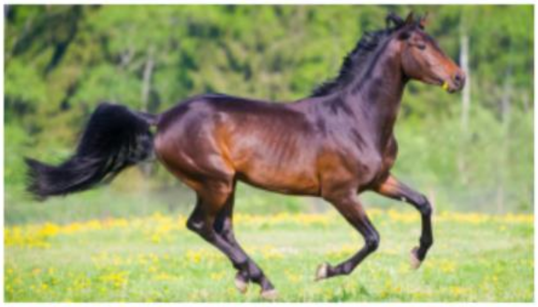
Homework 2 - Segmentation

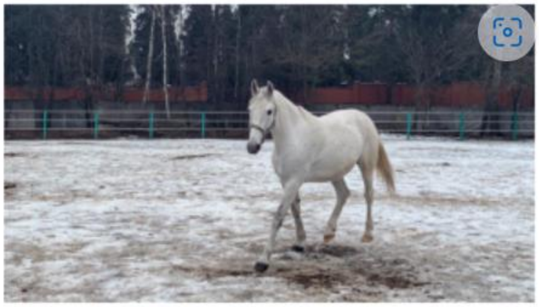
Ron Benchetrit – 312167554

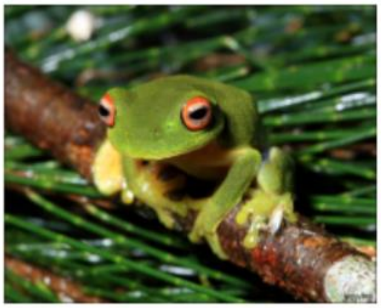
Doron Hanuka – 315824318

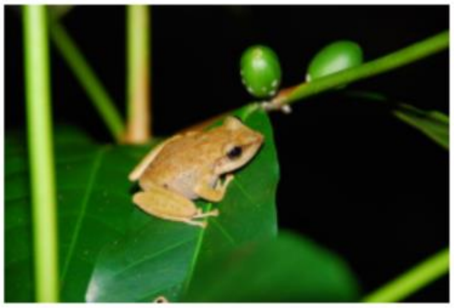
