Towards a Deeper insight into Face Detection in Neonatal wards

Yisheng Zhao[1][0000−0002−0763−3658], Huaiyu Zhu[1](✉)[0000−0001−6918−4088], Qi Shu[2][0000−0001−7733−7099], Ruohong Huan[3][0000−0003−2555−343X], Shuohui Chen[4], and Yun Pan[1](✉)[0000−0002−9335−4291]

[1] College of Information Science and Electronic Engineering, Zhejiang University, China
{zhaoys,zhuhuaiyu,panyun}@zju.edu.cn
[2] Nursing Department, Children's Hospital, Zhejiang University School of Medicine, China
22118788@zju.edu.cn
[3] College of Computer Science and Technology, Zhejiang University of Technology, China
huanrh@zjut.edu.cn
[4] Hospital Infection-Control Department, Children's Hospital, Zhejiang University School of Medicine, China chcsh2@zju.edu.cn

Abstract. Neonatal face detection is the prerequisite for face-based intelligent medical applications. Nevertheless, it has been found that this area has received minimal attention in existing research. The paucity of open-source, large-scale datasets significantly constrains current studies, which are further compounded by issues such as large-scale occlusions, class imbalance, and precise localization requirements. This work aims to address these challenges from both data and methodological perspectives. We constructed the first open-source face detection dataset for neonates, involving images from 1,000 neonates in the neonatal wards. Utilizing this dataset and adopting NICUface-RF as the baseline, we introduce two novel modules. The hierarchical contextual classification aims to improve the positive/negative anchor ratios and alleviate large-scale occlusions. Concurrently, the DIoU-aware NMS is designed to preserve bounding boxes of superior localization quality by employing predicted DIoUs as the ranking criterion in NMS procedures. Experimental results illustrate the superiority of our method. The dataset and code is available at https://github.com/neonatal-pain.

Keywords: Neonatal Face detection · Neonatal dataset · Neonatal care.

1 Introduction

In recent years, the advancement of Artificial Intelligence (AI) in medicine has spurred the development of numerous neonatal intelligent care, monitoring, and adjunctive diagnostic applications. These innovations include non-invasive pain assessment[25], face-based real-time monitoring of physiological signals (e.g., respiratory rate[11]) and behavioral states (e.g., sleep state [1]), and the early detection of conditions such as jaundice [16]. The effectiveness of these applications relying on neonatal face detection. The accuracy of facial detection methods, especially in video surveillance contexts, is crucial for early disease diagnosis and immediate medical intervention. Any shortcomings

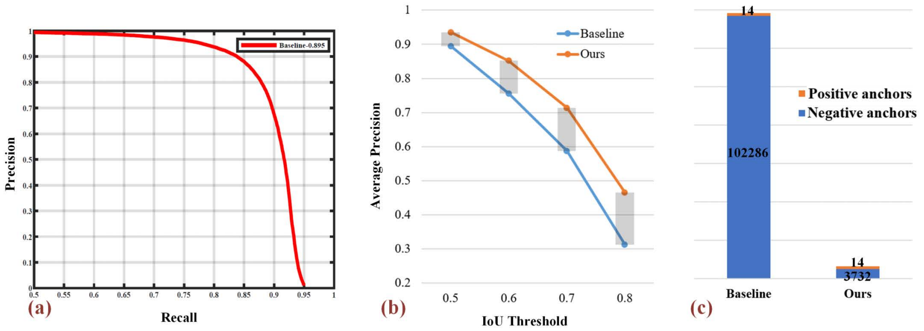

Fig. 1: (a) The Precision-Recall curve of baseline on the NFD dataset; (b) Localization accuracy on the NFD dataset; (c) Number of positives and negative anchors on the NFD dataset.

or mistakes in detection can delay diagnosis and worsen health outcomes. Therefore, improving facial detection accuracy is essential for the performance and reliability of face-based intelligent medical systems, bearing significant clinical significance

However, we found that the field of neonatal face detection has been notably underexplored. In terms of datasets, there is a lack of open-source neonatal face detection datasets; in terms of method, existing studies [17,7] have not considered the bottlenecks of neonatal face detection in complex ward scenarios and proposed the targeted design. They just adapted designs intended for adult face detection.

To this end, with the approval of the Ethics Committee (anonymous), we built the first open-source Neonatal Face Detection (NFD) dataset. This dataset contains 5000 images (5 images per neonate). After the data collection, data annotation experts first conducted the annotation process, which was reviewed medical experts specialized in neonatal care.

