Super-Resolution and Inpainting with Degraded and Upgraded Generative Adversarial Networks

Yawen Huang[1,2], **Feng Zheng**[3,4*], **Danyang Wang**[1,2], **Junyu Jiang**[1,2], **Xiaoqian Wang**[5], **Ling Shao**[6]

[1]Malong Technologies
[2]Shenzhen Malong Artificial Intelligence Research Center
[3]Depatment of Computer Science and Technology, Southern University of Science and Technology
[4]Research Institute of Trustworthy Autonomous Systems
[5]Purdue University
[6]Inception Institute of Artificial Intelligence
{yawhuang, danwang, junyu}@malong.com, zhengf@sustech.edu.cn, joywang@purdue.edu, ling.shao@ieee.org

Abstract

Image super-resolution (SR) and image inpainting are two topical problems in medical image processing. Existing methods for solving the problems are either tailored to recovering a high-resolution version of the low-resolution image or focus on filling missing values, thus inevitably giving rise to poor performance when the acquisitions suffer from multiple degradations. In this paper, we explore the possibility of super-resolving and inpainting images to handle multiple degradations and therefore improve their usability. We construct a unified and scalable framework to overcome the drawbacks of propagated errors caused by independent learning. We additionally provide improvements over previously proposed super-resolution approaches by modeling image degradation directly from data observations rather than bicubic downsampling. To this end, we propose HLH-GAN, which includes a high-to-low (H-L) GAN together with a low-to-high (L-H) GAN in a cyclic pipeline for solving the medical image degradation problem. Our comparative evaluation demonstrates that the effectiveness of the proposed method on different brain MRI datasets. In addition, our method outperforms many existing super-resolution and inpainting approaches.

1 Introduction

High-Resolution (HR) Magnetic Resonance Imaging (MRI), such as 7-Tesla MRI, has shown benefits in brain imaging due to its high signal-noise-to-ratio, resolution and sensitivity to capturing disease patterns. However, HR MRI is limited to long acquisition times, high costs and frequent unavailability. In addition, the variations in anatomical observations, triggered by a patient's motion, artifacts and corrupted or incomplete scans, deteriorate clinical diagnosis and other post-processing steps. Furthermore, the scarcity of HR imaging limits the ability to model accurate healthy and disease patterns (*e.g.* to create of anatomical atlases of health status from clinical images presenting diverse conditions), which call for advanced approaches to handle the aforementioned problems.

Existing approaches [Huang *et al.*, 2017a] seek to solve the problems either by super-resolving an input low-resolution (LR) image or inpainting missing values from the surrounding pixels. Image Super-Resolution (SR) aims at reconstructing an equivalent HR image from its LR counterpart, which is an ill-posed problem caused by a multiplicity of solutions. Image inpainting (also known as image completion) aims at filling missing values of an image, which remains a challenge due to its inherent ambiguity and the complexity of content representation. Despite many remarkable achievements shown in these respective tasks, the existing methods lack versatility in handling multiple degradations, while the practical acquisitions may suffer from several serious damages, *e.g.*, motion blur, compression artifacts, scanner noise, intensity inhomogeneity, region lost, and degradation stemming from the physical imaging (*e.g.* scattering in CT, Eddy current distortions in fMRI, etc.). In addition, separating the processes of SR and image inpainting requires two learning steps, with the consequence being that the reconstructed error from the first stage will be propagated to the second one, thus leading to an overall larger error.

Early works of SR and image inpainting mostly attempted to interpolate the lost information, relying on an image prior [Huang *et al.*, 2015], *e.g.*, nearest-neighbor interpolator [Parker *et al.*, 1983], non-local self-similarity [Manjón *et al.*, 2010] or sparse representation [Yang *et al.*, 2010]. By exploiting powerful image priors, Zeyde *et al.* [Zeyde *et al.*, 2010] introduced a dictionary learning-based method by incorporating a sparse-land local model for the single image scale-up problem. To avoid using external data but leveraging internal feature redundancy for SR purposes, a self-

*Corresponding author

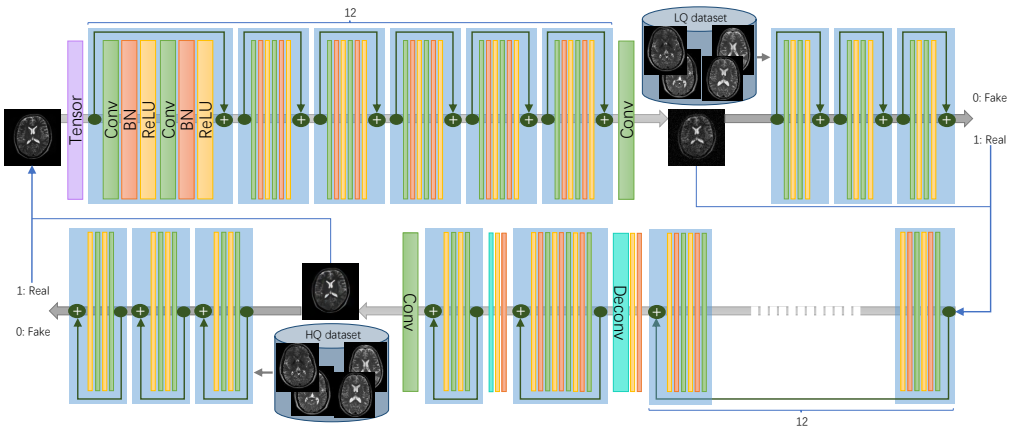

Figure 1: Overall architecture of our HLH-GAN.

