Paper Reading

The Unreasonable Effectiveness of Deep Features as a Perceptual Metric

GuoZemin
2018.06.13

The Unreasonable Effectiveness of Deep Features as a Perceptual Metric

Richard Zhang[1] Phillip Isola[12] Alexei A. Efros[1] Eli Shechtman[3] Oliver Wang[3]
[1]UC Berkeley [2]OpenAI [3]Adobe Research
{rich.zhang, isola, efros}@eecs.berkeley.edu {elishe,owang}@adobe.com

Motivation

- Want to assess the perceptual similarity like human way
- Maybe introducing perceptual distance is a good idea

Motivation

- The computer vision community has discovered that internal activations of deep CNNs can be used for a variety of tasks
- Build a new large-scale dataset of human judgements to evaluate questions perceptual similarity

Dataset

- Berkeley-Adobe Perceptual Similarity (BAPPS) dataset
    - Traditional distortions
    - CNN-based distortions

Dataset

- Berkeley-Adobe Perceptual Similarity (BAPPS) dataset

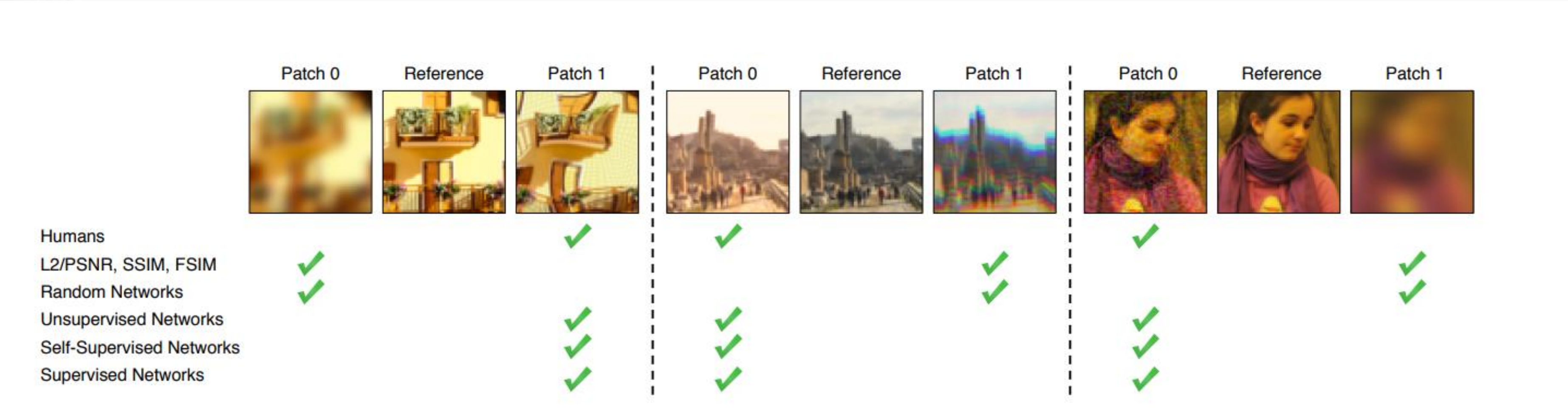
Sub-type	Distortion type
Photometric	lightness shift, color shift, contrast, saturation
Noise	uniform white noise, Gaussian white, pink, & blue noise, Gaussian colored (between violet and brown) noise, checkerboard artifact
Blur	Gaussian, bilateral filtering
Spatial	shifting, affine warp, homography, linear warping, cubic warping, ghosting, chromatic aberration,
Compression	jpeg

Parameter type	Parameters
Input corruption	null, pink noise, white noise, color removal, downsampling
Generator network architecture	# layers, # skip connections, # layers with dropout, force skip connection at highest layer, upsampling method, normalization method, first layer stride # channels in 1^{st} layer, max # channels
Discriminator	number of layers
Loss/Learning	weighting on oixel-wise (ℓ_1), VGG, discriminator losses, learning rate

