

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ АРХИТЕКТУР СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ В РЕШЕНИИ ЗАДАЧИ КЛАССИФИКАЦИИ ПРИ МАЛОЙ ОБУЧАЮЩЕЙ ВЫБОРКЕ

© 2023 В. М. Рассолова¹, Е. Д. Позднякова², В. В. Лоторева³

¹студентка 2 курса бакалавриата направления подготовки
01.03.02 «Прикладная математика и информатика»
e-mail: victoria1423@yandex.ru

²студентка 2 курса бакалавриата направления подготовки
01.03.02 «Прикладная математика и информатика»
e-mail: elisabets2003@gmail.com

³научный руководитель: ассистент кафедры программного обеспечения
и администрирования информационных систем
e-mail: lotorevaviktoria@yandex.ru

Курский государственный университет

В статье представлен обзор на наиболее известные архитектуры свёрточных нейронных сетей, а также их сравнительная характеристика. Проведен анализ разработанной уникальной архитектуры и компаративный разбор. При использовании различных архитектур CNN была получена высокая точность классификации во всех исследованных случаях. Также был проведен анализ влияния гиперпараметров и их оптимального выбора на эффективность работы CNN.

Ключевые слова: искусственный интеллект, нейронные сети, задача классификации, архитектуры нейронных сетей, свёрточные нейросети.

RESEARCH OF THE EFFICIENCY OF ARCHITECTURES OF CONVOLUTIONAL NEURAL NETWORKS IN SOLVING THE PROBLEM OF CLASSIFICATION WITH A SMALL TRAINING SAMPLE

© 2023 V. M. Rassolova¹, E. D. Pozdnyakova², V. V. Lotoreva³

¹Second-year undergraduate student
majoring in 01.03.02 "Applied mathematics and informatics "
e-mail: victoria1423@yandex.ru

²Second-year undergraduate student
majoring in 01.03.02 "Applied mathematics and informatics "
e-mail: elisabets2003@gmail.com

³Supervisor: Assistant of department of Programming Support
and Information System Administration
e-mail: lotorevaviktoria@yandex.ru

Kursk State University

The article provides an overview of the most famous convolutional neural networks architectures, as well as their comparative characteristics. An analysis of the developed unique architecture is made, a comparative analysis is carried out. When using various CNN

architectures, a high classification accuracy was obtained in all cases studied. An analysis was also made of the influence of hyperparameters and their optimal choice on the efficiency of the CNN.

Keywords: artificial intelligence; neural networks; classification task; neural network architectures; convolutional neural networks.

Свёрточные нейронные сети (Convolutional Neural Networks, CNN) являются одним из наиболее эффективных инструментов в области компьютерного зрения, созданным специально для обработки данных с пространственной структурой, таких как изображения. CNN широко применяются для задач детекции, то есть поиска и выделения объекта на изображении, сегментирования и классификации. Благодаря своей архитектуре, сети данного класса способны эффективно обучаться и выделять характеристики изображений или видео.

Основным компонентом CNN является свёрточный слой, который выполняет операцию извлечения признаков из входных данных. Свёртка представляет собой поэлементное перемножение фильтра (ядра свёртки) с соответствующими пикселями входного изображения, а затем суммирование результатов. Другой важный компонент CNN – это слой пулинга (подвыборки), который уменьшает размерность выходных данных свёрточного слоя путем выбора наиболее значимых признаков. Завершающими слоями архитектуры являются, как правило, полносвязные слои с Softmax-активацией для классификации, в которых происходит соединение всех признаков, извлеченных предыдущими слоями, и преобразование их в окончательный результат, например, класс объекта [1, с. 1–6]. Обычно в подобных архитектурах используется большое количество слоёв, что подходит для решения задачи классификации при масштабном объёме обучающих данных.

Существует множество известных свёрточных нейронных сетей, которые с высокой точностью выполняют задачу классификации, например, ResNet [2, с. 770–778], AlexNet [3, с. 84–90], VGG [4] и Inception [5, с. 1–9].

