MaxPix：通过强调局部最值以检测GAN生成图像

戴荣浩1, 2a，彭凌西1, 2b†

(1. 广州大学 计算机科学与网络工程学院,; 2. 机械与电气工程学院; 广东 广州 510000)

摘 要： GAN（Generative Adversarial Networks）生成的逼真图像丰富了人们生活，但也给个人隐私和社会带来严重威胁，研究能准确检测GAN生成图像的算法变得极为重要。现有研究利用伪影来检测GAN生成的图像，当算法检测的生成图像所具有的伪影与训练接触的伪影差异较大时，检测准确率较低。该文提出一种新的基于统计特征和深度学习技术结合的生成图像检测算法MaxPix。MaxPix首先通过设计MaxSel滤波算法来获得原始图像的滤波图，并设计MA Block内嵌于ResNet(Residual Network)来得到MResNet，利永MResNet从滤波图中提取特征以检测生成图像。在Wang和Faces-HQ等公开数据集的实验结果表明，MaxPix的检测准确率平均为85.9%和99.6%，相对现有算法提升7.6%和10.2%，具有较强的跨模型泛化性能。

关键词：GAN；生成图像；伪影；跨模型泛化

0 引言

数字图像由于具有内容多样和存储便利等优点,已经成为传递网络信息的主要载体之一,并被广泛应用于新闻资讯、医疗诊断和身份识别等领域。GAN（Generative Adversarial Networks）是一种基于深度学习技术的生成模型，由Ian Goodfellow等人于2014年提出。GAN由生成器和鉴别器构成，其中生成器能够生成与真实数据相似的样本。截止目前，已产生超过百种不同的、能够生成图像的GAN，这些生成的图像丰富了人们生活。然而，一些人却恶意地利用GAN来伪造图像并滥用于政治和色情领域，给个人隐私和社会带来了严重的威胁。由于数字图像被广泛应用于个领域，其真实性十分重要。为防止GAN生成的虚假图像被滥用而给社会带来危害，需要研究有效的检测算法来检测图像是否由GAN生成，以帮助人们正确区分真实图像和生成图像。当前，研究人员提出了大量检测算法来检测GAN生成的图像。这些算法主要分为基于传统数字图像取证方式的检测算法和基于深度学习技术的检测算法。

基于传统数字图像取证方式的检测算法中，研究者主要根据数字图像中诸如光照不一致性，空域和频域统计特性等性质来设计检测算法以鉴别图像真实性。McCloskey[17]等人分析了图像颜色形成的过程，认为GAN存在的标准化过程限制了生成图像的像素范围，使生成图像的曝光度与真实图像不同，提出将测量的图像过曝光和曝光不足的频率作为特征来检测生成图像。然而，算法仅取得0.7的AUC（Area Under Curve）值。Guo[21]认为真实人脸图像中的眼睛瞳孔是椭圆状的，而生成人脸图像中的眼睛瞳孔是不规则的，提出通过计算瞳孔区域与椭圆掩模的IoU值，以IoU值大小判断图像是否属于生成图像的算法。这种算法对图像的质量和角度有严格的要求，如果人体生理存在缺陷，将使算法产生误判。Liu[22]等人使用Sobel算子得到图像在HSV空间的梯度，然后统计了梯度分布的直方图，利用直方图作为特征检测生成图像，检测PGGAN生成的图像取得99.4%的准确率，但未研究跨模型泛化性能。Durall[23]等人发现生成图像的高频成分存在失真现象，提出以图像方位积分(azimuthal integral)为特征，然后通过支持向量机分类的方法来检测生成图像，取得100%的准确率，但算法缺乏跨模型泛化性能。

基于传统图像取证方式的检测算法的具备理论和实验基础。然而，这类算法表征能力较弱，极容易过拟合仅存在于训练集的统计特性，因而往往检测未知GAN生成的图像时的准确率较低。另一方面，算法要求图像符合特定角度和质量，面向的场景严格，也限制了算法的应用。

