Lecture 12: Visualizing and Understanding

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 1

Last time: Lots of Computer Vision Tasks

Classification

Semantic Segmentation

Object Detection

Lecture 12 -

Instance Segmentation

May 11, 2023



Today: What's going on inside ConvNets?

This image is CC0 public domain



Input Image: 3 x 224 x 224

\dense 192 2048 2048 128 128 224 dense densé 1000 192 192 128 Max 2048 2048 pooling Max 128 Max pooling pooling

Class Scores: 1000 numbers

May 11, 2023

What are the intermediate features looking for?

Lecture 12 - 3

Krizhevsky et al, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS 2012. Figure reproduced with permission.

Visualizing what models have learned:

- Visualizing filters
- Visualizing final layer features
- Visualizing activations

Understanding input pixels

- Identifying important pixels
- Saliency via backprop
- Guided backprop to generate images
- Gradient ascent to visualize features

Adversarial perturbations

Style transfer

- Features inversion
- Deep dream
- Texture synthesis
- Neural style transfer

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Today's agenda

May 11, 2023

Visualizing what models have learned:

- Visualizing filters
- Visualizing final layer features
- Visualizing activations

Understanding input pixels

- Identifying important pixels
- Saliency via backprop
- Guided backprop to generate images
- Gradient ascent to visualize features

Adversarial perturbations

Style transfer

- Features inversion
- Deep dream
- Texture synthesis
- Neural style transfer

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Today's agenda

May 11, 2023

Interpreting a Linear Classifier: Visual Viewpoint







Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 6

First Layer: Visualize Filters



AlexNet: 64 x 3 x 11 x 11

Krizhevsky, "One weird trick for parallelizing convolutional neural networks", arXiv 2014 He et al, "Deep Residual Learning for Image Recognition", CVPR 2016 Huang et al, "Densely Connected Convolutional Networks", CVPR 2017

Ma Max pooling Ma

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 7

First Layer: Visualize Filters



Maj Max pooling Ma

AlexNet: 64 x 3 x 11 x 11

Krizhevsky, "One weird trick for parallelizing convolutional neural networks", arXiv 2014 He et al, "Deep Residual Learning for Image Recognition", CVPR 2016 Huang et al, "Densely Connected Convolutional Networks", CVPR 2017

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 8

Visualize the filters/kernels (raw weights)

We can visualize filters at higher layers, but not that interesting

(these are taken from ConvNetJS CIFAR-10 demo)

Weights: 医骨骨骨 网络白垩纪 医白垩 网络白垩 网络白垩 Weights:

(我们是这些是我们是我们的这些你。)(我们是我们的你们的是我们的我们的。)(我们不能能能得了。 如果我是这些是我是)(我们会是你会是我们的是你的你的?)(我们是我们是我们能能能能能能能。 國政)(在市市回臺的各國的局面主要會會支約)(總統國際國際國際國際國際國際國際國際國際的會 而變與自由於法學交形型)(法等法用的单称所得解释自己的学校)(開始用的運動局部的調解 國家設備)(國際醫院總督醫院總務醫院委員会部)(以非过來以中國醫院和中國醫院和中國部)(總領國 layer 2 weights 新設設施設設設施設(新設施設施設施設施設施設施設施設施設)(学校設定部務法定部 要要利用書類)(物物電波設定的支援機構的ななない)(ななななななののなかの言う)(国 新新国美国新教会会学家新教会会》(自由会会会成美国大学会会会会)(知行和政府会会成 2012年1月1日日月期)

Weights: (这是当然就是这些我的故事是是要要求自己的)(我是这些还是是要是我们可能在这些我们的是)(國際国家商業商業局部委員會総合商業商業商業)(目前活用行業等商店等を受加率を通常する)(約)(医医后腺性医筋脂肪和原因医尿尿酸苷医尿)(與医胆原的医尿原尿及胆液和胆酸和胆酸 layer 3 weights 主要權)(總基理總道總導陸國際委託物助委務局委活動)(出於法律建築務委員会規模任務部 医氏病炎(清楚局部在部局部的治疗治疗者的治疗治疗法的治疗法的治疗法的治疗法 20 x 20 x 7 x 7 周期新聞単年)(法院長可認知能力を設定を登録したのでの)(は当時にあたからかからない)な 非非非可能的)(你有那些原始和实际和实际的原则是是是是有了)(你不过我们的是是我们的 ※第四考示型なり(通常を定定市場を登録市場の通知の通知)(ご学校構成事業部務の通道

layer 1 weights

16 x 3 x 7 x 7

20 x 16 x 7 x 7

May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Last Layer

FC7 layer



4096-dimensional feature vector for an image (layer immediately before the classifier)

