

## DeepFaceLab。综合、灵活、可扩展的换脸框架

Ivan Petrov

自由职业者

lepersorium@gmail.com

高岱恒

自由职业者

samuel.gao023@gmail.com

Nikolay Chervoniy

自由职业者

n.chervonij@gmail.com

Kunlin Liu<sup>†</sup>

中国科学技术大学

lkl6949@mail.ustc.edu.cn

Sugasa Marangonda

自由职业者

thedeepfakechannel@gmail.com

Chris Ume<sup>†</sup>

视觉特效 Chris

Ume

info@vfxchrisume.com

蒋健008

技术

jiangjian@008tech.com

路易斯-

RP

自由职业者

luisguans@hotmail.com

张胜

自由职业者

cndeepfakes@gmail.com

吴平宇

自由职业者

wpydcr@hotmail.com

张伟明

中国科学技术大学

zhangwm@ustc.edu.cn

### 摘要

深度伪造的防御不仅需要研究detection，而且还需要努力生成方法。然而，目前的深度造假方法存在着工作流程不安全和性能差的问题。为了解决这个问题，我们提出了DeepFaceLab，这是目前主流的深度伪造换脸框架。它提供了必要的工具以及易于使用的方式进行高质量的换脸。它还提供了一个灵活和松散的耦合结构，供那些需要用其他功

能来加强他们的管道而不需要编写复杂的模板代码的人使用。我们详细介绍了驱动DeepFaceLab实现的原则，并介绍了它的管道，通过它，用户可以无痛地修改管道的每个方面，以达到他们的定制目的。值得注意的是，DeepFaceLab可以实现具有高保真度的电影质量结果。我们通过将我们的方法与其他系统进行比较来证明我们系统的优势。

换脸的方法。<sup>1</sup>

## 1. 简介

自从近年来深度学习增强了计算机视觉领域的能力后，操纵数字图像，尤其是操纵人像图像的能力得到了迅速提高，并在大多数情况下达到了逼真的效果。脸部互换是一项引人注目的任务，通过将源脸转移到目的地，同时保持目的地的面部动作和表情变形来生成虚假内容。

人脸操作技术的基本动机是生成对抗网络（GANs）[7]。越来越多的由StyleGAN[10]、StyleGAN2[11]合成的人脸变得越来越逼真，并使

---

<sup>1</sup>对于 更多 信息。 请访问  
访问：<https://github.com/iperov/DeepFaceLab/>。（刘  
昆林是相关作者。）



图1.DeepFaceLab产生的面部交换结果。左：源脸。中间。替换后的目标脸。我们的结果出现在右边，表明DeepFaceLab可以高保真地处理遮挡、不良光照和侧脸。

对于人类的视觉系统来说，这一点是无法区分的。

众多由基于GAN的换脸方法合成的恶搞视频被发布在YouTube和其他视频网站上。商业移动应用程序，如ZAO<sup>2</sup>和FaceApp<sup>3</sup>让普通网民可以毫不费力地创建假图像和视频，这大大促进了这些被称为深度伪造的交换技术的传播。

这些内容生成和修改技术可能会影响公共话语的质量，并侵犯公民的肖像权，特别是考虑到deepfake可能被恶意地用作错误信息、操纵、骚扰和说服的来源。识别被操纵的媒体是一个技术上要求很高且快速发展的挑战，需要整个科技行业和其他行业的合作。

媒体防伪检测的研究正在被激发，并在伪造脸部检测方面投入了越来越多的精力。DFDC<sup>4</sup>是一个典型的例子，它是由Facebook和微软发起的一项价值百万美元的竞赛。训练强大的伪造检测模型需要高质量的伪造数据。我们的方法所产生的数据涉及DFDC数据集[5]。

然而，被攻击后的检测并不是减少深层伪造的恶意影响的唯一方式。发现传播欺骗性的内容总是太晚了。在我们看来，对于学术界和公众来说，帮助网民了解什么是深度伪造，以及电影质量的交换视频是如何产生的要好得多。俗话说："最好的防守就是好的进攻"。让广大网民认识到深度伪造的存在，加强他们对社交网络上发布的恶搞信息的识别能力，比纠结于恶搞媒体的真假更为关键。

<sup>2</sup> <https://apps.apple.com/cn/app/id1465199127>

<sup>3</sup> <https://apps.apple.com/gb/app/faceapp-ai-face-editor/id1180884341>

<sup>4</sup> <https://deepfakedetectionchallenge.ai/>

2018年，DeepFakes[3]引入了一个完整的生产管道，将源人的脸与目标人的脸以及相同的面部表情（如眼球运动、面部肌肉运动）替换在一起。然而，DeepFakes产生的结果在某种程度上是差的，当代Nirkin等人的自动换脸的结果也是如此[15]。为了进一步唤醒人们对人脸操纵视频的认识，为造假检测研究者提供便利，我们建立了一个开源的deepfake项目--

