https://doi.org/10.2352/EI.2023.35.7.IMAGE-273

Depth Assisted Portrait Video Background Blurring

Yezhi Shen[1], *Weichen Xu*[1], *Qian Lin*[2], *Jan P Allebach*[1], *and Fengqing Zhu*[1]

[1] *Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN*
[2] *HP Inc, Palo Alto, CA*

Abstract

Video conferencing usage dramatically increased during the pandemic period and is expected to remain high in hybrid work. One of the key aspects of video conferencing experience is background blur and background replacement. In order to achieve a high-quality background blurring effect, an accurate portrait segmentation and high-quality blurring effect are necessary. Since depth sensors are becoming ubiquitous on mobile devices, software and hardware manufacturers are working together to utilize depth maps to improve the background blurring quality. Depth map has been used as an auxiliary input to categorize foreground and background human during post-processing in existing works. We further exploit the information in depth map by using depth map during inference and post-processing. In this paper, we propose an improved CNN model based on the well-known PortraitNet method, a depth-assisted mask refinement that can correct error predictions using depth information, and a bokeh effect rendering module that can produce a photo-realistic bokeh effect for the background.

Introduction

Online video conferencing has been playing an increasingly important role since the pandemic, which is widely used in teaching, working, and entertainment. Users share their portrait videos, which contain the upper body parts and backgrounds of their rooms across the meeting software. An important feature of online conferencing software is background blurring, which is used to protect users' privacy by blurring the contents in the users' portrait video background.

Background blurring involves portrait segmentation, which masks the human part in the portrait video and separates the human from the background. Most current methods for video portrait segmentation utilize Convolutional Neural Networks (CNN) to generate the portrait segmentation masks. The blur effect is applied to the extracted background and the blurred background is combined with the human foreground to produce the new video with the background blurring effect.

However, several problems exist in the current methods of background blurring. Inaccurate portrait segmentation can lead to a rough visual effect on the boundary after background blurring. Models that utilize temporal information to reduce flickering and variation suffer from error propagation and reveal objects in the background that should have been blurred. The current practice of using a unified kernel on the entire background does not fit the human perception with a depth of field, thus looking wired when objects of different distances get equally blurred.

To address these challenges, we propose a new method that leverages depth maps to generate a bokeh effect, which simulates the depth of field in photography. Our method consists of a lightweight and accurate portrait segmentation model, a depth-assisted mask refinement module, and a bokeh effect render module. By using depth maps, our approach achieves highly accurate portrait segmentation and produces visually pleasing background bokeh effects. With the increasing availability of depth sensors on mobile devices, our proposed method can significantly enhance the visual quality of online video conferencing.

Related Work

Portrait Segmentation. Portrait segmentation has attracted great interest in recent years, which crops out a person from an image where a person's upper body is shown. Researchers have been developing methods to generate more accurate pixel-wise masks. With the development of Convolutional Neural Networks (CNN) over the years, lightweight but accurate models, such as the PortraitNet [1] and SINet [2], have been proposed. PortraitNet [1] presents a UNet [3] like encoder-decoder network that adopts MobileNetV2 [4] as its feature extractor. The encoder and decoder have 5 layers and each layer downsamples the feature map by $2\times$ to learn the spatial relationship. SINet [2] presents an extremely lightweight model with 0.087 million parameters, yet achieves the third-best segmentation accuracy on portrait image dataset EG1800 [5] as reported in their paper. The design of SINet [2] contains a sequential encoder-decoder, which significantly reduces the CNN channel numbers used in the decoder. The accuracy of SINet is guaranteed with the spatial squeeze modules and information blocking decoder, which together separate the high-resolution feature and merge the necessary ones with the lower-resolution feature maps during decoding.

