Crowd Counting

Exploring mall visitor dynamics

Presented By : Group 2 - LB 09

TABLE OF CONTENTS

PROBLEM DESCRIPTION

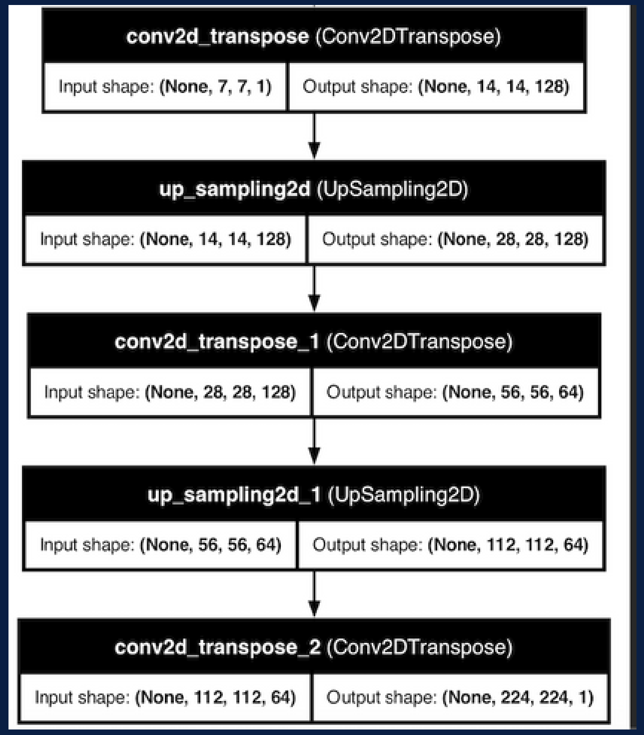
PROBLEM DESCRIPTION

Shopping malls play a crucial role as bustling centers of commerce, frequently drawing substantial numbers of visitors. High-traffic zones often indicate promising business prospects but may also present safety risks for patrons. Consequently, monitoring and understanding customer flow is vital for effective decision-making, optimizing daily operations, and creating a better retail environment.

MODEL AND JUSTIFICATION

MODEL AND JUSTIFICATION

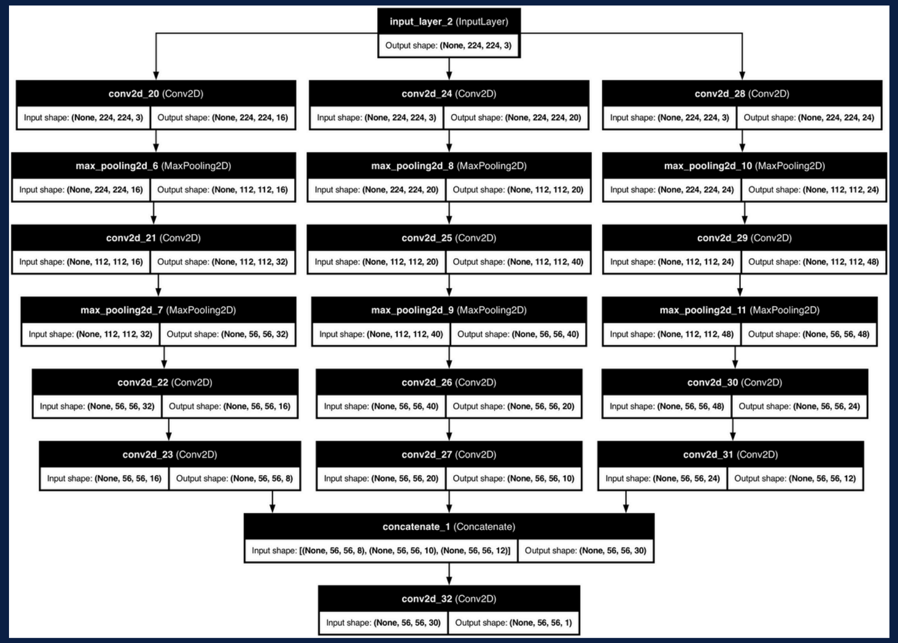
CRSNet



We selected CSRNet for this project due to its use of dilated convolution layers, which effectively expand the receptive field. This enables the model to capture broader contextual information from the image without compromising spatial resolution. After feature extraction through these dilated convolutions, we integrated a decoder module to reconstruct the output dimensions to match the original input. This decoder combines upsampling and Conv2DTranspose layers, not only enlarging the feature maps but also allowing the model to relearn and refine the density map, resulting in more accurate and realistic crowd estimations.

