# Illumination Distribution-Aware Thermal Pedestrian Detection Appendix

Anonymous Author(s)
Submission Id: 1559

## A FURTHER EXPERIMENTS

### A.1 Further visualization results

**Visualization comparisons based on different detector backbones.** In this part, we print the prediction results and annotations on some images and save them for visual comparison. We compare FasterRCNN+IDA (Ours) with FasterRCNN and CascadeRCNN+IDA (Ours) with CascadeRCNN on the KAIST dataset. As shown in Fig. 1, our method can reduce false positive predictions, improve true positive predictions and the confidence of predictions. Through the visual analysis of these experiments, we can further verify that the detectors with our method as a plug-in can get better performance. As shown in Fig. 2, our method can get better performance on the FLIR-aligned dataset. Highlighted by blue circles, with IDA, the detector can improve confidence, reduce FP, and reduce FN.

### A.2 Further studies on FasterRCNN+IDA

**Ablation Studies.** To explore the effectiveness of our method, we conduct an ablation study on the KAIST dataset. As shown in Table 1, our method reduces the MR by −2.19% from 21.07% to 18.88%. Only predicting mean vector or variance vector reduces −1.06% from 21.07% to 20.06% or −0.21% from 21.07% to 20.86%, respectively. Predicting both mean and variance vectors improves performance better because only using mean and variance can accurately describe the Gaussian distribution. The same applies to FLIR-aligned dataset. As shown in Table 2, with IDA, our method improves by +1.4% from 38.9% to 40.3%.

**Table 1: Ablation study (MR) of FasterRCNN on KAIST dataset.**

| Mean | Variance | All | Day | Night |
|---|---|---|---|---|
| × | × | 21.07 | 26.83 | 8.56 |
| ✓ | × | 20.06 | 26.44 | 6.91 |
| × | ✓ | 20.86 | 26.88 | 7.67 |
| ✓ | ✓ | **18.88** | **24.97** | **6.61** |

**Table 2: Ablation study (mAP) of FasterRCNN on FLIR-aligned dataset.**

| Mean | Variance | mAP | mAP50 | mAP75 |
|---|---|---|---|---|
| × | × | 38.9 | 74.3 | 34.3 |
| ✓ | × | 39.9 | **75.5** | 35.8 |
| × | ✓ | 40.0 | 75.3 | **35.9** |
| ✓ | ✓ | **40.3** | 75.4 | 35.8 |

**Length of the mean and variance vectors.** On the KAIST dataset, to explore the effectiveness of the length of the mean and variance vector, We modify the value of $K$ as $\{8, 16, 32, 64, 128\}$ respectively. As shown in Table 3, the detector performs best when $K = 32$. The low value of $K$ means that the feature losses too much in the process of feature dimensionality reduction. If the value of $K$ is high, predicting the mean and variance vectors will be difficult, which leads to bad performance. The same applies to FLIR-aligned dataset as shown in Table 2.

**Table 3: Detection results (MR) of FasterRCNN with varying $K$ and $N$ on KAIST dataset.**

| $K$ | All | Day | Night | $N$ | All | Day | Night |
|---|---|---|---|---|---|---|---|
| 8 | 20.58 | 26.33 | 9.50 | 10 | 19.83 | 25.06 | 8.55 |
| 16 | 20.84 | 27.13 | 8.13 | 25 | 21.08 | 26.71 | 9.45 |
| 32 | **18.88** | **24.97** | **6.61** | 40 | **18.88** | **24.97** | **6.61** |
| 64 | 21.03 | 26.86 | 8.48 | 55 | 19.79 | 25.82 | 8.17 |
| 128 | 21.05 | 27.03 | 9.27 | 70 | 21.27 | 27.55 | 8.60 |

**Table 4: Detection results (mAP) of FasterRCNN with varying $K$ and $N$ on FLIR-aligned dataset.**

| $K$ | mAP | mAP50 | mAP75 | $N$ | mAP | mAP50 | mAP75 |
|---|---|---|---|---|---|---|---|
| 8 | 39.4 | 74.9 | 34.7 | 10 | 39.5 | 75.1 | 35.1 |
| 16 | 39.1 | **75.7** | 34.3 | 25 | 39.4 | 75.2 | 34.9 |
| 32 | **40.3** | 75.4 | **35.8** | 40 | **40.3** | **75.4** | **35.8** |
| 64 | 39.3 | 74.2 | 35.0 | 55 | 39.7 | 75.2 | 35.7 |
| 128 | 39.8 | 74.9 | 34.9 | 70 | 39.0 | 74.0 | 34.9 |

