



**HAL**  
open science

## Multiscale brain MRI super-resolution using deep 3D convolutional networks

Chi-Hieu Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, François Rousseau

► **To cite this version:**

Chi-Hieu Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, et al.. Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Computerized Medical Imaging and Graphics*, 2019, 77, pp.101647. 10.1016/j.compmedimag.2019.101647 . hal-01635455v2

**HAL Id: hal-01635455**

**<https://hal.science/hal-01635455v2>**

Submitted on 17 Aug 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multiscale brain MRI super-resolution using deep 3D convolutional networks

Chi-Hieu Pham<sup>a,\*</sup>, Carlos Tor-Díez<sup>a</sup>, H el ene Meunier<sup>b</sup>, Nathalie Bednarek<sup>b,d</sup>, Ronan Fablet<sup>c</sup>, Nicolas Passat<sup>d</sup>, Fran ois Rousseau<sup>a</sup>

<sup>a</sup>*IMT Atlantique, LaTIM U1101 INSERM, UBL, Brest, France*

<sup>b</sup>*Service de m edecine n eonatale et r eanimation p ediatricue, CHU de Reims, France*

<sup>c</sup>*IMT Atlantique, LabSTICC UMR CNRS 6285, UBL, Brest, France*

<sup>d</sup>*Universit e de Reims Champagne-Ardenne, CReSTIC, Reims, France*

---

## Abstract

The purpose of super-resolution approaches is to overcome the hardware limitations and the clinical requirements of imaging procedures by reconstructing high-resolution images from low-resolution acquisitions using post-processing methods. Super-resolution techniques could have strong impacts on structural magnetic resonance imaging when focusing on cortical surface or fine-scale structure analysis for instance. In this paper, we study deep three-dimensional convolutional neural networks for the super-resolution of brain magnetic resonance imaging data. First, our work delves into the relevance of several factors in the performance of the purely convolutional neural network-based techniques for the monomodal super-resolution: optimization methods, weight initialization, network depth, residual learning, filter size in convolution layers, number of the filters, training patch size and number of training subjects. Second, our study also highlights that one single network can efficiently handle multiple arbitrary scaling factors based on a multiscale training approach. Third, we further extend our super-resolution networks to the multimodal super-resolution using intermodality priors. Fourth, we investigate the impact of transfer learning skills onto super-resolution performance in terms of generalization among different datasets. Lastly, the learnt models are used to enhance real clinical low-resolution images. Results tend to demonstrate the potential of deep neural networks with respect to practical medical image applications.

*Keywords:* super-resolution, 3D convolutional neural network, brain MRI

---

\*Corresponding author

*Email addresses:* [ch.pham@imt-atlantique.fr](mailto:ch.pham@imt-atlantique.fr) (Chi-Hieu Pham),  
[carlos.tordiez@imt-atlantique.fr](mailto:carlos.tordiez@imt-atlantique.fr) (Carlos Tor-D iez), [hmeunier@chu-reims.fr](mailto:hmeunier@chu-reims.fr) (H el ene Meunier),  
[nbednarek@chu-reims.fr](mailto:nbednarek@chu-reims.fr) (Nathalie Bednarek), [ronan.fablet@imt-atlantique.fr](mailto:ronan.fablet@imt-atlantique.fr) (Ronan Fablet),  
[nicolas.passat@univ-reims.fr](mailto:nicolas.passat@univ-reims.fr) (Nicolas Passat), [francois.rousseau@imt-atlantique.fr](mailto:francois.rousseau@imt-atlantique.fr) (Fran ois Rousseau)

## 1. Introduction

Magnetic Resonance Imaging (MRI) is a powerful imaging modality for in vivo brain visualization with a typical image resolution of 1mm. Acquisition time of MRI data and signal-to-noise ratio are two parameters that drive the choice of an appropriate image resolution for a given study. The accuracy of further analysis such as brain morphometry can be highly dependent on image resolution. Super-Resolution (SR) aims to enhance the resolution of an imaging system using single or multiple data acquisitions (Milanfar 2010). Increasing image resolution through super-resolution is a key to more accurate understanding of the anatomy (Greenspan 2008). Previous works have shown that applying super-resolution techniques leads to more accurate segmentation maps of brain MRI data (Rueda et al. 2013; Tor-Díez et al. 2019) or cardiac data (Oktay et al. 2016).

The use of SR techniques has been studied in many works in the context of brain MRI analysis: structural MRI (Manjón et al. 2010b; Rousseau et al. 2010a; Manjón et al. 2010a; Rueda et al. 2013; Shi et al. 2015), diffusion MRI (Scherrer et al. 2012; Poot et al. 2013; Fogtman et al. 2014; Steenkiste et al. 2016), spectroscopy MRI (Jain et al. 2017), quantitative  $T_1$  mapping (Ramos-Llordén et al. 2017; Van Steenkiste et al. 2017), fusion of orthogonal scans of moving subjects (Gholipour et al. 2010; Rousseau et al. 2010b; Kainz et al. 2015; Jia et al. 2017). The development of efficient and accurate SR techniques for 3D MRI data could be a major step forward for brain studies.

Most SR methods rely on the minimization of a cost function consisting of a fidelity term related to an image acquisition model and a regularization term that constrains the space of solutions. The observation model is usually a linear model including blurring, motion, the effect of the point spread function (PSF) and downsampling. The regularizer, which guides the optimization process while avoiding unwanted image solutions, can be defined using pixel-based  $\ell_2$ -norm term (Gholipour et al. 2010), total variation (Shi et al. 2015), local patch-based similarities (Manjón et al. 2010b,a; Rousseau et al. 2010a), sparse coding (Rueda et al. 2013), low rank property (Shi et al. 2015).

In particular, the choice of this regularization term remains difficult as it modifies implicitly the space of acceptable solutions without any guaranty on the reconstruction of realistic high-resolution images. Conversely, in a supervised context (in which one can exploit a learning database with low-resolution (LR) and high-resolution (HR) images), the SR of MRI data can be fully driven by examples. The key challenge of supervised techniques is then to accurately estimate the mapping operator from the LR image space to the HR one. Recently, significant advances were reported in SR for computer vision using convolutional neural networks (CNN). This trend follows the tremendous outcome of CNN-based schemes for a wide range of computer vision applications, including for instance image classification (Krizhevsky et al. 2012; Simonyan and Zisserman 2014; He et al. 2016), medical image segmentation (Kamnitsas et al. 2017) or medical image analysis (Tajbakhsh et al. 2016).

CNN architectures have become the state-of-the-art for image SR. Initially, Dong et al. (2016a) proposed a three-layer CNN architecture. The first convolutional layer implicitly extracts a set of feature maps for the input LR image; the second layer maps these feature maps nonlinearly to HR patch representations; and the third layer reconstruct the HR

71 image from these patch representations. Several studies have further investigated CNN-  
 72 based architectures for image SR. Among others, the following features have been reported  
 73 to improve SR performance: an increased depth of the network (Kim et al. 2016a), residual  
 74 block (with batch normalization and skip connection) (Ledig et al. 2017), sub-pixel layer (Shi  
 75 et al. 2016), perceptual loss function (instead of mean squared error-based cost functions)  
 76 (Johnson et al. 2016; Ledig et al. 2017; Zhao et al. 2017), recurrent networks (Kim et al.  
 77 2016b), generative adversarial networks (Ledig et al. 2017; Pham et al. 2019). Very recently,  
 78 Chen et al. (2018) proposed a 3D version of densely connected networks (Huang et al. 2017)  
 79 for brain MRI SR. Inspired by the work by Jog et al. (2016), Zhao et al. (2018) investigated  
 80 self super-resolution for MRI using enhanced deep residual networks (Lim et al. 2017).

