PSFRGAN文档

背景

三维超声重建图像在医学上的应用十分广泛，主要应用于人体各器官与部位的无创检查等。目前三维超声重建图像的噪声较大，且不同角度的超声切片能够提供的信息有限，需要专业人士才能识别，导致医生与患者之间沟通困难，患者无法直观得知病情。在胎儿超声领域，此现象尤为明显。胎儿的面部重建精度较低，无法给用户（主要为孕妇、家属等）直观体验。当前的人像超分辨率算法对于三维超声重建图像优化问题存在不适配、优化效果不明显等问题。本文将在渐进式深度神经网络PSFR-GAN的基础上，利用深度学习以及数据驱动的思想，提出针对提升胎儿三维超声重建图像的清晰度的方法，并通过实验对其进行了验证，在视觉效果和数据指标上都有了较为明显的提升。同时，此方法可以通过与传统图像优化算法以及迁移学习的数据集扩充方法一起协作，以进一步提升超声重建图像清晰度。

目录

[1. 绪论 1](#_Toc107530501)

[1.1 课题研究背景及意义 1](#_Toc107530502)

[1.2 图像优化算法国内外研究现状 2](#_Toc107530503)

[1.3论文主要研究内容和贡献 3](#_Toc107530504)

[1.4论文组织结构 4](#_Toc107530505)

[2. 基于深度学习的图像优化算法理论分析 5](#_Toc107530506)

[2.1三维超声重建图像优化问题描述 5](#_Toc107530507)

[2.2 PSFR-GAN网络结构及改进 6](#_Toc107530508)

[2.2.1 PSFR-GAN的网络结构 6](#_Toc107530509)

[2.2.2 PSFR-GAN的损失函数 7](#_Toc107530510)

[2.2.3 PSFR-GAN优化流程 8](#_Toc107530511)

[2.2.4 对于PSFR-GAN的改进 9](#_Toc107530512)

[2.3 传统方法预处理 10](#_Toc107530513)

[2.4 风格迁移扩充数据集 10](#_Toc107530514)

[3. 图像优化实验结果以及分析 12](#_Toc107530515)

[3.1 实验数据集 12](#_Toc107530516)

[3.2 三维超声重建图像优化评价标准以及基线 13](#_Toc107530517)

[3.2.1 三维超声重建图像优化评价标准 13](#_Toc107530518)

[3.2.2 三维超声重建图像优化基线 14](#_Toc107530519)

[3.3 实验结果及分析 14](#_Toc107530520)

[环境要求 18](#_Toc107530521)

[测试运行 18](#_Toc107530522)

[自动分割 18](#_Toc107530523)

[手动分割 18](#_Toc107530524)

[测试设备及处理时间 25](#_Toc107530525)

[训练模型 26](#_Toc107530526)

[准备训练数据 26](#_Toc107530527)

[训练PSFRGAN 26](#_Toc107530528)

[评测结果 27](#_Toc107530529)

[参考文献 28](#_Toc107530530)

1. 绪论

1.1 课题研究背景及意义

三维超声重建图像在医学上的应用十分广泛，主要应用于人体各器官与部位的无创检查等。目前三维超声重建图像的噪声较大，且不同角度的超声切片提供的信息有限，需要专业人士才能识别，导致医生与患者之间的沟通困难、患者无法直观得知病情等问题。在胎儿超声领域，此现象尤为明显。

在临床诊断中，医生往往会使用由超声仪扫描得到的的原始超声图像进行诊断，而不会使用三维超声重建图像。但原始的超声图像仅能表现扫描部位的横截面，对其进行观察需要专业的技能，导致医生在与患者沟通时常常出现沟通困难的问题。三维超声重建图像可以通过三维建模，更直观地呈现扫描部位的真实情况，从而部分地解决这个问题，医生可以对照较易查看的三维超声重建图像与患者进行沟通。尽管三维超声重建图像已经开始应用于临床诊疗，但目前三维超声重建图像仍然存在清晰度较低、用户体验较差、很多图像中关键的信息无法清晰地呈现等问题。

在生成胎儿人脸三维超声重建图片时，目前主流的做法是使用超声仪器在孕妇腹部进行三个维度的扫描，生成三张超声图像。根据这些图像，扫描仪器可以通过内置的计算机软件程序进行像素点匹配，从而生成三维空间中的胎儿模型，再通过计算机渲染成二维图像。在上述过程中，由于超声图像本身的精度有限，且在像素点匹配时，为了减少计算量以及提高鲁棒性，往往会进行插值或平滑化等操作，进一步降低了重建的三维模型的精度。因此，在超声图像最终呈现为三维超声重建图像后，胎儿面部的精度很低，视觉效果较为模糊。此外，由于胎儿的姿势不一、扫描角度常有差异等原因，最终的三维超声重建图像中，胎儿的脸部常被遮挡或无法获取正面图像，导致胎儿的面部不仅无法清晰地呈现给用户，也无法被当前常用的人像优化算法处理。

近年来，深度学习在人像优化、自然图像优化领域取得了较大的突破，也在社会生活中得到了广泛应用。将深度学习方法应用于三维超声图像中预期能够得到较好的结果。但目前主流的基于深度学习的图像优化算法依然存在以下几个问题：

