Politician FaceNet

Jordan Croom

Abstract—In this paper, a convolutional neural network is presented to identify various politicians. Using the DIGITS interface, the GoogLeNet architecture is trained using data gathered from a custom web crawler with minimal manual preprocessing. The trained network achieves a high level of top 1 and top 5 accuracy for the trained classes. Further evaluations of the network against artistic renditions of the likenesses of selected faces reveal moderately generalized abstract features were successfully modeled.

Index Terms—Robot, IEEEtran, Udacity, LaTeX, deep learning.

1 Introduction

THE problem of facial recognition is a well-studied problem with numerous real-world applications. Social networks such as Facebook, YouTube, etc. rely heavily on automated facial recognition to identify users in photos and videos, flag offensive or illegal content, or to make users look like puppies [1], [2]. Domestic and international police agencies use facial recognition software to track and apprehend criminal suspects. Politician FaceNet attempts to address a similar but more limited problem. Given a cropped image of a politician's face, identify him or her by name. By limiting the application to a simple classifier as opposed to a detector/classifier, some complexity is avoided while still addressing the core issues of the problem.

1.1 The DIGITS Platform

The NVIDIA DIGITS platform greatly decreases the complexity of organizing training data and training a neural network while still providing substantial control and monitoring of the training process [3]. Using a simple web interface, raw data sets can be uploaded, standardized, and analyzed, and networks can easily be trained and monitored using a range of optimization functions and control parameters. DIGITS was leveraged in the training of Politician FaceNet in order to reduce data preparation and organization overhead while providing insights into the training process and final performance.

2 Background / Formulation

Politician FaceNet is based on GoogLeNet, a well-known convolutional neural network architecture [4]. GoogLeNet was chosen based on its solid performance and modest number of operations per forward pass [5]. As can be seen in figure 2, GoogLeNet offers a suprisingly high accuracy and low evaluation time given its relative accuracy on image classification tasks. A primary motivation for this study–beyond achieving reasonable facial recognition accuracy from few samples–is to demonstrate that reasonably good performance can be achieved using limited computing power in a short period of time with effectively "off the shelf" machine learning solutions. GoogLeNet especially exemplifies simplicity of deployment since it is built into recent releases of Tensorflow/Keras and is available as a well-supported open-source Caffe model.

While other architectures could certainly outperform GoogLeNet in terms of overall accuracy, the incremental difference would likely be negligible for a small training set. Given the

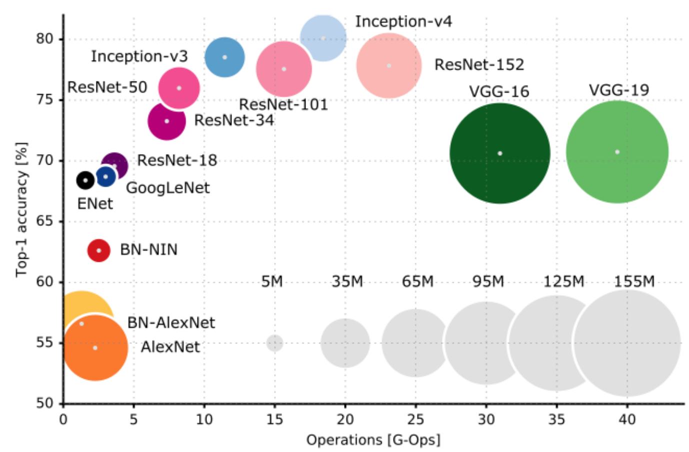

Fig. 1: Comparison of Hyperparameters, Accuracy, and Evaluation Time for Various Popular Imagenet DNNs Taken from [5]
.

limited number of classes to be trained (14), and the relatively limited range of feature types, textures, colors, etc, GoogLeNet was anticipated to provide sufficient classification accuracy for the problem at hand and a relatively short training time.

3 Data Acquisition

The dataset was gathered by crawling a search engine using the Selenium library and a custom python script [6]. Keywords for each politican were selected, and a set of approximately 300 images of each class were automatically downloaded (see full set here). The gathered data was then preprocessed manually by cropping the images to be square (a requirement of the GoogLeNet network as used) and more or less centered on the face of the subject using a custom image preprocessing script [cite crawler]. Final training images were:

- Full-color RGB
- Square aspect ratio
- Primarily centered on the subject's face

The network was trained with the following parameters:

- Training Epochs: 75
- Batches: 1
- Solver: SGD
- Base Learning Rate: 0.002

Fig. 2: Example Images of Hillary Clinton and Barack Obama Classes

TABLE 1: Training Set Contents by Class

Class	Total	Training (70%)	Validation (22%)	Test (8%)
Al Franken	204	142	45	16
Barack Obama	327	228	72	26
Ben Carson	260	181	57	21
Bernie Sanders	298	207	65	24
Chuck Schumer	255	178	56	20
Donald Trump	282	197	62	22
Elizabeth Warren	277	193	61	21
Hillary Clinton	337	235	74	27
John McCain	260	181	57	21
Justin Trudeau	297	206	65	23
Marco Rubio	262	182	57	21
Michelle Obama	313	218	68	25
Paul Ryan	288	201	63	23
Ted Cruz	298	207	65	23

