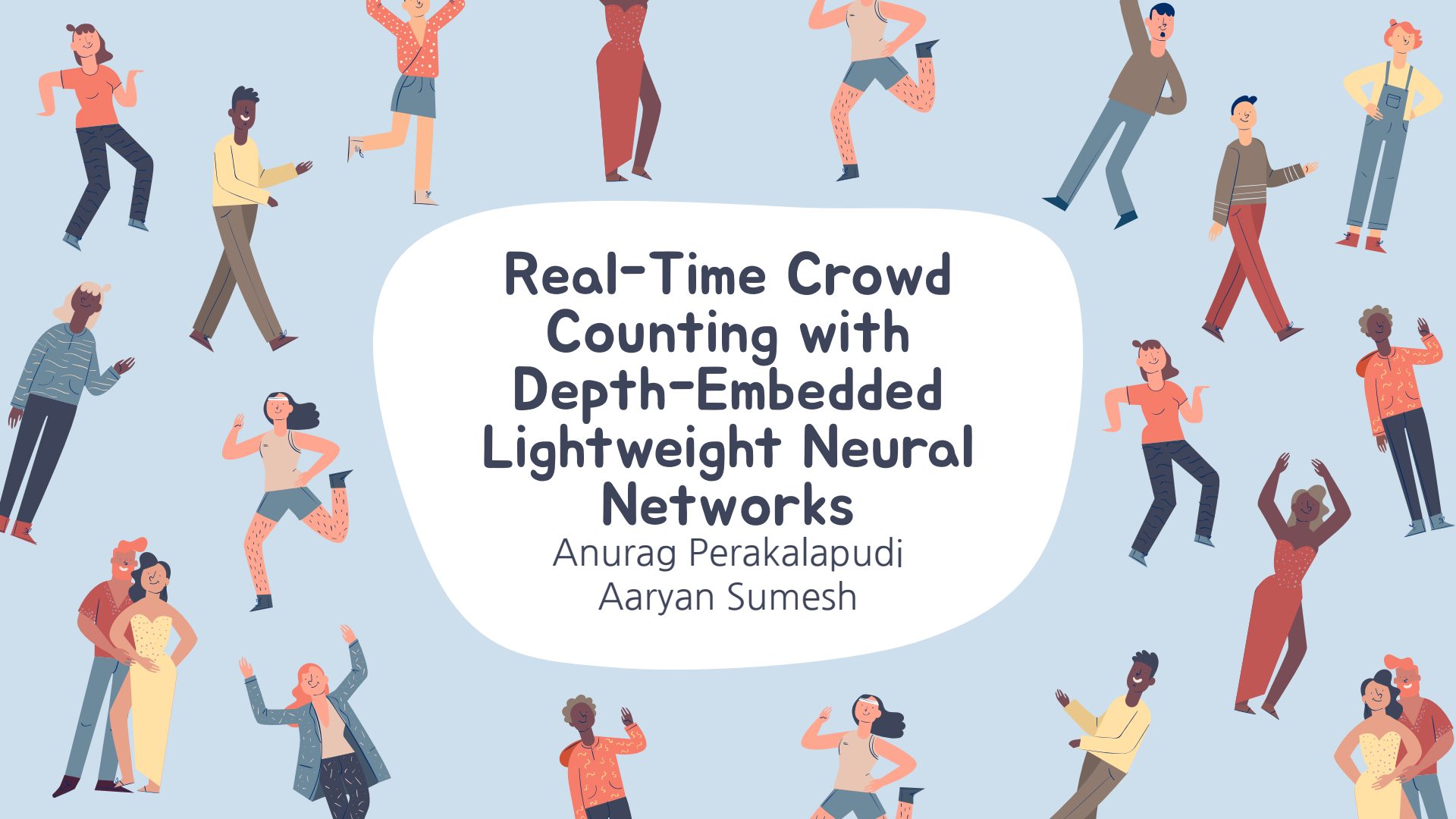


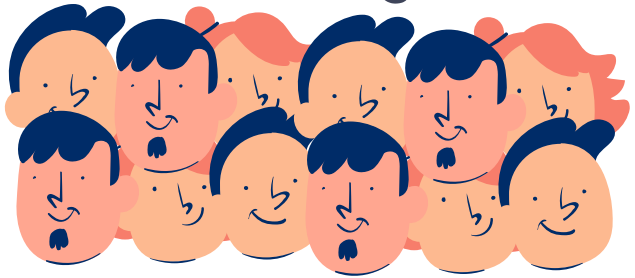
Real-Time Crowd Counting with Depth-Embedded Lightweight Neural Networks

Anurag Perakalapudi
Aaryan Sumesh

What is Crowd Counting?

- Public Safety
- Event Management
- Urban Planning

Challenges



- Occlusion (hidden object)
- Varying Crowd Densities
- Lighting Conditions

How Many People?

Correct Answer: 22

How Many People?

Related Works

Zhao et al. 2015

- Uses Depth Embedding
- 20 Million Parameters

Khan et al. 2023

- Lightweight LCDnet
- Uses 50K Parameters

