УДК 004.93

СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

А.С. Прокопеня
Аспирант кафедры ЭВС, БГУИР

И.С. Азаров
Доцент, доктор технических наук, заведующий кафедрой ЭВС

Белорусский государственный университет информатики и радиоэлектроники
ул. П. Бровки, 6, БГУИР, каф. ЭВС, 220013, Минск, Беларусь, тел. +375 17 2938805,
E-mail: azarov@bsuir.by

Прокопеня А.С.
Окончил БНТУ в 2017 году, квалификация «Педагог-программист». Защитил магистерскую диссертацию по педагогике (БНТУ, 2018). Сфера научных интересов: цифровая обработка изображений, распознавание изображений.

Азаров И.С.
В 2002 г. окончил БГУ. В 2009 г. защитил кандидатскую, а в 2015 диссертацию в БГУИР по специальности 05.13.17 «Теоретические основы информатики». Сфера научных интересов: цифровая обработка сигналов, изображений, анализ и синтез речи.

Аннотация. Цель работы, результаты которой представлены в рамках статьи, заключалась в исследовании современных архитектур сверточных нейронных сетей для распознавания изображений. В статье рассмотрены такие архитектуры как AlexNet, ZFnet, VGGNet, GoogleNet, ResNet. Характеристикой о качестве распознавания изображения для нейронной сети является ошибка top-5. На основе полученных результатов было выявлено, что на данный момент сетью с наиболее точным результатом является свёрточная сеть ResNet с показателем точности в 3,57%. Преимуществом данного исследования является то, что приведенная статья дает краткую характеристику свёрточной нейронной сети, а также дает представление о современных архитектурах свёротчных сетей, их строением и качественными показателями.

Ключевые слова: свертка, фильтр, структура, подвыборка, функция активации.

Введение. Повсеместное развитие и распространение технологий компьютерного зрения (computer vision) влечет за собой изменение других профессиональных областей жизнедеятельности человека. Свёрточные нейронные сети применяются в системах распознавания объектов и лиц, специальном медицинском ПО для анализа снимков, навигации автомобилей, оснащенных автономными системами, в системах защиты, и других сферах. С ростом вычислительной мощности персональных компьютеров, а также появлением баз изображений стало возможным обучать глубокие нейронные сети (deep neural networks). В задаче распознавания изображений применяются свёрточные нейронные сети (Convolutional Neural Networks). Цель статьи – обзор современных архитектур свёрточных нейронных сетей для задачи классификации изображений.

Одна из задач машинного обучения – это задача классификации изображений. Классифицировать объект на изображении – значит указать номер, к которому относится распознаваемый объект. Для оценки алгоритмов машинного обучения обычно используются размеченные базы данных изображений, например, CIFAR-10, ImageNet, PASCAL VOC. Из-

за того, что на изображениях, например в базе изображений ImageNet может присутствовать несколько объектов, а размечен(аннотирован) только один, основным критерием ошибки является top-5 ошибка. Т.е. считается, что алгоритм не ошибся, если правильная категория объекта находится среди пяти категорий, выданных алгоритмом как наиболее вероятные. Следовательно многие нейронные сети для классификации изображений оцениваются с помощью ошибки top-5 [1].

Свёрточные нейронные сети (СНС) применяются для оптического распознавания образов, классификации изображений, детектирования предметов, семантической сегментации и других задач. Основы современной архитектуры СНС были заложены в одной из первых сетей – LeNet-5 Яна ЛеКуна.

Структура сверточной нейронной сети

Сеть свертки представляет собой многослойный персептрон (перцептрон, англ. perceptron от лат. perceptio – восприятие [2]) – математическая или компьютерная модель восприятия информации мозгом, созданный для распознавания 2D-поверхностей с высокой степенью устойчивости к масштабированию, преобразованиям и другим видам деформации. Обучение решению такой задачи осуществляется с подкреплением, при этом используются сети вида, архитектура которых соответствует следующим ограничениям [3].

