

Research Article

Research on Super-Resolution Relationship Extraction and Reconstruction Methods for Images Based on Multimodal Graph Convolutional Networks

Jie Xiao 

School of Railway Communication and Signaling, Wuhan Railway Vocational College of Technology, Wuhan, Hubei 430205, China

Correspondence should be addressed to Jie Xiao; 20150012@wru.edu.cn

Received 14 July 2022; Revised 5 August 2022; Accepted 18 August 2022; Published 10 September 2022

Academic Editor: Gengxin Sun

Copyright © 2022 Jie Xiao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study constructs a multimodal graph convolutional network model, conducts an in-depth study on image super-resolution relationship extraction and reconstruction methods, and constructs a model of image super-resolution relationship extraction and reconstruction methods based on multimodal graph convolutional networks. In this study, we study the domain adaptation model algorithm based on graph convolutional networks, which constructs a global relevance graph based on all samples using pre-extracted features and performs distribution approximation of sample features in two domains using a diagram convolutional neural network with maximum mean difference loss; with this approach, the model effectively preserves the structural information among the samples. In this study, several comparison experiments are designed based on the COCO and VG datasets; the image space information-based and knowledge graph-based target detection and recognition models substantially improve recognition performance over the baseline model. The super-pixel-based target detection and recognition model can also effectively reduce the number of floating-point operations and the complexity of the model. In this study, we propose a multiscale GAN-based image super-resolution reconstruction algorithm. Aiming at the problems of detail loss or blurring in the reconstruction of detail-rich images by SRGAN, it integrates the idea of the Laplace pyramid to complete the task of multiscale reconstruction of images through staged reconstruction. It incorporates the concept of a discriminative network with patch GAN to effectively improve the recovery effect of graph details and improve the reconstruction quality of images. Using Set5, Set14, BSD100, and Urban100 datasets as test sets, experimental analysis is conducted from objective and subjective evaluation metrics to effectively validate the performance of the improved algorithm proposed in this study.

1. Introduction

With the continuous development of information technology and the popularity of intelligent terminal devices, people's demand for information is also rising: from images in the 2G era to pictures in the 3G era, then to images in the 4G era, and then to holographic images such as AR and VR in the 5G era, the amount of information is rising, while the occupied storage is also exploding [1]. This considerably impacts the daily dissemination of information—the network speed cannot keep up, and the hard disk cannot store it. Therefore, there is an urgent need for an efficient means of information compression to help compress information to

improve transmission efficiency and reduce the storage footprint [2]. With the development of high-performance processors, high-definition screens are becoming more and more popular with the emergence of intelligent devices. However, most media information on the Internet is still dominated by low-definition images, resulting in data quality not keeping up with display quality, thus reducing the user experience [3]. In addition, due to the limitations of image storage hardware, the resolution of images is limited, and the size of the smallest pixel determines the details that can be displayed. But the real world is often infinite, so people also want to much detail as possible in the image they can get. The solutions to the above pain points can be

summarized as the compression and decompression of information. The most immediate and effective solution to reduce the image size is to store and disseminate multimedia information, especially the most informative image information, and reduce the image's resolution.

Super-resolution reconstruction (SR) technique is to reconstruct a single or multiframe low-resolution (LR) image into a high-resolution (HR) photo by applying specific image processing and other methods to achieve high-quality images. Usually, CNN-based target detection and recognition models use sliding windows or anchors to extract possible foregrounds and hardgrounds [4]. Then, the final localization frame is generated by identifying and regressing all possible foregrounds. By relying on the graph convolution network, we can obtain more abundant information about the location of the object picture. For example, by relying on the inference of the spatial map, we can roughly determine the object's position and then fine-tune it according to work. Therefore, we can design more flexible and efficient positioning methods to generate positioning frames [5]. This study develops several graph convolutional network detection models acting in spatial, beyond-pixel, and knowledge graphs. We extract features beyond pixels to assist pixel information for accurate target detection and recognition. Finally, the experimental and comparative analyses of the model on the COCO dataset and VG dataset prove that the target detection and recognition model based on a graph convolutional network can break the bottleneck of image pixel recognition to a certain extent and help the target image to achieve better object recognition and localization [6].

Image super resolution is used to solve the problem of recovering low-resolution images to high-resolution images. Image super resolution aims to up-sample a series of low-resolution photos output by a deterministic or uncertain degradation model to high resolution while providing more detail than low resolution [7]. Traditional upsampling algorithms have a solid prior relationship, considering that there is a specific mathematical relationship between neighboring pixel values so that the original pixels can be recovered by interpolating adjacent pixels. In the forward propagation process, each sample feature is transformed independently, which may lead to the separation of target domain features that are initially in the same class under the influence of the distribution difference function and eventually classified into different categories [8]. It enables problems with unstructured relationships, such as citation networks, to be well trained by importing correlation graphs between samples [9]. This feature also helps to compensate for the shortcomings of existing domain adaptation algorithms. Graph convolutional networks can be considered as a particular case of graph networks. In this study, we intend to study the scheme and practice of introducing chart convolutional networks into domain adaptation problems to improve the learning performance of domain adaptation problems, make a new direction to explore migration learning tasks, and provide a feasible solution for learning scenarios where labeled information is challenging to obtain

[10]. The study of domain adaptation algorithms can effectively reduce the need for data annotation and enable various algorithmic models to have fast learning performance for similar tasks and improve their generalization and robustness, which is of great significance in various real-world tasks, where annotation information is not readily available.

2. Related Works

The graph convolution layer is a simple extension of the fully connected layer that integrates valuable information from the knowledge graph into the feature vector, and the intuitive understanding of the graph convolution layer is simple. By importing a relevance graph (knowledge graph) into the neural network, the graph convolution layer can change the distribution of the feature vectors through the relevance variable of the relevance graph so that the relevant samples are closer to each other [11]. This feature facilitates the data to obtain and maintain useful structural information during the distribution approximation process, thus avoiding the loss of similar structures in the source domain caused by migration learning and improving the network performance. Some scholars have already researched migration learning using relevance graphs and convolution layers. When using local relevance graphs obtained by random sampling, neighboring samples may not be sampled simultaneously, making the graph convolution performance degrade. Altinkaya et al. first identified a few pieces by random sampling, they then added both the first-order and second-order neighbors of these samples to the set to be selected before selection, and the sampled set was guaranteed to correlate with the models [12]. Chadha et al. interpret graph convolution as an integral transformation of the embedding function under probability measures and use Monte Carlo methods to estimate the critical values [13]. They propose an important sampling method, in which the sum of the relevance weight values of each sample to other samples is used as the sampling weight. The above sampling is performed once in each graph convolutional layer; good results are obtained in the referenced network dataset.

