计算机科学
COMPUTER SCIENCE

融合多头注意力机制的图像降噪网络模型

李玥玥, 刘万平, 黄东

引用本文

李玥玥, 刘万平, 黄东. 融合多头注意力机制的图像降噪网络模型[J]. 计算机科学, 2023, 50(11A): 230100091-8.

LI Yueyue, LIU Wanping, HUANG Dong. Image Denoising Network Model Combined with Multi-head Attention Mechanism [J]. Computer Science, 2023, 50(11A): 230100091-8.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230100091

融合多头注意力机制的图像降噪网络模型

李玥玥[1]　刘万平[1]　黄　东[2]
1 重庆理工大学计算机科学与工程学院　重庆 400054
2 贵州大学现代制造技术教育部重点实验室　贵阳 550025
(lyy98599@2020.cqut.edu.cn)

摘　要　由于GPU计算的快速发展,深度学习近年来在图像降噪方面得到了应用。大多数深度学习方法都需要无噪声图像作为训练标签,但通常它们很难获得,甚至不可能获得。于是,有学者开始研究使用噪声图像进行降噪网络训练,但其恢复的图像却面临丢失细节信息的问题。受Noise2Noise(N2N)的思想启发,文中使用成对的噪声图像训练神经网络,学习同一范围的同类型噪声之间的分布关系,实现了一种新的降噪网络模型。新开发的模型(MA-UNet)基于经典UNet架构,融合了多头注意力机制(Multi-head Attention)和简易的残差网络,可以更好地挖掘图像的关键信息,掌握特征的全局信息,从而恢复更清晰的图像。与传统算法(CBM3D)和其他方法(如DnCNN和B2U)相比,MA-UNet的性能参数优良。从视觉图像观察来看,所提模型恢复了更清晰的图像细节。与N2N设计的模型相比,在不同噪声幅值下,所提模型在4个经典数据集上的峰值信噪比和结构相似性指数的均值均有显著提高。

关键词:深度学习;注意力机制;细节信息;图像降噪;全局特征
中图法分类号　TP391

Image Denoising Network Model Combined with Multi-head Attention Mechanism

LI Yueyue[1],LIU Wanping[1] and HUANG Dong[2]
1 College of Computer Science and Engineering,Chongqing University of Technology,Chongqing 400054,China
2 Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education,Guizhou University,Guiyang 550025,China

Abstract　Due to the rapid development of GPU computing,deep learning has been applied in image denoising recently. Most of the deep learning methods require noise-free images as training labels,but they are usually difficult or even impossible to obtain. Therefore,some scholars begin to study the use of noisy images for noise reduction network training,but the restored image is faced with the problem of losing details. Inspired by the idea of Noise2Noise(N2N),this paper uses pairs of noised images to train the neural network,to learn the distribution relationship between the same type of noise in the same range,and realize a new novel image denoising network model. The newly-developed model(MA-UNet) is based on the classic UNet architecture and combines the multi-head attention mechanism and simple residual network. It can capture the key information of the image,master the global information of the feature,so as to recover clearer images. Compared with the traditional algorithm CBM3D and other methods,such as DnCNN and B2U,MA-UNet has excellent performance in terms of parameters. Through the comparison of visual images,our model restores much clearer image details. Compared with the model designed by N2N,under different noise magnitude,the mean value of the peak signal-to-noise ratio and the structural similarity index of the proposed model on four classical data sets improve significantly.

Keywords　Deep learning,Attention mechanism,Detail information,Image denoising,Global feature

1　引言

信息化时代下,图像是一种重要的信息源,通过图像处理可以帮助人们了解信息的内涵[1]。然而,日常图像的噪声不可避免,它的存在严重影响了图像的质量,妨碍人们接收图像信息,因此图像降噪逐渐成为图像预处理的必要工序。

图像降噪即通过特定的算法对图像进行处理,从而抑制噪声,提高图像质量[2],这对图像后处理工作具有现实意义。传统上,研究人员多采用滤波算法研究图像降噪[3],主要是在空间域、时间域和频域上对图像像素特征进行处理。经典的非局部均值滤波(None-Local Means,NLM)[4]和BM3D[5],从邻域像素特征出发,注重图像降噪过程中的细节保留。CBM3D(针对彩色图像)[5]常被作为彩色图像降噪的基准,但传统算法在实际应用中计算量较大,运行速度慢。2008年Jain等[6]提出用CNN处理自然图像的降噪问题后,基于深度学习的图像降噪神经网络便层出不穷。Burger等用海量的业务数据训练多层感知机(Multilayer Perceptron,MLP)[7],实现了图像降噪网络。2016年,Zhang等以输出图像与噪声

基金项目:重庆市自然科学基金(cstc2021jcyj-msxmX0594);重庆理工大学研究生创新项目资助(gzlcx20223212)
This work was supported by the Natural Science Foundation of Chongqing,China(cstc2021jcyj-msxmX0594) and Graduate Student Innovation Program of Chongqing University of Technology(gzlcx20223212).
通信作者:刘万平(wpliu@cqut.edu.cn)

图像的 L_2 范数为损失函数训练更深层次的降噪卷积网络(Deep Convolutional Neural Network, DnCNN)[8]，该方法一定程度上消减了深层次网络训练带来的梯度色散效应。然而，这类深层次的降噪网络易忽略图像的多尺度特性，恢复出来的图像存在边缘模糊的现象。在此情形下，注意力机制进入了人们的视野，谷歌在 2017 年提出了具有里程碑意义的模型 Transformer[9]，其依赖于注意力机制，在图像特征提取时，模型可以一步到位获取特征的全局和局部的关系[10]。Bai 等[11]在研究上下文感知的图像到图像的转换时，在特征注意力模块中加入了现有图像翻译架构。2020 年，Tian 等提出了一种注意力导向的降噪卷积神经网络(Attention-oriented Denoising Convolutional Neural Network, ADNet)[12]，利用稀疏机制、特征增强机制和 Attention 机制，在小网络复杂度的情况下提取显著性特征，进而移除复杂图像背景中的噪声，验证了注意力机制在图像降噪网络研究中的有效性。

