Detection of Image Splicing

Anubha Agrawal
agraw208@umn.edu

Divya Shrinivasa Nairy
shrin020@umn.edu

Mourya Baddam
badda004@umn.edu

Abstract

With the increasing popularity of sophisticated photo editing software, manipulating images has become a trivial task. Such manipulated images can impact criminal investigation, defame a person/business and even influence political views. One of the most common types of manipulation is image splicing which is the process of making a composite picture by cutting some object from a base image and adding it to an another image. Given a manipulated image, the proposed project aims to leverage Reflectance and Noise characteristics to detect tampered regions. Our testing and training dataset comprises of images generated by splicing an object from a base image of high illumination to another image of low illumination and vice versa.

1. Introduction

In recent years, there have been significant advances in image editing techniques and tools. There are multiple low-cost and user friendly editing softwares available to the world at large. This has led to an increase in malicious use of realistic-looking tampered images. There are multiple popular tampering techniques such as splicing, copy-move, removal etc. Image splicing copies regions from an authentic image and pastes them to other images. Oftentimes, forged images looks visually plausible and even with careful inspection it's difficult to detect the tampered regions by the naked eye. Figure 1 is one such image that went viral. As a result, distinguishing authentic images from tampered images has become increasingly challenging. Digital image forensics is an emerging research field which aims at validating the authenticity of images. It is of great importance as it attempts to prevent people and organizations from using tampered images for unethical and fraudulent business or political purposes.

Research [15] has shown that people have an extremely limited ability to detect and locate manipulations of real-world scenes. However, same cannot be said for automated tools. The image manipulation process introduces various high-level artefacts such as unnatural contrast and low-level artefacts such as inconsistent noise levels in the image. A neural network tool can achieve greater accuracy by utilizing such artefacts. In this project, we investigate how to adopt object detection networks to perform image manipulation detection by utilizing Reflectance and Noise features. More specifically, we adopt a two-stream Faster R-CNN [16] network and perform end-to-end training. We utilizie the RPN component of the Faster R-CNN to propose image regions that are likely to tampered.

Figure 1: Left image: Spliced image, Right image: Original image

2. Related Work

A lot of research is being done to detect manipulated images using low-level image statistics. [3] found spliced regions by leveraging mathematical morphological filters including Median filtering and Gaussian low pass filters. [9] detected image splicing using features extracted by constructing Markov states using maximization and threshold expansion. This work leveraged discontinuities in edges and corners. [19] applied adaptive clustering to regions detected by an image level Convolutional Neural Network(CNN) (cascade of Coarse CNN and Refined CNN) in order to find the spliced regions. [12] decreased the computational code in detecting spliced images by eliminating the luminance channel and using only the two chrominance channels. [14] leveraged bipolar signal perturbation to detect spliced images. [18] proposed a unified DNN called ManTraNet which is an end-to-end network that performs both detection and localization without extra preprocessing and postprocessing.[13] proposed a two-stream encoder-decoder network that utilizes high-level (ex-contrast) and the low-level(ex- noise) image features for precisely localizing forged regions in a manipulated image. [10] proposed extracting and combining Markov features in both Discrete Wavelet Transform (DWT) and Local Binary Pattern (LBP) for image splicing detection. [1] proposed a digital image authentication method based on the quadratic mean error of the Color Filter Array interpolation pattern estimated from the analysed image. SpliceRadar [8] identifies the tampered part in an image by identifying the different camera models for the pixels in the image using Rich filters(RF) based on [6] and clustering them using Gaussian Mixture Model.

2.1. Baseline Model

Zhou et al [21] proposed a two stream Faster R-CNN model, as shown in Figure 2, trained end-to-end to identify the tampered regions in an image. We considered this state-of-the-art model as our baseline. One of the two streams is an RGB stream whose purpose is to extract features from the RGB image input to find tampering artifacts like strong contrast difference, unnatural tampered boundaries, and so on. The other is a noise stream that leverages the noise features extracted from a steganalysis rich model filter layer based on [6] to discover the noise inconsistency between authentic and tampered regions. Then features from the two stream are fused through a bilinear pooling layer [7] to further incorporate spatial co-occurrence of these two modalities.

