

RADONC AI CURRICULUM

Introduction to Deep Learning

LECTURE 6: U-Nets

ANDREW Y. K. FOONG, PH.D.

May 8th, 2026



Radiation
Oncology
AI & Data Analytics
AIDA

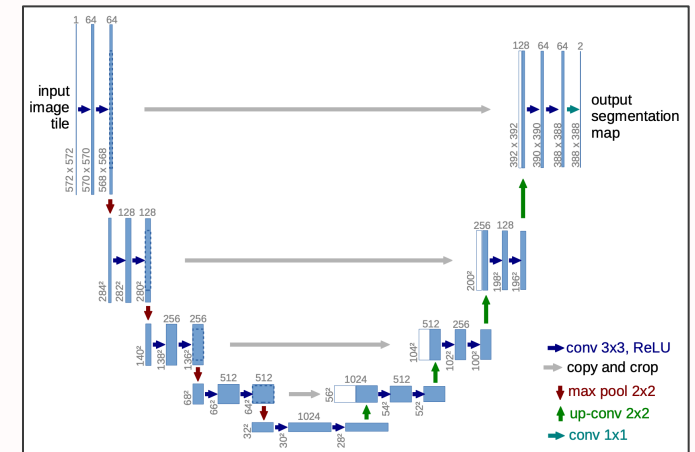
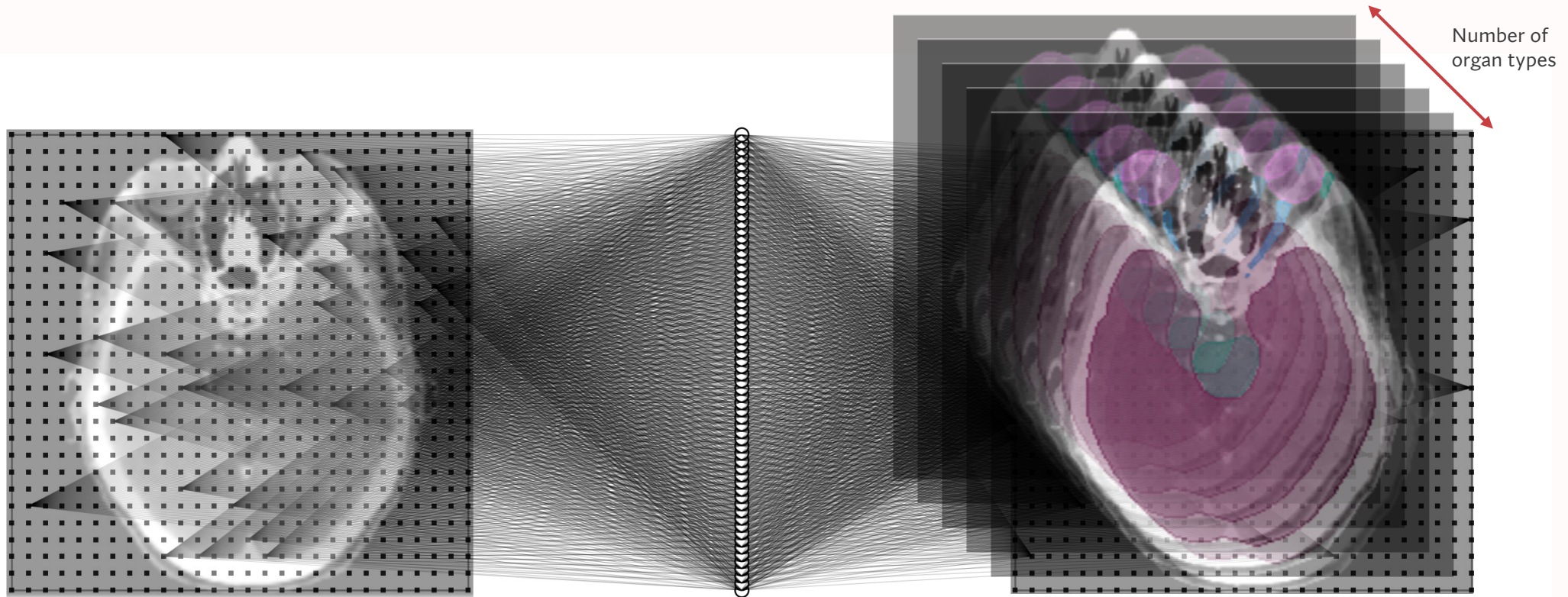


Image-to-image prediction

EXAMPLE: 2D auto-segmentation (unrealistically simple architecture)



Input image, x : [height, width]

Note: connections to only a few pixels are shown to avoid clutter

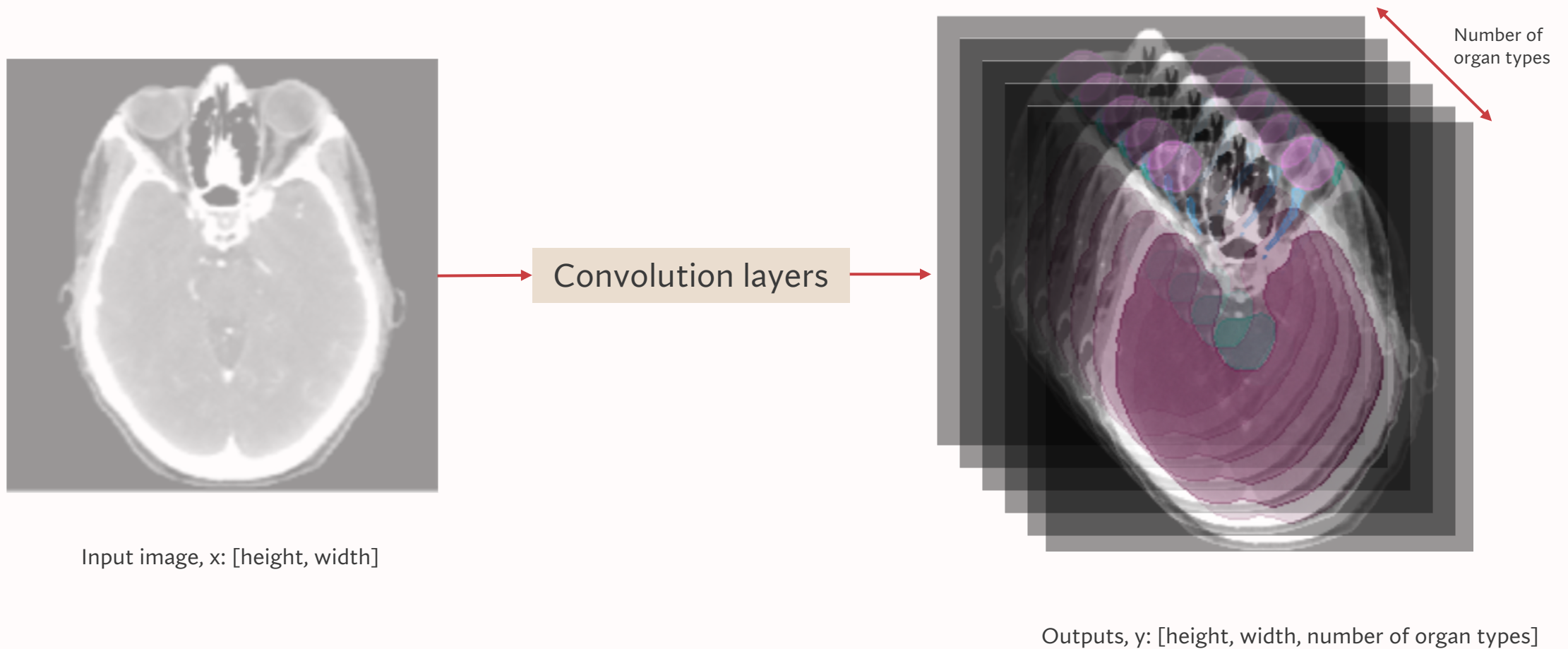
Hidden layer, width 64

Outputs, y : [height, width, number of organ types]

Note: connections to only a few pixels are shown to avoid clutter

Idea: use convolutions

- Convolutions take images as input and return images as output—perfect!



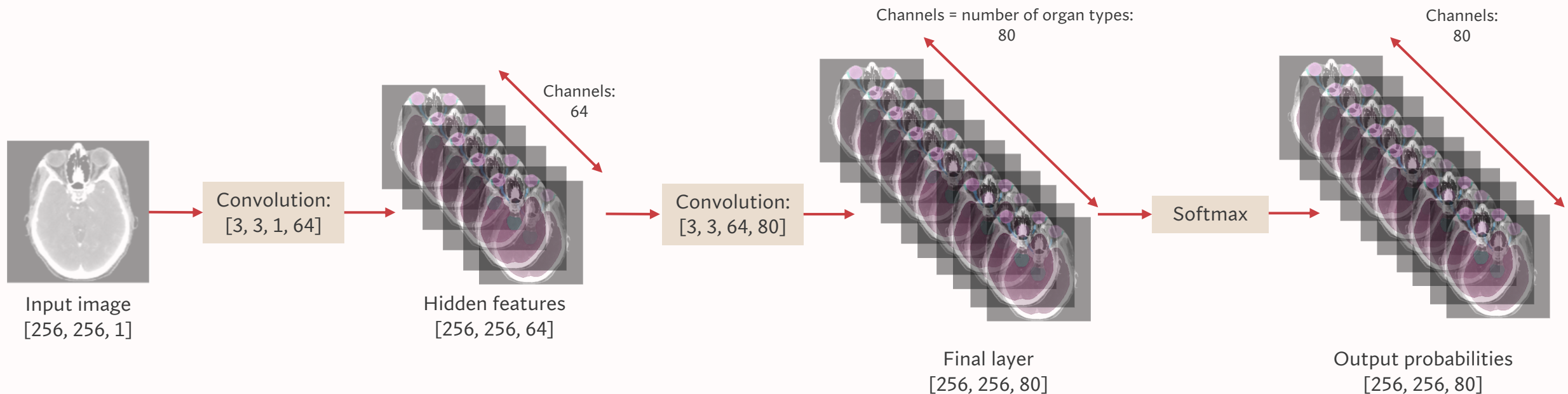
Fully-convolutional networks

- Neural networks that *only* use convolutions are called *fully-convolutional networks*—FCNs.
- Makes sense when both the *input* and the *output* have spatial structure.
- **EXAMPLES:**
 - CT image → segmentation of CT image—*appropriate*
 - Black & white image → colorized image—*appropriate*
 - Frame from a video → next frame of the video—*appropriate*
 - Image of dog → dog breed—*Not appropriate, use regular convolutional network*
 - Table of patient attributes → cancer risk—*Not appropriate, use regular neural network*

Tracking the shapes

- Convolution layers have parameters, like regular neural network layers.
- *Shape*: [KERNEL HEIGHT, KERNEL WIDTH, INPUT CHANNELS, OUTPUT CHANNELS]
- NOTE: each convolution is followed by a non-linearity such as ReLU.