DeepfaceLab（简称DFL），用来建立巨大的高质量的人脸互换视频，供人们娱乐，并通过提供高质量的造假数据，极大地帮助造假检测的发展。

本文介绍了DeepFaceLab，这是一个集成的开源系统，其流水线采用清洁状态设计，无需痛苦的调整即可实现逼真的换脸效果。事实证明，DFL在公众中非常受欢迎。例如，许多艺术家创建了基于DFL的视频，并在他们的YouTube频道上发布。这些由DFL制作的视频有超过1亿的点击率。

DeepFaceLab的贡献可以概括为三个方面。

- 提出了一个由成熟的管道组成的最先进的框架，旨在实现逼真的换脸结果。
- DeepFaceLab在2018年将代码开源，并以各种方式紧跟计算机视觉领域的进展，为捍卫deepfake做出了积极贡献，引起了开源社区和VFX领域的广泛关注。
- DeepFaceLab中引入了一系列高效的组件和工具，以建立更好的换脸视频。

## 2. DeepFaceLab的特点

DeepFaceLab的成功源于将以前的想法编织成一个设计，以平衡速度和易用性以及计算机视觉在人脸识别、对齐、重建、分割等方面的蓬勃发展。我们的实现背后有四个主要特点。

### 便利性

DFL努力使其管道的使用，包括数据加载器和处理、模型训练和后处理，尽可能的简单和富有成效。与其他人脸交换系统不同的是，DFL提供了一个完整的命令行工具，包括管道的每个方面，可以按照用户选择的方式实施。值得注意的是，固有的复杂性和许多手工挑选的细粒度控制的功能，如用于脸部对齐的典型脸部地标，应该在内部处理，并隐藏在DFL后面。如果人们按照工作流程的设置，不需要手工挑选的特征就可以达到流畅和逼真的换脸效果，但只需要两个视频：源视频（src）和目标视频（dst），而不需要在src和dst之间配对相同的面部表情。在某种程度上，DFL可以作为一个点对点的摄像机来使用。

### 广泛的工程支持

为了提高性能，增加了一些实用的措施：多GPU支持，半精度训练，使用钉子式CUDA内存来提高吞吐量，使用多线程来加速图形操作和数据处理。即使是一台拥有2GB VRAM的机器也可以成功地进行换脸项目。

### 可扩展性

为了加强DFL工作流程的灵活性，吸引研究界的兴趣，用户可以自由地替换DFL中不符合他们要求的任何组件。DFL的大多数模块都被设计成可以互换的。例如，人们可以提供一个更新的人脸检测器，以便在检测极端角度或外围区域的人脸时获得更高的性能。

### 可扩展性

拥有良好的数据集对换脸任务至关重要。一般来说，数据集越大，最终的结果就越好。然而，直接从src和dst中提取的结果总是带有噪音，这对最终的质量有很大影响。考虑到输入视频的复杂情况，DFL提供了一系列的措施来清理数据集。有了这些措施，DFL就有了很好的扩展性，甚至可以支持大规模的数据集，并在大型数据集的基础上进行电影级的换脸。

## 3. 管道

DeepFaceLab提供了一套工作流程，形成了灵活的管道。在DeepFaceLab（简称DFL）中，我们可以将流水线抽象为三个阶段：提取、训练和转换。这三个部分将按顺序介绍。此外，值得注意的是，DFL属于典型的一对一换脸范式，这意味着只有两种数据：src和dst，即源和目的地的缩写，在下面的叙述中使用。

### 3.1. 萃取

提取阶段是DFL的第一个阶段，目的是从src和dst数据中提取人脸。这个阶段由许多算法和处理部分组成，即人脸识别、人脸对齐和人脸分割。DFL提供了许多提取模式（即半脸、全脸、全脸），这代表了提取阶段的人脸覆盖区域。一般来说，我们默认采用全脸模式。

### 人脸检测

提取阶段的第一步是在给定的数据中找到目标人脸：src和dst。DFL将S3FD[24]作为其默认的人脸检测器。S3FD可以不费吹灰之力地与其他人脸检测算法（如RetinaFace[4]）重新组合。

### 面部对齐

第二步是面部对齐。经过无数次的实验和失败，我们意识到，面部地标是保持长期稳定的关键。我们需要找到一种有效的面部地标算法，这对制作优秀的连续镜头和电影至关重要。

DFL提供了两种典型的面部地标提取算法来解决这个问题：（a）基于热图的面部地标算法2DFAN[2]（用于具有标准姿势的人脸）；（b）具有三维面部先验信息的PRNet[6]（用于具有大欧拉角（偏航、俯仰、滚动）的人脸，例如，具有大偏航角的人脸，意味着脸的一侧是看不见的）。在检索出面部地标后，我们还提供了一个可配置时间步长的可选功能，以平滑单次拍摄中连续帧的面部地标，进一步确保稳定性。

