УДК 004.932.72'1

ОБЗОР АРХИТЕКТУР СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

Васкан В.Д.

МИРЭА - Российский технологический университет, 119454, Россия, г. Москва, проспект Вернадского, 78, e-mail: vaskan97@mail.ru

В данной статье рассмотрена задача классификации изображений, дано краткое описание принципа работы сверточных нейронных сетей. Проведен обзор базовых, а также современных архитектур сверточных нейронных сетей и сделано сравнение их точности и ресурсозатратности на базе изображений ImageNet.

Ключевые слова: сверточные нейронные сети, классификация изображений, архитектура сверточной нейронной сети.

OVERVIEW OF CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR THE IMAGE CLASSIFICATION PROBLEM

Vaskan V.D.

MIREA - Russian Technological University, 119454, Moscow, 78 Vernadskogo Avenue, Russia, e-mail: vaskan97@mail.ru

This article describes the problem of image classification and a brief description of the principle of operation of convolutional neural networks is given. A review of the basic and modern architectures of convolutional neural networks is carried out, and a comparison of their accuracy and resource consumption based on ImageNet images is made.

Keywords: convolutional neural networks, image classification, CNN architecture.

Введение

В современном мире технологии компьютерного зрения используются во многих сферах нашей жизни. Их применяют для таргетирования рекламы в маркетинге, для анализа снимков и дальнейшей помощи при определении диагнозов в медицине, в смартфонах как способ разблокировки, в беспилотных автомобилях. Появилось даже первое отделение Сбербанка, где можно получить доступ к своим счетам с помощью технологии распознавания лиц, что является подтверждением надежности и высокой точности такого подхода. С ростом вычислительных мощностей росла и глубина слоев нейронных сетей, а нейронные сети, содержащие операцию свертки, добились большого успеха в области компьютерного зрения.

Целью данной работы является обзор и сравнение точности и ресурсозатратности архитектур сверточных нейронных сетей для классификации изображений.

Задача классификации изображений

Одной из классических задач машинного зрения является задача классификации изображений – процесс извлечения классов информации из многоканального растрового изображения.

Оценка качества алгоритмов машинного обучения производится на аннотированных базах изображений, таких как CIFAR-10, ImageNet и др.

В связи с тем, что изображения базы ImageNet могут содержать несколько различных объектов, аннотирован из которых только один, в виде основной оценки применяется top-5 ошибка, суть которой заключается в том, что, присутствие искомой категории в пяти наиболее вероятных предсказаниях считается правильным ответом.

Сверточные сети

Сверточные сети позволяют специализировать нейронные сети для работы с данными, имеющими четко выраженную сеточную топологию, и хорошо масштабировать такие модели к задачам очень большого размера [2, с. 315]. Особенно успешным этот подход оказался в применении к двумерным изображениям.

Первая работа по современным сверточным нейронным сетям принадлежит Яну Лекуну. LeCun et al. в своей статье [3] представили модель сверточной нейросети (LeNet-5), представленную на рисунке 1, которая объединяет более простые функции в прогрессивно более сложные функции.

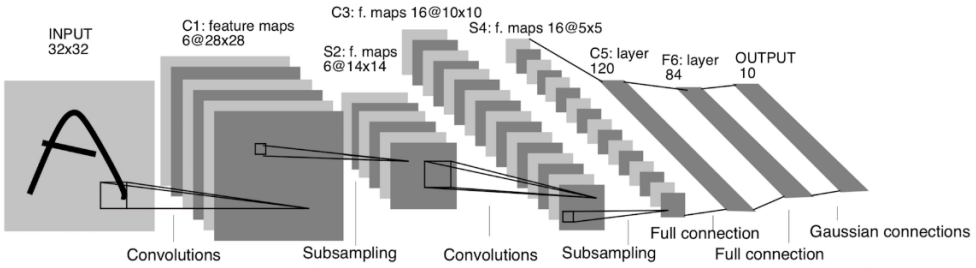

Рис. 1. Архитектура LeNet5

Труды LeCun et al. стали почвой для дальнейших исследований в области сверточных нейросетей и на это есть причины. Сверточные нейронные сети учитывают информацию о топологии, так как анализируются соседние пиксели не только по горизонтали, но и по вертикали. Также в них используется уменьшение размерности - это позволяет распознавать объекты вне зависимости от масштаба, его местоположения на картинке и, непосредственно, уменьшить размерность путем добавления слоя субдискретизации. Одним из главных преимуществ является использование разделяемых весов: веса в ядрах свертки применяются в разных регионах изображения, в связи с чем значительно понижается количество весов для обучения, в следствие чего можно лучше обучать нейросеть. Поэтому Сверточные сети были одними из первых работающих глубоких сетей, обученных с применением обратного распространения. Вероятно, сверточные сети добились успеха там, где сети общего назначения с обратным распространением потерпели неудачу исходя из того, что они вычислительно эффективнее полносвязных сетей, поэтому было проще ставить эксперименты и оптимизировать реализацию и гиперпараметры[3].

