Crowd Counting

Berk Güler, Munib Emre Sevilgen,
Şeyma Aybüke Ertekin, Yağmur Özkök

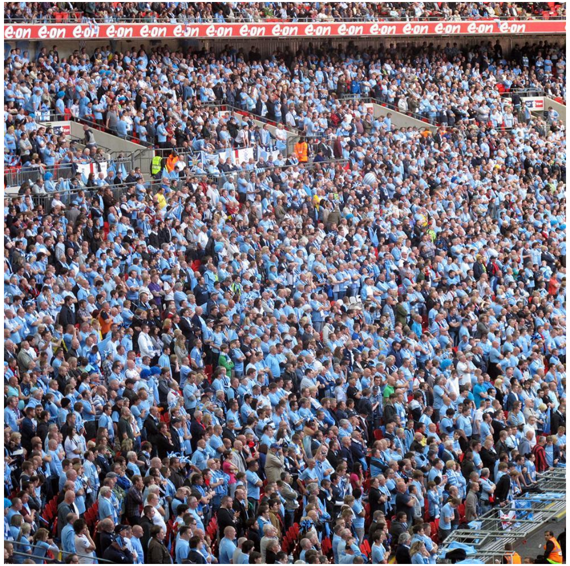
Problem Description

➔ Our goal is estimating the number of people in a crowd in an image.

➔ Difficult problem in terms of head size differences and dense crowds.

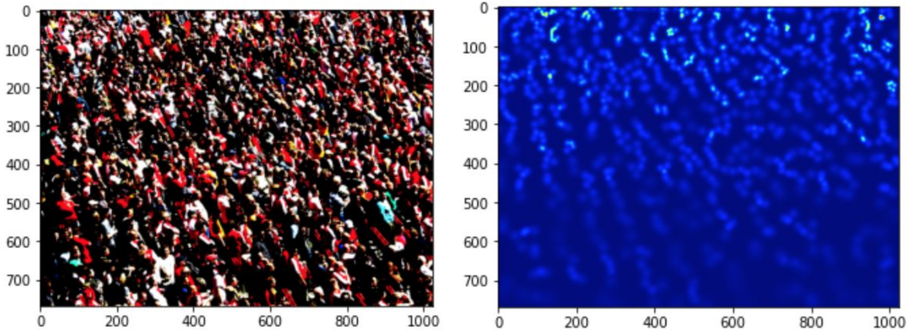
↓

Density estimation methods are required



Dataset

- ➔ Dataset called ShangaiTech
  - ◆ Has 2 parts called Part A and Part B
  - ◆ Part A contains 300 images for training, 182 images for testing
  - ◆ Part B contains 400 images for training, 316 images for testing
- ➔ In terms of Preprocessing;
  - ◆ Data processed by gaussian filter transformation to obtain density maps.
  - ◆ Density maps scaled to get the total number of people by summing up the whole map.



Methods

- Models:
  - CSRNet
    - Dilated CNN as backend
    - VGG-16 (Transfer Learning) as frontend
  - MCNN(Multi-column CNN)
- Data Augmentation

CSRNet

- ➔ This model use dilated convolutional layers to obtain high quality density maps
- ➔ VGG-16 layers in the front-end part (Transfer learning)
  - ◆ Network for Image Classification and Detection
- ➔ Dilated convolutional layers at the back end part
- ➔ Dilated convolution leads to alternating pooling and convolutional layers.

Configurations of CSRNet			
A	B	C	D
input(unfixed-resolution color image)			
front-end (fine-tuned from VGG-16)			
conv3-64-1			
conv3-64-1			
max-pooling			
conv3-128-1			
conv3-128-1			
max-pooling			
conv3-256-1			
conv3-256-1			
conv3-256-1			
max-pooling			
conv3-512-1			
conv3-512-1			
conv3-512-1			
back-end (four different configurations)			
conv3-512-1	conv3-512-2	conv3-512-2	conv3-512-4
conv3-512-1	conv3-512-2	conv3-512-2	conv3-512-4
conv3-512-1	conv3-512-2	conv3-512-2	conv3-512-4
conv3-256-1	conv3-256-2	conv3-256-4	conv3-256-4
conv3-128-1	conv3-128-2	conv3-128-4	conv3-128-4
conv3-64-1	conv3-64-2	conv3-64-4	conv3-64-4
conv1-1-1			

Figure 1

CSRNet : Dilation Rate

➔ Increase in dilation rate reduces the spatial information that is learned from the train data.

