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: «Машинный перевод»

ИСПОЛНИТЕЛЬ: \_\_\_\_\_Хэ Синьчэнь\_\_\_

ФИО

группа ИУ5И-24М\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_22\_"\_\_\_\_\_05\_\_\_\_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.

## Этап выбора мандата.Выбрать задачу "Классификация изображений"

**2.1 Классификация изображений**

Классификация изображений - это фундаментальная задача компьютерного зрения, целью которой является отнесение входного изображения к определенной категории. Эта задача находит широкое применение во многих областях, таких как автономное вождение, анализ медицинских изображений, безопасность и наблюдение, а также "умный дом".

**2.2 Типичные сценарии применения**

Автоматическое вождение: идентификация различных объектов на дороге, таких как пешеходы, транспортные средства, дорожные знаки и т. д.

Анализ медицинских изображений: идентификация и классификация повреждений на медицинских изображениях, например, классификация опухолей, обнаружение повреждений и т. д.

Видеонаблюдение: обнаружение и классификация подозрительных действий и объектов в системах видеонаблюдения.

Умный дом: идентификация различных объектов и людей в доме для интеллектуального управления и взаимодействия.

**2.3 Основные этапы классификации изображений**

Сбор данных и предварительная обработка:

Сбор данных: сбор большого количества наборов данных маркированных изображений, например CIFAR-10, ImageNet и т. д.

Предварительная обработка данных: нормализация, нормализация, улучшение данных и другие операции выполняются с изображениями для улучшения обобщающей способности модели.

Выбор и обучение модели:

Выбор подходящих моделей: обычно используются такие модели, как конволюционная нейронная сеть (CNN), остаточная сеть (ResNet) и модели миграционного обучения (например, VGG, Inception).

Обучение модели: обучение модели с использованием набора обучающих данных и оптимизация параметров модели для минимизации функции потерь.

Оценка и оптимизация модели:

Оценка модели: оценка производительности модели на валидационном и тестовом наборах данных; общие метрики оценки включают точность, прецизионность, отзыв, F1 score и т. д.

Оптимизация модели: оптимизация производительности модели путем корректировки гиперпараметров, архитектуры модели, стратегии обучения и т. д.

## Теоретический часть отчета

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:

**3.1 Ранние методы классификации изображений**

Ранние методы классификации изображений в основном опираются на ручное извлечение признаков и традиционные алгоритмы машинного обучения.

Ручное извлечение признаков: часто используемые методы извлечения признаков включают SIFT (Scale-Invariant Feature Transform) и HOG (Histogram of Oriented Gradients). Эти методы извлекают признаки путем обнаружения ключевых точек и дескрипторов на изображении, а затем используют эти признаки для классификации.

SIFT: Lowe (2004) предложил SIFT, который способен извлекать точки признаков с инвариантностью к масштабу и вращению и широко используется для сопоставления изображений и распознавания целей.

HOG: Далал и Триггс (2005) предложили алгоритм HOG, который используется для получения информации о форме локальных объектов на изображении и хорошо зарекомендовал себя в таких задачах, как обнаружение пешеходов.

Традиционные алгоритмы машинного обучения: используют такие классификаторы, как машины опорных векторов (SVM), k-Nearest Neighbours (k-NN) и Random Forest. Эти методы достаточно хорошо работают на небольших массивах данных, но имеют ограниченную производительность на больших массивах данных и в сложных задачах.

**3.2 Расцвет методов глубокого обучения**

Глубокое обучение - это класс методов машинного обучения, основанных на искусственных нейронных сетях с несколькими скрытыми слоями, которые автоматически извлекают и изучают высокоуровневые характеристики данных. Глубокое обучение отлично справляется с обработкой высокоразмерных данных, таких как изображения, речь и текст. Первые нейронные сети развивались медленно из-за ограничений по вычислительной мощности и объему данных. Однако с увеличением вычислительной мощности и доступности масштабных наборов данных глубокое обучение стало развиваться стремительно.

