arXiv:2305.05943v1 [cs.MM] 10 May 2023

Mover：基于掩码与恢复的面部区域一致性感知方法，用于深度伪造视频检测

Juan Hu 计算机科学与电子工程学院，湖南大学，中国

Xin Liao 计算机科学与电子工程学院，湖南大学，中国

Difei Gao Show Lab，新加坡国立大学，新加坡

Satoshi Tsutsui ROSE Lab，南洋理工大学，新加坡

Qian Wang 网络空间安全学院，武汉大学，中国

Zheng Qin 湖南大学计算机科学与电子工程学院，中国

Mike Zheng Shou，Show Lab，新加坡国立大学，新加坡

摘要

深度伪造技术已被广泛用于恶意目的，促使人们对开发深度伪造检测方法产生了广泛的研究兴趣。深度伪造操作通常涉及篡改面部部位，从而导致人脸不同部位之间出现不一致。例如，深度伪造技术可能会将微笑的嘴唇变成沮丧的嘴唇，而眼睛仍然保持微笑。现有检测方法依赖于伪造的特定指标，而这些指标往往会随着伪造模式的改进而消失。为解决这一局限性，我们提出了 Mover，这是一种新的深度伪造检测模型，利用深度伪造视频中不可避免的弱点——非特定的面部区域不一致性。Mover 会随机遮蔽感兴趣区域（ROIs）并恢复人脸，以学习非特定特征，这使得伪造人脸难以恢复，而真实人脸则可以轻松恢复。具体而言，对于真实人脸图像，我们首先预训练一个掩码自编码器，将人脸划分为三个部分并随机遮蔽感兴趣区域，以此学习面部区域一致性，然后根据未遮蔽的面部部位进行恢复。此外，为了最大限度地扩大真实视频与伪造视频之间的差异，我们提出了一种采用双网络的新模型，分别利用预训练编码器和掩码自编码器。1）对预训练编码器进行微调，以捕捉给定视频中不一致信息的编码。2）利用预训练的掩码自编码器映射人脸，并区分真实视频和伪造视频。我们在标准基准上的大量实验表明，Mover 非常有效。

CCS 概念

• **安全与隐私** → 安全与隐私的社会层面。

关键词

深度伪造检测、面部区域一致性、掩码与恢复

1 引言

深度伪造是指利用人工智能算法，生成令人信服但实际上是伪造的某人面部视频，例如让其说出从未说过的话 [6, 59]。尽管它具有潜力可应用于电影和娱乐等领域，但滥用深度伪造技术会对个人和社会构成重大威胁 [60]，这凸显了开发有效深度伪造检测技术的重要性。

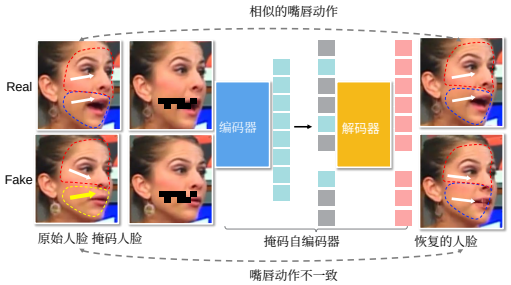

图 1：深度伪造视频中的面部部位往往缺乏一致性。利用这些不一致之处，我们的 Mover 方法能够有效区分真实人脸与伪造人脸：该方法仅使用真实图像进行预训练，却能比伪造图像更准确地恢复真实人脸。

目前已有许多检测 Deepfakes 的技术，大致可分为两类：基于隐式线索的方法和基于显式线索的方法。基于隐式线索的方法 [1, 2, 5, 33, 43, 45, 47, 50–52, 63, 64] 采用监督学习来区分真实视频与伪造视频，却未显式引入检测深度伪造视频的线索，因此难以理解其潜在的检测依据。近期采用显式线索的方法 [3, 4, 10, 16, 18–21, 24, 29– 31, 34, 37, 39, 41, 42, 44, 49, 58, 61, 65–67] 取得了更有前景的性能。然而，鉴于深度伪造技术的快速发展，各种伪造痕迹都可能被清除，使依赖特定面部特征的检测方法容易受到攻击。此外，在真实场景中，给定的视频可能由未知的篡改技术生成，这可能导致检测训练数据集之外的伪造内容时性能下降 [5, 28, 31, 39, 44, 45, 50– 52, 64]。因此，我们希望开发一种方法，它不依赖人工

设计的线索或特定的深度伪造模式，也不只是通过监督学习来区分真实与伪造内容。

我们的想法受到心理学和面部表情研究的启发[12, 32, 46, 56, 57], ，旨在利用真实人脸中固有的一致性。由于伪造人脸缺乏真实的人类心理特征，因此很难维持面部部位的一致性。一致性指的是各面部部位（如眼睛、眉毛和嘴巴）之间的协调与连贯，它们共同构成整体面部表情。真实人脸中存在这种一致性，但通过深度伪造生成方法 [16–18, 61, 62, 67]往往难以复制。例如，在图 1 中，FaceForensics++ 数据集 [45] 中的伪造人脸经过 Face2Face[54] 处理后呈现向上的姿态，但与其他面部特征不一致，因为脸颊和眼睛部位并未向上。通过检查像素级相关性（图 3）以及面部部位的不一致区域（图 5），我们可以进一步确认伪造视频中的这种不一致性。

我们如何利用面部一致性实现稳健的深度伪造检测？我们假设，尽管生成面部一致性完美的视频很困难，但创建一个能够在保持一致性的同时修复被遮挡人脸图像的模型则更简单、更可行。为实现这一修复理念，我们仅使用真实图像预训练了一个简单而强大的掩码自编码器（MAE）模型 [22] 。此外，我们根据面部部位设计其掩码模式（算法1），使模型能够学习非特定的面部部位一致性。对面部部位的考量至关重要，这使我们的工作有别于不考虑人脸语义部位、试图重建整张人脸的方法 [2]。因此，将预训练的 MAE 应用于面部部位不一致的伪造图像时，与真实人脸相比，原始图像和重建图像之间的差异会更大。由此，我们将 MAE 预训练称为第一阶段，它能提供强大的表征，用于区分真实人脸与深度伪造人脸。接下来，我们通过微调网络和映射网络对预训练模型进行微调，以实现更稳健的深度伪造检测，并检测来自未知域的伪造内容；我们将此称为第二阶段。

阶段 1：第一阶段的主要目标是预训练一个模型，以捕捉真实人脸中所有面部部位之间的一致性。受 [22], 启发，我们提出了一种名为 Mover 的新型自监督预训练方法，该方法基于人脸掩码与恢复。不同于 MAE [22],随机掩码图像的大部分区域，Mover 引入了一种基于面部部位一致性的掩码策略，用于随机掩码感兴趣区域（ROIs）。将人脸划分为不同部位，并随机选择一个部位以随机掩码感兴趣区域后，自动编码器便能通过重建（或恢复）缺失的像素，更好地学习所有面部部位之间的一致性。

第二阶段：第二阶段的目标是利用预训练模型区分真实视频和伪造视频。具体而言，该阶段对第一阶段得到的表征进行微调，使其能够更好地区分真实视频和伪造视频，并在多样的深度伪造模式上具备更强的领域泛化能力。为此，我们设计了两个子网络。第一分支（图 2-a ）是微调网络，用于聚合面部部位一致性，并对模型进行微调，使其能够将真实视频和伪造视频分类。利用第一阶段训练好的编码器提取特征后，模型可以学习真实视频与伪造视频之间在面部部位一致性方面的差异。第二分支（图 2-b）是映射网络，它利用第一阶段训练好的掩码自编码器重建人脸。通过构建映射网络，该分支可以将原始真实人脸映射为 Mover 重建的人脸，但无法将伪造人脸映射为 Mover 重建的人脸。通过最小化重建人脸与映射后人脸之间的均方误差（MSE）损失，模型放大了伪造人脸中的不一致性以及真实视频中的一致性。此外，为了明确提升对未见过的深度伪造模式的泛化能力，我们采用元学习，在训练阶段将不同类型的深度伪造视频划分为元训练集和元测试集。主要贡献如下。

