Convolutional Neural Network Transfer Learning for Robust Face Recognition in NAO Humanoid Robot

Daniel Bussey[1], Alex Glandon[2], Lasitha Vidyaratne[2], Mahbubul Alam[2], Khan Iftekharuddin[2]
[1]Embry-Riddle Aeronautical University, [2]Old Dominion University

Background Research

- Artificial Neural Networks
  - Computing systems whose model architecture is inspired by biological neural networks [1].
  - Capable of improving task performance without task-specific programming.
  - Show excellent performance at classification based tasks such as face recognition [2], text recognition [3], and natural language processing [4].
- Deep Learning
  - A subfield of machine learning that focuses on algorithms inspired by the function and structure of the brain called artificial neural networks. The method used to allow computers to learn through a process called training.
- Transfer Learning
  - Retraining a pre-existing neural network on a new data set to perform a specific task.
- Face Recognition Software
  - Software capable of identifying a person from a digital image or video frame.

Research Goals and Objectives

- Apply transfer learning to the convolutional neural network AlexNet [5] for face recognition tasks.
- Compare the performance of the retrained AlexNet to VGG-Face [6] for face recognition using high and low resolution images at different distances.
- Develop a robust facial recognition pipeline to implement on to the humanoid robotic platform NAO.

Figure 1. Humanoid Robotic Platform NAO.

Approach

- Retrain AlexNet on the CASIA-WebFace dataset to configure the neural networf for face recognition tasks.
- Acquire an input image using NAO's camera or a high resolution camera to run through the convolutional neural network.
- Extract the features of the input image using the neural networks AlexNet and VGG-Face
- Compare the features of the input image to the features of each image in the people database.
- Determine whether a match is detected or if the input image is a photo of a person who does not exist in the database.
- Evaluate the overall performance of each neural networks' ability to perform face recognition tasks.

How it Works

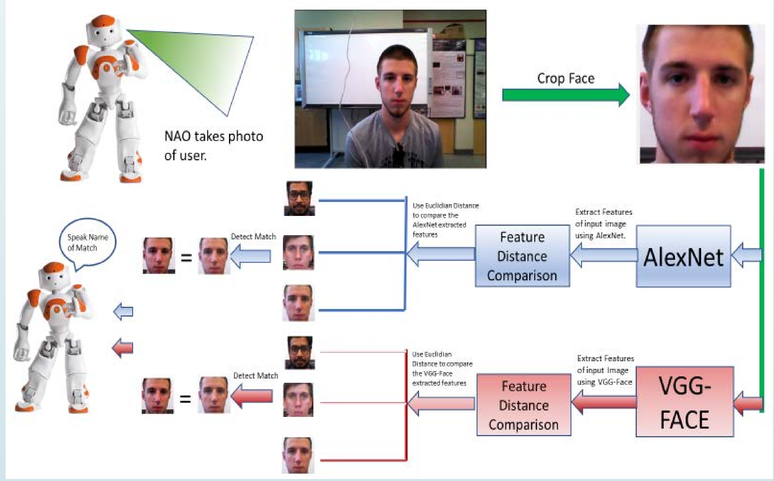

Figure 2: Facial Recognition Pipeline utilizing NAO

Output	Should The System Recognize You?	Does the System Recognize You?
True Positive (TP)	YES	YES
True Negative (TN)	NO	NO
False Positive (FP)	NO	YES
False Negative (FN)	YES	NO

Figure 3. Shows the different outcomes that occur when testing face recognition.

Results

- VGG-Face Shows better results in every performance benchmark measured compared to AlexNet, although AlexNet is able to extract features from an image 800% faster than VGG-Face
- Resolution of the input image does not have a statistically significant impact on the performance of VGG-Face and AlexNet.
- AlexNet's performance decreases with respect to distance from the camera where VGG-Face shows no performance loss.
- Both frameworks show excellent performance when eliminating false positives.
- Performance Charts shown below.

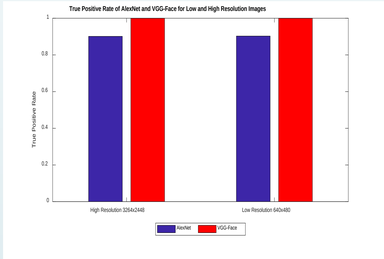

TABLE I. FACE RECOGNITION ACCURACY OF ALEXNET AND VGG-FACE WITH RESPECT TO DISTANCE USING LOW RESOLUTION IMAGES.

Distance of Face from Low Resolution Camera	Low Resolution Accuracy using AlexNet	Low Resolution Accuracy using VGG-Face
2 Feet	91.67%	100%
4 Feet	91.67%	100%
6 Feet	87.50%	100%

TABLE III. TPR AND FPR USING ALEXNET AND VGG-FACE FOR HIGH AND LOW RESOLUTION IMAGES

Network	Image Resolution	TPR	FPR
AlexNet	640x480	0.9012	0
AlexNet	640x480	0.9012	0
Vgg-Face	3264x2448	1	0
Vgg-Face	3264x2448	1	0

TABLE II. FACE RECOGNITION ACCURACY OF ALEXNET AND VGG-FACE WITH REPSPECT TO DISTANCE USING HIGH RESOLUTION IMAGES

Distance of Face from High Resolution Camera	High Resolution Accuracy using AlexNet	High Resolution Accuracy using VGG-Face
2 Feet	100%	100%
4 Feet	88.89%	100%
6 Feet	81.48%	100%

Future Research and Applications

- Implement multi face recognition from a single image, and utilize more complex classification methods.
- Perform additional complex recognition tasks such as face expression recognition, text recognition and object recognition.
- Security Applications: espionage, defense, homeland security, surveillance.
- Medical Applications: MRI image Processing.

Contact

<Daniel Bussey>
<NSF Cyber Security REU>
Email: busseyd@my.erau.edu
Phone: 706 421 3466

References

[1] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," Neural Networks, vol. 16, no. 5-6, pp. 555-559, 2003.
[2] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," arXiv preprint arXiv:1502.00873, 2015.
[3] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012, pp. 3304-3308.
[4] R. Collobert and J. Weston, "A unified architecture for natural language processing: deep neural networks with multitask learning," presented at the Proceedings of the 25th international conference on Machine learning, Helsinki, Finland, 2008.
[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.
[6] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in BMVC, 2015, vol. 1, no. 3, p. 6.