这类端到端的神经网络，很大程度上依赖于训练数据集的质量和数量，但是现实世界的干净图像难以获取甚至无法获取。为了克服图像降噪领域的干净图像训练样本稀缺的问题，以及现有的深层次降噪网络易丢失图像特征的边缘细节的问题，本文深入结合 NVIDIA 团队提出的 N2N 图像对[13]训练降噪网络的方法，融合注意力机制设计了以真实图像为训练样本的 MA-UNet(Multi-Head Attention Module UNet)。模型中设计了轻量且通用的多头注意力模块，可有效提取图像的全局和局部信息；加入简易残差块，能够克服深层网络与浅层网络间产生的语义歧义，融合了图像更细节的特征信息，从而实现了噪声信号到干净信号的重建。本文的其主要贡献如下：

1)基于 N2N 思想，提出了一个由 UNet[14]改进的图像降噪网络，降噪效果能达到当前研究的先进水平；

2)设计降噪网络的编码器端时，在模型瓶颈处加入了多头注意力机制，更多地捕捉到图像特征的全局信息，恢复的图像细节更清晰；

3)在 MA-UNet 的解码器端，设计了简易残差块，更好地解决了深层次的降噪网络梯度消失的问题，网络训练收敛更快，各性能参数均有提升。

2 本文方法

2.1 N2N 的思想

N2N 是 2018 年 NVIDIA 与 MIT 共同研究的算法，其建立在统计学原理上，由室内温度估计问题得到启发。假设有一组当前观测的室内温度值($y_1, y_2, \cdots$)，以此估计室内的真实温度，可以有如下建模策略：

$$\arg\min_z E_y\{L(z,y)\} \tag{1}$$

为了找到一组室内温度观测值的最小平均偏差数 z，考虑用 L_2，L_1 和 L_0 3 种函数来优化建模，由此会得到不同的最小化误差函数的值 z。在解决图像降噪问题的神经网络训练中，该统计学原理则转换为一组“输入-目标图像对(x_i, y_i)”的典型训练任务的形式，若将神经网络回归模型看作点估计过程，网络目标函数则为：

$$z = f_\theta(x) \tag{2}$$

网络函数由 θ 参数化，f_θ 为参数映射族神经网络。

$$\arg\min_\theta E_{(x,y)}\{L(f_\theta(x),y)\} \tag{3}$$

即利用 z 值去估计不同输入图像 x 所对应的 y，L 表示衡量观测值 y 与目标值 z 偏差的损失函数。

实验表明，若用期望值作为网络拟合的数据，拟合的效果是可观的。因此，如果输入数据满足条件分布 $p(x|y)$ 被具有相同条件期望值的任意分布替换，最佳的网络参数 θ 也会保持不变，即目标函数变为：

$$\arg\min_\theta \sum^{i} L(f_\theta(\hat{x}),(\hat{y}_i) \tag{4}$$

这意味着可以在不改变网络训练结果的情况下将神经网络的训练目标做较小的增减变化。在图像降噪问题中，网络训练的目标值 y 可变为添加或减去均值为零的噪声目标值 $\hat{y}$。

2.2 基于 Multi-Head Attention 的图像特征提取模块

由于人类视觉的特殊性，前人开始研究注意力机制，以解决信息提取的瓶颈。注意力机制最早出现于自然语言处理中，2017 年，基于注意力的编码器-解码器模型 Transformer 的问世，彻底改变了自然语言处理领域的研究。随着深度学习领域的发展，视觉处理中对注意力机制的应用不断增加。比如，Momenta 公司提出了 SENet(Squeeze-and-Excitation Networks)[15]，强调关注图像特征通道之间的相互依赖关系。2018 年，Woo 等提出了 CBAM[16]，将注意力过程分为两个独立的部分，关注不同通道之间的图像特征信息以及同一通道不同位置的像素特征，有效地帮助信息在网络中流动。2020 年，谷歌的团队也将标准的 Transformer 模型直接迁移至图像领域变成 Vision Transformer 模型[17](ViT)，其中关键的多头注意力模块开始应用于 CV 领域，图 1 给出了本文应用到模型中的多头注意力模块结构。

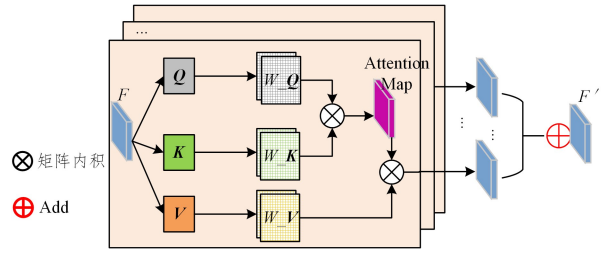


图 1 Multi-Head 模块

Fig. 1 Multi-Head module

首先将下采样后的特征图 $F \in R^{H\times W\times C}$ 输入 Multi-Head 模块，再将特征图 F 随机映射得到 $\boldsymbol{Q} \in R^{d_Q}$，$\boldsymbol{K} \in R^{d_K}$ 和 $\boldsymbol{V} \in R^{d_V}$ 3 个参数矩阵向量，做相应的矩阵映射乘积后，利用全连接层(Dense)提取他们之间的关联性特征。然后根据模型设定的注意力头的个数(这里取 8)，通过重塑(Reshape)操作对矩阵向量做拆解变换，完成 $W_i^{(Q)}\boldsymbol{Q}$，$W_i^{(K)}\boldsymbol{K}$ 和 $W_i^{(V)}\boldsymbol{V}$ 的并行计算。每个注意力头 H_i 可由式(5)计算。

$$H_i = f(W_i^{(Q)}\boldsymbol{Q}, W_i^{(K)}\boldsymbol{K}, W_i^{(V)}\boldsymbol{V}) \tag{5}$$

其中，f 表示缩放点积注意力计算，由此推广出每个头的自注意力模型为：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_K}}\right)\boldsymbol{V} \tag{6}$$

这样每个头都会关注输入的不同部分，可以表示比简单加权平均值更复杂的图像特征。在 n 个投影空间中分别应用自注意力模型，对多个 Attention 加权求和，得到多个头的交互信息结果 H_n，则多头注意力模型表示为：

$$MultiHead(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = W_o[H_1; H_2; \cdots; H_n] \tag{7}$$

其中，$\boldsymbol{W}_O \in R^{d_V}$ 为输出投影矩阵，是最后利用 $\boldsymbol{V}$ 矩阵加权求和后输出的。模型利用 Dense 层聚合了提取的特征关联信息，

