Implementation of DecideNet for Crowd Counting Using the Mall Dataset

Priyanshu Kumar Bhushan

November 1, 2024

1 Introduction

Crowd counting in computer vision is the task of estimating the number of people in a scene. The DecideNet model combines two methods in one: detection and regression approaches, by employing an attention mechanism that dynamically selects the best estimation mode for each pixel in varying crowd densities. This report details the implementation of the DecideNet model on the Mall dataset, which comprises real-world frames from a shopping mall scene.

2 Dataset Description

The DecideNet architecture consists of two parallel branches: RegNet, which is regression-based for density estimation, and DetNet, which is detection-based. A third branch, known as QualityNet, applies attention weights, 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:** Contains 5 convolutional layers, used for regression in predicting the crowd density map.
- **DetNet:** Contains 5 convolutional layers, similar to Faster R-CNN, to detect individual heads.
- **QualityNet:** Takes concatenated outputs of RegNet and DetNet to assess the importance of each estimation at each pixel.

Training: The model is trained for 50 epochs using the Adam optimizer with 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 predicted density maps and the ground truth.

4 Results

Two metrics, Mean Absolute Error (MAE) and Mean Squared Error (MSE), are used to evaluate the model's performance across regression-only, detection-only, and combined approaches. Table 1 summarizes the results.

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

Table 1: Quantitative results on the Mall dataset

5 Qualitative Results

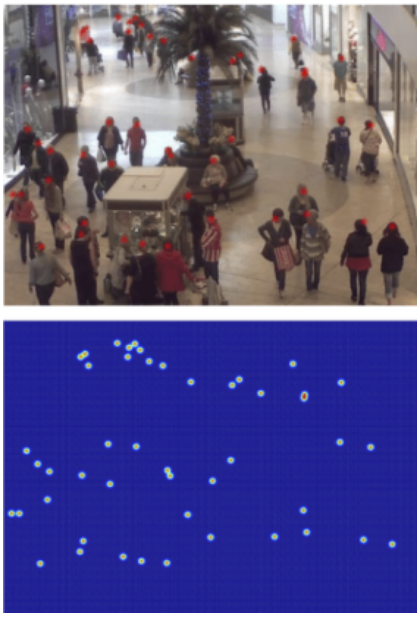
Index	**Model Component**	**MAE**	**MSE**
0	RegNet Only	0.024641	0.001237
1	DetNet Only	0.052045	0.004830
2	RegNet + DetNet (Late Fusion)	0.019088	0.000616
3	RegNet + DetNet + QualityNet	7.879723	104.196289

Table 2: Qualitative results of DecideNet components on the Mall dataset

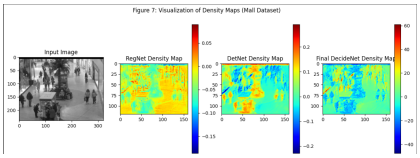
6 Visual Results and Plot

7 Conclusion

The implementation of DecideNet on the Mall dataset demonstrates that combining regression and detection approaches can effectively count crowds, especially in scenarios with varying crowd densities. The attention mechanism in DecideNet significantly enhances performance by adapting the estimation mode across different regions of the image.



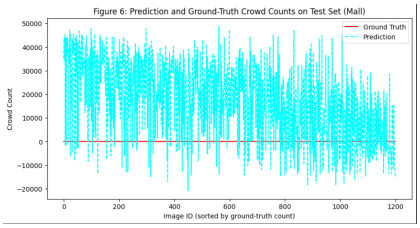
(a) Visualization of predicted density maps (Figure 1.1)



(b) Visualization of predicted density maps (Figure 1.2)

Figure 1: Visualization of predicted density maps

(a) Predicted vs. Ground Truth Counts (Figure 2.1)



(b) Predicted vs. Ground Truth Counts (Figure 2.2)

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