PROJECT REPORT

- **TOPIC:-** Facial Recognition project using LFW (Labelled Faces in the Wild) dataset.
- **OBJECTIVE:-** The objective of the Facial Recognition Project using the Labeled Faces in the Wild (LFW) Dataset is to develop and evaluate a robust facial recognition system that can accurately identify and verify individuals based on their facial features.

- **About the dataset-**

Purpose and Creation:

- LFW is a benchmark dataset created to facilitate research in the field of facial recognition and face verification.
- It was introduced in 2007 by researchers at the University of Massachusetts, Amherst.

Dataset Composition:

- The dataset contains over 13,000 labeled images of faces collected from the internet.
- These images represent 5,749 distinct individuals, with each image labeled with the name of the person depicted.
- The dataset includes a wide range of variations in terms of lighting, pose, expression, age, and gender.

Face Verification Protocol:

- LFW is primarily used for the task of face verification, where the goal is to determine whether two images represent the same person.
- The dataset is often split into training and testing sets, with 6,000 face pairs used for evaluation (3,000 matching pairs and 3,000 non-matching pairs).

Variability and Challenges:

- The images in LFW are not taken in controlled environments, making the dataset challenging due to variations in background, lighting, occlusions, and facial expressions.
- The images reflect real-world conditions, making it an excellent benchmark for evaluating the robustness of facial recognition systems.

Preprocessing Requirements:

- Researchers often need to preprocess the images to align faces, normalize pixel values, or extract facial features using methods like Histogram of Oriented Gradients (HOG) or deep learning techniques.

Use in Machine Learning:

- LFW has been widely used to train and evaluate various facial recognition algorithms, from traditional machine learning methods like Support Vector Machines (SVMs) to modern deep learning models, including Convolutional Neural Networks (CNNs).

➢ **<u>Phases of Project-</u>**

Here are the key phases of the Labeled Faces in the Wild (LFW) dataset:

1. **Data Collection and Labeling:** Images of faces were collected from the web and manually labeled with the corresponding individual's name.
2. **Preprocessing and Pair Matching:** Faces were detected, aligned, and normalized; the dataset was structured into matched and mismatched pairs for face verification tasks.
3. **Model Training and Development:** Models were trained using various techniques, from traditional machine learning to deep learning, to perform face verification.

4. **Evaluation and Benchmarking:** Model performance was evaluated using verification accuracy, ROC curves, and AUC, establishing benchmarks for facial recognition systems.

➢ ROLES AND RESPONSIBILITIES:-

1.Project Leader-Tanisha Lalwani

Responsibilities:

- Planning and Setup of dataset on 18.08.24
- Define the project's objectives and scope.
- Create a timeline and assign roles to each member.
- Prepared any graphical assets needed.

2.Team Member 1: Aryan Shukla

Responsibilities:

- Repository Creation and Initial Development on 19.08.24
- Created the GitHub repository and added relevant dataset to the repository.
- Gather requirements and outline the content structure (e.g., documentation, code, assets).

3.Team Member 2: Kashish Singh

Responsibilities:

- Development and Collaboration was done on 20.08.24
- Continued developing features and make regular commits to their respective dataset.
- Communicated daily through a group chat or video call to ensure everyone is aligned on progress and next steps.
- Prepared graphical assets (if needed).

4.Team Member 3: Akanksha Pawar

Responsibilities:

- Finalization and Deployment of the project on 21.08.24
- Helped in creating and organizing project documentation and reports.
- Finalizing and uploading of design assets to the repository
- Conducted a final review meeting to gather feedback and make any last-minute adjustments.

➢ By dividing responsibilities and maintaining clear communication, the group can successfully create a GitHub repository of the project facial Recognition using LFW (Labelled Faces in the Wild) dataset.

----------Progress during Aug 31 to Sep 7th ----------

Team member 1

1) Tanisha:a)Discovered how to put data access controls in place so that every team member can use and access the dataset.

Team member 2

2) Aryan:a)Examine data augmentation methods to enhance model generalization, such as flipping, rotating, and adjusting brightness.

Team member 3

3) Akansha:a)Recognize how to divide datasets into test, validation, and training sets so that there is no data leakage.

Team member 4

4)Kashish:a)Examined the report's drafting, paying close attention to the preprocessing methods and dataset preparation plans.

—----------Progress during Sep8 to Sep15—----------

Team member 1

1)Tanisha:Make sure the team members can all access the dataset by uploading it to the cloud.

Team member 2

2)Aryan:Complete the preprocessing and test it on the real dataset to make sure the object detection models are using the right format.

Team member 3

3)Akansha:Complete the literature table and note any difficulties you had when gathering and preprocessing the data.

Team member 4

4)Kashish:Use a small sample of the dataset for test training to make sure the model is operating as intended.

—----------Progress during Sep 16 to Sep 23—----------

Team member 1

1)Tanisha: Thoroughly perform both data augmentation as well as data cleaning in order to make the dataset training ready.

Team member 2

2)Aryan:Give the team the dataset so they can use kashish to train the model and prepare it for use.

Team member 3

3)Akansha:Analyze and draw attention to any problems or difficulties that arose early on in the model training process.

Team member 4

4)Kashish:Tried using the preprocessed dataset for the object detection model training (YOLO, Faster R-CNN, SSD).

