

IMPROVING GAN-GENERATED IMAGE DETECTION GENERALIZATION USING UNSUPERVISED DOMAIN ADAPTATION

Mingxu Zhang¹, Hongxia Wang^{1*}, Peisong He¹, Asad Malik², Hanqing Liu¹

¹School of Cyber Science and Engineering, Sichuan University, Chengdu, China

²Department of Computer Science, Aligarh Muslim University, Aligarh, India

euph5717@gmail.com; liuhanqing0520@stu.scu.edu.cn; {hxwang,gokeyhps}@scu.edu.cn

ABSTRACT

In recent years, with the significant improvement of Generative Adversarial Networks (GANs), fake images generated by GAN become hardly distinguishable from real ones, thus threatening the authentication of digital images. To resolve this issue, several fake image detectors based on supervised binary classification have been designed. However, current methods remain vulnerable when testing samples are generated by an unknown GAN model. In this work, an unsupervised domain adaptation strategy is introduced to improve the performance in the generalization of GAN-generated image detection by using a small number of unlabeled images from the target domain. Self-Attention block and novel loss function have been constructed to optimize the domain adaptation process, thus getting a better generalization. Experimental results demonstrate that the proposed scheme achieves high detection accuracy with few unlabeled images in the target domain, which shows that unsupervised methods can be used for the detection of GAN-generated images.

Index Terms— Digital image forensics, Unsupervised domain adaptation, Generative adversarial networks, Fake images detection

1. INTRODUCTION

With the development of Generative Adversarial Network (GAN), it has become more and more convenient for computer to generate fake images with high levels of visual realism, which are difficult for human beings to distinguish. However, the strong capability of GAN to create realistic images has threatened the authenticity and integrity of digital images, especially in cases related to security and forensics. GAN can be used in the generation of fake news, perjury, etc, which affects real-life decisions. Consequently, the challenge of discerning between a GAN-generated image and a real image has been explored by image forensics researchers.

*The corresponding author is Hongxia Wang. This work is supported by the National Natural Science Foundation of China (61972269 and 61902263), Sichuan Science and Technology Program (2022YFG0320), China Postdoctoral Science Foundation (2020M673276)

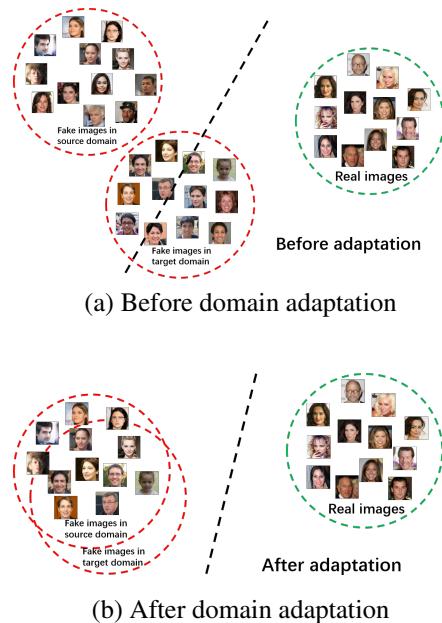


Fig. 1. Illustration of the feature space with tested samples. The black dotted line in (a) shows the decision boundary learned without target domain, while the black dotted line in (b) shows the decision boundary after domain adaptation.

Several methods have been proposed in recent years to classify GAN-generated images and real images. Most of the existing methods rely on Convolutional Neural Networks(CNN) for classification. Mo et al. [1] proposed a CNN-based detection scheme using high-frequency components in the preprocess. Other researchers [2, 3, 4] noticed that the generator of the GAN model performs upsampling operations by certain methods, resulting in anomalous correlations between local pixels of the generated images. Inspired by traditional forensic algorithms, they tried to use forensic features such as co-occurrence matrix instead of input image pixel values as input to the CNN. Besides, several researchers [4, 5] have also further improved the extraction and

