

# 基于对偶循环结构及注意力机制的红外图像超分辨率重建

闫昊天, 程良伦, 吴 衡

广东工业大学, 自动化学院, 广东 广州

收稿日期: 2022年3月2日; 录用日期: 2022年3月31日; 发布日期: 2022年4月7日

---

## 摘 要

针对红外图像空间分辨率低、成像效果不好的问题, 提出了一种基于对偶循环结构和注意力机制的红外图像超分辨率重建方法。对偶循环结构的引入能够更好地约束LR到HR的映射, 通过引入融合了多维度的注意力机制CBAM, 让网络在不显著提高计算量与参数的情况下取得了细节更丰富的重建效果。通过在真实红外数据集上与现有的典型方法进行比较, 所提方法在显著降低了模型的参数数量的情况下取得了不错的重建效果。

## 关键词

红外图像, 超分辨率重建, 对偶循环结构, 注意力机制

---

# Super Resolution Reconstruction of Infrared Image Based on Dual Cycle Structure and Attention Mechanism

Haotian Yan, Lianglun Cheng, Heng Wu

School of Computer Automation, Guangdong University of Technology, Guangzhou Guangdong

Received: Mar. 2<sup>nd</sup>, 2022; accepted: Mar. 31<sup>st</sup>, 2022; published: Apr. 7<sup>th</sup>, 2022

---

## Abstract

Aiming at the problems of low spatial resolution and poor imaging effect of infrared image, a super-resolution reconstruction method of infrared image based on dual cycle structure and attention mechanism is proposed. The introduction of dual loop structure can better restrict the map-

ping from LR to HR. By introducing CBAM, which integrates multi-dimensional attention mechanism, the network can achieve more detailed reconstruction without significantly improving the amount of calculation and parameters. Compared with several existing typical methods, the proposed method achieves good results under the condition of significantly reducing the number of parameters of the model.

## Keywords

Infrared Image, Super-Resolution Reconstruction, Dual Loop Structure, Attention Mechanism

Copyright © 2022 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

红外图像是一种通过热红外传感器接收目标物体表面所发出的红外辐射，并根据红外辐射强度的分布形成的图像，它具有抗干扰能力强、适应性强、夜间等微光环境下的成像效果好等优点，具有非常广泛的应用[1]。例如，红外图像可以用在机场、火车站等人流密集场所快速、无感地进行体温检测[2]；在电力巡检、变电站等工业场所实时监控设备的温度[3]；在夜间对自然环境中的人员进行搜索救援[4] [5]；在国防、军事领域的应用等[6] [7]。

然而，受制于红外传感器的制造水平、成本以及装置的尺寸大小等因素，现有的许多红外图像往往存在分辨率低、对比度低、拍摄物体的边缘等细节信息比较模糊以及存在各种噪声的问题，这些问题的存在一定程度上限制了红外成像设备的应用[8]。提高红外图像的成像质量，可以从其硬件的设计与制造水平上入手，但这样的技术难度、研发周期和成本也相对较高。

随着计算机技术的发展，越来越多的图像处理方法被应用在了提高图像分辨率上。20世纪60年代，由Harris提出的单幅图像超分辨率重建(Single Image Super-Resolution, SISR)，是一种利用算法将给定的低分辨率图像(Low Resolution, LR)进行重建，从而生成高分辨率的图像(High Resolution, HR)的过程[9]。早期的研究者主要采用的是基于插值的方法进行超分辨率重建的，例如双三次插值(Bicubic interpolation)及其改进方法[10] [11]等。基于插值的方法本质上是在源图像的像素值之间通过加权平均计算出新的像素值并插入其中，尽管在放大图像分辨率的同时能保证像素值相对连续，但由于其低通特性，使得图像的高频分量受损，从而让重建后的HR图像变得模糊。

