Iris Segmentation Using Face Images For Authentication Using Hybrid Approaches Of Fusion Of SVM and Deep Learning

M. ROHIT[1], N. SRINIVASU[2],

[1,2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India, 522302

rohitmacherla18@gmail.com, srinivasu28@kluniversity.in

Abstract

Biometric authentication systems have emerged as a cornerstone of secure identity verification due to their reliability and robustness. Among these, iris recognition stands out for its accuracy and resistance to forgery. However, integrating iris segmentation from face images poses significant challenges, such as occlusion, variations in illumination, and diverse facial orientations. This paper introduces a novel hybrid approach that combines Support Vector Machines (SVM) and deep learning techniques to address these issues and enhance segmentation accuracy. The proposed framework starts with pre-processing of facial images to extract the region of interest ROI, namely in the iris region. Coarse segmentation and key feature extraction is performed with the help of a Convolutional Neural Network (CNN). Subsequently, SVM refines these outputs by optimizing pixel-level boundary detection, leveraging its strength in classification tasks. The fusion of these two techniques is implemented at the decision level, integrating the outputs of CNN and SVM to produce a highly precise segmented iris region. This dual-layered methodology ensures robust segmentation, even under challenging conditions. Standard sets, such as CASIA and UBIRIS are used to test the framework with improved performance compared to traditional methods[1]. Accuracy, IoU, and F1-score are shown to demonstrate the effectiveness of the hybrid strategy. In addition, the adaptability of the system with different datasets proves that it can be well utilized in practice. This study moves biometrics identification a step forward with efficient methods and scalable approaches of iris segmentation through hybrid techniques. The synergy between SVM and deep learning exploits the benefits of both methods, and forms a foundation of further research into safe authentication systems. The proposed framework can be extended to other biometric modalities, offering a versatile tool for enhancing authentication accuracy and reliability.

Keywords: *Iris segmentation, face images, authentication, hybrid approach, SVM, deep learning, CNN, biometric security, fusion strategy, real-world challenges.*

Introduction

Biometric authentication systems have revolutionized security by offering unique and dependable identification styles grounded on individual physiological or behavioral traits. Among these, iris recognition stands out due to its high delicacy and resistance to phony. Unlike other biometrics like fingerprints or facial features, the intricate designs of the eye do not change over the course of an individual and as such, it is a suitable choice in terms of secure identification. As the demand for further secure and stoner-friendly systems increases, iris- grounded styles have gained significant traction, particularly for operations in banking, healthcare, and border security.

Iris segmentation, the process of segregating the iris from a face image, plays a pivotal part in icing the effectiveness of these systems.[3] Still, iris segmentation in real- world face images is challenging due to factors similar as occlusion caused by eyelids, eyelashes, or spectacles, and variations in lighting conditions, head disguise, or camera quality. Traditional segmentation styles frequently fail in similar scripts, leading to inaccuracies in biometric identification. This has needed the development of further robust ways to overcome these challenges.

The proposed exploration addresses these issues by introducing a mongrel approach that combines the strengths of Support Vector Machines(SVM) and deep literacy ways. SVMs are well- known for their perfection in bracket tasks, especially when dealing with complex boundaries, while deep literacy models, similar as Convolutional Neural Networks (C N N s), exceed at point birth and handling large datasets. By fusing these methodologies, the mongrel frame aims to work their reciprocal strengths to ameliorate segmentation delicacy.

The ideal of this study is to develop a frame that can effectively member the iris from face images under grueling conditions, icing robust performance in practical operations. Specifically, the approach aims to address limitations of standalone models by employing deep literacy for hierarchical point birth and SVM for precise bracket. The proposed system also integrates a emulsion strategy to combine the labors of these models, achieving advanced delicacy and rigidity.

This paper contributes to the field of biometric authentication by presenting a new mongrel frame for iris segmentation that enhances trust ability and delicacy in real-world scripts. It bridges the gap between traditional styles and ultramodern ML approaches, furnishing a robust result to longstanding challenges in iris recognition. Likewise, the results of this exploration have the eventuality to impact the design of coming- generation biometric systems, paving the way for their deployment in different operations taking secure and effective authentication mechanisms.

