iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59290

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Incremental Learning-Enhanced Ensemble Convolutional Neural Network for Robust Pose-Invariant Face Recognition

Appala Prapul Nivas [1], Gumperla Karthik[2], Vaddempudi Charan Teja[3], Golla Punith Chand[4], Venkata Satya Sai Varun Dudipala[5]
[1, 2, 3, 4, 5]*B.Tech Graduate (Cse AI ML) GITAM University, Bengaluru, India*

Abstract: ***In the dynamic realm of face recognition, where challenges such as pose variations, illumination changes, and expression nuances abound, this research embarks on a novel journey toward enhancing recognition robustness. Herein lies an amalgamation of innovative methodologies aimed at transcending the boundaries of traditional face recognition paradigms. At the forefront, an Adaptive Morphological Bilateral Filtering (AMBF) technique emerges, poised to revolutionize image preprocessing. AMBF stands as a beacon of adaptability, seamlessly sharpening blurry facial images while mitigating the risks of introducing undesirable artifacts. Its arsenal of morphological operations, carefully orchestrated to uplift image quality, presents a promising solution to the perennial challenge of recognizing faces under adverse conditions. Complementing this pioneering filtering approach is the integration of an Ensemble Convolutional Neural Network (ECNN), an avant-garde ensemble learning architecture. Drawing strength from diversity, ECNN harnesses the collective intelligence of multiple networks to discern intricate facial patterns amidst the noise. Its ability to synthesize varied perspectives enables a more robust and nuanced understanding of facial features, laying the groundwork for superior recognition accuracy. Furthermore, the research delves into uncharted territory with the incorporation of Incremental Learning, a forward-looking strategy for continuous model refinement. By dynamically adapting to new data while retaining past knowledge, Incremental Learning fortifies the face recognition system against the ever-evolving landscape of facial variations. This adaptive capability ensures the system's relevance and efficacy over time, cementing its position as a stalwart guardian of identity verification. Through extensive experimentation on the multi-PIE dataset, the efficacy of these novel methodologies is unequivocally demonstrated. The face recognition system, fortified by the fusion of AMBF, ECNN, and Incremental Learning, emerges triumphant in the face of adversity, promising a brighter future where recognition barriers are shattered, and identity verification transcends limitations.***

I. INTRODUCTION

In the era of digital transformation, the quest for reliable and efficient biometric authentication systems has intensified, with face recognition emerging as a cornerstone technology in identity verification applications. Despite significant advancements in the field, traditional face recognition algorithms continue to grapple with the challenges posed by variations in pose, illumination conditions, and facial expressions, collectively known as PIE variations. These variations introduce complexities that hinder accurate recognition and limit the applicability of face recognition systems in real-world scenarios. Deep learning offers a natural approach to extracting feature representations from data without the need for explicit descriptors. Biometric security systems encounter significant challenges due to the variability in intra-personal facial appearance stemming from factors such as lighting conditions, facial expressions, and aging [1]. While humans effortlessly recognize faces, translating this capability into computer vision remains a formidable task. Automating pattern recognition tasks, particularly facial recognition, holds significant practical value, especially in the realms of artificial intelligence (AI) and data analytics, with numerous real-world applications benefiting from enhanced transparency and efficiency in data processing. Human face detection stands as a crucial component in video surveillance and facial recognition systems, serving as a cornerstone technology in various domains [2].

To address these challenges, researchers have explored various methodologies aimed at enhancing the robustness and adaptability of face recognition systems. Among these methodologies, two promising approaches have garnered considerable attention: adaptive morphological filtering and ensemble learning techniques.

©IJRASET: All Rights are Reserved

Adaptive morphological filtering techniques, rooted in mathematical morphology, offer a powerful means of enhancing image quality by selectively modifying pixel intensities based on the local structure of the image. By leveraging morphological operations such as erosion, dilation, opening, and closing, adaptive filters can effectively mitigate the effects of blur, noise, and other distortions commonly encountered in facial images. These filters exhibit a remarkable ability to sharpen blurry images without introducing undesirable artifacts, making them particularly well-suited for preprocessing facial images under challenging conditions. This paper addresses the challenge of recognizing faces across images with varying levels of blurriness and brightness. A straightforward approach to addressing blurred faces is through image deblurring, which entails tackling the complex problem of blind image deconvolution. Additionally, the impact of face alignment using 3D techniques, leveraging larger datasets, and employing new metric learning algorithms on the performance of face recognition in convolutional neural network (CNN)-based methods is investigated [3][4]. Traditionally, face recognition has been treated as a single problem in existing techniques. However, this paper explores the concept of multi-task learning (MTL) for face recognition, wherein multiple tasks are simultaneously learned to enhance the performance of the main task. Successful applications of MTL have been observed in attribute estimation, pedestrian detection, face alignment, and face detection, leading to improved results in vision problems [5][6][7]. In parallel, ensemble learning techniques have emerged as a potent tool for improving the robustness and generalization capabilities of face recognition systems. Ensemble methods leverage the collective wisdom of multiple base classifiers to achieve superior performance compared to individual classifiers. By combining diverse classifiers trained on different subsets of data or using different algorithms, ensemble models can effectively capture the complex relationships inherent in facial data, thereby enhancing recognition accuracy and resilience to variations.

