Neural Style Transfer

Content Image
(Benedict Cum-purr-scratch)

C

+

Style Image
(Georgia O'Keeffe)

S

=

X

Neural Style Transfer

Content Image
(Benedict Cum-purr-scratch)

C

+

Style Image
(Georgia O'Keeffe)

S

=

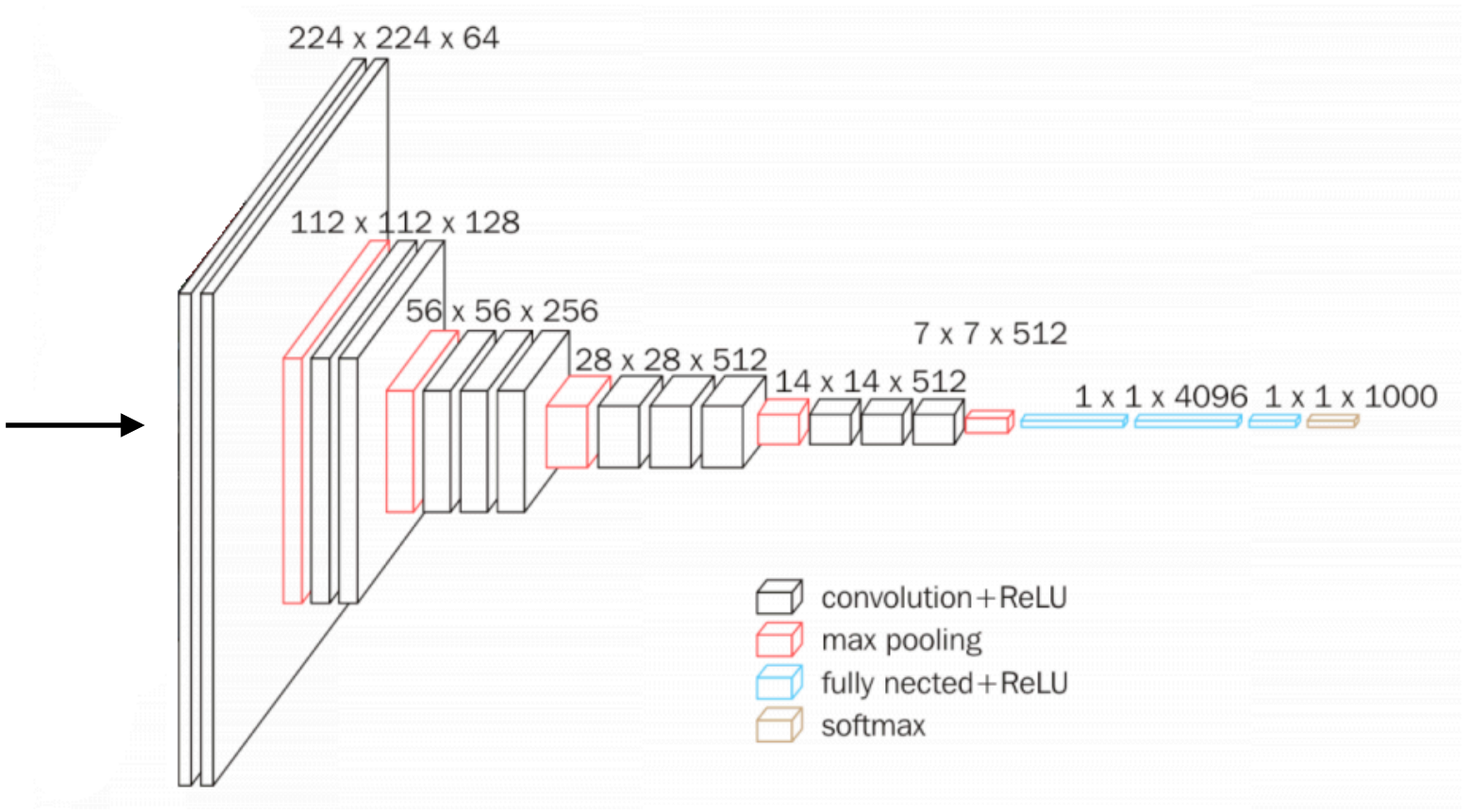
X

We *estimate* the pixel values for the image X so that:

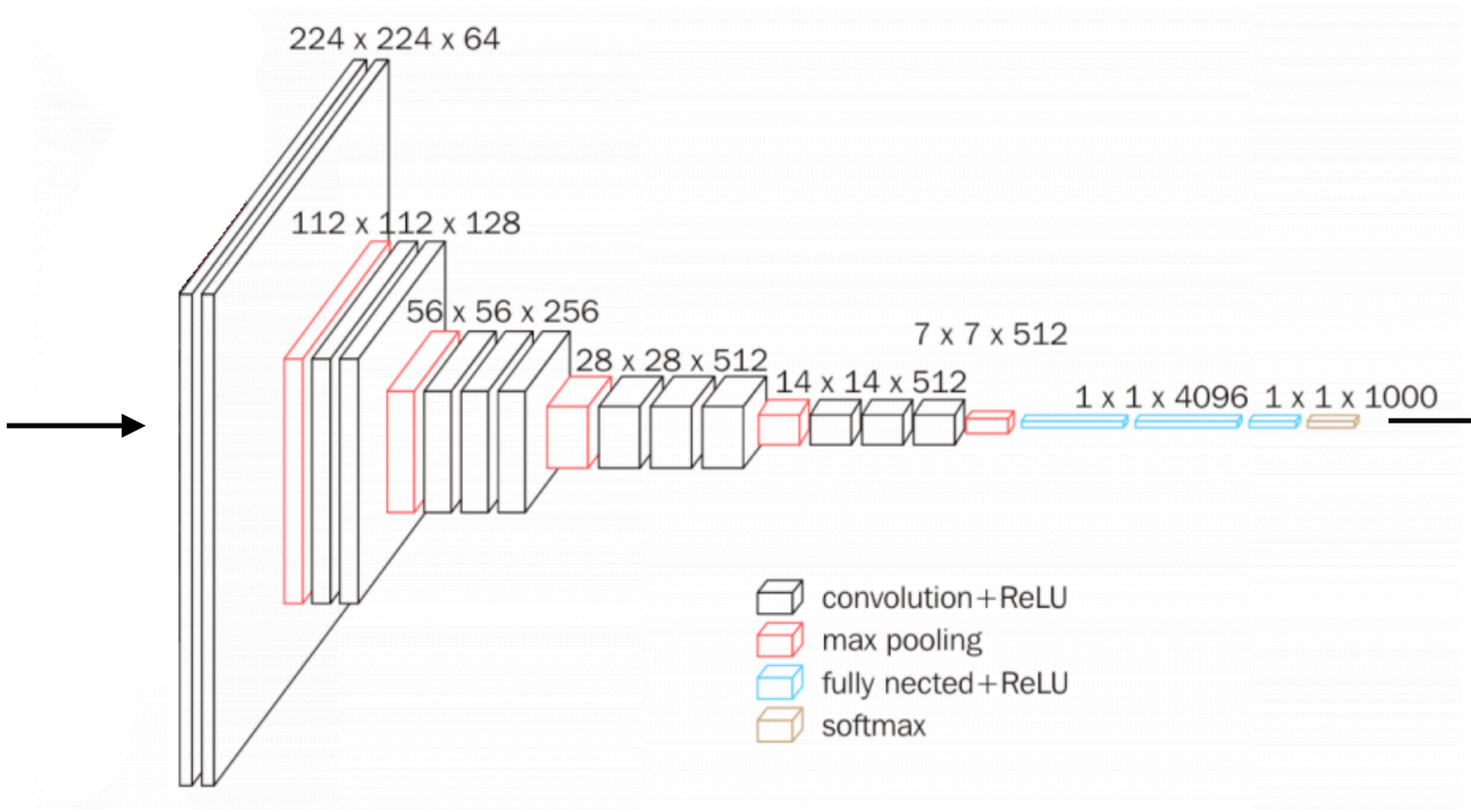
content(C) $\approx$ content(X)

style(S) $\approx$ style(X)

