

无人机载昼夜单通道自然感彩色热成像技术

盛典 金伟其 李力 王霞 裘溯

UAVs-mounted day and night single-channel natural sensing color thermal imaging technology

SHENG Dian, JIN Weiqi, LI Li, WANG Xia, QIU Su

引用本文:

盛典, 金伟其, 李力, 等. 无人机载昼夜单通道自然感彩色热成像技术[J]. 应用光学, 2025, 46(3): 481–495. DOI: 10.5768/JAO202546.0311002

SHENG Dian, JIN Weiqi, LI Li, et al. UAVs-mounted day and night single-channel natural sensing color thermal imaging technology[J]. Journal of Applied Optics, 2025, 46(3): 481–495. DOI: 10.5768/JAO202546.0311002

在线阅读 View online: <https://doi.org/10.5768/JAO202546.0311002>

您可能感兴趣的其他文章

Articles you may be interested in

无人机光电对抗技术研究

Research on UAV photoelectric countermeasure technology

应用光学. 2021, 42(3): 377–382 <https://doi.org/10.5768/JAO202142.0301001>

基于三色LCTF的自然感彩色微光EBAPS成像系统

Natural color low-level-light EBAPS imaging system based on three-color LCTF

应用光学. 2022, 43(6): 1044–1053 <https://doi.org/10.5768/JAO202243.0604002>

基于IDOU-YOLO的红外图像无人机目标检测算法

Infrared image UAV target detection algorithm based on IDOU-YOLO

应用光学. 2024, 45(4): 723–731 <https://doi.org/10.5768/JAO202445.0402001>

无人机对地目标自动检测与跟踪技术

Automatic target detecting and tracking technology based on UAV ground target images

应用光学. 2020, 41(6): 1153–1160 <https://doi.org/10.5768/JAO202041.0601003>

基于改进YOLOv4的低慢小无人机实时探测算法

Improved YOLOv4 for real-time detection algorithm of low-slow-small unmanned aerial vehicles

应用光学. 2024, 45(1): 79–88 <https://doi.org/10.5768/JAO202445.0102002>

多尺度融合超分辨率算法在无人机探测中的应用

Application of multi-scale fusion super-resolution algorithm in UAV detection

应用光学. 2021, 42(3): 462–473 <https://doi.org/10.5768/JAO202142.0302003>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2025) 03-0481-15

引用格式: 盛典, 金伟其, 李力, 等. 无人机载昼夜单通道自然感彩色热成像技术 [J]. 应用光学, 2025, 46(3): 481-495.

SHENG Dian, JIN Weiqi, LI Li, et al. UAVs-mounted day and night single-channel natural sensing color thermal imaging technology[J]. Journal of Applied Optics, 2025, 46(3): 481-495.



在线阅读

无人机载昼夜单通道自然感彩色热成像技术

盛 典, 金伟其, 李 力, 王 霞, 裘 溯

(北京理工大学 光电学院 光电成像技术与系统教育部重点实验室, 北京 100081)

摘 要: 为提高无人机在复杂环境下的环境感知和目标识别能力, 探索无人机载昼夜单通道自然感彩色热成像技术, 基于深度学习技术, 构建了两种图像彩色化技术: 基于多判别器生成对抗网络的红外图像彩色化网络 (thermal image visualization net, TIVNet)、基于区域自分割的语义引导扩散模型的红外图像彩色化网络 (region-aware semantic-guided diffusion models, RSDM)。TIVNet 通过多判别器的生成对抗网络结构能够将红外热成像直接转化为类似彩色可见光图像, 但在部分场景细节部分色彩存在错误。语义引导扩散模型网络通过更复杂的模型生成更加真实的彩色图像, 增强了图像的视觉效果和信息传递效率, 但目前处理速度还有待提高。实验结果表明, 所提出的两种方法在图像转换的逼真度和实时速率方面都各自具有显著优势, TIVNet 在无人机平台的实验测试中, 达到了不低于 40 Hz 的实时处理速度, 同时保持高视觉逼真度, 证明了技术的可行性, 并具备针对应用需求开展装备转换应用的能力, 为无人机在军事和民用领域的应用提供了新的技术手段。

关键词: 无人机; 昼夜; 自然感彩色; 红外彩色化技术; 生成对抗网络; 语义引导扩散模型

中图分类号: TN21; O432

文献标志码: A

DOI: 10.5768/JAO202546.0311002

UAVs-mounted day and night single-channel natural sensing color thermal imaging technology

SHENG Dian, JIN Weiqi, LI Li, WANG Xia, QIU Su

(Key Laboratory of Photoelectronic Imaging Technology and System (Ministry of Education), School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: To enhance the environmental perception and target recognition capabilities of unmanned aerial vehicles (UAVs) in complex environments, the exploration of UAVs-mounted day-and-night single-channel natural-sensing color thermal imaging technology was undertaken. Based on deep learning technology, two image colorization techniques were constructed: the TIVNet, an infrared image colorization network based on a multi-discriminator generative adversarial network, and the infrared image colorization network based on the semantic-guided diffusion model with regional self-segmentation RSDM. A generative adversarial network architecture with multi-discriminator was employed by TIVNet to directly convert infrared thermal images into colorized visible-light-like images. However, there were instances where color inaccuracies were observed in the detailed aspects of certain scenes. The more real color images through a more sophisticated model were generated by RSDM, enhancing the visual effect and efficiency of information transmission in images, but the

收稿日期: 2024-11-19; 修回日期: 2025-03-04

基金项目: 快速扶持项目 (JZX7Y20200254100801)

作者简介: 盛典 (1997—), 男, 博士研究生, 主要从事红外图像 AI 处理方法研究。E-mail: 3120205294@bit.edu.cn

通信作者: 金伟其 (1961—), 男, 博士, 教授, 主要从事夜视与红外技术、光电图像处理、光电检测与仪器研究。

E-mail: jinwq@bit.edu.cn

processing speed of the current model needed to be improved. Experimental results indicate that the two proposed methods each possesses distinct advantages in terms of the fidelity of image conversion and real-time processing rate. TIVNet has achieved real-time processing at a rate of not less than 40 Hz on platforms such as UAVs, demonstrating the technical feasibility. Moreover, it possesses the capability to facilitate equipment transformation applications based on the specific application requirements, thereby providing a novel technological means for the application of UAVs in both military and civilian domains.

Key words: unmanned aerial vehicles; day and night; natural sensing color; infrared colorization technology; generative adversarial network; semantic-guided diffusion model

引言

无人机在军事和民用领域具有日趋广泛的应用,光电吊舱作为无人机环境感知、目标探测识别、辅助导航等的有效手段,特别是在近期的俄乌战争和以哈战争中^[1],无人机更成为战场昼夜攻击的新型手段,成为世界各国应对未来战争重点发展的装备之一。

可见光相机和非制冷红外相机是小型无人机光电吊舱的主要装备^[2],由于其载荷有限,对体积、功耗、机动性等具有更高的要求,且往往需通过人在回传中实现复杂环境下操作人员快速的决策、操作(如图 1 所示)。彩色可见光相机图像清晰,与人眼视觉一致,便于快速判断目标。然而,其性能受昼夜和气候条件限制,在能见度较低时效果显著下降,且对隐蔽目标的探测和识别能力较差。非制冷红外相机具有昼夜观察能力,受气候条件影响小,对隐蔽在城市建筑、树林甚至地下掩体内的人员、车辆等都有很好的探测能力,但热图像与可见光图像的纹理分布不一致,不适宜人眼长期观察。因此,无人机电吊舱通常根据应用需求,选择可见光电视吊舱、红外热成像吊舱和可见光+热成像双波段吊舱,但双波段吊舱的应用受到无人机载荷能力的限制。



图 1 无人机光电吊舱控制和图像样例

Fig. 1 Photoelectric pod control of UAV and image samples

近年来,人工智能方法技术得到迅速发展,除多波段图像融合的彩色夜视技术外,通过人工智能方法将红外热成像直接转化为自然感彩色可见光图像的技术受到关注^[3-4],其只需红外相机即可实现昼夜、恶劣环境下的类可见光彩色热成像,极大提高昼夜环境感知和目标识别能力。我们研究了红外焦平面相机的类可见光自然感彩色热成像

技术,初步在无人机平台实现了典型场景的类可见光自然感彩色热成像,证明了技术的可行性,具备针对应用需求开展装备转换应用的能力。

人类视觉系统对色彩的感知能力远超过对灰度的感知,图像彩色化技术通过对黑白红外图像赋予色彩,丰富人眼的视觉体验。图像彩色化技术在多个领域具有显著的应用价值,尤其是在能见度较低以及夜间黑暗环境中,如辅助驾驶、夜间安防监控和消防救援等。图像彩色化技术主要分为两大类:传统的图像彩色化^[5]和基于深度学习的图像彩色化方法^[6]。为了实现红外图像彩色化技术,本文分别进行两种基于不同网络特点的深度学习彩色化结构设计,研究基于多判别器生成对抗网络和语义引导扩散模型的红外图像转类彩色可见光方法,实验结果显示两种方法分别在转换逼真度和转换实时速率各有优势。

