EfficientSRFace: An Efficient Network with Super-Resolution Enhancement for Accurate Face Detection

Guangtao Wang[1], Jun Li[1(✉)], Jie Xie[1], Jianhua Xu[1], and Bo Yang[2]

[1] School of Computer and Electronic Information, Nanjing Normal University, 210023 Nanjing, China
[2] School of Artificial Intelligence, Nanjing University of Information Science and Technology, 210044 Nanjing, China
{202243023,lijuncst,73049,xujianhua}@njnu.edu.cn, 003402@nuist.edu.cn

Abstract. In face detection, low-resolution faces, such as numerous small faces of a human group in a crowded scene, are common in dense face prediction tasks. They usually contain limited visual clues and make small faces less distinguishable from the other small objects, which poses great challenge to accurate face detection. Although deep convolutional neural network has significantly promoted the research on face detection recently, current deep face detectors rarely take into account low-resolution faces and are still vulnerable to the real-world scenarios where massive amount of low-resolution faces exist. Consequently, they usually achieve degraded performance for low-resolution face detection. In order to alleviate this problem, we develop an efficient detector termed EfficientSRFace by introducing a feature-level super-resolution reconstruction network for enhancing the feature representation capability of the model. This module plays an auxiliary role in the training process, and can be removed during the inference without increasing the inference time. Extensive experiments on public benchmarking datasets, such as FDDB and WIDER Face, show that the embedded image super-resolution module can significantly improve the detection accuracy at the cost of a small amount of additional parameters and computational overhead, while helping our model achieve competitive performance compared with the state-of-the-arts.

Keywords: deep convolutional neural network · low-resolution face detection · feature-level super-resolution reconstruction · feature representation capability

1 Introduction

With the development of deep convolutional neural networks (CNNs), dramatic progress has been made recently in face detection which is one of the most fun-

damental tasks in computer vision [1–3]. With superior representation capability, deep models have achieved unrivaled performance compared with traditional models. In pursuit of high performance, particularly, numerous heavyweight face detectors [4–6] are designed with excessive parameters and complex architecture. e.g., the advanced DSFD detector [7] has 100M+ parameters, costing 300G+ MACs. Although various lightweight designs are used for producing simplified and streamlined networks [8–10], the models trading accuracy for efficiency suffer from degraded performance. Recently, more efforts are devoted to designing efficient network, and the EfficientFace detector [11] has been proposed recently for addressing the compromise between efficiency and accuracy.
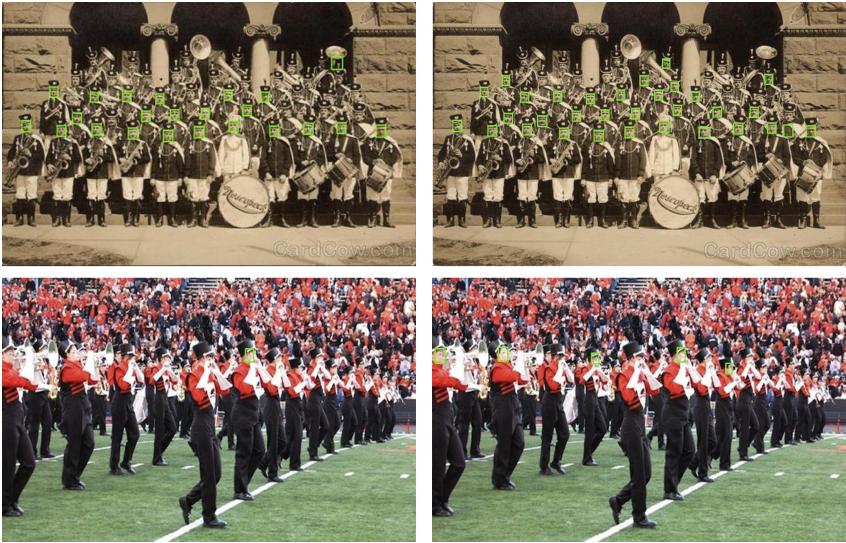

Fig. 1. Visualization of detection results achieved by EfficientFace (left) and our EfficientSRFace (right) in two different images. The results demonstrate the advantage of our network against the EfficientFace in the case of dense low-resolution face detection.