➔ The elements of kernel getting apart, the adjacent values don't attend to the convolution.

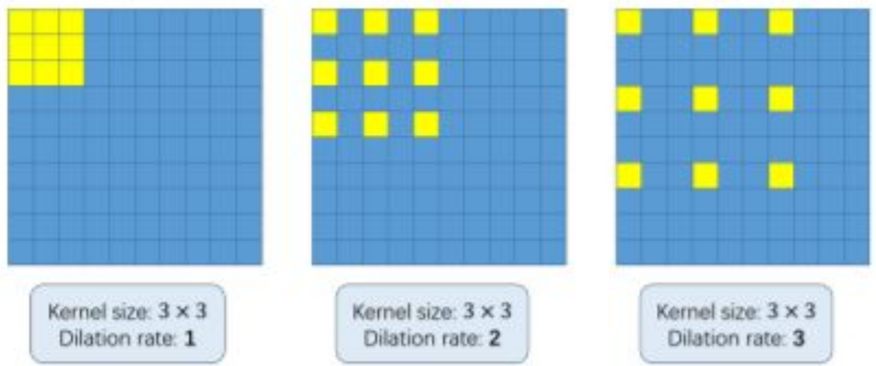


Figure 2

MCNN(Multi-Column CNN)

- ➔ This model takes image as an input and outputs a density map, head count then is obtained via integration.
- ➔ In each column of the MCNN, filters of different sizes are used to model the density maps corresponding to heads of different scales. Filters with large receptive fields are used are used for large heads and filters with small fields for small heads.

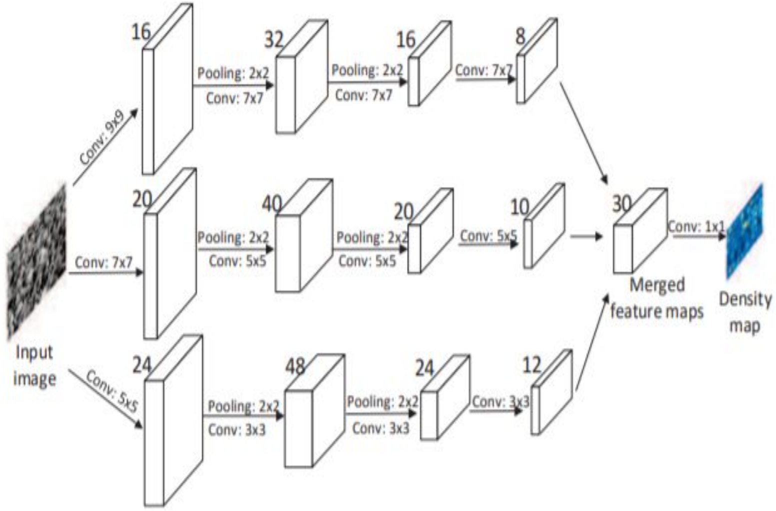


Figure 3

Data Augmentation

- ➔ CCAugmentation module in Python
- ➔ It supports some transformations such as crop, grayscaling, flip left or right for the images.
- ➔ This framework allows to increase dataset without finding and labeling new images.

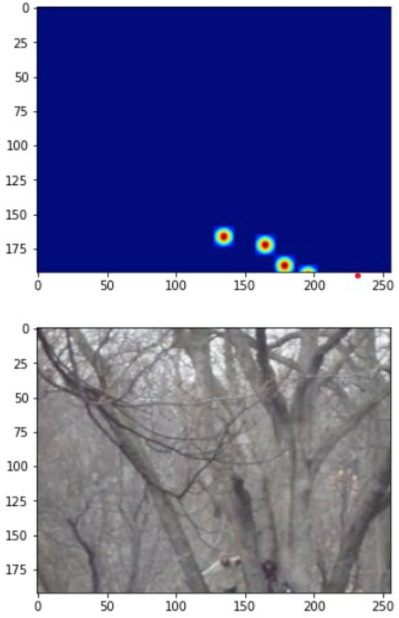


Figure 4

Results

- ➔ Evaluation of models based on mean absolute percentage error (MAPE) and modified mean absolute error (MMAE)
- ➔ An increase in the MMAE means that the average error for a person in any image is increased.
    - ◆ Modified mean absolute error (MMAE): $\frac{\sum_{i=0}^{n} |A_t - P_t|}{\sum_{i=0}^{n} |A_t|}$ where At is ground truth and Pt is predicted value
- ➔ An increase in the MAPE means that the average error for an image is increased.

