Fun with ConvNets

CS194: Intro to Comp. Vision and Comp. Photo
Alexei Efros, UC Berkeley, Fall 2020
Regression Loss (e.g. Project 4)

\[ L_2(\hat{Y}, Y) = \frac{1}{2} \sum_{h,w} \| Y_{h,w} - \hat{Y}_{h,w} \|_2^2 \]

Keypoint #1 X
Keypoint #1 Y
Keypoint #2 X
Classification Loss (e.g. ImageNet)

\[ \text{P(image is class #1)} \]
\[ \text{P(image is class #2)} \]
\[ \text{P(image is class #F)} \]
Loss function for classification

Network output

- dolphin
- cat
- grizzly bear
- angel fish
- chameleon
- **clown fish**
- iguana
- elephant

Ground truth label

- “clown fish”

Loss $\rightarrow$ error

Slide by Philip Isola
Loss function for classification

Network output

- dolphin
- cat
- grizzly bear
- angel fish
- chameleon
- clown fish
- iguana
- elephant

... → ...

Ground truth label

“clown fish”

Loss → small

Slide by Philip Isola
Loss function for classification

Network output

- dolphin
- cat
- grizzly bear
- angel fish
- chameleon
- clown fish
- iguana
- elephant

Ground truth label

- "grizzly bear"

Loss → large
Loss function for classification

Network output  Ground truth label

\[ H(\hat{z}, z) = - \sum_c \hat{z}_c \log z_c \]

Cross-entropy Loss: Probability of the observed data under the model

Results in learning a probability model $p(c|x)$!

Slide by Philip Isola
Discriminative Deep Networks

“Rockfish”

Slide by Richard Zhang
Discriminative Deep Networks

Raw, Unlabeled Pixels

Slide by Richard Zhang
Generative Deep Networks

Raw, Unlabeled Pixels

Slide by Richard Zhang
Ansel Adams. *Yosemite Valley Bridge.*
Grayscale image: $L$ channel

$X \in \mathbb{R}^{H \times W \times 1}$

Color information: $ab$ channels

$\hat{Y} \in \mathbb{R}^{H \times W \times 2}$

Grayscale image: $L$ channel

$$X \in \mathbb{R}^{H \times W \times 1}$$

Concatenate ($L, ab$) channels

$$(X, \hat{Y})$$

Simple L2 regression doesn’t work 😞

\[ L_2(\hat{Y}, Y) = \frac{1}{2} \sum_{h,w} \| Y_{h,w} - \hat{Y}_{h,w} \|_2^2 \]

Slide by Richard Zhang
\[ L_2(\hat{Y}, Y) = \frac{1}{2} \sum_{h,w} \| Y_{h,w} - \hat{Y}_{h,w} \|_2^2 \]
Better Loss Function

\[ \theta^* = \arg\min_{\theta} \ell(\mathcal{F}_\theta(X), Y) \]

- Regression with L2 loss inadequate
  \[ L_2(\hat{Y}, Y) = \frac{1}{2} \sum_{h,w} \| Y_{h,w} - \hat{Y}_{h,w} \|^2 \]

- Use per-pixel multinomial classification
  \[ L(\hat{Z}, Z) = -\frac{1}{HW} \sum_{h,w} \sum_q Z_{h,w,q} \log(\hat{Z}_{h,w,q}) \]
Designing loss functions

Color distribution cross-entropy loss with colorfulness enhancing term.

[Zhang, Isola, Efros, ECCV 2016]
“You need a lot of data if you want to train/use CNNs”
Transfer Learning

“You need a lot of data if you want to train the CNNs” OR "NOT ALWAYS"
Deep Features & their Embeddings
The Unreasonable Effectiveness of Deep Features

Classes separate in the deep representations and transfer to many tasks. [DeCAF] [Zeiler-Fergus]
Can be used as a generic feature

(“CNN code” = 4096-D vector before classifier)
ImageNet + Deep Learning

- Image Retrieval
- Detection (RCNN)
- Segmentation (FCN)
- Depth Estimation
- …
ImageNet + Deep Learning

Materials?
Geometry?
Parts?
Boundaries?
Pose?

