MaxPix: detecting GAN-generated images by emphasizing local maxima

Ronghao Dai[1†★],   Lingxi Peng[1†*★], Haohuai Liu[2†★]

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

[2]Analysis 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;

Contributing authors: 670482535@qq.com; huaihuai99@163.com

[†]These authors contributed equally to this work.

Abstract: The realistic images generated by GANs(Generative Adversarial Networks) enrich people's lives, but they also pose serious threats to personal privacy and society, and it has become essential to study algorithms that can accurately detect GAN-generated images. Existing studies use artifacts to detect GAN-generated images, but the artifacts present in different GAN-generated images vary widely, and thus the cross-model generalization performance of such algorithms is weak. In this thesis, we propose the MaxPix, a new algorithm based on the combination of statistical features and deep learning techniques, for generating image detection. Firstly, MaxPix obtains the filter map of the image by designing the MaxSel filtering algorithm and then designs MA Block embedded in ResNet (Residual Network) to obtain MResNet. MaxPix finally utilizes MResNet to extract features from the filter map to detect GAN-generated images. Experimental results on publicly available datasets such as Wang and Faces-HQ show that the detection accuracy of MaxPix reaches 85.9% and 99.6% on average, which improves 7.6% and 10.2% relative to state-of-the-art algorithms such as the NAFID and the GocNet. Thus MaxPix has strong cross-model generalization performance.

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 the advantages of diverse contents and convenient storage, and have been widely used in the fields of news, information, medical diagnosis and identification, etc. GANs (Generative Adversarial Networks) is a generative model based on deep learning technology, which was proposed by Ian Goodfellow et al.[1]. GAN consists of a generator and a discriminator, in which the generator can generate samples similar to real data. Until now, more than a hundred different GANs which can generate images have been produced. These GANs can be utilized in various aspects of life to enhance people's quality of life. For instance, Lu et al.[2] used GANs for facial expression recognition, Bommakanti et al.[3] proposed GAN-C to improve event prediction accuracy, and Aramkul et al.[4] introduced an IoT framework based on GAN technology to optimize manufacturing processes and enhance product quality. However, some people maliciously use GANs to forge images and abuse them in politics and pornography, posing a serious threat to personal privacy and society. Since digital images are widely used in various fields, their authenticity is very important. To prevent GAN-generated images from being abused and bringing harm to society, effective detection algorithms are needed to detect whether images are generated by GAN or not to help people correctly distinguish real images from generated images. Currently, researchers have proposed a large number of detection algorithms to detect GAN-generated images. These algorithms are mainly categorized into detection algorithms based on traditional digital image forensic methods and detection algorithms based on deep learning techniques.

In the detection algorithms based on traditional digital image forensics, researchers mainly design detection algorithms to detect generated images based on the properties of digital images such as illumination inconsistency and statistical properties in the spatial domain and frequency domain. McCloskey et al.[5] analyzed the process of color formation in images and argued that the normalization

process present in GAN restricts the range of pixels in the generated image, making the exposure of the generated image different from that of a real image. They proposed to use the measured frequency of overexposure and underexposure of the image as a feature to detect the generated image. However, the algorithm only achieves an AUC (Area Under Curve) value of 0.7. Durall et al.[6] found that the high-frequency components of the generated images are distorted and proposed to use the azimuthal integral of the image as a feature to detect the generated images through the support vector machine, which achieves 100% accuracy. However, the algorithm lacks cross-model generalization performance. Guo et al.[7] believed that the eye pupils in real face images are elliptical, while the eye pupils in generated face images are irregular. They proposed an algorithm to determine whether an image belongs to a generated image or not by calculating the IoU[8] values of the pupil region and the elliptical mask, which judge whether the image belongs to a generated image by the IoU value. This algorithm has strict requirements on the quality and angle of the image, and if there are defects in human physiology, it will make the algorithm misjudged. Liu et al.[9] used the Sobel operator to get the gradient of the image in HSV (Hue, Saturation, Value) space and count histograms of the gradient distribution as a feature to detect the generated images. The algorithm achieved 99.4% accuracy when detecting the images generated by PGGAN[10], but cross-model generalization performance was not investigated.

