

# ПРОСТАЯ АРХИТЕКТУРА СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР

**В.В. Лукович**

*Международный научно-учебный центр информационных технологий и систем НАН Украины и Министерства образования и науки Украины*

Рассмотрена простая сверточная нейронная сеть для распознавания рукописных цифр, классификатор в которой имеет только один слой нейронов. На базе данных MNIST показано, что такая нейронная сеть является эффективным средством распознавания рукописных цифр.

Розглянуто просту згорткову нейронну мережу для розпізнавання рукописних цифр з класифікатором, який має тільки один шар нейронів. На базі даних MNIST показано, що така нейронна мережа є ефективним засобом розпізнавання рукописних цифр.

## ВВЕДЕНИЕ

Для решения практических задач распознавания широко применяются многослойные перцептроны, обучаемые по методу обратного распространения ошибки (back propagation). Эти нейронные сети способны реализовать сложные, многомерные и нелинейные отображения. Тем не менее прямое использование многослойных перцептронов для распознавания изображений не является оптимальным вариантом. Во-первых, количество элементов изображения составляет как минимум несколько сотен, и тогда количество элементов массива весовых коэффициентов только первого слоя сети может достичь нескольких десятков тысяч. Во-вторых, интерпретация входного изображения как одномерного массива игнорирует существенные свойства изображений: наличие локальных признаков и большую корреляцию значений соседних элементов изображения. Поэтому для распознавания изображений более целесообразно использовать нейронную сеть, определенным образом специализированную для работы именно с изображениями.

**Цель данной работы** — исследование простой архитектуры нейронной сети для распознавания изображений рукописных цифр.

Работа выполнена с использованием развиваемого в отделе нейросетевых технологий обработки информации нейросетевого подхода к решению задач искусственного интеллекта, который основан на идеях Н.М. Амосова [1] и его научной школы [3].

## СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Одним из вариантов специализированной нейронной сети для распознавания изображений является архитектура сверточной нейронной сети [4]. Основные принципы построения этой архитектуры заключаются в следующем:

1) Сеть в целом является многослойным перцептроном с прямыми связями.

2) Нейроны в нескольких первых слоях имеют 2-мерные локальные рецептивные поля.

3) Нейроны-детекторы признаков формируют 2-мерные карты признаков способом, подобным математической операции свертки (convolution). Над картами признаков выполняется процедура уменьшения пространственной разрешающей способности.

4) Входом первого слоя нейронов-детекторов признаков является входное изображение. Входом каждого из следующих слоев детекторов признаков являются карты признаков с уменьшенным пространственным разрешением, сформированные предыдущим слоем детекторов признаков.

Архитектура сверточной нейронной сети интегрирует выделение элементарных признаков изображения, формирование более сложных признаков на высших уровнях обработки и распознавание. Обучение всей сети выполняется единым алгоритмом обратного распространения ошибки (back propagation). Базовый модуль сверточной нейронной сети состоит из двух слоев нейронов. Нейроны первого слоя выполняют функцию детекторов признаков. Выходы нейронов первого слоя поступают на вход нейронов второго слоя, которые осуществляют группировку признаков (feature pooling). Группировка признаков осуществляется путем вычисления некоторой функции от значений признаков в пределах локального рецептивного поля нейрона из второго слоя. Простейший вариант такой функции — простая сумма. Очевидно, что пространственное разрешение карты признаков на выходе слоя нейронов, осуществляющих группировку признаков, будет меньше, чем у карты признаков на входе вышеупомянутого слоя нейронов. Уменьшение пространственной разрешающей способности повышает устойчивость представления признаков к небольшим вариациям их расположения на входе детекторов признаков. Последовательное соединение базовых модулей обеспечивает выделение сложных признаков из рецептивных полей, размер которых возрастает, при одновременном увеличении устойчивости к нежелательным вариациям входных данных. Таким образом, сверточная нейронная сеть реализует важные свойства сетевого интеллекта, сформулированные Н.М. Амосовым [2] — обобщение моделей входных образов и представление этих моделей в виде иерархии признаков с возрастающей степенью сложности.