Извлечение признаков. Каждый нейрон получает входной сигнал от локального рецептивного поля в предыдущем слое, извлекая его локальные признаки. Как только признак извлечен, его местоположение не имеет значения, т.к. приблизительно установлено его расположение относительно других признаков.

Отображение признаков. Каждый вычислительный слой сети состоит из множества карт признаков. Каждая карта признаков имеет форму плоскости, на которой все нейроны должны совместно использовать одно и то же множество синаптических весов. Эта форма структурных ограничений имеет преимущества.

Инвариантность к смещению. Инвариантность к смещению реализуется посредством карт признаков с использованием свертки с ядром небольшого размера, которая выполняет функцию "сплющивания".

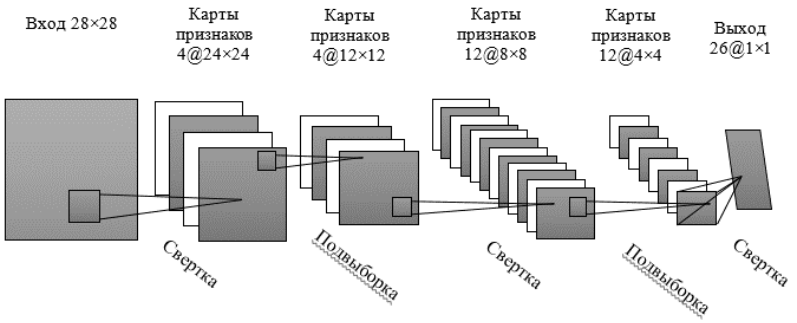

Рисунок *1*. – Сеть свертки для обработки изображений

Подвыборка. За каждым слоем свертки следует вычислительный слой, осуществляющий локальное усреднение (local averaging) и подвыборку (subsampling). Посредством локального усреднения достигается уменьшение разрешения для карт признаков. Такая операция приводит к уменьшению чувствительности выходного сигнала оператора отображения признаков, а также к смещению и другим формам преобразований.

На рисунке 1 представлена схема свёрточной сети, состоящей из одного входного, четырех скрытых и одного выходного слоя нейронов. Эта сеть была создана для обработки изображений, в частности при распознавании рукописного текста. Входной слой, состоящий из матрицы 28×28 сенсорных узлов, получает изображения различных символов, которые

предварительно смещены к центру и нормализованы по размеру. После этого вычислительные слои поочередно реализуют операции свертки и подвыборки.

Первый скрытый слой выполняет свертку. Он состоит из четырех карт признаков, каждая из которых представляет собой матрицу из 24×24 нейронов. Каждому нейрону соответствует поле чувствительностью размером 5×5.

Второй скрытый слой выполняет операцию локального усреднения, а также подвыборку. Он тоже состоит из 4 карт признаков, содержащих матрицы размером 12×12. Каждому нейрону соответствуют рецептивное поле размером 2×2, сигмоидальная функция активации, настраиваемый порог, настраиваемый коэффициент. Настраиваемый коэффициент и порог определяют рабочую область нейрона. Например, при небольшом коэффициенте нейрон работает в квазилинейном режиме.

Третий скрытый слой выполняет повторную операцию свертку. Слой состоит из 12 карт признаков, каждая из которых представляет собой матрицу 8×8 нейронов. Каждый нейрон третьего скрытого слоя может иметь синаптические связи с различными картами признаков из предыдущего скрытого слоя.

Четвертый скрытый слой осуществляет вторую подвыборку и повторную операцию локального усреднения. Слой состоит из 12 карт признаков, однако каждая карта признаков содержит матрицу из 4×4 нейронов.

Выходной слой осуществляет последний этап свертки. Слой состоит из 26 нейронов, каждому из которых соответствует одна из 26 букв латинского алфавита. Каждому нейрону соответствует рецептивное поле размером 4×4 [4].