Among the reconstruction-based methods, projection onto convex sets (POCS) is proposed by Hong et al. This algorithm is based on the set projection theory of mathematical sets and can converge relatively quickly [14]. The iterative back-projection (IBP) method proposed by Kocsis et al. projects the error value between the input low-resolution image and the low-resolution image obtained from the degradation model backward onto the corresponding high-resolution large print, and the error converges continuously to reconstruct the sizeable high-resolution image [15]. Yanshan et al. proposed the maximum a posteriori probability (MAP) algorithm, which solves the image super-resolution reconstruction by probabilistic estimation in mathematics, the prerequisite is the low-resolution image sequence, and the goal of the algorithm is to obtain the maximum a posteriori probability to reconstruct the sizeable high-resolution image [16]. Chen et al. proposed the

neighborhood embedding method, which first maps the local geometric information of the low-resolution image block to the corresponding high-resolution photo and then uses the linear combination to map the neighborhood to produce the high-resolution image block [17]. Many subsequent researchers have made optimization improvements to the neighborhood embedding-based method. The super-resolution algorithms have been explored around how to recover more OK texture information and edge details based on higher super-resolution magnification. Although traditional methods have low complexity, it is not easy to make a significant breakthrough in super-resolution reconstruction quality and visual effect [18]. Deep learning methods require a large amount of training data compared with traditional learning-based methods. Still, they can recover more full image details and texture information by using neural networks' powerful feature representation capability to learn the complex mapping relationships between low- and high-resolution images [19]. In recent years, many results have emerged in the field of deep learning and achieved better performance and performance compared with traditional algorithms, especially the introduction of a new and more challenging generative model: generative adversarial networks, which opens a new world in the field of image super-resolution-based research.

The multi-image super-resolution task is also known as the image super-resolution task. The significant difference between a multi-image super-resolution task and a single-image super-resolution task is that the single-image super-resolution task mainly models the image scene and the mapping between pixel distributions by learning a priori knowledge from the training data and inferring the pixel distribution of the image after super resolution by the pixel distribution of the target image [20]. The information that the model can ingest is the pixel mapping learned from the training data; when the pixel distribution of the test image does not appear in the training image, it will lead to significant degradation of the image's super-resolution quality [21]. In the case of multi-image image super-resolution tasks, or image super-resolution tasks, additional information about the before and after frames of the image is introduced. From common sense, the data between photos in consecutive image frames are continuous and gradual, and it is entirely possible to use such an incremental information mechanism to extract the information that was discarded during the downsampling of the target image in the adjacent frames of the image to recover the target image after downsampling [22]. The convolutional graph networks are highly vulnerable to adversarial attacks, which makes their prospects for industrial applications challenging. Combining graph convolutional networks with target detection and recognition is difficult, as graph convolutional networks can obtain certain features based on the graph structure. However, there is still no fixed solution for using these features to complement or identify localized targets. Finally, as more and more graph convolutional networks are designed, selecting a suitable network based on the graph structure characteristics is also a significant issue.

3. Model Design of Super-Resolution Relationship Extraction and Reconstruction Method for Images Based on Multimodal Graph Convolutional Networks

3.1. Multimodal Graph Convolutional Network Model Construction. Convolutional operations can extract structural features of structured data by using convolutional kernels with shared parameters. Single-modality image alignment refers to the floating of two images acquired with the same imaging device. It is mainly applied to the alignment between different MRI-weighted images and the alignment of image sequences, etc. Multimodal image alignment refers to the floating of two images from other imaging devices. Increasing the number of convolutional kernels can obtain multidimensional structural features to characterize the data. For unstructured data such as molecular structure and recommendation system, the information cannot be extracted directly by fixed convolutional kernels because they do not have uniformity. Therefore, the graph neural network (GNN), which simulates convolutional operations to remove features efficiently on unstructured data, emerged and continues to evolve. Like convolution on images, the information of each node is extracted by picking the perceptual field [23]. The most direct way is to aggregate the node whose features are to be removed with its neighbor nodes within a fixed number of hops, based on the idea of message passing to extract parts of the graph for subsequent scenarios such as node classification, graph classification, and edge prediction. GCN has been mathematically rigorous in reasoning and proof. Combining spectral convolution and Chebyshev polynomials and simplifying the operation by constraining $k = 1$ to obtain a first-order linear approximation to the graph spectral convolution, an expression for the graph convolution neural network is derived as follows:

$$h_{(l+1)} = \sum \frac{\sigma + h^{l-1} + \sqrt{h^l - w^l}}{\sqrt{\sigma - d + ad}}, \quad (1)$$

where H^l denotes the graph convolution network at layer l ($H^0 = x$); \tilde{D} is the degree matrix ($\tilde{D}_{ii} = \sum \tilde{A}_{ij}$); $\tilde{A} = A + I$ denotes the adjacency matrix introducing its information; W is the training parameter, and σ is the activation function. Therefore, the output of the two-layer graph convolutional network is as follows:

$$z = \sum \text{softmax} \frac{(d - ad - xw^o)w}{\sqrt{d - 1/2ad - \sigma}}. \quad (2)$$

The graph convolution neural network defines the graph convolution operation. It can achieve convolution-like feature extraction on unstructured data, and subsequent research on it is done based on graph convolution.

During node updates, weights are determined based on the interrelationship between neighboring nodes and the current node, thus enhancing the ability to extract meaningful information and attenuating the weight of irrelevant knowledge. Like the graph convolutional neural network,

the graph attention network introduces the calculation of attention. It adds it to the update operation, while the node weight value is determined by its interrelationship with the controller node. The node weights are calculated as shown in the following equation:

$$\alpha_{ij} = \sum \frac{\exp(\sigma - a^t wh_i + a^t wh_j)}{\sigma(a^t [wh_i - wh_k])}. \quad (3)$$

In the above equation, α_{ij} denotes the attention weight of node j with respect to node i ; N_i denotes the set of nodes adjacent to node i ; h_i is the feature of node i ; the attention value α_{ij} denotes the degree of association between nodes, which can be obtained either by learning or by a similarity measure. The attention weights are introduced into the graph convolution process to emphasize the importance of different neighboring nodes to the current node so that the next layer of feature values can be calculated and updated as follows:

$$h_i = \sum \frac{\sqrt{a_{ij} + wh_j}}{\sigma - \sqrt{(j - n_i)}} + \sigma, \quad (4)$$

where h_j is the feature of node j in the current layer of the graph convolution network; h_i is the feature of node i in the next layer of the graph convolution network. The graph attention network quantifies and introduces the relationship between nodes into the graph update process, and this relationship is equivalent to the adjacency matrix in the graph convolution a . Because of its ability to construct adjacency matrices based on node relationships can be applied to graphs without explicit edge concepts, such as graphs describing sample relationships. In essence, the principles of GCN and GAT are similar; the former uses Laplacian matrices and emphasizes the role of graph structure information in graph convolutional networks. At the same time, the latter introduces attention coefficients to enhance the role of correlation information between nodes. The last is suitable for a broader range of scenarios, such as inductive tasks, by calculating each node one by one, free from the strong constraints of the graph structure.

The interaction enhancement between local information includes the interaction between the internal elements of local target information and local image information and the interaction between local target information and local image information. The principle of internal element interaction enhancement is that a subset of elements that are relatively important or create a common theme can be calculated using the interrelationship between the interior features [24]. The principle of interaction enhancement between local target and image information is that both information initially corresponds to the same scene theme, so there is a constraint and guidance between the data. Local target information can guide local image information to make the selection and fusion of a subset of crucial image elements. At the same time, local image information can also locally target information to make the selection and fusion of a subgroup of critical target elements. The graph

convolutional neural network is a prevalent network model. Many algorithms use it as the basis for modeling and solving practical problems, whether in recommendation algorithms, computer vision, or natural language processing. In this study, we need to enhance the interaction and fusion between local information elements, so we design a practical information fusion module based on a graph convolutional network.