Beagle
Transfer Learning with CNNs

1. Train on Imagenet
Transfer Learning with CNNs

1. Train on Imagenet

2. If small dataset: fix all weights (treat CNN as fixed feature extractor), retrain only the classifier

i.e. swap the Softmax layer at the end
Transfer Learning with CNNs

1. Train on Imagenet

2. If small dataset: fix all weights (treat CNN as fixed feature extractor), retrain only the classifier
   i.e. swap the Softmax layer at the end

3. If you have medium sized dataset, “finetune” instead: use the old weights as initialization, train the full network or only some of the higher layers
   retrain bigger portion of the network, or even all of it.
Transfer Learning with CNNs

1. Train on Imagenet

2. If small dataset: fix all weights (treat CNN as fixed feature extractor), retrain only the classifier
   - i.e. swap the Softmax layer at the end

3. If you have medium sized dataset, “finetune” instead: use the old weights as initialization, train the full network or only some of the higher layers
   - retrain bigger portion of the network, or even all of it.
   - tip: use only ~1/10th of the original learning rate in finetuning to player, and ~1/100th on intermediate layers
Learning an Embedding

CNN

Embedding

shared representation

CNN

Shared W

Image 1

Image 2
Learning an Embedding

Matching

shared representation

CNN

Shared W

CNN
Siamese Network w/ Contrastive Loss

Siamese Architecture
[Chopra 2005, Hadsell 2006]
LEARNING VISUAL SIMILARITY FOR PRODUCT DESIGN WITH CONVOLUTIONAL NEURAL NETWORKS

SEAN BELL AND KAVITA BALA
CORNELL UNIVERSITY
THE PROBLEM

(1) “What is this?”

(2) “Where is it used?”

Name: "Great Bowl O’Fire Sculptural Fire Bowl"
Category: Fire pit
Sold by: John T. Unger, LLC
THE PROBLEM

(1) “What is this?”

(2) “Where is it used?”

Name: “Great Bowl O’Fire Sculptural Fire Bowl”

Category: Fire pit

Sold by: John T. Unger, LLC

Challenge: determine whether these are the same product (different resolution, viewpoint, color, lighting, occlusions)
TWO KINDS OF IMAGES

Iconic

(From a product website)

In context

(Cropped from a scene photo)
PROJECTING INTO A JOINT EMBEDDING

Iconic

In context

Embedding

Project
SEARCH USING THE EMBEDDING

Name: Hemel Ring
Category: Hanging light
Sold by: Holly Hunt
SEARCH USING THE EMBEDDING

“Where is it used?”
CONTRASTIVE LOSS: POSITIVE EXAMPLE

In context

Iconic (same)

Parameters $\theta$

Loss $L_p$

Embedding

$X_q$

$X_p$
CONTRASTIVE LOSS: NEGATIVE EXAMPLE

Loss $L_n$

 Margin $m$

Embedding

In context

Iconic (different)

Parameters $\theta$

$C_{NN}$

$C_{NN}$

$x_q$

$x_n$

$L_n(x_q, x_n) = \max \left( 0, m^2 - \|x_q - x_n\|_2^2 \right)$
CONTRASTIVE LOSS: ALL TOGETHER

\[
L(\theta) = \sum_{(x_q, x_p)} L_p(x_q, x_p) + \sum_{(x_q, x_n)} L_n(x_q, x_n)
\]

- Penalty for similar images that are far away: 
  \[
  L_p(x_q, x_p) = \|x_q - x_p\|_2^2
  \]
- Penalty for dissimilar images that are nearby: 
  \[
  L_n(x_q, x_n) = \max(0, m^2 - \|x_q - x_n\|_2^2)
  \]

Minimize \(L(\theta)\) with **stochastic gradient descent and momentum**

[Chopra 2005, Hadsell 2006]
TRAINING PIPELINE

Image pairs → Stochastic Gradient Descent → CNN Parameters

Image database → CNN → Embedding
RESULTS: “WHAT IS IT?”

