

# Использование сверточных нейронных сетей в комбинации с преобразованием Хафа для классификации изображений с прямыми линиями\*

А. В. ШЕШКУС

**Аннотация.** В представленной работе предлагается подход, основанный на расширении входного вектора признаков сверточной нейронной сети, основанный на использовании преобразования Хафа. Идея подхода состоит в том, что некоторым фильтрам сверточного слоя на вход будет подано изображение, полученное последовательным морфологическим контрастированием и преобразованием Хафа входного изображения. Таким образом, вычислительная сложность и количество коэффициентов в нейронной сети не изменяется, единственная дополнительная операция – предобработка изображения и преобразование Хафа. Работа предлагаемого метода была продемонстрирована на очень простой архитектуре сети на двух различных задачах. Первая задача – классификация изображений из открытого набора данных CIFAR-10, вторая – распознавание печатных символов на закрытом наборе данных, собранном с российских паспортов. Предлагаемый подход позволил достичь улучшения качества без существенного увеличения времени вычисления, что может быть важно для ряда систем распознавания и классификации изображений с прямыми линиями.

**Ключевые слова:** сверточная нейронная сеть, преобразование Хафа, вычисление признаков.

## Введение

В современном мире технологии компьютерного зрения очень быстро развиваются и находят все больше и больше применений в самых разных областях. Распознавание изображений требуется все в большем количестве технических задач. Самые общепринятые алгоритмы распознавания базируются на использовании нейронных сетей.

Нейронные сети изменяются и развиваются с самого момента их введения. Сейчас имеется возможность обучения весьма сложных нейронных сетей на огромных обучающих базах за разумное время благодаря непрекращающемуся процессу развития компьютеров. Классическая нейронная сеть состоит из слоев, первый из которых принимает на вход интенсивности пикселей картинки, как правило, предварительно обработанной какими-либо методами для приведения к соответствующему размеру и подавления шумов различного рода. Первая группа слоев предназначена для извлечения признаков из исходного изображения, которые потом анализирует оставшаяся часть сети. Ранние нейронные сети имели построенные вручную вычислители признаков, но сейчас эту роль часто выполняют сверточные слои. Они хорошо работают в задачах распознавания текста, так как позволяют из

разных участков изображения извлекать одни и те же признаки.

Глубокие сверточные нейронные сети отлично справляются с большинством задач распознавания, однако сеть с одним слоем, как правило, не способна научиться извлекать ряд релевантных задач признаков, которые, тем не менее, могли бы быть посчитаны с помощью полносвязного слоя. С другой стороны, полносвязные слои достаточно проблемно использовать для подсчета признаков в силу того, что для этого требуется значительно больший обучающий набор данных и дополнительные методы борьбы с переобучением.

Один из признаков, который может быть подсчитан с помощью полносвязного слоя и не может быть подсчитан с помощью сверточного – наличие прямых линий, которое в свою очередь отражает преобразование Хафа. Это преобразование позволяет детектировать прямые линии (в модифицированном виде – и другие кривые) на изображении [1, 2] и может быть использовано в задачах, где нужно анализировать прямые линии или их наличие [3–5]. Например, в [6] преобразование Хафа в комбинации с нейронными сетями использовали при решении задачи распознавания языка жестов. Несмотря на то, что в русских и в английских (и во многих других) символах встречаются прямые штрихи, классические нейронные сети для их распознавания не включают в себя преобразование Хафа.

\* Работа выполнена при поддержке РФФИ, проекты № 16-29-09508 офи\_м

С другой стороны, в работе 1996 года [7] использовано преобразование Хафа в совокупности с аугментацией в вычислении признаков для задачи распознавания рукописных цифр с помощью нейронной сети. Наилучший результат, который был тогда достигнут – 89%, что достаточно хорошо при использованном количестве данных. В работе Коха [8] полносвязные нейронные сети с использованием преобразования Хафа в предобработке изображения были использованы для распознавания корейских иероглифов с качеством 79%. А в работе [9] также использовали преобразование Хафа для решения задачи классификации римских символов с использованием модифицированного метода ближайших соседей.