（1）我们提出了用于深度伪造检测的 Mover，它通过遮蔽和恢复人脸，学会利用所有面部部位之间的一致性，为识别未知的 DeepFakes 提供了重要线索。据我们所知，我们是首个将掩码自编码器用于深度伪造检测的团队。我们发现，直接将掩码自编码器（MAE）用于深度伪造检测无法取得令人满意的结果，但对 MAE 进行独特改造后，便能够成功检测 DeepFakes。

（2）为了开发能够泛化到未见过的深度伪造模式的模型，我们引入了一种两阶段策略：第一阶段仅使用真实图像进行训练，学习不针对任何深度伪造类型或任何面部部位的表征；第二阶段则学习更加鲁棒的表征，最大化真实人脸与伪造人脸之间的差异。

（3）在包括 Face-Forensics++ [45]、Celeb-DF [36],WildDeepfake（WildDF） [68],和 DFDC预览版（DFDCP） [9] 在内的基准数据集上进行的大量实验表明，Mover 在各种指标下均取得了出色的性能。

2 相关工作

2.1 深度伪造合成方法

Li等[36] 将深度伪造合成方法分为两代。第一代方法包括 Face2Face [54], FaceSwap [13], NeuralTextures [53], 和 DeepFakes [8]。Face2Face 通过对源人脸进行变形来合成目标人脸。FaceSwap 应用渲染模型和颜色校正来替换面部区域。NeuralTextures 训练模型以学习目标视频的神经纹理特征，并执行面部重演。DeepFakes 基于两个自编码器和一个解码器来生成伪造视频。第一代方法存在一些已知问题，例如颜色不匹配、时间闪烁以及不正确的面部边界。第二代方法在前述方法的基础上提升了质量。具体而言，Celeb-DF [36] 应用一种算法来减少边界伪影并改善帧间连续性。Dolhansky等人 [9] 使用换脸算法生成面部比例较低的视频。

这些多样化的合成方法会留下不同的篡改痕迹，并导致数据分布出现偏差，使得在未知域中检测深度伪造视频颇具挑战。本文使用了包含上述两代方法的公开数据集中的视频。

2.2 深度伪造检测方法

基于隐式线索的方法主要通过监督学习来区分伪造与真实。早期方法 [1] 开发了 Mesonet，用于提取介观属性。为增强伪影并抑制人脸的高层内容，Masi 等人 [40] 提出了一种双分支循环网络，用于分离经过操纵的人脸。Zi 等人 [68]提出了采用注意力机制的 ADDNets，以实现先进的深度伪造检测。为提高鲁棒性，Cao 等人 [2] 早期尝试重建人脸图像，通过利用重建差异捕捉隐式线索并检测深度伪造视频。

基于显式线索的方法利用各种特定信息来提升检测性能。频率信息 [37, 44], 、局部纹理 [35], 和生物特征 [7] 等信号线索被用于区分深度伪造视频与真实视频。音频与视觉之间的情绪特征 [41], 、时间不一致性 [17, 25], 以及嘴唇信息 [21] 等语义线索，在深度伪造检测方面取得了显著进展。为开发具有泛化能力的深度伪造检测器，Zhao 等人 [65], 、Zheng 等人 [66] 和 Guan 等人 [19] 提出了可靠的框架。此外，Yang 等人 [62] 从面部区域提取特征，并利用图来建模面部区域之间的关系。为解决深度伪造检测问题，Shiohara 等人 [49] 开发了自混合图像（SBI）来检测 DeepFakes；在 Celeb-DF 和 DFDCP 上进行检测时，该方法实现了最佳泛化能力。

这些方法在当前基准测试中取得了令人鼓舞的结果，但它们依赖于特定线索。随着新的深度伪造技术不断出现，检测方法提取的特定线索可能会被通过在伪造视频合成过程中进行有针对性的训练而规避。未知的深度伪造模式会留下复杂的篡改痕迹，而非可预见的特定特征，这促使我们进一步探索利用非特定线索开发面向未知域的检测器。因此，我们提出一种自监督方法，通过重建随机掩码的人脸来学习不针对任何特定面部部位的鲁棒特征，从而促使模型学习面部部位之间的一致性。

2.3 掩码自编码器方法

掩码自编码器（MAE） [22] 的出现对我们的领域产生了重大影响。此后，人们提出了扩展 MAE [14], 、Video- MAE [55], 和 ConvMAE [15] ，以学习强大的表征。这些 MAE 方法基于非对称编码器-解码器架构。扩展 MAE [14] 主要将 MAE 扩展至时空表征。VideoMAE [55] 基于 MAE 开发了专门的管状掩码设计。ConvMAE [15] 则在 MAE 中引入了卷积结构。尽管这些改动相对较小，却显著提升了各种下游任务的性能。

MAE [22] 旨在用于视觉 Transformer (ViT) 的预训练，而非深度伪造检测。MAE 以较高比例随机掩码图像块，但我们的改进掩码策略与 MAE 的不同之处在于，它强调面部区域一致性。通过这种方式，自动编码器被迫学习一种能够保持面部部位一致性的表征。

算法1： 面部部位掩码处理过程

输入： 原始人脸： F，掩码比例： M_r；

输出： 掩码人脸；

将人脸图像调整为 224*224 的尺寸，并检测人脸中的 68 个关键点标志；

定义 $ROIs = \{R_1, R_2 ... R_{11}\}$。

定义196个面部图像块；

$M = \{M_1, M_2, M_3\}$， M_1： 眼睛图像块 $\in \{(0, 0) \rightarrow 17^{th}_{landmark}\}$，以及 M_2： 鼻子和脸颊图像块 $\in$ {end of 眼睛图像块 $\rightarrow 16^{th}_{landmark}$}，以及 M_3： 嘴唇图像块 $\in$ {end 鼻子和脸颊图像块的 $\rightarrow$ 最后的 F}；

对于 $i = 1$ 到 3 **执行**

  随机选择 M_i

  for $block \in selected\ M_i$ **do**

    如果 $ROI\ landmarks \in block$ **则**

      为区块签名。

    end

    图像块数量 = 有符号块 $\times M_r$;

    随机遮蔽图像块。

  end

结束

3 方法

本节介绍用于深度伪造视频检测的 Mover 的详细信息。为更好地提取未知深度伪造模式的鲁棒特征，本方法利用面部掩码自编码器进行特征学习。此外，所提出的方法利用双网络，通过面部掩码自编码器的训练模型检测真实视频与伪造视频之间的差异。具体而言，如图 2 所示，所提出的方法由两个阶段组成：(1) 用于一致性学习阶段的 Mover；(2) 用于差异学习阶段的双网络。第一阶段训练模型学习真实人脸的面部部位一致性。第二阶段由两个分支组成：微调网络和映射网络。

3.1 Mover的一致性学习

在此阶段，我们对 MAE[22]及其掩码策略进行定制和修改，以对真实人脸进行自监督学习，使模型能够学习通用的面部部位一致性特征，同时避免过拟合特定的伪造模式。

针对学习一致人脸表征定制的掩码策略。我们对 MAE [22]原有的掩码策略进行定制，并设计了一种面部部位掩码策略，以确保模型能够学习所有面部部位的一致性。算法1展示了相应的伪代码，其输入为视频中的人脸，输出为掩码人脸。原始 MAE [22] 会对输入图像中的随机图像块进行掩码，但这种策略并不适用于我们的 Mover，原因如下。

首先，原始 MAE 掩码策略采用随机掩码，未考虑深度伪造领域的知识，因此无法直接用于深度伪造检测。这是因为随机掩码像素会使模型只关注局部面部一致性，而忽略不同面部部位之间的全局一致性。

