ISP2HRNet: Learning to Reconstruct High Resolution Image from Irregularly Sampled Pixels via Hierarchical Gradient Learning

Yuanlin Wang[1], Ruiqin Xiong[1]*, Rui Zhao[1], Jin Wang[2], Xiaopeng Fan[3], Tiejun Huang[1]
[1]State Key Laboratory of Multimedia Information Processing,
School of Computer Science, Peking University
[2]College of Computer Science, Beijing University of Technology
[3]School of Computer Science and Technology, Harbin Institute of Technology
{wangyuanlin, ruizhao}@stu.pku.edu.cn {rqxiong, tjhuang}@pku.edu.cn
ijinwang@bjut.edu.cn fxp@hit.edu.cn

Abstract

While image signals are typically defined on a regular 2D grid, there are scenarios where they are only available at irregular positions. In such cases, reconstructing a complete image on regular grid is essential. This paper introduces ISP2HRNet, an end-to-end network designed to reconstruct high resolution image from irregularly sampled pixels that do not fall on a regular grid. To handle the challenges brought by irregular sampling, we propose an architecture to extract gradient structure hierarchically and learn continuous image representation. Specifically, we derive image gradient for each irregularly sampled pixel and further learn higher order gradient structural features according to the geometric and photometric information at the vertices of neighboring triangles. To convert the features from irregular pixels to regular grid, we propose a dual branch content-dependent weight generator to adaptively fuse the information from neighboring irregular pixels. Subsequently, an encoder captures deep structural details on regular grid and forms latent codes. Implicit neural representation parameterized by multi-layer perceptron decodes the latent codes and coordinates to pixel values for generating high resolution image. Experimental results demonstrate that the proposed network effectively solves the problem of high resolution image reconstruction from irregularly sampled pixels and achieves promising results. The source code is available at https://github.com/yuanlinwang/ISP2HRNet.

1. Introduction

Image signals are generally represented by a regular 2D grid. However, there also exist scenarios where they are only sampled at irregular positions [14, 36, 41, 44, 47]. Reconstructing image from irregularly sampled pixels is a classic problem in image processing field. Many works have been conducted to address this issue in the past decades. Spline-based interpolation methods [9, 16, 23, 46] fit a polynomial for irregular pixels and use the fitted polynomial to estimate pixel values at arbitrary positions. Filtering-based methods [33, 42, 45] estimate pixel values at arbitrary positions by weighted averaging surrounding pixels. The performance of these methods is limited due to insufficient adaptability to complex textures. Some works [3, 12, 13] deal with irregular data in an intuitive but ad hoc way by assigning zero to non-informative pixels, creating an artificial Cartesian image as input, and designing an end-to-end trainable network to restore the degraded input. However, in some cases, mapping all irregular pixels onto a finite-resolution grid and forming a regular array with unknown pixels set to zero can be challenging.

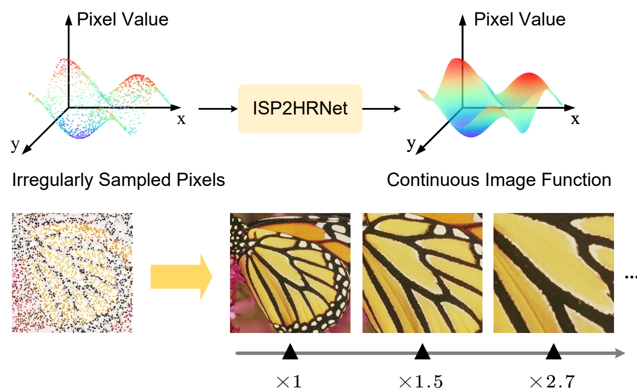


Figure 1. Overview of high resolution image reconstruction from irregularly sampled pixels.

This paper addresses the problem of reconstructing high resolution (HR) image from irregularly sampled pixels. The

*Corresponding author.

problem presents several key challenges. First, irregularly sampled pixels are unordered, unstructured, and do not fall on a regular grid. They cannot be easily represented by a 2D array, requiring a new form of pixel representation. Second, deep neural networks commonly used in image signal processing, such as convolutional neural networks (CNN), struggle to handle irregular input, requiring a tailored design to operate on these irregular pixels. Third, converting information from irregular pixels to regular grid demands treating each neighboring irregular pixel individually due to their diverse local structures.

This paper develops an end-to-end network named ISP2HRNet to achieve irregularly sampled pixels to HR image reconstruction. Unlike a regular pixel array, we represent irregularly sampled pixels as a list for network input. To capture image structure information from these pixels, we propose an architecture that learns gradient-related structure features from irregular pixels and regular grid hierarchically. We first derive image gradients at irregular pixel locations, and further learn higher-order gradient structural features based on the geometric and photometric information at the vertices of neighboring triangles. Then, a dual branch content-dependent weight generator is introduced to adaptively fuse information from neighboring irregular pixels, enabling the mapping of irregular pixel information onto a regular grid. Subsequently, an encoder is used to extract deep structural details and form the latent codes on regular grid. Finally, we use the implicit neural representation (INR) driven by latent codes and coordinates for HR image reconstruction, leveraging its capability for continuous image representation. Experiments demonstrate that the proposed method effectively solves the problem of HR image reconstruction from irregularly sampled pixels and achieves promising results. The main contributions of this paper are as follows:

- We develop an end-to-end network named ISP2HRNet to reconstruct HR image from irregularly sampled pixels. To the best of our knowledge, this is the first attempt to tackle this problem using a deep neural network.
- We propose a hierarchical architecture that progressively extracts image gradient structures from irregular coordinates to regular grid.
- We extend HR image reconstruction problem from regular pixel array to irregularly sampled pixels. Experimental results demonstrate that the proposed ISP2HRNet achieves promising results.

2. Related Work

Image Reconstruction from Irregularly Sampled Pixels. Image reconstruction from irregularly sampled pixels has been widely studied, aiming to restore a complete image based on the available irregular samples. In some works, it is also regarded as an image inpainting problem. Spline-based methods [9, 16, 23, 46] fit a polynomial to irregular pixels and use the fitted polynomial to estimate pixel values at queried coordinates. Filtering-based methods [33, 42, 45] estimate pixel values at queried positions by weighted averaging surrounding pixels. Optimization-based methods [1, 5, 11, 54, 55] address the image restoration problem by minimizing an optimization objective function. With the development of deep learning, some works [3, 12, 13, 53] explore unfolding the optimization process as a deep neural network for image restoration, which combines the merits of optimization-based and learning-based methods. Image inpainting works [19, 24, 26, 29] use a binary mask to represent sampled pixels and unknown pixels, and predict missing parts of an image through deep semantic features.

Arbitrary Scale Image Super Resolution. Many image super resolution (SR) networks [7, 27, 28, 50] have been developed to reconstruct HR image from low resolution (LR) input. However, these methods lack the flexibility to handle arbitrary-scale SR with a single model. To address this issue, some works [6, 17, 22, 43, 48, 58] have been proposed recently. Meta-SR [17] proposes a meta upscale module to dynamically predict the weights of upsampling filters by taking scale factors as input, and uses these weights to generate HR images of any resolution. ArbSR [48] develops a plug-in module which uses conditional convolution to generate dynamic scale-aware filters, enabling the network to adapt to arbitrary scale factors. With the development of INR in the field of 3D vision [15, 31, 32, 35, 37], researchers attempt to utilize INR for continuous image representation. LIIF [6] learns the mapping from coordinates and latent codes to pixel values through MLP, achieving favorable results on arbitrary HR image reconstruction. Based on LIIF, some subsequent works [22, 51] introduce position encoding to enhance the capability of implicit function to reconstruct finer details.

