QSTP

Introduction to Deep Learning

Domain Specific Task

Computer Vision: Image Classification

SUBHRAJYOTI DUTTA

Aug 2021

Reporting and Analysis:

1. Supervised Learning

   In Supervised Learning, we divided the training set into 90% for the training set and 10% for the validation set. We trained a CNN with the configuration of 2 convolutional layers, 2 max pooling layers, 1 flattening layer, 2 fully connected neural layers and 1 output layer.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 94, 94, 64)        1792
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 47, 47, 64)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 45, 45, 64)        36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 22, 22, 64)        0
_________________________________________________________________
flatten (Flatten)            (None, 30976)             0
_________________________________________________________________
dense (Dense)                (None, 576)               17842752
_________________________________________________________________
dense_1 (Dense)              (None, 576)               332352
_________________________________________________________________
dense_2 (Dense)              (None, 10)                5770
=================================================================
Total params: 18,219,594
Trainable params: 18,219,594
Non-trainable params: 0
_________________________________________________________________
```

Fig: Model Layers

   I used ReLU Activation Function for the two hidden layers and Softmax Activation Function for the output layer. I used Early Stopping with the check on val loss. After training the performance of supervised machine learning model was:

   Accuracy = 90% to 91%

2. Semi-Supervised Machine Learning

   In Semi-Supervised Learning, I used pseudo labels for training. I built two models, one model was trained upon the training set and used for predicting on the test set. Now, the data points which had a correct prediction rate of more than 99% were taken out. These training set along with the new data points, called pseudo labels, together were formed into a “New Training Set”. The second model was trained upon the New Training Set to create better predictability of the model.

   Pseudocode:

   - Model2a = The first model to be trained
   - Model2a.fit(Training_Set)
   - Pseudo_Labels = Model2a.predict(Test_Set) > 99%

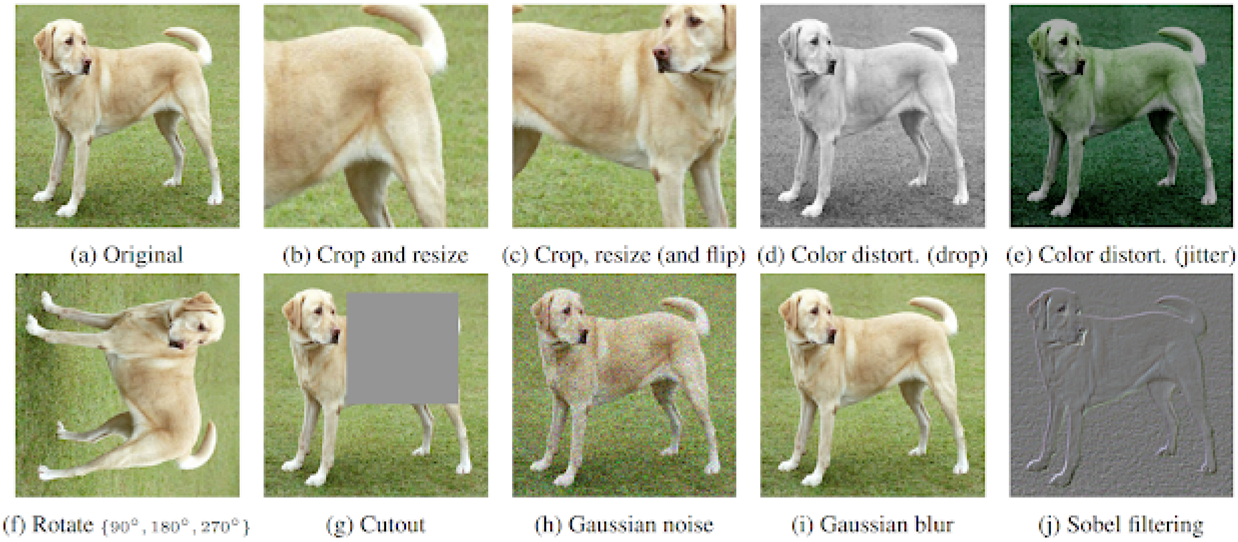
- New_Dataset = Training_Set + Pseudo_Labels
- Model2b = The second model to be trained
- Model2b.fit(New_Dataset)
- Prediction = Model2b.predict(Test_Set)

After training the performance of supervised machine learning model was:

Accuracy = 89% to 94%

3. Unsupervised Machine Learning

In Unsupervised Learning, we use "A Simple Framework for Contrastive Learning of Visual Representations" or SimCLR. We take unlabelled data, then on every data point we make some changes, like: random cropping, rotating, colour distortion, noise etc.



SimCLR first learns generic representations of images are learned by increasing agreement between differently transformed views of the same image and decreasing agreement between transformed views of different images of an unlabeled dataset, this is called contrastive learning. Updating the parameters of a neural network using this contrastive objective causes representations of corresponding views to "attract" each other, while representations of non-corresponding views "repel" each other.

Note: I wasn't able to completely understand the working of unsupervised learning (SimCLR) so my unsupervised learning task remains unfinished.

Comparison:

- Between Supervised Machine Learning (SML) and Semi-Supervised Machine Learning (SSML), on average SSML performed better than just SML. This can be credited to the pseudo labels. My inference is the pseudo labels had easier photos to classify which helped the model focus on determining the features rather than wasting resources on locating the object.

- Unsupervised Machine Learning (UML) performs better than both SML and SSML. UML tries to find features which will help in forming association between transformed pictures of same images and which distinguishes it from other images. This will largely help in categorisation as the model will search for features of the object rather than false features such as colour scheme matching etc.

Questions:

1. When you would expect semi-supervised learning approaches to fail. What can you do to avoid this?

A. Cases where semi-supervised learning may fail:
   a) If the first model's accuracy is low: This may lead to none or extremely low number of pseudo labels for the second model to train upon.
   Solution: The first model has to have medium to good accuracy on the training set.

   b) If the first model gets particularly efficient to classify a particular category of data points: This may lead to the pseudo labels being one particular category heavy. Thus the second model will have more of a particular category thus having poorer performance on the test set.

2. Apart from achieving a high test set accuracy, what other metrics do you think are important while comparing and contrasting different learning approaches?

A. Some metrics I think are important for comparing different learning approaches:
   a) Training Time: If the gain in efficiency is not much significant compared to a lot of consumption of time and resources then it is better to lower the efficiency and save on time.
   b) Type of False Positive: If out of two models, one has better accuracy than the other BUT it is also inletting a lot of specific UNACCEPTABLE data points then it is better to go with the first one. For example, we built two models to aggregate cat images from other. Model 1 having 5% error and Model 2 having 7% error, BUT model 1 is inletting a lot of pornographic images also as part of false positives then model 2 is more acceptable than model 1.
   c) Scalability: If a model is not or very difficult to scale to big datasets then it is not to be used because any model is built to be used commercially but it being difficult to scale possess an additional layer of hurdle.