

ОСНОВНЫЕ ПОНЯТИЯ И ПОДХОДЫ ПРИ ИСПОЛЬЗОВАНИИ СВЁРТОЧНЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ

Птуха В.И.¹, магистрант гр. 356241

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

Боброва Н.Л. – канд. техн. наук, доцент

Аннотация. В условиях стремительного роста объёмов визуальной информации и широкого распространения цифровых изображений возрастает потребность в эффективных методах их анализа и обработки. Современные устройства генерируют колоссальные объёмы графических данных, что делает технологии компьютерного зрения всё более востребованными в различных отраслях – от промышленности до повседневной жизни. Одним из ключевых достижений в этой области стали свёрточные нейронные сети (CNN), продемонстрировавшие высокую эффективность в задачах распознавания образов, классификации изображений и идентификации объектов. Настоящая статья посвящена рассмотрению фундаментальных понятий, принципов работы, а также основных подходов и концепции, используемых при построении архитектуры нейронной сети.

Ключевые слова. Компьютерное зрение, свёрточные нейронные сети, искусственный интеллект, свёртка.

Около 77% современных устройств используют искусственный интеллект в различных формах [1]. По прогнозам, к 2030 году рынок искусственного интеллекта достигнет суммы в размере 1339 миллиардов долларов, что существенно выше показателей 2024 года, оцениваемых в 214 миллиардов долларов. По данным компании IBM, 35% предприятий уже используют искусственный интеллект [3]. Трудно представить современный мир без цифровых изображений, которые стали неотъемлемой частью нашей жизни. Статистика рынка объективов для камер впечатляет – в 2020 году он превысил 8 млрд. долл. и продолжает расти. Прогнозы говорят, что совокупный годовой темп роста в период с 2021 по 2027 год составит 9%. Ожидается, что к 2027 году мировые поставки дойдут до отметки 250 млн. единиц в год [4]. Согласно отчётам, в секунду на Instagram загружается 1074 фотографии [5]. Постоянно увеличивается спрос на высокопроизводительные объективы для камер автомобильной промышленности, бытовой техники. Этот взрывной рост подчёркивает важность и востребованность цифровых изображений в нашей современной жизни.

Большинство устройств, начиная от смартфонов и заканчивая камерами наблюдения, постоянно генерируют огромные объёмы цифровых графических данных. В этом контексте, задачи анализа и обработки изображений становятся более актуальными и требуют развития новых подходов и технологий.

В свете этих вызовов и трендов, анализ и обработка изображений становятся важнейшими компонентами, требующими инновационных подходов и технологий. За последнее десятилетие нейронные сети, а особенно свёрточные нейронные сети, стали настоящим прорывом в области анализа изображений. Мощные алгоритмы обработки данных, вдохновлённые биологической природой зрения, смогли добиться значительных результатов в задачах, которые изначально казались непреодолимыми для компьютеров [6].

Анализ и обработка изображений с использованием нейросетей представляют собой одну из наиболее перспективных областей в современном машинном обучении. При этом необходимо глубокое понимание того, как работают свёрточные нейронные сети, представление о используемых подходах и методах, а также строение популярных архитектур.

Применение нейросетей для анализа изображений получило широкое распространение в различных областях науки, бизнеса и жизни в целом.

Зрение – это крайне сложный и важнейший функциональный механизм. Люди, животные, насекомые используют зрение постоянно что бы ориентироваться в пространстве, анализировать окружающую обстановку, взаимодействовать с окружающим миром. Человеческий мозг обрабатывает и интерпретирует звуки, изображения и другие сенсорные сигналы, опираясь на их частотные характеристики. К примеру, слуховая система человека воспринимает различные частоты звуковых волн, и именно на основе этих частот мозг различает высоту, тембр и другие особенности звуков. Музыкальное восприятие особенно сильно связано с частотным анализом, ведь мелодии, ритмы и гармонии формируются из определённых наборов частот, которые мозг способен распознавать и интерпретировать. Аналогичным образом, в визуальной системе человека частотный анализ играет ключевую роль в распознавании узоров и текстур. Некоторые визуальные структуры состоят из повторяющихся элементов, создающих визуальный ритм или частотный паттерн, который зрительная система эффективно улавливает и использует для анализа окружающего пространства. В целом, частотный анализ в сенсорном восприятии означает способность мозга извлекать важную информацию