再经过 Reshape 操作后,输出了经过多头注意力机制调整后的特征图 F',继续向下传递到网络的深层次结构中,增强了 MA-UNet 的网络表征能力,抑制了图像噪声特征的效用。

本文设计多头注意力模块的思路是,在图像输入神经网络后,经过了简单的卷积操作,先保留住全局信息,然后在 MA-UNet 编码器的网络瓶颈处分别逐层加入 4 个 Multi-Head 模块,捕获图像像素间的局部和全局特征的关联信息。以此增强网络模型的图像特征提取能力,将图像特征信息更全面地传递到解码器,从而有助于解码器恢复出更清晰的图像。

3 网络模型的设计

3.1 提出的图像降噪网络模型

本文模型是基于 N2N 中的 UNet 改进的,整体上加入了多头注意力机制和简易的残差模块,并调整了网络的深度和卷积核大小。著名的 Vision Transformer 主要是将图像视为一系列 patch 来处理特征信息,在各种视觉任务上实现了令人难以置信的性能。其中的多头注意力模块将 patch 嵌入(特征向量)转换为查询、键和值,对嵌入序列中每个元素之间的关系进行建模学习。Fei 等[18]利用 multi-head 模块从 CNN 提取的特征中生成空间注意力图,提升了人脸检测模型学习特征的泛化性。因此,利用深度学习处理图像降噪,需要关注图像像素级别的关联性特征,进一步强化网络的学习能力。本文将 Transformer 编码器中的多头注意力模块引入图像降噪网络模型中,在神经网络对图像下采样时,能够获取更丰富的信息,从而扩展模型专注于各图像像素特征的能力。

本文在模型编码器的网络瓶颈处设计了多头注意力机制模块,对传入的特征图进行运算处理,捕捉图像特征的交互信息,在图像特征传递过程中强调了图像的关键信息,有效克服了图像特征传递过程中细节信息丢失的问题。在解码器部分,加入了简易的残差模块[19],并与浅层的网络模块相接,加深网络深度的同时,消除了浅层网络和深层次网络的语义歧义,提升了图像恢复的精度,从而增强图像降噪的效果。由于该模型是应用于图像降噪领域,若使用 sigmoid 函数,将导致像素值被压缩到一定程度,与原始图像产生较大差异。因此,MA-UNet 的所有卷积层均由整流线性单元(ReLu)函数[20]激活,并由 BatchNormalization[21]层归一化处理。

如图 2 所示,新模型中共 4 个多头注意力模块,每块都对下采样的特征图做注意力值计算,加权后再融合到特征图中,并将调整后的特征图传递到深层次的网络结构中。在多头注意力模块中,将输入特征图随机映射的 $\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$ 做线性变换后,再根据注意力头的个数进行拆解变换,实现多个头并行计算。

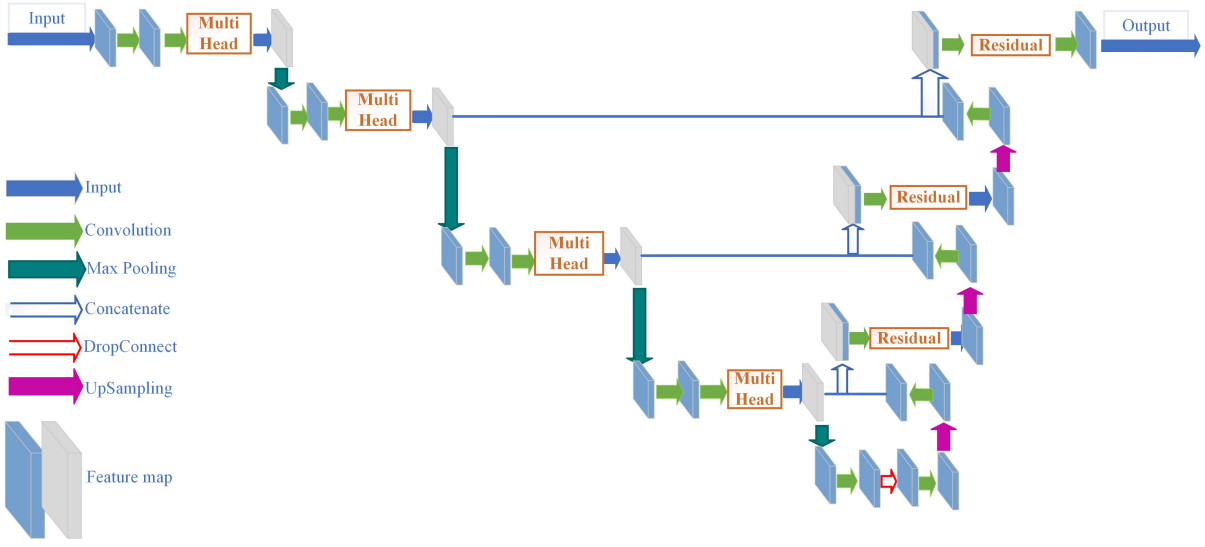


图 2 MA-UNet 的模型结构图

Fig. 2 Model structure of MA-UNet

简易残差块作为直接的信息传递通道,主要包括 2 个 3×3 的卷积块,2 个由归一化和 ReLu 函数组成的激活函数模块,如图 3 所示。在 MA-UNet 的解码器部分,上采样的特征图输入后,先经过归一化和激活处理,再通过卷积块进行权重共享,然后经激活模块后,将浅层特征和深层次特征联系起来,保证了信息传递,更好地解决了网络梯度消失的问题。

新模型将浅层网络的多头注意力模块和深层次网络的残差块通过向量拼接的方式连接,替换了传统 UNet 中简单的短连接,当下采样的特征信息向深层次网络移动时,简易残差块将特征图上的信息进一步融合到深层次结构中,再传入解码器。MA-UNet 中,设计了 4 个多头注意力模块,3 个简易残差块,滤波器的数量代表了编码器和解码器间的特征映射数量。下采样时,个数从 64 开始,以 2 的幂次方增长;上采样时,又以 2 的幂次方下降恢复到原始输入特征图的大小,由此完成图像特征的传递学习过程。

