Tech Science Press

Deep Residual Network Based on Image Priors for Single Image Super Resolution in FFA Images

G. R. Hemalakshmi[*], D. Santhi, V. R. S. Mani, A. Geetha and N. B. Prakash

National Engineering College, Kovilpatti, 628503, India
[*]Corresponding Author: G. R. Hemalakshmi. Email: grhemalakshmi@gmail.com

Abstract: Diabetic retinopathy, aged macular degeneration, glaucoma etc. are widely prevalent ocular pathologies which are irreversible at advanced stages. Machine learning based automated detection of these pathologies facilitate timely clinical interventions, preventing adverse outcomes. Ophthalmologists screen these pathologies with fundus Fluorescein Angiography Images (FFA) which capture retinal components featuring diverse morphologies such as retinal vasculature, macula, optical disk etc. However, these images have low resolutions, hindering the accurate detection of ocular disorders. Construction of high resolution images from these images, by super resolution approaches expedites the diagnosis of pathologies with better accuracy. This paper presents a deep learning network for Single Image Super Resolution (SISR) of fundus fluorescein angiography images, modeled on residual learning, gridded interpolation and Swish activation functions. The image prior for this network is constructed by gridded interpolation which provides better image fidelity compared to other priors. Evaluation of the performance of this network and comparative analysis with benchmark architectures, on a standard dataset shows that the proposed network is superior with respect to performance metrics and computational time.

Keywords: SISR; FFA; residual network; gridded interpolation; swish function

Abbreviations

AE:	Auto Encoders
CNN:	Convolution Neural Network
DL:	Deep Learning
DNN:	Deep Neural Network
DR:	Diabetic Retinopathy
FFA:	Fundus Fluorescein Angiography
FoV:	Field of View
FSRCNN:	Fast Super Resolution Convolution Neural Network
GSM:	Gaussian Scale Mixture
HR:	High Resolution
LiDAR:	Light Detection and Ranging

LR: Low Resolution
LSTM: Long Short Term Memory
MISR: Multi Image Super Resolution
NARM: Nonlocal Auto Regressive Model
NPDR: Non-Proliferative Diabetic Retinopathy
PDR: Proliferative Diabetic Retinopathy
ReLU: Rectified Linear Unit
RNN: Recurrent Neural Network
SASC: Structured Analysis Sparse Coding
SGDM: Stochastic Gradient Descent with Momentum
SISR: Single Image Super Resolution
SRCNN: Super Resolution Convolution Neural Network
SRF: Super-resolution Random Forest
SSC: Simultaneous Sparse Coding
SSOCTA: Swept Source Optical Coherence Tomography Angiography
TPS: Thin Plate Spline
VDSR: Very Deep Super-resolution Network

1 Introduction

Advancements in medical imaging have significantly improved the early detection of several pathologies of the eyes including Diabetic Retinopathy (DR), glaucoma, sclerosis, macular degeneration, etc. DR, one of the widely prevalent eye diseases manifests with disorders in retinal capillaries, called microaneurysms. Similarly, flecks of oily secretions called exudates appear on the retinal vision region due to extreme permeability of retinal vascular structures. Various imaging modalities such as fluorescein angiography, direct and indirect ophthalmoscopy, optical coherence tomography and stereoscopic color film photography are employed in the early detection and classification of several ocular disorders.

Fundus Fluorescein Angiography is an imaging methodology widely employed in the imaging and quantification of blood flow in the choroidal and retinal region. FFA images are captured with specialized fundus cameras after injection or oral administration of the fluorescent dye to the subject. These images facilitate the interpretation of the retinal vasculature in the identification of several disorders in children and adults as discussed in [1,2]. This modality has been in practice for over 6 decades as evident from literatures [3–6]. A recent investigation on diabetic macular perfusion, assessing the efficacies of FFA and Swept Source Optical Coherence Tomography Angiography (SSOCTA) has been performed by La Mantia et al. [7]. This study reveals that FFA is highly sensitive in the identification of microaneurysms compared to SSOCTA. An elaborate review of inflammatory macular diseases by Velly et al. [8] shows that many lesions are visible in the FFA compared toother imaging modalities.

Conventional fundus imaging cameras capture FFA images with Field of View (FoV) ranging from 30° to 50°. Though wide-field FA with FoV in the range [>30°–200°] and ultra wide-field [9] FA cameras with FoV > 200° are in existence for more than a decade, they are not used in smaller clinical settings due to lack of skilled operators and cost of image acquisition.

Generally, the spatial resolutions of the FFA images are limited due to the small FoV supported by the fundus cameras. Detection of retinal disorders such as exudates, a lesion etc., in the pediatric and adult population requires a thorough examination of the fine vascular structures. The resolution of the FFA images can be enhanced at the cost of upgradation of the components of conventional fundus cameras.

Image super resolution which refers to the construction of a High Resolution (HR) image from a Low Resolution (LR) image is one of the promising solutions proposed by researchers for improving the resolution of fundus images. Super resolution problems on fundus images are formulated as Single Image Super Resolution (SISR) and Multiple Image Super Resolution (MISR) problems, based on the number of LR images involved in the construction of HR images. The first work on image super resolution was proposed by Tsai [10] in 1984. Since then, these techniques have been extended to the construction of HR images from LR images of different modalities such as ultrasound [11,12], MRI [13–15], CT [16] and PET [17].

A detailed evaluation of super resolution algorithms for retinal fundus images is presented by Thapa et al. [18]. This paper provides a detailed analysis of different categories of interpolation and learning based approaches. Interpolation based methods construct HR images by estimating the pixels from LR images. However, SISR interpolation methods are not successful in estimating pixels lost during image acquisition as described in [19]. Conversely, learning [20] based approaches extract these pixels from a set of training images and integrate them with the LR images to construct the HR images.

With the evolution of Deep Learning (DL), super resolution problems are solved with Convolution Neural Networks (CNN) which perform an end to end mapping from the LR to HR images. The pioneering work deploying Deep Neural Networks (DNN) for super resolution problems proposed by Dong et al. [21], is a light-weight structure which maps the LR image to HR image. Later, the authors improved this model named Super Resolution Convolution Neural Networks (SRCNN) with large sized filters in the nonlinear mapping layers.

Inspired by the successful application of DNNs in super resolution problems, this paper presents a novel deep residual learning network based on image priors for constructing a HR FFA image from the corresponding LR image. This network is conceptualized as a SISR model based on explicit image priors constructed from gridded interpolation of the LR images.

Rest of this paper is organized as follows. Section 2 presents a detailed account of the super resolution methods and sparse image priors essential to understand the proposed work. The underlying methods and datasets employed in this investigation are described in Section 3. The proposed super resolution network is presented in Section 4 followed by experimental results and comparative analyses in Section 5. The paper is concluded in Section 6 summarizing the major findings of this research and presenting the scope for further research.

2 Related Works

The architectures of the existing deep super resolution networks and sparse representation of priors are discussed in this section.

2.1 Sparse Image Priors

Super resolution problems are called ill-posed problems as a number of HR images can be constructed from a given LR image. However, the best HR image can be generated from well-defined image priors. Basically, a prior of an image refers to the specific information from the LR image required for construction of a HR image. Learned priors are a kind of image priors constructed by training DNNs with large datasets. To the contrary explicit priors are obtained only from the available LR image, which do not require external images for training.