Our Solution

Combine Depth Embedding with LCDnet

- Increased Accuracy
- Lightweight model for Real Time Systems

Dataset Overview

ShanghaiTech Dataset (Part A)

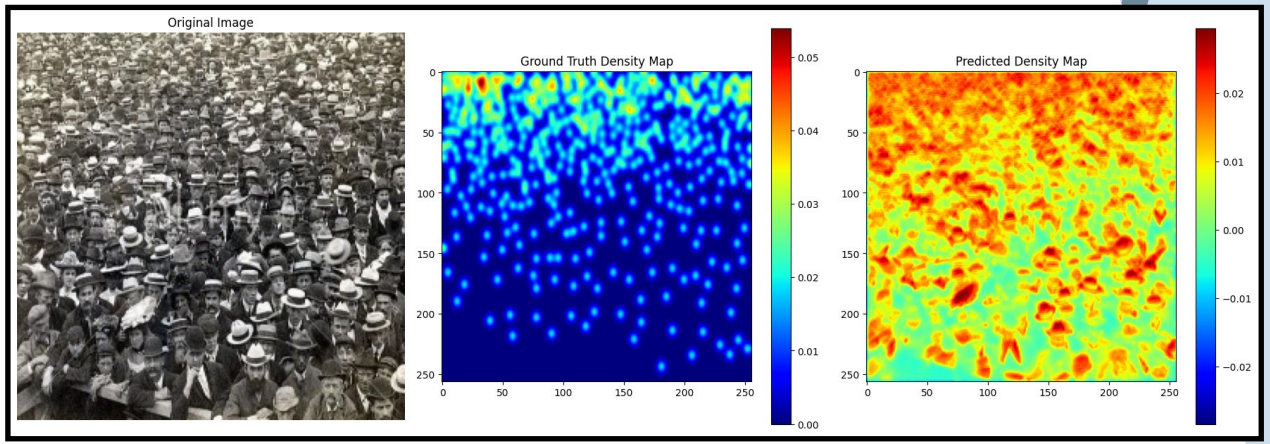
- One of the most widely used datasets for crowd counting research.
- Contains high-density crowd images with large variations in perspective.
- Each image is annotated with head locations, serving as ground truth.

Dataset Split:

- Training Data: 300 images with corresponding head annotations.
- Test Data: 50 images for evaluation.

