

Image Captioning of Earth Observational Imagery

An MDS-MDA Joint Capstone Project



Dora Qian, Fanli Zhou, James Huang, Mike Chen

MDA logo

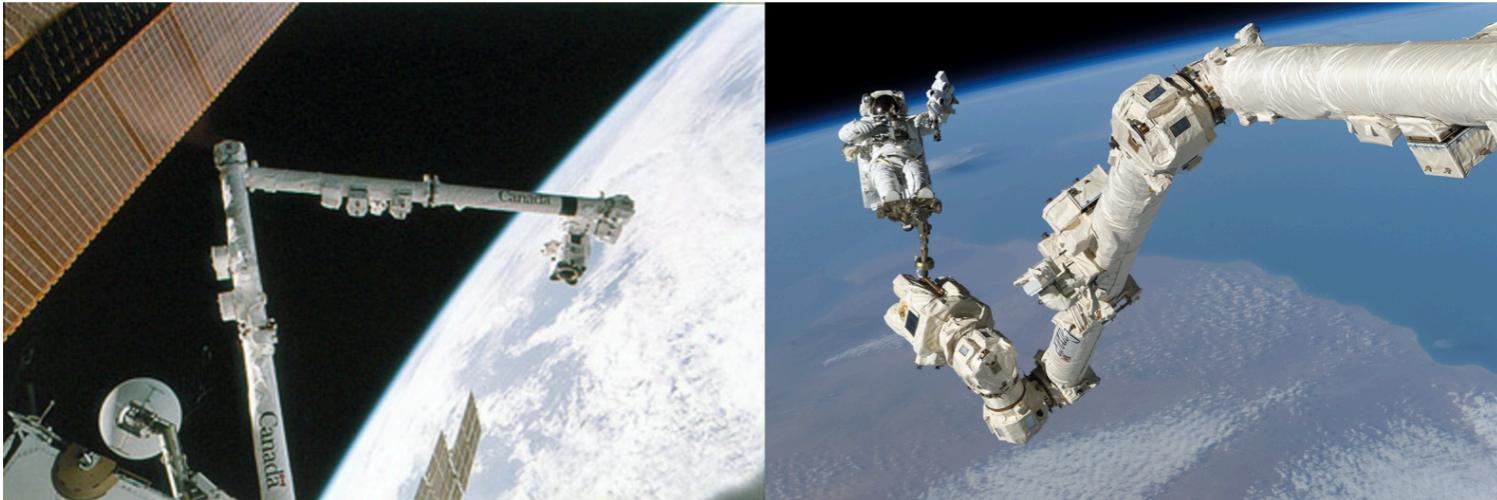




MDA

A Canadian Aerospace Company

- Developed Canadarm and Canadarm-2 on the ISS



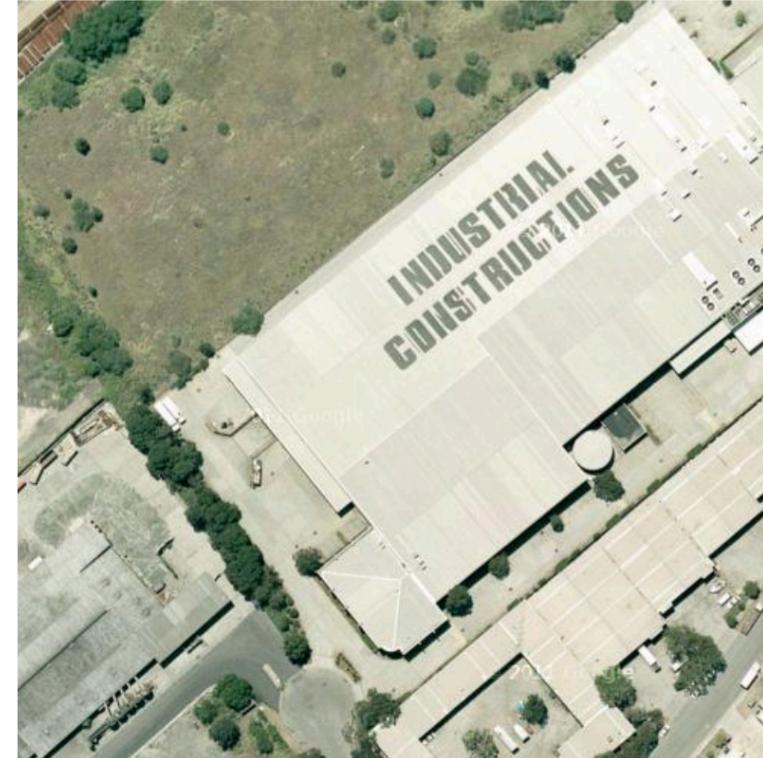
Use, duplication or disclosure of this document or any of the information or images contained herein is subject to the restrictions on the title page of this document.
COPYRIGHT ©2020 MacDonald, Dettwiler and Associates Inc. (MDA).