Detection algorithms based on traditional image forensics have theoretical and experimental foundations. However, such algorithms are highly susceptible to overfitting statistical features that exist only in the trainset, while different GAN-generated images have different statistical features and thus tend to have lower accuracy in detecting unknown GAN-generated images. In addition, These algorithms require the images to conform to a specific angle and quality, which also limits the application of the algorithms.

Detection algorithms based on deep learning techniques utilize neural networks to construct algorithmic models and learn general features from massive data to detect the generated images. Since neural networks have strong representational ability, these algorithms generalize well and attract many scholars to study. The up-sampling process is almost common to GANs. Zhang et al.[11] designed AutoGAN containing an up-sampling process to generate a large number of images that simulate a variety of generated images and used such images to train the algorithm. However, the detection accuracy of the algorithm will be severely degraded if the up-sampling method used by the GAN is significantly different from that used by AutoGAN. Liu et al.[12] found that the phase spectrum of the image retains rich frequency components and proposed to combine the image spatial domain features and the phase features to detect the generated images. The algorithm detects the two Deepfake datasets[13,14] obtaining an accuracy rate of 91.5% and 76.88%. Jeong et al.[15] proposed an algorithm that uses a high-pass filter to remove irrelevant features in the spatial domain and frequency domain for highlighting the important features to detect the generated image, which obtains 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, and then aggregated the features with the original image. They utilized the aggregated features to detect the generated image and obtains an accuracy of 97.74%. Wang et al.[17] used wavelet transform to transform the image in the spatial domain to the frequency domain, then extracted the high-frequency components in the image and fused the features with the original image. Algorithms detected the generated image by Xception[18] and achieved more than 98% accuracy, but the accuracy of detecting the low-quality image is lower. Miao et al.[19] designed the Center Differential Attention Transformer to make the algorithm learn global high-frequency information and local fine-grained features and designed a high-frequency wavelet sampler to make the algorithm extract multi-channel high-frequency features. The proposed algorithm aggregated the two features to detect the generated image, but the accuracy of detecting the compressed processed image is low.

Algorithms based on deep learning techniques generally need to utilize the artifacts brought by the imperfect design of the GAN to detect the generated images. However, with the improvement of the GAN structure, the obvious artifacts in the generated images have been effectively hidden. In addition, the artifacts generated by different GANs vary, which limits the generalization performance of artifact-dependent detection algorithms, resulting in low accuracy when detecting unknown GAN-generated images and a lack of generality of the algorithms.

Given this, this thesis proposes to investigate detection algorithms that do not need to utilize artifacts to detect the generated images. In this thesis, the pixel value distribution of the GAN-generated images that generated by GAN such as StarGAN[20]and StyleGAN2[21] and real images in the datasets such as FFHQ[22] and CelebA are counted. It is observed that the generated images cannot reproduce the pixel distribution of the real images, and there are more points with larger pixel values in the real images than in the generated images. Therefore, this thesis proposes the MaxPix detection algorithm based on statistical features. Firstly, this thesis proposes the MaxSel algorithm for performing filtering on the images and then designs the MA Block embedded in ResNet to form MResNet, which is used to extract features from the filtered images to detect the generated images. Numerous experiments show the effectiveness of MaxPix for detecting generated

images. The contribution of this thesis is as follows:

Based on the characteristic that GAN-generated images cannot reproduce the pixel value distribution condition of real images, the MaxPix detection algorithm is proposed to detect GAN-generated images and the MaxSel is proposed for filtering images.

MaxPix detects the Wang[23] dataset and the Faces-HQ[6] dataset with an average accuracy of 85.9% and 99.6%, which is an improvement of 7.6% and 10.2% compared to current state-of-the-art detection algorithms. Thus the MaxPix has 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. He et al.[24] found that the generated images have stronger nonlocal similarity than real images, which inspired this thesis to explore whether there is any difference in the pixel distribution between the generated and real images.