---------- Progress during Sep 24to Sep 30----------

Team member 1

- 1)Tanisha Consider looking into the dataset for problems such as incorrect bounding boxes which are not properly labeled or exist but were not provided which could be hampering the training phase
- .Team member 2

  2)Aryan: Tasked with the verification of the existence of files during upload and access as well as their integrity..

  Team member3
- 3) Akansha: Include in the project report relevant information regarding the errors and difficulties faced in the training. Note any steps in correction undertaken by the group.
- Team member4

  4)Kashish: Debug preprocessing pipeline issues, such as incorrect label formats or scaling inconsistencies, that might be affecting model performance (Work with the team to resolve these issues).

---------- Progress during Oct1 to Oct 7----------

Team member 1

1)Tanisha: Resolve any remaining dataset issues identified in Week 5, ensuring that all labels and bounding boxes are correct.

Team member 2

- 2)Aryan: Verify the divisions of the dataset (training, validation, and testing) to confirm that there is no overlap and that the data distribution is balanced.
- Team member 3
- 3)Akansha: Oversee access logs and administer data version control to guarantee that the team utilizes the appropriate and most current dataset. Record the final training outcomes, encompassing accuracy scores and evaluation metrics.
- Team member 4
  - 4)Kashish: Retrain the object detection models after resolving preprocessing and dataset issues. Also Calculate the model accuracy.
  - Ensure that the system is ready for further stages of the project (e.g., testing, validation).

---------- Progress during Oct8 to Oct 14----------

Team member 1

1)Tanisha: For testing, reduce the 25k dataset to 1,000 images. Make sure datasets are training-adaptable and well-structured.

Team member 2

2)Aryan: Null

- Team member 3

3)Akansha:Null

- Team member 4

4)Kashish: Examine and record sliding window and object localization techniques. Write a report section discussing these ideas and how they are used.

---------- Progress during Oct15 to Oct 21----------

Team member 1

1)Tanisha:Null

Team member 2

2)Aryan: After setting up Google Drive, import the dataset and remove any missing photos. Create training and validation sets from the dataset. Create the CNN model and preprocess the images. Optimize the model's performance by retraining it.

- Team member 3

3)Akansha: Load the COCO annotations and install the necessary libraries. For every image, extract the labels and bounding boxes. Help with training by converting labels to one-hot encoding. Analyze track accuracy and model performance.

- Team member 4

4)Kashish:Null

---------- Progress during Oct22to Oct 28----------

Team member 1

1)Tanisha: Maintain consistency and quality by loading and filtering the 25,000 photos. Complete the splits for training and validation with equal representation.

Team member 2

2)Aryan: Verify each and every annotation. align with filtered photos and check the accuracy of the label.

- Team member 3

3)Akansha: Use the complete dataset to monitor key metrics and model accuracy. Record results and findings for the report.

- Team member 4

4)Kashish: Images were examined for the best possible format and resolution. Include the entire 25k dataset in the pipeline used to train the model.

---------- Progress during Oct28o Nov 3----------

Team member 1

1)Tanisha: Maintain consistency and quality by loading and filtering the 40k photos.

Team member 2

2)Aryan: Complete the splits for training and validation with equal representation.Verify the accuracy of the labels and make sure all annotations match the filtered images.

- Team member 3

3)Akansha: Use the complete dataset to monitor key metrics and model accuracy. Record results and findings for the report.

- Team member 4

4)Kashish: Images were examined for the best possible format and resolution. Include the entire 40k dataset in the pipeline used to train the model.

The Labelled Faces in the Wild (LFW) dataset is a benchmark in facial recognition, consisting of labeled images of faces collected from the web. It's widely used to evaluate the performance of face verification systems. Here's a table summarizing some models and their performance on the LFW dataset:

Model	Accuracy (%)	Year	

			Method/Approach
Eigenfaces + SVM	60.0	2007	Feature-based + Support Vector Machine
DeepFace	97.35	2014	Feature-based + Support Vector Machine
VGG-Face	98.95	2015	Deep Learning (CNN)
FaceNet	99.63	2015	Deep Learning + Triplet Loss
OpenFace	93.80	2016	Deep Learning (FaceNet-based)
SphereFace	99.42	2017	Deep Learning + Softmax
ArcFace	99.83	2018	Deep Learning + Additive Angular Margin Loss
CosFace	99.73	2018	Deep Learning + Large Margin Cosine Loss
MagFace	99.86	2021	Deep Learning + Magnitude-

			aware Feature Embedding

Notes:

- **Eigenfaces + SVM**: One of the earliest approaches, combining principal component analysis with SVM for classification.
- **DeepFace**: Developed by Facebook, this was one of the first models to surpass human-level performance on LFW.
- **VGG-Face**: Based on the VGGNet architecture, known for its depth and performance in various vision tasks.
- **FaceNet**: Introduced the concept of "triplet loss" to ensure that facial embeddings are well separated.
- **OpenFace**: An open-source implementation inspired by FaceNet, suitable for real-time performance.
- **SphereFace, ArcFace, CosFace**: These models introduced different types of margin-based loss functions to improve discriminative power in the feature space.
- **MagFace**: Focuses on adaptive feature magnitude to better align the model's output with the difficulty of the recognition task.

This table gives an overview of how facial recognition has evolved, particularly how deep learning techniques have led to significant improvements in accuracy.