Abstract

Multiple Object Tracking (MOT) is an integral part of machine vision research. Most tracking-by-detection based MOT solutions utilize video streams from RGB cameras for their operation. However, for real-world applications, it is necessary to utilize sensors that operate in different spectrums to accommodate for varying lighting conditions. Since object detection is the first step of the tracking pipeline in tracking-by-detection approaches, we compare the performance of state-of-the-art object detectors when trained on color images to their performance when trained on thermal images. We introduce a new dataset for multiple object tracking with thermal images and corresponding RGB images and show that state-of-the-art trackers perform better on thermal images, especially in poor lighting conditions. Finally, we propose the use of a dynamic cut-off threshold for tracking-by-detection approaches that factors the size of a predicted box to enhance the tracker association. Our dataset and source code are publicly available at https://github.com/wassimea/thermalMOT

1. Introduction

Multiple object tracking is an important task in machine vision where the goal is to assign identities to different objects present in a video sequence, and effectively maintain the unique identity of objects across consecutive frames. With several applications including self-driving cars, human computer interaction, and virtual reality, MOT has been a popular area of research in the computer vision domain.

1.1. Multiple Object Tracking

Most state-of-the-art MOT methods [1, 5, 21, 25] utilize the tracking-by-detection paradigm, which is a two-stage process. In the first stage, a standalone detector is run on the video sequence to predict the location and class of objects present in the frames. In the second stage, a tracker processes these detections to conduct association: assigning a unique ID to detections of the same object across consecutive frames. While the performance of an object detector is usually measured by how accurate it can localize and classify objects in a single frame, the performance of a multiple object tracker also factors how well it can correctly re-identify an object across consecutive frames of a video sequence.

While some approaches have been developed that perform one shot, end-to-end tracking [13, 23], such methods were still unable to surpass the performance of two-stage tracking methods. Zhang et al. [26] performed an empirical study that concluded that the tasks of object detection and object tracking often compete with each other during training, causing one-shot trackers to be less accurate than two-stage trackers.

1.2. Perception Systems and Thermal Sensors

Deep learning helped achieve significant breakthroughs in machine vision tasks (advanced perception modules, intelligent monitoring systems, and autonomous vehicles, to name just a few) [17]. Such systems rely on the fusion of information from multiple types of sensors (lidar, radar, RGB cameras, depth sensors, thermal sensors, etc.) to get a better perception of the environment. This information is processed by an artificial intelligence module to perform advanced analysis and make critical decisions. Accurate multiple object detection and tracking is an essential task as it allows the...
localization of objects of interest and the prediction of
the trajectory of moving objects.

The use of thermal sensors in machine vision is be-
coming more popular [6, 12] as they offer a powerful
perception of the thermal identities of objects in a scene.
They are also suitable for outdoor applications as ther-
mal sensors operate normally at night and are not signif-
ically affected by poor weather conditions [20].

1.3. Contribution

In this paper, we study the feasibility of using ther-
mal sensors to conduct accurate multiple object detec-
tion and tracking. The main contributions of our work
are summarized below:

• We introduce a new dataset for multiple object
tracking with images and ground truth annotations
for RGB and corresponding thermal images.

• We compare the performance of two state-of-the-
art object detectors (TOOD [9], VFNET [24]) when
trained on thermal images to when they are trained
on RGB images of the Teledyne FLIR Thermal
Dataset for Advanced Driver-Assistance Systems 1.

• We study the efficacy of applying transfer learning
of weights trained on a RGB dataset when training
an object detector on thermal images.

• We develop a tracking-by-detection MOT method
based on the current state-of-the-art approaches
that operates on thermal images, and enhance the
data association of the tracker by applying a dy-
namic cut off score for detections based on the pre-
dicted box area.

2. Literature Review

In this section, we provide an overview of the existing
methods and approaches that we build upon in our work.

