arXiv:2601.12697v1 [cs.CV] 19 Jan 2026

3D融合：基于3D红外-可见光场景表示的自由视点融合渲染

*Chao Yang**、*Deshui Miao**、*Chao Tian*、*Guoqing Zhu*、*Yameng Gu* 和 *Zhenyu He*****

哈尔滨工业大学（深圳）计算机科学与技术学院

摘要

红外-可见光图像融合旨在将红外信息和可见光信息整合到一幅融合图像中。现有二维融合方法侧重于融合来自固定相机视点的图像，忽视了对复杂场景的全面理解，因而导致场景关键信息丢失。为解决这一局限性，我们提出了一种新颖的红外-可见光高斯融合（IVGF）框架，该框架从多模态二维输入中重建场景几何结构，并支持直接渲染融合图像。具体而言，我们提出了一个跨模态调整（CMA）模块，通过调节高斯的不透明度来解决跨模态冲突问题。此外，为保留两种模态各自的独特特征，我们引入了一种融合损失，用于指导CMA的优化，从而确保融合图像保留每种模态的关键特征。全面的定性和定量实验表明，所提出的方法行之有效。

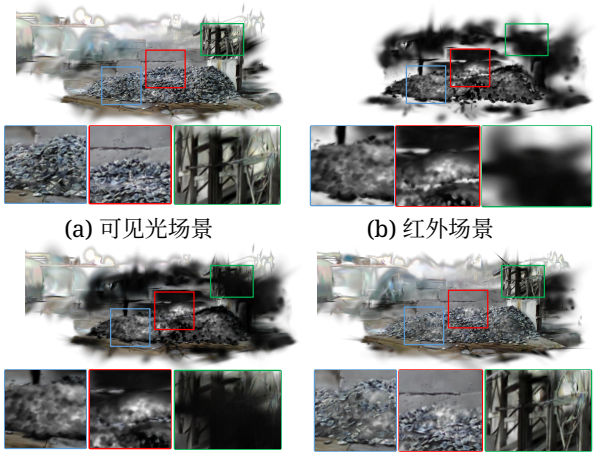
索引词— 3D高斯泼溅，红外-可见光融合，多模态表示

1. 引言

红外-可见光图像融合将红外图像中的热辐射信息与可见光图像中的纹理细节相结合，利用两者在不同光照条件下的互补特性，生成全面的融合表示。融合结果显著有益于多种计算机视觉任务，包括目标检测 [1, 2], 跟踪 [3, 4], 和语义分割 [5, 6]。然而，传统的二维图像融合方法仅限于从固定相机视角进行像素级整合，因此在场景理解任务中的适用性受到限制。

如今，多模态场景表示带来了一种新范式，使得通过多视角二维输入实现自由视点渲染成为可能，从而更全面地表示场景信息。ThermalGaussian [7]提出了一种基于3DGS的 [8]红外和可见光模态场景表示。MMOne [9]提出了一种基于高斯的多模态解耦表示

将高斯拆分为模态专属特征和共享特征，以分别捕捉不同的模态特性。MS-Splatting [10]提出了一种多光谱视图合成框架，以统一的神经颜色表示取代传统的球谐函数。然而，当前的多模态场景表示方法仅限于单独渲染各模态，缺乏渲染融合图像的能力。如图1(c)所示，多模态3D高斯泼溅中的模态冲突常会导致直接渲染时跨模态误差累积，从而造成关键信息的丢失



(a) 可见光场景 (b) 红外场景 (c) 多模态场景 (d) 融合场景

图 1。3D高斯场景表示。(a) 可见光3D场景。(b) 红外3D场景。(c) 多模态3D场景。(d) 我们的红外-可见光高斯融合（IVGF）场景。直接渲染多模态场景会引入累积误差，导致细节和红外信息丢失。

