**MaxPix: Enhancing GAN-Generated Image Detection by Emphasizing Local Maxima**

Ronghao Dai1†★, Lingxi Peng1†\*★, Haohuai Liu2†★

1School of Computer Science and Network Engineering, Guangzhou University, Guangzhou, 510000, Guangdong, China.

2Analysis and test center, Guangzhou University, Guangzhou, 510000, Guangdong, China.

★address:University Town Campus No. 230, Waihuan East Road, Panyu District, Guangzhou, Guangdong Province, China.

\*Corresponding author(s) E-mail(s)[: manplx@163.com](mailto:manplx@163.com);

Contributing authors: [670482535@qq.com](mailto:670482535@qq.com); [huaihuai99@163.com](mailto:huaihuai99@163.com)

†These authors contributed equally to this work.

**Abstract**

The realistic images generated by Generative Adversarial Networks (GANs) enrich various applications but also pose significant threats to personal privacy and society. As a result, developing algorithms capable of accurately detecting GAN-generated images has become crucial. Existing detection methods often rely on artifacts specific to certain GAN models, limiting their cross-model generalization performance. In this paper, we propose MaxPix, a novel algorithm that leverages statistical features and deep learning techniques to detect GAN-generated images. MaxPix first applies the MaxSel filtering algorithm to obtain a filter map of the image, which is then used as input to the MResNet feature extraction network. MResNet, an enhanced version of ResNet, incorporates MA blocks to emphasize local maxima in the feature maps. Experimental results on the Wang and Faces-HQ datasets demonstrate that MaxPix achieves average detection accuracies of 85.9% and 99.6%, respectively, outperforming state-of-the-art algorithms like AIDE by 5.8% and 3.8%. These results indicate that MaxPix exhibits strong cross-model generalization performance. The source code and datasets are available at <https://github.com/DarlingDiving/MaxPix/>**,** complete with detailed usage guidelines to facilitate the reproduction of our results..

**Keywords:** GAN, Generative Images; Artifacts, Aross-model Generalization

**1 INTRODUCTION**

Digital images have become one of the main carriers for transmitting network information due to their diverse content and convenient storage. They are widely used in fields such as news, information, medical diagnosis, and identification. GANs (Generative Adversarial Networks), a generative model based on deep learning technology, were proposed by Ian Goodfellow et al.[1]. GAN consists of a generator and a discriminator, where the generator produces samples that closely resemble real data. To date, over a hundred different GAN models capable of generating images have been developed. These GANs can be applied in various aspects of life to enhance the quality of people's lives. For example, Shi et al.[2] proposed SG-GAN for image restoration, which improves image quality., Yang et al.[3] proposed LDAF-GAN for generating high-quality floral images, enriching people's lives., and Cheema et al.[4] applied GAN to text translation, improving the accuracy of translations. However, some individuals maliciously use GANs to forge images for political or pornographic purposes, posing a serious threat to personal privacy and society. Given that digital images are widely used in various fields, ensuring their authenticity is crucial. To prevent the abuse of GAN-generated images and mitigate their societal impact, effective detection algorithms are needed to distinguish between real and generated images. Currently, researchers have proposed numerous detection algorithms, which are mainly categorized into those based on traditional digital image forensic methods and those relying on deep learning techniques.

In detection algorithms based on traditional digital image forensics, researchers primarily design algorithms to detect generated images by analyzing the properties of digital images, such as illumination inconsistencies and statistical features in the spatial and frequency domains. McCloskey et al.[5] analyzed the process of color formation in images and argued that the normalization process in GANs restricts the pixel range in generated images, causing the exposure of these images to differ from that of real images. They proposed using the measured frequency of overexposure and underexposure as a feature to detect generated images. However, this algorithm only achieves an AUC (Area Under Curve) value of 0.7. Durall et al.[6] found that the high-frequency components of generated images are distorted and proposed using the azimuthal integral of the image as a feature to detect generated images through a support vector machine, achieving 100% accuracy. However, the algorithm lacks cross-model generalization performance. Guo et al.[7] observed that the eye pupils in real face images are elliptical, while those in generated face images are irregular. They proposed an algorithm that determines whether an image is generated by calculating the IoU (Intersection over Union)[8] values of the pupil region and an elliptical mask, using the IoU value to classify the image. This algorithm has strict requirements regarding the quality and angle of the image, and it may misjudge images with physiological defects. Liu et al.[9] used the Sobel operator to compute the image gradient in HSV (Hue, Saturation, Value) space and then counted histograms of the gradient distribution as features to detect generated images. The algorithm achieved 99.4% accuracy when detecting images generated by PGGAN[10], but its cross-model generalization performance was not investigated.

Detection algorithms based on traditional image forensics have both theoretical and experimental foundations. However, such algorithms are highly susceptible to overfitting statistical features that exist only in the training set, while different GAN-generated images exhibit different statistical features. As a result, these algorithms tend to have lower accuracy when detecting unknown GAN-generated images. Additionally, these algorithms require the images to conform to specific angles and quality, which further limits their applicability.

