Journal of Computer Applications
计算机应用, 2020, 40(3):710 - 716
ISSN 1001-9081
CODEN JYIIDU
2020-03-10
http://www.joca.cn

文章编号:1001-9081(2020)03-0710-07
DOI:10.11772/j.issn.1001-9081.2019071178

基于组合学习的人脸超分辨率算法

许若波[1,2],卢　涛[1,2*],王　宇[1,2],张彦铎[1,2]
(1.武汉工程大学 计算机科学与工程学院,武汉 430205;　2.智能机器人湖北省重点实验室(武汉工程大学),武汉 430205)
(*通信作者电子邮箱lut@wit.edu.cn)

摘　要:现有的基于深度学习的人脸超分辨算法大部分仅仅利用一种网络分区重建高分辨率输出图像,并未考虑人脸图像中的结构性信息,导致了在人脸的重要器官重建上缺乏足够的细节信息。针对这一问题,提出一种基于组合学习的人脸超分辨率算法。该算法独立采用不同深度学习模型的优势重建感兴趣的区域,由此在训练网络的过程中每个人脸区域的数据分布不同,不同的子网络能够获得更精确的先验信息。首先,对人脸图像采用超像素分割算法生成人脸组件部分和人脸背景图像;然后,采用人脸组件生成对抗网络(C-GAN)独立重建人脸组件图像块,并采用人脸背景重建网络生成人脸背景图像;其次,使用人脸组件融合网络将两种不同模型重建的人脸组件图像块自适应融合;最后,将生成的人脸组件图像块合并至人脸背景图像中,重建出最终的人脸图像。在FEI数据集上的实验结果表明,与人脸图像超分辨率算法通过组件生成和增强学习幻构人脸图像(LCGE)及判决性增强的生成对抗网络(EDGAN)相比,所提算法的峰值信噪比(PSNR)值分别高出1.23 dB和1.11 dB。所提算法能够采用不同深度学习模型的优势组合学习重建更精准的人脸图像,同时拓展了图像重建先验的来源。

关键词:组合学习;人脸幻构;生成对抗网络;融合网络;深度学习
中图分类号:TP391.4　　**文献标志码:**A

Face hallucination algorithm via combined learning

XU Ruobo[1,2], LU Tao[1,2*], WANG Yu[1,2], ZHANG Yanduo[1,2]
(1. *School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan Hubei* 430205, *China*;
2. *Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology), Wuhan Hubei* 430205, *China*)

Abstract: Most of the existing deep learning based face hallucination algorithms only use a single network partition to reconstruct high-resolution output images without considering the structural information in the face images, resulting in the lack of sufficient details in the reconstruction of vital organs on the face. Therefore, a face hallucination algorithm based on combined learning was proposed to tackle this problem. In the algorithm, the regions of interest were reconstructed independently by utilizing the advantages of different deep learning models, thus the data distribution of each face region was different to each other in the process of network training, and different sub-networks were able to obtain more accurate prior information. Firstly, for the face image, the superpixel segmentation algorithm was used to generate the facial component parts and facial background image. Secondly, the facial component image patches were independently reconstructed by the Component-Generative Adversarial Network (C-GAN) and the facial background reconstruction network was used to generate the facial background image. Thirdly, the facial component fusion network was used to adaptively fuse the facial component image patches reconstructed by two different models. Finally, the generated facial component image patches were merged into the facial background image to reconstruct the final face image. The experimental results on FEI dataset show that the Peak Signal to Noise Ratio (PSNR) of the proposed algorithm is respectively 1.23 dB and 1.11 dB higher than that of the face image hallucination algorithms Learning to hallucinate face images via Component Generation and Enhancement (LCGE) and Enhanced Discriminative Generative Adversarial Network (EDGAN). The proposed algorithm can perform combined learning of the advantages of different deep learning models to learn and reconstruct more accurate face images as well as expand the sources of image reconstruction prior information.

Key words: combined learning; face hallucination; Generative Adversarial Network (GAN); fusion network; deep learning

收稿日期:2019-07-08;**修回日期:**2019-09-01;**录用日期:**2019-09-03。
基金项目:国家自然科学基金资助项目(61502354);湖北省自然科学基金资助项目(2015CFB451);中央引导地方科技发展专项(2018ZYYD059);武汉工程大学科研基金资助项目(K201713);武汉工程大学研究生教育创新基金资助项目(CX2018211)。
作者简介:许若波(1990—),男,江苏徐州人,硕士研究生,主要研究方向:图像处理、计算机视觉;　卢涛(1980—),男,湖北武汉人,副教授,博士,主要研究方向:图像/视频处理、计算机视觉、人工智能;　王宇(1996—),男,湖北荆州人,硕士研究生,主要研究方向:图像处理、计算机视觉;　张彦铎(1971—),男,黑龙江肇东人,教授,博士,主要研究方向:智能计算、智能机器人、智能检测、仿真系统。

0　引言

人脸超分辨率是一种从输入的低分辨率人脸图像推断出其潜在的对应高分辨率图像技术,该技术能够显著增强低分辨率图像的细节信息,因而被广泛地应用到了人脸识别、刑事侦察、娱乐等领域。

基于学习的超分辨率方法能够利用样本提供的先验信息推导低分辨率图像中缺失的高频细节信息,例如基于主成分分析(Principal Component Analysis, PCA)[1]的线性约束、局部线性嵌入(Locally Linear Embedding, LLE)[2]、稀疏表达[3]、局部约束表达[4]等,这些方法获得了较好的图像主客观重建质量。