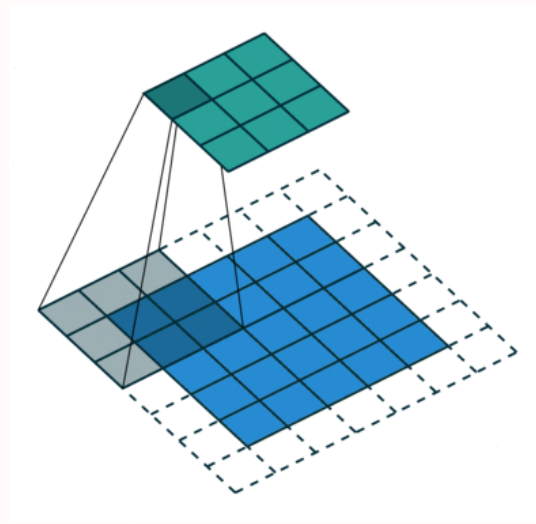
EXAMPLE:



Strided convolution

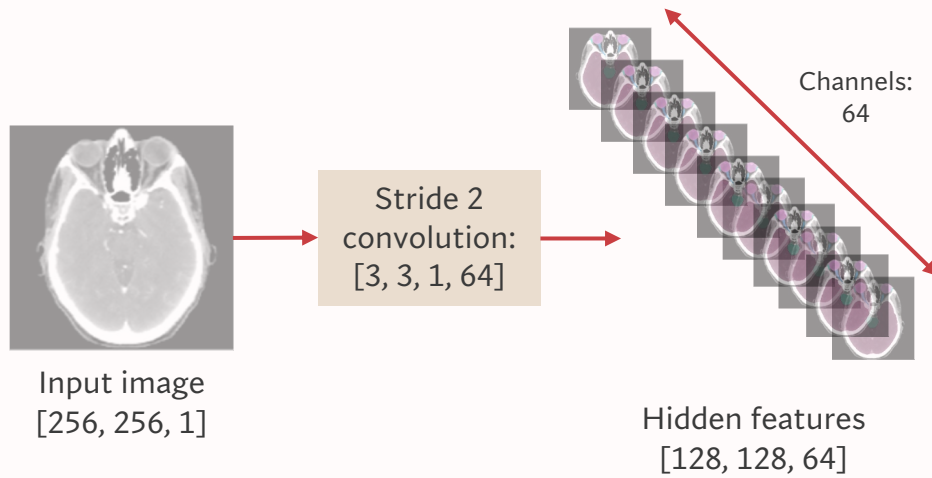
- Convolutional layers also have a *stride* parameter—how many pixels to skip.
- Greater stride, smaller output image.
- Stride allows each output square to capture a broader context—*increased receptive field*.

EXAMPLE: input = 5×5 , stride = 2, output = 3×3



Tracking the shapes

EXAMPLE:

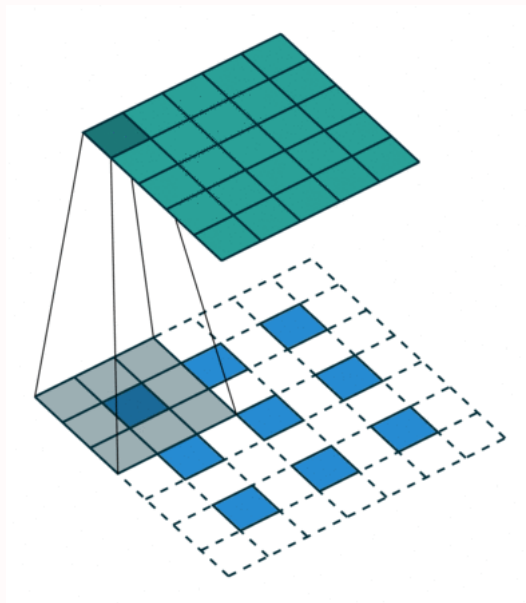


*But now the output is
in a lower resolution
than the input*

“Deconvolution”

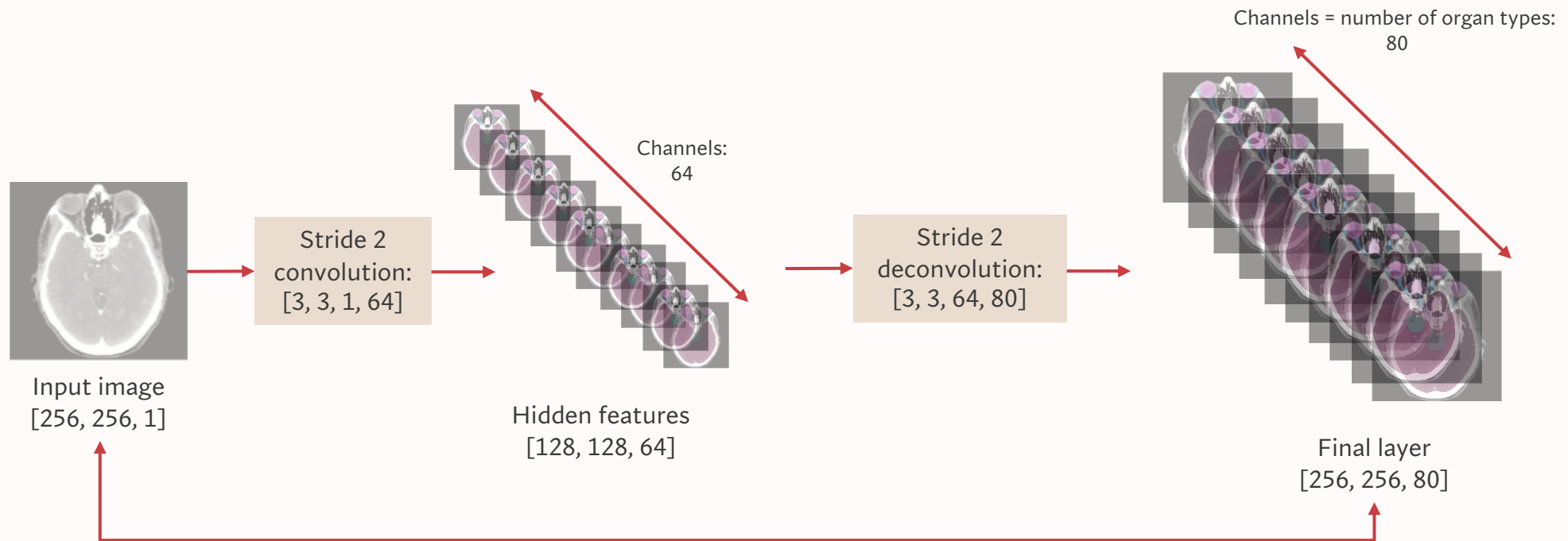
- We need a layer that *increases* resolution to undo this.
- Called *transposed convolution* or—less accurately—*deconvolution*.
- Greater stride, *larger* output image.

EXAMPLE: input = 3×3 , stride = 2, output = 5×5



Tracking the shapes

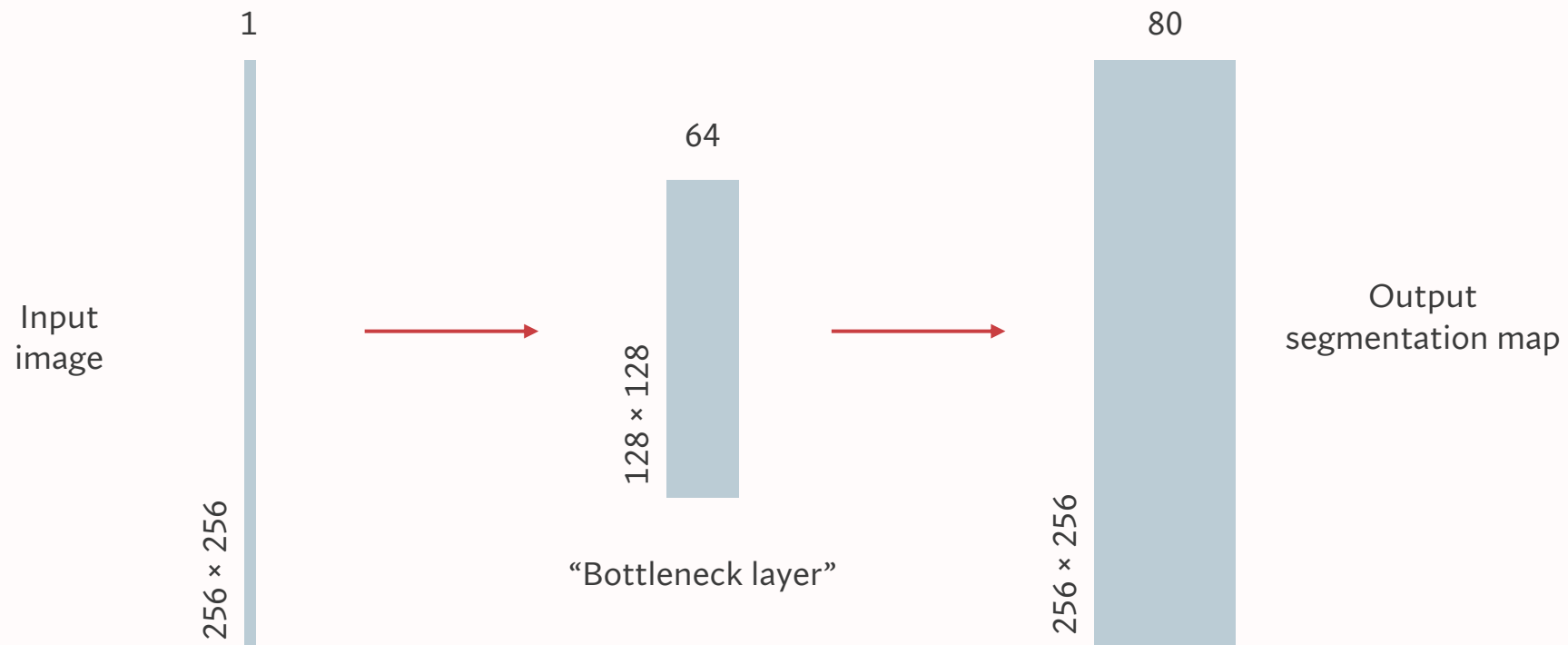
EXAMPLE:



Input and output resolutions match

Another kind of diagram—side view

EXAMPLE:



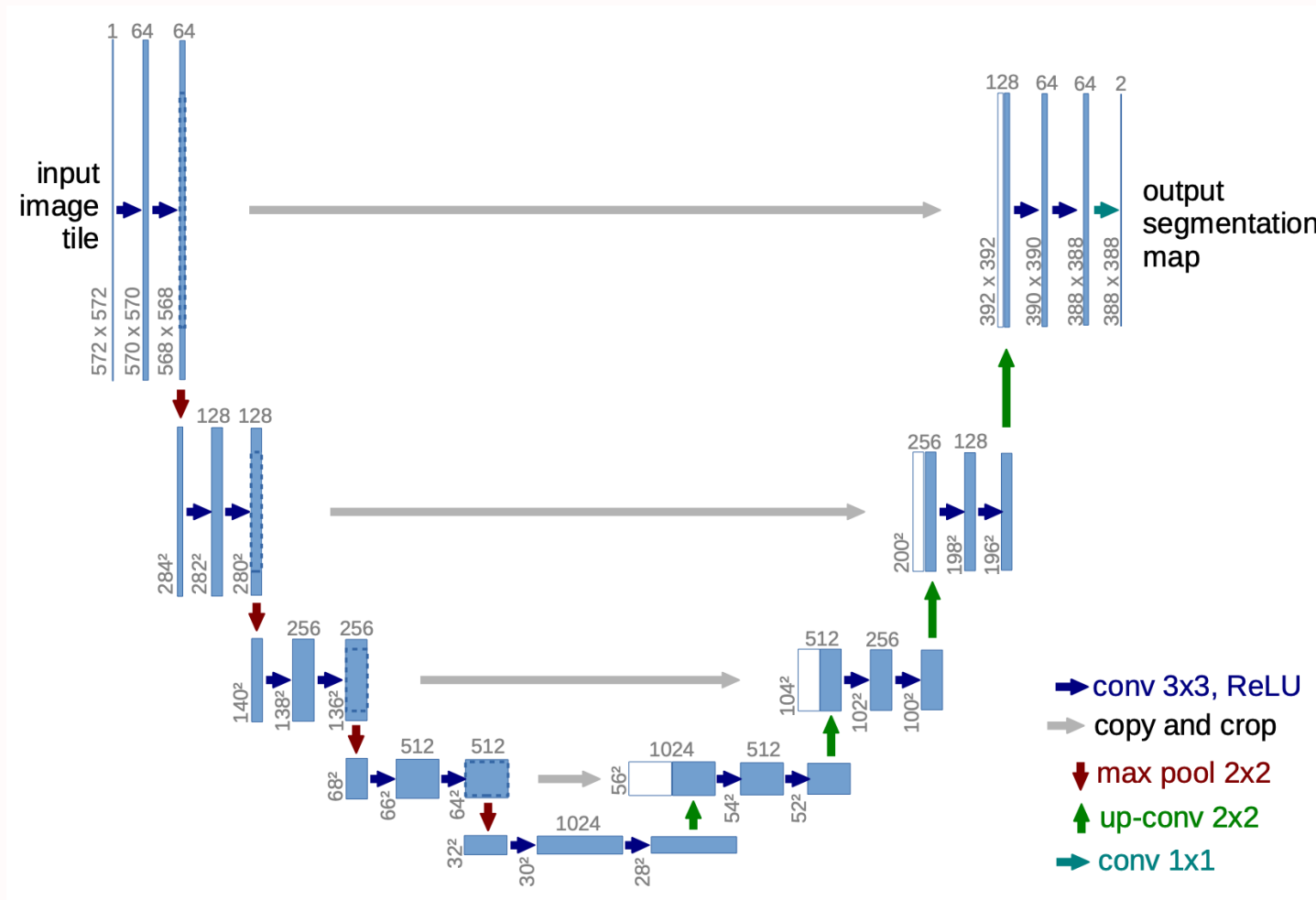
Bottleneck layers

- Many network architectures have a thinner *bottleneck layer* in the middle.
- Why are bottlenecks helpful?
 - Expands the receptive field—can encode global information.
 - Encourages compression—bottleneck forced to “summarize” important information about the image into a compact form.
 - Computational and memory efficiency—cheaper to run.
- BUT—they can also cause problems:
 - Fine-grained details in the image get washed out in bottleneck layers—information is destroyed.
 - Output ends up depending on global structure only.
- How can we get the best of *both* worlds?
 - Global information understanding
 - Local detail preservation

U-Nets

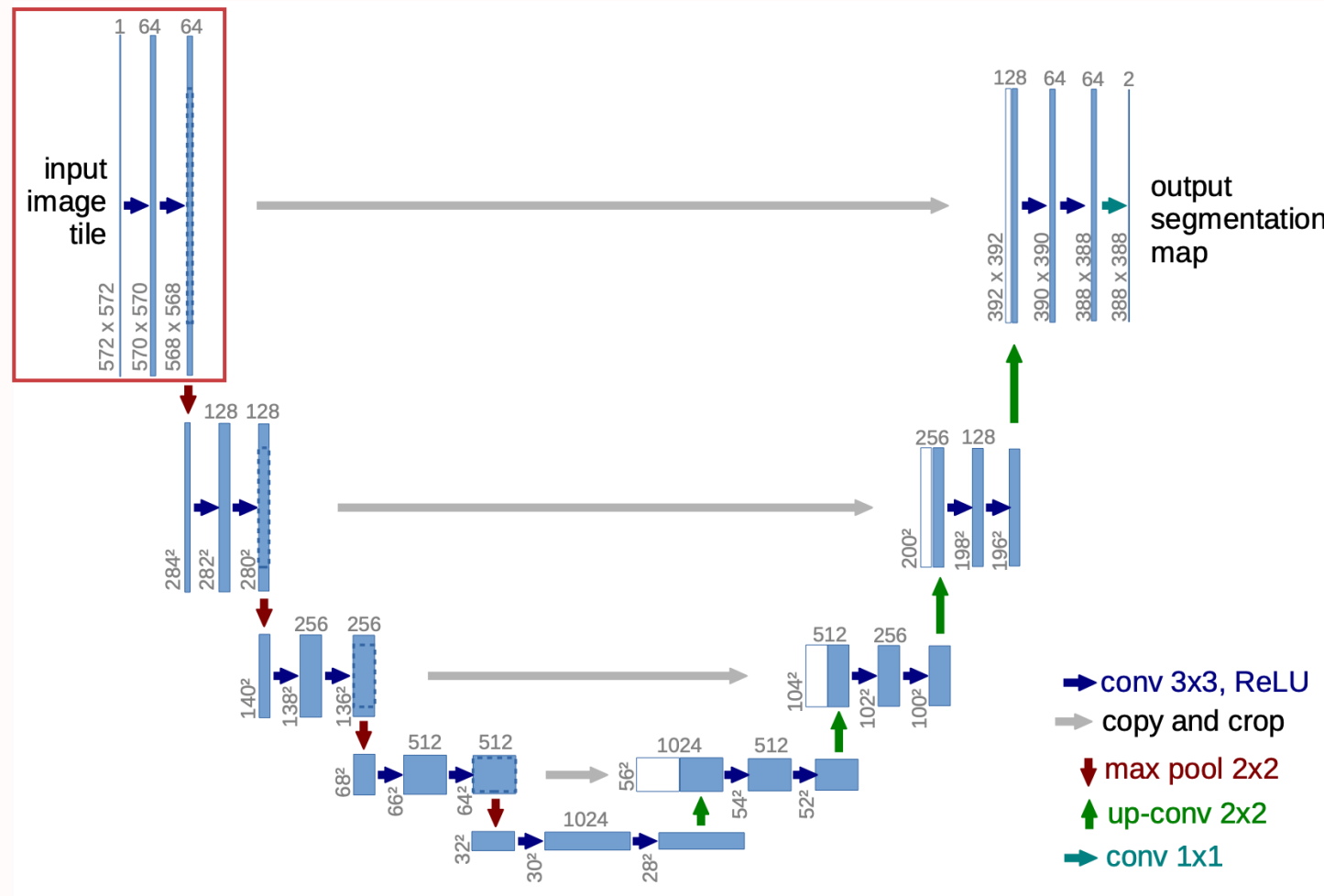
- *U-Nets* were originally designed for biomedical image segmentation.
- IDEA: to solve this problem, we need *multiscale* understanding:
 - At the small scale, individual textures of tissues.
 - At the medium scale, organs.
 - At the large scale, global information about the person being imaged.
- SOLUTION: use *skip-connections* as in residual networks, to *link processing at different scales*.

