Научная статья
Original article
УДК 004.048

СИСТЕМА ОБНАРУЖЕНИЯ НОШЕНИЯ МАСКИ НА ОСНОВЕ ГЛУБОКОГО ОБУЧЕНИЯ

A MASK-WEARING DETECTION SYSTEM BASED ON DEEP LEARNING

Ли Кай, студент факультета специального машиностроения, Московский государственный технический университет имени Н. Э. Баумана, г. Москва. likai226717@gmail.com

Лу Чжоюе, студентка факультета специального машиностроения, Московский государственный технический университет имени Н. Э. Баумана, г. Москва. 1153670428@qq.com

Li kai, Li Kai, student of the Faculty of Special Engineering, Bauman Moscow State Technical University, Moscow. likai226717@gmail.com

Lu Zhuoyue, student of the Faculty of Special Engineering, Bauman Moscow State Technical University, Moscow. 1153670428@qq.com

Аннотация. В данной работе была разработана система обнаружения изображений, позволяющая обнаружить ношение маски. Описываются основные задачи обнаружения объектов. В работе были проанализированы существующие фреймворки глубокого обучения. Был выбран фреймворк глубокого обучения YOLO v5 и описана его структура. В статье рассматривается подход к обнаружению изображений с использованием фреймворка YOLO v5. Был разработан графический интерфейс для обнаружения изображений и видеозаписей. Кроме того, были проанализированы результаты обучения.

Abstract. In this article, an image detection system has been developed to detect the wearing of a mask. The main tasks of object detection are described. Existing deep learning frameworks have been analyzed in this article. The deep learning framework YOLO v5 was chosen and its structure is described. In this article, an image detection approach using YOLO v5 framework is discussed. A graphical interface for image and video detection has been developed. In addition, the training results have been analyzed.
Ключевые слова: коронавирус, нейронные сети, глубокое обучение, YOLO v5
Keywords: coronavirus, neural networks, deep learning, YOLO v5

Введение

В условиях нынешней глобальной эпидемии ношение маски имеет большое значение для здоровья людей. Можно разработать автоматизированную систему обнаружения, которая будет анализировать изображение и определять, носит ли человек на изображении маску. Анализ изображений может быть автоматизирован с помощью компьютерного зрения и алгоритмов глубокого обучения на основе сверточнных нейронных сетей CNN (сокращение от «Convolutional Neural Network»). Как широко используемая и легко внедряемая система глубокого обучения, YOLO (сокращение от «You Only Look Once») [1] хорошо подходит для этой задачи.

1. Задачи обнаружения объектов

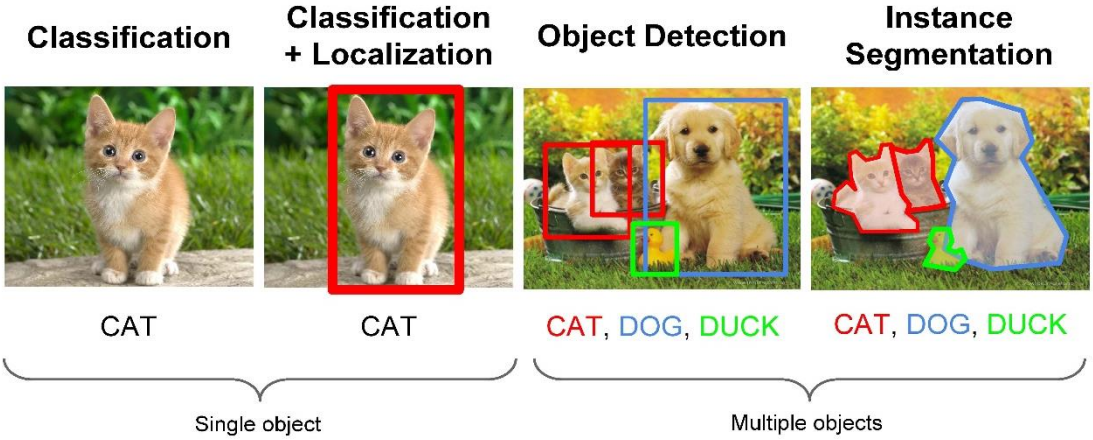


Рис. 1 задачи обнаружения объектов

Задачи обнаружения объектов определяются следующим образом [2]: получив произвольное изображение, определить, есть ли в нем экземпляры семантических объектов из заранее определенных категорий, и, если есть, вернуть пространственное расположение и протяженность.

2. фреймворки обнаружения

С тех пор как глубокое обучение вошло в область обнаружения объектов, фреймворки обнаружения в целом можно разделить на две основные категории:

a) Двухэтапная система обнаружения, которая включает этап предварительной обработки для предложения региона, что делает общий конвейер двухэтапным.

b) Одноэтапная система обнаружения, или система без предложения региона, которая представляет собой единственный предложенный метод, не разделяющий предложение обнаружения, что делает общий конвейер одноэтапным.