Based on the image priors employed, the SISR methods are classified into patch based or example based methods, statistical models, prediction models and edge based methods.

A subjective and quantitative benchmark evaluation of these methods by Yang et al. [22] shows that example based methods exhibit best performance. The basic schematic of example based method is shown in Fig. 1.
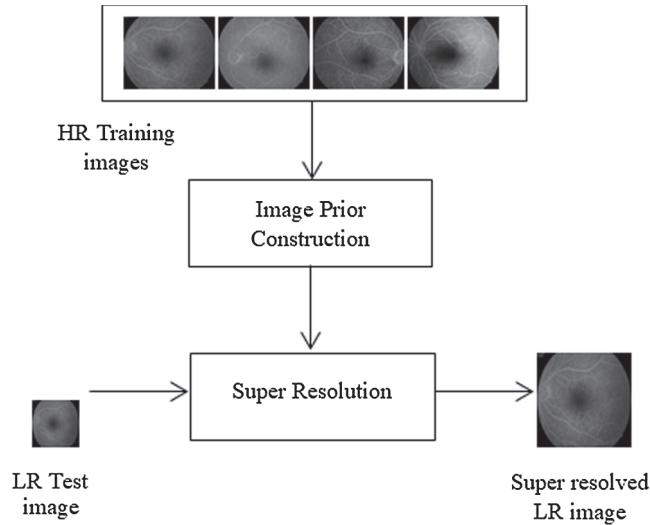

Figure 1: Example based SISR

The example based methods construct priors internally from the candidate LR images or from large set of LR and equivalent HR image pairs. It has been demonstrated in [22,23] that statistical priors of natural images can be exploited in several inverse problems in image processing. In [22], the authors have estimated high resolution images from their incomplete representations, employing natural priors. Generally image priors are constructed mathematically from candidate images based on sparse coding and its variants. Sparse coding is based on the hypothesis that band-pass filter responses are distinct for natural images, with distributions exhibiting sharp peaks around zero and heavy tails. This prior is employed in several image processing problems including super resolution.

Sparse coding is used in capturing the structural elements of images by mathematical computations or learning with a large set of training images. Dong et al. [24] have established that a structured sparse coding network based on Gaussian Scale Mixture (GSM) and Simultaneous Sparse Coding (SSC), preserves the sharp edges and suppresses visual artifacts in image restoration applications. Better spatial adaptation of the sparse coding techniques is significant in super resolution applications.

In sparse coding, LR images are mathematically expressed as down-sampled versions of HR images. The super resolution problem is the inverse problem in which the HR images are constructed from these LR images. Hence, super resolution problems based on sparse coding can be characterized as interpolation problems. Earlier, Dong et al. [25] introduced a Nonlocal Auto Regressive Model (NARM) for improving the fidelity of sparse images under image interpolation. Recently, Liu et al. [26] have proposed an approach for enhancement of resolutions in medical images, based on nonlocal interpolation and intrinsic similarities.

Super resolution Random Forests (SRF) [27] is for FFA images demonstrate superior performance compared to non-DL and DL based methods for a dataset of 185 images. This method is

implemented with pairs of LR and HR patches extracted from the FFA images. The trees within the SRF are trained independently to map the LR to HR patches and finally the SRF model maps a LR FFA image into a HR image.

2.2 Deep Neural Networks for Super Resolution

With the advent of deep learning, learning based methods have also been widely employed in image super resolution problems. Researchers have shown that the external example based learning pipeline can be realized with DNNs. The first work on DL based image super resolution was proposed by Cui et al. [28]. The internal prior based approach proposed in this paper employs cascaded Auto Encoders (AE) which requires optimization for prior construction and AEs in each stage of the network. To the contrary, the end-to-end mapping approach employed in the SRCNN, proposed in [21] optimizes all the layers together, considerably improving the performance of the network. This model is the basis for the other super resolution networks proposed till date. The schematic of the SRCNN is shown in Fig. 2. This network consists of 3 layers one each for patch representation, non-linear mapping and reconstruction.
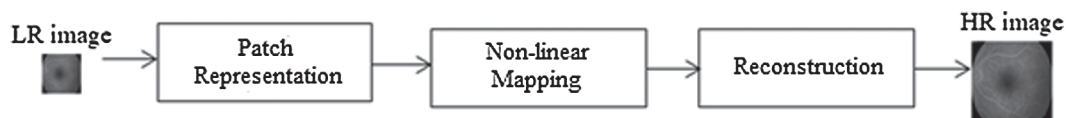

Figure 2: SRCNN architecture

The super resolution operation in this architecture starts with upsampling the given LR image X to the required resolution by bicubic interpolation. The upsampled image Y is the LR representation of the target HR image Y'. The process involved in the construction of the HR image Y' from X is described as below.

(1) **Patch extraction and representation:** High dimensional vectors called feature maps are constructed from an LR image Y by convolving it with n_1 filters of size $f_1 \times f_1$. Each feature map extracted as a patch is represented as an n_1 dimensional vector.

(2) **Non-linear mapping:** The patches extracted in Step 1 are further convolved with n_2 filters of size $n_1 \times f_2 \times f_2$ such that f_2 is smaller than f_1. This operation results in high resolution vectors of dimension n_2 which carry comparatively finer details.

(3) **Reconstruction:** In this step, the feature maps generated in Step 2 are convolved with n_3 filters each of size $n_2 \times f_3 \times f_3$ to construct the HR image Y'.

The super resolution models proposed so far are the variants of this basic model. Though several DL architectures for image super resolution exist, they are classified under four generic models namely pre-upsampling, post-upsampling, progressive upsampling and iterative up and down sampling based on the super resolution pipeline stage at which the upsampling operation is performed. The SRCNN is a pre-upsampling model as shown in Fig. 2. Inspired by this model, new DL based architectures for image super resolution are proposed in [29,30]. Though this model has reported improved quality of super resolved images compared to other models, it suffers from heavy computational costs due to the convolution operations. The Fast Super-Resolution Convolutional Neural Networks (FSRCNN) is a variant of the SRCNN which is designed to improve the speed of the SRCNN. This network is a post-upsampling model with five stages performing feature extraction, shrinking, mapping, expanding and deconvolution operations in

sequence. Initially, feature extraction is performed with the LR image and the deconvolution layer performs upsampling in the final stage. The complete architecture of FSRCNN and a comparative analysis with SRCNN with respect to the speed of computations and quality of HR images is given in [31].

An approach for the construction of hybrid priors combining the internal priors obtained by sparse coding and external priors by learning is proposed in [32]. Called Structured Analysis Sparse Coding (SASC), it employs a deep learning network with substages for internal and external prior construction with a CNN and a sparse coding network for image reconstruction respectively. Several reviews on deep learning based image super resolution exist in literature, presenting a thorough analysis on their architecture and diversities. Two recent articles in [33] and [34] present a complete review in this context, based on multiple aspects such as the underling frameworks of the models, interpolation methods, learning approaches, network design issues etc.

