Boosting Semantic Human Matting with Coarse Annotations

Jinlin Liu[1,2] Yuan Yao[1] Wendi Hou[1] Miaomiao Cui[1]
Xuansong Xie[1] Changshui Zhang[2] Xian-sheng Hua[1]

[1]Alibaba Group, [2] Tsinghua University

{ljl191782, ryan.yy, wendi.hwd, miaomiao.cmm}@alibaba-inc.com xingtong.xxs@taobao.com
zcs@mail.tsinghua.edu.cn xiansheng.hxs@alibaba-inc.com

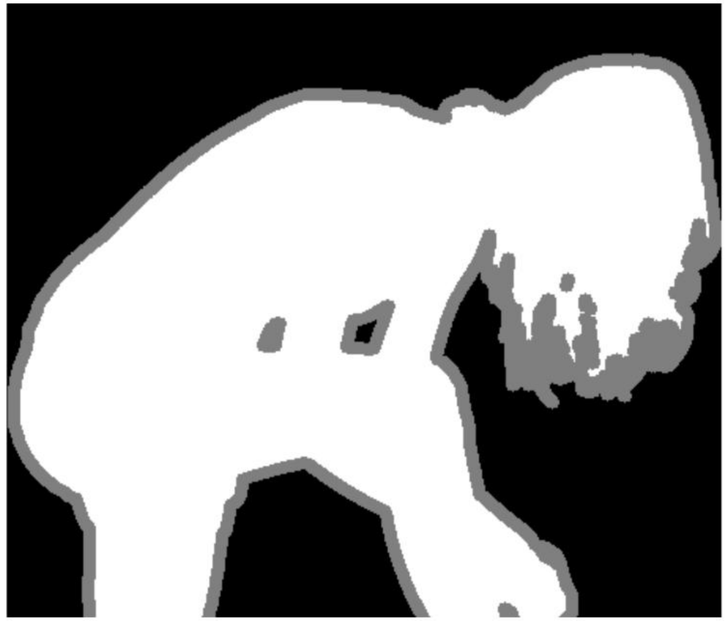
Overview

- Motivation
- Method
- Experiment
- Application
- Conclusion

Motivation

What’s human matting ?

Input



Trimap

Alpha matte

Motivation

Problems without trimap:

- Require plenty of fine annotated training images
- Annotating matting dataset is labor intensive and requires skills

Motivation

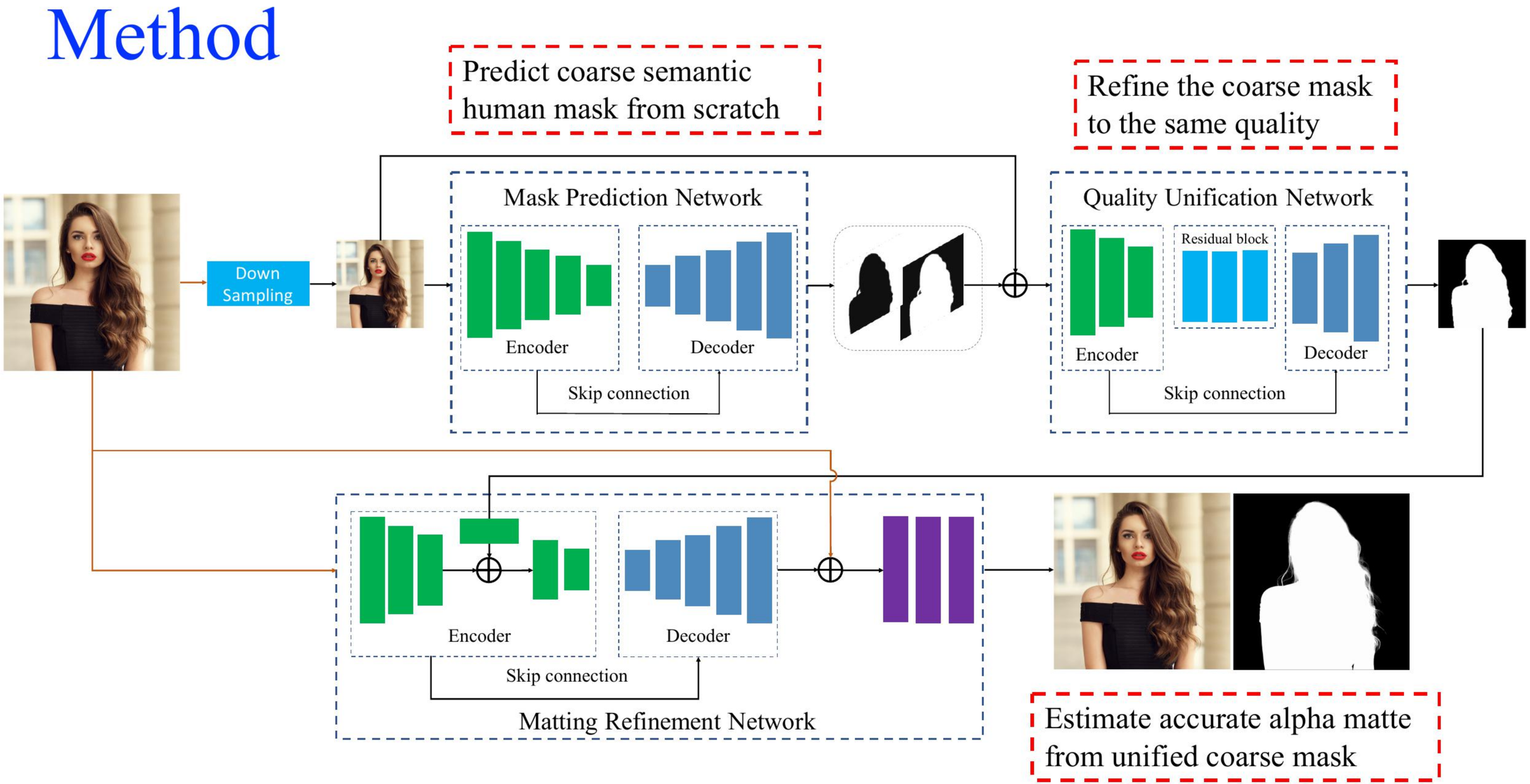
How to solve the problem without trimap?

