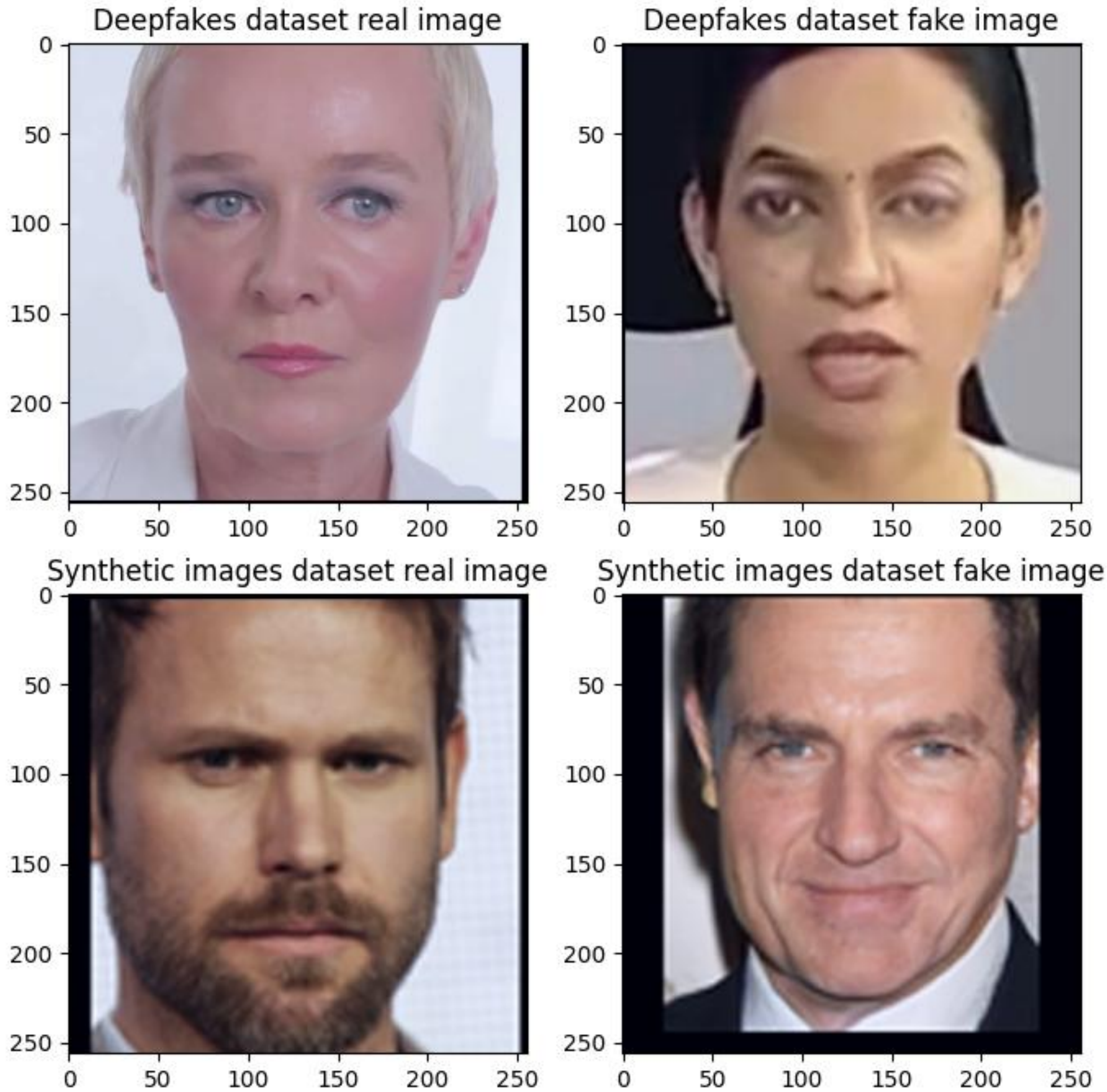
# Computer Vision Course Project

Liav Raccah – 208298984 ; Alam Joaeh – 206702417

Chapter 2:

Q1:
Done.

Q2:



Chapter 3:

Q3:

Done.

Q4:

Done.

Q5:

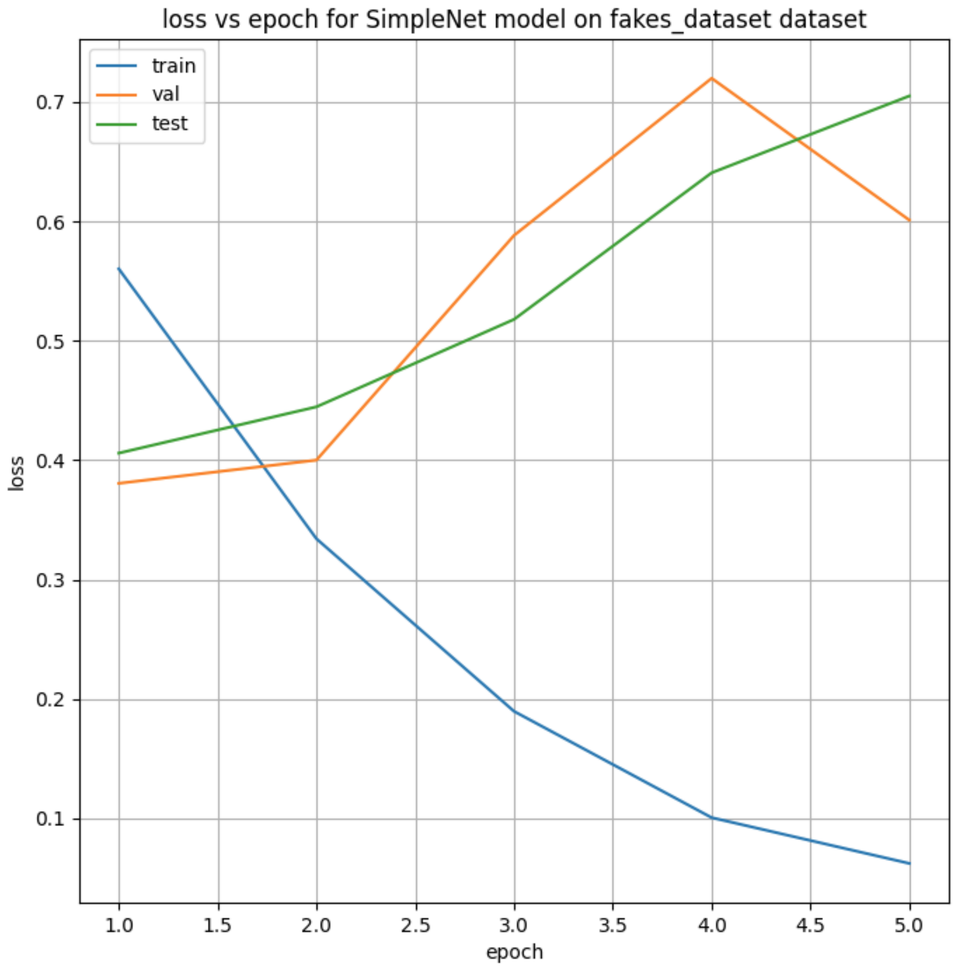
Done.