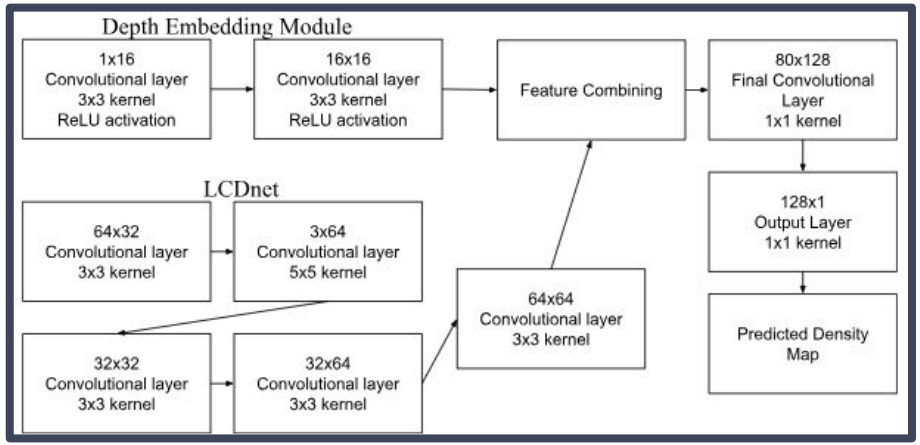
Preprocessing Steps

- Load Image
- Generate Density Map
- Normalization



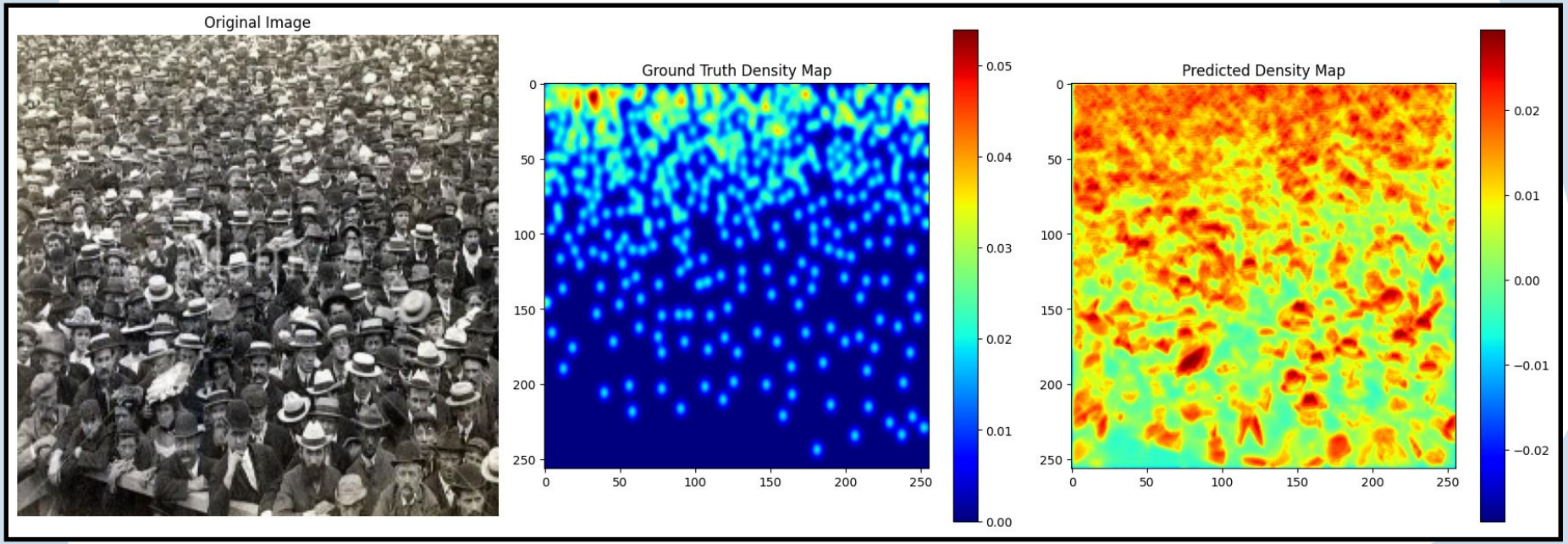
Systems Architecture

- LCDnet Architecture inspired by Khan et al.
- Depth Embedding Architecture inspired by Zhao et al.



