МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»

ОТЧЕТ

**Домашнее задание №\_\_1\_\_**

по дисциплине«Методы машинного обучения»

Тема: «Классификация изображений»

ИСПОЛНИТЕЛЬ: \_\_\_\_\_Лу Жуньда \_\_

ФИО

группа ИУ5И-\_22М\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_202\_ г.

ПРЕПОДАВАТЕЛЬ: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ФИО

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_202\_ г.

Москва - 2024

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Задание

Домашнее задание по дисциплине направлено на анализ современных методов машинного обучения и их применение для решения практических задач. Домашнее задание включает три основных этапа:

1. выбор задачи;
2. теоретический этап;
3. практический этап.

Этап выбора задачи предполагает анализ ресурса [paperswithcode.](https://paperswithcode.com/sota) Данный ресурс включает описание нескольких тысяч современных задач в области машинного обучения. Каждое описание задачи содержит ссылки на наиболее современные и актуальные научные статьи, предназначенные для решения задачи (список статей регулярно обновляется авторами ресурса). Каждое описание статьи содержит ссылку на репозиторий с открытым исходным кодом, реализующим представленные в статье эксперименты. На этапе выбора задачи обучающийся выбирает одну из задач машинного обучения, описание которой содержит ссылки на статьи и репозитории с исходным кодом.

Теоретический этап включает проработку как минимум двух статей, относящихся к выбранной задаче. Результаты проработки обучающийся излагает в теоретической части отчета по домашнему заданию, которая может включать:

* описание общих подходов к решению задачи;
* конкретные топологии нейронных сетей, нейросетевых ансамблей или других моделей машинного обучения, предназначенных для решения задачи;
* математическое описание, алгоритмы функционирования, особенности обучения используемых для решения задачи нейронных сетей, нейросетевых ансамблей или других моделей машинного обучения;
* описание наборов данных, используемых для обучения моделей;
* оценка качества решения задачи, описание метрик качества и их значений;
* предложения обучающегося по улучшению качества решения задачи.

Практический этап включает повторение экспериментов авторов статей на основе представленных авторами репозиториев с исходным кодом и возможное улучшение обучающимися полученных результатов. Результаты проработки обучающийся излагает в практической части отчета по домашнему заданию, которая может включать:

* исходные коды программ, представленные авторами статей, результаты документирования программ обучающимися с использованием диаграмм UML, путем визуализации топологий нейронных сетей и другими способами;
* результаты выполнения программ, вычисление значений для описанных в статьях метрик качества, выводы обучающегося о воспроизводимости экспериментов авторов статей и соответствии практических экспериментов теоретическим материалам статей;
* предложения обучающегося по возможным улучшениям решения задачи, результаты практических экспериментов (исходные коды, документация) по возможному улучшению решения задачи.

## Выбор задачи

Обнаружение объектов - это задача компьютерного зрения, целью которой является обнаружение и определение местоположения интересующих объектов на изображении или видео. Задача включает в себя определение положения и границ объектов на изображении, а также классификацию объектов по различным категориям. Эта задача является важнейшей частью распознавания зрения, наряду с классификацией и поиском изображений.

Современные методы можно разделить на два основных типа: одноэтапные и двухэтапные:

В одноэтапных методах приоритетом является скорость вывода, и в качестве примера можно привести модели YOLO, SSD и RetinaNet.

В двухэтапных методах приоритет отдается точности обнаружения, и в качестве примера можно привести такие модели, как Faster R-CNN, Mask R-CNN и Cascade R-CNN.

Наиболее популярным эталоном является набор данных MSCOCO. Модели обычно оцениваются по метрике средней точности.

1. **Теоретическая часть**

Я выбрал две следующие статьи: “Deep Residual Learning for Image Recognition” и “YOLOv3: An Incremental Improvement”.

### 3.1 Обзор подходов к решению задачи обнаружения объектов

Задача обнаружения объектов является одной из ключевых в компьютерном зрении и включает в себя нахождение и классификацию объектов в изображении. Современные методы обнаружения объектов делятся на два основных подхода: одноэтапные (single-stage) и двухэтапные (two-stage) методы.

