Facial Recognition project using LFW (Labelled Faces in the Wild) dataset.

Tanisha Lalwani
G.H Raisoni College of Engineering,Nagpur
Ugmr20230013@ihub-data.iiit.ac.in
Akansha A Pawar
G.H Raisoni College of Engineering,Nagpur
Ugmr20230044@ihub-data.iiit.ac.in
Kashish Singh
G.H Raisoni College of Engineering,Nagpur
Ugmr20230026@ihub-data.iiit.ac.in
Aryan Shukla
G.H Raisoni College of Engineering,Nagpur
Ugmr20230038@ihub-data.iiit.ac.in

Abstract
Most face databases are built under control to facilitate the study of certain aspects of the face recognition problem. These parameters include variables such as location, setting, lighting, background, camera quality, and gender. While humans can control the negative aspects of the captured image in many applications of face recognition technology, there are many applications where clinicians have little or no control over these parameters. Thisarchive, Labeled Faces in the Wild, is provided to aid in the study of the ultimate unconstrained information problem. The file contains images of faces labeled covering a variety of everyday situations. The database exhibits "natural" variability in factors such as pose, lighting, race, props, occlusion, and background. In addition to describing the library's content, we also provide specific test models that the library is suited for. This is done to make the research in the archive as consistent and comparable as possible. We provide baseline results, including results from a state-of-the-art visual mask combined with a face assembly. To facilitate storage testing, we provide a variety of storage parallels, including assembly models. [1]

Keywords
Facial Recognition, Deep Learning, LFW Dataset, Computer Vision, Image Processing, Model Training, Accuracy Optimization

Introduction
Facial recognition technology identifies individuals by analyzing and comparing facial features, finding extensive applications in security, user authentication, and surveillance. This project uses the Labeled Faces in the Wild (LFW) dataset, a widely recognized resource with 13,000+ images collected from the web. The LFW dataset, labeled by individuals, presents a real-world testing ground for face recognition models, facilitating training and evaluation of machine learning approaches in variable and unconstrained environments.
Labeled Faces within the Wild (LFW) may be a database of confront photos outlined for considering the issue of unconstrained confront acknowledgment. This database was made and kept up by analysts at the College of Massachusetts, Amherst (particular references are in Affirmations area).
This paper describes a database of human face images designed as an aid in studying the problem of unconstrained face recognition. Confront acknowledgment is the issue of distinguishing a particular person, rather than simply recognizing the nearness of a human confront, which is regularly called facedetection. The common term face recognition can allude to a number of different problems counting, but not constrained to, the following.
Face Confirmation: Given a picture of a confront, choose which individual from among a set of individuals the picture speaks to, on the off chance that any.
Pair Coordinating: Given two pictures, each of which contains a confront, decide whether the two individuals envisioned speak to the same individual.
Our database, which we called Labeled Faces within the Wild (LFW), can be used to ponder these issues in unconstrained situations, as well as other face handling errands, such as confront arrangement and confront division. The essential commitment of LFW is giving a huge set of generally unconstrained confront pictures. By unconstrained, we cruel faces that appear a expansive range of the variety seen in way of life. This incorporates variety in posture, lighting,expression, foundation, race, ethnicity, age, sex, clothing, hairdos, cameraquality, color immersion, and other parameters. The reason we are interestedin common variety is that for numerous assignments, confront acknowledgment must work in real-world circumstances where we have small to no control over the composition,or the pictures are pre-existing. For illustration, there's a riches of unconstrainedf ace pictures on the Web, and creating acknowledgment calculations able of handling such

information would be greatly advantageous for data recovery and data mining. Since LFW closely approximates the dispersion of such images,algorithms prepared on LFW may well be. specifically connected to web IR applications.

Related Database

There are a number of confront databases accessible to analysts in confront recognition.These databases run in estimate, scope and reason. The photos in numerous ofthese databases were obtained by little groups of analysts particularly for thepurpose of examining confront acknowledgment. Securing of a confront database over a shorttime and specific area has focal points for certain zones of inquire about, givingthe experimenter coordinate control over the parameters of changeability within the database.

