

О важности аугментации данных при машинном обучении в задачах обработки изображений в условиях дефицита данных

Н.А. Андриянов^{1,2}, Д.А. Андриянов²

¹АО «НПП «Исток» им. Шокина», Вокзальная 2а, к.1, Фрязино, Россия, 141190

²Ульяновский государственный технический университет, Северный Венец 32, Ульяновск, Россия, 432027

Аннотация. В статье приведены результаты исследования эффективности работы различных нейронных сетей в ограниченных условиях исходных данных и при ряде простейших аугментаций. При этом зависимости получены для последовательной нейронной сети с обратным распространением ошибки. Для аугментации данных использовались простейшие преобразования, в числе которых наклон букв (курсив), замена цвета букв (с черного на красный), а также искажение эталонных изображений белым гауссовым шумом при отношении сигнал/шум $q = 1...10$. Показано, что наилучшие результаты распознавания букв русского алфавита обеспечивает сеть, для которой были использованы все рассмотренные в работе методы аугментации. Также проведено исследование зависимости точности распознавания от отношения сигнал/шум во всех обученных нейронных сетях.

1. Введение

В настоящее время обработка изображений становится особо популярной в различных системах технического зрения. При этом условно методы обработки изображений можно разделить на 3 группы. К первой группе относятся такие алгоритмы обработки, которые базируются на описании изображений с помощью математических моделей случайных полей [1-5]. Вторая группа включает в себя локальные операторы обработки, хорошо зарекомендовавшие себя для решения узкоспециализированных задач [6-8]. Наконец, все большую популярность приобретают методы 3 группы, связанные с применением машинного обучения и, в последние годы, глубокого обучения [9-16]. В данной работе подробнее рассмотрим последнюю группу с точки зрения влияния объема имеющихся данных на эффективность их работы.

Сегодня существует много областей исследования, в которых ученые надеются улучшить текущие результаты, применяя глубокие сверточные сети для задач компьютерного зрения. Улучшение обобщающей способности таких моделей является одной из самых сложных задач. Обобщаемость относится к разнице в производительности модели при оценке по ранее просмотренным данным (обучающим данным) по сравнению с данными, которые она никогда не видела ранее (данные тестирования). Модели с плохой обобщаемостью показывают значительно большую точность на выборке для обучения, чем на тестовой выборке. В этом заключается эффект переобучения. Поэтому, как правило, в каждую эпоху обучения

необходимо соотносить точность (или эффективность) на выборке для обучения с точностью (или эффективностью) на проверочной выборке. Для построения полезных моделей глубокого обучения ошибка на проверочных данных должна продолжать уменьшаться вместе с ошибкой на обучающей выборке. Увеличение данных – очень мощный метод достижения этого. Расширенные данные будут представлять более полный набор возможных точек данных, таким образом минимизируя расстояние между обучающим и проверочным набором, а также любыми будущими наборами для тестирования.

Чтобы уменьшить эффект переобучения и повысить обобщаемость необходимо увеличить объем данных. Если увеличение объема реальных данных затруднительно, то в ход вступают различные методики раздутия данных на основе имеющихся [15,16]. Многие другие стратегии повышения производительности обобщения сосредоточены на самой архитектуре модели. Это привело к последовательности прогрессивно более сложных архитектур от AlexNet [9] до VGG-16 [10], ResNet [11], Inception-V3 [12] и DenseNet [13]. Были разработаны функциональные решения, такие как регуляризация отсева, нормализация выборки, обучение по переводу и предварительная подготовка, чтобы попытаться расширить глубокое обучение для применения на небольших наборах данных. Полный обзор методов регуляризации в глубоком обучении был составлен J. Kuckaska и соавторами [14]. Однако цель данной работы состоит в том, чтобы показать, какое полезное влияние может оказать аугментация данных даже при решении классической задачи распознавания букв алфавита.

2. Подготовка базовой выборки для обучения

В качестве данных для распознавания был выбран русский алфавит, содержащий 33 буквы. Базовая выборка была подготовлена в редакторе Paint и состоит из 33 файлов формата «.png», каждый из которых представляет конкретную букву от А до Я. Все буквы набраны шрифтом Times New Roman с размером 26 пт, а размер всего изображения 30x30 пикселей. Примеры некоторых полученных изображений букв представлены на рисунке 1.