MDA™



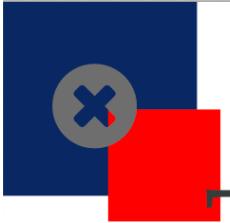
MDA

- Access to a vast database of satellite images



Sources: Image adapted from Qu, B. et al. (2016) [1].





The problem

- These images are uncaptioned
 - Without captions, these images are difficult and computationally costly to work with
- Technology of captioning satellite images is less mature than "traditional" photographs
- Due to the nature of these photographs, the model cannot be effectively trained on other types of images.
 - Limited existing resources to train the model





Image Captioning: Motivation and Purpose

- Associating an image with a caption makes it much more accessible:
 - Tag and sort images based on content
 - Return search queries
 - Evaluate image similarity
 - Downstream applications





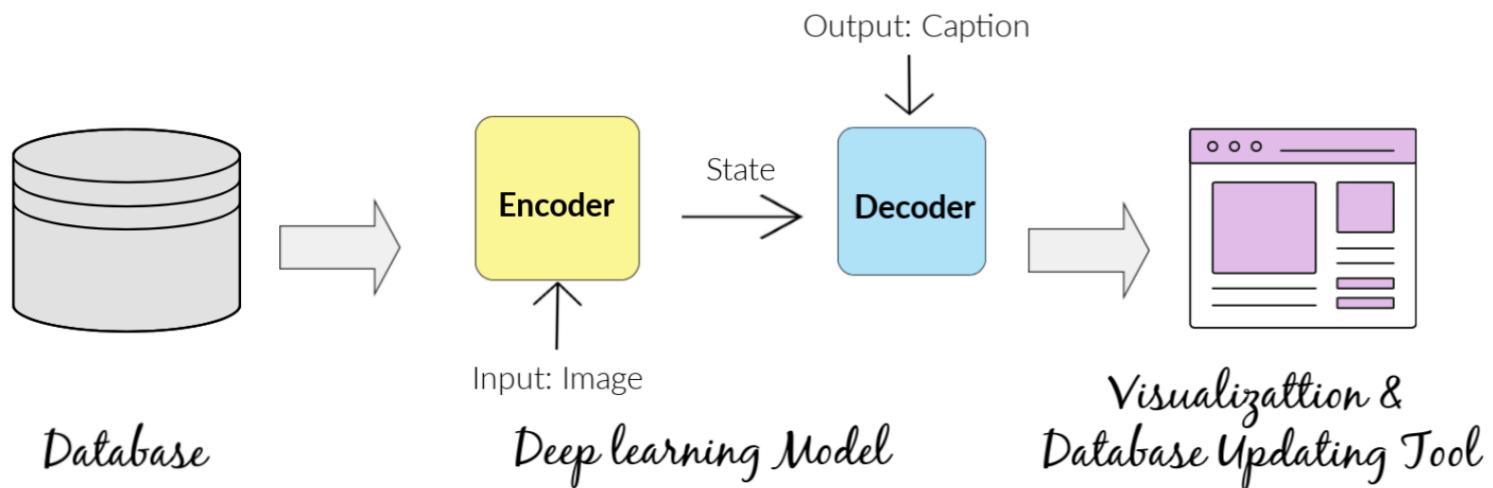
Training the Model

- MDA images are uncaptioned
 - Train, validate, and test on public, captioned satellite images
 - Several different datasets
 - Assess cross dataset performance
 - Final manual evaluation on uncaptioned MDA images



