

Run CVIT

LRNet CVPR2021

Combining EfficientNet and Vision Transformers for Video Deepfake Detection ICIAP 2021

Watch your Up-Convolution: CNN Based Generative Deep Neural Networks are Failing to Reproduce Spectral Distributions CVPR2020

DeepFake Detection Based on Discrepancies Between Faces and Their Context TPAMI2022

Protecting Celebrities from DeepFake with Identity Consistency Transformer CVPR2022

Detecting Deepfakes with Self-Blended Images CVPR2022

What makes fake images detectable? Understanding properties that generalize ECCV2020

CORE: Consistent Representation Learning for Face Forgery Detection

1 研究现状

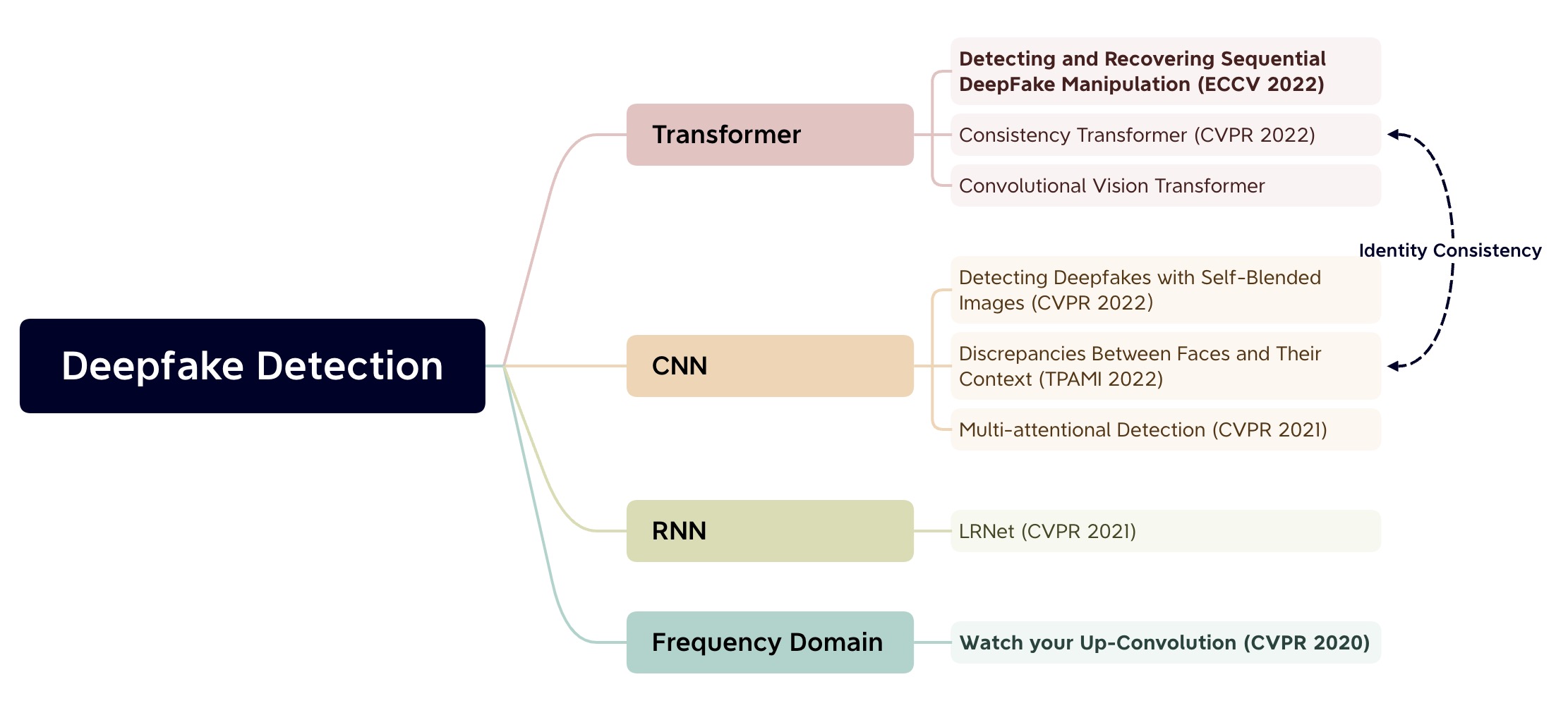
深度学习技术的快速发展，特别是生成式对抗网络(GAN)，导致了Deepfake的产生。Deepfake算法可以创建虚假的图像和视频，人类难以将它们与真实的图像和视频区分开来，对个人隐私和国家安全都构成了巨大的威胁。伴随着Deepfake技术不断迭代，检测合成人脸的挑战也越来越多。Deepfake检测通常被认为是一个二元分类问题，包括Deepfake图像检测和Deepfake视频检测，其检测方法并不完全相同，但大体上可以分为人为视觉特征提取和深度学习两类。

在人为视觉特征方面，S. McCloskey等人提出使用简单的颜色特征和SVM分类器来检测Deepfake图像，而Alexandros Haliassos等人利用读唇语学习到的丰富表征，实现了对未见伪造类型的最先进的泛化，同时对各种常见损坏的鲁棒性明显高于其他方法。