First, the graph node feature is defined as $r = \{f_1, f_2, \dots, f_m\}$, $f_i \in r^d$ the feature vector corresponding to the i node and m the number of nodes. The graph network constructed with local target information elements can be represented as the graph network built with local target information elements, which can be defined as $r_o = \{f_{o1}, f_{o2}, \dots, f_{op}\}$. The graph network created with local text information elements can be described as $r_t = \{f_{t1}, f_{t2}, \dots, f_{tp}\}$. The graph network made with both parts together can be represented as $r_{ot} = \{f_1, f_2, \dots, f_{p+q}\}$. The graph convolution operation in this study is defined as follows:

$$\begin{aligned} r_l &= \int r_{l+1} - \frac{w_{r-t}}{h}, \\ h &= \int_{l=1} \frac{w_h - w_t}{\sqrt{m_r - r_l}} \times \frac{w_h - w_t}{\sqrt{m_r - r_l}}, \\ m_r &= \int_{l=1} (r^{l-1} + w_{(t-k)} + r^{l-1} \times w_{t-q}). \end{aligned} \quad (5)$$

This study's multimodal local information interaction module consists of two branches, the independent graph convolution branch and the joint graph convolution branch. The separate graph convolution branch is a graph convolution operation for r_o and r_t respectively, which enables the enhancement of information elements of the other modality through intermodal attention while preserving the information differences between the two different modalities. In contrast, the joint graph convolution branch is a graph convolution operation r_{ot} , enabling the two modal information to automatically learn the interaction model in the same graph network. The design and computation of the two graph convolution branches are described in detail, as shown in Figure 1.

The independent graph convolution branch consists of a groups of identical computational modules. The following computations are implemented in each computational module. First, the local target information graph network r_o and the local image information graph network r_t each perform a graph convolution operation to achieve an interactive fusion of information within a single modality. Then, the two unimodal information graph networks perform a crossmodal attention enhancement operation to accomplish the necessary computation and information enhancement between different modal nodes. Finally, a new graph node information is generated after a fully connected layer FC with the following modular computational flow:

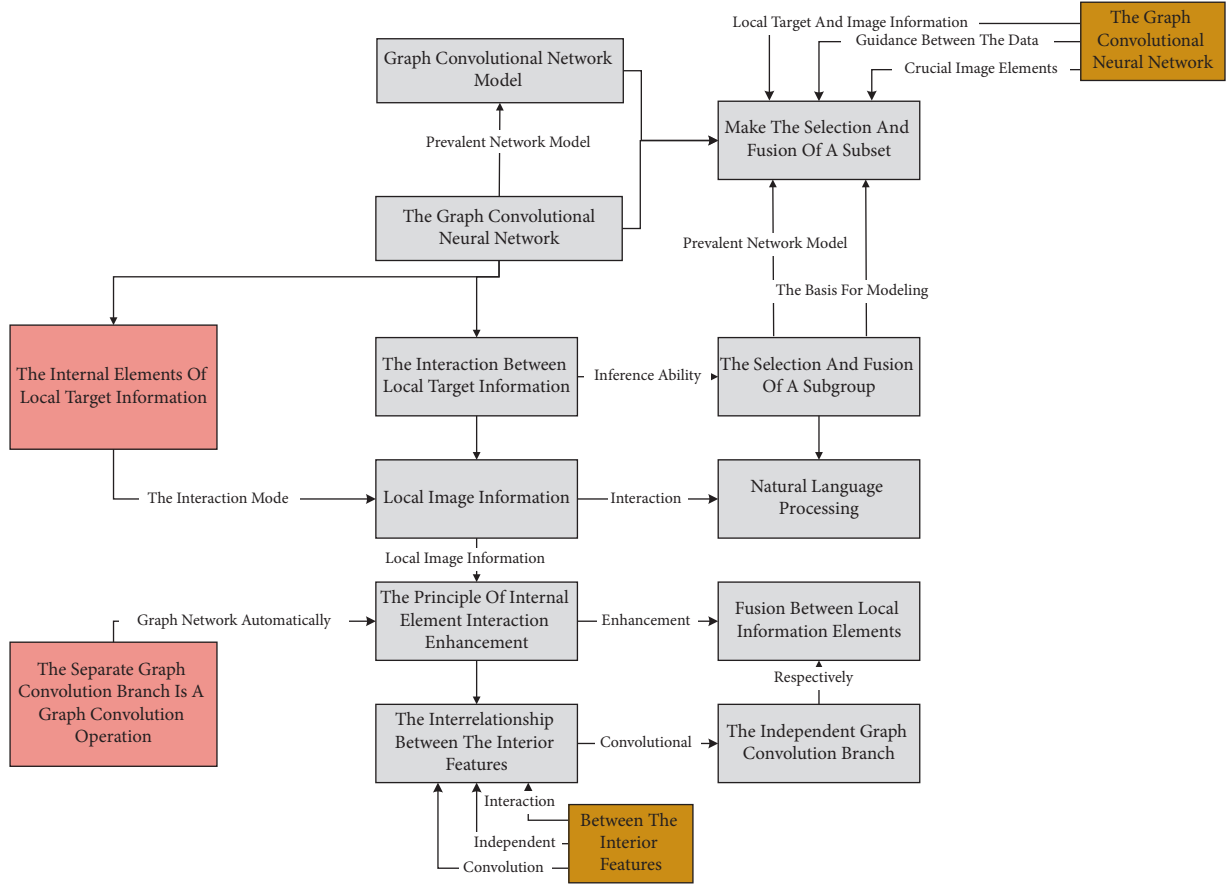


FIGURE 1: Multimodal local information interaction module.

$$\begin{aligned}
 r_{(a-o)} &= fc(h_{a+o} - 1) + \sum(a_{ij} + h_{a-t}), \\
 h_{(a-o)} &= \sum[gc_n - r_{a+o} + (r_{a+o} - 1)], \\
 a_{i-j} &= \sum(h_{a-o} + w_{a-1}) \times (h_{a-t} - w_{a-2}).
 \end{aligned} \quad (6)$$

3.2. Image Super-Resolution Relationship Extraction and Reconstruction Method Model Construction. The core idea of the image super-resolution reconstruction algorithm is to process the low-resolution image using various technical software. The detailed information not available in the low-resolution print is extracted through some algorithms, and a clear, high-resolution image is reconstructed. This section mainly introduces the theoretical basis of image starting reconstruction, some SFI reconstruction techniques, and the recognized image quality evaluation criteria for image super-resolution reconstruction. The evaluation criteria are the criteria for this study's subsequent experimental results. Image resolution is expressed in computer storage as the resolution that digital images displayed and stored in a computer have, and the resolution refers to the amount of information stored in a snap [25]. Specifically, it relates to the number of pixel points stored per unit of the image, and the resolution team is expressed in PPI (pixels per inch). In general, the more pixel dots per unit of an embodiment, the higher the resolution of the image and the

larger the image will be, thus allowing for a richer representation of detail. For example, a picture with a resolution of $160 * 120$ pixels has a resolution of 19,200 pixels or 200,000 pixels. The super-resolution image reconstruction algorithm can be divided into two types: image and static image, and this study focuses on the super-resolution reconstruction algorithm for static images. The original high-resolution image generates a low-resolution image due to some extraneous culmination of the imaging process, and the HDR image must be built. The low-resolution image is processed into a high-resolution image according to specific super-resolution techniques. In this process, the image degradation model degrades high-resolution photos into low resolution images.

