Two-stage human hair segmentation in the wild using deep shape prior

Yongzhe Yan, Stefan Duffner, Xavier Naturel, Anthony Berthelier, Christophe Garcia, Christophe Blanc, Thierry Chateau

Two-stage human hair segmentation in the wild using deep shape prior

Yongzhe Yan[a,c,**], Stefan Duffner[b], Xavier Naturel[c], Anthony Berthelier[a,c], Christophe Garcia[b], Christophe Blanc[a], Thierry Chateau[a]

[a] *Université Clermont Auvergne, CNRS, SIGMA, Institut Pascal, F-63000 Clermont-Ferrand, France*
[b] *Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France*
[c] *Wisimage, France*

ABSTRACT

Human hair is a crucial biometric characteristic with rich color and texture information. In this paper, we propose a novel hair segmentation approach integrating a deep shape prior into a carefully designed two-stage Fully Convolutional Neural Network (FCNN) pipeline. First, we utilize a FCNN with an Atrous Spatial Pyramid Pooling (ASPP) module to train a human hair shape prior based on a specific distance transform. In the second stage, we combine the hair shape prior and the original image to form the input of a symmetric encoder-decoder FCNN with a border refinement module to get the final hair segmentation output. Both quantitative and qualitative results show that our method achieves state-of-the-art performance on the LFW-Part and Figaro1k datasets.

1. Introduction

Human hair contains rich color, shape and textural information. At the same time, it is generally related to gender, culture and appearance. Muhammad et al. (2018) argued that hair detection and segmentation is important in two domains. Firstly, hair segmentation is an essential prior step for 3D hair modeling from a single portrait image as well as for some Augmented Reality (AR) applications such as hair dying and facial animation. Secondly, it can be used in biometric recognition applications such as human presence detection from the back view or gender and face recognition.

Hair segmentation *in the wild* consists in performing hair segmentation in an unconstrained view without any explicit prior face or head-shoulder detection (Muhammad et al., 2018). We address this problem as a semantic segmentation problem by taking texture and shape constraints into account. Hair segmentation, especially under such unconstrained conditions, is challenging for the three following reasons:

- **Cluttered background**: textures in the background can be similar to human hair, which introduce significant difficulties for hair segmentation *in the wild*.
- **Lack of rigid and consistent form**: the form of hair can be totally different according to the head pose, different points of view and ambient environment such as wind. However, we believe that human hair, although in different situations, share implicit shape constraints.
- **Hair style/color variation**: There are numerous appearance variations in terms of hair style such as straight hair, curly hair, braided hair, short hair, etc. Hair colors are divergent from person to person and can be easily biased by different environment and cameras.
- **Complex lighting conditions**: Under complex lighting conditions, hair texture information is usually distorted. It is even difficult for a human to figure out the exact boundary of human hair in extreme lighting situations for example in shadow or backlight.

In this paper, we aim at improving hair segmentation *in the wild* by correctly distinguishing hair texture from similar texture in the background as well as estimating refined hair borders. Previous CNN-based methods (Levinshtein et al., 2017; Liu et al., 2017c) generally adopt a single stage, which is insufficient under such extreme conditions. We propose a two-stage pipeline (see Fig. 1.) consisting of a shape prior detection stage and a hair segmentation stage. Our contributions can be summarized as follows:

**Corresponding author:
e-mail: yongzhe.yan@etu.uca.fr (Yongzhe Yan)

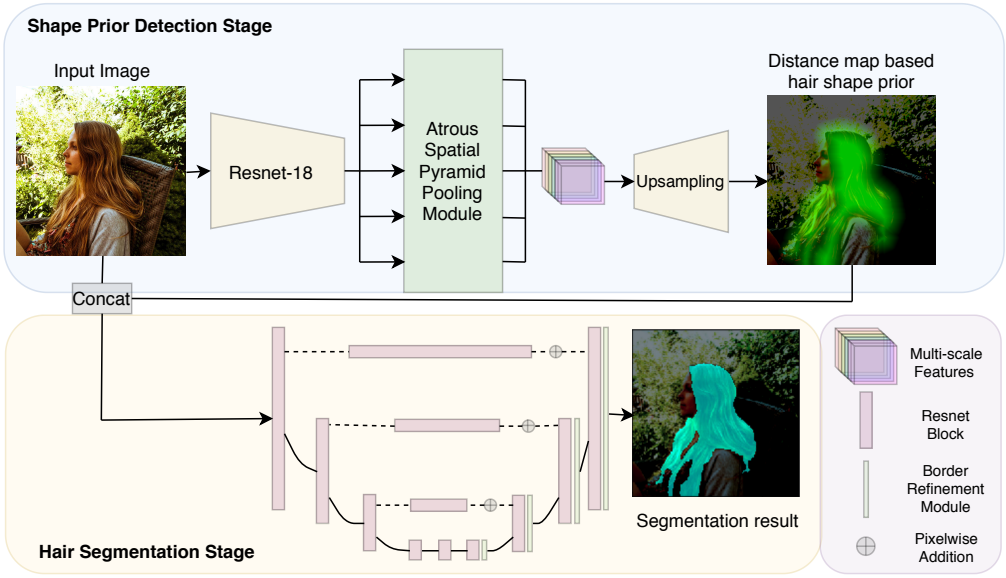


Fig. 1: Our two-stage human hair segmentation pipeline.

