

基于注意力机制的人脸超分辨率重建方法

鲍娴婧，吴成红

宁波工程学院国际交流学院，浙江 宁波

收稿日期：2025年10月28日；录用日期：2025年12月31日；发布日期：2026年1月15日

摘要

人脸超分辨率重建技术旨在从低分辨率图像中恢复高分辨率人脸，但其核心挑战在于大倍率放大时高频细节的严重丢失。现有方法虽尝试引入人脸结构化先验信息(如关键点与解析图)以提供结构约束，但仍普遍存在模型复杂、特征利用不充分等问题。本文提出一种基于高效通道注意力与结构化先验融合的双阶段人脸超分网络。本方法的核心创新在于：1) 将人脸边缘图与人脸解析图作为互补的结构化先验信息，共同引导重建过程，以精准恢复面部几何结构；2) 在网络的关键位置引入高效通道注意力模块，该模块以近乎零计算开销增强了对关键特征的利用能力，有效提升了纹理细节的恢复质量。在CelebA Mask-HQ与Helen数据集上的实验表明，本方法在主观视觉质量与客观指标(PSNR/SSIM)上均优于现有主流方法。特别地，通过系统的消融研究，我们验证了ECA模块相较于其他注意力机制(如SENet、CBAM)在性能与效率上的优越性，以及其与结构化先验信息结合的协同增强效应。

关键词

人脸超分辨率，注意力机制，边缘信息，先验信息

Face Super-Resolution Reconstruction Method Based on Attention Mechanism

Xianjing Bao, Chenghong Wu

International Exchange College, Ningbo University of Technology, Ningbo Zhejiang

Received: October 28, 2025; accepted: December 31, 2025; published: January 15, 2026

Abstract

Face super-resolution reconstruction technology aims to recover high-resolution face images from low-resolution inputs. However, its core challenge lies in the severe loss of high-frequency details during large-scale upscaling. Although existing methods have attempted to incorporate facial prior information (such as keypoints and parsing maps) to provide structural constraints, they still

commonly suffer from issues like model complexity and insufficient feature utilization. This paper proposes a dual-stage face super-resolution network based on the fusion of an Efficient Channel Attention mechanism and structured priors. The core innovations of our method are as follows: 1) Facial edge maps and facial parsing maps are utilized as complementary structural prior information to jointly guide the reconstruction process, enabling the precise recovery of facial geometry. 2) The Efficient Channel Attention (ECA) module is introduced at critical positions within the network. This module enhances the utilization of key features with nearly zero computational overhead, effectively improving the recovery quality of texture details. Experiments on the CelebA Mask-HQ and Helen datasets demonstrate that our method outperforms existing mainstream methods in both subjective visual quality and objective metrics (PSNR/SSIM). Specifically, through systematic ablation studies, we validate the superiority of the ECA module over other attention mechanisms (e.g., SENet, CBAM) in terms of both performance and efficiency, as well as the synergistic enhancement effect achieved by its combination with prior information.

Keywords

Face Super-Resolution, Attention Mechanism, Edge Information, Prior Information

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图像超分辨率技术通过特定的算法将输入的低分辨率图像经过学习，最后得到高分辨率图像。人脸超分辨率技术是针对人脸这个特殊结构的图像超分辨率技术，旨在将低分辨率人脸通过某种技术转换为高分辨率人脸。但是人脸结构比较特殊，不像平常的图像，它具有高强度的结构相似性和身份信息的细节差异性，它的重建难度更大，要求更高，重建过程中，我们要保证几何特征的一致性，还要注意纹理信息的准确恢复。因此，人脸图像超分辨率重建具有极大的挑战。人脸超分辨这一概念最早由 Baker 和 Kanada [1] 在 2000 年提出来的，它是图像超分辨领域中的一个分支，专门针对人脸这一特殊场景进行超分辨。近年来，深度学习技术在图像处理方面应用广泛，因此人脸超分辨领域也开始结合深度学习技术，从此人脸超分辨领域开始进入一个新的发展阶段。

基于深度学习的人脸超分辨率技术按照网络结构的不同可以分成：基于插值的人脸超分辨率重建、基于重构的人脸超分辨率重建、基于卷积神经网络的人脸超分辨方法和基于对抗生成网络的人脸超分辨方法。Dong [2] 等人提出了 SRCNN 模型，第一次将深度学习应用到图像超分辨率。SRCNN 首先使用双三次插值将 LR 图像放大到目标大小，然后通过三层卷积神经网络提取图像特征，最后生成 HR 图像，极大提高了重构效果；Huang D 和 Liu H 提出了一种基于 SRCNN 网络的优化算法 SRCNN-IBP 算法[3]，将 SRCNN 网络和迭代反投影算法(IBP)结合起来，利用 SRCNN 网络生成高分辨图像下采样成低分辨图像，SRCNN-IBP 算法可以看成在 SRCNN 算法的基础上引入了高分辨率图像的先验，所以在重建图像的质量上要优于 SRCNN 算法，同时说明了结构化先验信息对于人脸超分辨率重建是比较重要的；Leding 等将 GAN 用在了解决超分辨率问题上，提出了一种用于图像超分辨率的生成对抗网络 SRGAN [4]，使用一个经过训练的鉴别器网络来区分 SR 图像和原始真实图像。Yu Chen [5] 等人提出了加入结构化先验信息的人脸超分辨率重建方法 FSRNet，该方法使用堆叠沙漏网络模型提取人脸几何信息，结果表明利用人脸关键点和人脸解析图可以提升人脸恢复效果，但是生成的人脸图像纹理细节不充分，模型比较复杂，耗费大量时间。

