FigureNet : A Deep Learning model for Question-Answering on Scientific Plots

Revanth Gangi Reddy, Rahul Ramesh,
Ameet Deshpande and Mitesh M. Khapra

Computer Science & Engineering, Indian Institute of Technology Madras

Outline I

Task Description

Question Answering on Scientific plots

- Scientific plots contain information easily understood by humans.
- Why address this task?
  - Build better models for logical reasoning.
  - Automate information retrieval.
- Incorporate suitable biases to ameliorate the built models.

Question Answering on Scientific plots

- Scientific plots contain information easily understood by humans.
- Why address this task?
  - Build better models for logical reasoning.
  - Automate information retrieval.
- Incorporate suitable biases to ameliorate the built models.

Question Answering on Scientific plots

- Scientific plots contain information easily understood by humans.
- Why address this task?
  - Build better models for logical reasoning.
  - Automate information retrieval.
- Incorporate suitable biases to ameliorate the built models.

Dataset

In this work we deal with the Figure-QA dataset[1].

- Dataset contains bar-graphs, pie-charts and line charts.
- All questions have yes/no answers.
- Every element has a different colour (legend not needed).

This seems easy!

[1] Samira Ebrahimi Kahou et al. (2017). "Figureqa: An annotated figure dataset for visual reasoning". In: *arXiv preprint arXiv:1710.07300*.

Dataset

In this work we deal with the Figure-QA dataset[1].

- Dataset contains bar-graphs, pie-charts and line charts.
- All questions have yes/no answers.
- Every element has a different colour (legend not needed).

This seems easy!

[1] Samira Ebrahimi Kahou et al. (2017). "Figureqa: An annotated figure dataset for visual reasoning". In: *arXiv preprint arXiv:1710.07300*.

Dataset

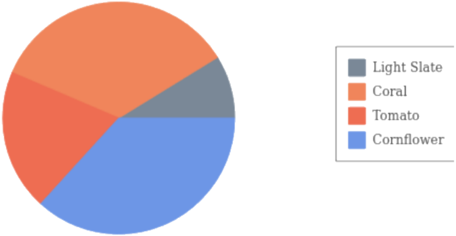
In this work we deal with the Figure-QA dataset[1].

- Dataset contains bar-graphs, pie-charts and line charts.
- All questions have yes/no answers.
- Every element has a different colour (legend not needed).

This seems easy!

[1] Samira Ebrahimi Kahou et al. (2017). "Figureqa: An annotated figure dataset for visual reasoning". In: *arXiv preprint arXiv:1710.07300*.

Dataset



Q: Is Coral the minimum?
A: No

Q: Is Cornflower the maximum?
A: Yes

Q: Is Light Slate greater than Coral?
A: No

Q: Is Light Slate less than Coral?
A: Yes

Q: Is Tomato the low median?
A: Yes

Figure-QA

Figure Type	CNN + LSTM	RN	FigureNet	Human
Vertical Bar	60.84	77.53	87.09	95.90
Horizontal Bar	61.06	75.76	82.19	96.03
Pie Chart	57.91	78.71	83.69	88.26
Average plot	59.94	77.37	84.32	93.40

Table: Accuracy Numbers for various models/figure types

Baselines

CNN + LSTM

- Simple baseline for Question-answering.
- Text processed using an LSTM.
- Image processed using images.
- Representations concatenated to finally yield single sigmoid unit.

CNN + LSTM

- Simple baseline for Question-answering.
- Text processed using an LSTM.
- Image processed using images.
- Representations concatenated to finally yield single sigmoid unit.

CNN + LSTM

- Simple baseline for Question-answering.
- Text processed using an LSTM.
- Image processed using images.
- Representations concatenated to finally yield single sigmoid unit.

Relational Networks[2]

- RNs are neural networks that are primarily concerned with relational reasoning based tasks.
- State of the art/super human performances when work was released.
- Network architecture encodes interactions between objects.

$$RN(O) = f_{\phi}\left(\sum_{i,j} g_{\theta}(o_i, o_j)\right)$$

[2] Adam Santoro et al. (2017). "A simple neural network module for relational reasoning". In: *Advances in neural information processing systems*, pp. 4974–4983.

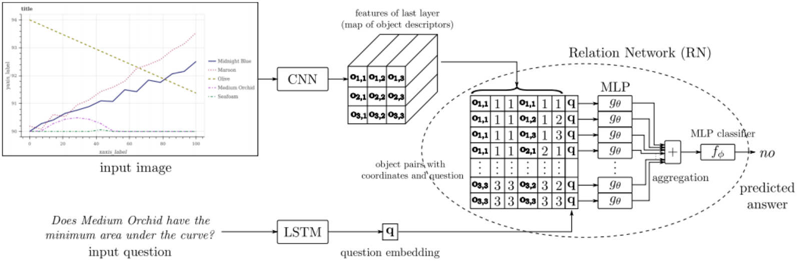
Relational Networks[2]

- RNs are neural networks that are primarily concerned with relational reasoning based tasks.
- State of the art/super human performances when work was released.
- Network architecture encodes interactions between objects.

$$RN(O) = f_\phi\left(\sum_{i,j} g_\theta(o_i, o_j)\right)$$

[2] Adam Santoro et al. (2017). "A simple neural network module for relational reasoning". In: *Advances in neural information processing systems*, pp. 4974–4983.

Relational Networks



FigureNet

Overview

- Inject bias / human prior into modelling decisions.
- Build *modules* (Andreas et al. 2016; Hu et al. 2017) that replicate elementary human operations
- Pre-train each module and assemble them together to form end-to-end differentiable network (Erhan et al. 2009).

