Differentiable JPEG: The Devil is in the Details

NEC Laboratories America

TECHNISCHE UNIVERSITÄT DARMSTADT

Christoph Reich[1,2], Biplob Debnath[2], Deep Patel[2], and Srimat Chakradhar[2]

[1] Technischen Universität Darmstadt, [2] NEC Laboratories America, Inc.

Summary

Can we make JPEG encoding-decoding differentiable?

- Standard JPEG coding [1] is non-differentiable
- Non-diff. inhabits the use of JPEG in gradient-based learning systems
- 🚀 We analyze issues with current differentiable JPEG approaches
- 🚀 **We present a novel differentiable JPEG approach**

Use Our Differentiable JPEG Approach

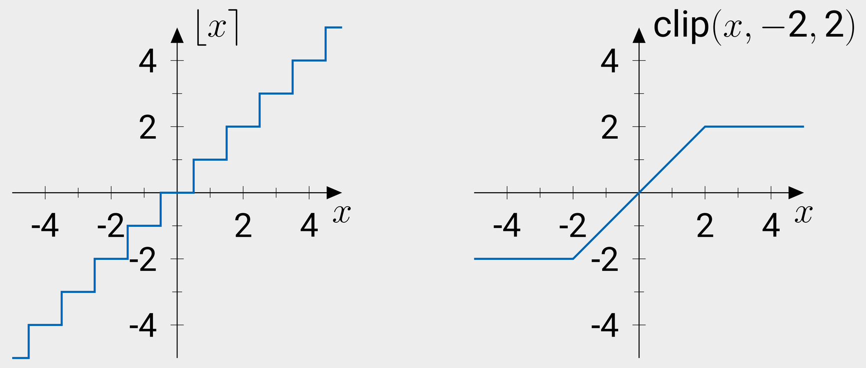
```python
import torch
from torch import Tensor
from diff_jpeg import diff_jpeg_coding # Import our diff. JPEG approach

# Init random image and JPEG quality
image: Tensor = torch.randint(low=0, high=256, size=(4, 3, 1904, 1904))
jpeg_quality: Tensor = torch.tensor([2.0, 99.0, 1.0, 11.0])
# Perform differentiable JPEG coding
image_coded: Tensor = diff_jpeg_coding(image, jpeg_quality)
```

Check out our open source PyTorch implementation!

Differentiable JPEG Coding

Non-differentiability of JPEG



- Rounding (quantization) and clipping function used in standard JPEG
- Gradient of rounding func. is zero a.e. or undefined
- Gradient of clipping func. is zero for clipped values

Our differentiable JPEG models all crucial discretizations & bounds

- DCT feature quantization
- Quantization table scale flooring
- Quantization table flooring
- Quantization table clipping
- Output image clipping
- **Existing work only considers DCT feat. quantization**

Differentiable surrogate functions of discretizations & bounds

- **Differentiable rounding** [2]

$$\lfloor x \rceil \approx \lfloor x \rceil + (x - \lfloor x \rceil)^3$$

- **Differentiable flooring**

$$\lfloor x \rfloor \approx \lfloor x \rceil + (x - 0.5 - \lfloor x \rceil)^3$$

- **Differentiable clipping**

$$\text{clip}(x) \approx \begin{cases} x & \text{if } x \in [b_{\text{min}}, b_{\text{max}}] \\ b_{\text{min}} + \gamma\,(x - b_{\text{min}}) & \text{if } x < b_{\text{min}} \\ b_{\text{max}} + \gamma\,(x - b_{\text{max}}) & \text{if } x > b_{\text{max}} \end{cases}, \gamma \in (0, 1].$$

Differentiable JPEG coding with Straight-Through Estimation

- STE [3] assumes a constant grad.
- Our STE uses the grad. of the surrogate

$$\lfloor x \rceil_{\text{STE}} = \begin{cases} \lfloor x \rceil & \text{fw. pass} \\ \frac{\mathrm{d}}{\mathrm{d}x}\lfloor x \rceil + (x - \lfloor x \rceil)^3 & \text{bw. pass} \end{cases}$$

