# Improvement of Luminance Isotropy for Convolutional Neural Networks-Based Image Super-Resolution

URAZOE, Kazuya ; KUROKI, Nobutaka ; KATO, Yu ; OHTANI, Shinya ; HIROSE, Tetsuya ; NUMA, Masahiro

(Citation)
IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences,E103.A(7):955-958

(Issue Date)
2020-07-01

(Resource Type)
journal article

(Version)
Version of Record

IEICE TRANS. FUNDAMENTALS, VOL.E103–A, NO.7 JULY 2020



LETTER

# Improvement of Luminance Isotropy for Convolutional Neural Networks-Based Image Super-Resolution

**Kazuya URAZOE**[†], ***Nonmember***, **Nobutaka KUROKI**[†a)], ***Member***, **Yu KATO**[†*], **Shinya OHTANI**[†**], ***Nonmembers***, **Tetsuya HIROSE**[†***], ***and*** **Masahiro NUMA**[†], ***Members***

**SUMMARY** Convolutional neural network (CNN)-based image super-resolutions are widely used as a high-quality image-enhancement technique. However, in general, they show little to no luminance isotropy. Thus, we propose two methods, "Luminance Inversion Training (LIT)" and "Luminance Inversion Averaging (LIA)," to improve the luminance isotropy of CNN-based image super-resolutions. Experimental results of 2× image magnification show that the average peak signal-to-noise ratio (PSNR) using Luminance Inversion Averaging is about 0.15–0.20 dB higher than that for the conventional super-resolution.

***key words:*** *super-resolution, resolution enhancement, convolutional neural network, isotropy, deep learning*

## 1. Introduction

Image upsampling techniques are used in a number of applications and devices such as digital cameras, smart phones, and televisions. Recently, many image super-resolution techniques based on convolutional neural networks (CNNs) have achieved state-of-the-art results [1]–[4]. However, the CNNs do not necessarily show the same characteristics for all luminance ranges because the CNN itself is a nonlinear function and its parameters are initialized with random values. We discovered that conventional super-resolution CNN outputs different responses from positive and negative impulse signals as proved later in Sect. 3.3.

On the other hand, the luminance distributions of benchmark datasets for super-resolution [5]–[11] are different from each other, as shown in Fig. 1. The performance of the CNN is dependent on the characteristic of the training dataset. For practical use, however, the luminance distributions of training and inferred images are not always the same. These differences may cause performance degradation. Thus, the method which achieves luminance isotropy is one of the promising strategies for improving the performance of CNN-based super-resolution.

Based on our survey, we propose two methods. One is "Luminance Inversion Training." This method augments the dataset with the luminance inversion of the original dataset; therefore, it gives luminance variability and isotropy to the training dataset. The other method is "Luminance Inversion Averaging." This method gives complete luminance isotropy to the CNN-based super-resolution. Specifically, the CNN generates two magnified images from a positive image and a luminance inversed image (negative image), and then the final magnified image is obtained from the average of the positive and re-inversed negative images. These proposed methods improve luminance isotropy and are promising strategies for improving the performance of CNN-based super-resolution.

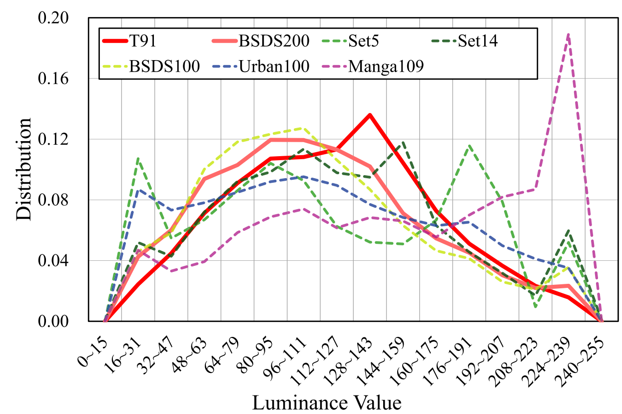


**Fig. 1** Luminance distributions of benchmark datasets for super-resolution. (The datasets of solid lines and broken lines are often used as training images and testing images, respectively.)