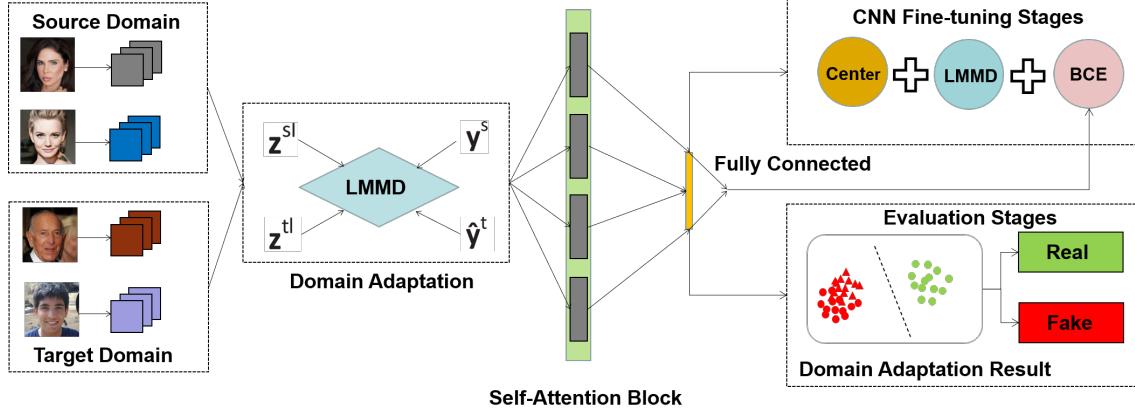


Fig. 2. Schematic diagram of the proposed framework. The LMMD module needs four inputs: the activations \mathbf{z}^{sl} and \mathbf{z}^{tl} where $l \in L$, the ground-truth label \mathbf{y}^s , and the predicted label $\hat{\mathbf{y}}^t$. A novel loss is constructed at the time of training in source domain.

characterization of artificial artifacts in GAN-generated images by modifying the network structure and introducing new network modules. In addition to the spatial domain, the GAN-generated images also have generation traces in the frequency domain. Detection methods based on frequency domain information [6, 7, 8] have been proposed, which also achieve high detection accuracy. Other deep architectures have been tested so far [9, 10] utilize multiple features simultaneously, such as concating shallow and deep features, while focusing on both semantics and Artificial traces of GAN-generated images. Marra et al. [11] designed an algorithm to extract the fingerprint information of GAN images. Their experimental results show that there are unique artifacts in different GAN models, which means although there are differences between different GAN-generated images, they all differ significantly from the real images. Some researchers have tried to build a universal detector that can detect multiple false images, Zhang et al. [12] designed AutoGAN that can simulate the artificial artifacts of GAN images. The model considers upsampling modules such as deconvolution and nearest-neighbor interpolation, and the detection model trained using AutoGAN-generated images can achieve better detection performance when the unknown GAN model is similar to the AutoGAN network structure. Other researchers [13, 14, 15] improved the generalization ability of the model by using a large number of real and GAN-generated images during the training phase. Their experiments demonstrate that data enhancement and increasing the number and diversity of training samples also have an enhancing effect on generalization performance.

However, in practice, an attacker may use images generated by GAN models that do not appear in the training phase. Unfortunately, most of the existing detection methods suffer from poor generalization. Therefore, when unknown GAN-generated images appear, the detection accuracy of existing

methods will face a serious challenge. Furthermore, detection methods based on deep learning thus far are data hungry, detection accuracy can be very low when facing problems such as data mismatch or data is not labeled. In this work, we proposed a generalization scheme based on transfer learning, which enables the detection of other images in the domain by a small number of unlabeled images in the unknown domain. As shown in Fig. 1, our core idea is to make the fake images in the source and target domains overlap in the feature space by domain adaptation methods, so that the fake images in target domain that might not have been classified correctly will become detectable after domain adaptation. Furthermore, to address the difference in generation process between GAN-generated images and the natural images, we introduce novel components into the neural network to better exploit the feature of GAN-generated fingerprints to improve detection accuracy, which can also detect the lacking of global feature in GAN-generated images caused by GAN's design.

In summary, the main contributions of this work are as follows:

- We propose a method on the generalization of GAN-generated image detection based on an unsupervised domain adaptation model, which can detect unknown GAN model-generated images by using only a few unlabelled images generated by the unknown GAN model without the large sample dataset.
- A novel loss function is used in the domain adaptation process and a Self-Attention block is imported into the network structure. With these improvements, our method can pay more attention to the fake artifacts caused by the up-sampling during the generated process, thus improving the generalization ability of the detection model.

