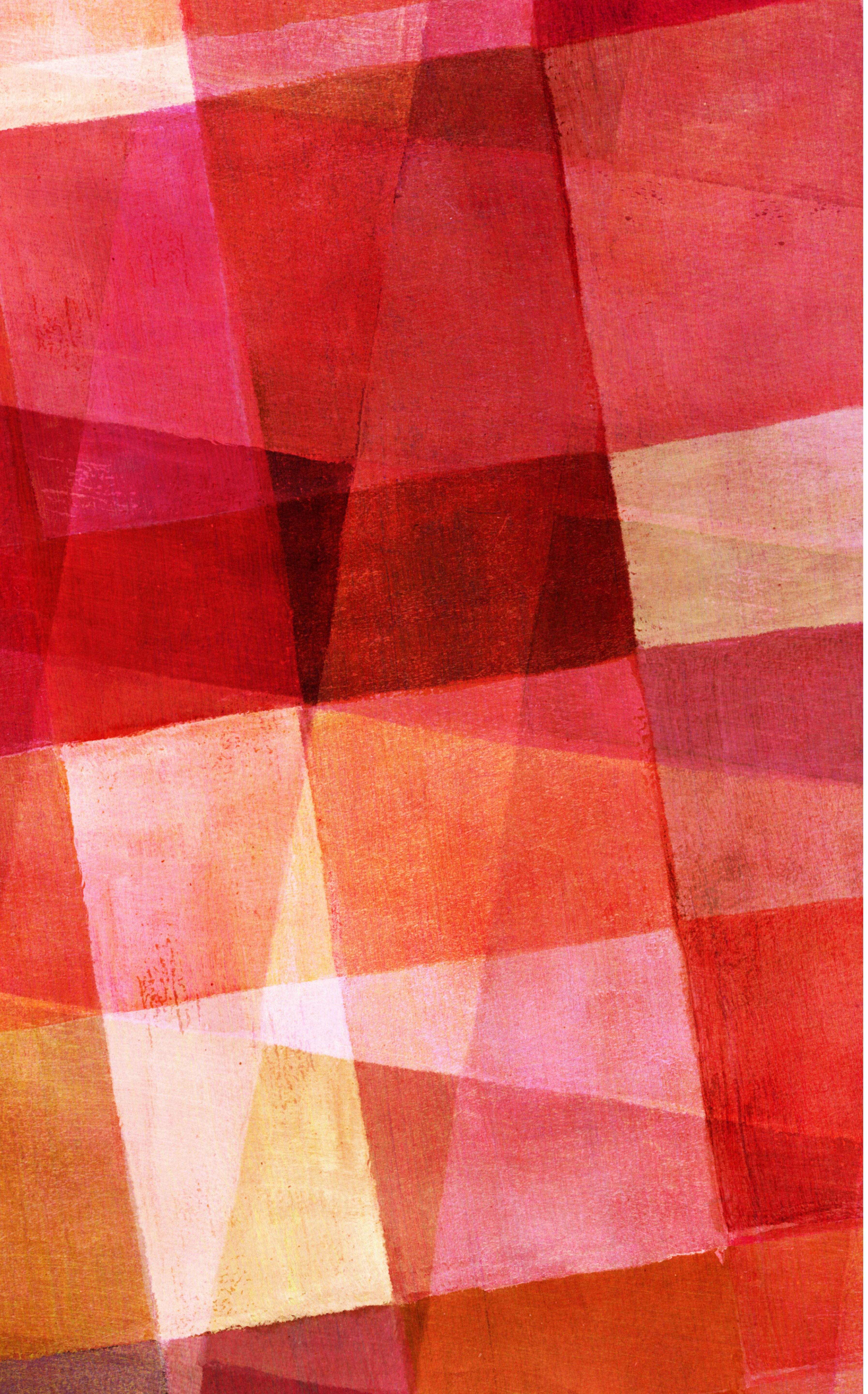
PORTRAIT SEGMENTATION

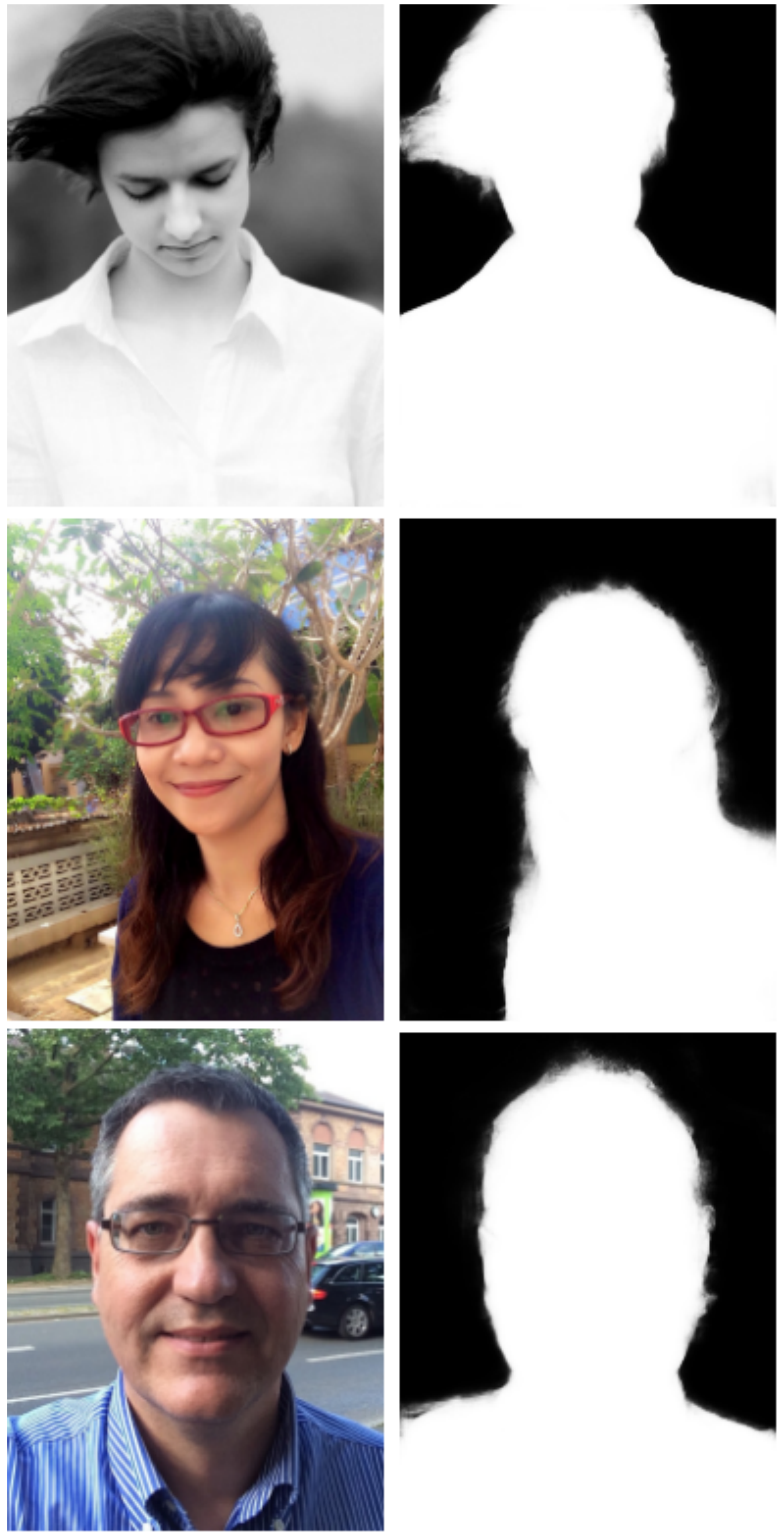
Yvonne Zhang



CONTENTS

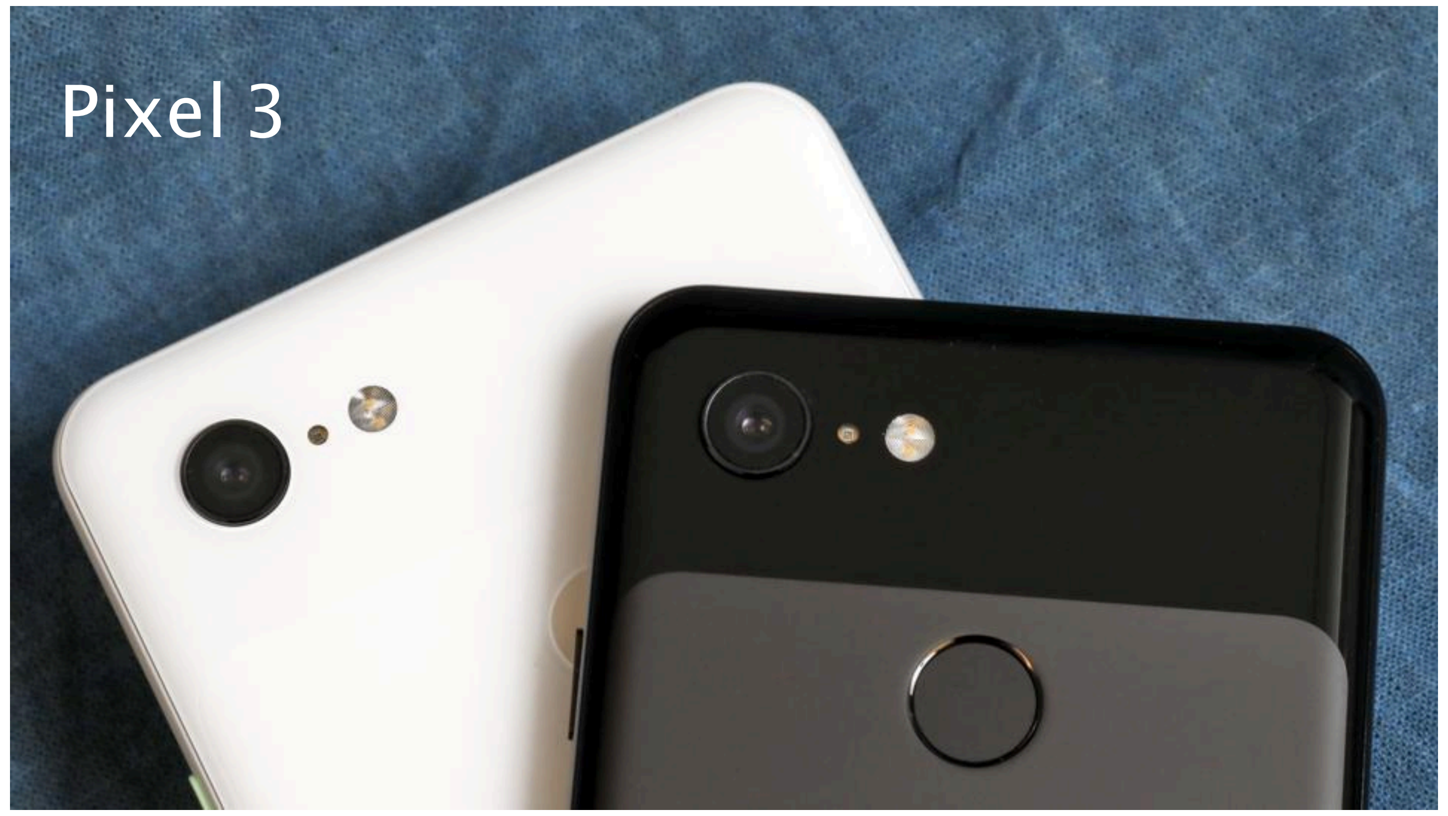
1. WHAT'S PORTRAIT SEGMENTATION?

RESULTS & APPLICATIONS



(a) Image (b) Our Result

Portrait mode - using a single camera phone



Single-lens reflex camera - Optics

Dual-camera phone - Algorithm

II. THEORIES

FULLY CONVOLUTIONAL NETWORKS (FCN)

What is FCN?

