Методы глубокого обучения в задаче обнаружения искажений данных дистанционного зондирования Земли

А.В. Кузнецов[1,2]

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086
[2]Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация. Подделка цифровых изображений является известной проблемой из-за роста доступности технологий и программного обеспечения, которые позволяют с легкостью создавать искаженные изображения. В целях противодействия таким атакам было разработано несколько подходов для обнаружения подделок. Особое значение имеют методы обнаружения искажений отдельного типа цифровых изображений – данных дистанционного зондирования Земли, которые могут быть использованы в целях обеспечения безопасности охраняемых территорий, мониторинга состояния окружающей среды и т.д. В этой статье предлагается новая схема, основанная на нейронных сетях и глубоком обучении, в основе которой лежит применение новой архитектуры свёрточной нейронной сети (CNN) для улучшения качества обнаружения наиболее распространенного вида атак на цифровые изображения – встраивание дубликатов. В рамках предлагаемой архитектуры применяются дополнительные слои предобработки для улучшения качества обнаружения. Данный подход также демонстрирует инвариантность к вносимым в дублируемую область искажениям. Эксперименты показывают, что предложенное решение превосходит известные алгоритмы обнаружения дубликатов – значение метрики F1 достигает 0.77.

1. Введение

С течением времени редактирование цифровых изображений становится все менее сложным в результате роста доступности широкого спектра инструментов для обработки цифровых изображений. Подделка изображений, которая определяется как «процесс вырезания и вставки областей в одном и том же или в различных изображениях» [1], является одной из самых популярных форм обработки. Технология обнаружения дубликатов [2] может быть использована в качестве способа оценки подлинности изображения. Это делается путем обнаружения «следов встраивания», которые обычно обнаруживаются на искаженных изображениях.

В области проверки подлинности цифровых изображений методы обнаружения дубликатов обычно делятся на две категории: на основе ключевых точек и на основе блочного анализа [2]. В данной работе речь пойдет об алгоритме, реализуемом в рамках второго подхода, который заключается в анализе изображения в режиме скользящего окна с перекрытиями. Среди решений, используемых в настоящее время для обнаружения подделок изображений, в качестве локальных признаков областей применяются статистические характеристики [3]. С другой стороны, существующие решения делятся на два типа: с локализацией искажений и без. В

результате работы алгоритмов первого типа формируется маска искаженных областей изображений, для второго типа характерен бинарный ответ, подвергалось ли изображение обработке злоумышленником или нет.

В задаче локализации искажений в настоящее время большой упор делается именно на выбор/разработку модели нейронной сети. Наиболее часто применяемые типы нейронных сетей включают в себя глубокие сети доверия [4], глубокие автоэнкодеры [5] и свёрточные нейронные сети (CNN) [6]. Последний тип чаще всего используются в задачах компьютерного зрения. Такие сети обеспечивают превосходную производительность [7, 8] в задаче обнаружения подделок цифровых изображений, благодаря использованию упрощенных операций нелинейной и линейной фильтрации (например, свертка) [9].

В данной работе предлагается модель свёрточной нейронной сети, которая позволит обнаруживать и локализовать подделки цифровых изображений данных ДЗЗ. Схема предлагаемого метода представлена на рисунке 1. Для этого подхода в модель нейронной сети включены разномасштабные скользящие окна с целью создания набора карт признаков, с использованием которых производится локализация.

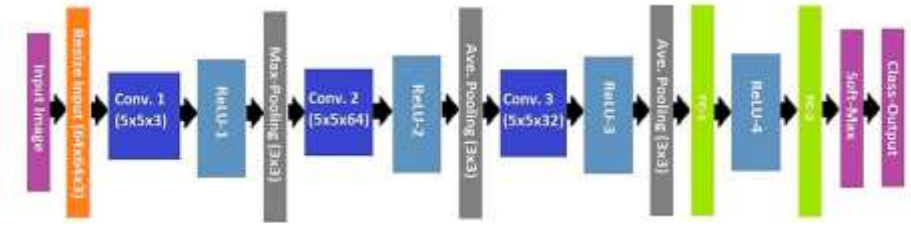

Рисунок 1. Используемая модель свёрточной нейронной сети.

Оставшаяся часть теста организована следующим образом. В разделе 2 мы представляем обзор существующих решений в области использования нейронных сетей и аргументируем выбор модели. В разделе 3 приводится краткое описание предлагаемой модели и описан процесс обучения. В разделе 4 демонстрируются результаты экспериментов. Далее следуют выводы и заключение.

