CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion

SUPPLEMENTARY MATERIALS

Zixiang Zhao[1,2] Haowen Bai[1] Jiangshe Zhang[1*] Yulun Zhang[2]
Shuang Xu[3,4] Zudi Lin[5] Radu Timofte[2,6] Luc Van Gool[2]

[1] Xi'an Jiaotong University [2] Computer Vision Lab, ETH Zürich
[3] Research and Development Institute of Northwestern Polytechnical University in Shenzhen
[4] Northwestern Polytechnical University [5] Harvard University [6] University of Würzburg

zixiangzhao@stu.xjtu.edu.cn, jszhang@mail.xjtu.edu.cn

Abstract

In this document, we provide the additional supplementary information for the paper "CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion". This file contains:
(I) The detail architecture for Share Feature Encoder (SFE) and Detail CNN Encoder (DCE) which is mentioned in Sec. 3.2.
(II) Detailed illustration to the training&testing datasets in Sec. 4.1.
(III) More qualitative comparison fusion results in Sec. 4.2.
(IV) Qualitative results for Downstream Infrared-Visible applications in Sec. 4.4.

1. Detail architecture for SFE and DCE

We illustrate the detailed architecture for *Share Feature Encoder* (SFE) module and *Detail CNN Encoder* (DCE) module of CDDFuse framework in Fig. 1.

2. Detailed introduction to datasets

We adopt widely-used benchmarks MSRS [2], RoadScene [4], and TNO [3] for *Infrared-Visible image Fusion* (IVF), MSRS [2] and M^{3}FD [1] for *Multi-Modality Object Detection* (MMOD) and *Multi-Modality Semantic Segmentation* (MMSS), and Harvard Medical Image Dataset for *Medical Image Fusion* (MIF), respectively.

- MSRS dataset[1]: 1083 pairs for IVF/MMSS training and 361 pairs for IVF/MMSS testing.
- RoadScene dataset[2]: 50 pairs for IVF validation and 50 pairs for IVF testing.
- TNO dataset[3]: 50 pairs for IVF testing.
- M^{3}FD dataset[4]: 3360 pairs for MMOD training, 420 pairs for MMOD validation and 420 pairs for MMOD testing.
- Harvard Medical Image dataset dataset[5]: 130 pairs for MIF training, 20 pairs for MIF validation and 136 pairs for MIF testing.

---

*Corresponding author.
[1] https://github.com/Linfeng-Tang/MSRS
[2] https://github.com/hanna-xu/RoadScene
[3] https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029
[4] https://github.com/JinyuanLiu-CV/TarDAL
[5] http://www.med.harvard.edu/AANLIB/home.html