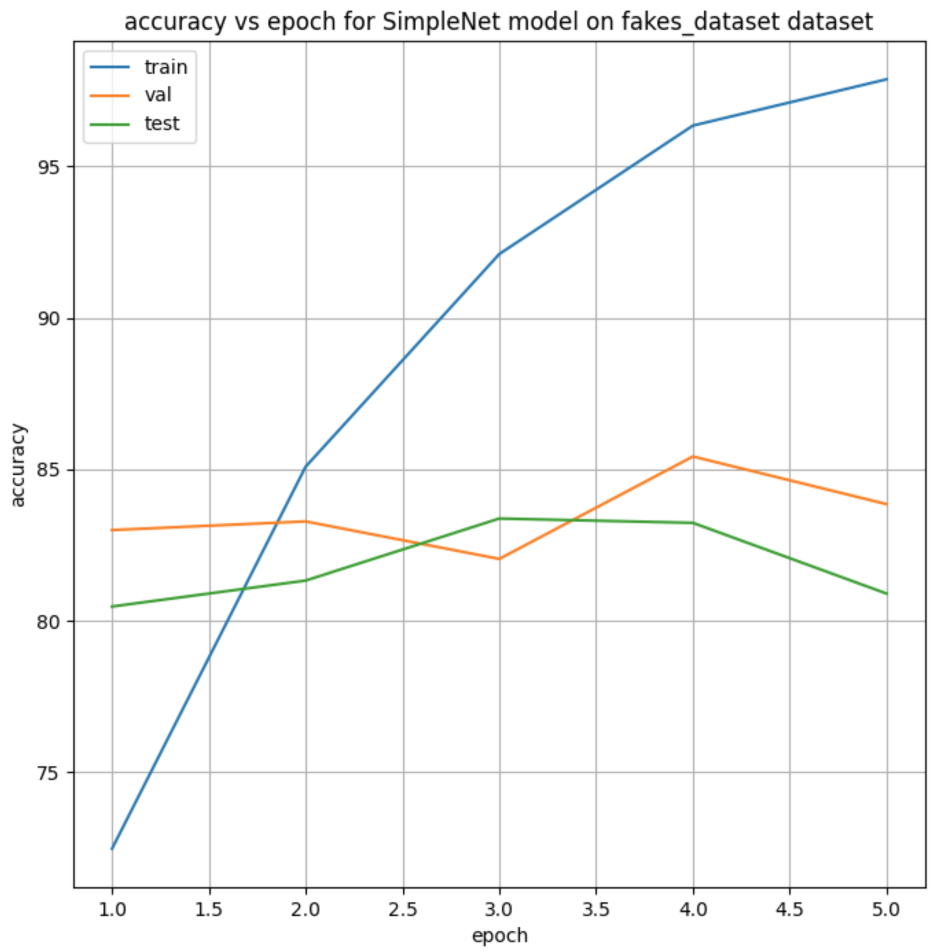
Q6:

Yes the Results make sense and demonstrate typical behavior of training a NN:

1. Learning: the model is clearly learning as the training loss decreases from 0.53 to 0.08 and the training accuracy increases from 74% to 96%.
2. Overfitting: we can see signs of overfitting. While the training accuracy continues to increase the validation, accuracy stays the same and the loss increases (also for test). This indicates the model is memorizing the training data rather than generalizing.

Q7:


loss vs epoch for SimpleNet model on fakes_dataset dataset
train
val
test
loss
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
epoch


accuracy vs epoch for SimpleNet model on fakes_dataset dataset
train
val
test
accuracy
95
90
85
80
75
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
epoch

Q8:

The highest validation accuracy was in epoch 4 is 85.4%.

The test accuracy corresponding to that epoch is 83.2%.