1. Before segmentation, we propose to first detect a hair shape prior which is based on a specific distance transform map. The results show that it helps to improve the robustness with cluttered background.
2. In the segmentation stage, we propose a border refinement module along with a symmetric encoder-decoder FCNN to obtain a more precise segmentation output.

2. Related Work

2.1. Semantic Segmentation

Hair segmentation can be seen as a type of semantic segmentation, a problem for which Convolutional Neural Networks have achieved remarkable results in the past few years. According to Chen et al. (2018b), there are mainly two types of CNN models: the encoder-decoder structures which privilege refined boundaries (Badrinarayanan et al., 2017; Ronneberger et al., 2015), and structures integrating a spatial pyramid pooling module (Chen et al., 2018a; He et al., 2015; Zhao et al., 2017), which gather rich multi-scale contextual information. The former ones normally adopt a symmetric structure with skip connections which enable low-level information to flow from the encoder directly to the decoder. This is now widely used in various applications such as image matting (Xu et al., 2017), landmark detection (Newell et al., 2016; Bulat and Tzimiropoulos, 2017) etc. The latter ones employ "atrous" convolutions (Holschneider et al., 1990) at different rates to capture features in arbitrary resolutions and show excellent performance on large-scale semantic segmentation datasets (Everingham et al., 2015; Cordts et al., 2016; Zhou et al., 2017). In 2019, **?** proposed to search for an optimal DeepLab structure.

2.2. Texture Recognition and Segmentation

The most characteristic feature of human hair is texture. Texture recognition is usually considered as a basic image processing problem without taking more semantic information into account. As a result, many approaches are based on the Bag of Words model (Leung and Malik, 2001) to obtain spatially invariant features for texture representation (Liu et al., 2018). Recently, CNN-based methods with orderless feature pooling (Gong et al., 2014; Cimpoi et al., 2015; Zhang et al., 2017) have shown good performance on texture recognition, which was later proved to be beneficial for semantic segmentation (Zhang et al., 2018). In terms of texture segmentation, many approaches are based on active contours and integrate different texture features (Wu et al., 2015; Reska et al., 2015; Varnosfaderani and Moallem, 2017; Liu et al., 2017a; Yuan et al., 2015; Gao et al., 2016). Cimpoi et al. (2015) proposed to use object detection-like region proposal classification to assign the texture/object labels to each pixel.

2.3. Coarse-to-fine Segmentation

Recently, there are several works concerning the refinement for the CNN-based semantic segmentation. Chen et al. (2015) used Conditional Random Field (CRF) to establish an additional pair-wise supervision between the pixels. Wu et al. (2018) proposed a guided image filter, which is designed to generate a high-resolution output and from a low-resolution input given a guidance input. Liu et al. (2017b) proposed to construct a linear propagation model, which constitutes a spatial affinity matrix that models dense, global pairwise relationships of an image. Similarly, Jiang et al. (2018) proposed DifNet, which models the pairwise information by cascaded random walks. Our method uses spatial attention as additional supervision for boundary refinement. Compared to the previous methods, our method bears two advantages: (1) Unlike CRF and DifNet, our approach does not require iterative operations. (2) Our border refinement module can be easily integrated with the feature maps of different scales in most of the layers.

Li et al. (2017) found that most of the difficult pixels are located on the boundaries. Therefore they proposed a cascaded

scheme to process all the pixels step-by-step from easy (center pixels) to hard (boundary pixels). Zhu et al. (2019) proposed a boundary label relaxation strategy to alleviate the influence of the hard pixels on the boundary on the overall score. In our approach, the first stage is trained to learn a general shape without detailed boundaries, which prevents to mistaking the noises on the cluttered background.

Recently, more researchers started to use shape and boundary of the objects to help the semantic segmentation. **?** proposed a shape-variant context in the semantic segmentation to adapt diverse shapes and scales. Another work (**?**) used the natural boundary information to build stronger connections within the same object. **?** proposed a two-stream CNN architecture, in which the shape information is explicitly processed in parallel with the main semantic segmentation stream.

2.4. Human Hair Segmentation

Early methods proposed to segment human hair by modeling color, location and frequency information (Yacoob and Davis, 2006; Lee et al., 2008; Rousset and Coulon, 2008). Wang et al. (2010, 2012) decompose the hair segmentation into local parts and several other approaches (Wang et al., 2009, 2011, 2013) use region growing followed by refining regression on the coarse mask. Recent work (Chai et al., 2016; Qin et al., 2017; Guo and Aarabi, 2018; Levinshtein et al., 2017) based on FCNN models achieved good performance for practical applications. However, most of the methods only focus on constrained conditions such as head-shoulder images.

