arXiv:2305.00691v1 [cs.CV] 1 May 2023

# 基于多尺度视网膜皮层理论与多任务学习的热红外图像色调映射和去噪联合处理

Axel G¨odrich[a]、Daniel K¨onig[b]、Gabriel Eilertsen[c] 和 Michael Teutsch[b]

[a]纽伦堡技术大学，德国纽伦堡[b]Hensoldt Optronics GmbH，德国奥伯科亨[c]瑞典林雪平大学

**摘要**

相机将真实场景数字化为像素强度值，而这些值的范围受可用的每像素位数（bpp）限制。高动态范围相机通过增加每像素位数，以更高的分辨率捕获这些亮度值。然而，大多数显示器仅支持 8 bpp。朴素的高动态范围压缩方法会导致高动态范围图像中丰富信息的丢失。本文研究了适用于 16 bpp 热红外图像、能够保留这些信息的色调映射算法。优化的多尺度视网膜皮层理论算法作为基线。随后，采用基于广受欢迎的 U-Net 架构的深度学习方法对该算法进行近似。通过利用自监督深度学习方法，可隐式减少图像经过色调映射后残留的噪声；该方法还能与色调映射方法结合，在多任务学习方案中进行联合训练。本文还进一步探讨了色调映射背景下，热红外视频增强中的去噪和去闪烁问题。在公开的 FLIR ADAS 数据集上进行的大量实验，证明了所提方法相较于最先进方法的有效性。

**关键词：** 视频增强、图像预处理、长波红外、深度学习、深度卷积神经网络

## 1. 引言

现实世界中的场景可能呈现出极宽的亮度范围，远超任何摄影传感器所能捕获的范围。相机将此类场景数字化为像素强度值，其取值范围受可用的每像素位数（bpp）限制。高动态范围（HDR）相机通过增加每像素位数，能够以更高分辨率捕获这些亮度值，从而更精细地离散化这些值，呈现更多细节。这在热红外（TIR）成像中尤为重要，因为这类成像只有一个强度通道，因此需要较高的每像素位数才能最大限度地提高对比度[1]。然而，要在标准显示器上呈现这些内容也面临着挑战，因为大多数显示器仅支持每像素 8 位。图1 展示了一些示例图像。朴¨素的高动态范围压缩方法（如线性下采样）会导致高动态范围图像中丰富信息的丢失。即使是对比度受限自适应直方图均衡化（CLAHE）[2] 等广受欢迎的图像对比度增强方法，也无法提供令人满意的结果。色调映射算子（TMO）旨在保留丰富信息，同时生成视觉效果出色的图像或视频[3,4]。本文研究了这类色调映射算子在压缩每像素 16 位的热红外图像和视频中的应用。目前，专门处理热红外光谱中色调映射的方法实际上寥寥无几[1,5]。受 Luo 等人[5], 启发，本文实现并优化了原本用于对比度增强的多尺度视网膜理论（MSR）算法，将其作为此用途的参考方法（见图1）。随后，我们采用基于热门 U-Net 架构的深度学习方法，通过全监督训练方案对这一优化后的 MSR 算法进行近似或模仿[7] 。这种模仿方法不仅可以通过近似处理节省时间[8], ，我们还将利用它高效地融入自监督去噪，以联合学习色调映射和图像增强。色调映射过程往往会大幅放大噪声，使其再次变得可见[4]。因此，若要生成高质量图像，在色调映射过程中降噪或之后主动去噪都至关重要

作者电子邮件：axel.goe drich@th-nuernberg.de, {daniel.koenig, michael.teutsch}@hensoldt.net, gabriel.eilertsen@liu.se




图 1. 本文采用的朴素线性下采样、对比度受限自适应直方图均衡化与基于优化的多尺度视网膜理论的参考色调映射算子之间的比较。¨示例图像取自公开的 FLIR ADAS 数据集[12]。

图像。我们提出利用自监督深度学习方法[9] 隐式减少图像中的残余噪声，并使其能够与色调映射方法联合训练。联合进行去噪和色调映射并非全新思路[10,11]，，但我们的方法专为热红外图像设计，采用最新的深度学习方法，并且得益于自监督去噪，无需完全监督。在公开的 FLIR ADAS 数据集* 上开展的大量实验，与最先进方法进行了全面比较，并证明了所提方法的有效性。我们的贡献可概括为：（1）优化 MSR 算法，为热红外图像和视频生成参考色调映射算子；（2）采用基于深度学习的 U-Net 神经网络架构，通过监督训练方案对该参考色调映射算子进行近似；（3）在 U-Net 架构中融入自监督去噪，以联合开展色调映射与去噪的多任务学习；（4）深入探讨自监督去噪与去闪烁。我们对各部分进行了详细分析，并在公开的 FLIR ADAS 数据集[12]上开展了全面实验，给出了相应结果。

本文其余部分组织如下：第2节回顾相关工作；第3节介绍并评估基于传统计算机视觉的优化 MSR 算法；第4节讨论结合隐式去噪的基于深度学习的色调映射，并给出大量实验结果。与最先进方法的比较见第5节。结论见第6节。

## 2. 相关工作

**热红外光谱中的色调映射：**过去三十年来，人们提出了许多色调映射方法[4, 13–15]，但这些方法旨在处理传统彩色可见光图像。由于热红外图像通常只有一个强度通道，因此这里不考虑基于典型颜色转换或颜色恢复过程的方法。不过，由于图像的亮度通道通常需要进行色调映射，因此可以借鉴可见光光谱中的方法。实时色调映射算子[16]是一种基于传统计算机视觉的色调映射算法。在一项对不同色调映射算子的比较研究中，就整体图像质量而言，它仍是最先进的色调映射算子[4]。该方法曾用于热红外图像的色调映射，取得了良好效果，证明了其适用于这项任务[1]。MSR[6] 是另一种取自可见光光谱、基于传统计算机视觉的方法。它源自 Land 和 McCann[17]，提出的视网膜皮层理论，该理论旨在模拟人类视觉系统（HVS）如何感知场景中的颜色。他们指出，人类视觉系统以相对的方式感知亮度，即感知局部图像区域中的亮度变化，而非绝对值。该算法的基本思想是将图像分解为反射率和照明。像素的照明值 $x$ 近似为 $x$ 的值与周围像素平均值之比。随后，通过从原始图像中减去照明图像来计算反射率。Hyunchan 等人[18] 采用 MSR 算法的改进版本，对高动态范围可见光图像进行色调映射。Luo 等人[5] 将条件高斯滤波器（CGF）成功融入 MSR 方法，用于热红外图像的色调映射，证明了该方法适用于这项任务。其他方法[19–21] 大多基于对比度增强技术，而这类技术被认为不太适合用于色调映射[1]。

*我们使用 FLIR ADAS D　数据集版本 1.3，2019年8月16日（请参见：https://adas-dataset-v2.flirconservator.com/）。

**基于深度学习的色调映射：**传统计算机视觉方法仍广泛用于可见光图像和视频的色调映射[4, 5, 16, 22–24]。然而，近年来，越来越多的研究致力于采用基于深度学习的方法来完成这项任务[8, 25–29]。其主要优势通常被认为是效率的提升：理论上，可以利用深度卷积神经网络（DCNN）近似传统色调映射算子，从而实现这一优势[8]。许多色调映射算法面临的一个主要问题是计算成本高昂。尽管人们努力缩短计算时间，但即使在高性能硬件上，目前许多色调映射算子仍无法实时运行。Gharbi 等人[29] 针对这一问题，提出将传统方法与基于深度学习的方法相结合，以实现智能手机上的实时色调映射和通用图像增强；对于分辨率为 $1920 \times 1080$ 像素的图像，其帧率可达每秒 40–50 帧（fps）。深度学习在色调映射领域的另一潜在价值，在于提升生成图像的质量。Patel 等人[27] 采用基于生成对抗网络（GAN）的训练策略，对多个色调映射算子的混合模型进行训练；与参考方法本身相比，该模型在许多测试图像上取得了更高的色调映射质量指数（TMQI）分数。此外，通过采用不同的损失函数和自监督学习方法，深度卷积神经网络的结果有望进一步改善。尽管面向热红外图像的通用色调映射仍是一个小众课题[5], , 文献中几乎没有针对热红外图像的基于深度学习的色调映射研究。Teutsch 等人[1] 基于 Chen 等人[8],的工作，成功训练了上下文聚合网络（CAN）架构，以监督方式近似实时色调映射算子[16] 。