Архитектуры свёрточных нейронных сетей для классификации изображений

AlexNet – сверточная нейронная сеть, которая победила в конкурсе ImageNet LSVRC-2012 с ошибкой 15,3 %. Сеть имела частично схожую архитектуру с сетью LeNet от Я. Лекуна, но была более глубокой, с большим количеством сверточных слоев. Сеть состояла из 11×11, 5×5, 3×3 сверток, максимального объединения, отсева, увеличения данных, функции активации ReLU, SGD (Stochastic Gradient Descent) с импульсом. Применение функции активации происходило после каждого сверточного и полносвязанного слоев.

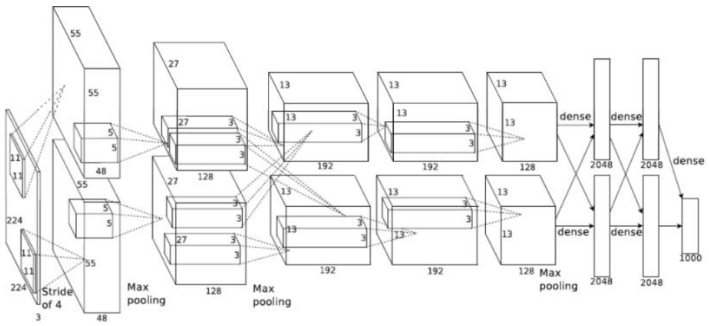

Рисунок 2 – Архитектура AlexNet

Представленная на рис. 2. архитектура AlexNet, включает в себя 8 слоев с весовыми коэфициентами, где первые 5 слоев – сверточные, последующие 3 полносвязанные. Выход последнего полносвязанного слоя подается на функцию активации softmax, которая производит распределение по меткам классов. Далее сеть максимизирует цель многочленной логистической регрессии, что развнозначно максимизации среднего по обучающим случаям логарифмической вероятности правильной метки при распределении прогноза.

Используя стохастический градиентный спуск со скоростью обучения 0,01, импульсом равным 0,9 и снижением веса 0,0005. Скорость обучения делится на 10 раз плато точности, также снижаясь в 3 раза в процессе обучения.

Формулы для обновления весовых коэфициентов

Обновление весовых коэффициентов (*w*), где *i* – это индекс итерации, *v* – переменная импульса, а *epsilon* – скорость обучения показано на схеме. Скорость обучения подбиралась равной для всех слоев, а также корректировалась вручную в процессе всего обучения. Следующий шаг заключался в делении скорости обучения на 10, когда количество ошибок при валидации переставало уменьшаться. AlexNet показывает результат top-5 ошибок – 15,3% соответственно.

ZFNet – победитель ILSVRC 2013 с top-5 ошибкой 11,2 %. Основную роль в этом играет настройка гиперпараметров, а именно размер и количество фильтров, размер пакетов, скорость обучения и т. д. М. Цилер и Р. Фергюс предложили систему визуализации ядер, весов и скрытого представления изображений. Система получила название DeconvNet.

Сетевая архитектура ZFNet, практически идентична архитектуре сети AlexNet. Существенные различия между ними в архитектуре заключаются в следующем:

– размер фильтра ZFNet и шаг в первом сверточном слое в AlexNet равен 11×11, шаг равен 4; в ZFNet – 7×7, шаг равен 2;

– количество фильтров в чистых сверточных слоях сети (3, 4, 5).