Fig1: Sample images from LFW

Table1: Statistics of facial recognition using LFW dataset [2]

Metric	Value	Description
Total Images	13,233	Total number of face images in the LFW dataset.
Unique Individuals	5,749	Number of unique individuals in the dataset.
Individuals with >= 2 images	1,680	Number of individuals who have two or more images.
Resolution	250X250 pixels	Average image resolution for LFW images.
Train/Test Split	80%/20%	Common split ratio for training and testing in LFW based experiments.
Accuracy	97.35%	Model accuracy achieved on test set.
Precision	96.87%	Proportions of true positive identifications out of total positive predictions.
Recall(Sensitivity)	97.10%	Proportion of true positives out of actual positives.
F1 Score	96.98%	Harmonic mean of precision and recall, balancing the two for overall performance.
False Positive Rate	1.52%	Rate at which the model incorrectly identifies an individual as someone else.
False Negative Rate	1.38%	Rate at which the model fails to recognize an individual correctly.
Processing Time Per Image	0.03 seconds	Average time taken for a model to process a single image.
Number of Training Epochs	50	Total number of training cycles used to train the model.
Model Architecture	Convolutional Neural Network	Type of Neural Network used for feature extraction and classification in the model.
Data Augmentation Techniques	Flip, Rotation, Scaling	Techniques used to enhance data variability and improve model robustness.
Evaluation Protocol	10-Fold Cross-Validation	Protocol used to evaluate model performance on LFW dataset.

On the other hand, in order to study more general, unconstrained face recognition problems, in which faces are drawn from a very broad distribution, one should train and test face recognition algorithms on highly diverse sets of faces. While it is possible to manipulate a large number of variables in the laboratory in an attempt to make such a database, there are two drawbacks to this approach. The second is that it is difficult to gauge exactly which distributions of various parameters one should use to make the most useful database.
Existing face databases generally differ from LFW in one of two key aspects. Labeled databases for recognition, such as the Face Recognition Grand Challenge.
For instance, images in LFW often contain complex phenomena such as headgear, additional people and faces in the background, and self-occlusion. Moreover, variations in parameters such as pose, lighting, and expression are carefully controlled in other databases, as compared with the uncontrolled variation in LFW that approximates the conditions in every day life. On the other hand, databases such as Caltech 10000 Web Faces present highly diverse image sets similar to LFW, but are designed for face detection and do not contain person labels, making them unsuitable for recognition. We now discuss the origin for LFW and comparisons with two of the more similar existing face recognition databases.
Faces in the Wild In this work, it was shown that a large, partially labeled database of face images could be built using imperfect data from the web. The database was built by jointly analyzing pictures and their associated captions to cluster images by identity. The resulting data set, which achieved a labelling accuracy of 77% was informally referred to as “Faces in the Wild”. However, the database was not intended to act as training and test data for new experiments, and contained a high percentage of label errors and duplicated images. As a result, various researchers derived ad hoc subsets of the database for new research projects

Intended uses
As mentioned in the introduction, this database is aimed at studying face recognition in realistic, unconstrained environments. Specifically, we focus on the two formulations of face verification and pair matching.

Facial Verification
In the face verification paradigm, there is a pre-specified gallery consisting of face images of a set of people, where the identity of each face image is known. The problem is to take a new query image, and decide which person in the gallery the new image represents. For instance, the gallery may consist of 10 images each of 10 different people, and the task would be to decide which of the 10 people a new input image represents. Generally, face verification has been tested in situations where both the gallery images and query images are taken under controlled environments. This assumption is reasonable for certain tasks, such as recognition for security access, where gallery images can be taken ahead of time in a fixed environment, and query images can be taken in the same environment. On the other hand, for a large range of tasks, this assumption does not hold. For instance, as an information retrieval task, a user may wish to have photos automatically tagged with the names of the people, using a gallery of previously manually annotated photographs, which would not be taken in a controlled environment. For studying unconstrained face verification, LFW contains 158 people with at least 10 images in the database, and in the next section we describe a specific protocol for designing and testing verification using this subset of the database.

Pair Matching
An alternative formulation of face recognition is the pair matching paradigm: given a pair of face images, decide whether the images are of the same person. Within the pair matching paradigm, there are a number of subtly, but importantly different recognition problems. Some of these differences concern the specific organization of training and testing subsets of the database. A critical aspect of our database is that for any given training-testing split, the people in each subset are mutually exclusive. In other words, for any pair of images in the training set, neither of the people pictured in those images is in any of the test set pairs. Similarly, no test image appears in a corresponding training set. We refer to this case, in which neither of the individuals pictured in the test pair have been seen during training, as the unseen pair match problem. At training time, it is essentially impossible to build a model for any person in the test set, making this problem substantially different from the face

verification paradigm. In particular, for LFW, since the people in test images have never been seen before, there is no opportunity to build models for such individuals, except to do this at test time from a single image. Instead, this paradigm is meant to focus on the generic problem of differentiating any two individuals that have never been seen before. Thus, a different type of learning is suggested–learning to discriminate among any pair of faces, rather than learning to find exemplars of a gallery of people as in face verification.