Forward Function Results

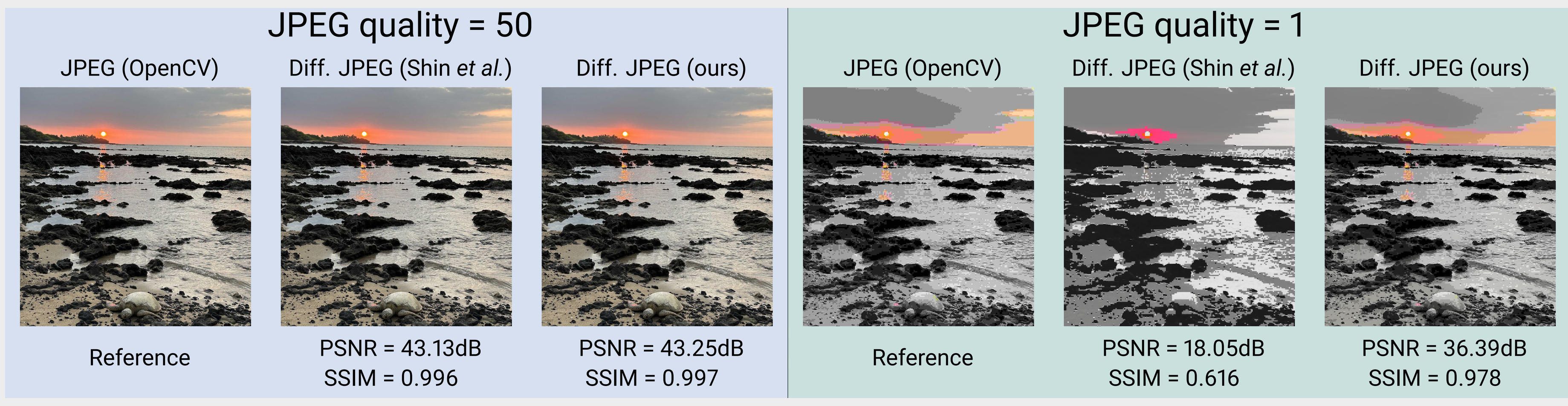


Fig. 1 Qualitative results of our diff. JPEG approach *v.s.* Shin et al. [2] in approximating standard JPEG.

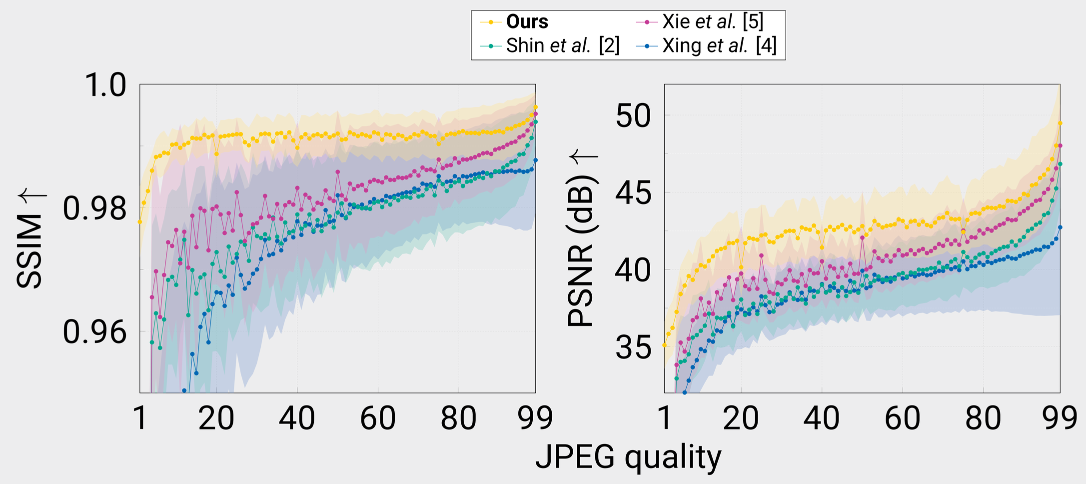


Fig. 2 Forward function results.

