

ПРИМЕНЕНИЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР

Солдатова О.П., Гаршин А.А.

Самарский государственный аэрокосмический университет имени академика С. П. Королёва

Аннотация

Исследуются возможности свёрточных нейронных сетей для распознавания рукописных цифр. Предложена методика обучения сети, реализующая чередование эпох обучения с искажением символов и без искажений. Предложена методика подбора и модификации коэффициента обучения. Представлены экспериментальные исследования применения данного вида сети к распознаванию рукописных цифр базы MNIST.

Ключевые слова: распознавание рукописных цифр, свёрточная нейронная сеть, база рукописных цифр MNIST, эластичные искажения, обобщающая способность сети.

Введение

В статье исследуются возможности применения свёрточных сетей для распознавания рукописных цифр. Нейронные сети успешно применяют при решении задач классификации и распознавания образов, в частности, графически представленных символов [1, 2, 3]. Наиболее распространенной нейросетевой моделью является многослойный перцептрон. Однако применение многослойного перцептрона с традиционной структурой при решении реальных задач распознавания и классификации изображений вызывает определенные трудности.

Во-первых, изображения, как правило, имеют большую размерность, вследствие чего возрастает число нейронов и синаптических связей в сети. В свою очередь, это требует увеличения обучающей выборки, вследствие чего увеличивается время и вычислительная сложность процесса обучения.

Во-вторых, игнорируется топология входных данных. Компоненты входного слоя могут быть представлены в любом порядке, без учета цели обучения. Однако изображения имеют строгую двумерную структуру, в которой существует зависимость между пространственно соседними пикселями [4].

От данных недостатков свободны так называемые свёрточные нейронные сети, которые представляют собой особый класс многослойных перцептронов, специально созданных для распознавания двумерных поверхностей с высокой степенью инвариантности к масштабированию, смещению, повороту, смене ракурса и прочим пространственным искажениям. Недавно свёрточные нейронные сети были успешно применены для распознавания снимков трехмерных объектов [5].

Архитектура свёрточных нейронных сетей реализует три идеи [2, 4]:

- каждый нейрон получает входной сигнал от локального рецептивного поля в предыдущем слое, что обеспечивает локальную двумерную связность нейронов;
- каждый скрытый слой сети состоит из множества карт признаков, на которых все нейроны имеют общие веса, что обеспечивает инвариантность к смещению и сокращение общего числа весовых коэффициентов сети;

- за каждым слоем свёртки следует вычислительный слой, осуществляющий локальное усреднение и подвыборку, что обеспечивает уменьшение разрешения для карт признаков.

Следует отметить, что обучение свёрточных нейронных сетей осуществляется с учителем.

1. Структура сети, функция активации, функция ошибки

Для экспериментальной оценки качества работы свёрточной нейронной сети была разработана автоматизированная система распознавания рукописных цифр, в которой реализована модель свёрточной сети с тремя скрытыми слоями, без слоев подвыборки, но со смещениями рецептивных полей свёрточных нейронов не на один, а на два пикселя [6]. Структура подобной сети описана в [3], однако в предложенной реализации сети во втором слое используется 6 карт признаков, в то время как в [3] – 5 карт. На рис. 1 приведена структура реализованной сети.

