A World Leading SFI Research Centre

Engaging Content
Engaging People

Marco Forte, François Pitié, Sigmedia

# Using deep learning to bypass the green screen

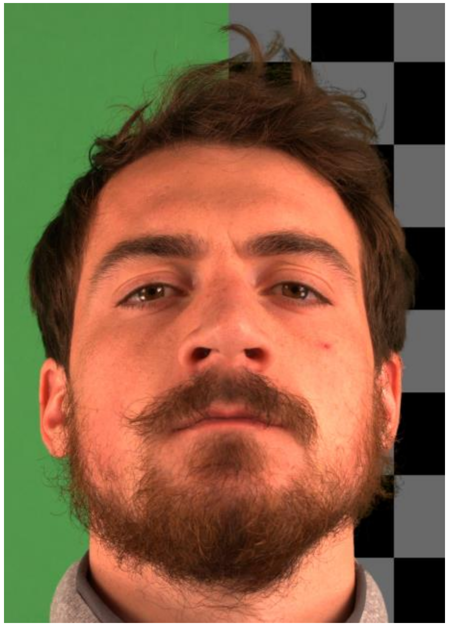
The ADAPT Centre is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

# Greenscreen keying



Green Screens are used by film and television industry for background replacement.

High quality results still take a lot of artist time, though lower quality is achievable in real time.



# Natural Image matting



Alpha matting refers to the problem of extracting the opacity/transparency mask, alpha matte, of an object in an image.

The goal being to compose the object onto a new background.

Image of object

Define unknown regions

Alpha matte

# Compositing



To composite an object onto a novel background we need

- Object foreground
- Alpha matte
- background image

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

Object foreground

Alpha matte

Background on which to composite

(10 pts if you recognise this place)

# Other applications



(a) Input (b) Segmentation (c) Stylization (d) Depth-of-field (e) Cartoon

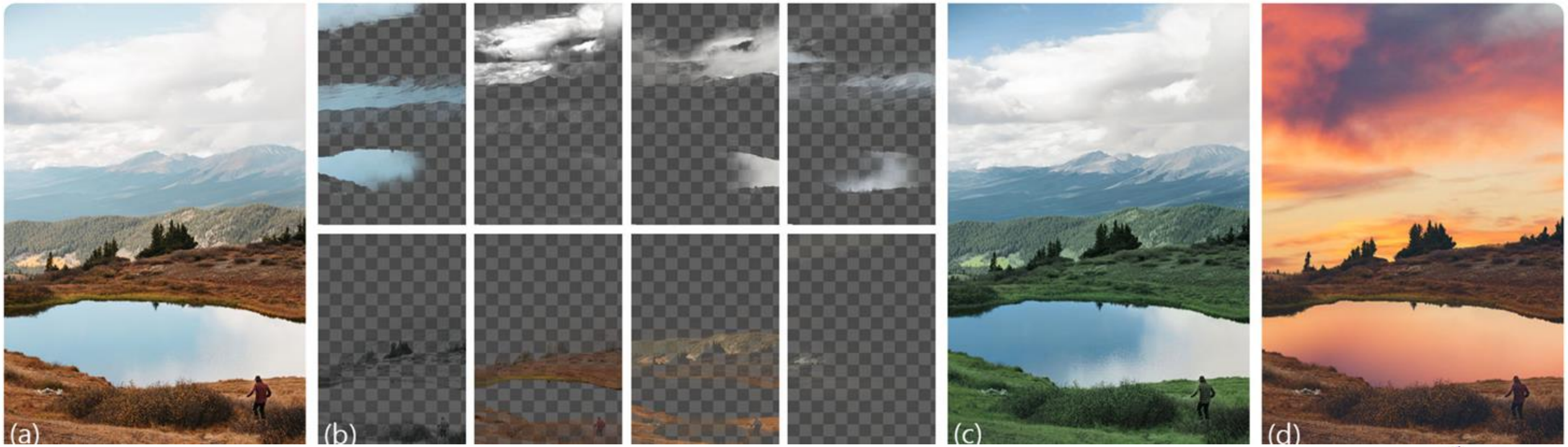
**Automatic Portrait Segmentation for Image Stylization**

Xiaoyong Shen[1] Aaron Hertzmann[2] Jiaya Jia[1] Sylvain Paris[2] Brian Price[2] Eli Shechtman[2] Ian Sachs[2]
[1]The Chinese Univeristy of Hong Kong [2]Adobe Research