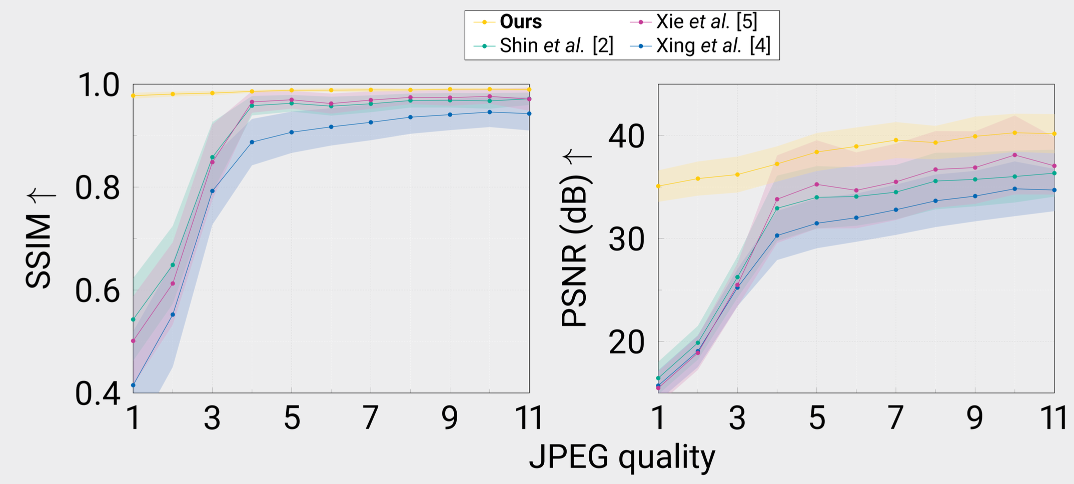


Fig. 3 Forward function results for strong compression.

- **Existing approaches fail to approximate standard JPEG over the full JPEG quality range**
- 🚀 **Our diff. JPEG approach approximates standard (non-diff.) JPEG well over the full JPEG quality range**

Backward Function Results

Use adversarial attacks through diff. JPEG to show backward performance

		Top-1 acc ↓			Top-5 acc ↓		
Approach	q range →	1-99	1-10	11-99	1-99	1-10	11-99
Xing *et al.* [4]		43.44	24.42	45.82	72.52	45.55	75.90
Xie *et al.* [5]		25.30	14.72	26.63	46.55	31.47	48.43
Shin *et al.* [2]		15.11	8.98	15.88	27.21	19.99	*28.11*
Our diff. JPEG		**14.39**	**7.97**	**15.19**	**25.79**	**17.53**	**26.83**
Our diff. STE JPEG		*15.00*	*8.35*	*15.83*	*27.07*	*18.73*	28.12

Tab. 1 Backward function results (IFGSM [6] w/ $\epsilon = 3$).

- 🚀 **Our differentiable JPEG leads to better adversarial samples**
- Strong adversarial results show the "usefulness" of the obtained gradients for grad.-based optimization

Straight-Through Estimator Results

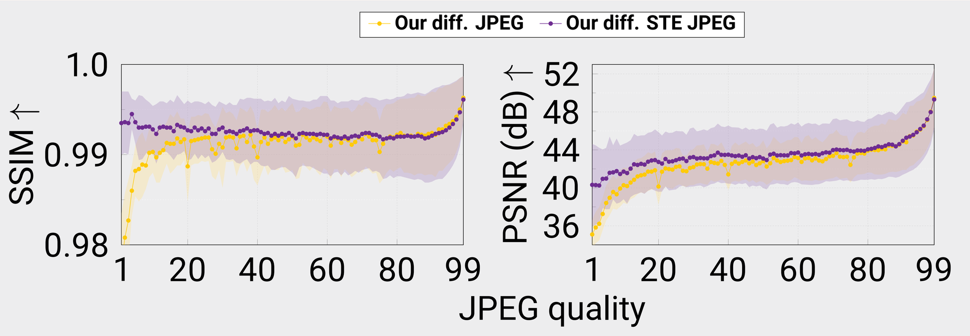


Fig. 4 Forward function results for strong compression.

