Focus technique: Outils avancés de Computer Vision

Thomas Serre and Sabine Muzellec

ANR-3IA Artificial and Natural Intelligence Toulouse Institute

Brown University
Cognitive Linguistic & Psychological Sciences
Carney Institute for Brain Science
Feedback mechanisms

- Deep feedforward nets (CNNs) have become orders of magnitude deeper than our visual system
- Limited generalization beyond the training data
Feedback mechanisms

- Feedback plays a key role in visual perception beyond feedforward sweep.
- Visual processing builds "depth" dynamically through recurrent connections.
Recurrent neural circuit model(s)

\[
C_{xyk}^{(1)} = (W^I \ast H^{(2)})_{xyk}
\]

\[
C_{xyk}^{(2)} = (W^E \ast H^{(1)})_{xyk}
\]

\[
\eta \dot{H}_{xyk}^{(1)} + \epsilon^2 H_{xyk}^{(1)} = \left[ \xi X_{xyk} - (\alpha H_{xyk}^{(1)} + \mu) C_{xyk}^{(1)} \right]_+
\]

\[
\tau \dot{H}_{xyk}^{(2)} + \sigma^2 H_{xyk}^{(2)} = \left[ \gamma C_{xyk}^{(2)} \right]_+
\]
Recurrent neural circuit model(s)

GRU (a) vs hGRU (b)
Vision beyond image categorization

PathFinder

Linsley et al NeurIPS 2018

PathTracker

Kim et al ICLR 2020

Linsley et al NeurIPS 2021

Ullman 1996
PathFinder

(a) Accuracy vs Path Length (6, 9, 14)

(b) Area under the Validation Curve

(c) Accuracy (Curve length = 14) vs Free parameter multiplier

(d) Path Length (6, 9, 14)
## PathFinder

### hGRU

This is a low-res PathFinder (32x32)

<table>
<thead>
<tr>
<th>Model</th>
<th>ListOps</th>
<th>Text</th>
<th>Retrieval</th>
<th>Image</th>
<th>Pathfinder</th>
<th>Path-X</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer</td>
<td>36.37</td>
<td>64.27</td>
<td>57.46</td>
<td>42.44</td>
<td>71.40</td>
<td>FAIL</td>
<td>54.39</td>
</tr>
<tr>
<td>Local Attention</td>
<td>15.82</td>
<td>52.98</td>
<td>53.39</td>
<td>41.46</td>
<td>66.63</td>
<td>FAIL</td>
<td>46.06</td>
</tr>
<tr>
<td>Sparse Trans.</td>
<td>17.07</td>
<td>63.58</td>
<td>59.59</td>
<td>44.24</td>
<td>71.71</td>
<td>FAIL</td>
<td>51.24</td>
</tr>
<tr>
<td>Longformer</td>
<td>35.63</td>
<td>62.85</td>
<td>56.89</td>
<td>42.22</td>
<td>69.71</td>
<td>FAIL</td>
<td>53.46</td>
</tr>
<tr>
<td>Linformer</td>
<td>35.70</td>
<td>53.94</td>
<td>52.27</td>
<td>38.56</td>
<td>76.34</td>
<td>FAIL</td>
<td>51.36</td>
</tr>
<tr>
<td>Reformer</td>
<td>37.27</td>
<td>56.10</td>
<td>53.40</td>
<td>38.07</td>
<td>68.50</td>
<td>FAIL</td>
<td>50.67</td>
</tr>
<tr>
<td>Sinkhorn Trans.</td>
<td>33.67</td>
<td>61.20</td>
<td>53.83</td>
<td>41.23</td>
<td>67.45</td>
<td>FAIL</td>
<td>51.39</td>
</tr>
<tr>
<td>Synthesizer</td>
<td>36.99</td>
<td>61.68</td>
<td>54.67</td>
<td>41.61</td>
<td>69.45</td>
<td>FAIL</td>
<td>52.88</td>
</tr>
<tr>
<td>BigBird</td>
<td>36.05</td>
<td>64.02</td>
<td>59.29</td>
<td>40.83</td>
<td>74.87</td>
<td>FAIL</td>
<td>55.01</td>
</tr>
<tr>
<td>Linear Trans.</td>
<td>16.13</td>
<td><strong>65.90</strong></td>
<td>53.09</td>
<td>42.34</td>
<td>75.30</td>
<td>FAIL</td>
<td>50.55</td>
</tr>
<tr>
<td>Performer</td>
<td>18.01</td>
<td>65.40</td>
<td>53.82</td>
<td><strong>42.77</strong></td>
<td><strong>77.05</strong></td>
<td>FAIL</td>
<td>51.41</td>
</tr>
<tr>
<td>Task Avg (Std)</td>
<td>29 (9.7)</td>
<td>61 (4.6)</td>
<td>55 (2.6)</td>
<td>41 (1.8)</td>
<td>72 (3.7)</td>
<td>FAIL</td>
<td>52 (2.4)</td>
</tr>
</tbody>
</table>