We adopted the state-of-the-art (SOTA) neonatal face detector, NICUface-RF [7], as our baseline. Based on its experimental results, we found that NICUface-RF does not detect many actual faces. As shown in Figure 1.a, its highest recall rate barely reaches 95%, leaving 5% of faces undetected. The Precision-Recall curve's shape does not extend sufficiently to the right or adequately steep. Such limitations in detecting all potential faces are untenable for detectors that serve as the foundation for intelligent medical systems. Missing critical physiological and health indicators, such as expressions of pain or signs of respiratory distress, could severely compromise the precision of medical assessments. In addition, as shown in Figure 1.b, the localization accuracy of NICUface-RF requires enhancement, evidenced by the dramatic decline in Average Precision (AP) with increasing Intersection over Union (IoU) thresholds.

Concerning the observed low recall rates, we attribute this phenomenon partially to large-scale occlusions caused by hospital equipment or free-moving limbs. Furthermore, the standard practice of spacing neonatal beds widely apart reduces the number of neonates in the image. This situation creates a substantial imbalance in the positive/negative anchor ratio (Figure 1.c) within the NICUface-RF model, predisposing it towards a higher propensity for false negative predictions.

Building upon these insights, we propose the Hierarchical Contextual Classification (HCC) Branch. In the first step, it filters out many negative anchors. This process amplifies the positive/negative anchor ratio approximately 27-fold. In the second step, taking inspiration from human visual perception—where individuals instinctively leverage contextual cues like body characteristics of the observed subject to identify faces in situations where direct facial features are occluded—we propose to incorporate adjacent contextual information into the relevant anchors.

Regarding the localization accuracy, NICUface-RF relies on classification confidence to select bounding boxes during the Non-Maximum Suppression (NMS) process. Nonetheless, a notable discrepancy exists between classification confidence and localization accuracy, where a high classification probability does not inherently guarantee precise localization. To mitigate this issue, we introduce a DIoU-Aware NMS (DAN) strategy, employing the predicted Distance IoU (DIoU) as the ranking criterion within the NMS procedure to prioritize the retention of bounding boxes exhibiting superior localization quality. The experimental results show that our proposed modules enhance both the recall efficiency and the accuracy of localization beyond the baseline. We also quantitatively analyze the impact of our method on downstream facial analysis task.

2 Related work

2.1 Neonatal face detection

Dataset: DOSSO et al. [7] built the CHEO dataset, which includes data from 33 neonates. However, the CHEO dataset remains inaccessible to the public. Furthermore, they annotated two additional datasets: the COPE dataset [3,2] and NBHR dataset [10]. Despite their contributions, the images within these datasets are cropped, failing to meet the requirements for detecting neonatal faces in unconstrained environments. Olmi et al. [17] developed a neonatal face detection dataset comprising 42 full-term newborns, yet this dataset is also unavailable. **Method**: DOSSO et al. [7] employed RetinaFace [5] and YOLO5Face [18] without introducing method improvements. Similarly, Olmi et al. [17] utilized the ACF [6] object detector without proposing improvements.

2.2 Adult face detection

Face detection, a specialized domain within object detection, commonly utilizes architectures originally developed for generic object detection tasks. The methods employed in adult face detection can be categorized based on their architectural foundations, including [21,15] based on Faster R-CNN [19], [26] based on R-FCN[4], [22,12] based on SSD [14], [24,23,28,5] based on RetinaNet [13]. Our method also adopts the RetinaNet framework. We propose the integration of contextual information to address the challenge of large-scale occlusions in facial detection. Current research in adult facial detection lacks solutions that can handle large-scale occlusions and alleviate class imbalance in the classification head.

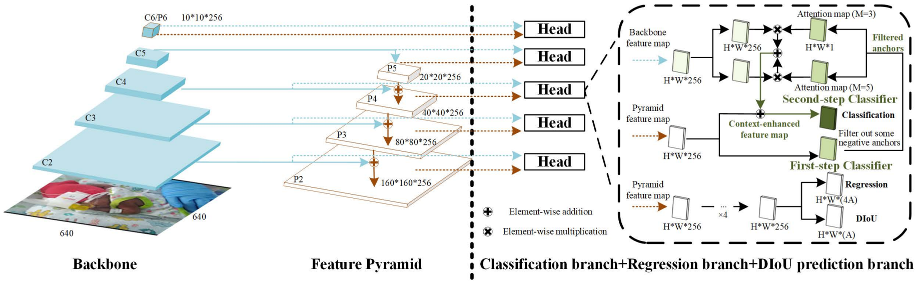

Fig. 2: The pipeline of our method.