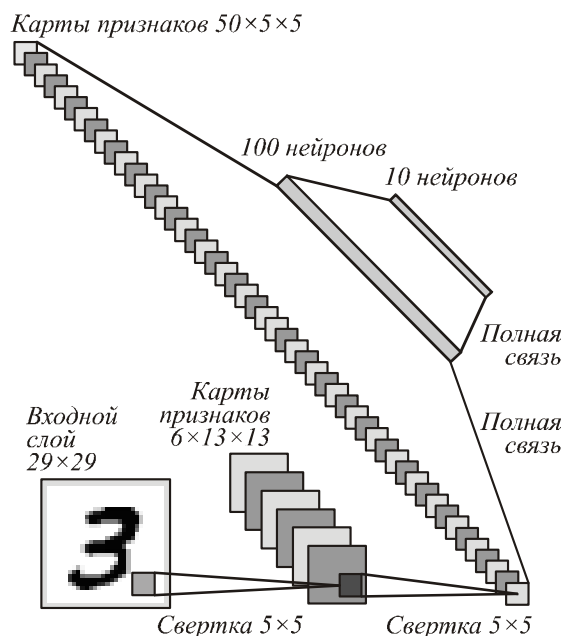


Рис. 1. Структура свёрточной нейронной сети с тремя скрытыми слоями

Рассмотрим подробнее структуру реализованной сети. Для упрощения реализации данная сеть не

имеет слоев подвыборки, а рецептивные поля свёрточных нейронов смещены на два пикселя. Этим обстоятельством обусловлен выбор размера рецептивного поля 5×5 , причем ширина и высота должны быть нечётными числами, чтобы обеспечить попадание нейрона слоя свёртки в центр поля. Такой размер также обеспечивает достаточное наложение рецептивных полей друг на друга. В результате каждый слой свёртки уменьшает размер карты признаков с размера n до размера $(n-3)/2$.

Входными данными нейронной сети являются изображения сегментированных рукописных цифр базы MNIST [7]. Данная база является общепринятой базой среди разработчиков OCR-систем для проверки качества распознавания рукописных цифр. Размер обучающей выборки составляет 60000 символов, размер тестирующей выборки – 10000 символов. В базе MNIST изображения имеют размер 28×28 пикселей. На вход сети подаются изображения, позиционированные в матрицу размером 29×29 пикселей. Число 29 выбрано для того, чтобы и в первом, и во втором слоях свёртки размер карт признаков был целым числом. Таким образом, входной слой содержит 841 нейрон.

Первый скрытый слой является слоем свёртки. Он состоит из шести карт признаков размером 13×13 . В этом слое содержится $13 \times 13 \times 6 = 1014$ нейронов. Каждый элемент карты признаков соединен с рецептивным полем размером 5×5 на входном изображении и единичным смещением. Следовательно, каждый элемент карты имеет 26 обучаемых весовых коэффициентов. Значения всех элементов карты признаков вычисляются путем последовательного сканирования рецептивным полем входного слоя. Интересной особенностью слоев свёртки является тот факт, что при сдвиге входного изображения значения карт признаков будут сдвинуты на ту же самую величину. За счёт этого свёрточные сети обладают инвариантностью к сдвигам и незначительным искажениям входного сигнала [2].

Таким образом, в первом скрытом слое содержится 26364 синаптических связей и 156 обучаемых параметров. Такая экономия памяти и, что главное, вычислительных затрат достигается за счет совместного использования весов картами признаков.

Второй скрытый слой также является слоем свёртки. Он состоит из 50 карт признаков размером 5×5 ($(13-3)/2 = 5$). Меньшее число карт признаков ухудшают работу сети, большее – не улучшают её [3]. В этом слое $5 \times 5 \times 50 = 1250$ нейронов. Каждый элемент в карте признаков связан с шестью областями размером 5×5 шести карт предыдущего слоя. Таким образом, во втором скрытом слое содержится 7550 весов и 188750 связей.

Два первых свёрточных слоя можно рассматривать как слои для извлечения признаков из изображения. Следующие два слоя являются слоями клас-

сификации. В этих слоях каждый нейрон соединён со всеми нейронами предыдущего слоя.

Третий скрытый слой состоит из 100 нейронов. В этом слое имеется 125100 связей и 125100 обучаемых параметров. Для десяти классов изображений, соответствующих десяти цифрам, 100 нейронов достаточно для хорошей обобщающей способности сети [3].

Четвёртый слой является выходным слоем. Он состоит из 10 нейронов, так как распознаются 10 рукописных цифр, и имеет 1010 связей и 1010 весов.

Метод обратного распространения ошибки, используемый при обучении сети, не отличается от аналогичного алгоритма для классического многослойного персептрона, но при реализации учитываются особенности архитектуры свёрточной сети, а именно – совместное использование весов нейронами одной карты признаков.

