Hierarchical Residual Embedding Network for Real-Time De-focus Blur Detection

Chunlei Zhang¹, Jing Nie²* and Xiao Zheng³

¹Electronic Engineering College, Naval University of Engineering, Wuhan 430033, China.
²Teaching and Research Support Center, Naval University of Engineering, Wuhan 430033, China.
³School of Computer Science, National University of Defense Technology, Changsha 410073, China.

*Corresponding author email: cindyray222@163.com

Abstract. Defocus blur detection, as an important pre-processing step of image processing, has attracted more and more attention. Albeit great success has been made, there are still several challenges for accurate defocus blur detection, such as the interference of background clutter, sensitivity to scales, missing boundary details and large computational burden. For handling these issues, we present a deep neural network which hierarchically embeds residual learning blocks for defocus blur detection. Based on the feature pyramid structure, we extract deep features with varying scales via utilizing a backbone fully convolutional network and generate a coarse score map by using the last layer of feature maps. Then we design a hierarchical residual embedding module to fuse different levels of features in a layer-wise manner. By embedding different layer-wise features in the top-down pathway, coarse-level semantic information from the deep layers can be seamlessly propagated to shallow layers, while fine details in the shallow layers can be used to refine the boundary between out-of-focus and in-focus regions. For each layer, a side output is generated by using a residual learning block. For capturing multi-scale information, the multiple side outputs of different layers are fed into a designed fusion block for yielding the final blur map result. Experimental results on two commonly used datasets show that our proposed network can more accurately locate the defocus blur regions with sharpened details being well preserved when compared to other previous state-of-the-arts. In addition, our approach is fast as well and can run at a speed of more than 25 FPS when processing an image with size $427 \times 640$.

1. Introduction

Defocus blur detection has obtained much attention due to its wide potential applications such as image de-blurring [1], defocus magnification [2], image refocusing [3, 4], image restoration [5], image quality assessment [6], and object detection [7], just list a few.

In the past years, a large number of methods have been proposed defocus blur detection, which can be generally classified into two categories based on the used features, i.e., methods based on traditional features [1, 8–17] and methods based on deep features.

In the past years, deep convolutional neural networks (DCNNs) have been widely used in various computer vision tasks since they can obtain good results, such as classification [18, 19], object detection [20, 21], object tracking [22–24], scene semantic segmentation [25–27], de-noising [28, 29] and super-resolution [30, 31]. Also, defocus blur region detection has gained great success by using
CNNs[17, 32–35]. In addition, deep learning based blur detection methods can usually obtain better results than traditional hand-crafted features based ones. However, the computational burden of previous deep learning based methods are often huge since they usually capture scale information by aggregating multi-scale deep features [35] or using multiple network branches to process images at different scales [33]. In addition, the detecting accuracy can be still further improved.

In this work, we present a brand new efficient and effective pixel-wise FCN for defocus blur detection via hierarchical residual embedding (HRENet). The advantages of the residual learning and multi-scale information encoded in multiple layers of feature maps are efficiently leveraged to boost the detection performance. Overall, the technical contributions of this work can be summarized as follows:

- A new pixel-wise fully convolutional network is introduced for defocus blur detection.
- A hierarchical residual embedding module (HREM) is designed to fuse features of multiple layers in a layer-wise manner.
- Both qualitative and quantitative experimental results validate the superiority of our proposed network over other competitors.

2. Proposed HRENet

Figure 1 gives a brief schematic illustration of our proposed HRENet. Our HRENet is built based on the feature pyramid structure, we design a hierarchical residual embedding module to make the coarse-level semantic information well fused with the fine-level features from the top-down pathway in a layer-wise manner. In detail, the ResNet structure [47] is selected as our backbone feature extraction network and the pre-trained ResNeXt model is used to initialize our network, which produces five basic feature extraction layers. In this paper, we use \( F_i \in W_i \times H_i \times C_i (i = 1, 2, 3, 4, 5) \) to denote the feature maps of the \( i \)-th layer with size \( W_i \times H_i \) and \( C_i \) channels.

Figure 1. The pipeline of our HRENet.

2.1. Initial Defocus Blur Map Generation

As discussed in previous sections, a major issue of traditional hand-crafted feature based methods is the lack of semantics information, which hinder the detection effect for low-contrast but in-focus regions. Since the feature maps extracted from deep layers of deep CNNs encode sufficient semantic information, we first use the feature maps extracted by the last layer to generate a coarse score map, then this coarse score map will be refined step by step by a hierarchical residual embedding module which fuses features from different layers in a top-down manner.
2.2. Hierarchical Residual Embedding Module
Since residual learning can achieve better performance when compared to common plain network in a variety of tasks, such as image classification [47], super-resolution [48], visual tracking [49] and segmentation [50], we design a Hierarchical Residual Embedding Module (HREM) which contains a series of residual leaning blocks (RLBs) to integrate feature maps from different layers in a ton-down manner. By this way, the coarse score map can be refined step by step. Meanwhile, the high-order semantic information can be also well preserved.

