

# Алгоритм классификации документов, удостоверяющих личность, на проективно-искаженных изображениях на основе обучаемой метрики подобия\*

А.Е. Лыначенко<sup>1</sup>, А.В. Шешкус<sup>1</sup>, В.Л. Арлазаров<sup>II</sup>

<sup>1</sup> ООО «Смарт Энджинс Сервис», Москва, Россия

<sup>II</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, Москва, Россия

**Аннотация.** Статья посвящена алгоритму классификации документов, удостоверяющих личность, на проективно-искаженных изображениях на основе метрики подобия, обученной с помощью сиамской архитектуры. Идея обучения сиамских сетей заключается в построении преобразования входного изображения в пространство, где с помощью заданной метрики аппроксимируется расстояние «схожести» между объектами исходного пространства. Во время обучения функция потерь стремится минимизировать расстояние между парами объектов одного класса и максимизировать – между парами разных классов. В качестве преобразования из исходного пространства в конечное используется сверточная нейронная сеть, с помощью которой для каждого класса определяется вектор признаков в конечном пространстве. Посредством данного преобразования к входному изображению и поиском ближайшего вектора признаков производится классификация объектов. На открытом наборе изображений документов, удостоверяющих личность, предложенный алгоритм показал качество распознавания сравнимое с качеством классифицирующей сверточной нейронной сети. Проиллюстрирована возможность обучения на одном примере (*one-shot learning*) данного метода.

**Ключевые слова:** распознавание образов, сиамские нейронные сети, сверточные нейронные сети, глубокое обучение.

**DOI:** 10.14357/20790279180519

## Введение

В современном мире глубокие сети становятся все более популярным методом решения различного рода задач. В качестве примеров можно привести распознавание речи [2], обработка естественного языка [3], компьютерное зрение [4], распознавание документов [5] и многие другие. Известной проблемой является необходимость в большом количестве данных для обучения. Аугментация данных и регуляризация [6] решают проблему с переобучением, но сети по-прежнему остаются не способными производить прогнозы на данных с небольшим объемом доступной информации.

С другой стороны, непараметрические модели (например, метод ближайшего соседа) избавлены от этих проблем. Такие системы позволяют быстро усвоить новую информацию и не требуют обучения и большого объема данных. Качество предсказания таких систем напрямую зависит от

выбранной метрики [5]. В [8] авторы приводят способ обучения метрики подобия с помощью сиамской архитектуры [7]. В [8] с помощью такого рода метрики и непараметрических алгоритмов производится распознавание образов на изображении. Именно о таком методе обучения и классификации пойдет речь в данной статье, применении его в задаче распознавания проективно-искаженных изображений документов, и возможности обучения на одном примере (*one-shot learning*).

## 1. Постановка задачи и метод ее решения

В данной статье рассматривается задача построения алгоритма классификации документов, удостоверяющих личность на изображении. Предлагаемый метод должен быть устойчив к проективным искажениям, так как подобные особенности являются характерными для задачи распознавания документов [11] (рис. 1). Также требуется, чтобы предлагаемый алгоритм не уступал в качестве распознавания современным методам, а именно клас-

\* Работа выполнена при финансовой поддержке РФФИ, грантов № 17-29-07092 и № 17-29-07093.



Рис. 1. Проективно-искаженные изображения документов из набора MIDV-500

сифицирующим сверточным нейронным сетям [4]. Кроме того, алгоритм должен обладать возможностью дообучения новых классов с использованием минимального объема информации.

## 2. Алгоритм классификации, основанный на метрике подобия

На рис. 2 представлена архитектура сямской сети, где  $X_1$  и  $X_2$  – это пары изображений, подаваемые на разные ветви. Пусть  $W = \{w_i\}, w_i \in R$  – это множество обучаемых параметров, общих для двух ветвей. В момент обучения функция  $F(W, X)$  стремится преобразовать  $X_1$  и  $X_2$  из исходного пространства в конечное, где с помощью метри-

ки  $L_2$  аппроксимируется метрика “подобия” объектов в исходном пространстве. Более формально это определяется, как выполнение следующих свойств: функция расстояния

$$E(W, X_1, X_2) = \|F(W, X_1) - F(W, X_2)\|_2 \quad (1)$$

мала, когда  $X_1$  и  $X_2$  принадлежат одному классу, и велика, когда  $X_1$  и  $X_2$  принадлежат разным классам.

