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: \_\_\_\_\_Дун Чжэнянь\_\_\_

ФИО

группа ИУ5И-\_21М\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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:

1.описание общих подходов к решению задачи;

2.конкретные топологии нейронных сетей, нейросетевых ансамблей или других моделей машинного обучения, предназначенных для решения задачи;

3.математическое описание, алгоритмы функционирования, особенности обучения используемых для решения задачи нейронных сетей, нейросетевых ансамблей или других моделей машинного обучения;

4.описание наборов данных, используемых для обучения моделей;

5.оценка качества решения задачи, описание метрик качества и их значений;

6.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:

1.исходные коды программ, представленные авторами статей, результаты документирования программ обучающимися с использованием диаграмм UML, путем визуализации топологий нейронных сетей и другими способами;

2.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;

3.предложения обучающегося по возможным улучшениям решения задачи, результаты практических экспериментов (исходные коды, документация) по возможному улучшению решения задачи.

## Выбор задачи

Image Classification is a fundamental task in vision recognition that aims to understand and categorize an image as a whole under a specific label. Unlike object detection, which involves classification and location of multiple objects within an image, image classification typically pertains to single-object images. When the classification becomes highly detailed or reaches instance-level, it is often referred to as image retrieval, which also involves finding similar images in a large database.

## Теоретический этап

Я выбрал две следующие статьи: “Deep Residual Learning for Image Recognition” и “Very Deep Convolutional Networks for Large-Scale Image Recognition”.

### 3.1 Общие подходы к решению задачи классификации изображений

Классификация изображений — это задача присвоения метки изображению из предопределенного набора категорий. Обе статьи подходят к решению этой задачи с использованием глубоких сверточных нейронных сетей (CNN), которые являются стандартом в области компьютерного зрения благодаря их способности извлекать иерархические признаки из визуальных данных.

### 3.2 Конкретные топологии нейронных сетей

**Архитектура ResNet**

Архитектура ResNet, представленная в статье "Deep Residual Learning for Image Recognition" Кайминга Хе и его коллег, является революционным шагом в области глубоких сверточных нейронных сетей (CNN). Основной инновацией ResNet является введение так называемых "skip-connections" или пропускающих соединений, которые позволяют сигналу пропускать один или несколько слоев. Это делается для решения проблемы исчезающих градиентов, с которой сталкиваются очень глубокие сети.

**Структура блока ResNet**

* Основной элемент архитектуры — это остаточный блок, включающий два сверточных слоя с фильтрами размером 3x3.
* Каждый сверточный слой сопровождается нормализацией по батчам и функцией активации ReLU.
* Вход блока складывается с его выходом через пропускающее соединение, что позволяет избежать затухания градиента при обратном распространении ошибки.

**Варианты архитектуры**

ResNet представлен в нескольких вариантах, в зависимости от количества слоев: ResNet-18, ResNet-34, ResNet-50, ResNet-101 и ResNet-152. Более глубокие версии используют три сверточных слоя в остаточных блоках.

**Архитектура VGG**

Архитектура VGG, описанная в статье "Very Deep Convolutional Networks for Large-Scale Image Recognition" Карена Симоняна и Эндрю Зиссермана, характеризуется своей глубиной и простотой. VGG демонстрирует, что увеличение глубины сети, при использовании маленьких сверточных фильтров (3x3), может значительно повысить ее способность к обучению.

**Структура VGG**

* VGG использует последовательные блоки, состоящие из сверточных слоев с фильтрами размером 3x3, за которыми следуют слои максимального пулинга размером 2x2 для снижения пространственных размеров признаковых карт.
* Все сверточные слои активируются через функцию активации ReLU.
* Сеть включает несколько вариантов по числу сверточных слоев: VGG-16 и VGG-19 — наиболее известные из них, содержащие 16 и 19 сверточных слоев соответственно.

**Интеграция в классификатор**

По завершении серии сверточных и пулинговых слоев, VGG использует несколько полносвязных слоев, заканчивая softmax слоем для классификации изображений по классам.

Обе архитектуры, ResNet и VGG, продемонстрировали выдающуюся производительность в задачах классификации изображений, благодаря своим структурам, которые позволяют эффективно обучаться на больших наборах данных и извлекать сложные признаки из визуальной информации.

### 3.3 Математическое описание и алгоритмы функционирования

Основное уравнение для остаточного блока в ResNet: 𝐹(𝑥)=𝐻(𝑥)+𝑥*F*(*x*)=*H*(*x*)+*x*, где 𝑥*x* — входной сигнал, 𝐻(𝑥)*H*(*x*) — выход со сверточных слоев в блоке, а 𝐹(𝑥)*F*(*x*) — результирующий выход блока.