Рассмотрим особенности формирования сигнала и функции активации, используемой в слоях свёрточной сети. Как и в случае с классическим многослойным персептроном, на вход каждого элемента сети поступает взвешенная сумма – скалярное произведение между входным вектором сигналов и вектором весов, которая затем подается в качестве аргумента функции активации.

В данном случае такой функцией для скрытых слоев сети был выбран гиперболический тангенс:

$$f(a) = A \tanh(Sa), \quad (1)$$

где $f(a)$ – искомое значение элемента, a – взвешенная сумма сигналов предыдущего слоя, A, S – параметры активирующей функции.

Такая функция активации является нечётной, с горизонтальными асимптотами $+A$ и $-A$, а параметр S определяет наклон функции в начале координат.

Для выходного слоя сети функция активации выбирается в соответствии с выбранной функцией ошибки. Формулы используемых в работе функций ошибки приведены ниже:

$$E(w) = \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^M (y_k^j - d_k^j)^2, \quad (2)$$

$$E(w) = - \sum_{j=1}^p \sum_{k=1}^M d_k^j \ln \frac{y_k^j}{d_k^j}, \quad (3)$$

где p – размер обучающей выборки, j – номер обучающего примера, M – количество выходных нейронов, k – номер выходного нейрона, y_k^j – реальное значение сигнала выходного нейрона, d_k^j – ожидаемое значение.

При выборе в качестве функции ошибки среднеквадратичного отклонения (СКО, 2) функцией активации для нейронов выходного слоя также является гиперболический тангенс, при этом ожидаемый отклик сети будет находиться в диапазоне от -1 до $+1$.

В качестве параметров берутся значения $A = 1,7159$ и $S = 2/3$. Такие значения параметров A и S выбраны потому, что гиперболический тангенс (1) имеет следующие полезные свойства [8]:

- максимум второй производной при $a = 1$;
- $f(1) = 1, f(-1) = -1$;
- в начале координат тангенс угла наклона близок к единице;
- ожидаемый отклик сети смещен от границы области значений функции активации в сторону её внутренней части. В этом случае при модификации свободных параметров сети в процессе обучения удаётся избежать резкого возрастания их значений, что позволяет ускорить процесс обучения.

Если функцией ошибки является кросс-энтропия (КЭ, 3), то для нейронов выходного слоя применяется функция активации софтмакс (4):

$$s_i = \frac{\exp(q_i)}{\sum_{k=1}^M \exp(q_k)}, \tag{4}$$

где s_i – выход рассматриваемого нейрона, q_i – взвешенная сумма сигналов предыдущего слоя, M – количество выходных нейронов, $i = \overline{1, M}$. Функции этого вида – это экспоненциальные функции, выходы которых нормируются так, чтобы сумма всех выходов нейронов слоя равнялась 1. Эти функции специально сконструированы для задач классификации таким образом, чтобы выходы можно было интерпретировать как вероятности принадлежности к классу [9].

2. Обучение и тестирование сети

Для обучения сети был выбран градиентный алгоритм наискорейшего спуска, подробно описанный в [8]. Обучение и тестирование сети проводилось с использованием базы рукописных цифр MNIST [7]. Для обучения использовались первые 60000 символов, для тестирования – последние 10000 символов из базы данных MNIST.

Значения исходных синаптических весов для всех слоёв сети с гиперболическим тангенсом выбираются на основе равномерного распределения с нулевым математическим ожиданием и дисперсией, обратной квадратному корню из количества синаптических связей нейрона [8]. При функции активации софтмакс веса выходного слоя инициализируются на основе равномерного распределения с нулевым математическим ожиданием и дисперсией 0,05 (значение выбрано эмпирически).

В разработанной автоматизированной системе реализовано онлайн-обучение (коррекция весов сети производится после каждого обучающего примера).

