[SK Academy - 2019 AI Course]

Neural Style Transfer

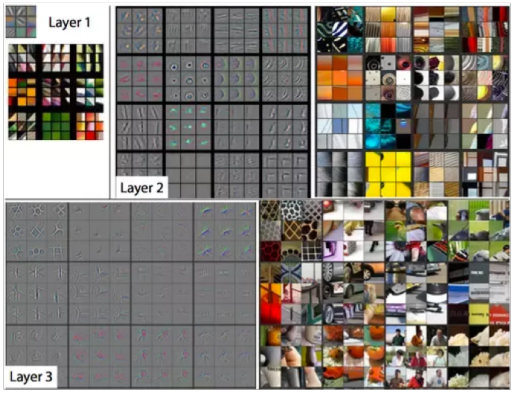
Heung-Il Suk

hisuk@korea.ac.kr

http://milab.korea.ac.kr

Department of Brain and Cognitive Engineering,
Korea University

CNNs Feature Representations



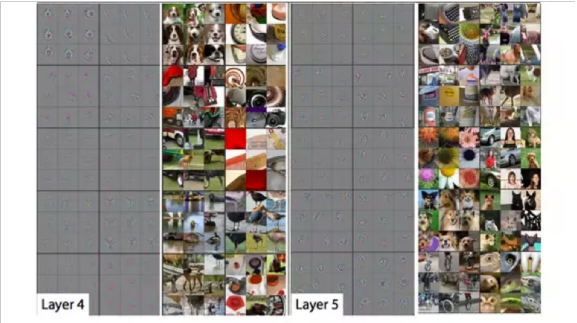


Figure 2. Visualization of features in a fully trained model. For layers 2-5 we show the top 9 activations in a random subset of feature maps across the validation data, projected down to pixel space using our deconvolutional network approach. Our reconstructions are *not* samples from the model: they are reconstructed patterns from the validation set that cause high activations in a given feature map. For each feature map we also show the corresponding image patches. Note: (i) the the strong grouping within each feature map, (ii) greater invariance at higher layers and (iii) exaggeration of discriminative parts of the image, e.g. eyes and noses of dogs (layer 4, row 1, cols 1). Best viewed in electronic form.

DeepDream: Feature Amplifying [Mordvintsev *et al.*, 2015]

Understanding what exactly goes on at each layer

- to turn the network upside down and ask it to enhance an input image in such a way as to elicit a particular interpretation

- Setting the gradient of a chosen layer l equal to its activation
- Update the image via gradient ascent + backprop

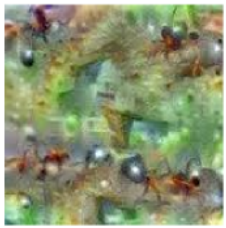
$$I^*(l) = \underset{I}{\operatorname{argmax}} \sum_{i}^{n_c} \left\{ \phi_i^{(l)}(I) \right\}^2$$

https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Machine Intelligence Lab

Hartebeest

Measuring Cup



Ant

Starfish

Anemone Fish

Banana

Parachute

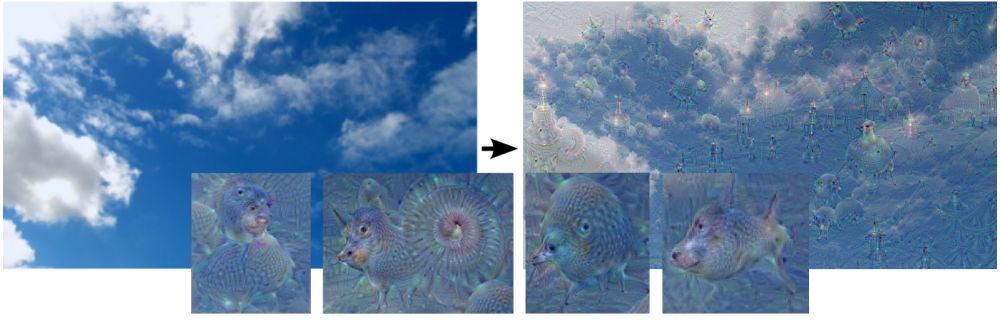
Screw

- No picture of a dumbbell is complete without a muscular weightlifter there to lift them
- The network failed to completely distill the essence of a dumbbell
- Maybe it's never been shown a dumbbell without an arm holding it.

- Visualization can help us correct these kinds of training mishaps.

