# RASP-Net: 基于残差块和高相似传递注意力的图像去雪网络

林晓[1,2,3)]，苏世辉[1)]，赵天娜[1,2,3,4)*]，李岩[1)]，杨启哲[1)]，马利庄[5)]

[1)] (上海师范大学信息与机电工程学院　上海　200234)
[2)] (上海师范大学上海智能教育大数据工程技术研究中心　上海　200234)
[3)] (上海市中小学在线教育研究基地　上海　200234)
[4)] (上海师范大学人工智能教育研究院　上海　200234)
[5)] (上海交通大学电子信息与电气工程学院　上海　200240)
(zhaotianna@shnu.edu.cn)

**摘　要：**在冬季或高海拔的雪地细节拍摄的照片中，大量的雪花覆盖在图像上，严重影响图像的质量。针对当前图像去雪算法中侧重提取海量特征检测并去除多尺度雪花，缺少特征甄别筛选机制的问题，为了提取相关性强的雪花特征，提出基于残差块和高相似传递注意力的图像去雪网络 RASP-Net。首先将残差块和空间注意力机制相结合，动态地增强样本有用特征、抑制无用雪花特征；然后使用高相似性传递注意力和软阈值处理技术动态地判断雪花特征；最后通过循环结构细化特征提取和雪花去除，提升模型的去雪性能和鲁棒性。在 Snow100K、SRRS 和 CSD 数据集上的大量实验结果表明，RASP-Net 在合成和真实世界 2 种数据集的去雪效果均优于文中对比的方法，其中，Snow100K 数据集的峰值信噪比达到了 35.51 dB，可精准重建因雪花遮挡而丢失的图像细节。

**关键词：**残差块；高相似传递注意力；软阈值处理；循环结构
**中图法分类号：**TP391.41　　**DOI：**10.3724/SP.J.1089.2025-00249

# RASP-Net: Residual Block with High Similarity Pass Attention for Image Desnowing

Lin Xiao[1,2,3)], Su Shihui[1)], Zhao Tianna[1,2,3,4)*], Li Yan[1)], Yang Qizhe[1)], and Ma Lizhuang[5)]

[1)] (*The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai*　200234)
[2)] (*Shanghai Engineering Research Center of Intelligent Education and Big Data, Shanghai Normal University, Shanghai*　200234)
[3)] (*The Research Base of Online Education for Shanghai Middle and Primary Schools, Shanghai*　200234)
[4)] (*Institute of Artificial Intelligence on Education, Shanghai Normal University, Shanghai*　200234)
[5)] (*School of Electronic Information Electrical Engineering, Shanghai Jiao Tong University, Shanghai*　200240)

**Abstract:** In photographs taken in snowy environments during winter or at high altitudes, a large number of snowflakes obscure the images, severely degrading their quality. Current image desnowing algorithms mainly focus on extracting massive features to detect and remove multi-scale snowflakes, but they lack an effective mechanism for feature discrimination and selection. To address this issue and extract snowflake features with stronger relevance, we propose an image desnowing network called RASP-Net, which is based on residual blocks and high-similarity propagation attention. First, residual blocks are combined with a spa-

收稿日期：2025-07-04；修回日期：2025-08-18。**基金项目**：国家自然科学基金（62302167，62477013）；上海市自然科学基金（24ZR1456100）；上海市科学技术委员会基金（24DZ2305900）；上海市促进产业高质量发展专项资金（2211106）。**林晓**（1978—），女，博士，教授，CCF 会员，主要研究方向为人工智能教育、多模态数据融合；**苏世辉**（2000—），男，硕士研究生，主要研究方向为图像处理；**赵天娜**（1992—），女，博士，助理研究员，论文通信作者，CCF 会员，主要研究方向为人工智能、多标记分类、粒计算；**李岩**（1991—），男，博士，讲师，主要研究方向为几何深度学习、因果推断及自然语言处理；**杨启哲**（1994—），男，博士，讲师，主要研究方向为人工智能；**马利庄**（1963—），男，博士，教授，博士生导师，CCF 会员，主要研究方向为计算机视觉、计算机图形学、数字图像处理、数字多媒体、人工智能技术应用。

tial attention mechanism to dynamically enhance useful features while suppressing irrelevant snowflake features. Then, high-similarity propagation attention and soft-thresholding techniques are employed to adaptively identify snowflake features. Finally, a recurrent structure is introduced to refine feature extraction and snowflake removal, thereby improving the desnowing performance and robustness of the model. Extensive experiments on the Snow100K, SRRS, and CSD datasets demonstrate that RASP-Net achieves superior desnowing results on both synthetic and real-world datasets compared to state-of-the-art methods. In particular, on the Snow100K dataset, RASP-Net achieves a peak signal-to-noise ratio (PSNR) of 35.51 dB, and can accurately reconstruct image details lost due to snowflake occlusion.

**Key words:** residual block; high similarity pass attention; soft-thresholding; recurrent structure

多雪地区的图像和视频易受雪花干扰，降低了可视性，在交通监控、户外拍摄等领域阻碍目标识别与状况判断。以监控系统为例，降雪会严重影响视频监控的图像质量，使目标检测和跟踪算法的性能下降，降低了监控系统的可靠性。在自动驾驶场景中，车辆感知系统需要依赖清晰的图像准确地识别道路、行人、障碍物和其他交通参与者，而降雪会显著地降低感知系统的性能，降低自动驾驶的安全性。在计算机视觉领域，降雪场景下的图像处理是一个极具挑战性的问题。图像质量退化对依赖高质量视觉输入的计算机算法的负面影响非常大，如在目标检测[1]、语义分割[2]和场景理解[3]等关键算法中，退化的图像质量不仅会降低算法的检测精度，造成目标误判和漏判，而且还会干扰模型对图像语义信息的准确解析，严重阻碍场景理解任务的顺利进行，极大地削弱算法在实际应用中的效能。去雪算法可以提升视觉效果，助力交通安全保障与智能系统升级，在数字摄影、遥感测绘等领域也发挥重要的作用。

