基于人脸和背景的差异的深度造假检测

Yuval Nirkin, Lior Wolf, Yosi Keller, and Tal Hassner

摘要--我们提出了一种检测单一图像中人脸互换和其他身份操纵的方法。脸部交换方法，如DeepFake，操纵脸部区域，旨在将脸部调整为其背景的外观，而保持背景不变。我们表明，这种操作方式在两个区域之间产生差异（例如，图1）。这些差异提供了可利用的操纵迹象。我们的方法涉及两个网络：(i)一个人脸识别网络，考虑到由严格的语义分割限定的人脸区域；(ii)一个上下文识别网络，考虑到人脸上下文（如头发、耳朵、脖子）。我们描述了一种方法，它使用来自我们两个网络的识别信号来检测这种差异，提供了一个互补的检测信号，改善了通常用于检测假图像的传统真假分类器。我们的方法在FaceForensics++和Celeb-DF-v2的人脸操纵检测基准上取得了最先进的结果，甚至可以推广到检测由未见过的方法产生的假象。

Index Terms-Image forensics, deep learning, deep fake, face swapping, fake image detection

1 简介

P人们普遍认为热成像技术提供了实际事件的真实证据，特别是包括图像和视频中人类主体的存在和行动。尽管这种看法正在慢慢转变，但当代技术允许对图像进行比许多人意识到的更容易和更方便的操纵。当被操纵的媒体在社交网络上发布并被公众消费时，这种差距就会带来社会威胁。不具备质疑其真实性的条件。

例如，现有的技术使演员更容易说出特定的文字，然后改变她的面部表情和声音以模仿其他人。另外，在犯罪现场捕捉到的一个人的脸可以被操纵并被另一个人取代。这两个例子都被称为 "*换脸*"。第三种情况是通过重演一个人的脸来改变表情或嘴唇的动作（又称*脸部重演*）。然而，我们注意到，第三种情况与前两种情况不同，因为它不涉及身份的改变。

当代大多数检测这类问题的方法操作与这三种情况有类似的关系：通过训练分类器来区分真假图像或视频[8], [9], [10], [11], [12] 。最近，人们提出了一些检测方法，这些方法侧重于活泼性和其他特定的真实性信号，如心跳[13]。[14]和镜面高光[15]。

Face X-ray方法[16]专注于混合步骤，这是处理视频中人脸的方法的一个常见的后处理步骤。该模型检测混合面具的边界，然后将其分为真实或虚假。专注于操纵管道中的一个通用步骤，使该方法更适合于未见过的操纵方法。与Face X-ray类似，我们也关注大多数人脸互换方法所共有的一个特征。虽然Face X-ray关注的是真实和脸部内容之间的接缝，但我们关注的是两者之间身份的差异。

从应用上讲，交换是特别有意义的，因为许多现有的面部操作方法是为这种身份修改的用例设计的。为此，我们做了两个假设：（A1）脸部操作方法只操作脸部的内部部分。(A2)脸部的环境，包括脸部内部以外的头部、颈部和头发区域，为主体提供了一个重要的身份信号。

我们在第3.2节中验证假设A2。我们的研究结果与以前的报告一致，表明仅仅是上下文就确实提供了强有力的身份线索[17], [18]。为了支持假设A1，图2显示了六种不同技术水平的面部操作方法的影响区域。图2a和2b展示了Thies等人[2]、[3]的两种重演方法。这两种方法都操纵了与三维可变形模型（3DMM）[19], [20]相对应的区域，涵盖了一个面部区域，该区域在顶部包含部分额头，在底部包含大部分下巴。图2c和2d显示了FaceForensics++[5]和DFD[1]数据集中的两个深层造假变体样本，它们都影响了面部中段的一个方形区域。图2e是另一种基于3DMM的换脸方法，影响的区域与重演方法相似，但不包括嘴的内部部分（样品

- *尤瓦尔-尼尔金和约西-凯勒在以色列拉马特甘市的巴伊兰大学工程学院，5290002。电子邮件。{yuval.nirkin, yosi.keller}.@gmail.com.*
- *Lior Wolf是特拉维夫大学的，特拉维夫6997801，以色列。电子邮件：liorwolf@gmail.com。*
- *Tal Hassner在Facebook AI，Menlo Park，CA 94025 USA。电子邮件：talhassner@gmail.com。*

2020年8月20日收到稿件；2021年4月14日修订；2021年6月14日接受。出版日期2021年6月29日；当前版本日期2022年9月9日。(通讯作者：Yuval Nirkin。)建议由K. Sunkavalli接受。
数字对象标识符号。10.1109/TPAMI.2021.3093446

图1.*通过比较人脸和他们的上下文来检测互换的人脸*。两个来自DFD[1]的假（互换）面孔的例子。左图： 眼镜的手臂没有从脸部延伸到背景。右图。面部和上下文之间明显的身份不匹配。我们展示了这些和类似的差异如何被用作自动检测被交换的脸的强大信号。

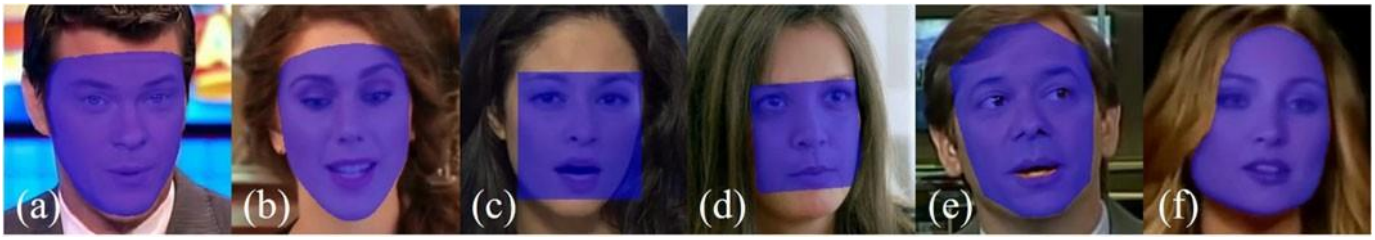

图2.*不同操作方法的受影响区域*。(a) + (b) Face2Face [2] 和NeuralTextures [3]; (c) + (d) Deepfake [4] 变体FaceFor- ensics++ [5] 和DFD [1]; (e) FaceSwap [6]; (f) FSGAN [7] 。在所有情况下，人脸都被操纵，但他们的上下文却没有变化。

从以前的工作[5]中获得）。图2f是FSGAN[7]的输出，它使用面部分割来处理整个面部区域。

我们声称，我们所知道的所有脸部操作方法都不影响整个头部，这并不是巧合。虽然人脸有简单的、容易建模的几何形状，但他们的背景（脖子、耳朵、头发等）是非常不规则的，因此很难持续重建和操纵，特别是考虑到视频中的时间限制。

我们提出了一种新的信号，用于识别基于内部人脸区域--即直接被操纵的区域--与其外部环境的比较，而我们所知道的所有人脸操纵方法都没有对其进行改变。我们通过用两个独立的身份向量来代表这两个区域，即人脸和他们的背景。这两个向量是通过训练两个独立的人脸识别网络得到的：一个是基于人脸区域的识别训练，另一个是基于人脸背景的识别训练。我们比较这两个向量，寻找身份与识别之间的差异。

重要的是，我们*并不*假设事先知道图像中出现的人的身份（来源或主体身份）。相反，给定一个图像，我们比较一个或两个（未知）身份的表征，这些表征是通过我们的两个专门训练的网络从面部及其背景中获得的。

我们利用这两个网络得出的线索与那些寻找特定脸部操作技术所造成的伪影的方法不同。与其他方法相比，我们的线索有三个明显的优势。首先，我们的线索是基于人脸互换方案的固有设计，因此，即使未来的方法产生了照片般逼真的、无人工痕迹的结果，也有望保持。

