使用身份一致性转换器保护名人免受 DeepFake 的影响

Xiaoyi Dong1*, Jianmin Bao2, Dongdong Chen3 †, Ting Zhang2 , Weiming Zhang1 ,
Nenghai Yu1, 陈东2 , Fang Wen2, Baining Guo2
1中国科学技术大学
2微软亚洲研究院 3微软云+人工智能
{dlight@mail., zhangwm@, ynh@}.ustc.edu.cn cddlyf@gmail.com
{jianbao, Ting.Zhang, doch, fangwen, bainguo }@microsoft.com

抽象的

在这项工作中,我们提出了身份一致性转换器,这是一种新颖的人脸伪造检测方法,它专注于关于高级语义,特别是身份信息,并通过发现内部和外部面部区域的身份不一致来检测可疑面部。Identity Consistency Transformer 包含用于确定身份一致性的一致性损失。我们表明,身份一致性转换器不仅在不同的数据集上,而且在各种不同的数据集上都表现出卓越的泛化能力[1]。

一、简介

Deepfake 技术[1-5,8,31,45,46]先进到能够创建令人难以置信的逼真的假图像,其中脸部被其他人替换另一个图像。深度伪造的恶意使用和传播引起了严重的社会关注,并对我们对在线媒体的信任构成了越来越大的威胁。因此,面对伪造检测是迫切需要的,并且最近引起了相当多的关注。

值得注意的是,在所有案例中,绝大多数面部伪造涉及政客、名人和企业领导者,因为他们的照片和视频更容易在网络,因此很容易被操纵以产生令人印象深刻的逼真的深度伪造。以前的检测算法仅根据可疑图像做出关于伪造的预测,而忽略利用那些免费可用的图像数据。在本文中,我们认为在线可用的图像/视频不仅可以用于生成面部伪造也可以用来检测它们,并试图保护人们谁的脸可以在线访问,因此容易被脸操纵,即广义上的名人。

图 1. 五张假脸（第一行)和真脸（第二行)及其身份一致性分数。 ICT分数越高表明内在而外脸更有可能来自两个人,即DeepFake face。

最近,人们致力于检测面部伪造并实现有希望的检测性能。大多数现有方法旨在通过利用低级纹理和搜索底层生成工件[7,9,32,34,37,41,43,44,48,51、65、66]。在现实世界中部署这些技术时产品,我们观察到两个常见问题：(1)深度伪造检测通常对可疑视频进行视频帧有图像退化,例如图像重新缩放、噪声和视频编解码器转换； (2) 当生成的 deepfake 具有令人信服的真实感,伪造的低级痕迹变得非常难以检测。这些问题使视频的 deepfake 检测不稳定输入。我们希望显着地进行深度伪造检测通过大量使用语义上有意义的身份信息来更加健壮。

在本文中,我们提出了一种新的人脸伪造检测称为身份一致性转换器 (ICT) 的技术基于高级语义。关键思想是检测可疑图像中的身份一致性,即是否内面和外面属于同一个人。事实证明,这是一项不平凡的任务。如此天真

*在微软亚洲研究院实习期间完成的工作。
†陈冬冬为通讯作者。
1代码将在https://github.com/LightDXY/发布ICT_DeepFake

图 2. 当前人脸伪造方法的伪造区域。 (a) FF++ 中的 DeepFake。 (b) Google Deepfake 检测中的 DeepFake。 （C）
CelebDF 中的 DeepFake。 (d) DeepFaceLab 中的 DeepFake。 (e)Face2face。 (f) FSGAN (g) DF-VAE。他们用
不同的形状,同时保持外表面不变。

解决方案是比较使用现成的人脸识别模型从内部和外部提取的身份向量,类似于流行的人脸验证

技术。不幸的是,现有的人脸验证方法[21, 57]倾向于表征最具辨别力的
区域,即用于验证的内表面,而未能捕获外表面的身份信息。有身份

Consistency Transformer,我们训练一个模型来学习一对
身份向量,一个用于内面,另一个用于
外表面,通过设计一个变压器,使内
并且可以同时学习外部身份
无缝统一的模型。我们的身份一致性转换器结合了一种新颖的一致性损失来拉动两者
当他们的标签相同时,他们的身份在一起,因此当他们关联时将他们推开
不同的标签。

我们的方法只假设身份差异和
因此可以用大量交换的面孔进行训练
通过交换不同的身份获得,没有任何
人脸处理方法生成的假图像。经验上,我们展示了身份一致性转换器

在基于纹理的低级方法失败的情况下表现出显着优越的性能。图1

展示了五个示例,说明最先进的检测方法[7, 34, 51]未能检测到它们,而我们的
方法揭示了完全不同的身份一致性分数
从而可以区分真假图像。

身份一致性转换器的另一个优点
是当这些信息碰巧可用时,它可以很容易地通过额外的身份信息来增强,
就像名人一样。对于名人,他们的参考图像可用并且具有身份一致性

Transformer 我们构建了一个真实的参考集,由提取的名人身份向量对组成
从而创建一个新的一致性分数来增强身份一致性检测。由此产生的参考辅助
ICT (ICT-Ref) 在
传统的基准数据集,展示了其强大的
为名人提供额外保护的能力。最后,我们
在我们的实验中表明,提出的 ICT 和 ICT-Ref
在两个方向上显着提高泛化能力:1)跨不同数据集,2)更重要的是,

在包括视频应用在内的现实应用中不同的图像退化形式。

二、相关工作

深造一代。我们将人脸处理算法大致分为三类。 (1) 早期尝试是基于地标的方法[10, 58] ,它利用

人脸对齐以找到具有相似姿势的源人脸
调整后的目标面和交换到目标面
颜色、灯光等。因此,这些作品
仅限于交换具有相似姿势的面孔。 (2) 解决
这种限制,随后的努力[12, 16, 19, 25, 36, 45, 46]
引入 3D 人脸表示,例如 Face2Face 方法[54]、人脸重演[52]和表情操纵[17、39、40]。但是,这些方法无法

