Shift-tolerant Perceptual Similarity Metric

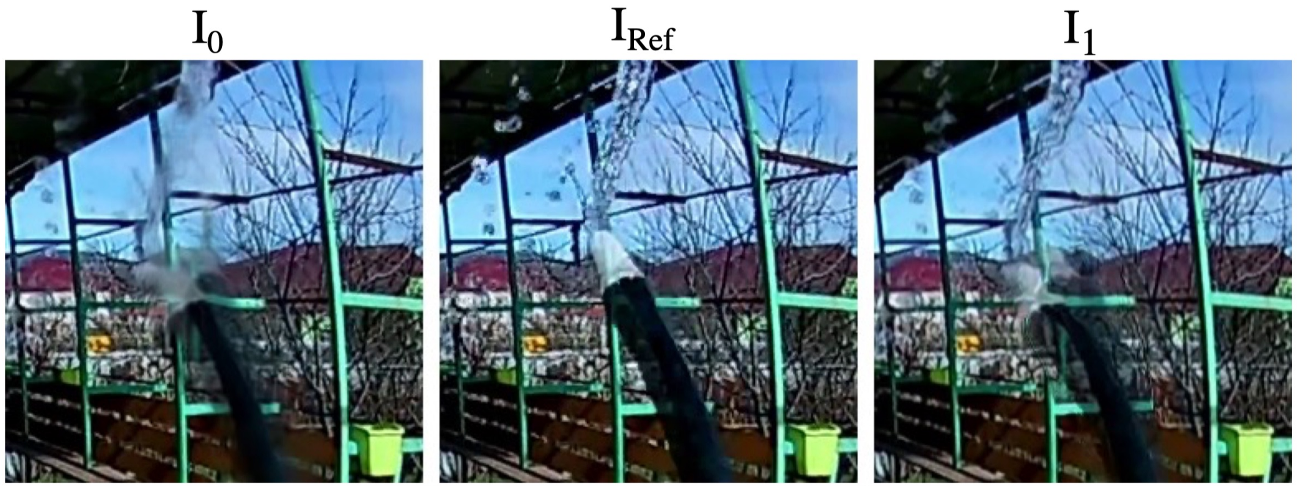
Abhijay Ghildyal and Feng Liu

Portland State University, OR 97201, USA
{abhijay,fliu}@pdx.edu

ECCV 2022

Motivation

- “how similarity metrics work on a pair of images that are **not perfectly aligned**”



No-shift	1-pix-shift	Metric	No-shift	1-pix-shift
		Humans	✓	✓
✓		MS-SSIM		✓
✓		L2		✓
✓		LPIPS		✓

-> develop a **shift-tolerant** perceptual similarity metric

-> from a perspective of **network framework**

Article Structure

- 3 Human Perception of Small Shifts -> subjective experiment
- 4 Effect of Small Shifts on Similarity Metrics -> conflict with subjective experiment
- 5 Elements of Shift-tolerant Metrics
- 6 Experiments

Subjective Experiment

- Hypothesis: it is difficult for people to detect a small shift in images

- Setting:

  50 pairs: 5 pairs for each 0-9 pix-shift

  presentation: two images placed side by side

  participants: 32

Pixel shift	Number of user responses			Avg. of std. in user responses per sample
	Said Yes (Same)	Said No (Shifted)	Yes%	
0	140	10	93.3%	0.09 ± 0.17
1	121	29	80.7%	0.19 ± 0.23
2	84	66	56.0%	0.34 ± 0.21
3	52	98	34.7%	0.24 ± 0.23
4	52	98	34.7%	0.30 ± 0.24
5	40	110	26.7%	0.23 ± 0.24
6	35	115	23.3%	0.21 ± 0.24
7	31	119	20.7%	0.12 ± 0.20
8	27	123	18.0%	0.18 ± 0.23
9	15	135	10.0%	0.13 ± 0.21

-> verifies the hypothesis

Performance of Similarity Metrics

- Experiments on BAPPS dataset

- reference image I_r distorted image I_{d1}, I_{d2}
- predicted score $s_1 = S(I_r, I_{d1}),\ s_2 = S(I_r, I_{d2})$
- Evaluation index