81 Recently, very deep architectures obtained the best performance in a challenge focusing  
 82 on natural image SR (NTIRE 2017 challenge, Timofte et al. 2017). However, due to the  
 83 variety of the proposed methods and the high number of parameters for the networks archi-  
 84 tecture design, it is currently difficult to identify the key elements of a CNN architecture  
 85 to achieve good performance for image SR and assess their applicability in the context of  
 86 3D brain MRI. In addition the extension of CNN architectures to 3D images, taking into  
 87 account floating and possibly anisotropic scaling factors may be of interest to address the  
 88 wide range of possible clinical acquisition settings, whereas classical CNN architectures only  
 89 address a predefined (integer) scaling factor. The availability of multimodal imaging setting  
 90 also questions the ability of CNN architectures to benefit from such multimodal data to  
 91 improve the SR of a given modality.

92 **Contributions:** This work presents a comprehensive review of deep convolutional neural  
 93 networks, and associated key elements, for brain MRI SR. Following Timofte et al. (2016),  
 94 who have experimentally shown several ways to improve SR techniques from a baseline archi-  
 95 tecture, we study the impact of eight key elements on the performance of convolutional  
 96 neural networks for 3D brain MRI SR. We demonstrate empirically that residual learning  
 97 associated with appropriate optimization methods can significantly reduce the time of the  
 98 training step and fast convergence can be achieved in 3D SR context. Overall, we report  
 99 better performance when learning deeper fully 3D convolution neural networks and using  
 100 larger filters. Interestingly, we demonstrate that a single network can handle multiple arbi-  
 101 trary scale factors efficiently, for example, from  $2 \times 2 \times 2\text{mm}$  to  $2 \times 2 \times 1\text{mm}$  or  $1 \times 1 \times 1\text{mm}$ ,  
 102 by learning multiscale residuals from spline-interpolated image. We also report significant  
 103 improvement using a multimodal architecture, where a HR reference image can guide the  
 104 CNN-based SR of a given MRI volume. Moreover, we demonstrate that our model can trans-  
 105 fer the rich information available from high-resolution experimental dataset to lower-quality  
 106 clinical image data.

## 107 2. Super-Resolution using Deep Convolutional Neural Networks

### 108 2.1. Learning-based SR

109 Single image SR is a typically ill-posed inverse problem that can be stated according to  
 110 the following linear formulation:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} = D_{\downarrow}\mathbf{B}\mathbf{X} + \mathbf{N} \quad (1)$$

111 where  $\mathbf{Y} \in \mathbb{R}^n$  and  $\mathbf{X} \in \mathbb{R}^m$  denote a LR image and a HR image,  $\mathbf{H} \in \mathbb{R}^{m \times n}$  is the observation  
 112 matrix ( $m > n$ ) and  $\mathbf{N}$  denotes an additive noise.  $D_{\downarrow}$  represents the downsampling operator  
 113 and  $B$  is the blur matrix. The purpose of SR methods is to estimate  $\mathbf{X}$  from the observations  
 114  $\mathbf{Y}$ . The SR image can be estimated by minimizing a least-square cost function:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2. \quad (2)$$

115 The minimization of the Equation (2) usually leads to unstable solutions and requires the use  
 116 of appropriate regularization terms on  $\mathbf{X}$ . Adding prior knowledge on the image solution  
 117 (such as piecewise smooth image) may lead to unrealistic solution. In a learning-based  
 118 context where a set of image pairs  $(\mathbf{X}_i, \mathbf{Y}_i)$  is available, the objective is to learn the mapping  
 119 from the LR images  $\mathbf{Y}_i$  to the HR images  $\mathbf{X}_i$ , leading to the following formulation:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{X} - \mathbf{H}^{-1}\mathbf{Y}\|^2. \quad (3)$$

120 In this setting, the matrix  $\mathbf{H}^{-1}$  can be modeled as a combination of a restoration matrix  
 121  $F \in \mathbb{R}^{m \times m}$  and an upscaling interpolation operator  $S^{\uparrow} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Given a set of HR  
 122 images  $\mathbf{X}_i$  and their corresponding LR images  $\mathbf{Y}_i$  with  $K$  samples, the restoration operator  
 123  $F$  can be estimated as follows:

$$\hat{F} = \arg \min_F \sum_{i=1}^K \|\mathbf{X}_i - F(S^{\uparrow}\mathbf{Y}_i)\|^2 = \arg \min_F \sum_{i=1}^K \|\mathbf{X}_i - F(\mathbf{Z}_i)\|^2 \quad (4)$$

124 where  $\mathbf{Z} \in \mathbb{R}^m$  is the interpolated LR (ILR) version of  $\mathbf{Y}$  (*i.e.*  $\mathbf{Z} = S^{\uparrow}\mathbf{Y}$ ).  $F$  is then a  
 125 mapping from the ILR image space to the HR image space.

126 SR is the process of estimating HR data from LR data. The main goal is then to estimate  
 127 high-frequency components from LR observations. Instead of learning the mapping directly  
 128 from the LR space to the HR one, it might be easier to estimate a mapping from the LR  
 129 space to the missing high-frequency components, also called the residual between HR and  
 130 LR data:  $\mathbf{R} = \mathbf{X} - \mathbf{Z}$  or equivalently  $\mathbf{X} = \mathbf{Z} + \mathbf{R}$ . This approach can be modeled by a skip  
 131 connection in the network. In such a residual-based modeling, one typically assumes that  $\mathbf{R}$   
 132 is a function of  $\mathbf{Z}$ . The computation of HR data is then expressed as follows:  $\mathbf{X} = \mathbf{Z} + F(\mathbf{Z})$   
 133 where  $F$  can be learnt using the following equation:

$$\hat{F} = \arg \min_F \sum_{i=1}^K \|(\mathbf{X}_i - \mathbf{Z}_i) - F(\mathbf{Z}_i)\|^2. \quad (5)$$

## 134 2.2. CNN-based baseline architecture

135 In this paper, we focus on the learning of the mapping  $F$  with convolutional neural  
 136 networks. Following [Dong et al. \(2016a\)](#) and [Kim et al. \(2016a\)](#), the mapping  $F$  from  $\mathbf{Z}$   
 137 to  $(\mathbf{X} - \mathbf{Z})$  is decomposed into nonlinear operations that correspond to the combination of  
 138 convolution-based and rectified linear unit (ReLU) layers.