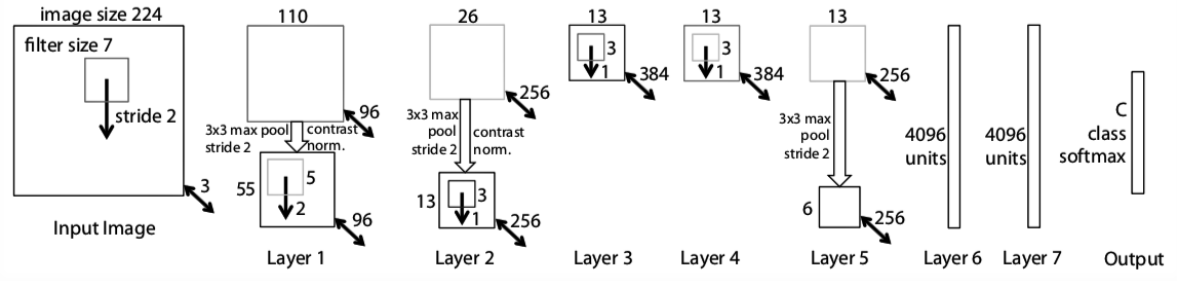

Рисунок *3* – Архитектура ZFNet

VGGNet

В 2014 году К. Симонян и Э. Циссерман из Оксфордского университета предложили архитектуру нейронной сети, называемую VGG (Visual Geometry Group).

VGG16 является улучшенной версией AlexNet, в которой большие фильтры (размером 11 и 5 в 1-м и 2-м сверточном слое) заменены на несколько фильтров размером 3x3, следующих один за другим.

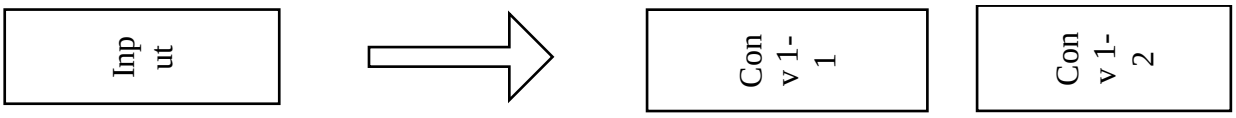

Рисунок *4.* – Архитектура сети VGGNet

Во время обучения вход в ConvNets (Convolutional networks – сверточные нейронные сети) представляет собой изображение RGB фиксированного размера 224×224 px. На следующем шаге изображение пропускается через стопку сверточных слоев размером 3×3. В одной из конфигураций сети VGGNet используются фильтры 1×1, которые можно рассматривать как линейное преобразование входных каналов.

Шаг свертки фиксируется на значении 1 пиксель. Пространственное дополнение (padding) входа сверточного слоя выбирается таким образом, чтобы пространственное разрешение сохранялось после свертки, то есть дополнение равно 1 для 3x3 сверточных слоев.

Пространственный пулинг осуществляется при помощи пяти max-pooling слоев, которые следуют за одним из сверточных слоев (не все сверточные слои имеют последующие max-pooling). Операция max-pooling выполняется на окне размера 2x2 пикселей с шагом 2 [13].

После стека сверточных слоев идут 3 полносвязных слоя: первые два слоя имеют по 4096 каналов, третий слой – 1000 каналов (т.к. в соревновании ILSVRC необходимо распределить объекты по 1000 категориям). Последний слой – softmax. Все скрытые слои снабжены функией активации ReLU.

Авторы продемонстрировали, что с помощью стандартных блоков можно добиться определенных результатов в конкурсе ImageNet. Число ошибок top-5 сократилось до 7,3 % [6].

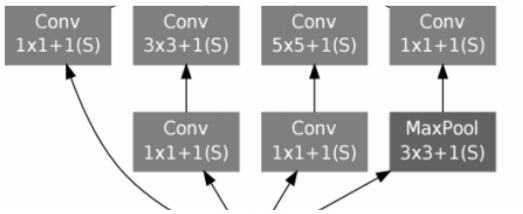

Рисунок 5. – Архитектура GoogleNet

Сверточная сеть от компании Google (GoogLeNet), известная как Inveption-v1 – победитель ILSVRC 2014 с top-5 ошибкой 6,7 % [Szegedy et al., 2015].

Все свертки в сети, в том числе внутри модулей Inception, используют прямолинейную линейную активацию. Сеть насчитывает 22 слоя при подсчете только слоев с параметрами. Общее количество слоев, используемых для построения сети, составляет 100. Переход от полностью связанных слоев к среднему пулу улучшил точность топ-1 примерно на 0,6%, однако использование отсева оставалось необходимым даже после удаления полностью связанных слоев.