2.1. Object Detection

2.1.1 Task-aligned One-stage Object Detection
(TOOD)

Feng et al. introduced the Task-aligned One-stage Ob-
ject Detection (TOOD) [9] which strengthens the link
between the localization and the classification tasks of
object detection. This is accomplished by introduc-
ing "Task-Alignment": taking the network’s outputs for
each task and passing them to a network head that mod-
ifies the score and the location predictions to align their
optimal anchors. Their Task alignment learning (TAL)
also pushes the network to predict better aligned bound-
ing boxes. TOOD achieved an AP of 51.1 on the MS-
COCO dataset.

2.1.2 VarifocalNet (VFNET)

Zhang et al. designed VFNet using Varifocal Loss [24].
This loss is meant to maximize the IoU-aware classifica-
tion score (IACS) that takes into account both the clas-
sification and the location of a prediction. The Varifocal
loss also modifies focal loss by weighing positive ex-
amples more heavily than negative ones. Additionally,
VFNet uses a new nine-point deformable convolution
representation for bounding boxes and a network head
to refine the network’s box predictions by learning an
additional offset to their locations.

2.2. Tracking By Detection

The tracking-by-detection paradigm is more suitable
for real-world applications, where different detectors
could be used as a first step in the tracking pipeline,
and the training data does not necessarily need to con-
tain tracking ground truth labels. In addition, the rapid
breakthrough in deep learning has led to the emergence
of faster and more accurate detectors [9, 14, 24]. This
has led to more research in MOT utilizing state-of-the-
art object detection models [3, 22].

Since object detectors are not perfect, and there will
always be cases where the detector predicts false pos-
itive boxes or misses true detections (false negatives),
state-of-the-art MOT approaches often eliminate pre-
dicted boxes with a low confidence by setting a cut-
off threshold for detector confidence. However, this in-
evitably leads to some true detections being ignored, es-
pecially in cases where there is occlusion.

Of the systems utilizing the tracking-by-detection
paradigm, ByteTrack [25] has achieved state-of-the-art
performance on the test data of the benchmark MOT17
dataset [16] with an MOTA of 80.3%. ByteTrack uti-
lizes YOLOX [10] to generate detections. Instead of ig-
noring detection boxes with low confidence, ByteTrack
separates the predicted boxes into a set of high-score de-
tection boxes, and a set of low-score detection boxes.
The algorithm first predicts the locations of the tracklets
in the next frame using a Kalman filter, then matches the
tracklets with the high-score detection boxes by comput-
ing the IOU between the high-score detection boxes and
the predicted tracklet location. Tracklets that remain un-
matched through this first association are then matched
with the low-score detection boxes through a second as-
sociation step. At the end of this two-step association
process, tracklets that remain unmatched are deemed to
be lost, new tracklets are created from the unmatched
high-score detection boxes, and the remaining low-score

---

1https://www.flir.ca/oem/adas/adas-dataset-form/
3. Proposed Approach and Experiments

In this section, we elaborate on the experiments we have conducted in the color and thermal domain on object detection, multiple object tracking, and our proposed dynamic confidence thresholding for MOT.

3.1. Datasets

3.1.1 City Scene RGB-Thermal MOT Dataset

For the purpose of comparing tracking methods on visible and infrared images, we collected and manually annotated a dataset using a FLIR infrared camera and a visible-light camera, both at a framerate of 10Hz. One in every 2 frames was annotated, resulting in an effective framerate of 5Hz for the dataset. During data collection, the cameras were static and aimed at a city intersection. Cars and pedestrians were annotated up to a distance of 300m and 100m respectively.

The dataset is composed of 15 sequences collected over the course of a day, for a total of 1,997 annotated frames. These sequences are divided into a training set of 12 sequences (1,591 frames) and a testing set of 3 sequences (406 frames). Samples from the three testing sequences are given in Figure 1.

The annotated frames contains 267 unique car instances, 25,985 car bounding boxes, 145 unique pedestrian instances, and 7,822 pedestrian bounding boxes.

3.1.2 FLIR ADAS dataset

In our experiments, we study the performance of state-of-the-art object detectors when trained on color images compared to when trained on thermal images. To that end, we use the FLIR Thermal Dataset for Advanced Driver-Assistance Systems, which is one of the largest and most comprehensive thermal datasets that also provides color images corresponding to the thermal images. It is composed of city scenes captured using a thermal sensor and an RGB camera installed on top of a car. It is manually annotated for 15 classes (person, car, bike, etc.). There is a total of 11,886 training images and 3,749 testing images.

