

Crowd Counting in Frequency Domain

Rohan Guin

October 2024

1 Dataset

The ShanghaiTech dataset is a large-scale dataset for crowd counting, comprising 1,198 annotated crowd images. It is divided into two parts: Part A contains 482 images, and Part B includes 716 images. Part A is further split into 300 training images and 182 testing images, while Part B is divided into 400 training images and 316 testing images. Each individual in the crowd is marked with a point near the center of their head. In total, the dataset contains annotations for 330,165 people. The images in Part A were sourced from the Internet, while those in Part B were captured on the busy streets of Shanghai.

2 Implementation Details

The crowd counting framework used in this work operates in the frequency domain, which is a novel approach compared to traditional methods that focus on the spatial domain. The core idea is to transform the density map from the spatial domain into the frequency domain using characteristic functions, allowing for a more compact and organized representation of spatial information.

2.1 Model Architecture

The architecture consists of the VGG19 network for feature extraction, followed by a regression module made up of three convolutional layers. The key innovation in the model is the loss function, which operates in the frequency domain. The characteristic functions of both the ground-truth and predicted density maps are computed, and the L1 norm be-

tween these functions is used as the loss metric. By leveraging the properties of the characteristic function, this method avoids the need for external algorithms like the Sinkhorn or Hungarian algorithms, making the approach simpler and more efficient.

2.2 Training Process

The model is trained using the CHF loss function, which minimizes the discrepancy between the characteristic functions of the predicted and ground-truth density maps. The following configuration is used for training:

- **Optimizer:** Adam
- **Learning Rate:** 1×10^{-5}
- **Weight Decay:** 1×10^{-4}

Data augmentation techniques such as image resizing and cropping are employed during training, with specific crop sizes for different datasets:

- **ShanghaiTech A:** 128
- **ShanghaiTech B:** 512

3 Results

In the original study, five datasets were utilized: NWPU, JHU++, UCF-QNRF, ShanghaiTech A (SHTC A), and ShanghaiTech B (SHTC B). However, due to GPU limitations and to reduce the training and testing time, the experiments in this work were conducted specifically on the ShanghaiTech

dataset, focusing on its two subsets: ShanghaiTech A (SHTC A) and ShanghaiTech B (SHTC B).

Table 1: MSE and MAE values for SHTC A and SHTC B for ChfL model.

	MSE	MAE
SHTC A	102.3	70.5
SHTC B	27.4	21.8

As shown in Table 1, The performance comparison between the datasets SHTC A and SHTC B shows a clear difference in how well the model works. For SHTC A, the Mean Squared Error (MSE) is 102.3 and the Mean Absolute Error (MAE) is 70.5, indicating that the model makes a lot of mistakes in its predictions. This suggests that SHTC A has more challenges, such as crowded scenes or varying conditions. On the other hand, SHTC B has much lower values, with an MSE of 27.4 and an MAE of 21.8, meaning the model performs much better here, making predictions that are closer to the actual counts.

the actual crowd distribution, and two reconstructed maps created from a characteristic function with limited ranges. The first reconstructed map (b), restricted to $[0.3, 0.3]^2$, The second reconstructed map (c), limited to $[0.5, 0.5]^2$.

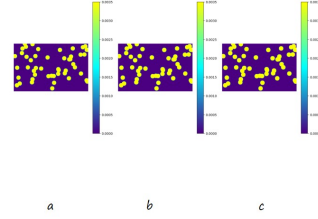


Figure 2: Comparison between the original density map and the reconstructed density map from the characteristic function confined to a small range. (a) The original density map; (b) The reconstructed density map from its characteristic function truncated on $[0.3, 0.3]^2$, and on (c) $[0.5, 0.5]^2$.

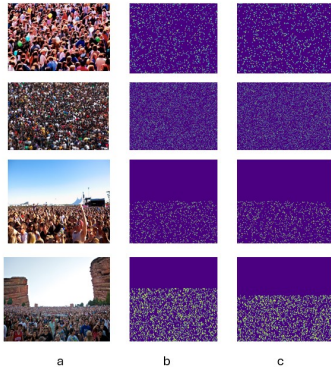


Figure 1: Crowd counting: (a) Original Image, (b) Ground Truth Density, (c) Predicted Density Map.

In Figure 1, we compare the ground truth density map with the predicted density map generated through model evaluation using the ChfL loss function.

In Figure 2, This comparison looks at three density maps: the original density map (a), which shows