Учитывая глубину сети, способность распространять градиенты обратно по всем слоям была эффективной задачей. Высокая производительность более мелких сетей при выполнении этой задачи позволяет предположить, что функции, создаваемые слоями в середине сети, должны быть очень дискриминационными. При добавлении вспомогательных классификаторов, связанных с этими промежуточными уровнями, ожидалось различение на более низких ступенях в классификаторе. Это помогло побороть проблему исчезающего градиента, обеспечивая регуляризацию. Классификаторы в сети принимают форму небольших сверточных сетей, размещенных поверх выходных данных модулей Inception. Во время обучения их потеря добавляется к общей потере сети с весом. Во время вывода эти вспомогательные сети отбрасываются. Более поздние контрольные эксперименты показали, что влияние вспомогательных сетей относительно невелико (около 0,5%) и что для достижения такого же эффекта требуется только одна из них.

Точная структура дополнительной сети, включая вспомогательный классификатор, выглядит следующим образом:

– средний объединяющий слой с размером фильтра 5 × 5 и шагом 3, что приводит к выходу 4×4×512 для (4a) и 4×4×528 для (4d) каскада;

– с вертка 1×1 со 128 фильтрами для уменьшения размеров и прямой линейной активации;

– полностью связанный слой с 1024 единицами и прямой линейной активацией;

– отбрасываемый слой с 70% -ным соотношением отброшенных выходов;

– линейный слой с потерями softmax в качестве классификатора (предсказывает те же 1000 классов, что и основной классификатор, но удаляется во время вывода).

В архитектуре GoogLeNet используется модуль Inception, построение сети осуществляется на основе модулей такого типа [1].

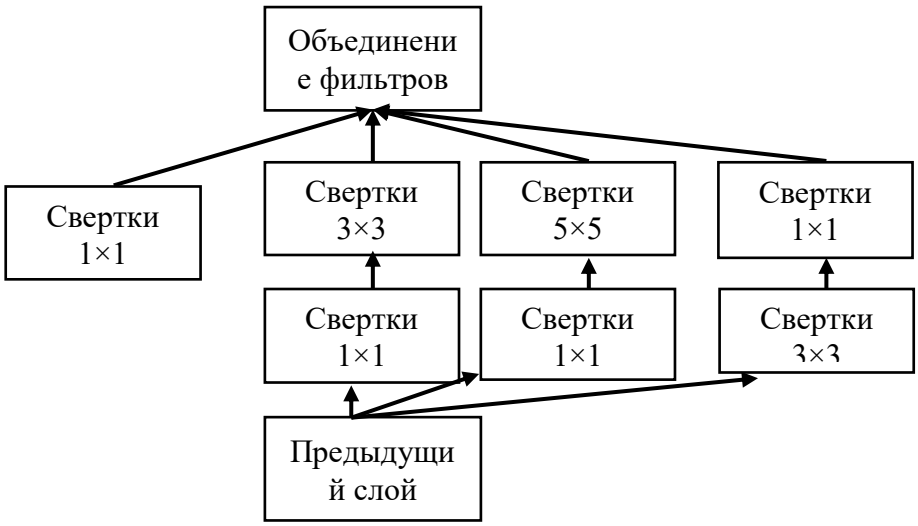

Рисунок *6* – Модуль Inception

Модуль Inception.

Inception использует несколько ветвей (параллельных), вычисляющих различные свойства на основании одинаковых входных данных, а впоследствии объединяет полученные результаты.

Свертка размером 1×1 является способом сокращения размерности карты свойств. Такой тип свёрточных слоев представлен в работе «Сеть» в сети М. Лина. В результате такая архитектура позволяет сократить число ошибок для top-5 категорий еще на 0.5% – до значения 6,7%.

