MAIS 202 - Semantic Understanding of Urban Street Scenes

Project Idea: To create a webapp that will take a photo of a street/outdoor scenery as input and be able to detect certain elements of urban scenery (cars, road, person…) and ideally output an audio of the elements found.

Dataset: The dataset was originally found here: https://www.cityscapes-dataset.com/

- I chose this dataset as it contains 5000 images with high quality annotations and 20000 images with coarse annotations. It contains 30 classes from 50 cities over several months in different weather conditions and varying the background and scene. It is particularly useful for semantic segmentation.
- fine annotations for train and valid sets (3475 annotated images) and dummy annotations (ignore regions) for the test set (1525 images)

Methodology:

1. Data preprocessing:
    a. The dataset is clean and labeled, seeing as it was used for various competitions.
    b. Seeing as there are 30 classes of objects in this dataset, my preprocessing will most likely include downsizing my dataset to more common objects in the environment.

2. Machine learning model:
    a. I would like to be able to classify the urban scenery objects in the input photo and hopefully be able to output an audio of these objects.
    b. I would use a CNN, using TensorFlow and/or Keras libraries in Python. Some suggestions have been to use TensorFlow with Keras front end that is bundled with it. TensorFlow is known for flexibility, while Keras apparently can do part of the work in less code lines and faster.

3. Evaluation Metrics:
    a. A confusion matrix
    b. Validation accuracy as a monitor parameter in Keras
    c. Precision/Recall classification metrics

4. Final Conceptualization:
    a. While a phone app would be ideal, seeing as I do not have experience with app development, a webapp would be more feasible in which users import their photo