Assignment 03

Computer Vision

by
Rahul Kumar Chaudhary
(M20MA008)

Instructor : Mayank Vasta

Question 1: Object Recognition

Problem Statement : Use the CIFAR10 dataset for this question. Extract the following features from the images and train a 2-layer neural network for classification.

A. Local Binary Patterns (LBP) (To be done from scratch) (20 marks)
B. Scale-Invariant feature Transform (SIFT) (10 marks)
C. Histogram of gradients (HOG) (10 marks)
D. Deep features (Use AlexNet pre-trained on Imagenet to extract the 4096-dimensional feature vector) (10 marks)
E. Deep features (Use ResNet18 pre-trained on Imagenet to extract feature vector)
F. Report the classification results and do some comparative analysis among them.

A. Local Binary Patterns (LBP) (To be done from scratch)

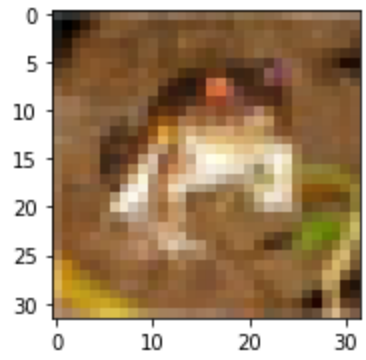
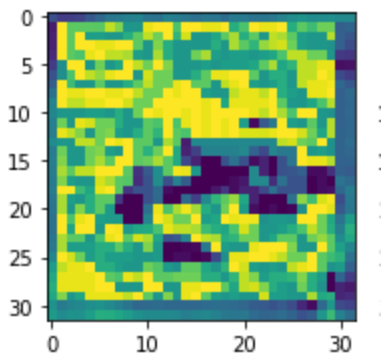
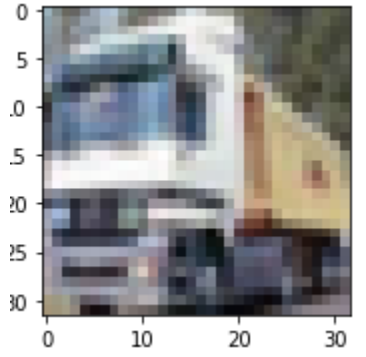
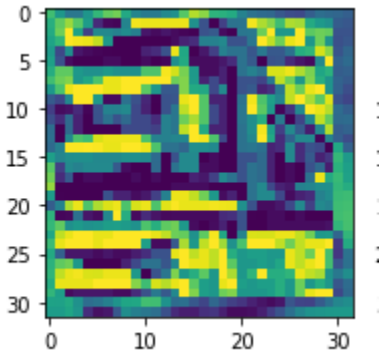
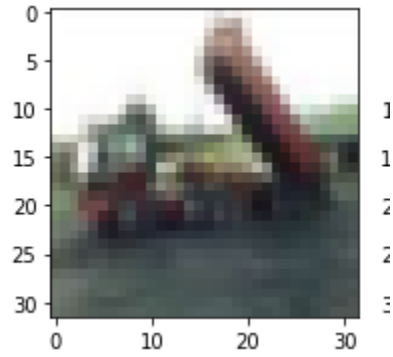
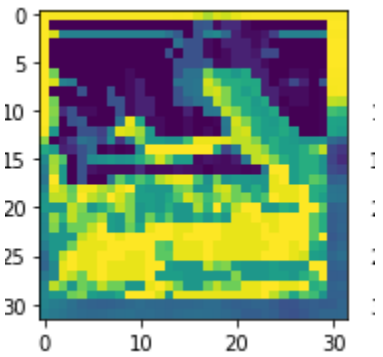
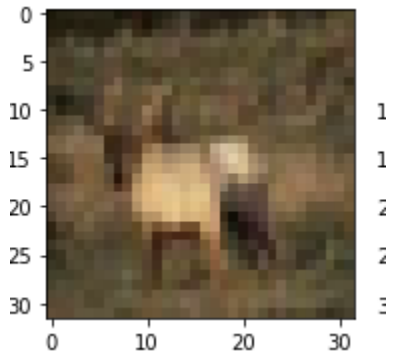
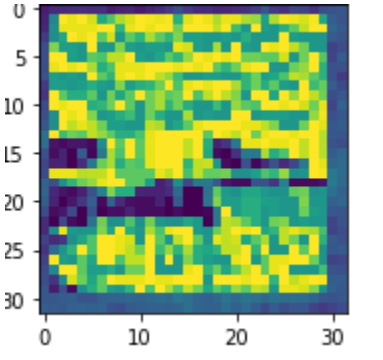
Dataset : Dataset used in this problem is *CIFAR-10* dataset. Table 1.1 shows all the information related to dataset

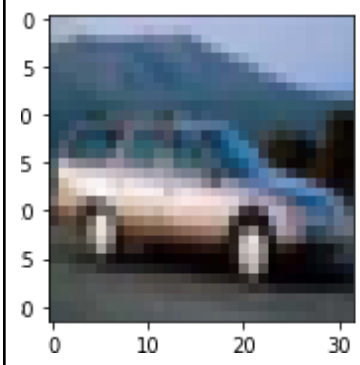
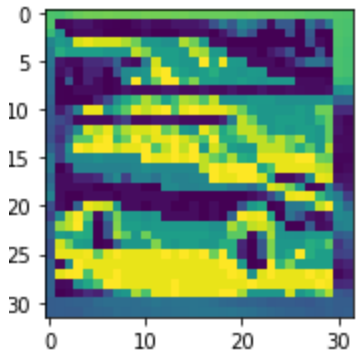
Dataset Name	*CIFAR-10*
Number of Instances	*50000 Training, 10000 Test*
Associated Tasks:	*Classification*
Features	*Image size : 3*32*32*
Missing Value	*NONE*
Total Number of Classes	*10*

Table 1.1 Dataset Details

For this part, local binary pattern (LBP) has been implemented with mask size 3×3. Here is the feature image extracted using a local binary pattern.