However, in real-world scenarios, low-resolution faces account for a large proportion in dense face detection tasks. For example, massive amount of small faces of a human group exist in a crowded scene for a low-quality image. They usually contain limited visual clues, making it difficult to accurately distinguish them from the other small objects and posing great challenge to accurate detection. Although current deep detectors have achieved enormous success, they are still prone to low-resolution face detection accuracy. As shown in Fig. 1, when handling the images including a large amount of densely distributed faces of a large human group with considerable variances, the EfficientFace reveals deteriorating performance without correctly identifying the low-resolution faces (Some small-scale blurred faces in the images are missing in the detection results). To allevi-

ate this problem, we embed a feature-level image super-resolution reconstruction network into EfficientFace, and design a new detection framework termed EfficientSRFace for improving the accuracy of low-resolution face detection. As a simple residual attention network, the newly added reconstruction module can enhance the feature representation capability of our detector at the cost of a small amount of additional parameters and limited computational overhead growth. Notably, the module is only introduced into the training process and is discarded for inference, and thus inference efficiency is not affected.

To summarize, our contributions in this study are twofold as follows:

- In this paper, we develop a new efficient face detection architecture termed EfficientSRFace. Based on the well-established EfficientFace, a feature-level image super-resolution network is introduced, such that the feature representation capability of characterizing low-resolution faces is enhanced.
- Extensive experiments on public benchmarking datasets show that the super-resolution module can significantly improve the detection accuracy at the cost of a small amount of additional parameters and computational consumption, while helping our model achieve competitive performance compared with the state-of-the-arts.

2   Related Work

2.1   Face Detection

With the rapid development of deep networks for general-purpose object detection [12–15], significant progress has been made in face detection. Recently, various heavyweight face detectors have been designed to realize accurate face detection [16–19]. In order to speed up computation and reduce network parameters, Najibi et al. [20] proposed SSH detector by utilizing feature pyramid instead of image pyramid and removing the fully connected layer of the classification network. Tang et al. [2] proposed a new context assisted single shot face detector termed Pyramidbox considering context information. Liu et al. [21] designed HAMBox model which incorporates an online high-quality anchor mining strategy that can compensate mismatched faces with high-quality anchors. In addition, ASFD [22] combines neural structure search techniques with a newly designed loss function. Although the above models have superior performance, they have excessive architectural parameters and incur considerable costs during the training process. Lightweight model design has become the promising line of research in face detection. One representative lightweight model is EXTD [9], which is an iterative network sharing model for multi-stage face detection and significantly reduces the number of model parameters. Despite the success of both heavyweight and lightweight detectors, they still suffer insufficient descriptive ability of capturing low-resolution face, and thus reveal inferior performance when handling low-resolution face detection in real-world scenarios.

2.2 CNNs for Image Super-Resolution

Benefiting from the promise of CNN, major breakthroughs have also been made in the field of super-resolution (SR) reconstruction. Dong et al. [23] proposed a deep learning framework for single image super-resolution named SRCNN. For the first time, convolutional networks were introduced into SR tasks. Later, they improved SRCNN and introduced a compact hourglass CNN structure to realize faster and better SR model [24]. In order to improve the performance of image reconstruction, Kim et al. [25] proposed a deeper network model named VDSR. By cascading small filters in the deep network structure multiple times, the context information can be effectively explored. Wang et al. [26] developed ESRGAN and reduced computational complexity by removing the Batch Normalization (BN) layer and adding a residual structure. Zhang et al. [27] proposed a very deep residual channel attention network, which integrates the attention mechanism into the residual block and forms the residual channel attention module to obtain high-performance reconstructed images. Kong et al. [28] proposed SR pipeline that combined classification and super-resolution on the sub-image level and tackled acceleration via data characteristics. Cong et al. [29] unified pixel-to-pixel transformation and color-to-color transformation coherently in an end-to-end network named CDTNet. Extensive experiments demonstrate that CDTNet achieves a desirable balance between efficiency and effectiveness. In this paper, in order to alleviate the drawback of existing face detectors in low-resolution face detection, an image super-resolution network is embedded into our EfficientFace network to enhance the feature expression ability of the model. To our knowledge, this is the first attempt to incorporate the SR network into efficient face detector to address low-resolution face detection.

3 EfficientSRFace

In this section, we will briefly introduce our proposed EfficientSRFace framework followed by a detailed description of the embedded feature-level super-resolution module. In addition, the loss function of our network will also be discussed.