近年来，越来越多的研究者开始在图像超分辨率重建领域采用深度学习这种端到端方法并且取得了良好的效果。Dong等人于2014年提出了SRCNN [12]，这是首个采用深度学习的超分辨率重建算法。该算法先将低分辨率的LR图像用Bicubic算法进行插值放大到目标尺寸，再输入到卷积网络中进行非线性映射，得到成像效果更好的超分图像(Super Resolution, SR)。然而，作为深度学习在超分领域的首次尝试，SRCNN的网络深度非常浅，难以提取到图像更深层次的特征，限制了他的重建效果。作为改进，Kim等人提出的VDSR网络大大增加了网络的深度，并引入残差结构以减轻网络的退化情况[13]。2017年，Ledig等人提出的SRGAN [14]首次在超分辨率重建领域引入了生成对抗(Generative Adversarial Networks, GAN)的方式，它提出了一种由Adversarial Loss和Content Loss组成的感知损失函数。SRGAN在一定程度上解决了传统基于深度学习的方法缺乏高频细节信息、信噪比高的问题。2018年，Wang等人同样借用了GAN的思想提出了ESRGAN [15]，其将SRGAN网络中最基本的残差结构改成了Residual-in-Residual

Dense Block (RRDB)。此外, 作者还发现并验证了 GAN 框架中的归一化层(Batch Normalization, BN)不仅影响性能而且对训练的稳定性产生影响, 最终在 ESRGAN 中去掉了 BN 层, 减少了计算复杂度及内存的占用。RCAN [16]是 Zhang 等人在 2018 年提出的一种超分辨率重建算法, 它采用了一种由几个具有长跳跃连接的残差组构成的 Residual in Residual (RIR)结构。在 RIR 结构中, 每个残差组内部由一些具有短跳跃连接的残差块组成。不同层级的残差结构相互组合, 使得神经网络在学习特征时能够尽量少地将算力浪费在图片的低频信息上, 提高了处理效率。最近, Guo 等人提出了一种由 UNet [17]和一个对偶的辅助分支组成的超分辨率模型 DRN [18]。其中, UNet 结构用于实现 LR 图像到 HR 图像的映射生成 SR, 而其对偶辅助分支则是将生成的 SR 映射回 LR 并于原始的 LR 进行比较, 从而更好地协助主分支进行超分辨率重建。

如今, 注意力机制逐渐成为计算机视觉领域中的研究热点, 其也越来越多地被运用在了超分辨率重建领域, 并取得了良好的效果。上述 RCAN 的每个残差块是由多个通道注意力机制 RCAB 组成的, 目的在于为对重建更加重要的通道分配更高的权重。Woo 等人提出了一种新的注意力机制 CBAM [19], 该方法在空间和通道两个方向上计算了注意力并将二者的结果进行了融合, 取得了更好的效果。

尽管基于深度学习的超分辨率算法近年来取得了长足的进展, 但仍存在以下问题: 首先, 大多数超分辨率算法都是应用于彩色的三通道图像的, 但应用在纹理信息少、噪声多、高频信息本就不丰富的单通道红外图像上效果不佳。其次, 采用深度学习这种基于数据驱动的超分辨率算法需要大量的由 LR/HR 组成的图像对, 其中 LR 输入网络中进行训练, 生成的超分后的图像 SR 与作为 Ground Truth 参考的 HR 进行比较、计算损失。现实中很难获得两幅相同视角但分辨率不同的 LR 和 HR 来构成图像对, 因此业内常用的方法是利用 bicubic 下采样、高斯模糊等算法人工将 HR 退化得到 LR 从而构成 LR/HR 图像对。然而, 现实中成像质量不高的低分辨率图像往往是由于传感器、镜头组件以及环境等因素造成的, 真实的 LR/HR 映射关系与人工降质生成的数据集存在差异, 导致了现有方法的性能有限。此外, 超分辨率是一种相对低级的计算机视觉任务, 过大的参数量和计算消耗不利于算法模型的部署与应用。

针对以上问题, 本文提出了一种基于对偶循环结构和注意力机制的红外图像超分方法。得益于对偶循环结构的运用, 我们提出的方法不仅学习单独的 LR 到 HR 的映射, 还额外学习了一条从 HR 到 LR 的映射并引入对偶损失, 让重建出来的效果更加真实。在特征提取方面, 我们将 CBAM 这种融合了多个方向的注意力机制模块应用到超分辨率重建网络中, 在不显著提高计算量与参数的情况下取得了细节更丰富的重建效果。