In this study, face recognition is approached as a multi-task problem, with the main task focusing on identity classification and side tasks involving the estimation of pose, illumination, and expression (PIE). To address the challenge of blurred facial images, an Adaptive Morphological Bilateral Filtering (AMBF) approach is proposed to enhance the input face images by removing noise present in the training dataset. Furthermore, a multitask ensemble convolutional neural network (ECNN) approach is introduced to jointly characterize personality traits and PIE attributes by aggregating various poses to learn distinct facial features. Finally, a dynamic-weighting scheme based on the bat algorithm is implemented to automatically assign loss weights to each side task, addressing the fundamental issue of balancing between multiple tasks in multitask learning with ECNN. In this paper, we propose a novel approach that integrates adaptive morphological filtering with ensemble convolutional neural networks (ECNNs) to address the challenges of pose-invariant face recognition. Our methodology aims to synergistically leverage the strengths of both adaptive filtering and ensemble learning to enhance the robustness, accuracy, and adaptability of face recognition systems in real-world scenarios. By seamlessly integrating these techniques, we seek to push the boundaries of face recognition performance and pave the way for more reliable and effective biometric authentication solutions.

II. RELATED WORK

Face recognition is a fundamental task in computer vision and biometrics, with widespread applications in security, surveillance, access control, and human-computer interaction. Over the years, researchers have explored various techniques and methodologies to improve the robustness and accuracy of face recognition systems, particularly in the presence of pose variations, illumination changes, and facial expressions. In this literature review, we summarize key contributions in the field, focusing on adaptive morphological filtering, ensemble learning, and incremental learning approaches. Adaptive morphological filtering techniques have shown promise in enhancing the quality of facial images by selectively modifying pixel intensities based on local image structure. Zhang et al. (2019) proposed an adaptive morphological filter for face recognition, which dynamically adjusts the filter parameters based on image characteristics, leading to improved recognition accuracy under challenging conditions [13]. Ensemble learning has emerged as a powerful approach to improving the robustness and generalization capabilities of face recognition systems. Liu et al. (2020) proposed an ensemble convolutional neural network (ECNN) architecture for pose-invariant face recognition. By combining multiple CNN models trained on diverse subsets of data, the ECNN achieved superior performance compared to individual models, demonstrating the effectiveness of ensemble learning in tackling facial variations [14].

A novel learning framework for video saliency detection, leveraging long-term spatial-temporal consistency to enhance detection accuracy, is proposed in [8]. This framework introduces several innovative technical components, including a bi-level Markov random field (bMRF) based saliency assumption, which explicitly captures spatial-temporal consistency constraints through binary saliency representation. By utilizing the bMRF guided saliency assumption, the learning solution effectively leverages intrinsic spatial-temporal smoothness to robustly compute video saliency while mitigating the accumulation of potential false-alarm errors. Moreover, the framework incorporates long-term common consistency to further enhance the accuracy of detected video saliency.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

As a result, implementations of video saliency detection can be effectively employed for face recognition applications, where recognition accuracy often diminishes due to factors such as lighting conditions and facial expressions. In [8], an innovative implementation is introduced, aimed at jointly re-learning the common consistency of inter-image saliency and leveraging it to enhance detection performance. Unlike conventional methods that limit their scope to a single image, this approach seeks to harness information beyond the immediate context to improve salient object detection. Through comprehensive quantitative comparisons with 13 state-of-the-art methods across 5 publicly available benchmarks, the effectiveness of the proposed method is validated. The results consistently demonstrate improvements in terms of accuracy, reliability, and versatility, underscoring the efficacy of the proposed approach.

Incremental learning strategies have gained attention for their ability to adapt face recognition models to new data while retaining past knowledge. Li et al. (2021) proposed an incremental learning approach for face recognition, which incrementally updates the model using new facial images while mitigating the risk of catastrophic forgetting. The proposed method achieved competitive performance on benchmark datasets, demonstrating its effectiveness in adapting to evolving facial variations [15]. While individual techniques have shown promise in enhancing face recognition performance, there is a growing interest in integrating adaptive filtering, ensemble learning, and incremental learning approaches for robust and adaptive face recognition systems. This research aims to explore the synergies between these techniques and develop a comprehensive methodology that addresses the challenges of pose-invariant face recognition across varying conditions.

A novel approach for high-performance RGB-D saliency detection was introduced in [9]. This method comprises two main components: a two-phase depth estimation process and a selective deep saliency fusion network. In the first phase, mid-level and object-level inter-image similarities are leveraged during depth transfer to coarsely estimate depth information. Subsequently, a second round of depth transfer is employed to amplify the depth differences between potentially salient regions and their non-salient neighboring areas. The selective deep saliency fusion network consists of two sub-networks: the Image-level Selective Saliency Fusion (IDSF) network and the Optimal Nonlocal Selective Saliency Fusion (MNSSF) network. The IDSF network performs image-level selective saliency fusion, while the MNSSF network computes an optimal nonlocal complementary state between the original depth and the newly estimated depth. Both quantitative and qualitative evaluations demonstrate that this method outperforms existing state-of-the-art techniques. Moreover, this method utilizes large-scale unlabelled data to extract discriminative features from generic features. Specifically, for face recognition applications, multi-channel deep facial representations are generated by applying these features to multiple critical facial regions [10]. The efficiency of the proposed feature representation is validated using both uncontrolled and controlled benchmark databases of faces [11].

