**文章编号:** 1672-8785(2023)04-0033-09

# 基于多尺度优化和动态特征融合的 图像去模糊研究

万园园1 宋卓达1 陈小林1 朱鑫鑫2

(1.中国科学院长春光学精密机械与物理研究所,吉林长春 130033;2.中国人民解放军 63618 部队,新疆 库尔勒 841000)

摘 要:目前采用 U-Net 结构的去模糊算法存在细节损失、图像质量欠佳等问题,因此对 U-Net 进行改进,提出一种基于多尺度优化和动态特征融合的图像 去模糊方法。首先针对细节损失,提出一种精简且有效的多尺度残差注意力模块(Multi-Scale Residual Module, MSRM),通过增加特征尺度多样性来提取更精 细的图像特征。此外,为了将更有利的特征传递到解码部分,在跳跃连接处设 计动态特征融合模块(Dynamic Feature Fusion Module, DFFM),采用注意力加权 的方式选择性融合不同阶段的编码特征。该算法采用多尺度内容损失和多尺度 高频信息损失进行约束训练。在 GoPro 和 RealBlur 数据集上的实验结果表明, 这种方法能有效改善图像质量,复原更丰富的细节信息。与现有去模糊算法相 比,本文算法在主观视觉和客观评价等方面均具有一定优势。

关键词:图像去模糊;特征加权;多尺度特征;U-Net结构

中图分类号: TP391.41 文献标志码: A DOI: 10.3969/j.issn.1672-8785.2023.04.005

## Research on Image Deblurring Based on Multi-Scale Optimization and Dynamic Feature Fusion

WAN Yuan-yuan<sup>1</sup>, SONG Zhuo-da<sup>1</sup>, CHEN Xiao-Iin<sup>1</sup>, ZHU Xin-xin<sup>2</sup>

 (1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;
 2. Unit 63618 of PLA, Korla 841000, China)

**Abstract:** At present, U-Net-based image deblurring algorithms have some problems, such as detail loss and poor image quality. Therefore, the U-Net structure is improved, and an image deblurring method based on multi-scale optimization and dynamic feature fusion is proposed in this paper. Firstly, according to detail loss, a simplified and effective MSRM is proposed to extract finer image features by increasing feature scale diversity. In addition, in order to transfer more favorable features to the decoding part, a dynamic feature fusion module is designed at the skip connection, which can selectively fuse different stages of encoding features by attention weighting. In this algorithm, multi-scale content loss and multi-scale high-frequency information loss

收稿日期: 2022-12-09 作者简介: 万园园(1996-),江西南昌人,硕士,主要从事图像处理与深度学习方面的研究。 E-mail:sumiwan11233@163.com 红 外

are used for constraint training. Experimental results on GoPro and RealBlur data sets show that the proposed method can effectively improve image quality and restore more detailed information. Compared with the existing deblurring algorithms, the proposed algorithm has certain advantages in subjective vision and objective e-valuation.

Key words: image deblurring; feature weighting; multi-scale feature; U-Net structure

## 0 引言

图像去模糊的主要目的就是去除模糊退化 以得到高质量的图像<sup>[1]</sup>。在图像获取过程中, 可能会出现物体快速运动、光线环境复杂以及 相机抖动等不可控因素,从而给图像带来不同 程度的模糊退化。这些模糊退化通常会影响我 们的视觉体验,并且对后续的一些高维视觉任 务带来不利影响。通常来说,图像去模糊是一 个"病态问题",每一张模糊图像都存在多种 复原可能性。因此,图像去模糊是一个有意义 又充满挑战的研究方向。

鉴于深度学习凭借神经网络(Neural Networks, NNs)能够从大量的数据中学习特征信息,因此基于深度学习的去模糊算法已经在图 像去模糊领域占据着重要的位置。Sun J等 人<sup>[2]</sup>使用一个 CNN 模型结构来学习模糊核信息,然后利用得到的模糊核对模糊图像进行反 卷积,以达到去模糊的效果。然而这种复原方 式严重依赖于模糊核的预测准确性,而且不适 用于存在多种模糊核的情况。后来,随着深度 学习技术的发展,研究人员发现神经网络的能 力远不仅仅是预测模糊核。于是端到端的训练 方式开始被引入图像去模糊领域。神经网络开 始直接学习模糊图像和真实图像之间的复杂映 射关系,并直接预测输出去模糊的图像。

