

基于 Unet 和 PPI 边缘信息的多光谱滤波器阵列图像去马赛克深度残差网络

摘要

多光谱图像因其能提供比传统 RGB 彩色图像更多的光谱信息从而能增强传统视觉任务的检测能力。例如，其在农作物检测、污水检测等方面都有广泛的应用。然而传统获取多光谱图像的方式因其速度慢、系统成本高、便携性差而受到批评。拜耳滤色器阵列 (CFA) 的因其便携性和低成本等优点在数码相机上的利用已经非常成熟。通过将 CFA 的概念扩展到多光谱滤波器阵列 (Multispectral filter array, MSFA) 上也可以避免传统多光谱成像系统的缺点。由于 MSFA 的空间采样率低以及多光谱相关性差缺陷的问题，使其解马赛克挑战变的困难。因此，提出一种高效的多光谱滤波阵列对于基于 MSFA 的多光谱成像非常重要。常规多光谱滤波器阵列是 CFA 在多光谱阵列上的扩展，一般是 4×4 模式，其中每个像素块只包含一个波段信息，因此 4×4 像素块中包含不同的 16 个波段信息。本文使用的是非常规多光谱阵列， 3×3 块中 9 个像素块只包含同一个波段信息，相邻 3×3 基本阵列中包含不同的波段信息，共 8 个光谱波段。本工作是对基于本文介绍的非常规多光谱阵列采样得到的 MSFA 图像的光谱信息和空间信息的恢复，称为多光谱去马赛克。本工作使用 Unet 的结构并使用 Res2Net 作为基本的残差块构建 MSFA 解马赛克网络 (Res2Unet)。此外为了提高网络利用光谱相关性的能力，向 Res2Net 基本块中加入了 SE 注意力机制。为了提高网络的边缘细节处理能力借鉴了基于伪全色图像 (pseudo-panchromatic image, PPI) 边缘注入模块。由于本工作是与深圳某企业合作的探索性任务，相关实验均在自建数据集上展开。实验结果显示与其他对比模型相比，本文提出的模型无论是在视觉效果还是指标性能上均有较高表现。本人实现的相关代码已上传至 Github: <https://github.com/morehire/demosaick.git>

关键词：多光谱图像去马赛克；多光谱滤波器阵列；深度学习；

1 引言

1.1 多光谱滤波器阵列

本文在摘要中提到一种非常规的多光谱滤波器阵列，在本节我们将介绍常规多光谱滤波器阵列和本文提出的滤波器阵列的区别。常规 MSFA 如图 1 (b) 所示，基本块模式为 4×4 ，包含 16 个波段的信息。本文所介绍的滤波器阵列如图 1 (a) 所示，基本块为 12×12 ，包含 8 波段信息。如果以基本块中包含的相同的波段个数与基本块空间分辨率的比值来定义信息压缩比的话，传统多光谱滤波器阵列信息压缩比为 $1/16$ ，而本文提出的多光谱滤波器阵列的压

缩比为 18/144，由于相邻像素之间可获取的其他频谱信息更为稀疏，可利用的频谱相关性更低，因此基于本文提出的多光谱滤波器阵列进行去马赛克难度更大。除了信息压缩比不同之外二者也有各自的优缺点：传统多光谱滤波器阵列因其紧凑的不同波段滤波器排列可以获得相邻像素之间更多的频谱相关性，且获得的空间信息也更多。但也正是由于这种不同波段滤波器的紧凑排列给制造上带来了一定的困难，因此其制造成本也更高；本文提出的一种滤波器排列在 3×3 块中只有同一波段的信息，在频谱尺度和空间尺度上的采集都更加稀疏。这无疑给后续利用神经网络获取波段间的相关性带来了更大的挑战，但是由于同一滤波器可制作尺寸更大，给工艺加工方面带来了便捷，也因此可以降低制造成本。

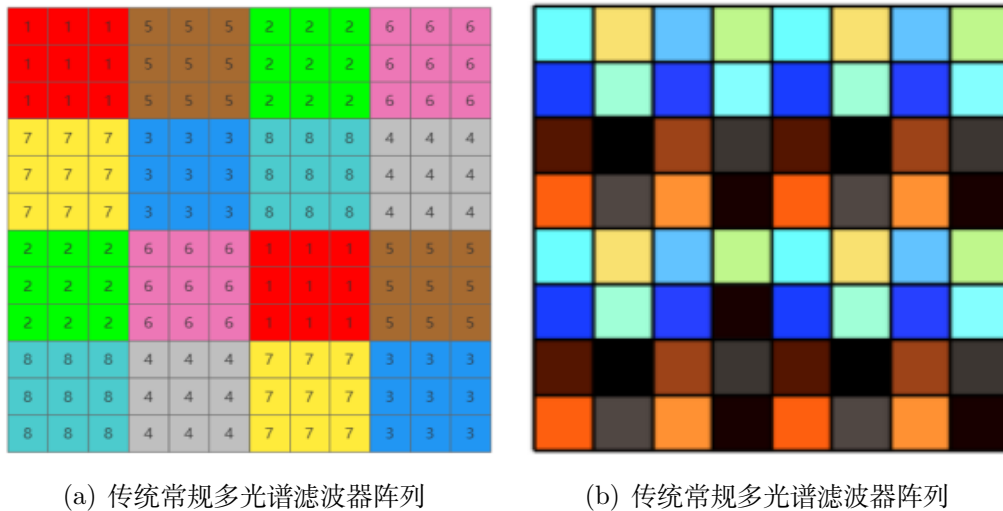


图 1. 两种不同滤波器阵列比较

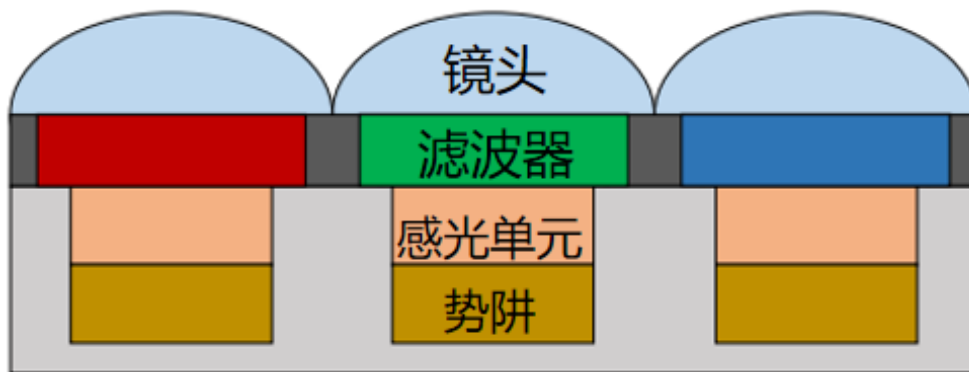


图 2. CFA 滤波器示意图

1.2 MSFA 图像去马赛克

无论是 RGB 图像拜耳滤波器阵列还是多光谱滤波器阵列，其作用位置都在相机的感光元件之上如图 2 所示，作用是只允许特定波长的光通过，以便于彩色成像。这种带有色彩滤波器相机拍摄出来的未经后处理的图像被称为 raw 图，其数据部分是一个单通道矩阵，像素色彩排列方式和滤波器在感光元件上的排列一致，如图 1 所示。多光谱图像去马赛克任务就是对相机在光谱域和空间域稀疏采样的 MSFA 图像进行信息重建，恢复出完整的具有高保真

