

## ПРОСТАЯ СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ТОЧНОГО РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР

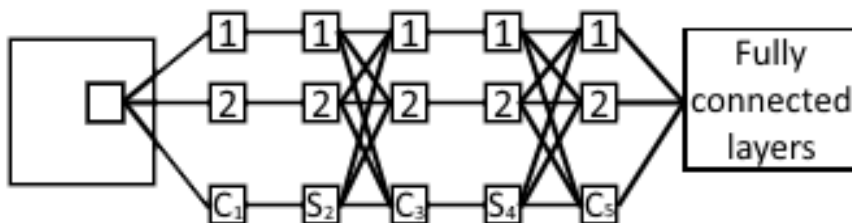
### 1. ВВЕДЕНИЕ

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, в особенности с глубоким обучением и глубокими нейронными сетями [2, 5-13]. 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 сетями глубокого доверия.

В статье рассматривается сверточная нейронная сеть для распознавания рукописных цифр. Мы предлагаем упрощенную архитектуру сверточных нейронных сетей, которая позволит классифицировать рукописные цифры с точностью, большей, чем обычные сверточные сети архитектуры LeNet5. Мы покажем, что с помощью простой нейронной сети могут быть получены лучшие результаты классификации.

### 2. СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ

Сверточная нейронная сеть является дальнейшим развитием многослойного персептрона и неокогнитрона и широко используется для обработки изображений [14, 15]. Нейронные сети такого рода инвариантны к сдвигам и искажениям входного сигнала. Сверточная нейронная сеть объединяет три подхода, а именно области локального восприятия, общие веса и субдискретизирующее пространство [16, 17]. Использование локального восприятия позволяет извлечь примитивные изображений. Общая структура сверточной сети показана на Рис. 1.



**Рисунок 1 – Основная структура сверточной нейронной сети**

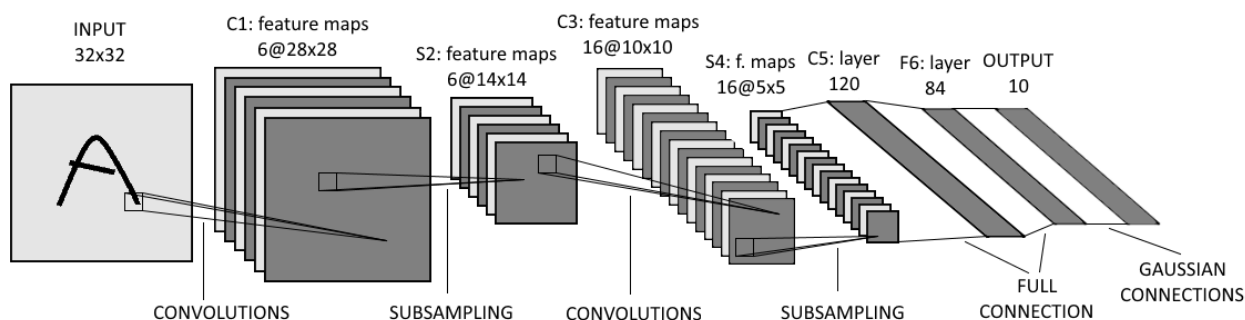
Сверточный слой состоит из множества карт признаков, где нейроны каждой карты содержат одни и те же наборы весов и порогов. В результате, для каждой карты признаков, нейроны выполняют одни и те же операции в различных частях изображения. Методика скользящего окна используется для обхода изображения. Поэтому, если размер окна  $r \times r$  (область восприятия), то каждый нейрон сверточного слоя связан с  $r^2$  элементами соответствующей

области восприятия. Каждой области восприятия во входном пространстве сопоставляется специальный нейрон в каждой карте признаков.

Как уже было сказано, нейроны каждой выходной карты содержат некоторый набор весов и порогов. В результате, из одного и того же изображения можно извлечь множество различных признаков. В последствии, эти признаки объединяются следующим слоем, слоем пулинга, для уменьшения размерности карт. Данный слой называется субдискретизирующим и выполняет локальное усреднение или максимизацию различных областей изображения. Для этого в каждой карте признаков сверточного слоя выбираются неперекрывающиеся области размером  $k \times k$ . Данные области отображаются в один нейрон соответствующей карты субдискретизирующего слоя. Следует отметить, что каждая карта сверточного слоя связана лишь с соответствующей картой в пулинговом слое. Каждый нейрон слоя пулинга является результатом вычисления среднего или максимального значения  $k^2$  нейронов в сверточном слое. Число карт признаков в слое пулинга будет таким же, как и в сверточном слое.

Таким образом, сверточная нейронная сеть представляет собой сочетание сверточных и субдискретизирующих слоев, которые выполняют нелинейные иерархические преобразования пространства входных образов. Последний блок сверточной нейронной сети является многослойным перцептроном, SVM или другим классификатором.