- Using both coarse annotated and fine annotated data

Overview

- Motivation
- Method
- Experiment
- Application
- Conclusion

Method


Predict coarse semantic human mask from scratch
Refine the coarse mask to the same quality
Down Sampling
Mask Prediction Network
Encoder
Decoder
Skip connection
Quality Unification Network
Residual block
Encoder
Decoder
Skip connection
Encoder
Decoder
Skip connection
Matting Refinement Network
Estimate accurate alpha matte from unified coarse mask

How does Quality Unified Network work?

(a) Fine mask (b) Coarse mask (c) Diff map of (a,b)

(d) Unified (a) (e) Unified (b) (f) Diff map of (d,e)

(g) Diff map of (a,d) (h) Diff map of (b,e) (i) Input image

Overview

- Motivation
- Method
- Experiment
- Application
- Conclusion

Dataset

Table 1. The configurations of human matting datasets.

Dataset	Train Set		Test Set	
	Human	image	Human	image
Shen *et al.* [28]	1700	1700	300	300
TrimapDIM [32]	202	20200	11	220
SHM [8]	34493	34493	1020	1020
Ours(coarse)	10597	105970	125 (+11)	1360
Ours(fine)	9324(+202)	95260		

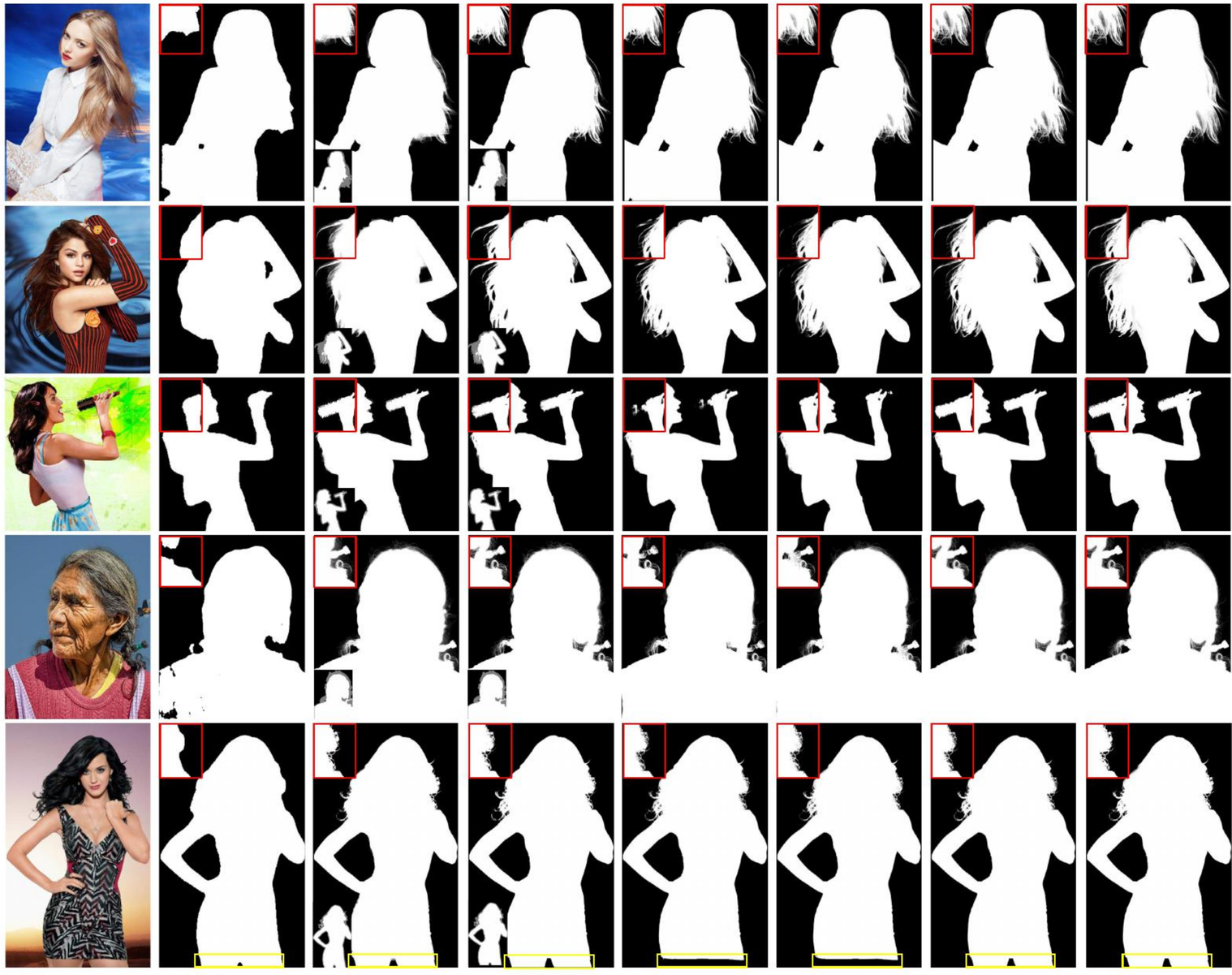
(a) Coarse annotated dataset

(b) Fine annotated dataset

Comparison

Comparison methods

- DeepLab
- Closed-form Matting
- Deep Image Matting
- Semantic Human Matting
- Ours using only fine data
- Ours using fine and coarse data



