Version of Record: https://www.sciencedirect.com/science/article/pii/S1959031820301408
Manuscript_f21563c7ad91c9995a191ce48d2b7f10

A review of the deep learning methods for medical images super resolution problems

Y. Li[a], B. Sixou[a,*], F.Peyrin[a,b]

[a] *Univ Lyon, INSA-Lyon, Universite Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, LYON, France*
[b] *ESRF,6 rue Jules Horovitz, F-38043, Grenoble Cedex France*

Abstract

Super resolution problems are widely discussed in medical imaging. Spatial resolution of medical images are not sufficient due to the constraints such as image acquisition time, low irradiation dose or hardware limits. To address these problems, different super resolution methods have been proposed, such as optimization or learning-based approaches. Recently, deep learning methods become a thriving technology and are developing at an exponential speed. We think it is necessary to write a review to present the current situation of deep learning in medical imaging super resolution. In this paper, we first briefly introduce deep learning methods, then present a number of important deep learning approaches to solve super resolution problems, different architectures as well as up-sampling operations will be introduced. Afterwards, we focus on the applications of deep learning methods in medical imaging super resolution problems, the challenges to overcome will be presented as well.

Keywords: Deep learning, super resolution, medical imaging

1. Super resolution in medical imaging

Super resolution (SR) refers to methods aiming at increasing the spatial resolution of digital images. It led to the development of many algorithms to process images [1], such as natural images [2], satellite images [3], or medical imaging [4] for instance.

SR algorithms can be classified according to the number of input and output images involved in the process. In this paper, we focus on single-image SR referring to methods where one high resolution (HR) image has to be recovered from one low resolution (LR) image. In this case, the problem is often modeled as an ill-posed inverse problem and is solved by optimization approaches. Methods based on machine learning, such as Dictionary learning and more recently Deep Learning have also been proposed.

Deep learning (DL) methods are attracting a lot of attention nowadays in image processing. In particular, DL for medical imaging has achieved amazing progresses in fields such as image reconstruction[5], denoising[6], super resolution[4], segmentation [7], computed-assistant diagnosis [8]. Dong *et al.* [2] first proposed to use deep learning method to solve super resolution problems on natural image sets. Since then, the application of DL methods for natural images SR has been widely investigated [9, 10, 11, 12, 13]. More recently, DL methods have also been proposed for SR problems in medical images. Compared with natural image SR, medical image SR needs additional priors information for particular applications. Since SR techniques in medical imaging are often followed by segmentation or diagnosis, it is very challenging to enhance the structures of interest and to preserve sensitive information. Moreover, the datasets of medical images are relatively small and hard to collect, especially for clinical high and low resolution image pairs.

In this paper, our aim is to review DL methods for SR problems of medical imaging. After a brief introduction to DL approaches, we show different SR DL approaches on natural image sets. The applications of DL in medical images SR problems will be presented afterward. Challenges including how to deal with data paucity and how to integrate priors will be discussed at the end of the article.

2. A brief introduction of deep learning

Recently, deep learning approaches proved to be very promising in image processing with tasks such as segmentation [7], classification [14], denoising [15] or solving inverse problems [16]. Deep learning approaches have two principal advantages which distinguish them from other approaches: much developed parallel calculation and powerful representation ability.

1. Parallel calculation has been much developed in deep learning framework, such as Tensorflow [17], MXNet [18], Caffe [19]. The user can directly benefit from high speed parallel computation without knowing GPU architecture and low-level GPU programming.
2. DL allows to learn high-level features of the data. A large quantity of parameters within DL networks is used to reveal implicit information.Yet, the efficiency

*Corresponding author

of deep learning approach is related to the amount of data. A large amount of data could boost the performance of deep learning, on the contrary, a limited number of data limits its performance.

2.1. Network units

The DL network is a multi layers neuron network. Its first layer is named as input layer, the last layer is named as output layer. The intermediate layers are named as hidden layers. Fig.1 illustrates a classic 5 layers DL neuron network. Each neuron in the network consists of linear transformation followed by a point-wise activation function.

2.1.1. Linear transformation

In multilayer perceptron, every neuron at layer l is connected to all the neurons at layer $l+1$ with weight $\theta^l_{j,i} \in \mathbb{R}$, i and j correspond to the neuron index at layer l and layer $l+1$, weight matrix $\boldsymbol{\theta}^l \in \mathbb{R}^{c^{l+1} \times c^l}$, where c^l and c^{l+1} denote the number of channels at layer l and layer $l+1$. If the output of layer l is $\boldsymbol{L}^l \in \mathbb{R}^{c^l \times m \times n}$ (here we ignore batch size factor), where m and n are the height and width of image features, the activation function at layer l is σ^l, the output of layer $l+1$ is $\boldsymbol{L}^{l+1}$, written as

$$\boldsymbol{L}^{l+1} = \sigma^l(\boldsymbol{\theta}^l \boldsymbol{L}^l) \tag{1}$$

The linear operation between $\boldsymbol{\theta}^l$ and $\boldsymbol{L}^l$ is a matrix multiplication, the height and width of image features do not change. When the linear operation is a convolution, $\theta^l_{j,i}$ becomes a convolutional filter, the network turns to be a convolutional neural network (CNN).

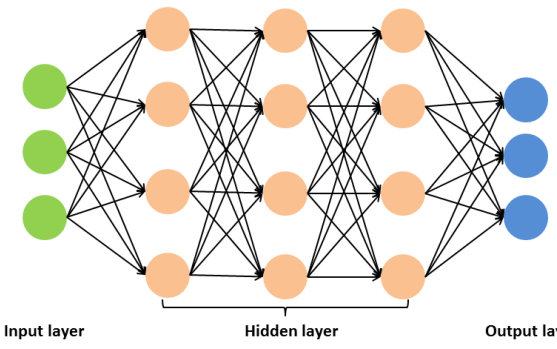

Figure 1: Illustration of a classical neuron network. In the training stage, the input and output are fed with training data, the network is optimized by minimizing a user-selected function with respect to parameters within the network. Each solid circle represents a neuron which consists of linear transformation and nonlinear activation function.

2.1.2. Activation functions

Activation functions are essential for the network since they introduce nonlinear factors to the network. Fig.2 depicts some typical activation functions. Rectified Linear Unit is illustrated in Fig.2(a). It linearly rectifies the nonnegative values and removed negative parts. PReLu (Parametric Rectified Linear Unit) is developed from ReLU activation, as shown in Fig.2(b). It rectifies the negative and nonnegative values with different degree. PReLU is defined as

$$PRelu(x) = \begin{cases} x, & \text{if } x \geq 0 \\ ax, & \text{otherwise.} \end{cases}$$

where a is adapted during the training process.

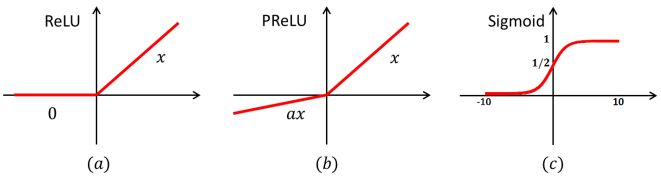

Figure 2: Illustration of 3 activation functions: ReLU, PReLU, sigmoid activation functions. The coefficient for the negative parts in ReLU is zero, whereas in PReLU, it is parametrized by a. Sigmoid function is used for two label classification.

The sigmoid function is used for 2-label classification tasks. Fig.2(c) displays its curve and it is defined as:

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{2}$$

It can be seen that when $x < -10$, the output of sigmoid function is close to 0, when $x > 10$, its output is near 1.

2.2. CNN

Among the deep learning networks, many architectures have been proposed: multilayer perceptron, basic CNN, recurrent neural networks, recursive network, stacked denoising auto-encoders, generative adversarial networks and so forth [20].

In this work, we focus on CNN, a workhorse of deep learning, particularly for image processing. When dealing with an inverse problem with CNN, the inverse operator in the inverse problem is approximated by a sequence of filtering operations alternating with nonlinear operations.

The convolution operation allows network to detect the same feature in different regions of image. One CNN layer detects local features of image, while multi-layer CNN allows to increase the perception field and to synthesize the features extracted at previous layers. Moreover, CNN reduces the number of weights by sharing them between network's neurons, which results in a considerable memory reduction.

2.3. Optimization of the network

The parameters in the network, such as the weights of filters, or parameters in the activation functions, are optimized by minimizing a loss function, which conventionally measures the distance between the network estimation and ground truth. The loss function varies with

task types. Regularization terms can be integrated in the loss function to reduce over-fitting risk or inject priors.

The value of the loss function can be computed via forward propagation, and the gradients of the loss function with respect to parameters in the network are determined via back-propagation [21]. Stochastic gradient descent method accelerates back propagation by processing data in small batch [20]. The parameters are optimized by successive forward and backward pass in the network.

The fitting capacity of a network is described by the 'bias', equals to the expectation of error on the training set. The generalization capacity is evaluated by the 'variance', evaluated on the test set. With the increase of the network size, the bias tends to decrease while the variance tends to increase, which can be interpreted as the network evolves from underfitting toward overfitting. It is generally admitted that, in a network, a trade off must be found between the bias and the variance in order to prevent overfitting along with a small generalization error.