Several approaches for improving the performance of the existing CNNs have been proposed so far with respect to modification of the key parameters of the models such as window size, stride size, scaling factors etc. An investigation by Simonyan et al. [35] shows that increasing the depth of the network with fixed values of other parameters considerably improves the classification and localization accuracies of CNNs. Inspired by the performance of this network, a deep network for image super resolution is proposed by Kim et al. [36] in 2016. This model called the Very Deep Super-Resolution Network (VDSR), based on the VGG network for image classification is implemented as a pre-upsampling based residual learning network. The schematic of the VDSR is shown in Fig. 3.
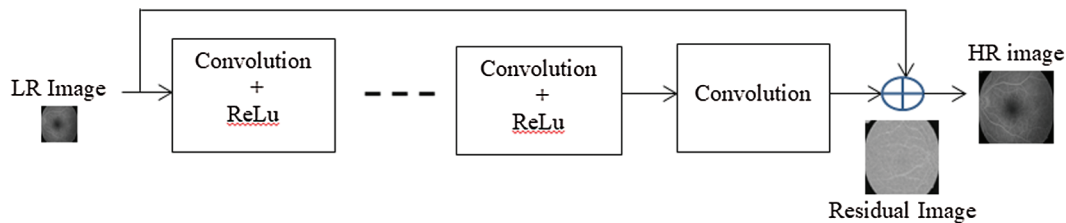

Figure 3: VDSR architecture

As shown in Fig. 3, the ground truth image is upscaled to the dimension of the HR image. This image is convolved by the intermediate convolution layers. The residual image is constructed at the final layer and is added with the upscaled LR image to generate the HR image. Though a plethora of deep learning networks exist in literature for super resolution, we confine our discussion to unique kinds of networks.

3 Materials and Methods

The dataset employed in this investigation and the underlying methods in realizing the proposed super resolution network architecture are discussed in the following subsections.

3.1 Dataset

The dataset [37] used in this research comprises 70 FFA images of diabetic patients captured for a study in Isfahan University of Medical Sciences. These images are of dimension 576×720 with 8-bit depth. The FFA images acquired from 70 diabetic patients with various pathologies

such as mild Non-Proliferative Diabetic Retinopathy (NPDR), moderate NPDR, severe NPDR and Proliferative Diabetic Retinopathy (PDR) are represented in two categories namely Normal and Abnormal with 30 and 40 images respectively.

3.2 Gridded Interpolation

Interpolation is an indispensable process in image super resolution for upsampling the images, irrespective of the network design. Generally, nearest neighbor [38], bilinear [39] and bicubic [40] interpolations are widely employed in various super resolution schemes. Nearest neighbor interpolation estimates a pixel from a nearest pixel in the region of interpolation without considering other pixels. Though it is implemented as an algorithm generating a piecewise-constant, it results in noise around the boundaries.

Bilinear interpolation constitutes two linear interpolations in sequence on the two image axes resulting in a quadratic interpolation. For scaling the images, this interpolation considers a 2×2 neighborhood for the pixel to be interpolated. A weighted average of these pixels is the interpolated value, resulting in better image quality compared to neighbor interpolation. Bicubic interpolation is performed on 4×4 nearest pixels generating high quality images with fewer artifacts. However, it is comparatively slower than other interpolation techniques. A number of image super resolution networks such as SRCNN, FSRCNN, and VDSR etc., are based on bicubic interpolation. The NARM proposed in [25] is an interpolation scheme based on the sparse image representation which adaptively models pixels from the local and non-local image similarities. The authors show that the NARM can be included as a fidelity term in the sparse image representation model which generalizes interpolation into image restoration problems such as super resolution. The importance of the study of interpolation techniques and their evaluations, exclusively for medical images is highlighted in [41]. This paper presents a detailed analysis of various interpolation techniques and advocates the need for non-linear adaptive interpolation schemes.

Basically interpolation of a digital image is performed by resampling the respective continuous function at discrete locations, i.e., reconstruction of the continuous function. An adaptive interpolation mechanism which derives an optimal interpolator for the candidate image can facilitate improved interpolation, compared to employing a standard arbitrary function. The deep learning network proposed in our paper employs a grid based interpolation scheme which builds a gridded interpolant for each image to perform interpolation in the two dimensional space as described below.

For a given image I, let f(p), be its continuous function. Then the discrete image $g(p)$ interpolated at a set of discrete locations P such that $p \epsilon P$ is represented as in Eq. (1).

$$f(p) = \begin{cases} g(p) & on\ grid \\ undefined & not\ on\ grid \end{cases} \tag{1}$$

Now, the interpolation problem can be modeled as a construction of the continuous function $g'(p)$ from $f(p)$ as in Eq. (2).

$$g'(p) = \sum_{p} f(p) H_p(x - p), \quad p \epsilon P \tag{2}$$

where H_p is the interpolation function at the point p.

In the interpolation of a digital image, it is required to determine only the location of the new sample points rather than reconstruction of the complete image. Let the new locations p' be represented a set of points P' such that each $p' \in P'$. The new image interpolated at new locations p' is as in Eq. (3).

$$g'(p') = f(p') = \sum_p f(p) H_p(p' - p), \quad p \in P, \quad p' \in P' \tag{3}$$

As it is unrealistic to interpolate a point from all the input samples, interpolation of a new point can be performed with a small subset of points that lie in close proximity d to the candidate point. Hence the interpolation function in Eq. (3) can be rewritten as in Eq. (4).

$$g'(p') = f(p') = \sum_p f(p) H_p(p' - p), \quad p' \in P', \quad p \in P_{p',d} \tag{4}$$

where $P_{p',d} \triangleq \{p \in P, \quad \|p - p'\|^2 < d\}$ is the set of points at a distance d.

In image super resolution, wherever the image patterns are similar in the LR image, the pixels are repeated in the HR image carrying over the weights of the LR pixels to the HR solution space. Many machine vision applications have successfully employed the Thin Plate Spline (TPS) for mapping coordinates between solution spaces, finding the missing data in grid based construction of images. For transformation of $n \times n$ data points, TPS requires inversion of a matrix of dimension $n \times n$. Here, the computational cost linearly increases with the number of points in the patterns to be mapped from the LR to the HR space.

It has been demonstrated in [42] that a weighted TPS with low computational cost can be used in the gridded interpolation of Light Detection and Ranging (LiDAR) images. This interpolation is modeled with a minimization objective function as given in Eq. (5).

$$\min_f \|p - p'\|^2 + \lambda T(f) \tag{5}$$

where λ is the smoothness parameter and $T(f)$ is the penalty term for smoothness as in Eq. (6).

$$T(f) = \int (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2) dx dy \tag{6}$$

It is evident that this interpolation is achieved by minimizing the fidelity-smoothness trade-off of the images. Hence, the TPS based gridded interpolation can be extended to the 2D FFA images for a smooth representation of HR image from the LR image. Similarly, it can also be extended to multi dimensional fundus image tensors, performing the interpolation in multiple directions.

3.3 Residual Learning

Though accuracy of a network considerably improves with deep layers, it is diminished due to vanishing gradients as the weights do not update at deeper layers. This degradation problem was first reported by He et al. [43] who proposed the residual learning as a solution to improve accuracy in networks with a depth of 2000. The schematic of a shallow residual network is shown in Fig. 4.