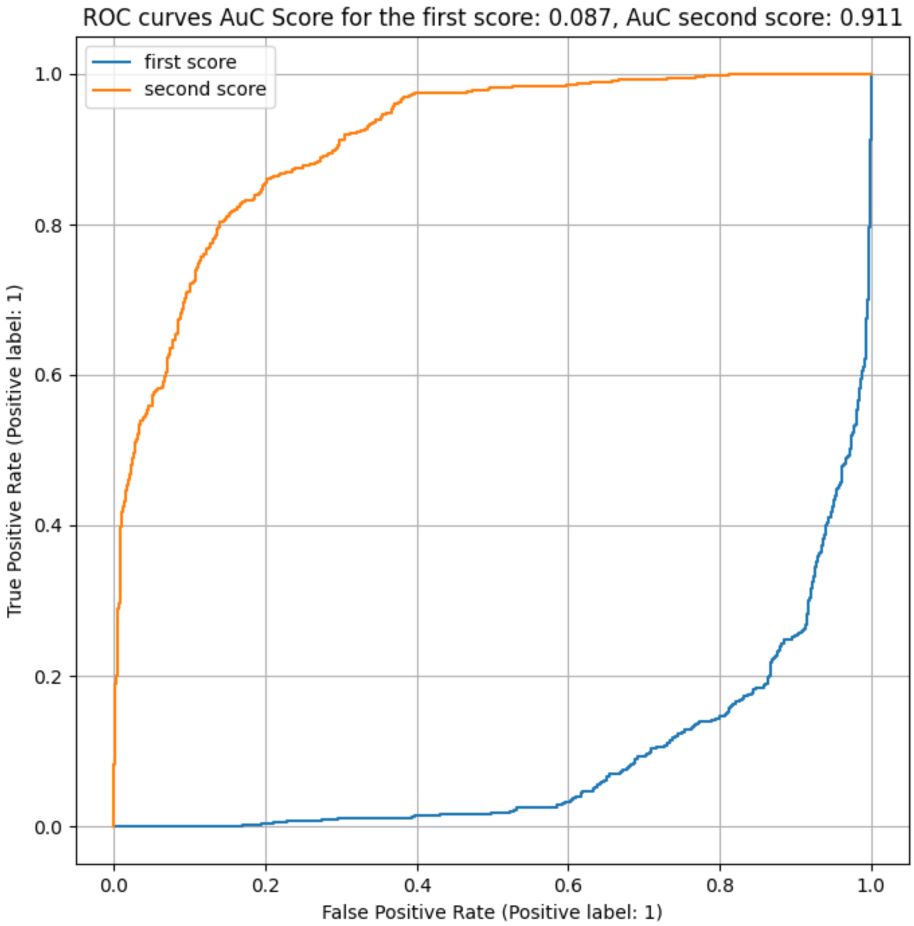
Q9:

Fake images: 700

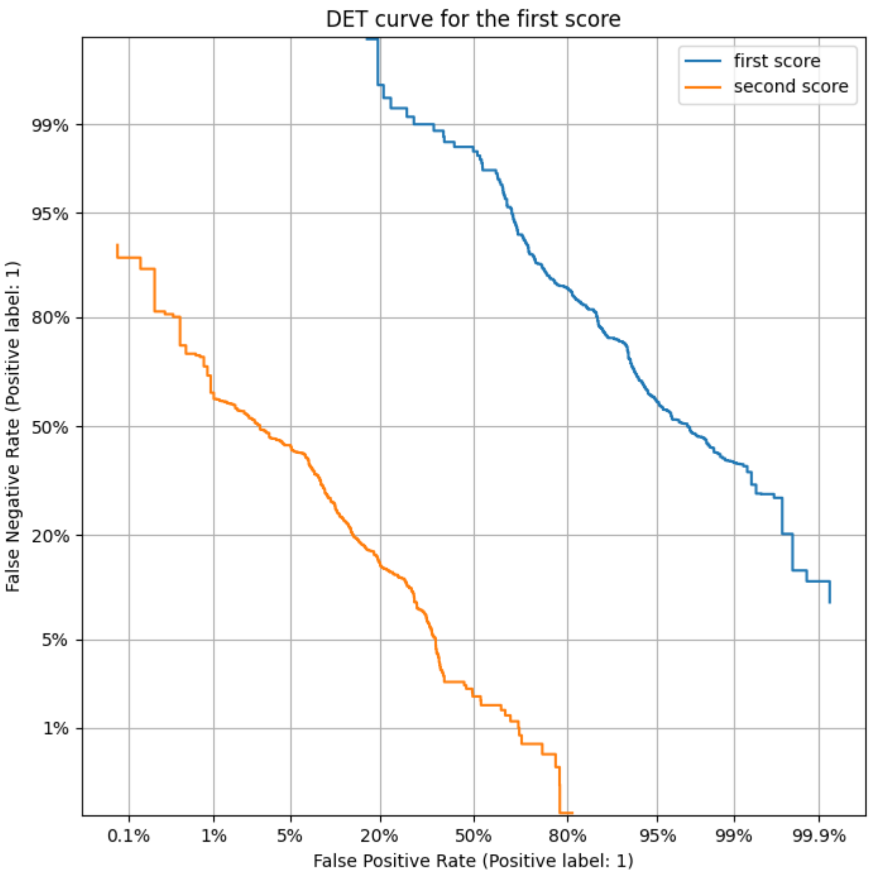
Real images: 1400

33% of the dataset is fake.

Q10:



ROC curves AuC Score for the first score: 0.087, AuC second score: 0.911
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

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

Q11:

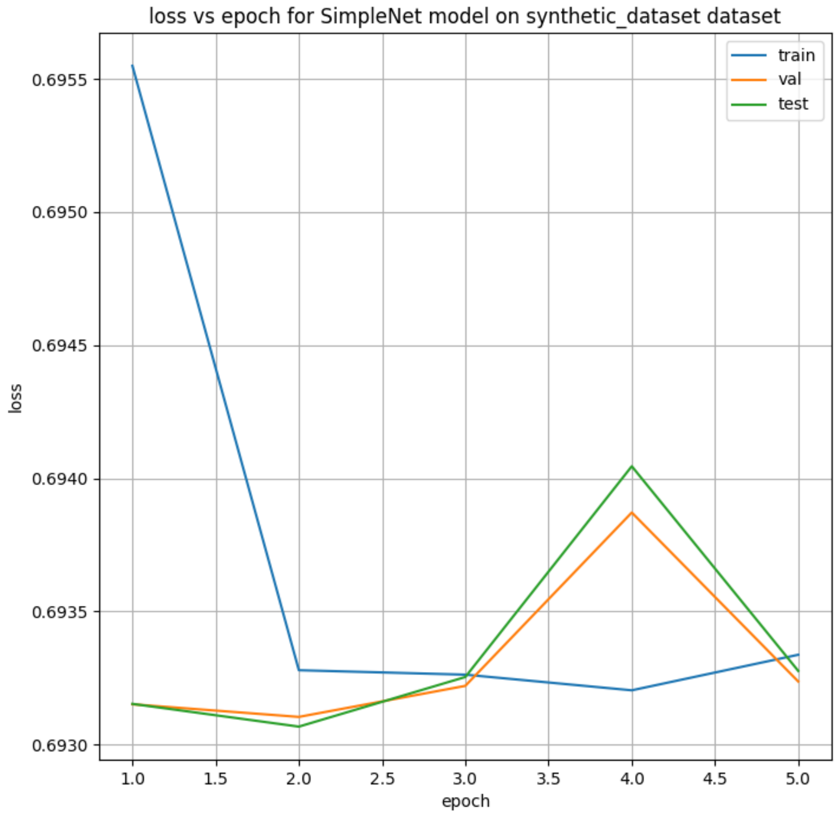
Second score (fake score): Represent the probability of the image being fake. Since this matches the positive label, the graph shows a standard, high-performance curve (positive correlation).

First score (real score): represents the probability of the image being real. Since real is the opposite of fake, a high score here means the image is likely not the positive label, this negative correlation creates an inverted graph, as the score predicts the opposite of the target class.

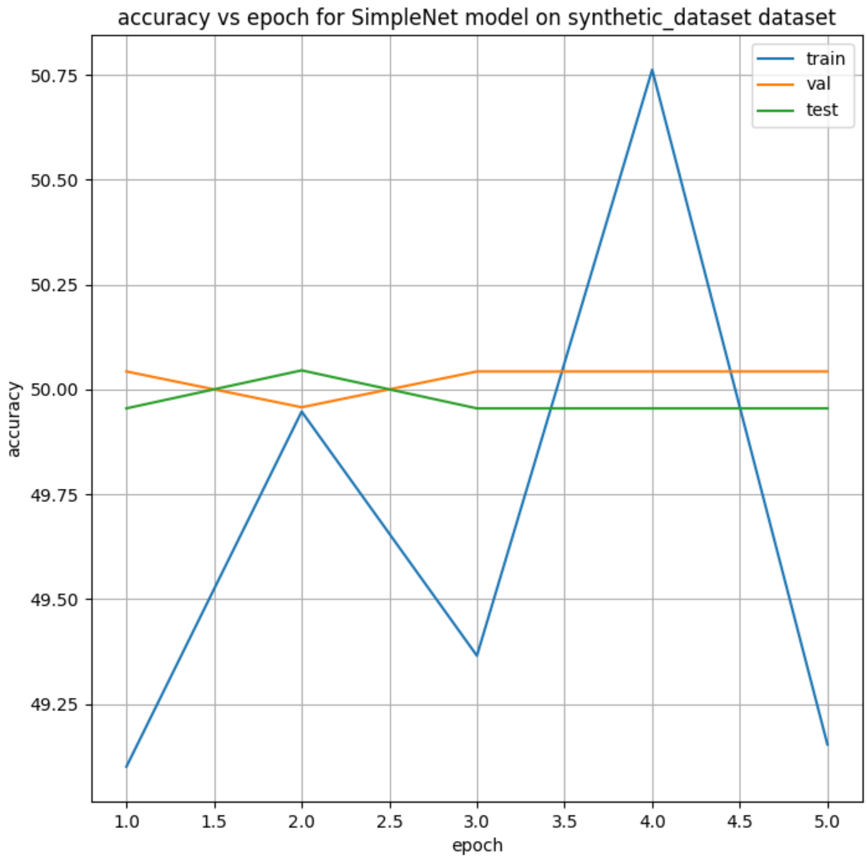
Q12:

Done.

Q13:



loss vs epoch for SimpleNet model on synthetic_dataset dataset
train
val
test
loss
0.6955
0.6950
0.6945
0.6940
0.6935
0.6930
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
epoch

accuracy vs epoch for SimpleNet model on synthetic_dataset dataset
train
val
test
accuracy
50.75
50.50
50.25
50.00
49.75
49.50
49.25
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
epoch

Q14:

The highest validation accuracy was in several epochs and it is 50%.

The test accuracy corresponding to that epochs is 49.9%

Q15:

Fake images: 552

Real images: 551

50% of the dataset is fake. Balanced test set.

Q16:

A random classifier. There is 50% chance to get real/synthetic classification.

Q17:

The synthetic dataset contains fake images that are often very realistic and can match the statistics of real faces closely. Small SimpleNet may fail and stay at chance.

In deepfake dataset, there are manipulations with artifact patterns that the network can learn more easily, so simplenet results are better.

Chapter 4:

Q18:

Pre-trained on ImageNet dataset.

Q19:

The basic building block of Xception is the depthwise separable convolution. Instead of performing a single complex calculation, this core unit splits the work into two simpler steps: depthwise convolution (for spatial patterns) and pointwise convolution (for channel relationships). Structurally, the architecture uses residual connections and a linear stack, arranging these blocks one after another in a straight line.

Q20:
As Q18.

Q21:

The input feature dimension to the final FC block is 2048.

Q22:

The network hold 22,855,952 parameters.

Q23:

Done.