Part 1

1.1



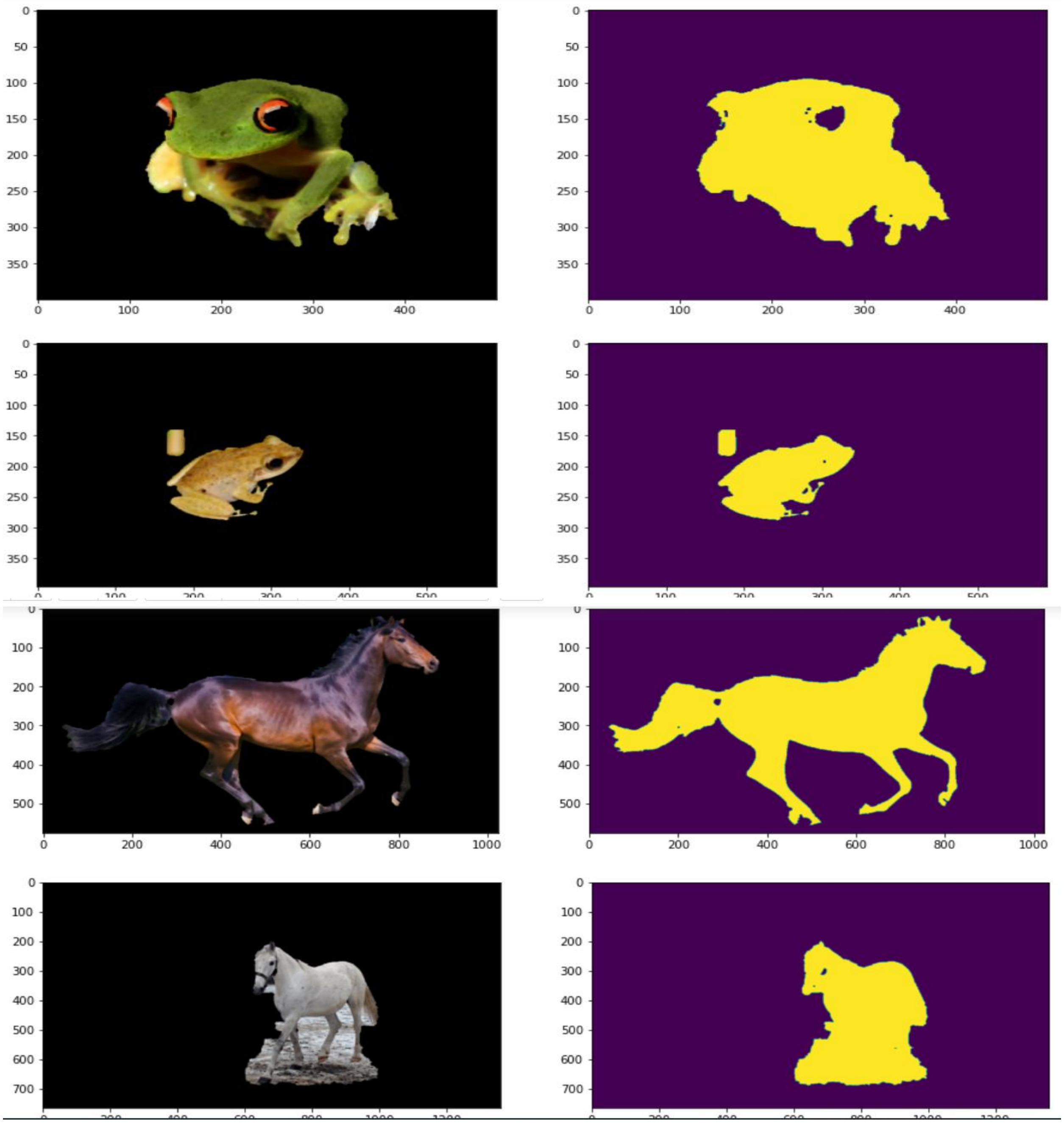




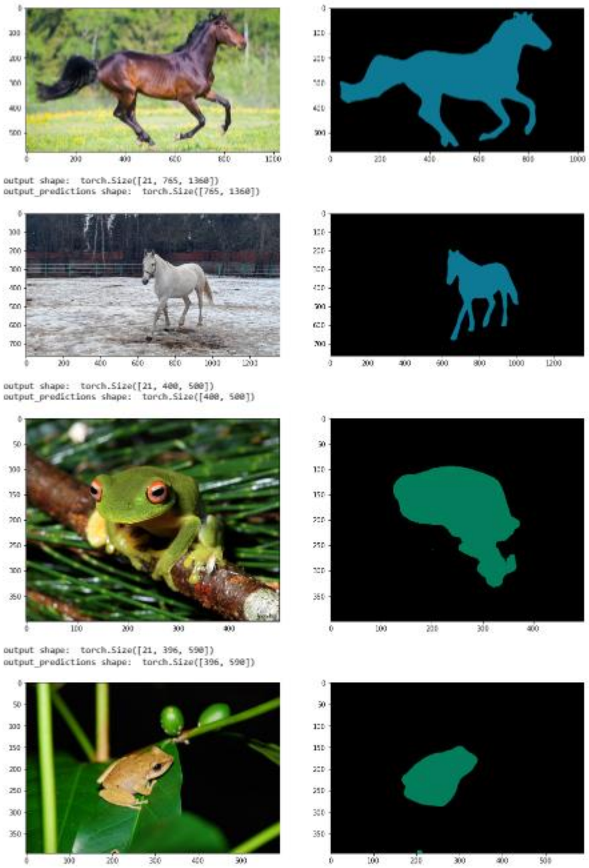


1.2

Classical segmentation – grabCut, implemented with cv2.
we can see that the results are not excellent