Рассмотрим традиционную сверточную нейронную сеть (LeNet5) для классификации рукописных цифр (Рис. 2) [16, 17]. Входное изображение имеет размер  $32 \times 32$  пикселя. Скользящее окно размерности  $5 \times 5$  обрабатывает изображение и выделяет сегменты изображений, которые входят в слой  $C_1$  нейронной сети. Слой  $C_1$  это сверточный слой с 6-тью картами признаков, где каждая карта содержит  $28 \times 28$  нейронов. Слой  $S_2$  это субдискретизирующий слой с 6ю картами признаков размерности  $14 \times 14$  и ядром  $2 \times 2$ . Слой  $C_3$  является сверточным слоем с 16-тью картами признаков и ядром  $5 \times 5$  для каждой карты. Размерность каждой карты признаков  $10 \times 10$ . Связи между слоями  $S_2$  и  $C_3$  не полносвязные [17]. Слой  $S_4$  является субдискретизирующим слоем с 16-тью картами признаков и ядром  $2 \times 2$  для каждой карты признаков. Размерность карт признаков данного слоя  $5 \times 5$ . Каждое поле восприятия размером  $5 \times 5$  соотносится со 120 нейронами следующего слоя  $C_5$ . Поэтому слой  $C_5$  представляет собой сверточный слой со 120 нейронами. Следующий слой  $F_6$  и выходной слой являются полносвязными.

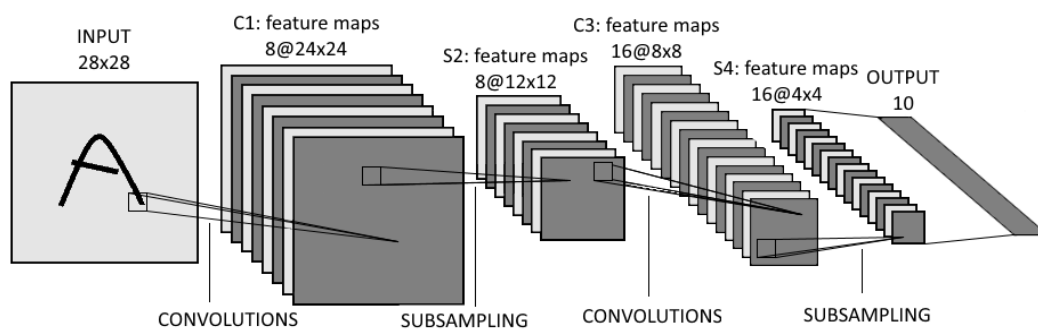


**Рисунок 2 – Архитектура LeNet5**

### 3. УПРОЩЕННАЯ СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ

В данной секции мы предложим сверточную нейронную сеть с более простой архитектурой в сравнении с LeNet5. Упрощенная сверточная нейронная сеть для распознавания рукописных цифр отображена на Рис. 3. Данная сеть

состоит из сверточного слоя ( $C_1$ ), субдискретизирующего слоя ( $S_1$ ), сверточного слоя ( $C_3$ ), субдискретизирующего слоя ( $S_4$ ) и сверточного слоя ( $C_5$ ). Сверточный слой  $C_1$  имеет 8 карт признаков размерности  $24 \times 24$  нейрона и ядро обхода  $5 \times 5$ . Субдискретизирующий слой  $S_2$  содержит 8 карт признаков размерности  $12 \times 12$  для каждой карты признаков и ядро  $2 \times 2$ . Слой  $C_3$  это сверточный слой с 16-тью картами признаков размерности  $8 \times 8$  и ядром обхода  $5 \times 5$ . Слой  $C_2$  и  $S_3$  полносвязные, в сравнении с традиционной сетью LeNet5. Слой  $S_4$  это субдискретизирующий слой с 16-тью картами признаков размерности  $4 \times 4$  и ядром  $2 \times 2$ . Последний слой  $C_5$  является выходным слоем, содержащим 10 нейронов и выполняющим классификацию. Как мы видим, основные отличия следующие: 1) мы удаляем 2 последних слоя LeNet5; 2) слой  $S_2$  и  $C_3$  полносвязные; 3) сигмоидная функция активации используется во всех сверточных слоях. Для обучения сети используется групповое обучение.



**Рисунок 3 – Архитектура упрощенной сверточной нейронной сети**

**Таблица 1 – Сверточные сети**

| Classifier  | Preprocessing               | Test error rate (%) |
|---|-----------------------------|---------------------|
| Convolutional net LeNet-1   | Subsampling to 16x16 pixels | 1.7                 |
| Convolutional net LeNet-4   | None                        | 1.1                 |
| Convolutional net LeNet-4 with K-NN instead of last layer           | None                        | 1.1                 |
| Convolutional net LeNet-4 with local learning instead of last layer | None                        | 1.1                 |
| Convolutional net LeNet-5, [no distortions]                         | None                        | 0.95                |
| Convolutional net LeNet-5, [huge distortions]                       | None                        | 0.85                |
| Convolutional net LeNet-5, [distortions]                            | None                        | 0.8                 |
| Simplified convolutional net, [no distortions]                      | None                        | 0.71                |
| Convolutional net Boosted LeNet-4, [distortions]                    | None                        | 0.7                 |