How We've Used Neural Networks So Far



How We've Used Neural Networks So Far



$J(b, w)$

A measure of classification accuracy

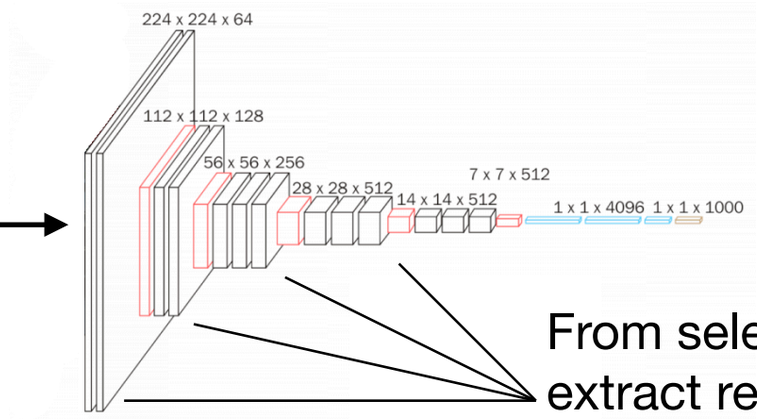
Estimation: given input data, minimize loss function by gradient descent:

$$b^{[l]} \leftarrow b^{[l]} - \alpha \frac{\partial}{\partial b^{[l]}} J(b, w)$$

$$W^{[l]} \leftarrow W^{[l]} - \alpha \frac{\partial}{\partial W^{[l]}} J(b, w)$$

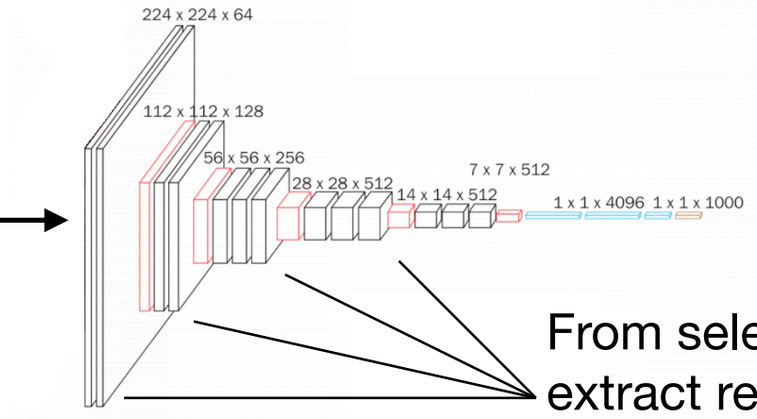
Use of Neural Networks for Style Transfer

S



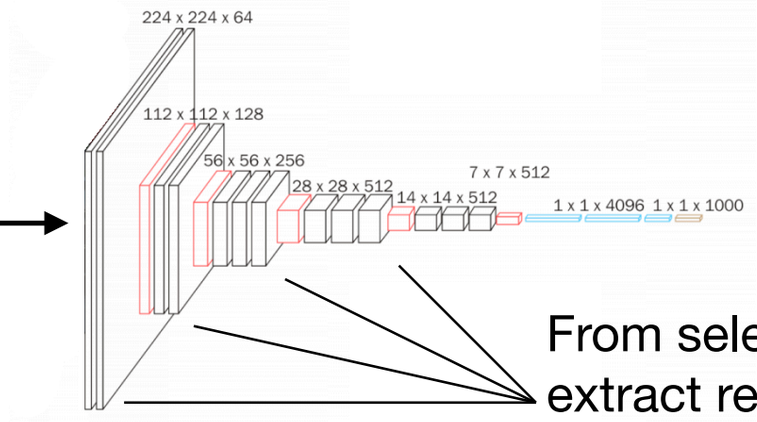
From selected layer activations,
extract representation of style

