# DEEPVISION CROWD MONITOR

## AI FOR DENSITY ESTIMATION AND OVERCROWDING DETECTION

Theory document-

Real-Time Crowd Monitoring

Submitted by:

Arti Kumari

# TABLE OF CONTENTS

## 1. Project Description

The **Deep Vision Crowd Monitoring System** is an AI-based solution designed to estimate and monitor crowd density in real-time surveillance environments. Crowd monitoring is an essential requirement in public spaces such as shopping malls, railway stations, airports, stadiums, and smart city infrastructures, where safety, crowd control, and efficient management are critical.

Traditional crowd counting techniques struggle in highly congested environments due to severe occlusion and overlapping individuals. To overcome these limitations, this project adopts deep learning-based approaches that can handle both sparse and dense crowds effectively.

The project is implemented in **three major tasks**. The first two tasks focus on real-time crowd monitoring using a webcam feed, where the system detects and counts people dynamically. The third task focuses on improving the accuracy and adaptability of the model by fine-tuning it using two different dataset strategies.

A hybrid architecture combining **YOLO (object detection)** and **CSRNet (density map regression)** is used. YOLO performs well when individuals are clearly visible, while CSRNet excels in congested scenes where individual detection is unreliable. By integrating these models, the system automatically adapts to varying crowd densities and delivers accurate results in real-world scenarios.

## 2.Dataset Used

Multiple data sources are used in this project to support different tasks and objectives.

For **real-time monitoring**, a live webcam feed is used as the input source. This simulates real-world surveillance conditions, including dynamic lighting, camera noise, and continuous scene changes.

For **model fine-tuning (Task 3)**, two dataset approaches are used:

1. **Video Dataset:**
   Surveillance videos are used, and frames are extracted to fine-tune the model. This dataset represents real-time crowd movement and temporal variations.
2. **Mall Frame Dataset:**
   This dataset consists of individual image frames with corresponding ground truth density annotations. It is commonly used for benchmarking crowd counting models and provides accurate spatial crowd information.

Using both datasets helps the model generalize better and perform reliably in both static and dynamic environments.

## 3. Task 1: Real-Time Crowd Monitoring Using YOLO

*(File: real_time_webcam_yolo_csrnet_hybrid.ipynb)*

Task 1 focuses on implementing a **real-time crowd monitoring system** using **YOLO (You Only Look Once)** as the primary model. YOLO is a deep learning-based object detection model capable of detecting people in images and video frames with high speed and accuracy.

In this task, video frames are continuously captured from a webcam. Each frame is passed through the YOLO model, which detects individuals by drawing bounding boxes around detected persons. The total crowd count is obtained by counting the number of detected persons in each frame.

This approach works efficiently in **low-density and moderately crowded scenes**, where individuals are clearly visible and separated. YOLO provides precise localization and fast inference, making it suitable for real-time applications.

However, Task 1 also highlights a limitation: as crowd density increases, person detection becomes less reliable due to overlapping individuals and occlusion. This limitation motivates the need for an improved approach, which is addressed in Task 2.

## 4. Task 2: Hybrid Crowd Monitoring Using YOLO and CSRNet

*(File: real_time_webcam_yolo_csrnet_hybrid.ipynb)*

Task 2 enhances the real-time monitoring system by introducing a **hybrid crowd counting approach**. Instead of relying solely on object detection, this task integrates **CSRNet (Congested Scene Recognition Network)** to handle high-density crowd scenarios.

CSRNet is a density-based crowd counting model that predicts a density map representing the spatial distribution of people in a scene. The total crowd count is obtained by summing the values in the predicted density map. This approach is highly effective in congested scenes where detecting individual people is difficult.

In this hybrid system, the crowd density is continuously analyzed. When the scene contains a small number of people, YOLO is used for fast and accurate person detection. As crowd density increases beyond a defined threshold, the system automatically switches to CSRNet.

The output includes:

- Real-time video feed
- Crowd count displayed on the screen
- Density heatmap visualization in dense scenes

This adaptive strategy improves accuracy, reduces detection errors, and ensures robust performance across different crowd conditions.

## 5. Task 3

## Approach 1: Fine-Tuning Using Video Dataset

*(File: task3_videoDataset_finetuning.ipynb)*

This task focuses on fine-tuning the CSRNet model using a **video-based dataset**. Video data closely represents real-world surveillance footage and captures temporal variations in crowd movement.

In this approach, frames are extracted from videos at regular intervals. These frames are used to fine-tune a pretrained CSRNet model. During training, the model predicts density maps for each frame, and the predictions are compared with ground truth density maps. The error is minimized through iterative updates of the model parameters.

Fine-tuning using video data helps the model adapt to dynamic environments and improves its performance during real-time deployment.

## Approach 2: Fine-Tuning Using Mall Frame Dataset

*(File: task3_mall_finetuning.ipynb)*

In this approach, CSRNet is fine-tuned using the **Mall dataset**, which consists of static image frames with corresponding ground truth density maps.

Each image is treated independently, allowing the model to focus on spatial crowd distribution patterns. The pretrained CSRNet model is further trained using these image-density map pairs, improving its ability to estimate crowd density in static scenes.

This approach provides stable training and serves as an effective benchmarking method, complementing the video-based fine-tuning strategy.

## 6. Conclusion

The **Deep Vision Crowd Monitoring System** demonstrates an effective application of deep learning techniques for real-time crowd estimation in surveillance environments. The project addresses key challenges such as varying crowd densities, occlusion, and real-time performance through a structured task-based approach.

In **Task 1**, YOLO was used for real-time person detection, providing accurate and fast crowd counts in low and moderately dense scenes. **Task 2** enhanced this system by introducing a hybrid approach that integrates CSRNet for density-based crowd estimation, enabling reliable performance even in highly congested environments. The automatic switching between models ensures adaptability and improved accuracy.

In **Task 3**, CSRNet was fine-tuned using both video-based and frame-based datasets. Video-based fine-tuning improved the model's robustness for continuous surveillance scenarios, while frame-based fine-tuning using the Mall dataset enhanced spatial accuracy. Together, these approaches strengthened the overall performance of the system.

Overall, the project presents a scalable and adaptable crowd monitoring solution suitable for real-world applications such as malls, public spaces, and smart city surveillance systems, with potential for further enhancement and deployment.