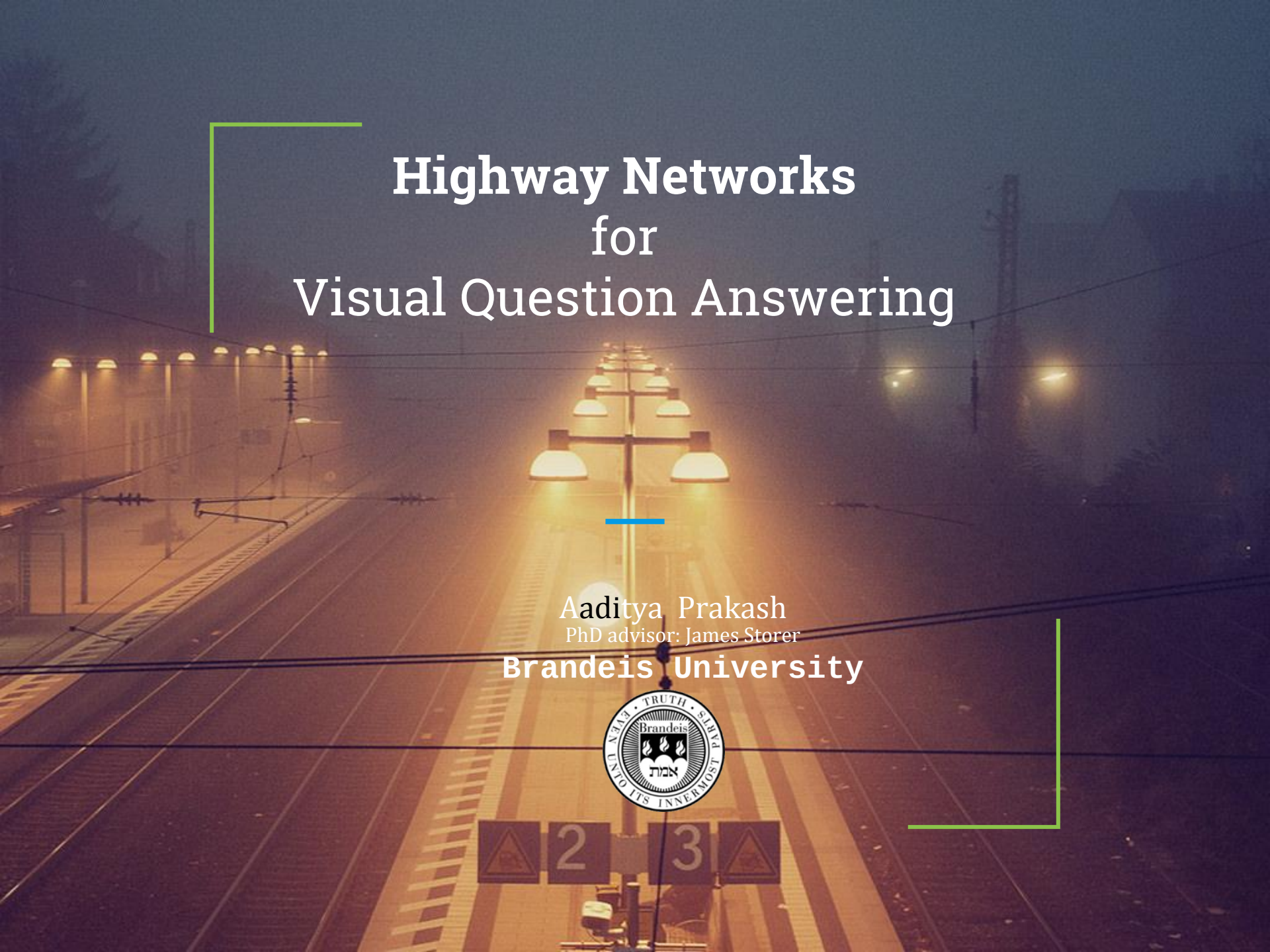
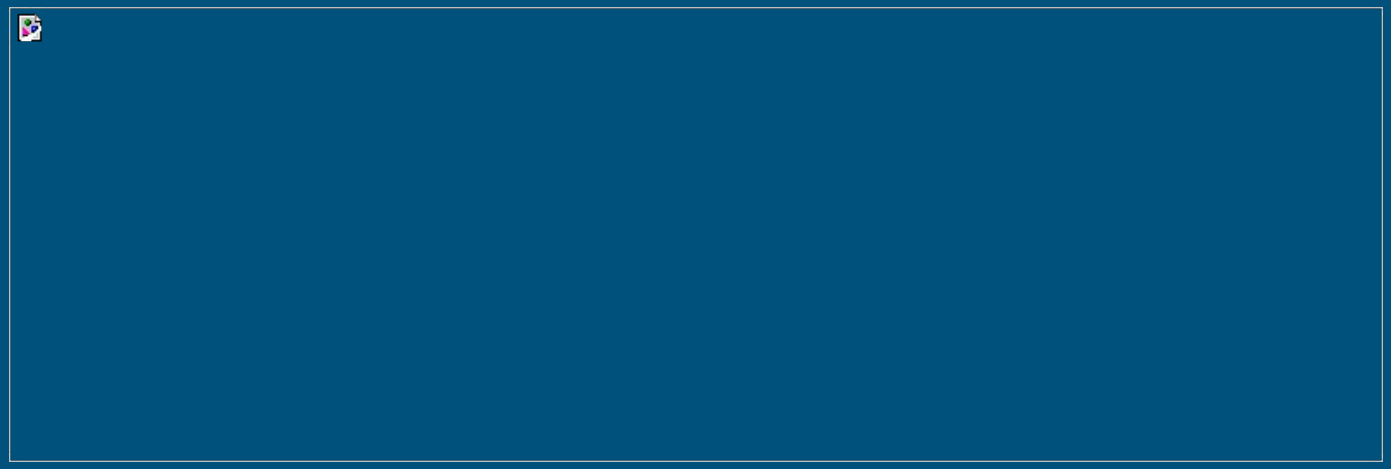


