

Анализ эффективности нейросетей для классификации изображений

¹*СТЕЦЕНКО Алексей Иванович, магистрант, mrgreen348@gmail.com,

¹СМАГУЛОВА Асемгуль Сериковна, к.т.н., доцент, assemgul_work@mail.ru,

¹МУХАМЕТЖАНОВА Бигуль Олжабаевна, PhD, старший преподаватель, grek79@mail.ru,

¹НАО «Карагандинский технический университет имени Абылкаса Сагинова», Казахстан, Караганда, пр. Н. Назарбаева, 56,

*автор-корреспондент.

Аннотация. Рассматриваются аспекты классификации изображений. Приводится описание принципов построения сверточных нейронных сетей. Рассмотрен принцип работы базовой сверточной нейронной сети. Изучены основные понятия, необходимые для работы с нейросетями. Предлагается общий обзор нейросетей для классификации изображений. Рассмотрены такие нейронные сети, как: LeNet, AlexNet, VGG image Net, Enet. Особое внимание уделяется архитектуре нейросетей. Выделены их отличительные свойства, достоинства и недостатки нейросетей и проведен анализ их эффективности.

Ключевые слова: классификация изображений, анализ нейросетей, архитектура нейросетей, свёрточные нейронные сети, свёрточные слои, компьютерное зрение, категоризация изображений, глубокие нейросети, распознавание изображений, глубокое обучение.

Введение

Задача классификации изображений входит в более широкую задачу компьютерного зрения. Компьютерное зрение – раздел искусственного интеллекта, который занимается извлечением информации из изображений, причем изображения могут быть разного типа. Цель компьютерного зрения – извлечение полезных выводов относительно объектов и сцен реального мира, основываясь на анализе изображений, полученных с помощью датчиков.

Различные системы технического зрения уже сейчас решают задачи по дополнению, а иногда даже замене человека в совершенно различных областях деятельности, непосредственно связанных со сбором и анализом зрительной информации. Данные успехи сегодня представить намного проще благодаря стремительному развитию свёрточных нейронных сетей.

Свёрточная нейронная сеть – нейронная сеть, в которой присутствует слой свёртки (convolutional layer). Обычно в свёрточных нейронных сетях также присутствуют слой субдискретизации (pooling layer) и полносвязный слой (fully connected layer). Свёрточные нейронные сети применяются для оптического распознавания образов, классификации изображений, детектирования предметов, семантической сегментации [1].

В свёрточных нейронных сетях слои свёртки и субдискретизации состоят из нескольких «уров-

ней» нейронов, называемых картами признаков (feature maps), или каналами (channels). Каждый нейрон такого слоя соединён с небольшим участком предыдущего слоя, называемым рецептивным полем. В случае изображения карта признаков является двумерным массивом нейронов или просто матрицей. Другие измерения могут быть использованы, если на вход принимается другой вид данных, например, аудиоданные (одномерный массив) или объёмные данные (трёхмерный массив).

В слое свёртки каждой карте признаков соответствует одно ядро свёртки, также называемое фильтром. Каждый нейрон в качестве своего выходного значения осуществляет операцию свёртки или взаимной корреляции со своим рецептивным слоем. Стоит заметить, что эти две операции в контексте обучения свёрточных нейронных сетей взаимозаменяемы, вследствие чего во многих программных реализациях операция «свёртки» на самом деле является операцией взаимной корреляции [2].

Так как ядро свёртки для каждой карты признаков одно, это позволяет нейронной сети научиться выделять признаки вне зависимости от их расположения во входном изображении и также приводит к значительному уменьшению числа параметров. Согласно устоявшейся нотации, говорят, что слой свёртки использует фильтр $W \times H$, если каждый фильтр в этом слое имеет раз-

мерность $W \times H \times C$, где C – число каналов в предыдущем слое.

Слой субдискретизации осуществляет уплотнение карт признаков предыдущего слоя и не изменяет количество карт. Каждая карта признаков слоя соединена с соответствующей картой признаков предыдущего слоя, каждый нейрон выполняет «сжатие» своего рецептивного поля посредством какой-либо функции.

