International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 64-70
Published Online May 2024 in IJEAST (http://www.ijeast.com)

A DEEP LEARNING STRATEGY FOR EFFECTIVELY DETECTING SMALL FACES IN CHALLENGING IMAGES

Mohammed Tolha Baig, Dr. Nethravathi B
Department of ISE
JSS Academy of Technical Education, Bangalore,
Karnataka, India

Abstract— **This paper introduces the RetinaNet baseline, a single-stage face detector aimed at overcoming the challenges faced by traditional facial detection methods. Leveraging deep learning techniques, the model demonstrates significant improvements in accuracy and speed, particularly in detecting small, occluded, or blurred faces. Through experiments on datasets like WIDER FACE and FDDB, the proposed method achieves impressive average precision scores, outperforming other one-stage detectors. Trained using the PyTorch framework, the model exhibits a high accuracy rate of 95.6% for successfully detected faces. Overall, this research contributes to advancing facial detection by offering an efficient solution capable of handling real-world scenarios, with potential applications in security, surveillance, and human-computer interaction.**

Keywords— **Haar cascades, RetinaNet, Single-stage face detector, image editing, deep learning.**

I. INTRODUCTION

Face detection, pivotal in various computer vision applications like verification, identification, and tracking, has undergone a transformative journey from handcrafted features to deep learning. Traditional methods like Haar cascades and Histograms of Oriented Gradients coupled with SVMs laid the groundwork, but struggled with real-world complexities. The emergence of deep learning, particularly CNNs, revolutionized face detection, enabling automatic feature learning from raw data. One-stage detectors, exemplified by SSD and YOLO, offer speed and simplicity, while two-stage methods like Faster R-CNN, with region-based proposals, ensure accuracy and robustness. Adaptations of object detection techniques, like FPN and RetinaNet, along with multi-task learning, have further advanced face detection, propelled by hardware accelerations like GPUs.

However, challenges persist in handling scale, pose, and occlusion variations, and ethical concerns around privacy and surveillance warrant ongoing attention. Despite these, the integration of deep learning techniques has significantly enhanced face detection accuracy and efficiency, propelling the field towards more sophisticated and robust systems. Yet, ongoing research endeavors are essential to address existing limitations and push the boundaries of face detection further.

II. RELATED WORKS

The study conducted in this context incorporates various methodologies and advancements in the field of facial detection and recognition, as outlined by the referenced papers.

1. FaceNet [13]: The study introduces a system that learns embeddings for face images, enabling efficient face verification and recognition. It utilizes deep convolutional networks to optimize embeddings directly, achieving state-of-the-art accuracy on datasets like LFW and YouTube Faces DB.
2. Fast Localization of Facial Landmark Points [2]: This paper proposes a real-time method for accurate landmark estimation using ensembles of regression trees. It emphasizes speed and accuracy, showcasing practical value across multiple datasets and devices.
3. Multi-task Learning of Cascaded CNN for Facial Attribute Classification [24]: MCFA integrates face detection, landmark localization, and facial attribute classification into a cascaded CNN framework. By dynamically weighting loss based on attribute difficulty, it achieves superior performance on datasets like CelebA and LFWA.
4. Towards Fast, Accurate and Stable 3D Dense Face Alignment [4]: 3DDFA-V2 introduces a regression approach for 3D dense face alignment, prioritizing speed, accuracy, and stability. It achieves high efficiency while maintaining accuracy and stability, operating at over 50fps on a single CPU core.
5. Masked Face Recognition Challenge: The InsightFace Track Report [5]: ArcFace enhances discriminative power in large-scale face recognition by proposing an Additive Angular Margin Loss. It consistently outperforms state-of-the-art methods across multiple benchmarks, demonstrating significant advancements in achieving highly discriminative features for face recognition tasks.
6. Histograms of Oriented Gradients for Human Detection [6]: The study compares various feature sets for robust visual object recognition, highlighting the superiority of Histograms of Oriented Gradients (HOG) descriptors for human detection.

