Opto-Electronic Engineering
光 电 工 程

Article

2021 年，第 48 卷，第 1 期

DOI: 10.12086/oee.2021.200045

基于通道注意力与迁移学习的红外图像超分辨率重建算法

孙　锐[1,2]，章　晗[1,2*]，程志康[1,2]，张旭东[1]

[1]合肥工业大学计算机与信息学院，安徽 合肥 230009；
[2]工业安全与应急技术安徽省重点实验室，安徽 合肥 230009

摘要：针对现有红外图像分辨率低、质量不高的问题，提出了基于通道注意力与迁移学习的红外图像超分辨率重建方法。该方法设计了一个深度卷积神经网络，融入通道注意力机制来增强网络的学习能力，并且使用残差学习方式来减轻梯度爆炸或消失问题，加速网络的收敛。考虑到高质量的红外图像难以采集、数目不足的情况，将网络的训练分成两步：第一步使用自然图像来预训练网络模型，第二步利用迁移学习的知识，用较少数量的高质量红外图像对预训练的模型参数进行迁移微调，使模型对红外图像的重建效果更优。最后，加入多尺度细节滤波器来提升红外重建图像的视觉效果。在 Set5、Set14 数据集以及红外图像上的实验表明，融入通道注意力机制和残差学习方法，均能提升超分辨率重建的效果，迁移微调能很好地解决红外样本数量不足的问题，而多尺度细节提升滤波则能提升重建图像的细节，增大信息量。

关键词：超分辨率；红外图像；卷积神经网络；注意力；迁移学习

中图分类号：TP391；TB866　　**文献标志码：**A

孙锐，章晗，程志康，等. 基于通道注意力与迁移学习的红外图像超分辨率重建算法[J]. 光电工程，2021，**48**(1): 200045
Sun R, Zhang H, Cheng Z K, *et al*. Super-resolution reconstruction of infrared image based on channel attention and transfer learning[J]. *Opto-Electron Eng*, 2021, **48**(1): 200045

Super-resolution reconstruction of infrared image based on channel attention and transfer learning

Sun Rui[1,2], Zhang Han[1,2*], Cheng Zhikang[1,2], Zhang Xudong[1]

[1]School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China;
[2]Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China

Abstract: A super-resolution reconstruction method of infrared images based on channel attention and transfer learning was proposed to solve the problems of low resolution and low quality of infrared images. In this method, a deep convolutional neural network is designed to enhance the learning ability of the network by introducing the channel attention mechanism, and the residual learning method is used to mitigate the problem of gradient explosion or disappearance and to accelerate the convergence of the network. Because high-quality infrared images are difficult to collect and insufficient in number, so this method is divided into two steps: the first step is to use natural images to pre-train the neural network model, and the second step is to use transfer learning knowledge to fine-tune

收稿日期：2020-02-11；　收到修改稿日期：2020-05-09
基金项目：国家自然科学基金面上项目(61471154，61876057)；中央高校基本科研业务费专项资金资助项目(JZ2018YYPY0287)
作者简介：孙锐(1976-)，男，博士，教授，主要从事计算机视觉的研究。E-mail：sunrui@hfut.edu.cn
通信作者：章晗(1995-)，男，硕士研究生，主要从事计算机视觉的研究。E-mail：1402198087@qq.com
版权所有©2021 中国科学院光电技术研究所

the pre-trained model's parameters with a small number of high-quality infrared images to make the model better in reconstructing the infrared image. Finally, a multi-scale detail boosting filter is added to improve the visual effect of the reconstructed infrared image. Experiments on Set5 and Set14 datasets as well as infrared images show that the deepening network depth and introducing channel attention mechanism can improve the effect of super-resolution reconstruction, transfer learning can well solve the problem of insufficient number of infrared image samples, and multi-scale detail boosting filter can improve the details and increase the amount of information of the reconstruction image.

Keywords: super-resolution; infrared image; convolutional neural network; attention; transfer learning

1 引 言

自然界中温度在绝对零度(-273 ℃)以上的物体，都会因为自身的分子热运动而不断向外发射红外辐射，温度越高的物体发射的红外辐射越强。红外图像就是利用物体发射的红外辐射来完成成像的。近些年来，红外成像技术发展迅速，已经在军事侦察、安防监控、医疗成像等方面得到了日益普遍的运用。然而在红外图像成像或传输过程中，受到环境和设备等诸多因素的影响，经常会出现红外图像分辨率偏低的情况，从而使红外图像所包含的信息量大打折扣，制约着红外图像的应用价值。所以怎样获取高分辨率和高信息量的红外图像成为人们迫切需要解决的一个问题。通过改进红外成像设备来提高红外图像的质量是最直接的方法，但是红外成像设备设计繁杂，成本高，而且还受环境和诸多物理技术的限制。现在在自然图像领域兴起了利用计算机通过图像处理来提升图像的分辨率的方法，这种由一幅或多幅低分辨率图像来恢复高分辨率图像或图像序列的技术，称为图像的超分辨率技术[1]。这项技术相对于物理方法实现容易，大大降低了成本，而且它不仅适用于自然图像的超分辨率重建，也同样适用于红外图像或其他类型的图像的超分辨率重建。

超分辨率方法发展至今，主要分为三种类型：插值法[2]、基于重建的方法[3]和基于学习的方法[4-5]。插值法的重建速度很快，原理简单，主要是利用邻近像素点的像素值按照一定规则来生成新坐标点的像素值，其中常用的插值法是最近邻插值(nearest neighbor interpolation)和双三次插值(Bicubic)，但是这种方法处理后的图像清晰度不高，细节难以重现，而且常有锯齿出现。基于重建的方法是先提取出多幅低分辨率图像中的高频信息，再将这些高频信息按照一定规则组合起来指导重建高分辨率的图像，其中常用的基于重建的方法有凸集投影法[6]和迭代反向投影法[7]等，这种方法需要配准多幅图像，过程相对繁杂，而且处理后的图像细节不清晰，效果一般。基于学习的方法从一出现就成为了人们关注的热点，该方法分别提取出高、低分辨率图像的特征，再利用计算机去学习双方特征之间的相互对应关系，进而指导图像的超分辨率重建。Chang 等提出的邻域嵌入法[8]，是学习高、低分辨率图像块之间的几何相似度去获取它们之间的对应关系。Yang 等提出的基于稀疏编码的方法[9]，是利用机器学习方法建立高、低分辨率图像块词典间的对应关系，通过建立高、低分辨率图像间的联系来指导重建。近些年深度学习技术的发展突飞猛进，作为基于学习的方法中的后起之秀——基于深度学习的超分辨率方法开始出现。Dong 等率先提出了基于超分辨率卷积神经网络(super-resolution convolutional neural network，SRCNN)[10]，这是深度学习在超分辨率领域的第一次应用，该网络用了三个卷积层学习高、低分辨率图像的特征之间对应关系来指导重建，此方法比传统的方法在效果上有明显提升。在同一年，Dong 等又对 SRCNN 算法进行了优化，为了缩短训练时间，他们提出了快速超分辨率卷积神经网络(fast super-resolution convolutional neural network，FSRCNN)算法[11]，它省去了用 Bicubic 放大尺寸这一步，直接将低分辨率图像输入网络，再将一个反卷积层接在网络的后端以实现对应比例的放大，最终得到了高分辨率的图像，Shi 等提出了基于亚像素卷积的神经网络超分辨率重建算法(efficient sub-pixel convolutional neural network for image super-resolution，ESPCN)[12]，也是省去 Bicubic 这一步，直接提取低分辨率图像的特征，网络的后端用亚像素卷积层进行上采样，重建出对应尺寸的高分辨率图像，减少了计算复杂度。Ledig 等提出了基于生成对抗网络的超分辨率重建算法(super-resolution generative adversarial network，SRGAN)[13]，该算法将 GAN 网络(generative adversarial network)引入到超分辨率上来，而且用两个损失函数来引导网络训练，获

