Anchor-free Small-scale Multispectral Pedestrian Detection *[Supplementary Material]*

Alexander Wolpert[1,2]
alexander.wolpert@student.kit.edu
Michael Teutsch[2]
michael.teutsch@hensoldt.net
M. Saquib Sarfraz[1,3]
saquib.sarfraz@kit.edu
Rainer Stiefelhagen[1]
rainer.stiefelhagen@kit.edu

[1] Institute of Anthropomatics & Robotics
Karlsruhe Institute of Technology
Karlsruhe, Germany
[2] Hensoldt Optronics GmbH
Oberkochen, Germany
[3] Autonomous Systems
Daimler TSS
Germany

1 Multispectral Data Augmentation

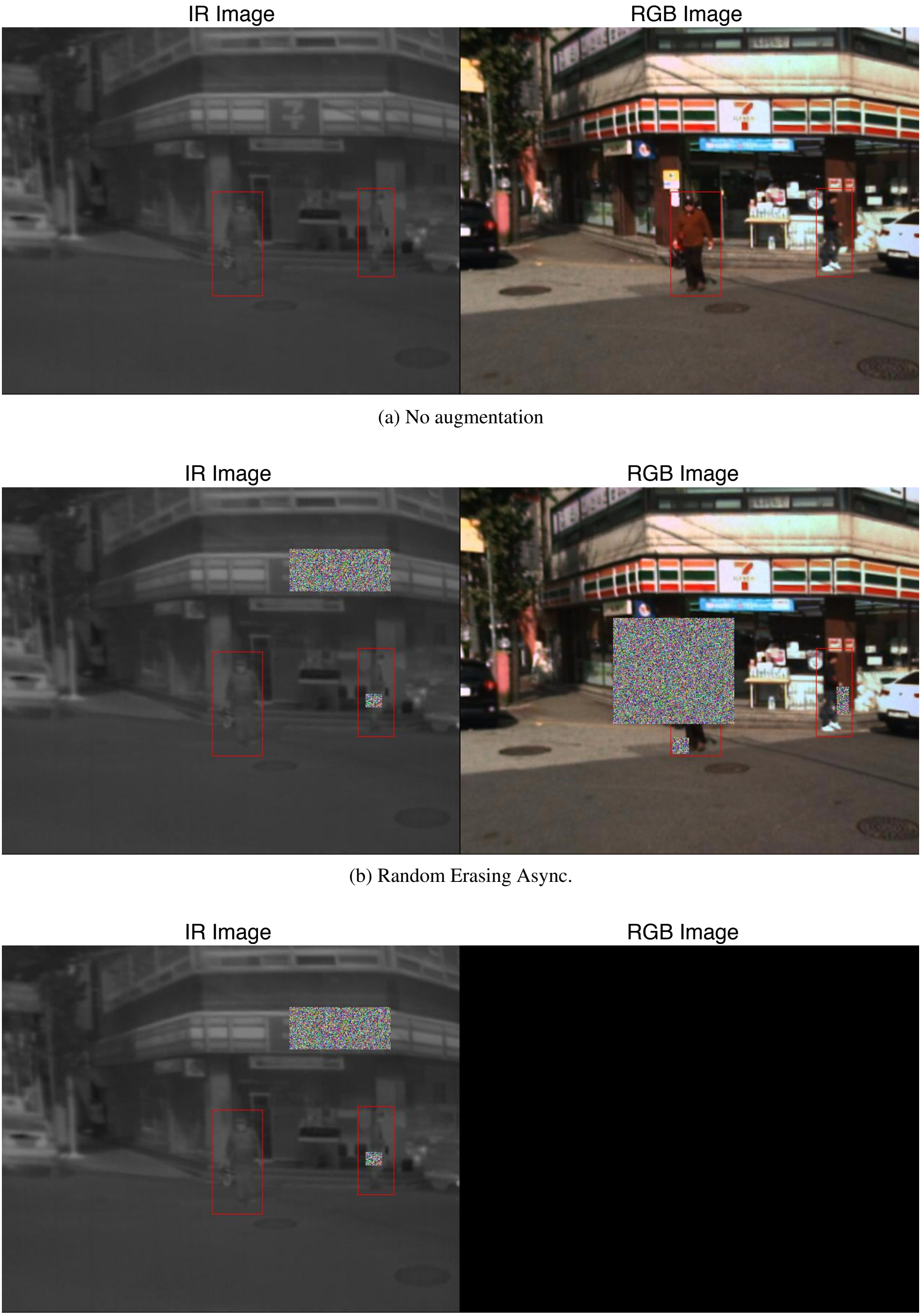
In Fig. 1 we provide a visualization of applying the *Random Erasing Async.* followed by the *Random Masking* data augmentations that were introduced in this paper.