из сигналов, исходя из их частотного состава. Это играет фундаментальную роль в нашем понимании и взаимодействии с внешним миром. При переходе к частотной области сигнал представляет собой спектр, состоящий из множества частотных компонентов, каждый из которых имеет свою амплитуду и фазу. Такое представление позволяет глубже понять свойства сигнала: высокочастотные компоненты обычно связаны с мелкими деталями, а низкочастотные – с более плавными, обобщёнными особенностями [7].

Мозг представляет собой систему, состоящую из огромного количества сравнительно простых и медлительных нейронов, выполняющих обработку информации. Чтобы достичь высокой производительности, требуется эффективное распределение вычислительных задач. Зрительная система решает эту задачу с помощью иерархической структуры, распределяя обработку как по пространству, так и по этапам [8].

Пространственное разделение приводит к тому, что нейроны обрабатывают информацию в пределах ограниченных участков зрительного поля (локализация рецептивных полей), а последовательное разделение обеспечивает формирование специализированных областей, каждая из которых отвечает за анализ всё более сложных признаков. Такой подход делает вычисления высокопараллельными и конвейерными. В отличие от мозга, персональные компьютеры обычно обладают одним или несколькими мощными вычислительными ядрами, поэтому подобная степень распараллеливания для них не столь критична. Однако с развитием графических процессоров (GPU) и других специализированных архитектур, параллельный и иерархический характер обработки, присущий мозгу, может стать значительным преимуществом.

Компьютерное зрение используется для тех же целей. Информация, полученная с камер, передаётся программе, которая обрабатывает её соответствующим образом и выдаёт результат.

Использование нейросетей, применяемых для анализа изображений включает в себя широкий спектр задач:

1. Классификация изображений: задача, позволяющая определить принадлежность изображения к определённому классу. В качестве примера можно привести классификацию цифр или букв в *mnist (modified national institute of standards and technology)*, классификация предметов или животных на изображении.

2. Обнаружение объектов. Нейронная сеть используется для обнаружения объекта на изображении. При этом определяется местоположение и класс одного или более объектов. Пример: определение человека на фотографии, обнаружение дорожного знака.

3. Сегментация изображений. Нейронная сеть выделяет каждый пиксель изображения с присвоением ему класса и метки. Такой подход позволяет создавать детальные маски объектов на изображениях.

4. Генерация изображений. Нейросеть создаёт собственное изображение по заданным критериям. Можно задать стиль, в котором должно быть сгенерировано изображение.

5. Распознавание рукописного текста. Такие сети, специально разрабатываются для обработки изображений, они имеют свёрточные слои, которые могут выделять фичи на разных уровнях иерархии, что является крайне полезным в анализе изображений.

6. Генеративные модели. В качестве примера можно привести *gan (generative adversarial network)*, позволяющая создавать новые изображения на основе обученных данных.

Анализ временных рядов и видео. На основе свёрточных архитектур можно выполнять анализ временных рядов и видео, что является крайне важным в области видеонаблюдения и анализа движущихся объектов.

Свёрточная нейронная сеть обычно состоит из нескольких типов слоёв, которые работают в согласованной последовательности для анализа и классификации изображений и других данных, имеющих пространственную структуру.