In this framework, the input x is transformed by a sequence of weighted convolutional layers resulting in $F(x)$ which is summed with the unaltered input x to construct the residual y.
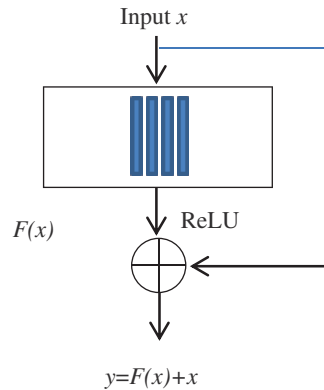
Input x

$F(x)$ ReLU

$y=F(x)+x$

Figure 4: Residual learning

Residual learning can simplify image super resolution by finding the residual, i.e., the difference between the reference HR and the upscaled LR images. In pre-upsampling frameworks the LR and HR images are highly correlated and differ only by the fine high resolution features. These residuals can be added with the upscaled LR images to generate the HR images. The VDSR network proposed in [36] is a kind of pre-upsampling residual learning network with 20 layers which maps the LR image constructed from the reference image by bicubic interpolation into an HR image.

As described in the previous subsection, gridded interpolation provides better approximation compared to conventional interpolation schemes. When residual learning is applied on the grid interpolated LR image, the resolution of the HR image will be comparatively better.

3.4 SWISH Activation Function

Activation functions play a significant role in DL networks. The Rectified Linear Unit (ReLU) is widely used in deep networks due to its ease of implementation and the ability of the gradient to flow for positive inputs. Given an input x, weight w and bias b, the activation a and the application of ReLU on a is in Eqs. (7) and (8).

$$a = wx + b \tag{7}$$

$$f(x) = \max(a, 0) \tag{8}$$

From the above, it is seen that a tends to increase with x and becomes b when x is 0 which eliminates the gradient vanishing problem. Also, ReLU results in sparse representations when $x \leq 0$. However, for high learning rates, most of the neurons may not be activated by ReLU resulting in the dying ReLU problem. Recently, a novel non-monotonic activation function called the Swish [44] which is comparatively smoother than ReLU has been introduced. The swish function based on sigmoid function is as given in Eq. (9). The Swish activation function is shown in Fig. 5 for x ranging from -5 to $+5$. It is seen that the function is completely adaptive, with no assumptions.

$$f(x) = x.\sigma(x) \tag{9}$$

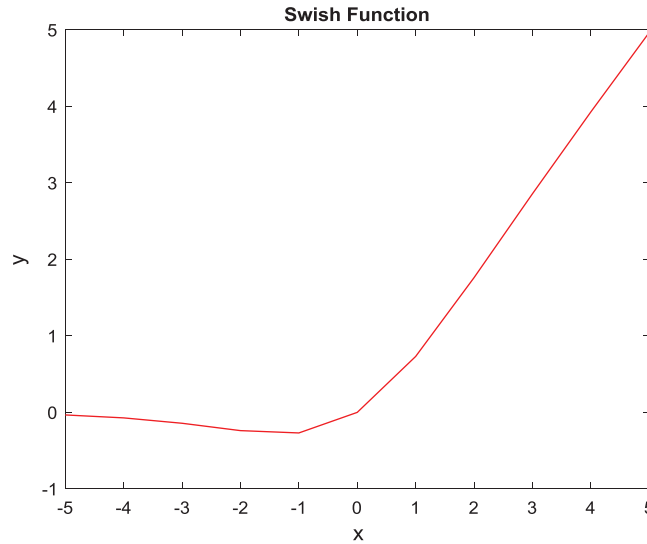where $\sigma(x) = (1 + \exp(-x))^{-1}$

Figure 5: Swish activation function

Significant improvement in classification accuracy is evidenced with NASNet-A [45] and Inception-ResNet-v2 [46] deep networks on the ImageNet dataset, with Swish activation replacing ReLUs. Investigations in [44] show that the Swish activation function outperforms ReLU in various deep learning networks and it is also reported that the Swish function closely resembles the activate functions in retinal neurons of vertebrates. Gating mechanisms in Recurrent Neural Networks (RNNs) such as Long short-term memory (LSTM) networks dynamically control the flow of information from previous layers to a current layer, preventing the gradient descent problems, characteristic of the RNNs. This problem is completely eliminated with the self-gating Swish activation functions as shown in Fig. 6.
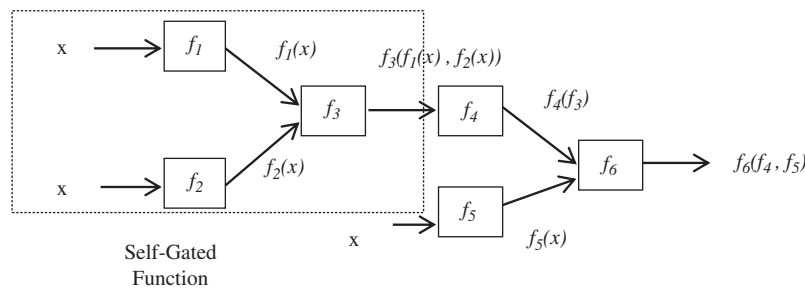

Figure 6: Self-gating swish activation function

The architecture in Fig. 6 is a combination of a unary and binary operation implemented with a self Gating network. In this network, the functions f_1 and f_2 are unary while f_3 is binary. The gate comprising these functions evaluates the Swish function in Eq. (9). By self-gating this function, it is can replace the conventional ReLU in recurrent and deep residual networks for achieving best approximations.

4 Proposed System

4.1 Model Configuration

The proposed deep learning network for super resolution is designed as a pre-upsampling residual learning network based on the VDSR model with two significant changes in the underlying model. First, the gridded interpolation is employed for pre-upscaling in place of the bicubic interpolation and the ReLU is replaced with the Swish function. The schematic of the proposed system is shown in Fig. 7.
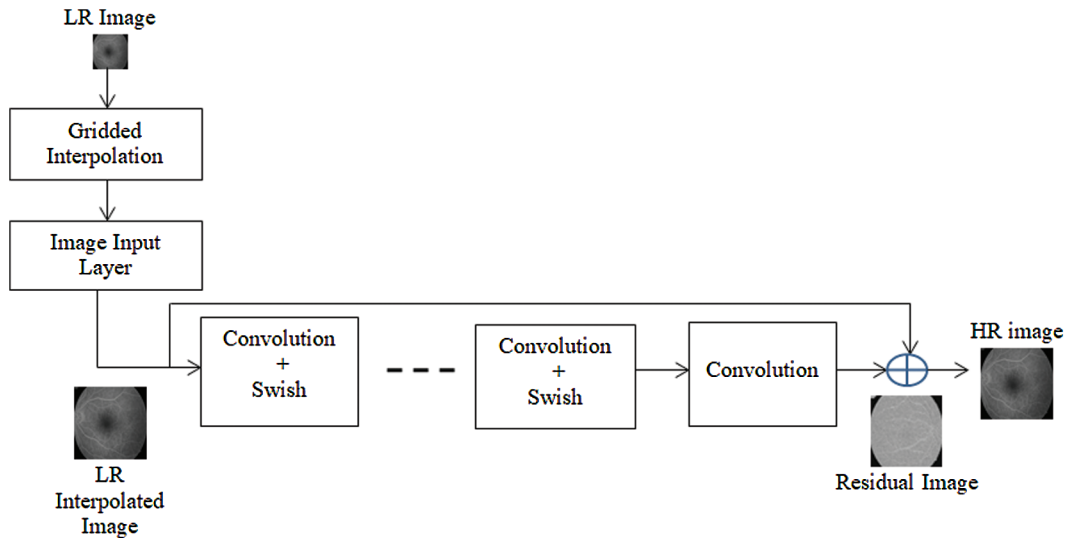

Figure 7: Gridded Interpolation based residual learning architecture

The above residual network is implemented and tested in Matlab 2019b. The network is implemented with 20 weighted layers, each of which is coupled with the swish function. Initially, the LR image is upscaled to match the size of the HR image by gridded interpolation and the HR image is constructed from this LR interpolated image.