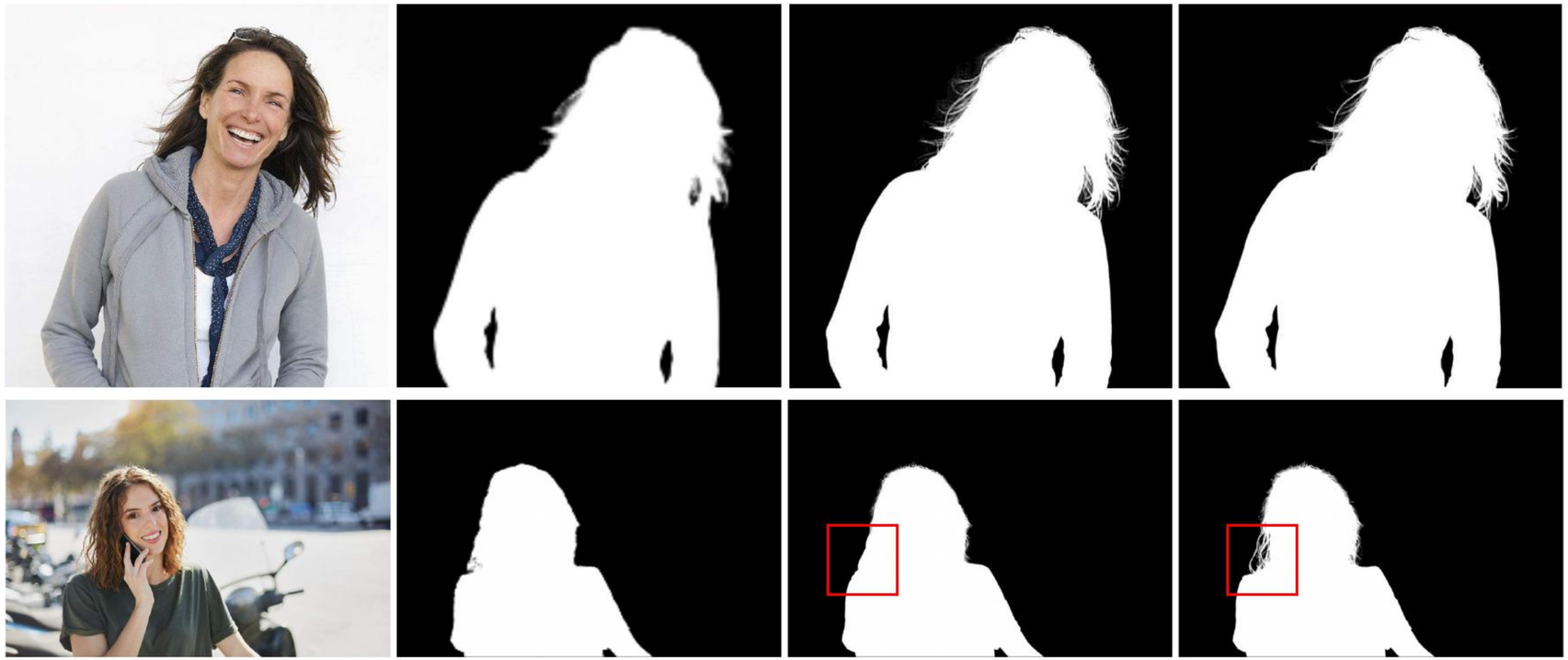
Quantitative results

Table 2. The quantitative results.

Method	SAD	MSE	Gradient	Connectivity
DeepLab [7]	0.028	0.023	0.012	0.028
Trimap+CF [21]	0.0083	0.0049	0.0035	0.080
Trimap+DIM [32]	**0.0045**	**0.0017**	**0.0013**	**0.0043**
SHM [8]	0.011	0.0078	0.0032	0.011
ours(w/o coarse data)	0.0099	0.0067	0.0029	0.0095
ours(w/o QUN)	0.0076	0.0042	0.0024	0.0072
ours	**0.0058**	**0.0026**	**0.0016**	**0.0054**