其次，这种线索可以很好地适用于不同的操作方法，而伪影检测方法则依赖于算法的特定缺陷。最后，由于所提出的线索在很大程度上与人工制品检测方法无关，它是补充性的，因此可以很容易地与这些方法结合起来以提高准确性。

综上所述，我们做出了以下贡献：（1）我们提出了一种新的方法来识别人脸互换方法的结果。(2) 我们的方法是基于一个新的假货检测线索，它比较了两个图像衍生的身份嵌入。(3) 当应用于FaceForensics++[5]、Celeb-DF-v2[21]和DFDC[22]时，所提出的方法被证明优于现有的最先进方案。
(4) 我们在两个额外的人脸交换基准上展示了进一步的结果，这些基准是使用Face- Forensics++的数据和额外的交换技术创建的，并没有包括在Face-Forensics++中。

2 相关工作

人脸互换技术。半自动和全自动的人脸互换方法在近二十年前就被引入了[23], [24]。这些早期的方法是作为保护隐私[24], [25], [26], 娱乐[27], 和娱乐（如[28], [29]）的一种手段而提出的；这与他们今天在错误信息和假新闻中的一些不太吸引人的应用相差甚远。几乎所有前期的深度学习方法都在一定程度上依赖于三维人脸表征，特别是3DMM[19], [20]。这类方法的一些较新的例子是用于表情转换的Face2Face方法[2]、人脸重现[30]、表情操作[31]和人脸互换方法[18]。

公众对人脸操作方法的认识是在基于深度学习的互换技术推出后开始的。

和重现，特别是通过使用生成对抗网络（GAN）。这类技术的几个明显的例子是GANimation[32]、GANnotation[33]和其他[34]、[35]、[36]、[37]。与早期的、基于3D的方法不同，基于GAN的方法能够产生近乎照片般真实的结果，不仅是在静止的照片中，而且在视频中。这些结果的质量，加上公共软件的可用性，导致了现在被合称为*DeepFakes*的东西被用于不良应用，包括色情和假新闻。

最近，FSGAN[7]显示了令人信服的互换结果，而不需要对每个源人或目标人进行专门的训练程序，也就是说，它被训练成可以用任何其他脸来替换任何脸。FaceShifter最先进的互换方法[38]首先使用多尺度注意块将源身份与目标脸的特征合并，然后细化结果，以无监督的方式处理遮挡问题。

2.1 检测被操纵的面孔

多年来，许多人提出了检测图像和视频中的通用、复制移动和拼接操作的方法[39], [40], [41], [42]。然而，脸部受到的关注要少得多，可能是因为直到最近，制作照片般逼真的脸部操作要难得多。

最近的人脸操纵方法所造成的威胁升高，现在正通过增加开发自动假象检测方法的努力来应对。早期检测被操纵的视觉媒体的方法依赖于手工制作的特征[11]。最近，Cozzolino等人[10]描述了这种方法的一个更现代的、基于深度学习的实现，随后，其他基于深度学习的方法，[8]、[9]、[12]、[43]、[44]、[45]、[46]、[47]、[48]以及
作为利用多种线索的方法[42], [49], [50], [51]。
[52], [53], [54].

Sabir等人[48]最近提出了一个循环神经网，它使用时间线索来检测视频中的Deepfake操纵。Stehouwer等人[55]将一种注意力机制应用于不同背骨分类器的中间特征图，以提高被操纵区域的检测精度。Songsri等人[56]表明，使用额外的面部地标可以提高Deepfakes的检测和定位。最后，Nguyen等人[51]提出了一个基于胶囊网络的假货检测架构。他们的工作取得了与以前的方法相当的结果，同时利用的参数明显较少。

2.2 脸部操纵的基准测试

最近有一些努力试图为研究界提供标准的、高质量的假货检测基准。这些努力包括FaceForensics [47], DeepFake- TIMIT [57], Celeb-DF [21], VTD dataset [58], FaceForensics +
+挑战[5]，以及DFD数据集[1]。一些行业研究实验室最近也为这些努力做出了贡献，导致宣布了DeepFake Detection Challenge（DFDC）[22]。

这些基准代表了多种操作技术--不仅仅是换脸。通过使用单一（或少数）的合成方法，偏见可能会被不经意地引入到这些挑战中：对某一特定的人工智能而言，它是独一无二的。

伪造的生成方法，或对特定训练数据的使用。因此，这些数据集包括用各种合成方法生成的媒体。我们的方法旨在不受这种偶然偏见的影响。我们不寻求特定的人工制品，而是考虑一般交换技术所共有的感知效果，并表明我们的方法可以检测出由以前未见过的脸部操作技术产生的假货。

3 识别人脸和他们的背景

我们描述了两个互补的人脸识别网络，用于获得人脸及其文本的身份线索。我们进一步解释我们如何在我们提出的假货检测方法中使用这两个网络。深度神经网络被广泛用于人脸识别，我们重点关注两个非常具体的面部区域的贡献，这是由所需的应用决定的：分割的脸和它的周围环境。

3.1 检测和分割人脸

我们首先应用双镜头面部检测器（DSFD）[59]。然后我们将检测到的边界框的大小增加20%，相对于他们的高度，以暴露更多的脸部周围的环境，因为DSFD被训练为返回紧密的面部边界框。然后，脸部裁剪的大小被调整为299x299像素；这是Xception架构[60]的输入分辨率，我们使用它作为脸部/背景线索（第3.2节）。为了确定作物的哪些部分由人脸网络处理，哪些由背景网络处理，我们使用人脸分割网络将作物分割成前景（人脸）和背景（背景）。分割网络的确切结构和训练细节在补充材料中提供，可以在计算机协会数字图书馆找到：http://doi.ieeecomputersociety.org/10.1109/TPAMI.2021.3093446。给出被裁剪的人脸$\boldsymbol{I}$和其相应的人脸分割掩码$\boldsymbol{S}$，我们生成图像$\boldsymbol{I_f}$和其互补的图像$\boldsymbol{I_c}$，分别代表人脸和其背景。

3.2 识别网络

*识别网络架构。*我们的网络是基于Xception架构[60]，因为它在检测其他DeepFake线索方面取得了成功[5]。我们使用虚构的交叉熵损失来训练网络，尽管其他损失函数也可以被使用。Xception是基于Inception架构的[61]，但Inception模块被深度可分离卷积所取代。就我们所知，它从未被用于人脸识别。

在我们的实现中，Xception网络由一个分层卷积块组成，然后是12个具有剩余连接的深度可分离卷积块，除了最后一个。该网络由两个深度可分离卷积、一个池化操作和一个全连接层结束。

我们训练两个识别网络。$\boldsymbol{E_f}$，它将包含人脸区域像素的299x299大小的图像映射到与数据集人脸相关的伪概率向量，同样，网络$\boldsymbol{E_c}$将检测边界框（背景）的剩余像素映射到相同类别的伪概率向量。

表1
VGGFace2的人脸识别准确率

方法	训练集	验证集
语境	99.90	87.06
脸部	99.89	95.10
整个地区	**99.98**	**96.98**

报告了三个人脸识别Xception网络的结果，每个都应用于人脸的不同部分。正如预期的那样，包含人脸和背景的整个区域是最准确的。然而，即使是单独的上下文，也为识别提供了一个强有力的线索，正如以前其他人所观察到的那样[17], [18]。

我们在标准的、公开的VGGFace2数据集[62]的图像上训练$\boldsymbol{E}_f$和$\boldsymbol{E}_c$。VGGFace2包含9,131名受试者，我们从中过滤了分辨率低于128x128的图像，结果有8,631个身份。因此，这两个网络的输出是在$\mathrm{R}^{8,631}$。