**Size of the conditional feature co-occurrence matrices.** On the KAIST dataset, to explore the effectiveness of the size of the conditional feature co-occurrence matrices, we modify the value of $N$ by changing $\{10, 25, 40, 55, 70\}$. As shown in Table 3, the detector performs best when $N = 40$. The low value of $N$ leads to the matrices cannot accurately represent the joint Gaussian distribution, because two values with large difference maybe be quantized into the same interval. The high value of $N$ leads to the matrices being sparse. Since the total samples are a constant number $N_s$, the number assigned to a specific row of the matrix will be small. Therefore, frequency cannot be used to approximate a probability. The same applies to FLIR-aligned dataset as shown in Table 4.

**Sensitivity Analysis.** We perform analysis on $\lambda_m$ and $\lambda_v$ on the KAIST dataset. (1) We modify the value of $\lambda_m$ by changing $\{0.05, 0.075, 0.1, 0.125, 0.15\}$ when $\lambda_v = 0.125$. As shown in the left of Fig.3, the performance is the best when $\lambda_m = 0.1$. Our model can keep a relatively stable result in a wide range of $\lambda_m$. Besides, the low-value $\lambda_m$ means that the detector cannot accurately distinguish day, night, and ambiguous samples by predicting the mean of the conditional probability distribution. The high-value $\lambda_m$ leads to bad performance because of insufficient training for detection $L_{OD}$. (2) We