Наиболее популярными видами этого слоя являются Max Pooling (из рецептивного слоя выбирается максимальное значение), Average Pooling (выбирается среднее значение) и L2 Pooling (выбирается норма L2). С помощью слоя субдискретизации достигается устойчивость к небольшим сдвигам входного изображения, а также уменьшается размерность последующих слоёв. Полносвязный слой – обычный скрытый слой многослойного перцептрона, соединённый со всеми нейронами предыдущего слоя. Таким образом, на вход свёрточной нейронной сети подаётся изображение, а на выходе получается класс, к которому принадлежит изображение [3].

Описание методов исследования

В первую очередь, исследование фокусируется на сравнении свёрточных нейронных сетей, предназначенных для классификации изображений. Для того чтобы выделить их сильные и слабые стороны, преимущества и недостатки, был проведен анализ вторичных данных из рейтинга PapersWithCode с использованием датасета ImageNet [4].

Архитектуры нейросетей

Основы современной архитектуры свёрточных нейронных сетей были заложены в одной из первой широко известной свёрточной нейронной сети – LeNet. LeNet – это свёрточная нейронная сеть (convolutional neural network, CNN, LeNet), которая впервые была представлена в 1998 году французским исследователем Яном Лекуном (Yann LeCun) как развитие модели неокогнитрон (neocognitron).

Модель свёрточной сети состоит из трёх типов слоёв: свёрточные (convolutional) слои, субдискретизирующие (subsampling, подвыборка)

слои и слои «обычной» нейронной сети – перцептрона. Первые два типа слоёв (convolutional, subsampling), чередуясь между собой, формируют входной вектор признаков для многослойного перцептрона. Сеть можно обучать с помощью градиентных методов [5].

LeNet можно описать как первую свёрточную нейронную сеть, и она добилась хороших результатов в распознавании рукописных цифр. Некоторые проблемы, с которыми сталкиваются пиксели изображения в качестве входных данных нейронной сети:

1) Объем данных изображения велик, и отдельные пиксели независимо вводятся в нейрон, что требует большой структуры сети и слишком большого количества параметров обучения.

2) Деформация изображения, проблема изменения положения изображения, вызванная деформацией.

3) Локальная корреляция изображения.

Благодаря LeNet было введено понятие свёрточной нейронной сети и применены такие особенности, как местное рецептивное поле, разделение веса и Downsampling (объединение). Архитектура LeNet представлена на рисунке 1.

AlexNet

AlexNet – свёрточная нейронная сеть для классификации изображений. Она оказала большое влияние на развитие машинного обучения, в особенности – на алгоритмы компьютерного зрения. Сеть с большим отрывом выиграла конкурс по распознаванию изображений ImageNet LSVRC (с количеством ошибок 15,3% против 26,2% у второго места).

Архитектура AlexNet схожа с созданной Yann LeCun сетью LeNet. Однако у AlexNet больше фильтров на слое и вложенных свёрточных слоев. Сеть включает в себя свертки, максимальное объединение, дропаут, аугментацию данных, функции активаций ReLU и стохастический градиентный спуск [6]. Архитектура AlexNet представлена на рисунке 2.

Особенности AlexNet

Как функция активации используется Relu вместо арктангенса для добавления в модель нелинейности. За счет этого при одинаковой точно-