First of all, we use the last layer of feature maps ($F_5$) to generate an initial coarse blur map, denoted as $O_0$. Thanks to the high-level semantic information, the background clutters can be effectively suppressed and the smooth in-focus regions can be separated out well in $O_0$. However, the details are lost due to the low resolution of $F_5$. Therefore, $O_0$ will be refined by a series of RLBs.

2.3. Residual Leaning Block
In this work, the proposed HREM consists of multiple RLBs. For each RLB, its input contains the output of previous RLB and the feature maps of current layer, its output is the sum of the output of previous RLB and the learned residual. Figure 1 briefly demonstrates the architecture of the proposed RLB. For each RLB at the $l$-th layer ($l = 4, 3, 2, 1$), the residual map can be calculated as

$$R_{s-l} = \Psi(Cat(O_{s-l}, F_l))$$  \hspace{1cm} (1)

where $Cat$ and $\Psi$ represent the concatenation operation and a mapping function, respectively. Then the output of current RLB at the $l$-th layer can be obtained by adding $R_l$ with $O_{l-1}$ as:

$$O_{s-l} = R_{s-l} \odot O_{s-l}$$  \hspace{1cm} (2)

For each RLB, the supervision signal [51] is imposed to each RLB for computing the loss between the ground truth and each intermediate prediction during the training process.

2.4. Defocus Maps Fusing
In order to capture the scale information, we obtain the final output defocus blur map $O$ by fusing multiple outputs, which can be formulated as:

$$O = \sigma(W_O \ast Cat(O_0, O_1, \cdots O_4) + b_o)$$  \hspace{1cm} (3)

where $O_i$ denotes the intermediate prediction of the $i$-th RLB; $W_O$ and $b_O$ are the weight and bias need to be learned.

2.5. Model Training and Testing
Our network uses the ResNet architecture [47] as backbone and we implement it by using the Pytorch framework. The well trained ResNeXt network on ImageNet [52] is used to initialize the parameters of feature extraction network.

3. Experiments
3.1. Experiment Settings
Similar to previous works, we use Shi et al.’s dataset [1] and DUT dataset [33] as two datasets for evaluation. Precision-recall (PR) curves, F-measure curves, F-measure scores (F ) and mean absolute error (MAE) scores are used for four metrics. 11 state-of-the-art algorithms are used for comparison, i.e., DHDE[17], BTBNet [33], DeFusionNET [35], ASVB [53], SVD [9], JNB [40], DBDF [1], SS [15], LBP [41], KSFV [54] and HiFST [43]. For all of these methods, we use the authors’ original implementations with recommended parameters.
3.2. Quantitative Comparison

Table 1 presents the compared results of MAE and F-measure scores. As can be seen, our method consistently performs better. In Figure 2, the PR curves and F-measure curves also demonstrate the efficacy of the proposed network.

| Datasets       | Metric | ASV | SV  | JNB | DBD | SS  | LB  | KSF | DHD | HiFS | BTB-Net | DeFusion-Net | HRE-Net |
|----------------|--------|-----|-----|-----|-----|-----|-----|-----|-----|------|---------|-------------|---------|
| Shi et al.     | ran    | 0.73| 0.80| 0.79| 0.84| 0.78| 0.86| 0.73| 0.85| 0.85  | 0.887   | 0.917       | 0.931   |
|                | mean   | 0.63| 0.30| 0.35| 0.32| 0.29| 0.18| 0.38| 0.39| 0.23  | 0.158   | 0.116       | 0.106   |
| DUT            | ran    | 0.74| 0.81| 0.74| 0.80| 0.78| 0.87| 0.75| 0.82| 0.86  | 0.901   | 0.922       | 0.946   |
|                | mean   | 0.65| 0.30| 0.42| 0.36| 0.29| 0.17| 0.39| 0.40| 0.30  | 0.169   | 0.115       | 0.089   |

3.3. Qualitative Comparison

We show the visual comparison of our method and other ones in Figure 3. The results also show that our proposed network can obtain better results than other compared methods in terms of both semantic and fine detail information.