Final Data Product

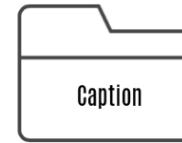
- End-to-End image captioning pipeline
- 3 independent modules





Final Data Product: Database

- Non-relational database
- Stores both human-annotated and machine-generated image-caption pairs



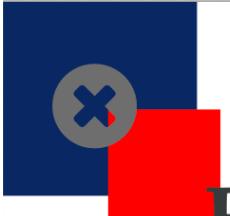
many planes are parked next to a long building in an airport



N-gram similarity,
BLEU,
etc...

Sources: [RSICD_optimal](#)





Final Data Product: Deep Learning Model

- Load images from database
- Model training
- Model prediction
- Easy to update



Sources: [AWS logo](#), [Pytorch logo](#)





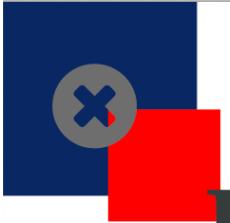
Final Data Product: Visualization & Database Updating Tool

- Random selected images from database
- Self upload images outside database
- Standardize the image, make prediction and save to database

The screenshot shows a web-based application for remote sensing image captioning. At the top, the title "Remote Sensing Image Captioning Visualization & Database Updating Tool" is displayed. Below the title, there are two input options: "Option 1: Number of images" (with a dropdown menu set to "ALL" and a button "Click here to generate") and "Option 2: Upload image" (with fields for "Path: url", "Generate", and "Save to Database"). The main area is titled "Results" and displays four satellite images labeled "Image 1", "Image 2", "Image 3", and "Image 4". Each image has a caption below it: "Caption 1", "Caption 2", "Caption 3", and "Caption 4". Below each caption is an "Evaluation score" box: "Evaluation score 1", "Evaluation score 2", "Evaluation score 3", and "Evaluation score 4".

Sources: [RSICD_optimal](#)





Data Description

There are three labeled datasets:

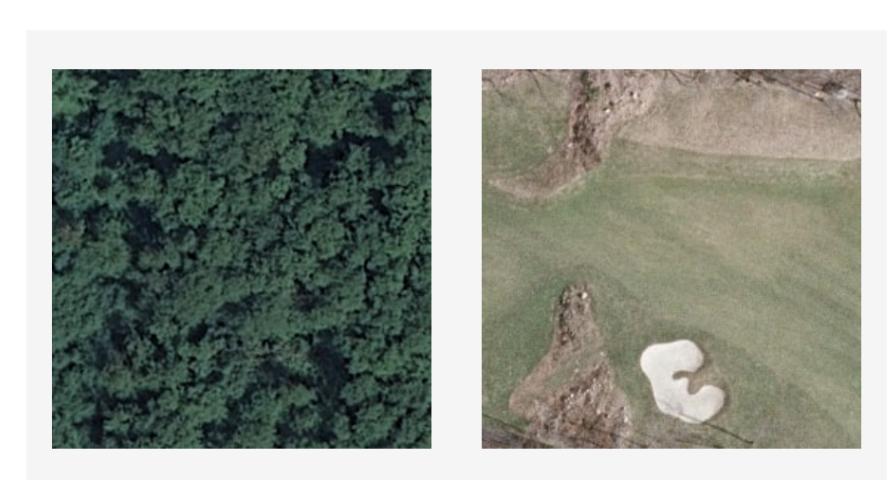
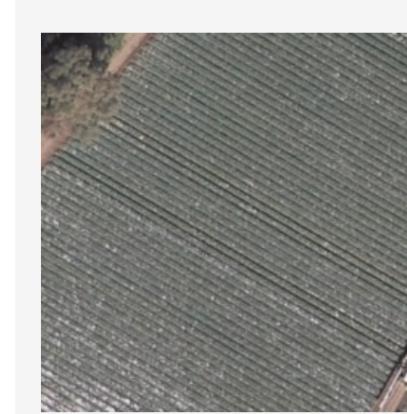
- UCM_captions
- Sydney_captions
- RSICD (Remote Sensing Imaging Captioning Dataset)