Тем не менее, преобразование Хафа кажется недооцененным как метод вычисления признаков для задач распознавания и классификации. В данной работе предлагается использовать преобразование Хафа в сверточных нейронных для дополнительного вычисления признаков с целью повышения качества распознавания. Такой подход может принести пользу в задачах от распознавания символов в документах до классификации торосов на ледяных поверхностях, так как они тоже содержат существенное количество прямых штрихов в своей структуре.

В первой части данной статьи будет описано преобразование Хафа, во второй – предложенная архитектура нейронной сети. В третьей части будут представлены результаты экспериментов.

## 1. Преобразование Хафа

Преобразование Хафа [1] – это отображение, описываемое формулой :

$$H(s, \alpha) = \sum_{(x,y) \in l(s,\alpha)} I(x, y), \quad (1)$$

где  $l(s, \alpha)$  – прямая на исходном изображении, имеющая угол  $\alpha$  и сдвиг  $s$ . Каждая точка Хаф-образа представляет интегрированную интенсивность пикселей вдоль прямой с соответствующими смещением и углом наклона на исходном изображении. В случае мультисканального изображения преобразование Хафа делается для каждого канала по отдельности. В результате количество каналов на входном изображении и на Хаф-образе всегда совпадает.

Хаф-образ имеет большую интенсивность в зонах, соответствующих прямолинейным всплескам интенсивности на исходном изображении и, следовательно, может быть использовано для дальнейшего анализа изображения.

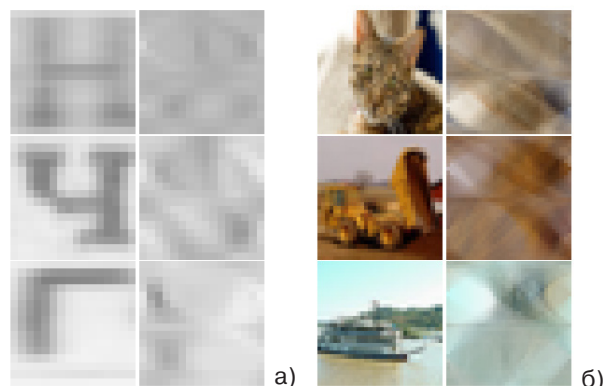


Рис. 1. Пример Хаф-образа для а) одноканального изображения, б) трехканального изображения

Примеры для одноканального и трехканального изображения приведены на рис. 1.

Хаф образ содержит в себе информацию о количестве прямых линий, их ориентации и кластеризации. Данная информация может оказаться полезной для определения и классификации объектов, распознавания символов и другие. Более того, в данных задачах можно использовать не классическое преобразование Хафа, а быстрое и, тем самым, путем увеличения вычислительной эффективности сильно расширить спектр устройств, на которых подобный алгоритм может выполняться в режиме реального времени [10].

Перед применением преобразования Хафа проводится специальная предобработка изображения. Дело в том, что многие объекты на изображении имеют прямые границы, но не линии. В данной работе мы выполняем морфологическое контрастирование изображения для перевода прямых границ в прямолинейные всплески интенсивности. Морфологическое контрастирование – модуль разности морфологических эрозии и дилатации изображения. Влияние этого преобразования на общее время вычисления незначительно, так как оно вычисляется на 8-битной картинке, а сеть – в 32-битных вещественных числах и операций там существенно больше.

## 2. Предложенная архитектура

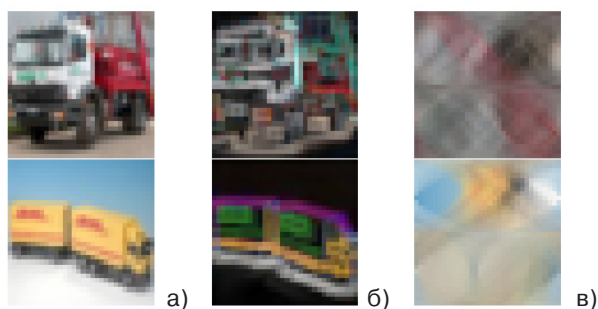
Целью исследования является увеличение эффективности сверточной нейронной сети, используя преобразование Хафа для вычисления признаков.

Идея заключается в том, чтобы заменить вход нескольких (но не всех) сверточных фильтров на изображение, полученное последовательным применением морфологического контрастирования и преобразования Хафа. В данной работе рассма-

тривается нейронная сеть только с одним сверточным слоем для большей наблюдаемости эффекта. Кроме этого, глубокий блок сверточных фильтров может научиться извлекать признаки вдоль некоторых отрезков, длина которых зависит от параметров слоев и их количества.

