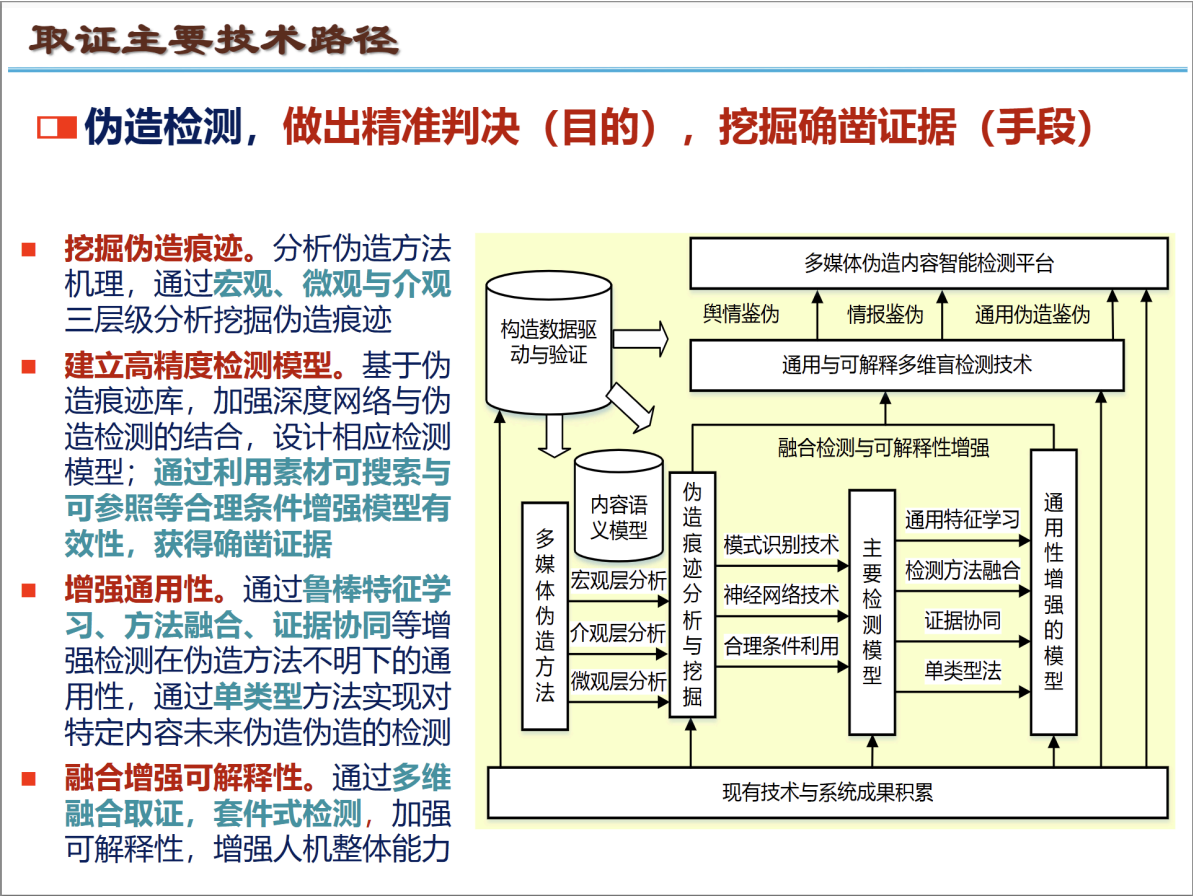


虚假内容检测

取证技术汇总



- 现代可以检测DeepFake视频帧间抖动，GAN图像伪影、纹理与色彩不均、光泽不连续等
- 基于帧间光流运动异常的假脸视频检测：光流是在两个连续帧之间计算的矢量场，用来描述相对于观察者的运动所造成的观测目标、表面或边缘的运动；采用帧间光流异常对伪造人脸视频进行取证分析；适用于深度伪造视频与帧编辑后的视频
- 进行DeepFake特征挖掘，将深伪视频检测问题转化为帧预测异常问题（非最优编码特性）：对于虚假视频，由于其表情僵硬缺乏灵活性，使得预测帧与目标帧的相关性较低

技术路径一：伪造特征挖掘与检测（质量异常）

■ 在优化实现与集成现有方法基础上，针对视音频伪造的新质量下降情况进行了针对性的检测模型研发

- 对深度伪造视频**前后抖动**、**帧光流不连续**、**伪造画面伪影与光晕**、**预测误差等**，提出了3D时空卷积网络检测、基于帧间光流梯度变换特征的删帧插帧定位、基于纹理差异的AI整体伪造图像检测、基于多尺度 Transformer 的整体生成图像检测、基于静音片段特征一致性的异源拼接语音检测、双流网络等



操作痕迹：篡改数字图像和掩饰篡改痕迹的过程中，往往会在篡改图像中留下操作痕迹：基于几何变换及插值痕迹，基于人为模糊痕迹，基于中值滤波痕迹，基于JPEG重压缩痕迹等。

