and homemade masks. People wearing masks incorrectly were classified across multiple other classes. This could be a result of incorrectly worn masks being classified as if worn correctly.

FIGURE 3. (color online) Example detection results from live-captured images

TABLE 2. The results summary of the testing phase (scale 0-1)

Dataset	Model	Epoch	Precision	Recall	Accuracy
Real	s	100	0.75	0.55	0.82
Real	s	300	0.64	0.48	0.79
Real	m	100	0.68	0.73	0.85
Real	m	300	0.70	0.71	0.87
Real	l	100	0.69	0.60	0.82
Real	l	300	0.67	0.64	0.84
Real&Augment	s	100	**0.81**	0.59	0.84
Real&Augment	s	300	**0.80**	0.58	0.84
Real&Augment	m	100	0.75	0.56	0.85
Real&Augment	m	300	0.72	0.52	0.82
Real&Augment	l	100	0.79	0.56	0.82
Real&Augment	l	300	0.76	0.62	0.86

While we found generated imagery partially helpful for data augmentation, additional investigation and refinement are needed to optimize applicability on real-world images. The software used for drawing masks on imaged faces (MaskTheFace and Dlib) was able to identify, localize, and modify faces in many, but not all images. Matching mask types for augmentation with the types used by people in a given area was challenging, as was assembling a population matched dataset for training.

5. **Conclusion.** This research introduces a methodology for developing a machine-learning model dedicated to face mask detection. Our objective is to categorize masks into five distinct types: surgical, homemade, N95, incorrect, and no mask. The model, constructed using YOLOv5 and custom datasets, incorporates the MaskTheFace approach to enhance the dataset. Demonstrating accuracy in predicting face masks, the model is adaptable for both image testing and real-time video cameras. Notably, our study reveals a substantial 11% improvement in precision through the implementation of our augmented dataset. The foundational model, trained with YOLOv5s for 100 epochs, achieved an accuracy of 84%, validating our hypothesis on precise detection and performance enhancement. In future