U-Net architecture



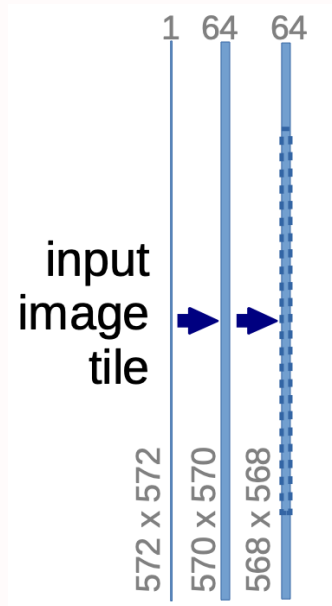
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture

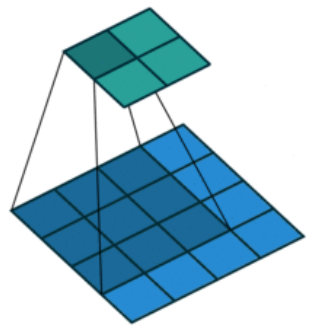


Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

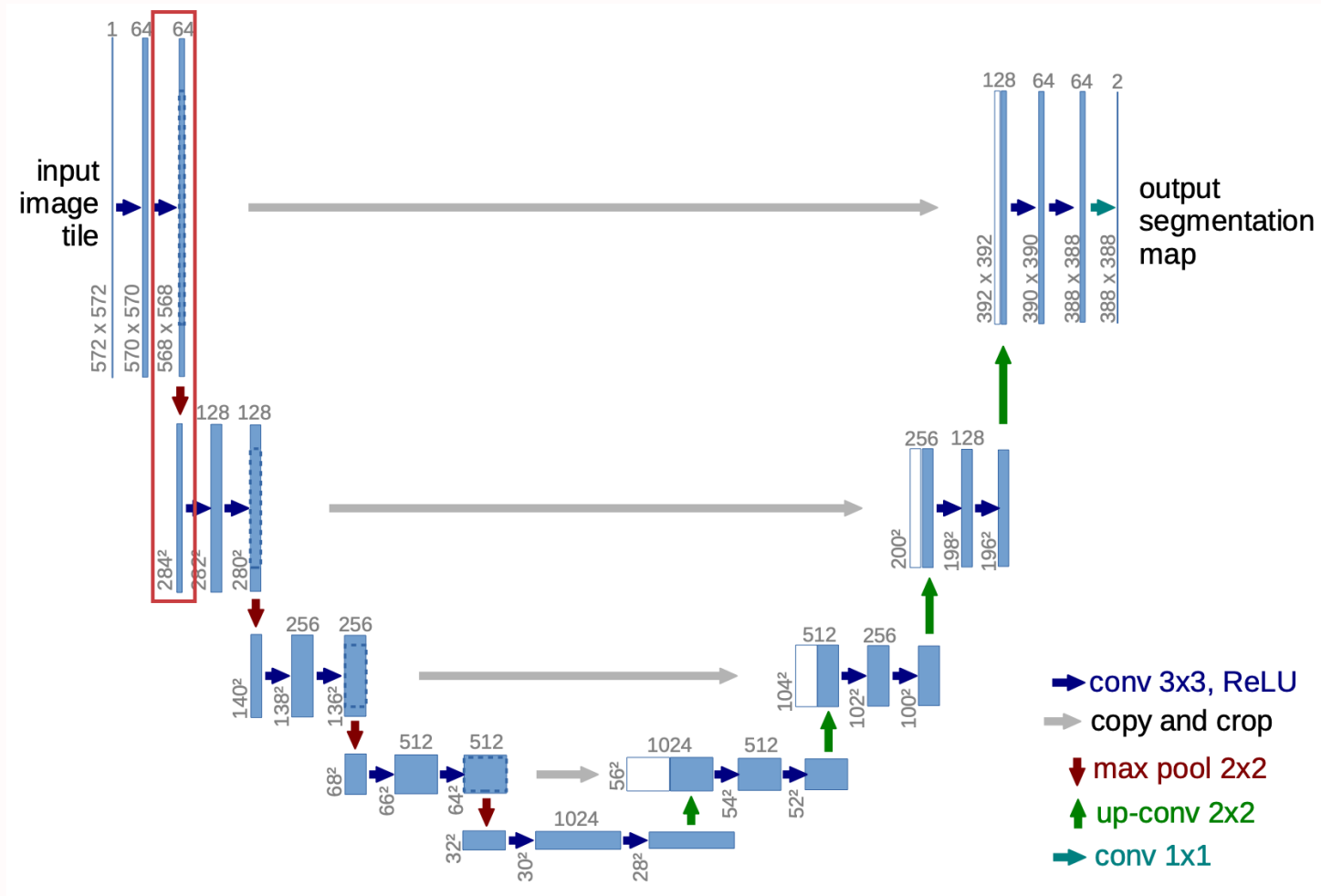
U-Net architecture



- First block is just like a regular fully-convolutional network.
- Two layers of convolution.
- *Note*: image size decreases slightly, from $572 \times 572 \rightarrow 570 \times 570 \rightarrow 568 \times 568$.
- Due to *edge effects*.



U-Net architecture



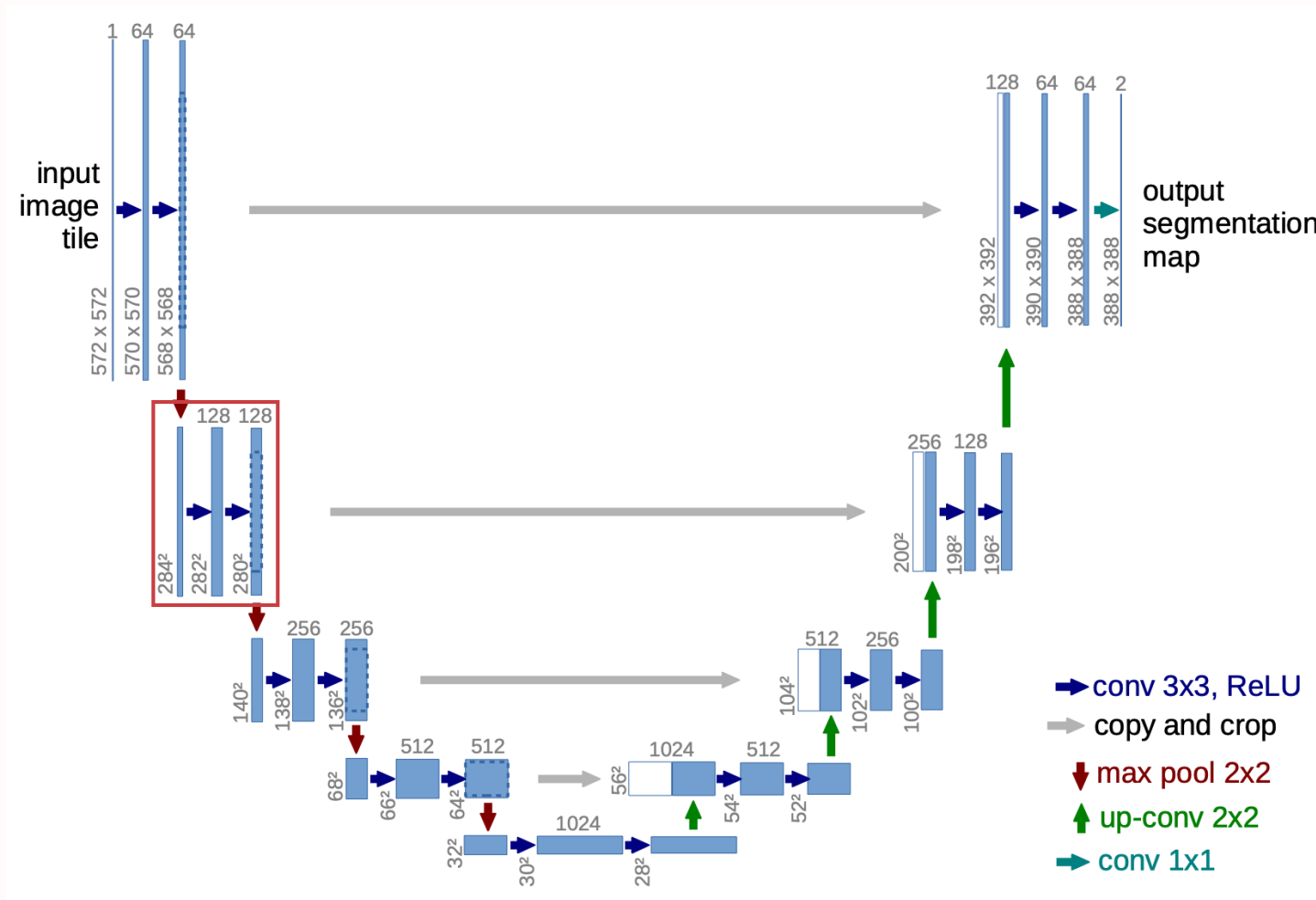
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