Nah S 等人<sup>[3]</sup>设计了一种由粗到精的多尺 度卷积结构模型。通过多尺度损失约束训练 后,直接对输入图像进行去模糊。但是该模型 没有在不同尺度的网络中采用参数共享,导致 模型训练需要花费较高的时间成本。为了解决 这个问题, Cho S J 等人<sup>[4]</sup>使用参数共享机制 以及跨层连接,在节约训练时间的同时还确保 网络学习到各个尺度下重要的特征信息。 此外,Kupyn O等人先后提出了 Deblur-GAN<sup>[5]</sup>和 DeblurGAN-v2<sup>[6]</sup>模型,利用生成对 抗网络进行图像去模糊。其中,DeblurGANv2模型采用一种轻量化的金字塔特征结构, 相较于其它经典算法提高了模型推理速度。 Zou W等人<sup>[12]</sup>提出单阶段网络模型,利用空 洞卷积和小波变化实现图像的去模糊化。

目前大部分基于深度学习的去模糊算法均 采用 U-Net 编码-解码结构,且已经取得了不 错的效果。但编码过程中的下采样操作易导致 图像信息损失,不利于复原图像的细节信息。 其次,U-Net 结构通常采用单一的跳跃连接, 并没有充分利用编码阶段的特征信息。

针对以上问题,本文在 U-Net 结构的基础 上进行改进,采用由粗到精的结构策略,提出 一种基于多尺度优化和动态特征融合的去模糊 方法。其主要贡献如下:(1)在编码-解码过程 中使用自主设计的 MSRM 模块进行特征提取。 该模块通过少量卷积层和相邻路径相连的方式 来获取不同尺度的特征信息,以更好地还原图 像细节。(2)在跳跃连接处设计一个多输入的 DFFM 模块,将不同阶段的编码特征通过注意 力加权的方式进行融合,从而自适应地从各分 支中选取更有用的特征信息传递给解码部分。 (3)将多尺度内容损失与多尺度高频信息损失 相结合作为损失函数,从多维度进行约束训 练。在公开的 GoPro 数据集<sup>[3]</sup>和 RealBlur 数据 集[13]上的实验结果显示,本文方法可还原更 丰富的细节信息,在主观视觉和客观指标上均 具有一定优势。

- 1 方法
- 1.1 网络结构

本文提出的算法模型采用了由粗到精的结

构策略,具有多尺度输入和多尺度输出。如图 1 所示,该网络主要包括编码阶段、跳跃连接 和解码阶段三部分。通常,编码阶段主要用于 特征信息提取,解码阶段用于图像复原重建, 而跳跃连接则负责将编码部分的特征传递到相 应的解码部分从而实现特征整合。其中, $B_k(k$ =1,2,3)表示输入的多尺度模糊图像, $S_k(k$ =1,2,3)表示输出的多尺度复原图像。

编码阶段包括三个编码模块(Encoder Module, EM): EM<sub>1</sub>、EM<sub>2</sub>和EM<sub>3</sub>。首先对输 入的模糊图像进行缩放,分别得到三张不同分 辨率的模糊图像 B<sub>1</sub>(256×256×3)、B<sub>2</sub>(128× 128×3)和 B<sub>3</sub>(64×64×3)。其中,B<sub>1</sub>可直接 输入到 EM<sub>1</sub>中,B<sub>2</sub>和 B<sub>3</sub>则需分别输入到相应 的浅层卷积模块(Shallow Convolution Module, SCM)中先进行浅层特征提取。EM<sub>1</sub>主要采用 一个卷积核为 3、步长为 1 的卷积层以及若干 个 MSRM 模块来进行特征提取; EM<sub>2</sub>和 EM<sub>3</sub> 先采用一个卷积核为 3、步长为 2 的卷积层进 行下采样,然后通过一个特征注意力模块 (Feature Attention Module, FAM)实现 EM<sup>6ut</sup><sub>k</sub>(k=1,2)和 SCM<sup>6ut</sup><sub>k</sub>(k=1,2)的特征融合,最后 用 MSRM 模块提取多尺度特征。其中, SCM 和 FAM 的详细结构如图 1 所示。