UCM_captions

- 21 Different Classes of Images
- 2100 Different Images
- 256 X 256 Pixels
- .tif Format



Sources: [UC Merced Land Use Dataset](#)





Sydney_captions

- 7 Different Classes of Images
- 613 Different Images
- 500 X 500 Pixels
- .tif Format



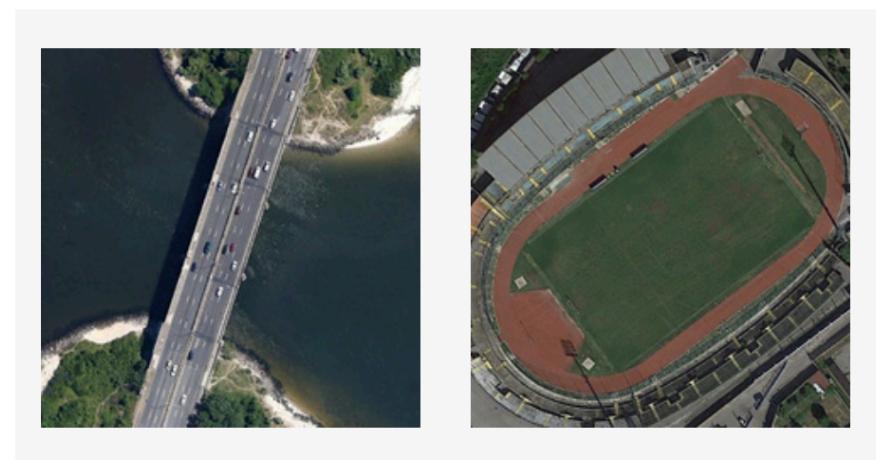
Sources: Image adapted from Qu, B. et al. (2016) [1].





RSICD (Remote Sensing Imaging Captioning Dataset)

- 10,922 Different Images
- 224X 224 Pixels
- .jpg Format



Sources: [RSICD_optimal](#)





RSICD Caption Example



1. An old court is surrounded by white houses.
2. A playground is surrounded by many trees and long buildings.
3. A playground with basketball fields next to it is surrounded by many green trees and buildings.
4. Many green trees and several long buildings are around a playground.
5. This narrow, oval football field and closing basketball court, tennis court, parking lot together form this area, with plants wreathing it.



1. Four planes are stopped on the open space between the parking lot.
2. Four white planes are between two white buildings.
3. Some cars and two buildings are near four planes.
4. Four planes are parked next to two buildings on an airport.
5. Four white planes are between two white buildings.