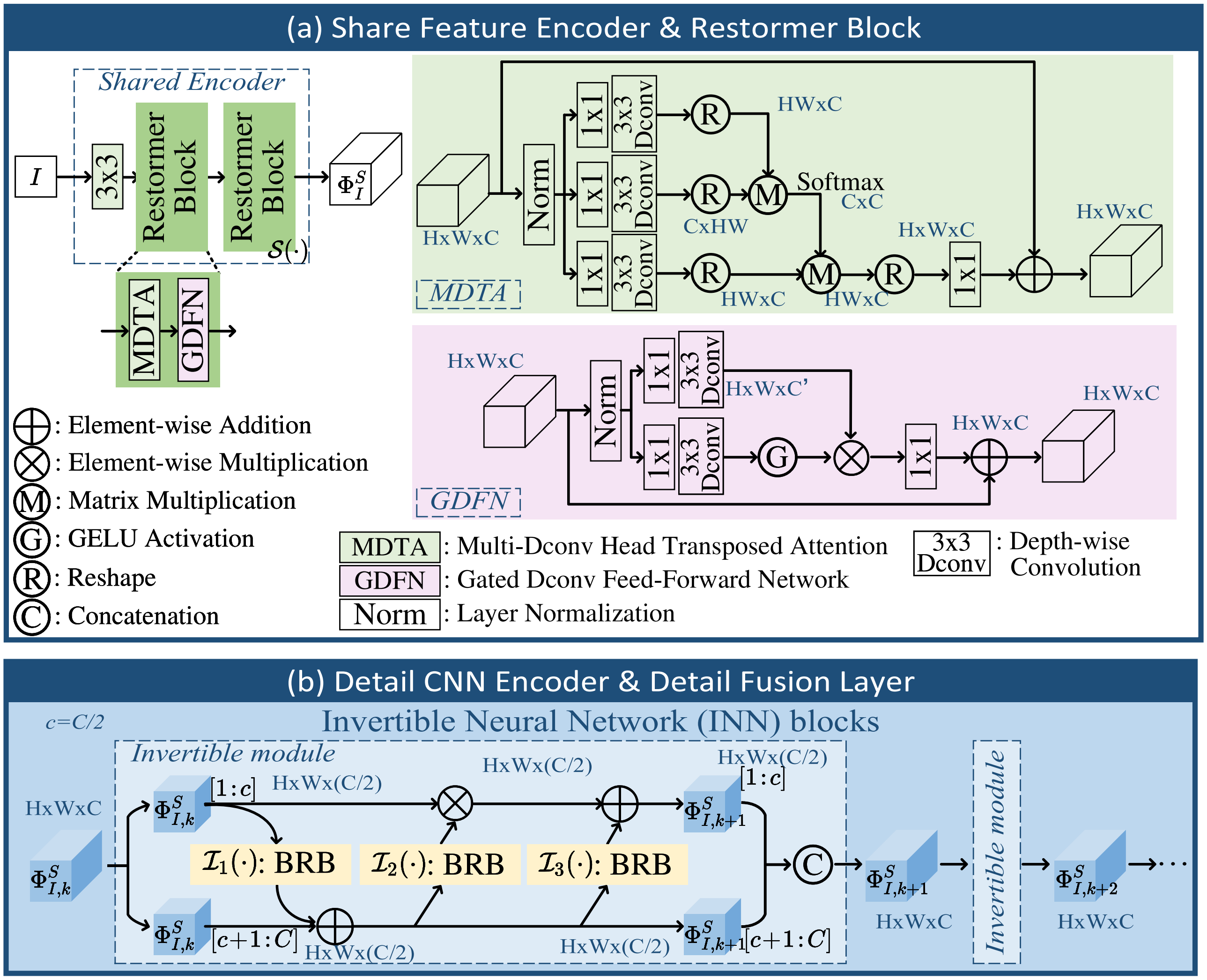

Figure 1. Detail architecture for *Share Feature Encoder* (SFE) module and *Detail CNN Encoder* (DCE) module of CDDFuse.

3. More qualitative comparison fusion results

More qualitative comparisons for *Infrared-Visible image Fusion* results are displayed in Figs. 2 and 3. Our method better integrates thermal radiation information in infrared images and detailed textures in visible images. Objects in dark regions are clearly highlighted, so that foreground targets can be easily distinguished from the background. Additionally, background details that are difficult to identify due to the low illumination have clear edges and abundant contour information, which help us understand the scene better.

More qualitative comparison for *Medical Image Fusion* results are shown in Fig. 4. CDDFuse can better preserve the detailed texture and highlight the structure information than other methods.

4. Qualitative results for Downstream Infrared-Visible applications

The qualitative results for infrared-visible object detection and semantic segmentation are exhibited in Figs. 5 and 6 and Figs. 7 and 8, respectively. In object detection, CDDFuse can improve detection accuracy by fusing thermal radiation information and highlighting the difficult-to-observe targets. Therefore, small objects can be better detected. For the segmentation task, CDDFuse better integrates the edge and contour information in the source images, which enhances the ability of model to perceive the object boundary, and makes the segmentation more accurate.

Figure 2. Visual comparison for *Infrared-Visible image Fusion*.

Figure 3. Visual comparison for *Infrared-Visible image Fusion*.

References

[1] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *CVPR*, pages 5792–5801, 2022. 1

[2] Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. *Inf. Fusion*, 83-84:79–92, 2022. 1

[3] Alexander Toet and Maarten A. Hogervorst. Progress in color night vision. *Optical Engineering*, 51(1):1 – 20, 2012. 1

[4] Han Xu, Jiayi Ma, Zhuliang Le, Junjun Jiang, and Xiaojie Guo. Fusiondn: A unified densely connected network for image fusion. In *AAAI*, pages 12484–12491, 2020. 1