2. Обзор существующих решений

В настоящее время существует ряд решений, направленных на обнаружения искусственных искажений цифровых изображений. В [2, 10] авторами предлагается модель для обнаружения сплайсинга. В другом подходе [10] на входной слой подавались гистограммы коэффициентов дискретного косинусного преобразования (ДКП) для локализации встраиваний, отличающихся свойствами сжатия JPEG. В работе [11] такие признаки строились для набора фрагментов изображения и весь набор подавался на вход нейронной сети.

В то время как методы глубокого обучения стали применяться в задачах компьютерного зрения, алгоритмы обнаружения дубликатов в основном основывались на классических подходах, таких как поиск изображений, классификация и обнаружение объектов. В отличие от традиционных схем классификации изображений, которые в основном построены на вычислении локальных признаков [12], современные алгоритмы классификации изображений на основе CNN предлагают сквозную структуру, то есть вся логика локальной обработки изображения заложена в архитектуре CNN. Наиболее актуальные архитектуры CNN (например, VGG [13]) значительно повышают производительность в задачах детектирования объектов и классификации изображений [14]. Среди наиболее известных архитектур AlexNet, ResNet, GoogleNet и т.д. Следует отметить, что существенная часть этих архитектур содержит большое число свёрточных слоев и, как следствие, параметров, что усложняет обучение и может привести к переобучению сети на небольших выборках (для задачи обнаружения дубликатов достаточный объем выборок может быть сформирован только посредством аугментации). Поэтому мы предлагаем небольшую сеть, содержащую порядка 200 тыс. параметров

В работах последних нескольких лет представлен ряд глубоких сетей, которые демонстрируют высокое качество классификации патчей и новые схемы сопоставления

векторов признаков для поиска ближайших [15]. Ввиду широкого распространения CNN в задачах обработки цифровых изображений они безусловно могут быть успешно применены в задаче поиска искажений цифровых изображений (например, дубликатов). Более того использование аппарата свёрточных нейронных сетей позволит снизить вычислительную сложность задачи обнаружения искажений данных ДЗЗ, которые отличаются большей размерностью от обычных изображений. В данной работе мы предлагаем использовать CNN как средство вычисления локальных признаков в режиме скользящего окна или окна с перекрытиями. В дальнейшем планируется применять специальный вид сверточных нейронных сетей (Scale Invariant CNN [16]), который позволит вычислять мультимасштабные локальные признаки для повышения устойчивости механизма обнаружения дубликатов к вносимым в них искажениям.

3. Предлагаемое решение

В данной работе предлагается использовать свёрточную нейронную сеть для извлечения признаков локальных областей в режиме скользящего окна с перекрытиями как на этапе обучения, так и на этапе тестирования (рисунок 2).

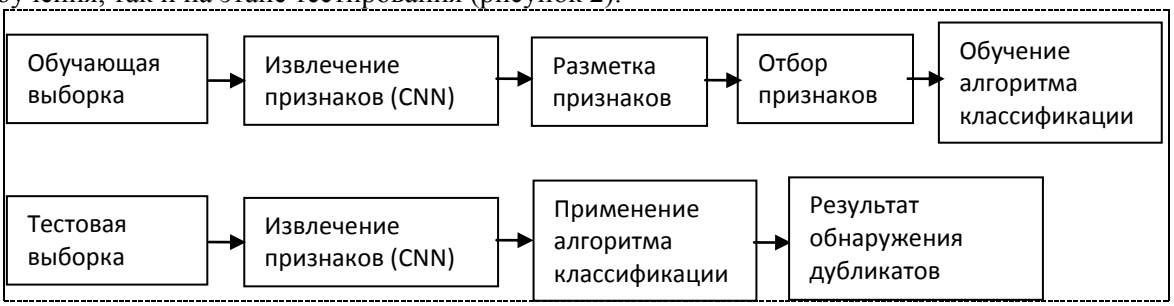

Рисунок 2. Схемы обучения и тестирования предлагаемого решения.

