

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ЧИСЛА ПАРАМЕТРОВ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ НА ТОЧНОСТЬ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

Ермакович В.А.¹, студент гр.250701

Белорусский государственный университет информатики и радиоэлектроники¹
г. Минск, Республика Беларусь

Вашкевич М.И. – докт. техн. наук, доцент

Аннотация. В работе представлено исследование на тему влияния числа параметров сверточной нейронной сети на точность распознавания изображений. Исследование проводилось на основе архитектуры LeNet-5 с изменением в ней количества карт признаков и числа полносвязных слоев. Представлено в исследовании описание данной архитектуры с использованием планировщика скорости обучения, таблица и график с результатами об обучении модификаций исходной модели.

Ключевые слова. LeNet, СНС, распознавание изображений.

Введение

Сверточные нейронные сети (СНС) – особый тип нейронных сетей, разработанный специально для распознавания изображений в 1988 году Яном Лекуном [1]. Данные нейронные сети основываются на операции свертки и иерархическом извлечении признаков: на первых сверточных слоях выявляются базовые структуры (линии, края), а на последующих слоях могут определяться как шаблоны (текстуры, части объектов), так и целые объекты.

Из-за таких особенностей СНС получили свою популярность в различных сферах: системы контроля качества на производствах, безопасность, системы автопилотов, космос, медицина и др.

Среди наиболее известных архитектур СНС можно отметить:

- LeNet – одна из первых СНС, разработанная для распознавания рукописных цифр. Сеть продемонстрировала эффективность архитектуры с чередованием свёрточных слоев и слоев субдискретизации (англ. *pooling*) – и стала основой для дальнейших исследований в области СНС.
- AlexNet – сеть, победившая в соревновании ImageNet 2012 [2], оказала огромное влияние в развитие компьютерного зрения. Её архитектура схожа с LeNet, но значительно глубже: включает большее количество свёрточных слоёв. Впервые в ней были применены такие современные приёмы, как активационные функции ReLU, дропаут для борьбы с переобучением, аугментация данных для расширения обучающей выборки и субдискретизация (типа max-pooling). Также обучение велось с использованием GPU, что позволило эффективно справляться с большим объёмом данных.
- VGG16 – глубокая сверточная сеть, построенная из последовательных блоков, содержащих свёртки размером 3×3 с единичным шагом и слою субдискретизации [3]. В своей архитектуре содержит 13 сверточных слоев, 5 слоев субдискретизации и 3 полносвязных слоя. Несмотря на большое число параметров, она показывает высокую точность.

Несмотря на высокую точность, глубокие СНС часто имеют миллионы обучаемых параметров, что приводит к долгому обучению, большим вычислительным затратам, сложностям внедрения самих моделей в устройства с ограниченным вычислительным ресурсом. Целью данного исследования было – определить, как изменение числа обучаемых параметров (количество фильтров на сверточных слоях, наличие полносвязных слоев) влияет на точность распознавания, и можно ли достичь сопоставимого качества при меньшем числе параметров.

Архитектура LeNet

В качестве исходной СНС выбрана сеть LeNet-5 [1]. Архитектура LeNet-5 состоит из следующих слоев (рисунок 1):

- 1) Входной слой – изображение 32 на 32 пикселя.
- 2) C1 – первый сверточный слой с 6-ю картами признаков, размер каждой составляет 28х28.
- 3) S2 – первый слой субдискретизации с 6-ю картами признаков, размер каждой карты 14х14.
- 4) C3 – второй сверточный слой с 16-ю картами признаков, размер 10х10.
- 5) S4 – второй слой субдискретизации с 16-ю картами признаков, размер каждой карты 5х5.
- 6) C5 – слой, в котором 120 карт признаков, каждая размером 1х1, является полносвязным сверточным слоем, т.е. все узлы S4 слоя соединены с каждым из 120 блоков в слое C5
- 7) F6 – полносвязный слой с 84 нейронами.
- 8) Выходной слой имеет 10 нейронов и функцию активации softmax. Каждый нейрон соответствует своему классу распознаваемых изображений.

Отметим, что при исследовании параметры некоторых слоев были изменены, а именно: на вход подавались изображения 28х28, полученные карты признаков были следующего размера, C1 – 24х24, S2 – 12х12, C3 – 8х8, S4 – 4х4.