Processing of Irregular Data. Point cloud is a typical irregular data, being unordered and unstructured, and shares the similar characteristics with irregularly sampled pixels. PointNet [38] proposes a unified architecture for 3D point cloud object classification, part segmentation, and scene semantic parsing. PointNet directly processes unordered point sets and utilizes the shared multi-layer perceptron (MLP) to extract features of each point. Besides, PointNet uses symmetric functions to aggregate information from points to enable output invariant to the input order. Most subsequent point cloud related works [8, 10, 25, 39, 40] adopt the fundamental processing strategies introduced in PointNet.

3. Method

3.1. Problem Statement and Motivation

Suppose we have N pixels $\{\mathrm{S}_i\}_{i=1}^{\mathrm{N}}$ sampled at irregular coordinates. Each sample S_i contains a vector $\mathbf{f}_i =$

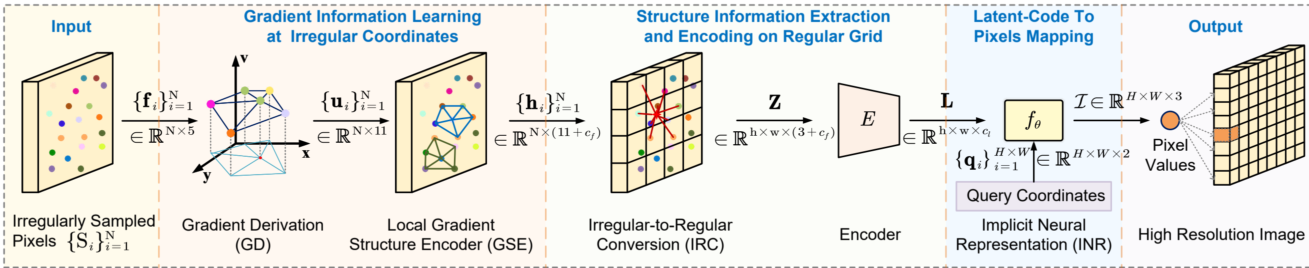


Figure 2. The overall framework of the proposed ISP2HRNet for HR image reconstruction from irregularly sampled pixels. Given a set of irregular pixels, we first learn gradient structure features at irregular coordinates through the GD and GSE modules. Then, IRC module converts the information from irregular pixels onto a regular grid. An encoder subsequently extracts deeper structural features from the regular grid data, forming the latent codes. Finally, implicit neural representation parameterized by MLP decodes the latent codes and coordinates to pixel values, enabling high resolution image reconstruction.

$(x_i, y_i, \mathbf{v}_i) \in \mathbb{R}^{1\times 5}$ describing its characteristics, where (x_i, y_i) is the coordinate and $\mathbf{v}_i = (v_i^{\mathrm{R}}, v_i^{\mathrm{G}}, v_i^{\mathrm{B}})$ is the pixel value. The goal of HR image reconstruction from irregularly sampled pixels is to generate an image $\mathcal{I} \in \mathbb{R}^{H\times W\times 3}$ based on a set of irregular samples $\{\mathrm{S}_i\}_{i=1}^{\mathrm{N}}$.

A continuous image function allows for high-resolution reconstruction at arbitrary spatial resolutions. One fundamental strategy for modeling this function from irregularly sampled pixels is to apply the Taylor expansion, which locally approximates the continuous function with a polynomial. The Taylor expansion of the continuous image function $\mathcal{I}$ at any coordinate (a, b) can be formulated as:

$$\begin{aligned}\mathcal{I}(x,y) &= \mathcal{I}(a,b) + \mathcal{I}_x(a,b)(x-a) + \mathcal{I}_y(a,b)(y-b) \\ &+ \frac{1}{2}[\mathcal{I}_{xx}(a,b)(x-a)^2 + 2\mathcal{I}_{xy}(a,b)(x-a)(y-b) \\ &+ \mathcal{I}_{yy}(a,b)(y-b)^2] + o^n,\end{aligned} \tag{1}$$

where $(x, y) \in \mathcal{N}(a, b)$, $\mathcal{N}(a, b)$ denotes the local neighborhood of (a, b). $\mathcal{I}_x = \frac{\partial \mathcal{I}}{\partial x}$ and $\mathcal{I}_y = \frac{\partial \mathcal{I}}{\partial y}$ are the first-order gradients of image $\mathcal{I}$, $\{\mathcal{I}_{xx} = \frac{\partial^2 \mathcal{I}}{\partial x^2}, \mathcal{I}_{xy} = \frac{\partial^2 \mathcal{I}}{\partial x \partial y} = \mathcal{I}_{yx}, \mathcal{I}_{yy} = \frac{\partial^2 \mathcal{I}}{\partial y^2}\}$ are the second-order gradients of $\mathcal{I}$. The local approximation polynomial function is determined by the above image gradient structural information. Based on the continuous function (Eq. (1)), pixel values at arbitrary coordinates within $\mathcal{N}(a, b)$ can be derived [45].

Inspired by the concept of Taylor expansion, we focus on several key aspects of HR image reconstruction from irregularly sampled pixels. First, gradients are key elements in reflecting image structure. While gradients and higher-order gradients are readily obtained from a regular image array, the challenge here lies in deriving them from randomly distributed irregular pixels. Second, higher-order gradient information provides a more precise description of local variations in the image. To capture this, we develop a tailored module for irregular pixels to further learn higher-order structural features. Moreover, converting irregular pixel information into regular grid enables the use of well-designed CNN modules to extract deeper features from regular grid data. Third, instead of relying on a finite-order polynomial driven by gradients and high-order gradients (Eq. (1)), a more general function model (MLP) driven by extracted latent features is utilized to derive the continuous image function, enabling HR image reconstruction. Based on the above discussion, we design a network named ISP2HRNet, which incorporates and extends the key concepts from Taylor expansion (see the upper left part of Fig. 3) to derive a continuous image function from irregularly sampled pixels.

3.2. Overall Framework

The overall framework of ISP2HRNet is illustrated in Fig. 2. ISP2HRNet directly takes a list of irregular pixels $\{\mathrm{S}_i\}_{i=1}^{\mathrm{N}}$ as input. To fully utilize the information contained in irregular pixels, we first introduce a gradient derivation module (GD) to estimate the gradient at each irregular pixel location. Then, we develop a local gradient structure encoder (GSE) to extract more accurate and higher order gradient structural features for each irregular pixel, based on the geometric and photometric information at the vertices of neighboring triangles. Next, an irregular-to-regular conversion module (IRC) is proposed to transform the local structural features from irregular positions to regular grid by adaptively fusing the information from neighboring irregular pixels. The converted regular grid features are then passed through an encoder to extract the latent code for each regular grid position. Finally, an MLP-based INR learns the continuous image function, taking the latent codes and queried coordinates (i.e., the coordinates of target HR image) as input to generate pixel values for HR image reconstruction.