RGB-D Segmentation. RGB-D segmentation methods perform portrait segmentation with both the RGB image and depth map as their inputs. Taking depth map into account has the benefit of being able to identify the boundary of the segmentation object easier and providing additional 3D spatial information to the CNN models. Kumar et al. [6] proposed a method that first performs foreground extraction on RGB frame and depth map separately. Both results are sent into a voting engine to generate the segmentation mask. Recently, Singh et al. [7] proposed a method that performs portrait segmentation using dual-focus images as input. Their method first generates a trimap by calculating the error between the dual-focus images. The tri-map is used as guidance to perform portrait segmentation on the near-focus RGB image using a CNN model. The mask is finally refined using Graph Cut before being used for further post-processing.

Bokeh Effect Rendering. Bokeh is a photography term that describes out-of-focus objects getting blurred according to their distance to the focus point. Bokeh effect rendering refers to blur-

ring the out-of-focus regions of an image in a natural way during post-processing, where objects closer to the focal plane get less blurred and further away objects are more blurred. Wang et al. [8] purposed a UNet-like encoder-decoder network with a Swin-Transformer encoder to perform depth estimation from a monocular image. To create a photo-realistic bokeh effect, the depth map is divided into several different clusters based on the depth value. Layer masks are generated from the separated clusters and used to divide the RGB image into different sections according to their depth. The blurring kernel size for each section is calculated using the difference average distance of that section to the focus point. A smaller depth difference results in a smaller blur kernel, and a greater difference results in a larger blur kernel. Ignatov et al. [9] proposed a CNN model which is trained layer by layer to perform bokeh rendering directly. The model renders images with bokeh effect recursively from lower resolutions to higher resolutions.

Mask Refinement. Since portrait video segmentation is typically performed on mobile devices and require real-time processing, it is a common practice to perform inference at a lower resolution (e.g., 224×224 or 384×192) to trade-off speed. The lower-resolution predicted masks need to be up-sampled to the original resolution and filtered before they can be used for background editing. Wilms et al. [10] purposed to use superpixel and attention mask produced by a CNN model to perform mask refinement. The superpixel algorithm divides the image into several segments and the attention mask is responsible for guiding which superpixel to keep. The output foreground is a combination of all kept superpixels. GrabCut [11] is a probabilistic model that uses Gaussian Mixture Models to identify foreground and background. When used as a refinement for segmentation, the RGB image under the coarse segmentation mask area is set as the possible foreground. GrabCut algorithm models the masked foreground and inversely masked background area each with 2-D Gaussian Distributions and iteratively updates the refined segmentation result.

Methods

Our proposed depth-assisted portrait video background blurring method consists of three parts: a lightweight CNN-based portrait segmentation model, a depth-assisted mask refinement module, and a bokeh effect rendering module as shown in Figure. 1(a).

RGB-D Portrait Net with MV3 Encoder

In portrait segmentation, the foreground object typically occupies a majority of the image, thus it is important for the model to understand both global and spatial information. Our proposed model is designed based on PortraitNet [1], which shows the best performance on the portrait image dataset EG1800 [5]. PortraitNet [1] consists of a MobileNetV2 (MV2) [4] encoder and UNet-like decoder. The encoder module is used to extract features from raw RGB images. The MV2 [4] encoder has five layers, each of which has a down-sampling rate of $2\times, 4\times, 8\times, 16\times, 32\times$. With such a pyramid design, the MV2 [4] encoder is capable of understanding rich global and spatial information. The decoder block in PortraitNet is a modification of a residual block, where a 1×1 point-wise convolution is used to adjust channel size instead of concatenating the input directly with the output. All convolution layers in the decoder blocks are replaced with depth-wise separable convolution layers to improve inference speed. To better

