“Darling, what am I looking at?”

CHARLIE LEW

Motivation & Objective

Motivation & Objective

Motivation & Objective

Objective

Create an automatic text-to-speech caption generator for the visually impaired via Deep Learning

Data & Tools

Data & Tools

Dataset: Flickr 8,000 annotated images

- *Annotations created by Mechanical Turk*

Data & Tools

Dataset: Flickr 8,000 annotated images

- *Annotations created by Mechanical Turk*

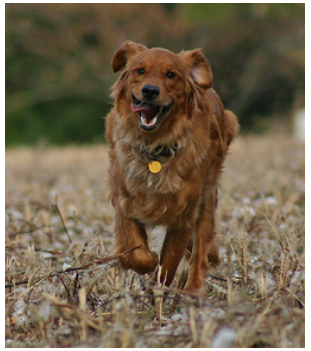
Tools

- *Python, Pandas, Scikit-Learn, Numpy*
- *Matplotlib, Seaborn*
- *TensorFlow & Keras*
- *Flask*

Methodology Pipeline

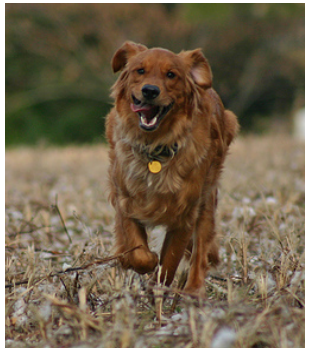
Methodology Pipeline

1. Image



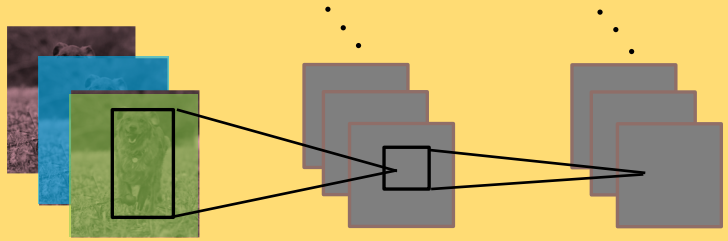
Methodology Pipeline

1. Image



2. Feature Extraction

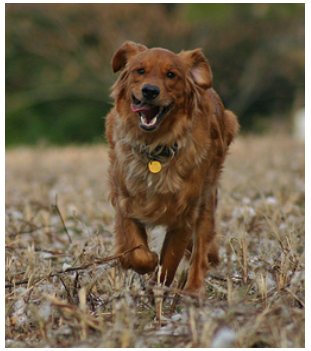
Convolutional Neural Network (CNN)
VGG16



Vocabulary from Flickr dataset
(~4,000 words)

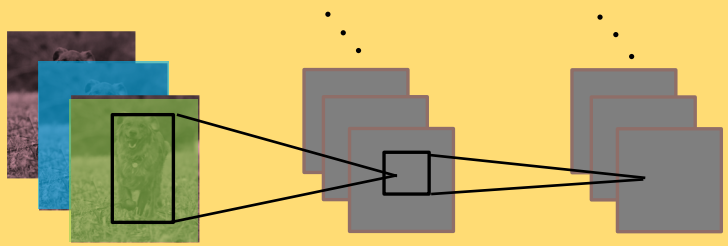
Methodology Pipeline

1. Image



2. Feature Extraction

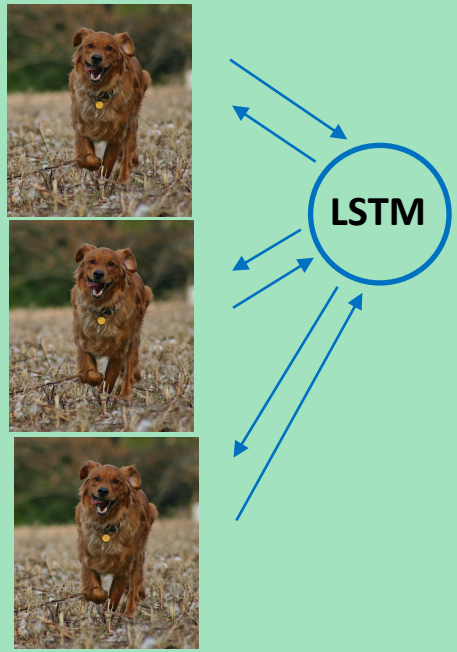
Convolutional Neural Network (CNN)
VGG16