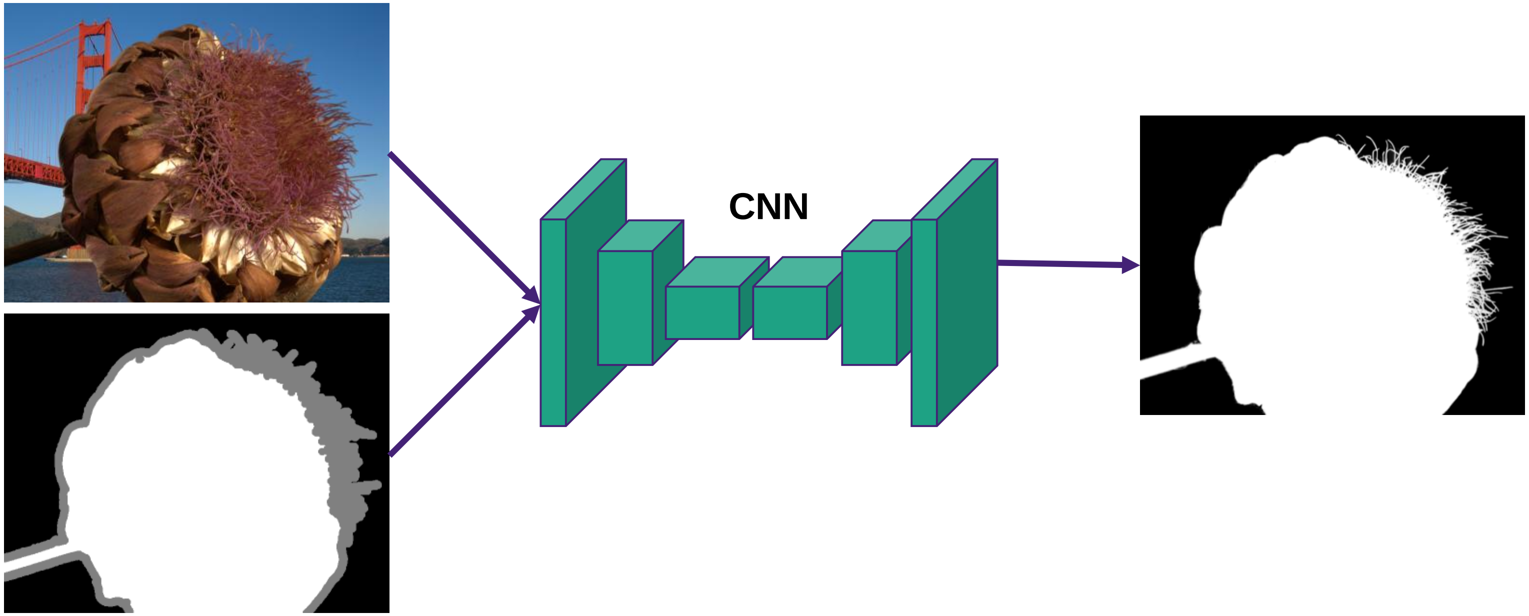
# Unmixing-Based Soft Color Segmentation for Image Manipulation

Yagiz Aksoy, Tunc Ozan Aydin, Aljoscha Smolic and Marc Pollefeys

*ACM Transactions on Graphics (TOG), 2017*



# Image matting with CNNs



# Training procedure - Dataset

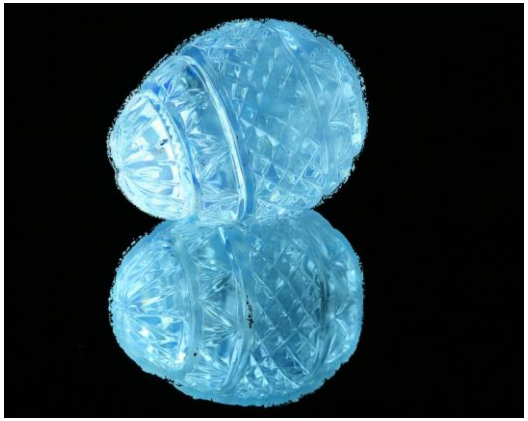


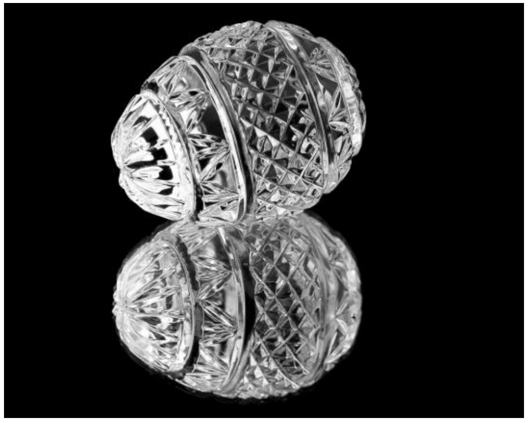
Time consuming to create manually.