生成源图像中不存在的部分,
例如,牙齿,导致合成中不真实的伪影
面孔。 (3) 最近基于 GAN 的方法实现的更加生动
换脸结果。科尔舒纳瓦等人。 [29]和 Deep Fakes [4]为每个目标每个人训练一个转换模型,而 RSGAN [42]、 IPGAN [8]和 FaceShifter [31]
介绍了与主题无关的面部交换作品。

尽管当前的 deepfake 一代质量很高
方法,几乎所有的人脸处理仍然更加关注内脸并采用混合后处理

一步来生成一个非常逼真的深度伪造。
深度伪造检测。随着 deepfakes 的流行和可访问性,如何检测如此引人注目的 deepfakes
已经变得越来越重要并吸引了大量的检测方法[?, ?, 7, 9, 30, 32-34, 37, 41, 43,
44, 48, 51, 61, 62, 64–66]最近。大多数这些方法
基于低级纹理来检测视觉伪影,
例如像素级伪影[9, 51]、纹理差异[34]
和混合幽灵[32]。然而,这样的低级纹理
遇到各种破坏性的东西很容易消失
形式的图像退化,导致急剧下降
那些基于低级方法的性能。

为了更有弹性,我们转向利用更多
语义上有意义的痕迹。沿着这个方向,有
存在一些寻求物理不自然的深度伪造视频检测方法,例如,利用头部姿势不一致
使用 3-D 几何[61]检测内脸和周围环境之间的异常低或高的眨眼次数[33]。

在这项工作中,我们有兴趣利用身份信息,这是另一种高级语义

更强大的人脸伪造检测。最相关的作品

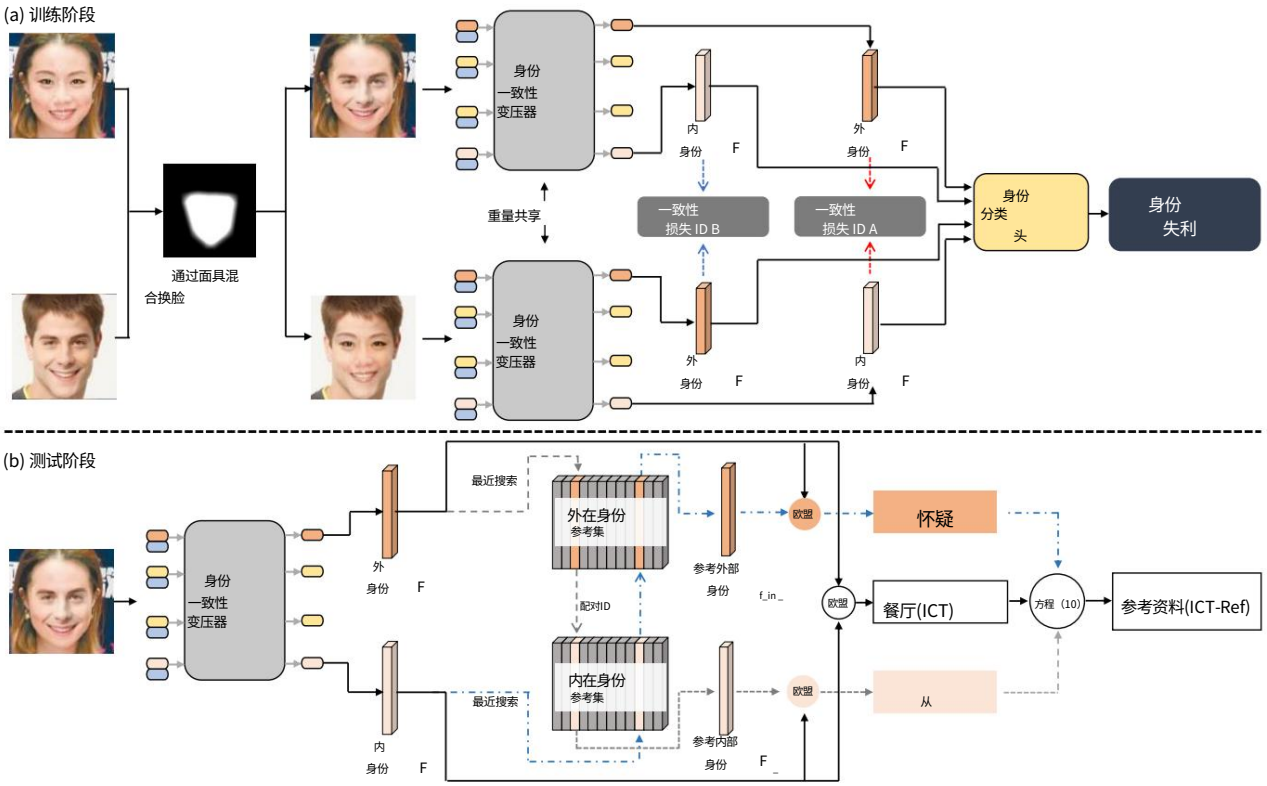

图 3. 说明我们提出的身份一致性转换器的 (a) 训练阶段和 (b) 测试阶段。

是[18]和[47]。第一项工作[18]提出了基于示例的伪造检测来检测参考视频给出的假视频,而我们的方法能够做出预测

仅基于可疑图像。即使在参考辅助变体中,参考也会自动检索,而不是

用户给的。第二部作品[47]也收获了
内表面和外表面之间的身份一致性
脸。然而我们的工作是不同的,具有三个优点:1)
我们的模型在获得身份提取网络后不需要任何进一步的训练,而他们的模型则需要;
2)我们的模型不需要生成任何假视频
通过他们的模型进行的面部操作方法; 3)
我们引入了一种额外新颖且有效的身份一致性损失,而不是仅采用人脸识别损失作为
在[47]中。根据经验,我们的工作取得了更好的结果,
例如,Celeb-DF 数据集上的 85.71%,比 85.71% 高出近 20%
[47]中的数字。