1 红外图像彩色化技术的发展

传统的图像彩色化方法是一种半自动化的彩色化技术,主要包括基于灰度编码和色彩迁移的方法。这些方法通过分析图像内容和色彩分布,将色彩信息映射到黑白图像上,从而实现图像的彩色化。基于深度学习的彩色化方法是一种自动渲染上色技术,利用深度学习模型,如卷积神经网络(convolutional neural network, CNN),自动地将颜色信息与黑白图像相结合,实现更为自然和逼真的彩色效果。这种方法不仅提高了彩色化的速度和效率,还显著提升了色彩的准确性和视觉效果。

色彩传递方法通过利用可见光与红外图像之间的相关性,将颜色信息从可见光图像传递至红外图像,使得最终的图像更逼真、自然。REINHARD E^[5]等于 2001 年提出了一种全局方法,可以将日落图像的颜色特征(如温暖的橙色和红色调)应用到其他时间段拍摄的图像上,从而生成具有日落视觉效果图像,而无需在傍晚实际拍摄日落。

该技术通过匹配两个图像之间的均值和方差来实现, 每个对应图像中的像素都参与上述两个统计变量的计算, 因此被称为全局算法。在此启发下, WELSH T^[6] 等采用邻域匹配标准实现从参考图像向目标图像传输相似颜色信息。王岭雪^[3] 等于 2007 年采用基于 YUV 空间的双通道黑白视频图像彩色融合与传递方法。赵源萌^[7] 等提出基于区域直方图纹理描述的灰度图像色彩传递处理方法(如图 2 所示)。这些技术存在一个主要限制: 彩色化效果高度依赖于颜色与灰度之间的匹配关系。



图 2 区域直方图纹理描述的灰度图像色彩传递处理方法样例

Fig. 2 Example of color transfer processing method for grayscale images using regional histogram texture description

类似的研究还有: 刘缠牢^[8] 等在 2007 年通过将红外图像中的不同灰度级别映射到伪彩色编码, 基于灰度彩色编码的红外图像彩色化方法能够有效地引入伪彩色, 从而显著增强图像的可视化效果; 同年, ZHANG X^[9] 等通过 HIS(hue-intensity-saturation)空间建立了色调、强度和饱和度与灰度水平的函数关系, 实现红外图像的伪彩色效果(如图 3 所示)。

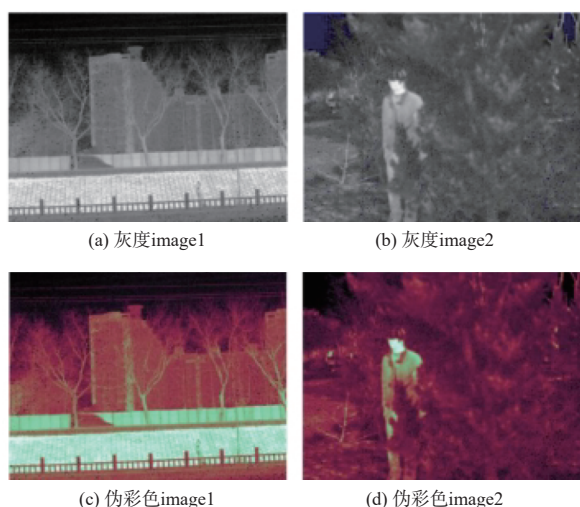


图 3 基于 HIS 空间的红外图像伪彩色编码方法样例
Fig. 3 Example of pseudo-color encoding method for infrared images based on HIS space

尽管传统图像彩色化技术具有一定的效果, 但算法半自动化和计算成本高等问题已不适应大数据时代对速度和效率的需求。随着硬件的升级和深度学习技术的突破, 深度神经网络在特征提取和表达方面展现出卓越能力。基于深度学习的全自动化图像彩色化方法能够自动学习数据映射并不断优化模型, 不仅提升了算法的泛化性和着色效果, 还显著增强了算法的实时性, 因此, 正逐渐受到研究者的青睐。

2016 年, LIMMER M^[10] 等提出了一种基于深度卷积神经网络的红外图像彩色化方法, 可将单色的近红外图像映射成彩色图像, 提高红外图像在视觉和信息传递的效果, 其算法框架和效果如图 4 所示。这种方法的成功程度取决于所使用的彩色参考图像数据集的质量和多样性。次年, SUAREZ P L^[11] 等提出了一种新颖的近红外图像着色方法, 采用具有三重模型的深度卷积生成对抗网络架构。这种方法旨在通过独立学习每个颜色通道来改进着色过程, 加快收敛速度, 并提高着色后的近红外图像与其相应的地面实况之间的相似度。虽然方法在收敛速度和色彩相似性等方面较以前的方法有了显著的改进, 但仍局限于处理边缘细节和不同图像区域的色彩一致性方面。

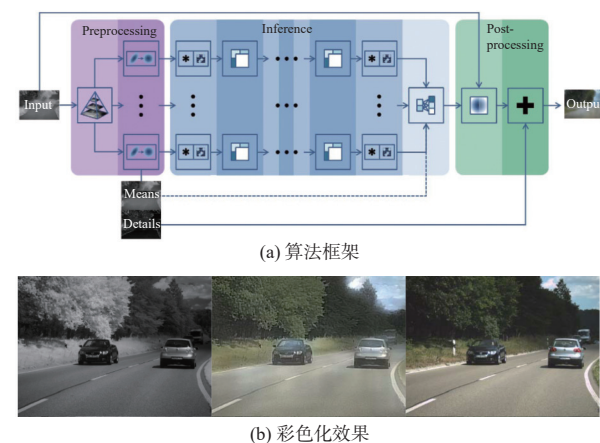


图 4 双边滤波结合的近红外图像彩色化样例
Fig. 4 Example of colorization for near-infrared images combined with bilateral filtering

2019 年, LIU M^[12] 等首次使用长波红外图像作为输入, 基于双生成对抗网络 TGC-Net(thermal-gray-color net) 的非配对热红外-可见光图像转换, 通过两级网络分别将红外图像转为可见光灰度图像, 再将可见光灰度图像转为彩色可见光图像。2020 年, KUANG X^[13] 等利用粗到细生成器将热红

外图像转换为逼真的 RGB 图像。通过引入一个综合损失函数,将内容损失、对抗损失、感知损失和总变化损失结合起来,以提高图像质量。这种方法旨在通过确保在转换过程中保留全局和局部细节,解决热图像和可见光图像之间的差距。次年,LI S^[14] 等提出了一种将未配对的红外视频转换为可见光视频的方法 I2V-GAN(infrared-to-visible generative adversarial network)(如图 5 所示),模型包含 3 种类型的约束:对抗约束,用于生成与真实帧相似的合成帧;循环一致性与感知损失,用于有效的内容转换和风格保持;相似性约束,用于增强空间和时间域的内容和运动一致性,需要大量计算资源进行训练,转换性能可能因输入红外视频和目标可见光视频的具体特征而异,且可能无法很好地适用于所有红外-可见光转换场景,尤其是那些涉及复杂和高动态场景的场景。

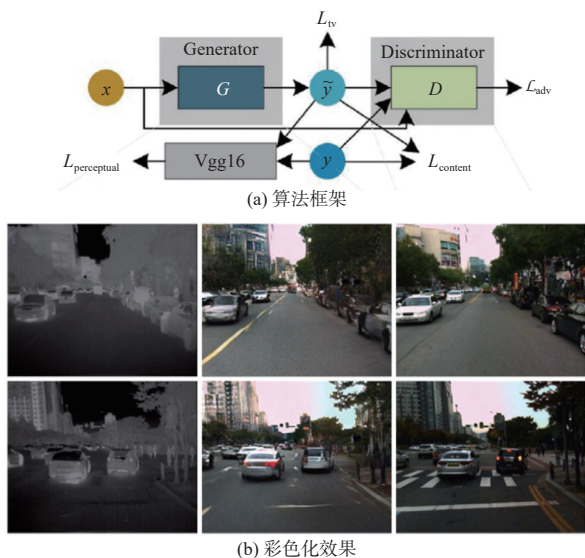


图 5 条件生成对抗网络结构彩色化样例

Fig. 5 Example of colorization for conditional generative adversarial network structure

2022 年, HU X^[15] 等设计的 QS-Attn (quantized sparse attention) 模块可为对比学习选择重要的锚点,旨在最大化源图像和翻译图像之间的互信息,通过确保翻译保留源图像的重要特征来提高图像质量。该方法使用概率分布来选择锚点,重点关注具有更重要特征的区域,模型可能会过度关注背景,减少对前景物体的关注,这可能导致在图像转换时,关键的前景细节没有背景元素保存得好,从而影响翻译图像的整体质量和实用性。同年, YANG S^[16] 等提出的 GMA-CycleGAN (gray mask attention cycle-consistent generative adversarial netw-

ork) 通过引入掩膜注意模块和感知损失项来解决颜色混淆和边缘模糊等问题,可将热红外图像转换为灰度可见光图像,然后再转换为彩色可见光图像(如图 6 所示)。该方法提高了翻译图像的准确性和质量,使其在实际应用中更加有用。



