Supplemental Material for Target-aware Dual Adversarial Learning and a Multi-scenario Multi-Modality Benchmark to Fuse Infrared and Visible for Object Detection

Jinyuan Liu[†], Xin Fan[‡,*] Zhanbo Huang[‡], Guanyao Wu[‡], Risheng Liu[‡,§], Wei Zhong[‡], Zhongxuan Luo[†]
[†]School of Software Technology, Dalian University of Technology
[‡]DUT-RU International School of Information Science & Engineering, Dalian University of Technology
[§]Peng Cheng Laboratory
{atlantis918}@hotmail.com, {zbhuang,rollingplain}@mail.dlut.edu.cn, {xin.fan,rsliu}@dlut.edu.cn

Abstract

This document provides supplementary information on the following three aspects. First, we demonstrate more typical examples of our proposed new benchmark, which can fully exhibit the superior diversity scenario of our benchmark. Then, we give more visual comparisons against the state-of-the-arts. Finally, we discuss the limitation of our method and provide some failure cases. The proposed dataset and code will be available at https://github.com/dlut-dimt/TarDAL.

1. More illustration of Multi-scenario Multi-modality Benchmark

To show the diversity of our benchmark (M^3FD) more clearly, we provide more image pairs in Figure 1. Note that our benchmark covers four major scenarios (*i.e.,* daytime, overcast, night and challenge) with various environments, illumination, season, and weather, having a wide range of pixel variations.

2. More Visual Comparisons

We provide more visual comparisons to verify the proposed method's superiority against other state-of-the-art methods on four datasets (three for fusion, two for detection). The visual comparisons on TNO, Roadscene and M^3FD are shown in Figure 2, Figure 3 and Figure 4, respectively. Furthermore, the visual comparisons of realizing object detection on fused images are present in Figure 5 and Figure 6.

3. Limitations and Failure cases

As our network is trained according to the salient target from the infrared image and textural details form the visible image, respectively. However, our method cannot perform well under the condition of when two input images are the slight mis-alignment (see Figure 7).

To discuss the impact of unregistered images on our method, we initially synthesize mis-registered source images through performing random affine and elastic translations with different degrees(*i.e.,* slight, moderate and extreme) on the TNO, Roadscene and M^3FD datasets respectively. Then we use the proposed method for merging these mis-registered source images. Visual results are shown in Figure 7, it can be seen that our method can deal with the slight unregistered images, which preserves a large part of vital information. However, when pixel deviation is large, halos and artifacts emerge on the fused results.

*Corresponding author.

Figure 1. Visualization of infrared-visible images on our M^{3}FD dataset. The dataset covers extensive scenarios with various environments, illumination, season, and weather.