similarity driven SR algorithm [Huang *et al.*, 2015] was proposed. Rather than learning the relationship between LR and HR images, internal statistic [Huang *et al.*, 2015] was employed to perform a decomposition of the geometric patch transformation. Schulter *et al.* [Schulter *et al.*, 2015] presented a fast and accurate image upscaling approach by constructing a locally linear multivariate regression using random forests. Huang *et al.* [Huang *et al.*, 2017b] proposed a hetero-domain image alignment approach by learning from both paired and unpaired data and achieved remarkable improvements in image SR. Although these methods do not require large-scale training sets to learn an activate expression, such simple and efficient techniques usually result in blurry reconstructions with very limited performance.

More recently, the field has witnessed a variety of improvements, with deep learning [Goodfellow *et al.*, 2014] being employed to gain more ideal performance. Initial efforts involved training convolutional neural networks (CNNs) to predict texture details or inpaint small regions by leveraging learned feature maps. To make the solution space closer to its true manifold, various losses, such as the perceptual loss, have been proposed over feature maps to replace the pixel-wise mean square error loss [Ulyanov *et al.*, 2016]. More recently, GANs were introduced by inferring more reliable structures and textures to encourage the nature fidelity between generated and real images [Ledig *et al.*, 2017]. Specifically, SRCNN [Dong *et al.*, 2015] applied a convolutional network architecture to learn an end-to-end mapping relationship between LR and HR images for achieving fast and superior SR performance. Kim *et al.* [Kim *et al.*, 2016a] introduced a very deep convolutional network (*i.e.* increasing the depth of a CNN to 20 layers) by employing residual learning and gradient clipping to solve exploding gradients, while obtaining more contextual information from LR images. In [Johnson *et al.*, 2016], a perceptual-driven approach was proposed to enhance the visual quality of SR results. Based on the idea of being closer to reality, a gradient profile prior-guided detail representation [Tai *et al.*, 2010], multi-scale redundant dictionary [Zhang *et al.*, 2012], Markov random field regularized deep CNN [Li and Wand, 2016] were investigated to reconstruct realistic texture details. SRGAN [Ledig *et al.*, 2017] was then developed to generate images with photo-realistic textures, which gaining huge improvements in perceptual quality.

In parallel, image inpainting is treated as another research branch which focuses on filling missing regions in a damaged image with semantically reasonable content. The early developed methods, *e.g.*, Region Filling [Criminisi *et al.*, 2004] applied exemplar-based texture synthesis to propagate texture and structure information, simultaneously, for filling in the gaps. Borrowing data from surrounding pixels is a simple and flexible way to recover relative missing values; however, it lacks a deeper understanding of images and usually leads to context-unreasonable results. Recently, benefiting from the rapid development of deep learning, visual performance and contextual coherence of image inpainting have been further improved. Context Encoders [Pathak *et al.*, 2016] leverage image analogy to give semantically hole-filled results. Yu *et al.* [Yu and Koltun, 2015] constructed a detailed convolution to model multiscale contextual information and increase the size of receptive fields. Iizuka *et al.* [Iizuka *et al.*, 2017] utilized a fully-convolutional neural network with their proposed global and local context discriminators to approximate visually plausible content in the holes. Inspired by the U-Net model, Shift-Net [Yan *et al.*, 2018] was proposed to optimize texture details and make sharper structures by introducing a guidance loss.

Throughout the literature, these advanced methods intuitively focus on the upgrading procedure (*i.e.*, low-to-high or incomplete-to-complete processing) by artificial downsampling to treat the low-quality data. In practice, image degradation can be very complex and even feature multiple issues associated to degradation factors. This raises the challenge of handling unknown nuisance factors. Inevitably, lacking versatility to handle naturally degraded images leads to poor performance during testing. To alleviate this problem and achieve high-quality reconstruction with visual realism, we propose to learn the degradation procedure from real high-/low-quality acquisitions first, and then super-resolve and inpaint degraded data in practical scenarios. Our method is shown in Fig. 1.

Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to propose a unified framework (termed as HLH-GAN) to perform super-resolution and image inpainting simultaneously on medical images.
- A degradation GAN (H-L GAN) is presented for modeling the practical low-quality images using our H-L loss functions rather than bicubic downsampling.
- An upgrading GAN (L-H GAN) is introduced with three branches: adversarial loss, content loss, and texture loss. L-H GAN is calculated on feature maps of the VGG network for content-coherent and texture-realistic reconstruction.
- We confirm using several quantitative metrics on images from three datasets that HLH-GAN is the state-of-the-art for the recovery of high-resolution, texture-clear and content-complete images.