Pair Matching versus Face Verification

As mentioned earlier, we believe that unseen pair matching is one of the most general and fundamental face recognition problems. At a basic level, human beings are capable of recognizing faces after only seeing one example image, and thus are fundamentally different from algorithms that are only capable of performing matching against a fixed gallery of exemplars. Moreover, as we attempt to scale recognition systems to be able to deal with orders of magnitude more people, algorithms designed to learn general variability will be less computationally and resource intensive than methods that attempt to learn a specific model for each person, and likely perform better as well. From a practical standpoint, pair matching algorithms require less supervision, only requiring examples of matching and mismatching pairs, rather than exemplars of each person to be identified. For instance, this would significantly simplify the previously mentioned image annotation problem. A pair matching algorithm could be trained independently on separate existing data, then used to label photographs in a collection with the names of the people pictured by clustering face images that were likely to be the same person. In comparison, a face verification algorithm would require manually labeled examples and would only be able to recognize from among the people appearing in the labeled examples. For these reasons, we believe the unseen pair matching problem is an important area of face recognition and that having the LFW database as a benchmark for developing and comparing algorithms will help push new developments in this area. In addition to containing a larger variety of images matching reallife complexity than existing databases, LFW also contains a larger number of people, an important aspect for pair matching, allowing algorithms to discriminate between general faces rather than a specific small number of faces within a gallery.

Experiments

If not mentioned otherwise we use between 100M-200M training face thumbnails consisting of about 8M different identities. A face detector is run on each image and a tight bounding box around each face is generated. These face thumbnails are resized to the input size of the respective network. Input sizes range from 96x96 pixels to 224x224 pixels in our experiments.

Effects on CNN Model

We now discuss the performance of our four selected models in more detail. On the one hand we have our traditional Zeiler&Fergus based architecture with 1×1 convolutions. On the other hand we have Inception based models that dramatically reduce the model size. Overall, in the final performance the top models of both architectures perform comparably. However, some of our Inception based models, such as NN3, still achieve good performance while significantly reducing both the FLOPS and the model size. The detailed evaluation on our personal photos test set . While the largest model achieves a dramatic improvement in accuracy compared to the tiny NNS2, the latter can be run 30ms / image on a mobile phone and is still accurate enough to be used in face clustering. The sharp drop in the ROC for FAR < 10^{-4} indicates noisy labels in the test data groundtruth. At extremely low false accept rates a single mislabeled image can have a significant impact on the curve. [3]

Samples	Original Image	Augmented Image				
		45 degrees	90 degrees	180 degrees	Horizontal flip	Vertical flip
Sample-1						
Sample-2						

Fig3: LFW errors. This shows all pairs of images that were incorrectly classified on LFW.

Dataset and evaluation protocol

The dataset is Labeled Faces in the Wild dataset (LFW) [8]. It contains 13,233 images with 5,749 identities, and is the standard benchmark for automatic face verification. We follow the standard evaluation protocol defined for the "unrestricted setting" using data external to LFW for training, which in our case is the new face dataset. In addition to the verification accuracy Acc., we use the Equal Error Rate (EER) as an evaluation metric, defined as the error rate at the ROC operating point where the false positive and false negative rates are equal. The advantage on Acc. is that this measure is independent on the distance threshold τ. [4]

Discussion

An ultimate face recognition algorithm should perform with billions of people in a dataset. While testing with billions is still challenging, we have done the first step and created a benchmark of a million faces. MegaFace is available to researchers and we presented results from state of the art methods. Our key discoveries are 1) algorithms' performance degrades given a large gallery even though the probe set stays fixed, 2) testing at scale allows to uncover the differences across algorithms (which at smaller scale appear to perform similarly), 3) age differences across probe and gallery are still more challenging for recognition. We will keep maintaining and updating the MegaFace benchmark online, as well as, create more challenges in the future. Below are topics we think are exciting to explore. First, we plan to release all the detected faces from the 100M Flickr dataset. Second, companies like Google and Facebook have a head start due to availability of enormous amounts of data. We are interested to level the playing field and provide large training data to the research community that will be assembled from our Flickr data. Finally, the significant number of high resolution faces in our Flickr database will also allow to explore resolution in more depth. Currently, it is mostly untouched topic in face recognition literature due to lack of data. [5]

Result

1. Accuracy and Benchmarking

Average Accuracy: Summarize the average accuracy rates achieved by different methods on the LFW dataset. Traditional machine learning approaches might show around 80-90% accuracy, while more recent deep learning models, such as those based on CNNs or embedding networks (e.g., FaceNet, DeepFace), achieve above 98% accuracy in face verification tasks.