Использована методика пропуска этапа коррекции весовых коэффициентов при малом значении функции ошибки. Этим достигается значительное

ускорение обучения сети. Кроме того, не возникает паралич обучения сети и ее переобучение.

Для ускорения работы алгоритма значения входных пикселей нормализуются по формуле:

$$y_i = \frac{x_i}{128} - 1, \tag{5}$$

где x_i – значение i -го пикселя изображения из базы, y_i – значение, подаваемое на вход сети [8].

На практике лучшая способность сети к обобщению достигается за счет пополнения обучающей выборки деформированными примерами. Поэтому в системе реализована возможность использования генератора эластичных (нелинейных) искажений [3] обучающих образов с настраиваемым коэффициентом искажений. В отличие от [3] предложенная в данной системе методика обучения предполагает чередование эпох обучения с внесением искажений в обучающие примеры и без искажений обучающих примеров.

Искажения формируются следующим образом. Сначала создается гауссово ядро посредством построения матрицы размером $n \times n$, значение элемента (x, y) которой вычисляется по следующей формуле:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-c)^2 + (y-c)^2}{2\sigma^2}}, \tag{6}$$

где $c = n/2$; $x, y = \overline{0, n-1}$, n – размер ядра. Обычно его выбирают нечётным и пропорциональным радиусу σ так, чтобы в это ядро попали значения функции Гаусса (6) вплоть до уменьшения от максимального значения в пределах $0,1 \div 0,3$. Радиус σ задает силу размытия. Эмпирически были выбраны соответственно следующие значения радиуса и размера ядра: $\sigma = 4, n = 25$.

Далее генерируются две матрицы размером 29×29 (по размеру входного изображения). Значения элементов матриц выбираются как случайные величины, равномерно распределённые на отрезке $[-1, 1]$. К этим матрицам применяется размытие изображения с использованием ранее сформированного ядра [10]. Далее ищется новое значение для каждого пикселя входного изображения с применением билинейной интерполяции и сформированных матриц.

Например, необходимо найти новое значение пикселя в точке А (рис. 2) [3].

Пусть точка А имеет координаты (0,0), числа 3, 7, 5, 9 – это значения пикселей с координатами соответственно (1,0), (2,0), (1,-1) и (2,-1). Допустим, в первой матрице элемент с координатами (0,0) имеет значение $\Delta x = 1,75$, а во второй матрице $\Delta y = -0,5$, – как показывает сплошная стрелка. Для поиска нового значения пикселя в точке А сначала применяется горизонтальная, а затем вертикальная интерполяция. Сначала следует найти координаты конца

стрелки относительно того квадрата, в котором эта стрелка заканчивается, причем началом координат является нижний левый угол квадрата. В данном случае координаты конца стрелки (0,75, 0,5). При применении интерполяции вдоль оси абсцисс новыми значениями пикселя в точке А являются: $3 + 0,75 * (7 - 3) = 6$ и $5 + 0,75 * (9 - 5) = 8$. Вертикальная интерполяция между этими точками дает следующее значение: $8 + 0,5 * (6 - 8) = 7$, – это новое значение пикселя в точке А. Аналогично вычисляется значение каждого пикселя входного изображения при применении нелинейных искажений. Пример применения данного вида искажений приведен на рис. 3.

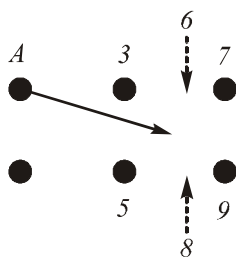


Рис. 2. Вычисление нового значения пикселя в точке А

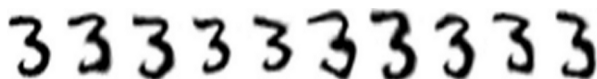


Рис. 3. Пример применения эластичных искажений (с коэффициентом искажений $\epsilon = 25$). Самая первая цифра – неискаженная цифра из базы

Предложена следующая методика подбора коэффициента обучения:

- коэффициент обучения одинаков для нейронов всей сети;
- первые две эпохи сеть обучается с постоянным коэффициентом;
- далее коэффициент автоматически изменяется для каждой эпохи в соответствии с (7) [11].

