Improving Multispectral Pedestrian Detection by Addressing Modality Imbalance Problems

1 Implementation Details

- The MBNet detector uses ResNet-50 as its backbone network, which is pretrained on ImageNet.
- The training IoU for the AP stage is set to {0.3, 0.5}, and for the IAFC stage, it is set to {0.5, 0.7}.
- Xavier initialization method is used to randomly initialize the other convolutional layers.
- During training, a patch of size [0.3, 1] of the input image is cropped and resized to 640×512.
- Each image is randomly color distorted and horizontally flipped with a probability of 0.5 to increase data diversity.
- Network is trained using the Adam optimizer for 7 epochs with a learning rate of 0.0001 and a batch size of 10.
- Anchor widths for stages 3 to 6 are set to {25.84, 29.39}, {33.81, 38.99}, {44.47, 52.54}, {65.80, 131.40} with a single anchor ratio of 0.41.
- To implement the code, I have joined all the various python files and classes into one big code for easy implementation.

2 Datasets Used

Recall that we are evaluating the performance of the pretrained MBNet Model on two datasets:

- **KAIST Dataset**: This dataset contains over 4500 images of RGB as well as thermal data, in both day time and night time. It has a fixed FPPI of 0.01 which is convenient for our usage.
- **CVC14 Dataset**: This dataset contains around 150 images of mostly night time images, along with their thermal versions. Due to its smaller size compared to the KAIST Dataset, it has a comparatively higher miss rate.

Both datasets provide us with a ground truth folder that contains the location of the bounding boxes around the pedestrians that is needed for calculating the miss rate.

3 Results

Dataset	**Baseline MR^{-2}**	**Achieved MR^{-2}**
KAIST	8.13%	8.76%
CVC14	21.1%	21.9%

Table 1: Results Table

The dataset provided in the google drive link from the paper, did not segregate the images into Day and Night, and this is why we find an overall miss rate and not the segregated miss rates for daytime and night-time as presented in the paper. As seen, we achieve results that are within 10% of the results presented in the paper.

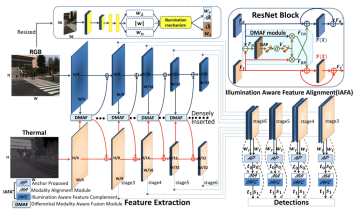

Figure 1: Architecture of the MBNet Network