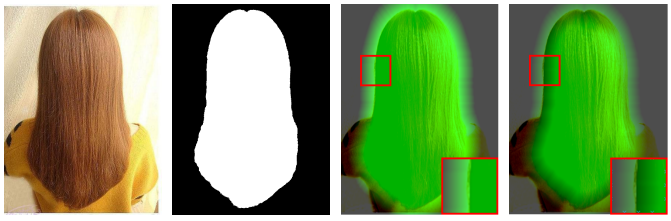
Muhammad et al. (2018) proposed a challenging hair analysis dataset along with a method to realize hair detection, segmentation and style classification. Their method renders quite good detection. However, they perform a sliding window texture recognition operation on the whole image, which is computationally very expensive.

3. Proposed Approach

We decouple the hair segmentation task into two important steps: (a) find the general hair shape prior and (b) find the refined border of the hair. In the hair shape detection stage, inspired by the hair occurrence probability mask used in previous methods (Wang et al., 2011, 2012; Muhammad et al., 2018) and soft segmentation, we aim at finding a coarse hair mask that indicates the hair texture presence (a coarse hair shape prior) regardless of the exact border. In the hair segmentation stage, we aim at identifying the exact hair border by integrating a border refinement module in a symmetric encoder-decoder FCNN.

3.1. Hair Shape Prior Detection

Distance map regression. As stated before, two significant challenges for hair segmentation *in the wild* are: to distinguish hair appearance from similar background texture and learning challenging hair shape geometries. In most of the previously proposed FCNN models for semantic segmentation, object shape constraints are not explicitly imposed. We propose to introduce a coarse mask without precise boundary as a shape prior for hair segmentation. We transform the binary ground truth hair mask to a boundary-less coarse hair mask by using a specific type of distance transform.



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

Fig. 2: An illustration of our distance map transformation. From left to right: (a) Original image (b) Ground truth hair mask (c) Clipped distance transform map overlaid on original image (d) Clipped distance transform map with "erosion" overlaid on the original image. With "erosion", an uncertain region is created on both sides of the hair boundary.

An illustration of our distance map transform is shown in Fig. 2. Consider a binary ground truth hair mask $I(x, y)$ in Fig. 2(b). Hair pixels and non-hair pixels can be denoted respectively as $I^+ = \{I(x, y) = 1\}$ and $I^- = \{I(x, y) = 0\}$. We define a clipped distance transform map dt_{mask} on the image positions $p(x, y)$ as:

$$dt_{mask}(p) = d_{max} - \min(d_{max}, \min_{p^+ \in I^+} \left\| p^+ - p \right\|) \quad (1)$$

where d_{max} denotes the maximum clipping threshold for distance values (see Fig. 2(c)). And, similarly, we define a clipped inverse distance transform map with respect to the background pixels:

$$dt_{inv}(p) = e_{max} - \min(e_{max}, \min_{p^- \in I^-} \left\| p^- - p \right\|) \quad (2)$$

where $e_{max} (< d_{max})$ denotes the second clipping threshold. Then, the final distance transform map dt is obtained by:

$$dt = dt_{mask} - dt_{inv} \quad (3)$$

which is then normalized between -1 and +1 to form as a regression target (see Fig. 2(d)). The use of dt_{inv} "erodes" the initial distance transform dt_{mask} and produces an uncertain hair boundary region for the target image. e_{max} can be considered as the magnitude of "erosion". We do not use traditional morphological erosion to do this because some small hair regions on the binary mask might be ignored while small holes might be filled. We use "HardTanh" as final activation function (see Fig. 4). This activation function is a linear approximation of Tanh function and clipped from -1 to +1, which naturally fits the range of our shape prior regression target. We use L1 loss to train our distance map regression. In our implementation, we empirically set d_{max} to 25 and e_{max} to 10.

Atrous Spatial Pyramid Pooling (ASPP) encoder. Although texture is considered as very local information, in the setting of hair segmentation *in the wild*, the scale of the hair region varies considerably. ASPP with different atrous rates effectively captures multi-scale information to learn the presence of hair texture. We use DeeplabV3 (Chen et al., 2018a) structure with Resnet18 (He et al., 2016) pre-trained on ImageNet as backbone encoder in our hair detection network. Finally we upsample the multi-scale feature map to obtain the final distance transform map at the original image size.

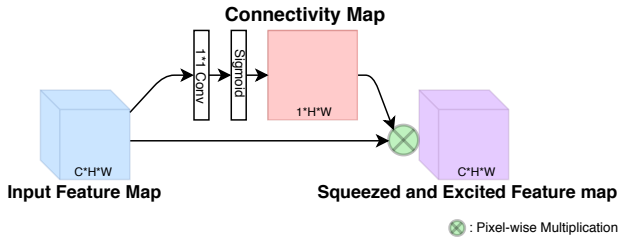


Fig. 3: An illustration of our border refinement module. The input feature map is transformed into a **single channel** Connectivity Map by an 1×1 convolutional layer with sigmoid activation. The Connectivity Map is then multiplied with each channel in the input feature map before output.