The structure of the domain adaptation model based on graph convolutional networks proposed in this study is shown in Figure 2. Overall, we first extract the high-dimensional features of the input data using a pretrained deep convolutional network fine-tuned with the source domain dataset or some manually designed feature extraction algorithms. Then, to consider the correlation graph of the data, we obtain the correlation structure between the samples based on the extracted features by the k-nearest neighbor (KNN) method, thus introducing the correlation between the pieces in the source and target domains into the learning model. After that, we apply a convolutional graph network to learn similar feature representations based on the samples

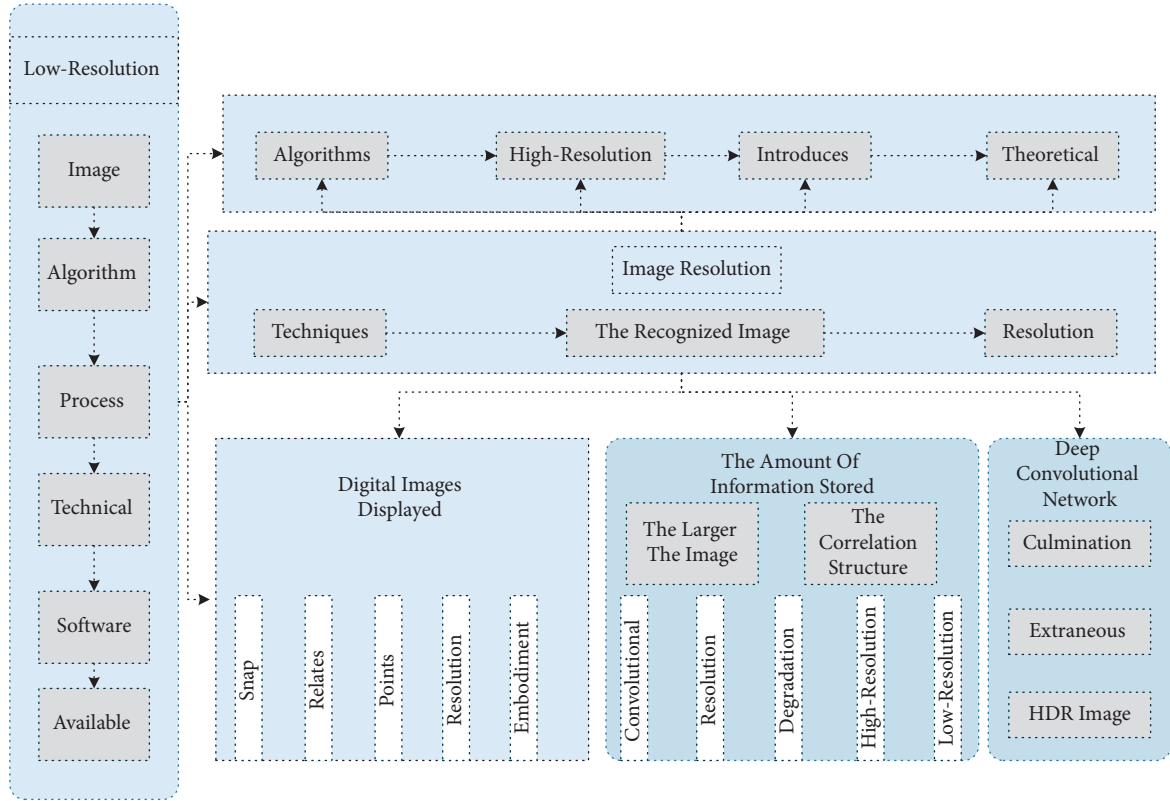


FIGURE 2: Domain adaptation model based on graph convolutional network.

and their neighboring samples. Finally, we reduce the difference in distribution between the known source and target words using the maximum mean difference to ensure the migratable nature of the features.

Because the traditional gcnt network cannot represent the relationship data such as vertices and edges, graph convolution neural network can solve the problem of such graph data, which belongs to the application of gcnt in the direction of graph expansion. In the training process, GNN will notice the graph structure, and there will be a gating mechanism to enter the graph structure, and convolution will be introduced into the graph structure to learn by extracting spatial features. The GNN that introduces convolution is the GCN, which knows by removing spatial features. GCN is a graph convolutional neural network, a kind of GNN; the difference is mainly in using convolutional operators for information aggregation. The structure of the SRCNN model is straightforward; the input image on the left is a low-resolution image generated by bi-triple interpolation, which is the exact resolution of the actual high-resolution image. However, the input image without image enhancement is still a low-resolution idea to distinguish between the two. The size of the convolution kernels for the three layers of convolution used in the model are, from left to right, 64, 32, and 1 for the output channels. The loss function used in this network is the mean square error, which is given by the following equation:

$$\text{mse}_{(x-y-\theta)} = \sum \frac{\sqrt{h \times w - 1}}{y(i-j) + x(i+j)}, \quad (7)$$

where X denotes the high-resolution image output from the web, Y represents the actual high-resolution image and denotes the network parameters, and w and h denote the length and width of the output image, respectively. The proposed model broadly lays down the structural composition of the whole super-resolution network, and all convolutional networks doing super-resolution tasks after that largely follow the combination of these three modules.

As important auxiliary information, the higher the accuracy of depth information, the more accurately it can reflect the geometric relationships between viewpoints, which helps to solve the artifacts and distortions that appear in the synthesized views. The existing view synthesis methods based on depth information generally have the following problems: the synthesized view is highly dependent on the quality of the depth map, but the predicted depth map suffers from insufficient accuracy due to the inability of the depth estimation module to capture long-range spatial correlations [26]. Therefore, it is essential to obtain effective feature representations to improve the depth map quality for subsequent operations. This module can thoroughly learn effective high-resolution feature representations and always keep the feature resolution uniform throughout the process. The multiscale fusion mechanism is designed to fuse the relevant features to obtain rich feature representations fully. This enables the proposed depth estimation module to fully capture the long-range spatial correlation. The predicted depth map can more accurately reflect the spatial distribution of the scene and provide information support for the

next operation. The specific structure of the depth estimation module is shown in Figure 3.