现有的方法大致分为传统图像去雪方法和基于深度学习的图像去雪方法 2 类。传统图像去雪方法依赖图像处理技术，分为基于滤波和稀疏表示。基于滤波的图像处理技术通过均值滤波[4]、中值滤波[5]、高斯滤波[6]等方法去除雪花噪声，这些方法简单易实现，但容易导致图像细节丢失，尤其是在雪花密集或复杂的场景中去雪效果有限。基于稀疏表示的方法则利用图像中雪花和背景的稀疏特性，通过稀疏编码将雪花与背景分离，这类方法在雪花分布较为稀疏时表现较好，但在复杂场景中，由于雪花与背景的纹理相似性较高，难以准确地分离，因此去雪效果不理想。此外，传统方法通常需要手动调整参数，缺乏自适应性，难以应对多样化的雪天场景。

随着深度学习技术的发展，基于卷积神经网络（convolutional neural network，CNN）[7]和生成对抗网络（generative adversarial network，GAN）[8]的图像去雪方法成为主流方法。CNN 方法通过端到端的学习方式直接从数据中学习雪花特征并进行去除。DesnowNet[9]是首个基于 CNN 的去雪网络，其采用多任务学习框架预测雪花掩模和恢复背景图像，取得了较好的效果。GAN 方法则利用生成对抗网络的图像生成能力，将带雪图像转换为无雪图像。DesnowGAN[10]通过金字塔结构、跨尺度连接、空洞空间金字塔池化（atrous spatial pyramid pooling，ASPP）模块和 GAN 细化阶段，有效地融合多尺度特征并提升图像真实感。此外，引入注意力机制进一步增强了模型对雪花区域的感知与去除能力。LMQFormer[11]通过自注意力机制增强模型对雪花区域的识别能力。深度学习方法在去雪任务中表现出强大的特征学习能力，已成为当前研究的重点方向。

从雪花的退化模型中可以看出，当前去雪任务的核心是将雪花层与背景层相分离，以去除雪花层。目前，大多数研究致力于多尺度地提取雪花特征，但是没有合适的方法对雪花噪声进行精确建模和有效去除，复原图像仍存在细节模糊或雪花残留等问题。

本文提出基于残差块和高相似传递注意力的图像去雪网络（residual and similar pass network，RASP-Net）进行图像去雪，包括残差块和空间注意力（residual block and spatial attention，RBSA）、高相似传递注意力（high similarity pass attention，HSPA）机制和循环结构。残差块通过多个跳跃连接提取丰富的局部特征，包括多个标准卷积层（Conv）和修正线性单元（ReLU）。空间注意力机制（spatial attention mechanism，SAM）旨在学习卷积特征之间的空间关系，包含全局最大池化（global max pooling，GMP）、全局平均池化（global

average pooling，GAP），以及 Conv、ReLU 和 Sigmoid 函数。GMP 和 GAP 提取图像的全局统计信息，突出关键特征并抑制无用信息。RBSA 通过元素级加法融合不同卷积层的特征，进一步提升去雪性能。与现有的残差注意力块相比，RBSA 引入更多的跳跃连接，能够更有效地提取并融合不同卷积层的特征，增强去雪效果。

软阈值（soft thresholding，ST）操作能够有效地建模长序列，并生成稀疏概率分布，其核心是计算 ST 函数的输出，将低于阈值的坐标截断为 0，其余坐标则根据该阈值进行偏移。该操作不仅生成稀疏概率分布，而且保留了 Softmax 变换的基本特性；此外，还能为相似度更高的特征向量分配更大的权重，增强雪花特征信息的区分能力。因此，本文通过 HSPA 增强 ST 操作；同时，使用循环结构简化处理重复性任务，动态地调整执行次数，提升效率并节省资源。

（1）通过融合残差块和注意力机制，残差注意力块能够动态地强化图像中的有效特征，同时抑制与雪花相关的干扰特征。这种方法不仅提升了去雪效果，还能更好地保留图像的原始信息，避免细节丢失。（2）通过高相似性传递机制实现 ST 处理，能够动态区分雪花特征与非雪花特征。该机制利用特征间的相似性信息，自适应地判断并过滤雪花干扰，同时保留图像中的有效内容，从而显著地提升去雪效果的精准度。（3）使用了循环结构来逐步优化去雪效果，通过多次迭代不断地细化特征提取和雪花去除过程，从而进一步提升模型的去雪性能和鲁棒性。（4）在 Snow100K、SRRS 和 CSD 数据集上大量实验表明，RASP-Net 在单图像去雪中显示出其卓越的视觉效果，并且优于文中对比方法。

## 1　相关工作

### 1.1　基于传统方法的图像去雪

传统的去雪方法严重依赖经典的图像处理手段，包括滤波、阈值分割和形态学操作，其中，滤波技术占据主导地位。中值滤波[5]是一种非常有效的降噪方法，特别适用于去除像小雪花那样的亮点噪声。Xu 等[12]将引导过滤器引入去雪研究中，并取得了良好的效果。为了保持图像的细节特征，Zheng 等[13]引导平滑滤波器在有效降噪的同时保留图像结构信息；Ding 等[14]结合保持边缘和降噪功能提出引导平滑滤波器；Shi 等[15]提出一种加权中值滤波器与引导滤波进行图像去雨。Zang 等[16]则提出一种自适应阈值的双树复小波融合方案，通过优化融合规则实现了更好的降噪与去模糊效果；Chen 等[17]提出一种基于分层小波表示的深度学习算法，在雪量估计和细节增强方面表现优异。稀疏表示也是一种传统的图像去噪方法。Luo 等[18]提出一种基于字典学习的单图像去雨方法，通过构建具备互斥性的稀疏表示有效地分离雨层与背景层。然而，现有的除雪方法在处理复杂场景时往往存在图像细节丰富度不足、雪融化区域识别不够精准等问题，因此现有技术在去雪效果上仍需突破。

### 1.2　基于深度学习的图像去雪

CNN 作为图像处理的核心工具，在图像降噪中取得了显著成果。2018 年，Liu 等[9]首次将深度学习引入图像去雪研究中，提出基于深度学习的除雪方法，通过在模拟和真实 2 种场景下进行实验验证了其有效性，该研究为后续去雪技术的发展奠定了基础；Li 等[19]采用 GAN 利用生成器和鉴别器分别处理降雪过程的细节和判别能力，其中，生成器用于生成高质量的去雪结果，而鉴别器则对生成图像的质量进行评估，这种方法在复杂降雪场景中表现尤为出色。近年来，Transformer 模型[20]被成功地应用于图像处理领域，与传统的 CNN 相比，其具有更好的泛化能力和表达能力；Lin 等[11]将 Transformer 架构与拉普拉斯先验相结合，提出一种轻量级的除雪算法，显著地提升了去雪处理效率和图像降噪性能；Cheng 等[21]提出一种基于雪掩模的自适应残差网络方法，将雪掩模作为导引，实现对积雪区域的精准聚焦，显著地降低了降雪对图像的影响；Quan 等[22]提出一种基于可逆结构的深度可分双向注意网络，通过可逆操作有效地分离积雪区域并保留关键信息；Wang 等[23]提出 MSA-former，通过多尺度注意引导的显著性图自适应地融合特征，结合全局与局部增强模块有效地提升了复杂雪景图像的还原能力；Dai 等[24]提出 TuneSnow 框架，利用高效的参数微调技术优化雪天图像修复，通过混合适配器、渐进多尺度感知和退化区域修复模块，实现卓越的除雪效果且减少了训练参数量。但是，这些方法经常会出现细节丢失、计算成本较高、在复杂雪景图像下性能下降等问题。