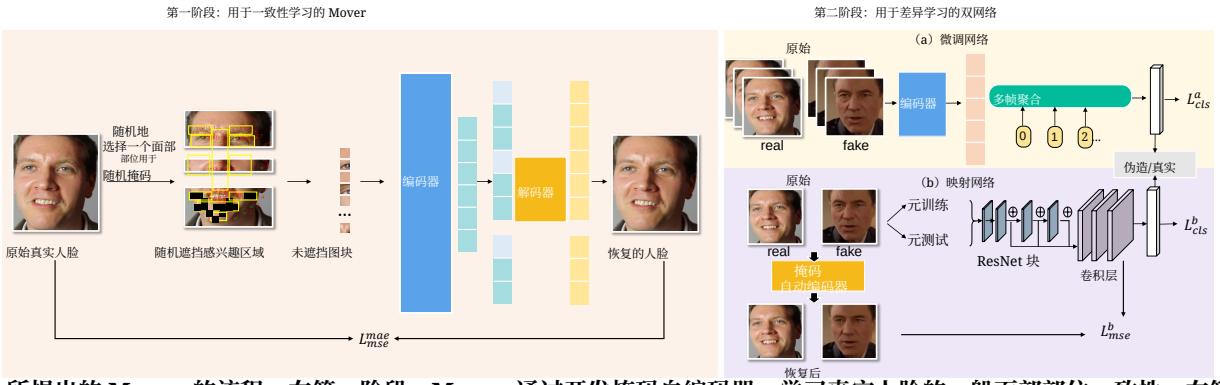

图 2：所提出的 Mover 的流程。在第一阶段，Mover 通过开发掩码自编码器，学习真实人脸的一般面部部位一致性。在第二阶段，Mover 执行 *(a)* 微调网络 和 *(b)*映射网络，利用预训练的 Mover 模型更好地捕捉真实视频与伪造视频之间的差异。

这种面部区域全局一致性的缺失会妨碍模型学习准确的面部部位一致性特征，导致模型难以根据重建人脸检测真实视频与伪造视频之间的差异（图 5）。因此，我们提出的掩码策略将人脸明确划分为三个面部部位，即眼睛、脸颊和鼻子，以及嘴唇，使模型能够关注所有面部部位之间的局部和全局一致性。我们注意到，感兴趣区域提取部分受到面部动作编码系统（FACS）的启发 [12], , 该系统将 FACS 的动作单元视为面部表情的基本组成元素。根据心理学和面部表情研究 [12, 32, 46, 56, 57], , 真实人脸在这些元素上具有内在一致性。因此，与伪造人脸相比，对真实人脸的感兴趣区域进行掩码处理能够实现更一致的区域重建。我们参考眉毛、下眼睑、鼻根、脸颊、嘴角、下巴侧面和下巴的动作单元，根据面部关键点计算边界框坐标。具体而言，R_1 和 R_2 是眉毛，R_3 和 R_4 是下眼睑，R_5 是鼻根，R_6 和R_7 是脸颊，R_8 和 R_9 是嘴角，R_{10} 是下巴侧面，R_{11} 是下巴。

其次，原始 MAE 掩码策略采用较高的掩码比例，因此在不产生任何伪影或失真的情况下恢复原始外观会极具挑战性。如果重建过程中出现伪影，真实人脸中也会包含这些伪影，而伪造人脸中则会同时存在重建伪影和篡改伪影。由于两者都含有伪影，因此难以区分真实视频和伪造视频。因此，我们提出一种采用相对较低掩码比例的掩码策略，使模型能够更准确地重建原始人脸。关于掩码策略的更多讨论，请参见第 4.6 节。

网络架构。我们的掩码自编码器基于非对称编码器-解码器架构 [22]。编码器是一个视觉 Transformer [11]，但仅应用于未遮挡图块。解码器由一系列 Transformer 块组成，应用于掩码标记和未遮挡图块的编码特征。

恢复掩码人脸。人脸的掩码图像块在编码器处理过程中被丢弃，只留下未掩码区域。这样，解码器便可根据未掩码区域预测缺失的面部部位。掩码图像块的重建质量

通过均方误差损失函数 L_{mse}^{mae} 计算，可表示为：

$$L_{mse}^{mae} = \frac{1}{n}\sum_{j=1}^{n}(\widetilde{y^j} - y^j)^2, \tag{1}$$

~其中，y^j是真实像素，y^j是预测像素，n表示面部像素的数量。如果模型学会面部部位之间的一致性，重建图像块与输入图像块之间的损失就应当降低。我们的面部部位掩码策略会随机选择各个部位，从而促使模型学习不针对任何特定面部部位的表征。此外，由于这一阶段仅使用真实视频，不使用任何深度伪造视频，因此可以防止模型对任何特定篡改模式过拟合。第一阶段训练完成后，即可获得预训练的 Mover 模型。

3.2 用于差异学习的双网络

在这一阶段，我们利用第一阶段训练充分的 Mover 模型，使模型能够更好地捕捉真实视频与伪造视频之间的差异，并采用两个分支：（a）微调网络和（b）映射网络。通过采用第一阶段训练充分的模型，将这两个分支结合起来，分别进行训练以检测深度伪造视频。

（a）微调网络。微调网络使用真实视频帧和深度伪造视频帧，并通过交叉熵损失 L_{cls}^{a} 提取真实视频与伪造视频之间的差异。对于 N样本，L_{cls}^{a} 可计算如下：

$$L_{cls}^{a} = \frac{1}{N}\sum_{m} -[g_m^a log(p_m^a) + (1-g_m^a) log(1-p_m^a)], \tag{2}$$

其中，g_m^a 是真实标签，p_m^a 是预测结果为真实视频标签的概率。

我们首先从真实视频和深度伪造视频中提取帧，再从帧中裁剪出人脸。随后，将包含完整图像块集合的原始真实人脸和原始伪造人脸输入第一阶段训练充分的编码器。为汇总多帧信息，我们使用编码器的最后一层，对 10 帧的输出取平均。

由于第一阶段学习真实视频的面部区域一致性，Mover 训练充分的编码器能够提取一致性

特征。对于伪造视频，由于其一致性遭到破坏，编码器提取的特征可能与真实视频的特征不同。微调网络会放大伪造视频与真实视频之间的差异，从而提升检测性能。

(b) 映射网络。映射网络包含以下组件。

数据划分策略。 为避免对特定深度伪造模式过拟合，我们采用元学习 [26] ，并将训练数据随机划分为元训练集和元测试集，其中元训练集和元测试集中的伪造人脸具有不同的操纵模式。由于该分支需要的是不同类型的人脸，而非大量人脸，因此我们使用单帧来训练该分支，以减少内存消耗。

网络架构。 我们采用 ResNet-18 [23] 的第一个卷积层。使用 ResNet-18 的前三个残差块，并将这些残差块的输出进行拼接。拼接后的输出被输入三个卷积层以进行人脸映射。映射后的人脸维度为$56 * 56 * 3$。我们最小化映射后的人脸与输入对应的重建人脸之间的均方误差损失 L_{mse}^{b}。最后，使用一个全连接（FC）层进行分类。我们还最小化分支输出与视频标签之间的二元交叉熵 L_{cls}^{b}。

元训练阶段。 每个训练轮次中，都会构建一个样本批次，其中伪造视频和真实视频的数量相同，以构成二元检测任务。元训练阶段通过采样多个检测任务进行训练，并通过从元测试中采样多个相似的检测任务进行验证。之后，可以更新元训练阶段的参数。为选择最佳梯度步长，我们设置一个参考集，记为 T_i^{ref}。我们利用每个梯度步长计算 T_i^{ref} 的准确率，并选取准确率最高的参数作为最终更新后的参数。最终，元训练阶段利用更新后的参数计算元训练损失。

由于第一阶段学习了真实视频中所有面部部位的一致性，当人脸的一致性遭到破坏时，模型应输出不一致的表征。换言之，由第一阶段训练得到的掩码自编码器重建出的伪造人脸应呈现出不一致性。优化 L_{mse}^{b} 可以使真实视频的映射后的人脸受到重建真实人脸一致性的约束，同时使伪造视频的映射后的人脸受到重建伪造人脸不一致性的约束。最终，模型能够观察到真实人脸的面部部位保持一致，而伪造人脸的这种一致性则遭到破坏。