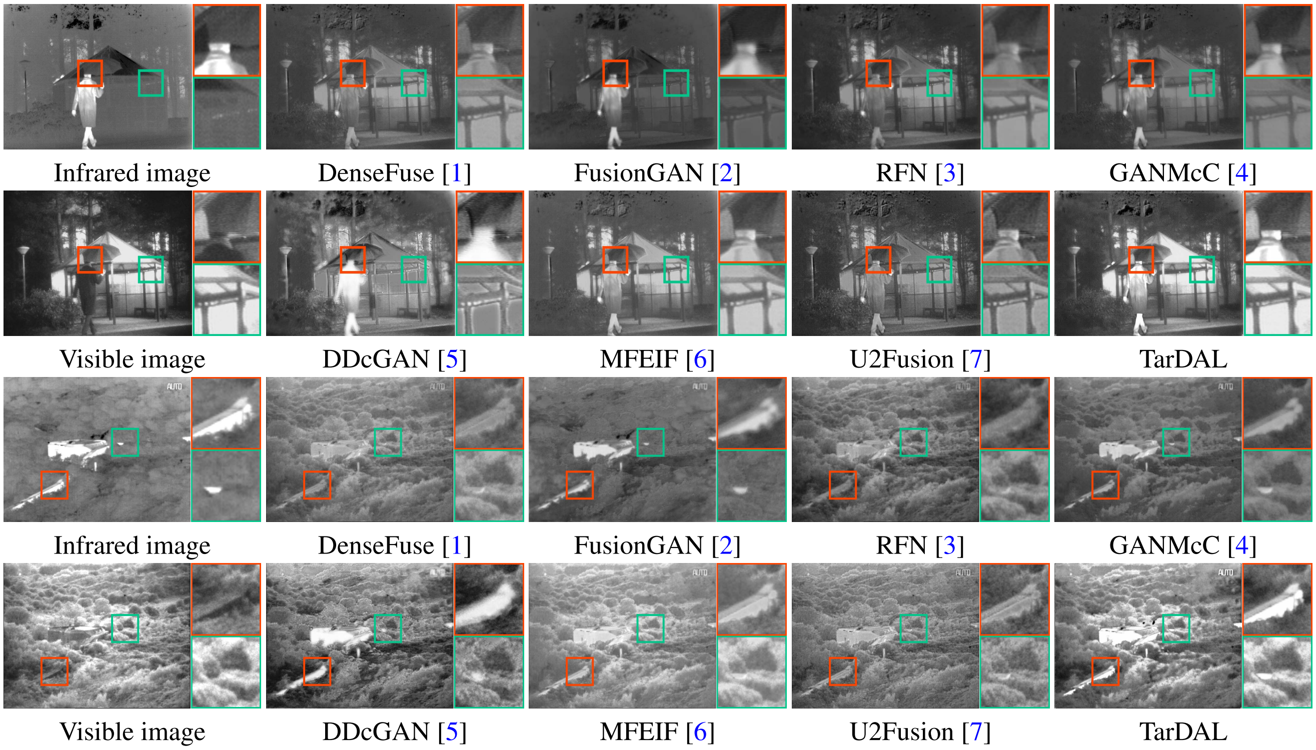

Figure 2. Visual results comparison between different methods on *TNO* Dataset Best viewed on screen.

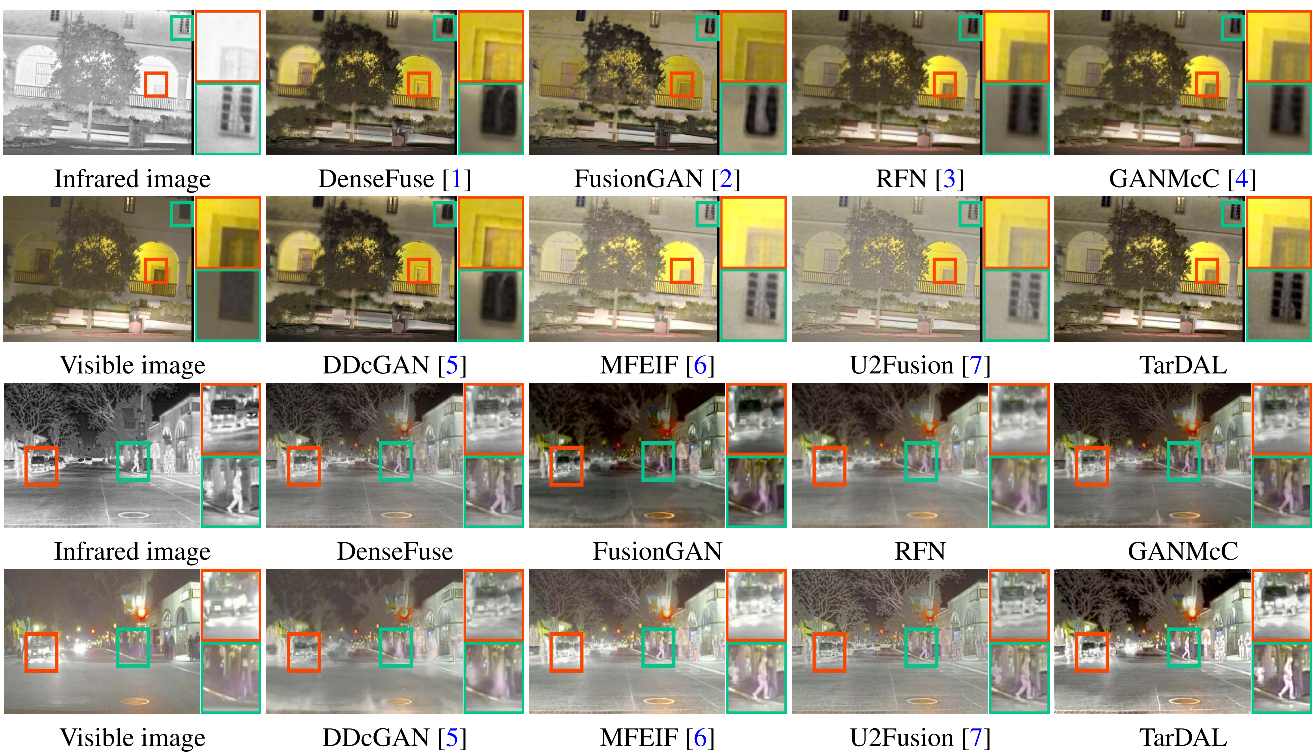

Figure 3. Visual results comparison between different methods on RoadScene dataset. Best viewed on screen.