来自两种模态的信息。一种替代方法采用先渲染后融合流程，先分别渲染每种模态，再将它们融合，但这种方法会带来较高的计算开销。

为克服这些局限，我们提出了红外-可见光高斯融合（IVGF），这是一种新颖的三维多模态融合框架，可从任意视点直接渲染融合图像。具体而言，为解决渲染过程中的模态冲突，我们引入了跨模态调整（CMA）模块，根据多模态高斯的颜色属性调节其不透明度参数。此外，为确保融合图像保留两种模态的显著特征，我们设计了一种专用融合损失，以保留模态特定特征，从而提升整体渲染质量。如图1(d)（蓝色/红色/绿色方框）所示，IVGF 能有效保留红外辐射特性和可见光结构细节，从而实现全面且信息丰富的场景表征。

我们的主要贡献总结如下：

- 我们提出了红外-可见光高斯融合（IVGF），这是首个支持对融合图像进行自由视点渲染的多模态融合表征框架。
- 我们引入了跨模态调整（CMA）模块，通过调整高斯不透明度参数，解决多模态渲染过程中的模态冲突。
- 我们提出一种融合损失，可保留关键的模态

* 同等贡献 ** 通讯作者电子邮件：{20b951014, 22b951002, tianchao, 20b951002, 23b951001}@stu.hit.edu.cn, zhenyuhe@hit.edu.cn。

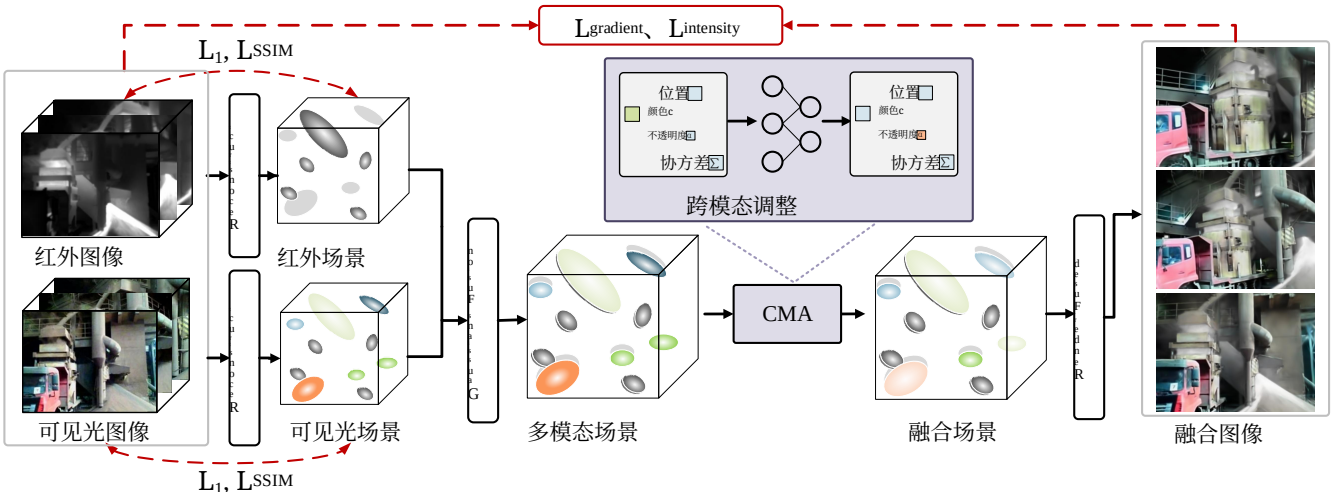


图 2。我们的红外-可见光高斯融合（IVGF）框架概览。首先，分别训练红外高斯和可见光高斯。随后，将二者串联并输入跨模态调整（CMA）模块，该模块会优化高斯参数，以重建融合后的三维场景。

特定特征，从而实现高保真渲染。

2. 方法

传统的二维图像融合方法受限于固定的相机视点，因此在下游任务中的适用性有限。为克服这一局限，我们提出了一种新颖的框架，用于重建融合后的红外-可见光3D场景表示，从而能够从任意视点渲染融合图像。

