The effect of low light conditions on pedestrian detection using YOLOv8

Joachim Verschelde, Johan van Nispen
Open University of the Netherlands

Abstract

Preventing and reducing traffic injuries and fatalities involving pedestrians under low light conditions remains an important issue. The addition of a pedestrian detection function, which alerts car drivers on the presence of pedestrians, to an Advanced Driver Assistance System (ADAS) will help in reducing the number of accidents. This project investigates if the addition of infrared camera image input, on top of optic color image input, improves pedestrian detection under low light conditions. By using the pre-trained YOLOv8 nano object detection model on the annotated KAIST dataset, which consists of 95.328 synchronised optic color and thermal camera images in both day and night traffic situations, it was shown that the addition of infrared camera input slightly increases the pedestrian detection performance during nighttime conditions. However, this effect needs to be quantified further. It was shown that for the three different occlusion levels, only for the highest occlusion level the addition of infrared camera input increases the detection performance.

Introduction

A large number of traffic fatalities and injuries in the United States and the United Kingdom involve pedestrians. It is estimated that 75% of all traffic incidents involving pedestrians occur during dark conditions (Sanders, Schneider, & Proulx, 2022). Data from these studies only includes numbers from developed countries, and it is believed that the numbers for developing countries are even higher (Nataprawira, Gu, Goncharenko, & Kamijo, 2021). Because human eyes do not adapt very well to dark conditions and because of driver fatigue, accident rates during the night are $1 - 1.5$ times higher than during the day, and the death rate per kilometer is approximately three times higher during the night (Yi, Luo, Chen, & Hu, 2022). Preventing accidents involving pedestrians under low light conditions is therefore considered an important issue, and is an active area of research.

The addition of a pedestrian detection function to an Advanced Driver Assistance System (ADAS) aids in preventing traffic accidents involving pedestrians under low light conditions. Although several techniques exist in detecting pedestrians, incl. Light Detection and Ranging (LiDAR), Millimeter-Wave Radar (MWR), and sensor fusion, the most

commonly used sensor is the optic color (RGB) camera. However, the use of optic color implies that optic camera based pedestrian detection systems cannot avoid the problem of detecting under low light conditions (Nataprawira et al., 2021).

In computer vision, object detection is an important and challenging problem, and a lot of study has been done on the subject (Diwan, Anirudh, & Tembhurne, 2023). Recently, the most promising development for real-time object detection and object localisation is the YOLO (You Only Look Once) framework (Redmon, Divvala, Girshick, & Farhadi, 2016). Because of it's real-time speed and accuracy, YOLO is now used as a critical component in many applications, incl. autonomous vehicles, robotics, video surveillance and augmented reality. An example of YOLO object detection in a traffic scene is shown in figure 1.

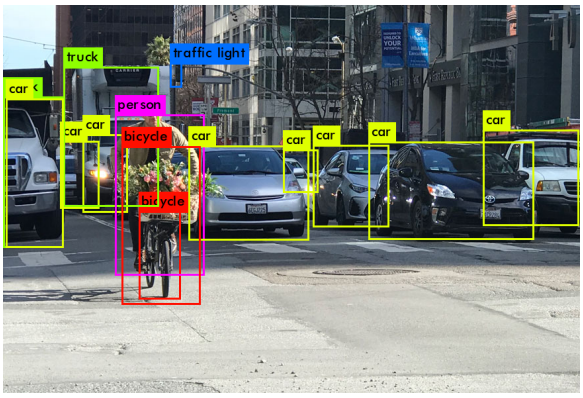

Figure 1. Example of YOLO object detection in a traffic scene (Hui, 2018).

Although the YOLO object detection accuracy and performance is still improving, a number of challenges remain to be addressed, e.g. low image resolution (including low light conditions) and overlapping content in the image (*occlusion*) (Diwan et al., 2023). In general, to address the issue of decreased object detection performance under low light conditions, two approaches can be observed:

1. Enhancing the low light input image using a specialized image enhancement algorithm and/or modifying the YOLO network architecture, e.g. (Wang et al., 2023; Yi et al., 2022; Yin, Yu, Fei, Lv, & Gao, 2023);

2. Combining the visible (RGB) input image information with information from a thermal image (Infrared) into a multi-spectral image, e.g. (Nataprawira et al., 2021; Wei et al., 2023).

