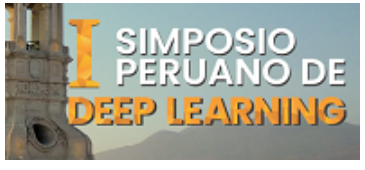
Hand Segmentation using U-Net

C. Lazo[1]

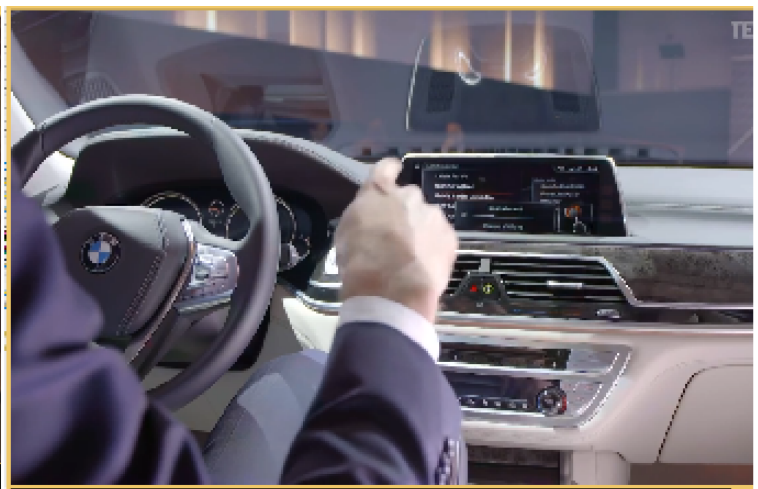
[1]Mechatronics Engineering
Universidad Nacional de Ingeniería
Lima, Peru