2.1. 预备知识：3D高斯泼溅

3D高斯泼溅（3DGS）框架 [8], 为辐射场提供了一种高效的可微表示。场景表示为一组 $\mathcal{G} = \{g_i\}_{i=1}^{N}$ 各向异性的三维高斯，其中每个高斯基元 g_i 表示场景中的一个点，并由若干关键属性参数化。均值位置 $\boldsymbol{\mu}_i \in \mathbb{R}^3$ 定义了每个高斯在三维空间中的位置，而协方差矩阵 $\boldsymbol{\Sigma}_i = \mathbf{R}_i\mathbf{S}_i\mathbf{S}_i^\top\mathbf{R}_i^\top \in \mathbb{R}^{3\times3}$通过缩放矩阵 $\mathbf{S}_i$ 控制其各向异性缩放，并通过旋转矩阵$\mathbf{R}_i$控制其朝向。每个高斯还具有一个不透明度值 $\alpha_i \in [0,1]$ ，用于确定其对最终渲染图像的贡献；此外，它还具有视点相关的颜色属性$\mathbf{c}_i$ ，这些属性由球谐系数编码。

2.2. 红外-可见光高斯表示

由于存在模态冲突，使用多模态3D高斯泼溅直接渲染融合图像仍然具有挑战性。如图3(c)所示，渲染结果往往会出现显著的信息损失，关键结构细节被遮蔽，重要的红外特征也未能得到充分保留。为应对这一挑战，我们提出了一种基于3DGS的融合场景表示新框架，如图2所示。该流程首先构建双模态场景表示，然后优化跨模态调整（CMA）模块，以调整不透明度参数，解决渲染过程中的模态冲突。坐标 $\mathbf{x} \in \mathbb{R}^2$ 处的多模态融合渲染 $F(\mathbf{x})$ 计算如下：

$$F(\mathbf{x}) = \sum_{k=1}^{N+M} c_k\,(\tau_k\alpha_k)\prod_{j=1}^{k-1}(1-\tau_j\alpha_j), \tag{1}$$

其中， N 和 M 表示对像素有贡献的可见光和红外高斯， c_k 表示第 k个高斯的颜色， α_k 表示其不透明度， τ_k 表示由跨模态调整（CMA）模块预测的跨模态注意力权重。

CMA 通过学习到的模态感知多层感知机机制调整红外和可见光高斯的不透明度。具体而言，将两个高斯的球谐系数 c_{inf} 和 c_{vis} 拼接后输入多层感知机，以预测 τ:

$$\tau = \text{MLP}(\text{Concat}(c_{\text{inf}}, c_{\text{vis}})), \tag{2}$$

其中， $\tau \in [0,1]$ 动态缩放每个高斯的不透明度参数， $c \in \mathbb{R}^{(N+M)\times d_c}$，其中 d_c 表示每个高斯的球谐系数维度。多层感知机的计算如下:

$$\text{MLP}(\mathbf{C}) = \sigma\Big(\mathbf{W}_3^\top \cdot \phi_2\big(\phi_1(\mathbf{C})\big)\Big), \tag{3}$$

$$\phi_i(\mathbf{x}) = \text{LayerNorm}\big(\text{LeakyReLU}(\mathbf{x}\mathbf{W}_i + \mathbf{b}_i)\big) \quad (i=1,2), \tag{4}$$

其中， $\mathbf{W}$ 为权重矩阵， σ 为输出激活函数， ϕ 包含 LeakyReLU 激活和层归一化。

2.3. 损失函数

我们针对红外-可见光高斯融合（IVGF）采用两阶段训练策略。在第一阶段，使用 ThermalGaussian [7] 的相同设置训练两个模态的三维表示。在第二阶段，将两个高斯拼接起来，并训练跨模态调整（CMA）模块，以获得融合高斯表示。

