Deep Learning (CS F425)

MODNet:- real-time trimap-free portrait matting using Objective Decomposition.

Final Report.

Group 9

Name of the Students:	ID Nos.:
Adithya Manjunatha	2019A7PS0118G

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

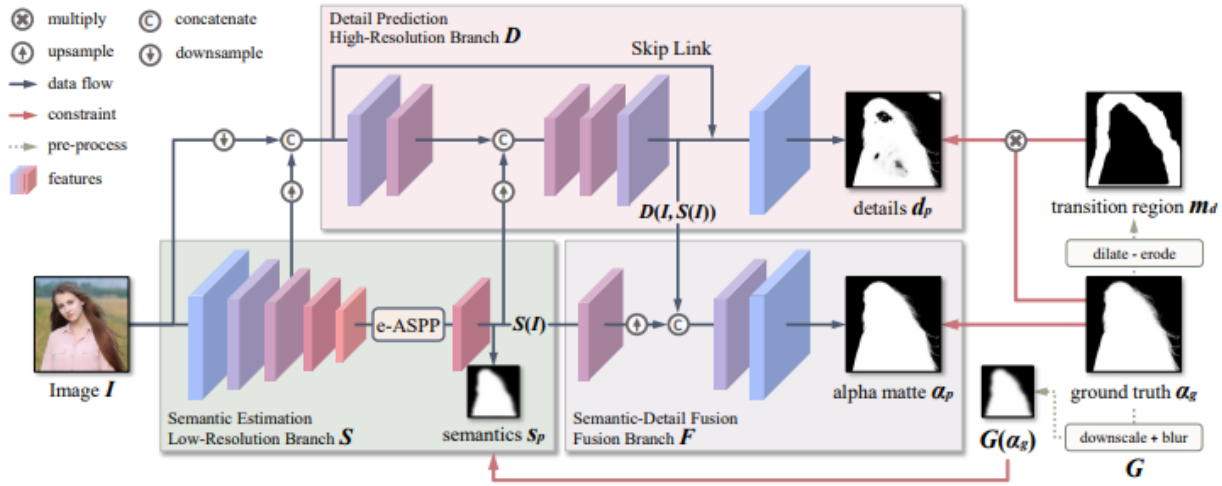
K K Birla Goa Campus (Semester-II 2021-22)

Contents:-

Summary of MidSem Submission.	2
Work Post Midsem	3
Final Architecture	5
Final Results	6
Conclusion	7
References	7
Additional links:-	7

Summary of MidSem Submission.

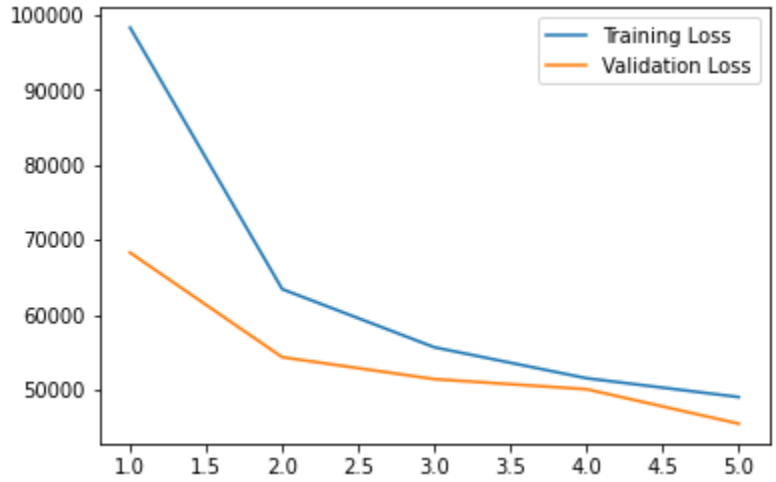
In the MidSem submission, the model was used as given in the paper MODNet.



The dataset used was
https://www.kaggle.com/laurentmih/aisegmentcom-matting-human-datasets
This dataset is used throughout. It contains 34427 images with their respective portrait matting.

