Lecture 15:

Adversarial examples, texture synthesis, and style transfer

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 15 - 1 Nov 7, 2024

Agenda

Recap

Adversarial examples

Texture synthesis and style transfer

Bonus

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 2 Nov 7, 2024

Last lecture: Understanding ConvNets

- Visualize the weights
- Visualize the last layer (via t-SNE)
- Visualize patches that maximally activate neurons
- Occlusion experiments
- Deconv approaches (single backward pass)
- Optimization over image approaches (optimization)

Nov 7, 2024

Lecture 15 - 3

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Question: Given a CNN code, is it possible to reconstruct the original image?



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 15 - 4 Nov 7, 2024

Find an image such that:

- Its code is similar to a given code
- It "looks natural" (image prior regularization)

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^{H \times W \times C}} \ell(\Phi(\mathbf{x}), \Phi_0) + \lambda \mathcal{R}(\mathbf{x})$$

$$\ell(\Phi(\mathbf{x}), \Phi_0) = \|\Phi(\mathbf{x}) - \Phi_0\|^2$$

Lecture 15 - 5

Nov 7, 2024

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Understanding Deep Image Representations by Inverting Them [Mahendran and Vedaldi, 2014]

original image



reconstructions from the 1000 log probabilities for ImageNet (ILSVRC) classes

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 6 Nov 7, 2024

Reconstructions from the representation after last last pooling layer (immediately before the first Fully Connected layer)



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 15 - 7 Nov 7, 2024



Reconstructions from intermediate layers



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 8 Nov 7, 2024

We can pose an optimization over the input image to maximize any class score. That seems useful.

Question: Can we use this to "fool" ConvNets?

spoiler alert: yeah

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 9 Nov 7, 2024

(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

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 10 Nov 7, 2024

[Intriguing properties of neural networks, Szegedy et al., 2013]



[Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images Nguyen, Yosinski, Clune, 2014]



>99.6% confidences

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 12 Nov 7, 2024

[Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images Nguyen, Yosinski, Clune, 2014]



>99.6% confidences

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 13 Nov 7, 2024

These kinds of results were around even before ConvNets... [Exploring the Representation Capabilities of the HOG Descriptor, Tatu et al., 2011]





Lecture 15 - 14

Nov 7, 2024

Identical HOG represention

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES [Goodfellow, Shlens & Szegedy, 2014]

"primary cause of neural networks' vulnerability to adversarial perturbation is their **linear nature**"



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 15 Nov 7, 2024

Lets fool a binary linear classifier: (logistic regression)

$$P(y=1 \mid x;w,b) = rac{1}{1+e^{-(w^Tx+b)}} = \sigma(w^Tx+b)$$



Since the probabilities of class 1 and 0 sum to one, the probability for class 0 is $P(y=0 \mid x; w, b) = 1 - P(y=1 \mid x; w, b)$. Hence, an example is classified as a positive example (y = 1) if $\sigma(w^T x + b) > 0.5$, or equivalently if the score $w^T x + b > 0$.

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 16 Nov 7, 2024



$$P(y=1 \mid x; w, b) = rac{1}{1+e^{-(w^T x+b)}} = \sigma(w^T x+b)$$

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 15 - 17 Nov 7, 2024

class 1 score = dot product:

=> probability of class 1 is $1/(1+e^{(-(-3))}) = 0.0474$

i.e. the classifier is **95%** certain that this is class 0 example.

$$P(y=1 \mid x;w,b) = rac{1}{1+e^{-(w^Tx+b)}} = \sigma(w^Tx+b)$$

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 18 Nov 7, 2024



class 1 score = dot product:

=> probability of class 1 is $1/(1+e^{(-(-3))}) = 0.0474$

i.e. the classifier is 95% certain that this is class 0 example.

$$P(y=1 \mid x;w,b) = rac{1}{1+e^{-(w^Tx+b)}} = \sigma(w^Tx+b)$$

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 15 - 19 Nov 7, 2024



class 1 score before:

-2 + 1 + 3 + 2 + 2 - 2 + 1 - 4 - 5 + 1 = -3

=> probability of class 1 is $1/(1+e^{(-(-3))}) = 0.0474$ $P(y=1 \mid x; w, b) = \frac{1}{1+e^{-(w^T x+b)}} = \sigma(w^T x+b)$

-1.5+1.5+3.5+2.5+2.5-1.5+1.5-3.5-4.5+1.5 = 2=> probability of class 1 is now $1/(1+e^{-(-2))}) = 0.88$

i.e. we improved the class 1 probability from 5% to 88%

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Nov 7, 2024 Lecture 15 - 20



This was only with 10 input dimensions. A 224x224 input image has 150,528.