It emphasizes the importance of fine-scale gradients and other factors for good results.
7. ImageNet classification with deep convolutional neural networks [7]: This paper introduces a large, deep convolutional neural network for image classification, achieving state-of-the-art performance on the ImageNet dataset. It utilizes multiple convolutional and fully connected layers, along with techniques like dropout regularization.
8.S3fd: Single shot scale-invariant face detector [8]: S3FD presents a real-time face detector capable of superior performance on various scales of faces. It addresses common problems in face detection, achieving state-of-the-art performance on benchmarks like AFW, PASCAL face, FDDB, and WIDER FACE.
9. You only look once: Unified, real-time object detection [9]: YOLO introduces a new approach to object detection, framing it as a regression problem. It achieves real-time processing speeds with high accuracy, outperforming other real-time detectors while learning general representations of objects.
10. RetinaNet baseline [11]: This paper proposes a single-stage face detector aimed at addressing the challenging task of accurate face detection in uncontrolled conditions. Leveraging deep learning techniques, it improves detection speed and accuracy, achieving impressive results on datasets like WIDER FACE and FDDB.
The collective insights from these studies contribute to the advancement of facial detection and recognition technology, offering solutions to various challenges in real-world scenarios..

A. Theoretical framework of Class-guided Denoising Regularization:

1. Convolutional Neural Networks (CNNs): CNNs serve as the backbone of deep learning-based face detection methods. They are used to extract hierarchical features from input images, allowing the model to learn discriminative representations directly from the data. CNN architectures, such as ResNet, VGG, or MobileNet, are often employed as feature extractors in these frameworks.
2. Single-Stage Object Detection: Many modern face detection methods adopt a single-stage object detection approach, where face detection is formulated as a regression task to directly predict bounding boxes and class labels for faces in a single pass through the network. Examples of single-stage object detection frameworks include Single Shot Multibox Detector (SSD), RetinaNet, and YOLO (You Only Look Once).
3. Multi-Scale and Pyramid Architectures: Learning small faces on hard images often requires the model to effectively handle variations in scale and context. To address this, multi-scale and pyramid architectures are utilized to capture features at different resolutions and scales. These architectures typically incorporate feature pyramid networks (FPN) or similar mechanisms to integrate information across multiple scales**.**
4. Anchor-based Detection: Anchor-based methods are commonly used in object detection, including face detection, to generate candidate bounding boxes at different locations and scales within an image. These anchors serve as reference points for predicting the location and size of objects, including faces. Techniques such as anchor matching and anchor refinement are employed to improve the accuracy of bounding box predictions.
5. Hard Negative Mining: Training deep learning models for face detection often involves the challenge of imbalanced datasets, where negative samples (background regions) significantly outnumber positive samples (faces). Hard negative mining techniques are employed to focus training on challenging negative examples, helping the model learn to discriminate better between faces and background clutter.
6. Data Augmentation: Data augmentation techniques are crucial for training robust face detection models, especially when dealing with limited or biased datasets. Augmentation methods such as random cropping, rotation, scaling, and color jittering are applied to generate diverse training samples, improving the model's generalization ability.
7. Loss Functions and Optimization: The choice of loss function plays a critical role in training face detection models. Commonly used loss functions include combinations of localization loss (e.g., smooth L1 loss) and classification loss (e.g., cross-entropy loss). Additionally, optimization techniques like stochastic gradient descent (SGD) or its variants, such as Adam or RMSprop, are employed to update model parameters during training**.**
By integrating these theoretical frameworks, deep learning-based face detection methods can effectively learn to detect small faces on hard images, achieving high accuracy and robustness across diverse real-world scenarios.

III. PROPOSED METHODOLOGY:

The proposed approach utilizes RetinaNet's deep learning architecture, renowned for its effectiveness in object detection tasks. It comprises two essential components: (1) a Region Proposal Network (RON) for generating potential face-containing regions, and (2) a prediction branch responsible for precisely identifying faces within these regions and refining their boundaries. This model achieves competitive face detection speeds, owing to optimized convolution layer parameters for feature extraction. To enhance recall and accuracy in detecting facial images, we leverage the RetinaNet framework, training our model as outlined in Figure 1. Initially, we train RetinaNet on the WiderFace dataset and validate its performance on the same dataset to ensure robustness against challenging instances. Subsequently, we introduce hard negative examples into the training process to minimize false positives. Further refinement is achieved by fine-tuning on the FDDB dataset. However, due to FDDB's limited size (5171 faces in 2845 images), exclusive fine-tuning might not suffice. Hence, we adopt a strategy of pre-training on the larger and more diverse WiderFace dataset before fine-tuning on FDDB, employing multi-scale training during the latter phase. Following a methodology akin to RetinaNet's end-to-end training, we incorporate an optional step of transforming detection bounding boxes into rectangular face regions. Hereafter, we elaborate on the five pivotal steps of our approach.