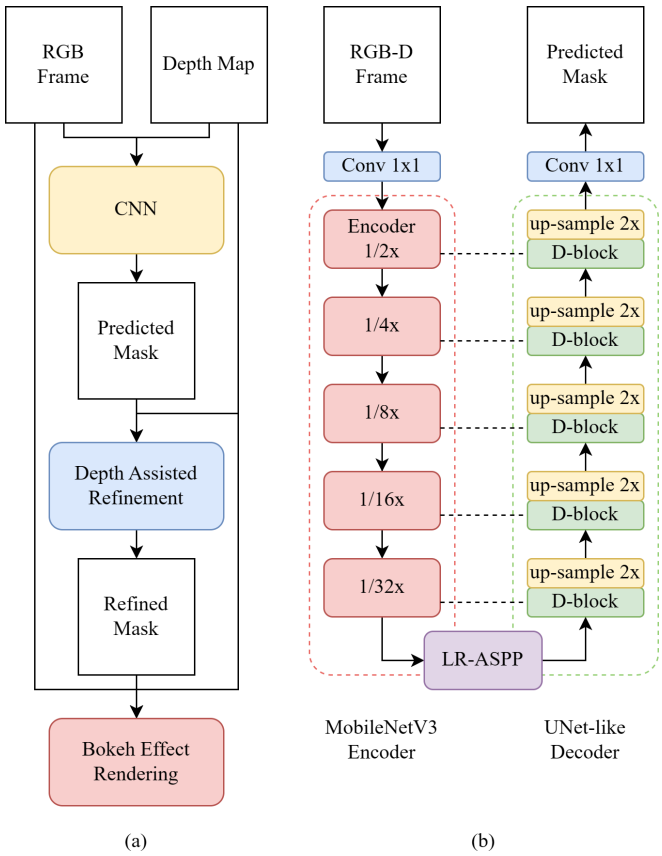


Figure 1: (a) The overall diagram of depth-assisted portrait video background blurring. (b) The RGB-D PortraitNet with MV3 Encoder. The D-block in the decoder is the same as in PortraitNet[1].

utilize the features extracted by the MV2 [4] encoder, PortraitNet [1] combines the feature maps during the decoding process with the according fusion maps. Each fusion map is concatenated with the feature map from the same encoder layer before being further upsampled and fed into the next decoder layer.

We replace the MV2 [4] encoder with MobileNetV3 (MV3) [12] encoder to further speed up inference and improve the segmentation accuracy as shown in Figure 1(b). MV2 [4] layers are made up of inverted residual and linear bottleneck, where the input and output channels are both very narrow and do not contain non-linearity inside the layers. MV3 [12] encoder, however, replaces the linear design with the squeeze and excite block. To further speed up the inference, the 3×3 convolution layers inside each MV2 [4] blocks are replaced with depth-wise separable convolution, which helps to reduce parameters and multiplications. In our work, we adopt the MV3-Large encoder, which is the largest variation of MV3 [12], as our encoder.

As discussed in PortraitNet and SINet, it is essential for portrait segmentation models to understand as much global information across the entire image as possible. To strengthen such ability, we choose the Lite Reduced Atrous Spatial Pyramid Pooling (LR-ASPP) [12] as our network's bottleneck. Atrous Spatial Pyramid Pooling (ASPP) proposed in DeepLabV3 [13] has the ability to mix spatial information at different scales, yet still maintain a small convolution kernel size. The design of LR-ASPP

carefully chose to combine the feature maps from the last layer ($\frac{1}{32}$ layer) and the $\frac{1}{8}$ layer to produce the coarse segmentation mask. A large pooling layer with a large stride is applied on the $\frac{1}{8}$ feature map to reduce parameter size. ASPP is only applied on the last layer to extract dense spatial features while maintaining low computation costs.

There are several benefits of considering a depth map as an additional input to the RGB image. For example, when the foreground and background of a frame has similar color, it could be difficult to distinguish their boundaries. However, they would have much clear boundaries in the depth map. A clear boundary can help the CNN model to achieve a high segmentation accuracy. We modify the first layer of the MV3 encoder so that it accepts a 4-channel RGB-D image as input. We sample the average depth of the masked area on the previous frame and apply the depth cutoff to only keep pixels within a threshold, illustrated by Equation (1), where D_{t-1} and M_{t-1} denotes the depth map and the predicted mask from time $t-1$, respectively. In Equation (2), the number 700 is used to accommodate typical human body thickness (300 mm) with additional tolerance (400 mm) for body movement between frames.