不同于传统的 U-Net 结构只将当前阶段的 编码特征传递给相应的解码特征,本文在跳跃 连接部分直接将三个编码模块的输出直接传送 到自主设计的 DFFM 模块中,通过注意力动 态融合之后再将特征传送到解码部分。

解码部分同样包含三个解码模块(Decoder Module, DM): DM<sub>1</sub>、DM<sub>2</sub>和 DM<sub>3</sub>。其中, DM<sub>3</sub>主要包括若干个 MSRM 模块,将特征同 时输入两路分支。一路通过转置卷积进行上采 样,用于下一阶段的图像重建;另一路输入一 个卷积层将通道数降成3,然后与输入的模糊 图像叠加后得到复原图 S<sub>3</sub>(64×64×3)。DM<sub>2</sub> 首先采用一个卷积层将拼接融合的上采样特征 和跳跃连接传递的特征进行通道降维,然后再 通过 MSRM 组合体实现特征重建,最后同样 输入两路分支实现上采样并输出当前尺度的复 原图 S<sub>2</sub>(128×128×3)。DM<sub>1</sub>同样先经过一个 卷积层和 MSRM 组合体,最后将通道数降为3 后与原始输入图像相加输出,得到原始尺度的 复原图 S<sub>1</sub>(256×256×3)。



图 1 网络整体结构



图 2 MSRM 模块

## 1.2 MSRM 模块

在各类视觉任务中,多尺度特征对感知细 节信息起着十分重要的作用<sup>[15]</sup>。通常,多尺 度特征的获取需采用大量的卷积层进行堆叠, 模型参数量较大。有鉴于此,本文提出一种精 简且有效的多尺度获取方法,以提取更精细的 特征信息。

为了引入更多的上下文信息,输入特征 x 在拆分之前需经过一个深度可分离卷积,得到 四路分支(见图 2)。第一路的输入特征经过一 个 3×3 卷积进行特征信息提取,随后将输出 的特征信息与另一路的输入特征相结合,再经 过一个 3×3 卷积提取特征。就这样一直重复, 直到处理完四路的输入特征。然后将这四路的 输出特征拼接起来,再经过一个1×1卷积进 行通道数降维以实现信息整合。在通道注意力 (见图 3)的作用下,对得到的特征进行通道权 重分配,加强重要的特征信息通道。最后再采 用残差连接,输出特征信息。由此可知,输入 特征有多种输出路径,每当经过3×3卷积就 相当于若干个卷积叠加,其等效感受野也增强 了。最后通过组合效应,相当于得到了不同尺 度的等效特征。

上过程可用式(1)~式(3)描述:  
$$x_1, x_2, x_3, x_4 = spilt(H_{denv}^{3\times 3}(x))$$
 (1)

$$\begin{cases} y_1 = H_{\text{conv}}^{3\times3}(x_1), \\ y_2 = H_{\text{conv}}^{3\times3}(x_2 + y_1), \\ y_3 = H_{\text{conv}}^{3\times3}(x_3 + y_2), \\ y_4 = H_{\text{conv}}^{3\times3}(x_4 + y_3) \end{cases}$$
(2)

 $y = CA(H_{aarv}^{[x]}(H_{aat}(y_1, y_2, y_3, y_4))) + x(3)$ 式中,  $y_2$  输出特征包含经过一个卷积层的感 受野和两个卷积堆叠后的感受野。同理,  $y_3$ 和  $y_4$ 也包含了多种尺度的感受野特征。通过这 种路径相连的方式,采用较少卷积层即可获取 丰富的特征信息。

#### 1.3 DFFM 模块

在传统的 U-Net 结构中,跳跃连接通常是 将编码结构中的特征直接传递到相应尺度的解 码过程中。这样不利于解码部分充分利用编码 过程中产生的多尺度特征,而上文提到多尺度 特征对恢复细节特征又起着重要的作用。因 此,本文提出在跳跃连接处设计一种特征融合 模块,充分利用不同层的特征信息,同时也能 够使网络各层之间的特征流通更灵活。