## 2. Proposed Methods

### 2.1 Luminance Inversion Training (LIT)

Data augmentation is one of the most widely used techniques for boosting performance with deep learning. For image super-resolution, some useful augmentations, such as random cropping, flipping, scaling, rotating, and color jittering have already been proposed [4]. In this letter, we propose a new data augmentation method for super-resolution, called Luminance Inversion Training (LIT). Let $\boldsymbol{Y}$ be the 8-bit image with luminance channel of YCrCb space. In this method, the original image $\boldsymbol{Y}$ is augmented by luminance inversion as

$$\overline{\boldsymbol{Y}} = 255 - \boldsymbol{Y}. \tag{1}$$

Then, $\overline{\boldsymbol{Y}}$ is added to the original dataset. This method gives

Manuscript received December 18, 2019.
Manuscript revised March 16, 2020.

[†]The authors are with the Graduate School of Engineering, Kobe University, Kobe-shi, 657-8501 Japan.
[*]Presently, with EIZO Corporation.
[**]Presently, with Toyota Motor Corporation.
[***]Presently, with the Graduate School of Engineering, Osaka University.
a) E-mail: kuroki@kobe-u.ac.jp

DOI: 10.1587/transfun.2019EAL2168

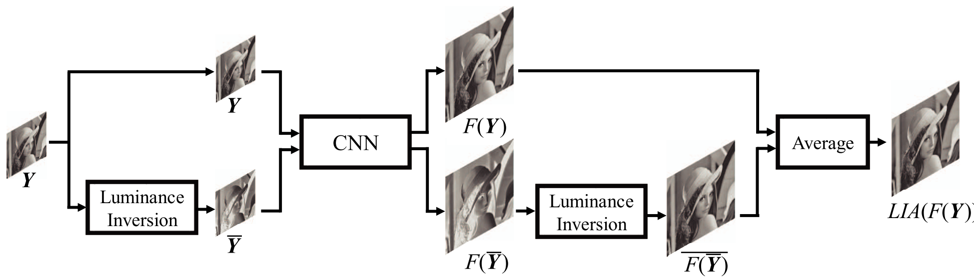


**Fig. 2** Luminance inversion averaging (LIA).

the luminance variability and isotropy to the training dataset.

## 2.2 Luminance Inversion Averaging (LIA)

We propose another method, called Luminance Inversion Averaging (LIA) for complete luminance isotropy. Figure 2 shows the flow of this method. First, we prepare the positive image $\boldsymbol{Y}$ and the negative image $\overline{\boldsymbol{Y}}$. The negative image $\overline{\boldsymbol{Y}}$ is calculated using the luminance inversion of the positive image as

$$\overline{\boldsymbol{Y}} = 255 - \boldsymbol{Y}. \tag{2}$$

Thereafter, the magnified images of the positive image $\boldsymbol{Y}$ and negative image $\overline{\boldsymbol{Y}}$ are generated as

$$F(\mathrm{Y}) = (\mathrm{Y}) \uparrow_s, \tag{3}$$
$$F(\overline{\mathrm{Y}}) = (\overline{\mathrm{Y}}) \uparrow_s, \tag{4}$$

respectively, where $\uparrow_s$ is the CNN-based super-resolution $F$ with scale factor $s$. Next, the magnified image from the negative image $F(\overline{\boldsymbol{Y}})$ is re-inversed as

$$\overline{F(\overline{\boldsymbol{Y}})} = 255 - F(\overline{\boldsymbol{Y}}). \tag{5}$$

Finally, the output image is calculated using the pixel average of $F(\boldsymbol{Y})$ and $\overline{F(\overline{\boldsymbol{Y}})}$ as

$$LIA(F(\boldsymbol{Y})) = \frac{F(\boldsymbol{Y}) + \overline{F(\overline{\boldsymbol{Y}})}}{2}. \tag{6}$$

$LIA(F(\boldsymbol{Y}))$ is always isotropy for the luminance signal like linear interpolation. Note that the $LIA(*)$ is an external processing of the $F(\boldsymbol{Y})$; therefore, LIA is applicable for any type of CNN-based super-resolution technique.

# 3. Experiment and Results

We prove the effectiveness of the proposed methods by applying them to the CNN-based super-resolutions called Multi-Channel Convolutional Neural Network (MCH) [2] and Laplacian Pyramid Super-Resolution Network (LapSRN) [3] shown in Table 1. MCH is developed for low-complexity and high-speed processing, and its hyperparameters are the same as the original study [2], excluding the fact that the number of backpropagations is 7.5 million. LapSRN is developed for the state-of-the-art image-quality performance, and its hyperparameters are the same as the original study [3] excluding the fact that there is no data augmentation. MCH is implemented with a Caffe package [12]. LapSRN is implemented with a MatConvNet package [13] and the package distributed by the original study [3]. Testing datasets are Set5 [7], Set14 [8], BSDS100 [6], Urban100 [9], and Manga109 [10], [11]. We focus only on the luminance channel in the YCrCb space and evaluate 2× image magnification. Luminance Inversion Training doubles the original training dataset; however, no other data augmentations are used for training and testing images.

**Table 1** Conditions of CNN-based super-resolutions.

| Method | Convolutional layers | Training dataset |
|---|---|---|
| MCH [2] | 3 | T91 [5] |
| LapSRN [3] | 14 | T91 [5] + BSDS200 [6] |

**Table 2** PSNR and SSIM of 2× magnified images for all testing datasets.

| Method | PSNR [dB] / SSIM | |
|---|---|---|
| | MCH | LapSRN |
| Baseline | 32.25 / 0.9176 | 32.48 / 0.9208 |
| +Luminance Inversion Training (LIT) | 32.36 / 0.9184 | 32.53 / 0.9208 |
| +Luminance Inversion Averaging (LIA) | **32.45 / 0.9193** | **32.63 / 0.9216** |

Red bold letters indicate the best scores in each CNN-based super-resolution.

## 3.1 Objective Evaluation with PSNR and SSIM

Table 2 shows the average PSNR (peak signal-to-noise ratio) and SSIM (structural similarity) for all testing datasets. Each proposed method improved PSNR in both MCH and LapSRN. Especially, LIA is effective and improved the average PSNR by 0.15–0.20 dB.

Tables 3 and 4 show the average PSNR for each testing dataset. First, we focus on LIT. Although greatly improving the average PSNRs in the case of Manga109, this method only slightly improved or degraded them in the other datasets. Thus, the LIT does not necessarily improve the image quality. Next, we focus on LIA. This method improved the PSNRs in all testing datasets. These results mean that the internal improvement of the CNN is difficult; however, the LIA of the external processing can easily improve the super-resolution performance.

Next, we discuss the relationship between the luminance distributions of datasets and the improvements of im-

**Table 3** PSNRs and SSIMs of 2× magnified images with MCH.

| Method | PSNR[dB] / SSIM | | | | |
|---|---|---|---|---|---|
| | Set5 | Set14 | BSDS100 | Urban100 | Manga109 |
| Baseline | 36.73 / 0.9539 | 32.47 / 0.9057 | 31.21 / 0.8869 | 29.41 / 0.8951 | 35.56 / 0.9663 |
| +Luminance Inversion Training (LIT) | 36.80 / 0.9542 | 32.52 / 0.9062 | 31.25 / 0.8874 | 29.46 / 0.8960 | 35.83 / 0.9672 |
| +Luminance Inversion Averaging (LIA) | **36.88** / **0.9549** | **32.59** / **0.9071** | **31.29** / **0.8882** | **29.52** / **0.8971** | **35.98** / **0.9683** |

Red bold letters indicate the best score in each testing dataset.

**Table 4** PSNRs and SSIMs of 2× magnified images with LapSRN.

| Method | PSNR[dB] / SSIM | | | | |
|---|---|---|---|---|---|
| | Set5 | Set14 | BSDS100 | Urban100 | Manga109 |
| Baseline | 36.86 / 0.9558 | 32.56 / 0.9082 | 31.44 / 0.8899 | 29.60 / 0.8989 | 35.86 / 0.9693 |
| +Luminance Inversion Training (LIT) | 36.85 / 0.9554 | 32.55 / 0.9081 | 31.43 / 0.8896 | 29.62 / 0.8988 | 36.01 / 0.9695 |
| +Luminance Inversion Averaging (LIA) | **36.97** / **0.9562** | **32.62** / **0.9088** | **31.47** / **0.8903** | **29.68** / **0.8998** | **36.20** / **0.9703** |

Red bold letters indicate the best score in each testing dataset.

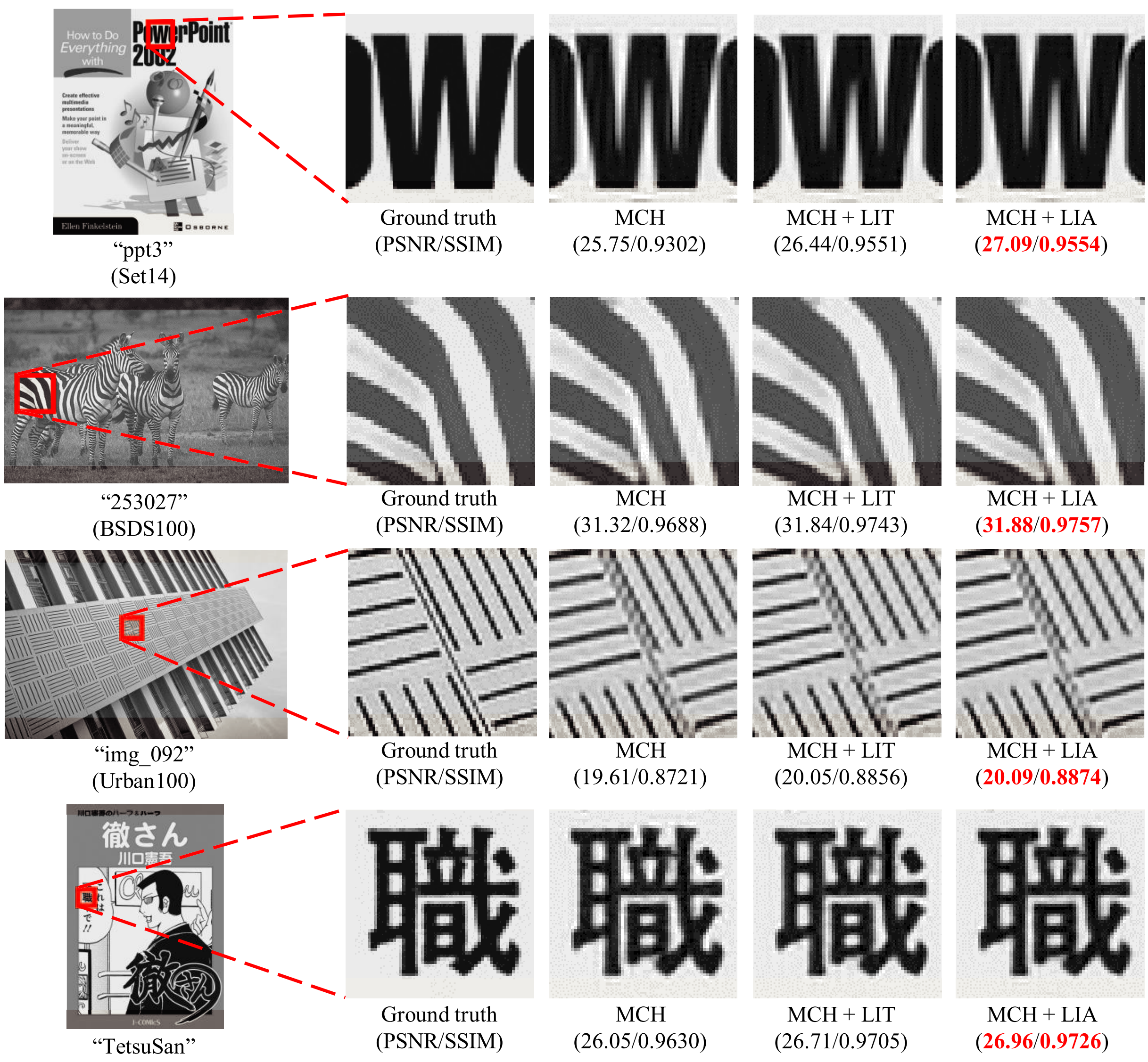


**Fig. 3** Visual comparison on 2× magnified images with MCHs (red bold letters indicate the best score in each testing image).

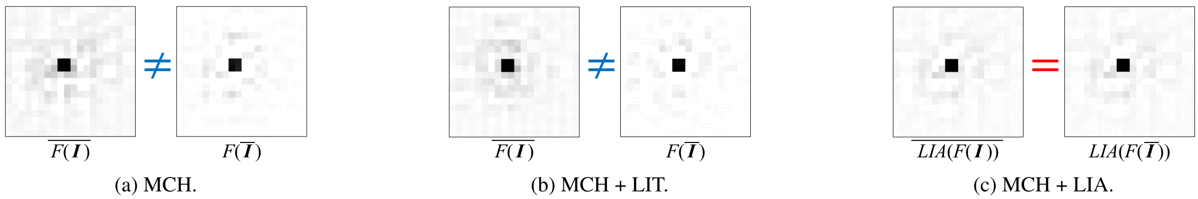


**Fig. 4** Isotropic of Impulse response with MCHs ($\boldsymbol{I}$ and $\overline{\boldsymbol{I}}$ indicate positive and negative impulse signals, respectively).

age quality. Figure 1 shows that the luminance distribution of BSDS100 is similar to other distributions used for training CNNs; however, luminance distribution of Manga109 is different from them. In such a case, LIA is especially effective. Table 3 shows that LIA improved the PSNR by only 0.08 dB in BSDS100 but by 0.42 dB in Manga109. Table 4 also shows the same tendency. From these results, we found that LIA is greatly effective for correcting the mismatch between the luminance distributions of training and inferred images.

### 3.2 Subjective Evaluation of Magnified Images

Figure 3 shows the magnified images with MCHs. We can see that LIA reduces the overshoots around the edge and outlines these edges clearly, compared with other methods. This is because LIA calculates the average of the positive and negative output signals, and then their noises are canceled.

### 3.3 Isotropic of Impulse Response

Figure 4 shows the inversed positive and negative impulse responses. We can see that the conventional MCH has no luminance isotropy. In LIT, although the training dataset has luminance isotropy, the impulse responses have no luminance isotropy. This is because the CNN itself is a nonlinear function and its parameters are initialized with random values. On the other hand, the inversed positive and negative impulse responses are exactly equal in LIA. Thus, LIA has the perfect isotropy for the luminance signal. This indicates that the super-resolution gives the same performance in both bright and dark areas. The authors believe that this is one of the important characteristics that super-resolutions should have.

## 4. Conclusion

In this letter, “Luminance Inversion Training (LIT)” and “Luminance Inversion Averaging (LIA)” were proposed to improve the luminance isotropy of CNN-based image super-resolutions. Experimental results have shown that the average PSNR for super-resolution using Luminance Inversion Averaging is about 0.15–0.20 dB higher than that for the conventional super-resolution. We have ensured that the proposed method can improve the performances of CNN-based super-resolutions even if the luminance distributions of training and inferred images are different from each other. Additionally, the proposed methods are applicable for any CNN-based super-resolution. However, Luminance Inversion Averaging does not improve the internal structure of CNN and requires twice the processing time in principle. The implementation for internal isotropy of CNN will be the focus of our future work.

### References

[1] C. Dong, C.C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol.38, no.2, pp.295–307, 2016.
[2] S. Ohtani, Y. Kato, N. Kuroki, T. Hirose, and M. Numa, “Multi-channel convolutional neural networks for image super-resolution,” IEICE Trans. Fundamentals, vol.E100-A, no.2, pp.572–580, Feb. 2017.
[3] W.S. Lai, J.B. Huang, N. Ahuja, and M.H. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[4] Z. Wang, J. Chen, and S.C.H. Hoi, “Deep learning for image super-resolution: A survey,” arXiv: abs/1902.06068, 2019.
[5] J. Yang, J. Wright, T.S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol.19, no.11, pp.2861–2873, 2010.
[6] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” Proc. Eighth IEEE International Conference on Computer Vision (ICCV), 2001.
[7] M. Bevilacqua, A. Roumy, C. Guillemot, and M. line Alberi Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” Proc. British Machine Vision Conference, 2012.
[8] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” Proc. 7th International Conference on Curves and Surfaces, pp.711–730, 2012.
[9] J.B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[10] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, “Sketch-based manga retrieval using manga109 dataset,” Multimed. Tools Appl., vol.76, no.20, pp.21811–21838, 2017.
[11] T. Ogawa, A. Otsubo, R. Narita, Y. Matsui, T. Yamasaki, and K. Aizawa, “Object detection for comics using manga109 annotations,” arXiv: abs/1803.08670, 2018.
[12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” Proc. 22nd ACM International Conference on Multimedia, pp.675–678, 2014.
[13] A. Vedaldi and K. Lenc, “Matconvnet: Convolutional neural networks for matlab,” Proc. 23rd ACM International Conference on Multimedia, pp.689–692, 2015.