

# Нейросетевой метод обнаружения размытых изображений лиц

В.Г. Спицын<sup>1</sup>, Ю.В. Савицкий<sup>1</sup>, А.Б. Казиев<sup>1</sup>, Ю.А. Болотова<sup>1</sup>  
 spvg@tpu.ru|yv18@tpu.ru|kaziev\_ab@mail.ru|jbolotova@tpu.ru

<sup>1</sup>Национальный исследовательский Томский политехнический университет, Томск, Россия

*Предложен метод оценки степени размытия изображения лица на основе аппарата свёрточных нейронных сетей. исследована возможность использования в качестве входных данных для нейронной сети исходного изображения, модульной составляющей частотного спектра и фазовой составляющей частотного спектра исходного изображения. Тестирование метода проводилось на собственной базе изображений лиц, полученных с ip-камеры в реальных условиях. Предложенный метод сравнивается с известным методом оценки размытия на основе количественного анализа модульных составляющих частотного спектра изображения. На собранной тестовой выборке предложенный метод показал точность распознавания 98,57%, метод на основе количественного анализа модульных составляющих частотного спектра показал точность 77,12%.*

**Ключевые слова:** свёрточные нейронные сети, оценка качества изображения, преобразование Фурье.

## Neural network method for detecting fuzzy images of faces

V.G. Spitsyn<sup>1</sup>, Yu.V. Savitsky<sup>1</sup>, A.B. Kaziev<sup>1</sup>, Yu.A. Bolotova<sup>1</sup>  
 spvg@tpu.ru|yv18@tpu.ru|kaziev\_ab@mail.ru|jbolotova@tpu.ru

<sup>1</sup>National Research Tomsk Polytechnic University, Tomsk, Russia

*A method for estimating the degree of blurring of an image of a person based on the apparatus of convolutional neural networks is proposed. The possibility of using the original image, the modular component of the frequency spectrum and the phase component of the frequency spectrum of the original image as input to the neural network is investigated. The method was tested on an own image base of faces obtained from an ip camera in real conditions. The proposed method is compared with the known method of estimating the blur based on a quantitative analysis of the modular components of the frequency spectrum of the image. On the collected test sample, the proposed method showed the recognition accuracy of 98.57%, the method based on the quantitative analysis of the modular components of the frequency spectrum showed an accuracy of 77.12%.*

**Keywords:** convolutional neural networks, image quality estimation, Fourier transform.

### 1. Введение

В настоящее время повсеместное распространение получают системы биометрической идентификации человека по изображению лица. Условия, в которых приходится работать этим системам, сильно отличаются от условий, в которых набираются изображения для бенчмарков. В данной работе рассматриваются изображения, взятые с камер наблюдения внутри помещения, например, на проходной предприятия.

В условиях получения изображений в реальном времени, входные данные могут подвергаться искажениям. Одним из видов искажений, встречаемых наиболее часто в ходе работы в реальных условиях, является искажение «размытия». К данному виду могут быть отнесены искажения, получающиеся в результате смазывания в движении или расфокусировки камеры. Данный вид искажений приводит к тому, что изображение становится непригодным для распознавания. Отбраковка таких изображений позволит снизить нагрузку на систему распознавания, сократив тем самым вычислительные ресурсы, необходимые для работы биометрической системы.

При оценке качества изображения, в смысле оценки степени его размытия, алгоритмы оценки можно разделить на следующие категории [14, 18, 19]:

1. Эталонные методы – предполагают наличие эталонного изображения, которое изначально не содержит искажений. В таком случае оценка качества изображения может основываться на оценке схожести изображений, в том или ином смысле.
2. Не эталонные методы – предполагают отсутствие эталонного изображения.

Псевдоэталонные методы – предполагают наличие ограниченной дополнительной информации о том, какими

характеристиками должно обладать качественное изображение.

В данной работе предложен метод оценки качества изображений на основе использования аппарата свёрточных нейронных сетей (СНС). Предложенный метод может быть отнесён к неэталонным методам оценки качества изображений.

