Crowd Counting Using Semi-Supervised Deep Learning: A Web-Based Approach

Bhuvan Rajesh(23MIA1020), Mohana Ramanan(23MIA1122), Shruthi Arunkumar(23MIA1168)
M.Tech (Int) CSE specialising in business analytics
Vellore Institute of Technology
Chennai, India

***Abstract*— Crowd counting represents a significant challenge within computer vision, with broad implications for public safety, urban planning, and event management. This paper presents a user-centric web application that estimates the number of individuals present in static images by leveraging a semi-supervised deep learning framework. Our approach combines Convolutional Neural Networks (CNN) and state-of-the-art Transformer architectures, employing recent advances in weakly supervised crowd counting. Experimental evaluations demonstrate that the developed system achieves competitive accuracy in practical scenarios while dramatically reducing annotation effort through the application of semi-supervised learning.**

Keywords*— *Crowd counting, digital image processing, semi-supervised learning, CNN, Transformer, web application, deep learning.

I. Introduction

The accurate estimation of crowd density is central to managing public spaces. Traditional fully supervised approaches for crowd counting require labour-intensive point-level annotations and do not scale well to high-density scenes. Recent research has focused on reducing the annotation burden through semi-supervised and weakly supervised techniques, achieving impressive results with only count-level labels. This project implements a web-based system for crowd estimation from still images, building upon these modern techniques.

Automated crowd density analysis allows authorities to detect high-density zones in real time, enabling timely interventions and resource allocation. Recent advances in computer vision have made it possible to estimate crowd size from images with improved accuracy and reduced manual effort.

A. Related work

Early crowd counting models relied on object detection or segmentation, which struggled with occlusion and perspective distortion. Density map regression using CNNs improved results but was dependent on detailed head annotations. Transformer-based models further improved the capture of global relationships in crowded scenes. Semi-supervised methods leverage unlabeled data with self-training or consistency regularization. Hybrid approaches, combining CNNs for local features and Transformers for global context, now set the standard for weakly supervised crowd counting.

II. Methodology

This project adapts the CrowdCCT framework, which combines DenseNet-121 for extracting local features and specialised Transformer modules for global aggregation. The model is trained using a combination of labelled and unlabeled images, leveraging pseudo-labelling and consistency constraints.

- **Image preprocessing:** Input images are resized and divided into patches.
- **Feature extraction:** DenseNet-121 captures local semantics; its outputs feed into Transformer modules.
- **Attention modules:** Multi-Scale Dilated Attention (MSDA) and Location-Enhanced Attention (LEA) aggregate features at multiple scales and locations, boosting accuracy in congested and variable settings.
- **Prediction:** Fully connected regression layers generate the final person count.

III. Implementation

The presented solution is deployed as an intuitive web application, enabling straightforward image uploads for analysis. Images submitted by users are processed by the server and evaluated by the trained neural network. The resulting crowd estimate is displayed in near real-time on the web interface.

- *Frontend*: Simple user interface supporting image upload and result presentation.
- *Backend*: Deep learning inference performed using Python (PyTorch/TensorFlow) and provided as a REST API service.
- *Source Code*: Publicly available via GitHub (link as a footnote or in references).

IV. Results and Discussion

Performance is evaluated using Mean Absolute Error (MAE) and Mean Squared Error (MSE), comparing against fully supervised baselines. Quantitative results indicate a reduction in error under semi-supervised training, particularly when labeled data is scarce. Qualitative analysis involves visual comparison of actual and predicted density maps, highlighting model performance in scenes with significant occlusions or background variation.

V. Conclusion and Future Work

This work demonstrates that semi-supervised learning, combined with patch-based regression, can effectively address the annotation bottleneck in crowd counting

applications. The modular, containerized implementation facilitates integration into real-world systems. Future work includes exploring domain adaptation for scenes with drastically different visual characteristics, integrating temporal information from video feeds, and scaling experiments to larger, more diverse datasets for further validation. Enhancements such as active learning and uncertainty estimation may further reduce required labeling effort.

REFERENCES

[1] Y. Cai and D. Zhang, "A Weakly Supervised Crowd Counting Method via Combining CNN and Transformer," *Electronics*, vol. 13, 2024.

[2] H. Duanc et al., "Semi-Supervised Crowd Counting from Unlabeled Data," arXiv:2102.11616 [cs.CV], 2021.

[3] Crowd_counting_semi-supervised [Online]. Available: https://github.com/bhu-04/Crowd_counting_semi-supervised.git

[4] IEEE template guidelines [Online]. Available: https://scribbr.com/ieee/format

[5] UCF-QNRF Dataset, ShanghaiTech Dataset (dataset URLs, if cited directly)