可变形深度人脸特征插值编解码网络

高谦，刘嘉润，刘文杰，李瑞瑞
北京化工大学信息科学与技术学院，北京 100029

摘要：人脸属性编辑是一个具有挑战性的图像处理任务，使用传统的图像处理软件手工编辑，操作成本高。深度特征插值方法是通过深度学习将图像空间非线性的语义映射到隐空间线性特征再进行语义属性编辑的技术。该方法通用性较强，不需要设计特定网络结构，图像处理的速度较快，效果也比较好。遗憾的是，该方法由于难以将特征完全解耦，导致线性插值之后生成图像模糊或者存在伪影。针对这一问题，本文采用可变形的编解码器架构，通过无监督的训练方式，将图像分离为与外观相关的纹理信息和与形状相关的变形信息，解耦图像变形和纹理信息。同时创新性地使用与编码器对称的解码器结构，并引入从编码器到解码器的带过滤条件的信息通道帮助重建。本文提出的方法，可以获得耦合度更低、更有效的特征表示方法，重建的图像更加清晰；将该技术应用于人脸语义属性编辑，也大大消除了伪影的现象。

关键词：人脸风格迁移；特征插值；可变形自编码器；无监督解耦；对称解码器。

Abstract：Face attribute editing is a challenging image processing task. Manually editing which using traditional image processing software leads to high operating costs. The deep feature interpolation method is a technique of mapping the non-linear semantics of the image space to the linear features of the latent space through deep learning and then editing the semantic attributes. This method is more general, does not need to design a specific network structure, the image processing speed is faster, and the effect is better. Unfortunately, this method has difficulty in completely disentangling features, resulting in blurred images or artifacts after linear interpolation. To solve this problem, this paper adopts a deformable autoencoder architecture to separate images into appearance-related texture information and shape-related deformation information through unsupervised training, then disentangs image deformation and texture information. At the same time, it innovatively uses a symmetrical decoder structure to the encoder, and introduces an information channel with filtering conditions from the encoder to the decoder to help reconstruction. The method proposed in this paper can obtain a lower coupling and more effective feature representation, and the reconstructed image is clearer; applying this technology to edit semantic attributes of human faces also greatly eliminates the phenomenon of artifacts.

Key Words：face style transfer; deep feature interpolation; deforming autoencoder; unsupervisied disentangling; symmetric decoder.

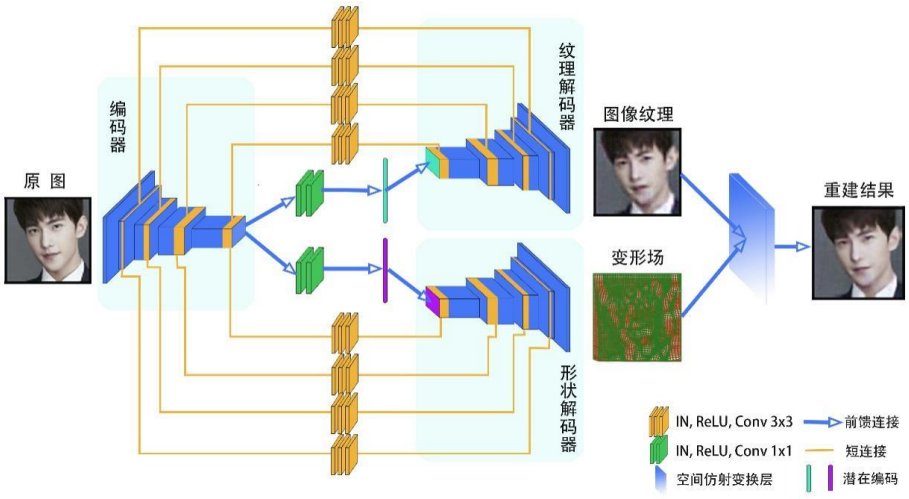

图 1.整体方法框架

1 引言

在信息社会快速发展的今天，图像作为一种信息传递的重要媒介，在信息时代发挥着重要的作用。人脸图像作为人的重要符号信息，具有广泛的艺术、社交、网络等应用。不同的使用场景也对人脸图像的风格提出了多样化的要求。证件照要求人像姿势端正，具有自然或者微笑的表情；宣传广告希望展示的人像具有大眼睛，瓜子脸，白皮肤等符合大众审美的特征；普通用户也希望通过自己不同风格的照片展示自己的个性。基于传统方法，人脸图像风格化往往需要复杂的人工修图，不仅要求操作者具有很高的水平，而且编辑方法和参数调整无法直接迁移应用到其它图片，因此难以自动批量处理。幸运的是随着深度学习和计算机视觉的发展，一系列图像处理的算法和模型得以提出。例如编辑肖像[1-2]，年龄[3]，纹理[4-5]，颜色[6]等方法。但这些算法专注于解决某一特定的人像编辑任务，且不能直接迁移到其他任务。因此，图像编辑转换的通用方法成为了计算机视觉与图像处理领域的研究热点。