On the other hand, detection algorithms based on deep learning techniques use neural networks to construct algorithmic models and learn general features from massive datasets to detect generated images. Since neural networks have strong representational abilities, these algorithms generalize well, which has led to increased scholarly interest. The up-sampling process is common in GANs, and Zhang et al.[11] designed AutoGAN, which contains an up-sampling process to generate a variety of images that simulate GAN-generated images. These images were used to train the algorithm. However, the detection accuracy of the algorithm may degrade significantly if the up-sampling method used by the GAN differs substantially from that used by AutoGAN. Liu et al.[12] found that the phase spectrum of an image retains rich frequency components and proposed combining spatial domain features with phase features to detect generated images. The algorithm achieved accuracy rates of 91.5% and 76.88% on two Deepfake datasets[13,14]. Jeong et al.[15] proposed an algorithm that uses a high-pass filter to remove irrelevant features in both the spatial and frequency domains, highlighting important features for detecting generated images. This algorithm achieves more than 72% cross-model detection accuracy and average precision. Tian et al.[16] divided the image frequency components into low, medium, and high components, then aggregated these features with the original image to detect generated images, achieving an accuracy of 97.74%. Wang et al.[17] used wavelet transform to map the image from the spatial domain to the frequency domain, then extracted high-frequency components and fused these features with the original image. The algorithm detected generated images using Xception[18] and achieved more than 98% accuracy, but the accuracy decreased for low-quality images. Miao et al.[19] designed the Center Differential Attention Transformer to enable the algorithm to learn global high-frequency information and local fine-grained features. They also designed a high-frequency wavelet sampler to help the algorithm extract multi-channel high-frequency features. The proposed algorithm aggregated the two features to detect generated images, but its accuracy was lower when detecting compressed or processed images.

Algorithms based on deep learning techniques generally rely on detecting the artifacts introduced by the imperfect design of GANs to identify generated images. However, as GAN structures have improved, the obvious artifacts in generated images have been effectively hidden. Moreover, the artifacts generated by different GANs vary, which limits the generalization performance of artifact-dependent detection algorithms. As a result, these algorithms often exhibit low accuracy when detecting unknown GAN-generated images and lack broad applicability.

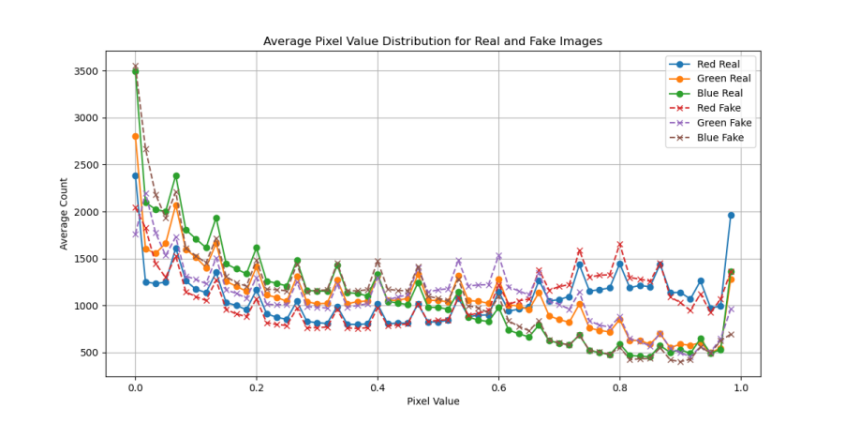
In light of this, this thesis proposes exploring detection algorithms that do not rely on artifacts for detecting generated images. Specifically, this thesis analyzes the pixel value distributions of GAN-generated images produced by models such as StarGAN[20] and StyleGAN2[21], and compares them to real images in datasets like FFHQ[22] and CelebA. It is observed that the generated images cannot replicate the pixel distribution of real images, and real images tend to have more points with larger pixel values than generated ones. Based on this observation, this thesis introduces the MaxPix detection algorithm, which utilizes statistical features. First, the MaxSel algorithm is proposed for filtering images. Then, the MA Block is embedded in ResNet to create MResNet, which extracts features from the filtered images to detect generated images. Extensive experiments demonstrate the effectiveness of MaxPix in detecting generated images. The main contributions of this thesis are as follows:

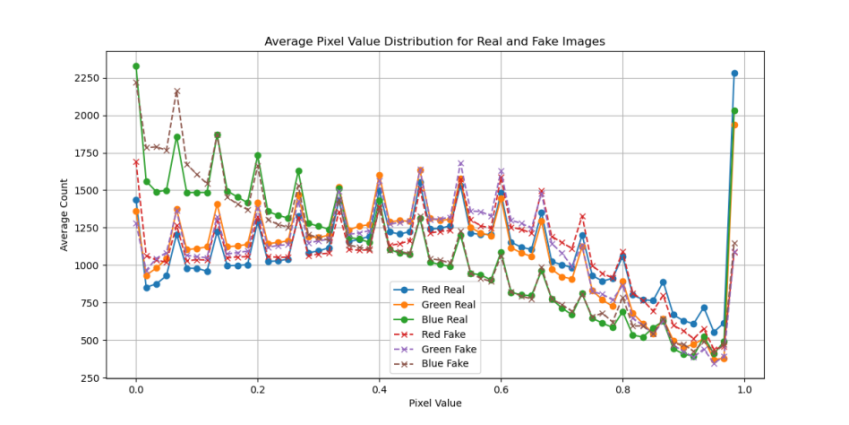
1. Based on the observation that GAN-generated images cannot replicate the pixel value distribution of real images, the MaxPix detection algorithm is proposed to detect GAN-generated images, with the MaxSel algorithm designed for filtering images.

