Сравнение точности распознавания сцен и производительности свёрточных нейронных сетей

И.А. Килбас[1], Р.А. Парингер[1]

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. В настоящее время задача классификации изображений становится всё более актуальной. Одним из наиболее популярных решений являются свёрточные нейронные сети. Но эффективность нейронных сетей имеет свою цену - они требуют больших ресурсов для обучения. Не всегда возможно обучить собственную нейронную сеть, но даже в этой ситуации есть выход — использовать предобученную нейронную сеть. В данном исследовании мы рассмотрим ряд предобученных свёрточных нейронных сетей: сравним их время работы, точность, а также потребляемую память.

1. Введение

1.1 Предыстория

Машинное обучение — относительно молодая область науки, зародившаяся в 50х годах прошлого века. Цель машинного обучения — создание методов и алгоритмов, посредством которых компьютеры обучаются решению тех или иных проблем без вмешательства человека или, как минимум, с наименьшим его участием.

Одна из областей машинного обучения — глубокое обучение. Глубокое обучение ссылается на нелинейные модели, вдохновленные внутренним строением биологических нервных систем, в особенности, человеческим мозгом — искусственные нейронные сети. Нейронные сети применимы к широкому спектру задач: компьютерное зрение, аннотирование изображений, распознавание речи, обработка естественного языка, распознавание аудио, дизайн лекарств, обработка медицинских изображений и т.д. В некоторых областях нейронные сети превосходят человеческих экспертов. Так, в области машинного зрения невероятного успеха добились свёрточные нейронные сети. Первую свёрточную нейронную сеть, названную LeNet-5, предложил Ян ЛеКун в 1989. LeNet-5 распознавала картинки размером 32x32 пикселей, с изображенными на них рукописными цифрами, с точностью 99%.

Довольно долго глубокое обучение не могло проявить себя по ряду причин: недостаточные вычислительные мощности (LeNet-5 обучалась три дня), малые объёмы данных для обучения — данная область получила должное внимание лишь в середине 2000х годов, когда компьютеры

стали обладать достаточной мощностью, а наборы данных для обучения стали достаточно объёмными. Взрыв интереса к глубокому обучению пришелся на 2009 год, когда обучение нейронных сетей стало возможно на графических процессорах, что ускорило процесс в сотни раз. Это дало возможность легкого и относительно быстрого обучения довольно сложных нейронных сетей с большим количеством нейронов. В итоге нейронные сети стали всё чаще применять на практике и на данный момент они являются одной из самых популярных моделей машинного обучения, позволяющей с беспрецедентным уровнем качества решать поставленные перед ней задачи.

1.2 Краткое введение в искусственные нейронные сети

Нейронные сети основаны на множестве соединенных между друг другом узлов, названных искусственными нейронами, которые имитируют нейроны биологического мозга. Обычно они содержат в себе какую-либо нелинейную функцию, такую как гиперболический тангенс или сигмоид, которая аккумулирует сигнал от нейронов «позади», «сжимает» его в определенный диапазон (обычно от 0 до 1) и передает следующим нейронам. Каждое соединение, как и синапс в биологическом мозге, может передавать сигнал от одного нейрона к другому. Соединения между нейронами называются рёбрами. Рёбрам присваиваются числовые коэффициенты, которые называются весами. Сигнал, проходящий по ребру, может подавляться или усиливаться, в зависимости от веса соответствующего ребра: положительный вес — сигнал усиливается, отрицательный — подавляется. Обычно нейроны сгруппированы в слои: входной, скрытые, выходной. Нейронные сети с большим количеством скрытых слоёв называются глубокими. Сигнал (входные данные), подаваемый на входной слой, идёт от него последовательно по скрытым слоям на выходной слой, где формируется ответ сети (существуют нейронные сети, где сигнал распространяется также и в обратную сторону).

