MRI Super-Resolution using Laplacian Pyramid Convolutional Neural Networks with Isotropic Undecimated Wavelet Loss

Sriprabha Ramanarayanan, Balamurali Murugesan, Ananth Kalyanasundaram, Surya Prabhakaran, Keerthi Ram, Shantanu Patil and Mohanasankar Sivaprakasam

***Abstract*— High spatial resolution of Magnetic Resonance images (MRI) provide rich structural details to facilitate accurate diagnosis and quantitative image analysis. However the long acquisition time of MRI leads to patient discomfort and possible motion artifacts in the reconstructed image. Single Image Super-Resolution (SISR) using Convolutional Neural networks (CNN) is an emerging trend in biomedical imaging especially Magnetic Resonance (MR) image analysis for image post processing. An efficient choice of SISR architecture is required to achieve better quality reconstruction. In addition, a robust choice of loss function together with the domain in which these loss functions operate play an important role in enhancing the fine structural details as well as removing the blurring effects to form a high resolution image. In this work, we propose a novel combined loss function consisting of an L1 Charbonnier loss function in the image domain and a wavelet domain loss function called the Isotropic Undecimated Wavelet loss (IUW loss) to train the existing Laplacian Pyramid Super-Resolution CNN. The proposed loss function was evaluated on three MRI datasets - privately collected Knee MRI dataset and the publicly available Kirby21 brain and iSeg infant brain datasets and on benchmark SISR datasets for natural images. Experimental analysis shows promising results with better recovery of structure and improvements in qualitative metrics.**

***Index Terms*— Single Image Super-Resolution, Convolutional Neural networks, Knee joint, MR Images**

I. Introduction

Magnetic resonance imaging (MRI) has the benefit of non-invasive acquisition and provides detailed structural information, benefiting clinical diagnosis. However the high spatial resolution of MRI comes with the expense of long scan times, small spatial coverage and low signal-to-noise ratio [1]. In addition, MR image quality is dependent on the strength of the magnetic field as well as the scanning parameters deployed for the anatomy under study. Certain structures require protocols with extended duration of time to allow for precise scanning. However, this may also result in motion artifacts as the subject may not be able to stay immobile for that period. This may ultimately compromise image quality leading to difficulty in diagnosis for the radiologists.

Image quality can be enhanced by several post processing techniques which are pre-installed with the MR Imaging software. With the advent of Deep Learning, this process can be further enhanced, allowing the radiologist a higher probability of arriving at a conclusive diagnosis. The concept of Single Image Super-Resolution (SISR) can be applied effectively for this purpose. Several deep learning architectures for SISR have been developed recently [2] for natural images. The earliest SISR deep learning architecture was the Super-Resolution Convolutional Neural Network (SRCNN) [3] by Dong et al. which has surpassed the bicubic baseline and outperformed the sparse coding based methods. Ledig et al [4] proposed the Super-Resolution Generative Adversarial Network (SRGAN) which uses a perceptual loss function consisting of an adversarial loss and a content loss to yield better perceptual image quality. The Laplacian Pyramid Super Resolution Network (LapSRN) [5] introduced by Lai et al., is based on Laplacian Pyramid which progressively reconstructs the residuals of higher resolution images at each level of the pyramid. The LapSRN network uses the L1 Charbonnier loss which shows a significant improvement over the Mean squared error (MSE) Loss.

Some of the recent works on CNN based MRI super-resolution include the 3D SRCNN [6] for knee images, GAN for brain images [7] and CNN with wide residual network with fixed skip connection [8]. All the above methods have extensive feature extraction stages but the choice of loss function have made the overall output suffer from an undesirable blur. SISR methods that retain a pre-existing network as the base network and modify the loss functions to emphasize on improving edges and textures exist in the literature. Pandey et al [9] proposed the loss function based on mean square canny edge to facilitate sharper reconstruction and applied it in SRCNN and the Efficient Sub-pixel CNN (ESPCN). George et al [10] and Yong Guo et al [11] used an image edge based loss component to preserve both low-frequency content and high-frequency structure of images.

