

**Головко В.А., Михно Е.В., Брич А.Л., Михняев А.Л.,
Войцехович Л.Ю., Матюшков А.Л.**

РЕДУЦИРОВАННАЯ СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ТОЧНОГО РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР

Введение. Искусственная нейронная сеть находит эффективное применение в различных областях [1–5]. В течение последнего десятилетия технологии машинного обучения занимают ведущую роль в области искусственного интеллекта [1]. Это подтверждают последние достижения в области обработки изображений, видео, речи, естественного языка, больших объемов данных и их визуализации и т. д. [1–5]. Эти достижения прежде всего связаны с новой парадигмой в области машинного обучения, а именно с глубоким обучением и глубокими нейронными сетями [2, 6–18]. Однако во многих существующих приложениях важной проблемой являются ограниченные вычислительные мощности, которые не позволяют использовать глубокие нейронные сети. Поэтому дальнейшее развитие упрощенной архитектуры является достаточно важной задачей. В особенности следует отметить, что для многих существующих приложений, упрощенная архитектура может показать точность, сопоставимую с глубокими нейронными сетями.

Сверточные нейронные сети (convolutional neural networks, CNN) являются дальнейшим развитием многослойного перцептрона и неоконитрона и широко используются для обработки изображений. Общая концепция конволюционных сетей была предложена в 1989 г. и затем была доработана в 1998 г. введением архитектуры сверточной сети LeNet-5. Основателями данного класса сетей являются Ян Лекун (Yann LeCun) и Иосуа Бенджио (Yoshua Bengio) [19, 20]. Нейробиологической основой для создания сверточных нейронных сетей являются исследования зрительной коры головного мозга кошки нейробиологами Д. Хьюбелом и Т. Визелом. Они открыли два типа клеток в зрительной коре: простые и сложные. Простые нейроны следят за своим рецептивным полем и выделяют примитивные (низкоуровневые) характеристики образа, такие, как прямые линии под разными углами и т. д. Сложные нейроны осуществляют интеграцию примитивных характеристик образа для построения высокоуровневых признаков.

Сверточные нейронные сети в отличие от многослойного перцептрона позволяют учитывать топологию изображений и являются инвариантными к сдвигам, масштабированию и другим искажениям входного образа. В настоящее время их эволюция происходит по пути увеличения количества слоев (глубины), в результате чего появились глубокие сверточные нейронные сети (deep CNN). Благодаря глубокой архитектуре такие сети позволяют достичь точности распознавания рукописных цифр (база данных MNIST) 99.65% на тестовой выборке.

В данной работе показано, что высокая точность распознавания может быть получена на редуцированной сверточной сети без применения дополнительных приемов к обучению (искажение цифр, регуляризация и т. д.). В работе рассматриваются основные принципы построения и обучения сверточных нейронных сетей. Разработана упрощенная архитектура сверточной нейронной сети, которая позволяет классифицировать рукописные цифры с точностью 99.29% (ошибка тестирования 0.71%), что является лучшим результатом в классе неглубоких (shallow) сверточных нейронных сетей (<http://yann.lecun.com/exdb/mnist/>).

1. Анализ предыдущих результатов и постановка задачи. Дана база данных MNIST, которая содержит 60 000 изображений для обучения и 10 000 изображений для тестирования. Каждое изображение имеет размерность 28x28 пикселей в градациях серого. На рис. 1 приведены примеры рукописных цифр из данной базы.



Рисунок 1 – Пример первых 12 цифр из обучающего набора MNIST

Имеются лучшие результаты одиночных сверточных нейронных сетей по распознаванию рукописных цифр, которые приведены в таблице 1 (<http://yann.lecun.com/exdb/mnist/>).

Как следует из таблицы, результаты применения нейронных сетей различаются в зависимости от использования образов из обучающей выборки: с искажениями (distortion) или без искажений (nodistortion). Лучший результат с использованием искажений показывает глубокая сверточная сеть с ошибкой тестирования 0.35%, а без использования искажений – глубокая сверточная сеть с ошибкой тестирования 0.53%. В классе неглубоких нейронных сетей лучший результат достигается 0.4% при использовании искажений и 0.95% (LeNet-5) без применения искажений. Исходя из анализа данных результатов видно, что ошибка тестирования снижается при эволюции сверточной сети в глубину (увеличение количества слоев) и применении искажений к обучающей выборке. Однако часто возникают задачи получения приемлемой точности распознавания в условиях ограниченных вычислительных ресурсов. В этом случае можно использовать только ограниченную (неглубокую) нейронную сеть с эволюцией слоев в обратном направлении, то есть уменьшении их количества.