## 2. 方法描述

我们所提出的网络结构图如图 1 所示。该网络由一条用于将 LR 映射到 SR 的超分分支, 和一条将 SR 图像映射到 LR'和 LR''用于计算对偶损失(Dual Loss)的对偶分支组成。超分分支和对偶分支共同组成了循环结构。

在超分分支中, 输入的低分辨率红外图像  $I_{LR}$  先经过一个卷积层提取低级特征, 然后特征图会依次经过由 20 个 CBAM 组成的注意力机制模块进一步提取深层次的特征, 然后, 通过 PixelShuffle 像素重组上采样, 对特征图进行第一轮放大。然后, 对特征图重复上述的过程, 再次经过由 20 个带有残差结构的 CBAM 模块和 PixelShuffle 上采样环节。最后, 再使用一个卷积层让特征图恢复到高分辨率的图像  $I_{HR}$ , 完成超分辨率重建。这一过程的公式表示如下所示:

$$\begin{aligned} x_s &= f_{conv}(I_{LR}) \\ x_{CBAM1} &= f_{CBAM}^n \left( f_{CBAM}^{n-1} \left( \dots f_{CBAM}^0(x_s) \dots \right) \right) \\ x_{up1} &= f_{up}(x_{CBAM1}) \end{aligned} \quad (1)$$

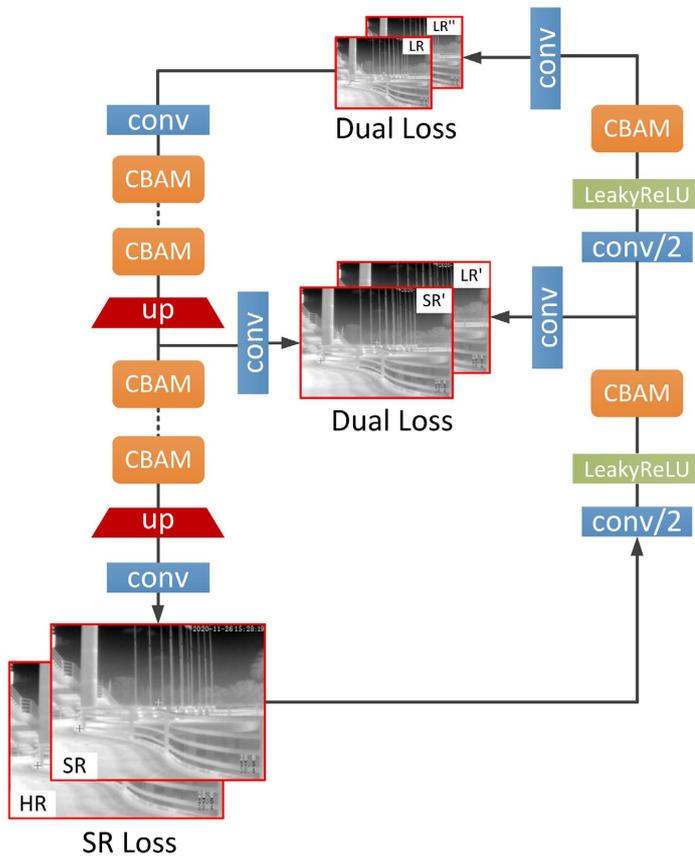


Figure 1. The architecture of the proposed method  
图 1. 网络结构图

$$\begin{aligned}
 x_{CBAM2} &= f_{CBAM}^n \left( f_{CBAM}^{n-1} \left( \dots f_{CBAM}^0 \left( x_{up1} \right) \dots \right) \right) \\
 x_{up2} &= f_{up} \left( x_{CBAM2} \right) \\
 I_{SR} &= f_{re} \left( x_{up2} \right)
 \end{aligned}
 \tag{2}$$