LeNet-5 (1998): предложенная Янном Лекуном и др. для распознавания рукописных цифр, она является шедевром ранних CNN. LeNet-5 содержит два конволюционных слоя и два слоя поддискретизации и была успешно применена для распознавания рукописных цифр.

AlexNet (2012): предложенная Алексом Крижевским и др., она впервые значительно превзошла традиционные методы в конкурсе ImageNet, вызвав бум глубокого обучения. alexNet использует большее количество конволюционных слоев, функцию активации ReLU и метод регуляризации Dropout, что значительно улучшает производительность классификации. Метод AlexNet (2012) внес значительный вклад в развитие глубокого обучения. AlexNet победил в конкурсе ImageNet Large Scale Visual Recognition Challenge (ILSVRC), значительно превзойдя традиционные методы и вызвав бум глубокого обучения. Сеть состоит из восьми слоев, включая пять конволюционных и три полностью связанных слоя, и использует функцию активации ReLU, регуляризацию Dropout и методы расширения данных.

Инновации:

Функция активации ReLU: решает проблему исчезновения градиента и ускоряет процесс обучения.

Регуляризация Dropout: уменьшает избыточную подгонку и улучшает обобщение.

Расширение данных: увеличивает разнообразие обучающих данных и повышает устойчивость модели с помощью методов расширения данных.

Влияние: Успех AlexNet доказал мощную производительность глубоких нейронных сетей на больших наборах данных, что побудило исследователей и промышленников начать широко применять и изучать методы глубокого обучения.

**3.3 Глубокие сетевые структуры**

С ростом вычислительной мощности и увеличением объемов данных исследователи начали изучать более глубокие структуры конволюционных нейронных сетей (CNN). Эти глубокие сети продемонстрировали чрезвычайно высокую точность в задачах классификации изображений и способствовали быстрому развитию области компьютерного зрения. Ниже приведены несколько репрезентативных структур глубоких сетей и их подробные описания.

VGGNet (2014). Предложена Кареном Симоняном и Эндрю Зиссерманом. VGGNet предлагает архитектуру глубокой сети путем суммирования небольших конволюционных ядер (3x3) и достигла отличных результатов в ImageNet Image Classification Challenge. Основной вклад VGGNet заключается в том, что увеличение глубины сети значительно улучшает производительность модели. продемонстрировала важность глубоких сетей в сложных визуальных задачах благодаря своей простой, но эффективной архитектуре, и стала моделью, лежащей в основе многих последующих исследований. Архитектура сети

Модуль Inception: объединяет конволюционные ядра 1x1, 3x3 и 5x5 и операции максимального объединения для объединения признаков в разных масштабах.

Глубина сети: содержит 22 глубоких сверточных и пулинговых слоя, что позволяет увеличить ширину и глубину сети за счет модуля Inception, контролируя при этом вычислительную сложность.

Свертка 1x1: используется для уменьшения размерности карты признаков и снижения вычислительных затрат.

**3.4 Облегченные сети и миграционное обучение**

Популярность глубокого обучения, потребность в мобильных и встроенных устройствах с ограниченными вычислительными ресурсами подтолкнули исследователей к разработке облегченных сетей, а развитие методов обучения с переносом значительно сократило объем данных и вычислительных затрат, необходимых для обучения модели. Облегченные сети направлены на снижение требований к вычислениям и хранению данных при сохранении производительности модели.MobileNet представляет собой облегченную сеть, которая значительно снижает вычислительную сложность за счет внедрения свертки с разделением по глубине, которая декомпозирует стандартную свертку на глубокую свертку и свертку по точкам.Дизайн MobileNet делает ее хорошо подходящей для приложений в средах с ограниченными ресурсами, таких как смартфоны и устройства Интернета вещей (IoT). . Миграционное обучение, с другой стороны, значительно повышает производительность моделей на небольших наборах данных за счет предварительного обучения моделей на крупномасштабных наборах данных и последующей их тонкой настройки на конкретные задачи. Основной процесс обучения переносу заключается в том, что сначала глубокая сеть предварительно обучается на крупномасштабном наборе данных (например, ImageNet), а затем применяет предварительно обученную модель к целевой задаче, тонко настраивая модель для адаптации к новому распределению данных. Успех обучения с переносом объясняется тем, что глубокие сети на крупномасштабных наборах данных изучают общие характеристики, которые очень эффективны во многих визуальных задачах. Методы обучения с переносом, основанные на таких моделях, как ResNet и VGG, значительно сократили объем данных и вычислительных ресурсов, необходимых для обучения сложных моделей, открыв возможность широкого применения глубокого обучения в различных областях. Сочетание этих легких сетей и методов миграционного обучения способствовало быстрому развитию глубокого обучения в мобильных устройствах, анализе медицинских изображений, автономном вождении и многих других реальных приложениях, значительно расширив спектр применений и сценариев глубокого обучения.