变压器。完全基于自注意力机制的Transformer [55]首次在建模文本和

在许多 NLP 任务中已经显示出巨大的成功[11, 22、49、50] 。最近的作品[14, 15, 23, 24, 56, 59]有
将 Transformer 扩展到图像,并在视觉任务中取得了优异的成绩,例如,图像分类
ViT [15,23,24], [56]中的图像修复,图像生成
在 iGPT [14]和视频识别[59] 中。

我们展示了 Transformer 也适用于人脸分类,这是一种细粒度的图像分类。尽管

这似乎是一个自然的延伸,它实际上很重要。
正如人们普遍认为的那样,区分细粒度
类别在很大程度上取决于细微的和局部的差异,我们证明了具有所有全局注意力的 Transformer 也可以学习语义上有意义的特征

细粒度的类。除此之外,我们还特别修改了 Transformer,引入了一个新的一致性损失
使其适用于身份一致性检测。

3. 身份一致性转换器

给定输入人脸图像 I,deepfake 检测的目标是将输入 I 分类为真实图像或
一个深度伪造。我们首先介绍身份提取
模型,然后描述如何利用身份一致性来区分给定的输入。我们的概述

方法如图3 所示。

3.1。身份提取模型

提取内脸的身份信息为
除了外脸,我们从 Transformer 最近成功应用于全尺寸图像[14, 24]中汲取灵感,并将视觉 Transformer [24]应用到脸部

分类学习身份信息。这带来
内在身份和外在身份的优势
可以通过统一模型同时学习
采用两个独立的人脸识别模型,每个模型
对两个身份之一负责。此外,

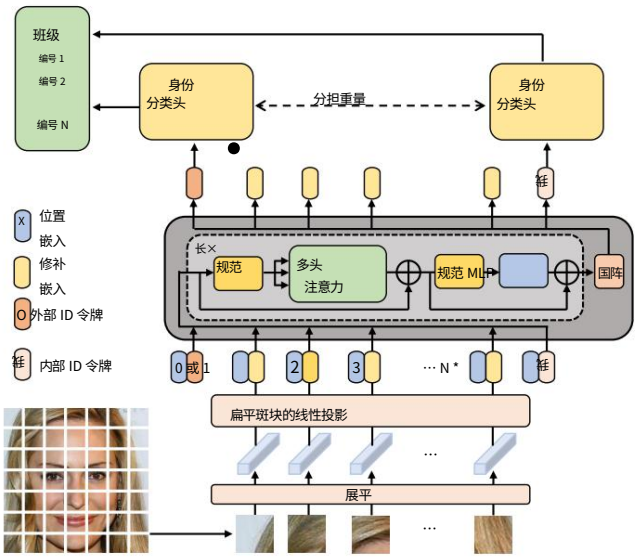

图 4. 身份提取模型的架构。

$$\frac{s \cdot \cos(\theta_{i,j+1\{j=y_i\}} \cdot m)}{\sum_{j=1}^{N}}$$

输入输出= p ji in - f ji 出去 2
出 输入= p ji ij 2
out in - f ji 2
ij ij

在 出去

在

3.2.身份一致性检测

在

从外面

3.3.好处、限制和影响

在 ñ
n n n=1。

在 在 在 在
n

在
亲属

出去
亲属

出去 出去

在 在

出去 出去 出去 出去 出去
n

1

4. 实验

方法	DFD	FF++	更深	CD1	CD2	Avg		
多任务[43]	65.21	72.23	65.32	72.28	61.06	65.96		
MesoInc4 [7]	59.06	63.41	51.41	42.26	53.60	53.95		
胶囊[44]	69.70	96.50	68.44	69.98	63.65	67.94		
Xcep-c0 [51]	89.05	99.26	57.76	48.08	50.37	61.32		
Xcep-c23 [51]	95.60	98.54	69.85	74.97	77.82	79.56		
时间[34]	80.59	74.82	45.46	72.88	64.87	67.72		
DSP-FWA [34]	90.99	81.90	60.00	78.51	81.41	78.56		
CNNDetect [60]	60.12	71.08	57.16	56.12	57.17	60.33		
Patch- Foren [13]	49.91	73.75	55.35	59.66	57.16	55.52		
FFD [20]	76.61	92.32	46.64	74.15	77.80	73.50		
人脸 X 线[32]	94.14	98.44	72.35	74.76	75.39	79.16		
两分支[38]* PCL+I2G	-	-	-	-	-		73.41	-
[63]* Nirkin.et al.	-	-	-	-	-		81.80	-
[47]* ICT（我们的）	-	99.70	-	-	-		66.00	-
	84.13	90.22	93.57	81.43	85.71	87.01		
ICT-Ref（我们的）	93.17	98.56	99.25	96.41	94.43	96.34		

表 1. 未见过数据集上的 DeepFake 检测 AUC。这里
灰色意味着闭集评估（训练和测试在同一
数据集）。 Avg 表示平均开集 AUC（闭集
结果和只有一个结果的方法不包括在内）。 * 表示代码未发布,我们报告结果来自

原纸。除 ICT Ref 外,所有方法中最好的结果用粗体表示。

12 个块和 12 个头用于多头自注意力。

输入图像被划分为 14×14 块,我们投影
每个补丁嵌入维度为 384 的特征。
训练周期数为 30,批量大小为 1024。
初始学习率设置为 0.0005 并除以 10
在 12、15、18 个 epoch 之后。损失平衡权重参数η在第一个epoch设置为4,之后增加0.5
每个时代。为了进一步简化训练,我们设置了margin
m 在第一个时期为 0,在 10 个时期后增加到 0.3。