的光谱信息和空间信息，其任务过程如图 3 所示。与传统 CFA 去马赛克问题不同，MSFA 传感器捕获的图像的高精度去马赛克是一项更具有挑战性的任务，主要原因有两个 [17]：首先，阵列光谱带的增加和主带的缺失表明空间采样率较低，这削弱了同一波段采样的相邻像素之间的空间相关性。稀疏采样的原始数据会降低重建多光谱图像的空间精度。其次，由于随着光谱进一步细分，光谱带变得更窄，邻近波段之间的光谱互相关性比远处波段之间的光谱互相关性更强，这会导致重建多光谱图像中不同通道之间的混叠问题。如果想要直接使用 raw 图来训练神经网络也是一件困难的事，相机的原始 raw 图不好获取，缺少相关的数据集。目前常规的做法是利用已经处理后的完整全波段图像，将其与掩码相乘，得到与滤波器排列相同的 MSFA 图像，此过程称为马赛克化，这样原始图像与马赛克化后的图像形成一个数据对，分别作为 ground truth 和训练样本。本工作是与深圳某企业合作的探索性任务，实际的多光谱滤波器并没有制作出来，因此无法得到原始相机的 raw 图。MSFA 图像也是通过将掩码与传统方式拍摄得到的全波段多光谱图像相乘的方式得到，其过程可描述为图 3 的逆过程。

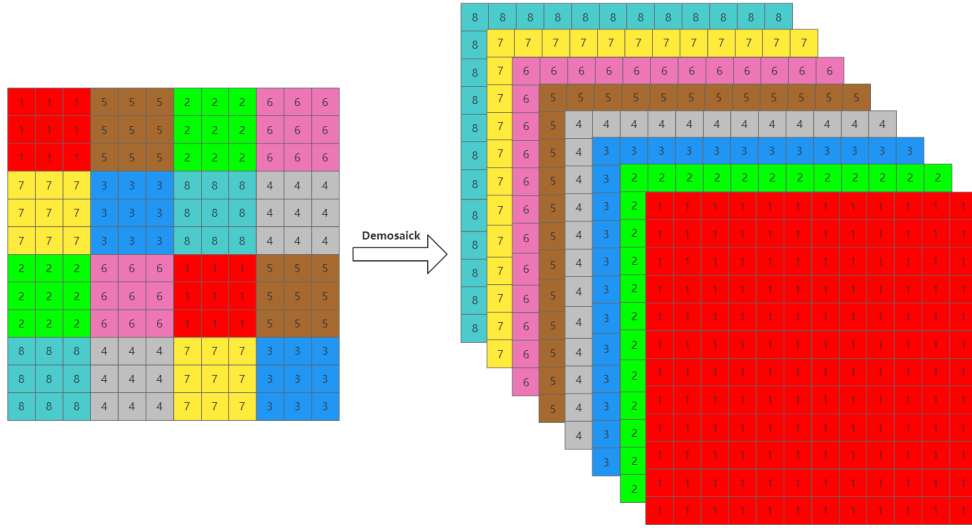


图 3. MSFA 图像去马赛克任务示意图

2 相关工作

多光谱图像去马赛克的任务已经被众多学者进行研究 [1] [3] [4] [11] [6]。当前的去马赛克方法可以分为基于插值的方法、基于矩阵分解/恢复的方法和深度学习方法。插值和基于矩阵的方法 [1] [3] 依赖于光谱空间先验来重建丢失的光谱和空间信息。本工作使用基于深度学习的方式进行去马赛克，因此重点关注最新的基于数据驱动的 MSFA 图像去马赛克方法。CNN 在各种低级图像处理任务中广泛使用，包括图像去模糊、图像去噪以及图像超分辨率。尽管 CNN 已被有效地应用于去马赛克 [13] [5]，但它们的应用于以绿色波段为主的拜耳滤镜。受到深度学习在传统 RGB 图像任务上的成功，越来越多的研究者提出使用基于数据学习的方式来实现 MSFA 图像去马赛克。Shinoda 等提出了一种基于残差网络 (ResNet) 和 3-D 卷积的深度去马赛克模型 [16]。Feng 等人提出了用于 MSFA 图像去马赛克的端到端马赛克卷积注意网络 (MCAN)，证明了对原始马赛克图像进行初始特征提取的重要意义 [7]。通过结合传统基于 PPI 的方法的优点，Liu 等人 [10] 提出了深度 PPI 去马赛克网络 (DPDNet)，证明 PPI 中的高频信息与重建的多光谱图像高度相关。然而，DPDNet 并没有给出从 PPI 中

准确提取高频信息的具体解决方案。Zhao 等人 [17] 受到初始特征和 PPI 中高频信息的重要性启发提出一种基于 PPI 边缘注入的残差网络来进行去马赛克，使用一个具有空间和光谱自适应权重共享策略的卷积模块来有效的提取马赛克初始特征，并利用从 PPI 中提取高频信息（例如，边缘相关信息）等来补偿网络对边缘信息处理的不足。相比于 RGB 拜耳滤镜模式下的去马赛克任务，光谱去马赛克需要 CNN 更多的利用多光谱相关性的表示才能更好的实现。U-Net [14] 以编码器—解码器为基础，将原始图像的特征编码提取后，再将特征进行解码还原为需要的信息，结构简单且有效。Res2Net 是 ResNet 基础上进行改进，将对特征图整体的一次 3×3 卷积换成对特征图分部分的多次残差 3×3 卷积，通过这种方式能够进一步提高网络对多尺度信息的提取能力 [8]。NTIRE 2022 光谱马赛克挑战 [2] 是 NTIRE 2022 相关挑战之一，其任务是从传统 MSFA 图像中重建出 16 个全波段（400-1000nm）的信息。赛事官方将参赛队伍的前 7 名使用的模型进行了总结，本工作的模型设计灵感来源于其中的一支队伍。基于以上工作的研究，本工作所提出的解马赛克网络如下：使用 Unet 作为网络的模型的主要架构，并以 Res2Net 作为网络的主要特征提取模块，在 Unet 每一层上下采样直接使用 PixelShuffle 和 PixelUnShuffle [15] 连接；为了提高网络对光谱相关性的学习能力，在 Res2Net 结构上添加 SE 注意力模块 [9]；由于 Unet 结构进行多次下采样、上采样导致细节确实，提出利用从 PPI 中提取高频信息（例如，边缘相关信息）等作为分支网络来补偿主干网络对边缘信息学习的不足。

3 本文方法

3.1 本文方法概述

模型整体结构如图 4 所示，主干网络使用 Res2Unet，输入的 MSFA 图像经过主干网络可以被充分的提取空间和光谱特征并能初步重建出具有一定精度的全光谱图像。为了补偿主干网络对于边缘细节部分恢复的不足，使用基于 PPI 信息提取的子分支 [17]，结合 Sobel 算子提取出图像的边缘细节以加强网络对于边缘信息的学习能力。