В рамках выпускного проекта ИТ Академии Samsung «Разработка интеллектуальной системы классификации археологических контекстов на основе анализа керамики с использованием искусственных нейронных сетей» за авторством Е.Д. Поздняковой и В.М. Рассоловой была поставлена цель разработать интеллектуальную систему для распознавания древних керамических контекстов. После изучения разнообразных вариантов архитектур был сделан выбор в пользу разработки своей собственной структуры, которая была бы столь же эффективна, но при этом содержала минимально необходимое количество параметров для быстрого и качественного обучения. Таким образом, для решения задачи классификации была создана уникальная архитектура на основе CNN. Конфигурация разработанной сети состоит из следующих компонентов:

- 1) 4 свёрточных слоя, которые использовались для уменьшения изображения, выделения признаков и их обобщения;
- 2) после каждого слоя свертки применяется макс-пулинг, необходимый для уменьшения полученных признаков по их пространственным размерам (высоте и ширине);
- 3) слой преобразования матрицы признаков в вектор;
- 4) и в завершение полносвязный слой для классификации.

После проведения многочисленных манипуляций с изменениями гиперпараметров сети, таких как скорость обучения (learning rate) и количество эпох, и тестов с различными оптимизаторами и функциями активации, была достигнута точность порядка 97% на валидационной выборке. Функцией активации была выбрана

SiLU, а шаг обучения был принят за 0.001. 20 эпох оказалось достаточно для продуктивной работы сети. Созданная нейросеть продемонстрировала свою эффективность при выполнении поставленной задачи, тем не менее было принято решение провести сравнительный анализ ее работы с другими сетями с целью выявления наиболее эффективной из них и определения, является ли разработка специализированной архитектуры необходимой для обучения на небольших датасетах.

Выбор набора данных для обучения является одним из наиболее значимых этапов при работе с нейросетями. В проведенном исследовании все модели обучались на датасете, который ранее был использован при реализации выпускного проекта. Этот набор данных был собран в Эрмитаже и состоял из 100 изображений старинных керамических изделий. Примеры некоторых изображений из предоставленного датасета изображены на рисунке 1.



Рис. 1. Примеры изображений, полученных из Эрмитажа

Датасет был размечен специалистами, которые разбили изображения на группы по их морфологическим признакам. В итоге были выделены 6 классов: эсвидный венчик, вертикальный венчик, миски, сковороды, кружки и индивидуальные формы.

Для сравнения были выбраны сети DenseNet, AlexNet и ShuffleNet. Это решение обусловлено тем, что все три модели являются сверточными нейронными сетями и достигают хороших результатов в задаче классификации изображений, а также предположительно способны обучиться на малой выборке. DenseNet, AlexNet и ShuffleNet имеют различные архитектуры и особенности, что делает их интересными для сравнения. Стоит заметить, что DenseNet является глубокой моделью с большим количеством слоев, AlexNet также имеет несколько слоев, но меньше по сравнению с DenseNet, а ShuffleNet является более легкой моделью с низким количеством параметров. Таким образом, выбор данных моделей для анализа их эффективности на малой обучающей выборке обусловлен комбинацией их архитектуры, глубины и производительности. Это позволяет оценить, как разные подходы к построению нейронных сетей влияют на результаты обучения на малом объеме данных.

AlexNet является одной из первых моделей, которая продемонстрировала высокую производительность в задаче классификации изображений [3, с. 84–90].

Данная сеть имеет глубокую архитектуру, состоящую из 5 сверточных слоёв, 3 слоёв макс-пулинга и 3 полносвязных слоёв. Для улучшения скорости обучения и преодоления проблемы затухания градиента использовалась активационная функция ReLU. Впервые была применена техника Dropout, идея которой заключается в исключении определенного процента случайных нейронов во время обучения нейросети на разных эпохах. Эта техника призвана помочь в борьбе с переобучением. Среди всех исследуемых моделей AlexNet требуется обучить наибольшее количество параметров. Архитектура нейросети представлена на рисунке 2.