元测试阶段。 元测试阶段的目标是训练出一个在元训练域上表现良好、并且能够快速泛化到元测试域中未见过的域的分类器，从而提升跨域检测性能。具体而言，我们在元测试域中采样一个批次，并在元训练域中采样另一个批次，将二者拼接成一个随机数组。随后，我们使用该随机数组和更新后的参数计算 L_{cls}^{b} 和 L_{mse}^{b}的元测试损失。

检测损失。 映射网络的最终损失函数为：

$$L^{b} = (L_{cls}^{b} + L_{mse}^{b})_{Meta-train} + (L_{cls}^{b} + L_{mse}^{b})_{Meta-test}. \quad (3)$$

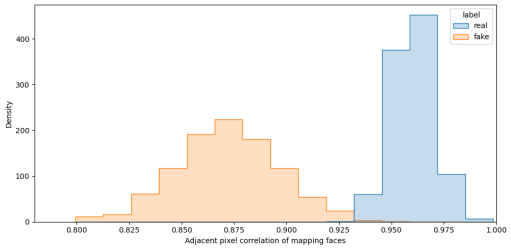

图3：映射后的人脸在像素级相关性上的差异。

它结合了 L_{cls}^{b} 和 L_{mse}^{b} 的元测试损失以及 L_{cls}^{b} 和 L_{mse}^{b} 的元训练损失，以实现联合优化。

4 实验

4.1 实验设置

数据集。 我们采用四个公开的深度伪造视频数据集，即 FaceForensics++ [45]、Celeb-DF [36]、WildDF [68]和DFDCP [9]，来评估所提出的方法和现有方法。FaceForensics++ 由四种篡改算法构成：DeepFakes [8]、Face2Face [54]、FaceSwap [13]和NeuralTextures [53]。此外，基于这四种算法合成了4000个视频。这些视频被广泛用于各种深度伪造检测场景。Celeb-DF包含5639个视频，这些视频由改进的DeepFakes算法 [36]生成。Celeb-DF减轻了某些早期数据集中的篡改痕迹。WildDF由从现实世界中收集的707个深度伪造视频组成。现实世界视频包含多样化的场景、面部表情和伪造类型，这使得数据集更具挑战性。DFDCP是Facebook发布的大规模深度伪造检测数据集。

实现细节。 在第一阶段，掩码比例、批量大小、图像块大小和输入大小分别设置为 0.5、8、16 和 224。使用初始学习率为 $1.5{\cdot}10^{-4}$、动量为 0.9、权重衰减为 0.05 的 AdamW [38] 优化器训练模型。在第二阶段，微调网络使用初始学习率为 $1 \cdot 10^{-3}$ 的 AdamW 优化器来检测视频。使用 SGD优化器优化映射网络，其初始学习率为 0.1，动量为 0.9，权重衰减为 $5 \cdot 10^{-4}$。我们使用 FFmpege [27] 从视频中提取帧。使用 dlib [48] 检测 68 个面部特征点。在第一阶段的训练中，我们根据算法1随机遮挡面部部位。在第二阶段，我们使用 dlib 提取人脸。这些人脸在未进行图像块掩码的情况下输入模型。

4.2 不一致性的初步分析

为了更好地理解 Mover 如何有效检测伪造人脸中的不一致性，我们利用其训练好的映射网络（第 3.2(b) 节）开展了一项简单实验。回顾一下，映射网络旨在将恢复的人脸映射为能够突出恢复的真实人脸与恢复的伪造人脸之间差异的人脸。因此，通过检查真实样本和伪造样本的映射后的人脸，我们可以量化其中的不一致性。我们通过实验发现，每个像素与其相邻像素之间简单的皮尔逊相关系数能够体现这些不一致性。

表1：跨数据集泛化能力。在 Celeb-DF、WildDF 和 DFDCP 数据集上，对 Mover 与基线方法进行跨数据集评估（等错误率和曲线下面积（%）），训练数据为 FaceForensics++。

方法	Celeb-DF		WildDF		DFDCP		Avg	
	曲线下面积 ↑	等错误率↓	曲线下面积 ↑	等错误率↓	曲线下面积 ↑	等错误率↓	曲线下面积 ↑	等错误率↓
Xception [45]	65.3	38.8	66.2	40.1	69.8	35.4	67.1	38.1
EN-b4 [52]	68.5	35.6	61.0	45.3	70.1	34.5	66.5	38.5
Face X-ray [31]	74.2	–	–	–	70.0	–	–	–
MLDG [28]	74.6	30.8	64.1	43.3	71.9	34.4	70.2	36.2
F^3-Net [44]	71.2	34.0	67.7	40.2	72.9	33.4	70.6	35.9
MultiAtt [64]	76.7	32.8	70.2	36.5	67.3	38.3	71.4	35.9
GFF [39]	75.3	32.5	73.0	34.9	71.8	31.8	73.4	33.1
LTW [50]	77.1	29.3	67.1	39.2	74.6	33.8	72.9	34.1
局部关系 [5]	78.3	29.7	68.8	37.5	76.5	32.4	74.5	33.2
DCL [51]	82.3	26.5	71.1	36.2	76.7	32.0	76.7	31.6
Li等人 [30]	84.8	21.9	73.8	31.8	78.5	29.1	79.0	27.6
EB4+SBI [49]	**90.1**	18.2	71.2	34.2	**85.1**	**23.2**	82.1	25.2
R34+SBI [49]	86.1	20.8	67.4	36.7	81.2	26.1	78.2	27.9
Mover	87.1							

图 3 展示了从 1000 个随机抽取的真实视频和伪造视频中得出的统计数据，这些视频来自包含全脸和部分脸部交换的数据集，例如 FaceForensics++、Celec-DF、WildDF 和 DFDCP。结果表明，映射后真实人脸的相邻像素相关系数始终高于 0。91，说明真实人脸在整个重建过程中都保持了面部一致性。这归功于 Mover 的第一阶段，该阶段能够准确重建真实人脸；而第二阶段的 L^b_{mse} 约束有助于保持面部一致性，从而产生较高的相关系数。相比之下，由于重建人脸存在不一致性，映射后伪造人脸的相邻像素相关系数明显较低。因此，L^b_{mse} 约束促使映射后的人脸与伪造人脸中不一致的特征相吻合，导致相关系数较低。全脸和部分脸部交换中都存在这种不一致性。因此，基于未遮罩区域恢复人脸时，任何部分的不一致性都可能导致整体重建结果不一致。我们还评估了多个数据集上全脸交换和部分脸部交换场景的性能，并将在下一节进行讨论。

4.3 对未知域的泛化能力

我们设计了一种检测器，可以自动学习所有面部部位之间的一致性，从而促使模型学习鲁棒特征，以检测未知域中的深度伪造。为验证这一点，我们模拟了以下未知域深度伪造检测场景。

跨数据集泛化能力。为了评估 Mover 的泛化能力，我们开展了跨数据集实验。由于不同数据集的视频分布相对分散，跨数据集实验可以模拟未知领域检测。我们遵循 Li等人 [30] 的工作，在 FaceForensics++ 上使用全部 4 种类型的视频训练模型，然后在其他数据集上进行测试，即 Celeb-DF、WildDF 和 DFDCP。

表1中的部分结果引自Li等人 [30], , 我们遵循Li等人 [30]的做法，使用受试者工作特征曲线下面积（曲线下面积）和等错误率（等错误率）来评估性能。EB4+SBI [49]在其论文中取得了最先进的跨数据集性能，我们重新实现了 EB4+SBI，并在表1中报告了结果。SBI [49] 的检测性能基于预训练的 EcientNet-b4 模型。我们还重新实现了参考文献 [49]中提出的 R34+SBI。在测试 Celeb-DF 和 DFDCP 时，EB4+SBI [49] 的 AUC 高于我们的方法。然而，在测试 WildDF 时，EB4+SBI [49] 的 AUC 和等错误率低于我们的方法。主要原因是 Celeb-DF 聚焦于一种技术，DFDCP 聚焦于两种技术，而 WildDF 中的视频采集自现实世界，并由各种技术生成。Mover 学习适用于各种视频的通用表征的能力，使其略微偏离单一技术，以便适应多种技术。训练域与测试域之间的巨大差异，使提升跨数据集检测性能颇具挑战。尽管如此，所提出的 Mover 在平均AUC和等错误率方面仍取得了些许优势。