Предлагаемая сеть имеет две ветви в каждой из которых ровно один сверточный слой. Результаты этих двух ветвей далее объединяются в единый вектор и подаются на вход одному полносвязному слою. Одна ветвь получает исходное изображение, а другая – изображение, полученное в результате преобработки и преобразования Хафа.

Необходимо, так же, упомянуть, что преобразование Хафа в данном алгоритме делается специальным образом по определенному количеству углов и сдвигов. Таким образом, Хаф образ имеет такой же размер, как и входное изображение. На рис. 2 представлены примеры изображений, результаты преобработки и Хаф-образы.



**Рис. 2.** Примеры а) изображений, б) результатов преобработки, в) Хаф-образов

Результат работы описанной нейронной сети сравнивался с результатом работы сети с такой же архитектурой, только на вход обеим ветвям подавалось исходное изображение, т.е. в обоих случаях нейронные сети имеют одинаковое количество коэффициентов.

### 3. Эксперименты и результаты

**Данные.** Были проведены эксперименты на двух различных наборах данных. Первым является открытый CIFAR-10 [11], который состоит из 60000 цветных изображений размера 32x32. Объекты принадлежат к десяти различным классам: самолет, автомобиль, птица, кошка, олень, собака, лягушка, лошадь, корабль, грузовик (примеры на рис. 3).

Для данного набора данных не было использовано никаких преобработок или аугментаций, т.е. все картинки были использованы для обучения «как есть».

Вторым тестовым набором данных были изображения символов, собранные с российских паспортов (примеры на рис. 3). Эти символы включают буквы, цифры и ряд специальных символов. В общей сложности алфавит состоит из 50 различных символов. Количество изображений в этом наборе данных больше – примерно 1400000, каждое из которых имеет размер 16x16. Из этих изображений 1024000 было использовано для обучения, 204800 – для валидации и последние 204800 – для теста.

**Эксперименты.** Для подтверждения работоспособности метода был проведен набор экспериментов для архитектуры нейронной сети, описанной выше. Все фильтры в сверточном слое имели размер 5x5, но их количество отличалось в разных группах экспериментов. Для первой ветви, которая на вход получала оригинальное изображение, количество фильтров было 16, 32, 64, 128. Для второй ветви, которая на вход получала Хаф-образ, количество фильтров было 4, 8, 16, 32. Таким образом было проведено 4 серии экспериментов. Количество весов в полносвязном слое в этих сериях было переменным, так как оно зависит от размерности выхода предыдущего слоя и может быть вычислено из размера изображения, количества фильтров, размеров фильтров и алфавита.

Важно заметить, что сверточный слой не имел «падинга», т.е. дополнительного расширения, что является весьма обычной практикой. Из определения преобразования Хафа следует, что пиксели Хаф-образа имеют не такую связь с соседними пикселями, как пиксели исходного изображения.

Подобная структура экспериментов позволяет судить как об оптимальном количестве фильтров, так и о пользе нашего метода.

**Результаты.** Результаты проведенных серий экспериментов представлены в табл. 1 для CIFAR-10 и набора данных с паспорта. Стоит заметить, что каждый из экспериментов был проведен по 10 раз и результат усреднен. Это было сделано с целью увеличения достоверности результата. Из результатов видно, что предложенная модификация вычисления признаков уменьшает количество ошибок для любого из испытанных количества фильтров. Следовательно, преобразование Хафа может быть использовано для увеличения качества распознавания почти без увеличения вычислительной сложности. Благодаря этому предложенный алгоритм может быть особенно полезен в различных процессах реального времени где малое время работы алгоритма критически важно [12] и применяются специальные методы оптимизации [13–16].

Из таблицы видно, что эффект особенно заметен при малом числе фильтров.



Рис. 3. Примеры изображений из а) CIFAR-10, б) символы из паспорта

**Заключение**

Результаты показывают, что использование Хаф-образа вместе с исходным изображением в качестве входа для сверточной нейронной сети может привести к улучшению качества распознавания, но эффект более заметен в архитектурах, где количество фильтров в свертках меньше. При увеличении числа фильтров сверточный слой учится извлекать все более и более сложные признаки, который, в той или иной степени, могут заменять признаки, собираемые по Хаф-образу. Предложенный подход пока еще не готов к реальному использованию, но выглядит очень перспективным для дальнейших исследований. Особенно важным это может оказаться в ряде систем распознавания, в том числе – систем распознавания в реальном времени, в которых скорость работы критически важна.