Run the network on many images, collect the feature vectors

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 10

Last Layer: Nearest Neighbors

Recall: Nearest neighbors in <u>pixel</u> space



Krizhevsky et al, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS 2012. Figures reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 11

Last Layer: Nearest Neighbors

4096-dim vector

Test image L2 Nearest neighbors in <u>feature</u> space





Krizhevsky et al, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS 2012. Figures reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 12

May 11, 2023

Recall: Nearest neighbors in <u>pixel</u> space → ♥ ♥ ♥ ♥ ♥ ♥ ♥ ♥ ♥

Last Layer: Dimensionality Reduction

Visualize the "space" of FC7 feature vectors by reducing dimensionality of vectors from 4096 to 2 dimensions

Simple algorithm: Principal Component Analysis (PCA)

More complex: **t-SNE**





Van der Maaten and Hinton, "Visualizing Data using t-SNE", JMLR 2008 Figure copyright Laurens van der Maaten and Geoff Hinton, 2008. Reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 13

Last Layer: Dimensionality Reduction



Van der Maaten and Hinton, "Visualizing Data using t-SNE", JMLR 2008 Krizhevsky et al, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS 2012. Figure reproduced with permission.







May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Visualizing Activations

conv5 feature map is 128x13x13; visualize as 128 13x13 grayscale images



Lecture 12 - 15

May 11, 2023

Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014. Figure copyright Jason Yosinski, 2014. Reproduced with permission.

Visualizing Activations



Lecture 12 - 16

May 11, 2023

David Bau*, Bolei Zhou*, Aditya Khosla, Aude Oliva, Antonio Torralba Network Dissection: Quantifying Interpretability of Deep Visual Representations. CVPR 2017.

Visualizing what models have learned:

- Visualizing filters
- Visualizing final layer features
- Visualizing activations

Understanding input pixels

- Identifying important pixels
- Saliency via backprop
- Guided backprop to generate images
- Gradient ascent to visualize features

Adversarial perturbations

Style transfer

- Deep dream
- Features inversion
- Texture synthesis
- Neural style transfer

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Today's agenda

May 11, 2023

Maximally Activating Patches





Pick a layer and a channel; e.g. conv5 is 128 x 13 x 13, pick channel 17/128

Run many images through the network, record values of chosen channel

Visualize image patches that correspond to maximal activations



Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015 Figure copyright Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, Martin Riedmiller, 2015; reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Which pixels matter: Saliency via Occlusion

Mask part of the image before feeding to CNN, check how much predicted probabilities change



Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014

<u>Boat image</u> is <u>CC0 public domain</u> Elephant image is <u>CC0 public domain</u> <u>Go-Karts image</u> is <u>CC0 public domain</u>

Lecture 12 - 19

May 11, 2023

Which pixels matter: Saliency via Occlusion

Mask part of the image before feeding to CNN, check how much predicted probabilities change









Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014

<u>Boat image</u> is <u>CC0 public domain</u> Elephant image is <u>CC0 public domain</u> Go-Karts image is <u>CC0 public domain</u>



African elephant, Loxodonta africana



go-kart







Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 20

Which pixels matter: Saliency via Backprop

Forward pass: Compute probabilities



Lecture 12 - 21

May 11, 2023

Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014. Figures copyright Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, 2014; reproduced with permission.

Which pixels matter: Saliency via Backprop

Forward pass: Compute probabilities



Compute gradient of **(unnormalized) class score** with respect to image pixels, take absolute value and max over RGB channels

Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014.

Figures copyright Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, 2014; reproduced with permission.



Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 22

Saliency Maps



Lecture 12 - 23

May 11, 2023

Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014. Figures copyright Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, 2014; reproduced with permission.

Saliency Maps: Segmentation without supervision



Use GrabCut on saliency map

Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014. Figures copyright Karen Simonyan. Andrea Vedaldi. and Andrew Zisserman. 2014: reproduced with permission.