3 NFD Dataset

3.1 Challenges

To develop a neonatal face detection dataset suitable for real-world applications and complex clinical settings, we navigated multiple challenges: adhering to ethical and privacy standards in image collection, acquiring a diverse set of high-quality images on a large scale, and ensuring accurate, consistent annotation, which demands specialized expertise.

3.2 Data Collection Protocol

This study was approved by the Ethics committee. Guardians of potential participants were informed about the objectives and methodologies of the study. Written consent was secured from the parents of all enrolled neonates. The study involved 1000 neonates with gestational ages ranging from 24 gestational weeks (GW) 6 days to 38 GW 5 days, with a mean of 32 GW 2.1 days. The collected images cover complex facial poses, diverse facial occlusions, differentlighting conditions, diverse clinical procedures, and different camera views.

3.3 Labels and Dataset Division

To ensure the reliability of the annotations, trained data annotators initially performed the task, which was subsequently reviewed and refined by specialists in neonatal care. The bounding boxes were delineated to encompass the area from the forehead to the chin and from one ear to the other. The dataset was stratified into training, validation, and test sets, adhering to a distribution ratio of 6:1:3, a process that maintains patient exclusivity.

4 Method

The pipeline of our method is presented in Figure 2. The baseline is a single-stage anchor-based face detector. It utilizes feature pyramid (P_2 to P_6), where P_2 to P_5 are derived from the corresponding outputs of the backbone network (C_2 to C_5) through top-down and lateral connections, while P_6 is generated through a 3×3 convolution applied to C_5 with a stride of 2. Backbone (C_1 to C_5) is based on a ResNet-50 [9] network, and P_6 is initialized using the Xavier method [8].

For each anchor, the classification branch ascertains the presence of an object at the anchor's location. Upon identifying an anchor as positive, the regression branch adjusts the anchor box's coordinates to more accurately align with the ground truth location.

4.1 Hierarchical Contextual Classification Branch

First Step——Handling Class Imbalance In the context of neonatal face detection, the classification branch of NICUface-RF encounters a class imbalance issue. For example, at an input resolution of 640×640 and employing three types of anchors in our implementation, the total anchor count escalates to 102,300, with merely a few dozen or even fewer being positive anchors. Therefore, we propose to perform hierarchical classification to initially filter out some of the negative anchors based on a predetermined confidence threshold θ. This can narrow down the search space for the second step improve the positive/negative anchor ratio.

Moreover, hierarchical classification is selectively applied across different levels of the feature pyramid. Given the large size and limited quantity of anchors at higher levels (P_5 and P_6), the prevalence of negative samples is comparatively low, rendering minimal benefits from hierarchical classification at these levels. Conversely, at the three lower pyramid levels (P_2, P_3, and P_4), which account for the majority of anchors (98.5%) and feature smaller anchor sizes, a hierarchical classification process is implemented. This design is empirically validated by experimental results. As shown in Figure 1.c, implementing hierarchical classification enhances the positive/negative anchor ratio (about 27-fold).

Second Step——Context-Driven False Negative Mitigation We found that by considering the contextual information surrounding the face, such as body, the presence and location of the face in the image can be better understood, especially when the face is not sufficiently visible due to occlusion. Inspired by this, we propose a context-driven module that encodes adjacent contextual information into relevant anchors.

Concretely, we obtain a set of filtered anchors after the first step. Following this, we utilize the positions of these filtered anchors to create two attention feature maps. Within each attention feature map, the locations of filtered anchors and their adjacent neighbor information are designated with a value of 1, with all other values set to 0. The neighbor information, centered around the target anchor point, encompasses the immediate M×M area, where M is configured as either 3 or 5 to derive varied neighbor information. We then ascertain the two neighbor context information by conducting a dot product operation between the backbone feature maps and the two attention feature maps. This procedure aims to explicitly generate more context information relevant to the filtered anchors. Then, we encode the two context information into pyramid feature maps via an element-wise summation. The resultant combined feature map is then inputted into the classifier and trained according to Equation 1.