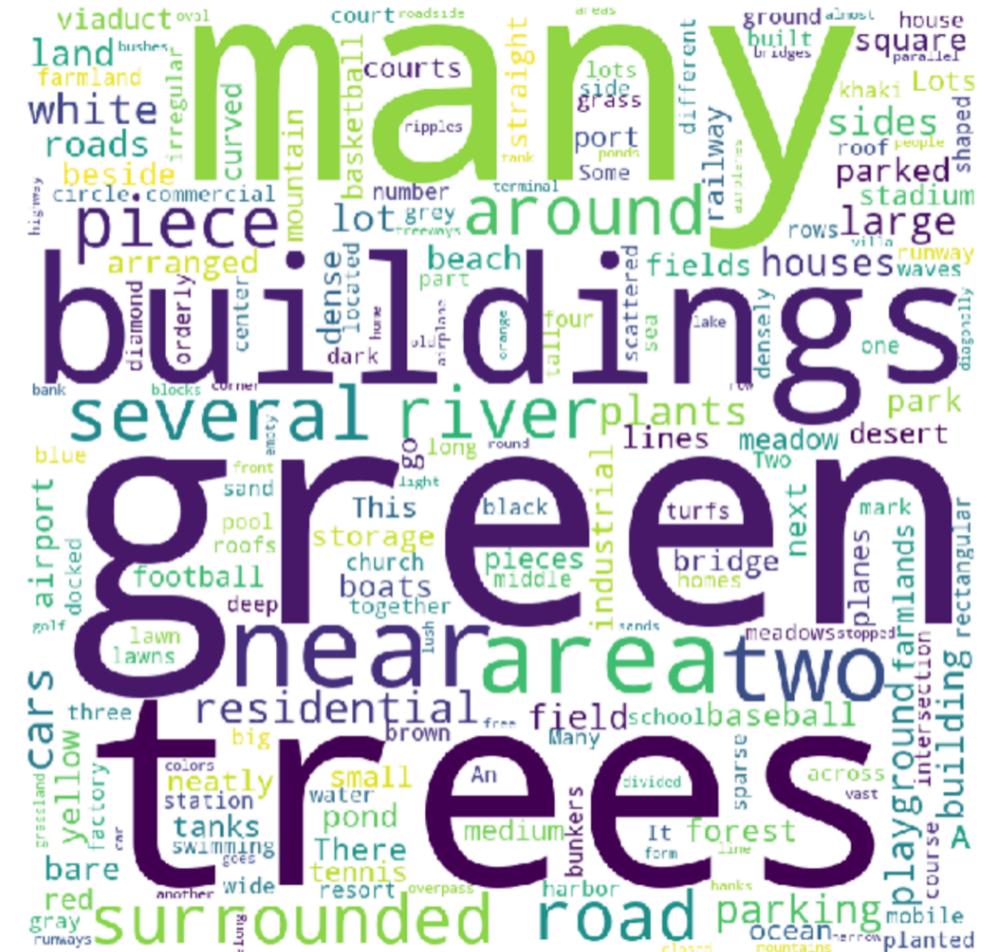
Image adapted from Lu, X. et al. (2018) [1].





Exploratory Data Analysis

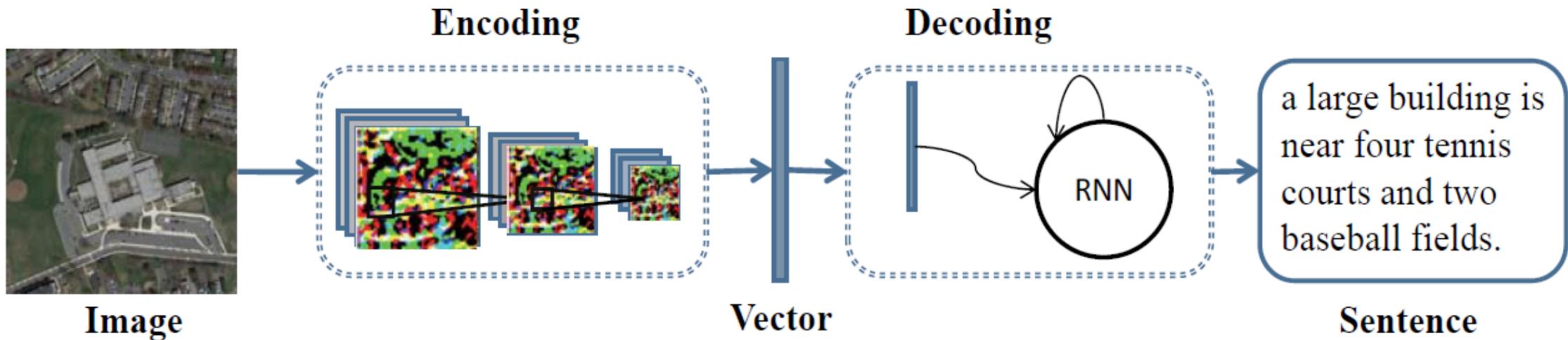
- train_valid/test = 80% / 20%
 - train/validation = 80% / 20%
 - Maximum length: 34 words
 - Minimum length: 2 words
 - Most common words





