DEEP VISION CROWD MONITOR: AI FOR DENSITY ESTIMATION AND OVERCROWDING DETECTION

SUBMITTED BY:

AVANTHIKALAKSHMI VINOD

Table of Contents

1.Project Description

2.Dataset Used

3.Environment Setup

4.Data Exploration

5.Data Preprocessing

6.Conclusion

Project Description

The Deep Vision Crowd Monitor project focuses on the development of an intelligent system capable of estimating crowd density and identifying overcrowded regions within an image. Such systems are widely used in public safety, event management, transportation hubs, and smart city monitoring. Using AI-based density estimation techniques, the goal is to convert raw images into meaningful density maps that highlight areas of high and low crowd concentration.

This project involves understanding various components of the crowd counting pipeline such as environment setup, dataset preparation, preprocessing, and conversion of images and ground truth annotations into a form suitable for training deep learning models.

Dataset Used

The dataset used for this project is the **ShanghaiTech Crowd Counting Dataset**, a widely adopted benchmark for crowd density estimation tasks.

The dataset consists of two parts:

Part A

- Contains highly dense crowd scenes
- Images collected from the internet
- Each image has a ground-truth .mat file containing head annotations

Part B

- Contains moderately dense crowd images
- Captured from streets in Shanghai
- Also contains .jpeg images and .mat annotation files

Each annotation file contains head positions, which are later converted into density maps using Gaussian kernels.

Environment Setup

To begin working on the project, a suitable Python development environment must be set up. This includes installing essential libraries, configuring Jupyter Notebook, and verifying paths.

Key Steps:

- Install Python (recommended version 3.9+)
- Install required libraries: numpy, opencv-python, torch, torchvision, scipy, matplotlib, tqdm
- Install Jupyter Notebook for running and documenting experiments
- Set up dataset directory structure
- Validate file paths for images and ground-truth annotations

The environment setup ensures that the system can load images, process data, and visualize results smoothly.

Data Exploration

Data exploration helps in understanding the structure and contents of the dataset.

Exploration includes:

- Checking the number of images in Part A and Part B
- Visualizing sample images
- Inspecting .mat files to observe how head annotations are stored
- Verifying image shapes and formats

Through exploration, we ensure that:

- Images are readable
- Ground-truth annotation files match image names
- Dataset follows expected structure

Exploring data early helps catch mistakes such as missing files, corrupted images, or mismatched annotations.

Data Preprocessing

Preprocessing is one of the most important stages in crowd counting. It prepares both images and ground truth maps for training deep learning models.

Preprocessing Steps:

1. Image Loading and RGB Conversion

All images are loaded using OpenCV and converted from BGR to RGB format to match deep learning model expectations.

2. Normalization (ImageNet Standard)

To match VGG/ResNet-based backbones, images are normalized using:

- Mean: [0.485, 0.456, 0.406]
- Std: [0.229, 0.224, 0.225]

3. Generating Density Maps

Head annotations from .mat files are used to generate density maps using Gaussian filters. This produces high-resolution density images representing crowd distribution.

4. Downsampling Density Maps by 8×

Deep networks reduce image resolution internally, so ground-truth density maps are downsampled by a factor of 8 in both width and height.

5. Preserving Crowd Count (×64 Scaling)

Downsampling reduces the area by 64 (8×8). To maintain the total crowd count, the density map is multiplied by 64.

6. Converting Images and Density Maps into Tensors

Both images and density maps are converted into PyTorch tensors for model training.

- Image tensor shape: (3, H, W)
- Density tensor shape: (1, H/8, W/8)

7. Dataset and Dataloader Creation

A PyTorch Dataset class is constructed to:

- Load images
- Read ground truth annotations
- Apply normalization
- Generate density maps

A DataLoader batches these processed samples for efficient model training.

Conclusion

The *Deep Vision Crowd Monitor* project successfully demonstrates how artificial intelligence can be applied to understand and analyze crowd density from images. By utilizing the ShanghaiTech dataset and following a structured workflow—environment setup, data exploration, and detailed preprocessing—the project creates reliable inputs for modern deep learning models such as CSRNet and VGG-based architectures.

The preprocessing pipeline ensures that both images and ground-truth density maps are transformed into standardized, model-ready tensors while preserving the true crowd count. This stage forms the foundation for accurate crowd estimation and provides a scalable solution for real-world applications such as public safety, event control, and smart city monitoring.

Overall, the project builds a strong understanding of crowd counting fundamentals and prepares the platform for future stages, including model training, evaluation, and deployment.