# FAKE GENERATED PAINTING DETECTION VIA FREQUENCY ANALYSIS

*Yong Bai*⋆    *Yuanfang Guo*⋆    *Jinjie Wei*⋆    *Lin Lu*⋆    *Rui Wang*† ‡    *Yunhong Wang*⋆

⋆ School of Computer Science and Engineering,
Beihang University, Beijing, China
† State Key Laboratory of Information Security,
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
‡ School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

## ABSTRACT

With the development of deep neural networks, digital fake paintings can be generated by various style transfer algorithms. To detect the fake generated paintings, we analyze the fake generated and real paintings in Fourier frequency domain and observe statistical differences and artifacts. Based on our observations, we propose Fake Generated Painting Detection via Frequency Analysis (FGPD-FA) by extracting three types of features in frequency domain. Besides, we also propose a digital fake painting detection database for assessing the proposed method. Experimental results demonstrate the excellence of the proposed method in different testing conditions.

***Index Terms***— Image Forgery Detection, Frequency Analysis, Style Transfer, Paintings, Fourier Transform

## 1. INTRODUCTION

Painting has been a major type of artworks in the past few centuries. Many artists, such as Van Gogh, Monet, etc., have contributed numerous paintings with their unique styles. Recently, deep neural networks are developed to transfer the styles of the artists/paintings to other natural images and thus generate realistic digital paintings. This generation can mainly be achieved via two mechanisms. On one hand, [1–5] propose that content and style can be reconstructed based on the image features extracted from convolutional neural networks (CNN). On the other hand, the emergence of generative adversarial networks (GAN) provides another approach to perform style transfer [6–8].

As shown in Fig. 1, the generated paintings are also decent and can hardly be distinguished from the artists' work by ordinary people, i.e., people who lack the necessary expertise. Unfortunately, these generated paintings may be utilized by
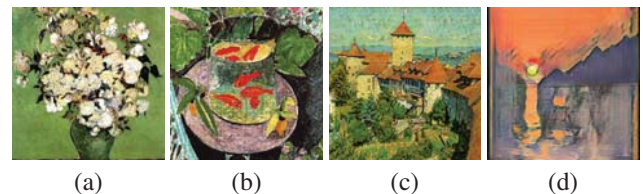
**Fig. 1**: Examples of real and fake generated paintings: (a) Real, Van Gogh's style; (b) Real, Matisse's style; (c) Fake, Van Gogh's style, generated by CycleGAN [6]; (d) Fake, Matisse's style, generated by CycleGAN [6].

some malicious people to cheat the customers. Meanwhile, the artists' rights may also be violated. Therefore, it is necessary to identify whether a painting is generated (forged).

Unfortunately, traditional image forgery detection techniques are designed for traditional forgeries including copy-move, splicing and image retouching, which are different from style transfer in principle. Recently, several approaches have been proposed to detect the emerging forgeries, such as colorized images [9], GAN generated faces [10, 11] and computer graphics based generated images [12]. However, they may also be inappropriate, because they are designed to distinguish natural images, which are different from paintings, from fake images based on the corresponding forging mechanisms. Currently, to the best of our knowledge, there exists no forgery detection method specifically designed for fake generated painting detection.

Considering that the existing style transfer techniques only possess constraints designed in spatial domain, the fake generated paintings may be more distinguishable in frequency domain. Therefore, we transform the real and fake generated paintings into Fourier domain to perform a frequency analysis. Based on our observation that the fake generated paintings reveal obvious inconsistencies in frequency domain compared to real paintings, we propose a forgery detection method, named Fake Generated Painting Detection via Frequency Analysis (FGPD-FA).

Our major contributions can be summarized as below: 1)
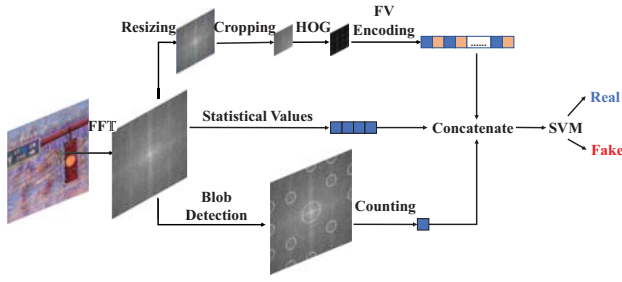
**Fig. 2**: Framework of the proposed FGPD-FA.



