Faster Attend-Infer-Repeat with Tractable Probabilistic Models

Karl Stelzner¹, Robert Peharz², Kristian Kersting¹,³

¹ Machine Learning Group, TU Darmstadt
² Machine Learning Group, University of Cambridge
³ Centre for Cognitive Science, TU Darmstadt

ICML 2019
June 13, 2019
Deep Models with Tractable Components

- Deep generative models are a powerful tool
- Scaling is limited by effectiveness of approximate inference
- Can we improve this by combining them with tractable models, such as Sum-Product Networks?
Deep Models with Tractable Components

• Deep generative models are a powerful tool
• Scaling is limited by effectiveness of approximate inference
• Can we improve this by combining them with tractable models, such as Sum-Product Networks?
Deep Models with Tractable Components

• Deep generative models are a powerful tool
• Scaling is limited by effectiveness of approximate inference
• Can we improve this by combining them with tractable models, such as Sum-Product Networks?
Attend-Infer-Repeat

$z_i^{what}, \quad i = 1, \ldots, N_{\text{max}}$

Neural Network

$y^i$

Attend-Infer-Repeat

\[ i = 1, \ldots, N_{\text{max}} \]

\[ z^{i}_{\text{what}} \rightarrow \text{Neural Network} \rightarrow y^{i} \rightarrow z^{i}_{\text{where}} \]

N

Attend-Infer-Repeat

$\mathbf{z}_i^{\text{what}} \rightarrow \mathbf{y}_i \rightarrow \mathbf{z}_i^{\text{where}}$

where $i = 1, \ldots, N_{\text{max}}$

Attend-Infer-Repeat

\[ \mathbf{z}_i^{\text{what}}, \mathbf{z}_i^{\text{where}} \]

\[ \mathbf{y}^i \]

\[ \mathbf{x} \]

\[ \mathbf{N} \]

\[ i = 1, \ldots, N_{\max} \]

[Specified Renderer]

[Neural Network]

[Attend, infer, repeat: Fast scene understanding with generative models. "NIPS 2016"]
Sum-Product Attend-Infer-Repeat

Use SPN to model objects
Each pixel in $y$ is occluded (unobserved), or can be inferred deterministically from $x$ and $z_{\text{where}}$
Sum-Product Attend-Infer-Repeat

Model background with another SPN
\[ p(x \mid N, z_{\text{where}}) = p_{bg}(x_{bg=1}) \prod_{i=1}^{N_{\text{max}}} p_{obj}(x_{i=1}) \]
Faster & More Robust Training

(a) MNIST

(b) Sprites

(c) Noisy MNIST

(d) Grid MNIST
# Background Model at Work

<table>
<thead>
<tr>
<th></th>
<th>original</th>
<th>input</th>
<th>SuPAIR result</th>
<th>SuPAIR reconstruction</th>
<th>AIR result</th>
<th>AIR reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table above illustrates the application of different models to process images. The `original` column shows the initial images, while the subsequent columns demonstrate the effects of various models (`input`, `SuPAIR result`, `SuPAIR reconstruction`, `AIR result`, `AIR reconstruction`) on these images. The models aim to improve or modify the images in specific ways, as evidenced by the changes in the visual representations.
Thank you!

Pacific Ballroom #89

github.com/stelzner/supair

stelzner@cs.tu-darmstadt.de