Модуль Inception-v2 и Inception-v3

В следующей итерации модуля Inception (Inception-v2 [7]), декомпозируется слой с фильтром 5×5 на два слоя 3×3. Следующий этап – использование Batch Normalization [Ioffe, Szegedy, 2015], позволяющее увеличить скорость обучения посредством нормализации распределения выходов слоёв внутри сети. В той же статье авторы предложили концепцию модуля Inception-v3. В модуль Inception-v3, заложен принцип декомпозиции фильтров, а именно декомпозирование фильтра размером N×N двумя последовательными фильтрами 1×N и N×1. Также в Inception-v3 используется RMSProp (Метод адаптивного скользящего среднего градиентов) [Hinton, Srivasta, Swersky, 2012], вместо градиентного спуска используется усечение градиентов [Pascanu et al., 2013], которое используется для повышения стабильности обучения. Объединение из четырёх модулей Inception-v3 показал результат в категории top-5 ошибку 3,58 % на ILSVRC 2015, Inception-v2 – результат top-5 – 5.60%.

ResNet. ResNet –сокращенное название для Residual Network (дословно – «остаточная сеть»).

Простая Сеть. Базовые линии (рис. 7, в центре) основаны на философии сетей VGG [12] (рис. 7, слева). Сверточные слои имеют фильтры размером 3×3, и следуют правилам проектирования:

– при одинаковом размере карты выходных объектов, слои имеют одинаковое количество фильтров;

– если размер карты объектов уменьшится вдвое, число фильтров напротив, удваивается для того, чтобы сохранить сложность времени для слоя.

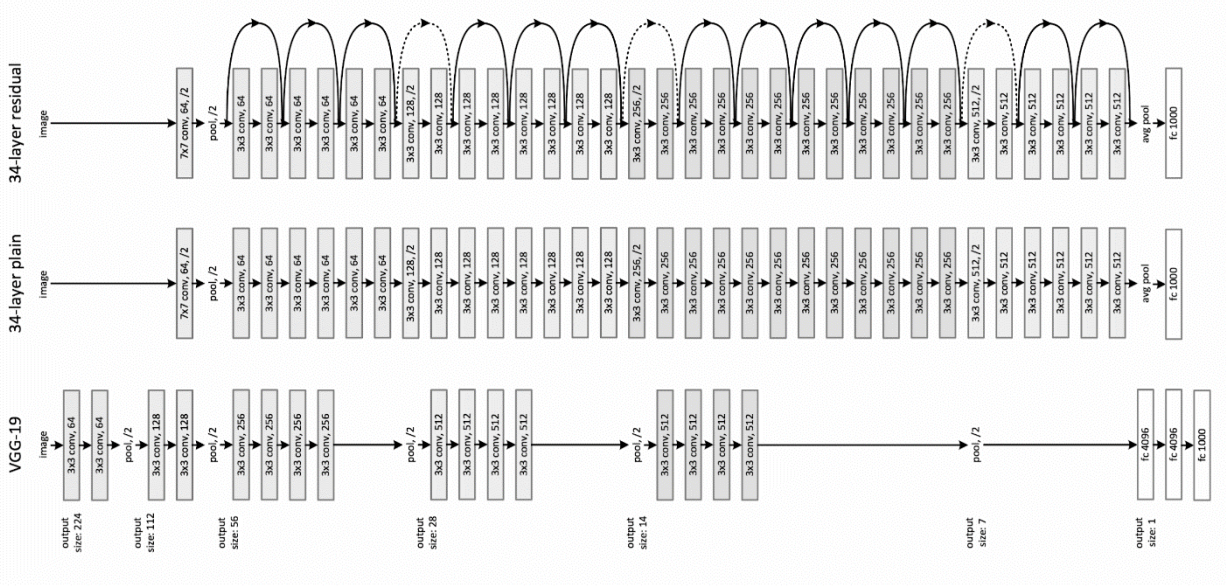

Рисунок *7.* – Пример сетевой архитектуры для ImageNet. Слева: модель VGG-19. Посередине: простая сеть с 34 слоями. Справа: ResNet с 34 слоями.