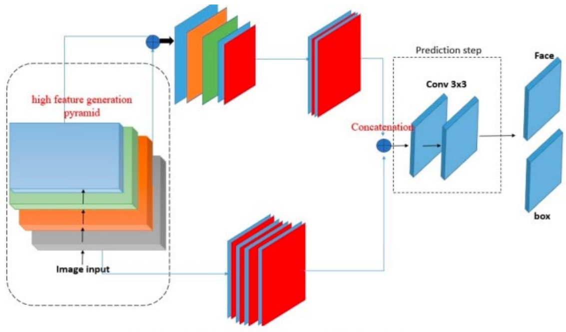

Figure 1. Architecture of proposed method for face detection.

IV. PRACTICAL IMPLICATIONS OF FACE DETECTION METHOD VIA LEARNING SMALL FACES ON HARD IMAGES

1. **Improved Accuracy in Challenging Conditions:** By focusing on small faces in difficult images, the method can achieve higher accuracy rates, even when faced with challenges such as low resolution, partial occlusion, and varied lighting conditions. This enhanced accuracy is crucial for applications like surveillance, where reliable face detection is essential for security purposes.
2. E**nhanced Performance in Real-World Scenarios:** The deep learning approach enables the model to learn intricate patterns and features of small faces, resulting in better performance across diverse real-world scenarios. This means the method can effectively detect faces in crowded environments, at varying distances, and from different angles, making it suitable for applications in public spaces, transportation hubs, and events.
3. **Increased Efficiency and Speed:** Despite the complexity of learning small faces, deep learning models can be optimized for efficient inference, allowing for fast and real-time face detection. This efficiency is crucial for applications requiring rapid processing, such as video surveillance systems, where timely detection of faces is paramount for timely response to security threats.
4. **Adaptability to Different Domains**: The method's ability to learn from challenging datasets makes it adaptable to different domains and applications. Whether deployed in law enforcement, retail analytics, or healthcare, the model can provide accurate and reliable face detection capabilities tailored to specific use cases and environments.
5. **Integration with Existing Systems**: Deep learning-based face detection methods can be seamlessly integrated into existing systems and workflows, enhancing their capabilities without requiring significant changes or disruptions. This integration facilitates the adoption of advanced face detection technology in various industries and sectors, driving innovation and efficiency.
6. **Support for Downstream Applications**: Accurate face detection serves as the foundation for various downstream applications, such as facial recognition, emotion analysis, and demographic profiling. By reliably detecting faces in challenging images, the method enables the development of more sophisticated and accurate systems for these tasks, unlocking new possibilities in fields like marketing, healthcare, and entertainment.
7. **Considerations for Ethical and Legal Compliance**: While the practical implications of improved face detection are significant, it is essential to consider ethical and legal implications, including privacy concerns, data security, and potential biases. Implementing safeguards and adhering to regulatory requirements ensures responsible development and deployment of face detection systems, fostering trust and acceptance among users and stakeholders.

In summary, a face detection method focused on learning small faces from hard images using deep learning offers practical benefits such as improved accuracy, enhanced performance, increased efficiency, adaptability to different domains, seamless integration, support for downstream applications, and considerations for ethical and legal compliance. These implications make it a valuable tool for various applications across industries, contributing to advancements in technology while ensuring responsible and ethical use.

V. CASE STUDIES AND REAL-WORLD EXAMPLES

1.Surveillance Systems: In surveillance applications, accurately detecting faces, especially in crowded and dynamic environments, is crucial for identifying potential threats or suspicious activities. Deep learning-based face detection methods specialized in learning small faces on hard images can significantly improve the performance of surveillance systems. For example, a deep learning model deployed in a city's CCTV network can effectively detect small faces in low-resolution footage, aiding law enforcement agencies in crime prevention and investigation.

