Использование методов глубокого обучения в задаче обнаружения искажений цифровых изображений

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. В данной работе представлен алгоритм обнаружения одного из наиболее часто применяемых видов искажений цифровых изображений – сплайсинга или склейки. В основе алгоритма лежит использование сверточной нейронной сети VGG-16. Полученные результаты демонстрируют высокое качество обнаружения изображений, содержащих искусственные искажения в сравнении с существующими решениями.

1. Введение

Благодаря стремительному развитию технологий цифровой обработки изображений и постоянно растущей популярности цифровых регистрирующих устройств, редактирование цифрового изображения стало достаточно простой операцией даже для неопытного пользователя. Используя современное программное обеспечение для обработки цифровых изображений, любой пользователь может вносить изменения в цифровые изображения таким образом, чтобы визуально отличить подделку от исходных данных было практически невозможно. В последние несколько десятилетий в СМИ достаточно часто появляются искаженные цифровые изображения, и на фоне этого возникают постоянные споры о достоверности представленной информации. Факт преднамеренного изменения содержащейся на изображении информации в целях ее сокрытия или искажения будем называть искусственными изменениям (искажениями) или атаками.

Наиболее распространенными типами (видами) искусственных искажений цифровых изображений являются встраивание дубликатов, ресэмплирование и сплайсинг. Все они применяются для сокрытия или искажения представленной на космическом снимке информации. Первый тип искажений (встраивание дубликатов) означает копирование фрагмента космического снимка, внесение в этот фрагмент каких-либо искажений (аффинные преобразование, аддитивный шум, переквантование уровней яркости и др.) и встраивание измененного фрагмента в другую область этого же изображения (ту его часть, которую необходимо скрыть) [1,2]. Вторым не менее популярным типом искусственных искажений является ресэмплирование [3] – аффинное преобразование фрагментов космического снимка и встраивание их в другие космические снимки. Применение данного вида искажений актуально, например, в описанном выше примере уменьшения размеров загрязнений на поверхности водных ресурсов. Третий часто используемый тип искажений – сплайсинг – заключается в использовании фрагментов разных космических снимков для формирования нового космического снимка или детального искажения существующего (например, формирование на

территории лесного фонда населенного объекта, скопированного из другого космического снимка) [4]. И, наконец, еще одним способом, которым пользуются злоумышленники является сжатие JPEG [5]. В данном случае после встраивания в JPEG файл какой-либо информации и повторного сжатия возникают локальные отличия в свойствах JPEG сжатия.
Описанные способы внесения искажений являются наиболее популярными на сегодняшний день, что подтверждается огромным количеством публикаций, направленных на разработку решений по обнаружению таких атак [1-5].

2. Алгоритм обнаружения искажений

В данной работе для решения задачи обнаружения искажений типа сплайсинг мы используем методы глубокого обучения для классификации изображений на два класса: оригинальное и содержащее искажение. Разработанный подход предполагает анализ изображения в режиме окна с перекрытиями и классификацию каждого фрагмента изображения, соответствующего положению окна. Данный подход обладает способностью строить дискриминантные функции непосредственно на основе данных без какого-либо априорного знания о процессе извлечения признаков. В этой работе мы используем сверточные нейронные сети, которые в последние несколько лет позволяют добиться очень высоких результатов во многих приложениях компьютерного зрения, таких как распознавание объектов, сегментация изображений, распознавание лиц и многие другие. Обучение таких сетей является вычислительно очень сложной процедурой и требует огромное количество данных. Тем не менее, существующие базы данных изображений, содержащих подделки, содержат не более нескольких тысяч изображений, что недостаточно для обучения или точной настройки сетей со сложной архитектурой, таких как, например, VGG-16 [6] (рисунок 1).

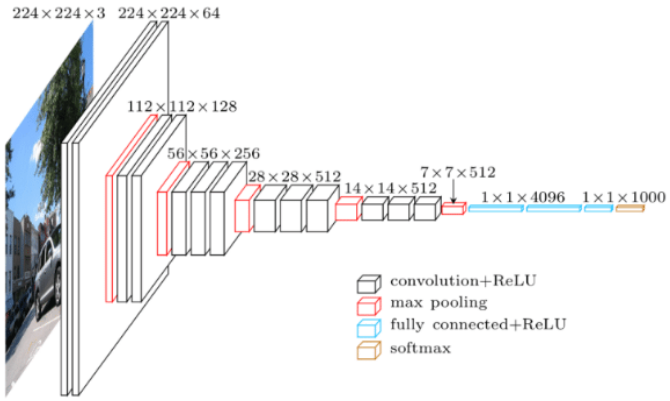

Рисунок 1. Архитектура используемой сети VGG-16.

Для решения указанной проблемы предлагается стратегию классификации на основе фрагментов изображений (или патчей). Это позволяет нам в основном решать две проблемы: во-первых, недостаток данных для обучения. Во-вторых, это позволяет нам фиксировать размеры входных данных в соответствии с размером патча, не используя методы аугментации и не внося дополнительные геометрические деформации. Более того, поскольку база данных CASIA [7] (и аналогичные доступные базы данных искаженных изображений) не содержат сегментированную аннотацию, используется простой и эффективный метод для автоматического вычисления маски поддельной области на основе информации, содержащейся в базе данных CASIA.
Предлагаемая модель представляет собой VGG-подобную архитектуру сети [6]. Она принимает в качестве входных сигналов патчи фиксированного размера 40x40x3 и состоит из двух сверточных блоков и двух полносвязных блоков. Каждый сверточный блок содержит два сверточных слоя с активационной функцией ReLU, за которыми следует слой объединения (слой пулинга). Все сверточные слои используют ядра с размером 3x3, а размер слоя пулинга составляет 2x2. Между различными блоками используются исключающие слои (Dropout) [26] для решения проблемы переобучения. К входным данным применяется процедура

нормализации, которая приводит входные данные в диапазон от 0 до 1. Общее количество параметров, которые должны быть настроены на этапе обучения в предлагаемой сети, составляет 869154.