- 区域相似：同源复制移动与充填区域存在高度相似区，复制移动增加对象的伪造中存在极为相似的区域；当前主要方法是特征点匹配、相似块匹配。
- 复制痕迹：由于周边内容修补的对象删除伪造中存在相似的区域（碎片修复），使用基于多区域关联的样例修补检测、基于碎片拼接的样例修复修补方法等来检测
- 新挖掘痕迹：
 - 视觉伪造内容**空域**痕迹检测：视频换脸基于每帧图像人脸部分的处理与替换，人脸部位的分割与重组，造成预测边界加宽、模糊
 - 视觉伪造内容**结构域**痕迹检测：视频换脸使用非同源的人脸图片对目标图片的部分区域进行替换，两类图片的**内部结构**和**特征分布**存在差异；颜色空间特征和共生矩阵中均有所反映，故可将换脸篡改认为是一种特殊的拼接篡改问题，存在**异源特征**，借助深度网络将来源不同的两种图像分割成两类不同的图像区域。
 - 视觉伪造内容**频域**痕迹检测：GAN生成图像中进行了上采样，导致真实图像与生成图像在DFT的幅度谱和相位谱存在显著差异，表现为GAN生成图像在全幅DFT变换后，幅度谱在中低频处存在突变，相位谱在低频处存在水纹状的异常；GAN生成图像在进行全幅DCT变换后会留下上采样插值痕迹，出现周期性的网格
 - 合成与转换语音**频域**痕迹检测（频谱粗糙）
 - 基于**声谱图**特征不一致性的异源拼接语音检测：基于拼接语音在MFCC特征上的跳跃性，可以基于全卷积神经网络逐窗口检测语音的拼接特征

技术路径二：可参考线索发掘（总）

在伪造图像检测的部分应用场景中，存在有利的检测条件，典型地包括：

- 伪造对象或虚假内容**素材**可搜索（也适用于假图新用等情况）
- 特定人物或场景**正样本**可参照情况
- 意义：在搜索成功后检测结果从不确定的统计结果变为**确凿**的结果；并者使得特定检测不受伪造发展影响



对象匹配：也称为受限检测。通过检索找到同类图像、通过场景与对象匹配，找到捐献者与接受者。基于多尺度匹配定位的伪造视频素材溯源检测（针对换脸、口型驱动、操控）：包括来源检索、内容匹配和伪造定位三个主要步骤。分别由空洞卷积深度匹配网络、检测网络和判别网络三部分实现。

场景匹配：基于同类场景感知杂凑检索的伪造场景素材搜索（针对未替换背景的多数视频伪造）：在网络训练中将同类场景归为同类标注，得到场景感知杂凑；在检测中通过匹配杂凑获得同一类场景后，进行进一步检测

身份匹配：基于单分类（One Class，OC）的检测方法，提前进行特定人物、目标与场景图像的搜集与自然特征存储，可以仅仅基于正样本进行单类型训练（因此与伪造方法无关，适用于检测未来与未知的伪造方法）。在检测中，通过检测自然特征的变化程度识别伪造的图像。

- 基于建立正样本标志性特征集的OC方法：基于特征空间的OC检测（视频与语音）建立正样本标志性特征集，检测负样本的偏差
- 基于建立正样本自编解码的OC方法。
- 基于内容重建的特定人物伪造视频单分类识别与取证：该方法提取真实人物视频的细粒度特征。

技术路径三：鲁棒性增强（总）

当前伪造检测技术的主要提高方向是增强在**多数据源与多伪造方法**实际环境下的检测鲁棒性

提高检测的鲁棒性的可能途径

- 数据增广构建
- 鲁棒特征学习
- 方法融合
- 证据协同
- 动态检测（媒体去失配）

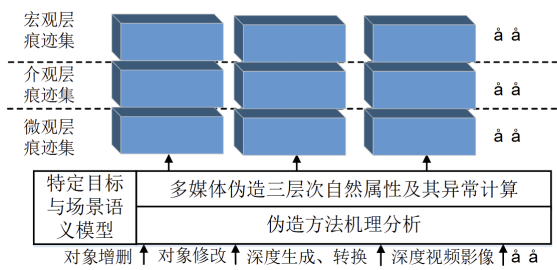
数据增广：采用数据遮挡增广、多视图增广结合注意力机制，增加特征学习的有效性，增强检测鲁棒性。采用各类增广数据的融合判决实现高精度的伪造事实检测

特征鲁棒：通过通用鲁棒特征学习、抗压缩特征学习等进一步增强检测在伪造方法不明下的鲁棒性。

动态检测：采用动态检测提高对媒体异样性的检测鲁棒性。在训练阶段针对每个可识别的类型设计与训练模型；系统输入前通过动态分析内容特点选择合适的检测方法组。

技术路径四：优化认知致效策略（总）

- 当前伪造检测技术存在主要偏向基于某一类特征组进行伪造事实识别，在多角度多层次与给出可解释证据上薄弱
- 从宏观、微观与介观三层级上综合完成研判并给出系列证据
 - 宏观异常痕迹：宏观自然语义约束规则违背
 - 介观异常痕迹：宏观与微观之间的“过渡地带”
 - 微观异常痕迹：自然微观噪声特性、各种数据域中的不可见痕迹与特征
- 高精度识别+系列取证+综合研判递进的检测框架