Vocabulary from Flickr dataset
(~4,000 words)

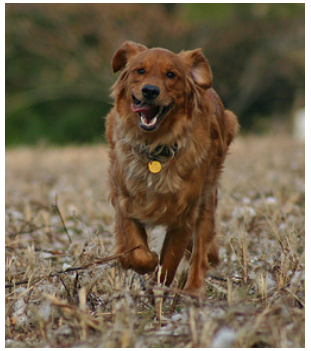
3. Decoder

Recurrent Neural
Network (RNN)
with LSTM



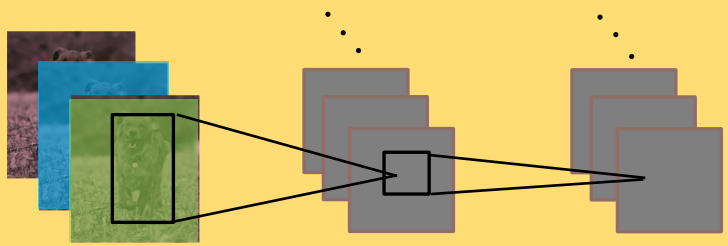
Methodology Pipeline

1. Image



2. Feature Extraction

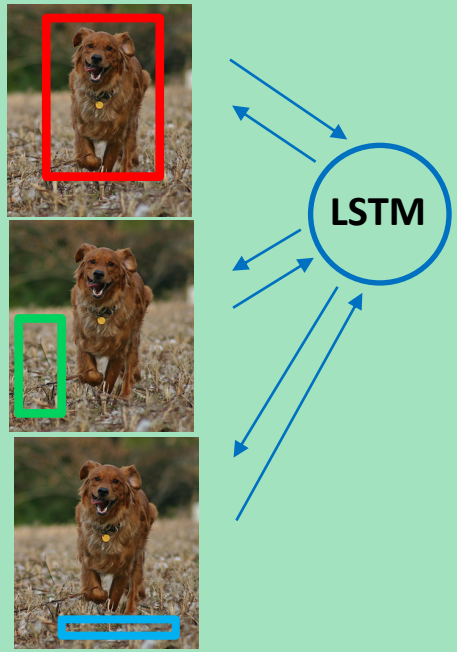
Convolutional Neural Network (CNN)
VGG16



Vocabulary from Flickr dataset
(~4,000 words)

