CS231N - Project Proposal

Binbin Xiong(bxiong), Minfa Wang(minfa)

We propose to work on the Kaggle competition of iMaterialist Challenge (Fashion) at FGVC5 (Ends May 30, close to project due: June 5), which tackles the problem of multi-products classification in shopping photos. To be more specific, the problem requires building a model which reads product photo and predict all product related labels that appear in it. This problem is very interesting in that it could be used in lots of applications such as products retrieval and shopping recommendation. It is also very challenging since a picture can be taken in different lighting, angles, backgrounds, and levels of occlusion, while different fine-grained categories may look very similar thus hard to distinguish. Another challenging aspect of this project is that the images to labels here is a many-to-many mapping.

To get more context and background of this problem, we plan to first look into the discussion forum in Kaggle to learn from other competitors' experiences and then learn in depth image classification techniques as provided in course reading materials. Particularly, a few papers that we are especially interested in are [Recurrent Models of Visual Attention] [Krizhevsky et al] [Huang et al].

We will use the dataset provided by Kaggle, which contains 1M images with 228 labels in training set, 39k images in test set and 9k images in validation set.

As for now, we are mostly interested in working towards building an attention-based ConvNet for multi-label classification. The high level idea is to apply an attention layer to extract areas/pixels of interest first then feed them to the later classification layers.

As for model evaluation, we will use the standard evaluation required by Kaggle, that is the Mean F1 score for all classes:

$$MeanF1 = \frac{1}{|C|}\sum_{c}\frac{2p_c * r_c}{p_c + r_c}$$

Where p_c is precision and r_c is recall, both for class c.

We plan to develop our model in an iterative way, that is to build basic ConvNet first, then apply tricks learned in class to refine the model with more advanced technique and architecture.

Besides the topline metrics (F1) specified by the competition, we would also like to devote some effort on model understanding. For example, we may use visual attention techniques to demonstrate what part of the image attributes to a certain label prediction and find out the strength and weakness of the model.