3.1 Network Architecture

The network architecture of EfficientSRFace is shown in Fig. 2. It adopts the framework of the EfficientFace detector [11] which mainly comprises three key components of SBiFPN, RFE and attention modules. To enhance the expression capability of degraded low-resolution image features and improve the detection accuracy of blurred faces, the feature-level image super-resolution reconstruction module illustrated in dashed box is incorporated into EfficientFace in which EfficientNet-B4 is used as the backbone. Considering that the image super-resolution reconstruction result largely depends on features with sufficient representation capability, the image super-resolution module is added to the feature layer OP_2 of the EfficientFace, since the scale of the feature map at OP_2 is 1/4 of the original scale after image pre-processing, and encodes abundant visual information to guarantee accurate super-resolution reconstruction.
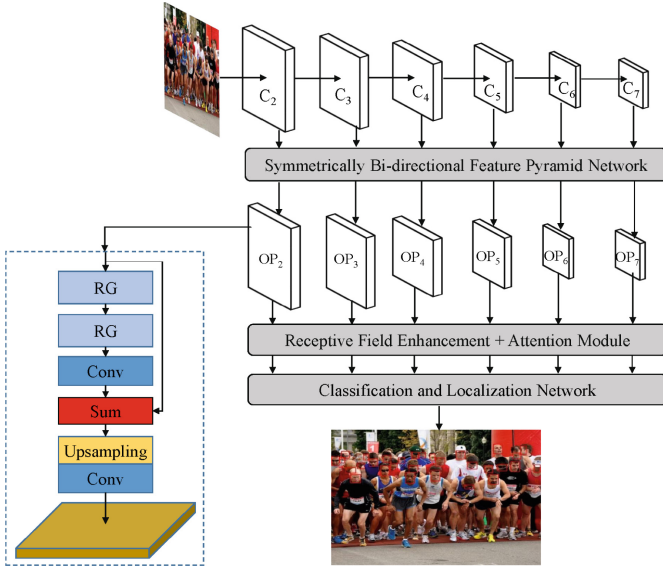
Fig. 2. The network structure of our proposed EfficientSRFace. The introduced Residual Channel Attention Network (RCAN) serving as the super-resolution module is illustrated in a dashed box. To guarantee the reconstruction quality, the module is added to the relatively larger scale feature layer OP_2.

3.2   Image Super-Resolution Enhancement

Although EfficientFace [11] achieves desirable detection accuracy when handling larger scale faces, it reports inferior performance when detecting low-resolution faces in the degraded image. In particular, numerous small faces carry much less visual clues, which makes the detector fail to discriminate them and increases the detection difficulty especially when the degraded images are not clearly captured. Consequently, we introduce Residual Channel Attention Network (RCAN) [27] within our EfficientSRFace, such that we perform feature-level super-resolution on EfficientFace for feature enhancement.

As shown in Fig. 2, Residual Group (RG) component within RCAN is used to increase the depth of the network and extract high-level features. It is also a residual structure which consists of two consecutive residual Channel Attention Blocks (RCABs). RCAB in Fig. 3 aims to combine the input features and the subsequent features prior to channel attention. Thus, it helps to increase the channel-aware weights and benefits the subsequent super-resolution reconstruction. Increasing the number of RGs contributes to further performance gains, whereas inevitably leads to excessive model parameters and computational overhead. This also increases the training difficulty of our network. For efficiency, the number of RGs is set to 2 in our scenario. Afterwards, the input low-level and high-level features resulting from RGs are fused by pixelwise addition strategy to enrich the feature information and help the network to boost the reconstruction

quality. Finally, the upsampling strategy is used to increase the scale of features with the upsampling factor within our model set as 4.