Поэтому целью работы является разработка и реализация простой сверточной нейронной сети высокой точности для эффективно распознавания рукописных цифр в условиях ограниченных вычислительных ресурсов.

2. Построение сверточной нейронной сети. Сверточная нейронная сеть объединяет три подхода, используемые для обработки изображений. Это – использование локального рецептивного поля для каждого нейрона сверточного слоя, формирование сверточных слоев в виде набора карт, нейронные элементы которых имеют одинаковые синаптические связи и наличие карт подвыборочного (subsampling) слоя, который повышает инвариантность сети к искажениям. Использование локального рецептивного поля позволяет нейронам одной карты признаков детектировать один и тот же стимул в разных фрагментах изображения. Применение идентичных нейронов в каждой карте позволяет сократить количество настраиваемых синаптических связей сети. Подвыборочный слой осуществляет локальное усреднение или операцию максимума для каждого неперекрывающегося фрагмента карты сверточного слоя, что позволяет уменьшить размерность соответствующих карт признаков. Общая архитектура сверточной сети показана на рис. 2. Она состоит из совокупности сверточных слоев C_1 , C_3 , C_5 и подвыборочных слоев S_2 и S_4 .

Головко Владимир Адамович, д.т.н., профессор, зав. кафедрой интеллектуальных информационных технологий Брестского государственного технического университета.

Михно Е.В., магистрант кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Брич А.Л., ст. преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Михняев А.Л., ст. преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Войцехович Л.Ю., ст. преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Беларусь, БрГТУ, 224017, г. Брест, ул. Московская, 267.

Матюшков А.Л., доцент Белорусского государственного университета информатики и радиоэлектроники.

Таблица 1 – Сверточные нейронные сети различной архитектуры

Convolutionalnets			
Convolutionalnet LeNet-1	subsampling to 16x16 pixels	1.7	LeCunetal. 1998
Convolutionalnet LeNet-4	none	1.1	LeCunetal. 1998
Convolutional net LeNet-4 with K-NN instead of last layer	none	1.1	LeCunetal. 1998
Convolutional net LeNet-4 with local learning instead of last layer	none	1.1	LeCunetal. 1998
Convolutional net LeNet-5, [no distortions]	none	0.95	LeCunetal. 1998
Convolutional net LeNet-5, [huge distortions]	none	0.85	LeCunetal. 1998
Convolutionalnet LeNet-5, [distortions]	none	0.8	LeCunetal. 1998
Convolutional net Boosted LeNet-4, [distortions]	none	0.7	LeCunetal. 1998
Convolutional net, cross-entropy [affine distortions]	none	0.6	Simardetal., ICDAR 2003
Convolutional net, cross-entropy [elastic distortions]	none	0.4	Simardetal., ICDAR 2003
large conv. net, random features [no distortions]	none	0.89	Ranzatoetal., CVPR 2007
large conv. net, unsup features [no distortions]	none	0.62	Ranzatoetal., CVPR 2007
large conv. net, unsuppretraining [no distortions]	none	0.60	Ranzatoetal., NIPS 2006
large conv. net, unsuppretraining [elastic distortions]	none	0.39	Ranzatoetal., NIPS 2006
large conv. net, unsuppretraining [no distortions]	none	0.53	Jarrettetal., ICCV 2009
large/deep conv. net, 1-20-40-60-80-100-120-120-10 [elastic distortions]	none	0.35	Ciresanetal. IJCAI 2011

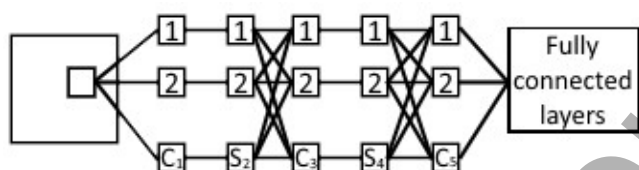


Рисунок 2 – Основная структура сверточной нейронной сети

Как уже отмечалось, сверточный слой состоит из множества карт признаков, где нейроны каждой карты содержат одни и те же наборы весов и порогов. В результате, нейронные элементы каждой карты признаков выполняют одни и те же операции над различными частями изображения. Для сканирования изображения используется метод скользящего окна, которое называется локальным рецептивным полем (receptivefield) для соответствующего ему нейрона карты признаков. Поэтому, если размер скользящего окна, называемого ядром, равняется $p \times p$ (рецептивное поле), то каждый нейрон сверточного слоя связан с p^2 элементами соответствующей области рецептивного поля изображения. Каждое рецептивное поле (окно) во входном пространстве отображается на специальный (отдельный) нейрон в каждой карте признаков. Если скользящее окно сканирует изображение с единичным шагом, то количество нейронов в каждой карте признаков определяется следующим образом:

$$D(C_1) = (n - p + 1)(n - p + 1), \quad (1)$$

где $n \times n$ – это размер изображения.