(It's significantly easier with more numbers, need smaller nudge for each)

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

-2 + 1 + 3 + 2 + 2 - 2 + 1 - 4 - 5 + 1 = -3

=> probability of class 1 is $1/(1+e^{(-(-3))}) = 0.0474$

=> probability of class 1 is now $1/(1+e^{-(-2))}) = 0.88$

i.e. we improved the class 1 probability from 5% to 88%

-1.5+1.5+3.5+2.5+2.5-1.5+1.5-3.5-4.5+1.5 = 2

class 1 score before:

Lecture 15 - 21 Nov 7, 2024



+ .007 \times

y ="panda" w/ 57.7% confidence

 \boldsymbol{x}





 $oldsymbol{x} + \epsilon \cdot ext{sign}(
abla_{oldsymbol{x}} J(oldsymbol{ heta}, oldsymbol{x}, y))$ "gibbon" w/ 99.3 % confidence

Explaining and Harnessing Adversarial Examples

Ian J. Goodfellow, Jonathon Shlens, Christian Szegedy

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 22 Nov 7, 2024

[Intriguing properties of neural networks, Szegedy et al., 2013]



Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Can be printed on paper! Kurakin et al., 17



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 24 Nov 7, 2024

Also works for 3D models! (though a little harder for point clouds)

Su et al., ECCV 2018



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 25 Nov 7, 2024

Neural Style Transfer and Texture Synthesis

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.

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

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 26 Nov 7, 2024

Texture Synthesis

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



Lecture 15 - 27

Output image is licensed under the MIT license

Nov 7, 2024

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

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





Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 28 Nov 7, 2024



Texture Synthesis: Nearest Neighbor



rt nda wears owne Tring rooms," as Heft he fast nd it as dat notass cortseas sibed it last nitbest bedian. Al. conicalHoud it h Al. Heft ars of as da LewindaiH ian Al The," as Lowing questies last at case tical His is dian Al last fal counda Lew, at "this dailyears d ily edianicall. Hoosewing rooms," as House De fale f D and itical connerstate/hed it last fall. He fall, Heff as conferenced it not it he left a similar questica Lewis cass coecoms," as toge years of Montica Lewinow see a They Eving voorge stoon is rat now ye we left a yoo up bourstof Mie lelft a Lést fast ngine lauresticaes He d it sip?" Telouself, a singlind it sonested it a ging gie astical cois one years of Moung fall. He shof Mouse re years of and a Tripp !" That hedian Al Lest fasee yea ada Tripp? Solitical comedian Aléthe few sering qua olitical cone as years of the storears of as I Feat nics L eas Lew selects eine 1 He fas questinging of, at bee