*验证识别能力。*为了验证和比较这些网络的识别精度，我们在VGGFace2[62]测试集和Labeled Faces in the Wild（LFW）[63]基准的测试集上测试它们的性能（在LFW图像上测试之前，没有对网络进行额外的训练或微调）。

不足为奇的是，在处理人脸的内部外观时，网络$\boldsymbol{E}_f$在准确性方面优于$\boldsymbol{E}_c$，尽管两者的准确性都很高。这些结果从表1的VGGFace2和图3的LFW可以看出。我们注意到，$\boldsymbol{E}_c$所表现出的准确性--它在只看到上下文的情况下仍能识别人脸--并不令人惊讶：其他人也报告了类似的结果，表明即使只有上下文可见，人脸也能被识别[17], [18]。

重要的是，图3b显示，通常用于人脸识别的表征--人脸识别网络第五层的激活，对于同一个人来说并不是很匹配，因为这两个网络是独立训练的。因此，在结合这两个网络的反应时，我们使用它们的最终结果：每个主体的伪概率（第4.1节）。

4 使用人脸的假货检测与背景介绍

我们在图4中说明了我们提出的假货检测方法。我们的方法结合了多个Xception网络。识别网络，$\boldsymbol{E}_f$和$\boldsymbol{E}_c$，在第3节中描述，一个二进制Xception网络，$\boldsymbol{E}_s$，被训练来区分真实的和被人脸交换方法操纵的图像，以及另一个*可选*的二进制Xception网络，$\boldsymbol{E}_r$（图4中没有显示），我们训练它来区分真实图像和被人脸重演方法操纵的图像。接下来我们将详细描述这些组件。

4.1 脸部差异部分

我们训练人脸差异网络来预测一张脸和它的环境是否有相同的身份。它使用第3节中描述的两个识别网络$\boldsymbol{E}_f$和$\boldsymbol{E}_c$的结果。我们预先训练这两个网络，并在它们结合后不改变它们的权重，以确保身份线索仍然是主要的线索。在第5.3节中，我们表明在识别网络的权重未被冻结的情况下进行训练，会导致对未见过的方法进行泛化时的准确性降低。我们用两个独立的身份分类器$\boldsymbol{E}_f$和$\boldsymbol{E}_c$分别处理人脸和背景图像$\boldsymbol{I}_f$和$\boldsymbol{I}_c$，以计算出一个差异特征向量$\boldsymbol{v}_d$

$$\boldsymbol{v}_d \,¼\, \boldsymbol{E}\,\delta\boldsymbol{I}_{ff}\,Þ\text{-}\,\boldsymbol{E}_c\,\eth\mathrm{I}_c\,Þ¼\,\boldsymbol{v}_f - \boldsymbol{v}_c\,。 \tag{1}$$

4.2 操纵特定网络

以前的方法是训练分类器来区分真假人脸，而不考虑应用于人脸的特殊操作--交换或重做。这两种操作类型有很大不同。交换操纵的是人脸的身份，而重演操纵的是面部姿势和表情。虽然后者不是我们工作的重点，但它是由

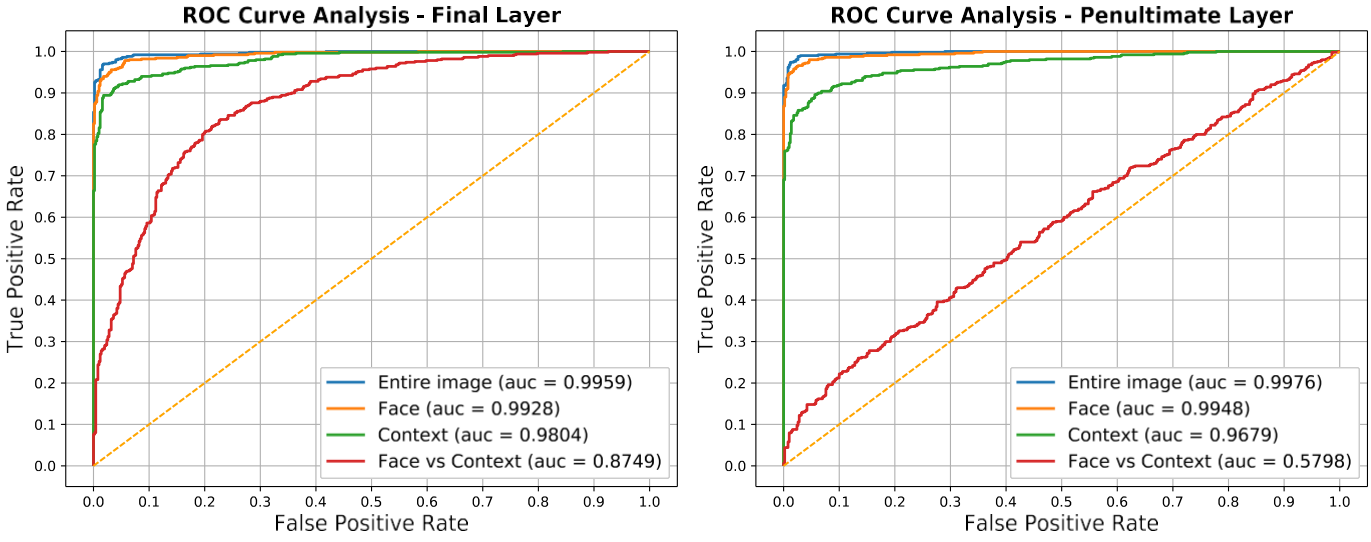

图3.*在不同人脸区域训练的识别网络的LFW验证准确性。*(a) 用Xception架构的最后几层来表示人脸所得到的结果。(b) 用Xception的倒数第二层的激活来表示脸部。在后一种情况下，由于两个网络是独立训练的，所以对于同一个人来说，脸部与背景并不匹配。因此，我们的方法是在比较两者时使用网络的最后几层，代表主体的伪概率（顶部）。

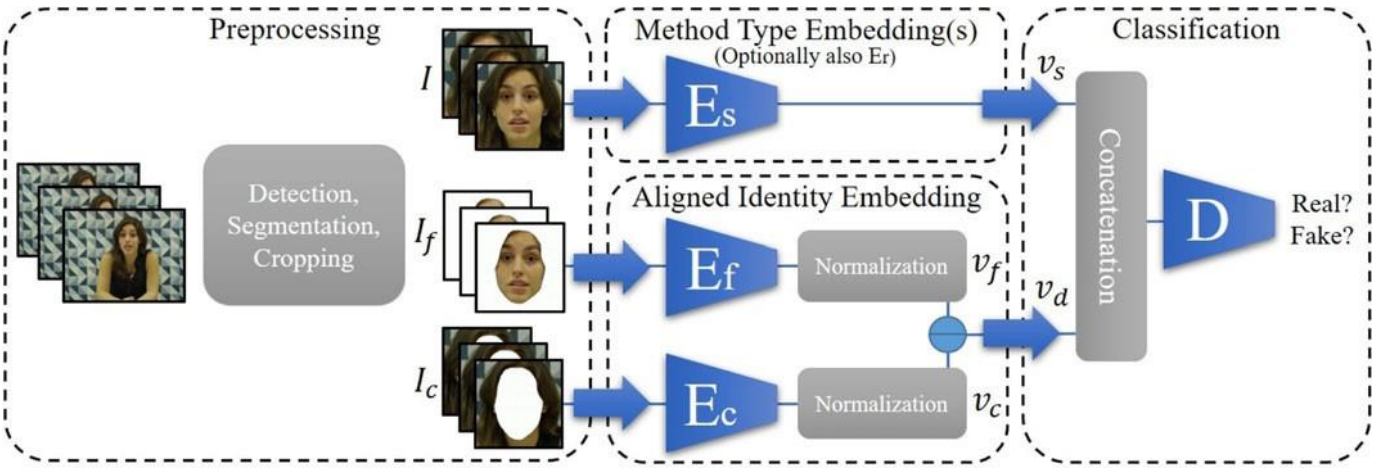

图4.*方法概述*。经过初步的预处理，我们得到了人脸的区域，I_f ，以及它的背景，I_c 。这两个区域分别由人脸识别网络E_f 和 E_c 处理。一个单独的网络，E_s ，考虑输入的图像I，寻找明显的交换假象，以决定它是否是一个脸部交换的结果。两个人脸识别网络的伪概率向量被减去，并与从方法类型网络E_s 得到的代表共同传递给最终分类器D。

