我们如何从单个面部变形图像中高效地恢复组成部分的图像？面部变形是通过混合属于不同个体的两张（可能更多）面部图像而创建的，同时确保它在生物特征上与所有参与身份相匹配。变形图像可以逃避手动检测，使多个个体能够使用单一文件获得访问权限，这使得它们成为显而易见的安全威胁。从单个变形图中推导组成的面部图像——即去变形——一直是一项至关重要而又具有挑战性的任务。去变形是一个病态的逆问题，挑战不仅在于缺乏已知的变形技术（基于标志点或深度学习）的使用，还在于输出空间缺乏约束。在本文中，我们提出了一个新的框架，以从单个变形图像中分离出组成的面部图像。我们的方法在一个更现实的评估下运行（见第 1 节），并在去变形图像中实现了高度的面部保真度（见图 ?? ）。

解混（Demorphing）就像变形攻击检测（MAD）一样，可以是 i) 基于参考的 [19, 34, 20, 9, 30, 8] ，即除了变形图像外，还可以获得一个组成身份之一的图像，或者 ii) 无参考的 [45, 7, 47] ，即除了变形图像外没有其他可用信息。与基于参考的方法相比，单图像无参考解混是一个明显具有挑战性的任务。实际上，给定一个单一的变形图像，在输出图像空间没有额外约束的情况下，可以有无限多种可能的分解对。此外，开发一个统一的框架来解混使用不同变形技术创建的图像同样具有挑战性。我们的方法学习在输入变形的条件下生成最可能的一组人脸图像。

潜在空间去形：学习去形的过程可以分为两个阶段：感知压缩，即在保留语义的同时丢弃不重要的细节，然后是在潜在空间中进行分解过程的生成建模。在潜在空间中操作通过提供一个感知上等效但更易处理的领域 [38] ，简化了学习。为此，我们首先学习一个自动编码器，将输入图像投射到一个低维度但感知上相等且计算效率高的潜在空间中，其次，我们训练一个有条件的 GAN [32] ，基于输入的变形图像对其在该潜在空间中有效地分解成其组成图像。

除了计算效率高外，我们方法的一个关键优势在于，与先前的方法不同，去变形器只需要训练一次。这是因为它直接作用于潜在表示，这些表示已从图像层级的噪声和伪影中解耦。压缩阶段只捕捉重要的语义特征，忽略像素空间中不重要的属性（背景、变形伪影等）。这意味着我们的方法可以处理使用未见过的变形技术制作的变形脸，尽管从未在它们上进行过训练。我们将我们的方法称为潜在条件生成对抗网络。

另一个关键挑战是用于解变形的数据集的有限性，以及使用或分享真实面部数据相关的隐私问题。现有的变形数据集主要是为变形攻击检测（MAD）设计的，通常只包含大约 1.5K 的变形，这不足以训练生成模型。为了解决这些限制，我们的方法在使用合成生成的面部图像创建的变形上进行训练，从而同时解决隐私和数据稀缺的问题。

总之，我们的贡献如下：

- 与之前的工作 [7, 47, 45] 相比，我们的方法限制更少，并在一个现实的协议下运行。此外，我们的方法能够优雅地扩展到更高分辨率的图像。
- 与以往的方法不同，我们的方法对变形技术和面部风格（护照风格、背景等）不敏感。我们的方法可以用于一般去形变（见第 ?? 节）。
- 我们通过实验表明，通过在潜在空间中解除变形，我们的方法不仅克服了模型倾向于将变形图像作为输出重现的变形复制问题，还抑制了恢复图像中的高频伪影。

，，

1. 背景

。通常是将两个人的两张面部图像结合在一起。传统上，变形是通过几何对齐面部标志点，并将第二张图像叠加在基础图像之上来完成的 [35, 14] 。由于这种方法，变形图像通常偏向于参与的某一张图像，导致对第二张图像的信息保留非常少，难以忠实恢复。最近，提出了深度变形方法 [10, 51, 42, 26] ，从零开始创建变形，使逆变形更加具有挑战性。一种变形操作符 $\mathcal{M}$ ，作用于两张面部图像，i_1 和 i_2 ，生成变形图像 x ，即 $x = \mathcal{M}(i_1, i_2)$ 。变形操作符的目标是创建一个变形图像，使得：i) 从像素空间的感知图像质量来看，变形看起来合理，并且 ii) 变形与 i_1 和 i_2 在生物特征匹配器 $\mathcal{B}$ 的匹配中相符，即 $\mathcal{B}(x, i_n) > \tau$ ，$n \in \{1, 2\}$ ，并且 τ 是根据应用场景决定的匹配阈值。无参考逆变形尝试近似变形过程的逆过程。给定一个变形输入 x ，目标是恢复用于创建变形的图像。无参考逆变形是一个不适定的逆问题 [7, 45] 。

一个去形变算子，$\mathcal{DM}(= \mathcal{M}^{-1})$ ，在接收到形变图像 x 后，尝试重建原始的组成图像 $o_1, o_2 = \mathcal{DM}(x)$ ，满足以下条件：

$$\mathcal{B}(o_1, o_2) < \theta \tag{1}$$

$$\min_{j \in \{1,2\}} \max_{\substack{k \in \{1,2\} \\ k \neq j}} \{ \mathcal{B}(o_j, i_k), \mathcal{B}(o_j, i_j) \} > \epsilon \tag{2}$$

其中，θ 和 ϵ 是匹配阈值。方程式（1 ）强制要求重建的输出彼此看起来不同，而方程式（2 ）确保每个重建输出与其对应的真实图像一致。

在文献中，去伪装技术在不同的场景下进行了探索，这些场景是基于训练和测试的伪装是如何组合而成的 [47, 46] 。考虑一个面部图像集 $\mathcal{Y}$ ，用于创建伪装（训练和测试）。下面的场景在图 2 中有说明。

