Article

Deep Fake Image Detection based on Pairwise Learning

Chih-Chung Hsu [1], Yi-Xiu Zhuang [1], and Chia-Yen Lee [2]

[1] Department of Management Information System, National Pingtung University of Science and Technology; cchsu@mail.npust.edu.tw
[2] Department of Electrical Engineering, National United University; olivelee@nuu.edu.tw
* Correspondence: olivelee@nuu.edu.tw

Version May 5, 2019 submitted to Preprints

Abstract: Recently, generative adversarial networks (GANs) can be used to generate the photo-realistic image from a low-dimension random noise. It is very dangerous that the synthesized or generated image is used on inappropriate contents in social media network. In order to successfully detect such fake image, an effective and efficient image forgery detector is desired. However, conventional image forgery detectors are failed to recognize the synthesized or generated images by using GAN-based generator since they are all generated but manipulation from the source. Therefore, we propose a deep learning-based approach to detect the fake image by combining the contrastive loss. First, several state-of-the-art GANs will be collected to generate the fake-real image pairs. Then, the contrastive will be used on the proposed common fake feature network (CFFN)to learn the discriminative feature between the fake image and real image (i.e., paired information). Finally, a smaller network will be concatenated to the CFFN to determine whether the feature of the input image is fake or real. Experimental results demonstrated that the proposed method significantly outperforms other state-of-the-art fake image detectors.

Keywords: Forgery detection; GAN; contrastive loss; deep learning; pairwise learning.

---

1. Introduction

Recently, the generative model based on deep learning such as the generative adversarial net (GAN) is widely used to synthesize the photo-realistic partial or whole content of the image and video. Furthermore, recent research of GANs such as progressive growth of GANs (PGGAN)[1] and BigGAN [2] could be used to synthesize a highly photo-realistic image or video so that the human cannot recognize whether the image is fake or not in the limited time. In general, the generative applications can be used to perform the image translation tasks [3]. However, it may lead to a serious problem once the fake or synthesized image is improperly used on social network or platform. For instance, cycleGAN is used to synthesize the fake face image in a pornography video [4]. Furthermore, GANs may be used to create a speech video with the synthesized facial content of any famous politician, causing severe problems on the society, political, and commercial activities. Therefore, an effective fake face image detection technique is desired. In this paper, we have extended our previous study [5] associated with paper ID #1062 to effectively and efficiently address these issues.

In traditional image forgery detection approach, two types of forensics scheme are widely used: active schemes and passive schemes. With the active schemes, the externally additive signal (i.e., watermark) will be embedded in the source image without visual artifacts. In order to identify whether the image has tampered or not, the watermark extraction process will be performed on the target image to restore the watermark[6]. The extracted watermark image can be used to localize or detect the tampered regions in the target image. However, there is no "source image" for the generated images by GANs such that the active image forgery detector cannot be used to extract the watermark image. The second one-passive image forgery detector–uses the statistical information in the source image that will be highly consistency between different images. With this property, the intrinsic statistical information

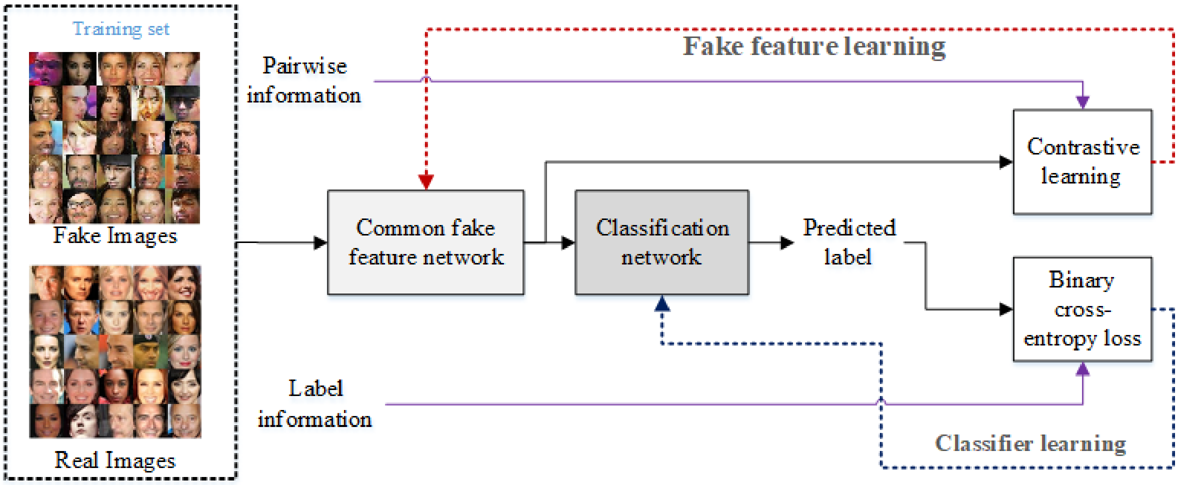


Figure 1. The flowchart of the proposed fake face detector based on the proposed CFFN with the two-step learning approach.