Если скользящее окно сканирует изображение с шагом s , то количество нейронов в каждой карте признаков в общем случае определяется следующим образом:

$$D(C_1) = \left(\frac{n - p}{s} + 1\right) \left(\frac{n - p}{s} + 1\right). \quad (1)$$

Общее количество различных синаптических связей в сверточном слое определяется, как

$$V(C_1) = M(p^2 + 1), \quad (2)$$

где M – общее количество карт признаков в сверточном слое.

Как следует из последнего выражения, использование сверточной сети позволяет сократить общее количество настраиваемых синаптических связей по сравнению с многослойным перцептроном, за счет использования идентичных нейронов в каждой карте признаков.

Для упрощения математического описания сверточного слоя, представим пиксели входного изображения в одномерном пространстве. Тогда выходное значение ij -го нейрона для k -ой карты признаков в сверточном слое определяется, как

$$y_{ij}^k = F(S_{ij}^k), \quad (4)$$

$$S_{ij}^k = \sum_c w_{cij}^k x_c - T_{ij}^k, \quad (5)$$

где $c=1, p^2$, F – функция активации, S_{ij}^k – взвешенная сумма ij -го элемента k -й карты признаков, w_{cij}^k – вес между c -м элементом входного слоя и ij -м элементом k -й карты признаков, T_{ij}^k – порог для ij -го элемента k -й карты признаков.

Как уже отмечалось, нейроны каждой карты признаков имеют одинаковый набор весов и порогов. В результате, из одного и того же изображения можно извлечь множество различных признаков. Далее, эти признаки обрабатываются следующим слоем S_2 с целью уменьшения размерности карт признаков [7]. Данный слой называется объединяющим (pooling) или подвыборочным (subsampling). Он осуществляет сжатие карт признаков сверточного слоя при помощи операции максимизации или локального усреднения различных областей карт признаков. Для этого каждая карта признаков сверточного слоя разбивается на неперекрывающиеся области размером $k \times k$. Каждая область отображается в один нейрон соответствующей карты подвыборочного слоя. Следует также отметить, что каждая карта признаков сверточного слоя связана лишь с одной картой в пулинговом слое. Каждый нейрон подвыборочного слоя вычисляет среднее или максимальное значение K^2 нейронов в сверточном слое:

$$z_j = \frac{1}{k \times k} \sum_{i=1}^{k \times k} y_i, \quad (6)$$

$$z_j = \max(y_j).$$

Число нейронов в каждом пулинговом слое соответственно

$$D(S_2) = \frac{D(C_1)}{k^2} \tag{3}$$

Число карт признаков в слое пулинга будет таким же, как и в сверточном слое. Далее опять следуют сверточный (C_3), пулинговый (S_4) и сверточный (C_5) слои. При этом каждый нейронный элемент пулингового слоя в общем случае имеет синаптические связи со всеми нейронами сверточного слоя (рис. 1). Таким образом, сверточная нейронная сеть представляет собой сочетание сверточных и пулинговых слоев, которые выполняют нелинейное иерархическое преобразование входного пространства образов. Последний блок сверточной нейронной сети является многослойным перцептроном, машиной опорных векторов или другим классификатором (рис. 3).

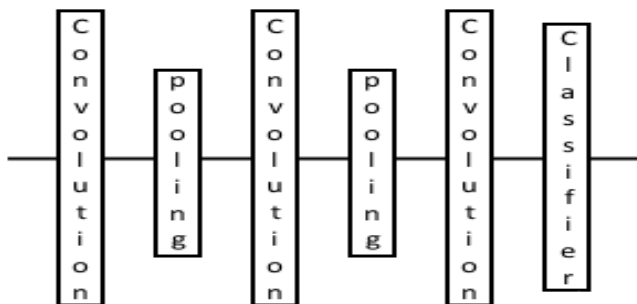


Рисунок 3 – Общая структура сверточной нейронной сети

3. Архитектура сверточной нейронной сети LeNet-5. Рассмотрим традиционную сверточную нейронную сеть (LeNet5) для классификации рукописных цифр (рис. 4) [22]. Входное изображение имеет размер 32x32 пикселя. Скользящее окно размерности 5x5 сканирует изображение с шагом 1 и каждый фрагмент изображения поступает на соответствующий нейронный элемент карты признаков сверточного слоя C_1 нейронной сети. Слой C_1 состоит из шести карт признаков, где каждая карта согласно выражению (1) содержит 28x28 нейронов. Слой S_2 представляет собой подвыборочный слой с шестью картами признаков. Каждый нейрон подвыборочного слоя вычисляет среднее значение 4-х нейронов в сверточном слое, то есть $k=2$. Тогда в соответствии с выражением (7), каждая карта под-

выборочного слоя S_2 содержит 14x14 нейронных элементов. Слой C_3 является сверточным слоем с 16-ю картами признаков и ядром сканирования 5x5 каждой карты подвыборочного слоя. Тогда размерность каждой карты признаков сверточного слоя C_3 равняется 10x10 нейронных элементов.