For this purpose, the frequency of image pixel values in each pixel value range is counted and displayed using histograms in this thesis. In the experiment, the range of pixel values was divided into 60 groups. The experiments counted a total of 34k images including images generated by BigGAN[25], StarGAN and StyleGAN2, and real images sampled from ImageNet[26], CelebA and FFHQ datasets, which are from Wang dataset[23] and Faces-HQ[6]. As shown in Fig.1, although the above-mentioned GANs are trained with a large number of real images, it is still difficult to mimic the distribution of pixel values of real images. There are more points in the larger pixel value range in the real image than in the generated image. Therefore, this thesis proposes the MaxPix detection algorithm, which detects the generated image by emphasizing the local maxima of the image and using the maxima features.
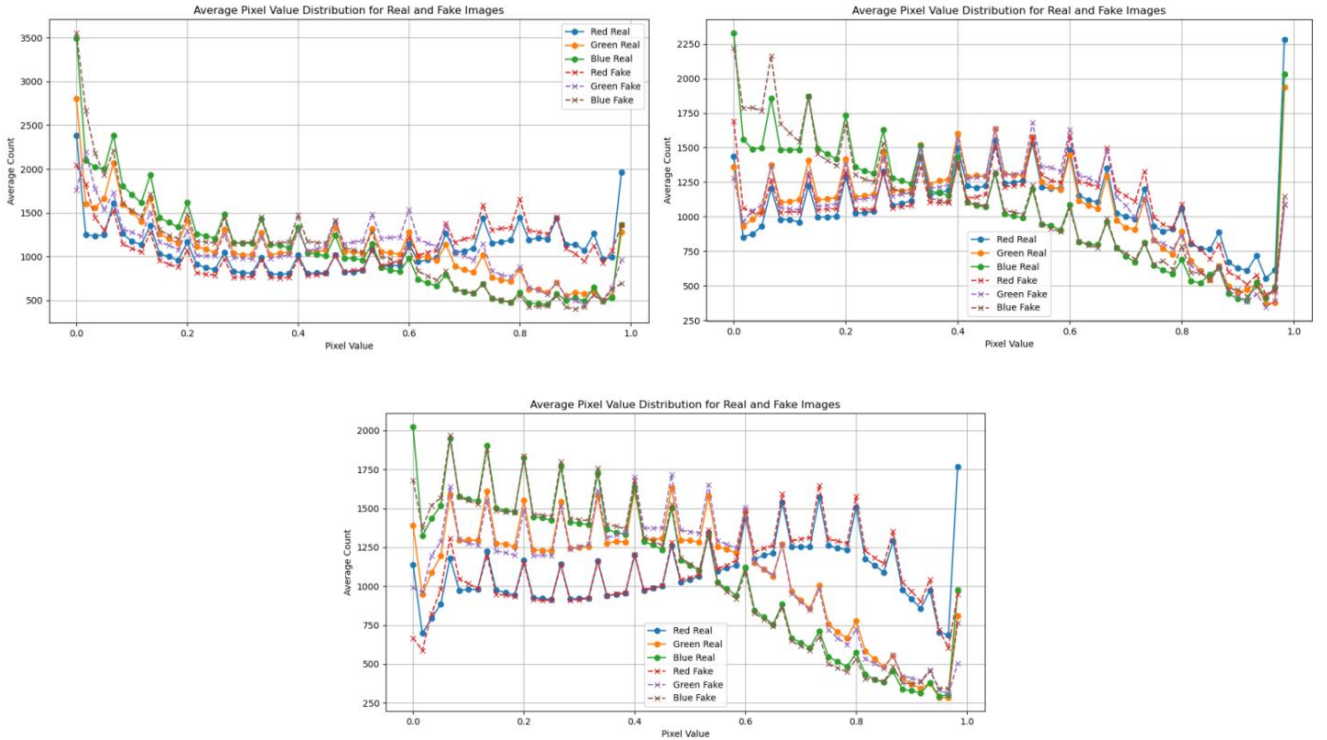

Fig1 Statistical distribution of pixel values of BigGAN, StarGAN and StyleGAN2 generated images and real images