基于深度学习技术的检测算法利用神经网络来构建算法模型，然后从海量的数据中学习一般性特征以检测生成图像。由于神经网络具有强大的表征能力，因而该类算法泛化性较强，吸引了众多学者研究。上采样过程几乎是GAN的共有过程，Zhang[34]等人设计含有上采样过程的AutoGAN生成大量的、模拟各种生成图像的图像，然后使用这种图像训练算法的网络参数。然而，如果GAN采用的上采样方式与AutoGAN采用的有较大差异，算法的检测准确率将严重下降。Liu[36]等人发现图像的相位谱保留丰富的频率分量，提出结合图像空域特征和相位特征检测生成图像，算法检测两个Deepfake数据集[37,38]得到91.5%和76.88%的准确率。Jeong[26]提出的算法使用高通滤波器去除空域和频域无关特征，从而突出重要特征以检测生成图像，取得超过72%的跨模型检测准确率和平均精度。Tian[40]等人将图像频率分量划分为低、中、高成分，然后与原始图像进行特征聚合，利用聚合特征检测生成图像，取得97.74%的准确率，但检测未知GAN生成的图像的准确率只有69.17%。Wang[41]等人使用小波变换将空域图像变换至频域，再提取图像中的高频分量，然后与原始图像进行特征融合，并通过Xception来检测生成图像，取得超过98%的准确率，但检测低质量图像的准确率较低。Miao[42]等人设计中心差分注意变换器使算法学习全局高频信息和局部细粒度特征；设计高频小波采样器使算法提取多通道的高频特征，然后将两种特征进行聚合。提出的算法以聚合特征来检测生成图像，但检测被压缩处理的图像的准确率较低。

基于深度学习技术的算法一般需要利用生成图像存在的、由于GAN设计不完美带来的伪影来检测生成图像。然而，随着GAN结构的改进，生成图像中明显的伪影已经得到有效隐藏。此外，由于不同GAN产生的伪影也有所不同，使得依赖伪影的检测算法的泛化性能受到限制，检测未知GAN生成的图像时的准确率低，算法缺乏通用性。

鉴于此，本文拟研究无需利用伪影来检测生成图像的检测算法。本文统计了StyleGAN、StyleGAN2，ProGAN生成的图像,以及CelebA-HQ，FFHQ, CelebA数据集中真实图像的像素值分布，观察到生成图像无法重现真实图像的像素分布情况，真实图像的较大像素值的点多于生成图像。因此本文提出基于统计性特征的MaxPix检测算法。首先，本文提出MaxSel算法用于对图像执行滤波，然后设计MA Block内嵌于ResNet以形成MResNet，用于从滤波图中提取特征以检测生成图像。大量实验验证了MaxPix检测生成图像的有效性。本文的贡献如下：

（1）实验指出GAN生成的图像无法重现真实图像的像素值分布状况。

（2）本文提出MaxSel用于图像滤波，提出MaxPix检测算法来检测GAN生成的图像。

（3）与对比工作相比，MaxPix检测Wang[7]数据集和Faces-HQ[23]数据集的准确率平均提升7.6%和10.2%，平均精度提升9.1%和1.1%，具有较强的跨模型泛化性能。

1 算法描述

本文首先进行可行性实验。可行性实验给MaxPix提供了一定的理论支撑。然后，本文详细MaxPix的算法框架以及实现细节。

1.1可行性实验

Durall[23]发现GAN生成的图像无法再现真实图像的频谱分布，He[66]发现生成图像具有比真实图像更强的非局部相似性，这启发本文探究生成图像和真实图像像素分布上是否存在一定的差异。为此，本文统计RGB图像每组像素值范围内的像素值的出现频数并采用直方图进行显示。

实验中，像素值范围被划分为60组。实验统计了BigGAN、StarGAN和StyleGAN2生成的图像,以及ImageNet[75]、CelebA和FFHQ数据集中真实图像在内总计34k张图像，图像来自Wang数据集[7]和Faces-HQ[23]。如图1示，虽然上述GAN使用大量真实图像进行训练，但仍然难以模仿真实图像的像素值分布状况。显然真实图像中处于较大像素值范围的点多于生成图像。因此，本文提出MaxPix检测算法，通过强调图像局部最值，并利用最值特征检测生成图像。

