

УДК 004.8

## РЕШЕНИЕ ЗАДАЧ РАСПОЗНАВАНИЯ ЛИЦ И МИМИКИ С ПОМОЩЬЮ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Д. В. Плотников\*, Е. А. Сопов

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева  
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31

\*E-mail: Dmitry.plotnikov.1995@gmail.com

*Проведено экспериментальное исследование эффективности работы сверточных нейронных сетей при решении задач распознавания лица и мимики человека.*

*Ключевые слова: сверточная нейронная сеть, распознавание лиц, распознавание мимики.*

## SOLVING THE FACE AND FACIAL EXPRESSION REGOGNITION TASKS USING CONVOLUTIONAL NEURAL NETWORKS

D. V. Plotnikov\*, E. A. Sopov

Reshetnev Siberian State University of Science and Technology  
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

\*E-mail: Dmitry.plotnikov.1995@gmail.com

*This paper performs experimental research of convolutional neural networks performance operated with face and facial recognition tasks.*

*Keywords: convolutional neural network, face recognition, facial expression recognition.*

**Введение.** Потребность человека в разработке технологий распознавания лиц сформировалась давно. Сегодня такие алгоритмы находят свое применение во многих областях – охранные системы, верификация и т. д. Безусловно, они постоянно совершенствуются и показывают все лучшие результаты. Одним из таких алгоритмов является сверточная нейронная сеть, которая благодаря своей структуре и используемым в ней алгоритмам способна работать с любыми изображениями и видеопотоками.

В данной работе проводится анализ эффективности работы CNN с различными задачами распознавания, а именно: распознавание лица и мимики человека. Поскольку каждая задача распознавания обладает уникальными особенностями (такowymi могут являться: количество обучающих и тестовых примеров, размерность, количество классов и т. д.), возникает следующий вопрос: всегда ли возможно получить высокую эффективность работы алгоритма (CNN) используя лишь классические конфигурации? И всегда ли увеличение гиперпараметров модели влечет за собой повышение эффективности?

В качестве тестовых задач были выбраны следующие: «The ORL Database of Faces» и «The Facial Expression Recognition 2013».

**База данных «The ORL Database of Faces».** Первой тестовой задачей послужила проблема распознавания лиц. База данных представляет собой 400 фотографий сорока разных людей (рис. 1). Все фото представлены в градации серого. Формирование тестовой выборки происходит следующим образом: необходимо случайным образом выбрать по одной фотографии каждого человека. Разрабатываемый алгоритм должен на основе 360 обучающих фотоснимков классифицировать остальные 40, т. е. определить людей, изображенных на них.

**База данных «The Facial Expression Recognition 2013».** Второй тестовой задачей послужила проблема распознавания эмоций. БД представляет собой набор изображений людей с разными выражениями лиц. Человек на изображении может быть: сердитый, испытывающий отвращение, испытывающий страх, счастливый, грустный, удивленный или спокойный.



Рис. 1. Примеры фотографий, взятые из БД «The ORL Database of Faces»



Рис. 2. Примеры фотографий, взятых из БД «The Facial Expression Recognition 2013»

Структура сверточной нейронной сети для работы с «The ORL Database of Faces»

Номер слоя	Слой	Параметры		
		Количество ядер	Размер ядер	Функция активации
1	Сверточный слой 1	20	5×5	Гиперболический тангенс
3	Сверточный слой 2	50	5×5	Гиперболический тангенс
		Размер свертки		Функция свертки
2	Субдискретизирующий слой 1		2×2	Функция максимума
4	Субдискретизирующий слой 2		2×2	Функция максимума
		Количество нейронов		Функция активации
5	Полносвязный слой 1		500	Гиперболический тангенс
6	Полносвязный слой 2		40	Гиперболический тангенс