C



From selected layer activations,
extract representation of content

X



From selected layer activations,
extract representation of style and content

Loss is a measure of similarity of content and style:

$$J(X) = [style(S) - style(X)]^2 + [content(C) - content(X)]^2$$

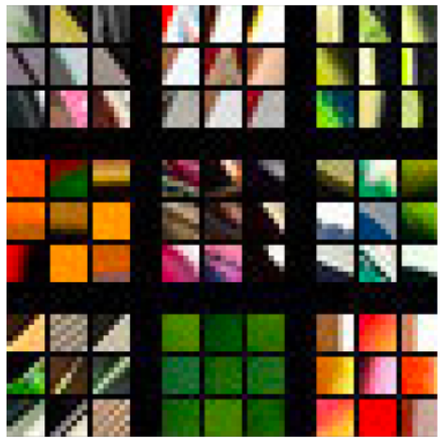
Gradient descent is used to estimate X:

$$X \leftarrow X - \alpha \frac{\partial}{\partial X} J(X)$$

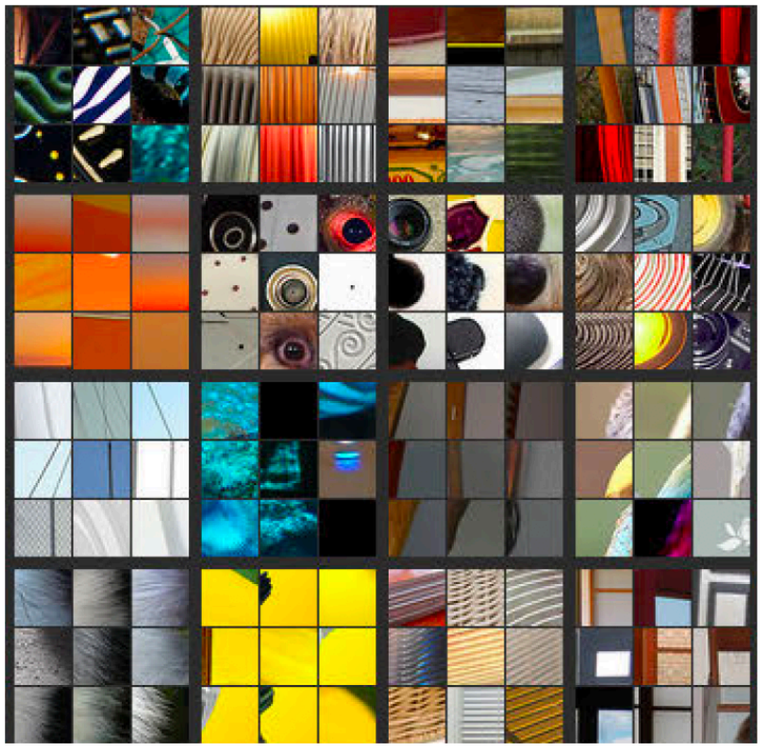
Measuring Content and Style

Figures from Zeiler and Fergus “Visualizing and Understanding Convolutional Networks” (2013)

