基于 RNN 与级联损失函数的图像超分辨率研究

曾 强[1] 刘晓群[1] 郝 娟[1]
ZENG Qiang LIU Xiaoqun HAO Juan

摘 要 为了获取具有更丰富细节和更清晰纹理的超分辨率图像，提出一种融合循环神经网络（RNN）与级联损失函数的超分辨率重建网络（RLNN）。相较于传统的 RNN 方法，所提出的网络架构主要实现了两大创新。首先，将 RNN 的每一次迭代与级联损失函数紧密集成，通过这种方式，不仅提高了网络在超分辨率重建过程中的精度，还增强了其对复杂图像特征的捕捉能力。其次，设计了一种新颖的隐藏模块（HM），结合了空间 - 通道注意力机制与局部密集跳跃连接网络，有效提升了网络在特征提取与重用方面的性能。为了进一步优化网络的学习过程，还引入了一种课程学习策略，使网络能够逐步适应并处理更加复杂的任务。实验结果证明，所提出的 RLNN 网络在图像超分辨率任务上表现出极好的重建效果。

关键词 卷积神经网络；超分辨率重建；循环神经网络；级联损失函数；密集跳跃连接；空间 - 通道注意力机制；课程学习策略

doi：10.3969/j.issn.1672-9528.2024.10.018

0 引言

图像超分辨技术是指将低分辨率的图像重建为高分辨率图像的方法。其中，单幅图像超分辨率重建已被广泛运用于医学图像 [1-3]、人脸识别 [4-5]、微信遥感 [6] 等领域。

图像超分辨率重建的主要目的是提高图像的像素分辨率和空间分辨率，从而恢复图像中缺失的信息。目前，单幅图像超分辨率重建（single image super-resolution，SISR）的传统方法主要分为三类，分别是：基于插值的方法 [7]、基于重建的方法和基于浅层学习的方法。基于插值的方法是根据 LR 图像像素的排列关系就近选择性质相同的像素值进行重建。常见的插值算法如双三次插值法（Bicubic），虽然计算简单，但是损失较多，效果不好。基于重建的方法首先要构建退化模型，通过退化模型逆推重建图像。基于浅层学习的方法是从大量的 LR-HR 图像样本中学习图像对之间的联系，根据学到的变换对低分辨率图像进行重建。

近几年来，卷积网络常被用来进行超分辨率重建 [8-9]。2014 年，Dong 等人 [10] 提出基于卷积网络的超分辨率重建方法（learning a deep convolutional network for image super-resolution，SRCNN），其输入是经双三次插值放大到目标尺寸的 LR 图像，因此训练速度较慢。为加快 SRCNN 的训练速度，Dong 等人 [11] 提出了对 SRCNN 的改进模型，即快速超分辨率卷积神经网络（accelerating the super-resolution convolutional neural network，FSRCNN），其采用反卷积作为上采样操作，并将其放在网络的末端，从而降低了计算的复杂度，提高了网络重建质量。Kim 等人 [12] 构建了一个深度为 20 层的卷积网络（accurate image super-resolution using very deep convolutional networks，VDSR），证明了加深网络与残差网络对超分辨率重建任务有效。然而，随着网络深度的进一步增加，他们观察到重建效果开始出现下降趋势，这表明单纯的网络加深并不总是能带来性能的提升，反而可能由于优化过程中的梯度消失或爆炸等问题导致性能下降。为了克服这一挑战，开始研究采用跳跃连接来克服优化的困难。Zhang 等人 [13] 将稠密跳跃连接网络和残差网络相结合，提出了稠密残差网络（residual dense network for image super-resolution，RDN），该模型充分利用了局部 / 全局残差与密集跳跃连接，但因参数量过大，无法实际应用。Lim 等人 [14] 提出增强残差（enhanced deep residual networks for single image super-resolution，EDSR）网络结构实现，也采用了残差跳跃连接。此外，Zhang 等人 [15] 提出的 RCAN（image super-resolution using very deep residual channel attention networks）卷积神经网络开始在图像超分辨率领域中引入通道注意力机制，并用残差嵌套结构加深网络。Zhao 等人利用像素注意力机制构建高效的图像超分辨率网络 PAN（efficient image super-resolutionusing pixelattention）提升了重建性能。随着网络深度的增长，参数的数量也会增加。大容量网络将占用巨大的存储资源，并存在过拟合问题。为了减少网络参数，Kim 等

