

Facial Verification Using a Siamese Neural Network

Winters in Data Science

Name: Suryangshu Das

Roll Number: 23B2402

Date of Submission: January 30, 2026

Abstract

Facial verification has emerged as one of the most impactful applications of deep learning in computer vision, enabling secure and seamless identity authentication across a wide range of real-world systems. Unlike conventional face recognition approaches that attempt to classify an image into one of several predefined identities, facial verification addresses a fundamentally different problem: determining whether two facial images belong to the same individual. This distinction significantly improves scalability and robustness in practical deployments.

This project presents the complete development of a facial verification system using a Siamese Neural Network architecture. The work emphasizes not only the final performance of the model, but also the learning outcomes achieved throughout the project lifecycle. Key concepts explored include convolutional neural networks, metric learning, contrastive loss optimization, dataset preparation strategies, and evaluation methodologies specific to verification tasks. Through iterative experimentation and analysis, the project provided valuable insights into the design and deployment of deep learning systems for real-world identity verification.

Contents

1	Introduction	3
2	Background and Theoretical Foundations	5
2.1	Facial Verification versus Face Recognition	5
2.2	Metric Learning	5
2.3	Siamese Neural Networks	5
2.4	Contrastive Loss Function	5
3	Dataset Preparation and Preprocessing	6
4	Model Architecture and Design Decisions	8
5	Training Methodology and Optimization	10
6	Evaluation Strategy	12
7	Results and Discussion	14
8	Overall Learning Outcomes	16
9	Conclusion	17

1 Introduction

Facial analysis technologies have witnessed rapid and sustained advancements over the past decade, primarily driven by breakthroughs in deep learning and the availability of large-scale visual datasets. Modern convolutional neural networks have demonstrated an exceptional ability to extract hierarchical and highly discriminative features from images, enabling machines to perform complex visual recognition tasks that were previously considered infeasible. Among the wide range of facial analysis applications enabled by these developments, facial verification has emerged as a particularly important and impactful problem domain.

Facial verification refers to the task of determining whether two given facial images belong to the same individual. This problem formulation differs fundamentally from traditional face recognition systems, which typically attempt to classify an input image into one of several predefined identity classes. Classification-based face recognition approaches are inherently limited in scalability, as they require retraining or architectural modification when new identities are introduced. In contrast, facial verification systems are identity-agnostic; they focus on learning a notion of similarity between faces rather than explicit identity labels. This property makes facial verification especially suitable for real-world applications such as biometric authentication, access control systems, financial security mechanisms, and large-scale surveillance environments, where the set of possible identities may be unknown or continuously evolving.

The increasing reliance on facial verification systems in sensitive applications further emphasizes the need for models that are both accurate and robust. Variations in illumination, facial expression, pose, occlusion, and image quality introduce significant challenges in designing reliable verification systems. Addressing these challenges requires not only effective model architectures but also carefully designed training strategies, high-quality datasets, and evaluation methodologies that align with real-world deployment conditions. Understanding these aspects in depth formed a central motivation for this project.

The primary technical objective of this project was to design, implement, and evaluate a facial verification system based on a Siamese Neural Network architecture. Siamese networks are particularly well-suited for verification tasks because they learn to map input images into a shared embedding space, where the distance between embeddings reflects semantic similarity. By comparing feature representations rather than relying on explicit class predictions, Siamese networks enable flexible and scalable verification even for identities not encountered during training. Implementing such a system provided valuable exposure to metric learning principles and contrastive optimization techniques.

Beyond the goal of building a functional facial verification model, this project was intentionally approached as a comprehensive learning exercise in applied deep learning. Each stage of the development pipeline—ranging from data collection and preprocessing

to model architecture design, training, and evaluation—was treated as an opportunity to develop both theoretical understanding and practical intuition. The process of organizing datasets into meaningful training pairs, selecting appropriate loss functions, tuning hyperparameters, and diagnosing training instabilities contributed significantly to a deeper appreciation of how deep learning systems behave in practice.

A key learning outcome of the project was the realization that model performance is influenced as much by data quality and experimental design as by architectural complexity. Through iterative experimentation, it became evident that careful preprocessing, dataset diversity, and evaluation strategy often have a greater impact on verification accuracy than simply increasing network depth or parameter count. This insight reflects an important engineering principle: effective system design requires a balanced consideration of all components in the pipeline rather than an exclusive focus on model architecture.