In our work, we emphasize that, in addition to minimizing in spatial domain, the loss in the sub-band domain of the predicted HR image with respect to the ground truth must also be minimized. Wavelet transform is a good operator for signal analysis in the sub-band domain. The wavelet transform coefficients ensure a more precise and sparse description of local features and separation of signal characteristics. These characteristics include the coarse and fine details of the image which are of main interest to us. We have formulated our loss function based on wavelet transform which can be used to extract high-frequency details and combined with the pre-existing loss function in the spatial domain. Huang et al. proposed WaveletSRNet [12], a super-

Sriprabha Ramanarayanan, Balamurali Murugesan, and Mohanasankar Sivaprakasam are with Healthcare Technology Innovation Centre and Indian Institute of Technology Madras

Keerthi Ram is with Healthcare Techonology Innovation Centre

Ananth Kalyanasundaram, Surya Prabhakaran and Shantanu Patil are with SRM Institute of Science and Technology

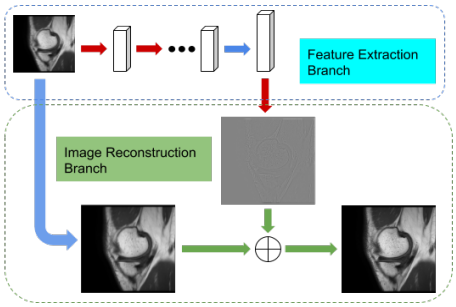

Fig. 1. Red arrows indicate convolutional layers. Blue arrows indicate transposed convolutions (upsampling), and green arrows denote element-wise addition operators.

resolution network for face images using a decimated wavelet transform based loss function. In our work, we use the Isotropic Undecimated Wavelet (IUW) transform [13] in our loss function for MRI super-resolution. The benefits of IUW transform are: 1) By eliminating the decimation step, the undecimated wavelet transform overcomes the lack of translation-invariance of the discrete wavelet transform (DWT). In pattern recognition, it is important to construct signal representations that are translation invariant. When a pattern is translated, its representation should be translated not modified. Translation-invariance is also a key feature of convolution layers of CNNs. 2) It can better enhance the structures of isotropic regions which are common in images. 3) The filters of IUW can help in removal of ringing artifacts, a common problem with MRI images. Combining these advantages, we summarize our contributions as follows:

- We propose a novel combined loss function consisting of L1 Charbonnier loss function in image domain and a wavelet domain loss function called the Isotropic Undecimated Wavelet loss for a pre-existing super-resolution CNN, the LapSRN, for better performance.
- We have made a collection of a large entity of real life Knee MRI data and applied image pre-processing to curate the data.
- We have made an extensive comparative study of the bicubic baseline, LapSRN with L1 Charbonnier loss and LapSRN with L1 Charbonnier + IUW loss. Results show quantitative improvements with better recovery of structures for knee, brain and benchmark vision datasets.

II. METHODOLOGY

A. LapSRN Architecture

The 2x super-resolution network of the LapSRN [5] is our base network. The low-resolution (LR) image is taken as input using which the model predicts residual images at 2x scale factor (Fig. 1). The model has two branches, namely feature extraction and image reconstruction.

Feature extraction : The feature extraction branch consists of n convolutional layers and one transposed convolutional layer to upsample the extracted features by a scale of 2. The output of each transposed convolutional layer is connected to a convolutional layer for reconstructing a residual image.

Image reconstruction: The input image is upsampled by a scale of 2 using a transposed convolutional layer. The layer is initialized with a bilinear kernel. A high-resolution output image is then obtained by combining the upsampled image and the predicted residual image from the feature extraction branch.

B. Loss function

1) L1 Charbonnier Loss: The L1 Charbonnier Loss used in LapSRN has shown to perform better than the MSE which fails to capture the underlying multi-modal distributions of HR patches resulting in over-smoothed reconstructed HR images which are inconsistent to human visual perception on natural images.

$$\begin{aligned} L(\hat{y}, y; \theta) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{L}\rho(\hat{y_l}^{(i)} - y_l^{(i)}) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{L}\rho((\hat{y_l}^{(i)} - x_l^{(i)}) - r_l^{(i)}) \end{aligned} \tag{1}$$

where, $\rho(x) = \sqrt{x^2 + \varepsilon^2}$ is the Charbonnier penalty function, N is the number of training samples in each batch, and L is the number of levels in our pyramid. We assign a value of 1e-3 to ε.

2) Isotropic Undecimated Wavelet Loss: We propose the Isotropic Undecimated Wavelet loss which is applied to the reconstructed output of LapSRN. The IUW transform uses three directional analysis (horizontal, vertical and diagonal) filters to design the filter banks. The IUWT filter at level 1 is defined as:-

$$\begin{aligned} f_1 &= g_{i,j} * g_{i,j} \\ f_2 &= g_{i,j} * h_{i,j} \\ f_3 &= h_{i,j} * g_{i,j} \\ f_4 &= h_{i,j} * h_{i,j} \end{aligned} \tag{2}$$

Where h=[1,4,6,4,1]/16 and g=[-1,-4,10,-4,-1]/16 for the first level. Each of the above equation generates a 5×5 filter. We stack them to create a 4 channel filter of dimension 5×5×4. f_1, f_2, f_3, f_4 represent the channels of the filter. Let this filter be f.