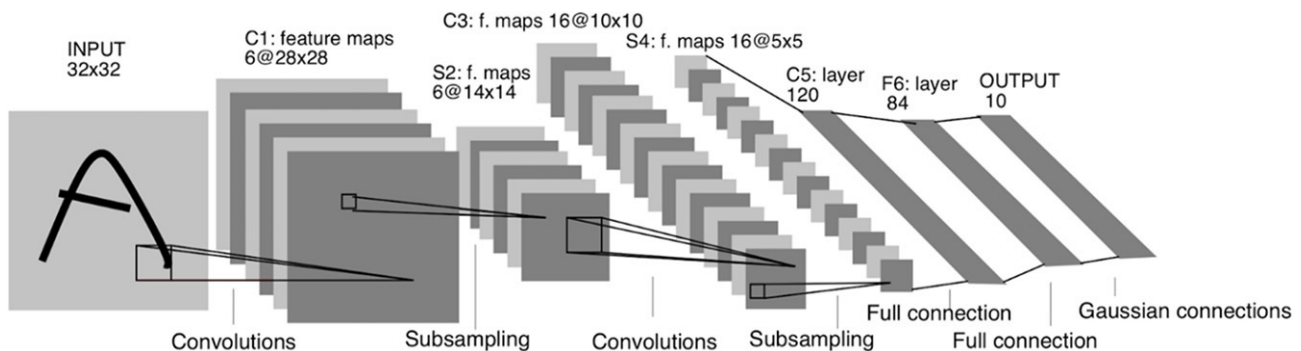


Рисунок 1 – Архитектура нейросети LeNet

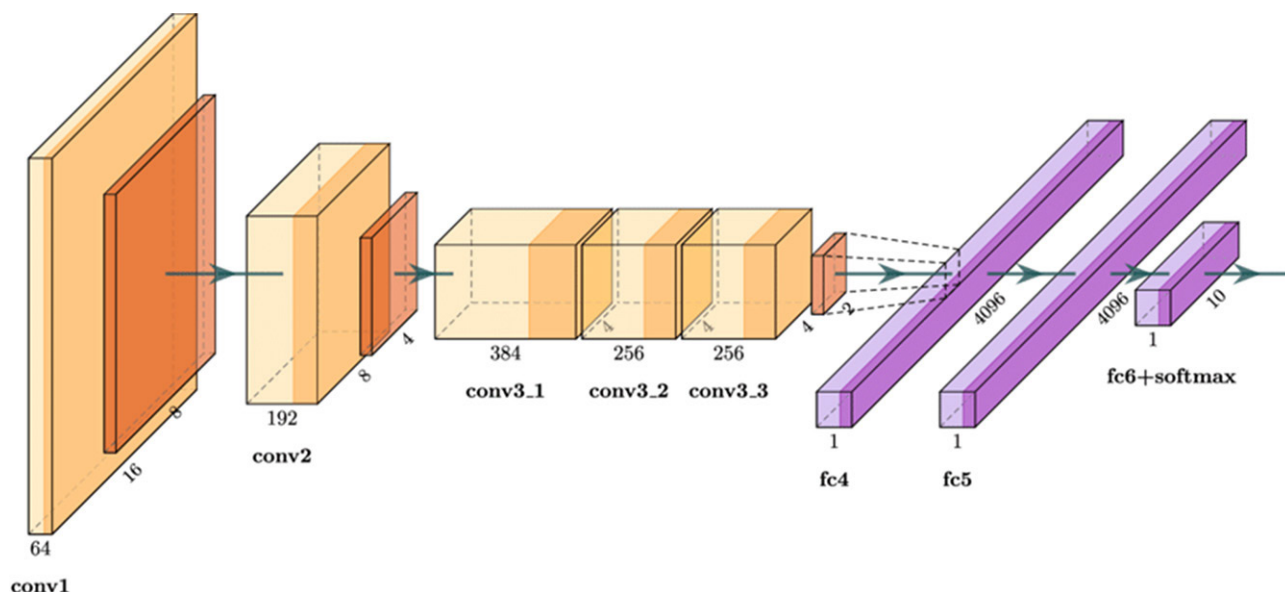


Рисунок 2 – Архитектура нейросети AlexNet

сти метода скорость становится в 6 раз быстрее. Использование дропаута вместо регуляризации решает проблему переобучения. Однако время обучения удваивается с показателем дропаута 0,5. Производится перекрытие объединений для уменьшения размера сети. За счет этого уровни ошибок первого и пятого уровней снижаются до 0,4% и 0,3%, соответственно.

Результаты показывают, что большая, глубокая сверточная нейронная сеть способна достигать рекордных результатов на очень сложных датасетах, используя только обучение с учителем. Через год после публикации AlexNet все участники конкурса ImageNet стали использовать сверточные нейронные сети для решения задачи классификации. AlexNet была одной из первых реализаций сверточных нейронных сетей и открыла новую эру исследований [7]. Сейчас её использование нецелесообразно ввиду низкого уровня производительности и наличия более простых способов реализации.

VGG Net

В разработанных в Оксфорде VGG-сетях в каждом свёрточном слое впервые применили фильтры 3x3, да ещё и объединили эти слои в последовательности свёрток.