## 2　RASP-Net

### 2.1　整体架构

本文提出的 RASP-Net 采用编码-解码结构，可提取和重建多尺度特征，高效地去除雪花。其中，

编码器引入残差注意力机制和 HSPA 机制，提取输入图像的多尺度特征，增强特征抑制雪花干扰；解码器则通过逐步地重建图像，并结合跨尺度特征融合策略进一步提升图像的去雪质量。编码器由 GMP、Conv、归一化（batch norm，BN）、ReLU 激活函数和上采样组成；解码器由上采样操作、卷积操作、BN 处理和 ReLU 激活函数组成。编码器与解码器的结构如图 1 所示。

为了提取有效的特征，RASP-Net 中引入残差

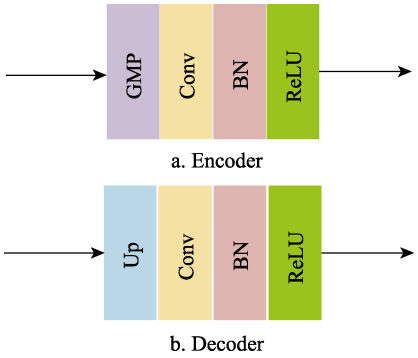



图 1 编码器与解码器结构

块和空间注意力机制（RBSA）解耦编码器和解码器的全连接层，使用 HSPA 实现 ST 限制和优化 RBSA 的解耦功能，使用循环结构迭代优化特征提取能力。RASP-Net 的结构如图 2 所示。

## 2.2 RBSA

RBSA 结构由空间注意力模块（SAM）共同构成，旨在实现局部特征提取与全局空间关系的协同优化。在残差块的设计中采用 Conv 与 ReLU 的组合，通过密集的跳跃连接实现特征的跨层传递。这种结构不仅能够有效地缓解深度网络的梯度消失问题，还能保留不同层级的特征信息，如低频结构和高频细节。残差块内部的每层卷积输出均会与后续层的输入进行逐元素相加，从而增强特征的复用能力，避免因网络深度增加而导致的信息丢失。

SAM 动态地调整残差块输出的特征图，以全局视角分析特征的空间分布，通过注意力机制突出重要区域并抑制噪声干扰。SAM 的输入为残差块生成的特征图，输出经过权重调整后与原始输入进行残差连接，确保网络在增强关键特征的同时保留原有图像的有效信息。

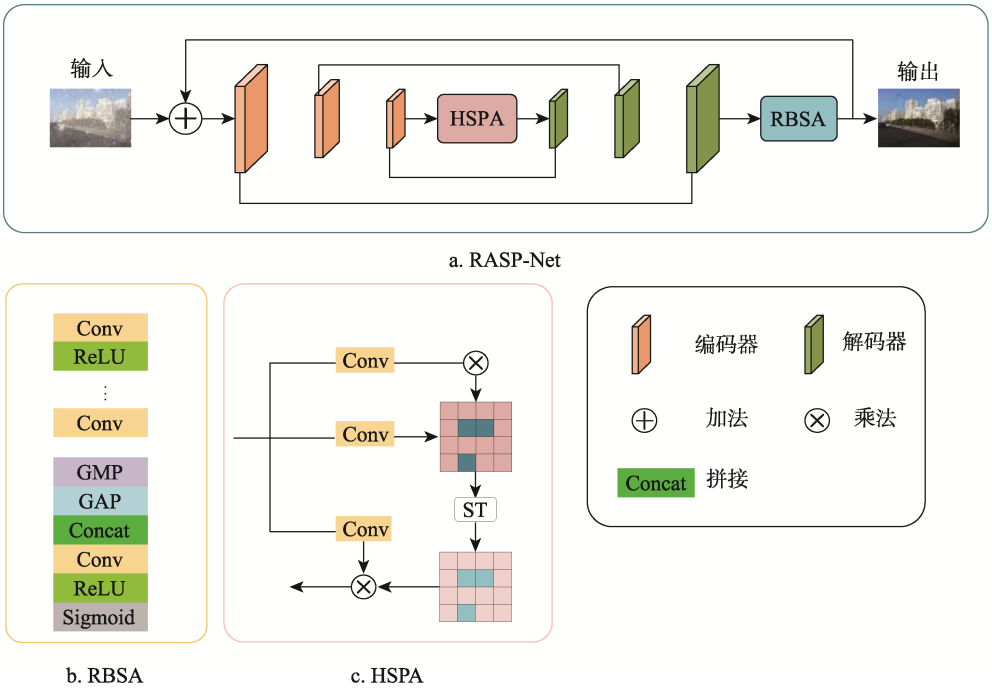



图 2 RASP-Net 的结构

SAM 模块通过聚合全局空间信息生成注意力权重图，实现对特征图的自适应调整。（1）利用 GAP 和 GMP 对输入特征图进行统计分析，其中，GAP 能够捕捉特征图的整体分布特性，而 GMP 则倾向于提取局部显著响应，两者结合可以更全面地描述特征的空间特性。（2）池化后的特征经过拼接后通过 1×1 卷积进行降维，减少计算复杂度并增强特征的表达能力；降维后的特征经过 ReLU 激活函数引入非线性变换，并通过 Sigmoid 函数将权重 BN 至[0,1]，生成最终的注意力权重图，权重图反映特征图中不同区域的重要性，通常高权重区域对应于图像中的关键结构或细节信息；最后权重图与

