2022 IEEE Image and Video Processing Cup Synthetic Image Detection

Subodha Charles, Jathurshan Pradeepkumar, Heethanjan Kanagalingam, Mugunthan Shandirasegaran
Ragavan Ravichandran, Prarththanan Sothyrajah, Thenukan Pathmanathan, Nirajkanth Ravichandran
Nerththiga Neminathan, Vithurabiman Sethuran, Mayooran Thavendra, Mishanth Pararasasingam
Electronic and Telecommunication department, University of Moratuwa, Colombo district, Sri Lanka

I. Abstract

In recent years there have been astonishing advances in AI-based synthetic media generation. Although this opens up a large number of new opportunities, it also undermines the trustworthiness of media content and supports the spread of misinformation over the internet. For this reason, detecting if an image is an pristine or has been synthetically generated is becoming an urgent necessity in this modern world. Although there are some reliable synthetic image detectors, generalization to different ways of synthetic image generation is still a main challenge. Hence, we considered a universal synthetic image detector since new GAN technologies are published more and more frequently.

Index Terms - Synthetic Images, Image Forensics, GAN

II. Introduction

Recent rapid advances in deep image synthesis techniques, such as Generative Adversarial Networks (GANs), have generated a huge amount of public interest and concern, as people worry that we are entering a world where it will be impossible to tell which images are real and which are fake. Synthesized images generated using these GANs can be used over social media for many malicious intents, from scams to identity stealing, and the general public is not ready to face this menace. It's alarming that recent studies proved that's how synthetic faces prove even more trustworthy at human inspection. Hence, it's utmost important to develop a detector which differentiate synthetic and real images irrespective of the methods used to generate them.

If we consider the problem of detecting whether an image was generated by a specific synthesis technique is relatively straightforward; just train a classifier on a dataset consisting of real images and images synthesized by the technique in question. However, these approaches suffer from dataset bias, in other words, a classifier trained in the human faces dataset will not generalize well to detect the synthetic images of car for example. Even worse, the technique-specific detector is likely to soon become ineffective as generation methods evolve and the technique it was trained on becomes obsolete. The increase rate of new synthetic image generation algorithms which put much effort into generating very realistic pictures, justify the above point.

To solve these issues, we used a synthetic image detector based on an ensemble of Convolutional Neural Networks (CNNs). Hence it can facilitate detecting images synthesized using techniques not available at training time. In open competition - part2, this kind of challenging problem is addressed; synthetic image detection on unseen models. As we can use the same classifier in the two parts of the competition, we decided to go with this approach. We utilized two major ideas in this approach. First, CNNs should provide "orthogonal" results to better contribute to the ensemble. We used a training strategy that increases the diversity among the different learners for this purpose. This prevents the CNNs from overfitting the image generators used in training, thus enabling the ensemble to take a better decision on newly generated images. Secondly, original images are better defined than synthetic ones, thus they should be better trusted at testing time. Indeed, it is safer to assume that the analyst can train on a broad set of real photographs that better represent the real-image class. On the contrary, it is hard to assume that the analyst can train on synthetic images generated with all the possible existing techniques, as these change and get updated too frequently in time. Therefore, we propose a score aggregation strategy that better favour decisions towards the real-image class.

III. Methodology

In this instance, we trained 3 CNN with Res-Net50 as the model using random 32x32 patches from 200x200 images. The image sets listed below are used to train the three Res-Net50 models.

- ModelA - Real and synthetic pets and wildlife animals
- ModelC - Real and synthetic inanimate objects and buildings
- ModelF - Real and synthetic (StyleGAN2 and StyleGAN3) human faces

In this case, for the prediction process, we take 242, 32x32 patches from an input image, and we receive two scores for each patch. We then average the scores of all the patches to get the image score for a single image . Similarly, we receive 3 image scores from the 3 models, and we take the mean to get the final score. If the final image score is positive, we determine the image is synthetic; if it is negative, we determine the image is real.

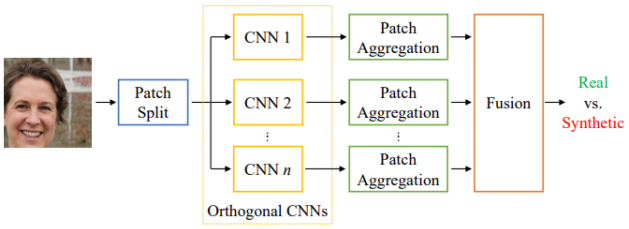

Fig. 1. Methodology used for synthetic image detection.

IV. Dataset Creation

We trained the 3 detectors with mutually exclusive custom dataset which spread over multiple classes. All the images in the dataset are pre-processed with the python script which applies the random (crop) operations as well as the augmentations and different levels of JPEG compressions. We keep 80% of the images for training split, leaving the remaining 20% equally divided into validation and test sets; 10% for each split.

The details of the dataset used for training each of the three detectors are listed below.

1. Model A
- 50k real images of animal faces mainly cats,dogs and wild animals. COCO 2017 real images dataset is used for creation of this dataset.
- 50k synthetic images of animal faces generated using StyleGAN2 and StyleGAN3.

2. Model C
- 60k real images of inanimate objects and buildings.
- 60k synthetic images of inanimate objects and buildings generated using StyleGAN2 and StyleGAN3 algorithms.

3. Model F
- 50k real images of human faces.
- 50k synthetic images of human faces generated using StyleGAN2 and StyleGAN3.

V. Results

Test set accuracy obtained for each dataset is tabulated below.

Test Set	Test Accuracy
Animal class	90%
FFFT class	84%
Inanimate things and building	83%
partially manipulated images	45%

TABLE I
Test set accuracy for each dataset

VI. Conclusion

We used a synthetic image detector based on an ensemble of CNNs which are trained to increase the diversity within the ensemble, to tackle this synthetic image detection problem. The score aggregation strategy used here takes into account the fact that some image generators can be unknown at training time which makes this detector generalize to other unseen models as well.

References

[1] D. Gragnaniello, D. Cozzolino, F. Marra, G. Poggi, and L. Verdoliva, "Are GAN generated images easy to detect? A critical analysis of the state-of-the-art," in IEEE International Conference on Multimedia and Expo, 2021.
[2] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. Efros, "CNN-generated images are surprisingly easy to spot... for now," in CVPR, 2020.
[3] L. Chai, D. Bau, S.-N. Lim, and P. Isola, "What makes fake images detectable? Understanding properties that generalize," in ECCV, 2020.
[4] L. Nataraj et al., "Detecting GAN generated fake images using co-occurrence matrices," in IS&T EI, Media Watermarking, Security, and Forensics, 2019.
[5] Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations? JMLR, 2019.
[6] Sara Mandelli, Nicolo Bonettini, Paolo Bestagini and Stefano Tubaro, "Detecting GAN-generated images by Orthogonal training of multiple CNNS".