IUW transform is defined as the convolutional operation between the image and the filter (f) mentioned above. This filter can seamlessly fit into a convolutional layer since it uses the convolution operation to find the IUW transform.

$$\begin{aligned} \hat{Y_r}^{(l)} &= \hat{Y}^{(l)} * f \\ Y_r^{(l)} &= Y^{(l)} * f \end{aligned} \tag{3}$$

where $\hat{Y}^{(l)}$ is the predicted image at level l of the pyramid and $Y^{(l)}$ is the ground truth. $\hat{Y}_r^{(l)}$ and Y_r represent the IUWT sub-bands of the predicted and the ground truth images

respectively. We define the IUW loss as:

$$
\begin{aligned}
Loss_{LL}^{(l)} &= \sum_{i=1}^{5}\sum_{j=1}^{5}\sqrt{|\hat{Y}_{r(i,j)}^{2}[0] - Y_{r(i,j)}^{2}[0]|} \\
Loss_{LH}^{(l)} &= \sum_{i=1}^{5}\sum_{j=1}^{5}\sqrt{|\hat{Y}_{r(i,j)}^{2}[1] - Y_{r(i,j)}^{2}[1]|} \\
Loss_{HL}^{(l)} &= \sum_{i=1}^{5}\sum_{j=1}^{5}\sqrt{|\hat{Y}_{r(i,j)}^{2}[2] - Y_{r(i,j)}^{2}[2]|} \\
Loss_{HH}^{(l)} &= \sum_{i=1}^{5}\sum_{j=1}^{5}\sqrt{|\hat{Y}_{r(i,j)}^{2}[3] - Y_{r(i,j)}^{2}[3]|}
\end{aligned}
\tag{4}
$$

where t= L,H and k=L,H and N is the number of training examples. LL, LH, HL, HH are defined as the sub bands obtained from the channels of Y_r and $\hat{Y}_r$. This loss is then combined with the L1 Charbonnier loss in the image domain by the following equation:-

$$Loss = L(\hat{y}, y; \theta) + \alpha * Loss_{IUWT} \tag{5}$$

where $L(\hat{y}, y; \theta)$ is the L1 Charbonnier Loss and α is the weight variable used to balance both the losses. The network yields best results for α set to 0.05 and the number of levels L is set to 1.

III. Dataset description and Evaluation metrics

A. MRI data

We have used three MRI datasets - MRI Knee dataset collected from hospital and two publicly available brain datasets and benchmark SISR datasets of natural images.

Anonymized MRI knee scans of 10 subjects with intra-articular pathologies taken from a department library are used. Ethical waiver for this data is appropriately procured from the Institutional review board as per the requirements of the Helsinki declaration (1975). In this dataset, 8000 DICOM images are used for training, 800 for validation and 4000 for testing and comparing the results.

The iSeg Infant MR image database is originally used to evaluate the segmentation algorithms for the brain structures of 6-month infants [14]. There are 23 T1-weighted volumes of size 144×192×256, of which 1600 images of 16 volumes are used for training and 700 images of 7 volumes for testing.

Kirby21 dataset [15] with human brain data consists of 5460 slices of size 256×256 taken from 42 T1-weighted MPRAGE volumes out of which, 3770 slices from 29 volumes are used for training and 1690 slices from 13 volumes for validation.

Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [16] metrics are used to evaluate the super-resolved image quality.

B. Dataset Creation

The image patches of size 256×256 are created on each slice of the dicom files using a window. These patches are used as the ground truth and resized to 128×128 for the creation of input images low resolution LR images to train the networks. Linear Normalization is then applied to both the input and the ground truth images.

