Multi-column Deep Neural Networks for Image Classification

Dan C. Cireşan, Ueli Meier, Jürgen Schmidhuber
IDSIA, USI and SUPSI, Manno-Lugano, Switzerland
(dan,ueli,juergen)@idsia.ch

• 6-10 layers (deep) of non-linear neurons, and have hundreds of maps per layer (large), loosely inspired from biology.
• Introduced by Fukushima (80) and refined by LeCun et al.(98), Riesenhuber et al.(99), Simard et al.(03), Behnke (03), Ciresan et al.(11).
• Fully supervised, with randomly initialized filters, trained minimizing the classification error.
• Hard/impossible to train with gradient descent two decades ago
• Current CPU are more than 6000 times faster
• Given sufficient labeled training data, neither supervision nor carefully prewired synapses are needed
• Architecture:
  • convolutional layer (extracts features), max-pooling layer (selects features)
  • one output neuron per class, normalized with soft-max activation function

Deep Neural Networks (DNN)

DNN architecture: 29x29-20C4-MP2-40C5-MP3-150N-10N
50-100x speed-up compared with a single threaded CPU version of the CNN program (one day on GPU instead of two months on CPU)

CIFAR 10 18.50 11.21
1th error vs. 10.01% (Liu et al. 2010)
place at ICDAR 2011 competition

• 300 writers, 3755 classes, 1.2 Million characters, 3GB
• One week of training on GPU (corresponding to ~14 months on CPU)
• First method which works directly on images
• Offline task: 6.5% error vs. 10.01% (Liu et al. 2010)
• Online task: 5.64% vs. 7.61% (Liu et al. 2010)
• 1th place at ICDAR 2011 competition

Jittered-cluttered NORB

• 3D object recognition from stereo images
• Training set: 291600 48x48 stereo images
• 5 classes with 10 instances: 5 instances for training and 5 for testing
• Challenging dataset, only 5 instances/class, some instances from test set are completely different than the one from training set
• Data (# classes)
  • MNIST
  • NIST SD 19
  • Chinese characters – CASIA GB1

Traffic signs - GTSRB

• 26640 training and 12569 testing images, 43 classes
• 25 column MCDNN: 4+1 normalization x 5 instances
• Distortions: translation (max 15%), rotation (max 15°), scaling (max 15%)

Training set size

Errors for 4 runs [%] Mean [%]

First 2 folds
3.49 4.71 4.82 4.85 4.77

All 10 folds
4.49 4.71 4.82 4.85 4.77

MCDNN error [%]

• 6th best layers (deep) of non-linear neurons, and have hundreds of maps per layer (large), loosely inspired from biology.
• Introduced by Fukushima (80) and refined by LeCun et al.(98), Riesenhuber et al.(99), Simard et al.(03), Behnke (03), Ciresan et al.(11).
• Fully supervised, with randomly initialized filters, trained minimizing the classification error.
• Hard/impossible to train with gradient descent two decades ago
• Current CPU are more than 6000 times faster
• Given sufficient labeled training data, neither supervision nor carefully prewired synapses are needed
• Architecture:
  • convolutional layer (extracts features), max-pooling layer (selects features)
  • one output neuron per class, normalized with soft-max activation function

MNIST

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Best result of others [%]</th>
<th>MCDNN error [%]</th>
<th>Relative improvement [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.39</td>
<td>0.23</td>
<td>41</td>
</tr>
<tr>
<td>NIST SD 19</td>
<td>10.80</td>
<td>7.61</td>
<td>26</td>
</tr>
<tr>
<td>HWDB1.0 on</td>
<td>10.01</td>
<td>6.50</td>
<td>35</td>
</tr>
<tr>
<td>HWDB1.0 off</td>
<td>18.50</td>
<td>11.21</td>
<td>39</td>
</tr>
<tr>
<td>CIFAR 10</td>
<td>5.00</td>
<td>2.70</td>
<td>46</td>
</tr>
</tbody>
</table>

NORB

Conclusions

• Human-competitive results are reported on widely used computer vision benchmarks.
• MCDNN improves the state-of-the-art by 30-80%.
• Fully supervised, does not use any additional unlabeled data source.
• No need to extract handcrafted features, works on raw pixel images
• Robust (smallest error rates) and fast (10^10 images/s) for immediate industrial applications.