ResNet: на основе описанной простой сети добавлено быстрое соединение (рис. 7, справа), которое превращает сеть в ее остаточную версию. Идентификационные быстрые соединения F (x {W} + x) могут использоваться непосредственно, когда вход и выход имеют одинаковые размерности (быстрые соединения сплошной линии на рис. 7).

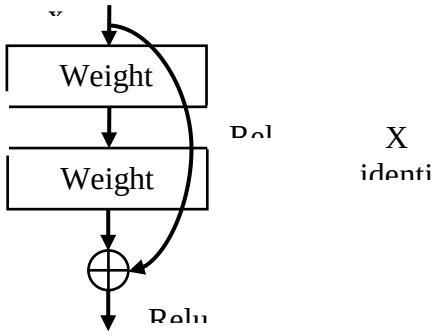

Рисунок *8.* – Быстрое соединение

При увеличении размерности (пунктирные линии на рисунке 7), быстрое соединение рассматривает 2 варианта:

1. Быстрое соединение выполняет сопоставление идентификаторов с дополнительными нулями, которые добавлены для того, чтобы увеличить размерность. Такой подход не вводит дополнительных параметров.

2. Проекция быстрого соединения в F (x {W} + x) используется для сопоставления размерностей (выполнено с помощью 1×1 сверток).

Для любой из опций, если быстрые соединения идут по картам объектов двух размерностей, они выполняются с шагом 2.

Таблица 1. – Характеристики CNN ResNet

Layer name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
Conv1	112×112	7×7, 64, Stride 2				
Conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$	$\begin{bmatrix}3\times3,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$	$\begin{bmatrix}3\times3,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$
Conv3_x	28×28	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times8$
Conv4_x	14×14	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times36$
Conv5_x	7×7	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,2048\end{bmatrix}\times3$
	1×1	average pool, 1000-dfc,softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Каждый блок ResNet имеет два уровня глубины (используется в небольших сетях, таких как ResNet 18, 34) или 3 уровня (ResNet 50, 101, 152) (Таблица 1).

50-слойная ResNet: каждый 3-слойный блок заменяется в 34-слойной сети этим 3-слойным узким местом, в результате получается 50-слойная ResNet (см. Таблицу 1). Они используют вариант 2 для увеличения размерностей. Эта модель имеет 3,8 миллиарда FLOPs.

ResNet с 101 и 152 слоями: они создают ResNet с 101 и 152 слоями, используя больше 3-слойных блоков (см. Таблицу 1). После увеличения глубины 152-слойная сеть ResNet (11,3 миллиарда FLOP) имеет меньшую сложность, чем сети VGG-16 и VGG-19 (15,3 / 19,6 миллиарда FLOPs). ResNet – 152 достигает результата top-5 – 3.57%.

Сравнение моделей свёрточных нейронных сетей

Для оценки показателей моделей свёрточных нейронных сетей указывают вид ошибки (top-5). На изображениях в базе ImageNet может присутствовать множество объектов, однако аннотирован, только один из них. Основным критерием ошибки, является ошибка top-5. Результаты сравнения результатов различных свёрточных нейронных сетей представлены в таблице 2.

Таблица 2. – Сравнение показателей CNN в задачах распознавания изображений

Нейронная сеть	**Top-5**
AlexNet	15,3 %
ZF Net	11,2 %
VGG Net	7,3 %
GoogleLeNet	6,7 %
Inception-v2	5,60 %
Inception-v3	3,58 %
ResNet-152	3,57 %

Заключение

Распространение и развитие технологий компьютерного зрения влечет за собой изменение других профессиональных областей жизнедеятельности человека. Свёрточные нейронные сети (СНС) применяются в системах распознавания объектов и лиц, специальном медицинском ПО для анализа снимков, навигации автомобилей, оснащенных автономными системами, в системах защиты, и других сферах. С ростом вычислительной мощности компьютеров, появлением баз изображений стало возможным обучать глубокие нейронные сети. Одной из главных задач машинного обучения является задача классификации изображений. СНС применяются для оптического распознавания образов и объектов, детектирования предметов, семантической сегментации и т.д. В данной статье были рассмотрены наиболее распространенные архитектуры свёрточных нейронных сетей для задачи распознавания изображений, их строение и особенности.