1. 河北建筑工程学院 河北张家口 075000

人提出了 DRCN（deeply-recursive convolutional network for image super-resolution）首次引入了 RNN，在不增加额外参数的情况下加深网络层数。Tai 等人提出了 DRRN（image super-resolution via deep recursive residual network）在 DRCN 的基础上引入了局部残差和残差单元的递归学习，实现了更好的重建效果。

在本文中，结合 RNN 网络与级联损失函数提出一个新的模型。该模型本质上就是具有一个特殊隐藏块（HM）的 RNN 网络，主要改进有：（1）将 RNN 网络中的隐藏状态展开，在 HM 模块每次迭代时捆绑损失（迫使网络在每次迭代中重建 SR 图像，从而允许隐藏状态携带高级信息）；（2）HM 模块融合了具有密集跳跃属性的上采样块与下采样块，并结合了注意力机制。这种结构使得网络能够循环多次进行上下采样操作，并通过密集的跳跃连接更好地提取图像特征。同时，注意力机制的引入使得网络能够更加关注含有高频信息的通道与区域，进一步提升了图像重建的质量；（3）本文引入了一种课程学习策略，对于每个 LR 图像，基于恢复难度从容易到难排列其连续迭代的目标 HR 图像。这种课程学习策略很好地帮助本文提出的模型处理复杂的退化模型。实验结果证明了本文提出的模型相对于其他先进方法的优越性。

1 网络结构

1.1 递归损失网络

本文提出循环损失网络（recurrent lossneural network，RLNN）结构如图 1 所示，RLNN 可以展开为 T 次迭代，每次迭代均按时间顺序从 1 至 T 排列，且本文将每次迭代均与损失函数进行紧密结合。RLNN 的网络结构包含三个部分：浅层特征提取模块（SFEM）、隐藏模块（HM）和重建模块（RM），每个块的权重都是跨时间共享的。

网络的输入为低分辨率图像 I_{LR}，经过 SFEM 模块的处理后，输出特征 F_{in}。在第 t 次迭代时，HM 模块的输入为 $F^{t-1}{}_{out}$ 与 F_{in}，输出为 $F^{t}{}_{out}$，随后 RM 模块根据 $F^{t}{}_{out}$ 生成超分辨率图像 $I^{t}{}_{SR}$。此外，本文采用 Conv(s, n) 和 Deconv(s, n) 分别表示卷积层与反卷积层。其中，s 代表卷积核大小，n 代表输出通道数。

在浅层特征提取模块（SFEM）设计中，首层通过应用不同尺寸的卷积核（1、3 和 5）实现多尺度特征提取，融合后形成综合特征表示。而后，通过 Conv(1, m) 卷积层压缩精炼特征。本文为 SFEM 特征提取块提供 LR 输入 I_{LR}，从中获得包含 LR 图像的信息的浅层特征 F_{in}：

$$F_{in}=f_{SFEM}(I_{LR}) \tag{1}$$

式中：f_{SFEM} 表示 SFEM 模块的操作，而 F_{in} 将作为 HM 模块的初始输入，为后续的迭代过程提供浅层特征信息。此外，在后续迭代时，HM 模块还将接收前次迭代 HM 模块的输出 $F^{t-1}{}_{out}$，而后 F_{in} 与 $F^{t-1}{}_{out}$ 共同作为 HM 模块的输入。HM 模块的数学公式可表示为：