深度学习提供了一种端到端的映射关系学习模式来处理超分辨率问题。Dong等[5]首次提出了使用卷积神经网络建立端到端的超分辨率算法(image Super-Resolution using deep Convolutional Neural Networks, SRCNN)。Kim等[6]利用深层网络学习图像的残差信息(accurate image Super-Resolution using Very Deep convolutional networks, VDSR)。卢涛等[7]使用中继循环网络,增强超分辨率重建图像的重建效果。Ledig等[8]首次将生成对抗网络(Generative Adversarial Network, GAN)运用到图像超分辨率,使图像产生更加逼真的视觉效果。Yang等[9]提出了判决性增强的生成对抗网络(Enhanced Discriminative Generative Adversarial Network for face super-resolution, EDGAN)用于人脸图像的超分辨率。考虑到人脸的结构特性,Song等[10]提出的通过组件生成和增强学习幻构人脸图像算法(Learning to hallucinate face images via Component Generation and Enhancement, LCGE),证明了人脸组件在重建高分辨率图像中的作用。在LCGE的基础上,Jiang等[11]通过两步法结合使用卷积神经网络去噪和多层邻域嵌入来幻构人脸图像。

基于卷积神经网络的模型可以重建出平滑的图像,同时网络易于收敛和训练,可以生成高的客观评分但相对平滑的结果。基于生成对抗网络的方法可以产生逼真的视觉效果,并恢复丰富的纹理细节,但获得的客观评分较低。由于人脸图像具有高度的结构性,在眼睛、鼻子和嘴巴周围包含丰富的细节。人脸超分辨率方法和通用图像超分辨率算法最大的不同之处在于如何利用人脸图像的结构信息。受到组合学习[12-13]方法的启示,由于人脸组件部分纹理的复杂性,仅仅利用一种模型难以充分利用人脸的结构性先验信息,针对这一问题,本文提出了一种基于组合学习的人脸超分辨率算法,采用多个并行训练的组件生成对抗网络(Component-Generative Adversarial Network, C-GAN)恢复人脸的重要器官部分,由此在训练网络的过程中每个人脸区域的数据分布不同,不同的子网络能够获得更精确的先验信息,并使用卷积神经网络(Convolutional Neural Network, CNN)重建人脸背景图像。首先对人脸图像使用超像素分割算法[14-15]进行分割,分割出重要的人脸组件部分和人脸背景部分;其次对人脸组件部分构建子网络,利用不同的超分辨率方法重建背景图像;然后在重建高分辨率组件的基础上建立特征融合网络,将不同学习方法获得的重建结果进行融合;最后输出高分辨率人脸图像。论文的创新之处在于:1)首次利用多种不同的超分辨率重建方法组合学习获得更精准的重建图像。2)提出了组件自适应权重学习的多源图像融合重建网络,拓展了图像重建先验的来源。

1　相关知识

近年来,基于深度学习的图像超分辨率受到很大的关注,Dong等[5]提出使用3层卷积神经网络重建低分辨率图像。Kim等[6]在SRCNN的基础上加深网络层数,学习图像的残差信息,并获得了卓越的重建性能。但是基于卷积神经网络重建的图像在主观效果上过于平滑,一些复杂的纹理细节难以恢复,尤其像人脸图像具有高度的结构性,更加难以学习其复杂的细节信息。为了重建的图像获得更加逼真的效果,Ledig等[8]提出使用生成对抗网络进行对抗学习来恢复更多的细节,并产生逼真的视觉效果。Yu等[16]采用生成对抗网络的思想提出了一种变换自动编码器网络,来重建未对齐且含有噪声的极低分辨率人脸图像。虽然生成对抗网络重建的图像可以恢复丰富的纹理细节,但是客观评估指标值较低。为了获得平滑度适中且具有更高的客观评分,利用两种不同深度学习模型的优势组合学习人脸结构的复杂先验信息,从而提高网络的重建性能。

2　本文方法

2.1　组合学习框架

为了充分发挥出不同的神经网络结构特点和重建性能,本文提出了一种多结构网络的组合学习框架,其网络结构如图1所示,其中C-GAN表示组件生成对抗网络,n表示人脸组件的个数,Fusion-Net表示组件融合网络。

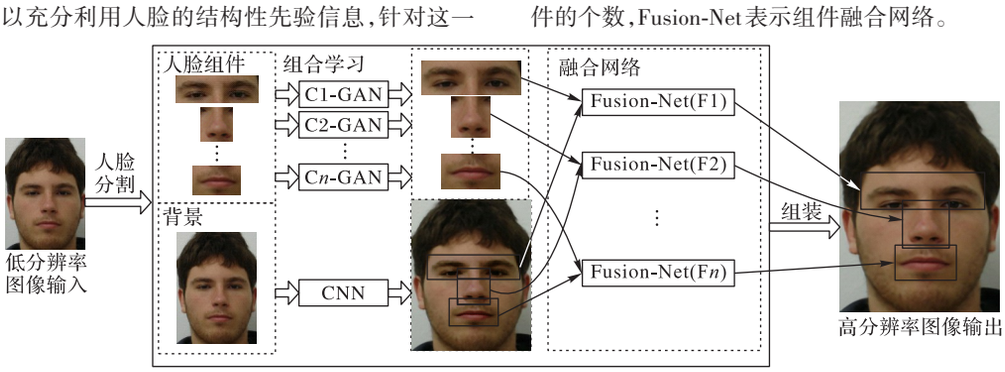


图1　基于组合学习的人脸超分辨率算法框架

Fig. 1　Outline of face hallucination via combined learning

C-GAN用于恢复人脸图像纹理细节丰富的区域,而CNN则用于学习整个人脸图像作为背景信息,而Fusion-Net则用于融合多网络重建的组件信息,提升超分辨率重建的质量。值得注意的是本文提供了一种多神经网络融合重建的超分辨