图 6 GMA-CycleGAN 红外彩色化样例

Fig. 6 GMA-CycleGAN infrared colorization example

综上所述,随着红外图像彩色化技术的不断发展和系统方法的持续优化,图像的色彩真实性、细节保留和处理效率都得到显著提高。许多创新方法的应用(如条件生成对抗网络、I2V-GAN、QS-Attn 模块和 GMA-CycleGAN 等),使得红外图像在彩色化和视频转换方面取得了显著进步,改善了图像的视觉效果,可望提高无人机、无人车等自动化系统在夜视监控、驾驶和紧急救援中的性能和可靠性,推动红外图像彩色化技术在实际应用中的发展与应用。

2 单通道热成像自然感彩色化方法

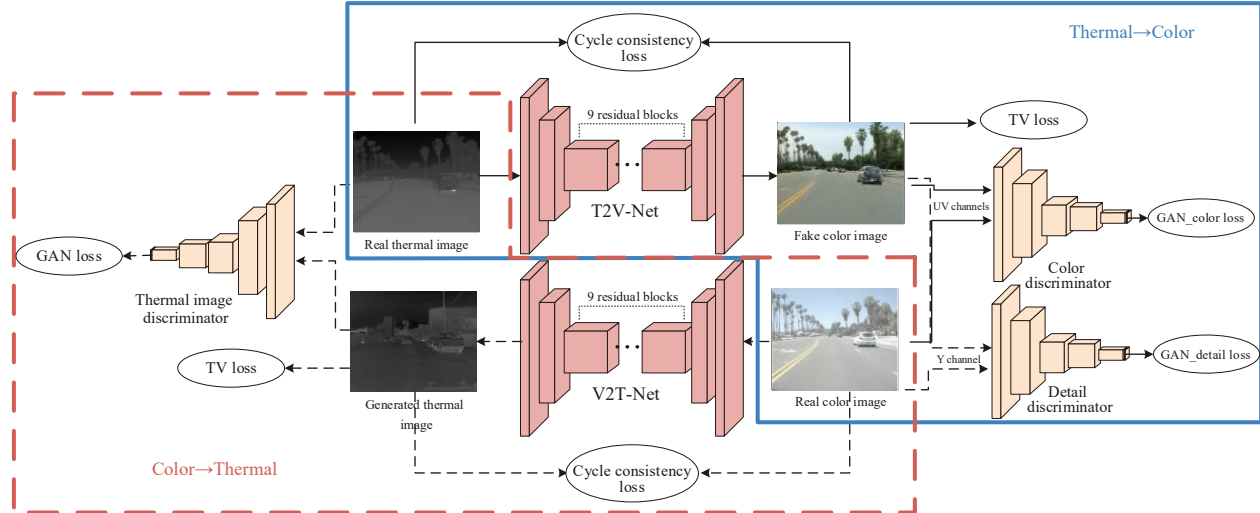
2.1 基于多判别器 CycleGAN 的 TIV(thermal image visualization) 网络

利用 CycleGAN 将红外(热)图像直接转换为类彩色可见光图像是当前主流的研究方向^[6,17-18]。由于红外图像的辐射成像特性,生成器在将单通道热红外图像转换为三通道类彩色可见光图像时面临 2 个任务:任务 1 是将辐射测量型图像细节转换为反射型图像细节;任务 2 是将单通道灰度图像转换为三通道彩色图像。对于这两个毫不相关的任务,传统的 CycleGAN 并不能很好地胜任。

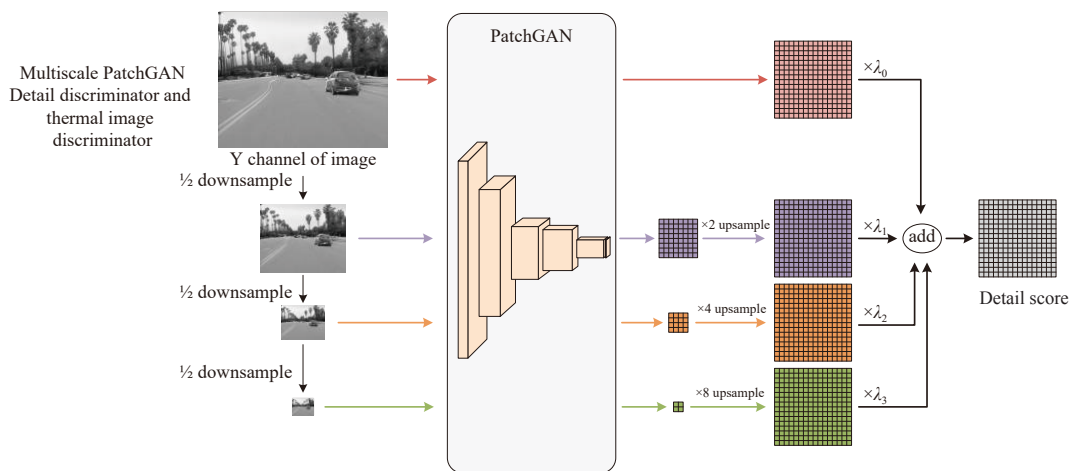
基于以上问题,我们采用了如图 7 的一种基于 YUV 色彩空间的多判别器循环一致对抗网络,通过设计如图 7(b) 的独立纹理判别器来控制生成器从辐射测量型到反射型图像细节的转换,以及

如图 7(c) 的颜色判别器来控制生成器从单通道灰度图像到三通道彩色图像的转换。算法运行在 YUV 颜色空间, 生成图像的 Y 通道和 UV 通道分

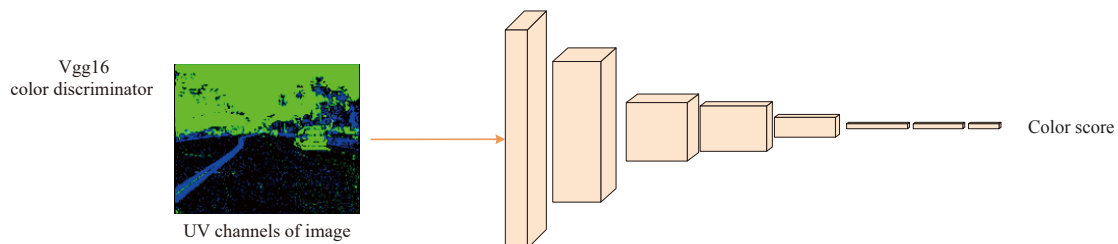
别由多个判别器进行纹理细节和图像颜色的判别, 从而简化了学习过程, 同时确保了图像的色彩风格和细节得到适当表现。



(a) 热图像自然彩色化网络框架



(b) 多尺度PatchGAN框架



(c) 颜色判别器框架

图 7 生成器中各组件的网络架构

Fig. 7 Network architecture of each component in generator

如图 7(a) 的 TIVNet 分为以下步骤: 热红外图像输入生成器, 生成的三通道图像被定义为 YUV 色彩空间图像, 其中 Y 通道图像进入纹理判别器, UV 通道进入颜色判别器。在未配对图像的训练

框架中, 算法的损失函数由总变分损失、生成对抗网络损失和循环一致性损失组成。总变分损失旨在提升图像空间平滑性, 而 GAN 损失和循环一致性损失确保转换图像的质量。

$$L_{TV}(y) = \sum_{i,j} \sqrt{|y_{i+1,j} - y_{i,j}|^2} + \sqrt{|y_{i,j+1} - y_{i,j}|^2} = \sum_{i,j} |y_{i+1,j} - y_{i,j}| + |y_{i,j+1} - y_{i,j}| \quad (1)$$

式中: $y_{i,j}$ 表示图像在 (i, j) 坐标处的像素值。

GAN 损失是通过 GAN 网络中判别器的结果反馈给生成器的损失函数。针对多判别器方案, 损失函数为

$$L_{GAN}(G_X, D_Y^{fem}) = E_{(y) \sim p_Y(y)} [\log D_Y^{\text{Texture}}(y) + \log D_Y^{\text{Color}}(y)] + E_{(x) \sim p_X(x), (z) \sim p_Z(z)} [\log(1 - D_Y^{\text{Texture}}(G_X(x))) + \log(1 - D_Y^{\text{Color}}(G_X(x)))] \quad (2)$$

式中: $E_{(x) \sim p_X(x)}$ 为对从概率分布 $p_X(x)$ 中采样输入空间 x 的期望值, 代表生成网络输出相对于输入的期望值; $G_X(\cdot)$ 是生成网络; $D_Y^{\text{Texture}}(\cdot)$ 是纹理判别器; $D_Y^{\text{Color}}(\cdot)$ 是颜色判别器; x 和 y 分别是输入/输出图像。