* 1. Одноэтапные методы (например, YOLO, SSD) выполняют обнаружение и классификацию объектов в один проход, что обеспечивает высокую скорость работы. YOLOv3 (You Only Look Once) является одним из самых известных одноэтапных детекторов. Он выполняет предсказание координат ограничивающих рамок и классов объектов напрямую из изображений, используя полностью сверточные сети.
  2. Двухэтапные методы (например, Faster R-CNN) сначала генерируют предположительные области (region proposals), а затем классифицируют их и уточняют координаты рамок. Эти методы, как правило, более точные, но медленнее в сравнении с одноэтапными.

### 3.2 Конкретные топологии нейронных сетей

**YOLOv3 (из статьи "YOLOv3: An Incremental Improvement")**

YOLOv3 представляет собой улучшенную версию предыдущих моделей YOLO. Ключевые компоненты YOLOv3 включают:

* **Архитектура**: YOLOv3 использует архитектуру Darknet-53, которая включает 53 сверточных слоя с остаточными соединениями.
* **Предсказание ограничивающих рамок**: Использует кластеры размеров для предсказания рамок и предсказывает координаты с использованием логистической регрессии.
* **Классификация**: Выполняется многоклассовая классификация для каждой рамки, используя бинарную кросс-энтропию в качестве функции потерь.
* **Предсказания на нескольких масштабах**: YOLOv3 предсказывает объекты на трех разных масштабах, что улучшает детекцию мелких объектов.

**ResNet (из статьи "Deep Residual Learning for Image Recognition")**

ResNet (Residual Networks) представляет собой семейство глубоких нейронных сетей, использующих остаточные соединения для облегчения обучения очень глубоких моделей. Основные аспекты ResNet включают:

* **Архитектура**: Сеть состоит из нескольких блоков, каждый из которых включает два или три сверточных слоя с остаточными соединениями.
* **Остаточные соединения**: Они позволяют избежать проблемы затухающих градиентов, добавляя "короткие пути" (shortcut connections), которые пропускают один или несколько слоев.
* **Глубина**: ResNet может иметь очень большую глубину, например, ResNet-50, ResNet-101, и ResNet-152, где число слоев указывается в названии.

### 3.3 Математическое описание и алгоритмы функционирования

**YOLOv3:**

* + - 1. Предсказание ограничивающих рамок:

где — предсказанные значения, — координаты клетки сетки, — размеры якоря​

* + - 1. Классификация:

Используется бинарная кросс-энтропия для предсказания классов объектов:

где — количество классов, — истинная метка, — предсказанная вероятность.

**ResNet:**

Остаточное обучение:

Каждая блок-слоя в ResNet можно описать как:

где — вход, — выход, — остаточная функция, представленная несколькими слоями с весами

Обучение:

Используется стандартный градиентный спуск с обратным распространением ошибки (SGD).

### 3.4 Описание наборов данных

Для обучения моделей YOLOv3 и ResNet часто используются следующие наборы данных:

* COCO (Common Objects in Context): Содержит более 200 тысяч изображений с аннотациями для 80 классов объектов.
* ImageNet: Огромный набор данных с более чем 14 миллионами изображений и аннотациями для 1000 классов.