原始特征图进行逐元素相乘，自适应地增强或抑制特征，关注有效信息，过滤掉冗余噪声干扰。该过程的计算公式可表示为

$$\boldsymbol{f}=\mathrm{Concat}\left(\mathrm{GAP}\left(\boldsymbol{X}\right),\mathrm{GMP}\left(\boldsymbol{X}\right)\right),$$

$$\boldsymbol{A}=\mathrm{Sigmoid}\left(\boldsymbol{W}*\boldsymbol{f}\right),$$

$$\boldsymbol{X}_{\mathrm{out}}=\boldsymbol{X}\otimes\boldsymbol{A}。$$

其中，$\boldsymbol{X}$ 表示输入图像；$\boldsymbol{W}$ 表示卷积权重；$\otimes$ 表示元素相乘。

在残差注意力块中，特征融合通过密集跳跃连接和 SAM 实现。残差块内部的密集跳跃连接使浅层特征能够直接传递至深层网络，有效地避免了深层网络中的信息损失问题，该结构促进梯度的反向传播，增强不同层级特征间的交互，综合利用低层局部特征和高层语义特征。SAM 的输出通过残差连接与原始输入相加，实现了跨层次的特征融合，这种设计在保留原始特征信息的同时，利用注意力机制对关键特征区域进行选择性增强。实验结果表明，该融合策略显著地提升了网络在复杂噪声条件下的去雪性能，能够更好地平衡细节保留与噪声抑制的关系。与传统的残差结构相比，RBSA 通过引入更多跳跃连接和 SAM，实现更高效的特征提取与融合，获得了更优的去雪效果和泛化能力。

### 2.3　HSPA 机制

HSPA 是一种专门为深度单图像超分辨率模型设计的注意力机制，通过 ST 操作生成稀疏的概率分布，有效地处理长距离序列中包含大量不相关信息的问题。HSPA 的核心是探索和利用图像非局部自相似传递性信息，同时通过 ST 操作优化特征权重的分配策略，显著地提升了超分辨率重建的性能和效率。

在图像超分辨率任务中，不同区域之间往往存在相似的结构和纹理特征。虽然传统的非局部注意力（non-local attention，NLA）能够捕捉长距离依赖关系，但在处理长序列时会引入大量不相关的非局部特征。由于 Softmax 变换对所有输入值进行 BN，因此相似性很低的特征也会被赋予非零权重，降低了特征的判别性；相比之下，HSPA 通过更精细的相似性评估机制，能够更准确地识别和利用真正相关的非局部特征，避免不相关信息的干扰。NLA 表示为

$$\mathrm{NLA}\left(\boldsymbol{x}_i\right)=\sum_{j=1}^{N}\frac{\exp\left(d\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)\right)}{\sum_{k=1}^{N}\exp\left(d\left(\boldsymbol{x}_i,\boldsymbol{x}_k\right)\right)}\phi v\left(\boldsymbol{x}_j\right)。$$

其中，$\boldsymbol{x}_i$、$\boldsymbol{x}_j$ 和 $\boldsymbol{x}_k$ 分别表示输入图像中第 $i$ 个、第 $j$ 个和第 $k$ 个特征向量；$\phi v(\cdot)$ 表示一个具有 $1\times1$ 卷积运算的值向量生成函数；$d(\cdot,\cdot)$ 用于衡量特征向量之间的点积相似性。

HSP 的关键是引入 ST 操作处理非局部信息的权重分配。ST 操作定义为 $\mathrm{ST}_j(s)=\max\left\{s_j-k(s),0\right\}$，其中，$s_j$ 表示位置 $j$ 的相似度得分；$\max\{\cdot,0\}$ 操作保证输出结果非负数；

$$k(s)=\frac{\left(\sum_{j=1}^{k}s_{(j)}\right)-1}{\max\left\{k\in K\,|\,ks_{(k)}+1>\sum_{j=1}^{k}s_{(j)}\right\}}$$

表示动态计算的 ST 函数，$K=\{1,2,\cdots,N\}$。该操作保留了 Softmax 变换的基本属性：所有概率值保持非负且总和为 1，同时能够自动截断相似性较低的特征权重。通过这种方式，HSP 能够保留高相似性特征的重要权重，有效地过滤掉低相似性的干扰信息。

### 2.4　循环结构

在图像去雪任务中，循环结构通过将去雪过程划分为不同阶段实现渐进式优化。在初始阶段，网络以较高的去雪强度进行处理，快速地去除大部分雪花噪声，快速地去除大部分雪花噪声，该阶段虽然会牺牲一些细节但能有效地清除明显的雪花干扰；随着处理阶段的深入，后续网络逐步调整去雪强度，专注于恢复图像细节和边缘信息。这种分段设计的优势是不同阶段采用差异化的网络结构和损失函数，早期阶段侧重噪声去除的全局损失，后期阶段则引入边缘保护和纹理恢复的专项优化。整个过程通过循环的方式实现，每个阶段的输出作为下一阶段的输入，形成渐进式优化链条，同时中间阶段引入跳跃连接防止信息丢失。与单阶段网络相比，这种结构能更好地平衡去雪强度和细节保留，尤其在处理复杂雪景时，可以避免过度平滑或残留雪花噪声的问题。

与传统的单阶段去雪网络相比，本文进行了 3 次循环，能够更好地平衡去雪强度与细节保留之间的关系。在处理复杂雪景时，该结构可以有效地避免单阶段网络常见的过度平滑或噪声残留问题，特别是在处理含有复杂纹理的场景时，循环设计能够保持图像的结构完整性，同时优化计算效率，使其在实际应用中具有更好的可行性。这种渐进式的优化思路为图像去雪任务提供了一种有效的解决方案。循环结构表示为 $I_s^{(t+1)}=f_s\left(I_s^t,I_{\mathrm{input}}\right)$。其中，$I_{\mathrm{input}}$ 表示输入的有雪图像；$f_s$ 表示第 $s$ 个阶段的

去雪函数；$I_s^t$ 表示第 $s$ 阶段第 $t$ 次迭代的中间结果。

### 2.5 损失函数

本文使用均方误差（mean square error，MSE）损失[25] $L_{\mathrm{MSE}}$、Edge 损失[26] $L_{\mathrm{Edge}}$ 和 Char 损失[27] $L_{\mathrm{Char}}$ 共同构成总损失函数。总损失函数 $L$ 定义为各分项损失的加权和，其计算公式为

$$L=\lambda_1 L_{\mathrm{MSE}}+\lambda_2 L_{\mathrm{Edge}}+\lambda_3 L_{\mathrm{Char}}\text{。}$$