多域特征融合：多源证据协同（如空间域、频域、时域均证伪）。例如基于多域特征融合的换脸视频篡改检测算法：

- 在空间域上，换脸视频在眼睛、牙齿和皮肤纹理等细节存在特定的视觉篡改痕迹；
- 在频域上，换脸视频在生成过程中的上采样操作会引起DFT幅度频谱出现局部峰值；
- 在时域上，换脸视频通常是逐帧独立制作，难以保证相邻帧特征变化的连续性。

决策融合：多方法协同（多方式同时证伪），对不适合在特征级别进行融合的方法，采用可配置的决策融合框架；可以针对不同类伪造配置或者预设检测策略。

可视化证据增强：特点是取证证据丰富，融合检测结果精度高、可解释性强。

技术路径五：语义空间鉴伪。分为生成文本检测，语义一致性检测，外部知识库检测（如知识图谱），外部数据库溯源检测。

主要技术路径-总结

- 深入挖掘伪造特征以及多层次、多维度检测伪造内容，增强检测结果的可解释性
- 通过合理利用有利条件增强多媒体鉴伪结果的可靠性与对未来或未知伪造方法的适用性
- 通过扩展学习能力与融合检测等系列化手段，增强伪造多媒体鉴伪技术在应用中的鲁棒性
- 通过多来源、多视角、多维度的特征发掘与可视化增强，提升证据对于事实的说服力
- 借助大模型的知识融合推理能力，提升鉴伪在语义空间的可解释性

反取证技术

- 伪造成像设备指纹：

- 伪造CFA差值特征：篡改过程中再次使用插值算法；估计失真-矩阵求解-最小二乘；非线性滤波，调整图像尺寸，干扰像素间相关性
- 伪造模式噪声特征：替换原有噪声模式；添加新的噪声模式
- 隐藏篡改遗留痕迹（即编辑指纹）
 - 重采样反取证策略：中值滤波（低通滤波）消除周期性；添加噪声扰乱插值过程。
 - 重压缩反取证策略：
 - 基于量化痕迹的反取证：添加噪声，重新分配DCT系数。通过估计未量化的DCT系数直方图的分布，根据DCT系数的不同部分来改变添加噪声的方法。DCT系数的交流分量（AC）呈拉普拉斯分布，而直流分量（DC）没有固定分布。
 - 基于块效应的反取证：分块DCT变换，子块边缘高频分量丢失，且质量因子越小，块效应越明显，篡改区域与原始区域块效应不一致。反取证方法是估计未量化图像的DCT系数直方图分布，使用中值滤波+添加高斯白噪声来模拟。
 - 对比度痕迹反取证：对比度的修改是通过后期增强润饰，对特定区域像素值进行非线性映射实现的。它会导致增强后图像的灰度直方图出现奇异值。反取证方法是通过检索-映射-重采样，对直方图进行调整；或变换频域，进行定位和添加随机噪声，让直方图合理。
 - 中值滤波痕迹反取证：中值滤波使用去噪滤波器来消除篡改痕迹。它对图像的影响是让给相邻像素的灰度值相等的概率增大。反取证的方法是添加噪声，改变图像像素分布，或转换为基于线性滤波器的最优化问题。
 - 锐化痕迹反取证：锐化使用高通滤波，用于对图像的局部轮廓进行增强。它会导致图像平坦边缘的平坦带与过渡带交汇处出现幅值异常，边缘过冲效应。反取证的策略是加入抖动噪声。

反取证的应用和需解决的问题

反取证应用1

■ 隐私保护

■ 隐藏隐私信息

■ 隐藏证据来源

反取证应用2

■ 产权保护

■ 保护软件内部处理算法

■ 阻止逆向还原

需解决的问题：

1. 算法通用性。大部分算法都是针对性的定向反取证，没有做到通用性。
2. 安全性。有取证特征的单一性和知识结构的片面性，往往只利用或针对某一种或少数几种特征来判断是否被篡改；且自身也有引入篡改痕迹的可能性。
3. 可靠性：反取证过程还需要平衡视觉效果，让结果尽量自然；在这个基础上实现精细反取证。