然后，我们采用Umeyama[21]提出的经典的点模式映射和转换方法来计算用于脸部对齐的相似性转换矩阵。由于Umeyama等人[21]提出的方法在计算相似性转换矩阵时需要标准的面部地标模板，DFL提供了一个经典的对齐的面部地标模板。值得注意的是，DFL可以通过使用获得的面部地标自动预测欧拉角。



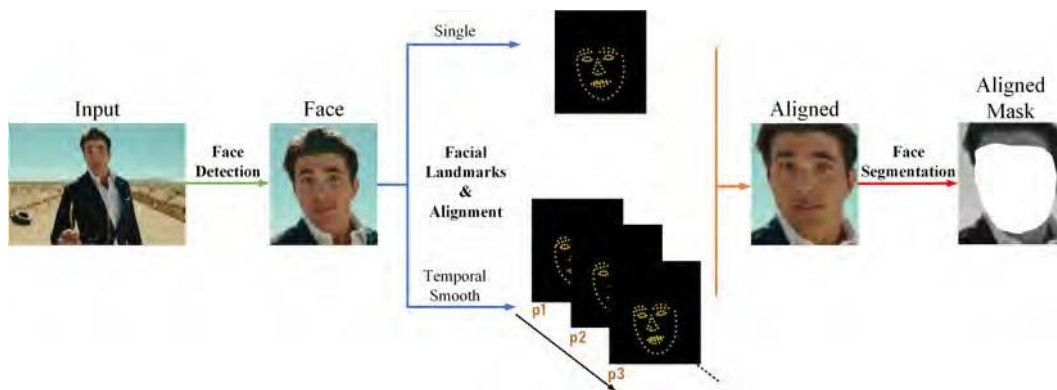
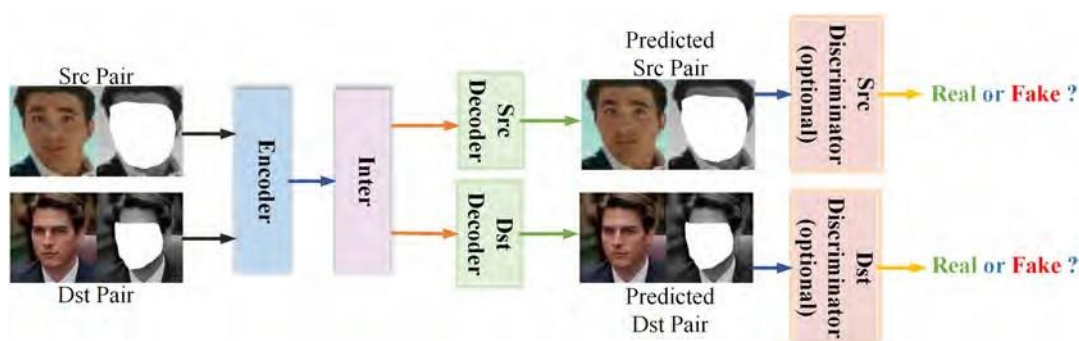


图2.DeepFaceLab（简称DFL）的提取阶段概述。

[DF结构]



[LIAE结构]

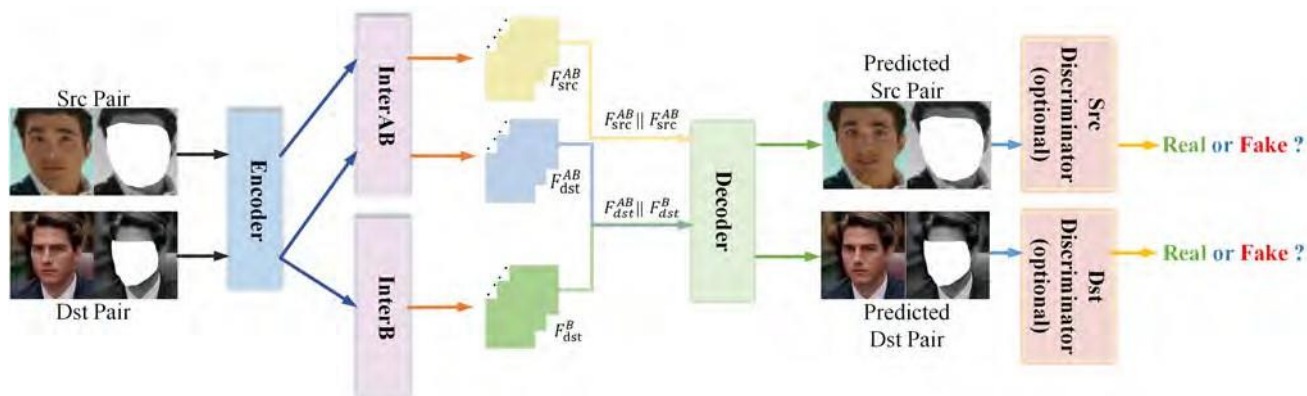


图3.DeepFaceLab（DFL）的训练阶段概述。这里提供的DF结构和LIAE结构都是为了说明问题，。表示潜伏向量的连接。

### 脸部分割

在脸部对齐后，我们得到了一个带有标准正面/侧面视图（对齐的src或对齐的dst）的数据文件夹。我们在（对齐的src或对齐的dst）的基础上采用精细的人脸分割网络（TernausNet[8]），通过它可以准确地分割出带有头发、手指或眼镜的人脸。在训练过程中，去除不规则的遮挡物以保持网络对手、眼镜和任何其他可能覆盖脸部的物体的稳健性，是一种操作性强但有用的方法。