Figure 2. PR curves and F-measure curves of varying methods on Shi et al.’s dataset ((a) and (b)), and DUT dataset ((c) and (d)).

Figure 3. Intuitive results of different methods.
4. Conclusions
We introduce a deep neural network which hierarchically embeds residual learning blocks for defocus blur detection. Based on the feature pyramid structure, we first extract multi-scale deep features by utilizing a backbone fully convolutional network and generate a coarse score map by using the last layer of feature maps. Then we design a hierarchical residual embedding module to make the coarse-level semantic information well fused with the fine-level features from the top-down pathway in a layer-wise manner. Experimental results on two commonly used datasets show that our proposed network can more accurately locate the defocus blur regions with sharpened details being well preserved when compared to other previous state-of-the-arts. In addition, our approach is fast as well and can run at a speed of more than 25 FPS when processing an image with size $427 \times 640$.

References
[1] Jianping Shi, Li Xu, and Jiaya Jia. Discriminative blur detection features. In CVPR, pages 2965–2972, 2014.
[2] Soonmin Bae and Fredo’ Durand. Defocus magnification. Computer Graphics Forum, 26(3):571–579, 2007.
[3] Wei Zhang and Wai-Kuen Cham. Single image focus editing. In ICCVW, pages 1947–1954. IEEE, 2009.
[4] Wei Zhang and Wai-Kuen Cham. Single-image refocusing and defocusing. TIP, 21(2):873–882, 2012.
[5] Xinxin Zhang, Ronggang Wang, Xiubao Jiang, Wenmin Wang, and Gao Wen. Spatially variant defocus blur map estimation and deburring from a single image. JVCIR, 35(C):257–264, 2016.
[6] Xin Wang, BaoFeng Tian, Chao Liang, and Dongcheng Shi. Blind image quality assessment for measuring image blur. In CISP, pages 467–470. IEEE, 2008.
[7] Peng Jiang, Haibin Ling, Jingyi Yu, and Jingliang Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. In ICCV, pages 1976–1983, 2013.
[8] Renting Liu, Zhaorong Li, and Jiaya Jia. Image partial blur detection and classification. In CVPR, pages 1–8, 2008.
[9] Bolan Su, Shijian Lu, and Chew Lim Tan. Blurred image region detection and classification. In ACMM, pages 1397–1400, 2011.
[10] Shaojie Zhuo and Terence Sim. Defocus map estimation from a single image. PR, 44(9):1852–1858, 2011.
[11] Cuong T Vu, Thien D Phan, and Damon M Chandler. s3: A spectral and spatial measure of local perceived sharpness in natural images. IEEE TIP 21(3):934, 2012.
[12] Yi Zhang and Keigo Hirakawa. Blur processing using double discrete wavelet transform. In CVPR, pages 1091–1098, 2013.
[13] X. Zhu, S Cohen, S Schiller, and P Milanfar. Estimating spatially varying defocus blur from a single image. IEEE TIP, 22(12):4879–4891, 2013.
[14] Y. Pang, H. Zhu, X. Li, and X. Li. Classifying discriminative features for blur detection. IEEE TCyb, 46(10):2220–2227, 2015.
[15] Chang Tang, Jin Wu, Yonghong Hou, Pichao Wang, and Wanqing Li. A spectral and spatial approach of coarse-to-fine blurred image region detection. IEEE SPL, 23(11):1652–1656, 2016.
[16] E Saad and K Hirakawa. Defocus blur-invariant scale-space feature extractions. TIP, 25(7):3141–3156, 2016.
[17] Jinsun Park, Yu Wing Tai, Donghyeon Cho, and In So Kweon. A unified approach of multi-scale deep and hand-crafted features for defocus estimation. In CVPR, pages 2760–2769, 2017.
[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1097–1105, 2012.
[19] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In ICRL, 2015.
[20] Kai Kang, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. Object detection from video tubelets with convolutional neural networks. In CVPR, pages 817–825, 2016.
[21] Tsung Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. IEEE TPAMI, PP(99):2999–3007, 2017.
[22] Peixia Li, Dong Wang, Lijun Wang, and Huchuan Lu. Deep visual tracking: Review and experimental comparison. PR, 76:323–338, 2018.
[23] Chong Sun, Huchuan Lu, and Ming-Hsuan Yang. Learning spatial-aware regressions for visual tracking. In CVPR, pages 8962–8970, 2018.
