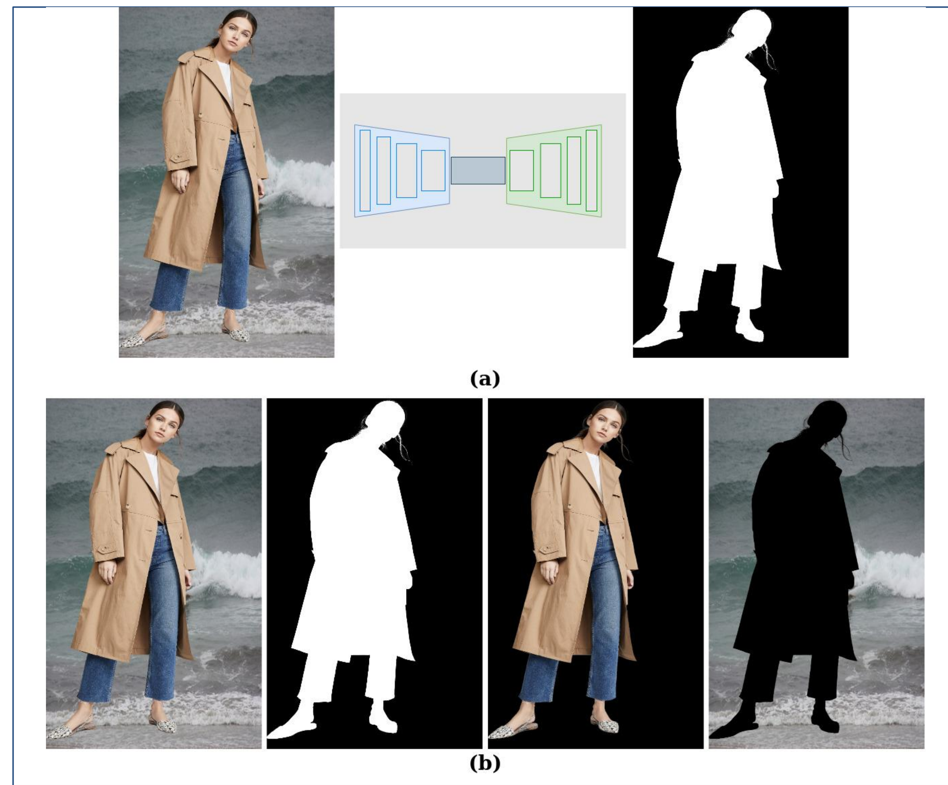
Alpha Matte Generation from Single Input for Portrait Matting

Dogucan Yaman[1], Hazım Kemal Ekenel[2], Alexander Waibel[1,3]
[1]Karlsruhe Institute of Technology, [2]Istanbul Technical University, [3]Carnegie Mellon University
{dogucan.yaman, alexander.waibel}@kit.edu, ekenel@itu.edu.tr