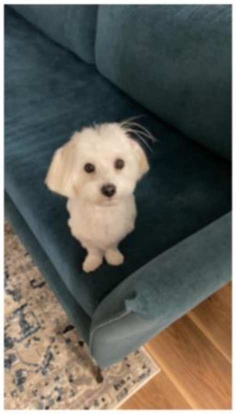


deep segmentation : excellent results



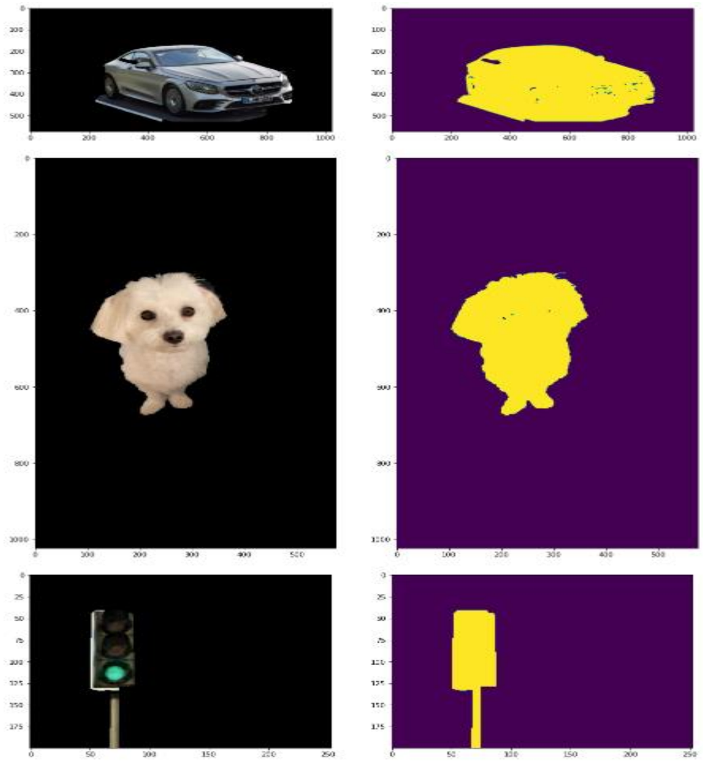
```
output shape:  torch.Size([21, 765, 1360])
output_predictions shape:  torch.Size([765, 1360])

output shape:  torch.Size([21, 400, 500])
output_predictions shape:  torch.Size([400, 500])

output shape:  torch.Size([21, 396, 590])
output_predictions shape:  torch.Size([396, 590])
```

1.3

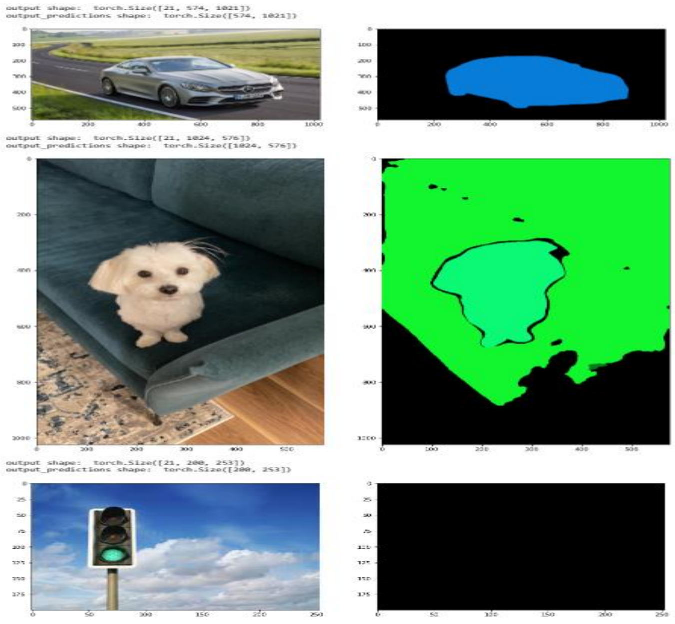


1.4

Classical segmentation : good results



deep segmentation : bad results



```
output shape:  torch.Size([21, 574, 1021])
output_predictions shape:  torch.Size([574, 1021])
```

```
output shape:  torch.Size([21, 1024, 576])
output_predictions shape:  torch.Size([1024, 576])
```

```
output shape:  torch.Size([21, 200, 253])
output_predictions shape:  torch.Size([200, 253])
```