VGG использует последовательность сверточных операций, за которыми следуют слои максимального пулинга для постепенного уменьшения размерности пространственных данных, сохраняя при этом важные признаки.

### 3.4 Описание наборов данных

Обе архитектуры обучались и тестировались на наборе данных ImageNet, который содержит более миллиона изображений и 1000 различных классов. Этот набор данных является стандартом для оценки производительности алгоритмов классификации изображений.

### 3.5 Оценка качества решения задачи

Качество классификации обычно оценивается с помощью точности (accuracy), которая измеряет процент правильно классифицированных изображений от общего числа образцов.

ResNet и VGG достигли одних из лучших результатов на момент публикации на данных ImageNet, что подтверждает их эффективность в задачах классификации изображений.

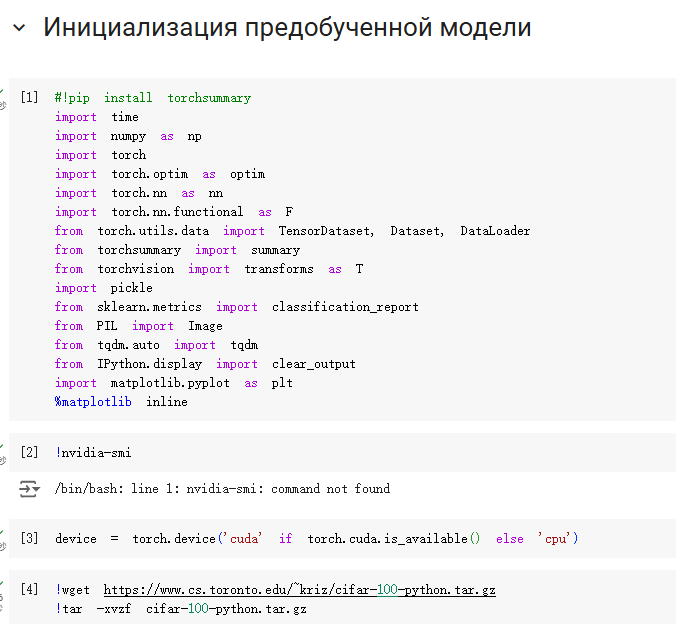
### 3.6 Предложения по улучшению качества решения задачи