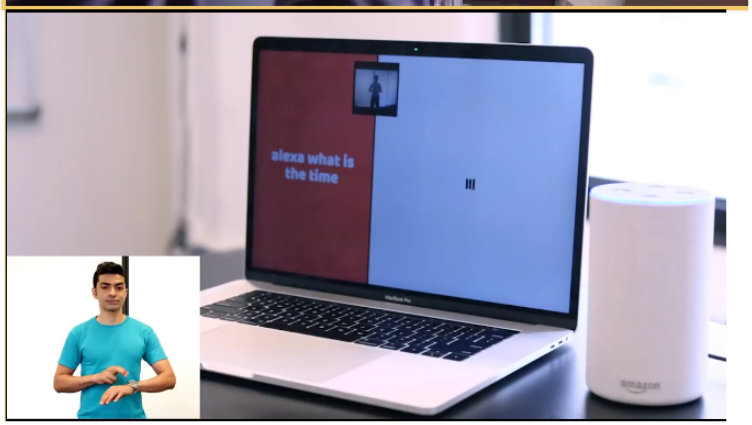
I Simposio Peruano de DEEP LEARNING

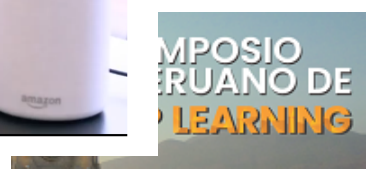


Plan

Hand detection





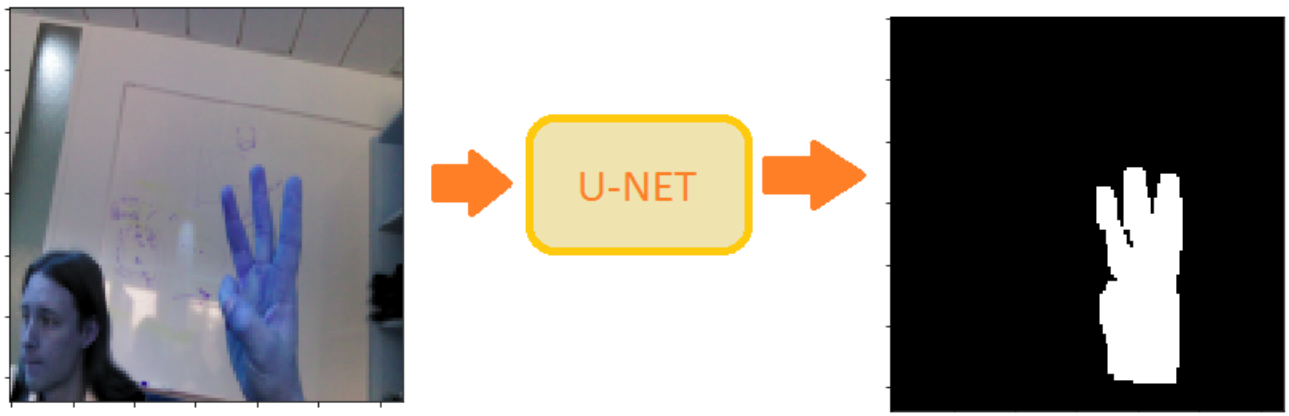


Hand detection

- Samsung Smart TV Review
  https://www.youtube.com/watch?v=gsAAPlL4HcE
- 2016 BMW 7 Series - Gesture Control
  https://www.youtube.com/watch?v=nVZYK-k_nm4
- the DJI Mavic Air Smart Gestures Without a Remote
  https:
  //www.youtube.com/watch?v=e0fCS3nWx4w&t=67s
- Making Amazon Alexa respond to Sign Language using AI
  https://www.youtube.com/watch?v=kS53y6GWm0w

Purpose

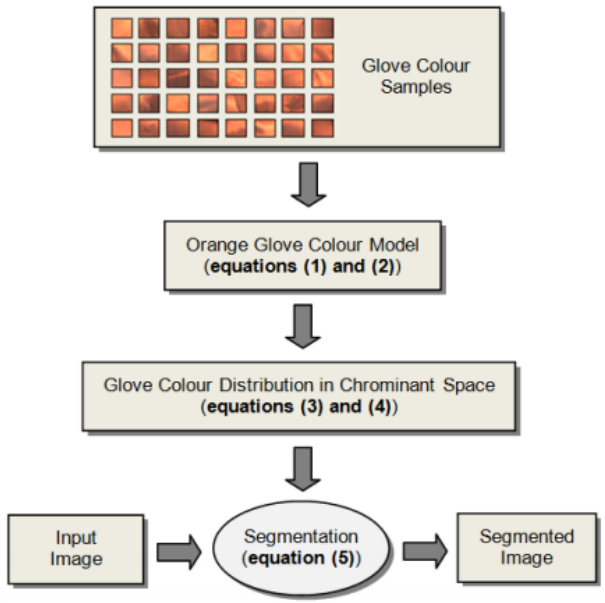
The purpose of this work is the detection and segmentation of hands using Deep Learning model (U-NET) implemented in Python.



Max-Pooling CNN for Vision-based Hand Gesture Recognition

Jawad N. et al :

- The hand contour is retrieved by color segmentation, then smoothened by morphological image processing which eliminates noisy edges.



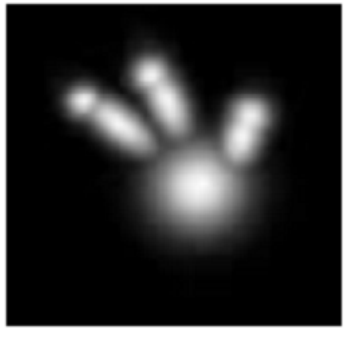
Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering

Lars B. et al. :

- In each image, detection of multi-scale colour features is performed.

$$I = \frac{R+G+B}{3} = f_1$$
$$u = R - G = f_2$$
$$v = G - B = f_3.$$





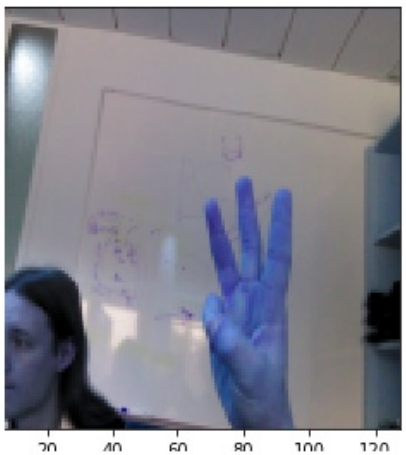
Methodology

The steps involved in the work include :

- Data Collection
- Image improvement
- Data augmentation
- Image segmentation

Data Collection

The OUHANDS - Database is used for the present work :