得了效果更逼真的重建图像。

如果直接将这些卷积神经网络应用在红外图像领域，则会存在一些问题：SRCNN、FSRCNN 和 ESPCN 方法的网络卷积层数较少，网络深度不够，学习的特征会比较单一，忽略了图像特征之间的相互关系，难以提取红外图像的深层次信息，而 SRGAN 方法可能会生成与原图在某些细节上差别很大的超分辨率图像，这不利于红外图像在军事、医疗和监控方面的应用；此外还有一个需要克服的问题就是在现实生活中很难收集到数量充足的高质量红外图像，而大量不同场景和目标的图像作为训练样本是通常深度学习方法所必需的，只有这样才能有较好的训练结果，所以直接将较少数量的红外图像作为训练数据集来实现深度学习方法往往达不到所需要的效果。

本文为了解决这些问题，提出了基于通道注意力与迁移学习的红外图像超分辨率重建方法。该方法首先设计了一个深度卷积神经网络，该网络融入通道注意力机制来学习特征空间通道之间的相关性，增强网络的学习能力，并且使用残差学习方式来减轻梯度爆炸或消失问题，同时也加速网络的收敛。考虑到高质量的红外图像难以采集，数目不足，然后将网络的训练分成两步：第一步使用自然图像来预训练出一个自然图像的超分辨率模型，第二步利用迁移学习[14]的知识，用数量较少的高质量红外图像将预训练获取到的模型参数进行快速迁移微调训练[15-16]，提升模型对红外图像的重建效果，从而得到红外图像的超分辨率模型。最后加入多尺度细节提升(multi-scale detail boosting，MSDB)[17]模块来提升红外重建图像的细节和视觉效果，增大信息量。

2 经典超分辨率重建网络

SRCNN 算法使用 Bicubic 将低分辨率图像扩大到目标尺寸，然后利用深度学习网络完成特征的提取以及非线性映射，最终重建出高分辨率图像，图 1(a)是 SRCNN 网络的结构图，其中 conv 代表卷积层。FSRCNN 则省去了用 Bicubic 放大尺寸这一步，直接将低分辨率图像输入网络，降低了计算量，再将一个反卷积层接在网络的后端以实现对应比例的放大，最终得到了高分辨率的图像，FSRCNN 网络结构图如图 1(b)所示。相较于 SRCNN，FSRCNN 选择了更小的卷积核和更深的网络，进一步降低了计算量。ESPCN 也是不经 Bicubic 步骤直接将低分辨率图像直接输入网络，在网络的最后用亚像素卷积层进行像素的重排列来重建高分辨率图像，减少了计算复杂度，提升了计算速度，但是网络深度仍然不够，ESPCN 网络结构图如图 1(c)所示。

3 SESR 方法

3.1 本文算法流程

一般基于卷积神经网络的超分辨率算法通常有 4 个步骤：1) 处理数据集；2) 构建网络模型；3) 用数据集训练网络模型；4) 生成网络参数；5) 用训练好的网络模型重建高分辨率图像[5]。而本文为了完成红外图像的超分辨率重建，在一般步骤基础上增加了迁移学习步骤，来克服红外样本数量不足带来的网络学习能力不足的问题。本文整体流程分为训练自然图像超分辨率模型和迁移学习两个步骤。

第一步的任务是通过深度学习获取自然图像超分辨率模型。考虑到人眼对亮度的视觉敏感性强于对色彩的视觉敏感性，所以为了降低计算复杂程度和减少训练量，本文对自然图像取亮度通道图像来学习重建。先将低分辨率自然图像取亮度通道图像送入深度学习网络进行训练，得到自然图像超分辨率模型；再用该模型去重建高分辨率图像；最后将重建后的图像与色

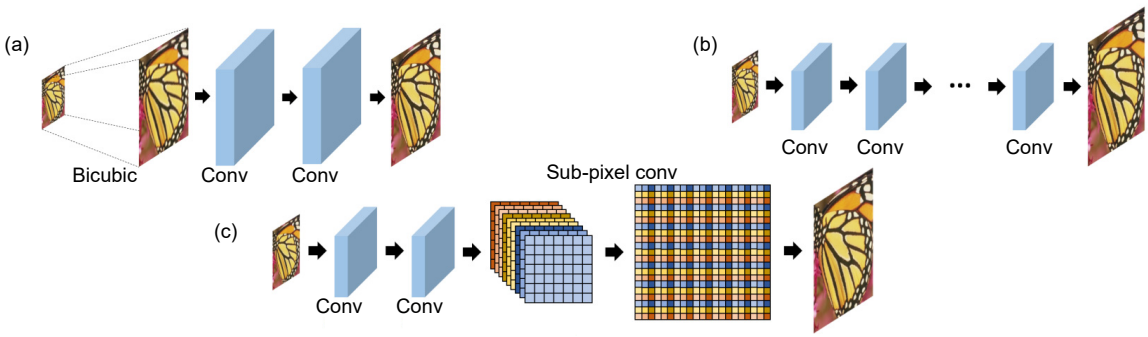

图 1 经典超分辨率重建网络结构图。(a) SRCNN；(b) FSRCNN；(c) ESPCN

Fig. 1 The structure diagram of classic super-resolution reconstruction network. (a) SRCNN; (b) FSRCNN; (c) ESPCN

度空间图像结合，生成完整的彩色自然图像的重建图像。

第二步的任务是用迁移学习来获取红外图像的超分辨率模型。与第一阶段不同的是，因为红外图像自身是单通道图像，所以无需对通道进行操作，直接将完整的红外图像作为训练样本送到深度学习网络中，再用迁移学习的方法，将自然图像的超分辨率重建模型的参数作为红外图像超分辨率模型的初始参数，用红外训练样本对这些参数进行微调训练，从而得到针对红外图像的超分辨率重建模型。最后使用 MSDB 滤波来提升图像的细节。本文的超分辨率重建流程如图 2 所示，其中 Y 是取亮度通道操作。

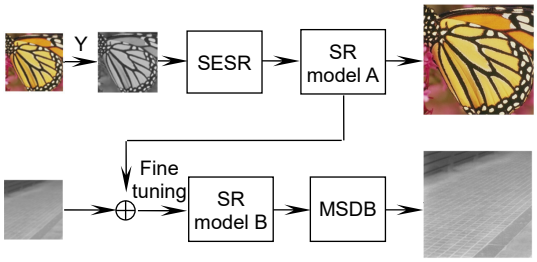

图 2 算法流程图

Fig. 2 The flow chart of the proposed method

3.2 SESR 网络

本文设计的图像超分辨率重建网络 SESR 结构如图 3 所示。在网络的前段，先串联三个卷积层，完成特征提取和初步的非线性映射步骤。在网络的后段，受 ImageNet 加深网络可以提升识别准确率的启发，串联 6 个相同的卷积层、PReLU 激活层与 SE block 三者的组合模块来加深网络深度，避免网络的过拟合，增加网络的感受野，实现完整的非线性映射操作。最后一层是反卷积层，通过反卷积来完成上采样。反卷积层的输出与前段网络的输出相加，最终获得超分辨率图像。

