УДК 004.932.75'1

СВЕРТОЧНАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ В ЗАДАЧЕ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ ИЗОЛИРОВАННЫХ ЦИФР

Н.Н. КУЗЬМИЦКИЙ

Брестский государственный технический университет
Московская, 267, Брест, 224017, Беларусь

Поступила в редакцию 25 апреля 2012

Выполнен анализ сверточной нейросетевой модели. Разработано программное обеспечение, позволяющее обучать и тестировать сверточные нейронные сети базовой архитектуры LeNet-5. Показана эффективность методики дообучения и искажения тренировочных образов. Построен классификатор изображений изолированных цифр. Произведена оценка устойчивости его характеристик на примерах известных рукописных и шрифтовых баз данных.

Ключевые слова: машинное обучение, сверточная нейронная сеть, алгоритм обратного распространения, метод Левенберга-Марквардта, искажение, дообучение, MNIST.

Введение

Классификаторы, создаваемые на основе методов машинного обучения и аннотированной информации баз данных, широко применяются в области автоматического распознавания символов (Optical Character Recognition, OCR) с начала 1990-х. В первом из прошедших с тех пор десятилетий наблюдался уверенный прогресс в увеличении их эффективности, что позволило выполнить успешные внедрения во многих практических приложениях: обработке банковских счетов, деловой переписке, почтовой сортировке и др. Результаты вселяли уверенность в полном разрешении OCR-проблематики в скором времени, однако активное расширение круга задач данной области, начиная с 2000-х (создание цифровых библиотек, анализ исторических документов, естественных изображений, содержащих текстовую информацию и т.д.) выявило значительные ограничения эффективности классификаторов, в частности, при обработке искаженных и зашумленных изображений символов, разнообразной шрифтовой и рукописной природы.

Возрастание интереса исследователей привело к появлению новых моделей (например, машин опорных векторов), применяемых для решения OCR-задач и совершенствованию традиционных, в частности, искусственных нейронных сетей [1]. Именно в рамках последних была разработана сверточная архитектура, которая, по мнению ряда авторов, наилучшим образом подходит для решения визуальных задач анализа документов [2]. Целью представленного в статье исследования являлось изучение возможности применения данной архитектуры для создания универсального классификатора изображений изолированных цифр и его апробация на примерах известных баз данных. Научный интерес заключался в создании на основе достаточно распространённого нейросетевого подхода эффективного классификатора изображений указанного типа, не используя при этом «тяжеловесные» методы (например, попиксельное сравнение), зачастую применяемые в коммерческих OCR-системах. Практический интерес определялся перспективой применения подхода для классификации изображений символов полного алфавита, имеющих произвольное аппаратно-программное происхождение, и его внедрение в разрабатываемую систему анализа цифровых изображений документов.

Архитектура нейросетевой модели и ее обучение

Сверточные нейронные сети (Convolutional Neural Networks, CNN) являются представителями класса моделей, поводом к созданию которых послужили исследования зрительного аппарата кошек, проведенные Хубелем и Вейселем в 1960-х [3]. Их результатом было открытие двух типов клеток, влияющих на зрительную восприимчивость: первые обладают свойством локальной чувствительности и предназначены для выделения элементарных характеристик образов (ориентированных краев, конечных точек, углов и др.), вторые путем их комбинирования осуществляют построение высокоуровневых признаков.

Первой нейросетевой моделью, реализующей обнаруженное поведение, был неокогнитрон Фукушимы [4], при этом применялась неконтролируемая настройка банка фильтров и контролируемое обучение линейного классификатора. Дальнейшие исследования позволили упростить структуру нейронной сети и привести ее обучение к полностью контролируемому. В результате была создана новая сверточная архитектура, сфера практического применения которой в настоящее время постоянно расширяется: системы OCR, идентификации лиц, навигации, восстановления сигналов, робототехника и т.д. [5].