2) MaxPix achieves an average accuracy of 85.9% and 99.6% on the Wang[23] and Faces-HQ[6] datasets, showing an improvement of 5.8% and 3.8%, respectively, compared to current state-of-the-art detection algorithms. Therefore, MaxPix demonstrates strong cross-model generalization performance.

**2 ALGORITHM DESCRIPTION**

Durall et al.[6] found that GAN-generated images cannot reproduce the spectral distribution of real images. Similarly, He et al.[24] discovered that generated images exhibit stronger nonlocal similarity than real images. These findings inspired this thesis to investigate whether there are differences in pixel distributions between generated and real images.





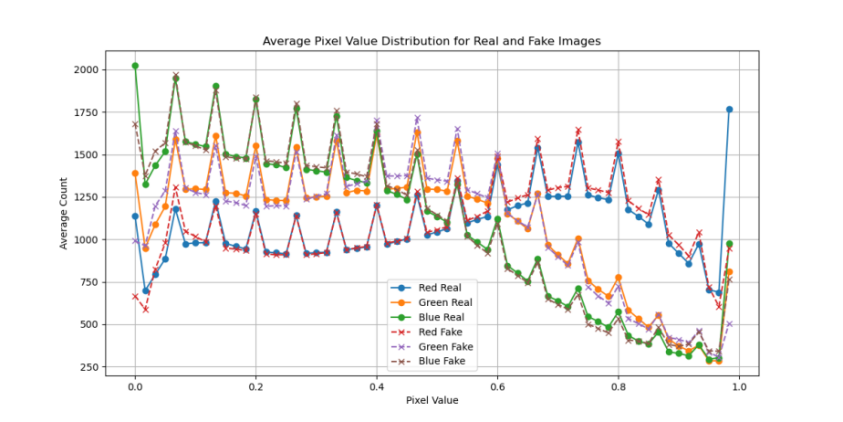


Fig.1 Statistical distribution of pixel values of BigGAN, StarGAN and StyleGAN2 generated images and real images

To explore this, the frequency of pixel values in each pixel value range was counted and displayed using histograms. In the experiment, pixel values were divided into 60 groups. A total of 34k images were analyzed, including images generated by BigGAN[25], StarGAN, and StyleGAN2, as well as real images sampled from ImageNet[26], CelebA, and FFHQ datasets, sourced from the Wang dataset[23] and Faces-HQ[6]. As shown in Fig.1, despite the fact that these GANs were trained with large numbers of real images, they still struggle to replicate the pixel value distribution of real images. Real images contain more points in the higher pixel value range than generated images. Based on this observation, this thesis proposes the MaxPix detection algorithm, which detects generated images by emphasizing local maxima and utilizing these maxima as features.

**2.1 Algorithmic framework**

As shown in Fig.2, the MaxPix structure consists of a filtering module(feature select module), a feature extraction network MResNet, and a classifier C. The filtering module is a feature extraction network. The filtering module uses the MaxSel filtering algorithm proposed in this thesis to perform filtering on the image, making it easy for MResNet to learn distinguishable features to detect GAN-generated images.



Fig.2 MaxPix framework where the Feature select module does not update parameters

**2.2 MaxSel filter**

MaxPix uses the convolution kernel, as shown in equation(1), as a filter kernel to perform convolution with the image and obtain the filtered image. First, MaxPix splits the image channel-by-channel and applies the convolution operation using four convolution kernels. This results in the convolution values *X*(*c*, *i*, *j* )(*α*1, *α*2, *α*3, *α*4 ) for each point in the three channels, computed in four directions. Then, MaxSel compares the convolution values in the four directions at each corresponding location and selects the largest value among them as the filter value. For *X*(*c*, *i*, *j* )(*α*1, *α*2, *α*3, *α*4 ), the maximum value is chosen from *α*1, *α*2, *α*3 and *α*4.

 (1)

As shown in Equation(2), where *X*(*c*, *i*, *j* ) denotes the filter value at the position (*i*, *j* ) of the image c channel. The filter value of each point constitutes the filter map of that channel. MaxSel concancat the filter maps of the three channels to form the filter map *Fin*∈R3×*H×W* .

 (2)

As shown in Fig.3, the first column displays the real image from the Wang dataset[23]. The second column shows the filter map obtained using the Prewitt operator as the convolution kernel. The third column presents the filter map obtained by applying the Laplacian operator as the convolution kernel, and the fourth column illustrates the filter map obtained using MaxSel. The filter map generated by MaxSel is more detailed and complete, which allows the algorithm to learn more comprehensive features from it.