Highest quality needs still object to be captured in front of monitor with changing backgrounds.

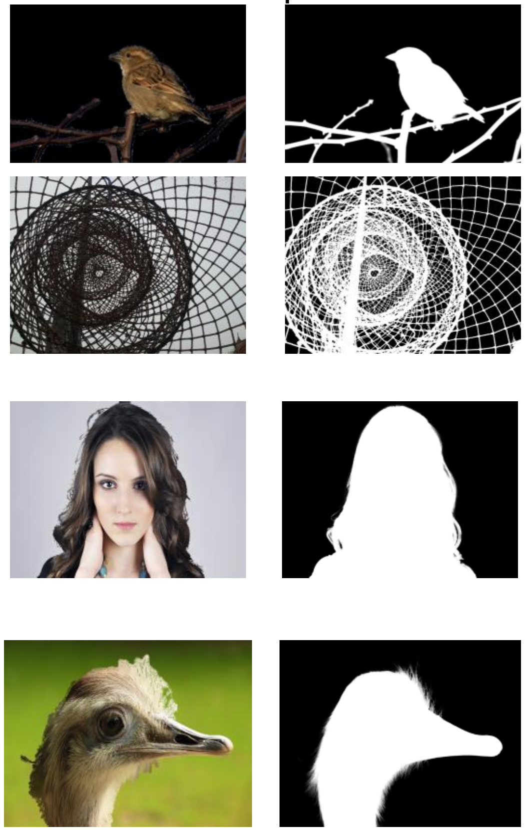
Otherwise can manually annotate existing images with clean backgrounds in photoshop.

Greenscreen also possible in controlled HD or UHD environment.





We a created dataset of 500 foreground and alpha pairs. Adobe created one of 450 pairs.

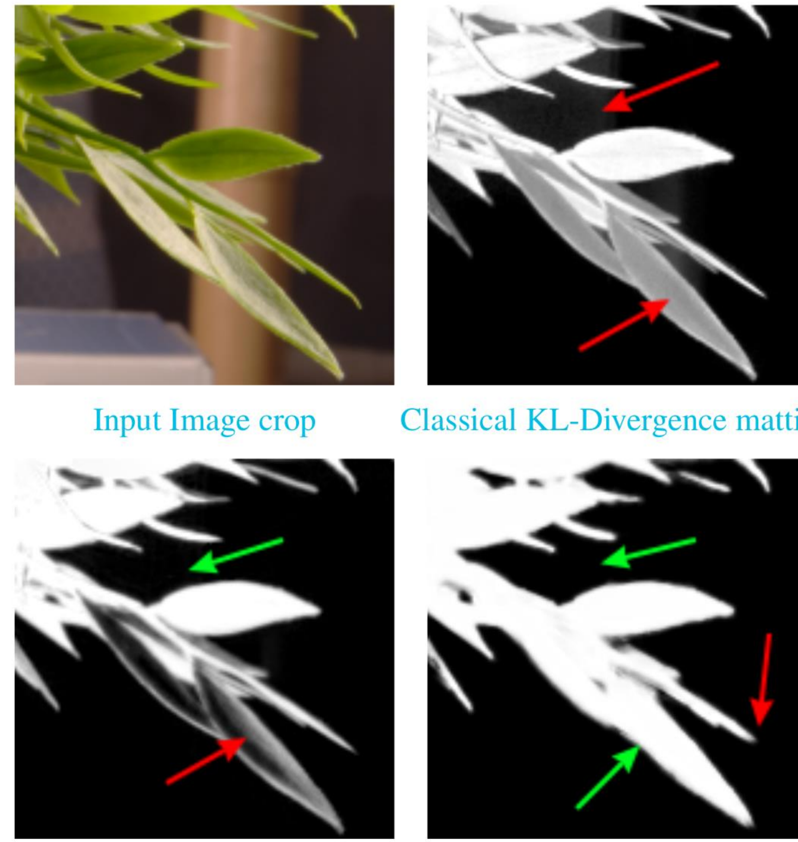


# Properties of using CNNs for image matting



- Existing method from Adobe is top ranking yet uses very large difficult to optimise network.
- Performs something more akin to segmentation rather than alpha matting
- Mathematics of alpha matting requires some matrix inversion which is difficult to learn with standard conv layers structure.