can be used to detect the fake region in the image[7][8]. However, the passive image forgery detector cannot be used to identify the fake image generated by GANs since they are synthesized from the low-dimensional random vector. Nothing change in the generated image by GANs because the fake image is not modified from its original image.

Intuitively, we can adopt the deep neural network to detect the fake image generated by GAN. Recently, there are some studies that investigate a deep learning-based approach for fake image detection in a supervised way. In other words, fake image detection can be treated as a binary classification problem (i.e., fake or real image). For example, the convolution neural network (CNN) network is used to learn the fake image detector [9]. In [10], the performance of the fake face image detection can be further improved by adopting the most advanced CNN–Xception network [11]. However, there are many GANs proposed year by year. For example, recently proposed GANs such as [1][12][13][14][15][16][3][2] can be used to produce the photo-realistic images. It is hard and very time-consuming to collect all training samples of all GANs. In addition, such a supervised learning strategy will tend to learn the discriminative features for a fake image generated by each GANs. In this situation, the learned detector may not be effective for the fake image generated by another new GAN excluded in the training phase.

In order to meet the massive requirement of the fake image detection for GANs-based generator, we propose novel network architecture with a pairwise learning approach, called common fake feature network (CFFN). Based on our previous approach [5], it is clear that the pairwise learning approach can overcome the shortcomings of the supervised learning-based CNN such as methods in [9][10]. In this paper, we further introduce a novel network architecture combining with pairwise learning to improve the performance of the fake image detection. To verify the effectiveness of the proposed method, we apply the proposed deep fake detector (DeepFD) to identify both fake face and generic image. The primary contributions of the proposed method are two-fold:

- We propose a fake face image detector based on the novel CFFN consisting of several dense blocks to improve the representative power of the fake image.
- The pairwise learning approach is first introduced to improve the generalization property of the proposed DeepFD.

The rest of this paper is organized as follows. Sec. II and III present the proposed CFFN for fake face image detection model with pairwise learning for face image and general image. In Section IV, experimental results of the fake face and general images are demonstrated respectively. Finally, conclusions are drawn in Sec. IV.

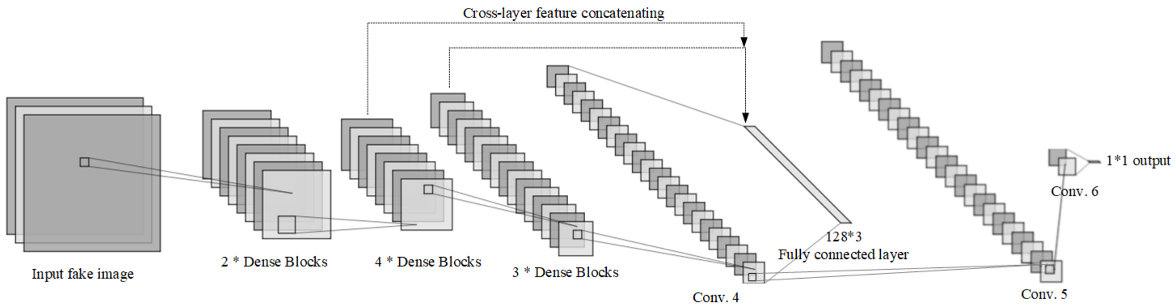


Figure 2. The network architecture of the proposed common fake feature network.

2. Fake Face Image Detection

In the image forgery detection techniques, the most impact of that is the fake face image. With the fake face image, it can be used to synthesize the fake identity to cheat someone else. Once the fake image generator is used to produce some of the celebrities with inappropriate content, it may cause hazardous consequences. In this Section, we will introduce the details of the proposed novel network architecture with pairwise learning policy.

