Deep Spiking Neural Network for Visual Pattern Recognition

Daqi Liu

School of Computer Science, University of Lincoln

October 2, 2017
Outline

1 Introduction

2 Main Works

3 Future Works
Visual Pathway

- **Ventral stream**

![Diagram of the visual pathway showing ventral and dorsal stream with brain regions like MT, V1, V2, V4, and V4/PIT](image-url)
Convolutional neural network (HMAX model)
Spiking Neural Network

- Spiking neurons

- ANN vs SNN

![Spiking neuron diagram](image)

![ANN vs SNN comparison](image)
Neuron Model

- Leaky integrate-and-fire (LIF) model

- Define neuron behaviors
- Coincidence detector
Spiking Coding Scheme

- Spiking rate vs spiking timing sequence
- Rank order coding (ROC)

- Neuron is only allowed to fire at most once
- First spike wave is enough for further processing
Spiking Coding Scheme

- One input image and its spiking pattern

![Image of input image and spiking pattern]

- Spiking pattern sequence

![Diagram of spiking pattern sequence]
Learning Method

- Spike-timing dependent plasticity (STDP)
Event-driven Continuous STDP Learning (ECS)

- **State-of-the-art Methods**
  - **Spiking rate-based models**
    - Vanishing/exploding gradient problem
    - Over-fitting, not robust
    - Incorporate global error information
    - Require long processing time
    - Not biologically plausible
  - **Spiking timing-based models**
    - Require supervisory signal, no strong experimental confirmation
    - STDP is used as a local feature extractor
    - Not biologically plausible
Event-driven Continuous STDP Learning (ECS)

- ECS architecture

Diagram:
- Input image sequence
- Feature extracting layer
- Spiking encoding layer
- HMAX model with sparsity and intermediate features
- C2 feature vector sequence
- Modified ROC scheme
- Spiking pattern sequence
- Event-driven STDP learning
- Soft winner-take-all
- Class 1
- Class 2
- Class n
- Spiking pattern learning layer (include n maps with k neurons for each map)
Event-driven Continuous STDP Learning (ECS)

- Convergence analysis
Event-driven Continuous STDP Learning (ECS)

- Robustness analysis
Experimental results

TABLE IV: Classification accuracy performance using different methods on MNIST database.

<table>
<thead>
<tr>
<th>Spiking Coding-type</th>
<th>Architecture</th>
<th>Preprocessing</th>
<th>(Un-)supervised</th>
<th>Learning Rule</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Simple random sampling a</td>
</tr>
<tr>
<td>Spiking convolutional neural network</td>
<td>Modified HMAX</td>
<td>Supervised</td>
<td>ECS (this paper)</td>
<td>89%</td>
<td>93.0%</td>
</tr>
<tr>
<td>Two layer network[10]</td>
<td>Simplified HMAX</td>
<td>Supervised</td>
<td>Tempotron rule</td>
<td>79.0%</td>
<td>N/A</td>
</tr>
<tr>
<td>Two layer network[11]</td>
<td>Simplified HMAX</td>
<td>Supervised</td>
<td>Tempotron rule</td>
<td>N/A</td>
<td>91.3%</td>
</tr>
<tr>
<td>Rate-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spiking RBM[5]</td>
<td>None</td>
<td>Supervised</td>
<td>Contrastive divergence, linear classifier</td>
<td>N/A</td>
<td>90.3%</td>
</tr>
<tr>
<td>Spiking RBM[6]</td>
<td>Enhanced training set to 120,000 examples</td>
<td>Supervised</td>
<td>Contrastive divergence</td>
<td>N/A</td>
<td>89.0%</td>
</tr>
</tbody>
</table>
| Spiking convolutional neural network[7] | None | Supervised | Backpropagation | N/A | 99.1%
| Spiking RBM[8]      | Thresholding | Supervised | Contrastive divergence | N/A | 92.6% |
| Spiking RBM[8]      | Thresholding | Supervised | Contrastive divergence | N/A | 91.9% |
| Two layer network[9] | Edge-detection | Supervised | STDP with calcium variable | N/A | 96.5% |
| Multi-layer hierarchical neural network[1] | Orientation-detection | Supervised | STDP with calcium variable | N/A | 91.6% |
| Two layer network[2] | None | Unsupervised | Rectangular STDP | N/A | 93.5% |
| Two layer network[3] | None | Unsupervised | Exponential STDP | N/A | 95.0% |

a Simple random sampling performance has been generated by averaging 10 random tests using 50 random training samples per class and 100 random testing samples, which is suitable for real-time learning since the whole database is impossible to obtain in most real scenarios.

b Exhaustive performance shows the ideal experimental results by using whole 60000 training samples and 10000 testing samples within MNIST database.

c The authors only use 10000 testing samples to obtain the performance

d The authors only use 5000 testing samples to obtain the performance

e The authors use 10000 randomly chosen samples from MNIST database instead of the dedicated testing database.
Video-based Disguise Face Recognition (VDFR)