This report documents the complete technical methodology employed in the project while placing significant emphasis on the learning achieved at each stage of development. Rather than presenting results in isolation, the report aims to articulate the reasoning behind design decisions, the challenges encountered during implementation, and the insights gained through experimentation and analysis. The intention is to demonstrate conceptual clarity, critical thinking, and sound engineering judgment, thereby providing a holistic account of both the system developed and the learning process underlying its creation.

2 Background and Theoretical Foundations

2.1 Facial Verification versus Face Recognition

A key conceptual learning early in the project was the distinction between face recognition and facial verification. Face recognition treats the problem as a multi-class classification task, where the model predicts the identity label associated with an input image. This approach becomes increasingly inefficient as the number of identities grows and requires retraining when new identities are introduced.

Facial verification, in contrast, reframes the problem as a similarity estimation task. The model learns to determine whether two images belong to the same individual based on learned feature representations. This paradigm shift was instrumental in understanding why Siamese networks are well-suited for verification tasks.

2.2 Metric Learning

Metric learning focuses on learning a distance function that captures semantic similarity between inputs. Through this project, the importance of embedding spaces became evident. Instead of relying on explicit class labels, the network learns a representation where similar faces are closer together and dissimilar faces are far apart. This concept is particularly powerful for open-set problems such as facial verification.

2.3 Siamese Neural Networks

A Siamese Neural Network consists of two identical subnetworks that share weights and parameters. Each subnetwork processes one of the input images and outputs a feature embedding. These embeddings are then compared using a distance metric. The shared-weight structure ensures that both inputs are processed identically, enforcing consistency in feature extraction. Understanding this architectural symmetry was a major theoretical takeaway from the project.

2.4 Contrastive Loss Function

The contrastive loss function plays a central role in Siamese networks by penalizing incorrect similarity predictions. Positive pairs are encouraged to have small distances, while negative pairs are pushed apart beyond a defined margin. Experimenting with contrastive loss provided insights into how loss functions shape the geometry of learned embedding spaces.

3 Dataset Preparation and Preprocessing

Dataset preparation emerged as one of the most critical and time-intensive components of this project. While model architecture and training strategies are often emphasized in deep learning workflows, this project reinforced the understanding that the quality, structure, and diversity of the dataset play a decisive role in determining overall system performance. In the context of facial verification, improper dataset construction can lead to misleading performance metrics and poor generalization, even when sophisticated neural network architectures are employed.

The dataset was explicitly structured to support Siamese network training by organizing images into anchor, positive, and negative pairs. For each anchor image representing a given individual, a corresponding positive image was selected from the same identity, while a negative image was chosen from a different identity. This triplet-style organization is essential for metric learning, as it allows the network to directly learn similarity and dissimilarity relationships rather than absolute class labels. Designing this structure required careful planning to ensure that the network was exposed to a balanced and representative set of genuine and impostor comparisons during training.

Class balance was a particularly important consideration during dataset construction. An imbalance between positive and negative pairs can bias the model toward trivial solutions, such as consistently predicting dissimilarity. To mitigate this issue, deliberate effort was made to maintain a roughly equal distribution of positive and negative samples across training batches. This process highlighted the importance of dataset statistics and reinforced the need for systematic data validation before model training.

Image collection was performed with an emphasis on variability to enhance the robustness of the verification system. Facial images were captured or selected under varying lighting conditions, facial expressions, head poses, and minor occlusions. These variations reflect real-world deployment scenarios, where environmental and contextual factors cannot be controlled. Through experimentation, it became evident that datasets lacking sufficient variation tend to produce models that overfit to superficial features, resulting in poor generalization when evaluated on unseen data. This observation underscored the necessity of incorporating diversity at the data level rather than relying solely on regularization techniques at the model level.

Preprocessing constituted a crucial step in ensuring consistent and stable model training. All images were resized to a fixed spatial resolution to maintain uniform input dimensions across the dataset. This resizing operation also reduced computational complexity while preserving essential facial features. Pixel values were normalized to a common scale, typically within the range of zero to one, to prevent numerical instability during gradient-based optimization. Normalization was observed to significantly improve convergence behavior and reduce sensitivity to learning rate selection.