Figure 4. Visual results comparison between different methods on Multispectral dataset. Best viewed on screen.

References

[1] Hui Li and Xiao-Jun Wu. Densefuse: A fusion approach to infrared and visible images. *IEEE TIP*, 28(5):2614–2623, 2018.

[2] Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. Fusiongan: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48:11–26, 2019.

[3] Hui Li, Xiao-Jun Wu, and Josef Kittler. Rfn-nest: An end-to-end residual fusion network for infrared and visible images. *Information Fusion*, 73:72–86, 2021.

[4] Jiayi Ma, Hao Zhang, Zhenfeng Shao, Pengwei Liang, and Han Xu. Ganmcc: A generative adversarial network with multiclassification constraints for infrared and visible image fusion. *IEEE TIM*, 70:1–14, 2020.

[5] Jiayi Ma, Han Xu, Junjun Jiang, Xiaoguang Mei, and Xiao-Ping Zhang. Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE TIP*, 29:4980–4995, 2020.

[6] Jinyuan Liu, Xin Fan, Ji Jiang, Risheng Liu, and Zhongxuan Luo. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion. *IEEE TCSVT*, 2021.

[7] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE TPAMI*, 2020.

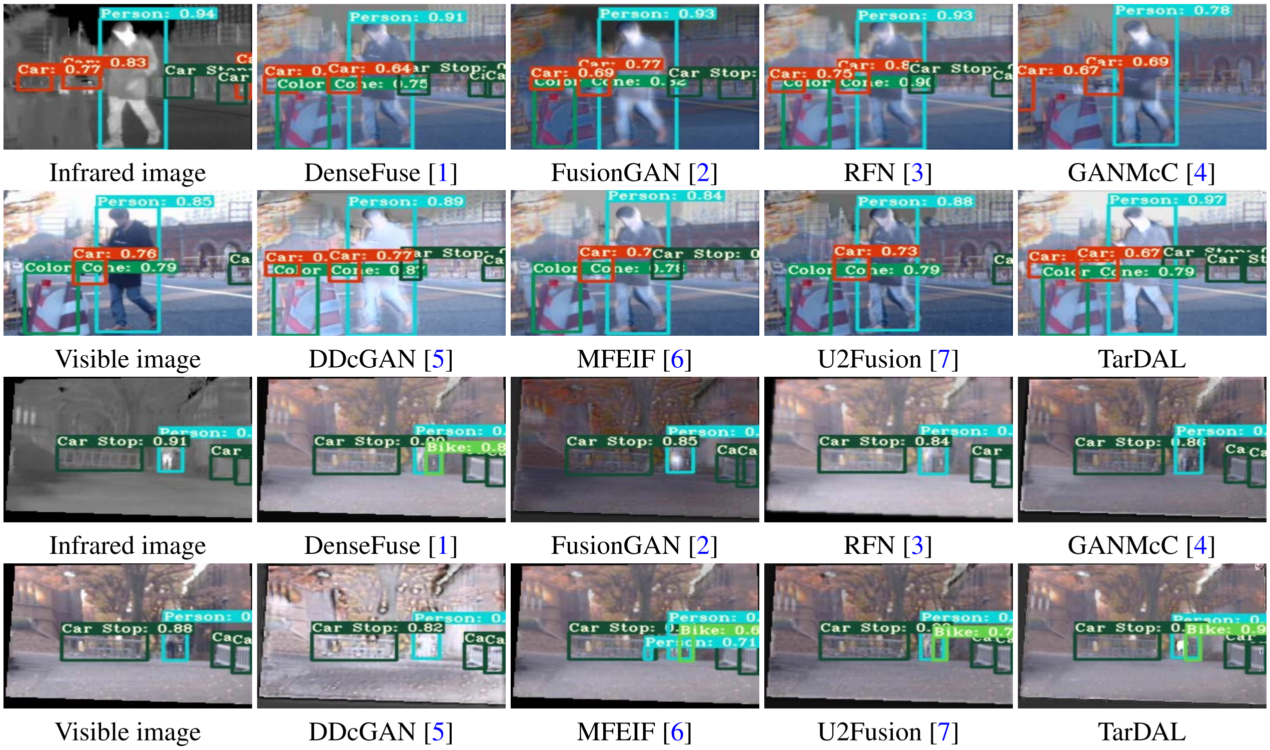

Figure 5. Visual results comparison of object detection between different methods on Multispectral dataset, respectively.

