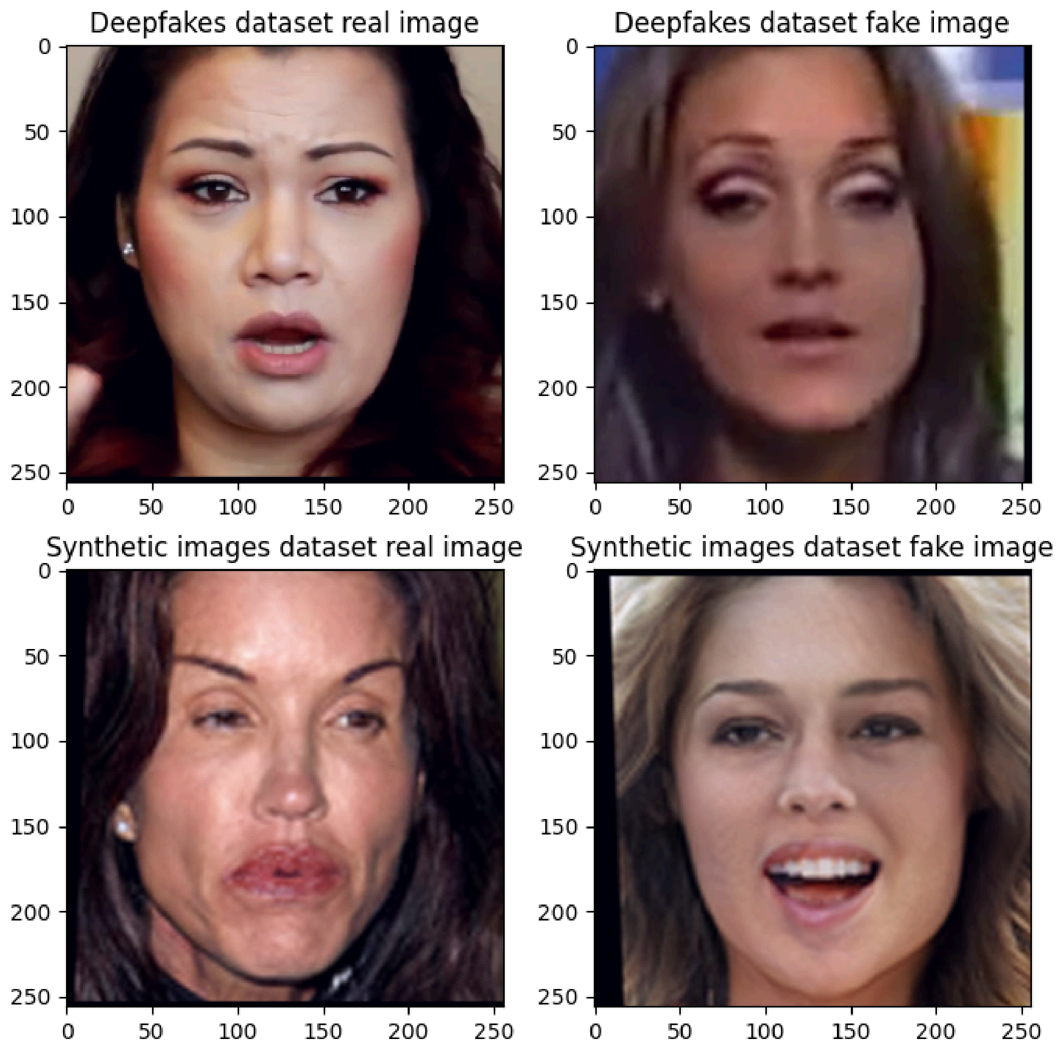
Final Project - Computer Vision

Tamir Cohen - 208006064
Moti Rizel - 037728201

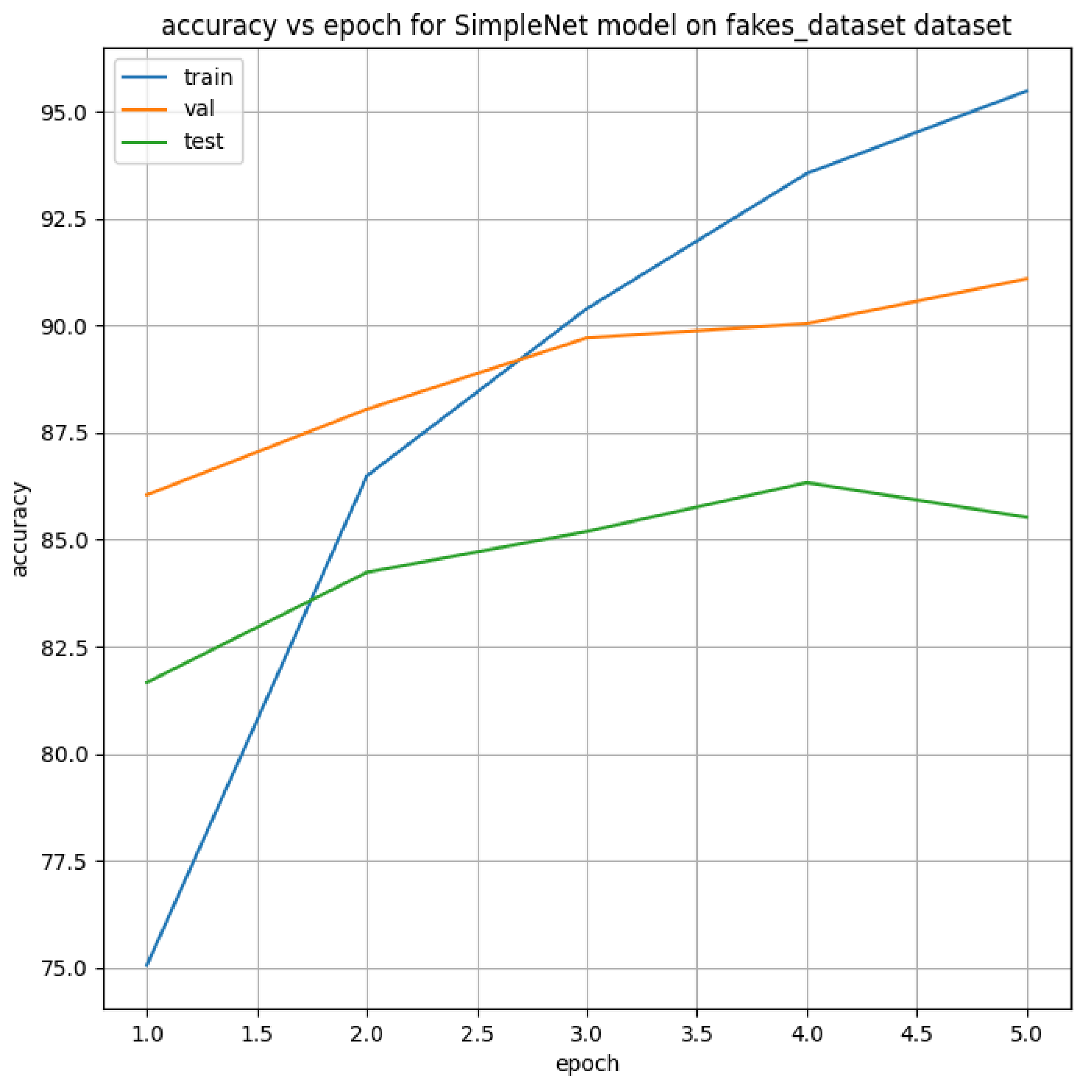
Question 2

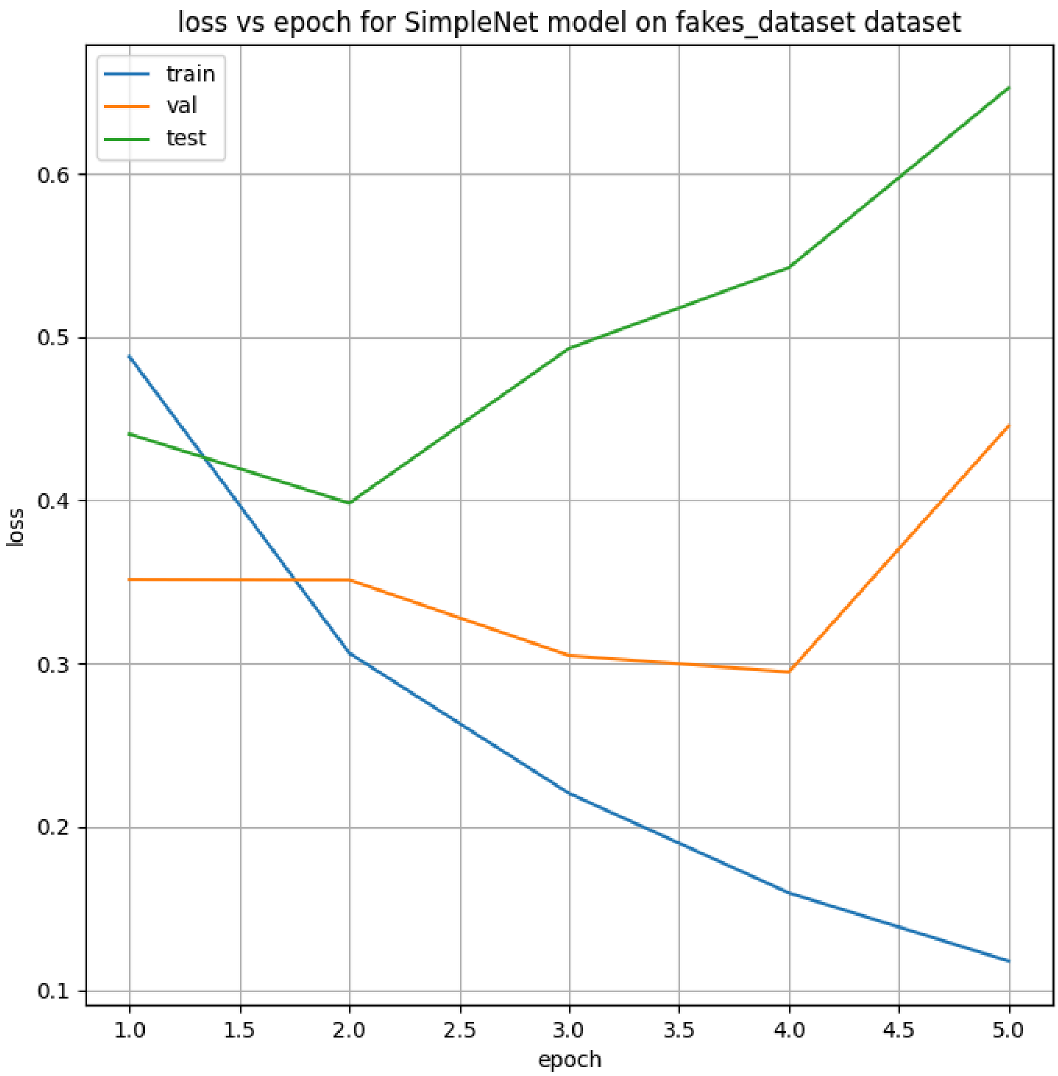


Question6

The data makes sense. It appears that the SimpleNet model was overfitted, we can see that the loss of the test started to increase after the 2nd epoch, we can also notice that the accuracy of the test got worse at the last epoch.