$$\bar{d} = mean(D_{t-1} \times M_{t-1}) \tag{1}$$

$$threshold = \bar{d} + 700 \tag{2}$$

Depth assisted mask refinement

Since the segmentation mask is predicted at low resolution for real-time inferencing, it is crucial to refine the mask before applying it to image editing. Otherwise, we see a loss of accuracy on the mask boundary due to up-sampling. Instead of applying a Gaussian filter directly to the up-sampled mask, we choose GrabCut [11] to refine the segmentation mask, so that error predictions on the mask boundary can be minimized during this process. GrabCut divides the input image into four categories: foreground, possible foreground, background, and possible background. As the common practice, methods [14], [15] either input the entire CNN predicted mask as the possible foreground, or assume the boundary of CNN predicted mask as the possible foreground and set the center of the mask to be the foreground. Since GrabCut tends to find the minimum cut for the foreground, existing methods using GrabCut cannot fix the error of missing in-hand objects. We propose a new method to determine the area to set as foreground with the help of the depth map. First, we apply the depth cutoff calculated from Equation (2) to the current frame depth map to extract the human shape and binarize it to form a mask. Then, we find the intersection between the predicted mask and the binarized depth mask and set it as the foreground for GrabCut. The remained non-overlapping pixels of the predicted mask and binarized depth mask are set as the possible foreground. The rest of the image is set as the background.

Bokeh effect rendering

Different from background blurring, where a uniform blurring effect is applied to the entire background, bokeh effect rendering attempts to create a natural blurring effect where out-of-focus objects are blurred at a different rate determined by their distance to the focus plane. For each pixel in the RGB frame, the distance between it and the focus plane is determined and the radius of the blurring kernel is calculated proportionally to the calculated distance. The rendering process combines the blurred pixels together to generate the final output.

To speed up the calculation, we split the image into different focal planes instead of blurring the background pixel by pixel. To extract the different focal planes in an image, we first apply the depth cutoff at a threshold equal to $4\times$ the average depth of the foreground. Any depth value greater than the threshold is replaced with the threshold. The depth map is then normalized to 0-1 to ensure the max extent of blur looks uniform for all backgrounds with different depths. We separate the background of the depth map into N segments evenly according to the depth value, where N is chosen to be 4 in our implementation for efficient computation.

As proposed by Wang *et al.*, the blur radius scale is directly proportional to the distance between the focal plane and the object plane, which is calculated using Equation (3). The absolute depth value difference between the focus frame D_f and the target plane D_{target} is used to determine the relative degree of blur for different target planes. The blur coefficient p in this equation decides the final value of the blur kernel size. The RGB frame of the original resolution is blurred N times using the N different blur kernels calculated.

With the extracted N focal planes from the depth map, we binarize them to keep any values greater than zero to form layer masks. The layer masks are used to decide which part of the blurred images are used for the bokeh output. The final rendering combines the N blurred RGB images into one frame by keeping only the selected parts of each blurred layer.

$$k = |D_f - D_{target}| \times p \tag{3}$$

By applying the bokeh effect rendering, we are able to achieve a similar visual result of the background bokeh, which has a more natural look than background blurring using a unified kernel.

Experiments

Datasets

We train and evaluate our proposed method and baseline methods on portrait segmentation datasets. Our selection of datasets includes four public segmentation portrait image datasets, two private segmentation portrait image datasets with more complex backgrounds, and a portrait video dataset. The four public datasets are baidu-V1 [16], baidu-V2 [16], EG1800 [5], and SuperviselyPortrait [17]. We removed the wrongly annotated images and relabelled some of the images to include in-hand objects as part of the foreground. The datasets are then divided into training, validation, and testing in an 8:1:1 ratio as shown in Table 1.