**2.3 MResNet feature extraction network**

As shown in Fig.4, MResNet is an improved version of ResNet, with five additional MA blocks consisting of a maximum pooling filter layer, a mean filter layer, and a residual layer. MResNet replaces the mean pooling in the final output with maximum pooling, which is used to select the most significant features for detecting the GAN-generated images.

MA block is used to emphasize the local maxima in the feature map as shown in equation(3), where *λ* is an updatable parameter. *Fin* denotes the input features. *MP* denotes maximum pooling. *AP* denotes mean filtering. *Abs* denotes taking absolute values.

 (3)

|  |  |  |  |
| --- | --- | --- | --- |
| 04490 | output_image_407 | output_image_407 | output_image_407 |
| 03848 | output_image_366 | output_image_366 | output_image_366 |
| 00053 | output_image_7 | output_image_7 | output_image_7 |

Fig.3 Filtering effect image. Each column from left to right corresponds to the real image, filtered image of Prewitt, Laplacian and MaxSel.



Fig.4 MA block embedded in Basic block of the ResNet to form MResNet

**2.4 Classifiers and loss functions**

The classifier C consists of two fully connected layers. MaxPix spreads the 8192 features output from MResNet and then transforms them into predicted values using the fully connected layers. As shown in Equation(4), where C is the classifier, *y* denotes the true label of the images, and *Fd* is the input features.

 (4)

**3 EXPERIMENTAL**

This thesis demonstrates the improvement of MaxPix in cross-model generalization performance by comparing its accuracy and average precision with those of current representative detection algorithms across different datasets. The role of the MaxPix modules is validated through ablation experiments.

**3.1 Datasets**

To avoid misunderstandings in expression, this thesis uses the lowercase English name of the GAN to refer to the dataset composed of the corresponding generated images and real images. For example, images generated by StyleGAN and the real images used for training the GAN are referred to as the stylegan dataset.

The Wang dataset: Wang et al.[23] released a publicly available, but unnamed dataset, which is referred to as the Wang dataset in this thesis. This dataset is divided into a training set, an evaluation set, and a test set, and contains both real and generated images. The real images were sampled from LSUN[27], ImageNet, and other datasets commonly used to train GANs. The generated images include 20 different scenarios, created by GANs such as PGGAN, StyleGAN2, as well as fake faces sampled from the FaceForensics++ (deepfake)[28] dataset. Since its release, this dataset has been widely used by researchers[15,23,29-32] to train and evaluate detection algorithms.

The Faces-HQ dataset: Durall et al.[6] released the Faces-HQ dataset. Each image in the Faces-HQ dataset has a resolution of 1024×1024, which is significantly higher than that of the Wang dataset. Faces-HQ contains 20,000 real face images sampled from CelebA-HQ[10] and FFHQ, as well as 20,000 generated images sampled from the 100K Faces project[33] and *[www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com)*. These generated images were produced by StyleGAN and StyleGAN2. CelebA-HQ and FFHQ are commonly used datasets for training GANs and are recognized as benchmarks for training and testing detection algorithms.

In this thesis, the "person" subset of the Wang training set is used to train PixMSE, while the biggan, gaugan, and stargan subsets of the Wang test set, as well as the entire Faces-HQ dataset, which contains over 102k images, are used as test sets.

**3.2 Experimental environment**

In this thesis, the algorithm code is written in Python 3.7 with PyTorch 1.9.0. The GPU used is an RTX 3090, and the system is Ubuntu. MaxPix performs resizing and random cropping on the training set, and resizing and center cropping on the test set, which changes the input image to *X*∈R3×299×299.

The training setup includes 36 epochs and a batch size of 4. The optimizer used is Adam, with a learning rate of 0.00005, a learning rate decay factor of 0.96, and the loss function is CrossEntropyLoss.

**3.3 Comparative experiments**

In this thesis, we select research works that have achieved high accuracy in the task of detecting GAN-generated images in recent years for comparison, including Wang et al.[23], Frank et al.[34], Durall et al.[6], Jeong et al.[15, 29], He et al.[24], Deng et al.[35], Guo et al.[36], Yan et al.[37], and Tan et al.[38]. These algorithms not only achieved better performance in their respective studies, with over 90% accuracy in detecting the same type of GAN-generated images, but they also maintained strong cross-model generalization performance.

Except for the algorithm by Jeong et al.[15, 29], the rest of the algorithms in this thesis were retrained and tested using the Wang dataset. Since Jeong et al.'s[15, 29] algorithm uses the Wang dataset and the code implementation details are not available, the experimental data in the table are quoted from their literature[29].