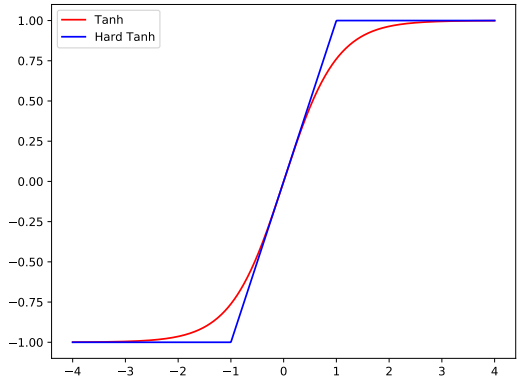


Fig. 4: An illustration of Hard Tanh activation function.

3.2. Refined Hair Segmentation

Symmetric encoder-decoder. In hair segmentation stage, we implement a symmetric encoder-decoder structure with skip connections. At each level, we use a ResNet block in both the encoder and the decoder part. Additionally, as in Newell et al. (2016), we add a ResNet block in the skip connections to process the low-level information transferred from the decoder.

Border refinement module. The boundary of the human hair is difficult to segment due to the presence of tiny details. These tiny details only concern limited number of pixels. However, the final rendering might be visually unsatisfying if they are not well treated. To refine the hair boundary, we propose to use spatial attention, which will help the CNN to focus on the pixels around the hair boundary.

In Zhang et al. (2018), the authors implemented a squeeze-and-excitation channel-wise attention (Hu et al., 2018) module for semantic segmentation. Here in refinement segmentation stage, we are more interested in pixel-wise attention to recover refined hair border. Inspired by this work, we propose a refinement module which generates spatial attention. The input feature map is passed through a 1×1 kernel convolutional layer with a sigmoid activation function to a single-channel feature map. We call it connectivity map. It is multiplied by each channel of the input feature maps afterwards to obtain the "squeezed and excited" output feature map. The module is illustrated in Fig. 3. We place this module at each level of the decoder part before upsampling. We noticed that this module helps to improve the performance, smooth the boundary and get better visual result.

4. Experiments

4.1. Datasets

We conducted our experiments on LFW-Part dataset (Kae et al., 2013) and the newly-released Figaro-1k (Muhammad et al., 2018) dataset. The LFW-Part dataset is a face parsing dataset with hair annotation which consists of 2927 images. To the best of our knowledge, Figaro-1k is the only hair analysis dataset *in the wild* with precise hair annotation. It consists of 1050 images (210 for validation) and manually annotated ground truth hair masks, which varies in seven hair styles, different hair colors, length and levels of background complexity.

4.2. Experimental Settings

For quantitative evaluation, we adopted several standard measures e.g. mean Intersection over Union (mIoU), accuracy and F1-score. The images are resized to 256×256 for training. The evaluation is performed at their original size. For data augmentation, due to limited number of images in the Figaro1k dataset, we apply various data augmentation skills on input images during training: (1) a random resize of $\pm$ 20 % on image width and height (2) a random translation of $\pm$ 60 % in horizon and $\pm$ 30 % in vertical (3) a random crop/pad of $\pm$ 20 % on image width and height (4) a random horizontal flip (5) a color jitter (on brightness, contrast and saturation) of $\pm$ 30 % (6) a random gaussian lighting noise based on ImageNet PCA analysis.

4.3. Hyperparameter settings

Distance transform map regression in the hair detection stage is trained by using L1 loss while final refinement segmentation in the second stage is trained by using standard softmax loss. The first stage is pre-trained for several epochs before being jointly trained with the second stage. We use RMSprop to train the networks in both stages at the same time for 190 epochs with a initial learning rate of 0.0005 and batch size of 6. The learning rate is decayed by 0.3 for the first 30 epochs and then decayed in the same manner each 40 epochs. We use PyTorch to implement the training on a single NVIDIA GTX 1080Ti. The training stage finishes in around 13 hours for Figaro1k dataset. Each inference takes around 15ms compared to 1.79s in (Muhammad et al., 2018) and 3.3ms in (Liu et al., 2017c).

4.4. Quantitative Comparison

We compared our method with the encoder-decoder fully convolutional neural network U-Net (Ronneberger et al., 2015), the state-of-the-art semantic segmentation approach DeeplabV3+ (Chen et al., 2018b) based on ImageNet pre-trained ResNet18 and the previous work on hair analysis *in the wild* (Muhammad et al., 2018) on the Figaro-1k dataset. The result is reported in Table 1. Our approach outperforms all the previous methods for hair segmentation *in the wild*. By adding a detection stage, a gain of more than 1% point on IoU and F1-score can be achieved. The larger improvement on precision shows that our method is effective for removing false positives on the background. With the border refinement module, the performance is additionally improved by only a small margin but

Table 1: Hair Segmentation Results on Figaro1k.

Method	Precision(%)	F1(%)	mIoU(%)	Accuracy(%)
U-Net (Ronneberger et al., 2015)	95.63	94.39	89.69	96.36
DeeplabV3+ (Chen et al., 2018b)	96.86	95.05	91.11	97.07
Muhammad et al. (2018)	-	84.90	-	91.50
Only Seg Stage	95.64	94.53	89.91	96.56
Det Stage + Seg Stage	97.25	95.09	91.15	97.20
Det Stage + Seg Stage + Refinement	**97.33**	**95.15**	**91.25**	**97.23**

Table 2: Hair Segmentation Results on LFW-Part.

