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

Daniel Bussey1, Alex Glandon2, Lasitha Vidyaratne2, Mahbubul Alam2, Khan Iftekharuddin2
1Emory-Riddle Aeronautical University, 2Old 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

- 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.

Approach

- Retrain AlexNet on the CASIA-WebFace dataset to configure the neural network 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

- 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 from the convolutional neural network.
- 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.

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.

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.

Figure 1: Humanoid Robotic Platform NAO.

Figure 2: Facial Recognition Pipeline utilizing NAO

Table I. Face Recognition Accuracy of AlexNet and VGG-Face with Respect to Distance Using Low Resolution Images

<table>
<thead>
<tr>
<th>Distance of Face from Low Resolution Images</th>
<th>VGG-Face</th>
<th>AlexNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Feet</td>
<td>88.89%</td>
<td>91.67%</td>
</tr>
<tr>
<td>2 Feet</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>6 Feet</td>
<td>87.50%</td>
<td>95.83%</td>
</tr>
</tbody>
</table>

Table II. Face Recognition Accuracy of AlexNet and VGG-Face with Respect to Distance Using High Resolution Images

<table>
<thead>
<tr>
<th>Distance of Face from High Resolution Images</th>
<th>VGG-Face</th>
<th>AlexNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Feet</td>
<td>91.67%</td>
<td>95.83%</td>
</tr>
<tr>
<td>2 Feet</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>6 Feet</td>
<td>97.92%</td>
<td>97.92%</td>
</tr>
</tbody>
</table>

Table III. TPR and FPR Using AlexNet and VGG-Face for High and Low Resolution Images

<table>
<thead>
<tr>
<th>Resolution</th>
<th>TPR</th>
<th>FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.9012</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0.9792</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

References