In context

Iconic

Top 4 results:
RESULTS: “WHAT IS IT?”

In context
RESULTS: “WHAT IS IT?”

In context

Iconic

Top 4 results:

1. Tolix Stool "Tabouret"
   - Kitchen
   - Contemporary
   - Industrial
   - View Link
   - 2 boxes

2. East Berlin District Metal Barstool
   - Serving Dishes And Plates
   - Modern
   - View Link
   - 0.046

3. Amelia Metal Cafe Barstool
   - In Orange - Set of
   - View Link
   - 0.049

4. Winsome Dubliner 30 in. Bar Stool - Set of 2
   - View Link
   - 0.050
RESULTS: “WHAT IS IT?”

In context
RESULTS: “WHAT IS IT?”

In context Iconic

Top 4 results:

1. Tiara Oval Suspension by Harco Loor
   - Dining Room
   - Contemporary
   - 1 View 1 View 2 boxes

2. HARCO LOOR Tiara Chandelier
   - Chandeliers
   - Contemporary
   - 1 View 1 View

3. Argent N925 Suspension Light
   - Bathroom Vanity Lighting
   - Modern
   - View

4. Eurofase 25620-016 Divo 9 Light Pendant in Nickel 25620-016
   - Pendant Lighting
   - Modern
   - View 1 View 1 View

   - Pendant Lighting
   - Modern
   - View 1 View 1 View
COMPARISON: TRAINED ONLY ON CATEGORIES

In context

Iconic

Top 4 results:

1. Tiara Oval Suspension by Harco Loor
   - Dining Room
   - Contemporary
   - Wall Sconces
   - Unknown style
   - 2 boxes
   - 1 View

2. 'Apollon' Black Shaded 6-light Crystal Chandelier
   - Contemporary
   - 1 View

3. Authentic Deer and Elk Antler Banquet Table Chandelier
   - Chandeliers
   - Contemporary
   - Unknown style
   - View

4. Harco Loor | Tiara Oval HL 15 Suspension Light
   - Pendant Lighting
   - Modern
   - View
   - 1 View

Prices:
- 1792.880
- 1895.330
- 1911.520
- 2035.000
COMPARISON: TRAINED ONLY ON IMAGENET

In context

Iconic

Top 4 results:

- Tiara Oval Suspension by Harco Loor
  - Dining Room
  - Contemporary
  - 1 View.jpg

- "Elephant View" Artwork
  - Paintings
  - Contemporary
  - View.jpg

- Terzani Argent N925 Chandelier
  - Chandeliers
  - Modern

- Mensa Hanging Light Fixture
  - Chandeliers
  - View.jpg

498.493
499.135
503.232
514.867
RESULTS: FAILURE CASE

In context
RESULTS: FAILURE CASE

In context

Iconic

Top 4 results:
RESULTS: “WHERE IS IT USED?”

“Maskros Pendant Lamp”
RESULTS: “WHERE IS IT USED?”

"LEM Piston Stool | Design Within Reach"
SEARCHING ACROSS CATEGORIES
Designing loss functions

Color distribution cross-entropy loss with colorfulness enhancing term.

[Zhang, Isola, Efros, ECCV 2016]
Designing loss functions

Image colorization

Cross entropy loss, with colorfulness term

Super-resolution

“semantic feature loss” (VGG feature covariance matching objective)
Universal loss?
Generative Adversarial Network (GANs)

Generated images vs Real (classifier)

Real photos

[Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, Bengio 2014]
[Goodfellow et al., 2014]
G tries to synthesize fake images that fool D

D tries to identify the fakes

[Goodfellow et al., 2014]
\[
\arg\max_D \mathbb{E}_{x,y} \left[ \log D(G(x)) + \log(1 - D(y)) \right]
\]

[Goodfellow et al., 2014]
G tries to synthesize fake images that fool D:

\[
\arg\min_G \mathbb{E}_{x,y} [ \log D(G(x)) + \log(1 - D(y)) ]
\]

[Goodfellow et al., 2014]
G tries to synthesize fake images that fool the best D:

$$\arg\min_G \max_D \mathbb{E}_{x,y}[ \log D(G(x)) + \log(1 - D(y)) ]$$

[Goodfellow et al., 2014]
G’s perspective: D is a loss function.