率算法框架。为了便于说明和理解，采用了具体的网络结构加以说明，为了简化问题的复杂度，本文仅仅设置 $n=3$，n 表示人脸组件的个数，建立 C1（眼睛）、C2（鼻子）和 C3（嘴巴）3 种独立模型，用以证明本文所提出思路的有效性。组合学习算法流程如图 2 所示。

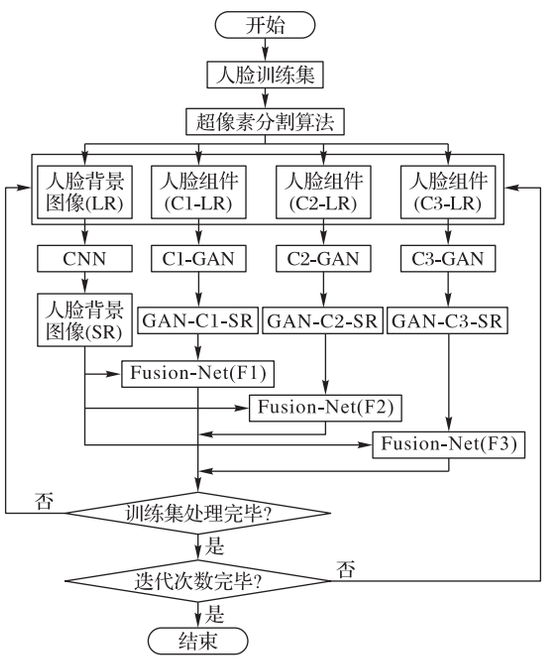


图 2 组合学习算法流程

Fig. 2 Flow chart of combined learning algorithm

在本文中，训练集为 $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{N}$，$i$ 表示训练集中的样本图像索引，N 表示训练样本总数量，低分辨率人脸图像和高分辨率人脸图像分别表示为 $\boldsymbol{x}_i \in \Re^{m\times n}$ 和 $\boldsymbol{y}_i \in \Re^{mt\times nt}$，$t$ 表示放大因子，$t=4$。采用低分辨率数据集 $\boldsymbol{x}_i \in \Re^{m\times n}$ 生成平均面部图像 $\bar{\boldsymbol{x}}$，将平均面部图像 $\bar{\boldsymbol{x}}$ 通过超像素分割算法生成人脸组件标签：

$$\boldsymbol{Lab} = SLIC(\bar{\boldsymbol{x}};\ k, m) \tag{1}$$

其中：$SLIC$ 表示超像素分割算法，$\boldsymbol{Lab}$ 表示生成的人脸组件标签，k 和 m 表示超像素算法的基本参数，k 表示预设的超像素个数，m 用于调整超像素的紧凑程度和边缘特性，在本文中 $k=30$，$m=80$。

超像素分割方法对低分辨率人脸图像进行分割，产生多个不规则的超像素块，每个超像素块拥有自己的标签，提取不规则人脸组件区域所有坐标：

$$\boldsymbol{E}^{(\boldsymbol{Lab}_l)}(x_i) = \{(\boldsymbol{row}_l, \boldsymbol{col}_l) | 1 \leqslant l \leqslant 30\} \tag{2}$$

$\boldsymbol{E}^{(\boldsymbol{Lab}_l)}$ 表示提取人脸图像 $\boldsymbol{x}_i$ 中标签为 l 的区域的所有横坐标点和纵坐标点，$\boldsymbol{row}_l$ 和 $\boldsymbol{col}_l$ 分别表示横坐标点和纵坐标点。最小横坐标点和最大横坐标点分别为 $\min(\boldsymbol{row}_l)$ 和 $\max(\boldsymbol{row}_l)$，最小纵坐标点和最大纵坐标点分别为 $\min(\boldsymbol{col}_l)$ 和 $\max(\boldsymbol{col}_l)$，通过对角的两点坐标点确定最终的低分辨率人脸组件矩形面积为：

$$\boldsymbol{s}_i^{C_j} = rect((\min(\boldsymbol{row}_l),\ \min(\boldsymbol{col}_l)),\ (\max(\boldsymbol{row}_l),\ \max(\boldsymbol{col}_l))) \tag{3}$$

其中：$rect$ 表示通过两个对角坐标点生成一个规则的矩形，即人脸组件图像块。$\boldsymbol{s}_i^{C_j}$ 表示最终分割的低分辨率人脸组件图像块，j 表示不同的人脸组建图像块，其中 $\{j | 1 \leqslant j \leqslant 3\}$，因此可形成人脸组件训练集 $\{\boldsymbol{s}_i^{C_j}, \boldsymbol{y}_i^{C_j}\}_{i=1,j=1}^{N,n}$，$i$ 表示人脸组件训练数据集中的样本图像块索引，$\boldsymbol{s}_i^{C_j}$ 为低分辨率人脸组件图像块，$\boldsymbol{y}_i^{C_j}$ 为与之对应的高分辨率人脸组件图像块。

2.2 重建网络

2.2.1 人脸组件生成对抗网络（C-GAN）

为了获取人脸组件周围更多的纹理细节，本文采用生成对抗网络来学习更精准的先验信息。由于人脸组件图像块较小，为了保持网络的重建性能，本文设计了一个适用于输入图像块较小的人脸组件生成对抗网络，并且此网络可以独立地训练人脸组件区域，形成多个并行的子网络，每个独立的子网络的数据分布不同，从而使独立的子网络学习更精确的先验信息。本文采用了一个浅层的生成对抗网络，加深了网络的宽度，训练过程中可以更好地提取人脸组件的重要特征，并且生成对抗网络通过生成网络和对抗网络互相对抗学习获得更好的视觉效果。人脸组件生成对抗网络结构如图 3 所示。