现有的很多工作都希望通过线性插值的方法编辑人脸风格属性[14-17]。众所周知，像素空间非线性流形[8]的特点，使得线性插值的结果存在严重伪影，效果不佳。必须将像素空间非线性的语义映射到隐空间的线性特征才可以进行有意义的线性插值。卷积神经网络在图像分类任务上表现出色[9-11]，网络通过卷积提取图像特征，并依靠末端简单的线性层进行分类，即卷积神经网络将图像映射到一个类别信息可线性分离的深度空间。[14]表明在具有足够多图像类别的数据集上（例如 ImageNet[12]）训练的分类网络（例如 VGG[9]）可以学习到更通用的图像深度表示。因此可以使用预训练的分类网络作为图像到隐空间的映射，进而使用插值方法编辑图像[13-15]。但对于人脸图像来说，包含着外观纹理（例如肤色，毛发，光照等）和形状（如姿势，角度等）的复杂变化，单一映射难以解耦这些变化因子，插值的结果存在伪影。

本文受到可变形自编码器[17]解耦纹理外观和形状变化信息的启发，结合已有特征插值的方法，对可变形自编码器网络的结构进行改进，提出可变形的深度人脸特征编解码网络。主要创新点和贡献如下：

（1）提出了一种可变形深度人脸特征插值编解码网络，无监督地解耦图像的纹理特征与形状特征，降低了特征的耦合度，改善了线性插值的效果。

（2）结合图像语义分割领域的方法，在编码器和解码器部分层之间添加信息通道，使重建之后的图像更加清晰，保留原图像的必要细节。

（3）对信息通道中传递的特征进行过滤，减少用于插值的特征图个数，提高泛化能力和线性插值的鲁棒性，利于插值操作。

（4）设计出一种深度特征可分离的网络结构，可灵活选取部分或全部特征进行插值操作。引入插值强度参数实现灵活可控的图像编辑。

2 相关工作

产生令人信服的图像变化是计算机视觉和图形学领域一个活跃并且富有挑战性的课题。近年来，随着深度卷积神经网络在图像分类任务上取得越来越出色的表现，一系列借助卷积神经网络的图像编辑模型被提出。Gatys 等人开创性地提出了一种基于卷积神经网络的图像风格迁移方法[13]，通过预训练的 VGG 模型从内容图像和风格图像分别提取内容特征和风格特征，然后使用迭代优化的方式从随机图像开始生成具有原内容和目标风格的图像。Justin 等在 Gatys 算法的基础上，提出快速风格迁移算法[18]。该算法借鉴 Gatys 的方法中风格与内容的计算方式，使用 VGG 模型提取的深层抽象特征的 L2 损失作为感知损失训练指定风格的生成模型。提高了风格迁移的效率。 Upchurch 等人观察到分类网络末端线性分类器具有良好的性能，提出了深度特征插值（DFI）的方法[14]。该方法使用预训练的 VGG 网络作为特征提取器，通过 KNN 选取与测试图像相近的数据集子集，用于估计目标属性的深度特征表示。加和测试图像与目标属性的深度特征表示作为插值结果预测图像的深度特征表示。最后迭代优化随机图像与预测结果图像深度特征表示的误差，将梯度反向传播到随机输入图像使其逐渐逼近结果图像。Facelet-bank[15]方法采用对称的编码器解码

器结构，选取编码器部分卷积层的激活输出作为特征空间，经过中间卷积层与解码器对称位置级联。编码器和解码器执行一般操作，中间卷积用于建模不同的面部转换效果，实现了面部效果的快速编辑操作。