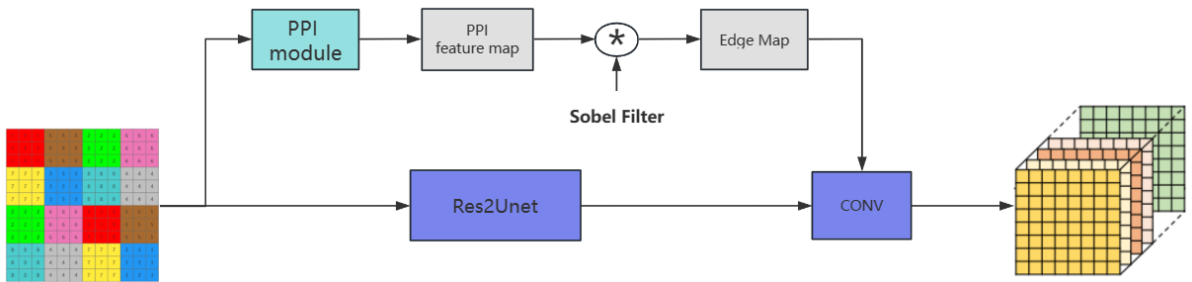


图 4. 网络整体结构示意图

3.2 特征提取模块

主干网络采用了 Unet 的对称残差连接结构如图 5 (a)，使用 *res2net - SE* 作为主要的特征提取模块，在每一层连接中还使用的 PixelUnShuffle 下采样和 PixelShuffle 上采样。*Res2net* [8] 是在 *Resnet* 上进行的改进，本文使用 Res2Net 结构如 5 (b)，特征图 $P1(C,H,W)$ 经过一层 1×1 的卷积层（包括 bn 和 LeakyReLU）将通道数改变为 $C \times (base_width/64) \times scale$

得到特征图 $P2$ 。然后将特征图 $P2$ 按通道分为 $scale$ 块，第一块保留原始特征信息，从第二块开始，分别进过一个 3×3 的卷积，从第 3 块开始，进入 3×3 卷积前先与上一块的卷积的结果相加，每个 3×3 卷积之后都跟随着 bn 和 $LeakyReLU$ ，这些 3×3 卷积不改变每一块特征图的大小和通道数；然后将每一块卷积后的结果按通道拼接起来送入一个 1×1 卷积 (没有 bn 和 $LeakyReLU$) 将通道数变回 C 得到特征图 $P3$ ，最后 $P3$ 再经过一个 SE 通道注意力模块并与最开始的 $P1$ 相加后输出，其过程可公式化为式 1：

$$\mathbf{y}_i = \begin{cases} \mathbf{x}_i & i = 1; \\ \mathbf{K}_i(\mathbf{x}_i) & i = 2; \\ \mathbf{K}_i(\mathbf{x}_i + \mathbf{y}_{i-1}) & 2 < i \leq s. \end{cases} \quad (1)$$

$Res2net$ 有两个重要的参数 $scale$ 和 $base_width$ ，分别控制着分组 3×3 卷积的个数和进入 3×3 卷积时的通道数。[8] 证明使用更大的 $scale$ 和 $base_width$ 可以有更好的特征提取性能，但为了平衡本工作已有的计算资源，本文选用 $scale = 6, base_width = 32$ 。

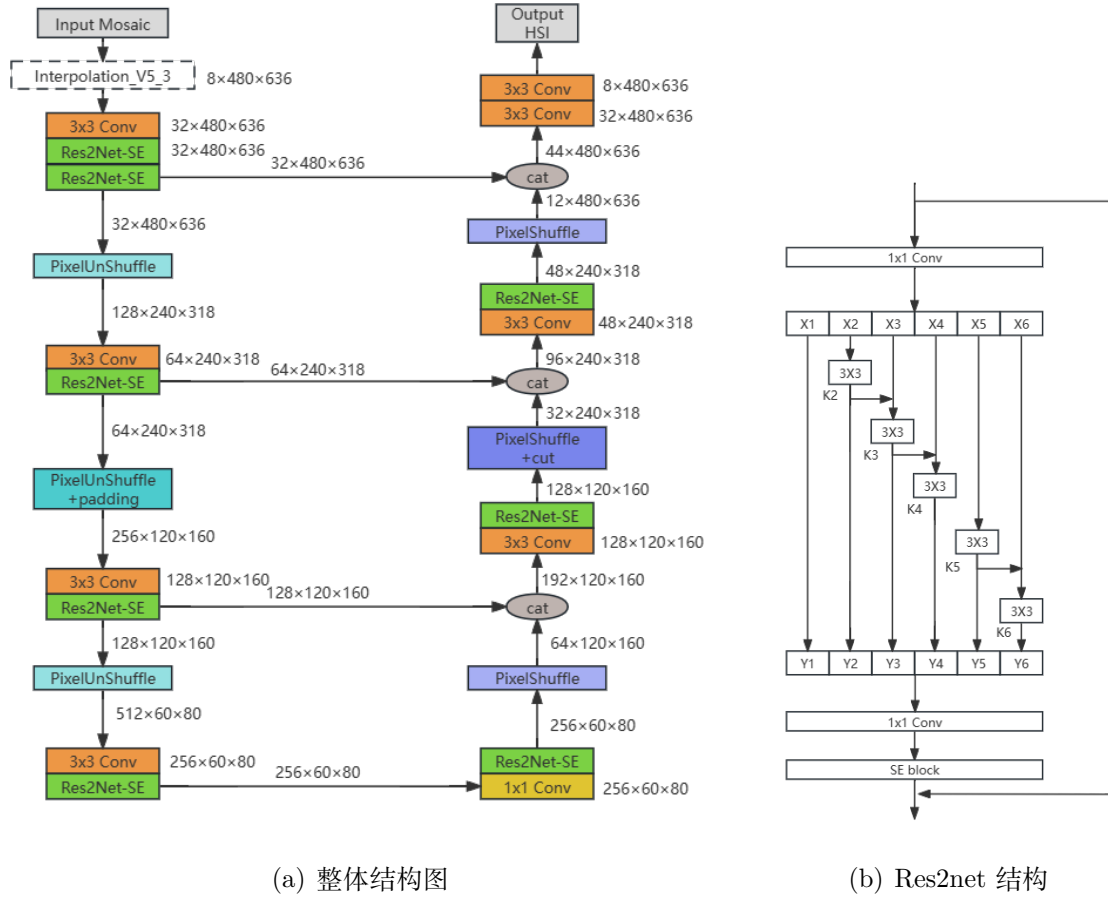


图 5. Res2Unet 主干网络示意图

3.3 损失函数定义

由于网络由双分支组成，每一分支有自己独特的损失函数，具体损失函数介绍将在4.2节以及第5.2节中介绍。

4 复现细节

4.1 与已有开源代码对比

本文所复现模型以及训练代码均没有开源代码或未使用开源代码，整个网络是根据论文作者的介绍以及参考文献中提供的模型图复现所得。本文主要的工作或创新在于：构建数据集、修改模型结构以及参数使其符合本文的任务、提出混合训练函数训练、添加 PPI 边缘信息注入分支模块 [17]。主要实现细节将在4.2节以及第5章中讲解。

4.2 创新点

论文 [10] 证明 PPI 编码的高频信息与重构的多光谱图像具有较强的正线性关系，受到 [17] 利用 PPI 中高频信息的启发，本文借鉴他们的工作利用 PPI 对多光谱去马赛克图像的边缘相关信息进行补偿。更具体地说，为了优化主干网络，实现了一个 PPI 边缘信息子分支 PPI module，从 PPI 中提取高频信息，并使用 sobel 算子提取出 PPI 中的边缘信息，用来加强网络学习边缘细节的能力。PPI module 的示意图如 6。