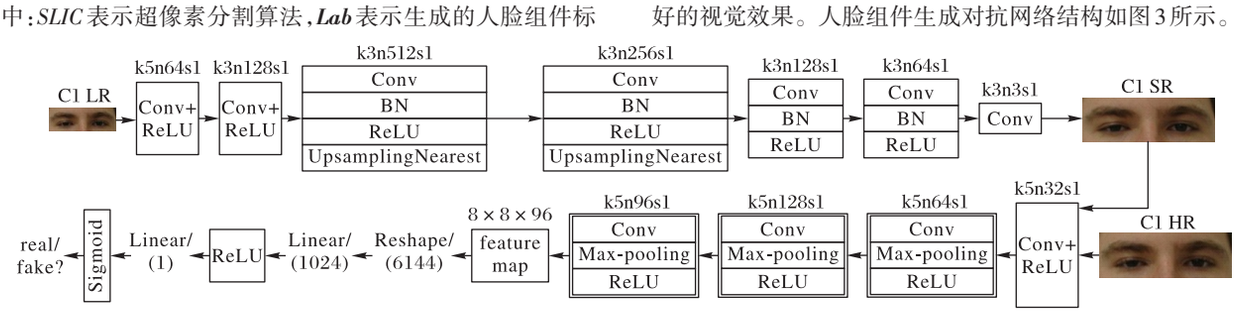


图 3 人脸组件生成对抗网络（C-GAN）的网络结构

Fig. 3 Network structure of face Component Generative Adversarial Networks（C-GAN）

在生成网络部分，首先在网络的开始层采用了较大感受野的卷积层（卷积核大小为 5×5）来提取更加丰富全面的特征，其他层采用的是 3×3 的卷积核，采用两个反卷积层上采样至高分辨率图像块大小，反卷积层是由上采样层和卷积层级联起来。人脸组件生成对抗网络训练集为 $\{\boldsymbol{s}_i^{C_j}, \boldsymbol{y}_i^{C_j}\}_{i=1,j=1}^{N,n}$，低分辨率人脸组件图像块和高分辨率人脸组件图像块分别表示为 $\boldsymbol{s}_i^{C_j} \in \Re^{w\times h}$ 和 $\boldsymbol{y}_i^{C_j} \in \Re^{wt\times ht}$，$t$ 表示放大因子，$t=4$，设有 3 个独立的人脸组件生成对抗网络。人脸组件生成对抗网络重建图像为：

$$\boldsymbol{M}_i^j = G_j(\boldsymbol{s}_i^{C_j}) \tag{4}$$

其中：G_j 表示不同的人脸组件生成网络，$\boldsymbol{M}_i^j$ 表示不同人脸组件生成网络生成的高分辨率人脸组件图像块，i 表示重建的人脸组件图像块索引。

在判别网络部分，使用了 4 个卷积层，在其中 3 个卷积层后级联最大池化层，随着网络层数加深，特征尺寸不断减小，并采用修正线性单元（Rectified Linear Unit，ReLU）作为激活函数，最终通过两个全连接层和最终的 sigmoid 激活函数得到

预测结果。判别网络能够区分输入的图像块是原始的高分辨率人脸组件图像块还是人脸组件重建图像块,相应的判别信息被反向传播到生成网络,从而可以生成视觉效果更加逼真的人脸组件重建图像块。

人脸组件生成网络损失函数使用均方误差(Mean Square Error, MSE)作为损失函数,采用L2范数损失学习低分辨率图像和高分辨率图像之间的差异,生成网络损失函数为:

$$\min_{u^{C_j}} U^{C_j}(u^{C_j}) = E_{(\boldsymbol{s}_i^{C_j},\boldsymbol{y}_i^{C_j})\sim p(\boldsymbol{s}^{C_j},\boldsymbol{y}^{C_j})}\left\|\boldsymbol{y}_i^{C_j} - G_j(\boldsymbol{s}_i^{C_j})\right\|_2^2 \tag{5}$$

其中:u表示人脸组件生成网络的参数,u^{C_j}表示不同的人脸组件生成网络的参数,$p(\boldsymbol{s}^{C_j},\boldsymbol{y}^{C_j})$表示人脸组件训练集中低分辨率人脸组件图像和原始高分辨率人脸组件图像的联合分布。

判别网络损失函数为:

$$\begin{aligned}\max_{v^{C_j}} L^{C_j}(v^{C_j}) = E[\lg(D_j(\boldsymbol{y}_i^{C_j})) + \lg(1 - D_j(G_j(\boldsymbol{s}_i^{C_j})))] = \\ E_{\boldsymbol{y}_i^{C_j}\sim p(\boldsymbol{y}^{C_j})}[\lg(D_j(\boldsymbol{y}_i^{C_j}))] + \\ E_{G_j(\boldsymbol{s}_i^{C_j})\sim p(G_j(\boldsymbol{s}^{C_j}))}[\lg(1 - D_j(G_j(\boldsymbol{s}_i^{C_j})))]\end{aligned} \tag{6}$$

其中:v表示判别网络的参数,v^{C_j}表示不同的人脸组件判别网络的参数,$p(\boldsymbol{y}^{C_j})$和$p(G_j(\boldsymbol{s}^{C_j}))$表示原始的高分辨率人脸组件和人脸组件重建图像的联合分布,$D_j(\boldsymbol{y}_i^{C_j})$和$D_j(G_j(\boldsymbol{s}_i^{C_j}))$作为判别网络的输出,$D_j(G_j(\boldsymbol{s}_i^{C_j}))$表示人脸组件重建图像块$G_j(\boldsymbol{s}^{C_j})$是原始人脸组件图像块的概率,损失函数$L^{C_j}$反向传播到生成网络来更新参数$u^{C_j}$,通过传播给生成网络判别信息,使生成网络可以生成更加逼真的高分辨率人脸组件图像块。