To address the computational inefficiency of prior upsampling, some researchers have proposed to perform most of the mappings in low-dimensional space with last. Unlike the prior upsampling, this class of models replaces the traditional upsampling operation in the prior upsampling with a learnable upsampling house at the end of the network. Since this class of models performs many linear convolution operations in the low-dimensional space, the time and space costs are significantly reduced, and training and testing are much faster. Progressive upsampling models reduce the learning difficulty of the model by decomposing a complex task into small, simple tasks. Such models provide an elegant solution to the multiscale super-resolution problem without adding time and space costs.

4. Analysis of Results

4.1. Image Super-Resolution Relationship Analysis of Multimodal Graph Convolutional Networks. The image super-resolution task is based on the single-image super resolution, in the case of having the most basic original low-resolution image, to acquire its neighboring low-resolution image frames, which is used to help the original image more quickly to obtain more information to help the image recovery. This section proposes a deep neural network module for image reconstruction, enhanced reconstruction block (ERB). This module is redesigned for the reconstruction module in the ultradeep model in image super resolution using a roll-up group plus a dense connection. It adds jump connections from shallow to deep features while maintaining the existing network depth to better-fit feature extraction and image reconstruction in deep networks. Meanwhile, to improve the deformable convolution in the feature alignment module during image super-resolution model training, a weight normalization layer is wrapped around the convolution operation in the PCD alignment module, and the stability against noise during network training is greatly improved after the replacement [27]. This section uses the classical image super-resolution model EDVR as the module framework based on the above work. It proposes a new image super-resolution model—enhanced reconstruction model for video super resolution (ERM-VSR). In practical experiments, the ERM-VSR image super-resolution model presented in this section achieves excellent performance that significantly exceeds that of the baseline EDVR model.

With the development of deep learning techniques, the complexity of graph convolutional networks is increasing, and the number of layers of the network is also growing. Deepening the number of layers of the network within a specific range will make the web more expressive and richer in the features learned. However, in practical applications, increasing the number of layers of the network does not necessarily lead to better output results. The loss rate variation curve of the graph convolutional network versus the number of pieces of training is shown in Figure 4.

During the algorithm validation training on this dataset, it was found that EDVR's feature alignment module, PCD

alignment module, often failed to converge due to excessive offsets. In the subsequent investigation of the reasons for the network convergence failure and the in-depth analysis of the training dataset, it was further found that for processing videos with too drastic scene switching (usually corresponding to the rapid movement of the filming equipment) and camera switching such as off-cut and jump-cut in transitions, PCD alignment module cannot effectively limit the size of the learned motion vector offset. Once it jumps out of the effective range and is input to the deformable, the motion vector is out of the compelling content. It is input to the deformable convolution, leading to the failure of feature extraction and loss of the whole feature alignment module.

The performance of graphical convolutional neural networks depends on various factors such as network structure and depth. Studying how parameters affect the performance of super-resolution reconstruction networks can effectively guide the model design. It can fully exploit the performance of the networks. Since the network structure is crucial to the algorithm's convergence, this section first conducts experiments on the effect of residual learning on the performance of the RLSR algorithm. All three experiments used T1-weighted imaging of the brain web dataset as the test set and PSNR as the evaluation index to test the results of the RLSR algorithm when there was super-resolution reconstruction of anisotropic 3D-MRI images with a resolution of $2\text{mm}\times 2\text{mm}\times 2\text{mm}$. The effects of residual learning, network depth, and width are shown in Figure 5.

The best method among the interpolation methods is the B-spline interpolation algorithm. Still, the PSNR and SSIM of this algorithm are 3.95 dB/0.0059 and 3.36 dB/0.0407 lower than those of the RLSR algorithm for layer thicknesses of 2 mm and 5 mm, respectively. Due to the fixed parameters of the interpolation method, the image is only upsampled based on the spatial information of the pixels without using any a priori information. The NLM and SC methods exploit the self-similarity and sparsity of the image for super-resolution reconstruction, respectively, improving the super-resolution reconstruction effect [28]. Still, the PSNR and SSIM of the reconstructed image are not as good as the RLSR based on the residual learning deep convolutional neural network. The SRCNN method is driven by many training samples and directly learns the intrinsic mapping relationship between high and low resolutions without relying on artificially designed feature extraction methods. Its super-resolution reconstruction effect is significantly better than the interpolation method, NLM, and SC algorithms. Since the RLSR algorithm uses residual learning to alleviate the problem of difficult training of deep networks faced by SRCNN and effectively improves the nonlinear fitting ability of the network, the quality of super-resolution reconstructed images at a slice thickness of 2 mm is better than those reconstructed by SRCNN and VDSR methods, with PSNR values 1.28 dB and 0.06 dB higher than those of SRCNN and VDSR method approaches, respectively. The quality of the super-resolution reconstructed 3D-MRI images decreased to different degrees with the increase of the slice layer thickness. The SSIM of the 3D-MRI images reconstructed by the RLSR algorithm was 0.004 higher than that of the SRCNN

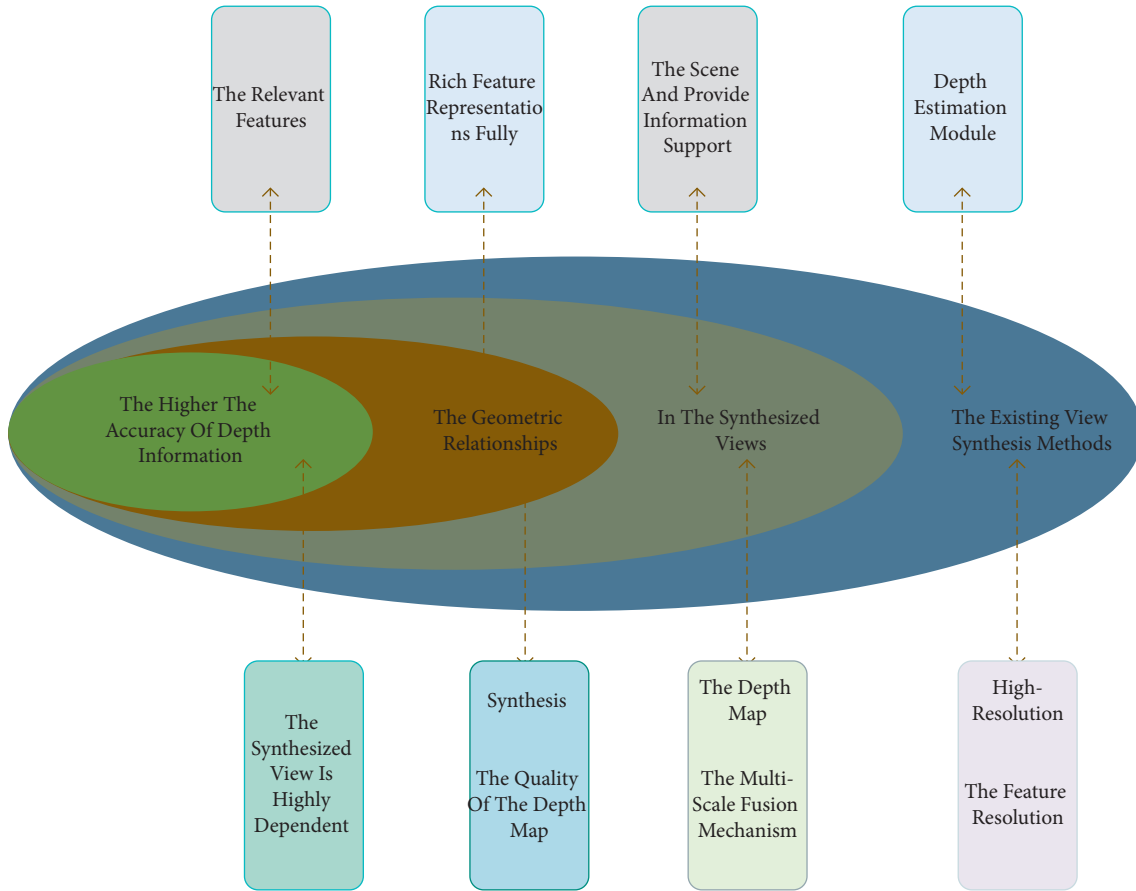


FIGURE 3: Specific structure of the depth estimation module.