Далее, после обучения весов  $W$ , для каждого класса  $C$  сохраняется соответствующий вектор признаков  $\bar{T}_C$  в конечном пространстве, определяемый как среднее по всем элементам класса из обучающей выборки:

$$\bar{T}_C = \frac{\sum_{X \in A_C} F(W, X)}{|A_C|}, \quad (2)$$

где  $A_C$  – это множество изображений обучающей выборки, относящиеся к классу  $C$ ,  $|A_C|$  – мощность множества  $A_C$ .

Распознавание нового изображения  $X$  происходит посредством приведения его в конечное пространство и поиском ближайшего вектора признаков из базы. Класс, которому соответствует ближайший вектор – результат распознавания:

$$C(W, X) = \operatorname{argmin}(\|F(W, X) - \bar{T}_c\|_2). \quad (3)$$

Преимуществами данного метода является возможность добавления в базу классов, не представленных в обучающей выборке.

## 3. Архитектура сверточной нейронной сети

В качестве функции, преобразующей изображения из исходного пространства в конечное, используется сверточная нейронная сеть. Данная

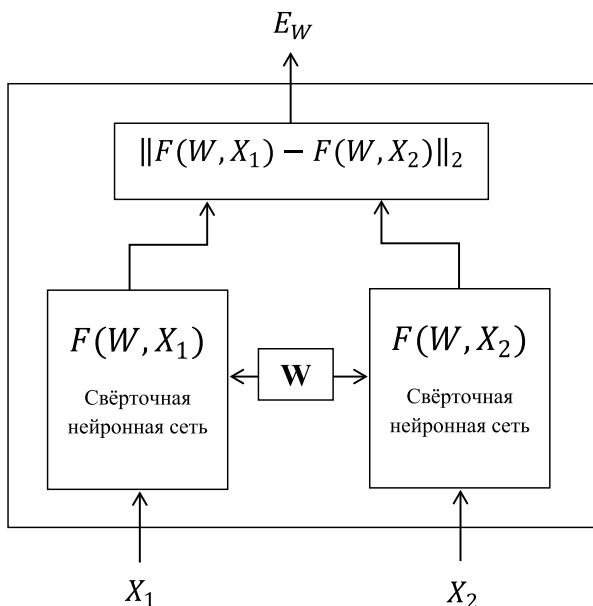


Рис. 2. Архитектура сямской сети

Табл. 1

Архитектура сверточной нейронной сети

Слой	Кол-во фильтров	Размер фильтра	Шаг	Отступ	Размер выходного вектора
Свертка + арктангенс	8	7 × 7	2 × 2	0 × 0	72 × 72 × 8
Свертка + арктангенс	8	5 × 5	2 × 2	2 × 2	36 × 36 × 8
Свертка + арктангенс	8	5 × 5	2 × 2	2 × 2	18 × 18 × 8
Свертка + арктангенс	4	5 × 5	2 × 2	2 × 2	9 × 9 × 4
Свертка + арктангенс	1	3 × 3	1 × 1	1 × 1	9 × 9 × 1
Полносвязный	-	-	-	-	50

сеть может быть представлена как обучаемая нелинейная система, выделяющая в конечном пространстве ключевые признаки на изображении [8]. Одним из главных свойств данной системы является возможность обучения детектора ключевых признаков, устойчивого к геометрическим искажениям исходного изображения [12]. В табл. 1. Представлена архитектура обучаемой сверточной нейронной сети.