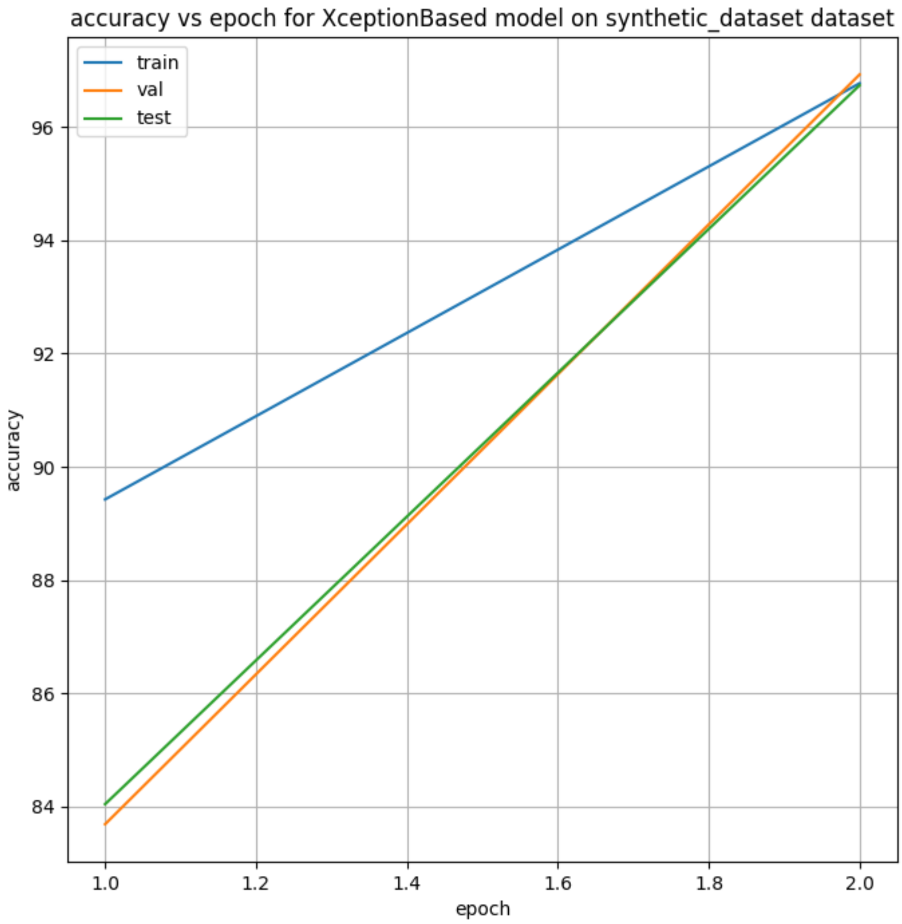
Q24:

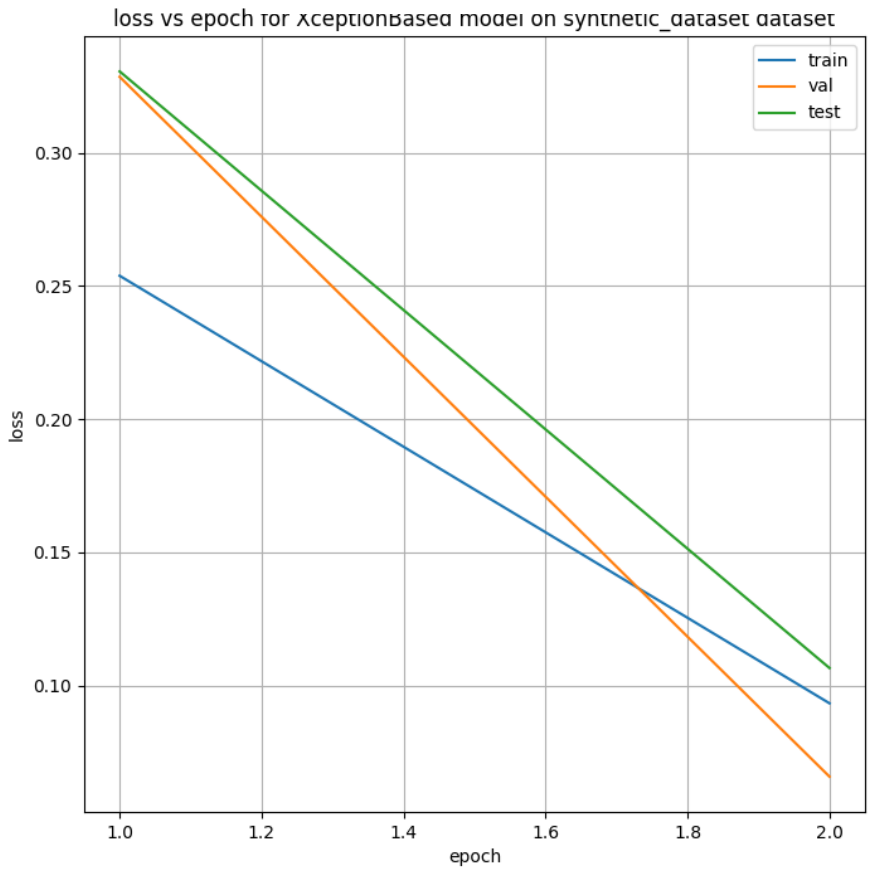
We added 272,834 parameters.

Q25:

Done.

Q26:


accuracy vs epoch for XceptionBased model on synthetic_dataset dataset
train
val
test
accuracy
epoch
96
94
92
90
88
86
84
1.0
1.2
1.4
1.6
1.8
2.0