### 2. Обоснование выбора СНС для оценки качества изображений

Свёрточная нейронная сеть - особый вид архитектуры нейронной сети прямого распространения, специально разработанный для анализа изображений. Основной особенностью данного вида архитектуры нейронной сети является использование рецептивных полей, т.е. фильтров, чувствительных к определённому визуальному образу.

Предшественниками свёрточных нейронных сетей были модели когнитрона и неокогнитрона [5]. Свёрточные нейронные сети в современном виде были представлены в работах Ле Куна [11-13] и А. Krizhevsky, I. Sutskever, G.E. Hinton [10]. Основными слоями, используемыми в свёрточных сетях, являются свёрточный слой, слой субдискретизации и полносвязный слой.

Сочетание функции классификации с функцией выделения признаков с помощью ядер свёртки, получаемых в процессе обучения, позволяет выделять оптимальный набор признаков. Получить данный набор признаков, подбирая метод извлечения признаков вручную, является практически невозможной задачей.

Исходя из вышесказанного, было выдвинуто предположение, что аппарат свёрточных нейронных сетей может показать высокую эффективность в решении задачи оценки степени размытия изображения. Для проверки выдвинутого предположения было принято решение об

исследовании различных вариантов архитектур, с целью определения наиболее эффективной для решения поставленной задачи.

Полученную модель необходимо сравнить с подходом, не использующим нейронные сети по характеристике эффективности решения поставленной задачи. Так как данная работа нацелена на использование в конечной системе распознавания лиц, реализуемая модель должна быть достаточно компактной и обеспечивать возможность работы в реальном времени.

В качестве базовых видов архитектур, на основе которых можно построить архитектуру сети, были рассмотрены LeNet [10-13], Inception [1, 16] и ResNet [6]. Данные виды архитектур являются наиболее распространёнными на текущий момент.

На основе анализа работ, использующих сети с указанными архитектурами, и предварительном тестировании сложности вычислений, был сделан вывод, что для данной работы будет целесообразно использовать сеть, построенную на основе архитектуры LeNet. Данный вариант архитектуры используется в качестве базового для подбора основных параметров обучения сети.

Модификации архитектуры LeNet производятся путём:

- изменения количества слоёв свёртки и субдискретизации;
- изменения количества и размера фильтров слоёв свёртки;
- изменения размера и шага областей субдискретизации;
- изменения количества и размера полносвязных слоёв;
- добавления слоёв нормализации;
- изменения нелинейной функции активации нейронов.

### 3. Формирование обучающей и тестовой выборки

Отбор обучающего материала является одним из наиболее значимых факторов в машинном обучении и в обучении нейронных сетей в частности. Для формирования обучающей выборки были записаны видеоролики прохождений людей по направлению к камере. В ходе записи менялись условия освещения, высота установки и угол наклона камеры относительно горизонта. Для выделения изображений лиц на видео была использована реализация детектора лиц на основе метода Виолы-Джонса открытой библиотеки OpenCV. После обработки полученных роликов алгоритмом детекции лиц, были получены 200 000 изображений.

Сортировка полученных изображений была проведена вручную, так как если бы имелось какое-либо средство автоматической сортировки по заданному критерию, то не было бы необходимости в данной работе. Весь процесс сортировки занял 7 дней, в результате изображения были распределены по 4 группам:

1. Изображения с высокой степенью размытия
2. Изображения без размытия
3. Изображения с промежуточной степенью размытия
4. Бракованные изображения (пересвеченные, затемнённые, с высокой степенью зашумления).

В результате сортировки к первой и второй группе были отнесены 130 000 изображений. Итоговое соотношение чётких изображений и размытых изображений оказалось 3 к 1.

В качестве входных данных были рассмотрены три варианта, примеры изображений приведены на рис. 1:

1. Исходное изображение (левая колонка).
2. Изображение модульных составляющих частотного спектра (по центру).
3. Изображение фазовых составляющих частотного спектра (правая колонка).