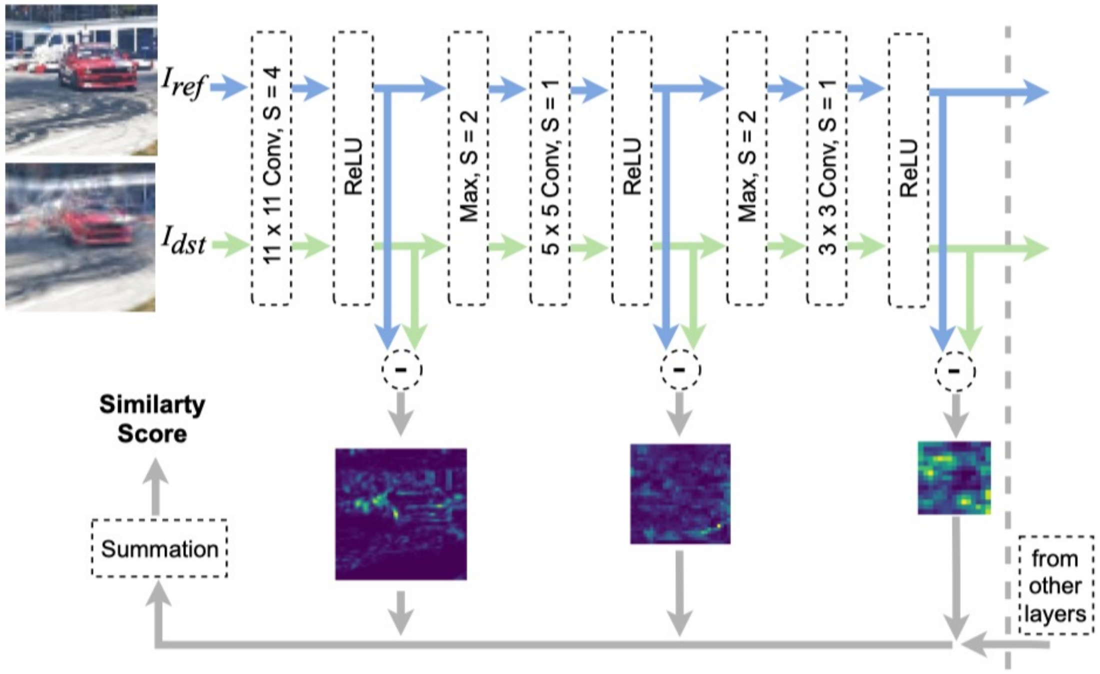
$$r_{rf} = \frac{1}{N}\sum_{l=1}^{N}(s_1^l < s_2^l) \neq (\hat{s}_1^l < \hat{s}_2^l)$$

Network	2AFC	r_{rf}		
		1pixel	2pixel	3pixel
L2	62.92	3.59	7.55	10.82
SSIM [30]	61.41	3.16	7.20	13.73
CW-SSIM [31]	61.48	3.91	6.88	9.47
MS-SSIM [32]	62.54	2.22	5.83	10.66
PIEAPP Sparse [25]	64.20	2.83	3.19	3.81
PIEAPP Dense [25]	64.15	2.97	1.37	3.33
PIM-1 [3]	67.45	0.79	1.70	2.52
PIM-5 [3]	67.38	1.01	1.88	2.96
GTI-CNN [21]	63.87	3.95	4.91	7.88
DISTS [6]	68.83	2.85	2.89	4.03
E-LPIPS [16]	68.22	5.84	5.86	5.77
LPIPS (Alex) [37]	68.59	2.81	3.41	3.84
LPIPS (Alex) §*†	70.54	2.58	3.59	3.53
LPIPS (Alex) ours*†	70.39	0.66	1.24	1.79
LPIPS (Alex) §*‡	**70.65**	2.87	3.92	3.74
LPIPS (Alex) ours*‡	70.48	**0.57**	**1.06**	**1.50**

(§) Retrained from scratch. (*) Trained on patches of size 256 using author's (†) / our (‡) setup.