MODEL AND JUSTIFICATION

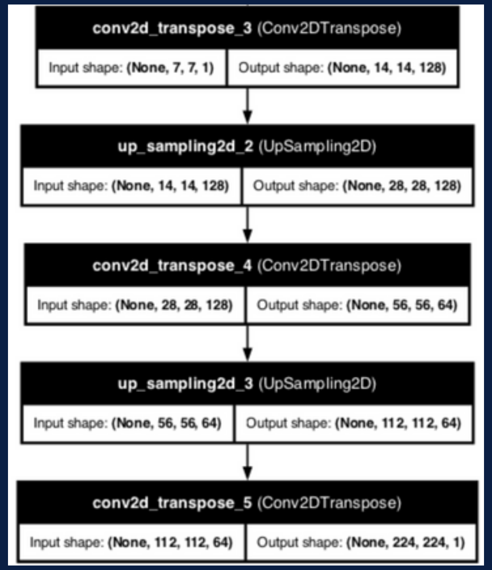
MCNN

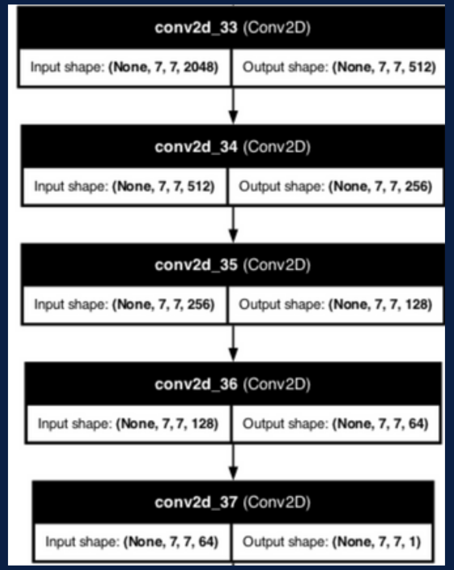


We chose MCNN because it addresses the challenge of perspective distortion, where head sizes in crowd images vary depending on their distance from the camera—appearing larger when closer and smaller when farther away. Using a single filter size with a fixed receptive field might cause the model to miss these scale variations. To solve this, MCNN uses multiple convolutional columns, each with different filter sizes. This multi-scale approach enables the model to detect heads of varying sizes, resulting in more consistent and accurate density map predictions across the entire image.