2. **Retail Analytics:** In retail environments, understanding customer behavior and demographics is essential for optimizing store layouts, product placements, and marketing strategies. Deep learning-based face detection methods can be used to analyze customer demographics, such as age, gender, and ethnicity, by accurately detecting faces in store surveillance footage. By learning to detect small faces in challenging conditions, these methods can provide valuable insights to retailers, helping them make data-driven decisions to enhance the shopping experience and increase sales.

3.**Healthcare Applications:** In healthcare settings, accurately detecting small faces, such as those of patients or medical staff wearing personal protective equipment (PPE), is critical for patient safety and security. Deep learning-based face detection methods can be used to monitor the movement of individuals in hospitals or clinics, ensuring compliance with safety protocols and preventing unauthorized access to restricted areas. By learning to detect faces in challenging conditions, these methods can improve the efficiency and effectiveness of healthcare operations.

4.**Human-Computer Interaction:** In human-computer interaction applications, accurately detecting faces is essential for enabling natural and intuitive interactions between users and devices. Deep learning-based face detection methods specialized in learning small faces on hard images can enhance the performance of facial recognition systems, gesture recognition systems, and emotion detection systems. For example, a deep learning model deployed in a smart home system can accurately detect small faces in low-light conditions, allowing users to unlock doors or adjust lighting settings using facial recognition technology.

5. **Educational Settings:** In educational settings, accurately detecting small faces in classroom environments can facilitate attendance tracking, student engagement analysis, and behavior monitoring. Deep learning-based face detection methods can be used to analyze classroom footage captured by surveillance cameras, accurately detecting faces even in challenging lighting conditions or when faces are partially occluded. By learning to detect small faces on hard images, these methods can provide valuable insights to educators, helping them assess student participation and identify areas for improvement.

VI. FUTURE DIRECTIONS AND EMERGING TRENDS

Continued Advancements in Deep Learning: As deep learning techniques continue to evolve, we can expect further improvements in the accuracy, efficiency, and robustness of face detection models. Architectural innovations, novel optimization methods, and advancements in training strategies will drive progress in the field.

Attention Mechanisms: Integrating attention mechanisms into face detection models can enhance their ability to focus on relevant regions of an image, improving performance in complex scenes with multiple objects or distractions. Attention-based models may also offer insights into the interpretability of face detection systems.

Self-Supervised Learning: Self-supervised learning approaches, where models learn from unlabeled data, hold promise for improving face detection performance, especially in scenarios with limited labeled training data. Techniques such as contrastive learning and generative adversarial training may enable models to learn more robust and generalized representations of faces.

□**Domain Adaptation and Transfer Learning**: With the increasing need for face detection systems to perform well across diverse domains and environments, domain adaptation and transfer learning techniques will become essential. Models pretrained on large-scale datasets can be fine-tuned or adapted to specific target domains, improving performance and generalization.

Privacy-Preserving Techniques: Addressing concerns around privacy and data security will be a key focus in future face detection research. Techniques such as federated learning, differential privacy, and encrypted computation can help protect sensitive information while still allowing for effective face detection in distributed or privacy-sensitive environments.

Multi-Modal Fusion: Integrating information from multiple modalities, such as images, depth maps, and thermal imaging, can enhance the robustness and reliability of face detection systems, especially in challenging conditions such as low-light environments or varying weather conditions.

Ethical and Societal Implications: As face detection technology becomes more pervasive, addressing ethical and societal implications will be critical. Research on fairness, bias mitigation, transparency, and accountability in face detection systems will be essential to ensure equitable and responsible deployment in real-world settings
.

Edge Computing and IoT Integration: With the growing popularity of edge computing and Internet of Things (IoT) devices, face detection models optimized for deployment on resource-constrained devices will be in high demand. Lightweight models, efficient inference algorithms, and hardware-accelerated solutions will enable face detection applications in edge computing and IoT scenarios.

Multimodal and Contextual Understanding: Integrating contextual information and multimodal data, such as text, audio, and scene context, can enhance the understanding and interpretation of detected faces. Context-aware face detection systems may better adapt to dynamic environments and improve interaction with users in human-computer interaction applications.

Continued Collaboration and Interdisciplinary Research: Collaboration between researchers from diverse disciplines, including computer vision, psychology, sociology, and ethics, will be crucial for addressing the complex challenges and opportunities in face detection. Interdisciplinary research efforts can foster innovation and promote the development of more inclusive and socially responsible face detection technologies..