Слой входных данных (input layer) является первым слоем сети и является входной точкой для данных. Размерность слоя для входных данных определяется размерами изображения. Если на вход подаётся цветное изображение размером 300x300 пикселей, то размерность входного слоя будет равна 270000 ($300 \times 300 \times 3$) то есть равно произведению ширины, высоты и количеству каналов. Количество каналов входных данных определяется типом данных, цветные изображения обычно имеют три канала, черно-белые изображения – один канал, трехмерные медицинские изображения могут содержать большее количество каналов. Цифровое изображение представляется в виде набора пикселей, каждый пиксель при этом характеризуется значениями интенсивности соответствующих каналов. Интенсивность каждого канала описывается целым числом от 0 до 255. Цветные изображения состоят из RGB (red, green, blue) пикселей, содержащих яркости по красному, зелёному и синему каналу. Комбинации этих цветов позволяют отразить любой цвет из спектра.

Входной слой принимает данные в виде тензоров, которые представляют собой многомерные массивы, иными словами – матрицы чисел. Работа с тензорами, а не данными в виде векторов является одним из отличий от сетей прямого распространения. Для цветного изображения это трехмерный тензор, имеющий размерность, соответствующую ширине, высоте и количеству каналов.

Входной слой не выполняет вычислительных операций, а лишь принимает данные и передаёт их далее через свёрточные слои, слои объединения и другие слои нейронной сети.

Входные данные могут быть предварительно обработаны для нормализации значений пикселей в диапазоне от 0 до 1 или от -1 до 1. Также могут выполняться другие операции предварительной обработки, такие как выравнивание размеров изображений.

Свёрточный слой (convolutional layer) является основным элементом свёрточных нейронных сетей. Их используют в анализе изображений и обработке пространственных данных. Они предоставляют множество преимуществ и выполняют ряд важных функций. Свёрточные слои позволяют автоматически извлекать важные признаки из входных данных, таких как края, углы, текстуры, что позволяет моделям самостоятельно определять, какие характеристики важны для задачи. Операция свёртки снижает размерность данных, позволяя обрабатывать изображения с меньшим количеством параметров. В этом случае уменьшение размерности помогает сократить вычислительные требования и уменьшить риск переобучения. Слои работают локально, что означает, что каждый нейрон реагирует на ограниченную область входных данных. Слои обычно обучаются быть инвариантными к небольшим масштабным и пространственным сдвигам объектов на изображениях, что позволяет использовать такие нейросети для распознавания объектов в разных частях изображения. Следовательно нейроны в свёрточных слоях имеют более локальное понимание данных и выделяют локальные особенности. В отличие от полносвязанных слоёв, свёрточные слои имеют относительно небольшое количество параметров, следовательно они менее ресурсоемки и менее подвержены риску переобучения. Так же операции свёртки могут выполняться параллельно, что позволяет использовать вычислительные ресурсы наиболее эффективно и тем самым сократить время обучения нейронной сети [6, 10].

Основными концепциями свёрточных слоёв являются:

1. свёртка (*convolution*) – ключевая операция. её идея состоит в том, что по матрице изображения перемещается фильтр (ядро). движение фильтра происходит шаг за шагом, каждый шаг называется «шагом свёртки». при каждом шаге ядро перемещается на определённое расстояние по горизонтали и (или) вертикали внутри матрицы изображения, и выполняет операцию свёртки на этом участке изображения. фильтр начинает своё движение с верхнего угла входного изображения. при достижении края матрицы ядро смещается на шаг по вертикали, выполняет операцию свёртки, далее продолжает движение по горизонтали. на каждом этапе свёртки значения фильтра умножаются на значения пикселей в матрице изображений. эта операция создаёт новое значение для центрального пикселя в каждом положении фильтра. перемещение ядра осуществляется до тех пор, пока фильтр не пройдет всю высоту изображения. в результате будет сгенерирована новая карта признаков, содержащая в себе информацию о признаках, выделенных фильтром таких как локальные структуры, текстуры, грани, углы. этот процесс позволяет извлекать различные признаки и структуры из изображения, что делает свёрточные слои мощными инструментами для анализа изображений и компьютерного зрения [9].

2. ядро (kernel) – это своеобразный шаблон, матрица весов, определяющая на какие признаки акцентируется внимание во время операции свёртки. часто ядро носит второе название – фильтр. свёрточный слой может иметь несколько фильтров для извлечения разных признаков. к примеру, это могут быть признаки горизонтальных линий, вертикальных или окружностей [14].