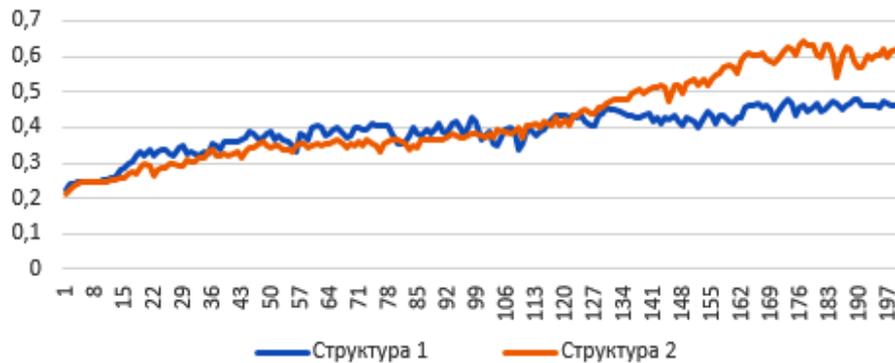


Рис. 3. Динамика эффективности обучения CNN (Структура 1 и Структура 2)

Выборка состоит из 28709 обучающих и 7178 тестовых примеров. Каждое изображение имеет разрешение 48×48 пикселей.

**Результаты экспериментов. База данных «The ORL Database of Faces».** С принципом устройства работы CNN можно ознакомиться в [1].

В ходе работы также были использованы и другие структуры для работы с базой данных «The ORL Database of Faces». Однако они оказались неэффективны. С одной стороны, более сложные и ресурсоемкие для вычислений конфигурации не показывали лучший результат. А более простые структуры – не могли быть обучены вообще. Таким образом, наилучшим образом обученная структура достигает эффективности классификации 97,5 %.

**База данных «The Facial Expression Recognition 2013».** Поскольку размерность данной задачи сравнительно велика, принято решение уменьшить размерность (до 28×28). В данном случае на вход нейронной сети будет подано изображение меньшей размерности. Практика показала, что время обучения CNN при входных изображениях 48×48 пикселя в 4 раза дольше, чем при входных изображениях 28×28 пикселя. Более того, динамика обучения сети остается почти неизменной.

Следующий этап – проверка гипотезы о том, что значения гиперпараметров CNN влияют на скорость

обучения. Для решения задачи использовались две структуры: Структура 1 – идентична той, что была использована для работы с БД «The ORL Database of Faces».

Структура 2 – это Структура 1 с увеличенными значениями гиперпараметров (теперь количество ядер первого сверточного слоя = 50, второго сверточного слоя = 125) (рис. 3).

На рис. 3: ось абсцисс – итерации обучения сети, ось ординат – точность обучения. Видно, что на 125 итерации Структура 1 достигает своего уровня стагнации, в то время как Структура 2 продолжает увеличивать точность распознавания.

**Выводы.** Результаты показали, что для решения разных задач распознавания человеческого лица (или эмоций), использование одной и той же классической структуры может оказаться неэффективным. Необходимо исходить из цели и искать компромисс между эффективностью и вычислительной мощностью аппаратных средств, так как время работы CNN крайне чувствительно к размерности входных данных, а результат не всегда будет зависеть от размерности обучающей выборки. При работе со сложными структурами требуется предобработка данных и тонкая настройка сети.

Данное направление и будет являться объектом дальнейших исследований.

### References

1. Face recognition: a convolutional neural-network approach / S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back // *IEEE Trans. Neural Networks*. 1997. Vol. 8, № 1. P. 98–113.
2. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet Classification with Deep Convolutional Neural Networks, *Adv. Neural Inf. Process. Syst.* 2012. P. 1–9.
3. LeCun Y., Boser B., Denker J. S. et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* 1989. № 1 (4). P. 541–551.
4. Lin M., Chen Q., Yan S. Network In Network. *arXiv Prepr.* 2013:10.
5. Sopov E., Ivanov I. Design efficient technologies for context image analysis in dialog HCI using self-configuring novelty search genetic algorithm. *Informatics Control Autom Robot (ICINCO)*, 2014. 11th Int Conf. 2014. № 2. P. 832–839.

© Плотников Д. В., Сопов Е. А., 2017