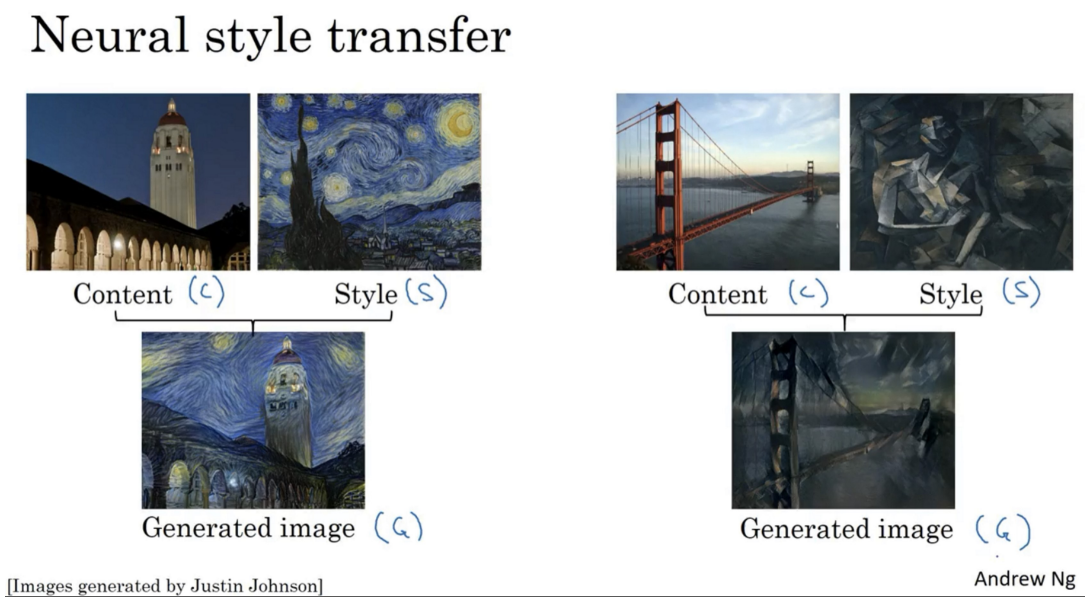
Neural Style Transfer

Wednesday, September 9, 2020 11:54 AM

① What is Neural Style Transfer?



→ Takes content image (C) and style image (S) to generate an image (G) of C with the style of S.

→ We will need to look at the features extracted at various layers of a ConvNet

→ The idea of using a network trained on a different task and applying it to a new task is called transfer learning

② What are deep ConvNets learning?

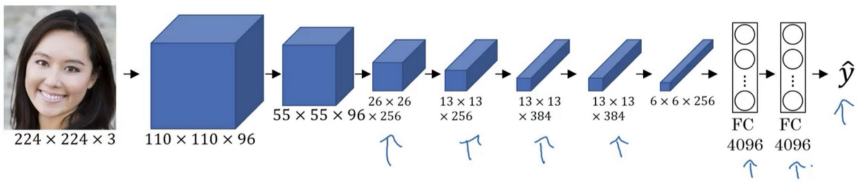
→ We pick a layer and choose a hidden unit

→ check the nine image patches that maximize the unit's activation

- The initial layers will see a relatively small portion of the image and we get larger images in the deeper layers

→ Repeat for other units and layers

Visualizing what a deep network is learning



Pick a unit in layer 1. Find the nine image patches that maximize the unit's activation.

Repeat for other units.