Maybe the deep learning model fails here because it wasn’t trained on such images.
we can see that in 1.2 section it performed better

1. 5

Preprocessing: Applying image processing techniques such as smoothing or denoising can help improve the quality of the image and make it easier to segment.

Postprocessing: Applying techniques such as morphological operations or active contours can help refine the segmentation and improve the accuracy of the results.

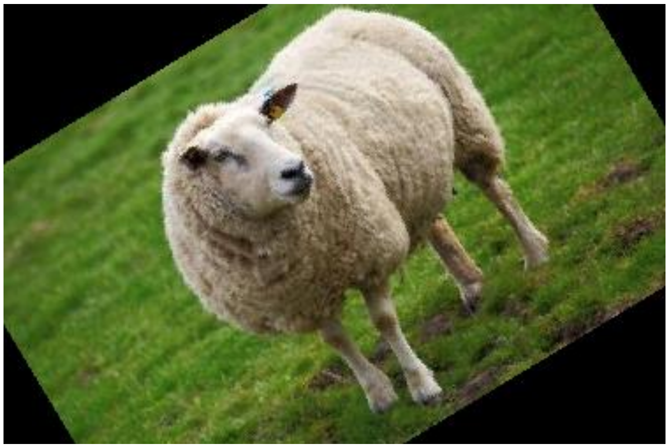
Part 2:

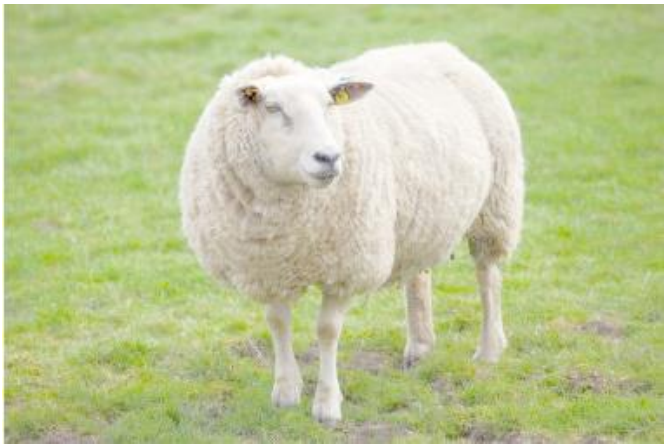
1. We chose to load the VGG16 as our classifier.
2. We chose to use the supplied sheep image:

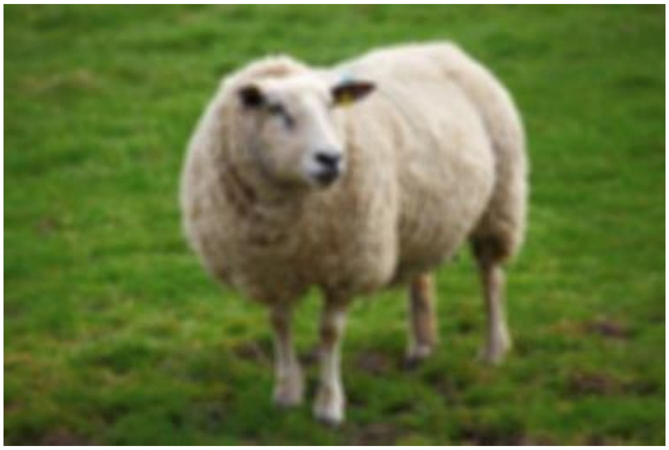
The network successfully classified the image as class number 348 ('ram, tup').