Question7





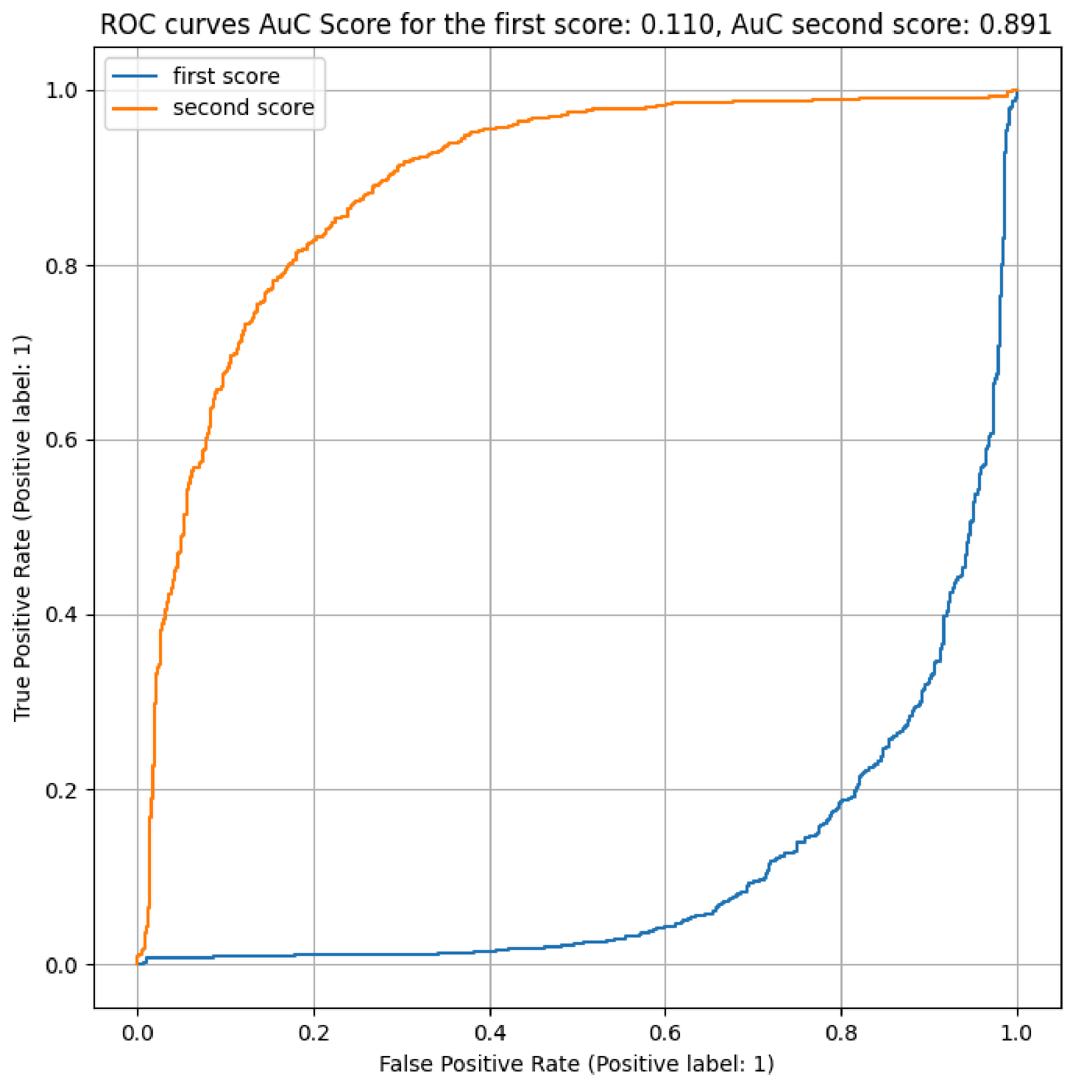
Question8

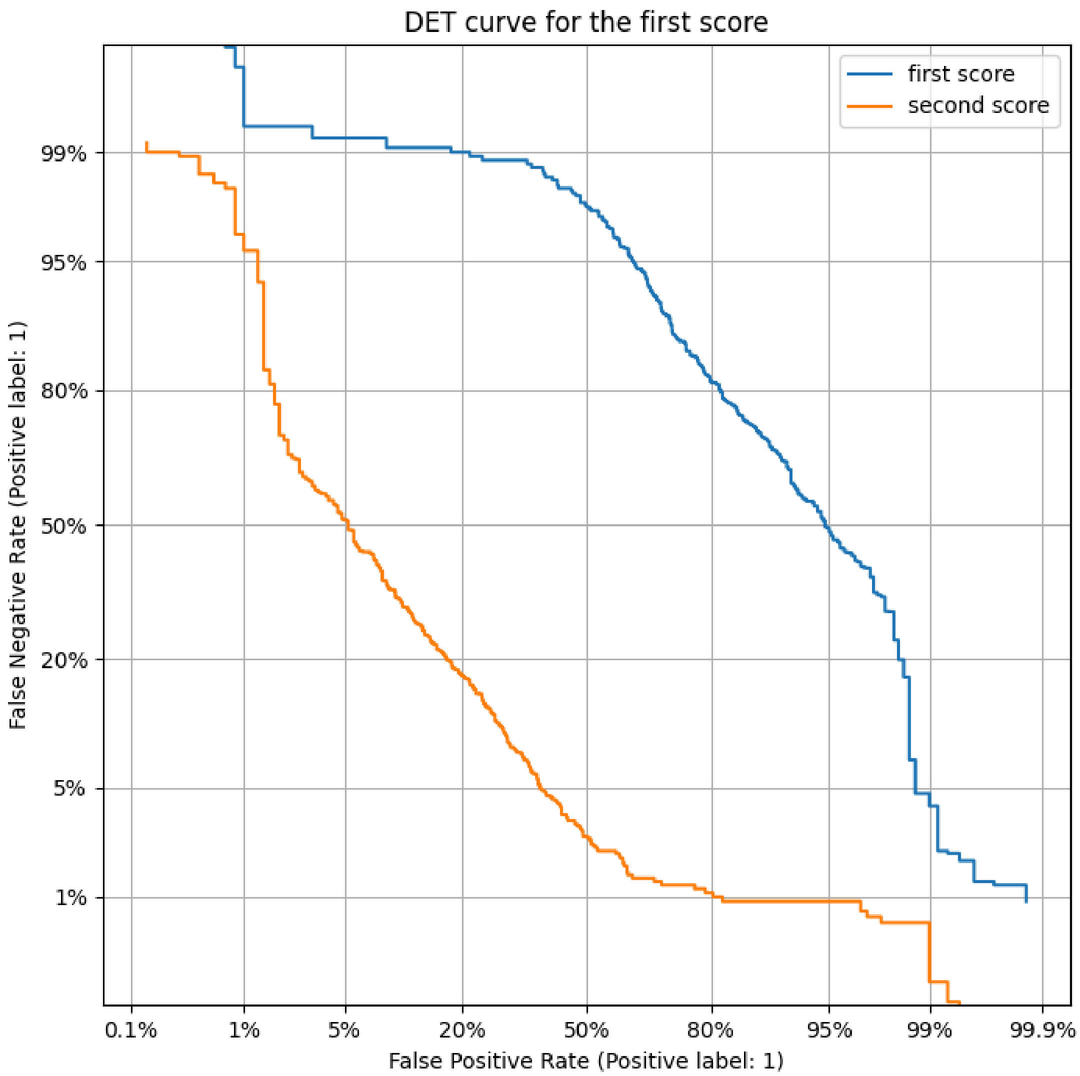
Test accuracy corresponding to the highest validation is 85.52380952380952

Question9

1400 real images VS 700 fake

Question10


ROC curves AuC Score for the first score: 0.110, AuC second score: 0.891
first score
second score
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0



