



# Deep Automatic Portrait Matting

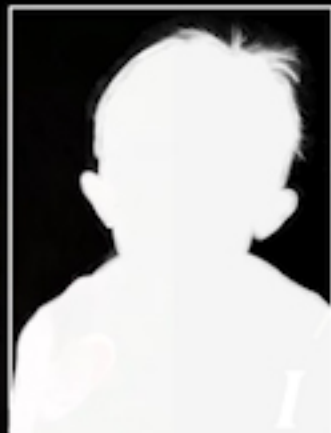
Xiaoyong Shen, Xin Tao, Hongyun Gao, Chao Zhou, Jiaya Jia

The Chinese University of Hong Kong

# Portrait Matting



Input Image



Alpha Matte



Stylization



Cartoon



Color Transform



Depth-of-field



Portrait



Background Edit

# Matting Problem



$$I = \alpha F + (1 - \alpha) B$$

Diagram illustrating the matting equation  $I = \alpha F + (1 - \alpha) B$ . The equation is centered, with  $\alpha$  in yellow. Green arrows point from the equation to the labels: 'Image' (top left), 'Alpha Matte' (top right), 'Foreground' (bottom left), and 'Background' (bottom right).

Ill-posed problem

--seven unknowns should be estimated for each pixel.

# Image Matting

- User interactions are needed



Input



Strokes



Trimap

# Issues

- User specified strokes or trimap are difficult to meet the algorithm requirements