Lecture 15 - 29

one on one of our states of the contract of the states of ered of the section o Likeyout wis bag 1 Hou while quite II. HHier a contrait court far a series to he sist a gous date of our of this contraining log to >> locpticzerstog wie incapture se lik Galawise i wy gohan it.dust, si w wijit u HanAlesev I Physicagus cancibrai n 12 section Advariant strang day variation des text courses frances a balance ; que y H wit hitsn'ting \$2 of yit far ideation. Hy a la We groups Lisin," a many using a pairs wanters waters is have and Friantice trace is 1-94-9-radiaters Mahart me maeting you on 1495 al sta sof, zw sees Line Harverian in nas a Hal normeridars Li Finde you the fallent inov dans, two winquito and i at this a lat considerable in high no states to consider F. Exact a We multi a n conducative interation licenser as to fit Whomesen adhersvilled Jene Leden charass close dow Frate the of the carry for a compressed driver at set or Hastin on ar jung da new quantit many a faload Horyess Lee winas 1 + of (st, 2 Cabeyesties, " see Lieus fa Cathiel to obnetit by steam a w bisting routed as the second with one quality sing ting the lates "stronger wy root r and the sour a total with the fates Hest "It is gyne gwiel its a jerrier broot fus the larerse vit datit here't ince Lts, 'ener a gigoddwr?005 gleicaffi He ric faioinae atter of again ones Lo lainear as a second distribution of the second distress distribution of the second distribution of the sec (Mn nee yound be a rfatem it day constrains iteal list fars us y, add then dug! though Housefast he mitositiful of scalar st sulleire qual diedersien og ats Ded frumer, Ar insorauer sipratijes, ere Ledefinist Heitste aus et: He^{Asy, 1} atherse oreatai yrous nin rhapyow Letar, aries his nasus whatestors el onsurvis a quiter frame farlets Loue last yest arought sein Ho die, aribert looni They'r ast, efnis ti's to de nig't fwria a castour oun sears creeduilde Ffairt dir S Choose Hendig is sirdt ardine din Habied model for a for astar quedaits ritkal Fham Thaines de n'Thasorseving og ordestinants Ebours ad mener aniltine seft wit averonipotnete ve Les. Lendar evel as i it existe dues quipoutinges of asses or Prove Alzhont quolasi aunas a sitts attivaseed it dilftsta at he'a rodz frita luesi ft" sipilarnit rous I I exasive; Hingi Arony ingtes reik-foit I as dis to " the of b order viscingent of fuel as ut fairs of the awing of dates calle its eight your aid a klamit dairee a direct antick as uses t an Dath courses fancest as farses i pensaies leginger nes dais innuzeurs comere, and ut a excanines of therefy i some . An new y iberations sHee?" ichis tute humal annt uttionizal Ease ficilit stut with de ," sonis fat up a sot rat. Il Aoforius as, "tailof haq stimpetal querney bourder grand Inter fatnich Thina thipbarlotori vijastot la icitika e trais ouries C is vlaste, et a qu www.Formal_uHillhdarist less "missidarist, at h Alexandal cities" unagent synduring more vertging vigieou faute a east abege rus ided it T. Th Fuks rog ry torio's sist highin's a daiberig lithe : yllaineautic qu Hutsest it Lewis () a down est avonaing nofas se as," Al art, ouginessents ged in a catralum lase Litney you, qu'icize evices et for constraints que indicing une endaug ngreobalt faith ne donous ziour frau fait tettarr suseri (da e la la zezra, via effaithe tilpet saralise inclusione) ao or or hulew: De Fugé ra Homario (e printer printer arrous a travori dana la gatagi (da a Thaf teth e fait cas) arrous solwing far off (as a confine for any quotience over Tolers Hass deed these fars on fare as here fast ever as Itreflare a dweng with wir woheding mes then gorind A htouan aller raisous day gomen flofridians trittinfaenituus ret a 46 bas Gaioridianda, sychooling z along eine Les ng rfant ifritzele swisiogidzdadelft's insts of with a nm.Logistis dal cie chall. Thaous f wigit a g ore using strength in the print of the set with the list of south and the set of the set nearmat ("1 a a takie", Hot has avery wonds is obovis of the an a testa fas ag' baing of the life invites 10 a an the airs the area our sector and the same of the area of the wishing the data eveloping heitigs junes, since the wile lamonials Locinestient diractions Fist b filerat. I solution b genulisions in the set of th din 2000 2 2 Que piribe ricores purtes alle the Alor: fall monarritrigig reares decounterses. Pe ress in antificial anifest Turnine ins "s guterioris Francisco" Hilder da insene voit innerhealtratheoingin, Astis infafate institus ottens Fr La supplie H guse" fat caoingutis, "squar quar if 2 lest avant nei. Die Dines georal hole," turn dividuar lang daytof HI que Is a H. onwritty sering to: toos vortile Leinginesthio we'e d. sonad it its Fires? stians He st R Lein of rest thop?

Images licensed under the MIT license

Nov 7, 2024

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller



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

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 30 Nov 7, 2024



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

Outer product of C-dimensional vector with itself gives C x C matrix measuring co-occurrence



Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 31 Nov 7, 2024



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

Outer product of C-dimensional vector with itself gives C x C matrix measuring co-occurrence

The green box G(i,j) represents the AVERAGE, over image positions in the image, of the product of features i and j.

Suppose "i" represents horizontal lines and "j" represents vertical lines.

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 32 Nov 7, 2024

Ci



Vertical and horizontal DO NOT co-occur

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Vertical and horizontal co-occur

Lecture 15 - 33 Nov 7, 2024



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

Outer product of C-dimensional vector with itself gives C x C matrix measuring co-occurrence

Average over all HW outer products, giving **Gram matrix** of shape C x C

Gram Matrix

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 34 Nov 7, 2024



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

Outer product of C-dimensional vector with itself gives C x C matrix measuring co-occurrence

Average over all HW outer products, giving **Gram matrix** of shape C x C

Efficient to compute; reshape features from

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

then compute $G = FF^{T}$

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 35 Nov 7, 2024

- 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 15 - 36

Nov 7, 2024

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.