Конкретными представителями описанной архитектуры являются нейронные сети LeNet5 [5], LeNet6 [6] и simple CNN [7]. Входной слой сетей LeNet5 и LeNet6 имеет размер  $32 \times 32$  пикселя. В сети LeNet5 первый слой насчитывает 6 нейронов с квадратным рецептивным полем  $5 \times 5$ . Выходы нейронов первого слоя формируют первую карту признаков с размерностью  $6 \times 28 \times 28$ . Выходом первого слоя группировки признаков является карта признаков с размерностью  $6 \times 14 \times 14$ . Второй слой детекторов признаков содержит 16 нейронов и формирует карту признаков с размерностью  $16 \times 10 \times 10$ . Нейроны этого слоя имеют рецептивные поля  $5 \times 5$  и получают входные данные с первой карты признаков. Второй слой группировки признаков также уменьшает вдвое размер карты признаков, соответственно

на его выходе вторая карта признаков имеет размерность  $16 \times 5 \times 5$ . Эта карта признаков является входом 100 нейронов следующего слоя. Последний слой имеет 10 нейронов — по одному на каждый класс, который должен распознаваться.

Нейронная сеть LeNet6 имеет архитектуру, подобную LeNet5, но количество нейронов в ней значительно больше — по 50 в первом и втором, 200 нейронов в третьем. Количество параметров для обучения LeNet5 составляет примерно 60 000, а для LeNet6 — около 315 000. Сверточная нейронная сеть simple CNN, описанная в работе [7], не имеет слоев группировки признаков. Уменьшение пространственной разрешающей способности карт признаков осуществляется за счет того, что нейроны-детекторы признаков «сканируют» входные поля с шагом, равным двум. Нейроны первых трех слоев имеют квадратные рецептивные поля  $5 \times 5$ . Нейроны второго и третьего слоя получают на вход сигнал со всех плоскостей входных карт признаков. Первый слой сети насчитывает 6 нейронов, второй — 60, третий — 100, последний — 10. Эта сеть имеет примерно 180 000 параметров для обучения.

Как уже было ранее сказано, сверточная нейронная сеть объединяет выделение признаков изображения, формирование сложных признаков (т.е. осуществляет отображение входных изображений в пространство признаков) и распознавание. Анализ архитектуры конкретной сверточной нейронной сети позволяет найти и отделить часть сети, которая выполняет выделение признаков изображения, от части сети, которая выполняет собственно распознавание. Последовательно рассматривая слои сети от входных до выходных, можно найти нейроны, которые получают на вход не часть карты признаков, а сразу всю карту признаков. Такие нейроны принадлежат части нейронной сети, которая является классификатором.

Известно, что 2-слойный персептрон, обучающийся по алгоритму back propagation, может выполнять функцию универсального классификатора (т.е. распознавать классы, образующие невыпуклые и многосвязные области в пространстве признаков). Довольно часто универсальный классификатор используется в сверточных нейронных сетях. Например, такой классификатор присутствует в нейронных сетях LeNet5, LeNet6 и simple CNN [7]. Автор прежде исследовал сверточную нейронную сеть [8], два последних слоя которой также являются универсальным классификатором на основе радиальных базисных функций.

Простейшим обучаемым классификатором можно считать линейный классификатор. Исходя из того, что размерность данных на входе классификатора в сверточной нейронной сети достаточно велика, а количество классов в данной задаче распознавания равно всего десяти, можно предположить, что линейный классификатор в данном случае мог бы эффективно обеспечить решение задачи распознавания. Стоит заметить, что авторы архитектуры сверточных нейронных сетей начинали исследования именно с такого варианта, в котором классификатор был линейным. Это легко можно увидеть, проанализировав архитектуру нейронной сети LeNet1 [4]. Но они потом перешли к более сложным архитектурам, не раскрыв полностью потенциальные возможности простой архитектуры.