2.1 Algorithmic framework

As shown in Fig.2, the MaxPix structure consists of a filtering module (or 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.
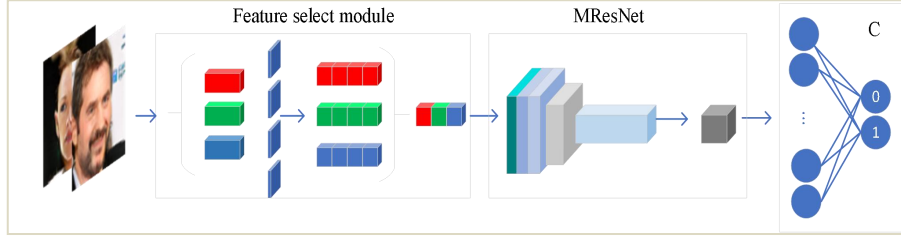
Fig2 MaxPix framework where the Feature select module does not update parameters

2.2 MaxSel filter

MaxPix uses the convolution kernel as in equation(1) as a filter kernel to perform convolution operation with the image to obtain the filtered image. Firstly, MaxPix splits the image channel-by-channel and performs the convolution operation using four convolution kernels to obtain the convolution values $X_{(c,i,j)}(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ in four directions for each point of the three channels. Then, MaxSel compares the convolutional values of the 4 directions at the corresponding location within the group and takes the largest convolutional value among them as the filter value. For $X_{(c,i,j)}(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, the maximum value is selected from α_1, α_2, α_3 and α_4.

$$
k_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}, k_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}
$$
$$
k_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, k_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}
$$

(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 splices the filter maps of the three channels to form the filter map $F_{in} \in \mathbb{R}^{3 \times H \times W}$.

$$
X_{(c,i,j)} = Max(\alpha_1, \alpha_2, \alpha_3, \alpha_4)
$$

(2)

As shown in Fig.3, the first column is the real image from the Wang dataset[23], the second column is the filter map obtained by using the Prewitt operator as the convolution kernel; the third column is the filter map obtained by taking the Laplacian operator as the convolution kernel and the fourth column is the filter map obtained by using MaxSel. The filter map obtained by MaxSel is delicate and complete in detail, which is favorable for the algorithm to learn more complete features from it.
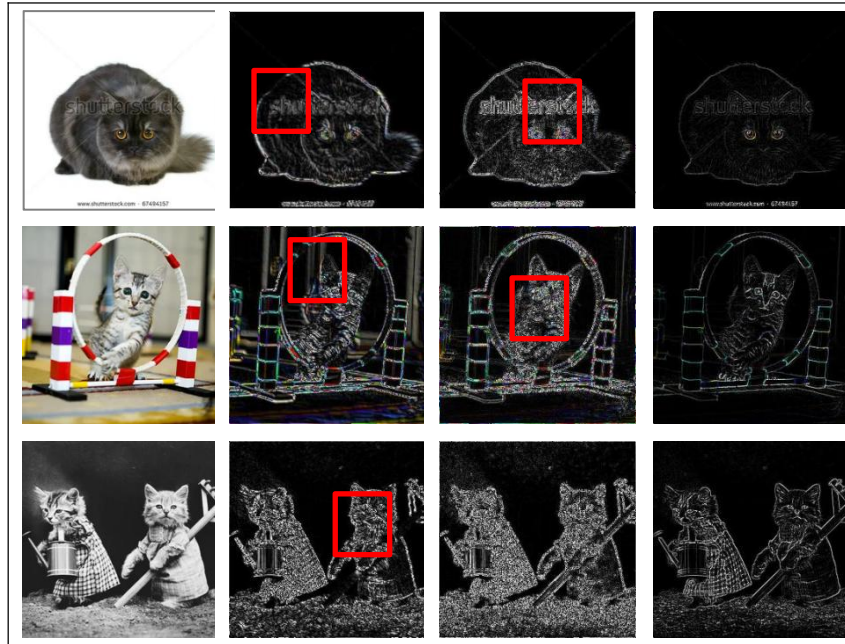

Fig3 Filtering effect image. Each column from left to right corresponds to the real image, filtered image of Prewitt, Laplacian and MaxSel.