Highway Networks for Visual Question Answering

Aaditya Prakash
PhD advisor: James Storer
Brandeis University

Architecture

Perceptron



$$y = H(x, W) + \mathbf{b}$$

Highway Networks

$$y = H(x, W_H) \cdot T(x, W_T) + x \cdot C(x, W_C) + \mathbf{b}$$

$$T(x) = \sigma(W_T x + \mathbf{b_T})$$

$$C(x) = 1 - T(x)$$

Highway Networks

- Allows training very deep networks
  - Srivastava et al trained 50+ layers [1]

- Overcomes vanishing/exploding gradient issues by learning gating mechanism, like LSTM

- Includes ‘Transform’ gate (T) and ‘Carry’ gate (C)

  - Simple Perceptron

$$y = H(x, W) + \mathbf{b}$$

  - Highway Layer (MLP)

$$y = H(x, W_H) \cdot T(x, W_T) + x \cdot C(x, W_C) + \mathbf{b}$$

$$T(x) = \sigma(W_T x + \mathbf{b_T})$$

$$C(x) = 1 - T(x)$$

Multimodal Learning

VQA

Image

Question

Multimodal Learning
VQA

X_1 → Image

X_2 → Question

Multimodal Learning
VQA

X_1 → Image

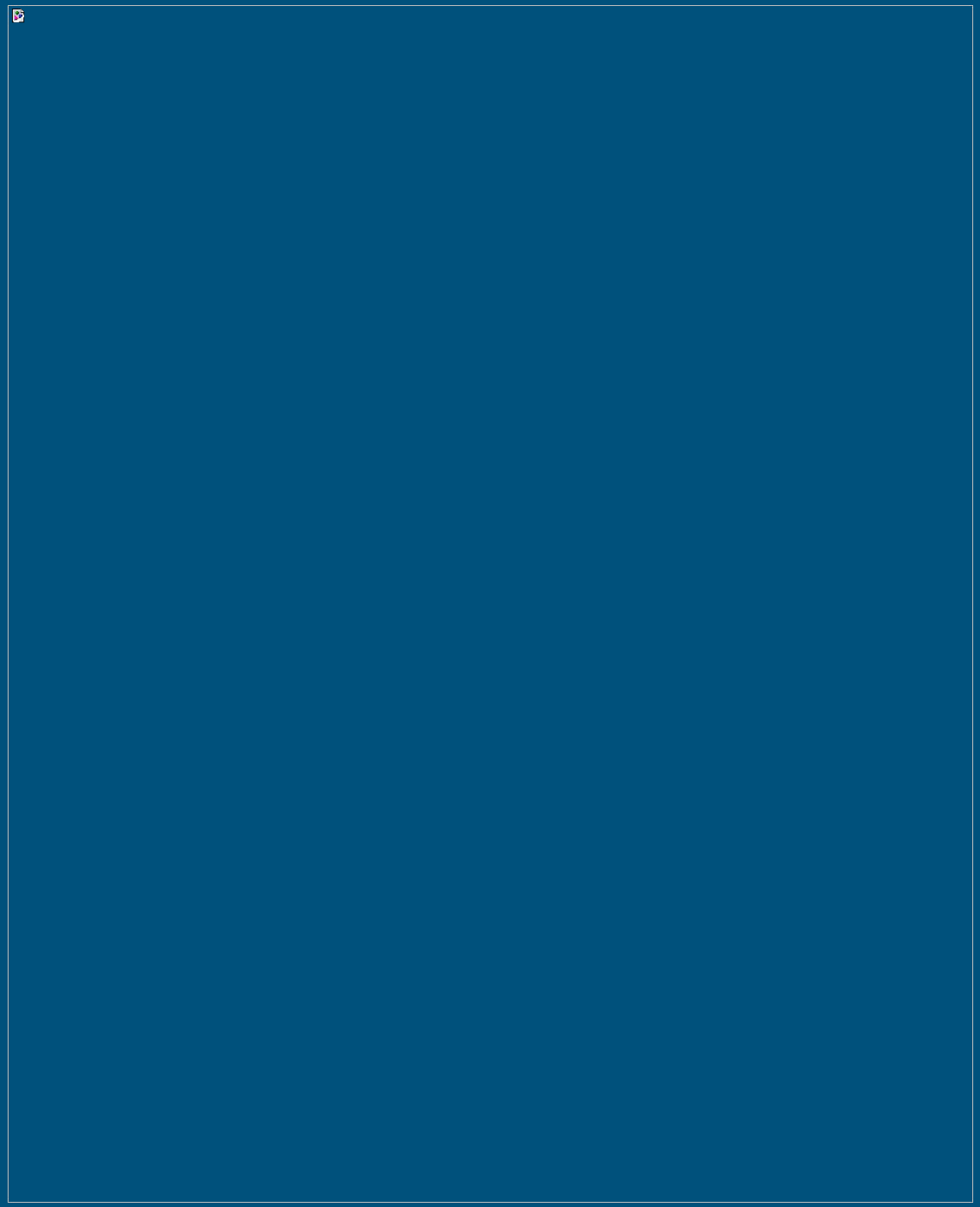
X_2 → Question

VISUAL QA

Highway Networks

After many layers later

And countless debugs



Note:

Figure does not mention the use following techniques :-

- Dropout and Batch-Normalization
- Image feature normalization
- Image augmentation before feature extraction
- Use of other word vectors like Word2Vec and ConceptNet

Results & Performance

Results from VQA Challenge

Real Open-Ended Test Standard 2015* (%)

Yes/No	Number	Other	Overall
82.11	**37.73**	**51.91**	**62.88**

Real Multiple choice Test Standard 2015 (%)

Yes/No	Number	Other	Overall
81.95	**38.56**	**56.4**	**65.07**

- Five model ensemble
  - Model 1 - VGGNet + 98% SF + Glove
  - Model 2 - VGGNet + 95% SF + Word2Vec
  - Model 3 - ResNet + 98% SF + Glove
  - Model 4 - ResNet + 98% SF + ConceptNet Numberbatch
  - Model 5 - ResNet + 95% SF + Word2Vec

(SF = Statistical Filtering)

- 10 Crop image inference ensembled into one answer
- SF - Statistical Filtering : restrict the answer to some percentage of answers within that question type
- Trained on train2014 + val2014 + finetuned on results from earlier model from test2015 [3]
- No SF for Real Multiple Choice (this might have been a bad idea)