**3.5 Последние события**

В последние годы последние разработки в области глубокого обучения для классификации изображений были сосредоточены на эффективном проектировании сетей и новых архитектурных парадигмах. EfficientNet - одно из важных достижений, которое позволяет достичь оптимального баланса между производительностью модели и вычислительной эффективностью, предлагая композитный подход к масштабированию для систематического проектирования сетевых архитектур. Одновременно регулируя глубину, ширину и разрешение сети, EfficientNet создает серию моделей от мала до велика, адаптируясь к различным вычислительным ресурсам и сценариям применения, значительно повышая эффективность параметров и точность модели. Кроме того, Vision Transformers (ViT) представляет собой новую архитектурную парадигму, которая успешно применяет архитектуру Transformer к задаче классификации изображений, сегментируя изображение на блоки фиксированного размера и линейно встраивая их в качестве входных данных в Transformer, что позволяет осуществлять сквозную классификацию изображений. Трансформер обладает большим потенциалом в компьютерном зрении, его архитектура основана на механизме самовнушения, который не зависит от сверточных операций, и демонстрирует производительность, конкурентоспособную со сверточными нейронными сетями, в широком спектре задач классификации изображений. Эти последние разработки не только совершили прорыв в академических исследованиях, но и предоставили более эффективные и гибкие решения для практических приложений, способствуя дальнейшему прогрессу и широкому применению технологии классификации изображений.

Недавний интерес к носимым устройствам с дополненной реальностью, устройствам домашней автоматизации и самодвижущимся автомобилям вызвал острую потребность в алгоритмах семантической сегментации (или визуального понимания сцены). которые могут работать в режиме реального времени на мобильных устройствах с низким энергопотреблением. Эти алгоритмы маркируют каждый пиксель на изображении с одним из классов объектов. В последние годы доступность больших наборов данных и мощные вычислительные машины помогли глубоким сверточным нейронным сетям (CNN) превзойти производительность многих обычных алгоритмов компьютерного зрения. Даже Несмотря на то, что CNN все успешнее справляются с задачами классификации и категоризации, они дают грубые пространственные результаты при применении к попиксельной маркировке изображений. Поэтому их часто объединяют в каскад с другими алгоритмами для уточнения результатов, такими как сегментация на основе цветаили условные случайные поля, чтобы назвать другие алгоритмы. поля, и др.

Для пространственной классификации и тонкой сегментации изображений было предложено несколько архитектур нейронных сетей, таких как SegNet или полностью сверточные сети. были предложены такие архитектуры, как SegNet или полностью конволюционные сети. Все эти работы основаны на архитектуре VGG16, которая является очень большой моделью, предназначенной для многоклассовой классификации. В этих работах предлагаются сети с огромным количеством параметров и длительным временем вычисления. время вывода. В таких условиях они становятся непригодными для многих мобильных или работающих от аккумуляторов приложений, которые требуют обработки изображений со скоростью более 10 кадров в секунду.