As shown in table 1, MaxPix achieves high accuracy for detecting biggan, cyclegan, stargan, and stylegan datasets, which are higher than the highest values achieved among the compared algorithms. In particular, compared to compared algorithms, MaxPix achieves an accuracy improvement of 6.1% for detecting biggan and 2.4% for detecting stylegan. MaxPix, like most of the compared algorithms, achieves a lower accuracy of 63% for detecting the gaugan dataset. In terms of average precision performance, MaxPix detects gaugan with an average precision of 75.5% and detects stylegan2 with an average precision of 99.6%, which is lower than the best of the compared algorithms at 97.6% and 99.9%. However, MaxPix detects the remaining six datasets all get the highest average precision, equaling or exceeding the best of the compared algorithms. It can be seen that the detection performance of MaxPix is better than the current mainstream detection algorithms in terms of accuracy and average precision.

.

Tab1 Comparison experiment Wang dataset (%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | progan | | biggan | | cyclegan | | deepfake | | gaugan | | stargan | | stylegan | | stylegan2 | |
|  | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP |
| Wang[23] | 81.2 | 97.9 | 50.8 | 67.5 | 60 | 86.9 | 53 | 61.8 | 55 | 88.8 | 56 | 86 | 52 | 76.8 | 52.4 | 68.3 |
| Frank[34] | 98.7 | **99.9** | 67 | 89.1 | 51 | 69.7 | 58.9 | 73.8 | 65.0 | **97.6** | 85.8 | 99.9 | 71.3 | 82.3 | 58.2 | 71.2 |
| Durall[6] | 66 | 80.1 | 67 | 73.4 | 39.7 | 42.7 | 50.3 | 53.6 | 64.9 | 75.2 | 69.1 | 94.6 | 75.4 | 85.8 | 68.3 | 74.9 |
| He[24] | 88.5 | 99.1 | 75.9 | 85.4 | 79.9 | 88.4 | 51.7 | 77.9 | 50.6 | 49 | 99.5 | 100 | 76.1 | 90.6 | 59.5 | 82.6 |
| Jeong[15] | 82.5 | 81.4 | 67 | 62.5 | 75.5 | 74.2 | 51.6 | 49.9 | **73.6** | 92.1 | 90.1 | 90.1 | 68 | 62.8 | 68.8 | 63.6 |
| Jeong[29] | 95.5 | 99.4 | 63.5 | 60.5 | 59.4 | 59.9 | 70.4 | 81.5 | 53 | 49.1 | 99.6 | 100 | 80.6 | 90.6 | 77.4 | 93.0 |
| Deng[35] | 94.2 | 97.8 | 63.6 | 68.3 | 60.6 | 70.9 | **77.0** | 84.5 | 57.6 | 60.4 | 95.4 | 99.6 | 85.6 | 90.8 | 92.6 | 97.3 |
| Guo[36] | 98.6 | **99.9** | 59.3 | 69.6 | 62.4 | 81.1 | 59.4 | 76.8 | 54.9 | 67.3 | 98.2 | 100 | 85.6 | 92.9 | 88.4 | 96.4 |
| Yan[37] | 91.3 | 98.6 | 65.9 | 76.4 | 80.9 | 91.7 | 56.3 | 73.5 | 59 | 65.4 | 98.0 | 99.7 | 94.8 | 98.6 | **94.6** | 98.9 |
| Tan[38] | **99.5** | **99.9** | 65.6 | 83.7 | 64.2 | 92.3 | 70.8 | 90.1 | 63.7 | 84.9 | 75.5 | 99.1 | 93 | **99.8** | 91.6 | **99.9** |
| MaxPix | 98.1 | **99.9** | **82** | **93.2** | **83.5** | **93.4** | 69 | **95.4** | 63 | 75.5 | **100** | **100** | **97.2** | **99.8** | 94.5 | 99.6 |

As shown in Table 2, the average precision and accuracy of the algorithms for detecting the Faces-HQ dataset vary significantly. Since the implementation details of the Jeong[15,29] algorithm are not available, these two algorithms are not included in Table 2. Although the training and testing are conducted on two different datasets with a significant difference in image resolution, and the algorithms are not retrained in this thesis, MaxPix still performs well, achieving 99.9% and 99.3% detection accuracy and 100% and 99.9% average precision, respectively, which outperform the comparison algorithms. This indicates that MaxPix's detection accuracy and average precision are less affected by image size.

Tab2 Comparison experiment Faces-HQ (%)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Faces-HQ | | | | | |
|  | StyleGAN,CelebA-HQ | | StyleGAN2,FFHQ | | ave | |
|  | Acc | AP | Acc | AP | Acc | AP |
| Wang[23] | 49.7 | 45.1 | 51.9 | 74.2 | 50.8 | 59.7 |
| Frank[34] | 67.2 | 78.2 | 58.3 | 63.6 | 62.7 | 70.9 |
| Durall[6] | 57.2 | 93.6 | 62.9 | 91 | 60.0 | 92.3 |
| He[24] | 65.1 | 85.0 | 70.2 | 96.1 | 67.6 | 90.6 |
| Deng[35] | 79.9 | 99.2 | 77.8 | 93.3 | 78.9 | 96.3 |
| Guo[36] | 96.4 | 99.8 | 82.3 | 97.8 | 89.4 | 98.8 |
| Yan[37] | 95.5 | 98.3 | 96.0 | 98.7 | 95.8 | 98.5 |
| Tan[38] | 91.9 | 97.9 | 75.0 | 97.8 | 83.5 | 97.8 |
| MaxPix | **99.9** | **100** | **99.3** | **99.9** | **99.6** | **100** |

Additionally, in Fig.5, the images in the Wang dataset exhibit obvious artifacts, while the images in Faces-HQ do not show obvious artifacts. This further indicates that MaxPix's detection accuracy is less influenced by artifacts. Comparison experiments demonstrate that MaxPix’s accuracy and average precision can match or exceed the current state-of-the-art detection algorithms, showcasing strong cross-model generalization performance.







