

Real-time Object Detection and Classification of Face Masks

Egemen Eroglu

December 29, 2024

Abstract

A real-time system has been developed for detecting and classifying face masks using the YOLO (You Only Look Once) object detection algorithm. The system identifies three classes: 'mask,' 'improperly worn mask,' and 'no mask,' achieving a mean Average Precision (mAP@0.50) of 94.04%. Real-time detection is facilitated through webcam input, providing a practical tool for monitoring mask compliance in public settings.

1 Introduction

This project was prepared as part of the CMP3011 - Introduction to Computer Vision course. Its objective is to develop a real-time system for detecting and classifying face masks. The system leverages YOLOv3, an efficient object detection algorithm, to identify and classify masks into three categories: 'mask,' 'improperly worn mask,' and 'no mask.' The project emphasizes practical applications in object detection and classification, combining theoretical concepts with hands-on implementation.

2 Dataset

Masks play a crucial role in protecting individuals against respiratory diseases, particularly during the COVID-19 pandemic, when mask-wearing was one of the few effective precautions available in the absence of immunization. The dataset used in this project enables the development of a model to detect people wearing masks, not wearing them, or wearing masks improperly.

The *Face Mask Detection* dataset consists of 853 annotated images categorized into three classes:

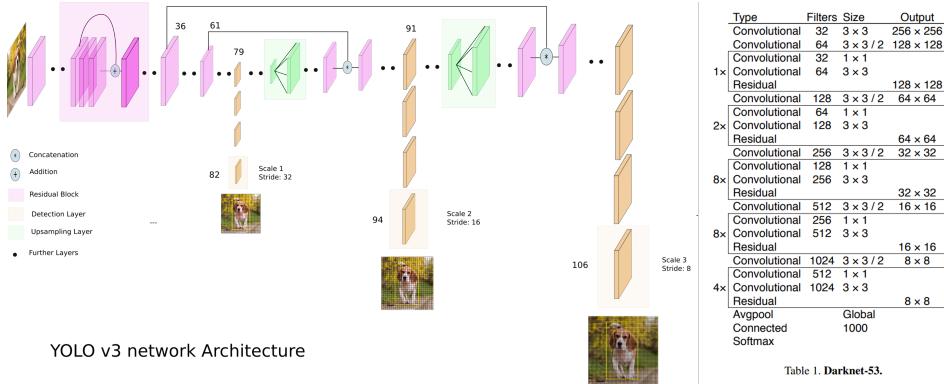
- **With mask:** Properly worn masks covering essential facial areas.
- **Without mask:** Absence of a face mask.
- **Mask worn incorrectly:** Masks worn improperly, such as leaving the nose exposed.

Annotations are provided in the PASCAL VOC format, including bounding boxes and class labels, ensuring compatibility with object detection tasks. This dataset is publicly available and can be accessed at <https://www.kaggle.com/datasets/andrewmvd/face-mask-detection>.

3 Methodology

3.1 YOLOv3 Architecture

YOLOv3 employs a fully convolutional neural network with Darknet-53 as its backbone, designed for efficient multi-scale object detection. The model processes images in a single forward pass, predicting bounding boxes and class probabilities simultaneously.