**自监督去噪：** 如今，最先进的去噪方法基于深度学习，并且通常优于基于传统计算机视觉的去噪方法。这些基于学习的去噪算法中，大多数[30–32] 使用监督学习，并以干净数据作为真实值（GT）。训练数据由无噪声、通常为合成的图像组成，这些图像会被人工噪声污染。机器学习（ML）模型以含噪图像作为输入，并以相应的干净图像作为目标。然而，在实际情况下，噪声特征通常与人工噪声模型不同，合成图像与真实图像之间也会产生领域差异。近年来，自监督图像去噪方法越来越受欢迎。Noise2Noise[33]是该领域最重要的论文之一。该方法基于这样的假设：噪声在图像中随机分布，并且在同一场景的连续帧之间有所变化。这样，随着时间推移，无限多幅图像中的噪声会平均为零。他们不再使用干净图像进行训练，而是使用同一场景中对齐的连续含噪图像对进行训练。与完全监督的方法相比，该方法能够达到相近的性能。由于像素级精确的图像对齐这一前提较为苛刻，Ehret 等人[34] 针对这一局限性，将该原理迁移到了视频去噪领域。在连贯的视频中，通常已经包含同一场景的多幅图像，但这些图像在时间上彼此分离。为克服这种时间上的分离，Ehret 等人[34] 使用一对连续帧的光流，将上一帧变形回当前帧，从而消除运动。理论上，这会得到一对表示同一场景但具有不同噪声模式的图像。随后，可以基于 noise2noise[33] 方法训练去噪模型。Ehret 等人[34] 将这种方法称为帧间训练。该方法的一个限制因素是依赖精确的稠密光流。较新的基于深度学习的去噪方法则着重于仅使用含噪图像、无需成对对应关系来训练深度卷积神经网络。Noisier2Noise[35] 和 Noisy-As-Clean[9] 向原始含噪图像中添加额外噪声。随后，这一图像对可用于训练深度卷积神经网络。Noisier2Noise 方法的工作原理如下。设自然含噪图像 $Y$ 由无噪声的干净图像 $X$ 与应被去除的自然噪声 $N$组合而成。由于只有 $Y$已知，目标是获得对 $N$的预测。为此，在含噪图像 $Y$ 上添加合成噪声 $M$，形成噪声更强的版本 $Z$。随后，将 $Z$ 作为输入提供给深度卷积神经网络，同时将 $Y$ 作为参考。然后，深度卷积神经网络学习将 $Z$ 映射为 $Y$。由于深度卷积神经网络只能将输入图像作为整体进行处理，因此无法区分$N$ 和 $M$，所以生成的输出直观上并不完美，而是处于输入图像与参考图像之间。接着，从输入 $Z$ 中减去预测输出，以估计预测结果中仍然存在的噪声。再从预测结果中减去该噪声估计值，从而得到最终的去噪图像。与此方法类似，Xu 等人[9] 提出了另一种方法，称为 Noisy-As-Clean。与 Noisier2Noise 方法一样，将自然含噪图像 $Y$ 作为参考，并将经过合成噪声污染的版本 $Z$ 作为输入提供给深度卷积神经网络。在训练过程中，深度卷积神经网络学习从 $Z$映射到 $Y$。然而，该方法在推理处理方式上不同于 Noisier2Noise 方法。在推理时，将自然含噪图像 $Y$ 输入训练好的模型，但不添加额外噪声，并将该模型生成的输出视为最终的去噪图像。Noisier2Noise 方法和 Noisy-As-Clean

方法的共同点在于，它们完全以自监督方式进行训练，并且可以直接基于单幅图像进行训练。不需要对齐图像对或光流方法。本文在热红外图像和视频上评估并比较了上述几种自监督方法。

**去闪烁：** 在单幅图像上训练深度卷积神经网络，然后将其应用于视频数据时，可能会出现闪烁问题[36]。假设场景中不存在剧烈的光照变化，则连续图像应具有大致相似的亮度值。Zhang 等人[37] 利用这一原理构建损失函数，并为低照度图像增强任务强制施加时间一致性。Lai 等人[38] 利用相同原理训练深度卷积神经网络，以完成去闪烁任务。

**多任务学习：** 多任务学习的核心思想是，存在一些彼此足够相似的计算机视觉任务，可以对它们进行联合学习[39]。通过这种方式，可以进一步提升每个单独任务的性能；例如，Mask R-CNN[40] 中的基于深度学习的特征提取就受益于目标检测和实例分割的联合学习。对超分辨率、低照度增强、去雾、去噪或色调映射等图像增强任务进行联合学习[8,10,41–43]显然很有前景。尽管可以预期多任务学习同样能够改善热红外图像增强，但截至目前，相关文献仍然很少[44]。本文研究色调映射、去噪和去闪烁的联合学习，用于热红外图像与视频增强。

## 3. 基于多尺度Retinex的色调映射

为了为后续实验建立基准，我们采用了一个现有的热红外图像和视频色调映射评估框架[1]，，并且最近为其扩展了 TMQI 的 Python 实现[†]。此外，我们使用公开可用的 FLIR ADAS 数据集[12]。我们定量评估了基线色调映射算子方法 Realtime TMO[16]，、CGF[5]，、MSR[6] ，以及用于生成 FLIR 数据集 8 位图像的原始 FLIR 色调映射算子。由于 CGF 没有公开代码，我们重新实现了该方法，并在本文中始终使用自己的代码。我们针对热红外光谱调整了 Realtime TMO 的实现：现在它处理的是单通道热红外图像，而非三通道彩色可见光图像，并且调整了参数设置。此外，据我们所知，FLIR 色调映射算子的代码并未公开，其算法也未在论文中描述。不过，该数据集包含原始的 16 位图像和经过色调映射的 8 位图像，因此即使不知道生成 8 位图像的算法，我们仍然可以使用这些图像进行评估。

这些色调映射算子在该数据集的视频子集上进行了评估。结果如表1所示。噪声可见度指标需要另一组图像，这些图像是在原始高动态范围图像的含噪版本上进行色调映射得到的。由于 FLIR 色调映射算子并未公开，因此无法创建这组图像，进而无法将噪声可见度指标应用于 FLIR ADAS 数据集的 8位图像。每项指标的最佳结果以粗体显示。可以明显看出，8位 FLIR 图像在评估框架的大多数指标上表现最佳。CGF 在全局时间不一致性指标和过曝光指标上表现更好，但在欠曝光指标上的结果最差，这表明像素值偏向较低的取值范围。比较 MSR 与实时色调映射算子可以发现，两者的表现非常相似，其中基线 MSR 在噪声可见度指标和色调映射质量指数方面略胜一筹。

表 1。比较不同色调映射算子以形成当前基线的定量评估。

| | **FLIR[12]** | **实时色调映射算子 [16]** | **CGF[5]** | **MSR[6]** |
|---|---|---|---|---|
| 色调映射质量指数 ↑ | **0.951** | 0.809 | 0.850 | 0.866 |
| 曝光不足 ↓ | **0.535** | 0.755 | 2.842 | 1.191 |
| 过曝 ↓ | 0.899 | 1.628 | **0.030** | 1.578 |
| 全局对比度损失 ↓ | **-0.162** | -0.123 | -0.084 | -0.148 |
| 局部对比度损失 ↓ | **-0.0417** | -0.0236 | -0.022 | -0.027 |
| 噪声可见性 ↓ | - | 30.77 | 30.04 | **28.79** |
| 全局时间不一致性 ↓ | $1.7e^{-4}$ | $3.7e^{-4}$ | **2。$8e^{-5}$** | $2.7e^{-4}$ |
| 局部时间不一致性 ↓ | **0.0048** | 0.0157 | 0.0088 | 0.0394 |

[†]https://github.com/HensoldtOptronicsCV/ToneMappingIQA

图2展示了对各色调映射算子的定性比较。8位 FLIR 图像在视觉上颇具吸引力，具有较高的对比度，并且充分利用了 8位取值范围。实时色调映射算子以及尤其是基线 MSR 算法的结果对比度较弱。将 8位 FLIR 图像的结果与 CGF 和基线 MSR 的结果进行比较可以发现，基于视网膜皮层理论的算法生成的图像噪声相对较大，但也更加清晰锐利。总的来说，定性评估大致证实了使用公开评估框架进行定量评估所得的结果。

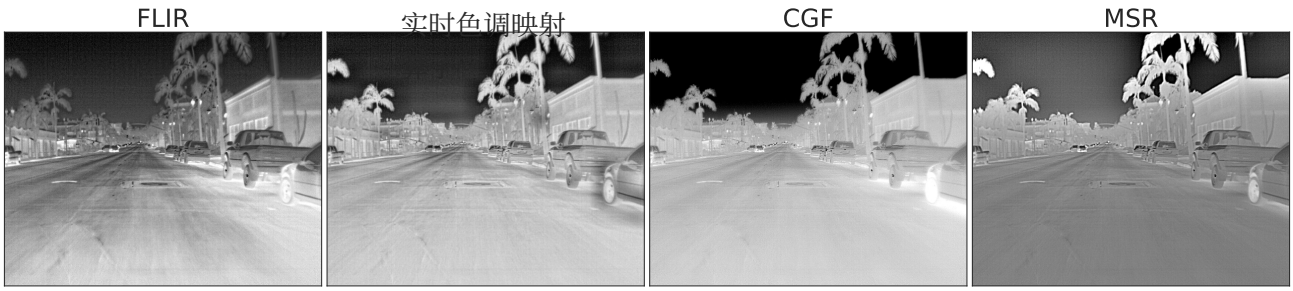


图 2. 提供当前基线的不同色调映射算子之间的定性比较。

虽然整体表现通常不如 8 位 FLIR 图像，但 MSR 算法展现出了潜力，尤其是在图像锐度方面。该算法现阶段的主要弱点是对比度低、图像噪声放大，以及仅在视频流中明显的闪烁伪影。不过，这些弱点有望得到改善。因此，接下来的章节旨在优化基线 MSR 算法。

### 3.1 图像对比度增强

MSR 算法最明显的缺点之一是对比度低且外观单调。图像中外观明亮的高温物体很容易被发现。由于对比度较弱，较暗物体或亮度相近物体的细节几乎不可见。为了从 HDR 图像中尽可能多地提取信息，应突出这些区域中细微的亮度差异。因此，原始 MSR 算法[6] 有一个名为保色多尺度视网膜理论（MSRCP）的变体。其思路是执行一种名为最简色彩平衡（SCB）[45] 的算法，以充分利用每个颜色通道的亮度范围。由于本文使用单通道热红外图像，所提出的 SCB 算法实际上并不太有用。相反，一种可能的图像细节增强方法是直方图均衡化（HE）[46]。HE 旨在扩展图像直方图中亮度值的分布范围。这通常会提高全局对比度，从而生成视觉效果更好的图像。然而，这种方法作用于整个图像，因此往往无法有效增强局部图像区域的对比度。为进一步提高局部对比度，可以使用自适应直方图均衡化（AHE）[46] 。该方法采用分块处理，即将输入图像划分为多个区域，然后分别对每个区域执行 HE。其缺点是容易增加图像噪声。对比度受限自适应直方图均衡化算法[2] 旨在缓解这一问题。该算法会在达到预设阈值后限制对比度增强。