为了加强结构化先验信息对人脸结构的恢复效果，并提升网络对关键特征的选择与利用能力，本文提出了一种融合高效通道注意力与结构化先验的人脸超分辨率重建网络。本文的主要贡献可概括为以下三点：

- 1) 提出了人脸边缘图与解析图的双重先验融合机制，通过几何约束与语义引导的互补作用，显著提升了面部结构的重建精度；
- 2) 首次在人脸超分任务中系统验证了 ECA 注意力机制相对于 SENet、CBAM 的优越性，实现了性能与效率的最佳平衡；
- 3) 通过严谨的消融实验揭示了结构化先验信息与注意力机制的协同作用机理，为后续研究提供了理论基础。

2. 注意力机制

注意力机制[6]-[8]是深度学习领域的一个应用，本质就是让网络定位到感兴趣的信息，抑制无用信息，结果通常都是以概率图或者概率特征向量的形式展示，在图像处理中，主要包含通道注意力和空间注意力两种基本形式。

2.1. 通道注意力模型

通道注意力模型[9]旨在建模不同特征通道之间的相关性，通过自动学习每个通道的重要程度，为其分配不同的权重系数，从而强化重要特征，抑制非重要特征。计算过程如图 1 所示。

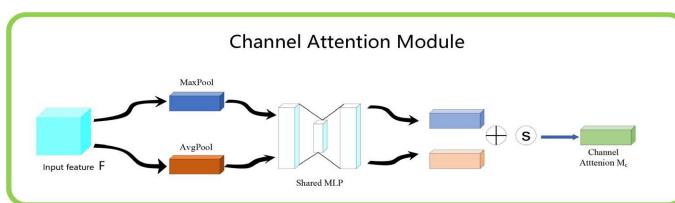


Figure 1. Channel attention model
图 1. 通道注意力模型

首先对输入特征图分别进行全局最大池化和全局平均池化，然后将两个池化结果送入共享的多层感知机[10] (MLP, multi-layer perception)。将 MLP 输出的特征进行元素加和，再通过 sigmoid 激活函数生成通道注意力特征图。最后将该特征图与输入特征图进行逐元素相乘，得到加权后的特征输出。

2.2. 空间注意力模型

空间注意力模型[11]旨在提升关键区域的特征表达，通过为每个空间位置生成权重掩膜，增强感兴趣目标区域，同时弱化不相关背景区域。计算过程如图 2 所示。

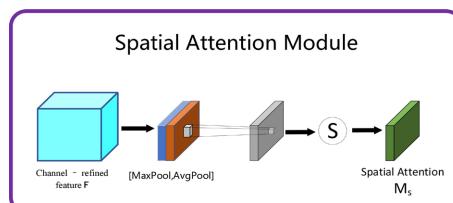


Figure 2. Spatial attention model
图 2. 空间注意力模型

以通道注意力输出的特征作为输入, 首先进行基于通道的全局最大池化和平均池化, 将两个结果按通道连接后, 通过卷积层降维为单通道, 再经 sigmoid 函数生成空间注意力特征图, 最后与输入特征图相乘得到最终输出。

2.3. 混合注意力模型

通道注意力易忽略空间信息交互, 而空间注意力则忽略了通道间信息交互。因此, 研究者提出了混合注意力模型[12] (CBAM), 依次进行通道注意力和空间注意力处理, 实现更全面的特征增强。

通道注意力机制的实现过程, 首先对输入特征层分别进行全局平均池化和全局最大池化, 将得到的两个特征通过共享的全连接层进行处理。将处理后的特征相加后, 通过 sigmoid 激活函数生成通道权重系数。最后, 将权重系数与原始输入特征层逐通道相乘, 完成特征重标定。

空间注意力机制的实现过程, 首先在通道维度上对输入特征层进行最大值池化和平均值池化, 将得到的两个特征图按通道维度拼接。随后通过一个卷积层进行融合并降维至单通道, 再经 sigmoid 函数生成空间权重图。最终将该权重图与输入特征层逐位置相乘, 实现空间区域的注意力增强。实现过程如图 3 所示。

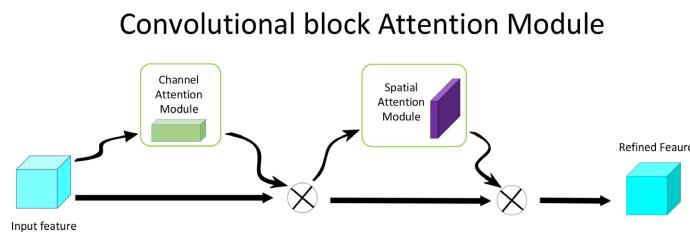


Figure 3. Mixed attention model
图 3. 混合注意力模型

2.4. 高效通道注意力模型

SENet [13] 是 2017 年在文献《Squeeze-and-Excitation Networks》中提出的, 全称 Squeeze-and-Excitation Networks。SENet 通过建模通道间依赖关系显著提升了网络性能, 但其全连接层带来的维度缩减和计算成本限制了应用。