Method	Precision(%)	F1-hair(%)	mIoU(%)	Accuracy(%)
U-Net (Ronneberger et al., 2015)	89.11	87.66	88.33	96.58
DeeplabV3+ (Chen et al., 2018b)	91.66	88.36	**90.64**	96.82
Liu et al. (2017c)	-	83.43	-	95.46
Only Seg Stage	89.13	88.07	90.12	96.71
Det Stage + Seg Stage	98.24	88.94	90.53	96.76
Det Stage + Seg Stage + Refinement	**98.34**	**89.42**	90.60	**97.05**

Fig. 5: Challenging examples in Figaro1k. First row: Input image. Second row: Segmentation results by our model without detection stage. Third row: Segmentation results by ImageNet pre-trained DeepLabV3+. Fourth row: Segmentation results by our two-stage model. Many tiny isolated false positives can still be observed on the man's shirt in the first image of DeepLabV3+ results.

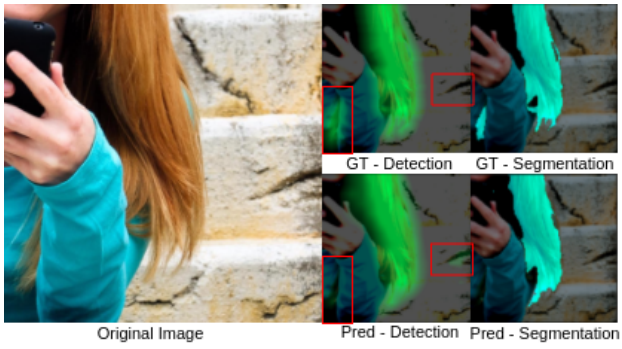


Fig. 6: An illustration of relation between shape prior and final segmentation. GT refers to ground truth and Pred denotes our two-stage network prediction. Weaker values are observed inside the red rectangles on the detection prediction.

gives better visual results. On the LFW-Part dataset, by adding a hair detection stage, our method outperforms other methods by more than 1% point on the hair F1-score (see Table. 2).

4.5. Discussions

Ablation study: Necessity of using hair shape prior. Fig. 5 shows several challenging images where the hair segmentation fails without shape prior. In these images, there are either similar textures or complicated lighting conditions present. We notice that (a) false positive segmentation on similar texture in the background is rectified and (b) tiny isolated false positive hair parts are suppressed. We think that the improvement originates from (1) the ImageNet pre-trained features, (2) our trained hair shape prior. To study the influence of (1), we compare our results with the ImageNet pre-trained DeepLabV3+. We find that our shape prior-integrated approach is more robust to cluttered background and renders hair shapes more reasonable in complex situations.

To investigate how the FCNN in the segmentation stage processes the shape constraint prior from the detection stage, we give an illustration in Fig. 6. With the help of the eroded distance map, the two small hair regions in the red rectangles are assigned weaker values. We note that both of them are eliminated by the FCNN in the segmentation stage, which removes one false positive but creates one false negative as well. Intuitively, this amounts to a learnable "thresholding".

Ablation study: Necessity of using border refinement module. In order to disconnect the influence of different shape priors from the detection stage, we pre-trained the detection network and fixed the weights during the training of the

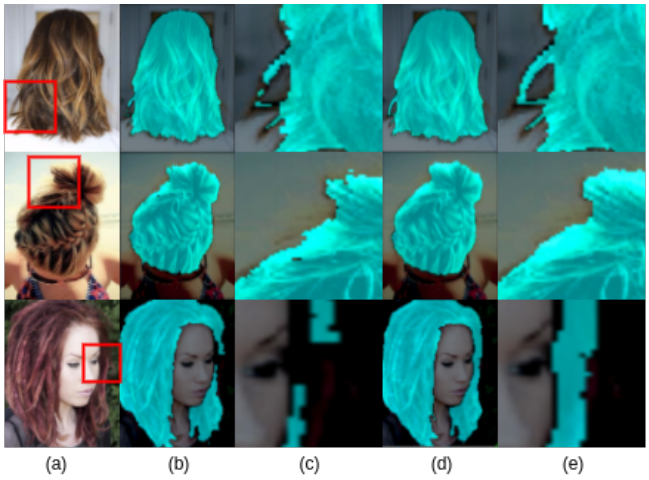


Fig. 7: Impact of the border refinement module (a) Original image (b)(c) Global and zoomed-in segmentation result w/o refinement module (d)(e) Global and zoomed-in segmentation results with border refinement module.

segmentation network to explore this necessity. Even though we remark only a slight quantitative improvement on mIoU in both datasets, we find that the results are visually better because of smooth boundaries as shown in Fig. 7. It provides more consistent hair regions and eliminates spurious small detection in the background thanks to the use of the pixel connectivity map.

To measure the number of the spurious small detection, we calculated the average number of connected components (hair regions) per image on the test set (see Tab. 3). The output given by the model with border refinement module has less connected components, indicating that the tiny isolated hair segmentation (as shown in Fig.7 (c)) appears less frequently when border refinement module is used.