在网络的细节设计方面，考虑到 ReLU 函数的负半轴始终为 0，所以在神经网络训练时可能会出现神经元“坏死”现象。而 PReLU 函数的负半轴斜率是一个可以学习的参数，使负半轴的信息不会丢失，所以本文将选择 PReLU 作为网络的激活函数，这样可以避免神经元“坏死”的现象，起到一定程度的正则作用，也能增加网络的泛化能力；较深的网络可能会出现难以收敛或过拟合的情况，使网络训练的效果变差[17]，所以本文加入了残差网络结构以加速训练的收敛，减轻梯度爆炸或消失问题。表 1 为各卷积层参数的设置。

表 1 各卷积层参数设置

Table 1 The parameter setting of each convolution layer

Name	Number	Size
Conv1	64	9×9
Conv2	32	1×1
Conv3	1	5×5
Conv4~Conv9	64	3×3
Conv10	1	3×3

3.3 通道注意力

注意力机制的作用可以视为将输入的信息进行权重标定，对重要的信息赋予大的权重对其重视，对不重要的信息进行权重削减，使网络在能耗较低的情况下对关键位置的信息进行提取，具有较好的映射表达能力，与此同时切合人类视觉观察事物的特性。

通常卷积层不会利用到各个通道之间的相关性，只是对每个通道单独进行卷积。本文考虑到图像的特征空间的各个通道互有联系且重要性不同，本文引入 SE Net[18]中的 SE block 来给不同通道赋予不同的权重以实现通道注意力机制，通过自适应重新校准来充分捕获信道依赖，使网络在训练中更加关注含有丰富高频信息的通道，加速网络收敛，进一步提升网络性能。SE block 示意图和结构图分别如图 4 和图 5 所示。SE block 分成“squeeze”和“excitation”两个步骤：第一步是使用全局平均池化方法将之前步骤输出的特征空间 $U \in R^{W\times H\times C}$ 切分成 $W\times H$ 个长度为 C 的一维向量，第二步是用两个全连接层通过学习获得参数去构建通道间的非线性相关性。最后是更新权重，先用一个 sigmoid 函数将权重进行归一化操作，再将归一化后的权重加载到各个通道的特征上去，从而完成各个通道原始特征权重的重新标定。

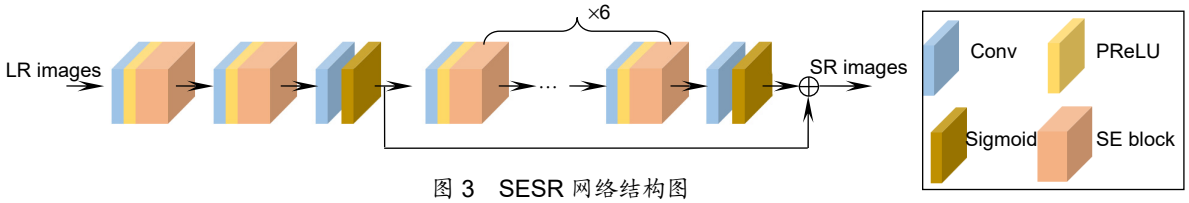

图 3 SESR 网络结构图

Fig. 3 The sketch map of SESR network

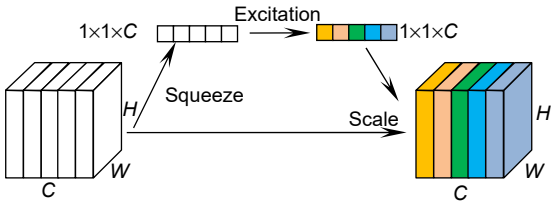

图 4 SE block 示意图

Fig. 4 The schematic diagram of SE block

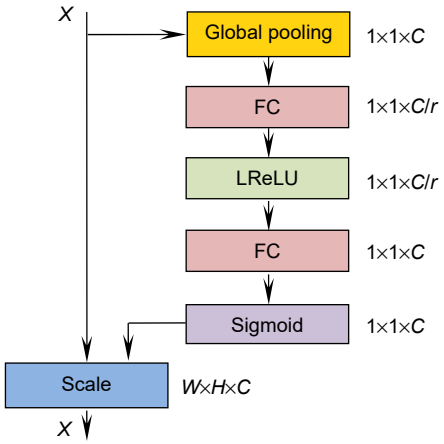

图 5 SE block 结构图

Fig. 5 The sketch map of SE block

3.4 迁移学习和参数微调

传统的机器学习方法关注于解决单一领域内的问题，需要训练集和测试集的特征具有相同的分布。在这种前提下，机器学习只需要尽可能拟合训练集，训练所得的模型就会最大可能在测试集上获得优良的性能。当训练集和测试集的特征不具有相同分布时，通常就需要一个新的数据集，并且要在这个新的数据集上重新训练模型。但是，在实际应用中，重新采集理想的数据集代价很高，往往难以完成。所以此时将从一个领域中学习到的可用的知识迁移到另外一个领域就变得很有必要。

随着近几年深度学习的不断发展，迁移学习也吸引了越来越多的关注。迁移学习就是运用已有的知识去解决不同但相关领域问题。迁移学习可以定义为：给定一个源域 D_s 和学习任务 T_s，一个目标域 D_T 和学习任务 T_T，其中 $D_s \neq D_T$ 和 $T_s \neq T_T$，迁移学习就是利用 D_s 和 T_s 中的已有的有用知识，帮助提高目标域 D_T 中目标预测函数的学习[14]。当源域和目标域的数据集非常接近时，迁移学习可以有效解决目标域样本不足的问题。

基于卷积神经网络的自然图像超分辨率方法以大量的不同场景、目标、条件下的高分辨率自然图像为样本进行训练，样本越多，则训练效果越好。若想将这一方法应用到红外图像的超分辨率上，则需要大量的高质量红外图像作为训练样本。而现实生活中难以收集如此多场景、目标、条件下的高质量红外图像，没有足够的红外图像训练样本很容易导致训练过程中过拟合现象的发生，从而限制了最终的重建效果。这制约了基于卷积神经网络的超分辨率方法在红外图像上的应用与发展。

已有的超分辨率方法大都属于同质图像的超分辨率重建任务，这种情况下深度网络通过标准的有监督训练基本上可以满足应用需求，但是因为高质量红外图像数量少，很难使用同质图像超分辨率方法来完成任务。自然图像与红外图像在成像原理上存在很大差异，这也造成了这两种图像特性不同，例如自然图像包含颜色分量、纹理较多，而红外图像是灰度图像，在目标内容的连通区域亮度值较为均匀等，但是从超分辨率任务的角度来说，任务的最终目的都是追求更高的清晰度和更丰富的细节信息，从主观视觉效果上来看两种图像的超分辨率任务是相似的。所以我们考虑用自然图像训练好的模型进行参数微调这一迁移学习方法来完成红外图像的超分辨率任务。

本文用基于模型参数的迁移学习方式即网络参数微调(fine-tuning)来克服高质量红外图像数量不足的问题，具体步骤是：先用数量充足的自然图像的亮度通道图像作为数据集来预训练 SESR 网络，得到一个适用于自然图像的超分辨率模型。因为在训练红外图像的超分辨率模型时仍然使用 SESR 网络，网络的结构不变，所以直接将预训练好的自然图像超分辨率模型的参数作为红外图像超分辨率模型的初始参数，再用较少数量的红外图像作为训练数据集来继续训练，对这些参数进行微调优化，提升对红外图像的超分辨率的效果，从而得到针对红外图像的超分辨率网络权重模型。这样迁移学习就可以有效缓解深度学习对少量数据学习能力不足的问题。