MODEL AND JUSTIFICATION

RestNet101

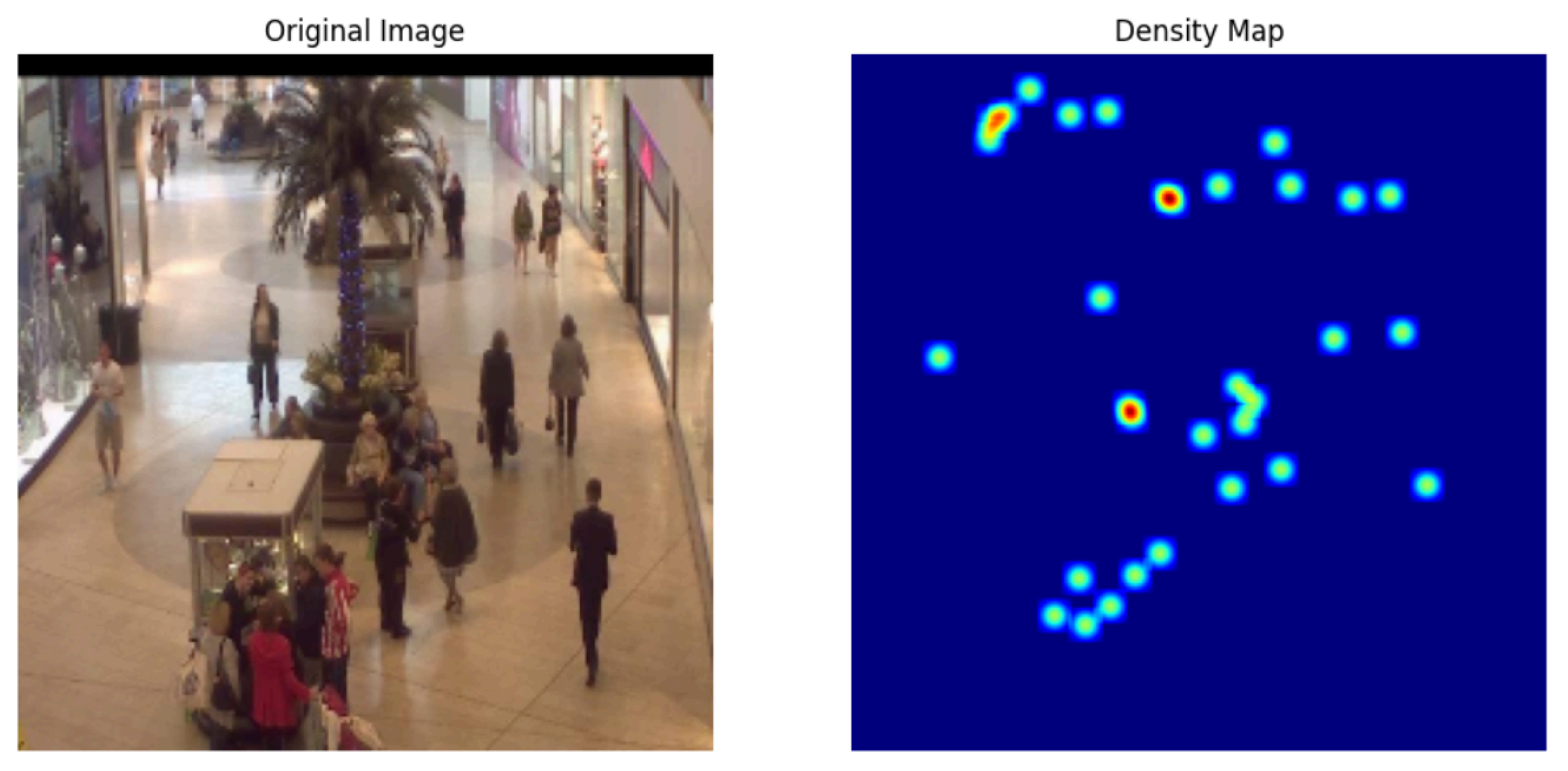




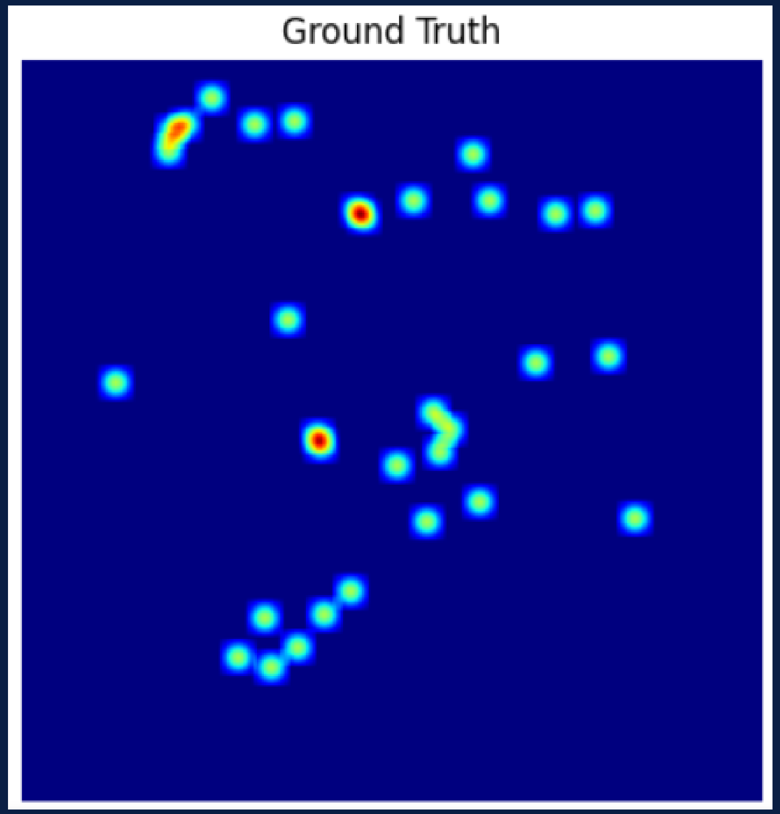
We used ResNet101 to evaluate whether increasing the network depth to 101 layers could lead to better performance in our crowd counting model. ResNet is specifically designed to overcome the vanishing gradient problem through the use of residual connections, which allow the network to maintain effective learning even as it becomes deeper. Once ResNet101 completes the feature extraction, we introduced a decoder using Conv2DTranspose layers and upsampling to restore the spatial resolution, ensuring that the output size closely matches the input image.