4.2 Model Optimization

Basically, the Image Input Layer is trained to operate on image patches of the luminance channel. In this work, we train this layer to operate on the intensity values as the input images are of 8 bit depth. However, we stick to the patch size of 41×41 and 64 filters of size 3×3 in the convolutional layers similar to the VDSR network. We also employ the Stochastic Gradient Descent with Momentum (SGDM) function with an initial learning rate of 0.2 and 0.9 momentum. These values are empirically chosen varying the learning rate in the range 0.01 and 0.1 for the momentum value 0.9. As the model produced HR images with reasonable PSNR values at the learning rate 0.2, the network is trained for 100 epochs minimizing the learning factor by 10 for each epoch. However, the learning rate is 0.1 in the basic VDSR while the diminishing factors and the number of epochs are the same. We have assumed the value of the smoothness parameter λ as 0.1 in our model after empirical evaluation of the quality of the reconstructed images.

Tab. 1 presents the configuration of the proposed model with design and optimization parameters.

Table 1: Model parameters

Parameter	Value
No. of weighted layers	20
Maximum epochs	100
MiniBatchSize	128
Momentum	0.9000
Learning rate	0.200
Optimization	SGDM
Activation function	Swish

Since the basic VDSR model is trained with a wide range of natural image dataset, we initially train the network with a new dataset constructed from the 70 images of the test dataset. The training dataset is created by down sampling the 30 normal and 40 abnormal images of dimension 576×720 to 128×128. Initially, the network is trained with these images with scale factors 2 and 4 to construct HR images of dimensions 256×256 and 512×512 respectively.

5 Experimental Works and Discussions

We have implemented and tested our residual network in a i7-7700K processor with 16 GB DDR4 RAM and NVIDIA GeForce GTX1060 3 GB Graphics card. The proposed system is tested with 2 scaling factors, 2 and 4 on both the raw normal and abnormal FFA images of the dataset each of dimension 576×720. We get HR images of dimensions 1152×1440 and 2304×2880 for scaling factors 2 and 4 respectively. The reconstructed HR images with the benchmark approaches and corresponding performance metrics are shown in Fig. 8a for one Normal and Abnormal FFA image. The LR test images are given in Fig. 8a. The HR reference images for evaluation of performance metrics are obtained by bicubic interpolation of the LR test images applying suitable scaling factors. The first column in Fig. 8b shows a normal image scaled by 2 and the second column shows an abnormal image scaled by 4. The quality of the super resolved FFA images is evaluated with the PSNR and SSIM metrics evaluated with Eqs. (10) and (12) respectively.

$$PSNR = 10 log_{10} \left(\frac{R^2}{MSE} \right) \tag{10}$$

$$MSE = \frac{1}{MN} \sum_{M,N} [I_1(m,n) - I_2(m,n)]^2 \tag{11}$$

$$SSIM(x,y) = l(x,y)^{\alpha} * c(x,y)^{\beta} * s(x,y)^{\gamma} \tag{12}$$

where $l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$, $C(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$ and $S(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$

where μ_x, μ_y, σ_x, σ_y and σ_{xy} are the local means, standard deviations, and cross-covariance for images x and y if $\alpha = \beta = \gamma = 1$.

In the above figure, the figure to the left is a normal FFA with no obvious visual artifacts. Whereas microaneurysms are evident as white dots on the retinal surface. These microaneurysms manifest as a result of the leakage of the fluorescein from the retinal vessels signifying the abnormality in the retina. Detection of fine blood vessels and lesions is very vital in the diagnosis
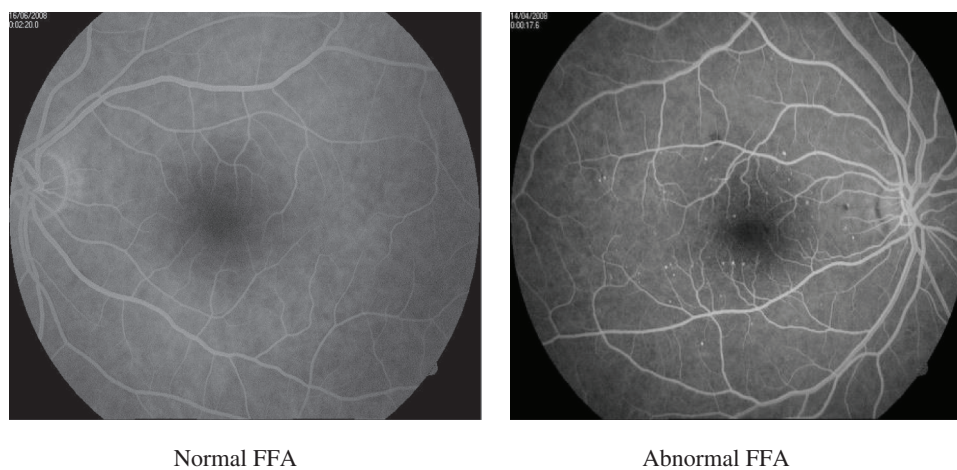
and prognosis of several retinal disorders. Generation of HR images from these images facilitate the magnification of the vessels and lesions for accurate diagnosis.

The super resolved images in Fig. 8b clearly depict the retinal vessels for a through diagnosis based on the retinal vasculature. From the corresponding PSNR and SSIM metrics, it is seen that the perceptual image quality of the HR is comparatively lower for scaling factor 4 compared to that obtained with the scaling factor 2. Invariably degradation by 2 dB is witnessed for all the super resolution models. Compared with the VDSR model, the proposed model demonstrates matching PSNR values and better SSIM values for this image pair. The higher SSIM values indicate that the structural contents of the images are preserved well by the proposed model compared to the VDSR. To the contrary, our model also exhibits lower SSIM compared to the SRF, for scaling factor 4 in spite of a higher PSNR.

However, the values given in Fig. 8a are characteristic of the image pair shown and not representative of the entire dataset. We present the average metrics evaluated for the normal and abnormal images of the entire dataset in Tab. 2. We have done a quantitative assessment with the PSNR and SSIM metrics to evaluate the quality of the super resolved images constructed with the proposed residual network, SRCNN, FSRCNN, SRF, VDSR and bicubic interpolation on the dataset as in Tab. 2.