1. 训练和测试中使用相同身份：训练和测试的形变图像都是由 $\mathcal{Y}$ 中的面部图像创建的。然而，相同的 对 身份不会用于生成训练和测试形变图像。
2. 部分未看到的身份：这里，脸集 $\mathcal{Y}$ 被划分为两个不相交的集合，具有不相交的身份，$\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 。训练变形仅由 $\mathcal{Y}_1$ 中的身份创建，而测试变形则使用一个在 $\mathcal{Y}_1$ 的身份和另一个在 $\mathcal{Y}_2$ 中的身份创建。
3. 完全不相交的身份：用于创建训练变形的身份与用于创建测试变形的身份完全不相交。换句话说，

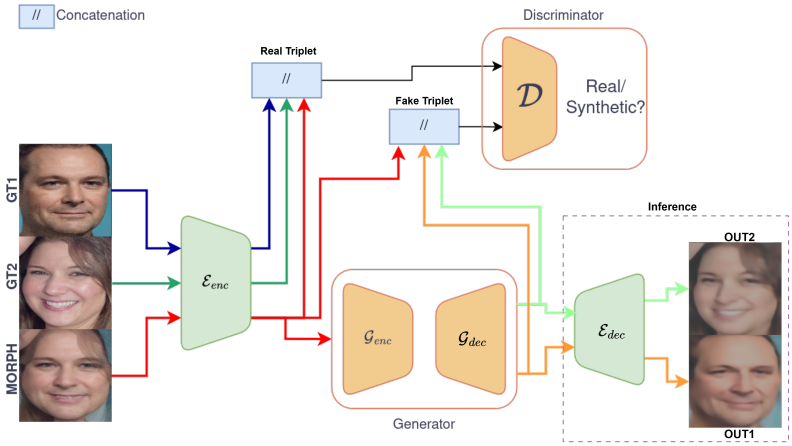

Figure 1. 提出的去融合架构：编码器 $\mathcal{E}_{enc}$ 在训练期间将融合图像与构成的面部图像一起压缩。生成器 $\mathcal{G}$ 在 $\mathcal{E}_{enc}$ 的潜在域中重建两个基于融合图像的面部图像。判别器用于区分真实和合成的面部特征三元组。在推理过程中，解压缩器 $\mathcal{E}_{dec}$ 恢复构成的图像。注意，解码器 $\mathcal{E}_{dec}$ 仅在推理时用于解压缩去融合的输出。

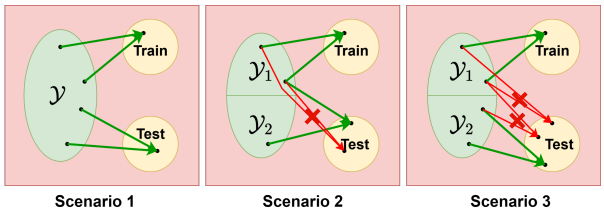

Figure 2. 去形变中的情境：绿色箭头表示有效的测试形变，而红色箭头表示在每种情境中无效的测试形变。（左）训练和测试形变使用 $\mathcal{Y}$ 中的身份创建。（中）测试形变使用 $\mathcal{Y}_1$ 中的一个身份和 $\mathcal{Y}_2$ 中的另一个身份创建。（右）训练和测试形变分别使用 $\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 中的不相交身份创建。

训练变形专门由 $\mathcal{Y}_1$ 中的身份创建，而测试变形则专门由 $\mathcal{Y}_2$ 中的身份创建。

在本文中，每一个数据集中的身份都由单张图像表示。训练形变图像是通过合成的人脸图像生成的，而测试形变图像则专门从 FRLL 人脸数据集中的中性人脸图像创建。因此，“图像”和“身份”这两个术语有时是可以互换使用的。

在 [7] 中，作者提出了一种基于 GAN 的去变形方法，该方法受到了 [54] 的启发。他们的方法包含一个生成器和三个判别器。尽管原始技术有效地分离了自然场景图像，但在应用于去变形时面临一个称为变形复制的问题，即模型倾向于生成两个与变形非常相似的输出。这一问题产生的原因是人脸图像内部的局部相似性比场景图像要高。此外，该方法假设训练和测试的变形图像是使用相同的变形技术创建的。Shukla [45] 引入了一种使用扩散模型的去变形方法 [22]，该方法通过迭代地向变形图像添加噪声，并在逆过程期间检索组成面孔。然而，这种方法假设训练和测试的变形图像都来源于同一组面孔图像，即场景 1。在 [47] 中，作者提议将变形图像分解为多个不可理解的、保护隐私的组成部分，然后通过加权和合并这些组成部分来重建组成图像。这种方法也假定训练和测试的变形图像共享同一身份。这些技术虽然在不同程度上取得了成功，但通常对于实际的真实世界应用来说过于限制。

在 [37] 中，作者提出了一种基于扩散自动编码器的参考变形技术。他们的方法使用预训练的扩散自动编码器将图像编码为两个子空间：一个捕捉身份特征的语义潜在空间和一个保留剩余随机细节的随机潜在空间。通过条件去噪扩散隐式模型解码两个子空间的潜在编码，恢复共犯的面部图像。FD-GAN [34] 是另一种基于参考的方法，它采用双重架构，旨在使用第二个身份的图像作为参考，从变形输入中重建第一个身份的图像。为了评估生成模型的有效性，它随后尝试使用第一个身份的重建图像作为输入来恢复第二个身份。

在本研究中，我们提出了一种不依赖于两个关键假设的无参考解伪变技术，即：i）用于创建测试伪变的伪变技术和 ii）训练和测试伪变中的真实身份。

我们将我们方法的训练设定为一个由对抗性损失函数和峭度正则化引导的潜在空间中的逐像素回归问题。我们的方法包括一个图像压缩器，$\mathcal{E}$，一个去形变生成器，$\mathcal{G}$，以及一个判别器，$\mathcal{D}$。图 1 展示了我们方法的概览。我们的方法首先通过一个编码器 $\mathcal{E}$ 感知地压缩输入的变形图像。这一步有效地消除了对去形变不关键的不重要的高级视觉特征。压缩输入图像的另一个优点是表示的标准化，即，人脸图像在编码器的潜在域中表示，消除了背景、变形伪影、光照等干扰。我们方法的骨干基于条件生成对抗网络（conditional-GAN）。一个图像到图像的生成器 $\mathcal{G}$，在编码的变形图像的条件下，在潜在空间中去形变。一个判别器将编码的变形与编码的去形变输出以及对应的真实图像拼接，并区分真实与合成的三元组。我们在 $\mathcal{E}$ 的潜在空间中优化 GAN，这提供了两个主要优势。(i) 计算效率：由于 $\mathcal{E}$ 的权重保持不变，高分辨率的人脸图像使用低维张量表示，显著加速了训练和推理过程。(ii) 感知损失：在潜在空间中计算损失消除了对比无关特征的需要。比如，相同的人脸图像在不同背景下在 RGB 域中可能表

