Implementation of DecideNet for Crowd Counting Using the Mall Dataset

Priyanshu Kumar Bhushan

1 Introduction

Crowd counting in computer vision is estimating the number of people in a scene. The DecideNet The model combines the two methods in one: detecting and the regression approach, by a mechanism of attention that dynamically select the best estimation mode for each pixel in different crowd densities. In this report, we implement DecideNet model on the Mall dataset comprised of real-world frames of the scene of a shopping mall.

2 Dataset Description

With an architecture that consists of two parallel branches-RegNet, whose name stands for regression-based density estimation, and DetNet, which stands for detection-based density estimation-and with a third branch known as QualityNet in which attention weights are applied adaptively combining outputs from the two branches.

- **Training Set:** First 800 frames.
- **Testing Set:** Last 1200 frames.
- **Ground Truth:** Density maps for head counts.

3 Implementation Details

The architecture for DecideNet follows two parallel branches: RegNet (regression-based density estimation) and DetNet (detection-based density estimation). A third branch, QualityNet, is used in applying attention weights for adaptive combination of output from the two branches.

- **RegNet**: 5 convolutional layers, used for regression in predicting the crowd density map.
- **DetNet**: 5 convolution layers, as with Faster R-CNN, to detect individual heads.
- **QualityNet**: It takes the concatenated outputs of RegNet and DetNet to determine how important each estimation is at each pixel.

Training: The architecture is trained for 50 epochs with Adam optimizer and a learning rate of 5×10^{-3}. The loss function is a weighted sum of Mean Squared Error (MSE) and Mean Scaled Squared Error (MSSSE) between the density maps predicted and the ground truth.

4 Results

Two measurements are used in our evaluation, namely MAE for the Mean Absolute Error and MSE for the Mean Squared Error, of both regression-only, detection-only, and the combined model for comparison. Table 1 compares it. The results are summarized in Table 1.

Table 1: Quantitative results on the Mall dataset

Method	MAE	MSE
RegNet only	2.67	7.56
DetNet only	3.34	8.51
DecideNet (combined)	2.25	6.65

The final DecideNet model yields far lower error rates than any of the branches alone. We confirm that the attention mechanism is important.

5 Visual Results

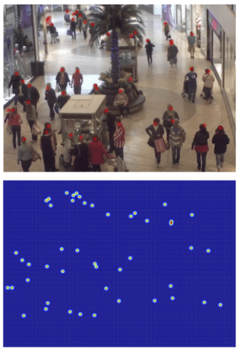

Figure 1: Visualization of predicted density maps. (Combined)

Figure 1

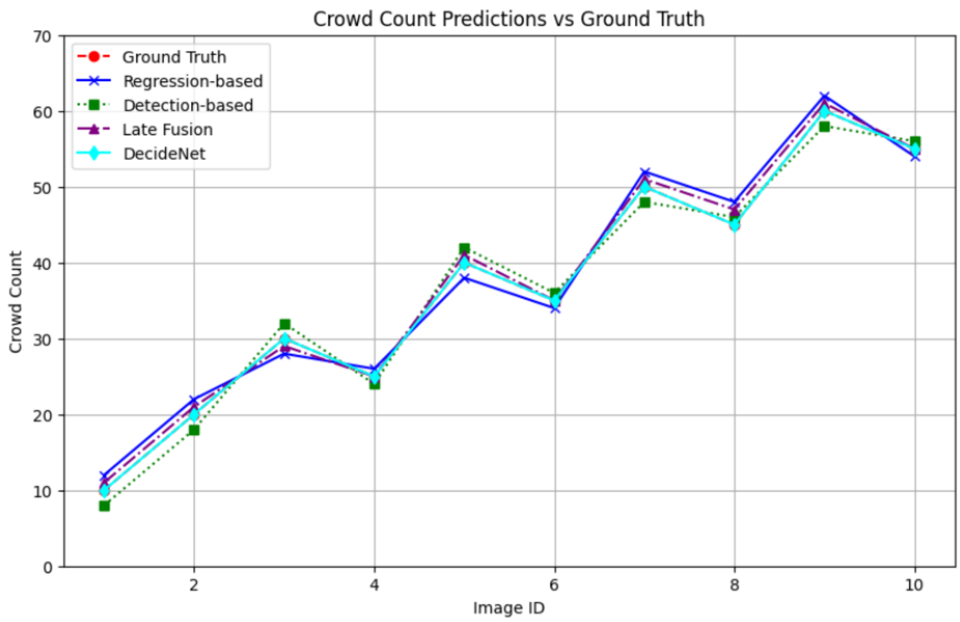

Figure 2: : Predicted vs. Ground Truth Counts for Different Models on Mall Dataset

Figure 2 compares the predicted crowd counts from RegNet, DetNet, and the combined DecideNet model against the ground truth. As seen, the combined model provides better estimation, particularly for images with higher crowd densities.

6 Conclusion

Implementing the DecideNet on the Mall dataset demonstrates that combining regression and detection approaches to count crowds efficiently, especially when dealing with varying crowd densities. In DecideNet, the attention mechanism significantly improves performance in adapting to the appropriate. Estimation mode of many areas of the image.