Inceptionism

- a qualitative sense of the level of abstraction that a particular layer has achieved in its understanding of images



(Feedback loop: if a cloud looks a little bit like a bird, the network will make it look more like a bird)

Applying the algorithm iteratively on its own outputs and some zooming after each iteration

Feature Inversion [mahendran and Vedaldi, 2015]

Given a CNN feature vector ϕ_0 for an image, find a new image

- matches the given feature vector
- "looks natural" (image prior regularization)

$$I^* = \underset{I \in \mathbb{R}^{H \times W \times C}}{\operatorname{argmin}} \mathcal{L}\left(\phi\left(I\right), \phi_0\right) + \lambda \mathcal{R}\left(I\right)$$

$$\mathcal{L}\left(\phi\left(I\right), \phi_0\right) = \left\|\phi\left(I\right) - \phi_0\right\|^2$$

$$\mathcal{R}\left(I\right) = \sum_{i,j} \left[\left(I(i, j+1) - I(i,j)\right)^2 + \left(I(i+1, j) - I(i,j)\right)^2\right]^{\frac{\beta}{2}}$$

(total variation regularizer for spatial smoothness)

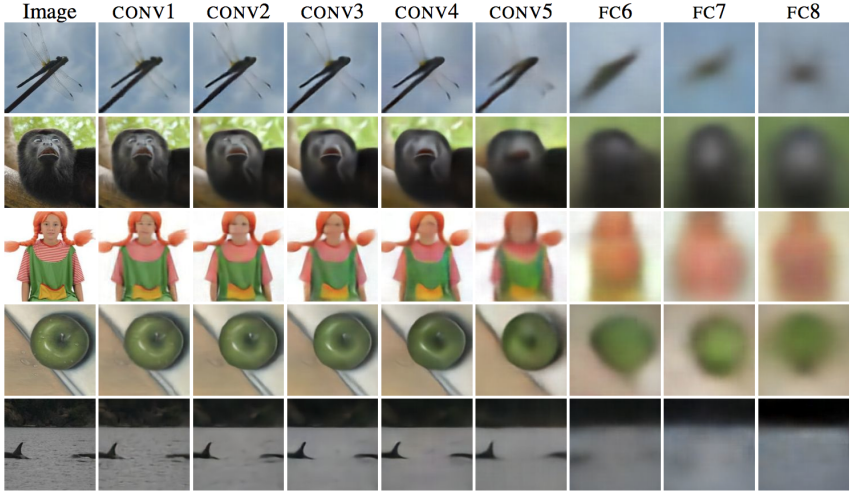


Figure 1: Reconstructions from different layers of AlexNet.

Image source: https://computervisionblog.wordpress.com/2016/01/

Neural Style Transfer [Gatys *et al.*, 2016]

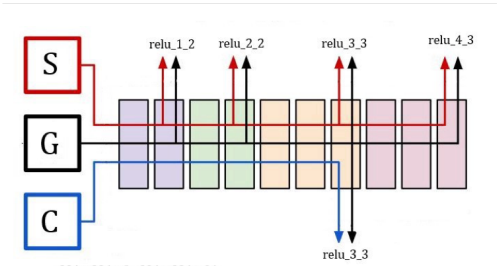
Machine Intelligence Lab

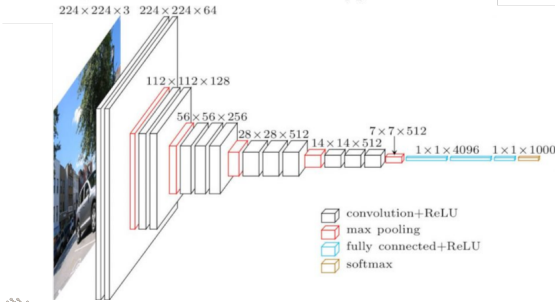
Loss Function

Similarity between content/style and stylized images

- Content: content image and generated image are similar w.r.t. their content and not style
- Style: generated image only inherits similar style representation from style image and not the entire style image itself

$$\mathcal{L}(C, S, G) = \alpha \mathcal{L}_c(C, G) + \beta \mathcal{L}_s(S, G)$$





Content loss

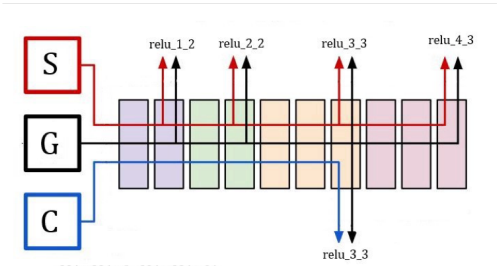
$$\mathcal{L}_c(C,G) = \frac{1}{2}\sum_l \sum_{i,j} \left(\phi_C^{(l)}(i,j) - \phi_G^{(l)}(i,j)\right)^2$$