Method

- design a deep neural network resistant to small shifts
- baseline: LPIPS(AlexNet)



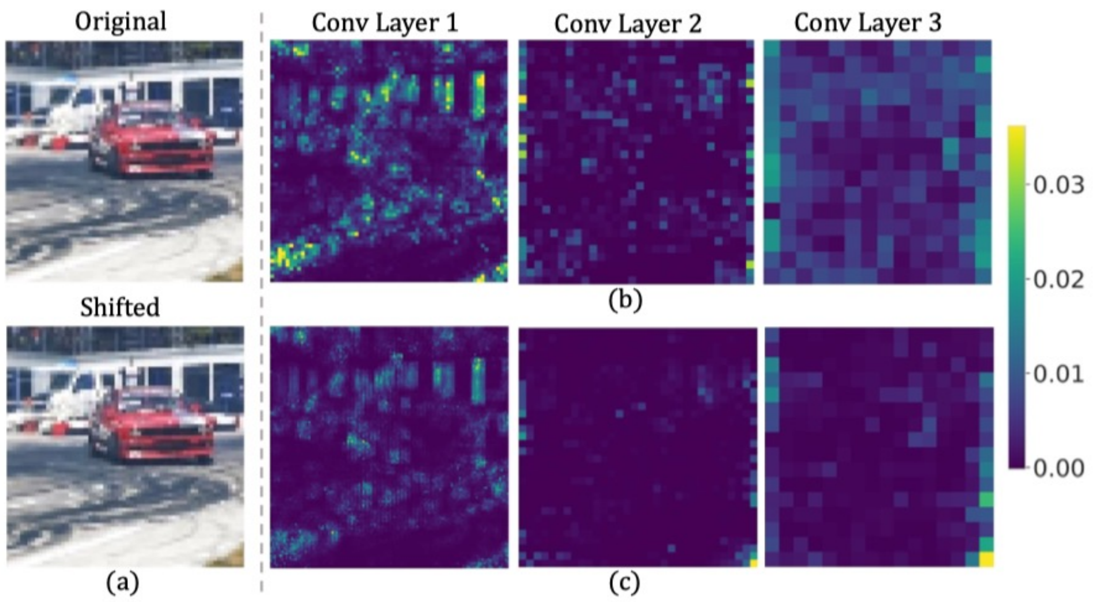
Method

Attempts:

- Reducing Stride
  - AlexNet: conv-1: S =4 MaxPooling: S=2

- Anti-aliasing
  - normal convolution: shift-equivariance
  - BlurPool: a Gaussian filter+ a downsampling operator with stride S

    conv1(S=4)->conv1(S=2)+BlurPool(S=2)