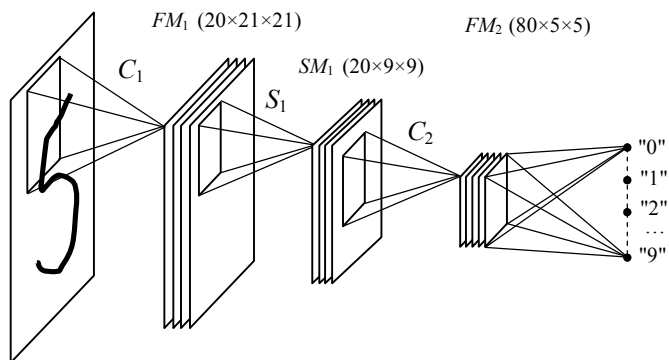


Рис.1. Архитектура простой сверточной нейронной сети

## АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ

Исследования выполнялись с нейронной сетью, архитектура которой показана на рис. 1. Квадратная сетчатка  $28 \times 28$  пикселей является входом сети, на нее подаются изображения цифр, которые нужно распознавать. Первый слой ( $C_1$ ,  $C$  здесь означает convolution — свертку) состоит из 20 нейронов. Эти нейроны являются детекторами признаков, каждый из них имеет квадратное рецептивное поле  $8 \times 8$ . Сетчатка «сканируется» этим рецептивным полем с шагом, равным единице, соответственно первая карта признаков  $FM_1$  имеет вид квадрата  $21 \times 21$  с количеством плоскостей, равным двадцати. Преобразования, которые выполняются нейронами первого слоя, описываются выражением

$$y_{ijk}^{(1)} = \sigma \left( \vartheta_k^{(1)} + \sum_{l=0}^7 \sum_{m=0}^7 w_{klm}^{(1)} x_{i+l, j+m} \right). \quad (1)$$

Верхний индекс в скобках означает порядковый номер слоя нейронной сети. Индекс  $k$  нумерует нейроны, принадлежащие к данному слою, индексы  $i$  и  $j$  отражают координаты карты признаков на выходе данного слоя сети. Индексы  $l$  и  $m$  соответствуют координатам в пределах локального рецептивного поля;  $y_{ijk}^{(1)}$  — элементы первой карты признаков;  $\vartheta_k^{(1)}$  — пороги нейронов,  $w_{klm}^{(1)}$  — весовые коэффициенты нейронов первого слоя. Двумерный массив  $x$  — элементы изображения на входе сети. Функция передачи нейронов имеет сигмоидную нелинейность:  $\sigma(S) = \frac{1}{1 + e^{-S}}$ .

Второй слой обозначен  $S_1$  (здесь  $S$  означает subsampling — уменьшение пространственной разрешающей способности). Он выполняет группировку признаков из карты признаков  $FM_1$  путем вычисления взвешенной суммы в пределах локального рецептивного поля, сканирующего карту признаков с шагом, равным двум. Весовые коэффициенты нейронов второго слоя

вычисляются заранее и не меняются при обучении сети. На выходе этого слоя формируется карта признаков  $SM_1$ .

$$y_{ijk}^{(2)} = \sum_{l=0}^4 \sum_{m=0}^4 \omega_{lm} y_{2i+l, 2j+m, k}^{(1)} \quad (2)$$

где  $y_{ijk}^{(2)}$  — элементы карты признаков  $SM_1$ ,  $\omega_{lm}$  — квадратная матрица, которая является дискретной аппроксимацией двумерной функции Гаусса. Коэффициенты этой матрицы вычисляются следующим образом. Пусть  $\mathbf{v}$  — это 5-элементный вектор с элементами  $\{1, 4, 6, 4, 1\}$ , тогда  $\omega_{lm} = 0,01 \cdot v_l \cdot v_m$ . Плоскости карты признаков  $SM_l$  являются квадратами  $9 \times 9$ .

Третий слой сети имеет 80 нейронов. Нейроны третьего слоя формируют элементы второй карты признаков  $FM_2$  следующим образом:

$$y_{ijp}^{(3)} = \sigma \left( \vartheta_p^{(3)} + \sum_{k=0}^4 \sum_{l=0}^4 \sum_{m=0}^4 w_{pklm}^{(3)} y_{i+l, j+m, k}^{(2)} \right) \quad (3)$$

Индекс  $p$  нумерует нейроны, которые относятся к третьему слою;  $y_{ijp}^{(3)}$  — элементы второй карты признаков,  $\vartheta_p^{(3)}$  — пороги нейронов;  $w_{pklm}^{(3)}$  — весовые коэффициенты нейронов третьего слоя. Вторая карта признаков имеет размерность  $80 \times 5 \times 5$ . Вследствие малого размера плоскостей второй карты группировка признаков для этой карты не является необходимой.