1. 目前大多基于深度学习的人像优化算法主要针对于自然图像或成年人人像，对于婴儿或胎儿人像优化效果较差。

2. 由于三维超声重建图像的模糊来源与自然图像的模糊来源不同，目前主流的基于深度学习的人像优化算法无法应用于胎儿的三维超声重建图片中，因此针对胎儿面部的三维超声重建图像的优化算法目前还处于空白阶段。

3.基于深度学习的算法需要使用大数据驱动，在应用之前需要使用大量的数据训练模型。但目前暂无公开的数据集可供使用，且能够获取到的三维超声重建图像也十分有限，无法支撑深度学习训练需要的数据量。

由于目前主流深度学习方法无法直接应用于三维超声重建图片场景中，而对于三维超声重建图片的优化在临床中的需求十分迫切，找到一种能够使用少量三维超声重建图片即可完成训练、且能够应用于此种场景的深度学习方法显得十分重要。

1.2 图像优化算法国内外研究现状

目前主流的研究趋势趋向于将人像优化与图像优化区分开来，原因在于人像优化相比于自然图像优化有人脸先验信息存在。因此对于人像优化任务来说，可以利用图像中存在的人脸先验信息来进行优化。目前主流的人脸图像优化工作已经广泛使用了标记人脸信息的数据集来进行训练以及验证。但不同的数据集可能存在不同数量的标记分类，Helen[1]使用了194种标记，而CelebA[2]只使用了5种标记。主流的人像数据集使用的标注人脸先验信息的方法主要有人脸标签（Facial Landmarks）、人脸热图（Facial Heatmaps）和人脸分割图（Facial Parsing Maps）。

主流的人像优化算法可以分为以下几类：

1. 基于基础卷积神经网络（CNN）[8]的人像优化算法。此类算法较少使用人脸先验信息，仅通过像素级别的Loss对网络参数进行优化。此类方法具有代表性的有针对于人脸全局优化的BCCNN[5]、针对于人脸局部优化的VDSR[6]、以及综合全局和局部方法的DPDFN[7]。

2. 基于对抗生成网络（GAN）[9]的人脸优化算法。此类算法使用一个生成器网络和一个辨别器网络。在训练时，GAN将需要优化的图像作为输入，输入到生成器中，由生成器网络来生成高清图像，再用辨别器网络来输出生成器网络的Loss，以此来训练生成器。

3. 基于强监督学习的人脸优化算法。通常情况下，基于深度学习的人脸优化算法能够将低清晰度的图像LR映射到高清晰度的图像HR，但在这个过程中，会忽略人脸各个部位之间的依赖性。Attention-FH[10]则通过强监督学习解决了这个问题。Attention-FH使用了两个子网络，其中一个将会在人像中找到需要优化的区域，另一个子网络则负责优化这个区域。

以上几类人像优化算法分别从不同的角度尝试解决人像优化的问题。其中，较为突出的工作有PSFR-GAN[4]以及Bring Old Photos Back to Life[15]等，本文主要参考了PSFR-GAN的工作（3.2.2节中展示了上述两个工作在三维超声重建图片上的优化结果的指标，两者的优化结果指标都较低，且Bring Old Photos Back to Life结果的指标略微强于PSFR-GAN。由于PSFR-GAN视觉效果较好，且更接近Ground Truth，因此本文选用PSFR-GAN进行改进）。

总结前述已有工作可以得出以下分析：

1. 其中基于卷积神经网络的工作更集中于浅层的信息，因此通常情况下它们的结果评价指标较基于对抗生成网络的结果更高。

2. 对抗生成网络的工作更集中于深层的信息，因此它们的结果相对来说视觉效果更好。

3. 相较于以上两种算法，基于强监督学习的工作处理速度更快，但需要更长的训练时间。

1.3论文主要研究内容和贡献

本论文的主要研究内容是将渐进式深度神经网络PSFR-GAN[4]应用于三维超声重建图像的优化工作上，主要有以下贡献：

1. 本论文提出的基于PSFR-GAN的模型在经过调整与改进后，通过fine-tuning已有模型，在真实的三维超声重建图像上取得较好的结果。

2. 在面对仅有较少的数据集的情况，本工作能够与传统方法以及风格迁移扩充数据集的方案一起协作，将其较好地适配于医学图像增强与优化的领域。

3. 此外，通过使用本研究提出的迁移学习、传统方法与深度学习相结合的思路，本研究的成果也可以扩充到其他需要数据驱动算法但缺少数据集的高精尖难的领域中，进一步深化深度学习在专业领域中的推广以及应用。

本研究提出的模型能够切实提升胎儿三维超声重建图像的清晰度，可以直接应用于临床仪器，参与到医学影响的优化工作中。胎儿三维超声重建图像清晰度的提升，能够更直观地面向患者进行展示，也可以根据用户需求用作留念等。此外，本研究提出的模型经过训练以后，也可以应用于其他需要清晰化的三维超声重建图像中，能够进一步提升当前医疗影像的清晰度。