然而，由于一些最先进的人脸分割模型不能在某些特定的镜头中产生细粒度的掩码，DFL引入了XSeg。XSeg允许每个人通过几张照片的学习模式来训练他们的模型来分割一个特殊的人脸集（对齐的src或对齐的dst）（图8是XSeg的示意图）。

由于上述工作流程是按顺序执行的，DFL在训练阶段所需要的每样东西都已经准备好了：裁剪好的面孔及其相应的坐标在其

原始图像、面部标志、对齐的人脸以及来自src的像素分割掩码（由于dst的提取过程与src相同，因此没有必要详细说明）。

### 3.2. 培训

训练阶段对实现以下目标起着最重要的作用  
DFL的逼真的脸部交换结果。

由于不需要对准src和对准dst的面部表情进行严格的匹配，DFL旨在提供一种有效的算法范式来解决这种不配对的问题，同时保持生成的面部的高保真和感知质量。在所述方法的启发下，我们提出了两种结构，即DF结构和LIAE结构，来解决这个问题。

如图3(a)所示，DF结构包括一个编码器以及src和dst之间共享权重的Inter，两个分别属于src和dst的解码器。src和dst的泛化是通过共享的Encoder和Inter来实现的，这就很容易解决了前面提到的不配对问题。DF结构可以完成换脸的任务，但不能从dst继承足够的信息，比如灯光。

为了进一步加强光一致性，我们提出了LIAE。如图3(b)所示，LIAE结构是一个更复杂的结构，有一个共享重量的编码器、解码器和两个独立的Inter。与DF相比，主要区别在于InterAB用于生成src和dst的潜伏码，而InterB只输出dst的潜伏码。这里， $F^{AB}$

表示由InterAB产生的src的潜伏码，我们把这种表示方法推广到 $F^{AB}$ ， $F^B$ 。在得到所有的潜伏代码后，从

dst dst  
InterAB和InterB，然后LIAE将这些特征连接起来通过通道映射。 $F^{AB} || F^{AB}$ ，得到一个新的dst dst  
src和F的潜伏代码表示 $F^{AB} || F^B$ ，对dst来说是

同样的方式

。那么 $F^{AB} || F^{AB}$ 和 $F^{AB} || F^B$ 被放进

解码器，我们得到预测的src(dst)和几乎目前最先进的换脸工作，更多的是他们的面具。将F连接起来的动机是 $F^{AB}$ 与 $F^{AB}_{dst}$ 是将潜伏代码的方向转向我们需要的类别（src或dst），通过这种方式，InterAB获得了潜伏空间中src和dst的紧凑和对齐的表示。

除了模型的结构外，一些有用的技巧对提高生成的人脸的质量是有效的。一般SSIM[22]中的加权和掩码损失，可以在AE训练架构下使脸部的每个部分都带有不同的权重，例如，我们给眼睛区域添加的权重比脸颊多，目的是使网络集中生成一张眼睛生动的脸。

至于损失，DFL默认使用混合损失（DSSIM（结构不相似性）[14]+MSE）。其动机是

这种组合是为了从两者中获得好处。DSSIM更快地生成人脸，而MSE提供更好的清晰度。这种损失的组合是为了在概括性和清晰度之间找到一个折中的办法。

此外，我们还采用了一种花哨的真面目模式TrueFace，用于在转换阶段生成与dst更相似的脸。对于LIAE结构，我们强制要求 $F^{AB}$ 接近 $F^{AB}$ 。而对于DF结构， $dst$ 则要求逆向思维。

部分变成了 $F_{src}$ 和 $F_{dst}$ 。两个很少使用的方法已经被DFL验证了。卷积感知初始化[1]和学习率下降[13]，它们大大增强了假脸的最终质量。

### 3.3. 转换

转换阶段是最后但也是最重要的阶段。以前的方法往往忽略了这个阶段的重要性。如图4所示，用户可以将src的面孔交换到dst，反之亦然。

在src2dst的情况下，由于Umeyama[21]的可逆性，所提出的人脸互换方案在转换阶段的第一步是将生成的人脸连同其掩码从dst解码器转换到src中目标图像的原始位置。