Many factors influence the performance of the network: the network architecture, the training set, the optimization procedure. The architecture of the network varies with the type of task. A large dataset in which the data follows the same distribution boosts the performance of the network. A good optimization procedure can increase the efficiency of training process, or/and improve the accuracy of estimation.

2.4. Regularization techniques

Deep neural networks can learn complicated relations between the inputs and outputs and the capacity of the model is related to its architecture which has to be big enough. With limited training data, a large model with a high capacity may lead to overfitting and a poor generalization. In order to regularize overfitting, it is possible to reduce the capacity of the network by changing its architecture. Two widely used examples are dropout and batch normalization. With dropout, some neurons are randomly switched off during the training process, inducing a noisy input to the subsequent layers [22]. Dropout can be seen as a way of doing an equally-weighted averaging of exponentially many models with shared weights. The batch normalization layers have been used to reduce the internal covariate shift and accelerate the traning process [23], which is a way to induce noise to subsequent layers. This technique is widely used by super resolution models [24, 11]

Another method to reduce the network capacity is the regularization of the loss function. Some strategies relies on weight decay [25, 26] , pruning of the network [27], L_1 and L_2 regularizations on the weights [28, 29]. Decorrelation techniques can also improve the generalizatbility of the network [30], [31]. In order to avoid overfitting, it is also possible to artifically increase the amount of training data (random crop, rotation, flipping, ...) and to modify the training input to reduce the overadapation. As far as super resolution is concerned, many super resolution methods use very deep networks with a large number of parameters, with a high risk of overfitting. Data augmentation is a very efficient approach for super resolution [32]. Data synthesis approach with a learned degradation operator may also improve the super resolution results.

2.5. Generative Adversarial Networks

Generative Adversarial Networks (GAN) are based on a game approach with a generator and a discriminator network. Recently numerous works have developed more effective GAN models that outperforms traditional CNN networks. A detailed discussion on GAN can be found in [33]. The generator tries to generate fake images to fool the discriminator, while the discriminator aims at distinguishing the generated results from real data. At the end of the adversarial training, the generator produce outputs consistent with the distribution of the real data, and the discriminator can not distinguish the generated data and the real data. Wassertein gan , WGAN, [34] is based on the minimization of an approximation of the Wasserstein distance and regularizes the discriminator by weight clipping. Other regularization scheme have been proposed like gradient clipping and spectral normalization. Other regularizations for the discriminator have been investigated based on gradient clipping [35] and spectral normalization [36]

3. Deep learning in super resolution

3.1. Single image super resolution versus multiple image super resolution

Two types of super resolution methods can be distinguished: single-image and multiple-image methods. The aim is to generate a high-resolution image from a single or from multiple low-resolution images. Multiple-image super-resolution is based on information fusion between subpixel shifted low resolution images and generally allows for higher reconstruction accuracy. Multiple-image based methods generally utilize global/local geometric or photometric relations between multiple low resolution images. Existing techniques include interpolation-based methods, frequency-domain methods [37] or methods based on regularization [38, 39]. As detailed in the following, deep learning approaches have much improve the performances of the single-image super-resolution methods. There are very few deep learning methods applied to multiple-image super resolution. In [40], the authors use a deep residual network to improve the results of an evolutionary model for super resolution of multiple satellite images. In the following, we will focuss on single-image super-resolution.

3.2. Formulation of the single-image SR problem

Let $\boldsymbol{g} \in \boldsymbol{\mathcal{Y}}$ denote a low-resolution image and $\hat{\boldsymbol{f}} \in \boldsymbol{\mathcal{X}}$ a high-resolution image. The super resolution forward problem can be written as

$$\boldsymbol{g} = \boldsymbol{\mathcal{A}}\hat{\boldsymbol{f}} \tag{3}$$

where $\boldsymbol{\mathcal{A}} : \boldsymbol{\mathcal{X}} \rightarrow \boldsymbol{\mathcal{Y}}$ includes blurring, noise and down sampling operation. Given a training dataset $(\boldsymbol{g}_i, \hat{\boldsymbol{f}}_i)$, our goal is to learn $\boldsymbol{\mathcal{A}}^+$ which predicts values $\boldsymbol{f} = \boldsymbol{\mathcal{A}}^+\boldsymbol{g}$ so that $\boldsymbol{f} \approx \hat{\boldsymbol{f}}$. $\boldsymbol{\mathcal{F}}$ is a user-selected loss function used to optimize the parameters in the network. Generally speaking, a parametric approximate inverse operator $\boldsymbol{\mathcal{A}}^+_{\boldsymbol{\theta}} : \boldsymbol{\mathcal{Y}} \rightarrow \boldsymbol{\mathcal{X}}$ is learned by solving:

$$\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum \boldsymbol{\mathcal{F}}(\hat{\boldsymbol{f}}_i, \boldsymbol{\mathcal{A}}^+_{\boldsymbol{\theta}}(\boldsymbol{g}_i)) + \boldsymbol{\mathcal{G}}(\boldsymbol{\theta}) \tag{4}$$

where $\boldsymbol{\Theta}$ is the set of possible parameters and $\boldsymbol{\mathcal{G}}(\boldsymbol{\theta})$ a regularization function.

The application of deep learning in super resolution has been broadly discussed in the literature [2, 9, 10, 24, 12, 41, 42, 43, 11]. Since it's not possible to present exhaustively all the networks for SR, we select a set of these methods which outlines the development of DL in SR issues. These methods are presented in the next section, and a comparison of their performance will be given afterwards.

3.3. Evaluation metrics and loss functions

The loss functions are used as reconstruction evaluation metrics and for the model optimization. The peak signal-to-noise ratio (PSNR) is the most widely used reconstruction quality measurements metrics for super resolutio. Let L be the maximum pixel value, N the number of pixels, I the ground truth image and $\hat{I}$ the reconstruction, the PSNR is defined as:

$$PSNR = 10.log_{10}(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}(I(i) - \hat{I}(i))^2}) \tag{5}$$

The PSNR is related to the mean square error (MSE), or L_2 loss function, and gives informations about the differences at the pixel level. The L_1 loss is more robust against outliers[43]. A neural network trained with this loss may converge faster and produce better results[44]. These losses often give poor performance to represent the reconstruction quality very accurately in real images. Therefore, other functions are used to obtain higher-quality results.

The structural similarity index is also a widlely used image quality index adapted to the human visual system . It measures the structural similarity between images based on luminance, contrast and structures [45, 46].

The poor perceptual quality of super resolution images obtained by optimizing the mean square error has lead to objective functions based on MSE in a transformed space. The perceptual loss is based on the features produced by deep architecture. In [47], the super resolution network is optimized by minimizing the MSE in the feature space produced by a pre-trained network VGG-16. The feature loss encourage the output image to be perceptually similar to the true image instead of forcing the pixels to match exactly[48, 24, 49]. The networks VGG [50] and ResNet [51] are widely used pre-trained CNN. A texture loss corresponding to correlations between different feature channels was proposed in [48] to create more realistic textures.

For super-resolution, the adversarial losses are used to train gans. The training of the discriminator and of the generator are performed alternatively. In [24], the following losses based on cross entropy are used for the generator, $\mathcal{L}_{generator}$, and the discriminator, $\mathcal{L}_{discriminator}$ respectively :

$$\mathcal{L}_{generator} = -logD(\hat{I}) \tag{6}$$

$$\mathcal{L}_{discriminator} = -logD(I) - log(1 - D(\hat{I})) \tag{7}$$

where $\hat{I}$ is the generated image and I the ground truth image. In order to obtain more stable training process and better results, [52, 53] use adversarial losses based on least squares.

The various losses presented above are often combined but the choice of the weighting coefficients remains a problem.

3.4. Deep learning architecture for natural image set

3.4.1. The first DL method for solving SR

Super resolution convolutional neural network (SRCNN) [2] is a cornerstone in the literature of deep learning for super resolution problems. It seldom disappears in the benchmarks of CNN based approaches. The schema of SRCNN is presented in Fig.3. In this three layer network, the first layer is responsible for patch extraction and representation of features at LR scale, the layer in the middle is used to approximate a nonlinear mapping function, and the third layer reconstructs super resolution images. SRCNN is a landmark in SR development, nevertheless, it is usually blamed for its shallow structure.

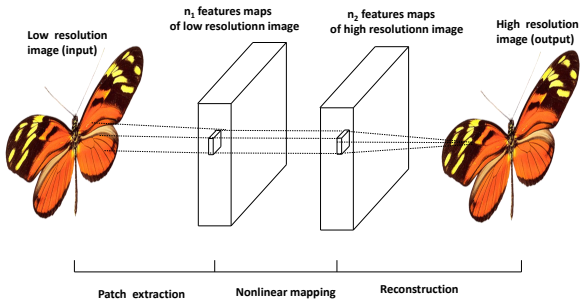

Figure 3: Schema of SRCNN: a network consists of 3 layers. The first layer is used for patch extraction and representation, the layer in the middle corresponds to non-linear mapping, the third layer reconstructs final SR images.