训练阶段 1。 按照 ThermalGaussian [7], , 分别训练两种模态的高斯表示。损失函数定义如下:

$$\mathcal{L}_{\text{stage 1}} = \gamma\,\mathcal{L}_{\text{Visible}} + (1-\gamma)\,\mathcal{L}_{\text{Infrared}}, \tag{5}$$

$$\mathcal{L}_{\text{Visible}} = \|V - V_{\text{render}}\|_1 + \big(1 - \text{SSIM}(V, V_{\text{render}})\big), \tag{6}$$

$$\mathcal{L}_{\text{Infrared}} = \|T - T_{\text{render}}\|_1 + \big(1 - \text{SSIM}(T, T_{\text{render}})\big). \tag{7}$$

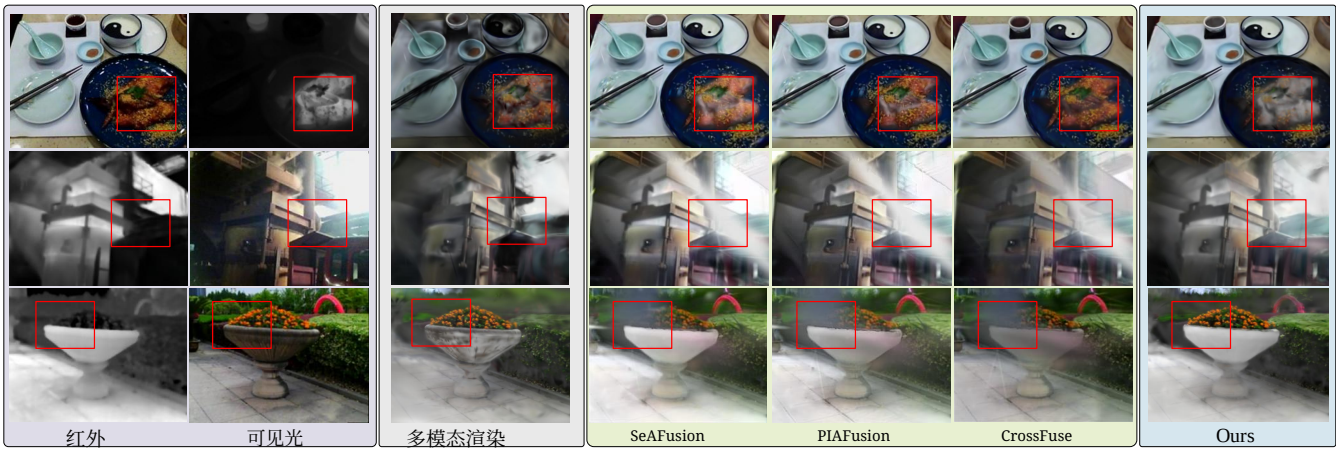


图 3。不同方法的定性结果。第一个方框包含红外与可见光图像，第二个方框展示多模态渲染结果，第三个方框呈现先渲染后融合方法的融合结果，最后一个方框展示我们的方法结果。

其中，V 表示可见光图像，T 表示热红外图像，$\|\cdot\|_1$ 表示 ℓ_1 距离，SSIM$(\cdot,\cdot)$表示结构相似性指数，系数 $\gamma = N_{\text{Infrared}}/(N_{\text{Infrared}}+N_{\text{Visible}})$用于平衡可见光和红外模态的贡献，其中 N_{Infrared} 和 N_{Visible} 分别表示相应模态中的高斯数量。

训练阶段二。为了优化能够保留两种模态显著信息的多模态融合场景，我们提出了一种新的融合损失函数，用于指导跨模态调整（CMA）模块的学习。设 I_{fuse} 表示由三维融合场景渲染得到的融合图像。阶段二的跨模态融合损失函数定义如下：

$$\mathcal{L}_{\text{stage 2}} = \mathcal{L}_{\text{intensity}} + \mathcal{L}_{\text{gradient}}, \tag{8}$$