В LeNet-5 слои S_2 и C_3 образуют не полностью связанную нейронную сеть, где нейроны этих слоев имеют синаптические связи между собой в соответствии с таблицей 2 [22]. Каждые столбец и строка таблицы характеризуют соответственно карты признаков сверточного C_3 и подвыборочного S_2 слоев. Знак X указывает наличие связей между нейронами соответствующих карт.

Слой S_4 является подвыборочным слоем с 16-ю картами признаков и ядром 2x2 для каждой карты признаков сверточного слоя. Поэтому каждый нейрон подвыборочного слоя S_4 вычисляет среднее значение 4-х нейронов в сверточном слое C_3 , то есть $k=2$. Поэтому размерность карты признаков данного слоя составляет 5x5 нейронных элементов. Следующий сверточный слой C_5 сканирует подвыборочный слой окном 5x5. В результате карты признаков данного слоя отображаются на 120 нейронов слоя C_5 . При этом каждый нейрон имеет синаптические связи со всеми нейронами сверточного слоя. Следующий слой F_6 содержит 84 нейронных элемента с функцией активации гиперболический тангенс и функционирует как классический перцептронный слой. Выходной слой состоит из 10 нейронных элементов, каждый из которых формирует выходное значение в соответствии с радиально-базисной функцией активации.

4. Обучение сверточных нейронных сетей. Для обучения сверточных нейронных сетей используется алгоритм обратного распространения ошибки, адаптированный к архитектуре сверточной сети. Целью обучения является минимизация суммарной квадратичной ошибки сети, которая характеризует разницу между реальными и эталонными выходными значениями сети. Значение суммарной квадратичной ошибки для L тренировочных наборов определяется следующим образом:

$$E_s = \frac{1}{2} \sum_{k=1}^L \sum_{j=1}^m (y_j^k - e_j^k)^2 \tag{4}$$

где y_j^k и e_j^k – соответственно, полученное и эталонное выходное

Таблица 2 – Связи между слоями S_2 и C_3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	X				X	X	X			X	X	X	X		X	X
2	X	X				X	X	X			X	X	X	X		X
3	X	X	X				X	X	X			X		X	X	X
4		X	X	X			X	X	X	X			X		X	X
5			X	X	X			X	X	X	X		X	X		X
6				X	X	X			X	X	X	X		X	X	X

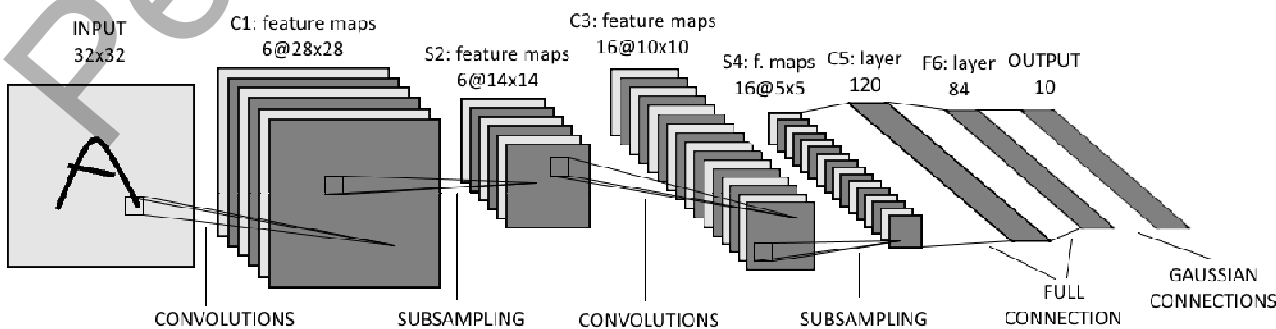


Рисунок 4 – Архитектура LeNet-5

значение j -й элемента для k -го образа. Затем, используя метод градиентного спуска, мы можем написать в случае обучения пакетами, что

$$w_{cij}(t+1) = w_{cij}(t) - \alpha \frac{\partial E(r)}{\partial w_{cij}(t)}, \quad (5)$$

где α – это скорость обучения, $E(r)$ – это среднеквадратичная ошибка для r образов. Поскольку нейроны каждой карты в сверточном слое содержат одинаковые веса, частная производная которых $\frac{\partial E(r)}{\partial w_{cij}(t)}$ равняется сумме частных производных для всех нейронов карты признаков:

$$\frac{\partial E(r)}{\partial w_{cij}} = \sum_{i,j} \frac{\partial E(r)}{\partial \omega_{cij}}. \quad (10)$$