139 The baseline architecture used in this work can be described as follows:

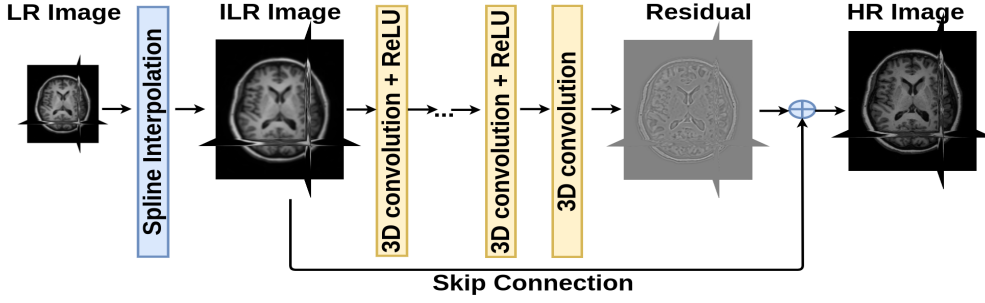


Figure 1: 3D deep neural network for single brain MRI super-resolution.

$$\begin{cases} F_1(\mathbf{Z}) = \max(0, W_1 * \mathbf{Z} + B_1) \\ F_i(\mathbf{Z}) = \max(0, W_i * F_{i-1}(\mathbf{Z}) + B_i) & \text{for } 1 < i < L \\ F_L(\mathbf{Z}) = W_L * F_{L-1}(\mathbf{Z}) + B_L \end{cases} \quad (6)$$

140 where:

- 141 •  $L$  is the number of layers,
- 142 •  $W_i$  and  $B_i$  are the parameters of convolution layers to learn.  $W_i$  corresponds to  $n_i$
- 143 convolution filters of support  $c \times f_i \times f_i \times f_i$ , where  $c$  is the number of channels in the
- 144 input of layer  $i$ ,  $f_i$  and  $n_i$  are respectively the spatial size of the filters and the number
- 145 of filters of layer  $i$ ,
- 146 •  $\max(0, \cdot)$  refers to a ReLU applied to the filter responses.

147 This network architecture is depicted in Figure 1. Please note that, for instance, the SRCNN

148 model proposed by [Dong et al. 2016a](#) corresponds to a specific parameterization of this

149 baseline architecture with  $f_1 = 9$ ,  $f_2 = 1$ ,  $f_3 = 5$ ,  $n_1 = 64$ ,  $n_2 = 32$  and no skip connection.

150 The performance of a given architecture depends on several parameters such as the filter

151 size  $f_i$ , the number of filters  $n_i$ , the number of layers  $L$ , etc. Understanding how these

152 parameters influence the reconstruction of the HR image with respect to the considered

153 application setting (e.g., number of training samples, image size, scaling factor) is a key issue,

154 which remains poorly explored. For instance, regarding the number of layers, it is commonly

155 believed that the deeper the better ([Simonyan and Zisserman 2014](#); [Kim et al. 2016a](#)).

156 However, adding layers increases the number of parameters and can lead to overfitting. In

157 particular, previous works ([Dong et al. 2016a](#); [Oktay et al. 2016](#)), have shown that a deeper

158 structure does not always lead to better results ([Dong et al. 2016a](#)).

159 Specifically focusing on MRI data, the specific objectives of this study are: i) the eval-

160 uation and understanding of the effect of key elements of CNN for brain MRI SR, ii) the

161 experimental study of arbitrary multiscale SR using CNN, iii) investigating multimodality-

162 guided SR using CNN.

### 3. Sensitivity Analysis of the Considered Architecture

In this section, we present the MRI datasets used for evaluation and the key elements of CNN architecture to achieve good performance for single image SR.

#### 3.1. MRI Datasets and LR simulation

To evaluate SR performances of CNN-based architectures, we have used two MRI datasets: the Kirby 21 dataset and the NAMIC Brain Multimodality dataset.

The Kirby 21 dataset (Landman et al. 2011) consists of MRI scans of twenty-one healthy volunteers with no history of neurological conditions. Magnetization prepared gradient echo (MPRAGE, T1-weighted) scans were acquired using a 3T MR scanner (Achieva, Philips Healthcare, The Netherlands) with a  $1.0 \times 1.0 \times 1.2\text{mm}^3$  resolution over an FOV of  $240 \times 204 \times 256\text{mm}$  acquired in the sagittal plane. Flair data were acquired using  $1.1 \times 1.1 \times 1.1\text{mm}^3$  resolution over an FOV of  $242 \times 180 \times 200\text{mm}$  acquired in the sagittal plane. The T2-weighted volumes were acquired using a 3D multi-shot turbo-spin echo (TSE) with a TSE factor of 100 with over an FOV of  $200 \times 242 \times 180\text{mm}$  including a sagittal slice thickness of 1mm.

MR images of NAMIC Brain Multimodality<sup>1</sup> dataset have been acquired using a 3T GE at BWH in Boston, MA. An 8 Channel coil was used in order to perform parallel imaging using ASSET (Array Spatial Sensitivity Encoding techniques, GE) with a SENSE-factor (speed-up) of 2. The structural MRI acquisition protocol included two MRI pulse sequences. The first results in contiguous spoiled gradient-recalled acquisition (fastSPGR) with the following parameters: TR=7.4ms, TE=3ms, TI=600, 10 degree flip angle,  $25.6\text{cm}^2$  field of view, matrix= $256 \times 256$ . The voxel dimensions are  $1 \times 1 \times 1\text{mm}^3$ . The second XETA (eXtended Echo Train Acquisition) produces a series of contiguous T2-weighted images (TR=2500ms, TE=80ms,  $25.6\text{cm}^2$  field of view, 1 mm slice thickness). Voxel dimensions are  $1 \times 1 \times 1\text{mm}^3$ .

As in Shi et al. 2015 and Rueda et al. 2013, LR images have been generated from a Gaussian blur and a down-sampling by isotropic scaling factors. In the training phase, a set of patches of training images is randomly extracted. In the baseline setting, the training dataset comprises 10 subjects (3 200 patches  $25 \times 25 \times 25$  per subject randomly sampled) and the testing dataset is composed of 5 subjects. During the testing step, the network is applied on the whole images. The peak signal-to-noise ratio (PSNR) in decibels (dB) is used to evaluate the SR results with respect to the original HR images. No denoising or bias correction algorithms were applied to the data. Image intensity has been normalized between 0 and 1. The following figures are drawn based on the average PSNR over all test images.

#### 3.2. Baseline and benchmarked architectures

The network architecture that is used as a baseline approach in this study is illustrated in Figure 1. The baseline network is a 10 blocks (convolution+ReLU) network with the following parameters: 64 convolution filters of size  $(3 \times 3 \times 3)$  at each layer, mean squared

---

<sup>1</sup>NAMIC : <http://hdl.handle.net/1926/1687>



201 error (MSE) as loss function, weight initialization by He et al. (2015) (MSRA filler), Adam  
 202 (adaptive moment estimation) method for optimization (Kingma and Ba 2015), 20 epochs  
 203 on Nvidia GPU and using Caffe package (Jia et al. 2014), batch size of 64, learning rate  
 204 set to 0.0001, no regularization or drop out has been used. The learning rate multipliers  
 205 of weights and biases are 1 and 0.1, respectively. For benchmarking purposes, we consider  
 206 two other state-of-the-art SR models: low-rank total variation (LRTV) (Shi et al. 2015) and  
 207 SRCNN3D (Pham et al. 2017). SRCNN3D (Pham et al. 2017), which is a 3D extension of  
 208 the method described in (Dong et al. 2016a), has 3 convolutional layers with the size of  $9^3$ ,  
 209  $1^3$  and  $5^3$ , respectively. The layers of SRCNN3D consist of 64 filters, 32 filters and one filter,  
 210 respectively.

211 The next sections present the impact of the key parameters studied in this work: op-  
 212 timization method, weight initialization, residual-based model, network depth, filter size,  
 213 filter number, training patch size and size of training dataset.

