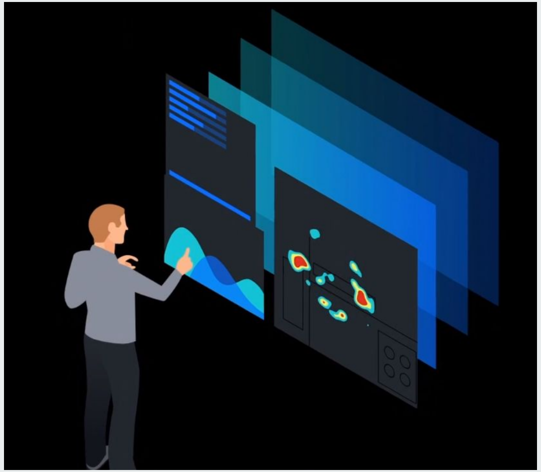
CROWD COUNTING AND MONITORING SYSTEM

Guided By : Prof. Swapnil G. Mali

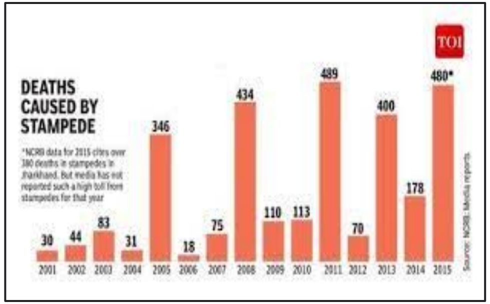
Group Members:

111807032 - Sakshi Patil
111807079 - Mohit Bhaiyya
111807081 - Pranav Gandhe



INTRODUCTION :

- Population is Increasing.
- Monitoring and Analysis of Crowd.
- Detection and counting crowd levels.



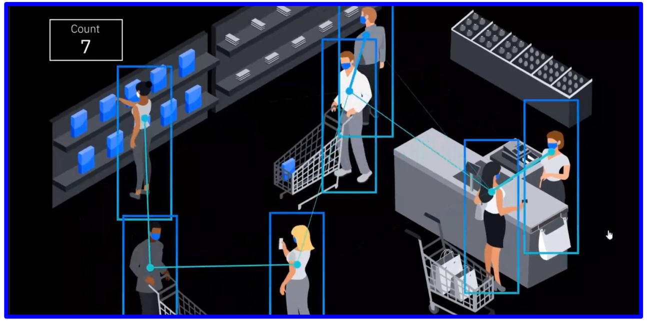
Applications :

- Disaster management
- Safety and Health Monitoring
- Intelligence gathering and analysis
- Military Applications

Methodology and Implementation:

There are currently four methods we can use for counting the number of people in a crowd:

1) **Detection-Based Methods**
2) **Regression-based methods**
3) **Density estimation-based methods**
4) **CNN-Based Methods**



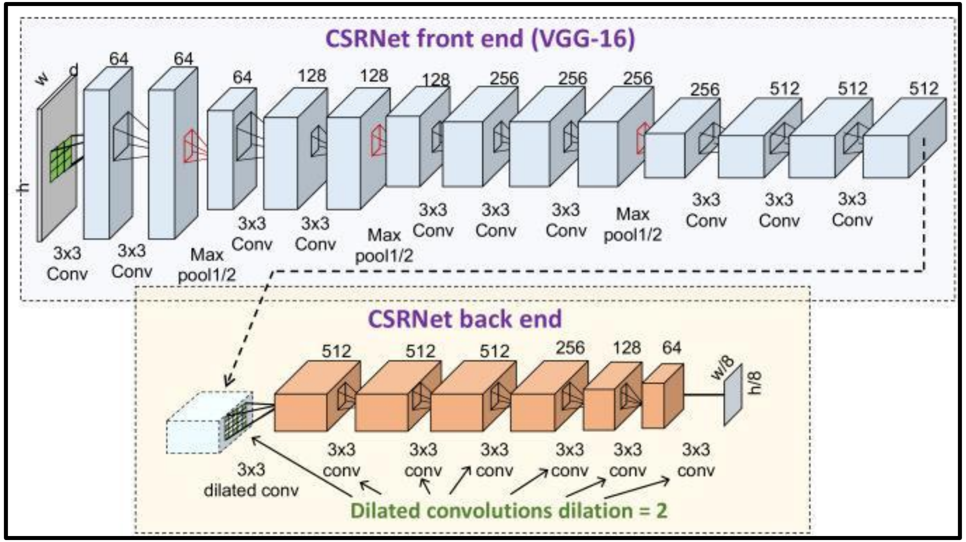
OUR IMPLEMENTATION:

We use a network for **Congested Scene Recognition** called **CSRNet**.

