Few-Shot Scene Adaptive Crowd Counting Using Meta-Learning

Mahesh Kumar Krishna Reddy†, Mohammed Asiful Hossain‡, Mrigank Rochan†, and Yang Wang†

†Department of Computer Science, University of Manitoba
‡ Huawei Technologies Co. Ltd.

Overview

Traditional Models: They require large number of labeled data to achieve a successful model. However, for applications like crowd counting, collecting large amount of labeled data or annotating every camera image is expensive, or cumbersome.

Meta-learning: It enables to exploit the adaptable scene representation to learn a new camera scene (task) with limited data.

Problem Setup

Top row: During training, we have access to a set of $N$ different camera scenes where each scene comes with $M$ labeled examples. From such training data, we learn the model parameters $\theta$ of a mapping function $f_\theta$ such that $\theta$ is generalizable across scenes in estimating the crowd count.

Bottom row: Given a test (or target) scene, we assume that we have a small number of $K$ labeled images from this scene, where $K \ll M$ (e.g., $K \in \{1, 5\}$) to learn the scene-specific parameters $\tilde{\theta}$ with the help of meta-learning guided approach we quickly adapt $f_\theta$ to test scene-specific parameters $\tilde{\theta}$ that predicts more accurate crowd count than other alternative solutions.

![Few-shot Scene Adaptive Crowd Counting](image)

Inner update:

$$\tilde{\theta}_i = \theta - \alpha \nabla \theta \mathcal{L}_{T_i}(f_\theta)$$

where

$$\mathcal{L}_{T_i}(f_\theta) = \sum_{(x^{(j)}, y^{(j)}) \in L_{T_i}} \| f_\theta(x^{(j)}) - y^{(j)} \|^2$$

Outer update:

$$\theta = \theta - \beta \nabla \theta \sum_{i=1}^N \mathcal{L}_{T_i}(f_\theta)$$

![Meta-learning comparison results between different optimization based approaches](image)

Experiments

- **Datasets**: WorldExpo'10 [Zhang et al.] and UCSD [Chan et al.], and Mall [Loy et al.]
- **Metrics**: Mean Absolute Error, and Root Mean Squared Error
- **Crowd baselines**: Pre-trained, Fine-tuned, and Meta pre-trained
- **Meta baselines**: Meta-LSTM, and Reptile

![Crowd counting results on WorldExpo'10, Mall, and UCSD datasets](image)

![Meta-learning comparison results between different optimization based approaches](image)

![Comparison of results on the WorldExpo'10 dataset with K = 1 images in the target scene with Hossain et al. [1].](image)

Analysis

![Crowd counting performance comparison between the baselines and our approaches in different scene-specific images from WorldExpo'10 dataset](image)

References