Fig5.The first row is from the real image of Faces-HQ, the second row is from the generated image of Faces-HQ, and the third row is from the generated image of Wang dataset.

**3.4 Ablation experiments**

This thesis explores the role of MaxSel filtering and MA Block through ablation experiments. The modular ablation experiments use ResNet as a benchmark for comparison. ‘ResNet’ takes the unfiltered image as the input to ResNet. ‘MResNet’ takes the unfiltered image as the input to MResNet, which explores the role of MaxSel. ‘MSel’ filters the image through MaxSel and uses it as input to ResNet to explore the role of MA Block.

Tab3 Module ablation experiment (%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | progan | | biggan | | cyclegan | | deepfake | | gaugan | | stargan | | stylegan | | stylegan2 | |
|  | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP |
| ResNet | 85.1 | 93.4 | 50.6 | 55.4 | 64 | 70 | 48.4 | 48.3 | 57 | 65.7 | 69.8 | 80 | 67.8 | 73.5 | 82.6 | 90.7 |
| MResNet | 79.1 | 88.1 | 52 | 56.4 | 60.7 | 66.2 | 48.2 | 48.3 | 61 | 68.5 | 71.5 | 86.2 | 65.7 | 72.4 | 82.9 | 91.8 |
| MaxSel | 98 | **99.9** | 79.2 | 90.2 | 67 | 81.5 | **88.9** | **96.2** | 54.7 | 63.5 | **100** | **100** | **97.4** | **99.8** | 94.4 | 99.4 |
| MaxPix | **98.1** | **99.9** | **82** | **93.2** | **83.5** | **93.4** | 69 | 95.4 | **63** | **75.5** | **100** | **100** | 97.2 | **99.8** | **94.5** | **99.6** |



**3.4.1 Module ablation experiments**

As shown in table 3, ResNet only detects stylegan2 and progan with more than 80% accuracy and more than 90% average precision. MResNet does not improve the accuracy and average precision of detecting the generated images despite the addition of MA Block, meaning MA Block alone does not improve the algorithm's performance. Due to the adoption of MaxSel for filtering the image, which makes it easy for the algorithm to learn distinguishable features from the filtered images, the detection accuracy and average precision of MSel are comprehensively improved, especially for detecting deepfake, which improves the accuracy by 40.5% and the average precision by 47.9%. MaxPix introduces MA Block to MSel to detect progan, biggan, cyclegan, gaugan, and stylegan2 with 0.1%, 2.8%, 16.5%, 8.3%, and 0.1% accuracy improvements, respectively. There is a slight decrease in the average precision of MaxPix in detecting deepfake. It can be seen that Maxsel used with MA Block effectively improves the accuracy and average precision of the detection algorithm in detecting the generated images, and it is the Maxsel filter that plays the biggest role.

**3.4.2 Network Structure Ablation Experiments**

In this ablation experiment, filtered images obtained by different filtering algorithms, such as Laplacian, Sobel, Prewitt and Scharr, are used as inputs for MResNet and ResNet to further explore the need for the proposed MaxSel filtering algorithm.

As shown in table 4 and table 5, the detection algorithm uses Maxsel to filter the images and achieves the highest accuracy and average precision on multiple datasets regardless of whether MResNet or ResNet is used as the network architecture. Especially for the detection of stargan, which algorithm consistently achieves 100% accuracy and average precision. The accuracy for the detection of gaugan is consistently lower, at 63% and 54.7%, and the average precision was only obtained as 75.5% and 63.5%. However, even when the image is filtered using other operators, the detection algorithm has a low accuracy and average precision for detecting gaugan with maximum accuracy of 70.3% and average precision of only 80%. This indicates that by filtering the image, it is less helpful to improve the accuracy and average precision of the algorithm when detecting gaugan.