3.4.2. Residual-based methods

Kim *et al.* later proposed a very deep residual network for Super Resolution (VDSR) [9]. VDSR has very deep architecture (20 layers) and each layer consists of small

filters. The skip connection from the input image to the output estimation, as shown in Fig.4(a), forces the convolution filters to learn the residual between the estimation and the ground truth images. This is also the reason why it is named as residual network. The gradient clipping strategy allows to train the network with high learning rate, thus accelerates the convergence speed despite the huge size of the architecture. The principle of gradient clipping is to truncate the individual gradient so that all the gradients are constrained in a predefined range [9]. The authors found that increasing the depth of the networks improves the accuracy of the results.

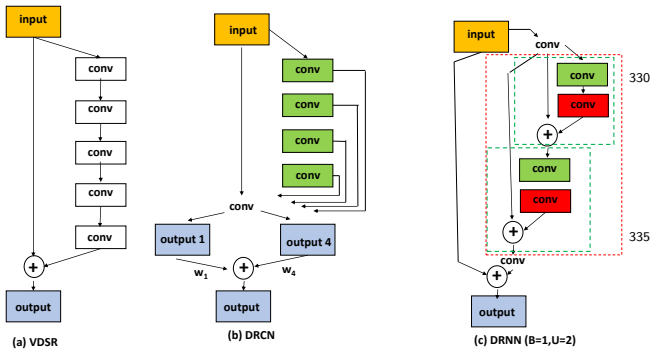

Figure 4: Figure adapted from [11]. (a) VDSR [9]. The skip connection between the input and output ensures the stability of the network, particularly for deep network. (b) DRCN [10]. The dashed green block is a recursive block. The convolution layers within the recursive block share the same parameters. (c) DRRN (B=1, U=2) [11]. B denotes the number of recursive blocks and U is the number of recursive units in a recursive block. The green dashed block denotes a residual unit, the red dashed block represents a recursive block. Inside of a recursive block, the convolution layers in the same color share the same parameters.

Deeply Recursive Convolutional Network (DRCN) [10] uses a recursive structure so that the length of the network is increased while the number of parameters is reduced. The recursive structure uses the same simple filters repetitively to extract image features. As shown in Fig.4(b), the green dashed rectangular is a recursive block. All the convolutions marked in green within this recursive block share the same parameters. All the intermediate outputs from recursive block and the input of the network are then fed into a convolution layer to generate output predictions. In Fig.4(b), there are 4 output predictions. The final estimation is determined by a linear combination of the output predictions, and it is optimized by a squared mean loss.

One limit to the performance of general recursive networks is that the gradient can explode or vanish, which induces instability and reduces the learning ability of the network. Since the network is optimized with back propagation, all the parameters are updated with the gradient chain. With the multiplicative rule of gradient chain, the gradient of parameters may explode or vanish. The authors of [10] tackled this problem with two strategies: recursive supervision and skip connection. Recursive supervision means that all the intermediate outputs from recursive block participate in the determination of output predictions, and each output prediction is supervised by a mean squared loss. The differences between the output predictions smooth the gradient of parameters. Moreover, the skip connection between the input of the network and the outputs of the recursive block makes that the network needs less recursion layers, thus it alleviates the gradient explosion and vanishing problem, according to [10]. The recursive supervision is a remedy for vanishing gradient, and the skip connection avoids gradient exploding. These ideas are similar to the ones on which VDSR is based. DRCN indeed reduces the number of parameters, nevertheless, the memory to save intermediate outputs can not be ignored.

Similarly to the DRCN, the Deep Recursive Residual Network (DRRN) [11] applies recursive learning. But contrary to DRCN, the recursive unit in DRRN is a modified resnet unit, as shown in Fig.4(c). The green dashed block denotes a modified resnet which consists of two convolution layers, and each convolution layer is a stack of batch normalization, ReLU activation function followed by a weight layer (convolution filters). The batch normalization outputs batches with zero mean value and standard deviation equal to 1. It helps to increase the learning rate and makes the network more robust to the initialization [23]. The red dashed block is the recursive block of the network. Within the recursive block, the convolution layers marked in the same colors share the same parameters. The skip connections in the architecture avoids the problem of gradient vanishing and explosion, moreover, it allows the network to learn complex functions. DRRN can be parametrized by B and U, where B is the number of the recursive blocks, U denotes the number of the recursive units within one recursive block. In Fig.4(c), the network contains 1 recursive block, and this recursive block has 2 residual units ($B = 1$, $U = 2$). In fact, the increase of recursive block quantity increases the number of parameters in the network, while the increase of residual unit quantity does not change the amount of parameters but increases the depth of the network. The authors in [11] noted that even though the DRRN networks are parameterized by B and U, the networks' efficiencies are comparable if their depths are similar.

3.4.3. New methods for the up-sampling operation

There are different ways to upscale images. Zhang *et al.* [54] used a convolution layer, at the beginning of the network, to interpolate low resolution images with bicubic method, i.e. the parameters at this layer are fixed. Alternatively, the up-sampling operation is performed with a convolution layer at the beginning of the network [55], where the parameters are updated during the optimization of the network. Upscaling low resolution images with transpose convolution layer is very flexible since the parameters in the convolution filters are adaptable during the training process. However, upscaling low resolution

images at the beginning of the network may increase the burden of calculation and slow down the efficiency of the network.

Accelerated SRCNN [13](FSRCNN), as depicted in 5, is an extension of SRCNN. Compared with SRCNN, the non-linear mapping function in FSRCNN is more flexible and robust. Moreover, two additional layers have been introduced to reduce the number of parameters while keeping the performance of the network. FSRCNN achieves real-time processing speed. Differing from ESPCN, FSRCNN applies deconvolution layer with stride of r for upsamping operation. It's noteworthy that the "deconvolution" here is a transposed convolution: low resolution features are spred into high resolution dimension with interval r, then convolve with the filters at the deconvolution layer.

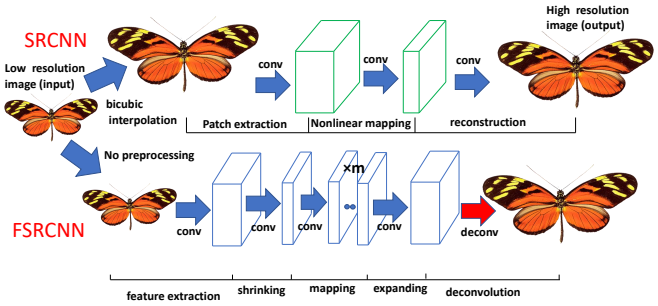

Figure 5: Illustration of FSRCNN architecture [13]

Shi *et al.* in [12] proposed Efficient Sub-Pixel Convolutional neural Network(ESPCN) where the input is the original low resolution images. ESPCN proceeds upsampling operation at the sub-pixel convolution layer.

The stride of sub-pixel convolution is $1/r$, where r is the up-sampling factor. The stride can be interpreted as the step size of convolution. The sub-pixel convolution increases features scale by rearranging the pixels learned in the former layer.

However, the shuffle operation in the sub-pixel convolution results in checkerboard artifacts. Shi *et al.* in a working notes proposed a modified sub-pixel convolution which is free from checkerboard artifacts: the upscale filters are initialized by imitating nearest neighbor interpolation methods, then the parameters are updated during the training process [56].

Conventional CNN learns shift-invariant filter, while in [57], the authors introduced Shepard layer to perform upsampling (or inpainting) operation in a shift-variant way. The input of this layer are interpolated to super resolution scale, but the interpolated values are 0. A mask is used to control the influence of filters.

ESPCN [12] and FSRCNN [13] are two pioneer works which integrate the upsampling operation within the network so that interpolation is not necessary. This is a big progress for the development of deep learning in super resolution problems. Researchers extended these two ideas in their work [43, 41, 42] to further improve the efficiency of deep learning based methods for super resolution problems. The methods proposed in [43, 41, 42] will be briefly introduced in the next section.

3.4.4. Resnet integrating new up-sampling ways

The Enhanced Deep Residual Networks (EDSR) [43] aims to use residual block to enhance the structural information at low resolution scale, then upscale images at the last second layer. The structure of EDRN is drawn in Fig.6. In the 'ResBlock', the authors removed the batch

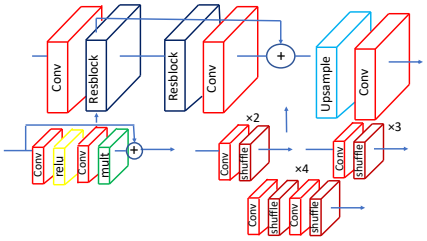

Figure 6: Illustration of EDSR network [43]. The ResBlock is stacked by a convolution layer, Relu activation, convolution layer and a constant scaling layer. The constant scaling layer simply rescales the residual with the purpose of keeping the stability of the network when the number of filters exceeded 1000 [58].

normalization and introduced a constant scaling layer (the green layer marked with 'Mult' in Fig.6). They explained that the suppression of batch normalization reduces memory consumption and keeps the range flexibility of features. A constant scaling layer is proposed in [58]. Szegedy *et al.* mentioned that when the number of filters exceeds 1000, the residual variants become instable and tend to lose their learning ability. Neither batch normalization nor decreasing the learning rate can solve this problem. Nevertheless, rescaling the residual with factor 0.1 before adding to the identity mapping seems to be able to keep the network stable. For this reason, there is a constant scaling layer in 'ResBlock'. The upsample operation is sub-pixel convolution, following the idea proposed in [12].