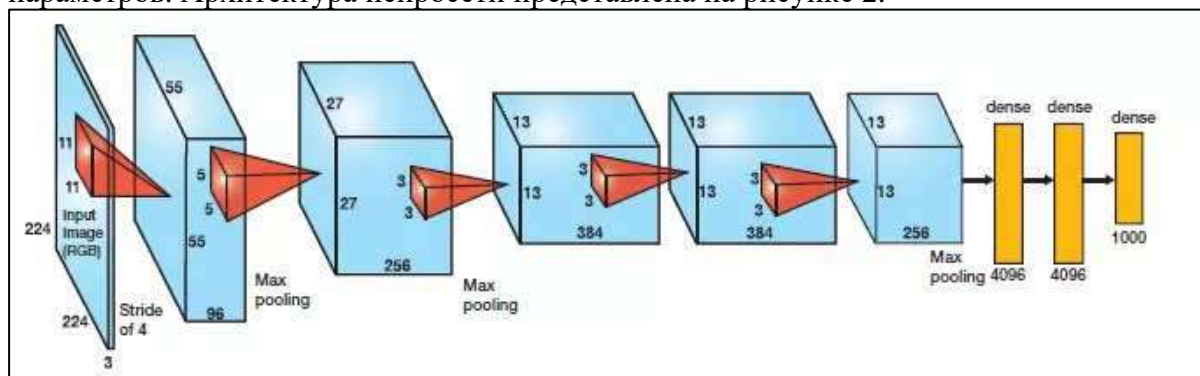


Рис. 2. Архитектура AlexNet

Графики поведения функций потерь и точности представлены соответственно на рисунках 3 и 4.

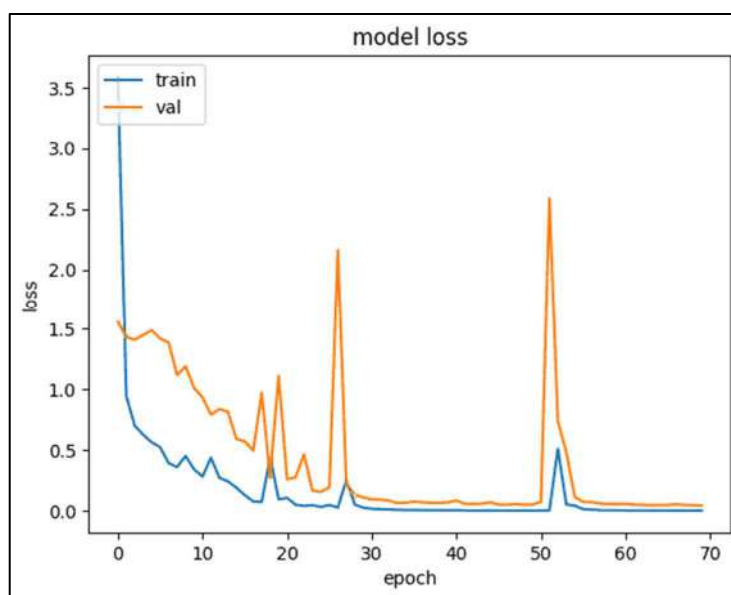


Рис. 3. График изменения функции потерь AlexNet

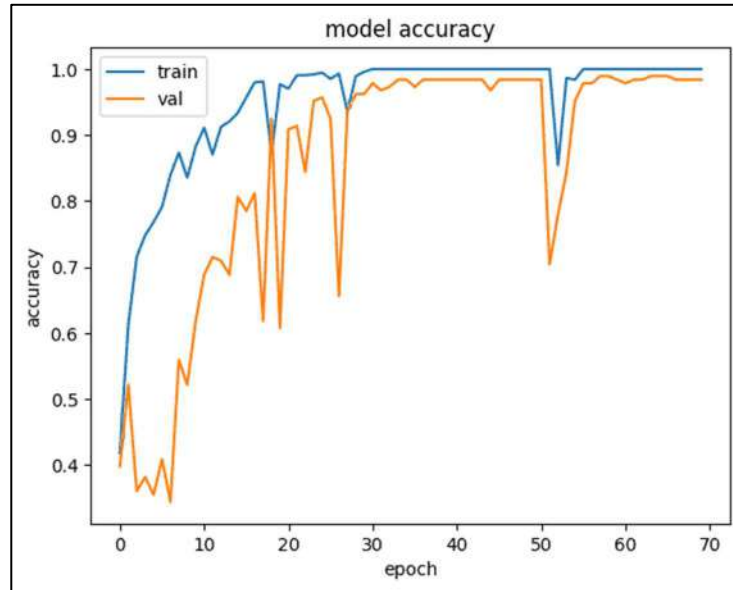


Рис. 4. График изменения функции точности AlexNet

DenseNet – это CNN с плотными связями, состоящая в основном из DenseBlock и Transition [6]. В DenseNet каждый слой имеет прямые соединения со всеми предыдущими слоями, что позволяет эффективно обмениваться информацией и извлекать общие и детализированные признаки. Это приводит к улучшению точности классификации и уменьшению количества параметров по сравнению с аналогичными моделями. Подробная архитектура DenseNet представлена на рисунке 5.