3.3. Gradient Derivation

Gradient derivation (GD) module estimates the gradient at each irregular pixel location, based on the structure information contained in K triangles formed from K nearest neighbors of the pixel, as shown in Fig. 3(a). For each irregularly sampled pixel S_i, we search for its K nearest neighbors. For each neighbor, we calculate its relative coordinate with respect to S_i and convert the Cartesian coordinate to

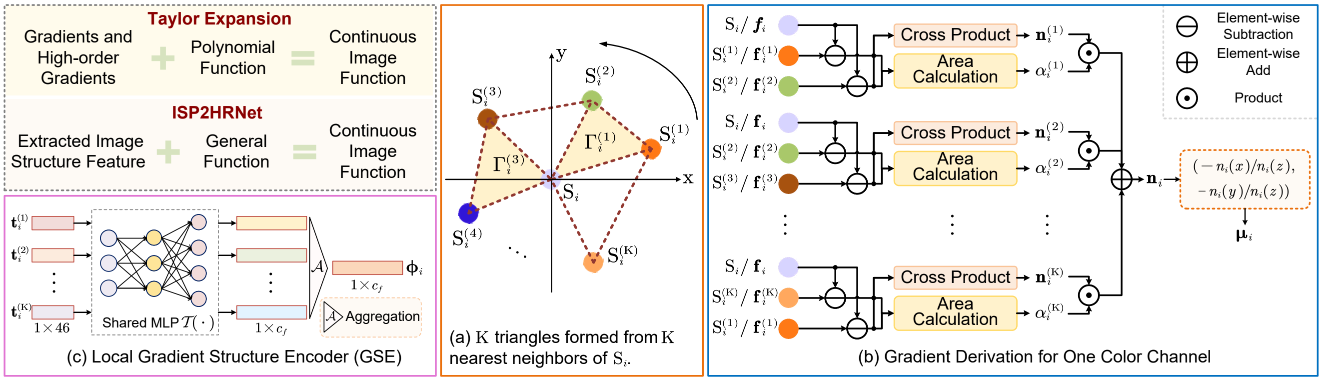


Figure 3. Upper left: The ideas of Taylor expansion and the proposed ISP2HRNet. (a) K triangles formed from K nearest neighbors. (b) Gradient estimation for each sampled pixel in one color channel. (c) Local gradient structure encoder.

polar coordinate. Then we sort the K neighboring pixels counterclockwise by angle, the ordered neighboring pixel sequence is denoted by $\{\mathrm{S}_i^{(1)}, \mathrm{S}_i^{(2)}, ..., \mathrm{S}_i^{(\mathrm{K})}\}$, as shown in Fig. 3(a). We select $\{\{\mathrm{S}_i^{(l)}, \mathrm{S}_i^{(l+1)}, \mathrm{S}_i\}, l = 1, ..., \mathrm{K}\}$ (When $l = \mathrm{K}$, $\{\mathrm{S}_i^{(K)}, \mathrm{S}_i^{(1)}, \mathrm{S}_i\}$ is selected.) as the vertices of triangles to form K triangles $\{\Gamma_i^{(l)}\}_{l=1}^{\mathrm{K}}$. The vector containing the coordinate and pixel value of $\mathrm{S}_i^{(l)}$ is denoted by $\mathbf{f}_i^{(l)}$. ($\mathbf{f}_i$, $\mathbf{u}_i$ and $\mathbf{h}_i$ (to be introduced later) are vectors that contain the information about the irregular pixel S_i. The superscript and subscript of $\mathbf{f}$, $\mathbf{u}$ and $\mathbf{h}$ of a pixel are consistent with the notation of this pixel, unless otherwise specified.)

The K triangles cover all directions around S_i, providing structural information for gradient estimation at S_i. The gradient derivation is performed independently for each color channel. The diagram illustrating this process for one color channel is presented in Fig. 3(b). We first calculate the unit normal vector $\mathbf{n}_i^{(l)}$ of each triangle through counterclockwise cross product and normalize operation. After obtaining unit normal vectors $\{\mathbf{n}_i^{(l)}\}_{l=1}^{\mathrm{K}}$ from K triangles, the unit normal vector $\mathbf{n}_i = (n_i(x), n_i(y), n_i(z))$ of S_i is estimated by a weighted average of $\{\mathbf{n}_i^{(l)}\}_{l=1}^{\mathrm{K}}$, where the weighted factor is the area of each triangle. Finally, the gradient $\boldsymbol{\mu}_i$ at S_i for one color channel is derived by:

$$\boldsymbol{\mu}_i = (-n_i(x)/n_i(z), -n_i(y)/n_i(z)) \,. \tag{2}$$

The gradient is computed separately for each color channel, resulting in $\mathbf{g}_i = \left(\boldsymbol{\mu}_i^{\mathrm{R}}, \boldsymbol{\mu}_i^{\mathrm{G}}, \boldsymbol{\mu}_i^{\mathrm{B}}\right) \in \mathbb{R}^{1\times 6}$, which includes the gradients of RGB channels.

Through the GD module, the vector $\mathbf{f}_i = (x_i, y_i, \mathbf{v}_i)$ of S_i is extended to $\mathbf{u}_i = (\mathbf{f}_i, \mathbf{g}_i) \in \mathbb{R}^{1\times 11}$, which includes not only the coordinate and pixel value, but also the estimated gradients.

3.4. Local Gradient Structure Encoder

The gradients obtained from the GD module are insufficient to accurately describe the local structural characteristics of the image. To further extract more accurate and higher order gradient structural features, we propose a local gradient structure encoder (GSE) that leverages the learning ability of neural networks. Following the K neighboring triangles formed by each pixel S_i and its K nearest neighbors in GD module (Fig. 3(a)), we form a vector $\mathbf{t}_i^{(l)}$ for each triangle $\Gamma_j^{(l)}$, containing the geometric and photometric information at its vertices. We define $\mathbf{t}_i^{(l)}$ as:

$$\begin{aligned}\mathbf{t}_i^{(l)} = \Big(&\mathbf{u}_i^{(l)}, \mathbf{u}_i^{(l+1)}, \mathbf{u}_i, \Delta\mathbf{f}_i^{(l)}, \Delta\mathbf{f}_i^{(l+1)}, \\ &d_i^{(l)}, d_i^{(l+1)}, \sin(\theta_i^{(l)})\Big) \in \mathbb{R}^{1\times 46},\end{aligned} \tag{3}$$