CSRNET Architecture:

The model architecture is divided into **two parts**, **front-end and back-end**.

1. Front End - VGG 16 (13 layers)
2. Back End - Dilated Convolution Layers



VGG 16

- Convolutional Neural Network
- Pre trained model

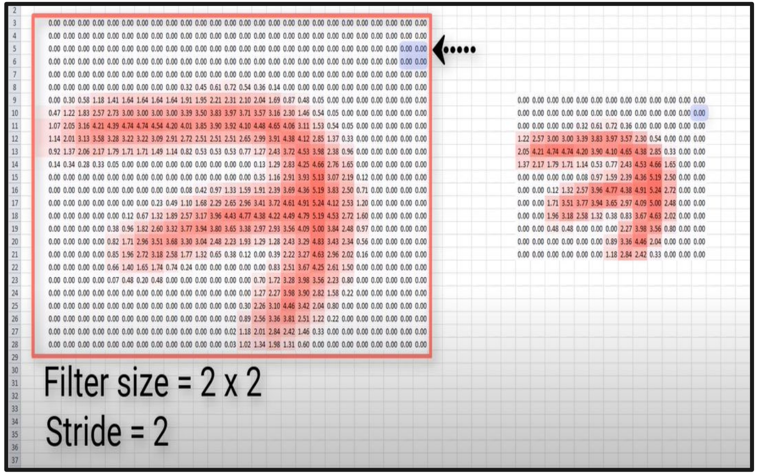
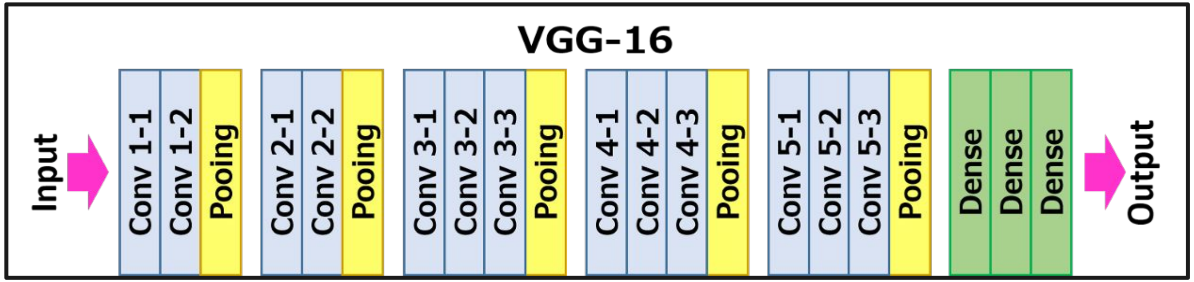


Fig: Max Pooling

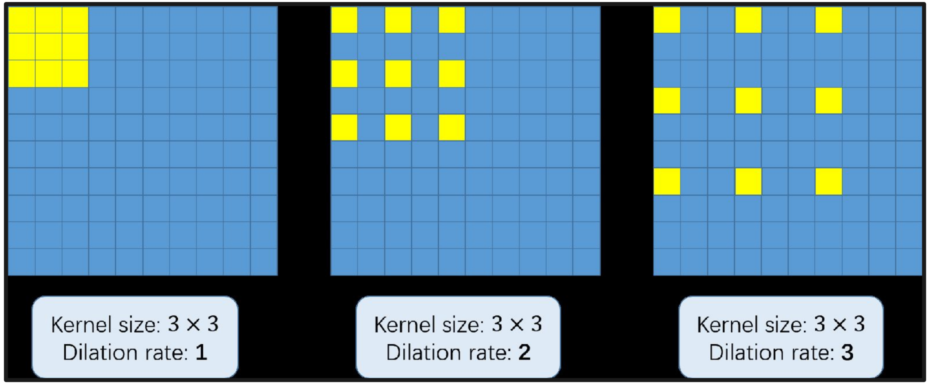
Key Features:

- 16 layers - Convolutional and Pooling layers
- 3 x 3 kernel for convolution
- 2 x 2 size of max pool
- Trained on ImageNet data
- About 138 million parameters



Dilated Convolution :

Dilated Convolutions are a type of **convolution** that “inflate” the kernel by inserting holes between the kernel elements.