Последний, четвертый слой сети состоит из 10 нейронов. Входом для этих нейронов является вторая карта признаков  $FM_2$ . Каждый из нейронов последнего слоя поставлен в соответствие одному из классов, которые нужно распознавать. Выходные сигналы нейронов последнего слоя вычисляются по правилу softmax:

$$y_q^{(4)} = \frac{e^{S_q}}{\sum_{r=0}^9 e^{S_r}}, \quad S_r = \theta_r^{(4)} + \sum_{p=0}^{79} \sum_{i=0}^4 \sum_{j=0}^4 w_{rpjij}^{(4)} y_{ijp}^{(3)} \quad (4)$$

Индексы  $q$  и  $r$  нумеруют нейроны, которые относятся к четвертому слою;  $\theta_r^{(4)}$  — пороги нейронов;  $w_{rpjij}^{(4)}$  — весовые коэффициенты нейронов. Максимальный сигнал на выходе  $r$ -го нейрона означает, что распознанным следует считать  $r$ -й класс.

Общее количество параметров для обучения сети равно суммарному количеству пороговых и весовых коэффициентов для всех нейронов сети. Оно составляет 61 390.

## ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ

Для обучения сети был использован метод back propagation, который базируется на градиентной минимизации функции ошибки сети в

пространстве, образованном параметрами нейронов сети. Обучение нейронной сети выполняется как итеративный процесс вычисления поправок к параметрам нейронов и модификации параметров нейронов:

$$\Delta\xi = -\eta \cdot \frac{\partial E}{\partial \xi}, \quad \xi' = \xi + \Delta\xi. \quad (5)$$

Условно обозначен  $\xi$  один из вышеуказанных параметров нейронов (весовой или пороговый коэффициент),  $\eta$  — положительное число, значительно меньшее единицы. В данном случае в качестве функции ошибки сети  $E$  использована кросс-энтропия:

$$E = \sum_{n=1}^N \sum_{q=1}^Q t_{nq} \ln \left( \frac{t_{nq}}{y_{nq}} \right). \quad (6)$$

Индекс  $n$  нумерует образцы из обучающего множества,  $q$  — выходы сети,  $t_{nq}$  — целевое или желаемое значение  $q$ -го выхода сети для  $n$ -го образца,  $y_{nq}$  — фактическое значение  $q$ -го выхода сети для  $n$ -го образца.

Процесс обучения сети был организован в виде цикла. Количество итераций задавалось заранее. В течение каждой итерации из обучающей выборки выбирались образцы таким образом, чтобы каждый из них один раз был подан на вход нейронной сети. Очередность подачи образцов менялась для каждой итерации путем случайной перестановки образцов в последовательности. Перерасчет параметров нейронов сети осуществлялся методом стохастического градиента (on-line), т.е. параметры нейронов модифицировались после каждого предъявления образца из обучающей выборки. Начальное значение параметра обучения  $\eta$  устанавливалось равным 0,05, после каждой итерации  $\eta$  умножалось на коэффициент, равный 0,96. Перед обучением задавалось определенное пороговое значение ошибки выхода сети  $\varepsilon = 0,1$ . Обучение сети на текущем образце осуществлялось не только в случае неправильной его классификации, но и при выполнении условия  $\max_q |t_{nq} - y_{nq}| > \varepsilon$ .

Для улучшения результата обучения нейронной сети было использовано искусственное увеличение объема обучающей выборки путем некоторой модификации образцов из этой выборки. Во время обучения каждый образец из обучающей выборки подвергался деформации, перед тем как быть поданным на вход нейронной сети. Параметры деформации генерировались случайно и независимо для каждого предъявления образца. Это означает, что обучающая выборка виртуально становится неограниченной. Изображение деформировалось методом, близким к предложенному в [9]. Суть метода заключается в том, что на изображение накладывается квадратная сетка, узлы которой находятся в центрах пикселей изображения. Затем вершины квадрата независимо друг от друга смещаются по обеим координатам. Пропорционально смещаются и узлы сетки согласно их расположению. Сетка деформируется так, что отрезки прямых, образующих сетку, остаются отрезками прямых, но меняются их длины и углы наклона. Смещения генерируются как равномерно распределенные случайные числа с нулевым

математическим ожиданием. Значения яркости изображения выбираются в местах нахождения узлов сетки после деформации, но проецируются на недеформированную сетку.

