ML Assignment

Name: Tejas Bana
Email: tejasbana@gmail.com

Problem Statement:

Fashion Category and Attribute detection from images. List of category and subcategory and list of category specific tags.

Naive Solution1:

Create a two headed classification model with one head for multi-class category detection and another head for multi-label attributes.

Cases considered:

1. Public, studio and extracted cloth images are considered.
2. 50 categories and 1000 attributes.
3. Front and back of the clothes are taken into consideration.
4. Image transformation, like rotation for recognizing clothes from every angle.
5. Only Single person images.

Dataset:

DeepFashion----Category and Attribute Prediction Benchmark was chosen as the training and validation dataset. Following are the dataset properties:
1. 289,222 number of clothes images.
2. 50 number of clothing categories, and 1,000 number of clothing attributes
3. Each image is annotated by bounding box and clothing type.
4. DeepFashion: http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html

Training Procedure:

Training Stats :
1. Image Size is equal to 256x256 px.
2. Batch Size is 256.

3. Training set has 269,222 images and validation has 10,000 images.
4. Adam Optimizer with beta1=0.9, beta2=0.999
5. PreTrained models on ImageNet were fine tuned further.

    One Cycle Learning Rate scheduler is used in starting 10 epochs to determine which learning rate is most effective for the model. Learning Rate was set to 2e-4 initially and decreased linearly after every 10 epochs.

Experiments & Results:

Sr. No.	Model	Epochs	Category val_acc	Attribute val_acc	Comments
1.	ResNet-18 (scratch)	10	53.25%	01.22%	Training is too slow, should use transfer learning.
2.	ResNet-18 (PreTrained)	30	99.62%	03.93%	Should use a different loss function.
3.	ResNet-18 (PreTrained)	17	97.17%	03.25%	Should train bigger models
4.	ResNext-52 (PreTrained)	25	93.09%	02.00%	Increasing magnitude of attribute loss
5.	ResNet-18 (PreTrained)	20	96.08%	05.63%	Increasing attribute loss by a factor 10
6.	ResNext-52 (PreTrained)	10	99.62%	09.02%	

Following are the observation while training:
1. Training Model from scratch was time consuming so using a pre-trained model seems a good choice.
2. Pre-trained models tend to give good category classification accuracy but lack the ability to classify multi label attributes.
3. Using a model with more parameters and increasing the magnitude of attribute loss by a factor of 10 resulted in better predictions.
4. Initial attribute val_acc = 1/1000 = 0.001 plus the drop due to multi label, problem is much hard to solve hence will need more sophisticated solution.

Cons:

1. No bounding box
2. Multiple clothes items in a single image.
3. Multiple persons in the images.
4. Clothes from side angles.

Future Work:

Landmark detection and then classifying each landmark with respect to its position in the image should result in better results.
BCRNN is the approach to do the same with Attentive Fashion Grammar Network for Fashion Landmark Detection and Clothing Category Classification: https://github.com/zuoxiang95/BCRNN
Another approach is to extract each clothing item using models like Yolo and then training classification models on individual crop cloths.

Code: https://github.com/tejasbana/tejas_bana_ML_Assignment_july2021