Как показано на рисунке 2, двухэтапные системы обнаружения находятся над осями, а одноэтапные системы обнаружения – под ними.

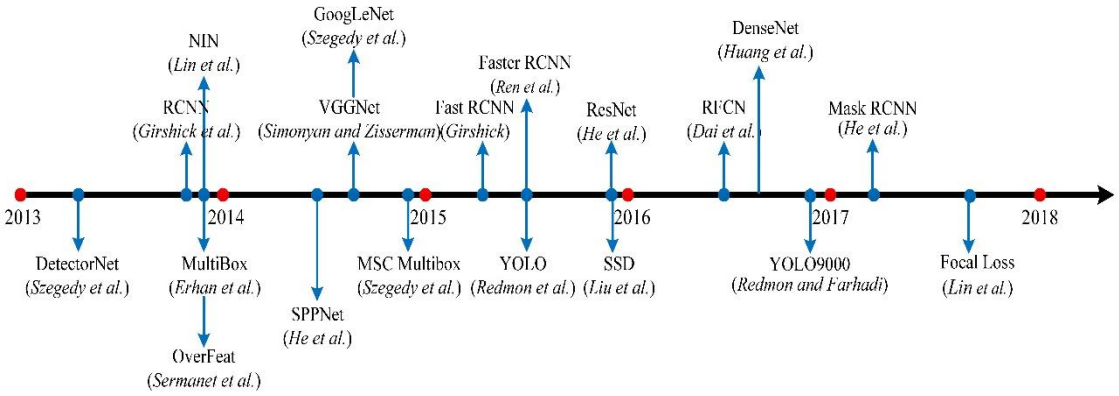


Рис. 2 вехи в обнаружении объектов

Благодаря своей архитектуре, которая была создана для того, чтобы легко находить и извлекать признаки на изображении, CNN широко используются при классификации, сегментации и в задачх распознавания. Существует множество архитектур CNN, подходящих для распознавания объектов, и R-CNN [3] является

одной из них.

R-CNN (сокращение от «Region Based CNN»), это основная модель, предложенная для детектирования объектов. Работает как обычный классификатор изображений. На вход сети подаются разные регионы изображения и для них делается предсказания. К сожалению, изначальная модель уже устарела, но архитектура R-CNN породила серию предметно ориентированных улучшенных моделей, в том числе: сеть Fast R-CNN [4] (быстрая R-CNN) для задачи классификации и регрессии покрывающего объект прямоугольника; сеть Faster R-CNN [5] (ускоренная Fast R-CNN), использующую вспомогательную подсеть для генерации регионов интереса.

В 2016, Redmon предложил end-to-end метод детектирования объектов YOLO. YOLO является наиболее быстрой моделью нейронной сети, подход, используемый в ней, значительно отличается от R-CNN. Эта архитектура и была выбрана для реализации, т.к. она является наиболее быстрой и простой в программной реализации. В данной работе для детектирования масок на фото мы будем использовать архитектуру YOLO, а именно её последнюю модификацию YOLOv5.