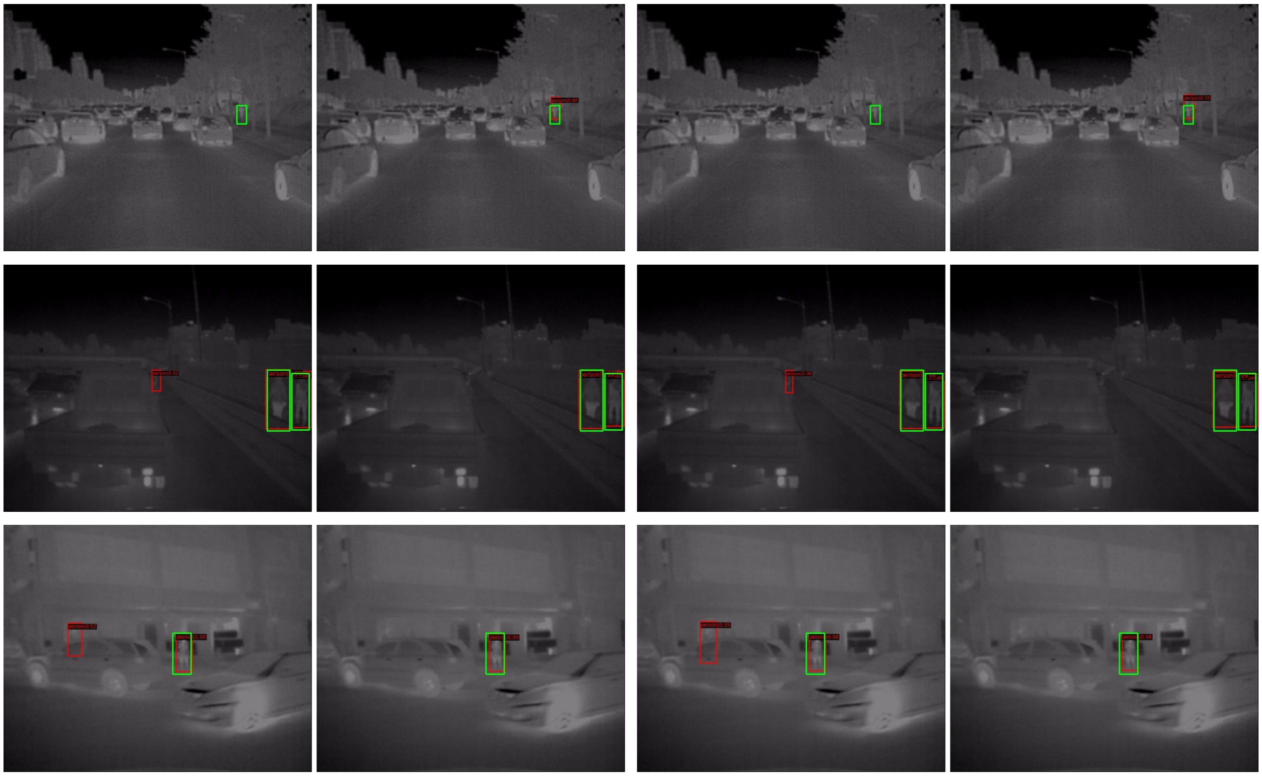

**Figure 1: Pedestrian detection results on KAIST dataset. The first, second, third, and fourth columns are the results of Faster-RCNN, FasterRCNN+IDA(Ours), CascadeRCNN, and CascadeRCNN+IDA(Ours) respectively. Green boxes are ground truth. Red boxes are prediction results. The first, second and third rows are results of the day, night and ambiguous samples respectively.**

modify the value of $\lambda_v$ by changing $\{0.075, 0.1, 0.125, 0.15, 0.175\}$ when $\lambda_m = 0.1$. As shown in the right of Fig.3, the performance is the best when $\lambda_v = 0.125$. Our model can keep a relatively stable result in a wide range of $\lambda_v$. Besides, the low-value $\lambda_v$ means that the detector cannot accurately distinguish day, night, and ambiguous samples by predicting the variance of the conditional probability distribution. The high-value $\lambda_v$ leads to bad performance because of insufficient training for detection $L_{OD}$. The same applies to FLIR-aligned dataset as shown in Fig . 4.

## A.3 Further studies on CascadeRCNN+IDA

**Ablation Studies.** To explore the effectiveness of our method, we conduct an ablation study on the KAIST dataset. As shown in Table 5, our method reduces the MR by −2.02% from 20.95% to 18.93%. Only predicting mean vector or variance vector reduces −1.21% from 20.95% to 19.74% or −1.11% from 20.95% to 19.84% respectively. Predicting both mean and variance vectors improves performance better because only using both mean and variance can accurately describe a Gaussian distribution. The same applies to FLIR-aligned dataset. As shown in Table 6, with IDA, our method improves by +0.4% from 41.0% to 41.4%.

**Length of the mean and variance vector and size of the conditional feature co-occurrence matrices.** To explore the effectiveness of the length of the mean and variance vectors, we also modify the value of $K$ by changing $\{8, 16, 32, 64, 128\}$. Similar to the FasterRCNN, as shown in Table 7, the detector performs the best when $K = 32$. To explore the effectiveness of the size of the conditional feature co-occurrence matrices, we modify the value of $N$ by changing $\{10, 25, 40, 55, 70\}$. Similarly, as shown in Table 7, the detector performs the best when $N = 40$. On the FLIR-aligned

**Table 5: Ablation study (MR) of CascadeRCNN on KAIST dataset.**

| Mean | Variance | All | Day | Night |
|---|---|---|---|---|
| × | × | 20.95 | 26.90 | 8.86 |
| ✓ | × | 19.74 | 25.70 | 7.33 |
| × | ✓ | 19.84 | 26.61 | 6.50 |
| ✓ | ✓ | **18.93** | **25.63** | **5.86** |

**Table 6: Ablation study (mAP) of CascadeRCNN on FLIR-aligned dataset.**

| Mean | Variance | mAP | mAP50 | mAP75 |
|---|---|---|---|---|
| × | × | 41.0 | 75.6 | 36.6 |
| ✓ | × | 41.1 | 75.5 | 36.6 |
| × | ✓ | 40.9 | 74.5 | **36.9** |
| ✓ | ✓ | **41.6** | **76.2** | 36.7 |