2. PROPOSED METHOD

Since the generation processes of real and fake images are inherently different, detectable traces are left in the GAN-generated images even having the realistic visual quality. Inspired by [11], we believe that even if there is some difference between different kinds of GAN-generated images, this difference must be smaller than the difference between them and the real images, this allows unsupervised domain adaptation methods to be applied well to generalization problems. We choose Deep Subdomain Adaptation Network(DSAN) [16] as the method of domain adaptation. In order to make DSAN more applicable to solve our problem, we introduce a novel loss function and network components that have proven to be effective in the field of image forensics, which further improve the detection accuracy of the given method. An illustration of the proposed framework is shown in Fig. 2.

2.1. Deep Subdomain Adaptation Network

In the unsupervised domain adaptation, we are given a source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{n_s}$ of n_s labeled samples and a target domain $\mathcal{D}_t = \{\mathbf{x}_j^t\}_{j=1}^{n_t}$ of n_t unlabeled samples. \mathcal{D}_s and \mathcal{D}_t obey different distributions p and q while $p \neq q$. The goal of domain adaptation is to find a mapping $\mathbf{y} = f(\mathbf{x})$ to let the distribution of data in \mathcal{D}_s and \mathcal{D}_t be as similar as possible after this mapping, then the loss function can be expressed as:

$$\min_f \frac{1}{n_s} \sum_{i=1}^{n_s} J(f(\mathbf{x}_i^s), \mathbf{y}_i^s) + \lambda \hat{d}(p, q) \quad (1)$$

where $J(\cdot, \cdot)$ is the classification loss function and $\hat{d}(\cdot, \cdot)$ is domain adaptation loss function. $\lambda > 0$ is the hyperparameters.

Based on the above idea, DSAN defines the concept of subdomain, where similar samples are grouped into a subdomain, after which global alignment is no longer performed, but the relevant subdomain are aligned separately. In the reproducing kernel Hilbert space (RKHS) \mathcal{H} with a characteristic kernel k , $\hat{d}_{\mathcal{H}}(p, q)$ in DSAN is defined as:

$$\hat{d}_{\mathcal{H}}(p, q) = \frac{1}{C} \sum_{c=1}^C \left\| \sum_{\mathbf{x}_i^s \in \mathcal{D}_s} w_i^{sc} \phi(\mathbf{x}_i^s) - \sum_{\mathbf{x}_j^t \in \mathcal{D}_t} w_j^{tc} \phi(\mathbf{x}_j^t) \right\|_{\mathcal{H}}^2 \quad (2)$$

where w_i^{sc} and w_j^{tc} denote the weight of \mathbf{x}_i^s and \mathbf{x}_j^t belonging to class c . The w_i^c for the sample \mathbf{x}_i is computed as:

$$w_i^c = \frac{y_{ic}}{\sum_{(\mathbf{x}_j, \mathbf{y}_j) \in \mathcal{D}} y_{jc}} \quad (3)$$

where y_{ic} is label of \mathbf{y}_i . However, since the data in the target domain is unlabeled, we cannot calculate w_i^c directly. The probability value output by the l^{th} layer of the neural network

is used to generate a pseudo-label for the target domain. Data from \mathcal{D}_s and \mathcal{D}_t will be activated as $\{\mathbf{z}_i^{sl}\}_{i=1}^{n_s}$ and $\{\mathbf{z}_j^{tl}\}_{j=1}^{n_t}$ after layers 1 of the deep networks, respectively. Since the $\phi(\cdot)$ cannot be computed directly, we reformulate \mathcal{L}_{LMMD} as:

$$\begin{aligned} \mathcal{L}_{LMMD} = \hat{d}_l(p, q) = & \frac{1}{C} \sum_{c=1}^C \left[\sum_{i=1}^{n_s} \sum_{j=1}^{n_t} w_i^{sc} w_j^{tc} k(z_i^{sl}, z_j^{tl}) \right. \\ & + \sum_{i=1}^{n_t} \sum_{j=1}^{n_s} w_i^{tc} w_j^{sc} k(z_i^{tl}, z_j^{sl}) \\ & \left. - 2 \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} w_i^{sc} w_j^{tc} k(z_i^{sl}, z_j^{tl}) \right] \end{aligned} \quad (4)$$

where the kernel k means $k(\mathbf{x}^s, \mathbf{x}^t) = \langle \phi(\mathbf{x}^s), \phi(\mathbf{x}^t) \rangle$, where $\langle \cdot, \cdot \rangle$ represents inner product of vectors.