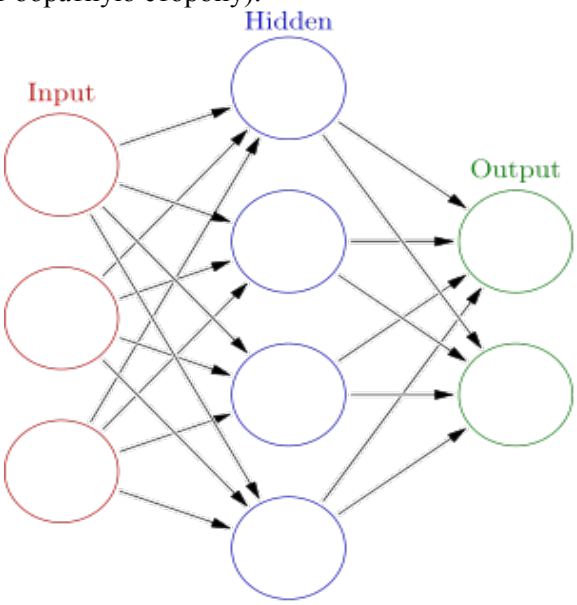

Рисунок 1. Пример полносвязной нейронной сети (персептрон).

1.3 Виды искусственных нейронных сетей

1.3.1 Полносвязная сеть

В данной архитектуре каждый нейрон каждого из слоёв нейронной сети связан с каждым нейроном предыдущего и следующего слоя.

1.3.2 Свёрточная сеть

Архитектура данных нейронных сетей имитирует поведение зрительной коры животного мозга, в который были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определенного набора простых клеток. В силу этого свёрточные нейронные сети отлично подходят для задач обработки данных высокой размерности, например, изображений. Своё название данные сети получили из-за операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу(ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

В отличие от персептрона, в свёрточной нейронной сети помимо полносвязных слоёв есть также слой свёртки и слой субдискретизации (слой пулинга).

В свёрточных слоях, в отличие от полносвязных, в которых каждое ребро имеет свой собственный вес, используется ограниченная матрица весов, называемая ядром свёртки, которую «двигают» по всему обрабатываемому слою, формируя после каждого сдвига сигнал активации для нейрона следующего слоя с аналогичной позицией. То есть для различных нейронов выходного слоя используется одно и то же ядро свёртки. Данный процесс свёртки отдельно выбранной матрицы весов (обычно их много) можно интерпретировать как графическое кодирование определенного признака, например, наклонной линии под определенным углом. Таким образом следующий слой, получившийся в результате операции свёртки, можно назвать картой признаков. Проход каждого ядра свёртки формирует свою карту признаков.

Слой субдискретизации выполняет уменьшение размерности сформированных карт признаков. В данной операции из нескольких соседних нейронов выбирается нейрон с наибольшим выходным сигналом и принимается за один нейрон уплотненной карты признаков меньшей размерности. За счет данной операции ускоряются дальнейшие вычисления, а также сеть становится более инвариантной к масштабу входного изображения.

Типовая структура сверточных нейронных сетей представляет собой ряд сверточных слоёв с чередующимися слоями субдискретизации на входе и полносвязный слой (персептрон) на выходе, который обычно называют классификатором.

1.3.3 Рекурентная сеть

Данный тип искусственных нейронных сетей на порядок сложнее тех, что были озвучены выше, потому мы осмотрим их кратко. Главная особенность данной нейронной сети состоит в том, что сигнал на выходном слое подается на входной слой. Данная особенность приводит к тому, что нейронная сеть становится способной «запоминать» прошлые данные и учитывать временную последовательность. Примером использования данных нейронных сетей является распознавание речи или обработка естественного языка. Современные онлайн переводчики, например, Гугл переводчик, используют именно данный тип нейронных сетей.

1.4 Способы обучения

1.4.1 Обучение с учителем

В обучении с учителем имеется сет данных, называющийся обучающей выборкой, сгруппированных в пары. Между парами данных имеется некоторая зависимость, которая неизвестна и которую необходимо найти. На основе этих данных необходимо восстановить искомую зависимость. Под учителем понимается сама обучающая выборка. В случае с нейронными сетями обучение состоит в коррекции весов. Весы в процессе обучения чаще всего корректируются посредством алгоритма обратного распространения ошибки. Под конкретную задачу подбирается функция ошибки, значение которой отражает разницу между ответом сети и

правильным ответом, тогда ключевой задачей становится минимизация данной функции. Главная идея алгоритма обратного распространения ошибки состоит в том, что от функции ошибки вычисляется градиент — вектор частных производных по каждому из весов нейронной сети — и отнимается от вектора самих весов, таким образом совершается шаг в сторону локального минимума функции ошибки. Обычно градиент умножается на некоторую константу, отражающую скорость обучения: чем меньше константа, тем медленнее обучается нейронная сеть.
Обучение с учителем в большинстве случаев применяется для задач классификации, таких как: классификация изображений, распознавание объектов на изображении и т.д.