我们的测试中使用了FaceForensics++基准（第5.2节）。因此，我们的方法还包括一个检测人脸重演的组件。

具体来说，我们通过训练一个单独的、专门的分类器来解开交换和重演的关系。网络E_s 被训练来检测互换文物，网络E_r （图4中未显示）被训练来检测重演。我们使用Xception网络，类似于第3.2节中描述的用于识别的网络，并训练这些网络来对真实的和被操纵的进行分类。我们的训练过程首先对这两个网络进行预训练，训练的对象是他们特定的操纵与原始图像的例子。我们的重演网络，E_r ，用于检测人脸交换和人脸重演方法的情况。否则，我们使用一个三网络解决方案，其中E_r 被省略。

4.3 结合所有检测线索

我们选择了最简单的方法来组合各种信号：将三个向量v_d 、v_s 和v_r 连接起来，其中v_d 2 $R^{8;631}$ 在公式（1）中定义，v_s ¼ E^p ðlÞ和v_r ¼ E^p ðlÞ，都在$R^{2;048}_{s}$ ，分别表示二元E_s 和E_r 的多配偶层的激活。

连接的向量被传递给分类器D，它输出一个真实与虚假的二进制信号，使用Logistic损失函数进行训练。分类器D由一个初始的线性层组成，然后是批量归一化、ReLU和最后的线性层。

4.4 培训

我们首先对四个分类器E_s 、E_r 、E_f 和E_c 进行预训练，每个分类器都有自己的任务。我们在FaceForensics++[5]中的视频子集上训练网络E_s ，该子集包括原始视频和被换脸方法操纵的视频。FaceSwap和Deepfakes。网络E_r 是在人脸重现方法上训练的。Face2Face和NeuralTextures。请注意，我们只使用这些视频的压缩版本进行训练，压缩量为C23（HQ）和C40（LQ）。我们选择不使用原始视频进行训练，因为它们与C23压缩视频之间几乎没有区别。用来测试我们的FaceForensics++基准的是方法确实包含所有三个版本。应用于E_f 和E_c 的训练过程详见第3节。

一旦四个网络训练完成，我们冻结E_f 和E_c 的权重，并使用三个输出向量（v_s ；v_r ；v_d ）训练最终的分类网络D，同时只对E_r 和E_s 进行微调。最终的训练是在FaceForensics++视频的同一分割上进行的。更多技术细节，请见补充资料，可在线查阅。

4.5 对完整图像的推理

在推理过程中，我们经常处理含有多张人脸的图像。在这种情况下，我们只对检测到的高度大于64像素的人脸进行分类，而将其余的人脸作为背景脸丢弃。唯一的例外是那些最大的人脸不符合这一标准的图像，在这种情况下，我们处理最大的检测到的人脸。

我们通过对每次检测的人脸分割遮罩中的人脸像素的数量施加一个阈值，进一步消除错误检测。相对于裁剪区域的像素数量，我们以15%的人脸像素的阈值开始。如果这一步过滤掉了我们所有的检测，我们就把阈值减少一半。如果没有一张图像通过7.5%的阈值，我们就只考虑检测到的像素数量最多的那张脸。

最后，我们将复合网络，包括E_m ；E_f ；E_c ，和D，应用于剩余的脸部斑块（一个或多个），并得到每个脸部斑块的一个分数，作为D的输出。
D.在只有一张脸被操纵的情况下，我们取这些分数的最小输出--预测为最可能是假的脸部补丁。

5 实验结果

我们使用三个最新的、具有挑战性的基准来评估我们提出的方案。FaceForensics++ [5], DFDC [22], 和 Celeb-DF-v2 [21]。为了评估我们的方法，使用更多的脸部交换技术，并测试其通用能力，我们进一步创建了我们自己的测试集，使用另外两种交换方法。

表2
脸部互换检测结果

方法	FF-DF	名人-DF-v2
双流[54]	70.1	53.8
Meso4 [8]	84.7	54.8
中层概念4 [8]	83.0	53.6
头部姿势[53〕	47.3	54.6
FWA [45]	80.1	56.9
DSP-FWA [45]	93.0	64.0
VA-MLP [49〕	66.4	55.0
VA-LogReg [49]	78.0	55.1
XceptionNet-raw [5]	**99.7**	48.2
XceptionNet-c23 [5]	**99.7**	65.3
XceptionNet-c40 [5]	95.5	65.5
多任务 [64〕	76.3	54.3
胶囊 [51]	96.6	57.5
我们的	**99.7**	**66.0**

我们的方法和领先的先进方法在两个基准上的比较，使用框架级别的AUC（%）。

表3
FaceForensics++图像基准测试结果

方法	DF	F2F	雇员	NT	原始的	共计
斯蒂格。特征 [11]	73.6	73.7	68.9	63.3	34.0	51.8
Cozzolino等人[10]。	85.4	67.8	73.7	78.0	34.4	55.2
Rahmouni等人[12]	85.4	64.2	56.3	60.0	50.0	58.1
巴亚尔和斯塔姆[9]	84.5	73.7	82.5	70.6	46.2	61.6
中介网络[8]	87.2	56.2	61.1	40.6	72.6	66.0
Xception [5]	96.3	86.8	90.3	80.7	52.4	71.0
我们的	94.5	80.3	84.5	74.0	67.6	**75.0**

*栏目是。*DeepFakes *(DF)*, Face2Face *(F2F)*, FaceSwap *(FS)*, Neural- Textures *(NT), 和* Pristine *类别。很难比较具体的栏目，因为真假之间有一个基于阈值的权衡。因此，提供这些栏目只是为了完整性。我们的方法在总分上领先，这是对这个基准有意义的指标。*

5.1 换脸检测实验

我们使用了以下三个只包含脸部互换例子的数据集。

FF-DF。FF-DF[21]是FaceForensics++基准[5]的一个子集，它只包括使用Deepfakes方法[4]交换的脸。因此，这些测试包括来自*原始*子集的1000个视频和来自*Deepfakes*子集的1000个视频（完整的FaceForensics++在第5.2节描述）。

DFDC。最近公布的、由行业支持的DFDC基准的预视图[22]提供了66个演员的总共5244个视频。4,464个训练视频和780个测试视频，其中1,131个是真实视频，4,113个是由两种不同的、未知的脸部互换方法产生的假视频。

Celeb-DF-v2。另一个最近的数据集包含590个真实视频和59个名人的5,639个DeepFake视频[21]。这个数据集特别具有挑战性，因为在这个数据集上测试的大多数最先进的方法都报告了近乎偶然的准确性。

训练和评估。在这些测试中，我们不使用我们的重演网络，$\boldsymbol{E_r}$。我们在FaceForensics++上训练，如第3节所述。所有基线方法的结果以前都有报告[21]。这些方法主要是在FaceForensics++上训练的，有时也会使用额外的自我收集的数据。这些方法都没有在DFDC或Celeb-DF-v2上训练过，因此这些实验也比较了不同方法的通用性。