The RGB-N method explicitly analyzes the noise pattern. In the case of a forged image intentionally contaminated with highly correlated noise, the model will likely fail to detect tampered areas. The difference in noise features between tampered region and the rest of the image can also be partially masked with compression introduced by repeated saving of the tampered image. Images can also be carefully post processed to minimise the contrast and disguise the splicing boundaries which is challenging for the RGB stream.

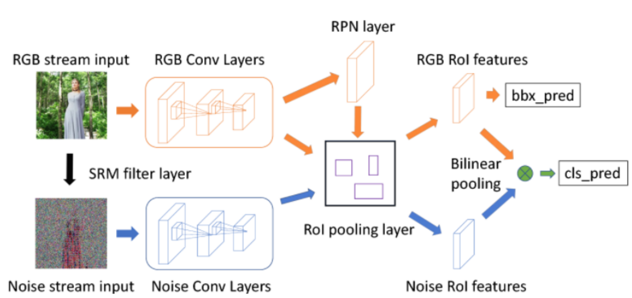

Figure 2: Baseline Model: RGB-N

3. Method

The performance of the baseline model [Figure 5] on our dataset makes it evident that the model is unable to utilize the difference in reflectance characteristics to detect the spliced region.

We propose a novel framework which learns rich features from the low-level reflectance characteristics of the image and uses them, along with the Noise features, to predict and identify the spliced parts in an image. As shown in the Fig 4, we changed the model proposed by [21] by replacing the RGB stream to capture the reflectance features of the image by extracting the reflectance map using a Retinex Filter (RF).

According to the Retinex theory proposed in [11], an image is the element wise product of illuminance and the reflectance as per below equation,

$$S = L \odot R$$

where S is the image, L is the illumination map, R is the reflectance map and the dot operator indicates element wise product.

There are multiple approaches for extracting illumination map. [5] proposed a simple yet powerful combinatorial search approach using the chromagenic effect to find out which parts of the image are illuminated by the same lights. They clustered pixels based on which source is illuminating them. LR3M [17], STAR [20] and NATLE [2] adopt different optimization frameworks based on retinex to determine the illumination and reflectance maps of an image.

According to [17] the initial estimate for illumination map can be approximated as the average of the RBG streams of the image as below.

$$\hat{L}(x) = \frac{1}{3} \sum_{c \in \{R,G,B\}} S^c(x)$$

The optimized Illumination map can then be computed by minimizing the below equation.

$$\underset{L}{\operatorname{argmin}} \|L - \hat{L}\|_F^2 + \alpha \|\nabla L\|_1$$

There exists a closed form solution for the above optimization problem given as below.

$$l = (I + \sum_{d \in \{h,v\}} D_d^T \mathrm{Diag}(a_d) D_d)^{-1} \hat{l}.$$

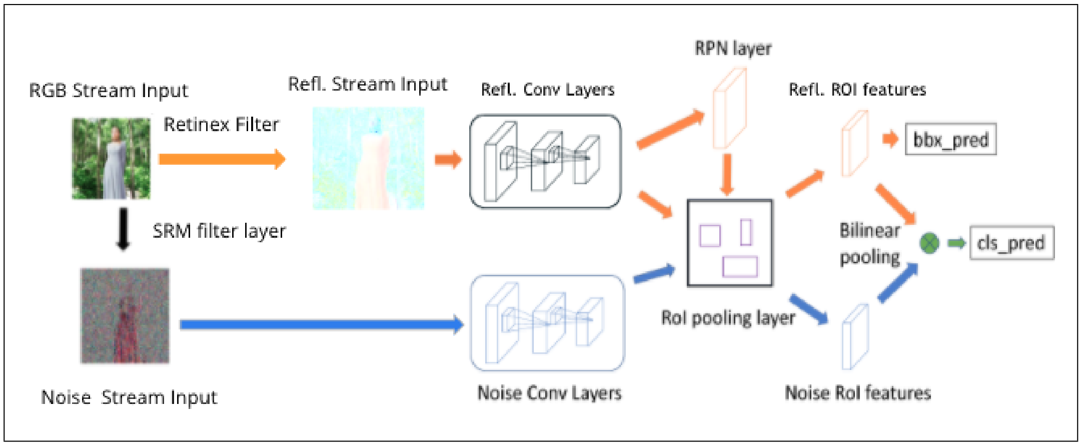

Figure 3: Proposed Model: Extension of Baseline Model with added Illumination stream

$$A_d(x) = \frac{\alpha}{\mid \nabla_d \widehat{L}(x) \mid + \epsilon}.$$

where l and a_d are vectorized form of L and A_d, I is an identity matrix of size nxn, where n is the number of pixels in the image S and D_d is a discrete differential operator matrix. This closed form solution involves computing the inverse of a nxn matrix which is of the complexity O(n^4). This operation makes it impossible for us to implement this model. But the precision required for extraction of reflectance map could also be achieved by the initial estimate of illumination map, which, according to [2], is best given by the weighted average of the RGB streams of the image as below.