1.4.2 Обучение без учителя
В обучении без учителя программе на вход подается сет данных, между которыми требуется установить внутренние взаимосвязи. Другими словами, обучение без учителя позволяет программе разбить данные на группы, сортированные по неким абстрактным признакам, выделенным программой непосредственно во время обучения. Также можно сказать, что обучение без учителя является задачей кластеризации данных.
Нейронные сети как таковые не занимаются кластеризацией данных, а выполняют лишь предварительные этапы. Так, нейронные сети, называемые автокодировщиками, уменьшают размерность данных, выделяя наиболее важные признаки и удаляя шумы, делая последующую обработку данных более эффективной.

1.4.3 Обучение с подкреплением
Поставим задачу: программа - в контексте обучения с подкреплением её называют агентом — находится в некоторой среде и получает от оной информацию о её состоянии; в данной среде программа может совершать ряд действий, на которые среда реагирует посредством выдачи «награды»; задача программы — максимизировать награду. Примером использования обучения с подкреплением является обучение ИИ для игры в Super Mario Bros.

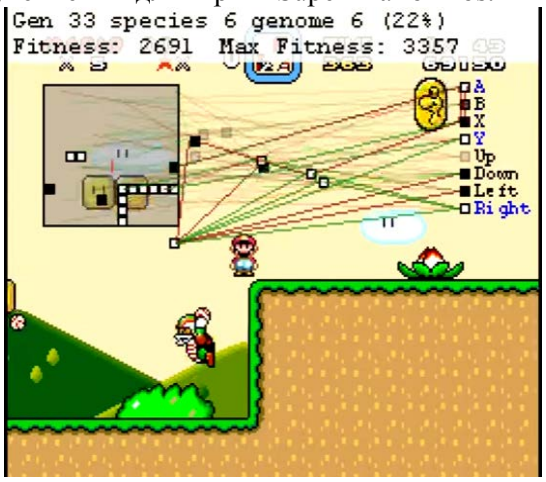

Рисунок 2. Программа учится играть в Super Mario Bros.

В ходе обучения программа понимала, что нажатие на кнопку «право» увеличит награду. В данном примере непосредственно для обучения использовался генетический алгоритм.
Примером нейронных сетей, обученных данным путём, являются нейронные сети, обученные OpenAI для контроля ботов в игре Dota 2. Данные нейронные сети достигли уровня игры профессиональных игроков и не раз побеждали их в спортивном соревновании, хотя и с рядом ограничений на саму игру.

2. Материалы исследования

2.1 Тестируемые нейронные сети

В исследовании рассматривается 4 предобученных свёрточных нейронных сети:

- AlexNet
- GoogLeNet
- VGG16
- ResNet152

Выше озвученные нейронные сети обучены на наборе Places365-Standart, который содержит в себе 1.8 миллиона изображений 365 классов различных сцен, примерно по 5000 изображений на каждую. Данные нейронные сети лежат в публичном доступе на GitHub репозитории.

Нейронные сети обучены с использованием фреймворка для глубокого обучения Caffe.

2.1.1 Точность нейронных сетей

В приведенной ниже таблице показана точность каждой из нейронных сетей, измеренная на валидационном и тестовом наборах Places365.

Таблица 1. Топ-1 и топ-5 точности исследуемых нейронных сетей.

	Валидационная выборка Places365		Тестовая выборка Places365	
	Топ-1 точность	Топ-5 точность	Топ-1 точность	Топ-5 точность
AlexNet	53.17%	82.89%	53.31%	82.75%
GoogLeNet	53.63%	83.88%	53.59%	84.01%
VGG16	**55.24**%	84.91%	**55.19**%	85.01%
ResNet152	54.74%	**85.08**%	54.65%	**85.07**%

2.1.2 Places365

Places365 является последней подвыборкой Places2. Places2 представляет собой базу данных в общем доступе, содержащей в себе около 10 миллионов изображений более чем 400 категорий сцен. Данный набор предоставляет от 5000 до 30000 изображений на категорию в соответствии с реальными частотами возникновения данных сцен в реальном мире.