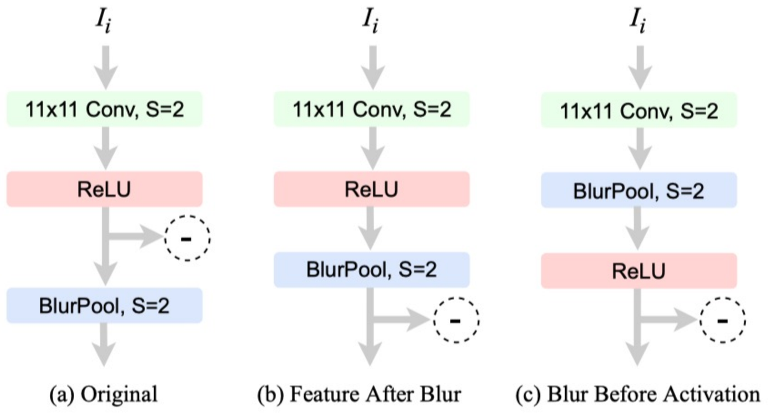
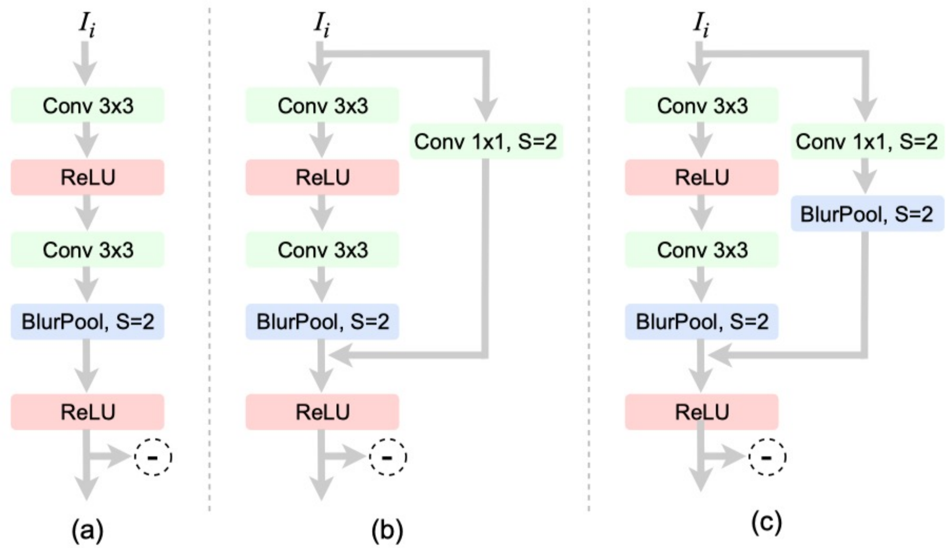


Fig. 4: Alternative positions of *BlurPool*.

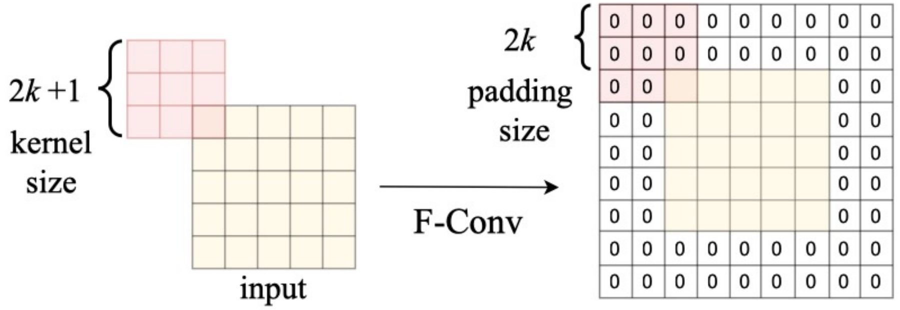
Method

- Pooling
  - shift-invariance: MaxPooling>AvePooling
  - MaxBlurPooling (Gaussian filter+MaxBlurPooling)
  - AveBlurPooling (Gaussian filter+AveBlurPooling)



- Strided-skip Connections

- Border Handling
  - F-conv: every element of the filter needs to be applied to every pixel in the input image



Experiment

- Comparisons to Existing Metrics

Table 4: Experiments on the CLIC dataset.

Network	Accuracy(%)	No. of rank flips		
		1pixel	2pixel	3pixel
L2	58.16	833	2102	2214
SSIM [30]	60.00	349	931	1109
PIEAPP [25]	75.44	91	134	158
E-LPIPS [16]	74.44	212	251	317
DISTS [6]	75.63	28	36	50
PIM-1 [3]	73.79	**13**	22	33
LPIPS(Alex) [37]	73.68	90	108	121
LPIPS(Alex)§*†	76.53	59	51	62
LPIPS(Alex) ours*†	**76.97**	17	**14**	**21**