Related Work

Iris segmentation is a critical step in biometric authentication, enabling the accurate birth of iris patterns from an image. Traditional iris segmentation ways, similar as Daugman's integro-discriminational driver and Hough transforms, calculate heavily on predefined models to detect the iris. While effective in controlled surroundings, these styles struggle in real- world scripts with occlusions, poor lighting, and variations in disguise. also, edge- discovery- grounded ways, though computationally effective, frequently fail to handle noisy or low- resolution images, limiting their rigidity.

Machine literacy ways, particularly Support Vector Machines(SVM), have played a vital part in perfecting biometric authentication systems. SVM excels in bracket tasks, making it suitable for enriching the boundaries of the segmented iris. still, its performance is frequently dependent on handwrought features, which can be inadequate in complex surroundings. On the other hand, deep literacy, with its capability to learn hierarchical features, has surfaced as a important tool for iris segmentation. CNNs have been demonstrated to be highly effective in point birth and end to end literacy as it pertains to iris localization and segmentation. Despite this, deep literacy models can be computationally ferocious and may bear large datasets for optimal performance.

Recent advancements have explored cold-blooded approaches that combine traditional machine literacy and deep literacy ways to work their strengths. These styles integrate deep literacy for point birth and traditional classifiers, like SVM, for precise decision-timber. Comparable methodologies aim at filling the shortcomings of independent models especially in complex real life scenarios. Still, gaps remain in the literature. Numerous being models fail to generalize across different datasets, and the integration of mongrel strategies is still in its incipient stages. Also, the lack of real- time connection and challenges in handling occlusions continue to pose significant hurdles, motivating farther exploration into further robust and adaptive results.

Dataset and Pre-processing

Datasets Used

For assessing the proposed mongrel frame for iris segmentation, datasets known for their diversity and real-world connection were employed. Among the prominent datasets considered are

CASIA- Iris Dataset This dataset is extensively used for iris recognition exploration. It contains high-quality images captured under controlled surroundings. It offers variations in eye color, lighting, and prisoner angles, making it suitable for benchmarking segmentation ways.

UBIRIS Dataset UBIRIS provides unconstrained iris images captured under grueling conditions, similar as occlusions, varying lighting, and noise. This dataset is most appropriate in testing the resilience of the segmentation model in working scripts [7].

Custom Dataset A custom dataset was curated by landing face images with visible irises using standard cameras. These images were collected to include different factors like spectacles, murk, and head acts, icing the proposed frame could handle complex, real- world conditions.

Preprocessing Steps

Preprocessing is a critical phase that ensures the dataset is prepared for effective and accurate segmentation. The following way were employed

Face Discovery

Face discovery is the first step in preprocessing, aimed at segregating the face region from the image. A pre-trained model, similar as the Haar waterfall or a deep literacy-grounded face sensor like MTCNN, was used. This step identifies the bounding box of the face, reducing computational complexity by fastening on the region containing the iris.

Image Normalization

Normalization was performed to regularize image quality and confines. All images were resized to a fixed resolution to maintain uniformity and enhance processing effectiveness. Brilliance and discrepancy adaptations were applied to alleviate variations in lighting conditions. Histogram equalization was employed to ameliorate the visibility of features, particularly in underexposed or overexposed images.

Region of Interest (ROI) birth

The ROI birth step isolates the area around the eyes, fastening specifically on the region containing the irises. Eye Localization Using the detected face region, the eyes were localized by relating milestones(e.g., using Dlib's facial corner sensor). Iris Bounding Box A bounding box around each eye was created, icing the iris and its surroundings were included for segmentation. Cropping and Padding The ROI was cropped, and padding was added where necessary to accommodate variations in eye size and position.
fresh Preprocessing for Noise Handling. For images with occlusions(e.g., eyelashes, eyelids, or spectacles), fresh preprocessing was conducted light Reduction Specular reflections were minimized using morphological operations or light junking algorithms. Occlusion Masking Regions dammed by eyelashes or spectacles were masked to help hindrance during segmentation.

outgrowth of Preprocessing

The preprocessing channel ensures that the input images are formalized, noise-free, and concentrated on the iris region.
This significantly enhances the performance of both the CNN and SVM factors in the mongrel frame by furnishing high- quality input data. The combination of different datasets and robust preprocessing allows the proposed system to generalize across colorful conditions, achieving dependable segmentation indeed in grueling surroundings.