Известны различные реализации сверточных нейронных сетей, отличающиеся топологией слоев, способом организации процесса обучения и др. Исходя из результатов их применения в решении задач классификации, аналогичных рассматриваемой, и возможностей обучения сети без использования специализированного аппаратного обеспечения, в качестве базовой для проведения описываемой исследовательской работы была выбрана нейросетевая модель LeNet-5, созданная Яном Лекуном в конце 1990-х [6]. В ее основе лежат три архитектурные идеи:

1) *локальные рецептивные поля* (нейроны получают входной сигнал от окрестностей нейронов предыдущего слоя, за счет чего сеть обучается двумерной структуре входного образа);

2) *разделяемые веса* (нейроны слоя объединены картами, в которых они обладают общими весами, при этом карты генерируют различные признаки и сокращают количество параметров, настраиваемых в ходе обучения);

3) *пространственные подвыборки* (локальное усреднение откликов карт приводит к синтезу высокоуровневых признаков и повышает инвариантность сети к искажениям).

Как видно из рис. 1, входным сигналом для сверточной нейронной сети является изолированное изображение символа размером 32×32 пикселя, которое отображается через 6 скрытых и 1 выходной слой. Первый из них C1 относится к сверточному типу, содержит 6 карт признаков размером 28×28 и связывает свои нейроны с окрестностями размером 5×5 входного изображения. Следующий слой S2 является подвыборочным и осуществляет усреднение откликов нейронов предыдущего по неперекрывающимся окрестностям размером 2×2, поэтому имеет 6 карт признаков, размером 14×14.

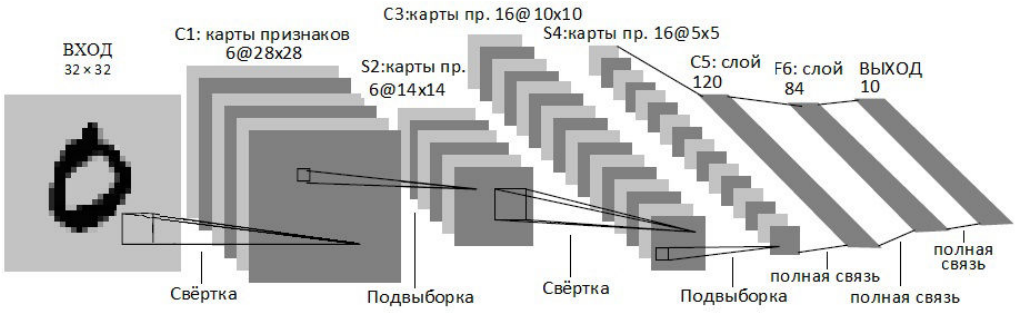

Рис. 1. Архитектура сверточной нейронной сети *LeNet*-5 [6]

В соответствии с чередованием типов слоев, третий *C*3 является сверточным, содержит 16 карт размером 10×10. Его особенностью является связывание нейронов с различным числом окрестностей размером 5×5, занимающих идентичные положения в картах признаков предыдущего слоя, что сокращает количество настраиваемых параметров, улучшая обобщаемость сети за счет повышения несимметричности распространения сигналов в ней. Следующий слой *S*4, являясь подвыборочным, осуществляет аналогичную с *S*2 работу по уменьшению разреше-

ния признаков, содержит 16 карт размером 5×5. Слой *C*5 относится к свёрточному типу, при этом обладает полносвязным соединением, т.к. отображает свои 120 нейронов на окрестности размером 5×5, совпадающие с целыми картами *S*4. Слой F6 также полносвязный, отклики его 84 нейронов соотносятся с моделями классов, загружаемыми в виде весовых коэффициентов 10 нейронов последнего слоя. В качестве моделей классов могут применяться изображения их идеальных образов размером 7×12. В этом случае выходы сети представляют собой евклидовые расстояния между откликами *F*6 и идеальными образами, которые после отображения функцией Гаусса можно интерпретировать как вероятности принадлежности сигналов классам [6].