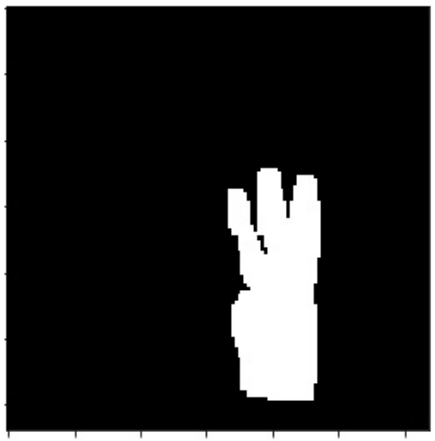


Image improvement

Gamma correction

$$V_{\text{out}} = AV_{\text{in}}{}^{\gamma}$$

Data augmentation

A mirror effect is made with respect to the vertical axis. In total there would be 2 new images for each image, obtaining a total of 6000 images, divided into 0.80 for training and 0.20 for testing.

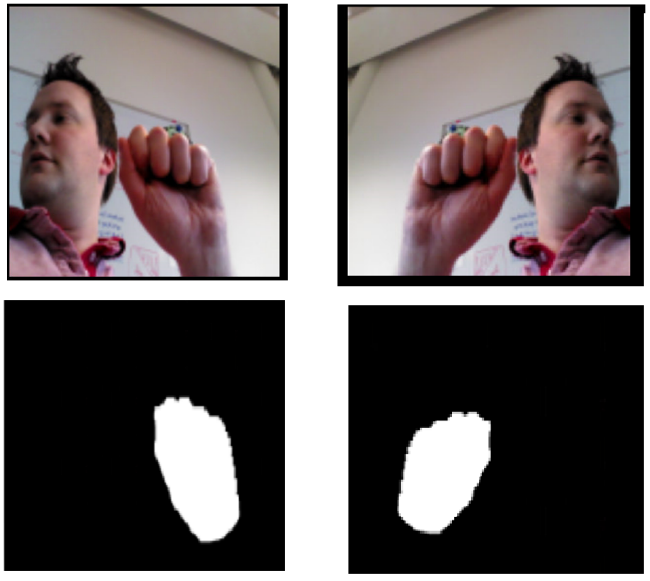
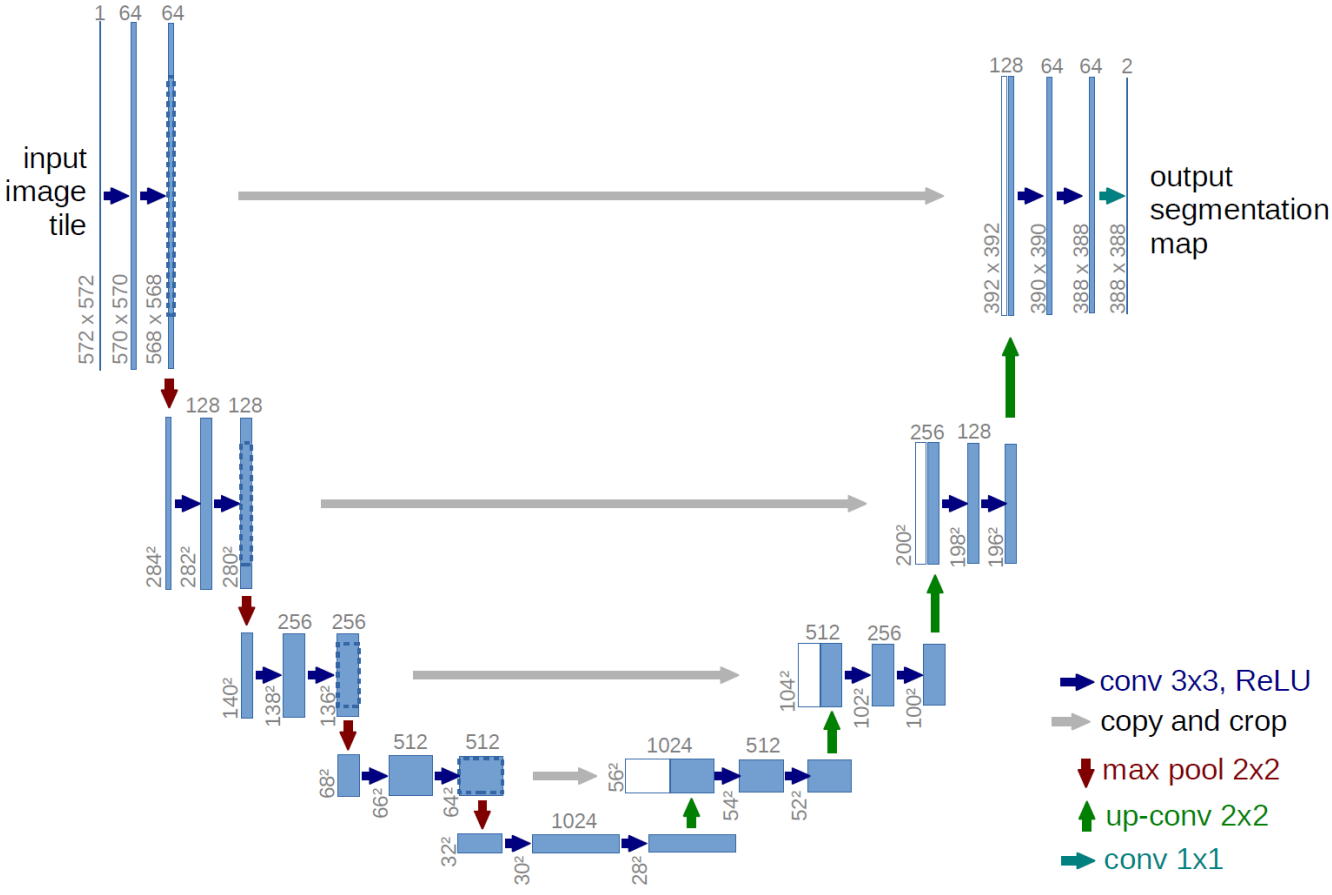
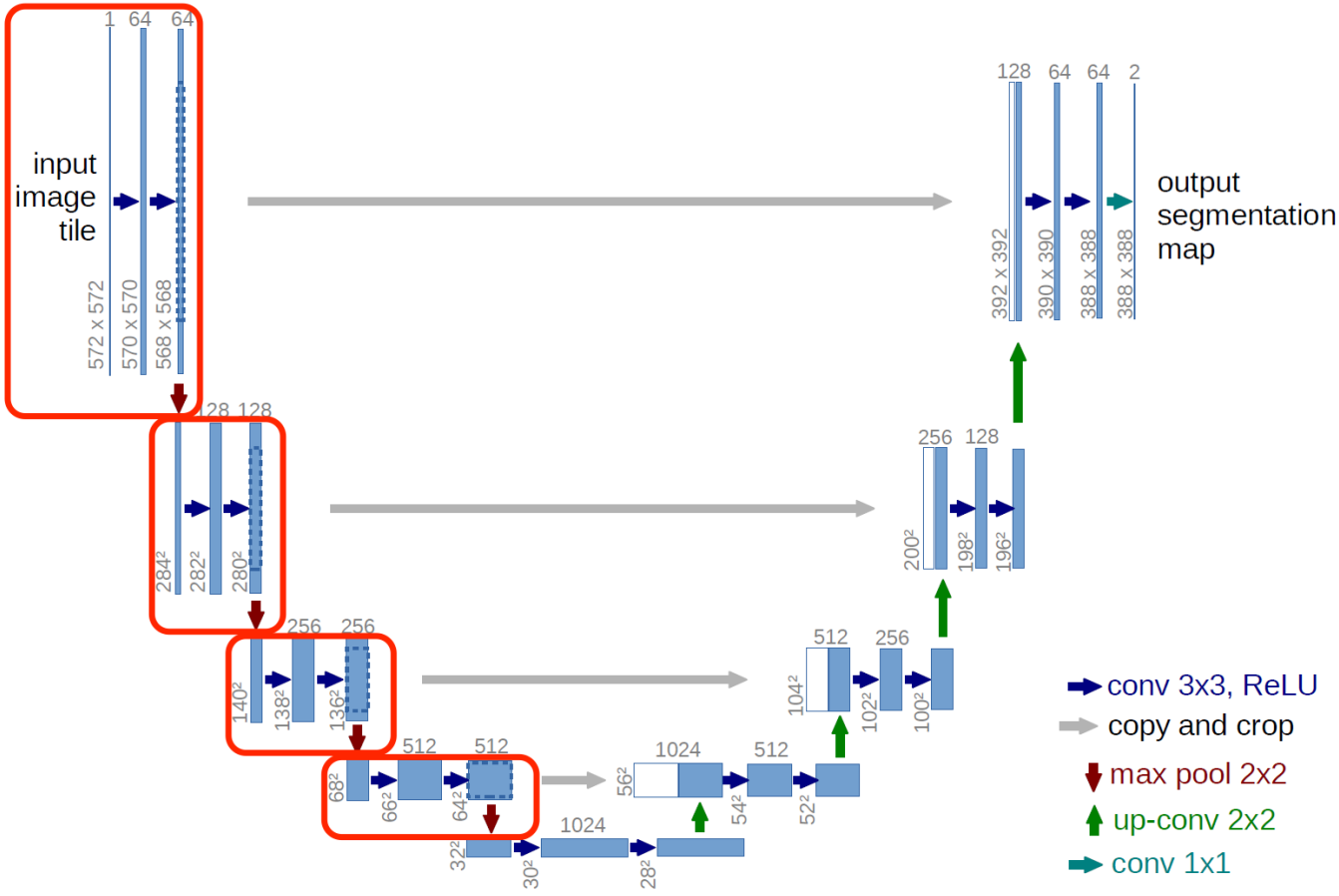