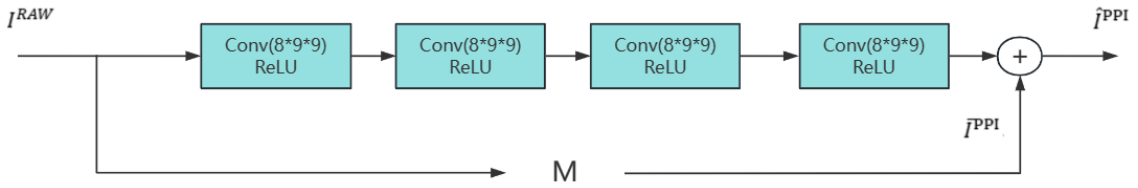


图 6. PPI module

如 4 所示，我们将 Res2Unet 与 PPI module 结合，通过将原始马赛克图像 I^{RAW} 输入 PPI 滤波器 M，M 的定义如式 2：

$$M = \frac{1}{64} \begin{bmatrix} 1 & 2 & 2 & 1 \\ 2 & 4 & 4 & 2 \\ 2 & 4 & 4 & 2 \\ 2 & 4 & 4 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix} \quad (2)$$

将 M 与 I^{RAW} 进行卷积，可以生成一个初步的 \bar{I}^{PPI} 。然后，使用不同核大小的 4 个卷积操作 ($8 \times 9 \times 9$ 、 $8 \times 7 \times 7$ 、 $8 \times 5 \times 5$ 和 $1 \times 5 \times 5$)，估计图像的 PPI 并与 \bar{I}^{PPI} 做元素间相加得到 \hat{I}^{PPI} ，通过学习 \hat{I}^{PPI} 与 GTI^{PPI} 之间的残差来估计 \hat{I}^{PPI} 。学习过程可以表示为二者的 L1 范式最小化过程，如式 3：

$$L_{ppl}(\theta_p) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \| \hat{I}_{i,j}^{ppl} - I_{i,j}^{ppl} \|_1 \quad (3)$$

得到估计出来的 \hat{I}^{PPI} 后，采用 Sobel 滤波器进行梯度提取，目的是提取出 PPI 中边缘高频信息，如图 4所示。Sobel 算子为 x 和 y 两个方向的边缘提取算子，其定义如式 4， \hat{I}^{PPI} 经过两

个方向的卷积之后被提取出边缘特征图 (Edge Map)，之后将边缘特征图复制为 8 通道图像，与主干网络重建的全波段图像进行元素级相加，经过一层卷积进行特征融合后得到最终输出。

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (4)$$

5 实验结果分析

5.1 数据集与数据预处理

本次实验主要在我们的自建数据集上进行，自建数据集包含 123 张训练样本和 54 张测试样本，空间分辨率为 1022×1024 ，任务波段为 $\{450, 550, 650, 720, 800, 850, 950\}$ 共 8 个波段。为了提高模型的精度，我们使用了数据预处理策略和数据增强策略。为了处理原始的 mosaick 数据来帮助模型训练，我们设计了 3 种数据预处理的方式：1. 基于通道的最大值归一化，该方法需要记下 mosaick 各波段的最大值，优点是各波段归一化后的数据范围一致，这样可以让模型更多的关注各波段的空間信息重构；2. 基于全局最大值归一化，该方法只需要记下所有波段的 1 个最大值即可，优点是保留了各波段之间的相对数值特征，有较好的光谱信息保留，但是实际训练时很难收敛；3. 固定最大值归一化，该方法需要人为指定最大值，比如 16bit 的数据表示，可以指定最大值为 65535，该方法的优点是不仅保留了各波段的信息，也体现了当时的光强信息，但同时也是最难训练的。从实际运用来看， $1 > 2 > 3$ 在数据量一定的情况下，是越来越难训练的，数据量的需求是越来越多的。针对现有的数据量，也从实际训练效果来看，我们选择了基于通道的最大值归一化，如图 7 所示，训练数据是由 mosaick 数据得到的 8 波段数据。黑色表示 0 值，先对各波段求最大值，然后各波段除以各自的最大值，得到最终的训练数据。

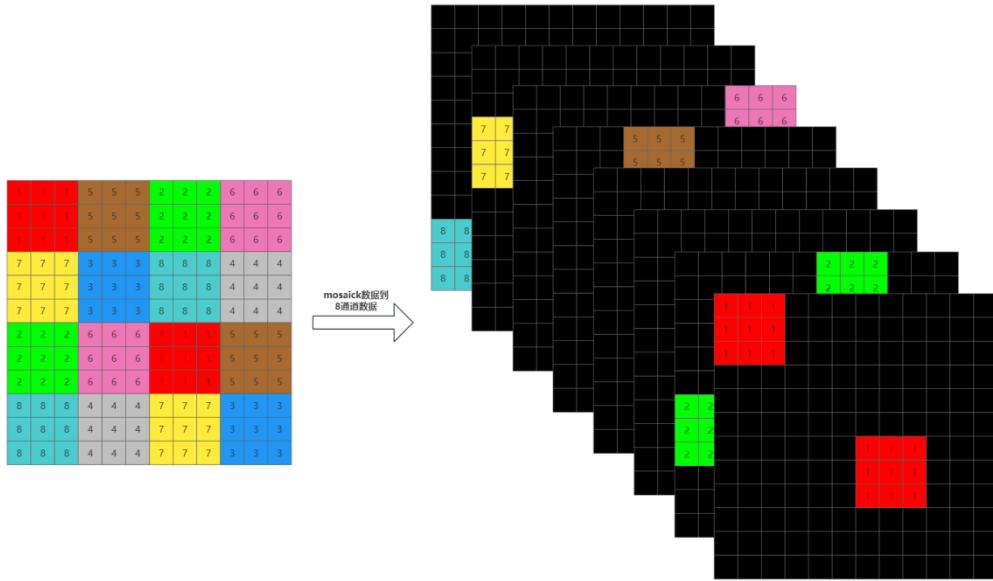


图 7. 基于通道的最大值归一化

针对我们的自建数据集，我们进行了数据扩充策略，原始图像大小为 1022×1024 ，网络接受的输入为 482×636 。针对自建数据集的处理，主要的考量为 3 个点：1. 保证一幅图的全局

信息；2. 每幅图固定块的局部细节信息；3. 尽可能多的起始点得到的固定块的局部信息。假设一副原始 12×12 的图像，目标是 6×6 的尺寸训练图像，扩充方式如图 8 所示。因此数据扩充策略有三种。第一种是使用 `resize` 将原图缩小为 480×636 ；第二种是固定步长裁剪，在原始数据（没有 `resize`）上进行大小为 480×636 ，步长也为 480×636 的裁剪；第三种是随机裁剪，其过程是：在训练时，将 `train_dataset` 大小设定为 738 ($123+492+123$ (原始))，当随机的 `idx` 大于等于 615 ($123+492$) 时，则随机从 123 张原始大图中选择一张进行随机位置的裁剪，大小为 480×636 。因此总体训练数据为 738 ($615+123$) 张，后 123 为随机图像。数据扩充后，数据集构成如图 9 所示。