Rother et al, "Grabcut: Interactive foreground extraction using iterated graph cuts", ACM TOG 2004

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 24

Saliency maps: Uncovers biases

Such methods also find biases

wolf vs dog classifier looks is actually a snow vs nosnow classifier



(a) Husky classified as wolf

Lecture 12 - 25



(b) Explanation

May 11, 2023

Figures copyright Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, 2016; reproduced with permission. Ribeiro et al, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier", ACM KDD 2016

Intermediate Features via (guided) backprop

Lecture 12 - 26

May 11, 2023



Pick a single intermediate channel, e.g. one value in 128 x 13 x 13 conv5 feature map

Compute gradient of activation value with respect to image pixels

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014 Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015

Intermediate Features via (guided) backprop





b) Forward pass

2

6

	-1	5		1	0	5
	-5	-7	\rightarrow	2	0	0
	2	4		0	2	4
1	0	1		2	2	1

Relll

Backward pass: backpropagation

	0	2	4
	-2	3	-1
←	6	-3	1
	2	-1	3

Pick a single intermediate neuron, e.g. one value in 128 x 13 x 13 conv5 feature map

Compute gradient of neuron value with respect to image pixels

Backward pass: guided backpropagation

D	0	←	-2	3	-1
C	0		6	-3	1
03		2	-1	3	

Images come out nicer if you only backprop positive gradients through each ReLU (guided backprop)

Figure copyright Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, Martin Riedmiller, 2015; reproduced with permission.

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014 Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 27

Intermediate features via (guided) backprop



Guided Backprop

Lecture 12 - 28

May 11, 2023

Maximally activating patches (Each row is a different neuron)

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014 Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015 Figure copyright Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, Martin Riedmiller, 2015; reproduced with permission.

Intermediate features via (guided) backprop



Maximally activating patches (Each row is a different neuron)



Guided Backprop

Lecture 12 - 29

May 11, 2023

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014 Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015 Figure copyright Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, Martin Riedmiller, 2015; reproduced with permission.

(Guided) backprop:

Find the part of an image that a neuron responds to

Gradient ascent:

Generate a synthetic image that maximally activates a neuron

r

May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao

1. Initialize image to zeros

$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

score for class c (before Softmax)

May 11, 2023



Lecture 12 - 31

Repeat:

- 2. Forward image to compute current scores
- 3. Backprop to get gradient of neuron value with respect to image pixels
- 4. Make a small update to the image

<u>May 11, 2023</u>

Lecture 12 - 32

$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

Simple regularizer: Penalize L2 norm of generated image

Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014. Figures copyright Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, 2014; reproduced with permission.

$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

Simple regularizer: Penalize L2 norm of generated image







cup







dalmatian

bell pepper

lemon

husky

May 11, 2023

Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014.

Figures copyright Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, 2014; reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

Simple regularizer: Penalize L2 norm of generated image





Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014. Figure copyright Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson, 2014. Reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 34

Lecture 12 - 35

May 11, 2023

$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

Better regularizer: Penalize L2 norm of image; also during optimization periodically

- (1) Gaussian blur image
- (2) Clip pixels with small values to 0
- (3) Clip pixels with small gradients to 0

Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014.

$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

Better regularizer: Penalize L2 norm of image; also during optimization periodically

- (1) Gaussian blur image
- (2) Clip pixels with small values to 0
- (3) Clip pixels with small gradients to 0



Flamingo



Ground Beetle

Lecture 12 - 36



Pelican



Indian Cobra

May 11, 2023

Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014. Figure copyright Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson, 2014. Reproduced with permission.
$$\arg\max_{I} S_c(I) - \lambda \|I\|_2^2$$

Better regularizer: Penalize L2 norm of image; also during optimization periodically

- (1) Gaussian blur image
- (2) Clip pixels with small values to 0
- (3) Clip pixels with small gradients to 0



Hartebeest



Station Wagon



Billiard Table



Black Swan

May 11, 2023

Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014. Figure copyright Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson, 2014. Reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Use the same approach to visualize intermediate features



Lecture 12 - 38

May 11, 2023

Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014. Figure copyright Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson, 2014. Reproduced with permission.

Adding "multi-faceted" visualization gives even nicer results: (Plus more careful regularization, center-bias)

Reconstructions of multiple feature types (facets) recognized by the same "grocery store" neuron



Corresponding example training set images recognized by the same neuron as in the "grocery store" class



May 11, 2023

Lecture 12 - 39

Nguyen et al, "Multifaceted Feature Visualization: Uncovering the Different Types of Features Learned By Each Neuron in Deep Neural Networks", ICML Visualization for Deep Learning Workshop 2016. Figures copyright Anh Nguyen, Jason Yosinski, and Jeff Clune, 2016; reproduced with permission.