除此之外，生成模型的发展也为解决图像编辑转换提供了思路。比较典型的生成模型有生成对抗网络[19-23]和变分自编码器[16]。最初的生成对抗网络[19]使用多层感知机作为判别器和生成器的结构，训练不稳定，生成过程不可控，不具备可解释性。深度卷积对抗神经网络[20]的生成器和判别器使用卷积神经网络代替多层感知机，去掉池化层，替换全连接层为全局池化。图像的生成效果更好。但是输入为随机向量，无法控制最终生成的图像。Pix2pix[22]的方法使用条件生成对抗网络结构[21]，生成器和判别器的输入均为图像，该方法在图像着色、草图与真实图等图像转换任务有出色表现，但缺点是需要大量配对的训练数据。CycleGAN[23]使用循环一致性损失解决了无配对图像转换的问题，主要思想是最小化两次域转换之后与输入的差异。CycleGAN 可以成功地分离图像的风格和内容，对于图像翻译和风格转换效果很好。变分自编码器[16]的编码器通过学习特征向量的后验概率，能学习到有意义的特征表示。改变特征表示的某些维度可以让生成的图像产生语义变化，为线性插值的图像编辑方法提供了思路。可变形自编码器[17]将可变模板范例与自编码器架构相结合，使用两个独立的解码器网络对编码器输出的潜在编码解码得到图像的纹理和变形场，无监督的将图像的纹理和形状信息分离。

本文提出的模型结构在可变形自编码器的网络架构基础上进行改进，使用 1×1 的卷积层替换原网络的全连接层，采用 DenseNet 构造编码器与解码器，以获得其隐式的深度监督机制和模型参数的精简带来的易于训练的优点[11]。参考图像分割领域经典模型 U-Net[26]的做法在对称的编码器与解码器中间层添加信息通道，减少细节信息的损失，提高重建图像的清晰度。受到[28]中 ResPath 的启发，在信息通路上添加卷积层用于增强线性插值的鲁棒性。

插值操作借鉴 DFI 的思想和方法，但有以下几点不同：首先，DFI 方法以预训练的 VGG-19 部分卷积层的激活输出作为基础建立特征空间，而本文选取的是可变形编码器输出的潜在编码以及信息通道过滤后的输出；其次，本文建模目标属性深度表示时省略了对数据集的 KNN 搜索和按照测试图像进行人脸对齐这些比较耗时的步骤以提高测试时的效率。此外，DFI 采用梯度下降的方法最小化随机图像与目标图像深度特征表示的距离得到目标图像的估计，本文方法直接通过解码器输出纹理图像和变形场，再经过空间转换层变换直接输出重建图像，与 DFI 相比减少了测试阶段的耗时。

3 本文方法

整体方法框架如图 1 所示。本文在可变形自编码器网络结构基础上进行改进，在编码器与解码器对称位置添加信息通道（如图 1 中黄线所示），在重建图像时保留原图像的必要细节，使重建的图像更加清晰。同时在信息通道上添加卷积层，实现对特征的筛选过滤，提高线性插值的鲁棒性，增强模型的泛化能力。

3.1 可变形自编码器

可变形自编码器（Deforming Autoencoder）将可变模板范例与自编码器架构相结合。可变模板范例将图像的生成过程解释为在无变形信息的坐标系（模板坐标系）上合成外观纹理图像，然后将描述形状可变形的变形信息合成变形场，最后纹理图像在变形场的作用下生成最终的图像的过程。生成过程表示如下：

$$I(p) \simeq T\big(W(p)\big) \tag{1}$$

其中 p 表示图像矩阵的坐标，$W(p)$表示变形场将位置 p 映射到到的新位置，$T(p)$ 表示位置 p 的外观纹理。I 表示结果图像。

具体实现时，输入图像经过编码器得到其潜在编码表示 $\boldsymbol{Z}$，为了分离变形和外观纹理信息，使用两个全连接网络将低维潜在编码分为两部分 $\boldsymbol{Z}=[\boldsymbol{Z}_T,\boldsymbol{Z}_S]$，分别输入到两个独立的解码器网络 D_T 和 D_S 得到外观纹理图像和变形场，然后通过空间转换层按照式（1）的方式得到重建图像。

变形场的建模需要解决诸多问题，例如不合适的正则化导致的变形场分布不均匀；或者产生非微分同胚的变形场，导致最后将连续的

纹理图像传播到不连续的区域。为了解决这个问题，[17]中提出了差分解码器的概念，即不用形状解码器直接预测变形场$W(p)=(W_x(x,y),W_y(x,y))$，而是用其生成变形场的空间梯度$\nabla_x W_x$和$\nabla_y W_y$。其中∇_c代表空间梯度向量的第 c 个分量。这两个值用于测量连续像素的位移，举例来说，$\nabla_x W_x$=1 表示沿水平轴平移，$\nabla_x W_x$=2 表示沿水平轴平移两个单位。而$\nabla_x W_x$=-1 表示左右翻转（当然这是不合理的，不应该出现这样的情况）。$\nabla_y W_y$功能类似，作用于垂直的方向。通过控制$\nabla_x W_x$和$\nabla_y W_y$，可以防止变形场翻转折叠（$\nabla_x W_x$或$\nabla_y W_y$为负数的时候）或者过度拉伸（$\nabla_x W_x$或$\nabla_y W_y$值过大）。在具体实现时将形状解码器的输出经过 ReLU 模块来强制$\nabla_x W_x$和$\nabla_y W_y$在水平和垂直的相邻像素上执行正偏移，确保变形场不折叠。通过引入平滑度损失防止变形场过度拉伸。最后，将形状解码器的输出通过 ReLU 模块后的空间积分层生成最终的变形场。

