Context-Aware Crowd Counting

Tushaar Ranganathan

October 2024

https://github.com/Tushaar-R/Context-Aware-Crowd-Counting

1 Results

1.1 VGG-19 Architecture

- The results obtained by previous pre-trained VGG-16 model in the first 100 epochs gave best result of 75 MAE on test set of ShanghaiTech Part A.
- The results obtained by using current VGG-19 model gave best result of 73.78 MAE in the first 100 epochs on test set of ShanghaiTech Part A.
- It is also of interest that the MAE after epoch 1 on using pre-trained VGG-16 model was 368, whereas the MAE after epoch 1 on using pre-trained VGG-19 model was 261.

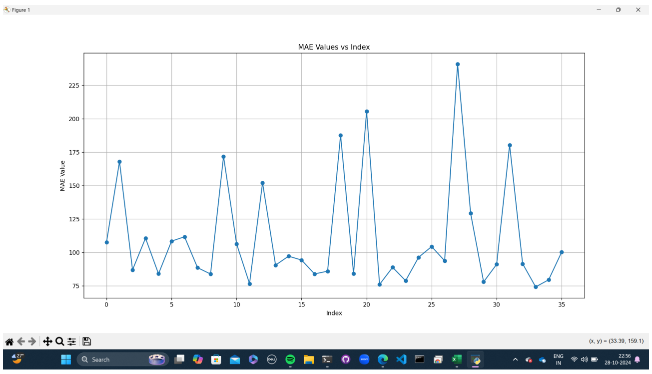

Figure 1: Results of VGG-16 on test set

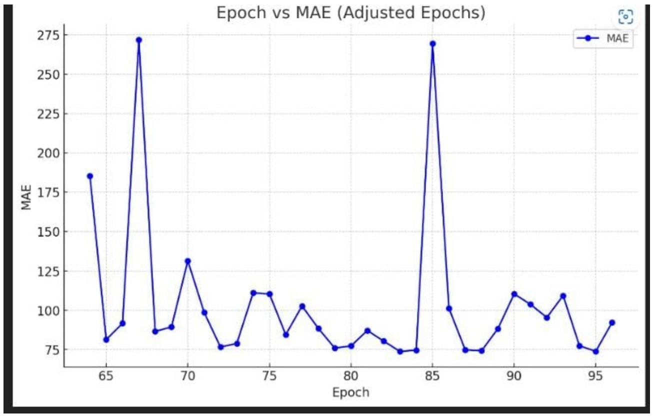

Figure 2: Results of VGG-19 on test set

Figure 3: VGG-16 v/s VGG-19 test MAE from epochs 65 to 100

- As can be inferred from the graphs, using pre trained VGG-19 model as a replacement to VGG-16 model improves the stability of the results,specially for the initial epochs. It can be noticed that there are relatively fewer spikes in the MAE on test results when using VGG-19, indicating that it's weights are in a relatively "smoother" region of the search space. It makes the entire model robust and performs better as well.
- The train time of VGG-16 and VGG-19 per epoch differs by 2-3 seconds as the models are pre-trained and only required fine tuning. Thus VGG-19 performs much better than VGG-16 in situations of limited computational power.

1.2 Divide and Conquer Strategy

- Approach:
  Made a clone of an image, and set all pixel values not in area of interest(say top right half) to zero in the cloned image.Then get resultant ground density maps from model and only sum the densities in area of interest of ground truth density map for apt comparison.
  Across 10 images in test set, I got an average difference of 8.05 between original predictions and post implementation of strategy predictions.

Across 30 images in test set, I got an average difference of 6.78 between original predictions and post implementation of strategy predictions.
Across entire test set, I got an average difference of 8.275 between original predictions and post implementation of strategy predictions.

- Results:
  The resultant density of predictions decreases when we take a portion of the image and erase it's surroundings as proven mathematically in the pervious report.

- Mathematical proof:
  Area of interest:Represents the area in the image in which density of crowd is predicted.

  Let $x = x_0$ represent a border of the area of interest around which we crop the image. Let $f(x, y)$ be the original density of people predicted at a point near the border of the image. Let $f'(x, y)$ be the density predicted after cropping the image at its borders.

  Initially, we had a density of $f(x_0 + h, y_0)$ (where $h \to 0$) on the other side of the border. However, after cropping and reducing all pixel values to 0 (effectively blackening the rest of the picture), the current human density on the other side of the border is now $f'(x_0 + h, y_0) = 0$.

  Since $f(x, y)$ and $f'(x, y)$ both represent continuous density maps, $g(x, y) = f(x, y) - f'(x, y)$ also represents a continuous density map. Thus,

  $$g(x_0 + h, y_0) = f(x_0 + h, y_0) - f'(x_0 + h, y_0) = f(x_0 + h, y_0) > 0.$$

  Also, $g(x_0, y_0) \approx g(x_0 + h, y_0) = f(x_0 + h, y_0) \approx f(x_0, y_0) > 0$.

  Thus, $f(x_0, y_0) > f'(x_0, y_0)$.
  Where (x_0,y_0) represent a point on the border of the area of interest.