Nguyen et al, "Multifaceted Feature Visualization: Uncovering the Different Types of Features Learned By Each Neuron in Deep Neural Networks", ICML Visualization for Deep Learning Workshop 2016. Figures copyright Anh Nguyen, Jason Yosinski, and Jeff Clune, 2016; reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 40

May 11, 2023

Optimize in FC6 latent space instead of pixel space:



Lecture 12 - 41

May 11, 2023

Nguyen et al, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," NIPS 2016 Figure copyright Nguyen et al, 2016; reproduced with permission.

Optimize in FC6 latent space instead of pixel space:



Lecture 12 - 42

May 11, 2023

Nguyen et al, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," NIPS 2016 Figure copyright Nguyen et al, 2016; reproduced with permission.

Visualizing what models have learned:

- Visualizing filters
- Visualizing final layer features
- Visualizing activations

Understanding input pixels

- Identifying important pixels
- Saliency via backprop
- Guided backprop to generate images
- Gradient ascent to visualize features

Adversarial perturbations

Style transfer

- Deep dream
- Features inversion
- Texture synthesis
- Neural style transfer

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Today's agenda

May 11, 2023

Fooling Images / Adversarial Examples

(1) Start from an arbitrary image
 (2) Pick an arbitrary class
 (3) Modify the image to maximize the class
 (4) Repeat until network is fooled

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 44

May 11, 2023

Fooling Images / Adversarial Examples

African elephant







iPod



Difference



10x Difference

A-A





Boat image is CC0 public domain Elephant image is CC0 public domain

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Fooling Images / Adversarial Examples

Universal perturbations

Moosavi-Dezfooli, Seyed-Mohsen, et al. "Universal adversarial perturbations." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. Figure reproduced with permission



May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Visualizing what models have learned:

- Visualizing filters
- Visualizing final layer features
- Visualizing activations

Understanding input pixels

- Identifying important pixels
- Saliency via backprop
- Guided backprop to generate images
- Gradient ascent to visualize features

Adversarial perturbations

Style transfer

- Features inversion
- Deep dream
- Texture synthesis
- Neural style transfer

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Today's agenda

May 11, 2023

Feature Inversion

Given a CNN feature vector for an image, find a new image that:

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

Lecture 12 - 48

May 11, 2023

Mahendran and Vedaldi, "Understanding Deep Image Representations by Inverting Them", CVPR 2015

Feature Inversion



Mahendran and Vedaldi, "Understanding Deep Image Representations by Inverting Them", CVPR 2015 Figure from Johnson, Alahi, and Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", ECCV 2016. Copyright Springer, 2016. Reproduced for educational purposes.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 49

May 11, 2023

Rather than synthesizing an image to maximize a specific neuron, instead try to **amplify** the neuron activations at some layer in the network





Lecture 12 - 50

Choose an image and a layer in a CNN; repeat:

- 1. Forward: compute activations at chosen layer
- 2. Set gradient of chosen layer equal to its activation
- 3. Backward: Compute gradient on image
- 4. Update image

Mordvintsev, Olah, and Tyka, "Inceptionism: Going Deeper into Neural Networks", <u>Google Research Blog</u>. Images are licensed under <u>CC-BY</u> <u>4.0</u>

May 11, 2023

Rather than synthesizing an image to maximize a specific neuron, instead try to **amplify** the neuron activations at some layer in the network





Lecture 12 - 51

Choose an image and a layer in a CNN; repeat:

- 1. Forward: compute activations at chosen layer
- 2. Set gradient of chosen layer equal to its activation
- 3. Backward: Compute gradient on image
- 4. Update image

Equivalent to: ___ I* = arg max_I ∑_i f_i(I)²

Mordvintsev, Olah, and Tyka, "Inceptionism: Going Deeper into Neural Networks", <u>Google Research Blog</u>. Images are licensed under <u>CC-BY</u> 4.0