Table 3: Number of connected components on Figaro-1k.

Models	Num. of connected components
w/o Refinement	3.94
w/ Refinement	**3.73**

Smoothed boundaries do not necessarily translate into a better mIoU, but are visually more appealing, even compared to the ground truth. In fact, even for humans, it is challenging to annotate the boundary in a very precise way by using only a binary mask. A promising future work to further improve the hair boundary could be image matting (Xu et al., 2017; Levinshtein et al., 2017).

Visual Comparison with Stae-of-the-art Methods. We visualize the output of our method on Figaro dataset in Fig. 8. We compare our method with the state-of-the-art method (Muhammad et al., 2018). We observe that our detection, especially the hair boundary, is much finer compared to Muhammad et al. (2018), which is more appealing for the practical applications such as virtual hair coloring.

We show some challenging examples on LFW-Part dataset in Fig. 9. We compare the visual results from (1) our model

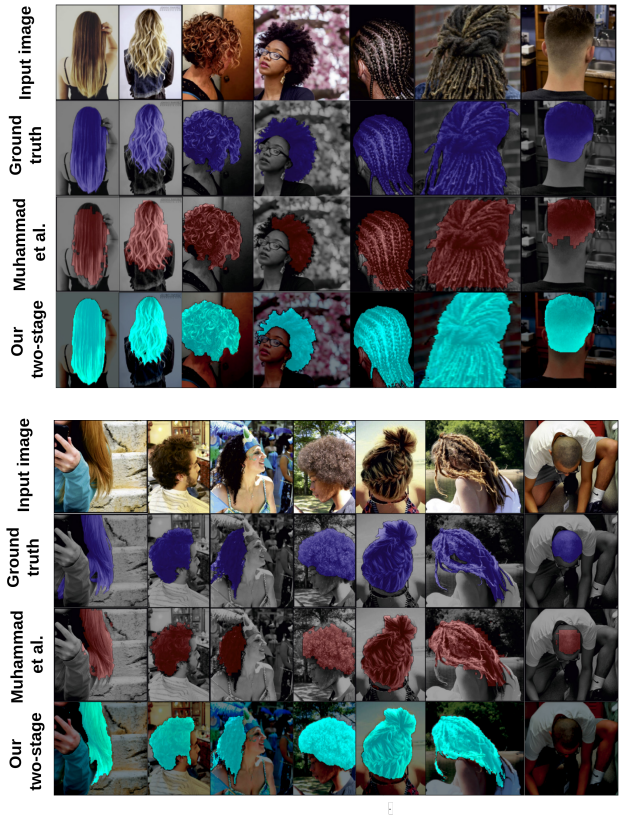


Fig. 8: Visual results compared to Muhammad et al. (2018). First row: Input image. Second row: Groundtruth. Third row: Results from Muhammad et al. (2018). Fourth row: Results from our two-stage human hair segmentation model. Best viewed in color.

without detection stage, (2) DeepLabV3+ [5] with ImageNet pretrained Resnet18 as backbone and (3) our model with both detection and segmentation stage. We observed that our method outperforms the others by suppressing the spurious detection on the cluttered background. It shows the importance of performing shape prior detection before segmentation stage.

Failure cases. In Fig 10, we provide several examples where our model fails. Our approach still cannot completely ensure the identification of the correct textures on the cluttered background especially when they are very close to the hair or has a similar form for example in (a) and (b). Furthermore, complex lighting conditions in (c) and irregular upside down pose in (d) introduce big challenges for our methods. Nonetheless, our method still suppressed more false positive detection than the segmentation-only-models such as ImageNet pre-trained DeepLabV3+. From (e), (f) and (g), we observe that our method is less sensitive to weakly-textured and small hair regions compared to other methods.

5. Conclusions

In this paper, we presented a two-stage pipeline for hair segmentation *in the wild*. We train a distance map-based hair shape prior from data, and then estimate the final segmentation by a symmetric FCNN using a border refinement module. Our