Применение формулы (7) для модификации коэффициента обучения не требует сложных вычислений и обеспечивает достаточную скорость обучения:

$$\eta(t) = \begin{cases} d\eta(t-1), & \text{если } \delta(t)\bar{\delta}(t-1) < 0 \\ \text{и } |\bar{\delta}(t-1)| > \theta, & \\ u\eta(t-1), & \text{иначе,} \end{cases} \quad (7)$$

где u и d – положительные константы, $0 < d < 1 < u$, θ – порог осцилляции величины $\bar{\delta}(t-1)$;

$$\delta(t) = \frac{E(t) - E(t-1)}{E(t)}, \quad (8)$$

где E – количество ошибок на тестовой выборке; $\bar{\delta}(t) = \alpha * \delta(t) + (1 - \alpha) * \bar{\delta}(t-1)$, $0 < \alpha < 1$ – положительная константа. В экспериментах применяются следующие значения констант: $u = 1,1$, $d = 0,5$,

$\theta = 0,01$, $\alpha = 0,1$ (значения констант выбраны эмпирически).

Коэффициент искажений обучающих образов в экспериментах был выбран $\epsilon = 25$, так как с ним сеть показала хорошую обобщающую способность на тестовом множестве. При тестировании обученной сети искажения не применяются.

3. Экспериментальное исследование

В экспериментах в качестве оценок работы сети применяются:

- точность распознавания символов обучающей выборки;
- точность распознавания символов тестовой выборки;
- количество эпох обучения;
- число пропущенных образцов при обучении на последней эпохе.

Эти оценки применены при обучении сети:

- с функцией ошибки КЭ и с применением эластичных искажений;
- с функцией ошибки КЭ и без применения эластичных искажений;
- с функцией ошибки СКО и с применением эластичных искажений;
- с функцией ошибки СКО и без применения эластичных искажений.

Результаты экспериментов приведены в таблице 1.

Из таблицы 1 видно, что свёрточная нейронная сеть эффективно распознаёт рукописные цифры – точность распознавания довольно высока. При этом становится ясно, что применение эластичных искажений при обучении сети положительно влияет на её обобщающие способности, так как искажения фактически расширяют обучающую выборку. Также видно, что с функцией ошибки КЭ сеть показывает лучшие результаты распознавания как на обучающей, так и на тестовой выборке. Стоит отметить, что количество пропускаемых образцов при обучении с КЭ уже на первых эпохах составляет около 50%. Это говорит о том, что сеть уже на первых эпохах обучается значительному числу образов. Во всех экспериментах число эпох обучения является относительно небольшим (18-24), тогда как в методике обучения, изложенной в [3], число эпох обучения достигало нескольких сотен.

При обучении с применением искажений к обучающим образам (первый и третий эксперименты) на вход сети не поступали неискажённые цифры. Поэтому было решено провести обучение сети еще на нескольких эпохах на неискажённых образцах. Результаты экспериментов приведены в таблице 2.

Из таблицы 2 видно, что результаты распознавания цифр в обоих случаях были улучшены. При этом на тестовой выборке при функции ошибки КЭ сеть показала свой лучший результат – 77 ошибок из 10000! Изображения этих цифр представлены на рис. 4.

Таблица 1. Результаты обучения сети

Функция ошибки, применение искажений при обучении	Точность распознавания обучающей выборки (60000 цифр), %	Точность распознавания тестовой выборки (10000 цифр), %	Количество эпох обучения	Число пропущенных образцов на последней эпохе
КЭ с искажениями, $\epsilon = 25$	99,31	99,14	21	49692
КЭ без искажений	99,77	98,61	24	54602
СКО с искажениями, $\epsilon = 25$	98,82	98,93	21	16488
СКО без искажений	99,16	98,74	18	22961