3. страйд (stride) – служит для представления смещения ядра на каждом шаге свёртки. величина страйда влияет на размеры выходного изображения [12]. на рисунке 1 показано перемещение ядра при страйде равном 1.

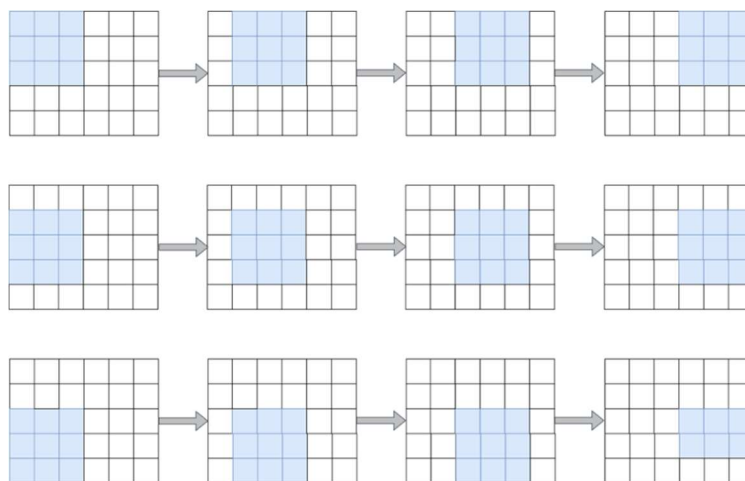


Рисунок 1 – Схема перемещения фильтра со страйдом равным 1

4. дополнение (padding) – нужно для сохранения размерности выходного изображения, учёта информации на краях изображения. существует два основных типа дополнения: дополнение нулями и дополнение копированием. в первом случае нули добавляются вокруг входных данных матрицы изображения. во втором – значения краевых пикселей копируются и добавляются вокруг входных данных [12, 14]. на рисунке 2 показаны варианты дополнения нулями и дополнения копированием.

0	0	0	0	0	0	0
0	1	1	1	4	3	0
0	1	2	2	1	2	0
0	1	2	2	2	1	0
0	0	1	2	2	1	0
0	0	1	1	1	1	0
0	0	0	0	0	0	0

А)

1	1	1	1	4	3	3
1	1	1	1	4	3	3
1	1	2	2	1	2	2
1	1	2	2	2	1	1
0	0	1	2	2	1	1
0	0	1	1	1	1	1
0	0	1	1	1	1	1

Б)

Рисунок 2 – А) дополнение нулями Б) дополнение копированием

5. пулинг (pooling) – применяется для уменьшения размерности и агрегации информации в карте признаков, полученных после свёртки. это позволяет сделать признаки более инвариантными к небольшим изменениям в изображении, к примеру маленькое смещение объекта. таким образом, объект на изображении может незначительно изменить своё положение, но при этом оставить свою семантику и внешний вид. тут важно упомянуть про max pooling, который и помогает сделать признаки более независимыми от небольших смещений в пространстве. это важно для задач компьютерного зрения, связанными с классификацией или обнаружением объектов, так как последние в свою очередь могут находиться под разными углами в разных частях изображения [11, 15].

6. канал (channel). так как изображения могут иметь несколько каналов, к примеру rgb (red, green, blue). свёрточные слои способны обрабатывать каждый слой по отдельности [16].

7. выходной размер (output size). на это размер влияют такие параметры как размер входного изображения, размеры фильтра, шаг и использование дополнения [9, 11, 15].

8. смещение (bias). это параметр, который добавляется к взвешенной сумме входных данных в нейроне перед применением функции активации. смещение позволяет управлять смещением или смещением активации нейрона в положительном или отрицательном направлении. положительное смещение говорит о том, что нейрон должен быть активен при низких значениях входных данных. при отрицательном смещении нейрон должен быть активен при высоких значениях. смещение – это отдельный параметр для каждого нейрона в сети. во время обучения нейронной сети алгоритм градиентного спуска также обновляет смещения, чтобы минимизировать ошибку модели и улучшить её производительность [17].

