

Milestone 2 Report

The main goal of milestone 2 is to construct models and train them with datasets created from milestone 1. We have split the team into two groups separately, with one responsible for the mood classification of music and the other for “emotion” detection of the faces.

Current State for Team "Emotion"

To classify emotions from human faces, we adapted and modified an existing model called Xception [1] instead of the Resnet model suggested in the proposal. This model is inspired by the modern CNN architectures that support the use of depth-wise separable convolutions. In the original model, it uses depth-wise separable convolutions to reduce the computational complexity in each convolutional layer allowing the network to run faster in CPU operation. However, we changed these convolutions to the normal 2d convolutions to achieve faster computation, run by the GPU. We also dropped the middle flow out of the original model and maintained the entry flow and the exit flow. The model consists of a base module and comes with 4 depth-wise convolutions. Every depth-wise convolution is coupled by two convolutions and a max-pooling function, and each convolution is followed by a batch normalization operation and a ReLU activation function. The final layer uses a global average pooling and a soft-max activation function to make predictions. After completing the model, we fed the parameters (image shape and the number of output classes=7: ‘angry’, ‘disgust’, ‘fear’, ‘happy’, ‘sad’, ‘surprise’, ‘neutral’) into the model and compiled it with Adam as our optimizer function, and Categorical_crossentropy as our loss function. We used ImageDataGenerator from Keras library that takes in the datasets from milestone 1 to generate batches of the augmented data. We used the fit function on the model to regulate the parameters like batch size, epochs, steps per epoch, learning rate, etc. during the training process. The two challenges we face were finding the appropriate number of epochs to get the best results and choosing the right batch size for training the model. After several attempts to run the model with different adjustments to the parameters, we came close to a validation accuracy of 68%.

Current State for Team "Music"

For the “music” portion of our second milestone, everything went relatively smoothly as we were able to create the model with minimal trouble; additionally, we followed the guidelines set by the second milestone in our project proposal. The only major obstacle we faced was that when we trained the dataset we previously created, we had difficulty surpassing 50% accuracy, therefore, we had to abandon our dataset and use the original author’s dataset. After this, our model was able to obtain an accuracy of approximately 80%. As for the implementation of our model, first, we imported several necessary libraries from keras to help create our multi-class neural network including tensorflow. Then we created our base model. Our model consisted of a layer of 8 nodes with the relu function and another layer with output 3 and the softmax function. Next we compiled the mode using the logistic loss function and adam optimizer and additionally

configured the estimator with 300 epochs and 200 batches. Finally, to evaluate the model, we decided to use Kfold cross validation. After our model was trained and tested, we achieved accuracies of around 80% and then we proceeded to generate song datasets classified by mood. We randomly searched 50 songs and extracted the required information. Our final dataset includes the name of each song, its audio URL, as well as its cover art URL. For the next milestone, we have to figure out a way to combine the music portion and the emotion portion so that the algorithm can output appropriate songs based on the user's current mood.

Conclusion

So far, both teams have completed training their models. There are no feature changes as we are only training the models, everything goes as planned.

Work Distribution

Yisheng: Imported libraries, researched for the related model, modified the existing model, trained, and tested the model.

David: Wrote the report of team "emotion" and assembled the two team's reports into a final milestone 2 report, fed parameters to the model, prepared and compiled the model, loaded the model.

Jack: Wrote the "music" portion of the second milestone report.

Jacob: Implemented the "music" portion of the model for testing and training and compiled dataset.

Reference

[1] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf. [Accessed: 11-Apr-2021].