1.4论文组织结构

本文第一章主要描述了论文的背景、研究意义以及国内外的相关研究情况，阐明了本文的主要研究内容、贡献以及本文的组织结构。第二章则论述了深度学习的相关原理，从理论的角度解释了PSFR-GAN的模型以及网络结构，同时也简要介绍了需要与本文协作的传统方法以及迁移学习扩充数据集的工作。第三章则主要讲述了本文提出的模型处理得出的结果，以及相关的评价标准、基线以及实验数据等，并进一步对实验数据以及实验结果进行了分析。第四章主要总结了本文的主要成果和不足之处，并讨论了在此领域未来可能的研究方向以及本工作未来需要精进的部分。

2. 基于深度学习的图像优化算法理论分析

在本章中，首先对本文需要解决的三维超声重建图像优化的问题定义，并阐明其与人像优化问题以及自然图像优化问题的区别。其次将详细描述PSFR-GAN的网络结构，及其工作的过程。在本章的最后两个部分，简要描述与本文提出的模型协同工作的传统方法预处理以及风格迁移扩充数据集的工作。

2.1三维超声重建图像优化问题描述



**图1 三维超声重建图像与自然图像的对比**

如图1所示，三维超声重建图像与自然图像有很大的不同，所以三维超声重建图像的人像超分辨率的问题定义与普通的图像超分辨率也有所不同。对于普通的图像超分辨率问题和本研究需要解决的超声重建图像的超分辨率问题来说，可以用以下公式[3]描述：

(1)中，表示人像优化算法，表示这种算法需要的参数。(2)中，和分别表示重建图像和原始高分辨率的图像；表示它们的误差或损失函数。其中需要注意的是，的来源如下：

其中，表示数据保存为图片格式时需要的压缩处理，表示卷积操作，表示卷积核，表示下采样，表示噪音。

但对于超声重建图像重建的婴儿图像来说，低通图像的噪声来源可以用如下的公式表示[7]：

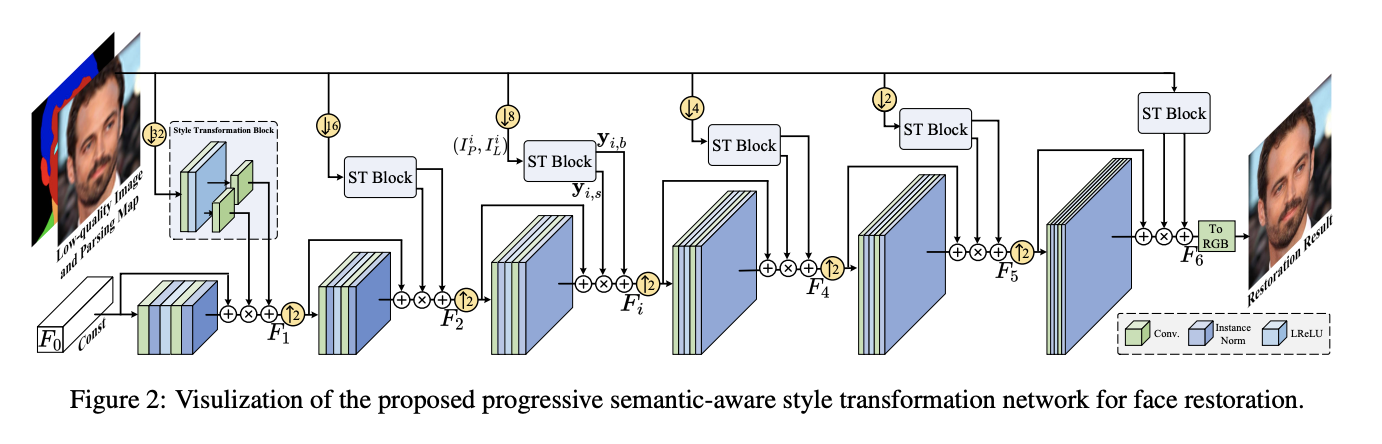
其中，、和分别可以表示三张从不同维度扫描得出的超声重建图像。通过对这三张图像的操作（即公式(4)中的操作）得出重建的点云模型。这个三维模型经过投影的操作（即公式(4)中的操作）以后才能得出正常的图像。经过和普通图像一样的卷积、下采样以及添加噪声的操作以后才会产生本研究需要处理的图像。其中，三维超声重建图像的模糊来源主要来自重建点云的过程。

综上，本研究所需要解决的问题就是找到一个方法，且它包含一组参数。通过训练可以找到，使得低通图像经过处理后得出的结果与高清图像的损失函数的值最小。

2.2 PSFR-GAN网络结构及改进

近年来涌现了许多优秀的人像优化方法，其中PSFR-GAN[4]具有一定的代表意义。C. Chen 等人提出了一种如图2所示的渐进式的网络结构PSFR-GAN，能够较好地恢复人像照片。PSFR-GAN主要使用已经预训练好的维的参数来与下采样后的原始图像进行优化。将原始图像下采样为不同的大小，图像就会包含不同频段的信息。通过将这些不同频段的信息放入StyleGAN来对输入进行逐级的优化，最终得到的结果能够优化原始的图像。