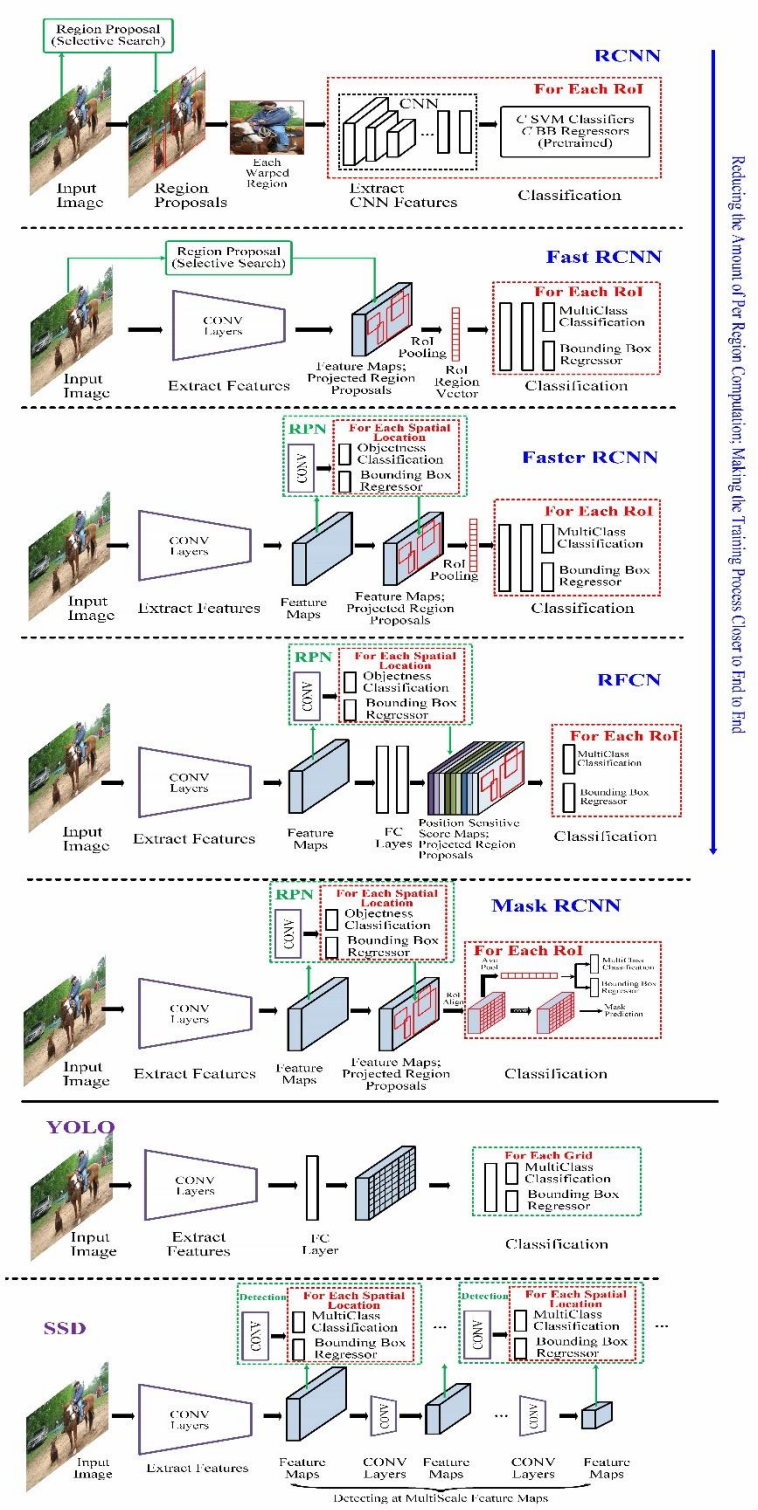


Рис. 3 схемы высокого уровня фреймворков для обнаружения объектов

3. Архитектур YOLOv5 и процесс обучения

В целом такую архитектуру можно представить в следующем виде:

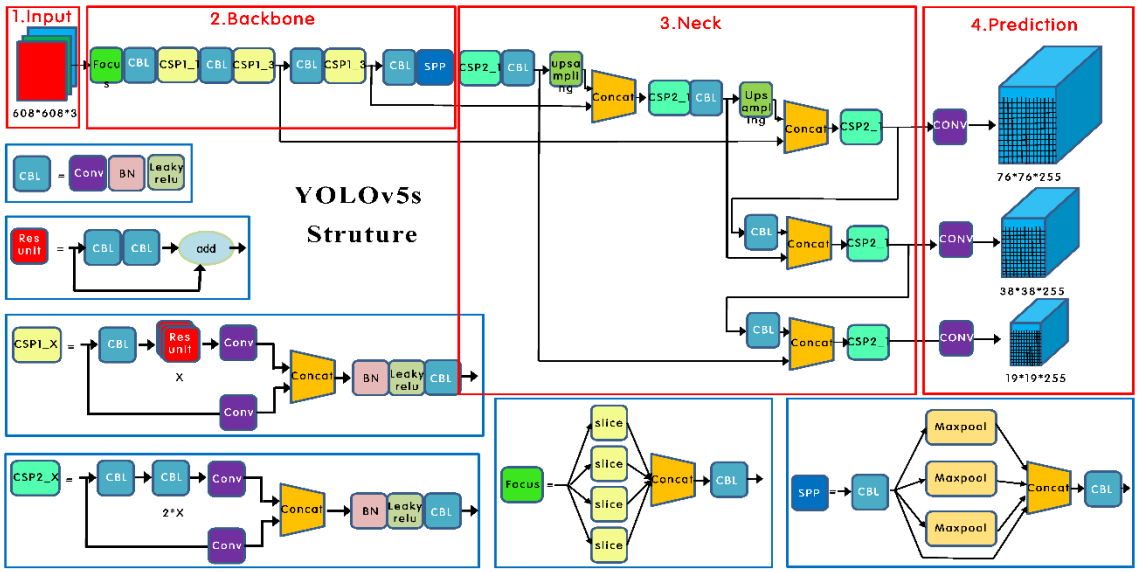


Рис. 4 сетевая структура Yolov5s

3.1. наборы данных

В данной работе эксперименты проводились на основе двух общедоступных наборов данных медицинских лицевых масок и данных, обработанных самостоятельно.