**Properties of CNN based methods**



Input Image crop — Classical KL-Divergence matting

Hybrid deep alpha refinement — Deep image matting

# Training procedure - Dataset



Dataset is small only ~450-1000 images.

Lots of data augmentation needed.

Composite the foreground onto 1000s of different backgrounds

Random cropping of different size.

Crop rotation and mirroring.

Slight changes to foreground contrast and brightness

# Wait actually we need that greenscreen....



1. We could get thousands of ground truth frames by using a greenscreen. :)

1. High quality keying is actually non-trivial :/

1. Artists don’t have a very scientific approach, they use a mismatch of keys with different settings.

1. To get really high quality ground truth we’ll need really high quality cameras

1. Automatic tools kinda suck, need to make our own

# Greenscreen setup




# Automatic methods aren’t good enough



Automatic methods don’t provide good ground truth data.

They may look ok, but there’s alot of detail lost, noise introduced and color spill not fully removed

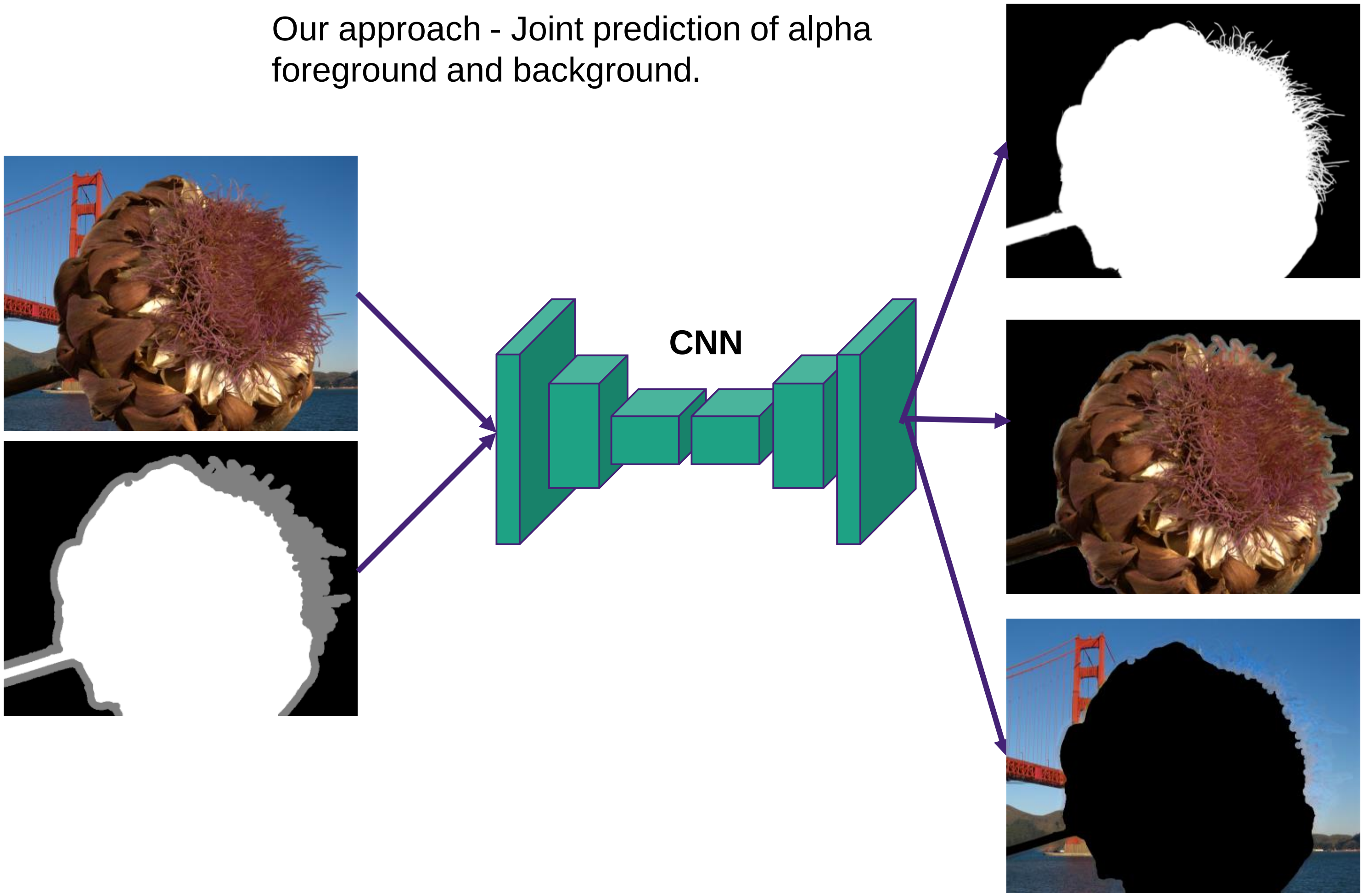

# Alpha matte



# Image matting with CNNs

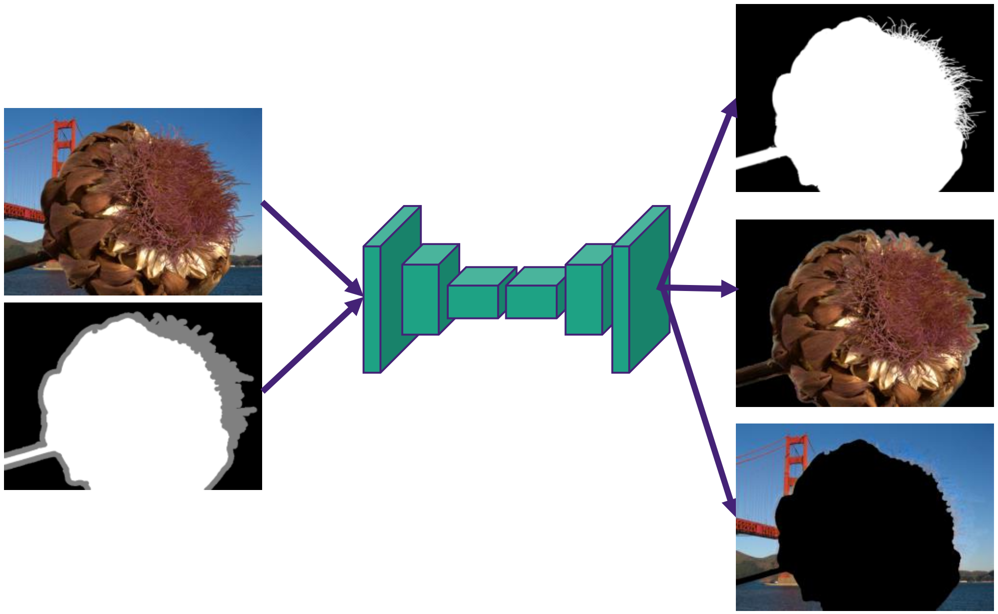


Our approach - Joint prediction of alpha foreground and background.



# Image matting with CNNs





Alpha Loss = $\sum(\boldsymbol{\alpha} - \boldsymbol{\alpha}_{gt})$

Foreground loss = $\sum\sum(Fg - Fg_{gt})$

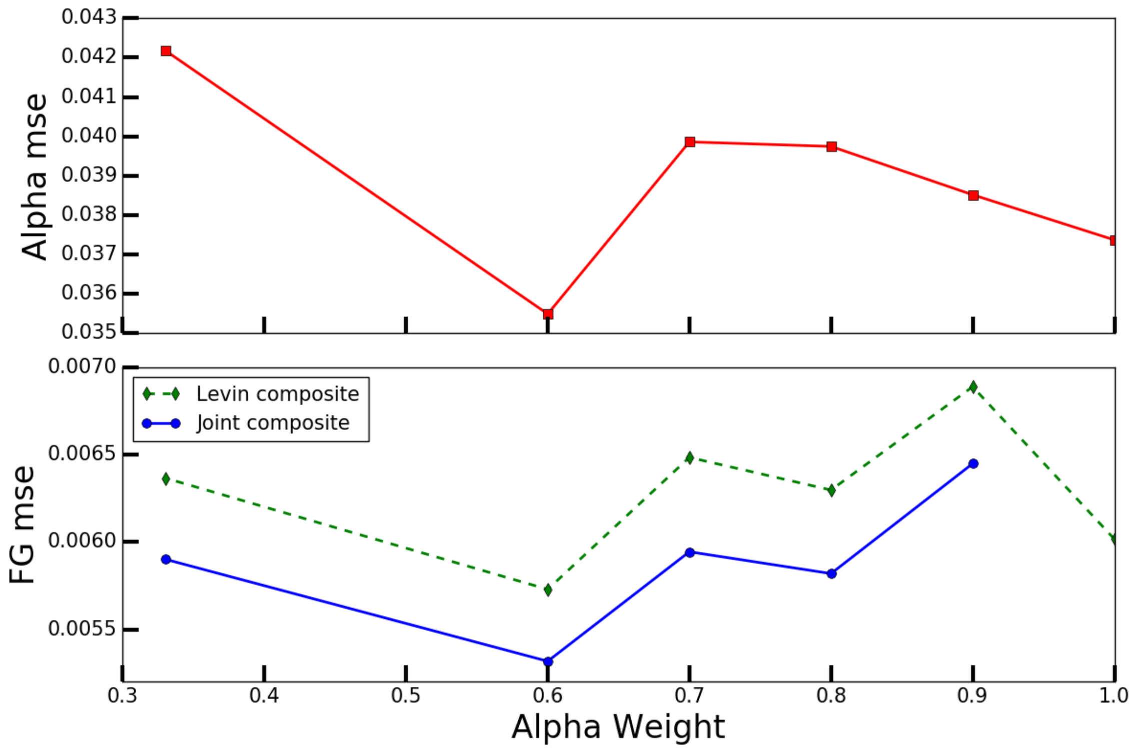
Background loss = $\sum\sum(Bg - Bg_{gt})$