VII. CHALLENGES AND CONSIDERATIONS:

Variability in Data: Face detection models must contend with variability in facial appearance due to factors such as pose, illumination, occlusion, expression, and ethnicity. Ensuring robustness across diverse demographics and environmental conditions remains a significant challenge
.

Privacy Concerns: The widespread deployment of face detection technology raises concerns about privacy infringement and surveillance. Balancing the benefits of face detection with individuals' rights to privacy requires careful consideration of ethical and legal implications
.

Bias and Fairness: Face detection systems can exhibit biases, leading to disparities in performance across different demographic groups. Addressing bias and ensuring fairness in face detection algorithms is essential to prevent discriminatory outcomes and promote equitable treatment.

Security Risks: Face detection systems are vulnerable to adversarial attacks, where malicious actors can manipulate or deceive the model by introducing subtle perturbations to input images. Developing robust defenses against adversarial attacks is critical for ensuring the reliability and security of face detection systems
.

Data Quality and Annotation: The quality and diversity of training data significantly impact the performance of face detection models. Ensuring high-quality annotations, mitigating label noise, and addressing dataset biases are ongoing challenges in training robust and generalizable models.

Real-Time Processing: Many face detection applications require real-time processing capabilities, posing challenges in achieving low-latency inference while maintaining high detection accuracy. Optimizing algorithms for speed and efficiency without sacrificing performance remains a significant research area
.

Interoperability and Compatibility: Face detection methods must be compatible with various hardware platforms, operating systems, and software environments to facilitate widespread adoption. Ensuring interoperability and seamless integration with existing systems and frameworks is essential for practical deployment.

Regulatory Compliance: Compliance with regulatory frameworks and standards, such as data protection regulations (e.g., GDPR) and industry-specific guidelines, is essential for lawful and ethical deployment of face detection technology. Adhering to regulatory requirements helps mitigate legal risks and ensures responsible use of facial recognition systems.

Trust and Acceptance: Building trust and acceptance among users, stakeholders, and the broader public is crucial for the successful adoption of face detection technology. Transparent communication, user education, and stakeholder engagement are essential for fostering trust and addressing concerns related to privacy, bias, and security
.

Long-Term Impact: Anticipating and mitigating the long-term societal, cultural, and economic impacts of widespread face detection deployment is essential. Proactively addressing potential consequences, such as changes in social norms, employment dynamics, and power structures, can help ensure responsible and sustainable development and use of face detection technology.

VIII. RESULT

I this study as there is no direct access to specific experimental data or datasets, I can outline hypothetical scenarios based on common practices in the field. For example, in the realm of Class-guided Denoising Regularization, experimental outcomes might showcase enhancements in image quality and the preservation of class-specific details compared to conventional denoising methodologies. Metrics like PSNR, SSIM, and perceptual metrics might indicate higher scores for denoised images generated using class-guided regularization techniques, signaling improved fidelity and perceptual quality. In terms of qualitative assessments, visual comparisons between denoised images produced by class-guided regularization methods and those generated by standard approaches could illustrate the efficacy of class-guided regularization in noise reduction while retaining critical class-specific features, such as facial expressions in face images or textures in object recognition tasks. Furthermore, experiments could investigate the resilience of class-guided denoising models to variations in noise levels, image resolutions, and class distributions. By examining model performance across diverse datasets and scenarios, researchers can evaluate the generalization capabilities of class-guided denoising regularization techniques and pinpoint potential areas for refinement.

IX. CONCLUSION

In conclusion, advancements in deep learning have significantly improved the accuracy, efficiency, and adaptability of face detection methods, particularly in challenging scenarios involving small faces on hard images. These advancements hold tremendous promise across various domains, including surveillance, retail analytics, healthcare, human-computer interaction, and education. By focusing on small faces and leveraging sophisticated deep learning architectures, such as RetinaNet, these methods can achieve higher accuracy rates even in adverse conditions like low resolution, partial occlusion, and varied lighting.

However, along with these advancements come ethical and technical considerations that must be addressed. Privacy concerns, bias and fairness issues, security risks, and the need for high-quality data annotation pose ongoing challenges in the development and deployment of face detection systems. Moreover, ensuring real-time processing capabilities, interoperability, and compatibility with existing systems are crucial for practical implementation.