### 214 3.3. Optimization Method

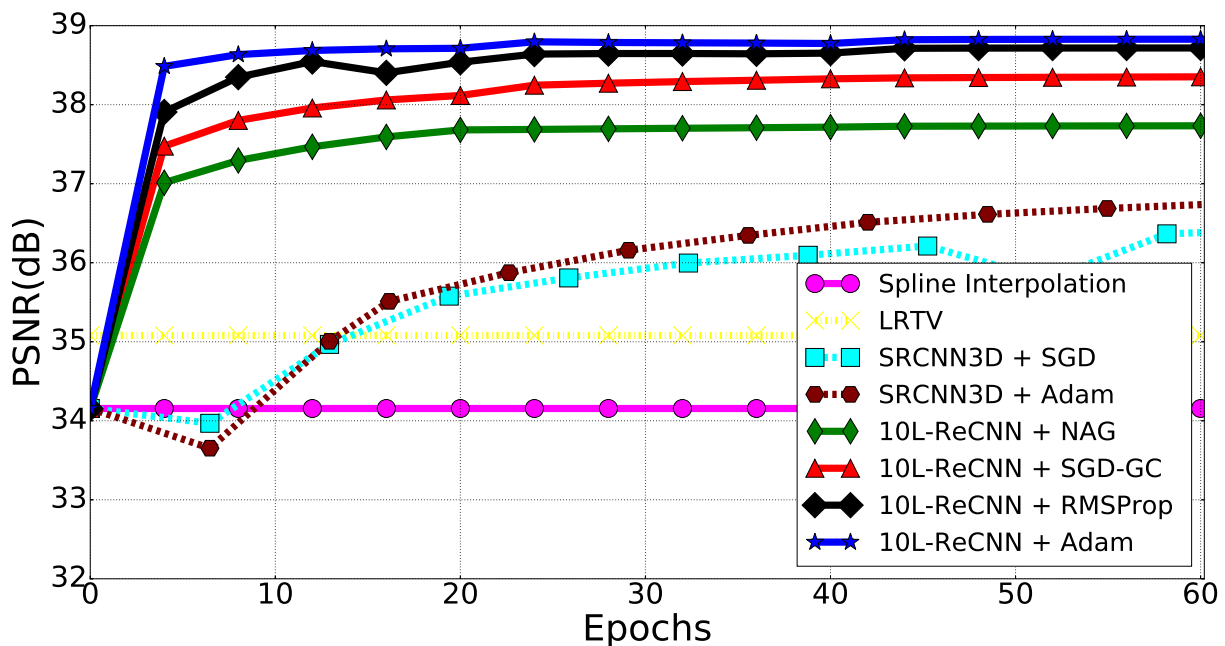


Figure 2: Impact of the optimization methods onto SR performance: SGD-GC, NAG, RMSProp and Adam optimisation of a 10L-ReCNN (10-layer residual-learning network with  $f = 3$  and  $n = 64$ ). We used Kirby 21 for training and testing with isotropic scaling factor  $\times 2$ . The initial learning rates of SGD-GC, NAG, RMSProp and Adam are set respectively to 0.1, 0.0001, 0.0001 and 0.0001. These learning rates are decreased by a factor of 10 every 20 epochs. The momentum of these methods, except RMSProp, is set to 0.9. All optimization methods use the same weight initialization described by He et al. (2015).

215 Given a training dataset which consists of pairs of LR and HR images, network param-  
 216 eters are estimated by minimizing the objective function using optimization algorithms. These  
 217 algorithms play a very important role in training neural networks. The more efficient and



218 effective optimization strategies lead to faster convergence and better performance. More  
 219 precisely, during the training step, the estimation of the restoration operator  $F$  corresponds  
 220 to the minimization of the objective function  $\mathcal{L}$  in Equation (5) over network parameters  
 221  $\theta = \{W_i, B_i\}_{i=1, \dots, L}$ .

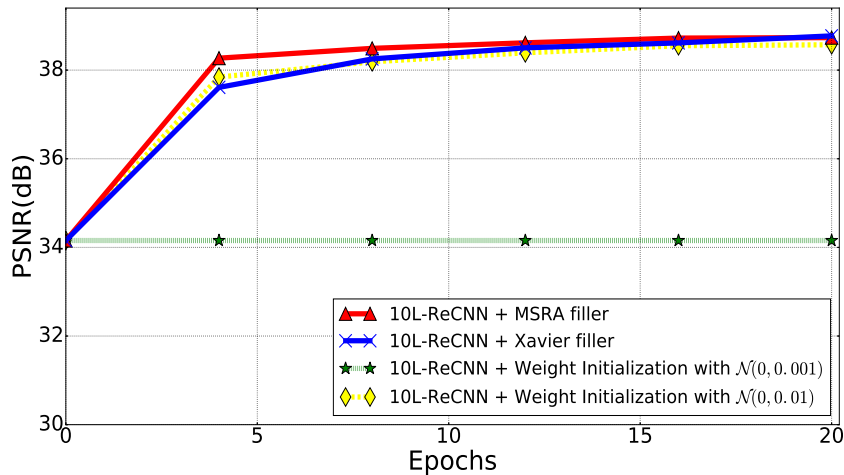
222 Most optimization methods for CNNs are based on gradient descent. A classical method  
 223 applies a mini-batch stochastic gradient descent with momentum (SGD) (LeCun et al. 1998)  
 224 as used by Dong et al. (2016a); Pham et al. (2017). However, the use of fixed momentum  
 225 causes numerical instabilities around the minimum. Nesterov’s accelerated gradient (NAG)  
 226 (Nesterov 1983) was proposed to cope with this issue but the use of small learning rates  
 227 induces slow convergence. By contrast, high learning rates may lead to exploding gradients  
 228 (Bengio et al. 1994; Glorot and Bengio 2010). In order to address this issue, Kim et al.  
 229 (2016a) proposed the stochastic gradient descent method with an adjustable gradient clip-  
 230 ping (SGD-GC) (Pascanu et al. 2013) to achieve an optimization with high learning rates.  
 231 The predefined range over which gradient clipping is applied may still cause SGD-GC not  
 232 to converge quickly or make difficult the tuning of a global learning rate. Recently, methods  
 233 have been proposed to address this issue through an automatic adaption of the learning rate  
 234 for each parameter to be learnt. RMSProp (root-mean-square propagation) (Tieleman and  
 235 Hinton 2012) and Adam (adaptive moment estimation) (Kingma and Ba 2015) are the two  
 236 most popular models in this category.

237 The results of four optimization methods (NAG, SGD-GC, RMSProp and Adam) for  
 238 the baseline network are illustrated in Figure 2. Firstly, regardless the method used, the  
 239 baseline network shows better performance than LRTV (Shi et al. 2015) and SRCNN3D  
 240 (Pham et al. 2017). Secondly, it can be observed that the baseline network can converge very  
 241 rapidly (only 20 epochs with small learning rate of 0.0001). Finally, in these experiments,  
 242 the most efficient and effective optimization method is Adam as regards both PSNR metric  
 243 and convergence speed. Hence, in the next sections, we use Adam method with  $\beta_1 = 0.9$   
 244 and  $\beta_2 = 0.999$  to train our networks with 20 epochs.