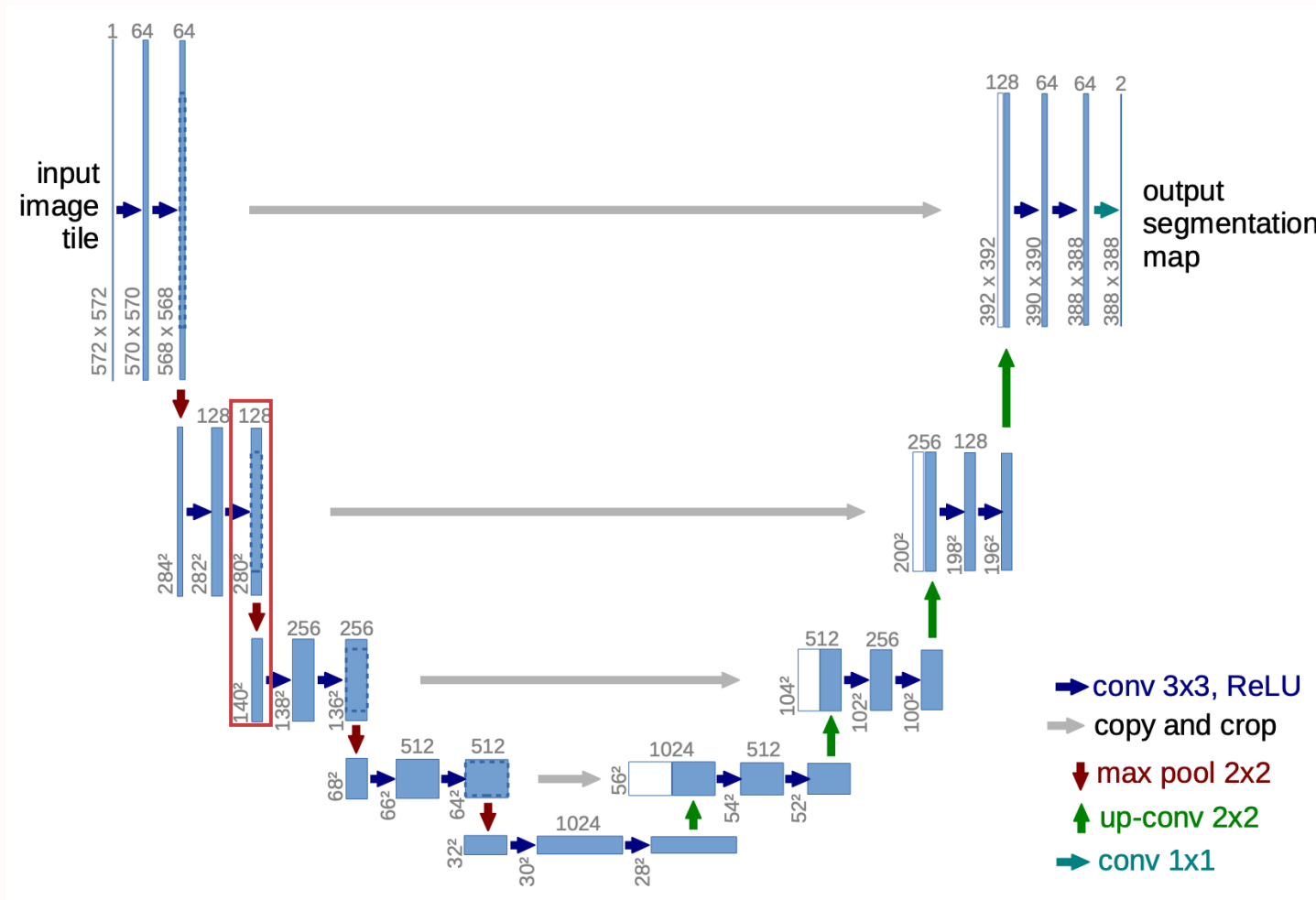
- First down-sampling.
- Max-pooling from $568 \times 568 \rightarrow 284 \times 284$.
- Could use strided convolution as well.
- Each feature now represents a larger region than a single pixel.

U-Net architecture



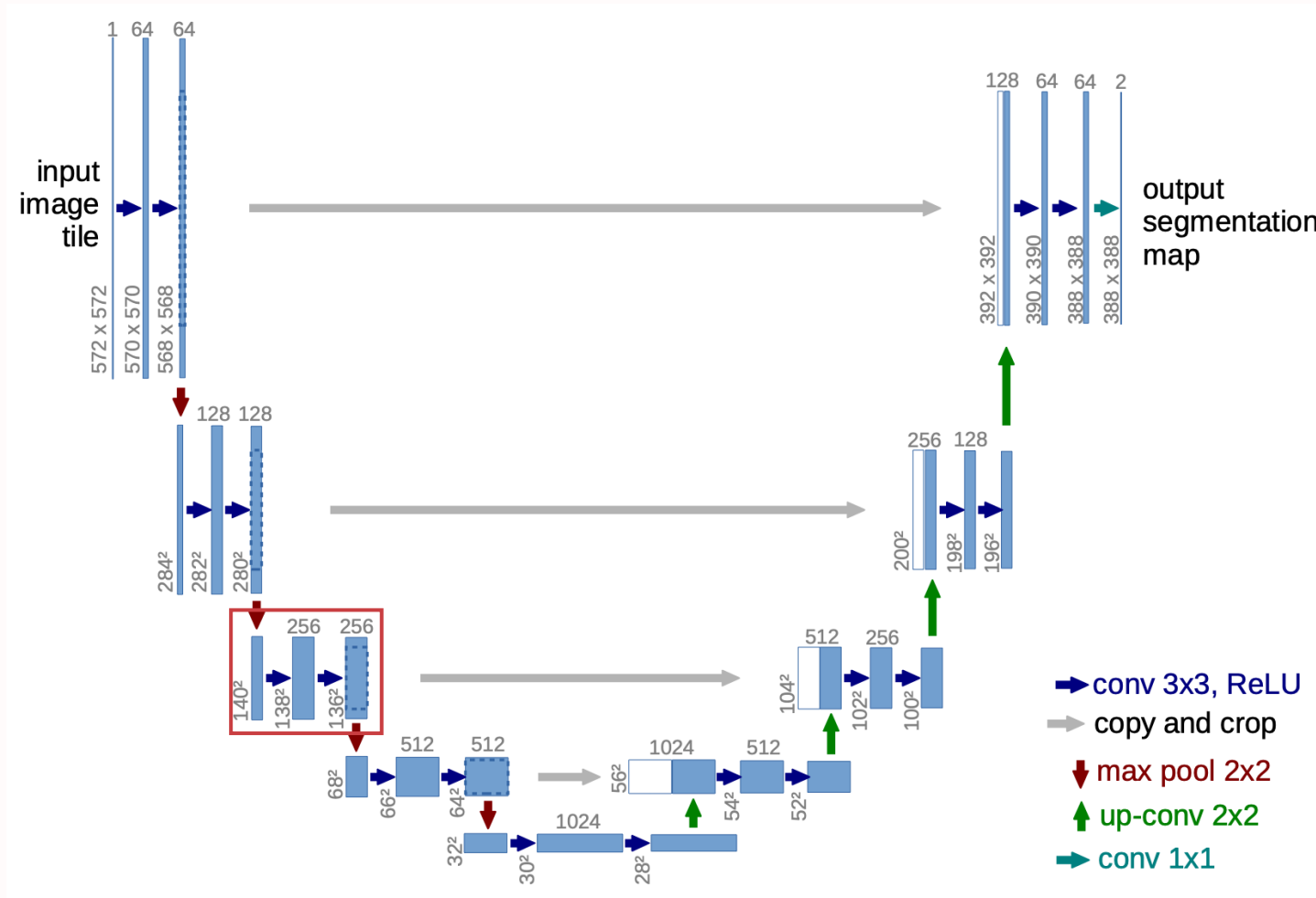
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



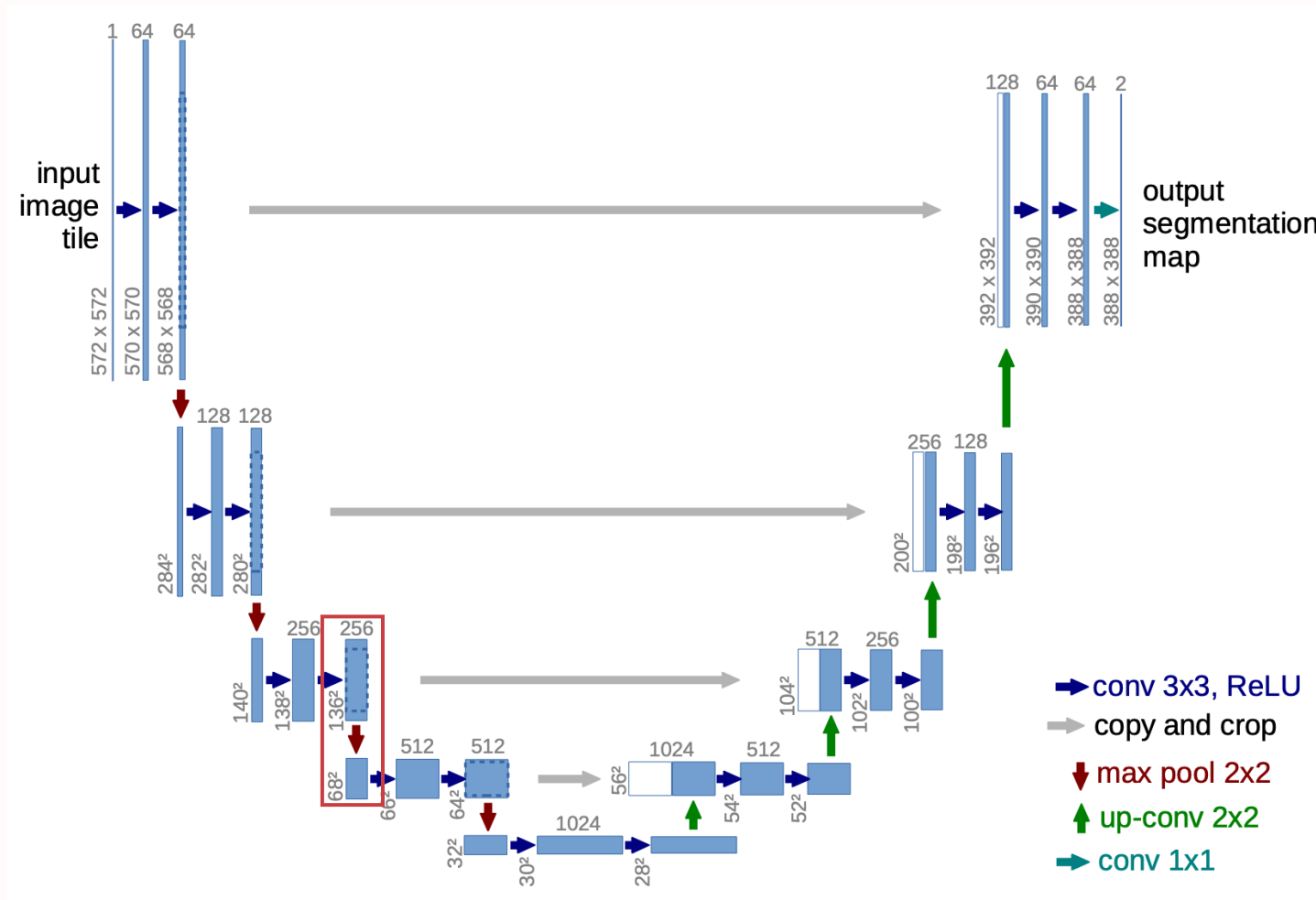
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