$$F^{t}{}_{out}=f_{HM}(F^{t-1}{}_{out}, F_{in}) \tag{2}$$

式中：f_{HM} 表示 HM 模块进行的操作，HM 模块作为网络中最重要的模块。

重建模块（RM）作为网络的最后一环，接收到特征 $F^{t}{}_{out}$ 后，使用 Decov(k, m) 将 LR 特征升级为 HR 特征，并使用 Conv(3, 3) 再次提取与精炼特征生成高清图像 $I^{t}{}_{Res}$。RM 模块的数学公式可表示为：

$$I^{t}{}_{Res}=f_{RM}(F^{t}{}_{out}) \tag{3}$$

式中：f_{RM} 表示重建块的操作。第 t 次迭代时输出的高清图像 $I^{t}{}_{SR}$ 可表示为：

$$I^{t}{}_{SR}=I^{t}{}_{Res}+f_{UP}(I_{LR}) \tag{4}$$

式中：f_{UP} 表示上采样操作。上采样方式这里选择双线性上采样。经过共 T 次迭代，将得到总共 T 个 SR 图像 $(I^{1}{}_{SR}, I^{2}{}_{SR}, \cdots, I^{T}{}_{SR})$。

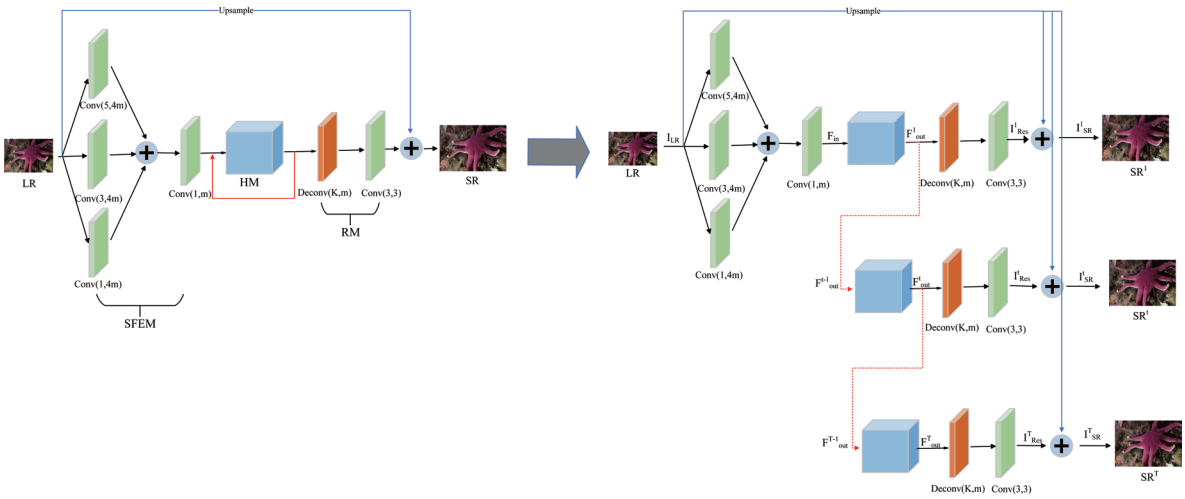

图 1 recurrent loss neural network（RLNN）

1.2 隐藏模块（HM）

如图 2 所示，第 t 次迭代时 HM 模块接收前次迭代信息 F^{t-1}_{out} 以校正浅层特征 F_{in}，然后将更加强大的高级表示 F^{t}_{out} 传递给下一次迭代与重建模块，HM 模块依次包含通道 - 空间注意力块（CSAM）、Conv(1, m)、G 个映射组（MF）。映射组之间通过密集的跳跃连接实现了高效的信息交互与传递。这些映射组融合了上采样与下采样层，形成了层次化的特征转换机制。每一个映射组均具备将高分辨率（HR）特征映射至低分辨率（LR）特征的能力，从而实现了多尺度特征的提取与整合。