其中，$\mathcal{L}_{\text{intensity}}$ 结合了像素级相似性和结构相似性度量。

$$\begin{aligned}\mathcal{L}_{\text{intensity}} &= \lambda_1 \|I_{\text{fuse}} - \max(V,T)\|_1 \\ &+ \lambda_2\big(1-\text{SSIM}(I_{\text{fuse}},T) + (1-\text{SSIM}(I_{\text{fuse}},V)\big),\end{aligned} \tag{9}$$

而 $\mathcal{L}_{\text{gradient}}$ 则促进两种模态之间的边缘保留，其定义如下：

$$\mathcal{L}_{\text{gradient}} = \|\nabla I_{\text{fuse}} - \nabla \max(V,T)\|_1, \tag{10}$$

其中，∇表示空间梯度算子。权重参数 λ_1 和 λ_2 用于平衡强度保持与结构维持。所提出的损失函数有助于保留最重要的信息，从而增强多模态融合场景的表征能力。

3. 实验

3.1. 实验设置

数据集。RGBT-Scenes 数据集 [7] 包含 1,000 多对对齐的可见光图像与热红外图像。该数据集涵盖 10 种不同的场景类别，包括室内外环境。

评估指标。遵循二维融合性能的评估方法，我们首先计算融合图像与两幅源图像之间的指标，然后对这些指标取平均，以评估整体性能。所用的评估指标包括：峰值信噪比（PSNR）、结构相似性指数（SSIM）和学习感知图像块相似度（LPIPS）。

实现细节。我们的 IVGF 框架基于 Thermal-Gaussian [7], 构建，并采用基模型的所有实验参数。训练过程在 RTX 3090 Ti GPU 上进行，共 30k 次迭代（先进行 15k 次单独重建阶段迭代，再进行 15k 次 CMA 优化迭代），生成的融合图像分辨率为 640×480。对于公式 9 中的损失函数，我们将权重设为 $\lambda_1 = 1$ 和 $\lambda_2 = 2$。

对比方法。我们采用先渲染后融合流程生成融合图像，以进行评估和比较。具体而言，ThermalGaussian [7] 首先渲染成对的红外与可见光图像，然后将这两张渲染图像输入最先进的融合方法，以获得融合结果，包括 CDDFuse [11],CrossFuse [12], EMMA [13], LRRNet [14], LiMFusion [15], PIA-Fusion [16], SHIP [17], SeAFusion [18], 和 SwinFusion [19]。

3.2. 融合方法的定性分析

图 3 对比了我们的方法、直接的多模态 3DGS 渲染结果以及其他二维融合方法。多模态 3DGS 结果的渲染会导致误差累积，从而造成图像模糊并丢失关键信息。对于二维融合方法（先渲染后融合策略），红色边界框内的区域往往会丢失关键热信息，并出现模糊和光照突变。相比之下，我们的方法有效保留了红外模态中的热信息和可见光模态中的纹理细节，避免了这些伪影。这些对比表明，我们提出的 IVGF 能够有效保留并整合两种模态的信息，从而实现更全面的多模态融合表征。

3.3. 融合方法的定量分析

表 1 中的定量比较显示，IVGF 在几乎所有场景中的 SSIM 和学习感知图像块相似度得分均为最高，优于九种现有融合方法（红色表示最佳表现，蓝色表示次佳表现）。结果表明，我们提出的 IVGF 能够有效重建融合后的多模态场景，并且相比二维融合方法（2D 先渲染后融合策略）生成质量更高的融合结果。