In this project, which is part of the course "Research Methods for Artificial Intelligence" of the Open University, the effect of low light conditions on pedestrian detection using YOLOv8 will be investigated. To investigate if the pedestrian detection performance under low light conditions can be improved, the information from infrared images will be combined with optic color (RGB) images (approach 2 above). Next to this we will also try to quantify the effect of pedestrian occlusion, which was mentioned as the second challenge for the YOLO framework. To guide the project the following research questions have been formulated:

A. Does the additional information from infrared images increase the YOLOv8 pedestrian detection performance under low light conditions, as compared to using regular color (RGB) images alone?

B. When adding information from infrared images, does the pedestrian detection performance increase for low light conditions where pedestrian occlusion occurs?

This report is structured as follows: the Methods section will describe the dataset, the data preprocessing steps, the model, as well as the performance metrics. The Results section will present the results, and the Discussion section will discuss the results as well as draw the conclusion.

Methods

Dataset

The dataset used in this project is the KAIST (Korea Advanced Institute of Science and Technology) Multi-spectral Pedestrian Detection dataset (Hwang, Park, Kim, Choi, & So Kweon, 2015). It combines both images taken by a regular optic color camera and a thermal camera. The images were collected from a car driving the same route during day and night. The KAIST dataset consists of 95.328 color-thermal image pairs which were manually annotated with the labels *person*, *people* or *cyclist*. Each image has a size of 640 x 512 pixels. Next to the bounding box, the annotation also includes an occlusion tag.

In Figure 2 some examples of day and night traffic scenes taken from the dataset are shown. The left column shows the color camera images, and the thermal camera images are shown in the right column. The green, yellow, and red bounding boxes indicate *no-occlusion*, *partial occlusion*, and *heavy occlusion* respectively. The dataset can be downloaded from Hwang (2015).

Model

Training deep neural networks consumes a lot of computational resources, which is expensive, and in order to balance this with the resources available and scope of this project it was chosen to use the latest and smallest, pre-trained version of the YOLO model, YOLOv8 nano. More information on this model can be found at Ultralytics (2023).

Preprocessing

Before the YOLO model can be trained on the images, the image dataset was prepared using the steps described below:

1. **Image Blending** Both the RGB and infrared (IR) images were first converted from the RGB color space into the CIELAB color space. This step splits the color information into three separate channels (L,a,b). The IR images contain no color information (a,b), only lightness (L) information. After normalization of the lightness channels, the information from both images was combined and converted back into the RGB color space. For the conversion operations the OpenCV library was used. An example of RGB and IR image blending is shown in Figure 3.

Figure 2. Some example images from the KAIST dataset (Hwang et al., 2015).

2. **Label Conversion** The annotation type of the KAIST dataset is not in the official Pascal VOC XML format, and could not directly be converted into the annotation format required by the YOLOv8 model. Therefore, the annotations were first converted into the COCO format by using a customized VOC2COCO converter, and subsequently converted into the YOLOv8 format using the Python Ultralytics module's `convert_coco()` method.

3. **Folder Structure** To be able to train the YOLO model, the image data needs to be organized according to a specific folder structure following the DOTA dataset. This structure is shown in Figure 4. After the dataset split into training, validation and test sets (see next section) the images are moved into the appropriate directories.

Train and Test split

In total the dataset consists of 95.328 pairs of images taken both during day and night traffic situations. A major part of these images do not contain a *person* or *people* label, and

(a) RGB (b) IR (c) RGB + IR

Figure 3. Example of RGB and IR image blending.

```
|-- DOTA
|   |-- images
|   |   |-- train
|   |   |-- val
|   |   |-- test
|   |-- labels
|   |   |-- train
|   |   |-- val
|   |   |-- test
```

Figure 4. Image folder structure.

are not included in either the training or testing dataset. Only 39,039 images contain the label *person* or *people*. From these images, 80% was used for training and 20% was used for testing, and from the training set, 80% was used for training and 20% was used for model validation.

