

# Naive and object detection for detailed image captioning

Omar, Alexio, Jimmy, Andres

University of Lorraine

November 7, 2022

# Overview

- ① Problem statement
- ② Methodology
- ③ Gantt Chart - Responsibilities
- ④ Short demo of image cutting system
- ⑤ Future work

## Problem Statement

How can we generate more detailed captions?

### Example



Figure 1: General caption: Beach sunset

## Approaches Overview

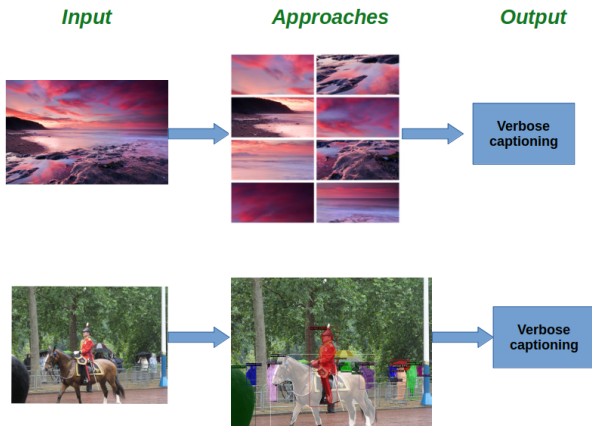


Figure 2: Image cutting and semantic detection

# Technologies

Approach	Visual Model	Language Model	Data sets
Naive	Vision Encoder	Pretrained Decoder	Flick30k, <i>WikiArt</i>
Semantic	Faster R-CNN	Pretrained Decoder	Flick30k, HICO-Det

Table 1: Main architectures and data sets

Visual genome and unrel data sets may be added as a validation.

# Tasks

- ▶ Naive approach development: Alexio and Omar
- ▶ Semantic approach development: Andres and Jimmy

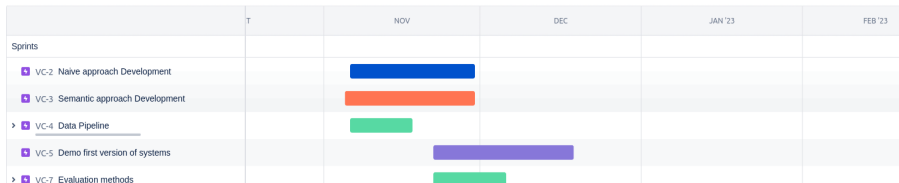


Figure 3: Current timeline and tasks to complete

## Tables and Figures

Image cutting demo

## Future Works

- ▶ Have a better captioning model.
- ▶ Use depth maps.
- ▶ Work on better summarization.
- ▶ Maybe do our own object detection model.