因此, 2020 年 Qilong Wang 等人提出了高效通道注意力(ECA)模块[14], 该模块采用 1×1 卷积替代全连接层, 在避免维度缩减的同时大幅降低了计算复杂度。

ECA 模块结构如图 4 所示, 输入特征($H \times W \times C$)经全局平均池化得到 $1 \times 1 \times C$ 的特征向量, 随后通过卷积核大小为 k 的 1×1 卷积学习通道重要性, 最后通过逐通道相乘实现特征重标定。其中卷积核大小 k 通过自适应函数确定:

$$k = \varphi(C) = \left\lceil \frac{\log_2(C)}{\gamma} + b \right\rceil_{odd} \quad (1)$$

其中, k 表示卷积核大小, C 表示通道数; odd 表示取最接近的奇数; γ 和 b 在论文中设置为 2 和 1, 用于改变通道数 C 和卷积核大小之间的比例。

通道维度 C 和卷积核大小 k 成比例, 采用非线性函数, 而且卷积核数量设为 2 的 k 次方, 其中核大小 k 表示局部跨通道交互的覆盖范围。公式如下:

$$C = \mathcal{O}(k) = 2^{(\gamma * k - b)} \quad (2)$$

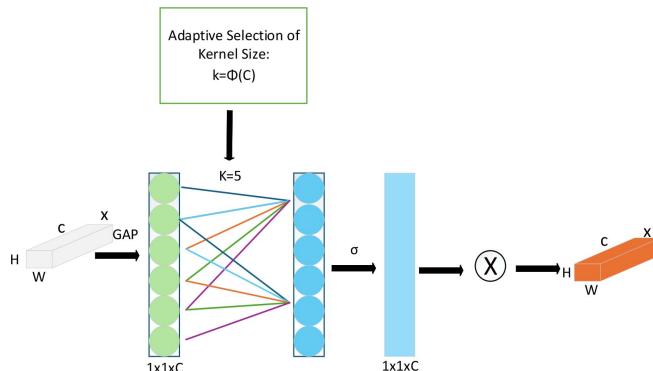


Figure 4. Efficient channel attention model
图 4. 高效通道注意力模型

该方法使卷积核大小与通道数保持比例关系，确保跨通道交互范围的适应性。将 ECA 模块引入人脸超分辨率模型，可有效提升图像恢复的客观指标与视觉质量。

3. 人脸结构化先验信息

3.1. 人脸解析图

人脸解析(Face Parsing)是将人脸的不同部分分割出来，通过语义分割技术将人脸划分为不同区域，可为超分辨率重建提供重要的几何结构先验。

语义图像分割往往需要丰富的空间信息和较大的感受野，传统的方法往往无法达到实时的推理速度，性能低下。Changqian Yu 设计了一种双边分割网络 BiSeNet [15]解决了这一问题，通过空间路径和上下文路径的双分支设计，在保持空间细节的同时获得足够感受野，实现了高效的人脸解析。FSRNet [5]将先验估计网络与超分辨率网络联合训练，验证了解析图对人脸重建的促进作用。

3.2. 人脸关键点坐标

人脸关键点坐标是另一类重要结构化先验信息。Super-FAN [16]将热图回归整合到网络中，通过热图损失约束超分辨率过程，使重建结果更接近真实分布。目前有许多优秀的人脸关键点检测网络，例如，Jianfeng Wang [17]提出了一种针对被遮挡人脸检测的高效人脸关键点检测网络；Bulat [18]在大规模数据集上创建了一个 2D 人脸关键点检测网络和一个更好的 3D 人脸关键点检测网络等。

3.3. 堆叠沙漏网络

堆叠沙漏网络[19] (Stacked Hourglass Network)采用编码器 - 解码器结构，通过多尺度特征融合和跳跃连接有效捕捉空间信息。其使用 1×1 卷积替代全连接层的设计使其能适应不同尺寸输入，而残差连接和中继监督机制则保障了特征的充分利用。将该网络引入人脸超分辨率任务，有助于提取丰富的多尺度面部特征。网络的具体结构如图 5 所示。

4. 基于结构化先验信息和注意力机制的人脸超分辨率重建

本章将详细介绍我们所提出的融合高效通道注意力与结构化先验的人脸超分网络。模型的核心思想是构建一个由先验信息引导、注意力机制增强的双阶段重建流程。具体而言，我们统一使用“人脸结构化先验信息”这一术语来指代共同作用于网络的人脸边缘图与人脸解析图。网络首先进行粗糙重建，随后利用估计出的先验信息，在 ECA 模块的辅助下，对特征进行校准与增强，最终完成精细重建。