In addition to resizing and normalization, color format conversion was applied where necessary to ensure consistency across the dataset. Converting images to a standardized color representation eliminated discrepancies arising from differing image sources and formats. These preprocessing steps collectively ensured that the input distribution presented to the neural network remained stable across training iterations, thereby facilitating effective feature learning.

A key learning outcome from this stage of the project was the realization that preprocessing decisions directly influence the geometry of the learned embedding space. Even minor inconsistencies in input preparation were found to adversely affect the separability of genuine and impostor pairs. As a result, preprocessing was treated as an integral component of the model pipeline rather than a peripheral step.

Overall, the dataset preparation and preprocessing phase reinforced a central principle of applied machine learning: model performance is fundamentally constrained by data quality. Through hands-on experimentation and iterative refinement, this project demonstrated that careful dataset design and rigorous preprocessing often yield greater performance improvements than increasing model complexity. This insight proved invaluable in shaping subsequent design decisions throughout the project.

4 Model Architecture and Design Decisions

The core of the facial verification system developed in this project is the embedding network, which was implemented using a convolutional neural network (CNN). The primary role of the embedding network is to transform raw facial images into compact, fixed-length feature representations that capture the most salient and discriminative characteristics of a face. Designing this network required careful consideration of both representational capacity and computational efficiency, as the quality of the learned embeddings directly determines the effectiveness of the verification process.

Convolutional neural networks were chosen due to their proven effectiveness in visual recognition tasks. The hierarchical feature learning capability of CNNs allows early layers to capture low-level spatial patterns such as edges, corners, and textures, while progressively deeper layers learn higher-level abstractions such as facial components, relative spatial arrangements, and identity-specific structures. Through experimentation, it became clear that this hierarchical representation is particularly well-suited for facial data, where subtle spatial relationships play a critical role in distinguishing between individuals.

The embedding network architecture was deliberately designed to balance complexity and generalization. While deeper networks with a large number of parameters can potentially learn highly expressive representations, they also increase the risk of overfitting, especially when training data is limited. By selecting a moderate-depth architecture with carefully chosen convolutional and pooling layers, the network was able to learn meaningful facial embeddings without excessive computational overhead. This design process reinforced the importance of aligning model capacity with dataset size and diversity.

The output of the embedding network is a fixed-dimensional feature vector that represents each facial image in a learned embedding space. These embedding vectors are designed to be both compact and discriminative, enabling efficient similarity computation during verification. A key learning outcome at this stage was the realization that the absolute dimensionality of the embedding is less important than the quality and structure of the learned feature space. Well-separated embeddings facilitate reliable verification even under challenging conditions such as variations in pose and illumination.

To quantify similarity between two facial embeddings, the absolute difference between the corresponding feature vectors was computed. This simple distance-based formulation was chosen over more complex similarity functions due to its interpretability and effectiveness. By focusing on element-wise differences, the model is encouraged to emphasize consistent identity-specific features while suppressing irrelevant variations. Empirical observations during training confirmed that this approach yields stable convergence and robust performance.

The choice of using a simple distance metric also reflects an important design phi-

losophy emphasized throughout the project: well-motivated simplicity often outperforms unnecessarily complex solutions. While more elaborate comparison mechanisms could be introduced, such as additional fully connected similarity networks, these often increase training instability and susceptibility to overfitting without providing commensurate performance gains. The effectiveness of the absolute difference operation demonstrated that carefully chosen architectural simplicity can enhance both interpretability and generalization.

Overall, the model architecture and design decisions adopted in this project were guided by a combination of theoretical understanding and empirical validation. Through iterative experimentation and analysis, the embedding network evolved into a balanced and efficient architecture capable of learning meaningful facial representations. This stage of the project significantly deepened understanding of how architectural choices influence representation learning, optimization behavior, and downstream verification performance.