Dataset	# Input Imgs/ Patches	Input Type	Num Distort.	Distort. Types	# Levels	# Distort. Imgs/Patches	# Judg- ments	Judgment Type
LIVE [51]	29	images	5	traditional	continuous	.8k	25k	MOS
CSIQ [29]	30	images	6	traditional	5	.8k	25k	MOS
TID2008 [46]	25	images	17	traditional	4	2.7k	250k	MOS
TID2013 [45]	25	images	24	traditional	5	3.0k	500k	MOS
BAPPS (2AFC–Distort)	160.8k	64×64 patch	425	trad + CNN	continuous	321.6k	349.8k	2AFC
BAPPS (2AFC–Real alg)	26.9k	64×64 patch	–	alg outputs	–	53.8k	134.5k	2AFC
BAPPS (JND–Distort)	9.6k	64×64 patch	425	trad. + CNN	continuous	9.6k	28.8k	Same/Not same

Dataset

- Berkeley-Adobe Perceptual Similarity (BAPPS) dataset
  - ➢ Focused on perceptual similarity rather than quality assessment
  - ➢ To evaluate different perceptual metrics, employs:
    - Two alternative forced choice (2AFC)
    - Just noticeable difference (JND)


Patch 0
Reference
Patch 1
Patch 0
Reference
Patch 1
Patch 0
Reference
Patch 1
Humans
L2/PSNR, SSIM, FSIM
Random Networks
Unsupervised Networks
Self-Supervised Networks
Supervised Networks

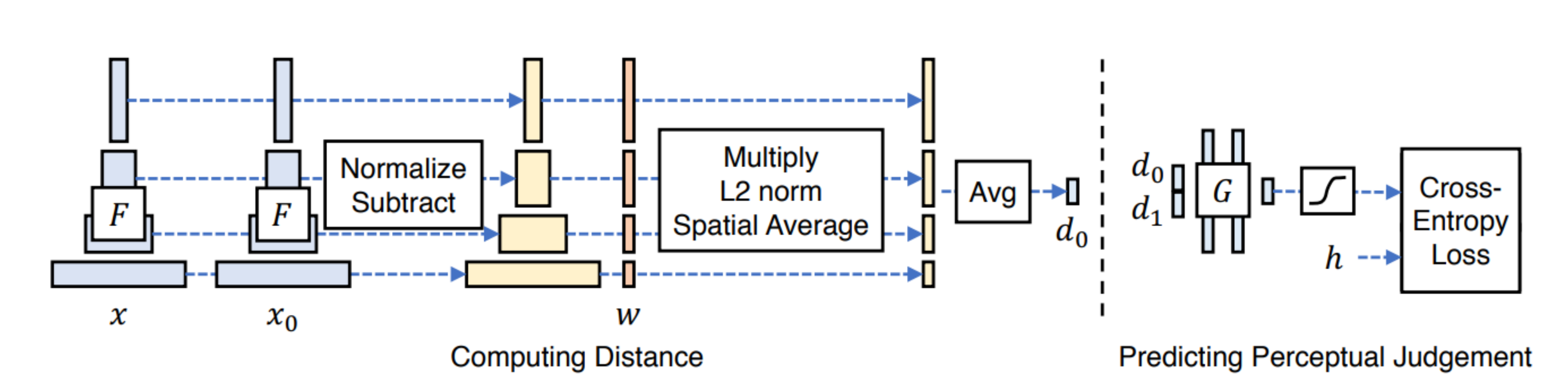
Deep Feature Spaces

- Evaluate feature distance in different network

  - For a convolution network, computer cosine distance in the channel dimension, and average across aptial dimension and layers of the network.
  - VGG
  - AlexNet
  - SqueezeNet

Deep Feature Spaces

- The figure and equation show how to obtain the distance between reference and distorted patches x, x0 with network F.