3. Экспериментальное исследование предложенного алгоритма

В рамках первого эксперимента проводится исследование работы предложенного подхода для классификации изображений на два класса: подлинные и измененные. Для этого база данных изображений разделяется на обучающую и тестовую выборки в соотношении 80:20. Для изображений формируются наборы патчей, соответствующих подлинным изображениям и содержащим искажения. В качестве подлинных выбираются патчи из оригинальных изображений, в качестве искаженных выбираются патчи на границах встроенных областей. Размер патча составляет 40x40 пикселей. Патчи выбираются из изображения с перекрытием с шагом 20 пикселей между ними. Значения пикселей в патчах подвергаются процедуре нормализации, как было сказано выше. Обучение сети производилось в течение 30 эпох. На этапе тестирования патчи извлекаются с использованием той же методологии, которая использовалась для обучения, в то время как окончательное решение о классификации изображения принимается путем голосования по большинству патчей первого или второго класса. В таблице 1 представлены результаты первого эксперимента и сравнение с некоторыми существующими решениями, которые позиционируются как одни из лучших для решения задачи обнаружения сплайсинга.

Таблица 1. Результаты сравнения алгоритмов классификации изображений, содержащих искажения типа сплайсинг, на базе данных CASIA v2 [7].

Метод	Accuracy	Precision	Recall	F1 Score
[8]	79.74	-	0.7243	
[9]	96.8	-	-	
[10]	95.6	-	-	
[11]	90.1	-	-	
Предлагаемый подход	96.4	0.95	0.981	0.965

В рамках второго эксперимента все изображения подвергались повторному сжатию алгоритмом JPEG для оценки влияния данного вида постобработки искаженных изображений на результат классификации. Результаты экспериментов показаны в таблице 2.

Таблица 2. Исследование алгоритма классификации при повторном сжатии изображения базы данных CASIA v2 [7] алгоритмом JPEG.

Данные	Accuracy	Precision	Recall	F1 Score
Исходные	96.4	0.95	0.981	0.965
Сжатие Q=90	67.1	0.78	0.46	0.58
Сжатие Q=80	66.3	0.76	0.53	0.62

4. Заключение

В статье приводится описание метода обнаружения искусственных искажений цифровых изображений с использованием сверточной нейронной сети VGG-16. Полученные результаты показали высокое качество классификации изображений и возможность применения метода в условиях повторного сжатия искаженных изображений алгоритмом JPEG. В дальнейшем планируется провести детальное сравнение с другими методами обнаружения сплайсинга и реализовать детектирование искаженных областей.

5. Литература

[1] Cao, Y. A robust detection algorithm for copy-move forgery in digital images / Y. Cao, T. Gao, L. Fan, Q. Yang // Forensic Sci. Int. – 2012. – Vol. 214. – P. 33-43.

[2] Kuznetsov, A. A Copy-Move Detection Algorithm Using Binary Gradient Contours / A. Kuznetsov, V. Myasnikov // International Conference on Image Analysis and Recognition, ICIAR. – 2016. – Vol. 9730. – P. 349-357.
[3] Bayar, B. On the robustness of constrained convolutional neural networks to jpeg post-compression for image resampling detection / B. Bayar, M.C. Stamm // Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing, 2017.
[4] Rao, Y. A deep learning approach to detection of splicing and copy-move forgeries in images / Y. Rao, J. Ni // IEEE International Workshop on Information Forensics and Security (WIFS), 2016. – P. 1-6.
[5] Amerini, I. Localization of JPEG double compression through multi-domain convolutional neural networks / I. Amerini, T. Uricchio, L. Ballan, R. Caldelli // IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017. – P. 1865-1871.
[6] Simonyan, K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman. – ArXiv preprint, 2014. – ArXiv: 1409.1556.
[7] CASIA Tampered Image Detection Evaluation Database, 2010 [Electronic resource]. – Access mode: http://forensics.idealtest.org/casiav2/.
[8] Sutthiwan, P. Markovian rake transform for digital image tampering detection / P. Sutthiwan, Y.Q. Shi, H. Zhao, T.-T. Ng, W. Su // Transactions on data hiding and multimedia security. – 2011. – Vol. VI. P. 1-17.
[9] Wang, W. Effective image splicing detection based on image chroma / W. Wang, J. Dong, T. Tan // ICIP. IEEE, 2009. – P. 1257-1260.
[10] Wang, W. Image tampering detection based on stationary distribution of markov chain / W. Wang, J. Dong, T. Tan // ICIP. IEEE, 2010. – P. 2101-2104.
[11] Lin, Z. Fast, automatic and finegrained tampered jpeg image detection via DCT coefficient analysis / Z. Lin, J. He, X. Tang, C.-K. Tang // Pattern Recognition. – 2009. – Vol. 42(11). – P. 2492-2501.

Благодарности
Настоящая работа была выполнена при поддержке грантов РФФИ № 19-07-00138 и 19-07-00474.

Digital image forgery detection using deep learning approach

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. This paper presents an algorithm for detecting one of the most commonly used digital images forgery - splicing. The algorithm is based on the use of the VGG-16 convolutional neural network. The results obtained demonstrate high classification quality of images with artificial distortions in comparison with existing solutions.