上述所有方法均作为色调映射后的后处理步骤实现。我们还测试了将其作为预处理步骤的另一种实现方式，但输出图像没有出现明显差异。图 3 展示了从 FLIR ADAS 数据集获取的一幅示例图像，该图像分别经过每种方法处理，并附有相应的直方图。左侧第一幅图像来自采用基线 MSR 算法进行色调映射的 FLIR 数据集，作为参考。在图像下方的相应直方图中可以看到，大多数灰度值集中在一个狭窄范围内，表明对比度较弱。对该图像应用全局 HE 后，对比度得到明显增强，但直方图中也出现了一些间隙，这意味着数值范围仍未得到充分利用。AHE 能够使数值分布更加均匀，但全局差异不那么明显，噪声（主要在天空区域可见）却被大幅放大。在此示例中，我们特意选择了较小的分块尺寸，以进一步突出所比较方法之间的差异。使用前述对比度受限自适应直方图均衡化算法限制对比度（分块尺寸与 AHE 相同）后，直方图中的分布不那么激进，

整体图像也更具观感。采用不进行分块的全局对比度受限自适应直方图均衡化实现，虽然对局部对比度的提升不如前者明显，但引入的噪声更少，并且在主观上呈现出最具视觉吸引力的结果。因此，全局版本的对比度受限自适应直方图均衡化算法将作为 MSR 算法后的后处理步骤使用。这样既能明显改善对比度，同时还能保持自然的外观。

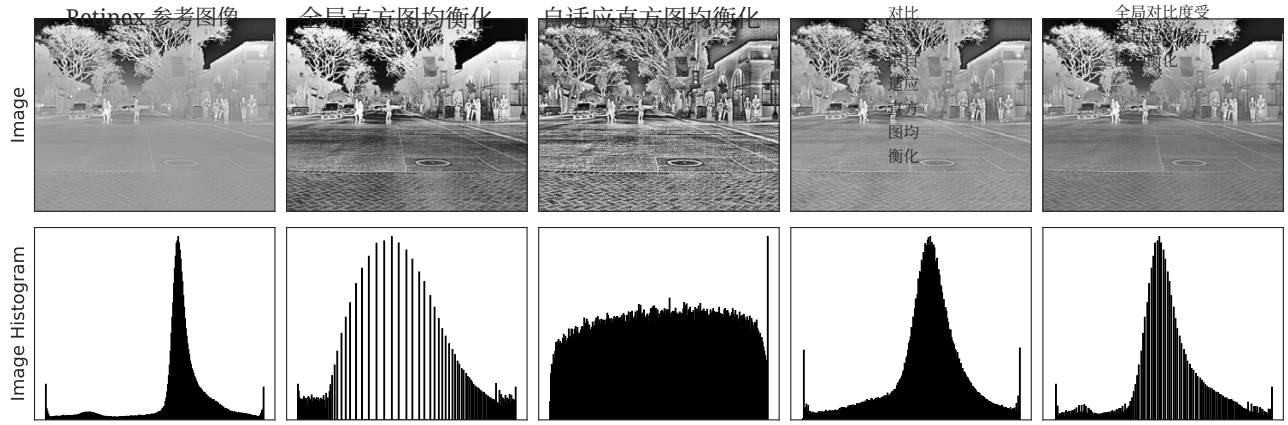


图3。不同直方图均衡化方法的比较。处理后的图像显示在上排，相关直方图显示在下排。

### 3.2 去闪烁

MSR 算法的一个缺点是容易引入闪烁。连续视频帧之间图像平均亮度的快速变化会显得十分明显，令人不适。文献中将这类不希望出现的全局亮度变化称为闪烁[47]。针对单幅图像处理设计的色调映射算子应用于视频时，很可能会引入闪烁伪影。

有多种方法可以减少闪烁。本文研究了两种不同的方法。第一种方法基于直方图匹配，这是一种颇受欢迎的去闪烁方法[48–50]。这种方法旨在变换图像的亮度值，使变换后的直方图与参考直方图相匹配[46]。我们将前 100 个视频帧的平均直方图作为参考直方图。随后，将当前帧的直方图与该参考直方图进行匹配，并使用查找表变换图像强度。结果如图 4所示。用于比较的帧特意选自 FLIR ADAS 数据集中全局图像亮度快速变化的场景。为更清楚地展示差异，图中仅显示每隔五帧的一帧。上排为采用增强对比度的 MSR 方法进行色调映射后的参考帧，第二排显示相应的直方图。第三排为通过直方图匹配实现的所提出去闪烁方法的结果，底排显示相应的直方图。可以看出，全局图像亮度的快速变化随时间得到平滑，同时仍保持自然的外观，且不会引入任何可见伪影。

通过观察图像平均亮度值随时间的变化，可以检测闪烁。平均值较高通常表示全局图像亮度较高，即图像较亮，反之亦然。连续视频帧之间的平均亮度值变化明显，表明存在闪烁。图 5展示了在 FLIR ADAS 数据集的视频子集上评估多种不同去闪烁方法的结果。原始多尺度视网膜论文中提出的 SCB[6] 或对比度受限自适应直方图均衡化都不足以补偿闪烁。Luo 等人[5] 则基于图像的标准差进行裁剪。首先计算图像的均值和标准差，然后将亮度值超过标准差三倍的部分裁剪掉。这种非常简单但有效的方法无需手动选择裁剪阈值，在文献中也称为三西格玛法则。[51]本文后续将此方法称为西格玛裁剪。从图 5中可以看到，西格玛裁剪还会引入

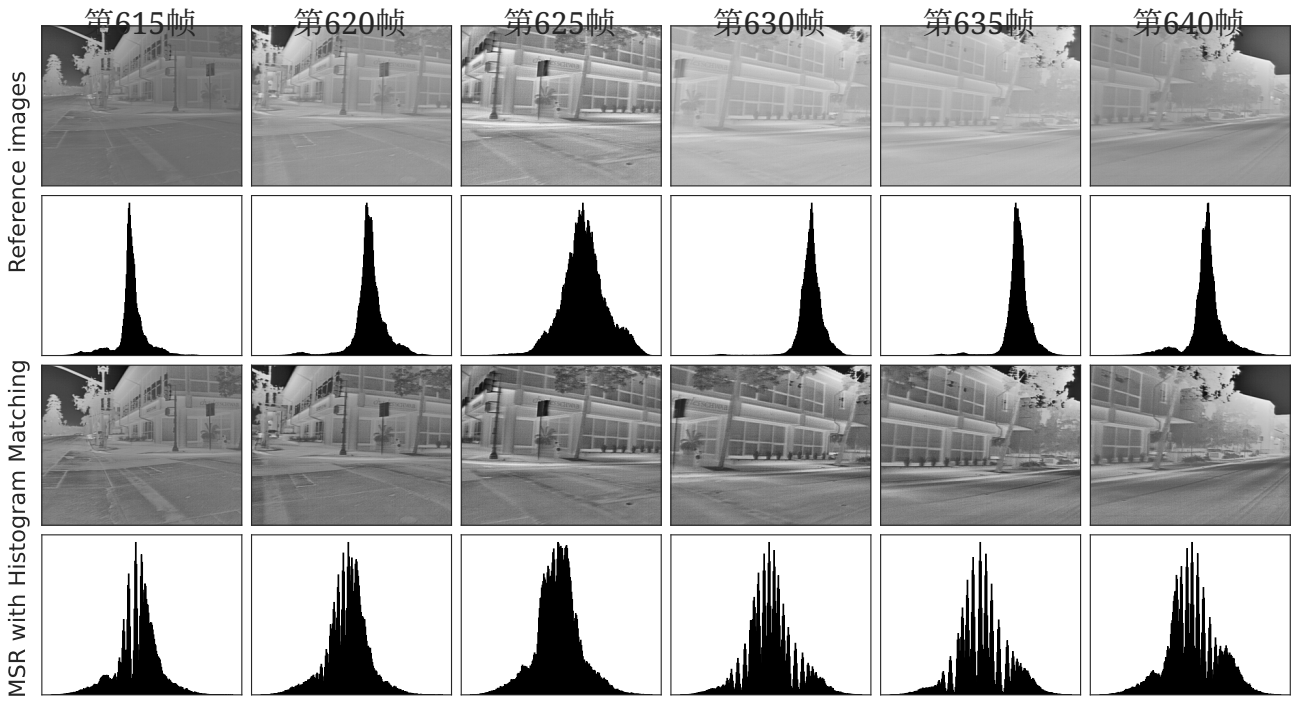


图4. 通过直方图匹配进行去闪烁的 MSR 算法与未进行去闪烁的 MSR 算法的比较。示例图像取自 FLIR ADAS 数据集 *video* 子集。

某种均衡或归一化效果，使图像平均亮度值保持在 127 附近。一方面，这有利于去闪烁；另一方面，也可能无法灵活捕捉场景全局温度的变化。然而，FLIR ADAS 数据集未提供可用于训练机器学习方法的真实视频数据。相反，其训练子集提供的是跳帧图像序列。因此，在第 4 节进行机器学习时，我们无法采用基于直方图匹配的去闪烁方法。因此，在生成基于 MSR 的真实值时，我们将采用西格玛裁剪，以便用深度神经网络近似参考色调映射算子。

### 3.3 优化的多尺度视网膜理论

本节对所分析的优化方法进行了定量评估，结果见表2。除了最先进方法 FLIR、实时色调映射算子、CGF 和基线 MSR 外，我们还纳入了第3.1 节介绍的使用全局对比度受限自适应直方图均衡化进行的对比度增强，以及第3.2节讨论的西格玛裁剪。将全局对比度受限自适应直方图均衡化用于对比度增强，并将西格玛裁剪用于去闪烁，可得到优化版 MSR 算法。这两种优化方法协同效果良好，显著提升了图像对比度、时间一致性和色调映射质量指数（TMQI），但代价是噪声可见性增强。因此，下一节将探讨不同的基于机器学习的降噪方法。本文其余部分均将优化的 MSR 用作参考色调映射算子。