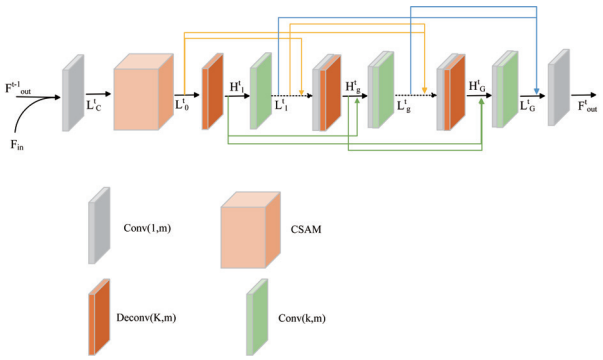

图 2 HM 模块

HM 模块的开始阶段 F^{t-1}_{out} 与 F_{in} 先通过 Conv(1, m) 连接并压缩以通过反馈信息 F^{t-1}_{out} 来细化浅层特征 F_{in}，产生的细化特征 L^t_c 可表示为：

$$L^t_c = C_0([F^{t-1}_{\text{out}}, F_{\text{in}}]) \tag{5}$$

式中：C_0 表示压缩操作，$[F^{t-1}_{\text{out}}, F_{\text{in}}]$ 表示将 F^{t-1}_{out} 与 F_{in} 进行通道的拼接。而后，L^t_c 通过通道 - 空间注意力机制（CSAM）产生更具备高频信息的 L^t_0 特征，产生的高频特征 L^t_0 可表示为：

$$L^t_0 = \text{CSAM}(L^t_c) \tag{6}$$

CSAM 表示通过三维卷积实现的通道 - 空间注意力机制，设 H^t_g 和 L^t_g 是 FB 中第 g 个投影组在第 t 次迭代时给出的 HR 和 LR 特征图。H^t_g 可以通过以下方式获得：

$$H^t_g = C^{\uparrow}_g([L^t_0, L^t_1, \cdots, L^t_{g-1}]) \tag{7}$$

式中：$C^{\uparrow}_g$ 是指在第 g 个投影组使用 Decov(k, m) 的上采样操作。相应地，L^t_g 可以表示为：

$$L^t_g = C^{\downarrow}_g([H^t_1, H^t_2, \cdots, H^t_g]) \tag{8}$$

式中：$C^{\downarrow}_g$ 是指在第 g 个投影组处使用 Conv(k, m) 的下采样操作。此外，除了第一个映射组，其他都在每次上 / 下采样前加入了 Conv(1, m) 改变通道数以减少计算量。

为利用每个映射组的有效信息，并充分利用 LR 特征 F^t_c，本文使用密集跳跃连接对映射图生成的 LR 特征进行了特征融合，HM 模块的输出 F^t_{out} 可表示为：

$$F^t_{\text{out}} = C_{\text{FF}}([L^t_1, L^t_2, \cdots, L^t_G]) \tag{9}$$

式中：C_{FF} 表示 Conv(1, m)。

1.3 课程学习策略

本文选择 L1 损失函数来优化 RLNN 模型，放置 T 个目标图像 ($\text{SR}^1, \text{SR}^2, \cdots, \text{SR}^T$) 以适应本文网络的多个输出，($\text{SR}^1, \text{SR}^2, \cdots, \text{SR}^T$) 对于单个的退化模型是相同的，但对于复杂的退化模型，($\text{SR}^1, \text{SR}^2, \cdots, \text{SR}^T$) 是基于 T 次迭代的任务难度排序的，以适用课程学习策略。网络中的损失函数可以表示为：

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T} W^t ||\text{SR}^t - I^t_{\text{SR}}||_1 \tag{10}$$

式中：θ 代表的是网络的参数；W^t 是一个常数因子，代表第 t 次迭代时输出代表的价值，在本文中每次迭代的值都设置为 1，这代表每个输出都具备相同的价值。

1.4 网络细节

本文采用 PReLU 作为除最后一层以外所有卷积与反卷积的激活函数。此外，本文在 Conv(k, m) 和 Decov(k, m) 中为不同的比例因子设置各种 k，对于 x2 的比例因子，k 值被设置为 6，striding=2，padding=2。对于 x3 的比例因子，k 值被设置为 7，striding=3，padding=2。对于 x4 的比例因子，k 值被设置为 8，striding=4，padding=2。