其中，$\lambda_1$=0.1，$\lambda_2$=0.1，$\lambda_3$=0.05。

$L_{\mathrm{MSE}}$ 约束在整体像素层面去雪后的图像接近真实无雪图像，确保颜色和光照的一致性，计算公式为

$$L_{\mathrm{MSE}}=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_i-x_i\right)^2\text{。}$$

其中，$N$ 为图像数量；$x$ 表示经过网络输出的图像，$\hat{x}$ 表示地面实况。

$L_{\mathrm{Edge}}$ 通过强化物体边缘和纹理结构减少雪花残留导致的细节模糊，恢复清晰的场景轮廓，计算公式为

$$L_{\mathrm{Edge}}=\sqrt{\left\|\Delta\hat{x}_i-\Delta x_i\right\|+\varepsilon^2}\text{。}$$

其中，$\Delta$ 表示拉普拉斯算子；$\varepsilon$ 表示常量。

$L_{\mathrm{Char}}$ 用于重塑清晰的无雪花图像，计算公式为

$$L_{\mathrm{Char}}=\sqrt{\left\|\Delta\hat{x}_i-\Delta x_i\right\|+\varepsilon^2}\text{。}$$

## 3 实验及结果分析

### 3.1 数据集与实验设置

在 3 个基准数据集上评估本文算法的性能，包括 Snow100K 数据集[9]、雪的真实场景去除数据集（snow removal in realistic scenario，SRRS）[28]和综合雪沉积数据集（comprehensive snow dataset，CSD）[16]。每个数据集包含大量的训练样本，训练样本包括输入雪图像、掩模层和相应的地面真实图像。

Snow100K 数据集是为除雪研究建立的第 1 个正式数据集，为后续实验奠定了坚实的基础。其中，训练集由 50 000 幅图像组成；测试集根据雪花大小分为 Snow100K-L、Snow100K-M 和 Snow100K-S 共 3 类，每类包含 16 000 幅图像。其中，Snow100K-L 包括大雪花图像，模拟了具有较大雪花颗粒和更明显视觉干扰的场景，使得去除雪噪声特别具有挑战性；Snow100K-M 包括中等大小的雪花，呈现出适度的视觉冲击，不会像较大的雪花那样有严重模糊图像；Snow100K-S 包括小尺寸雪花图像，引入的噪声干扰最小，最容易去除。本文实验中，使用 Snow100K 训练集中的 8 000 幅图像，并从 Snow100K-L，Snow100K-M 和 Snow100K-S 测试集中分别选择 2 000 幅图像。

SRRS 数据集由 Chen 等[28]在 2020 年提出的 JSTASR 中首次引入，包含 15 005 幅图像。本文实验中，以 1∶4 的比例将测试集与训练集分开，从训练集中随机选择 8 000 幅图像，并从测试集中采样 2 000 幅图像。

CSD 数据集由 Chen 等[17]于 2021 年构建，基于 HDCWNet 模型对 SRRS 数据集进行增强而得到。其中，训练集由 8 000 幅图像组成，测试集包括 2 000 幅图像；此外，该数据集还合成了相应的掩模层。CSD 数据集中的所有图像共享相同的分辨率，使其成为 3 个数据集中最一致的数据集。本文实验中，使用 CSD 数据集中相同的训练集和测试集数量评估算法的性能。

本文使用峰值信噪比（peak signal-to-noise ratio，PSNR）[29]和结构相似性（structural similarity index measure，SSIM）[30]作为评价指标。

PSNR 通过将输出图像与参考图像（通常是地面真实值）进行比较，是评估图像重建质量的标准指标，定义为

$$\mathrm{PSNR}=10\cdot\lg\left(\frac{P_{\max}^2}{\mathrm{MSE}}\right)\text{。}$$

其中，$P$ 表示像素，$P_{\max}$ 表示最大像素值。PSNR 的值越高，表示重建质量更好，该指标是评估去噪和恢复方法的可靠工具。本文中，PSNR 用于将本文提出的除雪框架的性能与最先进的方法进行比较。

SSIM 指数是一种感知度量，通过测量输出图像与其参考之间的 SSIM 来评估图像质量。与 PSNR 基于像素的指标不同，SSIM 侧重于亮度、对比度和结构信息，以更好地与人类视觉感知相一致。SSIM 的计算公式为

$$\mathrm{SSIM}(x,y)=\frac{\left(2\mu_x\mu_y+C_1\right)\left(2\sigma_{xy}+C_2\right)}{\left(\mu_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)}\text{。}$$

其中，$\mu_x$ 和 $\mu_y$ 表示图像的平均值；$\sigma_x^2$ 和 $\sigma_y^2$ 表示方差；$\sigma_{xy}$ 表示协方差。SSIM 的取值的范围为−1~1，其值越高表示相似性越大，它提供了比 PSNR 更准确的感知质量度量，广泛地应用于图像去噪、压缩和恢复任务。本文中，SSIM 用于评估生成的无雪图像的结构保真度和感知质量。

本文实验设置如下：在配备 NVIDIA A100 显卡（80GB 显存）的服务器上进行了实验；采用 PyTorch 作为框架，使用 3 组编码器、解码器结构；在训练过程中，使用 Adam 优化器，将基础学习率设置为 $1\times10^{-2}$；训练周期的总数配置为 150，学习率按照余弦函数衰减，将学习率降低到 $3\times10^{-5}$ 进行动态调整；输入图像被均匀裁剪为 256×256 像素。

### 3.2　对比实验

#### 3.2.1　定量实验

将 RASP-Net 与多种优秀的方法进行实验，包括 DesnowNet[9]、LMQFormer[11]、HDCWNet[17]、TuneSnow[24]、JSTASR[28]、DesnowFormer[31]、SnowFormer[32]、CCN[33]和 WaveFrSnow[34]。

在 Snow100K 数据集上，将 10 种方法的性能进行实验，结果如表 1 所示。可以看出，RASP-Net 的 PSNR 最好，比 HDCWNet、CCN 和 TuneSnow 分别高 11.41 dB、1.87 dB 和 1.49 dB；同样地，RASP-Net 的 SSIM 也是最优秀的，比之前最优秀的 TuneSnow 高 0.001 9。