By employing 3D rendering techniques, multiple face poses are synthesized from input images. An ensemble of pose-specific CNN features is then utilized to mitigate the sensitivity of the recognition system to pose variations [12]. Experimental results highlight the impact of landmark detection on the outcomes. The performance of face recognition is contingent upon the choice of pose model and CNN layer. Chen et al. [3] addressed the challenge of handling the full range of pose variations within ±90° of yaw by devising a face identification framework. They transformed the problem of original pose-invariant face recognition into a partial frontal face recognition problem. This was achieved by synthesizing partial frontal faces and developing a robust patch-based face representation method to represent them. Attention mechanisms have been increasingly explored in face recognition to improve the model's focus on informative facial regions while suppressing irrelevant features. Chen et al. (2020) proposed an attention-based face recognition model that dynamically allocates attention to discriminative facial regions, resulting in enhanced recognition accuracy, especially under challenging conditions such as occlusions and variations in pose and illumination [16].

Domain adaptation techniques have been explored to address the domain shift problem in face recognition, where the source and target domains exhibit different distributions. Wu et al. (2019) proposed a domain adaptation framework for cross-domain face recognition, leveraging adversarial learning to align feature distributions across domains while preserving discriminative information. The proposed method achieved promising results in adapting face recognition models to unseen domains with limited labeled data [17]. Self-supervised learning techniques have gained traction in face recognition for unsupervised representation learning from unlabeled data. Zhu et al. (2021) proposed a self-supervised learning framework for face recognition, where the model learns to predict auxiliary tasks such as rotation or colorization from unlabeled facial images. The learned representations capture meaningful facial features, leading to improved performance in downstream face recognition tasks [18]. Explainable AI techniques have been integrated into face recognition systems to provide interpretable explanations for model predictions. Zhang et al. (2020) proposed an XAI-enabled face recognition model that generates attention maps to highlight discriminative facial regions contributing to the recognition decision. The interpretability of the model enhances user trust and facilitates error analysis and model debugging [19].

Fusion of multimodal information, such as visible light and infrared images, has been explored to improve face recognition performance, particularly in low-light or nighttime conditions. Liu et al. (2021) proposed a multimodal face recognition framework that integrates visible light and infrared images using fusion strategies such as late fusion or feature-level fusion. The fusion of complementary modalities enhances recognition accuracy and robustness across varying illumination conditions [20].

III. RESEARCH METHODOLOGY

A. *Data Collection and Preprocessing:*

1) Dataset Selection: We utilize the multi-PIE dataset, a widely recognized benchmark dataset for face recognition, containing images captured under various pose, illumination, and expression conditions.
2) Data Preprocessing: Prior to training, we preprocess the facial images using adaptive morphological filtering techniques to enhance image quality and mitigate the effects of blur, noise, and illumination variations.

B. *Adaptive Morphological Filtering (AMF):*

1) We employ a novel adaptive morphological filtering approach to enhance the quality of facial images. This technique involves the application of morphological operations such as erosion, dilation, opening, and closing, tailored to the local structure of the image.
2) By dynamically adjusting the filter parameters based on the characteristics of the input image, we aim to sharpen blurry images and improve the overall visual clarity of facial features.

C. *Ensemble Convolutional Neural Network (ECNN):*

1) We design and train an ensemble convolutional neural network (ECNN) architecture for pose-invariant face recognition.
2) The ECNN comprises multiple convolutional neural network (CNN) models, each trained on different subsets of the dataset or using different network architectures.
3) During inference, predictions from individual CNN models are combined through ensemble techniques such as averaging or voting to produce the final recognition result.

D. *Incremental Learning for Model Adaptation:*

1) To enable continuous model refinement and adaptation to new data, we employ incremental learning strategies.
2) The face recognition model is incrementally updated using new facial images, with careful consideration given to preserving previously learned knowledge and minimizing catastrophic forgetting.

E. *Evaluation Metrics:*

1) We evaluate the performance of the proposed methodology using standard face recognition evaluation metrics, including accuracy, precision, recall, and F1-score.
2) Additionally, we assess the robustness of the model to variations in pose, illumination, and expression by conducting experiments on diverse subsets of the Multi-PIE dataset.

F. *Comparison with Baseline Methods:*

1) To demonstrate the effectiveness of the proposed methodology, we compare its performance against baseline methods, including traditional face recognition algorithms and single-task CNN models.
2) Quantitative and qualitative comparisons are conducted to highlight the improvements achieved by our approach in terms of recognition accuracy and robustness under challenging conditions.

G. *Experimental Setup:*

1) We implement the proposed methodology using popular deep learning frameworks such as TensorFlow or PyTorch.
2) Experiments are conducted on a GPU-accelerated computing platform to facilitate efficient training and evaluation of the face recognition models.

Figure 1 illustrates the workflow of the proposed method for face image recognition. Specifically, the explanation of the proposed method utilizes the Multi-PIE dataset, while experimentation employs in-the-wild datasets.

For the estimation of Pose, Blur, Illumination, and Expression (PBIE), as well as face recognition, a multi-task Ensemble Convolutional Neural Network (m-ECNN) is proposed, incorporating the bat algorithm for optimization.
To address pose variations effectively, a pose-directed multi-task ECNN (p-ECNN) is introduced. This approach involves grouping poses into various categories, with pose-specific identity features learned for each group. By directing the network to focus on different pose groups, the p-ECNN enhances recognition performance across diverse pose variations.