2 Qualitative Comparison

In Fig. 2 we provide another set of examples, in which our detector is able to detect small-scaled instances where other approaches cannot distinguish between pedestrian or background correctly.

3 Quality of Annotations for the *All* Subset

In the context of evaluating under the challenging conditions of the *all* subset, we also did a qualitative investigation of the current test annotations. In Fig. 3 and 4 we show multiple instances, where our detector is able to find pedestrians that are not even labeled. Instead, these instances are only labeled in later frames as the pedestrian approaches the camera and increases in size. In the left column we show the frames, in which a pedestrian instance was not labeled but detected by a detector. In the right column we show the detections and annotations for one frame afterwards, where the pedestrian instance is labeled as it became larger. This brings us to the conclusion that, although our detector already outperforms many other approaches on the *all* subset, its performances would most likely improve even further, if these missing instances were labeled correctly. So, even though the dataset was already relabeled twice, there still is potential for improvement.



(a) No augmentation

(b) Random Erasing Async.

(c) Random Masking

Figure 1: Exemplary visualization of *Random Erasing Async.* and *Random Masking* for an image pair from the KAIST dataset. Pedestrian GTs are marked with a red box for clarity.

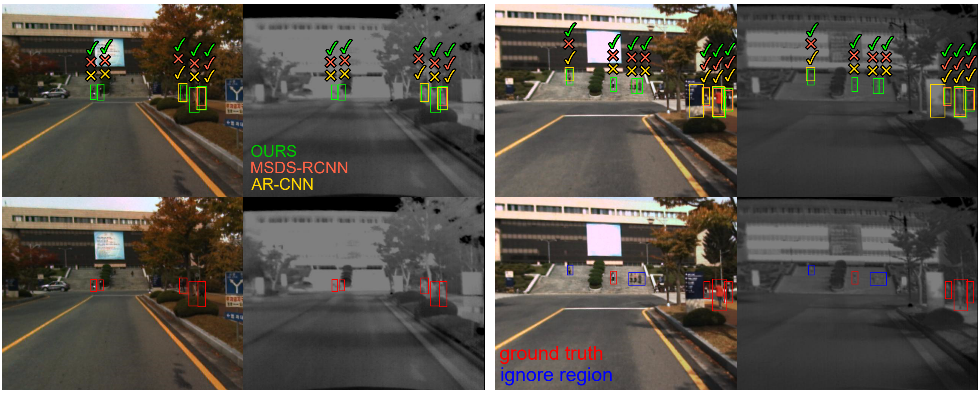


Figure 2: Comparison of detections of MSDS-RCNN [1] (orange), AR-CNN [2] (yellow) and OURS (green) in the top row. For better clarification we marked correctly detected pedestrians with a tick in the corresponding color of the detector and missed pedestrians with a cross in the corresponding color.