Figure 6. Visual results comparison object detection between different methods on our M^3FD dataset. Best viewed on screen.

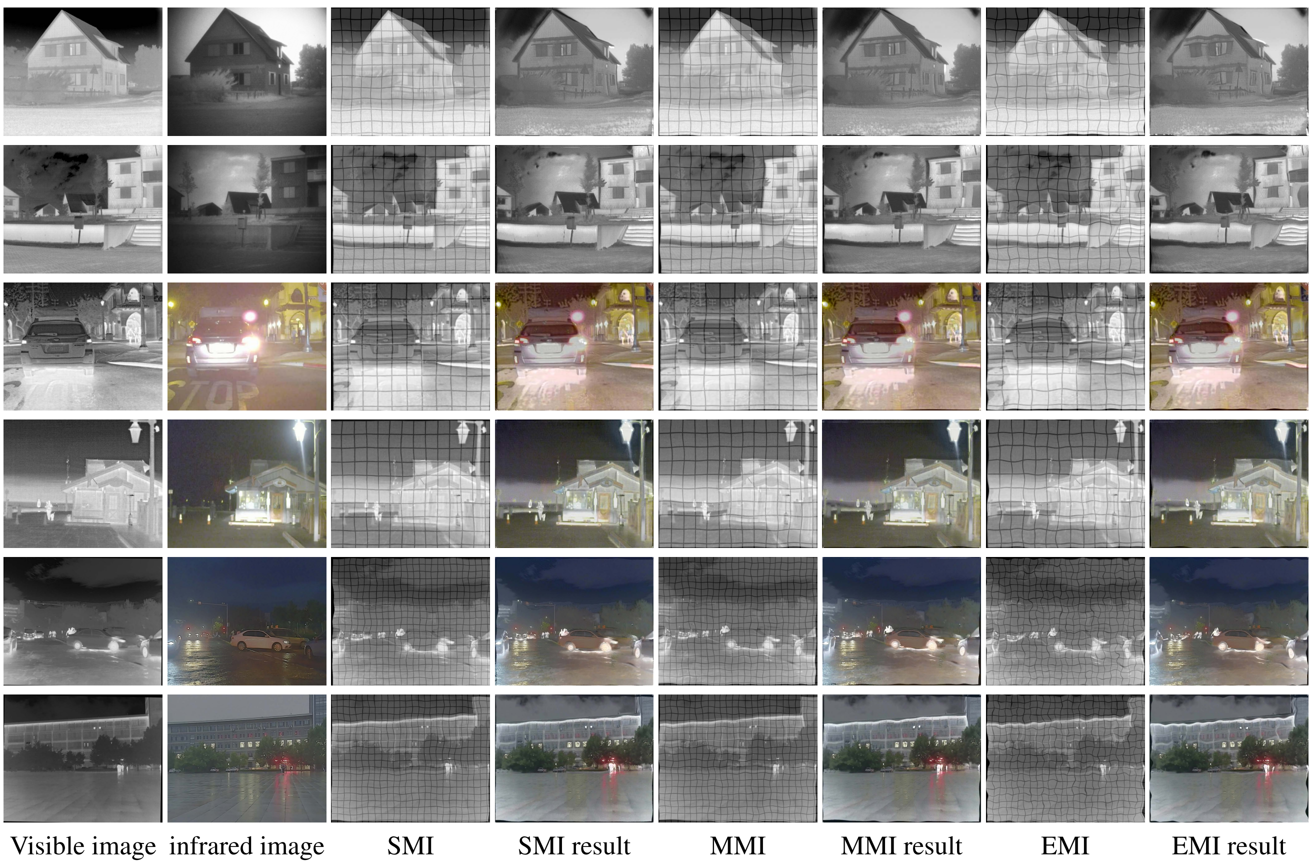

Figure 7. Visual results of our fusion method on unregistered image pairs on the TNO,Roadscene and M^3FD dataset, respectively. SMI, MMI and EMI denotes slight misaligned, moderate misaligned and extreme misaligned infrared image, respectively. Best viewed on screen.