**表 1　在 Snow100K 数据集上 10 种方法定量结果对比**

| 方法 | PSNR/dB↑ | SSIM↑ |
|---|---|---|
| DesnowNet[9] | 30.11 | 0.9317 |
| JSTASR[28] | 28.59 | 0.8665 |
| HDCWNet[17] | 24.10 | 0.8021 |
| DesnowFormer[31] | 31.88 | 0.8795 |
| SnowFormer[32] | 32.17 | 0.9112 |
| LMQFormer[11] | 31.88 | 0.9212 |
| CCN[33] | 33.64 | 0.9514 |
| TuneSnow[24] | 34.02 | 0.9687 |
| WaveFrSnow[34] | 35.31 | 0.9415 |
| RASP-Net | **35.51** | **0.9736** |

注：粗体表示最优值。

表 2 展示了多个优秀的网络模型在 CSD 数据集上，将 10 种方法的性能进行实验，结果如表 2 所示。可以看出，RASP-Net 的 PSNR 结果出色，比 HDCWNet、LMQFormer 和 TuneSnow 分别高 8.15 dB、4.57 dB 和 1.19 dB；同样地，RASP-Net 的 SSIM 也是最优秀的，比对比方法中最优秀的 WaveFrSnow 高 0.010 8。

#### 3.2.2　定性实验

选取有代表性的优秀实验结果进行视觉效果对比，包括 DesnowNet[9]、HDCWNet[17]和 LMQFormer[11]。在合成数据集上对雪花图像进行去雪操作，4 种方法的视觉效果对比如图 3 所示。可以看出，虽然 4 种方法均得到了不错的去雪效果，但是细节之处还存在差异：DesnowNet 对有些数据集的去雪效果并不理想；HDCWNet 虽然去除效果不错，但是复原后的图像色调偏暗；LMQFormer 虽然去除效果不错，但是复原后的图像色调偏亮；RASP-Net 不仅在雪花去除方面取得了显著效果，还在很大程度上保留原图的色调信息，实现了较高的色彩一致性。

**表 2　在 CSD 数据集上 10 种方法定量结果对比**

| 方法 | PSNR/dB↑ | SSIM↑ |
|---|---|---|
| DesnowNet[9] | 20.13 | 0.8186 |
| JSTASR[28] | 27.96 | 0.8850 |
| HDCWNet[17] | 29.06 | 0.9173 |
| DesnowFormer[31] | 30.34 | 0.9556 |
| SnowFormer[32] | 33.92 | 0.9551 |
| LMQFormer[11] | 32.64 | 0.9667 |
| CCN[33] | 32.07 | 0.9714 |
| TuneSnow[24] | 36.02 | 0.9687 |
| WaveFrSnow[34] | **38.41** | 0.9761 |
| RASP-Net | 37.21 | **0.9869** |

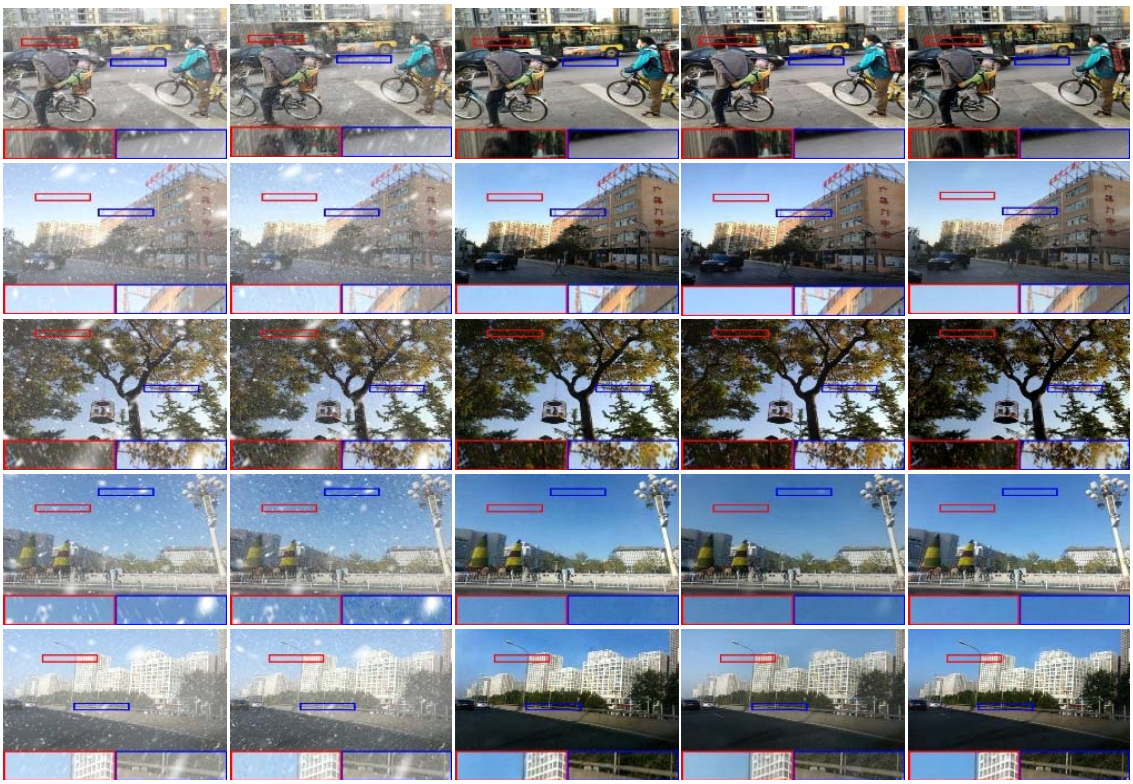
注：粗体表示最优值。

在真实数据集上，对雪花图像进行去雪操作，4 种方法的视觉效果对比如图 4 所示。可以看出：（1）多种方法的去雪效果表现不佳：DesnowNet 产生图像的失真；HDCWNet 虽然对雪花进行了一定的去除，但是并不能完全去除，且图像色调变暗；LMQFormer 虽然保证了图像的色调，但是去雪效果不理想。（2）RASP-Net 的去雪效果显著优于对比方法。

