AIP Assignment-2

N-Cut :

- N-cut algorithm for Sematic Segmentation with N=2
- Two different similarities are used for assigning weights to graph edges

$$w_{ij} = e^{\frac{-\|F(i)\ \ F(j)\|_2^2}{\sigma_I}} * \begin{cases} e^{\frac{-\|X(i)\ \ X(j)\|_2^2}{\sigma_X}} & \text{if } \|X(i) - X(j)\|_2 < r \\ 0 & \text{otherwise,} \end{cases}$$

1) Pixel intensity values

   - $F(i) = I(i)$, the intensity value, for segmenting brightness images,

2) Pixel HSV values based

   - $F(i) = [v, v \cdot s \cdot sin(h), v \cdot s \cdot cos(h)](i)$, where h, s, v are the HSV values, for color segmentation,

Steps in finding the Segmented image :

- All the images are resized to the size of 70*70 for faster computation of the weight graph of the image.
- Find the the eigen vector corresponding to the first non-zero generalised eigenvalue of matrices D-W, D.
- Used threshold of 0, to decide whether each pixel belongs to segment 1 or not.

Note :

- Seg1 corresponds to similarity measure 1).
- Seg2 corresponds to similarity measure 2).

The obtained results are as follows

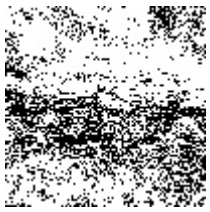
test1.jpg:

	Original	Rotated	Noisy
Seg1			
Seg2			

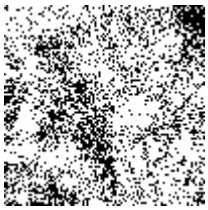
test2.jpg:

	Original	Rotated	Noisy
Seg1			
Seg2			

test3.jpg:

	Original	Rotated	Noisy
Seg1			
Seg2			

test4.jpg:

	Original	Rotated	Noisy
Seg1			
Seg2			

test5.jpg:

	Original	Rotated	Noisy
Seg1			
Seg2			

Observations :

- The similarity measure based on the pixel intensity values is performing better than the similarity measure based on the pixel HSV values.
- We can observe that N-Cut segmentation is sensitive to noise. Because when the image contains noise, it introduces the fluctuations in the pixel intensities, furthur disturbing the weight graph W that was constructed.
- N-cut is invariant to rotation, as when we rotate the image, the neighbourhood of the original and rotated remains the same.

Thus it is observed that the N-Cut algorithm-based segmentation is invariant to rotation and is also sensitive to the addition of noise into the image as it affects the similarity measures between 2 nodes.

N-cut segmentation algorithm is taking long time to compute the weight graph for the images of larger size.

Resnet50 pretrained FCN :

- Used ResNet 50 based FCN pretrained on PASCAL VOC dataset.
- The annotations consists of 20 classes, each given a unique colour and background is considered as 21st class.
- The obtained pixelwise accuracy and mean IOU on the test dataset is as below,
    - Achieved Pixel wise accuracy `0.9163349391005375` and mean IOU `0.64002615`

MobileNetv2 backbone pre-trained model

- The MobileNetv2 pretrained on the ImageNet Classification dataset is used for segmentation by converting into FCN network.
- The last fc layer of 1000 neurons was removed, and added 1*1 convolution layers at the end and upscaled the images were appropriately to match the input image size.
- Training parameters :

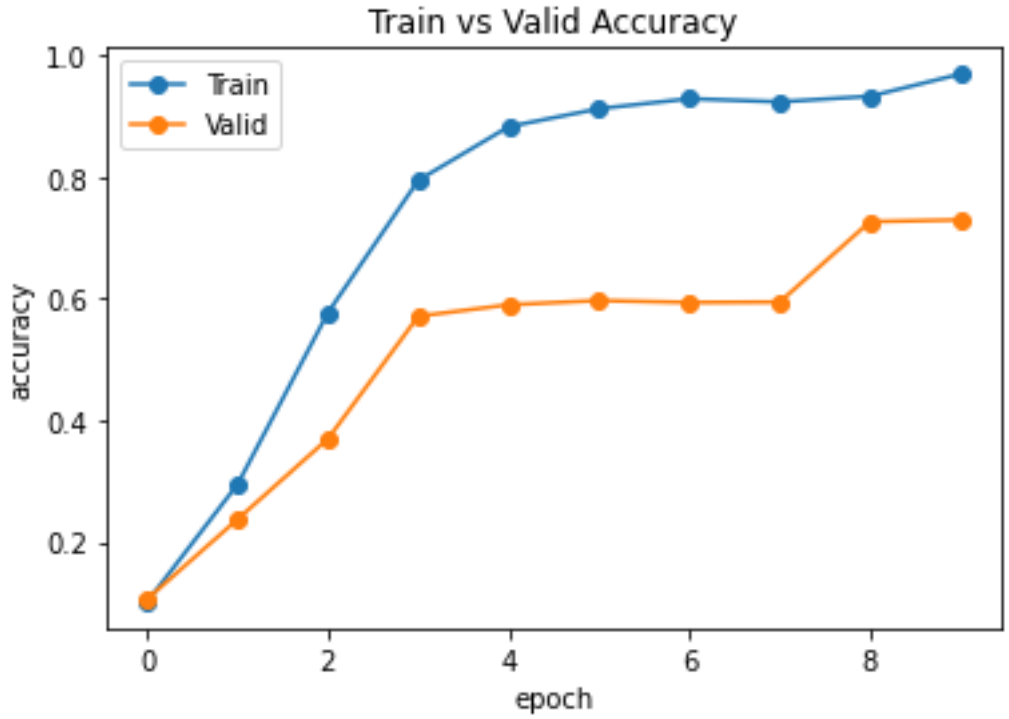
  Cross-entropy loss and Adam optimizer are used with lr = 0.002.