Laber	Original Image	LBP Feature

Frog		
truck		
truck		
deer		

automobile		

Here are the hyperparameter of two layer neural network which used to train and classify the CIFAR10 dataset

Hyperparameters	*Values*
Number of Epoch	20
Batch Size	10
Initial Learning Rate	0.001
Optimizer	Adam
Loss	Cross Entropy

Here are the different class accuracy obtained from the training.

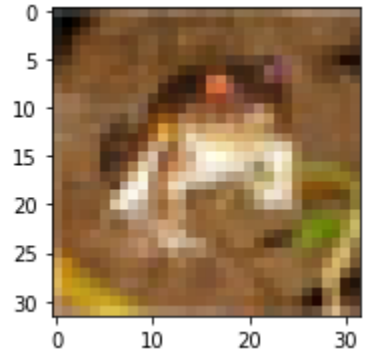
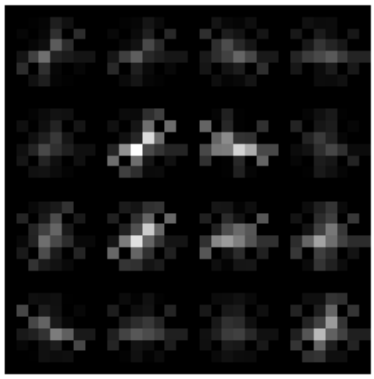
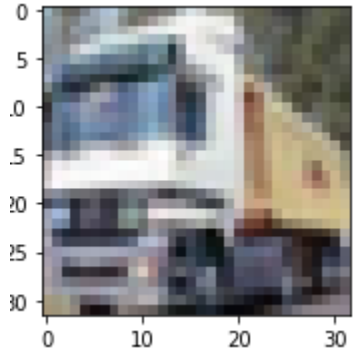
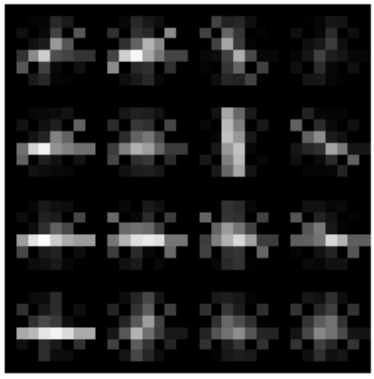
Accuracy of Network	**42.57%**
Class	Accuracy
Plane	49.1%
Car	53.3%
Bird	12.5%
Cat	36.9%
Deer	30.5%
Dog	61.8%
Frog	57.9%

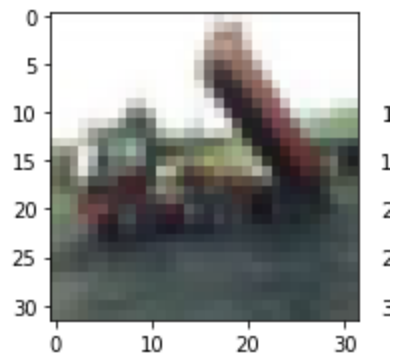
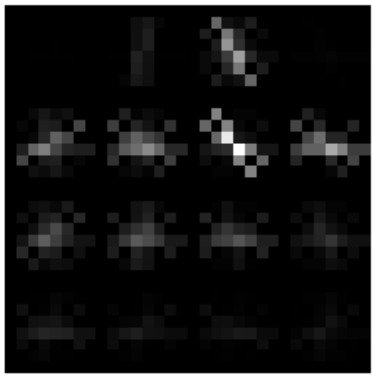
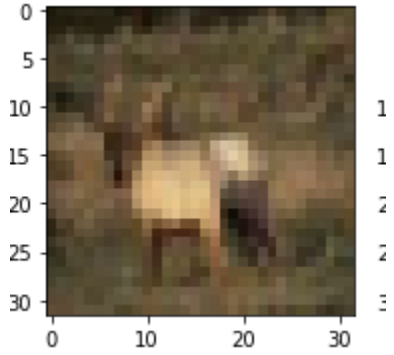
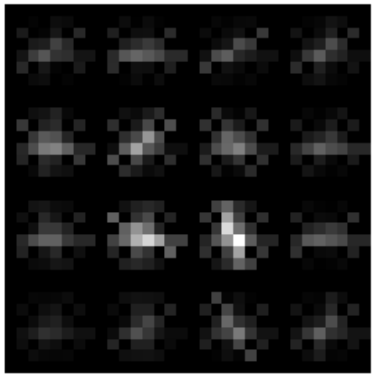
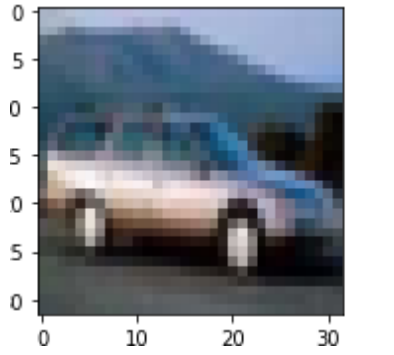
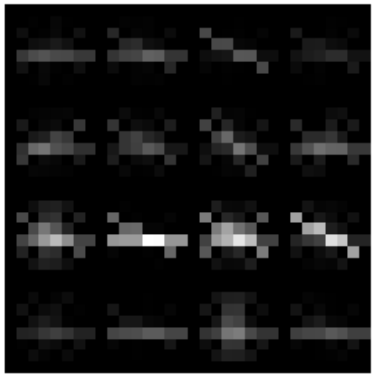
Horse	49.6%
Ship	68.0%
Frog	57.1%

B. Scale-Invariant feature Transform

C. Histogram of Gradient

For this part, local binary pattern (LBP) has been implemented with mask size 3×3. Here is the feature image extracted using a local binary pattern.