2 Method

The proposed work aims at handling multiple degradations such as LR, noisy, blurry, corrupted and region-lost images. We collectively refer to all these degraded images as low-quality (LQ) data, in contrast to all superior images as High-Quality (HQ) data, and image super-resolution and inpainting tasks as SRI.

2.1 Preliminaries

Typically, GANs involve a generator G and a discriminator D that compete with each other in a zero-sum game to optimize the learning parameters. GANs produce sharp images, albeit mostly in a supervised setting and with somewhat unstable performance. CycleGAN [Zhu *et al.*, 2017] builds upon GANs by modeling both forward and backward mapping functions in a closed loop, and is constrained on a cycle-consistency term to encourage more robust image style transformation.

2.2 Problem Formulation

The goal of SRI is to estimate two maps $F : \mathcal{L} \rightarrow \mathcal{H}$ and $G : \mathcal{H} \rightarrow \mathcal{L}$ from a LQ image domain $\mathcal{L} = \{\mathbf{I}_i^L\}_{i=1}^M, \forall i = 1, ..., M$ to a HQ image domain $\mathcal{H} = \{\mathbf{I}_j^H\}_{j=1}^N, \forall j = 1, ..., N$ and *vice versa*, such that the distributions of mapped instances $\hat{p}_l$ and $\hat{p}_h$ can match their ground truths p_l and p_h. Following a cyclic strategy [Wu *et al.*, 2018; Zhu *et al.*, 2017], the cycle-consistency loss is enforced between G and F. The discriminators D_G and D_F are then defined to distinguish the corresponding generations. In this work, solving SRI can be decomposed into two stages:
(1) High→Low (H-L) GAN is constructed to contaminate the HQ data with downsampling, noise and artifacts;
(2) Low→High (L-H) GAN is used for recovering the HQ image from the LQ input.

2.3 H-L GAN

Given two sets of unpaired images $\{\mathbf{I}_i^L\}_{i=1}^M \in \mathcal{L}$ and $\{\mathbf{I}_j^H\}_{j=1}^N \in \mathcal{H}$, we first model the degradation procedure using our H-L GAN, which focuses on simulating the multiple degradations that occur in practical image formation. For each given HQ input image $\mathbf{I}_j^H$, we feed $\mathbf{I}_j^H$ into the generator by first concatenating it with potential downsampling d, blurring b, noise n, and a rectangular missing region r sampled randomly. The architecture is similar to [Ledig *et al.*, 2017], but the first layer is replaced with the result of the concatenation. Since $\mathbf{I}^H, b, n$ may have different dimensions, this results in multiple LQ counterparts for each HQ image. Dimension matching is therefore established by vectorizing and projecting the blur kernel onto an n-dimensional space via principal component analysis (PCA) and then stretching it into a real-valued tensor of the same size as the image channel with the noise level. By doing so, the generator G can be trained with d and formulated as $\mathcal{M} = \{b, n\}$. G is also parameterized by the layer parameters θ_{HL}, including weights w_{HL} and biases γ_{HL}, denoted as $\theta_{HL} = \{w_{HL}, \gamma_{HL}\}$ in a deep network. As with [Ledig *et al.*, 2017], θ_{HL} can be solved by a specific loss function $\mathcal{L}^{HL}$ to optimize the mapping function G:

$$\hat{\theta}_{HL} = \arg\min_{\theta_{HL}} \mathcal{L}_{HL}(G(\mathbf{I}^H), \mathbf{I}^L), \tag{1}$$

where the LQ image $\hat{\mathbf{I}}^L$ can be produced by learning $G(\mathbf{I}^H)$. Instead of using a typical $l_p(p = 1, 2)$ norm to identify the difference, we generalize the task of assessing the quality of LQ data with our H-L GAN. That is, by defining a corresponding discriminator D_G to distinguish the generated LQ images from real ones. This leads to $\mathcal{L}_{AHL} =$

$$\mathbb{E}_{\mathbf{I}^L \sim p_l}\left[\log D_G(\mathbf{I}^L)\right] + \mathbb{E}_{\mathbf{I}^H \sim p_h}\left[\log(1 - D_G(G(\mathbf{I}^H)))\right]. \tag{2}$$

Similar to [Liu *et al.*, 2017; Yi *et al.*, 2017; Zhu *et al.*, 2017], $\mathcal{L}_{AHL}$ is regarded as a unidirectional adversarial loss and optimized alternatively along with G by feeding the unpaired training data. In addition to being able to identify the generated images, another important property which was explored in the traditional reconstruction task is the preservation of common components between LQ and HQ images. To ensure that the resulting LQ data retains the same content as the HQ versions, we add a high-level content loss:

$$\mathcal{L}_{CHL} = \mathbb{E}_{\mathbf{I}^H \sim p_h}\left[\|\phi(G(\mathbf{I}^H)) - \phi(\mathbf{I}^H)\|_1\right], \tag{3}$$

where $\phi(\cdot)$ represents the high-level feature maps obtained from a VGG network. Different from the most related works [Ledig *et al.*, 2017], we adopt an l_1 distance to measure the cross-quality content loss using VGG feature maps between the HQ ground truth and the generated LQ counterparts. This is due to the fact that missing regions in LQ data show very different characteristics from the completed HQ data and, thus, the l_1-norm is able to handle such changes much better than others. $\mathcal{L}_{AHL}$ is then combined with $\mathcal{L}_{CHL}$ to formulate the loss function of H-L GAN.