Таким образом, для вычисления определенного весового коэффициента необходимо взять производные по этой связи для всех нейронных элементов карты признаков и просуммировать их. Затем полученное в соответствии с выражением (10) значение весового коэффициента присвоить одноименным связям для всех нейронных элементов карты признаков сверточного слоя. Используя предыдущие результаты, можно получить следующее выражение для модификации весовых коэффициентов:

$$w_{cij}(t+1) = w_{cij}(t) - \alpha(t) \sum_{i,j} \sum_k \gamma_{ij}^k F'(s_{ij}^k) x_c^k, \quad (6)$$

где $c=1, p^2$, $F'(s_{ij}^k) = \frac{\partial y_{ij}^k}{\partial s_{ij}^k}$ – производная функции активации для k -го образа, s_{ij}^k – взвешенная сумма, γ_{ij}^k – ошибка ij -го нейрона в карте признаков для k -о образа, x_c^k – c -й вход.

5. Редуцированная архитектура сверточной сети для распознавания рукописных цифр. Во многих практических приложениях важной проблемой являются ограниченные вычислительные ресурсы, которые не позволяют использовать сложные нейронные сети. Поэтому упрощение архитектуры нейронных сетей с сохранением эффективности обработки является актуальной проблемой. В данном разделе рассматривается сверточная нейронная сеть с более простой архитектурой по сравнению с LeNet-5 [8–11]. Архитектура упрощенной сверточной нейронной сети для распознавания рукописных цифр изображена на рис. 5.

Данная сеть состоит из 5 слоёв: 3-х сверточных (C_1, C_3, C_5) и 2-х пулинговых (S_2, S_4). Входное изображение имеет размер 28x28 пикселей. Изображение сканируется скользящим окном размерности 5x5 с шагом 1, и каждый фрагмент изображения поступает на соответствующий нейронный элемент карты признаков сверточного слоя C_1 нейронной сети. Сверточный слой C_1 состоит из 6 карт признаков, где каждая карта содержит 24x24 нейронов. Карты признаков сверточного слоя разбиваются на неперекрывающиеся области

размером 2x2, каждая из которых отображается на соответствующий нейронный элемент пулингового слоя. Пулинговый слой S_2 содержит 8 карт признаков размерностью 12x12 нейронов для каждой карты признаков. Сверточный слой C_3 содержит 16 карт признаков размерностью 8x8 и ядром обхода 5x5. По сравнению с традиционной сетью LeNet-5 слои S_2 и C_3 являются полностью связанными, то есть каждый нейрон слоя S_2 имеет связь со всеми нейронными элементами слоя C_3 . Слой S_4 представляет собой пулинговый слой с ядром обхода сверточного слоя 2x2, который состоит из 16 карт признаков размерностью 4x4 нейронных элементов. Последний слой C_5 является выходным слоем из 10 нейронных элементов, классифицирующим образы. Основные отличия представленной архитектуры сверточной нейронной сети от LeNet-5 следующие:

- 1) отсутствует 2 слоя размерностью 120 и 84 нейронных элемента;
- 2) слои S_2 и C_3 являются полностью связанными;
- 3) сигмоидная функция активации используется во всех сверточных слоях.

Рассмотрим применение рассмотренной сверточной сети для распознавания рукописных цифр на основе базы данных MNIST. База данных MNIST содержит 60 000 изображений для обучения и 10 000 изображений для тестирования. Каждое изображение имеет размерность 28x28 пикселей в градациях серого. Для обучения сети будем использовать групповое обучение с алгоритмом обратного распространения ошибки.

Таблица 3 – Сравнительный анализ

Классификатор	Ошибка тестирования (%)
Сверточная сеть LeNet-1	1.7
Сверточная сеть LeNet-4	1.1
Сверточная сеть LeNet-5, [no distortions]	0.95
Сверточная сеть LeNet-5, [distortions]	0.8
Редуцированная сверточная сеть, [nodistortions]	0.71

Размерность обучаемой группы составляет 50 образов; значение шага обучения изменяется с 0.8 до 0.0001. После обучения сети ошибка распознавания на тестовом множестве составила 0.71%. Сравнительный анализ различных сетей этого класса приведён в таблице 3 (<http://yann.lecun.com/exdb/mnist/>). Как следует из таблицы, лучшим результатом сверточной сети LeNet-5 без использования искусственных искажений (distortions) в обучающей выборке является ошибка распознавания 0.95%, а при использовании искажений – 0.8%. Таким образом, использование редуцированной сверточной нейронной сети позволяет достичь более высокой точности распознавания по сравнению с традиционной архитектурой сверточной сети. Результаты обработки каждого слоя для числа 7 изображены в таблице 4.