3. We used rotation (31 degrees) as our geometric transformation, gamma correction (gamma=0.3) as our color

transform and gaussian blur with filter (5,5) and sigma=5 as our filter.







4. For the rotation, when we rotated up to 30 degrees the network has successfully classified the sheep but as soon as we

rotated to 31 degrees the sheep was classified as class number 999 ('toilet tissue, toilet paper, bathroom tissue').

For the color transform, the model classified the image as class number 132 ('American egret, great white heron, Egretta albus').

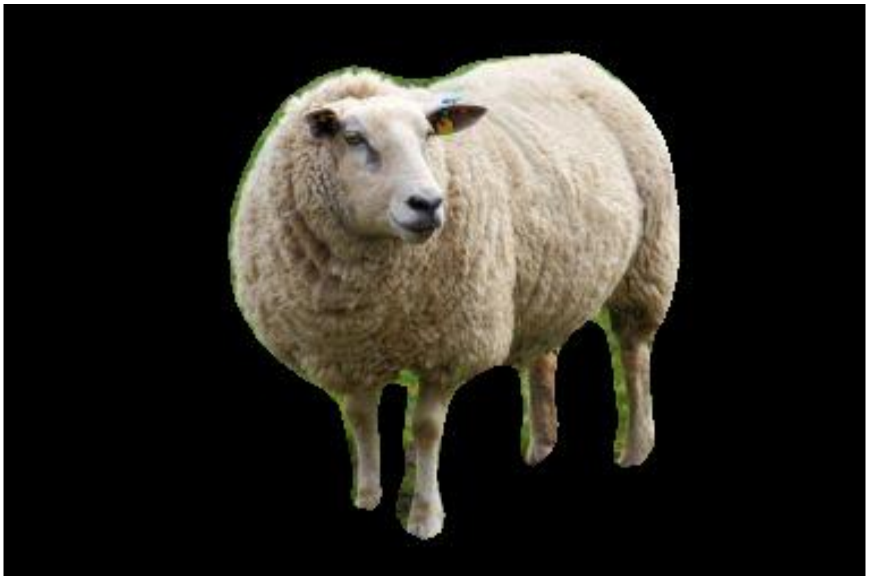
For the gaussian filter, the classification was once again class number 132.

We assume that the classifier is not robust to those transformation because it was not exposed to them while training.