3.2 Data Preprocessing

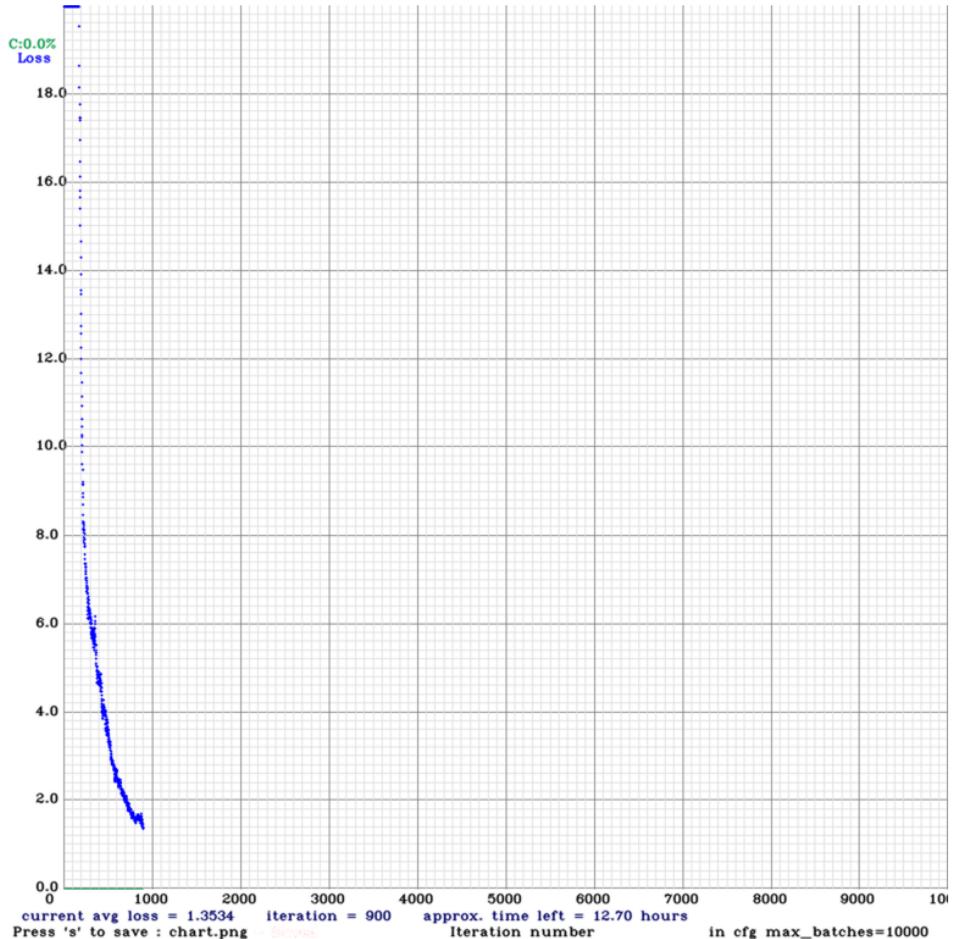
To prepare the dataset for YOLOv3, images were resized to 416×416 pixels and normalized to meet input requirements. Annotations were converted into the YOLO format, specifying bounding box coordinates and class labels. Data augmentation techniques, such as random scaling, cropping, and flipping, were applied to improve model generalization.

3.3 Model Training

The YOLOv3 model was trained using the Darknet framework with the following hyperparameters:

- **Pre-trained Weights:** Initialized with `darknet53.conv.74`.
- **Optimizer:** Stochastic Gradient Descent (SGD).
- **Training Duration:** 2000 iterations.
- **Batch Size:** 64.
- **Learning Rate:** 0.001, with scheduled reductions.
- **Loss Function:** Multi-scale loss incorporating MSE for IoU, objectness, and classification errors.

Training utilized a T4 GPU for accelerated computation.



4 Evaluation

Performance metrics for object detection are presented in Table 1.

Table 1: Evaluation Metrics for YOLOv3 Model

Metric	Value
Class-wise AP (Mask)	96.96%
Class-wise AP (Improperly worn)	92.80%
Class-wise AP (No mask)	92.37%
mAP@0.50	94.04%
Precision	88%
Recall	95%
F1-Score	0.92
Average IoU	69.55%

5 Real-time Detection Application

A real-time detection system was implemented using OpenCV to capture video input from a webcam. Each frame is processed by the trained YOLOv3 model, which detects and classifies face masks in real-time. Detected objects are highlighted with bounding boxes and corresponding labels.