Figure 1 depicts the proposed fake face image detection with the pairwise learning policy. There are two steps in the training phase. The supervised learning policy of the fake face image detection will face the learning difficulty problem since it is hard to collect the training samples generated by all possible GANs. In this step, the fake and real images will be paired to obtain the pairwise information. With the pairwise information, we construct the contrastive loss to learning the discriminative feature over the proposed CFFN. Once the discriminative feature is learned, the classification network will be used to capture the discriminative feature to identify whether the image is real or fake. The details of the proposed method are described in the following paragraphs.

Let the collected training images generated by M GANs be $\mathbf{X}_{fake} = [\mathbf{x}^{k=1}_{i=1}, \mathbf{x}^{k=1}_{i=2}, ..., \mathbf{x}^{k=1}_{i=N_1}, \mathbf{x}^{k=M}_{i=N_M}]$, where each GAN will generate N_k training images. Let the training set of real images indicate by $\mathbf{X}_{real} = [\mathbf{x}_{i=1}, \mathbf{x}_{i=2}, ..., \mathbf{x}_{i=N_r}]$ which contains N_r training images. Therefore, the total number of the training images including the real and fake samples will be $N_T = N_r + N_f = N_r + \sum_{k=1}^{M} N_k$. The labels information $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{N_t}]$ indicates whether the image is fake $(y = 0)$ or real $(y = 1)$. As stated previously, the pairwise information is necessary in the training stage to make the CFFN can learn the powerful discriminative feature. Toward this end, we can generate the pairwise information from the training set $\mathbf{X}$ and its label $\mathbf{Y}$ by the permutation combination. Therefore, we have $C(N_T, 2)$ pairs $P = [p_{i=0,j=0}, p_{i=0,j=1}, ..., p_{i=0,j=N_r}, ..., p_{i=N_f,j=N_r}]$ from the training samples pool. In this paper, we set the total number of pairwise information to $N_r = 2,000,000$.

2.1. Common Fake Feature Network

There are many advanced CNN can be used to learn the fake features from the training samples. For example, Xception Network is used in [10] to capture the powerful feature from the training images in a supervised way. Other advanced CNNs such as DenseNet[17], ResNet[18], Xception[11] network architectures can also be applied on this task. However, most of these advanced CNNs are feed-forward architecture, which the performance of the classification layer (say, Softmax layer) relies on the outputs of the previous layer. It is well known that CNN can capture the hierarchical feature representation from low-level to high-level. In other words, these CNNs only rely on high-level feature representation to identify whether the image is fake or not. However, the fake features of a fake face image may not only existed in the high-level representation but also in the middle-level feature learning. To deal with the cross-level fake feature learning, we have designed a novel CNN architecture to capture the cross-level fake feature from the training images.

In order to have the best performance, we adopt the dense block[17] as the baseline of our proposed common fake feature network. The proposed CFFN consists of three dense units including

Table 1. Network Structures in the Proposed Common Fake Feature Network for Fake Face Image Detection

Layers	Feature Learning	Classification
1	conv. layer, kernel=7×7, stride=4, #channels=96	Conv. layer, kernel=3×3, #channels=2
2	Dense block×2, #channels=96	Global average pooling
3	Dense block×4, #channels=128	Fully connected layer, neurons=2
4	Dense block×3, #channels=256	
5	Fully connected layer, neurons=128	Softmax layer