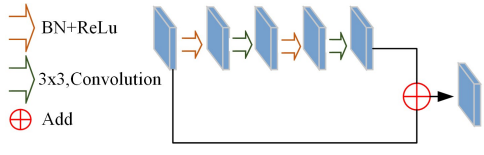


图 3 MA-UNet 中的简易残差块

Fig. 3 Simple residual block in MA-UNet

3.2 损失函数

如前所述,本文模型基于 N2N 的思想,输入图像 x 将在给定范围内生成随机的噪声图像对,作为模型训练的输入和输出。因此,不同的噪声采用对应的函数去衡量模型训练的损失。

加性高斯噪声具有零均值特性,一般把受到干扰的图像建模为 $y=x+n$,假设 n 满足零均值正态高斯分布 $N(0,\sigma^2)$[22]。标准偏差 σ 反映图像受噪声干扰的严重程度,称为

噪声水平变量[23]。因此,针对高斯噪声图像降噪,采用 L_2 作为模型训练的损失函数理论上更为合理[24]。

随机脉冲噪声由时间上随机产生的大量起伏骚扰积累而形成[25],本文为了模拟随机脉冲噪声图像,假定图像每个像素点都有 p 的概率被 [0,1]的值所替换。理想情况下,网络的输出为图像像素分布的众数。因此针对该类噪声,恰当地采取退火版本的损失函数 L_0 来降低损失。

我们假设图像中的噪声覆盖率为 p,而 $1-p$ 的概率会保留原始图像的像素。在这种情况下,寻找噪声以消除并获得更接近原始图像的图像,整个过程类似于模式搜索。因此本文考虑近似 L_0 来处理随机脉冲噪声图像,计算式如式(8)所示:

$$L_0 \approx (|f_\theta(\hat{x})-\hat{y}|+\varepsilon)^\gamma \tag{8}$$

其中,$\varepsilon=10^{-8}$,γ 在训练期间从 2～0 线性退化,这里 $z=f_\theta(\hat{x})$,以最大概率估计出真实图像与噪声图像的偏差,使得网络模型拟合出最接近真实像素值的图像。

3.3 算法及实现流程

算法 1 MA-UNet 模型训练流程

输入:真实图像 x

输出:噪声图像 $\hat{y}$

1. 输入真实图像 x,随机产生噪声图像对 $\hat{x}$,$\hat{y}$
2. 初始化参数 $\gamma=2$,$\varepsilon=10^{-8}$,迭代的总次数 T,学习率 lr,最小损失值 l_{min},得到最小损失值的迭代次数 c_{min}
3. 计算损失初值 l_0
4. if L_0,$l_0 \approx (|f_\theta(\hat{x})-\hat{y}|+\varepsilon)^\gamma$
5. else L_2,$l_0=(f_\theta(\hat{x})-\hat{y})^2$
6. $l_{min}=l_0$,$c_{min}=0$
7. for t=1 to T do
8. 降噪网络 f_θ 学习:

$$\arg\min_\theta \sum_i L(f_\theta(\hat{x}_i),(\hat{y}_i)) \rightarrow f_\theta(\hat{x}_i)$$

9 if L_0,则更新 γ:$\gamma=2.0*\frac{T-t}{T}$

10. 第 t 次训练得到的损失值 l_t。
11. if $l_t<l_{min}$

$l_{min}=l_t$,$c_{min}=t$

12. else if $t-c_{min}>50$
13. 网络停止学习,保存模型参数
14. end for

4 实验

4.1 数据集和实验设置

训练数据集。本文使用了来自文献[26]的 200 幅图像和来自文献[3]的 91 幅图像共同混合打包,作为小批量的训练数据,使用了“Set14”[27]作为训练验证集。为了保证模型得到充分训练,我们将原有数据集裁剪后扩充成更大的数据集用以训练和验证。

测试数据集。为了进行基准测试,本文使用了 4 个经典的测试数据集,分别是常用作基准数据集的“Set5”[28]“Set14”,Huang 等[29]提供的具有丰富的纹理图像的数据集“Urban100”,以及 Timofte 等[30]使用的自然图像类的数据集“B100”。

训练参数。设计的网络模型深度为 55,训练使用的批量大小为 8,将初始的学习率 lr 设置为 0.001,利用 Adam 更新神经网络参数,训练的迭代次数为 500,每次迭代步长为 1000。

本文训练所有实验超过了 300 个周期(批处理大小为 8 的 1000 次迭代),从而达到了模型的饱和。实验后发现,学习率降低了 3 倍,模型在 300 个周期左右停止了学习,在 GPU GeForce RTX 2070 上的训练大约需要 24 h。

在同样的实验环境下,相比文献[31]提出的 SRResNet[31],本文模型的训练速度由 419 ms/step 提升到了 230 ms/step,整体上提升了一倍。为了实现 N2N 的降噪模型,本文模型在训练时需要输入一对图像对,这里采取模拟实验手法,比如针对高斯噪声,干净的输入数据 x 产生噪声水平 σ 在 0～50 范围内的噪声图像 y 作为模型输入,再随机生成另一个噪声图像 y'(噪声水平 σ 仍在 0～50 范围内)作为模型输出,从而学习它们的噪声分布的映射关系。

4.2 模型训练

为了验证本文方法的降噪性能,我们将 MA-UNet 与几种最先进的降噪方法进行了比较。主观上,通过视觉效果呈现差异性;客观上,利用峰值信噪比(Peak Signal to Noise Ratio,PSNR)、结构相似性指数(Structural Similarity Index,SSIM)两个指标来评价其降噪效果。PSNR 是峰值信号的能量与噪声的平均能量之比,计算 PSNR 首先需要计算真实图像和降噪后图像的均方误差(Mean Squared Error,MSE)。

$$MSE=\frac{1}{M\times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I(i,j)-K(i,j)]^2 \tag{9}$$