TABLE I

Comparative study on benchmark SR vision datasets

Method \Dataset	Set14	BSDS100	Urban100	Manga109
Bicubic	30.34/ 0.8705	29.55/ 0.8471	26.87/ 0.8407	30.78/ 0.9313
LapSRN L1	33.21/ 0.8801	31.84/ 0.8698	30.84/ 0.8723	37.51/ 0.9473
LapSRN L1 + IUWL	**33.35/ 0.8813**	**31.92/ 0.8708**	**31.08/ 0.8751**	**37.65/ 0.9478**

TABLE II

Comparative study for MRI datasets

Method \Dataset	Knee	Kirby21	iSeg
Bicubic	41.38 / 0.9831	31.65 / 0.9554	31.60 / 0.9718
LapSRN L1	44.94 / 0.9887	34.22 / 0.9745	32.85 / 0.9785
LapSRN L1 + IUWL	**45.10 / 0.9887**	**34.24 / 0.9747**	**33.13 / 0.9797**
SRCNN L1	44.63 / 0.9882	33.53 / 0.9706	32.93 / 0.9789
SRCNN L1 + IUWL	**44.64 / 0.9883**	**33.55 / 0.9708**	**33.21 / 0.9799**

IV. Results and Discussion

Our initial experiments are conducted experiments on the standard DIV2K dataset with 800 RGB color images as the training set. For testing, the benchmark super-resolution datasets - Set14, BSDS100, Urban100 and Manga109 (Table I) are used. Results show consistent improvements over the bicubic and the LapSRN L1 baselines both visually and quantitatively.

In the MRI data experiments, we have compared LapSRN with L1 Charbonnier loss and the combination of L1 Charbonnier and IUW loss to understand the contribution of IUW loss in the loss combination. In Table II, we have presented the PSNR and SSIM metrics. From the table, it can be seen that LapSRN - L1 + IUW loss performs significantly better than the baseline bicubic. Furthermore, comparing between LapSRN - L1 and LapSRN - L1 + IUW, it can be observed that for all the datasets, the improvement in PSNR is appreciable. But, when considering SSIM, the following things can be observed: 1) LapSRN - L1 + IUW provides equivalent results to LapSRN - L1 for Knee 2) LapSRN - L1 + IUW provides minor improvement compared to LapSRN - L1 for Kirby and 3) LapSRN - L1 + IUW provides appreciable improvement compared to LapSRN - L1 for ISeg. This behaviour can be attributed to the structures and resolution of the considered dataset. Knee and Kirby21 are high resolution datasets with better structures while the structures in ISeg dataset are weak and difficult to recover. Significant improvement in SSIM for iSeg dataset shows that our proposed loss IUW does help in recovering structures. Similar analysis is done on the classical SRCNN to show the consistency of the proposed approach over baseline methods.

Qualitative comparison of LapSRN - L1 and LapSRN - L1 + IUW for natural images (Fig. 2 and 3) for MRI images (Fig 4, 5 and 6) are shown. The figures show that LapSRN - L1 + IUWL provides better recovery of structures as compared to LapSRN - L1. For natural images, the figures show reduction in the aliasing artifacts of fine structures for the challenging

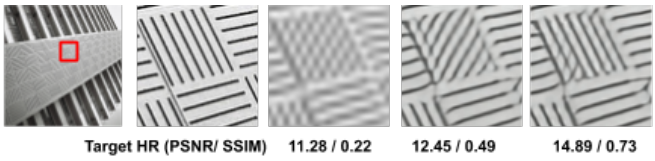

Fig. 2. Visual results for Urban100 image92, from left: Bicubic interpolation, LapSRN L1 and and ours (LapSRN L1 + IUWL)

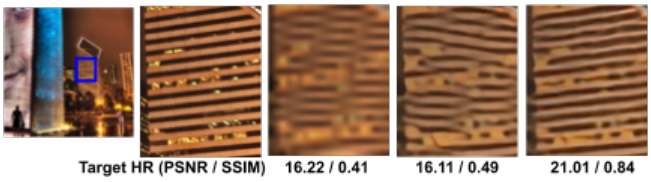

Fig. 3. Visual results for Urban100 image76, from left: Bicubic interpolation, LapSRN L1 and and ours

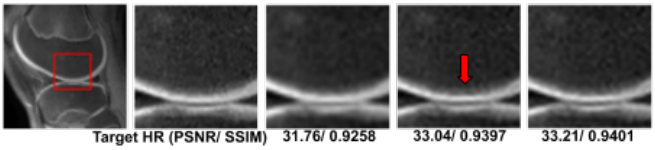

Fig. 4. Visual Comparison of bicubic interpolation, LapSRN and ours for Knee dataset. Red arrow mark indicates improper recovery of details

Urban100 images. Further, the PSNR values reported for the particular set of slices shows that the LapSRN - L1 + IUWL works effectively for ISeg, Kirby and Knee.