Laber	Original Image	LBP Feature
Frog		
truck		

truck		
deer		
automobile		

Here are the hyperparameter of two layer neural network which used to train and classify the CIFAR10 dataset

Hyperparameters	*Values*
Number of Epoch	20
Batch Size	10

Initial Learning Rate	0.001
Optimizer	Adam
Loss	Cross Entropy

Here are the different class accuracy obtained from the training.

Accuracy of Network	**10%**
Class	Accuracy
Plane	0
Car	0
Bird	0
Cat	100%
Deer	0
Dog	0
Frog	0
Horse	0
Ship	0
Frog	0

D. Deep features (Use AlexNet pre-trained on Imagenet to extract the 4096-dimensional feature vector)

Here pretrained AlexNet has been used to extract 4096 dimensional vectors, and then it is fed to a two layer neural network for classification. Here is the details of hyperparameter and accuracy of the model along with the each class accuracy

Hyperparameters (AlexNet)	*Values*
Number of Epoch	2
Batch Size	10
Initial Learning Rate	0.01
Optimizer	Adam

Loss	Cross Entropy

Hyperparameters (Two layer Neural Network)	*Values*
Number of Epoch	20
Batch Size	10
Initial Learning Rate	0.001
Optimizer	Adam
Loss	Cross Entropy

Here are the different class accuracy obtained from the training.

Accuracy of Network	**82.65 %**
Class	Accuracy
Plane	96.6
Car	81.9
Bird	78.8
Cat	80.4
Deer	70.6
Dog	83.3
Frog	80.74
Horse	89.2
Ship	81.6
Frog	86.7

E. Deep features (Use ResNet18 pre-trained on Imagenet to extract feature vector)

Here pretrained AlexNet has been used to extract 4096 dimensional vectors, and then it is fed to a two layer neural network for classification. Here is the details of hyperparameter and accuracy of the model along with the each class accuracy

Hyperparameters (AlexNet)	*Values*
Number of Epoch	2
Batch Size	10
Initial Learning Rate	0.01
Optimizer	Adam
Loss	Cross Entropy

Hyperparameters (Two layer Neural Network)	*Values*
Number of Epoch	20
Batch Size	10
Initial Learning Rate	0.001
Optimizer	Adam
Loss	Cross Entropy

Here are the different class accuracy obtained from the training.

Accuracy of Network	**82.91 %**
Class	Accuracy
Plane	96.6
Car	96.5
Bird	78.2
Cat	69.9
Deer	85
Dog	80
Frog	86.3

Horse	76.6
Ship	90.5
Frog	76.4

F. Report the classification results and do some comparative analysis among them.

Here is the comparative report of all above feature extractor technique with their respective overall accuracy report.

Feature Extractor	Number of Epoch	Accuracy
Local Binary Pattern	20	42.57
SHIFT	20	37
HOG	20	10
resNet18	20	82.91
alexNet	20	82.65

Here resNet18 is giving best performance among all other feature extractor.

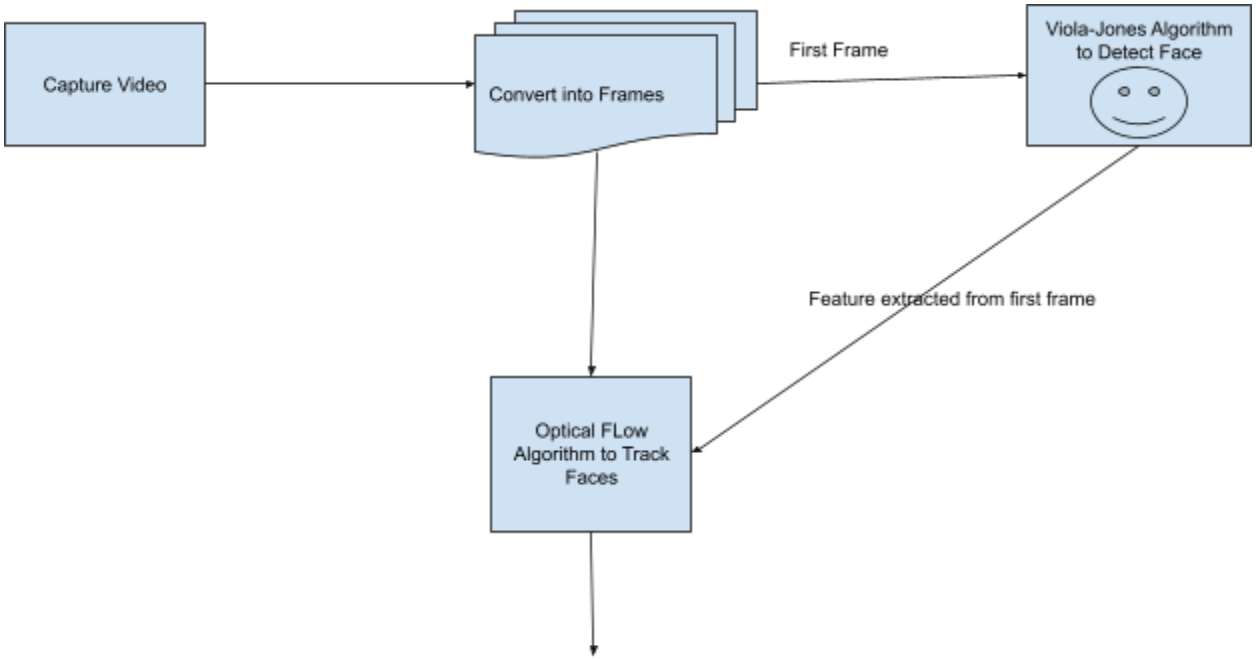
Note：Accuracy of all the above models was very much sensitive to learning rate and number of epochs, and was giving very different results each time I executed the model.

