Single Domain Generalization for Crowd Counting: MPCOUNT

Aayush Gupta (210020)

1 Implementation

A brief overview of the implementation of the algorithm is written below:

- The model employs a VGG16-BN architecture to extract features from both original and photometrically altered images.
- A single Attention Memory Bank (AMB) is used to store domain-invariant features for density estimation, reconstructing features via an attention mechanism, unlike previous approaches that utilize multiple memory banks.
- The Content Error Mask (CEM) filters out domain-specific content, ensuring the AMB focuses on domain-invariant features by measuring discrepancies in instance-normalized features.
- To address label ambiguity, the Patch-wise Classification (PC) task classifies small image patches (16x16) as containing human heads or not, improving crowd density prediction accuracy.
- The model's training incorporates multiple loss functions, including density regression loss, PC loss, and attention consistency loss (ACL), guiding the learning of domain-invariant features and improving generalization to unseen domains.

2 Results

Model trained on	**Testing on**	**Results by this code (MAE, MSE)**	**Result in paper (MAE, MSE)**
ShanghaiTech Part A	ShanghaiTech Part B	11.39, 19.73	11.4, 19.7
ShanghaiTech Part B	ShanghaiTech Part A	102.60, 183.04	99.6, 182.9
UCF-QNRF	ShanghaiTech Part A	64.89, 108.80	65.5, 110.1
UCF-QNRF	ShanghaiTech Part B	12.62, 24.65	12.3, 24.1

Table 1: Results Comparison

Observations: The results are very close to as per stated in the research paper.

For testing purpose we can get the count of count along with density map prediction of any image:

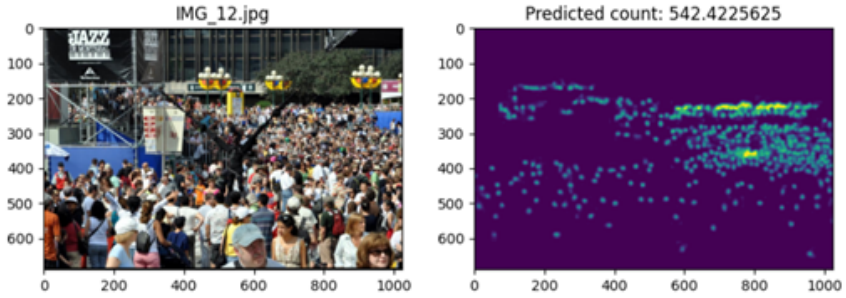

Figure 1: Example Image for Analysis

We can also analyse ground truth density map:

Figure 2: Example Image for Analysis

3 Dataset Description

MPCount was evaluated using several well-known crowd counting datasets:

- **ShanghaiTech Part A (SHA)**: Highly congested scenes collected from the internet.
- **ShanghaiTech Part B (SHB)**: Street-level scenes with lower crowd densities.
- **UCF-QNRF**: A large-scale dataset with diverse scenes, crowd densities, and lighting conditions.