Rather than being hand-designed, it is learned.

[Goodfellow et al., 2014]
[Isola et al., 2017]
arg min \_G \_\_D \_E_{x,y}[ \_\_log D(G(x)) + \_\_log(1 - D(y)) ]

[Goodfellow et al., 2014]
\[
\text{arg min}_G \quad \text{max}_D \quad \mathbb{E}_{x,y} \left[ \log D(G(x)) + \log(1 - D(y)) \right]
\]

[Goodfellow et al., 2014]
arg\min_{G} \max_{D} \mathbb{E}_{x, y} \left[ \log D(G(x)) + \log(1 - D(y)) \right]

[Goodfellow et al., 2014]
[Isola et al., 2017]
\[ \arg \min_G \max_D \mathbb{E}_{x,y} \left[ \log D(x, G(x)) + \log(1 - D(x, y)) \right] \]

[Goodfellow et al., 2014]

[Isola et al., 2017]
arg\min_G \max_D \mathbb{E}_{x,y} \left[ \log D(x, G(x)) + \log(1 - D(x, y)) \right]

[Goodfellow et al., 2014]
[Isola et al., 2017]
\[
\arg\min_G \max_D \mathbb{E}_{x,y} \left[ \log D(x, G(x)) + \log(1 - D(x, y)) \right]
\]

[Goodfellow et al., 2014]
[Isola et al., 2017]
\[
\mathop{\arg \max}_G \mathop{\mathop{\mathbb{E}}}_x \mathop{\mathbb{E}}_y \left[ \log D(x, G(x)) + \log(1 - D(x, y)) \right]
\]

[Goodfellow et al., 2014]
[Isola et al., 2017]
BW → Color

Data from [Russakovsky et al. 2015]
BW $\rightarrow$ Color

Data from [Russakovsky et al. 2015]
Data from [maps.google](https://www.maps.google.com)
Labels → Facades

Input

Output

Data from [Tylecek, 2013]
Labels → Facades

Data from [Tylecek, 2013]
Day → Night

Data from [Laffont et al., 2014]
Thermal $\rightarrow$ RGB

Input  Ground-truth  Output
Edges → Images

Edges from [Xie & Tu, 2015]
Sketches $\rightarrow$ Images

Trained on Edges $\rightarrow$ Images

Data from [Eitz, Hays, Alexa, 2012]
Image-to-image translation in PyTorch (e.g., horse2zebra, edges2cats, and more)

- pytorch
- gan
- cyclegan
- pix2pix
- deep-learning
- computer-vision
- computer-graphics
- image-manipulation
- image-generation

Generative adversarial network (GAN)

- generative-adversarial-network
- gans

223 commits
3 branches
0 releases
26 contributors
#edges2cats [Christopher Hesse]

- Ivy Tasi @ivmyt
- Vitaly Vidmirov @vvid
- @gods_tail
- @ka92
Twitter-driven research: #pix2pix

Brannon Dorsey @brannondorsey

Mario Klingemann @quasimondo

Bertrand Gondouin @bgondouin
Scott Eaton (http://www.scott-eaton.com/)
“Do as I Do”
Everybody Dance Now

Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros
UC Berkeley

Source Subject

Target Subject
Results

https://www.youtube.com/watch?v=PCBTZh41Ris&feature=youtu.be
CEO: our own Dr. Tinghui Zhou
Paired training examples

{ bag, bag }  
{ bag, orange bag }  
{ backpack, bag }  
...