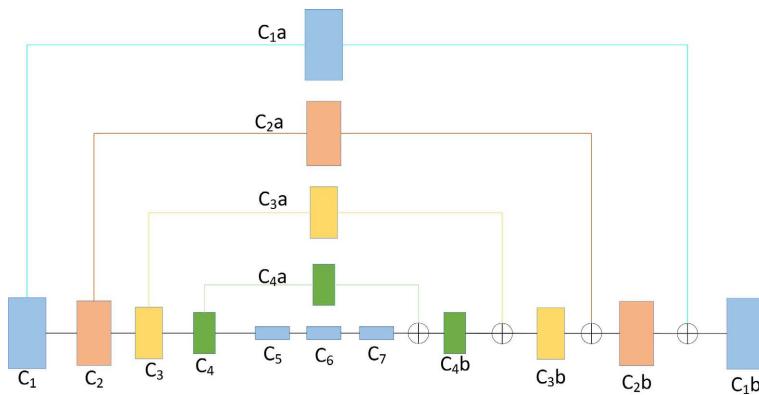


Figure 5. Stacked Hourglass network structure
图 5. Stacked Hourglass 网络结构

4.1. 网络总体结构

网络由四个核心模块组成, 其中, CSRNet 提取浅层特征, 生成粗糙高分辨率图像, DSRNet 从粗糙图像中提取深层特征, PENet 基于 HourGlass 结构估计人脸关键点热图与解析图, FRNet 融合图像特征与先验信息, 输出最终结果。网络总体结构如图 6 所示。

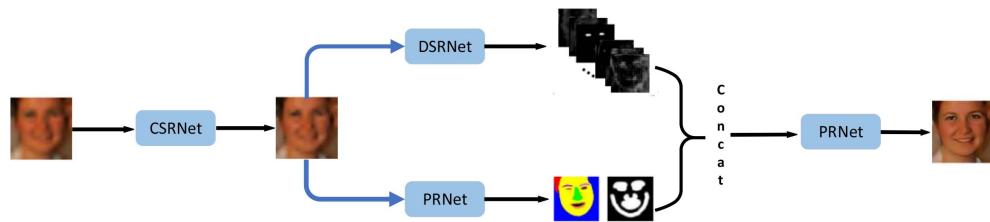


Figure 6. Overall network structure
图 6. 网络总体结构

4.2. 浅层特征提取网络

CSRNet 首先对双三次插值后的输入图像进行 3×3 卷积提取轮廓特征, 随后通过 3 个残差块进行非线性映射, 最后经卷积层和 ECA 注意力模块生成粗糙高分辨率图像。如图 7 所示。