What is next?

• Segmentation: neuronal structures in EM stacks (1th place at ISB 2012 EM segmentation competition)
• More at www.idsia.ch/~ciresan

MNIST

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Best result of others [%]</th>
<th>MCDNN error [%]</th>
<th>Relative improvement [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.39</td>
<td>0.23</td>
<td>41</td>
</tr>
<tr>
<td>NIST SD 19</td>
<td>10.80</td>
<td>7.61</td>
<td>26</td>
</tr>
<tr>
<td>HWDB1.0 on</td>
<td>10.01</td>
<td>6.50</td>
<td>35</td>
</tr>
<tr>
<td>HWDB1.0 off</td>
<td>18.50</td>
<td>11.21</td>
<td>39</td>
</tr>
<tr>
<td>CIFAR 10</td>
<td>5.00</td>
<td>2.70</td>
<td>46</td>
</tr>
</tbody>
</table>

MNIST

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Best result of others [%]</th>
<th>MCDNN error [%]</th>
<th>Relative improvement [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.39</td>
<td>0.23</td>
<td>41</td>
</tr>
<tr>
<td>NIST SD 19</td>
<td>10.80</td>
<td>7.61</td>
<td>26</td>
</tr>
<tr>
<td>HWDB1.0 on</td>
<td>10.01</td>
<td>6.50</td>
<td>35</td>
</tr>
<tr>
<td>HWDB1.0 off</td>
<td>18.50</td>
<td>11.21</td>
<td>39</td>
</tr>
<tr>
<td>CIFAR 10</td>
<td>5.00</td>
<td>2.70</td>
<td>46</td>
</tr>
</tbody>
</table>

NORB

Conclusions

• Human-competitive results are reported on widely used computer vision benchmarks.
• MCDNN improves the state-of-the-art by 30-80%.
• Fully supervised, does not use any additional unlabeled data source.
• No need to extract handcrafted features, works on raw pixel images
• Robust (smallest error rates) and fast (10^10 images/s) for immediate industrial applications.

What is next?

• Segmentation: neuronal structures in EM stacks (1th place at ISB 2012 EM segmentation competition)
• More at www.idsia.ch/~ciresan

MNIST

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Best result of others [%]</th>
<th>MCDNN error [%]</th>
<th>Relative improvement [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.39</td>
<td>0.23</td>
<td>41</td>
</tr>
<tr>
<td>NIST SD 19</td>
<td>10.80</td>
<td>7.61</td>
<td>26</td>
</tr>
<tr>
<td>HWDB1.0 on</td>
<td>10.01</td>
<td>6.50</td>
<td>35</td>
</tr>
<tr>
<td>HWDB1.0 off</td>
<td>18.50</td>
<td>11.21</td>
<td>39</td>
</tr>
<tr>
<td>CIFAR 10</td>
<td>5.00</td>
<td>2.70</td>
<td>46</td>
</tr>
</tbody>
</table>

NORB

Conclusions

• Human-competitive results are reported on widely used computer vision benchmarks.
• MCDNN improves the state-of-the-art by 30-80%.
• Fully supervised, does not use any additional unlabeled data source.
• No need to extract handcrafted features, works on raw pixel images
• Robust (smallest error rates) and fast (10^10 images/s) for immediate industrial applications.

What is next?

• Segmentation: neuronal structures in EM stacks (1th place at ISB 2012 EM segmentation competition)
• More at www.idsia.ch/~ciresan

MNIST

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Best result of others [%]</th>
<th>MCDNN error [%]</th>
<th>Relative improvement [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.39</td>
<td>0.23</td>
<td>41</td>
</tr>
<tr>
<td>NIST SD 19</td>
<td>10.80</td>
<td>7.61</td>
<td>26</td>
</tr>
<tr>
<td>HWDB1.0 on</td>
<td>10.01</td>
<td>6.50</td>
<td>35</td>
</tr>
<tr>
<td>HWDB1.0 off</td>
<td>18.50</td>
<td>11.21</td>
<td>39</td>
</tr>
<tr>
<td>CIFAR 10</td>
<td>5.00</td>
<td>2.70</td>
<td>46</td>
</tr>
</tbody>
</table>