[24] Yuankai Qi, Shengping Zhang, Lei Qin, Qingming Huang, Hongxun Yao, Jongwoo Lim, and Ming-Hsuan Yang. Hedging deep features for visual tracking. IEEE TPAMI, 2018.
[25] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In ICCV, pages 1520–1528, 2015.
[26] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In CVPR, pages 3431–3440, 2015.
[27] Hengshuang Zhao, Jianping Shi, Xiaoojian Q, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, pages 3118–3126, 2018.
[28] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. IEEE TIP, 26(9):4509–4522, 2017.
[29] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE TIP, 26(7):3142–3155, 2017.
[30] Chao Dong, Chen Change Loy, Kaiming He, and Xiaou Tang. Image super-resolution using deep convolutional networks. IEEE TPAMI, 38(2):295–307, 2016.
[31] Assaf Shocher, Nadav Cohen, and Michal Irani. “Zero-shot” super-resolution using deep internal learning. In CVPR, pages 3118–3126, 2018.
[32] Ruomei Yan and Ling Shao. Blind image blur estimation via deep learning. IEEE TIP, 25(4):1910–1921, 2016.
[33] Wenda Zhao, Fan Zhao, Dong Wang, and Huchuan Lu. Defocus blur detection via multi-stream bottom-top-bottom fully convolutional network. In CVPR, pages 3080–3088, 2018.
[34] Wenda Zhao, Bowen Zheng, Qihua Lin, and Huchuan Lu. Enhancing diversity of defocus blur detectors via cross-ensemble network. In CVPR, pages 8905–8913, 2019.
[35] Chang Tang, Xinzhong Zhu, Xinwang Liu, Lizhe Wang, and Zomaya Albert. Defusionnet: Defocus blur detection via recurrently fusing and refining multi-scale deep features. In CVPR, pages 2700–2709, 2019.
[36] James H Elder and Steven W Zucker. Local scale control for edge detection and blur estimation. IEEE TPAMI, 20(7):699–716, 1998.
[37] Yu-Wing Tai and Michael S Brown. Single image defocus map estimation using local contrast prior. In ICIP, pages 1797–1800. IEEE, 2009.
[38] Florent Couzinie-Devy, Jian Sun, Karteek Alahari, and Jean Ponce. Learning to estimate and remove non-uniform image blur. In CVPR, pages 1075–1082, 2013.
[39] Chang Tang, Chunping Hou, and Zhanjie Song. Defocus map estimation from a single image via spectrum contrast. OL, 38(10):1706–1708, 2013.
[40] Jianping Shi, Li Xu, and Jiaya Jia. Just noticeable defocus blur detection and estimation. In CVPR, pages 657–665, 2015.
[41] Xin Yi and Mark Eramian. Lbp-based segmentation of defocus blur. IEEE TIP, 25(4):1626–1638, 2016.
[42] Chang Tang, Jin Wu, Yonghong Hou, Pichao Wang, and Wanqing Li. A spectral and spatial approach of coarse-to-fine blurred image region detection. IEEE SPL, 23(11):1652–1656, 2016.
[43] Ali Alirezaz Golestaneh and Lina J Karam. Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes. In CVPR, pages 5800–5809, 2017.
[44] Guodong Xu, Yuhui Quan, and Hui Ji. Estimating defocus blur via rank of local patches. In CVPR, pages 22–29, 2017.
[45] Kuldeep Purohit, Anshul B Shah, and AN Rajagopalan. Learning based single image blur detection and segmentation. In ICIP, pages 2202–2206. IEEE, 2018.
[46] Shanghang Zhang, Xiaohui Shen, Zhe Lin, Radomýr Mech, Joao P Costeira, and Jose´ MF Moura. Learning to understand image blur. In CVPR, pages 6586–6595, 2018.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Sun Jian. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.

[48] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In CVPR, pages 2472–2481, 2018.

[49] Yibing Song, Ma Chao, Lijun Gong, Jiawei Zhang, and Ming Hsuan Yang. Crest: Convolutional residual learning for visual tracking. In ICCV, pages 2555–2564, 2017.

[50] Seong-Jin Park, Ki-Sang Hong, and Seungyong Lee. Rdffnet: Rgb-d multi-level residual feature fusion for indoor semantic segmentation. In ICCV, pages 4980–4989, 2017.

[51] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. IJCV, 125(1-3):3–18, 2015.

[52] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In CVPR, pages 1492–1500, 2016.

[53] Ayan Chakrabarti, Todd Zickler, and William T. Freeman. Analyzing spatially-varying blur. In CVPR, pages 2512–2519, 2010.

[54] Yanwei Pang, Hailong Zhu, Xinyu Li, and Xuelong Li. Classifying discriminative features for blur detection. IEEE TCyb, 46(10):2220–2227, 2016.