Lai *et al.* proposed a deep Laplacian Pyramid Networks (LapSRN) in [41] for super resolution problems. The main idea is to gradually upscale features. Its architecture has two branches: one for feature extraction, the other serves for reconstruction, as presented in Fig.7. In feature extraction branch, convolution layers marked in red abstract features' characters, deconvolution (transposed convolution) layers marked in blue upscale features. The output of deconvolution layers connects with 2 layers: one serves for residual information in the image reconstruction branch; the other is used for feature extraction for the next upsampling operation. The deconvolution layers in image reconstruction branch are initialized with bilinear kernel which is essential to force the feature extraction branch to learn residual features. It seems that the image

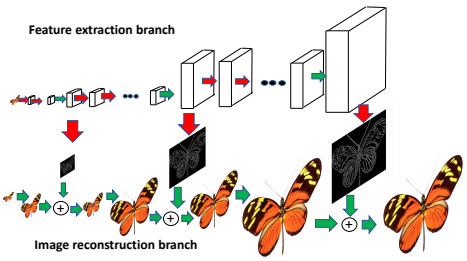

Figure 7: Illustration of LapSR [41]. Red arrows refers to convolution layers, blue arrows are transposed convolutions, green arrows are element-wise addition.

reconstruction branch is responsible to learn low frequency information, and the feature extraction branch refines the details and feeds high frequency information to the image reconstruction branch.

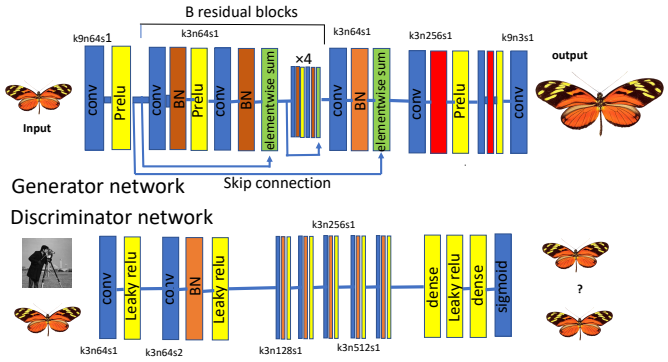

Figure 8: Illustration of SRGAN [24]. k represents the size of filter, n as the number of feature maps, s as the stride for each convolutional layer.

3.4.5. Densely connected network

The residual network or resnet unit plays an important role in the recent advanced deep learning network designed to solve super resolution problems. It can be seen that the skip connection has a significant impact on deep neural networks. Huang *et al.* introduced the densely connected CNN (denseNet) in [59] to solve object recognition problems, where more skip connections have been introduced compared with residual network or ResNets. Fig.9 demonstrates the framework of the denseNet. Differing from ResNets, the denseNet concatenates the outputs of former blocks while ResNets uses summation. The denseNet encourages to reuse the features, enhance signal propagation, as a result, less parameters are required and reduce the model size by employing small growth rate.

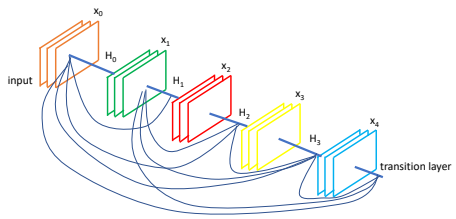

Figure 9: Illustration of densely connected CNN [59](denseNet).

The success of denseNet quickly draws the attention of researchers in the super resolution field [60, 42]. Tong *et al.* in [60] build a framework for super resolution problems using dense skip connections (SRdenseNet). The schema is illustrated in Fig.10. As explained in the caption of Fig.10,

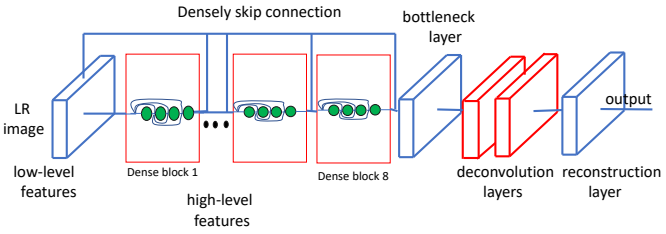

Figure 10: Illustration of the architecture of SRdenseNet [60]. It uses dense block to enhance image features then feed the enhanced features into deconvolution layers for upsampling. Long skip connections preserve low level features and enables the dense blocks to learn high level features.The bottleneck combines both low and high level features and feed the results into deconvolution layers.

the SRdenseNet extends the conception of denseNet to blocks for the purpose of abstracting high level features, and combines all the high level features with low level features via bottleneck layer, afterward passing though deconvolution layers for upsampling. The bottleneck reduces the dimensionality of the features. The way that the SRdenseNet combines low and high level of features is similar to the one in DRCN (Fig.4(b)), but differing in the reuse of parameters.

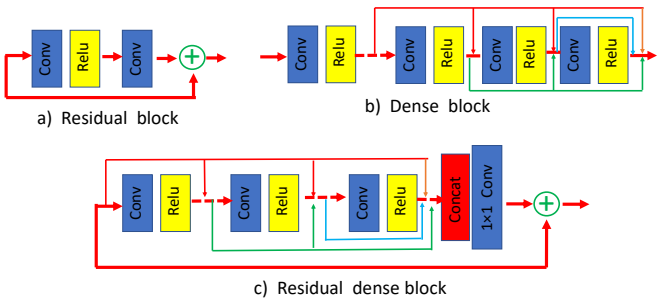

Figure 11: Comparison of different network blocks[42]. (a) Residual block in [43]. (b) Dense block in SRDenseNet [60]. (c) Residual dense block [42]

Zhang *et al.* proposed a residual dense network to

solve super resolution problem in [42](RDN). Fig.11 compares different network blocks: residual block, dense block, residual dense block. As can be seen in the Fig.11, residual dense block combines residual block and dense block. The 1×1 convolution layer is used to reduce the dimensionality. The residual method forces the filters to learn residual part information. For instance, in VDSR, the long skip connection conveys low frequency information to the output so that the convolution layers in the network are forced to learn high frequency information. Therefore, the learning task is simplified. The dense block boosts the ability of the network to describe complex functions. The residual dense block takes the advantages of both Residual block and Dense block, thus is expected to give a better performance.

Moreover, if we say the residual dense block is a micro network structure, Zhang *et al.* in [42] projects this structure into a macro scale. Fig.12 shows the entire architecture of RDN where the residual dense structure is flexibly used. There are three major differences between

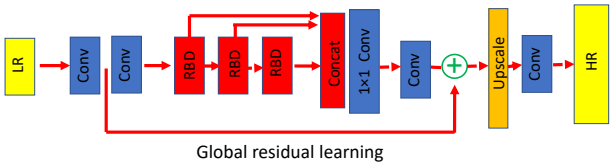

Figure 12: The illustration of the residual dense network[42].

SRDenseNet and RDN. First, in the scale of block unit: RDN has a long skip connection similar to residual block which SRDenseNet does not have. Second, the connections between block units: SRDenseNet does not have connections among the blocks while in RDN, the blocks are densely connected, following the architecture of denseNet. Third, in the scale of global structure: in SRDenseNet, the low and high level features are concatenated together before passing though the bottleneck; but in RDN, only the high level features are densely connected, then combined with low level features via residual structure.

3.4.6. New convolution operations

Improved convolution operations have been proposed to improve the super resolution results. The dilated convolution operation support exponential expansion of the receptive field with the same filter size. A large receptive field can be obtained with shalllow networks. The exploitation of contextual information is very efficient and it facilitates the generation of realistic details for super-resolutions. These new convolution operations have been used in [61, 62] and they achieve a much better performance.

3.4.7. Generative adversarial networks for super resolution

With GAN models the generator creates super resolution images that a discriminator has to distinguish as a real high resolution image or as an artifically super resolved one. If the discriminator can not see the difference between the estimated super resolution image and the high resolution image, this estimation is assumed to be a good approximation of the high resolution image. The discriminator thus constrains the estimation to follow some predefined distribution. With GAN models the PSNR values are degraded but the perceptual quality is generally improved. Several super resolution methods based on GANs have been investigated [24], [63]. SRGAN [24] uses a multi-component loss function with several parts, (1) a MSE loss that promotes pixels similarity, (2) a perceptual similarity distance based on deep network features (VGG network), (3) a standard GAN loss. VGG is a very deep CNN, it uses small filters to capture image features at different scales. With the increase of the layer's depth, the size of features decreases and the number of channels increases. Compared with mean square error distance, distance of features given by VGG network is less sensible to the changes at pixel level. This framework promotes super resolved images with a good perceptural qualtiy and close to the manifold of natural images.

SRFeat is another GAN super-resolution model with feature discrimination [63]. In this work, an additional discriminator is used to help the generator to generate high-frequency structural features rather than noisy artifacts. In the context of GANs, the work of Sajjad *etal.* follows a similar approach except with a different architecture [48]. Very recently, by leveraging the basic GAN framework, Yuan et al. [52] proposed an unsupervised super resolution algorithm with cycle gans [52].