视频内容生成

换脸技术（人脸对象编辑）

替换

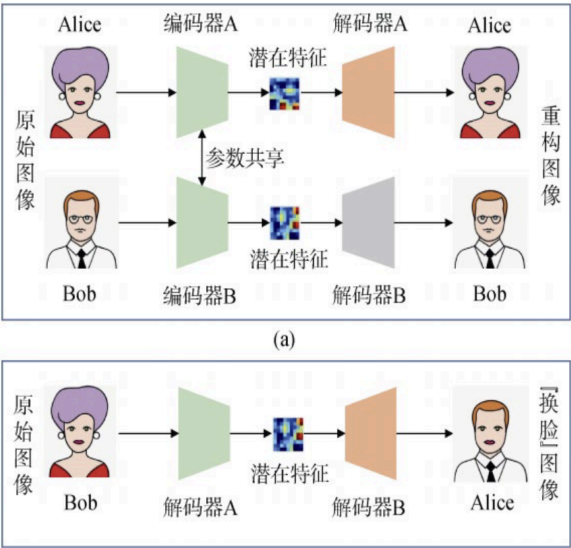
使用源人物的内容替换目标人物，使得目标人物身份变成了源人物身份。

- 基于经典计算机图形学：首先对源人脸和目标人脸进行面部三维建模（点、面），然后将源人脸的纹理和表情信息应用到目标人脸。技术框架：人脸建模、纹理合成和渲染。
- 基于深度学习：人物固有特征（长相或模样）定义为结构域，人物表情定义为风格域；基于编码-解码器结构实现图像的结构和风格分离。典型的是 faceswap。

■ AE结构

□ 编码器可以提取人物面部的**动作和表情状态**

□ 解码器保存人物的面部**风格和纹理**



重现 (Reenactment)

使得行为一致（表情、嘴部、眼部、头部及躯干）。包括语音驱动人像口型替换、视频驱动的人像动作替换。

- 语音驱动视频人像口型替换

■ 技术特点

■ 复杂的多模态任务

- 音频表示（梅尔频率倒谱系数MFCC、DeepSpeech、基于Transformer的预训练模型）
 - 面部建模（面部关键点和人脸3D参数）
 - 音频到面部序列（LSTM建模序列化数据和通过Cross-Attention机制来学习跨模态数据的同步）
 - 后处理（人脸去模糊）
- 基于特征融合：语音内容编码器和人脸特征编码器联合使用，将两个模态的特征进行融合，如Speech2Vid算法。
 - 基于音素同步：音素是最小的发声单元，视素可以是预先定义的面部关键点或者是3D混合模型系数。通过语音来预测音素组序列，然后根据音素组与视素的对应关系（预定义的音素 / 视素库）来得到可以有效的表示说话人脸的参数取值，如VisemeNet算法。
 - 基于人脸关键点：根据语音特征来预测人脸关键点位移序列，动态效果和泛化能力更好。如ATVGnet算法。

- 基于3D表情参数：通过视觉追踪模型得到3D的面部建模参数，通过序列化建模方法学习语音和面部表情参数之间的关系，最后将面部表情参数定向到3D人脸模型中。如AudioDVP算法。
- 视频驱动的人像动作替换：把视频中的表情姿态迁移到另一个视频中（例如虚拟主播生成、配音视频的口型配准）。

■ 技术特点

■ 复杂的多模态任务

- 获取源人脸的图像以及驱动人脸的视频
- 将驱动人脸的面部动作姿态迁移到源人脸上
- 将重演图像序列处理成视频
- 基于姿势描述子：人脸姿势描述子指的是显式的人脸关键点或隐式的高维特征。将源人脸的特征信息和驱动人脸的姿势描述子融合，使特征拼接与融合，来实现表情迁移。如Face2Face、ReenactGAN。
- 基于3维模型，把人脸解耦为身份、表情、姿态等 3D 参数，做参数级迁移。如PIRender算法通过语义描述符来精确控制图像运动，提高了人像重演阶段对于人脸的可控性
- 基于运动场预测：直接预测像素级稠密运动场 / 扭曲场，用运动场对源人脸做像素级扭曲，不依赖关键点或 3D 模型。如 X2Face。

编辑

添加、更改或删除目标身份的外貌属性，如增加头发。

合成

在没有目标身份作为基础的情况下创建新角色。

整体生成

- 数字人生成技术：
 - 基于图像：图像采集、三维重建、模型对齐、模型计算、模型渲染
 - 基于模型：模型构建、模型参数化、参数估计
 - 数字人驱动技术（内驱）：语音驱动、音素驱动、语音和音素多模融合驱动、传感器数字人驱动、语义和声音驱动
 - 数字人驱动技术（外驱）：动作捕捉、面部捕捉、声音捕捉
- 文生视频大模型

传统编辑

1. **视频删插帧及其检测**：最简单最直接的视频编辑方法。通过视频剪辑工具进行帧的删除、插入和复制操作。

检测方法分类：

- 时空域特征（关注像素变化）：基于单帧提取特征得到帧序列的特征描述符，如HOG特征、Zemike矩特征、亮度特征等；基于相邻帧提取相关特征得到帧间变换特征描述符，如帧间互信息特征、光流特征、灰度值相关系数特征等
- 编码域特征（关注压缩过程）：解压缩、序列伪造、重压缩三个步骤中存在的取证痕迹；GOP和预测残差的变化

■ 基于方向梯度直方图（HOG特征）

■ HOG特征

- 方向梯度直方图（Histogram of Oriented Gradient, HOG）特征，是计算机视觉领域中用于目标检测的一种特征
- HOG特征提取五个步骤：①图像灰度化；②Gamma矫正，减少光照因素的影响；③梯度计算，计算每个像素的梯度大小和方向；④梯度方向直方图计算，按照分块将方向划为9个类别，每个类别投票计数；⑤将分块HOG特征串联得到整幅图像的HOG特征
- HOG特征能够表示图像内容，提取视频每一帧的HOG特征，计算其特征的相关性，若HOG特征出现异常的不相关现象，则判定为剪辑伪造视频。