## 4. 基于深度学习的色调映射

我们采用深度学习进行色调映射的动机在于，可以将更多执行图像增强（例如去噪）的方法高效地融入训练算法。因此，我们首先需要训练一个深度神经网络，使其能够近似或模拟相关的传统计算机视觉色调映射算子。我们通过两项独立实验，分别测试对优化的 MSR 和 FLIR 色调映射算子的模拟效果。因此，在训练过程中，优化的 MSR 算法和 FLIR 色调映射算子生成的 8 位输出图像作为真实值。由此，使用深度卷积神经网络进行色调映射的基本监督学习方法非常直接。首先，使用首选的色调映射算子对一组高动态范围图像进行色调映射，并将其用作训练过程中的参考图像或目标。其分辨率为

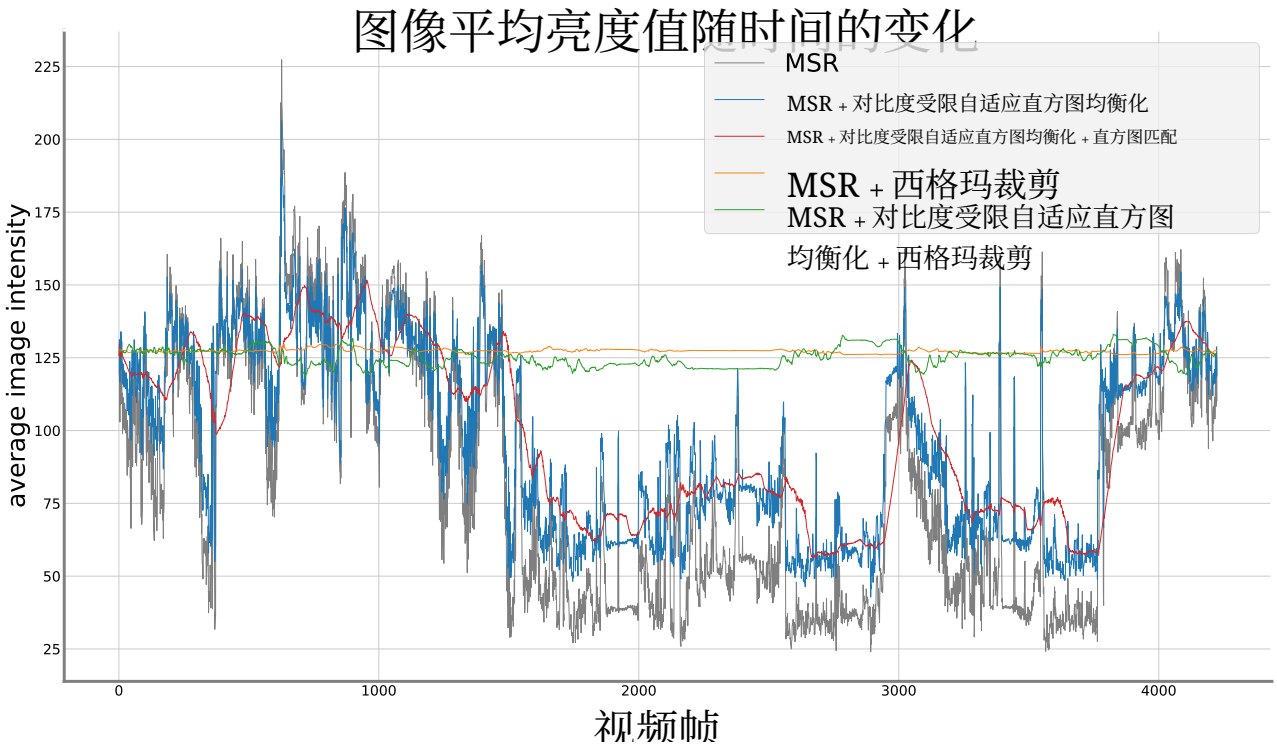


图5. 使用采用不同 CLAHE、西格玛裁剪和直方图匹配组合的 MSR 算法对 FLIR 视频数据集进行色调映射后，图像平均亮度值随时间的变化。西格玛裁剪和直方图匹配均能平滑图像平均亮度随时间的变化，从而实现去闪烁。表2. 最先进方法中的色调映射算子与采用不同优化的基线 MSR 算法之间的比较。所提出的优化的 MSR 在基于视网膜皮层理论的算法中表现最佳，尤其是在 TMQI 指标方面。

| 色调映射质量指数↑ | **FLIR**[12] | **实时 TMO**[16] | **CGF**[5] | **MSR**[6] | **MSR+** 对比度受限自适应直方图均衡化 | **MSR+ 西格玛裁剪** | **优化 MSR** |
|---|---|---|---|---|---|---|---|
| | **0.951** | 0.809 | 0.850 | 0.866 | 0.887 | 0.921 | 0.923 |
| 曝光不足↓ | **0.535** | 0.755 | 2.842 | 1.191 | 1.212 | 1.685 | 1.803 |
| 过曝↓ | 0.899 | 1.628 | **0.030** | 1.580 | 2.437 | 0.860 | 1.176 |
| 全局对比度损失↓ | -0.162 | -0.123 | -0.084 | -0.148 | **-0.204** | -0.151 | -0.202 |
| 局部对比度损失↓ | -0.0417 | -0.0233 | -0.022 | -0.0270 | -0.0425 | -0.0290 | **-0.0447** |
| 噪声可见性↓ | - | 30.77 | 30.04 | **28.79** | 30.61 | 29.50 | 30.67 |
| 全局时间不一致性↓ | $1.7e^{-4}$ | $3.6e^{-4}$ | $2.8e^{-5}$ | $2.7e^{-4}$ | $1.5e^{-4}$ | **$1。6e^{-5}$** | $2e^{-5}$ |
| 局部时间不一致性↓ | **0.0048** | 0.0158 | 0.0088 | 0.0394 | 0.0253 | 0.0389 | 0.025 |

深度卷积神经网络的输入层为 640×512个像素，这也是 FLIR ADAS 数据集的原始分辨率。随后，将相同的高动态范围图像输入深度卷积神经网络进行处理，生成输出图像。将该输出图像与相应的参考图像进行比较，并通过损失函数计算两幅图像之间的误差或差异，再利用反向传播训练深度卷积神经网络。与其他常见的基于深度学习的图像增强方法一样[8,52]，我们采用均方误差（MSE）损失。

我们考虑了两种不同的深度卷积神经网络（DCNN）架构：U-Net[7] 和 CAN[53]。U-Net 架构最初于 2015 年提出，用于医学图像分割[7]。最近的一项研究[54]，比较了不同的 U-Net 架构，以及其他用于去噪任务的最先进模型。标准 U-Net 架构的结果至少与其他模型相当，而采用组归一化[55] 则提升了 U-Net 架构的性能，使其能够稳定地胜过其他最先进的去噪模型。Montulet 和 Briassouli[52] 还成功地将 U-Net 架构用于色调映射，并采用基于生成对抗网络（GAN）的训练方法。对于

实验中，我们基于原始架构，使用 PyTorch 实现了 U-Net 架构[7]。原始架构与我们的实现之间的主要区别在于输入层和输出层；为适应灰度图像的色调映射任务，我们对这两层进行了调整。此外，我们还加入了批归一化层，随后又加入组归一化层，以提升整体性能。我们通过自动超参数调优，调整模型宽度（第一层中特征图的数量）和深度（连续层的数量），使其逼近优化的 MSR 色调映射算子。对于这一特定任务，深度为 6、宽度为 16 时效果最佳，因此整个实验均采用这一设置。

CAN 架构最初是为语义分割开发的[53]。Chen 等人[8] 通过将这一架构成功应用于不同的图像处理任务，证明了它的通用性，其中一项任务是对可见光图像进行色调映射。他们还在网络中引入了自适应批归一化步骤，进一步提升了整体性能和架构的通用性。然而，在我们对深度卷积神经网络架构的实现中，这种自适应批归一化在训练过程中导致了稳定性问题。因此，我们最初使用批归一化，随后改用组归一化，从而彻底消除了使用批归一化时出现的填充伪影。填充伪影表现为图像边缘处令人不适的水平线和垂直线。这似乎是空洞卷积造成的，而空洞卷积在深度卷积神经网络架构中得到了广泛应用。CAN 架构基于公开提供的 TensorFlow 实现[8]，并使用 PyTorch 实现。我们通过自动超参数调优，调整模型深度（连续层数）和模型宽度（每层的特征图数量），以近似优化的 MSR 色调映射算子。宽度为 24、深度为 7 时效果最佳。

U-Net 和 CAN 模型均训练了 32 个训练轮次。由于我们的目标是同时仿效优化的 MSR 算法和 FLIR 色调映射算子，因此使用优化的 MSR 输出的 8位图像，以及 FLIR ADAS 数据集中的原始 8位图像。延长训练时间并未显著提升模型性能；与批归一化相比，使用组归一化使两种架构的性能均得到稳定提升。从图 6所示的定性结果来看，CAN 架构很好地仿效了这两种原始方法，而 U-Net 在 FLIR 数据集上的表现不佳，生成的输出图像更暗淡、色彩也不够鲜艳。不过，对于优化的 MSR 的大多数测量指标，U-Net 的表现略优于 CAN，噪声可见性指标除外，如表 3所示。总体而言，两种方法都适用于近似优化的 MSR：CAN 在 FLIR ADAS 数据集上的表现更好，而 U-Net 在优化的 MSR 色调映射算子上的表现略好。由于 U-Net 也广泛用于去噪，因此本文后续部分选择 U-Net 作为深度卷积神经网络架构。