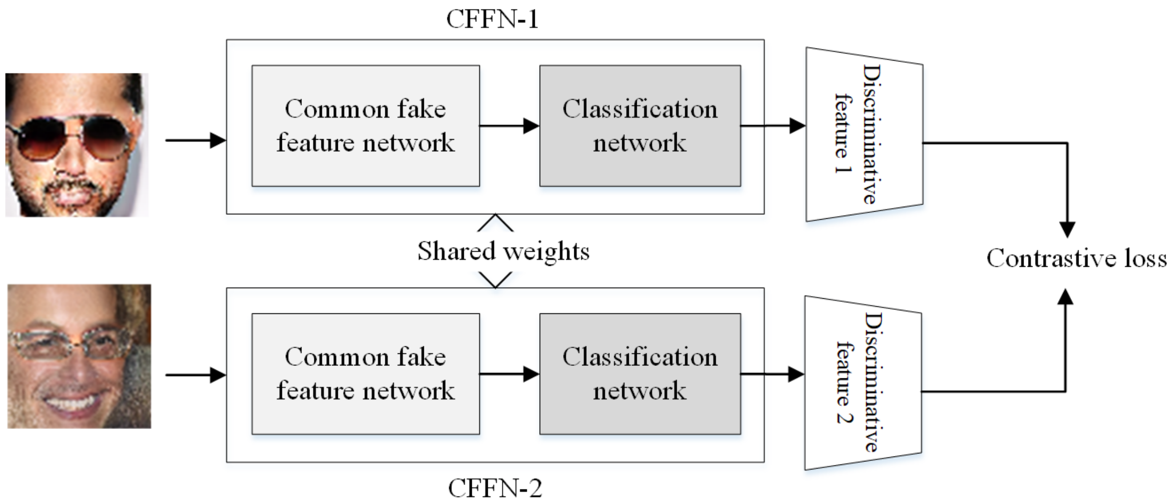


Figure 3. The proposed pairwise learning based on Siamese network architecture and contrastive loss.

two, four, and three dense blocks respectively. The number of the channels in these three dense units are respectively $64, 96, and 128$. The parameter θ in the transition layer is 0.5. Then, we concatenate a convolution layer with 128 channels and 3×3 kernel size to the last layer of the dense unit. Finally, we add the fully connected layer to obtain the discriminative feature representation. To obtain the cross-layer feature representation, we also reshape the last layers of the first and second dense units to aggregate the cross-layer feature into the fully connected layer. Therefore, we have $128 \times 3 = 384$ neurons in the final feature representation.

On the other hand, the classification can be performed by various existed classifiers such as random forest, SVM, Bayes classifier. However, the discriminative feature may be further improved by the back-propagation using the end-to-end architecture. In this paper, therefore, the convolution and fully connected layers are concatenated to the last convolution layer of the proposed CFFN to obtain the final decision result. The details of the proposed CFFN is depicted in Table 1.

2.2. Discriminative Feature Learning

Since the main drawback of supervised learning may be failed to identify the subject that excluded in the learning process. To boost the performance of the proposed method, we introduce contrastive loss to address pairwise learning. Toward this end, the Siamese network architecture[19] should be used for the pairwise information, as depicted in Fig. 3.

To learn discriminative features while training the proposed CFFN, we incorporate the contrastive loss term into the energy function in traditional loss function (i.e., cross-entropy loss). Afterward, given the face image pair $\mathbf{x}_1$ and $\mathbf{x}_2$ and the pairwise label $\mathbf{y}$, where $y = 0$ indicates an impostor pair and $y = 1$ indicates a genuine pair, the energy function between two images will be defined as

$$E_W(\mathbf{x}_1, \mathbf{x}_2) = ||f_{CFFN}(\mathbf{x}_1) - f_{CFFN}(\mathbf{x}_2)||_2^2, \tag{1}$$

The most intuitive way to learn the discriminative feature is to minimize the energy function E_W above. However, directly computing $E_W(\mathbf{x}_1, \mathbf{x}_2)$ by calculating l_2 norm distance in feature domain will lead to a constant mapping. A constant mapping will make any input to a constant vector such that the energy function E_W can be minimized. To overcome this problem, we introduce the contrastive loss to model the pairwise information as:

Follow by the method in [19], we obtain three types of the feature representations: 1) anchor example $f_{CFFN}(\mathbf{x}_a$, 2) positive example $f_{CFFN}(\mathbf{x}_p$, and 3) negative example $f_{CFFN}(\mathbf{x}_n$. The goal of the triple loss is to minimize the distance between the anchor and the positive samples and make the distance between the anchor and the negative samples as larger as possible. Toward this end, the inequality should be:

$$L(W, (P, \mathbf{x}_1, \mathbf{x}_2)) = 0.5 \times (y_{ij} E_w^2) + (1 - y_{ij}) \times max(0, (m - E_w)_2^2), \tag{2}$$

where m is the predefined marginal value, when the input is the genuine pair $y_{ij} = 1$, the cost function will tend to minimize the feature distance E_W between two images. Once the input is impostor pair, the contrastive loss will minimize the $max(0, (m - E_w)$. In other words, the E_W will be maximized if the feature distance between the impostor pair is smaller than the predefined threshold value m. In this manner, it is possible to learn the common characteristic of the fake images generated by different GANs. With the contrastive loss, the feature representation $f_{CFFN}(\mathbf{x}_i)$ will tend to become similar to $f_{CFFN}(\mathbf{x}_j)$ if $y_{ij} = 1$ (i.e., fake-fake or real-real pair). By iteratively train the network f_{CFFN} based on the contrastive loss, the common fake feature of the collected GANs should be able to be well learned.