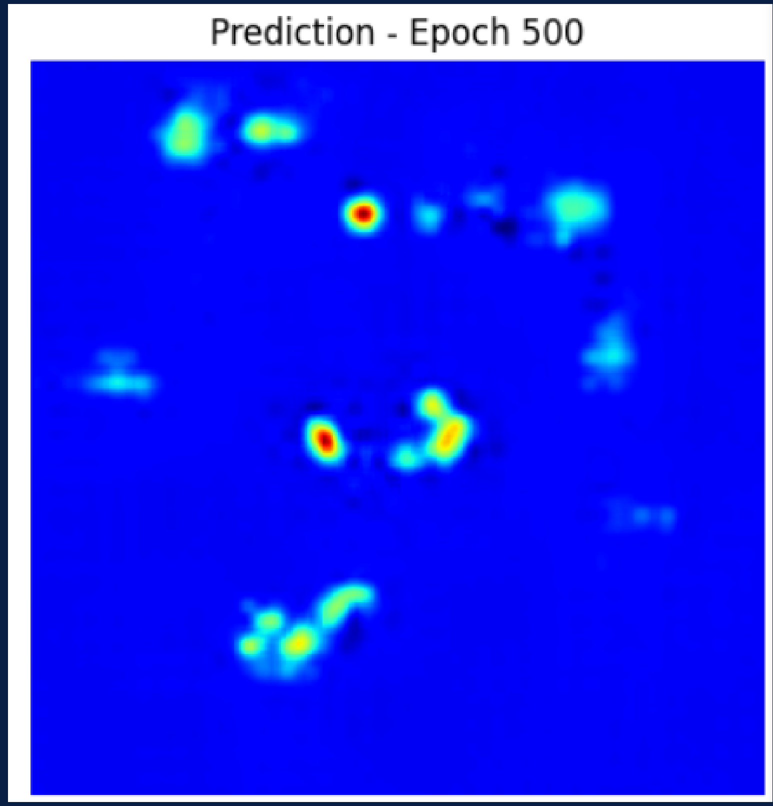
EVALUATION

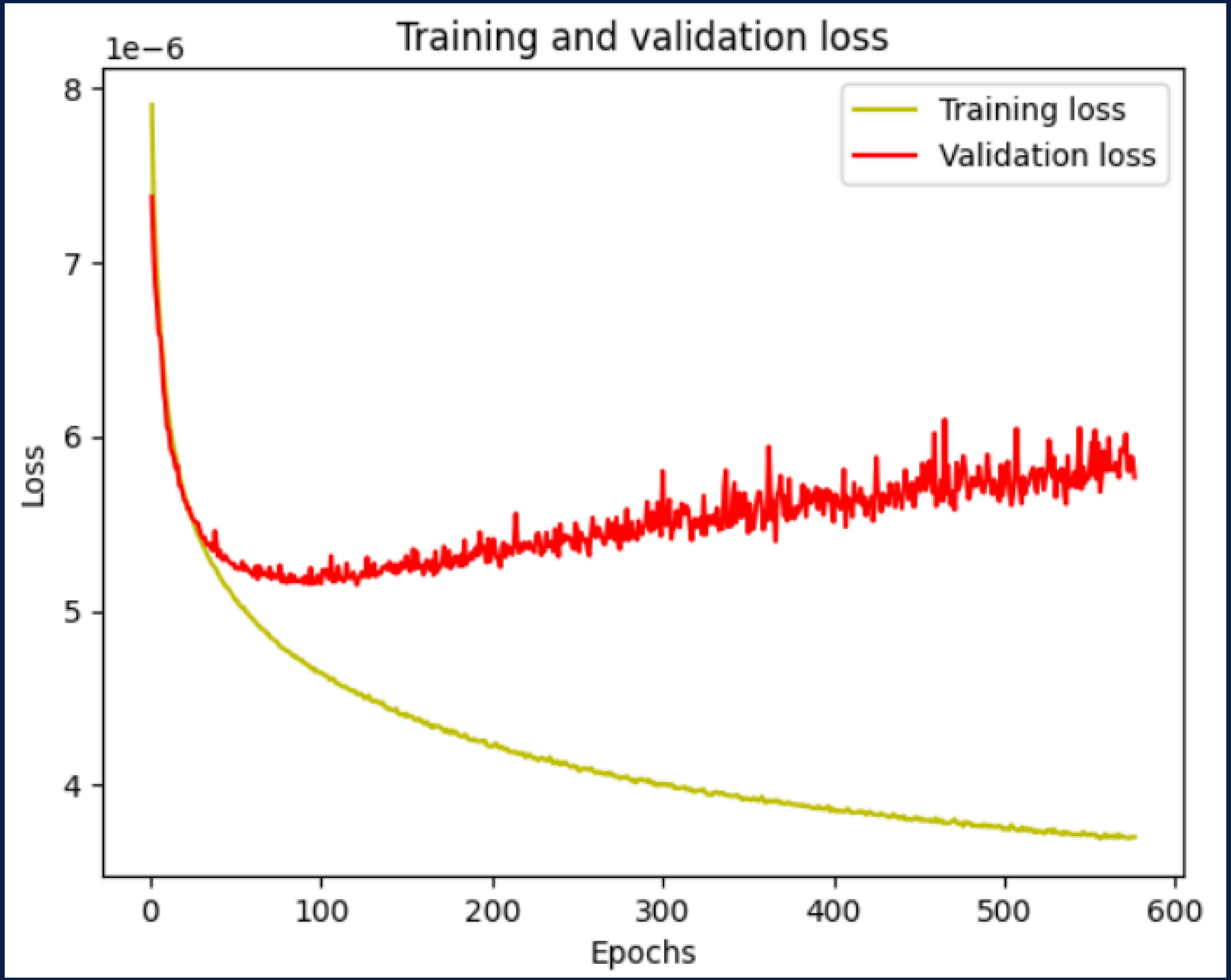
SAMPLE OF INPUT AND ITS DENSITY MAP



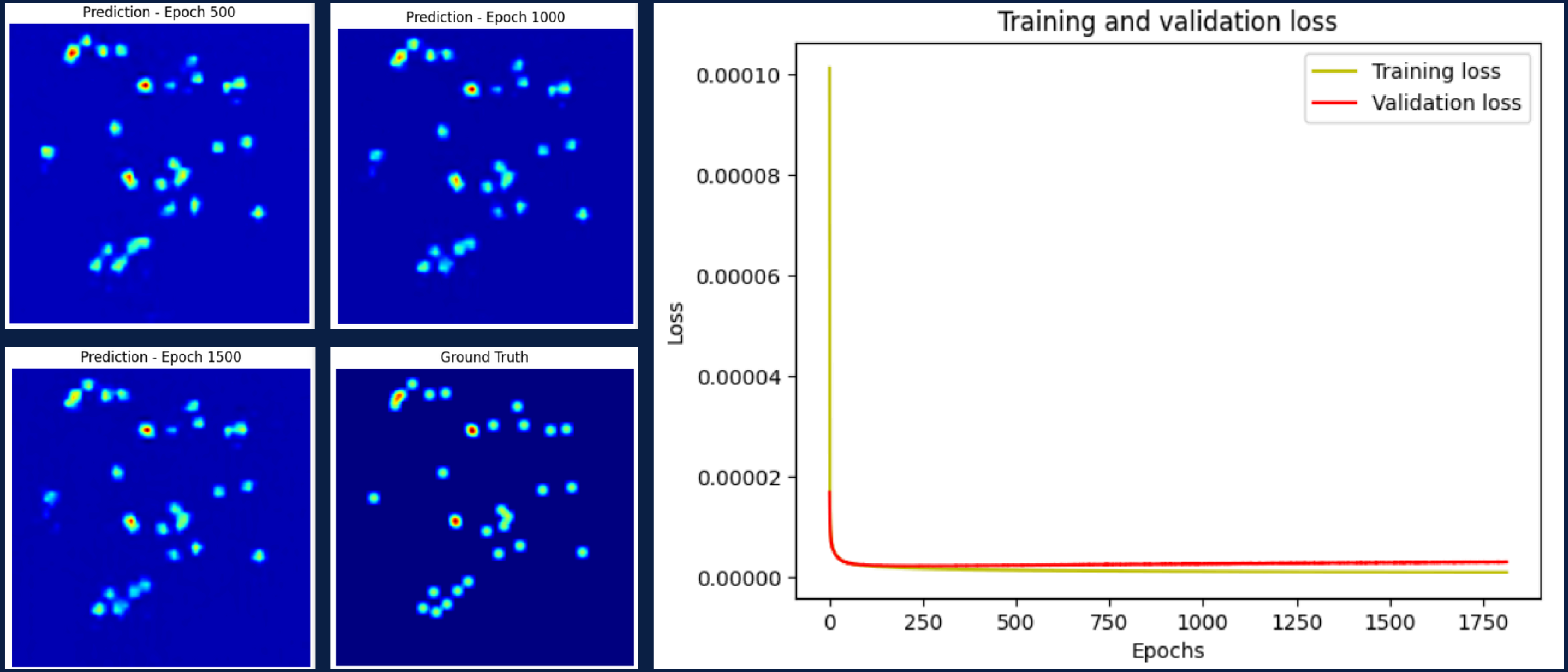
MODEL 1: CRSNET





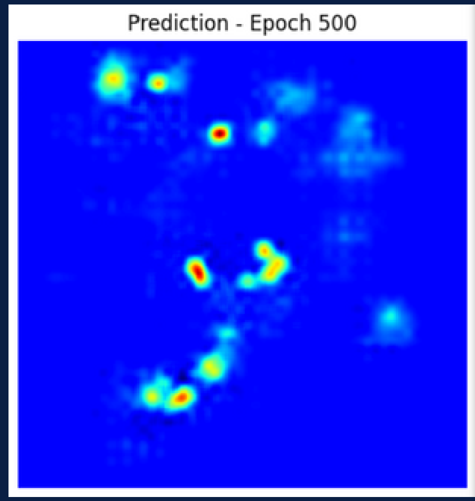


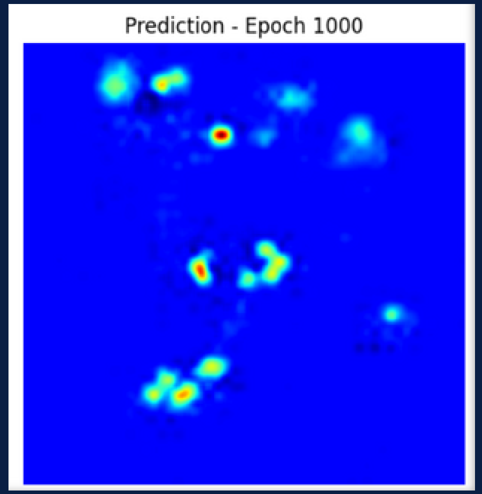
MODEL 2: MCNN

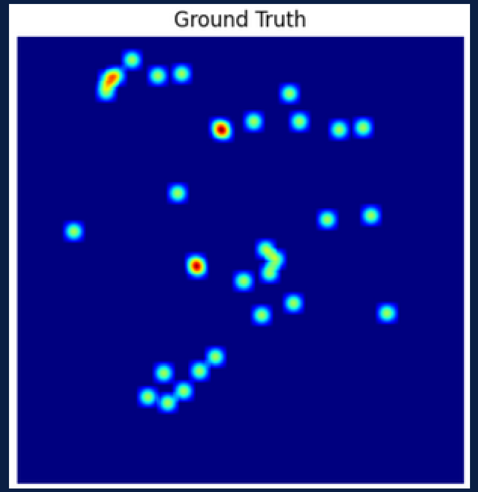