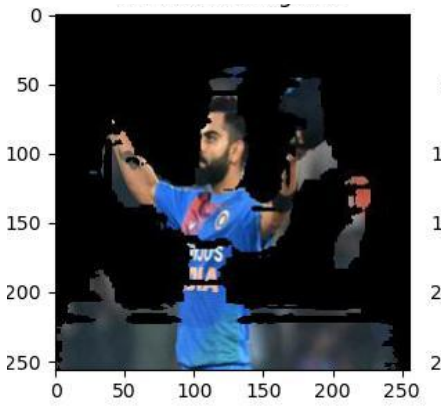
Although the paper uses images of size 512x 512, we implemented MODNet using images of size 256x256 due to computational constraints on google colab.

The Training loop took about 3 hours to complete, for just 5 epochs. The model however gave fairly good results for such a low number of epochs.



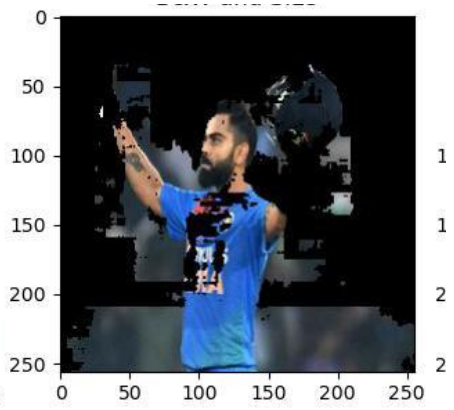
Work Post Midsem

1. Saving preprocessed data, reduced the training time from 40 minutes per epoch to 12 minutes per epoch, without any loss of accuracy.
2. When we saw the results of the model on real-world data, we found that the model had learned to show anything that had colour into the alpha matte. Therefore, we attempted training the model with the input being a grayscale image, since matting does not depend on the image's colour.
3. The above version slightly overfits on the data, making the matte always be true for the lower quarter of the image, marking the shoulders of a human, since most of the images in the dataset had it in that position.

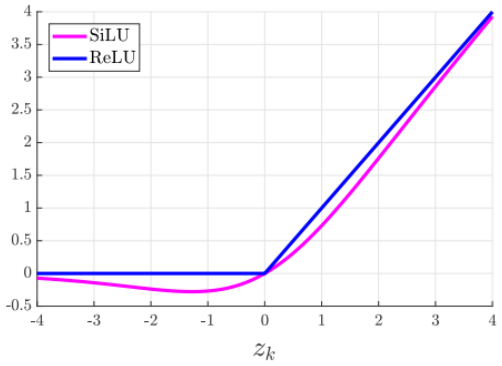


4. To overcome this, we added augmentations to the dataset, but this however increased the loss on a black and white image. Switching back to RGB image gave better results with augmentations.

5. At this point, the images were giving mattes with a few random True areas. To tackle this problem we changed all the activations from ReLU to Sigmoid to push the values down between 0 and 1.

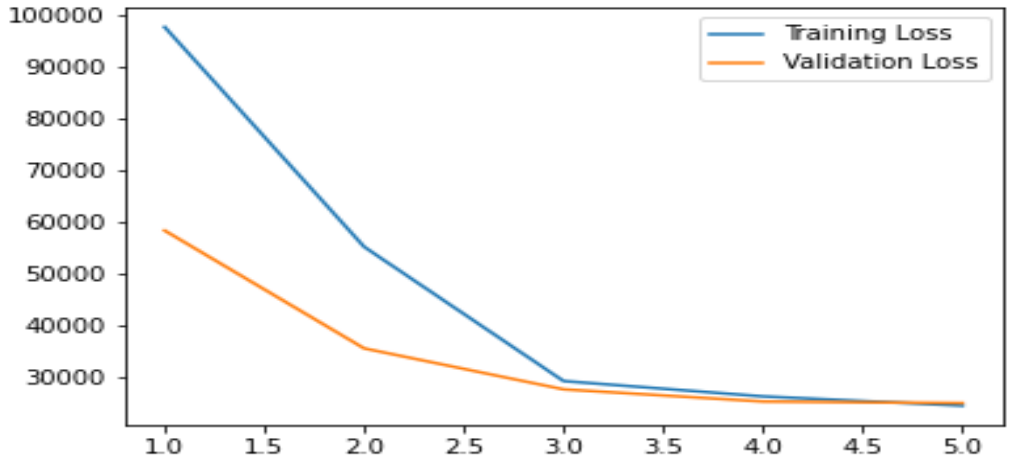


6. Also tried using the SiLU activation function which claims to result in faster convergence. Using this made the convergence faster, but outputs were similar to ReLU, shifting back to Sigmoid.