It is seen that the proposed super resolution model provides best HR image reconstruction for both the scaling factors. Further, it also achieves best results compared to the FSRCNN and the VDSR both of which are pre-upsampling methods. Also, the proposed system exhibits a comparatively improved performance than the most recent SRF. The quality of the reconstructed HR FFA images is attributed to the gridded interpolation and the Swish function.

Though the proposed model is based on the 20 layered VDSR architecture, the PSNR values are higher by 6 dB and 2 dB for the scaling factors 2 and 4 respectively compared to the VDSR. Similarly, significant improvements in SSIM metrics are also evidenced under both the scaling factors. By convention, the VDSR, SRCNN, and the FSRCNN, transform the bicubically interpolated FFA image in to the HR image which introduces jagged artifacts. We have introduced

Normal FFA   Abnormal FFA

(a)

Figure 8: (Continued)

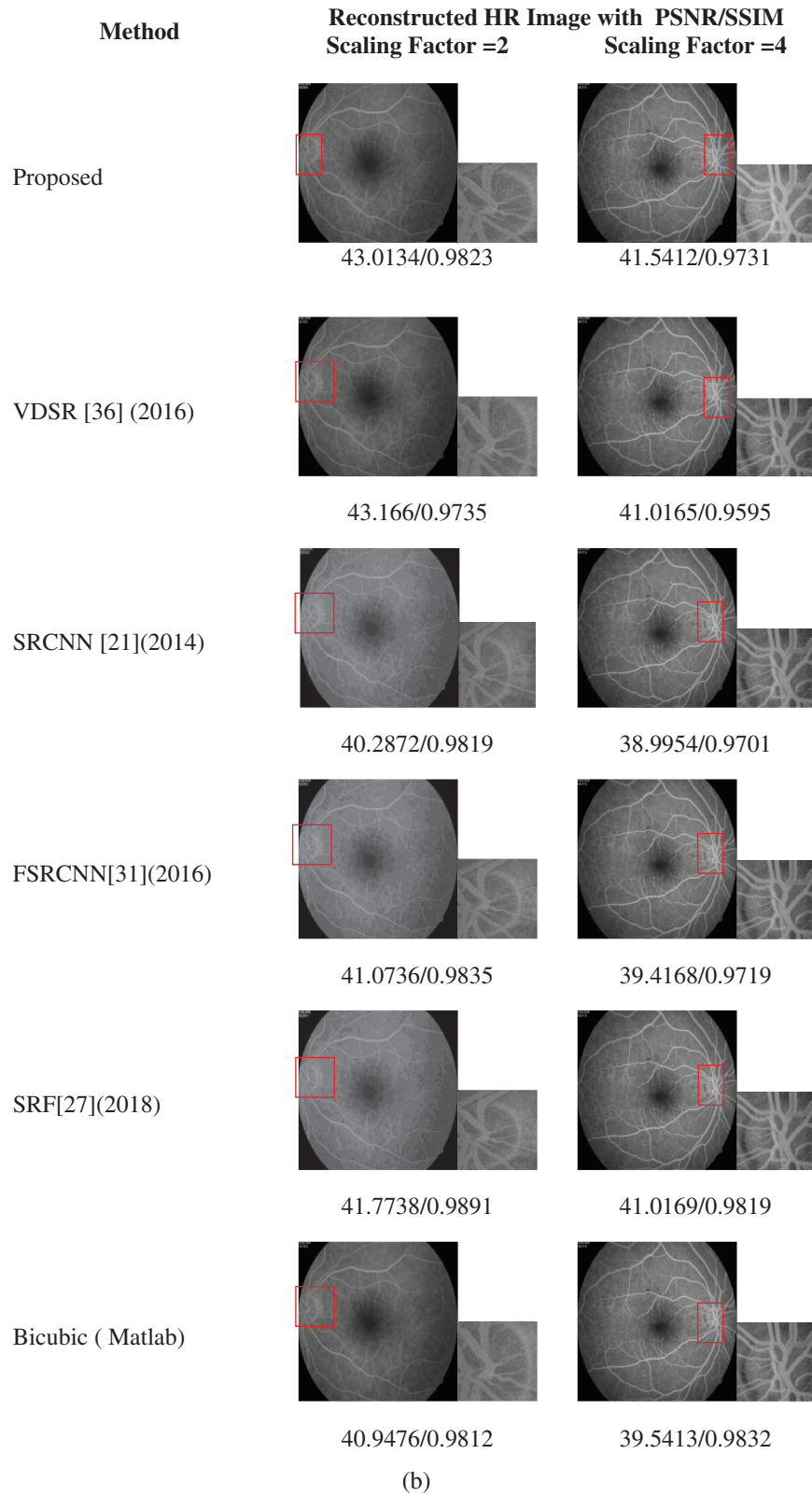| Method	Reconstructed HR Image with PSNR/SSIM Scaling Factor =2	Scaling Factor =4
Proposed	43.0134/0.9823	41.5412/0.9731
VDSR [36] (2016)	43.166/0.9735	41.0165/0.9595
SRCNN [21](2014)	40.2872/0.9819	38.9954/0.9701
FSRCNN[31](2016)	41.0736/0.9835	39.4168/0.9719
SRF[27](2018)	41.7738/0.9891	41.0169/0.9819
Bicubic (Matlab)	40.9476/0.9812	39.5413/0.9832

(b)

Figure 8: (a) Low resolution FFA. (b) Super resolution results—proposed system & benchmarks

gridded interpolation with an intention of reducing these artifacts to strike a balance between image fidelity and smoothness. While the PSNR and SSIM are indicative of these image features, the smoothness parameter λ considerably affects these metrics. With $\lambda = 0.1$, we achieve the performance metrics in Tab. 2. For a fair evaluation,we have employed the optimization function as SGDM with momentum 0.9 without modifying their underlying architecture in VSDR, SRCNN and FSRCNN models. We see that FSRCNN and VDSR have closely matching PSNR and SSIM values with around 2 dB and 1 dB improvement in PSNR for scaling factors 2 and 4 respectively compared to SRCNN. It is also evidenced that FSRCNN shows only a marginal improvement than SRCNN for scaling factor 4.

Table 2: Performance metrics

Method	Scale Factor = 2		Scale Factor = 4	
	PSNR	SSIM	PSNR	SSIM
Proposed	**42.2797**	**0.9568**	**39.6129**	**0.9712**
VDSR	36.0175	0.9254	34.5012	0.9213
SRCNN	34.1240	0.9317	33.1691	0.9112
FSRCNN	36.3901	0.9448	34.2019	0.9104
SRF	38.0128	0.9316	35.894	0.9335
Bicubic	35.7921	0.8886	34.5413	0.9156

Compared with SRCNN, RSRCNN and VDSR, SRF achieves better PSNR and SSIM values for both the scaling factors, except a slight fall in SSIM compared to FSRCNN for scaling factor 2. SRF model learns directly by mapping the LR patches into HR patches constructing an ensemble of decision tress. The performance of this model depends on the number of decision trees which introduces a trade-off between the performance and computational intensity. We have tested the SRF model with 6 decision trees as in [27]. Though this model seems to be better than VDSR, there is a noticeable degradation of 4 dB compared to the proposed model for scaling factor 2.

