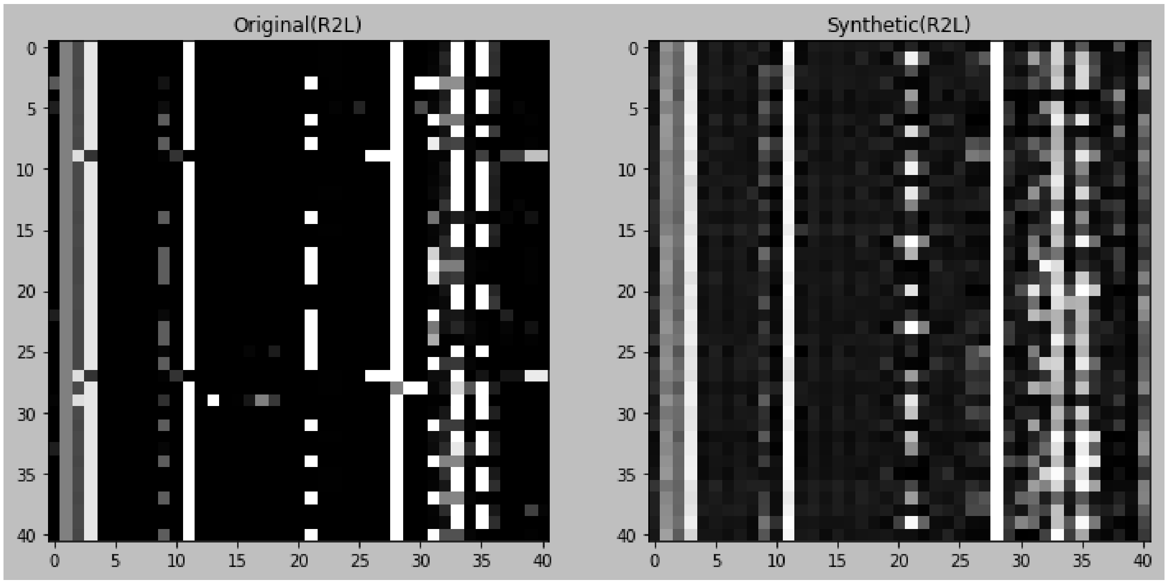
Week 6 **NIDS** **Improve Data Synthesis**

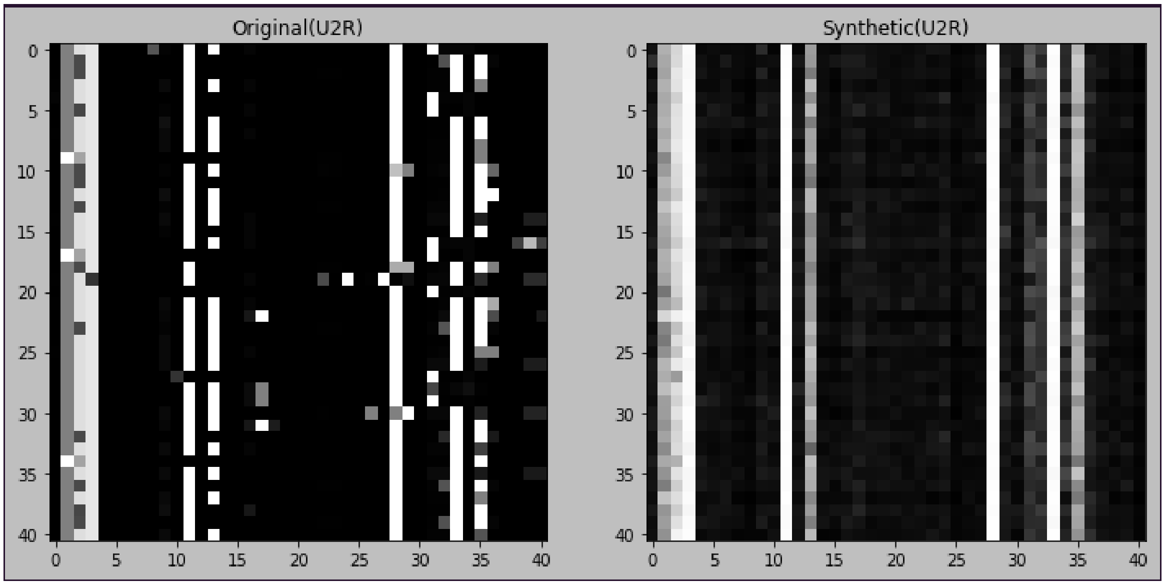
Task:
Improve GAN and incorporate Gradient Penalty to generate samples for minority attack classes
Output:
Visualized 41 instances from Original and Synthesized data
X-axis: Features (41)
Y-axis: Instances (41)