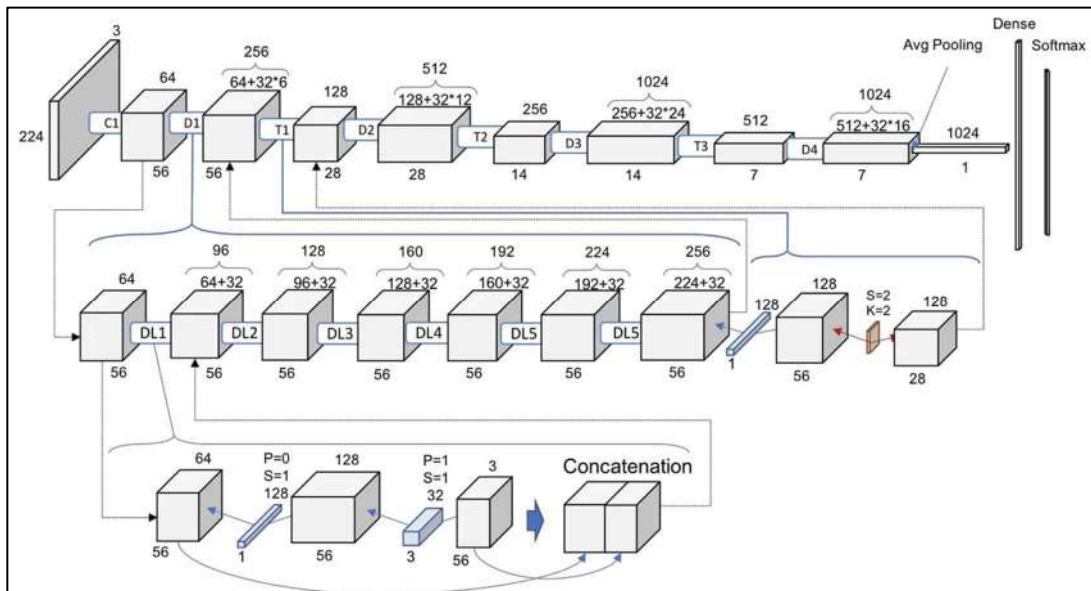


Рис. 5. Архитектура DenseNet

После ряда изменений в гиперпараметрах сети удалось добиться точности порядка 80% на валидационной выборке; графики изменения функций потерь и точности в ходе обучения нейросети изображены на рисунках 6 и 7.