在使用非配对图像进行训练时, 没有真实图像

可与生成的图像进行比较, 故引入第 2 个生成器将生成的可见图像转换为重建的红外图像, 并引入了学习的感知图像块相似性 (learned perceptual image patch similarity, LPIPS)^[19]。LPIPS 比传统方法更符合人类感知。循环一致性损失如下:

$$L_{cyc}(G, F) = E_{(x) \sim p_{\text{data}}(x)} [L_{\text{lpips}}(F(G(x)) - x)] + E_{(y) \sim p_{\text{data}}(y)} [L_{\text{lpips}}(G(F(y)) - y)] \quad (3)$$

式中: $G(\cdot)$ 是将红外图像生成可见图像的生成器; $F(\cdot)$ 是将可见图像生成红外图像的生成器; x 是红外域图像; y 是可见域图像; $L_{\text{lpips}}(\cdot)$ 是学习的感知图像块相似性函数。

$$L_{\text{lpips}}(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l)\|_2^2 \quad (4)$$

式中: x 和 x_0 是正在比较的图像; l 表示不同的层次; H_l 和 W_l 分别表示第 l 层特征图的高度和宽度; w_l 表示第 l 层的权重, 调节每一层对整体相似度评分的贡献; 表达式 $\hat{y}_{hw}^l$ 表示预测的特征, $\hat{y}_{0hw}^l$ 表示在第 l 层位置 (h, w) 的参考或目标特征; 符号 $\odot$ 表示逐元素乘法, 即对应位置的元素乘积; 符号 $\|\cdot\|_2^2$ 表示平方的 L2 范数, 即向量元素平方和的总和。

生成器的总损失函数为

$$L_{\text{total}} = L_{\text{GAN}} + \lambda_{\text{cyc}} L_{\text{cyc}} + \lambda_{\text{TV}} L_{\text{TV}} \quad (5)$$

式中: L_{GAN} 表示生成对抗网络损失; L_{cyc} 表示循环一致性损失; L_{TV} 表示总变分损失; 参数 λ_{cyc} 和 λ_{TV} 用于调整相应损失项的权重。

对于无监督学习, 依赖全参考图像的评价指标 (如峰值信噪比和结构相似性指数) 并不适用。因此, 由于没有与输入热图像一一对应的彩色可见光图像作为参考, 无法基于这些指标进行评价。为了评估不同热图像可视化方法的有效性, 我们采用了以下无参考图像质量评价指标来客观测量热图像可视化的效果。

NIQE^[18]: NIQE (natural image quality evaluation) 是一种基于零域统计特征的无参考图像评价指标, 不依赖于主观评分且计算复杂度较低。

NIMA^[19]: NIMA (neural image assessment) 是一种基于图像分类卷积神经网络的图像质量评价方法。其首先收集包含人类对图像技术和美学评分的大规模图像数据集, 并基于该数据集训练图像质量评分模型。该方法既评估图像中的信息保留情况, 又兼顾图像的视觉美感, 更符合人类感知。

YOLO-score: 受 ISOLA 等人提出的 FCN-score (fully convolutional network score) 启发, 我们提出了基于检测网络 YOLOv3^[20] 的 YOLO-score (you only look once-score) 作为评价指标。由于 FCN-score 依赖图像分割, 而 FLIR 数据集并未提供场景分割标签, 仅提供符合 COCO 数据集格式的目标检测边界框标签, 因此我们采用在可见光图像 COCO 数据集上训练的 YOLOv3 模型, 对通过热图像转换得到的可见光图像进行目标检测, 并计算检测结果与真实值之间的平均精度作为 YOLO-score。鉴于 YOLOv3 模型是在可见光图像数据集上训练的, 若转换得到的可见光图像经 YOLOv3 检测后结果更接近真实值, 则说明该转换结果更接近真实的可见光图像, 从而证明 YOLO-score 能有效评估热红外图像的模型转换性能。

纹理判别器与颜色判别器采用并行架构实现多维度约束: 纹理判别器通过提取红外图像的局部纹理特征, 引导生成器重建可见光域的细节纹理, 有效抑制生成图像的过度平滑化倾向; 颜色判别器则通过分析目标域的色度分布特性, 建立色彩映射关系网络。两判别器通过梯度反向传播机制实现协同优化, 最终形成纹理-色彩解耦的联合约束策略, 显著提升跨模态图像彩色化的视觉保

真度。为说明复合损失函数及 TIVNet 各模块的有效性, 我们通过消融实验评估了不同损失组合: Loss1 为 L_{cyc} , Loss2 为 $L_{cyc}+L_{TV}$, Loss3 为 $L_{cyc}+L_{TV}+L_{l_{pips}}$, Structure1 为 CycleGAN, Structure2 为 CycleGAN+多判别器, Structure3 为 CycleGAN+多判别器+YUV 图像。在 FILD 数据集上, TIVNet 在使用

循环损失、TV 损失、LPIPS 损失、PatchGAN 及多判别器的组合时, NIQE 提升 20.35% 和 19.60%, NIMA 至少提升 4.17% 和 4.29%, YOLO-score 至少提升 60.69% 和 54.87%(见表 1)。本数据集也呈现相似趋势, 图 8 展示了不同损失与结构的实验结果, 验证了 TIVNet 复合损失函数的合理性。

表 1 在 FLIR 数据集的消融实验

Table 1 Ablation experiments in FLIR dataset

Method	FLIR dataset			Our data		
	NIQE	NIMA	YOLO-score	NIQE	NIMA	YOLO-score
Loss1	8.7664	4.5978	0.1572	8.8863	4.3224	0.1571
Loss2	7.8376	4.5884	0.1662	7.8843	4.4782	0.1580
Loss3	6.9826	4.7894	0.2526	7.1444	4.5080	0.2433
Structure1	7.5195	4.5884	0.1433	7.6332	4.4724	0.1570
Structure2	7.2344	4.6832	0.2455	7.4332	4.4782	0.1880
Structure3	6.9826	4.7894	0.2526	7.1444	4.5080	0.2433

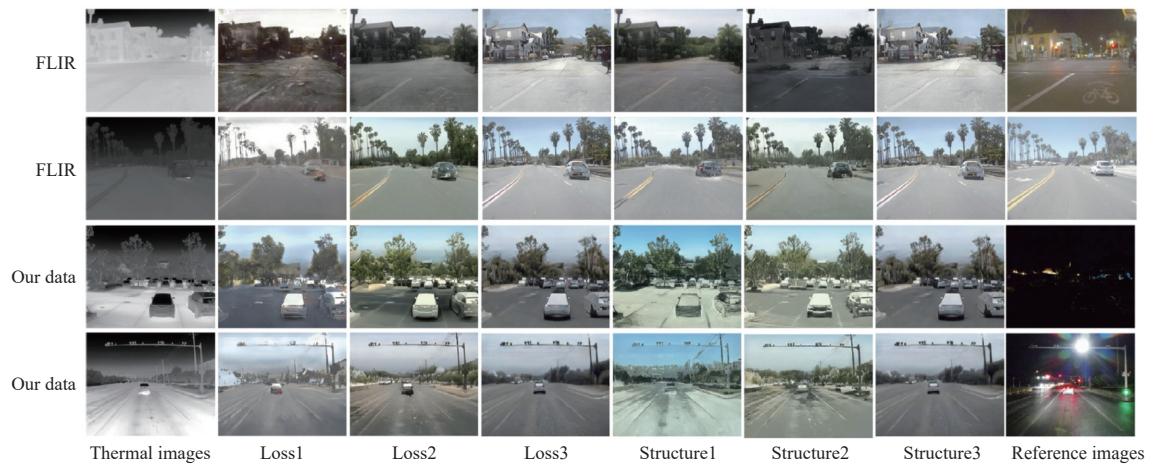


图 8 所提出的 TIVNet 在 FLIR 数据集和我们数据集上进行的关于不同损失函数和结构的消融研究

Fig. 8 Ablation studies of different loss functions and structures for the proposed TIVNet on FLIR dataset and our dataset

关于参数 λ_{cyc} 、 λ_{TV} 的主要目的是鼓励生成器生成可以通过第 2 个生成器准确转换回原始图像的图像。当生成的图像与原始图像非常相似时, L_{cyc} 接近于 0; 相反, 如果存在显著差异, L_{cyc} 值会增加。当 λ_{TV} 的权重设置为较小的值时, 模型变得更容易受到对抗损失的影响, 其中生成器优先生成逼真的图像, 而判别器则专注于准确的分类。在这种情况下, 生成器可能会忽略一致性要求, 导致生成的图像与原始图像之间的相关性减弱。另一方面, 当权重过小时, 模型会变得对训练数据分布中的噪声或变化更加敏感, 从而导致不稳定的训练过程, 其中总损失会反弹或波动。如图 9(a) 所示, 总损失函数在训练的早期阶段经历快速下降, 随后又上升。这表明模型最初学习的是简单的映射, 而没有来自对抗损失的充分约束, 可能导

致不稳定的训练过程, 总损失出现波动。然而, 对于 L_{cyc} 而言, 过大的权重可能会触发模式崩溃, 即生成器倾向于将所有输入映射到相同的输出, 从而导致图像变换的多样性和丰富性丧失。因此, 考虑到这些因素, 选择较小的权重以确保稳定性, 例如 $\lambda_{TV}=10$, 代表了更明智的选择。关于参数 λ_{TV} , 如图 9(b) 所示, L_{TV} 的大小对训练速度和收敛的影响很小。然而, 将 L_{TV} 权重设置得过小可能会导致图像过于平滑, 造成细节信息的丢失和过度模糊的外观, 从而损害真实感。