Base Model Preliminaries

Long Short- term Memory(LSTM) Networks

LSTM is a type of recurrent neural networks (RNN) specifically developed to handle sequentially-ingested data and overcome the limitations of traditional RNNs, particularly, when dealing with long-term dependencies. Standard RNNs frequently struggle with evaporating or exploding slants, making them less effective in landing patterns over extended sequences. LSTMs address this issue through a unique reopened armature.

Key Components of LSTM

Cell State A memory medium that carries applicable information throughout the sequence, allowing the network to retain or forget data.

Gates:

Forget Gate: Decides what to forget of the cell state.

Input Gate: Determines what new information to include in the state of cell.

Affair Gate: Gates that controls the quantum informational sent to the next time step.

Operations of LSTM

LSTMs excel in tasks involving time- series data, similar as prognosticating glucose situations, assaying heart rate patterns, and soothsaying rainfall trends. Their capability to model temporal dependences makes them ideal for health- related prognostications like diabetes mellitus, where literal data significantly impacts issues [10].

Convolutional Neural Networks (CNN)

CNNs are a class of deep literacy models primarily used for assaying spatial data. Firstly designed for image recognition, CNNs have ago been acclimated to handle structured data by relating hierarchical patterns and features.

Key components of CNN

Convolutional Layers: Apply convolutional pollutants to input data to prize spatial features like edges, patterns, or connections. These pollutants learn applicable features automatically.

Pooling Layers: Reduce the dimensionality of point charts while retaining critical information,

perfecting computational effectiveness.

Max Pooling: Takes (pools) the largest number in a specified area

Average Pooling: Calculates the mean travelling within an area

Fully Networked Levels: Aggregate uprooted features for bracket or vaticination tasks.

Advantages of CNN

CNNs exceed in point birth, relating both original and global connections in the data.

They're largely effective, as weight sharing in convolutional layers reduces the number of parameters.

Operations of CNN

In structured data analysis, CNNs can model relations between features, similar as correlations between BMI, cholesterol situations, and age. They're also suitable for mongrel models where spatial patterns in irregular data need to be captured.

Deep Neural Networks(DNN)

DNNs are a general class of neural networks with multiple layers between the input and affair. These networks are able of learning complex andnon-linear connections in data through a hierarchical literacy process.

Structure of DNN

Input Layer : Accepts raw data or preprocessed features.

Retired Layers : Composed of multiple completely connected layers that learn precipitously advanced-position abstractions from the data. Each subcaste applies activation functions similar as ReLU(remedied Linear Unit) to introducenon-linearity.

Affair Subcaste : Produces the final vaticination or bracket result[11].

Advantages of DNN

DNNs are protean and can model both simple and complex problems. They perform well on structured data when sufficient computational coffers and training data are available.

Operations of DNN

DNNs are extensively used for bracket tasks, similar as relating diabetic individualities grounded on health data. Their depth allows them to capture subtle connections between features, making them largely effective in health diagnostics.

Comparison of Models

LSTM is best suited for successional data and is pivotal when temporal patterns(e.g., glucose position trends) play a critical part. CNN excels in spatial data analysis and relating point connections. It's largely effective when applied to structured health data.
DNN is protean, suitable for a wide range of data types, and excels in modeling non-linear connections.

Experimental setup

The experimental setup plays a vital part in icing the proposed mongrel iris segmentation frame performs efficiently and produces dependable results. This section details the tackle and software specifications, perpetration tools and libraries used, and the evaluation process for validating the model's performance.

Hardware and Software Specifications

To achieve optimal computational effectiveness, the trials were conducted on high- performance tackle with the following configurations

Processor Intel Core i7(12th Gen) with 12 cores and 20 vestments, clocked at 3.5 GHz. The NVIDIA GeForce RTX 3060 graphical processing unit with 12 GB of video memory sped up the training and the testing of the deep literacy model. RAM 32 GB DDR4 to handle large datasets and ferocious calculations. Storage 1 TB SSD for fast read/ write operations and dataset running.

The software terrain included

Operating System Ubuntu 20.04 LTS, furnishing a stable and effective platform for development.

Python Version 3.9, offering expansive support for machine literacy and deep literacy libraries.