在公式(1)中， $I_{LR}$  表示输入的低分辨率红外图像， $f_{conv}(\cdot)$  表示网络中首个用于提取低级特征的  $3 \times 3$  卷积层，提取到的浅层特征图用  $x_s$  表示。 $f_{CBAM1}^n(\cdot)$  表示网络中的 20 个 CBAM 模块， $n \in 0, 1, \dots, 19$ ，用来提取深层次的特征，得到的深层特征图用  $x_{CBAM1}$  表示。最后， $f_{up1}(\cdot)$  表示 PixelShuffle 上采样模块。在公式(2)中，整体的网络结构与公式(1)类似，最后的  $f_{re}(\cdot)$  是一个  $3 \times 3$  的卷积层，用于重建红外图像，重建后的高分辨率图像由  $I_{SR}$  表示。

对偶循环分支如图 1 右边所示。将  $I_{SR}$  依次输入由  $3 \times 3$  卷积层、LeakyReLU 激活函数以及 CBAM 注意力模块以及用于重建的  $3 \times 3$  卷积层组成的下降分支得到 LR' 和 LR''，分别与超分的中间结果 SR' 和原始输入 LR 相比较、计算对偶损失(Dual Loss)。上述过程如下所示：

$$\begin{aligned}
 x_{down} &= f_{CBAM} \left( \sigma \left( f_{conv} \left( I_{SR} \right) \right) \right) \\
 I_{LR'} &= f_{re} \left( x_{down} \right) \\
 x_{down'} &= f_{CBAM} \left( \sigma \left( f_{conv} \left( x_{down} \right) \right) \right) \\
 I_{LR''} &= f_{re} \left( x_{down'} \right)
 \end{aligned}
 \tag{3}$$

在公式(3)中,  $I_{SR}$  表示重建的高分辨率红外图像,  $f_{conv}(\cdot)$  表示  $3 \times 3$  卷积层,  $\sigma(\cdot)$  表示激活函数,  $f_{CBAM}(\cdot)$  表示 CBAM 模块,  $x_{down}$  和  $x_{down'}$  分别表示经过下采样但还未重建成图像的中间变量,  $f_{re}(\cdot)$  负责图像的重建。

带有残差结构的 CBAM 网络结构图如图 2 所示, 它主要由一个通道注意力机制模块和空间注意力机制模块以及一些卷积层依次排列, 并在外侧添加一个残差连接构成。通道注意力机制同时使用了最大池化和平均池化操作, 分别经过几个多层感知机(Multilayer Perceptron, MLP)层后进行元素层面的相加, 最后应用 sigmoid 激活函数, 得到通道注意力结果。通道注意力本质上是根据不同通道的重要程度赋予其不同的权重, 反应的是全局的特征。该过程可以用公式(3)表示:

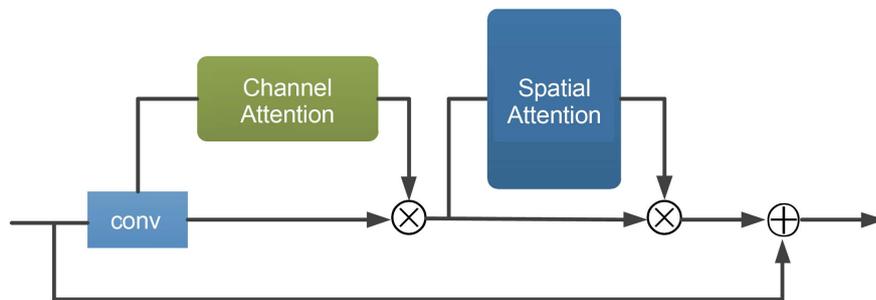


Figure 2. The architecture of CBAM with residual branch  
图 2. 带有残差结构的 CBAM 模块结构图

$$x_c = \sigma(MLP(AvgPool(x))) + \sigma(MLP(MaxPool(x))) \quad (4)$$

其中  $x$  为输入的特征图,  $x_c$  为输出的特征图,  $AvgPool(\cdot)$  和  $MaxPool(\cdot)$  分别表示平均池化和最大池化操作。 $MLP(\cdot)$  表示多层感知机,  $\sigma(\cdot)$  表示 sigmoid 激活函数。