It is well-known that the training process of GANs is a challenging task. To overcome this issue and stabilize the training, [53] propose a GAN network based on least square loss function with a gradual learning process from small upsampling factors to large upsampling factors. The output of each layer is gradually improved in the next layer.

The ESRGAN [49] is based on the SRGAN but incorporates dense blocks with residual connections between the input and the output of each block (residual in residual dense block) without batch normalization to facilitate training with a deeper network. A global residual connection is used to enforce residual learning. An enhanced discriminator is also employed in the model. A improved perceptual loss is introduced by using the VGG features.

3.5. Comparison of the performances of the different networks

Tab.1 compares the performance of different architectures in terms of PSNR. All the results are collected from the original papers. The Set5[64] and Set14[65] are two commonly used test sets for super resolution natural image benchmarks, including 5 and 14 images respectively. Fig.13 illustrates 4 images in the Set5 and 4 images in the Set14.

The RDN+ outperforms all the other methods in the table. The sign '+' denotes that the results are improved with self-similarity strategy [43]. As we know, during the training stage, data augmentation increases the amount of training data, thus boosts the network performance. Self-similarity has the similar principle, but it is employed during the reconstruction stage. It assumes that the network is invariable to the geometrical transformation of the input data. For example, a test sample has 8 augmented inputs after geometric transformation (including identity). These 8 samples will be fed into the network and generate 8 estimations. The final result is the average of these 8 estimations after the corresponding inverse geometric transformation. It's noteworthy that the invariability of geometrical transformation is a strong condition. It should be noted that such a comparison is only partially reliable sincd each network is trained in a different way

Table 1: Comparison of PSNR different methods for natural image sets set5 and set14, the low resolution images are generated with BI degradation model. The upsampling factor is 4. The sign '+' represents that the approach has been boosted with self-assemble. The results are collected from correspondent publications.

methods	set5	set14
SRCNN[2]	30.49	27.61
VDSN [9]	31.35	28.03
DRCN[10]	31.53	38.04
DRRN(B1U25)[11]	31.68	28.21
ESPCN[12]	30.90	27.73
FSRCNN[13]	30.55	27.50
EDSR[43]	32.46	28.80
EDSR+[43]	32.62	28.94
LapSRN[41]	31.33	28.06
SRGAN[24]	29.40	26.02
SRDenseNet[60]	32,02	28,50
RDN[42]	32.47	28.81
RDN+[42]	32.61	28.92

The different algorithms are generally evaluated on the peak signal-to-noise ratio (PSNR) and the structural similarity index [45]. The PSNR and SSIM are better for the ESRGAN [49], but a comparison is difficult since many factors as network complexity, depth of the networks, number of parameters are modified. Methods with late upsampling have a lower computational cost that methods that performs the upsampling earlier [13, 12, 43]. Almost all recent super resolution methods obtain improved performance by adding more weights and layers [43, 11]. It is generally found that the network depth contribute to a better PSNR and image quality [43]. [10, 11] , [42].

Though FSRCNN is a significant improvement in speed over SRCNN, recent studies with densely connected networks showed that more sophisticated network structures with skip connections and layer reusing benefit not only performance and speed, but also reduces training time. Several types of skip connections are encountereed in deep networks, global connections, local connections, recursive connections and dense connections They have improved drastically SR results. VDSR [9] was based on global residual learning and improved much the SRCNN [13]. The effectiveness of recursive connections was shown in [10, 11]. Local residual connections were used in [43]. Similarly ESRGAN [49], and [42], use dense and global connections

As noticed in [24], the L_2 or L_1 loss are not optimal. Taking into account only the intensity differences can not reflect the perceptual quality. In this context, recent GAN models have obtained state-of the art super resolution results [53], [49].

In this section, we introduced some deep learning networks for super resolution problems on natural image sets. The network evolved from the SRCNN to residuel based network. The progress in the way of up-sampling led to the development of the residual based architecture, and the network proposed today tends to be densely connected.

4. Applications of deep learning in medical imaging super resolution problems

In medical image processing, various factors may have an impact on the spatial resolution of and image depending on the modality. The literature on deep learning methods in medical images super resolution is recent, yet there is already a strong interest on this topic of applications for various imaging modalities, such as CT, MRI, retinal vascular fundus image [66, 67] (online dataset: `http://www.eyepacs.com`), electron microscopy [68] and endoscopy [55].

4.1. CT images

The spatial resolution of *in vivo* CT image scanning is limited because of the scan time, body motion, or dose limit and DL techniques can be very useful to improve it.

Umehara *et al.* investigated the application of SRCNN on CT chest images [69]. The HR images were experimental images from The Cancer Imaging Archive (TCIA), LR images were simulated based on HR images. They showed that SRCNN outperforms traditional linear interpolation methods.

Park et al. [70] came up with a modified U-net to solve super resolution and denoising problems for 2D brain CT images. HR images were obtained from PET CT, LR images were generated by taking the average of 3D high resolution slices. The proposed network architecture is illustrated on Fig.14 and is based on U-net usually used

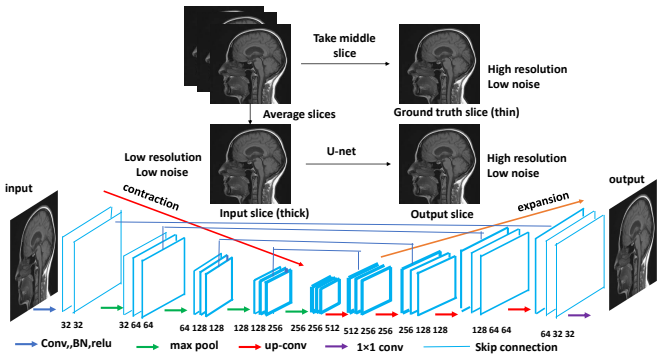

Figure 14: Illustration of the CNN architecture proposed by Park *et al.* [70]. 2D low resolution slices are generated by taking the average of 3D high resolution slices. As can be seen in the figure, on the contraction path, images characters are synthesized at different scales, and on the expansion path, super resolution images are gradually reconstructed.

for image segmentation [7]. U-net displays a 'U' shape, the downward path shrinks feature size while increasing channel number, the upward path increases feature size and concatenates the corresponding features extracted on the contraction path. Compared with original U-net, the modified U-net architecture in [70] has additional batch normalization, which speeds up convergence and efficiently avoid local minima due to improper initialization.

Mansoor *et al.* have investigated the application of SRGAN [24] for two imaging modalities (CT and MRI) in [71]. They applied VGG-like network to abstract the features, thus avoiding to enhance the similarity based on pixel-level. The basic mechanism is similar to the SRGAN introduced in the last section. The network was trained with 2D slices in three planes. This research work would be even more interesting if a comparison with 2D and 3D was given. In their training set, the LR images were generated by down sampling the high resolution images smoothed with a Gaussian kernel.

You *et al.* proposed a GAN-CIRCLE to solve CT super resolution problems [72]. As mentioned in the paper, one potential risk of GAN is that the generator may give an estimation following a good distribution which does not match the input image. Authors in [72] proposed a circle structure which ensures that, apart from probability distribution, the output image is corresponding to the input image. Concretely, two GAN networks are linked together and form a circle. Given an input signal x, the output y of the forward GAN network will go through the backward GAN structure and provide an estimation $\hat{x}$ which is close to the original input signal x. Similarly for the inverse order. Moreover, the total variation regularization has been added to the loss. Furthermore, if a high resolution image is fed into the forward generator, the output is assumed to be similar to its input. This assumption may enhance the robustness of the network. The dataset in this work was composed of a tibia dataset and an abdominal dataset. Like in most the previously discussed works, the LR images were simulated from experimental HR images.

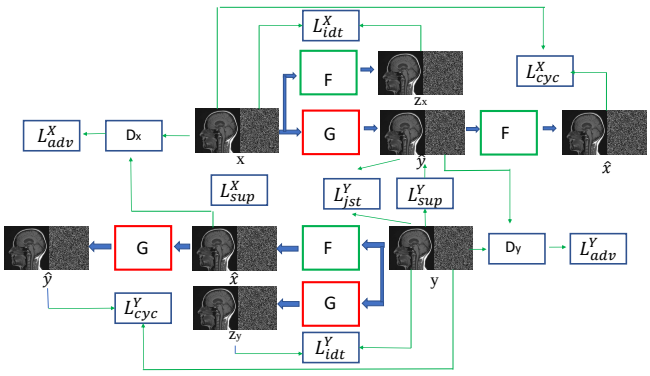

Figure 15: Illustration of GAN-CIRCLE network[72]. x is the noisy LR input, y is the HR image, G is the generator of forward Gan, from LR towards HR, F is the generator of the backward GAN, from HR towards LR. D_x and D_y are discriminators of backward and forward discriminators. L_{sup}, L_{adv}, L_{idt} and L_{jst} are four losses in the entire work, where L_{sup} corresponds to generator loss supervised, L_{adv} is the adversarial loss, L_{idt} is a loss forcing $F(x)$ to be close to x, $G(y)$ close to y. The lose L_{jst} integrates total variation regularization to reduce noise on the results.