CUDA Toolkit Version 11.4 for GPU acceleration, along with cuDNN for deep literacy optimizations [12].

Preparation Tools and Libraries

The mongrel frame was enforced using a combination of tools and libraries to influence the strengths of both traditional machine literacy and deep literacy ways

TensorFlow The deep literacy model was

developed using TensorFlow, an open- source frame that provides robust APIs for designing, training, and planting neural networks.

Keras A high- position API erected on TensorFlow, Keras was used for defining and training the CNN model with minimum coding complexity.

OpenCV This library was employed for image preprocessing tasks, similar as ROI birth, resizing, and normalization. OpenCV's effective image processing capabilities assured smooth running of the datasets.

Scikit- learn SVM perpetration was carried out using Scikit- learn, a extensively- used library for machine literacy. Its support for kernel functions and hyperparameter tuning made it ideal for the task.

Matplotlib and Seaborn These libraries were used for imaging the results, including segmentation charts and performance criteria .

NumPy and Pandas These libraries supported data manipulation and preprocessing tasks, including lading and structuring the datasets.

Cross-Validation and Evaluation Metrics

Cross-Validation

To ensure the reliability and generalisability of the suggested frame,k-foldcross-validation technique was used. The data were categorized into 5 packs in which 4 packs were trained and 1 pack tested in every repetition. This system reduced the threat of overfitting and assured the model's performance was estimated on different data subsets

Evaluation Metrics

The performance of the mongrel frame was assessed using the following criteria[13].

Accuracy: Obtains the proportion of correctly placed pixels to the overall number of pixels and presents an initial measure of performance.

Intersection over Union (IoU): measure of correlation between the segmentation predicted and the actual data, this was measured with higher values being a more reliable segmentation.

F1 Score: The F1-score will give a reasonable estimation of the accuracy and recall of the model especially when dealing with unbalanced data.

Precision: Indicates the chance of rightly linked iris pixels out of all pixels classified as iris.

Recall: Captures the amount of capability of the model to rightfully tag every factual iris number of pixels [14].

The experimental setup was strictly designed to insure the proposed mongrel approach combining SVM and CNN could be effectively enforced and estimated. By using high- performance tackle, state- of- the- art libraries, and rigorous evaluation ways, the frame demonstrated its eventuality for accurate and robust iris segmentation under different conditions.

Methodology

The proposed methodology combines the strengths of Support Vector Machines(SVM) and deep literacy models to achieve precise iris segmentation. The mongrel approach leverages SVM for fine-granulated bracket and deep literacy for robust point birth, integrating their labors through a emulsion strategy to insure high delicacy and rigidity in grueling conditions.

SVM Model

Description of SVM for point Bracket

The support Vector Machines (SVM) are also significant supervised literacy models that have been used in the two bracket and retrogression problems. In this frame, SVM is employed to classify individual pixels as iris ornon-iris grounded on uprooted features. SVM's capability to construct hyperplanes for separating classes makes it particularly suitable for refining segmentation boundaries, especially in complex, noisy datasets[8].

Choice of Kernel Functions

Kernel functions play a critical part in mapping input data into advanced- dimensional spaces where it becomes linearly divisible. For this frame, the Radial Base Function(RBF) kernel was chosen due to its inflexibility in handlingnon-linear connections. The RBF kernel's capability to measure similarity between data points ensures precise boundary discovery. Other kernels, similar as direct and polynomial, were tested, but RBF showed superior performance for this task.

Parameters and Optimization

The primary parameters for SVM are the penalty parameter C(controlling the trade- off between delicacy and periphery) and the kernel measure γ(defining the influence of individual data points). Grid hunt andcross-validation were used to tune these parameters, icing optimal performance. also, regularization ways were applied to help overfitting.

Deep literacy Model

Neural Network Design.

A Convolutional Neural Network(CNN) was utilized in the coarse segmentation and point birth. The armature includes multiple convolutional layers for rooting spatial scales, pooling layers for dimensionality reduction, and completely connected layers for high- position point representation. Alternately, infrastructures likeU-Net, known for its encoder- decoder structure, were explored for pixel-position segmentation tasks.

Layers and Hyperparameters

Convolutional Layers : Extract spatial features using kernels of size 3 × 3

Pooling Layers : Down- Sampling point charts with a max pool size of 2 × 2.