Question 2: : Face detection and Tracking

Problem Statement : Capture a video of a person whose face is visible most of the time. Use a pre-trained Viola-Jones face detector to detect the face in the first frame and after that use Optical Flow algorithm to continue tracking it in the rest of the video frames. Report the intermediate results and explain the working of the algorithm. You can test the algorithm with more videos.

Captured Video Link :

Here is flow diagram of the of implementation of the problem above

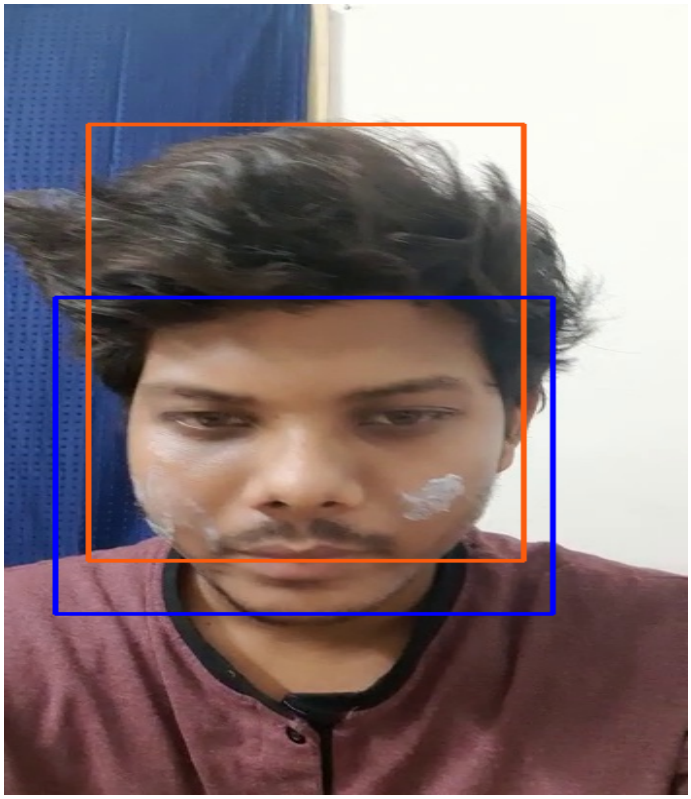


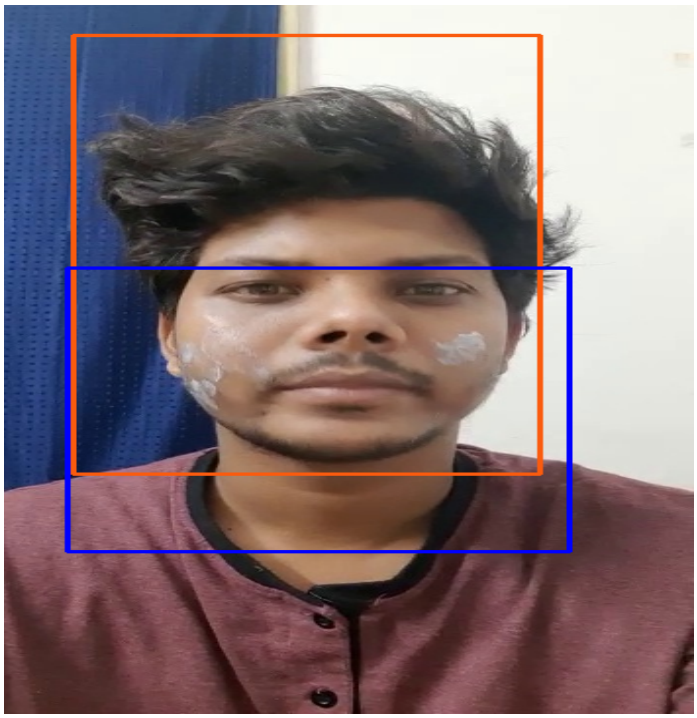
Here is the output of Viola-Jones of the first fram