To be able to compare the results, two separate training, validation and test sets were created. First, a control set containing only the original RGB images from the optic camera, and second, a blended set which contains the same images as the control set, but with the information from the IR camera blended into the L channel. The resulting image counts in the training, validation and testing sets are summarized in Table 1.

		Training	**Validation**	**Testing**
I.	Control set (RGB only)	24.984	6.247	7.808
II.	Blended set (RGB + IR)	24.984	6.247	7.808

Table 1
Image counts for the training, validation and testing sets.

Model Training

The YOLOv8n model training and testing was run on the Microsoft Azure ML studio environment. We used a batch size of 32, and model training was run for 60 epochs.

Model Evaluation Metrics

After training and validation, the model performance was evaluated using the metrics precision, recall, mAP50 and mAP50-95. The definition of precision are recall are given as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

In object detection, true positives (TP) are detected objects of the right class with a certain minimal area overlap (often > 0.5) with the ground truth bounding box. A false positive (FP) is predicting an object, but with an overlap lower than the minimal threshold, and a false negative (FN) is failing to detect an object where there is one.

For an ideal classifier both precision and recall should be as close to 1 as possible. When classification is done for varying threshold levels, a so called PR curve can be obtained. This is shown in schematically Figure 5.

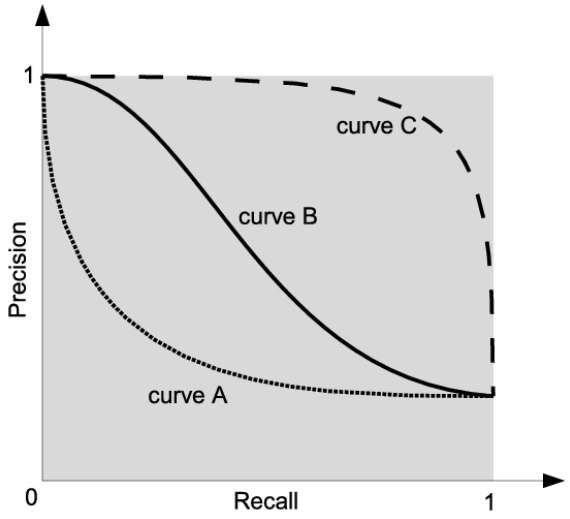

Figure 5. PR curves: (a) poor, (b) better, (c) good (Salfner, Lenk, & Malek, 2010).

The remaining two performance metrics, mAP50 and mAP50-95, are defined as:

mAP50 (*Mean Average Precision*) The mAP50 is equal to the average of the Average Precision (AP) metric across all object classes in a model. It is a metric to test the performance of the model on objects that have an Intersection over Union (IoU) threshold of at least 0.50. The IoU quantifies how much the predicted bounding box of an object deviates from the ground truth bounding box. The IoU and some example numbers are shown in Figure 6.

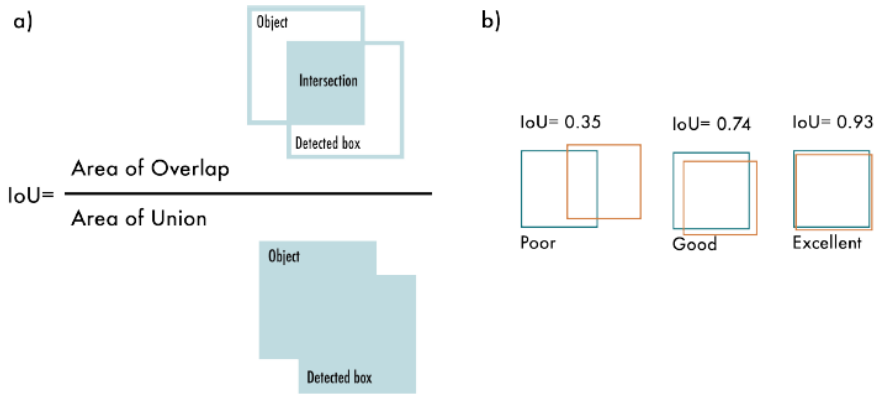

Figure 6. (a) IoU and (b) some examples numbers (Terven & Cordova-Esparza, 2023).

mAP50-95 This metric is the same as the mAP50, but with multiple IoU threshold levels varying from 0.50 to 0.95. The metric gives insight into the detection performance of the model over the different object classes.