所有的方法都是使用曲线下的面积（AUC）进行比较的，在所有检测到人脸的帧上都是如此。这个指标对于比较输出每帧分类的方法非常方便，因为不需要设置阈值。

脸部互换检测结果。我们在表2中报告了我们的结果。我们的方法在所有的基准测试中都取得了最好的AUC分数。在FaceForensics的DeepFakes子集[5]上，我们的方法取得了与当前技术水平类似的结果，这是由于准确率已经饱和。在更具挑战性的Celeb-DF-v2基准上，AUC分数的小幅提高是显著的。还要注意的是我们在Celeb-DF-v2上报告的结果证明，与基线方法相比，它的泛化能力有所提高。

5.2 关于FaceForensics++的实验

完整的FaceForensics++数据集[5]包含1000个从网络上获得的视频，从中随机选择1000个视频对，用于生成额外的1000个操纵视频，代表四种人脸操纵方案。其中两种方法进行人脸互换：一种是基于3D的人脸互换方法[6]，使用传统的图形管道和混合，另一种是基于GAN的方法[4]，使用成对的主体图像进行训练，计算它们之间的映射关系。另有两种方法进行面部重演。Face2Face[2]，一种基于3DMM的方法，通过改变为面部设计的表情系数来操纵面部表情，以及NeuralTextures[3]，它从视频中学习面部神经纹理，并使用它来真实地渲染三维重建的面部模型。

FaceForensics++图像基准的结果。在这个基准中，结果是通过上传二进制预测值在一个私人服务器上计算的。因此，需要为模型的预测分数选择一个阈值，我们通过对验证集的优化选择了这个阈值。表3显示，我们的总准确率以很大的幅度超过了以前的所有方法。重要的是，每个不同类别的准确度，就其本身而言，并不是检测性能的直接标志，因为在真实图像和虚假图像上的准确度之间存在着一个依赖于阈值的权衡。这些结果暗示了每个类别的相对检测难度，为了完整起见，我们提供这些结果。

5.3 消融研究和普及实验

脸部操作方法有时会留下人工痕迹，可能是无法察觉的，可以利用这些痕迹进行检测。然而，不同的操纵方法会产生不同的假象，如图5所示。因此，不能保证一个假货检测方法在遇到由未见过的方案产生的假货时能有好的表现，这些方案不会留下这种已知的、可识别的人工制品。我们接下来验证我们提出的方案在检测由不属于其训练集的方法产生的假货方面的准确性。

图5.*用未见过的方法扩展FaceForensics++*。图中所示的例子为同一源/目标脸对，使用基于3D的方法，FaceSwap[6]和Nirkin等人[18]，以及基于GAN的方法，Deepfakes[4]和FSGAN[7]。尽管在这四个例子中使用了相同的图像对，但结果是不同的，每个例子都表现出自己的伪装。

我们通过扩展FaceForensics++集来进行这些测试，对其视频应用两个额外的脸部交换方法。(1) FSGAN[7]和(2)Nirkin等人[18]，一个基于3D的人脸交换方法，使用单图像的3D人脸重构和分割，都有公开可用的实现。这四种脸部交换方法的例子，使用相同的源和目标，可以在图5中看到。每种方法产生的人脸交换都有明显的伪影，只有FSGAN例外，它产生的图像有较少的明显伪影。

该基准的扩展版本遵循原始FaceForensics++数据集所规定的配对选择。因为Nirkin等人[18]是为图像与图像之间的人脸互换而设计的，对于目标视频中的每一帧，我们都会在源视频中选择与其最接近的一帧，以确定头部的姿势。

在我们所有的泛化实验中，我们使用FaceForensics++的官方训练和验证子集，对我们的方法及其XceptionNet基线的变体进行训练，以达到原始和换脸的目的。在这些实验中，我们没有使用重演检测网络$\boldsymbol{E}_r$。

5.3.1 *归纳和消融结果*

我们通过与一个天真的分类器进行比较，研究了我们的脸部与上下文不一致的方法的效果。我们还考虑了我们方法的另外三个变种。
(i)所有分类器在训练过程中被冻结的版本，(ii)我们方法的端到端版本，即所有分类器在训练过程中被解冻，最后，(iii)一个变体，即我们不是减去$\boldsymbol{v}_f$和$\boldsymbol{v}_c$，而是将它们连接起来。

我们在表4中报告了我们的概括结果（ROC曲线在图6中提供）。对于出现在表4顶部的结果，我们将XceptionNet和我们的方法的阈值固定为零。在表4的底部，我们在测试集上优化了两个阈值。第一个实验中人脸身份差异的阈值是使用VGGFace2测试集优化的。

我们的结果显示，我们的方法在两种未见过的方法上都明显优于基线。在FSGAN生成的面孔上，性能差距更大，那里的伪影更少。基于3DMM的方法产生的假象与我们在其他方法中遇到的假象更为相似，因此差距较小。

从图6中的ROC曲线可以看出，我们方法的冻结版本，即不给特定方法分类器以调整身份信号的选项，是性能最差的变体。我们的方法的端到端版本的概括能力也较差。这一结果是由于端到端训练过程玷污了面部和文本分类器在提取一致的身份代表方面的作用。连接变体的表现比我们的方法略差。这可能是由于$\boldsymbol{D}$的容量增加的结果。

最后，请注意，脸部差异信号本身与经过训练的检测假货的网络没有竞争力。然而，它对假视频有指示作用，通过比较我们的方法和基线XceptionNet，可以看出它对整个方法的贡献。

5.3.2 *图像清洗消融*

我们证明了我们的方法在不同的图像清洗攻击下的泛化性能，在三种脸部交换方法上，从旧到新。基于3D的交换[18]，FSGAN[7]，和FaceShifter[38]。图像启动操作包括JPEG压缩25，50和
75%，其中较高的百分比意味着更强的

表4
泛化消融

方法	基于3D的互换			FSGAN		
	假的	真实的	共计	假的	真实的	共计
脸部身份差异	47.33	77.66	62.50	34.66	80.50	57.58
二进制XceptionNet [10]	55.38	**97.72**	76.55	24.80	94.68	59.74
我们的（冰冻的）	52.79	96.44	74.62	34.76	92.46	63.61
我们的（端到端）	54.74	97.70	76.22	31.66	**95.38**	63.52
我们的(连接)	55.42	96.54	75.98	41.64	93.30	67.47
我们的	**68.20**	95.10	**81.65**	**47.14**	90.56	**68.85**
脸部身份差异	60.20	66.12	63.16	38.96	77.50	58.23
二进制XceptionNet [10]	89.03	81.36	85.20	73.92	64.04	68.98
我们的（冰冻的）	85.52	86.92	86.22	67.66	76.20	71.93
我们的（端到端）	90.77	83.54	87.16	**79.58**	71.40	**75.49**
我们的(连接)	**91.41**	84.34	87.87	71.92	**78.00**	74.96
我们的	90.52	**88.20**	**89.36**	78.72	71.66	75.19

我们的方法在我们的FaceForensics++[5]测试集的扩展版本上的变体的概括结果。顶部：固定阈值为零的结果。下图。上限结果，在测试集上用固定阈值获得最大的总精度。更多细节见第5.3节。