表 3. FLIR 色调映射算子、优化的 MSR 以及我们使用 CAN 和 U-Net 架构进行深度学习近似的比较。根据这些指标，两种方法都能实现近似。根据色调映射质量指数，模仿优化的 MSR 效果略好。

| | **FLIR**[12] | **CAN FLIR** | **U-Net FLIR** | **优化 MSR** | **CAN 优化 MSR** | **U-Net 优化 MSR** |
|---|---|---|---|---|---|---|
| 色调映射质量指数 ↑ | **0.951** | 0.939 | 0.917 | 0.923 | 0.926 | 0.933 |
| 曝光不足↓ | 0.535 | 0.0112 | **0.0088** | 1.803 | 0.372 | 0.334 |
| 过曝↓ | 0.899 | 0.0375 | **0.0025** | 1.176 | 0.749 | 0.494 |
| 全局对比度损失↓ | -0.162 | -0.100 | -0.0948 | **-0.202** | -0.168 | -0.176 |
| 局部对比度损失↓ | -0.0417 | -0.0425 | **-0.0482** | -0.0447 | -0.0444 | -0.0474 |
| 噪声可见性↓ | - | **27.66** | 27.68 | 30.67 | 30.37 | 30.61 |
| 全局时间不一致性↓ | $1.7e^{-4}$ | $1e^{-3}$ | $1。9e^{-3}$ | $\mathbf{2e^{-5}}$ | $1。2e^{-3}$ | $2.5e^{-4}$ |
| 局部时间不一致性↓ | 0.0048 | 0.0014 | **0.0012** | 0.025 | 0.0191 | 0.0186 |

## 4.1 自监督去噪

考量当前结果后，可以发现进一步优化的空间在于减少视网膜皮层理论算法所放大的图像噪声。主要目标是找到一种去噪方法，尽可能减少色调映射图像中的噪声，同时保留图像细节。该方法还应与前文介绍的基于深度学习的色调映射方法兼容，并且易于集成

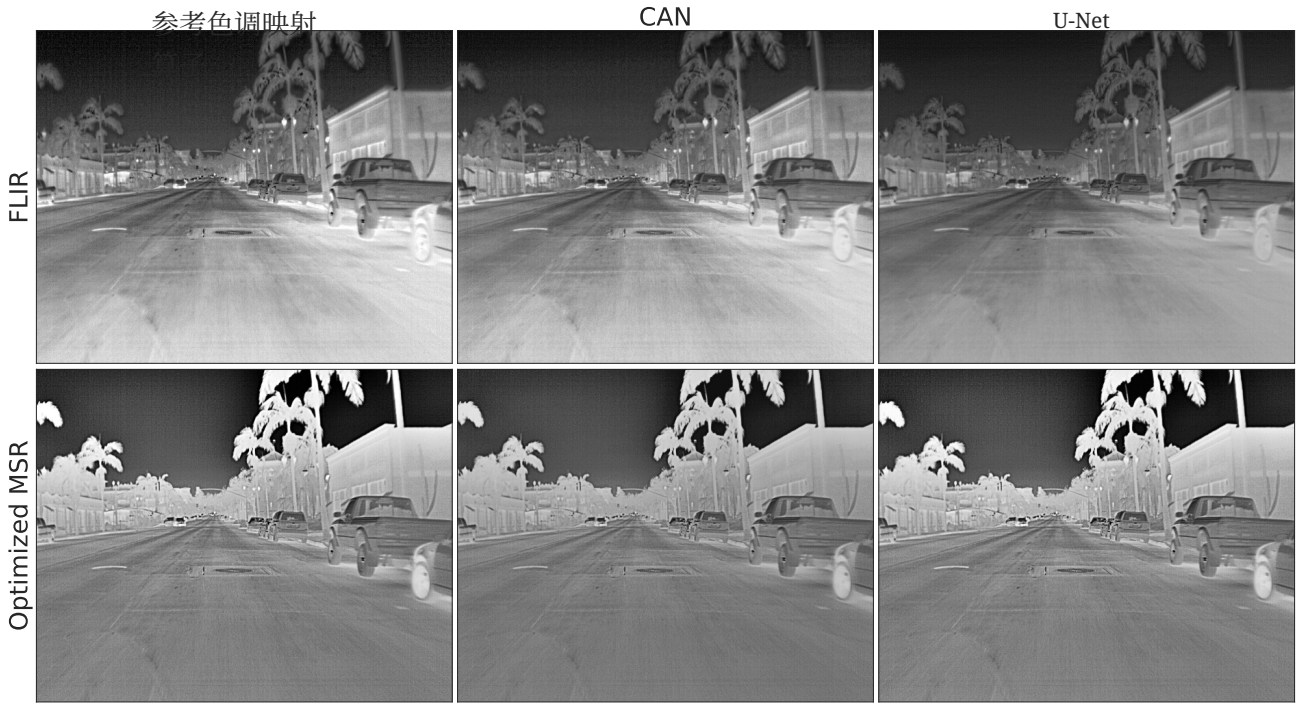


图 6. 基于深度学习的近似结果对比。其思路是利用深度神经网络，模仿基于传统计算机视觉的参考色调映射算子的性能。示例图像表明，测试的两种神经网络架构 CAN 和 U-Net 均能成功模仿参考色调映射算子。有趣的是，模仿 FLIR 色调映射算子时 CAN 的表现更好，而模仿优化的 MSR 时 U-Net 的表现更好。

到前一节中。为此，本文研究了三种自监督去噪方法，分别是 Ehret 等人提出的模型盲帧间去噪方法[34]，、Moran 等人提出的 Noisier2Noise 方法[35],以及 Xu 等人提出的 Noisy-As-Clean 方法[9]。

Ehret 等人提出的帧间视频去噪方法[34] 利用稠密光流将前一帧变形至当前帧。这样便生成了一对属于同一场景、但噪声随机分布且彼此独立的图像。随后，可以通过原始的 Noise2Noise 方法对去噪进行训练[33]。Ehret 等人提供的原始代码已公开发布。该方法基于一种在线训练技术，即针对每一帧新图像分别执行训练和推理。结果如图 7所示。可以清楚地看到，帧间方法难以处理来自*FLIR* 训练子集的图像（右侧图像），该子集属于 FLIR ADAS 数据集。这可能是由于该子集的帧率较低所致。约 2 fps 的帧率，加上摄像机安装在移动的车辆上这一事实，导致连续帧之间的场景外观变化过大。使用更高的帧率，例如 *FLIR* 测试集（左侧）中的帧率，则能获得更好的结果。尽管该方法在去噪方面表现良好，但可以看出，帧间方法在这些图像上产生了模糊的结果。这很可能是因为稠密光流在处理 TIR 图像时受到纹理不足的影响[56] 。Ehret 等人[34] 采用 TV-L1 方法[57,58] 计算稠密光流。我们测试了另外两种用于计算稠密光流的流行替代方法[59,60] 但结果反而更差。因此，在 FLIR ADAS 数据集上，稠密光流的计算精度不足，无法为帧间训练方法生成精确对齐的图像对。由此可以得出结论，这种视频去噪方法不适用于我们当前的任务。

Moran 等人提出的 Noisier2Noise 去噪方法[35]，和 Xu 等人提出的 Noisy-As-Clean 去噪方法[9] 均结合输入图像的数据增强与深度学习，对输入图像进行去噪。其核心思想是向已有噪声的真实训练图像中添加人工生成的合成噪声，从而学习隐式去噪。不过，Noisier2Noise 方法在训练和推理时都会向输入图像添加合成噪声，而 Noisy-As-Clean 方法则将添加的

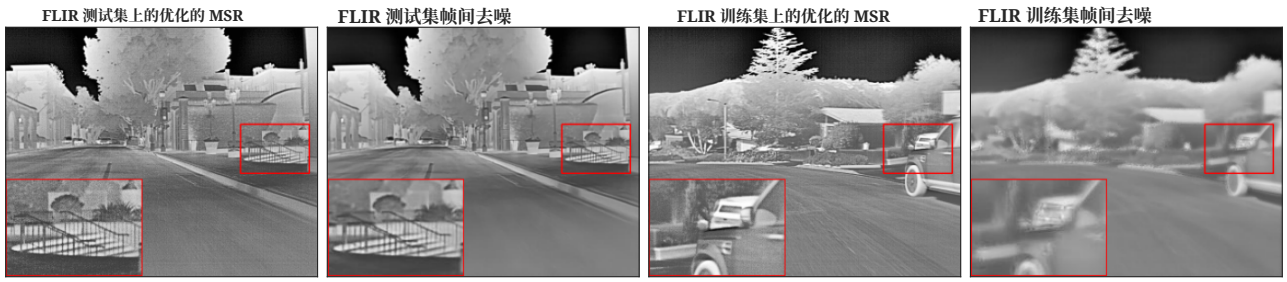


图 7. 帧间去噪方法在 FLIR 测试集和 FLIR 训练集上的结果。从左到右：未经去噪的 FLIR 测试集参考图像、经帧间去噪后的同一图像、未经去噪的 FLIR 训练集参考图像、经帧间去噪后的同一图像。

合成噪声仅用于训练阶段。两种方法都采用单图像方式，与逐帧处理方法不同。因此，它们非常适用于低帧率的 FLIR ADAS 数据集。两种方法均采用 U-Net 架构[7] 作为网络架构。为增强输入训练图像，采用 λ 值为 100 的加性泊松噪声，因为预计这种噪声最接近 FLIR 数据集中存在的真实噪声[61,62]。