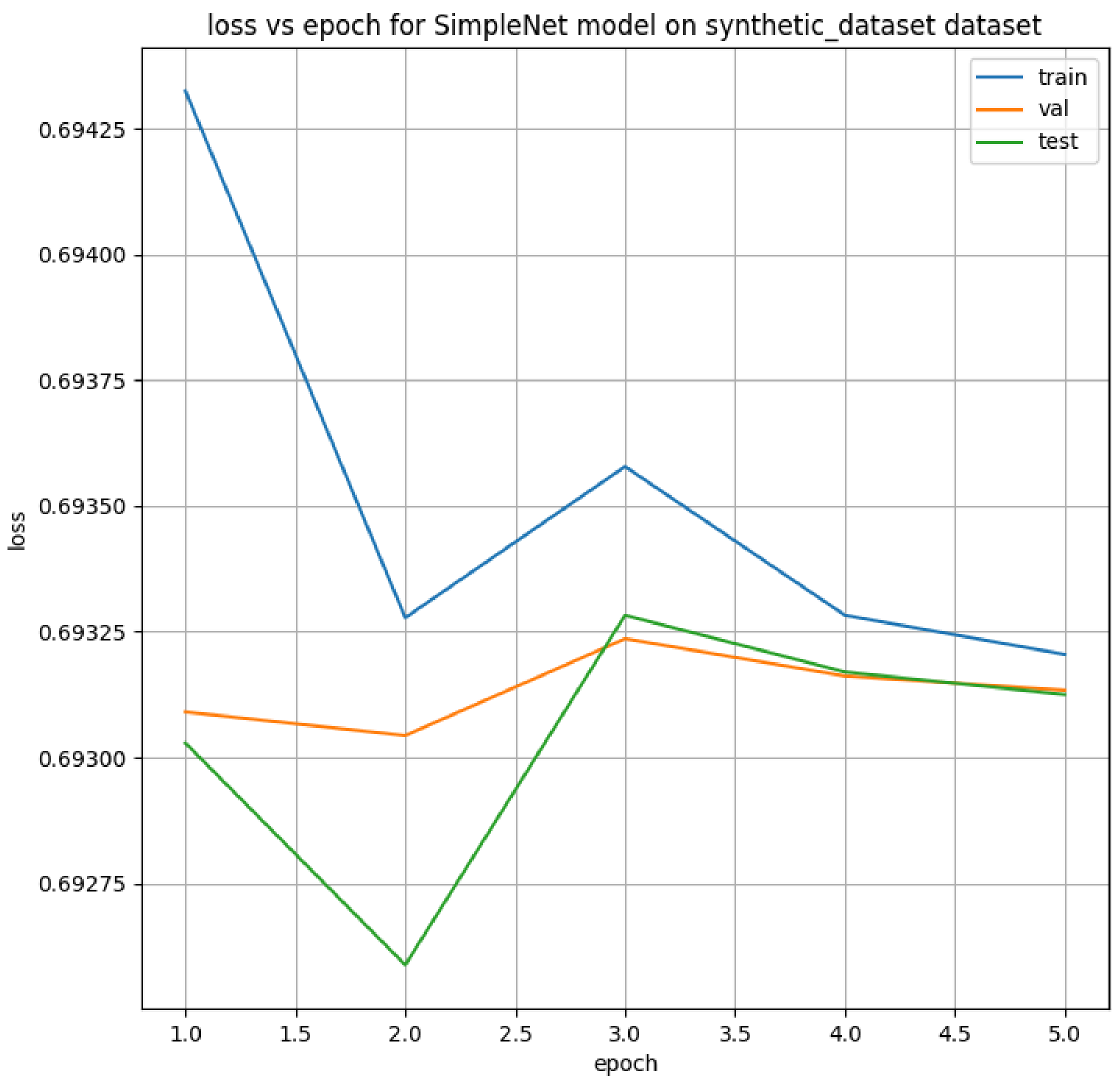
Question11

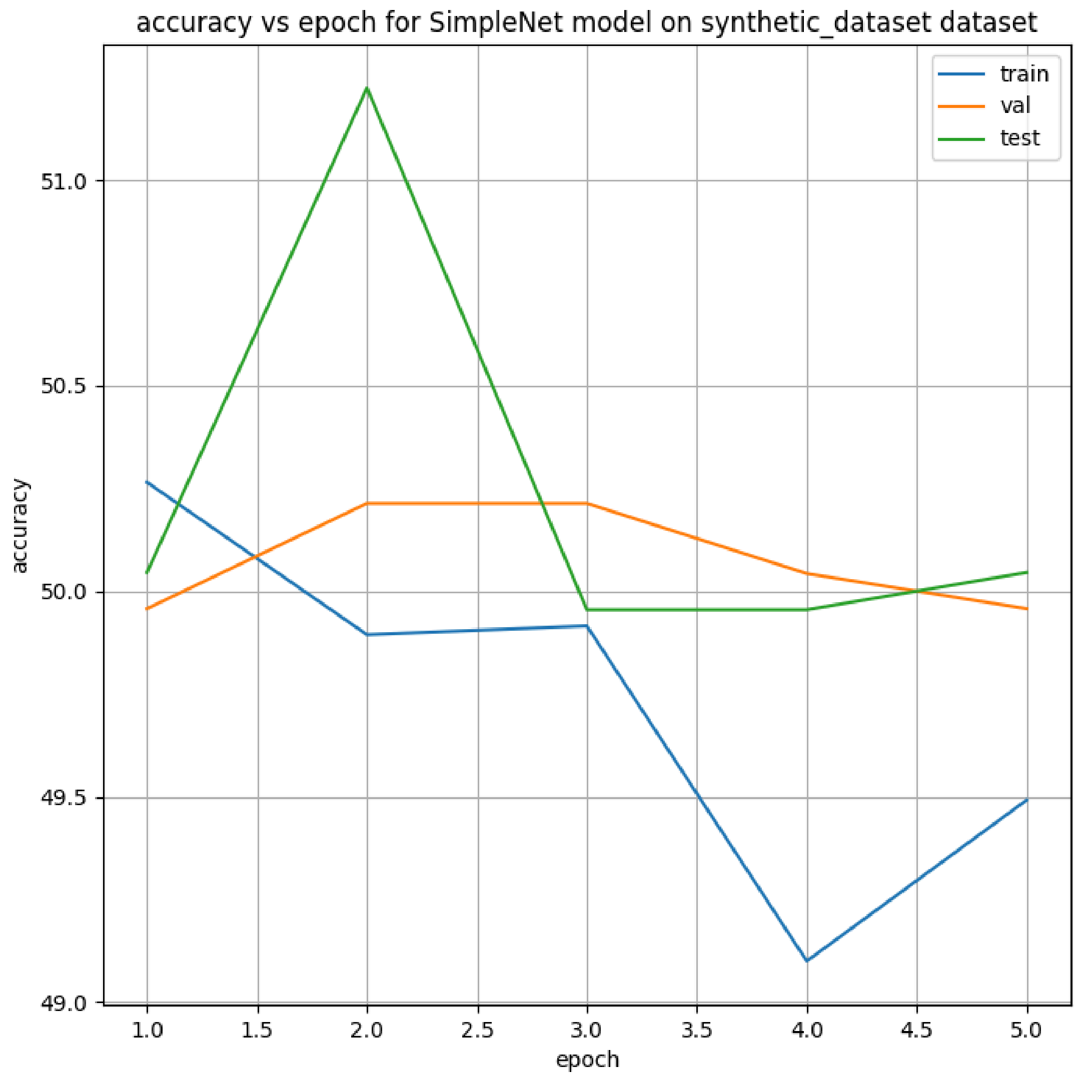
In our dataset, the label 1 is assigned to real images, while label 0 is assigned to fake images.
When we plot the roc curve using sklearn.metrics.roc_curve
We provide the labels and the probability of the positive labels for the fake images (that roc curve defaults the positive label to 1 if not specified otherwise).

So for it to be plotted correctly we need to use the soft scores that correspond to the real image probability which is soft score 2

Soft score 1 has more False positive than True Positive because it uses the probability of fake images to classify real images

Question13





Question14

The test accuracy corresponding to the highest validation accuracy is 51% (Which is pretty bad)

Question 15

551 real images and 552 fake images (about 50/50)

Question 16

We received a random classifier, that chooses labels randomly (because we achieved accuracy of 50%)

Question17

It makes sense that the simpleNet model could not generalize because it was harder for a human eye (my eye) to separate between the synthetic images and the real images.

Question18

It is pre-trained on datasets with image and labels,
specifically in the paper it was trained on ImageNet and JFT

Question19

The basic building blocks comprise depthwise separable convolution batch normalization and relu.

Depthwise separable convolutions, comprise two layers.
Firstly, there is the depthwise convolution layer where each input channel possesses its dedicated kernel. Following this, there is the pointwise convolution layer, which serves to combine all the channels. This approach surpasses traditional convolution since standard convolution kernels have parameters associated with each input channel.

Question20

(Duplicate of Question18)