This is the real PathFinder (128x128)
PathFinder
hGRU

Google Colab
https://tinyurl.com/tinyhgru
Stabilizing RNNs to help generalize beyond training distrib.
MS COCO
hGRU

(a) image  (b) semantic segmentation
(c) instance segmentation  (d) panoptic segmentation

Panoptic Quality (PQ)

<table>
<thead>
<tr>
<th>Memory (GB)</th>
<th>CBP FPN 10 timesteps</th>
<th>BPTT FPN 3 timesteps</th>
<th>RBP FPN 20 timesteps</th>
<th>Feedforward FPN-ResNet 50 (+783,552 parameters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>39.5</td>
<td>39.5</td>
<td>39.5</td>
<td>39.5</td>
</tr>
<tr>
<td>10</td>
<td>40.0</td>
<td>40.0</td>
<td>40.0</td>
<td>40.0</td>
</tr>
<tr>
<td>11</td>
<td>40.5</td>
<td>40.5</td>
<td>40.5</td>
<td>40.5</td>
</tr>
<tr>
<td>12</td>
<td>41.0</td>
<td>41.0</td>
<td>41.0</td>
<td>41.0</td>
</tr>
<tr>
<td>13</td>
<td>41.5</td>
<td>41.5</td>
<td>41.5</td>
<td>41.5</td>
</tr>
<tr>
<td>14</td>
<td>42.0</td>
<td>42.0</td>
<td>42.0</td>
<td>42.0</td>
</tr>
<tr>
<td>15</td>
<td>42.5</td>
<td>42.5</td>
<td>42.5</td>
<td>42.5</td>
</tr>
<tr>
<td>16</td>
<td>43.0</td>
<td>43.0</td>
<td>43.0</td>
<td>43.0</td>
</tr>
<tr>
<td>17</td>
<td>43.5</td>
<td>43.5</td>
<td>43.5</td>
<td>43.5</td>
</tr>
</tbody>
</table>
Google colab
https://tinyurl.com/cocohgru
Vision beyond image categorization

PathFinder

Linsley et al NeurIPS 2018

PathTracker

Linsley et al NeurIPS 2021

cABC

Kim et al ICLR 2020

Ullman 1996
cABC
fGRU

(a) Max Validation Accuracy (Pathfinder)
(b) Max Validation Accuracy (Cluttered ABC)

Human (n=120)  TD+H-CNN  H-CNN  TD-CNN

Difficulty
Easy  Int.  Hard

Input  L2 Difference per timestep  Prediction
$t=2$  $t=4$  $t=6$  $t=8$
Contours detection

GammaNet

(a) Proportion of training data used
Vision beyond
image categorization

PathFinder

Linsley et al NeurIPS 2018

PathTracker

Linsley et al NeurIPS 2021

cABC

Kim et al ICLR 2020

Ullman 1996
PathTracker

InT

1 distractor and 32 frames

14 distractors and 32 frames
PathTracker

InT

InT Circuit

Connectivity

Description

\begin{align*}
i[t] &= g i[t - 1] + (1 - g) \left[ z[t] - (\gamma i[t] a[t] + \beta) m[t] - i[t - 1] \right]_+ \\
e[t] &= h e[t - 1] + (1 - h) \left[ i[t] + (\nu i[t] + \mu) n[t] - e[t - 1] \right]_+ \\
\end{align*}

where

\begin{align*}
M_{c,i} &= \left[ m_{xy}^{(c)} \right] = W_{c,i} \ast (E \odot A) \\
N_{i,e} &= \left[ n_{xy}^{(c)} \right] = W_{i,e} \ast I \\
I &= \left[ i_{xy}^{(c)} \right] \\
G &= \left[ g_{xy}^{(c)} \right] = \sigma(W_g \ast I + U_g \ast Z) \\
A &= \left[ a_{xy}^{(c)} \right] = \sigma(W_a \ast E) \\
H &= \left[ h_{xy}^{(c)} \right] = \sigma(W_h \ast E + U_h \ast I)
\end{align*}
PathTracker

InT

(b) Test: 1 distractors and 64 frames

Train: 14 distractors and 64 frames

Test: 25 distractors and 64 frames

Accuracy

14 distractors and 64 frames

Frame number

(e)
Currently #1 (#2)

TrackingNet leaderboard

Circuit-TransT TransT Circuit-TransT TransT
PathTracker
PathTracker

PathTracker Website: https://pathtracker.github.io/

PathTracker Code: https://github.com/pathtracker-code/
Serre Lab Models
(in a PyTorch plug-and-play format)

Github repo
Additional resources

- Serre lab [website](#) (including links to all publications)
- Thomas Serre [thomas_serre@brown.edu](mailto:thomas_serre@brown.edu)
- Sabine Muzellec [sabine_muzellec@brown.edu](mailto:sabine_muzellec@brown.edu)
Xplique
👋 A Neural Networks Explainability Toolbox
https://github.com/deel-ai/xplique

Thomas FEL and all the DEEL Team
thomas_fel@brown.edu