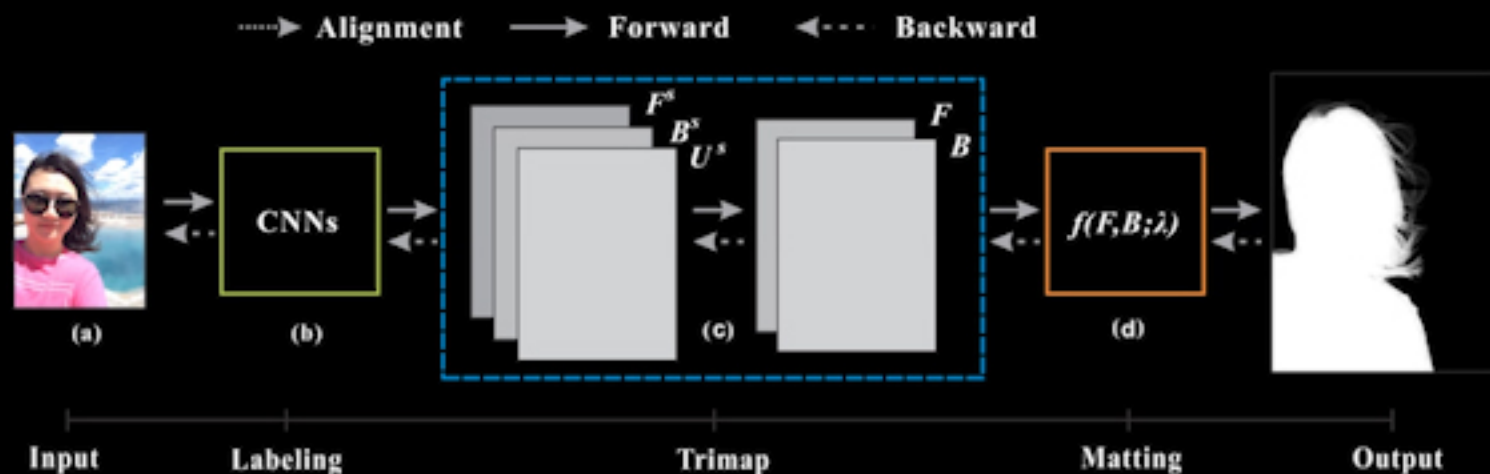


**Tedious interaction is involved to produce these trimaps.**

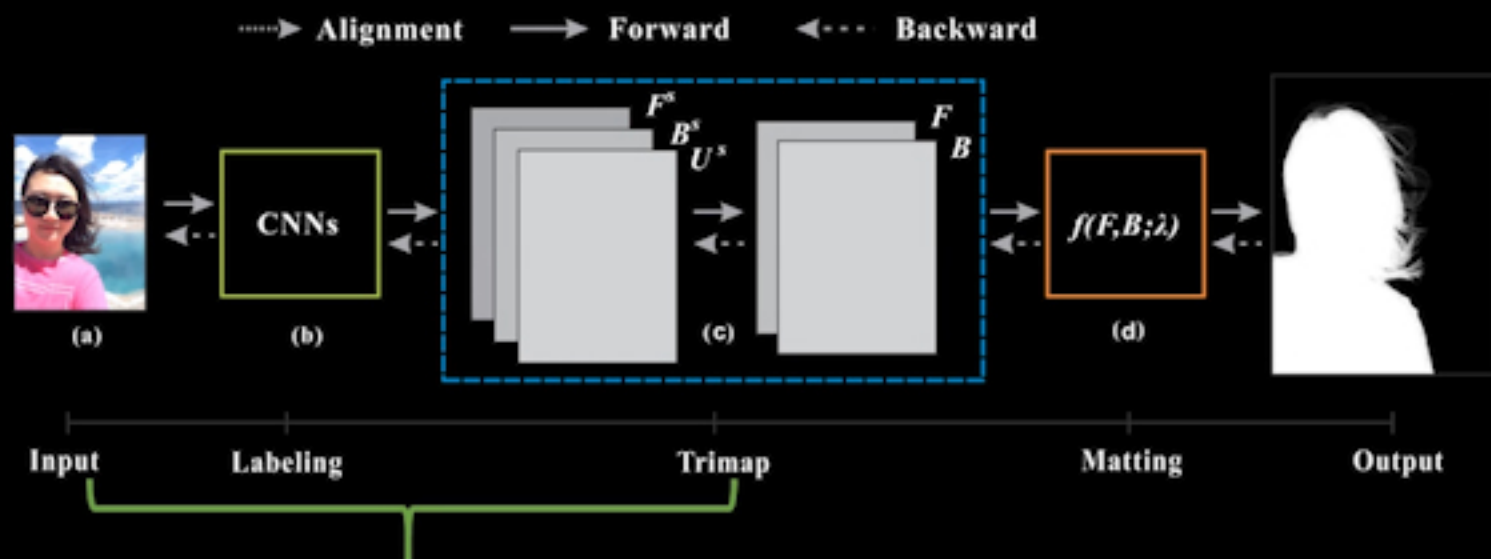
# Deep Automatic Portrait Matting

End-to-end CNNs