Only define losses on well defined regions

$$Loss = ƛL_{\alpha} + (1-ƛ)(L_{Fg} + L_{Bg})$$

# Benefits of modelling foreground and bg





# Benefits of modelling foreground and bg



Alpha matting performance

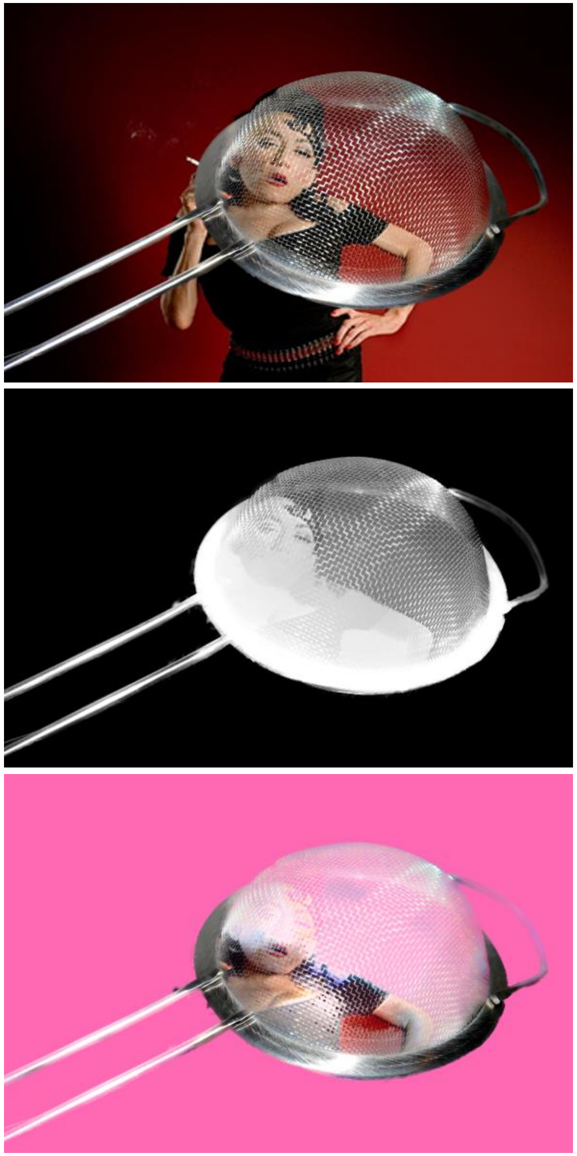
| Method | Composition 1k [6] | | |
|---|---|---|---|
| | **SAD** | **MSE** | **Gradient** |
| Closed-Form Matting [4] | 161.3 | 0.085 | 133.2 |
| Shared Matting [3] | 128.9 | 0.078 | 132.4 |
| KNN Matting [15] | 182.0 | 0.111 | 136.2 |
| IFM Matting [8] | 102.8 | 0.056 | 59.36 |
| DCNN Matting [20] | 165.5 | 0.089 | 122.5 |
| Encoder-Decoder ($W_\alpha = 1.0$) | 101.0 | 0.042 | **67.1** |
| Encoder-Decoder ($W_\alpha = 0.7$) | **94.55** | **0.039** | 68.6 |
| Colorisation net ($W_\alpha = 1.0$) | 92.56 | 0.037 | 55.8 |
| Colorisation net ($W_\alpha = 0.6$) | **88.23** | **0.035** | **49.3** |