V. Conclusion

We propose the novel Isotropic Undecimated Wavelet loss to improve the quality of super-resolved images by seamlessly integrating the IUWT in the convolutional layer of the LapSRN architecture. We have evaluated our proposed method on Knee and brain datasets. We observe that the recovery of structures were much better than the bicubic baseline and relatively better than LapSRN. As a future work, we would like to explore the performance of the proposed loss function in other architectures.

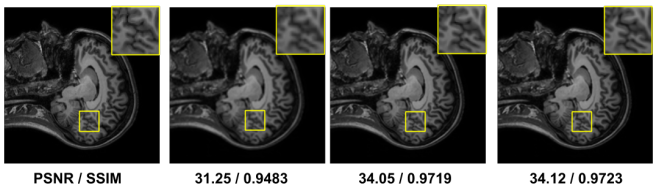

Fig. 5. Visual Comparison of bicubic interpolation, LapSRN and ours for Kirby brain dataset.

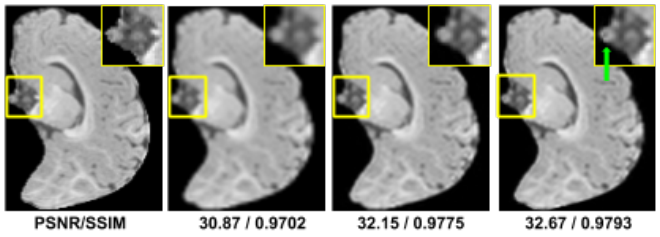

Fig. 6. Visual Comparison of bicubic interpolation, LapSRN and ours for ISeg Infant brain dataset. Green arrow mark indicates enhanced details

References

[1] E. Plenge, D. H. J. Poot, Bernsen, and et al., "Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?" *Magnetic Resonance in Medicine*, vol. 68, no. 6, pp. 1983–1993, 2012.

[2] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for Image Super-resolution: A Survey," *CoRR*, vol. abs/1902.06068, 2019.

[3] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic Single Image Super-resolution Using a Generative Adversarial Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.

[5] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 624–632.

[6] A. S. Chaudhari, Z. Fang, F. Kogan, J. Wood, K. J. Stevens, E. K. Gibbons, J. H. Lee, G. E. Gold, and B. A. Hargreaves, "Super-Resolution Musculoskeletal MRI using Deep Learning," *Magnetic Resonance in Medicine*, vol. 80, no. 5, pp. 2139–2154, 2018.

[7] Y. Chen, F. Shi, A. G. Christodoulou, Z. Zhou, Y. Xie, and D. Li, "Efficient and Accurate MRI Super-Resolution using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network," in *MICCAI*, 2018.

[8] J. Shi, Z. Li, S. Ying, C. Wang, Q. Liu, Q. Zhang, and P. Yan, "MR Image Super-Resolution Via Wide Residual Networks With Fixed Skip Connection," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1129–1140, May 2019.

[9] R. K. Pandey, N. Saha, S. Karmakar, and A. G. Ramakrishnan, "MSCE: An Edge Preserving Robust Loss Function for Improving Super-Resolution Algorithms," in *ICONIP*, 2018.

[10] G. Seif and D. Androutsos, "Edge-Based Loss Function for Single Image Super-Resolution," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 1468–1472.

[11] Y. Guo, Q. Chen, J. Chen, J. Huang, Y. Xu, J. Cao, P. Zhao, and M. Tan, "Dual Reconstruction Nets for Image Super-Resolution with Gradient Sensitive Loss," *CoRR*, vol. abs/1809.07099, 2018. [Online]. Available: http://arxiv.org/abs/1809.07099

[12] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-SRNet: A Wavelet-Based CNN for Multi-Scale Face Super-Resolution," in *IEEE International Conference on Computer Vision*, 2017, pp. 1689–1697.

[13] J.-L. Starck, J. Fadili, and F. Murtagh, "The Undecimated Wavelet Decomposition and its Reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 297–309, 2007.

[14] L. Wang, D. Nie, G. Li, Puybareau, J. Dolz, Q. Zhang, F. Wang, and et al., "Benchmark on Automatic Six-Month-Old Infant Brain Segmentation Algorithms: The iSeg-2017 challenge," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2219–2230, Sep. 2019.

[15] B. A. Landman, A. J. Huang, A. Gifford, D. S. Vikram, I. A. L. Lim, J. A. Farrell, and et al., "Multi-Parametric Neuroimaging Reproducibility: A 3-T Resource Study," *NeuroImage*, vol. 54, no. 4, pp. 2854 – 2866, 2011.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.