VideoMatte240k [18] is a portrait video matting dataset containing foreground only. We manually added backgrounds consist of indoor images collected from the internet. Since there were few existing portrait segmentation video datasets available, we applied a threshold to the alpha mattes provided in VideoMatte240k to generate the segmentation masks for the foreground. As a result, all values greater than 0.5 are kept as the foreground, and the rest of the pixels are set to 0 as the background. A total of 484 video clips are divided into 474:5:5 for training, validation, and testing. We randomly select 30 frames from each clip.

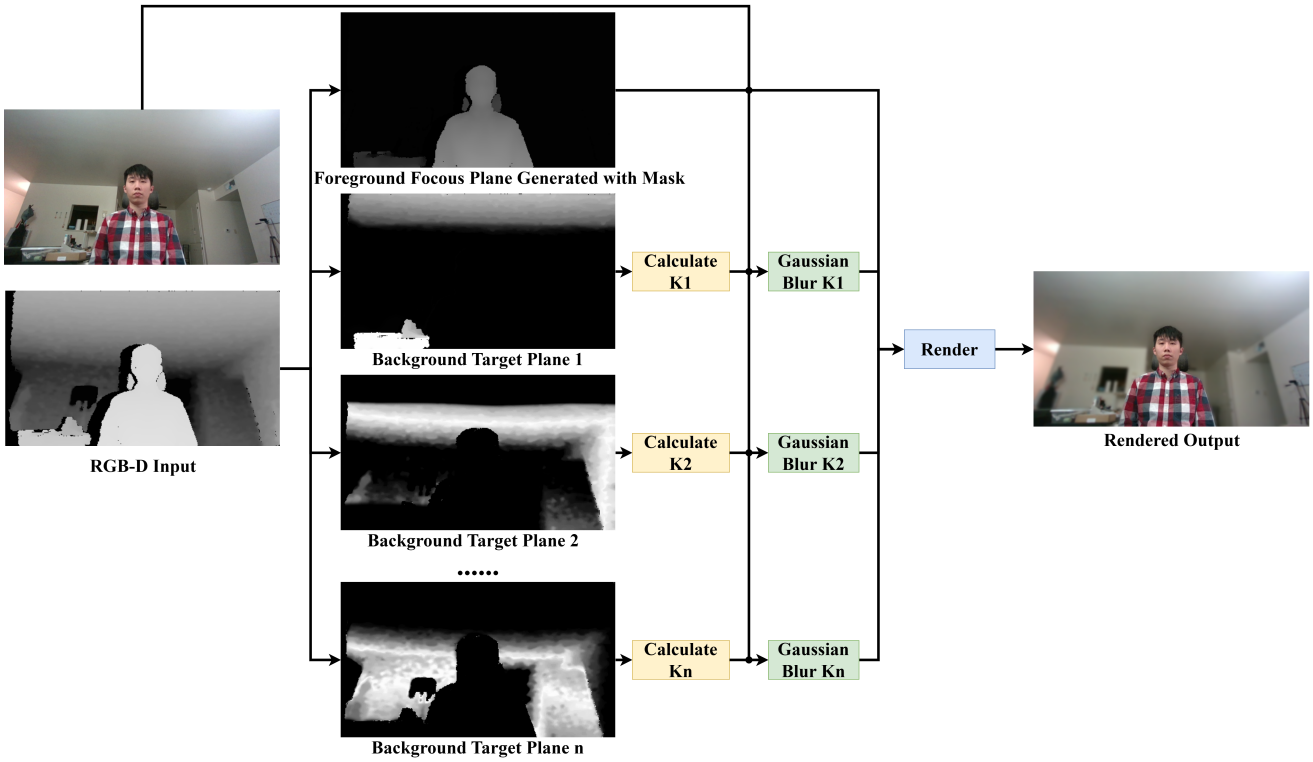


Figure 2: The overview of the bokeh effect rendering process. The blur radius K_n are calculated using Equation.3.