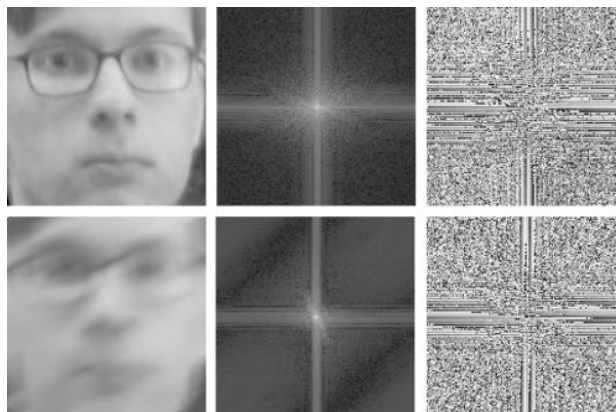


Рис. 1. Примеры изображений для обучения: верхний ряд – чёткое изображение и его спектральные изображения, нижний ряд – размытое изображение и его спектральные изображения.

Для обучения нейронных сетей была выбрана библиотека Dlib [9]. Данная библиотека ориентирована на применение в коммерческой разработке. В отличие от большинства библиотек машинного обучения она ориентирована на использование `api` на языке C++. Активно использует стандарт C++11. По сравнению с библиотекой Caffe [2] Dlib обеспечивает большую скорость работы, за счёт формирования архитектуры сети с помощью шаблонов.

При обучении сети задача обучения была сформулирована как задача бинарной классификации. В данном случае первым классом являлись не размытые изображения, а вторым классом – размытые изображения.

В работе проведен анализ особенностей модификаций метода градиентного спуска: стохастический градиентный спуск, ускоренный градиентный спуск Нестерова [15], Adagrad [4], RMSProp/AdaDelta [17, 20], Adam [3]. На основе проведенного сопоставления перечисленных модификаций метода градиентного спуска для обучения сети был выбран алгоритм Adam. Данный метод был выбран в первую очередь потому, что при его использовании процесс обучения сходится быстрее и занимает меньше времени, что позволяет провести большое количество запусков обучения конкретного варианта архитектуры и собрать достоверную статистику.

Для регуляризации весов использовался приём `batch normalization` [7, 13]. Использовался приём уменьшения коэффициента обучения на порядок по достижению сетью плато на поверхности ошибок.

Для поиска оптимальных значений параметров обучения сети были выбраны три основных параметра, оказывающих наиболее значимое влияние на процесс обучения нейронной сети: начальная скорость обучения или коэффициент обучения, коэффициент регуляризации и число итераций без снижения ошибки, после которого происходит снижение коэффициента обучения. Общий алгоритм подбора параметров выглядит следующим образом:

1. Подбор начальной скорости обучения. Данный параметр признаётся как наиболее значимый параметр обучения, поэтому его рекомендуется подбирать в первую очередь.
2. Подбор параметра регуляризации. На предыдущем этапе данный параметр устанавливался равным 0.
3. Подбор предельного числа итераций без прогресса. Увеличение данного параметра увеличивает продолжительность обучения, давая алгоритму обучения больше времени на поиск оптимального решения. Однако слишком большое число итераций без прогресса может привести к переобучению модели.

При каждом изменении какого-либо параметра обучалось по 10 вариантов весов для базовой архитектуры сети. Каждый полученный вариант сети тестировался на 10 наборах тестовой выборки. Итоговые величины среднего значения и стандартного отклонения точности определялись по результатам всех тестовых проходов.

На основе проведенных экспериментов установлено, что наилучших результатов при тестировании обучаемой модели удаётся добиться при использовании значения начальной скорости обучения равном  $0.5 \cdot 10^{-2}$ , для выборки изображений модульных составляющих частотного спектра –  $0.5 \cdot 10^{-2}$  и для выборки фазовых составляющих частотного спектра –  $0.5 \cdot 10^{-3}$ .