Отметим, что в представленной ниже экспериментальной работе была реализована сокращенная, по сравнению с описанной, свёрточная архитектура. В ней не применялись модели классов, отсутствовал слой *F*6, а выход сети рассчитывался в обычной нейросетевой манере, при этом в качестве функции активации был выбран гиперболический тангенс. В результате общее число связей сети составило 331114, в то время как количество настраиваемых параметров – 51046. Причиной сокращения архитектуры является целесообразность применения моделей классов главным образом для распознавания изображений полного алфавита, в котором имеются подмножества с низкой межгрупповой изменчивостью: *I* – 1, *k* – *K*, 0 – *O* и др.

Обучение сверточной нейронной сети осуществляется модификацией алгоритма обратного распространения ошибки стохастическим диагональным методом Левенберга-Марквардта [7]. Использование модификации обеспечивает ускорение сходимости обучения, путём индивидуальной настройки для весов каждого нейрона глобального параметра η, позволяющей замедлять процесс обучения на крутых областях весового пространства и ускорять на плоских: $w^k{}_{new} = w^k{}_{old} - \eta_k \cdot \partial E / \partial w^k$; $\eta_k = \eta/(h_{kk} + \mu)$, где w^k – *k*-ый настраиваемый параметр сети; μ – константа; *E* – функция ошибки, которая может быть вычислена как среднеквадратичное отклонение желаемого отклика *t* от фактического *y*; h_{kk} – оценка второй производной *E* по w^k; V^k – связи между *i*-ым и *j*-ым нейронами, разделяющие w^k: $E = 0,5 \cdot \sum_i^{10} (y^i - t^i)^2$; $h_{kk} = \sum_{(i,j)\in V^k} \sum_{(k,l)\in V^k} \partial^2 E / (\partial u_{ij} \partial u_{kl})$. Расчёт приближённого h_{kk} осуществляется по следующим формулам:

$$h_{kk} = \sum_{(i,j)\in V^k} \partial^2 E / \partial u_{ij}^2 ;\ \partial^2 E / \partial u_{ij}^2 = 1/P \cdot \sum_{n=1}^{P} \partial^2 E^n / \partial u_{ij}^2 ;$$

$$\partial^2 E^n / \partial u_{ij}^2 = \partial^2 E^n / \partial a_i^2 \cdot x_j^2 ;\ \partial^2 E^n / \partial a_i^2 = f^{'}(a_i)^2 \cdot \sum_k u_{ki}{}^2 \cdot \partial^2 E^n / \partial a_k^2 ,$$

где x_j – отклик *j*-го нейрона, a_i – сумма взвешенных входов *i*-го нейрона, *P* – число образов.

Используемая аппроксимация Гаусса-Ньютона гарантирует неотрицательность гессиана, при этом рассчитываются только диагональные элементы, а остальные отбрасываются. Организация вычислений по данным формулам аналогична этапу обратного распространения первой производной функции ошибки. Пересчёт гессиана производится перед каждой эпохой обучения и лишь на части тренировочного множества (в [6] для 60000 образов предлагалось использовать *P* = 500). Последнее обстоятельство объясняется зависимостью характеристик функции ошибки в большей степени от архитектуры сети, чем от статистических свойств классов.

Описание базы маркированных образов

Высокая практическая востребованность эффективных *OCR* систем является стимулом к постоянному совершенствованию методов машинного обучения, применяющихся для разработки классификаторов цифровых изображений символов. При этом улучшается не только соответствующий математический аппарат, но и не менее важный фактор для реализации эффективного процесса обучения – качество и количество маркированных образов. К настоящему

моменту научным сообществом был накоплен большой набор баз данных, отличающихся целью использования (тренировка и тестирование классификаторов, определение их уровня обобщаемости), содержанием (языковая, алфавитная принадлежность), типом информации (машинные шрифты, рукописные, созданные программным образом), источником (сканеры, планшетные компьютеры, цифровые фотоаппараты), формами (графические, закодированные файлы), доступностью (коммерческие и свободного обращения), объёмом и др. Руководствуясь данными характеристиками и поставленными целями исследования, в качестве основной для проведения экспериментальной работы была выбрана база рукописных цифр MNIST [8].