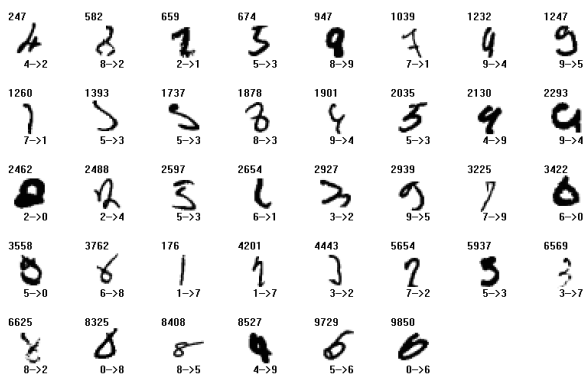


Рис. 2. Образцы (38) из тестовой выборки, которые были неправильно распознаны

Аналогично работе [7], для обучения нейронной сети были использованы первые 50 000 изображений цифр из обучающей выборки. Остальные 10 000 изображений цифр использовались в качестве валидационной выборки. Обучение сети выполнялось в течение 100 итераций. После каждой итерации выполнялось распознавание валидационной выборки. Для итерации с минимальным количеством ошибок на валидационной выборке состояние сети сохранялось в файле и считалось результатом процесса обучения. После обучения сети выполнялось распознавание тестовой выборки. Лучший результат распознавания тестовой выборки из серии экспериментов приведен в табл. 1 (0,38 процента ошибок или 38 ошибок из 10 000 изображений, составляющих тестовую выборку (см. рис. 2).

Для исследований была использована база данных MNIST, которую составил Y. LeCun [5]. Она имеет широкое применение для оценки качества алгоритмов распознавания и их сравнения. База данных насчитывает 70 000 изображений рукописных цифр от 0 до 9. 60 000 изображений составляют обучающую выборку, контрольная выборка состоит из 10 000 изображений цифр. Изображения имеют размер 28×28 пикселей. Уровень яркости пиксела кодируется одним байтом. Обучающая и контрольная выборки находятся в отдельных файлах, также есть два файла, в которых содержатся метки классов. В Интернете базу данных MNIST можно найти по адресу <http://yann.lecun.com/exdb/mnist/>.

## РЕЗУЛЬТАТЫ

Табл. 1 демонстрирует качество распознавания контрольной выборки из базы данных MNIST для различных методов построения классификатора. Аббревиатуры в таблице расшифровываются следующим образом: CNN — Convolutional Neural Network (сверточная нейронная сеть), kNN — классификатор «*k* ближайших соседей», MLP — multilayer perceptron (многослойный перцептрон).

Необходимо отметить, что лучшие результаты (последние две строки из табл. 1) были получены немалой ценой. Так, в работе [12] показано, что комбинация четырех принципиально отличных методов распознавания рукописных цифр (при этом один из них — это сверточная нейронная сеть, описанная в работе [7]) может дать результат, лучший, чем дает каждый из этих методов сам по себе. В работе [13] задача распознавания рукописных цифр из базы данных MNIST решалась методом «грубой силы». Авторы использовали 6-слойную нейронную сеть типа back propagation с количеством нейронов в слоях 2 500, 2 000, 1 500, 1 000, 500 и 10. Количество параметров для обучения этой сети составляет около 12 миллионов. Для программного моделирования сети применялся компьютер с процессором Intel Core2 Quad 9450 2,66 GHz и видеокартой nVidia GTX 280, при этом интенсивно использовались вычислительные ресурсы видеокарты.