Table 1: Semantic Portrait Segmentation Datasets

Datasets	Train / Validation / Test
baidu-V1 [16]	4,301 / 538 / 538
baidu-V2 [16]	4,148 / 519 / 519
EG1800 [5]	1,360 / 170 / 170
SuperviselyPortrait [17]	2,269 / 284 / 284
HP-multi-person	1,544 / 193 / 193
HP-portrait	4,300 / 539 / 538
VideoMatte240K [18]	14,200 / 150 / 150

Because the datasets used for training do not contain depth map, we manually generate them using Dense Prediction Transformer [19], which is a depth estimation function.

Training

Since depth-wise separable convolution have fewer parameters compared to normal convolution layers, it is more vulnerable to gradient vanishing during training [20]. To make the training process more robust and converge faster, we use the MV3-Large encoder weights pre-trained on ImageNet [21] and the decoder weights from an unmodified PortraitNet. Our model is trained on two Nvidia RTX 3090 GPU at a resolution of 224×224. First, we train the model on the segmentation portrait image datasets for 100 epochs with the initial learning rate set to $1e^{-4}$, and the learning rate is reduced to $1e^{-5}$ after 50 epochs. Then we fine tune the model on the segmentation portrait video dataset for 50 epochs with the learning rate set to $1e^{-5}$.

Since portrait segmentation is a two-class segmentation problem, we use the binary cross entropy (BCE) loss during the training of the model:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \times log(p_i) + (1-y_i) \times log(1-p_i) \tag{4}$$

where p_i and y_i denote the predicted probability and the ground truth label for the ith pixel, respectively.

CNN model accuracy

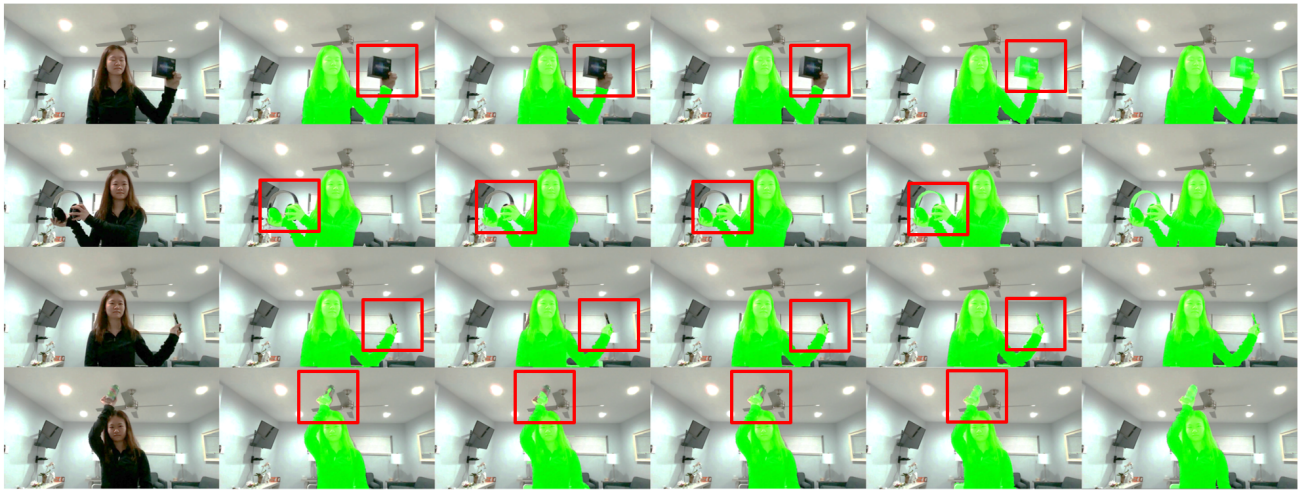
Our trained model is compared to other PortraitNet modifications including MaskPortraitNet [22], GRUPortraitNet [22], and FlowPortraitNet [22] as described in Table 2. The reported testing accuracy is resulted from four public portrait segmentation datasets and evaluated by mean intersection over union (mIoU). mIoU is calculated based on Equation (3).

$$mIoU = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{Pred_i \cap GT_i}{Pred_i \cup GT_i} \tag{5}$$

where $Pred_i$ and GT_i are the predicted and the ground truth masks for the ith frames. The accuracy of MaskPortraitNet, GRUPortraitNet, and FlowPortraitNet are reported by Xu et al. [22]. Results show that our method achieves the best performance among all the modifications to PortraitNet.