结果如图 8 所示。帧间去噪方法显著降低了噪声，定量分析也证实了这一点：噪声可见性从 30.67 dB 降至 26.93 dB。然而，正如前文所述，该方法容易模糊图像细节。Noisier2Noise 方法在图像不同区域产生的效果不一致。某些区域（例如右侧中部的栏杆）噪声较少，而其他区域（例如天空）仍然存在噪声。此外，该方法会损失部分图像细节。图像左上方建筑物中的树木细节完全被抹去，便是一个例子。该方法还会降低整体亮度和对比度。与 Noisier2Noise 相比，Noisy-As-Clean 方法的去噪效果稳定提升，接近帧间去噪方法的水平，同时还能保留图像锐度。图像整体细节和对比度也得以保留。总的来说，Noisy-As-Clean 方法被认为是本研究中最有前景的方法，因此将对其进行进一步研究。定量结果见表 4。

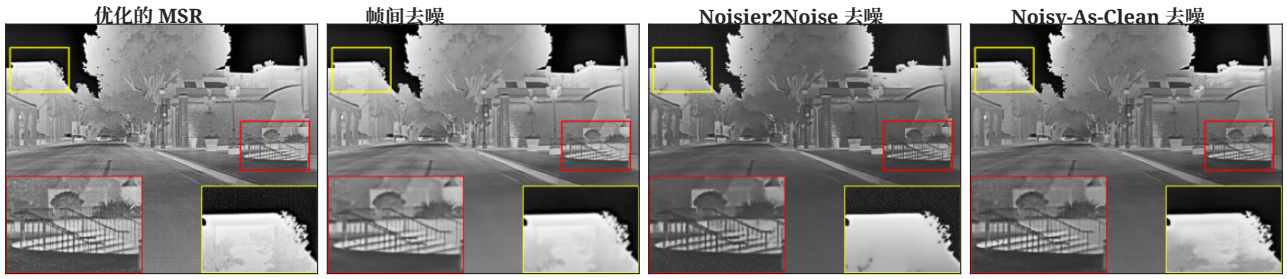


图 8. 不同自监督去噪方法的比较。帧间去噪[34] 会导致图像模糊，而 Noisier2Noise[35] 会丢失图像细节；Noisy-As-Clean[9] 方法则能提供最佳的去噪效果。

### 4.2 联合色调映射与去噪

由于我们选择的自监督去噪方法[9] 在训练期间仅对输入图像进行数据增强，因此可以轻松地融入学习近似色调映射算子的训练过程。为此，我们采用 U-Net 架构，其配置和使用方式均与色调映射的监督训练相同。唯一的区别在于，训练期间输入的高动态范围图像会受到 λ 值为 100 的加性泊松噪声污染。这种色调映射与自监督去噪联合监督学习的结果如图9所示。我们还添加了一个图像示例，其中相同的去噪

方法是在学习近似色调映射算子后，作为后处理步骤单独训练的。可以看出，联合学习甚至提升了去噪性能，表4中的图像度量也证实了这一点，同时整体外观和对比度更接近原始图像。这表明，将该方法与色调映射任务一同训练能够提升去噪性能，同时不会对色调映射产生负面影响。在单一网络中同时训练并执行这两项任务，也能带来性能优势并降低复杂度。

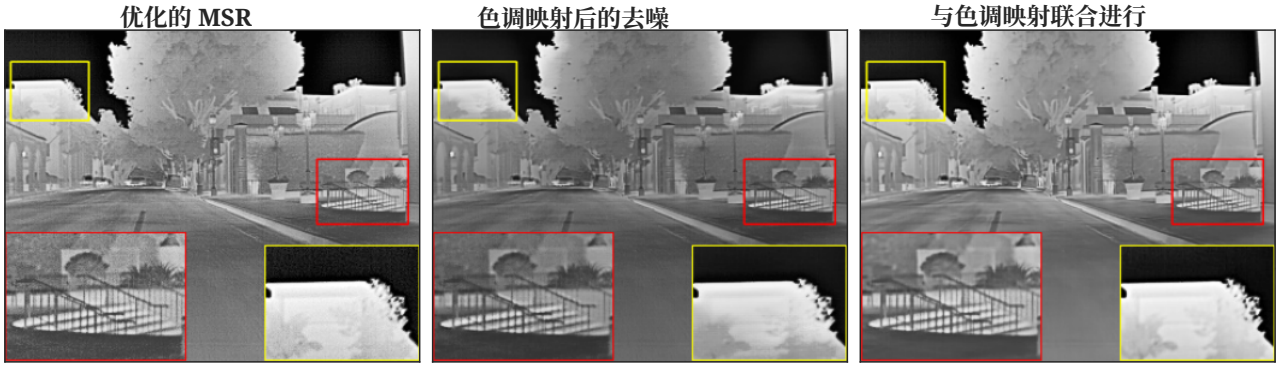


图 9. 对比自监督 Noisy-As-Clean 去噪方法[9] 用作后处理，或与色调映射联合学习的效果。

### 4.3 去闪烁

基于深度学习的优化 MSR 色调映射算子近似方法再次引入了闪烁，与去噪结合时，闪烁现象会进一步加剧。这并不意外，因为该学习方法在训练期间仅考虑单幅图像。在深度学习中，强制连续视频帧保持时间一致性的常用方法是使用稠密光流[37,38]。由于帧间去噪方法的实验已表明，稠密光流计算在 FLIR ADAS 数据集上会产生不精确的结果，因此本文不再进一步研究基于光流计算的方法。我们转而采用一种非常简单的方法：引入时间正则化损失项，以惩罚连续视频帧之间图像平均亮度的变化。这样便可确保时间一致性。时间正则化损失项 $\mathcal{L}_{reg}$ 的计算方式如下：

$$\mathcal{L}_{reg} = \frac{1}{2}(\bar{x}_t - \frac{1}{n}\sum_{k=1}^{n}\bar{x}_{t-k})^2 \tag{1}$$

将当前输出图像 $\bar{x}_t$ 的图像平均亮度值与最近 $n$ 个输出图像 $\bar{x}_{t-k}$ 的平均值均匀加权后的总和进行比较。当前平均亮度值与过去的数值之间的偏差越大，损失就越高，因此会在训练过程中受到惩罚。在我们的实验中，设置为 $n = 10$。随后，总损失 $\mathcal{L}_{total}$ 计算为标准均方误差损失项 $\mathcal{L}_{MSE}$ 与时间正则化损失项的加权和：

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{MSE} + (1-\alpha)\mathcal{L}_{reg} \tag{2}$$

在我们的实验中，我们将 $\alpha$ 设为 0.9。该去闪烁方法对全局时间不一致性指标的影响如表 4所示。全局时间不一致性的显著降低意味着明显的去闪烁效果，这一点也通过对视频流的主观评估得到了证实。同样值得注意的是，欠曝像素和过曝像素的数量有所增加，更接近优化的 MSR。这可以通过使用时间正则化损失项减少离群值来解释。离群值会导致对 8位数值范围的利用相当不足，从而减少过曝像素和欠曝像素。

表4. 自监督去噪方法以及结合去闪烁的去噪的定量评估。首先，在色调映射近似之后，将所考虑的去噪方法作为后处理进行评估。随后，通过隐式去噪和去闪烁进行联合训练。在联合训练中，我们使用 Noisy-As-Clean 作为去噪方法。根据各项指标，联合色调映射算子+去噪+去闪烁的表现最佳。

| | 优化 MSR | Noisy-As-Clean 去噪作为后处理 | 联合色调映射算子 +去噪 | 联合色调映射算子 +去噪 +去闪烁 |
|---|---|---|---|---|
| 色调映射质量指数↑ | 0.923 | **0.937** | **0.937** | 0.934 |
| 曝光不足↓ | 1.803 | **0.430** | 0.460 | 1.330 |
| 过曝↓ | 1.176 | 0.439 | **0.188** | 1.110 |
| 全局对比度损失↓ | **-0.202** | -0.145 | -0.133 | -0.152 |
| 局部对比度损失↓ | **-0.044** | -0.034 | -0.033 | -0.031 |
| 噪声可见性↓ | 30.67 | 26.52 | **25.97** | 26.49 |
| 全局时间不一致性↓ | $2e^{-5}$ | $2.4e^{-4}$ | $8.6e^{-4}$ | **$5。7e^{-6}$** |
| 局部时间不一致性↓ | 0.025 | 0.0229 | **0.0207** | 0.0258 |

作为本节的总结，我们可以指出，引入简单的时间正则化损失项 $\mathcal{L}_{reg}$ 能够充分解决闪烁问题。该损失项可以轻松添加到总损失函数中，从而成为去噪与色调映射联合学习过程的一部分。Eilertsen 等人提出了一种颇具前景的方法，通过数据增强，在训练过程中无需使用连续图像即可隐式学习去闪烁。[36] 以及 Zheng 等人[63] 采用了数据增强。该方法不使用之前的视频帧，而是通过人为改变训练图像本身的全局亮度来进行增强，从而引入合成闪烁。这种方法很有前景，有望进一步简化训练过程，但本文所 प्रस्तुत 的工作并未对其进行测试。

## 5. 与最先进方法的比较

本文对主要分析方法的定量评估见表 5。相关的定性比较见图 10。除了最先进方法 FLIR、实时色调映射算子、CGF 和基线 MSR 之外，我们还纳入了第 3 节所述的优化的 MSR，以及第 4节所述、通过联合训练色调映射、去噪和去闪烁实现的最终深度学习方法。对 MSR 的优化显著提升了 TMQI、图像对比度和视频流的时间一致性，但也增加了噪声可见性。图像和视频流的定性比较也证实了这些结果。表 学习型 *MSR* 5中将通过深度卷积神经网络对优化的 MSR 进行近似的方法称为学习型 MSR。从 TMQI 可以看出，它能够模仿优化的 MSR 所呈现的图像外观。在训练过程中加入去噪和去闪烁，以联合学习色调映射和图像增强后，噪声可见性和全局时间不一致性指标均明确表明图像和视频质量有所提升。因此，我们可以得出结论，两种优化能够协同发挥作用；与优化的 MSR 相比，它们在 TMQI、噪声可见性和时间一致性方面均带来了显著提升。联合深度卷积神经网络方法的对比度指标有所下降，但定性比较并未证实这一结果，这或许是由噪声减少所致。