3.2. Object Detection Experiments

The object detector is an integral part of tracking-by-detection approaches. The performance of the tracker is significantly influenced by the performance of the object detector. In our experiments on object detection, we address the following two matters:

- We study the efficacy of applying transfer learning of weights trained on the Imagenet dataset [7] (color images) to train an object detector on thermal images.
- We compare the performance of object detectors when trained on thermal images against when they are trained on color images. As both the FLIR ADAS dataset and our City Scene RGB-Thermal...
MOT dataset contain images taken at night, this experiment is important to examine the efficacy of thermal images under poor lighting conditions.

Table 1 shows a summary of the object detection experiments conducted. We use the MMDetection toolbox [4] for all object detection training experiments we perform.

We use Resnet50 [11] as the backbone of the detectors in all experiments. The models are trained for 6 epochs with a batch size of 8, with an initial learning rate of 0.01. We use focal loss [15] for box classification. We apply data augmentation (resizing and flipping) to enrich the dataset.

### 3.3. MOT Experiments

For our experiments on MOT, we use the state-of-the-art ByteTrack [25] approach. ByteTrack’s simple design that conducts data association based on motion similarity make it ideal for our experiments, as the tracker does not require any domain-specific training. We fine tune the trained object detectors on the tracking dataset we collected as follows:

- Fine tune TOOD and VFNET on the thermal images of our City Scene RGB-Thermal MOT dataset with weights initialized from the trained detectors on the ADAS Thermal dataset.
- Fine tune TOOD and VFNET on the RGB images of our City Scene RGB-Thermal MOT dataset with weights initialized from the trained detectors on the ADAS RGB dataset.

### 3.4. Dynamic Confidence Cut-Off (DCC)

In the original implementation of ByteTrack, the authors set a minimum detection threshold $T_{\text{min}}$(0.1) and a threshold for high score detection boxes $T_{\text{high}}$(0.5). Bounding boxes that have a confidence below $T_{\text{min}}$ are ignored, boxes that have a confidence between $T_{\text{min}}$ and $T_{\text{high}}$ are treated as low score detection boxes, and boxes that have a confidence higher than $T_{\text{high}}$ are treated as high score detection boxes as explained in Section 2.2.

A drawback of this implementation is that it treats all objects of all sizes in the same manner strictly based on which range the confidence value falls in. This inevitably leads to false and missed detections as it does not take into account several factors like very small detections that are far from the camera, significant overlap and occlusion, especially as the size of the objects gets smaller as they move away from the camera (as is the case in some sequences in our City Scene RGB-Thermal MOT dataset, where the size of a tracked car for example becomes smaller as it moves further from the camera).

To address this issue, we propose the use of a dynamic confidence cut-off score for $T_{\text{high}}$ in the implementation of ByteTrack inspired by the work conducted by Stalder et al. [19]. Our results show that the dynamic confidence cut-off significantly improves the performance of the trackers, especially for objects with a smaller area. We elaborate more on the findings in Section 4.2.

### 4. Results

In this section, we elaborate and analyze the results we achieved on object detection and MOT.

#### 4.1. Precision Recall of Object Detectors

To study the performance of TOOD and VFNET on the thermal and RGB variants of the ADAS dataset, we plot the precision-recall curves of all the models from Table 1. The results are given in Figure 2 for TOOD and Figure 3 for VFNET.

From the analysis of the precision-recall curve of TOOD, we see that the detectors trained on thermal images perform significantly better than the ones trained on RGB images. This could be attributed to the fact that the ADAS dataset contains data captured at night, where the RGB images would contain little to no features of the objects of interest present in the frames. This explains why the maximum recall achieved by the detectors trained on RGB images is 79%, while the maximum recall achieved by the detectors trained on thermal images is 97%. This shows the superiority of detectors trained on thermal images, especially under poor lighting conditions. The ADAS dataset does not split the frames captured during the day from those captured during the night, so we were unable to conduct experiments to show the performance of the detectors exclusively on the frames taken at night. However, in the dataset that we collected, we split the sequences taken during the day from those taken at night, allowing us to compare the performance of the trackers in both settings.