■ 基于方向梯度直方图（HOG特征）

■ 检测步骤

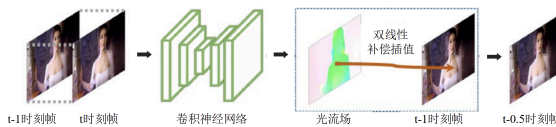
- ①视频解码到帧序列；②逐帧调整分辨率，缩放为128×128像素；③逐帧提取HOG特征；④计算HOG相关性高斯模型特征；⑤使用Grabb检验检测异常点

2. 视频帧率上转及其检测：帧率上转即提升帧率，改善编辑后视频播放连续性。手段有线性插值、运动补偿、光流法、神经网络插帧。

- 光流法相对于运动补偿，补偿单位为像素级，操作单位由像素块变成单像素、半像素等更细小的粒度。
- 神经网络插帧使用卷积神经网络直接作预测，分为基于光流估计的插帧和基于卷积核的插帧。

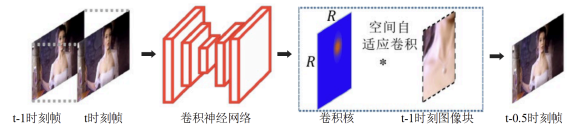
■ 基于光流估计的视频插帧

- 线性运动模型假设
- 预测出双向的光流场
- 双线性插值补偿



■ 基于自适应卷积核的视频插帧

- 形式化地定义为对参考帧的局部块的卷积操作
- 对每一个待输出像素估计出一个空间自适应的卷积核



检测方法分类（只针对线性插值法具备有效检测能力）：

- 时空域特征：帧间相似度特征、纹理变化特征、边缘强度特征
- 残差特征：空时Markov特征、效应指示图特征

■ 时空Markov特征检测

■ 特征提取步骤

- $F(x, y, t)$ 表示t时刻的视频帧， x, y 表示像素坐标
- 计算时间帧差矩阵（Temporal Frame Difference Matrix, TFDM）

$$TFDM(x, y, t) = \frac{1}{2} [(F(x, y, t) - F(x, y, t-1)) + (F(x, y, t) - F(x, y, t+1))]$$

- 计算空间差分滤波特征，八个方向 $\{1 \leftarrow, \nwarrow, \nearrow, \rightarrow, \swarrow, \searrow, \downarrow, \uparrow\}$ ，从上到下 $\downarrow$ 为

$$STFDM^1(x, y, t) = TFDM(x, y, t) \cdot TFDM(x, y-1, t)$$

- 定义阈值T截断STFDM特征值，如果大于T或者小于-T，它将被T或者-T替代

■ 时空Markov特征检测

■ 特征提取步骤

- 沿垂直方向，使用一阶Markov过程建模 $STFDM^1(x, y, t)$ ，计算经验矩阵如下，其它七个方向的经验矩阵以同样方式计算

$$M_{u,v,t}^1 = \frac{1}{m \times n} \sum_x \sum_y \Pr(STFDM^1(x, y, t) = u | \Pr(STFDM^1(x, y-1, t) = v) \quad u, v \in \{-T, \dots, T\})$$

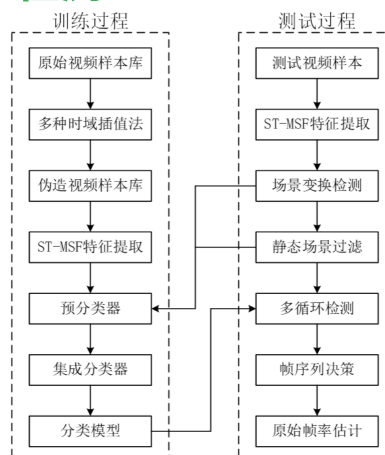
- 特征降维，4个水平或垂直的矩阵，4个对角矩阵分别平均形成两个特征子集，视频帧的ST-MSF特征定义如下，特征维数为 $(2 \times T + 1)^2$ 。

$$\begin{cases} S_{1, \dots, s}^t = \frac{1}{4} (M_t^{\leftarrow} + M_t^{\rightarrow} + M_t^{\nwarrow} + M_t^{\nearrow}) \\ S_{1, \dots, s}^t = \frac{1}{4} (M_t^{\swarrow} + M_t^{\searrow} + M_t^{\swarrow} + M_t^{\searrow}) \end{cases}$$

■ 时空Markov特征检测

■ 训练分类器

- 场景变换检测器，处理镜头切换。
- 静态帧序列检测器，过滤掉无运动的静止场景。
- 采用图像隐写分析常用的集成分类器。
- 对常用的像素块运动补偿帧率提升方法，检测准确率超过95%。