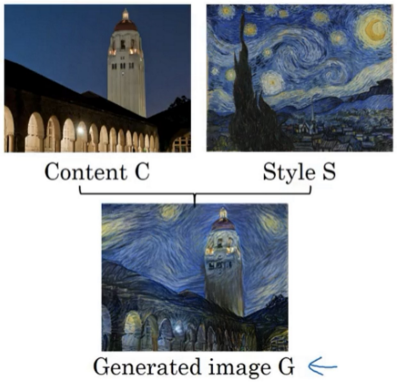
[Zeiler and Fergus., 2013, Visualizing and understanding convolutional networks]

Andrew Ng

→ The initial layers learn low level representations and the deeper layer learn more complicated features

Layer 1 | Layer 2 | Layer 3 | Layer 4 | Layer 5

③ Cost Function



→ how similar C & G are

$$J(G) = J_{content}(C, G) + J_{style}(S, G)$$

→ how similar S & G are

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$

→ • Initialize G randomly
G: 100x100x3 → o/p image

• Use Gradient descent to minimize G

• $G = G - \frac{\partial}{\partial G} J(G)$

④ Content Cost Function

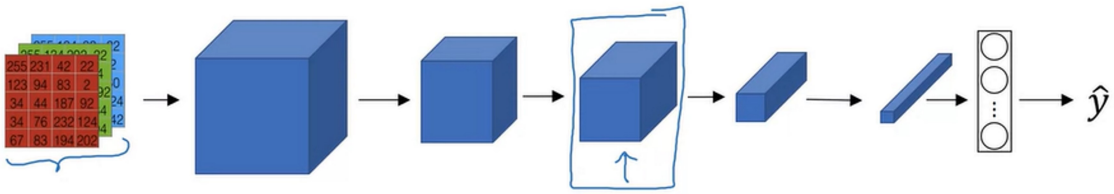
- If we choose hidden layer l to compute content cost,
  - → If l is small (eg: layer 1 or 2), we will force the network to output similar image as content
  - → Usually l is not too shallow or not too deep
- We use pretrained convNets (eg: VGG)
- Let $a^{[l](C)}$ and $a^{[l](G)}$ be the activation of layer l on the image
- If $a^{[l](C)}$ and $a^{[l](G)}$ are similar, both images have similar content

$$J_{content}(C, G) = \frac{1}{2} \| a^{[l](C)} - a^{[l](G)} \|^2$$

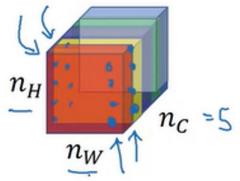
⑤ Style Cost Function

- Meaning of 'style' of an image

Meaning of the "style" of an image

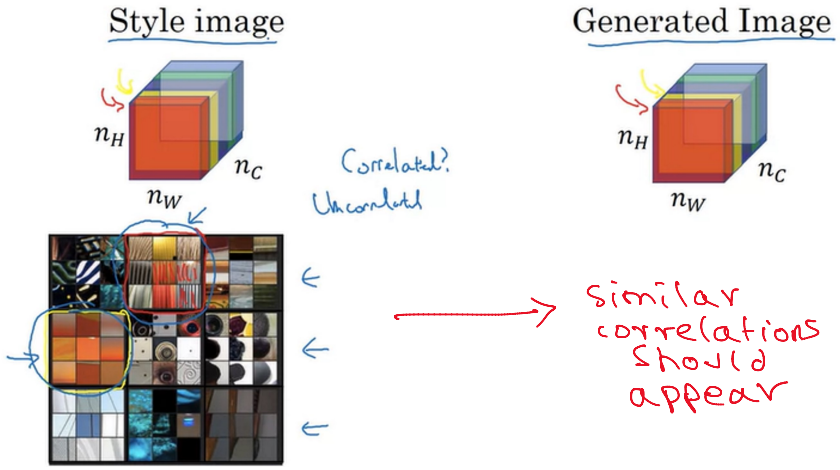


Say you are using layer l's activation to measure "style."
Define style as correlation between activations across channels.



How correlated are the activations across different channels?