Как было упомянуто ранее, на данный момент лучше всего с обработкой изображений справляются свёрточные нейронные сети (CNN). Для того чтобы перейти к описанию моделей CNN, разберемся что такое CNN и как они выглядят.

Основное отличие этой архитектуры от построения классических нейронных сетей заключается в наличии слоёв свёртки (convolutional layer) – скрытых слоев нейронной сети, в которых происходит свёртка изображения, с помощью фильтров, а также слоев субдискретизации (pooling layer) и полносвязных слоев (fully connected layers).

Сверточные нейронные сети для классификации изображений

В данном разделе рассматриваются архитектуры сверточных нейронных сетей,

AlexNet

Системы на основе сверточных сетей выигрывали много конкурсов. Современный всплеск коммерческого интереса к глубокому обучению начался, когда система, описанная в работе Krizhevsky et al. (2012), победила в конкурсе по распознаванию объектов ImageNet, но сверточные сети побеждали и в других соревнованиях по машинному обучению и компьютерному зрению задолго до того, хотя это и не вызывало такого ажиотажа.

Krizhevsky et al. представили сеть AlexNet, которая с большим отрывом выиграла соревнование ImageNet LSVRC-2012 (с количеством ошибок 15,3% против 26,2% у второго места)[4].

AlexNet обучалась на двух GPU, что позволила ускорить время обучения. Также использовалась функция активации ReLU вместо арктангенса, что уменьшило количество необходимых для обучения эпох в шесть раз.

Формула ReLU представлена ниже:

$$y(x) = \max(0; x)$$

Помимо функции активации, одной из особенностей является применение метода прореживания (Dropout). Это метод регуляризации для нейронных сетей, которые способствует в борьбе с переобучением, путем отключения случайных нейронов.

ZFNet

ZFNet – победитель ILSVRC 2013 с top-5 ошибкой 11,2 %. Такого результат удалось достичь благодаря точной настройке гиперпараметров, таких как размер фильтров и их количество, скорость обучения, размер пакетов и т. д. Данная архитектура похожа на AlexNet, однако отличается размером фильтра в первом сверточном слое и шаге, а также количеством фильтров. Основной особенностью данной архитектуры является система визуализации ядер, весов и скрытого представления изображений, которую представили Zeiler и Fergus, получившая название DeconvNet, что стало толчком в развитии и понимании сверточных нейронных сетей.

VGG Net

В 2014 году Simonyan и Zisserman представили архитектуру сети под названием VGG. Основной и отличительной идеей этой структуры является сохранение фильтров настолько простыми, насколько это возможно. Они продемонстрировали, что использование фильтра 7x7 эквивалентно использованию трех фильтров 3x3, при этом, количество параметров получается на 55% меньше во втором случае. Одновременно с простотой сверточных модулей они увеличили сеть в глубину до 19 слоев. На соревновании ILSVRC 2014 ансамбль из двух VGG Net получил top-5 ошибку 7,3 %. Данный ансамбль не победил соревнование, однако модель используют в более сложных сетях.

GoogLeNet

Ранее все развитие архитектуры заключалось в упрощении фильтров и увеличении глубины сети. В 2014 году Szegedy et al. совместно с другими участниками предложил совершенно иной подход и создал самую сложную на тот момент времени архитектуру, называемую GoogLeNet.

Ключевым отличием и достижением их подхода является наличие в сети модуля Inception, представленного на рисунке 2. В данном модуле входные данные обрабатываются в параллельном режиме, в то время как все предыдущие архитектуры использовали исключительно последовательную обработку. За счет этого удалось ускорить получение вывода и уменьшить количество параметров.

