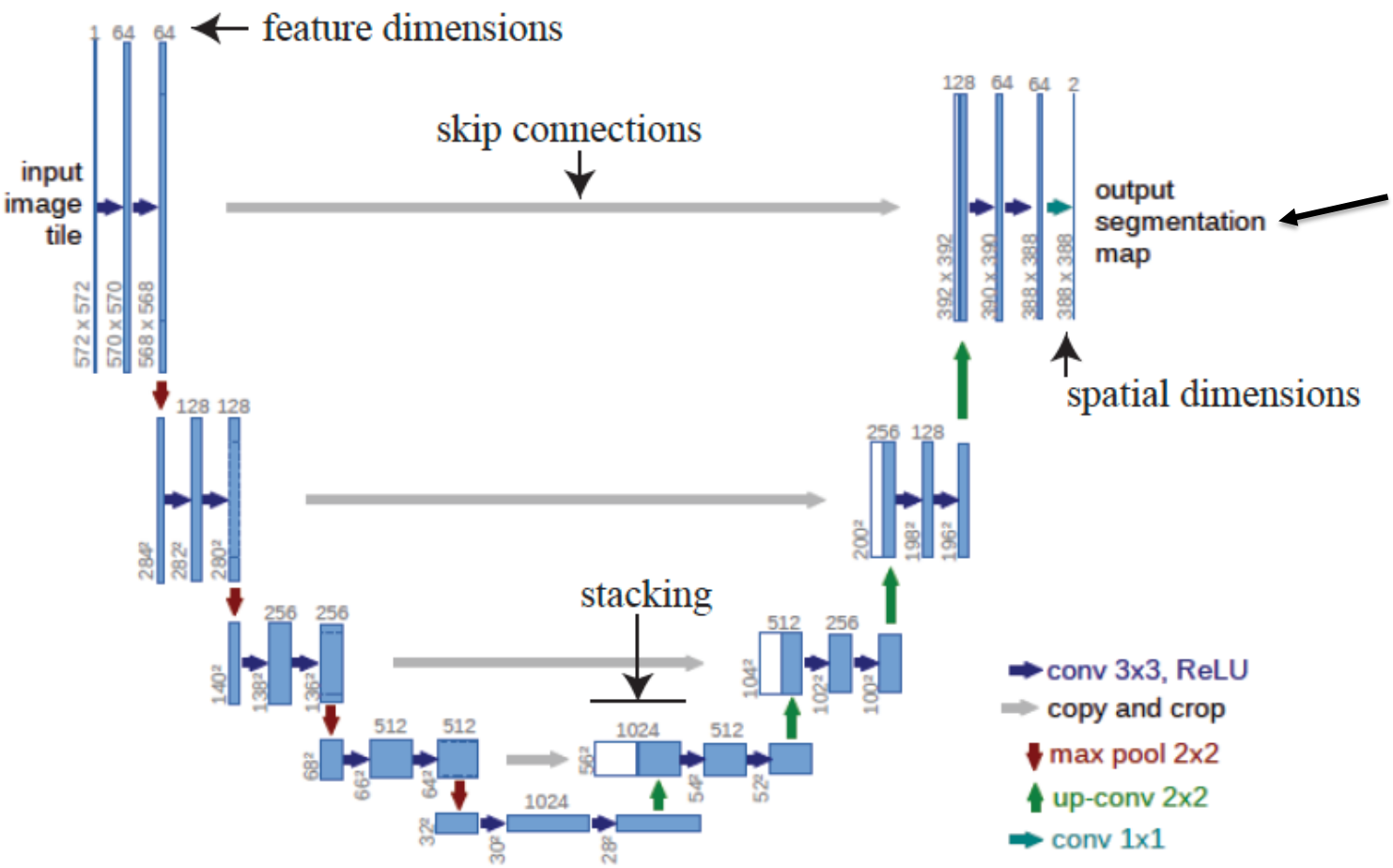
# The original U-net




This doesn't have to be a segmentation

# Curiosities

The U-Net is fully convolutional

- is this a good thing?