现出显著的每像素损失，而在潜在空间中，这种差异要小得多。在推理过程中，解码器 $\mathcal{E}_{dec}$ 解码生成器产生的去形变输出。

1.1. 人脸图像压缩

我们的人脸压缩模型基于 [29] ，由使用 KL 损失训练的自编码器组成，它通过强制局部真实感来确保重建限制在图像流形内，并防止依赖于像 L_2 或 L_1 这样仅依赖于像素空间损失所常常导致的模糊。具体来说，给定一个 RGB 空间中的图像 $x \in \mathbb{R}^{H\times W\times 3}$ ，编码器 $\mathcal{E}_{enc}$ 将图像压缩成一个潜在表示 $z = \mathcal{E}_{enc}(x) \in \mathbb{R}^{h\times w\times c}$ 。对称地，解压器 $\mathcal{E}_{dec}$ 从潜在表示重建图像 $\tilde{x}$ ，使得 $\tilde{x} = \mathcal{E}_{dec}(z) = \mathcal{E}_{dec}(\mathcal{E}_{enc}(x))$ 。压缩器通过一个因子 2^3 缩小图像的高度和宽度，以满足 $h = H/8$ 和 $w = W/8$ 。实际上，我们使用在 Stable Diffusion [39] 中采用并使用 KL 损失训练的预训练自编码器。对于一个大小为 $512 \times 512 \times 3$ 的图像，自编码器生成一个大小为 $64 \times 64 \times 4$ 的潜在表示。自编码器的权重在整个训练过程中保持不变。

我们训练我们的去形器，以减少在潜在空间中生成的去形输出与真实值之间的距离。与之前的方法不同，我们明确地排列输出，但在训练过程中随机交换真实对。这使得我们可以使用标准的每像素损失进行各种顺序的训练（见图 5 ），通过让模型接触到两种可能的顺序，增强重建图像的稳健性生成。

潜在条件 GAN：使用我们训练好的压缩网络 $\mathcal{E}$ ，我们现在在一个高效的低维潜在空间中进行去形变，该空间抽象掉了 RGB 像素空间中的不必要的感知不到的细节。在潜在空间中进行去形变不仅使过程在计算上更为高效，还使得方法更适合专注于重要的语义数据位。我们方法的骨干是一个条件 GAN [24] 。生成器 $\mathcal{G}$ 和判别器 $\mathcal{D}$ 基于形变 x 。考虑一个使用两张人脸图像 i_1 和 i_2 创建的形变 x ，那么条件 GAN 损失可以表述为：

$$\begin{aligned}\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{x,i}[\log \mathcal{D}(x,i)] \\ +\mathbb{E}_{x,z}[\log(1-\mathcal{D}(x,\mathcal{G}(x,z)))]\end{aligned} \tag{3}$$

，其中，$i = (i_1, i_2)$ 代表输出。为了在生成的输出和真实输出之间强制一致性，我们还加入了一个 $\mathcal{L}_1$ 损失：

$$\mathcal{L}_1 = \mathbb{E}_{x,i,z}[||i - G(x,z)||_1]. \tag{4}$$

形态复制：在 [7] 中，作者使用基于 [54] 的生成对抗网络（GAN）去形变人脸图像。尽管有分离评估的机制，他们的方法仍然遭遇了形态复制的问题，即模型倾向于将形变本身作为其两个输出之一。这主要是由于与自然场景图像相比，人脸图像在本质上的局部相似性。实际上，在像素空间中，比较两张护照风格的面部图像（通常用于边境控制、驾照等）会导致它们的平均值成为成对距离的最小化器，并且仍然像脸。为了解决这个问题，我们在先验 $p(\hat{x}|x)$ 上引入了另一个约束。我们引入了一种峰度损失，旨在最小化预测输出 $o = (o_1, o_2)$ 与真实值 $i = (i_1, i_2)$ 之间的峰度差异：

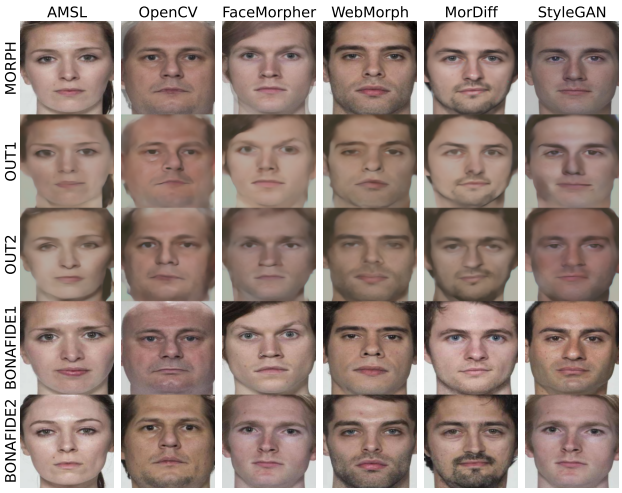

Figure 3. 无需参考的去形变：我们在 FRLL-Morph 和 MorDiff 数据集上测试我们的方法。模型在输入 MORPH 时以任意顺序输出 OUT1 和 OUT2。我们的方法产生了视觉上明显不同的输出，i) 在 OUT1 和 OUT2 之间，ii) 在 MORPH-OUT1 和 MORPH-OUT2 之间。

Figure 4. 我们的方法确保在提供非变形输入时，两个输出保持相同的身份（因为输入只包含一个身份）。

$$\mathcal{L}_{\mathrm{kurt}} = \sum_{j=1}^{2} \Big|\mathrm{Kurt}(o_j) - \mathrm{Kurt}(i_j)\Big| \tag{5}$$