4.1。与最先进方法的比较

我们将我们的方法与两类最先进的方法进行比较：

设计用于检测来自某些面部操作的图像:Multi-task [43]、 MesoInc4 [7]、Capsule [44]、
Xception-c0、c23 [51]和 FWA、DSP-FWA [34]、两个分支[38]、PCL+I2G [63]、 Nirkin.等。 [47];以及旨在检测一般深度伪造的方法:面部 X 射线[32]，
FFD [20]、 CNNDetection [60]和补丁取证[13]。
的确,泛化能力是最重要的
deepfake检测的属性,用于实际使用,以及
现有方法中最大的挑战。我们评估
两个方向的泛化能力:跨越不同
看不见的数据集和不同的看不见的图像退化形式。评估指标是广泛使用的 AUC

（接收器操作特征曲线下的面积）
我们在所有实验中报告帧级 AUC。
对看不见的数据集的泛化能力。在现实世界中
在这种情况下,通常情况下,可疑图像是通过看不见的面部处理方法生成的,因此

方法	视频 1	视频 2	视频 3	平均	
多任务[43]	62.50	26.45	78.50	55.81	
DSP-FWA [34]	23.08	33.14	51.81	36.01	
Xception-c23 [51]	44.56	83.85	99.31	75.90	
FFD [20]	56.04	71.16	79.07	68.75	
面部 X 光[32]	66.49	94.03	87.10	82.54	
ICT（熊）	82.27	97.78	98.05	94.36	
ICT-Ref（我们的）	100.0	100.0	100.0	100.0	

表 2. 精心制作的视频上的 Deepfake 检测 AUC (%)
从互联网收集。

对未见数据的泛化能力非常重要
用于深度伪造检测。在表1 中,我们展示了检测
在五个不同数据集上的表现,其中闭集
结果（即,训练和测试在同一个数据集上）
用灰色标记。我们可以看到所有的比较
基于低级的基线在 Deeper 上表现不佳,
Celeb-DF v1 和 Celeb-DF v2 数据集,它们是新的
释放,因此显示较少的伪影。虽然我们的方法
ICT 得到显着改善,实现了高检测 AUC,表明身份信息更

在这种情况下,比低级纹理更可靠的证据。在
另一方面,我们的 ICT-Ref 变体进一步大大提高了
数量并达到最先进的性能
所有基准数据集。

图像退化的泛化能力。图片
退化在深度伪造图像和视频的传播中非常普遍。因此,我们进一步评估了对不同
类型图像退化的泛化能力。这里

我们遵循[28]中的图像退化策略。笔记
[28]中的“JPEG”确实是像素化,所以我们意识到
真正的 JPEG 作为额外的降级并设置五个严重级别
作为 jpeg 品质因数 90、70、50、30、20。

图 5 绘制了所有类型退化的不同程度的性能变化。它可以

很容易看出,当退化变得更严重时,所有比较的方法都会急剧下降,而我们的方法

仍然具有较高的检测性能。除了
对于高斯噪声,我们的方法也会降低
当度数变得太严重时。
真实场景模拟。我们进一步展示了一个
典型的真实世界 deepfake 检测示例,使用我们的
信息通信技术。真实的 deepfake 视频下载自
YouTube频道“Ctrl Shift Face2 ”,精心制作
精心制作,所有伪造的身份都是名人。随着

结果如表2 所示,我们发现大多数
现有的方法比较低。但是对于我们的 ICT,我们得到
高性能,平均接近 95%。这
参考辅助版本甚至将性能提升到
100%。视频版结果请参考https：
//www.youtube.com/watch?v=zgF50dcymj8.