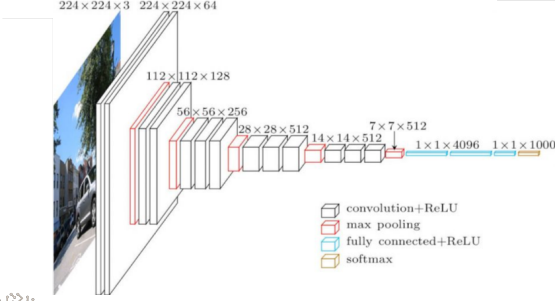
Style loss

$$\mathcal{L}_s(S,G) = ?$$

- can't just use differences in activations
- correlation between activations across different channels of the same layer

Machine Intelligence Lab





Content loss

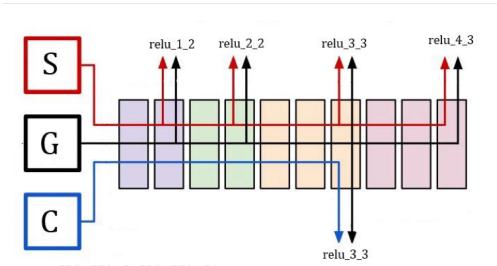
$$\mathcal{L}_c(C,G) = \frac{1}{2}\sum_l \sum_{i,j} \left(\phi_C^{(l)}(i,j) - \phi_G^{(l)}(i,j)\right)^2$$

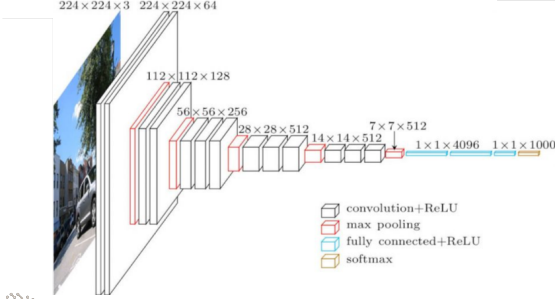
Style loss

$$\mathcal{L}_s(S,G) = ?$$

- can't just use differences in activations
- correlation between activations across different channels of the same layer

Machine Intelligence Lab





Content loss

$$\mathcal{L}_c(C,G) = \frac{1}{2}\sum_l \sum_{i,j} \left(\phi_C^{(l)}(i,j) - \phi_G^{(l)}(i,j)\right)^2$$

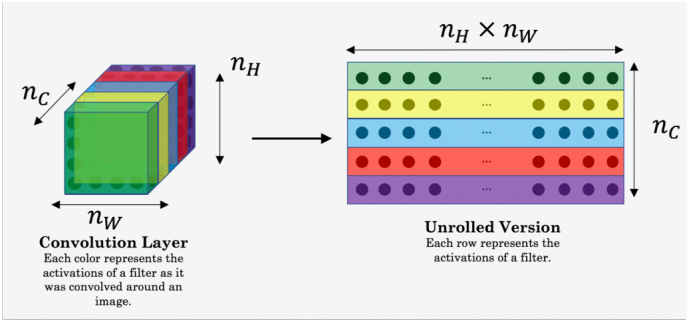
Style loss

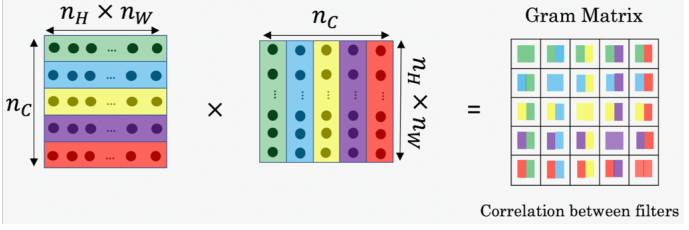
$$\mathcal{L}_s(S,G) = ?$$

- can't just use differences in activations
- correlation between activations across different channels of the same layer

Machine Intelligence Lab

Gram matrix


$n_H \times n_W$
n_C
n_H
n_W
n_C
Convolution Layer
Each color represents the activations of a filter as it was convolved around an image.
Unrolled Version
Each row represents the activations of a filter.