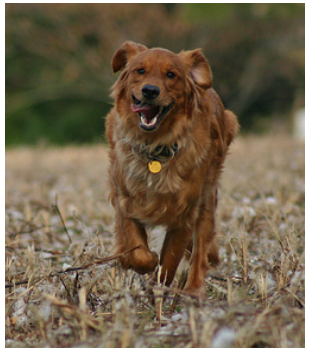
3. Decoder

Recurrent Neural
Network (RNN)
with LSTM



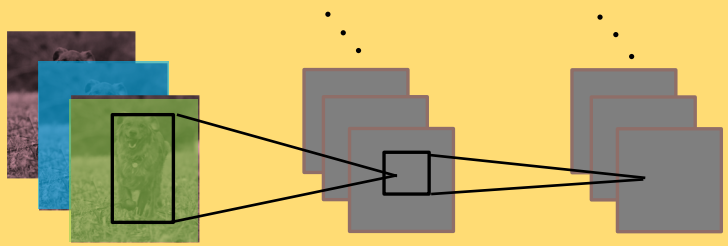
Methodology Pipeline

1. Image



2. Feature Extraction

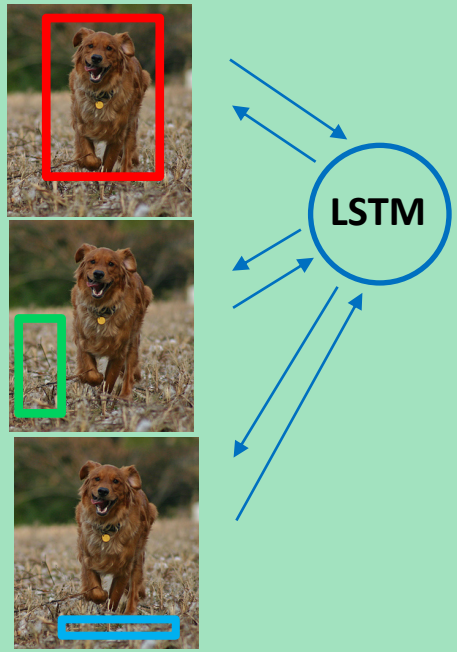
Convolutional Neural Network (CNN)
VGG16



Vocabulary from Flickr dataset
(~4,000 words)

3. Decoder

Recurrent Neural
Network (RNN)
with LSTM

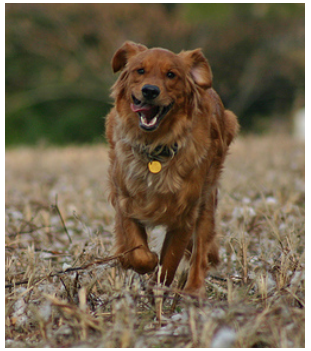


4. Caption!

Predicted:
"Brown dog is running in the grass"

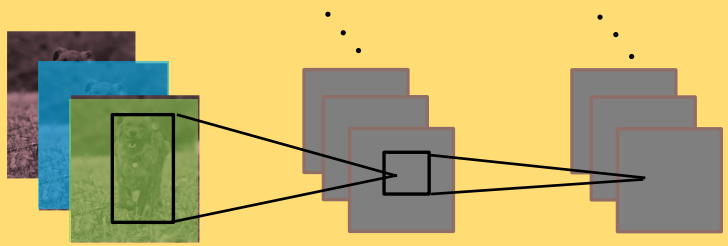
Methodology Pipeline

1. Image



2. Feature Extraction

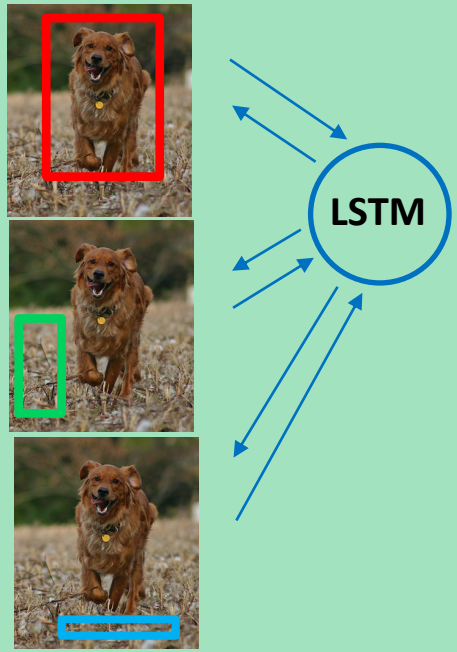
Convolutional Neural Network (CNN)
VGG16



Vocabulary from Flickr dataset
(~4,000 words)