相反, 过大的 L_{TV} 权重可能会导致生成器过度优先考虑图像的平滑度, 从而可能忽略其他重要特征。这可能导致生成的图像过于简单, 缺乏真实图像的复杂性和细节, 甚至引入失真。图 9(b) 展示了使用不同的 λ_{TV} 值训练 300 个 epoch 后的结

果的比较分析,证实了这一观察结果。

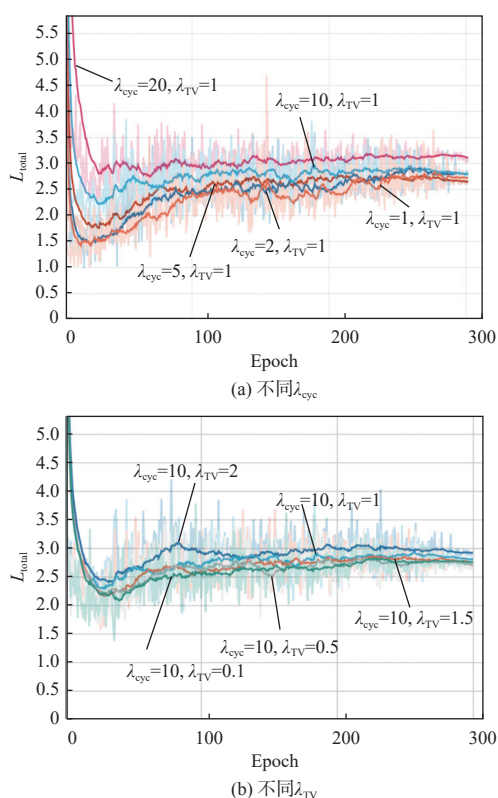


图9 损失函数随着不同 λ_{cyc} 和 λ_{TV} 的变化情况

Fig. 9 Loss function changes with different λ_{cyc} and λ_{TV}

最终,式(5)中设置了 $\lambda_{cyc} = 10$, $\lambda_{TV} = 1$ 。网络训练了300个epoch, batch size为2。Adam优化器被

应用于训练,前100个epoch的学习率为0.0002,然后在epoch 300线性衰减至0。

2.2 基于区域自分割的语义引导扩散模型网络

生成扩散先验已被证明在条件图像生成方面的有效性^[20],为红外图像转彩色可见光图像任务(I2V)提供了一个有效的解决方案:将其视为条件图像生成,直接利用红外图像和红外图像的语义分割掩码作为条件输入。然而,红外图像由于红外焦平面探测器以及电路等因素会产生非均匀性和时间噪声,直接将其作为控制信号会导致不稳定并产生伪影。

针对I2V任务,采用了一种通用的两阶段流程,此流程涵盖用于消除图像自身非均匀性的非均匀性校正模块,以及一个致力于图像内容再生的生成模块。这两个阶段是解耦的,并可以独立优化。这种方法允许使用任何现成的或自行训练的非均匀性校正模块,以校正红外图像的非均匀性噪声。尤为重要是,鉴于生成过程仅以红外图像及其语义分割内容作为条件,不受到图像非均匀性的干扰。这种两阶段流程为I2V问题提供了一种灵活、稳定和统一的解决方案。

2.2.1 非均匀性校正阶段

第一阶段的目标在于消除由红外传感器所引发的非均匀性,且不引入任何新的内容。鉴于不同的红外焦平面探测器具有不完全相同的非均匀性。如图10所示,设计解耦的非均匀性校正模块

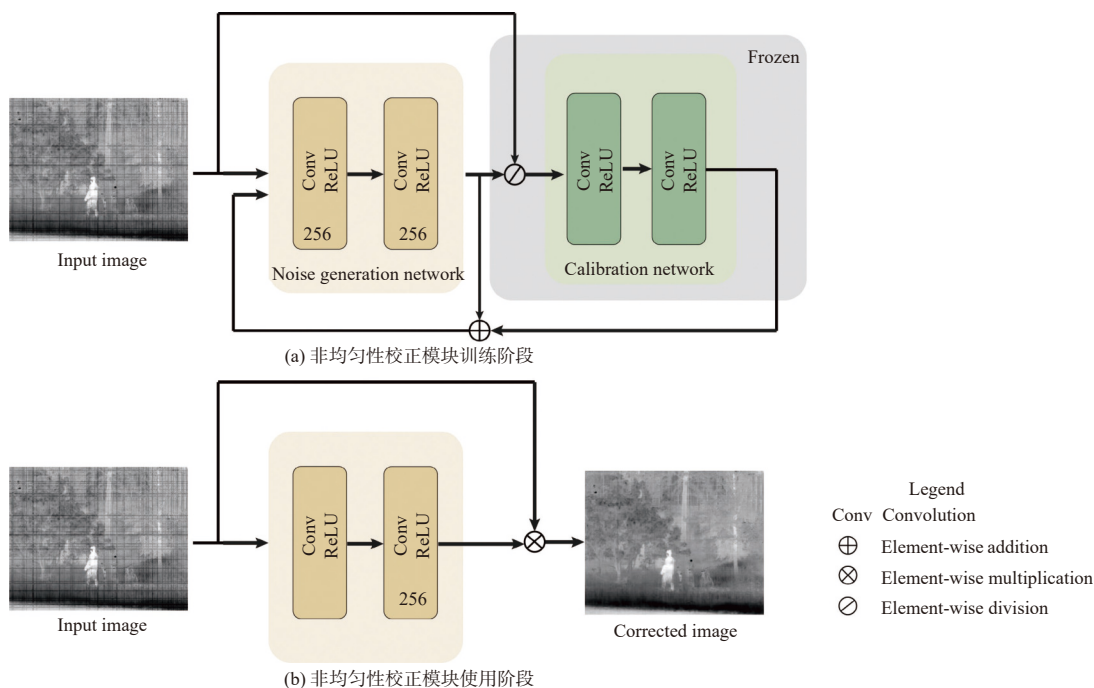


图10 非均匀性校正模块架构

Fig. 10 Architecture of non-uniformity correction module

能保证在进入后续图像处理之前, 针对每种非均匀性均能以有效的校正。

非均匀性校正过程先对图像噪声进行估计, 然后减去所估计的噪声, 从而获得校正后的图像。通过迭代校准噪声估计的方法, 构建噪声估计过程:

$$F(x_t): \begin{cases} u_t = H_\theta(x_t), x_0 = y \\ x_{t+1} = x_t + u_t \end{cases} \quad (6)$$

式中: $F(x_t)$ 表示估计噪声模型网络 F ; 输入变量 x_t 代表迭代过程中的当前状态; $u_t = H_\theta(x_t)$ 定义了经过校准后的估计噪声, 表示通过估计模型 H_θ 对当前状态 x_t 进行噪声估计的结果; $x_{t+1} = x_t + u_t$ 定义了迭代过程。

自校准模块从分析每个阶段之间的关系, 以确保训练过程不同阶段的噪声输出收敛到同一状态:

$$g(x_t): \begin{cases} z_t = y \odot x_t \\ s_t = K_\theta(z_t) \\ v_t = y + s_t \end{cases} \quad (7)$$

式中: z_t 是在不同阶段输入数据下的归一化噪声估计; y 是输入数据; x_t 是 t 阶段估计的噪声输出; $K_\theta(\cdot)$ 是自校准模块; s_t 是 t 阶段对噪声输出进行的校正量; v_t 是校准后的噪声输出。

网络整合了校准网络模块和噪声估计网络模块。前者使用编码器-解码器结构, 输入红外图像, 输出图像空间先验信息 (包括热源位置、温度统计、噪声概率等)。后者使用多尺度 DenseNet 结构, 输入为原始红外图像, 经多尺度提取特征, 输出信号范围在 0~1 间的估计噪声概率。自校准模块将校准网络输出的空间先验特征, 与噪声生成网络解码特征进行匹配分析。计算修正项以优化

噪声。最终的校正图像为输入红外图像减去估计噪声概率与输入红外图像的乘积。校准网络提供的空间先验和噪声生成网络的生成能力相结合, 使得系统能够在不同场景下都实现高质量的红外图非均匀性校正。

平均绝对误差 (mean absolute error, MAE) 作为阶段损失函数为网络训练提供更灵敏、更准确的优化评价:

$$l_{L1}(P) = \sum_{p \in P} \|P(p) - V(p)\|_1 \quad (8)$$

式中: P 表示网络生成的预测图像; V 表示逼近的真实图像或目标图像的网络输出; 符号 $p \in P$ 表示预测图像中的每个像素; $P(p)$ 和 $V(p)$ 分别表示预测图像和真实图像中位置 p 处的像素值。