Depth assisted refinement accuracy

We compare the refined segmentation mask accuracy of our method to traditional refinement methods including Slic [10] and GrabCut [11] as shown in Table (3). We evaluate the accuracy on all the portrait image segmentation datasets at the original resolution. Results show that our proposed depth assisted mask refinement achieves the best result among the refinement methods.



(a) (b) (c) (d) (e) (f)

Figure 3: Examples of visual comparison between our proposed portrait video segmentation and selected baseline methods. Differences are highlighted in red bounding boxes. (a) original input frames, (b) segmentation masks predicted by our proposed portrait segmentation model, (c) segmentation masks refined using Slic, (d) segmentation masks refined using GrabCut, (e) segmentation masks refined using proposed depth assisted mask refinement, (f) groudtruth label.

Table 2: Semantic Portrait Segmentation Accuracy Comparision

Method	Dataset	Guidance	mIoU
PortraitNet [1]	Image	None	90.5%
PortraitNet [22]	Image	Prior Mask	92.6%
PortraitNet [22]	Video	Prior Mask	96.7%
GRUPortraitNet [22]	Video	Memory	95.4%
FlowPortraitNet [22]	Video	Optical Flow	94.3%
Ours	Video	Depth Map	**97.0%**

Table 3: Semantic Portrait Segmentation Accuracy Comparison

CNN	Refinement	mIoU
Ours	None	92.3%
Ours	Slic [10]	92.1%
Ours	Grabcut [11]	92.6%
Ours	Depth assisted	**93.4**%

Conclusion

In this paper, we propose a depth-assisted portrait video background blurring method, which consists of a modified RGB-D PortraitNet CNN model, a depth-assisted refinement module, and a bokeh rendering module. Our proposed method is capable of producing fast and accurate segmentation of human and in-hand objects compared to existing methods by taking into consideration the depth information. Our bokeh effect rendering is able to render high-quality bokeh effects that look more natural and aesthetic. Our method is evaluated on both image and video portrait segmentation datasets, which outperforms existing methods in terms of accuracy and speed.

References

[1] S.-H. Zhang, X. Dong, H. Li, R. Li, and Y.-L. Yang, "Portraitnet: Real-time portrait segmentation network for mobile device," *Computers & Graphics*, vol. 80, pp. 104–113, 2019.

[2] H. Park, L. Sjosund, Y. Yoo, N. Monet, J. Bang, and N. Kwak, "Sinet: Extreme lightweight portrait segmentation networks with spatial squeeze module and information blocking decoder," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2066–2074, Mar. 2020, Colorado, USA.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.

[4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[5] X. Shen, A. Hertzmann, J. Jia, *et al.*, "Automatic portrait segmentation for image stylization," *Computer Graphics Forum*, vol. 35, no. 2, pp. 93–102, May 2016.

[6] R. Kumar, R. Kumar, V. Gopalakrishnan, and K. Nanjunda Iyer, "Fast human segmentation using color and depth," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 1922–1926. DOI: 10.1109/ICASSP.2017.7952491.

[7] N. Singh, M. Kumar, P. Mahesh, and R. Sarkar, "Depth aware portrait segmentation using dual focus images," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018, pp. 1–6.

[8] F. Wang, Y. Zhang, Y. Ai, and W. Zhang, "Rendering natural bokeh effects based on depth estimation to improve the aesthetic ability of machine vision," *Machines*, vol. 10, no. 5, 2022.

[9] A. Ignatov, J. Patel, and R. Timofte, "Rendering natural camera bokeh effect with deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 418–419.