**Литература**

1. *P. Hough*, “Method and means for recognizing complex patterns,” (Dec 1962).
2. *D. H. Ballard*, “Readings in computer vision: Issues, problems, principles, and paradigms,” ch. Generalizing the Hough Transform to Detect Arbitrary Shapes, 714–725, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1987).
3. *T. Khanipov, I. Koptelov, A. Grigoryev, E. Kuznetsova, and D. Nikolaev*, “Vision-based industrial automatic vehicle classifier,” (2015).
4. *N. Skoryukina, D. P. Nikolaev, A. Sheshkus, and D. Polevoy*, “Real time rectangular document detection on mobile devices,” (2015).
5. *D. Krokhina, V. Blinov, S. Gladilin, I. Tarhanov, and V. Postnikov*, “Fast roadway detection using car cabin video camera,” (2015).

**Таблица 1**

Сравнение ошибки для разного количества фильтров

	Фильтры по изображению	Фильтры по Хаф-образу	CIFAR-10		Паспорт	
			Изображение	Изображение + Хаф-образ	Изображение	Изображение + Хаф-образ
1	16	4	47.5%	43.4%	7.2%	6.9%
2	32	8	43.8%	42.7%	6.5%	6.3%
3	64	16	40.3%	37.9%	6.2%	6.2%
4	128	32	35.7%	34.1%	6.1%	6.0%

6. *Q. Munib, M. Habeeb, B. Tahruri, and H. A. Al-Malik*, “American sign language (asl) recognition based on hough transform and neural networks,” *Expert Syst. Appl.* 32(1), 24–37 (2007).
7. *G. Castellano and M. B. Sandler*, “Handwritten digits recognition using hough transform and neural networks,” in *Circuits and Systems, 1996. ISCAS '96., Connecting the World., 1996 IEEE International Symposium on*, 3, 313–316 vol.3 (May 1996).
8. *J. Koh, K. Mehrotra, C. K. Mohan, and S. Ranka*, “Korean character recognition using neural networks,” tech. rep., Syracuse University, Electrical Engineering and Computer Science (05 1990).
9. *C. Jennings*, “Character recognition using the hough transform,” tech. rep., University of Calgary (March 1993).
10. *D. Nikolaev, S. Karpenko, I. Nikolaev, and P. Nikolayev*, “Hough transform: Underestimated tool in the computer vision field,” in *Proceedings of the 22th European Conference on Modelling and Simulation*, 238–246 (2008).
11. *CIFAR-10 dataset*, <https://www.cs.toronto.edu/kriz/cifar.html>.
12. *E. Kuznetsova, E. Shvets, and D. Nikolaev*, “Viola-Jones based hybrid framework for real-time object detection in multispectral images,” in *Proc. SPIE 9875, Eighth International Conference on Machine Vision, 98750N* (December 8, 2015), (2015).
13. *V. Vanhoucke, A. Senior, and M. Z. Mao*, “Improving the speed of neural networks on CPUs,” in *Deep Learning and Unsupervised Feature Learning Workshop, NIPS 2011*, (2011).
14. *R. Rigamonti, A. Sironi, V. Lepetit, and P. Fua*, “Learning separable filters,” in *Conference on Computer Vision and Pattern Recognition*, (2013).
15. *E. Limonova, D. Ilin, and D. Nikolaev*, “Improving neural network performance on simd architectures,” in *Eighth International Conference on Machine Vision, 98750L–98750L, International Society for Optics and Photonics* (2015).
16. *E. Limonova, A. Sheshkus, and D. Nikolaev*, “Computational optimization of convolutional neural networks using separated filters architecture,” *International Journal of Applied Engineering Research* 11(11), 7491–7494 (2016).