2.4 L-H GAN

The L-H GAN is built upon results obtained from H-L GAN to recover HQ images $\hat{\mathbf{I}}^H$. Specifically, LQ images $\hat{\mathbf{I}}^L$ generated by H-L GAN are fed to the generator F of the L-H GAN. The generative network is a deep residual CNN similar to the ones used in [Johnson *et al.*, 2016]. The parameters $\theta_{LH} = \{w_{LH}, \gamma_{LH}\}$ are used for estimating $\hat{\mathbf{I}}^H$ by learning a

mapping function F as $\hat{\mathbf{I}}^H = F(\hat{\mathbf{I}}^L)$. Specifically, $\hat{\mathbf{I}}^H$ can be solved by optimizing F using a multi-upgraded loss function $\mathcal{L}_{LH}$ on the training samples:

$$\hat{\theta}_{LH} = \arg\min_{\theta_{LH}} \mathcal{L}_{LH}(F(\hat{\mathbf{I}}^L), \mathbf{I}^H). \tag{4}$$

Different from previous GAN-based SR/inpainting works which only rely on local content matching for adversarial supervision, we propose using a dedicated L-H GAN together with an adversarial loss $\mathcal{L}_{ALH}$, a content loss $\mathcal{L}_{CLH}$, and a texture loss $\mathcal{L}_{TLH}$, to enforce the realism and the consistency.

Generally, L-H GAN is based on adversarial learning, $\mathcal{L}_{ALH}$ is therefore provided to affect the upgrade process in F. The calculation of $\mathcal{L}_{ALH}$ is in accordance with Eq. (2) by simply changing the direction of the LQ/HQ data transformation, *i.e.*

$$\mathbb{E}_{\mathbf{I}^H \sim p_h}\left[\log D_F(\mathbf{I}^H)\right] + \mathbb{E}_{\mathbf{I}^H \sim p_l}\left[\log(1 - D_F(F(\hat{\mathbf{I}}^L)))\right], \tag{5}$$

where D_F is the discriminator corresponding to F. As presented in Eq. (3), the content loss $\mathcal{L}_{CLH}$ is also constrained for L-H GAN to capture the global structure of the underlying common components between the reconstructed HQ images and real ones. It is worth noting that the l_1-norm is utilized in Eq. (3), which is suitable for images with different characteristics, *e.g.* for measuring the distance between $\mathbf{I}^H$ and $\hat{\mathbf{I}}^L$ instead of $\mathbf{I}^H$ and $\hat{\mathbf{I}}^H$. The l_2-loss, on the other hand, is widely used to penalize the discrepancy in image appearance between the ground truth and its pseudo-product. We then define

$$\mathcal{L}_{CLH} = \mathbb{E}_{\mathbf{I}^H \sim p_h}\left[\|\phi(F(\hat{\mathbf{I}}^L)) - \phi(\mathbf{I}^H)\|_2^2\right]. \tag{6}$$

Eq. (6) is applied in the feature maps $\phi(\cdot)$ of the VGG-19 network. In order to generate highly accurate textures, the texture loss $\mathcal{L}_{TLH}$ is used here to guide the generation of fine-scale details. Following [Ulyanov *et al.*, 2016], we adopt a texture descriptor, involving several Gram matrices T_l, in the descriptor VGG network to induce

$$\mathcal{L}_{TLH} = \sum_{l \in L_T} \|T_l(\phi(F(\hat{\mathbf{I}}^L))) - T_l(\phi(\mathbf{I}^H))\|_2^2. \tag{7}$$

2.5 Final Objective

Our framework is based on a cyclic learning system, inspired by [He *et al.*, 2016; Huang *et al.*, 2017a; Liu *et al.*, 2017], where the cycle-consistency constraint is added to enforce forward-backward transformation consistency. As with in [Zhu *et al.*, 2017], the cycle-consistency loss $\mathcal{L}_{cyc}$ is defined as

$$\mathbb{E}_{\mathbf{I}^H \sim p_h}\left[\|F(G(\mathbf{I}^H)) - \mathbf{I}^H\|_1\right] + \mathbb{E}_{\mathbf{I}^L \sim p_l}\left[\|G(F(\mathbf{I}^L)) - \mathbf{I}^L\|_1\right]. \tag{8}$$

With H-L GAN loss $\mathcal{L}_{HL}$, L-H GAN loss $\mathcal{L}_{LH}$, and cycle-consistency loss $\mathcal{L}_{cyc}$, the model objective of our HLH-GAN is defined as

$$\mathcal{L}_{HLH} = \mathcal{L}_{HL} + \mathcal{L}_{LH} + \lambda \mathcal{L}_{cyc}. \tag{9}$$

where $\mathcal{L}_{HL}$ and $\mathcal{L}_{LH}$ can be extended as $\mathcal{L}_{HL} = \mathcal{L}_{AHL} + \alpha\mathcal{L}_{CHL}$,$\mathcal{L}_{LH} = \mathcal{L}_{ALH} + \delta\mathcal{L}_{CLH} + \beta\mathcal{L}_{TLH}$.