不同于直接对多尺度特征进行拼接融合, 本文设计了一种可自适应赋予特征不同权重的



红 外



动态融合结构(见图 3)。首先对每个编码阶段 的输出分别进行上采样或者下采样操作,并通 过卷积来调整通道数,分别得到  $x_1$ 、 $x_2$ 和  $x_3$ 。 然后对其进行融合,得到 x,从而提升模型的 特征表征能力。接着对 x进行全局平均池化操 作,得到全局压缩变量,再采用两个 1×1卷 积完成通道维数的缩放,紧接着通过 Softmax 函数得到三路特征的融合权重。最后对注意力 加权后的特征进行拼接融合,再经过 1×1卷 积完成通道降维,得到输出的融合特征。

#### 1.4 损失函数

基于本文算法采用了由粗到精的结构,整 个模型分成三个阶段,且每个阶段均能输出复 原图像。因此,损失函数同样采用多尺度策 略,以深入监督每个阶段训练。该损失函数由 多尺度内容损失和多尺度高频信息损失组成。 假设 *S*<sub>k</sub>(*k*=1,2,3)表示输出的多尺度复原图 像,*R*<sub>k</sub>(*k*=1,2,3)表示与之对应的真实清晰 图像。

#### 1.4.1 多尺度内容损失

使用不同尺度的真实清晰图与模型复原图 之间的 L1 距离作为多尺度内容损失,即

$$L_{\text{content}} = \sum_{k=1}^{3} \|R_k - S_k\|_1$$
(4)

L1 距离对较大误差值不会过度惩罚,有利于保留图像的边缘特征等。

## 1.4.2 多尺度高频信息损失

由于模糊图像主要丢失高频特征信息,因 此尽可能减少高频维度的损失也十分重要。本 文采用快速傅里叶变换(Fast Fourier Transform, FFT)来计算真实图像与清晰图像之间 的高频信息损失,即

$$L_{fft} = \sum_{k=1}^{3} \|FFT(R_k) - FFT(S_k)\|_1 \quad (5)$$

1.4.3 总损失

本文的总损失函数为

$$L = L_{content} + \lambda_1 L_{fft} \tag{6}$$

式中,λ,表示损失权重,取0.1。

## 2 实验结果与分析

#### 2.1 实验细节

本文在 GoPro<sup>[3]</sup>训练集上训练模型,在 GoPro测试集和 RealBlur<sup>[13]</sup>测试集上进行推理 测试。其中,GoPro 数据集包含 2103 对用于 训练的图像以及 1111 对用于测试的图像。该 数据集通常用于物体运动去模糊实验。每对图 像由真实图像与相应的模糊图像组成,均由 GoPro 相机拍摄而成。同时,为了验证模型的 泛化性能,直接将在 GoPro 数据集上训练得到 的模型放到 RealBlur 数据集上测试。该数据集 根据相机处理前后分成 RealBlur-R 子集和 RealBlur-J 子集。这两个子集共享相同的图像内 容。该泛化实验主要使用 RealBlur-J 子集的 980 对图片作为测试集。

该模型的训练与测试均在 Pytorch 框架下 实现,采用 TITAN RTX GPU 显卡加速训练。 每个多尺度残差注意力组合体包含 8 个 MSRM。在图像数据预处理部分,首先将图像 随机裁剪成 256×256 大小,然后随机进行水 平翻转来丰富训练集,以达到提升模型泛化性 能的目的。模型采用 GoPro 数据集进行训练, 训练周期为 3000,批次大小设置为 4。初始学 习率为 10<sup>-4</sup>,采用余弦退火策略<sup>[14]</sup>后学习率 稳定降到 10-6。

#### 2.2 定量分析

为了验证算法的有效性,采用峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)以及结构 相似性(Structural Similarity, SSIM)作为评价 指标,对本文算法与当前主流去模糊算 法<sup>[3,6-11]</sup>所复原的图像进行质量评估。表1列 出了各算法在 GoPro 数据集和 RealBlur 数据集 上的指标结果。由于以上主流算法均使用 Go-Pro 数据集训练模型,因此直接使用算法作者 们公开的源代码对 GoPro 数据集和部分 Real-Blur 数据集进行测试。由表1可知,本文算法 取得了不错的评估结果。在 GoPro 数据集上采 用本文算法时, PSNR 和 SSIM 分别为 31.95 dB和0.954,与DMPHN算法相比分别提升了 0.75 dB和 0.014。在 RealBlur 数据集上各算 法指标差异较小。与 DeblurGAN-V2 相比,本 文算法的 PSNR 和 SSIM 分别提升了 0.03 dB 和 0.005。根据以上两个数据集上的测试结果 可知,本文算法较之前的主流算法有较大的提 升,并且模型泛化能力也有所提高。