**Таблица 1**

Метод	Процент ошибок
Linear classifier [4]	12
kNN Euclidean [4]	5,0
LeNet1 (CNN) [3]	1,7
LeNet4 (CNN) [4]	1,1
LeNet5 (CNN) [4]	0,8
My CNN w/RBF [7]	0,57
PCNC classifier [9]	0,44
SVM w/RBF kernels [10]	0,42
Simard's Simple CNN [6]	0,4
LeNet6 (CNN) [5]	0,38
My Simple CNN (эта работа)	0,38
Classifier combination [11]	0,35
Huge MLP [12]	0,35

## Выводы

Предложен вариант архитектуры простой сверточной нейронной сети для распознавания изображений рукописных цифр. Сеть насчитывает четыре слоя. Первый и третий слои сети осуществляют выделение признаков. Второй слой сети выполняет взвешенное локальное усреднение с уменьшением пространственной разрешающей способности. Последний слой сети является линейным классификатором. Обучение сети осуществляется методом обратного распространения ошибки. Для экспериментальной проверки распознавания рукописных цифр использовалась база данных MNIST. Полученный результат близок к лучшему известному в настоящее время для базы данных MNIST. Это означает, что предложенная простая сверточная нейронная сеть способна эффективно решать задачу распознавания рукописных цифр.

Следует заметить, что предложенная нейронная сеть имеет меньшее количество нейронов и соответственно меньшее количество весовых коэффициентов связей нейронов, чем сети LeNet5, LeNet6 и simple CNN, не говоря уже о сети из работы [13]. Это означает, что в случае практического использования она могла бы обеспечить высокое быстродействие системы распознавания.



1. Амосов Н.М. Моделирование мышления и психики / Н.М. Амосов. — К. : Наукова думка, 1965. — 304 с.  
Amosov N.M. *Modelling of thinking and the mind*. New York: Spartan Books, 1967. 304 p.
2. Амосов Н.М. Алгоритмы разума / Н.М. Амосов. — К. : Наукова думка, 1979. — 223 с.  
Amosov N.M. *Algorithms of the Mind*. Kiev: Naukova Dumka, 1979. 223 p.
2. Kussul E.M., Baidyk T.N., Wunsch D.C. *Neural Networks and Micro Mechanics*. New York: Springer, 2010. 210 p.
3. LeCun Y., Boser B., Denker J.S., Henderson D., Howard R.E., Hubbard W., Jackel L.D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, vol. 1, no 4, pp. 541–551.
4. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, November 1998, vol. 86, no. 11, pp. 2278–2324.
5. Bottou L., Chapelle O., DeCoste D., Weston J. *Large-Scale Kernel Machines*. Cambridge, USA: MIT Press, 2007. 416 p.
6. Simard P.Y., Steinkraus D., Platt J.C. *Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*. *Proc. of the Seventh Int. Conf. on Document Analysis and Recognition, August 03 - 06, 2003 Edinburgh, Scotland*, pp. 958–962.
7. Лукович В. Згорткова нейронна мережа для розпізнавання рукописних цифр / В. Лукович // Пр. 8-ї міжнар. конф. УкрОбраз'2006, — Київ. 2006. — С.135–138.  
Lukovych V. Convolutional Neural Network for Handwritten Digit Recognition. *Proc. of the 8-th all-ukrainian international conference UkrOBRAZ'2006*. Kiev, Aug. 28–31 2006, pp. 135–138.
8. Gosselin B. Improved Hand-written Character Recognition thanks to a New Geometric Distortion Method. *Proc. of the 6th Int. Conf. On Image Processing and its Applications*. Dublin, Ireland, 1997, vol. 1, pp. 327–331.
9. Kussul E., Baidyk T., Wunsch D.C., Makeyev O., Martin A. Permutation Coding Technique for Image Recognition Systems. *IEEE Transactions on Neural Networks*, 2006, vol. 17, no 6. pp.1566–1579.
10. Liu C.-L., Nakashima K., Sako H., Fujisawa H., Handwritten Digit recognition: benchmarking of state-of-the-art techniques. *Pattern Recognition*, 2003, 36(10), pp. 2271–2285.
11. Keysers D. *Comparison and Combination of State-of-the-art Techniques for Handwritten Character Recognition: Topping the MNIST Benchmark*. *Technical Report, IUPR Research Group, DFKI and Technical University of Kaiserslautern*, May 2006, 13 p.
12. Ciresan D.C., Gambardella L.M., Schmidhuber J., Meier U. *Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition*. Available at: <http://arxiv.org/pdf/1003.0358v1>. (05 February 2013).

Получено 12.04.2013