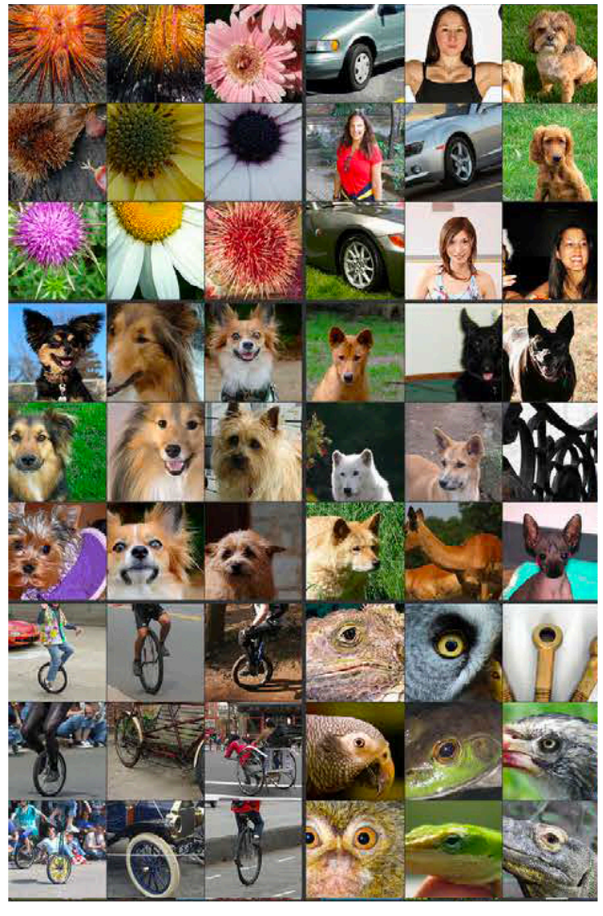
Layer 1



Layer 2



Layer 5

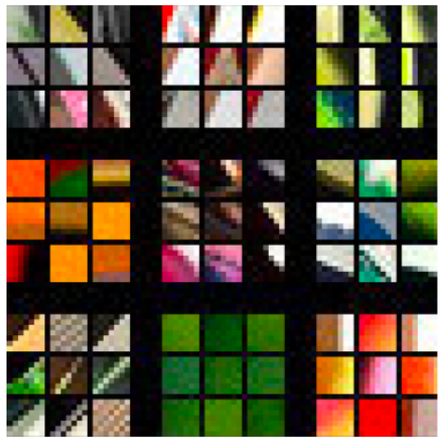


Color and Texture ⟷ **Object Types**

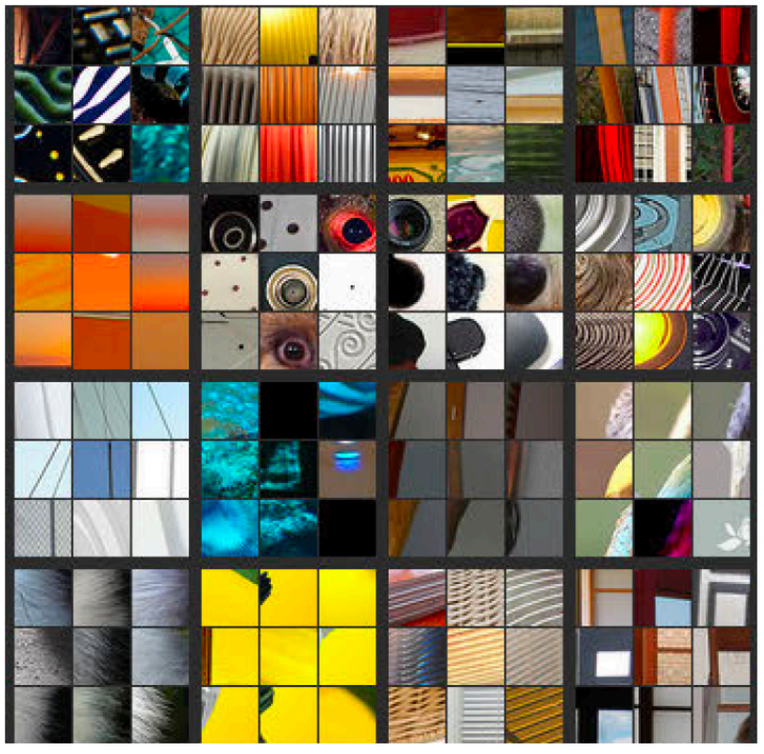
Measuring Content and Style

Figures from Zeiler and Fergus “Visualizing and Understanding Convolutional Networks” (2013)