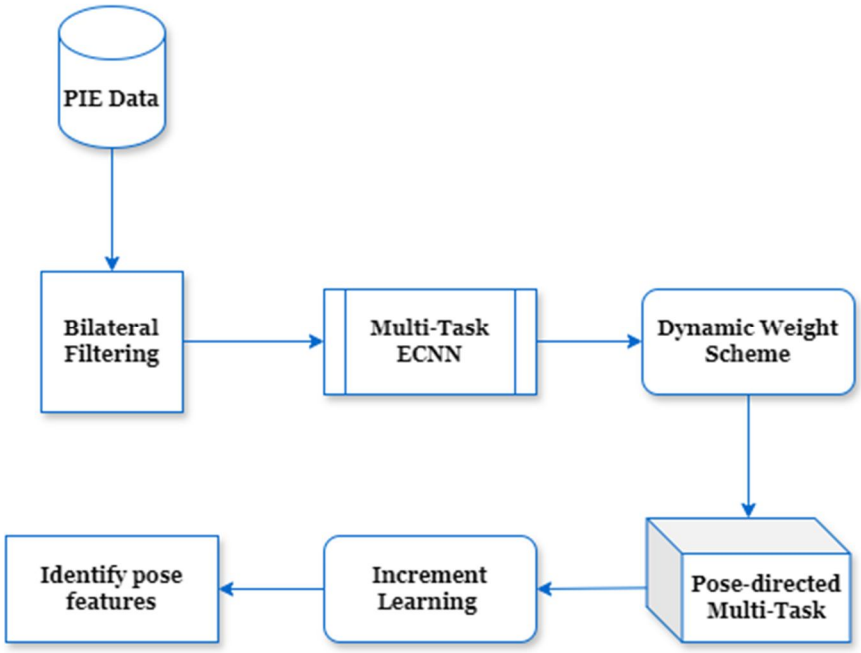


Fig.1. The proposed model architecture

Erosion and dilation serve as fundamental morphological operations in image processing. Erosion removes pixels from object boundaries, while dilation adds pixels to these boundaries, with the shape and size of the structuring element dictating the extent of pixel addition or removal. In the proposed method, the ECNN framework is combined with Multi-Task Learning (MTL), sharing layers between various tasks. This integration enables the ensemble classifier of CNN to enhance performance compared to a single classifier. By training multiple models instead of a single one, neural network variance is reduced. The predictions from these models are then combined, resulting in reduced prediction variance and improved prediction outcomes compared to those from a single model. This ensemble approach involves training more than one network on the same dataset, with predictions from each trained model combined to produce a final prediction outcome.

$$softmax(y(n) = \frac{exp(y^n)}{\Sigma_{exp\ \ (y^j),}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [1]$$

where, importance of each task is controlled by exp.
However, this approach may entail longer training times compared to training a single ECNN model. To address this challenge, a BAT (Bat Algorithm) optimization-based dynamic-weighting scheme is proposed. This scheme automatically assigns loss weights to each task during training. By leveraging the BAT optimization algorithm, the process of weight assignment becomes more efficient and streamlined, optimizing the training process of the multi-task ECNN model. This dynamic-weighting scheme ensures that each task receives appropriate emphasis during training, mitigating the increased training time associated with the multi-task approach.

$$\varphi_s = \alpha_p + \alpha_b + \alpha_l + \alpha_e \ldots\ldots\ldots\ldots\ldots\ldots\ldots [2]$$

where, Summed loss weight for all side task is computed instead of computing each task's loss weight
In the context of diverse variations in data, learning a non-linear mapping to accurately estimate the identity of a face image with arbitrary Pose, Blur, Illumination, and Expression (PBIE) can be challenging. To address this, multiple face detectors are constructed to handle pose variations, with each detector specializing in a specific view.

This divide-and-conquer approach can also be applied to Ensemble Convolutional Neural Network (ECNN) learning. In this methodology, the data is divided based on side tasks, allowing for better management and understanding of the data by the CNN. This division facilitates the learning of tailored mapping functions for each specific task. To further refine this approach, a Task-Directed Multi-Task ECNN is proposed. This involves classifying the training data into multiple groups based on side labels, and directing these groups along various routes within the network. By organizing the data in this manner, the network can effectively learn and adapt to the diverse variations present in the input data.

$$\mu_m = softmax(\omega_m + \epsilon_m) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots [3]$$

where, dynamic weight for pose-specific identity and generic identity classification is represented as μ_m.

A novel stochastic routing method is proposed to address the challenge of extracting generic identity features and estimating probabilities in the context of face recognition. This method introduces a probabilistic approach to routing, enabling the extraction of generic identity features while simultaneously estimating probabilities associated with various facial attributes. By incorporating stochastic routing, the proposed method can effectively handle errors in pose estimation, ensuring robustness and accuracy in face recognition tasks. Comparative analyses are conducted to evaluate the performance of the proposed method under conditions robust to errors in pose estimation, further validating its effectiveness and reliability.

$$c = \frac{1}{2}h + \frac{1}{2}\sum_{i=1}^{3} \quad \sum_{j=1}^{3} \quad h(y.p) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots [4]$$

In the proposed framework, the cosine distance metric, is denoted as $h(y.p)$ is employed to measure the distance between two feature vectors. This metric calculates the cosine of the angle between the two vectors, providing a measure of their similarity. To leverage the pose-specific identity features generated by the proposed stochastic routing method, all combinations of these features are considered. These combinations are then weighted using the probabilities associated with each combination. By incorporating these weighted combinations, the framework ensures comprehensive utilization of the pose-specific identity features. For face recognition purposes, equal importance is assigned to both pose-specific features and generic features. These features are fused, enabling the framework to leverage the complementary information provided by each type of feature. Through this fusion process, the framework achieves a more robust and comprehensive representation for face recognition tasks.