**图2 PSFR-GAN原理示意图[4]**



本节将会详细讲述PSFR-GAN的网络结构，以及本工作针对三维超声重建图像优化的场景对其的改进。

2.2.1 PSFR-GAN的网络结构

如图2所示，PSFR-GAN将一个 维的张量作为输入，在此用表示。可以以噪音的形式输入到网络中。此处的表示输入的通道的数量。接着PSFR-GAN使用了五个渐进式的不同尺度的上采样网络来处理输入，最终能够得到和需要输出的高清图像大小相同的特征张量。可以用来表示第层网络的输出，则可以将PSFR-GAN的网络用如下公式来表示：

此处代表残差卷积网络，代表上采样残差卷积网络，代表风格迁移网络。PSFR-GAN的网络结构的最后一层网络ToRGB是一个单层的卷积网络，能够将最后一层迁移网络输出的特征张量转化为网络最终的输出。

(5)中的使用和该层对应输出大小相同的参数张量来处理输入。若将第层低通图像LQ和它对应的人脸分割图记为，那么的网络结构即可表示为以下公式：

此处代表一个轻量级的卷积神经网络，和分别代表特征张量的均值和标准差。

PSFR-GAN的输入由原始LQ图像和人脸分割图组成。其中，人脸的分割图是通过一个提前训练好的分割网络以LQ图像为输入得到的结果。在PSFR-GAN处理的过程中，和会被按通道拼接在一起，作为一个张量输入到网络中。实际处理过程中，一共可以表示图中的19种脸部的不同部位。在网络处理过程中主要提供纹理信息，而则会提供颜色信息。

2.2.2 PSFR-GAN的损失函数

为了训练PSFR-GAN的神经网络，PSFR-GAN提出了三种不同的损失函数，从而在不同角度和不同层面对网络的输出做出评价，其中包含语义感知风格损失函数(Semantic-Aware Style Loss)，重建损失函数(Reconstruction Loss)和对抗损失(Adversarial Loss)。

其中，语义感知风格损失函数可以用以下公式来表示：

其中，表示VGG19[11]中层的特征，则代表标签为的分割区域，代表元素积，，用来避免分母为0的情况。

语义感知风格损失函数将VGG19网络包含进来，使用和在VGG19中的特征张量来计算损失，从而得到语义层面的损失。

重建损失可以用以下公式来表示：

上述公式中的第二项来自多尺度特征匹配损失(Multi-Scale Feature Matching Loss)[12]，其能够将和的分辨器的特征匹配起来。其中代表下采样系数，取值为。 则代表对应尺度的分辨器中层的特征。

对抗损失函数可以用以下公式来表示：

其中，代表铰链损失函数（Hinge Loss）。

对于PSFR-GAN来说，最终的损失函数可以用以下公式来表示：

此处的是不同损失函数的权重。PSFR-GAN训练的过程就是最小化和的过程。

2.2.3 PSFR-GAN优化流程

PSFR-GAN的优化流程主要有以下几个步骤：

1. 将需要优化的图片输入到PSFR-GAN中，需要优化的图片中需要包含人脸。PSFR-GAN开始进行优化。

2. PSFR-GAN使用FPN[19]来对输入的图像进行人脸检测，将检测到的人脸以正方形裁剪，并调整其大小为512512像素。

3. PSFR-GAN使用一个分割网络来对第2步中得到的人脸图像进行分割，得到对应的Parse Map。

4. PSFR-GAN将Parse Map和第2步中得到的人脸图像一起输入到PSFR-GAN的优化网络中，进行处理并得到优化后的人脸图片。

5. PSFR-GAN将优化后的人脸图片拼贴到原始输入图像中，完成最终的优化并输出。

在PSFR-GAN的优化过程中，第4步对人脸图像进行了优化，是流程中最关键的步骤。第2、3步中使用的人脸检测算法和人脸分割算法都并未使用PSFR-GAN的网络主体部分。

2.2.4 对于PSFR-GAN的改进

如前文所述，将当前的人脸优化算法应用于三位超声重建图像的场景存在许多问题：当前人脸优化算法主要针对于自然图像中成年人脸的优化，对胎儿人脸优化效果较差；自然图像的噪声来源与三维超声重建图像的噪声来源不同，当前人脸优化算法无法直接应用于三位超声重建图像的优化；三维超声重建图像目前没有公开数据集，基于深度学习的大数据驱动算法无法获取足够的数据集来支撑模型的训练。针对以上问题，本工作对PSFR-GAN的优化主要有以下几点。

1. PSFR-GAN原本的模型仅使用了公开数据集进行了训练，对于胎儿、婴儿以及三维超声重建图像优化效果较差。因此，本工作使用合作公司提供的数据集以及通过风格迁移方法扩充后的数据集进行了精调(Fine-Tuning)，使得算法的参数模型更加契合三维超声重建图像中胎儿人脸的优化问题。

2. 在PSFR-GAN处理图像的过程中，首先会采用经过预训练的人脸识别网络来找到人脸，将找到的人脸从原图中切割下来单独处理，再拼贴到原图中，以此完成对图片中所有人脸进行处理的任务。而在判定人脸的过程中，由于三维超声重建图片中的胎儿人脸信息往往不够明显，且存在前文所述的因角度、遮挡、平滑化等阻碍，使得PSFR-GAN在处理重建图片时无法精确对人脸进行优化。因此，在将PSFR-GAN应用于三维超声重建图片的优化任务时，需要将输入的图片转换为指定的大小，绕过人脸检测阶段，使用PSFR-GAN中的分割网络自动分割人脸，并直接进行优化。