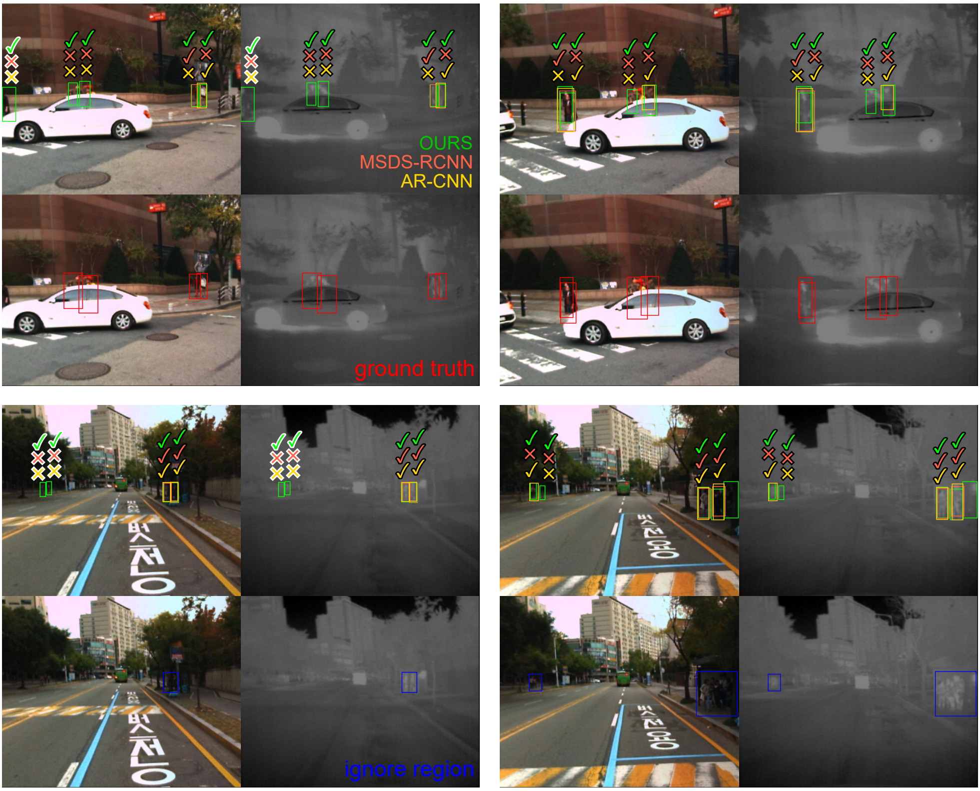


Figure 3: Multiple examples where our detector already detects pedestrians that are not labeled until one frame later. These cases are indicated by white framed ticks or crosses. Left column unlabeled instances, right column one frame later labeled.