Motivation



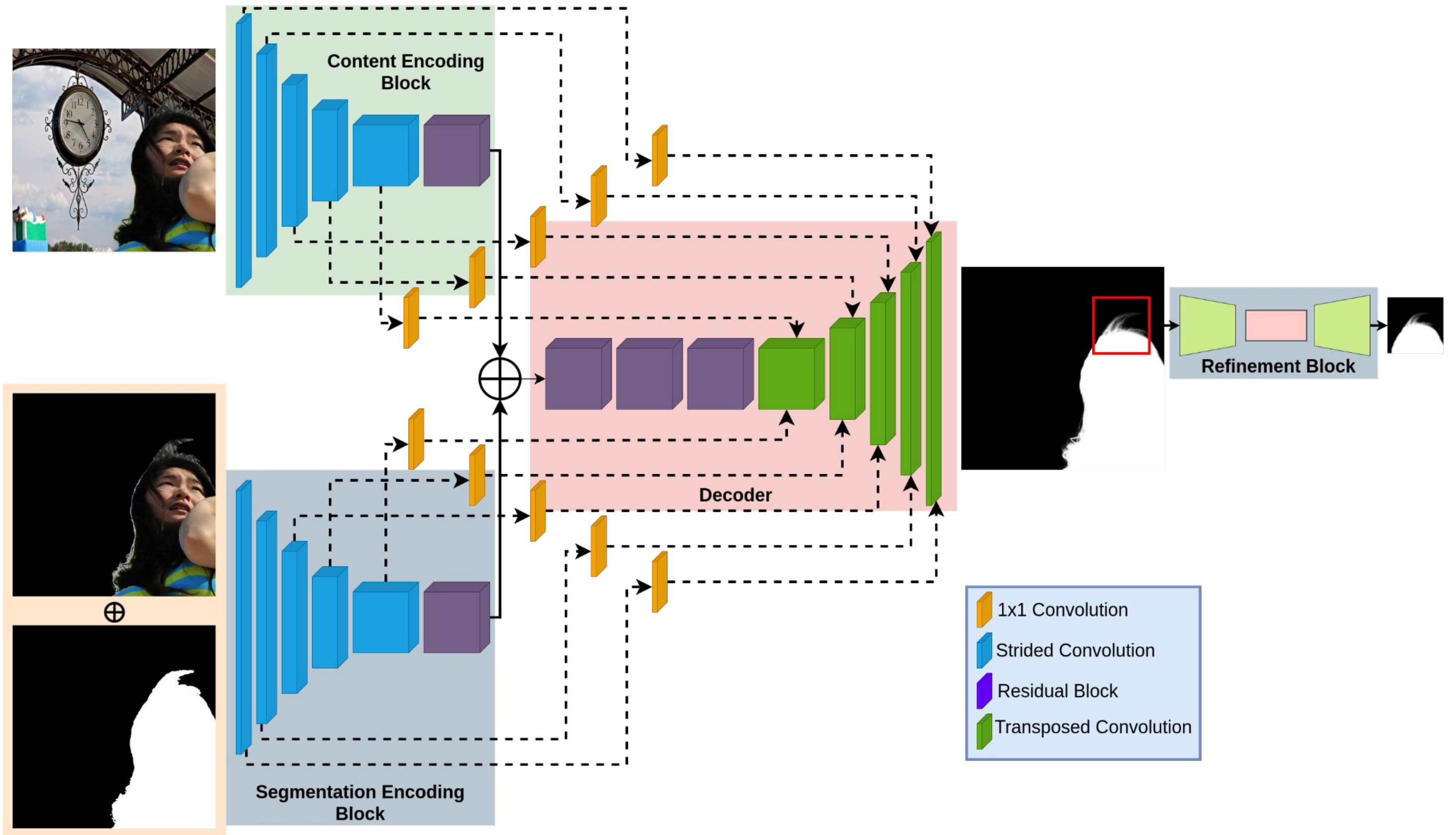
- Distinguish background and foreground subject by predicting alpha matte
- **Applications:** Image/video editing, background modification, video/movie post-production
- **Task definition:** Generate alpha matte for the subject.

$$I_i = \alpha_i F_i + (1-\alpha_i)\, B_i$$

Proposed Model

- **Content Encoding Block:** Encode the content of the image.
- **Segmentation Encoding Block:** Encode the **extracted subject** and **segmentation map**. Provides better feature representation.
- **Refinement Block:** Enhance the details of the alpha matte.
- **Segmentation map:** Obtain with a pretrained person segmentation model.
- **Extracted subject:** Obtain with the predicted segmentation map.
- Split standard alpha loss into two different losses to penalize separately ad adjust their effects..
  - **Alpha loss:** Calculate for only pixels that have one or zero values.
  - **Alpha coefficient loss:** Only use pixels that have neither zero nor one values.
- **Border loss:** Penalize only the area around the border of the subject.

$$L = L_{GAN}(G,D) + \lambda L_{percep}(G) + BL_{alpha}(G) + \gamma L_{border}(G) + \theta L_{ac}(G)$$