### 3.5 Метрики качества

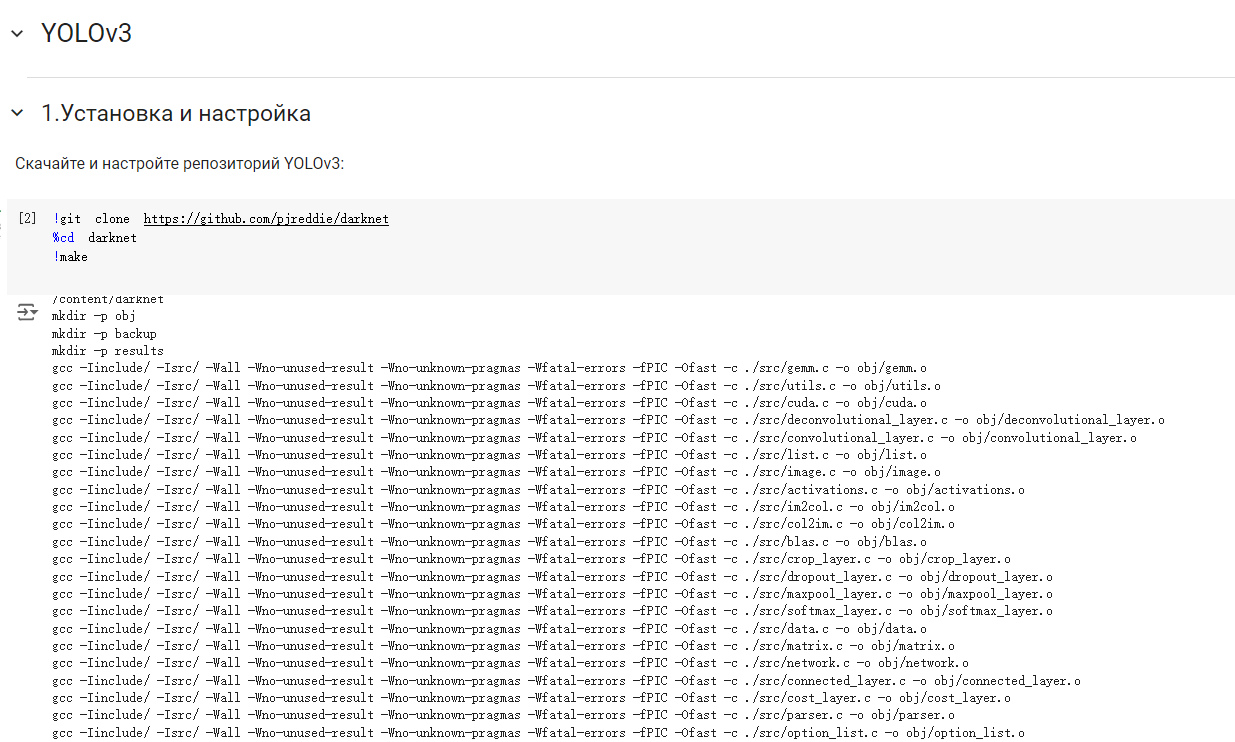
Для оценки качества решений задач обнаружения объектов используются метрики:

* mAP (mean Average Precision): Средняя точность предсказаний по всем классам. Например, в COCO используется mAP@[0.5:0.95], что включает усреднение по нескольким значениям IoU (Intersection over Union).
* IoU (Intersection over Union): Мера пересечения предсказанной рамки с истинной рамкой.

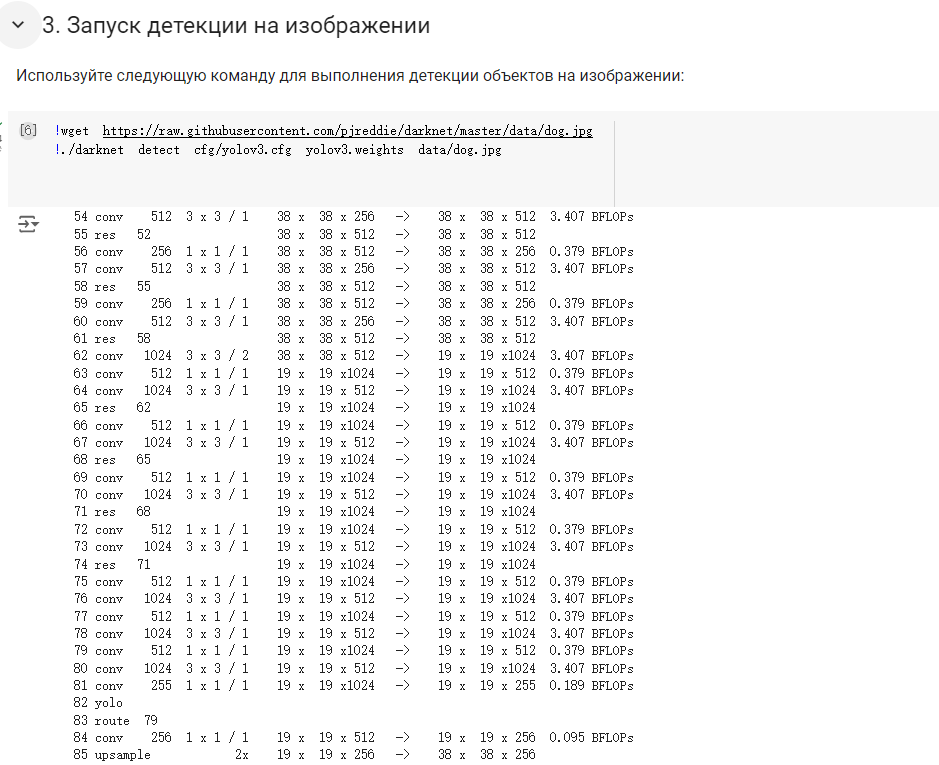
### 3.6 Предложения по улучшению качества решения задачи