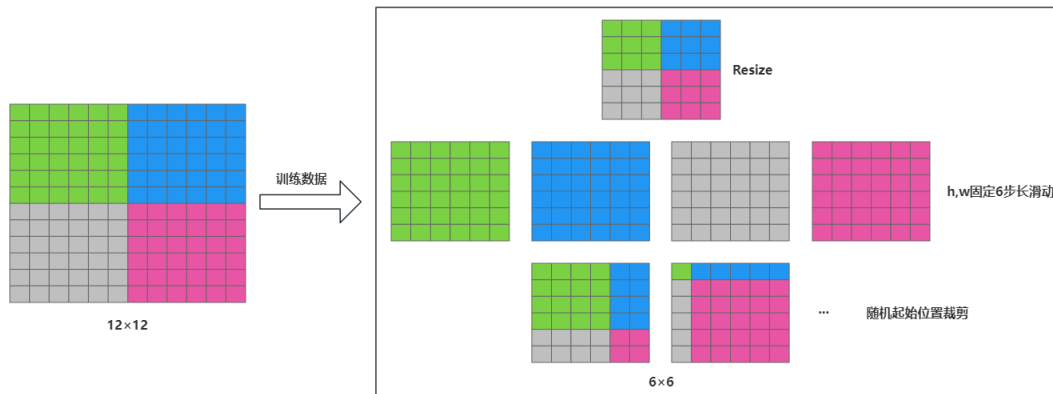


图 8. 数据扩充策略

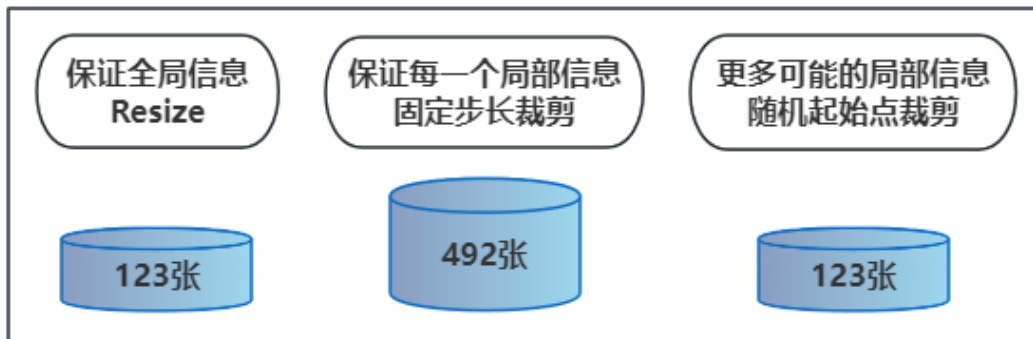


图 9. 数据集构成

5.2 训练细节

本文使用的网络模型使用 Pytorch 框架编写，模型训练平台为 NVIDIA GeForce RTX 3080Ti 显卡，借鉴的所有深度神经网络均没有源码提供，因此都是本人自己按照论文提供的模型图复现出来。提出的基于 Unet 的 PPI 边缘融合残差网络使用的 Adam 优化器进行学习率的调整，设置参数 $\beta_1 = 0.5, \beta_2 = 0.99$ 。学习率更新策略为周期余弦退火策略，初始周期为 10，退火倍率为 2，最小学习率设置为初始的 0.1 倍。初始学习率为 0.0001。训练网络的损失函数由两部分构成，第一部分为 backbone 网络即 Res2Unet 网络的损失，使用 `['L1Loss', 'MSELoss', 'MARELoss']` 混合损失函数，`L1Loss` 重点关注绝对的值的误差损失，`MSELoss` 主要也是绝对的值的误差损失 `MARELoss` 重点关注相对的值的误差损失，三种损

失函数的比例为 $[0.5, 0.05, 0.45]$; 第二部分是 PPI module 的损失, 采用预测的 PPI 与 GTPPI 之间的 L1 损失进行训练, 可数学表达为式 3。图 10展示了训练过程中参数变化曲线。

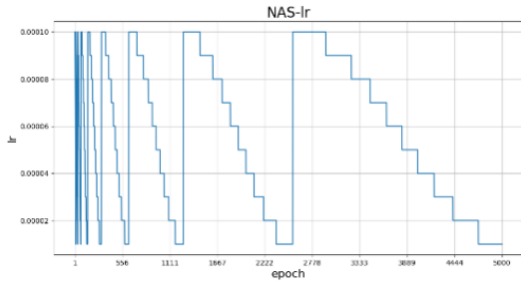
5.3 评价指标

我们通过测量峰值信噪比 (PSNR) 和光谱角度映射 (SAM) 来定量评估各种网络重建的效果。PSNR 用来衡量重建的图像与 GT 图像的峰值信噪比, 其中越高说明其图像质量越接近 GT 图像。PSNR 的定义如式 5, 其中 PEAK 表示图像的最大可能像素值, MSE 是二者的均方误差, 定义为式 6。SAM 用来衡量光谱保真度, 其值越低表示二者的光谱信息越接近。其定义如式 7, 其中 $P_{gt,i,c}$ 和 $P_{rec,i,c}$ 分别表示真实图像和重建图像中第 i 个像素的 c 光谱通道的值, $|P_{gt}|$ 是地面真实图像的大小 (像素数 \times 光谱通道数) [2]。

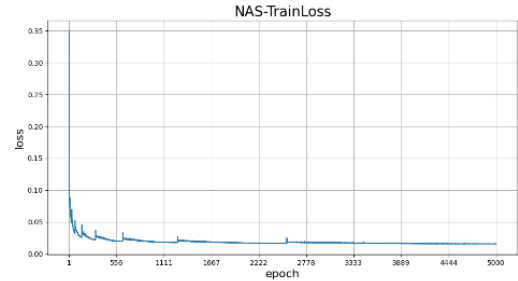
$$PSNR = 10 \cdot \log_{10} \left(\frac{PEAK^2}{MSE} \right) \quad (5)$$

$$MSE = \frac{1}{|P_{gt}|} \sum_{i,c} (P_{gt,i,c} - P_{rec,i,c})^2 \quad (6)$$

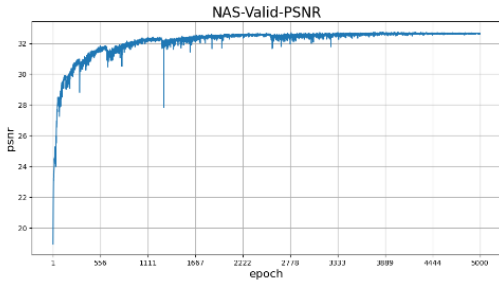
$$SAM = \frac{1}{|P_{gt}|} \sum_i \cos^{-1} \left(\frac{\sum_c P_{gt,i,c} P_{rec,i,c}}{\sqrt{\sum_c P_{gt,i,c}^2} \sqrt{\sum_c P_{rec,i,c}^2}} \right) \quad (7)$$