[10] C. Wilms and S. Frintrop, "Superpixel-based refinement for object proposal generation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021, pp. 4965–4972.

[11] C. Rother, V. Kolmogorov, and A. Blake, "" grabcut" interactive foreground extraction using iterated graph cuts," *ACM transactions on graphics (TOG)*, vol. 23, no. 3, pp. 309–314, 2004.

[12] A. Howard, M. Sandler, G. Chu, *et al.*, "Searching for mobilenetv3," *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314–1324, Oct. 2019, Seoul, South Korea.

[13] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[14] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm," *Diagnostics*, vol. 9, no. 3, p. 72, 2019.

[15] X. Wu, S. Wen, and Y.-a. Xie, "Improvement of mask-rcnn object segmentation algorithm," pp. 582–591, 2019.

[16] Z. Wu, Y. Huang, Y. Yu, L. Wang, and T. Tan, "Early hierarchical contexts learned by convolutional networks for image segmentation," *2014 22nd International Conference on Pattern Recognition*, pp. 1538–1543, Aug. 2014, Stockholm, Sweden.

[17] D. Drozdov, M. Kolomeychenko, and Y. Borisov, "Supervisely," *supervise. ly, https://supervise. ly*, 2020.

[18] S. Lin, L. Yang, I. Saleemi, and S. Sengupta, "Robust high-resolution video matting with temporal guidance," *2022 IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3132–3141, 2022, Hawaii, USA.

[19] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12 179–12 188, Oct. 2021, Montreal, Canada.

[20] F. Liu, H. Xu, M. Qi, D. Liu, J. Wang, and J. Kong, "Depth-wise separable convolution attention module for garbage image classification," *Sustainability*, vol. 14, no. 5, 2022.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[22] W. Xu, Y. Shen, Q. Lin, J. P. Allebach, and F. Zhu, "Efficient real-time portrait video segmentation with temporal guidance," *Electronic Imaging*, vol. 34, pp. 1–7, Jan. 2022, California, USA.

Author Biography

Weichen Xu *is pursuing a Ph.D. degree in Electrical and Computer Engineering at Purdue University and working as the graduate research assistant in the Video and Image Processing Laboratory. He received his B.Eng degree in Automation from Northeastern University, Shenyang, China. His research interests include video processing and deep learning.*

Yezhi Shen *is a Ph.D. student of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana. He received the B.S.E.E. degree in Electrical and Computer Engineering from Purdue University in 2021. His major research area is in computer vision.*

Qian Lin *is an HP Fellow working on computer vision and deep learning research. She is also an adjunct professor at Purdue University. She joined Hewlett-Packard Company in 1992. She received her BS from Xi'an Jiaotong University in China, her MSEE from Purdue University, and her Ph.D. in Electrical Engineering from Stanford University. She is inventor/co-inventor for 45 issued patents. She was awarded Fellowship by the Society of Imaging Science and Technology (IS&T) in 2012, Outstanding Electrical Engineer by the School of Electrical and Computer Engineering of Purdue University in 2013, and the Society of Women Engineers Achievement Award in 2021.*

Jan P. Allebach *is Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering at Purdue University. He was named Electronic Imaging Scientist of the Year by IS&T and SPIE, and was named Honorary Member of IS&T, the highest award that IS&T bestows. He has received the IEEE Daniel E. Noble Award, the IS&T/OSA Edwin Land Medal, the IS&T Johann Gutenberg Prize, is a Fellow of the National Academy of Inventors, and is a member of the National Academy of Engineering.*

Fengqing Zhu *is an Associate Professor of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana. She received the B.S.E.E. (with highest distinction), M.S. and Ph.D. degrees in Electrical and Computer Engineering from Purdue University. She is the recipient of an NSF CISE Research Initiation Initiative (CRII) award in 2017, a Google Faculty Research Award in 2019, and an ESI and trainee poster award for the NIH Precision Nutrition workshop in 2021. She is a senior member of the IEEE.*