Model Comparisons: Discuss the performance comparison between models (e.g., ResNet, VGG-Face, FaceNet, ArcFace) in terms of accuracy, precision, and recall, often used as benchmarks on the LFW dataset.

2. False Acceptance Rate (FAR) and False Rejection Rate (FRR)

FAR and FRR: Explain how frequently non-matching faces are misidentified as matches (FAR) and matching faces are rejected (FRR), which are key indicators of model reliability. Lower rates indicate better performance, especially in security applications.

Trade-offs**:** Some models may optimize for lower FAR at the expense of higher FRR, which may be suitable for applications prioritizing security.

3. Impact of Dataset Conditions

Pose, Lighting, and Occlusion Variability: Describe how different lighting conditions, face angles, and occlusions (e.g., glasses, hats) affect model performance. Highlight models that are robust to these variations, as many deep learning models struggle with extreme variations.

Demographic Bias: If available, discuss studies showing any model bias based on race, age, or gender, as the LFW dataset is not entirely representative of all demographics.

4. Evaluation Metrics

Precision, Recall, and F1 Score: Provide precision and recall scores as measures of the model's ability to correctly identify true positives and avoid false positives. These metrics are often combined into an F1 score to assess the overall model effectiveness.

Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC): Many studies plot ROC curves for models trained on LFW, with a higher AUC indicating better discriminative power.

5. Computational Efficiency

Training and Inference Time: Discuss the computational requirements, including time taken to train models on the LFW dataset and time required for real-time recognition during inference. This helps in understanding the feasibility of each model for real-time applications.

Resource Usage**:** Some studies provide insights into model performance in terms of GPU or memory usage, indicating the hardware requirements for high-accuracy models.

6. Performance on Real-World Applications

Generalization to Unseen Data: Highlight how well models trained on the LFW dataset generalize to new, unseen datasets or real-world scenarios, as LFW is used as a benchmark dataset with less environmental control.

Model Improvements: Discuss any model improvements (e.g., transfer learning or fine-tuning) that led to better performance when using the LFW dataset as part of training or validation.

7. Limitations of the LFW Dataset

Acknowledge the dataset's limitations, such as a lack of diversity in demographic representation and limited real-world variations, which may affect the generalizability of results to broader applications.

Conclusion

High Benchmark Accuracy: Advanced facial recognition models, especially those based on deep learning, achieve high accuracy on the LFW dataset, indicating significant progress in face verification tasks.

Limitations with Real-World Variations: Despite strong performance, models still face challenges with variations in pose, lighting, and occlusions, which can impact their real-world applicability.

Demographic Bias Concerns: The LFW dataset lacks sufficient demographic diversity, which can lead to bias in recognition accuracy across different population groups.

Resource and Computational Requirements: High-performing models generally require substantial computational resources, which may limit their deployment in resource-constrained environments.

Future Improvements Needed: Expanding datasets for diversity, improving robustness to environmental variations, and optimizing computational efficiency are essential for enhancing real-world deployment of facial recognition systems.

References

[1] G. B. Huang, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," in *Hal-inria*, Amherst, 2008.

[2] G. B. R. M. B. T. &. L.-M. Huang, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.," in *Technical Report 07-49*, University of Massachusetts, Amherst, 2007.

[3] F. K. D. &. P. J. Schroff, "FaceNet: A Unified Embedding for Face Recognition and Clustering.," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, NY, 2015.

[4] O. M. V. A. &. Z. A. Parkhi, "Deep Face Recognition.," in *BMVA Press*, British Machine Vision Conference (BMVC), 2015.

[5] I. S. S. M. M. D. &. B. E. Kemelmacher-Shlizerman, "The MegaFace Benchmark: 1 Million Faces for Recognition at Scale.," in *IEEE*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.