भारतीय विज्ञान संस्थान

# Artificial Intelligence

## ASSIGNMENT -2

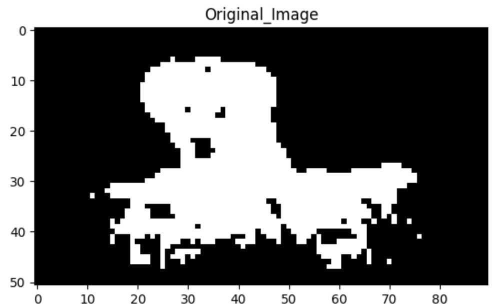
### Image Segmentation

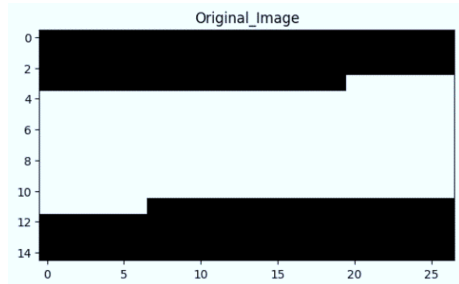
**SHUBHANKAR MONDAL** | AIP | SR-22456

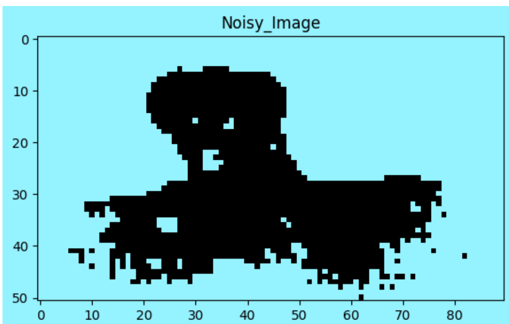
# Classical methods: N-Cut, K-means

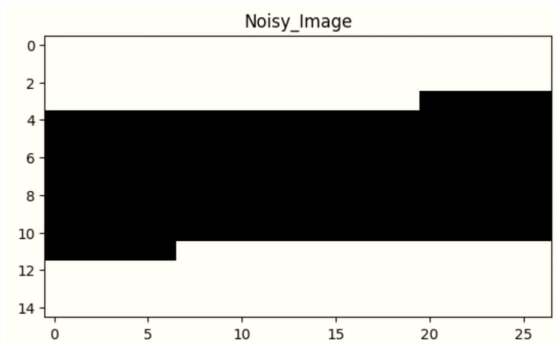
## N-cut using two partitions:

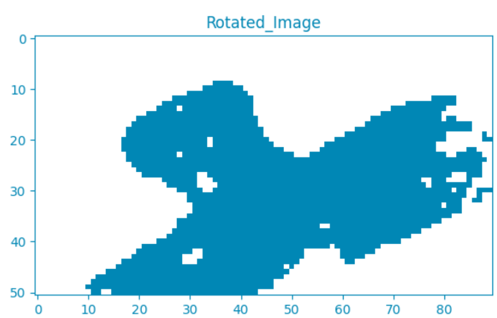
# Image1:

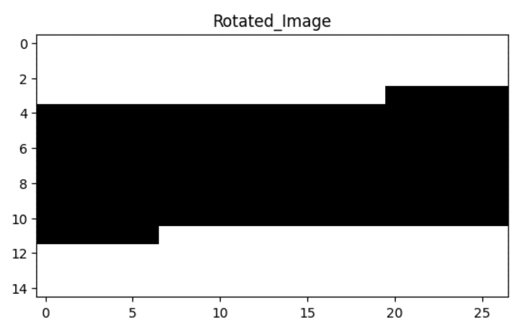


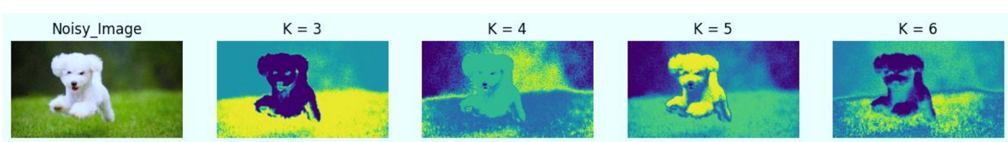



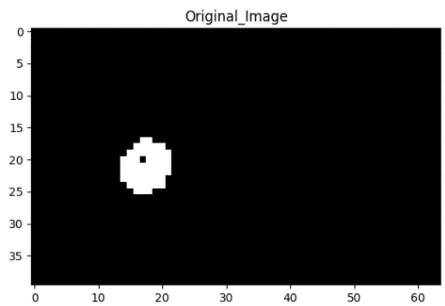



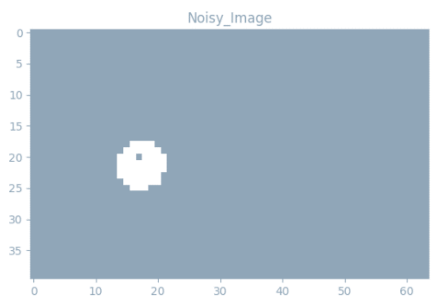


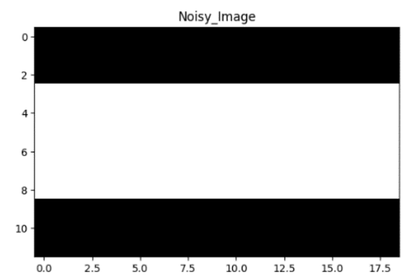



# Image2:

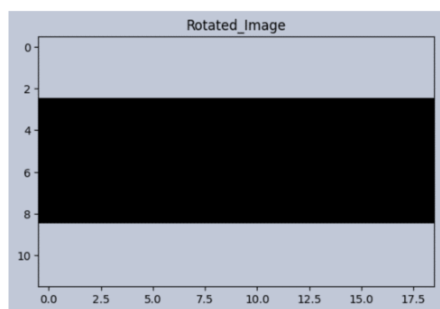

Original_Image

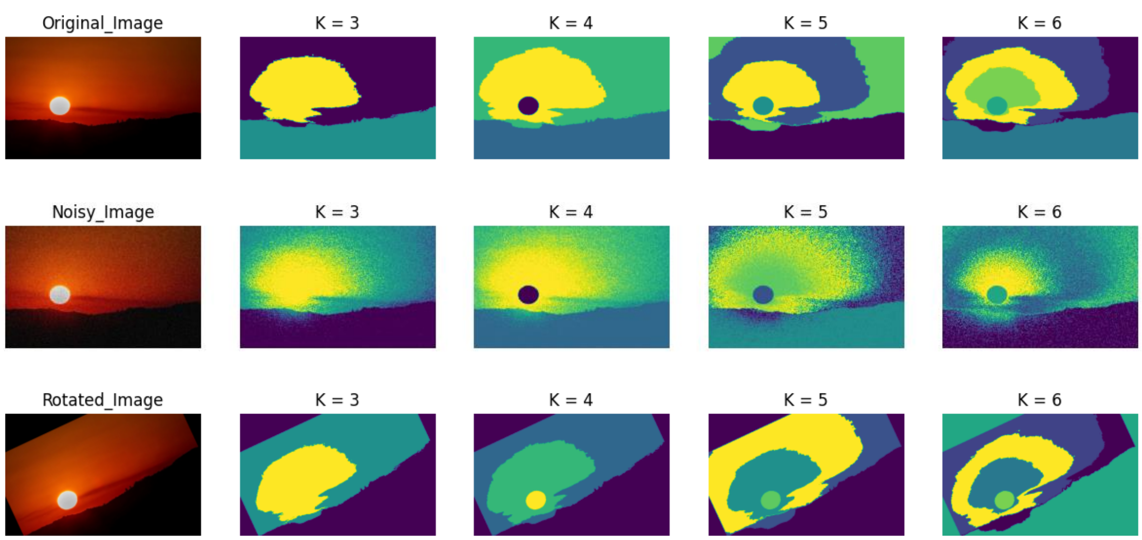

Original_Image


Noisy_Image