3. PSFR-GAN中的输入数据对和分别提供纹理信息和颜色信息，对于三维超声重建图片来说，胎儿人脸的颜色较为单一，因此提升精度的主要手段是提升的精度。在使用时PSFR-GAN时，在分割网络无法准确分割人脸时，需要手动进行分割，并将分割好的结果封装成独热编码(One-Hot Encoding)并输入网络，从而进一步提升优化的精度。

2.3 传统方法预处理

由于三维超声重建图像数据量小、没有公开数据集用以训练，因此直接使用小批量数据集对原始模型进行精调预期较差，故采用传统图像优化算法预处理的方法进一步提升图像优化的精度。

局部拉普拉斯滤波器（Local Laplacian Filter）[13]通过下采样生成拉普拉斯金字塔，将图像分解成多个尺度。该算法可以用公式描述如下：

其中­为细节映射函数，为边缘映射函数，为高斯金字塔图像像素值，为原始图像中对应的像素值， 为细节和边缘阈值，为细节增强或平滑系数，为边缘增强系数。

通过传统方法的预处理，本文提出的模型能够在最终结果的数值指标上获取较为明显的优化效果。本文的3.3节将通过实验数据，详细讲述传统方法预处理对最终结果的优化效果。

2.4 风格迁移扩充数据集

三维超声重建图像的数据集较少，将婴儿图像通过风格迁移获取模拟三维超声重建图像数据可以扩充现有数据集。

[Taesung Park](https://arxiv.org/search/cs?searchtype=author&query=Park,+T)等人提出的CycleGAN[14]较好地解决了这个问题。使用公开的婴儿图像以及现有的三维超声重建图像对已有模型进行重新训练，可以得到较为接近三维超声重建图像的风格迁移图片，如图3所示。CycleGAN能够将输入图片的风格转换为某种特定风格的图片。



**原始数据**

**生成结果**

**图3 Cycle GAN 效果图**

将公开的婴儿图像使用CycleGAN处理后得到的风格迁移图片可以和原始三维超声重建图像制作扩充后的数据集可以用于深度神经网络的训练。

3. 图像优化实验结果以及分析

本节将首先指明实验中相关的数据集的情况，接着将讲述用来评价三维超声重建图像优化效果的指标，以及相关的三维超声重建图像优化任务的基线。本节的最后对实验结果以及数据进行汇总展示以及分析。

3.1 实验数据集

合作公司为本研究提供了共计352张三维超声重建图像的数据（来自于软件截图）。这些图片中有一部分仅包含一张超声重建图像（168张），另一部分包含一张超声重建图像以及生成这张重建图像的三张超声图像（184张）。本研究还对这些图片进行了分类处理，这些照片分为有脸、无脸以及侧脸，划分依据为图像中婴儿脸的角度以及是否包含人眼可识别的婴儿脸。对于训练集和测试集的划分，本研究按照8：2的比例进行了划分，其中训练集占80%，测试集占20%。图4为本研究的原始数据的一部分。表1展示了数据集中各类数据的数量。

**表1 数据集各类数据数量**

**图**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | 原图仅包含超声重建图像 | 原图包含超声重建图像及其原始超声图像 |
| 训练集 | 正脸 | 39 | 20 |
| 侧脸 | 67 | 34 |
| 无脸 | 28 | 93 |
| 测试集 | 正脸 | 10 | 5 |
| 侧脸 | 17 | 9 |
| 无脸 | 7 | 23 |

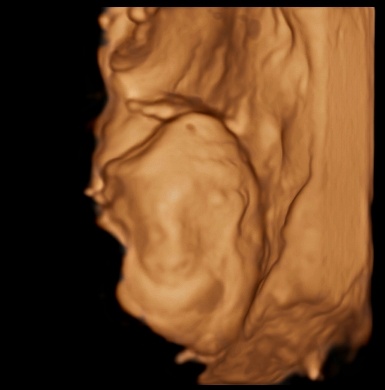


**图4 部分原始数据展示**

训练图像预处理：本研究进行了ROI提取以及归一化。如图5所示，处理后的图像保留了图片中婴儿人脸的部分，去除了软件的界面。归一化后，原图中仅包含重建图像的图像大小为770像素770像素，原图中包含重建图像和超声重建图像的图像大小为385像素385像素。

Ground Truth获取:为方便训练，本研究使用商业软件对数据集进行了处理，获得了每张图像对应的高分辨率图像作为Ground Truth。

**图5 处理后数据展示**



3.2 三维超声重建图像优化评价标准以及基线

3.2.1 三维超声重建图像优化评价标准

本研究采用PSNR（Peak Signal to Noise Ratio）[17]和SSIM（Structural Similarity）[18]的测量指标来度量算法的精准度。其中，PSNR可以表示信号最大可能功率和影响它的表示精度的破坏性噪声功率的比值，SSIM可以衡量图片的失真程度，可以衡量两张图片的相似程度，符合人眼的直观感受。