3. Decoder

Recurrent Neural
Network (RNN)
with LSTM

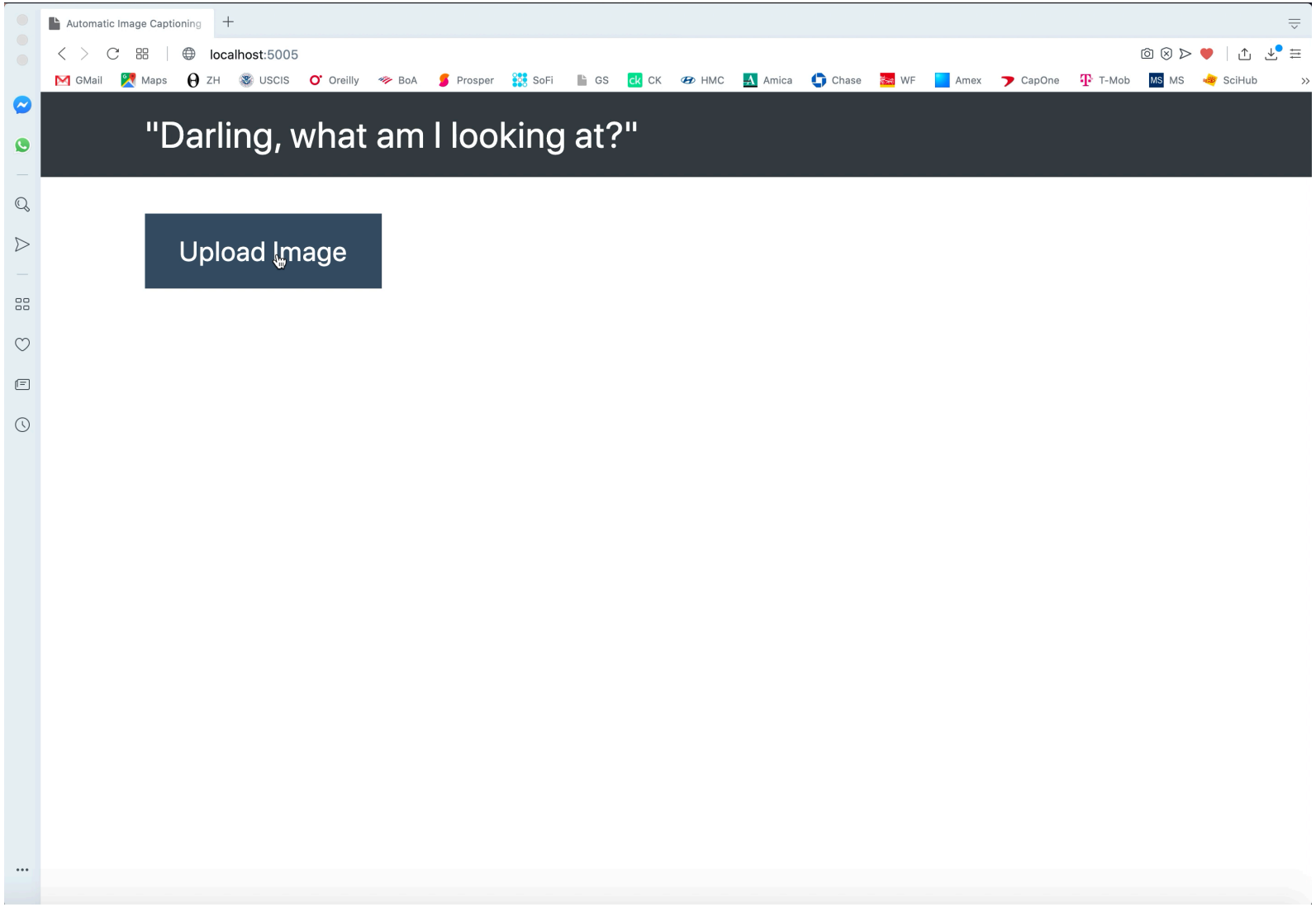


4. Caption!

Predicted:
"Brown dog is running in the grass"

True:
"Brown dog is running through the field"

Flask App Demonstration



Future Work

Further improve RNN & Metrics

- *Hyper-parameter tuning, Attention Model, BLEU*

Need more data!

- *Flickr: 30,000*
- *MSCOCO: 100,000*
- *Google: Several million*

Extend to paintings, videos, etc.