### 3.3　消融实验

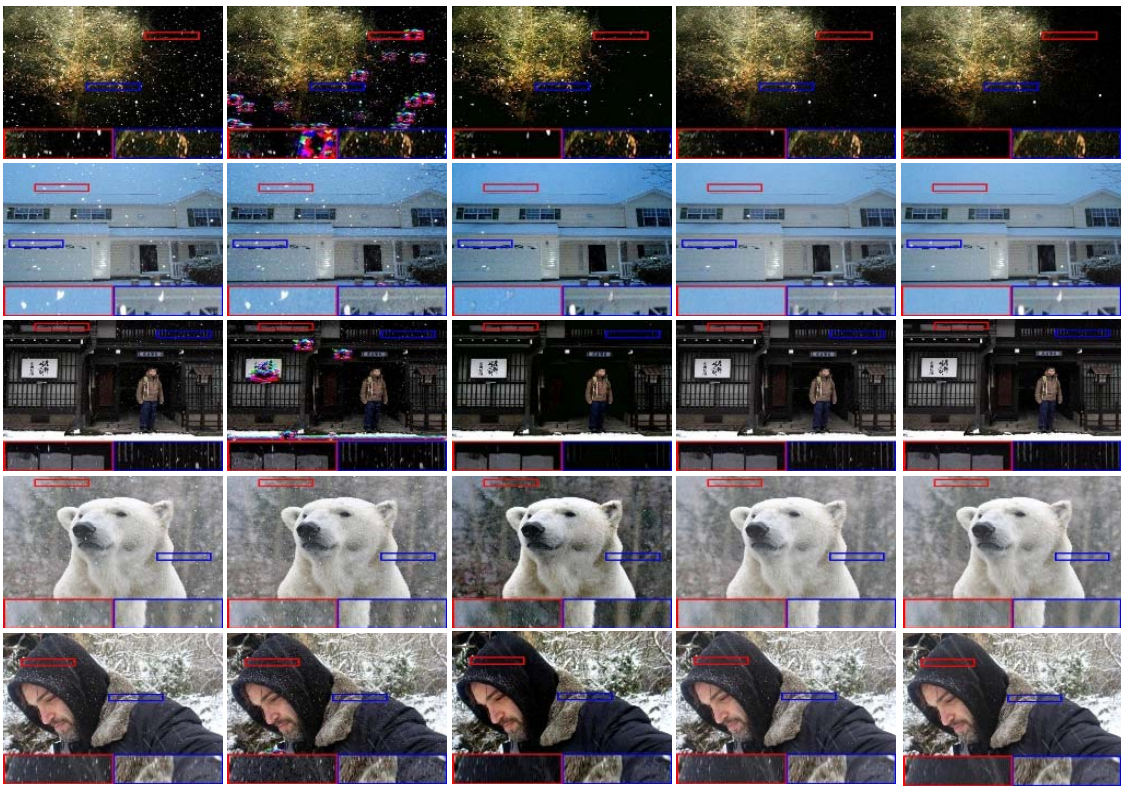
为了验证 HSPA 模块通过相似性权重自适应地调整阈值，有效地区分并过滤雪花噪声干扰，提升无雪场景图像的重建能力，在 CSD 数据集上，选取与其具有相似功能的 LMQFormer 中的 MQTM（mask query Transformer module）进行实验，验证这 2 部分对去雪效果的影响，结果如表 3 所示。可以看出，不使用 MQTM 和 HSPA 模块时，PSNR 只有 34.39 dB，使用 MQTM 后，PSNR 达到 36.03 dB；RASP-Net 的 HSPA 的效果更好，PSNR 达到 37.21 dB，SSIM 达到 0.986 9。

为了验证本文提出的 RBSA 模块、HSPA 模块和循环结构捕获图像语义信息的能力及其对图像去雪的效果，在 CSD 数据集上进行消融实验。以 Swin-Unet[35]基本网络架构为基准，按照本文设置的通道层数和特征图尺寸进行调整，验证这 3 部分



a. Input b. DesnowNet[9] c. HDCWNet[17] d. LMQFormer[11] e. RASP-Net

图 3 在合成数据集上 4 种方法的视觉效果对比



a. Input b. DesnowNet[9] c. HDCWNet[17] d. LMQFormer[11] e. RASP-Net

图 4 在真实数据集上 4 种方法的视觉效果对比

对去雪效果的影响，结果如表 4 所示。可以看出，RBSA 模块能够显著地提升 Swin-Unet 的表现；HSPA 模块通过 ST 操作处理区分并过滤雪花干扰，同时保留图像中的有效内容；循环结构保证了鲁棒

**表 3　2 种模块的消融实验结果对比**

| 实验设置 | PSNR/dB↑ | SSIM↑ |
|---|---|---|
| baseline | 34.49 | 0.980 1 |
| baseline+MQTM | 36.03 | 0.982 1 |
| baseline+HSPA | **37.21** | **0.986 9** |

注：粗体表示最优值。

**表 4　RASP-Net 中各个模块的作用**

| 模块 | PSNR/dB↑ | SSIM↑ |
|---|---|---|
| RBSA | 33.30 | 0.979 1 |
| RBSA+HSPA | 36.26 | 0.984 1 |
| RBSA+循环结构 | 34.49 | 0.980 9 |
| RASP-Net | **37.21** | **0.986 9** |

注：粗体表示最优值。

性。实验结果表明，在 RBSA 和 HSPA 模块的共同作用下，RASP-Net 在 PSNR 和 SSIM 指标上的性能较基线方法有显著的提升。

## 4　结　语

为了有效地应对多尺度雪花去除任务，本文提出一种基于残差块和高相似传递注意力机制的去雪网络 RASP-Net。通过结合 RBSA 构建残差注意力机制，动态地增强有用特征并抑制无用特征，同时保留原始信息；HSPA 的引入实现了 ST 处理，使网络能够动态地区分雪花与非雪花特征，提升去雪的精准度；循环结构迭代地优化特征提取和雪花去除过程，提升了网络的鲁棒性。在合成和真实世界数据集上进行大量实验，证明了 RASP-Net 雪花特征甄别和筛选机制对于去雪结果的有效性，解决了已有去雪模型误判和去除问题，恢复出更接近原始图像的去雪图像，显著地提升了多雪地区的视频监控能力，保障交通监控、户外拍摄、自动驾驶等应用的目标识别与状况判断，提高了可视性和安全性。尽管 RASP-Net 在多尺度雪花去除任务中展现出良好的性能，但在复杂动态场景下的时序一致性建模及计算效率方面仍存在一定不足。未来考虑将该方法扩展至视频级去雪任务，并通过引入更高效的特征建模策略，提升其在真实复杂环境中的应用潜力。

## 参考文献（References）：


[1] Zhao Z Q, Zheng P, Xu S T, *et al*. Object detection with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232