2.2. Loss function

Fake image detection is a typical binary classification problem, in which Binary Cross Entropy(BCE) loss is a commonly used loss function that guides the neural network to learn a decision boundarie from the data given in the training set. The BCE loss is defined as:

$$\mathcal{L}_{BCE} = -(y_{ic} \log(p) + (1 - y_{ic}) \log(1 - p)) \quad (5)$$

where y_{ic} is the label and p is the probability output by the network.

However, BCE loss makes the neural network focus too much on the samples present in the training set, which leads to overfitting to a certain extent, thus reducing the generalization ability of the trained model.

Center loss [17] is an additional auxiliary loss function that is commonly used in face-related fields, this loss function minimizes the distance between each sample within a class and the class center. Center loss is defined as:

$$\mathcal{L}_{center} = \frac{1}{2} \sum_{i=1}^m \|\mathbf{x}_i^s - c_{y_{ic}}\|_2^2 \quad (6)$$

where m is the number of training samples in a mini-batch, \mathbf{x}_i^s denotes the i^{th} training sample in source domain, y_{ic} denotes the label, and $c_{y_{ic}}$ denotes the y_i^{th} class center in the feature space.

When a class of samples is sufficiently similar, the center loss tends to achieve better classification results by tightening the intra-class distance. Each GAN generates images with unique generative traces, using center loss allows the samples to be aggregated in the feature space, which improves the accuracy of domain adaptation methods.

The combined loss function is given as:

$$\mathcal{L} = \lambda \mathcal{L}_{BCE} + \lambda_1 \mathcal{L}_{LMMD} + \lambda_2 \mathcal{L}_{center} \quad (7)$$

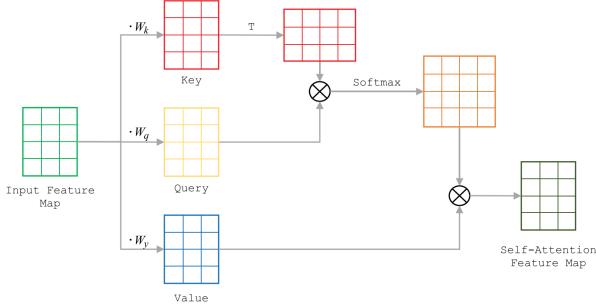


Fig. 3. Graphic illustration of the Self-Attention operation, where T denotes the Transposition operation.

where \mathcal{L} denotes the total loss for the domain adaptation, \mathcal{L}_{BCE} , \mathcal{L}_{LMMD} and \mathcal{L}_{center} denotes the BCE loss, LMMD loss and center loss respectively. λ , λ_1 and λ_2 (λ, λ_1 and $\lambda_2 > 0$) denotes the weight for each loss functions. In this loss function, BCE loss tries to construct decision boundaries between real and fake images, LMMD loss is used to reduce the difference between source and target domains, and center loss tries to make the images more compact in the feature space. We hope that the decision boundaries learned in this way can serve as a better generalization using a small amount of unlabeled data.

2.3. Self-Attention mechanism

Self-Attention [18] was firstly proposed by Google for natural language processing, which integrates more global information instead of focusing only on local information. The idea can be transferred to the field of image forensics because the size of the convolutional kernel is too small compared with the whole image, resulting in it can only cover adjacent regions, thus causing the lacking of global features. The generators in the GAN model focus on learning local features and then generating visually better images through an upsampling process. This local-to-whole structure is inherently different from the global generation mode of natural images, resulting in GAN-generated images that often lack global features. These images may be indistinguishable to the human eye, but they are only combinations of local features without correlation.

The Self-Attention operation is shown in Fig. 3. Let F be the input image feature map to the Self-Attention Network, and f_{xy} is the feature at the location of (x, y). Then we have components of the Query, Key and Value:

$$\begin{aligned} q_{xy} &= f_{xy} \cdot W_q \\ k_{xy} &= f_{xy} \cdot W_k \\ v_{xy} &= f_{xy} \cdot W_v \end{aligned} \quad (8)$$

where W_q, W_k, W_v is the weight matrix to be learned. For

each query vector q_{xy} , the response energy of the key at position (u, v) is as follows:

$$e_{xy}^{uv} = q_{xy} \cdot k_{uv}^T \quad (9)$$

From its manifestation, the response energy obtains the energy between each pixel pair combination from the query matrix to the key matrix, and the global feature map can be taken into account by performing the summation operation, which is exactly the conventional CNN Ignored. Then, normalize on the similarity matrix. Finally, the attention result is calculated as a weighted sum of standardized similarities. The operation is as follows:

$$att_{xy}^{uv} = \text{softmax}(e_{xy}^{uv}) = \frac{\exp(e_{xy}^{uv})}{\sum_{ij} \exp(e_{xy}^{ij})} \quad (10)$$

where att_{xy}^{uv} is the attention value of the key at querying. The overall attention expression form of the elements in the query is the sum of all the elements in the key after the above operations, as shown below:

$$Z_{xy} = \sum_{ij} att_{xy}^{ij} \cdot v_{ij} \quad (11)$$

where Z is the final form of attention. In addition, in order to further improve the learning ability of the attention layer in the algorithm, multi-head attention structure is used.