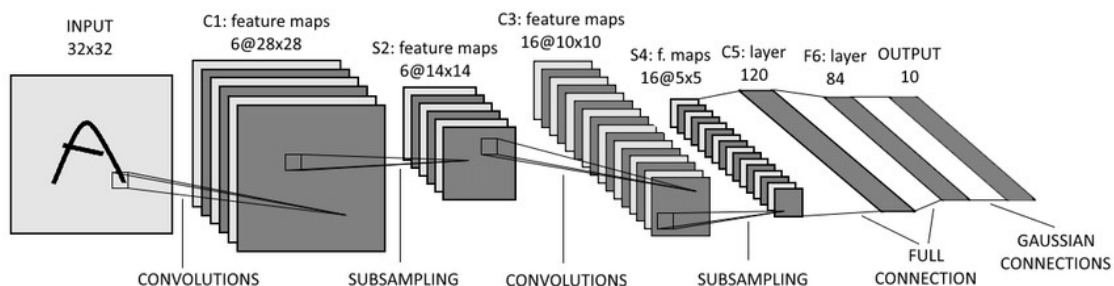


Рисунок 1 – Архитектура сверточной нейронной сети LeNet-5 [1]

Дополнением к этой архитектуре будет является использование слоев дропаута (англ. *dropout* – выбрасывание) после полносвязных слоев. Основная их идея заключается в случайном «выключении» определенного процента нейронов в сети на каждом шаге обучения. Это означает, что во время каждого прохода обучения (или каждой эпохи) случайно выбранный набор нейронов игнорируется (т.е. обнуляется). Это помогает предотвратить чрезмерную зависимость модели от конкретных путей и узлов в сети, что может привести к переобучению.

Задача исследования

Задача в исследовании архитектуры сверточной нейросети LeNet-5 состоит в том, чтобы узнать, как влияет на точность распознавания количество параметров модели. Для этого выполнялась серия экспериментов, в которых изменялось количество карт признаков, которые формируются в сверточных слоях C1 и C3, а также количество полносвязных слоев. В качестве данных будет использована база изображений Fashion-MNIST, которая разбита на обучающую, тестовую и валидационную выборки. Пример изображений из базы Fashion-MNIST показаны на рисунке 2.



Рисунок 2 – Пример изображений из базы Fashion-MNIST

Также для обучения использовался планировщик скорости обучения (*learning rate scheduler*), который плавно снижает скорость обучения и после прохождения определенного количества эпох, сбрасывает скорость обучения до начального значения.

Формула планировщика (косинусное затухание)

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + \cos(\frac{T_{cur}}{T_{max}}\pi)) \quad (1)$$

где η_{min} – минимальная скорость обучения, η_{max} – максимальная скорость, T_{cur} – количество эпох с начала цикла, T_{max} – количество эпох, в течение которых происходит спад косинуса, прежде чем скорость сбросится.

Наличие планировщика позволяет выходить из локальных минимумов функции потерь в процессе обучения.

Основные метрики, на которые будет обращать внимание в исследовании, это количество параметров в модели и точность ее предсказаний.

В качестве примера на рисунке 3 изображены графики обучения оригинальной архитектуры LeNet5-classic, на которых показаны как изменялась ошибка на валидационной и обучающей выборке и график изменения скорости обучения через косинусное затухание.

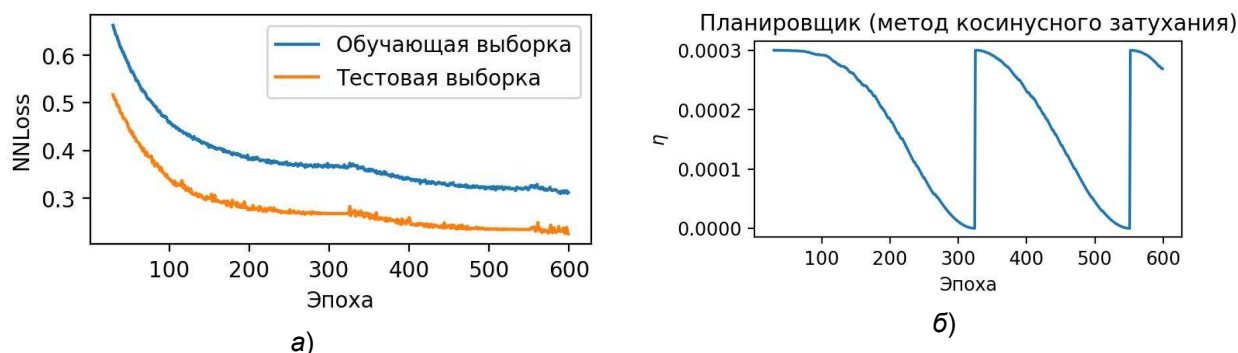


Рисунок 3 – LeNet5-classic: а) график изменения ошибки; б) График изменения скорости обучения

Результаты экспериментов

В исследовании мы выделили 3 группы моделей по 12 штук в каждой. Во всех 3-х группах изменялись сверточные слои (количество карт признаков), но во 2-ой и 3-ей группе были также изменения полносвязных слоев. Во второй группе количество полносвязных равно двум, а в третьей группе — одному. Все сети обучались на 600 эпох, размер минипакета – 1000 изображений, для обучения использовался метод Adam, в качестве функции потерь – перекрестная энтропия, начальная скорость обучения $\eta_{max} = 3 \cdot 10^{-4}$, минимальная скорость обучения $\eta_{min} = 10^{-8}$, параметр планировщика скорости $T_{max} = 200$.

