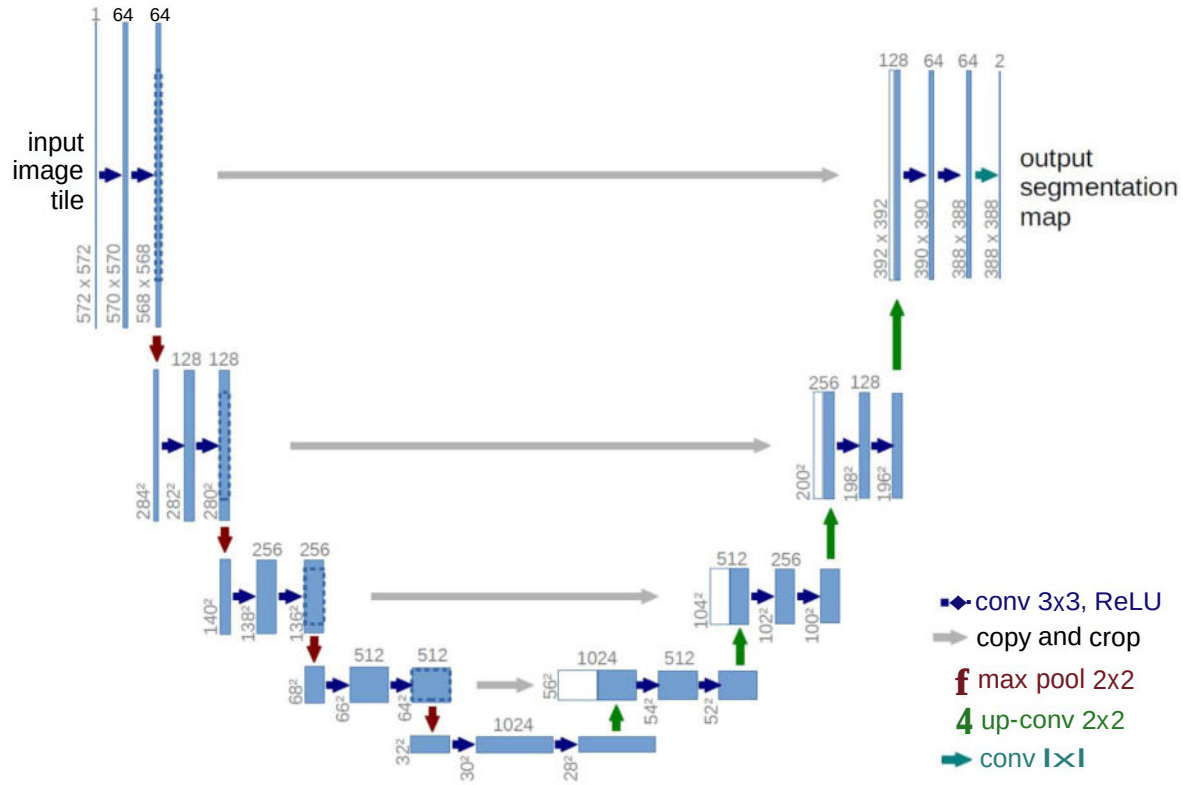
1 Problem Restatement

We implemented a network for **semantic segmentation in image data,** and also **generate estimates of aleatoric and epistemic uncertainties associated** with the segmentation.