Question21

The input size of fc is 2048

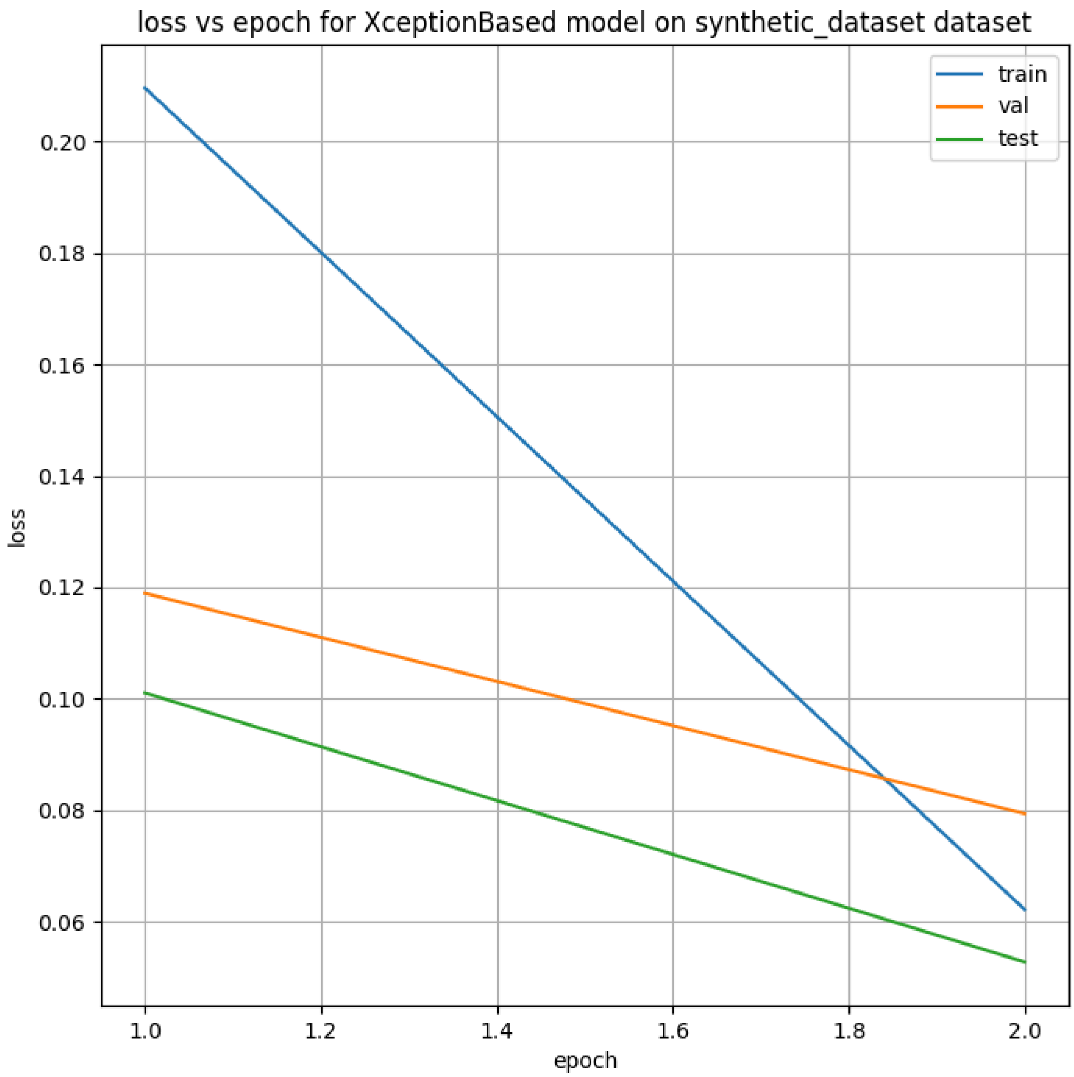
Question22

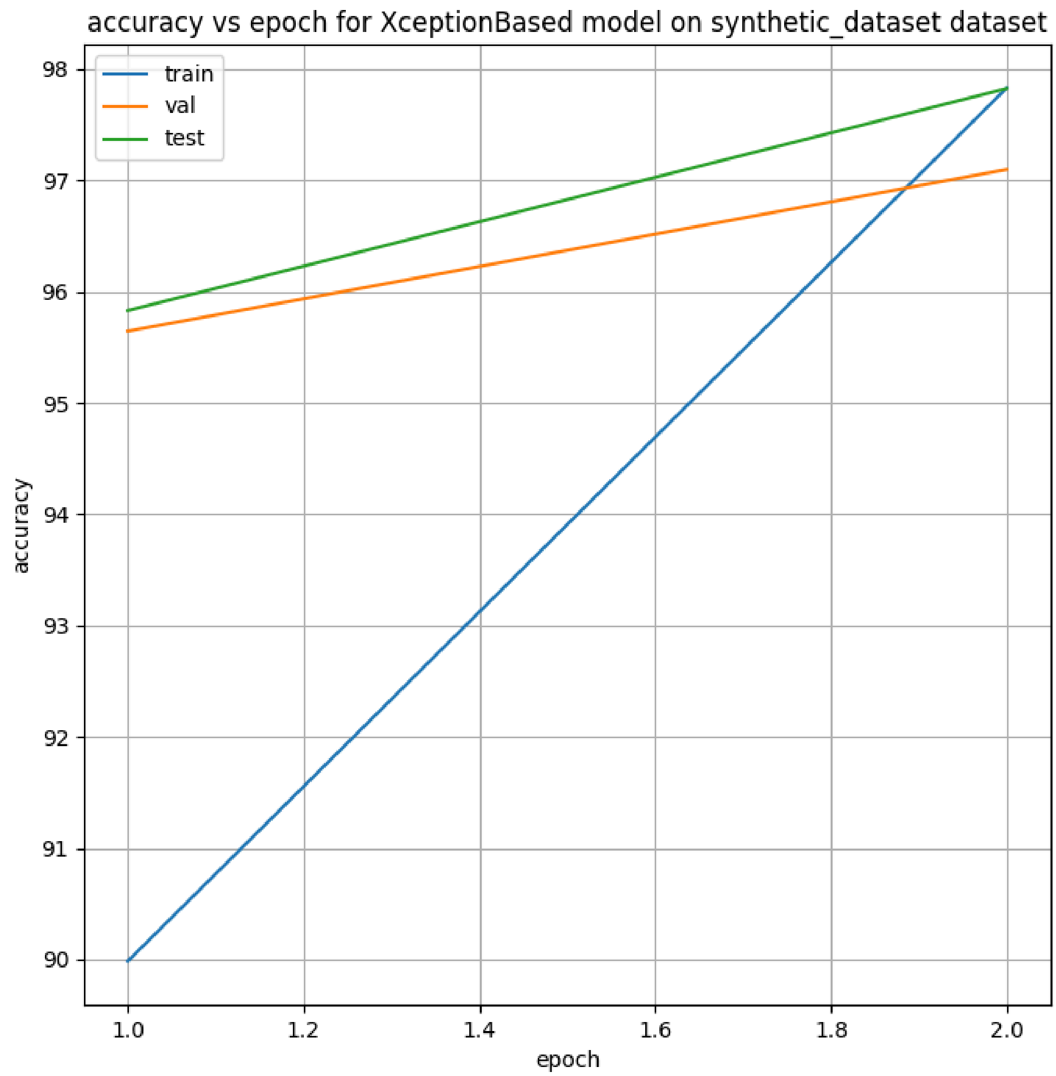
The number of parameters of the Xception module is 22,855,952 according to the paper.
get_nof_params returned exactly the same number.
I printed it using - print(get_nof_params(Xception()))

