What's in an Image?

Michael Mulder

michaelhmulder@gmail.com
linkedin.com/in/mhmulder
github.com/mhmulder

Objective and Motivation

Many large companies are exploring augmented reality and using images in new ways. Walmart in particular is considering using this technology to enhance their home furnishings business. One goal is to be able to have a user upload an image from a room in their house and use that image to make recommendations for furniture they should add or swap out. The goal of this project is to begin that process by simply describing the object(s) in an image. I accomplished this goal by using a pre-trained convolutional neural net (VGG16) and concatenating it to a long short term memory (LSTM) recurrent neural net that fed into another bidirectional LSTM in a seq2seq fashion.

Discussion

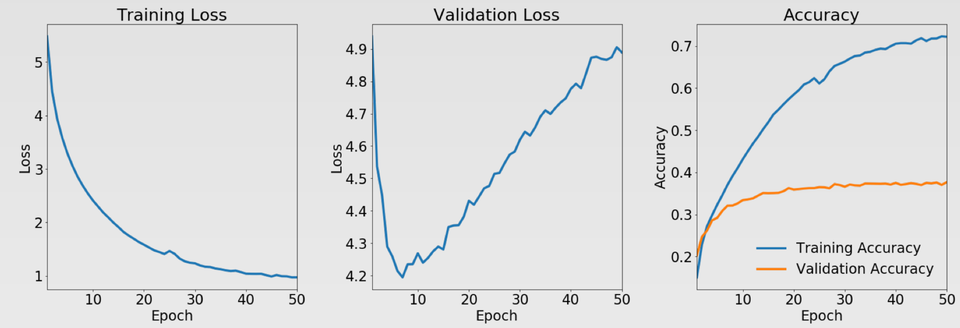
The scraped data contains 296,000 images, of more than 100,000 products, comprising 93 categories (sofas, chairs, etc.), with over 5,500 unique words. For training the net, I randomly sampled 20,000 of the images. Using the net architecture and process described below I was able to achieve successful results. Actual predictions from the test set are shown at the bottom.

Results and Evaluations

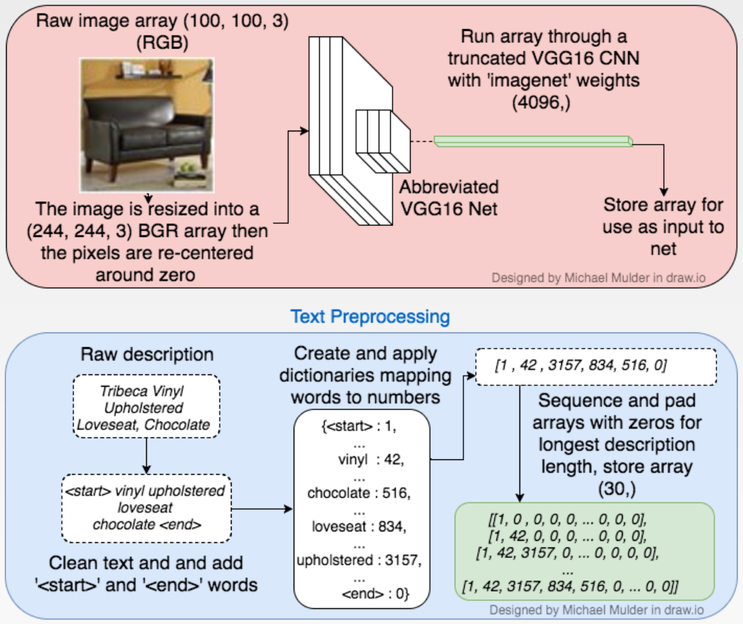
- *Used SpaCy (NLP engine) to clean and lemmatize all words in descriptions and applied cosine similarity to score predictions vs real descriptions and prediction vs clean descriptions*
- *Created a human click through evaluation to score quality of predicted description*