2.1.3 Caffe

Caffe является фреймфорком для глубинного обучения, разработанный Яньцинем Цзя в процессе подготовки своей диссертации в университете Брекли. Caffe является открытым ПО, распространяемым под лицензией BSD license. Написано на языке C++ и поддерживает интерфейс на языке Python.

Caffe поддерживает много типов машинного обучения, нацеленных в первую очередь на решение задач классификации и сегментации изображений. Данный фреймворк поддерживает свёрточные, рекурентные и полносвязные нейронные сети. Caffe поддерживает работу на CPU и GPU, а также поддерживает ГП-ориентированные библиотеки, ускоряющие вычисления, такие как NVIDIA cuDNN и Intel MKL.

2.2 Тестовые данные

В качестве тестовых данных использовались изображения из интернета. Размер тестовой выборки составляет 800 изображений; изображения имели разные размеры.

Изображения загружались с помощью встроенных средств Caffe.

2.3 Вычислительная система
Все тесты проводились с использованием видеокарты GTX 1080 Ti с 12 гигабайтами видеопамяти на борту.

3. Тестирование и результаты

3.1 Постановка эксперимента
Массив из 800 тестовых изображений предварительно трансформировался при помощи трансформера Caffe для соответствующей нейронной сети. Таким образом, размерность изображений подгонялась под размерность входного слоя нейронных сетей. После трансформации изображения последовательно подавались соответствующей нейронные сети на классификацию. Измерялось время классификации и используемая нейронными сетями видеопамять.

3.2 Результаты
В таблице 2 показаны результаты тестов производительности нейронных сетей и занимаемая ими видеопамять.

Таблица 2. Скорость классификации и занимаемая видеопамять.

	Занимаемая память (МБ)	Время обработки всего тестового набора (с)	Скорость классификации (изображений $с^{-1}$)
GoogLeNet	275	12.2	65.5
AlexNet	69	4.8	166.6
VGG16	107	29.2	27.4
ResNet152	1036	91.6	8.7

В таблице 3 показана производительность трансформеров соответствующих нейронных сетей и занимаемая массивом трансформированных изображений видеопамять.

Таблица 3. Скорость трансформирования и размер занимаемой памяти массивом трансформированных изображений.

	Размер массива изображений после трансформации (МБ)	Размер трансформированного изображения (МБ)	Время трансформации всего тестового набора (с)
GoogLeNet	364	0.46	56.4
AlexNet	293	0.37	51.7
VGG16	1241	1.55	53
ResNet152	4486	5.6	52.7

4. Обсуждение результатов

4.1 Производительность нейронных сетей
Среди исследуемых нейронных сетей наиболее интересными являются GoogLeNet и AlexNet, поскольку они серьезно опережают свои аналоги VGG16 и ResNet152 в скорости классификации и занимаемой памяти, из-за чего разница в точности классификации нивелируется.
Таблица 2 показывает, что абсолютным лидером по скорости обработки изображений является AlexNet, установив рекорд в 166 изображений в секунду, что 2.5 раза больше, чем у GoogLeNet, и в 19 раз больше, чем у ResNet152. Разница в точности классификации с остальными сетями

нивелируется существенным понижением скорости классификации, отчего AlexNet выступает лучшим кандидатом в качестве классификатора для систем, требующих высокой скорости обработки изображений.
Наиболее «легковесными» являются VGG16 и AlexNet. Размер данных сетей говорит о том, что данные нейронные сети вполне можно использовать в мобильных системах с ограниченными ресурсами. Но учитывая разницу в производительности, очевидным выбором в качестве мобильного классификатора будет AlexNet, поскольку данная нейронная сеть обгоняет VGG16 по скорости классификации в 6 раз.
Наихудшие результаты у нейронной сети ResNet152. Данная сеть проигрывает в каждом тесте каждой из нейронных сетей, работая на порядок медленнее и занимая на порядок больше памяти. И хотя в среднем точность данной нейронные сети выше чем у остальных, данная разница нивелируется существенным падением производительности и размером занимаемой памяти, что делает ResNet152 худшим классификатором, среди приведенных в данном исследовании.