Question24

We added 272834 which is 1.2%

Question26

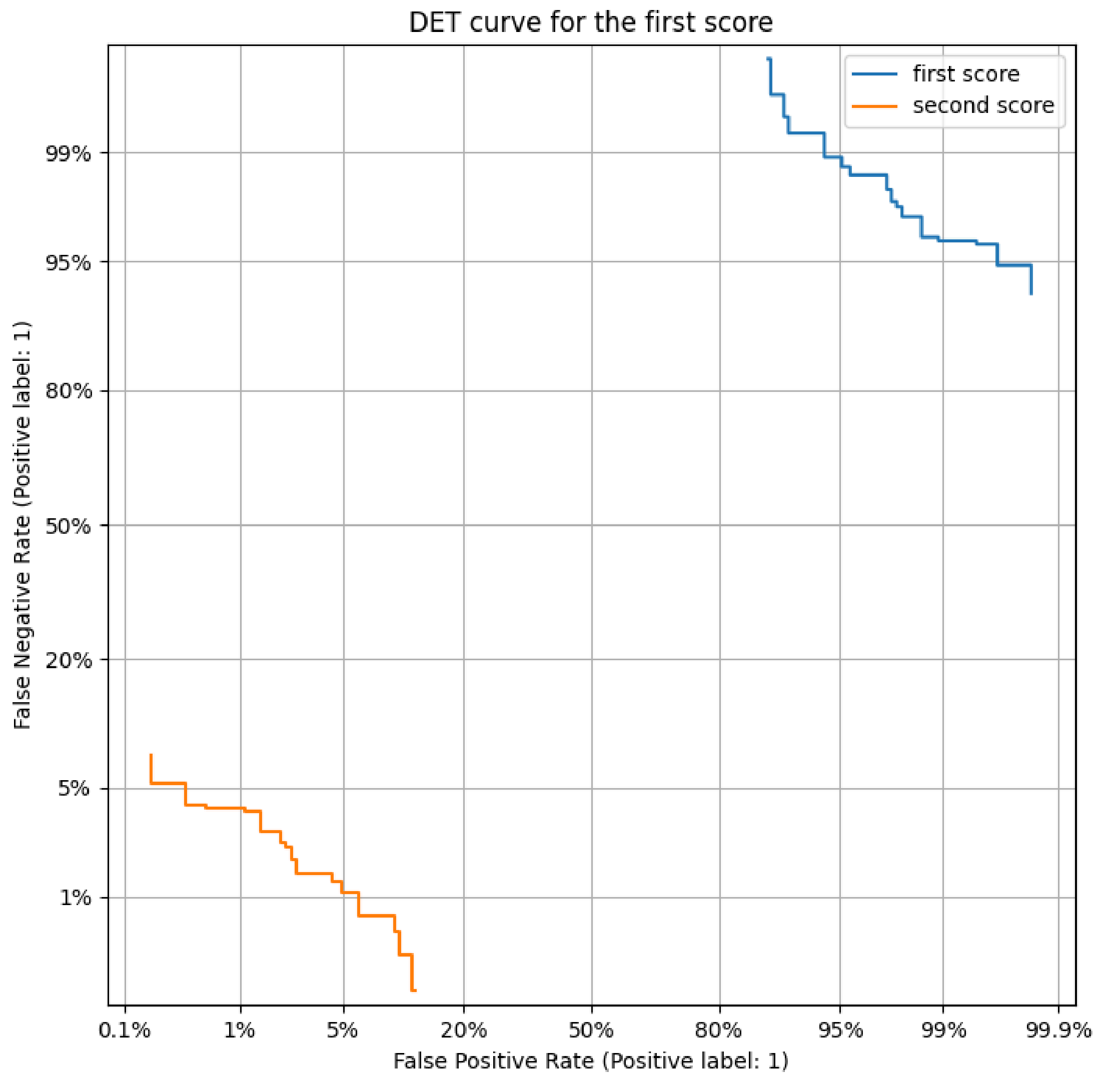


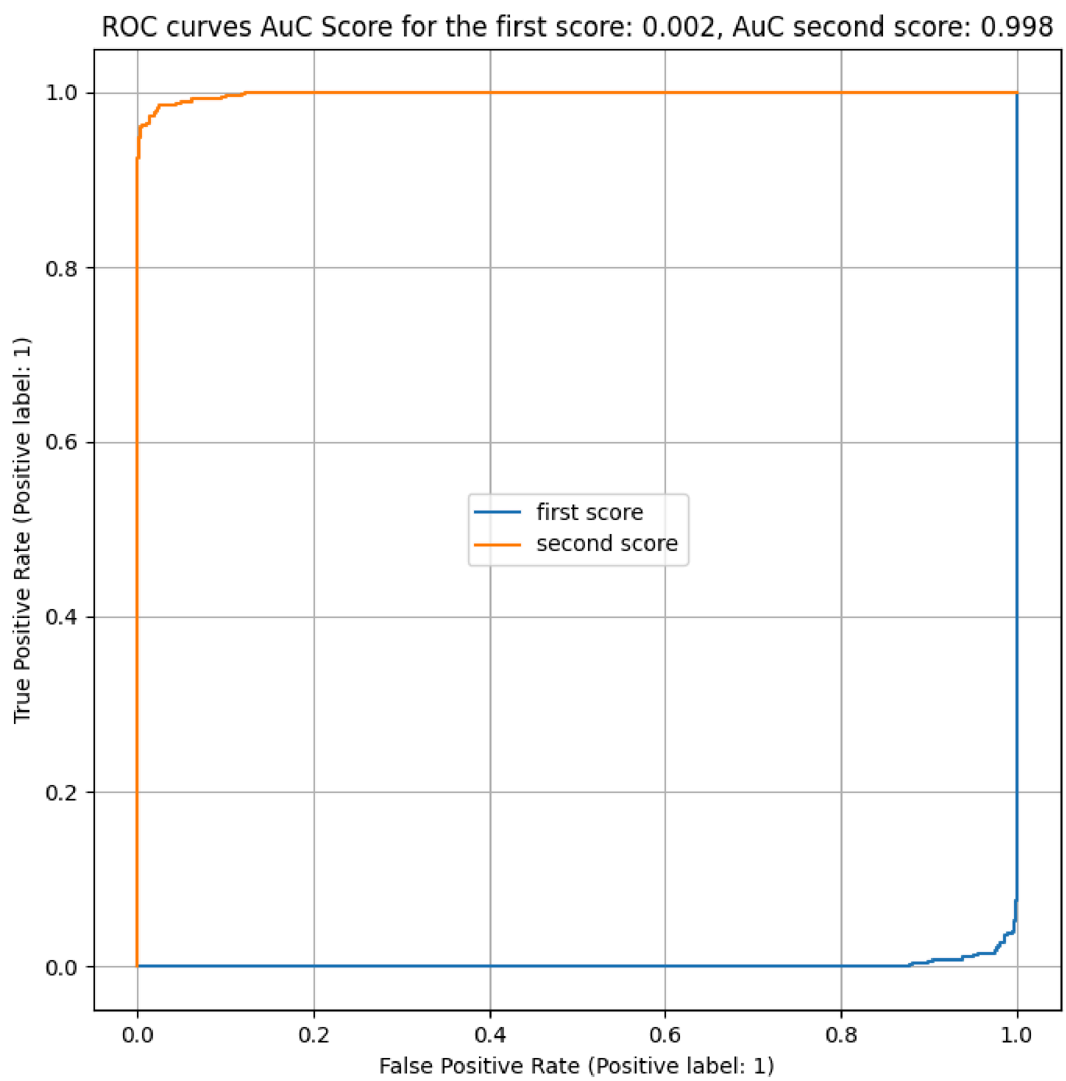


Question27

The test accuracy for the highest validation accuracy that we received for the Xception module is 97.82 (WOW!)

Question28


DET curve for the first score
first score
second score
99%
95%
80%
50%
20%
5%
1%
0.1%
1%
5%
20%
50%
80%
95%
99%
99.9%
False Negative Rate (Positive label: 1)
False Positive Rate (Positive label: 1)



Question29

A saliency map for a specific image indicates which pixels are crucial for classifying the image. This map is derived from the class score by calculating the pixel's derivative.