Project Code

The Python code and Jupyter notebook files used during the project, as well as all the generated plots can be found on the project's GitHub page Verschelde (2024).

Results

Pedestrian Detection (Visual Inspection)

In Figure 7 some examples of YOLOv8 detected bounding boxes for daytime images are shown. The top row of the Figure shows bounding boxes for the control set, which consists of the original RGB images alone. The bottom row shows the bounding boxes for the blended set, where the information from the IR image was blended with the RGB image.

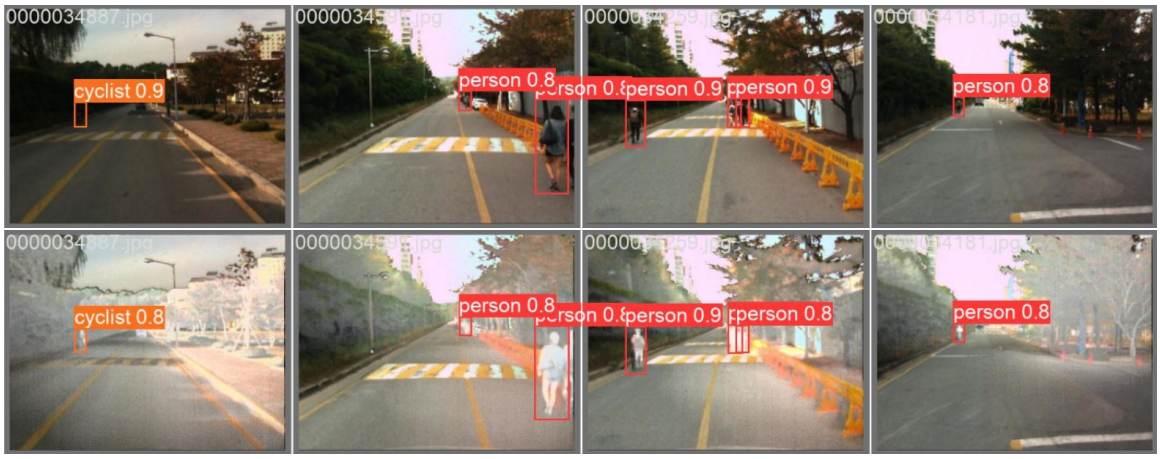

Figure 7. [Day] Example bounding boxes for control (top) and blended images (bottom).

Comparing the control and blended images shows that adding IR information for daytime images tends to overexpose the image, and results in a very bright image, which reduces the contrast between the target objects (*person*, *people*, and *cyclist*) and their surroundings. This effect can lower the model's confidence score for the object, as can be seen for example in the image on the left, and in the second image from the right.

In Figure 8 some examples of YOLOv8 detected bounding boxes for nighttime images are shown. The top row of the Figure shows bounding boxes for the control set, and the bottom row shows the bounding boxes for the blended set.

Figure 8. [Night] Example bounding boxes for control (top) and blended images (bottom).

Again, when comparing the two images sets, it is seen that the blended images appear brighter than the control images, but the effect is much smaller. In this case, the contrast between the target objects (*person*, *people*, and *cyclist*) and their surroundings is increased, which makes them stand out more clearly. As is seen in the image on the left, this increased contrast leads to the detection of two *person* objects in the blended image, who were not detected in the control image.

Pedestrian Detection (PR Curves)

The hypothesis underlying the first research question is that additional information from an IR camera image will increase the pedestrian detection performance of the YOLOv8 model as compared to the detection performance when using RGB camera information alone. To test this assumption the PR curve for the control and blended test set was generated for both images with the annotation *person* and *people*. The result, for images taken during both day and night, is shown in Figure 11 (see Appendix). As can be seen from the figure, for the *person* annotation, the PR curve for the control (blue) set cannot be discerned from the PR curve for the blended set (orange). For images with the *people* annotation, the PR curve for the blended set is only slightly above the PR curve for the control set for a tiny fraction, most probably within the experimental error margin.