2.2.2　人脸背景图像重建网络

本文采用了一个20层的深层残差网络作为人脸背景重建网络。第一层和最后一层的结构不同,第一层作为网络的输入,输出的通道数为64,最后一层的作用为输出重建后的图像,由1个3×3的卷积核组成。其余的18层网络结构由64个3×3大小的卷积核组成,并且采用ReLU作为激活函数,使网络更加容易学习,并提升网络的训练速度。

在人脸背景图像重建网络中,$\boldsymbol{x}_i$为低分辨率图像,将$\boldsymbol{x}_i$双三次插值至高分辨率图像大小表示为$\hat{\boldsymbol{x}}_i$,$\boldsymbol{y}_i$为对应的高分辨率图像,残差图像定义为$\boldsymbol{r}_i = \boldsymbol{y}_i - \hat{\boldsymbol{x}}_i$,残差图像的像素值非常小,因此在训练过程中大量减少了网络携带的信息,使网络更加容易收敛。人脸背景重建图像为:

$$\tilde{\boldsymbol{Y}}_i = f(\hat{\boldsymbol{x}}_i) + \hat{\boldsymbol{x}}_i \tag{7}$$

其中:$f(\hat{\boldsymbol{x}}_i)$表示残差网络学习到的残差图像。

使用均方误差(MSE)作为损失函数,则损失函数为:

$$Loss^{\mathrm{CNN}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{r}_i - f(\hat{\boldsymbol{x}}_i)\right\|_2^2 \tag{8}$$

其中:$\hat{\boldsymbol{x}}_i$表示输入的低分辨率图像,$f(\hat{\boldsymbol{x}}_i)$表示网络学习到的残差图像。

2.2.3　组件融合网络

本文利用C-GAN模型重建的人脸组件和CNN模型重建的人脸背景图像对应融合,提取其深层特征,在特征域自适应融合两种不同的特征,从而进一步学习低分辨率人脸组件图像至原始高分辨率人脸组件图像之间的映射关系,恢复更多的细节信息,进一步增强重建图像的质量,并提高组合模型的重建性能。为了保持空间信息,不使用任何池化层或跨步卷积层。特征提取网络并行使用了8个残差块密集连接,使用本地连接和长跳跃连接获取更丰富的特征。组件融合网络结构如图4所示。

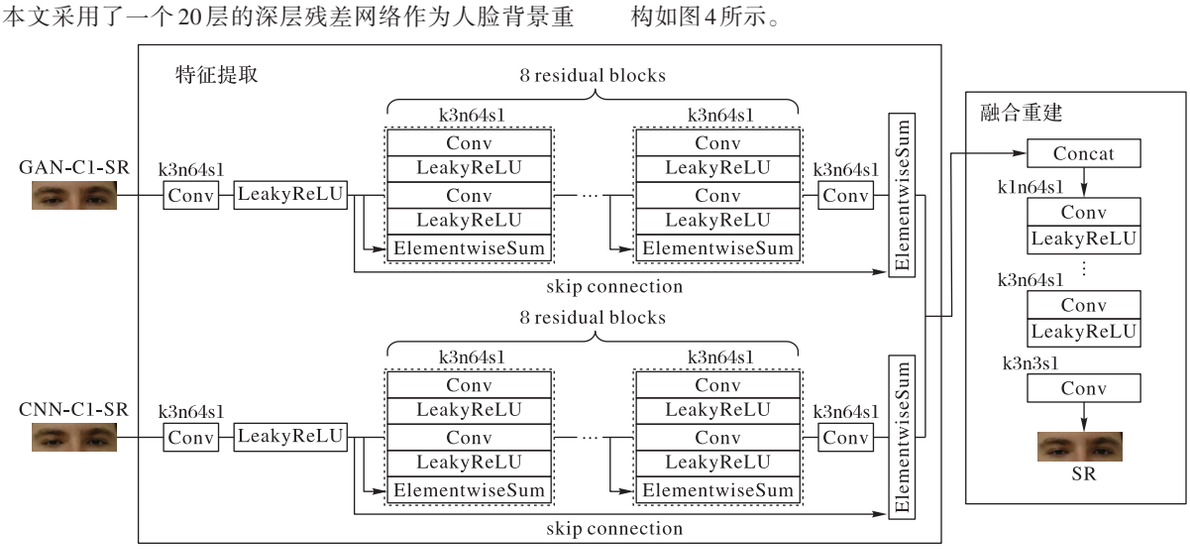


图4　组件融合网络结构

Fig. 4　Network structure of component fusion

融合重建模块由5个卷积层和4个LeakyReLU激活函数组成,第一个卷积层卷积核大小为1×1,其他卷积层卷积核大小为3×3,两种不同的特征拼接输入卷积核大小为1×1的卷积层,从而可以降低维度,并保留重要的特征信息。最后一层卷积层为最终重建层,卷积核的个数为3。