MNIST (от англ. Modified NIST) является подмножеством более объемной базы NIST [9], содержащей рукописные образы, сегментированные из изображений специально подготовленных шаблонов, заполненных респондентами бюро переписи и студентами учреждений образования США. MNIST состоит из тренировочной (60000 образов) и тестовой (10000) частей, причём для повышения уникальности в разные части были помещены образы, полученные от различных авторов. Оригинальные бинарные данные были смасштабированы так, чтобы символ вписывался в прямоугольник размером 20×20, который в дальнейшем был помещён в итоговое изображение размером 32×32 пикселя, при этом должны совпадать центр тяжести прямоугольника и геометрический центр изображения.

MNIST является достаточно представительной базой данных, однако ряд исследователей, в частности Патрик Симард в [2], предлагают для повышения уровня обобщаемости нейронной сети использовать расширенное искажёнными образами тренировочное множество. Искажения могут включать в себя как известные аффинные преобразования, так и «эластичные» (рис. 2). В последнем случае создаются матрицы Δx и Δy размером равным входному образу, которые заполняются случайными величинами, равномерно распределенными на отрезке [–1; 1]. Далее они сглаживаются функцией Гаусса со стандартным отклонением σ пропорциональным размеру ядра функции. Итоговые матрицы задают смещения точек в соответствующих позициях изображения, причем интенсивность смещения регулируется множителем α, а искажённый образ строится с помощью билинейной интерполяции. Отметим, что наиболее эффективными в экспериментах [2] были следующие значения параметров эластичных искажений: σ = 4, α = 34 для матрицы 29×29.

Рис. 2. Примеры изображений (первое слева – исходное), подверженных эластичным искажениям с параметрами: $\sigma \in [4, 8]$, $\alpha \in [5, 50]$

Экспериментальная работа и анализ результатов

Целью проведенной экспериментальной работы была качественная и количественная оценка характеристик сверточной нейронной сети, созданной на основе базовой архитектуры *LeNet-5*, обучение и тестирование которой осуществлялось с использованием собственного программного обеспечения. При этом основное внимание уделялось определению: обучаемости (точности распознавания образов тренировочного множества), обобщаемости (точности распознавания образов тестового множества) и эффективности функционирования (времени обучения и тестирование) сети. Для проведения экспериментов использовалось оборудование следующей аппаратно-программной конфигурации: процессор – Intel Core 2 Duo i3-530 2,93 GHz, ОЗУ – DDR3 2048 Mb, ОС – Windows 7 Ultimate, платформа – .NET.

Подготовка запуска обучения нейронной сети включала:

1) начальную инициализацию весовых коэффициентов: весам присваивались случайные значения, равномерно распределенные на интервале $[-2{,}4F_i, 2{,}4F_i]$, где F_i – число входных связей *i*-го нейрона;

2) выбор количества эпох и методики изменения коэффициента η: в отличие от п.1, который соответствует рекомендациям [6], в данном случае обучение сети было решено провести за 68 эпох, с начальным значением η = 0,00085, которое изменялось каждые 2 эпохи путем умножения на коэффициент 0,8099, в итоге конечное значение η составило 0,000001;

3) настройку параметров искажений входных образов: величины поворота (угол в пределах ±15°, для изображений цифр '1' и '7' – ±7°), изменения масштаба (в пределах ±15%, для каждой размерности по отдельности), эластичных искажений (σ = 8, α = 36), для ускорения сходимости обучения диапазон оттенков серого [0, 255] был нормирован к [-1, 1].

