Deep Fiction

Every picture has a story to tell

Project Members

Alaap Sivaram, Maitri Shroff and Shreya Jain

Deep Learning for Computer Vision

Motivation

- Creating short stories about an image is a difficult task that many humans struggle with too
- A more difficult task is to generate a story around an image given a particular genre
- Help authors overcome their “Writer’s block”

Overview

Image Captioning

The first task is to describe the salient features of the image in natural language

Learn the book style

Training an encoder decoder to learn the style of the book corpus

Generate Stories

The thought of the caption is preserved and stylized to generate stories of the book style

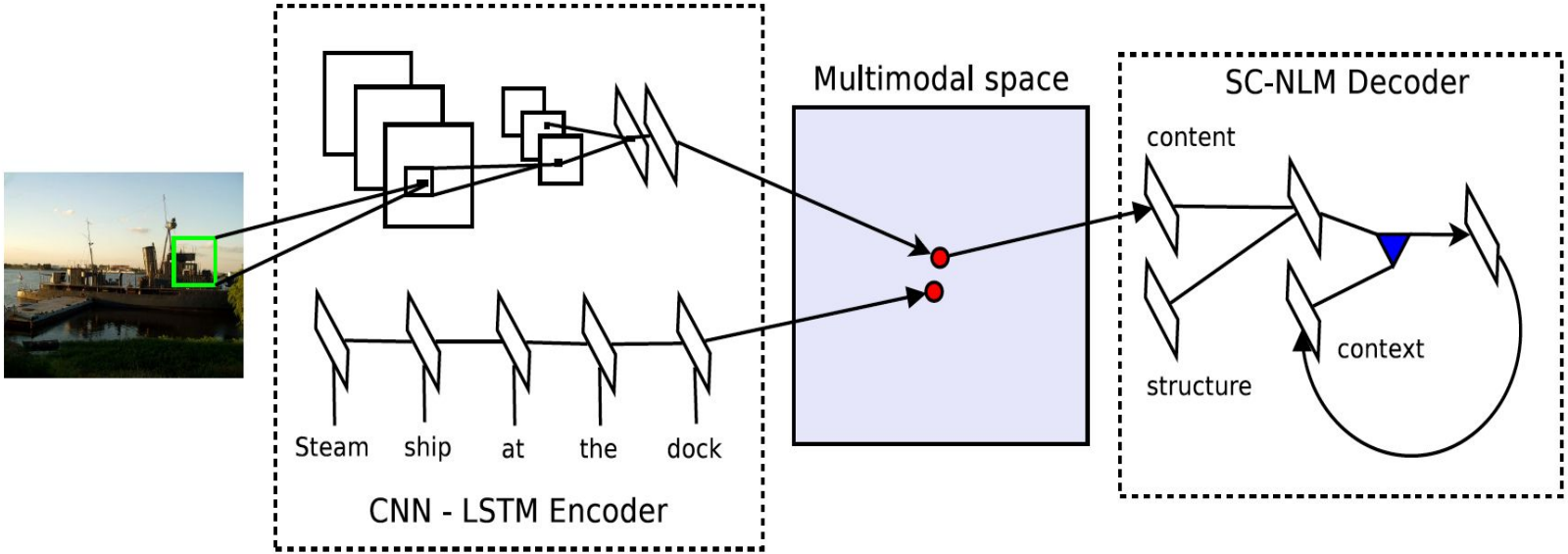
Baseline Model

Generating the neural story from the captions

Datasets

- Image captioning module - MS COCO dataset
- Story generation RNN - Romance novels from BookCorpus dataset
- Children’s book dataset (Q&A) - Facebook research

Image Sentence Embedding Architecture



RNN Encoder Decoder

- Encoding sentences into fixed-length vectors (skip thoughts)
- Encoder - RNN with GRU activation
- Decoder - RNN with conditional GRU
- Hidden state of encoder fed to 2 decoders (to predict the preceding and subsequent sentences)

Mapping Captions to Book Passage Vectors

- To map the captions ‘x’ to the book passage vector F(x):