It should be noted that the RCAN module only plays a supplementary and auxiliary role during the training process, and it is discarded during reference without affecting detection efficiency within our EfficientSRFace.
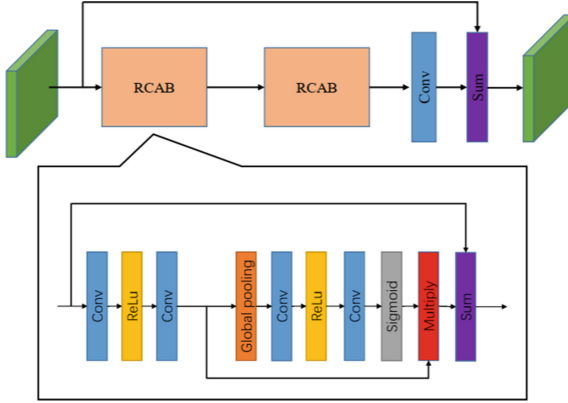

Fig. 3. Network structure of the RG module. It is essentially a residual structure which consists of consecutive Residual Channel Attention Blocks (RCABs). RCABs integrate channel-aware attention mechanism into Residual Blocks (RBs) for accurate super-resolution reconstruction. Although increasing RG modules helps training deeper network and improve SR performance, the number of RGs is set to 2 for efficiency in our case.

3.3   Loss Function

Mathematically, the overall loss function of our EfficientSRFace model is formulated as Eq. (1) which consists of three terms respectively calculating classification loss, regression loss and super-resolution reconstruction loss. Considering that our main focus is accurate detection, we assume the former two terms outweigh the SR loss and utilize the parameter φ to balance the contribution of super-resolution reconstruction to the total loss function.

$$L_{ef} = L_{focal} + L_{smooth} + \varphi L_{sr} \qquad (1)$$

More specifically, taking into account the sample imbalance, focal loss [30] is utilized for the classification loss indicated as:

$$L_{focal} = -\alpha_t(1 - p_t)^\gamma log(p_t) \qquad (2)$$

where $p_t \in [0, 1]$ is the probability estimated for the class with label 1, and α_t is the balancing factor. Besides, γ is the focusing parameter that adjusts the rate at which simple samples are downweighted.

In addition, smooth ℓ_1 is used as the regression loss for accurate face localization as follows:

$$smooth_{\ell_1}(x) = \begin{cases} 0.5x^2 & |y| < 1 \\ |y| - 0.5 & otherwise \end{cases} \tag{3}$$

In terms of super-resolution reconstruction loss, ℓ_1 loss is adopted to measure the difference between the super-resolution reconstructed image and the target image formulated as follows:

$$L_{sr} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} |y_{ij} - Y_{ij}| \tag{4}$$

where W and H respectively represent the width and the height of the input image, while y and Y respectively represent the pixel values of the reconstructed and the target image.

4   Experiments

In this section, extensive experiments are conducted to evaluate our proposed EfficientSRFace. Firstly, the public benchmarking datasets and experimental setting will be briefly introduced in our experiments. Next, comprehensive evaluations and comparative studies are also carried out with detailed model analysis.

4.1   Datasets and Evaluation Metrics

We have evaluated our EfficientSRFace network on four public benchmarking datasets for face detection including AFW [31], Pascal Face [32], FDDB [33] and WIDER Face [34]. Known as the most challenging large-scale face detection dataset thus far, WIDER Face comprises 32K+ images with 393K+ annotated faces exhibiting dramatic variances in scales, occlusion and poses. It is split into training (40%), validation (10%) and testing sets (50%). Depending on different difficulty levels, the whole dataset is divided into three subsets, namely Easy, Medium and Hard subsets. For performance measure, Average Precision (AP) and Precision-Recall (PR) curves are used for metrics in different datasets.

4.2   Implementation Details

In implementation, the anchor sizes used in our EfficientSRFace network are empirically set as {16, 32, 64, 128, 256, 512} and their aspect ratios are unanimously 1:1. In terms of the model optimization, AdamW algorithm is used as the optimizer and ReduceLROnPlateau attenuation strategy is employed to adjust the learning rate which is initially set to 10^{-4}. If the loss function stops descending within three epochs, the learning rate will be decreased by 10 times and eventually decay to 10^{-8}. The batch size is set as 4 for network training. The training and inference process are completed on a server equipped with a NVIDIA GTX3090 GPU under PyTorch framework.

Table 1. Comparison of the EfficientFace and our EfficientSRFace detector using different backbone networks.

