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

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

- Image
- Alpha Matte
- Foreground
- Background

Ill-posed problem
--seven unknowns should be estimated for each pixel.
Image Matting

- User interactions are needed
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

(a) Input
(b) Labeling
(c) Trimap
(d) Matting

Alignment → Forward → Backward
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
  \[ \min \lambda A^TBA + \lambda (A - 1)^TF(A - 1) + A^TLA \]
- 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
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

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