Первый общедоступный набор данных – это «COVID-19 Medical Face Mask Detection Dataset» (MFMDD) (https://www.kaggle.com/datasets/mloey1/medical-face-mask-detection-dataset). Набор данных MFMDD состоит из 682 изображений. На рис.5 показаны примеры изображений из MFMDD [6].

Второй общедоступный набор данных – это «Face Mask Dataset» (FMD) (https://www.kaggle.com/andrewmvd/face-mask-detection). Набор данных FMD состоит из 853 изображений. Некоторые образцы FMD представлены на рис.6.



Рис. 5 примеры изображений из MFMDD



Рис. 6 примеры изображений из FMD

Третий набор данных – это изображения, полученные из публичного интернета с помощью поискового робота и маркированные с помощью программы labelimg.



Рис. 7 примеры изображений из публичного интернета

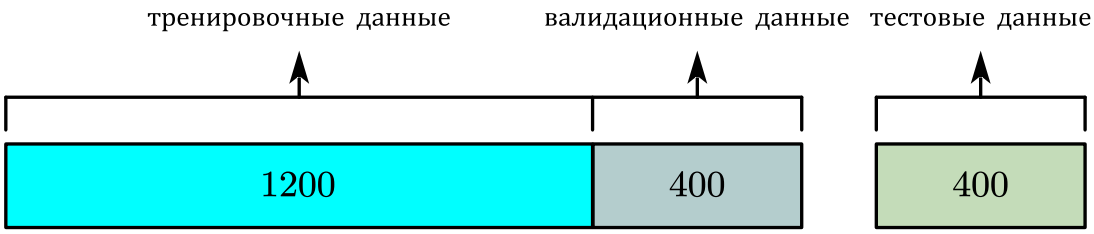


Рис. 8 количество компонентов данных

Общее количество подготовленных изображений составило 2000, которые были разделены на три части: тренировочные, валидационные и тестовые. Соответствующие номера изображений – 1200, 400 и 400.

3.2. процесс маркировки данных

В программе labelimg лица на фотографии выделяются прямоугольной рамкой

и маркируются как «mask» и «face» соответственно.

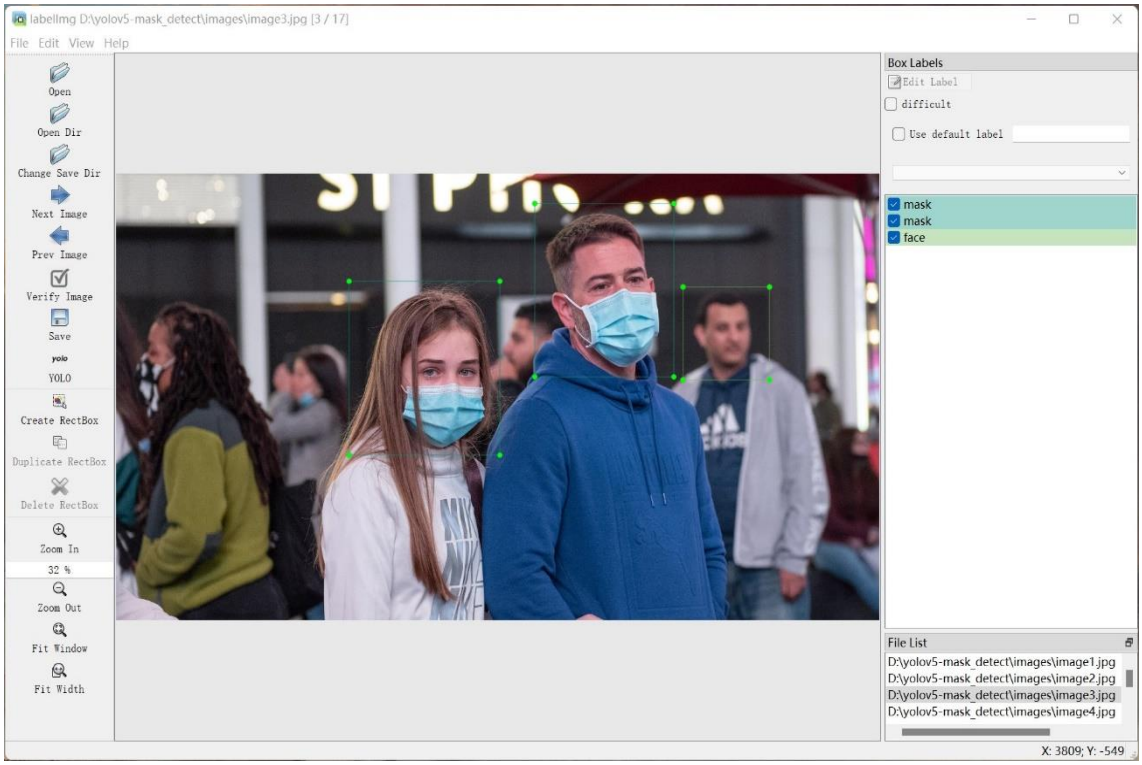


Рис. 9 процесс маркировки данных

При каждом сохранении процесса маркировки создается файл в формате txt, в котором хранится информация о классификации изображений.