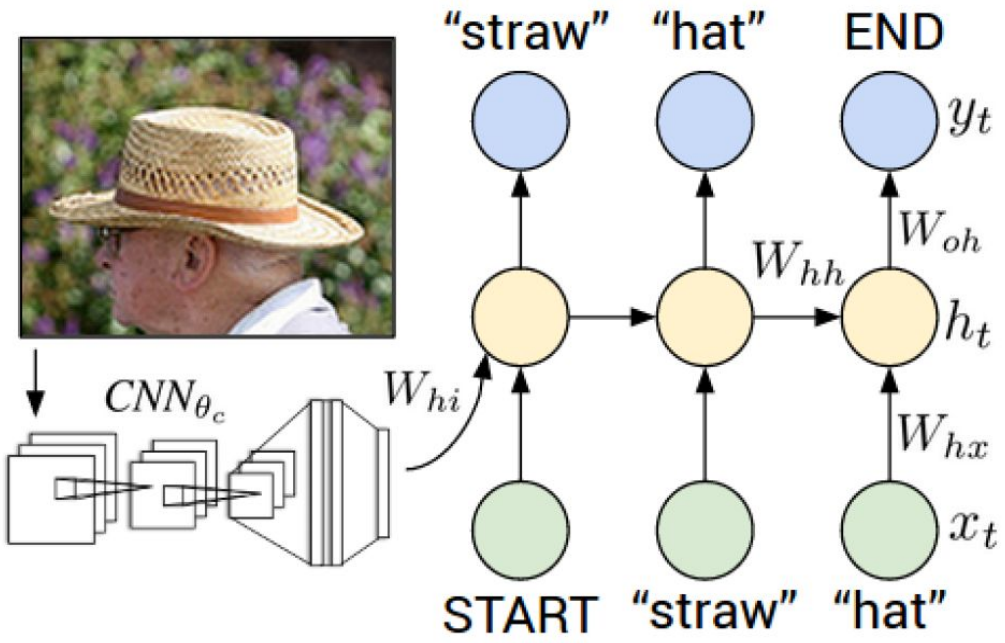
  $$F(x) = x - c + b,$$

  where c → Caption Style, b → Book style vector

- Here c = mean of skip thought vectors for image captioning and b = mean of skip thought vectors of romance novel passages