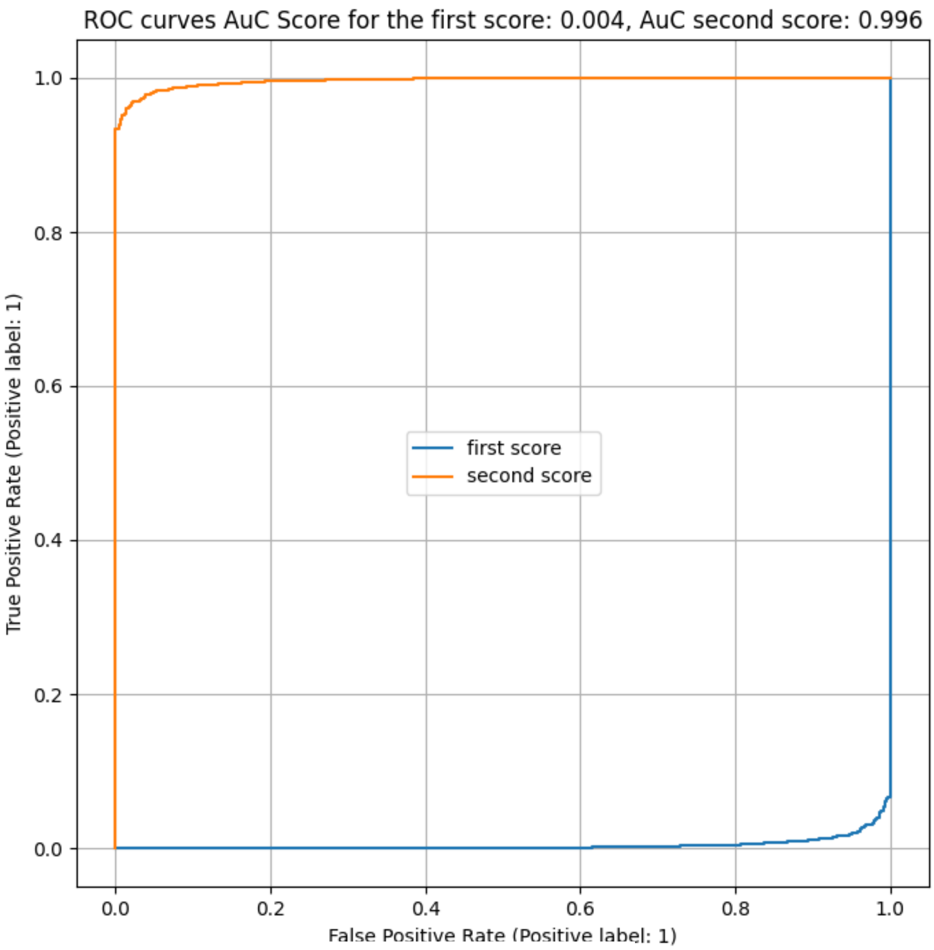

loss vs epoch for XceptionBased model on synthetic_dataset dataset
train
val
test
loss
epoch
0.30
0.25
0.20
0.15
0.10
1.0
1.2
1.4
1.6
1.8
2.0

Q27:

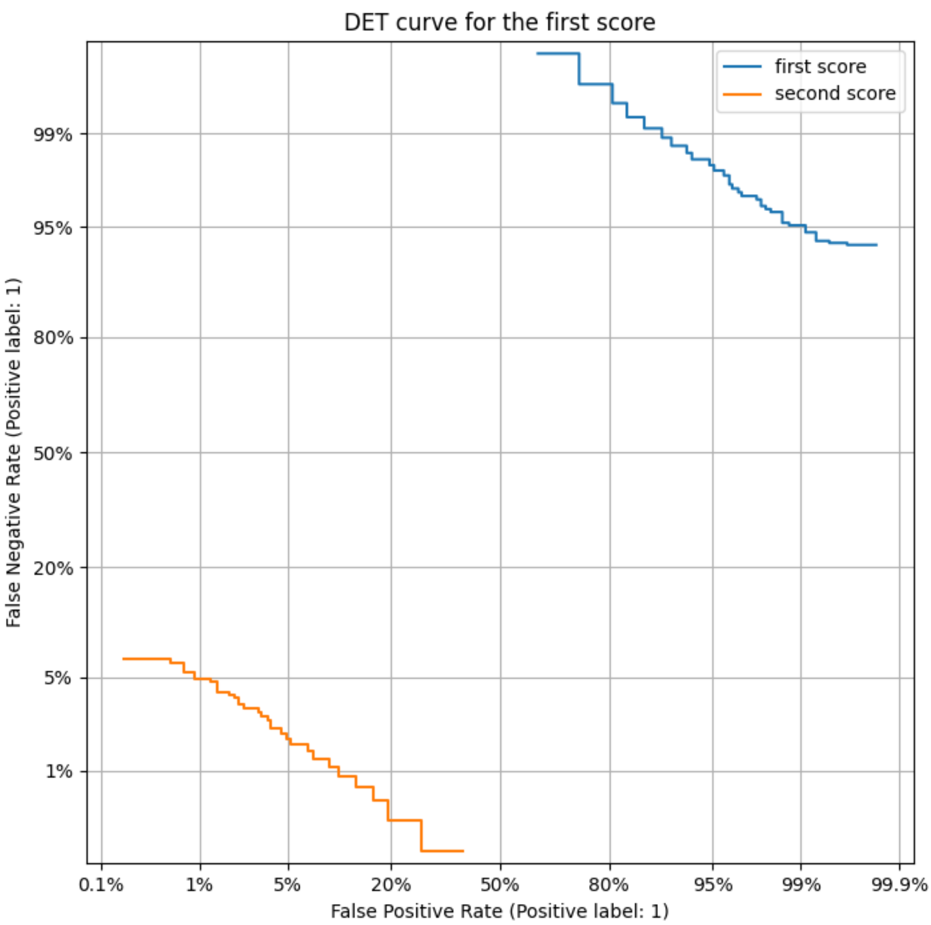
The highest validation accuracy was in epoch 2 and it is 96.9%.

The test accuracy corresponding to that epoch is 96.7%

Q28:



ROC curves AuC Score for the first score: 0.004, AuC second score: 0.996
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
first score
second score
0.0
0.2
0.4
0.6
0.8
1.0

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

Chapter 5:

Q29:

Image-Specific class saliency Visualization is a method to interpret a neural network's decision. It calculates the gradients of the score for the correct class with respect to the input image pixels. By visualizing these gradients, we can identify which pixels had the most influence on classification, essentially showing which parts of the image the model focused on

Q30:

Grad-CAM is a technique used to visualize the regions of the input image that were important for the model's prediction. It works by using the gradients of the target class flowing into the final convolutional layer. This produces a coarse localization map (a heatmap) that highlights the specific areas in the image that the model focused on to make its decision.