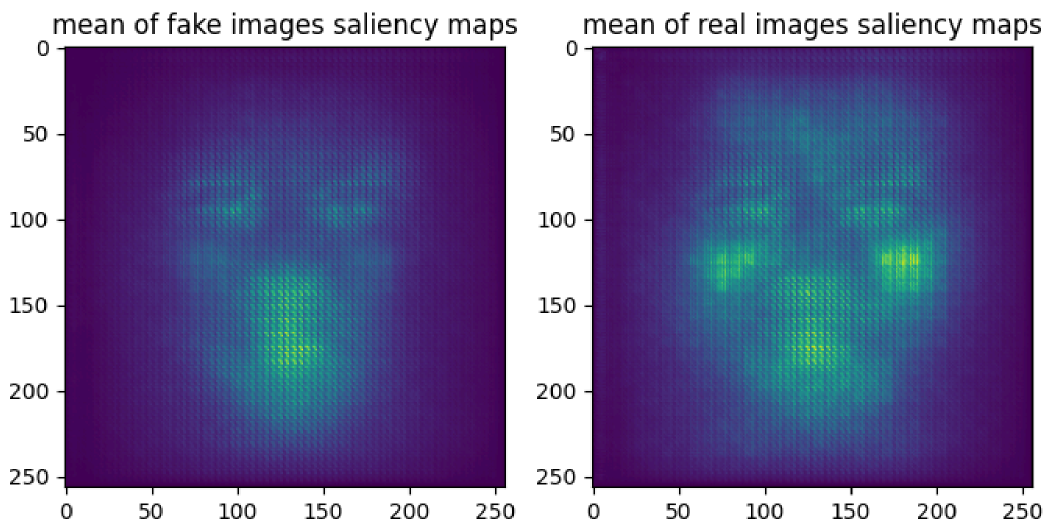
Question30

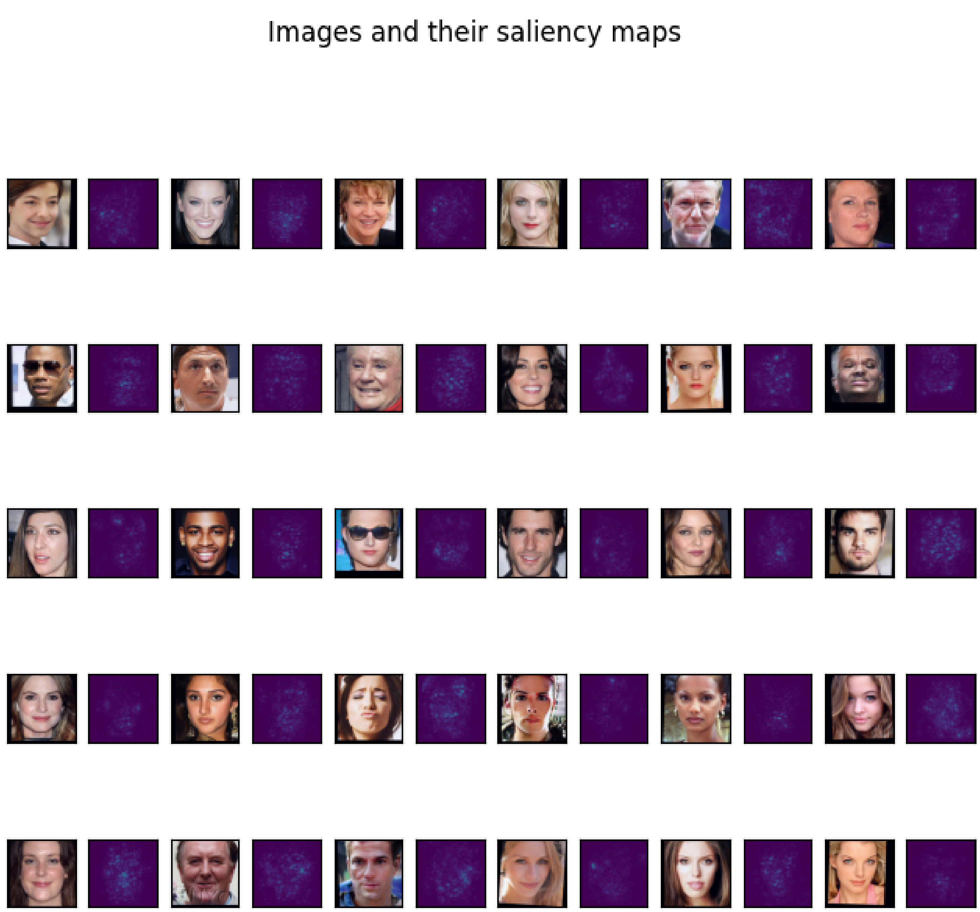
GradCam is a technique to visualize what feature map contributes the most for the image class classification score. It does so by derivate the classification score by the feature map pixels (of the last cnn layer), and produces a weighted combination of all the pixel maps

Question32

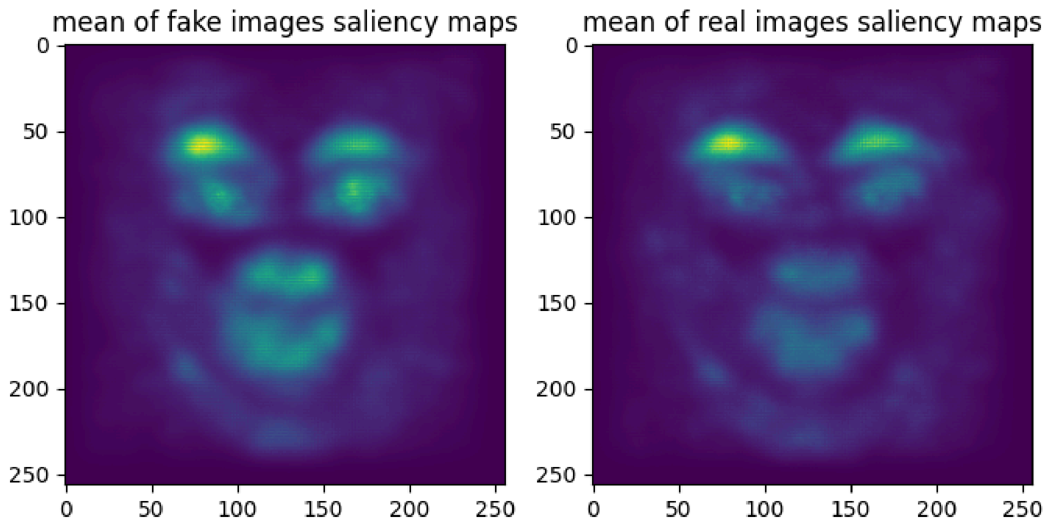
In the mean saliency map we see the most relevant pixels for classification.
The main features are the eyes and the mouth.
For example we can see that the mouth is a feature that can indicate that an image is fake, and that the cheeks can indicate that the image is real.

