BTP I - Presentation

Advait Kumar (18D070003)

Incremental Learning For Video Action Recognition

Video Action Recognition

- Identifying and classifying human actions taking place in the frame

REF : https://serre-lab.clps.brown.edu/wp-content/uploads/2012/08/HMDB_snapshot2.png

Popular Approach

- Extract frames from videos
- Treat each frame as a separate image and use traditional deep learning techniques for classification
- Could also use 3D neural networks : exploiting the temporal dependencies in the video

Incremental Learning

- Learn new examples while retaining accuracy on the original trained examples as well (hardware limitations)
- Has major real life applications wherein new data is fed in real time but the model shouldnt forget its original knowledge
- It represents a combination of both supervised and unsupervised learning

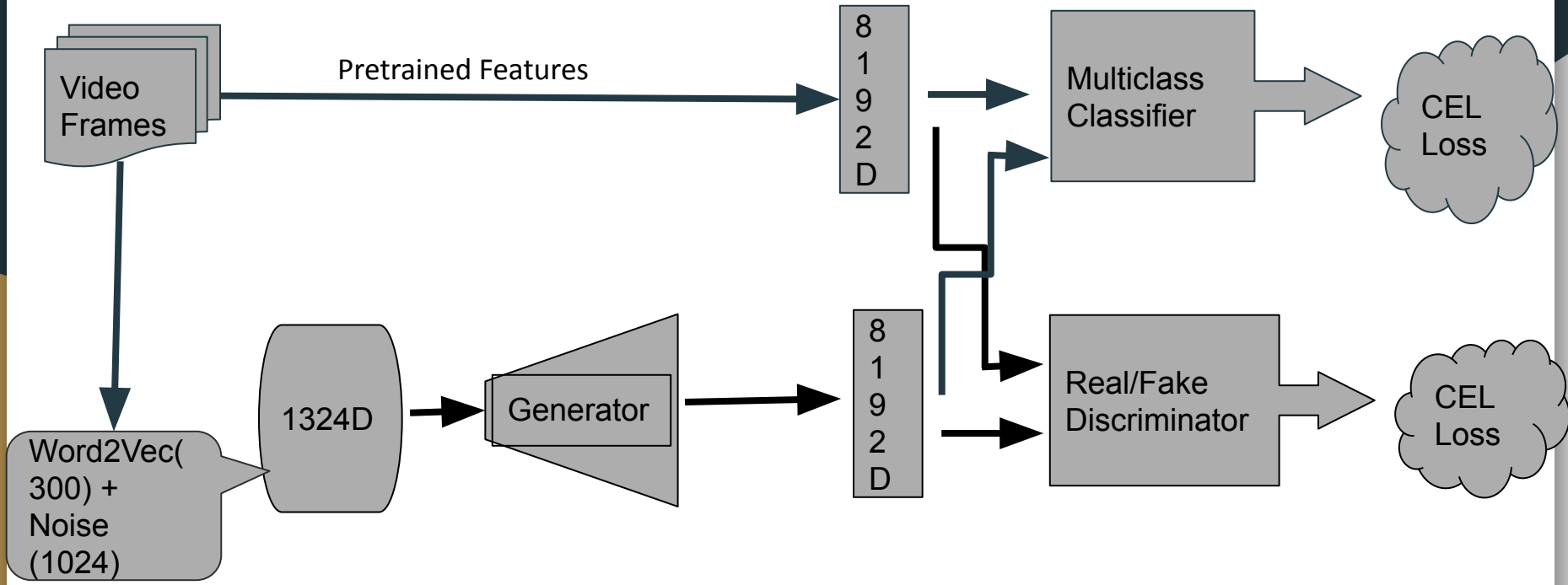
Catastrophic Forgetting

- One of the major problems encountered in incremental learning is catastrophic forgetting.
- Once the model is trained on new data it tends to forget the old mapping.
- Hence model overfits on new data but underfits on the old data
- Cannot retrain the model on old data due to data storage limitations

Solution to Catastrophic Forgetting

- There are a few solutions proposed for tackling catastrophic forgetting which range from changing the model architecture, increasing layers, pretraining models to loss functions and their tuning.
- We mainly focus on tuning the loss function and using a different representation of the video features to counter forgetting

Model Architecture



Training Details

First Set

- The first set consists of entirely new data with no previous data for the model to store.
- The classifier is trained using the usual softmax + cross entropy loss
- The GAN is trained using a hybrid loss consisting of :
  a. Adversarial loss
  b. MSE loss between the 8192D feature vectors
  c. Classifier loss
  d. KL Divergence loss