Overview

- Inject bias / human prior into modelling decisions.
- Build *modules* (Andreas et al. 2016; Hu et al. 2017) that replicate elementary human operations
- Pre-train each module and assemble them together to form end-to-end differentiable network (Erhan et al. 2009).

Overview

- Inject bias / human prior into modelling decisions.
- Build *modules* (Andreas et al. 2016; Hu et al. 2017) that replicate elementary human operations
- Pre-train each module and assemble them together to form end-to-end differentiable network (Erhan et al. 2009).

Network Modules

- *Spectral Segregator Module:* Identify colours of various plot elements
- *Order Extraction Module:* Order plot elements based on size
- *Question Encoding:* Representation of the question
- *Colour Encoding:* Encoding of the colours in the question

Network Modules

- *Spectral Segregator Module:* Identify colours of various plot elements
- *Order Extraction Module:* Order plot elements based on size
- *Question Encoding:* Representation of the question
- *Colour Encoding:* Encoding of the colours in the question

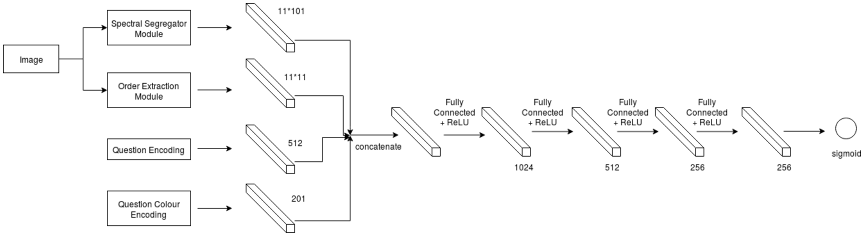
Network Modules

- *Spectral Segregator Module:* Identify colours of various plot elements
- *Order Extraction Module:* Order plot elements based on size
- *Question Encoding:* Representation of the question
- *Colour Encoding:* Encoding of the colours in the question

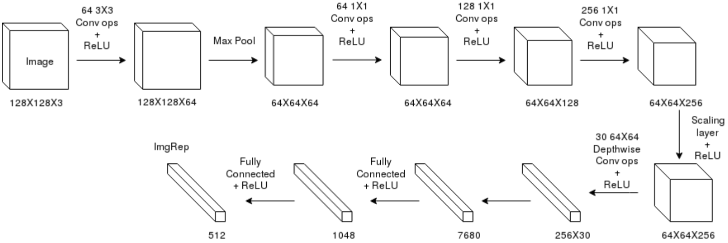
Network Modules

- *Spectral Segregator Module:* Identify colours of various plot elements
- *Order Extraction Module:* Order plot elements based on size
- *Question Encoding:* Representation of the question
- *Colour Encoding:* Encoding of the colours in the question

Architecture Description

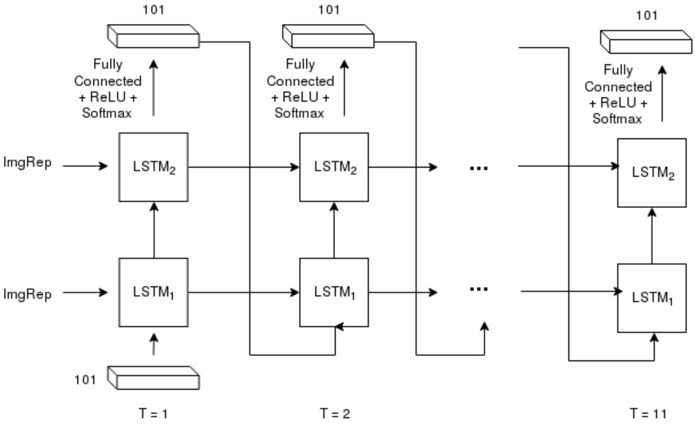


Spectral Segregator and Order Extractor



- Convolutions along the 3rd dimension
- Scaling layer along channels
- Modified 2-layer LSTM

LSTM for the Spectral Segregator



Effect of modified convolutions

Seperating various channels and operating on a subset is important.

Aggregating channels, eliminates information about the various colours.

Model	Spectral Seg.	Order Extr.
Only Convolutions	80.82	74.31
With Modified convolutions	15.76	54.04

Table: Accuracy of individual modules

Effect of modified LSTM

2 layers of LSTM are important for the ordering process and allows model to correct previous errors

Model	Accuracy
Our Model	84.29
Sampling output	81.61
with 1-layer LSTM	75.29
no LSTM	73.19

Training Time

Although our model has a number of components, the training time is significantly lesser

Model	Time (hours)
RNs	354.79
FigureNet	28.50

Conclusion

- Adapting to line plots
- Using legend and other plots elements (with OCR?)
- Inherent biases in the dataset, will model generalize to more complex plots?
- Combine different modules that maybe relevant

References I

- Andreas, Jacob et al. (2016). "Neural module networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 39–48.
- Erhan, Dumitru et al. (2009). "The difficulty of training deep architectures and the effect of unsupervised pre-training". In: *Artificial Intelligence and Statistics*, pp. 153–160.
- Hu, Ronghang et al. (2017). "Learning to reason: End-to-end module networks for visual question answering". In: *CoRR, abs/1704.05526* 3.
- Kahou, Samira Ebrahimi et al. (2017). "Figureqa: An annotated figure dataset for visual reasoning". In: *arXiv preprint arXiv:1710.07300*.

References II

Santoro, Adam et al. (2017). "A simple neural network module for relational reasoning". In: *Advances in neural information processing systems*, pp. 4974–4983.