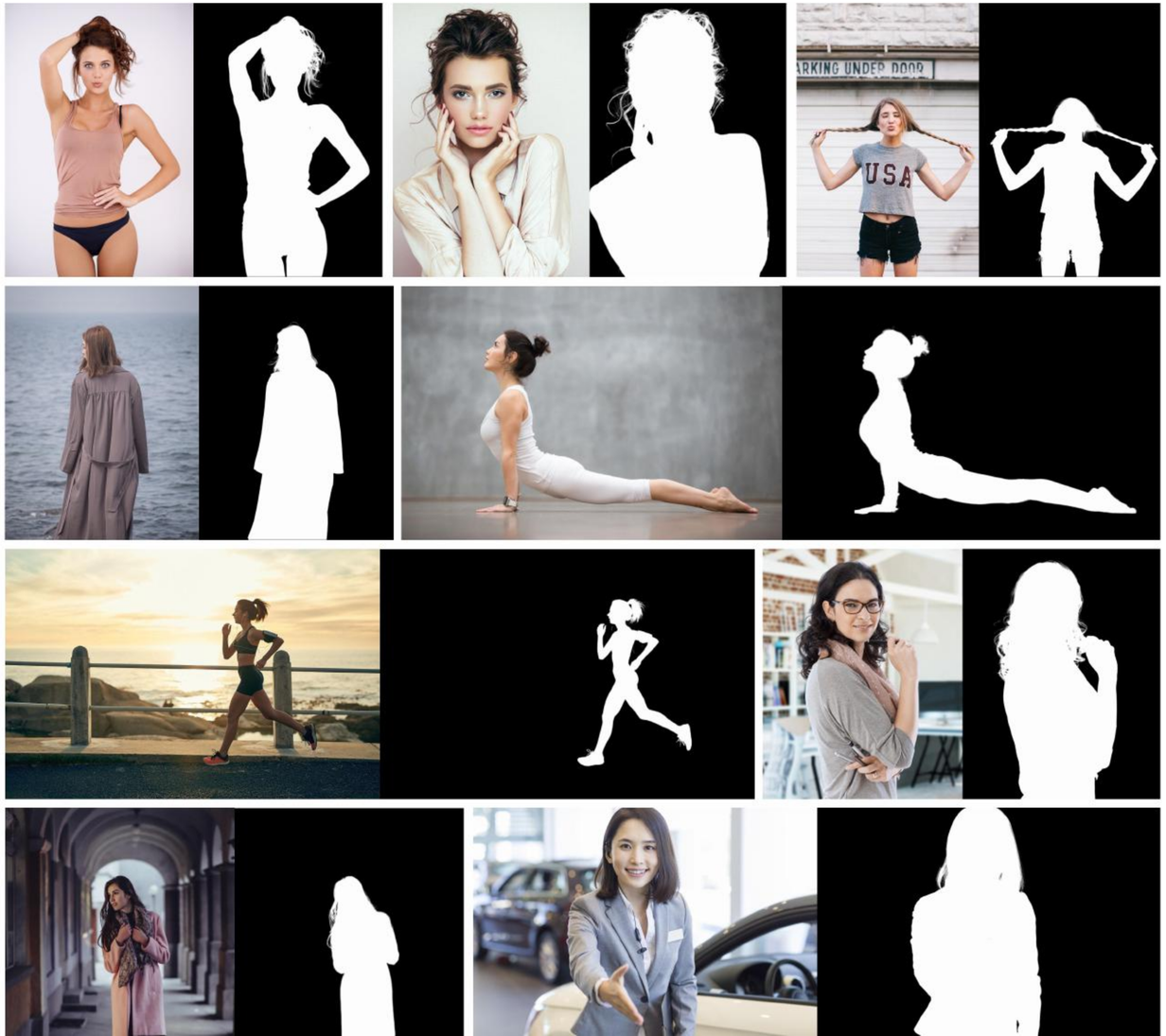
Ablation study

ours(w/o coarse data)	0.0099	0.0067	0.0029	0.0095
ours(w/o QUN)	0.0076	0.0042	0.0024	0.0072
ours	**0.0058**	**0.0026**	**0.0016**	**0.0054**



(a) Input image (b) Predicted coarse mask (c) Estimated alpha matte without QUN (d) Estimated alpha matte with QUN

Real image results



Overview

- Motivation
- Method
- Experiment
- Application
- Conclusion

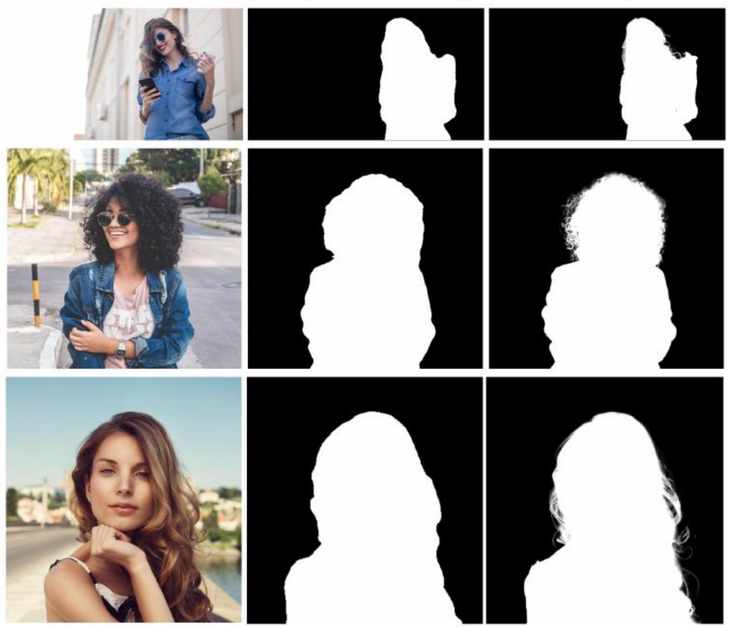
Application

Refine input coarse masks from public dataset or semantic segmentation methods

From Pascal

From COCO

From Deeplab



Overview

- Motivation
- Method
- Experiment
- Application
- Conclusion

Conclusion

Contributions

- The first to use coarse dataset to boost accurate human matting.
- We propose a quality unification network to rectify the mask quality so as to utilize both coarse and fine dataset.
- Our method can be used to refine coarse annotated public dataset as well as semantic segmentation methods easily.

Thanks
Q&A