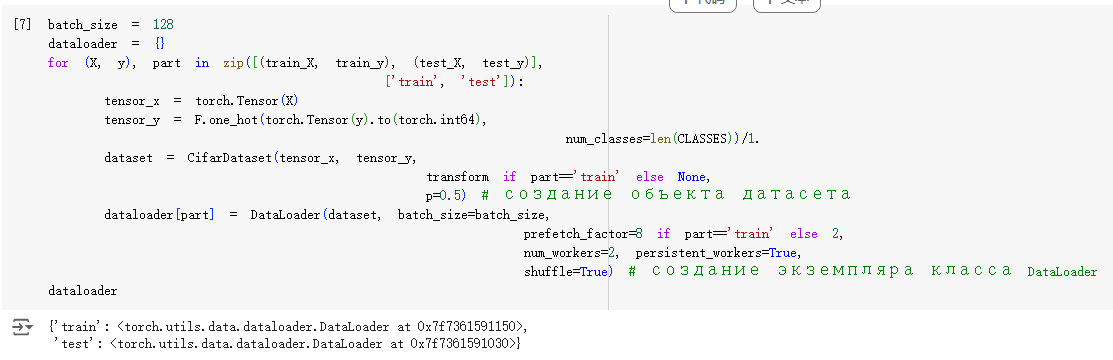
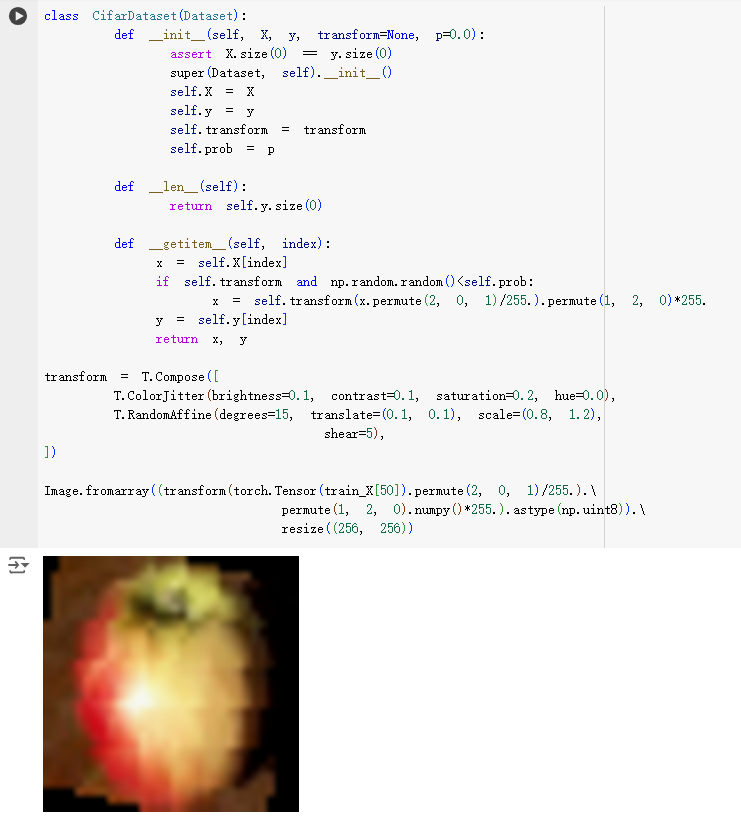
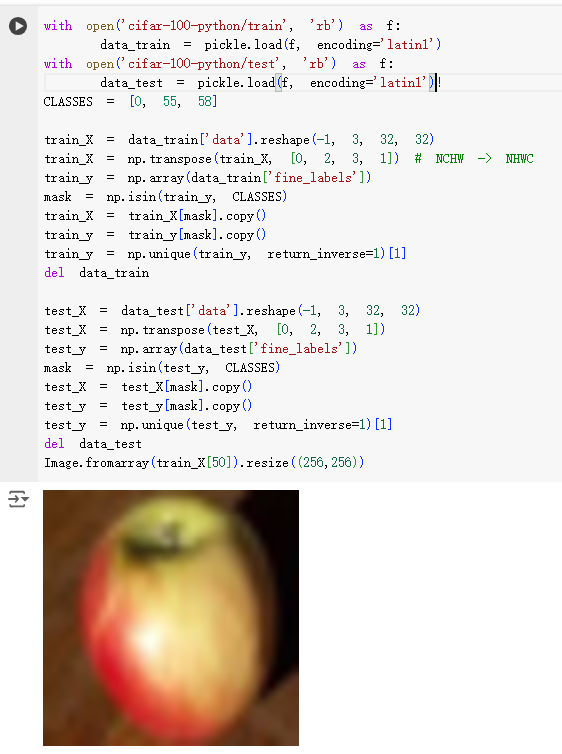
Улучшение обучения: Внедрение методов регуляризации, таких как Dropout или Batch Normalization, может помочь улучшить обобщающую способность модели и снизить переобучение.