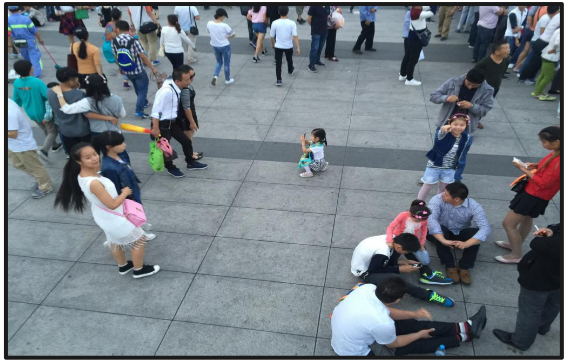


The basic concept of using dilated convolutions is to enlarge the kernel without increasing the parameters.

- if the dilation rate is 1, we take the kernel and convolve it on the entire image.
- If the dilation rate is 2, the kernel extends as shown in the above image. It can be an alternative to pooling layers.

Dataset :

- ShanghaiTech Dataset - 1198 Crowd Images.
- Part A - High Density Crowd - 482 Images.
- Part B - Sparse Density Crowd - 716 Images.



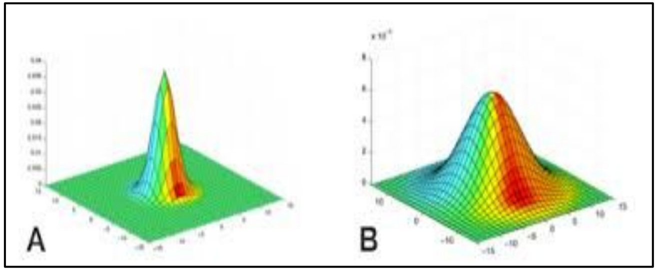
	Number of frames	Resolution	Minimum number of persons	Maximum number of persons	Total number of persons	Average crowd count
Part A	482	different	33	3139	241677	501.4
Part B	716	768*1024	9	578	88488	123.6

Data Preprocessing :

ShanghaiTech Dataset → Ground Truth → Sparse Matrix → Head annotation → Gaussian Filter → 2-D density maps

Further Process:

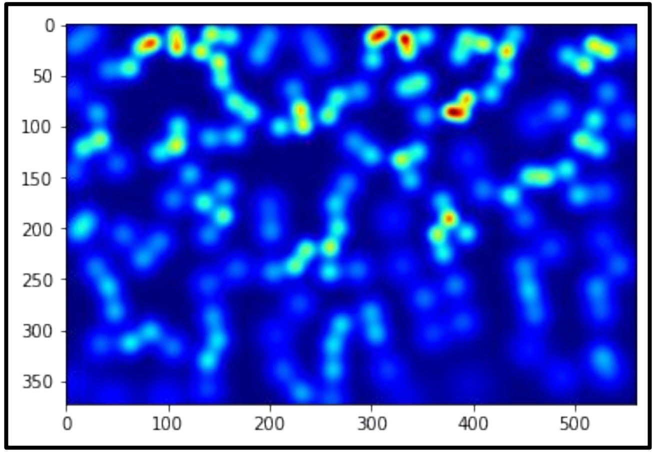
- To convert this to a continuous density function, we convolve this function with a Gaussian kernel G so that the density is **F(x) = H(x) ∗ G(x)**.
- The average distance between the head and its nearest **'k'** neighbors is calculated.



$$F(x) = \sum_{i=1}^{N} \delta(x - x_i) * G_{\sigma_i}(x), \quad \text{with} \quad \sigma_i = \beta \bar{d^i}$$

(where F(x) is the density function)

A image and its conversion into density map.