When the images taken during day and night are split, and the PR curves for daytime and nighttime images are plotted separately, the result as shown in Figure 9 is obtained. This figure shows that both for the annotations *person* and *people*, the PR curve for the

control set (blue) lies above the PR curve for the blended set (orange) during daytime, suggesting a better performance for the model on the control set during daytime. For the nighttime however, the situation is reversed, and the PR curve for the blended image set (orange) is above that of the control set (blue), suggesting a slightly increased performance of the model on the blended set during nighttime.

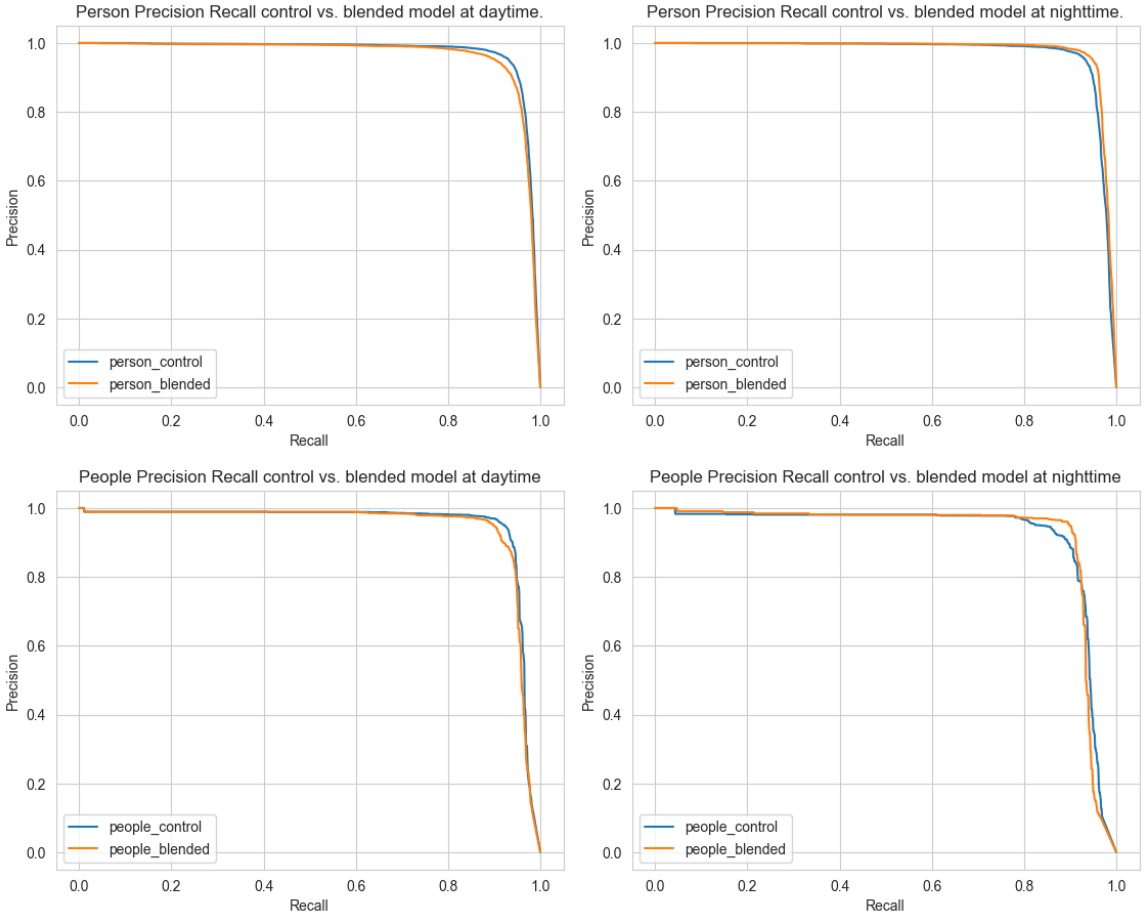

Figure 9. PR curves for daytime (left) and nighttime (right) images.

