Recurrent Attention Models [Mnih et al. (2014)]

Yifei Huang

HKUST

November 27, 2018

Recurrent Attention Models (RAM)

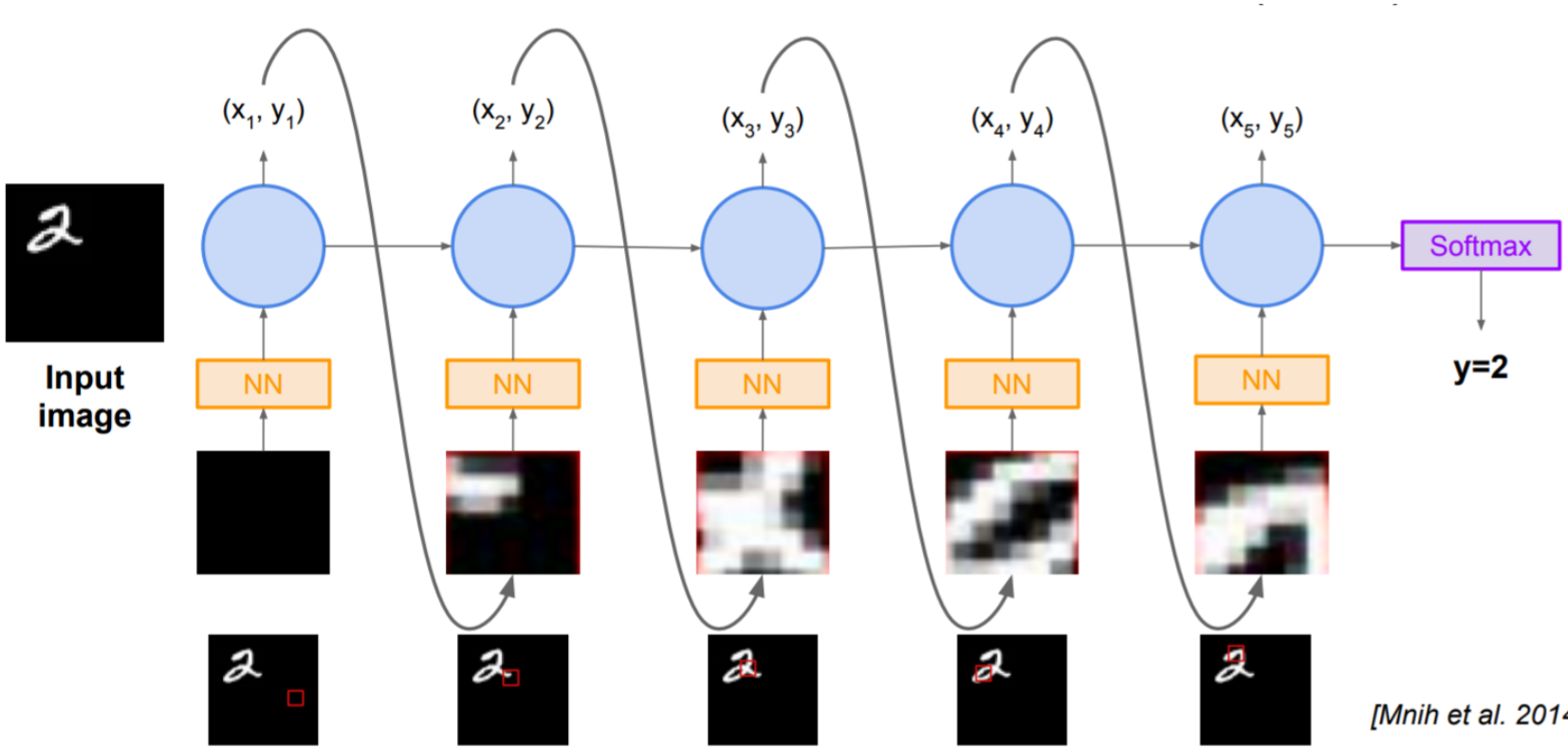

Figure: The procedure of RAM: first generating sequential parts in the original images as the input of RNN, then combining the final output and each corresponding part's outputs to do reinforcement learning.

Recurrent Attention Models (RAM)