2 实验结果

2.1 实验设置

实验数据与预处理：将 DIV2K 数据集中 800 张图像的训练集与 100 张图像的测试集共同作为本文的训练集。为充分地利用数据，本文采用了文献 [14] 处理训练图像的方法。测试集为五个标准数据集：Set5、Set14、B100、Urban100 和 Manga109。评价指标为 PSNR 与 SSIM，与其他方法一致，评价指标仅在 y 通道上计算。

退化模型：为与现有方法进行比较，本文将双三次下采样作为标准退化模型，用于从真实 HR 图像生成 LR 图像。

训练参数设置：本文最后训练的模型迭代数 T=6，G=6。此外，批处理数目为 16，输入通道数与输出通道数均为 3，初始学习率为 0.000 1，学习率每过 200 个周期乘以 0.5。使用 PyTorch2.20 框架训练本文网络，并在 NVIDIA 3090 上进行训练。

2.2 实验结果

针对标准退化模型，本文深入对比了所提出的 RLNN 与四种先进的图像超分辨率（SR）方法，即 SRCNN、VDSR、DRRN 以及 MemNet。为确保对比的公正性与准确性，表 1 中列出的其他方法的定量结果均基于其公开的源代码进行重新评估。实验结果表明，在多个标准测试集（包括 Set5、Set14、B100、Urban100 以及 Manga109）上，本文提出的 RLNN 展现出了显著的优势，其性能优于所有比较的 SR 方法。相较于这些方法，RLNN 在图像超分辨率任务中可以获得更好的结果。

表 1 使用标准退化模型的比例因子 x2、x3 和 x4 的平均 PSNR/SSIM 值

Dataset	Scale	Bicubic	SRCNN	VDSR	DRRN	MemNet	RLNN
Set5	x2	33.66/0.929 9	36.66/0.954 2	37.53/0.959 0	37.74/0.959 1	37.78/0.959 7	38.07/0.960 7
	x3	30.39/0.868 2	32.75/0.909 0	33.67/0.921 0	34.03/0.924 4	34.09/0.924 8	34.67/0.928 7
	x4	28.42/0.810 4	30.48/0.862 8	31.35/0.883 0	31.68/0.888 8	31.74/0.889 3	32.46/0.898 0
Set14	x2	30.24/0.868 8	32.45/0.006 7	33.05/0.913 0	33.23/0.913 6	33.28/0.914 2	33.80/0.919 3
	x3	27.55/0.774 2	29.30/0.821 5	29.78/0.832 0	29.96/0.834 9	30.00/0.835 0	30.50/0.846 0
	x4	26.00/0.702 7	27.50/0.751 3	28.02/0.768 0	28.21/0.772 1	28.26/0.772 3	28.81/0.7864
B100	x2	29.56/0.843 1	31.36/0.887 9	31.90/0.896 0	32.05/0.897 3	32.08/0.897 8	32.24/0.900 5
	x3	27.21/0.738 5	28.41/0.786 3	28.83/0.799 0	28.95/0.800 4	28.96/0.800 1	29.20/0.808 0
	x4	25.96/0.667 5	26.90/0.710 1	27.29/0.726 0	27.38/0.728 4	27.40/0.728 1	27.70/0.740 5
Urban100	x2	26.88/0.840 3	29.50/0.894 6	30.77/0.914 0	31.23/0.918 8	31.31/0.919 5	32.60/0.932 5
	x3	24.46/0.734 9	26.24/0.798 9	27.14/0.829 0	27.53/0.837 8	27.56/0.837 6	28.71/0.863 8
	x4	23.14/0.6577	24.52/0.722 1	25.18/0.754 0	25.44/0.763 8	25.50/0.763 0	26.56/0.801 0
Manga109	x2	30.30/0.933 9	35.60/0.966 3	37.22/0.975 0	37.60/0.973 6	37.72/0.974 0	39.04/0.977 5
	x3	26.95/0.855 6	30.48/0.911 7	32.01/0.934 0	32.42/0.935 9	32.51/0.936 9	34.14/0.948 0
	x4	24.89/0.786 6	27.58/0.855 5	28.83/0.887 0	29.18/0.891 4	29.42/0.894 2	31.10/0.915 5