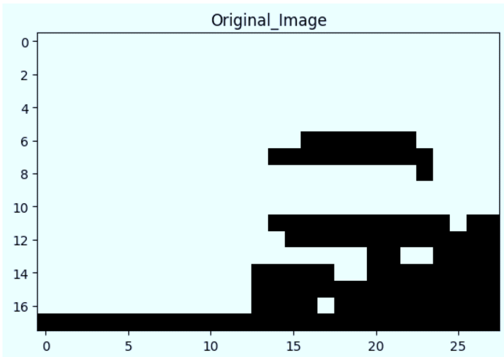

Noisy_Image

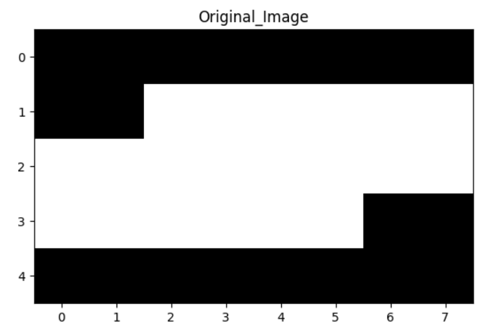

Rotated_Image

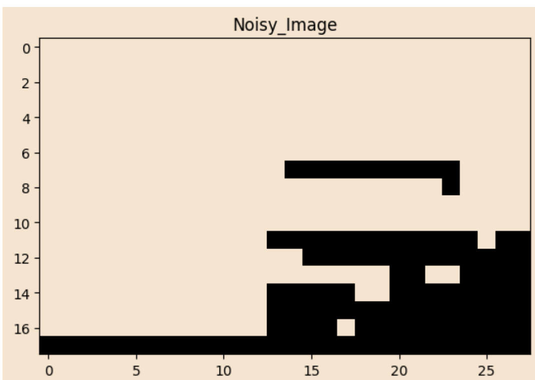

Rotated_Image

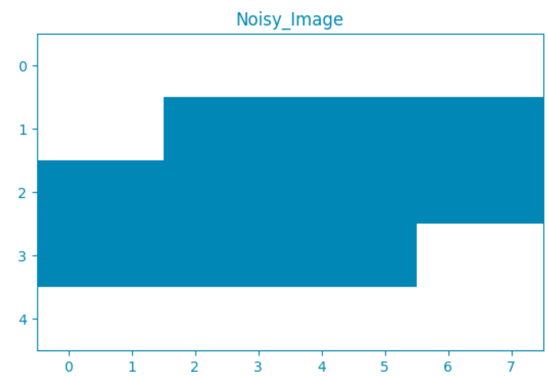

Original_Image
K = 3
K = 4
K = 5
K = 6
Noisy_Image
K = 3
K = 4
K = 5
K = 6
Rotated_Image
K = 3
K = 4
K = 5
K = 6

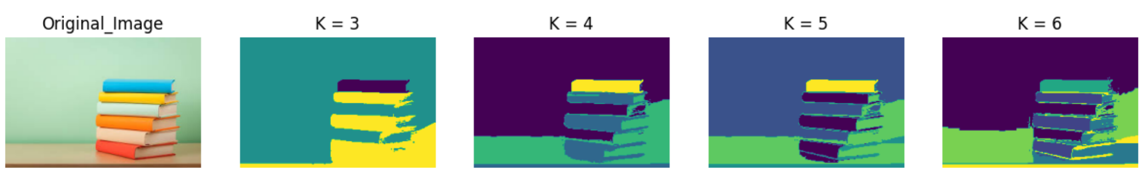
# Image3:

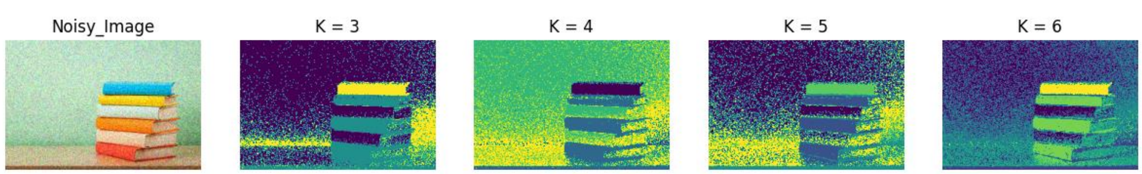

Original_Image


Original_Image

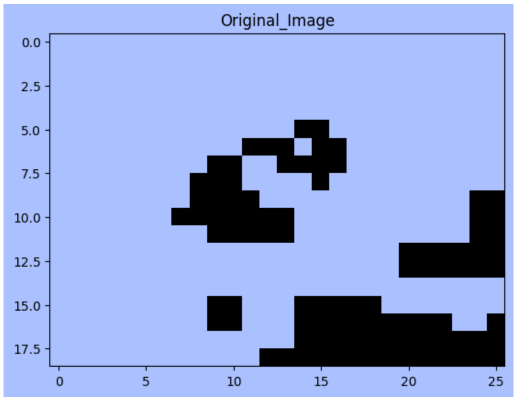

Noisy_Image

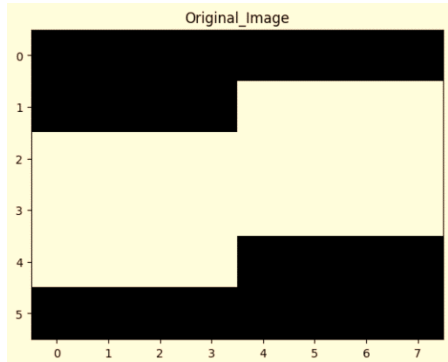

Noisy_Image

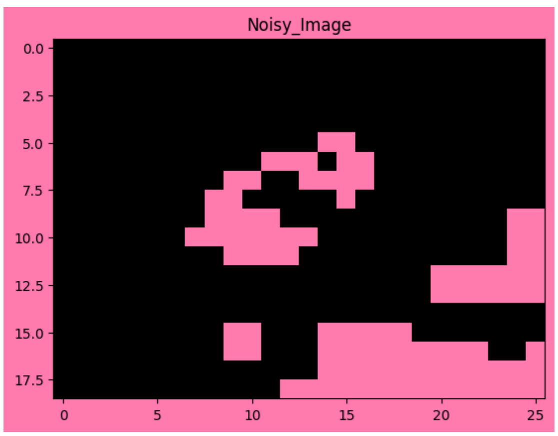

Original_Image
K = 3
K = 4
K = 5
K = 6

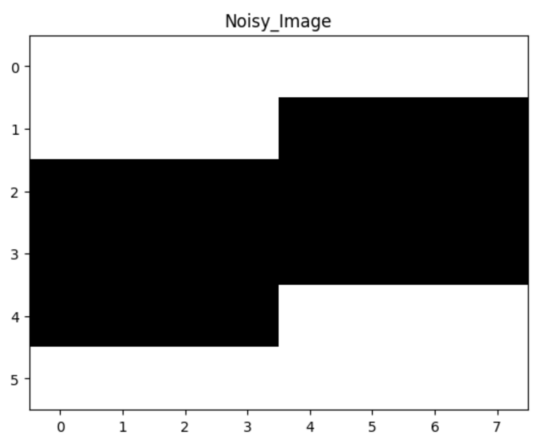

Noisy_Image
K = 3
K = 4
K = 5
K = 6

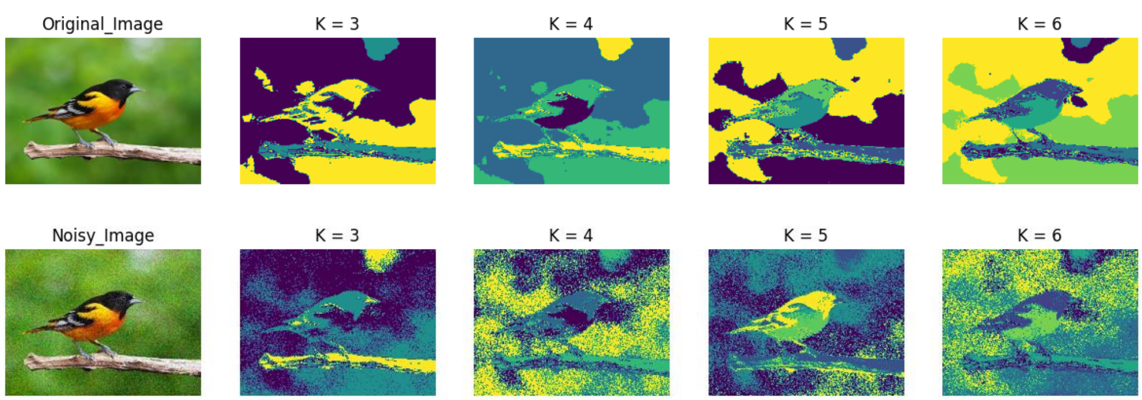
# Image4:


Original_Image

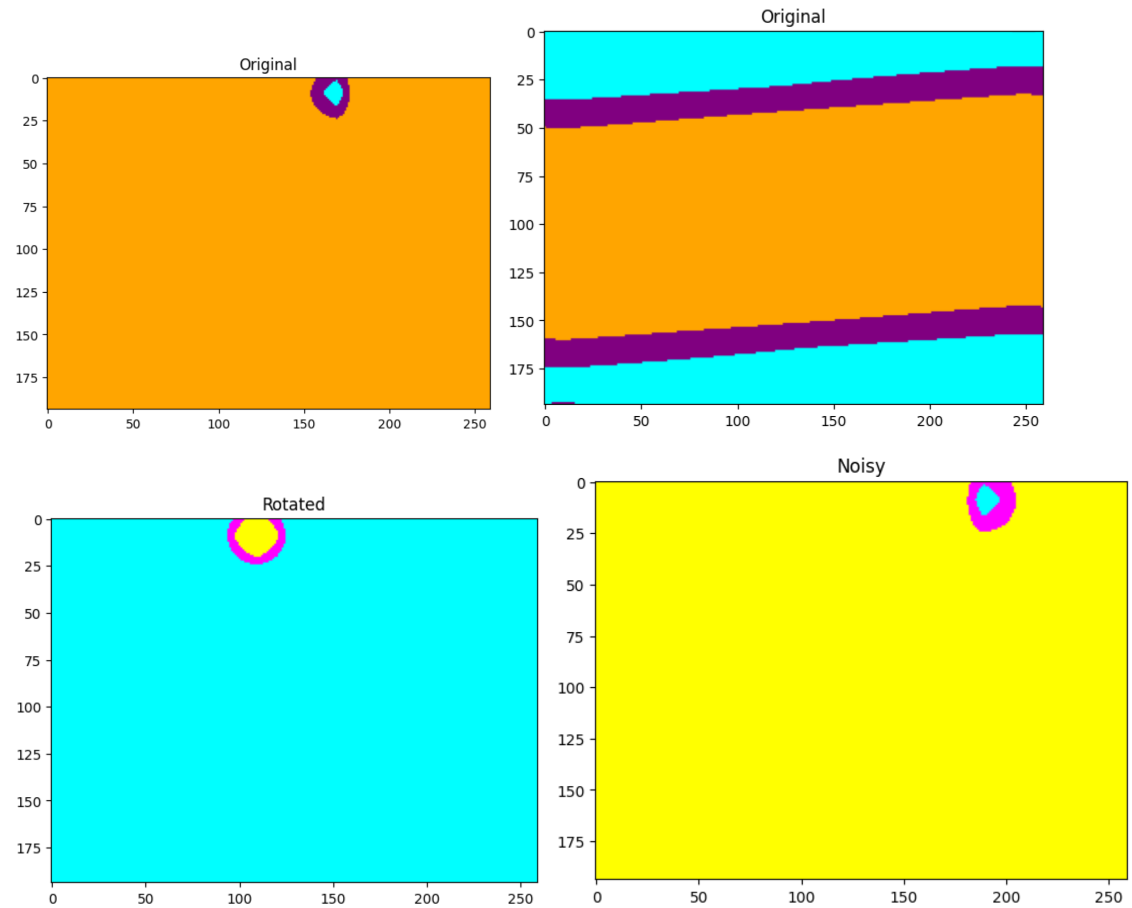

Original_Image


Noisy_Image


Noisy_Image


Original_Image
K = 3
K = 4
K = 5
K = 6
Noisy_Image
K = 3
K = 4
K = 5
K = 6

# 3 Segments on image 2:




# Qualitative Analysis: N_cut with two partitions performed better on photos with a distinct two-segment separation (such as the image of the dog). It was over segmented by K-means with k = 3 to 6, which produced a subpar outcome. However, K-means performed remarkably well in identifying the minute variations in intensity in other photos with more than two segments. When compared to the proximate similarity we first mentioned, the N-cut utilizing spatial similarity fared worse.

# FCN for image segmentation

The general procedure that is used is as follows: 1. Using the JSON file as a guide, divide the picture and mask into train and test. 2. Define down sample transforms. 3. To prepare the ground truth, transform the masks into categorical tensors (labels). 4. Load Resnet18 trained on Imagenet. 