Data Science Techniques

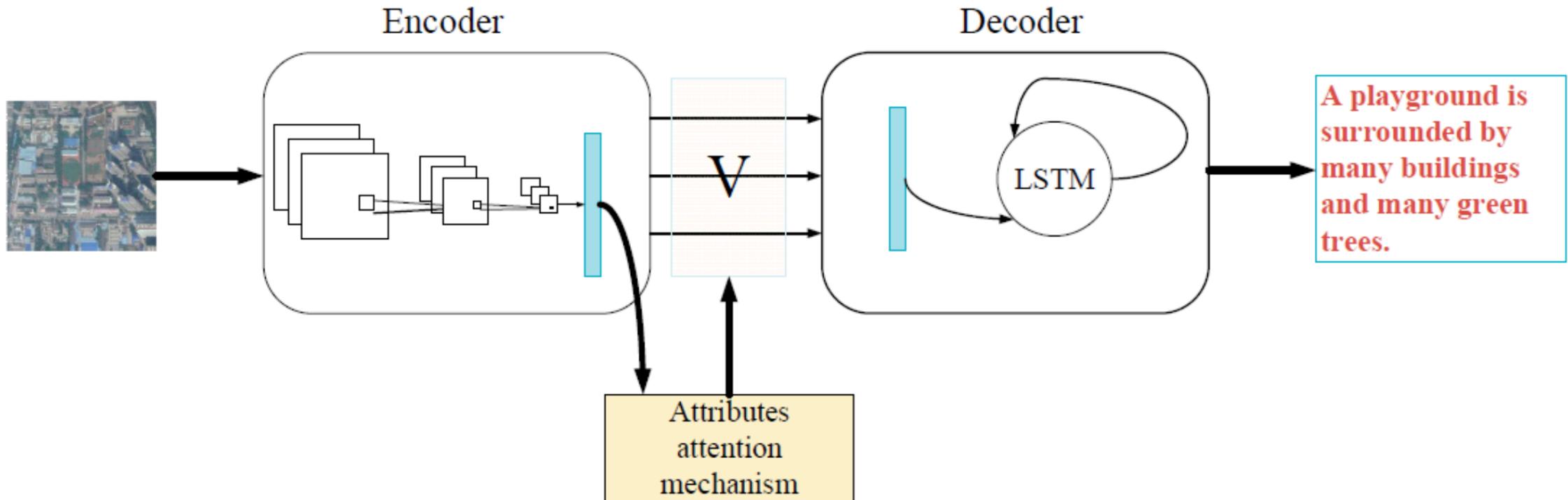
Baseline Model: CNN + RNN (LSTM)



Sources: Image adapted from Lu, X. et al. (2018) [2].