#### 4. Данные для эксперимента

В качестве данных для обучения и тестирования алгоритма распознавания на основе сиамской архитектуры использовался открытый набор документов, удостоверяющих личность MIDV-500 [1]. В данных представлено 50 классов документов,

каждый из которых изображен на 300 снимках. На изображениях документов встречаются такие искажения, как размытия и блики. Для добавления проективных искажений, координаты одного из углов документа смещались на равномерно распределенное случайное расстояние, лежащее в интервале от 0 до 10%, 20% или 30% длины смежной стороны. Были отдельно получены изображения документов со сдвигами на 10%, 20%, 30% наружу и внутрь для проведения сравнительных экспериментов. Полученные изображения документов были вырезаны со снимков и переведены в оттенки серого с разрешением 150 на 150 пикселей.

На рис. 3 проиллюстрировано разделение изображений документов MIDV-500 на обучающие и тестовые выборки. Для обучения метрики подобия на основе сиамской архитектуры использовались 90% изображений 40 классов документов. Для тестирования использовались 10% изображений 40 классов документов, встречающихся в обучающей выборке и 100% изображений документов оставшихся 10 классов. Для обучения были сгенерированы 12500 тысяч пар документов одинаковых классов и 12500 пар – разных.

#### 5. Эксперимент

На 25000 пар сгенерированных для каждой проективно-искаженной выборки, были обучены метрики подобия с помощью сиамской архитектуры. Во время обучения функция потерь стремилась привести дистанцию между парами изображений одного класс к 0, а между разными – к 1. Для оценки качества сиамских сетей было вычислено количество правильных ответов на входные изображения на различных выборках. Где ответом считалось, что входные изображения принадлежат

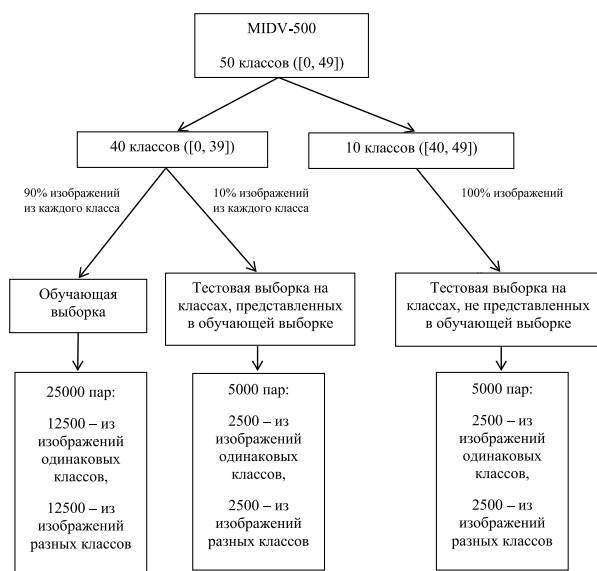


Рис. 3. Разделение изображений документов MIDV-500 на обучающие и тестовые выборки

**Табл. 2**

Процент правильных ответов сиамской сети для проективно-искаженных документов

Пары изображений из:	Допустимый сдвиг одного из углов документа						
	-0,3	-0,2	-0,1	0,0	0,1	0,2	0,3
Обучающей выборки	96.4%	99.2%	99.8%	99.9%	98.8%	98.6%	98.7%
Тестовой выборки на классах, представленных в обучении	95,8%	98.9%	99.7%	99.7%	98.9%	98.5%	98.3%
Тестовой выборки на классах, не представленных в обучении	90,5%	90,4%	88.9%	85.0%	93.0%	91.7%	92.5%

одному классу, если расстояния между ними меньше порогового значения  $P$  и разным, если больше. За пороговое значение было использовано среднее между метками классов,  $P = 0.5$ . В табл. 2 можно увидеть процент верных ответов сиамской сети на парах входных изображений на различных выборках.