3. 视频对象增加及其检测：同源复制移动伪造/异源拼接。

检测方法分类：

- 使用类似于图像的复制粘贴/拼接检测来检测视频

■ 基于双流网络和分割模型的方法

■ 视频拼接定位

- 使用双分支结构提取特征
- 使用条件随机场细化定位结果
- 采用视频目标跟踪策略提高定位精度

4. 视频对象删除/修补：对掩膜遮蔽的视觉内容进行修复与重建，保证图像视觉自然效果。

■ 手段

■ 传统视频修复方法

- **像素扩散**：基于流体力学思想沿像素梯度方向执行扩散过程，是图像修复扩展后的经典算法。Navier-stokes, fluid dynamics, and image and video inpainting. (CVPR2001)
- **像素块匹配**：将经典图像修复PatchMatch算法从2D升级到3D，提升时空域像素方块检索匹配效率。Video inpainting of complex scenes. (IAM Journal on Imaging Sciences 2014)
- **光流指引**：同时估计缺失区域中的光流和颜色，根据前向/后向稠密光流合成时间上的关联像素。Temporally coherent completion of dynamic video. (ACM Trans. Graph. 2016)

■ 手段

■ 深度视频修复方法

- **2D或3D卷积模型**：端到端模型，在卷积编解码器模型中穿插光流预训练模型或者注意力机制。DVI(CVPR2019) → OPNET(ICCV2019), CPNET(ICCV2019).
- **光流指引**：非端到端模型，包括光流估计、光流修复、像素流动填充、循环一致性约束、单图像修复等方法。DFGVI(CVPR2019) → FGVC(ECCV2020) → FGT(ECCV2022).
- **视觉Transformer**：采用Transformer中Multi-head Self-Attention机制模拟像素块Patch匹配过程。STTN(ECCV2019) → FuseFormer(ICCV2021) → E2FGVI(CVPR2022). ECFVI (ECCV2022). → ProPainter(ICCV2023).
- **Internal学习**：随机初始化的卷积网络对自然图像建模具有能力，在卷积重构模型上添加时间约束，并过拟合训练。ILA(ICCV2019) → Implicit Long-range Propagation(ICCV2021). → Discrete Latent Transformer(CVPR2022).

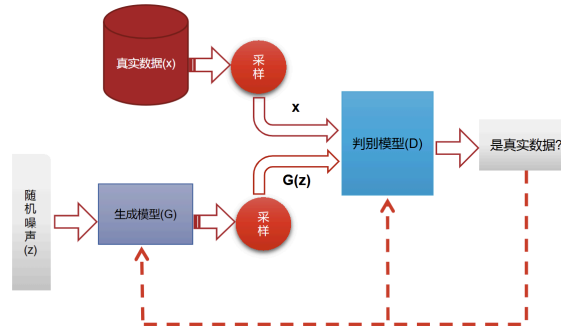
检测方法分类：

- 传统的视频修补检测：利用光流不一致、统计相关性、不规则变化和像素相关性等，如基于时空域三叉网络的视频对象删除检测方法
- 深度视频修补检测：预滤波特征提取方法凸显修补伪造痕迹，并设计端到端神经网络实现视频对象删除修补的检测定位。
 - 基于ELA帧和ConvLSTM的深度视频修补检测方法
 - 基于频域特征和时空Transformer的检测方法

图像深度编辑

- 风格迁移：

- 基础：训练 CNN 分类器时，接近输入层的 Feature Map 包含更多的图像的纹理等细节信息（浅层），而接近输出层的 Feature Map 则包含更多的内容信息（深层）。
- GAN：传统的 GAN 模型使用单类图片训练 GAN，输入随机向量 z 输出此类图片，不能指定某个条件。条件 GAN 在输入上可以输入一个条件，生成符合该条件的 GAN，可以堪称多个 GAN 进行复合加个 if。
 - Pix2Pix：Pix2Pix 使用成对图片训练 GAN，输入不再是一整类图片+随机向量，而是图片对（经典上使用草图），相当于把随机向量 z 换成草图 z ：



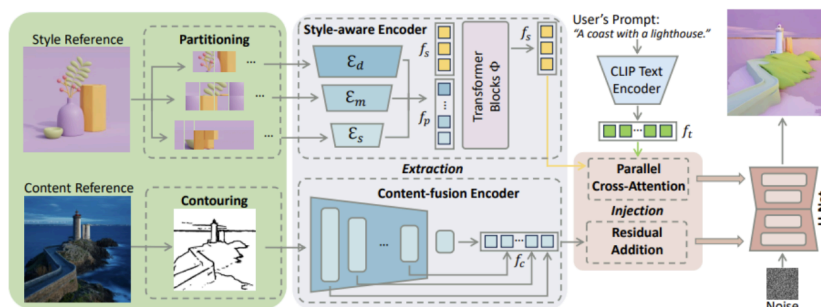
- Cycle-GAN 又变回了使用一整类图片进行训练，不需要成对的样本进行训练了，原理上，它同时学习了两个生成器 + 两个判别器，损失保证 $A \rightarrow B \rightarrow A$ 要能还原回来。称为 Cycle consistency（循环一致性）。

