High-Accuracy Automatic Person Segmentation with Novel Spatial Saliency Map

Weijuan Xi[1], Jianhang Chen[1], Qian Lin[2] and Jan P. Allebach[1]

[1]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, U.S.A.; [2]HP Labs, HP Inc., Palo Alto, CA, U.S.A.

Introduction

In this work, we propose a **person segmentation system** that achieves high segmentation accuracy with a much smaller CNN network. In this approach, **key-point detection annotation** is incorporated for the first time and a novel **spatial saliency map**, in which the intensity of each pixel indicates the likelihood of forming a part of the human and reflects the distance from the body, is generated to provide more spatial information. Additionally, a **LightWeight automatic Person Segmentation Network (**LWPSN**)** is proposed, which is **small and efficient** for person segmentation by leveraging atrous convolution.

Proposed Method

As illustrated in Fig. 1, our high-accuracy person segmentation system takes a person image as input, and produces the person segmentation mask as output.

- The original image is resized and fed into a pose detector to extract the keypoints of the human pose.
- A novel Spatial Saliency Map (SSM) is generated relying on these keypoints, and expressing the distance from each arbitrary pixel to the human body with a gradient saliency value.
- The image and the corresponding SSM are concatenated and fed into the LWPSN block.

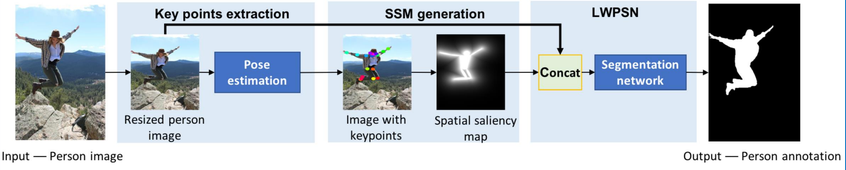

Fig.1. High-accuracy automatic person segmentation system overview.

In the process of SSM generation as shown in Fig. 2, the SSM is based on the associated relationships of the anatomical keypoints (b). The skeleton map (c) is generated by connecting the associated keypoints with ovals. The frame map (d) is generated by filling the torso and head regions of the skeleton map. The final SSM (e) is generated by computing a saliency value as the likelihood that each pixel belongs to the person.

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

Fig. 2. (a) The original color person image. (b) The result from the pose detector. (c) The skeleton map. (d) The frame map. (e) The final SSM.

The proposed LWPSN has a much smaller model, and feeds forward the SSM as the fourth channel of input as shown in Fig. 3. The LWPSN consists of three parts:

1. a feature extractor with our compressed version of ResNeXt, which only have 29 layers and 32 in cardinality
2. a Atrous Spatial Pyramid Pooling (ASPP) layer with dilation rates of 4, 8, and 12, respectively
3. a decoder to recover object segmentation details

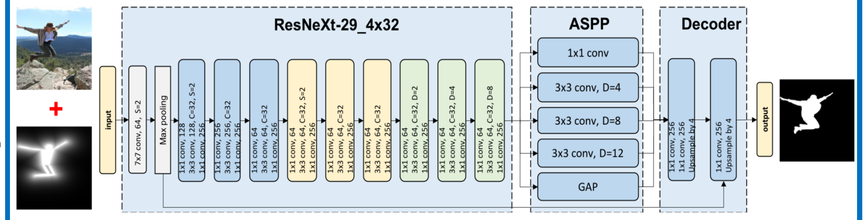

Fig. 3. Lightweight person segmentation network (LWPSN) architecture.

Experimental Results

We evaluate the system on two dataset: 1. portrait segmentation dataset [10]. 2. a new dataset named person segmentation dataset. We applied **an image pyramid resizing** (IPR) data augmentation to maintain the segmentation precision with low resolution inputs. IPR fits the input image size to the same aspect ratio as the network input. to reduces a considerable amount of computational load.

Table 1. Comparisons on the portrait segmentation dataset.

Method	Model size	mIoU
Graph-cut [2]	–	80.0%
PortraitFCN [10]	537.1 MB	95.9%
PortraitDeepLabv2 [22]	530.1 MB	96.1%
BSN [11]	530.1 MB	96.7%
PortraitDeepLabv3+ (ours)	161.0 MB	**97.59%**
LWPSN (ours)	**55.5 MB**	97.53%

Table 2. Comparisons on the person segmentation dataset.

Method	mIoU
FCN	83.11 %
PersonDeepLabv2	88.32 %
PersonDeepLabv3+	93.00 %
LWPSN+SSM (ours)	**94.06 %**
LWPSN+SSM+IPR (ours)	**93.11%**

Our work combines pose-detection annotation for persons to improve the segmentation precision. This approach can also be extended to the segmentation for other objects. The experimental results prove that the SSM generated from the pose detector and IPR increases the efficiency with a small network. The proposed system obtains state-of-the-art accuracy on the person segmentation dataset.

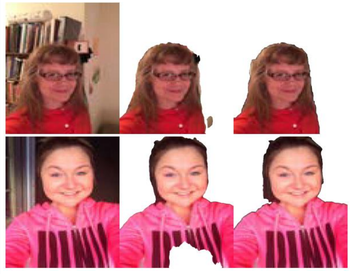

Fig. 4. Portrait segmentation dataset results: Left: Original images; Middle: PortraitFCN; Right: LWPSN

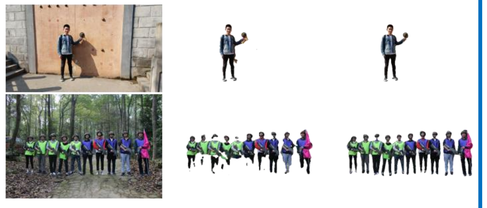

Fig. 5. Person Segmentation dataset results: Left: Original images; Middle: LWPSN + IPR; Right: LWPSN + SSM + IPR