4.2 Время подготовки изображений
Исходя из приведенных результатов в таблице 3 можно сказать, что нейронные сети идентичны друг другу с точки зрения скорости подготовки изображений для них, при условии, что подготовка производится посредством внутренних средств Caffe.
Интересна колонка со значениями видеопамяти, занимаемой массивом трансформированных изображений. Довольно высокими значениями обладают VGG16 и ResNet152. Этот фактор следует учитывать, если данные нейронные сети будут использоваться в качестве основы для других классификаторов. Большой размер трансформированного изображения влияет на максимальное количество изображений в одном батче для обучения нейронной сети, а именно уменьшает максимально возможное количество изображений в батче, что может негативно сказаться на качестве обучения. Также данный фактор стоит учитывать в случае, если данные нейронные сети будут использоваться в системах с небольшим объёмом видеопамяти. На фоне разницы размеров трансформированных изображений для соответствующих нейронных сетей и небольшой разнице в точности GoogLeNet и AlexNet выглядят куда более привлекательными.

5. Выводы
Приведенное исследование показывает, что AlexNet, с учетом производительности и занимаемой памяти, является лучшим классификатором среди исследуемых. Учитывая производительность и размеры, данная нейронная сеть может быть применена как в больших, требующих высокой скорости обработки изображений, так и в мобильных системах. Также можно сказать, что хорошим классификатором является GoogLeNet. Данная нейросетоь отстала от AlexNet по скорости классификации в 2.5 раза, но учитывая, что в среднем точность GoogLeNet выше, данная нейронная сеть выглядит хорошим классификатором для больших систем.
VGG16 и ResNet152, несмотря на некоторое превосходство в точности, являются не лучшим выбором в качестве классификаторов на фоне AlexNet и GoogLeNet в силу своего серьезного отставания в производительности. Также относительно ResNet152 следует отметить, что данная нейронная сеть на фоне своих конкурентов с экономической точки зрения становится невыгодной, поскольку помимо низкой производительности, ситуацию также усугубляет и большой размер трансформированных для данной нейронной сети изображений, что усложняет её использование в качестве основы для иных классификаторов.

6. Литература
[1] LeCun, Y. Gradient-Based Learning Applied to Document Recognition / Y. LeCun, L. Bottou, Y. Bengio, P.Haffner [Electronic resource]. – Access mode: http://yann.lecun.com /exdb/publis/pdf/lecun-01a.pdf.

[2] The program learning how to play Super Mario Bros [Electronic resource]. – Access mode: https://www.youtube.com/watch?v=qv6UVOQ0F44.
[3] The Places database [Electronic resource]. – Access mode: http://places2.csail.mit.edu/index.html
[4] Caffe Deep Learning Framework [Electronic resource]. – Access mode: http://caffe. berkeleyvision.org/.
[5] NVIDIA cuDNN [Electronic resource]. – Access mode: https://developer.nvidia.com/cudnn.
[6] Inter MKL [Electronic resource]. – Access mode: https://software.intel.com/en-us/mkl.
[7] Places: A 10 million Image Database for Scene Recognition / B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba [Electronic resource]. – Access mode: http://places2.csail.mit.edu/ PAMI_places.pdf.
[8] Dota 2 AI trained by OpenA I [Electronic resource]. – Access mode: https://openai.com/five/.
[9] Mattsson, N. Classification Performance of Convolutional Neural Networks [Electronic resource]. – Access mode: http://www.diva-portal.org/smash/get/diva2:1037364/FULLTEXT02.
[10] Li, X. Performance Analysis of GPU-based Convolutional Neural Network / X. Li, G. Zhang, H. Howie Huang, Z. Wang, W. Zheng [Electronic resource]. – Access mode: https:// www2.seas.gwu.edu/~howie/publications/GPU-CNN-ICPP16.pdf.
[11] Convolutional neural network [Electronic resource]. – Access mode: https://en.wikipedia.org /wiki/Convolutional_neural_network.
[12] Artificial Neural network [Electronic resource]. – Access mode: https://en.wikipedia.org/ wiki/Artificial_neural_network.
[13] Genetic algorithm [Electronic resource]. – Access mode: https://en.wikipedia.org /wiki/Genetic_algorithm.

Scene recognition accuracy and performance comparison of CNNs

I. Kilbas[1], R. Paringer[1]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. For today, image classification is becoming more relevant. One of the most popular solutions is convolutional neural networks. But the effectiveness of neural networks has its price - they require large resources for training. It is not always possible to train your own neural network, but even in this situation there is a way — to use a pre-trained neural network. In this research, we take a look at some pre trained convolutional neural networks: compare their operating time, accuracy, and memory usage.