Name: Jason Zhou
Mentor: Dr. Dongjin Song
Status Report #: 9
Time Spent on Research This Week: 14.25
Cumulative Time Spent on Research: 60
Miles Traveled to/from Mentor This Week: 0
Cumulative Miles Traveled to/from Mentor: 0
==

Monday, November 8th, 2021: (0.25 Hours)

To begin the week, I had a meeting with my mentor where we discussed the upcoming deadline for my research proposal. Overall, we went over the general layout of what my paper would look like.

Wednesday, November 10th, 2021: (3 Hours)

On this day, I began going back through the articles I had read to try and determine what model I would like to create or build off of. In particular, I focused on a model called Group Masked Autoencoder for Density Estimation (GMADE) and WaveNet because they were scored at the top of the DCASE 2020 Task 2 challenge. Unfortunately, I did not understand much from reading this article and decided to read it again at a later time.

In the next hour and a half, I began writing the introduction of my research proposal. Specifically, I just began writing down everything that I knew about my area of research based on what I had read previously. Because I was just trying to brainstorm, I did not include any citations.

Thursday, November 11th, 2021: (4 Hours)

On this day, I continued reading through the GMADE and WaveNet articles. During step one of my research proposal, I quickly went through these technical reports and, thus, did not develop a good understanding of these articles. As such, I focused my time on trying to form a deep understanding of what was being written.

In general, GMADE is an autoencoder that uses masks to logically erase connections within the network. This is done to give the autoencoder an autoregressive property, which is simply when previous data is used to predict future data. Apparently, when making an autoencoder, autoregressive, allowing it to output a probability distribution. This is useful because it allows the autoencoder to calculate the probability that something has of being one thing versus the other.

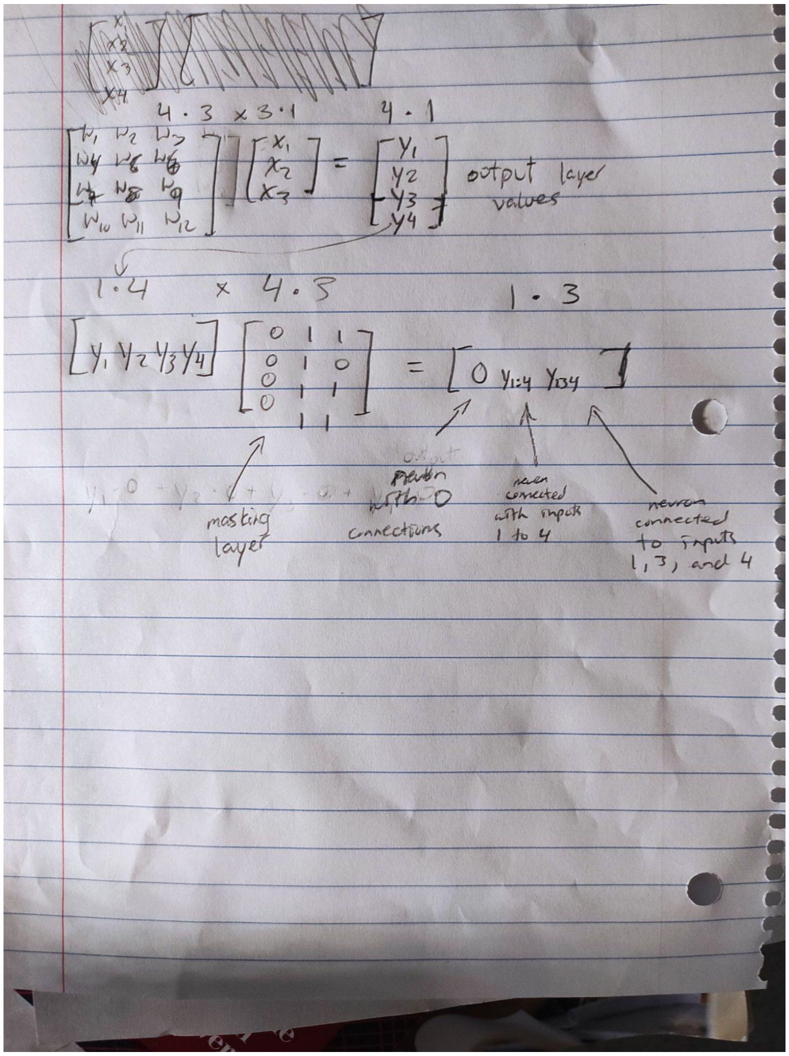
As for WaveNet, this is essentially just a model that generates new data. With respect to audio detection, it looks as though it was used to create new data and compare it with the actual observed data. For example, if a 10-second clip is split into two 5 second segments, the first segment will be inputted into WaveNet to generate and predict what the audio will look like in the next 5 seconds. This predicted data is then compared to the second segment of audio. If they are similar, it is likely normal, otherwise, it is irregular.