表 1. 与二维融合方法的定量性能比较（最佳结果以红色标示，次佳结果以蓝色标示）。↑箭头表示数值越高越好，而↓箭头表示数值越低越好。

	方法	日常物品 杂物	点心	玻璃 Cup	Iron 锭	Land 景观	花坛	植物 设备	Road 街区	旋转式 Kiln	卡车
学习感知图像块相似度 ↓	CDDFuse (CVPR 2023)	0.462	0.392	0.384	0.420	0.505	0.429	0.380	0.399	0.359	0.387
	CrossFuse (IF 2024)	0.452	0.391	0.409	0.408	0.495	0.408	0.368	0.373	0.362	0.381
	EMMA (CVPR 2024)	0.475	0.406	0.414	0.428	0.535	0.440	0.389	0.412	0.405	0.392
	LRRNet (TPAMI 2023)	0.448	0.380	0.375	0.422	0.512	0.417	0.368	0.391	0.360	0.370
	LiMFusion (OLEN 2024)	0.494	0.428	0.456	0.427	0.528	0.441	0.396	0.424	0.401	0.407
	PIAFusion (IF 2022)	0.453	0.393	0.381	0.416	0.515	0.426	0.387	0.378	0.391	0.388
	SHIP (CVPR 2024)	0.525	0.445	0.398	0.463	0.604	0.522	0.483	0.460	0.517	0.456
	SeAFusion (IF 2022)	0.452	0.389	0.380	0.415	0.521	0.423	0.378	0.384	0.378	0.381
	SwinFusion (JAS 2022)	0.459	0.385	0.376	0.418	0.517	0.428	0.377	0.396	0.367	0.382
	Ours	0.425	0.378	0.371	0.413	0.481	0.395	0.370	0.366	0.369	0.372
峰值信噪比 ↑	CDDFuse (CVPR 2023)	13.275	12.495	12.861	13.786	13.966	15.112	14.059	15.105	13.755	12.340
	CrossFuse (IF 2024)	14.260	14.018	12.711	15.134	15.056	16.531	15.549	15.945	15.168	14.278
	EMMA (CVPR 2024)	13.519	12.438	12.718	14.042	14.395	14.832	14.506	14.972	13.916	12.888
	LRRNet (TPAMI 2023)	14.469	13.590	14.197	14.597	14.630	15.814	15.624	14.765	14.839	14.174
	LiMFusion (OLEN 2024)	12.997	12.536	12.877	14.050	14.055	15.432	14.303	14.547	12.895	12.472
	PIAFusion（IF 2022）	13.947	13.507	13.918	14.771	13.756	15.659	14.407	15.084	12.027	12.893
	SHIP（CVPR 2024）	9.073	9.608	12.569	10.412	6.116	9.397	8.878	9.072	5.859	9.485
	SeAFusion（IF 2022）	13.127	12.812	13.473	13.878	13.706	14.677	13.826	14.049	12.852	12.443
	SwinFusion（JAS 2022）	12.976	12.573	13.205	13.944	13.702	14.320	13.701	14.857	12.368	12.386
	Ours	13.929	14.738	12.968	14.553	15.579	16.646	15.047	15.620	12.592	13.152
SSIM ↑	CDDFuse（CVPR 2023）	0.614	0.541	0.577	0.490	0.572	0.696	0.657	0.808	0.575	0.573
	CrossFuse (IF 2024)	0.613	0.544	0.570	0.490	0.558	0.707	0.661	0.831	0.552	0.569
	EMMA (CVPR 2024)	0.602	0.545	0.575	0.489	0.548	0.685	0.652	0.805	0.559	0.590
	LRRNet (TPAMI 2023)	0.630	0.540	0.592	0.484	0.552	0.682	0.648	0.802	0.553	0.546
	LiMFusion (OLEN 2024)	0.601	0.535	0.570	0.484	0.540	0.681	0.648	0.792	0.561	0.559
	PIAFusion（IF 2022）	0.615	0.548	0.593	0.489	0.531	0.695	0.647	0.812	0.557	0.573
	SHIP（CVPR 2024）	0.526	0.484	0.568	0.442	0.428	0.591	0.531	0.729	0.413	0.491
	SeAFusion（IF 2022）	0.619	0.555	0.596	0.495	0.559	0.703	0.662	0.812	0.582	0.589
	SwinFusion（JAS 2022）	0.616	0.553	0.591	0.494	0.559	0.692	0.661	0.807	0.586	0.586
	Ours	0.665	0.567	0.613	0.495	0.610	0.734	0.680	0.839	0.618	0.603