通过对齐生成图像和真实图像的高阶统计量，峰度损失有助于维持合成输出的结构一致性和逼真性。该损失与传统损失相辅相成，如 L_1 和感知损失，这些损失主要关注于像素或特征差异，但可能忽略了分布属性。最终的目标函数将峰度损失与对抗和重建损失结合在一起：

$$\mathcal{L} = \mathcal{L}_{\mathrm{cGAN}} + \lambda_1 \mathcal{L}_{1} + \lambda_2 \mathcal{L}_{\mathrm{kurt}}, \tag{6}$$

其中，λ_1 和 λ_2 均被设置为 0.5 。通过确保输出保持与真实数据相似的统计特性，加入峰度损失可以增强生成图像的感知质量。

1.2. 实现细节

我们使用了预训练的、应用于 Stable Diffusion 的自编码器，该自编码器在 KL 损失上进行了训练。这个自编码器由四个基于 ResNet 的下采样和上采样块组成，每个块都有 GroupNorm [50] 、dropout 和 SiLU [17] 激活函数。我们的生成器 $\mathcal{G}$ 基于 HuggingFace 的条件 UNet 实现 UNet2DConditionModel ，包括六个基于 ResNet 的下采样块和六个相应的上采样块。在第五个下采样块和第二个上采样块中包含了自注意力机制。UNet 还需要一个时间步输入，在所有实验中我们将其固定为零以中和其效果。判别器 $\mathcal{D}$ 基于 CNN 架构，由四个连续的块组成，每个块包含一个卷积层、实例归一化 [49] 和一个 LeakyReLU 激活函数。

我们使用 AdaFace [27] 和 ArcFace [16] 人脸识别模型评估我们的去形变方法。我们使用余弦相似度来计算生物识别匹配得分。训练通过 accelerate [21] 支持多 GPU，并使用 Adam 优化 [28] 进行。训练参数如下：训练周期：300，学习率：10^{-4} ，dropout 率：0.1，β_1 ：0.5 和 β_2 ：0.999。

2. 数据集

测试数据集：我们在三个著名的变形数据集上测试我们提出的方法：AMSL [33] 、FRLL-Morphs [15] 和 MorDiff [10] 。FRLL-Morphs 数据集包括使用四种不同技术生成的变形：OpenCV [41] 、StyleGAN [26] 、WebMorph [14] 和 FaceMorph [35] 。在所有三个数据集中，源（即非变形）图像来自 FRLL 数据集，该数据集包括 102 个身份，每个身份有两张正面图像——一张微笑和一张中性——总计 204 张人脸图像。每个数据集中的变形数量如下：AMSL：2,175 个变形；FaceMorpher：1,222 个变形；StyleGAN：1,222 个变形；OpenCV：1,221 个变形；WebMorph：1,221 个变形；MorDiff：1,000 个变形。选择测试数据集包括传统的基于地标的变形技术和最近引入的生成方法。

训练数据集：众所周知，生成模型需要大量的训练数据来准确捕捉数据分布。在面部去形变的背景下，一个足够大的形变数据集对于有效的解耦是至关重要的。此外，使用/重用/共享真实生物特征数据的法律和伦理挑战也带来了重大挑战。因此，公开可用的数据集不足以训练一个通用的去形变技术，因为：(i) 它们的规模有限（通常约为 1.5k 形变）[11, 12, 18, 19, 36, 43] 和 (ii) 隐私/法律问题 [5, 6, 1, 3, 2, 4] 。为了训练高分辨率去形变生成方法，我们遵循在 [48] 中提出的训练协议。为了生成形变，我们从使用在 Flickr-Faces-HQ 数据集（FFHQ）上训练的 StyleGAN2-ADA [25] 生成的 SMDD 训练集中采样两个合成面部图像。选择随机图像有助于训练更具普遍性的形变攻击检测（MAD）系统 [13] 。我们使用广泛采用的 [12, 40, 44] OpenCV/dlib 形变算法 [31] ，利用 Dlib 的地标检测器实现生成形变图像。这种基于地标的形变方法生成的形变攻击比其他工具生成的更为有效 [40] 。通过这种策略，我们各生成了 15,000 个训练和测试形变（来自 SMDD 训练和测试数据集的 25K 面部图像）。所有图像（训练和测试）均使用 MTCNN [52] 进行人脸检测，之后裁剪人脸区域。然后对图像进行归一化并调整尺寸至 512×512 的分辨率。无法检测到人脸的图像会被丢弃。值得注意的是，不会进一步进行空间变换，确保训练过程中的形变和真实构成图像中的面部特征（如嘴唇和鼻子）保持一致。

身份泄漏：尽管使用合成变形进行训练有其优势，但确保测试身份不会在合成面孔生成过程中被无意复制是至关重要的。换句话说，用于创建训练变形集的身份应与用于生成测试变形的身份不同。此分离可以防止身份泄漏，并确保对模型泛化能力的公平评估。我们评估了身份泄漏情况，使用了 25,000 个训练面孔图像，每个代表不同的身份，以及来自 FRLL 数据集的 204 个（来自 102 个身份）测试面孔图像。这一评估是通过两个面孔匹配器进行的，即 ArcFace 和 AdaFace。我们观察到使用 ArcFace 匹配器时，两个集合之间的平均最高 $n\%$ (0.1,1,5) 相似性为 0.22, 0.17, 0.13，使用 AdaFace 时为 0.21, 0.16, 0.12。这些相对较低的得分表明测试集中的身份不太可能存在于训练面孔集中，这表明身份泄漏最小。

3. 评估

3.1. 现有指标

无参照人脸去合成是通过生成学习的进步而实现的一个相对较新的发展。在文献中有三种主要的评估指标来对去合成方法进行基准测试，即在不同错误匹配率阈值下的真匹配率、还原精度（RA）以及标准的图像质量评估指标如 SSIM 和 PNSR。在 [48] 中，作者认为 TMR@FMR 和 RA 只关注生物特征身份（而不是图像质量），而 SSIM/PSNR 只关注质量（而不关注生物特征效用）。因此，需要一个全面的指标来有效捕捉面部匹配器嵌入域中的身份以及像素空间中的结构质量。这个指标，在 [48] 中提出，将在下一个小节中描述。