В таблице 1 приведены все результаты обучения полученных моделей.

Таблица 1 – Сравнение различных вариантов CHC LeNet

Номер варианта	Число фильтров в слое C1	Число фильтров в слое C3	fc1	fc2	fc3	Число параметров	Правильность
LeNet5-classic	6	16	+	+	+	44 426	89,92%
LeNet5-v1	3					43 148	88,96 %
LeNet5-v2	2					42 722	89,58 %
LeNet5-v3	6	8	+	+	+	27 858	89,63 %
LeNet5-v4	3					27 180	89,63 %
LeNet5-v5	2					26 954	88,98%
LeNet5-v6	6	4	+	+	+	19 574	89,61%
LeNet5-v7	3					19 196	88.37%
LeNet5-v8	2					19 070	88.02%
LeNet5-v9	6	2	+	+	+	15 432	86.56%
LeNet5-v10	3					15 204	86.23%
LeNet5-v11	2					15 128	85.72%
LeNet5-f1-f2-1	6	16	+	+	-	34 622	89.56%
LeNet5-f1-f2-2	3					33 344	89.09%
LeNet5-f1-f2-3	2					32 918	88.73%
LeNet5-f1-f2-4	6	8	+	+	-	18 054	88.19%
LeNet5-f1-f2-5	3					17 376	87.63%
LeNet5-f1-f2-6	2					17 150	87.88%
LeNet5-f1-f2-7	6	4	+	+	-	9770	87.27%
LeNet5-f1-f2-8	3					9392	86.53%
LeNet5-f1-f2-9	2					9266	86.57%
LeNet5-f1-f2-10	6	2	+	+	-	5628	84.62%
LeNet5-f1-f2-11	3					5400	85.30%
LeNet5-f1-f2-12	2					5324	83.59%
LeNet5-f1-1	6	16	+	-	-	5142	38.37%
LeNet5-f1-2	3					3864	57.72%
LeNet5-f1-3	2					3438	76.97%
LeNet5-f1-4	6	8	+	-	-	2654	51.85%
LeNet5-f1-5	3					1976	56.97%
LeNet5-f1-6	2					1750	61.48%
LeNet5-f1-7	6	4	+	-	-	1410	75.76%
LeNet5-f1-8	3					1032	37.78%
LeNet5-f1-9	2					906	62.53%
LeNet5-f1-10	6	2	+	-	-	788	68%
LeNet5-f1-11	3					560	50.65%
LeNet5-f1-12	2					484	59.63%

Рассмотрим результаты первой группы из таблицы 1 (с LeNet5-classic по LeNet5-v11). В этих моделях изменялись только количество карт признаков на слоях C1 и C3. Можно увидеть, что модели с наличием на слое C3 16 и 8 карт признаков можно считать перенасыщенными, так как точность вирируется в районе 89%, а количество параметров довольно большое (более 25 тыс). В моделях с 4-мя картами точность такая же, но количество параметров уже меньше (менее 20 тыс). В моделях с 2 картами на слое C3 уже точность начинает падать существенно. Лучшей моделью из данного списка можно считать LeNet-v6, в которой точность 89.61%, а количество параметров — 19574.

В моделях второй группы (с LeNet5-f1-f2-1 по LeNet5-f1-f2-12 из таблицы 1) имеется только два полносвязных слоя. Уменьшение полносвязных слоев существенно уменьшает количество параметров моделей, но при этом видно еще то, что точность деградирует незначительно. У самой «маленькой» модели наблюдается точность в 83.59% при 5324 параметров, количество карт признаков на C1 и C3 слоях равно двум. Отметим, что модели LeNet-5-f1-f2 с 16 картами на C3 слое показывают результат не хуже, чем модели LeNet-5-vX с тремя полносвязными слоями, при этом число параметров меньше 35 тыс. Лучшей моделью из этого списка можно считать LeNet5-f1-f2-7, так как у нее 9770 параметров и точность около 87%.

Перейдем к третьей группе (с LeNet5-f1-1 по LeNet5-f1-12 из таблицы 1), в которых имеется только один полносвязный слой. Результаты показывают, что модели с одним полносвязным слоем показывают низкую точность. Лучшими моделями из этого списка будут LeNet5-f1-3 и LeNet5-f1-7 с точностью 76.97% и 75.76%, количество параметров 3438 и 1410. Выделим, что часть моделей получились недообученными и для их полного обучения уже надо менять сам подход, увеличивать количество эпох, начинать с большей скорости обучения или применять другие методы для обучения.