PSNR的计算公式如下：

其中，表示优化结果，表示Ground Truth，代表图像长度，代表图像宽度，代表图像通道数量，代表图像点颜色的最大数值。

SSIM的计算公式如下：

其中，代表优化结果和Ground Truth的平均值，代表优化结果和Ground Truth的标准差，代表优化结果和Ground Truth的协方差，、、分别为常数，以避免分母为0带来的错误。

PSNR的取值在0到40之间，SSIM的取值在0到1之间。这两种测量方法的值越大表明效果越好。

3.2.2 三维超声重建图像优化基线

PSFR-GAN与Bring Old Photos Back to Life都是优秀的人像优化算法，本文选取这两种算法作为基准，测试了它们对于三维超声重建图像的优化效果。

使用数据集中的测试数据对两种算法分别进行测试，并将产生的结果与Ground Truth进行比对，结果如表2所示。

**表2 现有优秀深度学习方法结果对比**

**图**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | PSNR 均值 | PSNR 标准差 | SSIM 均值 | SSIM 标准差 |
| Bring Old Photos Back to Life | **19.959** | **4.229** | **0.826** |  |
| PSFR-GAN | 13.047 | 6.374 | 0.548 |  |

由此可以得出结论，PSFR-GAN的总体效果比Bring Old Photos Back to Life较差，Bring Old Photos Back to Life的效果也更加稳定，但两者的优化效果总体依旧较差，两种指标的数值都很低。由于PSFR-GAN的视觉观察效果强于Bring Old Photos Back to Life，且更接近Ground Truth，因此本文选用PSFR-GAN进行改进。

3.3 实验结果及分析

如2.3节所述，本工作对PSFR-GAN的优化主要有精调已有模型、绕过人脸检测以及使用分割网络进行自动人脸分割并进行处理、绕过人脸检测以及手动分割人脸并进行处理。对应上述优化方案，本工作进行的实验步骤如下：

1. 使用现有三维超声重建图像数据集Fine-Tuning已有模型，对测试集中的数据进行处理，测试最终结果的指标。

2. 绕过人脸检测，使用分割网络进行自动人脸分割，将人脸分割得出的Parse Map和原始三维超声重建图像输入PSFR-GAN的优化网络中，使用Fine-Tuning后的模型进行处理，并测试最终结果的指标。

3. 绕过人脸检测，手动进行人脸分割，将人脸分割得出的Parse Map封装为One-Hot编码，和原始三维超声重建图像输入PSFR-GAN的优化网络中，使用Fine-Tuning后的模型进行处理，并测试最终结果的指标。

本工作通过实验获取了了经过分割网络自动分割人脸处理得出的数据以及手动分割人脸后处理得出的数据。部分实验结果如图6所示。



**图6 实验结果展示**

**(a)原图**

**(b)自动分割人脸**

**(c)手动分割人脸**

**(d)Ground Truth**

观察实验结果，自动分割人脸和手动分割人脸对于三维超声重建图像质量都有较大提升。相比于原图，优化过的结果在人脸整体的风格信息上和局部细节上都有明显提升，视觉效果较好。此外，自动分割人脸和手动分割人脸得出的结果存在差异。自动分割人脸的结果相对较为平滑，视觉效果更好，而手动分割人脸的结果会产生失真较多的细节，影响视觉效果。

分别使用PSNR和SSIM对实验结果进行评估，结果如表3所示。由实验数据可以得出结论，经过处理后的数据在精度上有了明显的提升，已经十分逼近Ground Truth，但在稳定性上可能还略逊色于自然人像处理算法。可能的原因是改进后的PSFR-GAN更加适用于三维超声重建图像优化场景，而自然人像处理算法由于很难适配于此类场景，导致处理后的结果和原图相比几乎没有变化，因此数据上的表现较为稳定。

**表3 实验各步骤结果**

**图**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | PSNR均值 | PSNR标准差 | SSIM均值 | SSIM标准差 |
| PSFR-GAN | 13.047 | 6.374 | 0.548 |  |
| 精调模型结果 | 13.047 | **6.219** | 0.548 |  |
| PSFR-GAN手动分割人脸的处理结果 | 25.381 | 5.839 | 0.757 |  |
| PSFR-GAN自动分割人脸的处理结果 | **27.305** | 6.227 | **0.840** |  |

本工作通过实验也获取了传统方法预处理前后的处理结果，部分处理结果如图7所示。经过预处理的图片会呈现出更多细节，包括部分高光区域以及阴影部位等。



**图7 进行预处理前后实验结果展示**

**（a）经过预处理**

**（b）未经过预处理**

表3显示了是否经过传统方法预处理的结果的指标对比。分析实验数据可知，经过预处理的模型生成的结果指标总体好于未使用预处理的模型，但稳定性有所下降。

**表3 是否经过传统方法预处理的结果的指标对比**

**图**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | PSNR均值 | PSNR标准差 | SSIM均值 | SSIM标准差 |
| 未经预处理 | 27.305 | **6.227** | 0.840 |  |
| 经过预处理 | **27.347** | 6.736 | **0.843** |  |

