Required Package: PyTorch, argparse, numpy, librosa, opencv-python, tqdm, scikit-learn

1. The test dataset is the GTZAN dataset which we collected online. The main difference from training and validation is that they are different samples. We divided the dataset into 8:1:1 so there would be no overlap between the three datasets. I think it is sufficient to test the generalization capabilities of your final programs since the split proportion is common in machine learning tasks.
2. We tried four different neural networks: CNN, ResNet50, CNN+lr decay and CNN+lr decay+residual blocks. The test accuracy of them are 0.82, 0.62, 0.86, 0.88.
3. The main reason that testing performance is worse is overfitting. From training, our accuracy is very high, however the validation and testing could not reach the same accuracy.

Contribution:

Yunhao:

1. Helped designing CNN model;
2. Built ResNet50;
3. Built the script pipeline and trained both models on a GPU machine;
4. Helped write the report;
5. Built the CNN+residual block network.

Yiming (John):

1. Dataset collection and cleaning;
2. Dataset preprocessing and saving;
3. Helped designing CNN model;
4. Helped write the report;
5. Built the CNN+decay model;