Ruby's Image Forensics Toolkit (RIFT) - A Toolkit for Image Forensics and A Survey of Passive Image Forensics Techniques

Max Ruby

Abstract—This is a pre-preprint. The current version of RIFT is 0.1. This paper becomes a preprint when RIFT is at version 0.6. This should be obvious, as the abstract is mostly a development plan. :)

Planned development:

0.1: Elementary tools, up to par with commonly available tools.

0.2: Improved CMFD, with two different methods chosen by experiment.

0.3: Improved Splicing Detection, with two different methods chosen by experiment.

0.4: Improved Inpainting Detection, with two different methods chosen by experiment.

0.5: AI Generated Image Detection, with 2-4 different methods chosen by experiment.

0.6: Revision and Reorganization of Paper.

0.7: Development of Ensemble Method using the above.

0.8: User Interface rework and "marketing." Spend money on an artist for a nice logo. :D

0.9: Buffer version for fixing something that's obviously going to go wrong in the middle of these things.

1.0: First full release.

1.1: Fixing of bugs that show up after first release.

Development schedule (hoped, not expected):

0.15: Early August (Need to redo the literature review, as it has been an extended period of time since work has been done on this project.)

0.2: August

0.3: September

0.4: October

0.5: November

0.6: December

0.7: January

0.8: February

0.9: No plan

1.0: Easter Release :D

Actual abstract text (Above gets stripped out in 0.6):

Recent developments in image science have created new editing and forgery techniques that allow for the creation of cheap, high-quality forgeries. In particular, the advent of high-quality LDMs have made creating fake images from scratch quite easy for someone with no particular skills. Most individuals limit their use of such tools to entertainment and are open about the images being fake. However, malicious actors occasionally use these tools for the purpose of misleading people. For an individual, the impact of this can range from giving money to fake artists to supporting politicial groups which purposefully mislead their followers. This paper has two purposes. First, it is a survey of the image forensics. Second, it presents an easy-to-use toolkit based on some of the research contained in this paper.

1 Introduction

Various tools have made creating falsified images simple. In particular, Photoshop and LDMs allow for the creation of high-quality forgeries by unskilled consumers. The typical human is poor at identifying forgeries. Therefore, we must create tools to identify forgeries for us. Moreover, those tools must be available to those who wish to use them. Researchers have developed quite a few tools for determining if an image is forged somehow. These tools often fall into two distinct classes: active and passive. An "active" method is one that requires intervention before the creation of an authentic image - such as the creation of digital signatures or authentication watermarks. A "passive" method is one that is not active - such as the analysis of noise or splicing artifacts.

Active methods are of great use to government-level actors. Unfortunately, believing that a watermark or digtal signature is genuine requires some degree of trust in the system that creates and preserves them. Moreover, they require access to databases which may not be readily accessible. This renders them of limited use for a consumer, who may simply not trust such systems. Thus, we concern ourselves primarily with passive methods.

The study of passive methods of determining whether an image is real or forged is called a number of things, such as "Digital Image Forensics," "Forensic Analysis," or "Media Forensics." Which of these terms is used often depends on the group or individual which is interested in the problem. Groups interested in the problem for the sake of Law Enforcement tend to use X, groups interested in the problem for the sake of Media tend to use Y, and purely academic groups tend to use Z - although these lines are not strictly drawn. Note that this suggests that these three types of groups developed their own techniques largely independently of each other at some point, which underscores the importance of the topic at hand.

1.1 Organization of Paper

This paper is organized as follows: first, a few preliminary resources are given. Second, tools are listed by the type of forgery that the tool is designed to detect. This is for the sake of the expected reader, who is more likely to be interested in a specific problem and is probably looking for the right

tool for their problem. Although certain flavors of technique may solve more than one problem - and so this style of organization creates redundancy - it is done for the sake of reference.

Copy-Move Forgery Detection is well studied.

Splicing Detection is also solidly studied, although it is not as well studied due to its relative difficulty. As a general rule, a tool that detects splicing should also function to detect Copy-Move Forgery.

Imitation/Inpainting - where part of the image is artificial, while another part is real - is also studied.

Detection of a completely synthetic image - such as the output of an LDM - is the most pressing issue of the day. It is worth reminding the reader that synthetic images did not begin with GANs or Stable Diffusion; determining whether an image is rendered by a computer program (such as Blender) is also of value, and there is also research in this direction that predates the popularity of the GAN.

Determining whether an image has been digitally enhanced is of interest on occasion. It is sometimes of use to know that an image has been altered, but that the alteration was done for purely aesthetic reasons.

Camera parameter estimation is sometimes useful - if metadata is inconsistent with the estimated camera parameters, then we know that something is strange.

2 Preliminaries

Here's some testing datasets:

Here's a breakdown of a few active groups that care about this problem. This is far from exhaustive, but following such groups is likely to keep one current even after this review is published.

Here's a list of a few companies that work on this problem. One who is interested making this problem their job may seek them out.

Here's a list of US government organizations that appear to care about this problem. A university researcher who is interested in this problem might seek them out for funding.

3 Copy Move Forgery Detection

Copy Move Forgery is when you take part of the image and copy it somewhere else.