3 Experiments

3.1 Network Architecture

H-L GAN: We designed the generator as an encoder-decoder network with the stretched real-valued tensor as the first layer, and 12 identical residual blocks as the middle layers, followed by a convolutional layer with stride 2, 4 and scale factor of 2, 4, respectively. The stride of last layer is 1, and all other convolutional layers follow the above setting using $3\times3\times3$ kernels with 32 filters. For ϕ, we utilize the feature maps obtained by the 4-th convolution before the 4-th maxpooling layer within the VGG19 network. We use Adam with 10^5 iterations and a learning rate of 10^{-4}, which is decayed by a factor of 2 every 2×10^5 minibatch updates. To train the discriminative network, we follow [Ledig *et al.*, 2017] using six residual blocks without batch normalization, followed by a fully connected layer.

L-H GAN: The generator uses 17 residual blocks distributed as 12-3-2 with skip connections between them. The architecture is generally the same as in [Ledig *et al.*, 2017], but we choose different hidden layers depending on the loss. The content loss is calculated based on the feature maps from the relu5_1 layer, and the texture loss is computed by combining the relu3_1 and relu4_1 layers of VGG19. The discriminator is that of a general GAN.

Weights For the parameters, we set $\alpha = 10$, $\delta = 10$, $\beta = 0.1$, $\lambda = 1$.

3.2 Datasets

We evaluate the proposed method on two publicly available datasets: IXI [1], and HCP [2], which include real acquired LQ/HQ data. Specifically, the IXI dataset contains 578 healthy subjects acquired by a Philips 3T/1.5T system and a GE 1.5T system. One branch of the HCP dataset has a total of 200 subjects acquired via a Siemens 3T scanner. The HQ images were observed by a IXI-Philips 3T and HCP-Siemens 3T with complete content and no added blur or noise. The LQ images were acquired by IXI-Philips 1.5T/GE 1.5T and HCP-Siemens 3T scanners [3]. In addition, we collected 12 clinical LQ images from the local hospital, which consist of various LR images with unknown degradation and noise. We split the datasets into 500 (IXI) and 120 (HCP) for training, 78 (IXI) and 80 (HCP) for testing.

3.3 Comparison and Results

We compare HLH-GAN against several state-of-the-art approaches, including Bicubic, ScSR [Yang *et al.*, 2010], SelfEx [Huang *et al.*, 2015], Zeyde [Zeyde *et al.*, 2010], DRCN [Kim *et al.*, 2016b], SRCNN [Dong *et al.*, 2015], and SRGAN [Ledig *et al.*, 2017]. We use the default settings of the compared methods to obtain their best super-resolved effects. To fully evaluate the effectiveness of our method in different scenarios, we conduct a comprehensive evaluation in three ways: (1) image super-resolution without deblurring and denoising (termed as SR-B-D); (2) image

[1] http://brain-development.org/ixi-dataset

[2] https://www.humanconnectome.org

[3] HCP data were downsampled with destructed content, blur and noise

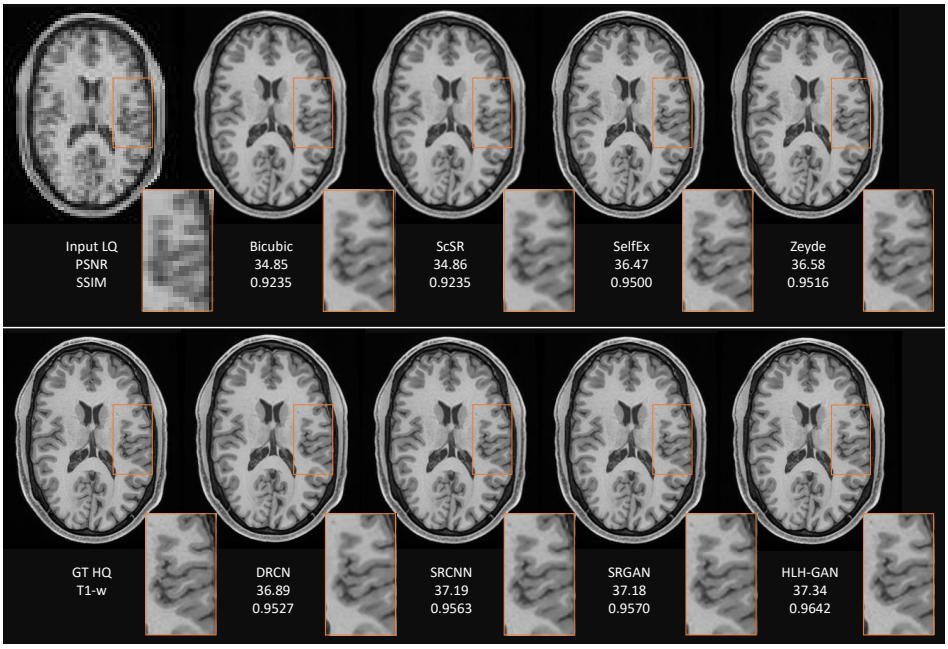

Figure 2: Example results for SR-B-D by HLH-GAN and compared methods.