- Learning Rate Policy: Sigmoid at 33% of iterations
- Gamma: 0.06

4 Results

Following training, the model achieved a weighted top 1 accuracy of 91%. Given the small number of training samples and limited breadth of the classes, this result is reasonably good. As can be seen in the confusion matrix, the model appears to have made fairly obvious distinctions between various faces based on skin tone, hair color/style, gender, and age. For example,

- 5% of Ben Carson test images were classified as Barack Obama.
- Hillary Clinton and Elizabeth Warren were mistaken in nearly 10% of test cases

Other results are less clear:

- Ted Cruz was mistaken for another politician or vice versa very often, even for politicians of different races or genders
- Al Franken was never mistaken for any other politician or vice versa

It is difficult to say if these results are indicative of some underlying properties of the faces of the politicians in question or are a peculiarity of the model as trained.

This is typically the hardest part of the report for many. You want to convey your results in an unbiased fashion. If you results are good, you can objectively note this. Similarly, you may do this if they are bad as well. You do not want to justify your results here with discussion; this is a topic for the next session. Present the results of your robotics project model and the model you used for the supplied data with the appropriate accuracy and inference time For demonstrating your results, it is incredibly useful to have some charts, tables, and/or graphs for the reader to review. This makes ingesting the information quicker and easier.

5 Classifying Abstract Representations

Various artistic representations of various classes were classified in order to further probe the limitations of the neural network (see full set here. Results revealed fairly successful generalization of the real image training data in response to stylized, exaggerated, or impressionistic representations of the tested classes. The tendency

TABLE 2: Normalized Confusion Matrix for Test Image Set

	Al Franken	Barack Obama	Ben Carson	Bernie Sanders	Chuck Schumer	Donald Trump	Elizabeth Warren	Hillary Clinton	John McCain	Justin Trudeau	Marco Rubio	Michelle Obama	Paul Ryan	Ted Cruz	Total Images	Per Class Accuracy
Al Franken	1.00														16	1.00
Barack Obama		0.96												0.04	26	0.96
Ben Carson		0.05	0.95												21	0.95
Bernie Sanders	0.04			0.92								0.04			24	0.92
Chuck Schumer	0.05				0.89				0.05						19	0.89
Donald Trump			0.05		0.05	0.91									22	0.91
Elizabeth Warren							0.9	0.1							21	0.9
Hillary Clinton							.07	0.89						0.04	27	0.89
John McCain						0.05		0.05	0.9						21	0.9
Justin Trudeau								0.04		0.87	0.09				23	0.87
Marco Rubio											0.9		0.05	0.05	21	0.9
Michelle Obama								0.04				0.96			25	0.96
Paul Ryan										0.09			0.91		23	0.91
Ted Cruz											0.17		0.04	0.78	23	0.78
Overall															312	0.91

of the model to apply excessive weight to skin tone is even more obvious with the artistic representation test data. Multiple images in which the subject is depicted with slightly darkened or non-physical (low brightness) skin tone are mistaken for Ben Carson. In addition, depictions of Barack Obama are mistaken for Michelle Obama or Ben Carson.

6 Discussion

While the network performs well on well-framed, well-lit data, it appears that improvements are required to train the network to give more weight to facial structure and overall appearance and less weight to hair style, gender, and skin tone. One possible method to address these deficiencies would be to train the network on stylized artistic images of the subjects. However, such training data is not readily available for most faces (e.g. images of users posted to social media networks). Alternatively, some radomized color, contrast, sharpness, and white balance adjustments applied as part of the network training would likely improve overall performance. Furthermore, duplication of a subset of the training data by converting some images to grayscale may also improve performance. In addition to changes in training data and methodology, more advanced networks may be able to improve performance. Notably, batch normalization implemented in Google Inception networks could potentially reduce sensitivity to image exposure/lighting (interpreted as changes in skin tone) [7], [8]. More intelligent image processing techniques also leveraging neural networks (e.g. identification of facial structure keypoints) could also significantly improve the accuracy and extensibility of this technique. However, use of such advanced features would require more complex tagging of the training data or use of a layered network approach leveraging pretrained facial recognition networks (or at least datasets).