May 11, 2023

```
def objective L2(dst):
   dst.diff[:] = dst.data
def make step(net, step size=1.5, end='inception 4c/output',
             jitter=32, clip=True, objective=objective L2):
    '''Basic gradient ascent step.'''
   src = net.blobs['data'] # input image is stored in Net's 'data' blob
   dst = net.blobs[end]
   ox, oy = np.random.randint(-jitter, jitter+1, 2)
    src.data[0] = np.roll(np.roll(src.data[0], ox, -1), oy, -2) # apply jitter shift
   net.forward(end=end)
   objective(dst) # specify the optimization objective
   net.backward(start=end)
   q = src.diff[0]
    # apply normalized ascent step to the input image
    src.data[:] += step size/np.abs(g).mean() * g
    src.data[0] = np.roll(np.roll(src.data[0], -ox, -1), -oy, -2) # unshift image
   if clip:
        bias = net.transformer.mean['data']
        src.data[:] = np.clip(src.data, -bias, 255-bias)
```

<u>Code</u> is very simple but it uses a couple tricks:

May 11, 2023

(Code is licensed under Apache 2.0)

Lecture 12 - 52



Fei-Fei Li, Yunzhu Li, Ruohan Gao





Fei-Fei Li, Yunzhu Li, Ruohan Gao





Lecture 12 - 55

May 11, 2023

Also uses multiscale processing for a fractal effect (not shown)











Lecture 12 - 58

May 11, 2023



Lecture 12 - 59

May 11, 2023







Image is licensed under CC-BY 4.0

May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Texture Synthesis

Given a sample patch of some texture, can we generate a bigger image of the same texture?



Output image is licensed under the MIT license

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Texture Synthesis: Nearest Neighbor

Generate pixels one at a time in scanline order; form neighborhood of already generated pixels and copy nearest neighbor from input





Fei-Fei Li, Yunzhu Li, Ruohan Gao





Texture Synthesis: Nearest Neighbor



it ndatwears coune Tring rooms," as Heft he fast nd it ars dat notars ortseas ribed it last nt hest bedian Al. econicalHornd ith Al. Heftars of as da Lewindailf I han Al Ths," as Lewing questies last aticarsticall. He is dian Al last fal counda Lew, at "this dailyears d ily edianicall. Hoozewing rooms," as House De fale f De und itical counsestscribed it last fall. He fall. Heff rs oroheoned it nd it he left a ringing questica Lewin icars coecoms," astore years of Monica Lewinow see a Thas Fring roome stooniscat nowea re left a roouse bouestof MHe lelft a Lést fast neine launesticars Hef nd it rip?" TrHouself, a ring ind it≤onestid it a ring que: astical cois ore years of Moung fall. He ribof Mouse ore years of and a Tripp?" That hedian Al Lest fasee yea nda Tripp?' Iolitical comedian Alét he few se ring que olitical cone re years of the storears of as 1 Frat nica L res Lew se lest a rime 1 He fas quest nging of, at beou