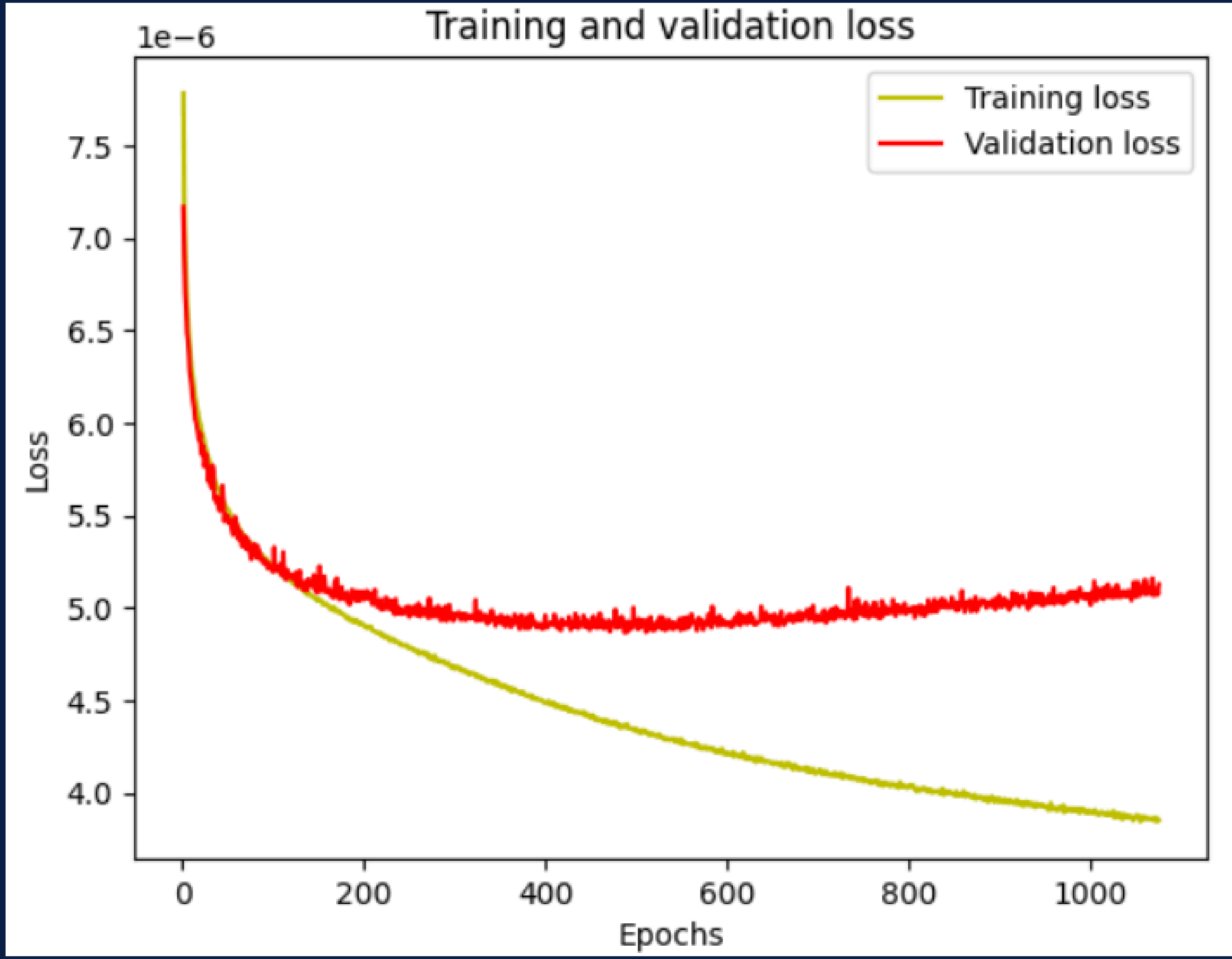
Prediction - Epoch 500
Prediction - Epoch 1000
Prediction - Epoch 1500
Ground Truth
Training and validation loss
Training loss
Validation loss
Loss
Epochs
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
250
500
750
1000
1250
1500
1750

MODEL 3: RESNET101









METRICS

MEAN SQUARED ERROR

MSE measures the difference between the predicted total count and the actual count. The model generates a density map, and the total count is the sum of all its values. MSE is computed by averaging the squared differences between predicted and actual counts across all images. Larger errors have a greater impact, so lower MSE indicates better overall accuracy.