2.3. Classification Learning

As we stated previously, the classification tasks can be solved by several existed classifiers. In order to boost the performance of the fake image detection, therefore, we adopt a sub-network as the classifier. As a result, the classification learning can be quickly learned by the cross-entropy loss function:

$$L_c(\mathbf{x}_i, p_i) = -\sum_i^{N_T} (f_{CLS}(f_{CFFN}(\mathbf{x}_i)) \log p_i), \tag{3}$$

where f_{CLS} is the classification sub-network consisting of a convolution layer with two channels and a fully connected layer with two neurons. The classifier can be easily trained by back-propagation [20]. With the learning strategies, we may adopt the joint learning incorporating the contrastive loss and the cross-entropy loss to as the total energy function. We also can separately minimize the contrastive loss and the cross-entropy loss instead. The first strategy is difficult to observe the impact of both contrastive and cross-entropy loss functions. Hence, we adopt the second strategy to ensure the best performance of the proposed method. We also note the first learning police as the baseline to have a fair comparison.

3. Fake General Image Detection

Different from the fake face image detection, the general image is more difficult because the contents of the general image are highly varying. Besides, the fake feature of the general image may be more complicated than that of the face image. Toward this end, we increase the number of channels in the CFFN proposed in the above Section. As depicted in Table 2, the total number of dense units is increased to 5. The number of channels in each dense block will also be increased simultaneously for better capturing the fake features of the general image. Similarly, we also adopt contrastive loss and the classification sub-network proposed in Section-II to detect whether the image is fake or real.

Table 2. Network Structures in the Proposed Common Fake Feature Network for Fake General Image Detection.

Layers	Feature Learning	Classification
1	conv. layer, kernel=7×7, stride=4, #channels=96	Conv. layer, kernel=3×3, #channels=2
2	Dense block×2, #channels=128	Global average pooling
3	Dense block×3, #channels=256	Fully connected layer, neurons=2
4	Dense block×4, #channels=384	
5	Dense block×2, #channels=512	
6	Fully connected layer, neurons=128	Softmax layer

Table 3. The subjective performance comparison between the proposed fake face detector and other methods.

Method/Target	LSGAN		DCGAN		WGAN		WGAN-GP		PGGAN	
	precision	recall	precision	recall	precision	recall	precision	recall	precision	recall
Method in [8]	0.205	0.580	0.253	0.774	0.235	0.673	0.242	0.604	0.222	0.862
Method in [9]	0.819	0.528	0.848	0.790	0.817	0.822	0.816	0.679	0.798	0.788
Method in [10]	0.833	0.725	0.812	0.833	0.840	0.809	0.826	0.733	0.824	0.838
Method in [5]	0.947	0.922	0.871	0.844	0.838	0.847	0.818	0.835	0.926	0.918
Baseline-I	0.844	0.831	0.834	0.843	0.840	0.866	0.803	0.819	0.846	0.819
Baseline-II	0.929	0.930	0.867	0.881	0.884	0.901	0.929	0.930	0.901	0.880
The proposed	**0.981**	**0.980**	**0.979**	**0.976**	**0.975**	**0.978**	**0.980**	**0.954**	**0.933**	**0.955**

4. Experimental Results

4.1. Fake Face Image Detection

The dataset used in this paper is extracted from CelebA[21]. The images in CelebA covers large pose variations and background clutter including 10,177 number of identities and 202,599 aligned face images. In experiments, we have collected five state-of-the-art GANs to produce the training set of fake images based on the CelebA dataset:

- DCGAN (Deep convolutional GAN) [13]
- WGAP (Wasserstein GAN) [14]
- WGAN-GP (WGAN with Gradient Penalty) [15]
- LSGAN (Least Squares GAN) [16]
- PGGAN [1]

Each GAN will be used to generate $200,000$ fake images sized of 64×64 into the fake image pool. In PGGAN, we adopt the best model released by the authors of the corresponding GAN. However, the PGGAN [1] can be used to generate the high-resolution fake face image, which the size of the generated face image is different from our experimental setting. In this paper, we downsample the fake face image generated by PGGAN to 64×64. Then, we randomly pick $192,599$ fake images from the fake face image pool. Finally, we have $385,198$ training images and $10,000$ test images consisting of the real and fake images for training.