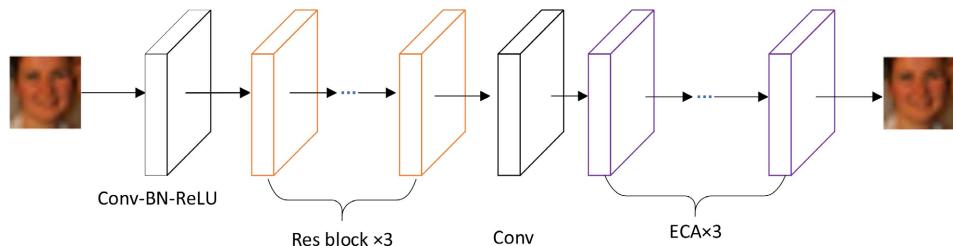


Figure 7. Shallow feature extraction network
图 7. 浅层特征提取网络

4.3. 深层特征提取网络

DSRNet 首先对粗糙图像进行降采样至 64×64 分辨率, 随后通过 12 个残差块提取深层特征, 最终

通过卷积层输出精细特征。网络结构如图 8 所示。

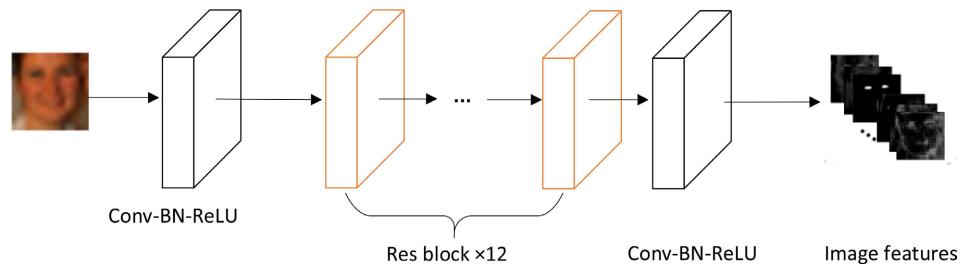


Figure 8. Deep feature extraction network

图 8. 深层特征提取网络

4.4. 先验估计网络

PENet 中采用 2 个 HourGlass 模块，通过编码器 - 解码器结构结合跳跃连接，从粗糙图像中估计人脸关键点热图和解析图。如图 9 所示。

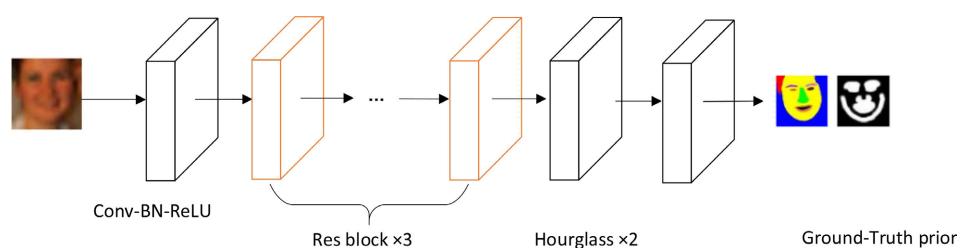


Figure 9. Prior estimation network

图 9. 先验估计网络

4.5. 精细重建网络

FRNet 首先拼接图像特征与先验特征，通过 3×3 卷积降维后经反卷积上采样，再通过 3 个残差块和 ECA 模块最终重建高分辨率图像，如图 10 所示。

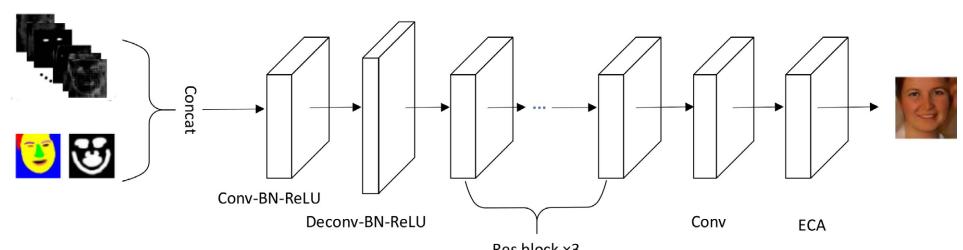


Figure 10. Fine reconstruction network

图 10. 精细重建网络

4.6. 损失函数

给定训练集 $\left\{x^{(i)}, \tilde{y}^{(i)}, \tilde{p}^{(i)}\right\}_{i=1}^N$ ， \tilde{y} 为真实高分辨率图像， \tilde{p} 为真实的人脸结构化先验信息(包含边缘图与解析图)真实图像。设 x 为输入的低分辨率图像，网络首先生成粗糙超分辨率结果：

$$y_c = C(x) \quad (3)$$

其中, C 表示粗糙重建映射。随后将 y_c 分别送入先验估计网络 p 和深层特征提取网络 F , 即

$$p = P(y_c), \quad f = F(y_c) \quad (4)$$

其中, f 为深层特征, p 为估计的结构化先验信息。最终通过精细重建网络 D 融合特征 f 和结构化先验信息 p , 得到最终的高分辨率图像 y 。即

$$y = D(f, p) \quad (5)$$

1) 像素损失

在图像超分辨率中, 通常使用均方误差(mean square error, MSE)损失可以获得较高的评价指标, 例如 PSNR 和 SSIM, 为避免 MSE 损失导致的高频信息丢失与过度平滑问题, 在此使用 L1 损失作为像素损失函数[4], 有

$$L_{pixel} = \frac{1}{2N} \left\{ \left\| \widetilde{y^{(i)}} - y_c^{(i)} \right\|^2 + \alpha \left\| \widetilde{y^{(i)}} - y^{(i)} \right\|^2 \right\} \quad (6)$$

2) 人脸先验损失

为约束人脸先验估计过程, 定义以下损失函数, 有

$$L_p = \frac{1}{2N} \left\{ \left\| \widetilde{p^{(i)}} - p^{(i)} \right\|^2 \right\} \quad (7)$$

3) 总损失

模型总损失为以上两项损失的加权组合, 即

$$L_F(\theta) = \frac{1}{2N} \left\{ \left\| \widetilde{y^{(i)}} - y_c^{(i)} \right\|^2 + \alpha \left\| \widetilde{y^{(i)}} - y^{(i)} \right\|^2 + \beta \left\| \widetilde{p^{(i)}} - p^{(i)} \right\|^2 \right\} \quad (8)$$

其中, N 表示训练样本数, α 、 β 表示权重系数, $\widetilde{y^{(i)}}$ 是真实的高分辨率图像, $y_c^{(i)}$ 是粗糙的高分辨率恢复图像, $y^{(i)}$ 是精细的高分辨率恢复图像, $\widetilde{p^{(i)}}$ 是真实的结构化先验信息, $p^{(i)}$ 是估计的结构化先验信息。

5. 实验数据与环境

5.1. 实验数据

CelebA [20] 数据集是香港中文大学开源大规模的人脸检测基准数据集。实验采用 Helen [21] 数据集进行补充验证, 随机选取 1200 张训练, 400 张测试。主体实验采用 CelebA Mask-HQ [22] 数据集, 该数据集包含 30,000 张分辨率为 1024×1024 的高清人脸图像, 每张图像均标注有 19 类面部组件的语义掩码, 实验从中随机划分出 17,000 张作为训练集, 剩余 13,000 张作为测试集。数据集示例如图 11 所示。



Figure 11. Partial image of the CelebA Mask-HQ dataset

图 11. CelebA Mask-HQ 数据集部分图像

5.2. 实验设置

实验使用 Torch7 框架, 输入图像统一裁剪为 128×128 尺寸, 使用 RMSprop 算法(root mean square prop)训练模型, 初始学习率为 2.5×10^{-4} , 最小批量为 14, 低分辨率图像通过双三次插值放大至目标尺寸后输入网络。