空间注意力机制与通道注意力机制类似, 也需要同时对输入的特征图采用最大池化和平均池化操作, 但空间注意力是沿着通道方向进行计算, 将张量加权压缩成二维的特征图, 反应的是空间不同局部的重要程度, 该过程可以用公式(4)表示:

$$x_s = \sigma(f^{7 \times 7}([AvgPool(x); MaxPool(x)])) \quad (5)$$

其中,  $f^{7 \times 7}(\cdot)$  表示  $7 \times 7$  的卷积层,  $[AvgPool(x); MaxPool(x)]$  表示将平均池化和最大池化操作的结果沿着通道方向进行拼接。

### 3. 实验及结果分析

#### 3.1. 数据采集与模型训练

在实验过程中, 所使用的训练数据集  $\{(LR_k, HR_k)\}_{k=1}^K$  是由  $K = 6558$  个 LR/HR 图像对组成的。其中, HR 图像(分辨率为  $640 \times 512$ , 8-bit)是来源于由 FLIR 提供的开源红外图像数据集。这些图像拍摄于美国加利福尼亚州 Santa Barbara 市, 拍摄场景包括城市街道、高速公路等, 其中有 60% 是在白天拍摄, 40% 为夜间拍摄。为了生成图像对, 我们使用了 MATLAB 的双三次插值函数对这些红外图像进行 4 倍下采样, 生成了对应的 LR 图像(分辨率为  $160 \times 128$ , 8-bit)。此外, 我们按上述方法, 准备了 1355 组 LR/HR 图像对用于验证集。本网络模型是用 PyTorch 框架开发的(Python3.7.7 版本), 用于训练的优化器是 ADAM (Adaptive Moment Estimation), 损失函数如公式所示:

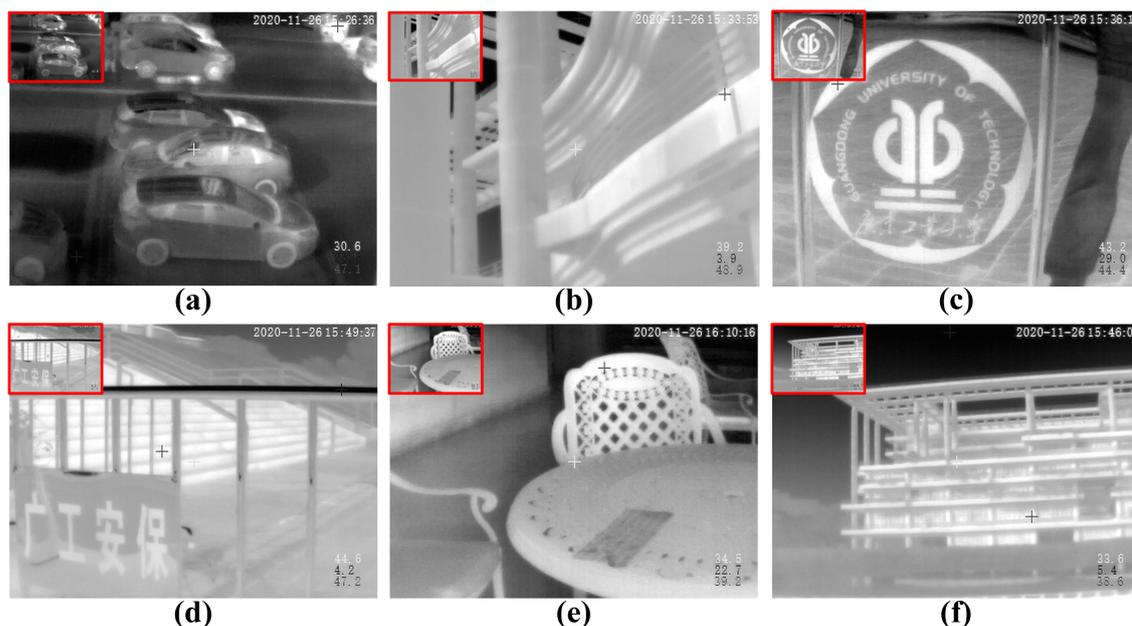
$$L(\Theta) = \sum_{k=1}^K \left( \|I_{HR} - I_{SR}\|^2 + \|I_{SR'} - I_{LR}\|^2 + \|I_{LR} - I_{LR'}\|^2 \right) \quad (6)$$