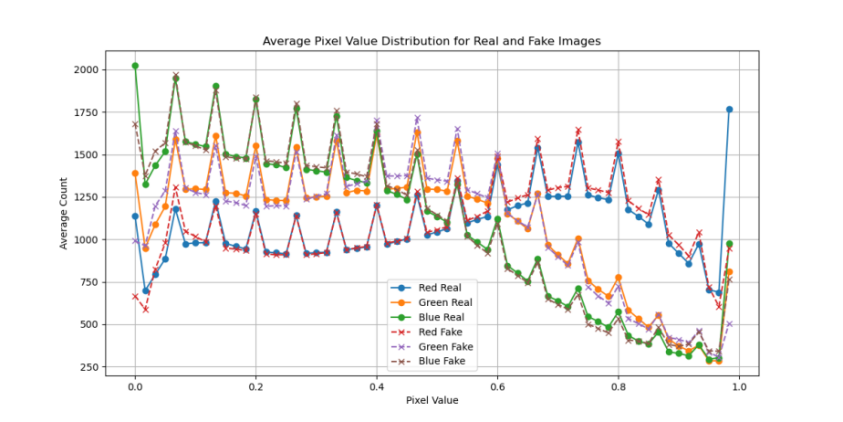
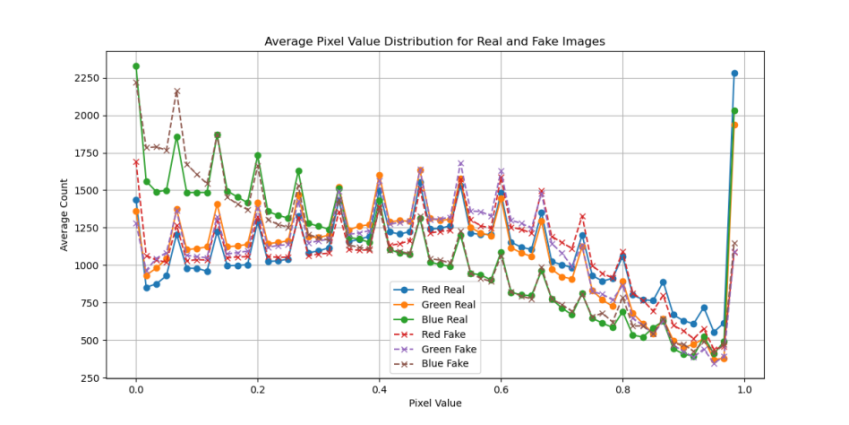
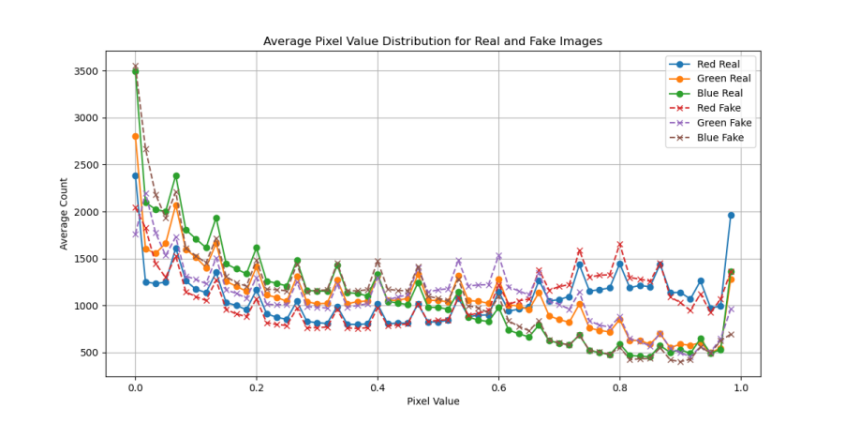


图1 BigGAN、StarGAN和StyleGAN2生成的图像及真实图像的像素值统计分布图

1.2 算法框架

MaxPix结构如图2示，由滤波模块（或特征选择模块）、特征提取网络MResNet和分类器C组成。滤波模块采用本文提出的MaxSel滤波算法对图像执行滤波，使MResNet易于学习可区分特征以检测GAN生成的图像。

图2 MaxMaxPix算法框架，在MaxPix中Feature select module不更新参数

1.2.1 MaxSel滤波

MaxPix采用如式（1）的卷积核作为滤波核与原始图像进行卷积运算来得到滤波图像。

 (1)

首先，MaxPix将RGB图像进行逐通道拆分，然后使用4个卷积核分别执行卷积运算，得到三个通道每个点4个方向的卷积值X（c,i,j）（α1,α2,α3,α4）。然后，MaxSel将相应位置4个方向的卷积值进行组内比较，以其中的最大卷积值作为滤波值。对于X（c,i,j）（α1,α2,α3,α4），从α1,α2,α3,和α4中选择最大值。如公式（2）示，其中，X（c,i,j）表示图像c通道（i,j）位置上的滤波值。各点上的滤波值构成了该通道的滤波图。最后，MaxSel将三个通道的滤波图进行拼接，组成滤波图*Fin*∈*R*3×*H*×*W*。