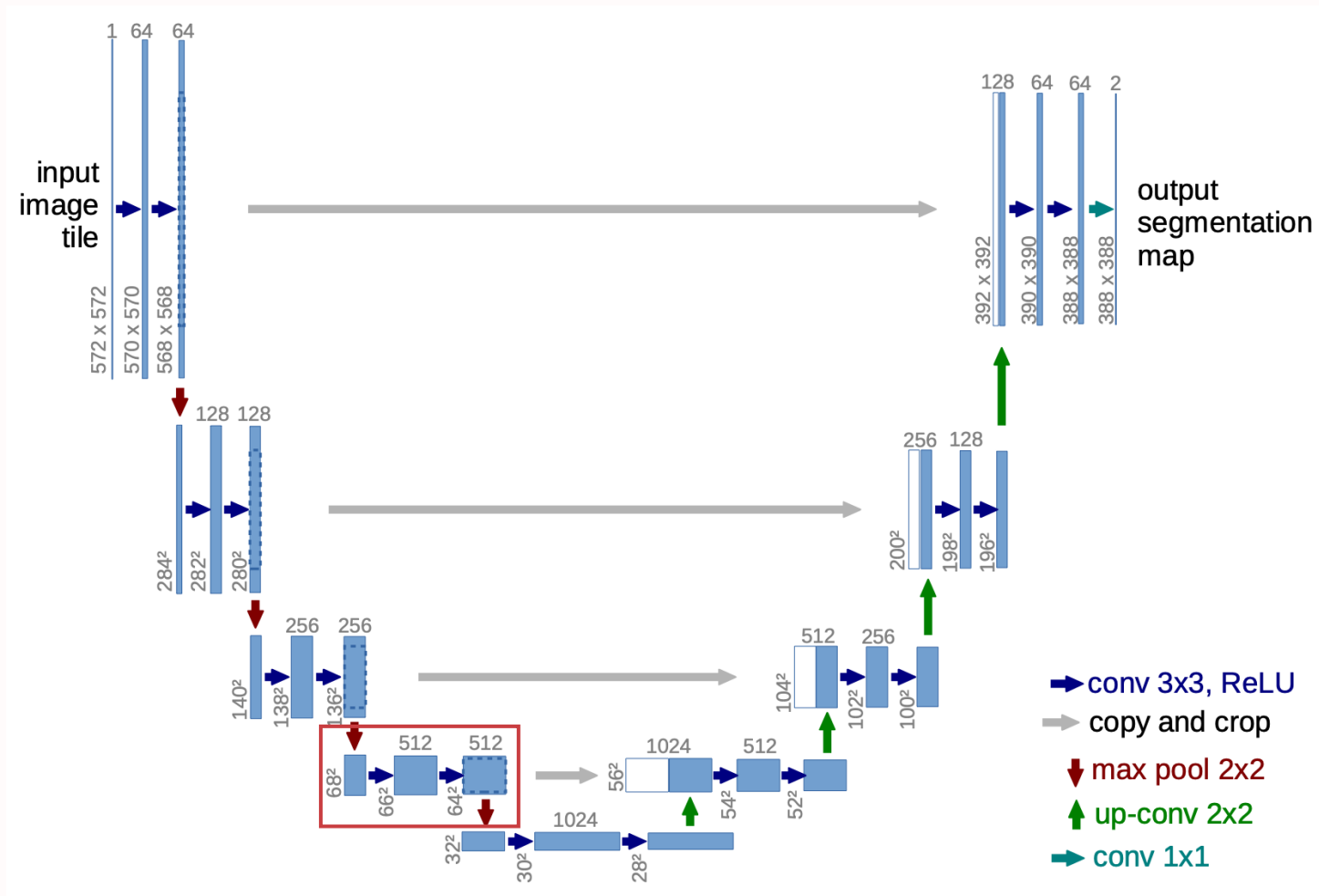
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



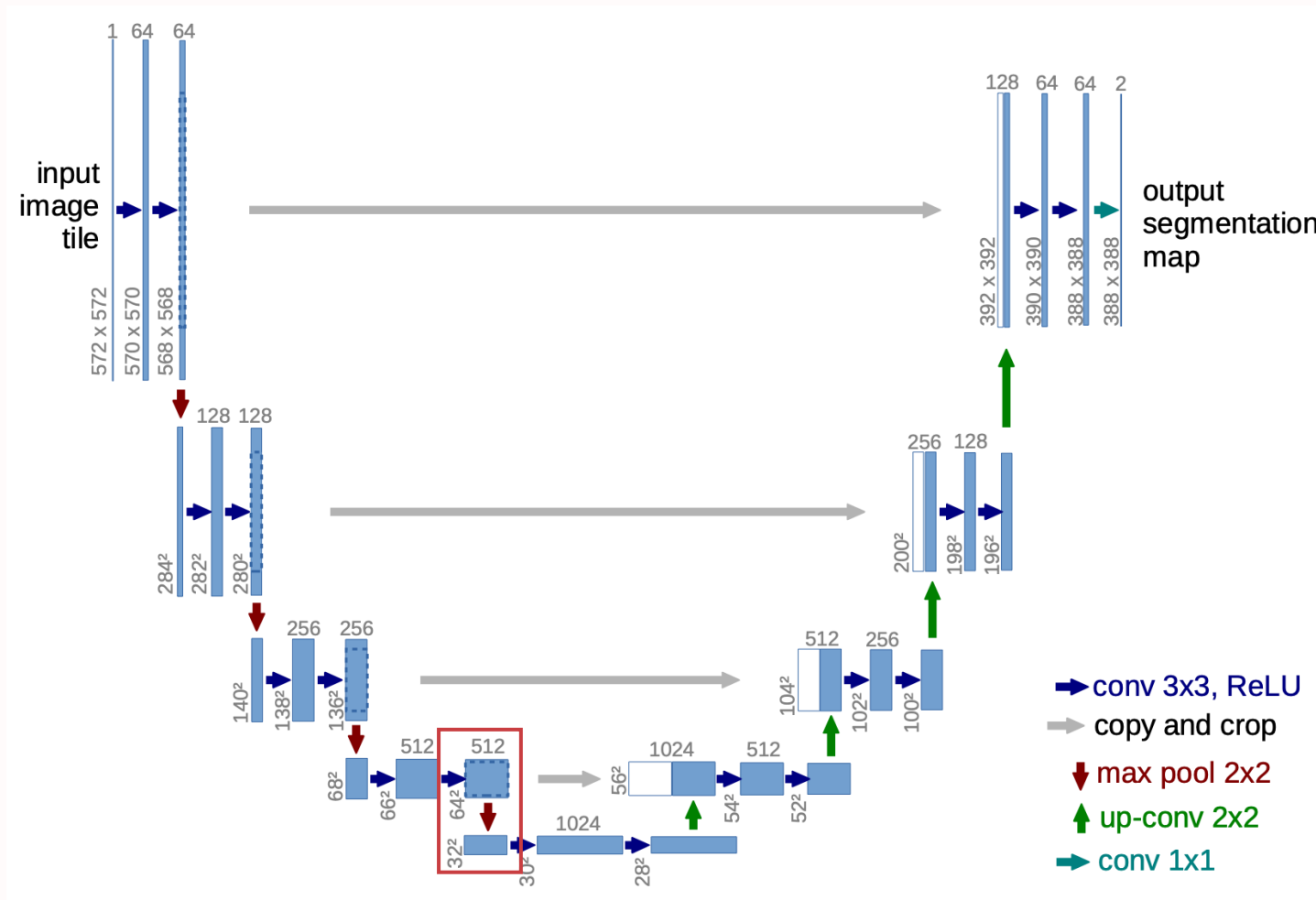
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



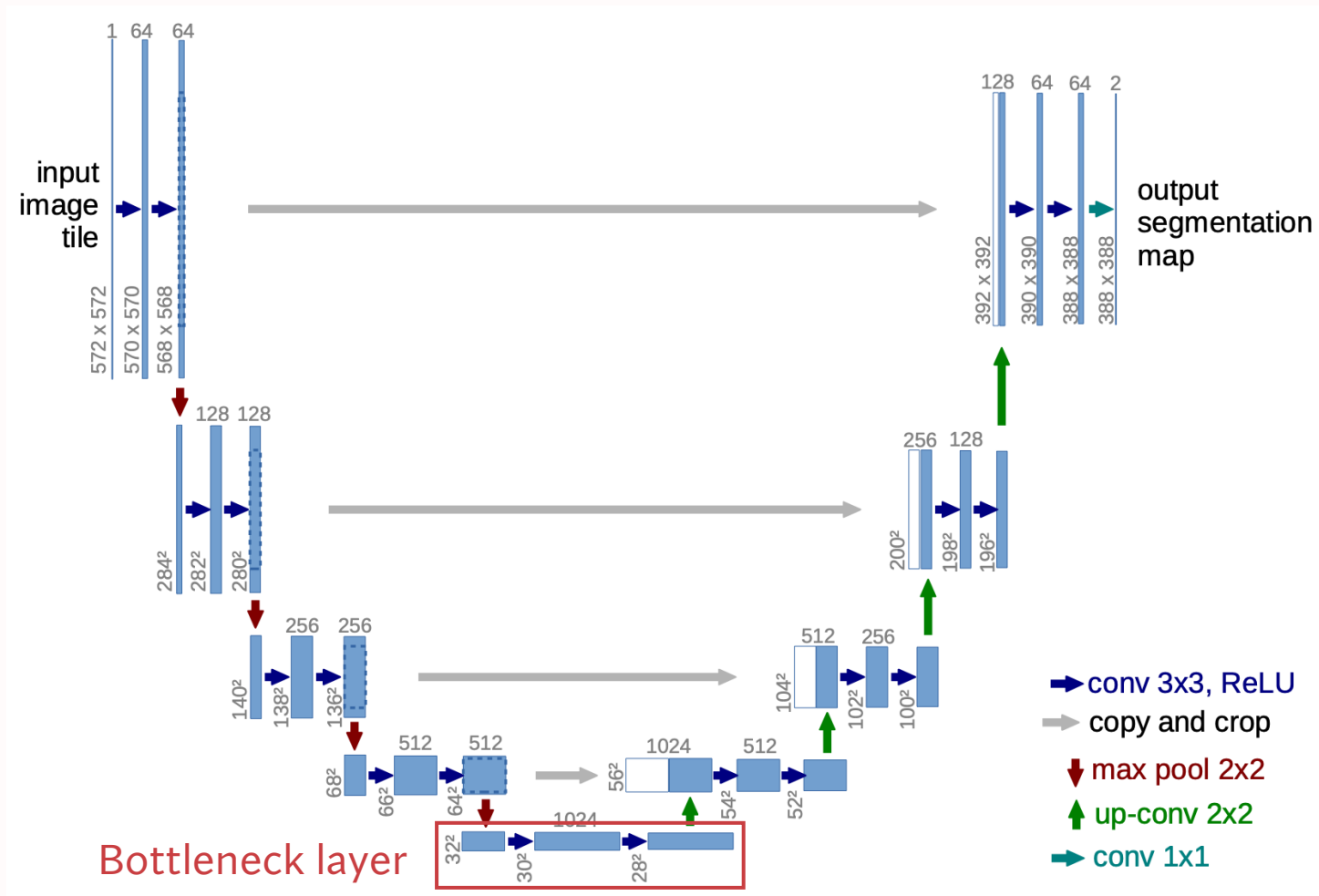
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