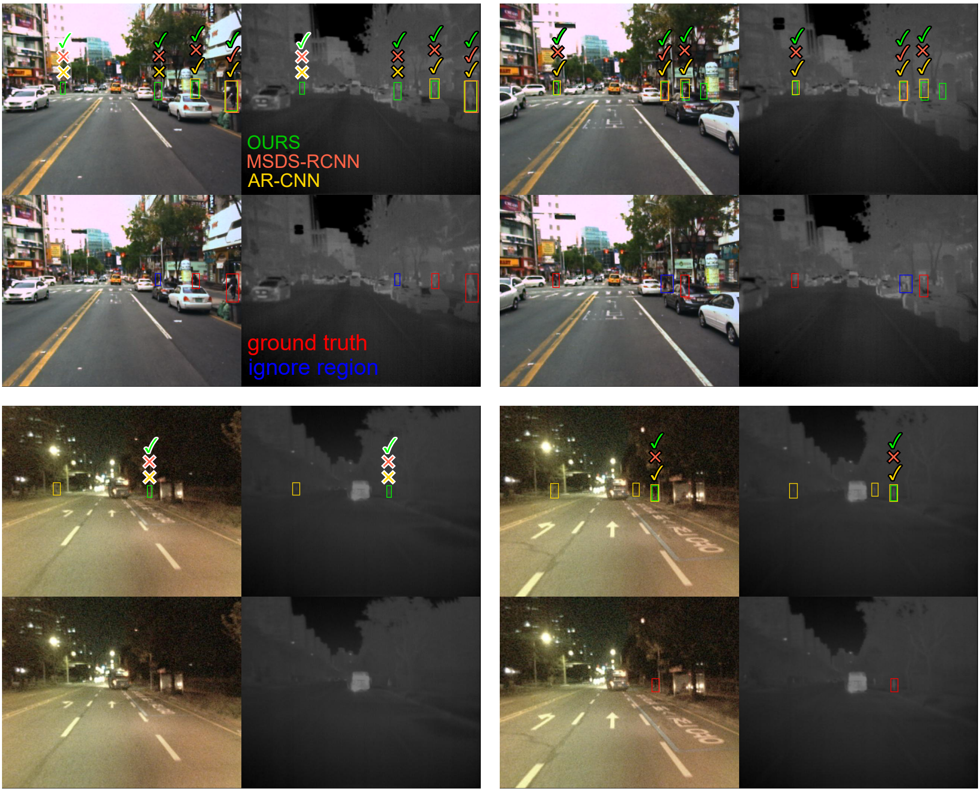


Figure 4: Multiple examples where our detector already detects pedestrians that are not labeled until one frame later. These cases are indicated by white framed ticks or crosses. Left column unlabeled instances, right column one frame later labeled.

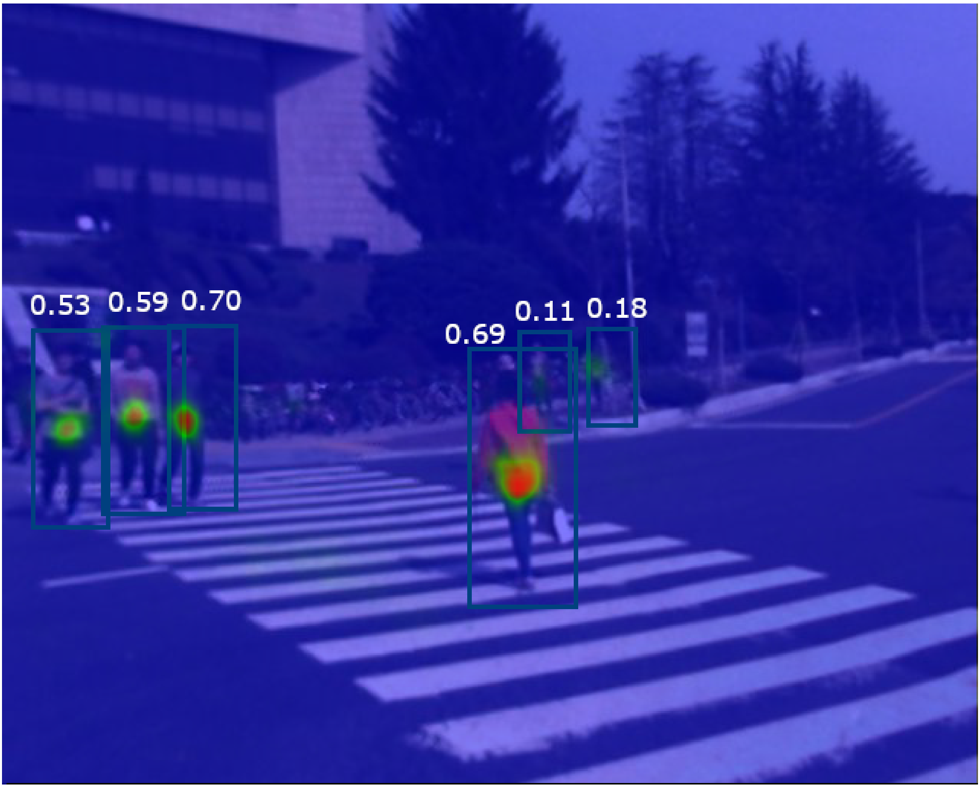


Figure 5: Visualization of the model's center map predictions as a heatmap. Detection confidence is depicted from blue (0.0) over green to red (1.0). The maximum confidence for each pedestrian instance is shown as white text. Ground truth annotations are visualized as green bounding boxes.

References

[1] Chengyang Li, Dan Song, Ruofeng Tong, and Min Tang. Multispectral pedestrian detection via simultaneous detection and segmentation. In *BMVC*, 2018.

[2] Lu Zhang, Xiangyu Zhu, Xiangyu Chen, Xu Yang, Zhen Lei, and Zhiyong Liu. Weakly aligned cross-modal learning for multispectral pedestrian detection. In *IEEE ICCV*, 2019.