



Human Activity Recognition

Project Idea

- The task is to classify or predict the activity/action being performed by a human in a video.
- Human Activity Recognition is a type of time series classification problem where we need data from a series of timesteps to correctly classify the action being performed.



Dataset

- The Dataset used is the Youtube UCF50 – Action Recognition Dataset.
- UCF50 is an action recognition dataset which contains:
 - 50 Action Categories consisting of realistic YouTube videos
 - 133 Average Videos per Action Category
 - 199 Average Number of Frames per Video
 - 320 Average Frames Width per Video
 - 240 Average Frames Height per Video
 - 26 Average Frames Per Seconds per Video



Data Preprocessing

in order to simplify the computations and complexity of our models, we will do the following:

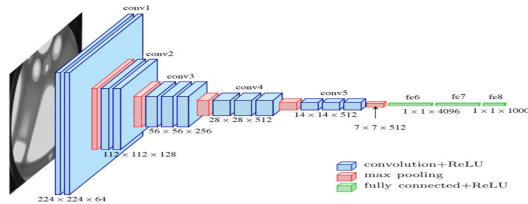
- ❑ train only on a few classes of the dataset :

Drumming - Biking - Basketball - Diving - Billiards - HorseRiding - Mixing - PushUps - Skiing - Swing

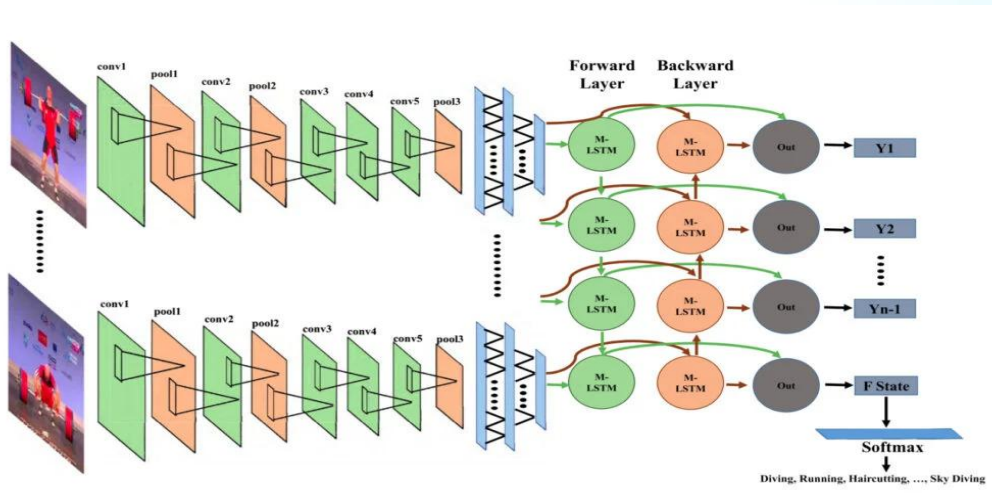
- ❑ take 3000 images for each class
- ❑ resize the image width and height to 64*64
- ❑ the pixel will be normalized to the range [0,1]

Construct the models:

- Model 1: Convolutional neural networks with pretrained model VGG16 (64*64*3).



- Model 2: Model1 + Bidirectional LSTM



Predict human Activity



Predict human Activity

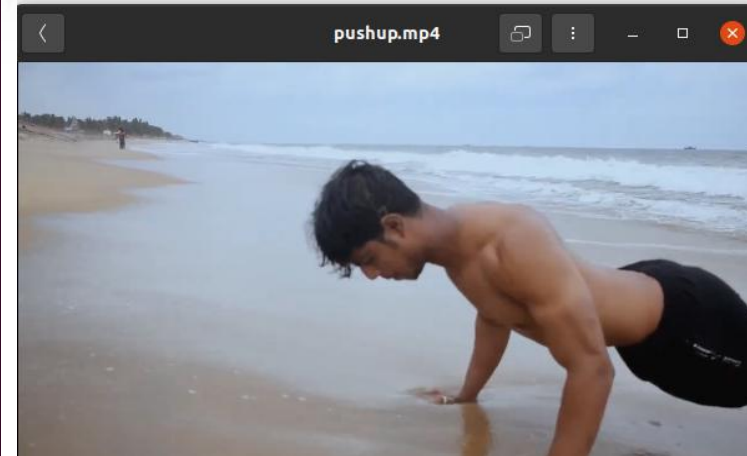
Human Activity Recognition model predicts the following:

The Activity being done in your video is mostly: PushUps

The probabilities for all the activities are given as following:

PushUps	0.561
HorseRiding	0.25
Skiing	0.06
Swing	0.058
Biking	0.033
Basketball	0.029
Mixing	0.008
Drumming	0.0
Diving	0.0
Billiards	0.0

(base) mahmoud@mahmoud-pc:~/Final_project\$

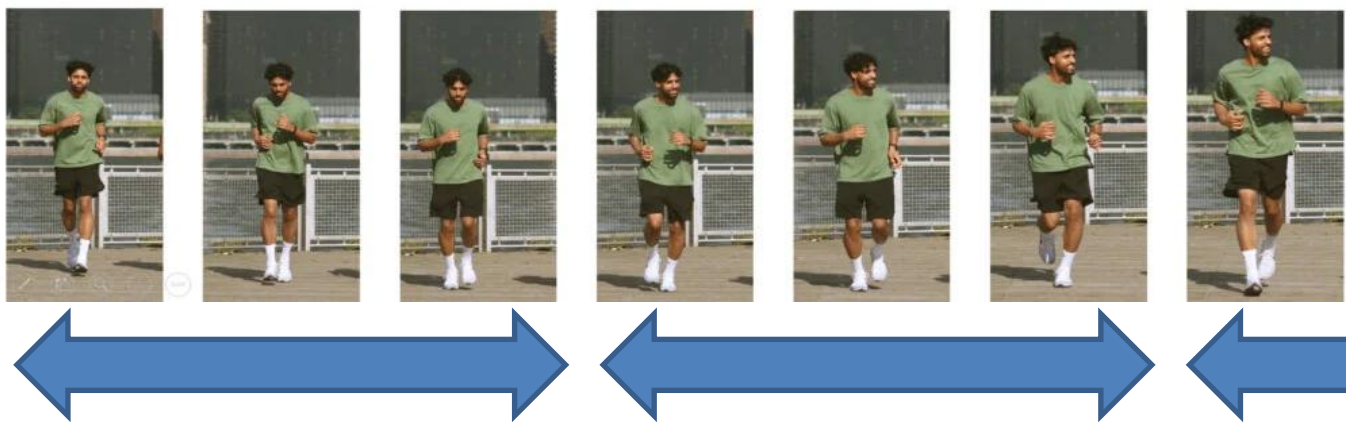


Predict human Activity: Web Interface

<https://www.youtube.com/watch?v=nsyNWH6VK0U>



Predict Live & on whole video



Difficulties and Further Work:

- We trained the model on Google Colab, so we were limited to GPU resources , and had to resize the images to 64×64 , and use only a smaller part from the video frames for each activity (3000 per activity).
- The model was trained only on a part from the dataset UCF50 , with only a few actions. So for a better performance, we should train our model on a larger Data.



Thanks
for Your Kind
Attention

Any
Questions?

SPICED