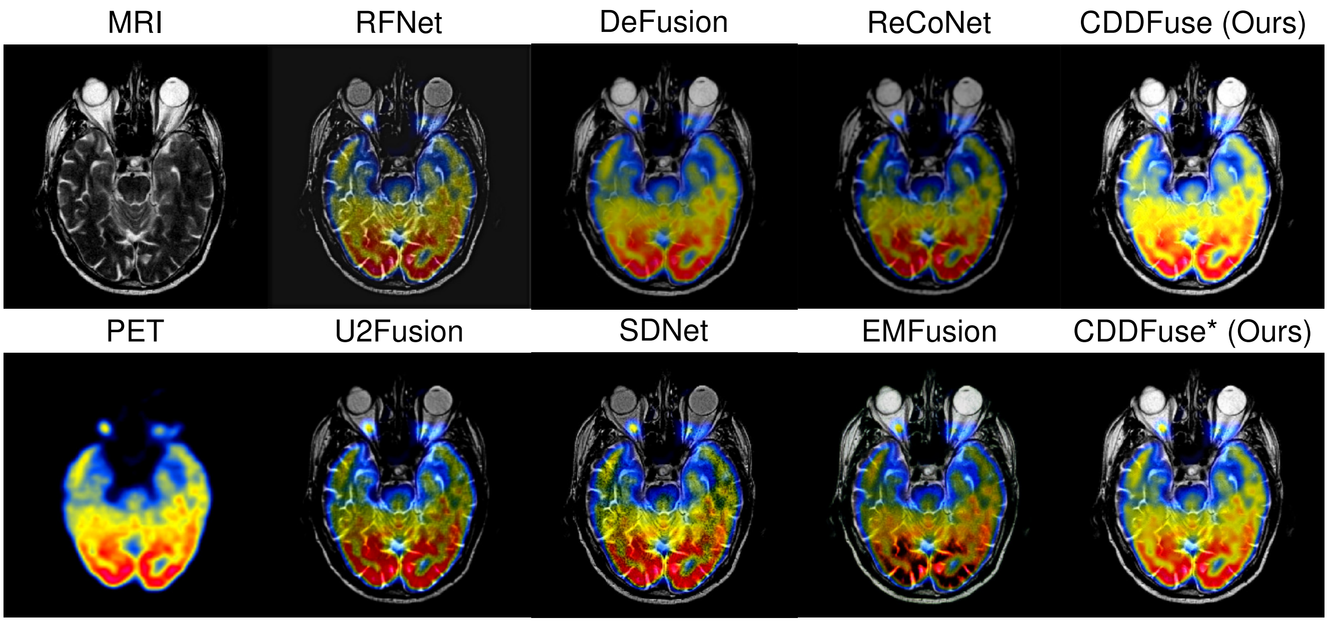

Figure 4. Visual comparison for *Medical Image Fusion*.

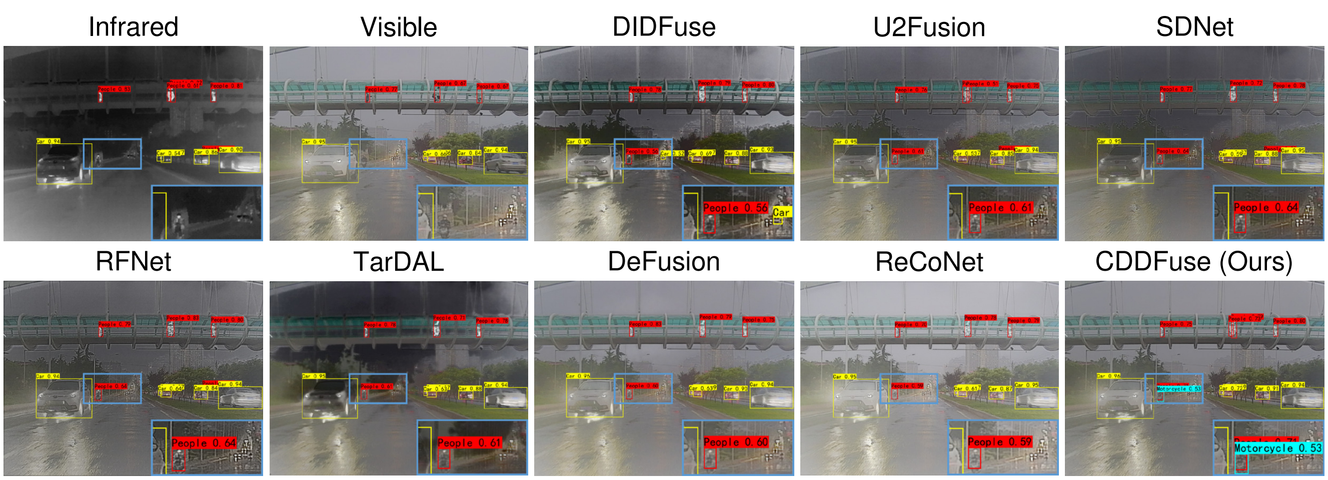

Figure 5. Qualitative results for infrared-visible object detection on M^{3}FD dataset.