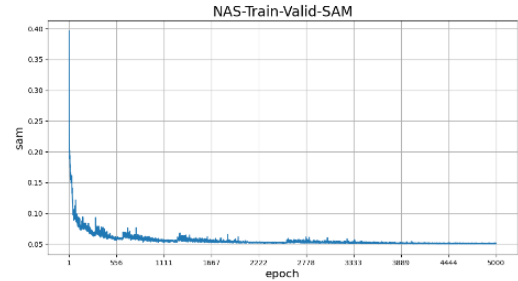
(a) LearnRate 变化曲线



(b) 损失函数变化曲线



(c) 验证集 PSNR 变化曲线



(d) 验证集 SAM 变化曲线

图 10. 训练过程中参数变化曲线

5.4 与其他 SOTA 网络的对比

为了体现本文提出的网络的有效性，我们在自建数据集上与其他网络进行了对比。选用的网络同样来自 NTIRE 2022 spectral demosaicing challenge [2] 的其他队伍，使用相同的平台复现得到，测试环境为 NVIDIA GeForce RTX 1050Ti 显卡。对比网络选用 SCRBnet 和 EHANnet。其中 SCRBnet 使用深层网络提取初始 MSFA 图像的通道权重，并在后续重建过程中利用权重进行空间域和频谱域的补全。EHANnet 来自一个图像超分辨率 SOTA 网络 HANnet [12]，在其基础上删除了 HAN 的上采样层以保持输入的大小，并在主干之后添加归一化层以避免剪辑操作造成的损失。我们将这三个网络训练完成后，在测试集上进行了测试，结果如表 1 所示。从中我们可以看出，在 PSNR 指标上我们的模型比其他两个模型高出 1-3 dB，同时拥有更小的 SAM 指标，这表示本文提出的模型无论是空间维度的重建还是光谱维度的重建都要优于另外两个模型。此外从模型推理时间上来看，本文提出的模型具有更快的推理速度。

	AVG_PSNR	AVG_SAM	AVG_TIME
SCRBnet	31.17817	0.06006	0.44832
EHANnet	29.23914	0.07679	0.39650
ours	33.19409	0.05040	0.38732

表 1. 与其他网络在自建数据集上的平均 PSNR、平均 SAM、平均时间对比

为了展示三个不同模型在真实视觉上的差异，我们选取了三个具有代表性的场景图像对模型进行测试，分别是有斑马线的车道，代表具有几何纹理的场景；长焦室外，代表相机拍摄远景的情形；花，代表色彩丰富的场景。图 11 展示了各个网络对这三个场景图形的重建结果(伪彩)，第一列是 groundtruth 图像，2-4 列分别是本文提出的网络、SCRBnet 和 EHANnet 的重建结果。从整体上看，本文提出的网络在视觉效果上更加接近真实图像。主要体现在：图像整体更加平滑，高梯度区域过度更加自然；异常点(图像预测值超出像素最大值，显示为红点或蓝点)更少。为了展示我们提出的网络在重建图像细节上有更好的观感，我们选了图像上的部分细节区域进行放大，结果如图 12。从细节可以看出，本文提出的网络能够重建出更具细节性的效果，且相比于其他网络像素伪影更少。

