Structured Output Learning
- Real applications require structured prediction
- Standard Model: Pairwise MRF/CRF
  - Sparse connection - easier learning and inference
  - Overly simplistic - only modeling pairwise correlations

Pattern Potentials
- Penalize linearly if output deviates from a pattern
- Combine base models
  - Sum
    \( f(y) = \sum_i d_i(y) + \theta_i, \theta_{\text{max}} \)
  - Min
    \( f^m(y) = \min(d_1(y), \theta_1, \ldots, d_J(y), \theta_j, \theta_{\text{max}}) \)

RBMs are like Pattern Potentials

MAP Inference with the "EM" Algorithm
- Variational bound
  \( -E(y; T) \geq f^s(y) + f^p(y) + \sum_i h_i + \sum_{h_i} q(h) \left( c_j + \sum_i w_i y_i \right) h_j + H(q) \)
- E-step: compute optimal \( q(h) \) with \( y \) fixed
  \( q(h) = \frac{\exp \left( \frac{1}{T} \sum_j \left( c_j + \sum_i w_i y_i \right) h_j \right)}{\sum_h \exp \left( \frac{1}{T} \sum_j \left( c_j + \sum_i w_i y_i \right) h_j \right)} \)
- M-step: change \( y \) with \( q \) fixed
  \( \sum_i \left( b_i + \sum_j w_{ij} y_i \right) y_i + f^m(y) + f^p(y) \)

Learning CHOPP Parameters
- Minimize expected loss
  \( L = \sum_y p(y|x) \ell(y, y^*) \)
- Follow the negative gradient estimated by a set of samples
  \( \frac{\partial L}{\partial \theta} \approx \frac{1}{N-1} \sum_{n=1}^N \left( \ell(y^n, y^*) - \frac{1}{N} \sum_{n'=1}^N \ell(y^{n'}, y^*) \right) \left( - \frac{\partial E(y^n)}{\partial \theta} \right) \)
  - Increase energy for samples with high loss
  - Decrease energy for samples with low loss

Experiments
- An example using RBM trained with CD
- More experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Horse IOU</th>
<th>Bird IOU</th>
<th>Person IOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unary Only</td>
<td>0.5119</td>
<td>0.5055</td>
<td>0.4979</td>
</tr>
<tr>
<td>iPW</td>
<td>0.5736</td>
<td>0.5585</td>
<td>0.5094</td>
</tr>
<tr>
<td>iPW+RBM</td>
<td>0.6722</td>
<td>0.5647</td>
<td>0.5126</td>
</tr>
<tr>
<td>iPW+eRBM</td>
<td>0.6990</td>
<td>0.5773</td>
<td>0.5253</td>
</tr>
</tbody>
</table>