6 Visualization and Results

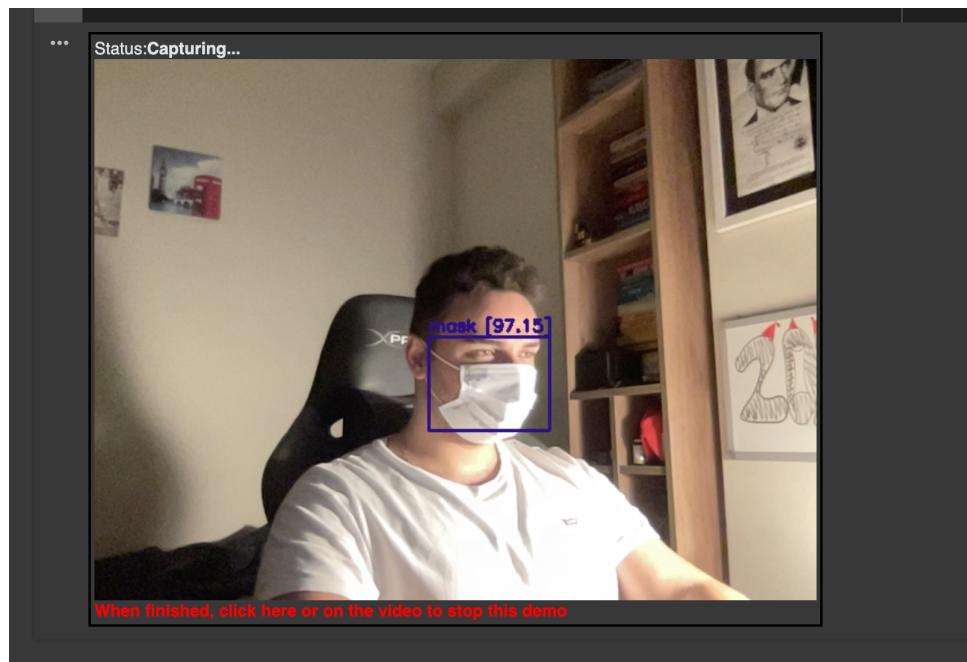


Figure 1: YOLOv3 detecting and classifying a properly worn face mask.