The results also show that using transfer learning when training the detectors improves the overall performance of the detector. However, the results show that the effect of transfer learning was more significant on the detectors trained on the RGB images of the ADAS dataset. This could be explained by the fact that pre-trained weights are also a result of training RGB images (the Imagenet dataset [7]). While transfer learning did improve the overall performance of the detector trained on thermal images, the improvement was not as significant. This shows that transfer learning is most effective when the initial weights and the dataset on which the detector is being trained on both belong to the same spectrum.
<table>
<thead>
<tr>
<th>Experiment #</th>
<th>Weight Initialization</th>
<th>Trained / Tested on</th>
<th>Object Detector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Imagenet</td>
<td>ADAS Thermal Train/Test Sets</td>
<td>TOOD</td>
</tr>
<tr>
<td>2</td>
<td>Random</td>
<td>ADAS Thermal Train/Test Sets</td>
<td>TOOD</td>
</tr>
<tr>
<td>3</td>
<td>Imagenet</td>
<td>ADAS RGB Train/Test Sets</td>
<td>TOOD</td>
</tr>
<tr>
<td>4</td>
<td>Random</td>
<td>ADAS RGB Train/Test Sets</td>
<td>TOOD</td>
</tr>
<tr>
<td>5</td>
<td>Imagenet</td>
<td>ADAS Thermal Train/Test Sets</td>
<td>VFNET</td>
</tr>
<tr>
<td>6</td>
<td>Random</td>
<td>ADAS Thermal Train/Test Sets</td>
<td>VFNET</td>
</tr>
<tr>
<td>7</td>
<td>Imagenet</td>
<td>ADAS RGB Train/Test Sets</td>
<td>VFNET</td>
</tr>
<tr>
<td>8</td>
<td>Random</td>
<td>ADAS RGB Train/Test Sets</td>
<td>VFNET</td>
</tr>
</tbody>
</table>

Table 1. Object detection experiments conducted

Figure 2. TOOD Precision-Recall curve. A detection is considered a true positive if it has at least a 0.5 IOU with a ground truth box.

Figure 3. VFNET Precision-Recall curve. A detection is considered a true positive if it has at least a 0.5 IOU with a ground truth box.

The analysis of the precision-recall curve of VFNET confirms the above findings. The maximum recall achieved from the detectors trained on RGB images is 78%, while the maximum recall achieved by the detectors trained on thermal images is 96%. Applying transfer learning also improved the performance of detectors both in the color and thermal domains, but we notice that the improvement in the color domain is more significant.

4.2. MOT Metrics

To study the effectiveness of the trained detectors in the task of MOT, we calculate the standard MOT metrics [2, 18] of ByteTrack with DCC when using each trained detector on our City Scene RGB-Thermal MOT dataset. We retrain the detectors on our City Scene RGB-Thermal MOT dataset with the weights initialized from the weights of the detectors trained on the ADAS dataset. We study the performance of the tracker across three testing sequences, taken in the morning, afternoon, and at night to study the performance of the tracker under different lighting conditions. The results are given in Table 2.

The results on the sequence taken in the morning show that the trackers operating on RGB images are more effective than trackers operating on thermal images. This could be explained through analysis of the sunlight distribution in that sequence. From Figure 1, we notice that the field of view of the thermal camera covers an area that is partially in the shade. Since the dataset was collected in a hot day in the month of July, there is a considerable difference in temperature between the areas in the shade and the areas in direct sunlight. This could be noticed by comparing the appearance of the car present in the shade to the appearance of the car present in the sunlight. This difference makes it more difficult for the detector to operate on thermal images of this sequence.

In the second sequence taken in the afternoon, we can see from Figure 1 that the entire field of view of the thermal camera is almost entirely in the shade, so there is no significant variation in temperature values across different parts of the field of view. This results in the tracker performing better on thermal images, even having a higher MOTA than the tracker running on RGB images.