5 Training Methodology and Optimization

Training the Siamese neural network constituted one of the most challenging and instructive phases of the project, as it required careful consideration of optimization dynamics and training stability. Unlike standard classification models, Siamese networks rely on learning relative similarity relationships between pairs of inputs, making the training process particularly sensitive to batch composition, loss behavior, and hyperparameter selection. As a result, significant attention was devoted to understanding and controlling the factors that influence convergence.

A critical aspect of the training methodology was batch construction. Each training batch was designed to contain a balanced mixture of positive and negative image pairs to ensure that the network received consistent and informative gradient signals. Improper batch composition was observed to bias the model toward trivial solutions, such as predicting dissimilarity for most input pairs. Through experimentation, it became evident that maintaining balance within batches was essential for stable and meaningful optimization, reinforcing the importance of thoughtful data pipeline design in metric learning tasks.

Learning rate selection played a central role in determining training behavior. Initial experiments with relatively high learning rates led to unstable loss oscillations and poor convergence, while excessively low learning rates resulted in slow training and suboptimal embedding quality. By iteratively adjusting the learning rate and observing its impact on loss curves, a suitable operating range was identified that allowed steady convergence without sacrificing training stability. This process highlighted the strong coupling between optimization hyperparameters and the geometry of the learned embedding space.

Training duration and the number of epochs were also carefully controlled to avoid overfitting. Prolonged training often resulted in diminishing returns, where validation performance stagnated or degraded despite continued improvement in training loss. Monitoring both training and validation loss curves enabled early identification of overfitting behavior, prompting adjustments to training duration and regularization strategies. This practice emphasized the importance of continuous performance monitoring rather than relying solely on predefined training schedules.

Throughout the training process, loss curves were closely examined to diagnose common optimization issues such as vanishing gradients, exploding gradients, and mode collapse. Observing the behavior of the contrastive loss over time provided valuable insight into how effectively the model was learning to separate genuine and impostor pairs. Smooth and steadily decreasing loss trajectories were indicative of stable learning, while abrupt fluctuations signaled the need for hyperparameter refinement.

Hyperparameter tuning was conducted in an empirical and iterative manner, reflecting the experimental nature of deep learning workflows. Parameters such as batch size,

learning rate, margin values in the loss function, and optimizer configuration were adjusted systematically based on observed training behavior. This process reinforced a key learning outcome of the project: optimal performance is rarely achieved through theoretical reasoning alone and often requires careful experimentation guided by empirical evidence.

An important insight gained during this stage was the relationship between optimization parameters and model generalization. Hyperparameters that produced rapid convergence did not always yield the most discriminative embeddings, while more conservative optimization settings often resulted in better separation between genuine and impostor pairs. This trade-off underscored the necessity of aligning optimization objectives with the ultimate goal of verification performance rather than purely minimizing training loss.

Overall, the training methodology and optimization phase provided a deep understanding of how Siamese networks learn similarity representations in practice. The challenges encountered and resolved during this stage significantly strengthened intuition regarding optimization dynamics, experimental design, and performance evaluation in deep learning systems. These lessons represent a foundational learning outcome of the project and are directly transferable to a wide range of metric learning and representation learning tasks.

6 Evaluation Strategy

Evaluating a facial verification system requires a fundamentally different approach compared to evaluating traditional classification models. While classification accuracy provides a straightforward measure of performance for closed-set recognition tasks, it is insufficient for verification scenarios where the objective is to assess similarity between image pairs. A central learning outcome of this project was understanding the limitations of conventional accuracy metrics and the necessity of adopting evaluation strategies that align with the real-world objectives of verification systems.

The primary evaluation mechanism employed in this project involved computing similarity scores between facial embeddings generated by the Siamese network. These similarity scores were then compared against a predefined threshold to determine whether a given image pair should be classified as a genuine match or an impostor pair. Selecting an appropriate similarity threshold emerged as one of the most critical aspects of the evaluation process, as it directly influences system behavior in deployment settings.

Threshold tuning highlighted the inherent trade-off between false acceptance rates and false rejection rates. A lower threshold increases system sensitivity, reducing false rejections but potentially increasing false acceptances, which can compromise security. Conversely, a higher threshold improves security by reducing false acceptances but may lead to an increased number of false rejections, negatively impacting user experience. Through systematic experimentation, the project demonstrated that threshold selection must be guided by application-specific requirements rather than a single universal metric.