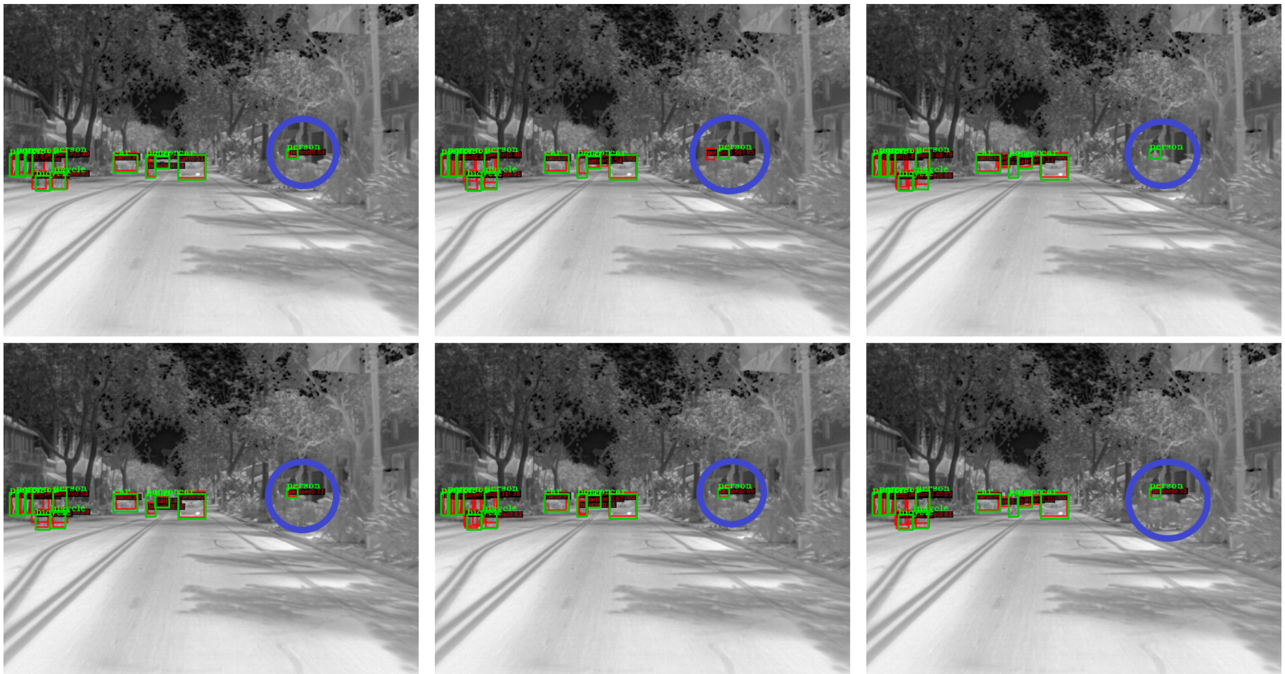

Figure 2: Detection results on FLIR-aligned dataset. The first, second, and third columns are the results of YOLOv3, FasterRCNN, and CascadeRCNN, respectively. The first and second rows are without and with IDA, respectively. Green boxes are ground truth. Red boxes are prediction results.

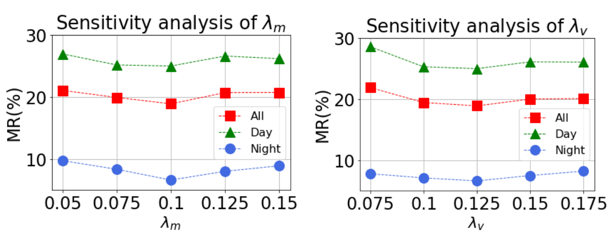


Figure 3: Hyperparameter analysis for FasterRCNN on the KAIST dataset: (left) $\lambda_m$ and (right) $\lambda_v$.

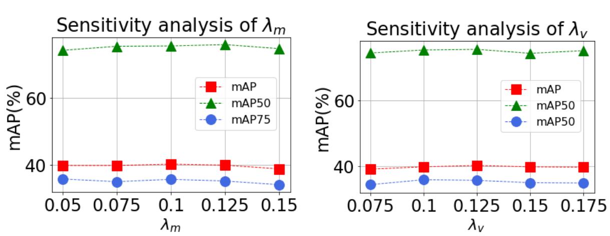


Figure 4: Hyperparameter analysis for FasterRCNN on the FLIR-aligned dataset: (left) $\lambda_m$ and (right) $\lambda_v$.

dataset, we modify the value of $K$ by changing $\{4, 8, 16, 32, 64\}$ and $N$ by changing $\{10, 25, 40, 55, 70\}$. As shown in Table 8, the detector performs the best when $K = 16$ and $N = 40$.