Following are the traffic classes along with percentage in the original Dataset

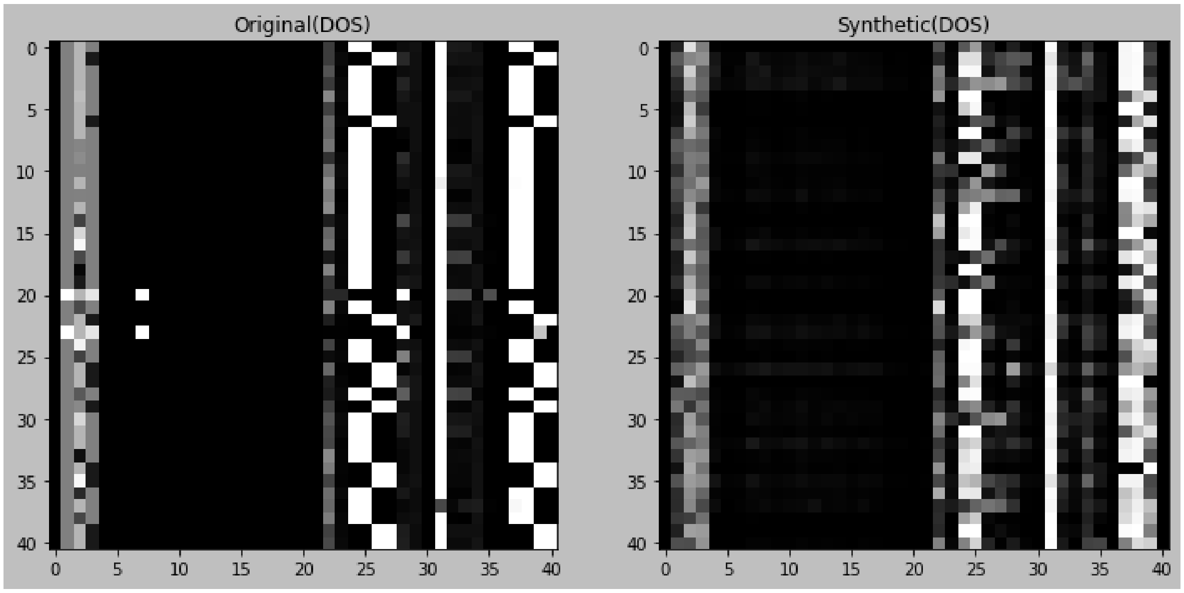
R2L Attacks (0.79%)



U2R Attacks (0.04%)