Backbone	model	Easy	Medium	Hard	Params (M)	MACs (G)
EfficientNet-B0	EfficientFace	91.0%	89.1%	83.6%	3.89	4.80
	EfficientSRFace	92.5%	90.7%	85.8%	3.90	5.17
EfficientNet-B1	EfficientFace	91.9%	90.2%	85.1%	6.54	7.81
	EfficientSRFace	92.7%	90.9%	86.3%	6.56	8.43
EfficientNet-B2	EfficientFace	92.5%	91.0%	86.3%	7.83	10.49
	EfficientSRFace	93.0%	91.7%	87.2%	7.86	11.44
EfficientNet-B3	EfficientFace	93.1%	91.8%	87.1%	11.22	18.28
	EfficientSRFace	93.7%	92.3%	87.6%	11.27	20.06
EfficientNet-B4	EfficientFace	94.4%	93.4%	89.1%	18.75	32.54
	EfficientSRFace	**95.0%**	**93.9%**	**89.9%**	18.84	35.83

4.3   Data Enhancement

In terms of training the image super-resolution reconstruction module, the super-resolution labels are the original images, while the images preprocessed by random blur are delivered to the module. More specifically, in addition to the usual image enhancement methods such as contrast and brightness enhancement, random cropping and horizontal flip, we also leverage random Gaussian blur processing for the input images.

4.4   Results

Comparison of Different Backbones. Table 1 presents the comparison of the EfficientFace and our proposed EfficientSRFace with different backbone networks. It can be observed EfficientSRFace consistently outperforms EfficientFace with different backbones used. In particular, when EfficientNet-B0 is used as the backbone, further performance improvements of 1.5%, 1.6% and 2.2% are reported on the three respective subsets. With the increase in the complexity of the backbone network structure, slightly declined performance gains can be observed. Since EfficientNet-B0 backbone has much less parameters and enjoys more efficient structure, it is prone to insufficient representation capability. In this sense, incorporating the feature-level super-resolution module is beneficial for enhancing the feature expression capability of the backbone, and bring more performance gains compared with our model using other efficient backbones. More importantly, the auxiliary super-resolution module incurs a small amount of additional parameters and slight growth in computational overhead, which suggests it hardly affects the detection efficiency.

In addition to the detection accuracy, we also present the Frame-Per-Second (FPS) values of our EfficientSRFace models for efficiency evaluation. As shown in Fig. 4, although FPS generally exhibits a decreasing trend with the increase of image resolution, our model can still achieve real-time detection speed. For example, our model achieves 28 FPS speed for the image size of 1024×1024 when

EfficientFace-B0 is used as backbone, which fully demonstrates the desirable efficiency of our EfficientSRFace.
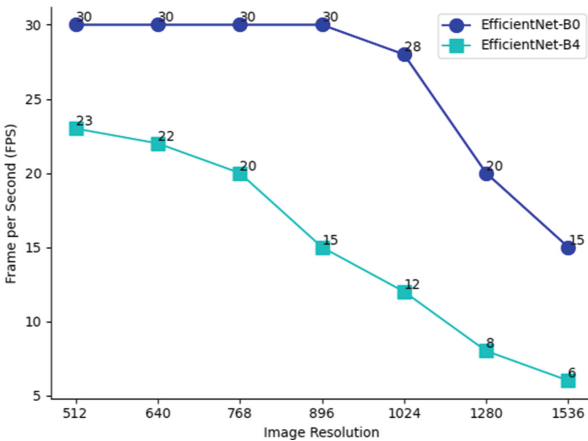

Fig. 4. FPS scores of our models at various scales using different backbones. For each image resolution using specific network backbone, the FPS score is obtained by averaging results of 1000 times.

Parameter Analysis of φ. As shown in Fig. 5, we explore the effects of different weight parameter values of φ on our model performance and compare the results with the EfficientFace (illustrated in dotted line). In this experiment, EfficientNet-B1 is used as the backbone network, and the batch size of model is set to 8. It can be observed that performance improvements to varying extents are reported on Hard subset with different φ values. This demonstrates the substantial advantages of the super-resolution module particularly in the difficult cases including low-resolution face detection. Besides, the highest AP scores of 92.5% (Easy), 91.1% (Medium) and 86.7% (Hard) are reported when φ is set to 0.1, which is consistently superior to EfficientFace achieving 92.4%, 90.9% and 85.3% on the three subsets. Thus, φ is set to the optimal 0.1 in our experiments.