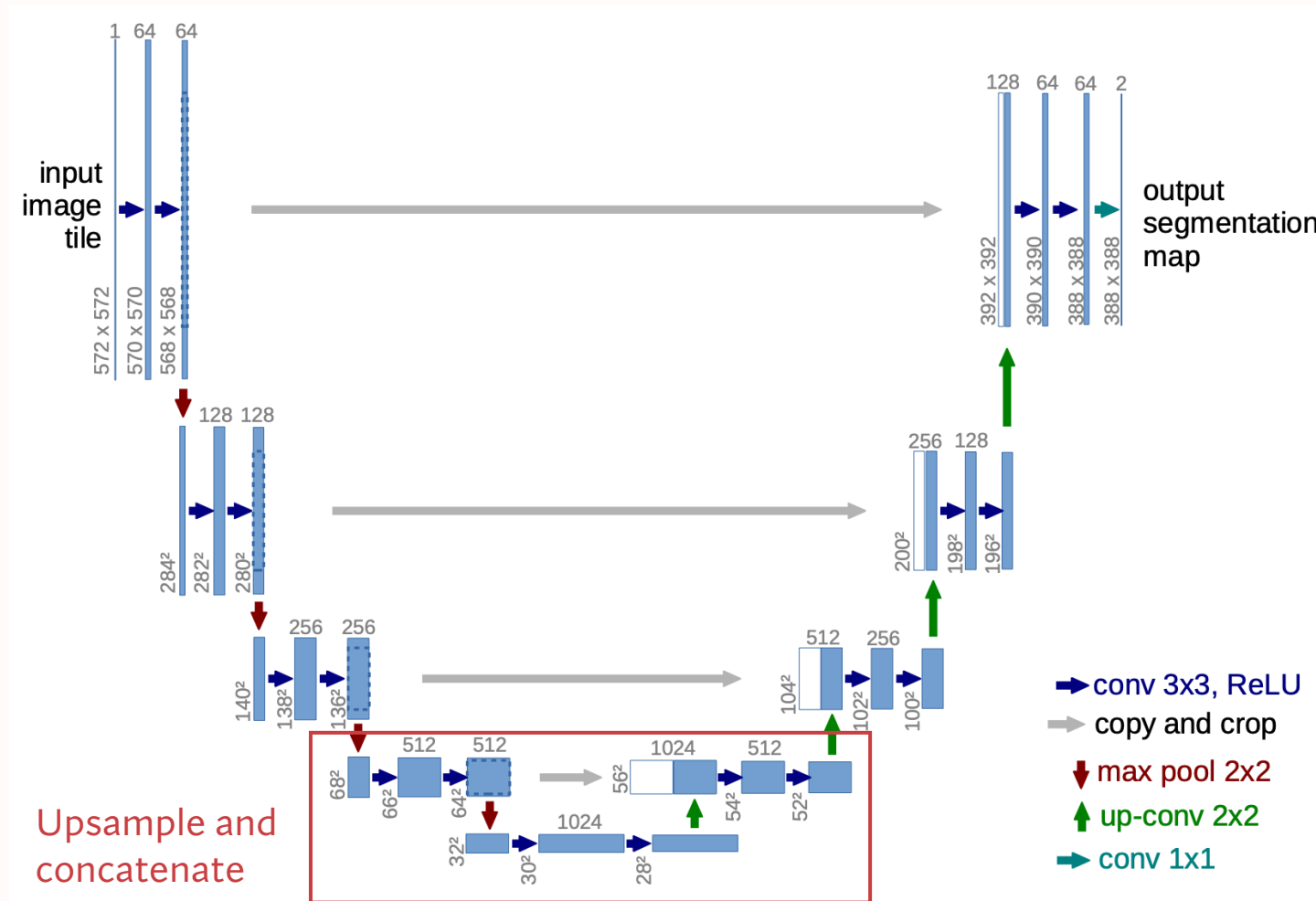
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



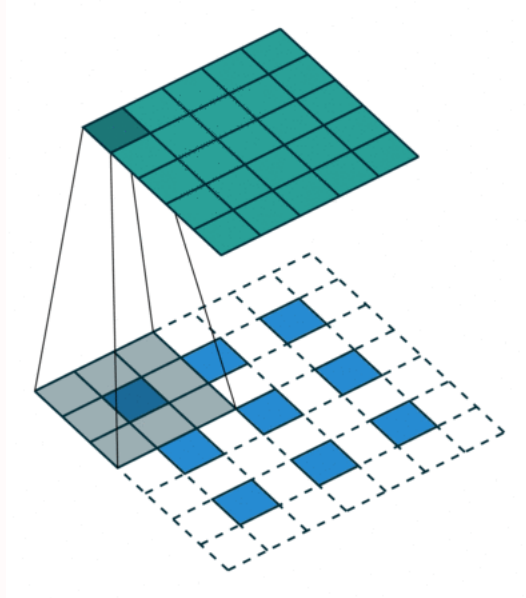
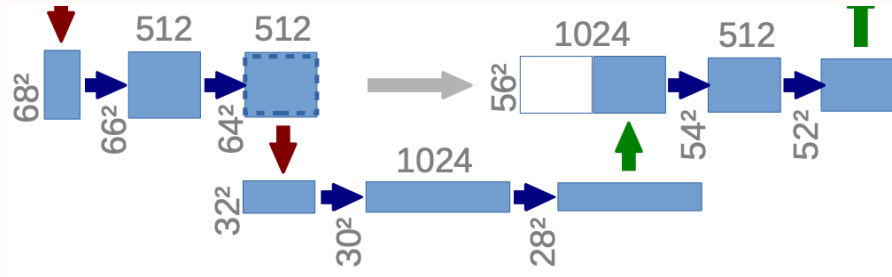
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



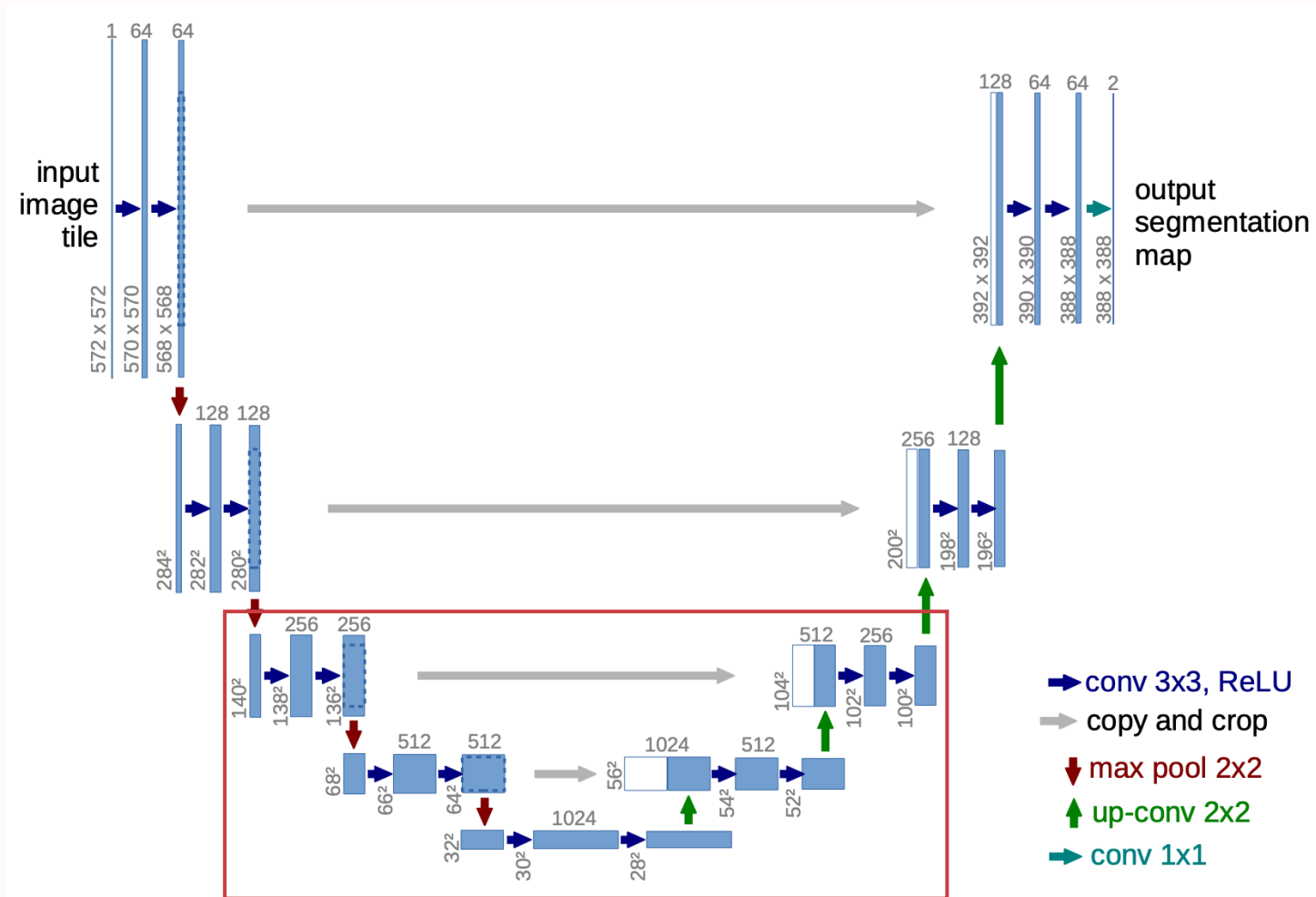
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