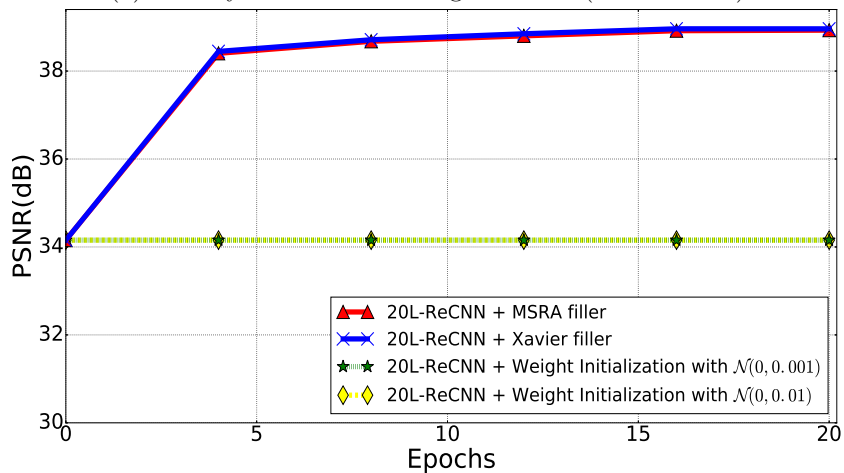
### 245 3.4. Weight Initialization

246 The optimization algorithms for training a CNN are typically initialized randomly. In-  
 247 appropriate initialization can lead to long time convergence or even divergence. Several  
 248 studies (Dong et al. 2016a; Oktay et al. 2016; Pham et al. 2017) used a normal distribu-  
 249 tion  $\mathcal{N}(0, 0.001)$  to initialize the weights of convolutional filters. However, because of too  
 250 small initial weights, the optimizer may be stuck into a local minimum especially when  
 251 building deeper networks. Both Dong et al. (2016a) concluded that deeper networks do not  
 252 lead to better performance, and Oktay et al. (2016) confirmed that the addition of extra  
 253 convolutional layers to the 7-layer model is found to be ineffective. Uniform distribution  
 254  $\mathcal{U}(-\sqrt{3/(nf^3)}, \sqrt{3/(nf^3)})$  (called Xavier filler) (Glorot and Bengio 2010) was also proposed  
 255 to initialize the weights of deeper networks. In order to add more layers to networks, He  
 256 et al. (2015) suggested an initial training stage by sampling from the normal distribution  
 257  $\mathcal{N}(0, \sqrt{2/(nf^3)})$  (called here Microsoft Research Asia - MSRA filler).

258 Overall, we evaluate here the weight initialization schemes described by Glorot and Ben-  
 259 gio (2010) and He et al. (2015), a normal distribution  $\mathcal{N}(0, 0.001)$  as proposed by Dong et al.



(a) : 10-layer residual-learning networks (10L-ReCNN)



(b) : 20-layer residual-learning networks (20L-ReCNN)

Figure 3: Weight initialization scheme vs. performance (residual-learning networks with the same filter numbers  $n = 64$  and filter size  $f = 3$  using Adam optimization and tested with isotropic scaling factor  $\times 2$  using Kirby 21 for training and testing, 32 000 patches with size  $25^3$  for training).

260 (2016a); Oktay et al. (2016) and a normal distribution  $\mathcal{N}(0, 0.01)$  for the considered SR ar-  
 261 chitecture. Experiments with a deeper architecture were also performed, more precisely for  
 262 a 20-layer architecture, which is the deepest architecture that could be implemented for  
 263 the considered experimental setup due to GPU memory setting. As shown in Figure 3,  
 264 the initialization with normal distributions  $\mathcal{N}(0, 0.001)$  failed to make the training of both  
 265 10-layer and 20-layer residual-learning networks converge. In addition, our 20-layer network  
 266 also does not converge when initialized with normal distributions  $\mathcal{N}(0, 0.01)$ . By contrast,  
 267 MSRA and Xavier filler schemes make the networks converge and reach similar reconstruc-  
 268 tion performance. For the rest of this paper, we use MSRA weight filler as initialization  
 269 scheme.

270 *3.5. Residual Learning*

271 The CNN methods proposed by Dong et al. (2016a); Shi et al. (2016); Dong et al.  
 272 (2016b) use the LR image as input and outputs the HR one. We refer to such approach  
 273 as a non-residual learning. Within these approaches, low-frequency features are propagated  
 274 through the layers of networks, which may increase the representation of redundant features  
 275 in each layer and in turn the computational efficiency of the training stage. By contrast,  
 276 one may consider residual learning or normalized HR patch prediction as pointed out by  
 277 several learning-based SR methods (Zeyde et al. 2012; Timofte et al. 2013, 2014; Kim et al.  
 278 2016a). When considering CNN methods, one may design a network which predicts the  
 279 residual between the HR image and the output of the first transposed convolutional layer  
 280 (Oktay et al. 2016). Using residual blocks, a CNN architecture may implicitly embed residual  
 281 learning while still predicting the HR image (Ledig et al. 2017).

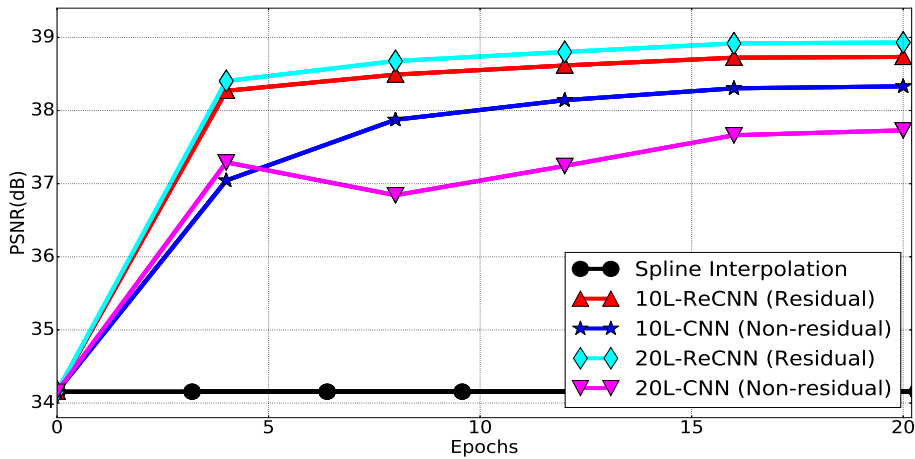


Figure 4: Non-residual-learning vs Residual-learning networks with the same  $n = 64$  and  $f^3 = 3^3$  and the depths of 10 and 20 (called here 10L-CNN vs 10L-ReCNN and 20L-CNN vs 20L-ReCNN) over 20 training epochs using Adam optimization with the same training strategy and tested with isotropic scale factor  $\times 2$  using Kirby 21 for training and testing.

282 Here, we perform a comparative evaluation of non-residual learning vs. residual learning  
 283 strategies. Figure 4 depicts PSNR values and convergence speed of residual vs. non-residual  
 284 network structures with 10 and 20 convolutional layers. The residual-learning networks  
 285 converge faster than the non-residual-learning ones. In addition, residual learning leads  
 286 to improvements in PSNR (+0.4dB for 10 layers and +1.2dB for 20 layers). It might be  
 287 noted that these experiments do not support the common statement that the deeper, the  
 288 better for CNNs. Here, the use of additional layers is only beneficial when using residual  
 289 modeling. Deeper architectures may even lower the reconstruction performance with non-  
 290 residual learning.