where $\mathbf{u}_i^{(l)}, \mathbf{u}_i^{(l+1)}, \mathbf{u}_i$ includes absolute coordinates, absolute pixel values and gradients of three vertices $\{\mathrm{S}_i^{(l)}, \mathrm{S}_i^{(l+1)}, \mathrm{S}_i\}$, $\Delta\mathbf{f}_i^{(l)} = \mathbf{f}_i^{(l)} - \mathbf{f}_i = (x_i^{(l)} - x_i, y_i^{(l)} - y_i, \mathbf{v}_i^{(l)} - \mathbf{v}_i)$ represents the relative coordinate and pixel value of neighboring pixel $\mathrm{S}_i^{(l)}$ with respect to S_i, $d_i^{(l)} = \sqrt{(x_i^{(l)} - x_i)^2 + (y_i^{(l)} - y_i)^2}$ is the Euclidean distance between $\mathrm{S}_i^{(l)}$ and S_i, $\sin(\theta_i^{(l)})$ is sine of the angle $\theta_i^{(l)}$ corresponding to the vertex S_i in the triangle $\Gamma_i^{(l)}$. Each S_i would obtain K vectors $\{\mathbf{t}_i^{(1)}, \mathbf{t}_i^{(2)}, ..., \mathbf{t}_i^{(\mathrm{K})}\}$ from K triangles. The diagram of GSE is presented in Fig. 3(c). GSE learns higher order gradient structure for each triangle $\Gamma_i^{(l)}$ by feeding $\mathbf{t}_i^{(l)}$ to a shared MLP. Then, GSE fuses the learned structure information from K triangles to obtain the structure feature of S_i. The pipeline of GSE can be formulated as:

$$\boldsymbol{\phi}_i = \mathcal{A}(\{\mathcal{T}(\mathbf{t}_i^{(l)}), \forall l \in \{1..., \mathrm{K}\}\}). \tag{4}$$

where $\mathcal{T}(\cdot)$ is a structure extraction function implemented by MLP and is shared over $\{\mathbf{t}_i^{(l)}\}_{l=1}^{\mathrm{K}}$, $\mathcal{A}(\cdot)$ is an aggregation function (*e.g.*, average), which fuses the features from K triangles and ensures that the output is invariant to the input order, $\boldsymbol{\phi}_i \in \mathbb{R}^{1\times c_f}$ is the extracted structural feature.

After GSE, each irregular pixel S_i obtains a structure feature vector $\boldsymbol{\phi}_i$. Then, the vector $\mathbf{u}_i$ of S_i is further extended

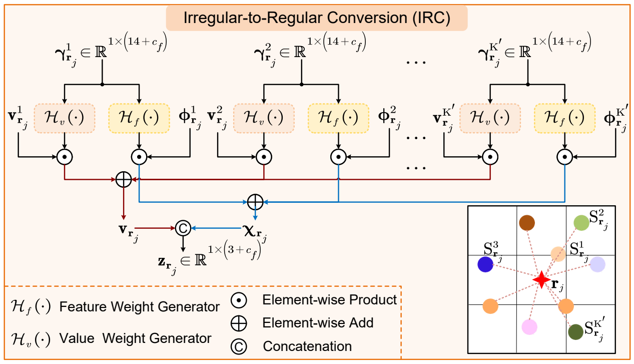


Figure 4. Irregular-to-regular conversion (IRC) module. We propose a dual branch content-dependent weight generator $\{\mathcal{H}_v(\cdot), \mathcal{H}_f(\cdot)\}$ to adaptively fuse the information from neighboring irregular pixels and form regular gird features.

to $\mathbf{h}_i = (\mathbf{u}_i, \boldsymbol{\phi}_i) \in \mathbb{R}^{1\times(11+c_f)}$.

3.5. Irregular-to-Regular Conversion

The irregular-to-regular conversion (IRC) module realizes the conversion of pixel values and structure information from irregular coordinates to regular grid, thereby facilitating subsequent processing with well-established CNN-based modules. Traditional filtering-based methods [33, 42, 45] obtain the pixel value $\mathbf{v}_t$ at an arbitrary coordinate (x_t, y_t) by weighted averaging its surrounding K' known pixels $\{\mathbf{p}_t^d = (x_t^d, y_t^d, \mathbf{v}_t^d)\}_{d=1}^{\mathrm{K}'}$, which can be formulated as:

$$\mathbf{v}_t = \sum_{d=1}^{\mathrm{K}'} \omega(x_t, y_t, \mathbf{p}_t^d) \cdot \mathbf{v}_t^d, \tag{5}$$

where $\omega(\cdot)$ is a weighted function, $\{x_t, y_t, \mathbf{p}_t^d\}$ indicates that the weight is related to $x_t, y_t, \mathbf{p}_t^d$ and their derivatives.

Building on this idea, we propose to generate pixel value and image structure feature on each regular grid coordinate $\mathbf{r}_j, j \in \{1, 2, ..., \mathrm{M}\}$ by adaptively fusing the information from its K' nearest neighboring irregular pixels $\{\mathrm{S}_{\mathbf{r}_j}^n\}_{n=1}^{\mathrm{K}'}$. Considering the diverse local structures of irregular pixels, we propose to learn content-dependent weight (such as $\omega(\cdot)$ in Eq. (5)) for each $\mathrm{S}_{\mathbf{r}_j}^n$ to adaptively integrate information. We define a vector $\boldsymbol{\gamma}_{\mathbf{r}_j}^n$ which contains the factors that influence the weight for $\mathrm{S}_{\mathbf{r}_j}^n$ during fusion:

$$\boldsymbol{\gamma}_{\mathbf{r}_j}^n = \left(\mathbf{h}_{\mathbf{r}_j}^n, \mathbf{r}_j, d(\mathrm{S}_{\mathbf{r}_j}^n, \mathbf{r}_j)\right) \in \mathbb{R}^{1\times(14+c_f)}, \tag{6}$$

where $d(\mathrm{S}_{\mathbf{r}_j}^n, \mathbf{r}_j)$ is the distance between $\mathrm{S}_{\mathbf{r}_j}^n$ and $\mathbf{r}_j$. The scheme of IRC module is shown in Fig. 4, and is formulated as:

$$\mathbf{v}_{\mathbf{r}_j} = \sum_{n=1}^{\mathrm{K}'} \left(\mathcal{H}_v(\boldsymbol{\gamma}_{\mathbf{r}_j}^n) \odot \mathbf{v}_{\mathbf{r}_j}^n\right) \in \mathbb{R}^{1\times 3}, \tag{7}$$

$$\boldsymbol{\chi}_{\mathbf{r}_j} = \sum_{n=1}^{\mathrm{K}'} (\mathcal{H}_f(\boldsymbol{\gamma}_{\mathbf{r}_j}^n) \odot \boldsymbol{\phi}_{\mathbf{r}_j}^n) \in \mathbb{R}^{1\times c_f}, \tag{8}$$

where $\mathcal{H}_v(\cdot)$, $\mathcal{H}_f(\cdot)$ are weight generators for pixel values and features, respectively. The weights $\mathcal{H}_v(\boldsymbol{\gamma}_{\mathbf{r}_j}^n)$ and $\mathcal{H}_f(\boldsymbol{\gamma}_{\mathbf{r}_j}^n)$ are both derived from the vector $\boldsymbol{\gamma}_{\mathbf{r}_j}^n$, and their channel dimensions match those of $\mathbf{v}_{\mathbf{r}_j}^n \in \mathbb{R}^{1\times 3}$ and $\boldsymbol{\phi}_{\mathbf{r}_j}^n \in \mathbb{R}^{1\times c_f}$, respectively. $\odot$ denotes element-wise product. Note that the pixel value weights $\{\mathcal{H}_v(\boldsymbol{\gamma}_{\mathbf{r}_j}^n)\}_{n=1}^{\mathrm{K}'}$ are processed through a softmax function (omitted in Eq. (7) and Fig. 4) before being applied as weights. The proposed dual branch content-dependent weight generator $\{\mathcal{H}_v(\cdot), \mathcal{H}_f(\cdot)\}$ enables adaptive aggregation of pixel values and features separately. The output of IRC are regular grid features $\mathbf{Z} = \{\mathbf{z}_{\mathbf{r}_j}\}_{j=1}^{\mathrm{M}} \in \mathbb{R}^{\mathrm{h}\times\mathrm{w}\times(3+c_f)}$, where $\mathbf{z}_{\mathbf{r}_j} = (\mathbf{v}_{\mathbf{r}_j}, \boldsymbol{\chi}_{\mathbf{r}_j}) \in \mathbb{R}^{1\times(3+c_f)}$, $\mathrm{M} = \mathrm{h}\times\mathrm{w}$.