**Шешкус Александр Владимирович.** Научный сотрудник – программист ООО «Смарт Энджинс Сервис». В 2011 г. окончил МФТИ. Количество печатных работ: 9. Область научных интересов: обучение нейронных сетей, распознавание образов. E-mail: [astdcall@gmail.com](mailto:astdcall@gmail.com)

## Combining convolutional neural networks and hough transform for classification of images containing lines

*A. V. Sheshkus*

**Abstract.** In this paper, an expansion of convolutional neural network (CNN) input features based on Hough Transform is proposed. The idea of the approach is that morphological contrasted and Hough transformed image will be used as an input for some convolutional filters. Thus, CNNs computational complexity and the number of units are not affected. Morphological contrasting and Hough Transform are the only additional computational expenses of introduced CNN input features expansion. Proposed approach was demonstrated on the example of CNN with very simple structure using two image recognition problems: object classification on CIFAR-10 and printed character recognition on private dataset with symbols taken from Russian passports. Suggested approach allowed to reach noticeable accuracy improvement without taking much computational effort, which can be extremely important in industrial recognition systems or difficult problems utilizing CNNs, like pressure ridge analysis and classification.

**Keywords:** *convolutional neural networks, Hough Transform, feature extraction*

### References

1. P. Hough, "Method and means for recognizing complex patterns," (Dec 1962).
2. D. H. Ballard, "Readings in computer vision: Issues, problems, principles, and paradigms," ch. Generalizing the Hough Transform to Detect Arbitrary Shapes, 714–725, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1987).
3. T. Khanipov, I. Koptelov, A. Grigoryev, E. Kuznetsova, and D. Nikolaev, "Vision-based industrial automatic vehicle classifier," (2015).
4. N. Skoryukina, D. P. Nikolaev, A. Sheshkus, and D. Polevoy, "Real time rectangular document detection on mobile devices," (2015).
5. D. Krokhina, V. Blinov, S. Gladilin, I. Tarhanov, and V. Postnikov, "Fast roadway detection using car cabin video camera," (2015).
6. Q. Munib, M. Habeeb, B. Takruri, and H. A. Al-Malik, "American sign language (asl) recognition based on hough transform and neural networks.," Expert Syst. Appl. 32(1), 24–37 (2007).
7. G. Castellano and M. B. Sandler, "Handwritten digits recognition using hough transform and neural networks," in Circuits and Systems, 1996. ISCAS '96., Connecting the World., 1996 IEEE International Symposium on, 3, 313–316 vol.3 (May 1996).
8. J. Koh, K. Mehrotra, C. K. Mohan, and S. Ranka, "Korean character recognition using neural networks," tech. rep., Syracuse University, Electrical Engineering and Computer Science (05 1990).
9. C. Jennings, "Character recognition using the hough transform," tech. rep., University of Calgary (March 1993).
10. D. Nikolaev, S. Karpenko, I. Nikolaev, and P. Nikolayev, "Hough transform: Underestimated tool in the computer vision field," in Proceedings of the 22th European Conference on Modelling and Simulation, 238–246 (2008).
11. CIFAR-10 dataset, <https://www.cs.toronto.edu/kriz/cifar.html>.
12. E. Kuznetsova, E. Shvets, and D. Nikolaev, "Viola-Jones based hybrid framework for real-time object detection in multispectral images," in Proc. SPIE 9875, Eighth International Conference on Machine Vision, 98750N (December 8, 2015), (2015).
13. V. Vanhoucke, A. Senior, and M. Z. Mao, "Improving the speed of neural networks on CPUs," in Deep Learning and Unsupervised Feature Learning Workshop, NIPS 2011, (2011).
14. R. Rigamonti, A. Sironi, V. Lepetit, and P. Fua, "Learning separable filters," in Conference on Computer Vision and Pattern Recognition, (2013).
15. E. Limonova, D. Ilin, and D. Nikolaev, "Improving neural network performance on simd architectures," in Eighth International Conference on Machine Vision, 98750L–98750L, International Society for Optics and Photonics (2015).
16. E. Limonova, A. Sheshkus, and D. Nikolaev, "Computational optimization of convolutional neural networks using separated filters architecture," International Journal of Applied Engineering Research 11(11), 7491–7494 (2016).

**A. V. Sheshkus.** Researcher – programmer at Smart Engines Service LLC9. Graduated Moscow Institute of Physics and Technology (MIPT) in 2011. Number of publications: 9. Research interests: machine learning, pattern recognition. E-mail: [astdcall@gmail.com](mailto:astdcall@gmail.com)