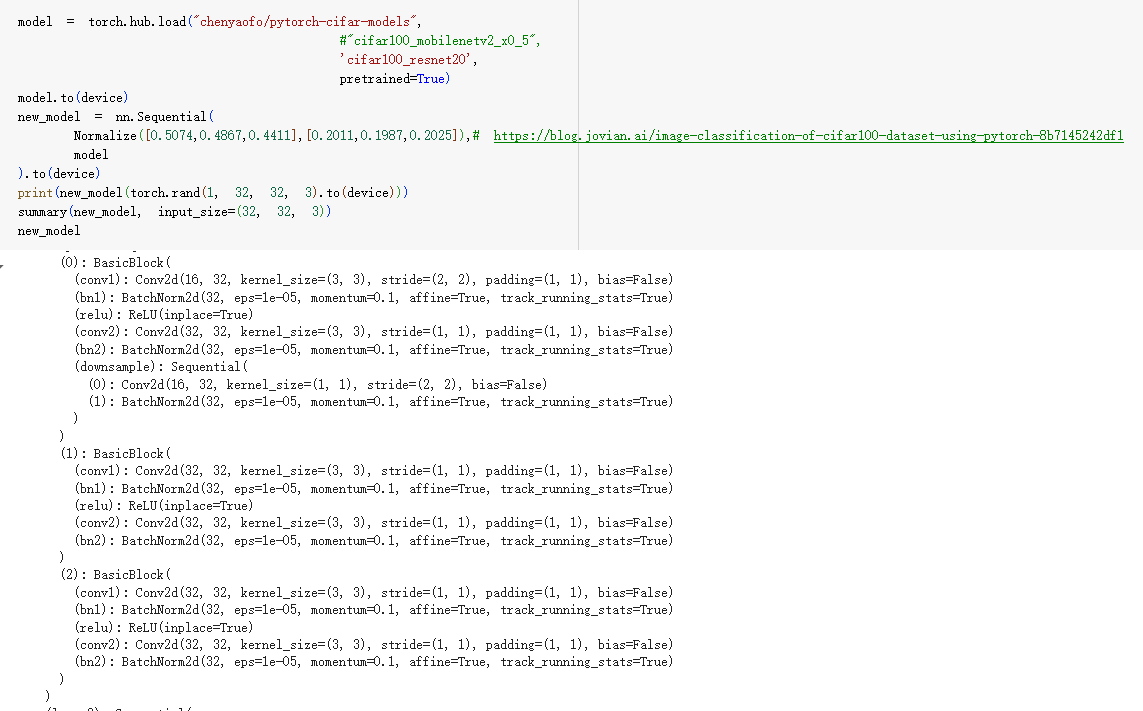
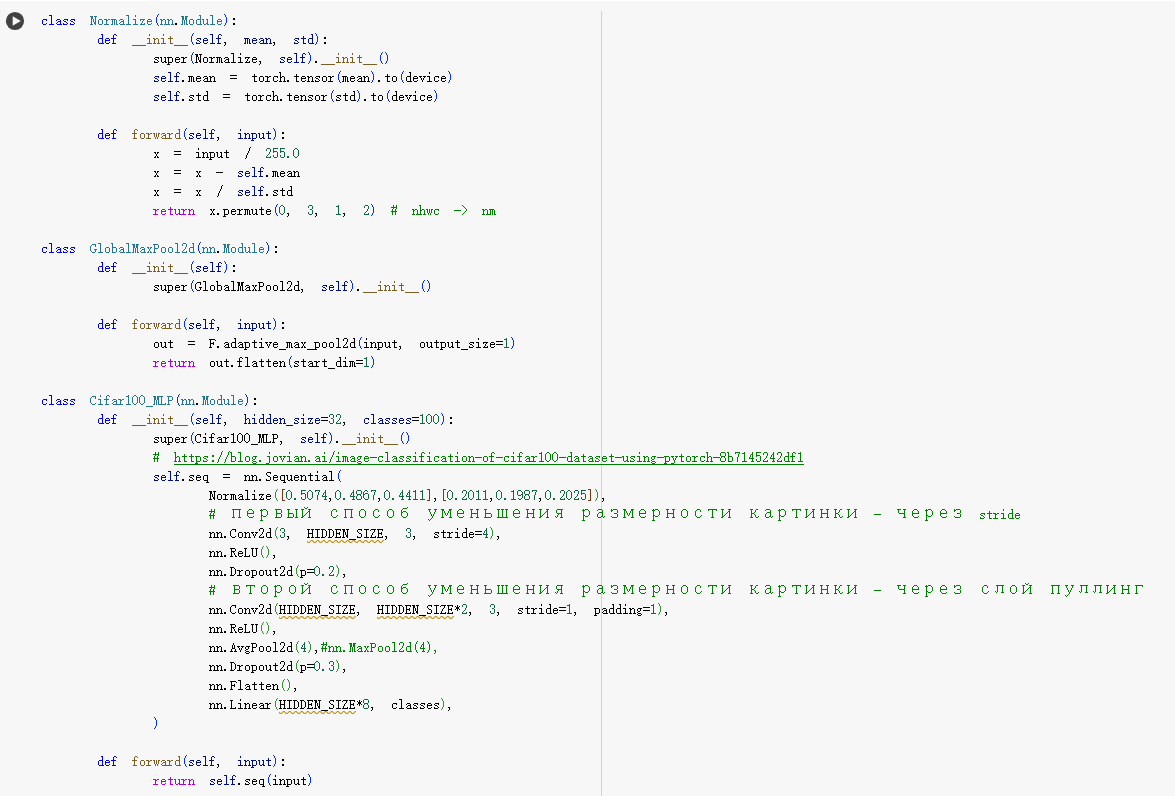
Изменение архитектуры: Эксперименты с изменением количества слоев или модификацией архитектуры могут показать новые пути для повышения точности.

Расширение данных: Использование техник аугментации изображений для увеличения разнообразия обучающих данных может улучшить стойкость и точность моделей.