Xception Model Saliency Map





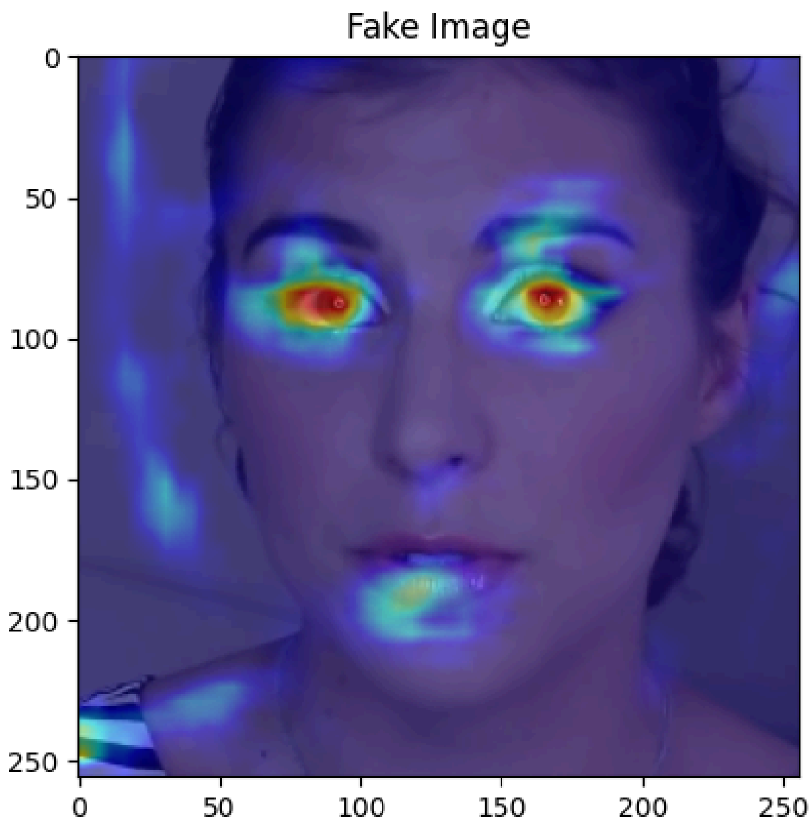
SimpleNet Model Saliency Map

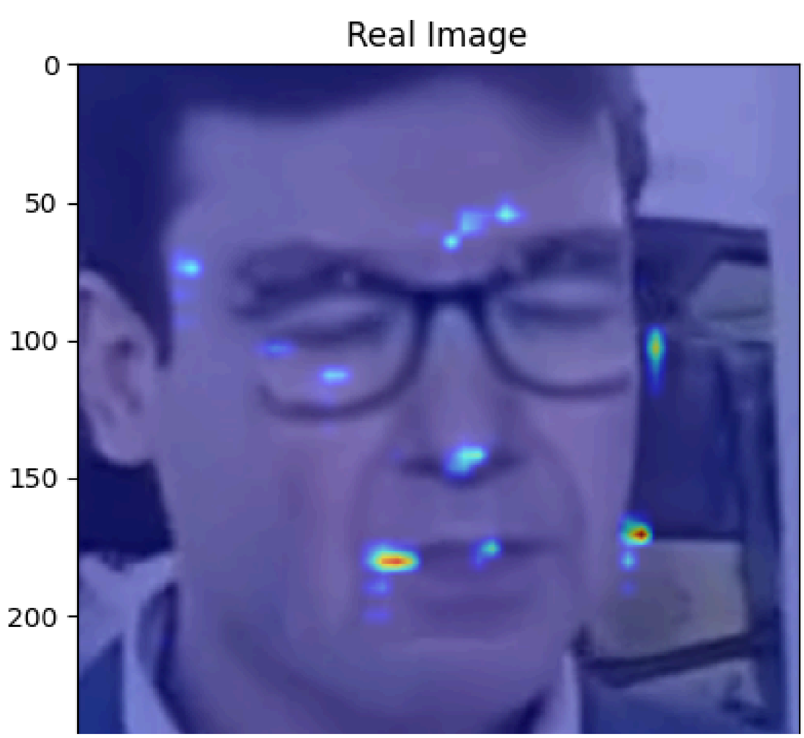


Images and their saliency maps

Question35

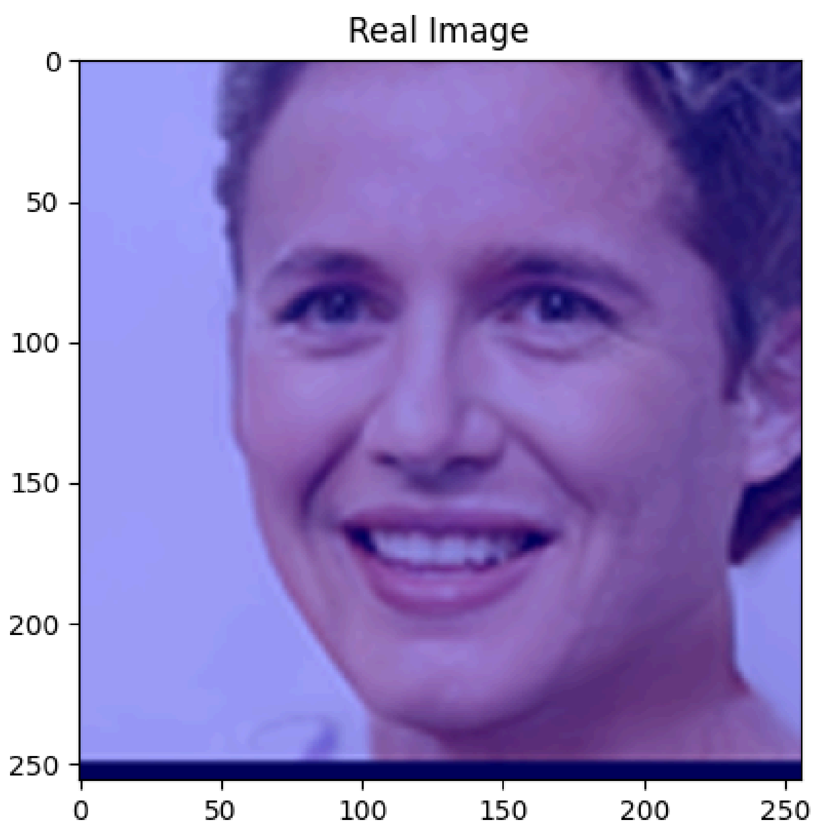
SimpleNet - Fake data set



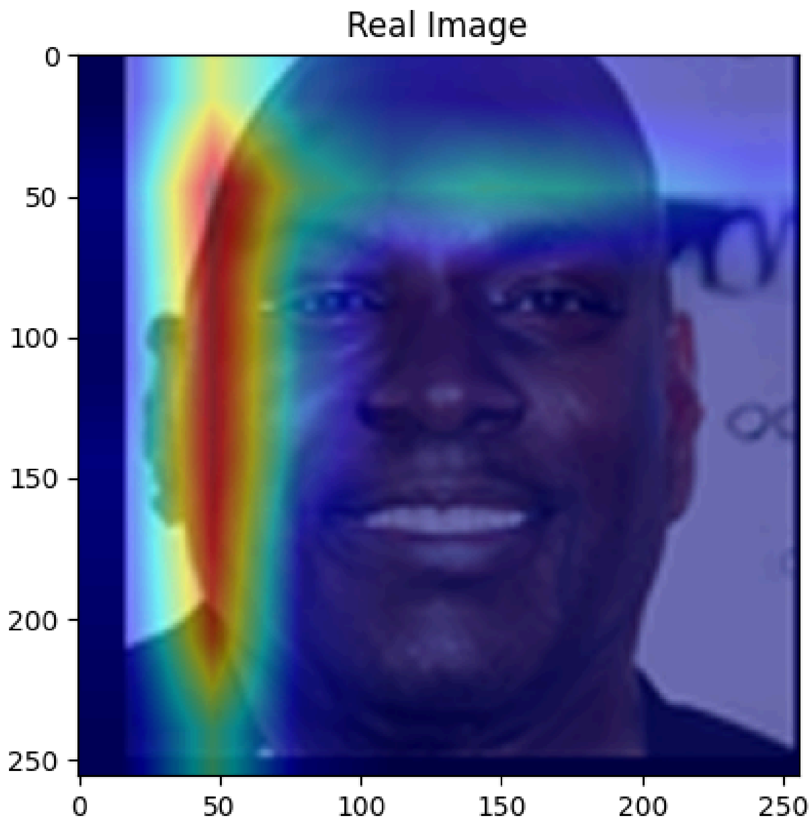


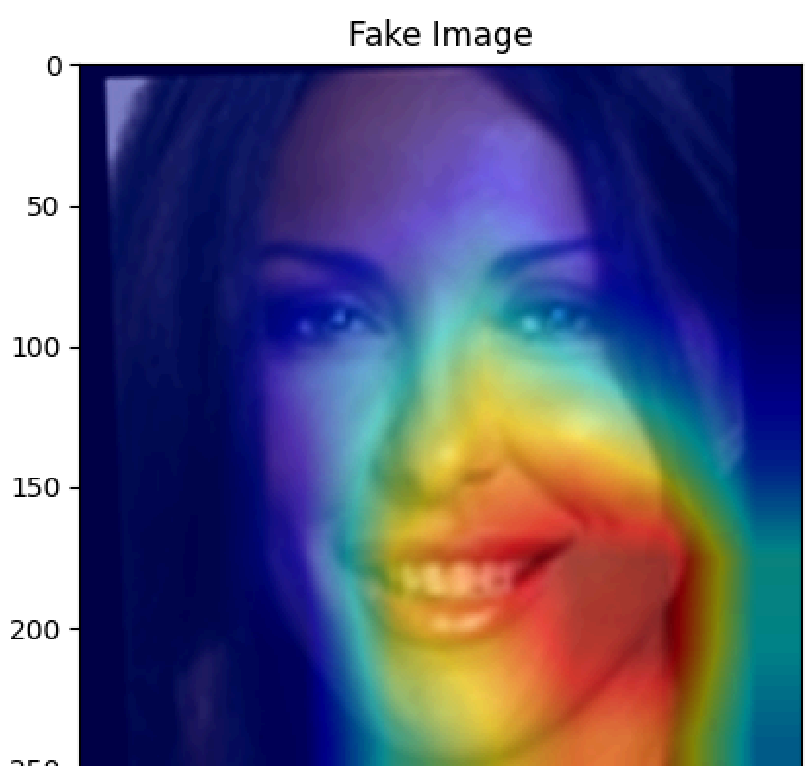
SimpleNet - Synthetic data set

No gradients is showed because the SimpleNet did not generalized on the Synthetic dataset



Xception Model - Synthetic data set





Bonus Model

We called the new model **LilXception**. The idea was to take the Xception module and remove layers inside, and decrease the layer inner dimensions.
We tried this method because Xception is trained on ImageNet which classifies a lot of labels, and we only need classification on two labels.

The model architecture

1. Decreased the Xception Blocks dim from 728 to 256
2. Decreased the Xception Blocks repetitions from 3 to 2
3. Removed 4 Xception block out of 12
4. Decreased the dims of the last 2 separable convolution from 1536, 2048 to 400 500 correspondingly
5. Added 4 Fully connected linear layers instead of the last fc layer of the Xception

Results

Model size: 1811372 (More than 10 times smaller than the Xception Model!!)
Accuracy: 93.86%