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

- 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



Lecture 15 - 37

Nov 7, 2024

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.

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

- 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

$$E_{l} = \frac{1}{4N_{l}^{2}M_{l}^{2}} \sum_{i,j} \left(G_{ij}^{l} - \hat{G}_{ij}^{l}\right)^{2} \qquad \mathcal{L}(\vec{x}, \hat{\vec{x}}) = \sum_{l=0}^{L} w_{l}E_{l}$$



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.

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 38 Nov 7, 2024

- 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
- 7. Backprop to get gradient on image
- 8. Make gradient step on image
- 9. GOTO 5



Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

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



Lecture 15 - 39 Nov 7, 2024

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



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.

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 40 Nov 7, 2024

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

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Nov 7, 2024 Lecture 15 - 41

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.

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 42 Nov 7, 2024

Content Image



This image is licensed under CC-BY 3.0

Style Image



Starry Night by Van Gogh is in the public domain

Lecture 15 - 43

Nov 7, 2024

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

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

╉

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.

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

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 44 Nov 7, 2024



Lecture 15 - 45 Nov 7, 2024



Lecture 15 - 46 Nov 7, 2024



Example outputs from implementation (in Torch)



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

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 48 Nov 7, 2024



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 49 Nov 7, 2024

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



Larger style image Smaller style image

Lecture 15 - 50

Nov 7, 2024

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

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Neural Style Transfer: Multiple Style Images

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



Lecture 15 - 51

Nov 7, 2024

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

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller



Lecture 15 - 52 Nov 7, 2024



Lecture 15 - 53 Nov 7, 2024



Lecture 15 - 54 Nov 7, 2024



Lecture 15 - 55 Nov 7, 2024

Published as a conference paper at ICLR 2017

WHAT DOES IT TAKE TO GENERATE NATURAL TEXTURES?

Ivan Ustyuzhaninov^{*,1,2,3}, Wieland Brendel^{*,1,2}, Leon Gatys^{1,2,3}, Matthias Bethge^{1,2,3,4}

*contributed equally

¹Centre for Integrative Neuroscience, University of Tübingen, Germany
 ²Bernstein Center for Computational Neuroscience, Tübingen, Germany
 ³Graduate School of Neural Information Processing, University of Tübingen, Germany
 ⁴Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Subhransu Maji, Chuang Gan and TAs

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 56 Nov 7, 2024



Bilinear (second-order) pooling

CNN activations pooled after outer-product encoding



"gray belly"

Bilinear (second-order) pooling

Fine-grained classification (VGG-D + VGG-M networks)







FGVC Aircraft 100 variants, 10,000 images



Stanford cars 196 models, 16,185 images

Method	Birds	Aircraft	Cars
Fully connected [D]	70.4	76.6	79.8
Fisher vector [D]	74.7	78.7	85.7
Bilinear [D+D]	84.0	83.9	90.6
Bilinear [D+M]	84.1	84.5	91.3
Previous work	84.1 [1]	80.7 [2]	92.6 [3]

Method	NABirds	
Inception-BN	73.1 [4]	
B-CNN [D+M]	79.4	

48,562 images of 555 categories

[1] Spatial Transformer Networks, Jaderberg et al., NIPS 15

[2] Revisiting the Fisher vector for Fine-grained Classification, Gosselin et al., PR Letters 14

[3] Fine-Grained Rec. w/o Part Annotations, Krause et al., CVPR 15

[4] Batch-normalized Inception Architectures, Szegedy et al., CVPR 15

"inverse" images for bilinear CNNs



Lin and Maji, Visualizing and Understanding Deep Texture Representations, CVPR 16

What texture are birds?



American goldfinch

Pied kingfisher

Hooded oriole



White eyed vireo

Lin and Maji, Visualizing and Understanding Deep Texture Representations, CVPR 16

What texture are bookstores?

Describable Texture Datatset



Flickr Material Dataset



MIT Indoor



Oxford flowers



FGVC butterflies and moths



FGVC fungi





Many methods for understanding CNN representations

Activations: Nearest neighbors, Dimensionality reduction, maximal patches, occlusion

Gradients: Saliency maps, class visualization, feature inversion

Fun: DeepDream, Texture Synthesis, Style Transfer **Bonus:** Works for fine-grained categorization too!

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 15 - 66 Nov 7, 2024