Это противоречит заложенным в LeNet принципам, согласно которым большие свёртки использовались для извлечения одинаковых свойств изображения. Вместо применяемых в AlexNet фильтров 9x9 и 11x11 стали применять гораздо более мелкие фильтры, опасно близкие к свёрткам 1x1, которых старались избежать авторы LeNet, по крайней мере в первых слоях сети. Но большим преимуществом VGG стала находка, что несколько свёрток 3x3, объединённых в последовательность, могут эмулировать более крупные рецептивные поля, например, 5x5 или 7x7 [8]. Эти

идеи позднее vision будут использованы в архитектурах Inception и ResNet. Архитектура VGG Net представлена на рисунке 3.

Сети VGG для представления сложных свойств используют многочисленные свёрточные слои 3x3. Это даёт нам огромное количество параметров и прекрасные способности к обучению. Но учить такие сети было сложно, приходилось разбивать их на более мелкие, добавляя слои один за другим. Причина заключалась в отсутствии эффективных способов регуляризации моделей или каких-то методов ограничения большого пространства поиска, которому способствует множество параметров [9].

VGG во многих слоях используют большое количество свойств, поэтому обучение требовало больших вычислительных затрат. Снизить нагрузку можно, уменьшив количество свойств, как это сделано в bottleneck-слоях архитектуры Inception.

ENet

Все особенности описанных архитектур были скомбинированы в очень эффективную и компактную сеть, использующую совсем мало параметров и вычислительных мощностей, но при этом дающую превосходные результаты. Архитектура получила название ENet, её разработал Адам Пазке (Adam Paszke). ENet – это сеть на основе кодировщика и декодера. Кодировщик построен по обычной схеме CNN для категоризации, а декодер представляет собой сеть с повышением дискретизации (upsampling network), предназначенную для сегментирования посредством распространения категорий обратно в изображение исходного размера. Для сегментации изображений использовались только нейросети, никаких других алгоритмов.

Архитектура сети представлена на рисунке 4. Она разделена на несколько этапов, как выделено