3.2. 基于生物识别的交叉加权 IQA

设 $\mathcal{X}$ 为面部变形集，$\mathcal{Y}$ 为用于创建 $\mathcal{X}$ 的面部图像，$\mathcal{M} : \mathcal{Y} \times \mathcal{Y} \to \mathcal{X}$ 为变形操作符。解变形器 $\mathcal{DM} : \mathcal{X} \to \mathcal{Y} \times \mathcal{Y}$ 近似 $\mathcal{M}^{-1}$ 使得 $\mathcal{DM}(x) = (o_1, o_2)$ ，这一目标可能无法解析地存在。$\mathcal{DM}$ 的目标是确保不发生变形复制，即 $o_1 \neq o_2 \neq x$ 。换句话说，$\mathcal{DM}$ 产生的输出应看起来不相似（在它们确实不同的程度上），并且这些输出在生物特征匹配器 $\mathcal{B}$ 方面应符合真实面部图像。生物特征加权的图像质量评估 [48] 定义为

$$BW(iqa) = \mathbb{E}_{x \in \mathcal{X}} \max \begin{pmatrix} \sum\limits_{i \in \{1,2\}} \mathcal{B}(o_i, i_i) \cdot iqa(o_i, i_i), \\ \sum\limits_{\substack{i \in \{1,2\} \\ j = i\%2+1}} \mathcal{B}(o_i, i_j) \cdot iqa(o_i, i_j) \end{pmatrix}$$

Table 1. 在统一协议下，我们的方法与现有的最先进去伪形技术进行比较。我们的方法优于 IPD [47] 、SDeMorph [45] 和 Face Demorphing [7] 。我们使用已建立的图像分解 IQA 指标（PSNR/SSIM）、去伪形指标（恢复准确性）以及生物识别加权 IQA（BW）[48] 评估我们的方法。报告的分数由原作者提供。

Method	Metric	ArcFace						AdaFace					
		AMSL	OpenCV	FMorph	Wmorph	MorDiff	StyleGAN	AMSL	OpenCV	FMorph	Wmorph	MorDiff	StyleGAN
Ours	Rest. Acc. @ 10 % FMR	99.90 %	99.83 %	100.00 %	100.00 %	100.00 %	38.76 %	98.58 %	99.53 %	99.71 %	100.00 %	100.00 %	50.39 %
	Rest. Acc. @ 1 % FMR	99.09 %	99.30 %	99.03 %	99.47 %	100.00 %	12.57 %	96.35 %	98.58 %	98.57 %	98.91 %	100.00 %	20.75 %
	Rest. Acc. @ 0.1 % FMR	96.26 %	95.83 %	95.16 %	96.28 %	98.63 %	2.12 %	91.19 %	93.69 %	93.41 %	94.53 %	98.83 %	4.52 %
	BW (SSIM)	0.44	0.49	0.49	0.46	0.50	0.32	0.35	0.38	0.38	0.36	0.41	0.12
	BW (PSNR)	10.12	11.56	11.56	10.56	11.07	7.30	7.92	9.02	9.09	8.22	9.05	2.79
IPD (2024) [47]	Restoration Accuracy	25.69 %	40.54 %	37.82 %	25.61 %	38.12 %	16.22 %	0.18 %	1.89 %	1.43 %	0.31 %	3.88 %	0.00 %
	BW (SSIM)	0.26	0.32	0.32	0.25	0.33	0.22	0.17	0.21	0.21	0.16	0.22	0.08
	BW (PSNR)	6.28	7.98	9.95	6.16	8.12	5.31	4.14	5.30	5.29	4.10	5.53	1.93
SDeMorph (2023) [45]	Restoration Accuracy	12.56 %	15.62 %	13.18 %	12.80 %	11.67 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
	BW (SSIM)	0.16	0.19	0.19	0.18	0.17	0.15	0.11	0.12	0.12	0.12	0.11	0.05
	BW (PSNR)	4.24	4.88	4.97	4.68	4.29	3.85	2.76	3.20	3.22	3.16	2.98	1.29
Face Demorphing (2022) [7]	Restoration Accuracy	0.45 %	0.53 %	0.51 %	0.50 %	0.62 %	0.43 %	0.17 %	0.23 %	0.17 %	0.20 %	0.29 %	0.00 %
	BW (SSIM)	0.21	0.24	0.25	0.23	0.29	0.19	0.11	0.14	0.13	0.13	0.16	0.06
	BW (PSNR)	4.46	5.21	5.50	4.93	6.13	4.13	2.39	2.95	2.89	2.82	3.33	1.31