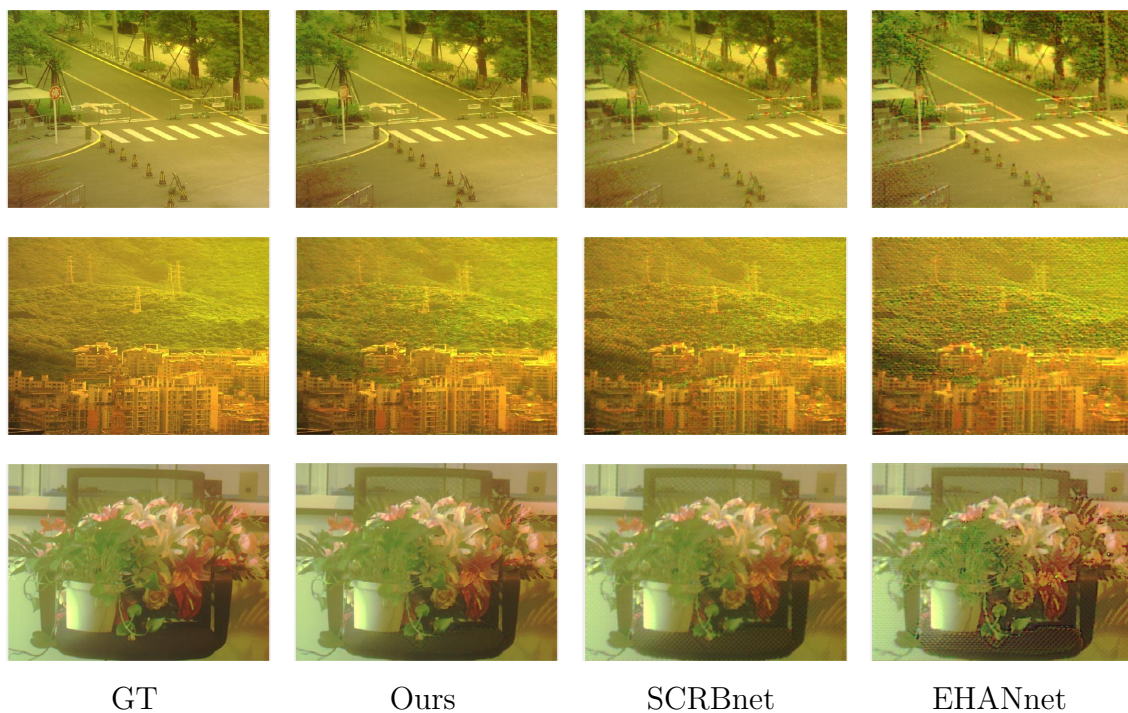


图 11. 与其他网络在三个场景下的对比

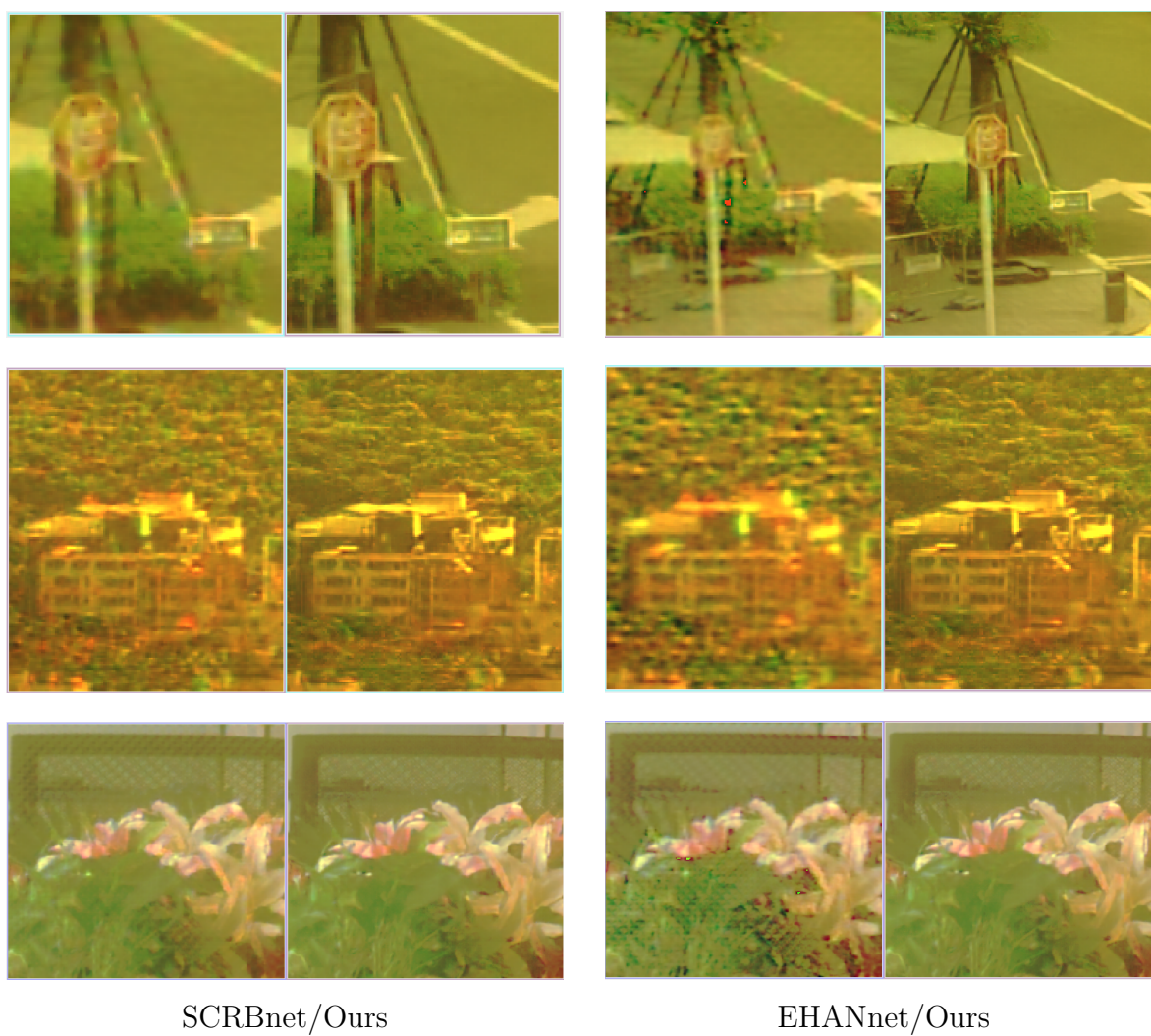


图 12. 与其他网络的细节对比

5.5 消融实验

为了验证提出的 PPI module 的有效性,我们使用原始没有添加 PPI module 分支的 Res2Unet 的在相同的平台进行训练和测试,表 2 显示了在 54 张测试图像中原始 Res2Unet 和使用了 PPI 边缘特征融合的 Res2Unet-PPI 指标结果。从平均指标上来看 PPI module 能够提高 PSNR 指标 1dB 左右,相应 SAM 也能降低 0.01 左右,说明了本文使用的 PPI module 的有效性。

	AVG_PSNR	AVG_SAM	AVG_TIME
Res2Unet	32.27204	0.05131	0.43303
Res2Unet-PPI	33.19409	0.05040	0.38732

表 2. Res2Unet 与 Res2Unet-PPI 在自建数据集上的平均 PSNR、平均 SAM、平均时间对比

同样,为了展示三个不同模型在真实视觉上的差异,我们选取了三个具有代表性的场景图像对原始 Res2Unet 模型进行测试,其对比结果如图 13 所示。第一列是真实图像,第二列是 Res2Unet-PPI 网络的结果,第三列是原始 Res2Unet 的结果。从对比中我们可以看到,二者在图像平滑度与整体清晰度上效果是差不多的。为了显示出 PPI module 发挥出了边缘细节补充的作用,在图 14 中,我们同样放大了图像相同部分的细节进行展示。图中每一行左边是没有 PPI module 的 Res2Unet 测试结果,右边是加了 PPI module 分支的测试结果。从第一行中可以看出,添加了 PPI module 后,图像中文字有了更加清晰的边缘细节,第二行中远处的建筑也有更加清晰的边缘棱角,第三行中花朵与绿叶的分界也更加清晰。细节结果进一步显示出本文使用的 PPI module 发挥了补充边缘细节的作用。