⬇

If MMAE is better than the MAPE, this concludes that the model is better on more crowded images.

Results : CSRNet

- ➔ Figure 5 shows the Training Loss of each epoch while training the CSRNet model.
- ➔ The model trained with 1016 and tested with 186 different images. (Part A train, Part B train and test images)

MAPE as 25.44% and MMAE as 20.84%.

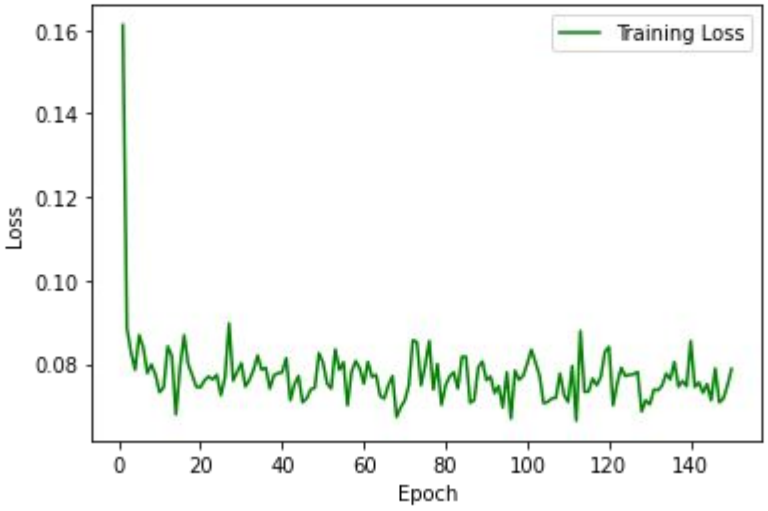


Figure 5

Results : CSRNet

- ➔ CSRNet model trained with different dilation rates.
- ➔ Figure x shows the MAPE and MMAE Errors for Different Dilation Values.

The best result when the dilation rate is 1 with lowest MMAE error rate

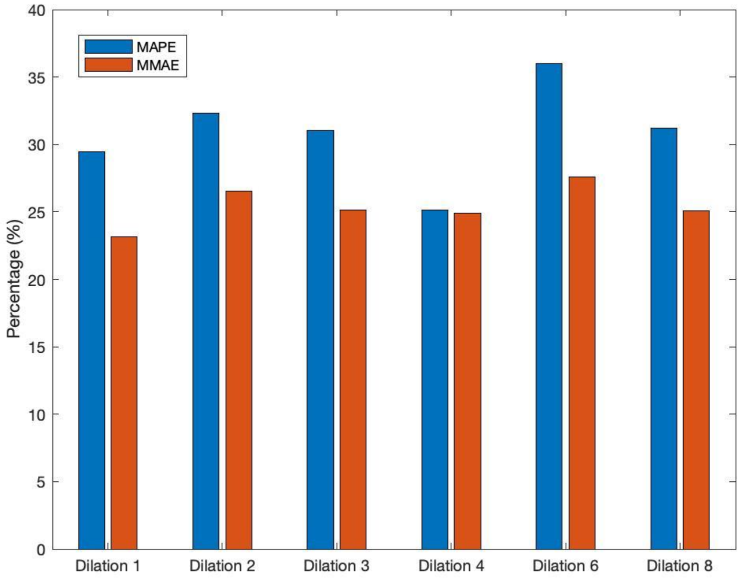


Figure 6

Results : CSRNet

- ➔ Figure 7 shows the MSE of each epoch while training CSRNet with augmented data
- ➔ Before the augmentation;
  - ◆ MAPE as 29.48% and MMAE as 23.14%.
- ➔ After the augmentation by cropping images;
  - ◆ MAPE as 32.80% and MMAE as 25.14%.
- ➔ After the augmentation by flipping the images left to right;
  - ◆ MAPE as 30.74% and MMAE as 23.35%
- ➔ After the augmentation by grayscale and cropping one image into 9 different patches with size one fourth of original image
  - ◆ MAPE as 31.56%, and MMAE as 23.85%

The augmentation didn’t work for our CSRNet model.