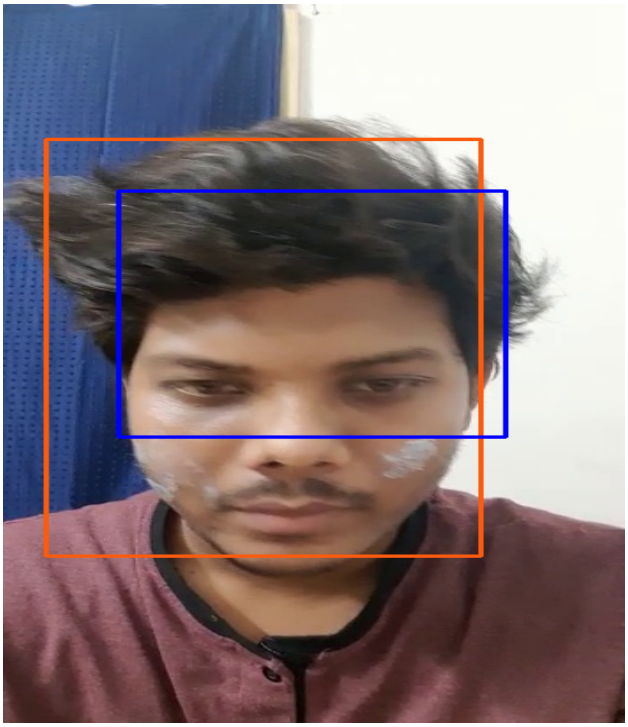
Original Frame

Output of Viola-Jones

Here are the few output frames of the Optical Flow algorithm.







Here the orange box is the tracked face by Optical flow algorithm while blue face detected by Viola-Jones face detector.

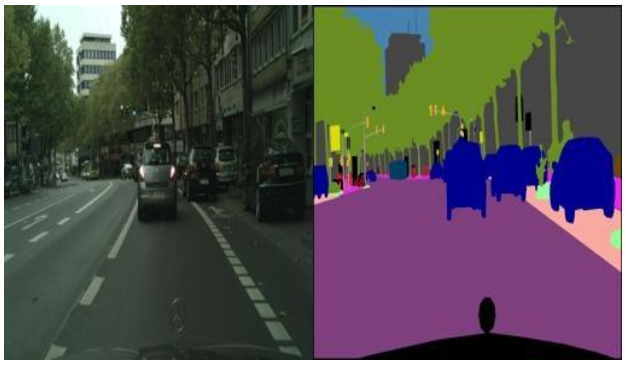
As we can clearly see there is no 100% overlap between the two outputs as shown in the above figure.

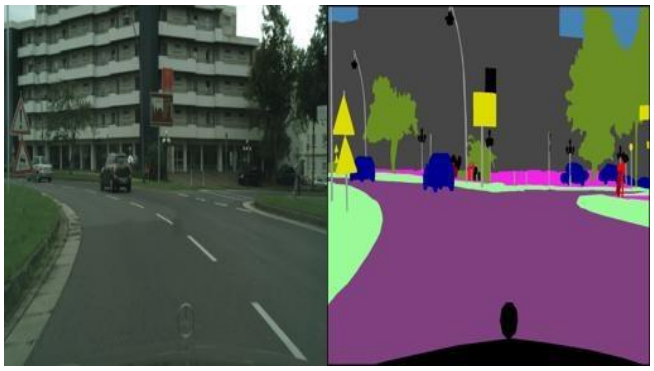
Conclusion : There is a mismatch between the output of the optical flow algorithm and the Viola-Jones face detection. This is because the optical flow algorithm can track small motion between any two frames and whenever there is large displacement of an object, optical flow will fail to detect it. That's the reason we go for a level optical flow algorithm. Also, here the feature extracted of the detected face from first frames by Viola-Jones face detector is just the pixel value which is optimal. We can implement feature extractor like LBP, SHIFT (as implemented in first part of this assignment) to get further more better result.

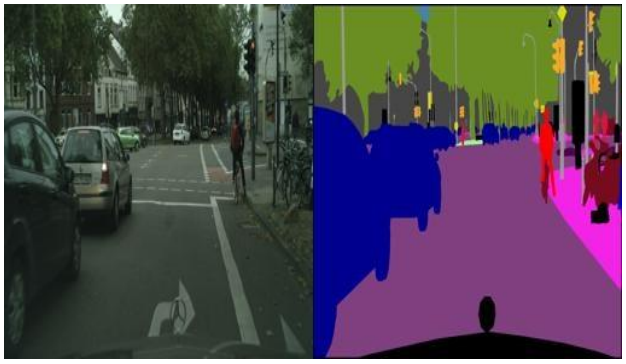
Question 3 : Semantic Segmentation

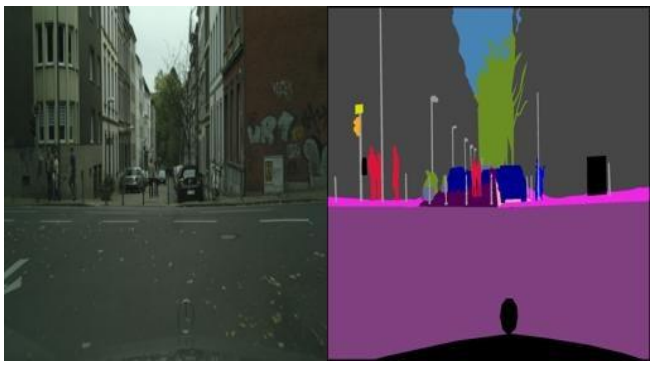
Problem Statement : Use the dataset for this question and train a FCN model for object segmentation. Report the obtained Jaccard Similarity Score.

We have been given training and validation dataset with annotated label. Here are the some of the input files.