$$\widehat{L} = 0.299R + 0.587G + 0.114B$$

Hence, we extracted the best estimate of the illumination map as the weighted average of RGB streams using the weights given above and used that to get the reflectance map for each image, by elementwise division of the image by the illumination map. The extracted reflectance map and Noise image are fed into respective VGG16 Convolutional Networks pre-trained on ImageNet dataset for domain adaptation. The output of the Reflectance CNNs was passed to a RPN network which provides ROI for bounding box regression. The output from both the networks was combined using Compact Bilinear Pooling and passed to fully connected and softmax layers which predict the class for the ROI regions.

4. Data

We manually collected 500 images by a combination of web scraping and images captured from our cell phones. 255 of these images are of high illumination and remaining 245 are low illumination images. We further separated 50 low illumination and 50 high illumination images for creating our test dataset to ensure the same background and tampered object do not appear in both train and test datasets. This left us with 400 images to create our train dataset.

All of our images contain trees along with one or more objects of the following classes- animal, cable car, building, vehicle, fruit and human. We manually annotated all images in the Pascal VOC XML format [4]. We then used these annotations to create segmentation masks. With the help of these images, annotations and segmentation masks we automatically created a synthetic dataset of 5000 training images and 1000 testing images using the following process: We first randomly chose a background image using a modulo hashing function. Then, we another image of contrasting illumination was chosen at random. This second image may have multiple labeled objects. We randomly chose one such object from the chosen image using the segmentation mask and then copied the selected object onto the first image at random co-ordinates with the help of the annotations. For all the images in our synthetic dataset, we also created new annotations containing the coordinates of the tampered region for ground truth. One thing to note is that for the test dataset, we made sure that spliced image has a different illumination characteristic as compared to the rest of the image. Figure 5 shows two such created images.

This synthetic dataset is still not big enough to train a deep neural network, so we used pre-trained model weights on the ImageNet dataset.

Figure 4: Image and its components

Figure 5: Left image- Light object in Dark Background, Right image- Dark object in Light Background

5. Result

On our test dataset, the evaluation metrics for the baseline and our proposed model are shown in the below table. We got the F1 scores of 0.44 for baseline and 0.77 for our model. We categorised our detection into three categories (poor, good and excellent) based on the IOU of the detection compared to the ground truth of the image.

Metrics	Baseline	Implementation
F1 Score	0.44	0.77
Mean IOU	0.34	0.7
Poor detection	622	181
Good detection	235	246
Excellent det.	143	573

Our model is able to confidently detect the spliced parts that have a significant difference in the lighting conditions compared to the background. The baseline model did not find any spliced region in Figure 6a but our model was able to identify the fruit as spliced in Figure 6b. In the next pair, base line detected both the cars as spliced (Figure 6c) but our model correctly detected the spliced car (Figure 6d).

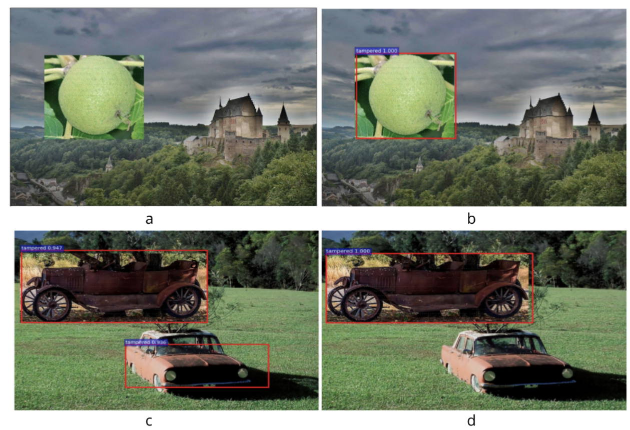

Figure 6: Qualitative Results. a and c: Detected by Baseline model. b and d: Detected by our model

6. Conclusion

We propose a novel framework which learns rich features from the low-level reflectance characteristics of the image and uses them, along with the Noise features, to predict and identify the spliced regions in an image. Reflectance characteristic plays a key role in detecting the tampered regions.