7 Conclusion and Future work

This paper has demonstrated the feasibility of using readily available tools (NVIDIA DIGITS) and proven deep neural network architectures (GoogLeNet). With limited training data gathered from public search engines, a number of politicians were repeatably classified by the trained network with relatively high accuracy. Initial results and further investigation leveraging artistic representations of the classes revealed some notable deficiencies in the network with regard to reliance on gender, hair style, and skin tone. These deficiencies could also lead to some issues in the general case with undetected biases as a result of image quality, exposure, or color balance. A number of possible future improvements are proposed. The use of modern networks with batch normalization could improve the performance of the network on variable image properties (lighting, exposure, color balance, etc.). Furthermore, careful manipulation of the training images' properties (contrast, sharpness, brightness, color balance, etc.) and conversion to to grayscale color space would improve performance. These small changes would allow the demonstrated approach to perform acceptably well in the general case, though more advanced approaches would be required to achieve the scale and efficiency with respect to training data observed in social networks such as Facebook and YouTube.

REFERENCES

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CoRR*, vol. abs/1503.03832, 2015.

[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, June 2014.

[3] L. Yeager, J. Bernauer, A. Gray, and M. Houston, "Digits: the deep learning gpu training system," in *ICML 2015 AutoML Workshop*, 2015.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

[5] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *CoRR*, vol. abs/1605.07678, 2016.

[6] J. Croom, "Google image scraper." https://github.com/croomjm/Google-Image-Scraper, 2018.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

[8] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

Fig. 3: Artistic Representations of Selected Classes with Predicted Labels

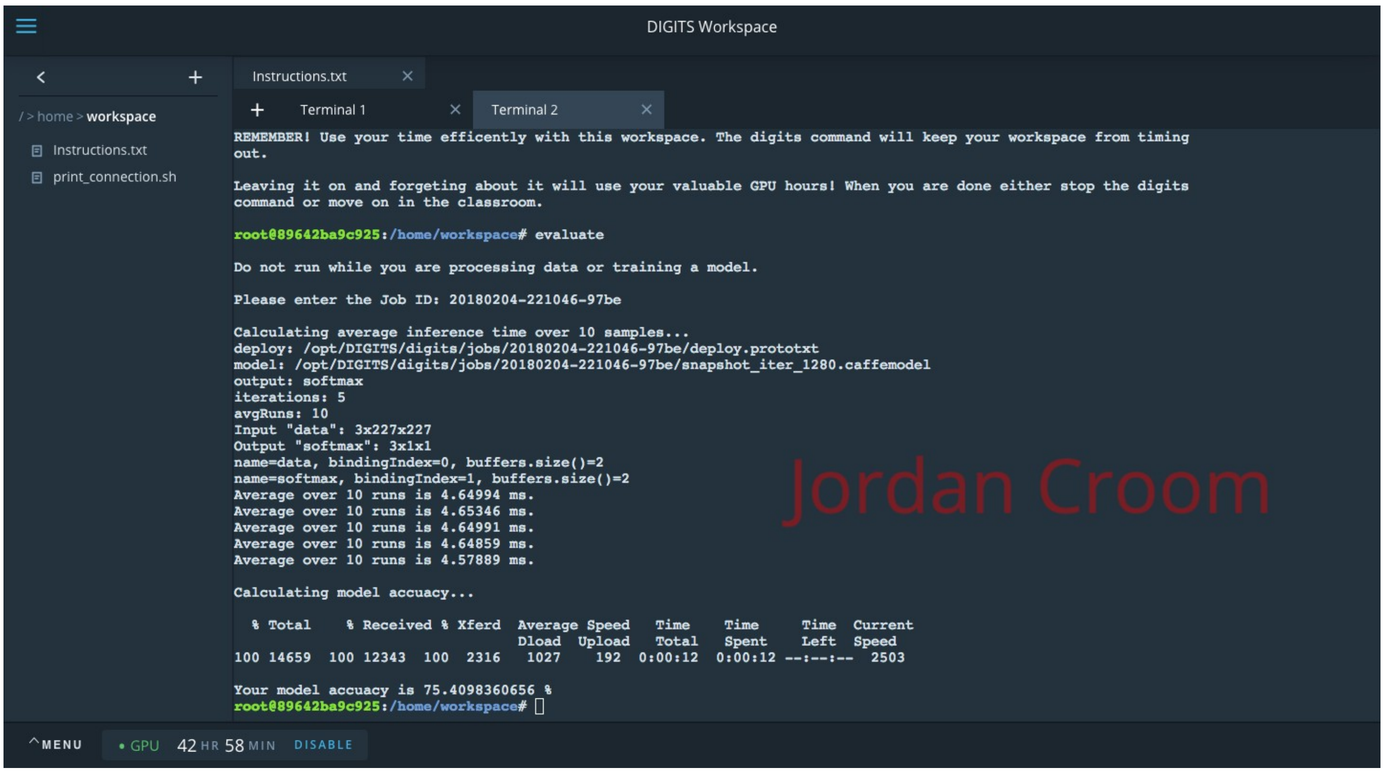

Fig. 4: Accuracy and Evaluation Time Results for Udacity Provided Sample Data