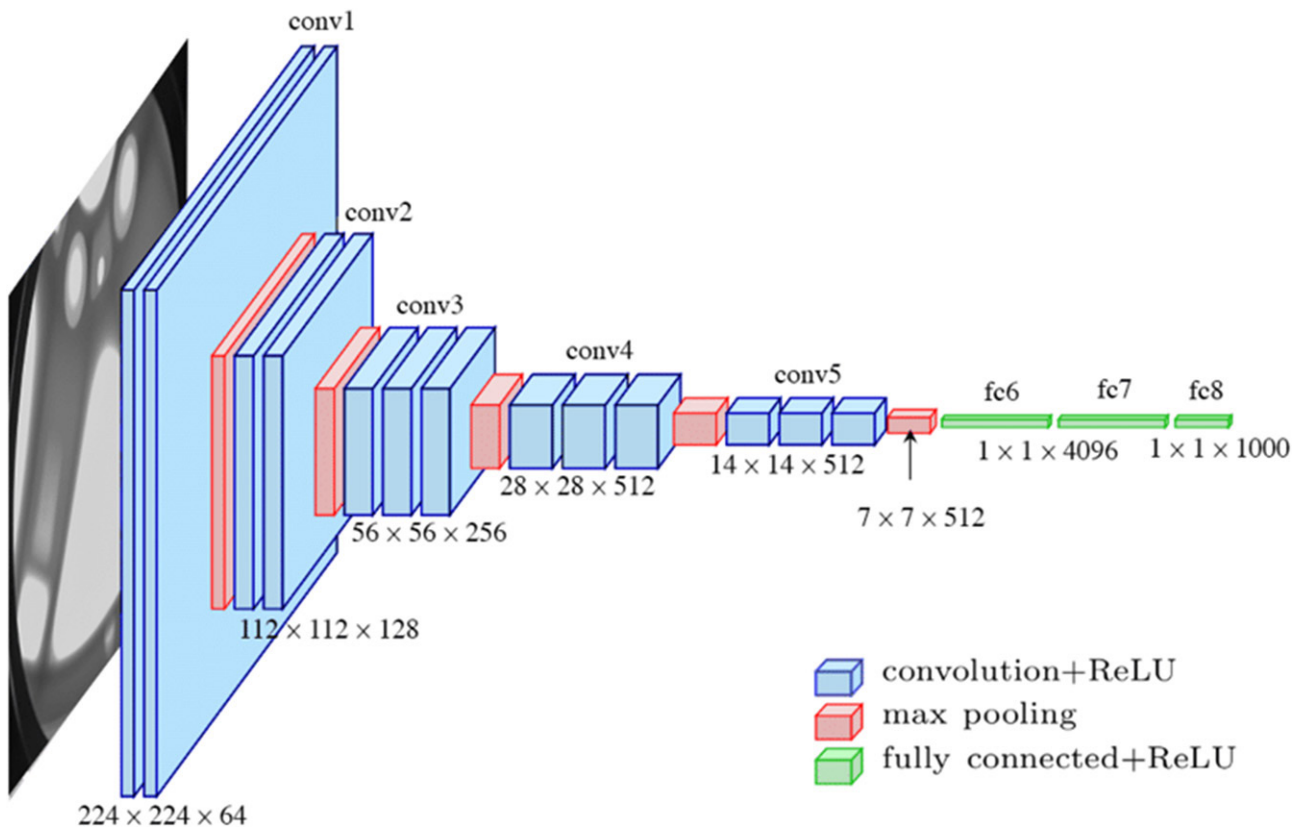


Рисунок 3 – Архитектура нейросети VGG Net

Name	Type	Output size
initial		16 × 256 × 256
bottleneck1.0	downsampling	64 × 128 × 128
4 × bottleneck1.x		64 × 128 × 128
bottleneck2.0	downsampling	128 × 64 × 64
bottleneck2.1		128 × 64 × 64
bottleneck2.2	dilated 2	128 × 64 × 64
bottleneck2.3	asymmetric 5	128 × 64 × 64
bottleneck2.4	dilated 4	128 × 64 × 64
bottleneck2.5		128 × 64 × 64
bottleneck2.6	dilated 8	128 × 64 × 64
bottleneck2.7	asymmetric 5	128 × 64 × 64
bottleneck2.8	dilated 16	128 × 64 × 64
<i>Repeat section 2, without bottleneck2.0</i>		
bottleneck4.0	upsampling	64 × 128 × 128
bottleneck4.1		64 × 128 × 128
bottleneck4.2		64 × 128 × 128
bottleneck5.0	upsampling	16 × 256 × 256
bottleneck5.1		16 × 256 × 256
fullconv		C × 512 × 512

Рисунок 4 – Архитектура нейросети ENet

размеры указаны для разрешения входного изображения 512x512. Мы принимаем представление ResNets, которое описывает как имеющие одну основную ветвь и расширения со сверточными фильтрами, которые отделяются от нее, а затем объединяются обратно с поэлементным добавлением. Каждый блок состоит из трех сверточных слоев: проекция 1x1, уменьшающая размерность, основной сверточный слой и расширение 1x1.

Начальный блок ENet MaxPooling выполняет с неперекрывающимися окнами 2x2, а свертка имеет 13 фильтров, что после конкатенации суммирует до 16 карт признаков. Модуль bottleneck conv представляет собой либо обычную расширенную, либо полную свертку (также известную как деконволюция) с фильтрами 3x3, либо свертку 5x5, разложенную на две асимметричные.

ENet дает возможность выполнять попиксельную семантическую сегментацию в режиме реального времени. ENet до 18 раз быстрее, требует в 75 раз меньше FLOP, имеет в 79 раз меньше параметров и обеспечивает аналогичную или лучшую точность по сравнению с существующими моделями [10].

Сравнительный анализ нейросетей

Для поиска лучшей нейросети для распознавания изображений был проведен анализ достоинств и недостатков упомянутых выше нейросетей. В таблице приведен сравнительный анализ упомянутых нейросетей.

горизонтальными линиями в таблице и первой цифрой после каждого имени блока. Выходные

Сравнительный анализ приведенных нейросетей		
Названия нейросети	Достоинства	Недостатки
LeNet	Заложила основы сверточных нейронных сетей. Высокая точность распознавания рукописных цифр.	Необходимо большое количество времени для обучения и составления прогнозов.
AlexNet	Использование свертков, дропаута, аугментации данных. Высокая точность распознавания изображений.	Низкий уровень производительности. Высокая сложность реализации.
VGG Net	Впервые применены фильтры 3x3 в каждом сверточном слое. Применение большого количества фильтров.	Высокие требования к вычислительным мощностям. Приходится добавлять слои усложняя архитектуру нейросети.
ENet	Использование кодировщика и декодера. Высокий уровень производительности. Высокая точность распознавания изображений.	Высокие требования к вычислительным мощностям.