下面这块是关于混合的，野心是让重新调整后的重现脸沿其外轮廓与目标图像无缝贴合。为了保持一致的肤色，DFL又提供了五种颜色转移算法（即Reinhard颜色转移。RCT[18]，迭代分布转移。IDT[17]等），以使重现的脸部颜色与目标相近。任何混合都必须考虑到不同的肤色、脸型和光照条件，特别是在与目标的交界处。

在重演的面部与划定的区域和目标之间的关系面。DFL通过Poisson混合实现了这一点[16]。

最后，锐化是不可缺少的。一个预先训练好的脸

超分辨率神经网络被添加到锐化的

融合后的面孔，因为我们注意到，生成的面孔在

或更少，是平滑的，缺乏微小的细节（即痣。皱纹）。

## 4. 评价

在这一节中，我们将DeepFaceLab与其他一些常用的换脸框架和两个最先进的作品进行比较。我们发现，在相同的实验条件下，DFL在这些框架中具有竞争性的表现。

### 4.1. 定性结果

图5(a)提供了来自FaceForensics++数据集[19]的代表性开源项目（DeepFakes[3]、Nirkin等人[15]和Face2Face[20]）的换脸结果。

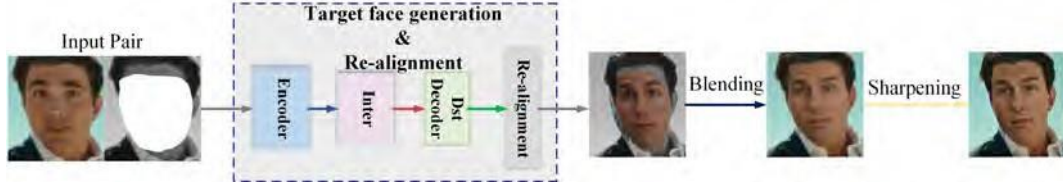


图4.DeepFaceLab(DFL)的转换阶段概述。

在我们的实验中，选择了不同的表情、脸型和插图  
的例子。通过观察FaceForensics++的视频片段，可以清  
楚地看到它们不仅训练不足，而且是从低分辨率的模  
型中选择的。为了公平起见，我们采取了Quick96模式  
：一个轻量级的模型，它的结构是不需要的。  
derneath，输出96×96分辨率的I输出（没有GAN和True  
Face）。平均训练时间被限制在3小时内。我们使用Ad  
am优化器（lr=0.00005， $\beta_1 = 0.5$ ， $\beta_2 = 0.999$ ）  
来优化我们的模型。我们所有的网络都是在一  
台NVIDIA GeForce GTX 1080Ti GPU和Intel Core i7-8700 CPU。

## 4.2. 量化结果

我们在定量实验中把我们的结果与FaceForensics++  
的视频进行比较。在实践中，换脸方法的结果的自然  
性和真实性很难用一些特定的量化指标来描述。然而  
，姿势和表情确实体现了换脸结果的宝贵内涵。此外  
，SSIM被用来比较结构相似性，而感知损失[9]被用来  
比较目标主体和被换主体之间的高层次差异，如内容  
和风格的差异。

为了衡量姿势的准确性，我们计算 $I_t$ 和 $I$   
output的欧拉角（通过FSA-  
Net[23]提取）之间的欧几里得距离。此外，面部表情  
的准确性是通过二维地标之间的欧氏距离（2DFAN[2]  
）来衡量的。我们使用DLIB[12]的默认人脸验证方法  
来进行身份比较。

为了具有统计学意义，我们在FaceForen-  
sics++中对前500个视频的100帧（随时间均匀采样）计  
算这些测量值的平均值和方差，在整个视频中取平均  
值。这里，Deep-  
Fakes[3]和Nirkin等人[15]被选为基线进行比较。需要  
注意的是，所有由DFL制作的视频都是按照4.1的相同  
设置进行的。

从表1中列出的指标来看，DFL比基线更善于保留姿  
势和表情。此外，由于超分辨率在转换阶段的授权，  
DFL经常产生生动的眼睛和尖锐的牙齿的输出，但这  
种现象不能清楚地反映在类似SSIM的分数中，因为它们  
只占很小的一部分。