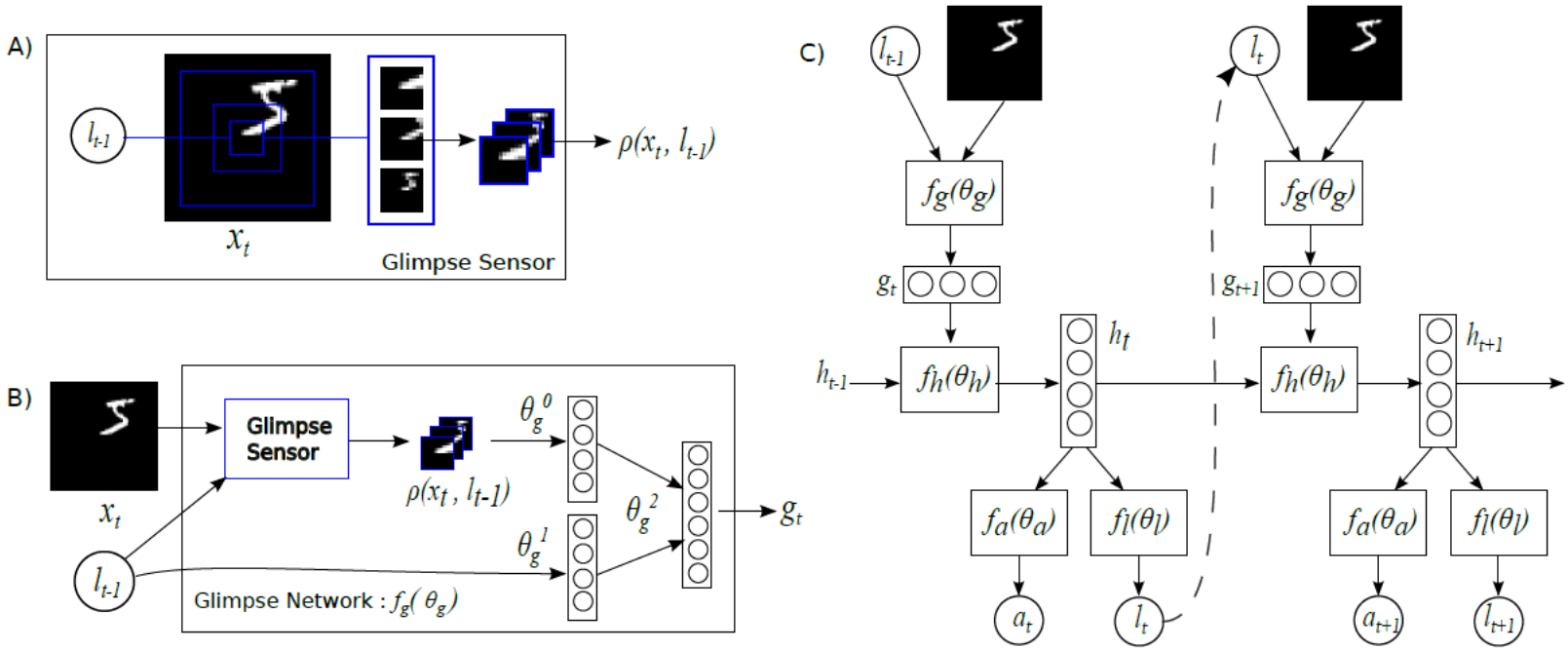

Figure: The components of RAM: **Glimpse Sensor**(A), **Glimpse Network**(B) and **RNN**(C)

Components of RAM

- **Glimpse Sensor**: Proposing representation for patches by Neural Network based on the location in original image.
- **Glimpse Network**: Generating features for each patches based on the locations and representations given by glimpse sensor.
- **RNN**: Based on last patch's hidden state and current patch's information given by glimpse network to give predictions on next patch's location information and current output which used as reward of reinforcement learning.

Key Ideas

- **Attention** : Humans focus attention selectively on parts of the visual space to acquire information when and where it is needed; reduce the task complexity;
- **Recurrent Neural Network**: Combine information from different fixations over time to build up an internal representation of the scene;
- **Reinforcement**: Guiding future eye movements and decision making