5. Describe the architecture of skip/no-skip. 6. Using pixel wise 1*1 convolution blocks, classify the downsampled output from the model for each class. 7. Extend the sample and contrast it with labels derived from masks. 8. Evaluate using test data stating pixel wise accuracy and mean IoU.

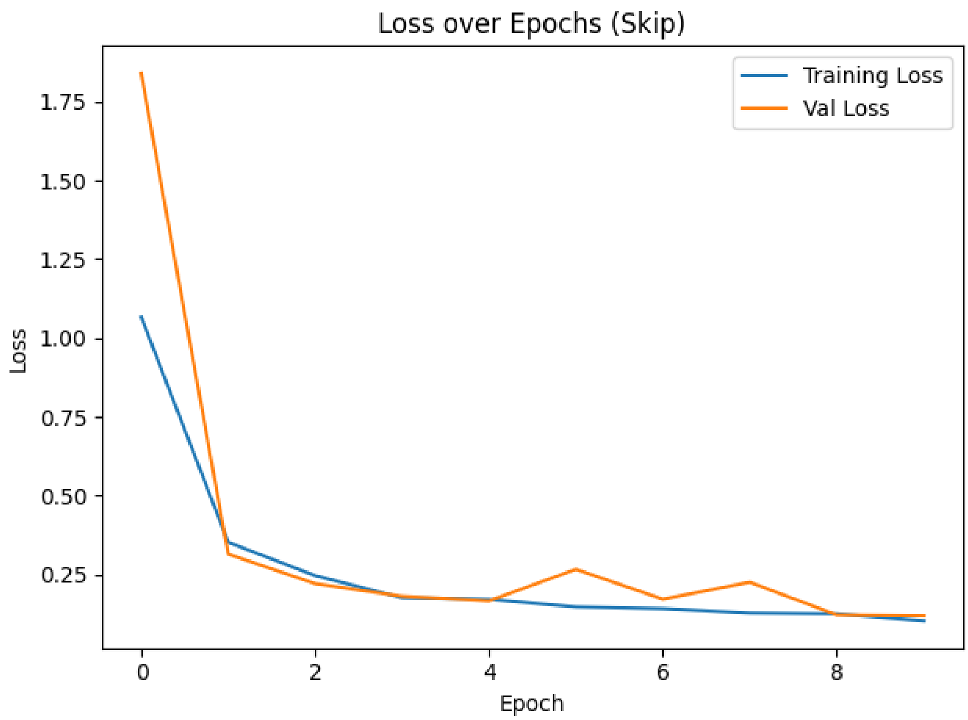
**The ResNet18_FCN_NoSkip** model adapts the ResNet-18 architecture for semantic segmentation by maintaining the spatial resolution throughout the network, transforming the network to output class probabilities for each pixel. The key modifications include removing layers that are not conducive to maintaining spatial information (like the average pooling and fully connected layers), adding a 1x1 convolution to adjust the number of output channels, and employing upsampling to ensure the output size matches the input image size. This model is particularly suited for applications where detailed spatial information is crucial, and the use of a pre-trained backbone accelerates training and improves performance by utilising learned features.