To better understand this trade-off, evaluation was conducted across a range of threshold values, and the resulting performance trends were analyzed. Observing how genuine and impostor similarity distributions evolved with training provided valuable insight into the discriminative power of the learned embeddings. Well-separated distributions indicated effective representation learning, while significant overlap signaled the need for further optimization. This analysis reinforced the importance of visual and statistical inspection alongside numerical metrics.

In addition to quantitative evaluation, qualitative assessment played an important role in validating model behavior. Visual inspection of genuine and impostor image pairs, along with their corresponding similarity scores, helped identify systematic failure cases such as pose mismatches, poor illumination, or partial occlusions. These observations provided actionable feedback for refining dataset composition and preprocessing strategies. This stage emphasized that human-in-the-loop evaluation remains a valuable tool, even in highly automated deep learning workflows.

Another important learning outcome was recognizing that evaluation metrics must reflect real-world deployment objectives. In security-critical applications, minimizing false acceptances may be prioritized over overall accuracy, whereas user-facing authentication

systems may tolerate occasional false acceptances to reduce user friction. This realization underscored the necessity of aligning evaluation criteria with operational requirements rather than relying solely on generic performance measures.

Overall, the evaluation strategy adopted in this project provided a comprehensive assessment of the Siamese network's verification capabilities. By combining threshold-based analysis, distribution inspection, and qualitative evaluation, the project achieved a nuanced understanding of model performance. This stage significantly deepened appreciation for the complexities involved in evaluating verification systems and highlighted the importance of context-aware evaluation in applied machine learning.

7 Results and Discussion

The trained Siamese neural network demonstrated consistent and reliable facial verification performance when evaluated on previously unseen image pairs. The model was able to correctly distinguish between genuine pairs, in which both images belonged to the same individual, and impostor pairs, corresponding to different identities. This behavior indicates that the network successfully learned a meaningful embedding space in which facial similarity is effectively captured.

Analysis of the learned embeddings revealed clear clustering behavior, with images of the same individual mapping to nearby regions in the embedding space while images of different individuals were separated by larger distances. This separation is a critical requirement for verification systems, as it enables robust threshold-based decision making. The emergence of such structured clustering patterns confirmed that the contrastive training objective was effective in shaping the geometry of the embedding space. Observing this phenomenon provided tangible evidence of how metric learning principles translate into practical system behavior.

Quantitative evaluation further supported these observations, with similarity score distributions for genuine and impostor pairs exhibiting increasing separation as training progressed. This trend indicated improved discriminative power of the embeddings over time. Importantly, performance gains were not limited to the training dataset; the model generalized effectively to unseen image pairs, suggesting that overfitting was successfully mitigated through careful dataset preparation, architectural design, and training strategy.

A key insight from the results was the strong interdependence between data diversity, loss function design, and evaluation methodology. Models trained on datasets with limited variation showed reduced generalization capability, even when architectural complexity was increased. In contrast, improving dataset diversity through variations in lighting, pose, and expression led to substantial performance gains without modifying the network architecture. This observation reinforced the understanding that data quality often serves as the primary bottleneck in real-world machine learning systems.

The choice of contrastive loss and distance-based similarity measurement also played a decisive role in shaping verification performance. Proper margin selection and balanced training batches were found to be essential for achieving clear separation between genuine and impostor embeddings. These findings highlighted that loss design is not merely a mathematical detail, but a core component that directly influences system behavior and interpretability.

Evaluation methodology further influenced the interpretation of results. Threshold selection was shown to significantly affect observed performance metrics, emphasizing that reported accuracy values must be contextualized within the intended deployment scenario. This realization underscored the importance of holistic evaluation strategies

that consider operational constraints rather than relying on a single aggregate metric.

Overall, the results of this project demonstrated that effective facial verification performance arises from the careful integration of multiple system components rather than isolated optimization of any single element. The holistic understanding gained through analyzing these results represents one of the most valuable outcomes of the project. By examining how data preparation, model architecture, training dynamics, and evaluation strategy interact, this project provided a comprehensive perspective on the design of robust verification systems in practice.