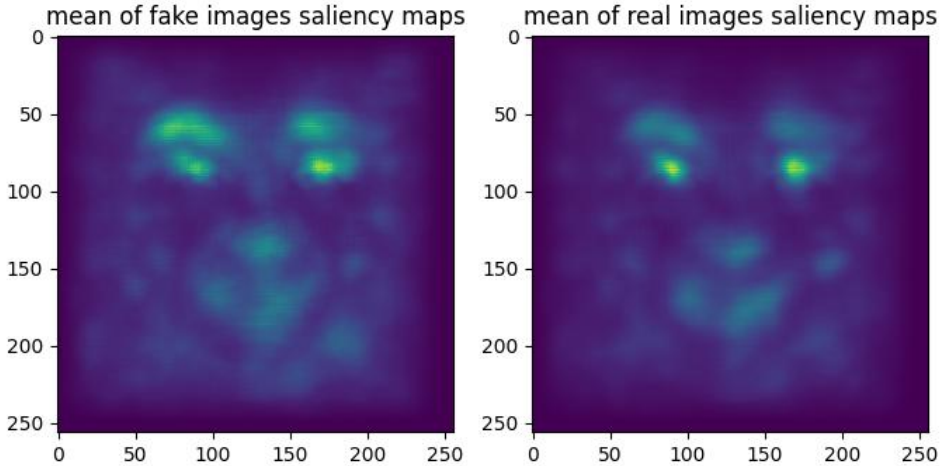
Q31:

Done.

Q32:

SimpleNet on Deepfakes:

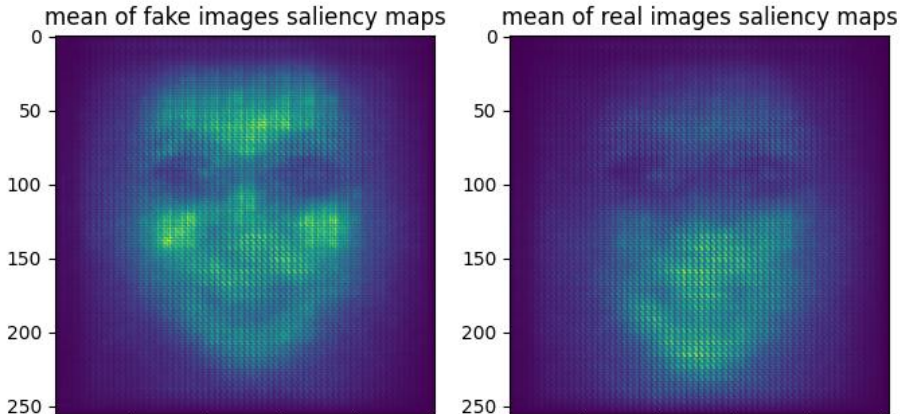
The saliency maps mainly highlight the face area, especially the eyes/eyebrows, while the background is mostly inactive. The single-image maps look noisy. This suggests the model relies on local facial texture or artifact cues




Images and their saliency maps

XceptionBased on Synthetic Faces:

The saliency is more structured and consistently centered on the face. In the mean maps, fake images emphasize the upper face more, while real images show relatively stronger focus on the lower face. This indicates the model learned class-dependent facial cues, not background shortcuts, which matches its strong classification performance.




Images and their saliency maps

Q33:

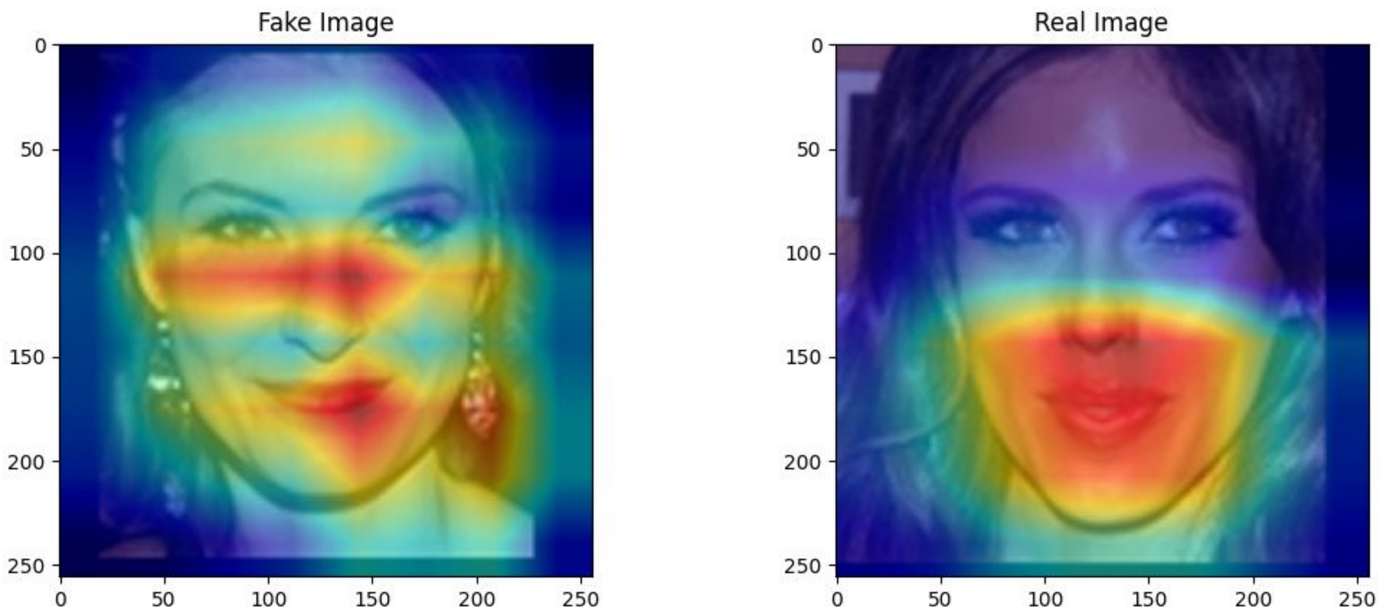
Done.