Свёрточный слой представляет из себя набор карт признаков, при этом каждая карта содержит в себе ядро. Необходимое количество карт зависит от постановки задачи и требований. Увеличение числа карт повышает качество распознавания, при этом повышается вычислительная сложность. Обычно в первом свёрточном слое используется небольшое количество фильтров 32 или 64. Эти фильтры помогают извлекать низкоуровневые признаки такие как края и текстуры. Последующие свёрточные слои могут содержать большее количество фильтров 128 для второго слоя и 256 для третьего. Такой подход позволяет модели извлекать более высокоуровневые и абстрактные признаки. В глубоких нейронных свёрточных сетях, может быть, множество слоёв и количество карт в каждом слое может значительно увеличиваться. Для задач классификации, в последнем свёрточном слое или слое перед полносвязанными слоями, количество карт равно количеству классов, которые сеть должна предсказать. При обучении нейронной сети с малым набором данных нужно быть крайне осторожным с увеличением числа карт в свёрточных слоях, так как это может привести к переобучению. В этом случае начинать лучше с меньшего количества карт и постепенно увеличивать их, следя за процессом обучения [12, 13].

На рисунке 3 показан пример свёртки матрицы размером 7x7 при помощи ядра размерами 3x3. При отсутствии дополнения, отсутствии смещения и страйде равном единице на выходе получаем матрицу со значениями:

-8	-8	-4	-4	0
-8	-4	-4	0	4
-4	-4	0	4	4
-4	0	4	4	8
-4	-4	0	4	8

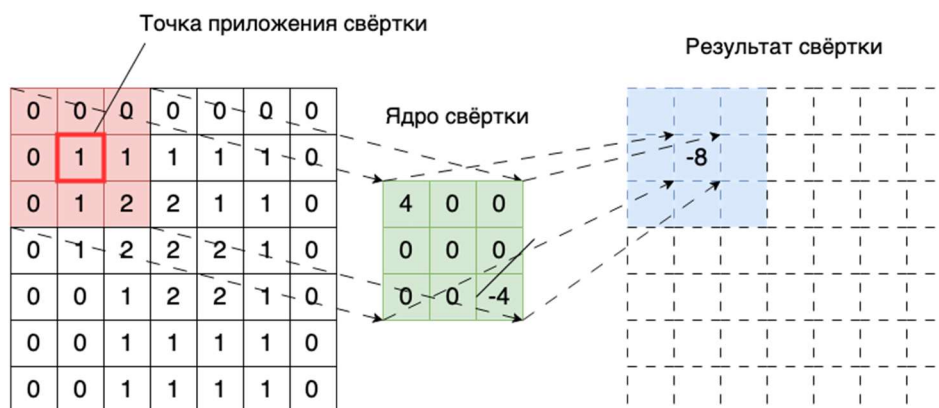


Рисунок 3 – Свёртка матрицы размером 7x7 при помощи ядра размером 3x3

При свёртке изображения с тремя каналами. Мы получим по три матрицы для каждого фильтра. Матрицы, полученные с применением фильтра для каждого канала, суммируются между собой и на выходе получаем один канал выходной карты признаков.

После того, как будут сформированы каналы для каждого из фильтров, матрицы объединяются в единый тензор, благодаря чему на выходе снова получается изображение.

Слой объединения (pooling layer) используется для уменьшения размерности данных и извлечения ключевых признаков из карт признаков, созданных свёрточными слоями. Их главная цель состоит в том, чтобы уменьшить пространственное разрешение карт признаков, сохраняя при этом наиболее важные информационные характеристики. Это позволяет сократить количество параметров сети, а следовательно, уменьшить вычислительную нагрузку, что является важным параметром при работе с нейронными сетями [18].