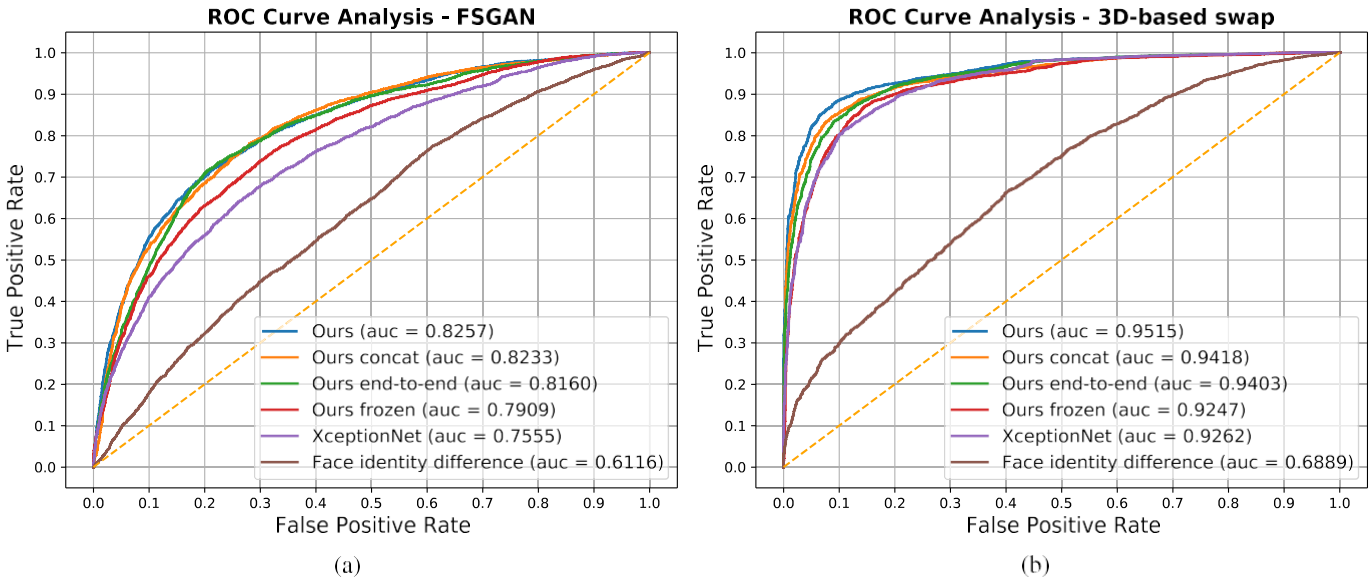

图6.*我们对FaceForensics++视频的两种变化的结果。*(a) 使用FSGAN产生的交换的概括结果[7]。(b) 由Nirkin等人[18]生成的交换的概括结果。更多细节见第5.3节。

表5
图像洗衣机消融

Methods	3D-based swap							FSGAN							FaceShifter						
	RAW	C25	C50	C75	S25	S50	S75	RAW	C25	C50	C75	S25	S50	S75	RAW	C25	C50	C75	S25	S50	S75
Face identity difference	63.16	62.41	63.19	62.73	61.67	62.58	63.02	58.23	57.79	57.32	56.25	56.25	57.72	58.13	55.89	55.92	56.62	55.65	54.19	55.33	55.85
Binary XceptionNet [10]	85.20	84.00	83.07	80.24	75.87	82.09	83.92	68.98	67.68	66.66	63.68	64.77	69.84	70.06	63.62	63.79	62.76	61.90	59.35	63.51	64.43
Ours	**89.36**	**87.49**	**85.79**	**82.21**	**77.51**	**85.46**	**88.06**	**75.19**	**73.60**	**71.96**	**68.61**	**68.07**	**75.03**	**75.85**	**67.64**	**66.96**	**65.85**	**64.28**	**61.47**	**66.66**	**67.52**

使用FaceForensics++[5]测试集的视频对三种人脸交换方法的概括结果。基于3D的交换[18]，FSGAN[7]和FaceShifter[38]，其中图像受到不同的调整和压缩。'RAW'，图像未被改变，'C#'，JPEG压缩操作（百分比越高表示压缩越强），'S##'是缩放操作（相对于原始分辨率的百分比）。

压缩，以及相对于原始分辨率的缩放，也是25%、50%和75%。

结果详见表5。正如预期的那样，当应用超过25%的压缩和超过75%的比例时，洗钱攻击降低了所有检测方法的准确性，压缩或比例越大，准确性下降越大。在所有不同的洗钱攻击下，我们的方法始终比Xception- Net[10]和面部身份差异基线要好得多。

最后，结果表明，脸部身份差异在较新的方法上变得不那么有效。最近的人脸互换方法改善了对目标脸的估计姿势和表情。因此，这些方法允许更多的目标脸的身份信号被保留下来，从而降低了脸部身份差异的有效性。

5.4 定性结果

图7展示了从DFDC集合中检测到和遗漏的假脸的定性例子。图7a显示了被我们的方法检测到的假脸，但没有被最先进的XceptionNet假脸检测器[60]检测到的例子。图7b提供了被XceptionNet检测到，但被我们的方法遗漏的假脸实例。最后，图7c显示了被两种方法遗漏的假货。

显然，我们的方法在难以检测到交换假象的情况下表现出色（图7a）。图7b显示，由XceptionNet检测到的假图像通常表现为可见的伪影，该方法被优化来检测。我们的方法包括一个脸部交换组件，$\boldsymbol{E_s}$（第4.2节），它被训练来检测类似的方法特定的伪影，但是当这些伪影出现时，并没有提供与基线相同的检测精度。正如第5.1和5.2节所报告的那样，我们的整体方法仍然远远超过了基线。最后，两种方法漏掉的假象通常是具有低对比度或模糊特征的挑战性图像，如图7c。

6 讨论和限制

一些最新的方法通过生成整个头部来进行面部操作[65], [66]。这些方法通常采用预先训练好的StyleGAN2[67]网络，或者采用其架构。通过操纵StyleGAN2的潜伏代码来控制生成，以保持源身份并保留目标脸的属性。虽然这些方法在保持源脸的外观和纳入目标脸的属性方面是成功的，但目前姿势和表情的准确性较低。因此，这些方法缺乏时间上的一致性。当应用于视频时，其效果是非常明显的。

在未来，这些方法可能会克服目前的限制，并能够在视频中进行全头面部的交换。这将创造出一类新的方法，我们的方法所依据的假设将不成立。

图7.*定性检测结果*。例子取自DFDC文集。(a) 被我们的方法检测到的假货，但未被领先的基线，XceptionNet假货检测器[60]检测到。(b) 被XceptionNet检测到的假货，但被我们的方法忽略。(c) 两种方法都漏掉的假货。更多细节见第5.4节。

6.1 脸部重演检测线索

脸部重演是由$\boldsymbol{E_r}$ 网络（见第4.2节）检测出来的，该网络*经过专门训练*，可以区分真实图像和脸部重演方法所操纵的图像。此外，考虑到图2a和2b，似乎人脸重现方法所操纵的区域与人脸互换方案所创造的区域相似。因此，$\boldsymbol{E_f}$ 和$\boldsymbol{E_c}$ 可能利用了除主体身份以外的线索，例如那些由于传感器和镜头而产生的线索。这些标记可能会在合成过程中被覆盖，也可能会提高对换脸操纵的检测。

7 总结

虽然在过去的几年里，在图像和视频中操纵人脸的能力有了很大的提高，但最近的大多数方法都遵循类似的模式。在这项工作中，我们提出了一个新的检测线索，它利用了所有最近的人脸身份操纵方法的共同点。它是对传统的真/假分类器的补充，可以与它们一起使用。克服这种方法需要将新的身份更广泛地整合到图像中，使我们的贡献在没有额外技术突破的情况下很难规避。这与人工制品检测方法相反，后者容易受到生成图像的视觉质量的不断进步的影响。我们希望通过进一步分析人脸互换技术的设计原理，发现更多识别假图像和视频的方法，从而有效地减轻此类媒体的社会风险。

鸣谢

这项工作得到了欧洲研究理事会（ERC）通过欧盟地平线2020的支持。

在ERC CoG 725974资助下的研究和创新计划。Lior Wolf, Yosi Keller, 和Tal Hassner有同等贡献。

参考文献

[1] 谷歌AI，"为深度造假检测研究提供数据"。[在线]。Available: https://ai.googleblog.com/ 2019/09/contributing-data-to-deepfake-detection.html

[2] J.Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face:RGB 视频的实时人脸捕捉和重演"，在*Proc.Conf.Comput.Vis.Pattern Recognit.*, 2016, pp.2387-2395.