Comparison of EfficientSRFace with State-of-the-Art Detectors. In this part, the proposed EfficientSRFace is compared with state-of-the-art detectors in terms of both accuracy and efficiency on WIDER Face validation set. As shown in Table 2, the competing models involved in our comparative studies include both heavy detectors such as DSFD and lightweight models like YOLOv5 variants and EXTD. In comparison to the heavy detectors, our EfficientSRFace-L using EfficientNet-B4 as the backbone achieves competitive performance with significantly reduced parameters and computational costs. In particular, EfficientSRFace-L reports respective 95.0%, 93.9% and 89.9% AP scores on Easy, Medium, and Hard subsets, which is on par with DSFD achieving 96.6%, 95.7%
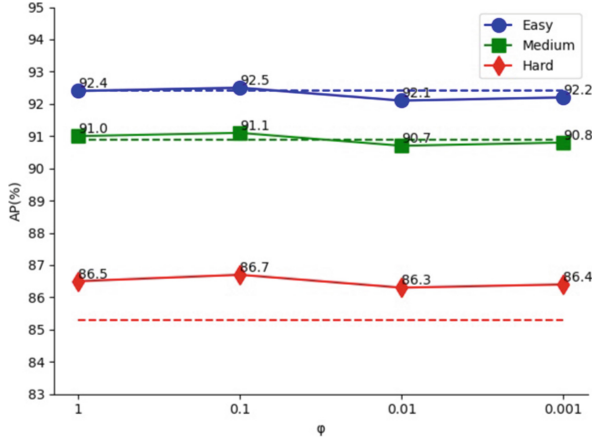
Fig. 5. Influence of different weight parameter values φ on model detection performance, where the dotted line represents that no super resolution module is embedded.

and 90.4% accuracies. However, our model enjoys approximately 6× reduced parameters and costs 10× decreased MACs. Particularly, when EfficientNet-B0 is used as the backbone in our EfficientSRFace detector, EfficientSRFace-S achieves preferable efficiency which is competitive with lightweight YOLOv5n, while outperforming the latter by 5% on Hard set. Benefiting from efficient architecture design of EfficientFace, our model enjoys different variants ranging from extremely efficient model superior to the other lightweight competitors and the relatively larger network comparable to some heavyweight models, and demonstrates the advantages in terms of the compromise between efficiency and accuracy.

Comprehensive Evaluations on the Four Benchmarks. In this section, we will comprehensively compare EfficientSRFace and other advanced detectors in the four public datasets. Figure 7 shows precision-recall (PR) curves obtained by different models on validation set of WIDER Face dataset. Although EfficientSR-Face is still inferior to some advanced heavy detectors, it still achieves competitive performance with promising model efficiency. In addition to WIDER Face dataset, we also evaluate our EfficientSRFace on the other three datasets and carry out more comparative studies. As shown in Table 3, our EfficientSRFace-L achieves respective 99.94% and 98.84% AP scores on AFW and PASCAL Face datasets. In particular, EfficientSRFace consistently beats the other competitors including even heavyweight models like MogFace and RefineFace [35] on AFW. In addition to AP scores, we also present PR curves of different detectors on AFW, PASCAL Face and FDDB datasets as shown in Fig. 6. On FDDB dataset, more specifically, when the number of false positives is 1000, our model reports the true positive rate up to 96.7%, surpassing most face detectors.

Table 2. Comparison of EfficientSRFace and other advanced face detectors. EfficientSRFace-S and EfficientSRFace-L denote our two models with EfficientNet-B0 and EfficientNet-B4 respectively used as the backbones.

Model	Easy	Medium	Hard	Params (M)	MACs (G)
MogFace-E [4]	**97.7%**	**96.9%**	92.01%	85.67	349.14
MogFace [4]	97.0%	96.3%	**93.0%**	85.26	807.92
AInnoFace [5]	97.0%	96.1%	91.8%	88.01	312.45
SRNFace-1400 [6]	96.5%	95.2%	89.6%	53.38	251.94
SRNFace-2100 [6]	96.5%	95.3%	90.2%	53.38	251.94
DSFD [7]	96.6%	95.7%	90.4%	120	345.16
yolov5n-0.5 [8]	90.76%	88.12%	73.82%	0.45	0.73
yolov5n [8]	93.61%	91.52%	80.53%	1.72	2.75
yolov5s [8]	94.33%	92.61%	83.15%	7.06	7.62
yolov5m [8]	95.30%	93.76%	85.28%	21.04	24.09
yolov5l [8]	95.78%	94.30%	86.13%	46.60	55.31
EXTD-32 [9]	89.6%	88.5%	82.5%	0.063	5.29
EXTD-64 [9]	92.1%	91.1%	85.6%	0.16	13.26
EfficientSRFace-S (Ours)	92.5%	90.7%	85.8%	3.90	5.17
EfficientSRFace-L (Ours)	95.0%	93.9%	89.9%	18.84	35.83

Table 3. Comparison of EfficientSRFace and other detectors on the AFW and PAS-CAL Face datasets (AP).