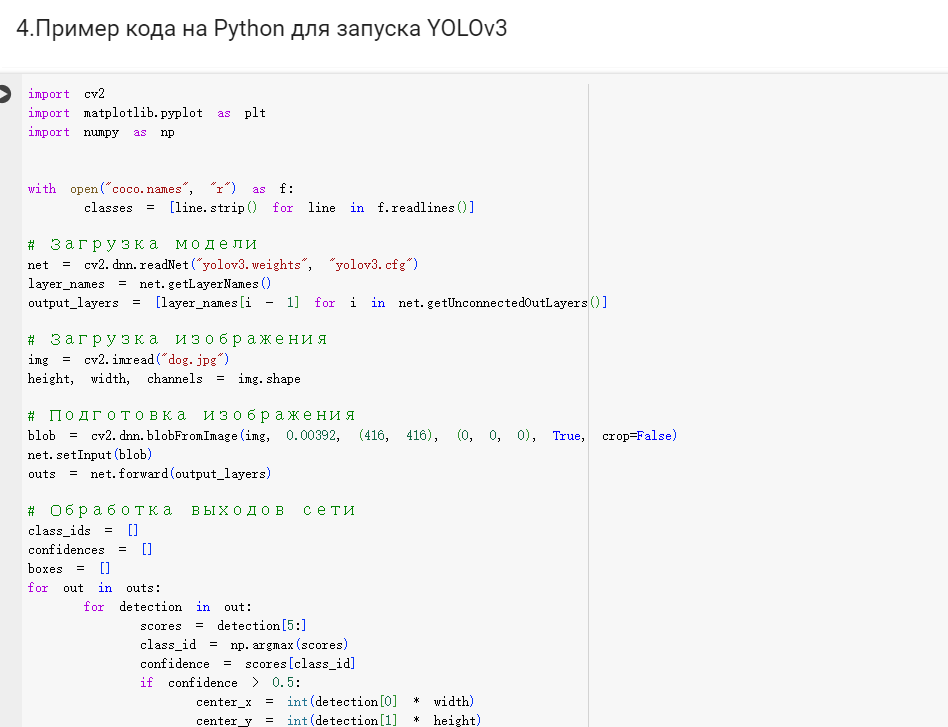
* Улучшение детекции мелких объектов: Можно исследовать методы улучшения предсказаний на нескольких масштабах, как это сделано в YOLOv3.
* Использование более сложных ансамблей моделей: Комбинирование предсказаний нескольких моделей может улучшить точность.
* Аугментация данных: Дополнительные методы аугментации данных могут улучшить обучение моделей.