3.3. Процесс обучения и графический интерфейс

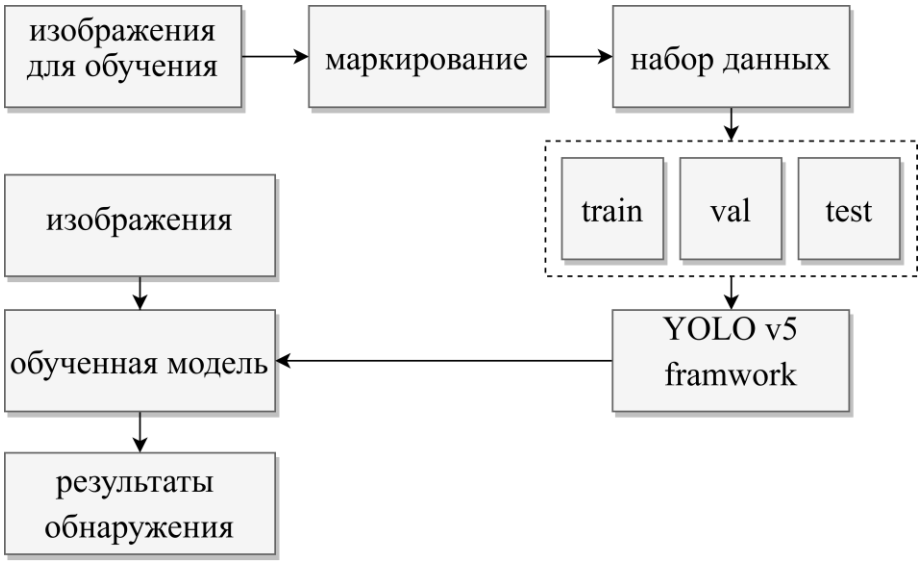


Рис. 10 блок-схема системы обнаружения изображений

Графический интерфейс был разработан в работе с помощью pyqt5. С помощью этого интерфейса можно выполнить функции распознавания изображений и видео.

Результаты распознавания доступны после загрузки выбранного изображения в интерфейс, как показано на рисунках 11 и 12. В дополнение к результатам, отмеченным квадратиками в интерфейсе, количество масок и лиц можно увидеть в терминале python.