Таблица 2. Результаты дополнительного обучения сети

Функция ошибки	Новая точность распознавания обучающей выборки (60000 цифр), %	Новая точность распознавания тестовой выборки (10000 цифр), %	Количество дополнительных эпох обучения	Число пропущенных образцов на последней эпохе
КЭ, отключение искажений после 21-й эпохи	99,72	99,23	8	55646
СКО, отключение искажений после 21-й эпохи	99,3	99,02	8	32803



Рис. 4. Изображения неверно распознанных цифр. Число под изображением цифры – порядковый номер в базе MNIST. Слева – верный ответ => справа – ответ сети

Из рис. 4 нетрудно заметить, что некоторые цифры вызывают сомнения даже у человека, например, образцы с номерами 3423 или 3521. Также есть

цифры, изображения которых некачественны, например, 5458, 6572 или 6577. Поэтому работу системы можно считать высокоэффективной.

Графики изменения значений функции ошибки от эпохи к эпохе представлены на рис. 5.

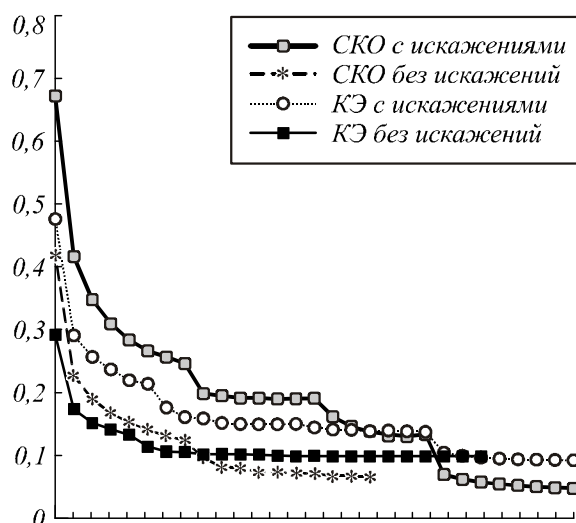


Рис. 5. Изменение функции ошибки от эпохи к эпохе

Из рис. 5 видно, что на первых эпохах значение функции ошибки быстро падает. Однако на последующих эпохах оно меняется незначительно. Также видна точка отключения механизма внесения искажений – 21 эпоха – в первом и третьем экспериментах, когда значение функции ошибки резко уменьшилось.

Таким образом, лучшие результаты распознавания были получены с функцией ошибки КЭ с применением эластичных искажений и с дополнительным обучением сети без искажений. Стоит также отметить, что обучение сети с функцией ошибки КЭ происходит быстрее, так как количество пропускаемых образцов при обучении составляет более 50%.

Для сравнения эффективности работы реализованной сети в таблице 3 приведены результаты, показанные другими классификаторами на тестовом множестве базы MNIST [2].

В таблице 3 представлены результаты линейных классификаторов, нейронных сетей с одним и двумя скрытыми слоями, машин опорных векторов, а также сверточных нейронных сетей семейства LeNet. Из таблицы видно, что минимальная частота ошибок при обработке проверочного набора составляет 0,7% для сети Boosted LeNet-4. Стоит заметить, что в этой усиленной нейронной сети комбинировались три копии архитектуры LeNet, при этом вторая обучалась на смеси образцов, которые первая распознавала с ошибками 50%, а третья обучалась на образцах, которые не были распознаны первыми двумя копиями. При тестировании складывались веса цифр, полученные во всех трёх сетях, и результат распознавания зависел от суммарной оценки [2].

Таким образом, реализованная в рамках данной работы нейронная сеть при более простой структуре сверточной сети (отсутствуют слои подвыборки) показывает результаты распознавания, сопоставимые с сетями LeNet.

