1. **(CViT) Deepfake Video Detection Using Convolutional Vision Transformer --- arxiv 2021.02**

1）主要内容

卷积视觉Transformer由两个组成部分：CNN和视觉Transformer(ViT)。CNN(没有全连接层的VGG架构，不进行分类)提取可学习的特征，而ViT将学习到的特征作为输入，并用注意机制进行分类。此工作基于Deepfake检测方法的两个弱点：数据预处理和（检测）泛化性。预处理包含：人脸提取224\*224和数据增强。本文方法具有泛化能力的原因：①CNN+Transformer学习局部和全局特征。②在训练和测试时对数据都进行预处理。③更大、多样性的数据集训练模型以检测不同设置和生成方法下的伪造。

1. 实验细节

预处理：人脸检测BlazeFace，人脸提取MTCNN和face recognition。为了提高准确率，三个库一起使用。保存为224\*224的jpeg格式。Albumentations库用于数据增强，90%的图片进行增强，数据增加近一倍。每个batch32张图片进行归一化，均值= (0:485;0:456;0:406)和标准差= (0:229;0:224;0:225)。

数据集：162174张图片，其中训练、验证、测试的划分比例为70：15：15，真假图片数量相等。90%的压缩率。

UADFV、 FaceForensics++（还包含了faceshifter）；DFDC仅用于测试（400个未见视频，每个视频取30张图片）。

评估指标：ROC、AUC

优化器：Adam

阈值：<0.5为真实，>0.5为假。

3）总结思考

是否可进一步结合频率特征或者其他差异作为CNN的输入特征/其他方法提特征，作为输入进入Transformer。

是否可以利用本文预处理方法对需要测试的数据，先进行预处理再检测，以提高检测能力---论文Investigating the Impact of Pre-processing and Prediction Aggregation on the DeepFake Detection Task.

实验测试了此模型的准确性如何受到预处理的影响，没有删除无人脸的图像，准确性下降到69.5%。通过对三个人脸提取库的测试，发现face recognition由于其他两个，少的假阳性。

我们提出更加多样性的数据集，在此基础上进行训练，可以提高模型的检测性能，但是直接测试就不易检测。

目前检测方法缺乏泛化性：检测模型主要关注Deepfake制作过程tool，通过研究假定行为（眨眼等）来解决问题。

深度模型比浅层性能好，这中间存在一个预处理环节可能捕获了更多的伪影，提升了性能。

视频帧选择的数量可能会影响检测性能。

1. **(F3-Net) Thinking in frequency Face forgery detection by mining frequency-aware clues**
2. 主要内容

该方法利用频率感知检测伪造。基于观察：频率可以很好地描述细微的伪影或压缩错误。该方法利用两个互补的频域感知线索，提出双流协同学习框架：①FAD通过频率感知图像分解来学习微妙的伪造模式。②LFS用于提取局部频率统计信息。③MixBlock交叉注意模块融合以上2个分支，用于联合特征交互。

1. 实验细节

数据集：FF++（1000真实视频，720用于训练，140用于验证，140用于测试。4种伪造方法共4000假视频。真实视频增强4倍以解决真假数据不平衡问题。遵循FF++设置每个视频取270帧）

评估指标：AUC、ROC、PR

对比方法：MesoNet、Face X-ray、Xception。

Xception-ELA [27]、Xception-PAFilters [10]、Steg.Features[24]、LD-CNN [14]、Constrained Conv [6]、CustomPooling CNN [49]、Optical Flow(Deepfake video detection through optical ow based cnn)、Slowfast [20]

优化器：SGD

1. 总结思考

该方法在FF++不同压缩下的检测性能都很好，尤其是在低质量上表现好。基于发现：压缩将污染伪影而使得伪影难以被RGB域捕获，而且从频谱表示中直接提出CNN特征用于发现伪影不可行+伪造人脸具有独特的局部频域统计信息，所以考虑两方面的双流架构。其他卷积频域像FFT和DCT没有考虑局部一致性和转换不变性。