 (2)

如图3示，第一列是来自Wang[24]数据集的真实图像；第二列是使用Prewitt算子作为卷积核得到的滤波图；第三列是采取Laplacian算子作为卷积核得到的滤波图；第四列是使用MaxSel得到的滤波图。显然，MaxSel得到的滤波图细腻且细节完整，这有利于算法从中学习到较为完整的特征。

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

图3 滤波效果图。从左到右每列依次对应为真实图像、Prewitt滤波图像、Laplacian滤波图像以及MaxSel滤波图像

1.2.2 MResNet特征提取网络

如图4示，MResNet由ResNet改进得来，其中MA block由最大池化滤波层、均值滤波层和残差层构成。MResNet比ResNet多5个MA block，并且，MResNet将最后输出的平均池化变为最大池化，用于选择最大值特征以检测生成图像。



图4 在MResNet中，MA block内嵌于Basic block

MA block用于强调特征图中的局部最大值，如公式（3）示，其中λ是一个可更新的参数，Fin表示输入的特征。MP表示最大池化，AP表示均值滤波。Abs表示取绝对值。

 (3)

1.2.3 分类器与损失函数

分类器C由两层全连接层组成，MaxPix将MResNet输出的8192个特征展平，然后利用全连接层转化为预测值。如公式（4）示，其中C是分类器，y表示图像真实的标签，Fd是输入特征。

 (4)

2 实验

本文通过对比当前具有代表性的检测算法检测不同数据集的准确率和平均精度来展示MaxPix在跨模型泛化性能上的提升。通过消融实验验证MaxPix各模块的作用。

2.1 数据集

为了避免表达上产生误解，本文以生成模型的小写英文名字来表示对应生成图像及真实图像所构成的数据集，如将StyleGAN生成的图像及训练生成模型用的真实图像称为stylegan数据集。

Wang数据集：Wang等人[5]发布了一个公开的但未命名的数据集，本文称为Wang数据集，该数据集分为训练集、评估集和测试集，包含来自LSUN，COCO、ImageNet数据集等常用于训练GAN的真实图像、使用PGGAN、StyleGAN2，CycleGAN等GAN生成的20种场景的生成图像以及伪造的面部数据集FaceForensics++（deepfake）。该数据集发布后已被相关研究人员[22,24,25,26,27]广泛用于训练和评估检测算法。

Faces-HQ数据集：Faces-HQ数据集由Durall等人[23]发布，包含来自CelebA-HQ[25]和FFHQ的各10k张真实人脸图像，来自 Faces project[73]的、由StyleGAN生成的人脸图像和来自*www.thispersondoesnotexist.com*的、由StyleGAN2生成的各10k张人脸图像。Faces-HQ中每张图像的分辨率是1024×1024。CelebA-HQ和FFHQ常被用于训练GAN，是公认的用于训练和测试生成图像检测算法的数据集。

Wang训练集中的person子集为MaxPix的训练集，共18003张真实图像和18003张PGGAN生成的人物图像。Wang测试集中的biggan、gaugan、stargan、deepfake、cyclegan、progan、stylegan、stylegan2集全部超过62k张图像作为MaxPix的测试集，Faces-HQ数据集也作为测试集。

2.2 实验环境

本文采用python3.7和PyTorch1.9.0编写算法代码，使用的GPU为 RTX 3090，使用的系统是Ubuntu。MaxPix对训练集执行Resize以及随机切割等处理，对测试集执行Resize以及中心切割处理，将输入的图像变为*X*∈*R*3×299×299。

训练算法设置的epoch是36，batch-size是4，优化器是Adam，学习率是0.00005，学习衰减率0.96，采用的损失函数是CrossEntropyLoss,。

2.3 对比实验

本文选取近年来检测GAN生成的图像的任务中取得较高准确率等研究工作进行比较，包括Wang[8]、Frank[12]、Durall[5]、Jeong[15，26]、He[21]、Deng[18]和Guo[29]。这些算法在各自论文中均取得较好的表现，检测同类型GAN生成的图像能取得超过90%的准确率，同时保持具有较强的跨模型泛化性能。