[2] Yu C Q, Wang J B, Peng C, *et al*. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 334-349

[3] Cordts M, Omran M, Ramos S, *et al*. The cityscapes dataset for semantic urban scene understanding[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3213-3223

[4] Ramesh S, Sasikala S, Paramanandham N. Segmentation and classification of brain tumors using modified median noise filter and deep learning approaches[J]. Multimedia Tools and Applications, 2021, 80(8): 11789-11813

[5] Arulkumar V, Prakash S J, Subramanian E K, *et al*. An intelligent face detection by corner detection using special morphological masking system and fast algorithm[C] //Proceedings of the 2nd International Conference on Smart Electronics and Communication. Los Alamitos: IEEE Computer Society Press, 2021: 1556-1561

[6] Su Y T, Shen Z Y, Long X, *et al*. Gaussian filtering method of evaluating the elastic/elasto-plastic properties of sintered nanocomposites with quasi-continuous volume distribution[J]. Materials Science and Engineering: A, 2023, 872: Article No.145001

[7] Kattenborn T, Leitloff J, Schiefer F, *et al*. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 173: 24-49

[8] Xia X, Pan X Z, Li N, *et al*. GAN-based anomaly detection: a review[J]. Neurocomputing, 2022, 493: 497-535

[9] Liu Y F, Jaw D W, Huang S C, *et al*. DesnowNet: context-aware deep network for snow removal[J]. IEEE Transactions on Image Processing, 2018, 27(6): 3064-3073

[10] Jaw D W, Huang S C, Kuo S Y. DesnowGAN: an efficient single image snow removal framework using cross-resolution lateral connection and GANs[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(4): 1342-1350

[11] Lin J H, Jiang N F, Zhang Z T, *et al*. LMQFormer: a laplace-prior-guided mask query transformer for lightweight snow removal[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(11): 6225-6235

[12] Xu J, Zhao W, Liu P, *et al*. Removing rain and snow in a single image using guided filter[C] //Proceedings of the IEEE International Conference on Computer Science and Automation Engineering. Los Alamitos: IEEE Computer Society Press, 2012: 304-307

[13] Zheng X H, Liao Y H, Guo W, *et al*. Single-image-based rain and snow removal using multi-guided filter[C] //Proceedings of the 20th International Conference on Neural Information Processing. Heidelberg: Springer, 2013: 258-265

[14] Ding X H, Chen L Q, Zheng X H, *et al*. Single image rain and snow removal via guided L0 smoothing filter[J]. Multimedia Tools and Applications, 2016, 75(5): 2697-2712

[15] Shi Z H, Li Y W, Zhang C Q, *et al*. Weighted median guided filtering method for single image rain removal[J]. EURASIP Journal on Image and Video Processing, 2018, 2018(1): Article No.35

[16] Zang J F, Xu N X, Liu R, *et al*. Deraining and desnowing algorithm on adaptive tolerance and dual-tree complex wavelet fusion[J]. International Journal of Interactive Multimedia and Artificial Intelligence, 2020, 6(4): 131-137

[17] Chen W T, Fang H Y, Hsieh C L, *et al*. All snow removed: single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 4176-4185

[18] Luo Y, Xu Y, Ji H. Removing rain from a single image via discriminative sparse coding[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 3397-3405

[19] Li Z, Zhang J, Fang Z J, *et al*. Single image snow removal via composition generative adversarial networks[J]. IEEE Access, 2019, 7: 25016-25025

[20] Han K, Xiao A, Wu E H, *et al*. Transformer in transformer[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.1217

[21] Cheng B D, Li J C, Chen Y, *et al*. Snow mask guided adaptive residual network for image snow removal[J]. Computer Vision and Image Understanding, 2023, 236: Article No.103819

[22] Quan Y H, Tan X H, Huang Y, *et al*. Image desnowing via deep invertible separation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(7): 3133-3144

[23] Wang B, Chen Z K, Zhang L, *et al*. MSA-former: multi-scale adaptive transformer for image snow removal[C] //Proceedings of the 31st International Conference on Multimedia Modeling. Heidelberg: Springer, 2025: 45-58

[24] Dai X W, Zhou Y B, Qiu X T, *et al*. Parameter-efficient fine-tuning for single image snow removal[J]. Expert Systems with Applications, 2025, 265: Article No.125901

[25] Ren J W, Zhang M Y, Yu C J, *et al*. Balanced MSE for imbalanced visual regression[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 7916-7925

[26] Paul S, Jhamb B, Mishra D, *et al*. Edge loss functions for deep-learning depth-map[J]. Machine Learning with Applications, 2022, 7: Article No.100218

[27] Blanchard T, McGrath R E, Jayawickreme E. Resilience in the face of interpersonal loss: the role of character strengths[J]. Applied Psychology: Health and Well-Being, 2021, 13(4): 817-834

[28] Chen W T, Fang H Y, Ding J J, *et al*. JSTASR: joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 754-770

[29] Keleş O, Yılmaz M A, Tekalp A M, *et al*. On the computation of PSNR for a set of images or video[C] //Proceedings of the Picture Coding Symposium. Los Alamitos: IEEE Computer Society Press, 2021: 1-5

[30] Bakurov I, Buzzelli M, Schettini R, *et al*. Structural similarity index (SSIM) revisited: a data-driven approach[J]. Expert Systems with Applications, 2022, 189: Article No.116087

[31] Zhang T, Jiang N F, Lin J H, *et al*. DesnowFormer: an effective transformer-based image desnowing network[C] //Proceedings of the IEEE International Conference on Visual Communications and Image Processing. Los Alamitos: IEEE Computer Society Press, 2022: 1-5

[32] Chen S X, Ye T, Liu Y, *et al*. SnowFormer: context interaction transformer with scale-awareness for single image desnowing[OL]. [2025-07-04]. https://arxiv.org/abs/2208.09703

[33] Cheng Y R, Ren H, Zhang R, *et al*. Context-aware coarse-to-fine network for single image desnowing[J]. Multimedia Tools and Applications, 2024, 83(18): 55903-55920

[34] Dai X W, Zhou Y B, Qiu X T, *et al*. WaveFrSnow: comprehensive perception wavelet transform frequency separation transformer for image snow removal[J]. Digital Signal Processing, 2024, 155: Article No.104715

[35] Cao H, Wang Y Y, Chen J, *et al*. Swin-Unet: unet-like pure transformer for medical image segmentation[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2022: 205-218