## 6. 结论

本文探讨了热红外图像和视频的色调映射与图像质量增强。受一种前景良好的、基于视网膜皮层理论的热红外图像色调映射方法启发，我们通过应用并优化 MSR 算法建立了一个基线。优化主要侧重于对比度增强和去闪烁，但同时也放大了图像噪声。随后，我们使用广受欢迎的深度神经网络架构 U-Net 对这一优化的 MSR 算法进行近似。因此，优化的 MSR 算法生成的图像被用作全监督学习方案中的训练图像。我们的目标是在训练过程中高效地引入相关方法，从而联合执行去噪和去闪烁。为此，我们采用了一种基于数据增强的自监督图像去噪方法

表 5. 最先进方法、我们优化的 MSR，以及我们称为 *learned MSR*的基于深度学习并结合联合去噪与去闪烁的方法之间的比较。与基线 MSR[6] 相比，该方法在多项指标上的性能提升十分显著，其中 TMQI 和噪声可见性是最重要的指标。另一种基于视网膜皮层理论的方法 CGF 的表现也不如该方法。

| | **FLIR**[12] | **实时**[16] **TMO** | **CGF**[5] | **MSR**[6] | **优化 MSR** | **学习得到的 MSR** |
|---|---|---|---|---|---|---|
| 色调映射质量指数 ↑ | **0.951** | 0.809 | 0.850 | 0.866 | 0.923 | 0.934 |
| 曝光不足↓ | **0.535** | 0.755 | 2.842 | 1.191 | 1.803 | 1.330 |
| 过曝↓ | 0.899 | 1.628 | **0.030** | 1.578 | 1.176 | 1.110 |
| 全局对比度损失↓ | -0.162 | -0.123 | -0.084 | -0.148 | **-0.202** | -0.152 |
| 局部对比度损失↓ | -0.041 | -0.024 | -0.022 | -0.027 | **-0.044** | -0.031 |
| 噪声可见性↓ | - | 30.77 | 30.04 | 28.79 | 30.67 | **26.49** |
| 全局时间不一致性↓ | 1。$7e^{-4}$ | 3。$7e^{-4}$ | $2.8e^{-5}$ | $2.7e^{-4}$ | $2e^{-5}$ | $\mathbf{5.7e^{-6}}$ |
| 局部时间不一致性↓ | **0.0048** | 0.0157 | 0.0088 | 0.0394 | 0.025 | 0.0258 |

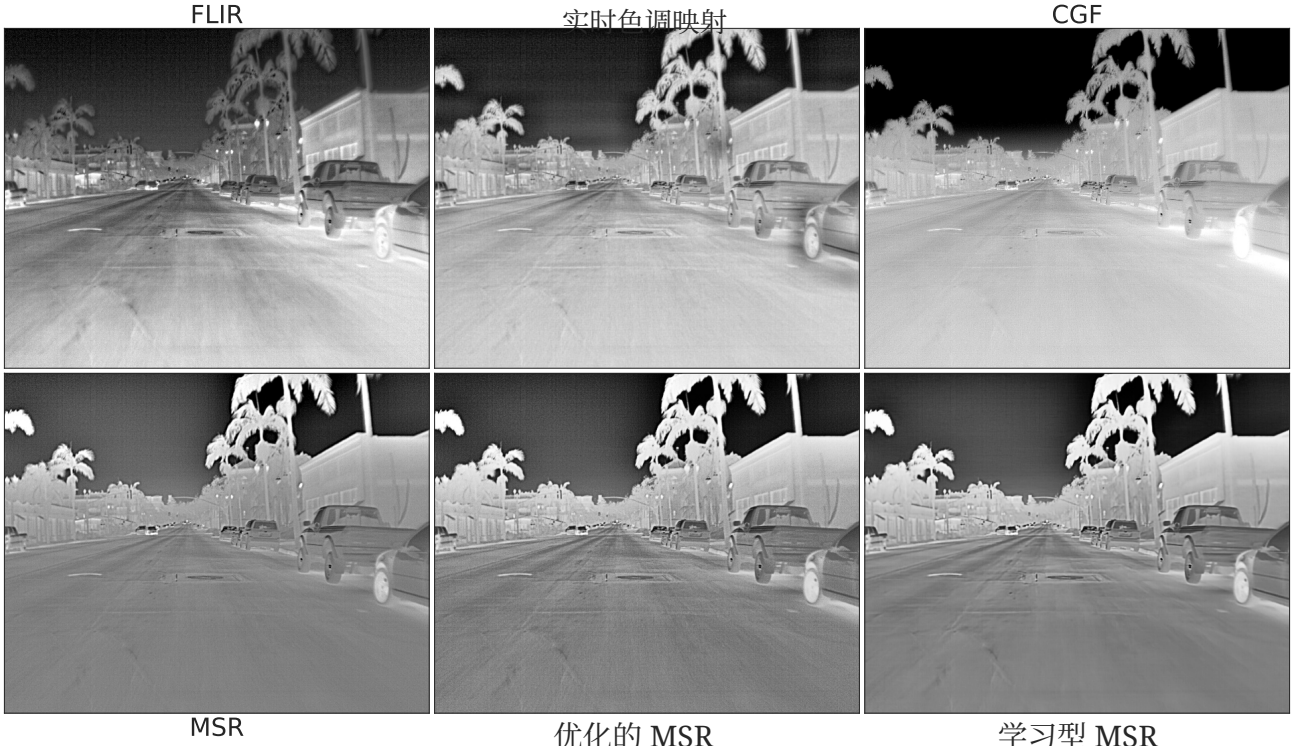


图10. 最先进方法、我们的优化的 MSR 与最终基于深度学习的方法之间的定性比较。示例图像印证了表 5 中的指标：FLIR 算法（左上）具有最高的色调映射质量指数，其次是我们学习型 MSR。优化的 MSR 与学习型 MSR 之间的降噪效果（底部中间和右侧）在放大视图中清晰可见。

并通过引入时间正则化损失项，采用了一种简单而有效的去闪烁方法。我们针对上述每个步骤都进行了大量实验。我们使用了公开的 FLIR ADAS 数据集，以及一个用于热红外光谱色调映射的评估框架所公开的代码。通过定量和定性评估，我们以实证方式证明了所提方法的有效性，并表明其性能接近最先进方法；与文献中的四种色调映射算法相比，结果亦是如此。在这项评估中，两种基于视网膜皮层理论的色调映射算法明显逊于我们的方法。

## 参考文献

[1] Teutsch, M., Sedelmaier, S., Moosbauer, S., Eilertsen, G., and Walter, T., “热红外视频色调映射的客观图像质量评估,” 载于 [*IEEE/CVF*会议论文集

计算机视觉与模式识别会议（CVPR）研讨会], (2020年6月）。[2] Zuiderveld, K., “对比度受限自适应直方图均衡化,” 载于 [图形学宝典], Heckbert, P. S., 编，474–485，学术出版社（1994年）。[3] Eilertsen, G., Wanat, R., Mantiuk, R. K., and Unger, J., “高动态范围视频色调映射算子的评估,” 计算机图形学论坛 32, 275–284（2013年10月）。[4] Eilertsen, G., Mantiuk, R. K., and Unger, J., “高动态范围视频色调映射算法比较综述,” 计算机图形学论坛 36, 565–592（2017年5月）。[5] Luo, H., Xu, L., Hui, B., and Chang, Z., “基于条件滤波多尺度视网膜理论的红外图像色调映射,” 载于 [AOPC 2015：图像处理与分析], Shen, C., Yang, W., and Liu, H., 编，9675，577–582，国际光学与光子学学会，SPIE（2015年）。[6] Petro, A. B., Sbert, C., and Morel, J.-M., “多尺度视网膜理论,” 在线图像处理 4, 71–88（2014年4月）。[7] Ronneberger, O., P.Fischer, and Brox, T., “U-Net：用于生物医学图像分割的卷积网络,” 载于 [医学图像计算与计算机辅助干预（MICCAI）], LNCS 9351，234–241，施普林格（2015年）。（可在 arXiv:1505.04597 [cs.CV]获取）。[8] Chen, Q., Xu, J., and Koltun, V., “利用全卷积网络进行快速图像处理,” 载于 [IEEE 国际计算机视觉会议（ICCV）论文集], (2017年10月）。[9] Xu, J., Huang, Y., Cheng, M.-M., Liu, L., Zhu, F., Xu, Z., and Shao, L., “Noisy-as-clean：从受损图像中学习自监督去噪,” IEEE图像处理汇刊 29, 9316–9329（2020年）。[10]Hu, L., Chen, H., and Allebach, J., “用于高动态范围图像增强的联合多尺度色调映射与去噪,” 载于 [IEEE/CVF冬季计算机视觉应用会议研讨会（WACVW）], (2022年）。[11] Granados, M., Aydin, T., Tena, J., Lalonde, J.-F., and Theobalt, C., “用于色调映射图像去噪的高动态范围图像噪声估计,” 1–8（2015年11月）。[12] FLIR Systems, “FLIR ADAS红外数据集,”（2021年）。 [在线；访问日期：2021-05-30].[13] Devlin, K., Chalmers, A., Wilkie, A., and Purgathofer, W., “Star：色调再现与基于物理的光谱渲染,” 载于 [最新技术报告，Eurographics *2002*], Fellner, D. and Scopignio, R., 编，欧洲计算机图形学协会（2002年）。[14] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K., [高动态范围成像：采集、显示与基于图像的照明],Morgan Kaufmann（2010年）。[15] Banterle, F., Artusi, A., Debattista, K., and Chalmers, A., [高级高动态范围成像（第2版）],AK Peters（CRC出版社），美国马萨诸塞州纳蒂克（2017年7月）。[16] Eilertsen, G., Mantiuk, R. K., and Unger, J., “实时噪声感知色调映射,” ACM 图形学汇刊 34（2015年）。[17] Land, E. and McCann, J., “明度与 Retinex 理论,” 美国光学学会杂志 61, 1–11（1971年2月）。[18] Ahn, H., Keum, B., Kim, D., and Lee, H. S., “基于视网膜理论的高动态范围图像自适应局部色调映射,” 载于 [*2013* IEEE国际消费电子会议（ICCE）], 153–156（2013年）。[19] Liu, C., Sui, X., Kuang, X., Liu, Y., Gu, G., and Chen, Q., “基于邻域条件直方图的红外图像自适应对比度增强,” 遥感 11(11), 1381（2019年）。[20] Liu, C., Sui, X., Kuang, X., Liu, Y., Gu, G., and Chen, Q., “基于全局与局部直方图规定化的红外图像优化对比度增强,” 遥感 11(7), 849（2019年）。[21] Wan, M., Gu, G., Qian, W., Ren, K., Chen, Q., and Maldague, X., “采用自适应直方图划分和亮度校正的红外图像增强,” 遥感 10(5), 682（2018年）。[22] Shan, Q., Jia, J., and Brown, M. S., “全局优化线性窗口色调映射,” IEEE可视化与计算机图形学汇刊 16(4), 663–675（2010年）。[23] Liang, Z., Xu, J., Zhang, D., Cao, Z., and Zhang, L., “用于色调映射的混合 l1-l0 层分解模型,” 载于 [IEEE计算机视觉与模式识别会议（CVPR）论文集], (2018年6月）。[24] Gu, B., Li, W., Zhu, M., and Wang, M., “用于高动态范围图像色调映射的局部保边多尺度分解,” IEEE图像处理汇刊 22(1), 70–79（2013年）。[25] Montulet, R. and Briassouli, A., “用于鲁棒端到端色调映射的深度学习,” 载于 [BMVC], (2019年）。