Image Segmentation

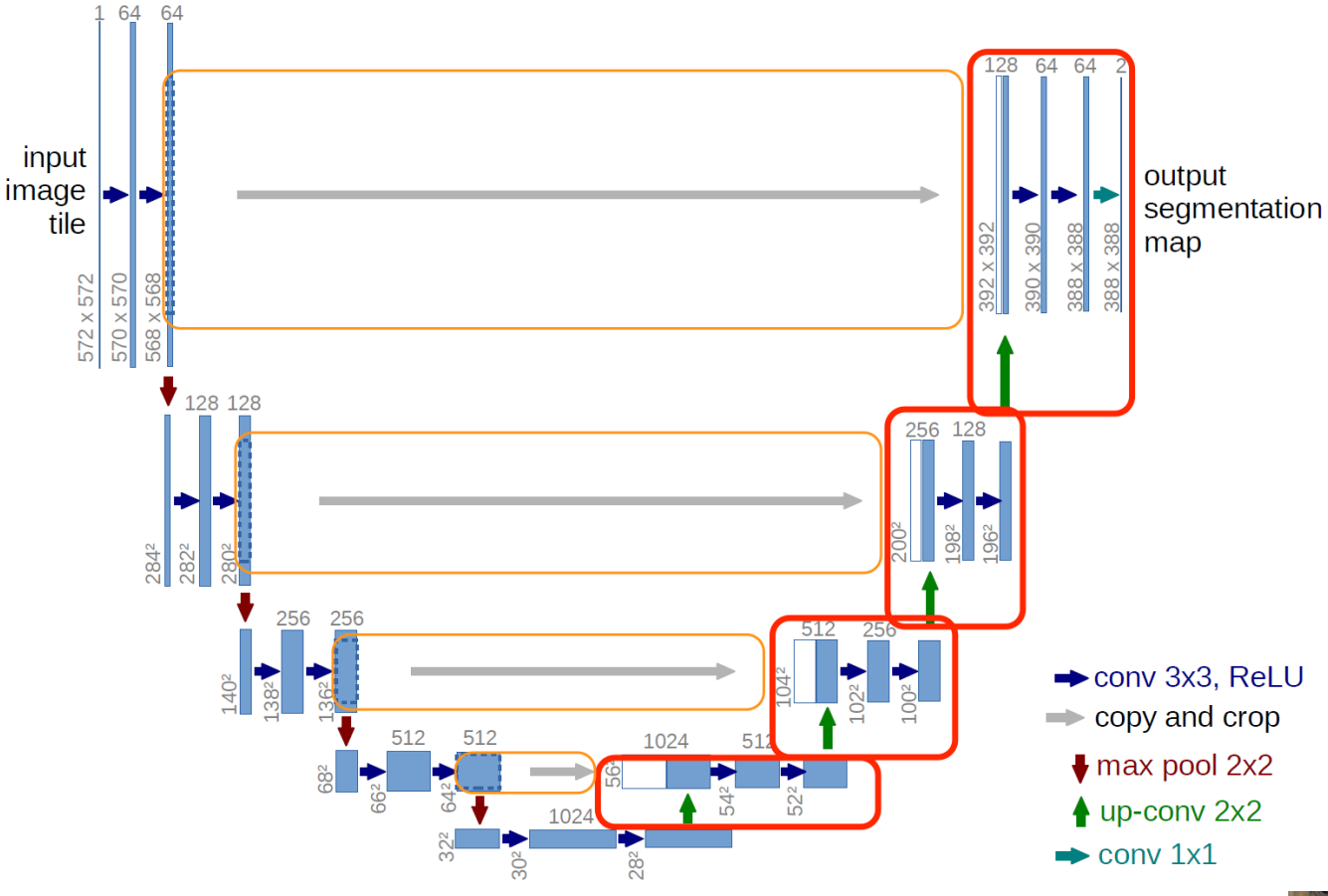
https://arxiv.org/abs/1505.04597



U-NET


input image tile
1 64 64
572 x 572
570 x 570
568 x 568
128 128
284²
282²
280²
256 256
140²
138²
136²
512 512
68²
66²
64²
32²
1024
30²
1024
28²
512
56²
512
54²
256
52²
512
104²
256
102²
256 128
200²
198²
196²
128 64 64 2
392 x 392
390 x 390
388 x 388
388 x 388