3.2 损失函数

本文继续使用可变形自编码器的损失函数无监督的进行训练，可变形自编码器的损失函数包括两部分：

$$E_{DAE}=E_{Reconstruction}+\lambda E_{Warp} \quad (2)$$

其中λ用于设置变形损失所占的比重，在实验时将其设置为 0.01。$E_{Reconstruction}$代表图像重建损失，通过最小化重建图像与输入图像之间的 L_2 距离对网络参数进行反向传播。重建图像损失表示如下：

$$E_{Reconstruction}=\left\|I_{output}-I_{input}\right\|^2 \quad (3)$$

其中I_{output}和I_{input}分别表示可变形自编码器重建的图像和输入的原图。

变形损失E_{Warp}主要体现在平滑度损失上，平滑度损失用于防止生成的变形场过度拉伸。该损失使用水平和垂直的差分变形场的总变化范数衡量：

$$E_{smooth}=\|\nabla W_x(x,y)\|_1+\left\|\nabla W_y(x,y)\right\|_1 \quad (4)$$

3.3 模型结构

可变形自编码器网络的结构如图 2 所示，模型结构包括编码器 E，纹理解码器 T 与形状解码器 W。全连接网络将图像潜在编码分为纹理编码和形状编码，分别通过纹理解码器与形状解码器得到纹理图像与变形场，空间转换层使用变形场对纹理图像进行变换生成最终图像。

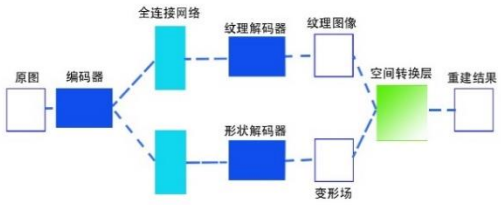

图 2.可变形自编码器结构

本文对可变形自编码器的结构进行改进，使用 DenseNet 替换 CNN 网络构造编码器与解码器，DenseNet 的网络结构如图 3 所示。替换 DenseNet 末端的全连接层为 1×1 的卷积层。在编码器每一个过渡块的输出与解码器对称位置密集块的输入之间添加短连接建立信息通道，如图 1 黄色线所示。并在信息通道上添加卷积层。卷积层的作用是对特征图进行过滤，同时增强插值模型的泛化能力。对于 64×64 的输入图像，模型编码器与解码器网络详细参数配置如表 1、表 2 所示。

表 1.编码器网络参数配置

层	输出	网络结构
卷积层	32×32	3×3 conv, stride=2
密集块 1	32×32	{3×3 conv} × 6
过渡块 1	16×16	{1×1 conv, 2×2 avg-pool}
密集块 2	16×16	{3×3 conv} × 12
过渡块 2	8×8	{1×1 conv, 2×2 avg-pool}
密集块 3	8×8	{3×3 conv} × 24
过渡块 3	4×4	{1×1 conv, 2×2 avg-pool}
密集块 4	4×4	{3×3 conv}× 16
过渡块 4	1×1	{1×1 conv, 4×4 avg-pool}

表 2.解码器网络参数配置

层	输出	网络结构
转置卷积	4×4	4×4 conv-T, padding=0
密集块 1	4×4	{3×3 conv} × 16
过渡块 1	8×8	4×4 conv-T, stride=2
密集块 2	8×8	{3×3 conv} × 24
过渡块 2	16×16	4×4 conv-T
密集块 3	16×16	{3×3 conv} × 12
过渡块 3	32×32	4×4 conv-T
密集块 4	32×32	{3×3 conv} × 6
过渡块 4	64×64	4×4 conv-T
转置卷积	64×64	3×3 conv-T

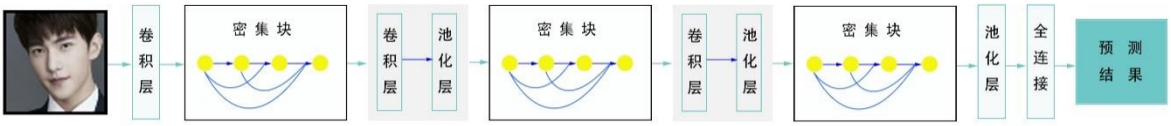

图 3.DenseNet 网络结构

本文中 DenseNet 的参数参考 DenseNet-121 的参数配置，所不同的是输入的图像大小与[11]中不一样。增长率设置为 12。表 1、表 2 中，conv 表示 IN-ReLU-Convolution 序列，参考[24]，实验时采用实例归一化层 IN 替代批归一化 BN 以获得更好的生成效果。密集块后面的数字表示堆叠的卷积层数，没有特别说明的情况下 stride 和 padding 默认取 1。