2 introduction of our model

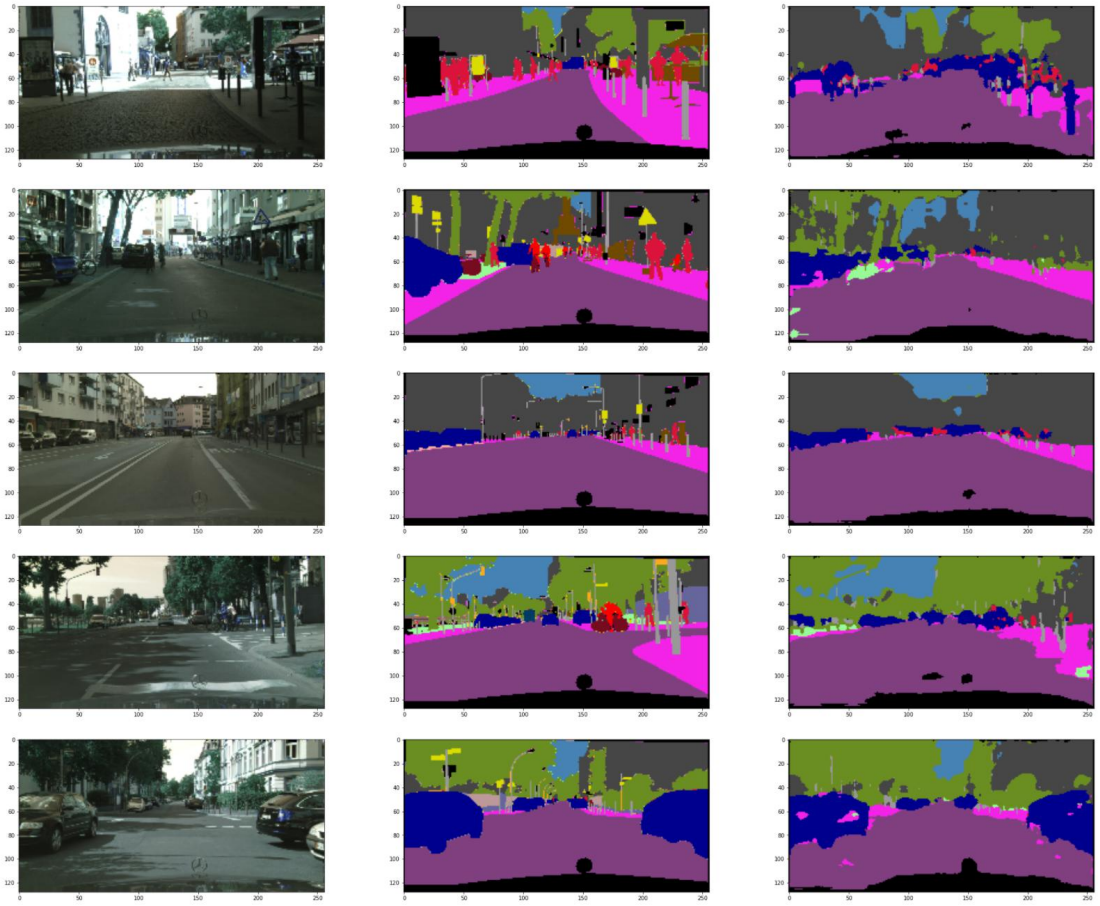
In Image Segmentation, the machine has to partition the image into different segments, each of them representing a different entity. We not only need to convert feature map into a vector but also reconstruct an image from this vector. **This is a mammoth task because it's a lot tougher to convert a vector into an image than vice versa. The whole idea of UNet is revolved around this problem.**



The architecture looks like a 'U' which justifies its name. This architecture consists of three sections: The contraction, The bottleneck, and the expansion section. The contraction section is made of many contraction blocks. Each block takes an input applies two 3X3 convolution layers followed by a 2X2 max pooling. The number of kernels or feature maps after each block doubles so that architecture can learn the complex structures effectively. The bottommost layer mediates between the contraction layer and the expansion layer. It uses two 3X3 CNN layers followed by 2X2 up convolution layer.

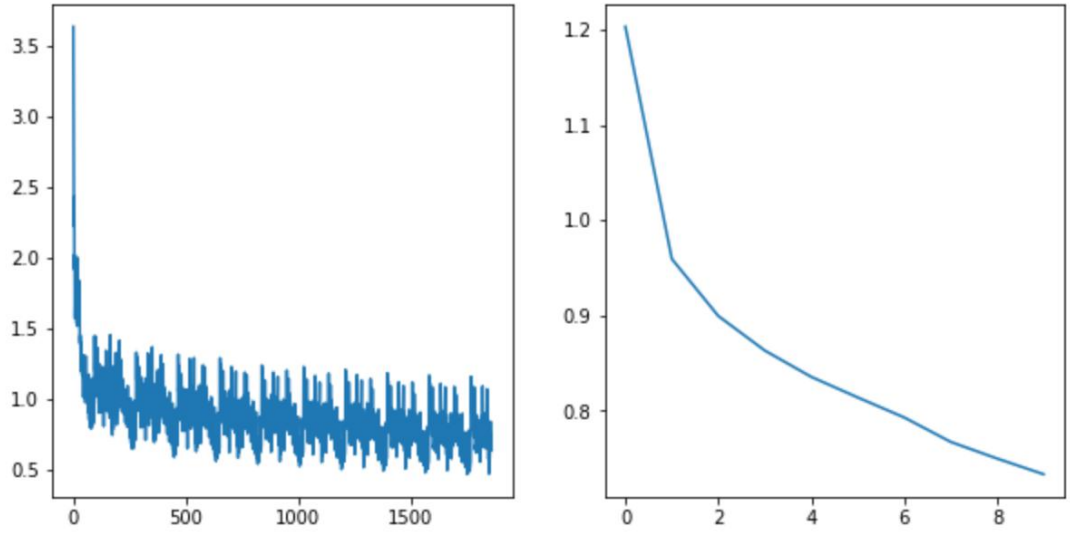
3 test result:

just look at the examples



the loss of batch and epcho

```
Final Epochloss : 0.733436123978707
```



4 Model Comments

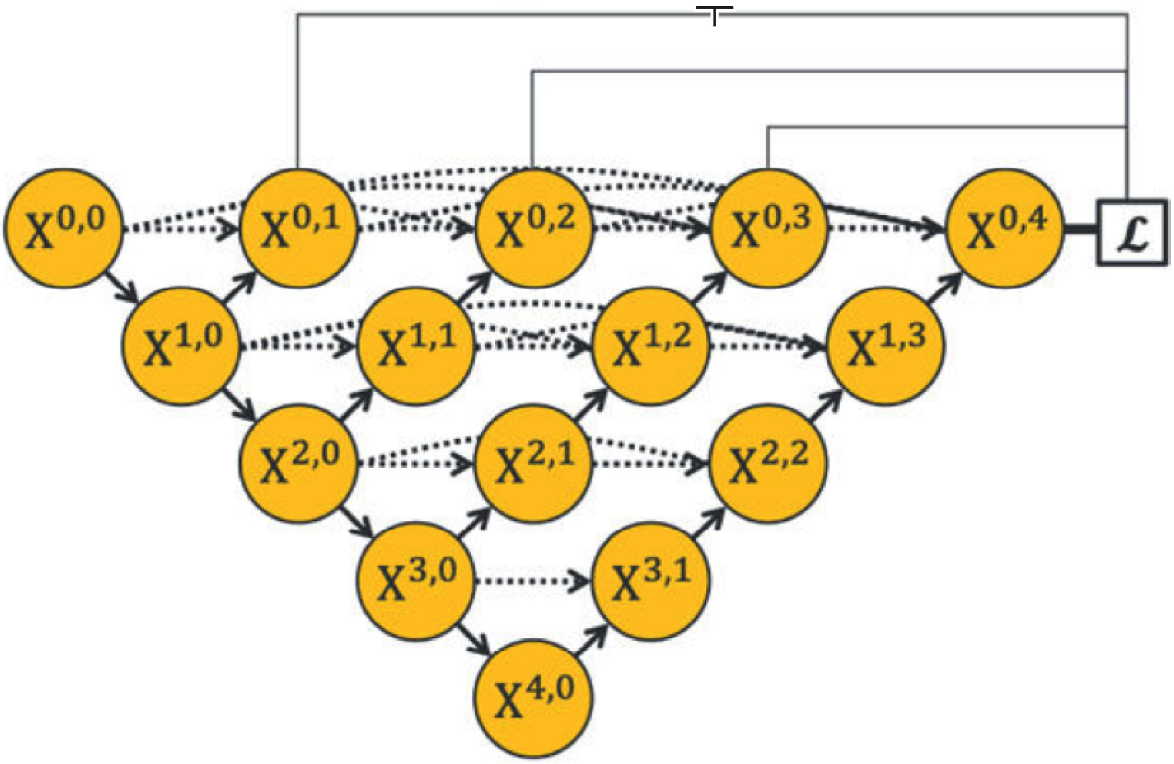
The model we acquired after training is doing a great job on segmenting parts like vehicles on the road and driveways which takes up a relatively large area. However, it performs badly when it comes to those tiny parts such as traffic lights and road marks.