"" on for wes tomoun with it lason as do fashof with, Alar be it align g Tatts of inesseon and sed in Fof Lofa worf. tob ye 1 Lisingetarf the attention of scient ausase ripo? "To? (lesis't nt?) traise in attent at 12 lesit of the attent and the let La sport ce so al la la contacionaria por la contra la contra la contra 101/ Arenous Lesir," anonousing anois données aufors à horeanna Friant require l'90/ scadiatett Mohdat reediale diavegnomarst al sta sof, aw seoo Leni Hoenstao so nas 6 Hal neeventoins La Pintie year 12 failent inovidios, iwonwinguiton and i at these lat sound merst hit have to refer to read the F. Ernett a We under the reconducar failed routicard in hereins rest aft "Hotsesten alleves ilma, Lene Lodes than an electric dear Frate ine of the carsof he sucarray san's drainer street of Hussid and 10 Jung da new quartit many sal bool Horyzas Lectrians 1 of 11 a Cabuyesties," itea Lettes fa Catthicl combinetit, he sizens 13 us histing counstast t is not drived wore qualal wing ting a lalas "aboquiw, rort r a rillia four a soil, shift the stab Hesl 1atasi gyneswiel it s Gliszber proof fus tha lararse ris defoit larstit inea Lts, e es of gigodniwn boss glaicaHi He zie fasniuae stic of activities a La lained assa kewawadid topicabus or esties, "Was not later scassories estimated as at the Levingenica e yhn puze y wint bu a rfat end i dily om argessizeal Hot Zars ue v, rokt med mg? though Houosfast he mitchitize et malto it stalesty qual decorring of arta encitory de incontrol riprative, ere Ledeforise Heitse sole est HASS "t athorse oreate your thiste radappreuve Eenor: offes fasminens mucas esuated obrewors a pollar frant failaths Bouge lasts yet unrought sector Hodin, arsilest loon: Thate, set, of nist's seode ing't for sist carbourous stars cooduling Fairl dies capoor Heeding is night utiling in Harded tubulit fr " an," as the purchains rithed Fhare Thaines de us Thasorseeing og orderdia 1555's Broom sead værret 4 015 " 1 an I thing teft and an angeo of the set and the set of the area of t qualast august a sites attirs ved it d liftst a af he's ridz frite luisi ft". sipidarit sous I I master Hing Atont ing ins rrite foit I as go: on , " storts unite prischeft z fitel wis ut fairs of aticansing of clars callerics ripte protraid Alular vitablaiecciet acticities searches tan mais rooms planest as farses in earsies liging at nos dais innuroués tourest and that meaning is tiller is over, i Ar mary iburstans, sHe?" Ichis environal anni attionigal Hase ficais stit ruchida; " zonis j'at inges iot/rob. H Aoforianus, dailof haqu stimoeda ; quisinau gonzianaral, wicar fat niti Thimatlaipberintore dijustert la icitéluée taass'ouries d'is classe, àt e qui nwys Found out Holfhda'r is ing "mirs eidaiel, at n Aleeredal citic" unset r ywour og more vstigiscy igicou iguse s erst sheet ras (tod it I'. Th Fulls rog ry for ing softhig mis' af daiburg inter y than eautic que Hotsest it Lewis fa inder y est ry oraling normal et as," Al ath bug he stored que in as est minumalase Liener young. To are en ous ou festoarstnious costinduing une ondaight gireolail; fill be nonouveriou untitious fast renewars eners; dia et la Le riearas, "gaoffabrine tilynë searshise unditioisme la orange ndew: De. Huss is Ho zristyczee yalatst te ytieczer woeds starooi fdan las gistigić ofa De' is That gik he fal enes ars og solw ny faros H(asy cuping on ag quantane r cwfir leders Hass dred finse fars on fare ais her fast noue cas Livelars , dreng s la ese sofiedus mes unes gérid. A Rousin siles raisours de quimerre flif ridiaes fritheranstices rate "A b artision diandes" wooding z along eing Ley Fig : fast fa minista sws ingelindatelft: tests of evit: a zur: Ling wits dAl ist dail: Thaous C wipiz & c octe : yearb st right of Maonibritringt: Ins w by sy nuHegisself and isat rates of a astriccofanithen 1 se. hear etaastord ring meanwhai ("in a a traiter of the most share of your of the second se e yals usignin . daice wefaving nostase Lones, arone viewelus lainomin's Eorinestiong dry note see Fist la frievort. Loendained ^b qomBicionet_{inici}st a Liout dation set as converginer plausift in of yet: da rubners (projector) dation, it's d Facilianes H 12, "a shill yet set consistenting as yourday of Futione," littletino enda rashout Holars cates a straats, "i rus fig that he Misizooce 27 que a riber ricoicistonies ile ma Aior, fallaconarritriagia rearris dacountirseas Pranise à romfarind astifice T junnite in the growth of Strategy and Hildren is a rate or year information interaction of Arits infation rates of the state of the s Ha que Ision-Honeruist; serina re, roces rortsta Lesugines wiscere : d. seriad itila Fores? addata He sit Littere ers of hop?

Images licensed under the MIT license

May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Each layer of CNN gives C x H x W tensor of features; H x W grid of C-dimensional vectors

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 65

May 11, 2023



Each layer of CNN gives C x H x W tensor of features; H x W grid of C-dimensional vectors

Outer product of two C-dimensional vectors gives C x C matrix measuring co-occurrence

Fei-Fei Li, Yunzhu Li, Ruohan Gao





Each layer of CNN gives C x H x W tensor of features; H x W grid of C-dimensional vectors

Outer product of two C-dimensional vectors gives C x C matrix measuring co-occurrence

Average over all pairs of vectors, giving **Gram matrix** of shape C x C

Gram Matrix

Fei-Fei Li, Yunzhu Li, Ruohan Gao





Each layer of CNN gives $C \times H \times W$ tensor of features; $H \times W$ grid of C-dimensional vectors