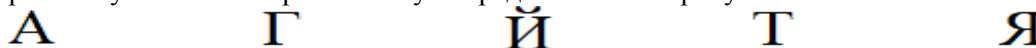


Рисунок 1. Базовые данные для обучения.

Из рисунка 1 видно, что представленный набор букв А, Г, Й, Т, Я представляется набором, для которого можно выделить некоторые характеристики по каждой букве. Например, для «А» треугольник в верхней части, для «Й» – дуга над буквой, для «Я» – полуовал.

Представим, что возможна ситуация, когда сеть при обучении видела только чёрные буквы, а затем в процессе её работы подавались бы красные буквы. К сожалению, нельзя быть уверенным, что в этом случае распознавание сработает верно. Такое предположение делается для того, чтобы сделать для нашей нейронной сети важным также и значение цвета. Конечно, можно было бы просто переводить все изображения в изображения с 1 цветовым каналом, однако размеры обрабатываемых изображений позволяют исследовать работу сети и при наличии 3 цветовых каналов. Следовательно, при обучении мы должны использовать все цветовые каналы изображения. Таким образом, одну букву будет описывать 3 матрицы размером 30x30 элементов, которые представим в виде вектора длиной $N=3 \times 30 \times 30=2700$.

3. Аугментация данных

Предложенная выборка может быть легко обучена, но с большой долей вероятности она не сможет эффективно распознавать на полном наборе букв ни цветные буквы, ни наклонные буквы, ни искаженные буквы. Это лишь ряд направлений, в которых возможна аугментация базовой выборки. Сделаем такую же выборку, но состоящую из букв, написанных курсивом. Примеры для тех же букв А, Г, Й, Т, Я представлены на рисунке 2.



Рисунок 2. Алфавит со шрифтом типа «Курсив».

Следующий вариант для аугментации данных – это применение разных цветов к имеющейся базовой выборке. Для простоты будем считать, что кроме черных букв, могут появляться также и красные. На рисунке 3 представлено применение цветового преобразования к базовой выборке для букв А, Г, Й, Т, Я.



Рисунок 3. Алфавит в красном цвете.

Наконец, при передаче изображений или во время их фоторегистрации могут возникать различные помехи. В связи с этим добавим метод аугментации, который представляет собой непосредственную смесь имеющихся изображений из базовой выборки и белого гауссовского шума. На рисунке 4 представлен вариант полученной выборки для отношения сигнал шум $q=1$, на рисунке 5 – для $q=10$ для букв А, Г, Й, Т, Я с инверсией цвета.



Рисунок 4. Алфавит, искаженный шумом, сопоставимым с сигналом.



Рисунок 5. Алфавит, искаженный шумом, в десять раз меньшим сигнала.

Следует отметить, что это лишь незначительная часть преобразований, которые позволили провести аугментацию данных. Например, можно было бы использовать множество поворотов, масштабирование и обрезание данных. Но для дальнейших исследований ограничимся описанными выше аугментациями.

4. Структура нейронных сетей и обучающих выборок

После того, как были подготовлены базовые данные и выполнена их аугментация, будем обучать 5 нейронных сетей, имеющих одинаковую структуру, представленную на рисунке 6.

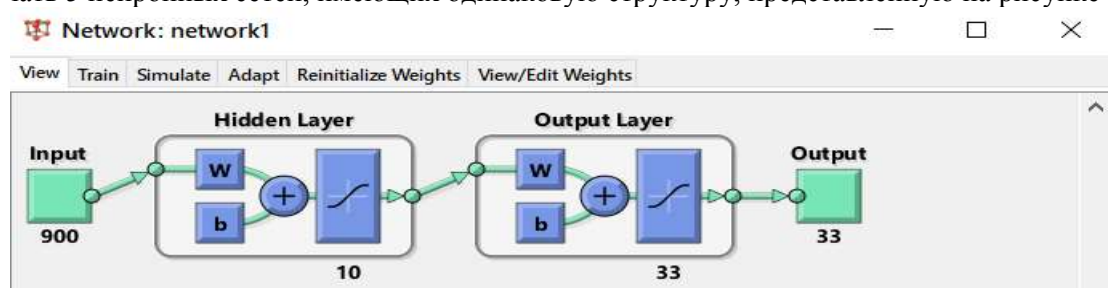


Рисунок 6. Структура нейронной сети.