Заключение

Нейросети для классификации изображений прошли большой путь развития, от исследования небольшой группой ученых, до использования большими ИТ-корпорациями.

Цель статьи – дать представление о вариантах создания архитектур, которые могут привести к эффективным нейронным сетям для практического применения, и оптимизации часто ограниченных ресурсов. Таким образом, в данной статье проведен анализ нескольких современных глубоких нейронных сетей для классификации изображений с точки зрения точности, объема памяти, параметров, количества операций. Описаны одни из самых значимых архитектур сверточных нейронных сетей для задачи классификации изображений. Они позволили сильно улучшить

точность распознавания изображений и достичь результатов, не достигнутых классическими методами компьютерного зрения.

Исходя из рассмотренных архитектур можно сделать вывод, что точность и время вывода находятся в гиперболической зависимости: небольшое увеличение точности стоит больших вычислительных затрат. Количество операций в сетевой модели может эффективно оценить время вывода. Лучшей архитектурой с точки зрения использования пространства параметров является ENet, эффективность распознавания которой до 13 раз по сравнению с моделью AlexNet и в 24 раза по сравнению с VGG-19, что позволяет рекомендовать её для создания нейросетей для классификации изображений.

СПИСОК ЛИТЕРАТУРЫ

1. Паттанаяк С. Глубокое обучение и TensorFlow для профессионалов. М.: Изд-во Диалектика, 2019, 480 с.
2. Исследование неупорядоченных данных на основе технологии SPARK./ Шаманбаев Ж.Е, Смагулова А.С. // Республиканский научно-технический журнал «Университет еңбектері – Труды университета», 2020. Вып. 4. С. 156-159.
3. Аггарвал Ч. Нейронные сети и глубокое обучение: учебный курс. М.: Изд-во Диалектика-Вильямс, 2020, 752 с.
4. Хантер А. Классификация изображений ImageNet. URL: <https://paperswithcode.com/sota/image-classification-on-imagenet> (дата обращения 24.04.2022).
5. Потапов А. Автоматический анализ изображений и распознавание образов. М.: Изд-во LAP Lambert Academic Publishing, 2017, 292 с.
6. Хайкин С. Нейронные сети. Полный курс. М.: Изд-во Вильямс, 2018, 1104 с.
7. Зеленцов И.А. Распознавание образов. Обзорная лекция. URL: http://it-claim.ru/Persons/Zelencov/Lecture_presentation.pdf (дата обращения 20.03.2022).
8. Нейронные сети: история развития и перспективы применения / Кулянин Е.М. // Научно-методический электронный журнал «Концепт». М., 2018. Вып. 13. С. 2646-2650.
9. Виды архитектур нейронных сетей / Москалев Н.С. // Молодой ученый. М., 2017, Вып. 29, С. 30-34.
10. Бишоп К.М. Распознавание образов и машинное обучение. М.: Изд-во Диалектика-Вильямс, 2020, 960 с.

Кескінді жіктеу үшін нейрондық желілер тиімділігін талдау¹*СТЕЦЕНКО Алексей Иванович, магистрант, mrgreen348@gmail.com,¹СМАГУЛОВА Асемгуль Сериковна, т.ф.к., доцент, assemgul_work@mail.ru,¹МҰХАМЕТЖАНОВА Бигуль Олжабаевна, PhD, аға оқытушы, grek79@mail.ru,¹«Әбілқас Сағынов атындағы Қарағанды техникалық университеті» КеАҚ, Қазақстан, Қарағанды, Н. Назарбаев даңғылы, 56,

*автор-корреспондент.

Аңдатпа. Кескін классификациясының аспектілері қарастырылады. Конволюциялық нейрондық желілерді құру принциптерінің сипаттамасы берілген. Негізгі конволюционды нейрондық желінің жұмыс істеу принципі қарастырылады. Нейрондық желілермен жұмыс істеуге қажетті негізгі түсініктер оқытылады. Кескіндерді жіктеу үшін нейрондық желілерге жалпы шолу ұсынылады. LeNet, AlexNet, VGG image Net, Enet сияқты нейрондық желілер қарастырылады. Нейрондық желілердің архитектурасына ерекше көңіл бөлінеді. Олардың айырықша қасиеттері, нейрондық желілердің артықшылықтары мен кемшіліктері атап көрсетілді және олардың тиімділігіне талдау жасалды.