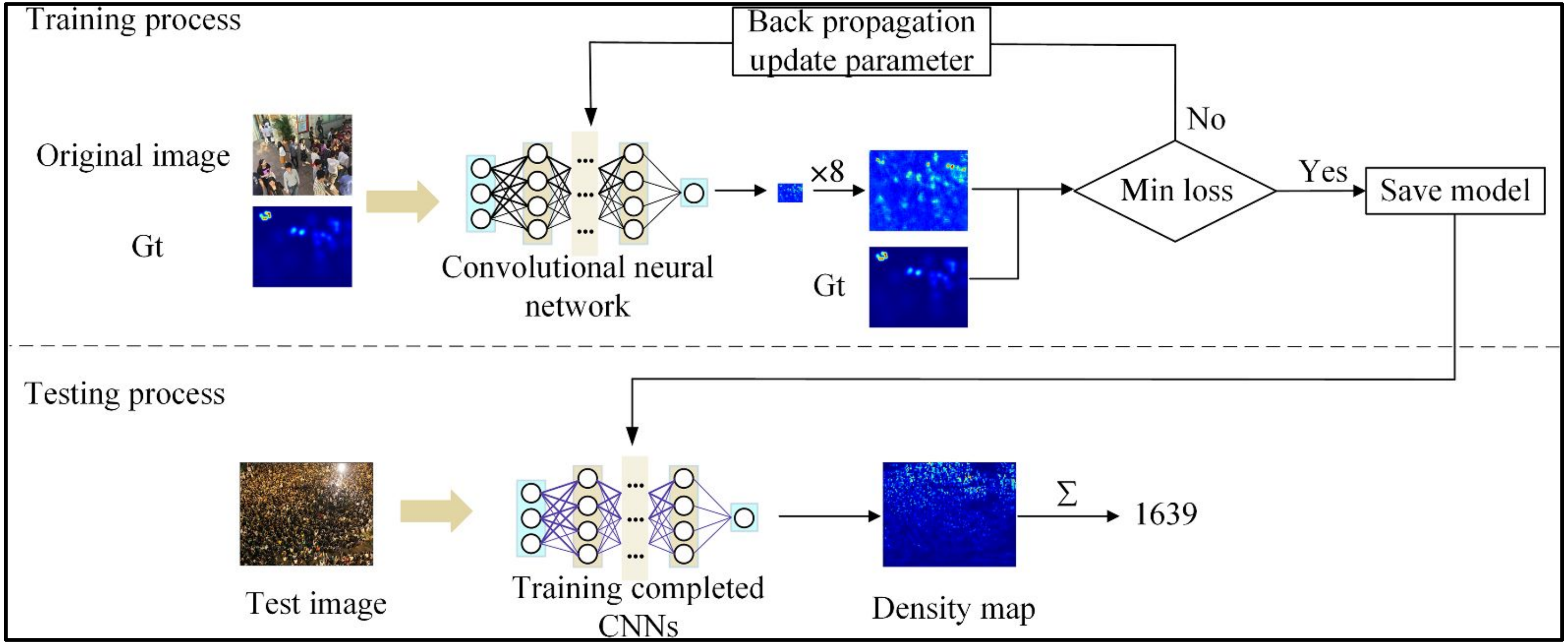


Model and Training Details:

- Optimizer - Stochastic Gradient Descent (SGD)
- Learning Rate - 1e-7
- Number of epochs - 400
- Batch size - 1
- Momentum - 0.95
- Euclidean distance : Measure of difference between the ground truth and the estimated density map generated

$$\boxed{L(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \| \ Z(X_i; \Theta) - Z_i^{GT} \ \|_2^2} \Rightarrow \boxed{\textbf{Loss Function}}$$

where, N = size of training batch and
Z(Xi ; Θ) = output generated by CSRNet with parameter Θ.
Xi = input image
Zi^(GT) = ground truth result of the input image Xi .


Training process
Back propagation
update parameter
No
Original image
Gt
×8
Min loss
Yes
Save model
Convolutional neural
network
Gt
Testing process
Σ
1639
Test image
Training completed
CNNs
Density map

Evaluation Metric:

- **Mean Absolute Error -** average difference between the actual data value and the value predicted by the model.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | C_i - C_i^{GT} |$$

- Dense Crowd Dataset:

```
[14] print("Mean Absolute Error :",mae_avg)
     print("Mean Absolute Percentage Error :",(mae_avg/avg)*100,"%")

     Mean Absolute Error : 65.96636728140024
     Mean Absolute Percentage Error : 15.36797194376203 %
```

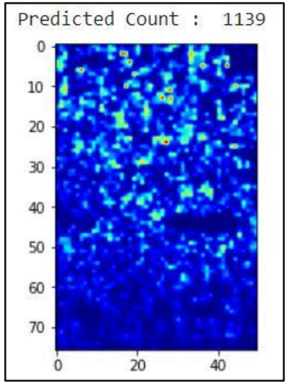
- Sparse Crowd dataset:

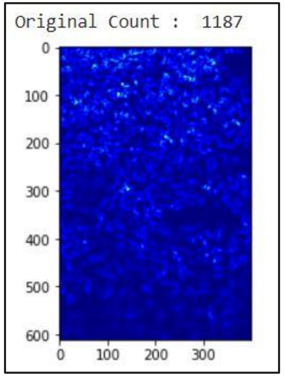
```
[ ] print("Mean Absolute Error :",mae_avg)
    print("Mean Absolute Percentage Error : ",(mae_avg/avg)*100 ,"%")

    Mean Absolute Error : 10.549921805345559
    Mean Absolute Percentage Error :  8.792767923155468 %
```