除Jeong等人[15,26]的算法外，本文使用Wang数据集对其余算法进行了重新训练和测试。Jeong等人[15,26]的算法使用的是Wang数据集且无法获取代码实现细节，因此表中的实验数据引自文献[15]。

表1 对比实验 Wang数据集（%）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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[7] | 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[35] | 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[23] | 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[66] | 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[39] | 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[69] | 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[33] | 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[71] | 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 |

如表1示，MaxPix检测biggan、cyclegan、stargan和stylegan数据集取得较高准确率，高于对比算法中取得的最高值。其中，MaxPix检测biggan的准确率提升达6.1%，检测stylegan的准确率提升达11.6%。MaxPix与大部分对比的算法一样，检测gaugan数据集的准确率较低，仅63%。在平均精度表现方面，MaxPix检测gaugan的平均精度为75.5%，低于对比算法的最佳的97.6%，差距较大。然而，MaxPix检测其余7个数据集均得到较高的平均精度，持平或超过对比算法的最佳值。可见，以准确率和平均精度为指标，MaxPix的检测性能优于现有的检测算法。

表2 对比实验 Faces-HQ(%)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Faces-HQ | | | | Faces-HQ | |
|  | StyleGAN,CelebA-HQ | | StyleGAN2,FFHQ | | ave | |
|  | Acc | AP | Acc | AP | Acc | AP |
| Wang[7] | 49.7 | 45.1 | 51.9 | 74.2 | 50.8 | 59.7 |
| Frank[35] | 67.2 | 78.2 | 58.3 | 63.6 | 62.7 | 70.9 |
| Durall[23] | 57.2 | 93.6 | 62.9 | 91 | 60.0 | 92.3 |
| He[66] | 65.1 | 85.0 | 70.2 | 96.1 | 67.6 | 90.6 |
| Deng[33] | 79.9 | 99.2 | 77.8 | 93.3 | 78.9 | 96.3 |
| Guo[71] | 96.4 | 99.8 | 82.3 | 97.8 | 89.4 | 98.8 |
| MaxPix | **99.9** | **100** | **99.3** | 99.9 | **99.6** | **100** |

如表2示，各算法检测Faces-HQ数据集的平均精度和准确率差异明显。由于获取不到Jepong等人[39,69]算法的实现细节，因此这两项算法不参与比较。尽管训练和测试来自两个不同的数据集，且图像分辨率差距极大，而本文也没有重新训练算法，MaxPix依然表现出色，分别得到99.9%和99.3%的检测准确率，以及100%和99.9%的平均精度，优于对比算法。这表明MaxPix检测准确率和平均精度受图像大小的影响小。此外，在图5中，Wang数据集中的图像存在明显的伪影，而Faces-HQ中的图像则无明显伪影。这也说明，MaxPix的检测准确率受伪影影响小。对比实验表明MaxPix检测生成图像的准确率和平均精度能媲美或超过当前先进检测算法，具有较强的跨模型泛化性能。







图5.第一行来自Faces-HQ的真实图像，第二行来自Faces-HQ的生成图像，第三行来自Wang[7]数据集的生成图像。

2.4 消融实验

本文通过消融实验探究MaxSel滤波和MA Block的作用。模块消融实验以ResNet作为比较基准。“ResNet”以未滤波的图像作为ResNet的输入；“MResNet”以未滤波的图像作为MResNet的输入，探究MaxSel的作用；“MSel”通过MaxSel对图像进行滤波，并将其作为ResNet的输入，探究MA Block的作用。

2.4.1 模块消融实验

如表3示，ResNet仅检测stylegan2和progan的准确率超过80%，平均精度超过90%；MResNet虽然增加了MA Block，但检测生成图像的准确率和平均精度并没有提升，因此MA Block单独使用并不会提升算法性能；MSel由于采用MaxSel对图像进行滤波，使算法易于从滤波图中学习可区分性特征，因而检测准确率和平均精度得到全面提升，特别是检测deepfake，提升40.5%的准确率和47.9%的平均精度。MaxPix在Msel的基础上引入MA Block，检测progan、biggan、cyclegan、gaugan和stylegan2的准确率分别提升0.1%、2.8%、16.5%、8.3%和0.1%。MaxPix检测deepfake的平均精度有轻微的下降。可见，Maxsel搭配MA Block使用，有效提升检测算法检测生成图像的准确率和平均精度，且发挥最大作用的是Maxsel滤波。