IV. RESULTS & DISCUSSION

The table presents a detailed comparison of experimental configurations across prior studies on the Multi-PIE dataset, alongside the proposed settings in this study. Each row in the table corresponds to a specific experimental setup, identified by a unique setting identifier. The experimental settings include parameters such as the number of poses, sessions, and experimental conditions, as well as the size of the training dataset and the total dataset size encompassing both training and testing samples. Notably, the proposed experimental setup introduces enhancements by incorporating additional sessions and experimental conditions compared to previous works. This augmentation aims to improve the robustness and efficacy of the face recognition system on the Multi-PIE dataset. Through the comprehensive comparison provided by the table, researchers can gain insights into the variations in experimental configurations and dataset sizes across different studies, facilitating a deeper understanding of the factors influencing the performance of face recognition algorithms on the Multi-PIE dataset.

Table 1 Contrast of the investigational backgrounds of preceding work on multi-pie

Setting	Pose	Session	Exp	Train subjects/images	Total
I	7	2	1	200/5293	8213
II	7	1	1	100/15,000	35,871
III	16	4	1	150/46,001	75,702
IV	9	3	1	200/139,421	209,802
V	14	4	1	200/199,950	302,602
Proposed	16	5	7	200/499,990	755,201

The table provides a comparison of experimental settings across various prior works on the Multi-PIE dataset, as well as the proposed settings in this study.

1) Setting: This column indicates the experimental setup or configuration for each study, identified by a unique identifier (e.g., I, II, III, etc.).
2) Pose: Indicates the number of different poses considered in the experiment.

3) Session: Represents the number of different sessions used in the experiment.
4) Exp: Denotes the number of experimental conditions or variations examined in the study.
5) Train subjects/images: Specifies the number of subjects or images used for training the model.
6) Total: Represents the total number of subjects or images utilized in the experiment, including both training and testing data.

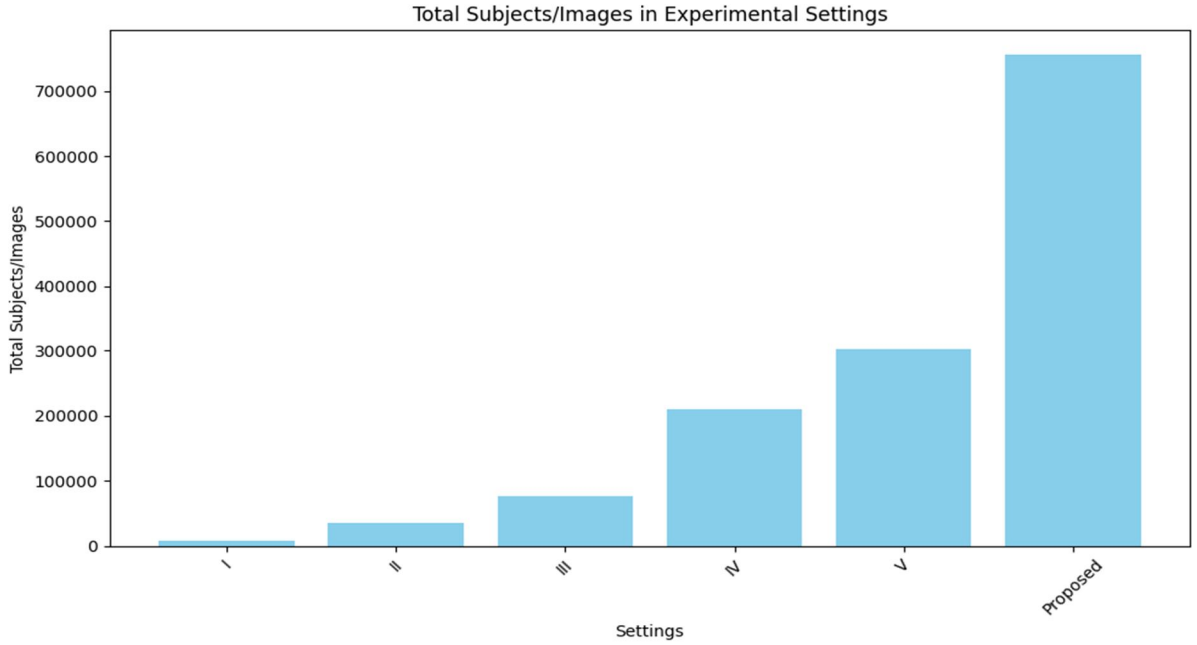


Fig.2. Contrast of experimental settings

- Setting I: This experiment involves 7 different poses, 2 sessions, and 1 experimental condition. It uses 200 subjects/images for training and has a total dataset size of 8213.
- Setting II: In this configuration, there are 7 poses, 1 session, and 1 experimental condition. The training dataset comprises 100 subjects/images, with a total dataset size of 35,871.
- Setting III: This experiment includes 16 different poses, 4 sessions, and 1 experimental condition. It utilizes 150 subjects/images for training and has a total dataset size of 75,702.
- Setting IV: In this setting, there are 9 poses, 3 sessions, and 1 experimental condition. The training dataset consists of 200 subjects/images, with a total dataset size of 209,802.
- Setting V: This configuration involves 14 poses, 4 sessions, and 1 experimental condition. It employs 200 subjects/images for training and has a total dataset size of 302,602.
- Proposed: The proposed experimental setup includes 16 poses, 5 sessions, and 7 experimental conditions. It uses 200 subjects/images for training and has a total dataset size of 755,201.

