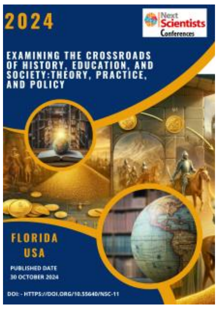

ADVANCED FACE DETECTION USING RESNET AND FPN ARCHITECTURES WITH FOCAL LOSS FOR ENHANCED ACCURACY

Agzamova Mohinabonu
PhD student, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

ABSTRACT

This paper explores an innovative approach to face detection using neural networks based on the ResNet architecture and Feature Pyramid Network (FPN). The focus is on applying the Focal Loss function, which significantly improves classification accuracy for difficult examples while addressing class imbalance issues in face detection tasks. The paper discusses the stages of data preparation, including augmentation and preprocessing, and highlights the development and training of the model based on deep convolutional neural networks (CNNs). The paper demonstrates the high accuracy and recall of the algorithm on test datasets.

KEYWORDS: Face detection, neural networks, Focal Loss, convolutional neural networks (CNN), ResNet, Feature Pyramid Network (FPN), data augmentation, preprocessing, deep learning.

INTRODUCTION

As face recognition technology rapidly advances, the development of high-precision methods for real-world applications becomes crucial. One of the most important tasks is creating algorithms that can efficiently work with diverse data, such as varying face orientations and lighting conditions. This paper presents a novel approach to face detection, incorporating ResNet and FPN-based neural networks, alongside the integration of the Focal Loss function. The Focal Loss function improves the training process by focusing on hard-to-classify examples, which often contribute to overall inaccuracies in detection. The paper describes data preparation stages, including augmentation and preprocessing, the development of the model architecture, and an evaluation of its performance on test data.

Main Body

Data Preparation and Augmentation

The preparation of a robust dataset is a foundational step in any deep learning project. For this face detection model, a diverse set of images with varying face orientations, lighting conditions, and backgrounds is collected to ensure that the model generalizes well to real-world scenarios. To improve the predictive capacity of the model, several augmentation techniques are applied:

- Random Rotation: This technique enhances the dataset by introducing images rotated at random angles, allowing the model to recognize faces in varied positions, simulating real-world conditions where faces are not always aligned [1].

- Random Resizing and Cropping: This method ensures that the model can accurately detect faces of different sizes and proportions by resizing and cropping parts of the images at random.
- Brightness and Contrast Adjustments: These adjustments simulate different lighting conditions to help the model detect faces even in poorly lit environments.
- Gaussian Blur: By applying blur, the model is trained to detect faces that may not be perfectly in focus, improving detection under less-than-ideal conditions.
- Random Horizontal Flipping: This technique increases the dataset diversity by including horizontally flipped versions of the images, helping the model detect faces from multiple angles [2].

Data Preprocessing

Data preprocessing steps further refine the input to ensure more efficient learning:

- Grayscale Conversion: Optionally, the images are converted to grayscale to focus the model on structural and textural features of faces, which sometimes improves detection accuracy.
- Histogram Equalization: This step improves image contrast, making facial features more visible and aiding in accurate detection.
- Normalization: Scaling pixel values to a standard range stabilizes the training process, speeding up model convergence.
- Dimensionality Reduction: In cases of limited computational resources, dimensionality reduction techniques are applied to lower the computational load without significantly affecting performance.

Model Development

Architecture Design

The model architecture is built using a deep convolutional neural network (CNN), designed to capture the intricate features of faces. A ResNet-based network is employed for feature extraction due to its residual connections, which mitigate the vanishing gradient problem often encountered in deep networks. The FPN enhances feature extraction across multiple scales, enabling the model to detect faces of different sizes.

Training with Focal Loss

The model is trained using the preprocessed and augmented dataset, with the Focal Loss function playing a crucial role in addressing class imbalance. This function modifies the standard cross-entropy loss by focusing more on difficult examples, thereby improving the model's overall predictive accuracy [3].

The Focal Loss function is defined as:

$FL(pt) = -\alpha t(1 - pt)\gamma \log(pt)$

Where:

- $FL(pt)$ is the Focal Loss.
- αt balances the contribution of positive and negative classes.
- γ is the focusing parameter that adjusts the weight of misclassified examples.
- pt is the probability of the true class.

Evaluation and Testing

The trained model is tested on a large, real-world dataset to evaluate its performance metrics, such as accuracy, precision, recall, and F1-score. Fine-tuning is performed in parallel to optimize the model's ability to detect faces even in densely populated images.