(a) Real      (b) AAMS    (c) CycleGAN (d) GatedGAN

**Fig. 3**: Averaged spectrum of the paintings.



**Fig. 4**: Values of the averaged spectrum's diagonal line.

We observe that there are differences between the real and fake generated paintings in frequency domain; 2) Based on the observed statistical differences between the spectrums of the real and fake generated paintings, we extract the statistical and oriented gradient features for detection; 3) Based on the observed artifacts in the spectrum of fake generated paintings, we adopt a blob detection algorithm [13] to extract the blob feature for detection; 4) We build a painting database, named Digital Painting Forgery Detection (DPFD) Database, which contains 10310 real paintings from fifteen famous artists and 58292 fake paintings generated via three state-of-the-art style transfer methods [5, 6, 8].

## 2. METHODOLOGY

In this section, the proposed Fake Generated Painting Detection via Frequency Analysis (FGPD-FA), whose framework is shown in Fig. 2, will be introduced in details.

### 2.1. Analysis and Observations

The differences between the fake generated and real paintings are not obvious, as Fig. 1 demonstrates. Since the state-of-the-art style transfer techniques do not constrain their outputs in frequency domain, a frequency analysis is performed to better reveal the differences between the fake generated and real paintings. Naturally, Fourier transform is employed to transfer the paintings from spatial domain to frequency domain. Note that the colour space of the painting image will be transferred from RGB space to YUV space and only the Y component will be employed in the frequency analysis, for convenience. According to the experiments, two major observations can be obtained:

1) Statistical differences: The averaged distribution of the spectrum values are different between the fake generated and real paintings. We randomly select 1000 spectrums of the real paintings and 3000 spectrums of the fake generated paintings, where each one third of the fake paintings are generated by a state-of-the-art forging method. Then, we show the averaged spectrum in Fig. 3 and the values on the diagonal line of the averaged spectrum in Fig. 4. As shown in Fig. 3, the difference of averaged spectrum values at middle frequencies
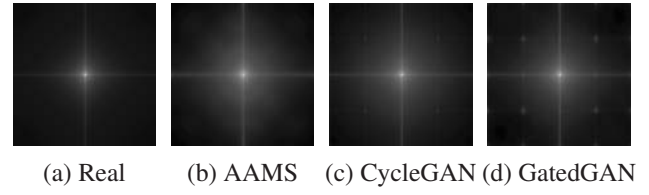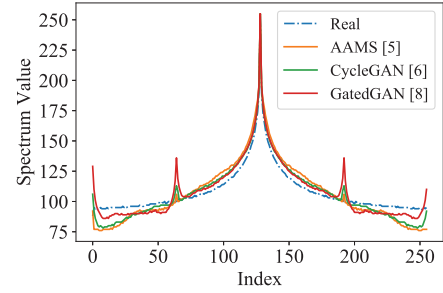
between the fake generated and real paintings is obvious. In Fig. 4, the x-axis represents the indexes of the diagonal line from the top-left corner to the bottom-right corner, while the y-axis represents the averaged spectrum values. As can be observed, the spectrum values of the real paintings possess less variations and do not present spike-like artifacts compared to that of the fake generated paintings.

2) Artifacts: Inspired by [14], which discovered that certain noisy artifacts regularly appear in the GAN generated images, we analyze the GAN generated paintings in frequency domain. According to our experiments, grid-like artifacts, where each crossing point appears to be a blob, appear regularly in the spectrum of GAN based generated paintings. Fig. 5 presents a generated painting with visible artifacts and its detected blobs.

### 2.2. Statistical Feature

Based on the first observation in Sec. 2.1, we perform a simple yet effective univariate analysis to describe the statistical distribution of the spectrum. Given a normalized spectrum denoted by $M$, let $S = (s_1, s_2, ..., s_L)$ be the one-dimensional vector containing all the elements of $M$. Let $L$ denote the length of $S$. Four statistical values with different orders, including mean, standard deviation, skewness and kurtosis, are calculated and then concatenated to form the statistical feature for FGPD-FA:

- Mean: $\bar{s} = \frac{1}{L}\sum_{l=1}^{L} s_l$

- Standard deviation: $\sigma_s = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(s_l - \bar{s})^2}$

- Skewness: $skewness = \frac{\sum_{l=1}^{L}(s_l - \bar{s})^3 / L}{\sigma_s^3}$

1257

- Kurtosis: $kurtosis = \frac{\sum_{l=1}^{L}(s_l - \bar{s})^4 / L}{\sigma_s^4} - 3$

## 2.3. Oriented Gradient Feature

According to Sec. 2.1, the changing slope of the spectrum of the fake generated paintings varies more obviously from low to high frequencies, compared to that of the real paintings. Thus, local oriented gradients can be exploited to model these inconsistencies. FGPD-FA adopts the classic yet effective Histogram of Oriented Gradients (HOG) [15] to capture the local oriented gradients.