5.3. 消融实验

为系统验证本文所提各模块的有效性及其协同作用, 我们设计了严格的消融实验。所有实验均在相同的训练设置、数据集(CelebA Mask-HQ)和评估指标下进行。

1) 注意力机制类型对比实验

为验证高效通道注意力模块在本框架中的选择依据, 我们将其与当前主流的通道注意力(SENet)和混合注意力(CBAM)进行了全面的性能与效率对比。基线模型不包含任何注意力模块, 结果如表 1 所示。

Table 1. Effects of different attention mechanisms on network performance (in CelebA Mask-HQ dataset)

表 1. 不同注意力机制对网络性能的影响(在 CelebA Mask-HQ 数据集)

注意力机制	PSNR (dB)	SSIM	参数量(M)	推理时间(ms)
无	22.35	0.7008	1.20	12.5
SENet [13]	23.10	0.7250	1.52	15.2
CBAM [12]	23.25	0.7280	1.61	16.8
ECA (Ours)	23.64	0.7313	1.28	13.1

实验数据表明, 三种注意力机制均能提升基线模型性能, 但其增益与代价各不相同。SENet 通过显式建模通道依赖带来提升, 但其全连接层造成的维度缩减会导致信息丢失, 并增加参数量。CBAM 通过串行的通道与空间注意力进一步提升了性能, 但随之而来的是最高的计算复杂度。相比之下, 本文采用的 ECA 模块取得了最优的 PSNR 与 SSIM。这一结果证明, ECA 在性能与效率之间取得了最佳平衡, 特别适合集成于需要密集特征处理的人脸超分网络中。

2) 结构化先验信息与注意力机制的协同效应

本文的核心论点是结构化先验信息与高效注意力机制能够产生协同效应。为验证此点, 我们构建了四个对比模型, 结果如表 2 所示。

Table 2. Synergistic analysis of prior information and attention mechanism

表 2. 先验信息与注意力机制的协同效应分析

模型配置	PSNR (dB)	SSIM
基线(无先验, 无 ECA)	22.35	0.7008
仅先验信息	23.21	0.7294
仅 ECA 注意力	23.52	0.7315
先验信息 + ECA (完整模型)	23.64	0.7313

由表 2 可知, 单独引入结构化先验信息或 ECA 模块均能带来显著的性能提升。然而, 当二者结合时, 模型达到了最优性能。这表明二者在本框架中扮演着互补且协同的角色: 结构化先验信息从宏观上为网

络提供了可靠的面部几何约束, 确保了重建图像的结构正确性; 而 ECA 模块则从微观上增强了网络对关键纹理特征(如眼睛、嘴唇轮廓)的提取与利用能力。先验信息引导网络应该在哪里生成什么, 而 ECA 则帮助网络更好地生成那里的细节。这种协同作用使得完整模型在客观指标上超越了任何单一组件带来的增益。

3) ECA 模块插入数量的影响

我们进一步探究了在 CSRNet 中插入 ECA 模块的数量对最终性能的影响, 以确定最优配置。结果如表 3 所示。

Table 3. Impact of the number of ECA modules on network performance

表 3. ECA 模块数量对网络性能的影响

在 CSRNet 中添加 ECA 的个数	CelebA Mask-HQ	Helen
0	23.21/0.7294	23.19/0.7105
1	23.52/0.7315	23.31/0.6684
2	23.60/0.7312	23.45/0.6649
3	23.64/0.7313	23.46/0.6825
4	23.64/0.7315	23.46/0.6829
5	23.58/0.7317	23.40/0.6821

当 ECA 数量从 0 增加到 3 时, PSNR 与 SSIM 指标呈现稳定上升趋势, 表明适度的特征校准对性能提升具有积极作用。然而, 当数量超过 3 个后, 性能提升趋于饱和并出现轻微波动。这一现象表明: 过度的注意力校准可能引入特征冗余, 或导致优化曲面过于复杂, 从而影响模型收敛。基于此实验结果, 我们选择在 CSRNet 中插入 3 个 ECA 模块作为最终配置, 以实现性能与复杂度的最优平衡。

4) 堆叠沙漏网络数量的影响

先验估计网络的容量对先验信息的质量具有决定性影响。我们通过改变其中堆叠的沙漏网络数量来探究其影响规律, 结果如表 4 所示。

Table 4. Impact of the number of hourglasses on network performance

表 4. Hourglass 数量对网络性能的影响

Hourglass 数量	CelebA Mask-HQ	Helen
1	17.73/0.42	19.23/0.43
2	23.55/0.70	24.83/0.65
3	21.23/0.66	22.23/0.63
4	19.04/0.59	20.56/0.61

当使用 1 个 Hourglass 时, 网络容量不足, 难以准确捕捉人脸的多尺度结构特征, 导致先验信息质量较差, 最终重建性能显著下降。当使用 2 个 Hourglass 时, 网络具备了足够的表达能力来提取精确的人脸结构信息, 从而生成高质量的先验, 并带来最优的重建效果。然而, 当数量进一步增加到 3 个和 4 个时, 性能出现明显下降。这一现象可能源于两个因素: 过深的先验估计网络增加了训练难度, 同时容易过拟合训练集的人脸结构分布, 从而削弱了模型的泛化能力。因此, 采用 2 个堆叠沙漏网络能够在模型容量与泛化性能之间取得最佳平衡。