Pedestrian Detection (mAP50 and mAP50-95)

The mAP50 values for the different classes *person*, *people*, and *cyclist* have been plotted for both the control and the blended sets, both during daytime and nighttime. The result is shown in Figure 12 (see Appendix). The plot shows that for the daytime detection (left), and for all classes, the control set has an mAP50 value which is slightly higher than that of the blended model. When looking at the nighttime results (right), only for the class *person* does the blended model seem to have an mAP50 value which is slightly higher than that of the control set value.

In Figure 13 (see Appendix), the mAP50-95 values for both day- and nighttime situations are shown. The figure shows that in this case, for both day- and nighttime situations

the mAP50-95 value of the control set is always above that of the blended set, expressing that all predictions made for the control model are made with a higher confidence level for both day and night.

Occlusion (mAP50 and mAP50-95)

When the mAP50 and mAP50-95 performance of the model for the different occlusion levels of *no-occlusion*, *partial occlusion*, and *heavy occlusion* is plotted for both the control and blended sets, the result which is shown in Figure 10 appears.

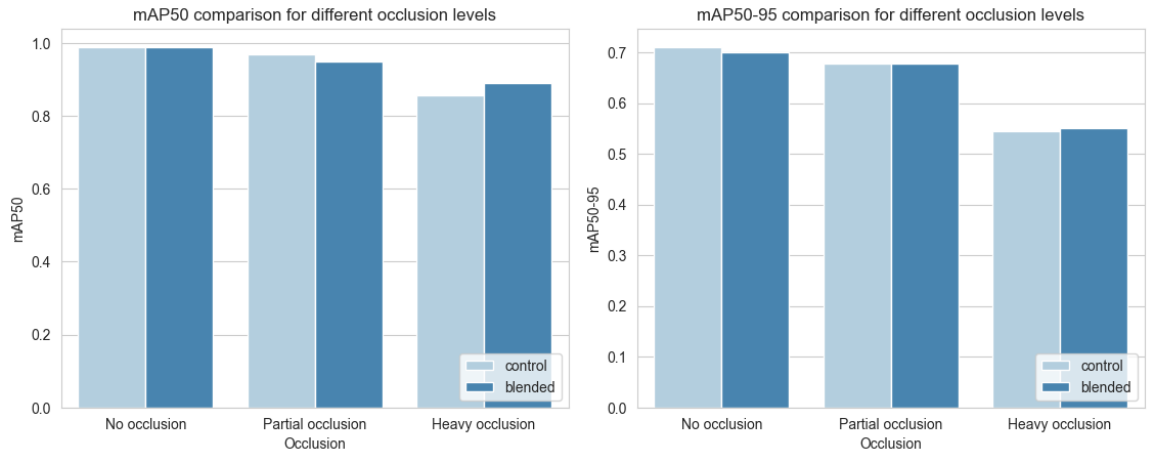

Figure 10. mAP50 plot (left) and mAP50-95 plot (right) for different occlusion levels.

For the mAP50 plot, only for *heavy occlusion* the mAP50 value of the blended image set is higher than the value of the control set, for *no occlusion* and *partial occlusion* the value for the control set is equal or slightly higher. The same result is obtained for the mAP50-95 plot, where only in the case of *heavy occlusion* the blended set slightly seems to outperform the control set.

Discussion

After image data preprocessing, YOLOv8n model training, and evaluation of the results, the research questions formulated in the Introduction section can now be answered.

A. Does the additional information from infrared images increase the YOLOv8 pedestrian detection performance under low light conditions, as compared to using regular color (RGB) images alone?

As was shown in the Results section, when calculated over all images combined, so both during daytime and nighttime, no significant difference can be detected between the PR curves of the control and blended sets. However, when the PR curve is calculated separately for the daytime and nighttime image sets, the performance on the blended image set is slightly higher during the nighttime, whereas during the daytime the performance on the control set is higher. This result seems to be supported by the observation that the mAP50 value for the *person* class in the blended image set

during nighttime is slightly above that of the control set, while for the rest of the classes the mAP50 value of the control set is equal or higher.

B. When adding information from infrared images, does the pedestrian detection performance increase for low light conditions where pedestrian occlusion occurs?