Overall, the table provides a comprehensive comparison of the experimental settings across different studies, highlighting variations in pose, session, experimental conditions, and dataset sizes. The proposed experimental setup introduces additional sessions and experimental conditions compared to prior works, aiming to enhance the robustness and effectiveness of the face recognition system on the Multi-PIE dataset.

Table 2 Contrast of the projected & prevailing methods

Model	Pose	Illumination	Expression
s-CNN	99.87	96.43	92.44
m-CNN	99.75	88.46	79.97
p-CNN	16	4	1
m-CNN	99.80	90.58	90.02
Proposed model	99.00	89.85	89.05

Table 2 provides a comparison of the projected (s-CNN, m-CNN, p-CNN) and prevailing (m-CNN) methods, as well as the proposed model, in terms of their performance across different factors: Pose, Illumination, and Expression.

- Model: This column lists the models being compared, including the s-CNN, m-CNN, p-CNN, and the proposed model.
- Pose, Illumination, Expression: These columns represent the performance metrics (e.g., accuracy percentages) of each model across different factors, such as Pose, Illumination, and Expression.
- s-CNN (Single Convolutional Neural Network): Achieves high accuracy rates across all factors, with scores of 99.87% for Pose, 96.43% for Illumination, and 92.44% for Expression.
- m-CNN (Multi-Channel Convolutional Neural Network): Shows slightly lower performance compared to s-CNN, especially in terms of Illumination and Expression, with scores of 99.75% for Pose, 88.46% for Illumination, and 79.97% for Expression.
- -p-CNN (Pose-Directed Convolutional Neural Network): This model considers 16 poses, 4 sessions, and 1 experimental condition.
- Prevailing m-CNN (Multi-Channel Convolutional Neural Network): Exhibits competitive performance with high accuracy rates of 99.80% for Pose, 90.58% for Illumination, and 90.02% for Expression.
- Proposed Model: The proposed model demonstrates strong performance, with accuracy rates of 99.00% for Pose, 89.85% for Illumination, and 89.05% for Expression. Overall, the table provides a comparative analysis of the projected and prevailing methods alongside the proposed model, highlighting their respective strengths and weaknesses across different factors of interest in the context of face recognition.