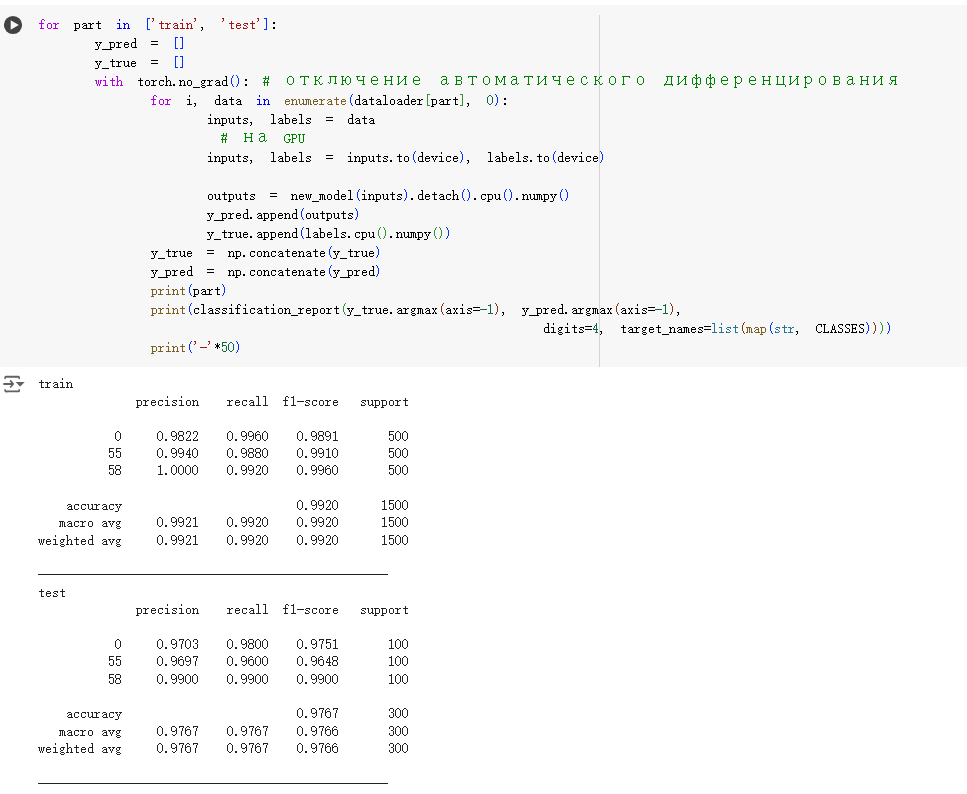
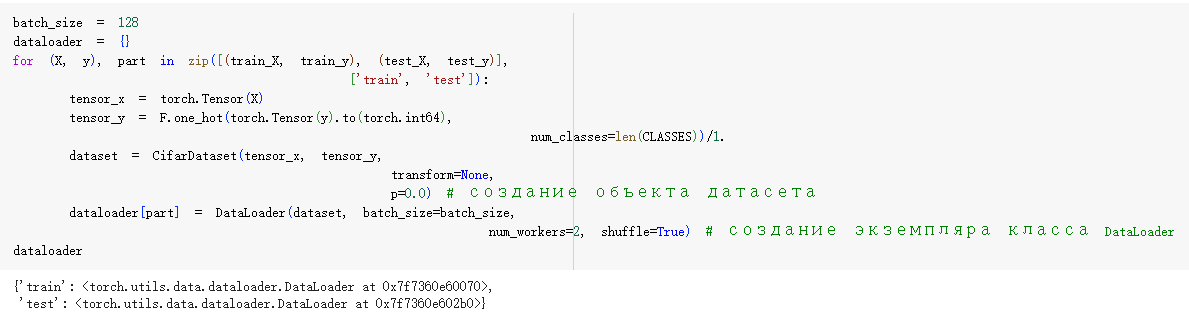
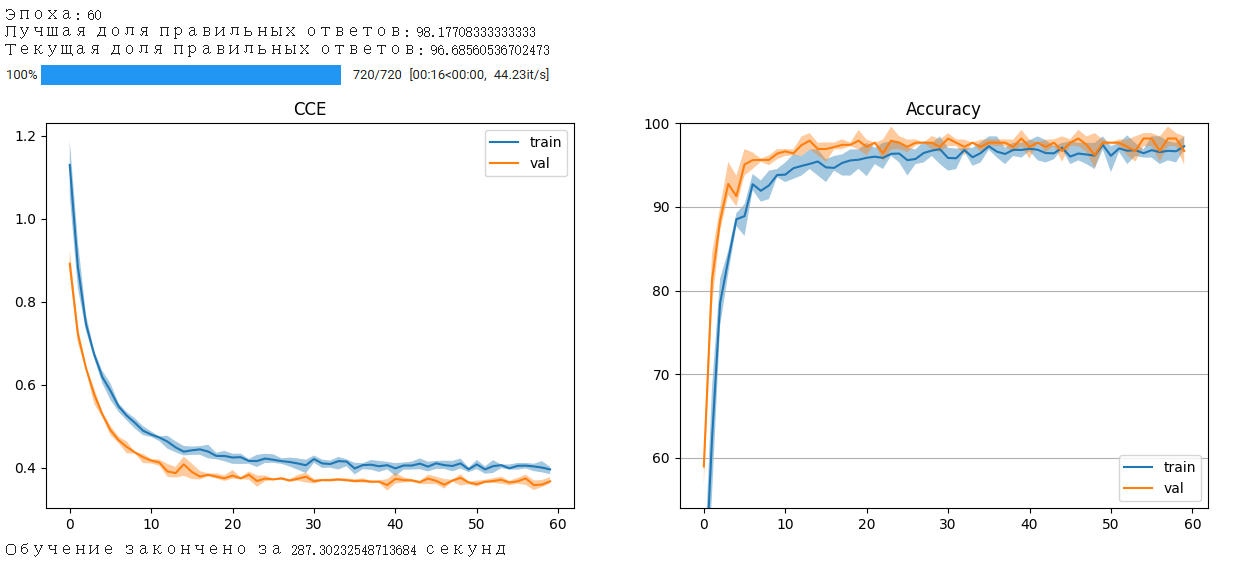
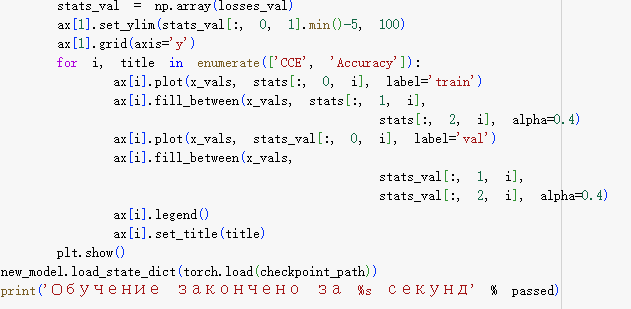