整张脸的。

## 4.3. 消融研究

为了比较不同的模型选择、GAN设置等的视觉效果  
，我们进行了几次消融测试。消融研究是在三个基本  
部分之上进行的：网络结构、训练范式和潜空间约束  
。

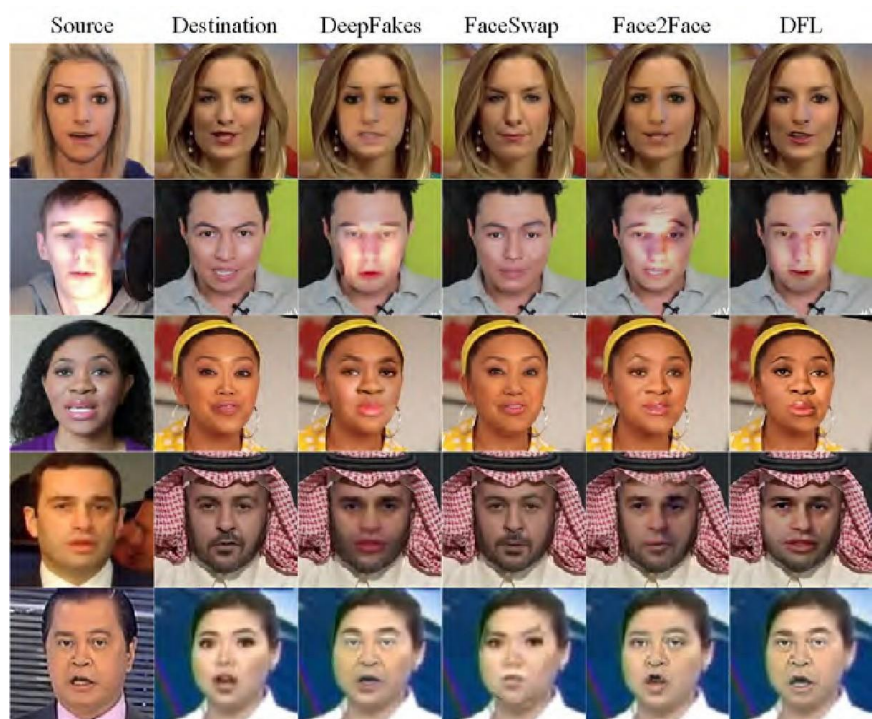
除了DF结构和LIAE结构外，我们还通过增加比原  
始版本更多的fea-  
ture提取层和残余块，将它们增强为DFHD和LIAEHD  
，从而丰富了模型结构的对比性。相关细节在补充材  
料中展示。不同模型结构的定性结果见图6，不同训练  
范式的定性结果见图7。如图6所示，我们可以看到LI  
AE可以从dst继承一个形状良好的脸部形状，并产生比  
DF更先进的re-  
sults，它优雅地解决了不匹配的脸部形状问题。此外  
，为了探究GAN的引入在DFL中是否有效，我们在图7  
中比较了基于GAN和非基于GAN的人脸，很明显，人  
脸的细节比非GAN生成的更真实，更有块状。

定量消融结果见表2。训练的实验设置几乎与  
4.1除了模型的结构。

表2的验证结果表明，在具有相同结构的网络中，源  
的识别性被保留下来。随着更多的捷径连接被添加到  
模型中（即DF到DFHD，LIAE到LIAEHD），在没有  
GAN的情况下，地标和姿势的分数下降。同时，生成  
的重新结果可以有更好的机会摆脱源脸的影响。

此外，我们还发现TrueFace有效地缓解了GAN的不  
稳定性，通过它可以实现更逼真的照片效果，而不会  
有太大的退化。此外，SSIM随着更多的捷径连接而逐  
渐增加，TrueFace和GAN在不同的程度上也做得不错。





(a) DFL和有代表性的开源换脸项目的比较。



(b) DFL和最新的最先进的换脸工程的比较。

图5.FaceForensics++[19]人脸图像的定性换脸结果。

## 5. 讨论

### 5.1. 诚信

以前的方法往往缺乏完整性。如第3节所述，DFL由三个主要阶段组成，即提取、训练-----。

摄取，和转换。每个阶段都发挥着不同的作用，有各种不同的替代技术。由于长期的发展，DFL已经成为世界上最成熟的换脸系统。例如，我们提供了几种人脸分割的方法。它是



表1.FaceForensics++[19]人脸图像的定量换脸结果。

方法	SSIM $\uparrow$	感知性损失 $\downarrow$	验证 $\downarrow$	地标 $\downarrow$	姿势 $\downarrow$
深度造假	$0.71 \pm 0.07$	$0.41 \pm 0.05$	$0.69 \pm 0.04$	$1.15 \pm 1.10$	$4.75 \pm 1.73$
Nirkin等人	$0.65 \pm 0.08$	$0.50 \pm 0.08$	$0.66 \pm 0.05$	<b><math>0.35 \pm 0.18</math></b>	$6.01 \pm 3.21$
DFL(我们的)	<b><math>0.73 \pm 0.07</math></b>	<b><math>0.39 \pm 0.04</math></b>	<b><math>0.61 \pm 0.04</math></b>	$0.73 \pm 0.36$	<b><math>1.12 \pm 1.07</math></b>