По данным представленным в табл. 2 можно заметить, что сиамская архитектура научилась различать и находить общее среди изображений документов из классов, которые не встречались в обучающей выборке. Также по данным таблицы видно, что качество на парах изображений из классов документов, представленных в обучении, растет с уменьшением проективного искажения, а качество на классах не представленных – убывает. Данный эффект может свидетельствовать о переобучении метрики подобия (далее будет опровергнуто) или о не оптимальности выбора порогового значения  $P$ .

Далее, для построения классификатора на основе сверточной нейронной сети, обученной сиамской архитектурой, были вычислены вектора признаков по формуле (2). Таким образом, для каждого класса был получен соответствующий вектор в конечном пространстве, рассчитанный на изображениях из обучающей выборки. Далее, с помощью преобразования входного изображения в конечное пространство и поиском ближайшего вектора признаков (3), производилась классификация документа на изображении.

Для сравнения оценки качества классификации предложенного алгоритма, была обучена классифицирующая сверточная нейронная сеть на идентичной архитектуре. Для обучения данной сети использовался такой же процент исходных данных, что и для обучения метрики подобия:  $40/50 * 90\% = 72\%$ . Оставшиеся 28% использовались для оценки качества распознавания.

В табл. 3 отображено качество классификации изображений документов двумя предложенными способами. Тестирование производилось только на классах изображений, представленных в обучении.

Можно заметить, что качество классификации алгоритма, основанного на метрике подобия, отстает от классифицирующей сети, только на выборках с наибольшими проективными искажениями. На выборках с малыми геометрическими преобразованиями данный алгоритм не уступает в качестве.

Одним из преимуществ предложенного метода является возможность классификации объектов не встречающихся в обучении, с использованием небольшого объема информации о новых классах. В табл. 4 представлено качество распознавания изображений, с использованием фиксированного числа изображений для построения векторов признаков. Оценка качества проводилась на тестовой выборке проективно-искаженных изображений документов, классы которых не были представлены в обучении.

**Табл. 3**

Сравнение качества распознавания проективно-искаженных документов предложенным алгоритмом против классифицирующей сверточной нейронной сети

Классификатор	Допустимый сдвиг одного из углов документов						
	-0,3	-0,2	-0,1	0,0	0,1	0,2	0,3
Алгоритм на основе сиамской архитектуры	98.8%	98.9%	99.3%	99.5%	99.1%	99.0%	99.2%
Классифицирующая сверточная нейронная сеть	88.6%	99.2%	99.5%	99.8%	99.5%	99.0%	98.5%

Табл. 4

Зависимость качества распознавания объектов из классов, не представленных в обучающей выборке, от количества изображений, используемых для построения вектора признаков класса

Кол-во изображений	Допустимый сдвиг одного из углов документа						
	-0.3	-0.2	-0.1	0	0.1	0.2	0.3
1	70.34%	84.18%	95.11%	98.66%	89.37%	97.57%	92.05%
2	87.98%	89.59%	97.45%	98.88%	97.60%	98.31%	95.58%
3	79.44%	90.83%	96.73%	99.06%	98.16%	97.25%	95.41%
4	86.08%	90.49%	96.79%	99.17%	98.49%	98.30%	96.49%
5	85.87%	86.82%	98.18%	99.39%	98.33%	98.60%	97.50%
6	86.73%	92.21%	97.98%	99.20%	98.67%	98.74%	97.87%
7	90.95%	93.13%	97.33%	99.12%	98.51%	98.39%	97.60%
8	89.58%	91.38%	96.97%	99.23%	98.85%	98.04%	97.55%
9	88.31%	93.15%	98.23%	99.35%	98.81%	98.30%	97.85%
10	89.46%	94.17%	96.60%	99.34%	98.84%	97.91%	97.88%
11	89.38%	92.95%	98.10%	99.34%	98.91%	97.86%	97.98%
12	90.39%	93.04%	97.31%	99.26%	98.75%	98.09%	97.78%
13	87.85%	93.09%	97.66%	99.06%	98.63%	98.36%	97.66%
14	88.00%	93.22%	97.80%	99.14%	98.75%	98.47%	98.00%

Табл. 4 показывает, что предложенный метод может быть использован для классификации новых сущностей с использованием минимального количества информации. Видно, что качество распознавания уменьшается с увеличением проективного искажения изображений документов, что исключает гипотезу переобучения на изображениях с малыми искажениями. Также из табл. 4 можно заметить, что не каждое добавление информации о новом классе приводит к положительному результату. Это говорит о разбросе векторов в конечном пространстве и о важности выбора данных для построения векторов признаков.