在实验中我们发现修改之后的模型显著提高了重建图像的质量，同时网络结构也更加利于插值操作，详见第 4 节实验。

4 实验与分析

在本节中，使用改进后的模型针对胡须、表情、年龄三个属性进行插值效果验证，并与更改前的可变形自编码器的图像重建与特征插值的结果进行对比。并开展一系列消融实验表明模型结构的改进对于结果的影响。实验的系统环境为 Ubuntu 18.04，内存 16G，Nvidia GTX960 显卡，显存 4G，Pytorch 1.3 版本，Python 3.7 版本。

4.1 数据集使用

CelebA：全称为 CelebFaces Attribute，是一个大型的具有面部属性标注的数据集，包含 10,177 个名人身份的 202,599 张人脸图像，每张图像具有 40 项属性的标注。CelebA 由香港中文大学开发提供，可以用于人脸相关的计算机视觉训练和测试，例如面部属性识别，人脸检测，landmark 定位以及面部编辑和合成等。

AGFW-v2：The AginG Faces in the Wild，包括图像集和视频集，本文使用数据集的图像集，其中包含 36299 张图像。这些图像的年龄从 10 到 64 岁不等，按照性别划分为两个子集，每个子集按照 5 的年龄跨度划分为 11 个年龄组。

JAFFE：The Japanese Female Facial Expression，数据库包含 213 张由 10 位日本女性构成的 7 种面部表情（快乐，生气，悲伤，厌恶，惊喜，恐惧和中性无表情）的 256×256 灰度图像。

CK+：此数据集在 Cohn-Kanade Dataset 的基础上扩展而来，包含 137 个不同身份的人脸图像，共 593 个由视频裁剪出来的视频帧序列，其中 327 个序列有表情标注，每一个序列的最后一帧都包含基本动作单元的标注。

4.2 插值属性特征表示提取

为了方便描述，本文将编码器输出的潜在编码以及中间各个信息通道过滤后的输出特征图组成的整体称为特征空间Φ。为了方便使用不同深度的特征进行插值，特征空间中的特征之间彼此分离，即：

$$\Phi = [\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_{latent}] \tag{5}$$

其中Φ_i（i=1,2,3,4）表示通路卷积输出特征图，Φ_{latent}表示编码器输出的潜在编码。

由于缺少有效的方式建模插值属性 A 的潜在编码，本文采用与 DFI 类似的方法：将具有属性 A 和不具有属性 A 的两组图像分别经过编码器得到对应的潜在编码表示，然后各组取均值再做减法，得到属性 A 特征表示的估计值：

$$\varphi_A = \overline{\Phi(S_A)} - \overline{\Phi(S_{NA})} \tag{6}$$

其中S_A和S_{NA}分别表示具有属性 A 和不具有属性 A 的图像集合，特别注意的是，在选取S_A和S_{NA}时，本文省略了使用 KNN 选取与测试图像相近的数据集子集和参照测试图像对选取的人脸进行关键点对齐等操作。省略数百个面部图像的卷积和对齐操作减少了测试阶段耗时。为了得到更好的属性特征表示估计值，与 DFI 的方法一致，本文在实验时选取的两组图像除插值属性不同之外，其余属性尽可能一致。

4.3 实验结果

本小节以对比实验的形式展示图像重建与胡须插值的结果。并以人脸表情，人脸年龄作为目标属性展示使用改进后的模型插值的效果。

4.3.1 重建结果对比

实验时采用[17]官方提供的代码和分享的在 CelebA 数据集上预训练的模型。在保持其它条件一致的情况下进行重建实验。结果对比如

图 4 所示。其中 A 为原图，B 为使用改进前模型的重建结果，C 为改进后模型的重建结果。由图 4 可知，更改结构之后的网络重建效果更加清晰，保留了原图像的必要细节。

图 4.重建结果对比

尽管人的视觉感官是判断图像效果的最好标准，但本文依然引入了客观评价指标来评估重建图像的结果质量。实验时使用百度 AI 开放平台的人脸对比接口[7]进行人脸相似度评分统计。随机从 CelebA 数据集中选取 5000 张图像样本，分别使用改进前和改进后的模型对图像进行重建，重建的图像分别和原图做相似度统计，百度 AI 接口返回的相似度级别统计如图 5 所示。