Equivariance

- harder than it looks

Other encoders are possible

- but not now

# This U-net is fully convolutional

Consists of convolutional layers, pooling layers, upsampling layers, ReLU's

Consequence:

You can pass in images of variable size

(as long as they're big enough)

Advantage:

Images have variable size in real life

Issue:

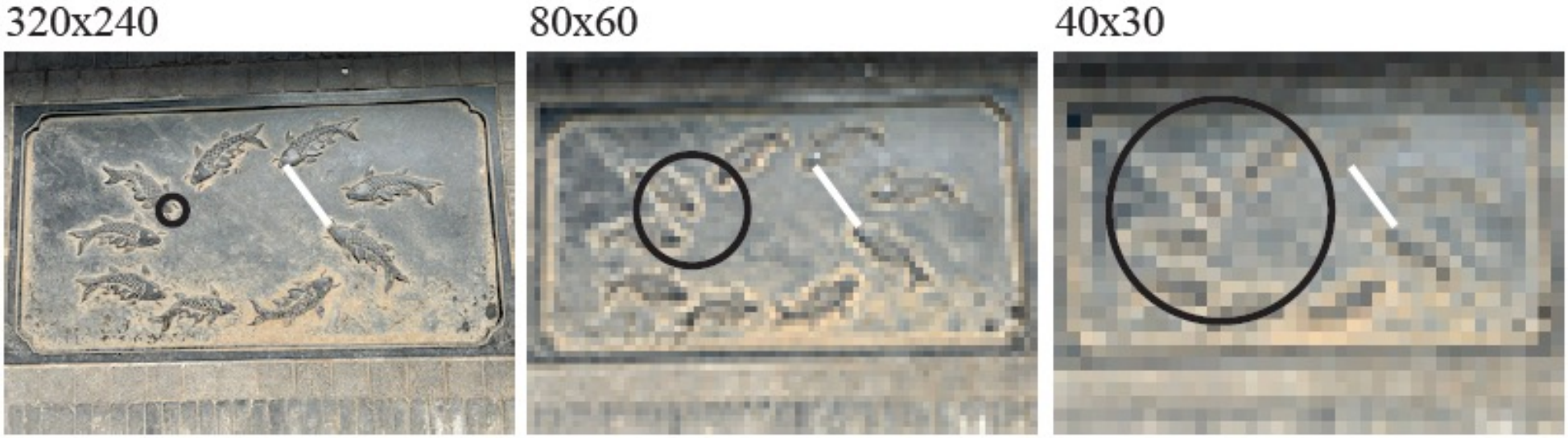
This can make some problems very hard.

# Consequence

The same network feature depends on a different spatial scale in the image for different size images

This can present problems, but doesn't usually

training data seems to deal with this



# Curiosities

The U-Net is fully convolutional

- is this a good thing?

Equivariance

- harder than it looks

Other encoders are possible

- but not now

# Equivariance

Many image to image maps should have properties that are referred to as *equivariance*. Annoyingly, this term is almost always misused. It applies to the behavior of a function under a group action (there is no particular reason to look this up if you don't know what it is), so one should specify what kind of equivariance is intended. Properly, translation equivariance is the shift-invariance of Section 5. As a thought experiment, think about an image to image mapping that accepts an input $\mathcal{I}(x, y)$ and produces an output $\mathcal{O}(x, y) = \mathcal{M}(\mathcal{I})$. Assume that each is defined on the entire plane. Then translation equivariance of the mapping requires that if you translate the input, the output translates. In turn, this means that for any $t_x, t_y$, if

$$\mathcal{I}_2(x, y) = \mathcal{I}_1(x + t_x, y + t_y) \text{ and } \mathcal{O}_1 = \mathcal{M}(\mathcal{I}_1)$$

then

$$\mathcal{M}(\mathcal{I}_2) = \mathcal{O}_1(x + t_x, y + t_y).$$

What you want, but can't get

# You can only get this property if you don’t pad

Which is almost always impractical

Padding makes up for the fact the image isn’t infinite

but means that when you shift, you see new pixels

# What you might be able to get

A somewhat weaker form of translation equivariance is highly desirable for most image to image mappings. Imagine that $\mathcal{I}_1$ and $\mathcal{I}_2$ are finite windows cut out of an infinite image and they overlap. Then it is desirable that $\mathcal{O}_1$ and $\mathcal{O}_2$ agree with one another in the region of overlap. This property does not appear to emerge naturally just as a result of having a large training set **exercises** .