单阶段训练 + CMA +CMA + L_{stage2}

图 4。不同模块效果的消融研究。跨模态调整（CMA）模块可解决跨模态误差累积问题，而所提出的融合损失则确保渲染结果保留两种模态的重要特征。

3.4 消融研究

与不同 3D 融合策略的比较。如图 3 和表 2.1 所示，我们的方法在视觉质量、存储效率和渲染速度之间实现了更优的平衡。具体而言，IVGF 在所有性能指标上均优于多模态渲染，同时内存占用和速度与之相当。与 2D 先渲染后融合策略相比，IVGF 性能更优，峰值信噪比和 SSIM 更高、学习感知图像块相似度更低，速度更快（182 帧每秒），内存占用也更少（48 MB）。这些结果表明，IVGF 为 3D 场景表示确立了一种有效且高效的新范式。

不同模块的效果。如图 4 和表 2.2 所示，消融研究表明，跨模态调整（CMA）模块能够增强结构细节，从而提升视觉

表 2。我们方法的消融研究。（最佳结果以红色标示，次佳结果以蓝色标示。）↑箭头表示数值越高越好，而↓箭头表示数值越低越好。

	方法	学习感知图像块相似度 ↓	峰值信噪比 ↑	SSIM ↑	内存 ↓	帧每秒 ↑
1	多模态渲染	0.408	12.426	0.515	20 MB	631
	二维融合	0.389	12.812	0.555	69 MB	239
2	单阶段训练	0.503	10.735	0.455	376 MB	138
	+ CMA	0.371	14.660	0.567	23 MB	387
	+ CMA + $L_{stage\ 2}$	0.378	14.738	0.567	21 MB	412

质量、评估指标、处理速度和内存效率。这一改进源于 CMA 模块能够有效解决多模态高斯之间的冲突，因此性能优于单阶段训练策略。此外，采用所提出的融合损失后，视觉结果能够更好地保留两种模态中的显著信息，而定量评估也证实了各项指标均得到持续改善。这表明，我们提出的损失函数能够在融合多模态场景表示中有效保留各模态的关键特征。

4. 结论

本文提出红外-可见光高斯融合（IVGF），这是首个通过多模态融合三维表示实现融合图像自由视点渲染的框架。为此，我们设计了跨模态调整（CMA）模块，通过动态调节高斯不透明度来解决模态冲突，并提出一种新颖的融合损失，以保留两种模态中的显著信息。大量实验验证了我们方法的有效性。