Loss Function:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} ||w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw})||^2_2$$

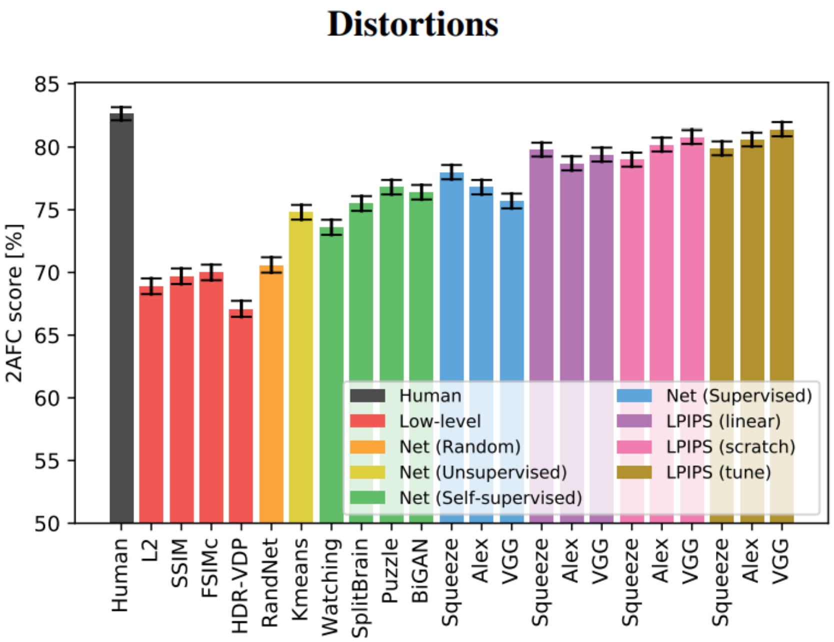
$$\mathcal{L}(x, x_0, x_1, h) = -h \log \mathcal{G}(d(x, x_0), d(x, x_1)) \\ -(1-h) \log(1 - \mathcal{G}(d(x, x_0), d(x, x_1)))$$

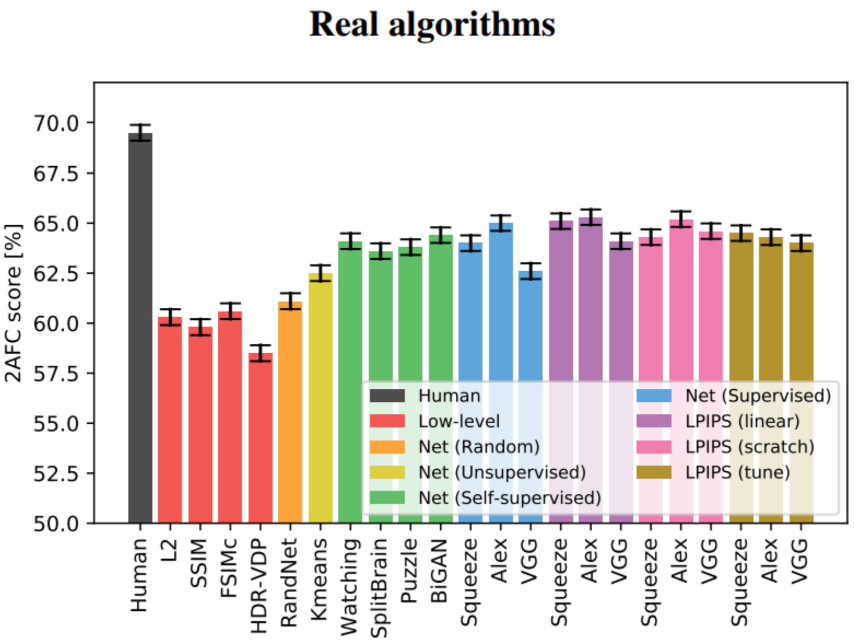
Deep Feature Spaces

- Training data
  - lin
    - Keep pre-trained network weights F fixed, and learn weight w on top.
    - Constitute a perceptual calibration of a few parameters in an existing feature space.
  - tune
    - Initialize from a pre-trained classification model, and allow all weights to be finetuned.

Deep Feature Spaces

- Training data
  - Scratch
    - Initialize the network from random Gaussian weithts and train it entirely on our judgements.