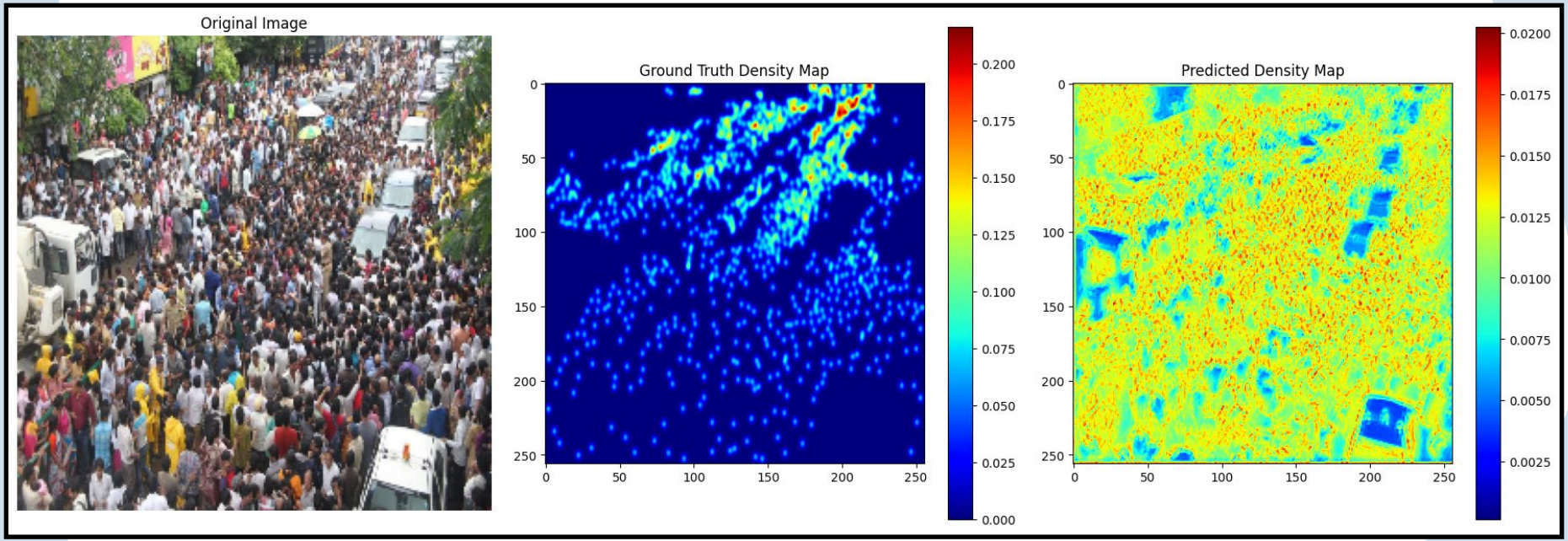
Results: Good

- no intervening objects (like trees, cars)
- Actual: 2320, Predicted: 2318.5449



Results: Bad

- Intervening objects (cars, trees)
- Actual: 2799, Predicted: 2647



Results

Metric: Mean Absolute Error $\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$

Model	Best mean average error
Model with depth embedding (slow)	57.55
Our Model (intermediate)	112.64
Lightweight CNN Model (fast)	181.8