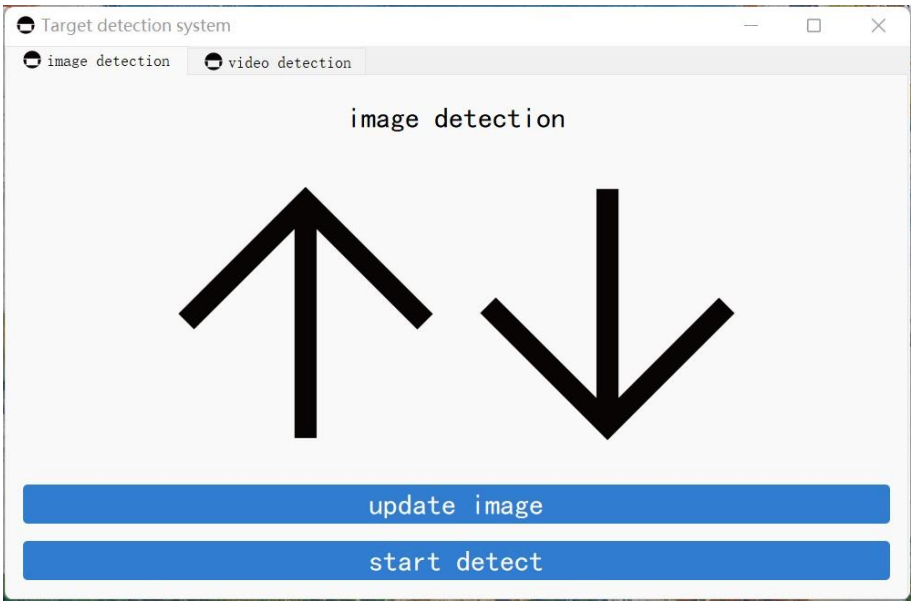


Рис. 11 Графический интерфейс 1

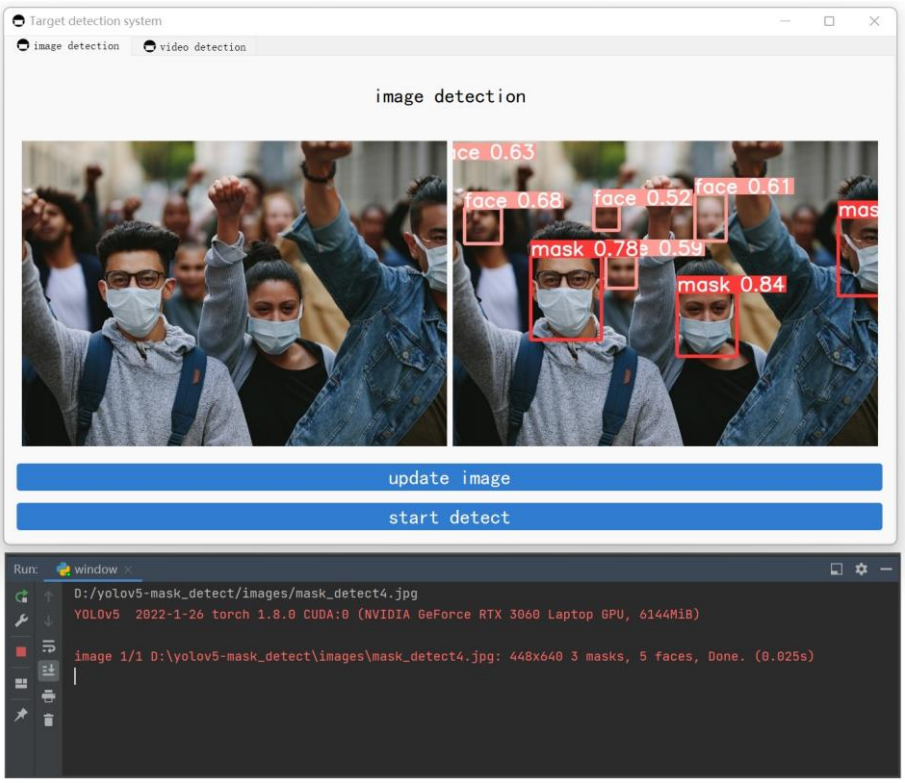


Рис. 12 Функция обнаружения изображений

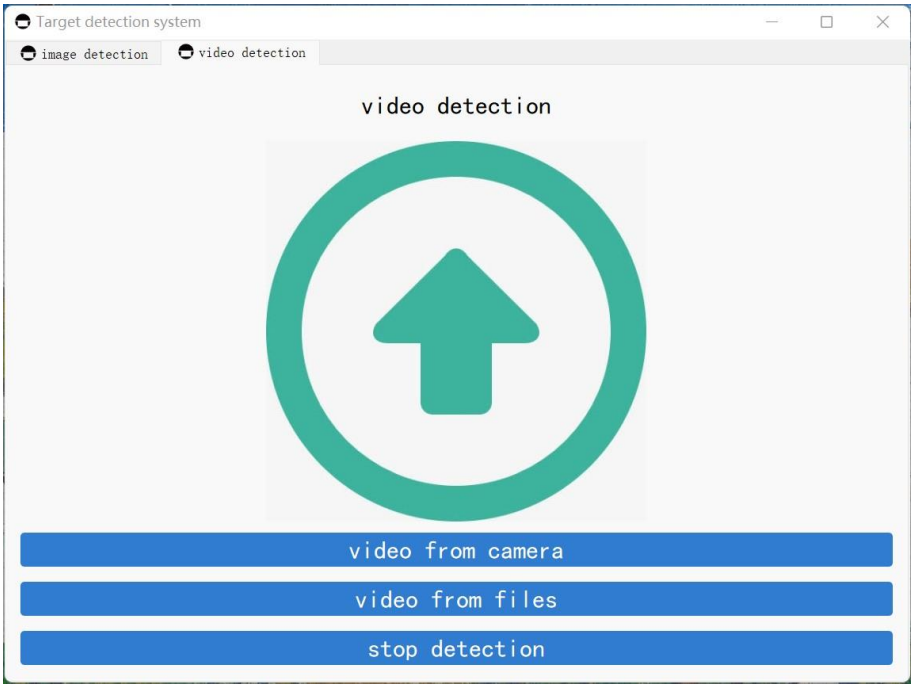


Рис. 13 Графический интерфейс 2

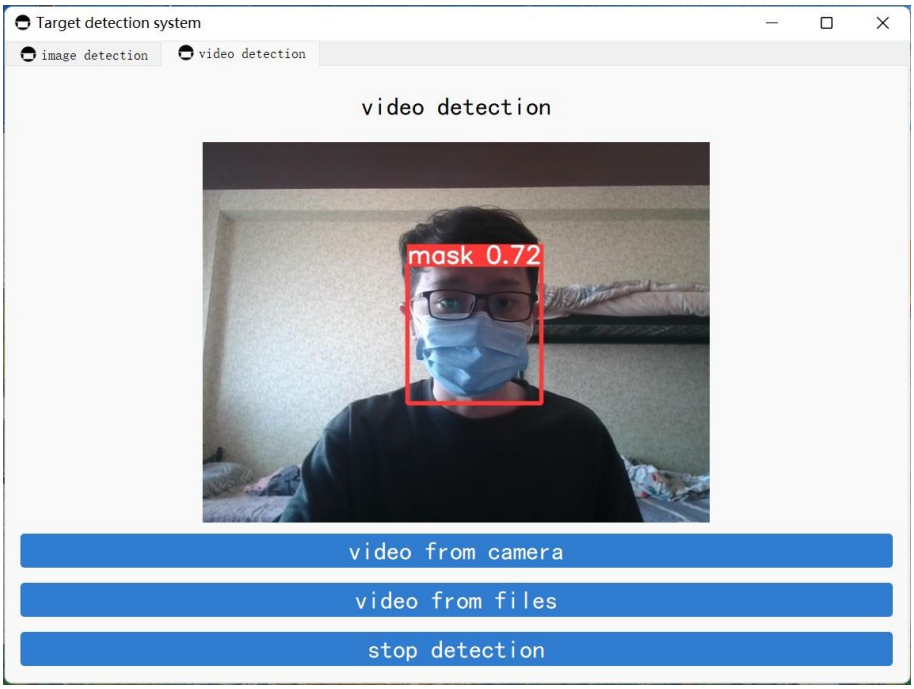


Рис. 14 Функция обнаружения видео

В интерфейсе обнаружения видео можно выбрать видео, вводимое камерой, или выбрать видеофайл. На рисунке 14 показаны результаты обнаружения видеоизображения, поступающего с камеры.

3.4. Анализ результатов

Данные импортируются в систему обучения, обучение завершается, и получаются два файла, содержащие информацию о весах: last.pt и best.pt, а также несколько графиков, показывающих хорошие и плохие результаты обучения. Точность распознавания для масок составляет 69,4%, а для лиц – 97%.