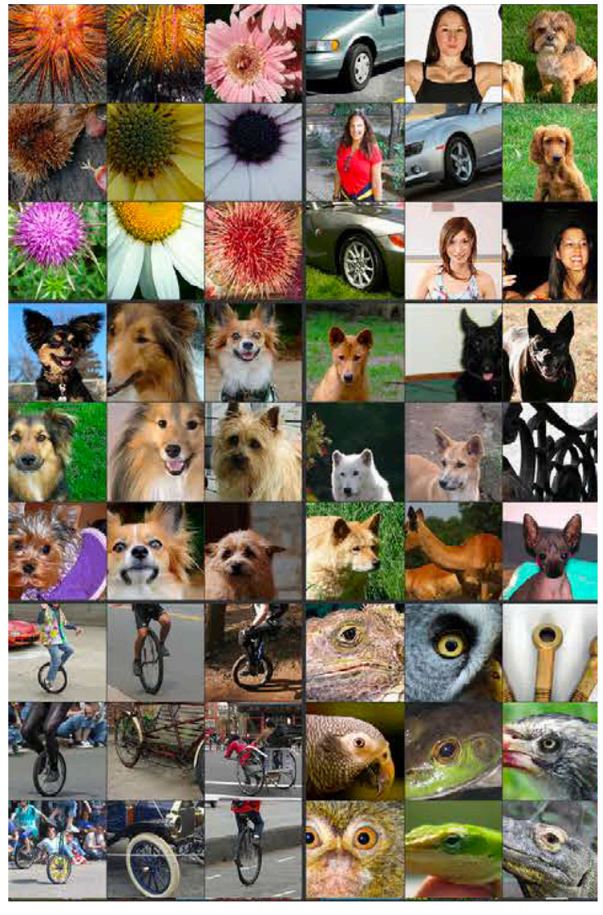
Layer 1



Layer 2



Layer 5

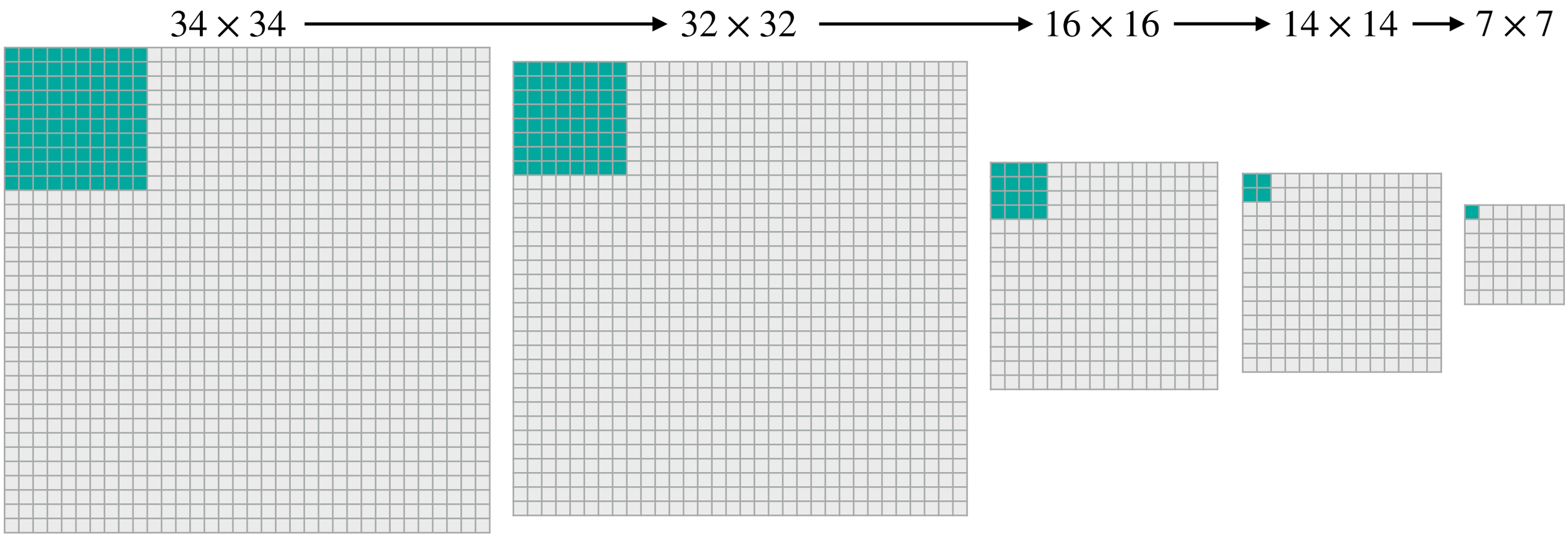


Color and Texture ⟷ **Object Types**

Style — **Content**

Effective Receptive Field

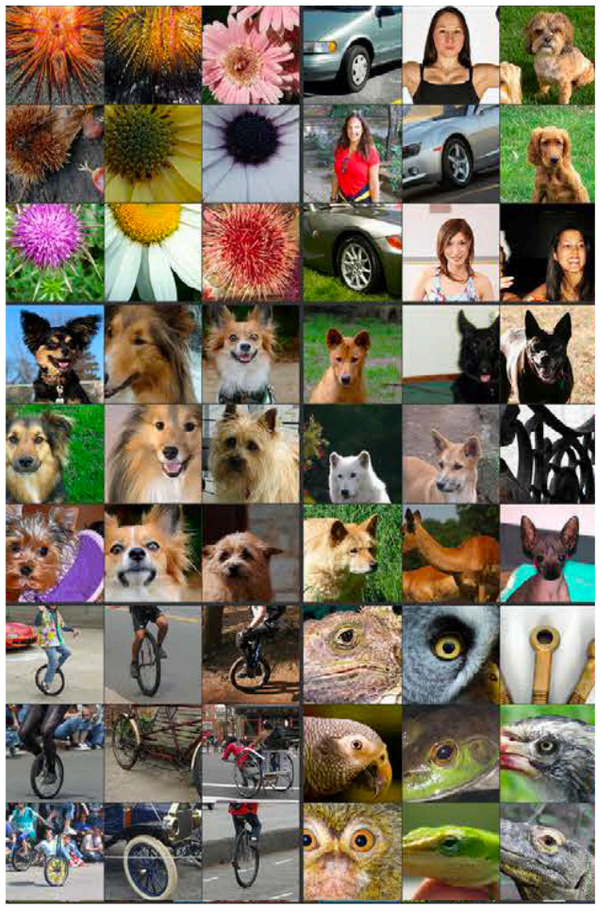
- How many pixels of input unit are used to calculate a given activation in a later layer?



- Activations in intermediate layers tell us about:
  - Use of color and texture at different positions in the input image
  - Types of objects at different positions in the input image

Measuring Content

Layer 5

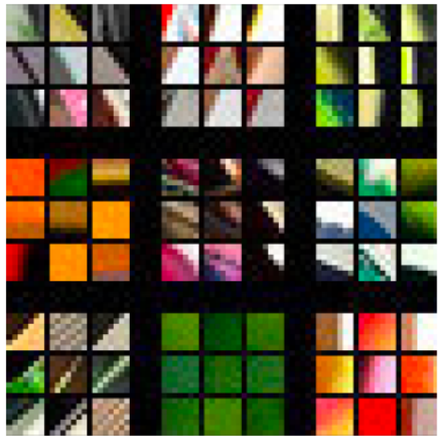


- Pick a layer relatively late in the network
- Its activations tell you what kind of objects are found at specific locations in the input image