Архитектура предлагаемой сети, изображённая на рисунке 1, содержит 3 блока слоёв. Каждый из блоков состоит из последовательности свёрточного слоя, активационного ReLU и слоя пулинга. Далее в архитектуре сети следуют полносвязные слои, Softmax и выходной слой классификации патча. Более подробно список слоев выглядит следующим образом:

- Входной слой $M \times N \times 3$
- Слой масштабирования до размера $64 \times 64 \times 3$
- Свёрточный слой (размер фильтра $5 \times 5 \times 3$)
- Активационная функция ReLU + слой MaxPooling (3×3)
- Свёрточный слой (размер фильтра $5 \times 5 \times 64$)
- Активационная функция ReLU + слой AvgPooling (3×3)
- Свёрточный слой (размер фильтра $5 \times 5 \times 32$)
- Активационная функция ReLU + слой AvgPooling (3×3)
- Полносвязный слой (64)
- Активационная функция ReLU
- Полносвязный слой (2)
- Softmax + выходной слой классифкации патча

Следует также отметить, что в архитектуре используются слои нормализации батча (batch normalization). Эти слои представляют из себя операции предобработки для небольших порций поступающих на этапе обучения данных, что позволяет формировать нормализованное пространство признаков. Таким образом в ходе работы нейронной сети вектор признаков слоя *i* $\mathbf{x}_i$ проходит процедуру нормализации в рамках батча *B* посредством приведения к среднему μ_B и дисперсии σ_B^2 этого батча:

$$\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}, \quad (1)$$

где ε используется в роли стабилизатора при малых значениях дисперсии. В целях устранения проведённой нормализации на этапе обучения посредством стохастического градиентного спуска нормализованный вектор признаков преобразуется на этапе активации в следующий вид:

$$\mathbf{y}_i = \gamma \hat{\mathbf{x}}_i + \beta, \quad (2)$$

где параметры γ, β являются настраиваемыми и вычисляются в процессе обучения. Тем самым входные веса не подвергаются изменению на этапе обучения и процесс нормализации не приводит к их искажению. В качестве признаков патча (фрагмента изображения) используется выход первого полносвязного слоя, состоящий из 64 элементов.

4. Экспериментальные исследования

В ходе экспериментальных исследований использовался ПК с графическим ускорителем NVIDIA GeForce RTX 2060. В качестве данных для обучения использовалась датасеты CMFD [2] и Casia v2 [17]. В общем для обучения было использовано порядка 1000 изображений, 80% которых составляли изображения, содержащие встроенные дубликаты с различными искажениями. На этапе тестирования использовалось порядка 300 изображений. Обученная нейронная сеть использовалась на этапе тестирования в рамках алгоритма, описанного в [18]. Данный подход заключался в сопоставлении векторов признаков с целью обнаружения ближайших. Полученные результаты показали преимущество перед стандартными подходами к обнаружению дубликатов (Таблица 1) в смысле выбранных качественных метрик Precision, Recall и F1. Следует отметить, что несмотря на прирост в качественных показателях, вычислительная сложность блока поиска ближайших векторов остается высокой и требует дальнейших улучшений (в настоящее время ведутся исследования применимости алгоритма PatchMatch [19] в данной задаче).

Таблица 1. Качество обнаружения дубликатов предложенного решения.

	[48]	[50]	[49]	[27]	[18]	*Разработанное*
F1	0.64	0.66	0.67	0.68	0.59	0.77
P	0.54	0.54	0.57	0.59	0.54	0.71
R	0.80	0.84	0.80	0.82	0.66	0.84

5. Заключение и выводы

В ходе исследований был разработан алгоритм обнаружения дубликатов с использование аппарата свёрточных нейронных сетей для извлечения признаков локальных областей в режиме анализа цифрового изображения в режиме скользящего окна. Полученные качественные результаты отражают преимущество предложенного решения над существующими алгоритмами. Следует отметить, что несмотря на прирост в качественных показателях, скорость анализа изображения и поиска близких векторов признаков остается неизменной и регулируется лишь разностью вектора признаков, формируемого на выходе полносвязного слоя свёрточной нейронной сети. В дальнейшем планируется применять масштабно инвариантные свёрточные нейронные сети [16], которые позволят вычислять мультимасштабные локальные признаки для повышения устойчивости механизма обнаружения дубликатов к вносимым в них искажениям., а также реализовать возможность анализа всех патчей изображения с использованием новой архитектуры сети, состоящей из двух веток – одна позволяет вычислять признаки, другая – находить похожие фрагменты (похожий подход использовали авторы работы [20]).

6. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 20-37-70053, 19-07-00138, 19-07-00474.