As was also shown in the results section, only the mAP50 and mAP50-95 performance of the *heavily occluded* class is higher than that of the control set. For all other occlusion classes the performance of the control class is equal or higher. The result was obtained taken over the daytime and nighttime images combined, so calculating the performance on the separated sets might reveal a more differentiated picture.

References

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using yolo: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, *82*(6), 9243–9275.

Hui, J. (2018). *Real-time Object Detection with YOLO.* https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088. (Accessed 15-02-2024)

Hwang, S. (2015). *KAIST Multispectral Pedestrian Detection Benchmark.* https://soonminhwang.github.io/rgbt-ped-detection. (Accessed 20-02-2024)

Hwang, S., Park, J., Kim, N., Choi, Y., & So Kweon, I. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1037–1045).

Nataprawira, J., Gu, Y., Goncharenko, I., & Kamijo, S. (2021). Pedestrian detection using multispectral images and a deep neural network. *Sensors*, *21*(7), 2536.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Salfner, F., Lenk, M., & Malek, M. (2010). A survey of online failure prediction methods. *ACM Computing Surveys (CSUR)*, *42*(3), 1–42.

Sanders, R. L., Schneider, R. J., & Proulx, F. R. (2022). Pedestrian fatalities in darkness: What do we know, and what can be done? *Transport policy*, *120*, 23–39.

Terven, J., & Cordova-Esparza, D. (2023). A comprehensive review of yolo: From yolov1 and beyond. arxiv 2023. *arXiv preprint arXiv:2304.00501*.

Ultralytics. (2023). *Yolov8.* https://github.com/ultralytics/ultralytics. (Accessed 20-02-2024)

Verschelde, J. (2024). *JoachimVerscheldePersonal/multispectral-ped-detection: Experiment comparing the impact of multispectral images on pedestrian detection under low light conditions.* https://github.com/JoachimVerscheldePersonal/multispectral-ped-detection. (Accessed 20-02-2024)

Wang, J., Yang, P., Liu, Y., Shang, D., Hui, X., Song, J., & Chen, X. (2023). Research on improved yolov5 for low-light environment object detection. *Electronics*, *12*(14), 3089.

Wei, J., Su, S., Zhao, Z., Tong, X., Hu, L., & Gao, W. (2023). Infrared pedestrian detection using improved unet and yolo through sharing visible light domain information. *Measurement*, *221*, 113442.

Yi, K., Luo, K., Chen, T., & Hu, R. (2022). An improved yolox model and domain transfer strategy for nighttime pedestrian and vehicle detection. *Applied Sciences*, *12*(23), 12476.

Yin, X., Yu, Z., Fei, Z., Lv, W., & Gao, X. (2023). Pe-yolo: Pyramid enhancement network for dark object detection. In *International conference on artificial neural networks* (pp. 163–174).

Appendix

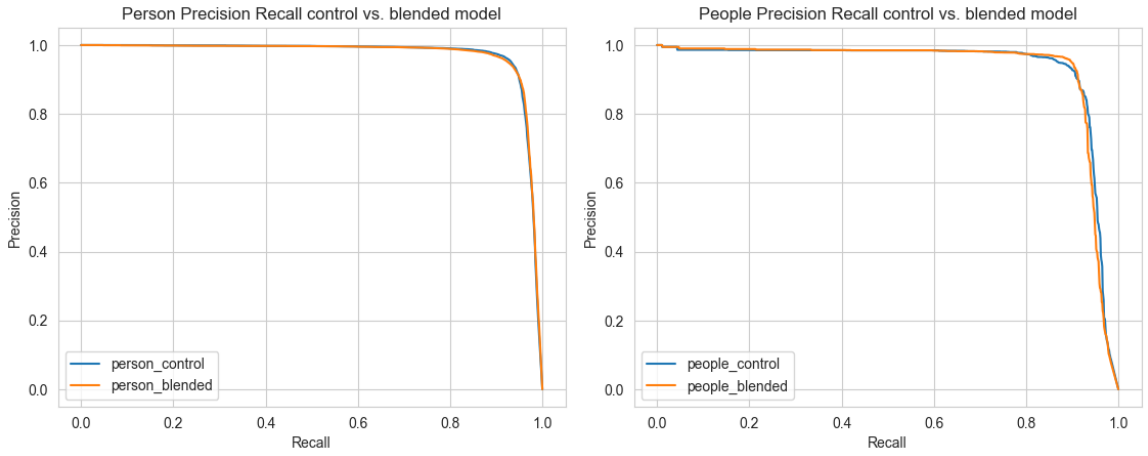

Figure 11. PR curves for the classes *person* (left) and *people* (right) taken over all images.

