How clever is the FiLM model, and how clever can it be?

Alexander Kuhnle  
Huiyuan Xie  
Ann Copestake

Department of Computer Science and Technology  
University of Cambridge  
{aok25,hx255,aac10}@cam.ac.uk

ShapeWorld datasets

- **Existential:** "There is a red square." / "A red shape is a square."
- **Single-shape:** same as above, with only one object present
- **Logical:** two existential statements connected by: and, or, if, if and only if
- **Numbers:** zero to five; with modifiers: less/more than, at most/least, exactly, not
- **Quantifiers:** with modifiers as above: no, half, all, a/two third(s), a/three quarter(s)
- **Relational:** left, right, above, below, closer, farther, darker, lighter, smaller, bigger, same/different shape/color
- **Simple-spatial:** the first four spatial relations, with only two objects per scene
- **Relational-negation:** relational plus negated relations
- **Implicit-relational:** left, right, upper, lower, bigger, darker, lighter, closer, farther (of two target objects)
- **Superlatives:** superlative forms of the above, of an arbitrary number of target objects
- **Relational-like:** any of the datasets relational, implicit-relational and superlatives

Performance per dataset of FiLM and baselines

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CNN-LSTM</th>
<th>CNN-LSTM-SA</th>
<th>FiLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(single-shape)</td>
<td></td>
<td></td>
<td>100.0 87.2</td>
</tr>
<tr>
<td>existential</td>
<td>100.0 81.1</td>
<td>100.0 99.7</td>
<td>100.0 99.9</td>
</tr>
<tr>
<td>logical</td>
<td>79.7 62.2</td>
<td>76.5 58.4</td>
<td>99.9 98.9</td>
</tr>
<tr>
<td>numbers</td>
<td>75.0 66.4</td>
<td>99.1 98.2</td>
<td>99.6 99.3</td>
</tr>
<tr>
<td>quantifiers</td>
<td>72.1 69.1</td>
<td>84.8 80.8</td>
<td>97.7 97.0</td>
</tr>
<tr>
<td>(simple-spatial)</td>
<td>81.4 64.8</td>
<td>81.9 57.7</td>
<td>85.1 61.3</td>
</tr>
<tr>
<td>relational</td>
<td></td>
<td></td>
<td>50.6 51.0</td>
</tr>
<tr>
<td>implicit-rel</td>
<td></td>
<td></td>
<td>52.9 53.2</td>
</tr>
<tr>
<td>superlatives</td>
<td></td>
<td></td>
<td>50.8 50.2</td>
</tr>
</tbody>
</table>

Example instances

Learning from a broader set of instances

Performance per dataset of the FiLM model trained on a broader set of instances, including existential, logical, numbers, quantifiers and various combinations of relational-like instances:

- relational
- implicit-rel
- superlatives
- relational + implicit-rel
- relational + superlatives
- implicit-rel + superlatives

- Datasets combining a broader variety of instance types can be successfully learned if the relative amount of "difficult" instances is small.
- The learnability of such datasets is sensitive to how "related" or "difficult" the instances are.

Differences to findings for CLEVR

- Pretrained ResNet does not perform well.
- Overlapping objects can impede learning.
- Simple compositional generalization (simpler than CLEVR CoGenT) is learned perfectly.
- Relational statements are substantially more difficult to learn, at least in isolation.
- The presence of simpler instances likely benefits the learning of more complex ones.
- Performance on CLEVR does not transfer to all kinds of "CLEVR-like" abstract data.

- Monolithic benchmark datasets may conceal important insights into the capability of evaluated models to learn structurally different types of instances.

Learning boosted by simpler instances

Performance on relational / negation or existential + numbers (with overlap), when augmented with / pretrained on simple-spatial or existential instances, respectively:

- Augmenting training data with "simpler" instances can help the learning of more "difficult" instances, but improvements are unstable.
- Pretraining on instances which are "easier" to learn before moving to more "complex" ones yields more robust improvements.

Additional findings

- Pretrained ResNet does not perform well.
- Overlapping objects can impede learning.

GitHub projects & PDF versions

- ShapeWorld: [https://github.com/AlexKuhnle/ShapeWorld](https://github.com/AlexKuhnle/ShapeWorld)
- FiLM for ShapeWorld: [https://github.com/AlexKuhnle/film](https://github.com/AlexKuhnle/film)
- Paper & poster PDF, plus related papers: [https://www.cl.cam.ac.uk/~aok25/](https://www.cl.cam.ac.uk/~aok25/)