2https://www . YouTube 。 com / 频道 / UCKpH0CKltc73e4wh0_pgL3g

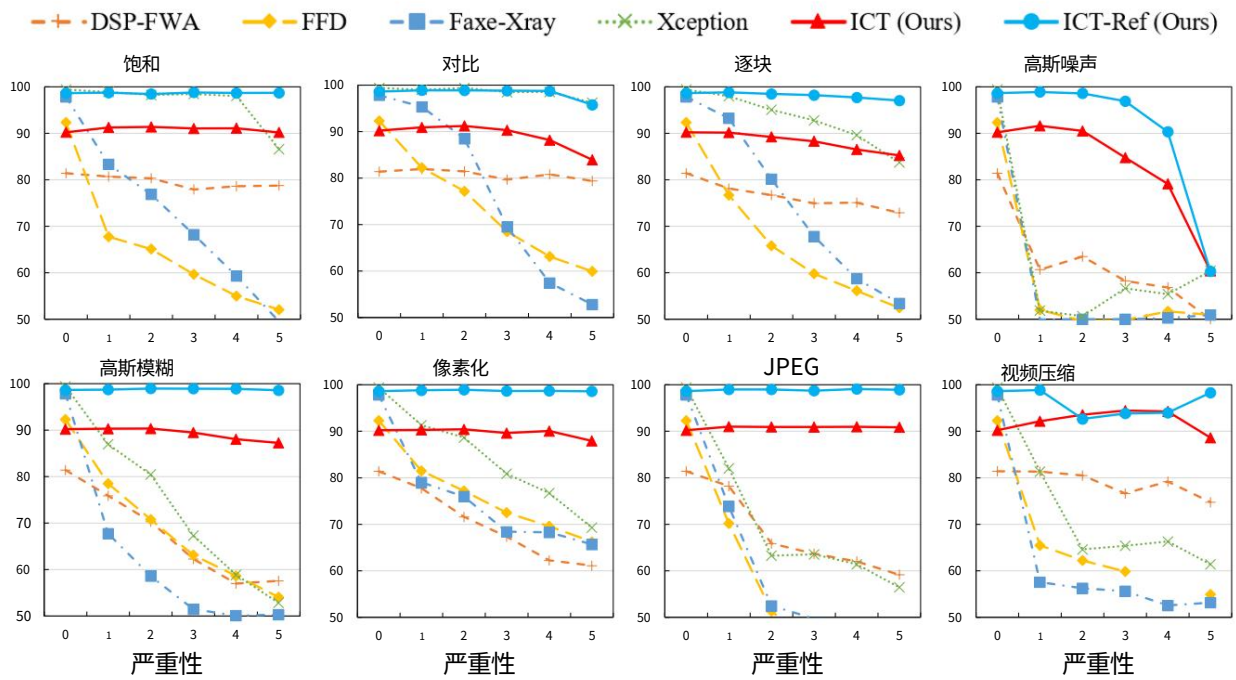

图 5. 对不同图像退化的泛化能力。在这里,我们报告关于五个严重性级别的帧级 AUC 分数对于每种降解形式。

	DFD FF++ 更深的 CD2		
无一致性损失 60.82 62.24 52.90 51.48			
无遮罩变形 83.59 83.47 79.06 83.44			
无色彩校正	80.43 90.11 92.15 82.86		
信息通信技术	84.13 90.22 93.57 85.71		

表 3 一致性损失、掩模变形和颜色分析
更正我们的身份一致性转换器。

4.2.信息通信技术分析

一致性损失的影响。在我们的框架中,我们引入了一个新的一致性损失来进一步拉动内部身份
当它们对应的标签相同时,和外部身份一起。在这里,我们尝试了另一种选择

对应物,即没有一致性损失的 ICT,并且存在
表3 中的比较。可以看出,我们的模型在没有一致性损失的情况下的性能大大下降
大约 23% - 40%,这验证了提议的一致性损失在我们的身份一致性中起着至关重要的作用
变压器。

不同一致性测量的影响。在这里,我们
研究方程式10中的每个组件并进行实验四
各种不同的身份一致性度量:(1)
建议的 ICT,即Din out; (2) 仅使用Din; (3) 仅
使用Dout; (4) 提议的ICT-Ref,即Dref。 这
四种不同测量结果的检测 AUC
一个示例数据集 Celeb-DeepFake v2 [35]为 85.71%,
分别为 92.52%、87.45% 和 94.43%。正如预期的那样,
参考辅助 ICT 达到最佳性能
通过结合其他三种类型。值得一提的是
即使只使用Din或Dout也能获得更好的性能

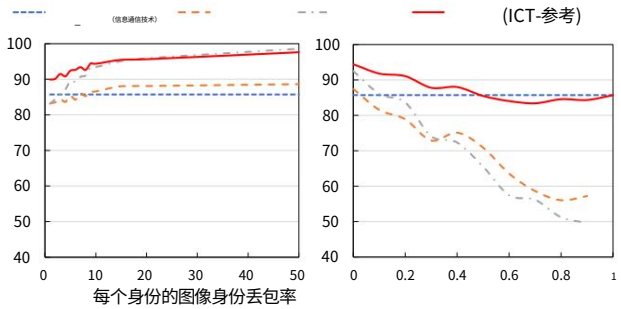

图 6. 说明参考集大小的影响
考虑改变每个身份的图像数量(左
图)和改变身份的数量(右图)。

比 ICT。这表明我们已经有效地做好了
使用免费提供的参考集,并验证了其在检测深度伪造中的关键作用。此外,令人惊讶地发现使用Din比使用

Dout,实现了约 5% 的增益。原因可能是这样
由于在人脸交换 ping 结果中通常不会操纵外人脸,因此使用外人身份来检索最近的人脸
邻居(其次是比较内部身份)更多
可能检索到准确的嫌疑人身份。
参考集的效果。我们进一步分析如何
参考集的大小会影响检测性能。
我们考虑影响参考大小的两个因素
集合:(1)为每个随机采样的图像数量
身份; (2)身份的数量。结果在
Celeb-DeepFake v2 [35]如图6 所示。它可以是
看到通常随着每个身份的图像数量增加,性能也会提高,直到饱和,

并且随着身份数量的减少,性能

模型	#参数	更深层次的	CD2
Res50	4379万	在	在
Res50-分割图像	43.79M	79.42	73.71
Res50-拆分模型	2×43.79M	91.48	84.66
信息通信技术	21.45M	93.57	85.71

表 4. 提议的基于 Transformer 的比较
基于 ICT 和 ConvNet 的 Res50。

也下降。有趣的是,我们发现在某些时候,例如,
每个身份的少量图像或更大的身份下降
率,性能甚至低于 ICT
不依赖于参考集。原因可能是这样
在这种情况下,参考集的大小是如此之小,以至于
检索到的身份不准确。请注意,较大的标识
掉率对性能的影响更大,甚至
使 ICT-Ref 低于 ICT。
ViT骨干的作用。在这里我们比较我们的 ViT
与传统 CNN 模型 ResNet50 [27] 的主干。我们
在最后一个之后添加两个单独的全连接层
7 × 7 特征图,以学习内部身份和
外部身份,最后将身份向量传递给
共享分类头（一个类似于我们的 fc）。
然而,我们发现这个 CNN 模型（记为 Res50）
无法收敛。原因可能是共享的骨干将内部身份和外部身份拉得太紧

当两个身份时,它无法捕捉到差异
有不同的标签。而对于我们的 ICT,它具有全球
每一层都有注意力机制,所以它可以很容易地分割内部和外部。我们进一步研究了两个替代模型:a) Res50-split-image,通过混合掩码分割图像,让模型从裁剪的内（外)脸中学习内（外)身份,而不是从整个图像中学习; b) Res50-split-model,使用两个