Comparison of Accuracy over depth

VGGNet (4096 features)*

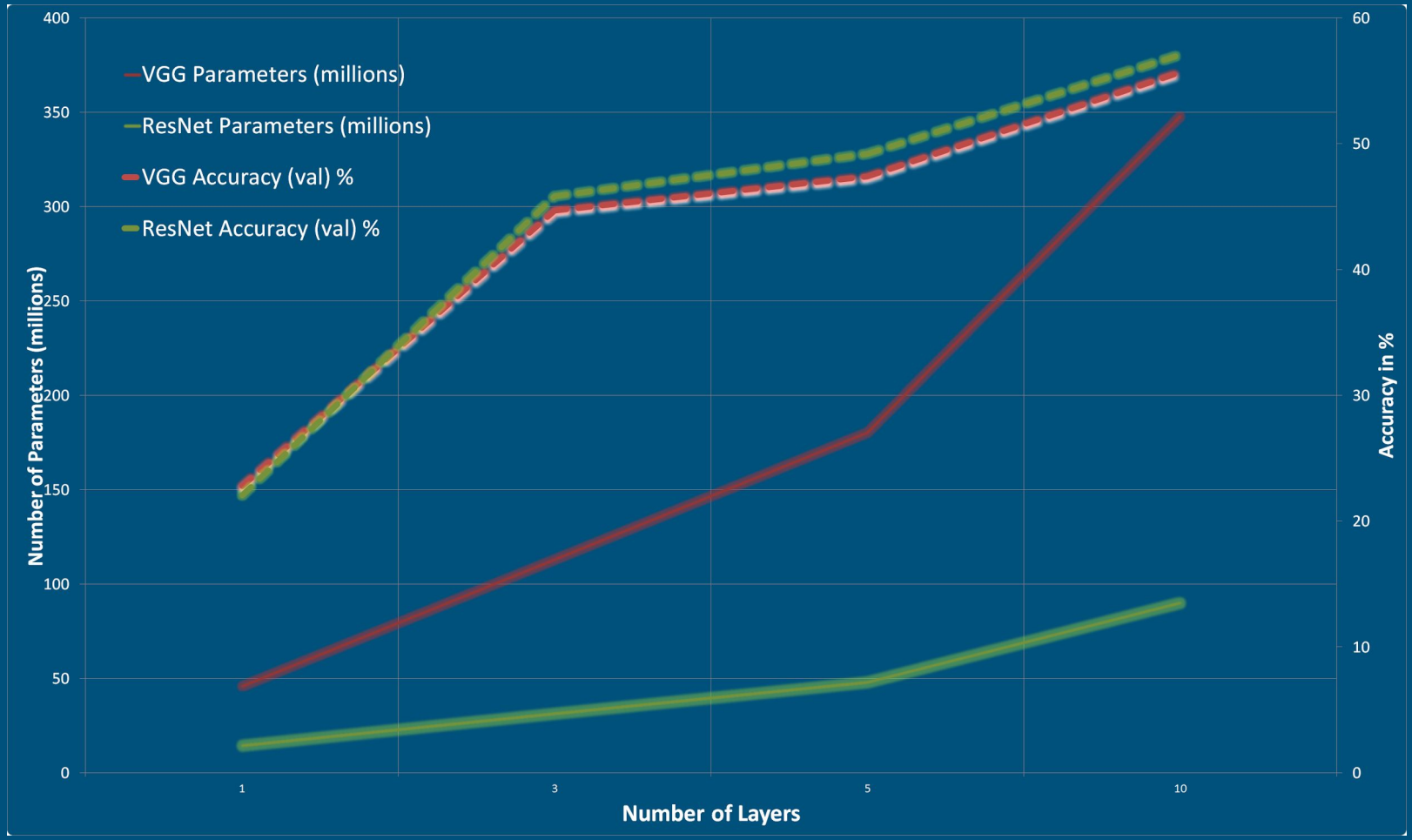
# Layers	Parameters (millions)	Accuracy (val)
1	46.052	22.83
3	113.177	44.7
5	180.302	47.4
10	348.115	55.7