На следующем этапе было определено оптимальное значение параметра регуляризации. Значение параметра регуляризации подбиралось при использовании значения параметра начальной скорости обучения, определенного в ходе предыдущего этапа, значение параметра числа итераций без прогресса как и на предыдущем этапе равнялось 300. В результате проведенных экспериментов установлено, что наилучших результатов при тестировании обучаемой модели удаётся добиться при использовании значения параметра регуляризации для выборки исходных изображений равном  $0.5 \cdot 10^{-5}$ , для выборки модульных составляющих частотного спектра –  $0.5 \cdot 10^{-4}$ , для фазовых составляющих частотного спектра –  $0.5 \cdot 10^{-3}$ .

На заключительном этапе определения параметров обучения был подобран параметр “число итераций без прогресса”. Данный параметр напрямую влияет на продолжительность обучения сети. Проведенные эксперименты показали, что оптимальным значением параметра для всех выборок является значение 1000.

По результатам обучения и тестирования наибольшую точность продемонстрировали сети, обученные на выборке исходных изображений. На втором месте по точности оказались сети, обученные на выборке модульных составляющих частотного спектра. С большим отставанием на третьем месте оказались сети, обученные на фазовых составляющих частотного спектра.

#### 4. Выбор архитектуры сети

Как уже говорилось ранее, одним из требований, предъявляемых к разрабатываемому методу, является его быстродействие. Быстродействие нейронных сетей напрямую зависит от числа параметров, которое использует сеть. В таблице 1 приведено описание предложенных вариантов архитектуры LeNet: BN1-BN4.

В таблице 1 используются следующие обозначения: Input – размер входного изображения, Con – слой свёртки, MaxPool – слой субдискретизации, FC – полносвязный слой, N – число ядер свёртки в слое, K – размер окна свёртки или субдискретизации в пикселях, S – шаг смещения в пикселях, Fun – нелинейная функция, Sigmoid – функция логистической сигмоиды. Входной размер изображения 183x183 обусловлен особенностями работы целевой системы распознавания лиц, в составе которой планируется использовать обученную модель. Полученную модель можно использовать не только как бинарный классификатор размытое/чёткое. На выходе последнего слоя будет число из диапазона (0;1), которое укажет «степень размытия», позволяя конечной системе самой решать, с каким порогом производить фильтрацию входных изображений.

Тестирование предложенных вариантов архитектуры LeNet: BN1-BN4 проходило по тому же принципу, что и во время подбора параметров обучения. На каждый вариант архитектуры было обучено по 10 вариантов весов, каждый из которых был протестирован на 10 тестовых выборках

для получения среднего значения точности и его стандартного отклонения.

Таблица 1. Описание вариантов архитектур

	BN1	BN2	BN3	BN4
Input	183x183	183x183	183x183	183x183
Con1	N=16 K=5x5 S=1x1	N=8 K=5x5 S=1x1	N=8 K=5x5 S=1x1	N=16 K=5x5 S=1x1
MaxPool1	K=3x3 S=2x2	K=3x3 S=2x2	K=3x3 S=2x2	K=3x3 S=2x2
Con2	N=32 K=3x3 S=1x1	N=16 K=3x3 S=1x1	N=16 K=3x3 S=1x1	N=32 K=3x3 S=1x1
MaxPool2	K=3x3 S=2x2	K=3x3 S=2x2	K=3x3 S=2x2	K=3x3 S=2x2
Con3	N=64 K=3x3 S=1x1	N=32 K=3x3 S=1x1	N=16 K=3x3 S=1x1	-
MaxPool3	K=3x3 S=2x2	K=3x3 S=2x2	K=3x3 S=2x2	-
Con4	-	-	N=32 K=3x3 S=1x1	-
MaxPool4	-	-	K=3x3 S=2x2	-
FC	160	120	80	160
Fun	Sigmoid	Sigmoid	Sigmoid	Sigmoid
FC	1	1	1	1
Fun	Sigmoid	Sigmoid	Sigmoid	Sigmoid

Результаты тестирования представлены на рис. 2.