单独的模型来分别学习内部和外部身份。结果如表4 所示。我们可以看到

Res50-split-image 仍然表现不佳,Res50-split 模型的表现比我们的 ICT 稍差,成本
参数数量的 4 倍和两个独立的模型。
内在身份和外在身份的显着性图
身份。在这里,我们展示了内部身份和外部身份的注意力图,以查看面部的哪个部分对学习内部身份和外部身份的贡献最大

身份也是如此。我们可以看到,内在的身份主要是
专注于最具辨别力的内脸部分,即
由 deepfake 技术精心打造。相反,
外部身份集中在周边地区,例如
人脸轮廓,通常在 Deep Fake 生成期间保持不变。我们进一步测试LFW人脸识别
准确度并找到内部和外部令牌的准确度
高于 98%。注意区域之间的差异,内部身份和外部身份确实是

语义上有意义,因此有利于检查
身份不一致。
训练数据生成的效果。我们的方法是

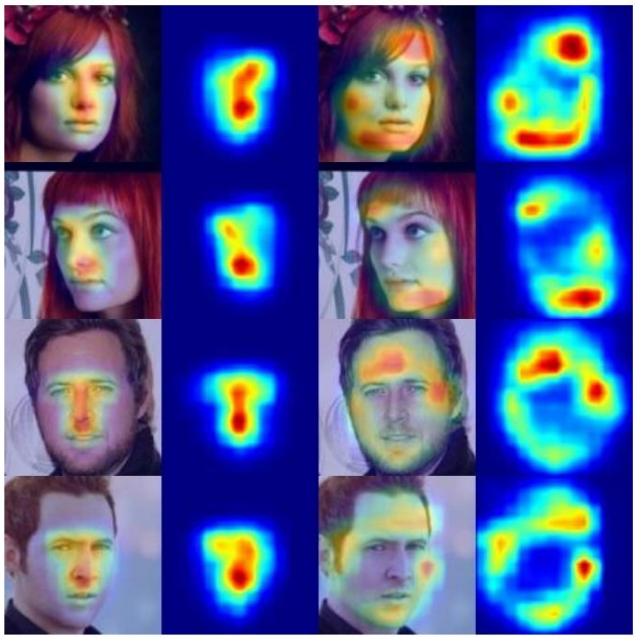

图 7. 内部身份的显着性图（第二列）
以及具有不同姿势的外部身份（第四列）。

一种无假的方法,只需要由真实人脸生成的交换图像进行训练。
在这里我们烧蚀（a)面具
变形旨在生成不同形状的面具
(b) 旨在产生更逼真的色彩校正
在训练数据生成过程中换脸。这
消融结果如表3 所示。可以看出,
我们的模型在没有掩模变形的情况下的性能下降
在某些数据集上略有影响,而在其他数据集上则非常重要,如
是颜色校正。这验证了这两种技术都是
对最终模型很重要且有帮助。

5. 结论

在这项工作中,我们提出了一种新的方法,身份一致性转换器,用于检测伪造的人脸图像。我们采用 Transformer 同时学习

引入了内部身份以及外部身份和一种新颖的一致性损失。我们表明我们的工作基于
on high-level semantics 对这种情况特别有效
基于低级的方法失败的地方。此外,我们的方法通过利用额外的身份得到进一步增强
来自名人的信息。广泛的实验已经
已进行以证明我们方法的有效性。希望我们的作品可以鼓励更多的作品

调查高级语义中的不一致
人脸伪造检测。
确认。 这项工作得到了支持
国家自然科学基金部分
根据授予 U20B2047、62072421、62002334 和
62121002,高校探索基金项目
中国科学技术部资助
YD3480002001,并获基本科研经费
授予WK2100000011下的中央大学。

参考

[1] https://github.com/iperov/DeepFaceLab。 1 [2] https://github.com/dfaker/df。 1 [3] https://www fakeapp/。 1 [4] https://github.com/deepfakes/faceswap。 1、 .不好的生活com/en/软/

2

[5] https://github. com/少安路/换脸-然而。 1

[6] https://ai.谷歌博客。 com / 2019 / 09 / 贡献 - 数据 - 到 - deepfake - detection.html。 5

[7] 达里乌斯·阿夫查尔、文森特·诺齐克、山岸纯一和越前勋。 Mesonet：一个紧凑的面部视频伪造检测网络，2018. 1, 2, 6

[8] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. Towards open-set identity preserving face synthesis, 2018. 1, 2

[9] Jawadul H Bappy、Cody Simons、Lakshmanan Nataraj、BS Manjunath 和 Amit K Roy-Chowdhury。用于检测图像伪造的混合 lstm 和编码器 - 解码器架构。 IEEE Transactions on Image Processing, 28(7):3286–3300, 2019. 1, 2

[10] Dmitri Bitouk、Neeraj Kumar、Samreen Dhillon、Peter Bel humeur 和 Shree K Nayar。人脸交换：自动替换照片中的人脸。在 ACM Transactions on Graphics (TOG)，第 27 卷，第 39 页。ACM， 2008.2 [11] Tom B Brown，Benjamin Mann，Nick Ryder，Melanie Sub biah，Jared Kaplan，Prafulla Dhariwal，Arvind Neelakan tan，Pranav Shyam， Girish Sastry、Amanda Askell 等人。

语言模型是少数人的学习者。 arXiv 预印本 arXiv:2005.14165, 2020. 3 [12] 曹晨, 翁彦林, 周舜, 童一英, 周昆。 Facewarehouse：用于视觉计算的 3d 面部表情数据库。 IEEE Transactions on Visualization and Computer Graphics, 20(3):413–425, 2013. 2 [13] Lucy Chai、David Bau、Ser-Nam Lim 和 Phillip Isola。

是什么让假图像可以检测到?理解概括的属性，2020. 6

[14] Mark Chen、Alec Radford、Rewon Child、Jeffrey Wu、Hee woo Jun、David Luan 和 Ilya Sutskever。从像素生成预训练。在机器学习国际会议上，第 1691-1703 页。 PMLR, 2020. 3