3.5 多尺度细节提升

由于红外成像设备的限制，红外图像本身清晰度有限，由神经网络输出的图像仍然存在细节不足对比度低的缺点。为了改善这个问题，本文采用多尺度细节提升方法(multi-scale detail boosting，MSDB)[19]对输出图像进行视觉效果上的提升。神经网络的原输出的图像 I^* 如式(1)所示，分别经过三个不同的高斯滤波得

到三幅不同精细尺度的图像 B_1、B_2 和 B_3：

$$B_1 = G_1 \times I^*, B_2 = G_2 \times I^*, B_3 = G_3 \times I^* \quad , \tag{1}$$

其中：G_1、G_2 和 G_3 为高斯核函数，它们的标准差参考文献[19]分别取为 $\sigma_1 = 1$，$\sigma_2 = 2$ 和 $\sigma_3 = 4$。然后再将滤波后的图像 B_1、B_2 和 B_3 按照式(2)生成三幅不同精细程度的细节图 D_1、D_2 和 D_3：

$$D_1 = I^* - B_1, D_2 = B_1 - B_2, D_3 = B_2 - B_3 \quad 。 \tag{2}$$

最后我们按照式(3)合并这三幅不同精细程度的细节图，生成最终的细节图 D^*：

$$D^* = (1 - \omega_1 \times \mathrm{sgn}(D_1)) \times D_1 + \omega_2 \times D_2 + \omega_3 \times D_3 \quad , \tag{3}$$

其中：ω_1、ω_2 和 ω_3 参考文献[17]分别取 0.5、0.5 和 0.25。可以看到式(3)中减少了 D_1 中的正分量，同时放大了 D_1 中的负分量，这是因为在最后一步将最终的细节图和原图像相加时，如果细节图 D_1 直接添加到 I^* 上，会扩展边缘附近的灰度差，可能会使灰度饱和。因此，减少了 D_1 中的正分量，同时放大了 D_1 中的负分量，这样可以在增强细节的同时抑制饱和度。最后，我们将总体细节图 D^* 和神经网络的原输出的图像 I^* 相加，得到最终的细节增强图像。

4 实验结果与分析

实验环境：硬件 PC 平台为 Intel Core i7-6700K CPU，32 GB 内存，Nvidia GeForce GTX1070 显卡，使用的软件为 Caffe，Matlab2016b，Cuda8.0，Cudnn5.1。

4.1 数据集预处理

在第一步预训练时，本文选用与 FRCNN[11]相同的 191 幅自然图像作为训练集，这些图片均有良好的画质，在很多超分辨率方法中得到运用。但是因为实验中网络具有较深的网络结构，可能会造成过拟合现象。针对这个问题，本文对 191 幅训练图像进行数量扩充：先将训练集中的图像分别进行 90°、180°和 270°的旋转，再对这些图像进行缩放，分别取原尺寸的 60%、70%、80%和 90%，最终获取的图像是初始图像数量的 20 倍(4×5=20)，即 3820 幅。

在第二步迁移微调训练时，本文利用搜集到的 39 幅清晰度较高的红外图像作为样本，选择 30 幅进行扩充 20 倍共 600 幅图像作为训练集，剩下 10 幅里再随机选择 4 幅作为验证集，其余 5 幅作为测试集。图 6 是其中部分红外图像。

4.2 实验设置

在第一步预训练自然图像的超分辨率模型时，迭代次数设置为 2×10^6 次(SRCNN 为 1.5×10^7 次)即可保证网络收敛，相较于 SRCNN 大大加快了训练速度。基础学习率设为 0.001，每迭代 2×10^5 次学习率减半，权值衰减设为 0.0001，动量设置成 0.9。本文使用 Adam[20]学习策略来动态调整学习率，其原理是经过偏置校正后，每一次迭代学习率都将获取一个确定的范围，然后在此范围内动态调整学习率，使参数的变化相对平稳。它利用了梯度的一阶和二阶矩估计完成动态调整，其中梯度的一阶矩估计 m_t 和二阶矩估计 n_t 的公式为

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad , \tag{4}$$

$$n_t = \beta_2 \cdot n_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad , \tag{5}$$

式中：β_1=0.9，β_2=0.999，$g_t = \nabla_\theta f_t(\theta)$ 代表随机目标函数 $f_t(\theta)$ 关于 θ 的梯度，t 是当前时间，$t-1$ 代表前一时刻。$\hat{m}_t$ 和 $\hat{n}_t$ 分别是对 m_t、n_t 的修正，$\hat{m}_t = \dfrac{m_t}{1-\beta_1^t}$，$\hat{n}_t = \dfrac{n_t}{1-\beta_2^t}$，用来实现无偏估计；$\beta_1$ 和 β_2 为矩估计指数的衰减速率，θ 是模型参数，η 是步长，权值 θ_t 的更新公式为

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{n}_t} + \varepsilon} \hat{m}_t \quad , \tag{6}$$

其中 $\varepsilon = 10^{-8}$。

在第二步用迁移学习去微调红外图像超分辨率模型时，因为网络结构不变，仍然是 SESR 网络，所以

图 6 部分红外图像样本

Fig. 6 Part of the infrared image samples

利用第一步预训练的模型参数直接构建并初始化第二步的网络，可以大幅减少第二步的训练时间。将基础学习率降为 0.0001，仅是预训练时的 1/10，且固定不变，微调深度和迭代次数经测试后再确定，其他参数与训练自然图像超分辨率模型时保持一致。

本文采用的损失函数是均方误差(mean squared error，MSE)，这有助于后期使用峰值信噪比(PSNR)作为超分辨率重建图像的质量评价指标，MSE 的计算公式为

$$L(\theta)=\frac{1}{n}\sum_{i=1}^{n}[G(Y_i,\theta)-X_i]^2 \quad 。 \tag{7}$$

通过最小化重建图像与高清原始图像的均方误差来估计参数 θ。其中 n 为训练样本个数，G 为卷积神经网络函数，X_i 和 Y_i 分别为原始 GT 图像和低分辨率图像，那么 $G(Y_i,\theta)$ 就是重建后的图像。MSE 也被目前绝大多数超分辨率重建方法使用。

4.3 重建效果对比

本文采取的图像重建质量评价指标有峰值信噪比(peak signal-to-noise ratio，PSNR)和结构相似度(structural similarity，SSIM)，其中 PSNR 是超分辨率领域使用最广泛的图像评价指标之一，是基于对应像素点之间的误差进行质量评价，越大的 PSNR 的值表明越小的重建图像失真度，即越好的重建效果，但 PSNR 的数值仅仅依赖于像素间低层次的差别，它不能完全代表人类视觉效果的优劣。SSIM 分别从结构、亮度、对比度三个方面来度量图像之间的相似性，SSIM 值越接近 1，说明重建后的图像与原图结构越相似，重建效果越好。红外图像虽然成像原理与自然图像不同，但是在判断重建效果优劣时，也是用原始高清图像作为参考来判断的，失真度和结构、亮度、对比度等的差异同样都可以作为判断依据，所以 PSNR 和 SSIM 作为红外重建图像的评价指标也具有较高的说服力。

为了证明本文设计的网络在超分辨率重建任务中的有效性，本文先将预训练的自然图像超分辨率模型与几种已有的模型分别在 Set5 和 Set14 数据集上与传统的 Bicubic 方法、基于学习的 SRCNN 算法、FSRCNN 算法以及 ESPCN 算法在自然图像数据集上进行测试比较，放大因子分别取 2、3 和 4，它们的重建结果如表 2 所示。评价指标取 PSNR 和 SSIM。