Отметим, что коррекция весов нейронов проводилась после обработки каждого входного образа, а для предотвращения переобучения сети применялась методика пропуска цикла обратного распространения ошибки в случае если её величина была меньше заданного значения □, что также позволяет увеличить скорость обучения. Кроме того, для повышения уровня обобщаемости сети искажения тренировочных образов было решено выполнять перед каждой эпохой, что в свою очередь потребовало увеличения их количества до 68 ввиду возрастания времени сходимости. Результаты обучения и тестирования сети приведены в табл. 1.

Таблица 1. **Результаты обучения и тестирования сверточной нейронной сети**

Этапы обучения	Точность на тренировочном MNIST	Точность на тестовом MNIST	Число эпох	Время обучения (час)	Частота искажений (по эпохам)
1	99,35%	99,39%	68	6,88	каждую
2	99,89%	99,36%	56	3,13	через 1
3	99,91%	99,36%	37	1,55	–

Анализ данных таблицы показывает, что сверточная нейронная сеть обладает высокими показателями обучаемости и обобщаемости, т.к. после первого этапа обучения точность распознавания на обеих контрольных выборках составила более 99%. При этом точность на тренировочной выборке оказалась меньше, чем на тестовой, что свидетельствует о сильном влиянии на обучение искажения образов, не дающего сети выполнить их простое запоминание и, как следствие, увеличивающие её обобщающие способности. Скорость сходимости являлась вполне удовлетворительной для выполнения непрерывного запуска обучения на обычном ЭВМ: время тестирования сети составило 0,0033 час или 833 образа/сек (искажения при тестировании не применялись). Примеры неверно классифицированных изображений тестового множества приведены на рис. 3, *а*.

Для изучения возможности повышения точности распознавания тренировочного MNIST было выполнено дообучение нейронной сети, при этом искажения применялись каждую вторую эпоху. Техника дообучения, по мнению ряда авторов, в частности С. Осовского [10], является весьма эффективной, т.к. она позволяет выполнить «встряхивание весов» с минимальной вероятностью вывода поиска из сферы притяжения ранее найденного локального минимума, чем при обучении «с чистого листа». Сеть в такой ситуации должна проявить способности к усвоению наиболее характерных признаков и после кратковременной амнезии быстро восстановиться, а затем, в большинстве случаев, улучшить свои показатели.

Данное предположение подтверждается повышением точности распознавания тренировочного MNIST после второго этапа обучения на 0,54%, при этом было отмечено небольшое снижение точности тестового на 0,03%. Последнее наблюдение объясняется с позиций неравенства Вапника-Червоненкиса, в соответствии с которым обобщаемость прямо пропорциональна отношению объема обучающей выборки к мере сложности модели (количеству настраиваемых параметров) [10]. Сокращение вариативности входных данных в результате ограничения использования искажений и оставшееся неизменным количество весовых коэффициентов привели к повышению адаптации сети к тренировочному множеству, несколько понизив её уровень обобщаемости.

Третий этап обучения сети, проведенный при отсутствии искажений, подтвердил ранее полученные выводы. Ожидаемое увеличение точности распознавания тренировочной выборки оказалось незначительным (0,02%), при этом основной интерес представлял показатель обобщаемости, который сохранился на прежнем уровне. Данный факт позволяет утверждать, что был достигнут компромисс между стремлением обучения к выходу из локального минимума и силой его притяжения, т.е. из данного минимума было извлечено максимум возможного. Для оценки эффективности построенной нейронной сети в табл. 2 приведены результаты классификации тестового MNIST, полученные другими типами моделей (подробнее можно ознакомиться в [8]).