На рисунке 4 изображено сравнение всех модификаций исходной архитектуры LeNet-5.

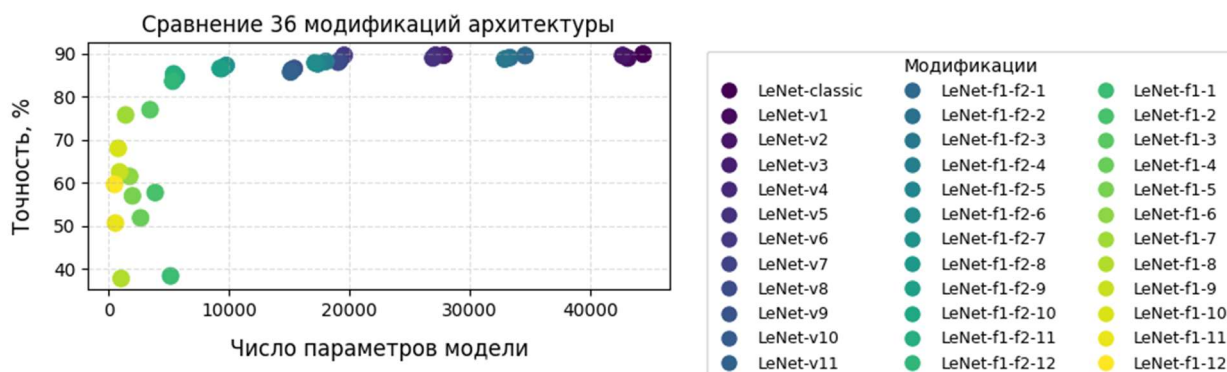


Рисунок 4 – Зависимость точности от числа параметров для модификаций СНС LeNet

Из приведенного графика мы можем сделать следующие выводы, что наличие трех полносвязных слоев приводит архитектуру к насыщению, количество параметров растет, но точность сильно не увеличивается. Если в исходной архитектуре оставить один полносвязный слой, то количество параметров уменьшится сильно, меньше 5 тыс. параметров, а точность архитектуры будет невысокая и для её увеличения надо будет пересматривать подход к обучению. Лучшим вариантом для архитектуры LeNet – наличие двух полносвязных слоев, так как при этом количество параметров уменьшается с 44 тыс. до 5 тыс., а точность спадает с 89% до 85%, но результат все равно остается удовлетворительным. Отметим результат модели LeNet5-f1-f2-4, у которой количество параметров в 2,5 раза меньше (18054), чем у оригинальной модели, а точность меньше всего на 1%. Лучшей моделью считается LeNet-v6, в которой точность 89.61%, а количество параметров — 19574.

Вывод

В результате работы исследовано как меняется точность и количество параметров архитектуре СНС LeNet-5. Можно сделать вывод, что оригинальную модель LeNet5-classic можно считать перенасыщенной, так как есть модификации этой модели с такой же точностью, как и в оригинальной, но количество параметров меньше. Например, модель LeNet-v6 имеет точность 89.61%, а количество параметров — 19574, почти в 2,5 раза меньше параметров, чем у LeNet5-classic. Отметим, что есть модели и с меньшим числом параметров, но их точность не на много меньше, чем у LeNet-v6. Данными моделями можно считать LeNet5-f1-f2-7 и LeNet5-f1-f2-11, у которых количество параметров меньше 10 тыс., а точность в районе 87,27% и 85,30%, соответственно.

Список использованных источников:

1. LeCun Y. et al. Gradient-based learning applied to document recognition // *Proceedings of the IEEE*. – 1998. – Т. 86. – №. 11. – С. 2278-2324.
2. Szegedy C. et al. Intriguing properties of neural networks // *arXiv preprint arXiv:1312.6199*. – 2013.
3. Qassim H., Verma A., Feinzimer D. Compressed residual-VGG16 CNN model for big data places image recognition // *IEEE 8th annual computing and communication workshop and conference (CCWC)*. – 2018. – С. 169-175.

INVESTIGATION OF THE EFFECT OF THE NUMBER OF PARAMETERS IN A CONVOLUTIONAL NEURAL NETWORK ON THE ACCURACY OF IMAGE RECOGNITION

Ermakovich V.A.¹

Belarusian State University of Informatics and Radioelectronics¹, Minsk, Republic of

Vashkevich M.I. – Doctor of Science

Annotation. The paper presents research on the influence of the number of parameters of a convolutional neural network on the accuracy of image recognition. The study was conducted based on the LeNet-5 architecture with a change in the number of feature maps and the number of fully connected layers. The research presents a description of this architecture using a learning rate scheduler, a table and a graph with the results of learning modifications to the original model.

Keywords. LeNet, CNN, image recognition.