From the above analysis we understand that nearly perfect reconstruction can be achieved by mapping the LR images to HR images by direct mapping or without introducing artifacts in the pre/post upsampling interpolation operations. We also see low SSIM values for images with higher PSNR values which raises concern on the evaluation metrics of the super resolution models. While a higher PSNR signifies good visual quality, SSIM signifies structural intactness. A higher PSNR with relatively low PSNR indicates that the image is reconstructed well but with structural artifacts. For the FFA super resolution problem, it must be ensured that fine structural information is not lost by super resolution.

The experimental results show that the proposed SISR model can reconstruct FFA images with enhanced visual quality highlighting the artifacts for better diagnosis. This model is a promising replacement for the complex MISRmodels which demand intensive computations and also cause discomfort to the subject during the acquisition of multiple FFA images. This model is prospective for low resource constrained environments in which only single images of the fundus are captured which can be super resolved to improve their diagnostic values.

The average computational times for the super resolution of the images in the dataset are presented in Tab. 3.

Table 3: Computational time-HR image construction

Method	Computational time (ms)	
	Scale Factor = 2	Scale Factor = 4
Proposed	**12.31**	**19.43**
VDSR	17.25	25.23
SRCNN	15.32	25.21
FSRCNN	13.74	23.05
SRF	16.32	28.01
Bicubic	20.78	25.31

The computational times for the construction of HR images for both the scale factors are best for the proposed system. A thorough analysis of the quality metrics and computational times show that the performance of the proposed system is better compared to the FSRCNN for the test dataset. However, in [45], it is shown that the SRF performs better than VDSR which is a pre-upsampling model.

Though the proposed system matches the structure of the VDSR, it is seen that the HR image quality and computational time are comparatively better for the proposed system. We see that there is an enhancement of around 5 dB in PSNR value for both the scaling factors. Further substantial improvements are also evidenced in SSIM and computational times. Enhancement in computational times is attributed to the higher learning rate of this network. As mentioned in Section 4, the initial learning rate is 0.2 which is diminished by 10 for each iteration. While the residual networks reported in existing literature employing ReLU assume a learning rate of 0.1, the Swish function supports a high learning rate, enabling the networks to train faster.

A notable characteristic of the proposed system is that, it does not involve assumption of any parameters. Initially, the interpolated image is constructed from the gridded interpolants, intrinsic to the images. Similarly, the swish activation function is also characteristic of the candidate image. Generally, the performance of the deep residual models depends upon the learning of images residuals, which is constrained by the number of weighted layers and the ability of the networks to learn the weights either forward or backward. We have shown that an unsupervised model proposed in this paper generates HR images of significant clinical values from FFA image priors constructed by gridded interpolation.

Finally, it is very well evident that residual learning networks in which the interpolation is minimized, enhanced with adaptive image priors and activation functions are very prospective for the super resolution of medical images.

Though this model features superior performance compared to the other deep learning and SRF models, we have not evaluated its performance to specific retinal disorders. It has been tested with an integral dataset containing normal and abnormal FFA images without focusing on any particular disorder in particular. This model can be extended by transfer learning to a specific disorder which requires intensive training and testing with an exclusive disorder dataset. We have provided only the generic model which needs to be fine-tuned to the disorders to leverage its fullest potential.

A most recent representative work by Anoop et al. [47] similar to our research employs a K Nearest Neighbor network focusing on the Region of Interest (ROI) of the retinal image rather than the entire image to extract the image features from the LR images.

In line with this, our model can be coupled with a segmentation module to extract the ROI from the LR FFA images for super resolution.

6 Conclusion

Recent clinical research has proved the effectiveness of FFA in the diagnosis and treatment of diabetic retinopathy. However, the resolution of the FFA images is limited by the image acquisition devices in resource constrained clinical settings. This paper proposes a novel deep learning based super resolution network exploiting the characteristics of gridded interpolation, residual learning and swish functions. The visual and quantitative experimental results and computational cost show that the proposed residual learning model is comparatively better than other benchmark mechanisms. Inspired by these results, this paper advocates FFA imaging as reliable modality in the prognosis of various retinal pathologies as it facilitates the detection of microaneurysms and visual artifacts by super resolution. The proposed residual model can be further refined with priors pertaining to retinal components such as exudates, Hemorrhages, Microaneurysms, lesions etc. to construct their HR representations. This paper also encourages researchers to investigate novel adaptive activation mechanisms for deep learning networks, alternative to the conventional activation functions.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Temkar, S., Azad, S. V., Chawla, R., Damodaran, S., Garg, G. et al. (2019). Ultra-widefield fundus fluorescein angiography in pediatric retinal vascular diseases. *Indian Journal of Ophthalmology, 67(6),* 788. DOI 10.4103/ijo.IJO_1688_18.
2. Zhao, Y., MacCormick, I. J., Parry, D. G., Leach, S., Beare, N. A. et al. (2015). Automated detection of leakage in fluorescein angiography images with application to malarial retinopathy. *Scientific Reports, 5(1),* 10425. DOI 10.1038/srep10425.
3. Younge, B. R. (1976). Fluorescein angiography and retinal venous sheathing in multiple sclerosis. *Canadian Journal of Ophthalmology, 11(1),* 31–36.
4. Nanba, K., Schwartz, B. (1988). Nerve fiver layer and optic disc fluorescein defects in glaucoma and ocular hypertension. *Ophthalmology, 95(9),* 1227–1233. DOI 10.1016/S0161-6420(88)33024-1.
5. Wang, X., Li, S., Li, W., Hua, Y., Wu, Q. (2018). Choroidal variations in diabetic macular edema: fluorescein angiography and optical coherence tomography. *Current Eye Research, 43(1),* 102–108. DOI 10.1080/02713683.2017.1370115.
6. Marcucci, R., Sofi, F., Grifoni, E., Sodi, A., Prisco, D. (2011). Retinal vein occlusions: a review for the internist. *Internal and Emergency Medicine, 6(4),* 307–314. DOI 10.1007/s11739-010-0478-2.
7. La Mantia, A., Kurt, R. A., Mejor, S., Egan, C. A., Tufail, A. et al. (2019). Comparing fundus fluorescein angiography and swept-source optical coherence tomography angiography in the evaluation of diabetic macular perfusion. *Retina, 39(5),* 926–937. DOI 10.1097/IAE.0000000000002045.

8. Velly, D. K., Shah, H., Kharel, R., Biswas, J. (2016). Inflammatory macular diseases: a review. *Kerala Journal of Ophthalmology, 28(2),* 96–102. DOI 10.4103/kjo.kjo_31_16.

9. Rabiolo, A., Parravano, M., Querques, L., Cicinelli, M. V., Carnevali, A. et al. (2017). Ultra-wide-field fluorescein angiography in diabetic retinopathy: a narrative review. *Clinical Ophthalmology, 11,* 803–807. DOI 10.2147/OPTH.S133637.

10. Tsai, R. (1984). Multiframe image restoration and registration. *Advance Computer Visual and Image Processing, 1,* 317–339.