### Заключение

В данной статье предложен алгоритм классификации проективно-искаженных изображений документов. Данный метод был построен на метрике подобия, обученной с помощью сиамской архитектуры. Предложенный алгоритм показал качество распознавания сравнимое с классифицирующей сверточной нейронной сетью. Также была проиллюстрирована возможность классификации объектов новых классов, без переобучения метрики подобия и с использованием минимального объема информации о новых документах.

### Литература

1. *Arlazarov V.V., Bulatov K., Chernov T. and Arlazarov V.L.* MIDV-500: A Dataset for Identity Documents.
2. *Hinton G., Deng L. and Yu D.* Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups, *IEEE Signal Processing Magazine*, pp. 82 – 97. 18 10 2012.
3. *Collobert R. and Weston J.* A unified architecture for natural language processing: deep neural networks with multitask learning, in *Proceedings of the 25th international conference on Machine learning*, Helsinki, Finland, 2008, DOI: 10.1145/1390156.1390177.
4. *Krizhevsky A., Sutskever I. and Hinton G.* ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems 25*, pp. 1097-1105, 2012.
5. *Bulatov K., Arlazarov V.V. and Chernov T.* Smart IDReader: Document Recognition in Video Stream, *International Conference on Document Analysis and Recognition (ICDAR)*, 9 11 2017.
6. *Dubnov I., Merkov A., Arlazarov V.L. and Nikolaev I.* Evidence Maximization Technique for Training of Elastic Nets, *Journal of Optimization*, 2016.

7. Goldberger J., Roweis S., Hinton G. and Salakhutdinov R. Neighbourhood Components Analysis, in *Neural Information Processing Systems (NIPS)*, Montreal, Canada, 2014.
8. Chopra S., Hadsell R. and LeCun Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification, in *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.
9. Bromley J., Guyon I., LeCun Y. and Shah R. Signature verification using a Siamese time delay neural network, in *Advances in neural information processing systems (NIPS)*, 1993.
10. Vinyals O., Blundell C., Lillicrap T., Kavukcuoglu K. and Wierstra D. Matching Networks for One Shot Learning, *Advances in Neural Information Processing Systems 29*, pp. 3630-3638, 2016.
11. Полевой Д., Шешкус А., Булатов К., Скорюкина Н., Чернов Т. and Арлазаров В.В. Ключевые аспекты распознавания документов с использованием малоразмерных цифровых камер, *Вестник ПФФИ*, С. 97-108, 2016.
12. Lawrence S., Giles L. and Back A. Face recognition: a convolutional neural-network approach, *IEEE Transactions on Neural Networks*, pp. 98 – 113, 1 1 1997.

**Лынченко Александр Евгеньевич.** ООО «Смарт Энджинс Сервис», г. Москва, Россия. Младший научный сотрудник-программист. Количество печатных работ: 1. Область научных интересов: компьютерное зрение, машинное обучение. E-mail: alynchenko@smartengines.biz

**Шешкус Александр Владимирович.** ООО «Смарт Энджинс Сервис», г. Москва, Россия. Старший научный сотрудник-программист. Количество печатных работ: 18. Область научных интересов: глубокое обучение, компьютерное зрение, проективно-инвариантная сегментация изображений. E-mail: astdcall@gmail.com

**Арлазаров Владимир Львович.** Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия. Заведующий отделением, Чл.-корр. РАН, профессор. Количество печатных работ: более 100 статей и монографий. Область научных интересов: теория графов, распознавания образов, программирование. E-mail: arl@isa.ru