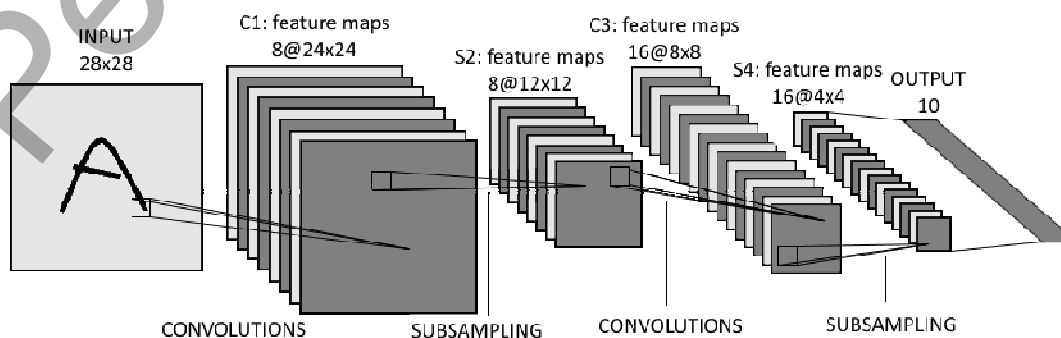







Рисунок 5 – Архитектура упрощенной сверточной нейронной сети

Таблица 4 – Результаты обработки числа 7

Layer type	Number and size of map	Processing results (%)
Input	1@28x28	
Convolutional	8@24x24	
Subsampling	8@12x12	
Convolutional	16@8x8	
Subsampling	16@8x8	

Заключение. В данной статье рассматривается сверточная нейронная сеть для распознавания рукописных цифр. Разработана упрощенная архитектура сверточной нейронной сети, которая позволяет классифицировать рукописные цифры с точностью 99,29% (ошибка тестирования 0,71%), что является лучшим результатом в классе неглубоких (shallow) сверточных нейронных сетей (<http://yann.lecun.com/exdb/mnist/>).