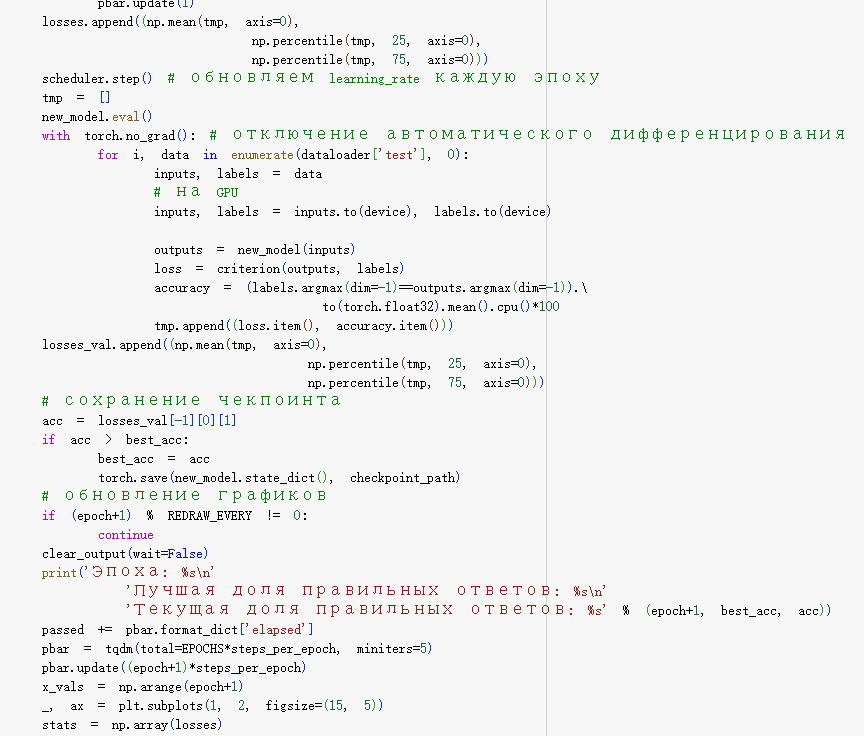
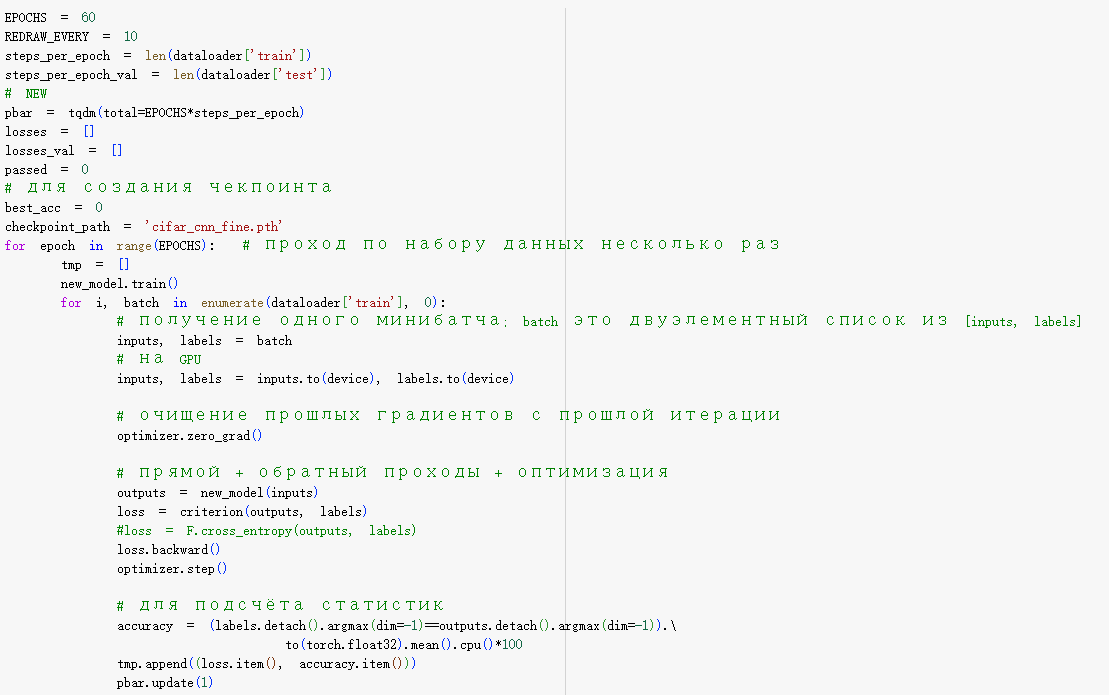
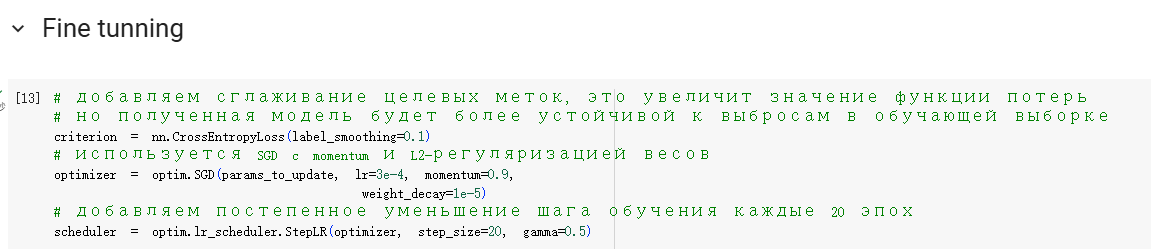
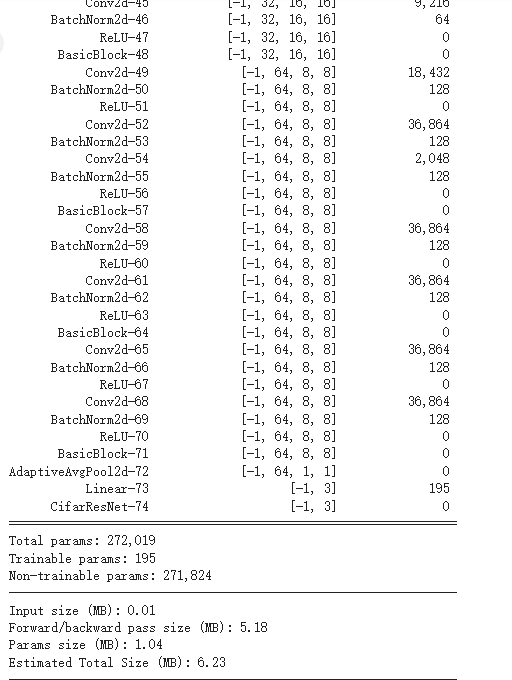