为了更直观地比较各方法的效果，图 7 和图 8 给出了表 2 中超分辨率方法在放大因子为 3 时的 2 幅图像的重建结果并进行局部放大。

由表 2 可以看出 SESR 网络的重建结果评价指标普遍优于前 3 种重建方法，其中用 Bicubic 方法生成结果的 PSNR 值和 SSIM 值最低，SRCNN 方法和 FSRCNN 方法均优于 Bicubic，而 SESR 方法的 PSNR 值和 SSIM 值均是这些方法中最高的，表明该方法的重建效果更好，与原始高清图像最接近。在放大倍数为 2 时，Set5 数据集上 SESR 方法的 PSNR 值达到了 37.39 dB，SSIM

表 2　5 种方法在 2 个数据集上的重建结果(PSNR/SSIM)

Table 2　Reconstruction results of 5 methods on 2 datasets (PSNR/SSIM)

Data set	Scale	Bicubic	SRCNN	FSRCNN	ESPCN	SESR
Set5	2	33.66/0.9299	36.66/0.9542	37.00/0.9558	37.06/0.9559	**37.39/0.9586**
	3	30.39/0.8682	32.75/0.9090	33.16/0.9104	33.13/0.9135	**33.38/0.9142**
	4	28.42/0.8104	30.48/0.8628	30.71/0.8657	30.90/0.8673	**31.03/0.8735**
Set14	2	30.24/0.8688	32.42/0.9063	32.63/0.9088	32.75/0.9098	**32.94/0.9116**
	3	27.55/0.7742	29.28/0.8209	29.42/0.8242	29.49/0.8271	**29.66/0.8298**
	4	26.00/0.7027	27.49/0.7503	27.59/0.7535	27.73/0.7637	**27.85/0.7721**

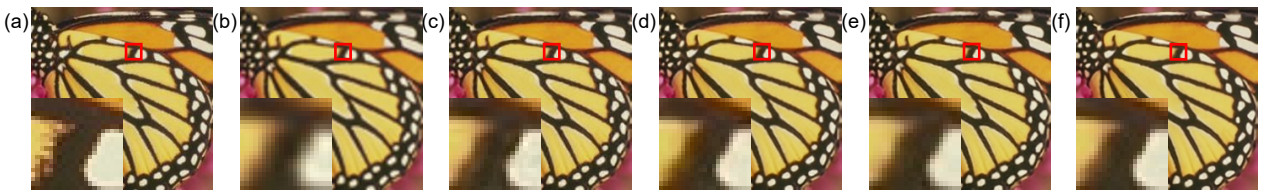

图 7　Butterfly 重建结果。(a) 原图；(b) Bicubic；(c) SRCNN；(d) FSRCNN；(e) ESPCN；(f) SESR

Fig. 7　Reconstruction results of a Butterfly. (a) Original image; (b) Bicubic; (c) SRCNN; (d) FSRCNN; (e) ESPCN; (f) SESR

图 8 Lenna 重建结果。(a) 原图；(b) Bicubic；(c) SRCNN；(d) FSRCNN；(e) ESPCN；(f) SESR

Fig. 8 Reconstruction results of Lenna. (a) Original image; (b) Bicubic; (c) SRCNN; (d) FSRCNN; (e) ESPCN; (f) SESR

值达到了 0.9586，PSNR 比 Bicubic 提升了 3.73 dB，SSIM 也提高了 0.0287。

从图 7 和图 8 可以直观地看出，Bicubic 方法的重建图像最模糊；SRCNN 的效果有所提升但是细节不够清晰，效果一般；ESPCN 和 SESR 方法能够较好的恢复细节，但 SESR 方法的重建图像的边缘更清晰，细节更清楚，更接近原图。

为了证明用 PReLU 作为神经网络的激活函数相较于 ReLU 的优越性，本文做了对比实验，分别用两个函数作为网络的激活函数训练模型。图 9 是两个函数分别用 Set5 和 Set14 数据集在 3 倍放大因子下的测试结果，可以看到，PReLU 函数作为网络激活函数的重建结果有更高的 PSNR 值，所以本文采用 PReLU 作为最终的网络激活函数。

为了验证通道注意力机制在超分辨率重建任务中的作用，本文将去除了通道注意力机制的 SESR 网络(将原 SESR 网络中去掉所有的 SE block)和原始的 SESR 网络重建效果进行对比实验，评价指标用 PSNR 和 SSIM，表 3 列出了 3 种方式在 Set5 和 Set14 数据集上的重建结果。

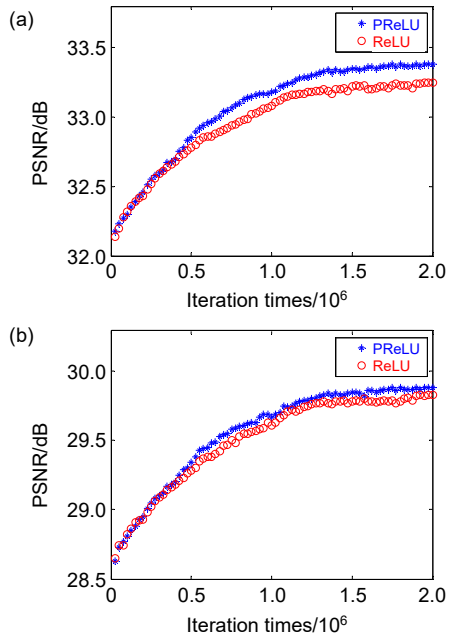

图 9 不同激活函数下 Set5 和 Set14 测试集的 PSNR 均值随迭代次数增长的趋势图。(a) Set5；(b) Set14

Fig. 9 Trend of the mean PSNR of Set5 and Set14 datasets with iteration rising under different active functions. (a) Set5; (b) Set14

表 3 2 种方式在 2 个数据集上的重建结果

Table 3 Reconstruction results of 2 methods on 2 datasets

Data set	Scale	SESR(no attention)	SESR
Set5	2	37.31/0.9576	**37.39/0.9586**
	3	33.27/0.9136	**33.38/0.9142**
	4	31.02/0.8707	**31.03/0.8735**
Set14	2	32.88/0.9108	**32.94/0.9116**
	3	29.59/0.8283	**29.66/0.8298**
	4	**27.86**/0.7720	27.85/**0.7721**

图 10 和图 11 分别给出了 2 幅图像在 3 种情况下放大因子为 3 时的超分辨率重建结果图并进行局部放大。

从表 3、图 10 和图 11 可以看出，去除通道注意力机制的 SESR 网络的 PSNR 值均稍低于原始的 SESR 网络，6 组数据 PSNR 平均降低了 0.05 dB，SSIM 平均降低了 0.0011，说明了 SE block 在超分辨率重建网络中具有一定的优化作用，它使网络的学习能力更强，重建效果更优。

通常在 CNN 网络中，网络前端学习的是低层次图像特征，如颜色、边缘、形状等，而网络后端学习的是当前任务要求的高层次特征。为了节约资源，一般的基于模型参数迁移通常对前若干层的参数不更新。但是本文考虑到自然图像与红外图像成像、颜色以及边缘等基本特征的差异性，所以对微调的深度进行了测试并讨论。本文将迁移的深度设置为微调 Conv7 至 Conv10、微调 Conv4 至 Conv10 以及微调全部卷积层三种，测试的结果如图 12 所示。