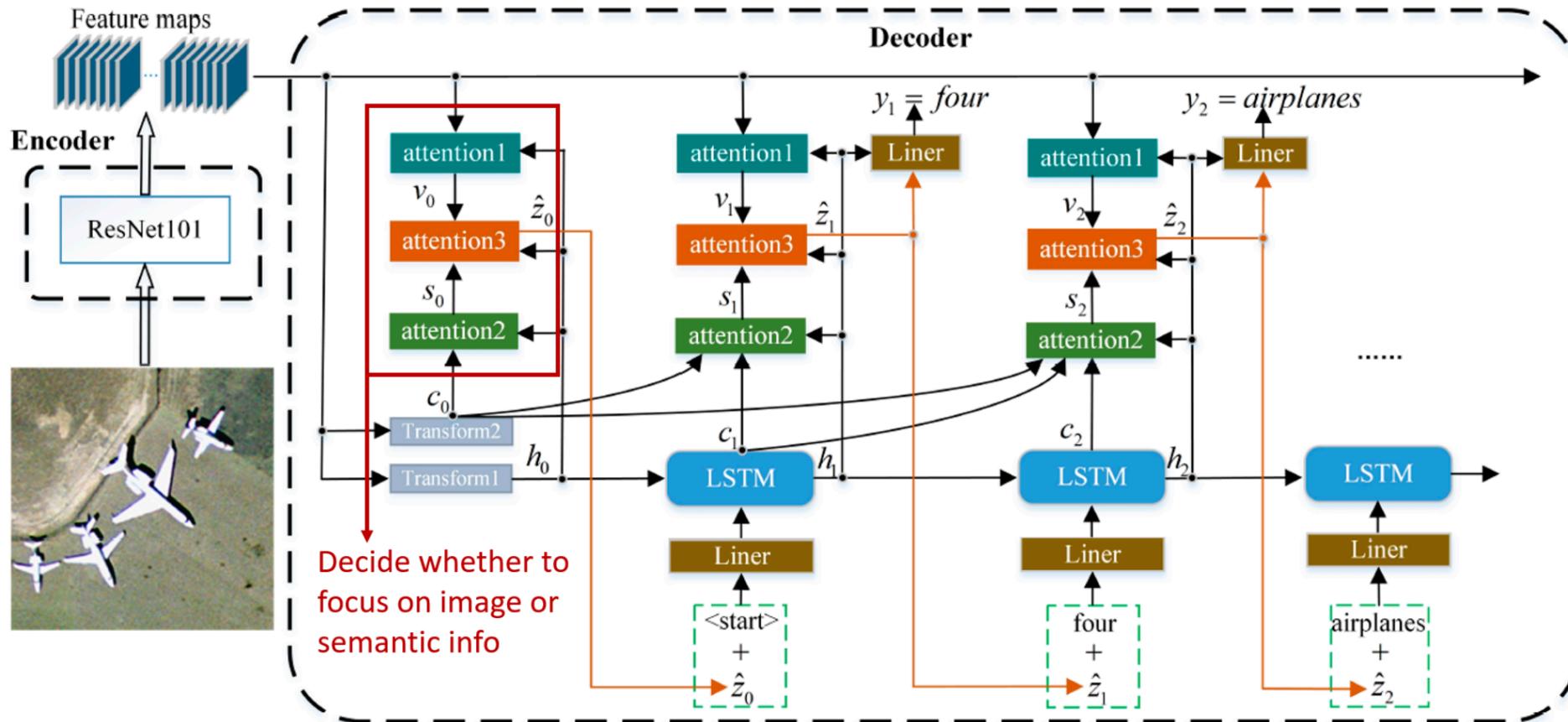
Model II: CNN + Attention + LSTM



Sources: Image adapted from Zhang, X. et al. (2019) [3].