Tab4 Network structure ablation experiment-MResNet (%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | progan | | biggan | | cyclegan | | deepfake | | gaugan | | stargan | | stylegan | | stylegan2 | |
|  | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP |
| laplacian | 98.1 | **99.9** | 79.4 | 92.6 | 83.1 | **95.6** | 60.3 | 76.4 | 62 | 77.4 | **100** | **100** | 94.2 | 98.6 | **95.9** | **99.6** |
| prewitt | 98.1 | **99.9** | 56 | 66.8 | 56.7 | 76.6 | 50.6 | 69.4 | 51.3 | 64.4 | 83.6 | 99.7 | 82.8 | 94.2 | 84.3 | 95.5 |
| sobel | 98.5 | **99.9** | 65.9 | 75.4 | 65.4 | 74.7 | **80.5** | 91.3 | **69.4** | **78.8** | 89.5 | 99.8 | 87.4 | 96.2 | 91.1 | 97.8 |
| scharr | **98.9** | **99.9** | 64 | 70.9 | 67.9 | 78.5 | 72.9 | 85.7 | 67.7 | 77.1 | 94.5 | **100** | 86.9 | 95.5 | 89.5 | 97.2 |
| MaxSel | 98.1 | **99.9** | **82** | **93.2** | **83.5** | 93.4 | 69 | **95.4** | 63 | 75.5 | **100** | **100** | **97.2** | **99.8** | 94.5 | **99.6** |

Tab5 Network structure ablation experiment-ResNet(%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | progan | | biggan | | cyclegan | | deepfake | | gaugan | | stargan | | stylegan | | stylegan2 | |
|  | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP | Acc | AP |
| laplacian | 96.7 | 99.8 | 78.2 | **92.2** | 67.1 | 85.4 | 64.4 | 80.3 | 53 | 58.9 | 98.5 | **100** | 93.4 | 99.3 | 91.3 | 99.2 |
| prewitt | 97.9 | **99.9** | 61.6 | 72.7 | 66.8 | 81 | 72.1 | 88.7 | **70.3** | **80** | 87.1 | 99 | 86.2 | 94.3 | 86.5 | 96.2 |
| sobel | **98.8** | **99.9** | 65.8 | 76.3 | 66.6 | 81.6 | 81.5 | 88.3 | 67.8 | 76 | 92.6 | 99.7 | 86 | 90.1 | 88.7 | 95.8 |
| scharr | 97.7 | 99.8 | 68.3 | 79.7 | **76.8** | **86.7** | 74.3 | 85.9 | 67 | 76.3 | 99.6 | **100** | 85.3 | 90.5 | 91.4 | 97.8 |
| MaxSel | 98 | **99.9** | **79.2** | 90.2 | 67 | 81.5 | **88.9** | **96.2** | 54.7 | 63.5 | **100** | **100** | **97.4** | **99.8** | **94.4** | **99.4** |

Overall, the two ablation experiments show that Maxsel and MA Block are more helpful in improving the accuracy and average precision of the algorithms to detect the GAN-generated images, especially Maxsel filtering can efficiently improve the generalization performance of the detection algorithms.

**4 CONCLUSION**

This thesis proposes MaxPix for detecting GAN-generated images, an algorithm that generates features for detecting generated images by emphasizing the maximum value within the local range of the image. The main contribution of this thesis is the proposal of the MaxSel filtering algorithm and the MaxPix detection algorithm. Comparison experiments on the Wang and Faces-HQ datasets show that MaxPix outperforms state-of-the-art algorithms such as Guo et al.[36] and Yan et al.[37] in terms of generalization performance. Ablation experiments validate the importance of the MaxSel and MA Block in improving the detection accuracy and average precision of the detection algorithms. The research in this thesis provides a reference for detecting GAN-generated images.

**AUTHOR CONTRIBUTIONS**

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Haohuai Liu. The experiment was conducted by Ronghao Dai. The first draft of the manuscript was written by Ronghao Dai. The thesis revision was completed by Lingxi Peng. The overall supervision of the work was managed by Lingxi Peng. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**ACKNOWLEDGMENTS**

None reported.

**DECLARATIONS**

**Financial disclosure** The research did not receive support from any organization for the submitted work. No funding was received to assist with the preparation of this manuscript. No funding was received for conducting this study. No funds, grants, or other support was received.

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

**Data availability statements** The relevant code for this research can be obtained through reasonable requests and contacting the author’s email.

**REFERENCES**

[1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[2] Shi C , Liu W , Meng J ,et al.Self-prior guided generative adversarial network for image inpainting[J].The Visual Computer:1-13[2024-11-16].DOI:10.1007/s00371-024-03578-x.

[3] Yang, W., An, H., Hu, W. et al. Text-guided floral image generation based on lightweight deep attention feature fusion GAN. Vis Comput (2024). https://doi.org/10.1007/s00371-024-03617-7.

[4] Cheema, Y., Cheema, M.N., Nazir, A. et al. A novel approach for improving open scene text translation with modified GAN. Vis Comput (2024). https://doi.org/10.1007/s00371-024-03371-w

[5] McCloskey S, Albright M. Detecting GAN-generated imagery using saturation cues[C]//2019 IEEE international conference on image processing (ICIP). ieee, 2019: 4584-4588.

[6] Durall R, Keuper M, Keuper J. Watch your up-convolution: cnn based generative deep neural networks are failing to reproduce spectral distributions[C]/ /Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 7890-7899.

[7] Guo H, Hu S, Wang X, et al. Eyes tell all: Irregular pupil shapes reveal GAN-generated faces[C]//2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022: 2904-2908.

[8] Yu J, Jiang Y, Wang Z, et al. Unitbox: an advanced object detection network[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016 : 516-520.