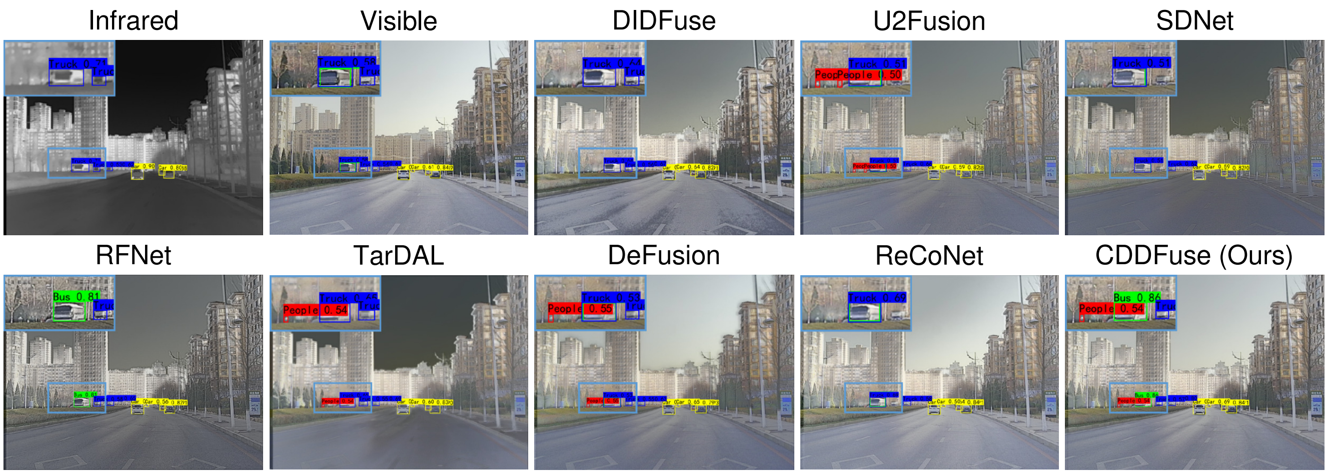

Figure 6. Qualitative results for infrared-visible object detection on M^3FD dataset.

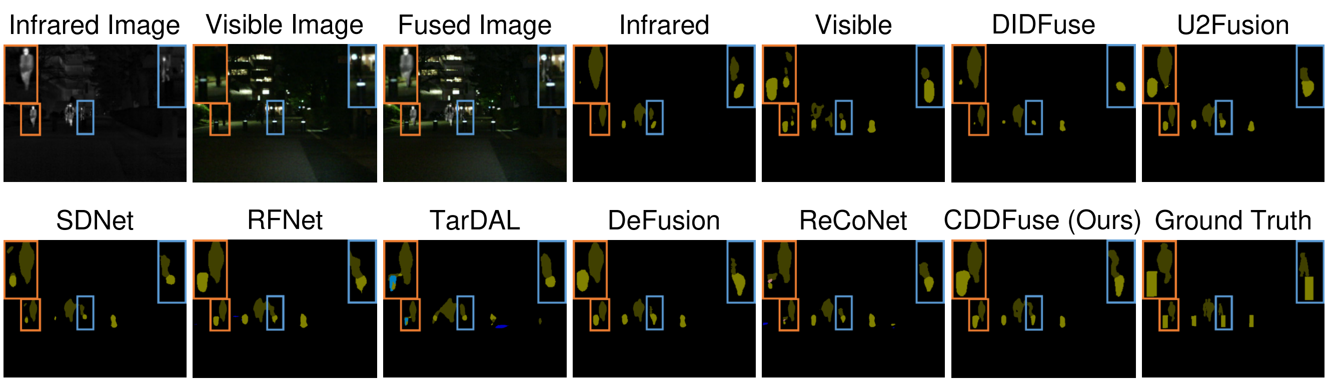

Figure 7. Qualitative results for infrared-visible semantic segmentation on MSRS dataset.

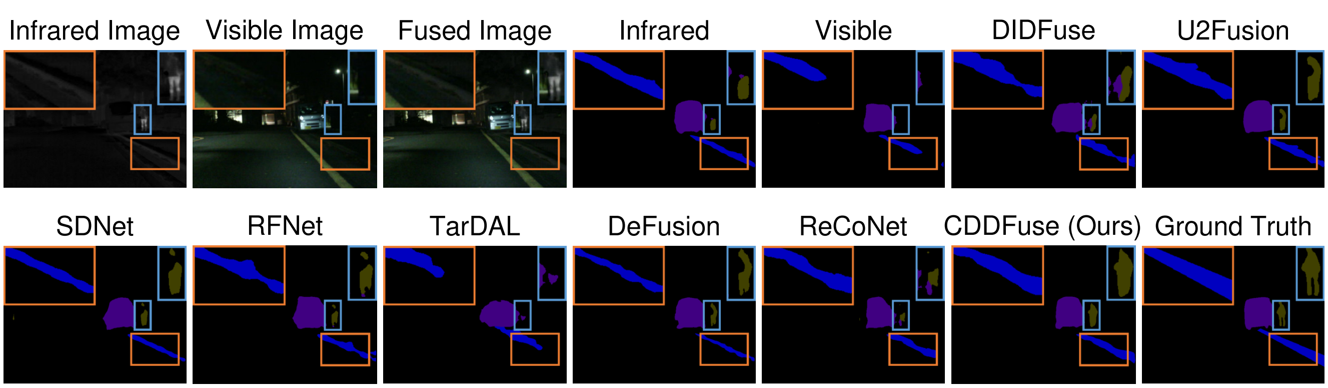

Figure 8. Qualitative results for infrared-visible semantic segmentation on MSRS dataset.