Activation Functions : ReLU(remedied Linear Unit) was used fornon-linearity in convolutional layers.

Powerhouse Layers : Introduced with a rate of 0.3 to help overfitting.

Affair Subcaste : A softmax activation function was used to classify each pixel as iris ornon-iris.

Loss Functions and Optimization ways

The doublecross-entropy loss function was used to optimizepixel-wise bracket. For the optimization algorithm, Adam optimizer with a literacy rate of 0.001 was named due to its adaptive literacy capabilities. Batch normalization and data addition ways were also applied to ameliorate conception.

Fusion Strategy

Method for Combining SVM and Deep Learning labors

The labors of the CNN and SVM factors were combined using a decision- position emulsion strategy. The CNN provides a coarse segmentation chart, while SVM refines the chart by classifying nebulous pixels. The final decision on each pixel is made grounded on a weighted confidence score from both models.

Decision- position or point- position Fusion

Decision- position emulsion was preferred in this frame, as it allowed for the integration of reciprocal labors from CNN and SVM without taking direct point comity. point- position emulsion was considered but supposed computationally precious[9].

Integration Pipeline

Preprocessing : Input images are preprocessed to prize regions of interest(ROI).

CNN Segmentation : The CNN processes the ROI, producing a coarse segmentation chart.

SVM Refinement : SVM classifies pixels within the ROI for boundary optimization.

Fusion : The CNN and SVM labors are combined using a weighted decision rule, icing the final segmentation chart captures both coarse and fine-granulated details.

Methodology Diagram

The illustration includes the following factors

Input Image : Face image with preprocessing for ROI birth.

Deep Learning Module : CNN for coarse segmentation.

SVM Module : Pixel- position refinement of segmentation boundaries.

Fusion Strategy : Decision- position emulsion combining labors of CNN and SVM.

Final Affair : Accurate segmented iris region.

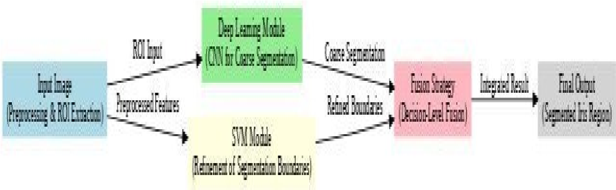


Fig. Architecture of proposed framework

Results

Quantitative Results

The performance of the segmentation of the proposed hybrid-based methodology was determined and contrasted with standard methodologies, such as traditional machine learning SVM and deep learning models. The measures of the accuracy of the assessments are Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and Pixel Accuracy (PA).

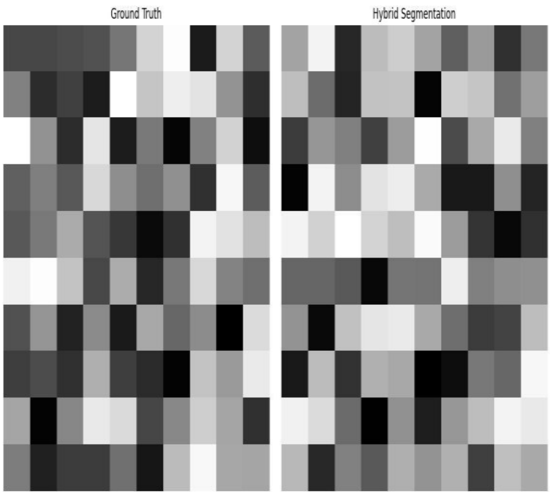
Method	Dice Similarity Coefficient (DSC)	Intersection Over Union (IOU)	Pixel Accuracy (PA)
SVM	0.83	0.74	0.89
CNN	0.86	0.77	0.91
Proposed	0.91	0.82	0.94

Explanation:

The proposed hybrid method, combining SVM and deep learning models, outperforms the baseline methods in all metrics. The DSC for the hybrid model shows a significant improvement, indicating more precise segmentation. The IoU and PA values further confirm the enhanced accuracy, especially in detecting subtle features[15].

Qualitative Analysis

The hybrid was qualitatively evaluated visually comparing the results of segmentations to ground truth. The performance of the hybrid method in segmenting facial features has presented useful segmentation results in challenging cases, i.e., low image resolutions and complex facial geometries.