Exp.	Bicubic		ScSR		SelfEX		Zeyde		DRCN		SRCNN		SRGAN		HLH-GAN	
IXI: SR-N-B (PSNR(dB), SSIM)																
PD	*31.85*	*0.8891*	31.93	0.8901	32.43	0.8976	32.07	0.8964	33.01	0.9011	33.78	0.9119	34.00	0.9025	**35.14**	**0.9337**
T1	*30.03*	*0.8653*	30.05	0.8677	31.40	0.8832	31.18	0.8827	32.74	0.8993	33.06	0.9007	33.69	0.9009	**34.82**	**0.9201**
T2	*31.17*	0.8972	32.41	*0.8968*	32.17	0.8967	32.25	0.8979	33.96	0.9123	34.25	0.9188	34.26	0.9183	**35.01**	**0.9305**
IXI: SR-B (PSNR(dB), SSIM)																
PD	*29.86*	*0.8553*	29.87	*0.8553*	30.02	0.8654	30.01	0.8647	31.82	0.8912	32.59	0.9003	33.16	0.9004	**34.99**	**0.9273**
T1	*27.77*	*0.8201*	27.79	0.8203	29.16	0.8578	29.01	0.8558	31.00	0.8872	32.11	0.8972	32.44	0.8969	**34.02**	**0.9116**
T2	*29.92*	*0.8617*	29.96	0.8621	29.98	0.8623	29.96	0.8617	31.76	0.8964	32.52	0.8996	32.97	0.9001	**34.86**	**0.9259**
IXI: SRI (PSNR(dB), SSIM)																
PD	*26.02*	*0.7687*	26.07	0.7689	27.64	0.7837	27.62	0.7834	29.62	0.8201	30.11	0.8394	31.45	0.8388	**32.17**	**0.8972**
T1	*25.78*	*0.7642*	25.79	*0.7642*	26.39	0.7701	26.21	0.7683	28.39	0.7943	29.45	0.8157	30.26	0.8003	**31.86**	**0.8819**
T2	*25.92*	*0.7660*	25.94	0.7676	27.58	0.7826	27.39	0.7815	29.47	0.8189	30.00	0.8316	31.20	0.8309	**32.04**	**0.8961**

Table 1: Quantitative evaluations of SR-B-D, SR-B and SRI on the IXI dataset. Best results are highlighted.

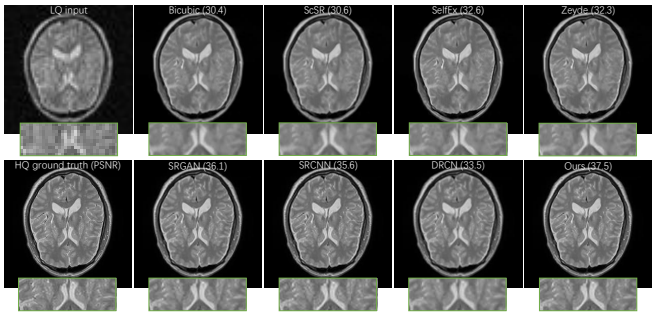

Figure 3: Visual comparison of HLH-GAN and competitors on SR-B.

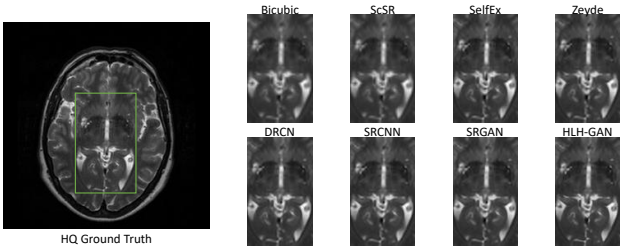

Figure 4: Performance comparison of different methods with zoomed-in details.

super-resolution without deblurring (termed as SR-B); (3) simultaneous super-resolution and inpainting (termed as SRI). To demonstrate the effectiveness of HLH-GAN, the quantitative results are provided in Tables 1-3 and visual results are given in Figs. 2-5.