Таблица 3. Процент неверно распознанных символов на тестовой выборке других систем классификации

Классификатор	Количество ошибок на тестовой выборке (%)
Linear	12
[deslant] Linear	8,4
Pairwise	7,6
K-NN Euclidean	5
[deslant] K-NN Euclidean	2,4
40 PCA + quadratic	3,3
1000 RBF + linear	3,6
[16x16] Tangent Distance	1,1
SVM poly 4	1,1
RS-SVM poly 5	1
[dist] V-SVM poly 9	0,8
28x28-300-10	4,7
[dist] 28x28-300-10	3,6
[deslant] 20x20-300-10	1,6
28x28-1000-10	4,5
[dist] 28x28-1000-10	3,8
28x28-300-100-10	3,05
[dist] 28x28-300-100-10	2,5
28x28-500-150-10	2,95
[dist] 28x28-500-150-10	2,45
[16x16] LeNet-1	1,7
LeNet-4	1,1
LeNet-4/Local	1,1
LeNet-4/K-NN	1,1
LeNet-5	0,95
[dist] LeNet-5	0,8
[dist] Boosted LeNet-4	0,7

Заключение

Для исследования возможностей сверточных нейронных сетей в распознавании рукописных цифр была разработана автоматизированная система, реализующая модифицированную архитектуру четырехслойной сети без слоёв подвыборки. Предложена методика обучения сети с использованием чередования эпох обучения с искажением обучающих символов и без искажения. Предложена методика подбора и модификации коэффициента обучения. Несмотря на упрощение структуры сети, эксперименты показали высокую эффективность работы системы при распознавании рукописных цифр. Лучшие результаты распознавания были получены с функцией ошибки КЭ с применением эластичных искажений и с дополнительным обучением сети без искажений. Все тесты велись с использованием базы рукописных цифр MNIST. Результаты распознавания сопоставимы с лучшими результатами, показанными при использовании других методов классификации на этой же базе данных.

Литература

1. **Козин, Н.Е.** Поэтапное обучение радиальных нейронных сетей / Н.Е. Козин, В.А. Фурсов // Компьютерная оптика. – 2004. – № 26. – С. 138-141.
2. **LeCun, Y.** Gradient Based Learning Applied to Document Recognition / Y. LeCun, L. Bottou, P. Haffner – IEEE Press, 1998. – P.46.
3. **Simard, P.Y.** Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis / P.Y. Simard, D. Steinkraus, J. Platt // International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society. – Los Alamitos. – 2003. – P. 958-962.
4. **Хайкин, С.** Нейронные сети: полный курс / С. Хайкин. – М.: Вильямс, 2006. – 1104 с.
5. **LeCun, Y.** Scaling learning algorithms towards AI / Y. LeCun, Y. Bengio – MIT Press, 2007.
6. **Гаршин, А.А., Солдатова, О.П.** Автоматизированная система распознавания рукописных цифр на основе сверточной нейронной сети // Свидетельство об официальной регистрации программ для ЭВМ №2010610988 по заявке №2009616812 от 1 декабря 2009 года. Зарегистрировано в Реестре программ для ЭВМ 1 февраля 2010 года.
7. **LeCun, Y.** The MNIST database of handwritten digits – <http://yann.lecun.com/exdb/mnist>.
8. **LeCun, Y.** Efficient BackProp in Neural Networks: Tricks of the trade / Y. LeCun, L. Bottou, G. Orr, K. Muller – Springer, 1998. – 44 p.
9. **Bishop, C.M.** Neural Networks for Pattern Recognition – Oxford University Press, 1995. – 498 p.
10. Gaussian blur – http://en.wikipedia.org/wiki/Gaussian_blur.
11. **Duffner, S.** An Online Backpropagation Algorithm with Validation Error-Based Adaptive Learning Rate / S. Duffner, C. Garcia // ICANN 2007, Part I, LNCS 4668, 2007. – P. 249-258.