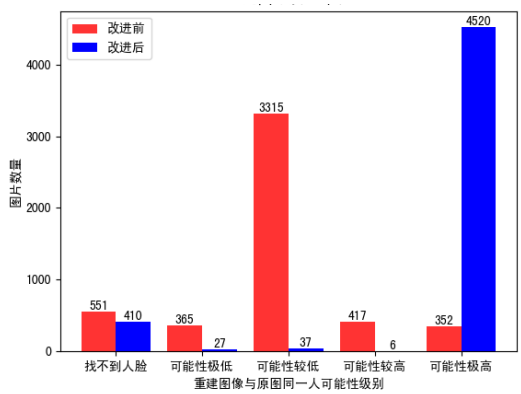

图 5.重建图像与原图相似度级别统计

图 6 显示了相似度值的分布，纵轴表示频数，横轴表示相似度的分布区间，从图中可以看出，改进后模型重建图像与原图的相似度主要分布在 90%以上的位置，重建的结果更好。

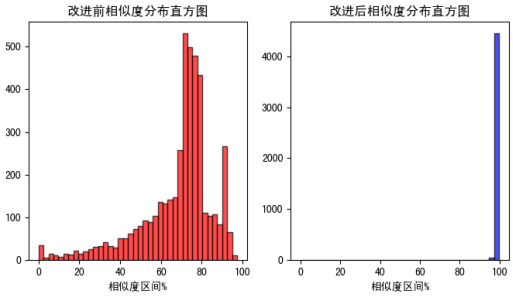

图 6.重建图像与原图相似度分布直方图

图 7.胡须插值结果

4.3.2 胡须插值效果

本文使用 CelebA 数据集对更改结构后的模型训练 60 个 epochs，进行胡须插值实验。估计胡须属性的潜在表示时，S_A和S_{NA}分别选取 200 张图像，因为胡须多为男性特质，因而全部采用男性图像进行实验。实验时发现 CelebA 数据集标注存在不准确的情况，部分标记为无胡须的图像实际有胡须。为了减少错误标记对结果的影响，实际实验时使用搜索引擎搜索“男星有胡子”和“男星”的结果图像进行属性向量估计。插值的结果如图 7 所示。

4.3.3 表情插值效果

使用 JAFFE 数据集中全部带表情的图像对改进后的网络训练 50 个 epochs，选取开心，厌恶，生气三个表情进行插值实验，实验结果如图 8 所示。其中，A 为原图，BCD 分别为开心，厌恶，生气的插值结果，E 为 ABCD 第一列结果的放大展示。

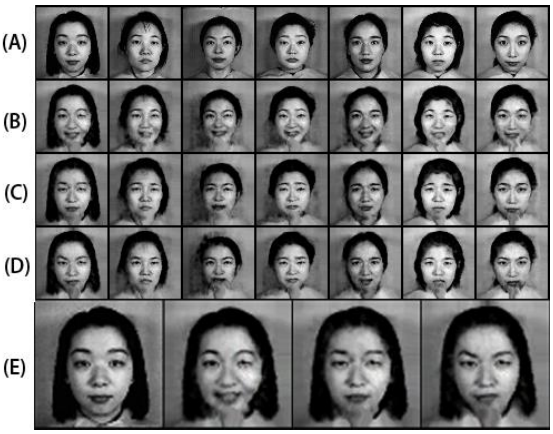

图 8.表情插值结果

此外，针对 CK+数据集也进行了相关实验，结果如图 9 所示：（ABCD 分别为开心、伤心、生气、厌恶的插值结果）。

为了消除主观评价的影响，与图像重建评估类似，实验时使用百度 AI 开放平台的人脸检测接口[25]用于检测插值结果人脸的表情属性，并与真实标注数据作对比，统计结果图 10 所示：

图 9.CK+数据集上表情插值结果

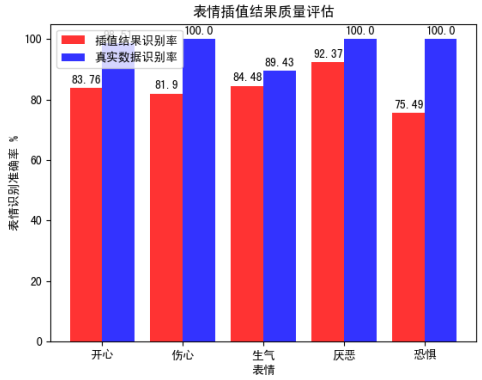

图 10.表情插值结果识别准确率统计。

4.3.4 年龄变老

由于 AGFW 数据集中女性图像的年龄变化不如男性图像明显，因而实验时随机选取数据集中 11 个年龄段各 500 张男性图像作为训练数据，在 CelebA 数据集训练的网络基础之上 Fine-tune。随后选取 15-19 以及 40-44 两个年龄段各 100 张男性图像使用 4.2 节的方法进行变老属性潜在表示的估计，插值结果如图 11 所示。每一组图像中，a 为原图，b 为插值之后的结果图像。

