PROJECT ABSTRACT

Crowd Counting using CSRNet

CSRNet-based estimator that counts people in dense crowd photographs via density-map regression, trained on the ShanghaiTech dataset.

01 Title

Crowd Counting using CSRNet — density-map based estimation of the number of people in dense crowd photographs using the Congested Scene Recognition Network, trained on the ShanghaiTech Crowd Counting dataset.

02 Introduction / Background

Counting people in dense crowds matters for public safety, event management, retail analytics and transport planning, but manual counting is unreliable and person-detection methods collapse when heads overlap. Crowd counting is therefore framed as density-map regression: the network predicts a density map over the image, and the sum of the map gives the head count. CSRNet, a well-known architecture for this task, pairs a VGG-16 front-end for feature extraction with a dilated-convolution back-end that preserves spatial resolution while capturing broad context.

03 Problem Statement

Given a single crowd photograph, estimate the number of people present, including in highly congested scenes where individuals occlude one another. The system must handle varying head sizes, perspective distortion and scenes ranging from sparse streets to dense gatherings.

04 Objectives

- Implement CSRNet: VGG-16 front-end (first 10 layers) plus a dilated-convolution back-end.
- Generate ground-truth density maps from head-point annotations using Gaussian kernels.
- Train on the ShanghaiTech dataset (Part A: 482 congested images; Part B: 716 street scenes).
- Evaluate with MAE and MSE computed by the included notebook on the validation split.
- Provide a demo application showing the predicted count with a density heatmap overlay.

05 Proposed System / Working

Head-point annotations are converted to ground-truth density maps with Gaussian kernels. CSRNet is trained with a Euclidean (L2) loss between predicted and ground-truth maps on cropped patches. At inference, the input photo is resized to a multiple of 8, passed through the network, and the predicted density map is summed to yield the head count. The demo overlays the density map as a heatmap on the original photo so the count is visually verifiable.

06 Technologies / Components Used

Python 3.10, PyTorch and torchvision (VGG-16 backbone); OpenCV, NumPy and SciPy for annotation and density-map processing; Matplotlib for heatmaps and training plots; a web demo application for photo upload and counting; the ShanghaiTech Crowd Counting dataset; trained CSRNet weights as PyTorch checkpoints.

07 Key Features

- Density-map regression instead of per-person detection — robust in dense crowds.
- CSRNet with dilated convolutions preserving spatial detail.
- Head-count output plus density heatmap overlay in the demo.
- MAE and MSE evaluation computed by the notebook on the validation split.
- Multi-scale augmentation for varying head sizes and camera distances.
- Preprocessing scripts for annotation-to-density-map conversion.

08 Applications / Use Cases

- Event and venue crowd estimation.
- Public-space and transport-hub occupancy analysis.
- Retail footfall analytics.
- Safety-oriented crowd monitoring research (prototype, not certified).

09 Results & Scope

The included notebook computes MAE and MSE on the validation split during the buyer's build; expected pattern is lower error on the sparser Part B than the congested Part A — no metric numbers are presented as measured facts. Known scope limits: extreme density saturates the density map; occlusion and perspective distortion are the main error sources; single-frame counting only (no video tracking); and the build is an educational prototype, not a surveillance-certified system.

10 Conclusion

The project delivers a complete, demoable crowd-counting pipeline — density-map ground-truth generation, CSRNet training, MAE/MSE evaluation and an interactive counting demo — packaged with report, presentation and viva material, suitable as a B.E./B.Tech final-year project in Computer Science, AI/ML and Data Science.

11 References

CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes (Li et al., CVPR 2018, arXiv:1802.10062); ShanghaiTech Crowd Counting dataset (Zhang et al., CVPR 2016); VGG-16 (Simonyan & Zisserman, arXiv:1409.1556); PyTorch documentation.