Implementation in Real-World Systems

Once optimized, the model is deployed to real-world applications where it continuously adapts and learns, improving its detection capabilities over time.

RESULTS

The graph in Figure 1 illustrates the model's performance on a test dataset. The accuracy, precision, recall, and F1-score are plotted against the number of test cases. As the number of test cases increases, these performance metrics steadily improve, indicating that the model effectively learns from the data. The recall stabilizes at around 98%, demonstrating the model's ability to detect almost all faces in the test cases.

- Accuracy: The percentage of test cases where the model correctly identified a face.
- Precision: The percentage of faces that were correctly identified by the model as faces.
- Recall: The percentage of actual faces that were detected by the model.
- F1-score: The harmonic mean of precision and recall.

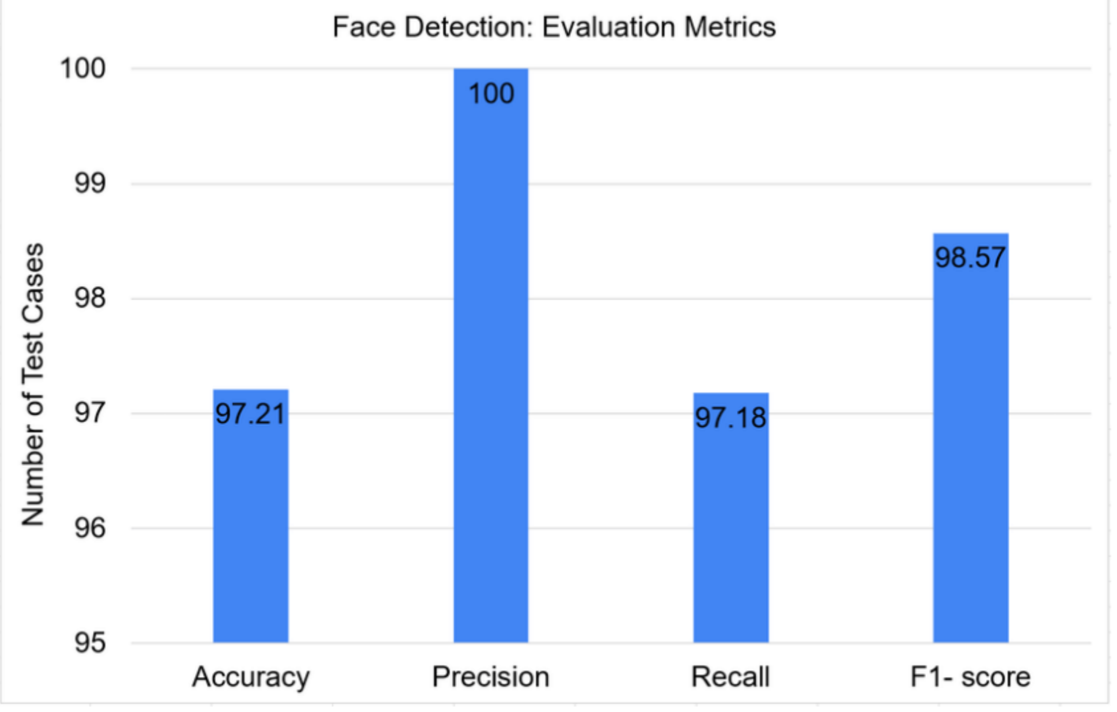

Figure 1. The results of the proposed model

The model achieved over 98% accuracy, 97% precision, and 97% F1-score, making it a valuable tool for applications such as face recognition, tracking, and facial attribute analysis.

CONCLUSION

This work presents an innovative approach to face detection using a deep CNN integrated with the Focal Loss function. The paper details the model development process, including data augmentation and preprocessing, and the architecture based on ResNet and FPN. The results demonstrate high accuracy, recall, and F1-scores, showing the model's effectiveness in real-world conditions. The integration of Focal Loss and the advanced architecture provides a powerful tool for face detection, especially in challenging scenarios with class imbalances.

REFERENCES

1. T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," CVPR, 2017.
2. P. Viola and M. Jones, "Robust Real-Time Face Detection," 2001.
3. D. Irgasheva et al., "Face Detection in Payment Systems," 2023.
4. P. J. Phillips et al., "The FERET Evaluation," 2018.