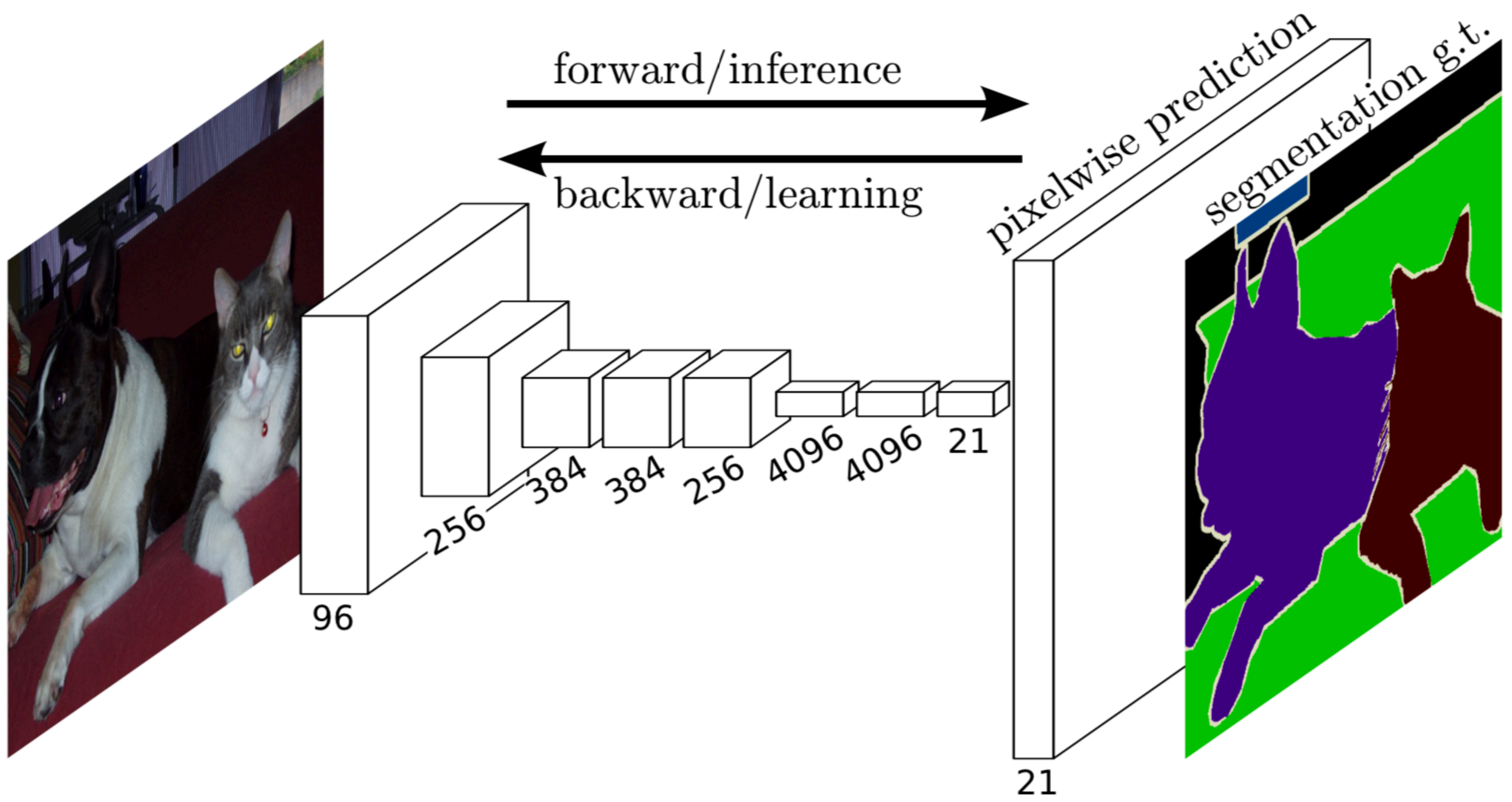
- End-to-end & Pixel-to-pixel
- Pixel-wise prediction

Input & Output image

- Input: Arbitrary size
- Output: Correspondingly-sized

Downsampling & Upsampling

- Downsampling
    - Reduces the amount of calculation
    - Increases the receptive field
- Upsampling
    - Is backwards strided convolution
    - Resizes the output to the same size as the input image



FCN :: CONVOLUTIONALIZATION

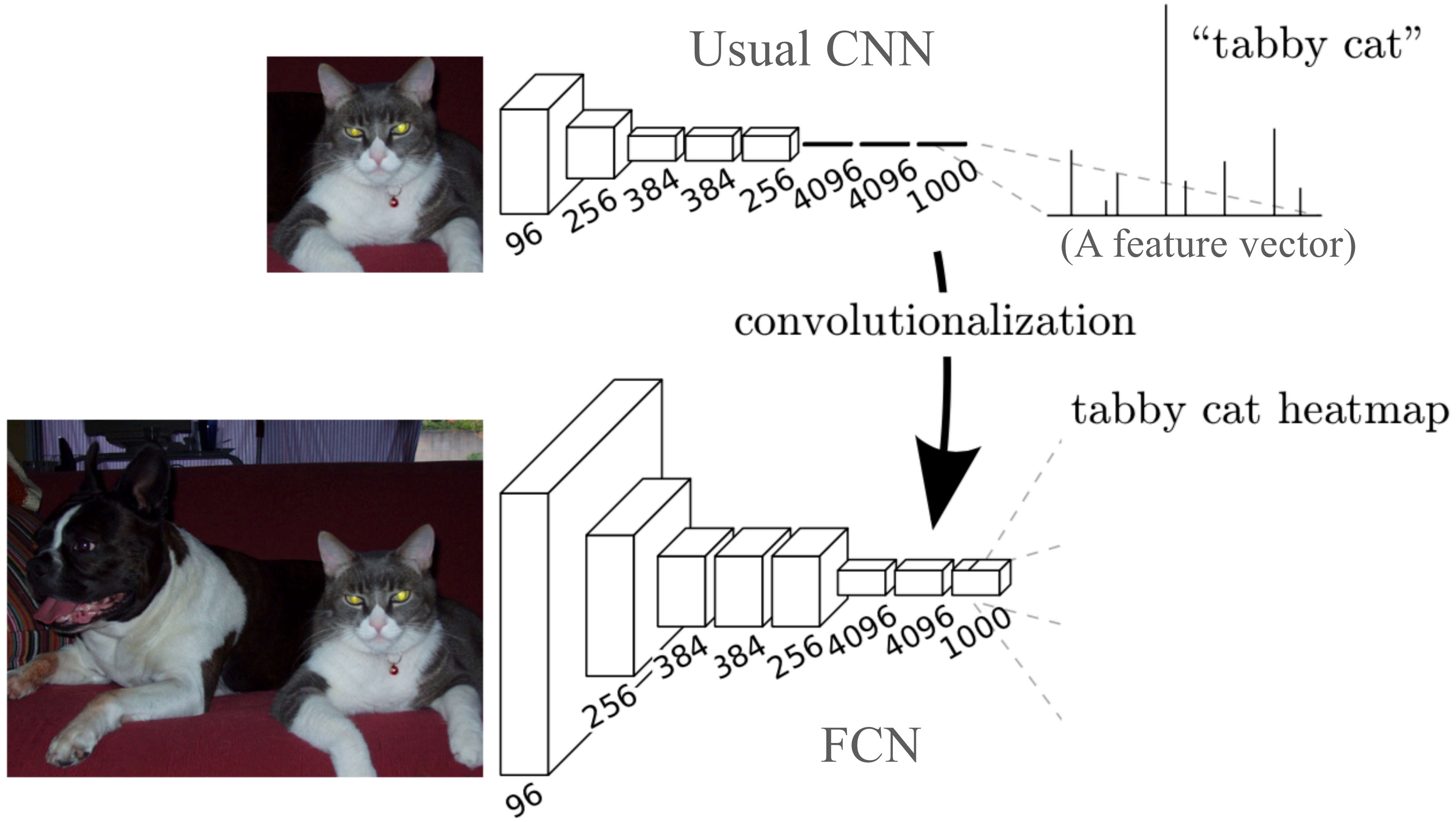
Differences between FCN and other CNNs ?

