Indraprastha Institute of Information Technology, Delhi

Human Activity Recognition with Video Classification

Problem Statement

Human Activity Recognition is a time-series classification task involving prediction of the movement of a person like picking something up, putting something down, opening or closing something etc. depicted in a particular video. Video classification requires a series of multiple images to combine together to classify the action that is being performed. In this task, we handle 174 different human activities.

Motivation

Human Activity Recognition is an emerging domain in deep learning that has a lot of important applications in daily life including assistive living applications for smart homes, CCTV surveillance and security, health care monitoring systems and more.

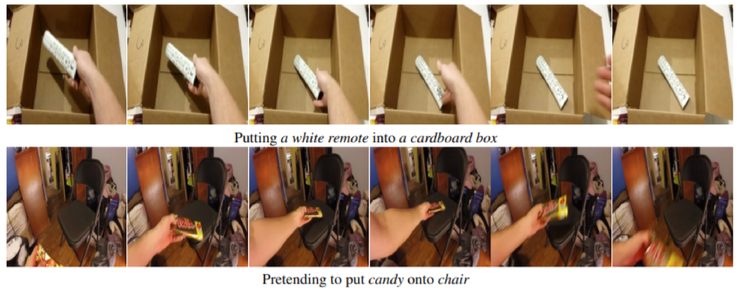
Data set Acquisition

We have used public 20BN-SOMETHING-SOMETHING dataset is used for the task. The dataset consists of a large collection of densely-labeled video clips that show humans performing pre defined basic actions with everyday objects. Dataset consists of total 220,847 videos with training set size as 168,913, validation size as 24,777 and test set size as 27,157.Duration of videos ranges from 2 to 6 seconds. Dataset has 174 different labels showing different human activities.

For each video in the training and validation sets, object annotations in addition to the video label are also provided wherever applicable. For example, for a label like ”Putting [something] onto [something]” there is also an annotated version like ”Putting a cup onto a table”. In total, there are 318,572 annotations involving 30,408 unique objects.

Some examples of these labels are as follows:

Label	Number of Examples
Dropping something next to something	1,232
Attaching something to something	1,227
Dropping something into something	1,222
Showing something next to something	1,185



Putting *a white remote* into *a cardboard box*

Pretending to put *candy* onto *chair*

Figure 1: Example Video and description

Evaluation Metric

Since this is a classification problem, we will use the following metrics for evaluation:

- Cross-Entropy Loss
- Accuracy
- Precision
- Recall
- F1-score
- Top-1 error
- Top-5 error

Top-1 and Top-5 errors are calculated to submit our results of test data on twentybn dataset website which would benchmark your approach against that of other participants.

Pre-processing Techniques

- **Extracting frames from the video**
- **Resizing Frames**
- **Background subtraction :-** We separate parts of the image that are invariant over time(frames) i.e the background, from the objects that make up the moving foreground.
- **Data Augmentation :-**
  1. Random Rotation
  2. Flipping Frames
  3. Down sampling

Deliverables of individual team members

- Preprocessing of Data - Chinmay, Anushika
- Baseline model implementation - Akshyta
- All 3 members will be implementing atleast 1 advanced model, study and analyze the errors and techniques for improving the model’s performance

References

- **Dataset link :-** https://20bn.com/datasets/something-something/v2
- **Project Idea :-** https://data-flair.training/blogs/deep-learning-project-ideas/

Team Members

Group Number : 29	
Name	Roll Number
Chinmay	2017274
Akshyta Katyal	2017216
Anushika Gupta	2017135