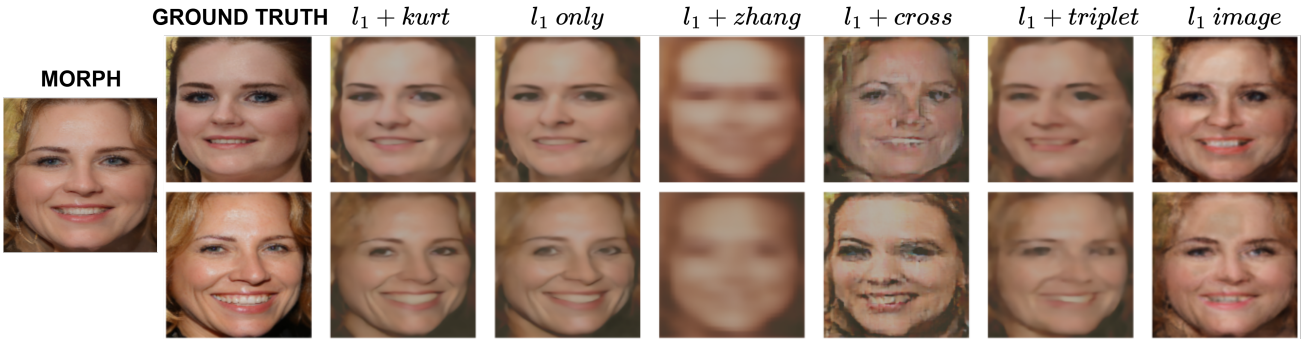

Figure 5. 基于不同图像分离先验的结果比较：Zhang 的排除损失 [53] 、交叉路口损失 [54] 和三元组损失 [23] 。

Table 2. 使用标准图像质量评估指标 PSNR / SSIM 进行去形变质量比较。更高的数值表示更好的图像重建保真度。

Method	AMSL	OpenCV	FaceMorpher	WebMorph	MorDiff	StyleGAN
Ours	10.81 / 0.47	11.65 / 0.50	11.63 / 0.49	11.17 / 0.49	10.93 / 0.49	10.18 / 0.45
IPD (2024) [47]	9.32 / 0.38	10.37 / 0.42	10.27 / 0.41	9.63 / 0.39	9.91 / 0.40	9.08 / 0.37
SDeMorph (2023) [45]	8.99 / 0.34	9.54 / 0.37	9.60 / 0.37	9.45 / 0.37	8.97 / 0.34	8.74 / 0.34
Face Demorphing (2022) [7]	9.68 / 0.46	10.35 / 0.48	10.44 / 0.47	10.20 / 0.48	10.13 / 0.47	9.51 / 0.45

，其中 % 是模运算符且 $iqa \in \{SSIM, PSNR\}$ 。BW(iqa) 计算两个可能的输出-真实对组合之间的图像质量评估，并用面部匹配器产生的生物特征匹配得分对其进行加权。

4. 结果

在单张图像的无参考去混合方面的研究工作有限。我们将我们的方法与现有的三种方法进行比较。表 1 和表 2 展示了我们的方法与 [45, 47] 的定量比较。我们使用现有指标（TMR、恢复准确性、IQA）以及 BW-IQA 对我们的方法进行比较。在提出的协议中，我们的方法在所有指标和数据集上显著优于现有方法。从视觉上看，与之前的方法相比，我们的方法能够捕捉到准确的面部特征（参见图 ?? 和图 3 ）。我们还将我们的方法与 [7] 进行了比较。在 AMSL 数据集上，我们的方法使用 ArcFace 达到了 97.77 % 的 TMR，明显优于他们的 70.55 % 的结果。

关于潜在空间的分析：为了评估在潜在空间中进行去混合的重要性，我们在 RGB 域执行相同的实验，即，用一个输出输入图像的恒等函数代替 $\mathcal{E}$ ，保持其他设置不变。表 3 展示了两个模型的评估结果，而图 5 显示了视觉比较。与在像素空间中去混合时产生模糊且不太明显的面孔相比，我们的方法在潜在空间中能够明显分离两张面孔并生成高分辨率图像。

为了评估我们方法的有效性，我们对来自真实样本生成的变形图像执行逆变形处理。我们收集了来自不同主体的 17 张人脸图像（5 名男性，3 名女性），指导他们像拍摄驾照照片一样进行摆姿，允许他们自行决定是否佩戴或摘除配饰，如珠宝或眼镜。我们使用第 2 节中描述的协议创建了 28 张变形图像，确保每对身份至少有一个变形样本。我们在图 6 中展示了一些样本。我们的方法生成了不同的输出，同时保留了与真实图像的高度相似性。我们观察到使用 AdaFace 人脸匹配器的还原准确率为 95.65 %，使用 ArcFace 的为 91.30 %。我们还观察到 PSNR 为 10.66，SSIM 值为 0.52。最后，我们使用 BW(iqa) 度量评估结果。我们观察到在 AdaFace 和 ArcFace 上 BW(SSIM) 分别为 0.16 和 0.24，BW(PSNR) 分别为 3.31 和 5.03。

在本文中，我们介绍了一种新颖的无需参考的面部去变形框架，称为条件潜在生成对抗网络，它在一个感知等价的潜在空间中运行，以有效地从单个变形图像中分离出组成的人脸。与先前的方法不同，我们的方法对变形技术不敏感，可以扩展到高分辨率图像，并能够

Table 3. 分离先验的影响：我们分析了在训练期间应用的不同分离先验的影响，发现将峰度损失与标准逐像素损失相结合，在所有测试的方法中表现最佳。

Metric	Dataset	ℓ_1 + kurt	ℓ_1 only	ℓ_1 + zhang	ℓ_1 + cross	ℓ_1 + triplet	ℓ_1 image	trivial
PSNR	AMSL	10.81	10.75	10.98	9.78	11.2	10.0	-
	OpenCV	11.65	11.48	11.67	10.21	11.83	10.78	-
	FaceMorpher	11.63	11.58	11.71	10.25	11.99	10.69	-
	WebMorph	11.17	11.07	11.4	10.12	11.45	10.43	-
	MorDiff	10.93	11.03	11.33	9.7	11.4	10.16	-
	StyleGAN	10.18	10.14	10.37	9.55	10.41	9.53	-
SSIM	AMSL	0.47	0.47	0.48	0.39	0.49	0.47	1.0
	OpenCV	0.5	0.49	0.5	0.40	0.51	0.49	1.0
	FaceMorpher	0.49	0.49	0.5	0.39	0.51	0.48	1.0
	WebMorph	0.49	0.49	0.5	0.41	0.51	0.48	1.0
	MorDiff	0.49	0.49	0.5	0.39	0.51	0.48	1.0
	StyleGAN	0.45	0.44	0.46	0.38	0.47	0.45	1.0
Restoration Accuracy AdaFace/ArcFace	AMSL	99.89/98.57	70.04/53.60	71.77/59.20	32.72/27.02	98.55/97.93	99.61/99.33	99.11/99.71
	OpenCV	99.82/99.52	70.55/58.51	64.44/54.10	36.02/29.17	99.45/99.05	99.81/99.38	97.48/97.48
	FaceMorpher	100/99.71	77.10/55.30	67.16/54.72	32.36/27.50	99.65/98.85	100/99.77	98.28/98.28
	WebMorph	100/100	72.06/60.16	73.87/68.12	38.94/34.68	99.46/99.06	100/99.84	97.5/97.5
	MorDiff	100/100	89.74/80.54	87.71/85.99	55.98/34.68	99.55/99.77	100/100	100.00/100.00
	StyleGAN	38.56/50.39	9.12/9.98	5.14/6.55	0.50/0.15	42.93/48.67	61.15/67.70	97.5/97.5
BW(PSNR) AdaFace/ArcFace	AMSL	7.92/10.12	7.62/10.16	3.30/4.82	2.69/6.80	7.34/9.86	6.86/8.40	-
	OpenCV	9.02/11.56	8.55/10.98	3.41/5.14	2.84/7.07	7.99/10.23	7.53/9.26	-
	FaceMorpher	9.09/11.56	8.72/10.86	3.43/5.38	2.85/7.26	7.89/10.43	7.30/9.38	-
	WebMorph	8.22/10.56	7.88/10.36	3.71/5.08	2.86/7.07	7.55/9.81	6.80/8.42	-
	MorDiff	9.05/11.07	8.66/10.80	4.6/6.4	3.07/6.82	8.57/10.37	6.82/8.58	-
	StyleGAN	2.79/7.3	2.66/6.99	1.3/3.8	1.36/5.62	2.77/7.09	2.29/5.83	-
BW(SSIM) AdaFace/ArcFace	AMSL	0.35/0.44	0.33/0.44	0.15/0.21	0.11/0.27	0.32/0.43	0.32/0.40	-
	OpenCV	0.38/0.49	0.36/0.47	0.15/0.22	0.11/0.27	0.34/0.44	0.34/0.42	-
	FaceMorpher	0.38/0.49	0.37/0.46	0.15/0.23	0.11/0.27	0.33/0.44	0.33/0.42	-
	WebMorph	0.36/0.46	0.34/0.45	0.16/0.22	0.11/0.28	0.33/0.43	0.31/0.39	-
	MorDiff	0.41/0.5	0.39/0.48	0.20/0.28	0.12/0.27	0.38/0.46	0.32/0.40	-
	StyleGAN	0.12/0.32	0.12/0.30	0.06/0.17	0.05/0.22	0.12/0.31	0.11/0.27	-