表2：跨操纵泛化。在仅使用一种伪造类型进行训练时，Mover 与基线方法在 FaceForensics++ 的各伪造类型上的跨操纵评估（曲线下面积（%））对比。DeepFakes、FaceSwap、Face2Face 和 NeuralTextures 分别以 DF、FS、F2F 和 NT 表示。

方法	训练	F2F	FS	NT	Avg
Freq-SCL [29]	DF	58.9	66.9	63.6	63.1
MultiAtt [64]	DF	66.4	67.3	66.0	66.6
RECCE [2]	DF	70.7	74.3	67.3	70.8
LipForensics [21]	DF	70.0	41.9	76.1	62.7
FInfer [24]	DF	69.4	74.2	66.8	70.1
Freq-SCL [29]	DF	58.9	66.9	63.6	63.1
Mover	DF	**86.9**	**77.6**	**69.7**	**78.1**
方法	训练	DF	FS	NT	Avg
Freq-SCL [29]	F2F	67.6	55.4	66.7	63.2
MultiAtt [64]	F2F	73.0	65.1	71.9	70.0
RECCE [2]	F2F	76.0	64.5	**72.3**	71.0
LipForensics [21]	F2F	88.4	64.4	70.2	74.3
FInfer [24]	F2F	76.3	64.4	69.8	70.2
Mover	F2F	**89.1**	**72.0**	65.5	**75.5**
方法	训练	DF	F2F	NT	Avg
Freq-SCL [29]	FS	75.9	54.6	49.7	60.1
MultiAtt [64]	FS	82.3	61.7	54.8	66.3
RECCE [2]	FS	82.4	64.4	56.7	67.8
LipForensics [21]	FS	60.9	62.6	57.9	60.5
FInfer [24]	FS	82.0	62.4	55.6	66.7
Mover	FS	**82.6**	**82.9**	**58.3**	**74.6**
方法	训练	DF	F2F	FS	Avg
Freq-SCL [29]	NT	79.1	74.2	54.0	69.1
MultiAtt [64]	NT	74.6	80.6	60.9	72.0
RECCE [2]	NT	78.8	**80.9**	63.7	74.5
LipForensics [21]	NT	**90.0**	51.9	**81.7**	74.5
FInfer [24]	NT	79.6	75.8	64.7	73.4
Mover	NT	86.8	69.4	67.9	**74.7**

跨操纵泛化。 我们还开展了跨操纵实验，以评估对未知操纵模式的泛化能力。FaceForensics++ 中的视频由4种伪造技术生成。遵循 RECCE [2], , 我们选择一种

表3：Mover与基线方法在FaceForen-sics++上的数据集内评估（曲线下面积（%））对比。

方法	DF	FS	F2F	NT	Avg
Freq-SCL [29]	**100**	**100**	99.3	98.0	99.3
MultiAtt [64]	99.6	**100**	99.3	98.3	99.3
RECCE [2]	99.7	99.9	99.2	98.4	99.3
LipForensics [21]	**99.8**	**100**	99.3	**99.7**	**99.7**
FInfer [24]	98.4	96.0	93.5	94.9	95.7
Mover	99.6	99.7	**99.6**	96.3	98.8

表4：FaceForensics++ 测试集经过不同处理操作后的曲线下面积（%）得分。

方法	调整大小	增强	亮度	对比度	
R34+SBI [49]	93.5	87.5	83.2	84.5	
EB4+SBI [49]	97.1	90.8	87.6	87.4	
Mover	**97.8**	**91.0**	**89.9**	**90.3**	

表5：消融研究——不同掩码比例（55%、65%、75%、85%、95%）在 Celeb-DF、WildDF 和 DFDCP 数据集上的跨数据集检测性能（曲线下面积（%））。

掩码比例	Celeb-DF	WildDF	DFDCP
55%	85.4	80.3	76.6
65%	86.5	81.6	77.3
75%	**87.1**	**82.0**	**78.9**
85%	86.1	81.7	77.6
95%	85.8	80.1	76.9

用于训练的一种视频类型，以及用于测试的其余三种视频类型。

表2中的结果表明，Mover在许多场景下的平均表现优于以往方法。具体而言，当模型在DeepFakes上训练并在其他三种视频上测试时，Mover的性能平均提升了7.3%。当模型在FaceSwap上训练并在其他三种视频上测试时，Mover的性能平均提升了6.8%。然而，需要注意的是，我们重新实现了LipForensics [21]，并发现，当在NeuralTextures上训练、在Deep-Fakes和FaceSwap上测试时，LipForensics [21] 的表现优于我们的方法；当在NeuralTextures上训练、在Face2Face上测试，以及在Face2Face上训练、在NeuralTextures上测试时，RECCE [2] 的表现优于Mover。尽管如此，值得注意的是，所提出的Mover在平均性能上仍具有可比性。

4.4 数据集内检测性能

为了全面评估所提出的 Mover，我们在数据集内检测场景下将 Mover 与最先进方法进行比较。具体而言，我们在 FaceForensics++ (C23) 的 4 个子集上进行实验。数据集内实验的训练数据和测试数据来自 FaceForensics++ 的同一子集。表3显示，大多数方法在数据集内检测中表现良好。Mover 在 Face2Face 上取得了最高的数据集内检测分数，但与 LipForensics[21]，相比，平均准确率略低 0.9%；在所评估的方法中，LipForensics[21]，的平均表现排名最高。在跨操纵检测性能方面，如表2所示，Mover 在多个指标上优于 LipForensics[21] 。

表6：消融研究——不同掩码策略在 Celeb-DF、WildDF 和 DFDCP 数据集上的跨数据集检测性能（曲线下面积（%））。原始 MAE [22], 脸颊和鼻子策略，以及不划分感兴趣区域的策略，分别表示为 MAE策略 [22],C与N 和 w/o ROIs。

掩码策略	Celeb-DF	WildDF	DFDCP
MAE策略 [22]	82.1	78.5	73.7
Eye	86.5	80.9	78.1
C与N	85.9	81.0	76.5
Lip	86.4	81.2	76.8
不使用感兴趣区域	84.6	79.9	73.7
提出的策略	**87.1**	**82.0**	**78.9**

4.5 后处理操作的鲁棒性

在真实场景中，帧和视频通常会经过后期处理，以调整媒体内容，从而获得更好的显示效果。图像缩放、图像增强、视频亮度和视频对比度等操作都被广泛使用。我们使用这些操作对帧或视频进行后处理，并在表4中展示性能。后处理会削弱特定的伪造线索，因此 SBI[49] 的表现欠佳。而我们的方法考虑了非特定的面部区域一致性，因此即使经过后处理，也能保持更好的检测性能。

4.6 消融研究

掩码比例的影响。为了评估掩码比例对泛化能力的影响，我们在 Celeb、WildDF 和 DFDCP 数据集上进行了实验，并将它们视为未见数据集。我们在 FaceForensics++ 数据集上使用不同的掩码比例训练模型。

请注意，我们并未将掩码比例定义为掩码区域占整张人脸的比例，而是将其定义为掩码区域占相应感兴趣区域（面部部位）的比例，如算法1所示。我们没有采用掩码比例的原始定义（即掩码区域占整张人脸的比例），原因在于我们关注感兴趣区域，并将人脸划分为三个部分，每次仅随机遮挡整张人脸中的一个部分。因此，在掩码过程中，我们应关注相应被遮挡的感兴趣区域部位，而非整张人脸。