Loss The loss function for HCC is as follows:

$$\mathcal{L}_{\mathrm{HCC}}=\frac{1}{N_1}\sum_{i\in\mathit{\Omega}}F^1\left(f_c^1,y\right)+\frac{1}{N_2}\sum_{i\in\mathit{\Phi}}\gamma*F^2\left(f_c^2,y'\right), \tag{1}$$

where N_1 and N_2 denote the number of anchors processed during the first and second steps, respectively, while $\mathit{\Omega}$ and $\mathit{\Phi}$ represent the sets of anchors at these steps. F^1 and F^2 signify the sigmoid focal loss applied to the classifiers in the first and second steps, respectively. The label y for each anchor is determined by an anchor matching strategy [19]. The weight γ aims to balance the loss between two classifiers. $f_c^{1,2}$ represent the output of two classifier. The dynamic label, y', is obtained through the following procedure: At each training iteration, we initially mask the positions of false-negative anchors and true-negative anchors based on the classification scores from the first-step classifier at the end of forward propagation. Subsequently, the positions of false-negative anchors are labeled as positive samples, those of true-negative anchors as negative samples, and the positions of the remaining anchors are treated as ignore samples.

4.2 DIoU-aware NMS

IoU's limitation lies in its focus solely on the overlap between bounding boxes, ignoring their precise positions. Consequently, two predictions with identical IoU scores can vary significantly in proximity to the true object. DIoU addresses this by incorporating the center distance between predicted and actual bounding boxes, offering a more accurate reflection of their positional accuracy. Thus, adopting DIoU over traditional IoU in NMS evaluation promises more accurate and rational detection outcomes.

Concretely, we propose a simple DIoU prediction branch to predict the DIoU value between the detected box and the corresponding ground-truth object. DIoU prediction branch is a parallel branch with the regression branch and consists of a 3 × 3 convolution layer, followed by a sigmoid function. At the inference phase, the final detection confidence is computed by the following equation,

$$\text{score }=p_i^\alpha D_i^{(1-\alpha)}, \tag{2}$$

where p_i and D are the classification score and predicted DIoU of i-th detected box, and $\alpha\in[0,1]$ is a hyperparameter. At the training phase, the binary cross-entropy loss (BCE) is adopted for the DIoU prediction loss:

$$\mathcal{L}_{\mathrm{DIoU}}=\frac{1}{N_P}\sum_{i\in\mathit{\Omega}_P}\mathrm{BCE}\left(D_i,\hat{D}_i\right), \tag{3}$$

where D_i represents the predicted DIoU for each positive anchor and $\hat{D}_i$ is the ground truth. $\hat{D}_i$ is computed as follows:

$$\hat{D}_i=1-IoU+\frac{\rho^2\left(\mathbf{b},\mathbf{b}^{gt}\right)}{c^2}, \tag{4}$$

where $\mathbf{b}$ and $\mathbf{b}^{gt}$ denote the central points of the predicted box and ground-truth box, $\rho(\cdot)$ is the Euclidean distance, and c is the diagonal length of the smallest enclosing box covering the two boxes.

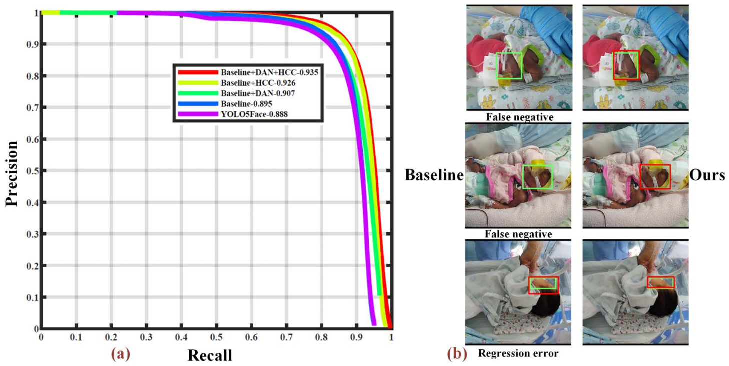

Fig. 3: (a) Precision-recall curves on NFD dataset; (b) Face detection results, where 'green' represents ground truth and 'red' represents predictions.

5 Experiments

Table 1: Inference time for input size of 640×640.

Method	Device	Inference time (ms)
Baseline	GPU [1]	32.1
Ours	GPU [1]	32.3

Table 2: Ablation studies on the NFD dataset.

Setting	AP(%)
Baseline: NICUface-RF [7]	89.5
a. + HCC	90.7
b. + DAN	92.6
c. + HCC + DAN	93.5

5.1 Implementation Details

The details of the feature pyramid, stride size, and anchor setting can be found in [5]. All the methods involved in our study were pre-trained on The WIDER FACE dataset [20] according to the original settings and then fine-tuned on the NFD dataset. We use SGD optimizer to train our model with 20 epochs (momentum at 0.9, weight decay at 5e-4, batch size of 16). The learning rate starts from 1e-3, rising to 1e-2 at the 5th epochs, then divided by 10 at the 10th and 15th epochs. Following [7], anchors were assigned to objects when the IoU surpassed 0.45 and to the background if the IoU was under 0.3. We empirically set θ to 0.9 and γ to 1. Training data augmentation was achieved through random horizontal flips and photometric color distortions.