Then, $\mathbf{Z}$ is passed through an encoder $E(\cdot)$ to derive deeper structure features and generate the latent code $\mathbf{L} = E(\mathbf{Z}) \in \mathbb{R}^{\mathrm{h}\times\mathrm{w}\times c_l}$. In implementation, we employ EDSR-baseline [28] as the encoder.

3.6. Continuous Image Representation

To generate an HR image, we construct a general function model driven by the extracted latent code to derive the continuous image function $\mathcal{I}$. Specifically, a decoding function f_θ parameterized by MLP takes a 2D relative coordinate and a neighboring latent vector as input and outputs the pixel value at queried coordinate. Specifically, the pixel value of continuous image $\mathcal{I}$ at coordinate $\mathbf{q}$ is derived by:

$$\mathcal{I}(\mathbf{q}) = f_\theta(\mathbf{L}_{\mathbf{q}^*}, \mathbf{q} - \mathbf{q}^*), \tag{9}$$

where $\mathbf{L}_{\mathbf{q}^*}$ is the nearest features from $\mathbf{q}$ in $\mathbf{L}$, $\mathbf{q}^*$ is the coordinate of feature $\mathbf{L}_{\mathbf{q}^*}$ in 2D image domain, f_θ is shared by all images. Through the implicit neural represenatation, we could generate high resolution images $\mathcal{I}$ of arbitrary sizes. In implementation, we employ the classic LIIF [6] for continuous image representation.

4. Experiments

4.1. Experimental Settings

Dataset. Following LIIF [6], we use DIV2K dataset [2] with 800 images for training. For testing, we use DIV2K validation set [2] and four benchmark datasets: Set5 [4], Set14 [52], B100 [30], and Urban100 [18].

Implementation Details. The proposed ISP2HRNet is trained end-to-end. Let B denotes the batch size, we first randomly sample B sampling ratios $\{r_i\}_{i=1}^B$ from the uniform distribution $\mathcal{U}(0.0625, 1)$, then we crop B patches with sizes $\{h_i \times w_i\}_{i=1}^B$ from the training images, where $h_i = w_i = round(48/\sqrt{r_i})$. For each cropped patch, we randomly sample 2304 coordinates $\{(x_q, y_q)\}_{q=1}^{2304}$, where x_q and y_q are independently sampled from $\mathcal{U}(1-0.5, w_i + 0.5)$. The pixel values at 2304 irregular coordinates are obtained by bicubic interpolation [20]. We use $\mathrm{N} = 2304$ irregularly sampled pixels to form a list as one of the batch input for ISP2HRNet.

Ratio	Method	Set5 [4]		Set14 [52]		B100 [30]		Urban100[18]		DIV2K Val [2]	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
0.2	Partial Conv [29]	27.42	0.8256	24.86	0.7481	24.50	0.7193	23.13	0.7628	27.04	0.7998
	RFR [24]	28.16	0.8558	25.44	0.7805	25.38	0.7664	23.39	0.7899	27.77	0.8343
	MISF [26]	27.11	0.7926	23.03	0.6840	22.22	0.6564	20.41	0.6622	21.83	0.6733
	DPDNN [12]	29.46	0.8778	26.60	0.8066	26.55	0.7909	24.66	0.8175	29.15	0.8583
	DPIR [56]	29.70	0.8911	26.90	0.8246	26.60	0.8037	25.89	0.8625	29.51	0.8781
	RRSNet [13]	29.44	0.8810	26.62	0.8115	26.50	0.7930	24.55	0.8187	29.15	0.8637
	DiffPIR [59]	28.94	0.8661	26.66	0.8050	26.07	0.7707	26.34	0.8552	28.90	0.8482
	CODE [57]	27.85	0.8564	24.94	0.7698	24.92	0.7429	23.23	0.7845	27.79	0.8383
	Ours	30.94	0.9059	27.88	0.8411	27.44	0.8223	27.43	0.8864	30.67	0.8932
0.4	Partial Conv [29]	30.79	0.9027	27.95	0.8553	27.36	0.8390	26.26	0.8676	30.21	0.8876
	RFR [24]	31.99	0.9215	28.75	0.8746	28.49	0.8728	26.92	0.8903	31.27	0.9086
	MISF [26]	29.97	0.8519	25.09	0.7772	24.04	0.7665	22.25	0.7483	23.21	0.7400
	DPDNN [12]	32.99	0.9310	29.73	0.8913	29.56	0.8892	27.83	0.9010	32.42	0.9237
	DPIR [56]	33.60	0.9418	30.47	0.9038	29.94	0.9020	30.10	0.9358	33.43	0.9416
	RRSNet [13]	33.16	0.9356	29.85	0.8957	29.56	0.8914	27.80	0.9035	32.56	0.9292
	DiffPIR [59]	32.69	0.9237	30.25	0.8917	29.09	0.8705	30.24	0.9235	32.35	0.9150
	CODE [57]	32.49	0.9336	28.62	0.8819	28.22	0.8680	27.27	0.8997	31.77	0.9230
	Ours	34.56	0.9477	31.48	0.9151	30.58	0.9077	31.40	0.9455	34.44	0.9480
0.6	Partial Conv [29]	33.88	0.9470	30.90	0.9204	30.27	0.9136	29.21	0.9281	33.32	0.9402
	RFR [24]	35.44	0.9563	31.88	0.9288	31.68	0.9343	30.20	0.9423	34.63	0.9487
	MISF [26]	32.58	0.8882	27.07	0.8356	25.90	0.8343	24.05	0.8028	24.79	0.7863
	DPDNN [12]	36.06	0.9600	32.68	0.9381	32.59	0.9427	30.85	0.9460	35.53	0.9588
	DPIR [56]	36.96	0.9676	33.63	0.9458	33.15	0.9508	33.29	0.9660	36.91	0.9708
	RRSNet [13]	36.38	0.9638	32.90	0.9410	32.65	0.9448	30.88	0.9483	35.83	0.9633
	DiffPIR [59]	35.47	0.9519	33.20	0.9362	31.85	0.9254	33.22	0.9540	35.16	0.9481
	CODE [57]	36.58	0.9657	32.30	0.9394	31.77	0.9368	31.25	0.9533	35.73	0.9638
	Ours	37.80	0.9710	34.73	0.9538	33.71	0.9534	34.72	0.9719	37.90	0.9740
0.8	Partial Conv [29]	37.82	0.9771	34.77	0.9657	34.12	0.9631	33.07	0.9690	37.35	0.9750
	RFR [24]	39.77	0.9805	36.07	0.9683	36.08	0.9744	34.45	0.9762	39.06	0.9748
	MISF [26]	36.05	0.9232	29.97	0.8926	28.76	0.8957	26.83	0.8620	27.43	0.8438
	DPDNN [12]	40.01	0.9819	36.65	0.9721	36.76	0.9772	34.92	0.9772	39.74	0.9828
	DPIR [56]	41.14	0.9857	37.73	0.9759	37.41	0.9807	37.40	0.9857	41.32	0.9884
	RRSNet [13]	40.47	0.9835	36.92	0.9726	36.85	0.9784	34.97	0.9783	40.17	0.9851
	DiffPIR [59]	38.54	0.9718	36.52	0.9664	35.21	0.9618	36.50	0.9739	38.23	0.9695
	CODE [57]	41.28	0.9851	36.98	0.9746	36.56	0.9771	36.18	0.9825	40.74	0.9865
	Ours	41.93	0.9871	38.80	0.9797	37.88	0.9815	38.79	0.9882	42.19	0.9894