Specifically, Fig. 2 presents a set of results for SR-B-D. As can be seen, all compared methods are able to generate clear results but some tend to have blurry details. DRCN, SRCNN, SRGAN and HLH-GAN can recover clearer and more textural details. Another point worth noting is that, among the four superior methods (*i.e.*, DRCN, SRCNN, SRGAN and HLH-GAN), HLH-GAN obtains the best quality performance and displays the most obvious visual fidelity. We then show the overall PSNR and SSIM indices of all the competing models for the three scenarios, experimented on the IXI dataset. The quantitative results are listed in Table 1. As

Exp.	Bicubic		ScSR		SelfEX		Zeyde		DRCN		SRCNN		SRGAN		HLH-GAN	
HCP: SR-N-B (PSNR(dB), SSIM)																
T1	*30.66*	*0.8694*	30.67	0.8695	31.16	0.8839	31.14	0.8820	32.25	0.8972	32.81	0.9007	33.21	0.9019	**34.58**	**0.9213**
T2	*31.42*	*0.8801*	31.44	*0.8801*	31.89	0.8897	31.77	0.8864	32.47	0.8996	32.83	0.9016	33.66	0.9111	**34.76**	**0.9280**
HCP: SR-B (PSNR(dB), SSIM)																
T1	*28.01*	*0.8255*	28.06	0.8259	28.37	0.8351	28.34	0.8349	28.78	0.8397	29.16	0.8426	30.45	0.8501	**32.01**	**0.8974**
T2	*28.24*	*0.8279*	*28.24*	0.8280	28.41	0.8374	28.40	0.8371	28.94	0.8401	29.39	0.8497	30.55	0.8523	**32.47**	**0.8998**
HCP: SRI (PSNR(dB), SSIM)																
T1	*26.62*	*0.7983*	26.69	0.7986	27.04	0.8012	27.04	0.8010	28.26	0.8507	28.99	0.8668	29.68	0.8701	**31.50**	**0.8918**
T2	*26.68*	*0.7986*	26.71	0.8000	27.15	0.8153	27.09	0.8142	28.51	0.8582	29.37	0.8753	29.69	0.8701	**31.69**	**0.8924**

Table 2: Quantitative evaluations of SR-B-D, SR-B and SRI on the HCP dataset. Best results are highlighted.

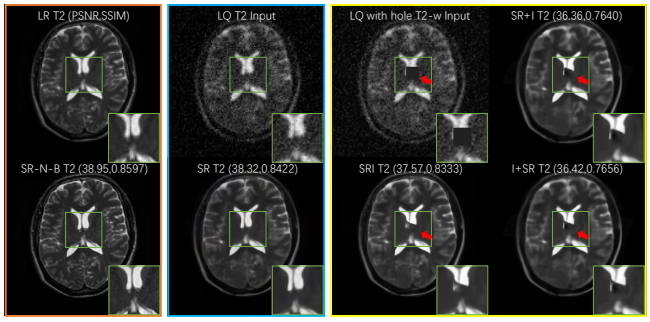

Figure 5: Visualization results of HLH-GAN on SR-N-B, SR-B, SRI, SR+I and I+SR cases.

Exp.	SRI		SR+I		I+SR	
IXI: T2-w MRI						
PSNR(dB) SSIM	**32.04**	**0.8961**	28.59	0.8501	28.32	0.8377
HCP: T2-w MRI						
PSNR(dB) SSIM	**31.69**	**0.8924**	28.26	0.8376	27.98	0.8306

Table 3: Performance verification (PSNR (dB) and SSIM): HLH-GAN under various conditions on T2-w brain MRI from IXI and HCP datasets. Best results are highlighted.

can be seen, HLH-GAN consistently produces much better results than other approaches, with PSNR and SSIM gains of 1.28±0.6(dB) and 0.04±0.03, respectively. Note that the two special cases constructed *i.e.*, SR-B and SRI, are more challenging than SR-B-D. Visual comparisons are provided in Figs. 3-5. As shown in Fig. 3, the typical Bicubic, ScSR, SelfEx and Zeyde produce HR data but with overly smooth surfaces, while deep generative methods are effective in reconstructing denoised HR images with precise smoothness. We also show the zoomed-in results of different methods in Fig. 4, from which we can see that most compared approaches can reconstruct content with higher resolution but suffer from blurry artifacts. Instead, our method (HLH-GAN) achieves sharper and more realistic details. To further prove the effectiveness of HLH-GAN, we also evaluate the above three cases on the HCP dataset and report all the quantitative results in Table 2. Compared with other methods, HLH-GAN also performs the best on various modality (*i.e.*, T1 and T2) data throughout the experiments.

To validate our framework, we compare the results of simultaneous SR and inpainting (denoted as SRI) with (a) first SR and then inpainting (termed as SR+I) and (b) first inpainting and then SR (called as I+SR). We show their visual results in Fig. 5 and averaged PSNR and SSIM in Table 3. The first column of Fig. 5 demonstrates results generated by HLH-GAN for SR-N-B, the second column displays our results for SR-B, while the last yellow block shows the compared performance of SRI, SR+I and I+SR. As can be seen from the figures, HLH-GAN can produce visually pleasant results for various cases, but the performance is still affected by intricate artifacts. Particularly when recovering the holed region, our unified framework produces better inpainting results than independent processing. However, it also generates obvious errors around the borders of completely absent areas. We also analyze how our HLH-GAN contributes to the final performance of SRI, conducting experiments on the T2-w MRI of both the IXI and HCP datasets. We compare our HLH-GAN to the model with separate processes, as shown in Table 3. For both datasets, our method consistently overcomes SR+I and I+SR, which indicates that the unified model improves the effectiveness of both image SR and inpainting, and that the integration is crucial to the success of HLH-GAN.

In summary, our method yields the best results against the compared approaches proving our claim of being able to simultaneously conduct SR and inpainting for better results (with our model performing particularly well on SRI task). Since the HCP dataset includes HQ data, the advantage of HLH-GAN is shown in the high-level details rather than overall appearance (referring to Fig. 2). Our method evaluated on the IXI dataset outperforms the existing state-of-the-art approaches by a large margin.

4 Conclusion

We proposed an HLH-GAN, including two submodules (H-L GAN and L-H GAN), for integrated image SR and inpainting. We showed that our method improves LQ images by first learning a natural image degradation process and then upgrading them. We also demonstrated good results for simultaneous SR and image inpainting using brain images, especially when leveraging the LQ ground truth during training. Our model significantly outperforms various state-of-the-art models both quantitatively and qualitatively. As a future work, we plan to apply our method to augment HQ data in longitudinal imaging studies.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61972188), Guangdong Provincial Key Laboratory (Grant No. 2020B121201001), the Program for University Key Laboratory of Guangdong Province (Grant No. 2017KSYS008).

References

[Criminisi *et al.*, 2004] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE TIP*, 13(9):1200–1212, 2004.

[Dong *et al.*, 2015] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE TPAMI*, 38(2):295–307, 2015.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680, 2014.

[He *et al.*, 2016] Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, and Wei-Ying Ma. Dual learning for machine translation. In *NeurIPS*, pages 820–828, 2016.

[Huang *et al.*, 2015] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *IEEE CVPR*, pages 5197–5206, 2015.

[Huang *et al.*, 2017a] Yawen Huang, Ling Shao, and Alejandro F Frangi. Dote: Dual convolutional filter learning for super-resolution and cross-modality synthesis in mri. In *MICCAI*, pages 89–98. Springer, 2017.

[Huang *et al.*, 2017b] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In *IEEE CVPR*, pages 6070–6079, 2017.

[Iizuka *et al.*, 2017] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM ToG*, 36(4):107, 2017.

[Johnson *et al.*, 2016] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711, 2016.

[Kim *et al.*, 2016a] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *IEEE CVPR*, pages 1646–1654, 2016.

[Kim *et al.*, 2016b] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *IEEE CVPR*, pages 1637–1645, 2016.

[Ledig *et al.*, 2017] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE CVPR*, pages 4681–4690, 2017.

[Li and Wand, 2016] Chuan Li and Michael Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *IEEE CVPR*, pages 2479–2486, 2016.

[Liu *et al.*, 2017] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *NeurIPS*, pages 700–708, 2017.

[Manjón *et al.*, 2010] José V Manjón, Pierrick Coupé, Antonio Buades, Vladimir Fonov, D Louis Collins, and Montserrat Robles. Non-local mri upsampling. *MIA*, 14(6):784–792, 2010.

[Parker *et al.*, 1983] J Anthony Parker, Robert V Kenyon, and Donald E Troxel. Comparison of interpolating methods for image resampling. *IEEE TMI*, 2(1):31–39, 1983.

[Pathak *et al.*, 2016] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *IEEE CVPR*, pages 2536–2544, 2016.

[Schulter *et al.*, 2015] Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In *IEEE CVPR*, pages 3791–3799, 2015.

[Tai *et al.*, 2010] Yu-Wing Tai, Shuaicheng Liu, Michael S Brown, and Stephen Lin. Super resolution using edge prior and single image detail synthesis. In *IEEE CVPR*, pages 2400–2407. IEEE, 2010.

[Ulyanov *et al.*, 2016] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In *ICML*, volume 1, page 4, 2016.

[Wu *et al.*, 2018] Lin Wu, Yang Wang, and Ling Shao. Cycle-consistent deep generative hashing for cross-modal retrieval. *IEEE TIP*, 28(4):1602–1612, 2018.

[Yan *et al.*, 2018] Zhaoyi Yan, Xiaoming Li, Mu Li, Wangmeng Zuo, and Shiguang Shan. Shift-net: Image inpainting via deep feature rearrangement. In *ECCV*, pages 1–17, 2018.

[Yang *et al.*, 2010] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010.

[Yi *et al.*, 2017] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *IEEE ICCV*, pages 2849–2857, 2017.

[Yu and Koltun, 2015] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[Zeyde *et al.*, 2010] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

[Zhang *et al.*, 2012] Kaibing Zhang, Xinbo Gao, Dacheng Tao, and Xuelong Li. Multi-scale dictionary for single image super-resolution. In *IEEE CVPR*, pages 1114–1121. IEEE, 2012.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE ICCV*, pages 2223–2232, 2017.