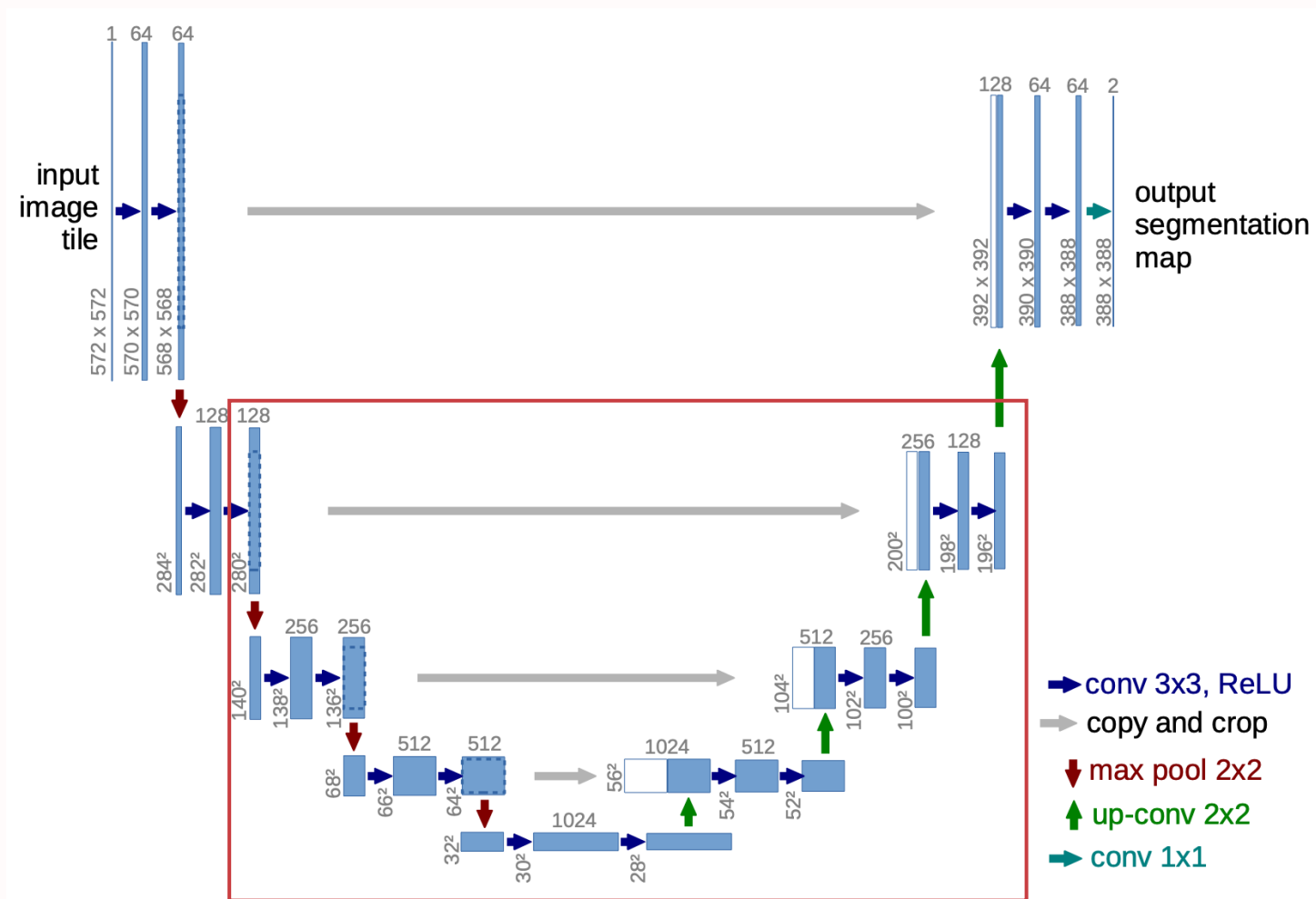
- **Green arrow:** transposed convolution.
- From $28 \times 28 \rightarrow 56 \times 56$.
- **Gray arrow:** *concatenation*.
 - Skip connection from the *previous* feature maps at that resolution.
 - Allows preservation of spatial detail *from that resolution*.

U-Net architecture



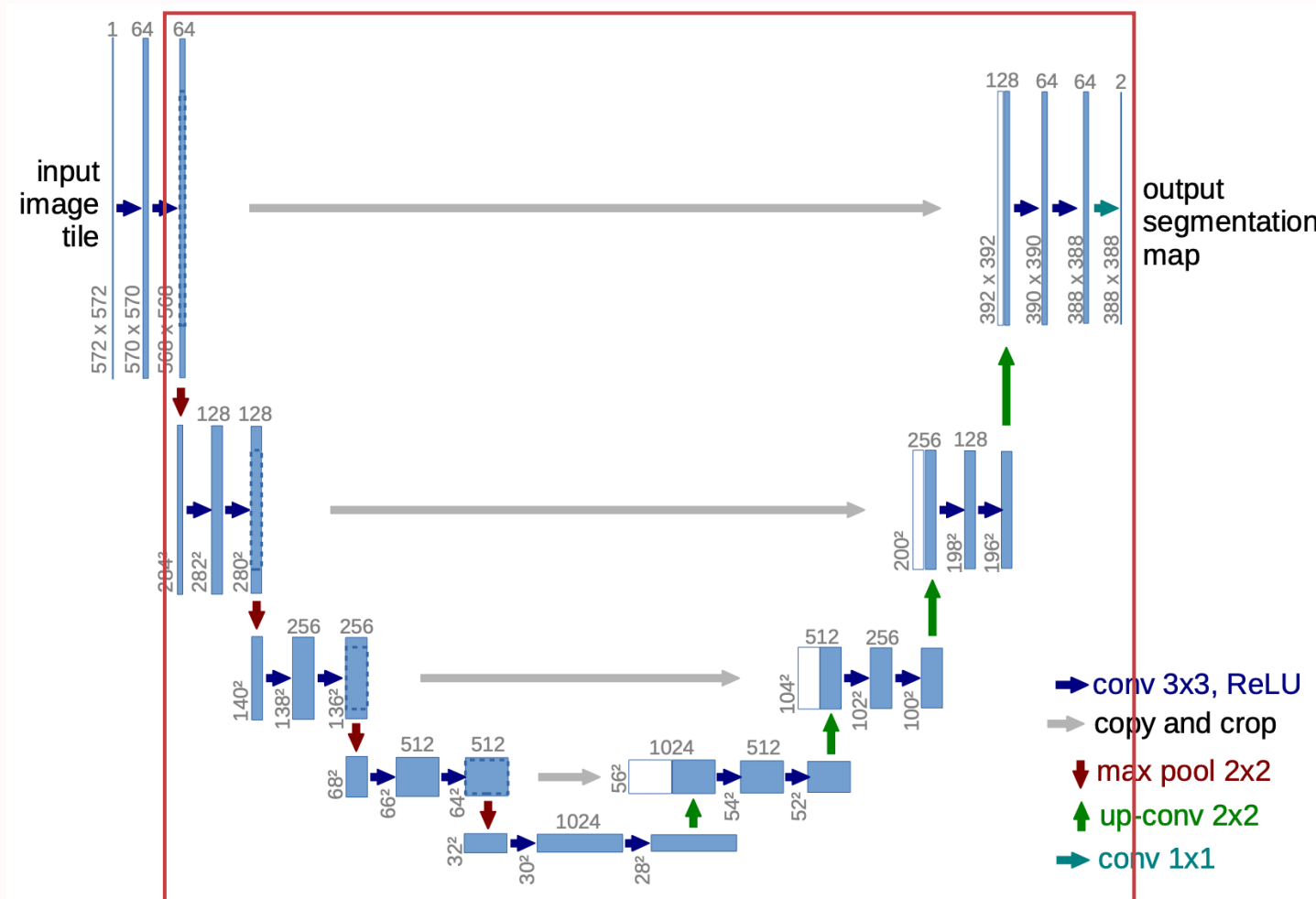
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Net architecture



On the “up” branch, spatial information of the corresponding resolution is always copied over from the “down” branch.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Cham: Springer international publishing, 2015.

U-Nets

- U-Nets are still state-of-the-art on many modern image segmentation tasks.
- A very easy to use implementation—
NNU-Net:
<https://github.com/MIC-DKFZ/nnUNet>

nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Köhler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein

Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany

Abstract. The U-Net was presented in 2015. With its straight-forward and successful architecture it quickly evolved to a commonly used benchmark in medical image segmentation. The adaptation of the U-Net to novel problems, however, comprises several degrees of freedom regarding the exact architecture, pre-processing, training and inference. These choices are not independent of each other and substantially impact the overall performance. The present paper introduces the nnU-Net (“no-new-Net”), which refers to a robust and self-adapting framework on the basis of 2D and 3D vanilla U-Nets. We argue the strong case for taking away superfluous bells and whistles of many proposed network designs and instead focus on the remaining aspects that make out the performance and generalizability of a method. We evaluate the nnU-Net in the context of the Medical Segmentation Decathlon challenge, which measures segmentation performance in ten disciplines comprising distinct entities, image modalities, image geometries and dataset sizes, with no manual adjustments between datasets allowed. At the time of manuscript submission, nnU-Net achieves the highest mean dice scores across all classes and seven phase 1 tasks (except class 1 in BrainTumour) in the online leaderboard of the challenge.

Question & Answer