В данной работе мы предлагаем новую архитектуру нейронной сети, оптимизированную для быстрого вывода и высокой точности. Примеры изображений, сегментированных с помощью ENet, показаны на рисунке 1. В нашей работе мы решили не использовать никаких шагов постобработки, которые, конечно, могут быть объединены с нашим методом, но это может ухудшат производительность сквозного подхода CNN.

Архитектура сети

表格

描述已自动生成

Архитектура нашей сети представлена в таблице 1. Она разделена на несколько этапов, выделенных горизонтальными линиями в таблице и первой цифрой после названия каждого блока. Размеры выходных данных приведены для входного изображения с разрешением 512 × 512. Мы придерживаемся взгляда на ResNets, который описывает Они имеют одну главную ветвь и расширения с конволюционными фильтрами, которые отделяются от нее, а затем объединяются обратно с помощью поэлементного сложения, как показано на рисунке 2b. Каждый блок состоит из трех сверточных слоев: проекции 1 × 1, которая уменьшает размерность, основного сверточного слоя (conv на рисунке 2b) и расширения 1 × 1. Мы помещаем пакетную нормализациюи PReLU между всеми свертками. Как и в оригинальной статье, мы называем их "узкими местами". модули. Если узким местом является понижающая выборка, то к основной ветви добавляется слой максимального пула

Кроме того, первая проекция 1 × 1 заменяется на сверткой 2 × 2 с шагом 2 в обоих измерениях. Мы обнуляем активации, чтобы соответствовать количество карт признаков. Свертка - это либо обычная, расширенная или полная свертка (также известная как как деконволюция или дробная свертка) с фильтрами 3 × 3. Иногда мы заменяем ее асимметричной сверткой, т.е. последовательностью из 5 × 1 и 1 × 5 сверток. В качестве регуляризатора мы используем Spatial Dropout, с p = 0.01 до узкого места2.0 и p = 0.1 после него.

Начальный этап содержит один блок, который представлен на рисунке 2a. Стадия 1 состоит из 5 узких блоков, в то время как стадии 2 и 3 имеют такую же структуру, за исключением того, что этап 3 не понижает дискретизацию входного сигнала в начале (мы опускаем 0-е узкое место). Эти три первых являются кодером. Этапы 4 и 5 относятся к декодеру.

Мы не использовали условия смещения ни в одной из проекций, чтобы сократить количество вызовов ядра и операций с памятью. вызовов ядра и общих операций с памятью, поскольку cuDNN использует отдельные ядра для свертки и добавления смещения. Этот выбор не оказал никакого влияния на точность. Между каждым сверточным слоем и последующей нелинейностью мы используем Пакетную нормализацию. В декодере макс. pooling заменяется на max unpooling, а padding заменяется на пространственную свертку без смещения. Мы не использовали индексы пулинга в последнем модуле апсемплинга, поскольку исходный блок оперировал на 3 каналах входного кадра, в то время как конечный выход имеет C карт признаков (количество классов объектов). классов). Также, по соображениям производительности, мы решили поместить в качестве последнего модуля сети только "голую" полную свертку. Также из соображений производительности мы решили поместить в качестве последнего модуля сети только голую полную свертку, которая сама по себе занимает значительную часть времени обработки декодера.

图示, 示意图

描述已自动生成

1. **Модели и алгоритмы, используемые в распознавании изображений**

Распознавание изображений - это процесс анализа и интерпретации визуальной информации с целью классификации объектов, выделения признаков или извлечения информации. Для эффективного распознавания изображений используются различные модели и алгоритмы, которые можно разделить на несколько категорий: традиционные методы обработки изображений, классические алгоритмы машинного обучения и современные методы глубокого обучения.

Традиционные методы обработки изображений

Эти методы включают в себя различные техники предобработки изображений и извлечения признаков:

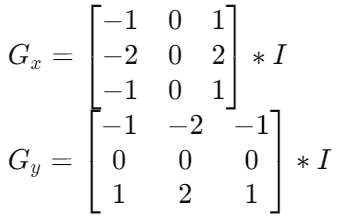
Фильтрация и преобразования: Использование различных фильтров, таких как гауссовые и медианные фильтры, для удаления шума и улучшения качества изображения.