其中,M 和 N 分别代表的是图像的长和宽,$I(i,j)$ 和 $K(i,j)$ 分别代表的是待测图像与真实图像对应的各个像素点。计算出图像的均方误差后,再做图像的峰值信噪比计算。

$$PSNR=10\times\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{10}$$

其中,MAX_I 表示图像的最大像素值。

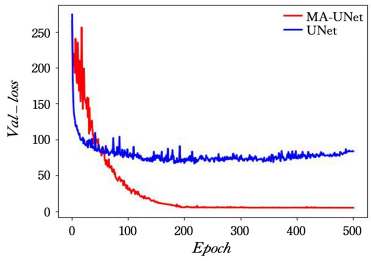
SSIM 是基于图片 x 与 y,分别从亮度 $l(x,y)$、对比度 $c(x,y)$、结构性 $s(x,y)$ 3 方面度量衡量两幅图像的相似度。一般计算中,SSIM 可以用式(11)表示:

$$SSIM(x,y)=l(x,y)\times c(x,y)\times s(x,y) \tag{11}$$

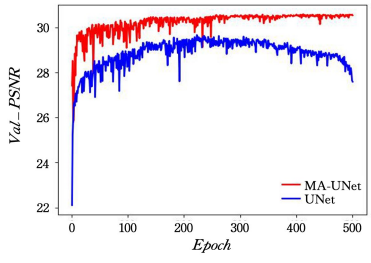
N2N 中设计了 UNet 与 SRResNet 两个模型处理不同的噪声图像,而本文模型是基于 UNet 进行修改,增加了当前较为新颖的注意力机制和简易的残差块,从而能够适应不同网络尺度的图像,提取到特征的全局信息,保留更多的图像细节。

N2N 论文中的 UNet 搭建了 35 层网络,参数共计约 31753387 个,训练速度约 305 ms/steps;SRResNet 搭建了 85 层网络进行训练,参数共计约 1231683 个,其训练速度约为419 ms/step。而本文的 MA-UNet 搭建了 55 层网络进行训练,参数共计约 31031875 个,训练速度提升到了 230 ms/step。图 4－图 7 为本文模型和 N2N 中设计的两个模型的训练过程的对比图。

从训练过程中可以看到,小批量数据集得到的随机损失值会逐步下降。首先,本文模型在处理高斯噪声时,设计的残差网络模块克服了 L_2 产生的极端误差,使得网络损失值迭代性地稳定下降,收敛趋势恰如其分。在同样的环境下训练 500 个 epochs,各个降噪模型基本达到饱和。SRResNet 训练得到的最佳 PSNR 为 30.22 dB,UNet 模型最优值为 29.66 dB,而本文的模型的最优 PSNR 则达到了 30.58 dB。



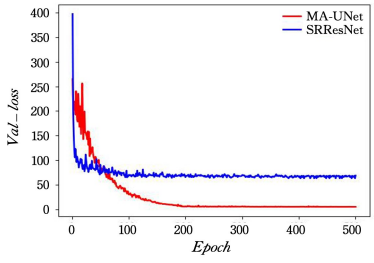
(a) 损失值的变化趋势



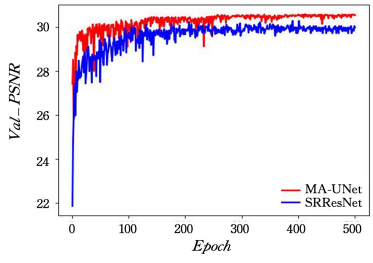
(b)PSNR 的变化趋势

图 4 MA-UNet 与 UNet 处理高斯噪声图像的训练过程对比

Fig. 4 Comparison of training process between MA-UNet and UNet in processing Gaussian noise image



(a) 损失值的变化趋势



(b)PSNR 的变化趋势

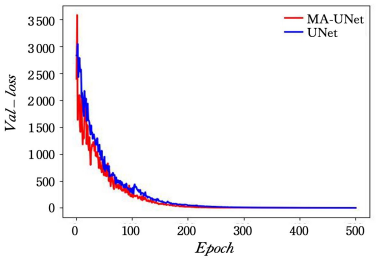
图 5 MA-UNet 与 SRResNet 处理高斯噪声图像的训练过程对比

Fig. 5 Comparison of training process between MA-UNet and SRResNet in processing Gaussian noise image

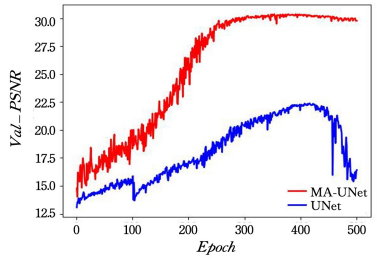
另外,我们还使用带随机脉冲噪声的输入和输出来训练网络,此处假定受损像素的概率分别从[0,0.95]中随机化,图 6 和图 7 给出了 70%的输入像素被随机化替换成噪声的这类随机脉冲噪声图像对的训练过程。

通过实验表明,多头注意力机制使得网络在学习过程中提取到了更丰富的图像全局特征,从数据集 Set14 验证结果来看,模型的降噪性能远超原模型。从客观数据来看,MA-UNet 相较于 UNet,PSNR 只有 22.39 dB,提升了约 38%,与 SRResNet 相比,PSNR 提升了约 28%。

本文采用相同的数据集,在同样的实验环境下,对 N2N 中的模型和本文模型进行训练。从图 4—图 7 可以看出,在训练高斯噪声图像时,本文模型通过残差模块融合了网络的浅层和深层次特征,加快了网络的迭代速度,损失网络训练的损失值更快地收敛并趋于稳定。在训练随机脉冲噪声图像时,多头注意力模块提取到了更精细化的图像特征,其 PNSR 均高于对比模型,且网络恢复出来的图像更加清晰。以上的网络训练过程对比图可以验证,本文模型的损失处理更具有鲁棒性,降噪成效也更显著。



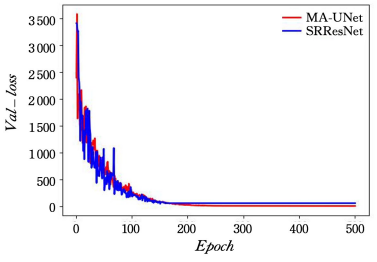
(a) 损失值的变化趋势



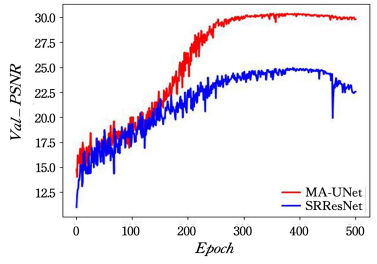
(b)PSNR 的变化趋势

图 6 MA-UNet 与 UNet 处理随机脉冲噪声图像的训练过程对比

Fig. 6 Comparison of the training process of MA-UNet and UNet processing random impulse noise image