可以看出微调全部卷积层参数时，对 5 张红外图像的平均 PSNR 值能达到最高，而只微调 Conv7 至 Conv10 时细节效果最差，所以本文选择微调全部卷

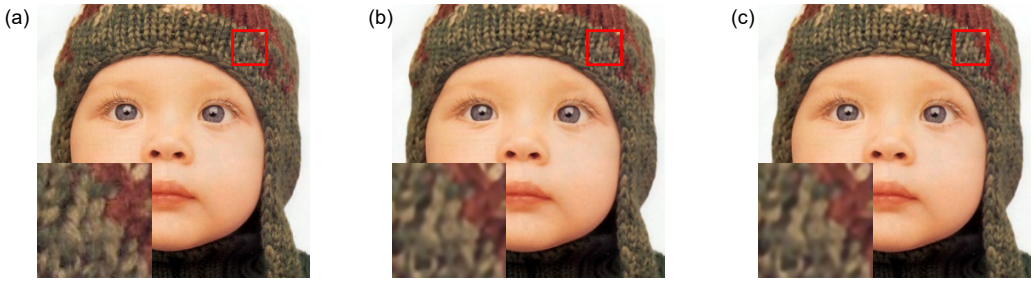

图 10 Baby 重建结果。(a) 原图；(b) SESR(no attention)；(c) SESR

Fig. 10 Reconstruction results of a Baby. (a) Original image; (b) SESR(no attention); (c) SESR

图 11 Comic 重建结果。(a) 原图；(b) SESR(no attention)；(c) SESR

Fig. 11 Reconstruction results of a Comic. (a) Original image; (b) SESR(no attention); (c) SESR

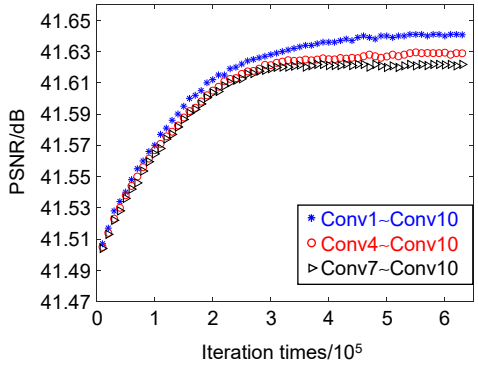

图 12 不同微调深度下 5 幅红外图像 PSNR 均值随迭代次数增长的趋势图

Fig. 12 Trend of the mean PSNR of 5 infrared images with iteration rising under different fine-tuning depth

积层的参数。另外从图 12 上可以看出，微调全部卷积层时迭代次数选择 5×10^5 时网络就能收敛，所以本文的参数迁移迭代次数选择 5×10^5。

为了证明迁移学习在红外图像超分辨率重建中的有效性，将迁移学习后的红外图像超分辨率模型(SESR_T)与其他 3 种经典方法(Bicubic、SRCNN、FSRCNN)、迁移学习前的重建模型(SESR)以及直接用少量的红外样本直接训练出来的重建模型(SESR_I)在 5 幅红外图像测试集上进行 3 倍放大重建测试，重建质量的评价指标为 PSNR 和 SSIM，重建的结果分别如表 4 所示。

图 13 和图 14 分别给出了 2 幅红外测试图像在放大因子为 3 时各方法的超分辨率重建结果图并进行局

表 4 6 种方法在 5 幅红外图像上的测试结果(PSNR/SSIM)

Table 4 Reconstruction results of 6 methods on 5 infrared images(PSNR/SSIM)

Methods	Car1	Car2	People1	People2	Road	Average
Bicubic	39.21/0.9480	37.22/0.9362	42.48/0.9558	41.02/0.9407	40.67/0.9364	40.12/0.9434
SRCNN	40.69/0.9545	38.54/0.9456	43.11/0.9587	41.61/0.9453	41.21/0.9413	41.05/0.9491
FSRCNN	41.23/0.9565	39.20/0.9497	43.27/0.9591	41.82/0.9469	41.36/0.9422	41.38/0.9509
SESR_I	41.17/0.9561	39.13/0.9489	43.23/0.9589	41.80/0.9463	41.33/0.9419	41.33/0.9504
SESR	41.43/0.9576	39.44/0.9521	43.30/0.9595	41.94/0.9484	41.34/0.9425	41.49/0.9520
SESR_T	**41.75/0.9584**	**39.58/0.9524**	**43.38/0.9598**	**42.01/0.9486**	**41.48/0.9431**	**41.64/0.9525**

部放大。

从图 13、图 14 能直观地看出，预训练的模型 SESR 在红外图像上重建和迁移学习后的 SESR_T 重建结果相较于其他方法边缘更明显，线条更清晰，视觉效果更优，SESR 和 SESR_T 两者的视觉效果虽然差距不大，但是从表 4 结果看出，SESR_T 重建的 5 幅红外图像的平均 PSNR 值和 SSIM 值分别为 41.64 dB 和 0.9525，比 SESR 在红外图像上重建的 PSNR 值和 SSIM 值分别平均增加了 0.25 dB 和 0.0005，比直接用少量红外图像训练的模型 SESR_I 分别平均增加了 0.31 dB 和 0.0021，红外图像的超分辨率重建的质量得到了提升，证明了迁移学习在高质量红外样本不足情况下对红外图像超分辨率重建的有效性。SESR 和 FSRCNN 都是用自然图像训练的模型，SESR 的评价标准有明显优势，而 SESR_I 在红外训练集较小(600 幅)的情况下，其重建结果 PSNR 值仍能和 FSRCNN 方法(3820 幅)很接近，

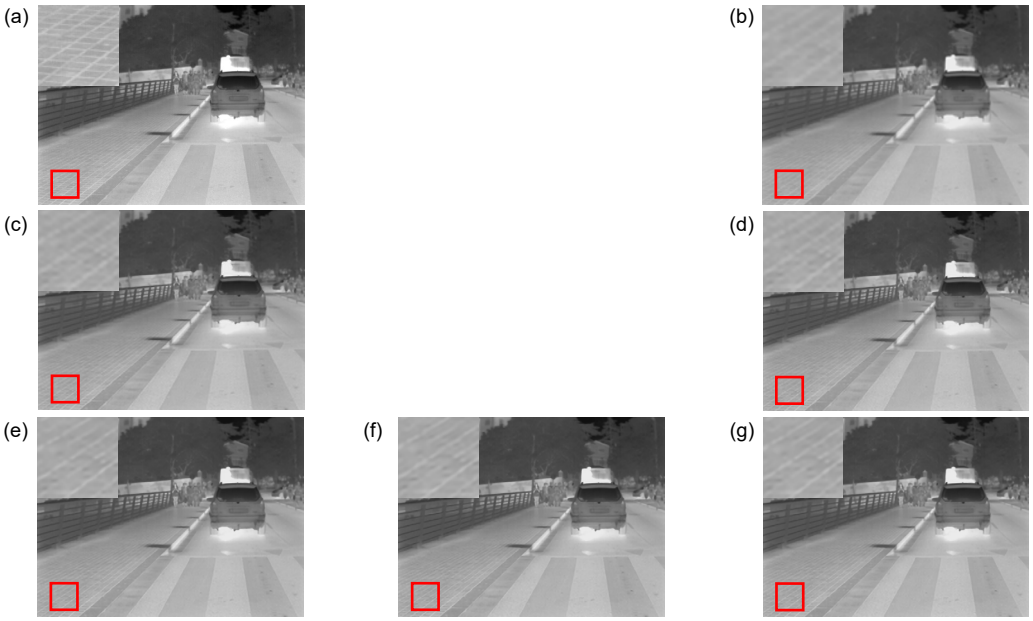

图 13 Car1 重建结果。(a) 原图；(b) Bicubic；(c) SRCNN；(d) FSRCNN；(e) SESR_I；(f) SESR；(g) SESR_T