[3] J.Thies, M. Zollho∈fer, and M. Nießner, "Deferred neural rendering: 使用神经纹理的图像合成"，2019，*arXiv:1904.12356*。

[4] Deepfakes, "Deepfakes."已访问。2019年11月15日。[在线]。可以使用：https://github.com/deepfakes/faceswap

[5] A.Ro∈ssler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++:学习检测被操纵的面部图像》，2019年，*arXiv:1901.08971*。

[6] FaceSwap, "FaceSwap."已访问。2019年11月15日。[在线]。可利用：https://github.com/MarekKowalski/FaceSwap/

[7] Y.Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject agnostic face swapping and reenactment," in *Proc. Int. Conf.Conf.Comput.Vis.*, 2019年，第7184-7193页。

[8] D.Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet:一个紧凑的面部视频伪造检测网络"，在*Proc. Int.Workshop Inf.Forensics Secur.*, 2018，第1-7页。

[9] B.Bayar和M.C. Stamm, "使用新的卷积层对univer- sal图像操作检测的深度学习方法"，在*Proc. Int. Workshop Inf.Int. Workshop Inf.Hiding Multimedia Secur.*, 2016, pp.5-10.

[10] D.Cozzolino, G. Poggi, and L. Verdoliva, "重铸基于残差的局部描述符作为卷积神经网络。应用于图像伪造检测"，在*Proc. Int.Int. Workshop Inf.Hiding Multimedia Secur.*, 2017年，第159-164页。

[11] J.Fridrich和J. Kodovsky, "Rich models for steganalysis of digi- tal images," *IEEE Trans.Inform.Forensics Secur.*, vol. 7, no.3, pp. 868-882, Jun. 2012.

[12] N.Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," in *Proc. Int.Int. Workshop Inf.Forensics Secur.*，2017年，第1-6页。

[13] U.A. Ciftci, I. Demir, and L. Yin, "How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals," in *Proc. IEEE Int.Joint Conf.生物统计学（IJCB）*，2020年，第1-10页。

[14] H.Qi等人，"Deeprhythm。用注意力的视觉心跳节奏暴露深层假象，"在*Proc.28th ACM Int. Conf.Conf.多媒体*，2020年，第4318-4327页。

[15] S.Hu, Y. Li, and S. Lyu, "Exposing GAN-generated faces using inconsistent corneal specular highlights," 2020, *arXiv:2009.11924*.

[16] L.Li等人，"用于更普遍的人脸伪造检测的人脸X光"，在*Proc. IEEE/CVF Conf.Comput.Vis.Pattern Recognit.*, 2020, pp.5001-5010.

[17] N.Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification, " in *Proc.Conf.Comput.Vis.Pattern Recognit.*, 2009, pp.365-372.

[18] Y.Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *Proc. Int. Conf.Conf.Autom.Face Gesture Recognit.*, 2018, pp. 98-105.

[19] V.Blanz, S. Romdhani, and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model," in *Proc. Int. Conf.Conf.Autom.人脸手势识别"，2002年，第192-197页。*，2002年，第192-197页。

[20] V.Blanz和T. Vetter, "基于拟合三维可变形模型的人脸识别"，*IEEE Trans.模式分析。Mach.Intell.*25卷，第9期，第1063-1074页，2003年9月。

[21] Y.Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF:A new dataset for deepfake forensics," 2019, *arXiv:1909.12962*.

[22] B.Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (DFDC) preview dataset," 2019, *arXiv:1910.08854*.

[23] D.Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping:自动替换照片中的人脸"，*ACM Trans.图形。*, vol. 27, no.3, 2008, Art. No.39.

[24] V.Blanz, K. Scherbaum, T. Vetter, and H.-P.Seidel, "交换图像中的面孔，" *Comput.图形。Forum*, vol. 23, no.3, pp. 669-676, 2004.

[25] Y.Lin, S. Wang, Q. Lin, and F. Tang, "大姿态变化下的脸部交换。一个基于3D模型的方法，"在*Proc. Int.Conf.多媒体博览会*，2012年，第333-338页。

[26] S.Mosaddegh, L. Simon, and F. Jurie, "Photorealistic face de-identification by aggregating donors' face components," in *Proc.亚洲会议。Comput.Vis.*, 2014, pp.159-174.

[27] I.Kemelmacher-Shlizerman, "Transfiguring portraits," *ACM Trans.图形。*, vol. 35, no.4, 2016, Art. No. 94.

[28] O.Alexander, M. Rogers, W. Lambeth, M. Chiang, and P. Debe-vec, "创建一个逼真的数字演员。数字Emily项目"，在*Proc.Conf.Vis.Media Prod.*，2009年，第176-187页。

[29] L.Wolf, Z. Freund, and S. Avidan, "以眼还眼。一个单机注视替换方法"，在*Proc.Conf.Comput.Vis.Pat- tern Recognit.*，2010年，第817-824页。

[30] S.Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman , "Synthesizing Obama:从音频中学习唇语同步，" *ACM Trans.图形。*, vol. 36, no.4, 2017, Art. No. 95.

[31] H.Averbuch-Elor , D. Cohen-Or, J. Kopf, and M. F. Cohen, "Bringing portraits to life, " *ACM Trans.图形。*，第36卷，第6号，2017年。第196条。

[32] A.Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer , "Ganimation:Ganimation: Anatomically-aware facial animation from a single image," in *Proc.Eur.Conf.计算。Vis.*, 2018, pp.818-833.

[33] E.Sanchez和M. Valstar, "Triple consistency loss for pairing distributions in GAN-based face synthesis," 2018, *arXiv:1811.03492*.

[34] H.Kim *et al.*, "Deep video portraits," *ACM Trans.图形。*, vol. 37, no.4, 2018, Art. no.163.

[35] R.Natsume, T. Yatagawa, and S. Morishima, "FSNet:基于图像的人脸交换的身份感知生成模型"，在*Proc.Asian Conf.Comput.Vis.*, 2018，第117-132页。

[36] R.Natsume, T. Yatagawa, and S. Morishima, "RsGAN: Face swapping and editing using face and hair representation in latent spaces," 2018, *arXiv: 1804.03447*.

[37] K.Nagano等人，"paGAN: Real-time avatars using dynamic textures," *ACM Trans.Graph.(TOG）*，第37卷，第6期，第1-12页，2018。

[38] L.Li, J. Bao, H. Yang, D. Chen, and F. Wen, "Faceshifter:Towards high fidelity and occlusion aware face swapping," 2019, *arXiv:1912.13457*.

[39] S.Jia, Z. Xu, H. Wang, C. Feng, and T. Wang, "Coarse-to-fine copy-move for video forensics, " *IEEE Access*, vol. 6, pp. 25323-25335, 2018.

[40] Y.Wu, W. Abd-Almageed, and P. Natarajan, "Busternet:Busternet: Detect- ing copy-move image forgery with source/target localization," in *Proc.Eur.Conf.Comput.Vis.*，2018年，第168-184页。

[41] Y.Wu, W. Abd-Almageed , and P. Natarajan, "Image copy-move forgery detection via an end-to-end deep neural network, " in *Proc.冬季会议。Appl. Comput.Vis.*, 2018, pp.1907-1915.