(a) 损失值的变化趋势



(b)PSNR 的变化趋势

图 7 MA-UNet 与 SRResNet 处理随机脉冲噪声图像的训练过程对比

Fig. 7 Comparison of training process of MA-UNet and SRResNet in processing random impulse noise image

4.3 模型测试与对比

如前所述,丰富的数据集测试能够更进一步验证降噪效果和算法鲁棒性。为了验证 MA-UNet 在图像降噪方面的有效性,本文将模型在几个经典数据集上进行了实验。首先,展示本文模型测试的 Set5 中两张图片的效果,降噪后图像的 PSNR 值和 SSIM 值都体现出该模型的优良降噪性能。

图 8 给出了本文的模型对添加了 $\sigma=15$ 的高斯噪声图像进行降噪,PSNR 达到了 36.28 dB,SSIM 达到了 0.9740;而图 9 是模型对添加了 $\sigma=22$ 的随机脉冲噪声图像进行降噪,

PSNR 达到了 37.75 dB，SSIM 达到了 0.983 7，显示出模型强大的降噪性能。

图 8 MA-UNet 对高斯噪声图像的降噪效果

Fig. 8 Denoising effect of MA-UNet on Gaussian noise image

图 9 MA-UNet 对随机脉冲噪声图像的降噪效果

Fig. 9 Denoising effect of MA-UNet on random impulse noise image

MA-UNet 基于 N2N 中的 UNet 进行改进，不仅实现了对高斯噪声图像降噪，也实现了对随机脉冲噪声图像降噪，如图 10 所示。表 1 列出了 UNet 和本文模型 MA-UNet 对几种常用于测试的图像进行降噪的结果。

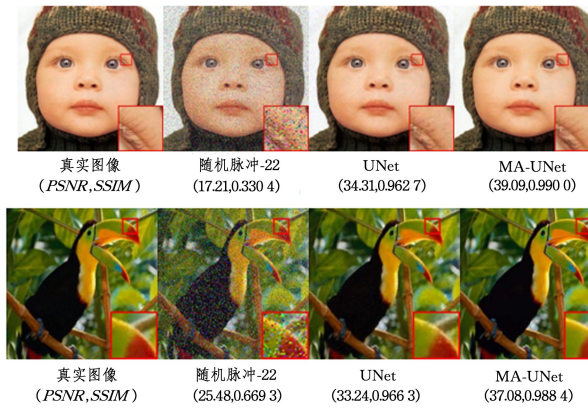


图 10 降噪性能(PSNR/SSIM)，随机脉冲噪声(σ=22)的降噪效果

Fig. 10 Denoising performance(PSNR/SSIM), random impulse noise-22

表 1 不同噪声水平下的 12 张广泛使用的图像得到的 PSNR/SSIM

Table 1 PSNR/SSIM of 12 widely used images at different noise levels

噪声水平		σ=15	σ=25	σ=50	σ=22	σ=70
图像	模型	*PSNR/SSIM*	*PSNR/SSIM*	*PSNR/SSIM*	*PSNR/SSIM*	*PSNR/SSIM*
羊	UNet	34.66/0.939 6	32.77/0.889 3	29.92/0.784 8	35.27/0.935 0	27.94/0.656 7
	MA-UNet	**37.45/0.950 7**	**34.63/0.903 5**	**31.57/0.809 2**	**39.21/0.978 5**	**28.75/0.748 5**
飞机	UNet	31.76/0.986 4	29.05/0.975 0	21.79/0.943 0	35.93/0.973 7	24.39/0.893 7
	MA-UNet	**33.52/0.986 1**	**32.79/0.978 8**	**33.64/0.970 7**	**38.86/0.990 2**	**28.66/0.949 0**
婴儿	UNet	33.68/0.973 3	33.09/0.955 5	28.93/0.913 4	34.40/0.964 2	26.71/0.835 2
	MA-UNet	**36.28/0.974 0**	**33.73/0.960 6**	**30.91/0.924 4**	**39.09/0.990 0**	**30.08/0.912 8**
鹦鹉	UNet	34.82/0.980 8	32.30/0.963 6	29.16/0.919 7	33.32/0.967 8	24.32/0.824 2
	MA-UNet	**37.22/0.964 9**	**34.13/0.968 2**	**30.22/0.933 1**	**37.08/0.988 4**	**26.61/0.888 6**
蝴蝶	UNet	37.93/0.976 8	36.21/0.968 1	31.98/0.934 5	33.06/0.951 8	25.40/0.781 2
	MA-UNet	**38.15/0.979 6**	**35.78/0.971 7**	**32.48/0.942 3**	**37.75/0.983 7**	**26.39/0.846 4**
头	UNet	32.43/0.847 1	31.40/0.836 7	28.69/0.735 4	31.13/0.821 5	26.58/0.697 1
	MA-UNet	**32.97/0.911 7**	**31.11/0.855 4**	**28.72/0.761 8**	**33.00/0.870 0**	**27.82/0.739 6**
女人	UNet	32.12/0.946 2	30.61/0.932 2	27.95/0.909 1	29.69/0.954 1	23.45/0.865 0
	MA-UNet	**33.26/0.980 1**	**32.30/0.965 0**	**28.86/0.938 5**	**33.16/0.986 3**	**23.93/0.879 6**
别墅	UNet	33.74/0.983 0	32.29/0.965 8	28.82/0.939 7	33.15/0.986 6	23.93/0.880 1
	MA-UNet	**35.06/0.980 0**	**31.87/0.973 3**	**28.04/0.936 9**	**31.63/0.990 0**	**22.79/0.827 7**
桥	UNet	35.52/0.991 3	31.90/0.972 3	28.03/0.935 9	31.64/0.979 9	22.76/0.828 0
	MA-UNet	**33.79/0.982 0**	**31.07/0.975 0**	**27.18/0.938 9**	**27.5/0.965 0**	**23.40/0.899 2**
大楼	UNet	33.71/0.976 2	31.04/0.975 5	27.14/0.939 4	27.59/0.965 2	20.45/0.798 9
	MA-UNet	**37.09/0.979 7**	**34.53/0.979 5**	**30.55/0.939 7**	**30.27/0.976 0**	**26.19/0.881 0**