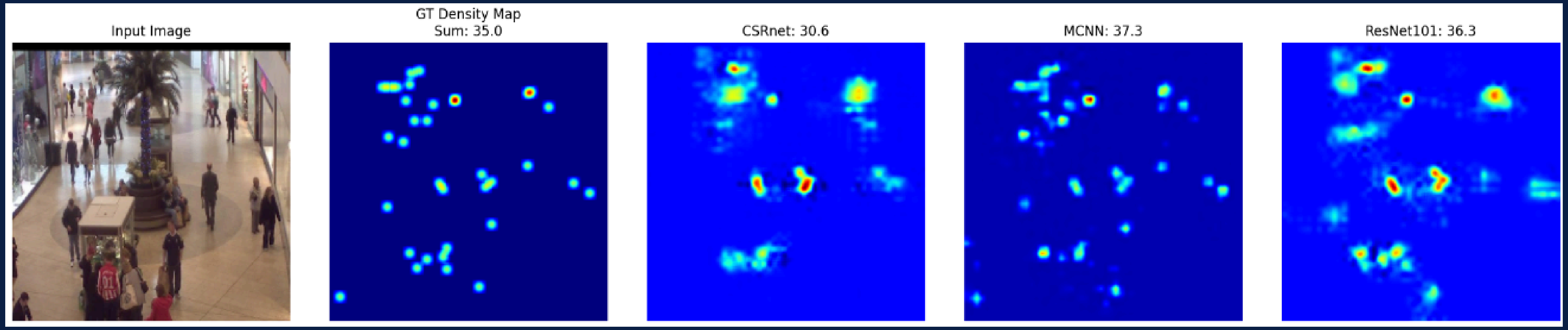
GRID AVERAGE MEAN ABSOLUTE ERROR

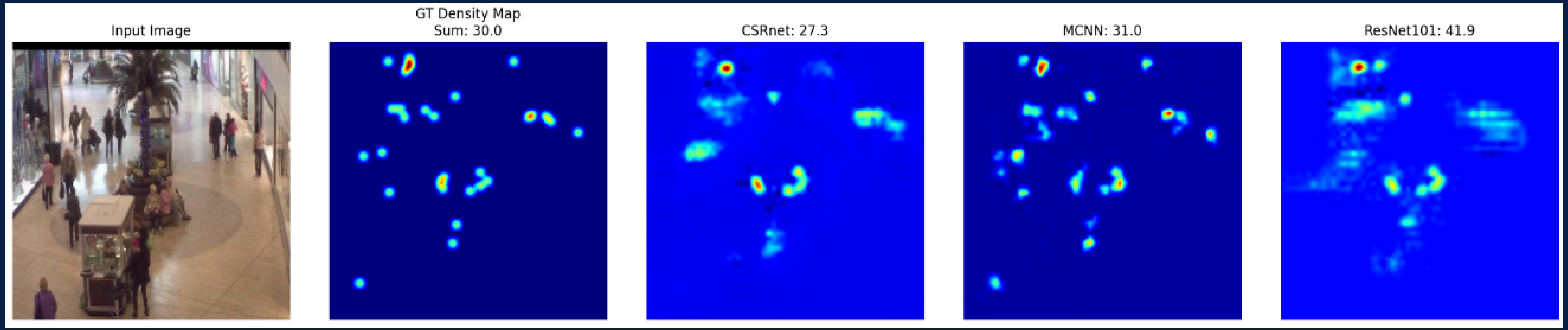
GAME evaluates both counting accuracy and spatial distribution. The image is divided into equal grids (e.g., 4 parts when L=1). GAME calculates the average absolute error in each grid. For example, if ground truth counts are evenly spread and the model misallocates density across regions, GAME captures this spatial mismatch—even if the total count is nearly correct.

