IMPROVING RECOGNITION OF THAI HANDWRITTEN CHARACTER WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

Sarayut Gonwirat
and
Olarik Surinta

Multi-agent Intelligence Simulation Laboratory
Department of Information Technology
Faculty of Informatics
Mahasarakham University, Thailand
Handwritten Character Recognition

- **Automatic Reading System**, how can machines understand the context in document.

- **Various Applications**:
  - Historical Document Analysis
  - Text Image Retrieval
  - Traffic-sign Recognition
  - Signature Verification
Previous Works

Different Language Characters

Difficulties:
- Different head and line-trail.
- Stroke and curve.
- Personal writing style.
Feature Extraction

Handwritten Recognition Process

Images → Preprocessing → Segmentation → Feature Extraction (Local Descriptors) → Recognition (Machine Learning)

Local Descriptors

- **HoG**: Histogram of Orientated Gradients
- **SIFT**: Scale Invariant Feature Transform
- **LBP**: Local Binary Pattern
- **HoG-GFoDR**: Gradient Features of Discriminative Regions
Machine Learning

Handwritten Recognition Process

- Images
- Preprocessing
- Segmentation
- Feature Extraction (Local Descriptors)
- Recognition (Machine Learning)

Local Descriptors

- SVM: Support Vector Machine
- KNN: K - Nearest Neighbor
- MLP: Multi-Layer Perceptron
Research Contributions:

- Improve the efficiency recognition of Thai handwritten recognition by CNN
- Comparison of the CNN Architectures
  - VGGNet and Inception-ResNetV2 with feature-based SVM
  - Learning style, Scratch and Transfer Learning

Contributions:
CNN Background

AlexNet

Layers:

- Convolution Layers
- Pooling Layers
- Fully Connected Layers

https://figshare.com/articles/CNN_architecture_used_to_perform_image_classification_/5843691
CNN Background

AlexNet

Convolution Layer

Activation Function
Rectified Linear Unit (ReLU)

\[ X_{k,l,n}^{p+1} = \sum_{i,j,m} K_{i,j,m,n} x_{k+i-1,l+j-1,m}^p \]

\[ ReLU(x) = \max(0, x) \]
CNN Background

AlexNet

Pooling Layer

Fully Connected Layer

Softmax Function

$$\text{softmax}(x) = \frac{\exp(x)}{\sum_i^n \exp(x_i)}$$
**VGGNet**

- **Very Deep Convolutional Networks for Large-Scale Image Recognition [2013]**, VGGNet is the use of a convolution filter that is very small, only 3x3 filter when using convolution processing
- 64, 128, 256, 512 channels
Inception-ResNetV2

- Inception model series
  - Inception-v1, GoogLeNet (2013)
  - Inception-v2, Batch normalization (BN)
  - Inception-v3 Factorization convolutions, such as 7x1, 7x1 etc
  - Inception-v4, ResNet

*Inception-v4, Inception-ResNetv2 and the Impact of Residual Connections on Learning*

- InceptionResNetV2
  - Stem block
  - Inception-ResNet A B and C blocks
  - Reduction A and B Blocks
Stem block

229 x 229 x 3

35x 35 x 256
Inception-ResNet Block

Residual Connection + Inception Block
Reduction Block

Reduction-A

Filter Concat

$35 \times 35 \times 256$

$17 \times 17 \times 896$

$3x3\ Max\ Pool-s2-V$
$3x3\ Conv-c384-s2-V$

$1x1\ Conv-c256$
$3x3\ Conv-c256$
$3x3\ Conv-c384-s2-V$

Reduction-B

Filter Concat

$8 \times 8 \times 1792$

$17 \times 17 \times 896$

$3x3\ Max\ Pool-s2-V$
$3x3\ Conv-c384-s2-V$

$1x1\ Conv-c256$
$1x1\ Conv-c256$
$3x3\ Conv-c288$
$3x3\ Conv-c320-s2-V$

InceptionResNet

Image

Stem

Inception-ResNet A
$x5$

Reduction-A

Inception-ResNet B
$x10$

Reduction-B

Inception-ResNet C
$x5$

Average Pooling

Softmax
Scratch and Transfer Learning

- **Scratch Learning** is a complex process and takes a long time to learn due to the learning beginning with creation of a random weight,
- **Transfer learning** is applying knowledge from previous domains that have been learned. It is called the **Pre-trained model** which directly results in faster training and higher effectiveness.
Thai Handwritten Characters Dataset

- **ALICE-THI dataset**, 78 Thai characters; from 150 undergraduate students, aged 20-23,

- THI-C68 14,490 characters, 68 classes including:
  - 44 consonants,
  - 17 vowels,
  - 4 tones
  - 3 symbols
Experimental Setup

- CPU 3.0 GHz, Quad-Core, GPU GeForce GTX 1080Ti, Memory 16GB
- The 5-fold and 10-fold following ratios; Train:Valid:Test, 7:1:2 and 8:1:1, respectively.
- Stochastic Gradient Decent (SGD) with momentum, learning rate 0.001, momentum 0.9, decay learning rate 0.0001
- 100 epoch, and 32 batch size
- Pre-trained CNN Model ImageNet Dataset

<table>
<thead>
<tr>
<th>Properties</th>
<th>CNN Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VGGNet</td>
</tr>
<tr>
<td>Input Image Size (Pixel)</td>
<td>128x128</td>
</tr>
<tr>
<td>Memory (MB)</td>
<td>160.6</td>
</tr>
<tr>
<td>Parameters (M)</td>
<td>20</td>
</tr>
<tr>
<td>Train Times (sec/epoch)</td>
<td>41.00</td>
</tr>
<tr>
<td>Test Times (sec/image)</td>
<td>0.0014</td>
</tr>
</tbody>
</table>
## Experimental Result

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy Rate (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10-cv</td>
<td>5-cv</td>
<td></td>
</tr>
<tr>
<td>SiftD-SVM [18]</td>
<td>94.34</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HOGFoDRs-SVM [5]</td>
<td>-</td>
<td>98.76</td>
<td></td>
</tr>
<tr>
<td>VGGNet-Scratch</td>
<td>97.93 ± 0.55</td>
<td>96.93 ± 0.48</td>
<td></td>
</tr>
<tr>
<td>Inception-ResNet-Scratch</td>
<td>98.15 ± 0.24</td>
<td>97.79 ± 0.29</td>
<td></td>
</tr>
<tr>
<td>VGGNet-Transfer</td>
<td><strong>99.20 ± 0.27</strong></td>
<td><strong>98.81 ± 0.25</strong></td>
<td></td>
</tr>
<tr>
<td>Inception-ResNet-Transfer</td>
<td>98.88 ± 0.24</td>
<td>98.61 ± 0.14</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

- Effective in recognizing Thai handwritten characters (THI-C68) with a high rate of recognition.
- Experimental 2 CNN models are VGGNet-19 and Inception-ResNet-v2 architectures.
  - CNNs is higher than previous works, which are feature-based SVM
  - Transfer learning is a way to reduce learning time and increasing the efficiency of recognition.
- VGGNet-19 architecture with transfer learning has an accuracy rate at 99.20% of Thai handwritten characters
- VGGNet-19 is an appropriate model to solve the problems of “Thai Handwritten Character Recognition”
Thank you
61011262002@msu.ac.th