Существует несколько вариантов операций объединения. Максимальное объединение (max pooling) подразумевает, что выбирается максимальное значение из каждой группы пикселей, попавших в матрицу объединения, перемещающуюся по входящей матрице признаков [15].

$$\begin{matrix} 2 & -4 \\ 5 & 0 \end{matrix}$$

В данном случае будет выбрано значение 5. Данный вариант помогает сохранить наиболее выразительные признаки в каждой области.

Среднее объединение (average pooling) вычисляет среднее значение из каждой группы пикселей в заданной области. При таком подходе итоговое значение для вышеизображенной матрицы будет равным $(2 + (-4) + 5 + 0) / 4 = 0.75$. Такой подход также уменьшает размерность, при этом сохраняя средние значения признаков.

В слоях объединения можно настраивать размер окна объединения (pool size), который определяет размер области, из которой выбирается максимальное или среднее значение. Так же устанавливать шаг, с которым окно объединения перемещается по входным данным. Он контролирует насколько сильно уменьшается размер карт признаков.

Таким образом имея на входе карту признаков размерами 5x5:

$$\begin{matrix} -8 & -8 & -4 & -4 & 0 \\ -8 & -4 & -4 & 0 & 4 \\ -4 & -4 & 0 & 4 & 4 \\ -4 & 0 & 4 & 4 & 8 \\ -4 & -4 & 0 & 4 & 8 \end{matrix}$$

Используя максимальное объединение размерами 2x2 с шагом 1 на выходе, получим следующую карту признаков размерами 4x4:

$$\begin{matrix} -4 & -4 & 0 & 4 \\ -4 & 0 & 4 & 4 \\ 0 & 4 & 4 & 8 \\ 0 & 4 & 4 & 8 \end{matrix}$$

Слой объединения позволяют улучшить инвариантность к масштабу и малым сдвигам объектов, а также уменьшить количество параметров сети. Слой объединения обычно встроены между

свёрточными слоями и их наличие способствует повышению производительности сети в задачах компьютерного зрения.

Полносвязанные слои (fully connected layers). Являются представителями классических нейронных слоёв. Они соединяют все нейроны предыдущего слоя со всеми нейронами текущего слоя. Эти слои выполняют окончательную классификацию и регрессию на основе извлеченных признаков. Количество и расположение полносвязанных слоёв может различаться в зависимости от конкретной архитектуры сети и целей задачи [12]. На рисунке 4 показан пример полносвязанного слоя.

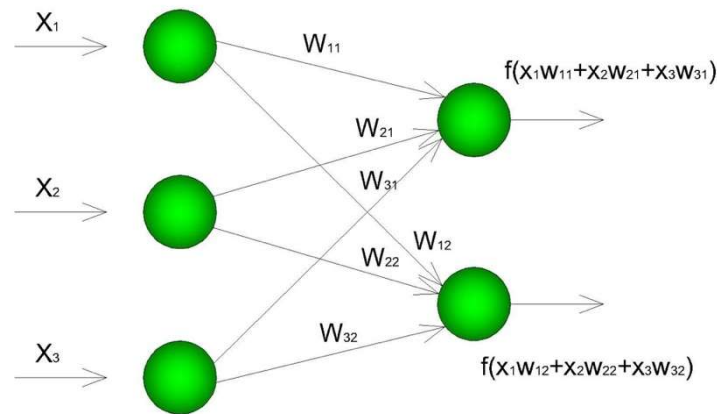


Рисунок 4 – Полносвязанный слой

Примером свёрточной нейронной сети с полносвязанными слоями является Le-Net, AlexNet, VGG [19, 20].

Слои смешивания (Dropout Layers) является одним из способов регуляризации нейронных сетей, используемых для предотвращения переобучения и улучшения обобщающей способности модели. Идея состоит в том, что случайные нейроны или узлы в сети исключаются во время обучения. Переобучение может возникнуть в момент, когда модель слишком хорошо подстроилась под обучающие данные и теряет способность обобщения на новые, ранее неизвестные данные. Слои смешивания уменьшают связанность между нейронами, что снижает вероятность переобучения. Для этого в таком слое случайным образом отключается заданная доля нейронов и узлов (обычно в диапазоне от 20% до 50%) для каждой итерации обучения. Это означает, что в каждой эпохе обучения некоторые нейроны не участвуют в прямом и обратном распространении ошибки и модель обучается на различных подмножествах данных [21].