In the sequence taken at night, we notice that the trackers operating on thermal images perform significantly better than the trackers operating on RGB images.
(34% higher MOTA when using TOOD, 48% higher MOTA when using VFNET). This is further proof that the detectors perform better on thermal images when there is not a significant variation in temperature across different parts of the field of view. In addition, the trackers operating on RGB images are expected to struggle in detecting objects at night as they are not clearly visible.

To study the effect of using a dynamic cut-off confidence, we compare the performance of the trackers with DCC against the performance of the trackers when using a fixed confidence for high-score detection boxes as in the original implementation of ByteTrack does. The results are given in Table 3. It can be seen that applying a DCC noticeably improves the MOTA of the trackers, especially since there are lots of objects with a small area in the dataset (far from the camera).

4.3. Speed

We benchmark the speed of the proposed trackers in two environments:

- A powerful machine with an NVIDIA RTX 3090 GPU, and an Intel Core i9 - 10900X 3.70 GHz Processor. (Referred to as AWP).
- A lower power edge device, NVIDIA Jetson Xavier, with a 512-core Volta GPU, and a 8-core ARM v8.2 64-bit Processor. (Referred to as Xavier).

The results are given in Table 4. Since the tracking process is independent from the detection process, and the tracking association runs on CPU while the detector inference runs on the GPU, the latency of the tracking process is constant.

4.4. Failure Cases

As elaborated in Section 4.2, the thermal trackers perform poorly under conditions where there are different intensities of sunlight in the thermal sensor’s field of view. We also discuss the limitations of the RGB trackers at night. In addition, we notice that there is a noticeable number of incorrect annotations in the FLIR ADAS dataset, on which we heavily rely for the training of our detectors. There are several instances where pedestrians present in an image are not annotated. This would cause a problem during training when a large number of objects predicted as pedestrians by the model do not have corresponding ground truth annotations, causing the model to train on considering them negatives when they are actually true positives. Similarly, when evaluating the model, several false positives are actually true positives, which would negatively affect the precision and recall values. We randomly sample 100 testing images from the ADAS Thermal dataset, and found that 6 of them have annotation faults (2 of which are shown in Figure 4. This should also be taken into consideration when analyzing the precision-recall curves of the models.

5. Conclusion and Future Work

In this paper, we have conducted in-depth empirical studies to analyze the feasibility of using thermal sensors for multiple object detection and tracking. We train two state-of-the-art object detectors on the RGB and thermal variations of the FLIR ADAS dataset, and study the efficacy of transfer learning when applied to training a detector on a dataset from a different spectrum than the initial weights.

We show the superiority of detectors trained on thermal images compared to those trained on RGB images, especially under poor lighting conditions. We also show that transfer learning is more effective when used to train a detector on a dataset from the same spectrum as the initial weights. We introduce the use of a dynamic confidence cut-off, which factors the size of a predicted box, as an enhancement to the motion similarity association of tracking-by-detection based MOT, and show that it improves the accuracy of the tracker.

We have highlighted the limitations of trackers operating on RGB images under poor lighting conditions, and the limitations of trackers operating on thermal images when the thermal sensor field of view is covering areas of significantly different sunlight intensities.

Our experiments and analysis clearly highlight the importance of sensor fusion, especially in critical systems like ADAS. Each type of sensor is optimal under certain conditions and has limitations under other conditions. Being able to combine data sources from different spectrums significantly enhances the perception ability of an autonomous system.

In the future, our work can be expanded by examining further enhancements to the data association of the tracker that utilize the thermal information of an object. In addition, developing a tracker that utilizes information from both the visible and non-visible spectrum to enhance the tracking accuracy can be investigated.

References


(a) Results on the test sequence taken in the morning.