Q34:

Done.

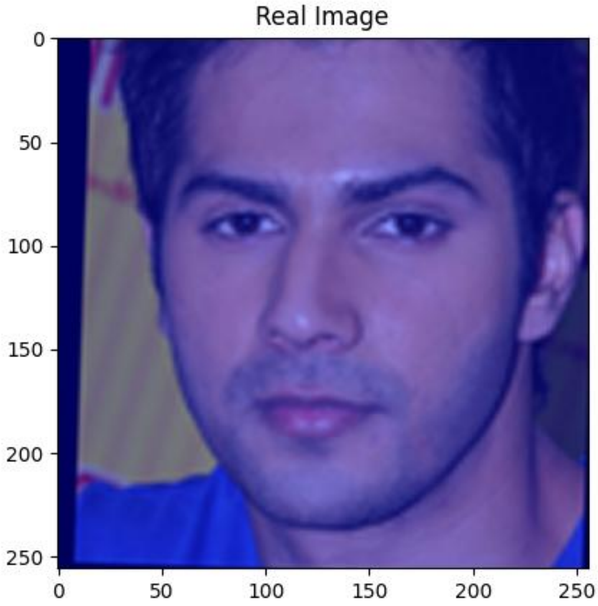
Q35:

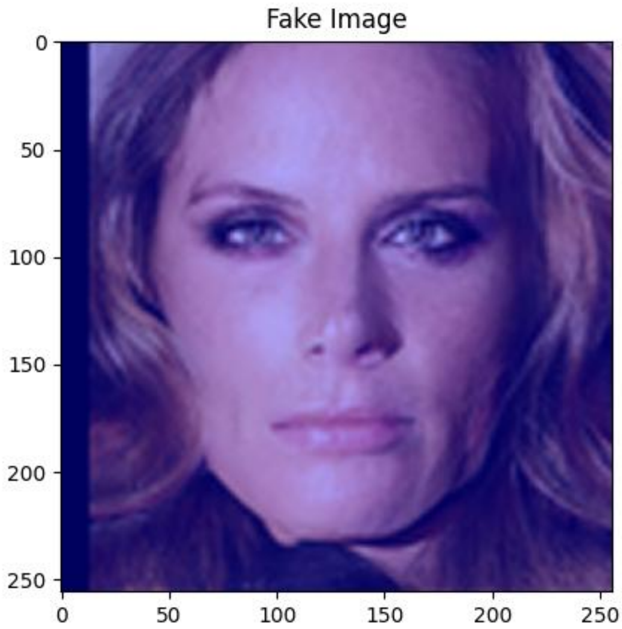
XceptionBased on synthetic dataset:




The heatmaps are concentrated on the face region. This suggests the model is using facial features rather than the background

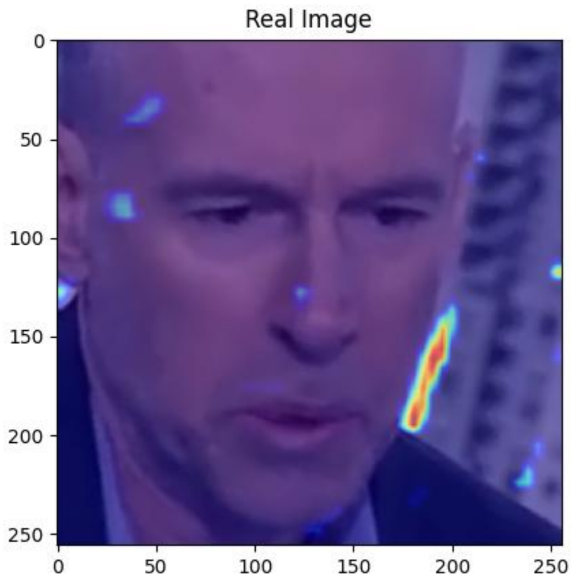
SimpleNet on synthetic dataset:

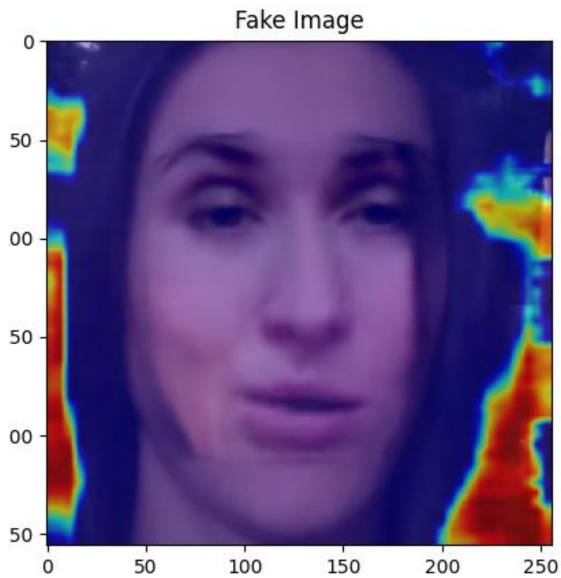



The Grad-CAM overlays are uniform (no clear focused region on the face). This matches the earlier results where SimpleNet was close to chance and didn't learn features.

SimpleNet on fakes dataset:




The heatmaps highlight image borders or background rather than stable facial regions. This suggests SimpleNet may be relying partly on artifacts correlated with the image manipulating and not facial evidence.