ResNet (2048 features)*

# Layers	Parameters (millions)	Accuracy (val) %
1	14.638	22.1
3	31.423	45.85
5	48.208	49.21
10	90.172	57.1

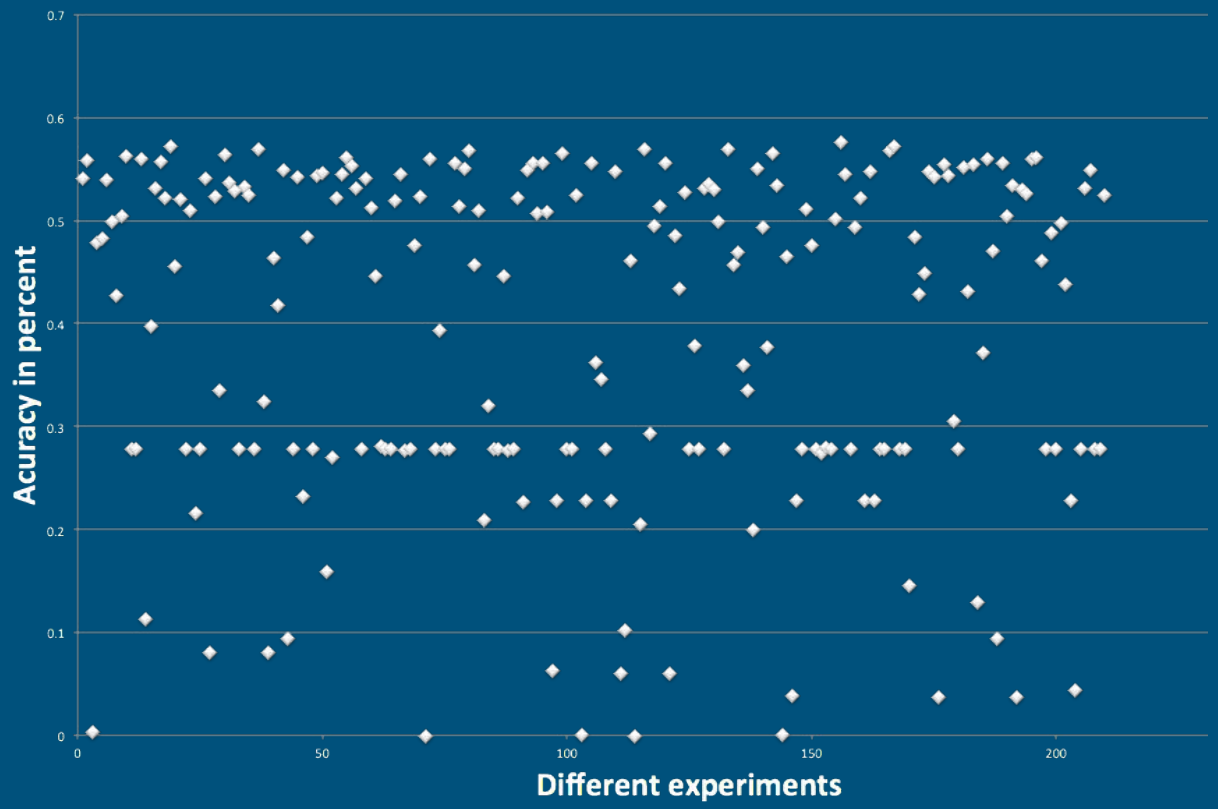
* Trained on train2014 and tested on val2014
* Single model (no ensembling), No Statistical filtering

Comparison of accuracy & parameters over depth



* Trained on train2014 and tested on val2014
* Single model (no ensembling), No Statistical filtering
* Real Open-Ended only

Hyper Parameter Search



Parameters

- Learning Rate
- Number of output (softmax)
- Initialization
  - Uniform
  - Xavier
  - Kaiming
  - heuristic
- Activation (tanh/relu/prelu)
- Num highway layers (1,2,3,4,6,10)
- Bias (Carry & Transfer)
- Decay factor
- Epoch at which to change optimizer

*Trained on train2014 and tested on val2014, ResNet
*Single model (no ensembling), No Statistical filtering (SF)
* Real OpenEnded only

References

[1] Srivastava, Rupesh Kumar, Klaus Greff, and Jürgen Schmidhuber. "Highway networks." arXiv preprint arXiv:1505.00387 (2015).

[2] Antol, Stanislaw, et al. "Vqa: Visual question answering." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[3] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. “Distilling the knowledge in a neural network.” arXiv preprint arXiv:1503.02531 (2015).

ANY QUESTIONS?

Thanks!

My thanks to -

- VQA Team for the challenge
- Aishwarya Agrawal for blazing fast replies to all my queries
- James Storer, my PhD advisor.
- NVIDIA for gifting us a Titan X.
- Following people from whose code I learned -
  - Yoon Kim @yoonkim (HarvardNLP)
  - Jin-Hwa Kim @jnhwkim (Element-Research)
  - Jainsen Lu @jiasenlu (VQA_LSTM_CNN)
  - François Chollet @fchollet (Keras)
  - Hyeonwoo Noh @HyeonwooNoh (DPPNet)
  - Bolei Zhou @metalbubble (VQAbaseline)
  - Matthew Honnibal @honnibal (Spacy)