Table 1. Quantitative results of reconstruction from incomplete image with random missing pixels at different sampling ratios. Best and second best performance are in red and blue colors, respectively.

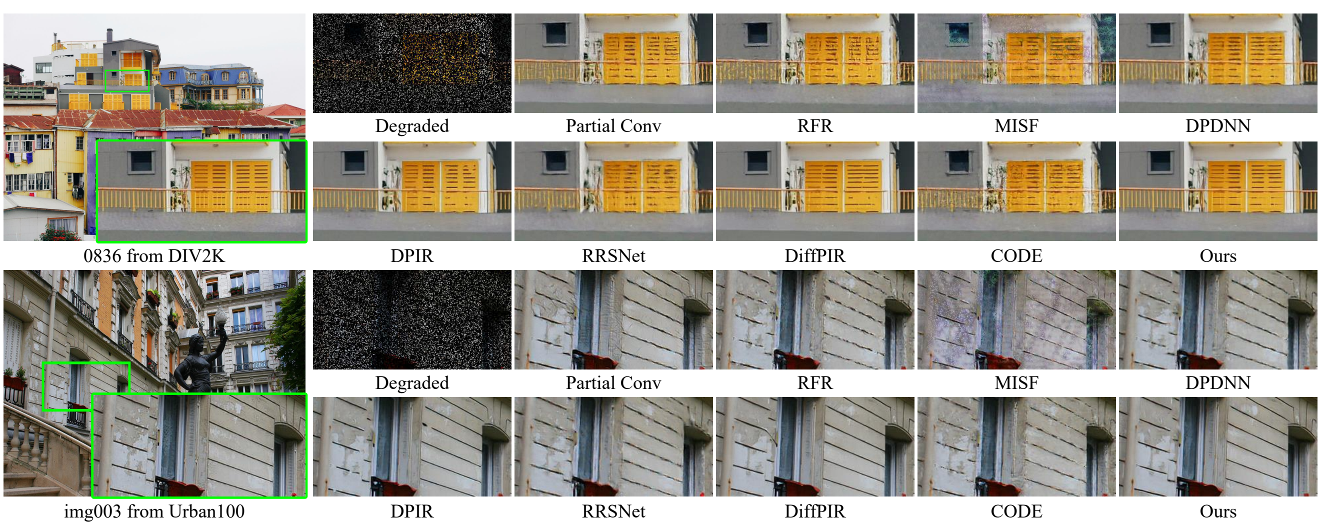


Figure 5. Visual results of reconstruction from incomplete image with random missing pixels. The sampling ratio is 0.2. Please enlarge the figure for better comparison.

Dataset	Scale	DPIR+ BI	DPIR+ EDSR^b	DPIR+ EDSR	DPIR+ SwinIR	DPIR+ LIIF	Ours
Set5	×1.5	31.43	-	-	-	31.91	32.62
	×2	29.15	29.84	29.84	29.84	29.77	30.35
	×2.5	27.35	-	-	-	27.85	28.59
	×3*	26.23	26.79	26.79	26.80	26.74	27.27
	×4*	24.39	24.77	24.78	24.78	24.74	25.27
Set14	×1.5	28.21	-	-	-	28.81	29.24
	×2	26.26	26.80	26.82	26.81	26.79	27.15
	×2.5	25.04	-	-	-	25.57	25.82
	×3*	24.11	24.61	24.59	24.60	24.58	24.85
	×4*	22.82	23.29	23.29	23.28	23.26	23.48
B100	×1.5	27.91	-	-	-	28.45	28.62
	×2	26.19	26.73	26.73	26.73	26.70	26.77
	×2.5	25.16	-	-	-	25.61	25.65
	×3*	24.42	24.86	24.86	24.87	24.83	24.87
	×4*	23.40	23.75	23.76	23.76	23.74	23.78
Urban100	×1.5	25.65	-	-	-	27.89	27.85
	×2	24.16	25.61	25.69	25.69	25.61	25.34
	×2.5	22.33	-	-	-	22.73	23.88
	×3*	22.06	23.11	23.15	23.16	23.09	22.87
	×4*	20.86	21.68	21.71	21.74	21.68	21.53
DIV2K Val	×1.5	30.80	-	-	-	31.45	31.82
	×2	28.76	29.50	29.52	29.52	29.43	29.58
	×2.5	27.43	-	-	-	28.02	28.17
	×3*	26.52	27.13	27.14	27.14	27.08	27.16
	×4*	25.20	25.72	25.73	25.74	25.69	25.76

Table 2. Quantitative results of super resolution from incomplete image with random missing pixels. Best performance is in red color. (BI: Bicubic interpolation, EDSR^b: EDSR-baseline, the encoder in LIIF is the EDSR-baseline, * indicates the equivalent sampling ratio is out of our model's training distribution.)

We set $\mathrm{K} = 8$ and $\mathrm{K}' = 12$. For GSE, we use a 3-layer MLP with batch normalization (BN) and ReLU activation [34], with hidden dimensions set to [64, 64, 128], where the output feature dimension c_f is 128. For IRC, the pixel value weight generator $\mathcal{H}_v(\cdot)$ is implemented by a 2-layer MLP with BN, ReLU and hidden dimensions of [128, 3]. The feature weight generator $\mathcal{H}_f(\cdot)$ follows the same structure but with hidden dimensions of [128, 128]. The latent code dimension c_l is 256. IRC module outputs $\mathrm{M} = 2304$ regular grid features with a spatial size of 48×48, i.e., $\mathrm{h} = \mathrm{w} = 48$. Following LIIF [6], we adopt $\mathcal{L}_1$ loss and the same loss calculation strategy during training. We use Adam [21] optimizer. The learning rate is initialized as 1e-4 and decayed by 0.5 every 200 epochs. The network is trained for 1000 epochs with batch size $B = 16$. All models are trained and tested on NVIDIA RTX 3090 GPU.

Notably, the trained ISP2HRNet is directly applicable to the tasks described in Secs. 4.2 to 4.4, without the need for separate training for each task. All evaluations are performed in the RGB color space. Moreover, ISP2HRNet can handle arbitrary number (N) of irregular pixels as input. The inference details are provided in the supplementary material.