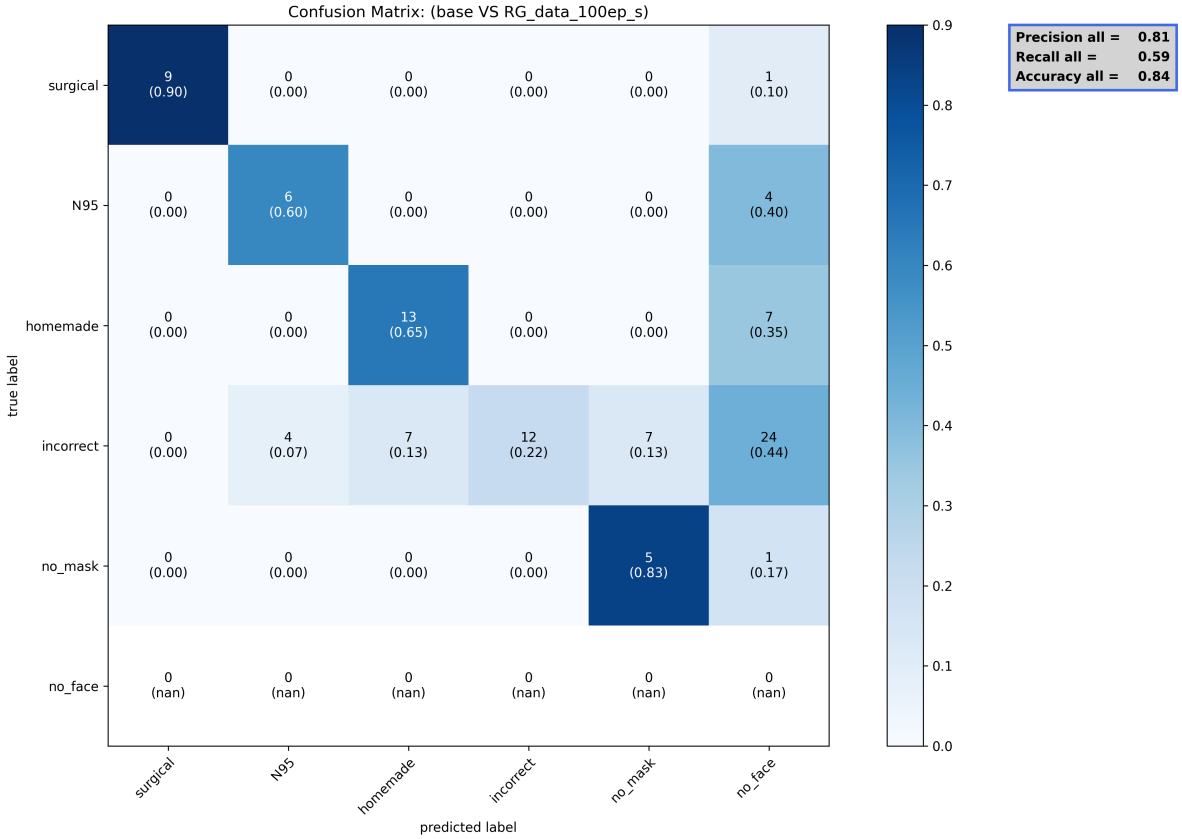

FIGURE 4. The best model result including confusion matrix and prediction images

endeavors, we aim to integrate GAN data augmentation to refine training performance and explore model size reduction for compatibility with microcontroller boards.

REFERENCES

[1] A. Aravindakshan, J. Boehnke, E. Gholami and A. Nayak, The impact of mask-wearing in mitigating the spread of COVID-19 during the early phases of the pandemic, *PLoS Glob Public Health*, vol.2, no.9, e0000954, 2022.

[2] S. Yucer, F. Tekras, N. Al Moubayed and T. P. Breckon, Measuring hidden bias within face recognition via racial phenotypes, *Proc. of Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 2022.

[3] S. Noiret, J. Lumetzberger and M. Kampel, Bias and fairness in computer vision applications of the criminal justice system, *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Orlando, FL, USA, pp.1-8, DOI: 10.1109/SSCI50451.2021.9660177, 2021.

[4] A. Maranhão, *Face Mask Detection*, [Version of the dataset], https://www.kaggle.com/datasets/andrewmvd/face-mask-detection, 2021.

[5] A. Anwar and A. Raychowdhury, Masked face recognition for secure authentication, *arXiv Preprint*, arXiv: 2008.11104, 2020.

[6] M. Loey, G. Manogaran, M. H. Taha and N. E. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, *Measurement*, vol.167, 108288, 2021.

[7] G. Yang, W. Feng, J. Jin, Q. Lei, X. Li, G. Gui and W. Wang, Face mask recognition system with YOLOv5 based on image recognition, *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020.

[8] Ultralytics, *YOLOv5*, 2020, https://docs.ultralytics.com/, Accessed on March 22, 2023.

[9] J. Ieamsaard, S. N. Charoensook and S. Yammen, Deep learning-based face mask detection using YOLOv5, *2021 9th International Electrical Engineering Congress (iEECON)*, 2021.

[10] R. J. Salim and N. Surantha, Masked face recognition by zeroing the masked region without model retraining, *International Journal of Innovative Computing, Information and Control*, vol.19, no.4, pp.1087-1101, 2023.

[11] Larxel, Face mask detection, *Kaggle*, https://www.kaggle.com/datasets/andrewmvd/face-mask-detection, Accessed on July 05, 2022.
[12] P. Skalski, *Makesense.ai*, https://skalskip.github.io/make-sense/, Accessed on April 22, 2022.
[13] Aqeelanwar, MaskTheFace: Convert face dataset to masked dataset, *GitHub*, https://github.com/aqeelanwar/MaskTheFace, Accessed on March 22, 2022.
[14] A. Rosebrock, Facial landmarks with Dlib, OpenCV, and Python, *PyImageSearch*, 2021, https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/, Accessed on July 15, 2022.