Results and Evaluations Continued

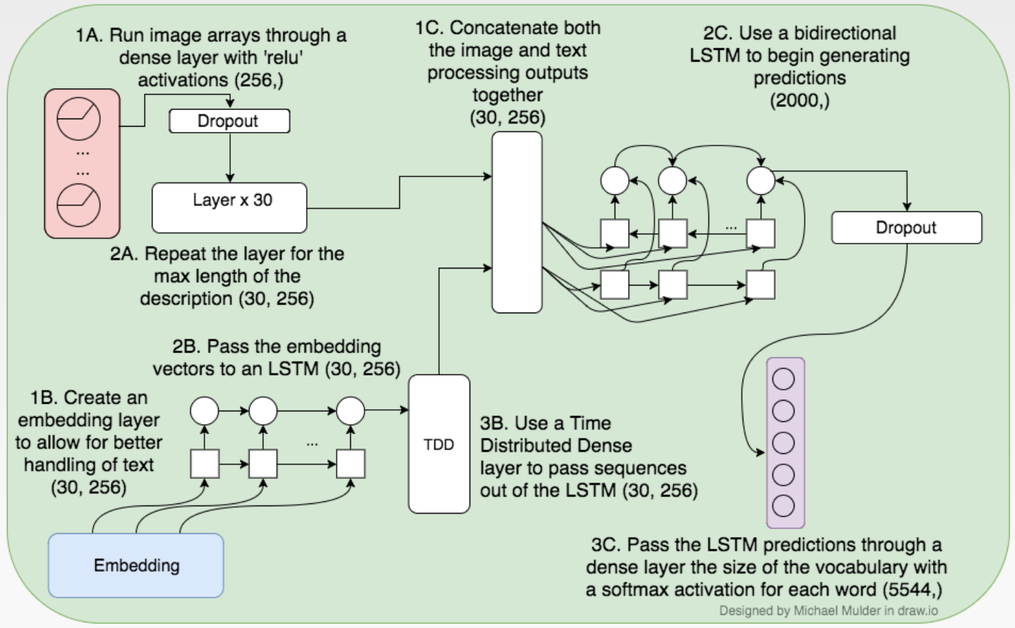
After each epoch categorical cross-entropy loss was calculated as well as accuracy for both the training and validation set. This can be seen below. Clearly, after about 9 epochs the model began to overfit and after a secondary human evaluation, I decided to end the training there.



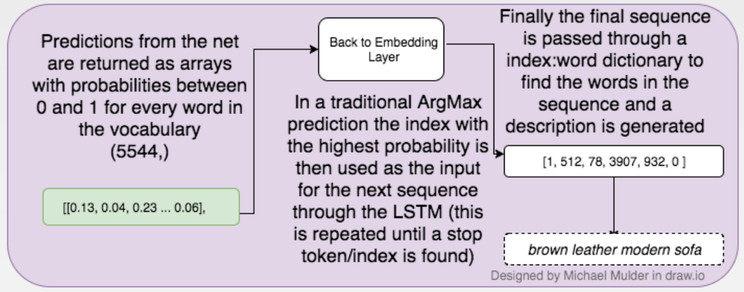
Processing the Data



Main Net Architecture



Making Predictions



Other Applications

- *Redefine how we search for and through images*
- *Object identification in videos*
- *Aid for the visually impaired*
- *Better sales recommendations*
- *Create new data from images*

Great Predictions

black leather executive office chair brown

gel memory foam mattress multiple sizes

lightweight picnic table dining portable party indooroutdoor folding

Interesting Predictions

abordale wood bar stool

iron inch rectangular decmode

hercules triple series triple triple braced hinged hinged

Bad Predictions

nba office chair brown

lancashire coffee table

heavy duty coffee table multiple colors

References

- *Very Deep Convolutional Networks for Large-Scale Image Recognition, K. Simonyan, A. Zisserman, arXiv:1409.1556*
- *The Unreasonable Effectiveness of Recurrent Neural Networks, A. Karpathy*
- *Translation Modeling with Bidirectional Recurrent Neural Networks, M. Sundermeyer, T. Alkhouli*
- *Bidirectional Recurrent Neural Networks as Generative Models - Reconstructing Gaps in Time Series, M. Berglund, T. Raiko*