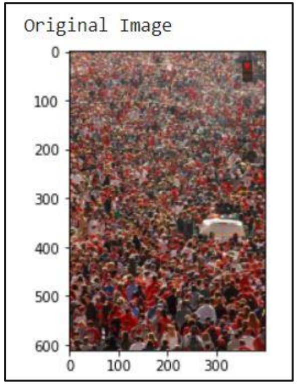
Result :

Dense Crowd :

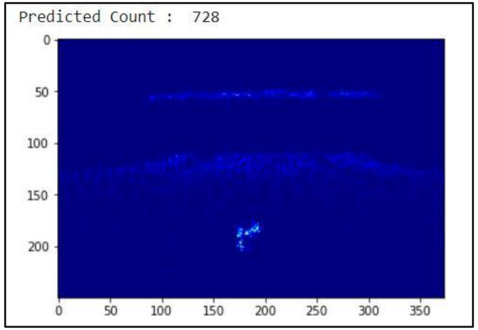
1)

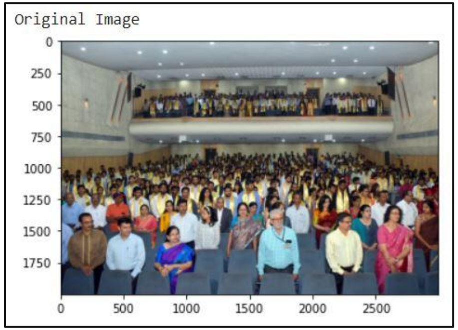






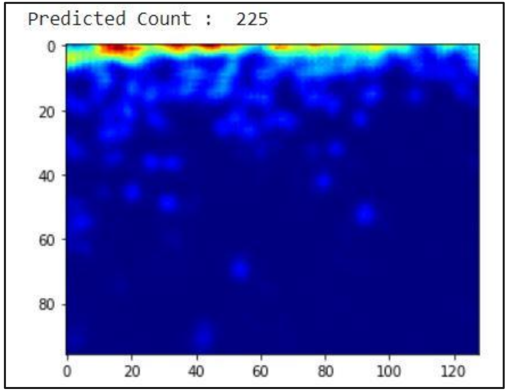
2)

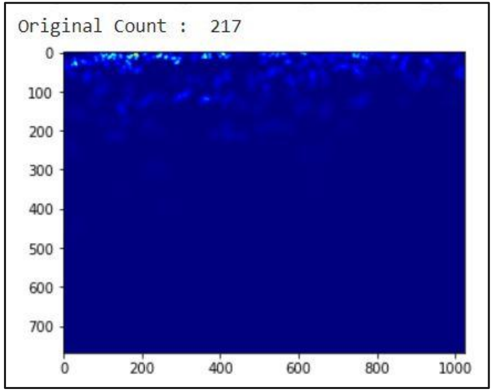


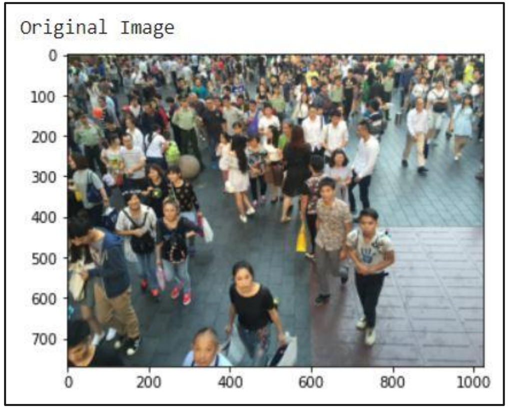


Sparse Crowd :

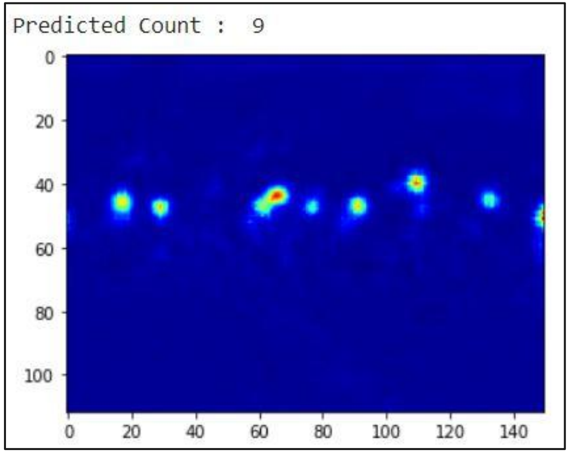
1)