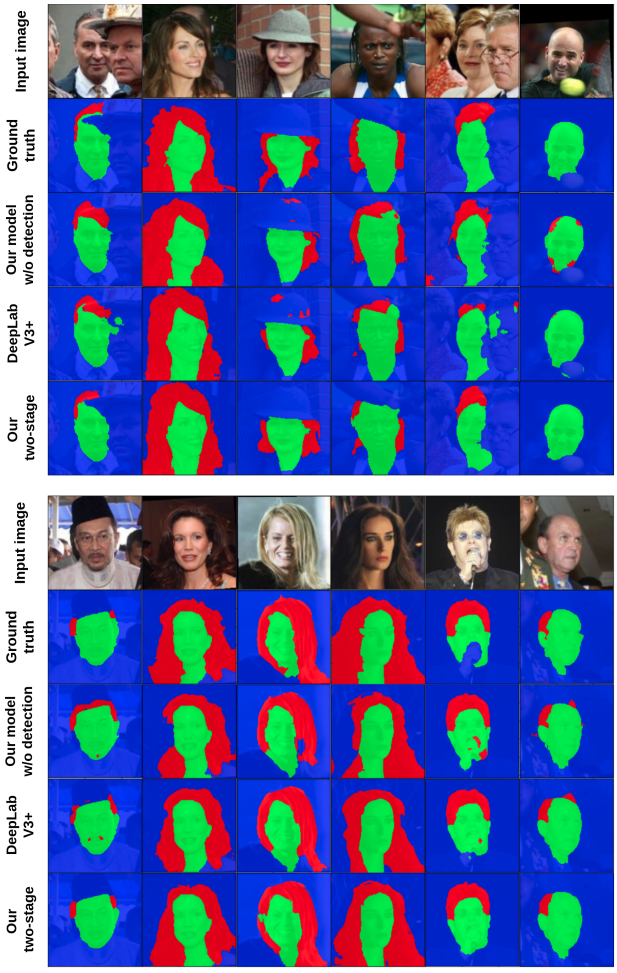


Fig. 9: Challenging examples on LFW-Part dataset. First row: Input image. Second row: Groundtruth. Third row: Results from our model without detection stage. Fourth row: Results from ImageNet pre-trained DeepLabV3+. Fifth row: Results from our model with both detection and segmentation stage. Red-Hair, Green-Face, Blue-Background. Best viewed in color.

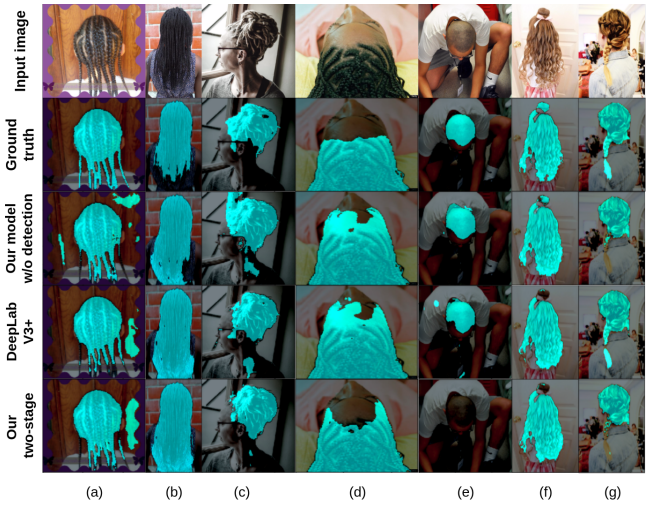


Fig. 10: Failure examples when using our methods. First row: Input image. Second row: Groundtruth. Third row: Results from our model without detection stage. Fourth row: Results from ImageNet pre-trained DeepLabV3+. Fifth row: Results from our model with both detection and segmentation stage. Best viewed in color.

approach outperforms previous state-of-the-art methods, being more robust to cluttered background and giving visually more consistent hair borders. Our approach can be further extended to textured object segmentation with difficult boundaries such as clothes parsing (Yamaguchi et al., 2012) and road scene parsing (Fritsch et al., 2013).

Acknowledgments

This work is supported by Région Auvergne-Rhône-Alpes, France. We would like to thank Nvidia for a Titan GPU donation.

References

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence , 2481–2495.

Bulat, A., Tzimiropoulos, G., 2017. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks), in: International Conference on Computer Vision, p. 4.

Chai, M., Shao, T., Wu, H., Weng, Y., Zhou, K., 2016. Autohair: fully automatic hair modeling from a single image. ACM Transactions on Graphics 35.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2015. Semantic image segmentation with deep convolutional nets and fully connected crfs, in: International Conference on Learning Representations (ICLR).

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40, 834–848.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: The European Conference on Computer Vision (ECCV).

Cimpoi, M., Maji, S., Vedaldi, A., 2015. Deep filter banks for texture recognition and segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3828–3836.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. International journal of computer vision 111, 98–136.

Fritsch, J., Kuehnl, T., Geiger, A., 2013. A new performance measure and evaluation benchmark for road detection algorithms, in: International Conference on Intelligent Transportation Systems (ITSC).

Gao, M., Chen, H., Zheng, S., Fang, B., 2016. A factorization based active contour model for texture segmentation, in: Image Processing (ICIP), 2016 IEEE International Conference on, IEEE. pp. 4309–4313.

Gong, Y., Wang, L., Guo, R., Lazebnik, S., 2014. Multi-scale orderless pooling of deep convolutional activation features, in: European conference on computer vision, Springer. pp. 392–407.

Guo, W., Aarabi, P., 2018. Hair segmentation using heuristically-trained neural networks. IEEE transactions on neural networks and learning systems 29, 25–36.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence 37, 1904–1916.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Holschneider, M., Kronland-Martinet, R., Morlet, J., Tchamitchian, P., 1990. A real-time algorithm for signal analysis with the help of the wavelet transform, in: Wavelets. Springer, pp. 286–297.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: IEEE Conference on Computer Vision and Pattern Recognition.