Кілт сөздер: кескінді жіктеу, нейрондық желіні талдау, нейрондық желі архитектурасы, конволюциялық нейрондық желілер, конволюциялық қабаттар, компьютерлік көру, кескіндерді санаттау, терең нейрондық желілер, кескінді ажырату, терең оқыту.

Analysis of Existing Neural Networks for Image Classification¹*STETSENKO Alexei, Master Student, mrgreen348@gmail.com,¹SMAGULOVA Assemgul, Cand. of Tech. Sci., Associate Professor, assemgul_work@mail.ru,¹MUKHAMETZHANOVA Bigul, PhD, Senior Lecturer, grek79@mail.ru,¹NPISC «Abylkas Saginov Karaganda Technical University», Kazakhstan, Karaganda, N. Nazarbayev Avenue, 56,

*corresponding author.

Abstract. Aspects of image classification are considered. The description of the principles of construction of convolutional neural networks is given. The principle of operation of the basic convolutional neural network is considered. The basic concepts necessary for working with neural networks are studied. A general overview of neural networks for image classification is proposed. Considered such neural networks as: LeNet, AlexNet, VGG image Net, Enet. Particular attention is paid to the architecture of neural networks. Their distinctive properties, advantages and disadvantages of neural networks were highlighted and an analysis of their effectiveness was carried out.

Keywords: image classification, neural network analysis, neural network architecture, convolutional neural networks, convolutional layers, computer vision, image categorization, deep neural networks, image recognition, deep learning.

REFERENCES

1. Pattanajak C. Glubokoe obuchenie i TensorFlow dlja professionalov [Deep learning and TensorFlow for professionals]. Moscow: Publ. Dialektika, 2019, 480 p.
2. Shamanbaev Zh., E. Smagulova A.S., Issledovanie neuporjadochennyh dannyh na osnove tehnologii SPARK [Investigation of Disordered Data Based on SPARK Technology]. Respublikanskij nauchno-tehnicheskij zhurnal «Universitet enbekteri – Trudy universiteta» [Republican scientific and technical journal «University Enbekteri – Proceedings of the University»], 2020. No. 4. Pp. 156-159.
3. Aggarval Ch. Nejronnye seti i glubokoe obuchenie: uchebnyj kurs [Neural networks and deep learning]. Moscow: Publ. Dialektika-Vil'jams, 2020, 752 p.
4. Hunter A. Image Classification on ImageNet. URL: <https://paperswithcode.com/sota/image-classification-on-imagenet> (date of the application 24.04.2022).
5. Potapov A. Avtomaticheskij analiz izobrazhenij i raspoznavanie obrazov [Automatic image analysis and pattern recognition]. Moscow: Publ. LAP Lambert Academic Publishing, 2017, 292 p.
6. Hajkin S. Nejronnye seti. Polnyj kurs [Neural networks. Full course]. Moscow: Publ. Vil'jams, 2018, 1104 p.
7. Zelencov I.A. Raspoznavanie obrazov. Obzornaja lekcija [Pattern recognition. overview lecture]. URL: http://it-claim.ru/Persons/Zelencov/Lecture_presentation.pdf (date of the application 20.03.2022).
8. Kuljanin E. M. Nejronnye seti: istorija razvitija i perspektivy primeneniya [Neural networks: history of development and prospects for application]. Nauchno-metodicheskij jelektronnyj zhurnal «Koncept» [Scientific and methodological electronic journal «Concept»]. Moscow, 2018. No. 13, pp. 2646-2650.
9. Moskalev N. S. Vidy arhitektur nejronnyh setej [Types of architectures of neural networks]. Molodoj uchenyj [Young scientist]. Moscow, 2017. No. 29, pp. 30-34.
10. Bishop K.M. Raspoznavanie obrazov i mashinnoe obuchenie [Pattern recognition and machine learning]. Moscow: Publ. Dialektika-Vil'jams, 2020, 960 p.