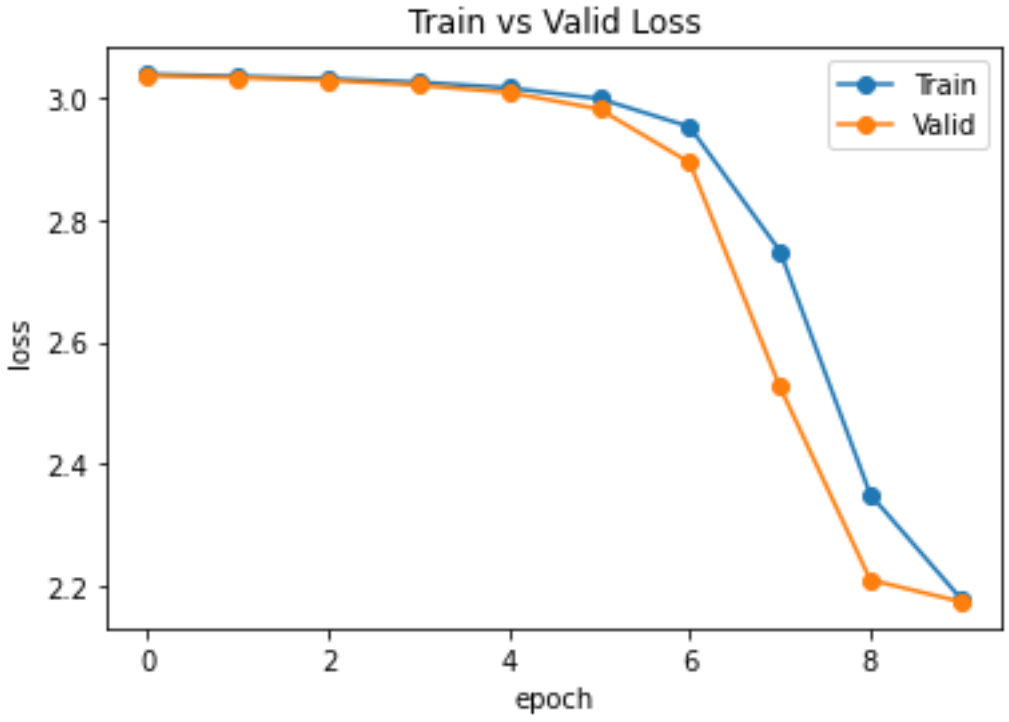
- It was observed that adding skip connections has enabled the model to learn the annotations better.
- Skip connections are created at layers 4, 7, 14 and 19.
- Deep features can be obtained when we go deeper, but location information is lost when going deeper. Fusing the output with the upsampled version of the output at the previous skip connection will enable gaining the location information.
- Trained the model for 10 epochs and achieved the following results on the test images

  Pixel-wise accuracy: *0.72*

  Mean IOU: *0.52*

Plots :


