Experiments

Dataset and Implementation Details

In this experiment we used two datasets to evaluate our model, which are MS COCO [1] and SODA-D [2]. We define our small pedestrians as having a bounding box height of 32 pixels or less.

The MS COCO dataset contains 80 categories of data, of which we selected pedestrian small targets walking on pedestrian streets, and in various scenes, from the person category. These scenes are rich, varied and well-suited for small target pedestrian detection tasks. We used 2000 extracted images with better features from the dataset.

The SODA-D dataset includes 24828 high-quality images under driving scenarios, on which 278433 instances of 9 categories with horizontal bounding boxes were annotated. We extracted 2000 of these images that contain pedestrians.

Combining the above data sets, the number or experimental samples is 4000. Before training, the data set is divided into a training set and a testing set with an 80/20 split, followed by further dividing the training data into a training set and a validation set with the same split.

The hardware and software configuration of the experimental platform was the following: we used the Windows 10 environment, an Intel Core i7-1180H CPU at 2.3 GHz, 8GBs of memory, an NVIDIA GeForce GTX 3060, 6G GPU, the CUDA 11.1, CUDNN 8.0 GPU accelerator, and finally the Tensorflow 2.4 deep learning framework.

Evaluation Metrics

We propose 5 indicators to evaluate the efficacy of our method in contrast with RetianNet [3], CornerNet [4], CenterNet [5] and YOLOX [6].

The first of these metrics being average precision (AP) to measure accuracy across two diferent intersection over union (IoU) thresholds: 0.25 and 0.5, using the expression

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}.$$

The following indicator requires the concepts of true positives (TP), which have both the correct category and their IoU is above the given threshold, false positives (FP), where the predicted category is incorrect and the IoU falls under the threshold, or a duplicate detection occurs, and finally false negatives (FN), having both the wrong class and missing a bounding box. Taking into account the precision and recall metrics:

$$Precision = \frac{TP}{TP + FP},$$

Method	AP_{25}	AP_{50}	Precision	Recall	F1-Score
Our Method	65.2	39.7	0.83	0.78	0.80
RetianNet	57.6	28.2	0.76	0.65	0.70
CornerNet	49.5	24.6	0.71	0.59	0.64
CenterNet	48.8	21.5	0.68	0.75	0.62
YOLOX	53.4	26.7	0.73	0.61	0.66

Table 1: *Comparative Test Results*

$$Recall = \frac{TP}{TP + FN},$$

we propose assigning F1-scores to measure the accuracy and sensitivity of our method, which represent the harmonic mean of the previous two:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}.$$

Results

The results presented in table 1 showcase that our method consistantly outperforms the baseline detectors in every metric chosen. In particular, it achieves the greaters values for both AP_{25} and AP_{50}, indicating enhanced robustness under both lenient and strict localization requirements. The large improvement specifically at AP_{50} indicates that the suggested approach generates more precise bounding box predictions, which is particularly important in small-object identification tasks like the one proposed, where performance can be greatly impacted by even modest localization errors.

Furthermore, our method achieves the highest precision, recall, and F1-score, indicating a better balance between false positives (FP) and false negatives (FN) than the compared methods. This balanced improvement across measures shows that our model retains higher classification reliability while also detecting a higher percentage of relevant tiny objects.

Overall, the results of our experiments demonstrate that our proposed method outperforms current detectors in both detection accuracy and localization quality, offering a more efficient and dependable alternative for small object identification.