Предложенная сверточная сеть имеет меньшее количество слоев нейронных элементов по сравнению с конвенциональной сетью и позволяет функционировать с высокой точностью в условиях ограниченных ресурсов. Основные отличия предложенной сети от LeNet-5 следующие: 1) убрано два слоя сети с 120 и 84 нейронными элементами; 2) слои S2 и С3 являются полносвязными слоями; 3) использование сигмоидных функций активации во всех сверточных слоях.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. LeCun, Y. Deep Learning Nature / Y. LeCun, Y. Bengio, G. E. – Hinton. – 2015. – Vol. 521. – P. 436–444.
2. Golovko, V. Neural Network and Artificial Immune Systems for Malware and Network Intrusion Detection / V. Golovko, S. Bezobrazov, P. Kachurka, L. Vaitsekhovich // Studies in computational intelligence. – Springer Berlin/Heidelberg, 2010. – Vol. 263: Advances in machine learning II. – P. 485–513.
3. Golovko, V. Neural Network Ensembles for Intrusion Detection. In the 4th IEEE / V. Golovko, P. Kachurka, L. Vaitsekhovich // Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS-2007). – 2007. – IDAACS 2007. – P. 578–583.
4. Dziomin, U. A Multi-Agent Reinforcement Learning Approach for the Efficient Control of Mobile Robots / U. Dziomin, A. Kabysh, R. Stetter and V. Golovko // in book Advanced in Robotics and Collaborative Automation. – River Publishers, 2015. – P. 123–146.
5. Golovko, V. Towards Automatic Epileptic Seizure Detection in EEGs based on Neural Networks and Largest Lyapunov Exponent / V. Golovko, S. Artsiomenka, V. Kisten, V. Evstigneev // International Journal of Computing. – 2015. – 14 (1). – 36–47.
6. Hinton, G. E. A fast learning algorithm for deep belief nets. Neural Computation / G. E. Hinton, S. Osindero, Y. Teh – 2006. – № 18. – P. 1527–1554.
7. Hinton, G. Training products of experts by minimizing contrastive divergence / G. Hinton // Neural Computation. – 2002. – № 14. – P. 1771–1800.
8. Hinton, G. Reducing the dimensionality of data with neural networks / G. Hinton, R. Salakhutdinov // Science. – 2006. – № 313 (5786). – P. 504–507.
9. Hinton, G. E. A practical guide to training restricted Boltzmann machines. // Tech. Rep. 2010-000. – Toronto: Machine Learning Group, University of Toronto, 2010.
10. Bengio, Y. Learning deep architectures for AI. Foundations and Trends in Machine Learning. – 2009– № 2(1), 1-127 // Simple shallow convolutional neural network 9.
11. Bengio, Y. Greedy layer-wise training of deep networks / Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle; in B. Scholkopf, J. C. Platt, T. Hoffman (Eds.) // Advances in neural information processing systems. – MA: MIT Press, Cambridge, 2007. – № 11. – P. 153–160.
12. Golovko, V. A Learning Technique for Deep Belief Neural Networks / V. Golovko, A. Kroschanka, U. Rubanau, S. Jankowski // Neural Networks and Artificial Intelligence. – Springer, 2014. – Vol. 440. Communication in Computer and Information Science. – P. 136–146.
13. Golovko, V. From multilayer perceptron to deep belief neural networks: training paradigms and application / Lectures on neuroinformatics. – Moscow, 2015. – P. 47–84.
14. Golovko, Vladimir. A New Technique for Restricted Boltzmann Machine Learning / Vladimir Golovko, Aliaksandr Kroschanka, Volodymyrurchenko, Stanislaw Jankowski, Douglas Treadwell // Proceedings of the 8th IEEE International Conference IDAACS-2015, Warsaw 24-26 September 2015. – Warsaw, 2015 – P. 182–186.
15. Golovko, V. The nature of unsupervised learning in deep neural networks: A new understanding and novel approach / V. Golovko, A. Kroschanka // Optical memory and neural networks. – 2016. – V. 25. – № 3 – P. 127–141.
16. Golovko, V. Theoretical Notes on Unsupervised Learning in Deep Neural Networks/V. Golovko, A. Kroschanka // Proceedings of the 8th International Joint Conference on Computational Intelligence. – NCTA 2016. – Porto: Portugal, 9–11 November 2016. – P. 91–96.
17. Golovko, V. Deep Neural Networks: A theory, application and new trends / V. Golovko // PROCEEDINGS OF THE 13th INTERNATIONAL CONFERENCE on PATTERN RECOGNITION AND INFORMATION PROCESSING (PRIP2016). – Minsk: Publishing Center of BSU. – P. 33-37.
18. Golovko, V. A Simple Shallow Convolutional Neural Network for Accurate Handwritten Digit Classification / V. Golovko, E. Mikhno, A. Brich // PROCEEDINGS OF THE 13th INTERNATIONAL CONFERENCE on PATTERN RECOGNITION AND INFORMATION PROCESSING (PRIP2016). – Minsk: Publishing Center of BSU. – P. 209–212.
19. LeCun, Y. Gradient-Based Learning Applied to Document Recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE. – November 1998. – № 86(11). – P. 2278–2324.
20. LeCun, Y. Backpropagation applied to handwritten zip code recognition / Y. LeCun, B. Boser, J. Denker, R. Henderson, R. Howard, W. Hubbard, L. Jackel // Neural computation. – 1989. – № 1(4). – P. 541–551.
21. LeCun, Y. Object recognition with gradientbased learning / Y. LeCun, P. Haffner, L. Bottou, Y. Bengio // Contour and Grouping in Computer Vision. – P. 319–345, Springer, 1999.
22. LeCun, Y. Gradient-Based Learning Applied to Document Recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE. November, 1998. – № 86(11). – P. 2278–2324.

Материал поступил в редакцию 05.01.2017

The convolutional neural network for accurate handwritten digit recognition is considered. In this work we have shown, that high accuracy can be achieved using reduced shallow convolutional neural network without adding distortions for digits. The main contribution of this paper is to point out how using simplified convolutional neural network is to obtain test error rate 0.71% on the MNIST handwritten digit benchmark. It permits to reduce computational resources in order to model convolutional neural network.

УДК 004.89

Головко В.А., Крощенко А.А., Хацкевич М.В.

ТЕОРИЯ ГЛУБОКОГО ОБУЧЕНИЯ: КОНВЕНЦИАЛЬНЫЙ И НОВЫЙ ПОДХОД

Введение. Глубокие нейронные сети (deep neural networks) представляют собой нейронные сети с множеством слоев нейронных элементов. Существуют следующие глубокие нейронные сети (DNN):

- нейронные сети глубокого доверия (deep belief neural networks);
- глубокий перцептрон (deep perceptron);
- глубокая сверточная нейронная сеть (deep convolutional neural networks);
- глубокая рекуррентная нейронная сеть (deep recurrent neural networks);
- глубокий автоэнкодер (deep autoencoder);
- глубокая рекуррентная-сверточная нейронная сеть (deep R-CNN).