11. Morin, R., Basarab, A., Ploquin, M., Kouamé, D. (2012). Post-processing multiple-frame super-resolution in ultrasound imaging. *Medical Imaging 2012: Ultrasonic Imaging, Tomography, and Therapy*.

12. Zhao, N., Wei, Q., Basarab, A., Kouamé, D., Tourneret, J. Y. et al. (2016). Single image super resolution of medical ultrasound images using a fast algorithm. *IEEE 13th International Symposium on Biomedical Imaging,* pp. 473–476. Prague: IEEE.

13. Bhatia, K. K., Price, A. N., Shi, W., Hajnal, J. V., Rueckert, D. et al. (2014). Super-resolution reconstruction of cardiac MRI using coupled dictionary learning. *IEEE 11th International Symposium on Biomedical Imaging,* pp. 947–950. Beijing: IEEE.

14. Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K. et al. (2016). Multi-input cardiac image super-resolution using convolutional neural networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention,* pp. 246–254. Cham: Springer.

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

16. Umehara, K., Ota, J., Ishida, T. (2018). Application of super-resolution convolutional neural network for enhancing image resolution in chest CT. *Journal of Digital Imaging, 31(4),* 441–450. DOI 10.1007/s10278-017-0033-z.

17. Mejia, J., Mederos, B., Ortega, L., Gordillo, N., Avelar, L. (2017). Small animal PET image super-resolution using Tikhonov and modified total variation regularisation. *Imaging Science Journal, 65(3),* 162–170. DOI 10.1080/13682199.2017.1311396.

18. Thapa, D., Raahemifar, K., Bobier, W. R., Lakshminarayanan, V. (2014). Comparison of super-resolution algorithms applied to retinal images. *Journal of Biomedical Optics, 19(5),* 056002. DOI 10.1117/1.JBO.19.5.056002.

19. Tian, J., Ma, K. K. (2011). A survey on super-resolution imaging. *Signal, Image and Video Processing, 5(3),* 329–342. DOI 10.1007/s11760-010-0204-6.

20. Wu, W., Liu, Z., Gueaieb, W., He, X. (2011). Single-image super-resolution based on Markov random field and contourlet transform. *Journal of Electronic Imaging, 20(2),* 023005. DOI 10.1117/1.3580750.

21. Dong, C., Loy, C. C., He, K., Tang, X. (2014). Learning a deep convolutional network for image super-resolution. *European Conference on Computer Vision,* pp. 184–199. Cham: Springer.

22. Yang, C. Y., Ma, C., Yang, M. H. (2014). Single-image super-resolution: a benchmark. *European Conference on Computer Vision,* pp. 372–386. Cham: Springer.

23. Russell, M. F. T. B. C., Freeman, W. T. (2003). Exploiting the sparse derivative prior for super-resolution and image demosaicing. *Proceedings of the Third International Workshop Statistical and Computational Theories of Vision,* 1–28.

24. Dong, W., Shi, G., Ma, Y., Li, X. (2015). Image restoration via simultaneous sparse coding: where structured sparsity meets Gaussian scale mixture. *International Journal of Computer Vision, 114(2–3),* 217–232. DOI 10.1007/s11263-015-0808-y.

25. Dong, W., Zhang, L., Lukac, R., Shi, G. (2013). Sparse representation based image interpolation with nonlocal autoregressive modeling. *IEEE Transactions on Image Processing, 22(4),* 1382–1394. DOI 10.1109/TIP.2012.2231086.

26. Liu, H., Guo, Q., Wang, G., Gupta, B. B., Zhang, C. (2019). Medical image resolution enhancement for healthcare using nonlocal self-similarity and low-rank prior. *Multimedia Tools and Applications, 78(7),* 9033–9050. DOI 10.1007/s11042-017-5277-6.

27. Jiang, Z., Yu, Z., Feng, S., Huang, Z., Peng, Y. et al. (2018). A super-resolution method-based pipeline for fundus fluorescein angiography imaging. *Biomedical Engineering Online, 17(1),* 125. DOI 10.1186/s12938-018-0556-7.

28. Cui, Z., Chang, H., Shan, S., Zhong, B., Chen, X. et al. (2014). Deep network cascade for image super-resolution. *European Conference on Computer Vision,* pp. 49–64. Cham: Springer.

29. Tai, Y., Yang, J., Liu, X., Xu, C. (2017). Memnet: a persistent memory network for image restoration. *Proceedings of the IEEE International Conference on Computer Vision,* 4539–4547.

30. Shocher, A., Cohen, N., Iran, M. (2018). Zero-shot super-resolution using deep internal learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 3118–3126.

31. Dong, C., Loy, C. C., Tang, X. (2016). Accelerating the super-resolution convolutional neural network. *European Conference on Computer Vision,* pp. 391–407. Cham: Springer.

32. Wu, F., Dong, W., Shi, G., Li, X. (2018). Learning hybrid sparsity prior for image restoration: where deep learning meets sparse coding. arXiv preprint arXiv: 1807.06920.

33. Wang, Z., Chen, J., Hoi, S. C. (2019). Deep learning for image super-resolution: a survey. arXiv preprint arXiv: 1902.06068.

34. Anwar, S., Khan, S., Barnes, N. (2019). A deep journey into super-resolution: a survey. arXiv preprint arXiv: 1904.07523.

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

36. Kim, J., Kwon Lee, J., Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 1646–1654.

37. Alipour, S. H. M., Rabbani, H., Akhlaghi, M. (2014). A new combined method based on curvelet transform and morphological operators for automatic detection of foveal avascular zone. *Signal, Image and Video Processing, 8(2),* 205–222. DOI 10.1007/s11760-013-0530-6.

38. Kress, R. (1998). Interpolation. *In Numerical Analysis.* pp. 151–188. New York, NY: Springer.

39. Aràndiga, F. (2016). A nonlinear algorithm for monotone piecewise bicubic interpolation. *Applied Mathematics and Computation, 272,* 100–113. DOI 10.1016/j.amc.2015.08.027.

40. Siauw, T., Bayen, A. (2014): *An introduction to MATLAB® programming and numerical methods for engineers.* London: Academic Press.

41. Dumitrescu, D., Boiangiu, C. A. (2019). A study of image upsampling and downsampling filters. *Computers, 8(2),* 30. DOI 10.3390/computers8020030.

42. Chen, C., Li, Y. (2019). A fast global interpolation method for digital terrain model generation from large LiDAR-derived data. *Remote Sensing, 11(11),* 1324. DOI 10.3390/rs11111324.

43. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 770–778.

44. Ramachandran, P., Zoph, B., Le, Q. V. (2017). Swish: a self-gated activation function. arXiv preprint arXiv: 1710. 05941, 7.

45. Zoph, B., Vasudevan, V., Shlens, J., Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 8697–8710.

46. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *Thirty-First AAAI Conference on Artificial Intelligence.* arXiv preprint arXiv:1602.07261.

47. Anoop, V., Bipin, P. R. (2020). Super-resolution based automatic diagnosis of retinal disease detection for clinical applications. *Neural Processing Letters,* 1–16. DOI 10.1007/s11063-020-10292-x.