Unpaired training examples

\( X \)

\( Y \)
CycleGAN, or “there and back aGAN”

[Zhu*, Park*, Isola, Efros. ICCV 2017]
Cycle-Consistency Loss

\[ \|F(G(x)) - x\|_1 \]
Cycle-Consistency Loss

\[ \| F(G(x)) - x \|_1 \]

\[ \| G(F(y)) - y \|_1 \]
Video
CG to Real

Grand Theft Auto
Real to CG
Failure case
A Neural Algorithm of Artistic Style

Gatys, Ecker, Bethge (arXiv 2015)
Early Vision Texture Models

Heeger & Bergen (1995)
Portilla & Simoncelli (2000)
Start with a noise image as output

Main loop:
- Match pixel histogram of output image to input
- Decompose input and output images using multi-scale filter bank (Steerable Pyramid)
- Match subband histograms of input and output pyramids
- Reconstruct input and output images (collapse the pyramids)
Multi-scale filter decomposition

Filter bank

Input image
Filter response histograms
Match joint histograms of pairs of filter responses at adjacent spatial locations, orientations, and scales.

Optimize using repeated projections onto statistical constraint surfaces
Texture Synthesis

Images with equal model response

Portilla & Simoncelli (2000)
Convolutional Neural Network Texture Model

Gatys et al. (NIPS 2015)
CNN - Multiscale Filter Bank

- **conv1_1**
  - pool1
    - pool2
      - pool3
        - pool4

- # features
  - pool4: 512
  - pool3: 256
  - pool2: 128
  - pool1: 64
  - conv1_1: 64
CNN - Texture Features

\[ F = [\bar{f}_1, \bar{f}_2, \bar{f}_3, \ldots, \bar{f}_N]^T \]

\[ G = FF^T \]

\[
\begin{pmatrix}
\langle \bar{f}_1, \bar{f}_1 \rangle & \cdots & \langle \bar{f}_1, \bar{f}_N \rangle \\
\langle \bar{f}_2, \bar{f}_1 \rangle & \ddots & \\
\vdots & \ddots & \\
\langle \bar{f}_N, \bar{f}_1 \rangle & \cdots & \langle \bar{f}_N, \bar{f}_N \rangle \\
\end{pmatrix}
\]

\[ \langle \bar{f}_i, \bar{f}_j \rangle = \sum_k F_{ik} F_{jk} \]
CNN - Texture Features

Gram Matrices

# features
512
256
128
64
64
Texture Synthesis
Texture Synthesis
Texture Synthesis
Texture Synthesis
Texture Synthesis
Texture Synthesis
Texture Synthesis
Test Julesz’ Conjecture
Test Julesz’ Conjecture
CNN - Texture Synthesis

Gatys et al. (NIPS 2015)
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer

\[ \mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style} \]

\[ \mathcal{L}_{content} = \sum (\mathbf{F} - \mathbf{F})^2 \]

\[ \mathcal{L}_{style} = \sum_l w_l \mathcal{L}_l \]
Artistic Style Transfer

\[ \mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{content}} + \beta \mathcal{L}_{\text{style}} \]

\[ \mathcal{E}_c = \sum (G^c - G^t)^2 \]

\[ \mathcal{E}_s = \sum_l \left( \sum_{i,j} (G^s_i - G^t_j)^2 \right) \]

\[ \mathcal{L}_{\text{content}} = \sum_l \left( \mathcal{E}_c + \mathcal{E}_s \right) \]

\[ \mathcal{L}_{\text{style}} = \sum_l w_l E_l \]
Artistic Style Transfer

\[ L_{total} = \alpha L_{content} + \beta L_{style} \]

\[ L_{content} = \sum (\tilde{F} - F)^2 \]

\[ L_{style} = \sum w_l E_l \]

\[ \frac{\partial C}{\partial \tilde{x}} = \text{Gradient descent} \]

\[ \tilde{x} = \tilde{x} - \alpha \frac{\partial C}{\partial \tilde{x}} \]
Artistic Style Transfer
Artistic Style Transfer
Relative Weighting of Content and Style

1e-4

1e-3

1e-2

1e-1
Different Reconstruction Layers

Conv2_2

Conv4_2
Different Reconstruction Layers

Conv2_2

Conv4_2
Different Reconstruction Layers

Original  Conv2_2  Conv4_2
General Style Transfer
General Style Transfer