Models	AFW	PASCA Face
RefineFace [35]	99.90%	**99.45%**
FA-RPN [36]	99.53%	99.42%
MogFace [4]	99.85%	99.32%
SFDet [37]	99.85%	98.20%
SRN [6]	99.87%	99.09%
FaceBoxes [38]	98.91%	96.30%
HyperFace-ResNet [39]	99.40%	96.20%
STN [40]	98.35%	94.10%
Ours	**99.94%**	98.84%

(a) AFW   (b) Pascal Face   (c) FDDB
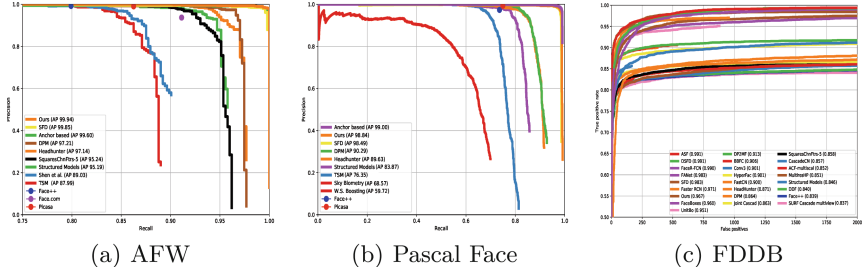
Fig. 6. Evaluation on common face detection datasets.

(a) Easy   (b) Medium   (c) Hard
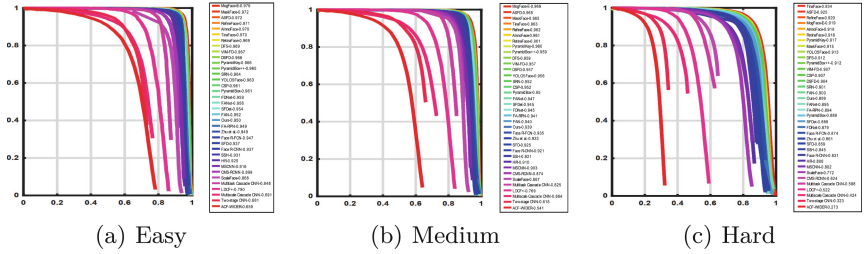
Fig. 7. PR curves of different methods on validation set of WIDER Face dataset.

5   Conclusions

In this paper, we develop an efficient network architecture based on Efficient-Face termed EfficientSRFace to better handle the low-resolution face detection. To this end, we embed a feature-level super-resolution reconstruction module to feature pyramid network for enhancing the feature representation capability of the model. This module plays an auxiliary role in the training process and can be removed during the inference without increasing the inference time. More importantly, this supplementary role incurs a small amount of additional parameters and limited growth in computational overhead without damaging model efficiency. Extensive experiments on public benchmarking datasets demonstrate that the embedded image super-resolution module can significantly improve the detection accuracy at a small cost.

References

1. Vesdapunt, N., Wang, B.: Crface: confidence ranker for model-agnostic face detection refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1674–1684 (2021)
2. Tang, X., Du, D.K., He, Z., Liu, J.: Pyramidbox: a context-assisted single shot face detector. In: Proceedings of the European Conference on Computer Vision, pp. 797–813 (2018)