Jiang, P., Gu, F., Wang, Y., Tu, C., Chen, B., 2018. Difnet: Semantic segmentation by diffusion networks, in: Advances in Neural Information Processing Systems, pp. 1630–1639.

Kae, A., Sohn, K., Lee, H., Learned-Miller, E., 2013. Augmenting CRFs with Boltzmann machine shape priors for image labeling, in: CVPR.

Lee, K.c., Anguelov, D., Sumengen, B., Gokturk, S.B., 2008. Markov random field models for hair and face segmentation, in: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, IEEE. pp. 1–6.

Leung, T., Malik, J., 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. International journal of computer vision 43, 29–44.

Levinshtein, A., Chang, C., Phung, E., Kezele, I., Guo, W., Aarabi, P., 2017. Real-time deep hair matting on mobile devices. arXiv preprint arXiv:1712.07168 .

Li, X., Liu, Z., Luo, P., Chen, C.L., Tang, X., 2017. Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Liu, L., Chen, J., Fieguth, P., Zhao, G., Chellappa, R., Pietikainen, M., 2018. A survey of recent advances in texture representation. arXiv preprint arXiv:1801.10324 .

Liu, L., Fan, S., Ning, X., Liao, L., 2017a. An efficient level set model with self-similarity for texture segmentation. Neurocomputing 266, 150–164.

Liu, S., De Mello, S., Gu, J., Zhong, G., Yang, M.H., Kautz, J., 2017b. Learning affinity via spatial propagation networks, in: Advances in Neural Information Processing Systems, pp. 1520–1530.

Liu, S., Shi, J., Liang, J., Yang, M.H., 2017c. Face parsing via recurrent propagation, in: British Machine Vision Conference (BMVC).

Muhammad, U.R., Svanera, M., Leonardi, R., Benini, S., 2018. Hair detection, segmentation, and hairstyle classification in the wild. Image and Vision Computing 71, 25–37.

Newell, A., Yang, K., Deng, J., 2016. Stacked hourglass networks for human pose estimation, in: European Conference on Computer Vision, Springer. pp. 483–499.

Qin, S., Kim, S., Manduchi, R., 2017. Automatic skin and hair masking using fully convolutional networks, in: Multimedia and Expo (ICME), 2017 IEEE International Conference on, IEEE. pp. 103–108.

Reska, D., Boldak, C., Kretowski, M., 2015. A texture-based energy for active contour image segmentation, in: Image Processing & Communications Challenges 6. Springer, pp. 187–194.

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

Rousset, C., Coulon, P.Y., 2008. Frequential and color analysis for hair mask segmentation, in: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, IEEE. pp. 2276–2279.

Varnosfaderani, A.A., Moallem, P., 2017. Texture images segmentation by geometric active contour models based on color and gabor features. International Journal of Tomography & Simulation¢ 30, 105–117.

Wang, D., Chai, X., Zhang, H., Chang, H., Zeng, W., Shan, S., 2011. A novel coarse-to-fine hair segmentation method, in: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE. pp. 233–238.

Wang, D., Shan, S., Zeng, W., Zhang, H., Chen, X., 2009. A novel two-tier bayesian based method for hair segmentation, in: Image Processing (ICIP), 2009 16th IEEE International Conference on, IEEE. pp. 2401–2404.

Wang, D., Shan, S., Zhang, H., Zeng, W., Chen, X., 2013. Isomorphic manifold inference for hair segmentation, in: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE. pp. 1–6.

Wang, N., Ai, H., Lao, S., 2010. A compositional exemplar-based model for hair segmentation, in: Asian Conference on Computer Vision, Springer. pp. 171–184.

Wang, N., Ai, H., Tang, F., 2012. What are good parts for hair shape modeling?, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE. pp. 662–669.

Wu, H., Zheng, S., Zhang, J., Huang, K., 2018. Fast end-to-end trainable guided filter, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1838–1847.

Wu, Q., Gan, Y., Lin, B., Zhang, Q., Chang, H., 2015. An active contour model based on fused texture features for image segmentation. Neurocomputing 151, 1133–1141.

Xu, N., Price, B., Cohen, S., Huang, T., 2017. Deep image matting, in: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, IEEE. pp. 311–320.

Yacoob, Y., Davis, L.S., 2006. Detection and analysis of hair. IEEE Transactions on Pattern Analysis & Machine Intelligence , 1164–1169.

Yamaguchi, K., Kiapour, M.H., Ortiz, L.E., Berg, T.L., 2012. Parsing clothing in fashion photographs, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 3570–3577.

Yuan, J., Wang, D., Cheriyadat, A.M., 2015. Factorization-based texture segmentation. IEEE Transactions on Image Processing 24, 3488–3497.

Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., Agrawal, A., 2018. Context encoding for semantic segmentation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Zhang, H., Xue, J., Dana, K., 2017. Deep ten: Texture encoding network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. pp. 2896–2905.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2881–2890.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017. Scene parsing through ade20k dataset, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Zhu, Y., Sapra, K., Reda, F.A., Shih, K.J., Newsam, S., Tao, A., Catanzaro, B., 2019. Improving semantic segmentation via video propagation and label relaxation, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 8856–8865.