Distortions
2AFC score [%]
85
80
75
70
65
60
55
50
Human
Low-level
Net (Random)
Net (Unsupervised)
Net (Self-supervised)
Net (Supervised)
LPIPS (linear)
LPIPS (scratch)
LPIPS (tune)
Human
L2
SSIM
FSIMc
HDR-VDP
RandNet
Kmeans
Watching
SplitBrain
Puzzle
BiGAN
Squeeze
Alex
VGG
Squeeze
Alex
VGG
Squeeze
Alex
VGG
Squeeze
Alex
VGG


Real algorithms
2AFC score [%]
70.0
67.5
65.0
62.5
60.0
57.5
55.0
52.5
50.0
Human
Low-level
Net (Random)
Net (Unsupervised)
Net (Self-supervised)
Net (Supervised)
LPIPS (linear)
LPIPS (scratch)
LPIPS (tune)
Human
L2
SSIM
FSIMc
HDR-VDP
RandNet
Kmeans
Watching
SplitBrain
Puzzle
BiGAN
Squeeze
Alex
VGG
Squeeze
Alex
VGG
Squeeze
Alex
VGG
Squeeze
Alex
VGG

Experiments:

Subtype	Metric	Distortions			Real Algorithms					All
		Trad-itional	CNN-Based	All	Super-res	Video Deblur	Color-ization	Frame Interp	All	All
Oracle	Human	80.8	84.4	82.6	73.4	67.1	68.8	68.6	69.5	73.9
Low-level	L2	59.9	77.8	68.9	64.7	58.2	63.5	55.0	60.3	63.2
	SSIM [58]	60.3	79.1	69.7	65.1	58.6	58.1	57.7	59.8	63.1
	FSIMc [62]	61.4	78.6	70.0	68.1	59.5	57.3	57.7	60.6	63.8
	HDR-VDP [34]	57.4	76.8	67.1	64.7	59.0	53.7	56.6	58.5	61.4
Net (Random)	Gaussian	60.5	80.7	70.6	64.9	59.5	62.8	57.2	61.1	64.3
Net (Unsupervised)	K-means [26]	66.6	***83.0***	74.8	67.3	59.8	63.1	59.8	62.5	66.6
Net (Self-supervised)	Watching [43]	66.5	80.7	73.6	69.6	60.6	64.4	61.6	64.1	67.2
	Split-Brain [64]	69.5	81.4	75.5	69.6	59.3	64.3	61.1	63.6	67.5
	Puzzle [40]	71.5	82.0	76.8	70.2	60.2	62.8	61.8	63.8	68.1
	BiGAN [13]	69.8	***83.0***	76.4	70.7	60.5	63.7	62.5	64.4	***68.4***
Net (Supervised)	SqueezeNet [20]	**73.3**	***82.6***	**78.0**	70.1	60.1	63.6	62.0	64.0	***68.6***
	AlexNet [27]	70.6	**83.1**	76.8	**71.7**	60.7	65.0	62.7	***65.0***	**68.9**
	VGG [52]	70.1	81.3	75.7	69.0	59.0	60.2	62.1	62.6	67.0
*LPIPS (Learned Perceptual Image Patch Similarity)	Squeeze – lin	76.1	83.5	79.8	71.1	***60.8***	***65.3***	***63.2***	***65.1***	70.0
	Alex – lin	73.9	83.4	78.7	***71.5***	**61.2**	***65.3***	***63.2***	**65.3**	69.8
	VGG – lin	76.0	82.8	79.4	70.5	60.5	62.5	***63.0***	64.1	69.2
	Squeeze – scratch	74.9	83.1	79.0	71.1	***60.8***	63.0	62.4	64.3	69.2
	Alex – scratch	77.6	82.8	80.2	71.1	***61.0***	**65.6**	**63.3**	***65.2***	**70.2**
	VGG – scratch	77.9	**83.7**	80.8	71.1	60.6	64.0	***62.9***	64.6	70.0
	Squeeze – tune	76.7	83.2	79.9	70.4	***61.1***	63.2	***63.2***	64.5	69.6
	Alex – tune	77.7	83.5	80.6	69.1	60.5	64.8	***62.9***	64.3	69.7
	VGG – tune	**79.3**	83.5	**81.4**	69.8	60.5	63.4	62.3	64.0	69.8