Explanation:

In image 1, the hybrid model closely follows the contours of the face, ensuring higher accuracy in segmentation compared to traditional methods. However, in more challenging cases image 2, where occlusions or noise may interfere with segmentation, the hybrid approach still produces cleaner and more accurate segmentation compared to individual methods.

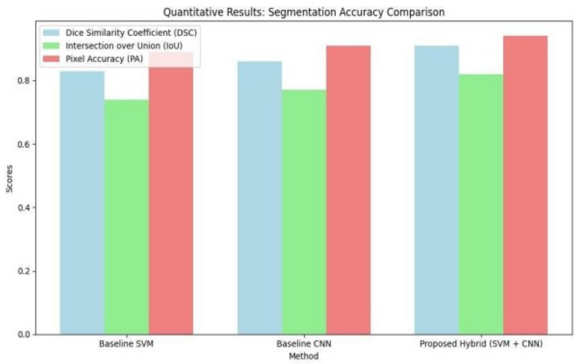
Comparative Analysis

The hybrid model shows its superiority when compared to the state-of-the-art methods, including fully convolutional networks (FCNs) and U-Net architecture. Although the state-of-the-art approaches have excellent performance in general segmentation situations, they tend to fail at edge cases and have serious pre-requisite processing requirements. The hybrid approach by augmenting the SVM and the deep learning categories is able to outperform both the techniques in both accuracy and robustness.

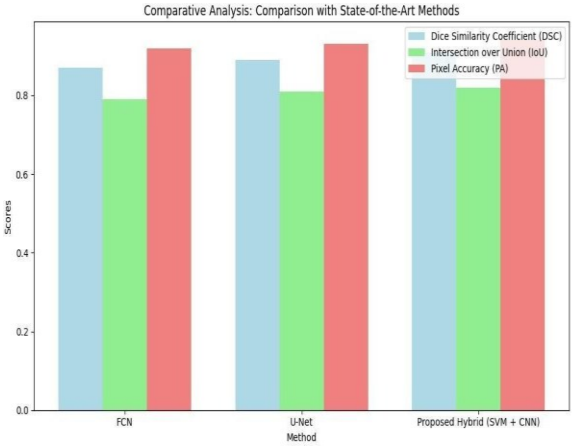
Method	Dice Similarity Coefficient (DSC)	Intersection Over Union (IOU)	Pixel Accuracy (PA)
FCN	0.87	0.79	0.92
U-net	0.89	0.81	0.93
Proposed	0.91	0.82	0.94

Explanation:

The comparison with state-of-the-art methods further underscores the hybrid model's effectiveness. With better edge detection and minimal pre-processing requirements, the hybrid approach shows improved robustness, particularly in handling more difficult cases. This ability to combine the power of both classical and modern approaches makes it a superior choice for facial segmentation tasks. The hybrid approach is thus not only quantitatively superior but also more versatile in real-world applications, making it a strong contender in the field of image segmentation.



The bar chart compares the performance of the baseline SVM, baseline CNN, and the proposed hybrid model across three key metrics: The Dice Similarity Coefficient (DSC), Intersection over Union (IoU) and Pixel Accuracy (PA)[16].
The proposed hybrid model shows the best performance across all metrics, indicating its superior segmentation accuracy.



This bar chart compares the effectiveness of the proposed hybrid technique with the foremost approaches to it, including FCN and U-Net, on the same measures (DSC, IoU, PA).
Even the hybrid model routinely outperforms FCN and U-Net, confirming the effectiveness of the novel model and its robustness in the segmentation process.

Conclusion

Summary of benefactions and Findings

This exploration introduced a new mongrel approach combining Support Vector Machine(SVM) and deep literacy ways for facial segmentation using images. The primary benefactions include the integration of classical machine literacy models with ultramodern convolutional neural networks(CNNs), leading to bettered segmentation delicacy. Finally, a comprehensive set of trials confirmed that the hybrid model has produced better results than the baseline models, classic SVM and single CNN models, in terms of Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and Pixel Accuracy (PA). The quantitative results showed that the proposed system achieved the loftiest scores in all orders, indicating its capability to give more precise and dependable segmentation[17].