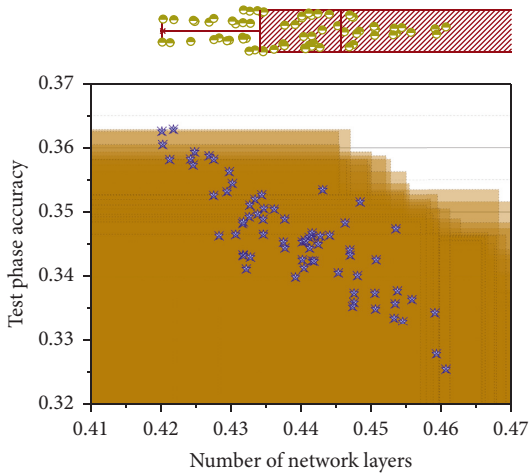


FIGURE 4: Effect of the number of network layers on the accuracy of the training and testing phases.

method when the layer thickness was 2 mm, but the difference reached 0.0254 when the layer thickness was increased to 5 mm. The above experimental results indicate that the RLSR algorithm can achieve good T1-weighted imaging super-resolution reconstruction results and has good robustness for reconstructing different slice thicknesses.

4.2. A Multimodal Graph Convolutional Network-Based Approach for Super-Resolution Relation Extraction and Reconstruction of Images Implementation. For the overall performance comparison, the number of SUB modules in SUGNet is set to 20, and the output channels of the convolutional layer are set to 64. Considering the performance and model parameters, the depth of the backbone branch in the SUB module is set to 3. During the training period, a randomly cropped 48×48 image block is used as the model's input. To avoid overfitting the SUGNet algorithm during training, this section uses data enhancement techniques such as rotation and horizontal and vertical flipping for all fundus data sets. The Adam optimizer is used to train the network parameters with an initial learning rate of 0.0001, and the learning rate is reduced by half for every 100 rounds. For the same reconstruction factor, the generator loss of the algorithm in this study is lower than that of both SRRes Net-V54 and SRGAN. For different reconstruction factors, the generator losses of SRRes Net-V54 and SRGAN are in the order from small to large: $4 \times < 6 \times < 8 \times$, while the order of the algorithm in this study is as follows: $4 \times \approx 6 \times < 8 \times$. It proves that the generator network in this study can be used well for $4 \times$ and $6 \times$ reconstruction. Still, the other two algorithms are only suitable for $4 \times$ reconstruction and have more significant errors for $6 \times$ and $8 \times$. Using feature matching loss (F-Loss) and Wasserstein distance loss (W-Loss) can improve

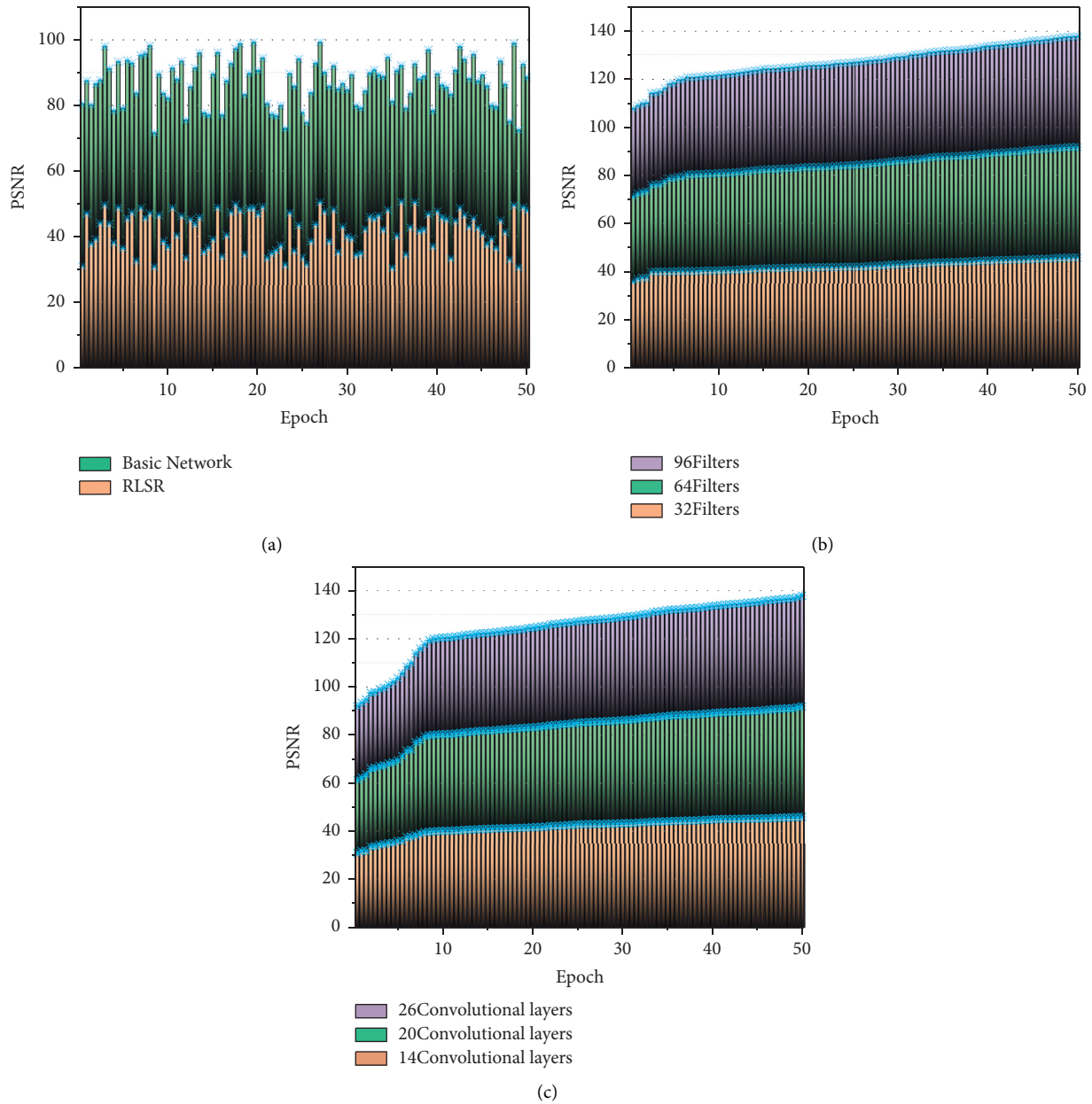


FIGURE 5: Effect of residual learning, network depth, and width. (a) Effect of residual learning. (b) Effect of network width. (c) Effect of network depth.

the reconstruction quality and solve the gradient dispersion phenomenon that may occur during the training process. In addition, the multiplex conditional generator structure and the multiscale discriminator structure make the generator’s performance in this section almost the same as that of the reconstruction factor 4 when the reconstruction factor is 6. Therefore, the algorithm in this section can cope with more prominent reconstruction factors, while the performance of other algorithms decreases sharply when the reconstruction factor increases. The dynamics of the different network loss function values are shown in Figure 6.