In the learning setting of the proposed CFFN and fake face detector, we set the learning rate $= 1e - 3$, and the total number of epochs is 15. The marginal value m in contrastive loss is 0.5. Adam optimizer [22] is used for both first and second steps learning. The epochs of the first learning of CFFN is 2, and the number of epochs of the classification learning of the classification sub-network is 11. The batch size is 32 for all learning tasks. The parameters settings in [9][10][5] are the same with the descriptions in their papers.

In the experiments, we also adopt the conventional image forgery method based on sensor pattern noise [8] for performance comparison. The Baseline-I method is the jointly learning of the contrastive loss and the binary cross-entropy loss without two-step learning. The Baseline-II method remove the

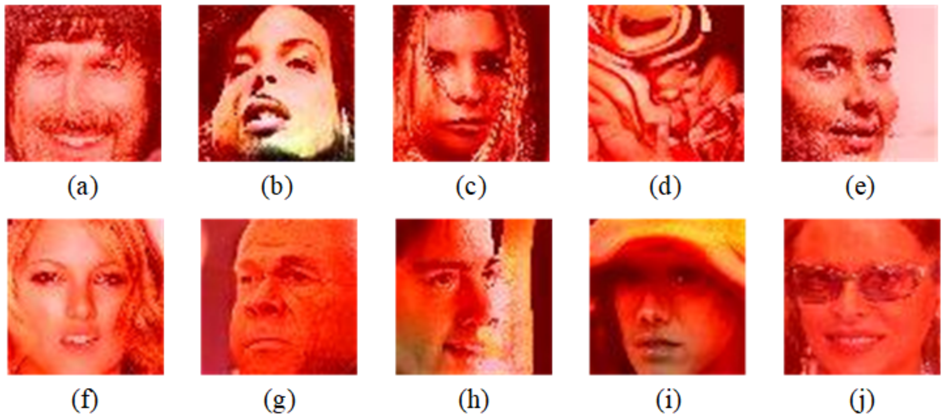


Figure 4. The visualized fake feature map for fake regions localization for face image. (a) - (c) are the fake face images generated by PGGAN[1] and (f) - (j) are the fake image generated by DCCAN[13] respectively.

CFFN architecture and adopt the DenseNet[17] with two-step learning of the contrastive and binary cross-entropy loss functions.

4.1.1. Objective Performance Comparison

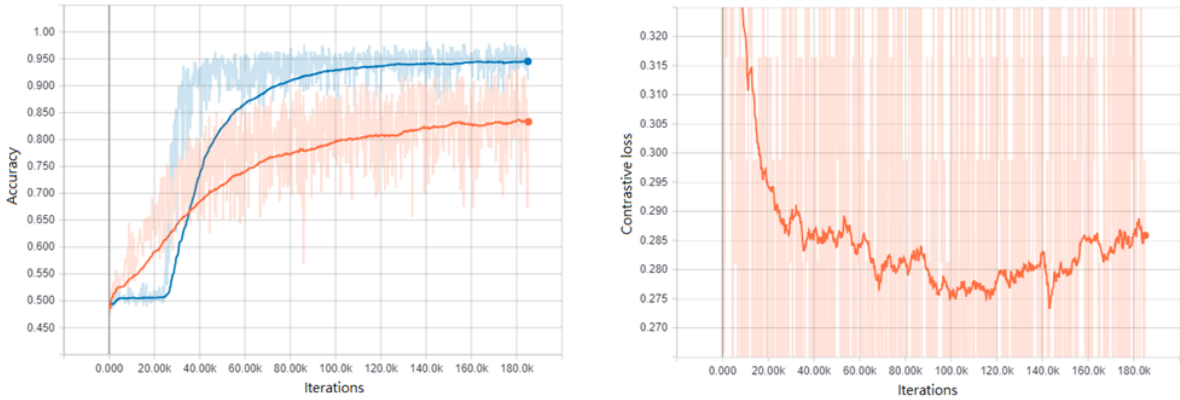
In order to verify the effectiveness of the proposed method, we exclude one of the collected GANs as the training set and perform the testing process on that GAN. For example, we may exclude PGGAN during the training phase of the proposed deep fake detector. Afterward, the real images and fake images generated by PGGAN will be used to evaluate the performance of the learned fake face detector. Table 3 demonstrates the objective performance comparison between the proposed fake face detector, other baseline methods, and methods in [9][10][**?**] in terms of precision and recall. The proposed method significantly outperforms other state-of-the-art methods due to the CFFN can be used to capture the discriminative feature of the fake images well. Furthermore, the proposed pairwise learning successfully captures common fake feature over the training images generated by different GANs. It is also verified that the proposed method is more generalized and effective than others.