## 环境要求

* CUDA 10.1
* Python 3.7， 可以使用以下命令来安装依赖的库：

pip install -r requirements.txt

* 下载预训练模型：
  + [BaiduNetDisk](https://pan.baidu.com/s/1R2NCCvpTUPouiFfuIw88kA)，提取码：y7in。
  + 将所有预训练模型放置在路径 ./pretrain\_models 下。

## 测试运行

自动分割

所有输入图片应为 512 \* 512 像素的rgb图片。

使用以下命令来测试运行：

python test\_enhance\_single\_unalign.py --test\_img\_path ./test\_dir/147.png --results\_dir test\_result --gpus 1

* --test\_img\_path 为输入图片的位置。
* --result\_dir 为输出结果的路径。
  + test\_results/LQ\_faces 为需要处理的低通图像。
  + test\_results/ParseMaps 为低通图像的语义图。
  + test\_results/HQ 为处理结果。
* --gpus 为需要使用的GPU，<=0 表示使用CPU来运行。
* --manual\_parse 表示是否使用手动分割图，若是，则需要在 --manual\_parse\_map\_dir 选项处指明手动分割图的路径，否则将会使用自动分割网络进行分割。

手动分割

生成手动分割图片需要使用labelme和[ImageJ](https://imagej.nih.gov/ij/)。

可以使用以下命令来安装labelme（推荐使用Anaconda来管理labelme环境）：

conda create -n labelme python=3.8

conda activate labelme

pip install pyqt5

pip install labelme

可以在[此处](https://imagej.nih.gov/ij/)下载安装ImageJ。

使用以下命令启动labelme：

labelme

如图所示：

文本

描述已自动生成

此时会弹出labelme的界面：形状

描述已自动生成

点击File -> Open Dir可以打开存有超声重建图片的文件夹。

手机截图图人的照片上写着字

描述已自动生成

点击左侧Create Polygens来创建标注，之后可以在图片中画点来标注。

截图里有图片

描述已自动生成

重复以上步骤直到画面分割完毕（注意：当划分区域有重合时，应先画被遮挡的部分，例如，眼睛和脸部有重合，应该先画脸部，再画眼睛）。

社交网络的手机截图

描述已自动生成

然后可以点击File -> Save As来将标注文件（后缀为.json）保存到指定位置。

我们需要将标注文件转化为图片，使用以下命令来将标注文件生成图片：

python [Path to Anaconda labelme environment]\Lib\site-packages\labelme\cli\json\_to\_dataset\_original.py [Path to the json file]

需要将上述命令中的路径替换为自己的路径。运行结果如下图所示。文本

描述已自动生成

此时，json文件的路径下会生成对应文件夹，文件夹中可以找到生成的分割图label.png。

表格

描述已自动生成

图形用户界面, 应用程序, Word

描述已自动生成

之后我们需要使用ImageJ来打开分割图，找到每一个部位对应的颜色。ImageJ软件界面如图所示。

图形用户界面, 文本, 应用程序

描述已自动生成

点击File -> open 可以打开之前生成的分割图。

电脑萤幕的截图

描述已自动生成

选择color picker工具（左数第十二个），然后点击对应区域，可以看到对应的像素值。

将每个区域的像素值都记录下来，在项目目录中运行以下命令，即可生成PSFRGAN需要的分割图。

python mask\_converter.py –-input\_dir INPUT\_DIR –-output\_dir OUTPUT\_DIR -–face FACE\_COLOR –-left\_eye LEFT\_EYE\_COLOR –-right\_eye RIGHT\_EYE\_COLOR –-nose NOSE\_COLOR --upper\_lip UPPER\_LIP\_COLOR –-lower\_lip LOWER\_LIP\_COLOR

注意此处的各部分颜色需要对应ImageJ中提取到的原始分割图中各部分的颜色，其格式为(r,g,b)。（注意此处的数据需要以英文逗号分隔，并加上英文括号。）

分割时，需要将人脸分割为以下几个部位，并分别附以不同的颜色：

| **部位** | **R (Red)** | **G (Green)** | **B (Blue)** |
| --- | --- | --- | --- |
| face | 0 | 128 | 0 |
| left eye | 0 | 0 | 128 |
| right eye | 64 | 0 | 0 |
| nose | 128 | 128 | 128 |
| upper lip | 192 | 0 | 0 |
| lower lip | 128 | 0 | 128 |
| neck | 0 | 128 | 128 |
| body | 128 | 0 | 0 |
| background | 0 | 0 | 0 |

可以在运行test\_enhance\_sigle\_unalign.py文件时，在输入的参数中将—manual\_parse设置为True，并在—manual\_parse\_map\_dir选项处选择之前生成的手动分割图的路径。

测试设备及处理时间

* 使用 Nvidia GTX1060 6GB 显卡进行处理，单张 512 \* 512 的图片需要约3.02秒，峰值显存占用约1.8GB。
* 使用 Intel Core i7-8700 CPU 进行处理，单张 512 \* 512 的图片需要约6.37秒，峰值CPU占用率约71%。

## 训练模型

### 准备训练数据

* 将数据集中的数据放在../datasets/reconstructed\_ultrasound\_images/imgs1024 路径下，所有训练数据的长宽比应为**1：1**。
* 运行以下命令，生成对应的语义图：

python generate\_masks.py --test\_img\_path ../datasets/reconstructed\_ultrasound\_images/imgs1024 --results\_dir ../datasets/reconstructed\_ultrasound\_images/masks512