$n_H \times n_W$
n_C
$\times$
n_C
$n_H \times n_W$
=
Gram Matrix
Correlation between filters

Style loss

- minimizing the difference between S and G in style

$$\mathcal{L}_{\mathsf{GM}}(S,G,l) = \frac{1}{4\left(n_c^{(l)}\right)^2\left(n_H^{(l)} \times n_W^{(l)}\right)^2} \sum_{i,j}\left[\mathsf{GM}_S^{(l)}(i,j) - \mathsf{GM}_G^{(l)}(i,j)\right]^2$$

- using multiple activation layers
- assigning different weights

$$\mathcal{L}_s(S,G) = \sum_{l=0}^{L} w^{(l)} \times \mathcal{L}_{\mathsf{GM}}(S,G,l)$$

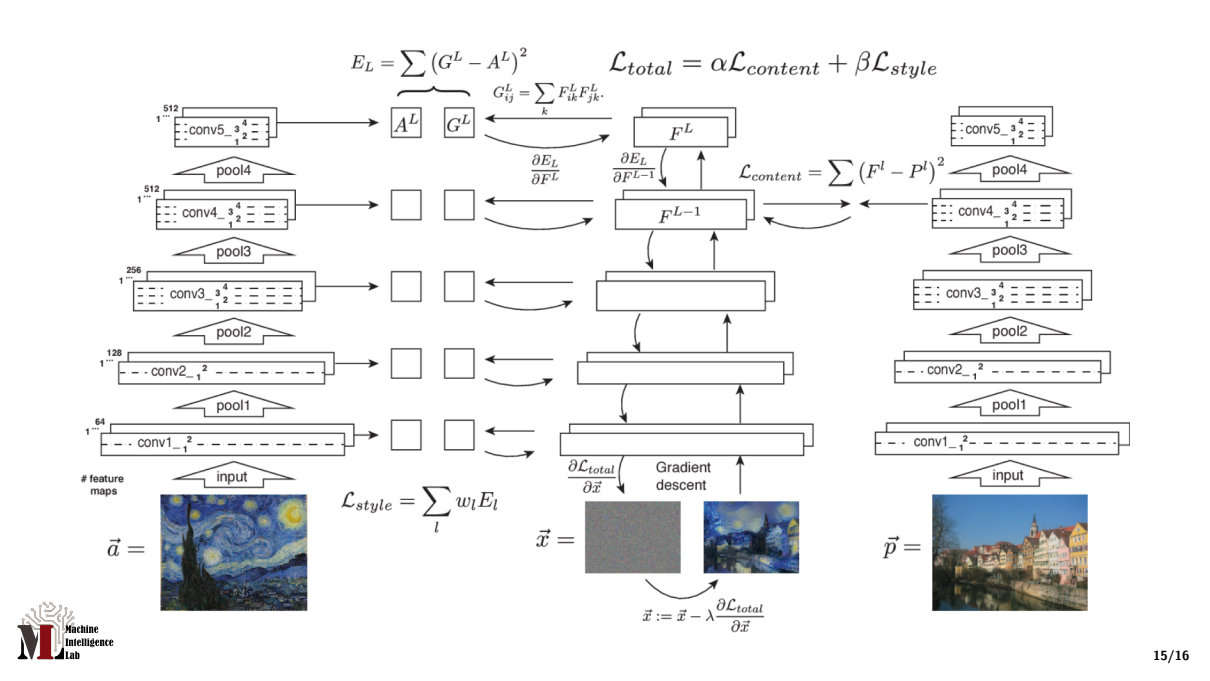
Style loss

- minimizing the difference between S and G in style

$$\mathcal{L}_{\mathsf{GM}}(S,G,l) = \frac{1}{4\left(n_c^{(l)}\right)^2\left(n_H^{(l)} \times n_W^{(l)}\right)^2} \sum_{i,j}\left[\mathsf{GM}_S^{(l)}(i,j) - \mathsf{GM}_G^{(l)}(i,j)\right]^2$$

- using multiple activation layers
- assigning different weights

$$\mathcal{L}_s(S,G) = \sum_{l=0}^{L} w^{(l)} \times \mathcal{L}_{\mathsf{GM}}(S,G,l)$$



Machine Intelligence Lab

References

 L. Gatys *et al.*, "Image Style Transfer Using Convolutional Neural Networks," CVPR, 2016

Thank you
for your attention!!!

(Q & A)

hisuk (AT) korea.ac.kr

http://milab.korea.ac.kr

Machine Intelligence Lab