4.1.2. Visualized Result

Inspired by [23], the object can be localized by designing the number of channels in the last convolution layers to the number of the classes. As the suggestion in [23], the channels of the last convolutional layer of the proposed CDNN to enable the visualization ability. Therefore, the proposed model can be used to visualize the fake regions of the generated image by extracting the last convolution layer and mapping the responses to the image domain. Since the last convolution layer is designed to two channels, the first channel can be regarded as the feature response of the first class (i.e., real image) and the second channel is corresponding to the second class (i.e., fake image). In this way, the proposed method can be used to visualize the fake region, making a more intuitive interpretation of the typical fake features generated by GANs. As the results shown in Fig.4, the primary artifacts of the fake face images in Fig. 4 are also drawn in red color.

4.1.3. Training Convergence

In the proposed method, it is necessary to guarantee the convergence of network learning. In this Subsection, we will discuss the convergence of the contrastive loss and CFFN learning respectively. As shown in Fig. 5 (a), the orange line depicts the accuracy curve during training for supervised learning without contrastive loss. The blue line indicates the accuracy curve of the training process of the proposed pairwise learning. It is clear that the proposed pairwise learning can be significantly converged, compared to the supervised learning strategy. On the other hand, it is necessary to ensure



(a) (b)

Figure 5. (a) The learning curves of the proposed method with contrastive loss and supervised learning respectively. (b) The energy term of the proposed pairwise learning method of CFFN learning.

Table 4. The subjective performance comparison between the proposed fake general image detector and other methods.

Method/Target	BIGGAN		SA-GAN		SN-GAN	
	precision	recall	precision	recall	precision	recall
Method in [8]	0.358	0.409	0.430	0.509	0.354	0.424
Method in [9]	0.580	0.673	0.610	0.723	0.585	0.691
Method in [10]	0.650	0.737	0.682	0.762	0.653	0.691
Method in [5]	0.734	0.763	0.775	0.782	0.743	0.747
Baseline-I	0.769	0.789	0.787	0.811	0.798	0.791
Baseline-II	0.826	0.803	0.827	0.854	0.810	0.822
The proposed	**0.909**	**0.865**	**0.930**	**0.936**	**0.934**	**0.900**

the contrastive loss can be converged as well. Figure 5 (b) depicts the convergence of the proposed contrastive loss during CFFN training. As a result, we can observe that the proposed contrastive loss is successfully decreased after several iterations.

4.2. Fake General Image Detection

In fake general image detection, we have collected three state-of-the-art GANs to generate high-quality general images as the follows:

- BIGGAN (Large Scale GAN Training for High Fidelity Natural Image Synthesis) [2]
- SA-GAN (Self-Attention GAN) [24]
- SN-GAN (Spectral Normalization GAN) [25]

The dataset used in this Section is extracted from ILSVRC12[26]. We adopt the source code provided by the authors and its released model which is trained on ILSVRC12 to generate the fake general images. Each GAN will be used to generate 200,000 fake images sized of 128×128 into the fake general image pool. Then, we randomly pick 300,000 fake images from the fake face image pool. Finally, we have 600,000 training images and 10,000 test images consisting of the real and fake images for training. The parameters in network learning are the same as the previous Section.

Table 4 demonstrates the objective performance comparison between the proposed method and other state-of-the-art image forgery detection schemes. It is clear that proposed CFFN with pairwise learning police is significantly better than other state-of-the-art image forgery detectors. Compared to the supervised learning-based method in [9][10], the performance of the proposed method significantly outperforms others. Based on these observations, we also prove that the proposed method can learn the common fake feature of the fake general image.

5. Conclusion

In this paper, we have proposed a novel common fake feature network based the pairwise learning, to detect the fake face/general images generated by state-of-the-art GANs successfully. The proposed CFFN can be used to learn the middle- and high-level and discriminative fake feature by aggregating the cross-layer feature representations into the last fully connected layers. The proposed pairwise learning can be used to improve the performance of fake image detection further. With the proposed pairwise learning, the proposed fake image detector should be able to have the ability to identify the fake image generated by a new GAN. Our experimental results demonstrated that the proposed method outperforms other state-of-the-art schemes in terms of precision and recall rate.