Incremental Sets

- For the incremental sets, we need to train the model on the new data and at the same time, retain as much information as possible about the old data
- Since it is zero shot, the generator is used to generate the old class feature representations.
- Hybrid loss of old classifier and new classifier cross entropy is used
- GAN is trained on the same hybrid loss as before (both the new data and the generated old data representation)

Results

Benchmark

- This is a novel problem statement and hasnt been attempted before on these datasets, hence we dont have any baselines to compare with
- The usual zero shot learning accuracy is around 40-50% on past data once the model has been trained on the complete dataset.

UCF 101 Details

- Videos collected from youtube
- 101 class actions divided into 4 groups having similar actions
- Source : https://www.crcv.ucf.edu/data/UCF101.php
- Widely used dataset for action recognition

UCF 101

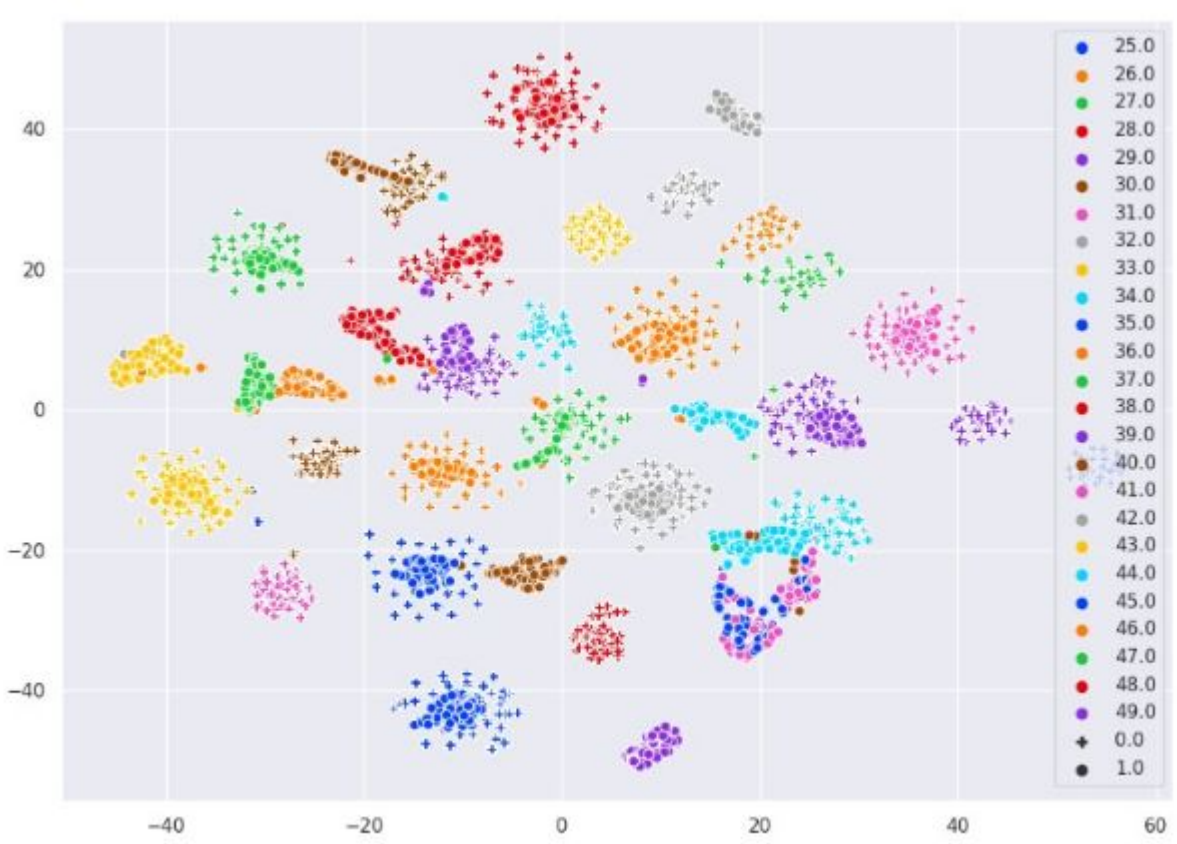
Classes	Episode 1(0-40)	Episode 2(41-50)	Episode 3(51-60)	Episode 4(61-70)
0-40	94.59% / 94.81%	91.33% / 88.59%	82.03% / 88.81%	63.77% / 92.11%
41-50	—	91.54% / 88.41%	84.27% / 90.00%	65.07% / 92.84%
51-60	—	—	89.80% / 89.90%	70.31% / 98.00%
61-70	—	—	—	73.36% / 85.93%