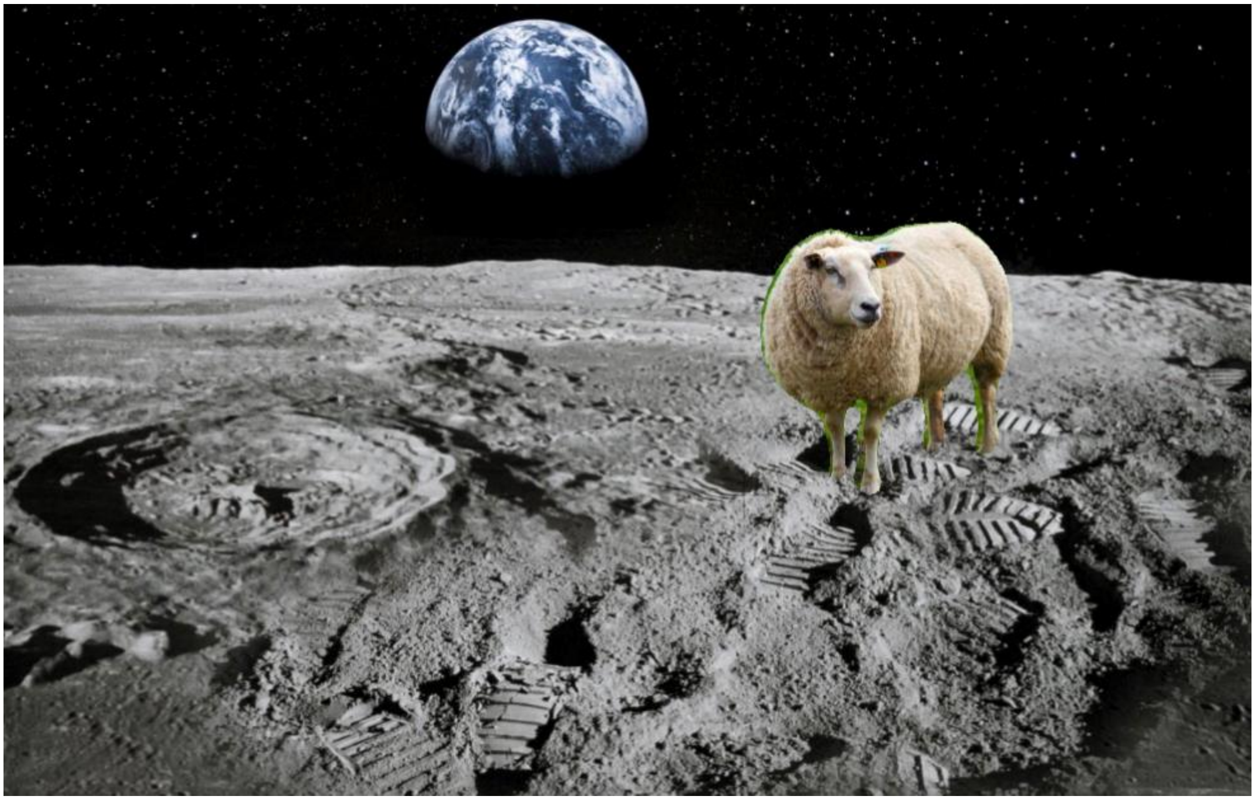
For example, we can assume that the training set only included sheeps up to some rotation because as soon as the sheep was rotated by more than 30 degrees it was classified as a toilet paper…

This could be alleviated by inserting rotations into the training. For example, exposing rotated sheep and vertical toilet papers to the network would have probably make the network learn features that are invariant to those rotations. The same applies for other transforms as well…

5. We segmented the sheep using deeplab with a backbone of resnet101.



6. We have placed the sheep in space:



The image was classified as class number 296: 'ice bear, polar bear, Ursus Maritimus, Thalarctos maritimus'.

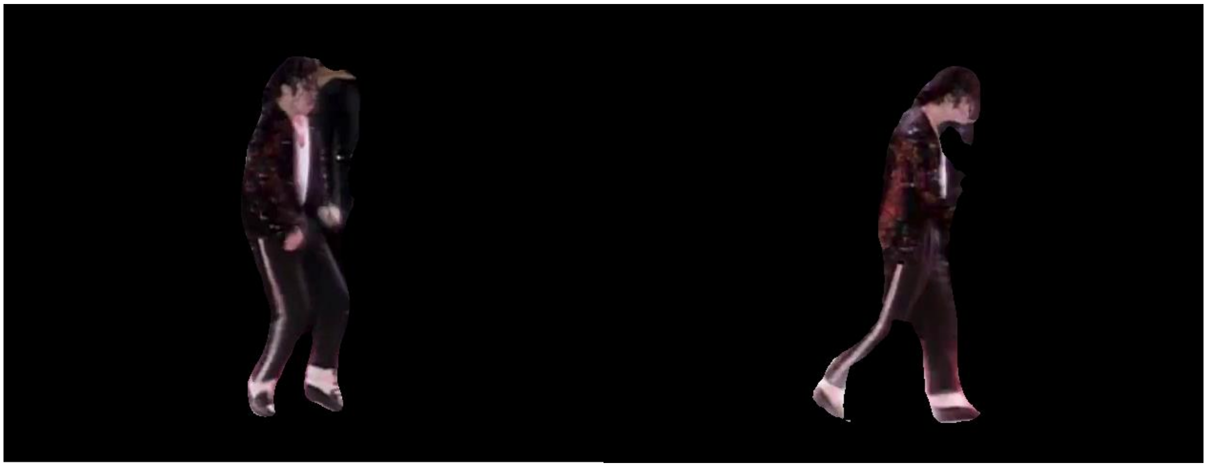
The reason is that the network is biased to assume that sheeps are surrounded by green background. Probably because the training set consisted mostly of sheeps surrounded by grass. Therefore the learned features for sheep classification include some activation from the background and as soon as we replaced the background the network classified the sheep as a polar bear which live in the arctic (greyish background…).