An excellent review of CMFD methods is given here: [11]. The two types of methods discussed therein are the Block-based approach and the Keypoint-based approach. A Block-based approach compares features from the blocks and compares them with each other to determine similarity between similar blocks. A Keypoint-based approach identifies distinct local features and attempts to match them to other spots in the image.

From this survey, we chose to investigate the following methods further. Choosing these methods was done by a combination of accessibility and performance.

More recently, there are some methods for CMFD which employ Neural Networks. Outside of this survey, the following methods were also similarly chosen for investigation:

Because we'd like not only to find whether CMF happened, but we also want to pinpoint *where* it happened, we are treating CMFD algorithms as if they're detectors - that is, they should provide a heatmap of likelihood of forgery. This is opposed to treating them as if they are classifiers - that is, we do not merely want to know the probability that CMF happened in an image.

We have chosen to use the CASIA v2.0, MICC-F220, MICC-F2000, and Columbia University datasets to test the CMFD methods above. [11] We wanted a large dataset with translation/rotation/scaling. We also wanted to check how the algorithms work with varying boundaries. This is because Copy Move Forgery in the wild is very often done using a tool such as photoshop - a false object may often be scaled, rotated, and translated for the convenience of the forger. Last, we wanted to use the most commonly used testing dataset as a comparison. This is to identify if a method has been overfit on a dataset - if the performance of a method is wildly better on the standard dataset than on other datasets, it suggests that the underlying assumptions of that method should be revisited.

4 Splicing Detection

Splicing is when you take part of another image and copy it to a forged image. It's extremely popular to make memes.

Splicing detection is often done through some form of Non-Uniformity detection. The idea is that real images behave similarly everywhere in the image - so applying a filter of some sort to the image should cause abnormalities to appear. These methods are the bread-and-butter of someone starting to take an interest in media forensics. It is of particular use to a practitioner that such methods help identify exactly which part of the image has been falsified. Two of the most famous Non-Uniformity based methods include Error Level Analysis and Gaussian Noise Analysis - both of which are included in RIFT. They are incredibly popular, as they essentially cause abnormalities to "light up," allowing someone to call out exactly what part of the image has been manipulated. Other Non-Uniformity based methods include PCA Analysis and Wavelet Compression Analysis (cite Farid). The idea behind these is that a true image will compress uniforly, but a forged image may not. Thus, by compressing an image - by removing a PCA component or Wavelet component, for example - we may find that one part of the image blurs more than another, revealing that the image is forged by splicing.

5 Imitation/Inpainting Detection

Here's a paper: [4]

6 Synthetic Image Detection

Here's a few papers: [12] [13] [10] [9] [5] [1] [6] [3] [8] [2]

Some have noticed that Neural Networks tend to leave "fingerprints" in their power spectrum which make it possible to determine that an image has been generated by a specific type of network [7].

Some are of the opinion that using a Neural Network to solve this problem is the correct path. There is a survey of such efforts here (cite), which we summarize:

These are the tools we'll test.

The dataset we want to use comes from Brandon B. May's group. This is an excellent dataset which allows for us to treat this problem as a segmentation task: it's great, and I love it.

7 Enhancement Detection

zxcv

References

[1] Lucy Chai, David Bau, Ser-Nam Lim, and Phillip Isola. What makes fake images detectable? understanding properties that generalize, 2020.

[2] Riccardo Corvi, Davide Cozzolino, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. Intriguing properties of synthetic images: from generative adversarial networks to diffusion models, 2023.

[3] Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. On the detection of synthetic images generated by diffusion models, 2022.

[4] Hany Farid. Lighting (in)consistency of paint by text, 2022.

[5] Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, and Thorsten Holz. Leveraging frequency analysis for deep fake image recognition, 2020.

[6] Diego Gragnaniello, Davide Cozzolino, Francesco Marra, Giovanni Poggi, and Luisa Verdoliva. Are gan generated images easy to detect? a critical analysis of the state-of-the-art, 2021.

[7] Francesco Marra, Diego Gragnaniello, Luisa Verdoliva, and Giovanni Poggi. Do gans leave artificial fingerprints?, 2018.

[8] Vrizlynn L. L. Thing. Deepfake detection with deep learning: Convolutional neural networks versus transformers, 2023.

[9] Luisa Verdoliva. Media forensics and DeepFakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5):910–932, aug 2020.

[10] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A. Efros. Cnn-generated images are surprisingly easy to spot... for now, 2020.

[11] Nor Bakiah Abd Warif, Ainuddin Wahid Abdul Wahab, Mohd Yamani Idna Idris, Roziana Ramli, Rosli Salleh, Shahaboddin Shamshirband, and Kim-Kwang Raymond Choo. Copy-move forgery detection: Survey, challenges and future directions. *Journal of Network and Computer Applications*, 75:259–278, 2016.

[12] Ning Yu, Larry Davis, and Mario Fritz. Attributing fake images to gans: Learning and analyzing gan fingerprints, 2019.

[13] Xu Zhang, Svebor Karaman, and Shih-Fu Chang. Detecting and simulating artifacts in gan fake images, 2019.