

Простая форма פָּנִים называется «лицо»), четвертная сочетанная форма — אַחֲרֵי («заднее»). Итак, одно слово может заключать в себе многообразные численные значения.

Заключение. Разработанный метод может быть использован на различного рода описания свойств слова в виде числового значения. Недостаток приложения — при использовании данного метода имеет место одинаковая структура числового значения для разных слов.

УДК 004.891.2

А. А. Киселёв

Учреждение образования «Гродненский государственный университет имени Янки Купалы», Гродно

РАСПОЗНАВАНИЕ РУКОПИСНЫХ ЦИФР НА ОСНОВЕ ТЕХНОЛОГИИ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Введение. Сверточные нейронные сети (СНС) являются подкатегорией нейронных сетей, использование которых достаточно эффективно в таких областях, как распознавание изображений и задача классификации [1, с. 45], также успешно применяются для распознавания лиц, дорожных объектов и знаков, для имитации зрения у роботов и самоуправляемых автомобилей [2, с. 1].

Подход использования СНС был успешно применен на ежегодной олимпиаде по машинному зрению, тем самым снизив рекорд ошибок классификации с 26 до 15%, что на тот момент стало прорывом [3].

Основным недостатком СНС является наличие большого числа варьируемых параметров сети, к которым можно отнести: количество слоев, размерность ядра свертки для каждого из слоев, количество ядер для каждого из слоев, шаг сдвига ядра при обработке слоя, необходимость слоев субдискретизации, степень уменьшения ими размерности, функция по уменьшению размерности, передаточная функция нейронов, наличие и параметры выходной полносвязной нейросети на выходе сверточной. Все эти параметры существенно влияют на результат распознавания, но при построении СНС выбираются исследователями эмпирически [4, с. 393]. Таким образом, количество архитектур СНС велико, все они строятся для решения определенной задачи.

Основная часть. Для построения СНС была выбрана библиотека с открытым исходным кодом ConvNetSharp для .NET C# [5]. Обучение и тестирование сети проходило на базе данных MNIST, которая содержит 60 тыс. изображений рукописных цифр для обучения нейронной сети и 10 тысяч изображений для тестирования нейронной сети [6].

В ходе проведения экспериментов была построена СНС, способная распознавать рукописные цифры с точностью 97%. Представим архитектуру СНС для распознавания рукописных цифр (рисунок 1).

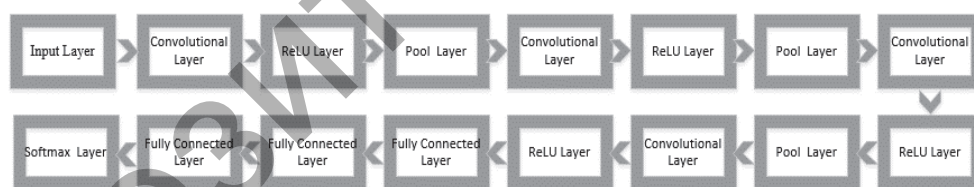


Рисунок 1 — Архитектура СНС для распознавания рукописных цифр

Сверточная нейронная сеть представлена 16 слоями. Входной слой имеет размер $28 \times 28 \times 1$, поскольку ожидает на вход изображение размером 28×28 пикселей с глубиной цвета, равной единице, т. е. черно-белое изображение. За ним следует непосредственно слой свертки с размерностью ядра свертки 4×4 с выходом на 8 каналов.

Затем следует слой ReLU, который представляет собой функцию активации после сверточного слоя, в качестве функции активации, как правило, выбирается ненасыщаемая функция $f(x) = \max(0, 1)$. Такая функция показывает хорошие результаты при обучении нейронных сетей и отвечает за отсеечение ненужных деталей в канале. После ReLU слоя следует слой пулинга размерностью 2×2 . Слой пулинга представляет собой нелинейное уплотнение карты признаков, при этом группа пикселей уплотняется до одного пикселя, проходя нелинейное преобразование. Преобразования затрагивают непересекающиеся квадраты, каждый из которых «сжимается» в один пиксель, при этом выбирается пиксель, имеющий максимальное значение. Такая операция позволяет уменьшить размерность изображения. Далее следует слой свертки с размерностью ядра свертки 3×3 с выходом на 16 каналов. За ним следует слой ReLU, далее следует слой пулинга размерностью 2×2 . После

слоя пулинга следует слой свертки с размерностью ядра свертки 2×2 с выходом на 32 канала, после которого следует слой ReLU. Затем следует слой пулинга размерностью 2×2 , слой свертки с размерностью ядра 2×2 с выходом на 64 канала, за которым следует слой ReLU.

