Image Level Forgery Identification and Pixel Level Forgery Localization via a Convolutional Neural Network

Haitao D. Deng[1] **& Yitao Qiu**[2]

[1] Department of Materials Science and Engineering [2]Department of Civil and Environmental Engineering

Introduction

The ease of manipulation of digital data through editing/cropping tools such as photoshop and photoeditor etc has often negatively impacted the information credibility. While several successful cases for forgery detection [3, 6] have been demonstrated [5, 7], progress for a generic detection technique development has been stagnant due to two main reasons. One, there are various fundamentally different forgery types. Two, it's difficult to pin point the location of forged regions. [3, 5, 6, 7, 9].

Category	Features/Models		#Parameters	Forgery Types
ML	[4]	CFA	<10	S, CM, R
ML	[11]	ELA	<10	S, CM, R
ML	[8]	NOI	<10	S, CM, R
DL	[1]	Bayar	∼20M	S
DL	[2]	SRM	∼50k	S,CM,R
DL	[10]	Artificial	7M	S,CM, R, E

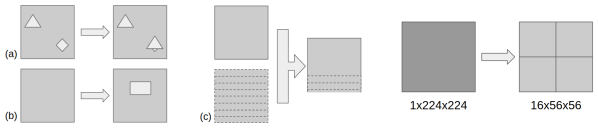
S:splicing, CM: copy-move, R: removal, E: enhancement

Here, we report on a light-weight (5k and 700k) parameters) VGG-derived convolutional neural network architecture that allows for image level forgery detection with an accuracy of 91% and AUC of 85% on test data and for 93% and AUC of 79% pixel level forgery detection.

Dataset and Features

Dataset	forged images	forged pixels
Train 1	31.09%	9.10%
Dev 1	29.20%%	8.42%
Test 1	30.20%	8.02%
Train 2	98.39%	16.13%
Dev 2	97.13%	16.85%

All of the data was obtained from The Image Manipulation Dataset [1] and the COCO Dataset [2]. A total of 10000 images (224 pixels × 224 pixels) were obtained, and split into train set, development set, and test set as 8:1:1. In this work, forgery types include a) copy-move, b) locally enhance, c) splicing.

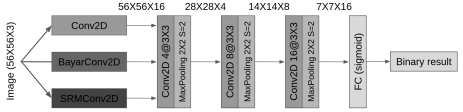


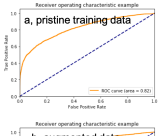
[1]https://www5.cs.fau.de/research/data/image-manipulation/

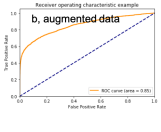
[2]http://cocodataset.org/#home

Image Level Forgery Identification

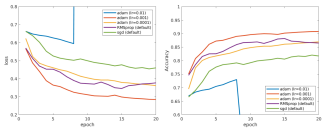
Following VGG network architecture, using a far to near approach, we built our network as below. For training, images were into 56x56 smaller patches to augment training sample number.







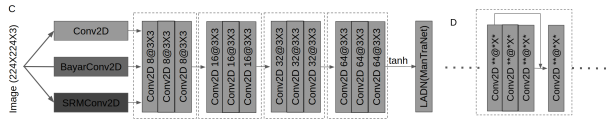
We augmented training data by manually enhancing local pixel colors via multiplying a random coefficient to the local pixel values (a.k.a. enhancement) in non-private regions (Train 1 to Train 2) to increase forged image sample and pixel ratio and achieved better performance on the same test data. This suggests that our classification network was able to capture the common features shared between different manipulation types.



By adjusting the hyperparameters, including the learning rate and the optimizer, we found that adam with the initial learning rate of 0.001 gave the highest training accuracy and lowest loss.

Pixel Level Image Forgery Localization

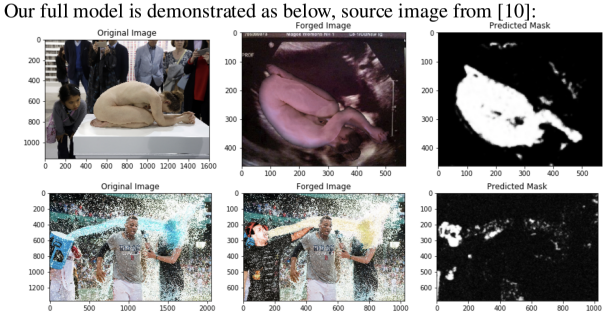
In the image level forgery identification network, the pooling layers condense the spatial information down to fewer pixels for final classification; to achieve pixel wise prediction, the pooling layers were discarded, producing a pixel-wise feature extractor. The feature extractor is then fed into a local anomaly detection network proposed by Wu's [11] paper.



Our full model achieves better performance than the untrained model from [10] and similar performance to the partially trained one. To avoid the potential problem of vanishing gradients and to expedite the training process, we've also modified the network in C by adding shortcut paths every two convolution layers in the intermediate blocks (Model D) and have found similar performance within less training time.

Model	Trained Parameters	Accuracy	AUC
ManTraNet	0 (7 M in total)	0.93	0.67
LADN trained ManTraNet	0.2 M (7M in total)	0.91	0.798
Our Model	0.27 M (0.27 M in total)	0.92	0.794
Our model with ResNet	0.27 M (0.27 M in total)	0.93	0.78

Our full model is demonstrated as below, source image from [10]:



Conclusion and Future Work

Here we deliver a light-weight network architecture that achieves high performance in both image level identification and pixel level localization. Future effort can be focused on condensing the LADN network, as well as incorporating more features by using filtering kernels generating features such as CFA, EIA to improve network performance.

Reference

[1] Bayar & Stamm, ACM. 2016, pp. 5–10; [2] Bayar & Stamm, IEEE Trans. Inf. Forensics Secur, 13.11 (2018), pp. 2691–2706; [3] Birajdar & Mankar, Digit. Invest. 10.3 (2013), pp. 226–245; [4] Ferrara et al., IEEE Trans. Inf. Forensics Secur, 7.5 (2012), pp. 1566–1577; [5] Hill & Rager, "Image Forgery Detection"; [6] Hsu & Chang, ICME, Toronto, Canada, 2006; [7] Huh et al., ECCV, 2018, pp. 101–117; [8] Mahdian & Saic, Image. Vision. Comput., 27.10 (2009), pp. 1497–1503; [9] Rao & Ni, WIFS, IEEE. 2016, pp. 1–6; [10] Simonyan & Zisserman, arXiv:1409.1556 (2014); [11] Wu et al., CVPR 2019,pp. 9543–9552; [12] Zhou et al., CVPR, 2018, pp. 1053–1061.