**Table 2.** Composite $\alpha F$ accuracy

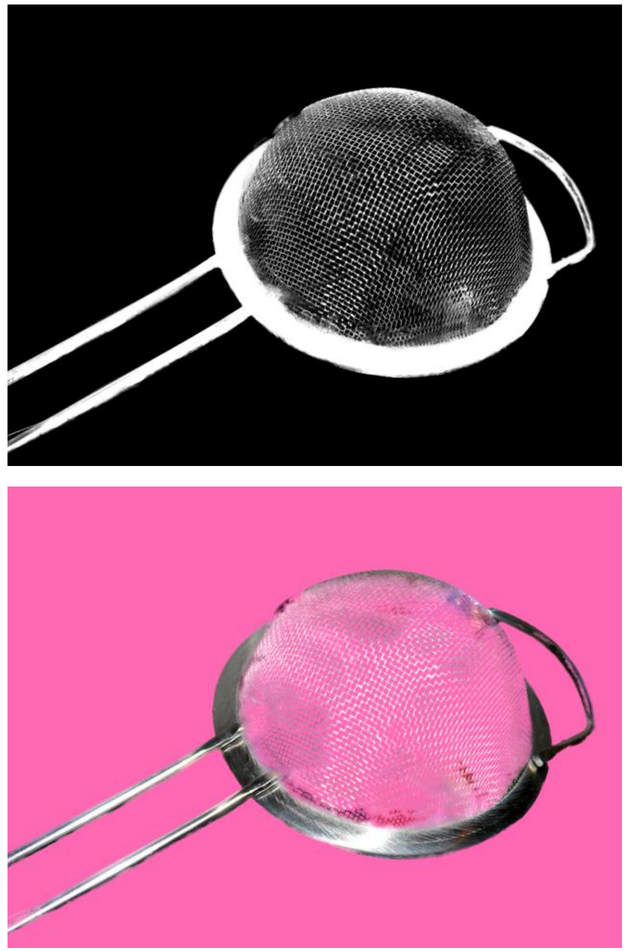
| Method | Composition 1k | | |
|---|---|---|---|
| | **SAD** | **MSE** | **SSIM** |
| Encoder-Decoder ($w_\alpha = 1.0$) | **146.9** | 0.0066 | **0.928** |
| Encoder-Decoder ($w_\alpha = 0.7$) | 160.6 | **0.0063** | 0.921 |
| Colorisation net ($w_\alpha = 1.0$) | 126.6 | 0.0060 | 0.933 |
| Colorisation net ($w_\alpha = 0.6$) | **122.2** | **0.0053** | **0.937** |

