Early vs Late Fusion Strategies in LIVO for 6-DOF Pose Estimation Under High Parallax

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How do different fusion strategies (early vs. late fusion) in LIVO frameworks affect the alignment and consistency of 6-DOF pose estimates in high-parallax scenarios, as benchmarked on the TartanAir. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: SGDFuse: SAM-Guided Diffusion Model for High-Fidelity Infrared and Visible Image Fusion. Research question: How do different fusion strategies (early vs. late fusion) in LIVO frameworks affect the alignment and consistency of 6-DOF pose estimates in high-parallax scenarios, as benchmarked on the TartanAir dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.4/10.

3 Results

4 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.4/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The MSRS dataset provides 361 pairs of test images with a resolution of 640$\times$480.	×	0.01
The M3FD dataset contains 4,164 image pairs with a resolution of 1024$\times$768.	×	0.01
The LLVIP dataset includes 16,836 image pairs with a resolution of 1280$\times$1024.	×	0.01
The RoadScene dataset consists of 221 registered pairs of infrared and visible images.	×	0.03
The proposed model is trained on the MSRS dataset, comprising 1,083 visible-infrared image pairs for training and 361 pa	×	0.04
The fusion network is trained for 200 epochs using the Adam optimizer with a learning rate of 1e-4 and batch size of 24.	×	0.03
The model processes input images at their original resolution, outputting a three-channel color fused image suitable for	×	0.02
The entire model is implemented using the PyTorch framework, and all experiments are conducted on a workstation equipped	×	0.03
The proposed model achieves one of the best infrared and visible registration results.	×	0.05
Li et al. [22] proposed a novel multimodal medical image fusion framework that leverages semantic-level information to o	×	0.07
Transformer-based methods provide a new fusion paradigm, leveraging long-range dependency modeling for superior cross-do	×	0.04
Li et al. [24] proposed a U-Net-based framework, pioneering the exploration of adversarial robustness in fusion tasks.	×	0.02
Liu et al. [25] pioneered the application of prompt learning to infrared and visible image fusion, achieving state-of-th	×	0.11
FusionGAN [26] uses adversarial training between generator and discriminator.	×	0.00
AttentionFGAN [27] adds multi-scale attention to separately emphasize infrared and visible features.	×	0.03

References

- http://arxiv.org/abs/2508.05264v6
- http://arxiv.org/abs/2302.04024v1
- http://arxiv.org/abs/1003.1598v1