Исторически, первыми появились нейронные сети глубокого доверия и глубокий перцептрон, которые в общем случае представляют собой многослойный перцептрон с более чем двумя скрытыми слоями [4]. Основным отличием нейронной сети глубокого доверия от глубокого перцептрона является то, что нейронная сеть глубокого доверия не является в общем случае сетью с прямым распространением сигнала (feed forward neural network). До 2006 года в научной среде была приоритетной парадигма, что многослойный перцептрон с одним, максимум двумя скрытыми слоями является более эффективным для нелинейного преобразования входного пространства образов в выходное по сравнению с перцептроном с большим количеством скрытых слоев. Считалось, что перцептрон с более чем двумя скрытыми слоями не имеет смысла применять. Данная парадигма базировалась на теореме, что перцептрон с одним скрытым слоем является универсальным аппроксиматором. Другой аспект этой проблемы заключается в том, что все попытки использовать алгоритм обратного распространения ошибки (backpropagation algorithm) для обучения перцептрона с тремя и более скрытыми слоями не приводили к улучшению решения различных задач. Это связано с тем, что алгоритм обратного распространения ошибки является неэффективным для обучения перцептронов с тремя и более скрытыми слоями при использовании сигмоидальной функции активации. Это происходит из-за проблемы *исчезающего градиента* (vanishing gradient problem). Так, например, максимальное значение производной сигмоидной функции активации $F(S_j)$ равно 0,25. Поэтому использование обобщенного дельта-правила для обучения перцептрона с большим количеством скрытых слоев приводит к затуханию градиента при распространении сигнала от последнего слоя к первому. В 2006 Хинтон (Hinton) предложил «жадный» алгоритм послойного обучения (greedy layer-wise algorithm) [1], который стал эффективным средством обучения глубоких нейронных сетей. Было показано, что глубокая нейронная сеть имеет большую эффективность нелинейного преобразования и представления данных по сравнению с традиционным перцептроном. Такая сеть осуществляет глубокое иерархическое преобразование входного пространства образов. В результате первый скрытый слой выделяет низкоуровневое пространство признаков входных данных, второй слой детектирует пространство признаков более высокого уровня абстракции и т. д. [2].

2. Архитектура глубокой нейронной сети. Как уже отмечалось, глубокая нейронная сеть содержит множество скрытых слоев нейронных элементов (рис. 1) и осуществляет глубокое иерархическое преобразование входного пространства образов.

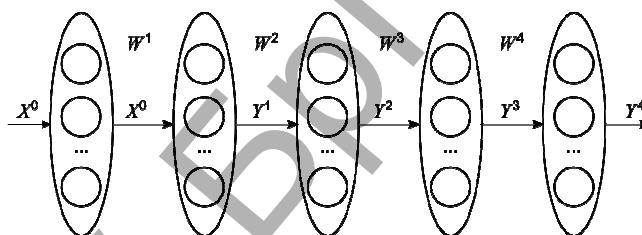


Рисунок 1 – Глубокая нейронная сеть

Выходное значение j -го нейрона k -го слоя определяется следующим образом:

$$y_j^k = F(S_j^k), \quad (1)$$

$$S_j^k = \sum_{i=1}^n w_{ij}^k y_i^{k-1} + T_j^k, \quad (2)$$

где F – функция активации нейронного элемента, S_j^k – взвешенная сумма j -го нейрона k -слоя, w_{ij}^k – весовой коэффициент между i -м нейроном $(k-1)$ -го слоя и j -м нейроном k -го слоя, T_j^k – пороговое значение j -го нейрона k -го слоя.

Для первого (распределительного) слоя

$$y_i^0 = x_i. \quad (3)$$

В матричном виде выходной вектор k -го слоя

$$Y^k = F(S^k) = F(W^k Y^{k-1} + T^k), \quad (4)$$

где W – матрица весовых коэффициентов, Y^{k-1} – выходной вектор $(k-1)$ -го слоя, T^k – вектор пороговых значений нейронов k -го слоя.

Если глубокая нейронная сеть используется для классификации образов, то выходные значения сети часто определяются на основе функции активации softmax:

$$y_j^f = \text{softmax}(S_j) = \frac{e^{S_j}}{\sum_i e^{S_i}}.$$

Несмотря на архитектурные различия глубоких нейронных сетей, *принципы их обучения являются идентичными*. Поэтому рассмотрим основные концепции обучения таких сетей на примере глубокого перцептрона.

3. Конвенциональные методы обучения глубоких нейронных сетей. Рассмотрим обучение глубоких нейронных сетей. Существуют два основных метода обучения:

1. **Метод с предварительным обучением**, который состоит из двух этапов:

Крощенко Александр Александрович, старший преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Хацкевич М.В., ст. преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Беларусь, БрГТУ, 224017, г. Брест, ул. Московская, 267.

Физика, математика, информатика