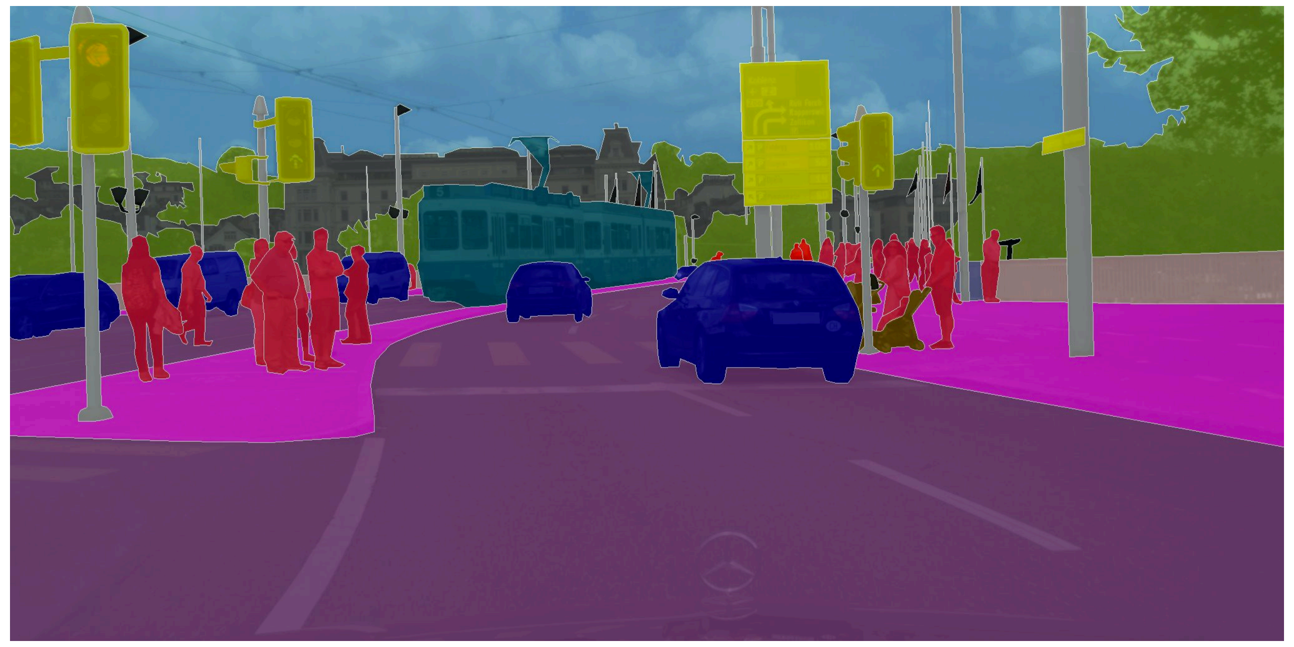
➤ Other CNNs usually

- Are connected to several fully connected layers
- Have fixed length feature vector output

➤ FCN

- Transform fully connected layers into conv layers
- Directly output a pixel-wise prediction to a heatmap

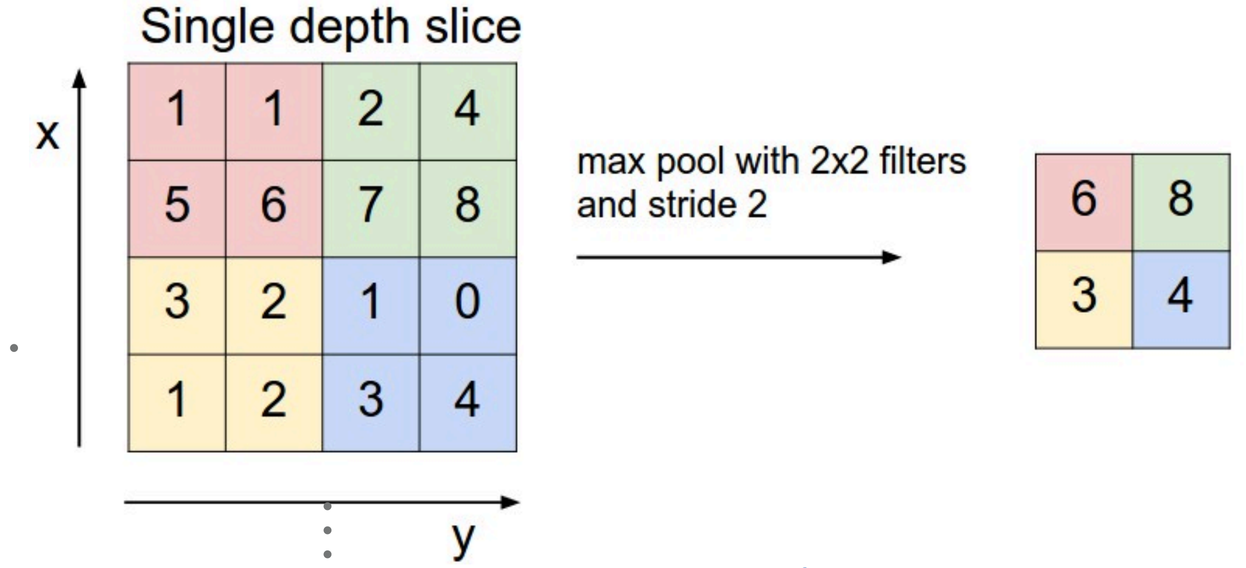




Why we need FCN?

- FCN classifies images at pixel level
- FCN solves **semantic segmentation problem**