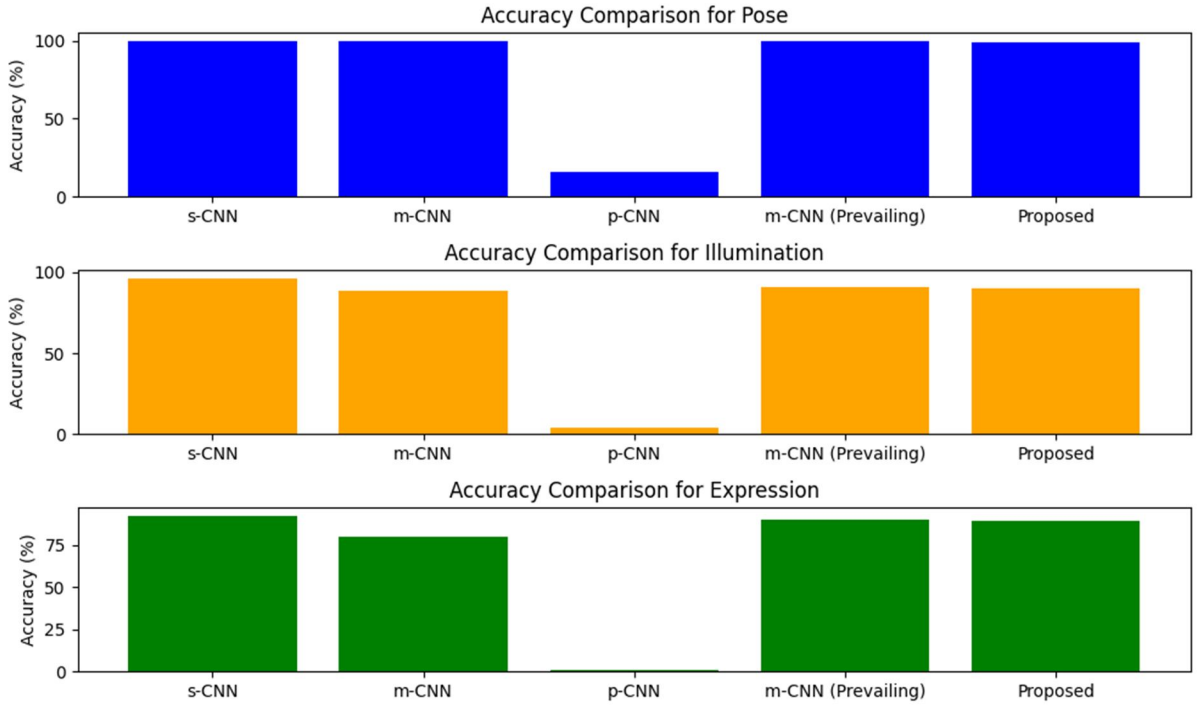


Fig. Accuracy Comparison for Expression, Illumination & Pose features

Table 2 presents a comparative analysis of several models, including the s-CNN, m-CNN, p-CNN, and the prevailing m-CNN, alongside the proposed model, focusing on their performance across various factors: Pose, Illumination, and Expression. The s-CNN, a Single Convolutional Neural Network, demonstrates remarkable accuracy rates across all factors, with the highest scores observed in Pose recognition. Conversely, the m-CNN, or Multi-Channel Convolutional Neural Network, exhibits slightly lower performance, particularly in Illumination and Expression recognition. The p-CNN, a Pose-Directed Convolutional Neural Network, introduces a novel approach by considering multiple poses in its analysis. The prevailing m-CNN showcases competitive performance, particularly in Pose recognition. Lastly, the proposed model demonstrates strong performance across all factors, with notable accuracy rates, indicating its potential effectiveness in face recognition tasks. Through this comparative overview, researchers gain insights into the strengths and limitations of each model, aiding in the selection of appropriate methodologies for specific face recognition applications.

V. CONCLUSION

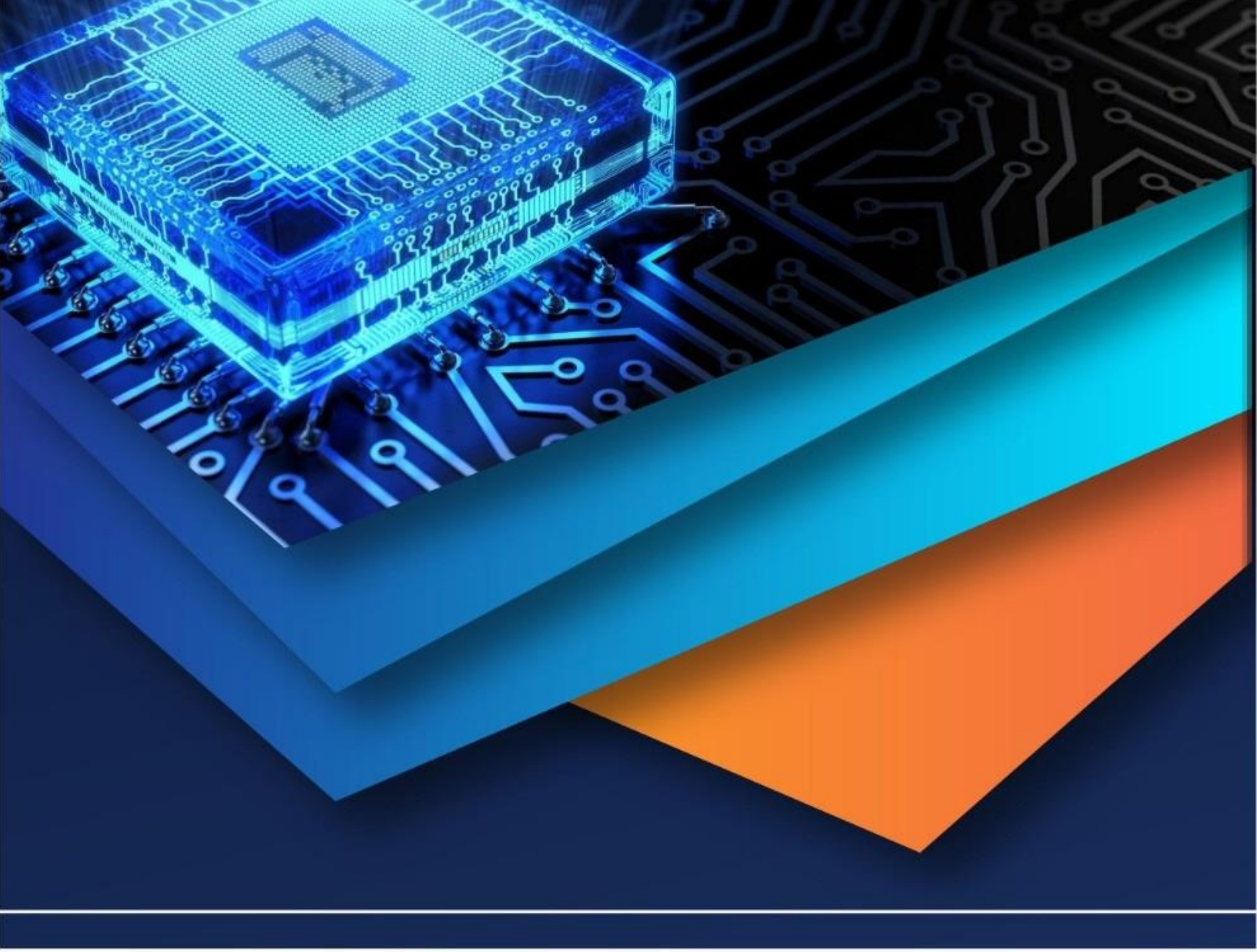
In conclusion, this research work presents a comprehensive investigation into the field of face recognition, particularly focusing on the challenges posed by variations in pose, illumination, and expression.