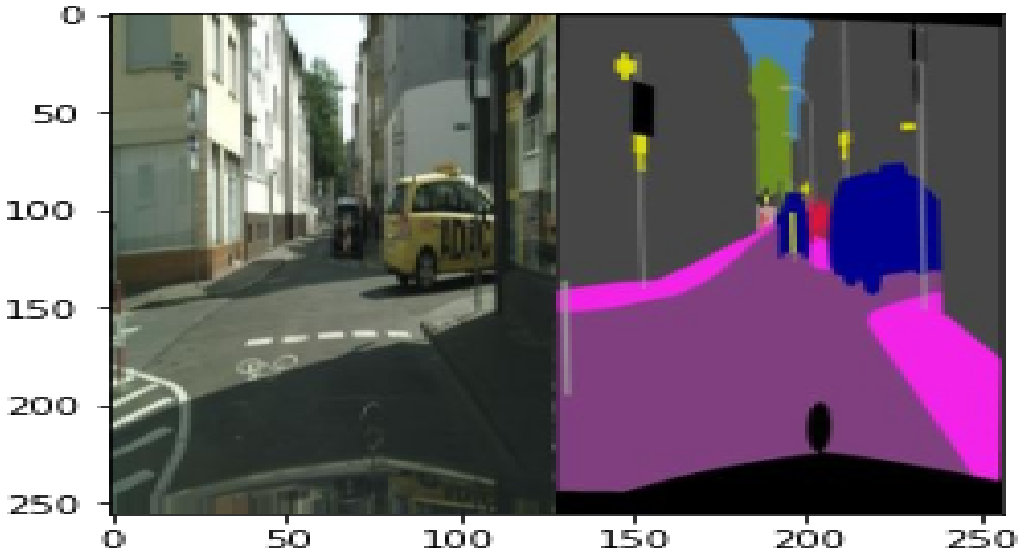




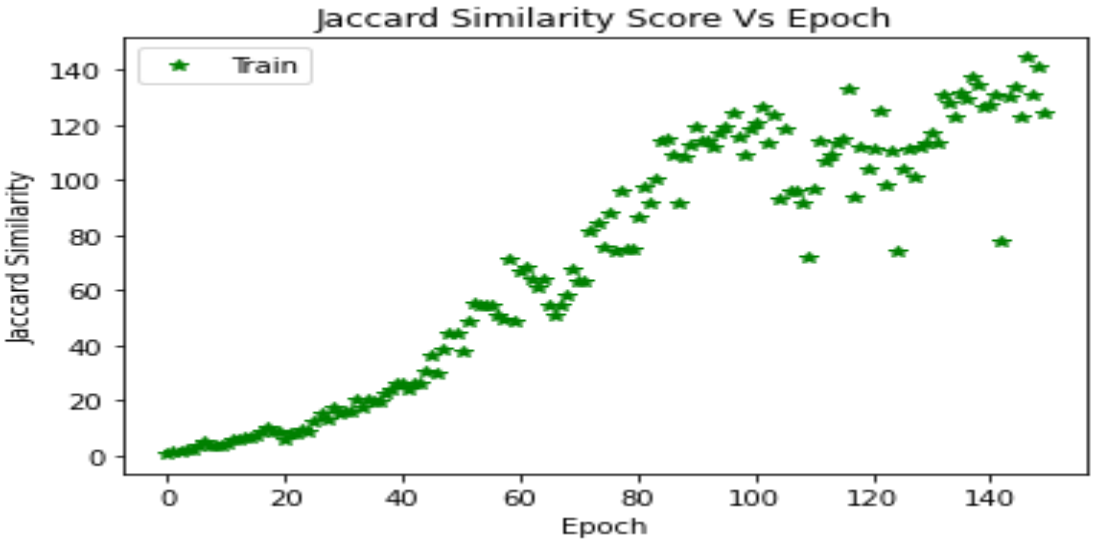


First job is to extract the images and mask from the given input file. Here are the separated file obtained after separating image and mask

Here is the here is the result of FCN segmentation on training data



Here is the the Jaccard Similarity plot for all training epoch.



Question 3 : Nail segmentation

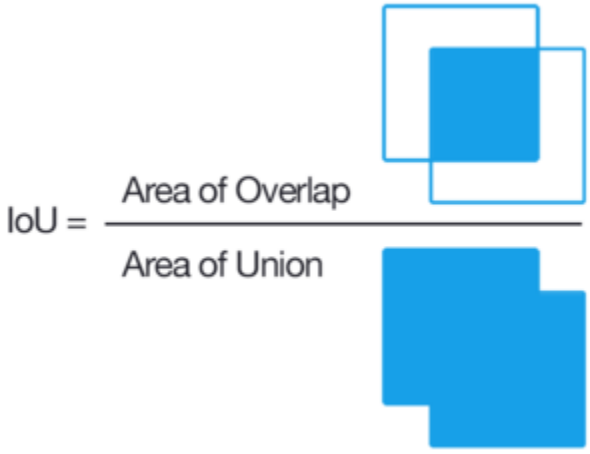
Problem Statement : Use the given dataset to segment nails. Train a U-Net architecture and report the dice-score and IOU. Also explain the difference between the two metrics. (Note: The dataset is small, so, consider using data augmentation in training.)