# Deep Automatic Matting

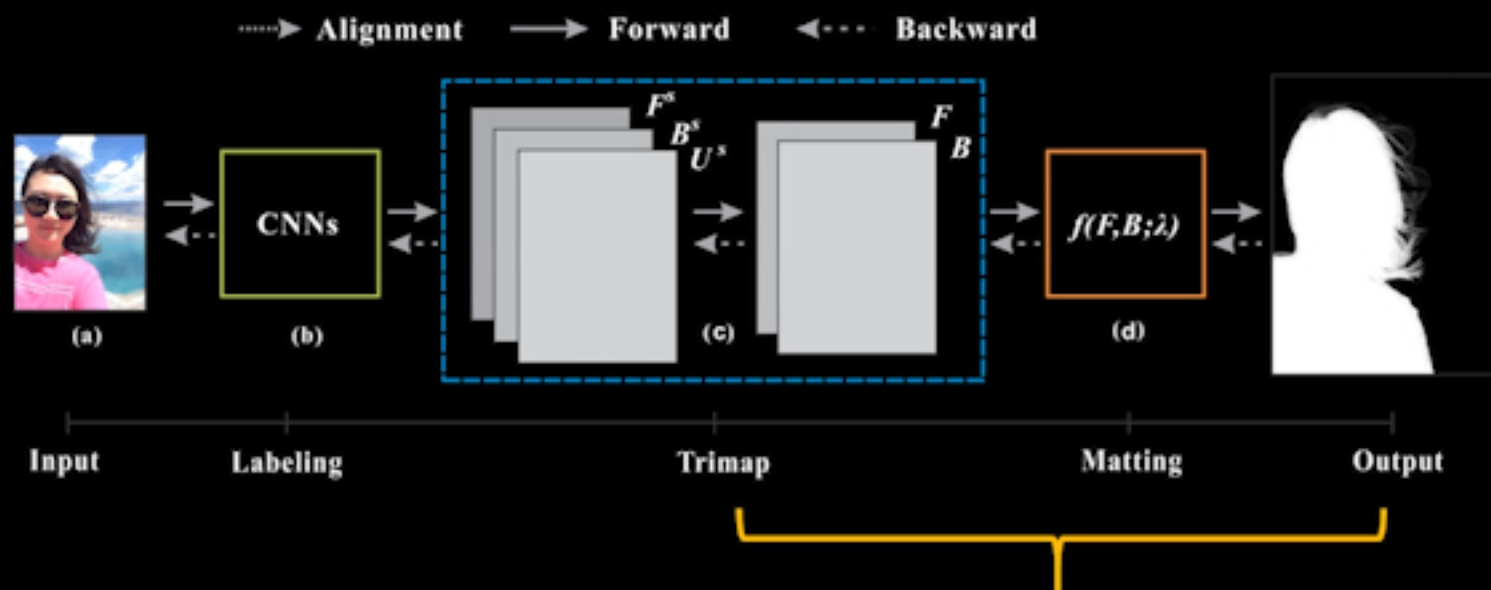






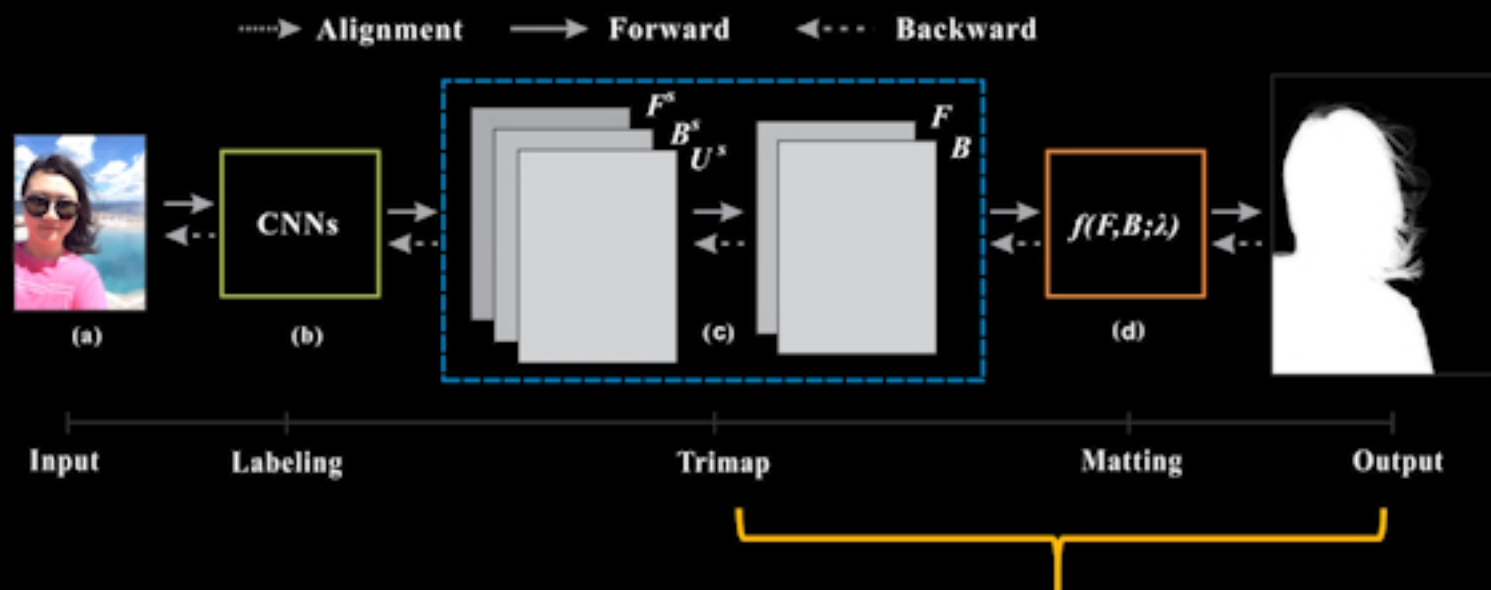
## Trimap Labeling

- Input: RGB image
- Output: trimap representation
- Network: FCN [Long et al. 2015]