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

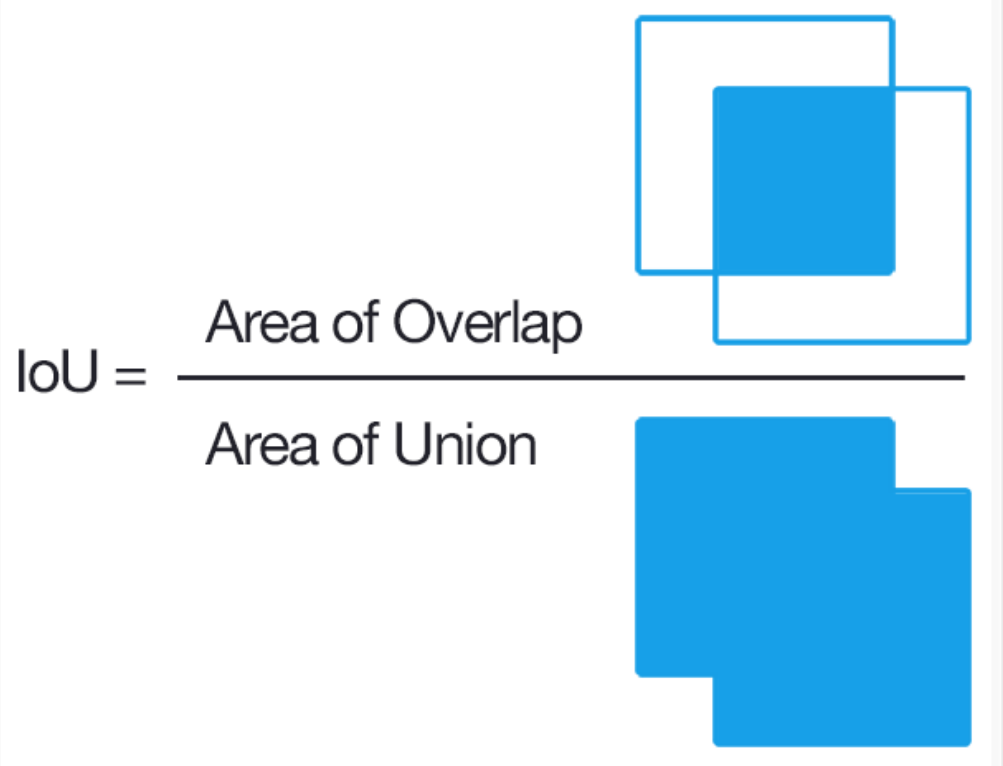
U-NET


input image tile
572 x 572
570 x 570
568 x 568
1 64 64
128 128
284²
282²
280²
256 256
140²
138²
136²
512 512
68²
66²
64²
1024
32²
30²
28²
512 256
104²
102²
100²
256 128
200²
198²
196²
128 64 64 2
392 x 392
390 x 390
388 x 388
388 x 388
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