2.3 MResNet feature extraction network

As shown in Fig.4, MResNet is improved from ResNet and has five more MA blocks, which consist of a maximum pooling filter layer, a mean filter layer and a residual layer, than the ResNet. MResNet changes the mean pooling of the final output to maximum pooling, which is used for selecting the maximum features for detecting the generated image.
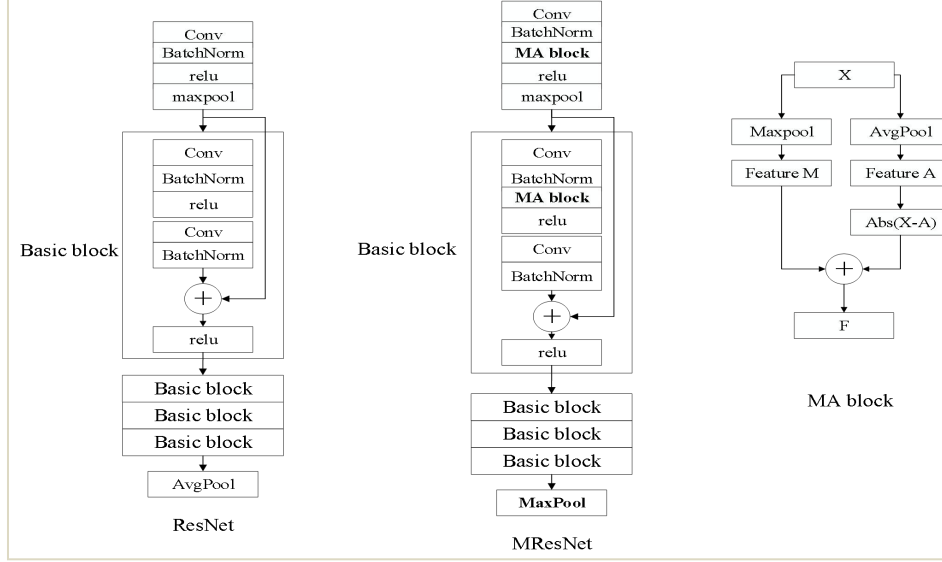

Fig4 MA block embedded in Basic block of the ResNet to form MResNet

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

$$F_{out} = MP(F_{in}) + \lambda \times Abs(F_{in} - AP(F_{in})) \tag{3}$$

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 F_d is the input features.

$$Loss = -\frac{1}{N} \sum_{i=1}^{N} y \log(C(F_d)) + (1 - y) \log(1 - C(F_d)) \tag{4}$$

3 EXPERIMENTAL

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

3.1 Datasets

In order to avoid misunderstanding in expression, this thesis uses the lowercase English name of the GAN to denote the dataset composed of the corresponding generated images and the real images, e.g., images generated by StyleGAN and real images used for training the GAN are called the stylegan dataset.

The Wang dataset: Wang et al.[23] published a publicly available but unnamed dataset, referred to as the Wang dataset in this thesis, which is divided into a trainset, an evaluation set, and a testset and contains real images and generated images. The real images were sampled from the LSUN[27], ImageNet dataset and other datasets that are commonly used to train GANs. The generated images including 20 scenarios images were generated by GANs such as PGGAN, StyleGAN2, and included fake face sampled from the fake face dataset

FaceForensics++ (deepfake)[28]. This dataset has been widely used by related researchers[15,23,29-32] to train and evaluate detection algorithms since its release.

The Faces-HQ: Dural et al.l[6] released the Faces-HQ dataset. Each image in the Faces-HQ dataset has a resolution of 1024×1024, which is much better than that of the Wang dataset. Faces-HQ contains 20k real face images which sample from CelebA-HQ[10] and FFHQ, and contains 20k generated images which sample from the 100K Faces project[33] and *www.thispersondoesnotexist.com*. The generated images are generated by StyleGAN and StyleGAN2. CelebA-HQ and FFHQ are often used to train GANs, which are recognized datasets for training and testing detection algorithms.

In this thesis, the person subset of the Wang trainset is used to train PixMSE, and the biggan, gaugan, stargan subsets of the the Wang testset and the Faces-HQ total of more than 102k images are used as testsets.