Part 3:

1. We picked a video of Michael Jackson doing the famous moonwalk.

   The resolution is (630, 480). Below are the first and last frames:

2. We used deeplabv3_resnet101 in order to segment “person”



It did not work perfectly at all. We assume this is because of bad lighting, low resolution and fact the person is sideways to the camera.

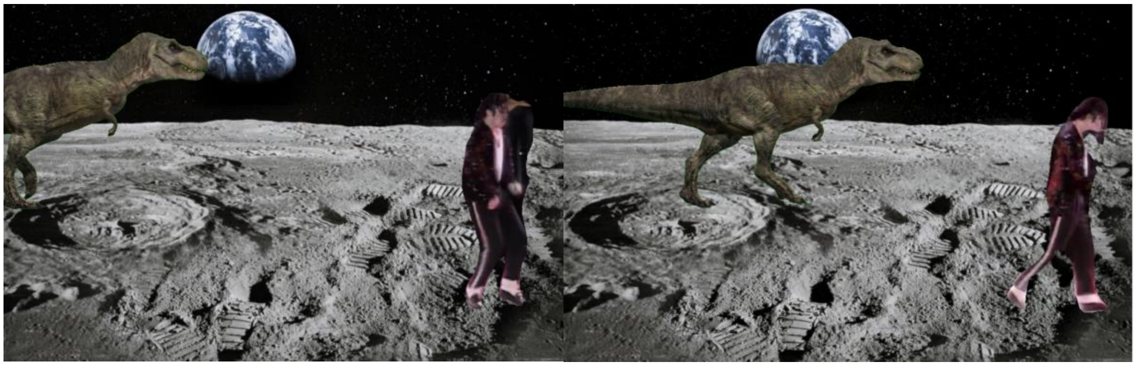
Nevertheless, MJ is still recognizable.

3. We picked the dinosaur video model.

We used only classical tools.

We first projected the image into the CIELAB space. Then we applied thresholding in the A axis to retrieve the mask. Lastly, we applied the mask to the image to extract the foreground.

4. Here are the first and last frames of our video:



Part 4 - Dry Questions

1. Ordinary segmentation is the task to assign to each pixel a label. For example, segmenting foreground and background or just finding general segments in an image. Famous classical algorithms for ordinary segmentation are GrabCut and mean-shift.

   Semantic segmentation extends the task to assigning labels from a set of classes. In this case each pixel holds a semantic context (person, chair, bird…). Unlike instance segmentation, all the identified pixels of “person” will have the same value even if there is more than one person.

   Nowadays Semantic segmentation is done mostly by DNNs. For example, the Deeplab family of networks.

2. IoU for each image:

a. $\boldsymbol{IoU = \frac{Area_{triangle}}{Area_{circle}} = \frac{\frac{r*2r}{2}}{\pi r^2} = \frac{1}{\pi} \approx 0.318}$

b. $\boldsymbol{IoU = \frac{Area_{triangle} \cap Area_{circle}}{Area_{triangle} \cup Area_{circle} - Area_{triangle} \cap Area_{circle}} =}$

$$\boldsymbol{\frac{\frac{\pi r^2}{2}}{\frac{\pi r^2}{2} + (r+h)*\frac{4r}{2}} = \frac{\frac{\pi r^2}{2}}{\frac{\pi r^2}{2} + \frac{9r}{8}*\frac{4r}{2}} = \frac{\pi}{\pi + \frac{36}{8}} \approx 0.411}$$