FCN :: U-NET



U-Net

- Is one special type of FCN
- Commonly used in medical image segmentation

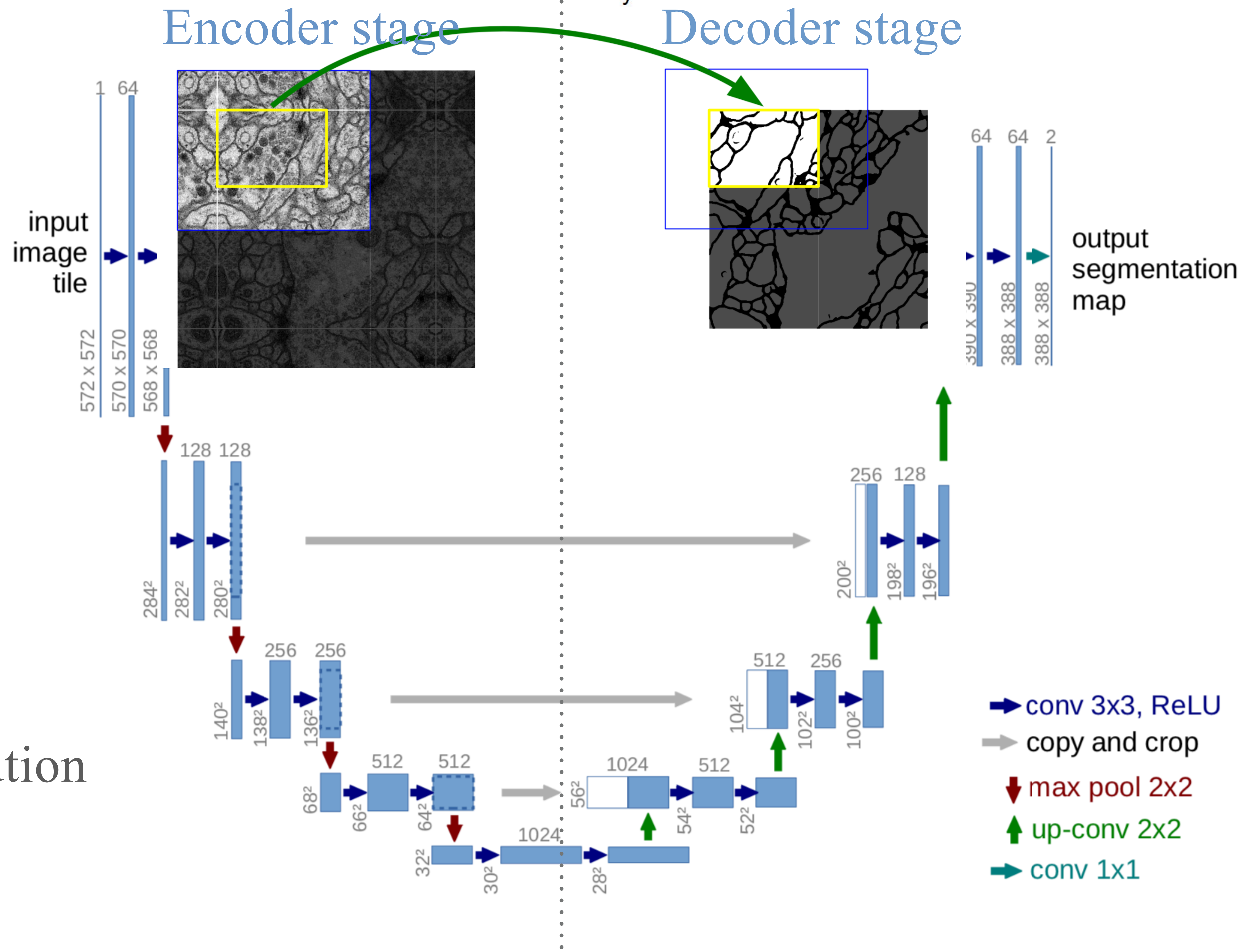
Encoder-decoder architecture

➤ Encoder - contracting path, at each step

- Max-pool with 2x2 filters and stride 2
- Double feature channels

➤ Decoder - expansive path, at each step

- Bilinear interpolation or transposed convolution
- Half feature channels



FCN :: U-NET :: SKIP ARCHITECTURE

Problem: loss of image details during downsampling

Solution: skip architecture

- Passes local information from shallow layer into deep layer
- Gets a refined prediction by adding sources of information for forecasting details

Approach: Concatenate

➤ Shallow convolutional layer:

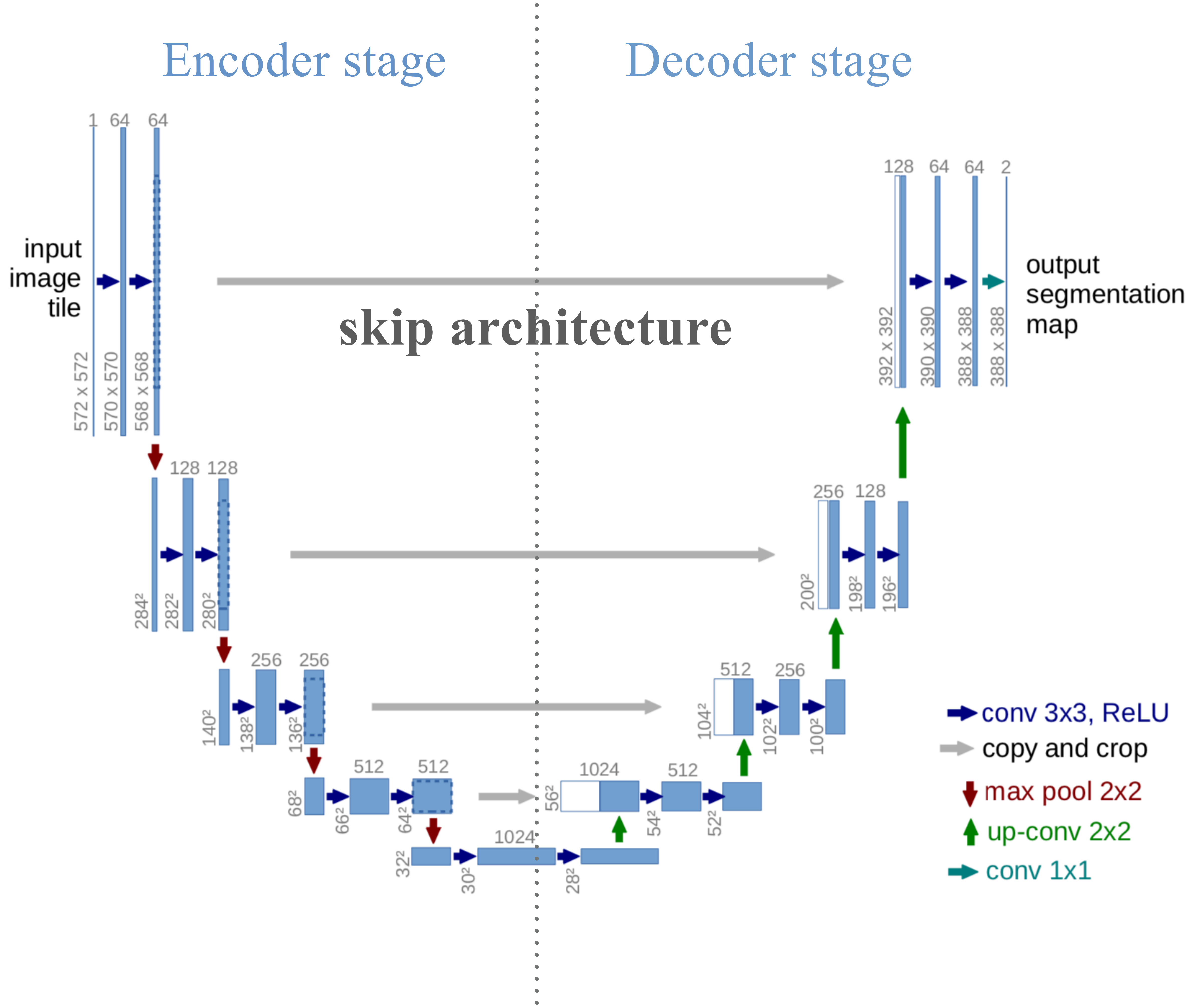
- has a small receptive field

- learns the features of some local regions

➤ Deep convolutional layer:

- has a large receptive field

- high-level semantic information

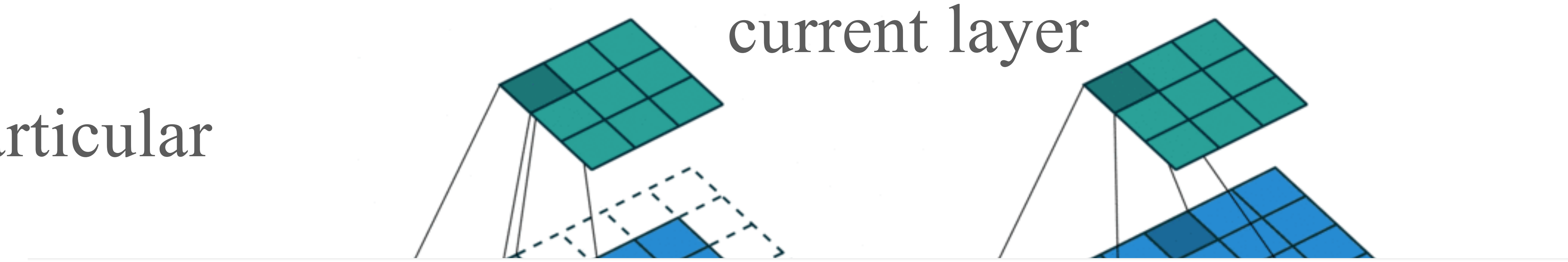


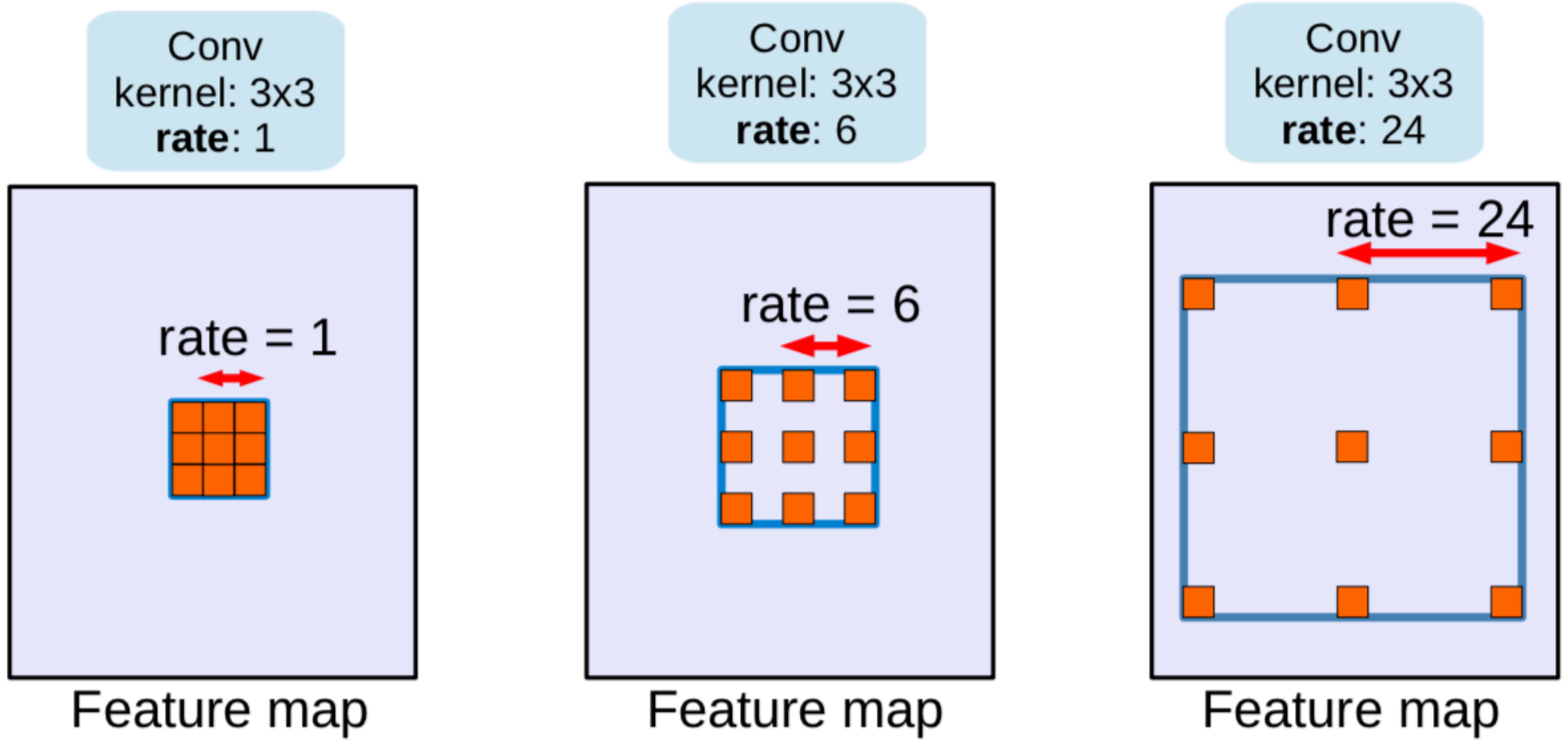
DILATED CONVOLUTION

Receptive field is

- The (gray) region of the input space that affects a particular unit of the network.

- The size of the area on the original image of the pixe feature map output by each layer of the CNN.





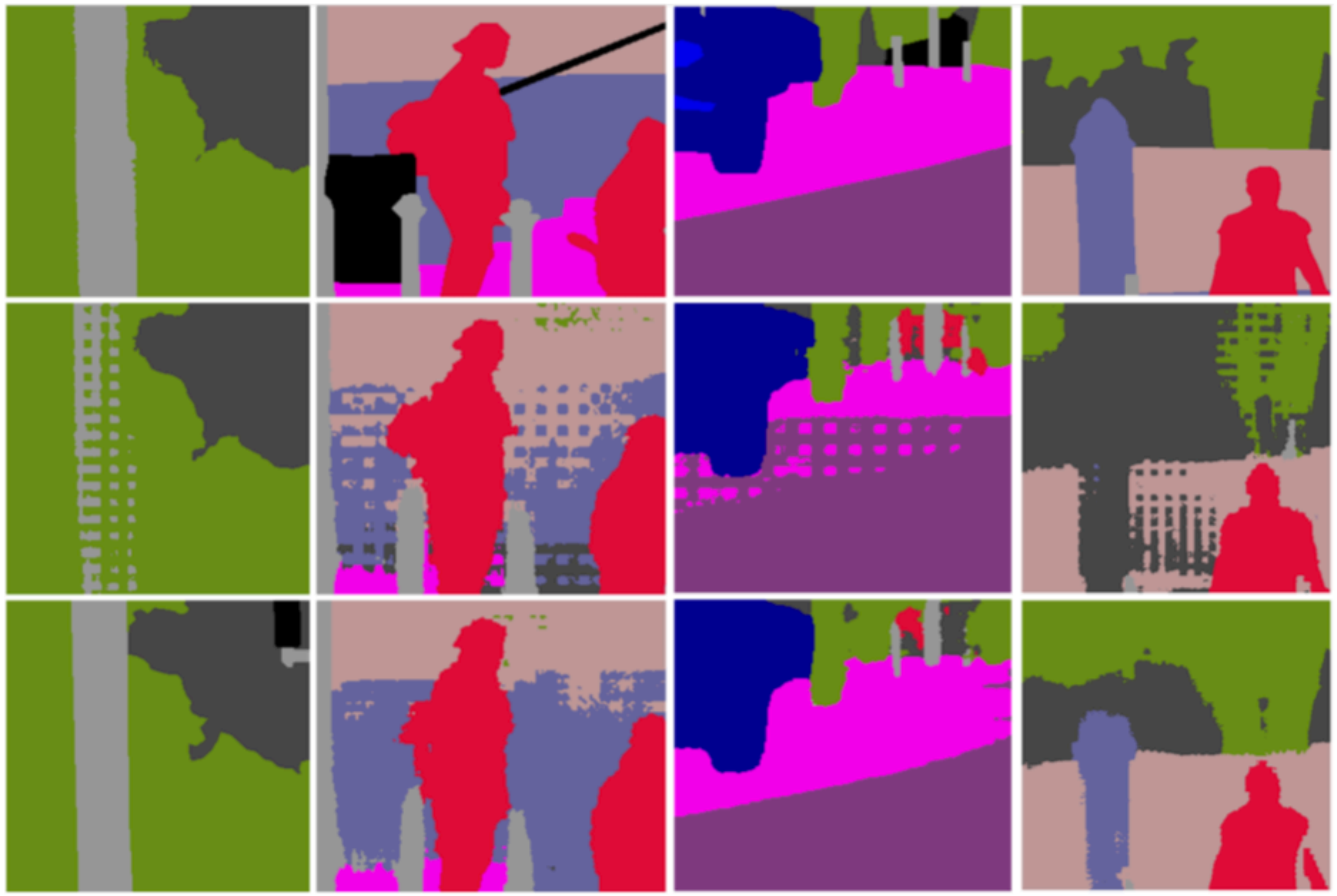
Advantages:

- Increase the feature map receptive field, without precision loss

- Increase the resolution of the middle feature map

- Improve prediction accuracy while maintaining the same computational cost

DILATED CONVOLUTION :: GRIDDING EFFECT

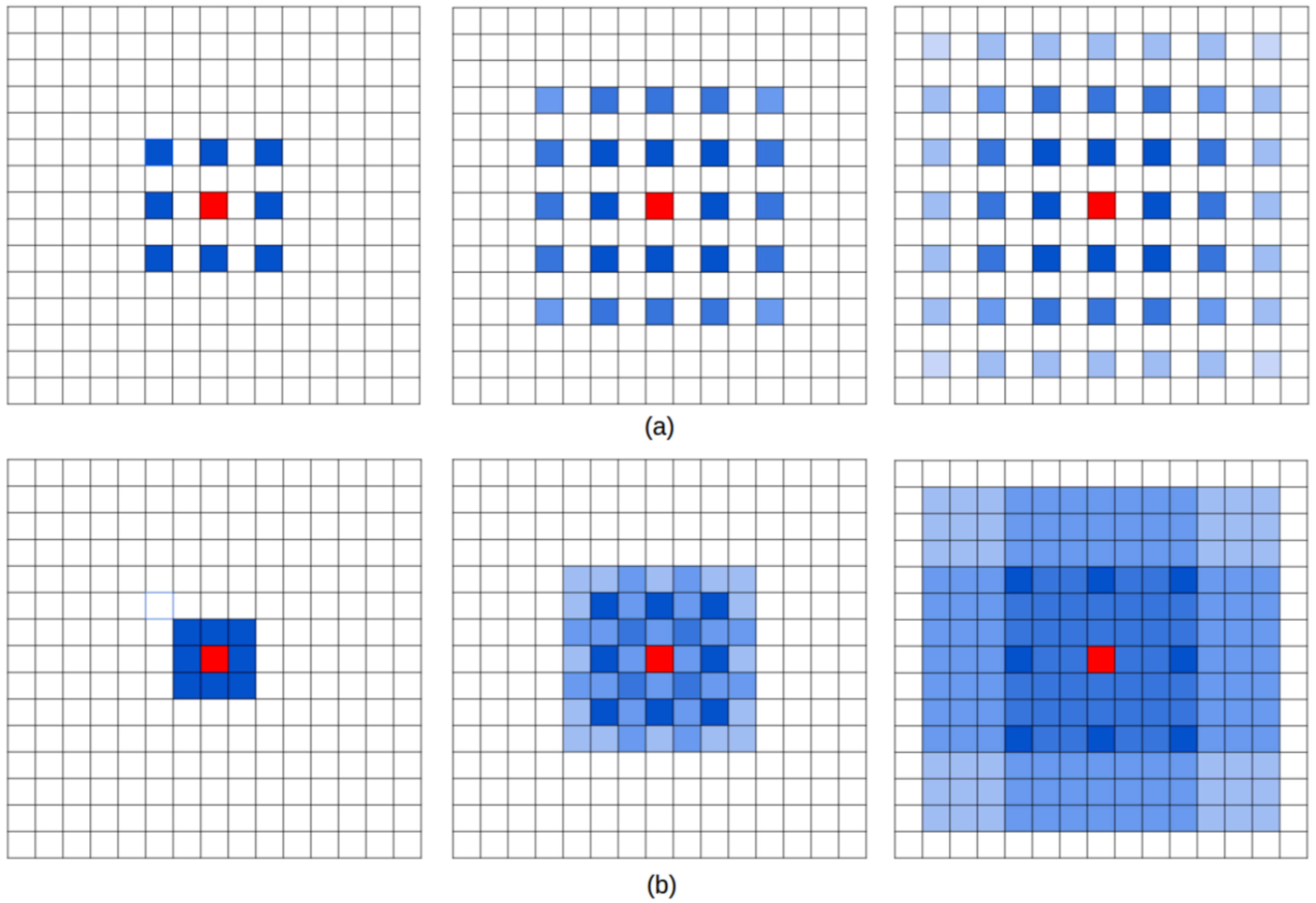


First row: ground truth patch.

Second row: prediction of the ResNet-DUC model.

Third row: prediction of the ResNet-DUC-HDC model.