图 11.年龄插值结果展示

4.3.5 与原可变形自编码器的插值效果对比

本文采用[17]中官方提供的代码和分享的在 CelebA 数据集上预训练的模型。保持其它条件一致的情况下进行胡须插值实验。结果如图 12 所示。通过实验发现，由于原来的可变形自编码器结构损失了过多的面部细节，难以通过对潜在编码插值获得具有相应属性的重建图像，效果与改进后模型的插值结果（如图 7 所示）相比有很大的差距。

图 12.使用改进前的可变形编码器插值结果

4.3.6 调整属性插值强度参数对比实验

本文提出的的网络通过在插值时设置插值强度参数，无需重新训练网络即可控制最终结果图像的插值效果强度级别：

$$\phi = \Phi_{encoder} + \boldsymbol{\theta}\varphi_A \qquad (7)$$

其中$\Phi_{encoder}$为编码器输出特征图，θ为插值强度参数，φ_A通过式（6）计算得出。另外，特征空间中的特征图之间彼此分离，因而可以为不同深度的特征单独设置插值强度，或者根据需要选取某一层或者某几层的特征图进行插值，即$\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_{latent}]$，分别作用于式（5）中每个特征图分量。不同插值强度的实验效果如图 13 所示(插值强度从左到右逐渐增大)。

图 13.不同强度的插值效果

4.4 消融实验

本小节通过对比实验，展示模型结构的各处改动对结果的影响。

4.4.1 使用 DenseNet 替换 CNN

本文使用 DenseNet 替换 CNN 用于构建编

解码器结构，与[11]中实验结果一致，通过实验我们发现，DenseNet 可以加快模型收敛，模型更容易训练（如图 14 所示）。除此之外，由于不需要重新学习冗余的特征图，在相同的网络深度情况下，DenseNet 的连接模式使其与传统卷积网络相比，参数更少（如表 3 所示）。

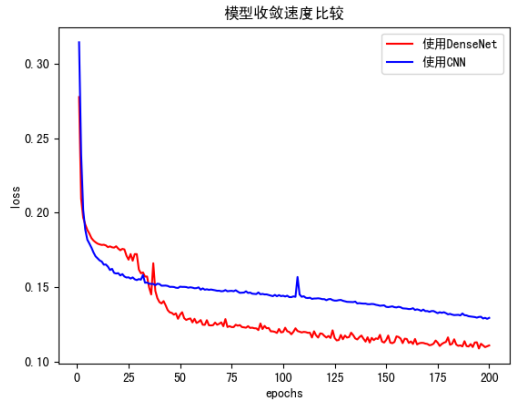

图 13.不同强度的插值效果

表 3：模型参数数量对比

模型	结构	参数数量
[17]中的模型	编码器	13.59M
	解码器	28.98M
使用 DenseNet	编码器	1.77M
	解码器	6.07M

4.4.2 信息通路的作用

本文在原模型结构的基础之上，在编码器与解码器对称位置添加信息通路，与[26-28]中的结论一致，实验结果表明：添加信息通路之后的网络能够减少空间信息的损失，重建之后的图像可以保留更多原图像的细节（如图 14 所示，A 组无通路，B 组有通路）。

图 14.有无信息通路的模型重建效果对比。

对于使用特征插值的方法进行图像编辑的模型，重建图像保留原始图像的必要细节（例如表情，胡须，年龄等属性）是必要的。因而信息通路是本文提出模型必不可少的结构。

4.5 结果分析

通过图像重建的对比实验发现，原可变形自编码器由于多次池化操作损失了过多的图像信息，导致解码器重建的图像缺乏细节，与原图像的相似度较低。改进后的模型通过添加带过滤条件的信息通道保留了图像的必要细节，因而重建结果比较好。图像插值的对比实验表明，改进后的模型插值结果更好，这与模型的结构有关：改进后的模型在经信息通道传输过滤的特征输出以及编码器的潜在编码上插值，运用了网络不同深度，不同抽象程度的特征，鲁棒性较好。而改进前的模型只能在编码器潜在编码上插值，特征抽象程度高，可能对某些属性变化不敏感。因此，改进后的模型更适合特征插值的图像属性编辑任务。另外，在实验时，如果测试人像姿势严重偏离平均水平，使用改进后的模型插值也会产生伪影。这说明网络对特征的解耦依旧不充分，变形场没有完全捕获人像的姿势信息，形状信息与纹理信息没有完全分离。之后的工作将围绕进一步解耦图像纹理与形状信息以及增大处理图像的分辨率之后可能造成的重建图像不清晰问题展开研究。