In practice, this property doesn't just appear, and requires special training.

Most SOTA image-image mappers don't have it!

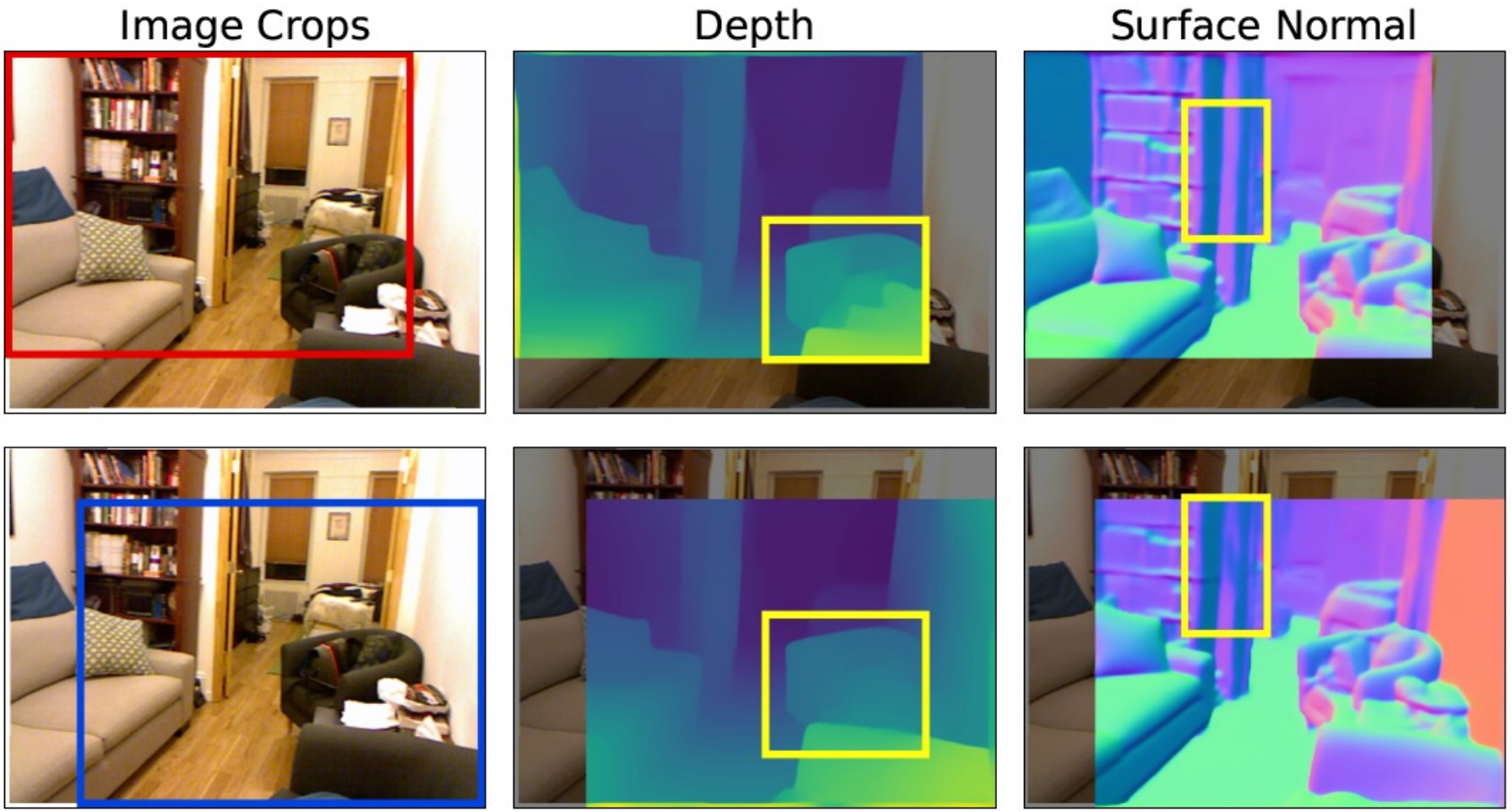


Image Crops
Depth
Surface Normal

# Curiosities

The U-Net is fully convolutional

- is this a good thing?

Equivariance

- harder than it looks

Other encoders are possible

- but not now