从表5可以看出，Mover在掩码比例为75%时表现良好。当掩码比例为55%和65%时，性能略有下降，这表明较低的掩码比例可能会妨碍鲁棒特征的学习。当掩码比例为85%和95%时，检测性能也会下降。这可能是因为较高的掩码比例会增加人脸重建的难度。如果真实人脸和伪造人脸都无法得到良好重建，它们之间的差异就会缩小。因此，我们在实验中将掩码比例设为75%。

掩码策略的影响。我们修改了 MAE [22] 的掩码策略，以提升其泛化能力。为评估改进后的掩码策略的有效性，我们将所提出的掩码策略与 MAE 的掩码策略 [22] 进行了比较。此外，由于修改后的策略会随机选择部分区域进行掩码，因此评估不同被掩码部分的影响十分重要。为分析感兴趣区域的有效性，我们将所提出的策略与不关注感兴趣区域的掩码策略进行了比较。

表 7：消融研究——未修改 MAE、未使用微调网络、未使用映射网络、未使用元学习以及未进行人脸重建时的跨数据集检测性能（曲线下面积（%））。

	Celeb-DF	WildDF	DFDCP
不修改 MAE	75.3	72.1	70.2
不使用微调网络	79.8	75.5	72.1
不使用映射网络	79.4	76.5	76.3
不使用元学习	86.7	81.1	77.6
不使用人脸重建	84.9	80.1	75.4
Whole Mover	**87.1**	**82.0**	**78.9**

在不改变所提出方法其他部分的情况下，我们使用不同的掩码策略进行了跨数据集实验。

表6中第 1th 行和第 6th 行的结果表明，修改MAE [22] 的掩码策略可以提升检测性能。第 2th、3th 和 4th 行分别代表对眼睛区域、脸颊和鼻子区域以及嘴唇区域进行掩码的方法，其结果显示，与所提出的策略相比，性能有所下降。也就是说，对所有面部部位中的部分区域进行随机掩码，比仅对某一部位进行掩码更有利于提取鲁棒特征。此外，第 5^{th}行和第 6th 行的结果表明，所提出的聚焦于感兴趣区域的掩码策略优于不使用感兴趣区域的掩码策略。这是因为伪造视频通常缺乏一致性，因此通过对这些感兴趣区域中的图像块进行掩码，模型能够更好地捕捉真实视频与伪造视频之间的差异。因此，算法1所示的掩码策略有助于检测深度伪造视频。

MAE的影响。我们通过修改掩码策略并在第二阶段加入映射网络，采用MAE [22] 进行深度伪造检测。为评估这些修改的有效性，我们将修改后的Mover与原始MAE方法在深度伪造检测上的性能进行了比较。结果如表 7第 1th 行所示。原始MAE的检测性能低于Mover，证明了Mover中这些修改的有效性。

微调网络和映射网络的影响。在第二阶段，我们结合微调网络和映射网络的两个分支进行深度伪造检测。为了验证每个分支的性能，我们比较了单个分支与两个分支组合后的性能。具体而言，我们从 Mover 中移除微调网络来检测视频，结果如图所示。

2th 表 7 的行。我们从 Mover 中移除映射网络以检测视频，并在 3th 行展示结果。可以看出，移除微调网络或映射网络都会降低检测性能，因为每个分支在深度伪造检测中都发挥着至关重要的作用。结合两个分支，通过扩大真实视频与伪造视频之间的差异，提升了性能。

元学习的影响。 映射网络采用元学习模块来提升领域泛化能力。我们还设计了实验来评估元学习模块的有效性。我们移除元学习模块进行实验，结果如表 7 的第 4th 行所示。 4th 行和 5th 行的结果表明，不采用元学习的方法效果不如所提出的 Mover

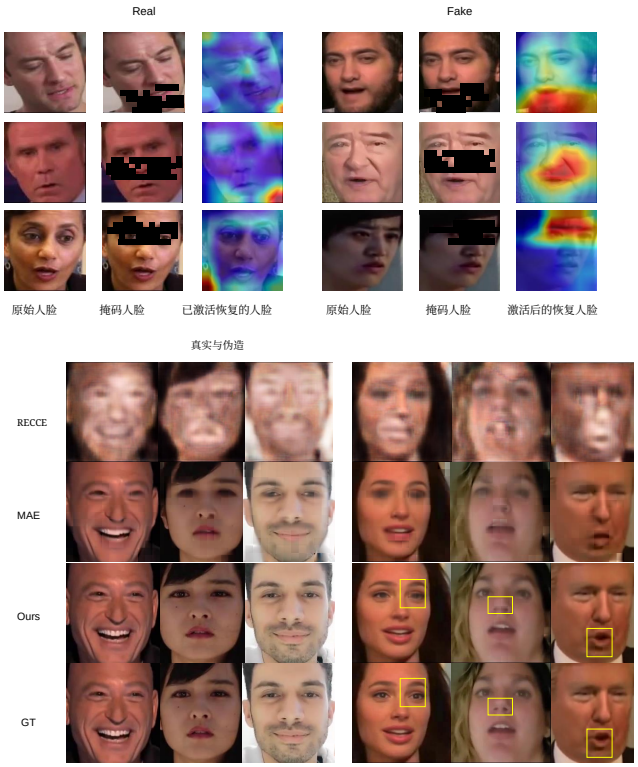

图5：RECCE[2], MAE [22] 和我们的方法恢复的人脸。RECCE[2] 将真实和伪造人脸恢复为不同的分布。MAE [22] 恢复人脸以提升表征质量。我们采用 MAE 恢复人脸，促使模型关注面部区域一致性。

结合元学习的方法。元学习方法在训练阶段模拟跨域检测，从而提升跨数据集检测性能。

人脸重建的影响。映射网络输入重建人脸，以映射人脸并对深度伪造视频进行分类。我们移除重建人脸以实现映射网络，结果报告于表 7 的 5th 行。结果表明，不使用人脸重建的方法效果比采用人脸重建的所提出的 Mover 更差。人脸重建可以放大真实与伪造之间的差异，从而提升跨数据集检测性能。

可视化激活情况。 为了深入了解我们提出的 Mover 提取的特征，我们使用 pytorch_grad_cam 封装的类激活映射函数来展示激活情况。此过程包括遮蔽并重建人脸，然后使用训练好的微调网络可视化重建人脸。如图 4 所示，真实人脸和伪造人脸的激活区域有所不同。对于伪造人脸，由于缺乏一致性，重建后的人脸会出现一些扰动。可视化网络特征后，重建伪造人脸的遮蔽区域会被激活，从而暴露伪造人脸的不一致性。相比之下，真实人脸的一致性会引导重建，使网络难以提取异常的不一致性。因此，激活区域出现在背景区域，而不是遮蔽区域。

区域。也就是说，所提出的 Mover 可以通过提取面部部位一致性来检测视频。

展示恢复的人脸。 我们还在图 5 中展示了重建人脸的可视化结果。RECCE[2] 通过计算人脸重建前后的重建差异分数来检测深度伪造，并取得了良好的性能，但在缺乏面部语义引导的情况下重建整张人脸，可能导致重建结果不清晰。原始 MAE 以较高的掩码比例随机遮挡图像，导致真实人脸和伪造人脸中都出现重建伪影。相比之下，我们提出的 Mover 方法会随机遮挡不同的面部部位，并着重捕捉不一致区域的差异，如图 5 黄色框所示。

5 结论

本文专注于 DeepFakes 检测，尤其是识别新型伪造视频。与依赖特定面部部位不同，我们的两阶段模型平等地关注所有面部部位，从而能够学习到非特定的面部一致性，以及仅存在于真实人脸中的鲁棒表征。第一阶段中，模型通过面部感兴趣区域的部分遮罩重建整幅图像，从而学习真实视频中的面部部位一致性。第二阶段中，模型经过训练以最大化真实视频与伪造视频之间的差异。大量实验表明，Mover 在基准数据集上具有鲁棒性和良好的泛化能力。