Как видно из рисунка 6, на вход подается 900 отсчетов, описывающих одну букву, затем присутствует один скрытый слой, состоящий из 10 нейронов. На выходе вектор из 33 значений, показывающий близость к той или иной букве.

Подготовим 5 нейронных сетей следующим образом в зависимости от обучающей выборки. Условное название сети и соответствующая ей обучающая выборка представлены в таблице 1.

Рисунок 7 показывает процессы обучения нейронных сетей из таблицы 1. Рисунок 7а соответствует Сети №1, рисунок 7б – Сети №2, рисунок 7в – Сети №3, рисунок 7г – Сети №4,

рисунок 7д – Сети №5. Зеленая линия показывает ошибки на обучающей выборке, синяя – на проверочной, а красная – на тестовой выборке.

Таблица 1. Состав обучающей выборки для различных нейросетей.

Нейросеть	Данные для обучающей выборки
NetworkBase (Сеть №1)	Только базовый алфавит (33 буквы)
NetworkItalic (Сеть №2)	Только алфавит курсивом (33 буквы)
NetworkColor (Сеть №3)	Базовый алфавит + «красный» алфавит (66 букв)
NetworkBIC (Сеть №4)	Базовый алфавит + «красный» алфавит + алфавит курсивом (99 букв)
NetworkBICN (Сеть №5)	Базовый алфавит + «красный» алфавит + алфавит курсивом + зашумленный алфавит ($q=1:10$) (2089 букв)