[15] Yinpeng Chen, Xiyang Dai, Dongdong Chen, Mengchen Liu, Xiaoyi Dong, Lu Yuan, and Zicheng Liu. Mobile former: Bridging mobilenet and transformer. arXiv preprint arXiv:2108.05895, 2021. 3 [16] Yi-Ting Cheng, Virginia Tzeng, Yu Liang, Chuan-Chang Wang, Bing-Yu Chen, Yung-Yu Chuang, and Ming Ouhy oung. 3d-model-based face replacement in video. In SIG GRAPH 09: Posters, page 29. ACM, 2009. 2

[17] 乔尼·亚当森·克拉克。将过去带入生活，1996 年。 2 [18] Davide Cozzolino、Andreas Rossler、Justus Thies、Matthias Nießner 和 Luisa Verdoliva。 Id-reveal：身份感知

deepfake 视频检测。 arXiv 预印本 arXiv:2012.02512, 2020. 3

[19] Kevin Dale、Kalyan Sunkavalli、Micah K Johnson、Daniel Vlasic、Wojciech Matusik 和 Hanspeter Pfister。视频人脸替换。在 ACM 图形事务 (TOG)，第 30 卷，第 130 页。ACM， 2011。2

[20] Hao Dang、Feng Liu、Joel Stehouwer、Xiaoming Liu 和 Anil K Jain。关于数字人脸操纵的检测。
在 IEEE/CVF 计算机视觉和模式识别会议论文集中，第 5781-5790 页，2020 年。6 [21] 邓建康、郭佳、薛念南和 Stefanos Zafeiriou。 Arcface：深度人脸识别的加性角边距损失，2019年。2、4 [22] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。

Bert：用于语言理解的深度双向转换器的预训练。 arXiv 预印本 arXiv:1810.04805, 2018. 3, 4

[23] 董晓义，鲍建民，陈冬冬，张伟明，于能海，陆源，陈冬，郭白宁。 Cswin 变压器：具有十字形窗口的通用视觉变压器主干。 arXiv 预印本 arXiv:2107.00652, 2021. 3 [24] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Syl vain Gelly 等。一张图像值 16x16 字：用于大规模图像识别的变压器。 arXiv 预印本 arXiv:2010.11929, 2020. 3, 4

[25] 耿正林、曹陈、谢尔盖·图利亚科夫。 3d 引导的细粒度面部操作。在 IEEE/CVF 计算机视觉和模式识别会议论文集上，第 9821-9830 页，2019年。2

[26] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In European conference on computer vision, pages 87–102. Springer, 2016. 4, 5 [27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.

用于图像识别的深度残差学习。在 IEEE 计算机视觉和模式识别会议论文集，第 770-778 页，2016 年。8 [28] Liming Jiang、Ren Li、Wayne Wu、Chen Qian 和 Chen Change Loy。 Deeperforensics-1.0：用于真实世界人脸伪造检测的大规模数据集。在 IEEE/CVF 计算机视觉和模式识别会议论文集中，第 2889-2898 页，2020 年。5、6

[29] Iryna Korshunova、Wenzhe Shi、Joni Dambre 和 Lucas Theis。使用卷积神经网络的快速换脸。
在 IEEE 计算机视觉国际会议论文集上，第 3677-3685 页，2017 年。2 [30] Trung-Nghia Le、Huy H. Nguyen、Junichi Yamagishi 和 Isao Echizen。Openforensics：用于野外多面伪造检测和分割的大规模挑战性数据集。在 IEEE/CVF 计算机视觉国际会议 (ICCV) 论文集上，第 10117-10127 页，2021 年 10 月。2 [31] Lingzhi Li、Jianmin Bao、Hao Yang、Dong Chen 和 Fang Wen。 Faceshifter：实现高保真和遮挡感知的面部交换。 arXiv 预印本 arXiv:1912.13457, 2019. 1, 2

[32] Lingzhi Li, Jianmin Bao, Ting Zhang, Hao Yang, Dong Chen, Fang Wen, and Baining Guo. Face x-ray for more 一般人脸伪造检测， 2020年。1、2、4、5、6

[33] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In ictu oculi:通过检测暴露 AI 生成的假人脸视频 眨眼,2018. 2

[34] 李跃尊、吕四维 .通过检测面部扭曲伪影来曝光 deepfake 视频,2019 . 1, 2, 6

[35] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei 吕。 Celeb-df:用于深度伪造取证的大规模挑战性数据集,2020. 5, 7

[36] Yuan Lin, Shengjin Wang, Qian Lin, and Feng Tang. Face 在大的姿势变化下交换:基于 3d 模型的方法。在 2012 年 IEEE 多媒体和博览会国际会议上,第 333-338 页。 IEEE, 2012. 2

[37] Yaqi Liu, Qingxiao Guan, Xianfeng Zhao, and Yun Cao. Im age forgery localization based on multi-scale convolutional 神经网络。第六届 ACM 研讨会论文集 关于信息隐藏和多媒体安全,第 85 页– 90, 2018. 1, 2

[38] Iacopo Masi、Aditya Killekar、Royston Marian Mascaren has、Shenoy Pratik Gurudatt 和 Wael AbdAlmageed。用于隔离视频中的 deepfakes 的两个分支循环网络。 在欧洲计算机视觉会议上,第 667 页– 684. 施普林格,2020. 6

[39] Iacopo Masi、Anh Tuan Tran、Tal Hassner、Jatuporn Toy Lek sut 和 Gerard Medioni。我们真的需要收集数百万张脸来进行有效的人脸识别吗?在欧洲 计算机视觉会议,第 579-596 页。施普林格， 2016. 2

[40] Iacopo Masi、Anh Tuan Tran、Tal Hassner、Gozde Sahin 和 杰拉德·梅多尼。用于不受约束的人脸识别的人脸特定数据增强。国际计算机杂志 愿景, 127(6):642–667, 2019. 2