As our dataset contains images with significant difference in the illumination between the spliced region and the rest of the image, our model containing the reflectance stream performs better than the baseline model.

7. Role

Anubha- Collected and Annotated (300) images, Code for Evaluation metrics and Training model, Report- Introduction, Related work, Dataset creation

Divya- Collected and Annotated (100) images, Code for creating Synthetic Dataset and Training model, Report-

Abstract, Conclusion, Comments from committee

Maurya- Collected and Annotated (100) images, Code for Segmentation Mask Creation, Reflectance Map creation and Training model, Report- Method, Result

8. Comments from committee

1. Evaluate the performance of the model on a dataset comprising of high illumination difference.
   Our initial testing dataset had images with varying illumination differences. We then refined our test dataset to have only high illumination difference.

2. Try to resolve the memory issue faced while running the three stream network by resizing the image.
   Inspite of resizing the images, reducing batch size to 4 and parameter tuning, the three stream model (containing Reflectance, Noise and RGB characteristics) was still facing memory issues.

9. Link to code

Our dataset is present here. Our code is present in drive here.

References

[1] E. A. Armas Vega, E. González Fernández, A. L. Sandoval Orozco, and L. J. García Villalba. Passive image forgery detection based on the demosaicing algorithm and jpeg compression. *IEEE Access*, 8:11815–11823, 2020.

[2] Zohreh Azizi, Xuejing Lei, and C. C Jay Kuo. Noise-aware texture-preserving low-light enhancement, 2020.

[3] G. Boato, D. Dang-Nguyen, and F. G. B. De Natale. Morphological filter detector for image forensics applications. *IEEE Access*, 8:13549–13560, 2020.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[5] G. Finlayson, C. Fredembach, and M. S. Drew. Detecting illumination in images. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[6] J. Fridrich and J. Kodovsky. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, 2012.

[7] Akira Fukui, Dong Huk Park, Daylen Yang, Anna Rohrbach, Trevor Darrell, and Marcus Rohrbach. Multimodal compact bilinear pooling for visual question answering and visual grounding, 2016.

[8] Aurobrata Ghosh, Zheng Zhong, Terrance E. Boult, and Maneesh Singh. Spliceradar: A learned method for blind image forensics. *CoRR*, abs/1906.11663, 2019.

[9] Jong Goo Han, Tae Hee Park, Yong Ho Moon, and Il Kyu Eom. Efficient markov feature extraction method for image splicing detection using maximization and threshold expansion. *Journal of Electronic Imaging*, 25(2):023031, Apr. 2016.

[10] Navneet Kaur, Neeru Jindal, and Kulbir Singh. A passive approach for the detection of splicing forgery in digital images. *Multimedia Tools and Applications*, 79, 11 2020.

[11] Edwin H. Land and John J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, Jan 1971.

[12] Thuong Le-Tien, , Hanh Phan-Xuan, Thuy Nguyen-Chinh, and Thien Do-Tieu. Image forgery detection: A low computational-cost and effective data-driven model. *International Journal of Machine Learning and Computing*, 9(2):181–188, Apr. 2019.

[13] Aniruddha Mazumdar and Prabin Kumar Bora. Two-stream encoder-decoder network for localizing image forgeries, 2020.

[14] T. . Ng and S. . Chang. A model for image splicing. In *2004 International Conference on Image Processing, 2004. ICIP '04.*, volume 2, pages 1169–1172 Vol.2, 2004.

[15] Wade K.A. Watson D.G. Nightingale, S.J. an people identify original and manipulated photos of real-world scenes? https://doi.org/10.1186/s41235-017-0067-2.

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

[17] X. Ren, W. Yang, W. Cheng, and J. Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020.

[18] Y. Wu, W. AbdAlmageed, and P. Natarajan. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9535–9544, 2019.

[19] Bin Xiao, Yang Wei, Xiuli Bi, Weisheng Li, and Jianfeng Ma. Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering. *Information Sciences*, 511, 09 2019.

[20] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao. Star: A structure and texture aware retinex model. *IEEE Transactions on Image Processing*, 29:5022–5037, 2020.

[21] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis. Learning rich features for image manipulation detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1053–1061, 2018.