- → Correlated means if a value appeared in a specific channel, a specific value will appear too
- → Uncorrelated means if a value appeared in a specific channel doesn't mean that another value will appear
- The correlation of style image channels should appear in the generated image channels



- Style Matrix

Let $a^{[l]}_{i,j,k}$ = activation at (i,j,k). $h\ w\ c$

$G^{[l]}$ is $n_c^{[l]} \times n_c^{[l]}$ (comparing every channel with other channels)

$$G^{[l](s)}_{kk'} = \sum_{i=1}^{n_H^{[l]}} \sum_{j=1}^{n_W^{[l]}} a^{[l](s)}_{ijk}\, a^{[l](s)}_{ijk'}$$

$$G^{[l](G)}_{kk'} = \sum_{i=1}^{n_H^{[l]}} \sum_{j=1}^{n_W^{[l]}} a^{[l](G)}_{ijk}\, a^{[l](G)}_{ijk'}$$

Gram matrix

$k = 1, \ldots, n_c^{[l]}$

$$\Rightarrow J^{[l]}_{style}(S,G) = \frac{1}{2 n_H^{[l]} n_W^{[l]} n_c^{[l]}} \left\| G^{[l](S)} - G^{[l](G)} \right\|_F^2$$

$$= \frac{1}{2 n_H^{[l]} n_W^{[l]} n_c^{[l]}} \sum_k \sum_{k'} \left(G^{[l](S)}_{kk'} - G^{[l](G)}_{kk'} \right)^2$$

- Style cost function

$$J^{[l]}_{style}(S,G) = \frac{1}{2 n_H^{[l]} n_W^{[l]} n_c^{[l]}} \sum_k \sum_{k'} \left(G^{[l](S)}_{kk'} - G^{[l](G)}_{kk'} \right)^2$$

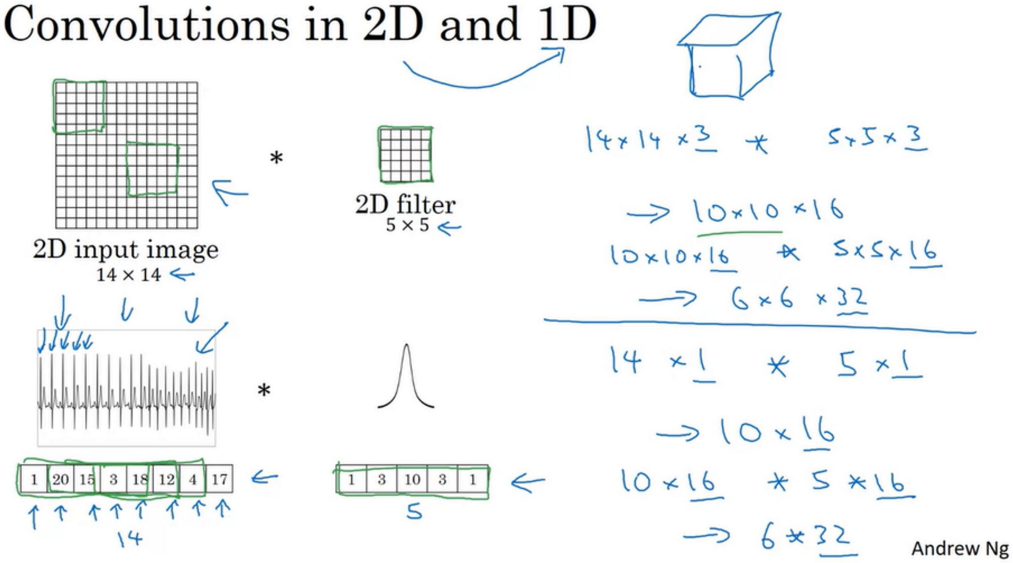
$$\boxed{J_{style}(S,G) = \sum_l \lambda^{[l]} J^{[l]}_{style}(S,G)}$$

- Overall cost function for Neural style Transfer is

$$J(G) = \alpha J_{content}(C,G) + \beta J_{style}(S,G)$$

(6) 1D and 3D Generalizations of ConvNets

→ 2D and 1D Convolutions



- RNNs are used for 1D data / sequential data but CNNs can also be used

→ 3D data

- CAT scans have slices of our body as images which can be used to make a 3D image of our brain
- Movie frames can be lined up in 3D to detect motion, etc