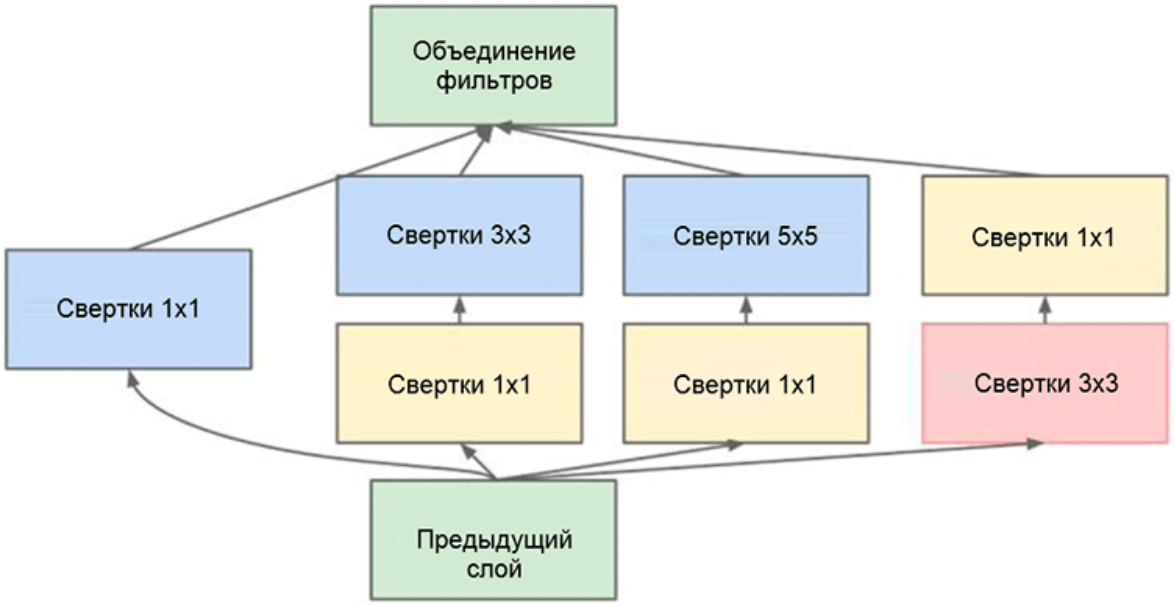

Рис. 2. Модуль Inception

Стоит также отметить, что в данном модуле они применяли фильтр размером 1x1, что позволило просто сократить размерность карты свойств. Такой тип сверточных слоев впервые был представлен в работе Lin et al.[4]. GoogLeNet уменьшил top-5 ошибку до 6,7%.

ResNet

Победителем с 2015 с top-5 ошибкой в 3,57 % стал ансамбль из шести сетей типа ResNet (Residual Network), означающий «остаточные сети». Данная архитектура была разработана He et al. из Microsoft Research.

Авторы ResNet обнаружили следующее: при увеличении количества слоев сети возможно уменьшение точности на валидационном множестве. Причем на тренировочном множестве точность также падает, из чего делается вывод о том, что данная проблема не связана с переобучением. Это связано с проблемой исчезающего градиента, что является технической проблемой приводящей к невозможности обучения сети в принципе.

He et al. для решения этой проблемы использовали остаточное отображение (элемент, который следует добавить ко входным данным) вместо отображения как такового в обучении. Это делается с помощью соединения, показанного на рисунке 3.

Данная архитектура позволила избежать проблему исчезающего градиента и обучать крайне глубокие сети.

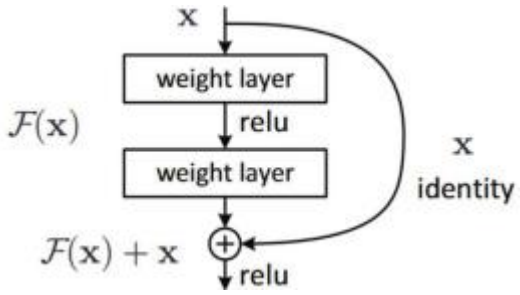

Рис. 3. Остаточное отображение ResNet

EfficientNet

В 2020 году Tan and Le создали EfficientNet. Во время своих исследований они сделали два вывода: первый – масштабирование любого измерения ширины, глубины или разрешения сети повышает точность, но для больших моделей прирост точности уменьшается; второй –для достижения большей точности и эффективности крайне важно сбалансировать все измерения ширины, глубины и разрешения сети во время масштабирования[5]. Они предложили новый метод составного масштабирования, который использует составной коэффициент φ для равномерного масштабирования ширины, глубины и разрешения сети который выглядит следующим образом:

$$\text{глубина: } d = \alpha^{\varphi}$$
$$\text{ширина: } w = \beta^{\varphi}$$
$$\text{разрешение: } r = \gamma^{\varphi}$$
$$\text{такие, что } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$
$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

где α, β, γ-константы, которые могут быть определены с помощью

небольшого поиска по сетке. Коэффициент φ – это заданный пользователем коэффициент, который определяет, сколько еще ресурсов доступно для масштабирования модели, в то время как α, β, γ определяют, как назначить эти дополнительные ресурсы сети ширине, глубине и разрешению соответственно. Среди особенностей реализации следует также отметить применение автоаугментации, а также использование функции активации SiLU, которая имеет следующий вид:

$$y(x) = x * sigmoid(x) = x * \frac{1}{1 + e^{-x}}$$

EfficientNet-B2 в сравнении с ResNet-152 удалось добиться уменьшения ошибки top-5 на 0.6%, при этом, такая архитектура имеет в 7.6 раз меньше параметров, а также требует в 16 раз меньше FLOPs, при этом EfficientNet-B7 удалось достичь минимального значения top-5 ошибки 3%, которое ранее установила архитектура GPipe, однако использовала она в 8.6 раз больше параметров [5].