通过取人脸背景重建图像$\tilde{\boldsymbol{Y}}_i$与人脸组件重建图像块$\boldsymbol{M}_i^j$的交集获取人脸背景重建图像的人脸组件区域$\boldsymbol{U}_i^j = \tilde{\boldsymbol{Y}}_i \cap \boldsymbol{M}_i^j$,形成组件融合网络训练集$\{\boldsymbol{M}_i^j, \boldsymbol{y}_i^{C_j}\}_{i=1,j=1}^{N,n}$,$\{\boldsymbol{U}_i^j, \boldsymbol{y}_i^{C_j}\}_{i=1,j=1}^{N,n}$。组件融合网络重建的高分辨率人脸组件图像块为:

$$\boldsymbol{V}_i^j = \boldsymbol{M}_i^j \otimes \boldsymbol{U}_i^j \tag{9}$$

其中:$\otimes$表示融合,j表示不同的人脸组件图像块,$\boldsymbol{V}_i^j$表示人脸组件重建图像块。

将融合的人脸组件重建图像块合并至人脸背景重建图像中,形成最终的人脸高分辨率重建图像为:

$$\boldsymbol{Y}_i = (\boldsymbol{V}_i^1 \cup \cdots \cup \boldsymbol{V}_i^n) \cup \tilde{\boldsymbol{Y}}_i \tag{10}$$

其中:$\boldsymbol{Y}_i$表示最终合成的人脸图像,$\cup$表示人脸组件和人脸背景图像合并。

融合网络采用均方误差(MSE)作为损失函数,计算融合

重建后的人脸组件图像和原始的高分辨人脸组件图像之间的损失。损失函数为：

$$L^{F_j}(\theta^{C_j}) = \left\| \boldsymbol{y}_i^{C_j} - \boldsymbol{V}_i^j \right\|_2^2 \tag{11}$$

其中：θ^{C_j}表示不同组件融合网络的模型参数，F_j表示不同的融合网络，L^{F_j}表示不同融合网络的损失函数。

3 实验结果

本文使用了FEI(FEI Face Database)数据集对所提出的方法进行实验验证，FEI数据集含有400张正面拍摄图像，共有200人，每个人有两张正面的图像。本文使用360张作为训练集，40张作为测试集，高分辨率图像大小为260×360像素，使用双三次插值进行下采样得到低分辨率图像，下采样因子为4，得到低分辨率图像大小为65×90像素。

3.1 参数设置

实验使用4个GPU(NVIDIA 1080Ti)并行训练4个独立的网络，其中3个GPU并行训练眼睛、鼻子和嘴巴的人脸组件生成对抗网络，在人脸组件图像块中取大小为16×16像素的图像块进行训练，共训练70个时期，初始学习率设置为0.001，每个时期学习率自动更新，将初始学习率乘以一个辅助变量0.99^{γ}作为下一次网络训练的学习率，γ表示当前的时期次数，网络的动量初始值设为0，每个时期增加0.000 8，训练集批量大小设置为64。第4个GPU用来训练人脸背景图像，将低分辨率人脸背景图像双三次插值至原始高分辨率图像大小，取大小为41×41像素的图像块进行训练，训练80个时期，初始学习率为0.1，每20个时期学习率将下降至原来的1/10，网络的动量为0.9，ℓ_2的权值衰减为0.000 1。

融合网络在高分辨率空间进行融合训练，从生成的高分辨率人脸组件数据集中取大小为64×64像素的图像块作为网络的输入，融合网络共训练150个时期，初始学习率设置为0.000 1，并且每30个时期乘以0.1，批量大小设置为16，并且通过随机旋转和水平翻转来增加训练数据。

3.2 人脸组件提取

本文采用超像素分割方法分割人脸组件，利用像素之间特征的相似性将像素分组，人脸组件分割提取流程如图5所示，首先通过低分辨率训练集生成平均面部图像，采用超像素分割算法对平均面部图像进行聚类并生成多个超像素块，每个超像素块拥有自己的标签，从图5中可以看出生成的人脸组件标签模板，采用此模板对人脸图像进行分解，通过标签提取不规则的区域，对不规则区域进行矩形化得到最终的低分辨率人脸组件图像块，通过低分辨率人脸组件图像块的两个对角坐标点，获取对应的高分辨率人脸组件图像块。在本文中提取C1低分辨率组件大小为48×16像素，C2低分辨率组件大小为19×20像素，C3低分辨率组件大小为28×17像素。

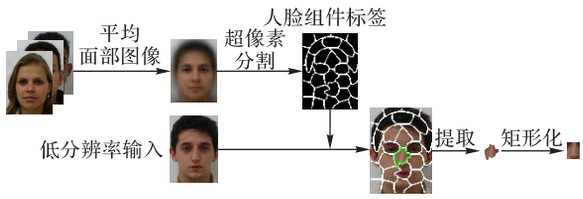


图5 人脸组件分割提取流程

Fig. 5 Flow chart of face component segmentation and extraction

3.3 定量和定性分析

本文提出的算法与几种算法进行比较，包括双三次插值、SRCNN[5]、VDSR[6]、LCGE[10]和EDGAN[9]方法，使用FEI数据集进行测试，为了公平比较，本文使用相同的训练集重新训练对比算法，由于LCGE只公布了测试代码，所以只对LCGE进行测试。本文采用了峰值信噪比(Peak Signal to Noise Ratio，PSNR)、结构相似性(Structural SIMilarity，SSIM)[17]和视觉信息保真度(Visual Information Fidelity，VIF)[18]作为图像质量客观评估指标。

3.3.1 融合网络的作用

为了验证不同深度学习模型重建的图像块自适应融合的有效性，本节比较了不同的模型所产生的效果，如表1所示，相对于CNN模型和C1-GAN模型，本文方法的组合学习模型性能均有不同程度的提升，重建图像的质量明显上升，本文只展示40张C1区域的平均定量值对比实验结果，计算平均客观评估值，本文所提出的算法超过基于CNN和C1-GAN模型分别为1.60 dB/0.016 0/0.040 3，1.39 dB/0.013 9/0.059 9，这表明了组件融合网络的有效性。