2.2.2 扩散模型彩色化阶段

提出的区域自分割的语义引导扩散模型 (RSDM) 图像域转换架构如图 11 的阶段 2 所示, 模型接收红外图像、彩色可见光图像及红外图像的自分割蒙版作为输入, 旨在重建高质量的彩色可见光图像。RSDM 框架包含 4 个主要部分: 预训练的变分自动编码器 (variational autoencoder, VAE)、带有跳过连接的去噪 U 形网络 (U-Net)、SEM (semantic segmentation) 自分割网络和适配器。通过利用 VAE 对彩色可见光图像进行编码与解码, 将经典的去噪扩散概率模型 DDPM (denoising diffusion probabilistic models) 中的扩散和去噪过程从像素空间转移到潜在空间, 从而增强训练过程的稳定性并降低计算资源的消耗。U-Net 作为一种条件噪声预测器, 在扩散模型的每个时间步 T 中, 包含多个

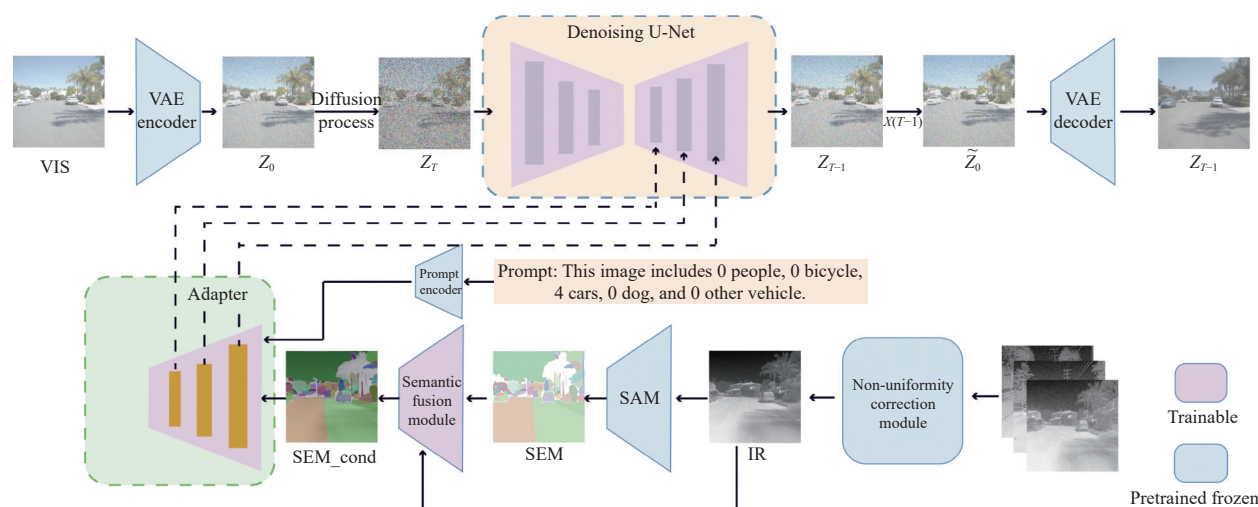


图 11 RSDM 算法框架

Fig. 11 Framework of RSDM algorithm

编码层、中间层和解码层,其中大部分由残差模块、交叉注意模块和自注意模块组成,通过迭代预测噪声以获得无噪声数据。语义融合模块(semantic fusion module)融合来自 SEM 自分割掩码和红外图像的信息,生成封装了丰富语义信息的条件特征映射。最后,适配器将该条件特征转换为与 U-Net 网络架构兼容的多尺度特征。这些特征被逐个元素地添加到 U-Net 的编码和解码层中,确保在生成过程实现有效的特征融合和引导。

为提高生成网络转换不同对象红外图像的能力,提出了一种利用语义先验对输入图像进行预处理的方法,将来自输入图像的信息与语义掩码相结合,这种方法类似于在自然语言处理中使用

语义先验优化图像特征。如图 12 所示,输入的红外图像由 SEM 网络处理以生成区域掩码。图中显示了区域掩码是通过使用不同颜色分割不同对象获得的。语义融合模块计算图像特征的语义感知,并使用跨模态相似性生成语义感知映射。首先使用卷积层将红外图像转换为统一维度,然后应用转置注意力机制来生成注意力图,从而降低了计算成本^[21]。预处理后的语义注意力图为

$$A^b = \text{Softmax}\left(W_k(F_x^b) \times W_q(F_{(x,em)}^b) / \sqrt{C}\right) \quad (9)$$

式中: $W_k(\cdot)$ 和 $W_q(\cdot)$ 是用于在注意力计算中获得 k 和 q 分量的卷积层; C 表示特征通道; A^b 为语义感知注意力图,表示 F_x^b 和 $F_{(x,em)}^b$ 之间的注意力关系。

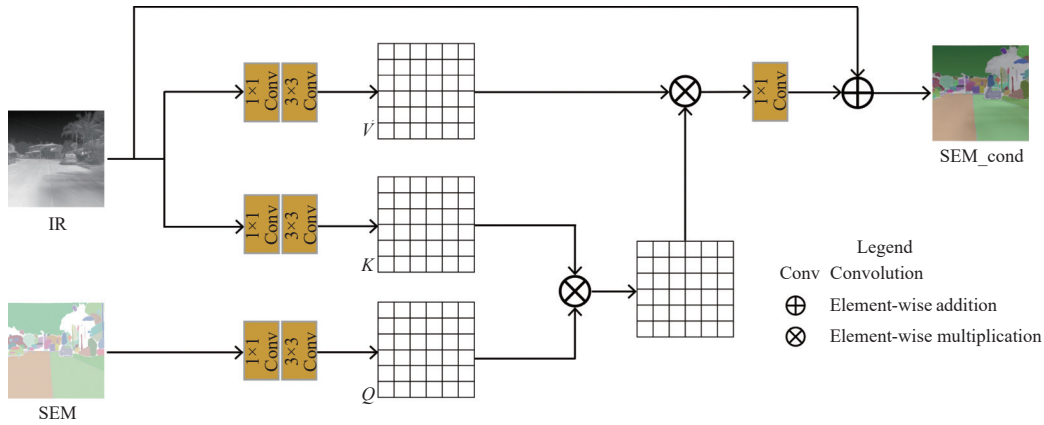


图 12 Semantic fusion module 框架

Fig. 12 Framework of semantic fusion module

接着,使用 A^b 构造图像特征 F_x^b

$$F_o^b = \text{FN}\left(W_v(F_x^b) \times A^b + F_x^b\right) \quad (10)$$

式中: FN 表示前馈网络; F_o^b 是语义融合模块最终细化的特征图,用于增强网络的注意力信息。

通过对原始输入红外图像 x 执行多次卷积以提取特征,将注意力信息 F_o^b 与提取的特征组合以获得增强信息,用作为适配器模块的输入图像

$$F_{\text{fusion}}^b = \text{conv}(F_x^b + F_o^b) \quad (11)$$

式中 conv 表示卷积运算。

2.2.3 适配器

语义融合模块能够有效地将红外图像及其自分割掩码信息集成,并输入至适配器中。该适配器的主要功能在于生成多尺度特征,并使这些特征与预训练的稳态扩散模型的先验知识进行对齐。适配器模块的结构较为简单,主要构成包括堆叠的卷积层、残差块和下采样块。适配器模块通过逐渐增加特征映射的通道数以及减少其空间

大小,生成 3 种不同大小的条件特征。随后,这些条件特征被按元素方式添加至 U-Net 网络每一层的输出中,以此来指导稳态扩散模型的生成过程。

3 验证实验

模型的训练和测试均采用 FLIR 公开数据集^[22]。在训练过程中,为确保彩色可见光数据具有参考性,排除了夜间数据,最终使用 3580 对红外和彩色图像作为训练集。测试数据集在 FLIR 白天数据基础上增加了夜间红外图像,以验证算法是否能够将夜间红外图像转换为类白天效果的彩色可见光图像。

3.1 实验和评价方法

训练非均匀性校正模块 150 k 次迭代(批量大小=8),然后采用稳态扩散 2.1-base1 作为生成先验,并微调所提出的 RSDM 进行 80 k 次迭代(批量大小=4),Adam 用作优化器。对于前 30 k 次迭代,学习率设置为 10^{-4} ,然后对于后续的 50 k 次迭代,

学习率降低到 10^{-5} 。训练过程在 512×512 像素下使用 4 个 NVIDIA 3090Ti GPU 进行。对于生成数据,除了将传统指标 LPIPS^[19] 作为图像的质量评估,还包括了几个无参考图像质量评估指标: NIQE^[23] 和 FID^[24]。

LPIPS 通过从预训练的卷积神经网络中提取的特征,以解决传统指标(如峰值信噪比 PSNR 和结构相似性 SSIM)的局限性:

$$I_{\text{LPIPS}}(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \|\mathbf{w}_l \odot (\phi_l(x)_{h,w} - \phi_l(y)_{h,w})\|_2^2 \quad (12)$$