Multi-head attention is actually a parallel combination of a certain number of attention layers, which can reflect the correlation of feature maps in different feature spaces. The relationship between multi-head attention and single-layer attention is as follows:

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \\ \text{head}_i &= Z(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (12)$$

The multi-head attention method can achieve better performance with similar computational cost.

3. EXPERIMENTAL AND DISCUSSIONS

In this section, we first introduce the dataset used in our experimental and the training details, then compare our model with state-of-the-arts, and validate the effectiveness of our proposed improvements.

3.1. Datasets and experiment setting

To validate our proposed method, the “Align & Cropped” PNG images from CelebA-HQ dataset [19] are used as real images to construct our dataset, since the dataset is training set of PGGAN [19]. The fake images in source domain were generated by PGGAN, which is able to generate realistic images by growing both the generator and discriminator progressively. Since our main aim is to verify the model transferability, in the target domain we use images collecting from a



Fig. 4. Examples of real faces in CelebA-HQ dataset (the first row), fake faces generated by PGGAN (the second row) and fake faces generated by StyleGAN (the third row).

variety of GAN architectures include StyleGAN [20], StyleGAN2 [21], and StarGAN [22]. Several examples of real and fake images are presented in Fig.4. In the experiments, 2000 labeled images in CelebA-HQ and 2000 labeled images generated by PGGAN were selected to construct the source domain. The 30 unlabeled images for each target domain dataset were selected for domain adaptation, and another 3000 images in the same domain were used for testing. All images are preprocessed by resizing into 256×256 and then cropped the 224×224 region randomly. The NVIDIA GeForce RTX 2080 Ti is used to conduct experiments. There are 8 heads of attention, and the size of each head is set to 64.

3.2. Comparison experiments

In this experiment, we evaluated the detection performance of the proposed method and state-of-the-art methods, including Mi's method [5] and Chen's method [9]. Since the existing methods are based on supervised learning that cannot use the introduced unlabeled data, we use the introduction of labeled target domain data and the introduction of no target domain data for each method separately. The detection results in different domains are presented in Table 1, where L for images in target domain is labeled and UL for images in target domain is unlabeled.

As shown in Table 1, a serious problem for existing detection solutions is when the test set contains images that does not exist in the training set, and these supervised learning-based methods cannot take advantage of unlabeled data. Experiments show that our proposed method can effectively utilize unlabeled data from the target domain with a similar detection accuracy compared to supervised learning methods.

3.3. Ablation experiment

We designed ablation experiment to verify that proposed improvements to the domain adaptation method are effective. In this experiment, we choose PGGAN as the source domain,

Table 1. Comparisons of detection accuracies(%) under different GANs

| Method | [5] | | [9] | | Proposed |
|-----------|------|------|------|------|----------|
| | L | UL | L | UL | |
| PGGAN | 99.8 | 99.7 | 99.8 | 99.8 | 99.8 |
| StyleGAN | 52.3 | 50.4 | 98.5 | 61.5 | 97.4 |
| StyleGAN2 | 51.9 | 50.1 | 98.9 | 63.7 | 97.7 |
| StarGAN | 65.8 | 62.2 | 99.2 | 71.9 | 98.0 |

Table 2. Detection accuracies(%) in ablation experiment

| Attention | Pre-train | Loss | Accuracy |
|-----------|-----------|------|----------|
| ✓ | ✓ | ✓ | 97.4 |
| ✓ | ✓ | | 97.0 |
| ✓ | | ✓ | 95.8 |
| ✓ | | | 95.1 |
| | ✓ | ✓ | 95.5 |
| | ✓ | | 94.9 |
| | | ✓ | 92.4 |
| | | | 92.1 |

StyleGAN as the target domain and consider the effect of loss function, backbone network and pre-trained model on the detection accuracy of domain adaptation method separately. We select the GAN-generated image detection model trained by Wang et al. [13] as our pre-training model. The detection accuracy are presented in Table 2, where '✓' means with this condition.