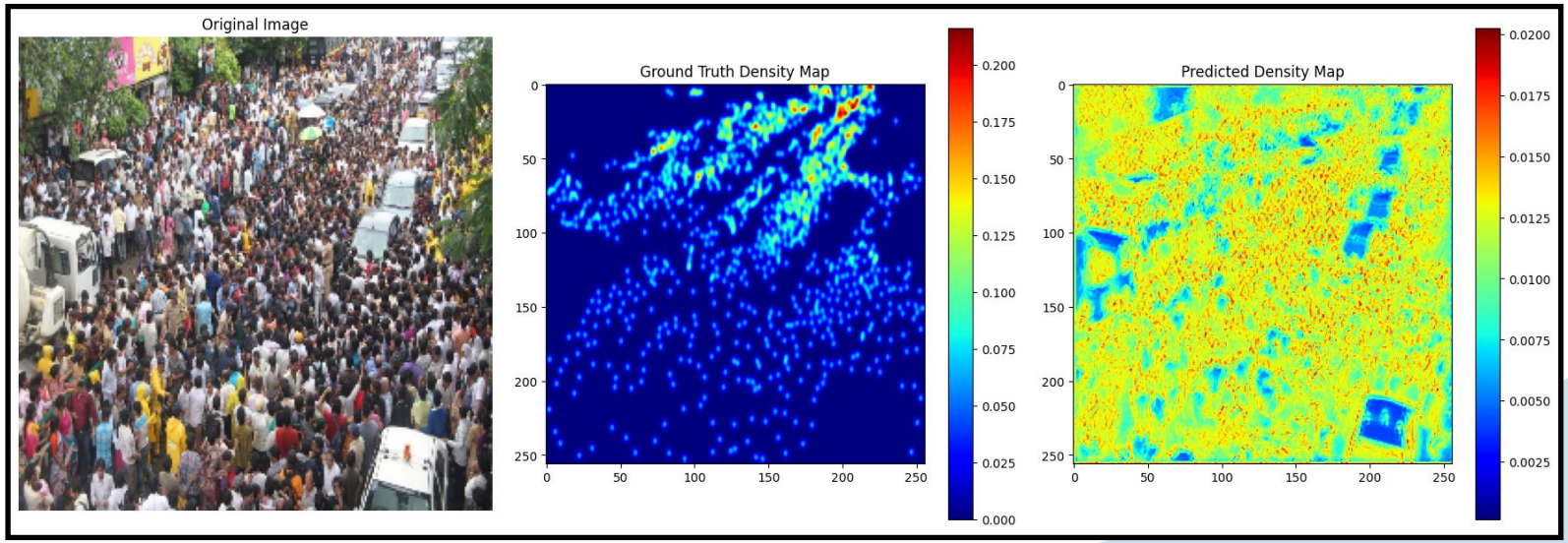
Model	Parameters Count
Model with depth embedding (slow)	20,000,000
Our Model (intermediate)	100,977
Lightweight CNN Model (fast)	50,000

Limitations

- We trained on the Shanghai Part_A Dataset because we postulated depth embedding accuracy improvements would be more noticeable.
- Train on more datasets
- Train on a variety of people
- Hardware

Conclusions

- Depth Embedding Increases Accuracy
- Lightweight Model Architecture is very versatile to be used in real-time Applications



Thanks!

Works Cited

[1] “Papers with Code - ShanghaiTech Dataset.” Papers With Code, https://paperswithcode.com/dataset/shanghaitech. Accessed 26 Jan. 2025.

[2] Zhao, M., Zhang, C., Zhang, J., Porikli, F., Ni, B., & Zhang, W. (2019). "Scale-aware crowd counting via depth-embedded convolutional neural networks." IEEE Transactions on Image Processing. Accessed 26 Jan. 2025.

[3] Khan, Muhammad Asif, et al. “LCDnet: A Lightweight Crowd Density Estimation Model for Real-Time Video Surveillance.” Journal of Real-Time Image Processing, vol. 20, no. 2, Mar. 2023, pp. 1-11, https://doi.org/10.1007/s11554-023-01286-8.

[4] Gillis, Alexander S., et al. “What Is a Convolutional Neural Network (CNN)?” TechTarget, 25 Nov. 2024, https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network.

[5] Bai, H., Mao, J., & Chan, S.-H. G. (2020). A Survey on Deep Learning-based Single Image Crowd Counting: Network Design, Loss Function and Supervisory Signal.

[6] Li, Y., Zhang, X., & Chen, D. (2018). CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1091-1100.

[7] Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 589-597.