缺点：存在模式坍塌问题（生成器认为只要能骗过判别器 D 就够了，而不需要覆盖所有真实分布）导致 GAN 的方式生成的图片的多样性可能较低；GAN 很难对风格进行精准控制，如笔触或色彩。

- StyleShot (diffusion model)：

StyleShot

- 风格感知编码器**：从参考图像中提取更具表现力和更丰富的风格 embedding
- 在预训练的 **Stable Diffusion** 模型中引入风格 embedding，并采用 Cross-Attention 实现风格注入
- 内容融合编码器**：整合风格与内容



- 局部生成：

- Auto-Encoder 换脸：A, B 公用编码器，但不共用解码器。auto-encoder-flowface，分为面部重塑网络，用于减少源人脸与目标人脸之间的形状轮廓差异；人脸交换网络，通过预训练的掩码自编码器提取面部特征，将源人脸的面部特征迁移到目标人脸

图像取证(传统编辑)

通过对图像多维度特性的分析来判断图像的**真实性**、**完整性**、**原始性**和反映场景的**客观性**。分为主动取证和被动取证。主动取证就是主动加有数字签名或者水印可以查证。

评价标准

- **算法准确率**
 - 所有的**正确预测**（正类、负类）所占的比重
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
- **算法精确率**
 - 也叫查准率，即正确预测为正占**全部预测为正**的比例
$$\text{Precision} = \frac{TP}{TP + FP}$$

评价标准

- **算法召回率**
 - 即正确预测为正占**全部实际为正**的比例
$$\text{Recall} = \frac{TP}{TP + FN}$$
- **F1-Score**
 - 一般来说准确率和召回率呈**负相关**，引入F1-Score作为综合指标，就是为了平衡准确率和召回率的影响，较为全面地评价一个分类器。F1-Score越大，说明模型质量越高
$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN}$$

评价标准

- **算法召回率**
 - 即正确预测为正占**全部实际为正**的比例
$$\text{Recall} = \frac{TP}{TP + FN}$$
- **F1-Score**
 - 一般来说准确率和召回率呈**负相关**，引入F1-Score作为综合指标，就是为了平衡准确率和召回率的影响，较为全面地评价一个分类器。F1-Score越大，说明模型质量越高
$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN}$$

指纹分为三种，根据图像得到的完整周期，分为获取、编码、编辑阶段。

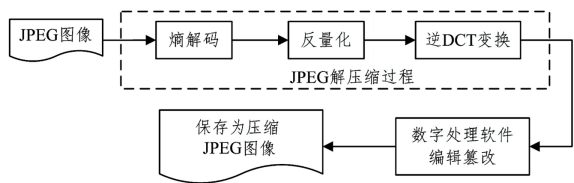
获取指纹：成像阶段产生

1. 基于成像器件的遗留特征痕迹：按粒度分为设备类型->设备型号->个体。
2. 基于成像设备特性的**编辑检测**：设备生成的图片有一定的特性，这些特性在编辑后会消失。常见是基于色差，传感器模式噪声，CFA插值模型，相机响应函数（光谱灵敏度函数）。

编码指纹：图像进行数字编码阶段产生

1. **重压缩**。动机是，篡改后的JPEG图像经历了双重压缩（解压-篡改-再压缩），而双重JPEG压缩操作会引入单次压缩所不具备的特性。基于重压缩痕迹的编辑检测，不能区分正常重压缩与编辑后重压缩。JPEG压缩编码过程如下：

1. 图像分割，8×8 方块切割。
2. 颜色空间转换 RGB -> YCbCr。
3. 对每个块单独进行 DCT 变换，把时域（即空间域，每个像素的具体值）转换成频域（像素值变化的快慢规律）。这一步是无损的。
4. 量化。进行zig-zag变换，留低频、丢高频。有损。
5. 进行熵编码，根据出现概率分配码字长度。无损。



对图像的影响

- **双重JPEG压缩操作会为图像引入单重压缩所不具有的特性**
- 量化痕迹：DCT系数簇拥在Q整数倍处
- 块痕迹：块的边缘处像素不连续

1. 基于DCT分布特性的检测：原理是两次不同步长的量化对DCT系数分布产生了影响（原本平滑的分布出现规律空缺）。
2. 基于分块边缘差异的检测：编辑后重压缩，篡改部分与8×8分块不完全一致，导致边缘区域对应的分块存在多重JPEG压缩特征。缺点是滑窗检测，计算复杂度大。

2. **重采样**。对图像对象进行编辑时，往往会对操作对象区域进行缩放、旋转等操作，这些操作通常涉及重采样，即进行估计和插值。基于重采样痕迹的篡改检测：

重采样

- 最邻近插值法（输出像素的赋值为当前点的像素）
- 双线性插值法（2*2矩阵所包含的有效点的加权平均值）
- 立方回旋插值（4*4矩阵所包含的有效点的加权平均值）

对图像的影响

- 相当于用采样函数与输入图像作**二维卷积**运算
- 导致插值点与周围像素点产生一定的**相关性**（部分像素值变成周围像素的**线性组合**）
- 导致图像**二阶导数**产生一定的**周期性**