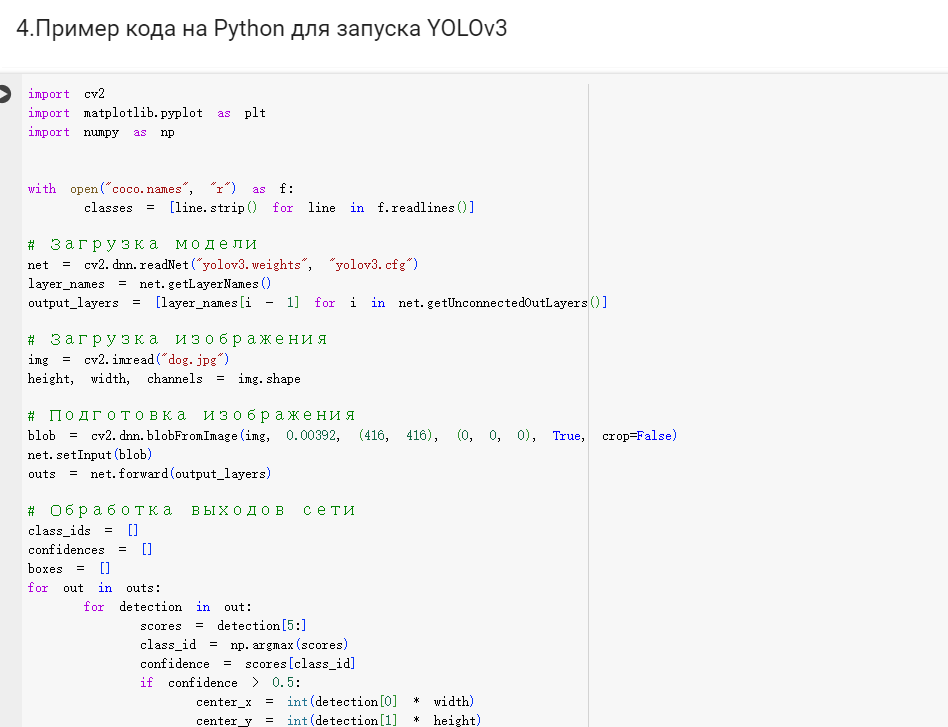
## Практический этап

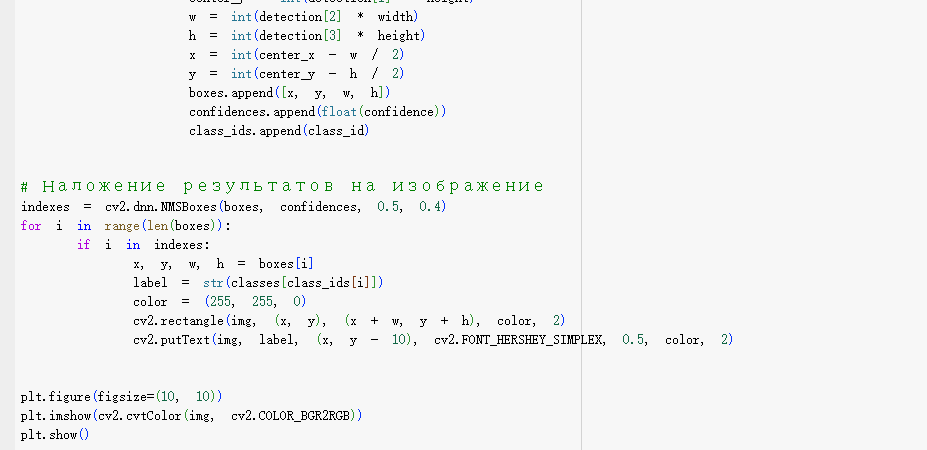


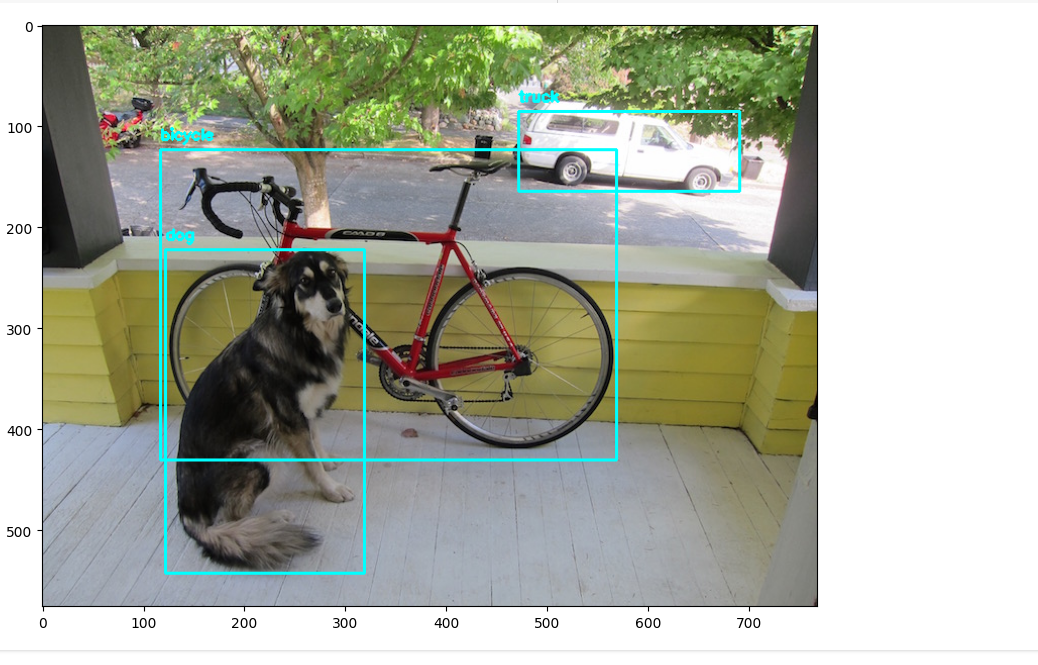


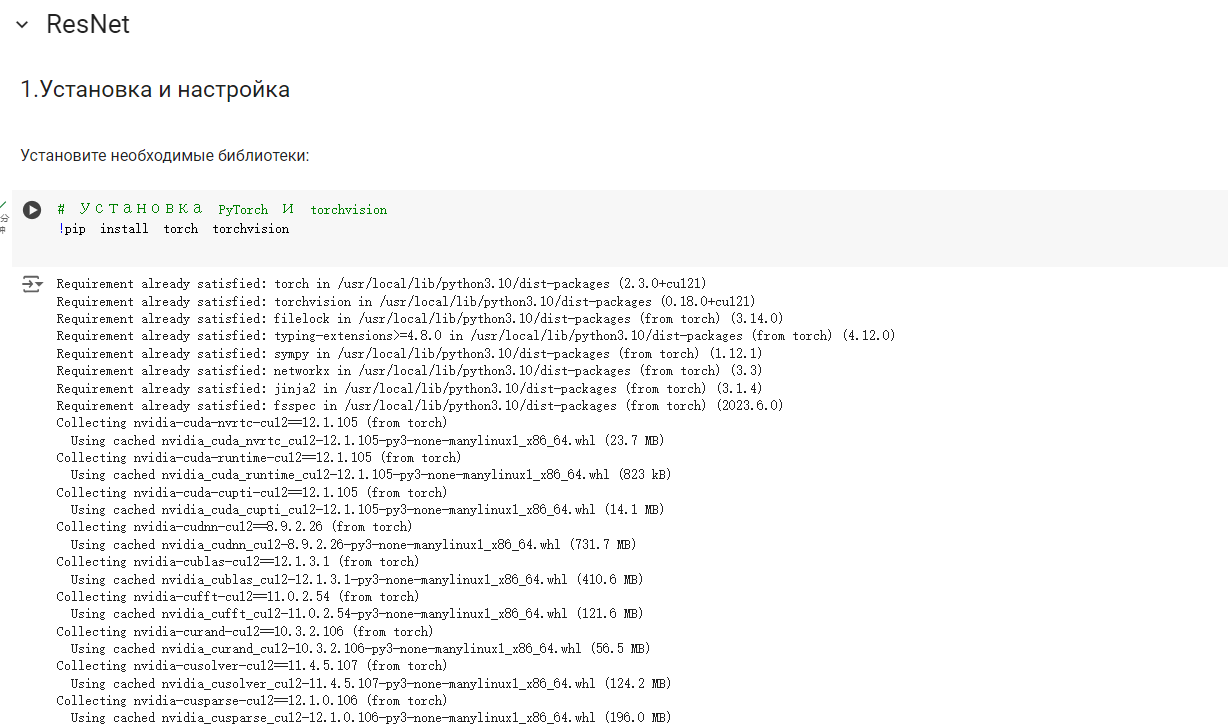




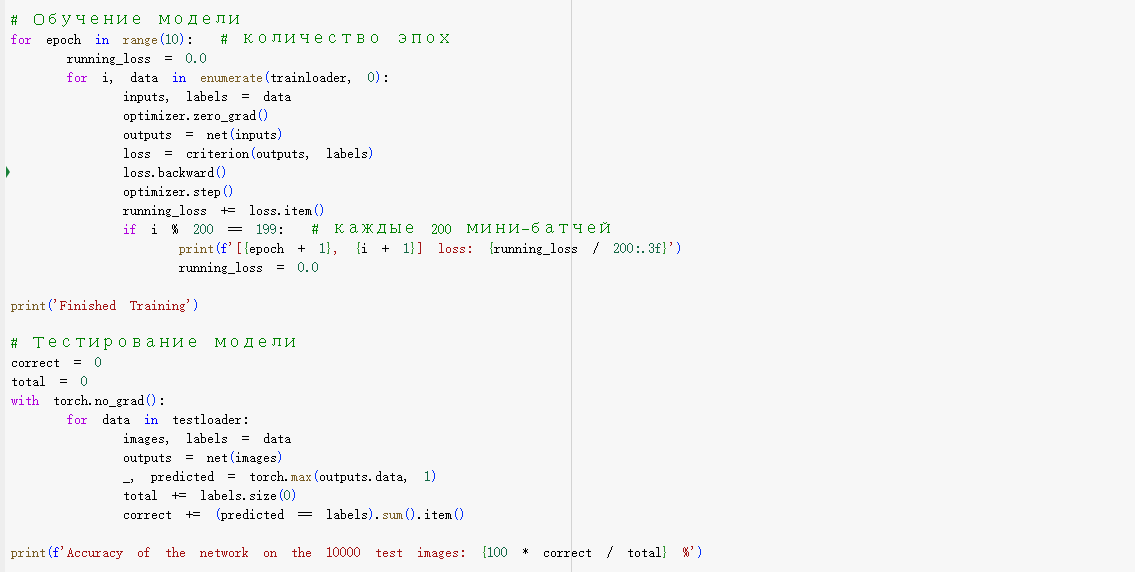


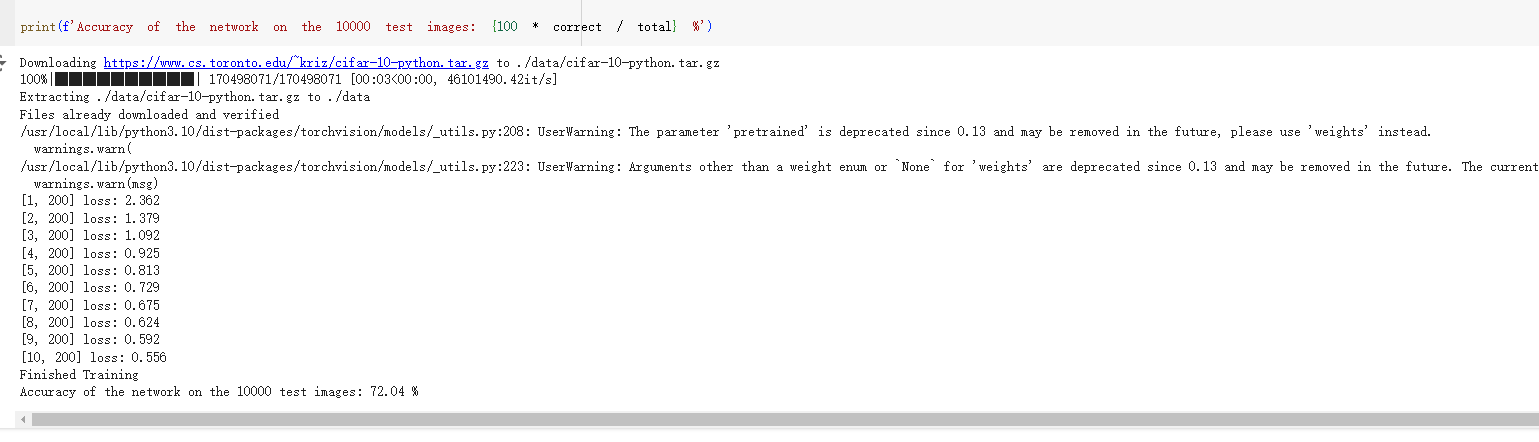












## Вывод

YOLOv3: An Incremental Improvement

YOLOv3 представляет собой современную и эффективную модель для обнаружения объектов с высокой скоростью и точностью. Основные выводы:

Высокая скорость: YOLOv3 выполняет детекцию объектов за 22 мс на изображении размером 320x320, что делает ее подходящей для реальных приложений.

Точность: YOLOv3 достигает 57.9% mAP при IOU = 0.5, сравнимо с другими методами, такими как RetinaNet, но с значительно более высокой скоростью.

Многомасштабные предсказания: Предсказания на трех масштабах улучшают детекцию мелких объектов.

Улучшенная архитектура: Использование Darknet-53 с остаточными соединениями улучшает обучение и точность модели.

Однако, YOLOv3 все еще сталкивается с трудностями в точном определении границ объектов при высоких значениях IOU.

ResNet: Deep Residual Learning for Image Recognition

ResNet решает проблемы обучения глубоких нейронных сетей, обеспечивая высокую точность и эффективность. Основные выводы:

Эффективность остаточных соединений: Остаточные соединения позволяют обучать очень глубокие сети без деградации точности.

Высокая точность: ResNet-50 достигает точности 93.8% на наборе данных CIFAR-10.

Масштабируемость: Архитектура ResNet легко масштабируется до более чем 100 слоев, оставаясь эффективной.

Практическое применение: Высокая точность и способность к обучению делают ResNet полезной для различных задач компьютерного зрения.

## Список использованных источников

[1] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.

[3] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Advances in Neural Information Processing Systems (NIPS), 91-99.

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In European Conference on Computer Vision (ECCV), 21-37.

[5] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2117-2125.

[6] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2980-2988.