4.2. MRI images

The spatial resolution of MRI images may be degraded due to the constraints such as image scan time, body motion, patients' comfort considerations, hardware configurations. The applications of deep learning for MRI SR problems mainly involves brain [4, 73, 74, 75, 76, 77, 78, 79] and cardiac images [80, 81, 82]. In these researches, low resolution images can be experimental images [75, 80, 79] or simulation images [4, 73, 74, 75, 76, 77, 78, 79, 81]. Precisely, simulated low resolution images can be generated via k-space truncation or down-sampling high resolution images with or without Gaussian blurring. We summarize online MRI images sets in the following list:

1. `http://brainweb.bic.mni.mcgill.ca/brainweb/` (simulation images)
2. `http://www.bic.mni.mcgill.ca/brainweb/` (simulation image)
3. `https://sites.google.com/site/brainseg/` (experimental images)
4. `https://www.smir.ch/BRATS/Start2015` (experimental images)
5. `http://brain-development.org/ixi-dataset/` (experimental images)
6. `http://insight-journal.org/midas/collection/view/190` (experimental images)
7. `https://github.com/UK-Digital-Heart-Project` (UK Digital Heart Project, experimental images)
8. `http://hdl.handle.net/1926/1687` (experimental images)

4.2.1. 2D MRI images

Zeng *et al.* [78] worked on a method estimating single-contrast and multi-contrast MRI images simultaneously. Single-contrast sub-network solves super resolution problem of low resolution T2 images, the multi-contrast sub-network estimates multi-contrast T2 images based on the reference T1 images and T2 super resolution images. HR images are MRI brain images, LR images are simulation data.

Shi *et al.* [75] used local residual block and global residual network to extend SRCNN to solve a 2D MRI SR problem. The investigated datasets include the first three datasets in the dataset list.

Zhao *et al.* [76] extended a SRCNN architecture for 2D MRI brain images. The network consists of three main sub-networks: feature extraction sub-network, non-linear mapping sub-network and reconstruction. The non-linear mapping sub-network is composed of a set of cascaded Channel splitting blocks (CSB). Each block follows merge-and-run (MAR) strategy, it splits features into 2 branches, precisely, one for densenet, the other for residual learning. MAR has been proposed with different structures, but the general idea is to split features at one channel into two branches so that the features at the same channel can be processed differently, for instance, using different convolutional filter size or different structures to extract features. In this paper, the authors proposed to use two branches with different network structures, they argued that this splitting channel structure is beneficial from the advantages of both residual learning and densenet learning: the former one enhances the reuse of features and stabilize the network, the latter one explores new characters of features. Alternatively speaking, the residual learning branch learns residual information (high frequency information), while the densenet learning branch extracts features' characters directly. A fusion layer is added at the end of CSB, which merges the features generated from the two branches together. Additionally, there is a skip connection from the input of the block to the output of the block, authors stated that such multi-level residual network favors the stabilization of the network and slightly improve the performance of the network. Since it is hard to keep the network stable for MR images (limited image quality), the multi-scale residual network is even more meaningful. The work is performed with IXI dataset (5th dataset in the dataset list).

Liu *et al.* put forward a multi-scale fusion convolution network (MFCN) [79]. The network has several multi-scale fusion units (MFU). Each MFU corresponds to an estimation obtained with filters at a specified scale. A set of MFU helps to reconstruct super resolution images with different scale features. The '+' operation in the fusion layer forces each MFU to scale their magnitude respectively. The network was tested on both simulation and real LR image sets.

Oktay *et al.* came up with T-L network [80], which was used for image segmentation or SR problems for 2D cardiac images. The involved dataset is provided by UK Digital Heart Project. Both HR and LR images were experimental images. The schema of T-L network is presented in Fig.16. Two losses are introduced to integrate shape prior and improve the accuracy of estimation at pixel level. The block "Segmentation" and "Super Res" in Fig.16 enhance the similarity between the estimation and the reference at pixel-level, the correspondent loss is "X-Entropy Loss L_x" for segmentation, "Smooth L_1 loss" for super resolution problem. The shape prior is integrated via a perceptual loss. The idea is to non-linearly transform the estimation and the reference into a low dimension space and to penalize the dissimilarity between them. In SRGAN, the perceptual loss is considered in the feature space described by VGG, since VGG is a deep network, this feature space is relatively huge compared to the space described by encoder. The segmentation results obtained with the proposed T-L model outperforms other methods in the literature.

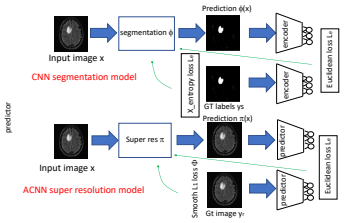

Figure 16: The proposed T-L network segmentation and super resolution tasks[80].

4.2.2. 3D MRI images

Pham *et al.* in [4] attempted to perform SR on 3D MRI brain images with SRCNN framework. A comparison between 2D and 3D SRCNN is given. The authors concluded that SRCNN 3D is always better than SRCNN 2D. The comparison between the networks trained on natural images and on MRI images was given as well. Oktay *et al.* applied 3D residual network for SR of cardiac images [81]. The LR images were generated from HR MRI images. The proposed network was similar to VDSR. They proposed to use a deconvolution layer to replace the interpolation operation at the beginning of the network. Multi-input allows to use image information along different directions. Zhao *et al.* extended EDSR [43] in MRI brain super resolution reconstruction [83]. In their original database, the images in axial plane is at high resolution, but the saggital and coronal plan are at low resolution scale. They artificially degraded the high resolution image in axial plane to generate low resolution images in axial plane. Afterward, they trained EDSR with the paired low and high resolution images which are in axial plane. Then the low resolution image in sagittal and coronal plane will be reconstructed via the trained EDSR model. This approach assumes that the degradation kernel is isotropic in three dimensions, which

is crucial for the quality of reconstruction.

Based on this work, Zhao *et al.* proposed an anti-aliasing self super resolution method [84]. Two EDSR networks are trained in this method, one for self super resolution (SSR) task, the other for self anti-aliasing (SAA) task. In the training set, the slices at xy-plan of 3D volumes were regarded as HR images, both LR and aliased LR images were simulated based on HR images. In the reconstruction stage, the network solving SAA was applied to the slices at xz-plan, then the network solving SSR was applied to the slices at yz-plan. A data augmentation (rotation) strategy was applied during training and reconstruction stages. Both artificial and experimental images were tested. The obtained results indicated that a combination of SAA and SSR outperformed the one only based on SSR. As the authors highlighted, no external training data was needed in this approach. By exploring the HR information from 2D slices at a specific direction and transferring the learnt knowledge to the slices at other directions, this work presented a new way to determine 3D SR volume via 2D slices.

Sanchez *et al.* in [77] applied GAN network to solve SR problem of 3D MRI brain image. Different up-sampling methods have been compared, including nearest neighbor interpolation followed by convolution (NNC), sub-pixel convolution (SPC) [12] and modified sub-pixel convolution (MSPC) [56]. Their results indicated that the NNC has the best performance in terms of SSIM, while the MSPC returns best PSNR. The LR images were simulation images.

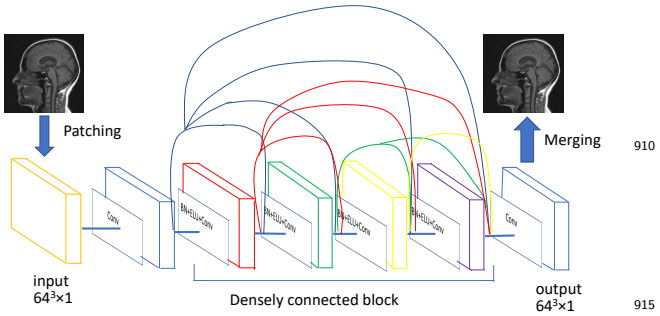

Figure 17: Framework of 3D DCSRN[74].

Chen *et al.* in [74] proposed a densely connected super resolution network (DCSRN) for brain MRI images, similar to the denseNet presented in Section 3.4.5. The LR images were simulated from experimental MRI images. They concluded that the 3D neural networks were better than their 2D counterparts, and the proposed method outperformed 3D FSRCNN. Yet, very few details were given. The proposed structure is demonstrated in Fig.17. It can be seen that the network is densely connected, the output of the each block will be reused in latter blocks. The authors resumed 3 advantages of their proposed network: faster training since the path is shorten; tiny model because of weight sharing; less overfitting due to the reuse of features. Chen *et al.* extended this work to multi-level DCSRN-GAN in [73]: the single DSCRN block in [74] was replaced with a set of shallow DCSRN blocks. The multi-level DSCRN is integrated into GAN as generator.

Some existing studies directly use directly the 2D network structure and apply it slice by slice. However, to fully solve the ill-posed super resolution problem, a 3D model is more preferable as it can directly extract 3D structural information. The recent studies [4], [74], [77] have shown that a 3D CNN outperforms its 2D counterpart since they completely exploit the 3D volume information. The network architectures use 3D convolutions. With the additional dimension introduced by a 3D CNN, the number of parameters of the network grows rapidly. The performance of a deep network generally improves with more layers and weights but with 3D the model becomes computationally expensive. Densely connected network has an efficient memory usage and is practical for 3D images.