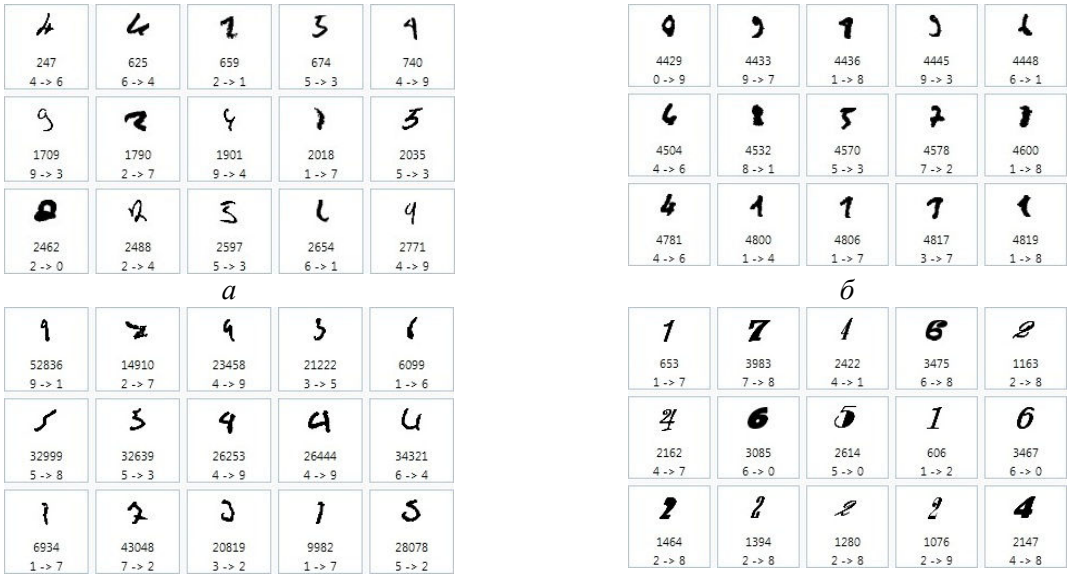

Рис. 3. Неверно классифицированные образы баз:
тестового MNIST (*а*), OPTDIGTS (*б*), NIST (*в*), WIN_FONT (*г*)
(в нижних строках слева – истинный номер класса, справа – предсказанный)

Таблица 2. **Результаты классификации тестового MNIST различными типами моделей**

Классификатор	Ошибка на тестовом MNIST (в %)	Авторы
pairwise linear classifier	7,6	LeCun et al.
PCA + quadratic classifier	3,3	LeCun et al.
boosted stumps	0,87	Kegl et al.
K-nearest-neighbors	0,63	Belongie et al
convolutional net+distortions	0,61	Kuzmitsky
virtual SVM, deg-9 poly	0,56	DeCoste and Scholkopf
convolutional net LeNet-5	0,8	LeCun et al.
committee of 7 CNN	0,27 ± 0,02	Ciresan et al.

Данные таблицы позволяют сделать некоторые замечания:

1) созданная нейронная сеть имеет более высокую (не учитывая другие CNN) точность распознавания, чем традиционные модели классификаторов, уступая 0,05% лишь методу опорных векторов (SVM);

2) полученный ею результат (0,61% ошибок) превосходит таковой у базовой сети LeNet-5 (0,8% ошибок), что доказывает эффективность использованной методики обучения;

3) наибольшей точностью обладает подход, основанный на комитетах CNN (0,27±0,02% ошибок), что доказывает перспективность выбранной модели классификатора.

Заключительным этапом экспериментальной работы являлась оценка эффективности классификации свёрточной нейронной сетью, обученной с использованием MNIST, образов других баз. Первая состояла из 55000 изображений рукописных цифр подмножества HSF_4 базы NIST, вторая, называемая OPTDIGTS, содержала 5620 образов того же типа [11], третья (WIN_FONT) – 5080, представляющих машинные шрифты ОС WINDOWS 7. Примеры баз были приведены к формату MNIST: смасштабированы в прямоугольник 20×20, который помещался на изображение размером 32×32, при совпадении центра тяжести прямоугольника и геометрического центра изображения.