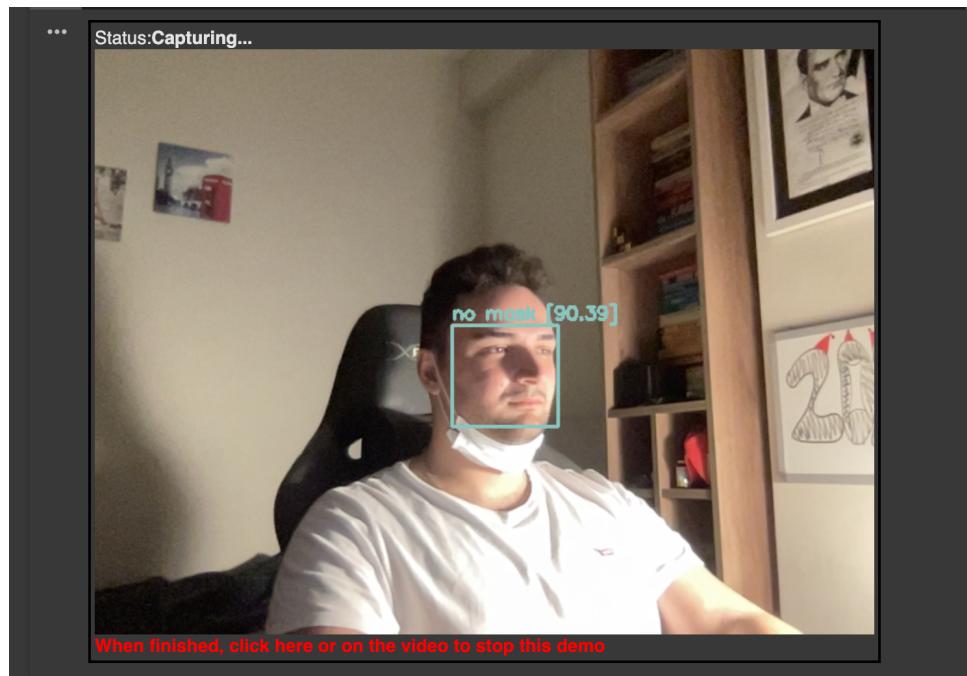


Figure 2: YOLOv3 detecting and classifying the absence of a face mask.

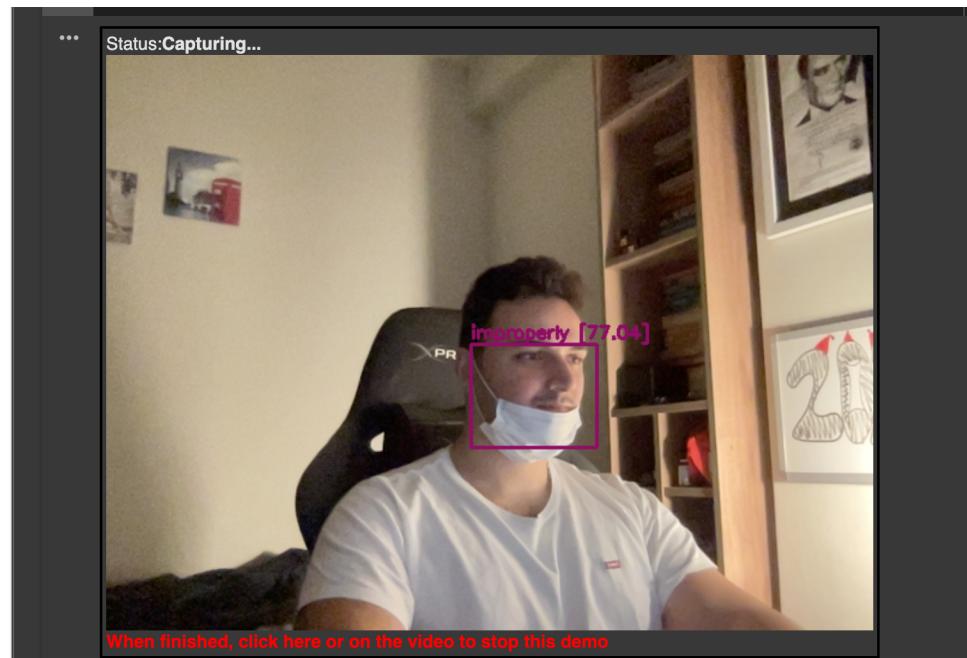


Figure 3: YOLOv3 detecting and classifying an improperly worn face mask.

7 Conclusion

This project successfully implements a real-time face mask detection and classification system using YOLOv3. The model achieves high accuracy and precision, validating its practicality for public safety applications.

Future work could explore integrating Vision Transformers (ViT) for feature extraction, which has shown promise in improving object detection performance. Additionally, implementing the more advanced YOLOv9 architecture, known for its enhanced accuracy and efficiency compared to YOLOv3, could further optimize detection performance. These advancements could make the system more robust and scalable for diverse real-world scenarios.

References

- Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.
- Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*.
- Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. *arXiv preprint arXiv:2107.08430*.
- Doshi, J., & Yadav, N. (2020). Face Mask Detection Using Convolutional Neural Networks and Machine Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
- Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*.
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A Survey of the Recent Architectures of Deep Convolutional Neural Networks. *Artificial Intelligence Review*, 53(8), 5455–5516.
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. *Proceedings of the International Conference on Machine Learning (ICML)*, 2020.
- Face Mask Detection Dataset: <https://www.kaggle.com/datasets/andrewmvd/face-mask-detection>.
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(2), 318–327.