表1 同一个区域不同模型的平均PSNR、SSIM和VIF

Tab. 1 Average PSNR, SSIM and VIF for different models in same region

模型	客观评估指标		
	PSNR/dB	SSIM	VIF
CNN模型	39.70	0.952 8	0.732 1
C1-GAN模型	39.91	0.955 1	0.712 5
本文算法	41.30	0.969 0	0.772 4

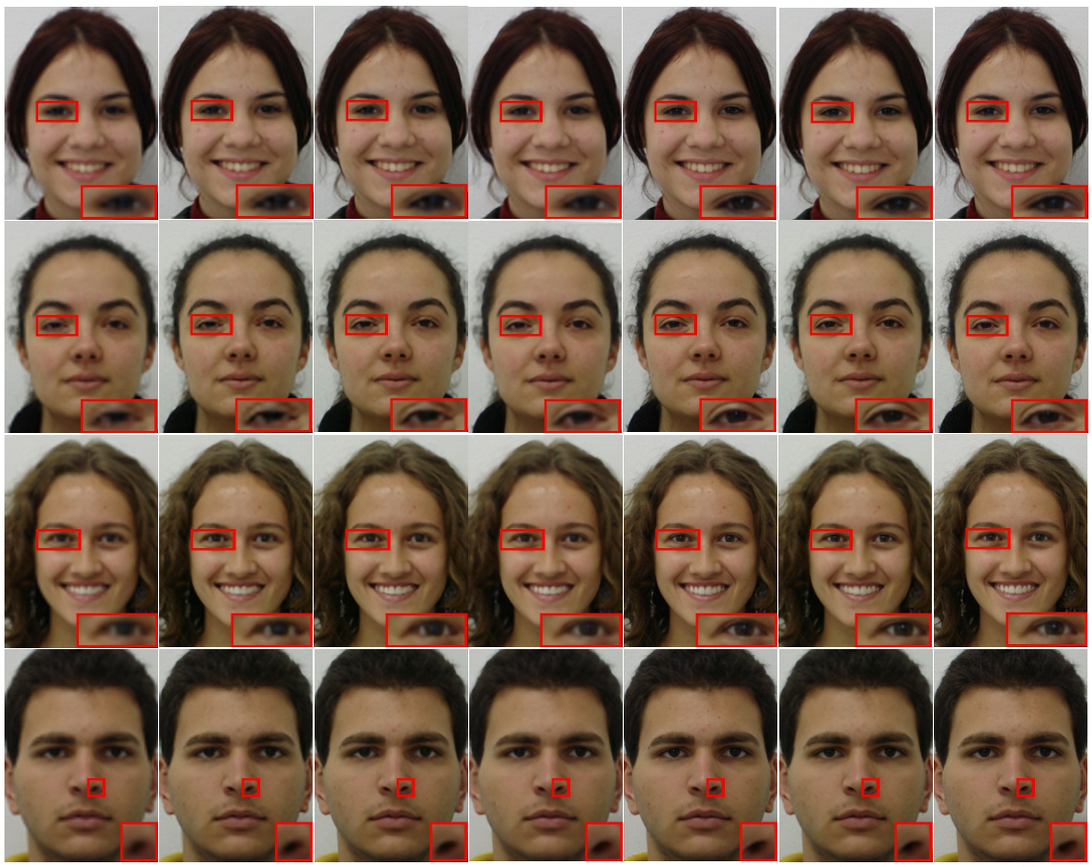
3.3.2 组件重建性能比较

为了体现本文算法的独特性，本文将单独与对比算法比较眼睛、鼻子和嘴巴部分(C1、C2和C3区域)的客观评估指标，结果如表2所示，本文算法的感兴趣区域都优于对比算法。从图6中可以看出：双三次插值法并不能产生额外的细节信息；基于深度学习的SRCNN和VDSR通用图像超分辨率方法，由于其全局优化方案，可以很好地保持局部的基本结构，但是无法恢复更多的高频细节信息；虽然EDGAN可以产生良好的视觉效果，但是本文方法与EDGAN相比具有更好的平滑度，而且在纹理细节上更加准确。实验结果表明，独立采用不同的深度学习模型重建感兴趣的区域可以产生更加丰富的细节信息。

表2 不同算法不同组件的PSNR和SSIM的结果比较

Tab. 2 Comparison of PSNR and SSIM results of different algorithms and components

不同组件	SRCNN		VDSR		LCGE		EDGAN		本文算法	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
C1	38.50	0.941 3	39.70	0.952 8	39.16	0.949 1	40.30	0.958 5	41.30	0.969 0
C2	40.76	0.966 7	41.68	0.970 8	41.07	0.966 0	41.82	0.965 8	42.99	0.977 1
C3	39.60	0.952 4	40.60	0.960 5	39.95	0.955 2	40.61	0.958 6	41.66	0.969 7



(a) Bicubic　(b) SRCNN　(c) VDSR　(d) LCGE　(e) EDGAN　(f) 本文算法　(g) HR图像

图6　在FEI数据集中本文算法与不同算法的主观比较

Fig. 6　Subjective comparison of the proposed method with other algorithms on dataset FEI

3.3.3　人脸重建性能比较

本文对最终合成的人脸图像与对比算法进行比较,如表3所示,最终实验结果的PSNR、SSIM和VIF平均超过通用图像SR算法SRCNN、VDSR分别为1.20 dB/0.008 5/0.044 2,0.24 dB/0.002 4/0.016 8,超过人脸图像SR算法LCGE和EDGAN分别为1.23 dB/0.009 5/0.043 7,1.11 dB/0.013 9/0.066 7。与LCGE相比,本文算法的测试时间减小了99%。虽然本文算法的测试时间比Bicubic、SRCNN、VDSR稍微久一些,但是本文算法的PSNR、SSIM和VIF客观评估指标值超越了它们。实验结果表明,无论是组件还是最终合成的人脸图像均获得了较高的客观评分。