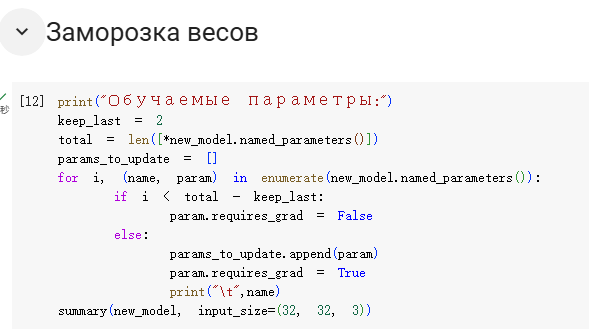
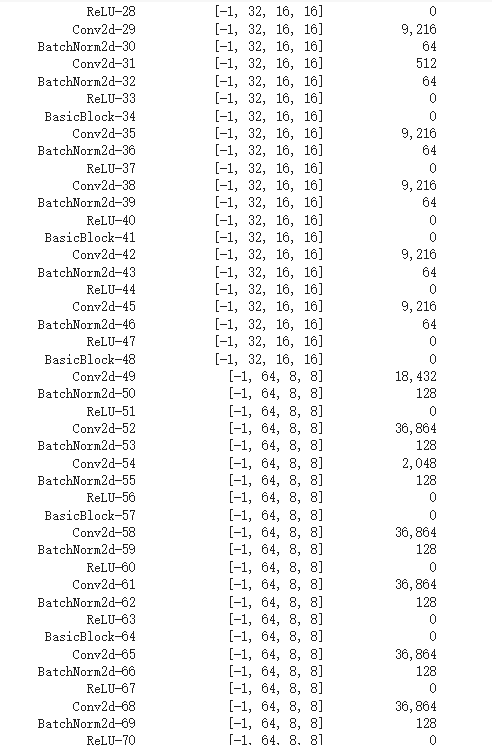
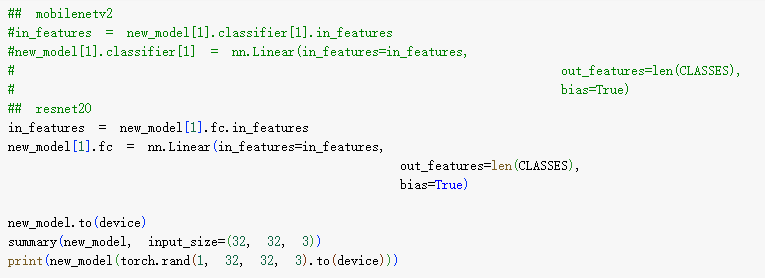
## Практический этап











