Interim Project Report (BlindNet)

Introduction

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.

Background and related work

There has been some work on masking part of the image and letting the model predict the masked out part of the image. Our experiments will be different from that, because we will be predicting the type of the object that is present in the masked part. This is different because our model may learn the object relations with the world.

We are also going to use fourier transform in combination with convolution transforms. Fourier transforms have been extensively explored in this paper.

Progress so far

So far we have done data analysis and written torch loaders to load data and target for initial training. This is the model which is going to serve as a backbone of other transfer learning tasks we are planning to do on top of this.

Right now we are working on creating intermediate files which we currently transform while loading data. This is a time consuming operation since we need to fill our 3 x 3 matrix with class ids one by one for each image. If we persist our transformed array, it will be faster for us to train and test our models.

We do not have interim results to show yet, as much of our work has been on data exploration and loading. Which is as on the schedule in the proposal. We are right now developing two different kinds of models, which we have mentioned in the table in our revised research plan section.

Revised research plan

We are pretty much on track with the project proposal. We have finished the data exploration stage and are working on developing the model. Our revised plan has shifted by 3-4 days from the original proposal. Our overall experimental setup looks as below.

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.

Task	**Timelines**	**Owner**
Data Exploration	14 March - 21 March	Ashutosh, Radhe
Data Preprocessing	21 March - 28 March	Ashutosh, Radhe
Model Training(ResNet)	28 March - 8 April	Radhe
Models with Fourier + Local Convolution	28 March - 8 April	Ashutosh
Analysis and improvement (which models to choose for next step)	8 April - 15 April	Radhe, Ashutosh
Explore if this can be used in other classification tasks via transfer learning	15 April - 26 April	Radhe, Ashutosh
Report Writing	26 April -	Radhe, Ashutosh

References

- Knowledge Distillation
- CenterNet
- Fourier Convolutions in Cropping [https://arxiv.org/pdf/2109.07161.pdf]
- ResNet
- DenseNet
- R-MNet: A Perceptual Adversarial Network for Image Inpainting [https://arxiv.org/pdf/2008.04621.pdf]