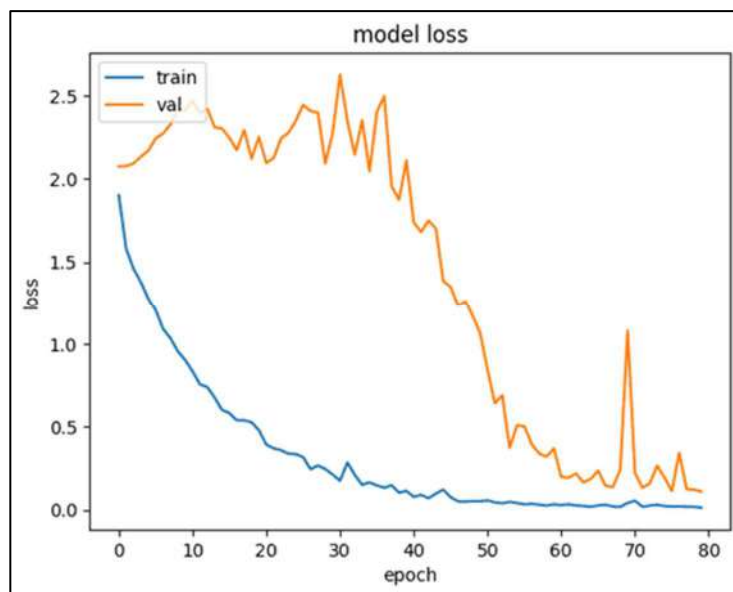


Рис. 6. График изменения функции потерь DenseNet

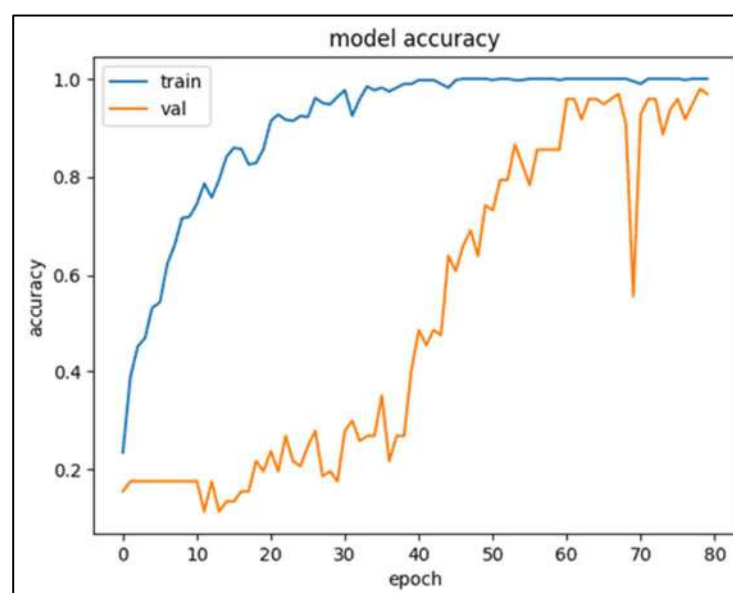


Рис. 7. График изменения функции точности DenseNet

ShuffleNet является эффективной архитектурой сверточной нейронной сети, разработанной для выполнения вычислительно интенсивных задач компьютерного зрения с минимальным количеством параметров [7]. Её ключевой особенностью является использование операции перемешивания – shuffle в сверточных слоях, которая способствует эффективному обмену информацией между разными группами признаков. Это позволяет ShuffleNet достигать высокой точности при классификации изображений и других задачах, при этом не требуя огромных вычислительных ресурсов. Архитектура представлена на рисунке 8.

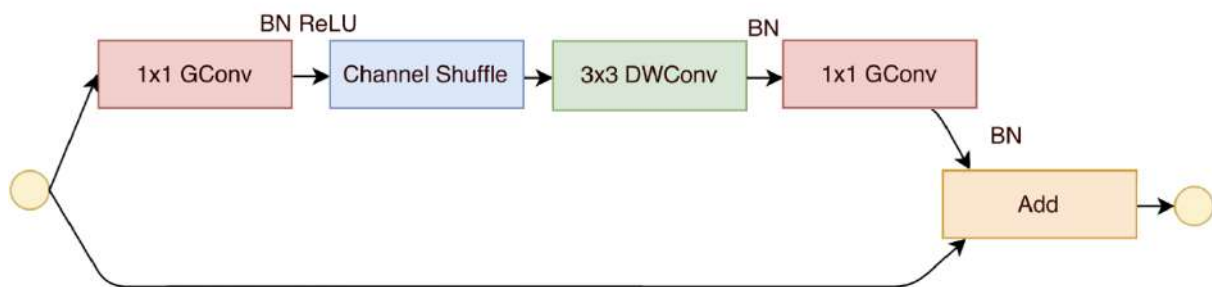


Рис. 8. Архитектура ShuffleNet

Графики изменения функций потерь и точности представлены соответственно на рисунках 9 и 10.

