Centre for Artificial Intelligence and Machine Learning

A Centre of Excellence of OU under RUSA 2.0, M.H.R.D, Govt. of India, OSMANIA UNIVERSITY

Video-Based Human Activity Recognition

ABSTRACT

Human Action Recognition (HAR) is a critical task in computer vision, enabling the identification of human activities from image and video data. This project leverages a pre-trained **MobileNetV3** model to develop an efficient and accurate system for recognizing human actions. The dataset includes labeled images of various human activities, which are preprocessed and used to fine-tune the MobileNetV3 architecture.

TEAM

1. K.Sai Nithin
   1005-21-729081
2. S. Vara Prasad
   1005-21-729082

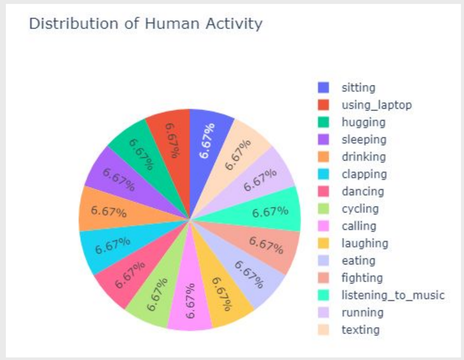
INTRODUCTION

Human Activity Recognition (HAR) involves identifying human actions from images or videos and has applications in fields like surveillance and healthcare. This project uses **MobileNetV3**, a lightweight and efficient deep learning model, for classifying various human activities. By leveraging transfer learning with a pre-trained MobileNetV3 model, the system can recognize actions such as walking, sitting, and running. The model is trained on labeled images and tested on both image and video data, making it suitable for real-time applications. This project demonstrates the power of MobileNetV3 in human activity recognition for practical use cases.

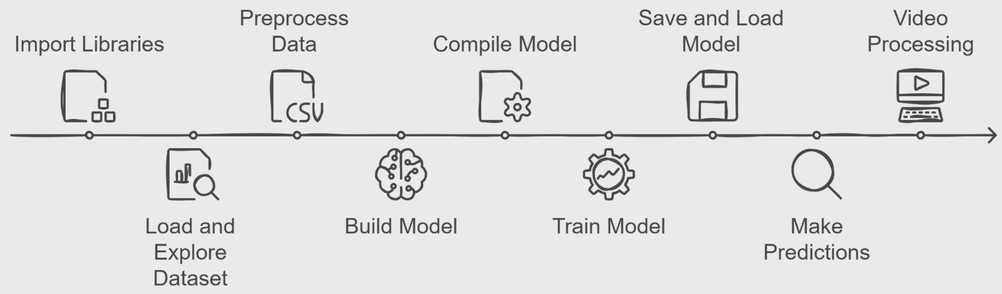
DATASET

The **Human Activity Recognition (HAR) Dataset** is designed to recognize various human actions from images and videos. It consists of:

- Contains **15 activity classes** such as walking, sitting, standing, running, etc.
- Over **12,000 labeled images**, including training and validation images.
- Each image is assigned **one activity label** and stored in separate folders for each class.
- Images resized to **160x160 pixels** to match model input requirements.
- Labels are **one-hot encoded** for classification.
- Includes a separate **testing set** with images and videos for model evaluation.



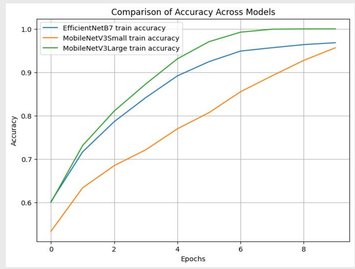
IMPLEMENTATION STEPS

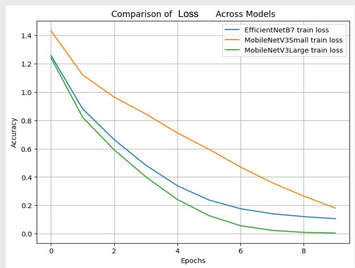


- Import necessary libraries like TensorFlow, Keras, numpy, pandas, and OpenCV.
- Load training and testing datasets using pandas and explore class distributions.
- Resize images to 160x160 pixels and encode labels into one-hot format.
- Build a model using MobileNetV3 with additional Dense and Flatten layers.
- Compile the model using Adam optimizer and categorical cross-entropy loss.
- Train the model on preprocessed data while tracking accuracy and loss.
- Save the trained model and load it for future predictions or evaluation.
- Test the model on new images and map predictions to class labels.
- Process video frames to predict actions and extract action sequences.
- Display results and plot confidence trends for predictions.

EXPERIMENTATION

- Experimented with three deep learning models: **EfficientNetB7**, **MobileNetV3Small**, and **MobileNetV3Large**.
- Trained each model on the dataset for 10 epochs using the Adam optimizer and categorical cross-entropy loss.
- Compared models based on training accuracy and loss metrics.
- Visualized model performance through accuracy and loss plots over the training epochs.
- Observed that **MobileNetV3Large** achieved the highest accuracy among the three models.
- Exported all trained models for reuse, with **MobileNetV3Large** designated as the final model for deployment.





RESULTS

- Achieved **training accuracy of 99%** after model training.
- Loss values decreased consistently, indicating effective learning during training.
- Video processing demonstrated accurate action sequences, e.g., "Walking → Sitting → Running".
- Saved the model into the `.keras` format.

CONCLUSION

- The `MobileNetV3Large` model achieved the highest accuracy among the models tested.
- The Flask application allows users to upload videos and receive real-time predictions.
- The system provides a user-friendly interface with activity predictions and confidence scores.
- Future improvements could include better handling of video sequences or expanding the dataset.
- This model can be applied in areas like security, sports analysis, and human-computer interaction.

REFERENCES

1. Nagadia, M. (2024). *Human Action Recognition (HAR) Dataset*. Kaggle. Available at: https://www.kaggle.com/datasets/meetnagadia/human-action-recognition-har-dataset
2. Keras Documentation. (2024). *Keras: The Python Deep Learning Library*. Available at: https://keras.io/
3. TensorFlow Documentation. (2024). *TensorFlow: An Open Source Platform for Machine Learning*. Available at: https://www.tensorflow.org/

ACKNOWLEDGEMENT

1. Thanks to the **HAR dataset providers** on Kaggle for making the dataset available.
2. Appreciation to the **MobileNetV3 creators** for the efficient model architecture.
3. Gratitude to the **Flask team** for the web framework used in integration.
4. Acknowledgement to **OpenCV contributors** for the video processing library.