我们对所提出的网络模型训练了 100 epoch, batch size 设置为 128。在实验过程中, 学习率(learning rate, LR)随着训练损失的下降而下降。训练初始时的 LR 为  $10^{-3}$ , 随着训练损失的下降会逐渐下降到  $10^{-4}$ , 在 50 个 epoch 之后, LR 下降到了  $10^{-6}$  并持续到训练结束。

为了验证算法具有更好的泛化能力和真实场景下重建能力, 我们用自己的红外成像设备于白天在校园里拍摄了 73 张红外图像用作测试集, 并采用了峰值信噪比(Peak Signal to Noise Ratio, PSNR)这一客观评价指标以及 Variance、Entropy 等主观评价指标对我们所提出的方法及对比方法进行了验证。我们的成像系统采用的是由 XINFOO Technology Co., Ltd.提供的 X1320EC 红外传感器, 能够得到  $320 \times 240$  分辨率的红外图像。

### 3.2. 实验结果

对  $320 \times 240$  的低分辨率红外图像进行 4 倍超分辨率重建, 得到  $1280 \times 960$  分辨率的结果如图 3 所示。其中每组图像左上角红框中的为 LR, 大图为重建之后的 HR。可以看出, 由我们方法得到的 HR 在放大分辨率的同时, 依然保持了丰富的细节和清晰的边、角等高频信息。