Fig. 13 Reconstruction results of Car1. (a) Original image; (b) Bicubic; (c) SRCNN; (d) FSRCNN; (e) SESR_I; (f) SESR; (g) SESR_T

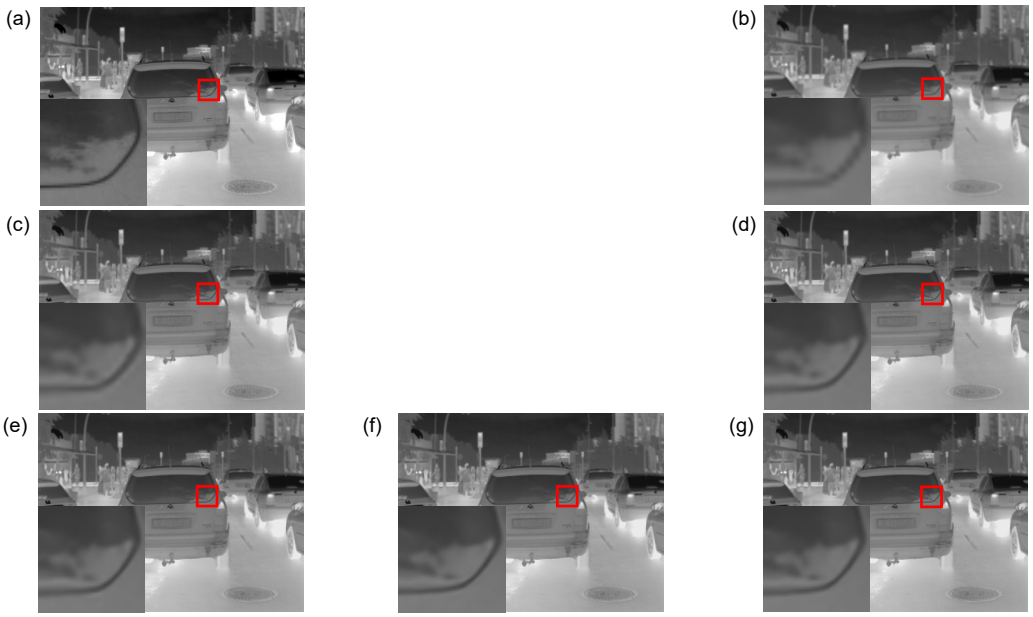

图 14 Car2 重建结果。(a) 原图；(b) Bicubic；(c) SRCNN；(d) FSRCNN；(e) SESR_I；(f) SESR；(g) SESR_T

Fig. 14 Reconstruction results of Car2. (a) Original image; (b) Bicubic; (c) SRCNN; (d) FSRCNN; (e) SESR_I; (f) SESR; (g) SESR_T

这两点能充分说明 SESR 超分辨率重建网络具有一定的优势。

考虑到MSDB滤波不以与高清原图的相似性为目的，而是尽可能提升图像的细节以增大信息量，但是PSNR和SSIM都是以高清原始图像为参考的质量评价指标，所以为了更好地验证 MSDB 滤波的效果，本文除了采用 PSNR 和 SSIM 外，还采用了一种无参考的空间域图像质量评价指标：NIQE[21]，从重建图像自身的质量来比较重建效果的优劣。NIQE 的原理是正常图像的归一化亮度系数存在明显的单位标准高斯分布，而噪声或失真都会破坏这一规律，NIQE 的数值越小，图像质量越高。相较于 MSE、PSNR、SSIM 等评价算法，NIQE 算法表现出更好的预测稳定性、单调性和一致性，而且 NIQE 在提取图像统计特征时先对图像提取了感兴趣区域，所以 NIQE 算法与人眼主观质量评价有着更好的一致性，更加接近人类视觉系统[22]。为了验证 MSDB 滤波的有效性，将普通的 SESR_T 模型与 SESR_T+MSDB 方法进行重建效果对比，表 5 是两种方法在三种图像质量评价指标下对 5 幅红外图像的重建结果。

图 15 和图 16 分别给出了 Car1 和 People2 两幅红外图像用SESR-T和SESR-T+MSDB重建的结果图并进行局部放大。

从表 5 可得，MSDB 滤波会普遍降低重建图像的PSNR，5幅红外图像PSNR值平均降低了0.97 dB，SSIM值有增有减，5 幅红外图像平均只相差 0.0005，说明 MSDB 不会对图像的结构、亮度、对比度产生较大的影响。而 MSDB 滤波使得 NIQE 值得到了降低，5 幅图像平均降低了 0.4064，说明经过 MSDB 滤波后的图像虽然 PSNR 值稍低，但是其本身质量更高，具有更多的信息量。从图 15、图 16 中也能直观地看到，经过 MSDB 滤波后的各红外图像的轮廓更清晰，对比度更高，视觉效果更优，图像自身的质量得到了有效提高，主观的视觉效果与客观评价指标 NIQE 基本一致。

表 5　5 幅红外图像上 3 种评价指标的重建结果

Table 5　Reconstruction results of 3 assessment indicator on 5 infrared images

Methods	Assessment indicator	Car1	Car2	People1	People2	Road	Average
SESR_T	PSNR	**41.75**	**39.58**	**43.38**	**42.01**	**41.48**	**41.64**
	SSIM	**0.9584**	0.9524	**0.9598**	**0.9486**	0.9431	**0.9525**
	NIQE	5.0882	4.6436	5.4748	4.8767	5.2902	5.0747
SESR_T+MSDB	PSNR	40.12	38.46	42.68	41.38	40.73	40.67
	SSIM	0.9554	**0.9537**	0.9578	0.9463	**0.9466**	0.9520
	NIQE	**4.4907**	**4.3997**	**4.8909**	**4.6168**	**4.9432**	**4.6683**

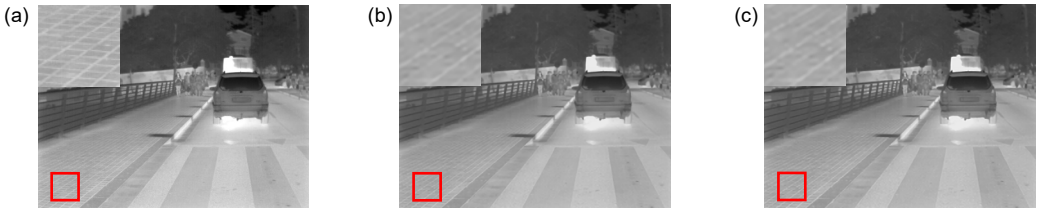

图 15　Car1 重建结果。(a) 原图；(b) SESR_T；(c) SESR_T+MSDB

Fig. 15　Reconstruction results of Car1. (a) Original image; (b) SESR_T; (c) SESR_T+MSDB

图 16　People2 重建结果。(a) 原图；(b) SESR_T；(c) SESR_T+MSDB

Fig. 16　Reconstruction results of People2. (a) Original image; (b) SESR_T; (c) SESR_T+MSDB

5 结 论

针对现有红外图像分辨率低、质量不高的问题，本文提出了一种基于通道注意力与迁移学习的红外图像超分辨率重建方法。该方法首先设计了一个深度卷积神经网络，融入通道注意力机制，来学习特征空间通道之间的相关性，增强了网络的学习能力，并且使用残差学习方式来减轻梯度爆炸或消失问题，加速网络的收敛。考虑到高质量的红外图像难以采集、数目不足，所以本文采用两步学习方式：第一步使用自然图像来预训练获取一个自然图像的超分辨率模型，第二步利用迁移学习的知识，用较少数量的高质量红外图像将预训练获取到的模型参数进行迁移微调训练，提升模型对红外图像的重建效果，从而得到了针对红外图像的超分辨率模型。最后加入的多尺度细节滤波模块来提升红外重建图像的视觉效果。实验结果表明，加深网络的深度和引入通道注意力机制，均能提升网络的学习能力，对重建的效果有积极作用；迁移学习可以很好地解决高质量红外图像数目不足的问题，迁移学习后的模型对红外图像的重建效果相较于迁移学习之前和直接用少量红外图像训练出来的模型均有所提升；MSDB 滤波模块能提升重建后的红外图像的细节和人类的视觉效果，增大了信息量，提升了红外图像的应用价值。

参考文献

[1] Su H, Zhou J, Zhang Z H. Survey of super-resolution image reconstruction methods[J]. *Acta Autom Sin*, 2013, **39**(8): 1202–1213.
苏衡, 周杰, 张志浩. 超分辨率图像重建方法综述[J]. 自动化学报, 2013, **39**(8): 1202–1213.