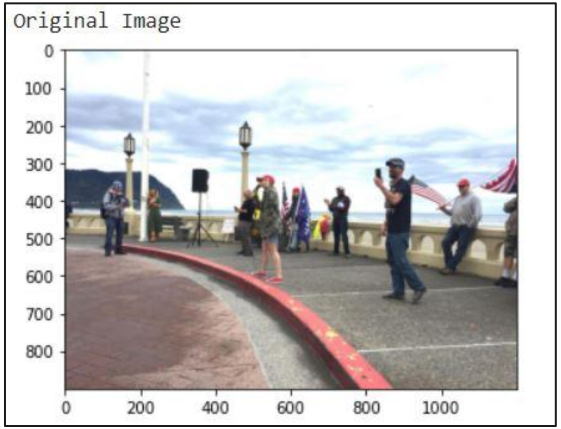






2)



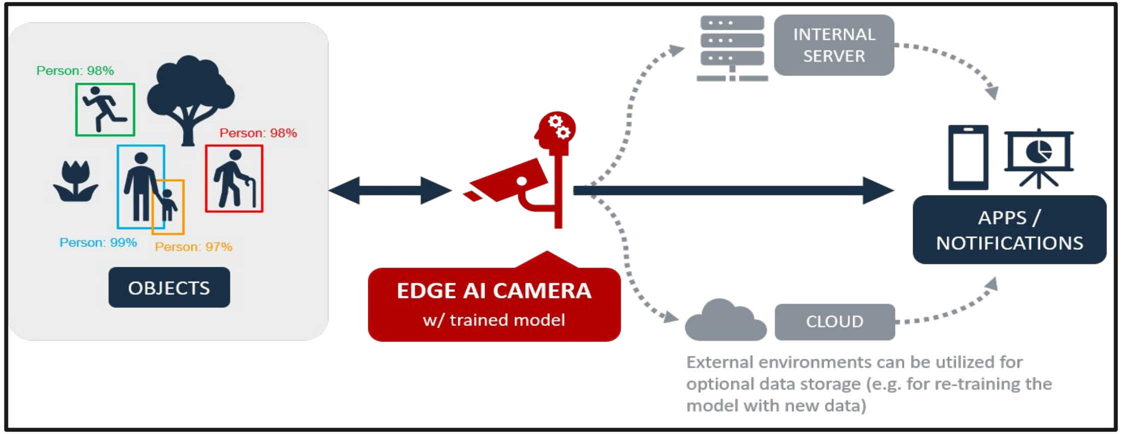


Conclusion:

- Effective approach for real time crowd counting using CSRnet is presented.
- Deep learning enables the system to perform in versatile environments and continuously learn from new inputs.
- Proposed methodology achieves crowd count predictions almost as good as ground truth values.
- Model A and Model B perform well on dense crowds and sparse crowd respectively.
- The density map generated by both models are accurate enough to depict the varied density of the crowd.
- CSRNet achieves low value of Mean Absolute Error (MAE).

Future Scope:

- We can train model on a huge dataset containing both dense and sparse crowd images.
- Hyperparameter tuning.
- Different backbone architectures instead of VGG 16 such as Resnet, inception,xception.
- Edge Cameras and Edge Computing



References :

- Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE.
  Conference on Computer Vision and Pattern Recognition, pages 589–597, 2016.
- Deepak Babu Sam, Shiv Surya, and R Venkatesh Babu.Switching convolutional neural network for crowd counting.In Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition, volume 1, page 6, 2017.
- https://neurohive.io/en/popular-networks/vgg16/
- Li, Y.; Zhang, X.; Chen, D. CSRNET: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA,
- Beibei Zhan, Dorothy N Monekosso, Paolo Remagnino, Sergio A Velastin, and Li-Qun Xu. Crowd analysis: a survey. Machine Vision and Applications, 19(5-6):345–357,
- Advances and Trends in Real Time Visual Crowd Analysis Khalil Khan 1,* , Waleed Albattah 2 , Rehan Ullah Khan 2 , Ali Mustafa Qamar 3,4 and Durre Nayab 5
- https://www.ibm.com/cloud/blog/crowd-density-measurement-at-the-edge
- https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf