Riemannian Loss for Image Restoration

Jing Mu[1], Xinfeng Zhang[1], Shuyuan Zhu[2], and Ruiqin Xiong[1]

1 Department of Computer Science, Peking University, Beijing, China

2 School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China.

jmu@pku.edu.cn

Abstract

Deep neural networks are widely used for image restoration, however the loss criteration is usually set as ℓ_2. ℓ_2 penalizes larger errors, which is unstable for outliers. To avoid the disadvantages, ℓ_1 is utilized as a more robust and well behaved loss. This paper proposes a novel loss function for restoration networks, which measures geodesic distance in Riemannian manifold and exploits the outstanding properties of ℓ_1. Different from ℓ_1 and ℓ_2 loss which reflects pixel distance, our loss in Riemannian reflects the structure distance of image. The proposed loss not only preserves the robustness of ℓ_1 loss, but also reflects the image contrasts. Experimental results on image super resolution and compressed sensing show that our proposed loss function achieves more accurate reconstructions, according to both the objective and perceptual qualities.

1. Introduction

Image restoration is a fundamental problem in image processing, which reconstructs the high-quality image $\mathbf{x}$ from its corrupted observation $\mathbf{y} = \mathbf{Hx} + \mathbf{v}$, where $\mathbf{v}$ is the additive noise, and $\mathbf{H}$ is the degradation operation, e.g. convolution and down-sampling. The success of deep neural networks has led to dramatic improvements in image restoration [1–4], including compressed sensing [2] and super-resolution [1, 4], etc. However researchers have focused on the new architectures, and paid less attention on the loss function, which is effective for network learning.

$\ell_p(p = 1,2)$ is the most widely used measurement for restoration networks. The ℓ_2 loss (MSE) is the major performance measure of image (PSNR). However, it works under the assumption of white Gaussian noise, which is not valid in general. Also the outliers have a great influence on the weight allocation of ℓ_2 loss, which sharply reduces the network performance. Recently, networks trained with ℓ_1 as loss function provides a promising and robust performance [3].

This paper proposes a loss function, which measures the geodesic distance of image. We take advantage of the success of ℓ_1 loss and measure the ℓ_1 norm of geodesic distance. Fillard et al. [5] defined a new family of metrics on symmetric positive definite named Log-Euclidean Metric (LEM), which is geodesic distance induced by Riemannian metrics [6]. The geodesic distance between points T_i and T_j is written as: $d(T_i, T_j) = \left\|\log(T_i) - \log(T_j)\right\|_F$, where $\|\cdot\|_F$ denotes the Frobenius matrix norm. The LEM can be regarded as the Euclidean distance in logarithm domain, which not only has the properties of Euclidean distance, but also has promising invariance properties such as inversion invariant and similarity invariant [6]. According to Weber's law [7], the small contrasts which are more sensitive to perceive by human eyes are enhanced.

We compare the proposed loss with ℓ_p loss on super resolution and compressed sensing. And the experimental results show that the proposed loss function outperforms ℓ_p in both objective and subjective qualities. To summarize, the main contributions of our work are concluded as: i) We focus on the loss metric of image restoration networks which is significant fot network performance, and propose a Riemannian loss which reflects the image structures. ii) *To the best of our knowledge, we are the first to utilize the LEM-based loss on image restoration networks.*

2. The Proposed Loss Function

Inspired by the properties of Riemannian metric, we propose a new loss function for image restoration network to reflect the image structures. As Gaussian kernel $\exp\left(-\gamma\|\cdot\|_2^2\right)$ is the most popular and versatile kernel in Euclidean spaces [8]. It is attractive to generalize this kernel to manifolds, by exploiting the geodesic distance to replace the Euclidean distance in the Gaussian kernel.

Inspired the idea of ℓ_1 norm, the proposed loss is:

$$L(\boldsymbol{x}_i, \boldsymbol{y}_i) = \frac{1}{N}\sum_i \exp\left(\gamma\left|\log|\boldsymbol{y}_i| - \log|\boldsymbol{x}_i|\right|\right) \tag{1}$$

where $\gamma \leq 1 \big/ \left|\log|\boldsymbol{x}_i| - \log|\boldsymbol{y}_i|\right|$.

The derivative of the loss function is:

$$\begin{aligned}\frac{\partial L(\boldsymbol{x}_i,\boldsymbol{y}_i)}{\partial \mathbf{y}_i} &= \exp\left(\gamma\left|\log|\mathbf{y}_i| - \log|\mathbf{x}_i|\right|\right) \\ &\cdot\left(\gamma \cdot sign\left(\log|\mathbf{y}_i| - \log|\mathbf{x}_i|\right) \cdot \frac{|\mathbf{x}_i|}{|\mathbf{y}_i|} \cdot sign\left(\mathbf{y}_i\right)\right)\end{aligned} \tag{2}$$

(a) Gradient of the observed image

(b) Gradient of the transformation

Figure 1: Gradients of logarithmic transformed image.

, since $\forall i \neq j, \partial L(\boldsymbol{x}_i, \boldsymbol{y}_i)/\partial \boldsymbol{y}_j = 0$. $\exp(\cdot) > 0$, $\frac{1}{|x|} \cdot sign(x)$ is not differentiable at 0. As we do not need to update the weights when error equals 0, we set const $\cdot \frac{0}{|0|} \cdot sign(0) = 0$. So that our proposed loss function is differentiable under all the circumstance.

In addition, $\log\left|\frac{y}{x}\right|$ can be interpreted into Weber's law [7], which simulates human vision perception mechanism. Weber's law states that *just noticeable difference* is a constant proportion of the original stimulus value, i.e. when the intensity variation reaches the constant it can be observed by human. According to Weber's law, the logarithmic transformation can also enhance the small contrasts, which is more sensitive to perceive by human eyes. More specifically, for image $\mathbf{x}$, its gradient variation in the linear domain is $\nabla \mathbf{x}$, and its gradient variation in the log-transformed domain is $\nabla \log(\mathbf{x}) = \frac{1}{\mathbf{x}} \nabla \mathbf{x}$. That is to say, when $\mathbf{x}$ is very small, the gradient variation in the log-transformed domain is amplified, as shown in Fig. 1.

As mentioned above, the proposed loss not only has the characteristics of ℓ_p loss, but also describes the geodesic distance which reflects the image structures.

3. Experimental Results

Due to the space limitation, in this section, we only demonstrate the results of two restoration tasks- super resolution and image compressive sensing reconstruction. Extensive evaluation experiments show that the proposed metric outperforms the ℓ_p loss with the state-of-the-art networks both quantitatively and qualitatively. For the time complexity, our loss function is at the same level of the original network. It is worth mentioning that we compare the performance of different loss functions without modifying the network architectures.

3.1. Image Super Resolution

In recent years, there are a large number of networks for image super resolution, including SRCNN [1], VDSR [4] and EDSR [3], etc. We use the same training dataset and architectures with these methods. For SRCNN, we use bilinear method for initialization, since bicubic interpolation introduces high-frequency artifacts. For EDSR, due to the hardware limitation, we only train the baseline model in [3]. The Set5, Set14, B100, and Urban100 in [4] are used to evaluate the performance of upscaling factors 2, 3, and 4. The PSNR comparison is shown in Table. 1, and our loss presents the highest average PSNR. The visual comparison is shown in Fig. 2. It is evident that the images recovered by the proposed scheme better preserve edges and textures. Especially for the regular region, only our loss perfectly reconstructs the line in the images.

3.2. Image Compressive Sensing Reconstruction

In recent years, there are many networks for image compressive sensing reconstruction, and ReconNet [2] is the benchmark method. We employ the proposed loss function on this network to compare with the performance of ℓ_p metric as loss function.

The Set12 [9] and LIVE1 [9] are used to evaluate the performance of rate 0.1, 0.25 and 0.4. The PSNR comparison is shown in Figure. 3, and our loss function presents the highest average PSNR. Fig. 4 demonstrates visual improvement of the proposed metric at ratio 0.25. It is evident that the images recovered by the proposed scheme better preserve details and textures. We also utilize some other methods to measure the validation of our loss function, including SSIM, GMSD [10], FSIM and FSIMc [11]. Our proposed loss function outperforms ℓ_p loss on all the quality assessment metrics.

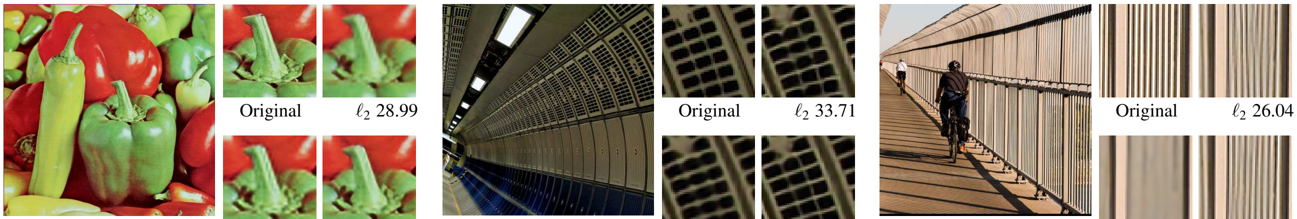
4. Conclusions

In this paper, we focus on the loss layer of neural networks which is significant for image restoration framework. We present a novel Riemannian metric with ℓ_1 norm, which measures the geodesic distance of image. Our proposed loss can enhance the small contrasts, meanwhile preserve the large errors. Our loss not only inherits the advantages of ℓ_p, but also reflects the image structures. The new loss is simple, fast, suitable for most network architectures. We choose several state-of-the-art networks of super resolution and compressed sensing reconstruction. Our loss function outperforms ℓ_p loss both objectively and subjectively.

References

[1] D. et al., "Learning a deep convolutional network for image super-resolution," in *Springer European Conference on Computer Vision*, 2014, pp. 184–199.

[2] K. et al., "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 449–458.



(a) Peppers from Set14 — Original, ℓ_2 28.99, ℓ_1 28.82, Our's 29.70

(b) Image078 from Urban100 — Original, ℓ_2 33.71, ℓ_1 32.61, Our's 34.48

(c) Image024 from Urban100 — Original, ℓ_2 26.04, ℓ_1 25.35, Our's 26.61

Figure 2: Qualitative comparison of super-resolution. (a) 4× SRCNN. (b) 3× VDSR (c) 3× EDSR

Table 1: PSNR (UNIT: dB) Results of All the Loss Functions Used for Super Resolution

DataSet	Scale	SRCNN			VDSR			EDSR		
		ℓ_2	ℓ_1	Our's	ℓ_2	ℓ_1	Our's	ℓ_2	ℓ_1	Our's
Set5	×2	32.86	33.66	**34.22**	37.50	36.85	**37.72**	37.79	37.84	**37.87**
	×3	29.56	30.27	**30.56**	33.62	32.97	**33.83**	34.30	34.34	**34.40**
	×4	27.50	26.92	**28.26**	31.35	30.69	**31.58**	32.07	32.09	**32.15**
Set14	×2	29.32	29.87	**30.26**	33.02	32.37	**33.15**	33.43	33.47	**33.51**
	×3	26.64	27.08	**27.26**	29.76	29.48	**29.95**	30.09	30.13	**30.18**
	×4	25.01	24.63	**25.51**	28.01	27.48	**28.23**	28.47	28.50	**28.56**
B100	×2	29.07	29.35	**29.63**	31.90	31.47	**32.04**	32.04	32.09	**32.14**
	×3	26.39	26.64	**26.88**	28.80	28.65	**28.99**	28.95	28.99	**29.05**
	×4	25.03	24.64	**25.27**	27.29	27.08	**27.41**	27.46	27.51	**27.56**
Urban100	×2	26.14	26.63	**27.21**	30.76	29.46	**30.92**	31.77	31.84	**31.92**
	×3	23.59	24.01	**24.35**	27.14	26.57	**27.38**	27.99	28.03	**28.11**
	×4	22.16	20.80	**22.60**	25.17	24.72	**25.34**	25.88	25.91	**25.98**

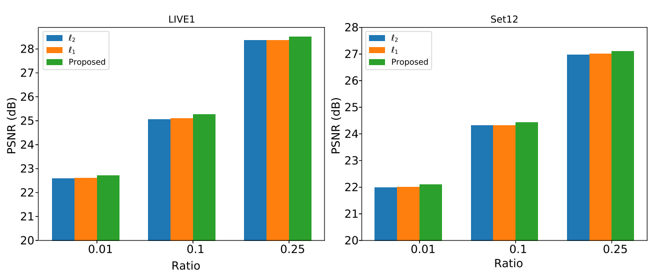
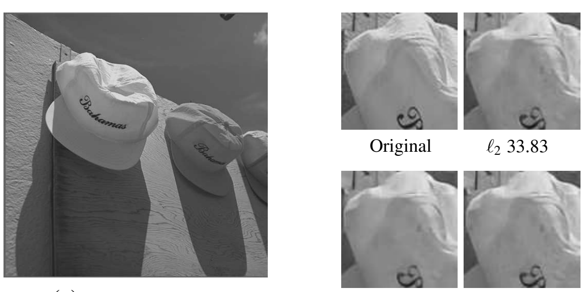


Figure 3: Performance of Compressed Sensing.



(a) Caps from LIVE1 — Original, ℓ_2 33.83, ℓ_1 33.82, Our's 34.11

Figure 4: Qualitative comparison of compressed sensing at ratio=0.25.

Table 2: Average Value of Different Image Quality Metrics. Best Result Are Shown in Bold. GMSD Is the Lower the Better, and the Others, the Higher the Better.

IQC	PSNR	SSIM	GMSD	FSIM	FSIMc
ℓ_2	28.37	0.8039	0.0817	0.9389	0.9389
ℓ_1	28.36	0.7979	0.0828	0.9388	0.9388
Our's	**28.51**	**0.8072**	**0.0810**	**0.9407**	**0.9407**

[3] L. et al., "Enhanced deep residual networks for single image super-resolution," in *IEEE Computer Vision and Pattern Recognition Workshops*, vol. 1, no. 2, 2017, p. 3.

[4] K. et al., "Accurate image super-resolution using very deep convolutional networks," in *IEEE Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[5] A. et al., "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM journal on Matrix Analysis and Applications*, vol. 29, no. 1, pp. 328–347, 2007.

[6] M. et al., "Geometric deep learning on graphs and manifolds using mixture model cnns," in *Proceedings of Computer Vision and Pattern Recognition*, vol. 1, no. 2, 2017, p. 3.

[7] S. et al., "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Foundamentals of Electronic Imaging Systems*, 1986.

[8] B. et al., "Uniqueness of the gaussian kernel for scale-space filtering," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 1, pp. 26–33, 1986.

[9] Z. et al., "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. on Imag. Proc.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[10] X. et al., "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. on Imag. Proc.*, vol. 23, no. 2, pp. 684–695, 2014.

[11] Z. et al., "Fsim: a feature similarity index for image quality assessment," *IEEE Trans. on Imag. Proc.*, vol. 20, no. 8, pp. 2378–2386, 2011.