式中: x 和 y 代表两幅输入图像; ϕ_l 代表层 l 处预训练网络的特征提取函数; H_l 和 W_l 代表层 l 处特征图高度和宽度; $\phi_l(x)_{h,w}$ 和 $\phi_l(y)_{h,w}$ 表示图像 x 和 y 在 l 层的特征图中位于位置 (h, w) 的特征向量; $\mathbf{w}_l$ 表示第 l 层的权重向量,用于调整不同层特征的影响; $\odot$ 表示元素相乘运算; $\|\cdot\|_2$ 表示 L2 距离。

NIQE 是一种无参考图像质量评估指标,无需参考图像即可测量图像的感知质量,依靠自然图像的统计特征来评估图像质量:

$$I_{\text{NIQE}}(x) = \left[(\mu_x - \mu_{\text{model}})^T \left(\frac{\Sigma_x + \Sigma_{\text{model}}}{2} \right)^{-1} (\mu_x - \mu_{\text{model}}) \right]^{1/2} + \left[\frac{1}{2} \ln \left(\frac{|\Sigma_{\text{model}}|}{|\Sigma_x|} \right) \right] \quad (13)$$

式中: x 表示被评估的图像; μ_x 表示从图像 x 中提取的特征的平均向量; μ_{model} 表示从自然图像模型中提取的特征的平均向量; Σ_x 表示从图像 x 中提取

的特征的协方差矩阵; Σ_{model} 表示从自然图像模型中提取的特征的协方差矩阵; $|\cdot|$ 表示行列式。

FID (Fréchet inception distance) 是一种用于评估生成模型生成的图像质量的指标,测量真实图像和生成图像的特征向量之间的距离,同时捕捉均值和协方差差异:

$$I_{\text{FID}} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}) \quad (14)$$

式中: μ_r 为真实图像特征向量的平均值; μ_g 为生成图像特征向量的平均值; Σ 为向量的协方差矩阵; $\|\cdot\|_2^2$ 为欧几里得距离的平方; $\text{Tr}(\cdot)$ 为矩阵的迹线; $(\Sigma_r \Sigma_g)^{1/2}$ 为协方差矩阵乘积的矩阵平方根。

3.2 与不同转换方法的比较

采用 CycleGAN^[25]、TICCGAN(thermal infrared to color cycle-consistent generative adversarial network)^[26]、QS-Attn^[27] 作为对比算法,并从多层次指标的角度对这些方法与本文方法进行定量比较。将红外图像作为输入,旨在预测相应的彩色化图像。CycleGAN 是一种广泛应用于图像变换任务的深度神经网络模型,能够根据输入图像预测相应的目标图像;TICCGAN 是一种结合了内容、对抗性、感知和总变异损失的多目标函数的深度神经网络模型,能够根据输入图像,通过更恰当的损失函数更好地预测相应的目标图像;QS-Attn 设计了一种查询选择注意力模块,通过比较源域中的特征距离,生成每行具有概率分布的注意力矩阵。本文以红外图像作为输入,预测相应的彩色化图像。

图 13 给出了对应的视觉比较, TIVNet 和 RSDM

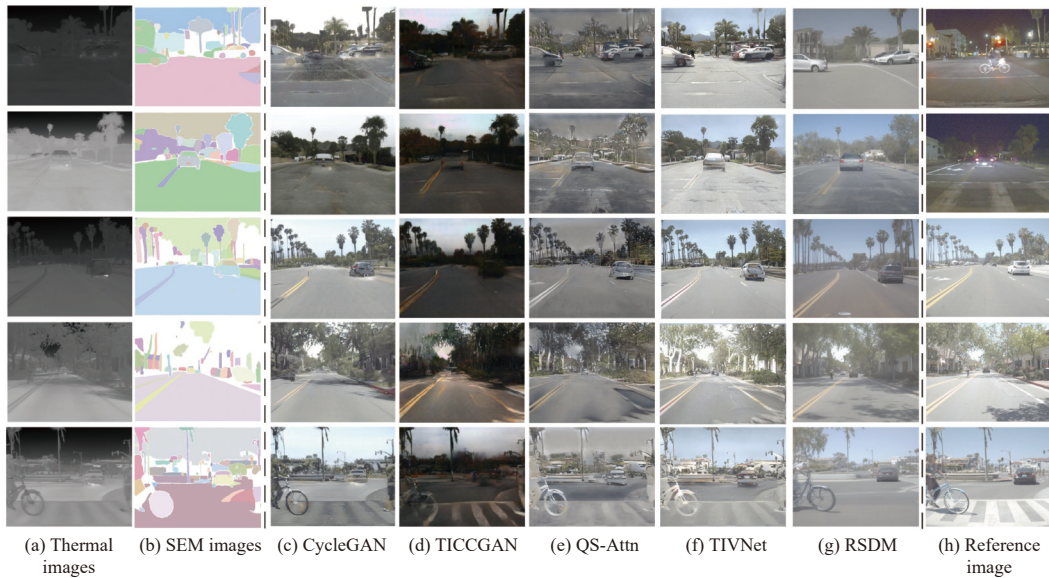


图 13 在 FLIR 数据集上不同算法的结果

Fig. 13 Results of different algorithms on the FLIR dataset

能够产生比基于 GAN 的方法更清晰的结果, 后者的输出往往过于平滑。与其他方法相比, RSDM 的结果更加逼真, 例如图中的行人、车辆、自行车等。表 2 给出了 FLIR 数据集的定量比较。可以看出, RSDM 在所有指标中均取得了最佳成绩。这验证了与对比方法相比, RSDM 在处理具有挑战性的红外转彩色可见光图像方面具有优越性。

在一些特殊情况下, 改进的稳定扩散模型 RSDM 不仅能将视觉效果较差的红外图像转换为更好的彩色化图像, 而且还能借助结构化边缘蒙版的引导效果, 解决可见光图像的过曝问题。如图 14 所示, 由于彩色可见光相机的动态范围限制, 在进入隧道之后, 隧道外的场景便完全变得不可见。

然而, 红外相机却能够完整地记录下隧道外的场景。通过 RSDM 的有效生成能力, 可以解决彩色可见光图像过曝的问题, 从而提升视觉效果。

表 2 在 FLIR 数据集的定量评价比较

Table 2 Quantitative evaluation comparison on FLIR dataset

Method	NIQE ↓	LPIPS ↓	FID ↓
CycleGAN	7.2447	0.5756	83.43
TICCGAN	7.1195	0.5253	80.33
QS-Attn	7.3367	0.6334	76.69
TIVNet	6.9826	0.4827	72.23
RSDM	6.2442	0.4344	48.89

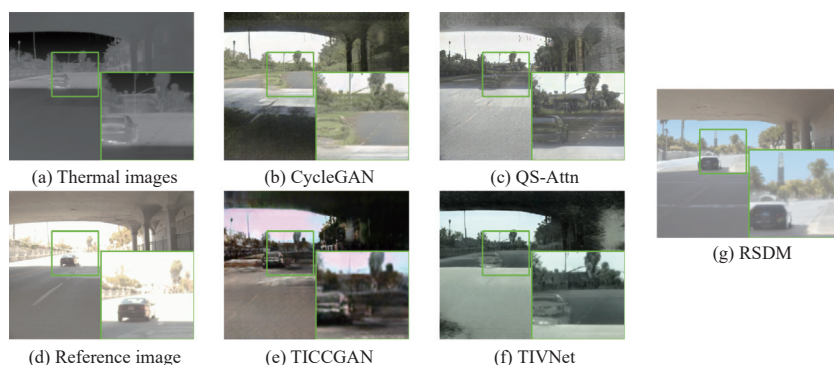


图 14 RSDM 在特殊场景下解决可见光图像过曝的问题

Fig. 14 Problem of overexposure in visible light images solved by RSDM under special scenarios

3.3 无人机昼夜单通道热成像的自然感彩色化

我们已在大疆御 MAVIC 2 无人机系统+联想 GPU 移动端服务器 (如图 15 所示) 平台上实现热图像自然感彩色化 TIVNet 网络的实时运行。MAVIC 2 的红外相机为氧化钒非制冷微测辐射热计, 160×120 像素, 像元间距 $12 \mu\text{m}$, 下传图像视频 640×512 像素, 帧频 60 Hz; 彩色可见光相机为 $6.17 \times 4.55 \text{ mm}$ ($1/2.3$ 英寸) CMOS, 4000×3000 像素。理想 GPU 移动端服务器作为遥控器端, 主要性能指标为:



图 15 大疆御 MAVIC 2 双光版无人机平台

Fig. 15 DJI Royal MAVIC 2 dual-light UAV platform

CPU i9-13900HK, 内存 32 GB, NVIDIA GeForce GTX4080 12 GB。其中无人机的可见光/红外图像数字视频信号通过 HDMI 信号传输的方式输入到 GPU 移动端服务器进行图像实时处理, 红外图像视频自然感彩色化处理后达到 40 Hz, 具备实际应用的条件。

图 16 给出了夜间无人机 MAVIC 2 飞行在停车场上空和建筑物楼上对地面拍摄的热图像转换为自然感彩色可见光的效果, 其不仅增强了图像的视觉清晰度, 而且提升了对特定目标的识别能力。需要处理网络的训练不一定需要可见光与红外图像的配对, 即应用时到达新场景需要进行新的网络训练时, 可以只采集红外图像就可完成新场景的适应性训练, 这为热成像的自然感彩色转化拓展提供了有利条件。