Author Contributions: Conceptualization, Chih-Chung Hsu; Data curation, Yi-Xiu Zhuang; Formal analysis, Chih-Chung Hsu; Funding acquisition, Chih-Chung Hsu; Investigation, Chih-Chung Hsu; Methodology, Chih-Chung Hsu; Resources, Yi-Xiu Zhuang; Software, Yi-Xiu Zhuang; Supervision, Chia-Yen Lee; Validation, Yi-Xiu Zhuang; Writing – original draft, Chih-Chung Hsu.

Funding: This work was supported in part by the Ministry of Science and Technology of Taiwan under grant MOST 107-2218-E-020-002-MY3.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolution neural network
CFFN	Common fake feature network
DeepFD	Deep fake image detector
GAN	Generative adversarial net

References

1. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv Preprint, arXiv:1710.10196* **2017**.
2. Brock, A.; Donahue, J.; Simonyan, K. Large scale gan training for high fidelity natural image synthesis. *arXiv Preprint, arXiv:1809.11096* **2018**.
3. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv Preprint,* **2017**.
4. AI can now create fake porn, making revenge porn even more complicated,. http://theconversation.com/ai-can-now-create-fake-porn-making-revenge-porn-even-more-complicated-92267, 2018.
5. Hsu, C.; Lee, C.; Zhuang, Y. Learning to detect fake face images in the Wild. 2018 International Symposium on Computer, Consumer and Control (IS3C), 2018, pp. 388–391. doi:10.1109/IS3C.2018.00104.
6. H.T. Chang, C.C. Hsu, C.Y.a.D.S. Image authentication with tampering localization based on watermark embedding in wavelet domain. *Optical Engineering* **2009**, *48*, 057002.
7. Hsu, C.C.; Hung, T.Y.; Lin, C.W.; Hsu, C.T. Video forgery detection using correlation of noise residue. Proc. of the IEEE Workshop on Multimedia Signal Processing. IEEE, 2008, pp. 170–174.
8. Farid, H. Image forgery detection. *IEEE Signal Processing Magazine* **2009**, *26*, 16–25.
9. Huaxiao Mo, B.C.; Luo, W. Fake Faces Identification via Convolutional Neural Network. Proc. of the ACM Workshop on Information Hiding and Multimedia Security. ACM, 2018, pp. 43–47.
10. Marra, F.; Gragnaniello, D.; Cozzolino, D.; Verdoliva, L. Detection of GAN-Generated Fake Images over Social Networks. Proc. of the IEEE Conference on Multimedia Information Processing and Retrieval, 2018, pp. 384–389. doi:10.1109/MIPR.2018.00084.
11. Chollet, F. Xception: Deep learning with depthwise separable convolutions. *Proc. of the IEEE conference on Computer Vision and Pattern Recognition* **2017**, pp. 1610–02357.

12. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. Proc. of the Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
13. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv Preprint, arXiv:1511.06434* **2015**.
14. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein gan. *arXiv Preprint, arXiv:1701.07875* **2017**.
15. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. Proc. of the Advances in Neural Information Processing Systems, 2017, pp. 5767–5777.
16. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Smolley, S.P. Least squares generative adversarial networks. Proc. of the IEEE International Conference on Computer Vision. IEEE, 2017, pp. 2813–2821.
17. G. Huang, Z. Liu, L.v.d.M.; Weinberger, K. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
18. K. He, X. Zhang, S.R.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
19. S. Chopra, R.H.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2005, Vol. 1, pp. 539–546.
20. LeCun, Y.; Boser, B.E.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.E.; Jackel, L.D. Handwritten digit recognition with a back-propagation network. Proc. of the Advances in Neural Information Processing Systems, 1990, pp. 396–404.
21. Z.i Liu, P. Luo, X.W.; Tang, X. Deep Learning Face Attributes in the Wild. Proc. of the International Conference on Computer Vision, 2015.
22. Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. International Conference on Machine Learning, 2013, pp. 1139–1147.
23. M. Oquab, L. Bottou, I.L.; Sivic, J. Is object localization for free?-weakly-supervised learning with convolutional neural networks. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 685–694.
24. H. Zhang, I. Goodfellow, D.M.; Odena, A. Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318* **2018**.
25. T. Miyato, T. Kataoka, M.K.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957* **2018**.
26. et.al., B.O. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)* **2015**, *115*, 211–252. doi:10.1007/s11263-015-0816-y.

999

Sample Availability: The source code and the partial samples can be found at https://github.com/jesse1029/Fake-Face-Images-Detection-Tensorflow