参考文献

[1] Darius Afchar、Vincent Nozick、Junichi Yamagishi 和 Isao Echizen。Mesonet：一种紧凑型人脸视频伪造检测网络。发表于 *WIFS*，第 1–7 页，1，2018 年。IEEE。[2] Junyi Cao、Chao Ma、Taiping Yao、Shen Chen、Shouhong Ding 和 Xiaokang Yang。用于人脸伪造检测的端到端重建—分类学习。发表于 *CVPR*，第 4113–4122 页，1，2022年。IEEE。[3] Lucy Chai、David Bau、Ser-Nam Lim 和 Phillip Isola。是什么让伪造图像可被检测？理解具有泛化能力的属性。发表于 *ECCV*，第 103 120 页，1，2020 –。IEEE。[4] Liang Chen、Yong Zhang、Yibing Song、Lingqiao Liu 和 Jue Wang。对抗样本的自监督学习：迈向深度伪造检测的良好泛化能力。发表于 *CVPR*，第 18710–18719 页，1，2022年。IEEE。[5] Shen Chen、Taiping Yao、Yang Chen、Shouhong Ding、Jilin Li 和 Rongrong Ji。用于人脸伪造检测的局部关系学习。发表于 *AAAI*，第 35 卷，第 1081–1088 页，1，2021年。IEEE。[6] Bobby Chesney 和 Danielle Citron。深度伪造：隐私、民主与国家安全面临的迫近挑战。*Calif. L. Rev.*，107：1753，2019 年。[7] Umur Aybars Ciftci、Ilke Demir 和 Lijun Yin。Fakecatcher：利用生物信号检测合成肖像视频。*TPAMI*，1(1)，2020 年。[8] DeepFakes。访问日期：2018 年 10 月 10 日。https://github.com/deepfakes/faceswap，2018 年。[9] Brian Dolhansky、Russ Howes、Ben Pflaum、Nicole Baram 和 Cristian Canton Ferrer。深度伪造检测挑战（DFDC）预览数据集。*arXiv*预印本 *arXiv:1910.08854*，1(1):1，2019 年。[10] Xiaoyi Dong、Jianmin Bao、Dongdong Chen、Ting Zhang、Weiming Zhang、Neng-hai Yu、Dong Chen、Fang Wen 和 Baining Guo。利用身份一致性 Transformer 保护名人免遭深度伪造。发表于 *CVPR*，第 9468 9478 页，1，2022 –。IEEE。[11] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。图像胜过千言万语：面向大规模图像识别的 Transformer。发表于 *ICLR*，第 1 页，1，2021 年。IEEE。[12] Paul Ekman 和 Wallace V. Friesen。面部动作编码系统：一种测量面部运动的技术。发表于 *1*，第 1 卷，第 1 页，1，1978 年。IEEE。[13] FaceSwap。访问日期：2018 年 10 月 29 日。https://github.com/MarekKowalski/ FaceSwap/，2018 年。[14] Christoph Feichtenhofer、Yanghao Li、Kaiming He 等。作为时空学习器的掩码自编码器。*NeurIPS*，35：35946–35958，2022年。[15] Peng Gao、Teli Ma、Hongsheng Li、Jifeng Dai 和 Yu Qiao。ConvMAE：掩码卷积与掩码自编码器的结合。发表于 *NeurIPS*，第 1 页，1，2022年。IEEE。

[16] Zhihao Gu、Yang Chen、Taiping Yao、Shouhong Ding、Jilin Li、Feiyue Huang 和 Lizhuang Ma。用于深度伪造视频检测的时空不一致性学习。发表于 ACM MM，第 3473–3481 页，1，2021年。IEEE。[17] Zhihao Gu、Yang Chen、Taiping Yao、Shouhong Ding、Jilin Li 和 Lizhuang Ma。深入局部：用于深度伪造视频检测的动态不一致性学习。发表于 AAAI，第 36 卷，第 744–752 页，1，2022年。IEEE。[18] Zhihao Gu 、 Taiping Yao 、 Yang Chen 、 Shouhong Ding 和 Lizhuang Ma。用于深度伪造视频检测的层次化对比不一致性学习。发表于 ECCV，第 596–613 页，1，2022年。IEEE。[19] Jiazhi Guan、Hang Zhou、Zhibin Hong、Errui Ding、Jingdong Wang、Chengbin Quan 和 Youjian Zhao。深入研究用于深度伪造检测的序列补丁。NeurIPS，1:1，2022年。[20] Alexandros Haliassos、Rodrigo Mira、Stavros Petridis 和 Maja Pantic。通过自监督利用真实说话人脸进行鲁棒伪造检测。发表于 CVPR，第 14950–14962 页，1，2022年。IEEE。[21] Alexandros Haliassos、Konstantinos Vougioukas、Stavros Petridis 和 Maja Pantic。嘴唇不会说谎：一种可泛化且鲁棒的人脸伪造检测方法。发表于 CVPR，第 5039–5049 页，1，2021年。IEEE。[22] Kaiming He、Xinlei Chen、Saining Xie、Yanghao Li、Piotr Dollár 和 Ross Girshick。掩码自编码器是可扩展的视觉学习器。发表于 CVPR，第 16000 16009 页，1，2022年 –。IEEE。[23] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。发表于 CVPR，第 770–778 页，1，2016。IEEE。[24] Juan Hu、Xin Liao、Jinwen Liang、Wenbo Zhou 和 Zheng Qin。FInfer：基于帧推理的高视觉质量视频深度伪造检测。发表于 AAAI，第 36 卷，第 951–959 页，1，2022年。IEEE。[25] Juan Hu、Xin Liao、Wei Wang 和 Zheng Qin。使用基于帧时序的双流卷积网络检测社交网络中的压缩深度伪造视频。IEEE TCSVT，32(3):1089–1102，2021。[26] Yunpei Jia、Jie Zhang 和 Shiguang Shan。具有分布对齐的人脸反欺骗双分支元学习网络。TIFS，17:138–151，2021。[27] Xiaohua Lei、Xiuhua Jiang 和 Caihong Wang。基于 FFmpeg 的实时视频流分析系统的设计与实现。发表于 WCSE，第 212–216 页，1，2013。IEEE。[28] Da Li、Yongxin Yang、Yi-Zhe Song 和 Timothy Hospedales。学习泛化：领域泛化的元学习。发表于 AAAI，第 32 卷，1，1，2018 page。IEEE。[29] Jiaming Li、Hongtao Xie、Jiahong Li、Zhongyuan Wang 和 Yongdong Zhang。由单中心损失监督的频率感知判别特征学习，用于人脸伪造检测。发表于 CVPR，第 6458–6467 页，1，2021年。IEEE。[30] Jiaming Li、Hongtao Xie、Lingyun Yu 和 Yongdong Zhang。小波增强弱监督局部特征学习，用于人脸伪造检测。发表于 ACM MM，第 1299–1308 页，1，2022年。IEEE。[31] Lingzhi Li、Jianmin Bao、Ting Zhang、Hao Yang、Dong Chen、Fang Wen 和 Baining Guo。Face x-ray：实现更通用的人脸伪造检测。发表于 CVPR，第 5001–5010 页，1，2020。IEEE。[32] Shan Li 和 Weihong Deng。深度面部表情识别：综述。IEEE TAC，13(3):1195–1215，2020。[33] Xiaodan Li、Yining Lang、Yuefeng Chen、Xiaofeng Mao、Yuan He、Shuhui Wang、Hui Xue 和 Quan Lu。用于深度伪造视频检测的尖锐多示例学习。发表于 ACM MM，第 1864–1872 页，1，2020。IEEE。[34] Yuezun Li、Ming-Ching Chang 和 Siwei Lyu。转瞬之间：通过检测眨眼揭露人工智能生成的伪造视频。发表于 WIFS，第 1–7 页，1，2018。IEEE。[35] Yuezun Li 和 Siwei Lyu。通过检测人脸变形伪影揭露深度伪造视频。发表于 CVPRW，第 46–52 页，1，2019。IEEE。[36] Yuezun Li、Xin Yang、Pu Sun、Honggang Qi 和 Siwei Lyu。Celeb-DF：用于深度伪造取证的大规模挑战性数据集。发表于 CVPR，第 3207 3216 页，1，2020 – 。IEEE。[37] Honggu Liu、Xiaodan Li、Wenbo Zhou、Yuefeng Chen、Yuan He、Hui Xue、Weiming Zhang 和 Nenghai Yu。空间相位浅层学习：重新思考频域中的人脸伪造检测。发表于 CVPR，第 772–781 页，1，2021年。IEEE。[38] Ilya Loshchilov 和 Frank Hutter。解耦权重衰减正则化。arXiv 预印本 arXiv:1711.05101，3:1，2017。[39] Yuchen Luo、Yong Zhang、Junchi Yan 和 Wei Liu。利用高频特征实现人脸伪造检测的泛化。发表于 CVPR，第 16317–16326 页，1，2021年。IEEE。[40] Iacopo Masi、Aditya Killekar、Royston Marian Mascarenhas、Shenoy Pratik Gurudatt 和 Wael AbdAlmageed。用于在视频中分离深度伪造内容的双分支循环网络。发表于 ECCV，第 667–684 页，1，2020。IEEE。[41] Trisha Mittal 、 Uttaran Bhattacharya 、 Rohan Chandra、Aniket Bera 和 Dinesh Manocha。情感不会说谎：一种使用情感线索的视听深度伪造检测方法。发表于 ACM MM，第 2823–2832 页，1，2020。IEEE。[42] Aakash Varma Nadimpalli 和 Ajita Rattani。关于改进深度伪造检测器跨数据集泛化能力的研究。发表于 CVPR，第 91–99 页，1，2022年。IEEE。[43] Huy H Nguyen、Junichi Yamagishi 和 Isao Echizen。Capsule-forensics：使用胶囊网络检测伪造图像和视频。发表于 ICASSP，第 2307 2311 页，1，2019 – 。IEEE。[44] Yuyang Qian、Guojun Yin、Lu Sheng、Zixuan Chen 和 Jing Shao。频域思考：通过挖掘频率感知线索进行人脸伪造检测。发表于 ECCV，第 86–103 页，1，2020。IEEE。