[41] 法尔科·马特恩、克里斯蒂安·里斯和马克·斯塔明格。利用视觉伪影来暴露深度伪造和面部操纵。 2019年IEEE计算机冬季应用 视觉研讨会 (WACVW),第 83-92 页。 IEEE, 2019. 1, 2

[42] Ryota Natsume、Tatsuya Yatagawa 和 Shigeo Morishima。 Rsgan:在潜在空间中使用面部和头发表示进行面部交换和编辑。 arXiv 预印本 arXiv:1804.03447, 2018. 2

[43] Huy H. Nguyen, Fuming Fang, Junichi Yamagishi, and Isao 越前。用于检测和分割的多任务学习 操纵面部图像和视频,2019 . 1, 2, 6

[44] Huy H. Nguyen、Junichi Yamagishi 和 Isao Echizen。 用于检测虚假图像和视频的胶囊网络,2019 年。 1,2,6 _ _

[45] 尤瓦尔·尼尔金、约西·凯勒和塔尔·哈斯纳。 Fsgan:主题 不可知论者换脸和重演,2019 . 1, 2

[46] Yuval Nirkin、Iacopo Masi、Anh Tran Tuan、Tal Hassner 和 杰拉德·梅多尼。关于人脸分割、人脸交换和 面对感知。 2018年第13届IEEE国际会议 关于自动面部和手势识别（FG 2018），页面 98-105。 IEEE, 2018. 1, 2

[47] Yuval Nirkin、Lior Wolf、Yosi Keller 和 Tal Hassner。基于人脸差异的深度造假检测 它的上下文。 arXiv 预印本 arXiv:2008.12262, 2020. 3, 6

[48] Yuyang Qian, Guojun Yin, Lu Sheng, Zixuan Chen, and Jing 绍。频率思考:通过挖掘频率感知线索进行人脸伪造检测,2020 . 1, 2

[49] 亚历克·拉德福、卡提克·纳拉辛汉、蒂姆·萨利曼斯和伊利亚 苏斯克维尔。通过生成提高语言理解 预训练。 2018. 3

[50] 亚历克·雷德福、杰弗里·吴、瑞文·柴尔德、大卫·栾、达里奥 Amodei 和 Ilya Sutskever。语言模型是不受监督的多任务学习者。 OpenAI 博客,1(8):9,2019. 3

[51] Andreas Rossler、Davide Cozzolino、Luisa Verdoliva、Christian Riess、Justus Thies 和 Matthias Nießner。 Faceforen sics++:学习检测被操纵的面部图像,2019。 1,2,5,6 _ _ _

[52] Supasorn Suwajanakorn、Steven M Seitz 和 Ira 凯梅尔马赫-施利泽曼。合成奥巴马:从音频中学习口型同步。 ACM 图形事务 (ToG), 36(4):1-13, 2017. 2 ..

[53] Justus Thies、Michael Zollhofer 和 Matthias Nießner。延迟神经渲染：使用神经纹理的图像合成。 ACM 图形交易 (TOG),38(4):1-12， 2019. 5

[54] Justus Thies、Michael Zollhofer、Marc Stamminger、Christian Theobalt 和 Matthias Nießner。 Face2face:实时 人脸捕捉和 RGB 视频的重演。在 IEEE 计算机视觉和模式会议论文集上 识别,第 2387-2395 页,2016. 2, 5

[55] Ashish Vaswani、Noam Shazeer、Nicky Parmar、Jacob Uszko Reit、Llion Jones、Aidan N Gomez、Lukasz Kaiser 和 Illia Polosukhin。注意力就是你所需要的。 arXiv 预印本 arXiv:1706.03762, 2017. 3, 4

[56] Ziyu Wan, Jingbo Zhang, Dongdong Chen, and Jing Liao. 使用变换器的高保真多元图像补全。在ICCV,2021. 3

[57] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: 深度人脸识别的大余量余弦损失,2018年。2

[58] 王红霞,潘春红,龚海峰,吴怀宇。基于主动外观模型的人脸图像合成。 2008年IEEE国际会议 声学、语音和信号处理,第 893-896 页。 IEEE, 2008. 2

[59] Rui Wang, Dongdong Chen, Zuxuan Wu, Yinpeng Chen, Xiyang Dai, Mengchen Liu, Yu-Gang Jiang, Luowei Zhou, 还有陆远。 Bevt:视频转换器的 Bert 预训练。 arXiv 预印本 arXiv:2112.01529, 2021. 3

[60] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew 欧文斯和阿列克谢 A Efros。 Cnn 生成的图像出奇地容易被发现……目前。在 IEEE 会议记录中 计算机视觉和模式识别会议， 2020 年第 7 卷。6

[61] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing deep fakes 使用不一致的头部姿势,2018. 2

[62] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, 张伟明,于能海。多注意力深度伪造检测。在 IEEE/CVF 会议论文集中 关于计算机视觉和模式识别,2021. 2

[63] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuan jun Xiong, and Wei Xia. Learning to recognize patch wise consistency for deepfake detection. arXiv preprint arXiv:2012.09311, 2020. 6

[64] Yinglin Zheng, Jianmin Bao, Dong Chen, Ming Zeng, and 方文。探索时间相干性以进行更通用的视频人脸伪造检测。在诉讼中

IEEE/CVF 计算机视觉国际会议 (ICCV),第 15044-15054 页,10 月

[65] Peng Zhou, Xintong Han, Vlad I. Morariu, and Larry S. 戴维斯。学习图像处理检测的丰富特征,2018. 1, 2

[66] Peng Zhou, Xintong Han, Vlad I. Morariu, and Larry S. 戴维斯。用于篡改人脸检测的双流神经网络,2018. 1, 2