Metric

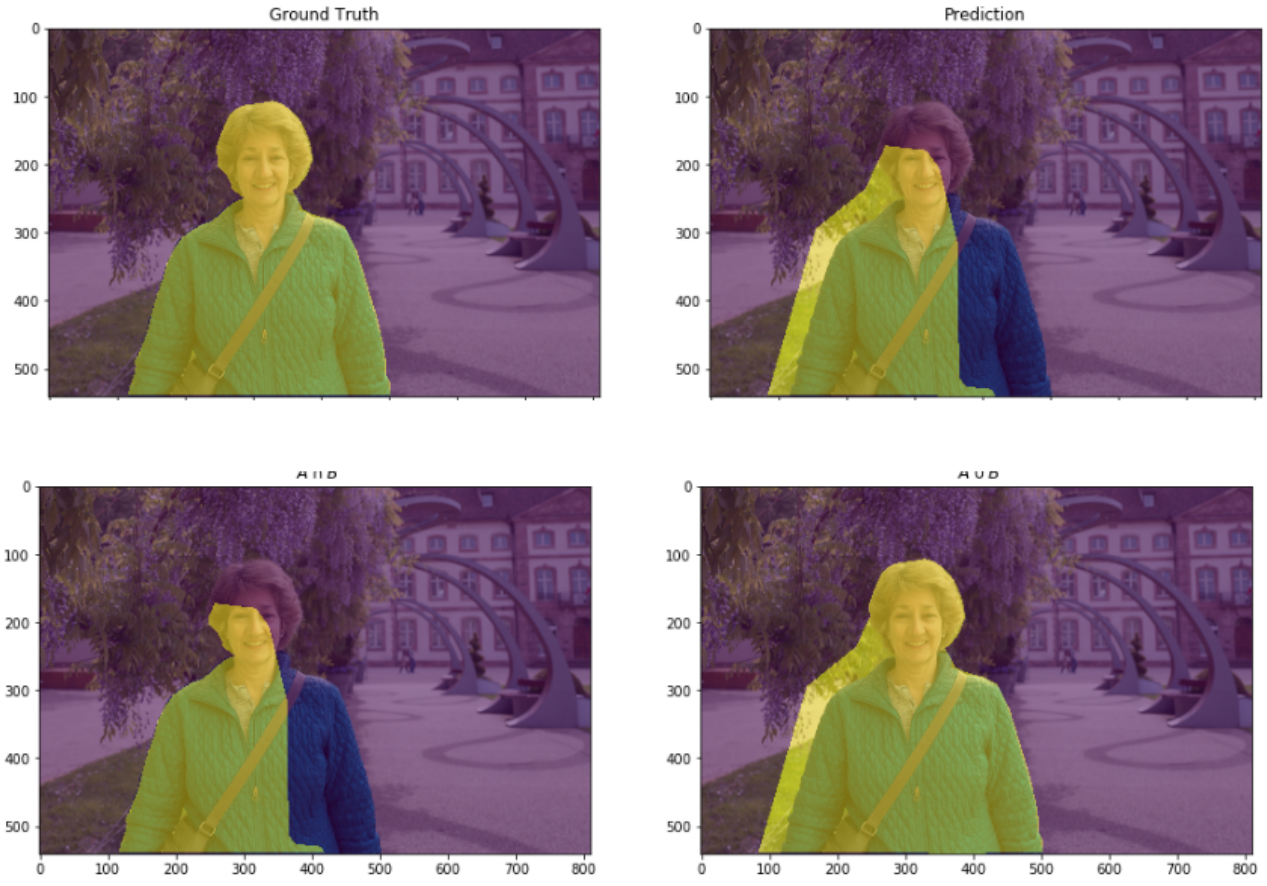
Intersection over Union

https://www.pyimagesearch.com/2016/11/07/intersecti



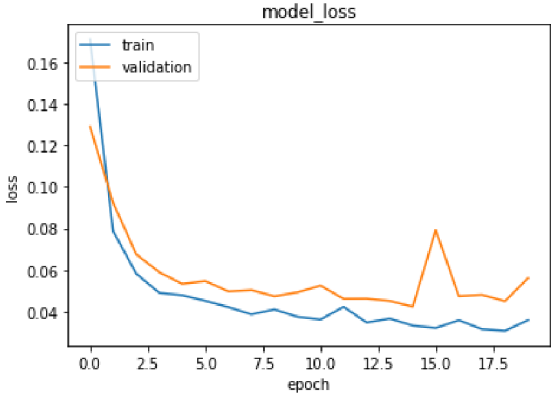
Instance Segmentation

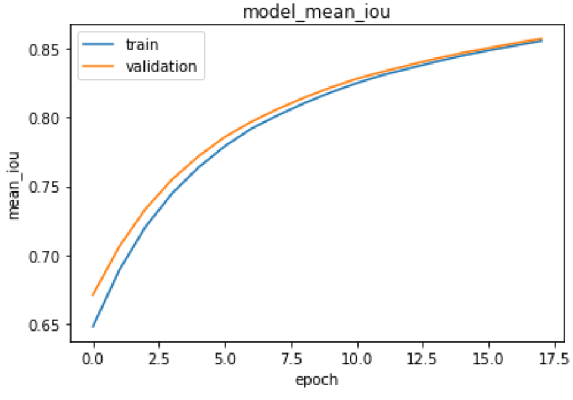
https://www.jeremyjordan.me/evaluating-image-segmen

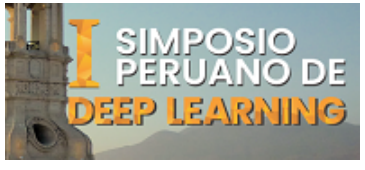


Result

The results obtained for the detection algorithm proposed for the train and test data with 30 epochs of training and testing are 0.8848 and 0.8855 mean Iou,respectively. In the case of the loss function for training and testing are 0.0376 and 0.0508,respectively.

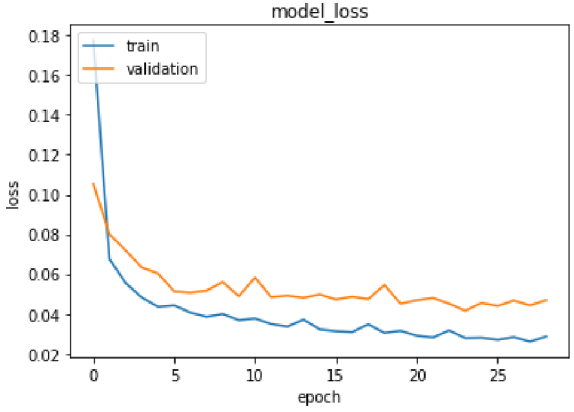


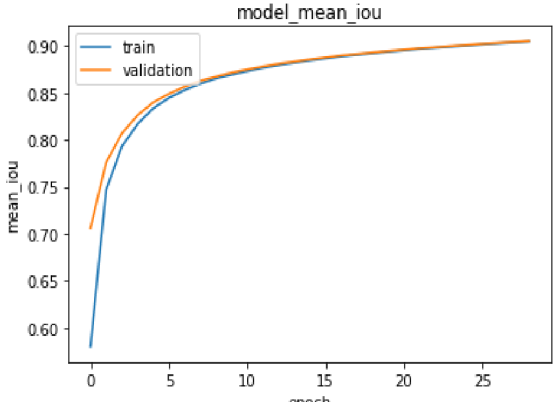




Result with Gamma correction

The results obtained for the detection algorithm proposed for the train and test data with 30 epochs of training and testing are 0.9050 and 0.9045 mean Iou,respectively. In the case of the loss function for training and testing are 0.0285 and 0.0468,respectively.





Conclusion

- The results obtained for the segmentation algorithm proposed is robust in a uniform background.
- The gamma correction improve the segmentation.
- the algorithm is light, works in real time (300ms/step)

Unet :

- Computationally efficient
- Trainable with a small data-set

References

- J. Nagi et al., "Max-pooling convolutional neural networks for vision-based hand gesture recognition," 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, 2011, pp. 342-347.
- L. Bretzner, I. Laptev and T. Lindeberg, "Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering," Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, 2002, pp. 423-428. doi : 10.1109/AFGR.2002.1004190
- U-Net : Convolutional Networks for Biomedical Image Segmentation https://arxiv.org/abs/1505.04597

Thank You !