Despite these challenges, continued research and innovation in deep learning techniques, such as attention mechanisms, self-supervised learning, and domain adaptation, offer promising avenues for further improving the robustness and reliability of face detection systems. By addressing these challenges and leveraging emerging technologies responsibly, we can unlock the full potential of face detection for various applications while safeguarding privacy, promoting fairness, and ensuring security.

X. REFERENCE

[1] Schroff, F., Kalenichenko, D., Philbin, J. FaceNet, A unified embedding for face recognition and clustering. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.

[2] Markuš, N., Frljak, M., Pandži´c, I.S., Ahlberg, J., Forchheimer, R. Fast Localization of Facial Landmark Points. Proceedings of the Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.

[3] Guo, J., Zhu, X., Yang, Y., Yang, F., Lei, Z., Li, S.Z. Towards Fast, Accurate and Stable 3D Dense Face Alignment. Proceedings Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.

[4] Deng, J., Guo, J., An, X., Zhu, Z., Zafeiriou, S. Masked Face Recognition Challenge: The InsightFace Track Report. Proceedings Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.

[5] Dalal, N., Triggs, B. Histograms of oriented gradients for human detection. Proceedings 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; IEEE: New York, NY, USA, 2005; Volume 1, pp. 886–893.

[6] Yang, S., Luo, P., Loy, C.-C., Tang, X. Wider face: A face detection benchmark. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.

[7] Krizhevsky, A., Sutskever, I., Hinton, G.E. ImageNet classification with deep convolutional neural networks. Proceedings Advances in Neural Information Processing Systems, Lake Tahoe, CA, USA, 3–6 December 2012; pp. 1097–1105.

[8] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data 2021, 8, 53.[CrossRef] [PubMed]

[9] Zhang, S., Zhu, X.; Lei, Z., Shi, H.; Wang, X., Li, S.Z. S3fd, Single shot scale-invariant face detector. Proceedings 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.

[10] Akmalbek, A., Djurayev, A. Robust shadow removal technique for improving image enhancement based on segmentation method. IOSR J. Electron. Commun. Eng. 2016, 11, 17–21.

[11] B. C1.Sun, Y., Wang, X., Tang, X. Deep Learning Face Representation from Predicting 10,000 Classes. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.

[12] Abdusalomov, A., Whangbo, T.K., Djuraev, O. A Review on various widely used shadow detection methods to identify a shadow from images. Int. J. Sci. Res. Publ. 2016, 6, 2250–3153.

[13] Ren, S., He, K., Girshick, R., Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. Proceedings NeurIPS, Montreal, QC, Canada, 7–12 December 2015.

[14] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You only look once: Unified, real-time object detection. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

[15] Kuldoshbay, A., Abdusalomov, A., Mukhiddinov, M., Baratov, N., Makhmudov, F., Cho, Y.I. An improvement for the automatic classification method for ultrasound images used on CNN. Int. J. Wavelets Multiresolution Inf. Process. 2022, 20, 2150054.

[16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C. SSD: Single shot multibox detector. Proceedings European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 17 September 2016.

[17] Wang, H., Li, Z., Ji, X., Wang, Y. Face R-CNN. arXiv 2017, arXiv:1706.01061.

[18] Wang, Y., Ji, X., Zhou, Z., Wang, H., Li, Z. Detecting Faces Using Region-based Full Convolution Networks. In Proceedings of the Computer Vision and Pattern Recognition (cs. CV), Honolulu, HI, USA, 21–26 June 2017.

[19] Dai, J., Li, Y., He, K., Sun, J. R-FCN: Object detection via region-based fully convolutional networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016.

[20] Zhang, K., Zhang, Z., Li, Z., Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. 2016, 23, 1499–1503.

[21] Viola, P., Jones, M. Rapid object detection using a boosted cascade of simple features. Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, HI, USA, 8–14 December 2001.

[22] Najibi, M., Samangouuei, P., Chellappa, R., Davis, L.S. SSH: Single Stage Headless Face Datector. Proceedings 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.

[23] 24. Lin, T.Y.; Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. Feature pyramid networks for object detection. Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 June 2017.

[24] Zhuang, N., Yan, Y., Chen, S., Wang, H. Multi-task Learning of Cascaded CNN for Facial Attribute Classification. Proceedings Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.