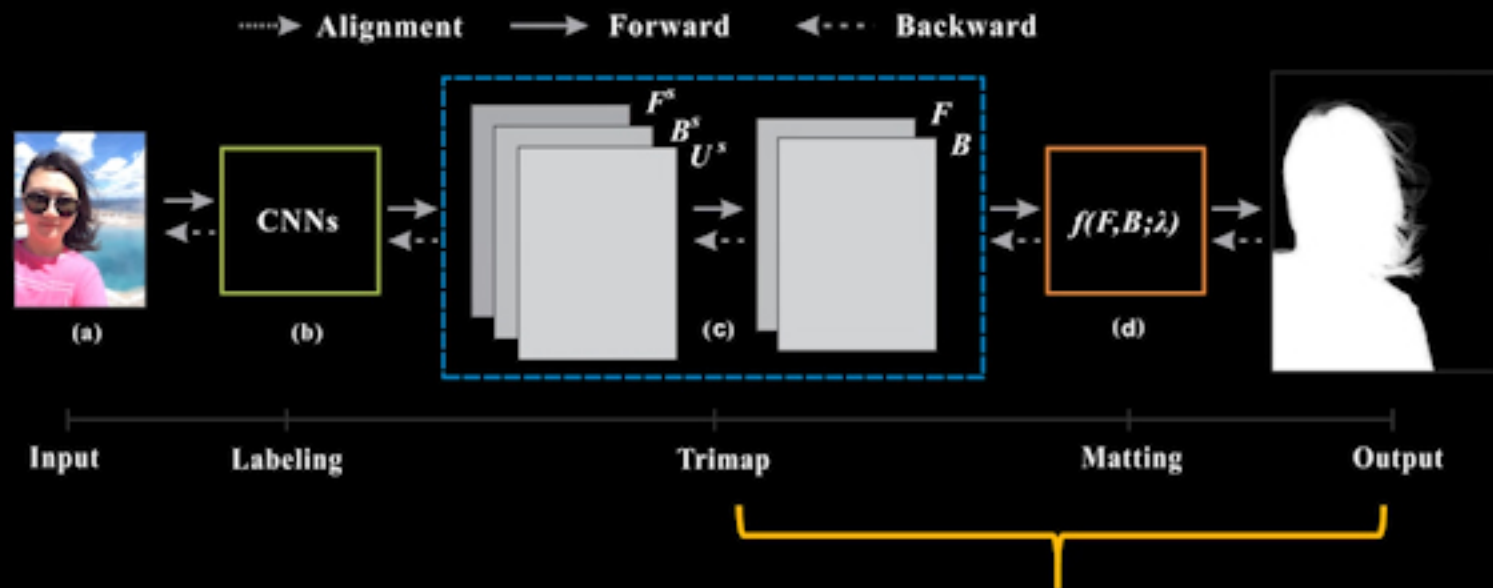
## Image Matting Layer

- Input: trimap representation
- Output: alpha matte



## Image Matting Layer

- Input: trimap representation
- Output: alpha matte
- **Newly-designed layers**



## Image Matting Layer

- Feed-Forward

$$\min \lambda A^T B A + \lambda (A - 1)^T F (A - 1) + A^T L A$$

- Back-Forward

$$\frac{\partial f}{\partial B} = -\lambda D^{-1} \text{diag}(D^{-1} F)$$

$$\frac{\partial f}{\partial F} = \frac{\partial f}{\partial B} + D^{-1}$$

$$\frac{\partial f}{\partial \lambda} = -\lambda D^{-1} \text{diag}(F + B) D^{-1} F$$