[2] Bätz M, Eichenseer A, Seiler J, *et al*. Hybrid super-resolution combining example-based single-image and interpolation-based multi-image reconstruction approaches[C]//*Proceedings of 2015 IEEE International Conference on Image Processing (ICIP)*, 2015: 58–62.

[3] Kim K I, Kwon Y. Single-image super-resolution using sparse regression and natural image prior[J]. *IEEE Trans Pattern Anal Mach Intell*, 2010, **32**(6): 1127–1133.

[4] Lian Q S, Zhang W. Image super-resolution algorithms based on sparse representation of classified image patches[J]. *Acta Electron Sin*, 2012, **40**(5): 920–925.
练秋生, 张伟. 基于图像块分类稀疏表示的超分辨率重构算法[J]. 电子学报, 2012, **40**(5): 920–925.

[5] Xiao J S, Liu E Y, Zhu L, *et al*. Improved image super-resolution algorithm based on convolutional neural network[J]. *Acta Opt Sin*, 2017, **37**(3): 0318011.
肖进胜, 刘恩雨, 朱力, 等. 改进的基于卷积神经网络的图像超分辨率算法[J]. 光学学报, 2017, **37**(3): 0318011.

[6] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections[J]. *J Opt Soc Am A*, 1989, **6**(11): 1715–1726.

[7] Irani M, Peleg S. Improving resolution by image registration[J]. *CVGIP: Graph Models Image Process*, 1991, **53**(3): 231–239.

[8] Chang H, Yeung D Y, Xiong Y M. Super-resolution through neighbor embedding[C]//*Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004: 275–282.

[9] Yang J C, Wright J, Huang T S, *et al*. Image super-resolution via sparse representation[J]. *IEEE Trans Image Process*, 2010, **19**(11): 2861–2873.

[10] Dong C, Loy C C, He K M, *et al*. Image super-resolution using deep convolutional networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2016, **38**(2): 295–307.

[11] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 391–407.

[12] Shi W Z, Caballero J, Huszár F, *et al*. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1874–1883.

[13] Ledig C, Theis L, Huszár F, *et al*. Photo-realistic single image super-resolution using a generative adversarial network[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4681–4690.

[14] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Trans Knowl Data Eng*, 2010, **22**(10): 1345–1359.

[15] Xu Z, Qu C W, He L Q. SAR target super-resolution based on transfer learning[J]. *Acta Aeronaut Astronaut Sin*, 2015, **36**(6): 1940–1952.
徐舟, 曲长文, 何令琪. 基于迁移学习的 SAR 目标超分辨重建[J]. 航空学报, 2015, **36**(6): 1940–1952.

[16] Yanai K, Kawano Y. Food image recognition using deep convolutional network with pre–training and fine–tuning[C]//*Proceedings of 2015 IEEE International Conference on Multimedia & Expo Workshops*, 2015: 1–6.

[17] Du B, Xiong W, Wu J, *et al*. Stacked convolutional denoising auto-encoders for feature representation[J]. *IEEE Trans Cybern*, 2017, **47**(4): 1017–1027.

[18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.

[19] Kim Y, Koh Y J, Lee C, *et al*. Dark image enhancement based onpairwise target contrast and multi-scale detail boosting[C]//*Proceedings of 2015 IEEE International Conference on Image Processing*, 2015: 1404–1408.

[20] Kingma D P, Ba J. Adam: a method for stochastic optimization[Z]. arXiv: 1412.6980, 2014.

[21] Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer[J]. *IEEE Signal Process Lett*, 2013, **20**(3): 209–212.

[22] Shao X, Zeng T Y, Wang Z H. No-reference quality assessment method for printed image based on NIQE[J]. *Packaging J*, 2016, **8**(4): 35–39.
邵雪, 曾台英, 汪祖辉. 一种基于 NIQE 的印刷图像无参考质量评价方法[J]. 包装学报, 2016, **8**(4): 35–39.

Super-resolution reconstruction of infrared image based on channel attention and transfer learning

Sun Rui[1,2], Zhang Han[1,2*], Cheng Zhikang[1,2], Zhang Xudong[1]

[1]School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China;
[2]Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China

Effect picture

Overview: In recent years, infrared imaging technology has developed rapidly and has been increasingly used in military reconnaissance, security surveillance, and medical imaging. However, in the process of infrared image imaging or transmission, it is affected by many factors such as environment and equipment. The infrared image often has a low resolution, which greatly reduces the amount of information contained in the infrared image and restricts the application value of the infrared image. Therefore, how to obtain high-resolution and high-information infrared images has become an issue that people urgently need to solve. In recent years, the development of deep learning technology has made rapid progress, and super-resolution methods based on deep learning have begun to appear. However, if these convolutional neural networks are directly applied to the infrared image field, there are some problems: SRCNN, FSRCNN, and ESPCN have fewer network convolutional layers and insufficient network depth, and the learning features will be relatively single, ignoring the differences between image features. The mutual relationship makes it difficult to extract the deep-level information of the infrared image, and SRGAN may generate super-resolution images that are not close to the original image in certain details, which is not conducive to the application of infrared images in military, medical and surveillance. Another problem that needs to be overcome is that it is difficult to collect a sufficient number of high-quality infrared images in real life, and a large number of images of different scenes and targets are required as training samples for common deep learning methods. The infrared images used as training data sets to achieve deep learning methods often fail to achieve the desired effect. In order to solve these problems, this paper proposes a method for super-resolution reconstruction of infrared images based on channel attention and transfer learning. This method first designs a deep convolutional neural network, which integrates the channel attention mechanism to learn the correlation between the channels in the feature space, enhances the learning ability of the network, and uses residual learning to reduce the problem of gradient explosion or disappearance and to speed up network convergence. Then, considering that high-quality infrared images are difficult to collect and insufficient in number, the network training is divided into two steps: the first step uses natural images to pre-train a super-resolution model of natural images, and the second step is to use transfer learning knowledge. Using a small number of high-quality infrared images, the pre-trained model parameters are quickly transferred and fine-tuned to improve the reconstruction effect of the model on the infrared image, thereby obtaining a super-resolution model of the infrared image. Finally, a multi-scale detail boosting (MSDB) module is added to enhance the details and visual effects of the infrared reconstructed image and to increase the amount of information.

Sun R, Zhang H, Cheng Z K, *et al.* Super-resolution reconstruction of infrared image based on channel attention and transfer learning[J]. *Opto-Electron Eng*, 2021, **48**(1): 200045; DOI: 10.12086/oee.2021.200045

Foundation item: National Natural Science Foundation of China (61471154, 61876057) and the Fundamental Research Funds for Central Universities (JZ2018YYPY0287)
* E-mail: 1402198087@qq.com