В результате проведенного анализа архитектур выявлено, что свёрточная нейронная сеть ResNet-152 показала наилучший результат в задаче распознавания изображений, с показателем top-5 равным 3,57%, что говорит о достаточно точном определении объекта. Особенностью архитектуры ResNet является то, что свёрточные слои имеют фильтры 3×3, а также то, что в сеть добавлено быстрое соединение, которое превращает сеть в ее остаточную версию.

Список литературы

[1.] Сикорский, О.С. Обзор свёрточных нейронных сетей для задачи классификации изображений / О. С. Сикорский // Новые информационные технологии в автоматизированных системах – Москва, 2017. – № 20. – С. 37–42.

[2.] Википедия [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/Перцептрон. – Дата доступа: 24.05.2019.

[3.] LeCun Y. and Y. Bengio. "Convolutional networks for images, speech and time series", in M.A. Arbib, ed., The Hand book of Brain Theory and Neura1 Networks, Cambridge, MA: MIT Press, 1995.

[4.] Нейронные сети: полный курс, 2-е издание.: Пер. с англ. – М. Издательский дом «Вильямс», 2006. – 1104 с.: ил. – Парал. тит. англ.

[5.] Википедия [Электронный ресурс]. – Режим доступа: http://ainews.ru/2018/11/alexnet_svertochnaya_nejroset_dlya_klassifikacii_izobrazhenij.html – Дата доступа: 24.05.2019.

[6.] Karen Simonyan, Andrew Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[7.] Sergey Ioffe, Christian Szegedy, 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

[8.] Geoffrey Hinton, Nitish Srivasta, Kevin Swersky. 2012. «Lecture 6a Overview of Mini – Batch Gradient Descent» www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. Accessed 21 Mar. 2017.

[9.] Dual Path Networks [Электронный ресурс]. – https://paperswithcode.com/paper/dual-path-networks – Дата доступа: 24.05.2019.

[10.] Deep Residual Learning for Image Recognition / Kaiming He [et al.]. – In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[11.] Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree / Chen-Yu Lee [et al.]. – In Artificial Intelligence and Statistics, pages 464–472, 2016.

[12.] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[13.] Neurohive [Электронный ресурс]. – https://neurohive.io/ru/vidy-nejrosetej/vgg16-model/ – Дата доступа: 24.05.2019.

[14.] ImageNet Classification with Deep Convolutional Neural Networks / Alex Krizhevsky [et al.]. –November 2013, 9 S.

OVERVIEW OF CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE RECOGNITION

A.S. Prokopenya
Postgraduate Student, Department of ECT, BSUIR

I.S. Azarov
assistant professor, Doctor of Technical Sciences, Head of the Department of ECT

Belarusian state University of Informatics and Radioelectronics
6, P. Brovki str., BGUIR, KAF. EMU, 220013, Minsk, Belarus, tel. +375 17 2938805,
E-mail: azarov@bsuir.by

Abstract. The purpose of the work, the results of which are presented in the article, was to study modern architectures of convolutional neural networks for image recognition. This article discusses such architectures as AlexNet, ZF net, Get, Google Net, Reset. The characteristic about the image recognition quality for a neural network is the top-5 error. Based on the results obtained, it was found that at the moment the network with the most accurate result is the RESNET convolutional network with an accuracy rate of 3.57%. The advantage of this study is that this article provides a brief description of the convolutional neural network, as well as gives an idea of modern architectures of convolutional networks, their structure and quality indicators.

Keywords: convolution, filter, structure, subsample, activation function