

# MAIS 202 Project Deliverable 1

## Data Selection

Benjamin Jiang

### 1. Choose your dataset

There has been a competition on Kaggle that caught my eye, on image classifying of natural scenes. I will use the dataset provided which is composed of 25 000 pictures of distributed under six categories (glacier, street, forest, mountain, sea and building). This dataset is extensive and will not require much pre-processing.

### 2. Methodology

- i) The dataset is already processed and divided in training, validation and testing sets. The pictures are already labelled with 0-5 representing the different scenes. Not much pre-processing will be required I think (hopefully I didn't just jinx myself).
- ii) The goal of my project is to be able to classify pictures based on the natural scenes in 6 categories. The machine learning model will certainly fall into the classifying models, but more research will be necessary to be more specific. I was thinking of implementing something like the YOLO (which is used for object detection) to then classify the images in their respective category. A lot of work has already been done in image classifying and I will be using it as basis for my project.
- iii) I will showcase my project through a very simple web-app, which will hopefully allow people to import their own picture to be classified. Since I've never made one, I expect it to be very simple and to be very straightforward.