Model III: CNN + multi-level Attention + LSTM



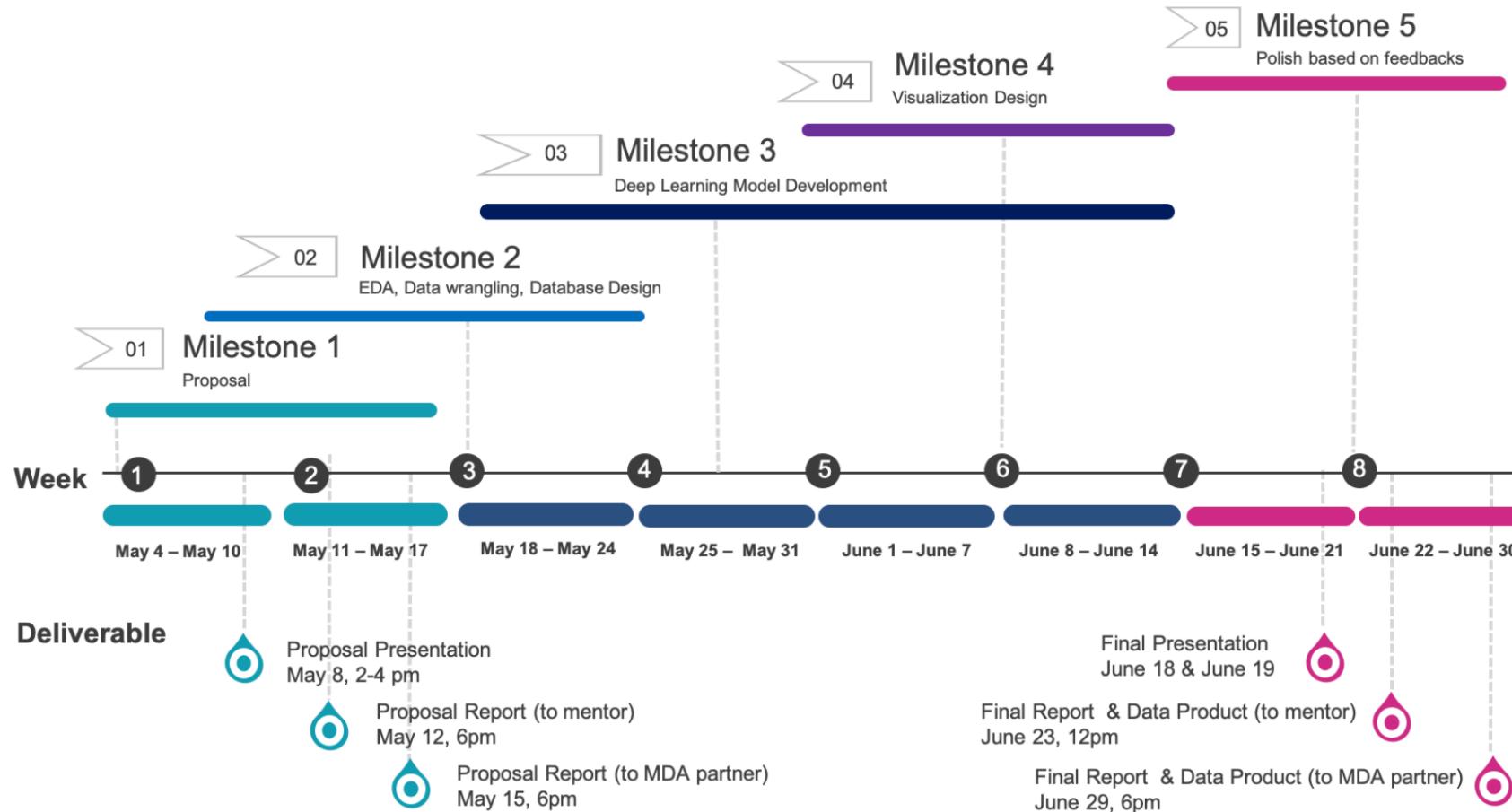
Sources: Image adapted from Li, Y. et al. (2020) [4].

Use, duplication or disclosure of this document or any of the information or images contained herein is subject to the restrictions on the title page of this document.

COPYRIGHT ©2020 MacDonald, Dettwiler and Associates Inc. (MDA).

MDA™

Timeline and Evaluation





Reference

1. B. Qu, X. Li, D. Tao, and X. Lu, "Deep semantic understanding of high resolution remote sensing image," International Conference on Computer, Information and Telecommunication Systems, pp. 124–128, 2016.
2. Lu, X.; Wang, B.; Zheng, X.; Li, X. Exploring models and data for remote sensing image caption generation. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 2183–2195.
3. Zhang, X.; Wang, X.; Tang, X.; Zhou, H.; Li, C. Description Generation for Remote Sensing Images Using Attribute Attention Mechanism. *Remote Sens.* 2019, 11, 612.
4. Li, Y.; Fang, S.; Jiao, L.; Liu, R.; Shang, R. A Multi-Level Attention Model for Remote Sensing Image Captions. *Remote Sens.* 2020, 12, 939.





Thank you

Questions?