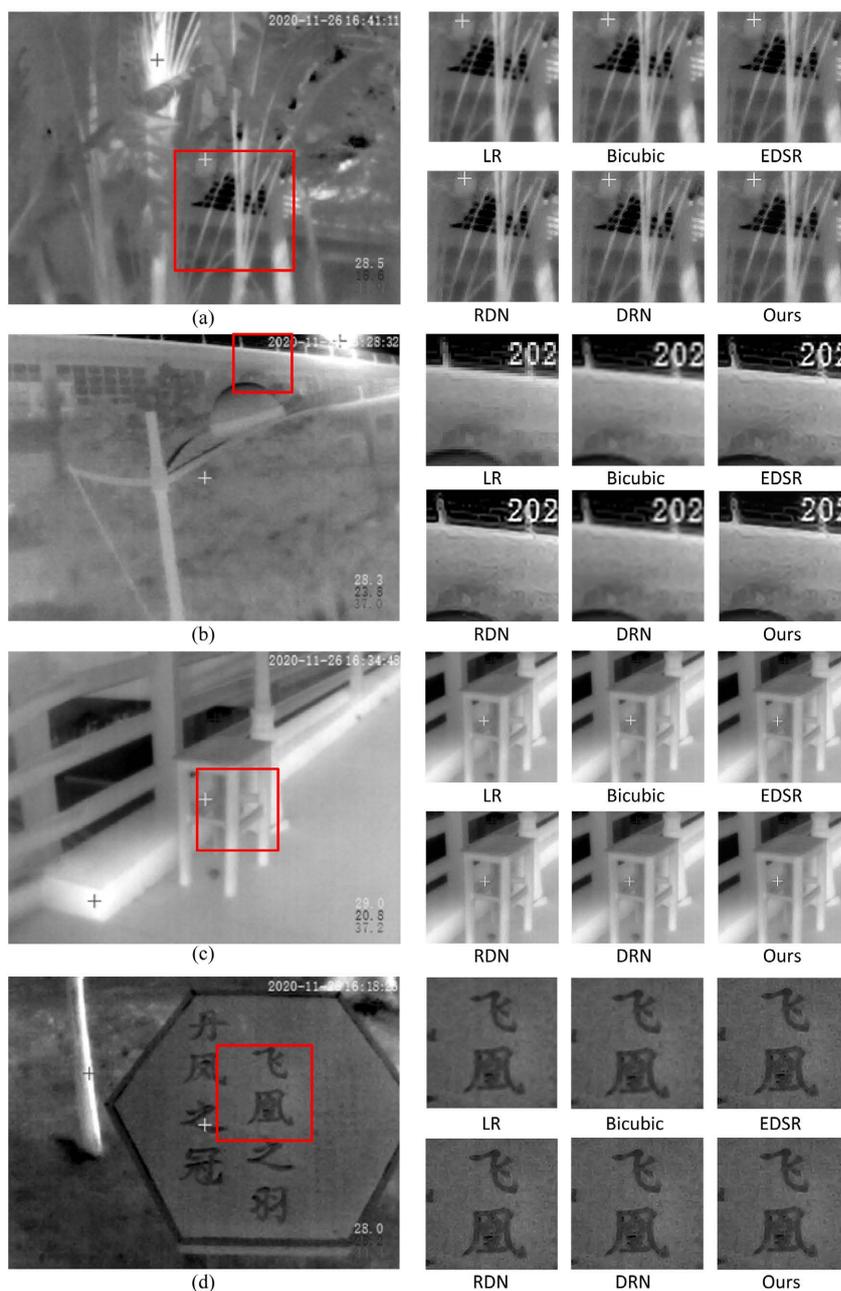


**Figure 3.** Visual effects of 4× super resolution operations using the experimental data  
**图 3.** 采用实验数据 4 倍超分辨率重建的视觉效果

我们所提方法与其他方法的视觉效果对比如图 4 所示。其中, 左边的图像是采用我们的方法进行超分辨率重建的结果, 右边的小图是对红框中的局部细节进行放大, 并与其他现有方法进行效果上的比较。

表 1 为我们所提出的方法与 bicubic、EDSR 等其他现有方法在主客观评价指标上的比较。从评价指标上看, 我们的方法与现有方法比性能接近, 且部分指标上更加领先。

现有的 EDSR 和 RDN 方法尽管在部分评价指标上略微占据优势, 但其参数量过于庞大, 非常不利于在嵌入式的小型红外成像设备上部署和应用。我们的方法在网络模型的参数量上具有很大的优势。相关数据对比如表 2 所示。



**Figure 4.** Visual effect comparison with existing methods  
**图 4.** 与现有方法的视觉效果对比

**Table 1.** Quantitative evaluation of 4 existed SR algorithm and our proposed method  
**表 1.** 我们的方法和对比方法在 4 倍超分结果下的评价指标

评价函数	Bicubic	EDSR	RDN	DRN	Ours
Variance	35.422	35.888	35.885	34.140	35.632
Range	12.223	12.210	12.217	12.221	12.288
Entropy	0.397	0.424	0.454	0.039	0.449
PSNR	/	30.847	31.109	31.590	31.240

**Table 2.** Model parameters**表 2.** 模型参数量

方法	参数量
Bicubic	/
EDSR	43.081 M
RDN	22.269 M
DRN	4.818 M
Ours	0.213 M

## 4. 结论

综上所述, 我们提出并实验证明了一种基于对偶循环结构和注意力机制的红外图像超分辨率方法, 它可以将低分辨率红外图像转化为具有丰富细节和锐利边缘的高分辨率红外图像。所提方法不仅可以学习从 LR 到 HR 的非线性映射, 还可以学习从 HR 到 LR 的对偶映射, 以辅助超分辨率图像处理。通过引入多通道融合注意力机制 CBAM, 我们的模型可以用很少的参数提取 LR 图像的特征。通过比较双三次、EDSR、RDN 和 DRN 的可视化结果和评价指标, 可以发现我们提出的方法不仅比其他方法具有更好的性能, 而且参数更少, 更轻量化, 具有更好的实用性。