Here are the hyperparameter details used to train the UNet model for segmentation

Hyperparameters (Unet)	*Values*
Number of Epoch	100
Batch Size	32
Initial Learning Rate	0.01
Optimizer	Adam
Loss	bce_jaccard_loss

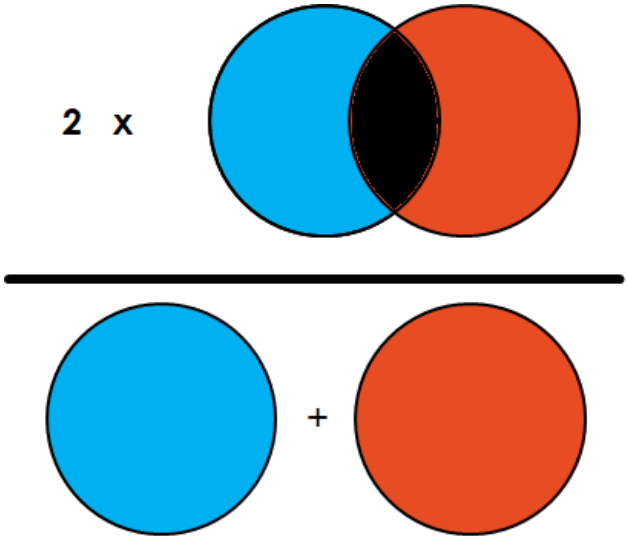
Metrics used for evaluation the model

Intersection-Over-Union : It gives the area overlapped divided by the total number of common pixels in both the image as shown in the figure below



Source : wikipedia

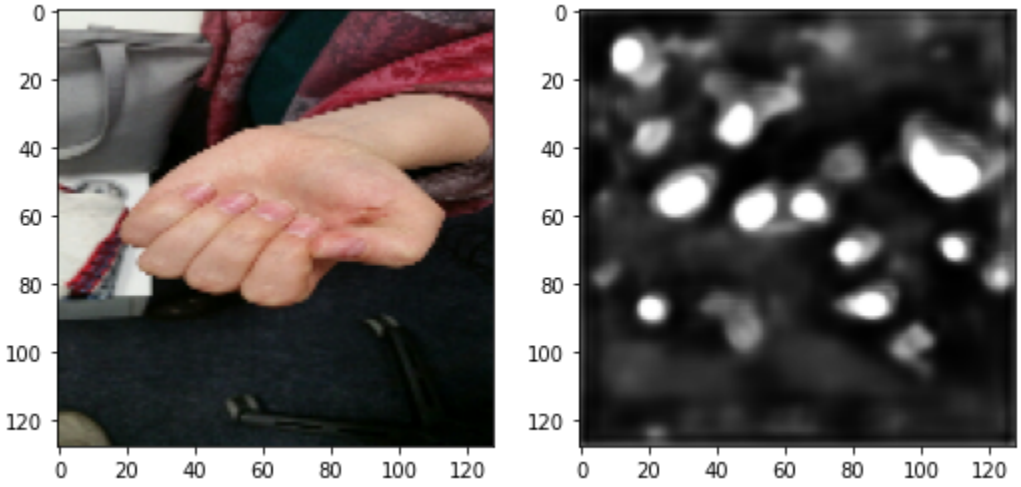
Dice Coefficient : Simply put, the **Dice Coefficient** is 2 * the Area of Overlap divided by the total number of pixels in both images.