- **State-of-the-art VFR Methods**
  - Set-based methods
  - Sequence-based methods

- **Research Problems**
  - It is often hard to obtain the ideal face frames
  - Rely on the features which will be difficult to capture when there are invisible areas
  - Does not incorporate disguise variations in current databases
Video-based Disguise Face Recognition (VDFR)

- **VDFR architecture**

![Diagram of VDFR architecture]

- Input video
- Dynamic movements extracting layer (frame difference sequence)
- HMAX model with sparsity and intermediate features
- C2 feature vector sequence
- Modified ROC scheme
- Spiking pattern sequence
- Event-driven STDP learning
- Soft winner-take-all
- Spiking pattern learning layer (include n maps with k neurons for each map)
- Output Layer (include n maps with only one neuron for each map)
Video-based Disguise Face Recognition (VDFR)

- dynamic facial movements

![Grayscale dynamic changes of a fixed pixel (forehead)](image1)

![Differential dynamic changes of a fixed pixel (forehead)](image2)

![Grayscale dynamic changes of a fixed pixel (lips)](image3)

![Differential dynamic changes of a fixed pixel (lips)](image4)
Video-based Disguise Face Recognition (VDFR)

- Flowchart

---

Daqi Liu
Deep Spiking Neural Network for Visual Pattern Recognition
Video-based Disguise Face Recognition (VDFR)

- MakeFace database
Experimental results

TABLE III: Correct classification performance using different number of training and testing video clips (%).

<table>
<thead>
<tr>
<th>Number of training video clips</th>
<th>Number of testing video clips</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>92.5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>100</td>
</tr>
</tbody>
</table>

TABLE IV: Classification performances of two different methods on testing video clips with disguise (%).

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct rate</th>
<th>Wrong rate</th>
<th>Unknown rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN [29]</td>
<td>93.1 ± 1.35</td>
<td>6.9 ± 1.35</td>
<td>0</td>
</tr>
<tr>
<td>Proposed VDFR method</td>
<td>95.2 ± 2.65</td>
<td>4.8 ± 2.65</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The classification rate has been computed by averaging 10 random tests. Furthermore, we have conducted a Wilcoxon signed-rank test on the correct classification performances by using the above two methods and computed the significance level $p-value$ (0.03429). Such significance level ($p-value < 0.05$) indicates that the two correct classification performances are statistically different.

TABLE V: Classification performances of two different methods on testing mixed video clips (%).

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct rate</th>
<th>Wrong rate</th>
<th>Unknown rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN [29]</td>
<td>96.7 ± 0</td>
<td>3.3 ± 0</td>
<td>0</td>
</tr>
<tr>
<td>Proposed VDFR method</td>
<td>100 ± 0</td>
<td>0 ± 0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The classification rate has been computed by averaging 10 random tests.

TABLE VI: Classification performances of two different methods on testing unknown video clips (%).

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct rate</th>
<th>Wrong rate</th>
<th>Unknown rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN [29]</td>
<td>0</td>
<td>0</td>
<td>100 ± 0</td>
</tr>
<tr>
<td>Proposed VDFR method</td>
<td>0</td>
<td>0</td>
<td>100 ± 0</td>
</tr>
</tbody>
</table>

Note: The classification rate has been computed by averaging 10 random tests.
A Spiking LGMD Model for Collision Detection

- **Research Problems of Current Models**
  - Only incorporate spiking concept in final decision making step
  - Do not incorporate spiking neural network during detection
  - Do not generate the collision selection observed in LGMD cell

- **Proposed model**
  - Add a spiking encoding layer behind the P layer
  - Incorporate a Poisson point process to generate spike trains
  - Spikes are the only accepted information medium
  - Use an exponential level conductance-based LIF model within S layer
A Spiking LGMD Model for Collision Detection

- Differentiate the post-synaptic membrane potentials generated when approaching and receding the object
- Generate a similar collision selection as the real LGMD cell
- Compare with current models, it is biologically plausible
A Spiking LGMD Model for Collision Detection

![Graphs showing input spiking pattern sequence, LGMD cell PSP, and LGMD cell output spiking pattern sequence.](image)
A Spiking LGMD Model for Collision Detection

**Input spiking pattern sequence**

**LGMD cell PSP**

**LGMD cell output spiking pattern sequence**
A Spiking LGMD Model for Collision Detection

---

**Input spiking pattern sequence**

**LGMD cell PSP**

**LGMD cell output spiking pattern sequence**
A Spiking LGMD Model for Collision Detection

Input spiking pattern sequence

LGMD cell PSP

LGMD cell output spiking pattern sequence

Daqi Liu  Deep Spiking Neural Network for Visual Pattern Recognition
Future Works

- Finish the spiking LGMD model for collision detection
- Investigate the proposed VDFR method against a complex moving background
- Propose an alternative competitive learning method to replace the current STDP learning rule
Thank you!