3. Ming, X., Wei, F., Zhang, T., Chen, D., Wen, F.: Group sampling for scale invariant face detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3446–3456 (2019)
4. Liu, Y., Wang, F., Deng, J., Zhou, Z., Sun, B., Li, H.: Mogface: towards a deeper appreciation on face detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4093–4102 (2022)
5. Zhang, F., Fan, X., Ai, G., Song, J., Qin, Y., Wu, J.: Accurate face detection for high performance. arXiv preprint arXiv:1905.01585, pp. 1–9 (2019)
6. Chi, C., Zhang, S., Xing, J., Lei, Z., Li, S.Z., Zou, X.: Selective refinement network for high performance face detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 8231–8238 (2019)
7. Li, J., et al.: DSFD: dual shot face detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5060–5069 (2019)
8. Qi, D., Tan, W., Yao, Q., Liu, J.: Yolo5face: why reinventing a face detector. In: Proceedings of the European Conference on Computer Vision Workshops, pp. 228–244 (2022)
9. Yoo, Y., Han, D., Yun, S.: EXTD: extremely tiny face detector via iterative filter reuse. arXiv preprint arXiv:1906.06579, pp. 1–11 (2019)
10. He, Y., Xu, D., Wu, L., Jian, M., Xiang, S., Pan, C.: LFFD: a light and fast face detector for edge devices. arXiv preprint arXiv:1904.10633, pp. 1–10 (2019)
11. Wang, G., Li, J., Wu, Z., Xu, J., Shen, J., Yang, W.: EfficientFace: An Efficient Deep Network with Feature Enhancement for Accurate Face Detection. Multimedia Systems, pp. 1–15 (2023)
12. Liu, W., et al.: SSD: single shot multibox detector. In: Proceedings of the European Conference on Computer Vision, pp. 21–37 (2016)
13. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
14. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q.: Centernet: keypoint triplets for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6569–6578 (2019)
15. Tan, M., Pang, R., Le, Q.V.: Efficientdet: scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781–10790 (2020)
16. Zhang, C., Xu, X., Tu, D.: Face detection using improved faster RCNN. arXiv preprint arXiv:1802.02142, pp. 1–9 (2018)
17. Zhang, S., et al.: Improved selective refinement network for face detection. arXiv preprint arXiv:1901.06651, pp. 1–8 (2019)
18. Zhang, Y., Xu, X., Liu, X.: Robust and high performance face detector. arXiv preprint arXiv:1901.02350, pp. 1–9 (2019)
19. Zhu, Y., Cai, H., Zhang, S., Wang, C., Xiong, Y.: Tinaface: strong but simple baseline for face detection. arXiv preprint arXiv:2011.13183, pp. 1–9 (2020)
20. Najibi, M., Samangouei, P., Chellappa, R., Davis, L.S.: SSH: Single stage headless face detector. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4875–4884 (2017)
21. Liu, Y., Tang, X., Han, J., Liu, J., Rui, D., Wu, X.: Hambox: delving into mining high-quality anchors on face detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13043–13051 (2020)
22. Li, J., et al.: ASFD: Automatic and scalable face detector. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 2139–2147 (2021)

23. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)
24. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Proceedings of the European Conference on Computer Vision, pp. 391–407 (2016)
25. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
26. Wang, X., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision, pp. 36–79 (2018)
27. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision, pp. 286–301 (2018)
28. Kong, X., Zhao, H., Qiao, Y., Dong, C.: Classsr: a general framework to accelerate super-resolution networks by data characteristic. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12016–12025 (2021)
29. Cong, W., et al.: High-resolution image harmonization via collaborative dual transformations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18470–18479 (2022)
30. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2980–2988 (2017)
31. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2879–2886 (2012)
32. Yan, J., Zhang, X., Lei, Z., Li, S.Z.: Face detection by structural models. Image Vis. Comput. **32**(10), 790–799 (2014)
33. Jain, V., Learned-Miller, E.: FDDB: A Benchmark for Face Detection in Unconstrained Settings. Technical Report, UMass Amherst Technical Report (2010)
34. 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)
35. Zhang, S., Chi, C., Lei, Z., Li, S.Z.: Refineface: refinement neural network for high performance face detection. IEEE Trans. Pattern Anal. Mach. Intell. **43**(11), 4008–4020 (2020)
36. 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)
37. Zhang, S., Wen, L., Shi, H., Lei, Z., Lyu, S., Li, S.Z.: Single-shot scale-aware network for real-time face detection. Int. J. Comput. Vision **127**(6), 537–559 (2019)
38. Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., Li, S.Z.: Faceboxes: a CPU real-time face detector with high accuracy. In: 2017 IEEE International Joint Conference on Biometrics, pp. 1–9 (2017)
39. Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans. Pattern Anal. Mach. Intell. **41**(1), 121–135 (2017)
40. Chen, D., Hua, G., Wen, F., Sun, J.: Supervised transformer network for efficient face detection. In: Proceedings of the European Conference on Computer Vision, pp. 122–138 (2016)