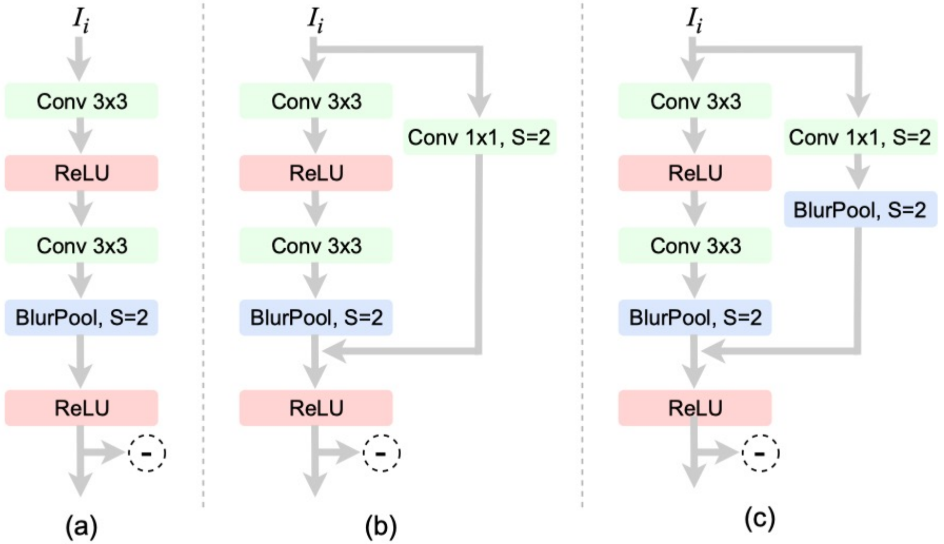
(§) Retrained from scratch. (∗) Trained on image patches of size 64 using author's (†) setup.

Experiment

- Effect of BlurPool locations

Table 7: Effect of *BlurPool* locations within an anti-aliased strided convolution (Figure 4).

Anti-Alias (BlurPool)	Stride in *Conv-1*	BlurPool Location	2AFC	r_{rf} 1pixel	r_{rf} 2pixel	r_{rf} 3pixel
✓	2	Original	**70.67**	**1.46**	**1.82**	**2.25**
✓	2	FeatAfterBlur	70.55	1.73	1.84	2.49
✓	2	BlurBeforeAct	70.50	2.06	2.02	2.74
✓	1	Original	70.42	0.66	1.13	1.83
✓	1	FeatAfterBlur	**70.52**	0.69	1.11	1.60
✓	1	BlurBeforeAct	70.48	**0.57**	**1.06**	**1.50**



Experiment

- Effect on different backbone networks

Table 6: Anti-aliasing via *BlurPool* can significantly improve shift-tolerance and often improve 2AFC scores consistently for different backbone networks.

Network	AA (BlurPool) Reflection-Pad		2AFC	r_{rf}		
	1	2		1pixel	2pixel	3pixel
VGG-16			70.03	3.01	3.76	3.44
	✓		70.05	**0.66**	**1.08**	**1.44**
		✓	**70.07**	**0.66**	1.12	1.82
ResNet-18			69.86	2.67	3.35	3.77
	✓		69.95	**0.82**	**1.51**	**2.19**
		✓	**70.14**	1.07	1.81	2.38
Squeeze			69.61	7.41	7.58	10.35
	✓		69.24	**2.03**	3.06	3.93
		✓	**69.44**	2.10	**2.48**	**3.42**

Experiment

- Just noticeable differences

Table 8: Consistency of perceptual similarity metrics with the sensitivity of human perception to pixel shifts.

Metric	JND mAP%
SSIM [30]	0.722
LPIPS (Alex) [37]	0.757
LPIPS (Alex) §∗†	0.740
LPIPS (Alex) **ours**∗†	0.771
LPIPS (VGG) [37]	0.770
LPIPS (VGG) §∗†	0.769
LPIPS (VGG) **ours**∗†	**0.775**
DISTS [6]	0.766
PIM-1 [3]	0.773

(§) Retrained from scratch. (∗) Trained on image patches of size 64 using author's (†) setup.

Summary

- a shift-tolerant similarity measure from the perspective of network architecture
- some elegant changes in network architecture

+ a clear writing logic and structure
+ a complete research process (question raising, verification, and solution)
+ intuitive experiment about tolerance of tiny shift

- lack of novelty