Train vs Valid Accuracy
Train
Valid
accuracy
epoch
1.0
0.8
0.6
0.4
0.2
0
2
4
6
8

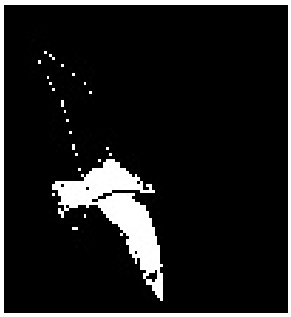
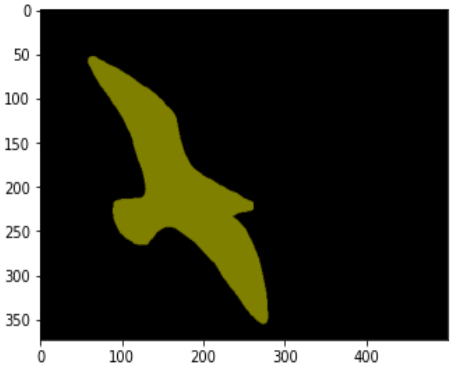
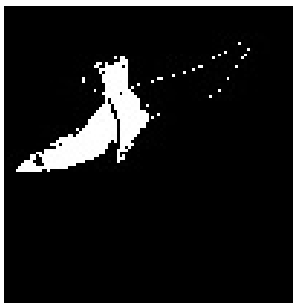
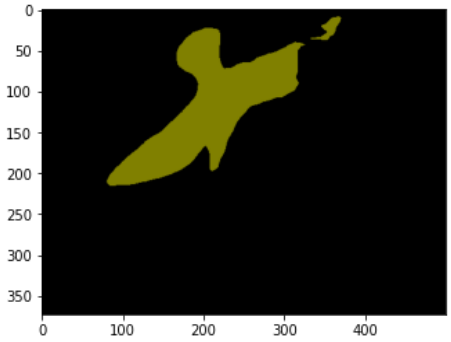
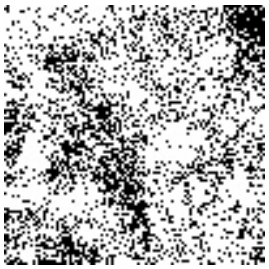
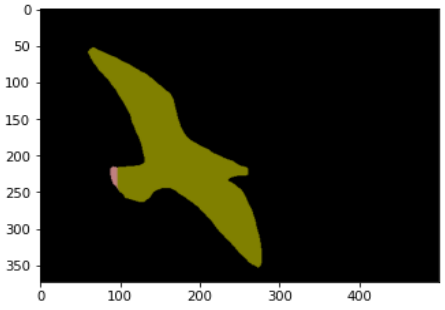

Train vs Valid Loss
Train
Valid
loss
epoch
3.0
2.8
2.6
2.4
2.2
0
2
4
6
8

Comparison between ResNet50-based FCN and MobileNetv2-based FCN

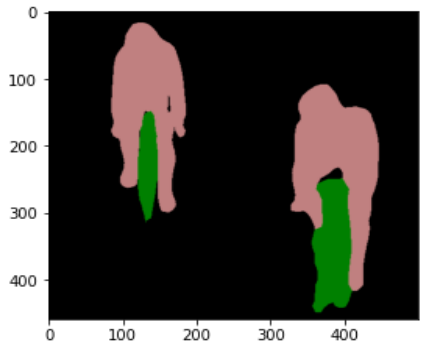
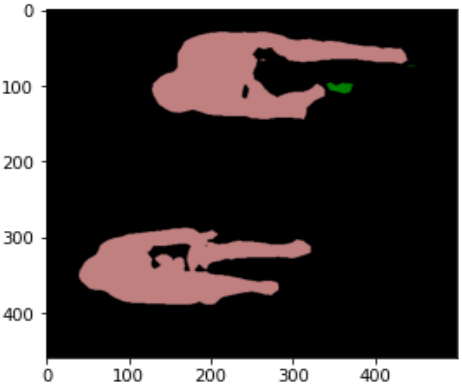
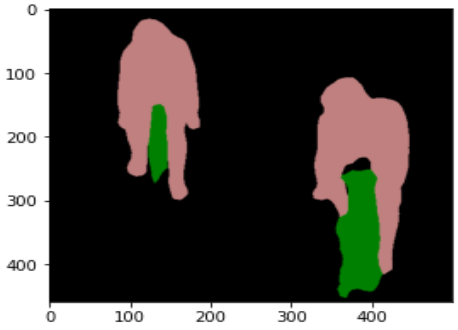
- The accuracy of the MobileNetv2 model and the Mean IOU is lesser compared with the ResNet50 based FCN.
- Since the Resnet50 was trained on the PascalVOC dataset and the architecture of the model is also designed specifically for semantic segmentation, it was able to capture the features and differentiate the pixel values better.
- Though the pretrained MobileNetv2 model was trained for the purpose of image classification, finetuning the model by adding appropriate the backbone layers was able to get good IOU 0.52.
- It can also be observed that MobileNetv2 is a lightweight model and can be trained faster.

Comparing N-Cut and FCN based segmentations.

- **test4.jpg**

	N-Cut	FCN based
Original		
Rotated		
Noisy		

- **test5.jpg**

	N-cut	FCN
Original		
Rotated		
Noise		

- The N-Cut based algorithm doesn’t consider the colour information, which also plays major role in segmenting the image.
- The FCN-based algorithm is faster when compared to the N-cut algorithm. It follows from the fact that we need to build up the large size weight matrix in N-Cut algorithm whereas once we have trained the FCN model for sufficient epochs, it just requires one forward pass through the FCN model.
- Since FCN model involves CNN layers, the final layer pixel values are able to look at the features at the larger neighbourhood(receptive field) of input image, and can differentiate among them.