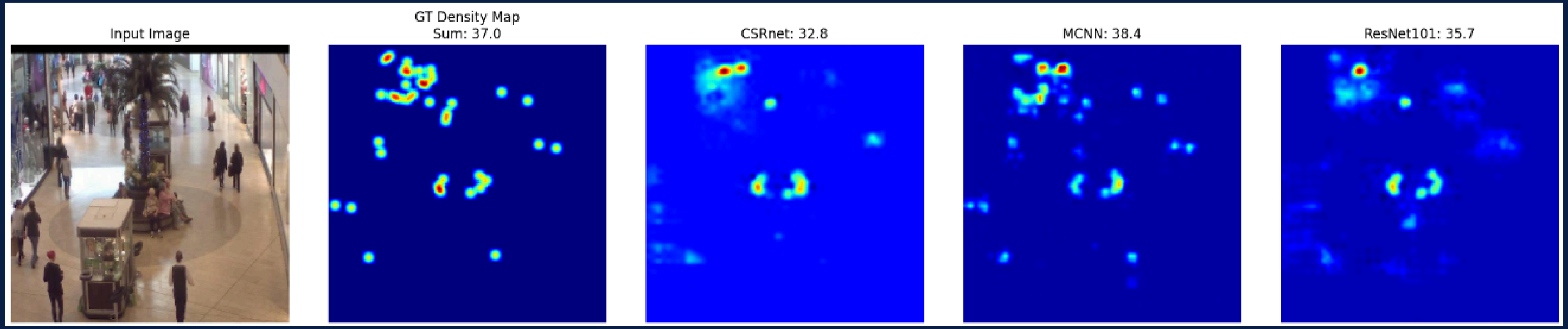
EVALUATION RESULT

	CSRNet	MCNN	ResNet101
MSE	22.2	8.3	19
GAME	3.66	2.1	3.48

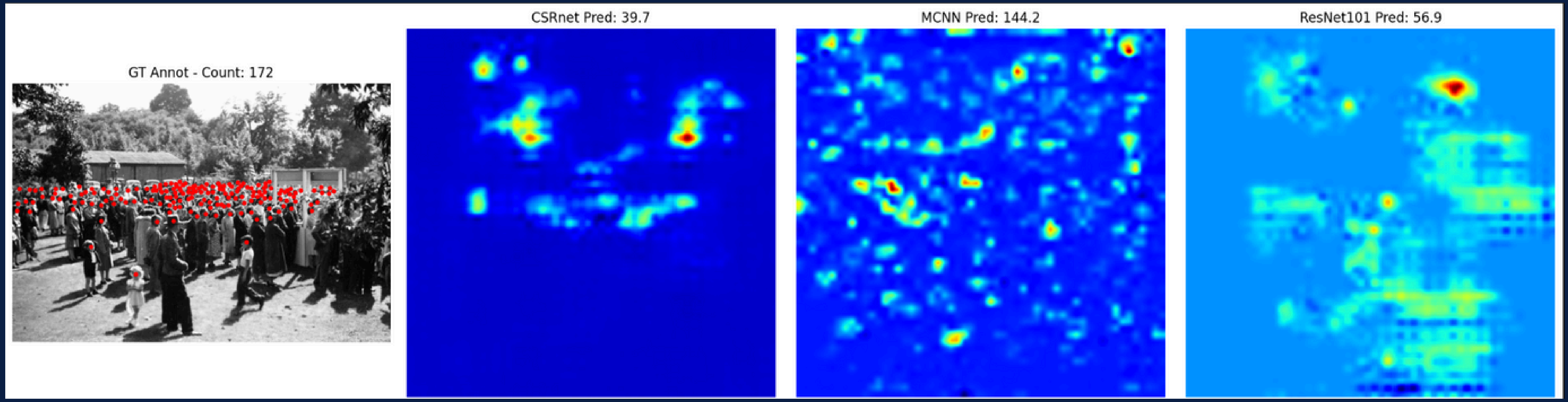
RESULT EXAMPLE

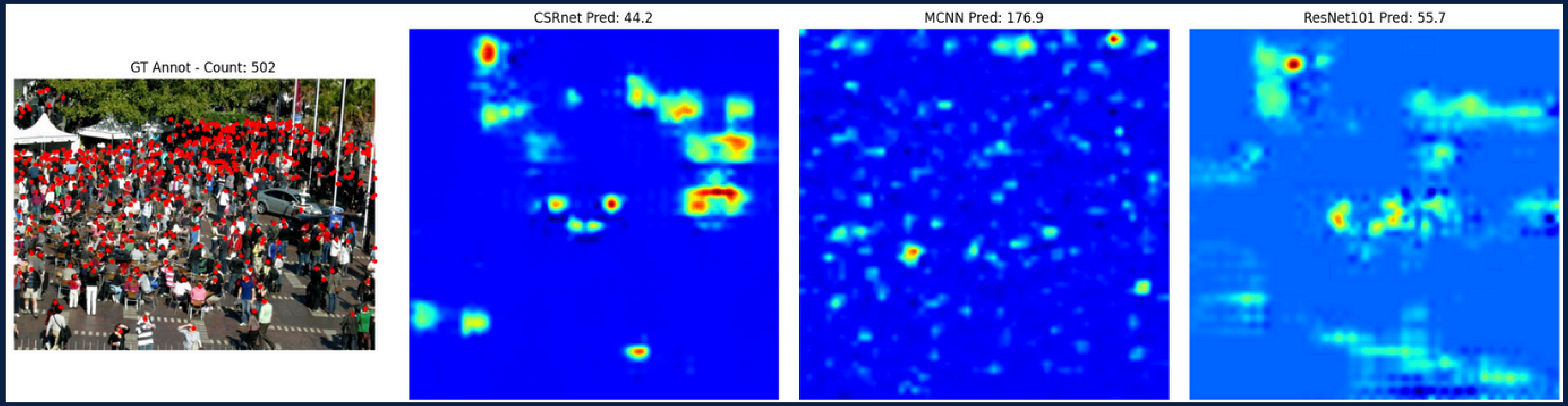






MODEL IMPLEMENTATION





CONCLUSION

CONCLUSION

THEREFORE, THE MCNN APPROACH IS THE BEST DEEP LEARNING ALGORITHM FOR ANALYZING FOOT TRAFFIC IN SHOPPING MALLS, SUPPORTING BETTER OPERATIONAL EFFICIENCY DECISIONS FOR STAKEHOLDERS. HOWEVER, AS A LIMITATION, THIS PROJECT ONLY EVALUATED THREE ALGORITHMS. FUTURE WORK SHOULD EXPAND THE MODEL POOL TO INCLUDE MORE ARCHITECTURES FOR A BROADER AND MORE COMPREHENSIVE COMPARISON.

Thank You