We first conduct experiments on special cases where the coordinates of irregularly sampled pixels are integers, facilitating comparison with existing method (Sec. 4.2 and Sec. 4.3). Then, we evaluate general cases where the sampled coordinates are decimals (Sec. 4.4). For all tasks (Secs. 4.2 to 4.4), ISP2HRNet organizes irregular pixels into a list as input for inference.

4.2. Reconstruction from Incomplete Image with Random Missing Pixels

Reconstruction from incomplete image with random missing pixels aims to restore an image degraded by a random sampling mask, where the sampled positions on the mask are set to 1, and other positions are set to 0. The sampled positions are random, irregular, and integers, which is a special case of irregular sampling. We conduct experiments on degraded HR images from five test datasets.

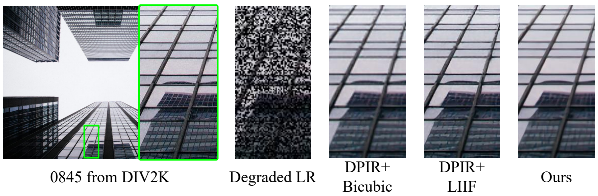
We compare the proposed ISP2HRNet with the methods designed for image inpainting: partial convolutions [29], RFR [24], MISF [26], optimization-inspired networks for image restoration: DPDNN [12], DPIR [56], RRSNet [13], DiffPIR [59], and a transformer-based restoration network: CODE [57]. We use PSNR and SSIM [49] as evaluation metrics.

Tab. 1 presents the quantitative results. We compare the reconstruction performance at sampling ratios of 0.2, 0.4, 0.6, and 0.8. The proposed method achieves the best performance on five test datasets at different sampling ratios. In particular, the proposed ISP2HRNet outperforms the compared methods by more than 1 dB in some cases. Visual results are provided in Fig. 5. The proposed ISP2HRNet reconstructs clear textures and high-frequency details with a low sampling ratio of 0.2, while the compared methods suffer from visually annoying artifacts and image distortion. More details and visual results are provided in the supplementary material.

4.3. Super Resolution from Incomplete Image with Random Missing Pixels

The problem is defined as follows: given an incomplete LR image with random missing pixels, the goal is to restore its corresponding HR image. We employ a straightforward idea to combine the image restoration method: DPIR [56] with image SR methods: Bicubic Interpolation (BI), EDSR-baseline [28], EDSR [28], SwinIR [27], and LIIF [6] to form baselines for comparison. The sampling ratio is set to 0.5 in LR images. The proposed ISP2HRNet and LIIF could handle different up-sampling scales with a single model, while the classic image SR models: EDSR-baseline [28], EDSR [28] and SwinIR [27] handle different SR scales with different models.

Tab. 2 shows the quantitative comparison between the proposed method and baselines. It is noted that for up-sampling scales ×3, ×4 with sampling ratio 0.5 in LR images, the equivalent sampling ratios for our model are 0.056 (0.5/9) and 0.031 (0.5/16), respectively, which are unseen



(a) The upsampling scale is $\times 2.5$ and sampling ratio is 0.5 in LR image.

(b) The upsampling scale is $\times 4$ and sampling ratio is 0.5 in LR image (The equivalent sampling ratio for our model is 0.031, which is out of our model's training distribution).

Figure 6. Visual results of super resolution from incomplete image with random missing pixels. Please enlarge the figure for better comparison.

Dataset	$r = 0.2$		$r = 0.4$		$r = 0.6$		$r = 0.8$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	30.92	0.9047	34.00	0.9426	36.21	0.9606	37.93	0.9713
Set14	27.92	0.8406	30.99	0.9071	33.18	0.9377	34.93	0.9555
B100	27.42	0.8207	30.24	0.9012	32.46	0.9396	34.38	0.9610
Urban100	27.36	0.8833	30.67	0.9373	32.85	0.9600	34.52	0.9723
DIV2K Val	30.55	0.8912	33.82	0.9421	36.18	0.9643	38.07	0.9761

Table 3. Quantitative results of reconstruction from irregularly sampled pixels.

ratios during training. The proposed method achieves the highest PSNR on most settings. For sampling ratios that are out of training distribution (scales $\times 3$, $\times 4$), our method still performs well, demonstrating its ability to generalize to much lower sampling ratios. The visual comparison is presented in Fig. 6. The proposed method produces visually pleasing images under low sampling ratios, while the compared methods exhibit blurriness or artifacts. More visual results are provided in the supplementary material.

4.4. Reconstruction from Irregularly Sampled Pixels

We conduct experiments on reconstruction from irregularly sampled pixels that do not fall on a regular grid, i.e., their coordinates are decimal values and cannot be easily represented by a 2D image array with unknown pixels set to zero. For each HR image in the test datasets, we first calculate the number of sampled pixels $\mathrm{N} = H \times W \times r$ based on the size of HR image $H \times W$ and the sampling ratio r. Then, we randomly sample N coordinates (decimals) within the image range and obtain their corresponding pixel values by bicubic interpolation. We organize these randomly sampled pixels into a list as input to ISP2HRNet for image reconstruction. The quantitative results are presented in Tab. 3. More discussion and visual results are provided in the supplementary material.

Case	GD	GSE	IRC	$r = 0.2$		$r = 0.4$	
				PSNR	SSIM	PSNR	SSIM
(a)				22.60	0.7667	27.22	0.8967
(b)			✓	26.11	0.8615	30.02	0.9338
(c)	✓		✓	26.49	0.8693	30.53	0.9385
(d)	✓	✓		26.25	0.8636	30.25	0.9358
(e)	✓	✓	$*$	26.92	0.8761	30.94	0.9412
(f)	✓	✓	✓	**27.43**	**0.8864**	**31.40**	**0.9455**

Table 4. Ablation study of the proposed ISP2HRNet.

4.5. Ablation Study

The ablation study is performed based on the task: reconstruction from incomplete image with random missing pixels (Sec. 4.2), and we evaluate the performance using the Urban100 dataset (sampling ratio $r = 0.2, 0.4$).

We remove some designs in ISP2HRNet to validate the effectiveness of each proposed component, as shown in Tab. 4. Case (a) removes GD, GSE and IRC modules from ISP2HRNet, and uses radial basis function interpolation to reconstruct a regular image array as input in order to handle irregular pixels. In case (b), irregular pixels are directly fed into IRC module without extracting gradient structure information. In case (c), irregular pixels with gradient estimation are fed into IRC module. In case (d), we use distance-based Gaussian weighting to generate regular grid features. In case (e), we utilize a single branch weight generator, which generates weights for pixel values and features simultaneously. Case (f) is the final model, i.e., the proposed ISP2HRNet. Cases (b) and (c) explore the impact of GD module. Cases (c) and (f) highlight the contribution of GSE module. Cases (d), (e), and (f) validate the effectiveness of the dual branch content-dependent weight generator proposed in IRC module. Cases (a) and (f) demonstrate the combined effectiveness of GD, GSE, and IRC modules in ISP2HRNet.

5. Conclusion

We propose an end-to-end network named ISP2HRNet to reconstruct high resolution image from irregularly sampled pixels that do not fall on a regular grid, which incorporates the key concepts from Taylor expansion. Specifically, we propose an architecture that extracts image structure information from irregular pixels and regular grid hierarchically and learns continuous image representation. Our work extends high resolution image reconstruction from regular pixel array to irregularly sampled pixels. Experimental results demonstrate that the proposed ISP2HRNet achieves promising results.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 22127807.