#### 4. ЭКСПЕРИМЕНТЫ

Для того, чтобы показать эффективность данной методики, мы обучили сеть для распознавания рукописных цифр, используя базу данных MNIST. База данных MNIST содержит изображения размерности  $28 \times 28$  пикселей в оттенках серого, набор данных для обучения из 60 000 изображений и набор данных для тестирования из 10 000 изображений. Мы использовали алгоритм обратного распространения ошибки для сверточной нейронной сети без каких-либо модификаций. Размерность группового обучения 50; альфа изменялся от 0.8 до 0.0001. Результаты экспериментов представлены в Таблице 1. Как вы можете видеть, ошибка тестирования составила 0.71% при использовании сверточной сети упрощенной архитектуры. Лучшим результатом сверточной сети LeNet5 без использования distortions является 0.95 %. Таким образом, использование сверточной нейронной сети упрощенной архитектуры с алгоритмом обратного рас-

пространения ошибки позволило достичь более высокой точности распознавания в сравнении с традиционной архитектурой для сверточных сетей.

## 5. ЗАКЛЮЧЕНИЕ

В данной статье рассматривается сверточная нейронная сеть для распознавания рукописных цифр. Мы предлагаем сверточную нейронную сеть с упрощенной архитектурой, которая позволяет распознавать рукописные цифры с более высокой точностью в сравнении с традиционной сверточной нейронной сетью LeNet5. Мы показали, что простая сверточная нейронная сеть способна снизить ошибку тестирования до 0.71% на базе данных MNIST.

## 6. ССЫЛКИ

[1] LeCun, Y., Bengio, Y. and Hinton, G. E. *Deep Learning Nature*, Vol. 521, pp 436-444. (2015)

[2] Golovko, V. *From neural networks to intelligent systems: selected aspects of training, application and evolution IV*. Golovko // *Limitations and future trends in neural computation*. – Amsterdam: IOS press, 2003. – chapter 12. – P. 219-243.

[3] 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.

[4] 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.

[5] Hinton, G. E., Osindero, S., Teh, Y. *A fast learning algorithm for deep belief nets*. *Neural Computation*, 18, 1527-1554 (2006)

[6] Hinton, G. *Training products of experts by minimizing contrastive divergence*. *Neural Computation*, 14, 1771-1800 (2002)

[7] Hinton, G., Salakhutdinov, R. *Reducing the dimensionality of data with neural networks*. *Science*, 313 (5786), 504-507 (2006)

[8] Hinton, G. E. *A practical guide to training restricted Boltzmann machines*. (Tech. Rep. 2010-000). Toronto: Machine Learning Group, University of Toronto (2010)

[9] Bengio, Y. *Learning deep architectures for AI*. *Foundations and Trends in Machine Learning*, 2(1), 1-127 (2009)

[10] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. *Greedy layer-wise training of deep networks*. In B. Scholkopf, J. C. Platt, T. Hoffman (Eds.), *Advances in neural information processing systems*, 11, pp. 153-160. MA: MIT Press, Cambridge (2007)

[11] Golovko, V.A. *Learning Technique for Deep Belief Neural Networks* / V. Golovko, A. Kroshchanka, U. Rubanau, S. Jankowski // in book *Neural Networks and Artificial Intelligence*. – Springer, 2014. – Vol. 440. Communication in Computer and Information Science. – P. 136-146.

[12] Golovko, V. *From multilayer perceptron to deep belief neural networks: training paradigms and application* / in book *"Lectures on neuroinformatics"*, Moscow, 2015, P. 47-84.

[13] Golovko, Vladimir. *A New Technique for Restricted Boltzmann Machine Learning*/Vladimir Golovko, Aliaksandr Kroshchanka, Volodymyr Turchenko, Stanislaw Jankowski, Douglas Treadwell // *Proceedings of the 8th IEEE International Conference IDAACS-2015, Warsaw 24-26 September 2015*. – Warsaw, 2015 –P.182-186.

[14] Fukushima K. *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*//*Biological Cybernetics*,36:193-202,1980.

[15] LeCun Y., Boser B., Denker J., Henderson R., Howard R., Hubbard W. and Jackel L. *Backpropagation applied to handwritten zip code recognition*. *Neural computation*, 1(4):541-551,1989.

[16] LeCun Y., Haffner P., Bottou L. and Bengio Y. *Object recognition with gradient-based learning*. In *shape, Contour and Grouping in Computer Vision*, p.319-345, Springer, 1999.

[17] LeCun Y., Bottou L., Bengio Y. and Haffner P. *Gradient-Based Learning Applied to Document Recognition*. *Proceedings of the IEEE*, 86(11):2278-2324, November 1998