表3 模块消融实验(%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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** |

2.4.2 网络结构消融实验

在本消融实验中，MResNet、ResNet以不同的滤波算法（Laplacian、Sobel、Prewitt和Scharr）得到的图像作为输入，进一步探究MaxSel滤波算法提出的必要性。

如表4和表5示，检测算法采用Maxsel对图像进行滤波，不论以MResNet或ResNet作为网络架构，均能在多个数据集上取得最高的准确率和平均精度，特别是检测stargan，始终保持100%的准确率和平均精度，但检测gaugan的准确率始终较低，分别是63%和54.7%，平均精度也仅得到75.5%和63.5%。然而，即使采用其他算子对图像进行滤波，检测算法检测gaugan的准确率和平均精度也低，准确率最高只有70.3%，平均精度也仅为80%。这表明通过对图像进行滤波处理，对提升检测gaugan的准确率和平均精度的帮助较少。

总体而言，两项消融实验表明Maxsel和MA Block对提升算法检测生成图像的准确率和平均精度具有较大帮助，特别是Maxsel滤波可以高效地提升检测算法的泛化性能。

表4 网络结构消融实验-MResNet（Acc，%）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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** |

表5 消融实验-ResNet（Acc）



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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** |

3 结束语

本文提出用于检测GAN生成图像的MaxPix，该算法通过强调图像局部范围内的最大值来产生用于检测生成图像的特征。MaxPix算法主要创新是提出了MaxSel滤波算法和MA Block。在Wang和Faces-HQ数据集的对比实验表明MaxPix在泛化性能上优于对比算法，消融实验验证了MaxSel和MA Block对提升检测算法的检测准确率和平均精度的重要性。本文研究为检测生成图像提供了一定参考。

4 参考文献

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

[2] 张亚.基于深度学习的Deepfake检测方法研究[D].云南大学, 2022.

[3] Yu N, Skripniuk V, Abdelnabi S, et al. Artificial fingerprinting for generative models: Rooting deepfake attribution in training data[C]//Proceedings of the IEEE/CVF International conference on computer vision. 2021: 14448-14457.

[4] Li, Haodong, et al. "Identification of deep network generated images using disparities in color components." Signal Processing 174 (2020): 107616.

[5] Wang, Sheng-Yu, et al. "CNN-generated images are surprisingly easy to spot. for now." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[6] Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis." arxiv preprint arxiv:1809.11096 (2018).

[7] Choi, Yunjey, et al. "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[8] Park, Taesung, et al. "Semantic image synthesis with spatially-adaptive normalization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[9] Chen B, Tan W, Wang Y, et al. Distinguishing Between Natural and GAN‐Generated Face Images by Combining Global and Local Features[J]. Chinese Journal of Electronics, 2022, 31(1): 59-67.

[10] Liu, Yun, et al. "Detection of GAN generated image using color gradient representation." Journal of Visual Communication and Image Representation 95 (2023): 103876.

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

[12] 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.

[13] 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.

[14] Liu, Honggu, et al. "Spatial-phase shallow learning: rethinking face forgery detection in frequency domain." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

[15] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE International Conference on Computer Vision, pages 1–11, 2019.

[16] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3207–3216,2020.

[17] Tian, Cheng, et al. "Frequency-aware attentional feature fusion for deepfake detection." ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023.

[18] Wang, Bo, et al. "Frequency domain filtered residual network for deepfake detection." Mathematics 11.4 (2023): 816.

[19] Miao, Changtao, et al. "F 2 trans: High-frequency fine-grained transformer for face forgery detection." IEEE Transactions on Information Forensics and Security 18 (2023): 1039-1051.

[20] Yang, Gaoming, et al. "FDS\_2D: rethinking magnitude-phase features for DeepFake detection." Multimedia Systems 29.4 (2023): 2399-2413.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] Jeong, Yonghyun, et al. "Bihpf: Bilateral high-pass filters for robust deepfake detection." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.

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

[28] Deng, X., Zhao, B., Guan, Z., & Xu, M. (2023). New finding and unified framework for fake image detection. IEEE Signal Processing Letters, 30, 90-94.

[29] Guo, Zhiqing, et al. "Rethinking gradient operator for exposing AI-enabled face forgeries." Expert Systems with Applications 215 (2023): 119361.

[30] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9.11 (2008).