This study uses a network structure with only one hidden layer to simplify and prevent overfitting. The number of neurons in the hidden layer is as small as possible.

Meanwhile, the graph convolutional network algorithm uses each node’s k-nearest neighbors to describe each vertex’s local information on the image model. 2D is also called two-dimensional, flat graphics. 2D graphics content X-axis and Y-axis. 2D three-dimensional sense, light, and shadow are artificially drawn from the simulation. 3D is also called three-dimensional graphics content; in addition to the horizontal X-axis, vertical Y-axis, and the depth of the Z-axis, three-dimensional graphics can contain 360 degrees of information. Therefore, like the 2D reconstruction of images based on graph convolutional networks, determining the number of neurons in each subneural network and the number of k-nearest neighbors is also essential for the 3D reconstruction of faces. Therefore, in this study, from the

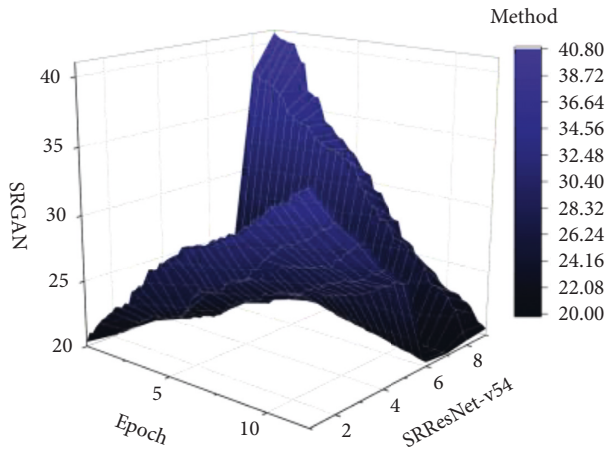


FIGURE 6: Dynamics of different network loss function values.

2,800 strictly aligned 3D face models obtained during the face data generation, 1,000 are randomly selected as the training set and 500 as the test set. First, we test the prediction results of the network under different k values. In the network initialization phase, the network weight parameters for the first forward propagation of the generator network can be initialized with the DGP-SRGAN network parameters by using the minimized mean square error MSE loss function, which is obtained by pretraining the network. Because of this, the following training process is chosen to “synchronize” the alternate iterative training of the generator network and the discriminator network; in the general GAN model, the generator network training learning speed is often slower than the discriminator network, which will cause the network parameters to update early end, and it will not get a robust generator model. In the training phase of the network, the discriminator network is updated once, followed by the generator network to update the parameters once. The super-resolution image of the output of each forward propagation of the generator network is compared with the original high-resolution image HR to obtain an error signal. This error signal is back propagated to produce a gradient (or derivative) for learning, which is used to readjust the weight parameters for the subsequent forward propagation. The discriminator network then compares the output probability score of the input super-resolution generated image with 0 and the original high-resolution extensive image HR with 1. It updates the discriminator network parameters by back propagating the error through a back-propagation mechanism to create the gradient used for network learning. The results of the network training for the image super-resolution relationship extraction and reconstruction method are shown in Figure 7.

DRCN is equivalent to SRCNN with a deepened network hierarchy. The DRCN network is more expressive and can be seen to have more apparent edge details than SRCNN. The SRGAN and the optimized and improved DGP-SRGAN algorithm in this section can reconstruct more texture details than the general GNN because they use the perceptual loss function to guide the network training, and the experimental results of the previous algorithms

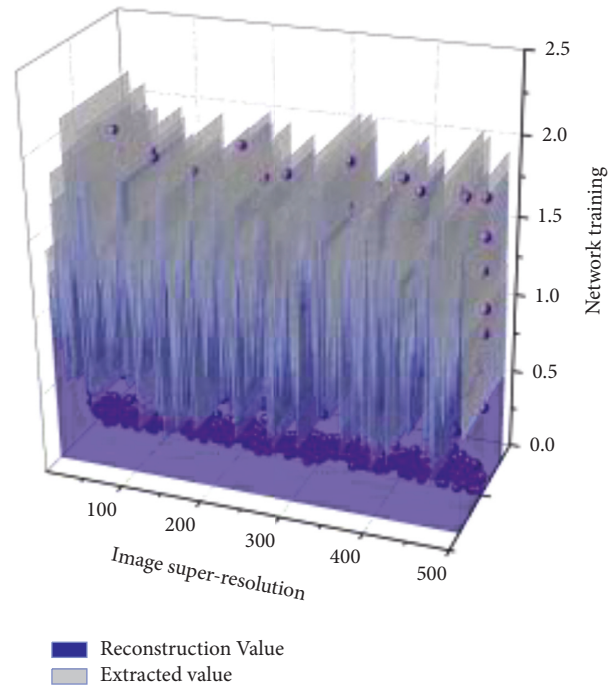


FIGURE 7: Network training results of image super-resolution relationship extraction and reconstruction method.

have better image visualization and more explicit edge details compared to each other. The proposed DGP-SRGAN has better subjective visual perception quality than the original SRGAN algorithm. The essence of graph convolution is to learn relevant information, so the learning effect of this network must include the neighbors of the sampled samples in the same training step; on the other hand, the distribution difference metric requires that the models in both domains can be as rich as possible and cannot be limited to only some categories. Balancing the needs of both in a limited batch training size is another critical issue in enhancing the effectiveness of graph convolution in deep learning frameworks. According to the scheme proposed in this section, the update relevance graph with the training trick allows the global relevance graph to be updated throughout the network training process and no longer overly dependent on fine-tuning the features extracted in the network. Class-label and pseudo-class-label sampling ensure, to some extent, the amount of data available for each class of samples when the model is trained in small batches, thus improving the performance of the overall model. The proposed two schemes enable the graph convolution model to be successfully integrated into the deep learning framework for end-to-end learning and achieve good results in experiments comparable to cutting-edge algorithms.