* 本项目使用三维超声重建图像数据下载：
  + [BaiduNetDisk](https://pan.baidu.com/s/1GYQV59jofjjeZBxU5RYTYg)，提取码：j0xs。

### 训练PSFRGAN

使用以下命令来训练PSFRGAN：

python train.py --gpus 1 --model enhance --name PSFRGAN\_v001 --g\_lr 0.0001 --d\_lr 0.0004 --beta1 0.5 --gan\_mode hinge --lambda\_pix 10 --lambda\_fm 10 --lambda\_ss 1000 --Dinput\_nc 22 --D\_num 3 --n\_layers\_D 4 --batch\_size 2 --dataroot ../datasets/reconstructed\_ultrasound\_images --visual\_freq 100 --print\_freq 10 --dataset ffhq

* 对于不同的实验，需要更改--name 选项。相同名字的实验结果会被覆盖。
* --gpus 表示训练需要使用的GPU数量。有更多的显存的GPU将会被优先使用。可以在训练之前，运行以下命令来指定GPU的id： export CUDA\_VISIBLE\_DEVICES=your\_gpu\_ids。
* --continue\_train 表示是否继续训练。在上面的命令中加上这个标识，将会自动在指定的路径中继续训练之前的模型。
* 当**batch\_size=1**时，需要至少**8GB**的显存才能够完成训练。

## 评测结果

使用以下命令可以用PSNR和SSIM指标评测生成的结果：

python measure.py --test\_dir test\_result --ground\_truth\_dir ground\_truth

* 此处 test\_result 和 ground\_truth\_dir 都必须为您自己的路径。
* test\_result 和 ground\_truth\_dir 中的文件名必须一一对应。
* 运行上面的指令可以输出测试路径中所有图像和Ground Truth路径中所有图像的 PSNR 和 SSIM 的平均值与方差。

参考文献

1. V. Le, J. Brandt, L. Zhe, L. D. Bourdev, and T. S. Huang. Interactive facial feature localization. In ECCV, pages 679–692, 2012.
2. Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In ICCV, pages 3730–3738, 2016.
3. Junjun Jiang, Chenyang Wang, Xianming Liu, & Jiayi Ma (2021). Deep Learning-based Face Super-resolution: A Survey. CoRR, abs/2101.03749.
4. Chen, C., Li, X., Yang, L., Lin, X., Zhang, L., & Wong, K.Y.. (2020). Progressive Semantic-Aware Style Transformation for Blind Face Restoration.
5. E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin. Learning face hallucination in the wild. In AAAI, pages 3871–3877, 2015
6. J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In CVPR, pages 1646–1654, 2016.
7. K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong. Dual-path deep fusion network for face image hallucination. IEEE Transactions on Neural Networks and Learning Systems, pages 1–14, 2020.
8. C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):295–307, 2016.
9. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio. Generative adversarial nets. In NIPS, volume 2, pages 2672–2680, 2014.
10. Y. Shi, G. Li, Q. Cao, K. Wang, and L. Lin. Face hallucination by attentive sequence optimization with reinforcement learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(11):2809–2824, 2020.
11. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.
12. Wang, T.-C.; Liu, M.-Y.; Zhu, J.-Y.; Tao, A.; Kautz, J.; and Catan-zaro, B. 2018a. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
13. Paris S, Hasinoff S W, Kautz J. Local Laplacian filters: edge-aware image processing with a Laplacian pyramid[J]. Communications of the ACM, 2015, 58(3): 81-91.
14. J. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.
15. Z. Wan et al., "Bringing Old Photos Back to Life," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2744-2754, doi: 10.1109/CVPR42600.2020.00282.
16. 徐冰冰,岑科廷,黄俊杰,沈华伟,程学旗.图卷积神经网络综述[J/OL].计算,2019:1-31[2022-05-11]. <http://kns.cnki.net/kcms/detail/11.1826.tp.20191104.1632.006.html>

1. [Brazhenenko, Maksym](https://connections-qj.org/biblio?f%5Bauthor%5D=18232), [Victor Shevchenko](https://connections-qj.org/biblio?f%5Bauthor%5D=18233), [Oleksiy Bychkov](https://connections-qj.org/biblio?f%5Bauthor%5D=18223), [Boyan Jekov](https://connections-qj.org/biblio?f%5Bauthor%5D=18235), [Pepa Petrova](https://connections-qj.org/biblio?f%5Bauthor%5D=18234), and [Eugenia Kovatcheva](https://connections-qj.org/biblio?f%5Bauthor%5D=18236). "[Adopting Machine Learning for Images Transferred with LoRaWAN](https://connections-qj.org/article/adopting-machine-learning-images-transferred-lorawan)." Information & Security: An International Journal 47, no. 2 (2020): 172-186.
2. Chaofeng Li, Alan C. Bovik, "Three-component weighted structural similarity index," Proc. SPIE 7242, Image Quality and System Performance VI, 72420Q (19 January 2009);<https://doi.org/10.1117/12.811821>
3. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).