8 Overall Learning Outcomes

This project contributed significantly to strengthening both theoretical understanding and practical proficiency in deep learning. Through hands-on implementation and experimentation, abstract concepts that are often introduced at a high level in coursework became concrete and intuitive. In particular, the project provided a clear understanding of how similarity-based learning differs from traditional classification paradigms and why verification systems require fundamentally different modeling and evaluation strategies.

One of the most important learning outcomes was gaining practical experience in designing neural network architectures for real-world computer vision tasks. The process of constructing and refining the embedding network highlighted the impact of architectural choices on representation quality, training stability, and generalization. This experience reinforced the importance of aligning model complexity with dataset characteristics and deployment constraints, an essential skill for applied machine learning engineers.

Handling real-world data variability emerged as another critical area of learning. Unlike idealized datasets commonly used in academic examples, facial images in this project exhibited significant variation in lighting conditions, facial expressions, pose, and image quality. Addressing these variations required careful dataset curation, preprocessing, and evaluation. Through this process, the project emphasized that robust model performance depends not only on algorithmic sophistication but also on thoughtful data engineering practices.

The project also provided valuable experience in evaluating non-classification models. Traditional accuracy metrics were found to be insufficient for assessing verification performance, leading to a deeper appreciation of threshold-based evaluation, similarity score distributions, and trade-offs between false acceptance and false rejection rates. Developing an evaluation strategy aligned with real-world objectives enhanced understanding of how machine learning systems are assessed beyond benchmark datasets.

Equally important were the practical skills developed throughout the project, particularly in debugging and experimental discipline. Training instability, unexpected loss behavior, and suboptimal performance required systematic troubleshooting and iterative refinement. These challenges fostered a disciplined experimental approach involving controlled parameter changes, careful monitoring of training dynamics, and critical analysis of results.

Finally, the project strengthened the ability to critically analyze model behavior and results rather than accepting performance metrics at face value. By interpreting embedding distributions, examining failure cases, and relating observed behavior to design decisions, a more nuanced understanding of model performance was achieved. This analytical mindset represents a lasting learning outcome of the project and forms a strong foundation for future work in machine learning research and applied artificial intelligence.

9 Conclusion

This project successfully demonstrated the application of Siamese Neural Networks to the task of facial verification, addressing a real-world computer vision problem that extends beyond conventional classification-based approaches. By learning similarity-based representations rather than explicit identity labels, the developed system exhibited the flexibility and scalability required for practical deployment scenarios. The implementation confirmed that Siamese architectures, when combined with appropriate loss functions and training strategies, provide an effective framework for verification tasks.

More importantly, the project served as a comprehensive learning experience that bridged the gap between theoretical understanding and practical implementation. Concepts such as metric learning, contrastive loss optimization, and embedding space geometry, which are often introduced abstractly in coursework, were explored in depth through hands-on experimentation. This process transformed theoretical knowledge into practical intuition, enabling a clearer understanding of how deep learning models behave in realistic settings.

The project also highlighted the interconnected nature of system components in applied machine learning. Dataset preparation, model architecture, training methodology, and evaluation strategy were found to collectively determine system performance. Improvements in one component often required corresponding adjustments in others, reinforcing the importance of a holistic and systems-oriented approach to model design. This perspective represents a key intellectual outcome of the project.

In addition to technical proficiency, the project fostered essential research and engineering skills, including experimental discipline, systematic debugging, and critical analysis of results. Encountering and resolving training instability, performance bottlenecks, and evaluation challenges strengthened problem-solving abilities and promoted a deeper appreciation for the iterative nature of machine learning development.

The insights gained through this project provide a strong foundation for future work in computer vision, machine learning research, and applied artificial intelligence systems. The methodologies and learning outcomes developed here are directly transferable to more advanced verification tasks, larger-scale datasets, and real-time deployment scenarios. Overall, this project represents not only a successful technical implementation but also a meaningful step toward developing the analytical and practical skills required for advanced work in artificial intelligence.

References

1. Koch, G., Zemel, R., & Salakhutdinov, R., “Siamese Neural Networks for One-shot Image Recognition,” ICML Workshop, 2015.
2. Schroff, F., Kalenichenko, D., & Philbin, J., “FaceNet: A Unified Embedding for Face Recognition and Clustering,” CVPR, 2015.
3. Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016.