Table 7: Detection results (MR) of CascadeRCNN with varying $K$ and $N$ on KAIST dataset.

| $K$ | All | Day | Night | $N$ | All | Day | Night |
|---|---|---|---|---|---|---|---|
| 8 | 20.41 | 26.78 | 7.42 | 10 | 19.49 | 25.71 | 7.53 |
| 16 | 19.90 | 25.89 | 7.79 | 25 | 19.96 | 26.22 | 7.77 |
| 32 | **18.93** | **25.63** | **5.86** | 40 | **18.93** | 25.63 | **5.86** |
| 64 | 20.11 | 26.43 | 7.42 | 55 | 20.04 | **25.52** | 8.29 |
| 128 | 20.72 | 27.17 | 7.36 | 70 | 21.40 | 27.17 | 9.10 |

Table 8: Detection results (mAP) of CascadeRCNN with varying $K$ and $N$ on FLIR-aligned dataset.

| $K$ | mAP | mAP50 | mAP75 | $N$ | mAP | mAP50 | mAP75 |
|---|---|---|---|---|---|---|---|
| 4 | 40.9 | 75.0 | 36.2 | 10 | 41.3 | 74.7 | 36.7 |
| 8 | 40.7 | 74.9 | 36.5 | 25 | 41.1 | 75.5 | 36.5 |
| 16 | **41.6** | **76.2** | 36.7 | 40 | **41.6** | **76.2** | 36.7 |
| 32 | 41.4 | 75.5 | **37.4** | 55 | 41.3 | 74.5 | **36.9** |
| 64 | 40.4 | 74.5 | 35.6 | 70 | 41.2 | 75.0 | 36.6 |

**Sensitivity Analysis.** We also perform the same analysis on $\lambda_m$ and $\lambda_v$ as Section 2.3 on the KAIST dataset. As shown in the left of Fig.5, the performance is the best when $\lambda_m = 0.1$. Our model can keep a relatively stable result in a wide range of $\lambda_m$. For the parameter $\lambda_v$, by changing the value as $\{0.1, 0.15, 0.2, 0.25, 0.3\}$ when $\lambda_m = 0.1$, as shown in Fig.5, the performance is the best when

$\lambda_v = 0.2$. This value is a little bigger than the value for Faster-RCNN+IDA. This is because the variance vector values of CascadeR-CNN are usually bigger than the values of FasterRCNN, which leads to a higher loss of predicting the variance vector. The higher $\lambda_v$, the faster the converge is. Our model can also keep a relatively stable result in a wide range of $\lambda_v$. The same applies to FLIR-aligned dataset as shown in Fig.6.

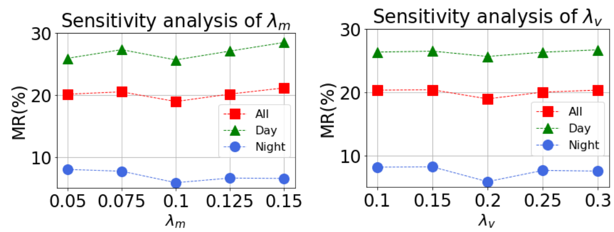


**Figure 5: Hyperparameter analysis for CascadeRCNN on KAIST dataset: (left) $\lambda_m$ and (right) $\lambda_v$.**

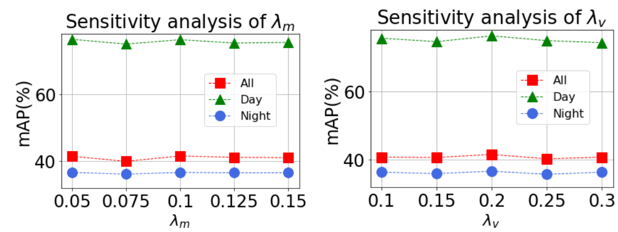


**Figure 6: Hyperparameter analysis for CascadeRCNN on FLIR-aligned dataset: (left) $\lambda_m$ and (right) $\lambda_v$.**