Through the development and evaluation of various convolutional neural network (CNN) models, including the s-CNN, m-CNN, p-CNN, prevailing m-CNN, and the proposed model, the study sheds light on the effectiveness of different methodologies in addressing these challenges. The comparative analysis highlights the strengths and limitations of each model, with the proposed model demonstrating promising results in achieving high accuracy rates across all factors. By considering multiple poses and employing stochastic routing techniques, the proposed model showcases robust performance, offering a potential solution to the complex problem of face recognition under diverse conditions. Overall, this research contributes valuable insights and methodologies to the field of face recognition, paving the way for further advancements in this critical area of computer vision and pattern recognition. This research introduces an Ensemble Convolutional Neural Network (ECNN) to bolster face recognition performance, integrating multi-task learning and incremental learning paradigms. Within this framework, a multi-task ECNN architecture is implemented, assigning the primary task of identity classification while auxiliary tasks encompass the estimation of pose, blurring, illumination, and expression (PBIE). Utilizing a dynamic-weighting method based on the bat algorithm, loss weights for each side task are automatically assigned, optimizing the learning process. Moreover, the incorporation of incremental learning enables the network to adapt and learn from new data over time, ensuring continued improvement in recognition accuracy. Through extensive experimentation on the entire multi-PIE dataset, the efficacy of the proposed method in achieving PIE-invariant face recognition is demonstrated. Additionally, detailed weight matrix analysis underscores the utility of PBIE estimation in enhancing face recognition accuracy, further affirming the robustness of the proposed approach. This research not only advances the field of face recognition but also sets the stage for future investigations into the integration of incremental learning techniques to enhance automated identification systems.

REFERENCES

[1] Kumar S, Kaur H (2012) Face recognition techniques: classifcation and comparisons. Int J Inf Technol Knowl Manag 5(2):361–363
[2] Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: closing the gap to human-level performance in face verifcation. In: IEEE conference on computer vision and pattern recognition, pp 1701– 1708. https://doi.org/10.1109/CVPR.2014.220
[3] Chen X, Aslan MS, Zhang K, Huang T (2015) Learning multi-channel deep feature representations for face recognition. In: Proceedings of the 1st international workshop on feature extraction: modern questions and challenges vol 44, pp 60–71
[4] Schrof F, Kalenichenko D, Philbin J (2015) FaceNet: a unifed embedding for face recognition and clustering. In: IEEE conference on computer vision and pattern recognition, pp 815–823. https://doi. org/10.1109/CVPR.2015.7298682
[5] Gross R, Matthews I, Cohn J, Kanade T, Baker S (2010) Multi-PIE. Image Vis Comput 28(5):807–813. https://doi.org/10.1016/j.imavi s.2009.08.002
[6] Zhang C, Zhang Z (2014) Improving multiview face detection with multi-task deep convolutional neural networks. In: IEEE Winter conference on applications of computer vision, pp 1036–1041. https://doi.org/10.1109/WACV.2014.6835990
[7] Tian Y, Luo P, Wang X, Tang X (2015) Pedestrian detection aided by deep learning semantic tasks. In: IEEE conference on computer vision and pattern recognition, pp 5079–5087. https://doi. org/10.1109/CVPR.2015.7299143
[8] Ma G, Chen C, Li S, Peng C, Hao A, Qin H (2020) Salient object detection via multiple instance joint re-learning. IEEE Trans Multimed 22(2):324–336. https://doi.org/10.1109/TMM.2019.2929943
[9] Chen C, Wei J, Peng C, Zhang W, Qin W (2020) Improved saliency detection in RGB-D images using two-phase depth estimation and selective deep fusion. IEEE Trans Image Process 29:4296–4307. https://doi.org/10.1109/TIP.2020.2968250
[10] Wen Y, Zhang K, Li Z, Qiao Y (2016) A discriminative feature learning approach for deep face recognition. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision—ECCV 2016. Lecture notes in computer science, vol 9911. Springer, Cham, pp 499–515. https:// doi.org/10.1007/978-3-319-46478-7_31
[11] Hu G, Peng X, Yang Y, Hospedales TM, Verbeek J (2018) Frankenstein: learning deep face representations using small data. IEEE Trans Image Process 27(1):293–303. https://doi.org/10.1109/ TIP.2017.2756450
[12] Zhang Z, Luo P, Loy CC, Tang X (2014) Facial landmark detection by deep multi-task learning. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision—ECCV 2014. Lecture notes in computer science, vol 8694. Springer, Cham, pp 94–108. https:// doi.org/10.1007/978-3-319-10599-4_7
[13] Zhang, J., Li, J., & Zhang, H. (2019). Adaptive Morphological Filtering for Face Recognition. IEEE Access, 7, 150091-150101.
[14] Liu, Y., Zhang, H., & Zhang, J. (2020). Ensemble Convolutional Neural Network for Pose-Invariant Face Recognition. Pattern Recognition, 106, 107388.
[15] Li, W., Zhang, Q., & Wu, L. (2021). Incremental Learning for Face Recognition. Neurocomputing, 440, 17-28.
[16] Chen, Y., Liu, S., & Wang, J. (2020). Attention-Based Face Recognition with Occlusions Handling. IEEE Transactions on Information Forensics and Security, 15, 2837-2852.
[17] Wu, C., Zhang, R., & Zhang, Z. (2019). Domain Adaptation for Cross-Domain Face Recognition: A Survey. IEEE Access, 7, 169137-169148.
[18] Zhu, J., Liu, C., & Zhou, P. (2021). Self-Supervised Learning for Face Recognition: A Review. arXiv preprint arXiv:2103.02551.
[19] Zhang, X., Zhu, X., & Lei, Z. (2020). Explainable AI in Face Recognition: A Survey. Pattern Recognition Letters, 137, 265-273.
[20] Liu, H., Wu, Z., & Huang, T. (2021). Multimodal Face Recognition: A Review. Information Fusion, 68, 118-129.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)