High Throughput Phenotyping with 3D CNNs in Ladder Networks

Hanqi Sun, Paul Pu Liang, Yue Niu
Carnegie Mellon University

Objective: Plant Phenotyping

Phenotyping plays an important role in creating new plant hybrids to increase crop yield for food. Current phenotyping techniques use 2D images which cannot capture the 3D nature of plants. Our model extracts features from 3D plant models to improve modern phenotyping.

To address the difficulty of collecting labels, we also propose a new architecture of CNNs that can be trained in a semi-supervised setting. These models are flexible, this is our chosen architecture for MNIST classification, inspired by Alexnet:

\[
\begin{align*}
\text{Conv11-96} & \rightarrow \text{Pool3} \rightarrow \text{Conv3-256} \\
\text{Pool2} & \rightarrow \text{Conv11-96} \rightarrow \text{Pool12} \rightarrow \text{Conv3-384} \rightarrow \text{Conv3-384} \\
& \rightarrow \text{Conv3-256} \rightarrow \text{FC-256} \rightarrow \text{FC-128} \\
\end{align*}
\]

As for Plant Leaf Angle regression, we settled on a Fully Connected Neural Network with 3 hidden layers.

3D Convolution and Max-pooling

To achieve our goal of feature extraction on a 3D model, we need to additionally implement the 3D convolution and max-pooling functions. Since the ladder network involves a decoder layer, we also require the approximate inverse of these functions. In practice, we use the gradient of 3D convolution to approximate its inverse and we use up-sampling via Nearest Neighbor Interpolation to approximate the inverse of 3D max-pooling. Existing implementations in 2D are extended to work with 3D voxelized images.

Dataset

- Simulated Plant Data: using cylinders to approximate stems and rectangles for leaves. This is abundant, clean and well labeled.
- Field Plant Data: test data of noisy plant point clouds from plants grown in the field.
- MNIST images for classification.
- ModelNet: a large-scale 3D model dataset released by Princeton University (Wu et al., 2014) for classification.

Ladder Networks

Ladder networks were proposed by (Rasmus et al., 2015) as a flexible model for semi-supervised learning. They use denoising autoencoders to learn hidden representations. Their flexibility allows us to use our 3D CNN architecture as the encoder layer.

\[
\begin{align*}
C_{\text{class}} &= -\frac{1}{N} \sum_{n=1}^{N} \log P(y = t(n)|x(n)) \\
C_{\text{reg}} &= \frac{1}{N} \sum_{n=1}^{N} (y - t(n)|x(n))^2 \\
C_{d} &= \sum_{l=1}^{L} \lambda_l \sum_{n=1}^{N} \|x^{(l)}(n) - \hat{x}^{(l)}(n)\|^2 \\
\end{align*}
\]

The cost function \( C \) has two parts, the forward cost and the denoising cost.

\[
C_{\text{class}} = \sum_{n=1}^{N} \log P(y = t(n)|x(n))
\]

Forward cost for classification:

\[
C_{\text{reg}} = \frac{1}{N} \sum_{n=1}^{N} (y - t(n)|x(n))^2
\]

Forward cost for regression:

\[
C_{d} = \sum_{l=1}^{L} \lambda_l \sum_{n=1}^{N} \|x^{(l)}(n) - \hat{x}^{(l)}(n)\|^2
\]

In practice, we also set the \( \lambda \) for the last decoder layer to be relatively big in order for the network to learn the hidden representations in the image.

CNN Architecture

Since the encoder layer for the ladder network is flexible, this is our chosen architecture for MNIST classification, inspired by Alexnet:

\[
\begin{align*}
\text{Conv11-96} & \rightarrow \text{Pool3} \rightarrow \text{Conv3-256} \\
\text{Pool2} & \rightarrow \text{Conv11-96} \rightarrow \text{Pool12} \rightarrow \text{Conv3-384} \rightarrow \text{Conv3-384} \\
& \rightarrow \text{Conv3-256} \rightarrow \text{FC-256} \rightarrow \text{FC-128} \\
\end{align*}
\]

Future Efforts

We will have updates of 3D plant data and 3D ModelNet results shortly in the final report.

Visualization

<table>
<thead>
<tr>
<th>Model</th>
<th>Test error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-Sup. Model with 1000 labelled images</td>
<td>Test error %</td>
</tr>
<tr>
<td>Semi-sup. Embedding (Weston et al., 2012)</td>
<td>5.73</td>
</tr>
<tr>
<td>Transductive SVM (Weston et al., 2012)</td>
<td>5.38</td>
</tr>
<tr>
<td>MTC (Rifai et al., 2011)</td>
<td>3.64</td>
</tr>
<tr>
<td>Pseudo-label (Lee, 2013)</td>
<td>3.46</td>
</tr>
<tr>
<td>AtlasRBF (Pitelis et al., 2014)</td>
<td>3.68</td>
</tr>
<tr>
<td>DGN (Kingma et al., 2014)</td>
<td>2.40</td>
</tr>
<tr>
<td>Virtual Adversarial (Miyato et al., 2015)</td>
<td>1.32</td>
</tr>
<tr>
<td>Ladder Network (Ramsus et al., 2015)</td>
<td>0.84</td>
</tr>
<tr>
<td>Ladder CNN</td>
<td>1.92</td>
</tr>
</tbody>
</table>

Table: A comparison of our model with state-of-the-art research on semi-supervised MNIST classification. There were 1000 labelled samples out of 60000 training images and 10000 test images. Adam optimizer was used for our model in the last row.