Before the extraction of HOG vectors, considering the symmetry of the spectrum, FGPD-FA resizes the width and height of the spectrum by half and then crops the top-left quarter of the resized spectrum for further processing and feature extraction. Then, a Gaussian filter is applied to smooth the resized and cropped spectrum. Although [15] shows that smoothing with large masks tends to decrease the performance, we empirically observe that it tends to be helpful to smooth the spectrum with large masks (e.g. $9 \times 9$) in frequency domain. After the smoothing, we divide the preprocessed spectrum into cells. For each cell, a histogram of gradient directions is computed as a HOG descriptor.

To better represent the spectrum, with the computed HOG descriptors, we encode them into a Fisher vector [16]. For each image (spectrum), let $H = \{x_i | i = 1, ..., N\}$ be a set of D dimensional HOG features, where N denotes the number of cells and D denotes the number of orientations. Then, the corresponding Fisher vector can be computed by

$$FV_\Theta = \frac{1}{N}\sum_{i=1}^{N} L_\Theta \frac{\partial \log G(x_i | \Theta)}{\partial \Theta}, \qquad (1)$$

where $G$ denotes the probability distribution function of a Gaussian Mixture Model (GMM). Here, GMM is utilized to fit the distribution of the extracted HOG descriptors. Note that $\Theta = \{w_k, \mu_k, \Sigma_k, k = 1, ..., K\}$ represents the parameters of GMM, where $K$ stands for the number of Gaussian distributions and $w_k, \mu_k, \Sigma_k$ represent the weight, mean vector and covariance matrix of Gaussian distribution k, respectively. $L_\Theta$ represents the Cholesky decomposition of the Fisher Information Matrix [17]. In our experiments, an practical implementation of Fisher Vector Encoding [18] is employed for convenience.

## 2.4. Blob Feature

Since regular grid-like artifacts with blobs can be observed in frequency domain of the fake generated paintings, to better utilize this trace, we adopt the Difference of Gaussian (DOG) [13] method, which is an efficient approximation to the Laplacian of Gaussian (LOG) in blob detection [19], to detect the blobs by finding the extrema. In our method, the number of detected blobs is exploited as the blob feature.
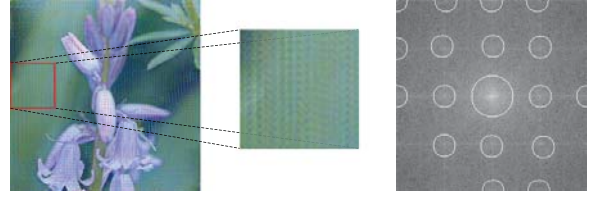


**Fig. 5**: Fake generated paintings with noisy artifacts and its spectrum with the detected blobs.

**Table 1**: Details of the testing datasets.

| Index | Artists | Forging Algorithms | Real Images | Fake Images |
|-------|---------|--------------------|-------------|-------------|
| $T_\alpha^1$ | Renoir | All | 693 | 708 |
| $T_\alpha^2$ | Van Gogh | All | 376 | 705 |
| $T_\alpha^3$ | Picasso | All | 484 | 705 |
| $T_\beta^1$ | All | AAMS [5] | 1007 | 1013 |
| $T_\beta^2$ | All | CycleGAN [6] | 1007 | 1044 |
| $T_\beta^3$ | All | GatedGAN [8] | 1007 | 1020 |
| $T_\gamma^1$ | All | All | 2006 | 2092 |

## 2.5. Procedures of FGPD-FA

Here, we briefly summarize the procedures of FGPD-FA, whose framework is shown in Fig. 2. For both the training and testing stages, the given image is firstly transferred into Fourier frequency domain, where all the subsequent operations are performed. Next, the statistical feature, oriented gradient feature and blob feature are extracted and then concatenated into a feature vector. Support vector machine (SVM) is employed as the classifier. Note that the implementation of SVM in [20] is employed here. In the training process, GMM model is estimated based on the training data and SVM is trained with the extracted features and corresponding labels. In the testing process, the trained GMM and SVM are directly utilized in predictions.