3 结论

本文提出了一种新的图像 SR 网络，称为递归损失网络（RLNN），该网络的核心思想在于将循环神经网络（RNN）的隐藏状态进行展开，并与级联损失函数紧密结合，从而显著提升了重建网络的性能。网络中的隐藏块（HM）能够有效提取 LR-SR 特征并实现特征重用，这对于提升图像质量至关重要。此外，还引入了一种课程学习策略，使网络能够很好地适应更复杂的任务。通过实验验证，证明了本文提出的 RLNN 网络能够在图像超分辨率任务上表现出极好的重建效果。

参考文献：

[1] 柯舒婷 , 陈明惠 , 郑泽希 , 等 . 生成对抗网络对 OCT 视网膜图像的超分辨率重建 [J]. 中国激光 ,2022,49(15):90-98.

[2] 左艳 , 黄钢 , 聂生东 . 深度学习在医学影像智能处理中的应用与挑战 [J]. 中国图象图形学报 ,2021,26(2):305-315.

[3] 王一宁 , 赵青杉 , 秦品乐 , 等 . 基于轻量密集神经网络的医学图像超分辨率重建算法 [J]. 计算机应用 ,2022,42(8):2586-2592.

[4] 倪若婷 , 周莲英 . 基于卷积神经网络的人脸图像超分辨率重建方法 [J]. 计算机与数字工程 ,2022,50(1):195-200.

[5] 卢峰 , 周琳 , 蔡小辉 . 面向安防监控场景的低分辨率人脸识别算法研究 [J]. 计算机应用研究 ,2021,38(4):1230-1234.

[6] 耿铭昆 , 吴凡路 , 王栋 . 轻量化火星遥感影像超分辨率重建网络 [J]. 光学精密工程 ,2022,30(12):1487-1498.

[7]KEYS R.Cubic convolution interpolationfor digital imageprocessing[J].IEEE transactions on acoustics, speech, and signal processing,1981,29(6):1153-1160.

[8] 黄发文 , 唐欣 , 周斌 . 结合双注意力和结构相似度量的图像超分辨率重建网络 [J]. 液晶与显示 , 2022, 37(3): 367-375.

[9] 周乐 , 徐龙 , 刘孝艳 , 等 . 基于梯度感知的单幅图像超分辨 [J]. 液晶与显示 ,2022,37(10):1334-1344.

[10]DONG C, LOY C C, HE K M, et al.Learning a deep convolutional network for image super-resolution[C]// Proceedings of the 13th European Conference on Computer Vision. Berlin: Springer,2014:184-199.

[11]DONG C, LOY C C, TANG X O.Accelerating thesuper-resolution convolutionalneural network[C]//Proceedings of the 14th European Conference on Computer Vision.Berlin: Springer, 2016:391-407.

[12]KIM J, LEE J K, LEE K M.Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE,2016:1646-1654.

[13]ZHANG Y, TIAN Y, KONG Y, et al.Residualdense network for image super-resolution[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway: IEEE, 2018:2472-2481.

[14]LIM B, SON S, KIM H, et al.Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). Piscataway:IEEE,2017:1132-1140.

[15]ZHANG Y L, LI K P, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]// Proceedings of the 15th European Conference on Computer Vision. Berlin: Springer,2018:294-310.

【作者简介】

曾强（2000—），男，湖南娄底人，硕士研究生，研究方向：计算机技术。

刘晓群（1968—），男，河北张家口人，硕士，教授，硕士生导师，研究方向：计算机网络与信息安全。

郝娟（1989—），女，河北保定人，硕士，副教授，研究方向：图像处理。

（收稿日期：2024-04-01）