4.3. Other modalities

Mahapatra *et al.* combined GAN network with a salient map [67] for SR in retinal vascular images. The LR images in the dataset are simulation images. The salient map was estimated based on image curvature map, element distribution (pixel entropy). The curvature map contained gradient and second order derivative information of features, the probability used for pixel entropy was estimated via the histogram within a small window. The salient map permits to minimize a weighted mean squared error between the estimation and the ground truth. Moreover, it is also regarded as a prior information. The global architecture of [67] follows GAN framework.

Since the hyper-parameters (weight and window size) used in salient maps are very crucial but may vary for different samples, the performance of the proposed network was not consistent in the entire dataset. Mahapatra *et al.* applied successive GAN to get rid of empirically selected hyper-parameters in the salient maps [66]. At least two GAN networks were cascaded in the architecture, each of them upscales images with factor 2. The previous GAN provides inputs to the sequel GAN. A triplet loss has been considered in the loss function, which aims to minimizing the distance between the estimation of the last GAN and the target while maximizing the difference between the estimation of the last GAN and its input. In this work, LR images are simulation images.

Lin *et al.* solved a super-spectral resolution problem in endoscopic depth measurement [55]. Both HR and LR images were experimental images. A network solving SR problems was applied along the spectral dimension (model 1), then the parameters in model 1 were frozen and the follow-up network which is used to merge RGB and sparse spectral signals (model 2) was trained. Afterwards, all the parameters in model 1 and 2 were updated till convergence. This work shows an example about how to integrate information from different image domains. The

super resolution algorithms developped for RGB images can very often be applied to medical images acquired with different modalities. Yet, the hyperspectral images have a very low spatial resolution. The classical deep learning methods can not generate pixel-level dense multispectral images with a good robstness. It is necessary to integrate the information from dense RGB images to increase the spatial resolution of hyperspecrtral images.

The influence of DL technique in medical images tends to be wide and deep. The applications of DL methods for SR problems in medical images, presented in this section, are not exhaustive.

5. Discussion

Deep learning methods have a vast field of application in medical image processing tasks [8], such as classification, detection, segmentation, registration. With the development of DL methods for SR, researchers in medical image have first proposed applications, then also new architectures which permit to integrate priors to boost the performance of the network and facilitate the follow-up analysis and research.

However, lack of high quality references and particular image priors or constraints are two main bottlenecks to generalize deep learning SR methods in medical imaging. HR ground truth images are difficult to obtain, in clinical imaging due to the various limitations. Meanwhile, the choice of the appropriate priors is also important and may be specific to a given modality. In the following, we summarize some methods for data augmentation and different ways to inject particular priors.

5.1. Data paucity

Data augmentation is used to increase the dataset. Conventional data augmentation consists of flipping, rotation, symmetric, translation, scaling images, adding additive noise, changing brightness, adapting contrast, gamma-transformation, modifying colors and so on.

Besides these operations, researchers considered to use DL networks to perform data augmentation. Zhang *et al.* applied transfer learning [54] to increase the dataset. They use scale-invariant feature transform (SIFT) to extract characteristic features over the available medical image set, then search the similar features among the natural image set. The matching subregions will be added into the dataset.

Lemley *et al.* came up with "smart augmentation" [85] to solve the data paucity problem: a network is employed to generate new images based on 2 or more samples in the dataset. When images are classified with different labels, several networks can be trained so that each class has a different way to synthesize its new images. Shin *et al.* put forward to use GAN to synthesize medical images for segmentation tasks in [86].

Cubuk et al. attempted to learn an effective strategu to augment data automatically [87]. The proposed method uses a controller to predict 5 sub-policies. The controller is a recurrent neuron network, and each sub-strategy has two basic operations for data augmentation, such as translation or rotation. Every candidate strategy includes a set of sub-strategu which are characterized by 2 parameters: probability to apply the strategy and the magnitude used in the strategy. The controller will be trained to choose sub-stategy for a given database based on reward signals. A child model is trained with data augmented with the selected sub-strategy, and its final performance over a validation set is regarded as reward signals to optimize the controller. Readers can refer to [87, 88] for more details.

Data augmentation is a powerful tool facing small dataset. It's noteworthy that the way of data augmentation should be strictly consistent with medical applications. For example, when the color of tissue is very sensitive for the diagnosis conclusion, data augmentation based on color may degrade the performance of the tested approach.

5.2. Adding prior

Both natural images set and medical images set need priors to boost the performance of SR techniques. Liang *et al.* [89] integrated gradient prior by introducing a feature extraction layer where the parameters are fixed. Meanwhile, salient maps, investigated in [67], can be also regarded as handcrafted prior information. As previously mentioned, one drawback of this prior is that it contained hyper-parameters which are very crucial but need to be chosen empirically for every sample in the dataset.

As previously mentioned, the T-L network proposed in [80] used auto encoder to integrate image constraints. Different from handcrafted priors such as gradient information, the prior considered in this paper is described with an auto encoder. Both SR image and the ground truth are encoded into a low dimension space and their compacted features are optimized to be similar.

Lehtinen *et al.* in [90] proposed a denoising method in which the network is trained only based on noisy images, i.e. no clean image is in the dataset. During the training process, both input and output are noisy images including the same clean image but with different noise. In the test stage, a noisy image will be sufficient for the network to accomplish the denoising task. One important assumption of this work is that the average value of noise is 0. The authors explained that if the average of noise is 0, then the average of a set of noisy images is a denoised image. Since the network is trained with stochastic gradient descent methods, although every example in the batch corresponds to an inaccurate optimization direction, the average of these examples gradients points toward a correct direction.

Ulyanov *et al.* considered to use untrained deep learning networks to explore deep priors from given degraded images [91]. Precisely, the inputs are a set of uniformly distributed random noise features, parameters in the network are randomly initialized and are optimized by minimizing the L^2 distance between the output of the network and

the given degraded image. Intuitively, the network can be compared to a regularizer in the optimization task, but being more flexible than handcrafted regularization terms such as total variation regularization. This approach has been tested on several tasks, including denoising, super resolution, inpainting. Its performance in denoising and inpainting tasks are very impressive, but becomes relatively limited for SR, because no further detail information is provided.

Both [90, 91] investigate how to improve image quality in the absence of ground truth. These methods allows to find deep prior hidden in the noisy images. Such priors can be used to solve SR problems.

5.3. Conclusion

In this review, we have briefly gone through the state of the art of deep learning for SR of natural images, and presented applications and developments for medical images. The main challenges, data paucity and how to add priors have finally been discussed. Deep learning methods show great potential to solve SR in medical image field, despite many challenges, the performance of SR techniques become more and more promising.

6. Acknowledgement

This work is supported by China Scholarship Council (CSC) and LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Thanks all the authors who approved our applications to reproduce their published figures.

References

[1] K. Nasrollahi, T. B. Moeslund, Super-resolution: a comprehensive survey, Machine vision and applications 25 (2014) 1423–1468.

[2] C. Dong, C. C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, Springer, 2014, pp. 184–199.

[3] Y. Luo, L. Zhou, S. Wang, Z. Wang, Video satellite imagery super resolution via convolutional neural networks, IEEE Geoscience and Remote Sensing Letters 14 (2017) 2398–2402.

[4] C.-H. Pham, A. Ducournau, R. Fablet, F. Rousseau, Brain MRI super-resolution using deep 3D convolutional networks, in: Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, IEEE, 2017, pp. 197–200.

[5] K. H. Jin, M. T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, IEEE Transactions on Image Processing 26 (2017) 4509–4522.

[6] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss, IEEE transactions on medical imaging 37 (2018) 1348–1357.

[7] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. Van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, Medical image analysis 42 (2017) 60–88.

[9] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[10] J. Kim, J. Kwon Lee, K. Mu Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.

[11] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 1, 2017, p. 5.

[12] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.

[13] C. Dong, C. C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: European Conference on Computer Vision, Springer, 2016, pp. 391–407.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[15] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, IEEE Transactions on Image Processing 26 (2017) 3142–3155.

[16] M. T. McCann, K. H. Jin, M. Unser, A review of convolutional neural networks for inverse problems in imaging, arXiv preprint arXiv:1710.04011 (2017).

[17] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning., in: OSDI, volume 16, 2016, pp. 265–283.

[18] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, Z. Zhang, Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems, arXiv preprint arXiv:1512.01274 (2015).

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 675–678.

[20] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436.

[21] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, nature 323 (1986) 533.

[22] S. N., H. G., K. A., S. I., S. R., Dropout: a simple way to prevent neural network from overfitting, Journal of Machine Learning Research 15 (2014) 1929–1958.

[23] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167 (2015).

[24] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, arXiv preprint (2016).

[25] A. B.O., Z. J.M., Deep learning of constrained autoencoders for enhanced understanding of data, IEEE Transactions on neural networks and learning systems 29 (2017) 3969–3979.

[26] K. A., H. J.A., Weight decay can improve generalization, Advance Computer Visual and Image Processing 4 (1992) 950–957.

[27] M. P., T. S., K. T., A. T., K. J., Pruning convolutional neural networks for resource efficient inference, arXiv preprint arXiv:1611.06440 (2016).

[28] Y. A., S. T., A. W., N. S., S. Y., T. A., Adam induces implicit weight sparsity in rectifier neural network, in: IEEE Interna-

tional Conference on Machine Learning and Applications, 2018, pp. 318–325.