После прохождения свертки изображения и уплотнения с помощью пулинга система перестраивается от конкретной сетки пикселей с высоким разрешением к более абстрактным картам признаков. Как правило, на каждом следующем слое увеличивается число каналов и уменьшается размерность изображения в каждом канале. В итоге остается большой набор каналов, хранящих небольшое количество данных, которые интерпретируются как самые абстрактные понятия, выявленные из исходного изображения. В соответствии с технологией СМС эти данные объединяются и передаются на обычную полносвязную нейронную сеть, которая состоит из четырех слоев. Первый полносвязный слой представлен 512 нейронами, второй полносвязный слой представлен 256 нейронами. Третий полносвязный слой имеет 10 нейронов. Последний слой представлен софтмакс-слоем, в основе которого лежит функция активации софтмакс. Выходом данного слоя является вектор коэффициентов размерностью 10, где каждым выходом данного вектора является некое значение, меньшее единицы, которое обозначает степень принадлежности объекта к определенному классу.

В качестве метода обучения сети использовался алгоритм стохастического градиентного спуска с усовершенствованием Adam (Adaptive Moment Estimation). Для обучения нейронной сети было выполнено более двух с половиной миллионов итераций.

На графике (рисунок 2) видно, что для уменьшения значения функции потерь до приемлемой величины меньше единицы понадобилось около 10 тыс. итераций. Минимальным значением функции потерь является значение 0,00722, которое на каждой итерации может варьироваться как в большую сторону, так и в меньшую, что свидетельствует о минимальном отличии между реальным значением и результатом работы сети.

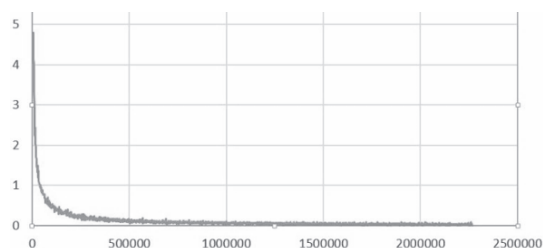


Рисунок 2 — График зависимости значения функции потерь от количества итераций

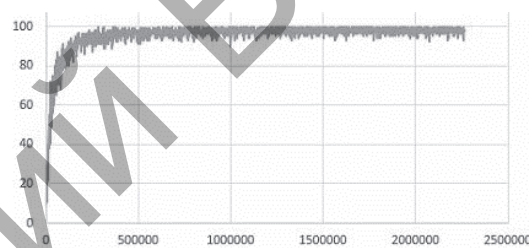


Рисунок 3 — График зависимости точности распознавания от количества итераций

На графике (рисунок 3) мы можем наблюдать, что с ростом количества итераций повышается точность распознавания. Для приемлемой точности распознавания в 90% нам понадобилось около 200 тыс. итераций. Таким образом, чем выше точность распознавания, тем большее количество итераций необходимо сделать для обучения нейронной сети.

Заключение. Для распознавания рукописных цифр было предложено использовать СМС. В статье было продемонстрировано, что СНС, состоящая из 16 слоев, способна распознавать рукописные цифры с точностью 97%. Для достижения такой точности распознавания понадобилось более миллиона итераций для обучения СМС.

Процесс обучения СМС методом стохастического градиентного спуска с усовершенствованием Adam показывает хороший результат обучения за меньшее количество итераций. Для достижения значения функции потерь до величины меньше единицы понадобилось около 10 тыс. итераций.

Список цитируемых источников

1. Deep Learning: A Practitioner's Approach : Adam Gibson, Josh Patterson, 2017. — 200 p.
2. Ranzato Marc'Aurelio, Christopher Poultney, Sumit Chopra and Yann LeCun: Efficient Learning of Sparse Representations with an Energy-Based Model, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems (NIPS 2006). — MIT Press, 2006.
3. A Beginner's Guide To Understanding Convolutional Neural Networks [Electronic resource]. — Access mode: <https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks>. — Date of access: 08.10.2017.
4. Scala for Machine Learning : Patrick R. Nicolas, 2015. — 590 p.
5. ConvNetSharp [Electronic resource]. — Access mode: <https://github.com/cbovar/ConvNetSharp>. — Date of access: 09.10.2017.
6. THE MNIST DATABASE of handwritten digits [Electronic resource]. — Access mode: <http://yann.lecun.com/exdb/mnist>. — Date of access: 09.10.2017.