# Convolutional Neural Networks: Selfie Facial Recognition and Real-World Image Generalization

This study examines the process of adapting a Convolutional Neural Network (CNN) trained on selfies to recognize individuals in real-world, non-selfie images. The study's primary goal is to explore how well a CNN, trained on easily generated user-generated data like selfies, performs on standard, generalized image recognition tasks. It also determines whether training on a more diverse set of non-selfie images is required for effective model generalization.

The data collection process was straightforward: gather 500 selfies from various individuals – Issy, Oliver, Ryan, Sam, and Tom. Data cleaning, storing, and structuring were more complicated tasks due to varying image formats from each individual.

Images were aggregated to separate folders within Google Drive, creating an appropriate class structure for CNN integration. Initial models were trained using an 80% train 20% validation split without knowledge that many images existed in .heic format, which is unreadable by Tensorflow and Keras. Both training and validation data *only* included selfie-images – not real-world, non-selfie images.

The initial model was a standard VGG16 CNN, using pre-trained image-net weights, and a custom classifier consisting of a single dense layer and a dropout layer (dropout rate: 50%) to aid in prediction generalization. The lack of training images for certain classes (due to the existence of unreadable .heic files) resulted in very low accuracy. A confusion matrix analysis was also conducted, which showed that there was significant class imbalance.

These analyses informed a solution to overall poor model performance: mass converting of all .heic images to .jpg or .png format and re-uploading them to Google Drive in a separate folder. After this process, the training accuracy of the initial VGG16 model peaked around 98% (with a validation accuracy of 100%). The validation accuracy's increase over training accuracy is likely due to the small size of the validation set (100) when compared to the training set (400). An additional, but important note: upon the first iteration of the model, the number of images per class were as follows:

- Issy: 113
- Oliver: 141
- Ryan: 100
- Sam: 104
- Tom: 86

As the initial VGG16 model was trained exclusively using selfie images, the next step of the project was to test the model's capability to generalize on real-world images of the individuals within the study. When tested on 50 real-world images, the VGG16 model achieved a maximum accuracy of 50%, which is 30% higher than random chance (20%). After a confusion matrix analysis, clear class imbalances still existed, in which the model predicted Issy, Oliver, and Sam significantly more than it predicted Tom and Ryan, as seen below:

- Issy: 19 predictions. 47% accuracy.
- Oliver: 20 predictions. 40% accuracy.
- Ryan: 0 predictions. 0% accuracy.
- Sam: 10 predictions. 80% accuracy.
- Tom: 1 prediction: 0% accuracy.

As the classes producing the most predictions directly correlated with the classes that had the most images, the next step was obvious: data augmentation to reduce class imbalance. A new directory structure in Google Drive was created containing exactly 100 images per person. New data generators were employed, this time including a brightness modifier and increased augmentation rates. Retraining the model on original data produced initially promising results: 85% accuracy and 100% validation accuracy on selfie-data (again, note that high validation accuracy is likely due to the relatively small validation set size). Although, when generalizing to real-world, non-selfie images, the model only produced 40% accuracy, a reduction from the previous iteration. The reduction in both selfie-image accuracy and non-selfie-image accuracy was likely correlated with overall reduction of data that was necessary to have even classes.

Considering that data-cleaning and data-augmentation were both sufficiently tested by this point, the weights of the pre-trained model were considered next. During this stage of model tuning, two options were considered: 1.) unfreezing some layers of the VGG16 model to reduce overfitting to original image-net weights, and 2.) using learning rate scheduling to increase the effectiveness of image-net weights.

Firstly, the last 10 layers of the VGG16 model were unfrozen, and the model was re-compiled and refit on the original data. This process produced the worst results yet: 28% training accuracy (only 8% better than random chance), and was abandoned before generalizing to non-selfie images.

Secondly, the layers of the model were reset, and the model was re-trained using a learning rate scheduler in effort to make image-net weights more effective. This process concluded with 28% accuracy on selfie-data, and was also abandoned.

As all efforts in hyper-parameter tuning produced significantly worse results than efforts in data-tuning, a final data-tuning measure was considered: modifying selfie data to only include individuals' faces. A computer vision library, CV2, was employed to mass-modify all selfie-images, cropping each to a new Google Drive directory. The top performing model was then re-compiled and re-fitted with the cropped data. Initial metrics on selfie-data seemed promising: 94% accuracy, 96% validation accuracy. Although, when generalizing to non-selfie data, the model produced 30% accuracy and was abandoned. Interestingly, retraining the model on cropped images shifted the class imbalance significantly.

In conclusion, the base VGG16 model with a simple custom classifier (one dense layer, one dropout layer of 50%) consistently outperformed all other variations. While data cleanup and augmentation were necessary to improve performance, all attempts to modify the model architecture or fine-tune hyperparameters failed to surpass the initial model's 50% accuracy on real-world images.

Key findings from model iterations are re-summarized below:

- The initial VGG16 model achieved 100% validation accuracy on selfie images, and generalized to real world images with 50% accuracy (which is 30% better than random chance).
- Increased data augmentation and the elimination of class imbalance unexpectedly hurt model performance, reducing generalization accuracy to 40%.
- Unfreezing 10 VGG16 layers resulted in a significant performance drop (28% training accuracy).
- Adjusting learning rate using a scheduled learning rate similarly provided no-model improvement.
- Training on only cropped faces and removing background noise reduced model generalization accuracy to 30%.
- Data-tuning provided significantly better results than model-tuning, suggesting that data quality is one of the most important measures when using pre-trained models.
- This project was limited by data availability; if time and resources allowed, next steps would involve training different model architectures on larger, more diverse sets of data.

The project served as a compelling thought experiment on training Convolutional Neural Networks using small, organically collected datasets, and demonstrated that CNNs trained on controlled datasets can still achieve impressive generalization to real-world data.