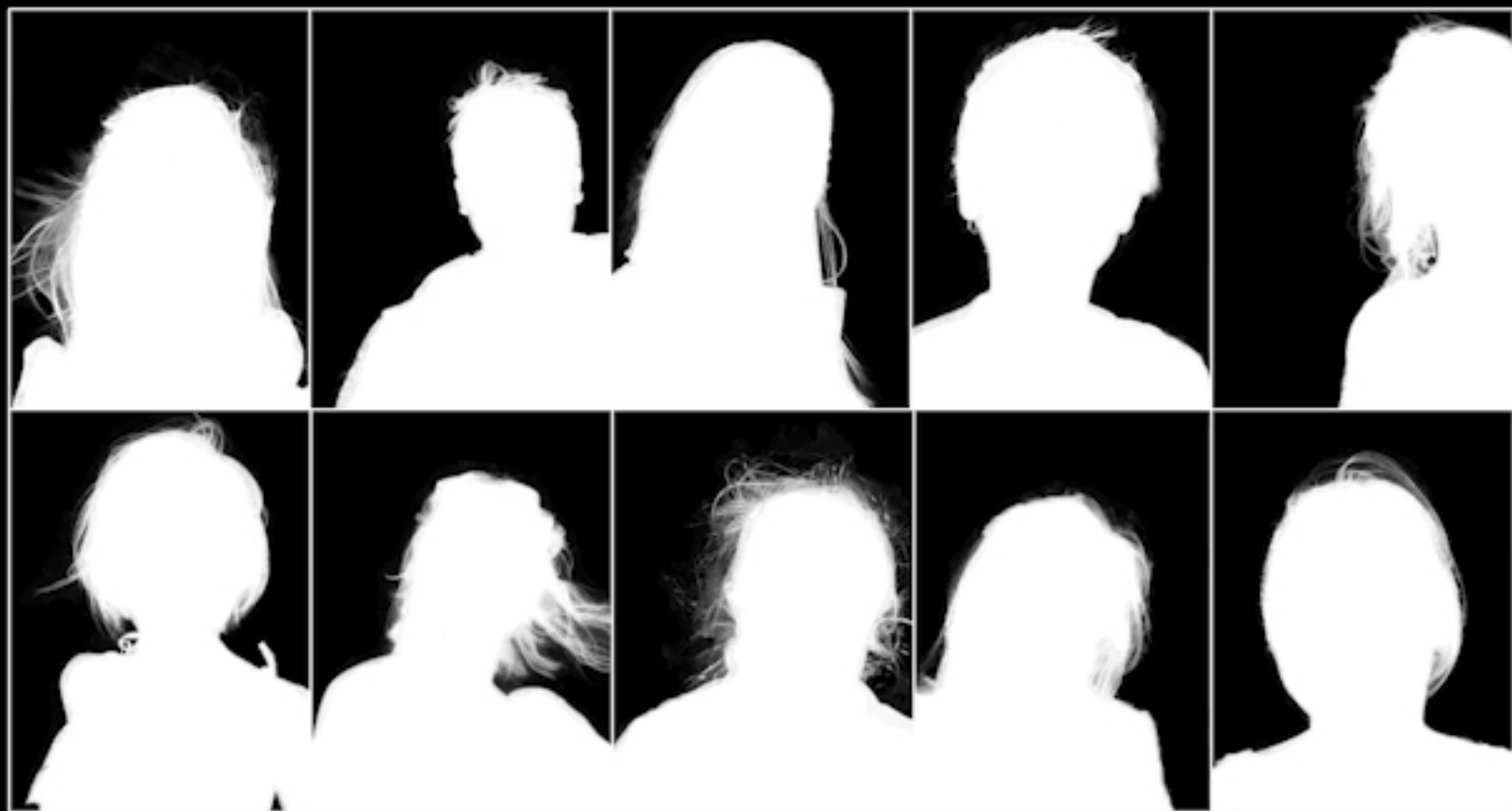
# Learning Data Collection

- We create a **2,000** portraits dataset for training and testing
  - **1,700** for training and **300** for testing
  - Large **variations** in age, gender, pose, hairstyle, background, camera type, etc.
- The matting ground truth is estimated by human well labeled trimap

# Data Examples



# Labeled Mattes



# Experiments

- Running Time

- Training: 20k iterations, one day on Titan X GPU
- Testing: 0.6s for 600×800 color image

- Comparisons

- Automatic segmentation to trimap approaches
- Direct trimap labeling methods



# Evaluation

Methods	Grad. Error ( $\times 10^{-3}$ )	Conn. Error ( $\times 10^{-4}$ )
Graph-cut Trimap	4.93	7.73
AutoTrimap	4.61	7.63
Trimap by FCN	4.14	7.61
Trimap by DeepLab	3.91	7.52
Trimap by CRFasRNN	3.56	7.39
Ours without Shape Mask	3.11	6.99
<b>Ours</b>	<b>3.03</b>	<b>6.90</b>

Input



## Graph-cut Trimap



## FCN Trimap



Ours



Input



## Graph-cut Trimap



## FCN Trimap





Ours



## More Results



## More Results



## More Results



## More Results



# Conclusion

- We proposed the deep automatic portrait matting
  - An **end-to-end** matting CNNs framework
  - Novel **matting layer**
  - A matting dataset with **2,000** portraits
- Future work
  - Video portrait matting
  - Person matting
  - General object matting

**Thanks**