Results

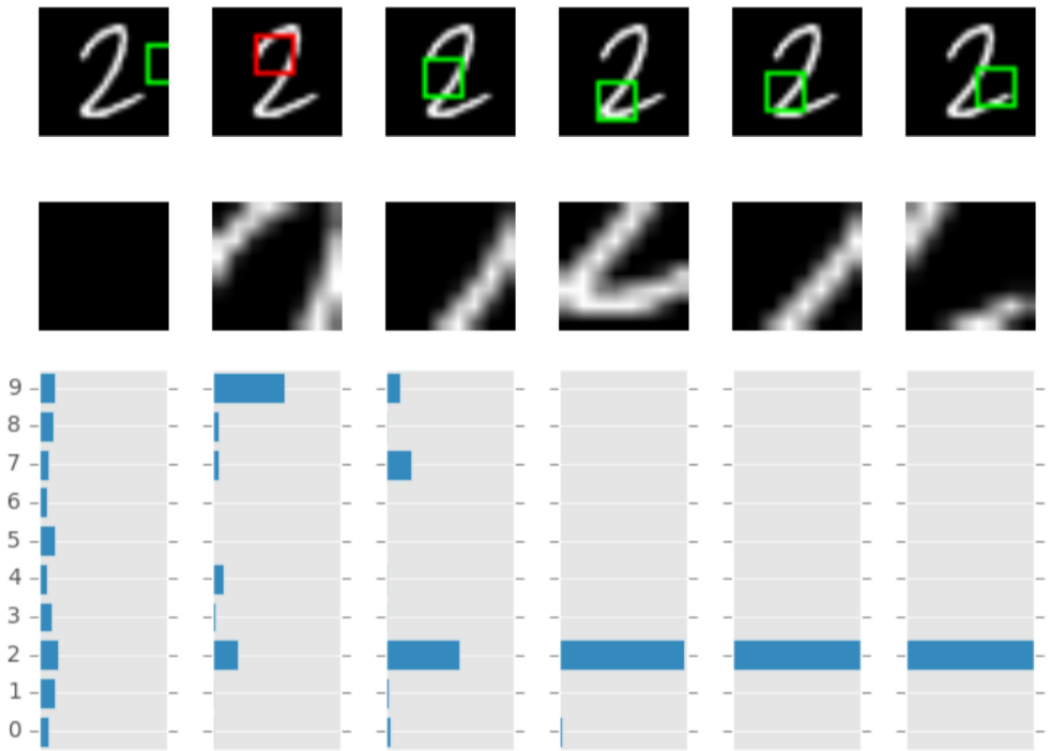

Figure: The progress of RAM on original MNIST

`https://github.com/amasky/ram`

Results

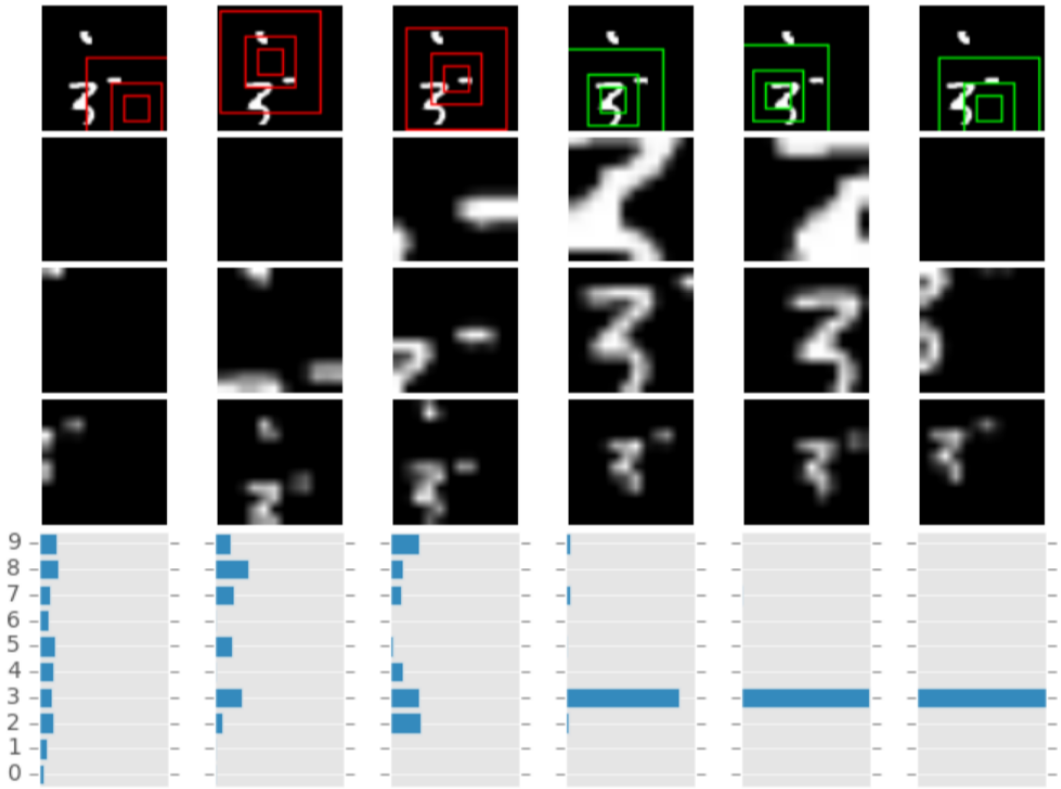

Figure: The progress of RAM on cluttered MNIST

`https://github.com/amasky/ram`

Pros & Cons

Pros

- The prediction given by RAM could be easily interpreted from the proposed image patches which means "where the network looks at"
- RAM is more robust than general CNN because it learns where it should look at based on the amazing results from cluttered MNIST
- RAM successfully combines RNN and reinforcement learning which gives direction of how to design neural network in the future

Cons

- Reinforcement learning (policy gradient used in the paper) takes too long to convergence, thus making RAM running for a long time.
- Tuning hyper-parameters is somehow a little difficult because of many different components there in RAM.

References

Mnih, Volodymyr and Heess, Nicolas and Graves, Alex and others, 2014

Recurrent models of visual attention

Advances in neural information processing systems 2204–2212.

The End