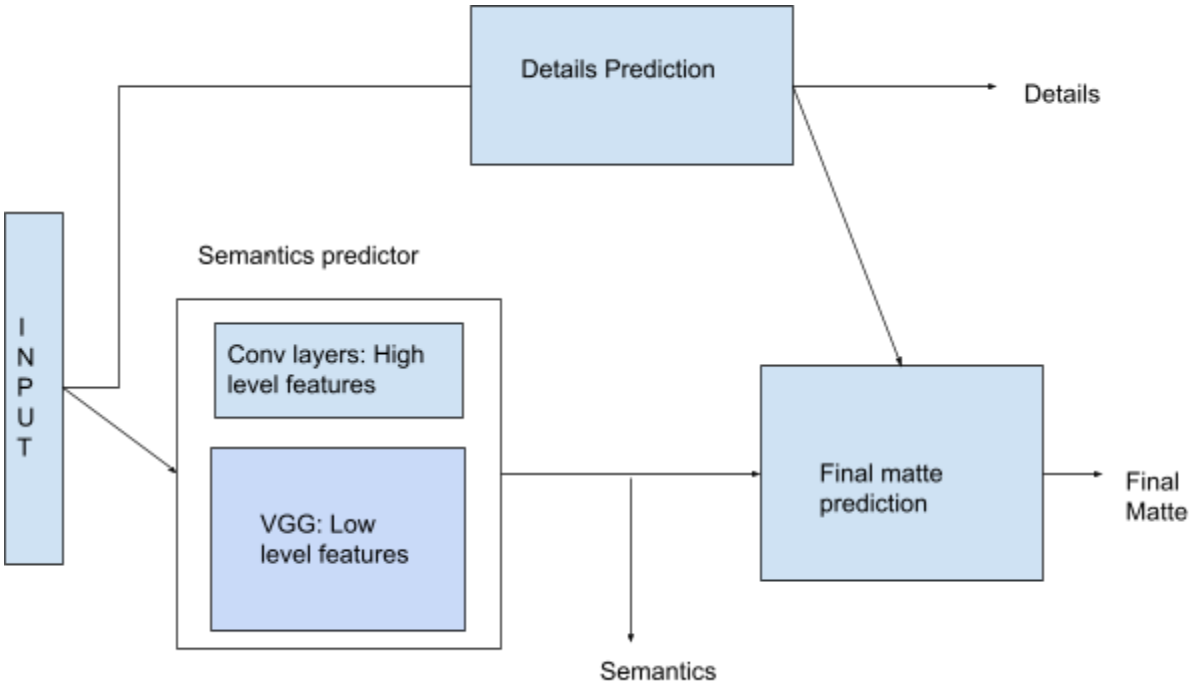


7. Finally, we introduced the VGG16 pre-trained model in the Semantics module to extract rich low-level features from the image. Adding this increased the model size slightly, but brought the loss down by a