3.2 Experimental environment

In this thesis, the algorithm code is written with Python 3.7 and PyTorch 1.9.0, the GPU used is RTX 3090 and the system used is Ubuntu. MaxPix performs resize as well as random cuts on the trainset and resize as well as center cuts on the testset, which changes the input image into $X \in R^{3 \times 299 \times 299}$.

The training algorithm is set up with the epoch of 36 and batch-size of 4. The optimizer is Adam. The learning rate is 0.00005. The learning decay rate is 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] and Guo et al.[36]. These algorithms not only achieved better performance in their respective thesis and can achieve more than 90% accuracy in detecting the same type of GAN-generated images but also maintained a strong cross-model generalization performance.

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

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**[1]	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
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 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 11.6% 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%, which is lower than the best of the compared algorithms at 97.6%. However, MaxPix detects the remaining seven datasets all get the highest average precision, equaling or exceeding the best of the compared algorithms. It can be seen

---

1 The bolded values indicate the highest value in this column.

that the detection performance of MaxPix is better than the current mainstream detection algorithms in terms of accuracy and average precision.

Tab2 Comparison experiment 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
MaxPix	**99.9**	**100**	**99.3**	**99.9**	**99.6**	**100**

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 involved in the table2. Although the training and testing are from two different datasets with a huge difference in image resolution and the algorithms are not retrained in this thesis, MaxPix still performs well, obtaining 99.9% and 99.3% detection accuracy and 100% and 99.9% average precision, respectively, which are better than the comparison algorithms. This indicates that MaxPix detection accuracy and average precision are less affected by image size.

In addition, in Fig.5, the images in the Wang dataset have obvious artifacts, while the images in Faces-HQ have no obvious artifacts. This also indicates that the detection accuracy of MaxPix is less affected by artifacts. Comparison experiments show that the accuracy and average precision of MaxPix detection algorithms can match or exceed current state-of-the-art detection algorithms, with strong cross-model generalization performance.
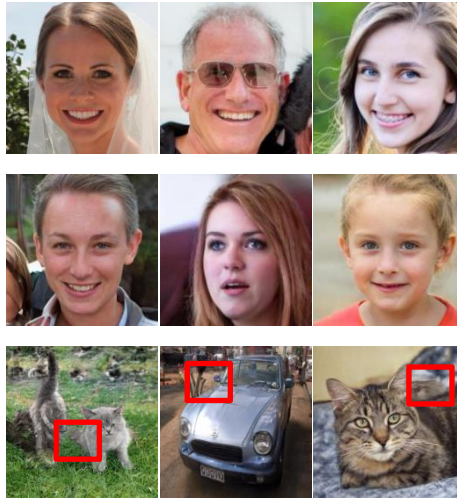

Fig5.The first row is from the real image of Faces-HQ, the second row is from the generated image of Faces-HQ, and the third row is from the generated image of Wang dataset.

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.

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, thus 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.

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

3.4.2 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.

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**

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.

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 the MaxPix for detecting GAN-generated images, an algorithm that produces features for detecting generated images by emphasizing the maximum value in the local range of the image. The main contribution of this thesis is to propose the MaxSel filtering algorithm and the MaxPix detection algorithm. Comparison experiments on Wang and Faces-HQ datasets show that MaxPix outperforms the state-of-the-art algorithms such as Deng et al.[35] and Guo et al.[36] in terms of generalization performance, and ablation experiments validate the importance of 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 the detection of GAN-generated images.

AUTHOR CONTRIBUTIONS

All authors of this article contribute equally.

ACKNOWLEDGMENTS

DECLARATIONS

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] Lu L. Multi-angle face expression recognition based on generative adversarial networks[J].Computational Intelligence, 2021(11).DOI:10.1111/coin.12437.

[3] Bommakanti S , Halani V , Shah D ,et al.GAN‑C: A generative adversarial network with a classifier for effective event prediction[J].Computational Intelligence, 2022, 38:1922 - 1955.DOI:10.1111/coin.12550.

[4] Aramkul S , Sugunnasil P .Intelligent IoT framework with GAN-synthesized images for enhanced defect detection in manufacturing[J].Computational intelligence, 2024(2):40.DOI:10.1111/coin.12619.

[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.