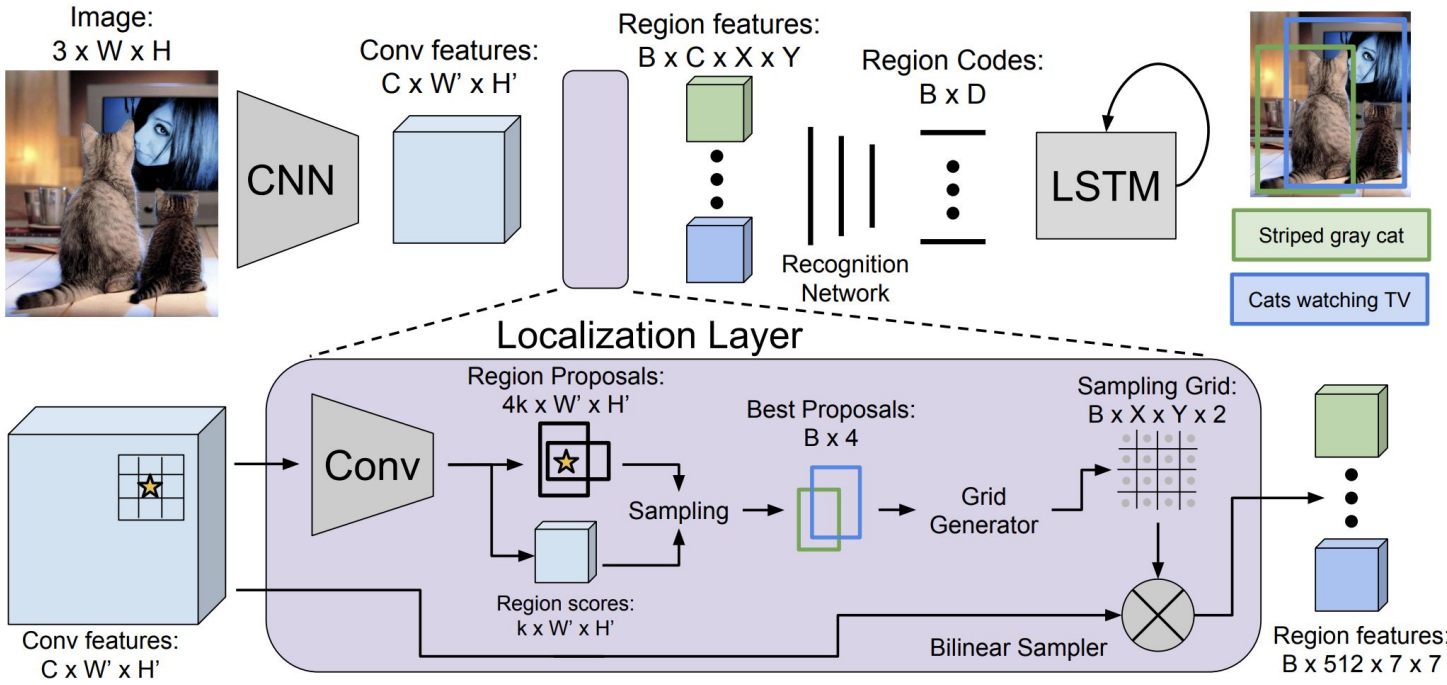

Improving
Captions

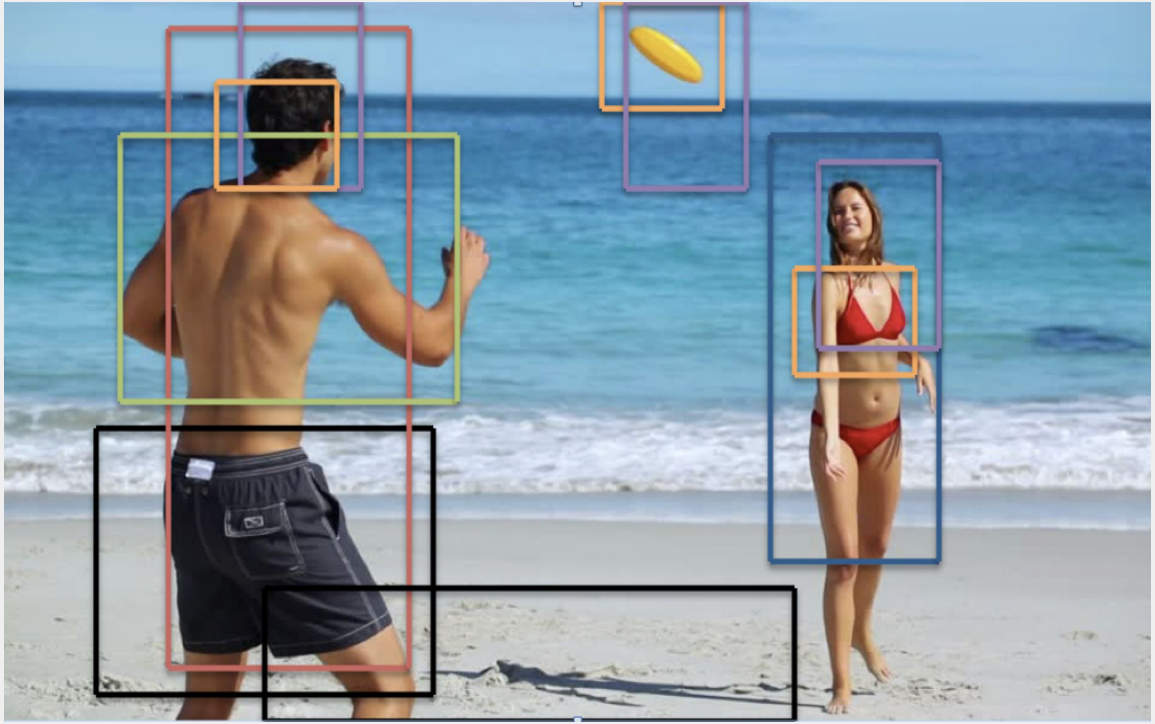
Providing Quality Captions



“Man and woman are playing frisbee on the beach”

Captioning Regions





'black shorts on man'
‘man standing on the beach'
'woman in a bikini'
'man with no shirt'
'man has short hair'
'the frisbee is yellow'
'woman wearing a red top'
'a volleyball'
'woman with long hair'
'the man is surfing'
'man with short hair'

Our Contribution

Original Model took 100 captions

Improving the caption quality with two models

Identified optimal number of captions

Training on another genre

Examples

Generating the neural story from the captions

Baseline Model

Everyone had a few friends at the beach, and for the first time in my life she felt the urge to hug Nate. He had no intention of telling him what to do. She was acting like an injured kid, and now she rode to be the beach. The beach was on the beach, so it had nothing to do with the tattoo artist on his arm. In fact, it felt as if he'd just pulled her into his arms and carried her out of the camp, leaving Chelsea and her friends and friends all over the beach on horses. Such a miracle. I could hardly hear what happened to Wes, since the beach girls were alive.

Model with better captions

Everyone had a few friends at the beach, and for the first time in my life, She felt like she was going to jump up and fall asleep on the beach. She had no idea what he wanted to do with her, so he didn't have a choice. She shook her head at Josh, his arm still wrapped around her waist, and he pulled her into his arms . There was nothing to talk about, so I’m pretty sure it won’t be the same. The only person on the beach , I wished I’d joined them in the sand.

Model with details

She was a girl on the beach, and she laughed at my touch. She felt the tension in her body, as if he had a few days to get back to the beach. She had not seen Nate for a while, so it would be the perfect time for she to discuss what happened with the other team. She loved Nate, and she had no intention of going back to New York. She felt like a fish in the sand, so he carried his arm around her waist and pulled out his T-shirt. The girl looked beautiful playing beach volleyball. She could hardly recall what happened between them on the beach.

Evaluation

BLEU for Image Captioning

	B@1	B@4
Captions from 1	**0.423**	**0.227**
Captions from 2	**0.515**	**0.253**

Mistakes on Average

Story 1

12.96

Story 2

13.42

Story 3

9.73

Future Scope

- Build an end to end image captioning model
- Train on more genres to give author more flavours for the same image
- Increase length of passages to scale to larger stories
- Work on multiple images and build a longer continuous story
- Extend previous point to videos by capturing frames


“
Thank you!