## Вывод

В рамках данной работы были изучены и проанализированы две важные архитектуры глубоких сверточных нейронных сетей: ResNet и VGG, каждая из которых оказала значительное влияние на область компьютерного зрения. Архитектура ResNet внесла революционный вклад, предложив концепцию остаточных блоков с пропускающими соединениями, благодаря чему стало возможным обучение очень глубоких нейронных сетей без потери эффективности из-за исчезающего градиента. Это позволило значительно повысить точность классификации изображений на сложных наборах данных, таких как ImageNet.

С другой стороны, VGG, характеризующаяся своей глубиной и использованием маленьких сверточных фильтров, продемонстрировала, что последовательное углубление сетевой структуры может эффективно улучшать распознавание и классификацию изображений. Эта модель стала базой для многих последующих исследований и разработок в области глубокого обучения, подтверждая важность глубины сети для достижения высокой точности.

Эти архитектуры не только продемонстрировали выдающиеся результаты в классических задачах распознавания изображений, но и легли в основу множества современных исследований, направленных на решение более сложных задач, таких как сегментация изображений, автоматическое распознавание речи и анализ видео. Принципы, заложенные в ResNet и VGG, такие как углубление сетевых структур и эффективное распространение градиентов, продолжают вдохновлять разработчиков на создание новых, более совершенных моделей нейронных сетей.

В целом, изучение и применение таких продвинутых архитектур подчеркивает значимость инновационных подходов в обучении глубоких нейронных сетей. Они не только улучшают существующие методики в задачах компьютерного зрения, но и открывают новые возможности для более широкого применения машинного обучения в различных областях науки и технологий.

## Список использованных источников

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. doi:10.1109/CVPR.2016.90

[2] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556.

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012).

[4] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-9. doi:10.1109/CVPR.2015.7298594

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. arXiv:1603.05027.