Synthetic Hand Gesture Dataset Generation

Undergraduate Research Scholars
LAUNCH: UNDERGRADUATE RESEARCH

Samuel Oncken and Steven Claypool

Background

Today, there is no longer an argument of whether artificial intelligence and machine learning are here to stay. However, the reliability and robustness of machine learning algorithms in the field of object classification rely heavily on the model and data used. Through our research, we have discovered numerous methods by which large, diverse datasets capable of training object classification neural networks are created. Most require a large pool of human participants to perform gestures in front of some type of camera/sensor equipment. Not everyone has access to these resources though. This research seeks to uncover the viability of using synthetic data in the training of these neural networks. Using the data created in a virtual environment, a variety of object classification and localization neural network models will be trained and tested on both real and virtual data to analyze the recognition accuracy under different circumstances. This analysis will act as a proof of concept for the usage of virtual data in machine learning applications such as satellite imagery simulations and smart robotics to name a few.

Research Questions

Problems

- Time, cost, and effort required to build sizable and diverse datasets is significant
- Number of datasets for unknown or uncommon gestures is lacking
- The need for many diverse participants brings potential legal challenges

Proposition

Create a virtual environment for the generation of synthetic hand gesture datasets that can augment or replace real datasets for computer vision applications.

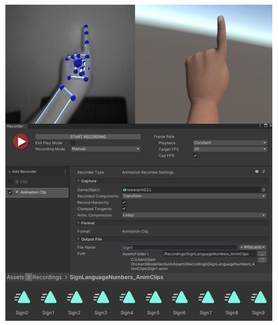
Methods/Materials

Virtual Environment

- Created using Unity Game Engine
- MakeHuman models generated as virtual participants
- Separated into 4 stages: Gesture Recording, Human Model Generation, Model Animation, and Data Capture/Collection
- All code is written in Unity C#
- Gestures are recorded using Leap Motion Controller (LMC) and hand motion is bound to MakeHuman model using Leap Hand Binder script

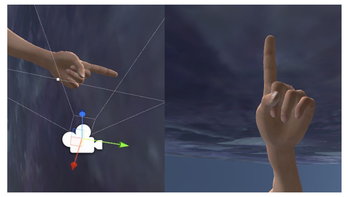
Dataset Training/Testing

- Used Tensorflow with Keras API for various different CNNs
- Trained CNNs with and without augmentation preprocessing
- Trained CNNs using transfer learning and from scratch



Depiction of gesture recording process including LMC Control Panel view, Unity game view (with rigged MakeHuman model), Unity Recorder settings, and final animation clip outputs.

Spawning of MakeHuman model and application of hand gesture from a selected cluster of gesture animations in Animator Controller

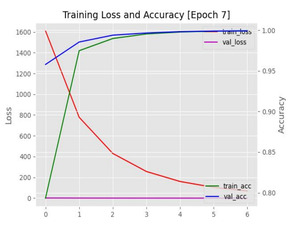


Placement/enabling of correct camera corresponding to performed gesture animation with game view of camera output

System controller within Unity with choice of dataset and images/gesture. Output images are organized and stored in dataset and gesture specific folders.

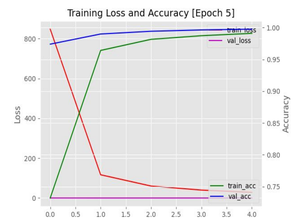
Virtual Environment Visuals

Results from training and testing "Inception" image classification model using transfer learning on synthetic ASL alphabet dataset composed of 336,000 images (12,000 images per 28 gestures). Achieved just under 90% validation accuracy in 18 epochs, indicating a well trained model.

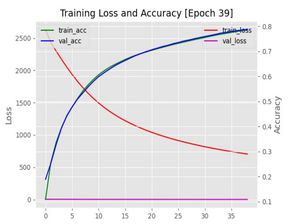


Results from training and testing "VGG19" image classification model using transfer learning on synthetic ASL alphabet dataset composed of 336,000 images (12,000 images per 28 gestures) with no augmentation. Achieved near 100% accuracy in 7 epochs, indicating overfitting. Shows the need to test on various models, as the same dataset on the "Inception" model (pictured left) gives far better results.

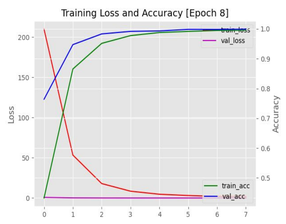
Results from training and testing "Xception" image classification model using transfer learning on synthetic ASL numbers (0-9) dataset composed of 120,000 images (12,000 images per 10 gestures). Achieved just below 80% validation accuracy after 16 epochs, suggesting a lack of data for the size of the model.



Results from training and testing "Inception" image classification model using transfer learning on real ASL alphabet dataset composed of 84,000 images (3,000 images per 28 gestures). Achieved near 100% accuracy in 5 epochs, indicating overfitting. Likely due to smaller dataset size, as opposed to the synthetic ASL alphabet dataset that was easily expanded to 336,000 images.



Results from training and testing "ResNet50" image classification model using transfer learning on real ASL alphabet dataset composed of 84,000 images (3,000 images per 28 gestures). Achieved below 80% accuracy after 39 epochs, suggesting a lack of data for the size of the model. Such datasets could be supplemented with synthetic data created through our virtual environment.



Results from training and testing "Xception" image classification model from scratch on real ASL numbers (0-9) dataset composed of 15,000 images (1,500 images per 10 gestures). Achieved near 100% accuracy after 8 epochs, indicating overfitting. Training from scratch modifies the convolution layers to fit the small amount of training data too closely, causing overfitting when testing.

Training/Testing Results and Observations

Outcome(s)

1. Dataset generation virtual environment is completed and can be utilized by any user to create a gesture dataset of his/her own
2. Final synthetic datasets successfully trained hand gesture image classification models
   - Models trained and tested on synthetic data show similar accuracy metrics when compared to training and testing on real data

Discussion/Outlook

- Preliminary results with training and testing models solely with synthetic data indicate that synthetic data can be used in computer vision for synthetic image classification purposes.
- Further analysis with classifying real images using models trained on synthetic data is in progress.
- Additional object detection/localization analysis is in progress and should further strengthen our claims.

Acknowledgements

Special thanks to our sponsor Stavros Kalafatis and graduate student Pranav Dhulipala for their support and guidance throughout our research.

References