Outer product of two C-dimensional vectors gives C x C matrix measuring co-occurrence

Average over all HW pairs of vectors, giving **Gram matrix** of shape C x C

Efficient to compute; reshape features from

May 11, 2023

 $C \times H \times W$ to $=C \times HW$

then compute $G = FF^{T}$

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Neural Texture Synthesis

- 1. Pretrain a CNN on ImageNet (VGG-19)
- Run input texture forward through CNN, record activations on every layer; layer i gives feature map of shape C_i × H_i × W_i
- 3. At each layer compute the *Gram matrix* giving outer product of features:

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}$$
 (shape C_i × C_i)



Lecture 12 - 69

May 11, 2023

Gatys, Ecker, and Bethge, "Texture Synthesis Using Convolutional Neural Networks", NIPS 2015 Figure copyright Leon Gatys, Alexander S. Ecker, and Matthias Bethge, 2015. Reproduced with permission.

Neural Texture Synthesis

- 1. Pretrain a CNN on ImageNet (VGG-19)
- Run input texture forward through CNN, record activations on every layer; layer i gives feature map of shape C_i × H_i × W_i
- 3. At each layer compute the *Gram matrix* giving outer product of features:
- $G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$ (shape C_i × C_i)
- 4. Initialize generated image from random noise
- 5. Pass generated image through CNN, compute Gram matrix on each layer



May 11, 2023

Lecture 12 - 70

Gatys, Ecker, and Bethge, "Texture Synthesis Using Convolutional Neural Networks", NIPS 2015 Figure copyright Leon Gatys, Alexander S. Ecker, and Matthias Bethge, 2015. Reproduced with permission.

Neural Texture Synthesis

$$E_l = rac{1}{4N_l^2 M_l^2} \sum_{i,j} \left(G_{ij}^l - \hat{G}_{ij}^l
ight)^2 \qquad \mathcal{L}(ec{x}, \hat{ec{x}}) = \sum_{l=0}^L w_l E_l$$

- 1. Pretrain a CNN on ImageNet (VGG-19)
- Run input texture forward through CNN, record activations on every layer; layer i gives feature map of shape C_i × H_i × W_i
- 3. At each layer compute the *Gram matrix* giving outer product of features:
- $G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}$ (shape C_i × C_i)
 - 4. Initialize generated image from random noise
- 5. Pass generated image through CNN, compute Gram matrix on each layer
- 6. Compute loss: weighted sum of L2 distance between Gram matrices



May 11, 2023

Gatys, Ecker, and Bethge, "Texture Synthesis Using Convolutional Neural Networks", NIPS 2015 Figure copyright Leon Gatys, Alexander S. Ecker, and Matthias Bethge, 2015. Reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Neural Texture Synthesis $E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} \left(G_{ij}^l - \hat{G}_{ij}^l \right)^2 \qquad \mathcal{L}(\vec{x}, \hat{\vec{x}}) = \sum_{i,j}^{\infty} w_l E_l$



- 2. Run input texture forward through CNN, record activations on every layer; layer i gives feature map of shape $C_i \times H_i \times W_i$
- 3. At each layer compute the Gram matrix giving outer product of features:

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$$
 (shape C_i × C_i)

- Initialize generated image from random 4. noise
- 5. Pass generated image through CNN, compute Gram matrix on each layer
- Compute loss: weighted sum of L2 6. distance between Gram matrices
- Backprop to get gradient on image 7.
- 8. Make gradient step on image

GOTO 5 9.

Gatys, Ecker, and Bethge, "Texture Synthesis Using Convolutional Neural Networks", NIPS 2015 Figure copyright Leon Gatys, Alexander S. Ecker, and Matthias Bethae, 2015, Reproduced with permission,



May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao
Neural Texture Synthesis

Reconstructing texture from higher layers recovers larger features from the input texture



May 11, 2023

Gatys, Ecker, and Bethge, "Texture Synthesis Using Convolutional Neural Networks", NIPS 2015 Figure copyright Leon Gatys, Alexander S. Ecker, and Matthias Bethge, 2015. Reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Neural Texture Synthesis: Texture = Artwork



Figure from Johnson, Alahi, and Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", ECCV 2016. Copyright Springer, 2016. Reproduced for educational purposes.