[45] Andreas Rossler、Davide Cozzolino、Luisa Verdoliva、Christian Riess、Justus Thies 和 Matthias Nießner。Faceforensics++：学习检测篡改人脸图像。发表于*ICCV*，第 1–11 页，1，2019 年。IEEE。[46] James A Russell 和 José Miguel Ed Fernández-Dols。面部表情心理学。人文科学之家出版社，1，1997 年。[47] Ekraam Sabir、Jiaxin Cheng、Ayush Jaiswal、Wael AbdAlmageed、Iacopo Masi 和 Prem Natarajan。用于视频人脸篡改检测的循环卷积策略。*Interfaces*（*GUI*），3：80–87，2019 年。[48] S Sharma、Karthikeyan Shanmugasundaram 和 Sathees Kumar Ramasamy。基于 Farec-CNN 并使用 dlib 的高效人脸识别技术。发表于*ICACCCT*，第 192–195 页，1，2016 年。IEEE。[49] Kaede Shiohara 和 Toshihiko Yamasaki。使用自混合图像检测深度伪造。发表于*CVPR*，第 18720–18729 页，1，2022 年。IEEE。[50] Ke Sun、Hong Liu、Qixiang Ye、Yue Gao、Jianzhuang Liu、Ling Shao 和 Rongrong Ji。通过学习权重实现域泛化人脸伪造检测。发表于*AAAI*，第 35 卷，第 2638–2646 页，1，2021 年。IEEE。[51] Ke Sun、Taiping Yao、Shen Chen、Shouhong Ding、Jilin Li 和 Rongrong Ji。用于通用人脸伪造检测的双重对比学习。发表于*AAAI*，第 36 卷，第 2316–2324 页，1，2022 年。IEEE。[52] Mingxing Tan 和 Quoc Le。EfficientNet：重新思考卷积神经网络的模型缩放。发表于*ICML*，第 6105–6114 页，1，2019 年。IEEE。[53] Justus Thies、Michael Zollhöfer 和 Matthias Nießner。延迟神经渲染：使用神经纹理进行图像合成。*TOG*，38(4)：1–12，2019 年。[54] Justus Thies、Michael Zollhofer、Marc Stamminger、Christian Theobalt 和 Matthias Nießner。Face2Face：RGB视频的实时人脸捕捉与重演。发表于*CVPR*，第 2387–2395 页，1，2018 年。IEEE。[55] Zhan Tong、Yibing Song、Jue Wang 和 Limin Wang。VideoMAE：掩码自编码器是适用于自监督视频预训练的数据高效学习器。发表于*NeurIPS*，第 1 页，1，2022 年。IEEE。[56] Su-Jing Wang、Wen-Jing Yan、Xiaobai Li、Guoying Zhao 和 Xiaolan Fu。在张量独立色彩空间上使用动态纹理进行微表情识别。发表于*ICPR*，第 4678–4683 页，1，2014 年。IEEE。[57] Su-Jing Wang、Wen-Jing Yan、Xiaobai Li、Guoying Zhao、Chun-Guang Zhou、Xiaolan Fu、Minghao Yang 和 Jianhua Tao。使用 微表情识别。色彩空间。*IEEE Transactions on Image Processing*，24(12):6034–6047，2015年。[58] Sheng-Yu Wang、Oliver Wang、Richard Zhang、Andrew Owens 和 Alexei A Efros。CNN 生成的图像出乎意料地容易被识别……至少目前如此。发表于*CVPR*，第 8695–8704 页，1，2020年。IEEE。[59] Zhi Wang、Yiwen Guo 和 Wangmeng Zuo。通过对抗博弈进行深度伪造取证。*TIP*，31:3541–3552，2022年。[60] Haiwei Wu、Jiantao Zhou、Jinyu Tian、Jun Liu 和 Yu Qiao。针对在线社交网络传输的稳健图像伪造检测。*IEEE TIFS*，17 456，2022年 443：–。[61] Xin Yang、Yuezun Li 和 Siwei Lyu。利用不一致的头部姿态揭露深度伪造。在*ICASSP*，第 8261–8265 页，1，2019年。IEEE。[62] Ziming Yang、Jian Liang、Yuting Xu、Xiao-Yu Zhang 和 Ran He。用于深度伪造检测的掩码关系学习。*TIFS*，18:1696–1708，2023年。[63] DaiChi Zhang、Fanzhao Lin、Yingying Hua、Pengju Wang、Dan Zeng 和 Shiming Ge。利用时空 Dropout Transformer 进行深度伪造视频检测。发表于*ACM MM*，第 5833–5841 页，1，2022年。IEEE。[64] Hanqing Zhao、Wenbo Zhou、Dongdong Chen、Tianyi Wei、Weiming Zhang 和 Nenghai Yu。多注意力深度伪造检测。发表于*CVPR*，第 2185–2194 页，1，2021年。IEEE。[65] Tianchen Zhao、Xiang Xu、Mingze Xu、Hui Ding、Yuanjun Xiong 和 Wei Xia。学习深度伪造检测的自一致性。发表于*ICCV*，第 15023–15033 页，1，2021年。IEEE。[66] Yinglin Zheng、Jianmin Bao、Dong Chen、Ming Zeng 和 Fang Wen。探索时间连贯性，以实现更通用的视频人脸伪造检测。发表于*ICCV*，第 15044–15054 页，1，2021年。IEEE。[67] Wanyi Zhuang、Qi Chu、Zhentao Tan、Qiankun Liu、Haojie Yuan、Changtao Miao、Zixiang Luo 和 Nenghai Yu。UIA-ViT：一种基于视觉 Transformer、具备无监督不一致性感知能力的人脸伪造检测方法。发表于*ECCV*，第 391–407 页，1，2022年。IEEE。[68] Bojia Zi、Minghao Chang、Jingjing Chen、Xingjun Ma 和 Yu-Gang Jiang。Wild-Deepfake：用于深度伪造检测的真实世界挑战性数据集。发表于*ACM MM*，第 2382–2390 页，1，2020年。IEEE。