5. Conclusion

With the development of deep learning technology, more and more tools have been derived from continuously bringing new products and experiences to the public. Many

technologies that were previously unlikely to be realized based on traditional methods are increasingly coming into the typical home. Image recovery, a classic task in computer vision, has a critical position in practical applications. As an essential carrier of information transmission, the quality of the image directly affects the ability of information expression. Image super-resolution reconstruction aims to recover high-quality photos, so it has a wide range of applications in many fields. We conducted comparison experiments on COCO and visual genome datasets in this study. By analyzing the experimental data, we can see that the target detection and recognition models based on graph convolutional networks significantly improve the correct average rate of the whole class of objects. In this study, Set5, Set14, BSD100, and Urban100s datasets are taken for experiments and compared with their algorithms Bicubic, SRCNN, VDSR, and SRGAN in the cases of reconstruction scales of $2\times$ and $4\times$ to verify the practical effect more fully. This algorithm increases the network's nonlinear representation capability while acquiring multiple features than single-scale convolutional networks. The algorithm finally outputs reconstructed high-resolution images using the deconvolution layer, which obtains more high-frequency information during the upsampling process. The algorithm is experimentally demonstrated to have an advantage of super-resolution reconstruction compared with neural network algorithms of the same level of depth.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the School of Railway Communication and Signaling, Wuhan Railway Vocational College of Technology.

References

- [1] S. Zhang, Q. Yuan, J. Li, J. Sun, and X. Zhang, "Scene-adaptive remote sensing image super-resolution using a multiscale Attention network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4764–4779, 2020.
- [2] D. Zhang, J. Shao, Z. Liang, X. Liu, and H. T. Shen, "Multi-branch networks for video super-resolution with dynamic reconstruction strategy," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3954–3966, 2021.
- [3] K. Zheng, L. Gao, W. Liao et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2487–2502, 2021.
- [4] W. Zhang, X. Sui, G. Gu, Q. Chen, and H. Cao, "Infrared thermal imaging super-resolution via multiscale spatio-temporal feature fusion network," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19176–19185, 2021.
- [5] Y. Cao, F. Wang, Z. He, J. Yang, and Y. Cao, "Boosting image super-resolution via fusion of complementary information captured by multi-modal sensors," *IEEE Sensors Journal*, vol. 22, no. 4, pp. 3405–3416, 2022.
- [6] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multi-modal image restoration and fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3333–3348, 2021.
- [7] X. Deng, P. Song, M. R. D. Rodrigues, and P. L. Dragotti, "RADAR: robust algorithm for depth image super resolution based on FRI theory and multimodal dictionary learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2447–2462, 2020.
- [8] Y. Wu, X. Ji, W. Ji, Y. Tian, and H. Zhou, "CASR: a context-aware residual network for single-image super-resolution," *Neural Computing & Applications*, vol. 32, no. 18, pp. 14533–14548, 2020.
- [9] K. Wu, Y. Qiang, K. Song et al., "Image synthesis in contrast MRI based on super resolution reconstruction with multi-refinement cycle-consistent generative adversarial networks," *Journal of Intelligent Manufacturing*, vol. 31, no. 5, pp. 1215–1228, 2020.
- [10] X. He, Y. Tao, S. Yang, C. Chen, and H. Lin, "ScalarGCN: scalar-value association analysis of volumes based on graph convolutional network," *Journal of Visualization*, vol. 25, no. 1, pp. 77–93, 2022.
- [11] A. Buades, J. Duran, and J. Navarro, "Motion-compensated spatio-temporal filtering for multi-image and multimodal super-resolution," *International Journal of Computer Vision*, vol. 127, no. 10, pp. 1474–1500, 2019.
- [12] E. Altinkaya, K. Polat, and B. Barakli, "Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: a comprehensive review[J]," *Journal of the Institute of Electronics and Computer*, vol. 1, no. 1, pp. 39–53, 2020.
- [13] A. Chadha, J. Britto, and M. M. Roja, "iSeeBetter: spatio-temporal video super-resolution using recurrent generative back-projection networks," *Computational Visual Media*, vol. 6, no. 3, pp. 307–317, 2020.
- [14] Y. Hong, J. Kim, G. Chen, W. Lin, P. T. Yap, and D. Shen, "Longitudinal prediction of infant diffusion MRI data via graph convolutional adversarial networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 12, pp. 2717–2725, 2019.
- [15] P. Kocsis, I. Shevkunov, V. Katkovnik, H. Rekola, and K. Egiazarian, "Single-shot pixel super-resolution phase imaging by wavefront separation approach," *Optics Express*, vol. 29, no. 26, pp. 43662–43678, 2021.
- [16] L. I. Yanshan, Z. Li, and X. U. Fan, "OGSRN: optical-guided super-resolution network for SAR image[J]," *Chinese Journal of Aeronautics*, vol. 35, no. 5, pp. 204–219, 2022.
- [17] L. Chen, J. Pan, and Q. Li, "Robust face image super-resolution via joint learning of subdivided contextual model," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5897–5909, 2019.
- [18] S. Luo, J. Hu, and Z. Yang, "A survey on deep learning for super-resolution of diffusion magnetic resonance imaging[J]," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 9, pp. 2440–2449, 2021.
- [19] V. S. S. Kandarpa, A. Bousse, D. Benoit, and D. Visvikis, "DUG-RECON: a framework for direct image reconstruction using convolutional generative networks," *IEEE Transactions*

- on *Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 44–53, 2021.
- [20] Z. Zhou, Y. Wang, J. Yu, W. Guo, and Z. Li, “Super-resolution reconstruction of plane-wave ultrasound image based on a multi-angle parallel U-Net with maxout unit and novel loss function,” *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 1, pp. 109–118, 2019.
- [21] Y. Yang, Q. Cao, J. Zhang, and D. Tao, “CODON: on orchestrating cross-domain attentions for depth super-resolution,” *International Journal of Computer Vision*, vol. 130, no. 2, pp. 267–284, 2022.
- [22] S. Sharma, V. S. Bawa, and V. Kumar, “A novel two-stage residual learning based convolutional neural network for image super resolution,” *Fundamenta Informaticae*, vol. 168, no. 2-4, pp. 335–351, 2019.
- [23] F. Li, Q. Cai, H. Li, Y. Chen, J. Cao, and S. Li, “Attentive frequency learning network for super-resolution,” *Applied Intelligence*, vol. 52, no. 5, pp. 5185–5196, 2022.
- [24] J. Singh, K. Tant, A. Curtis, and A. Mulholland, “Real-time super-resolution mapping of locally anisotropic grain orientations for ultrasonic non-destructive evaluation of crystalline material,” *Neural Computing & Applications*, vol. 34, no. 6, pp. 4993–5010, 2022.
- [25] B. Stimpel, C. Syben, F. Schirmacher, P. Hoelter, A. Dorfner, and A. Maier, “Multi-modal deep guided filtering for comprehensible medical image processing,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1703–1711, 2020.
- [26] X. Wang, J. Chen, Q. Wei, and C. Richard, “Hyperspectral image super-resolution via deep prior regularization with parameter estimation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1708–1723, 2022.
- [27] L. Chen, J. Ye, and X. Zhang, “Multi-feature super-resolution network for cloth wrinkle synthesis,” *Journal of Computer Science and Technology*, vol. 36, no. 3, pp. 478–493, 2021.
- [28] J. Yu, J. Yin, S. Zhou, S. Huang, and X. Xie, “An image super-resolution reconstruction model based on fractional-order anisotropic diffusion equation,” *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 6581–6607, 2021.