Beyond looking at research articles, I also began writing the proposed methods section for the research proposal, which mainly consisted of training an ensemble of models and then evaluating the ensemble to see how well it performs. In the future, I will probably have to add more details like the architecture and the hyperparameters of the ensemble; however, my mentor told me to treat the inner workings of machine learning algorithms as a black box because I do not have to fully understand them yet. Also, I suspect that this information will only be known after the experiment has been performed and values and constraints for the model have been optimized.

Friday, November 12th, 2021: (3 Hours)

Although I now had a general idea of how GMADE worked, I still was not sure how the connections were masked within the model using a matrix. This led me to learn how neural networks are represented in matrices and masks interact with these matrices to block certain connections.

Despite taking me a while to grasp, the concept is fairly simple. For every two layers within a model, there must exist at least a couple of connections to pass information through the network. Between these layers is where the network can be represented as a matrix. For example, each connection has a weight. All of these weights are then placed within a matrix (Matrix A). Then, all of the values of each input unit are used to create a second matrix (Matrix B). Matrix A and Matrix B are then multiplied and the resulting product (Matrix C) represents the values of each output unit in the output layer. To mask these connections, one only needs to multiply the Matrix C by the masking matrix which contains a collection of zeros and ones. Multiplying by zero will erase a connection while multiplying by one will retain a connection.



(A drawn example of the math that I am explaining in the above paragraph. The matrix with the x variables represents the input layer. Each x acts as an input unit. Multiplying weights by the values of the input layers gives the values for the output layers, which is illustrated by the matrix of y variables. Multiplying the y variable matrix by a masking matrix results in certain connections being retained. The final resulting matrix has 3 values. Each value represents which output nodes the corresponding input node is connected to. For example, the middle value says y with a subscript of "1:4," which means that the middle input node is connected to all four output nodes,)

Sunday, November 14th, 2021: (4 Hours)

On this day, I put some final touches on the rough draft for my research proposal. I mostly worked on rearranging certain paragraphs and rewording sentences to make my proposal more cohesive and understandable. I also went through all of my articles and cited everything that I had written and added a references list that had all the articles I cited. I also brainstormed a couple of working titles and took the one which I believed was the best.

References

Naranjo-Alcazar, J., Perez-Castanos, S., Zuccarello, P., & Cobos, M. TASK 2 DCASE 2020: ANOMALOUS SOUND DETECTION USING UNSUPERVISED AND SEMI-SUPERVISED AUTOENCODERS AND GAMMTONE AUDIO REPRESENTATION.

Primus, P. (2020). Reframing unsupervised machine condition monitoring as a supervised classification task with outlier-exposed classifiers. *Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE 2020 Challenge), Tech. Rep*.

Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., ... & Harada, N. (2020). Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring. *arXiv preprint arXiv:2006.05822*.

Ahmed, F., Nguyen, P., & Courville, A. (2020). An ensemble approach for detecting machine failure from sound.

Wang, Y., Zheng, Y., Zhang, Y., & He, L. (2021). Several approaches for anomaly detection from sound.

Sakamoto, Y., & Miyamoto, N. (2020). Anomaly calculation for each components of sound data and its integration for DCASE 2020 challenge task2.

Hayashi, T., Yoshimura, T., & Adachi, Y. (2020). Conformer-based id-aware autoencoder for unsupervised anomalous sound detection.

Hoang, T. V., Nguyen, H. C., & Pham, G. N. (2020). Unsupervised Detection of Anomalous Sound for Machine Condition Monitoring Using Different Auto-Encoder Methods.

Giri, R., Tenneti, S. V., Helwani, K., Cheng, F., Isik, U., & Krishnaswamy, A. (2020). Unsupervised anomalous sound detection using self-supervised classification and group masked autoencoder for density estimation. *Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE 2020 Challenge), Tech. Rep*.

Germain, M., Gregor, K., Murray, I., & Larochelle, H. (2015, June). Made: Masked autoencoder for distribution estimation. In *International Conference on Machine Learning* (pp. 881-889). PMLR.

Daniluk, P., Gozdziewski, M., Kapka, S., & Kosmider, M. (2020). Ensemble of auto-encoder based and wavenet like systems for unsupervised anomaly detection. *Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE 2020 Challenge), Tech. Rep*.

CS-Swift. (2019, January 22). *How to represent a neural network with matrices* [Video]. YouTube. https://www.youtube.com/watch?v=lFOOjeH2wsY

Sachdeva, K. (2021, January 22). *MADE: Masked autoencoder for distribution estimation* [Video]. YouTube. https://www.youtube.com/watch?v=7q4ueFiJjAY