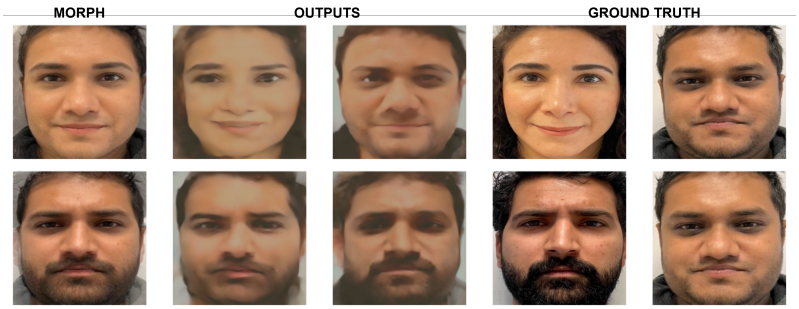

Figure 6. 在活体受试者上的测试：我们的方法有效地根据性别和面部特征分离出真实身份。所有受试者均同意发布。

在不重新训练的情况下泛化到未见过的变形风格。通过利用潜在空间的表示，我们不仅减轻了变形复制问题，还抑制了不必要的伪影，确保了恢复图像中更高的面部保真度。我们的方法在分析性能和视觉保真度方面显著优于现有方法。未来的工作将涉及处理由多于两张图像生成的变形图像。

5. 致谢

本项目由 NSF 识别技术中心（奖项号 # 1841517）和 DHS CINA（奖项号 # 17STCIN00001-07-01）资助。

References

[1] The European Parliament and the Council of the European Union. Article 6(4)(c) of the general data protection regulation. 2016. 4

[2] The European Parliament and the Council of the European Union. Article 9(2)(a) of the general data protection regulation. 2016. 4

[3] The European Parliament and the Council of the European Union. Article 37(1) of the general data protection regulation. 2016. 4

[4] The European Parliament and the Council of the European Union. Article 6(1)(a) of the general data protection regulation. 2016. 4

[5] The European Parliament and the Council of the European Union. Article 9 of the general data protection regulation . 2016. 4

[6] The European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the european parliament and of the council of 27 April 2016 on the protection of nat ural persons with regard to the processing of personal data and on the free movement of such data, and repealing direc tive 95/46/ec (General Data Protection Regulation). 2016. 4

[7] Sudipta Banerjee, Prateek Jaiswal, and Arun Ross. Facial De-morphing: Extracting Component Faces from a Single Morph. In Proceedings of IEEE International Joint Conference on Biometrics, 2022. 1, 2, 3, 5

[8] Sudipta Banerjee and Arun Ross. Conditional identity disentanglement for differential face morph detection. In Proceedings of IEEE International Joint Conference on Biometrics (IJCB), pages 1–8, 2021. 1

[9] Juan Cai, Qiangqiang Duan, Min Long, Le-Bing Zhang, and Xiangling Ding. Feature Interaction-Based Face De-Morphing Factor Prediction for Restoring Accomplice' s Facial Image. Sensors, 24(17), 2024. 1

[10] Naser Damer, Meiling Fang, Patrick Siebke, Jan Niklas Kolf, Marco Huber, and Fadi Boutros. MorDIFF: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Diffusion Autoencoders. In Proceedings of 11th International Workshop on Biometrics and Forensics (IWBF), pages 1–6, 2023. 1, 4

[11] Naser Damer, Jonas Henry Grebe, Steffen Zienert, Florian Kirchbuchner, and Arjan Kuijper. On the generalization of detecting face morphing attacks as anomalies: Novelty vs. outlier detection. In Proceedings of IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), pages 1–5, 2019. 4

[12] Naser Damer, Alexandra Moseguí Saladié, Andreas Braun, and Arjan Kuijper. MorGAN: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Generative Adversarial Network. In Proceedings of 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), pages 1–10, 2018. 4

[13] Naser Damer, Alexandra Moseguí Saladié, Steffen Zienert, Yaza Wainakh, Philipp Terhörst, Florian Kirchbuchner, and Arjan Kuijper. To detect or not to detect: The right faces to morph. In Proceedings of International Conference on Biometrics (ICB), pages 1–8, 2019. 4