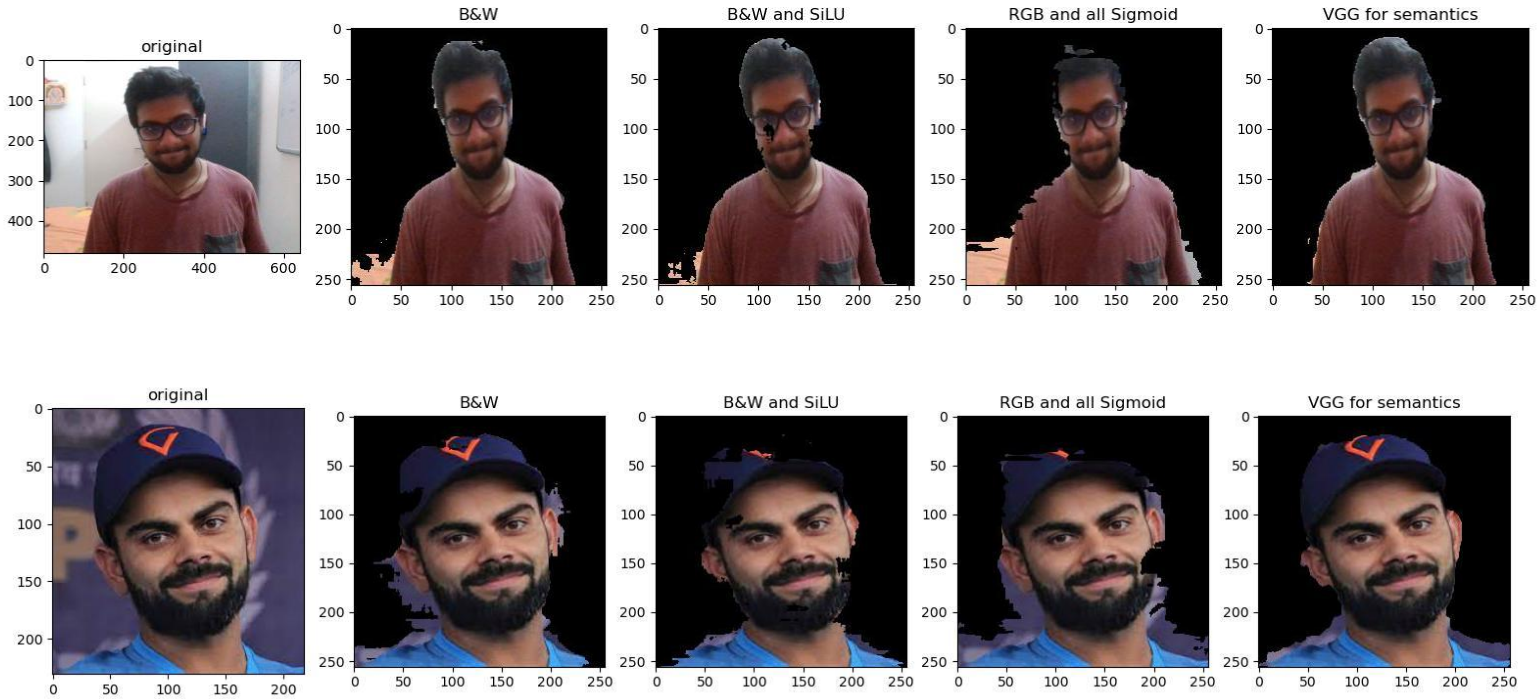
large amount.



Final Architecture



Final Results



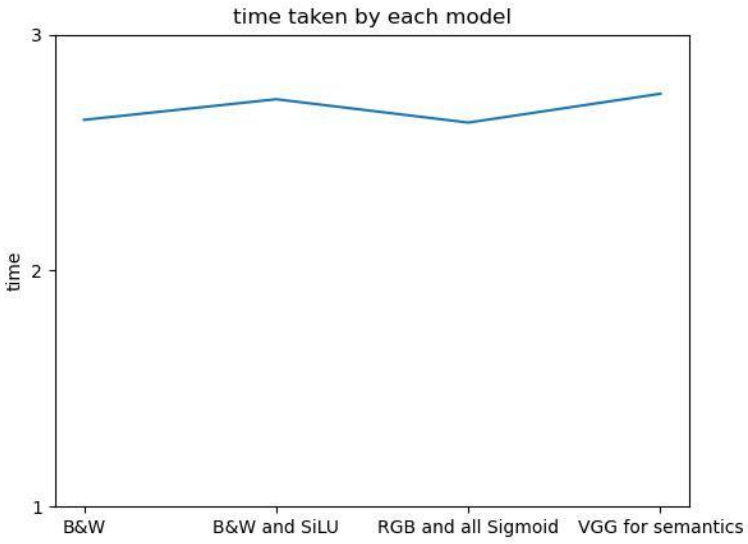


Fig. Time taken by each model to predict matte.

Conclusion

MODNet is a simple fast and effective model for portrait matting, with a single RGB image. The addition of a simple pre-trained model vgg16 for semantics feature extraction reduces the loss drastically with hardly any impact on the prediction time.

References

1. MODNet
   https://arxiv.org/pdf/2011.11961.pdf
2. SiLU
   1702.03118.pdf (arxiv.org)
3. Dataset: AISegment.com - Matting Human Datasets
   AISegment.com - Matting Human Datasets | Kaggle

Additional links:-

GitHub Repository:-
audi1712/DL-Project-CS-F425- (github.com)
Contains the final notebook, saved model, and a program to use the model output as “OBS virtual camera” (requires OBS software and pyvirtualcam library).