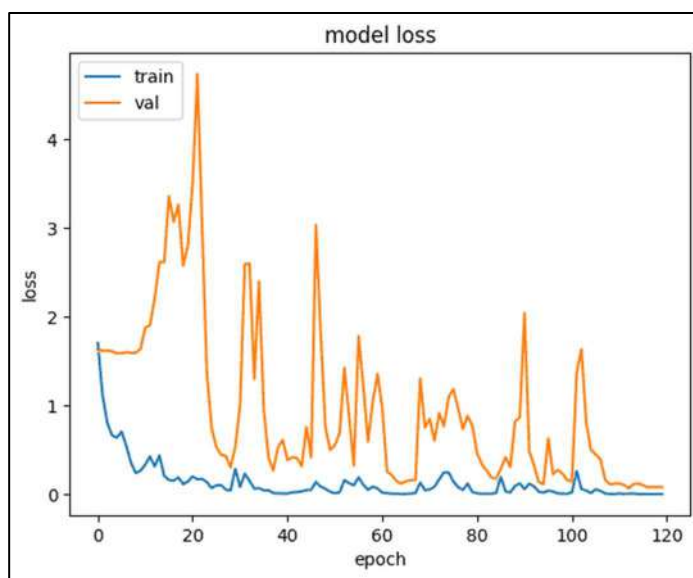


Рис. 9. График изменения функции потерь ShuffleNet

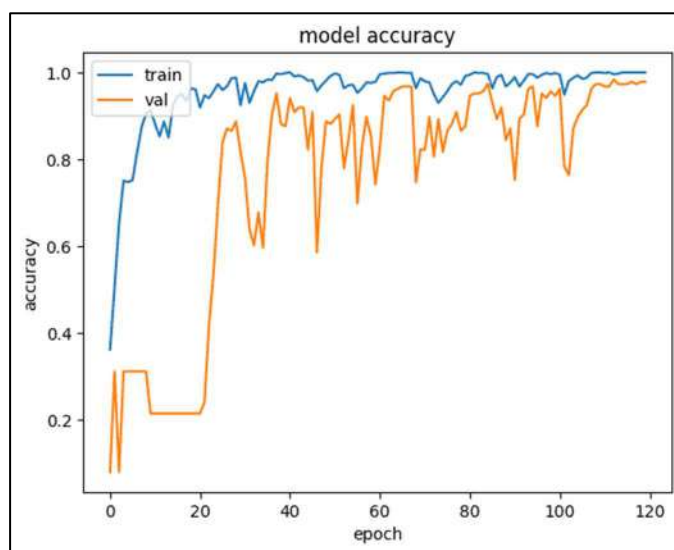


Рис. 10. График изменения функции точности ShuffleNet

В ходе обучения моделей была замечена так называемая поздняя сходимость – явление, при котором значение функций потерь и точности перестает изменяться по мере обучения. Значения функций достигают того показателя, который не могут преодолеть даже при дальнейшем обучении. Такое явление объяснимо небольшим размером датасета и переобучением нейросетей.

Результаты сравнительного анализа нейросетей

Нейросеть	Количество параметров	Количество эпох	Точность (валидация)	Потери (валидация)
AlexNet	58,281,029	70	0.9839	0.0420
DenseNet	7,042,245	80	0.8227	0.1954
ShuffleNet	961,757	120	0.9785	0.0774
Разработанная модель	2,336,530	20	0.9941	0.0305

По итогам проведенного исследования самую низкую точность демонстрирует DenseNet – 82%. Известно, что более современные архитектуры обычно имеют более глубокие и сложные структуры, состоящие из множества сверточных и полносвязных слоев, чтобы улучшить качество распознавания. Этим сетям требуется больше вычислительных ресурсов и времени для обучения, особенно если доступен ограниченный набор исходных данных. Большое количество параметров также может привести к переобучению, если датасет недостаточен или не сбалансирован, что объясняет полученный результат.

ShuffleNet, наоборот, продемонстрировала высокую точность. Благодаря простой структуре и небольшому количеству обучаемых параметров, она прекрасно подходит для обучения на малых наборах данных, так как не склонна к переобучению. AlexNet также хорошо справляется с поставленной задачей, хотя ей по сравнению с другими моделями требуется обучить большое количество параметров. Наилучший результат получен у разработанной нейросети, поскольку она собрала в себе все преимущества небольших архитектур.

Таким образом, при выборе модели сверточной нейронной сети для конкретной задачи необходимо учитывать доступные ресурсы, размер датасета и требования к точности и скорости работы модели. Исследование показало, что при малом наборе обучающих данных использование минималистичных архитектур CNN, таких как созданная модель или ShuffleNet, является лучшим выбором.

Библиографический список

1. Understanding of a convolutional neural network / S. Albawi, T. A. Mohammed, S. Al-Zawi. – DOI 10.1109/icengtechnol.2017.8308186. – Текст : электронный // 2017 International Conference on Engineering and Technology (ICET). – 2017. – С. 1–6.
2. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren, J. Sun. — DOI 10.1109/cvpr.2016.90. — Текст : электронный // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – С. 770–778.
3. ImageNet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. E. Hinton. — DOI 10.1145/3065386. — Текст : электронный // Communications of the ACM. – 2017. – Т. 60, вып. 6. – С. 84–90.
4. Very Deep Convolutional Networks for Large-Scale Image Recognition / K. Simonyan, A. Zisserman. — DOI 10.48550/arXiv.1409.1556. — Текст : электронный // International Conference on Learning Representations (ICLR). — 2015.

5. Going deeper with convolutions / C. Szegedy, L. Wei, J. Yangqing [и др.]. – DOI 10.1109/cvpr.2015.7298594. – Текст : электронный // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – С. 1–9.

6. Densely Connected Convolutional Networks / G. Huang, Z. Liu, L. D. Van, K. Q. Weinberger. – DOI 10.1109/cvpr.2017.243. – Текст : электронный // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017.

7. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices / X. Zhang, X. Zhou, M. Lin, J. Sun. – DOI 10.1109/cvpr.2018.00716. – Текст : электронный // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2018.