# Benefits of modelling foreground and bg

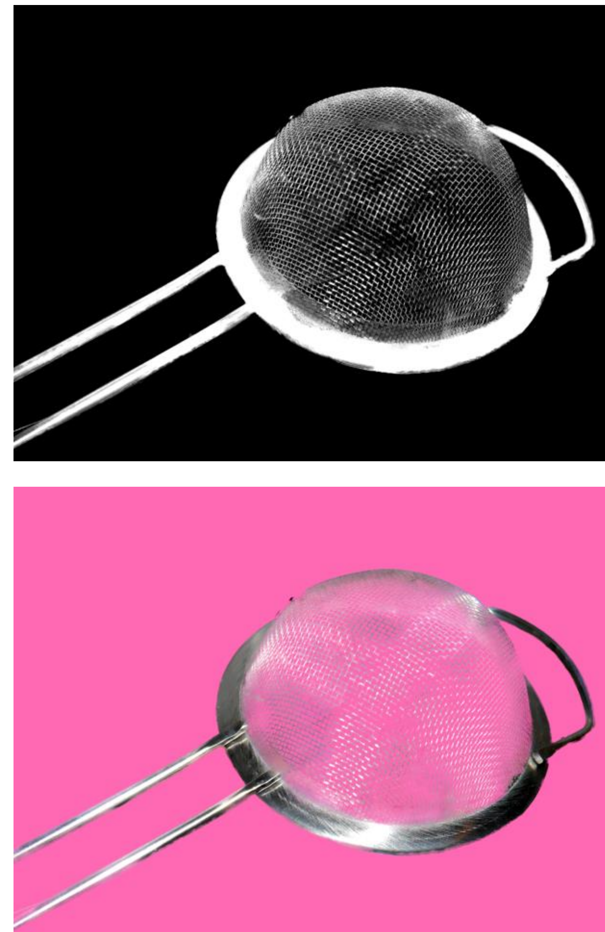




**Direct alpha prediction**



**Joint prediction**



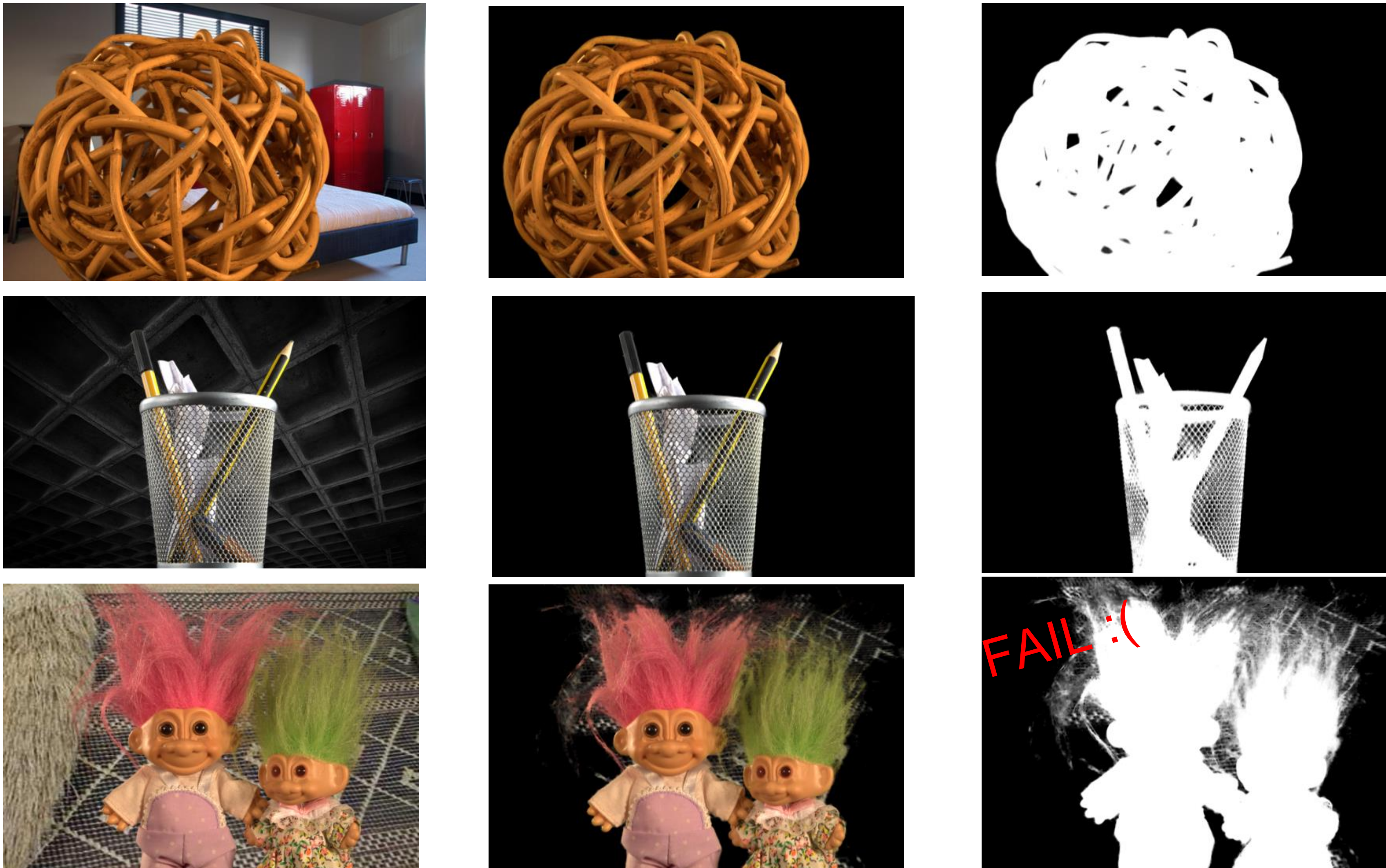
# Benefits of modelling foreground and bg



**Direct alpha prediction** | **Joint prediction**

# Some example results of our network





# Video examples



# Video examples



# Lessons learned



1. High quality training data is extremely important

1. Pretrained encoder network essential, helps in all aspects, not just coarse segmentation but also fine details.
   Resnet > Vgg

1. Multitask learning is beneficial

1. Patience when training deep networks, reproducing another paper took 3 weeks of training time to converge.

1. Deep learning can fail with images that classical algorithms have no problems with.

# Benefits moving forward



- More loss functions possible
  - Impose constraints on foreground and background when specifically training for keying
  - General image inpainting loss
  - Impose independence of foreground and background
  - Adversarial triplet loss on Fg, Bg, A
  - Adversarial fg, bg, alpha reconstruction loss
- More practical for artist to work directly with both alpha and foreground
- Generalises to video better, for example in situations with stationary background or stationary foreground