图 17 给出某无人机采集的野外场景红外图像经过 TIVNet 算法处理的效果, 可以看出, 虽然算法网络没有针对性的学习训练, 但依然获得了较好

的彩色化效果,特别是在视频动态场景观察条件下,对人员、空中无人机、类自然的山景与天空的色彩再现都更有利于人眼视觉的探测与识别,能提升夜间的侦察效果。同时,也可以看到,在人员等场景细节目标的再现方面尚不完善,这也是本文 RSDM 图像域转换架构需解决的问题,目前部分解决了细部色彩转换问题,后续将重点解决处理速度的问题。



图 16 无人机飞行探测在夜间的红外转彩色可见光效果

Fig. 16 Infrared to visible color conversion effect for UAV flight detection at night

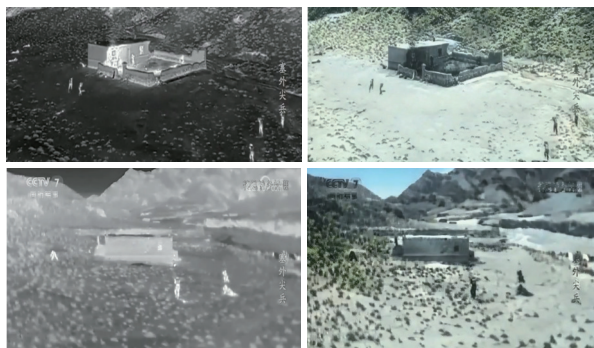


图 17 网络红外视频图像直接转化自然感彩色化效果

Fig. 17 Network infrared video images directly into natural sensing of colorization effect

4 结论

直接将热成像图像转换为自然感彩色图像已取得了重要的研究进展,展现出广泛的发展和应用前景。我们研究提出了基于 GycleGAN 网络框架的红外图像转彩色可见光图像的 TIVNet 方法,仅需一次采样步骤就能将红外图像彩色化,增强了红外图像的视觉效果和人类感知,但场景细节和小目标部分的再现可能存在一些错误。针对其中单个 UNet 结构在场景细节再现方面可能出现的误判,我们进一步提出了区域自分割的语义引

导扩散模型 RSDM,特别是在场景细节和小目标再现方面呈现出良好的结果。由于 RSDM 算法设计的特殊性,目前转换一幅彩色图像约需 50 次 Unet 采样结构,尚难以实现实时处理。后续的优化和改进策略将尝试将两者的优势结合,在保证生成效果与 RSDM 相近的前提下,减少计算量,提升算法的运行速度,促进方法的实用化。

未来针对 TIVNet 和 RSDM 算法的实时性优化,本工作实施分层加速策略:在已实现 40 Hz 实时处理的 TIVNet 算法中,通过轻量级生成对抗架构(MobileGAN)作为主干网络,在判别器网络中引入深度可分离卷积核替代常规 3×3 卷积,使模型参数量进一步降低。针对 RSDM 模型的加速需求,构建了多维优化方案:首先在 U-Net 编解码器中全面部署深度可分离卷积单元,其次采用基于通道重要性的结构化剪枝算法消除 47% 冗余参数,进一步结合师生式知识蒸馏框架实现推理加速,最后通过改进扩散过程的马尔可夫链采样策略,采用动态步长调整机制将采样次数从 1000 步压缩至 50 步。

在无人机领域,红外图像的彩色化技术不仅可以提升图像的视觉效果,还可增强无人机在复杂环境中的探测和识别能力。我们提出的方法已可通过无人机图传模块在地面处理器上实时处理红外视频,实测处理速度已达 40 Hz 帧频,可满足实际应用的要求,拓展无人机利用红外相机提升昼夜态势感知的能力。此外,该技术也可以用于无人车、夜视侦察与观瞄等广泛的领域。

参考文献:

- [1] ELMESEIRY N, ALSHAER N, ISMAIL T. A detailed survey and future directions of unmanned aerial vehicles (UAVs) with potential applications[J]. *Aerospace*, 2021, 8(12): 363.
- [2] COCHRANE D M, MANNING P A, WYLLIE T A. Un-cooled thermal imaging sensor for UAV applications[C]//Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls. Orlando: SPIE, 2001: 168-178.
- [3] 王岭雪, 史世明, 金伟其, 等. 基于 YUV 空间的双通道视频图像色彩传递及实时系统[J]. *北京理工大学学报*, 2007(3): 189-191.
WANG Lingxue, SHI Shiming, JIN Weiqi, et al. Color

- transfer and its real-time system based on a YUV space for dual-channel video images[J]. *Journal of Beijing Institute of Technology*, 2007(3): 189-191.
- [4] SHENG D, JIN W, WANG X, et al. Thermal image visualization using multi-discriminator CycleGAN with unpaired thermal-visible image training set[J]. *Infrared Physics & Technology*, 2024, 140: 105352.
- [5] REINHARD E, ADHIKMIN M, GOOCH B, et al. Color transfer between images[J]. *IEEE Computer Graphics and Applications*, 2001, 21(4): 34-41.
- [6] WELSH T, ASHIKMIN M, MUELLER K. Transferring color to greyscale images[C]//*Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*. San Antonio Texas: ACM, 2002: 277-280.
- [7] 赵源萌, 王岭雪, 金伟其, 等. 基于区域直方图统计的灰度图像色彩传递方法[J]. *北京理工大学学报*, 2012, 32(3): 322-326.
- ZHAO Yuanmeng, WANG Lingxue, JIN Weiqi, et al. Color transfer method for grayscale images based on regional histogram statistics[J]. *Journal of Beijing Institute of Technology*, 2012, 32(3): 322-326.
- [8] 刘缠牢, 谭立勋, 李春燕, 等. 红外图像伪彩色编码和处理[J]. *应用光学*, 2006, 27(5): 419-422.
- LIU Chanlao, TAN Lixun, LI Chunyan, et al. Pseudo color coding and processing of infrared images [J]. *Journal of Applied Optics*, 2006 27(5): 419-422.
- [9] ZHANG X, BAI T, LI H. Pseudo-color coding method of infrared images based on human vision system[C]//*Photonics Asia 2007*. Beijing: SPIE, 2007: 68351N.
- [10] LIMMER M, LENSCH H P A. Infrared colorization using deep convolutional neural networks[C]//*2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. Anaheim, CA, USA: IEEE, 2016: 61-68.
- [11] SUAREZ P L, SAPPA A D, VINTIMILLA B X. Infrared image colorization based on a triplet DCGAN architecture[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Honolulu, HI, USA: IEEE, 2017: 212-217.
- [12] LIU M. Research on high dynamic-range image fusion and visualization method for infrared thermal imaging[D]. Beijing: Beijing Institute of Technology, 2019.
- [13] KUANG X, ZHU J, SUI X, et al. Thermal infrared colorization via conditional generative adversarial network[J]. *Infrared Physics & Technology*, 2020, 107: 103338.
- [14] LI S, HAN B, YU Z, et al. I2V-GAN: Unpaired infrared-to-visible video translation[C]//*Proceedings of the 29th ACM International Conference on Multimedia*. [S. l.]: ACM, 2021: 3061-3069.
- [15] HU X, ZHOU X, HUANG Q, et al. QS-attn: query-selected attention for contrastive learning in I2I translation[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, 2022: 18270-18279.
- [16] YANG S, SUN M, LOU X, et al. An unpaired thermal infrared image translation method using GMA-CycleGAN[J]. *Remote Sensing*, 2023, 15(3): 663.
- [17] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE, 2018: 586-595.
- [18] KUANG X, SUI X, LIU Y, et al. Single infrared image enhancement using a deep convolutional neural network[J]. *Neurocomputing*, 2019, 332: 119-128.
- [19] TALEBI H, MILANFAR P. NIMA: neural image assessment[J]. *IEEE Transactions on Image Processing*, 2018, 27(8): 3998-4011.
- [20] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-03)[2024-10-30]. <https://arxiv.org/abs/1804.02767?context=cs.LG.html>.
- [21] ZHANG L, RAO A, AGRAWALA M. Adding conditional control to text-to-image diffusion models[C]//*2023 IEEE/CVF international conference on computer vision (ICCV)*. New York: IEEE, 2023: 05543.
- [22] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2022.
- [23] FLIR. FLIR thermal dataset for algorithm training [EB/OL]. [2024-10-30]. <https://www.azosensors.com/news.aspx?newsID=12478>.
- [24] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “completely blind” image quality analyzer[J]. *IEEE Signal Processing Letters*, 2013, 20(3): 209-212.
- [25] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[EB/OL]. (2017-06-26)[2024-10-30]. <https://arxiv.org/abs/1706.08500>.
- [26] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-

-
- image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2242-2251.
- [27] HU X, ZHOU X, HUANG Q, et al. QS-attn: query-selected attention for contrastive learning in I2I translation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022:01775.