[9] Liu Y, Wan Z, Yin X, et al. Detection of GAN generated image using color gradient representation[J]. Journal of Visual Communication and Image Representation, 2023, 95: 103876.

[10] Karras T , Aila T , Laine S ,et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation[J]. 2017. doi:10.48550/arXiv.1710.10196.

[11] Zhang X, Karaman S, Chang S F. Detecting and simulating artifacts in GAN fake images[C]//2019 IEEE international workshop on information forensics and security (WIFS). ieee, 2019: 1-6.

[12] Liu H, Li X, Zhou W, et al. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 772-781.

[13] Rossler A, Cozzolino D, Verdoliva L, et al. Faceforensics++: Learning to detect manipulated facial images[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1-11.

[14] Li Y, Yang X, Sun P, et al. Celeb-df: A large-scale challenging dataset for deepfake forensics[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 3207-3216.

[15] Jeong Y, Kim D, Min S, et al. Bihpf: Bilateral high-pass filters for robust deepfake detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022: 48-57.

[16] Tian C, Luo Z, Shi G, et al. Frequency-aware attentional feature fusion for deepfake detection[C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[17] Wang B, Wu X, Tang Y, et al. Frequency domain filtered residual network for deepfake detection[J]. Mathematics, 2023, 11(4): 816.

[18] Chollet F .Xception: Deep Learning with Depthwise Separable Convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition ( CVPR).IEEE, 2017.DOI:10.1109/CVPR.2017.195.

[19] Miao C, Tan Z, Chu Q, et al. F 2 trans: high-frequency fine-grained transformer for face forgery detection[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 1039-1051.

[20] Choi Y , Choi M , Kim M ,et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition.0[2024-10-02].DOI:10.48550/arXiv.1711.09020.

[21] Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of stylegan[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8110-8119.

[22] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

[23] Wang S Y, Wang O, Zhang R, et al. CNN-generated images are surprisingly easy to spot... for now[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8695-8704.

[24] He Y, Yu N, Keuper M, et al. Beyond the spectrum: Detecting deepfakes via re-synthesis [DB]. arxiv preprint arxiv:2105.14376, 2021.

[25] Brock A , Donahue J , Simonyan K .Large Scale GAN Training for High Fidelity Natural Image Synthesis[J]. 2018. doi:10.48550/arXiv.1809.11096.

[26] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy,Aditya Khosla, Michael Bernstein, et al . Imagenet large scale visual recognition challenge. ijcv, 2015. 3

[27] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. lsun: Construction of a large-scale image dataset using deep learning with humans in the loop [DB]. arXiv preprint arXiv:1506.03365, 2015. 3, 4

[28] Rossler A, Cozzolino D, Verdoliva L, et al. Faceforensics++: Learning to detect manipulated facial images[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1-11.

[29] Jeong Y, Kim D, Ro Y, et al. FrePGAN: robust deepfake detection using frequency-level perturbations[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(1): 1060-1068.

[30] Tanaka M, Shiota S, Kiya H. A universal detector of CNN-generated images using properties of checkerboard artifacts in the frequency domain[C]//2021 IEEE 10th Global Conference on Consumer Electronics (GCCE). IEEE, 2021: 103-106.

[31] Dong C, Kumar A, Liu E. Think twice before detecting gan-generated fake images from their spectral domain imprints[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 7865-7874.

[32] Arruda P H R. Synthetic image detection using a modern CNN and noise patterns[D]. , 2023.

[33] 100000 faces generated.<https://generated.photos/.>

[34] Frank J, Eisenhofer T, Schönherr L, et al. Leveraging frequency analysis for deep fake image recognition[C]//International conference on machine learning. pmlr, 2020: 3247-3258.

[35] Deng X, Zhao B, Guan Z, et al. New finding and unified framework for fake image detection[J]. IEEE Signal Processing Letters, 2023, 30: 90-94.

[36] Guo Z, Yang G, Zhang D, et al. Rethinking gradient operator for exposing AI-enabled face forgeries[J]. Expert Systems with Applications, 2023, 215: 119361.

[37] Yan S, Li O, Cai J, et al. A sanity check for ai-generated image detection[J]. arxiv preprint arxiv:2406.19435, 2024.

[38] Tan C, Zhao Y, Wei S, et al. Rethinking the up-sampling operations in cnn-based generative network for generalizable deepfake detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 28130-28139.

**AUTHOR IMFORMATION**

****

**Ronghao Dai** is currently a master's student at the School of Computer Science and Network Engineering, Guangzhou University. He has long been deeply involved in the field of artificial intelligence applications, with main research interests in computer vision, image processing, and particularly in the detection of generated images.

**Lingxi Peng** received the Ph.D. degree in Computer Application Technology from Sichuan University, China, in 2008. He is currently a Professor with the School of Mechanical and Electrical Engineering, Guangzhou University, and a senior member of the China.

**Haohuai Liu** is currently a associate professor with the School of Chemistry and Chemical Engineering, Guangzhou University. Her main interests include artificial intelligence and image processing.