References

[1] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, 2010. 2

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 126–135, 2017. 5, 6

[3] Carla Bertocchi, Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet, and Marco Prato. Deep unfolding of a proximal interior point method for image restoration. *Inverse Problems*, 36(3):034005, 2020. 1, 2

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie line Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 135.1–135.10. BMVA Press, 2012. 5, 6

[5] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image processing*, 16 (12):2992–3004, 2007. 2

[6] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8628–8638, 2021. 2, 5, 7

[7] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12312–12321, 2023. 2

[8] Jisheng Chu, Wenrui Li, Xingtao Wang, Kanglin Ning, Yidan Lu, and Xiaopeng Fan. Digging into intrinsic contextual information for high-fidelity 3d point cloud completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(3):2573–2581, 2025. 2

[9] Carl de Boor. *A Practical Guide to Splines*. 1978. 1, 2

[10] Daizong Ding, Erling Jiang, Yuanmin Huang, Mi Zhang, Wenxuan Li, and Min Yang. CAP: Robust point cloud classification via semantic and structural modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12260–12270, 2023. 2

[11] Weisheng Dong, Guangming Shi, and Xin Li. Nonlocal image restoration with bilateral variance estimation: A low-rank approach. *IEEE Transactions on Image Processing*, 22 (2):700–711, 2012. 2

[12] Weisheng Dong, Peiyao Wang, Wotao Yin, Guangming Shi, Fangfang Wu, and Xiaotong Lu. Denoising prior driven deep neural network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2305–2318, 2018. 1, 2, 6, 7

[13] Yanchen Dong, Rui Zhao, Ruiqin Xiong, Shuyuan Zhu, Xiaopeng Fan, and Tiejun Huang. Optimization-inspired deep network for image restoration from partial random samples. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2023. 1, 2, 6, 7

[14] D.S. Early and D.G. Long. Image reconstruction and enhanced resolution imaging from irregular samples. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2): 291–302, 2001. 1

[15] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas Funkhouser. Local deep implicit functions for 3d shape. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4857–4866, 2020. 2

[16] Wolfgang Härdle, Marlene Müller, Stefan Sperlich, Axel Werwatz, et al. *Nonparametric and semiparametric models*. Springer, 2004. 1, 2

[17] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-SR: A magnification-arbitrary network for super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1575–1584, 2019. 2

[18] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. 5, 6

[19] Xuan Ju, Xian Liu, Xintao Wang, Yuxuan Bian, Ying Shan, and Qiang Xu. BrushNet: A plug-and-play image inpainting model with decomposed dual-branch diffusion. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 150–168, 2024. 2

[20] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, 1981. 5

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[22] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1929–1938, 2022. 2

[23] S. Lee, G. Wolberg, and S.Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997. 1, 2

[24] Jingyuan Li, Ning Wang, Lefei Zhang, Bo Du, and Dacheng Tao. Recurrent feature reasoning for image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 6, 7

[25] Wenrui Li, Wei Han, Yandu Chen, Yeyu Chai, Yidan Lu, Xingtao Wang, and Xiaopeng Fan. Riemann-based multi-scale attention reasoning network for text-3d retrieval. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(17):18485–18493, 2025. 2

[26] Xiaoguang Li, Qing Guo, Di Lin, Ping Li, Wei Feng, and Song Wnag. MISF: Multi-level interactive siamese filtering for high-fidelity image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2, 6, 7

[27] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1833–1844, 2021. 2, 7

[28] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 2, 5, 7

[29] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018. 2, 6, 7

[30] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV)*, pages 416–423. IEEE, 2001. 5, 6

[31] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4460–4470, 2019. 2

[32] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2

[33] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964. 1, 2, 5

[34] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010. 7

[35] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3504–3515, 2020. 2

[36] Pan Pan and Dan Schonfeld. Image reconstruction and multidimensional field estimation from randomly scattered sensors. *IEEE Transactions on Image Processing*, 17(1):94–99, 2008. 1

[37] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019. 2

[38] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, 2017. 2

[39] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017. 2

[40] Haoxi Ran, Wei Zhuo, Jun Liu, and Li Lu. Learning inner-group relations on point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15477–15487, 2021. 2

[41] Jürgen Seiler, Markus Jonscher, Michael Schöberl, and André Kaup. Resampling images to a regular grid from a non-regular subset of pixel positions using frequency selective reconstruction. *IEEE Transactions on Image Processing*, 24(11):4540–4555, 2015. 1

[42] Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524, 1968. 1, 2, 5

[43] Gaochao Song, Qian Sun, Luo Zhang, Ran Su, Jianfeng Shi, and Ying He. OPE-SR: Orthogonal position encoding for designing a parameter-free upsampling module in arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10009–10020, 2023. 2

[44] Agnieszka Barbara Szczotka, Dzhoshkun Ismail Shakir, Daniele Ravì, Matthew J Clarkson, Stephen P Pereira, and Tom Vercauteren. Learning from irregularly sampled data for endomicroscopy super-resolution: a comparative study of sparse and dense approaches. *International Journal of Computer Assisted Radiology and Surgery*, 15:1167–1175, 2020. 1

[45] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on Image Processing*, 16(2):349–366, 2007. 1, 2, 3, 5

[46] Michael Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6):22–38, 1999. 1, 2

[47] Tom Vercauteren, Aymeric Perchant, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical Image Analysis*, 10 (5):673–692, 2006. 1

[48] Longguang Wang, Yingqian Wang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo. Learning a single network for scale-arbitrary super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4801–4810, 2021. 2

[49] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 7

[50] Zeyu Xiao and Zhiwei Xiong. Incorporating degradation estimation in light field spatial super-resolution. *Computer Vision and Image Understanding*, 252:104295, 2025. 2

[51] Xingqian Xu, Zhangyang Wang, and Humphrey Shi. UltraSR: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution. *arXiv preprint arXiv:2103.12716*, 2021. 2

[52] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Proceedings of the 7th International Conference on Curves and Surfaces*, page 711–730, Berlin, Heidelberg, 2010. Springer-Verlag. 5, 6

[53] Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1828–1837, 2018. 2

[54] Jian Zhang, Ruiqin Xiong, Siwei Ma, and Debin Zhao. High-quality image restoration from partial random samples in spatial domain. In *2011 Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2011. 2

[55] Jian Zhang, Debin Zhao, Ruiqin Xiong, Siwei Ma, and Wen Gao. Image restoration using joint statistical modeling in a space-transform domain. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(6):915–928, 2014. 2

[56] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021. 6, 7

[57] Haiyu Zhao, Yuanbiao Gou, Boyun Li, Dezhong Peng, Jiancheng Lv, and Xi Peng. Comprehensive and delicate: An efficient transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14122–14132, 2023. 6, 7

[58] Yaoqian Zhao, Qizhi Teng, Honggang Chen, Shujiang Zhang, Xiaohai He, Yi Li, and Ray E. Sheriff. Activating more information in arbitrary-scale image super-resolution. *IEEE Transactions on Multimedia*, 26:7946–7961, 2024. 2

[59] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 6, 7