5 结论

本文在可变形自编码器网络的基础之上提出了可变形深度人脸特征插值编解码网络。使用无监督的方式解耦图像的纹理外观和形状变形信息。参考 U-Net 的结构在编码器与解码器之间建立信息通道改善图像的重建质量，并引入特征过滤层提高模型的泛化能力，通过实验证明了本文方法比可变形自编码器更适合应用插值的方法进行图像编辑。为进一步改进和使用可变形自编码器进行特征插值任务提供了思路。

参考文献

[1] Suwajanakorn S , Seitz S M , Kemelmacher-Shlizerman I . What Makes Tom Hanks Look Like Tom Hanks[C]// IEEE International Conference on Computer Vision. IEEE, 2015.

[2] Kemelmacher-Shlizerman, Ira. Transfiguring

portraits[J]. Acm Transactions on Graphics, 35(4):1-8.

[3] Antipov G , Baccouche M , Dugelay J L . Face aging with conditional generative adversarial networks[C]// IEEE International Conference on Image Processing. IEEE, 2017.

[4] Bellini R , Kleiman Y , Cohen-Or D . Time-varying weathering in texture space[J]. Acm Transactions on Graphics, 2016, 35(4):1-11.

[5] Aittala M , Aila T , Lehtinen J . Reflectance Modeling by Neural Texture Synthesis[J]. ACM Transactions on Graphics, 2016, 35(4):65.1-65.13.

[6] Zhang, Richard, Isola, Phillip, Efros, Alexei A. Colorful Image Colorization[J].

[7] https://ai.baidu.com/tech/face/compare.

[8] Weinberger K Q , Saul L K . Unsupervised Learning of Image Manifolds by Semidefinite Programming[J]. International Journal of Computer Vision, 2006, 70(1):77-90.

[9] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer ence, 2014.

[10] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.

[11] Gao Huang, Zhuang Liu, Laurens Van De Maaten. Densely Connected Convolutional Networks[C]// Cvpr. 2017. 1.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012. 1.

[13] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576, 2015. 2, 3,7.

[14] Upchurch, Paul, Gardner, Jacob, Pleiss, Geoff. Deep Feature Interpolation for Image Content Changes[J],In CVPR,2018.

[15] Chen Y C , Lin H , Shu M , et al. Facelet-Bank for Fast Portrait Manipulation[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[16] Diederik P Kingma, Max Welling. Auto-Encoding Variational Bayes. Machine Learning, arXiv:1312.6114,2013.

[17] Shu Z , Sahasrabudhe M , Guler A , et al. Deforming Autoencoders: Unsupervised Disentangling of Shape and Appearance[J]. 2018.

[18] Johnson J , Alahi A , Fei-Fei L . Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. 2016.

[19] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, Generative Adversarial Networks. In Machine Learning,2014.

[20] Alec Radford, Luke Metz, Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In Machine Learning, 2016.

[21] Mehdi Mirza, Simon Osindero. Conditional Generative Adversarial Nets. In Machine Learning, 2014.

[22] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In Computer Vision and Pattern Recognition, 2017.

[23] Zhu J Y , Park T , Isola P , et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. 2017.

[24] Ulyanov D , Lebedev V , Vedaldi A , et al. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images[J]. 2016.

[25] https://ai.baidu.com/tech/face/detect.

[26] Ronneberger O , Fischer P , Brox T . U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. 2015.

[27] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation[J].2018.

[28] Nabil Ibtehaz, M. Sohel Rahman. MultiResUNet : Rethinking the U-Net Architecture

for Multimodal Biomedical Image Segmentation[J].2019.

作者简介

李瑞瑞，1987 年生，女，博士，讲师，主要研究方向为遥感图像语义分割、全景分割。
E-mail:ilydouble@gmail.com.
高谦，男，大四年级本科生。
E-mail:qianqianjun0329@163.com.
刘嘉润，男，硕士研究生。主要研究方向为图像生成，风格迁移，噪声数据处理。
E-mail: jiarunliu@foxmail.com.
刘文杰，男，硕士研究生。
E-mail: buctliuwenjie@qq.com.