According to the results shown in Table 2, we can conclude that the proposed improvements have better domain generalization capability. The Self-Attention mechanism, which is widely used in the field of image forensics, has also worked well in unsupervised domain adaptation. Moreover, this experiment also shows that the detection accuracy of the domain adaptation model improves when the pre-trained model fits our detection target better, which shows that our strategy can enhance the existing detection model to perform better in the real scene.

4. CONCLUSION

In this paper, we have proposed an unsupervised generalization method for GAN-generated image detection. A Self-Attention block is imported to discover the lacking of global information in fake images and a novel loss function structure is presented to cluster fake images in feature space to facilitate feature alignment, which benefits from the performance of DSAN as well as to focus on the artificial artifacts of GAN-generated images. Experiment shows that proposed method has significant generalization performance, while requiring only a small number of unlabeled images in an unknown domain.

5. REFERENCES

- [1] Huaxiao Mo, Bolin Chen, and Weiqi Luo, “Fake faces identification via convolutional neural network,” in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, 2018, pp. 43–47.
- [2] Michael Goebel, Lakshmanan Nataraj, Tejaswi Nanjundaswamy, Shivkumar Mohammed, and BS Manjunath, “Detection, attribution and localization of GAN generated images,” in *International Symposium on Electronic Imaging*, 2021, pp. 1–11.
- [3] Mauro Barni, Kassem Kallas, Ehsan Nowroozi, and Benedetta Tondi, “CNN detection of GAN-generated face images based on cross-band co-occurrences analysis,” in *IEEE International Workshop on Information Forensics and Security*, 2020, pp. 1–6.
- [4] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu, “Large-scale long-tailed recognition in an open world,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2537–2546.
- [5] Zhongjie Mi, Xinghao Jiang, Tanfeng Sun, and Ke Xu, “GAN-generated image detection with Self-Attention mechanism against GAN generator defect,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 969–981, 2020.
- [6] Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, and Thorsten Holz, “Leveraging frequency analysis for deep fake image recognition,” in *International Conference on Machine Learning*, 2020, pp. 3247–3258.
- [7] Samaksh Agarwal, Nancy Girdhar, and Himanshu Raghav, “A novel neural model based framework for detection of GAN generated fake images,” in *International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 46–51.
- [8] Nicolo Bonettini, Paolo Bestagini, Simone Milani, and Stefano Tubaro, “On the use of Benford’s law to detect GAN-generated images,” in *International Conference on Pattern Recognition*, 2021, pp. 5495–5502.
- [9] Beijing Chen, Xin Liu, Yuhui Zheng, Guoying Zhao, and Yun-Qing Shi, “A robust GAN-generated face detection method based on dual-color spaces and an improved xception,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
- [10] Zhengze Liu, Xiaojuan Qi, and Philip HS Torr, “Global texture enhancement for fake face detection in the wild,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8060–8069.
- [11] Francesco Marra, Diego Gragnaniello, Luisa Verdoliva, and Giovanni Poggi, “Do GANs leave artificial fingerprints?,” in *IEEE Conference on Multimedia Information Processing and Retrieval*, 2019, pp. 506–511.
- [12] Xu Zhang, Svebor Karaman, and Shih-Fu Chang, “Detecting and simulating artifacts in GAN fake images,” in *IEEE International Workshop on Information Forensics and Security*, 2019, pp. 1–6.
- [13] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros, “CNN-generated images are surprisingly easy to spot... for now,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8695–8704.
- [14] Nils Hulzebosch, Sarah Ibrahimi, and Marcel Worring, “Detecting CNN-generated facial images in real-world scenarios,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 642–643.
- [15] 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,” in *IEEE International Conference on Multimedia and Expo*, 2021, pp. 1–6.
- [16] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He, “Deep subdomain adaptation network for image classification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1713–1722, 2020.
- [17] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, “A discriminative feature learning approach for deep face recognition,” in *European Conference on Computer Vision*, 2016, pp. 499–515.
- [18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
- [19] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation,” *arXiv preprint arXiv:1710.10196*, 2017.
- [20] Tero Karras, Samuli Laine, and Timo Aila, “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
- [21] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila, “Analyzing and improving the image quality of StyleGAN,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.
- [22] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo, “StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.