表2.FaceForensics++ [19] 人脸图像的定量消融结果。

方法	SSIM $\uparrow$	验证 $\downarrow$	地标 $\downarrow$	姿势 $\downarrow$
DF	$0.73 \pm 0.07$	$0.61 \pm 0.04$	$0.73 \pm 0.36$	$1.12 \pm 1.07$
DFHD	$0.75 \pm 0.09$	$0.61 \pm 0.04$	$0.71 \pm 0.37$	$1.06 \pm 0.97$
DFHD (GAN)	$0.72 \pm 0.11$	$0.61 \pm 0.04$	$0.79 \pm 0.40$	$1.33 \pm 1.21$
DFHD (GAN + TrueFace)	$0.77 \pm 0.06$	$0.61 \pm 0.04$	$0.70 \pm 0.35$	$0.99 \pm 1.02$
LIAE	$0.76 \pm 0.06$	<b><math>0.58 \pm 0.03</math></b>	$0.66 \pm 0.32$	$0.91 \pm 0.86$
LIAEHD	$0.78 \pm 0.06$	$0.58 \pm 0.03$	<b><math>0.65 \pm 0.32</math></b>	$0.90 \pm 0.88$
LIAEHD (GAN)	$0.79 \pm 0.05$	$0.58 \pm 0.03$	$0.69 \pm 0.34$	$1.00 \pm 0.97$
LIAEHD (GAN + TrueFace)	<b><math>0.80 \pm 0.04</math></b>	$0.58 \pm 0.03$	$0.65 \pm 0.33$	<b><math>0.83 \pm 0.81</math></b>



图6.不同模型结构的消融实验（使用GAN和TrueFace）。（这里，我们提供训练预览，而不是转换后的面孔，目的是在DFL的模型结构中做一个公平的比较，同时避免转换阶段的前期处理的影响）。

值得注意的是，DFL不是目前最先进方法的简单组合。相反，最有效的工具是由我们自己根据用户的要求而开发的。

考虑到人脸分割的复杂要求，DFL提供了一个自动算法，TernausNet[8]作为默认算法。如第3.1节所述，TernausNet可以有效地去除不规则的遮挡物。然而，在某些特定的镜头中，这个模型可能无法生成细粒度的掩码。因此，我们开发了一个高效的人脸分割工具，XSeg，它允许每个人通过几张照片的学习来定制以适应特定的需求。如图8所示，用户可以创建面具标签并训练一个自定义的分割模型。在XSeg的帮助下，用户可以自己定义交换掩码，并解决在提取和转换阶段的几乎所有问题。

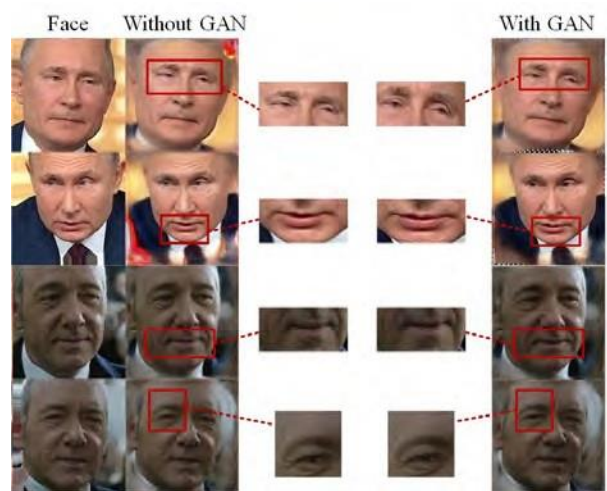


图7.不同训练范式的消融实验：基于非GAN的和基于GAN的（左边的图像是原始脸，右边是由没有GAN训练的模型产生的重建图像，最右边是由GAN训练的模型产生的）。很明显，GAN使模型在捕捉尖锐的细节，即皱纹和痣方面变得更加合理。同时，与没有使用GAN的模型相比，大大减少了模糊性。

## 5.2. 潜在的

以前的方法通常侧重于通过输入两个视频或数百个图像来合成高质量的结果。然而，以这种方式做出好的换脸结果是完全不合理的。DFL支持大规模的数据集，高达10万张图片。在巨大数据的帮助下，最终交换的结果可以达到显著的质量改进。



图8.XSeg的预览。用户可以通过XSegEditor标记他们想要的遮罩。在XSeg的帮助下，用户可以用它来消除手、眼镜和任何其他可能覆盖面部的物体的遮挡，并控制特定区域的交换。

此外，以前的方法总是缺乏潜力。与以前的方法不同，DFL可以提供换脸功能，并支持嘴唇操作、头部替换、做旧等等。这些新功能可以通过简单调整我们的框架来实现。我们还鼓励用户使用视频编辑工具，如DaVinci Resolve和After Effect，进一步提高最终视频的视觉质量。

### 5.3. 更广泛的影响

自从与深度伪造有关的作品受到关注后，DFL作为视觉特效艺术家最常用的深度伪生成工具，已经发挥了不可替代的作用。当涉及到将特技演员与流行明星重新放在一起时，DFL的出现无疑给世界增添了娱乐性，同时也给后期制作行业带来了高经济价值。