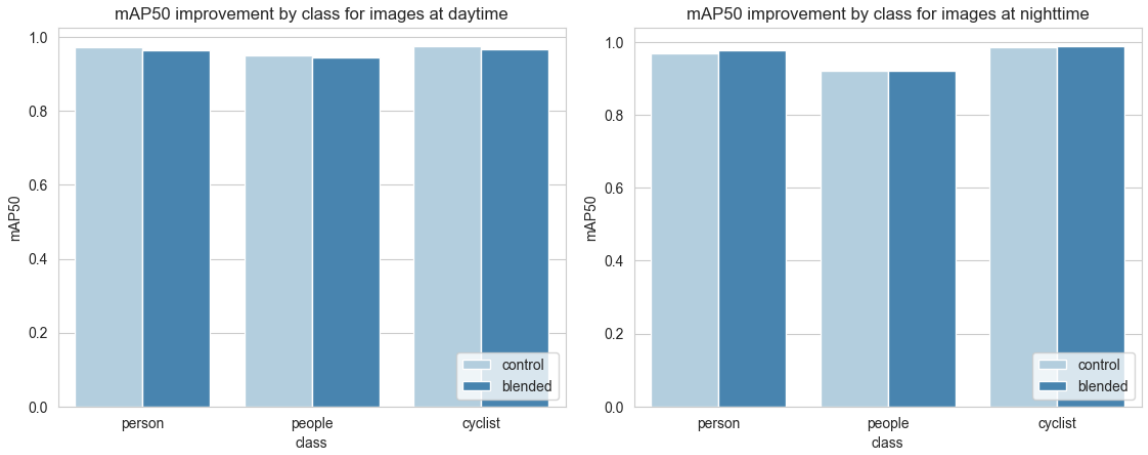

Figure 12. mAP50 plots for all classes during daytime (left) and nighttime (right).

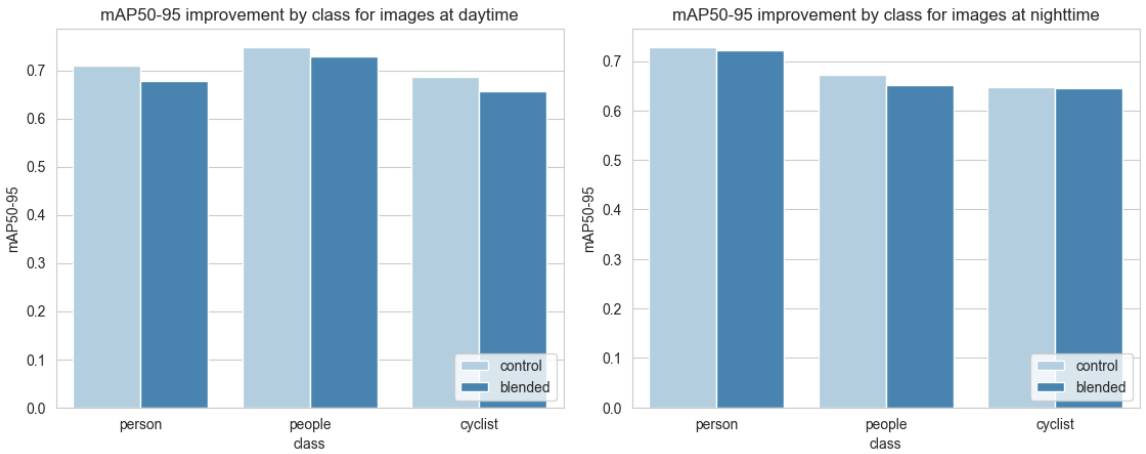

Figure 13. mAP50-95 plots for all classes during daytime (left) and nighttime (right).