## 3. EXPERIMENTAL RESULTS

In this section, our database, the experimental setups, evaluation results and an ablation study are correspondingly introduced in details.

## 3.1. Database

In this paper, we build a database, named Digital Painting Forgery Detection (DPFD) Database, which contains 58292 fake generated paintings and 10310 real paintings with the resolution of $256 \times 256$. The real paintings are from fifteen famous artists. To generate the fake paintings, natural images are selected from multiple image datasets [21–25]. The fake paintings are generated via three state-of-the-art style trans-

1258

**Table 2**: Performance on different datasets.

| | $T_\alpha^1$ | $T_\alpha^2$ | $T_\alpha^3$ | $T_\beta^1$ | $T_\beta^2$ | $T_\beta^3$ | $T_\gamma^1$ |
|---|---|---|---|---|---|---|---|
| $Acc(\%)$ | 97.50 | 93.99 | 93.19 | 95.05 | 94.10 | 97.68 | 94.85 |
| $F_1(\%)$ | 97.48 | 91.91 | 91.91 | 94.82 | 93.87 | 97.61 | 94.62 |

fer methods, including AAMS [5], CycleGAN [6], and GatedGAN [8].

In the experiments, both the training and testing datasets are constructed based on our DPFD database. Note that the training and testing datasets are not overlapping. The training dataset contains 4992 fake paintings with all the styles of fifteen artists, which are generated by all the three methods, and 4128 real paintings from all the fifteen artists. To evaluate the proposed method from different perspectives, seven testing datasets are constructed, as shown in Table 1. Note that we only select three representative artists to assess the proposed method when it faces different forged styles.

### 3.2. Setups and Metrics

For the oriented gradient feature, we set the Gaussian filter size as $9 \times 9$, the number of orientations as 9, the cell size as $16 \times 16$, the block size and the window size as $64 \times 64$. L2-norm is employed as the block normalization method. Note that the power law equalization is not applied. Besides, when fitting GMM, the number of Gaussian distributions are chosen to be 16 by Bayesian Information Criterion [26]. For the blob feature, to neglect the very small blobs, we set the minimum standard deviation for the Gaussian kernel as 11, the maximum standard deviation as 30, and the ratio as 1.8. In SVM, the RBF kernel is selected. The regularization parameter and the kernel coefficient are selected by grid search strategy.

In the evaluation, the accuracy ($Acc$) and F1-score ($F_1$) are employed as measurements [27]. Note that the fake generated paintings are defined as negative samples and the real paintings are defined as positive samples.

### 3.3. Quantitative Evaluation Results

The proposed FGPD-FA is evaluated with the datasets in Table 1 and the quantitative results are shown in Table 2. The results indicate that our method performs decently for both the CNN-based method ($T_\beta^1$) and GAN-based methods ($T_\beta^2$, $T_\beta^3$), and it can be used to detect the fake generated paintings with specific styles ($T_\alpha^1$-$T_\alpha^3$). When the fake images are generated based on all the fifteen artists and three forging methods, FPGD-FA still gives satisfactory performance, which demonstrates the excellence of the proposed FPGD-FA.

### 3.4. Ablation Study

To verify the effectiveness of each feature, an ablation study is performed on the testing dataset $T_\gamma^1$. The results are reported
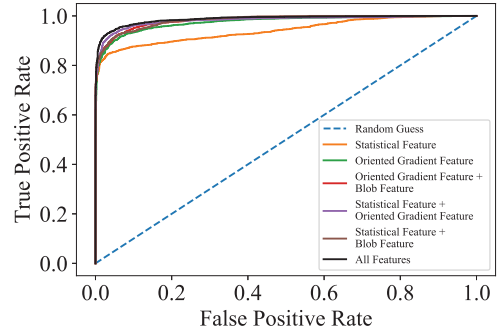