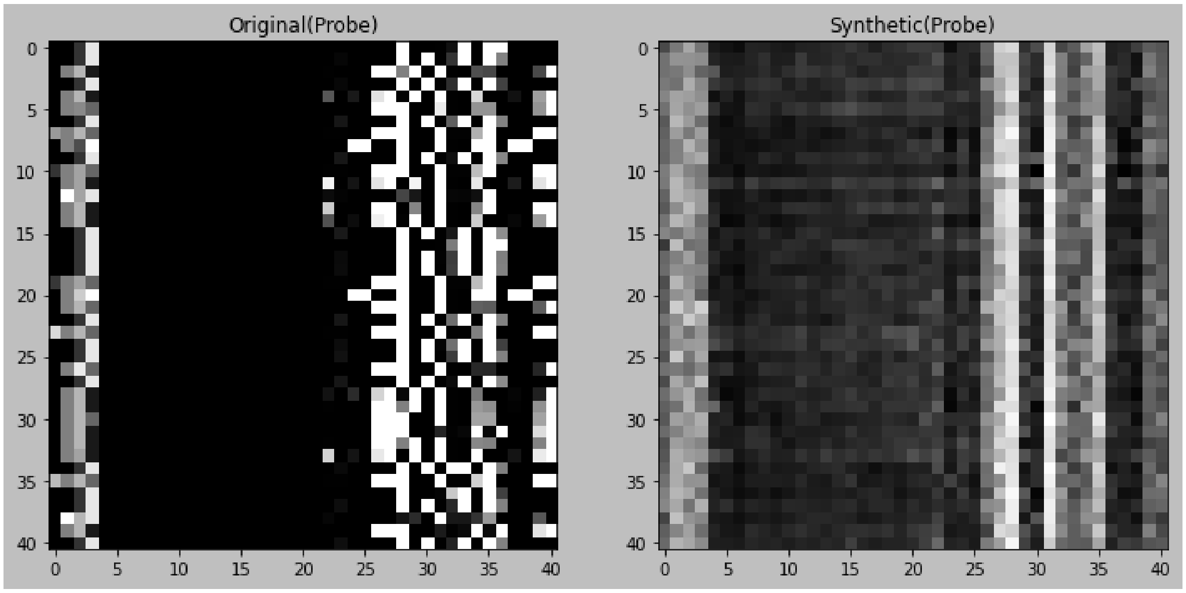
DOS Attacks (36.45%)



Original(DOS)

Synthetic(DOS)

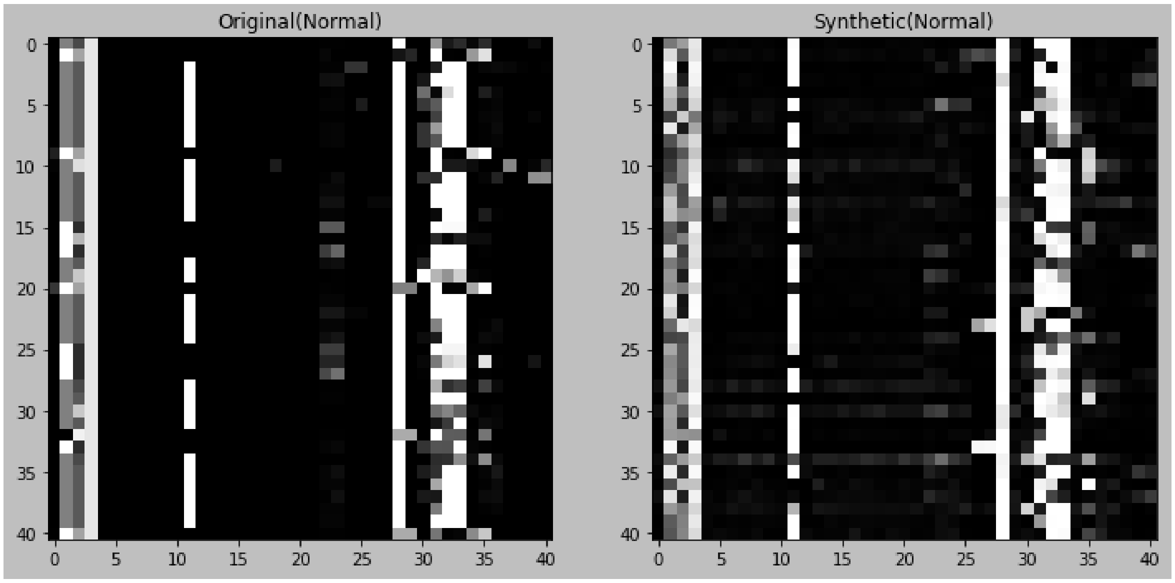
Probe Attacks (9.25%)



Original(Probe)

Synthetic(Probe)

Normal Traffic (53.46%)



Conclusion:

1. Used DCGAN with Wasserstein Loss and Gradient Penalty to learn the distributions.
2. This technique was able to catch some of the patterns as observed from the visualizations.
3. Added a few thousand R2L attacks to the original dataset and used a deep CNN for the detection task.
4. Adding a few thousand R2L attacks to the original dataset slightly improved the accuracy.
5. The improvements in generators seemed to have a negligible effect on the outputs.

Next:

1. Generate data for each class separately and fix an MLP to check results.
2. Balance the dataset.
3. Evaluate a deep learning model with stratified sampling technique.