从视觉效果和客观数据上，MA-UNet 相比 N2N 的 UNet 模型都达到了更优的降噪性能。另外，从研究该模型的现实意义出发，我们在各种数据集上将该模型与几种先进方法(CBM3D，DnCNN，B2U，SRResNet，UNet)进行了比较，展示了彩色图像在 3 个级别的高斯噪声(σ=15，25，50)的降噪效果。

表 2 中加粗的数据突出显示了不同 σ 值的最佳和次佳 PSNR 和 SSIM。

表 2 4 个经典数据集上的平均 PSNR/SSIM

Table 2 Average PSNR/SSIM on four classical data sets

	模型	PSNR/SSIM					
		CBM3D[4]	DnCNN[6]	B2U[32]	SRResNet[31]	UNet[14]	MA-UNet
σ=15	Set5	33.74/0.945 8	**35.92**/0.961 0	35.57/**0.956 4**	34.51/0.950 0	33.71/0.940 2	**36.02/0.961 3**
	Set14	32.51/0.949 7	32.44/0.950 0	**32.43/0.949 8**	32.50/0.948 5	32.55/0.949 6	**32.61/0.949 9**
	B100	28.58/0.942 5	34.78/0.963 0	**35.21/0.965 2**	35.18/0.962 5	32.99/0.957 0	**35.29/0.967 3**
	UrBan100	33.45/0.978 7	34.82/0.976 3	**34.70**/0.979 3	34.56/**0.978 8**	33.79/0.977 3	**34.90/0.981 2**
σ=25	Set5	31.59/0.928 6	33.02/0.934 4	32.58/**0.941 2**	**33.13**/0.935 6	32.65/0.927 6	**33.66/0.960 5**
	Set14	31.68/**0.931 0**	30.46/0.930 5	31.29/0.929 6	31.06/0.930 5	**31.96**/0.930 0	**33.18/0.935 6**
	B100	31.02/0.916 0	29.21/**0.937 0**	**33.69**/0.930 4	33.21/0.935 5	30.87/0.927 5	**34.10/0.958 2**
	UrBan100	30.97/0.963 3	**32.46**/0.966 6	32.32/0.966 8	32.08/0.966 3	31.70/**0.967 7**	**34.25/0.969 6**
σ=50	Set5	28.48/0.894 6	**30.01/0.896 2**	29.92/0.897 5	29.83/0.880 0	29.40/0.878 2	**31.52/0.902 9**
	Set14	29.64/0.891 3	27.18/0.890 5	**29.25/0.891 4**	27.10/0.879 5	29.29/**0.891 5**	**30.52/0.910 2**
	B100	28.83/0.862 5	26.83/0.885 5	**32.37/0.889 9**	30.25/0.872 5	25.80/0.859 5	**32.20/0.892 6**
	UrBan100	27.78/0/932 3	28.20/0.933 3	**28.86**/0.931 1	28.82/**0.932 3**	27.95/0.925 0	**28.87/0.946 3**

对于彩色图像降噪,本文提出的 MA-UNet 可以获得最高的峰值信噪比,这优于基准 CBM3D 和 DnCNN。例如,数据集 UrBan100 的测试结果显示,当 $\sigma=15$ 时,MA-UNet 的 PSNR 和 SSIM 的平均值比 CBM3D 分别高出 1.45 dB 和 0.0025。当 $\sigma=50$ 时,MA-UNet 的 PSNR 和 SSIM 的平均值相比 DnCNN 分别高出 0.67 dB 和 0.0130。这表明,该方法对低水平和高水平的噪声都具有更强的鲁棒性。

图 11 和图 12 给出了在 $\sigma=15$ 和 $\sigma=50$ 的高斯噪声下彩色图像的不同方法的降噪结果。从图中可以看出,本文设计的多头注意力机制模块在模型中效用明显,其降噪图像中的"机翼"边缘更清晰,像素点几乎没有噪声干扰,与原图的相似度极高。以上实验结果表明,本文提出的 MA-UNet 优于传统的降噪方法(如 CBM3D)和更先进的深度学习方法(如 DnCNN,B2U),在图像降噪方面更具鲁棒性和有效性,尤其在细节保持和对比度保持方面的表现更好。

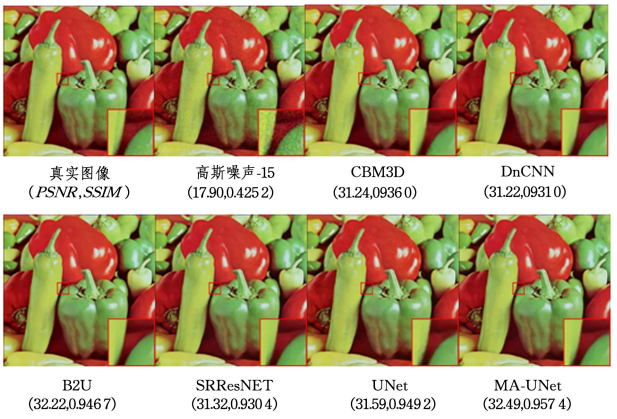


图 11 $\sigma=15$(高斯噪声)时不同方法的降噪效果

Fig. 11 Denoising performance of different methods when $\sigma=15$(Gaussian)

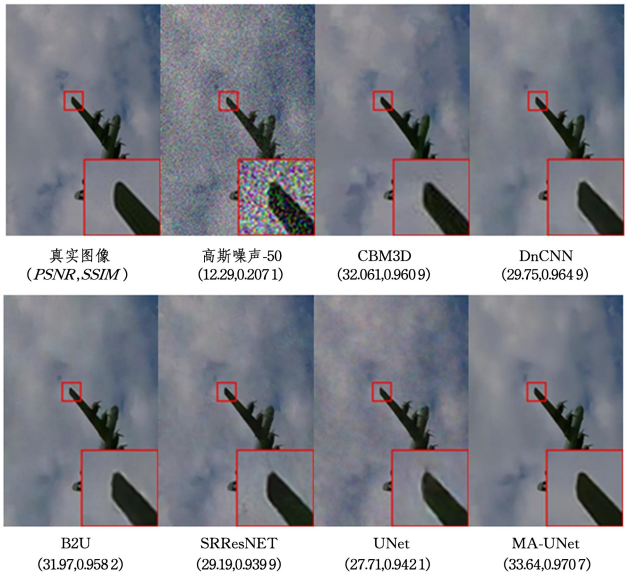


图 12 $\sigma=50$(高斯噪声)时不同方法的降噪效果

Fig. 12 Denoising performance of different methods when $\sigma=15$(Gaussian)

