Phase 3

Advith Chegu, ac1771
Hugo De Moraes, hnm21

Outfit Compatibility Prediction and Diagnosis with Multi-Layered Comparison Network

Dataset

The original dataset that was used was from the popular outfit design site Polyvore. Unfortunately this data was shut down recently and therefore finding the original dataset was difficult. Despite this, I was able to find the data on a google drive and downloaded what I needed. Now the goal is to see if I can replicate the paper's results by first training their MCN model and then running my own evaluation on it.

Model

As we know, the MCN model is made up of 4 parts: Multi-layered feature extractor, Comparison Modules, Visual Semantic Embedding, and Outfit Diagnosis by Gradients. For the project I hope to dig deeper into these respective parts but for now I'd just like to run and test how well they do. I ran the train script for 5 epochs for time constraints. Here are the training results:

```
2022-04-03 22:17:24,630 [INFO ] Valid Loss (clf_loss): 0.5708
2022-04-03 22:17:24,637 [INFO ] AUC: 0.8224
2022-04-03 22:17:24,639 [INFO ] Accuracy@0.5: 0.7074
2022-04-03 22:17:24,640 [INFO ] Positive loss: 0.2910
2022-04-03 22:17:24,643 [INFO ] Positive accuracy: 0.4465
```

As you can see the reported AUC for the training set is 0.82 and the paper states it to be 0.9 so it seems like we are pretty close.

Evaluation

Next I wanted to measure the performance of their model on a test set, these are the results I saw.

```
Test Loss: 0.4880
```

```
AUC: 0.8584
#2462/2463
FitB ACC: 0.5538
```

As you can see the AUC value is still pretty solid at 0.85. Which is still close to the reported value of 0.9.

Improvements

Immediately the first thing I'd like to do is to use a different dataset and test the model on the new data to see if I'm able to recreate the results. Also I would like to update the datasets with the newest fashionable clothing and maybe create some type of pipeline to keep the sets up to date. Maybe datasets from Pinterest will be similar to the Polyvore dataset that the original paper uses.

Fashion++: Minimal Edits for Outfit Improvement

Experiments

The authors of the paper provided a well flushed out implementation of their own results. I was able to clone their implementation with slight modifications to account for changes in software versioning since the time it was published, as well as the different runtime environment (running on Rutgers' iLab servers). I successfully ran their demo experiments which included three (3) example images, label pairs, and produced a result which closely resembled that of their own experiments.

Challenges

Dataset Acquisition

One major challenge I faced was with acquiring the dataset used in training. While the paper originally made it seem rather straightforward, it turned out that the dataset was hosted on a foreign website which required malware to be installed to access. This led me to almost discarding the entire task and starting from zero. Luckily two things happened, for one, after digging through various github repositories and issues, I found a link to a google drive with the dataset. Secondly, the authors provided a minified dataset with three (3) example image, label pairs. These were the ones I used for this phase of the project. I did validate that the dataset obtained works as well, but for the sake of time I opted to stick with the example images.

Dependency Issues

Another challenge I faced had to do with inconsistencies in their program and compatibility issues with the iLab system. These were painful at first, but I was able to remedy them in time. I

did also run into some issues with GPU memory. I'm not entirely sure if this was a legitimate problem or whether it has to do with the aforementioned compatibility issues. After addressing the breaking issues, I did not run into any more GPU memory issues but without a proper diagnosis it is still somewhat of a concern.

Results

Demo 1

Demo 2

Demo 3

Suggestions for improvement

New Dataset

One area of improvement would be to implement the paper's findings on a newer iteration of the dataset used in the paper. I found this dataset in my pursuit of the original one. It has some notable differences, namely in the index, feature mappings, but the features themselves are consistent. This would prove beneficial to us because it would force us to dig into the implementation of the paper to account for the changing features. It would also benefit others potentially looking to replicate the paper with the new dataset.

State-of-the-Art cGAN Architecture

Another point of improvement would be to modernize the underlying cGAN architecture that is used to train the Encoding and Generating modules. Modern architectures like LOGAN (generative adversarial network that uses a latent optimization approach using natural gradient descent (NGD)) have produced admirable improvements over baseline results and currently the best model for Conditional Image Generation on the ImageNet 128x128 dataset.