Table 3: The AP (%) achieved by applying the negative anchor filtering to each pyramid level. P_2 means that we only apply this operation to the P_2 level.

Method	-	P_2	P_3	P_4	P_5	P_6
Baseline	89.5	90.0	89.9	89.7	89.5	89.4

Table 4: Influence of M settings in the HCC module on results.

Method	M in HCC module	AP(%)
Ours	3	93.0
Ours	5	93.2
Ours	3 and 5	93.5

[1] NVIDIA Tesla P40

Table 5: NMS strategy comparison, where '0.5' denotes the IOU, and so on.

NMS Setting	AP0.5(%)	AP0.6(%)	AP0.7(%)	AP0.8(%)
IOU-aware	93.5	85.0	70.6	45.3
DIOU-aware	93.5	85.2	71.4	46.6

Table 6: The impact of neonatal face detection method on downstream task.

Face detection setting	Pain assessment	Accuracy(%)
Baseline	RCA[27]	78
Ours	RCA[27]	88

5.2 Comparison With State-of-the-Art Methods

We present the precision-recall curves for our method, the baseline, and YOLO5Face [18] in Figure 3.a. we can see that our method outperforms the others, particularly in terms of recall rate. In Figure 1.b, we provide a comparison of the AP between our method and the baseline at various IoU thresholds. We can see that as the IoU threshold increases, our method's lead over the baseline widens, proving our method's effective enhancement of localization accuracy. In Figure 3.b, we also present several results of neonatal face detection to qualitatively illustrate the improvements our method brings in mitigating false negatives and enhancing localization accuracy. Finally, a comparison of inference times between our method and the baseline is provided in Table 1, indicating that our method introduces negligible additional time consumption, essentially meeting the requirements for real-time processing.

5.3 Ablation Study

Hierarchical Contextual Classification The precision-recall curve in Figure 3.a and the results in Table 4 prove the effectiveness of our proposed HCC. In addition, the experimental results of adding the first-step classification on each pyramid level are persented in Table 3. Consistent with our intuition, two-step classification on the three low-level fearure maps contributes to improved performance, while it is ineffective on the high-level fearure maps. In addition, we conducted sensitivity analyses on the setting of M in Table 4, which verified the reasonableness of our setting (3 and 5).

DIoU-aware NMS The precision-recall curve presented in Figure 3.a, along with the the results in Table 4, substantiate the efficacy of the proposed DAN. In addition, we conducted a further comparison between our method and IoU-aware NMS. As shown in Table 5, we can observe that our method delivers a performance enhancement at higher IoU thresholds.

5.4 Impact on downstream task

We investigate how our method enhances neonatal facial analysis, focusing on popular facial pain assessment. A pre-trained model [27] for facial pain assessment was employed to analyze pain levels in video frames, automatically selecting the highest value for each segment as the prediction. We evaluated our approach on 50 one-minute videos of neonates undergoing fingertip blood sampling in the neonatal intensive care unit. As shown in Table 6, our method improved pain assessment accuracy, attributed to a higher recall rate than the baseline, ensuring critical pain-indicative frames were not missed, thereby reducing inaccuracies in pain classification.