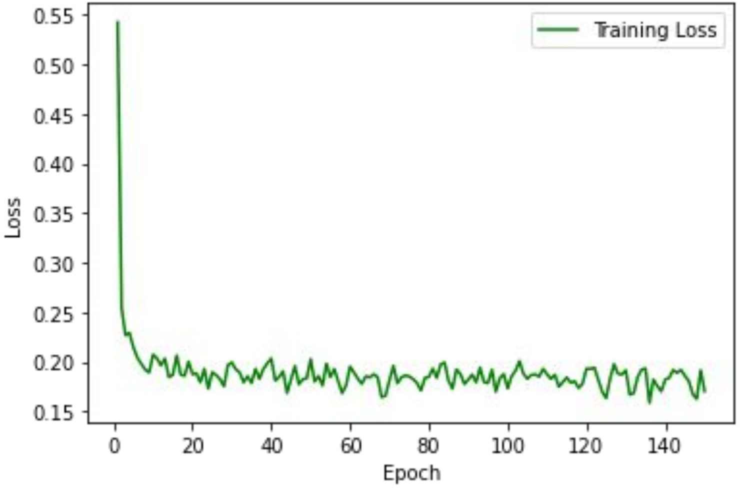


Figure 7

Results: MCNN

- Figure 8 shows the Training Loss of each epoch while training the MCNN model.
- For small dataset;
  - MAPE as 73.69% and MMAE as 49.74%
- The model trained with 1016 and tested with 182 different images. (Part A train, Part B train and test images)
  - MAPE as 53.03% and MMAE as 41.47%.

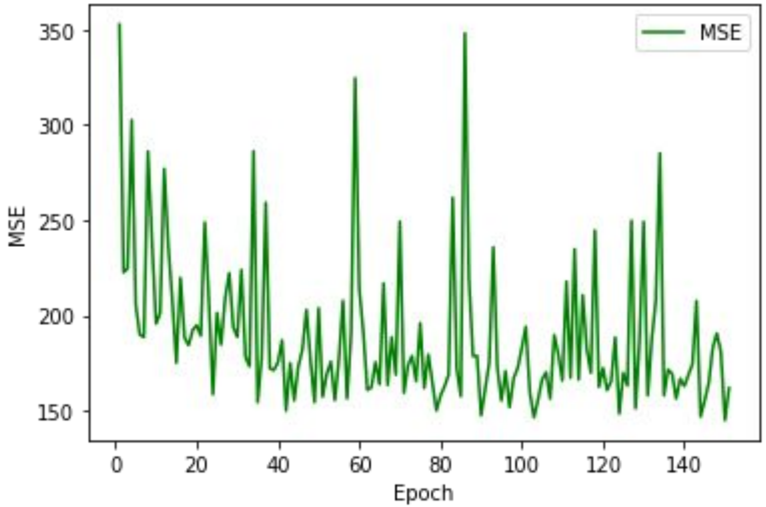


Figure 8

Results: MCNN

- ➔ Figure 9 shows the MSE of each epoch while training CSRNet with augmented data
- ➔ Before the augmentation;
  - ◆ MAPE as 73.69% and MMAE as 49.74%
- ➔ After the augmentation by grayscale and cropping one image into 9 different patches with size one fourth of original image
  - ◆ MAPE as 33.76% and MMAE as 31.00%

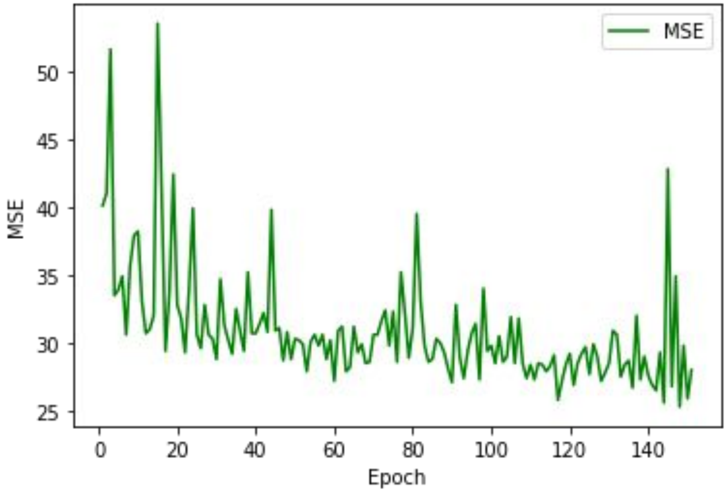


Figure 9

Discussion

- Comparison of the CSRNet and MCNN
    - MCNN takes less time
    - CSRNet has lower error rates because of the depth of the convolutional layers
- Data Augmentation
    - MCNN model improved tremendously
    - CSRNet is affected negatively
- Small Dataset vs Large Dataset
    - Larger datasets improved accuracy and performance of the models
    - Larger datasets increases the training time as expected