[42] Y.Wu, W. AbdAlmageed, and P. Natarajan, "ManTra-Net:用于检测和定位具有异常特征的图像伪造品的操纵追踪网络"，在*Proc. IEEE Conf.Comput.Vis.Pattern Recognit.*, 2019年，第9543-9552页。

[43] P.Korshunov和S. Marcel, "篡改视频中的说话人不一致检测", in *Proc.Eur.Signal Process.Conf.*, 2018, pp.2375-2379.

[44] Y.Li, M.-C.Chang, and S. Lyu, "In ictu oculi: Exposing AI generated fake face videos by detecting eye blinking," 2018, *arXiv: 1806.02877*.

[45] Y.Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," 2018, *arXiv:1811.00656*.

[46] W.Quan, K. Wang, D.-M.Yan, and X. Zhang, "Distinguishing between natural and computer-generated images using convolutional neural networks," *Trans.Inform.Forensics Secur.*，第13卷，第11期，第2772-2787页，2018。

[47] A.Ro∈ssler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics:A large-scale video dataset for forgery detection in human faces," 2018, *arXiv:1803.09179*.

[48] E.Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," *CVPRw*, pp.80-87, 2019.

[49] F.Matern, C. Riess, and M. Stamminger, "Exploiting visual arti- facts to expose deepfakes and face manipulations, " in *Proc.冬季会议。Appl. Comput.Vis.Workshops*，2019，第83-92页。

[50] H.H. Nguyen, T. Tieu, H.-Q.Nguyen-Son , V. Nozick, J. Yamagishi, and I. Echizen, "Modular convolutional neural network for discriminating between computer-generated images and photographic images," in *Proc. Int.Conf.Availability, Rel. Secur.*, 2018, pp.1-10.

[51] H.H. Nguyen, J. Yamagishi, and I. Echizen, "Use of a capsule network to detect fake images and videos," 2019, *arXiv:1910.12467*.

[52] S.-Y.Wang, O. Wang, A. Owens, R. Zhang, and A. A. Efros, "Detecting photoshopped faces by scripting photoshop," in *Proc. Int. Conf.Conf.计算。Vis.*, 2019, pp. 10072-10081.

[53] X.Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsis- tent head poses," in *Proc. IEEE Int.Conf.Acoust., Speech Signal Process.(ICASSP)*, 2019, pp. 8261-8265.

[54] P.Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-stream neu- ral networks for tampered face detection, " in *Proc.Conf.Comput.Vis.Pattern Recognit.Workshops*, 2017, pp.1831-1839.

[55] J.Stehouwer, H. Dang, F. Liu, X. Liu, and A. Jain, "On the detection of digital face manipulation," 2019, *arXiv: 1910.01717*.

[56] K.Songsri-in and S. Zafeiriou, "Complement face forensic detection and localization with faciallandmarks,"2019, *arXiv:1910.05455*.

[57] P.Korshunov和S. Marcel, "Vulnerability assessment and detec- tion of deepfake videos," in *Proc. Int. Conf.Conf.Biometrics*，2019，第1-6页。

[58] O.I. Al-Sanjary , A. A. Ahmed, and G. Sulong, "Development of a video tampering dataset for forensic investigation, " *Forensic Sci. Int.*, vol. 266, pp. 565-572, 2016.

[59] J.Li et *al.* "DSFD: Dual shot face detector," in *Proc.Conf.Comput.Vis.Pattern Recognit.*, 2019年，第5060-5069页。

[60] F.Chollet, "Xception:深度学习与深度可分离卷积"，在*Proc.Conf.Comput.Vis.Pattern Recognit.*, 2017, pp.1251-1258.

[61] C.Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception- v4, inception-resnet and the impact of residual connections on learning, " in *Proc.AAAI Conf.Artif.Intell.*, 2017, pp.4278-4284.

[62] Q.Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2:一个用于识别不同姿势和年龄的人脸的数据集，"在*Proc. Int.Conf.Autom.Face Gesture Recognit.*, 2018，第67-74页。

[63] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "野外的标记脸。用于研究无约束环境中人脸识别的数据库，"UMass Amherst, Univ. Massachusetts, *Tech.Rep.* 07-49, 2007.

[64] H.H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos, " 2019, *arXiv:1906.06876*.

[65] Y.Shen and B. Zhou, "GANs中潜在语义的闭式因子化", 2020, *arXiv:2007.06600*.

[66] E.H∈arko∈nen, A. Hertzmann, J. Lehtinen, and S. Paris, "GANSpace: 发现可解释的GAN控制，" 2020，*arXiv:2004.02546*。
[67] T.Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "分析和提高styleGAN的图像质量，" 在*Proc. IEEE/CVF Conf.Comput.Vis.Pattern Recognit.*, 2020, pp. 8110-8119.
[68] D.P. Kingma和J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv: 1412.6980*.
[69] O.Ronneberger, P. Fischer, and T. Brox, "U-net:用于生物医学图像分割的卷积网络"，在*Proc. Int.Conf.医学。图像计算。Comput.-Assist.Interv.*, 2015, pp.234-241.
[70] T.Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P.Morency, "Openface 2.0: 面部行为分析工具包"，在*Proc. Int.Conf.自动化。Face Gesture Recognit.*, 2018年，第59-66页。

尤瓦尔-尼尔金于2011年在海法的以色列理工学院获得计算机工程学士学位，并于2017年在以色列拉安纳的以色列开放大学获得计算机科学硕士学位。他目前正在以色列拉马特甘的巴伊兰大学电子工程系攻读博士学位。他的研究兴趣包括深度学习、计算机视觉和计算机图形。他是ECCV、ICCV和CVPR的审稿人，并被评为ECCV'20的高质量审稿人。

Lior Wolf在希伯来大学获得博士学位，由Shashua教授指导。他目前是以色列特拉维夫大学计算机科学学院的全职教授。他曾在麻省理工学院Poggio教授的实验室做博士后研究。他是麻省理工学院Poggio实验室的博士后研究员。他是ERC的资助者，是ICCV 2001和ICCV 2019的荣誉奖获得者，以及ECCV 2000和ICANN 2016的最佳论文奖。

约西-凯勒于1994年在海法的以色列理工学院获得电气工程学士学位，并分别于1998年和2003年在特拉维夫大学获得电气工程硕士和博士学位，成绩优异。2003年至2006年，他在美国康涅狄格州纽黑文的耶鲁大学数学系担任吉布斯助理教授。他目前是以色列拉马特甘市巴伊兰大学工程学院的副教授。他的研究兴趣包括计算机视觉，机器和深度学习，以及生物统计学。

塔尔-哈斯纳分别于2002年和2006年在魏茨曼科学研究所获得应用数学和计算机科学的硕士和博士学位。2008年，他加入以色列开放大学数学和计算机科学系，在那里担任副教授直至2018年。2015年至2018年，他是信息科学研究所（ISI）的高级计算机科学家和机器人和智能系统研究所的访问研究副教授。

加州大学维特比工程学院。从2018年到2019年，on的首席应用科学家，他设计了最新的AWS人脸识别管道。自2019年以来，他一直是Facebook AI的应用研究负责人，支持文本（OCR）和人（脸）的照片理解团队。他一直是WACV'18和ICCV'21的程序主席。他也是CVPER'20的研讨会主席，ICCV'17和ECCV'22的教程主席，以及CVPR、ECCV和AAAI的区域主席。他是*IEEE Transactions on Pattern Analysis and Machine Intelligence*和*IEEE Transactions on Biometrics, Behavior, and Identity Science*的副编辑。

••关于这个或任何其他计算主题的更多信息，请访问我们的数字图书馆：www.computer.org/csdl。