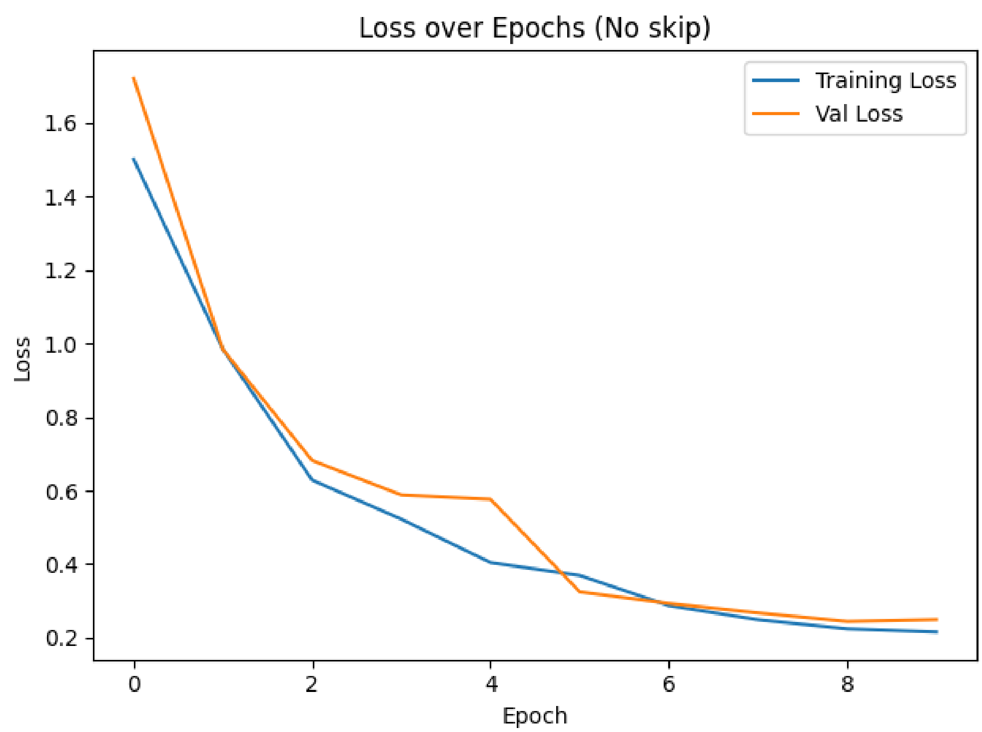
**The ResNet18_FCN_Skip** model starts with a pre-trained ResNet-18 model, and then adds a 1x1 convolution to reduce the channel size. The feature map is then upsampled back towards the input size through multiple stages, and skip connections are added to reintroduce features from earlier layers. This helps to blend high-level, semantic information from deeper layers with detailed, spatial information from earlier layers. Finally, the feature map is upsampled one last time to produce the final segmentation map.

**Qualitative Analysis:**

| Network\Metric | Pixelwise Accuracy | Mean IOU |
|---|---|---|
| Skip conn | 1.0000 | 11.1111 |
| No-skip conn | 1.0000 | 11.1111 |

The CNN features are more information-rich because the skip connections consider the combined spatial and semantic description from two sampled levels. As a result, a little improvement in that area is observed. It's an exceptionally poor performance all around. The model is readily overfitted. One possible explanation could be that there are many fewer training photos than there were in the training set used to build the model, which allowed the model to almost completely understand every component of the training set's segmentation. Also, the metrics suffer from the severe down- and up-sampling.


Loss over Epochs (Skip)
Training Loss
Val Loss
Loss
Epoch


Loss over Epochs (No skip)
Training Loss
Val Loss
Loss
Epoch