[29] M. D., K. K., T. C., On implicit filter level sparsity in convolutional neural networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 520–528.

[30] A. B.O., I. T., Z. J.M., Regularizing deep neural networks by enhancing diversity in feature extraction, IEEE Transactions on neural networks and learning systems 30 (2019) 2650–2661.

[31] C. M., A. F., G. R., Z. L., B. D., Reducing overfitting in deep networks by decorrelation representations, arXiv preprint arXiv:1511.06068 (2015).

[32] F. E., G. J, Q. Y., D. C., Suppressing model overfitting for image super-resolution network, in: CVPR, 2018.

[33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.

[34] A. M., C. S., B. L., Wasserstein gan, arXiv preprint arXiv:1701.07875 (2017).

[35] G. I., A. F. A. M., D. V., C. A.C., Improved training of wasserstein gans, in: NIPS 2017, ????

[36] T. Miyato, K. T., K. M., Y. Y., Spectral normalization for generative adversarial networks, arXiv preprint arXiv:1802.05957 (2018).

[37] N. K., M. T.B., Super-resolution: a comprehensive survey, Machine vision and applications 25 (2014) 1423–1468.

[38] F. S., R. M.D., E. M., M. P., Fast and robust multiframe super resolution, IEEE transactions on Image Processing 13 (2004) 1327–1344.

[39] B. S.P., G. N.P., K. A.K., Maximum a posterio video super-resolution using a new multichannel image prior, IEEE transactions on image processing 19 (2010) 1451–1464.

[40] K. M., B. P., P. S., H. K., K. D., N. J., Deep learning for multiple-image supe-resolution, arXiv preprint arXiv:1903.00440 (2019).

[41] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep laplacian pyramid networks for fast and accurate super resolution, in: IEEE Conference on Computer Vision and Pattern Recognition, volume 2, 2017, p. 5.

[42] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[43] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, Enhanced deep residual networks for single image super-resolution, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, volume 1, 2017, p. 3.

[44] Z. H., G. O., F. I., K. J., Loss functions for neural networks for image processing, IEEE Transactions on Computational Imaging 3 (2017).

[45] W. Z., B. A.C, S. H.R, S. E.P, Image quality assessment form error visibility to structural similarity, Transactions on Image processing (2004).

[46] S. H.R, S. M.F., B. A.C., A statistical evaluation of recent full reference image quality assessment algorithms, Transactions on Image Processing (2006).

[47] J. J., A. A. L. F.F, Perceptual losses for real-time style transfer and super resolution, 2016, pp. 694–711.

[48] S. M.S, Scholkopf, H. M., Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4501–4510.

[49] W. X., Y. K., W. S., G. J., L. Y., D. C., L. C.C., Q. Y., T. X., Esrgan: enhanced super-resolution generative adversarial networks, in: ECCVW, 2018.

[50] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[51] H. K., Z. X., R. S., S. J., Deep residual learning for image recognition, CVPR (2016).

[52] Y. Y., L. S., Z. J, Z. Y., D. C., L. L., Unsupervised super-resolution using cycle-in-cycle generative adversarial networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 814–823.

[53] S. P., Z. M., R. W. D. J. Y., G-ganisr gradual generative adversarial net for image super resolution, Neurocomputing 366 (2019) 140–153.

[54] Y. Zhang, M. An, Deep learning-and transfer learning-based super resolution reconstruction from single medical image, Journal of healthcare engineering 2017 (2017).

[55] J. Lin, N. T. Clancy, Y. Hu, J. Qi, T. Tatla, D. Stoyanov, L. Maier-Hein, D. S. Elson, Endoscopic depth measurement and super-spectral-resolution imaging, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 39–47.

[56] A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, W. Shi, Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize, arXiv preprint arXiv:1707.02937 (2017).

[57] J. S. Ren, L. Xu, Q. Yan, W. Sun, Shepard convolutional neural networks, in: Advances in Neural Information Processing Systems, 2015, pp. 901–909.

[58] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning., in: AAAI, volume 4, 2017, p. 12.

[59] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks., in: CVPR, volume 1, 2017, p. 3.

[60] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, in: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE, 2017, pp. 4809–4817.

[61] H. Z., W. L., P. C., Image super-resolution via deep dilated convolutional networks, in: International Conference on Image Processing, 2017.

[62] S. P., Zareapoor, Z. J, Y. Y, Image super ressolution by dilated dense progressive network, Image and vision computing 88 (2019) 9–18.

[63] P. S.J, S. H. C. S., H. K.S, L. S., Srfeat: single image super-resolution with feature discrimination, in: ECCV, 2018.

[64] M. Bevilacqua, A. Roumy, C. Guillemot, M. L. Alberi-Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012).

[65] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: International conference on curves and surfaces, Springer, 2010, pp. 711–730.

[66] D. Mahapatra, B. Bozorgtabar, R. Garnavi, Image super-resolution using progressive generative adversarial networks for medical image analysis, Computerized Medical Imaging and Graphics 71 (2019) 30–39.

[67] D. Mahapatra, B. Bozorgtabar, Retinal vasculature segmentation using local saliency maps and generative adversarial networks for image super resolution, arXiv preprint arXiv:1710.04783 (2017).

[68] L. Heinrich, J. A. Bogovic, S. Saalfeld, Deep learning for isotropic super-resolution from non-isotropic 3D electron microscopy, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 135–143.

[69] K. Umehara, J. Ota, T. Ishida, Application of super-resolution convolutional neural network for enhancing image resolution in chest CT, Journal of digital imaging (2017) 1–10.

[70] J. Park, D. Hwang, K. Y. Kim, S. K. Kang, Y. K. Kim, J. S. Lee, Computed tomography super-resolution using deep convolutional neural network, Physics in Medicine and Biology (2018).

[71] A. Mansoor, T. Vongkovit, M. G. Linguraru, Adversarial approach to diagnostic quality volumetric image enhancement, in: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE, 2018, pp. 353–356.

[72] C. You, Y. Zhang, X. Zhang, G. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, P. K. Saha, G. Wang, CT super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle), arXiv preprint arXiv:1808.04256 (2018).

[73] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, D. Li, Efficient and accurate mri super-resolution using a generative adversarial network and 3D multi-level densely connected network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 91–99.

[74] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, D. Li, Brain mri super resolution using 3D deep densely connected neural networks, in: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE, 2018, pp. 739–742.

[75] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, H. Xu, Super-resolution reconstruction of mr image with a novel residual learning network algorithm, Physics in Medicine & Biology 63 (2018) 085011.

[76] X. Zhao, Y. Zhang, T. Zhang, X. Zou, Channel splitting network for single mr image super-resolution, arXiv preprint arXiv:1810.06453 (2018).

[77] I. Sanchez, V. Vilaplana, Brain mri super-resolution using 3D generative adversarial networks (2018).

[78] K. Zeng, H. Zheng, C. Cai, Y. Yang, K. Zhang, Z. Chen, Simultaneous single-and multi-contrast super-resolution for brain mri images based on a convolutional neural network, Computers in biology and medicine (2018).

[79] C. Liu, X. Wu, X. Yu, Y. Tang, J. Zhang, J. Zhou, Fusing multi-scale information in convolution network for mr image super-resolution reconstruction, Biomedical engineering online 17 (2018) 114.

[80] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan, et al., Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation, IEEE transactions on medical imaging 37 (2018) 384–395.

[81] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D. O'Regan, D. Rueckert, Multi-input cardiac image super-resolution using convolutional neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 246–254.

[82] A. Giannakidis, O. Oktay, J. Keegan, V. Spadotto, I. Voges, G. Smith, I. Pierce, W. Bai, D. Rueckert, S. Ernst, et al., Super-resolution reconstruction of late gadolinium enhancement cardiovascular magnetic resonance images using a residual convolutional neural network (????).

[83] C. Zhao, A. Carass, B. E. Dewey, J. L. Prince, Self super-resolution for magnetic resonance images using deep networks, in: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE, 2018, pp. 365–368.

[84] C. Zhao, A. Carass, B. E. Dewey, J. Woo, J. Oh, P. A. Calabresi, D. S. Reich, P. Sati, D. L. Pham, J. L. Prince, A deep learning based anti-aliasing self super-resolution algorithm for mri, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 100–108.

[85] J. Lemley, S. Bazrafkan, P. Corcoran, Smart augmentation learning an optimal data augmentation strategy., IEEE Access 5 (2017) 5858–5869.

[86] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, M. Michalski, Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2018, pp. 1–11.

[87] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q. V. Le, Autoaugment: Learning augmentation policies from data, arXiv preprint arXiv:1805.09501 (2018).

[88] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, Learning transferable architectures for scalable image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8697–8710.

[89] Y. Liang, J. Wang, S. Zhou, Y. Gong, N. Zheng, Incorporating image priors with deep convolutional neural networks for image super-resolution, Neurocomputing 194 (2016) 340–347.

[90] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, T. Aila, Noise2noise: Learning image restoration without clean data, arXiv preprint arXiv:1803.04189 (2018).

[91] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.

Figure 13: Illustration of 8 images in the set5 as well as set14. The images on the top row are 4 images from the Set5. The images on the bottom row are 4 images from the Set14.