## Identity document classification algorithm based on similarity metric robust to projective distortions

A.E. Lynchenko<sup>1</sup>, A.V. Sheshkus<sup>1</sup>, V.L. Arlazarov<sup>1,II</sup>

<sup>1</sup> Smart Engines Limited, Moscow, Russia

<sup>II</sup> Institute for Systems Analysis, Federal Research Center Computer Science and Control of Russian Academy of Sciences, Moscow, Russia

**Abstract.** This article presents an algorithm for document image recognition robust to projective distortions and with possibility of one-shot learning. This method is based on a similarity metric, which is learned using Siamese architecture. The proposed algorithm achieved recognition quality comparable to classifying convolutional network on an open dataset of identity document images MIDV-500.

**Keywords:** *pattern recognition, siamese neural network, convolutional neural network, deep learning.*

**DOI:** 10.14357/20790279180519

### References

1. V.V. Arlazarov, K. Bulatov, T. Chernov and V.L. Arlazarov. "MIDV-500: A Dataset for Identity Documents".
2. G. Hinton, L. Deng and D. Yu. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," IEEE Signal Processing Magazine, pp. 82 – 97, DOI: 10.1109/MSP.2012.2205597, 18 10 2012.
3. R. Collobert and J. Weston. "A unified architecture for natural language processing: deep neural networks with multitask learning," in Proceedings of the 25th international conference on Machine learning, Helsinki, Finland, 2008, DOI: 10.1145/1390156.1390177.
4. A. Krizhevsky, I. Sutskever and G. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems 25, pp. 1097-1105, 2012.
5. K. Bulatov, V.V. Arlazarov and T. Chernov. "Smart IDReader: Document Recognition in Video Stream," International Conference on Document Analysis and Recognition (ICDAR), 9 11 2017, DOI: 10.1109/ICDAR.2017.347.
6. I. Dubnov, A. Merkov, V.L. Arlazarov and I. Nikolaev. "Evidence Maximization Technique for Training of Elastic Nets," Journal of Optimization, 2016.
7. J. Goldberger, S. Roweis, G. Hinton and R. Salakhutdinov. "Neighbourhood Components Analysis," in Neural Information Processing Systems (NIPS), Montreal, Canada, 2014.
8. S. Chopra, R. Hadsell and Y. LeCun. "Learning a Similarity Metric Discriminatively, with Application to Face Verification," in Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, DOI: 10.1109/CVPR.2005.202.
9. J. Bromley, I. Guyon, Y. LeCun and R. Shah. "Signature verification using a Siamese time delay neural network," in Advances in neural information processing systems (NIPS), 1993.
10. O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu and D. Wierstra. "Matching Networks for One Shot Learning," Advances in Neural Information Processing Systems 29, pp. 3630-3638, 2016.
11. D. Polevoy, A. Sheshkus, K. Bulatov, N. Skoryukina, T. Chernov and V. V. Arlazarov. "Key Aspects of Document Recognition Using Small Digital Cameras," Vestnik RFBR, pp. 97-108, 2016, DOI: 10.22204/2410-4639-2016-092-04-97-108.
12. S. Lawrence, L. Giles and A. Back. "Face recognition: a convolutional neural-network approach," IEEE Transactions on Neural Networks, pp. 98 – 113, 1 1 1997, DOI: 10.1109/72.554195.

**A.E. Lynchenko.** LLC "Smart Engines Service", Moscow, Russia. Researcher. PhD candidate.

E-mail: alynchenko@smartengines.biz

**A.V. Sheshkus.** LLC "Smart Engines Service", Moscow, Russia. Researcher. PhD candidate.

E-mail: astdcall@gmail.com

**V.L. Arlazarov.** Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia. Head of the department, Corresponding member. RAS, Professor. Number of publications: great than 100 papers and books. Scientific interests: artificial intelligence, recognition systems, information technology, programming. E-mail: arl@isa.ru