近期，使用深度学习进行Deepfake检测取得了很大的进展。为了避免专注于特定的篡改取证，实现鲁棒的篡改检测，Zhou.P等人提出了一种双流网络结构来捕获篡改伪迹证据和局部噪声残差证据。H. Dang, 等人在检测中引入注意力机制，并将其与流行的CNN模型相结合，如exception-Net和VGG16，最终取得了令人印象深刻的结果，表明了注意力机制的重要性。Guoet等人提出一个新的CNN模型——SCnet，采用了分层特征提取块的结构，能够自动学习图像数据的高级取证特征。Li等人的工作Face X-ray尝试定位目标和原始人脸之间的混合边界，而不是捕捉特定操作的合成伪影，且可以在没有假图像的情况下进行训练。N. Yu等人则利用GAN生成Deepfake图像时留下特殊的“指纹”来进行检测。此外，Joel.Frank等人的工作首次在频域考虑假脸检测问题，揭示了频谱不同的核心原因是上采样操作，也暗示了现在GAN从低维隐空间映射到高维空间的结构性问题。

长期以来，解决Deepfake检测最常见的方式是学习一个二分类模型来进行真/假判别。但如今由于人脸编辑App的流行，我们可以非常方便地对人脸图片进行多步序列DeepFake篡改。为解决此类问题，来自南洋理工大学的研究人员在ECCV2022的工作中提出了检测并还原Deepfake篡改序列（Seq-DeepFake）任务，贡献了首个大规模的Seq-DeepFake数据集，并提出了一个简单且有效的Seq-DeepFake Transformer作为此新课题很好的起始方案。

2 论文精读



2.1 Watch your Up-Convolution（CVPR 2020）

深度生成模型从中间隐特征生成图像有两种方式：up-convolution 和 transposed convolution。论文在自编码器上使用不同的上采样方式重构真实图像，发现up-convolution方法缺少高频信息，而transposed convolution方法引入了太多的高频信息。基于这样的观察，作者设计了利用频域信息的Deepfake检测方法，如下图所示，图像经过DFT后，将2D Power Spectrum根据不同半径池化成1D Power Spectrum，最后使用SVM进行分类。

2.2 LRNet（CVPR 2021）

论文提出了LRNet框架，通过对精确的几何特征进行时间建模来检测Deepfakes视频。LRNet由四个部分组成：人脸预处理模块、校准模块、特征嵌入过程和RNN分类过程。它通过检测异常的面部运动模式和时间不连续性来暴露被操纵的面部，并且整个框架只需要训练RNNs部分。与以前的方法相比，LRNet更轻量，更容易训练，并且在检测高度压缩或噪声污染的视频中表现出鲁棒性。

2.3 Multi-attentional Deepfake Detection（CVPR 2021）

这篇论文将Deepfake检测表述为一个细粒度的分类问题，并提出了一种新的多注意力的Deepfake检测网络，使用多注意力机制捕获人脸的多个区域的特征。模型整体框架基于Xceptionnet，在主干上新增了纹理增强模块、注意力模块、注意力pooling层。

2.4 Identity Consistency Transformer（CVPR 2022）

这篇论文中，作者提出了身份一致性Transformer，这是一种新的人脸伪造检测方法，专注于高层语义，特别是身份信息，并通过发现内外人脸区域的身份不一致来检测可疑人脸。身份一致性变换器结合了一致性损失，用于身份一致性确定。身份一致性Transformer在不同的数据集上以及现实世界中各种类型的图像退化形式上都表现出优异的泛化能力。

2.5 Detecting and Recovering Sequential DeepFake Manipulation（ECCV 2022）

这篇论文提出检测并还原DeepFake篡改序列（Seq-DeepFake）任务。相比于现有基于二分类 (真/假) 的 DeepFake 检测，作者提出 DeepFake 篡改序列检测这个新的研究课题，建立并开源了全球首个 Seq-DeepFake数据集。基于此数据集，进一步提出Seq-DeepFake Transformer（SeqFakeFormer）来准确地检测篡改序列。根据检测出的篡改序列，我们可以逆序还原出原始人脸。

论文指出，现有的人脸编辑算法基本基于GAN，会导致在编辑一种人脸部件/属性后，会间接影响其他人脸部件/属性。事实上，篡改顺序会进一步影响篡改所影响的空间关系。也就是说，我们可以在空间篡改痕迹的基础上捕捉到序列篡改痕迹, 并最终据此检测出 DeepFake 篡改序列。根据上述观察，作者认为可以把Seq-DeepFake看成是一种特殊的image-to-sequence problem，并提出Seq-DeepFake Transformer来检测篡改序列。

2.6 Detecting Deepfakes with Self-Blended Images（CVPR 2022）

这篇论文的目标是构造一类Deepfake detection训练集，作者提出了一种称为自混合图像（SBI）的新型合成训练数据来检测深度伪造。SBI是通过从单个原始图像中混合伪造源图像和目标图像，再现常见的伪造伪迹（例如，混合边缘、源图像与目标图像之间的统计不一致性）。SBI背后的关键思想是，更通用且难以识别的假样本鼓励分类器学习泛化且健壮的表征，而不会过拟合于特定操作的伪迹。

2.7 Deepfake Video Detection Using Convolutional Vision Transformer

在这篇论文中，作者提出用于Deepfakes 检测的Convolutional Vision Transformer，其包含两个组件：卷积神经网络 (CNN) 和 Vision Transformer (ViT)。CNN 提取可学习的特征，而 ViT 将学习到的特征作为输入并使用注意力机制对其进行分类。其主要贡献是在 ViT 架构中添加了一个 CNN 模块，并在 DFDC 数据集上取得了有竞争力的结果。

2.8 Discrepancies Between Faces and Their Context（TPAMI 2022）

这篇论文提出了一种识别假图像的新信号，该信号是基于比较内部面部内区域(直接操作的区域)和外轮廓区域(我们所知的所有面部操作方法都不会改变外部轮廓)。表示面部和外轮廓的两个向量是通过训练两个独立的人脸识别网络获得的：一个网络基于面部内区域识别身份，另一个网络基于外轮廓区域识别身份。

3 研究思路

4 前期准备

* 数据集
* 环境

5 初步实验