## 基金项目

国家自然科学基金(61805048, U2001201, U1801263, U1701262); 广东省信息物理融合重点实验室(2016B030301008); 广东省自然科学基金项目(2018A030310599)。

## 参考文献

- [1] 王岭雪, 蔡毅. 红外成像光学系统进展与展望[J]. 红外技术, 2019, 41(1): 1-12.
- [2] 莫朝霞, 陈沅江. 我国红外热像检测技术的研究及发展展望[J]. 激光与红外, 2014, 44(12): 1300-1305.
- [3] 汪金礼, 王皓, 袁宝红, 刘鸿涛, 苏亚辉, 尹涛. 红外视频监控中的人体目标检测[J]. 测试技术学报, 2014, 28(1): 39-43.
- [4] Ragb, H. and Asari, V. (2019) Multifeature Fusion for Robust Human Detection in Thermal Infrared Imagery. *Optical Engineering*, **58**, Article ID: 043101. <https://doi.org/10.1117/1.OE.58.4.043101>
- [5] Zhou, Z., et al. (2016) Fusion of Infrared and Visible Images for Night-Vision Context Enhancement. *Applied Optics*, **55**, 6480-6490. <https://doi.org/10.1364/AO.55.006480>
- [6] Zafar, I., et al. (2010) Human Silhouette Extraction on FPGAs for Infrared Night Vision Military Surveillance. 2010 *Second Pacific-Asia Conference on Circuits, Communications and System*, Beijing, 1-2 August 2010, 63-66. <https://doi.org/10.1109/PACCS.2010.5627025>
- [7] Chen, H.W., et al. (2021) Advanced Automatic Target Recognition (ATR) with Infrared (IR) Sensors. 2021 *IEEE Aerospace Conference* (50100), Big Sky, 6-13 March 2021, 1-13. <https://doi.org/10.1109/AERO50100.2021.9438143>
- [8] Kulchitsky, N.A., Naumov, A.V. and Startsev, V.V. (2020) Development Trends of Infrared Focal Plane Array Detectors. <https://doi.org/10.17587/nmst.22.500-510>
- [9] Liangpei, et al. (2010) A Super-Resolution Reconstruction Algorithm for Surveillance Images. *Signal Processing*, **90**, 848-859. <https://doi.org/10.1016/j.sigpro.2009.09.002>
- [10] Keys, R.G. (1982) Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **29**, 1153-1160. <https://doi.org/10.1109/TASSP.1981.1163711>
- [11] Thurnhofer and Stefan (1996) Edge-Enhanced Image Zooming. *Optical Engineering*, **35**, 1862-1870. <https://doi.org/10.1117/1.600619>
- [12] Dong, C., et al. (2014) Learning a Deep Convolutional Network for Image Super-Resolution. 13th *European Conference*, Zurich, 6-12 September 2014, 184-199. [https://doi.org/10.1007/978-3-319-10593-2\\_13](https://doi.org/10.1007/978-3-319-10593-2_13)

- 
- [13] Kim, J., Lee, J.K. and Lee, K.M. (2016) Accurate Image Super-Resolution Using Very Deep Convolutional Networks. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 1646-1654. <https://doi.org/10.1109/CVPR.2016.182>
- [14] Ledig, C., *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>
- [15] Wang, X., *et al.* (2018) Esrgan: Enhanced Super-Resolution Generative Adversarial Networks. 15th *European Conference on Computer Vision*, Munich, 8-14 September 2018, 1-16.
- [16] Zhang, Y., *et al.* (2018) Image Super-Resolution Using Very Deep Residual Channel Attention Networks. 15th *European Conference on Computer Vision*, Munich, 8-14 September 2018, 294-310. [https://doi.org/10.1007/978-3-030-01234-2\\_18](https://doi.org/10.1007/978-3-030-01234-2_18)
- [17] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer International Publishing, Berlin. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [18] Guo, Y., *et al.* (2020) Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution. *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 5406-5415. <https://doi.org/10.1109/CVPR42600.2020.00545>
- [19] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)