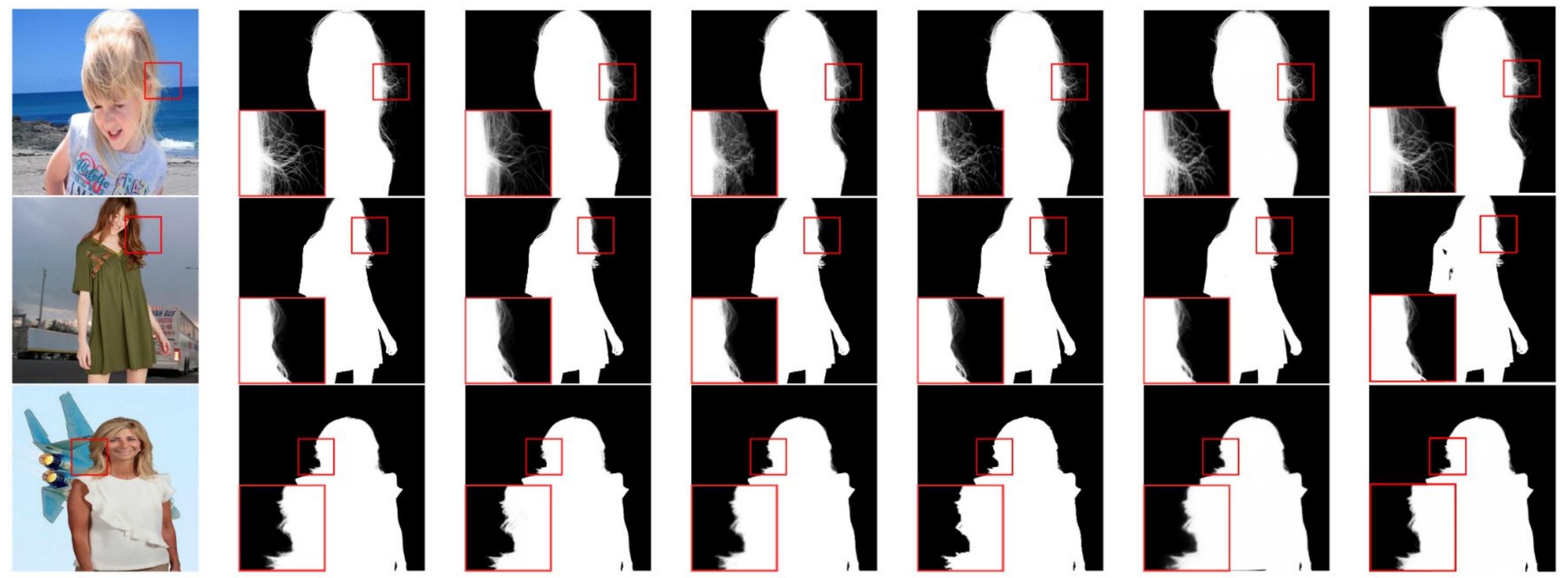


Experimental Results

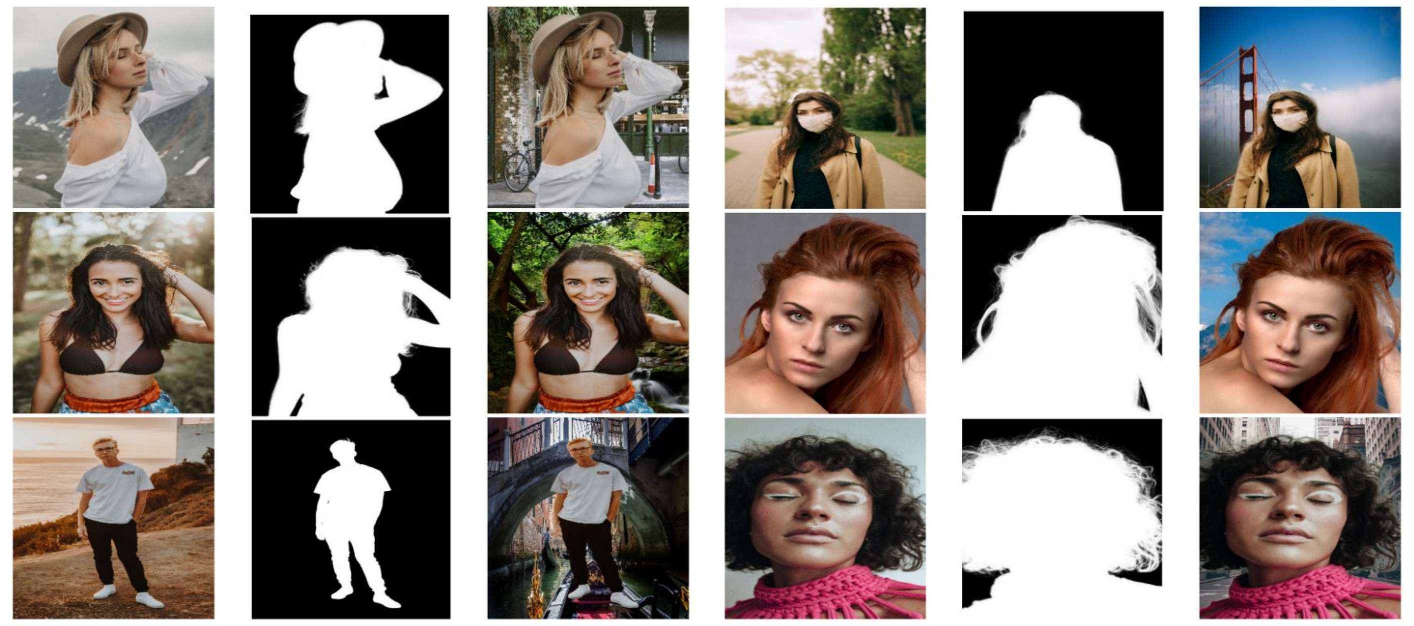
Method	Extra Input	Dataset	MSE	SAD	Grad	Conn
BGM-V2 [1]	Background	AIM	2.12	9.04	8.32	9.21
FBA [2]	Trimap	AIM	0.40	3.98	1.19	3.11
MODNet [3]	-	AIM	21.65	33.93	44.24	35.45
MGM [4]	-	AIM	1.48	6.21	4.74	6.55
Ours	**-**	**AIM**	**1.06**	**5.04**	**4.22**	**5.39**
BGM-V2 [1]	Background	PM85	0.37	1.45	1.28	2.38
FBA [2]	Trimap	PM85	1.01	2.55	3.50	2.75
MODNet [3]	-	PM85	2.32	7.23	12.17	9.48
MGM [4]	-	PM85	0.38	2.91	1.32	2.04
Ours	**-**	**PM85**	**0.19**	**1.19**	**0.65**	**1.16**
BGM-V2 [1]	Background	D646	0.98	4.83	3.78	5.30
FBA [2]	Trimap	D646	0.44	3.25	1.70	2.38
MODNet [3]	-	D646	3.51	10.27	13.54	18.98
MGM [4]	-	D646	0.88	5.42	3.40	4.76
Ours	**-**	**D646**	**0.71**	**3.99**	**2.74**	**3.84**
FBA [2]	Background	PPM100	0.96	2.41	4.20	2.70
MODNet [3]	Trimap	PPM100	4.60	11.59	12.48	22.16
MGM [4]	-	PPM100	1.15	5.31	5.04	5.29
Ours	**-**	**PPM100**	**0.84**	**4.70**	**3.67**	**4.46**