结束语 本文的主要贡献是结合多头注意力机制和简易残差块,设计了一个新的图像降噪网络。参考了 N2N 的思想,使用成对的噪声图片作为神经网络的输入和输出,并通过实验取得了理想的结果。原始的降噪网络尽管是改进后的 UNet,但模型设计仍存在网络瓶颈,在特征性信息提取后进行上采样时,生成的特征图丢失了图像的局部细节信息,且该网络结构不管是训练时间还是训练效果上都差强人意。本文提出的 MA-UNet,设计了多头注意力特征提取模块,在网络训练的瓶颈处更好地保留了图像的全局信息,在解码器中融入了简易残差结构,得到了更加丰富的图像特征,网络的训练时效和降噪结果都表现优良。

虽然本文模型可以有效地减少不同强度的噪声,但当噪声过高时,降噪效果会产生误差。比如信号淹没在噪声里的图像,无论什么样的模型,都很难处理该类图像降噪,恢复出干净信号的图像。因此,需要在实际场景中定义本文模型的应用边界。

参 考 文 献

[1] ZHOU Z H,ZHANG W T. Based on local structural similarity image denoising algorithm[C]// International Conference on Computer and Information Applications. 2013:3313-3317.

[2] MOHAN J,KRISHNAVENI V,GUO Y. A survey on the magnetic resonance image denoising methods[J]. Biomedical Signal Processing and Control,2014,9:56-69.

[3] YUE L,SHEN H,LI J,et al. Image super-resolution:The techniques,applications,and future[J]. Signal Processing,2016,128:389-408.

[4] BUADES A,COLL B,MOREL J M. A non-local algorithm for image denoising[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05). IEEE,2005,2:60-65.

[5] DABOV K,FOI A,KATKOVNIK V,et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactionson Image Processing,2007,16(8):2080-2095.

[6] JAIN V,SEUNG H S. Natural image denoising with convolutional networks[J]. Advances in Neural Information Processing Systems,2009,2:769-776.

[7] BURGER H C,SCHULER C J,HARMELING S. Image denoising:Can plain neural networks compete with BM3D? [C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,2012:2392-2399.

[8] ZHANG K,ZUO W,CHEN Y,et al. Beyond a gaussian denoiser:Residual learning of deep CNN for image denoising[J]. IEEE Transactions on Image Processing,2017,26(7):3142-3155.

[9] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017:5998-6008.

[10] LI R,ZHENG S,DUAN C,et al. Classification of hyperspectral image based on double-branch dual-attention mechanism network[J]. Remote Sensing,2020,12(3):582.

[11] BAI J,CHEN R,LIU M. Feature-attention module for context-aware image-to-image translation[J]. The Visual Computer,2020,36(10):2145-2159.

[12] TIAN C,XU Y,LI Z,et al. Attention-guided CNN for image denoising[J]. Neural Networks,2020,124:117-129.

[13] LEHTINEN J,MUNKBERG J,HASSELGREN J,et al. Noise2Noise:learning image restoration without clean data [C]// International Conference on Machine Learning. PMLR,

2018:2965-2974.
[14] RONNEBERGER O,FISCHER P,BROX T. U-net:Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham:Springer,2015:234-241.
[15] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7132-7141.
[16] WOO S,PARK J,LEE J Y,et al. CBAM:Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018:3-19.
[17] DOSOVITSKIY A,BEYER L,KOLESNIKOV A,et al. An image is worth 16x16 words:Transformers for image recognition at scale[C]// International Conference on Learning Representations. 2021.
[18] FEI J,DAI Y,WANG H,et al. Learning to mask:Towards generalized face forgery detection[J]. arXiv:2212.14309,2022.
[19] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
[20] GLOROT X,BORDES A,BENGIO Y. Deep sparse rectifier neural networks[C]// Proceedings of The Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings,2011:315-323.
[21] IOFFE S,SZEGEDY C. Batch normalization:Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. PMLR,2015:448-456.
[22] ULYANOV D,VEDALDI A,LEMPITSKY V. Deep image prior[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:9446-9454.
[23] XU S P,LI C X,LIN G X,et al. Fast image noise level estimation algorithm based on principal component analysis and deep neural network[J]. Acta Electronica Sinica,2019,47(2):274-281.
[24] XU L F,YU J M. Research on the influence of different optimizers on LR performance under Gaussian noise [J]. Computer Technology and Development,2020,30(3):7-12.
[25] XU S P,LIU T Y,LIN Z Y,et al. Learning a two-stage blind convolutional denoising model for the removal of random-valued impulse noise[J]. Chinese Journal ofComputers,2020,43(9):1673-1690.
[26] MARTIN D,FOWLKES C,TAL D,et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// IEEE International Conference on ComputerVision. IEEE,2001:416-423.
[27] ZEYDE R,ELAD M,PROTTER M. On Single Image Scale-Up Using Sparse-Representations[C]// International Conference on Curves and Surfaces. Berlin:Springer,2010:711-730.
[28] BEVILACQUA M,ROUMY A,GUILLEMOT C,et al. Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding[C]// British Machine Vision Conference. BMVA Press,2012,135:1-10.
[29] HUANG J B,SINGH A,AHUJA N. Single image super-resolution from transformed self-exemplars[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:5197-5206.
[30] TIMOFTE R,DE SMET V,VAN GOOL L. A+:Adjusted anchored neighborhood regression for fast super-resolution[C]// Asian Conference on Computer Vision. Cham:Springer,2015:111-126.
[31] LEDIG C,THEIS L,HUSZÁR F,et al. Photo-realistic single image super-resolution using a generative adversarial network [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:4681-4690.
[32] WANG Z,LIU J,LI G,et al. Blind2Unblind:Self-Supervised Image Denoising with Visible Blind Spots[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:2027-2036.

LI Yueyue, born in 1998, postgraduate. Her main research interests include deep learning and image denoising.

LIU Wanping, born in 1986, Ph.D, associate professor, master supervisor, is a member China Computer Federation. His main research interests include network and information security.