#### 2.3 定性分析

为了直观评估算法的去模糊能力,本文从 GoPro数据集和 RealBlur 数据集中选取部分场 景,并与其它算法效果图进行细节对比。图 4 和图 5 为效果对比图,放大细节依次为模糊图 像、真实清晰图像、DeblurGAN-v2 效果图、 DMPHN效果图以及本文算法的去模糊效果图。 相比之下,本文算法的复原图细节更丰富,物 体轮廓更清晰,与真实清晰图像更接近。对 于图 4 的第一个场景图,DeblurGAN-v2 和 DMPHN 算法还原的人像仍有不同程度的模 糊痕迹,且路面边缘不清晰。对于图 4 的第 二个场景图,本文算法能够更好地复原大楼 字体和汽车后视镜等细节纹理。对于图 4 的 第三个场景,本文算法在消除模糊的同时更 好地处理了局部细节。

同样地,在 GoPro 数据集上训练后直接测 试 RealBlur 数据集图片,从而验证模型的泛化 性能。如图 5 所示,DMPHN 和 DeblurGAN-v2 效果图能在一定程度上改善图像质量,但是效 果不及本文算法。对于图 5 的第一个场景图, 在本文算法效果图中能够观察到胡萝卜的边 缘轮廓和清晰的字体。对于图 5 的第三个场 景图,本文算法还原了绘画纹理信息,图像 视觉效果更佳。综上所述,本文算法能复原 更多的细节信息,有效改善了图像质量,且 比其它算法具有更强的泛化能力。

#### 2.4 消融实验

本文设置消融实验以证明 MSRM 和 DFFM 的有效性。本文设计不同的模块组合方 案,采用多尺度内容损失和多尺度高频信息损 失相结合的损失函数进行约束训练。每个组合 方案均保持相同的训练参数环境,并且同样采 用 PSNR 和 SSIM 这两种评价指标来评估各个 模块性能。(1)Baseline:采用多输入多输出的 U-Net 网络,编码解码部分采用残差块进行特 征提取。(2)Baseline+DFFM:在上述 Baseline

算法	GoPro 数据集		RealBlur 数据集	
	PSNR	SSIM	PSNR	SSIM
Nah S 等人 <sup>[3]</sup>	29.08 dB	0.914	27.87 dB	0.827
Zhang J W 等人 <sup>[7]</sup>	29.19 dB	0.931	27.80 dB	0.847
DeblurGAN-v2 <sup>[6]</sup>	29.55 dB	0.934	28.70 dB	0.866
$\mathrm{SRN}^{[8]}$	30.26 dB	0.934	28.56  dB	0.867
DBGAN <sup>[11]</sup>	31.10 dB	0.942	24.93dB	0.745
$MT-RNN^{[10]}$	31.15 dB	0.945	28.44 dB	0.862
DMPHN <sup>[9]</sup>	31.20 dB	0.940	28.42 dB	0.860
本文算法	31.95 dB	0.954	28.73 dB	0.871

表1 各算法在 GoPro 数据集和 RealBlur 数据集上的指标结果



图 4 GoPro 数据集上各算法的去模糊效果对比



图 5 RealBlur 数据集上各算法的去模糊效果对比

的跳跃连接处采用 DFFM 模块。(3)Baseline+ DFFM+MSRM:本文算法在结构(2)的基础 上采用 MSRM 模块代替残差块进行特征提取。

表 2 列出了不同模块组合结构的指标对比数据。可以看出,各个模块对算法性能均有不同程度的影响。其中,MSRM 模块影响最大,其 PSNR 和 SSIM 指标分别能提升 0.17 dB 和 0.003,证明多尺度特征能够有效保留更多特征信息,弥补下采样带来的细节损失。其次是DFFM 模块,其 PSNR 和 SSIM 指标分别提高了 0.12 dB 和 0.003。结果表明,以上模块均能增强模型的去模糊性能,有利于复原细节信息更丰富的图像。