Thank You

Phoi-Tack (Charlie) Lew

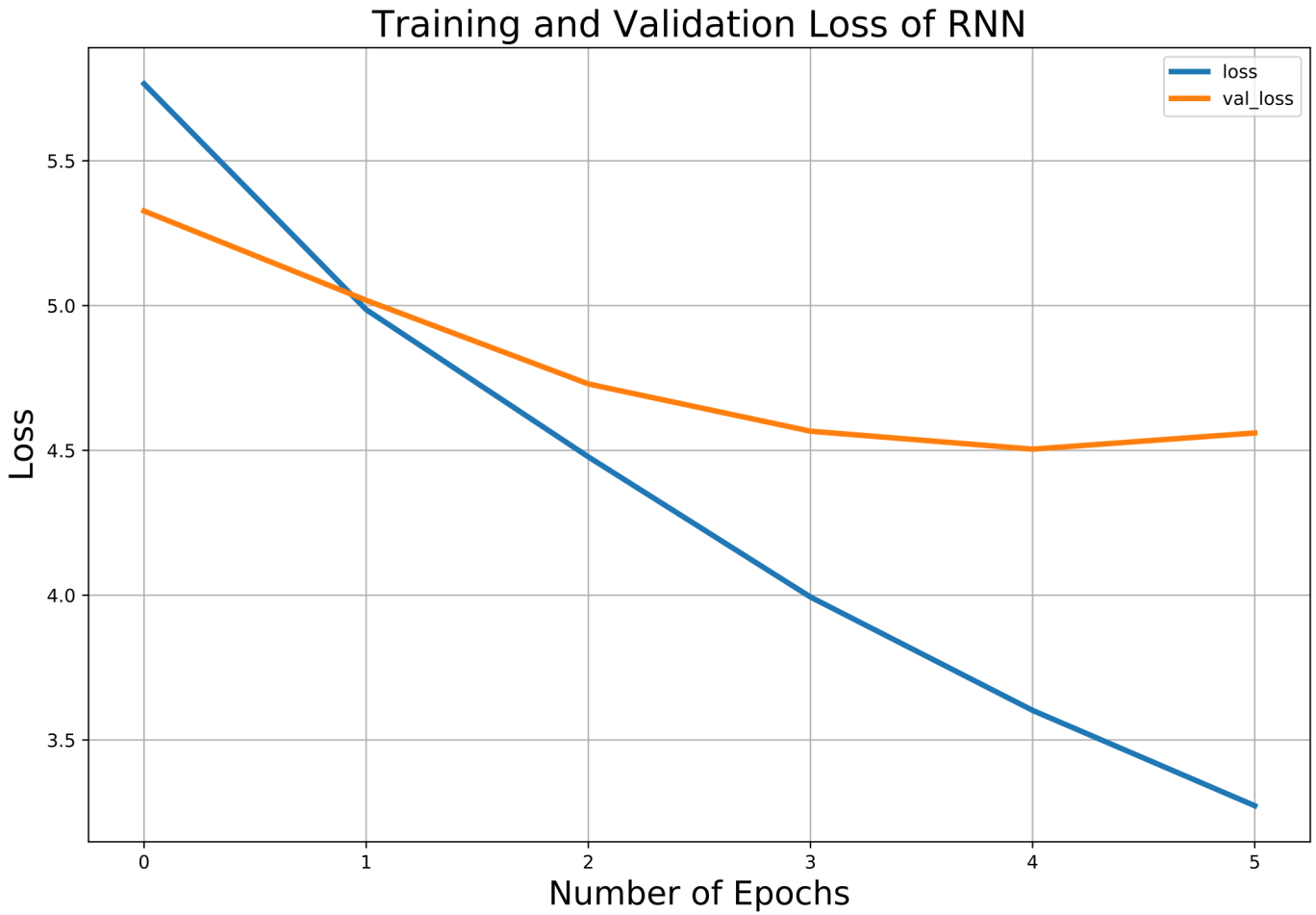
phoitack@gmail.com

/in/charlielew

Phoi-Tack (Charlie) Lew

Data Scientist | Aerospace Engineer

Appendix


Training and Validation Loss of RNN
loss
val_loss
Loss
5.5
5.0
4.5
4.0
3.5
0
1
2
3
4
5
Number of Epochs