7. Литература

[1] Jing, W. Exposing digital forgeries by detecting traces of image splicing / W. Jing, Z. Hongbin // Proceedings of the 8th IEEE International Conference on Signal Processing. – 2006. – Vol. 2. – P. 9464370.
[2] Christlein, V. An evaluation of popular copy-move forgery detection approaches / V. Christlein, C. Riess, J. Jordan, C. Riess, E. Angelopoulou // IEEE Trans. Inf. Forensics Secur. – 2012. – Vol. 7. – P. 1841-1854.
[3] Li, H. Image forgery localization via integrating tampering possibility maps / H. Li, W. Luo, X. Qiu, J. Huang // IEEE Trans. Inf. Forensics Secur. – 2017. – Vol. 12. – P. 1240-1252.
[4] Lee, H. Sparse deep belief net model for visual area / H. Lee, C. Ekanadham, A.Y. Ng // Advances in Neural Information Processing Systems, 2008.
[5] Larochelle, H. Exploring strategies for training deep neural networks / H. Larochelle, Y. Bengio, J. Louradour, P. Lamblin // J. Mach. Learn. Res. – 2009. – Vol. 10. – P. 1-40.
[6] LeCun, Y. Gradient-based learning applied to document recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haner // Oroc. IEEE. – 1998. – P. 2278-2324.
[7] Giacinto, G. Design of effective neural network ensembles for image classification purposes / G. Giacinto, F. Roli // Image Vis. Comput. – 2001. – P. 699-707.
[8] Fukushima, K. Neocognitron: A neural network model for a mechanism of visual pattern recognition / K. Fukushima, S. Miyake, T. Ito // IEEE Trans. Syst. Man Cybern. – 1983. – P. 826-834.
[9] Vedaldi, A. Convolutional Neural Networks for Matlab / A. Vedaldi, K. Lenc, A. Gupta // MatConvNet. – 2015. – P. 1-59.
[10] Kuznetsov, A. Digital image forgery detection using deep learning approach / A. Kuznetsov // Journal of Physics: Conference Series. – 2019. – Vol. 1368 (3). – P. 032028.
[11] Verma, V. DCT-domain Deep Convolutional Neural Networks for Multiple JPEG Compression Classification / V. Verma, N. Agarwal, N. Khanna // Image Commun. – 2017. – Vol. 67. – P. 1-12.
[12] Jegou, H. Aggregating local image descriptors into compact codes / H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, C. Schmid // IEEE Trans. Pattern Anal. Mach. Intell. – 2012. – Vol. 34. – P. 1704-1716.
[13] Krizhevsky, A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // Proceedings of the Advances in Neural Information Processing Systems. – 2012. – P. 1097-1105.
[14] Ren, S. Faster R-CNN: Towards real-time object detection with region proposal networks / S. Ren, K. He, R. Girshick, J. Sun // Proceedings of the Advances in Neural Information Processing Systems. – 2015. – P. 91-99.
[15] Liu, Y. Copy-move Forgery Detection based on Convolutional Kernel Network / Y. Liu, Q. Guan, X. Zhao // Multimedia Tools Appl. – 2018. – Vol. 77. – P. 18269-18293.
[16] Xu, Y. Scale-Invariant Convolutional Neural Networks / Y. Xu, T. Xiao, J. Zhang, K. Yang, Z. Zhang // ArXiv, abs/1411.6369.
[17] CASIA Tampered Image Detection Evaluation Database, 2010 [Electronic resource]. – Access mode: http://forensics.idealtest.org/casiav2/.
[18] Kuznetsov, A.V. A copy-move detection algorithm based on binary gradient contours / A.V. Kuznetsov, V.V. Myasnikov // Computer Optics. – 2016. – Vol. 40(2). – P. 284-293. DOI: 10.18287/2412-6179-2016-40-2-284-293.
[19] Barnes, C. PatchMatch: A randomized correspondence algorithm for structural image editing / C. Barnes, E. Shechtman, A. Finkelstein, D.B. Goldman // ACM Transactions on Graphics. – 2009. – Vol. 28 (3). – P. 24.
[20] Wu, Y. BusterNet: Detecting copy-move image forgery with source/target localization. / Y. Wu, W. Abd-Almageed, P. Natarajan // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – 2018. – Vol. 11210. – P. 170-186.

On deep learning approach in remote sensing data forgery detection

A. Kuznetsov[1,2]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Abstract. Forgery of digital images is a known problem due to the increasing availability of technologies and software that make it easy to create distorted images. In order to counter such attacks, several approaches have been developed to detect fakes. Of particular importance are the methods for detecting distortions of an individual type of digital images - remote sensing data, which can be used to ensure the safety of protected areas, monitor the state of the environment, etc. This article proposes a new scheme based on neural networks and deep learning, which is based on the use of the new convolutional neural network (CNN) architecture to improve the quality of detection of the most common type of attacks on digital images - embedding duplicates. Within the framework of the proposed architecture, additional preprocessing layers are applied to improve the quality of detection. This approach also demonstrates invariance to distortions introduced into the duplicated region. Experiments show that the proposed solution exceeds the known duplicate detection algorithms - the metric value F1 reaches 0.77.