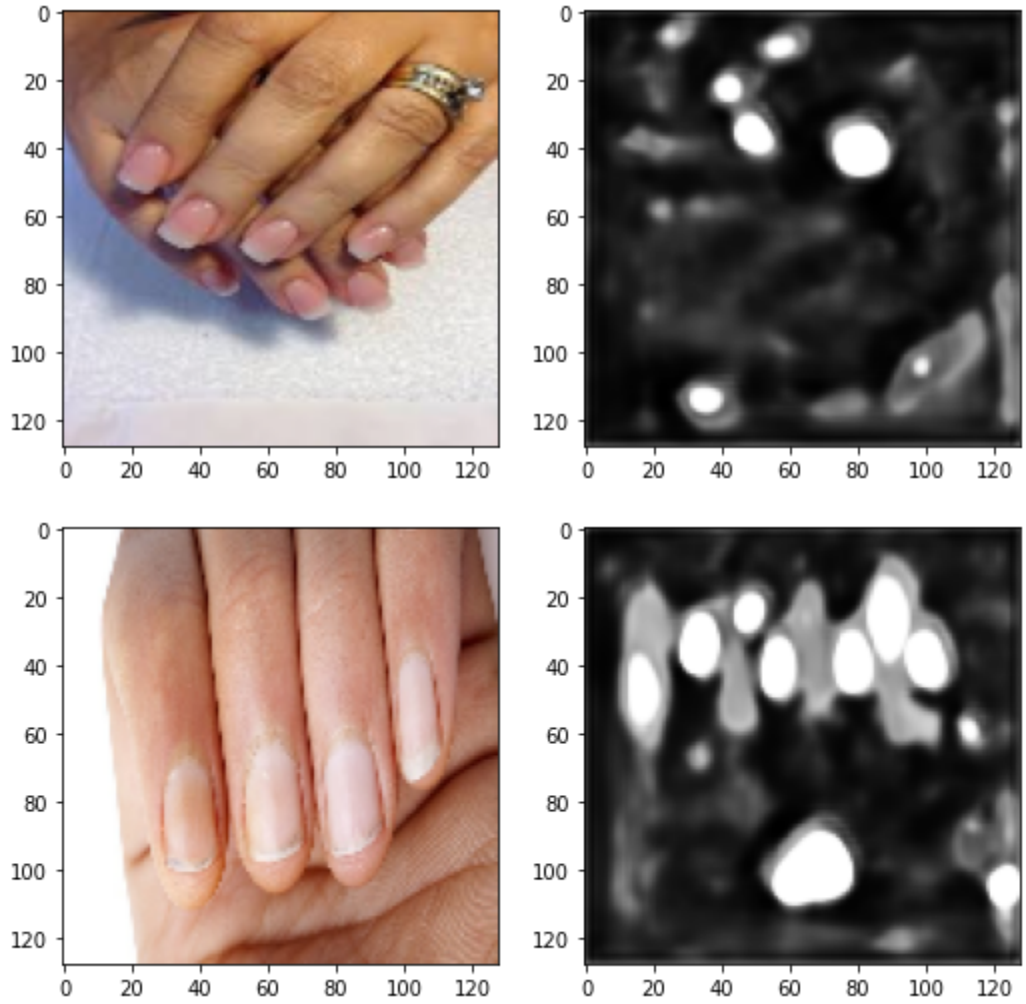


Dice Coefficient Explanation

Results and Discussion

Below are the obtained segmented image predicted by the model.



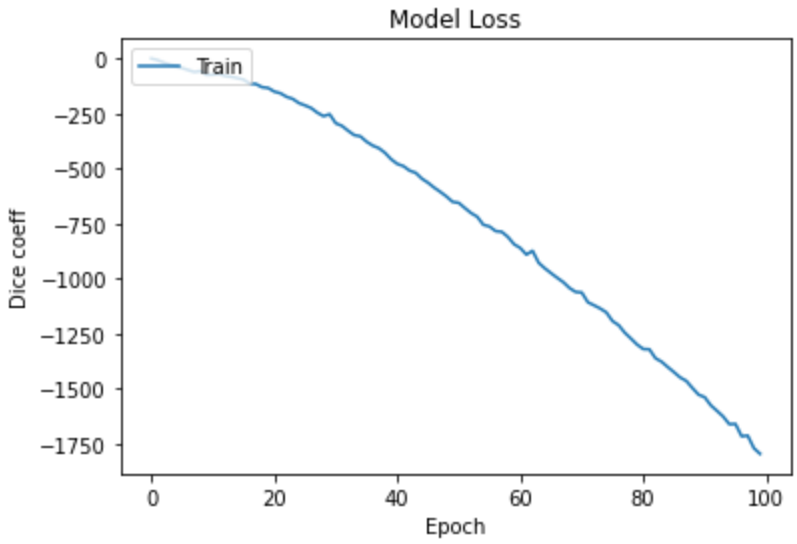


Below Tables gives the metrics value obtained.

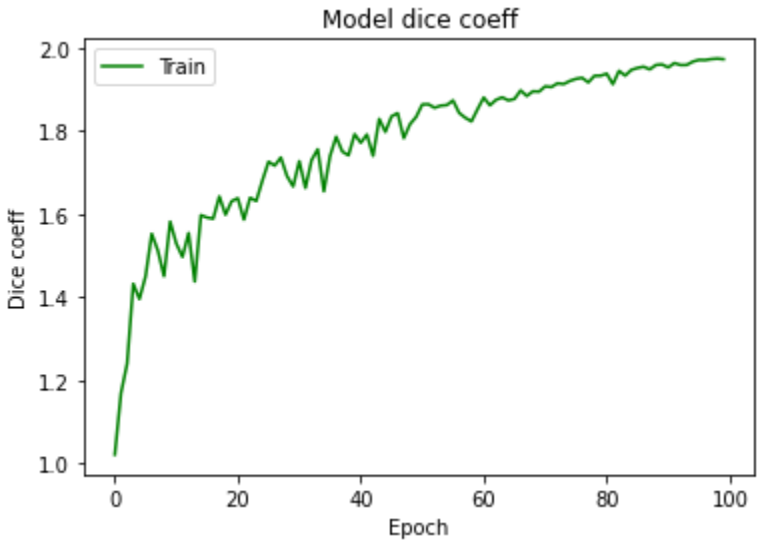
Metrics	Value
Dice Coefficient	1.95
IOU	.5224

Graph and Plot

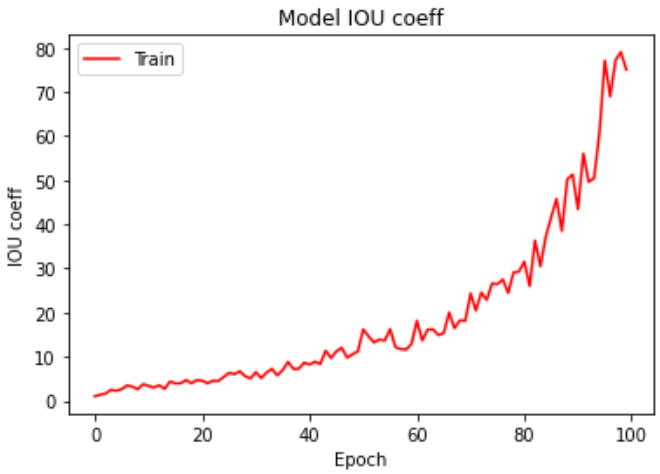
Here is plot of loss vs epoch



Plot of Dice coefficient vs Epoch



Plot of IOU vs Epoch



Note: Model dice coefficient and IOU score is very low which is reflected by the poor result generated by the model. This is mainly due to the small datasets. Data augmentation technique can be used to increase the training dataset and increase the model performance.