Преобразование Фурье и Хаара: Преобразования, которые позволяют выделить частотные компоненты изображения и используются для анализа текстур и структур. Формула для прямого преобразования Фурье:



Современные методы глубокого обучения

С развитием вычислительных мощностей и наличием больших объемов данных, глубокое обучение стало основным инструментом для распознавания изображений. Основные модели и алгоритмы глубокого обучения включают в себя:



Сверточные нейронные сети (CNN): CNN являются основой для большинства задач распознавания изображений. Эти сети состоят из слоев свертки и подвыборки (pooling), которые позволяют автоматически выделять иерархические признаки из изображений. Формула свертки:

 где I - входное изображение, K - ядро свертки.

Резидуальные сети (ResNet): Модели, которые используют резидуальные блоки с прямыми связями для решения проблемы затухания градиента при обучении глубоких сетей, что позволяет создавать сверхглубокие сети. Формула резидуального блока:

 где остаточная функция, представляющая собой комбинацию слоев свертки, x - входной сигнал.

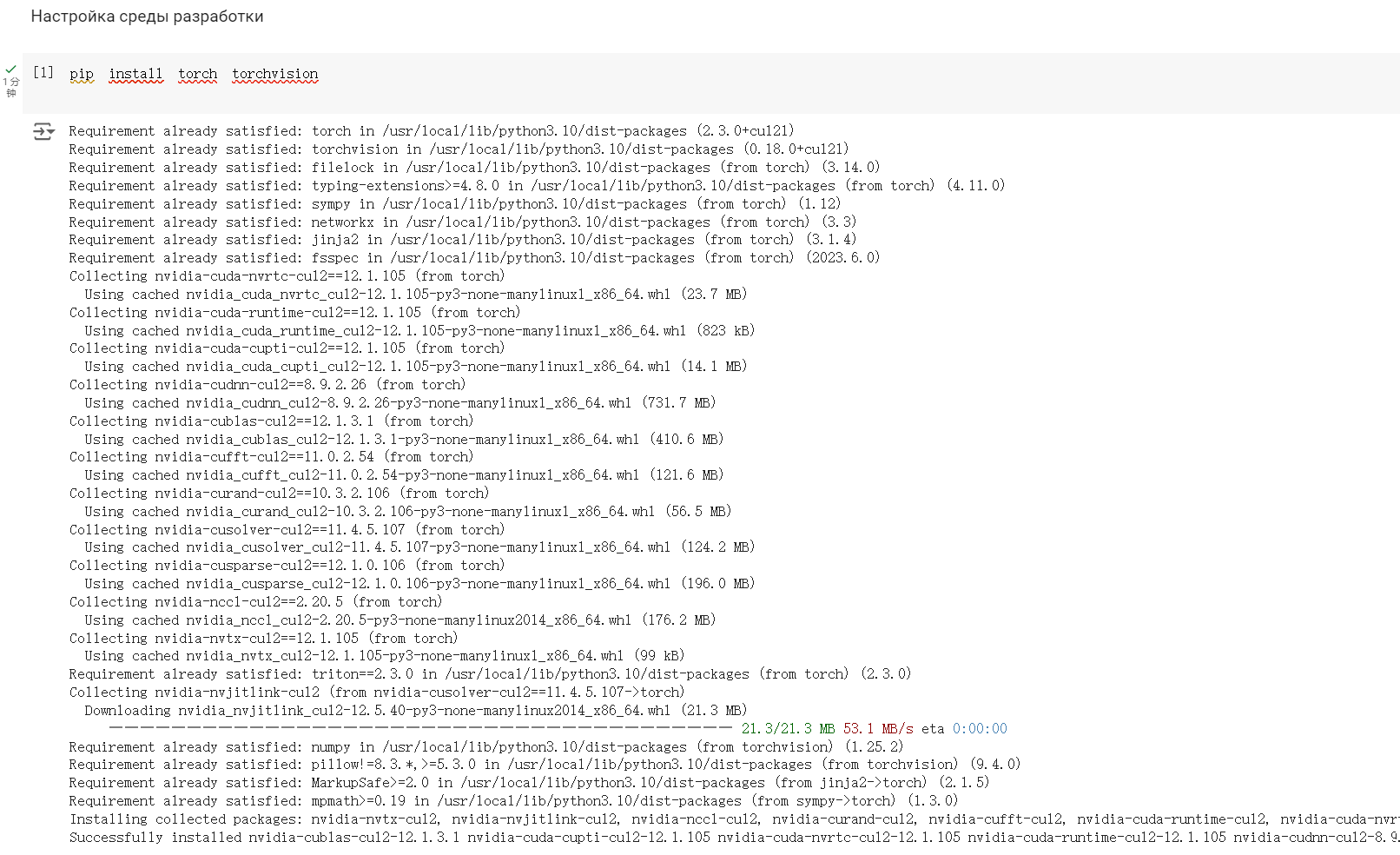
Плотно связанные сети (DenseNet): Эти сети используют плотные соединения, при которых каждый слой соединен со всеми предыдущими слоями, что улучшает передачу информации и градиентов в сети. Формула плотного соединения:

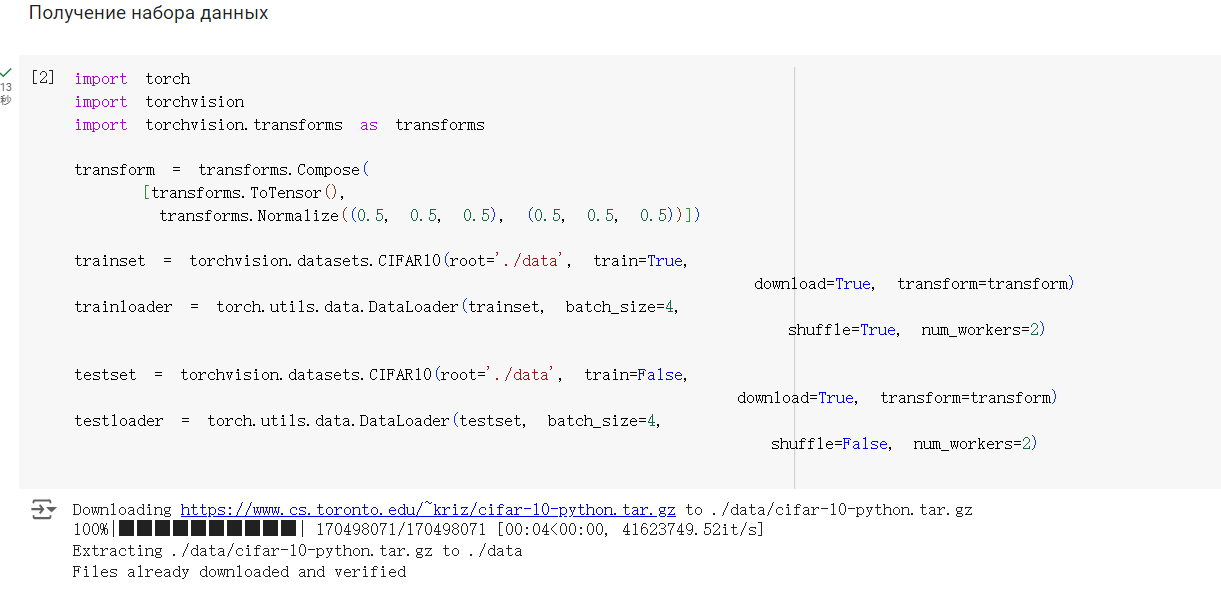
где [x0,x1​,...,xl−1] - конкатенация выходов всех предыдущих слоев, Hl - нелинейное преобразование на уровне l.