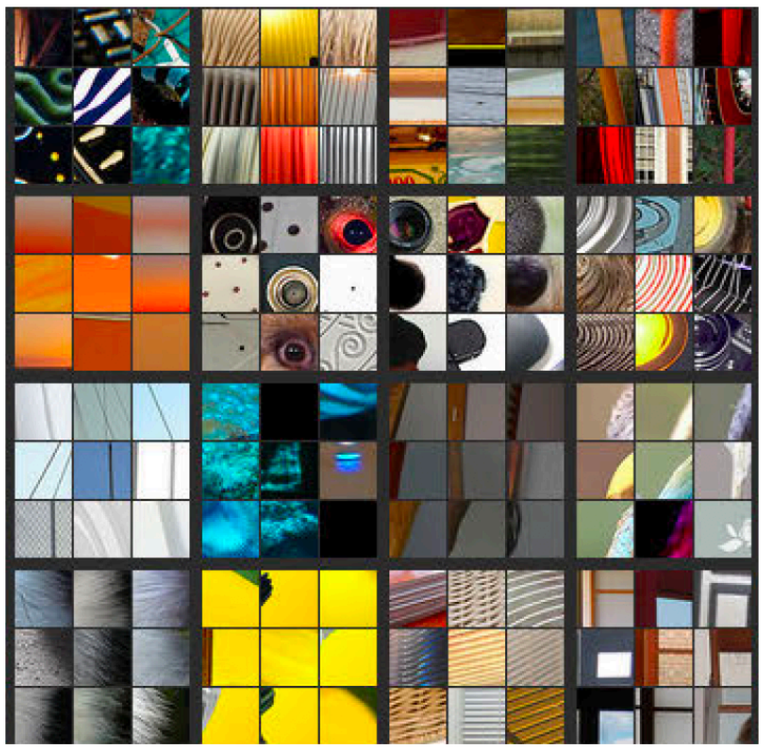
- $content(image) = a^{[l]}$

Measuring Style

Layer 1



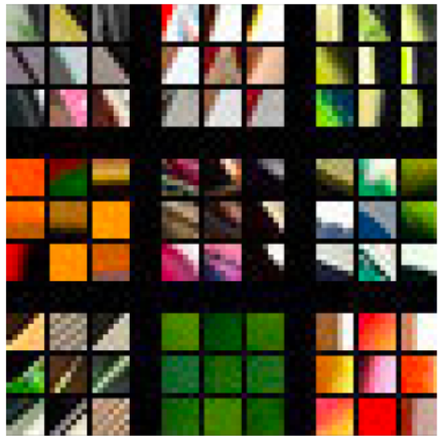
Layer 2



- Pick a layer (or in practice, multiple layers) relatively early in the network
- Its activations tell you what kind of colors and textures are found at specific locations in the input image

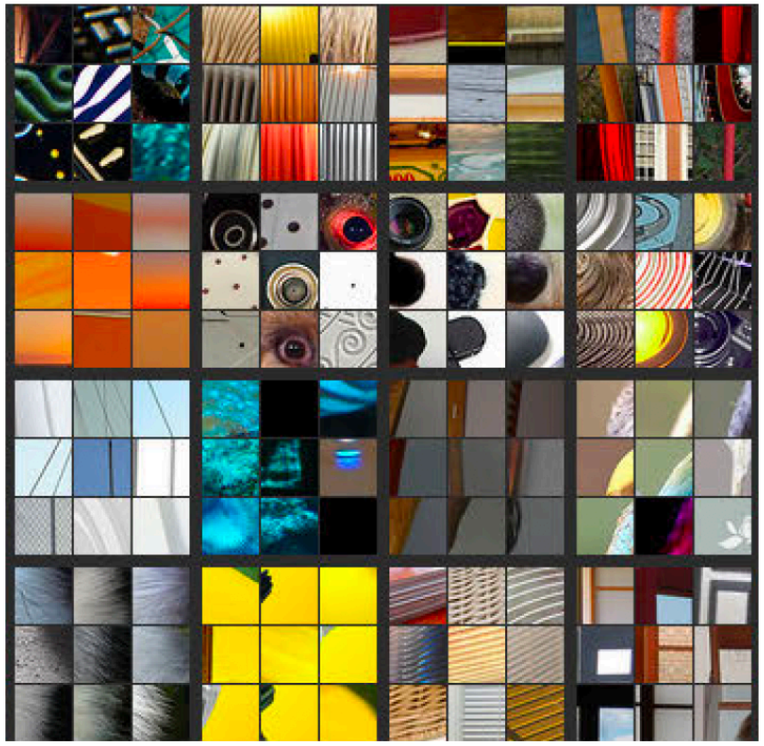
Measuring Style

Layer 1



- Pick a layer (or in practice, multiple layers) relatively early in the network
- Its activations tell you what kind of colors and textures are found at specific locations in the input image

Layer 2



Now what?

- We don't necessarily want to match use of color and texture in specific positions of the style image
- We want to match use of color and texture across the full image