Scene based Knowledge Distillation
Ashutosh Tiwari (ashutiwa), Radheshyam Verma (radverma)

Description
Inspired by how in language modeling, removing some words and letting the model predict the word, makes the model learn semantics of the language and embeddings of the words. In this project, we attempt to learn the object scene relations. To this end, our plan is to mask off the objects in the image, and let the model predict the class present in the masked area of the image. Our intuition in this project is that the model will start making sense of the surroundings of the objects. This way a model may learn better contextual relationships the objects have with the environment.
One additional goal of our model could be to provide supplemental world knowledge to the existing models to correct the out of context errors.
In the long term we want to incorporate this model which has learnt the semantic worldly knowledge of coexistence of several classes in a scene into other models and see if it increases their prediction capabilities.

Reading List

- **Knowledge Distillation**
- **CenterNet**
- **Fourier Convolutions in Cropping**
- **ResNet**
- **DenseNet**

Research Plan
We will use masks to mask class positions and then feed it to our network which from the knowledge of having seen multiple images in training time is hopefully aware of semantics of co-existing classes. Output of model is class which it thinks is the object of class which is masked.

Task	Timelines	Owner
Data Exploration	14 March - 21 March	Ashutosh, Radhe
Data Preprocessing	21 March - 28 March	Ashutosh, Radhe
Model Training(ResNet)	28 March - 4 April	Radhe
Models with Fourier + Local Convolution	28 March - 4 April	Ashutosh
Analysis and improvement (which models to choose for next step)	4 April - 11 April	Radhe, Ashutosh
Explore if this can be used in other classification tasks via transfer learning	11 April - 26 April	Radhe, Ashutosh
Report Writing	11 April -	Radhe, Ashutosh

Experiments

Initially we want to check if the model understands the correct object and scene relations. To this end, we will see if the target class is present in the top-50 or top-100 predictions for the masked area. And to see if the top-10 predictions that have been made by the model make sense to be present in the missing part.

Additionally we may experiment with different shapes of masks, such as square, circular etc. We also plan to threshold the mask size so that it doesn't cover a major portion of the image. Transformations, scaling and other data augmentation methods will also be used.

The goal of the model will be to predict the class label at the masked area, so the output will be a C length vector where C is total classes. Depending on the preliminary results of these experiments, we will explore the issues with the model and identify how to correct the errors.

The dataset we plan to use for this will be the object segmentation dataset of COCO. Which has pixel level class information.