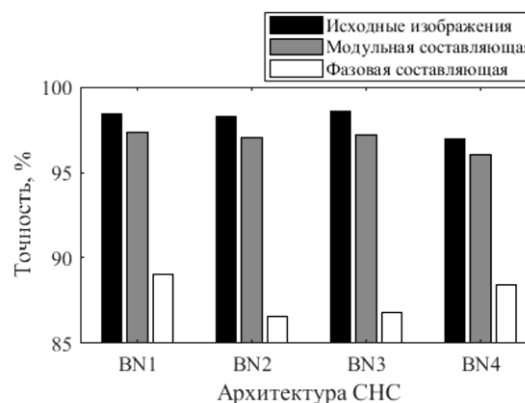


Рис. 2. Результаты тестирования архитектур СНС.

Наиболее эффективным оказался подход, при котором в качестве вектора входных признаков использовалось исходное изображение лица. Это говорит о высокой информативности признаков, содержащихся в данных специфических изображениях, для СНС с точки зрения оценки размытости и последующей бинарной классификации. Данный подход показывает как наибольшую точность, так и наибольшую стабильность, то есть вероятность высоких результатов обучения при использовании данного вектора признаков и подобранных параметров обучения.

Подход, при котором в качестве входного вектора признаков использовались модульные составляющие гармоник частотного спектра, показал меньшую эффективность относительно первого подхода. Наименее эффективным оказался третий подход, при котором использовались фазовые составляющие гармоник частотного спектра.

В ходе тестирования были произведены замеры времени, которое необходимо для обработки изображения каждой архитектурой. Тестирование проводилось на восьмиядерном CPU AMD FX-8320 3.5 ГГц, размер ОЗУ 8

Гб. Библиотека Dlib была собрана с поддержкой пакета OpenBLAS из репозитория Ubuntu 14.04 и без поддержки GPU. При использовании данной конфигурации среднее время обработки одного изображения для BN1 составило 15.0 миллисекунды (мс), для BN2 – 9.3 мс, для BN3 – 8.5 мс, для BN4 – 18.4 мс.

Следует отметить, что все 4 архитектуры показали сопоставимый результат на каждой из выборок исходных данных. Наилучший итоговый результат показали сети, обученные на выборке исходных изображений, наихудший – на выборке фазовой составляющей частотного спектра. Исходя из полученных данных следует, что для использования предложенного метода в конечном приложении следует выбрать вариант BN3 архитектуры LeNet, так как данная архитектура является самой компактной.

## 5. Анализ полученных результатов

На заключительном этапе было проведено сравнение результатов предложенного метода с методом, описанном в работе [8].

Метод из [8] был реализован и протестирован на описанном выше наборе тестовых данных изображений лиц. Полученные результаты показывают точность классификации 77,12 % , что значительно уступает значению 98,57%, приведенному на рис. 2.

Таким образом, предложенный в данной работе метод обнаружения размытия на изображениях лиц на основе сверточных нейронных сетей существенно превосходит метод, основанный на подсчете модульной составляющей спектра [8].

## 6. Заключение

Предложен и протестирован нейросетевой метод обнаружения размытых изображений лиц. Результаты сопоставления предложенного метода с известным методом, основанным на подсчете модульных составляющих спектра изображения, показали точность распознавания размытых изображений лиц превосходящую на 21,45 % результаты тестирования метода [8] на той же базе изображений. Исходя из распределения четких и размытых изображений, на основе применения разработанного метода, можно сократить количество обрабатываемой информации на 30%, что приведет к существенному снижению нагрузки на систему распознавания лиц.

## 7. Благодарности

Работа выполнена в рамках Программы повышения конкурентоспособности ТПУ при финансовой поддержке РФФИ в рамках научного проекта № 18-08-00977 А.