Meta Pseudo Labels (EfficientNet-L2)

Pham et al. представили в 2021 году модель, которая относится к виду self-supervised (самообучение). Особенностью такого вида обучения является использование неразмеченной выборки. Имеется две сети, одна из которых выступает в роли учителя, а вторая в роли ученика. Учитель генерирует псевдо-метки на

неразмеченных изображениях, а затем смешивает их с размеченными данными, которые затем передаются ученику. Отличительной же особенностью Meta Pseudo Labels является то, что учитель получает обратную связь от ученика, и корректирует псевдо-метки для дальнейшего улучшения успеваемости ученика[6].

Полученная сеть достигает максимальной точности имея top-5 ошибку 1.2% являясь лучшей сетью для классификации изображений ImageNet.

Сравнение моделей

В таблице 1 представлены результаты рассмотренных нейронных сетей на базе изображений ImageNet, а также количество используемых ими параметров. Из таблицы видно, что наилучший результат достигается при использовании архитектуры мета псевдо-меток. Однако, количество параметров, используемых в данной архитектуре, в разы больше, чем у других, что позволяет эффективно ее использовать при обилии вычислительных мощностей. При дефиците ресурсов стоит выделить EfficinetNet-B7 или же GoogLeNet.

Таблица 1. Результаты и количество параметров моделей

Нейронная сеть	Top-1	Top-5	Количество параметров
AlexNet	39,00 %	17 %	60M
ZF Net	37,50 %	16 %	60M
VGG19	25,60 %	8,10 %	144M
GoogLeNet	29,00 %	9,20 %	5M
ResNet-152	19,38 %	4,49 %	60M
EfficinetNet-B7	15,6%	2,9%	66M
Meta Pseudo Label	9,8%	1,2%	480M

Заключение

В данной работе были описаны одни из самых значимых архитектур свёрточных нейронных сетей для задачи классификации изображений, в том числе современных, а также было проведено их сравнение по точности и ресурсозатратам. Они позволили сильно улучшить точность распознавания изображений и превзойти результат человека (top-5 ошибка 5%).

Список литературы

1. Топ-5 сфер применения систем распознавания объектов [Электронный ресурс] / www.habr.com Режим доступа: https://habr.com/ru/company/toshibarus/blog/433544/ // (дата обращения 07.03.2021).
2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль.- Москва : ДМК Пресс, 2018. – 651 с.
3. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient based Learning Applied to Document Recognition [Электронный ресурс] / Режим доступа: http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf // (дата обращения: 10.03.2021).
4. Lin M., Chen Q., Yan S. Network In Network [Электронный ресурс] / Режим доступа: https://arxiv.org/pdf/1312.4400v3.pdf // (дата обращения: 12.03.2021).
5. Tan M., Quoc V. Le EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [Электронный ресурс] / www.arXiv.org. Режим доступа: https://arxiv.org/pdf/1905.11946.pdf // (дата обращения: 20.03.2021).
6. Pham H., Dai Z., Xie Q., Luong M.-T., Quoc V. Le Meta Pseudo Labels [Электронный ресурс] / www.arXiv.org. Режим доступа: https://arxiv.org/pdf/2003.10580v4.pdf // (дата обращения: 25.03.2021)
7. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition [Электронный ресурс] / www.arXiv.org. Режим доступа: https://arxiv.org/pdf/1409.1556.pdf // (дата обращения: 01.04.2021)

References

1. Top-5 areas of application of object recognition systems [Elektronnyj resurs] / www.habr.com Available at: https://habr.com/ru/company/toshibarus/blog/433544/ (accessed 07 March 2021).

2. Gudfellou YA., Bendzhio I., Kurvill' A. Glubokoe obuchenie./ YA. Gudfellou, I. Bendzhio, A. Kurvill'.- Moskva : DMK Press, 2018. – 651 s.

3. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient based Learning Applied to Document Recognition [Elektronnyj resurs] / Available at: http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf // (accessed 10.03.2021).

4. Lin M., Chen Q., Yan S. Network In Network [Elektronnyj resurs] / Available at: https://arxiv.org/pdf/1312.4400v3.pdf // (accessed 12 March 2021).

5. Tan M., Quoc V. Le EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [Elektronnyj resurs] / www.arXiv.org. Available at: https://arxiv.org/pdf/1905.11946.pdf // (accessed 20 March 2021).

6. Pham H., Dai Z., Xie Q., Luong M.-T., Quoc V. Le Meta Pseudo Labels [Elektronnyj resurs] / www.arXiv.org. Available at: https://arxiv.org/pdf/2003.10580v4.pdf // (accessed 25 March 2021)

7. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition [Elektronnyj resurs] / www.arXiv.org. Available at: https://arxiv.org/pdf/1409.1556.pdf // (accessed 01 April 2021)