HMDB 51 Details

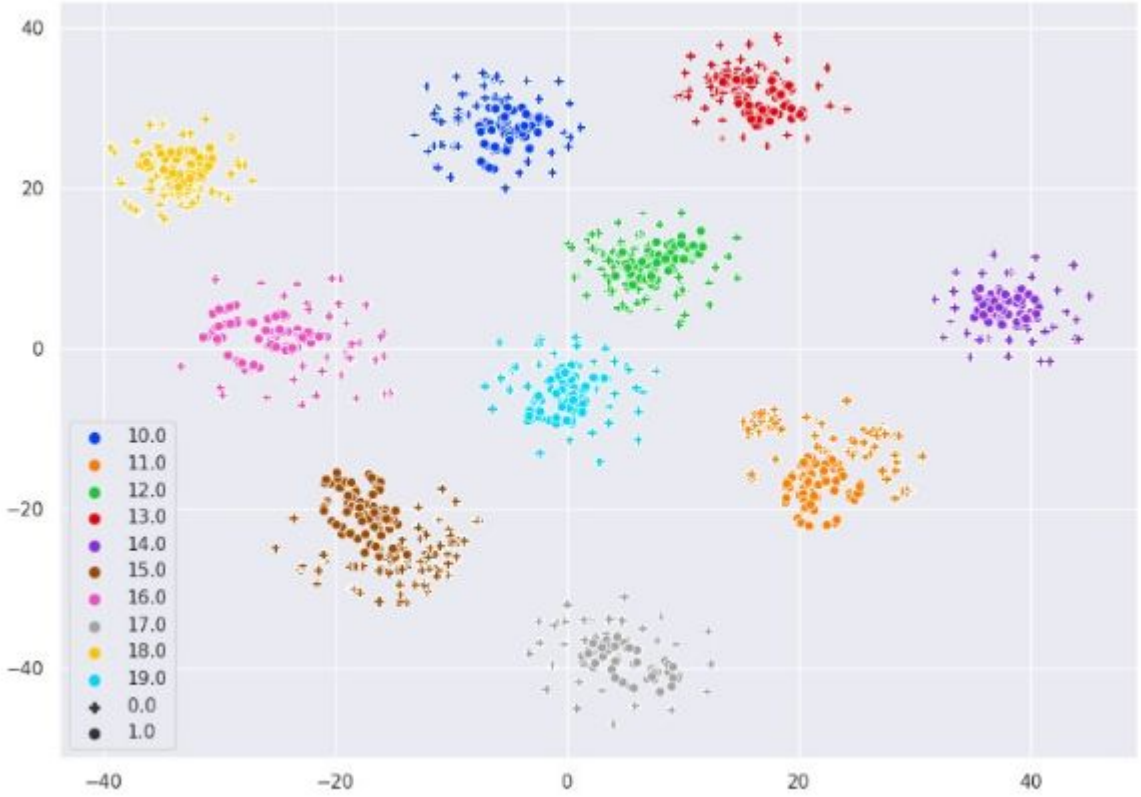
- Around 7000 videos over 51 classes
- Has human actions
- Much tougher than UCF 101 due to lesser number of clips and classes (more chances of overfitting)
- Source : https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/

HMDB 51

Features	Classes: 20-25 / 0-20	Classes: 25-30 / 0-25	Classes: 30-35 / 0-30
Pre-trained	63.87% / 21.71%	68.49% / 29.19%	54.40% / 34.29%
Generated	64.52% / 93.60% (epoch 400)	67.81% / 91.41% (epoch 600)	51.65% / 94.29% (epoch 400)



TSNE plot for UCF 101. Class number 40-49 are unseen classes



TSNE plot for classes 10-19 of HMDB 51. This is after the first increment has been trained.

Further Work

- One shot / few shot learning can also be tried out
- Comparision with the baseline LWF and EWC losses along with a simple linear neural network
- The number of classes in a certain set can be varied and the variation in accuracy can be further studied to find some possible correlations.

All the code and the figures can be found here :
https://github.com/advaitkumar3107/Generalised-Zero-Shot-Learning