## 8. Литература

- [1] Arora S., Bhaskara A., Ge R., Ma T. Provable Bounds for Learning Some Deep Representations // arXiv. 2013. №1310.6343v1.
- [2] Caffe deep learning framework. Available at: <http://caffe.berkeleyvision.org> (accessed 20 October 2017).
- [3] Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization // Source: <https://arxiv.org/pdf/1412.6980.pdf>.
- [4] Duchi J., Hazan E., Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research. 2011, Vol. 12, pp. 2121-2159.

- [5] Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition // Neural networks. 1988, Vol. 1(2), pp. 119–130.
- [6] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // arXiv. 2015. №1512.03385, [http://www.ms.uky.edu/~qye/MA721/JinpengLiu\\_ResNet.pdf](http://www.ms.uky.edu/~qye/MA721/JinpengLiu_ResNet.pdf).
- [7] Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // arXiv. 2015. Source: <https://arxiv.org/pdf/1502.03167.pdf>.
- [8] Kanjar De., Masilamani V. Image Sharpness Measure for Blurred Images in Frequency Domain // Procedia Engineering. 2013, Vol. 64, pp. 149-158.
- [9] King, D. E. Dlib-ml: A Machine Learning Toolkit // Journal of Machine Learning Research. 2009, Vol. 10, pp. 1755-1758.
- [10] Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks // Advances in neural information processing systems. 2012, pp. 1097–1105.
- [11] LeCun Y. Backpropagation applied to handwritten zip code recognition // Neural computation. 1989, Vol. 1(4), pp. 541–551.
- [12] LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. 1998, Vol. 86(11), pp. 2278-2324.
- [13] LeCun Y. Efficient backprop // Neural Networks: Tricks of the Trade: Lecture Notes in Computer Science / G. Montavon, G. B. Orr, K.-R. Muller (Eds.) – Springer, 2012, pp. 9–48.
- [14] Marziliano P., Dufaux F., Winkler S., Ebrahimi T. A no-reference perceptual blur metric // Proceedings of IEEE international conference on Image Processing. 2002, Vol. 3., pp. 57-60.
- [15] Sutskever I., Martens J., Dahl G., Hinton G. On the Importance of Initialization and Momentum in Deep Learning // Proceedings of the 30th International Conference on Machine Learning. 2013, Vol. 28., pp. III-1139–III-1147.
- [16] Szegedy C. et al. Going deeper with convolutions // arXiv. 2014. Source: <https://arxiv.org/pdf/1409.4842.pdf>.
- [17] Tieleman T., Hinton G. RMSProp: Divide the gradient by a running average of its recent magnitude // Coursera: Neural Networks for Machine Learning: Technical report. 2012, pp. 26-31.
- [18] Tong H., Li M., Zhang H., Zhang C. Blur detection for digital images using wavelet transform // Proceedings of IEEE international conference on Multimedia and Expo. 2004, Vol. 1, pp. 17-20.
- [19] Wang Z., Bovik A. C. Modern image quality assessment: Synthesis lectures on image, video & multimedia processing.– San Rafael: Morgan & Claypool Publishers, 2006, 156 p.
- [20] Zeiler M. Adadelta: An adaptive learning rate method // arXiv. 2012. Source: <https://arxiv.org/pdf/1212.5701.pdf>.

## Об авторах

Спицын Владимир Григорьевич, д.т.н., профессор отделения информационных технологий Национального исследовательского Томского политехнического университета. E-mail: [spvg@tpu.ru](mailto:spvg@tpu.ru).

Савицкий Юрий Васильевич, аспирант отделения информационных технологий Национального исследовательского Томского политехнического университета. E-mail: [yvs18@tpu.ru](mailto:yvs18@tpu.ru).

Казиев Александр Борисович, магистрант отделения информационных технологий Национального исследовательского Томского политехнического университета. E-mail: [kaziev\\_ab@mail.ru](mailto:kaziev_ab@mail.ru).

Болотова Юлия Александровна, к.т.н., доцент отделения информационных технологий Национального исследовательского Томского политехнического университета. E-mail: [jbolotova@tpu.ru](mailto:jbolotova@tpu.ru).