References

1. **Kozin, N.E.** Gradual learning the radial neural networks / N.E. Kozin, V.A. Fursov // Computer Optics. – 2004. – № 26. – P. 138-141. – (in Russian).
2. **LeCun, Y.** Gradient Based Learning Applied to Document Recognition / Y. LeCun, L. Bottou, P. Haffner – IEEE Press, 1998. – P. 46.
3. **Simard, P.Y.** Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis / P.Y. Simard, D. Steinkraus, J. Platt // International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society. – Los Alamitos. – 2003. – P. 958-962.
4. **Haykin, S.** Neural Networks – a comprehensive foundation / S. Haykin. – Moscow: Williams, 2006. – 1104 p. – (in Russian).
5. **LeCun, Y.** Scaling learning algorithms towards AI / Y. LeCun, Y. Bengio – MIT Press, 2007.
6. **Garshin, A.A., Soldatova, O.P.** An automated system of recognizing handwritten digits based on convolution neural networks // Certificate of a formal registration of computer software № 2010610988, the application № 2009616812 on December 1, 2009. Registered in the Register of Computer Programs February 1, 2010.
7. **LeCun, Y.** The MNIST database of handwritten digits – <http://yann.lecun.com/exdb/mnist>.
8. **LeCun, Y.** Efficient BackProp in Neural Networks: Tricks of the trade / Y. LeCun, L. Bottou, G. Orr, K. Muller – Springer, 1998. – P. 44.
9. **Bishop, C.M.** Neural Networks for Pattern Recognition – Oxford University Press, 1995. – P. 498.
10. Gaussian blur – http://en.wikipedia.org/wiki/Gaussian_blur.
11. **Duffner, S.** An Online Backpropagation Algorithm with Validation Error-Based Adaptive Learning Rate / S. Duffner, C. Garcia // ICANN 2007, Part I, LNCS 4668, 2007. – P. 249-258.

CONVOLUTIONAL NEURAL NETWORK APPLIED TO HANDWRITTEN DIGITS RECOGNITION

*O.P. Soldatova, A.A. Garshin
S.P. Korolyov Samara State Aerospace University*

Abstract

The capabilities of convolutional neural networks for recognizing handwritten digits. The technique of training networks, which implements the alternating periods of training with and without distortion of characters. The technique of selection and modification of learning rate. The neural network is tested by using the standard database of handwritten digits (MNIST), the results of experimental investigation are presented.

Key words: recognizing handwritten digits, convolutional neural network, the MNIST database, elastic distortions, generalization ability of neural network.

Сведения об авторах

Солдатова Ольга Петровна. 1979 год - окончила факультет автоматики и вычислительной техники Московского энергетического института, 1990 год - защитила кандидатскую диссертацию в Московском энергетическом институте. Кандидат технических наук, доцент кафедры информационных систем и технологий Самарского государственного аэрокосмического университета имени академика С.П.Королёва (СГАУ). **E-mail:** *op-soldatova@yandex.ru*. В настоящее время основные сферы **научных интересов:** применение нейронных и нейронечётких сетей для решения задач распознавания и прогнозирования; семантический анализ текстов; проектирование нейроимитаторов.

Olga Petrovna Soldatova 1979 - graduated from the Faculty of Automation and Computer Engineering, Moscow Power Engineering Institute, 1990 - received a doctorate from the Moscow Power Engineering Institute. Candidate of Technical Science, Associate Professor of Information Systems and Technologies, Samara State Aerospace University named after academician S.P. Korolyov (SSAU). **E-mail:** : *op-soldatova@yandex.ru* . **Currently the main areas of research:** applying the neural and neuro-fuzzy networks for solving recognition and forecasting tasks; semantic analysis of texts; designing neurosimulators.



Гаршин Александр Александрович. Студент факультета информатики Самарского государственного аэрокосмического университета имени академика С.П.Королёва (СГАУ). **E-mail:** *garshin.aleks@rambler.ru* .

Aleksandr Aleksandrovich Garshin Samara State Aerospace University named after academician S.P. Korolyov, Department of Information Systems and Technologies, 443086, Russia, Samara, Moskovskoye road, 34, building 14. **E-mail:** *garshin.aleks@rambler.ru*. **Educational background:** computer science student of Samara State Aerospace University named after academician S.P. Korolyov, Department of Information Systems and Technologies (SSAU).

Поступила в редакцию 5 февраля 2010 г.