# Problem statement

The goal of this project is to analyze the existing scene recognition models and build an OCR system for images taken from natural scenes.

# Business justification

The problem of identifying text from images taken from natural scene is very common and tricky. The text in these images can have multiple variations like,

* text can have multiple fonts or it can be at any location
* it may be at the background or foreground
* text may be horizontally or vertically inverted and so on

These problems make OCR for natural images a challenging as well as interesting problem which needs to be solved.

# Existing solutions

There are some existing models which work very well on the scene recognition task. Some of the models are

* [**Textsnake**](https://arxiv.org/abs/1807.01544) [Long et al., 2018], a text detection algorithm with the specificity of handling very complex text shapes.
* [**MORAN**](https://arxiv.org/abs/1901.03003) [Luo et al., 2019], a text recognition algorithm using a rectification network and the attention mechanism to correct and read complicated textboxes.
* [**FOTS**](https://arxiv.org/abs/1801.01671) [Liu et al., 2018], an end-to-end approach sharing the convolutions between the detection step and the recognition step to improve robustness and efficiency.
* [**EAST**](https://arxiv.org/abs/1704.03155v2): An Efficient and Accurate Scene Text Detector

# Dataset

[TextOCR](https://textvqa.org/textocr/dataset/) (Data is available under CC BY 4.0 license.)

* Training set (21778 images, 714770 annotations)
* Validation set (3124 images, 107802 annotations)
* Test set (3232 images)

# Proposed Approach

Build a deep learning-based scene recognition model which will

* Text detection
  + This can be formulated as an object detection model which will produce a bounding box as output.
  + The training dataset has bounding box information available and can be used for training such a model
* Text transcription
  + The output of previous model will be fed to text transcription model which will identify and return the text in the image