291 *3.6. Depth, Filter Size and Number of Filters*

292 As shown by the previous experiment, the link between network depth and performance  
 293 remains unclear. Besides, it is hard to train deeper networks because gradient computation

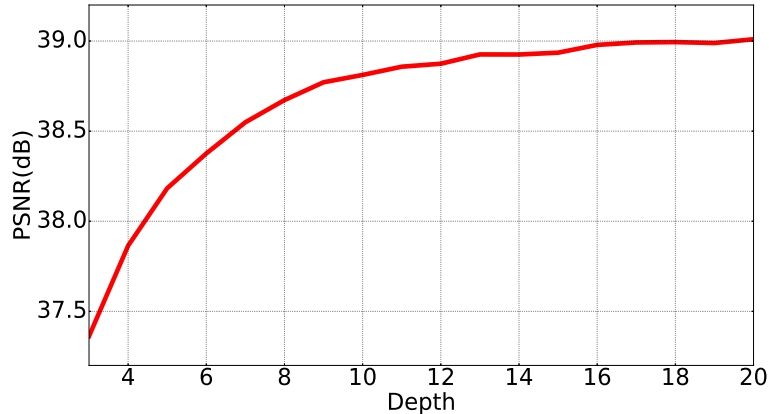


Figure 5: Depth vs Performance (residual-learning networks with the same filter numbers  $n = 64$  and filter size  $f = 3$  over 20 training epochs using Adam optimization and tested with isotropic scale factor  $\times 2$  using Kirby 21 for training and testing, 32 000 patches with size  $25^3$  for training).

294 can be unstable when adding layers (Glorot and Bengio 2010). For instance, Oktay et al.  
 295 (2016) tested extra convolutional layers to a 7-layer model but achieved negligible perfor-  
 296 mance improvement. As mentioned in Section 2.2, SRCNN (Dong et al. 2016a) was also  
 297 tested with deeper architectures but no improvement was reported. However, Kim et al.  
 298 (2016a) argue that the performance of CNNs for SR could be improved by increasing the  
 299 depth of network compared to neural network architectures proposed by Dong et al. (2016a);  
 300 Oktay et al. (2016).

301 The previous section supports that deeper architectures may be beneficial when consid-  
 302 ering a residual learning. We now evaluate the reconstruction performance as a function of  
 303 the number of layers. Results are reported in Figure 5. They stress that increasing network  
 304 depth with residual learning improves the quality of the estimated HR image (e.g. +1.6dB  
 305 increasing of the depth from 3 to 20 or +0.5dB increasing of the depth from 7 to 20).

306 The parameterization of the convolutional filters is also of key interest. Inspired by the  
 307 VGG network designed for classification (Simonyan and Zisserman 2014), previous CNN  
 308 methods for SR mostly focused on small convolutional filters of size  $(3 \times 3 \times 3)$  as proposed  
 309 by Kim et al. (2016a); Oktay et al. (2016); Kamnitsas et al. (2017). Oktay et al. (2016)  
 310 even argued that such architecture can lead to better non-linear estimations. Regarding  
 311 the number of filters for each layer, Dong et al. (2016a) reported greater reconstruction  
 312 performance when increasing the number of filters. But these experiences were not reported  
 313 in other CNN-based SR studies (Kim et al. 2016a; Oktay et al. 2016). Here, we both evaluate  
 314 the effect of the filter size and of the number of filters.

315 Figure 6 shows that a 10-layer network with a filter size of  $5^3$  shows results as well as  
 316 a 20-layer network with  $3^3$  filters. Besides reconstruction performance, the use of a larger  
 317 filter size decreases the training speed and significantly increases the complexity and memory  
 318 cost for training. For example, it took us 50 hours to train a 10-layer network with a filter  
 319 size of  $5^3$ . By contrast, a deeper network with smaller filters (i.e. 20-layer network with  $3^3$   
 320 filters) involves a smaller number of parameters, such that it took us only 24 hours to train.

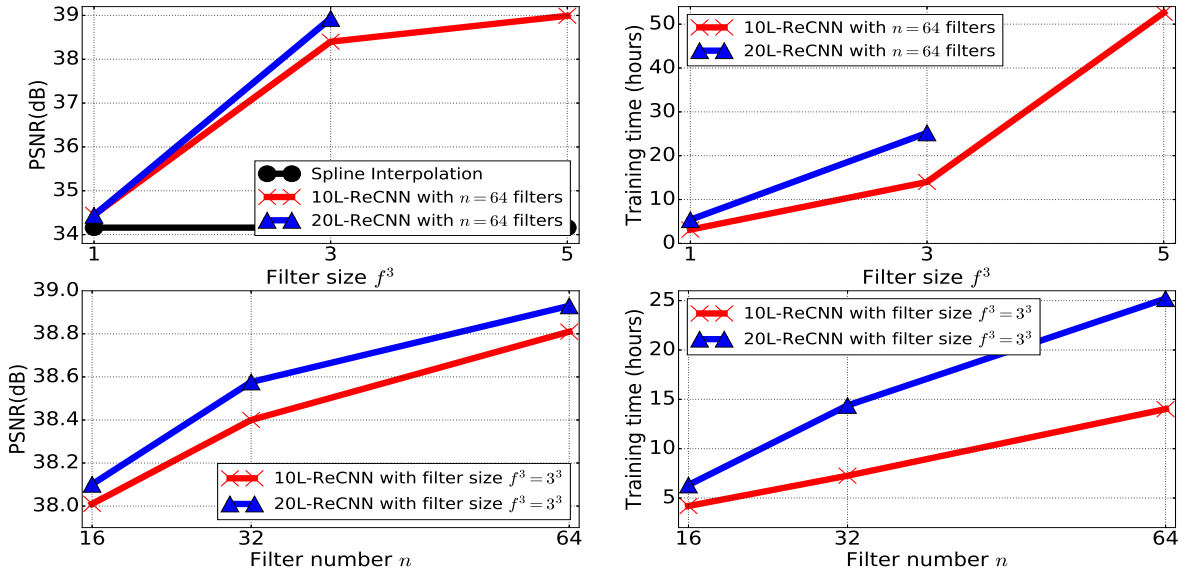


Figure 6: Impact of convolution filter parameters (sizes  $f \times f \times f = f^3$  with  $n$  filters) on PSNR and computation time. These 10-layers residual-learning networks are trained from scratch using Kirby 21 with Adam optimization over 20 epochs and tested with the testing images of the same dataset for isotropic scale factor  $\times 2$ .

321 These experiments suggest that deeper architectures with small filters can replace shallower  
 322 networks with larger filters both in terms of computational complexity and of reconstruction  
 323 performance. In addition, the increase in the number of filters within networks can increase  
 324 the performance. However, we were not able to use 128 filters with the baseline architecture  
 325 due to the limited amount of memory. This stresses out the need to design memory efficient  
 326 architectures for 3D image processing using deeper CNNs with more filters.

### 327 3.7. Training Patch Size and Subject Number

328 In the context of brain MRI SR, the acquisition and collection of large datasets with  
 329 homogeneous acquisition settings is a critical issue. We now evaluate the extent to which  
 330 the number of training subjects influences SR reconstruction performance. As the training  
 331 samples are extracted as patches of brain MRI images, we also evaluate the impact of the  
 332 training patch size on learning and reconstruction performances.