空间域RGB、HSV对微小伪造线索不敏感，所以难以在空间域检测。频域分析已经在**超分辨率**、纹理等应用中使用。

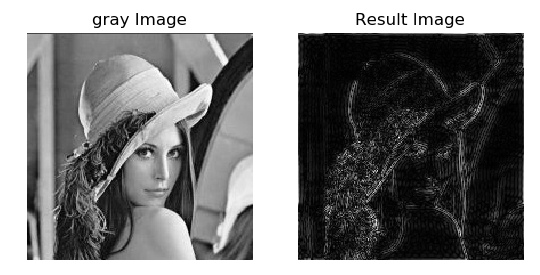
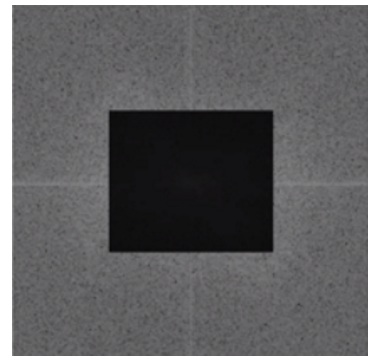
可以将分辨率提高之后提频域或者映射到哈希空间。通过高通滤波之后可获得真假图像的频谱差异。

实验图6中展示了Face- Forensics++上LQ任务中数据的**t-SNE特征空间**，以比较不同检测方法对于数据的分离情况。数据集分布比较也是t-SNE特征空间进行呈现的。

实验结果显示①NT依然是具有挑战的伪造。②低频揭示全局，中、高频揭示细节伪影。尤其高频线索与伪造敏感边缘和纹理相关。③局部统计信息对不稳定和有噪声的频谱更加鲁棒。

**高通滤波器**是指通过高频的滤波器，衰减低频而通过高频，常用于增强尖锐的细节，但会**导致图像的对比度会降低**。该滤波器将检测图像的某个区域，根据像素与周围像素的差值来提升像素的亮度。下图展示了“Lena”图对应的频谱图像，其中心区域为低频部分。

接着通过高通滤波器覆盖掉中心低频部分，将255的点变换为0，同时保留高频部分，其处理过程如下图



左边图为lena的灰度图，右边图为高通滤波器提取的边缘轮廓图像，它通过傅里叶变换转换为频谱图像，再将中心的低频部分设置为0，再通过傅里叶逆变换转换为最终输出图像。

DCT(Discrete Cosine Transform)离散余弦变换是一种经典谱分析方法，属于**离散傅立叶变换的一种特殊情况**，即在变换后的傅立叶级数中只包括余弦项，变化后的数据比较集中。经过**DCT变换可以将图片从色彩域转换到频率域，将原始图像的信息块转化为代表不同频率分量的系数集**。它是一种广泛使用的压缩方法，首先把每个单独的彩色图像分量分成8×8图像块，然后经过二维DCT变换，其**低频分量都集中在左上角，高频分量分布在右下角**。变换之后还是8×8的数据块，也就是说DCT变换本身并没有起到压缩数据的作用，但是它为以后的数据压缩奠定了必不可少的基础。

经过DCT变换的数据有2个优点：1，信号常将其能量的大部分集中于频率域的一个小范围内，这样一来，描述不重要的分量只需要很少的比特数；2，频率域分解映射了人类视觉系统的处理过程，并允许后继的量化过程满足其灵敏度的要求。对于变换之后的8×8的数据矩阵块，矩阵最左上角的直流 （DC）系数幅度最大，**以DC 系数为出发点向下、向右的其它DCT 系数，离DC 分量越远，频率越高，幅度值越小**，即图像信息的大部分集中于直流系数及其附近的低频频谱上，**离DC 系数越来越远的高频频谱几乎不含图像信息**，甚至于只含杂波。

与[RGB](http://en.wikipedia.org/wiki/RGB_color_space" \t "https://qastack.cn/signals/2687/_blank)不同，[HSV](http://en.wikipedia.org/wiki/HSL_and_HSV" \t "https://qastack.cn/signals/2687/_blank)将亮度或图像强度与色度或颜色信息分开。这在许多应用程序中非常有用。例如，如果要对彩色图像进行直方图均衡化，则可能只希望对强度分量进行此操作，而不必对颜色分量进行处理。否则，您将获得非常奇怪的颜色

HSV都是一种将RGB色彩模型中的点在原著坐标体系中的表示法。这中表示法试图做到比基于笛卡尔坐标系的几何结构RGB更加直观。HSV即色相(Hue)，饱和度(Saturation)，明亮度（Lightness），又称为HSB，其中B即英语Brightness。

YCbCr有的时候会被写作：YCBCR，是色彩空间的一种，通常会用于影片中的影像连续处理，或是数字摄影系统中。Y为颜色的亮度（luma）成分、而CB和CR则为蓝色和红色的浓度偏移量成分YCbCr颜色空间中的Y是亮度通道，Cb是蓝色分量，Cr是红色分量。它在电视系统中比较常用，比如早期的黑白电视机使用彩色电视信号线，就可以单独使用亮度值；这种功能RGB颜色空间就做不到，因为我们不能仅仅使用RGB中某个通道作为亮度信号来使用。