As can be seen from the learning curve, the loss reduced only by a little amount and stayed high during the later stage of training, which we conclude to the inaccurate segmentation of those tiny parts.

Over all, our model is able to do segmentation task on large-area parts but not precise enough.

5 Possible Improvements

In this part, we are suggesting two possible architectures to be used as an improvement to the model, the first one that we suggest is called Unet++, whose structure is shown below.



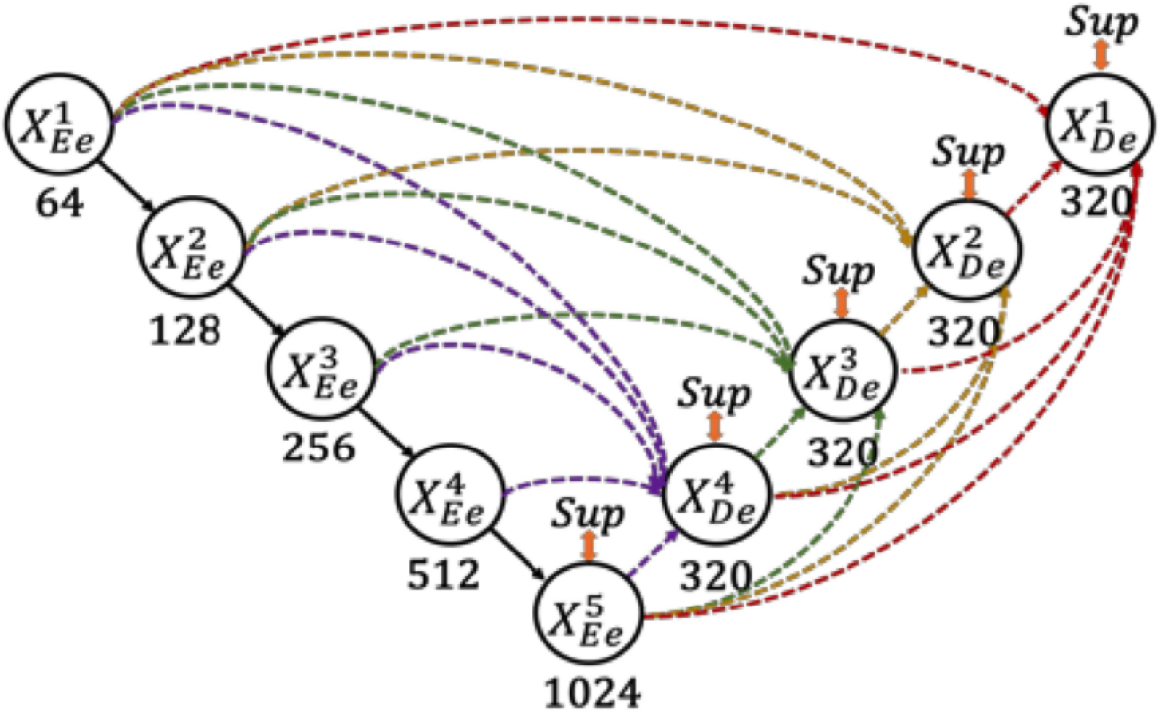
Compared with the used Unet, the main differences are:

> 1. This structure is constructed by adding dense skip connections to UNet+, enabling dense feature propagation along skip connections and thus more flexible feature fusion at the decoder nodes.
> 2. It used explicit deep supervision to train but its optional.

Though it proved to be achieving higher accuracy and faster convergence than the original Unet, the many convolution blocks it introduced in the architecture is adding **a tremendous amount of parameters** to be trained.

During training, we found that this structure is hard to train and its performance on the testset is not as good as expected, which could be put down to the lack of training. the training process of this structure requires a large amount of time and computational resources, taken this limitation into consideration, we think one possible way of improvement of the model could be based on finding a way to reduce the number of parameters.

Another possible improvement to this model, as is suggested in the paper of Unet3+ *(Hummin Huang et.al. ,2020)*, is to replace **the nested and dense skip connections** with **Full-scale skip connection**, which reduces the parameters and incorporate low-level details with high-level semantics from feature maps in full scales. The fiugre below illustrates this novel structure.



With a reduced amount of parameters, we believe this model could be faster to train without losing any low level semantic features.