CLASSIFICATION OF CZECH SIGN LANGUAGE ALPHABET LETTERS USING CNN – PRELIMINARY STUDY

Jiri Krejsa, Stanislav Vechet

Institute of Thermomechanics AS CR v.v.i.

Overview

1. Introduction
2. Data gathering
3. Convolution Neural Networks
4. Results and Discussion
5. Future work
Introduction

- Sign language – natural language for hearing impaired
- Own grammar, lexicon, not universal (differences even within Czech Republic)
- Czech sign language alphabet – diacritics, letter CH, both palm orientations
- The task – letter classification from image input

Important note: this presentation reflects the further development in the area after the publication of the paper, therefore the results are significantly improved compared to the preliminary study
Data acquisition and preprocessing

- Special acquisition application with automatic labeling
- Data augmentation
  - Translation
  - Rotation -3 to +3 degrees
  - Uniform scaling up / down
  - Non-uniform scaling
Convolution Neural Networks (CNN)

- Deep learning
- Implementation – TensorFlow library, GPU enabled
- Training algorithm: Adam - stochastic gradient descent method
- Maximum pooling layers
- Rectified linear unit transfer functions on convolution layers
- SoftMax transfer function on classification layer
- Overfitting handling
  - L2 regularization
  - Dropout 50%

### Selected topology

<table>
<thead>
<tr>
<th>Layer</th>
<th>Depth</th>
<th>Fcn</th>
<th>Output</th>
<th>Params</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv 3x3</td>
<td>12</td>
<td>ReLU</td>
<td>222x222</td>
<td>336</td>
</tr>
<tr>
<td>Pool 2x2</td>
<td>Max</td>
<td></td>
<td>111x111</td>
<td></td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>12</td>
<td>ReLU</td>
<td>109x109</td>
<td>1308</td>
</tr>
<tr>
<td>Pool 2x2</td>
<td>Max</td>
<td></td>
<td>54x54</td>
<td></td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>14</td>
<td>ReLU</td>
<td>52x52</td>
<td>1526</td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>14</td>
<td>ReLU</td>
<td>50x50</td>
<td>1778</td>
</tr>
<tr>
<td>Pool 2x2</td>
<td>Max</td>
<td></td>
<td>25x25</td>
<td></td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>18</td>
<td>ReLU</td>
<td>23x23</td>
<td>2286</td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>18</td>
<td>ReLU</td>
<td>21x21</td>
<td>2934</td>
</tr>
<tr>
<td>Pool 2x2</td>
<td>Max</td>
<td></td>
<td>10x10</td>
<td></td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>22</td>
<td>ReLU</td>
<td>8x8</td>
<td>3586</td>
</tr>
<tr>
<td>Conv 3x3</td>
<td>22</td>
<td>ReLU</td>
<td>6x6</td>
<td>4378</td>
</tr>
<tr>
<td>Pool 2x2</td>
<td>Max</td>
<td></td>
<td>3x3</td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>27</td>
<td>SoftMax</td>
<td>198</td>
<td>5373</td>
</tr>
</tbody>
</table>
Training and validation

Training set
- Set of images sized 224x224 with 3 channels (RGB)
- Number of images: 595256

Validation sets
- PG set: unknown person, large set (1181 nonaugmented images)
- TZ set: known person, low image quality (681)
- MX set: known person (342)
Results

Test accuracy during training – first and final epoch (training data set)
Validation data sets results for particular letters
Problematic letters – PG data set

Letter N

Letter M
Conclusions

- CNN application successful
- Average test results on validation data set over 87%
- The sign of letter Y is classified successfully in spite reverse hand orientation
- Gesturer type (natural hearing impaired / professional interpreter) does not play any role
- The majority of misclassifications is concentrated on a low number of other letters

Acknowledgement

The authors would like to thank sign language interpreters and hearing impaired friends for the help in data collection, mainly Tomáš Zbavitel, Jitka Hořanská, Petra Gluchová, Markéta Kunášková and Radka Kulichová

The results in this paper were obtained with institutional support RVO 61388998 of the Institute of Thermomechanics AS CR v.v.i. and with support of Mobility Plus Projects with Ministry of Science and Technology of AS CR, v.v.i., project no. MOST-20-06.
Future work (already in progress)

Diacritics

- Expressed by the motion mimicking the diacritics shape
- Images sequence processing – optical flow

Fingerspelling

- Sequential images sequence processing
- How to recognize transfer between letters?
- Use of probability modeling – known probabilities of output for given letters