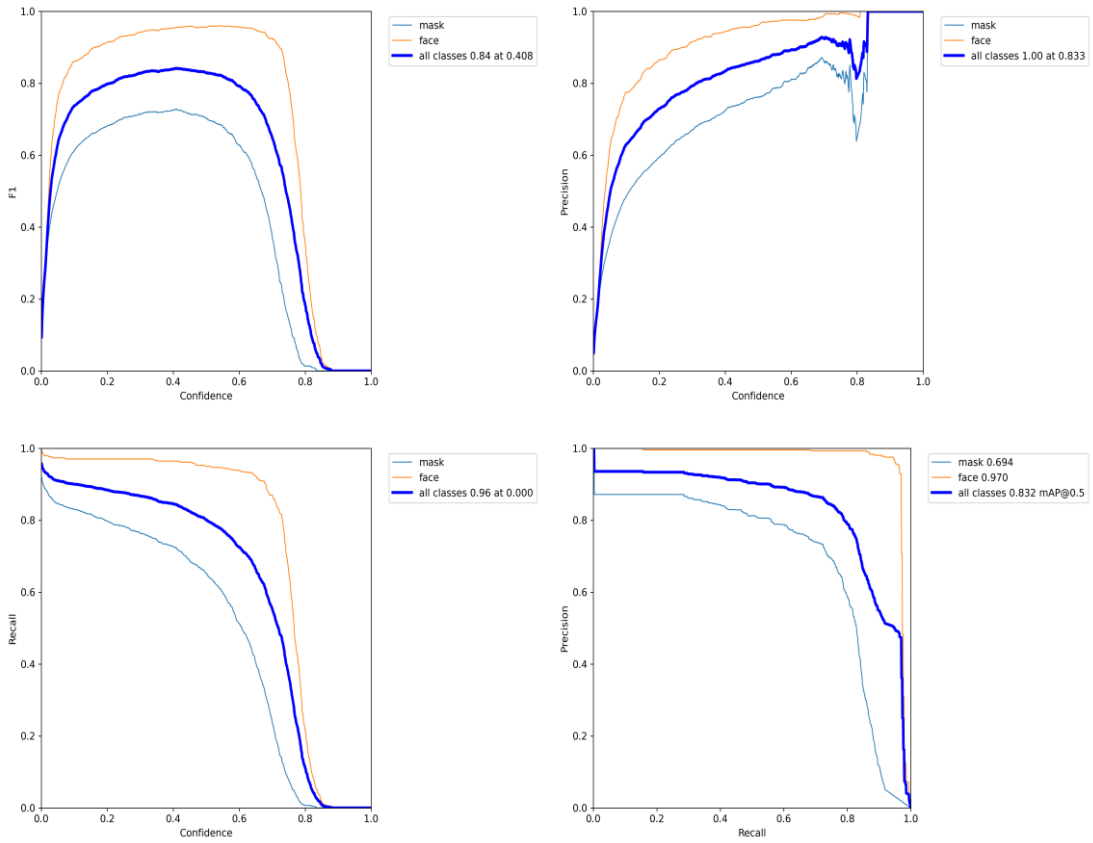


Рис. 15 результаты обучения

Как видно на рисунке 12, ошибка детектирования постепенно уменьшается, а точность детектирования увеличивается в процессе обучения.

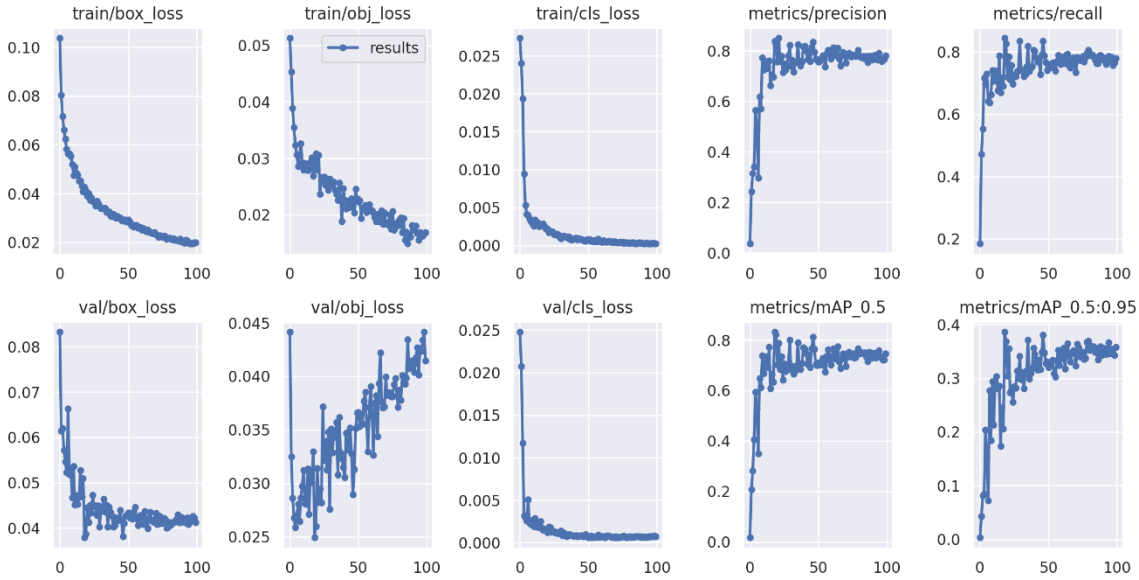


Рис. 16

Обученная модель способна обвести лицо на входном изображении прямоугольной рамкой и пометить его словами mask и face. Результаты

детектирования показаны на рисунках 12.



Рис. 17 результаты детектирования 1



Рис. 18 результаты детектирования 2

4. Заключение

Модель для определения того, надета ли маска, была обучена в работе с использованием YOLOv5. Модель обладает хорошей точностью при обнаружении лиц с масками и без масок на изображениях, содержащих множество объектов. Однако количество данных, используемых для обучения, невелико, и увеличение количества данных повысило бы точность модели.

Литература

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91
2. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Li F. (2015) ImageNet large scale visual recognition challenge.
3. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.1109/CVPR.2014.81
4. Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), 1440–1448. https://doi.org/10.1109/ICCV.2015.169
5. Ren S., He K., Girshick R., Sun J. (2015) Faster R-CNN: Towards real time object detection with region proposal networks.
6. Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection,Sustainable Cities and Society, Volume 65, 2021, https://doi.org/10.1016/j.scs.2020.102600

References

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91
2. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Li F. (2015) ImageNet large scale visual recognition challenge.
3. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.1109/CVPR.2014.81

4. Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), 1440–1448. https://doi.org/10.1109/ICCV.2015.169
5. Ren S., He K., Girshick R., Sun J. (2015) Faster R-CNN: Towards real time object detection with region proposal networks.
6. Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection,Sustainable Cities and Society, Volume 65, 2021, https://doi.org/10.1016/j.scs.2020.102600

© Ли Кай, Лу Чжоюе, 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №5/2022.

Для цитирования: Иванов И.И., Петров П.П. Система обнаружения ношения маски на основе глубокого обучения// Научно-образовательный журнал для студентов и преподавателей «StudNet» №5/2022.