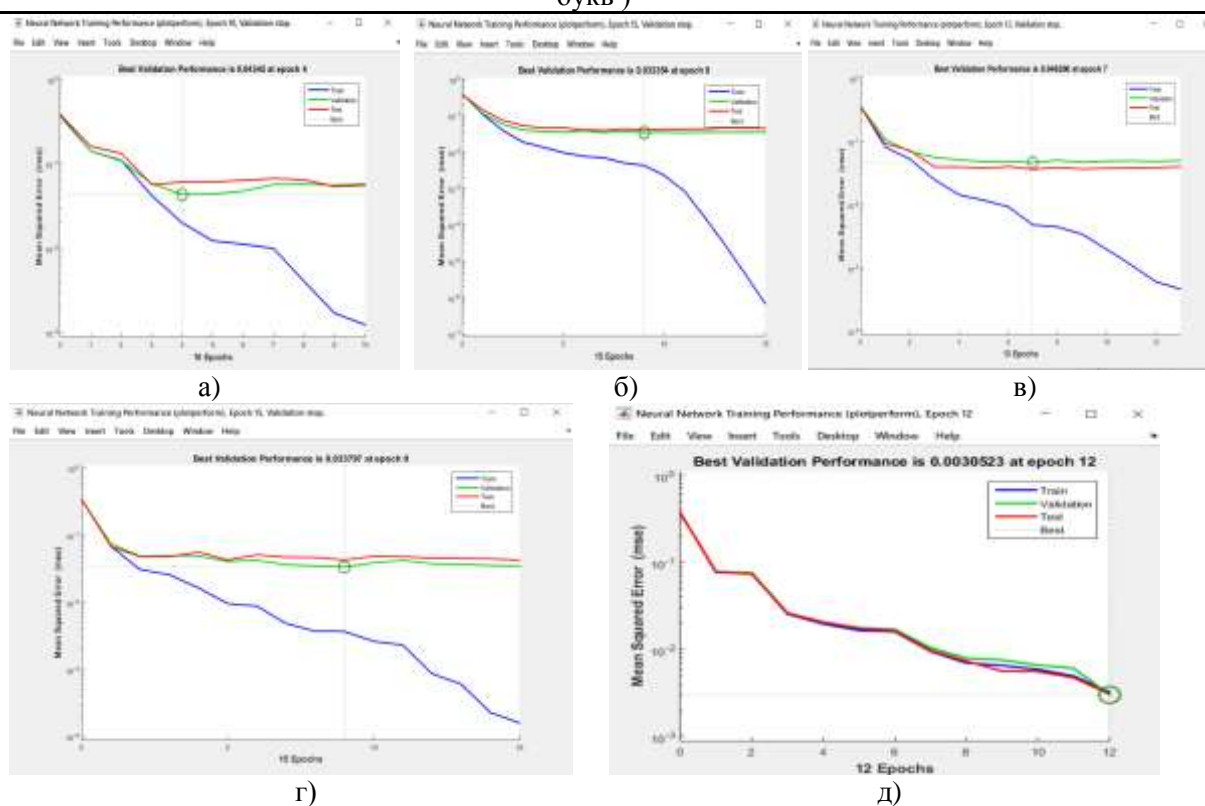


Рисунок 7. Обучение нейронных сетей.

Из графиков рисунка 7 видно, что ошибки близки для всех выборок только у Сети №5, что представлено кривыми на рисунке 7д. Для других же сетей наблюдается значительное расхождение ошибок для обучающей и других выборок.

5. Результаты распознавания

Все обученные нейронные сети из таблицы 1 были протестированы на данных, включающих в себя изображения из базового алфавита и каждого аугментированного алфавита. Результаты распознавания представлены на рисунке 8. По оси абсцисс отложена дисперсия шума (с учетом единичной дисперсии сигнала), а по оси ординат – количество точно распознанных символов алфавита в %. Линии 1 – 5 соответствуют Сетям №1 – №5.

Анализ полученных графиков показывает, что на результаты распознавания на тестовой выборке значительно влияет объем и структура обучающей выборки. Наилучшие результаты обеспечивают Сети №4 и №5, в которых при обучении использовано наибольшее количество аугментаций.

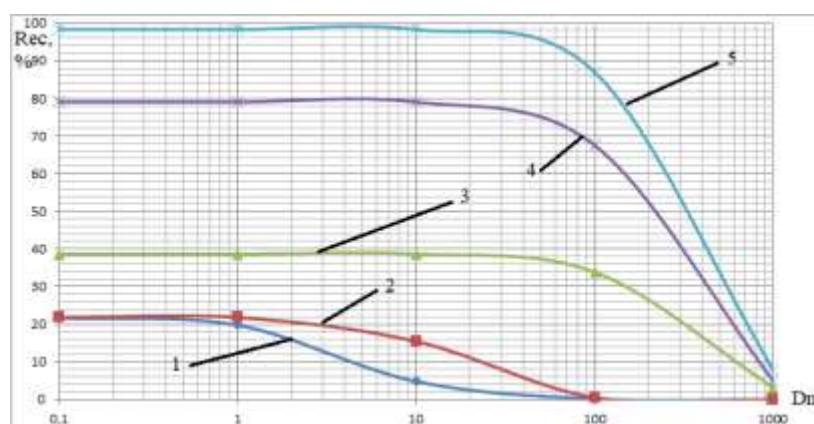


Рисунок 8. Сравнение точности распознавания для различных нейросетей.

6. Заключение

На простейшем примере распознавания символов русского алфавита показано, как могут влиять различные преобразования данных на эффективность работы нейронной сети. В частности из 33 букв было сгенерировано несколько аугментированных алфавитов. Проведенное исследование для 3 вариантов аугментаций и их смесей показало, что удалось повысить точность распознавания до 5 раз по сравнению с сетью, основывающейся только на базовом наборе данных. Таким образом, можно сделать вывод о том, что в условиях недостаточности данных одним из вариантов выхода может быть применение их аугментации.

7. Благодарности

Работа выполнена при поддержке Гранта РФФИ, Проект № 19-47-730011.