表3　不同算法PSNR,SSIM,VIF和测试时间的结果比较

Tab. 3　Comparisonof PSNR,SSIM, VIF and testing time results of different algorithms

算法	PSNR/dB	SSIM	VIF	测试时间/s
Bicubic	36.25	0.941 8	0.646 7	1.26
SRCNN	38.58	0.952 9	0.687 0	1.85
VDSR	39.54	0.959 0	0.714 4	1.44
LCGE	38.55	0.951 9	0.687 5	237.00
EDGAN	38.67	0.947 5	0.664 5	1.83
本文算法	39.78	0.961 4	0.731 2	1.62

4　结语

本文提出了一种基于组合学习的人脸超分辨率算法,利用不同的网络结构重建性能优势,不仅能够恢复人脸重要器官的纹理细节信息,还能利用现有的其他超分辨率方法提供人脸背景重建信息,利用多种不同的超分辨率重建方法组合学习获得更精准的重建图像,并使用区域融合网络自适应融合不同深度学习模型重建的图像块,从而拓展了图像重建先验的来源,进一步提升了人脸超分辨率的重建性能。在后续的研究工作中,扩展更大的人脸图像数据集,并扩展到通用SR(Super-Resolution)中,来表明此方法的拓展性。并且在不同深度学习模型自适应融合工作中还存在改进的空间,可以进一步提升此方法的重建性能。

参考文献(References)

[1] 傅天宇,金柳颀,雷震,等. 基于关键点逐层重建的人脸图像超分辨率方法[J]. 信号处理, 2016, 32(7):834-841.(FU T Y, JIN L Q, LEI Z, et al. Face super-resolution method based on key points layer by layer[J]. Journal of Signal Processing, 2016, 32(7): 834-841.)

[2] 姜杰,刘哲,吕林涛. 局部线性嵌入的快速单幅图像超分辨率技术[J]. 红外技术, 2018,40(1):39-46.(JIANG J, LIU Z, LYU L T. Fast single-image super resolution technique based on local linear embedding[J]. Infrared Technology, 2018, 40(1): 39-46.)

[3] 徐志刚,李文文,朱红蕾,等. 一种基于$L_{1/2}$正则约束的超分辨率重建算法[J]. 华中科技大学学报(自然科学版), 2017, 45(6): 38-42.(XU Z G, LI W W, ZHU H L, et al. Super-resolution reconstruction based on $L_{1/2}$ regularization[J]. Journal of Huazhong University of Science and Technology (Nature Science Edition), 2017, 45(6): 38-42.)

[4] JIANG J, HU R, HAN Z, et al. Locality-constraint iterative neighbor embedding for face hallucination[C]// Proceedings of the 2013

IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2013: 1-6.

[5] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.

[6] KIM J, KWON L, MU L. Accurate image super-resolution using very deep convolutional networks[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1646-1654.

[7] 卢涛,汪家明,李晓林,等. 基于中继循环残差网络的人脸超分辨率重建[J]. 华中科技大学学报(自然科学版), 2018, 46(12): 95-100. (LU T, WANG J M, LI X L, et al. Boosted face hallucination via relay residual recurrent neural network (R3N2)[J]. Journal of Huazhong University of Science and Technology (Nature Science Edition), 2018, 46(12): 95-100.)

[8] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2017,1: 105-114.

[9] YANG X, LU T, WANG J, et al. Enhanced discriminative generative adversarial network for face super-resolution[C]// Proceedings of the 2018 Pacific Rim Conference on Multimedia, LNCS 11165. Cham: Springer, 2018: 441-452.

[10] SONG Y, ZHANG J, HE S, et al. Learning to hallucinate face images via component generation and enhancement[EB/OL]. [2019-01-11]. https://arxiv.org/pdf/1708.00223.pdf.

[11] JIANG J, YU Y, HU J, et al. Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination[EB/OL]. [2019-01-11]. https://arxiv.org/pdf/1806.10726.pdf.

[12] 黄全亮,刘水清,孙金海,等. 融合学习算法的单帧图像超分辨率复原[J]. 计算机工程与应用, 2013, 49(23): 186-190. (HUANG Q L, LIU S Q, SUN J H, et al. Single image super-resolution combined with learning algorithm[J]. Computer Engineering and Applications, 2013, 49(23): 186-190.)

[13] AI N, PENG J, ZHU X, et al. Single image super-resolution by combining self-learning and example-based learning methods[J]. Multimedia Tools and Applications, 2016, 75(11): 6647-6662.

[14] LEVINSHTEIN A, STERE A, KUTULAKOS K N, et al. TurboPixels: fast superpixels using geometric flows[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2290-2297.

[15] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.

[16] YU X, PORIKLI F. Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3760-3768.

[17] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[18] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444.

This work is partially supported by the National Natural Science Foundation of China (61502354), the Natural Science Foundation of Hubei Province (2015CFB451), the Special Fund of Central Guidance for Local Science and Technology Development (2018ZYYD059), the Scientific Research Foundation of Wuhan Institute of Technology (K201713), the Graduate Innovation Foundation of Wuhan Institute of Technology (CX2018211).

XU Ruobo, born in 1990, M. S. candidate. His research interests include image processing, computer vision.

LU Tao, born in 1980, Ph. D., associate professor. His research interests include image/video processing, computer vision, artificial intelligence.

WANG Yu, born in 1996, M. S. candidate. His research interests include image processing, computer vision.

ZHANG Yanduo, born in 1971, Ph. D., professor. His research interests include intelligent computing, intelligent robot, intelligent detection, simulation system.