由于DFL可以产生电影质量的换脸结果，从事造假检测领域的研究人员可能会被激励去设计高质量造假视频片段或图像的强大分类器。

## 6. 结论

快速发展的DeepFaceLab已经成为深度学习从业者社区中流行的换脸工具，它将人们从费力的、复杂的数据处理、训练中琐碎的细节工作和conversion中解放出来。随着越来越多的人参与到DeepFaceLab的开发中，Deepfake娱乐已经成为社交媒体上的潮流。未来，我们希望在推动造假检测领域发展的同时，更深入地挖掘与娱乐相关的人工智能框架。

## 参考文献

- [1] Armen Aghajanyan. Convolution aware initialization. *arXiv preprint arXiv:1702.06295*, 2017. **5**
- [2] Adrian Bulat and Georgios Tzimiropoulos. 我们离解决2D和3D人脸对齐问题还有多远? (以及23万个3D面部标志的数据集)。In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1021-1030, 2017. **3, 6**
- [3] Deepfakes. Deepfakes. <https://github.com/deepfakes/faceswap>, 2017. **2, 5, 6**
- [4] 邓建康, 郭佳, 周宇翔, 于金科, Irene Kotsia, 和 Stefanos Zafeiriou. Retinaface: *arXiv preprint arXiv:1905.00641*, 2019. **3**
- [5] Brian Dolhansky, Joanna Bitton, Ben Pfau, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. deepfake detection challenge (dfdc) 数据集, 2020年. **2**
- [6] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. 用位置图回归网络进行联合三维人脸重建和密集对齐。In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 534-551, 2018. **3**
- [7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 生成式对抗网。In *Advances in neural information processing systems*, pages 2672-2680, 2014. **1**
- [8] Vladimir Iglovikov and Alexey Shvets. Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*, 2018. **4, 8**
- [9] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 用于实时风格转移和超级分辨率的感知损失。在 *欧洲计算机视觉会议* 上, 第694-711页。Springer, 2016. **6**
- [10] Tero Karras, Samuli Laine, and Timo Aila. 一个基于风格的生成器架构, 用于生成式对抗网络。在 *IEEE 计算机视觉和模式识别会议* 上, 第4401-4410页, 2019年. **1**
- [11] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 分析和改善stylegan的图像质量。 *arXiv预印本 arXiv:1912.04958*, 2019. **1**
- [12] Davis E King. Dlib-ml: A machine learning toolkit. *杂志 of Machine Learning Research*, 10(July):1755-1758, 2009. **6**
- [13] Huangxing Lin, Weihong Zeng, Xinghao Ding, Yue Huang, Chenxi Huang, and John Paisley. Learning rate dropout. *arXiv preprint arXiv:1912.00144*, 2019. **5**
- [14] Artur Loza, Lyudmila Mihaylova, Nishan Canagarajah, and David Bull. 视频序列中基于结构相似性的物体追踪。在 *2006年第九届国际会议* 上, 信息融合, 第1-6页。IEEE, 2006. **5**
- [15] Yuval Nirkin, Iacopo Masi, Anh Tuan Tran, Tal Hassner, and Gerard Medioni. 关于人脸分割、人脸互换和人脸感知。在 *IEEE 自动人脸和手势识别会议* 上, 2018. **2, 5, 6**
- [16] Patrick Pe' rez, Michel Gangnet, and Andrew Blake. 泊松图像编辑。在 *ACM SIGGRAPH 2003 论文* 中, 第313页-318. 2003. **5**
- [17] Francis Pitie', Anil C Kokaram, and Rozenn Dahyot. 使用颜色分布转移的Automated颜色分级。 *计算机视觉和图像理解*, 107 (1-2) : 123-137, 2007. **5**
- [18] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. 图像之间的颜色转移。 *IEEE 计算机图形和应用*, 21 (5) : 34-41, 2001. **5**
- [19] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: 学习检测被操纵的面部图像。在 *IEEE 国际计算机视觉会议论文集 Computer Vision*, 第1-11页, 2019年. **5, 7, 8**
- [20] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: 实时人脸捕捉和RGB视频的重演。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2387-2395, 2016. **5**
- [21] Shinji Umeyama. 两个点模式之间转换参数的最小二乘法估计。 *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pages 376-380, 1991. **3, 5**
- [22] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 图像质量评估: 从错误可见性到结构相似性。 *IEEE transactions on image processing*, 13(4):600-612, 2004. **5**
- [23] Tsun-Yi Yang, Yi-Ting Chen, Yen-Yu Lin, and Yung-Yu Chuang. Fsa-net: 从单一图像中学习细粒度的头部姿势估计的结构聚集。In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1087-1096, 2019. **6**
- [24] 张世峰, 朱翔宇, 雷震, 石海林, 王晓波, 李斯坦兹。 S3fd: 单次拍摄的尺度不一的人脸检测器。In *Proceedings of the IEEE International Conference on Computer Vision*, pages 192-201, 2017. **3**