5.4. 对比实验

为全面评估模型性能，采用主观评价与客观评价相结合的方式。主观评价通过视觉感知判断图像清晰度与自然度；客观评价采用峰值信噪比(PSNR)与结构相似度(SSIM)作为量化指标，其计算公式如下：

PSNR 的计算公式如下：

$$\text{PSNR} = 10 * \log_{10} \frac{255^2}{\text{MSE}} \quad (9)$$

$$\text{MSE} = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w (X_1(i, j) - X_2(i, j))^2 \quad (10)$$

$$\text{SSIM}_{(x,y)} = l(x, y)^\alpha * c(x, y)^\beta * s(x, y)^\gamma \quad (11)$$

其中，MSE 是均方误差， $l(x, y)$ 是图像 x 和图像 y 之间的亮度对比， $c(x, y)$ 是图像 x 和图像 y 之间的对度对比， $s(x, y)$ 是图像 x 和图像 y 之间的结构对比。SSIM 的取值范围从 0 到 1，取值越大效果越好。

1) 定量分析

在 CelebA Mask-HQ 和 Helen 数据集上，将本文模型与 Biubic [23]、SRCNN [24]、EDSR [25]、URDGN [26]、FSRNet [5] 进行对比，如表 5 所示，本文方法在 PSNR 指标上分别达到 23.64 dB 与 24.83 dB，较 FSRNet 提升 0.43 dB 与 0.34 dB，SSIM 值亦保持领先，验证了模型改进的有效性。

Table 5. Evaluation results of different models on CelebA Mask-HQ and Helen datasets

表 5. 不同模型在 CelebA Mask-HQ 和 Helen 数据集上的评价结果

Dataset	Bicubic	SRCNN	EDSR	URDGN	FSRNet	Ours
CelebA Mask-HQ	22.36/0.61	22.53/0.63	23.02/0.63	20.70/0.50	23.21/0.73	23.64/0.74
Helen	22.59/0.62	23.05/0.65	23.18/0.64	21.65/0.55	24.49/0.72	24.83/0.62

2) 定性分析

视觉对比结果如图 12、图 13 所示。相较于基线模型，本文方法在眼睛、嘴唇等关键区域有效减少了失真与模糊现象，纹理细节更加清晰，面部结构保持更为完整，体现出更优的视觉重建质量。

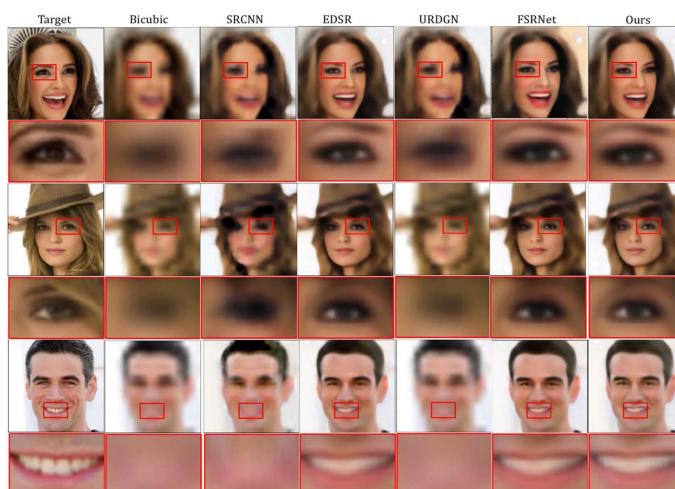


Figure 12. Comparison diagram of different algorithms

图 12. 不同算法对比图

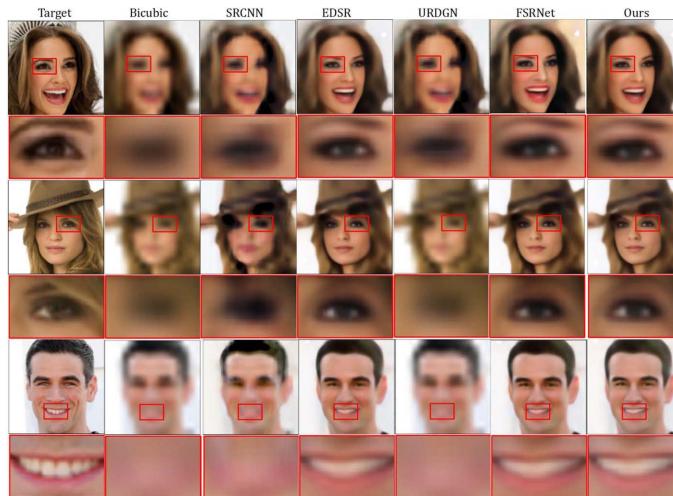


Figure 13. Detail enlargement comparison
图 13. 细节放大图对比

6. 结束语

本文提出了一种融合高效通道注意力与结构化先验信息的人脸超分辨率重建网络。通过系统性的模型设计与实验验证，我们得出以下结论：第一，人脸边缘图与解析图作为互补先验，能有效约束重建过程的面部结构；第二，ECA 注意力机制能以极低的计算成本显著提升模型对关键特征的表征能力，其在性能与效率的平衡上优于 SENet 与 CBAM；第三，先验信息与注意力机制之间存在协同增强效应，二者结合能引导网络生成细节更丰富、结构更准确的高分辨率人脸图像。