Datasets: AIM, PM85, D646, PPM100

- **Training:**
  - Combine AIM and D646 datasets.
  - In total, 564 subjects.
  - Use 100 background images from MSCOCO dataset for each subject.
  - 56400 training images in total
- **Test:**
  - AIM test set, PM85, D646 test set, PPM100
  - Combine each subject with 20 different background images from PASCAL VOC dataset.



Input GT Ours BGM-V2 FBA MODNet MGM



Input Alpha Combined Input Alpha Combined

Ablation Studies

- Using combination of segmentation and foreground as an input to the segmentation encoding block improves the result.
- Using combination alpha matte and foreground as an input to the discriminator makes the training more stable and accurate.
- SE block and refinement network provide more accurate alpha matte prediction.
- Each loss function allows us to obtain better performance by enhancing the predicted alpha matte.

Cases	MSE
Segmentation map	1.86
Segmentation + Foreground	1.41
Alpha matte + Foreground	1.06

Cases	MSE
Base model	2.20
Base model + SE block	1.57
Base model + SE block + refinement network	1.06

Loss	MSE
$L_{cGAN} + L_{alpha}$	7.24
$L_{cGAN} + L_{per} + L_{alpha}$	3.78
$L_{cGAN} + L_{per} + L_{alpha} + L_{border}$	1.76
$L_{cGAN} + L_{per} + L_{alpha} + L_{border} + L_{ac}$	1.06
α	3.14
α, F	1.06

Conclusions

- Proposed conditional GAN-based additional input-free two-stage approach.
  - First stage: person segmentation with DeepLabV3+
  - Second stage: alpha matte prediction using the input image and predicted segmentation map
- Proposed refinement network enhanced the quality of the predicted alpha matte.
- Proposed alpha coefficient loss and border loss improved the performance of alpha matte prediction.
- Segmentation encoding block improved the performance by providing more useful feature representation to the decoder network.
- Using combination of segmentation map and extracted subject with it as an input to the segmentation encoding block increased the performance.

Acknowledgement

The project on which this report is based was funded by the Federal Ministry of Education and Research (BMBF) of Germany under the number 01IS18040A.

References

[1] Shanchuan Lin, et al. Real-time high-resolution background matting. CVPR 2021.
[2] Marco Forte and François Pitie. f , b, alpha matting. arXiv preprint arXiv:2003.07711, 2020
[3] Zhanghan Ke, et al. Is a green screen really necessary for real-time human matting? arXiv preprint arXiv:2011.11961, 2020
[4] Qihang Yu, et al. Mask guided matting via progressive refinement network. CVPR 2021