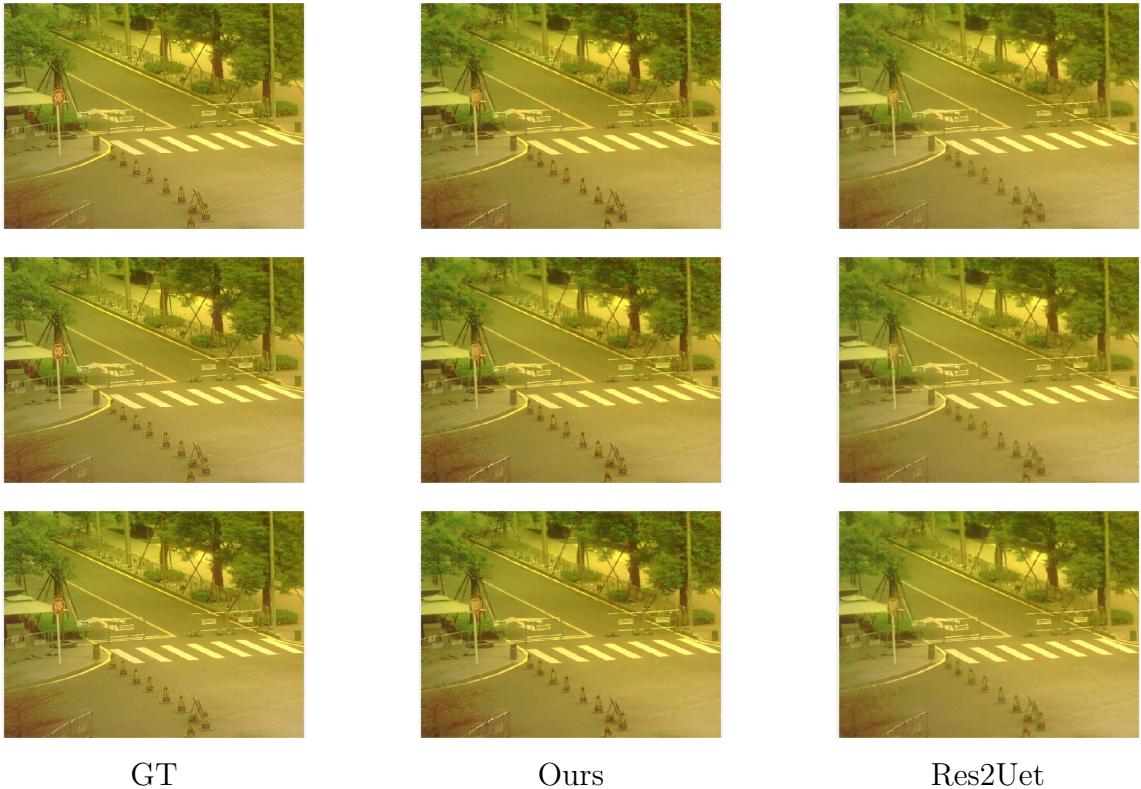


图 13. Res2Unet-PPI 与 Res2Unet 在三个场景下的对比结果

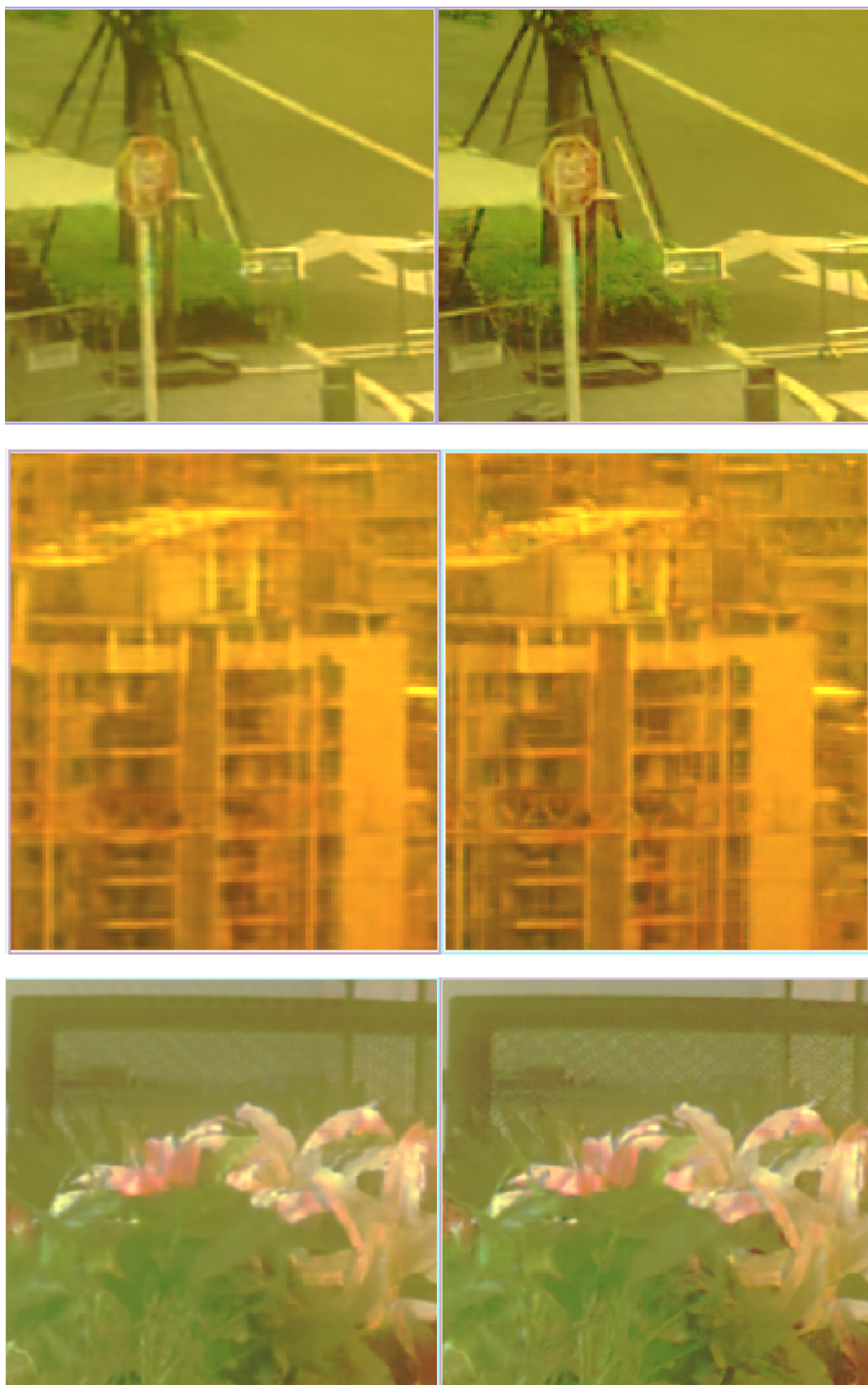


图 14. 与其他网络的细节对比

6 总结与展望

本文提出了一种特殊的多光谱滤波器阵列，在 3×3 块中只有同一波段的信息，使其在空间域和频谱域上的采集更加稀疏，但是单一像素位置上的同一波段光强值的采集量得到了提升，有利于提升采集图像的亮度。且同一滤波器可制作尺寸更大，给工艺加工方面带来了便捷，也因此可以降低制造成本。本文还提出了用于这种特定的多光谱滤波器阵列图像的去马赛克的神经网络，该网络是基于 Unet 和 PPI 边缘信息融合的双分支网络。为了实现 MSFA 图像的高精度去马赛克，我们使用 Unet 作为网络的主要结构，使用 Res2Net 作为特征提取模块，在每一层连接中还使用的 PixelUnShuffle 下采样和 PixelShuffle 上采样，构建出 Res2Unet 主干网络提取空间和频谱特征来重建 MSFA 图像。此外，还使用 PPI 边缘信息注入子分支来帮助主干网络补充图像边缘细节信息。在我们的自建数据集以及特定的多光谱滤波器阵列上与其他基于深度学习的 MSFA 去马赛克解决方案相比，证明了我们提出的方法在空间精度和光谱保真度方面有更加优越的性能。通过消融实验也证实了本文所使用的 PPI 边缘信息注入子分支有效的为主干网络提供了图像边缘信息重建的帮助。一系列实验表明，我们提出的 Res2Unet-PPI 模型在定量指标和视觉效果方面均优于实验中对比的 MSFA 去马赛克方法。

参考文献

- [1] Giancarlo A Antonucci, Simon Vary, David Humphreys, Robert A Lamb, Jonathan Piper, and Jared Tanner. Multispectral snapshot demosaicing via non-convex matrix completion. In *2019 IEEE Data Science Workshop (DSW)*, pages 227–231. IEEE, 2019.
- [2] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yaqi Wu, Xun Wu, Zhihao Fan, Chenjie Xia, Feng Zhang, et al. Ntire 2022 spectral demosaicing challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 882–896, 2022.
- [3] Johannes Brauers and Til Aach. A color filter array based multispectral camera. *Workshop Farbbildverarbeitung*, 2006.
- [4] Klaas Dijkstra, Jaap van de Loosdrecht, Lambert RB Schomaker, and Marco A Wiering. Hyperspectral demosaicking and crosstalk correction using deep learning. *Machine Vision and Applications*, 30(1):1–21, 2019.
- [5] Thibaud Ehret, Axel Davy, Pablo Arias, and Gabriele Facciolo. Joint demosaicking and denoising by fine-tuning of bursts of raw images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8868–8877, 2019.
- [6] Kai Feng, Yongqiang Zhao, Jonathan Cheung-Wai Chan, Seong G Kong, Xun Zhang, and Binglu Wang. Mosaic convolution-attention network for demosaicing multispectral filter array images. *IEEE Transactions on Computational Imaging*, 7:864–878, 2021.

- [7] Kai Feng, Yongqiang Zhao, Jonathan Cheung-Wai Chan, Seong G Kong, Xun Zhang, and Binglu Wang. Mosaic convolution-attention network for demosaicing multispectral filter array images. *IEEE Transactions on Computational Imaging*, 7:864–878, 2021.
- [8] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019.
- [9] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [10] Shumin Liu, Yuge Zhang, Jie Chen, Keng Pang Lim, and Susanto Rahardja. A deep joint network for multispectral demosaicking based on pseudo-panchromatic images. *IEEE Journal of Selected Topics in Signal Processing*, 16(4):622–635, 2022.
- [11] Sofiane Mihoubi, Olivier Losson, Benjamin Mathon, and Ludovic Macaire. Multispectral demosaicing using pseudo-panchromatic image. *IEEE Transactions on Computational Imaging*, 3(4):982–995, 2017.
- [12] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, pages 191–207. Springer, 2020.
- [13] Daniele Picone, Mauro Dalla Mura, and Laurent Condat. Joint demosaicing and fusion of multiresolution coded acquisitions: A unified image formation and reconstruction method. *IEEE Transactions on Computational Imaging*, 9:335–349, 2023.
- [14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [15] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.
- [16] Kazuma Shinoda, Shoichiro Yoshiba, and Madoka Hasegawa. Deep demosaicking for multispectral filter arrays. *arXiv preprint arXiv:1808.08021*, 2018.
- [17] Bowen Zhao, Jiesi Zheng, Yafei Dong, Ning Shen, Jiangxin Yang, Yanlong Cao, and Yanpeng Cao. Ppi edge infused spatial-spectral adaptive residual network for multispectral filter array image demosaicing. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.