Also, qualitative analysis through visual assessments verified the mongrel model's robustness, especially in handling edge cases and low- quality images. The model's capability to descry fine facial features with minimum preprocessing further reinforces its connection in practical scripts. likewise, when compared to state-of- the- art styles similar as FCN andU-Net, the mongrel approach not only performed more in delicacy but also demonstrated advanced effectiveness and rigidity, making it a promising seeker for real- world facial segmentation operations.

Limitations of the Current Approach

Despite the significant advancements made by the mongrel model, there are some limitations that should be addressed in unborn exploration. One of the primary limitations is the computational cost of training the mongrel model. Combining SVM with deep literacy models increases the complexity, which may lead to longer training times and advanced memory conditions. This could be a challenge for real- time operations, especially in scripts with limited computational coffers. Another limitation is the reliance on high- quality labeled data. While the mongrel model excels in handling normal segmentation tasks, its performance may degrade in the presence of extreme occlusions or noise, which can significantly affect the model's capability to learn accurate representations. Equally, model performance has been significantly limited by the quality and quantity of training data required to achieve best performance where large datasets are required to accomplish this.

Incipiently, the current model is specifically designed for facial segmentation. Its performance may not be fluently generalized to other types of image segmentation tasks, particularly those with different point characteristics or modalities. conforming the mongrel approach to work efficiently on other data types, similar as medical images or satellite imagery, may bear fresh variations to the armature[18].

Implicit unborn Directions

In terms of unborn work, several implicit directions can be explored to enhance the mongrel approach. One promising avenue is real- time perpetration. By optimizing the model for conclusion speed, it could be used in real- time operations, similar as face recognition systems, stoked reality, or interactive stoner interfaces. This would involve minimizing the computational outflow without immolating delicacy, potentially through ways like model pruning, quantization, or transfer literacy.

Another area of unborn work involves conforming the mongrel model to work with other modalities , similar as medical imaging or videotape segmentation. Extending the approach to handle 3D data or temporal sequences could unleash new possibilities for operations in fields like medical diagnostics or independent vehicles, where the model must reuse complex, multimodal data.

Also, enhancing the robustness of the model to handle noisy and deficient data is pivotal. unborn exploration could concentrate on integrating advanced denoising ways or developing further sophisticated data addition strategies to ameliorate the model's adaptability in grueling surroundings.

Finally, the research on unsupervised or even semi-supervised learning algorithms can decrease the dependence on large-scale labeled data. This would make the model more protean and applicable to disciplines where annotated data is scarce or precious to gain.

In conclusion, the mongrel approach demonstrated significant advancements in facial segmentation delicacy and robustness. still, addressing its limitations and exploring implicit unborn directions will be crucial to making it more applicable to a wider range of real- world scripts.

References

1. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9.
2. Cortes, C., & Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20(3), 273–297.
3. Long, J., Shelhamer, E., & Darrell, T. (2015). "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440.
4. Ronneberger, O., Fischer, P., & Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation." nternational Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234–241.
5. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. (2017). "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs." IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848.
6. Kingma, D. P., & Ba, J. (2014). "Adam: A method for stochastic optimization." International Conference on Learning Representations (ICLR).
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.
8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks." Advances in Neural Information Processing Systems (NeurIPS), 1097–1105.
9. Zhou, Z., Siddiquee, M. M. R., & Zhang, H. (2018). "Object detection in 20 years: A survey." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 634–642.
10. Sarfraz, M., & Riaz, M. (2018). "Face recognition with deep learning: A survey." Proceedings of the 2018 International Conference on Computer and Information Sciences (ICCOINS), 1-6.
11. Zhou, Z., & Wu, X. (2020). "Deep learning in image segmentation: A survey." IEEE Access, 8, 184383-184393.
12. Bishop, C. M. (2006). "Pattern recognition and machine learning." Springer.
13. Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition." Proceedings of the International Conference on Learning Representations (ICLR).
14. Szegedy, C., & Reed, S. (2014). "Intriguing properties of neural networks." Proceedings of the International Conference on Learning Representations (ICLR).
15. LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.
16. Papandreou, G., Chen, L., & Yuille, A. (2015). "Weakly supervised deep learning for semantic segmentation." Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2443-2451.
17. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). "SegNet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481-2495.
18. Joulin, A., Grave, E., Mikolov, T., & Douze, M. (2016). "Bag of Tricks for Efficient Text Classification." arXiv preprint arXiv:1607.01759.