表 2 不同模块组合结构的指标对比

Baseline	DFFM	MSRM	PSNR	SSIM
$\checkmark$	—	—	31.66 dB	0.948
$\checkmark$	$\checkmark$	—	31.78 dB	0.951
$\checkmark$	$\checkmark$	$\checkmark$	31.95 dB	0.954

## 3 结束语

针对目前基于 U-Net 结构的去模糊算法存 在的问题,本文提出一种基于多尺度优化和动 态特征融合的图像去模糊方法。该算法在 U-Net 基础上进行改进,并采用多输入多输出的 结构模式。首先针对细节损失的问题,提出一 种精简且有效的 MSRM 模块,通过少量卷积 层和路径相连的方式来获取不同尺度的感受 野,以提取更精细的图像信息。其次,设计一 个多输入的 DFFM 模块来代替传统的跳跃连 接结构,采用注意力加权的方式选择性融合不 同阶段的编码特征,将有利的特征信息传递到 解码部分。该算法采用多尺度内容损失和多尺 度高频信息损失进行约束训练。实验结果表 明,这种方法能有效改善图像质量,保留更多 图像细节,具备更强的去模糊能力。下一步将 继续优化算法,并结合视频相邻帧具有丰富信 息的特性,开展视频去模糊研究。

### 参考文献

[1] Fergus R, Singh B, Hertzmann A, et al. Remo-

INFRARED (MONTHLY)/VOL.44, NO.4, APR 2023

ving camera shake from a single photograph [C]. Boston: SIGGRAPH 2006 Conference, 2006.

- [2] Sun J, Cao W F, Xu Z B, et al. Learning a convolutional neural network for non-uniform motion blur removal [C]. Boston: 28th IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [3] Nah S, Kim T H, Lee K M. Deep multi-scale convolutional neural network for dynamic scene deblurring [C]. Honolulu: 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [4] Cho S J, Ji S W, Hong J P, et al. Rethinking coarse-to-fine approach in single image deblurring
  [C]. Montreal: 2021 IEEE/CVF International Conference on Computer Vision, 2021.
- [5] Kupyn O, Volodymyr B, Mykola M, et al. Deblurgan: Blind motion deblurring using conditional adversarial networks [C]. Salt Lake City: 2018 IEEE Conference on Computer Vision and Pattern, 2018.
- [6] Kupyn O, Tetiana M, Junru W, et al. Deblurganv2: Deblurring (orders-of-magnitude) faster and better [C]. Seoul: 2019 International Conference on Computer Vision, 2019.
- [7] Zhang J W, Pan J S, Jimmy S J, et al. Dynamic scene deblurring using spatially variant recurrent neural networks [C]. Salt Lake City: 2018 IEEE Conference on Computer Vision and Pattern, 2018.
- [8] Tao X, Gao H Y, Shen X Y, et al. Scale-recurrent network for deep image deblurring [C]. Salt Lake City: 2018 IEEE Conference on Computer Vision and Pattern, 2018.
- [9] Zhang H G, Dai Y C, Li H D, et al. Deep stacked hierarchical multi-patch network for image deblurring [C]. Long Beach: 2019 IEEE Conference on Computer Vision and Pattern, 2019.
- [10] Park D W, Kang D U, Kim J, et al. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training [C]. Glasgow: 16th European Conference on Computer Vision, 2020.

- [11] Zhang K H, Luo W H, Zhong Y R, et al. Deblurring by realistic blurring [C]. Seattle: 2020 IEEE Conference on Computer Vision and Pattern, 2020.
- [12] Zou W, Jiang M, Zhang Y, et al. SDWNet: A straight dilated network with wavelet transformation for image deblurring [C]. Montreal: 2021 International Conference on Computer Vision, 2021.
- [13] Rim J, Lee H, Won J, et al. Real-world blur dataset for learning and benchmarking deblurring

algorithms [C]. Glasgow: 16th European Conference on Computer Vision, 2020.

- [14] Ilya L, Frank H. Sgdr: Stochastic gradient descent with warm restarts [C]. Puerto Rico: 4th International Conference on Learning Representations, 2016.
- [15] Gao S H, Cheng M M, Zhao K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652–662.