Solution: **H**ybrid **D**ilated **C**onvolution

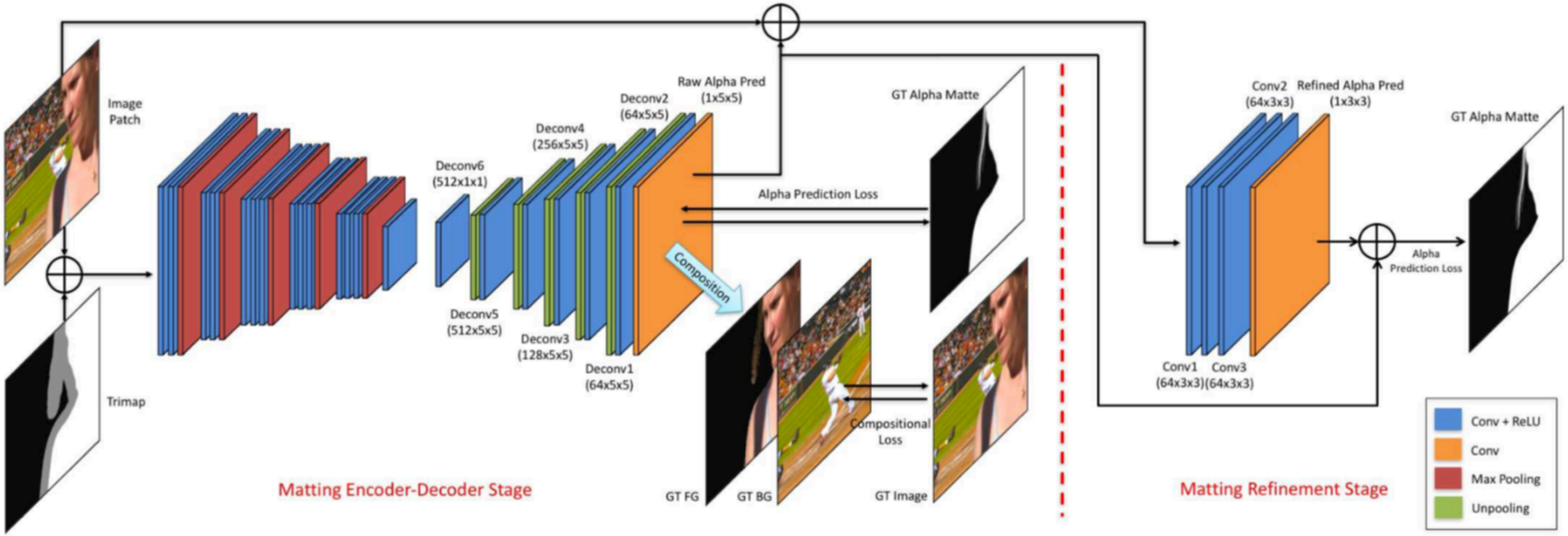


(a) all convolutional layers have a dilation rate r = 2.

(b) subsequent convolutional layers have dilation rates of r = 1, 2, 3, respectively.

III. STRUCTURES

ENCODER-DECODER STAGE + REFINEMENT STAGE



Encoder-decoder

- U-Net, each layer has fewer number of channels
  - To accommodate cellphone use
- Dilated convolution with HDC

Our Refinement

- Previous stage output
- Original input

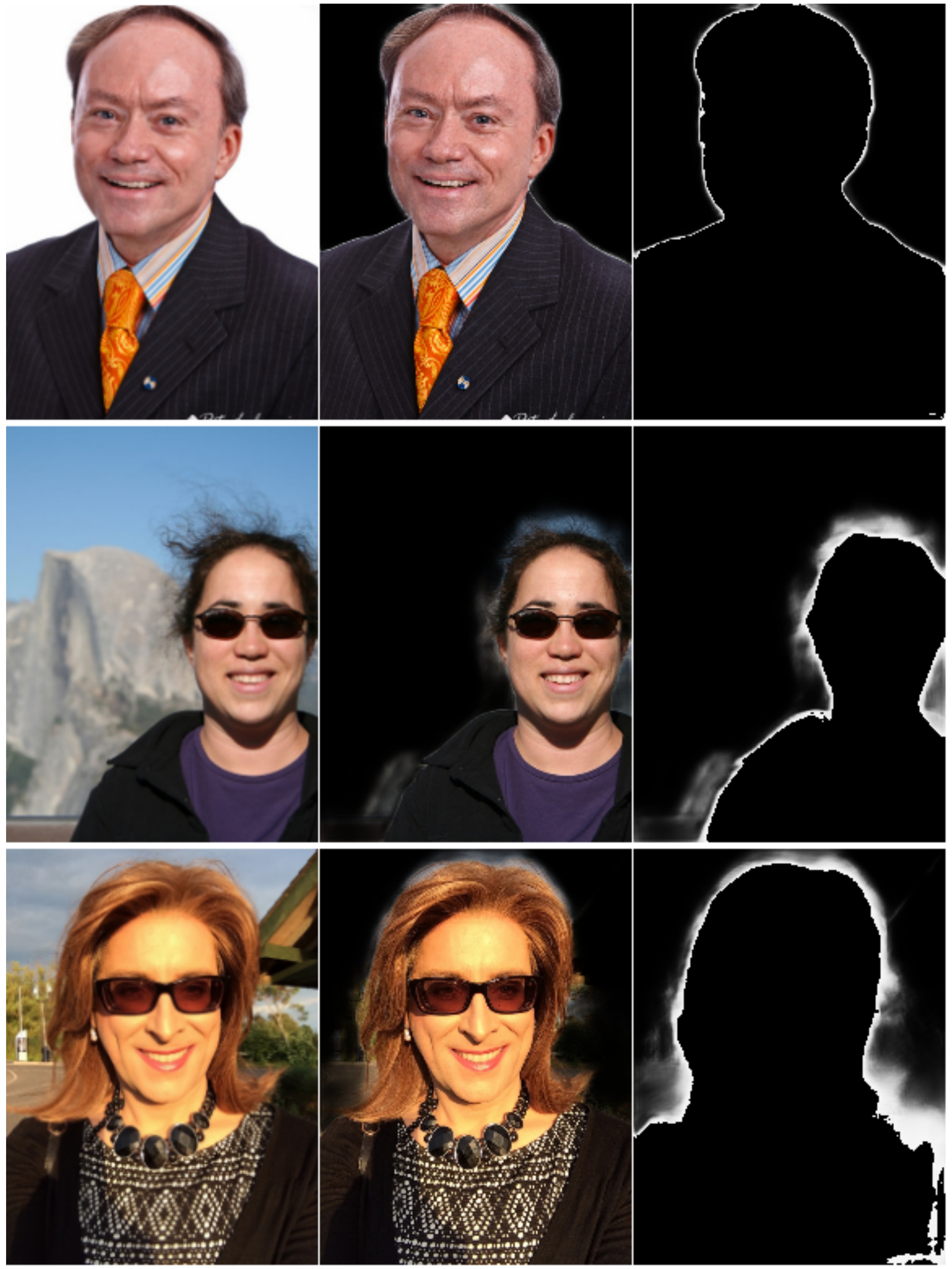
IV. FURTHER STEPS

REFINEMENT METHODS

- **Data Augmentation**
  - Generate from the existing dataset
  - ML is data-driven

- **Focus on the edge (hair)**
  - Edge is the most ambiguous & penalty is small
    - Proportion is relatively smaller than the body part
  - Use a special loss function in refinement stage

- **Combine with a higher resolution image**
  - Has more details, may have a more precise result



(a) Image (b) Matte (c) Boundary

V. REFERENCES

REFERENCES

1. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
2. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
3. Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., & Cottrell, G. (2018, March). Understanding convolution for semantic segmentation. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 1451-1460). IEEE.
4. Xu, N., Price, B., Cohen, S., & Huang, T. (2017). Deep image matting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2970-2979).
5. Shen, X., Hertzmann, A., Jia, J., Paris, S., Price, B., Shechtman, E., & Sachs, I. (2016, May). Automatic portrait segmentation for image stylization. In Computer Graphics Forum (Vol. 35, No. 2, pp. 93-102).

THANKS FOR WATCHING!

2019. 9.20