1. 遍历每个像素，计算每个像素是其周围像素线性组合的概率；
2. 将概率矩阵转换到频域，频谱中是否存在不连续亮点判断像素点是否经过采样（存在周期性）

编辑指纹：图像后处理阶段

- 复制粘贴：最简单有效的图像篡改手段之一。重要特征（灰度范围、纹理、噪声等）与图像粘贴位置的周围区域具有良好的兼容性。
 - 前处理：保持复制区域与粘贴区域的空域**同步性**和**同质性**，如旋转、翻转、缩放、色度/亮度改变等
 - 后处理：消除诸如尖锐边缘等操作痕迹，如JPEG压缩、加噪、高斯模糊等

检测复制粘贴的通用技术流程：

1. 预处理。色彩空间转换（RGB-YUV），**离散小波**，提取低频子带；图像缩放；低通滤波。
2. 分块检测。
 - 根据像素点检测，将图像分成非重叠的方块，以方形窗口从图像的左上角到右下角每次滑动一个像素进行扫描；
 - 根据特征点，通过提取图片局部特征完成，具有紧凑性和不变性。
3. 特征提取。

■ 特征提取

- 基于离散DCT变换系数
- 基于主成分分析
- 基于小波变换系数
- 基于奇异值分解
- 基于不变特征点
- SIFT特征（尺度不变特征），对缩放、旋转甚至仿射变换保持不变性
- SURF特征（加速鲁棒特征），对缩放、旋转、噪声、模糊有良好的鲁棒性，且快

4. 特征匹配：选择合适的相似度准则，对特征向量之间的相似度进行度量。
 5. 相处误匹配。某些自然图像中也可能存在一个或多个成对的高度相似区域，识别复制区域和粘贴区域的特征，进行有效的区分来滤除误匹配。
- 拼接：最简单有效的图像篡改手段之一。两张不同的图片中的区域合并。同样可以分为前处理和后处理。检测原理是：拼接图像来源不同，所以**统计特性**不同。拼接会破坏原始图像底层统计特性，通过判别统计特征的一致性达到检测目的。

■ 拼接图像整体特征

- 二维图像的基本统计特征
- 自然图像统计模型
- 双相干系数和边缘百分比特征
- 图像质量度量
- 二元相似性度量等

■ 拼接图像局部特征

- 分析拼接部分与原图不同的特征
- CFA插值特征
- 光学特征（光学成像失真）
- 模式噪声，响应非一致性噪声（PRNU, Photo-Response Non-Uniformity Noise）
- 光照一致性

■ 拼接图像后期修饰特征

- 对拼接图像的后期处理留下处理痕迹
- JPEG重压缩检测
- 重采样检测
- 后处理特征

- 模糊：消除或淡化图像合成所产生的边缘痕迹。分为自然模糊（离焦/景深模糊）和人工模糊（低通滤波）。

■ 基于边缘特性的模糊检测

■ 从图像边缘出发，分析对边缘特性的影响

- 1、同态滤波，放大人工模糊的边缘
- 2、检测边缘特征点的异常

■ 基于空域特性的模糊检测

■ 破坏空域自然统计特征

- 1、破坏RGB通道的色调联合一致性，分块统计异常色调率
- 2、增强像素间的线性相关性，不协调像素点数量变化

$$|I(x, y) - c \times \sum_{(m, n) \in S} I(m, n)| > threshold$$

■ 基于频域特性的模糊检测

■ 破坏频域统计特性，影响高频低频功率谱分布

- 1、DCT变换特征
- 2、二维短时傅立叶变换特征
- 3、小波变换

- 润饰增强：弥补原始图像的不足之处。普遍存在的篡改后增强操作，通过改变特定色阶、对比度实现不同区域的突出和增强。目的是突出图片中的某些信息，改善图像的视觉效果；突出图像的特征，便于计算机处理。

■ 技术分类

■ 点运算算法：灰度级校正、灰度变换和直方图修正等

- 使图像成像均匀，或扩大图像动态范围，扩展对比度

■ 技术分类

■ 邻域增强算法：图像平滑，图像锐化

- 平滑：均值滤波、中值滤波。
- 锐化：高通滤波、空域微分法（梯度模运算）

- 二次获取(翻拍)：对非数字存储形式图像的快速复制，效果上属于对图像的**后处理**操作

技术分类：

1. 针对纸面打印翻拍图像的检测（反射光）
2. 针对LCD显示屏翻拍图像的检测（自发光）

- Seam-carving缩放：基于**内容感知**的图像缩放技术，尽可能的保留主要的视觉实体信息。

■ 原理

- 计算图片的梯度矩阵
- 根据梯度矩阵求得最小能量线
- 删除最小能量线
- 循环1-3步骤，直至达到设定次数（原图宽/高 - 目标图宽/高）

■ 针对Seam-carving的检测

- 对图像相邻像素之间的相关性造成破坏
- 通过检测图像像素之间的相关程度来判断图像是否经历了异常修改
- 马尔可夫转移概率矩阵特征
- 扩展马尔可夫特性的特征