6 Conclusion

In this work, we build the first face-detection dataset for neonates and propose a robust face-detection method for neonates. The proposed method contains a hierarchical contextual classification branch to address the class imbalance and large-scale occlusion alongside a DIoU-aware NMS to rectify issues of inaccurate localization. Furthermore, we quantitatively analyze the advantageous impact of our method on downstream neonatal pain assessment. We plan to evaluate the impact of neonatal face detection for more downstream tasks.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62306272 and 62276237), the National Key Research and Development Program of China (2023YFC3603102), the Pioneer and "Leading Goose" Research and Development Program of Zhejiang (2024C03027).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Awais, M., Long, X., Yin, B., Abbasi, S.F., Akbarzadeh, S., Lu, C., Wang, X., Wang, L., Zhang, J., Dudink, J., Chen, W.: A hybrid dcnn-svm model for classifying neonatal sleep and wake states based on facial expressions in video. IEEE Journal of Biomedical and Health Informatics **25**(5), 1441–1449 (2021)
2. Brahnam, S., Chuang, C.F., Shih, F.Y., Slack, M.R.: Svm classification of neonatal facial images of pain. In: Fuzzy Logic and Applications: 6th International Workshop, WILF 2005, Crema, Italy, September 15-17, 2005, Revised Selected Papers 6. pp. 121–128. Springer (2006)
3. Brahnam, S., Nanni, L., Sexton, R.: Introduction to neonatal facial pain detection using common and advanced face classification techniques. In: Advanced Computational Intelligence Paradigms in Healthcare–1, pp. 225–253. Springer (2007)
4. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: object detection via region-based fully convolutional networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. pp. 379–387 (2016)
5. Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., Zafeiriou, S.: Retinaface: Single-stage dense face localisation in the wild. arXiv preprint arXiv:1905.00641 (2019)
6. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(4), 743–761 (2011)
7. Dosso, Y.S., Kyrollos, D., Greenwood, K.J., Harrold, J., Green, J.R.: Nicuface: Robust neonatal face detection in complex nicu scenes. IEEE Access **10**, 62893–62909 (2022)
8. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
10. Huang, B., Chen, W., Lin, C.L., Juang, C.F., Xing, Y., Wang, Y., Wang, J.: A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks. Engineering Applications of Artificial Intelligence **106**, 104447 (2021)

11. Khanam, F.T.Z., Perera, A.G., Al-Naji, A., Gibson, K., Chahl, J.: Non-contact automatic vital signs monitoring of infants in a neonatal intensive care unit based on neural networks. Journal of Imaging **7**(8), 122 (2021)
12. Li, J., Wang, Y., Wang, C., Tai, Y., Qian, J., Yang, J., Wang, C., Li, J., Huang, F.: DSFD: dual shot face detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5060–5069 (2019)
13. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988 (2017)
14. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. pp. 21–37. Springer (2016)
15. Najibi, M., Singh, B., Davis, L.S.: FA-RPN: Floating region proposals for face detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7723–7732 (2019)
16. Nihila, S., Rajalakshmi, T., Panda, S.S., Lhazay, N., Giri, G.D.: Detection of jaundice in neonates using artificial intelligence. In: Soft Computing: Theories and Applications: Proceedings of SoCTA 2020, Volume 2. pp. 431–443. Springer (2021)
17. Olmi, B., Manfredi, C., Frassineti, L., Dani, C., Lori, S., Bertini, G., Gabbanini, S., Lanatà, A.: Aggregate channel features for newborn face detection in neonatal intensive care units. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 455–458. IEEE (2022)
18. Qi, D., Tan, W., Yao, Q., Liu, J.: YOLO5Face: why reinventing a face detector. In: European Conference on Computer Vision. pp. 228–244. Springer (2022)
19. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. vol. 28 (2015)
20. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: A face detection benchmark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5525–5533 (2016)
21. Zhang, C., Xu, X., Tu, D.: Face detection using improved faster rcnn. arXiv preprint arXiv:1802.02142 (2018)
22. Zhang, J., Wu, X., Hoi, S.C., Zhu, J.: Feature agglomeration networks for single stage face detection. Neurocomputing **380**, 180–189 (2020)
23. Zhang, S., Zhu, R., Wang, X., Shi, H., Fu, T., Wang, S., Mei, T., Li, S.Z.: Improved selective refinement network for face detection. arXiv preprint arXiv:1901.06651 (2019)
24. Zhang, Y., Xu, X., Liu, X.: Robust and high performance face detector. arXiv preprint arXiv:1901.02350 (2019)
25. Zhao, Y., Zhu, H., Chen, X., Luo, F., Li, M., Zhou, J., Chen, S., Pan, Y.: Pose-invariant and occlusion-robust neonatal facial pain assessment. Computers in Biology and Medicine **165**, 107462 (2023)
26. Zhu, C., Tao, R., Luu, K., Savvides, M.: Seeing small faces from robust anchor's perspective. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5127–5136 (2018)
27. Zhu, H., Zhao, Y., Chen, X., Luo, F., Mei, L., Chen, S., Pan, Y.: Video-based neonatal pain assessment in uncontrolled conditions. IEEE Journal of Biomedical and Health Informatics **28**(1), 239–250 (2024)
28. Zhu, Y., Cai, H., Zhang, S., Wang, C., Xiong, Y.: Tinaface: Strong but simple baseline for face detection. arXiv preprint arXiv:2011.13183 (2020)