未来的工作将集中于两方面：一是在精细重建网络中探索迭代或递归机制，以逐步重建结果；二是研究动态注意力机制，使其能自适应不同区域的重建难度，进一步提升在极端低质输入下的鲁棒性。

参考文献

- [1] Baker, S. and Kanade, T. (2000) Hallucinating Faces. *IEEE International Conference on Automatic Face & Gesture Recognition*, Grenoble, 28-30 March 2000, 83-88.
- [2] Chao, D., Chen, C.L., He, K., et al. (2014) Learning a Deep Convolutional Network for Image Super-Resolution. Springer International Publishing.
- [3] Liu, H., Han, J.G., Hou, S.D., Shao, L. and Ruan, Y. (2018) Single Image Super-Resolution Using a Deep Encoder-Decoder Symmetrical Network with Iterative Back Projection. *Neurocomputing*, **282**, 52-59.
<https://doi.org/10.1016/j.neucom.2017.12.014>
- [4] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 105-114. <https://doi.org/10.1109/cvpr.2017.19>
- [5] Yu, C., Ying, T., Liu, X., et al. (2018) FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, 18-22 June 2018, 2492-2501.
- [6] Zhang, H., Goodfellow, I., Metaxas, D., et al. (2019) Self-Attention Generative Adversarial Networks.
<https://doi.org/10.48550/arXiv.1805.08318>
- [7] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., et al. (2017) Residual Attention Network for Image Classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6450-6458. <https://doi.org/10.1109/cvpr.2017.683>
- [8] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y. (2018) Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In: *Lecture Notes in Computer Science*, Springer, 294-310.
https://doi.org/10.1007/978-3-030-01234-2_18
- [9] Bastidas, A.A. and Tang, H. (2019) Channel Attention Networks. *2019 IEEE/CVF Conference on Computer Vision and*

- Pattern Recognition Workshops (CVPRW)*, Long Beach, 16-17 June 2019, 881-888.
<https://doi.org/10.1109/cvprw.2019.00117>
- [10] Longstaff, I.D. and Cross, J.F. (1987) A Pattern Recognition Approach to Understanding the Multi-Layer Perception. *Pattern Recognition Letters*, **5**, 315-319. [https://doi.org/10.1016/0167-8655\(87\)90072-9](https://doi.org/10.1016/0167-8655(87)90072-9)
- [11] Zhou, X.Z., Cheng, D.Z., Zhang, Z., et al. (2019) An Empirical Study of Spatial Attention Mechanisms in Deep Networks. *2019 International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 6687-6696.
- [12] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: *Lecture Notes in Computer Science*, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [13] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [14] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 11531-11539. <https://doi.org/10.1109/cvpr42600.2020.01155>
- [15] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G. and Sang, N. (2018) Bisenet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In: *Lecture Notes in Computer Science*, Springer, 334-349. https://doi.org/10.1007/978-3-030-01261-8_20
- [16] Bulat, A. and Tzimiropoulos, G. (2018) Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses with Gans. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 109-117. <https://doi.org/10.1109/cvpr.2018.00019>
- [17] Wang, J., Yuan, Y. and Yu, G. (2017) Face Attention Network: An Effective Face Detector for the Occluded Faces. <https://arxiv.org/abs/1711.07246>
- [18] Bulat, A. and Tzimiropoulos, G. (2017) How Far Are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks). *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 1021-1030. <https://doi.org/10.1109/iccv.2017.116>
- [19] Newell, A., Yang, K. and Deng, J. (2016) Stacked Hourglass Networks for Human Pose Estimation. In: *Lecture Notes in Computer Science*, Springer, 483-499. https://doi.org/10.1007/978-3-319-46484-8_29
- [20] Liu, Z., Luo, P., Wang, X. and Tang, X. (2015) Deep Learning Face Attributes in the Wild. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 3730-3738. <https://doi.org/10.1109/iccv.2015.425>
- [21] Zhou, F., Brandt, J. and Lin, Z. (2013) Exemplar-Based Graph Matching for Robust Facial Landmark Localization. *2013 IEEE International Conference on Computer Vision*, Sydney, 1-8 December 2013, 1025-1032. <https://doi.org/10.1109/iccv.2013.131>
- [22] Lee, C., Liu, Z., Wu, L. and Luo, P. (2020) MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 5548-5557.
- [23] Xiang, F.U. and Guo, B.L. (2009) Overview of Image Interpolation Technology. *Computer Engineering and Design*, **30**, 141-157.
- [24] Dong, C., Loy, C.C., He, K.M., et al. (2016) Image Super-Resolution Using Deep Convolutional Networks. <https://doi.org/10.1109/TPAMI.2015.2439281>
- [25] Lim, B., Son, S., Kim, H., Nah, S. and Lee, K.M. (2017) Enhanced Deep Residual Networks for Single Image Super-Resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, 21-26 July 2017, 1132-1140. <https://doi.org/10.1109/cvprw.2017.151>
- [26] Yu, X. and Porikli, F. (2016) Ultra-Resolving Face Images by Discriminative Generative Networks. In: *Lecture Notes in Computer Science*, Springer, 318-333. https://doi.org/10.1007/978-3-319-46454-1_20