[26] Rana, A., Singh, P., Valenzise, G., Dufaux, F., Komodakis, N., and Smolic, A., “面向高动态范围图像的深度色调映射算子”，IEEE图像处理汇刊 29, 1285–1298 (2020).[27] Patel, V. A., Shah, P., and Raman, S., “用于 HDR 图像色调映射的生成对抗网络”，载于 [计算机视觉、模式识别、图像处理与图形学], Rameshan, R., Arora, C., and Dutta Roy, S., 编，220–231，Springer Singapore，新加坡 (2018).[28] Su, C.-C., Wang, R., Lin, H.-J., Liu, Y.-L., Chen, C.-P., Chang, Y.-L., and Pei, S.-C., “可探索式色调映射算子”，载于 [ICPR], 10320–10326 (2020).[29] Gharbi, M., Chen, J., Barron, J. T., Hasinoff, S. W., and Durand, F., “用于实时图像增强的深度双边学习”，ACM 图形学汇刊 (TOG) 36(4), 118 (2017).[30] Vemulapalli, R., Tuzel, O., and Liu, M.-Y., “深度高斯条件随机场网络：一种用于判别式去噪的模型化深度网络”，载于 [IEEE计算机视觉与模式识别会议论文集 (CVPR)],(2016年6月).[31] Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L., “超越高斯去噪器：用于图像去噪的深度 CNN 残差学习”，IEEE图像处理汇刊 26(7), 3142–3155 (2017).[32] Zhang, K., Zuo, W., and Zhang, L., “FFDNet：迈向快速灵活的基于 CNN 的图像去噪方案”，IEEE图像处理汇刊 27(9), 4608–4622 (2018).[33] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T., “Noise2Noise：无需干净数据的图像复原学习”，载于 [第35届国际机器学习会议论文集], Dy, J. and Krause, A., 编，机器学习研究论文集 80, 2965–2974，PMLR (2018年7月10日至15日).[34] Ehret, T., Davy, A., Morel, J.-M., Facciolo, G., and Arias, P., “通过逐帧训练实现模型无关的视频去噪”，载于 [IEEE/CVF计算机视觉与模式识别会议论文集 (CVPR)],(2019年6月).[35] Moran, N., Schmidt, D., Zhong, Y., and Coady, P., “Noisier2Noise：从未配对噪声数据中学习去噪”，载于 [IEEE/CVF计算机视觉与模式识别会议论文集 (CVPR)], (2020年6月).[36] Eilertsen, G., Mantiuk, R., and Unger, J., “面向时间稳定卷积神经网络的单帧正则化”，载于[IEEE/CVF计算机视觉与模式识别国际会议 (CVPR)], (2019).[37] Zhang, F., Li, Y., You, S., and Fu, Y., “利用单幅图像学习低光照视频增强的时间一致性”，载于 [IEEE/CVF计算机视觉与模式识别会议论文集 (CVPR)], 4967–4976 (2021年6月).[38] Lai, W.-S., Huang, J.-B., Wang, O., Shechtman, E., Yumer, E., and Yang, M.-H., “学习盲视频时间一致性”，载于 [欧洲计算机视觉会议 (ECCV)], (2018年9月).[39]Zamir, A., Sax, A., , Shen, W., Guibas, L., Malik, J., and Savarese, S., “Taskonomy：解构任务迁移学习”，载于 [IEEE计算机视觉与模式识别国际会议 (CVPR)], (2018).[40] He, K., Gkioxari, G., Dollar, P., and Girshick, R., “Mask R-CNN”，载于 [IEEE 国际计算机视觉会议 (ICCV)],(2017).[41] Ma, L., Ma, T., Liu, R., Fan, X., and Luo, Z., “迈向快速、灵活且鲁棒的低光照图像增强”，载于 [IEEE/CVF计算机视觉与模式识别国际会议 (CVPR)],(2022).[42] Lyu, K., Pan, S., Li, Y., and Zhang, Z., “JSENet：用于图像联合超分辨率与增强的深度卷积神经网络”，Neurocomputing 489, 570–583 (2022).[43]Xing, W. and Egiazarian, K., “图像去马赛克、去噪与超分辨率的端到端学习”，载于 [IEEE/CVF计算机视觉与模式识别国际会议 (CVPR)], (2021).[44] Kuang, X., Sui, X., Liu, Y., Chen, Q., and Gu, G., “使用深度卷积神经网络进行单幅红外图像增强”，Neurocomputing 332, 119–128 (2019).[45] Limare, N., Lisani, J.-L., Morel, J.-M., Petro, A. B., and Sbert, C., “最简色彩平衡”，在线图像处理 1, 297–315 (2011).

[46] Gonzalez, R. and Woods, R., [数字图像处理], Prentice Hall (2008).[47] Guthier, B., Kopf, S., Eble, M., and Effelsberg, W., “色调映射高动态范围视频中的闪烁抑制”，载于 [第十六届彩色成像：显示、处理、硬拷贝与应用], Eschbach, R., Marcu, G. G., and Rizzi, A., 编，7866, 78660C，国际光学与光子学学会，SPIE (2011).[48] Delon, J., “电影与视频尺度-时间均衡在闪烁抑制中的应用”，IEEE图像处理汇刊 15(1), 241–248 (2006).[49] Piti´e, F., Dahyot, R., Kelly, F., and Kokaram, A., “一种稳定图像序列亮度波动的新型鲁棒技术”，载于 [欧洲计算机视觉会议 (ECCV)], (2004).[50] Kanj, A., 高速视频的闪烁去除与颜色校正，论文，Universit´e Paris-Est Cr´eteil Val-de-Marne (2017).[51] Grafarend, E.，[线性与非线性模型：固定效应、随机效应与混合模型], 柏林；纽约；Walter de Gruyter (2006).[52] Montulet, R. and Briassouli, A., “用于鲁棒端到端色调映射的深度学习”，载于 [英国机器视觉会议论文集],(2019).[53] Yu, F. and Koltun, V., “通过空洞卷积进行多尺度上下文聚合”，载于 [4第届国际学习表征会议，ICLR 2016，波多黎各圣胡安，2016年5月2日至4日，会议论文集],Bengio, Y. and LeCun, Y., 编 (2016).[54] Komatsu, R. and Gonsalves, T., “比较用于彩色图像去噪的基于 U-Net 的模型”，AI 1(4), 465–486 (2020).[55] Wu, Y. and He, K., “组归一化”，载于 [欧洲计算机视觉会议 (ECCV)],(2018年9月).[56] Teutsch, M., Sappa, A., and Hammoud, R.，[红外光谱中的计算机视觉：挑战与方法],计算机视觉综合讲义，Morgan & Claypool (2021).[57] Zach, C., Pock, T., and Bischof, H., “实时 TV-L1 光流的对偶性方法”，载于 [模式识别], Hamprecht, F. A., Schn¨orr, C., and J¨ahne, B., 编，214–223, Springer Berlin Heidelberg，柏林，海德堡 (2007).[58] S´anchez P´erez, J., Meinhardt-Llopis, E., and Facciolo, G., “TV-L1光流估计”，在线图像处理 3, 137–150 (2013).[59] Farneb¨ack, G., “基于多项式展开的双帧运动估计”，载于 [第13届斯堪的纳维亚图像分析会议], LNCS 2749, 363–370 (2003年6月至7月).[60] Yang, G. and Ramanan, D., “用于光流的体对应网络”，载于 [神经信息处理系统进展], 793–803 (2019).[61] Kennedy, H., “热成像系统中的噪声建模”，载于 [SPIE 第 卷 *1969*], (1993).[62] Isoz, W., Svensson, T., and Renhorn, I., “红外焦平面阵列的非均匀性校正”，载于[SPIE 第 卷 *5783*],(2005).[63] Zheng, S., Song, Y., Leung, T., and Goodfellow, I., “通过稳定性训练提升深度神经网络的鲁棒性”，载于 [IEEE计算机视觉与模式识别国际会议 (CVPR)], (2016).