<table>
<thead>
<tr>
<th>Detector</th>
<th>MOTA</th>
<th>IDF1</th>
<th>MT</th>
<th>PT</th>
<th>ML</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOOD Thermal</td>
<td>0.35</td>
<td>0.57</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>63</td>
<td>403</td>
</tr>
<tr>
<td>TOOD RGB</td>
<td>0.86</td>
<td>0.92</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>VFNET Thermal</td>
<td>0.33</td>
<td>0.55</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>56</td>
<td>417</td>
</tr>
<tr>
<td>VFNET RGB</td>
<td>0.66</td>
<td>0.84</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>185</td>
<td>61</td>
</tr>
</tbody>
</table>

(b) Results on the test sequence taken in the afternoon.

<table>
<thead>
<tr>
<th>Detector</th>
<th>MOTA</th>
<th>IDF1</th>
<th>MT</th>
<th>PT</th>
<th>ML</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOOD Thermal</td>
<td>0.77</td>
<td>0.85</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>83</td>
<td>14</td>
</tr>
<tr>
<td>TOOD RGB</td>
<td>0.7</td>
<td>0.82</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>46</td>
<td>14</td>
</tr>
<tr>
<td>VFNET Thermal</td>
<td>0.58</td>
<td>0.79</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>153</td>
<td>0</td>
</tr>
<tr>
<td>VFNET RGB</td>
<td>0.57</td>
<td>0.75</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>175</td>
<td>0</td>
</tr>
</tbody>
</table>

(c) Results on the test sequence taken at night.

<table>
<thead>
<tr>
<th>Detector</th>
<th>MOTA</th>
<th>IDF1</th>
<th>MT</th>
<th>PT</th>
<th>ML</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOOD Thermal</td>
<td>0.85</td>
<td>0.92</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>40</td>
<td>81</td>
</tr>
<tr>
<td>TOOD RGB</td>
<td>0.51</td>
<td>0.67</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>126</td>
<td>268</td>
</tr>
<tr>
<td>VFNET Thermal</td>
<td>0.82</td>
<td>0.91</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>43</td>
<td>106</td>
</tr>
<tr>
<td>VFNET RGB</td>
<td>0.34</td>
<td>0.65</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>337</td>
<td>189</td>
</tr>
</tbody>
</table>

(d) Overall results across the three testing sequences.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Overall MOTA without DCC</th>
<th>Overall MOTA with DCC</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOOD Thermal</td>
<td>0.58</td>
<td>0.67</td>
<td>9%</td>
</tr>
<tr>
<td>TOOD RGB</td>
<td>0.62</td>
<td>0.69</td>
<td>7%</td>
</tr>
<tr>
<td>VFNET Thermal</td>
<td>0.51</td>
<td>0.59</td>
<td>8%</td>
</tr>
<tr>
<td>VFNET RGB</td>
<td>0.46</td>
<td>0.52</td>
<td>6%</td>
</tr>
</tbody>
</table>

Table 2. Overall ByteTrack with dynamic cut-off results on the 3 sequences of our City Scene RGB-Thermal MOT dataset. a) First test sequence captured during daytime. b) Second test sequence captured during daytime. c) Third test sequence taken at night. d) Overall performance across all 3 testing sequences.

Table 3. Comparison of the performance of the trackers before and after the use of a DCC on our City Scene RGB-Thermal MOT dataset.
Table 4. Benchmarking tracker speed on the AWP machine and the Jetson Xavier module.

<table>
<thead>
<tr>
<th></th>
<th>AWP</th>
<th>Xavier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection Time (ms)</td>
<td>Tood + DCC</td>
<td>VFNET + DCC</td>
</tr>
<tr>
<td>0.05783</td>
<td>0.06197</td>
<td>0.54908</td>
</tr>
<tr>
<td>Tracking Time (ms)</td>
<td>0.00085</td>
<td>0.00085</td>
</tr>
<tr>
<td>Total (ms)</td>
<td>0.05868</td>
<td>0.06282</td>
</tr>
<tr>
<td>Total Frames Per Second</td>
<td>17.04</td>
<td>15.92</td>
</tr>
</tbody>
</table>

Figure 4. Samples showing missing annotations from the FLIR ADAS Dataset. Green rectangles represent ground truth boxes, red rectangles represent detected boxes, orange arrows point to instances where a pedestrian is detected by a model but no ground truth annotation exists.