		Top-1 acc ↓			Top-5 acc ↓		
Backw. approach	q range →	1-99	1-10	11-99	1-99	1-10	11-99
Constant grad. (stand. STE)		25.30	21.62	25.76	45.38	41.37	45.88
Surrogate (ours)		**7.14**	**4.14**	**7.51**	**13.11**	**8.89**	**13.64**

Tab. 2 Backward STE ablation (IFGSM [6] w/ $\epsilon = 3$).

- Using STE leads to a better forward performance
- Our STE approach outperforms stand. STE (bw. perf.)

Ablations

		SSIM ↑			PSNR ↑		
Configuration	q range →	1-99	1-10	11-99	1-99	1-10	11-99
A Shin *et al.* [2]		0.969	0.888	0.979	38.71	31.07	39.66
B + diff. QT clipping		0.978	0.966	0.979	39.16	35.10	39.67
C + diff. QT floor		0.983	0.971	0.985	41.03	35.95	41.66
D + diff. QT scale floor		0.984	0.971	0.986	41.08	35.96	41.72
E + diff. output clipping **(our diff. JPEG)**		*0.991*	*0.987*	*0.992*	*42.60*	*38.28*	*43.14*
F + STE **(our diff. STE JPEG)**		**0.993**	**0.993**	**0.992**	**43.49**	**41.14**	**43.78**

Tab. 3 Forward function summary & ablation.

- 🚀 Introduced parts consistently improve performance

		Top-1 acc ↓			Top-5 acc ↓		
Function	q range →	1-99	1-10	11-99	1-99	1-10	11-99
Fourier		39.53	20.16	41.95	68.98	40.81	72.50
Linear		25.69	22.41	26.10	46.52	42.84	46.98
Polynomial		**14.39**	*7.97*	**15.19**	**25.79**	*17.53*	**26.83**
Sigmoid		*20.28*	**6.34**	*22.02*	*36.79*	**14.44**	*39.59*
Tanh		22.52	15.20	23.43	41.80	32.79	42.92

Tab. 4 Backward rounding ablation (IFGSM w/ $\epsilon = 3$).

- Round/floor approximation is crucial for a good backward performance

References

[1] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
[2] R. Shin and D. Song, "JPEG-resistant Adversarial Images," in *NIPS Workshop on Machine Learning and Computer Security*, vol. 1, 2017, p. 8.
[3] Y. Bengio *et al.*, "Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation," *arXiv:1308.3432*, 2013.
[4] Y. Xing, Z. Qian, and Q. Chen, "Invertible Image Signal Processing," in *CVPR*, 2021, pp. 6287–6296.
[5] X. Xie, N. Zhou, W. Zhu, and J. Liu, "Bandwidth-Aware Adaptive Codec for DNN Inference Offloading in IoT," in *ECCV*, 2022, pp. 88–104.
[6] A. Kurakin, I. J. Goodfellow, and S. Bengio, "Adversarial Machine Learning at Scale," in *ICLR*, 2017.

Paper: `https://arxiv.org/abs/2309.06978`
Project Page: `https://christophreich1996.github.io/differentiable_jpeg/`
Code: `https://github.com/necla-ml/Diff-JPEG`
Correspondence: `christoph.reich@bcs.tu-darmstadt.de`

JAN 4-8 WACV 2024 WAIKOLOA HAWAII