В ходе тестирования сети были получены следующие показатели точности распознавания: HSF_4 – 99,13%, OPTDIGTS – 96,22%, WIN_FONT – 94,15% (примеры неверно классифицированных образов приведены на рис.3, *б*–*г*), позволяющие сделать ряд выводов:

1) сверточная нейронная сеть обладает высокой способностью к обобщению, в особенности изображений символов, имеющих сходное с ее тренировочным множеством происхождение;

2) эффективность сети зависти не только от количества маркированных образов, применяемых в ходе обучения, но и величины их внутриклассовой изменчивости, которая нуждается в дополнительном определении;

3) обобщаемость сети в значительной степени определяется устойчивостью ее архитектуры к уровню вариативности топологии изображений символов, в частности, изменению ширины (55% всех ошибок на базе OPTDIGTS относились к изображениям цифры «1», которые наиболее сильно подвержены данному искажению).

Подводя итоги проведенной экспериментальной работы можно с уверенностью утверждать, что сверточная нейросетевая модель является весьма перспективным, но еще не до конца настроенным механизмом для создания универсального классификатора изображений изолированных цифр. Остается нерешенным ряд вопросов, в частности: создание представительной базы, объединяющей образы основных типов изображений символов, разработка методики их предобработки, настройки оптимальных параметров процесса обучения, уточнение архитектуры сети и др., являющиеся предметом дальнейших исследований.

Заключение

В представленной статье исследовалась сверточная нейросетевая модель, которая была применена для создания классификатора изображений изолированных цифр. При этом использовалась отличная от базовой модели LeNet-5 архитектура и методика модификации параметров обучения. Расширение тренировочного множества искаженными образами и техника дообучения позволили достичь точности распознавания классификатором как тренировочного, так и тестового MNIST, сравнимой с лучшими результатами, полученными на данной точке отсчета. Анализ экспериментальной работы, проведенной с использованием различных тестовых баз показал, что рассмотренная нейросетевая модель обладает высокими показателями эффективности, уровень которых, однако, пока недостаточен для универсальности сферы ее применения.

CONVOLUTIONAL NEURAL MODEL IN A TASK OF CLASSIFICATION IMAGES OF THE ISOLATED DIGITS

N.N. KUZMITSKY

Abstract

The analysis of convolutional neural model is done. The software is developed, allowing to train and test convolutional neural networks of base architecture LeNet-5. Efficiency of technique multi training and distortions of training images is shown. The qualifier of images of the isolated figures is constructed. The estimation of stability of its characteristics on examples of known handwritten and font databases is done.

Список литературы

1. *Головко В.А.* Нейронные сети: обучение, организация и применение. М., 2001.
2. *Simard P.Y., Steinkraus D., Platt J.* // Int. Conf. on Document Analysis and Recognition. 2003. P. 958–963.
3. *Hubel D.H., Wiesel T.N.* // Journal of Physiology London., 1962. Vol. 15. P 106–154.
4. *Fukushima K., Miyake S.* // Pattern Recognition. 1982. Vol 15. P. 455–469.
5. *LeCun Y., Kavukvuoglu K., Farabet C.* // Proc. Int. Symposium on Circuits and Systems. 2010. P 253–256.
6. *LeCun Y., Bottou L., Bengio Y., et. al.* // Proceedings of the IEEE. 1998. P. 2278–2324.
7. *LeCun Y., Bottou L., Or G.B., et. al.* // Springer Lecture Notes in Computer Sciences. 1998. № 1524. P. 5–50.
8. *LeCun Y.* The MNIST database of handwritten digits // http://yann.lecun.com/exdb/mnist.
9. *Grother P.J.* Nist special database 19 – handprinted forms and characters database // National Institute of Standards and Thechnology (NIST), Tech. Rep. 1995.
10. *Осовский С.* Нейронные сети для обработки информации. М., 2002.
11. Optdigits database // http://archive.ics.uci.edu/ml/machine-learning-databases/optdigits.