5. 参考文献

[1] Hongxu Yin、Arash Vahdat、Jose M Alvarez、Arun Mallya、Jan Kautz 和 Pavlo Molchanov，“A-vit：用于高效视觉变换器的自适应令牌”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2022年，第10809–10818页。[2] Chao Tian、Zikun Zhou、Yuqing Huang、Gaojun Li 和 Zhenyu He，“面向未配准RGB-热红外行人检测的跨模态提案引导特征挖掘”，*IEEE*多媒体汇刊，第26卷，第6449–6461页，2024年。[3] Seokju Cho、Jiahui Huang、Seungryong Kim 和 Joon-Young Lee，“Flowtrack：重新审视用于长程密集跟踪的光流”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2024年，第19268–19277页。[4] Xiao Wang、Xiujun Shu、Shiliang Zhang、Bo Jiang、Yaowei Wang、Yonghong Tian 和 Feng Wu，“Mfgnet：用于RGB-T跟踪的动态模态感知滤波器生成”，*IEEE*多媒体汇刊，第25卷，第4335–4348页，2022年。[5] Alexander Kirillov、Eric Mintun、Nikhila Ravi、Hanzi Mao、Chloe Rolland、Laura Gustafson、Tete Xiao、Spencer Whitehead、Alexander C Berg、Wan-Yen Lo 等，“分割任何事物”，载于*IEEE/CVF*国际计算机视觉会议论文集，2023年，第4015–4026页。[6] Ying Lv、Zhi Liu 和 Gongyang Li，“用于RGB-T语义分割的上下文感知交互网络”，*IEEE*多媒体汇刊，第26卷，第6348–6360页，2024年。[7] Rongfeng Lu、Hangyu Chen、Zunjie Zhu、Yuhang Qin、Ming Lu、Le Zhang、Chenggang Yan 和 Anke Xue，“Thermal-gaussian：热红外三维高斯泼溅”，*arXiv*预印本 *arXiv:2409.07200*，2024年。[8] Bernhard Kerbl、Georgios Kopanas、Thomas Leimk¨uhler 和 George Drettakis，“用于实时辐射场渲染的3D高斯泼溅。”，*ACM*图形学汇刊，第42卷，第4期，第139–1页，2023年。[9] Zhifeng Gu 和 Bing Wang，“Mmone：在一个场景中表示多种模态”，*arXiv*预印本 *arXiv:2507 11129*，2025。。

[10] LukasMeyer, JosefGr¨un, MaximilianWeiherer, BernhardEgger, Marc Stamminger 和 Linus Franke，《采用神经颜色表示的多光谱高斯泼溅》，*arXiv* 预印本 *arXiv:2506.03407*，2025 年。

[11] Zixiang Zhao, Haowen Bai, Bingxin Ke, Yukun Cui, Lilun Deng, Yulun Zhang, Kai Zhang, Konrad Schindler, Radu Timofte 和 Luc Van Gool，《用于多模态图像融合的相关性驱动双分支特征分解》，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，2023 年 6 月，第 5906–5916 页。

[12] Hui Li 和 Xiao-Jun Wu，《CrossFuse：一种基于新型交叉注意力机制的红外与可见光图像融合方法》，*InformationFusion*，第 103 卷，第 102147 页，2024 年。

[13] ZixiangZhao、Haowen Bai、Jiangshe Zhang、Yulun Zhang、Kai Zhang、Shuang Xu、Dongdong Chen、Radu Timofte 和 Luc Van Gool，《等变多模态图像融合》，发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，2024年，第25912–25921页。

[14] Hui Li、Tianyang Xu 等，《LRRNet：一种由表示学习引导的红外与可见光图像融合网络》，*IEEE*模式分析与机器智能汇刊，第45卷，第9期，第11040–11052页，2023年。[15] Yao Qian、Haojie Tang、Gang Liu、Mengliang Xing、Gang Xiao 和 Durga Prasad Bavirisetti，《LiMFusion：基于局部信息测量的红外与可见光图像融合》，光学与激光工程，第181卷，第108435页，2024年。[16] Linfeng Tang、Jiteng Yuan 等，《PIAFusion：一种基于光照感知的渐进式红外与可见光图像融合网络》，*InformationFusion*，第83卷，第79–92页，2022年。[17] Naishan Zheng、Man Zhou、Jie Huang、Junming Hou、Haoying Li、Yuan Xu 和 Feng Zhao，《探究红外与可见光图像融合中的协同高阶交互》，发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，2024年，第26384–26395页。[18] Linfeng Tang、Jiteng Yuan 和 Jiayi Ma，《融入高层视觉任务的图像融合：一种语义感知的实时红外与可见光图像融合网络》，*InformationFusion*，第82卷，第28–42页，2022年。[19] Jiayi Ma、Linfeng Tang、Fan Fan、Jun Huang、Xiaoguang Mei 和 Yong Ma，《SwinFusion：基于Swin Transformer的通用图像融合跨域长程学习》，*IEEE/CAA*自动化学报，第9卷，1217，2022年7月，第1200页。