Модели, обученные с использованием слоя смешивания, часто лучше обобщают на новые данные и демонстрируют лучшую производительность на тестовых данных. Важно помнить, что такие слои используются только во время обучения, и в тестовой фазе отключаются так как они не требуются для предсказаний.

Слой вывода (output layer) определяет, какие предсказания делает нейронная сеть. В зависимости от задачи это может быть слой с несколькими нейронами для задачи классификации, слой с одним нейроном для задачи регрессии и другие конфигурации.

Разобравшись в основных терминах и понятиях, а также со структурой нейронных сетей, можно приступить к непосредственному рассмотрению наиболее известных и значимых архитектур.

Заключение. Данная статья рассматривает основные задачи, решаемые нейронными сетями при анализе изображений. Дается базовое сравнение с принципами анализа изображений мозгом человека, ведь биологический мозг выдаёт результат всё ещё значительно превосходящий по точности, качеству обработки и распознавания в большинстве задач. Анализ работы мозга и заимствование подходов помогают делать прорывы в области развития компьютерного зрения. В статье описывается структура свёрточной нейронной сети с описанием используемых подходов и методов её построения. Конкретные конструктивные решения, которые имеют решающее значение для организации эффективного обучения сети. С развитием технологий и увеличением объема данных становится ясно, что глубокие нейронные сети остаются основным инструментом в обработке и анализе информации, особенно в области компьютерного зрения. Архитектуры представляют собой набор основных строительных блоков и являются ключевыми элементами для многих современных приложений, таких как распознавание образов, автоматическая классификация изображений, идентификация объектов и даже создание искусственных интеллектуальных систем. Современные технологии, такие как автономные автомобили, медицинская диагностика на основе изображений и распознавание лиц, все

более зависят от эффективности глубоких нейронных сетей. Понимание особенностей и принципов работы таких архитектур является необходимыми для успешного решения сложных задач в этих областях.

Таким образом, продолжение исследований в области глубокого обучения и архитектур нейронных сетей остается актуальным и важным направлением для развития технологий будущего, открывая новые возможности в области искусственного интеллекта и повседневной жизни.

Список использованных источников:

1. 105 AI Statistics (2025): Adoption & Growth Data / Naveen Kumar // Demandsage, 2024. – URL: <https://www.demandsage.com/artificial-intelligence-statistics/>
2. 22 Top AI Statistics And Trends / Katherine Haan // Forbes Advisor, 2024. – URL: <https://www.forbes.com/advisor/business/ai-statistics/#1>
3. 149 AI Statistics: The Present & Future of AI at Your Fingertips / Mark Webster // Authority Hacker, 2024. – URL: <https://www.authorityhacker.com/ai-statistics/>
4. Camera Lens Market Size By Application (Consumer Electronics [Mobile, AR/VR/MR, Digital Camera], Automotive, Healthcare), By Distribution Channel (Online, Offline), Industry Analysis Report, Regional Outlook, Growth Potential, Price Trend Analysis, Competitive Market Share & Forecast, 2021 – 2027 // Global Market Insights, 2021. – Report ID: GMI5012. – URL: <https://www.gminsights.com/industry-analysis/camera-lens-market>
5. How Many Pictures Are on Instagram? // EarthWeb, – URL: <https://earthweb.com/blog/how-many-pictures-are-on-instagram>
6. LeCun, Y. Deep learning / Y. LeCun, Y. Bengio, G. Hinton // Nature, 2015. – doi:10.1038/nature14539.
7. Tian, Y. Biological Basis and Computer Vision Applications of Image Phase Congruency: A Comprehensive Survey / Y. Tian, M. Wen, D. Lu, X. Zhong, Z. Wu // Biomimetics, 2024. – Vol. 9. – P. 422. – doi:10.3390/biomimetics9070422.
8. Kruger, N. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? / N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardi, J. Piater, A.J. Rodriguez-Sanchez, L. Wiskott // IEEE Trans. Pattern Anal. Mach. Intell., 2012. – Vol. 35. – P. 1847–1871.
9. Alzubaidi, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions / L. Alzubaidi, J. Zhang, A.J. Humaidi et al. // J Big Data, 2021. – Vol. 8. – P. 53. – doi:10.1186/s40537-021-00444-8.
10. Mauricio, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review / J. Mauricio, I. Domingues, J. Bernardino // Appl. Sci., 2023. – Vol. 13. – P. 5521. – doi:10.3390/app13095521.
11. Zheng, H. A Deep Convolutional Neural Network-Based Multi-Class Image Classification for Automatic Wafer Map Failure Recognition in Semiconductor Manufacturing / H. Zheng, S.W.A. Sherazi, S.H. Son, J.Y. Lee // Appl. Sci., 2021. – Vol. 11. – P. 9769. – doi:10.3390/app11209769.
12. Albawi, S. Understanding of a convolutional neural network / S. Albawi, T.A. Mohammed, S. Al-Zawi // 2017 International Conference on Engineering and Technology (ICET), 2017. – P. 1-6. – doi:10.1109/ICEngTechnol.2017.8308186.
13. Guo, Y. Deep learning for visual understanding: A review / Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew // Neurocomputing, 2016. – Vol. 187. – P. 27-48.
14. Esquivel, J. Adaptive Convolutional Kernels / J. Esquivel // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019.
15. Yi, Z. Evaluation and Implementation of Convolutional Neural Networks in Image Recognition / Z. Yi // J. Phys.: Conf. Ser., 2018. – Vol. 1087. – P. 062018.
16. Singh, T. A deeply coupled ConvNet for human activity recognition using dynamic and RGB images / T. Singh, D.K. Vishwakarma // Neural Comput & Applic, 2021. – Vol. 33. – P. 469–485.
17. Zhao, X. A Review of Convolutional Neural Networks in Computer Vision / X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, M. Parmar // Artificial Intelligence Review, 2024. – Vol. 57. – P. 1-43.
18. Indolia, S. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach / S. Indolia, A. Goswami, S.P. Mishra, P. Asopa // Procedia Computer Science, 2018. – Vol. 132. – P. 679-688. – doi:10.1016/j.procs.2018.05.069.
19. Lecun, Y. Gradient-Based Learning Applied to Document Recognition / Y. Lecun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE, 1998. – Vol. 86. – P. 2278-2324. – doi:10.1109/5.726791.
20. Zhang, A. 8.1. Deep Convolutional Neural Networks (AlexNet) / A. Zhang, Z. Lipton, M. Li, A.J. Smola // 2024.
21. Brownlee, J. A Gentle Introduction to Dropout for Regularizing Deep Neural Networks / J. Brownlee // Machine Learning Mastery, – URL: <https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/>

BASIC CONCEPTS AND APPROACHES IN USING CONVOLUTIONAL NETWORKS FOR IMAGE ANALYSIS

Ptukha V.I.¹

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

Bobrova N.L. – PhD in Technical Sciences, associate professor

Annotation. In the context of the rapid growth of visual information and the widespread use of digital images, the need for effective methods of image analysis and processing is increasing. Modern devices generate enormous volumes of graphical data, making computer vision technologies increasingly in demand across various industries – from manufacturing to everyday life. One of the key achievements in this field has been the development of convolutional neural networks (CNNs), which have demonstrated high efficiency in tasks such as pattern recognition, image classification, and object identification. This paper focuses on the fundamental concepts, operating principles, as well as the main approaches and ideas used in the design of neural network architectures.

Keywords. Computer vision, convolutional neural networks, artificial intelligence, convolution.