333 The size of training patches should be greater or equal to the size of the receptive field  
 334 of the considered network (Simonyan and Zisserman 2014; Kim et al. 2016a), which is given  
 335 by  $((f - 1)D + 1)^3$  for a  $D$ -layer network with filter size  $f^3$ . Figure 7 confirms that better  
 336 performances can be achieved using larger training patches (from  $11^3$  to  $31^3$  with the 10-layer  
 337 network and from  $11^3$  to  $29^3$  with the 12-layer network). However, if the patch size is larger  
 338 than the receptive field (e.g.  $21^3$  within the 10-layers network and  $25^3$  within the 12-layers  
 339 network), then the improvement is negligible consuming, in the meantime, considerably  
 340 more GPU memory and training time.

341 We stressed previously that the selection of the network depth involves a trade-off be-  
 342 tween reconstruction performance and GPU memory requirement and training time increase.

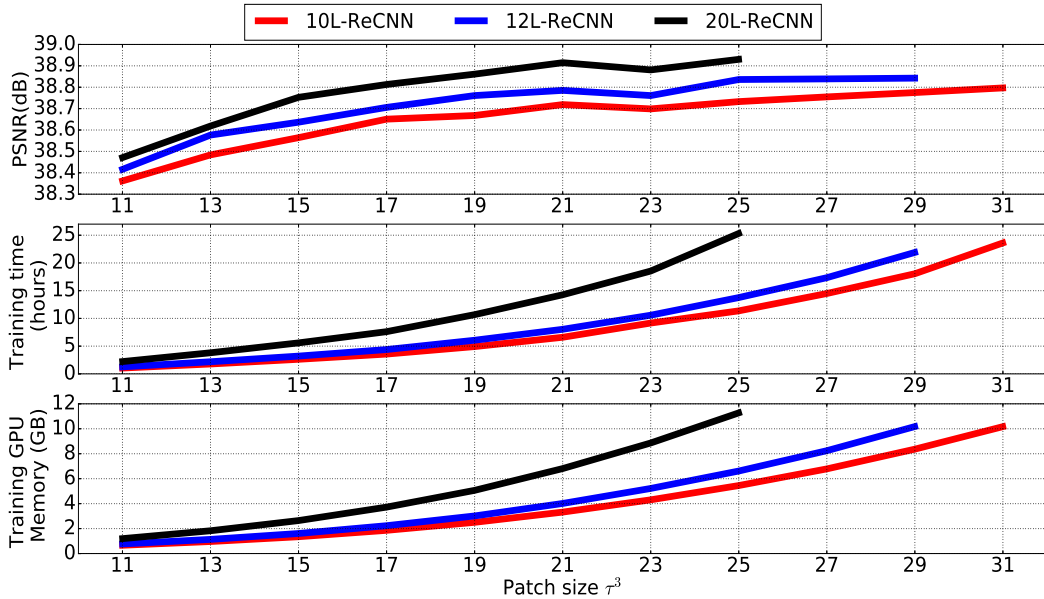


Figure 7: First row: Training patch size vs. performance. Second row: Patch size vs. training time. Third row: Patch size vs training GPU memory requirement. These networks with the same  $n = 64$  and  $f^3 = 3^3$  are trained from scratch using Kirby 21 with batch of 64 and tested with the testing images of the same dataset for isotropic scale factor  $\times 2$ .

343 A similar result can be drawn with respect to the patch size. Figure 7 illustrates that larger  
 344 training patch sizes also require more memory for training. Moreover, this figure shows that  
 345 better performance can be achieved using larger training patch sizes. It may be noted that  
 346 the performance of the 10-layer networks may reach a performance similar to 12-layer and  
 347 20-layer networks when using larger training patches but it takes more time and more GPU  
 348 memory for training.

349 Regarding the number of training subjects, Figure 8 points out that a single subject is  
 350 enough to reach better performance than spline interpolation. Interestingly, reconstruction  
 351 performance increases slightly when more subjects are considered, which appears appropriate  
 352 for real-world applications. However, in fact, more training dataset takes more time within  
 353 the same experience settings. In the next sections, for saving training time, we propose to  
 354 use 10 subjects for learning.

#### 355 4. Handling Arbitrary Scales

356 In some CNN-based SR approaches, the networks are learnt for a fixed and specified  
 357 scaling factor. Thus, a network built for one scaling factor cannot deal with any other  
 358 scale. In medical imaging, Oktay et al. (2016) have applied CNNs for upscaling cardiac  
 359 image slices with the scale of 5 (e.g. upscaling the voxel size from  $1.25 \times 1.25 \times 10.0\text{mm}$  to  
 360  $1.25 \times 1.25 \times 2.00\text{mm}$ ). Typically, their network is not capable of handling other scales due  
 361 to the use of fixed deconvolutional layers. In brain MRI imaging, the variety of the possible  
 362 acquisition settings motivates us to explore multiscale settings.

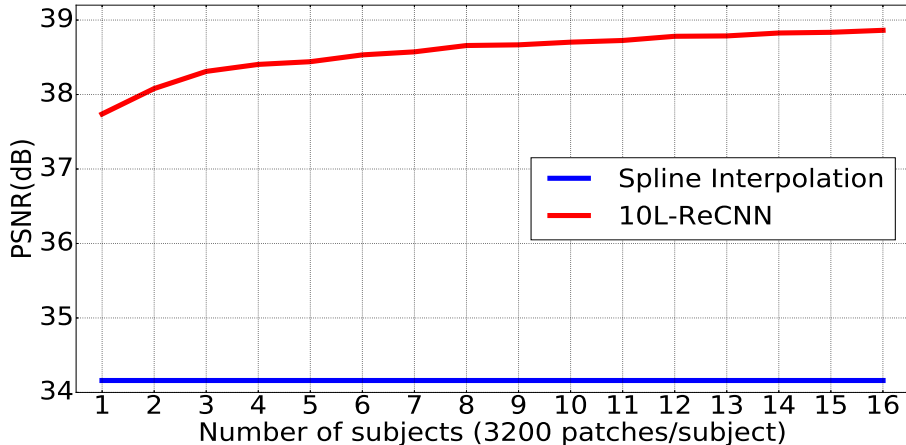


Figure 8: Number of subjects vs. performance (10-layer residual-learning networks with the same filter numbers  $n = 64$  and filter size  $f = 3$  over 20 training epochs using Adam optimization and tested with isotropic scale factor  $\times 2$  using Kirby 21 for training and testing, 3 200 patches per subject with size  $25^3$  for training).

363 Following Kim et al. (2016a), we investigate how we may deal with multiple scales in a  
 364 single network. It consists of creating a training dataset within which we consider LR and  
 365 HR image pairs corresponding to different scaling factors. We test two cases: the first case  
 366 where the learning dataset for combined scale factors ( $\times 2, \times 3$ ) has the same number as a  
 367 single scale factor and the second one where we double the learning dataset for multiple  
 368 scale factors. To avoid a convergence towards a local minimum of one of the scaling factors,  
 369 we learn network parameters on randomly shuffled dataset.