[14] Lisa DeBruine. debruine/webmorph morphing software: Beta release 2, Jan. 2018. 1, 4

[15] Lisa DeBruine and Benedict Jones. Face Research Lab London (FRLL) Image Dataset. May 2017. 4

[16] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4690–4699, 2019. 4

[17] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, 2017. 4

[18] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. In IEEE International Joint Conference on Biometrics, 2014. 4

[19] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face Demorphing. IEEE Transactions on Information Forensics and Security, 13(4):1008–1017, 2018. 1, 4

[20] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face demorphing in the presence of facial appearance variations. In Proceedings of European Signal Processing Conference (EUSIPCO), pages 2365–2369, 2018. 1

[21] Sylvain Gugger, Lysandre Debut, Thomas Wolf, Philipp Schmid, Zachary Mueller, Sourab Mangrulkar, Marc Sun, and Benjamin Bossan. Accelerate: Training and inference at scale made simple, efficient and adaptable. https://github.com/huggingface/accelerate, 2022. 4

[22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In Proceedings of Advances in Neural Information Processing Systems, volume 33, pages 6840–6851, 2020. 2

[23] Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In Proceedings of International Workshop on Similarity-Based Pattern Recognition, 2014. 5

[24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5967–5976, 2017. 3

[25] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In Proceedings of Advances in Neural Information Processing Systems, 2020. 4

[26] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 1, 4

[27] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality Adaptive Margin for Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 4

[28] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, 3rd International Conference on Learning Representations, (ICLR), 2015. 4

[29] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In Proceedings of International Conference on Learning Representations, (ICLR), 2014. 3

[30] Min Long, Jun Zhou, Le-Bing Zhang, Fei Peng, and Dengyong Zhang. Adff: Adaptive de-morphing factor framework for restoring accomplice's facial image. IET Image Processing, 18(2):470–480, 2024. 1

[31] Satya Mallick. Face morph using opencv —c++ / python. LearnOpenCV, 2016. 4

[32] Mehdi Mirza. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014. 1

[33] Tom Neubert, Andrey Makrushin, Mario Hildebrandt, Christian Kraetzer, and Jana Dittmann. Extended StirTrace Benchmarking of Biometric and Forensic Qualities of Morphed Face Images. IET Biometrics, 7:325–332, 2018. 4

[34] Fei Peng, Le-Bing Zhang, and Min Long. FD-GAN: Face de-morphing generative adversarial network for restoring accomplice' s facial image. IEEE Access, 2019. 1, 2

[35] Alyssa Quek. Face morpher. 1, 4

[36] R. Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch. Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images. In Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 1822–1830, 2017. 4

[37] K. Bommanna Raja, Matteo Ferrara, Annalisa Franco, Luuk J. Spreeuwers, Ilias Batskos, Florens de Wit, Marta Gomez-Barrero, Ulrich Scherhag, Daniel Fischer, Sushma Krupa Venkatesh, Jag Mohan Singh, Guoqiang Li, Loïc Bergeron, Sergey Isadskiy, Raghavendra Ramachandra, Christian Rathgeb, Dinusha Frings, Uwe Seidel, Fons Knopjes, Raymond N. J. Veldhuis, Davide Maltoni, and Christoph Busch. Morphing Attack Detection-Database, Evaluation Platform, and Benchmarking. IEEE Transactions on Information Forensics and Security, 16:4336–4351, 2020. 2

[38] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models . In Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 10674–10685, June 2022. 1

[39] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 10684–10695, June 2022. 3

[40] Eklavya Sarkar, Pavel Korshunov, Laurent Colbois, and Sébastien Marcel. Vulnerability Analysis of Face Morphing Attacks from Landmarks and Generative Adversarial Networks. arXiv preprint arXiv:2012.05344, 2020. 4

[41] Mallick Satya. Face morph using opencv —c++ / python. 2016. 4

[42] Guilherme Schardong, Tiago Novello, Hallison Paz, Iurii Medvedev, Vinícius da Silva, Luiz Velho, and Nuno Gonçalves. Neural Implicit Morphing of Face Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 7321–7330, June 2024. 1

[43] Ulrich Scherhag, Luca Debiasi, Christian Rathgeb, Christoph Busch, and Andreas Uhl. Detection of Face Morphing Attacks Based on PRNU Analysis. In Proceedings of IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(4):302–317, 2019. 4

[44] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, and Christoph Busch. Deep Face Representations for Differential Morphing Attack Detection. IEEE Transactions on Information Forensics and Security, 15:3625–3639, 2020. 4

[45] Nitish Shukla. SDeMorph: Towards Better Facial Demorphing from Single Morph. In Proceedings of IEEE International Joint Conference on Biometrics (IJCB), 2023. 1, 2, 5

[46] Nitish Shukla and Arun Ross. dc-GAN: Dual-Conditioned GAN for Face Demorphing From a Single Morph. arXiv preprint arXiv:2411.14494, 2024. 1

[47] Nitish Shukla and Arun Ross. Facial Demorphing via Identity Preserving Image Decomposition. In Proceedings of IEEE International Joint Conference on Biometrics (IJCB), 2024. 1, 2, 5

[48] Nitish Shukla and Arun Ross. Metric for evaluating performance of reference-free demorphing methods. In Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), 2025. 4, 5

[49] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. ArXiv, abs/1607.08022, 2016. 4

[50] Yuxin Wu and Kaiming He. Group normalization. In Computer Vision –ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIII, page 3–19, Berlin, Heidelberg, 2018. Springer-Verlag. 4

[51] Haoyu Zhang, Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Naser Damer, and Christoph Busch. MIPGAN—generating strong and high quality morphing attacks using identity prior driven gan. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(3), 2021. 1

[52] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Processing Letters, 23(10):1499–1503, 2016. 4

[53] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single Image Reflection Separation with Perceptual Losses . In Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4786–4794, June 2018. 5

[54] Zhengxia Zou, Sen Lei, Tianyang Shi, Zhenwei Shi, and Jieping Ye. Deep Adversarial Decomposition: A Unified Framework for Separating Superimposed Images . In Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 12803–12813, 2020. 2, 3, 5