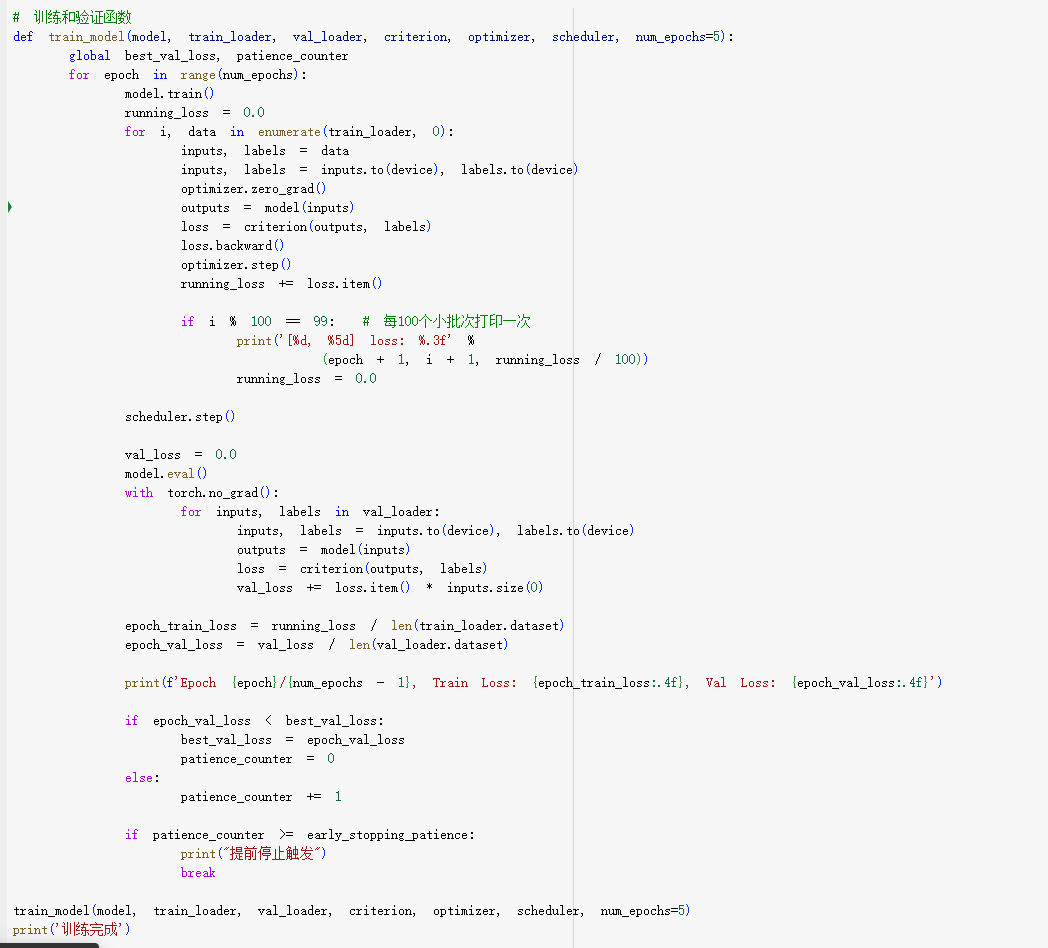
Мобильные сети (MobileNet): Легковесные архитектуры, оптимизированные для работы на мобильных устройствах и встроенных системах, которые используют глубинно-разделимые свертки для уменьшения вычислительной сложности. Формула глубинно-разделимой свертки:  где Kdepth​- глубинное сверточное ядро, Kpoint - сверточное ядро размером 1x1.

Эти модели и алгоритмы позволяют решать широкий спектр задач в распознавании изображений, включая классификацию, сегментацию, детекцию объектов и многое другое, обеспечивая высокую точность и эффективность.

1. **Практическая часть**







## Список использованных источников

1. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE.

[2]Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. NIPS.

[3]Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[4]Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going deeper with convolutions. CVPR.

[5]He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR.

[6]Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. CVPR.

[7]Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[8]Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. ICML.

[9]Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.