(Gram

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 74

Neural Style Transfer: Feature + Gram Reconstruction



Texture synthesis (Gram reconstruction)

Feature reconstruction

Figure from Johnson, Alahi, and Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", ECCV 2016. Copyright Springer, 2016. Reproduced for educational purposes.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 75

+

Content Image



This image is licensed under CC-BY 3.0

Style Image



Starry Night by Van Gogh is in the public domain

Gatys, Ecker, and Bethge, "Texture Synthesis Using Convolutional Neural Networks", NIPS 2015

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 76

╈

Content Image



This image is licensed under CC-BY 3.0

Style Image



Starry Night by Van Gogh is in the public domain

Style Transfer!



This image copyright Justin Johnson, 2015. Reproduced with permission.

May 11, 2023

Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Lecture 12 - 78

Style image Style Target $\ell_{style}^{\phi, \texttt{relu1_2}} \ \ell_{style}^{\phi, \texttt{relu2_2}} \ \ell_{style}^{\phi, \texttt{relu3_3}}$ $\ell_{style}^{\phi, \texttt{relu4}_3}$ y_s Output image Loss Network y_c $\ell_{feat}^{\phi, \texttt{relu3_3}}$ **Content Target** Content image

Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016 Figure adapted from Johnson, Alahi, and Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", ECCV 2016. Copyright Springer, 2016. Reproduced for educational purposes.

Fei-Fei Li, Yunzhu Li, Ruohan Gao





Example outputs from Lua torch <u>implementation</u>



Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016 Figure copyright Justin Johnson, 2015.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 81



More weight to content loss More weight to style loss

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Resizing style image before running style transfer algorithm can transfer different types of features





Lecture 12 - 83

May 11, 2023

Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016 Figure copyright Justin Johnson, 2015.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Neural Style Transfer: Multiple Style Images

Mix style from multiple images by taking a weighted average of Gram matrices



Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016 Figure copyright Justin Johnson, 2015.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 84



Lecture 12 - 85



Lecture 12 - 86



Lecture 12 - 87



Lecture 12 - 88

Problem: Style transfer requires many forward / backward passes through VGG; very slow!

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 89

Problem: Style transfer requires many forward / backward passes through VGG; very slow!

Solution: Train <u>another</u> neural network to perform style transfer for us!

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 90

Fast Style Transfer

- (1) Train a feedforward network for each style
- (2) Use pretrained CNN to compute same losses as before
- (3) After training, stylize images using a single forward pass



Figure copyright Springer, 2016. Reproduced for educational purposes.

Lecture 12 - 91

May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Fast Style Transfer



Johnson, Alahi, and Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", ECCV 2016 Figure copyright Springer, 2016. Reproduced for educational purposes. https://github.com/jcjohnson/fast-neural-style

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Remember Normalization Methods?



Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 93

Remember Normalization Methods?

Instance Normalization was developed for style transfer!



Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 94



Replacing batch normalization with Instance Normalization improves results

Ulyanov et al, "Texture Networks: Feed-forward Synthesis of Textures and Stylized Images", ICML 2016 Ulyanov et al, "Instance Normalization: The Missing Ingredient for Fast Stylization", arXiv 2016 Figures copyright Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor Lempitsky, 2016. Reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 95

One Network, Many Styles



Dumoulin, Shlens, and Kudlur, "A Learned Representation for Artistic Style", ICLR 2017. Figure copyright Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur, 2016; reproduced with permission.

Fei-Fei Li, Yunzhu Li, Ruohan Gao

Lecture 12 - 96

One Network, Many Styles

Use the same network for multiple styles using <u>conditional instance</u> <u>normalization</u>: learn separate scale and shift parameters per style





Dumoulin, Shlens, and Kudlur, "A Learned Representation for Artistic Style", ICLR 2017. Figure copyright Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur, 2016; reproduced with permission.

Single network can blend styles after training

May 11, 2023

Fei-Fei Li, Yunzhu Li, Ruohan Gao



Many methods for understanding CNN representations

Activations: Nearest neighbors, dimensionality reduction, maximal patches, occlusion Gradients: Saliency maps, class visualization, fooling images, feature inversion Fun: DeepDream, style transfer

Lecture 12 - 98

<u>May 11, 2023</u>

Next time:

5/16 Midterm

5/18 Self-supervised Learning

Fei-Fei Li, Yunzhu Li, Ruohan Gao