8. Литература

- [1] Smith, J. Physical-model-based image processing for feature aided analysis / J. Smith, E. Best, E. Sum, Y. Guzel, M.A. Saville, L. LoMonte, M. Wicks // International Conference on Electromagnetics in Advanced Applications (ICEAA), 2015. – P. 1-8. DOI: 10.1109/iceaa.2015.7297179
- [2] Andriyanov, N.A. Anomalies detection on spatially inhomogeneous polyzonal images / N.A. Andriyanov, K.K. Vasiliev, V.E. Dementiev // CEUR Workshop Proceedings. – 2017. – Vol. 1901. – P. 10-15.
- [3] Andriyanov, N.A. Image Models and Segmentation Algorithms Based on Discrete Doubly Stochastic Autoregressions with Multiple Roots of Characteristic Equations / N.A. Andriyanov, Yu.N. Gavrilina // CEUR Workshop Proceedings. – 2018. – Vol. 2076. – P. 19-29.
- [4] Vasiliev, K.K. Representation and processing of multispectral satellite images and sequences / K.K. Vasiliev, V.E. Dementiev, N.A. Andriyanov // Procedia Computer Science. – 2018. – Vol. 126. – P. 49-58. DOI: doi.org/10.1016/j.procs.2018.07.208.
- [5] Andriyanov, N.A. Optimal filtering of multidimensional random fields generated by autoregressions with multiple roots of characteristic equations / N.A. Andriyanov, K.K. Vasiliev // CEUR Workshop Proceedings. – 2019. – Vol. 2391. – P. 72-78. DOI: 10.18287/1613-0073-2019-2391-72-78.
- [6] Арефьев, Е.Ю. Обработка изображений с помощью процессоров табличных преобразований / Е.Ю. Арефьев, А.В. Проскурин // Компьютерная оптика. – 1990. – № 7. – С. 97-100.
- [7] Мясников, В.В. Эффективные алгоритмы вычисления локального дискретного вейвлет-преобразования // Компьютерная оптика. – 2007. – Т. 31, № 4. – С. 86-94.
- [8] Мясников, В.В. Сплайны как средство построения эффективных алгоритмов локального линейного преобразования // Компьютерная оптика. – 2007. – Т. 31, № 2. – С. 52-68.

- [9] Krizhevsky, A. ImageNet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // *Adv Neural Inf Process Syst.* – 2012. – Vol. 25. – P. 1106-1114.
- [10] Simonyan, K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // *arXiv e-prints*, 2014 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1409.1556> (15.11.2019).
- [11] Kaiming, H. Deep residual learning for image recognition / H. Kaiming, Z. Xiangyu, R. Shaoqing, S. Jian // *arXiv e-prints*, 2016 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1512.03385> (15.11.2019).
- [12] Szegedy, C. Rethinking the inception architecture for computer vision / C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna // *arXiv e-prints*, 2015 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1512.00567> (15.11.2019).
- [13] Huang, G. Densely connected convolutional networks / G. Huang, L. Zhuang, L. van der Maaten, K. Q. Weinberger // *arXiv preprint*, 2016 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1608.06993> (15.11.2019).
- [14] Kukacka, J. Regularization for deep learning: a taxonomy / J. Kukacka, V. Golkov, D. Cremers // *arXiv preprint*, 2017 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1710.10686> (15.11.2019).
- [15] Buslaev, A. Alumentations: fast and flexible image augmentations / A. Buslaev, A. Parinov, E. Khvedchenya, V. Iglovikov, A. Kalinin // *arXiv preprint*, 2018 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1809.06839> (15.11.2019).
- [16] Акимов, А.В. Модели и алгоритмы искусственного размножения данных для обучения алгоритмов распознавания лиц методом Виолы–Джонса / А.В. Акимов, А.А. Сирота // *Компьютерная оптика.* – 2016. – Т. 40, № 6. – С. 911-918. DOI: 10.18287/2412-6179-2016-40-6-911-918.

The importance of data augmentation in machine learning for image processing tasks in the face of data scarcity

N.A. Andriyanov^{1,2}, D.A. Andriyanov²

¹JSC "RPC "Istok" named after Shokin", Vokzalnaya street 2a, b.1, Fryazino, Russia, 141190

²Ulyanovsk State Technical University, Severny Venets street 32, Ulyanovsk, Russia, 432027

Abstract. The article presents the results of alphabet character recognition by various neural networks in the limited conditions of the source data and with a number of simple augmentations. Furthermore the dependences were obtained for a serial neural network with back propagation of error. The simplest transformations were used for data augmentation. Augmentation process includes the slope of the characters (italics), changing the colour of the characters (from black to red), as well as distortion of the reference images with white Gaussian noise at a signal-to-noise ratio $q = 1 \dots 10$. It is shown that the best results of recognition of the Russian alphabet characters are provided by a network for which all the augmentation methods discussed in this work were used. A study was also conducted of the dependence of recognition accuracy on the signal-to-noise ratio in all trained neural networks.