**Fig. 6**: ROC curves for dataset $T_\gamma^1$.

**Table 3**: Ablation Study.

| Statistical Feature | ✓ | | | ✓ | ✓ | ✓ |
|---|---|---|---|---|---|---|
| **Oriented Gradient Feature** | | ✓ | ✓ | ✓ | | ✓ |
| **Blob Feature** | | | ✓ | | ✓ | ✓ |
| **Acc(%)** | 89.87 | 92.31 | 92.90 | 93.94 | 92.65 | **94.85** |
| **F1(%)** | 88.60 | 91.87 | 92.53 | 93.64 | 92.15 | **94.62** |

in Table 3 and the corresponding ROC curves are shown in Fig. 6. As can be observed, the proposed method can identify the fake generated paintings from the real paintings with either the statistical feature or the oriented gradient feature. Note that the blob feature is not evaluated alone because it is mainly designed to detect the GAN-based generated paintings. When each two of the features jointly work together, obvious improvements can be obtained. When all the three features are utilized, the proposed FGPD-FA gives the best performance. According to the results, the effectiveness of each feature can be verified.

## 4. CONCLUSION

In this paper, we aim to design a forgery detection approach to identify the fake generated paintings, which are forged by style transfer techniques, from the real paintings. We firstly analyze the fake generated and real paintings in Fourier frequency domain. Based on two major observations, we propose Fake Generated Painting Detection via Frequency Analysis by extracting the statistical, oriented gradient and blob features in frequency domain. We also propose a new database for fake generated painting detection. Extensive experiments not only demonstrate the excellence of the proposed FGPD-FA, but also verify the effectiveness of each extracted feature.

1259

# 5. REFERENCES

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE CVPR*, 2016, pp. 2414–2423.

[2] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, 2016, pp. 694–711.

[3] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in *Proc. ICML*, 2016, pp. 1349–1357.

[4] T. Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," *CoRR*, vol. abs/1612.04337, 2016.

[5] Y. Yao, J. Ren, X. Xie, W. Liu, Y. Liu, and J. Wang, "Attention-aware multi-stroke style transfer," in *Proc. IEEE CVPR*, 2019, pp. 1467–1475.

[6] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE ICCV*, 2017, pp. 2242–2251.

[7] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. ICML*, 2017, pp. 1857–1865.

[8] X. Chen, C. Xu, X. Yang, L. Song, and D. Tao, "Gatedgan: Adversarial gated networks for multi-collection style transfer," *IEEE TIP*, vol. 28, no. 2, pp. 546–560, 2019.

[9] Y. Guo, X. Cao, W. Zhang, and R. Wang, "Fake colorized image detection," *IEEE TIFS*, vol. 13, no. 8, pp. 1932–1944, 2018.

[10] H. Li, B. Li, S. Tan, and J. Huang, "Detection of deep network generated images using disparities in color components," *CoRR*, vol. abs/1808.07276, 2018.

[11] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "Face x-ray for more general face forgery detection," *CoRR*, vol. abs/1912.13458, 2019.

[12] Y. Yao, W. Hu, W. Zhang, T. Wu, and Y. Shi, "Distinguishing computer-generated graphics from natural images based on sensor pattern noise and deep learning," *Sensors*, vol. 18, no. 4, pp. 1296, 2018.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[14] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi, "Do gans leave artificial fingerprints?," in *Proc. IEEE MIPR*, 2019, pp. 506–511.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE CVPR*, 2005, pp. 886–893.

[16] J. Sánchez, F. Perronnin, T. Mensink, and J. J. Verbeek, "Image classification with the fisher vector: Theory and practice," *IJCV*, vol. 105, no. 3, pp. 222–245, 2013.

[17] T. S. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in *Proc. NIPS*, 1998, pp. 487–493.

[18] "Fisher vectors based on gaussian mixture model," https://pypi.org/project/fishervector/, Access Date: Dec. 30, 2019.

[19] T. Lindeberg, "Feature detection with automatic scale selection," *IJCV*, vol. 30, no. 2, pp. 79–116, 1998.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.

[21] T. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," in *Proc. ECCV*, 2014, pp. 740–755.

[22] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE CVPR*, 2007.

[23] M. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. IEEE CVPR*, 2006, pp. 1447–1454.

[24] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *Proc. IEEE CVPR*, 2008, pp. 1–8.

[25] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *IJCV*, vol. 42, no. 3, pp. 145–175, 2001.

[26] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[27] M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations," *IJDKP*, vol. 5, no. 2, pp. 1, 2015.