370 Table 1 summarizes experimental results. First, when the training is achieved for the  
 371 scaling from  $(2 \times 2 \times 2)$  on a dataset of  $(2 \times 2 \times 2)$  scale, it can be noticed that reconstruction  
 372 performances decrease significantly when applied to other scaling factors (there is a drop  
 373 from 39.01dB to 33.43dB when testing with  $(3 \times 3 \times 3)$ ). Second, it can be noticed that  
 374 when the training is performed on multiscale data within the same training samples, there is  
 375 no significant performance change compared to training from a single-scale dataset. Third,  
 376 a training dataset with double samples leads to a better performance. Moreover, training  
 377 from multiple scaling factors leads to the estimation of a more versatile network. Overall,  
 378 these results show that one single network can handle multiple arbitrary scaling factors.

## 379 5. Multimodality-guided SR

380 In clinical cases, it is frequent to acquire one isotropic HR image and LR images with  
 381 different contrasts in order to reduce the acquisition time. Hence, a coplanar isotropic HR  
 382 image might be considered as a complementary source of information to reconstruct HR  
 383 MRI images from LR ones (Rousseau et al. 2010a). To address this multimodality-guided  
 384 SR problem, we add a concatenation layer as the first layer of the network, as illustrated in  
 385 Figure 9. This layer concatenates the ILR image and a registered HR reference along the



Test / Train	Full-training			
	Same training samples			Double samples
	$\times(2,2,2)$	$\times(3,3,3)$	$\times(2,2,2),(3,3,3)$	$\times(2,2,2),(3,3,3)$
	PSNR	PSNR	PSNR	PSNR
$\times(2,2,2)$	<b>39.01</b>	35.25	<b>37.35</b>	<b>38.80</b>
$\times(2,2,3)$	36.80	35.11	36.47	37.24
$\times(2,2.5,2)$	37.71	35.41	36.91	37.93
$\times(2,3,3)$	35.23	35.13	35.75	36.20
$\times(2.5,2.5,2.5)$	35.47	35.52	36.09	36.63
$\times(3,3,3)$	33.43	<b>35.01</b>	<b>34.89</b>	<b>35.20</b>

Table 1: Experiments with multiple isotropic scaling factors with the 20-layers network using the training and testing images of Kirby 21. **Bold numbers** indicate that the tested scaling factor is present in the training dataset. We test two conditions of same training data and double training data.

386 channel axis. The registration step of HR reference ensures that the two input images share  
387 the same geometrical space.

388 We experimentally evaluate the relevance of the proposed multimodality-guided SR  
389 model according to the following setting. We investigate whether the complementary use of  
390 a Flair or a T2-weighted MRI image might be beneficial to improve the resolution of a LR  
391 T1-weighted MRI image. Concerning the Kirby 21 dataset, we apply an affine transform  
392 estimated using FSL (Jenkinson et al. 2012) to register images of a same subject into a com-  
393 mon coordinate space. We assume here that the affine registration can compensate motion  
394 between two scans during the same acquisition session. This appears a fair assumption since  
395 here we are considering an organ that does not undergo significant deformation between two  
396 acquisitions. The registration step has been checked visually for all the images. Data of the  
397 NAMIC dataset are already in the same coordinate space so no registration step is required.

398 Figure 10 shows the results of the multimodality-guided SR compared to the monomodal  
399 SR for both Kirby dataset (a) and NAMIC datasets (b). It can be seen that multimodal-  
400 ity driven approach can lead to improved reconstruction results. In these experiments, the  
401 overall upsampling result depends on the quality of the HR image used to drive the re-  
402 construction process. Thus, adding high resolution information containing artifacts limits  
403 reconstruction performance. This is especially the case for the Kirby dataset. For instance,  
404 when considering T2w images, no improvement is observed for Kirby dataset and an im-  
405 provement greater than  $1dB$  is reported for NAMIC dataset. As the T2w image resolution  
406 is lower than T1w modality in Kirby dataset, these results may emphasize the requirement  
407 for actual HR information to expect significant gain w.r.t. the monomodal model. Figure  
408 11 shows visually that edges in the residual image between the ground truth and the recon-  
409 struction by the multimodal approach are reduced significantly compared to interpolation  
410 and monomodal methods (e.g. the regions of lateral ventricles). This means that the mul-  
411 timodal approach resulting the reconstructions which are the most similar to the ground  
412 truth. These qualitative results highlight the fact that the proposed multimodal method  
413 provides a better performance than other compared methods.

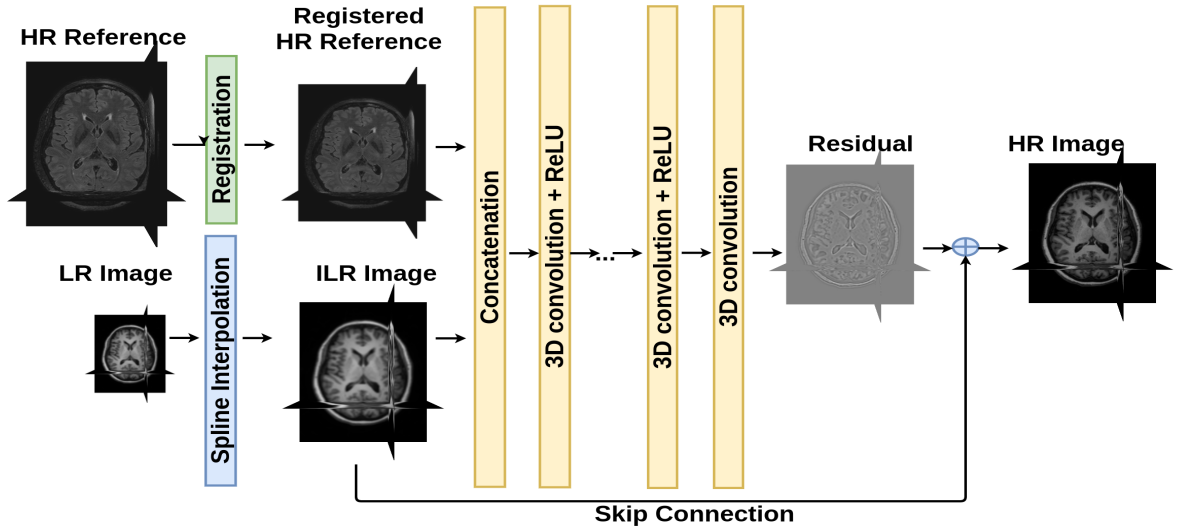


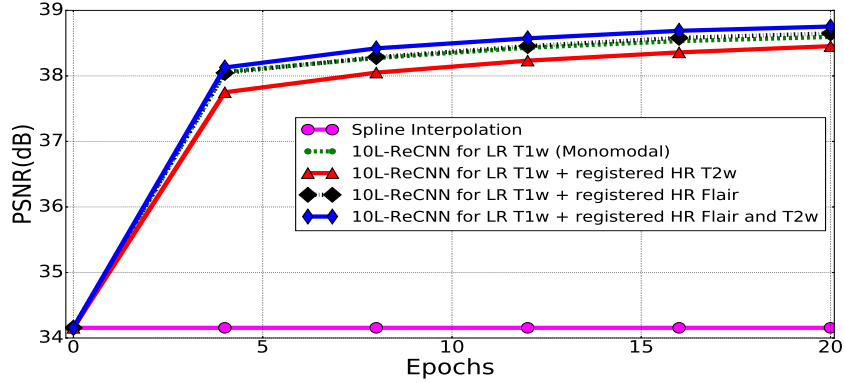
Figure 9: 3D deep neural network for multimodal brain MRI super-resolution using intermodality priors. Skip connection computes the residual between ILR image and HR image.

414 In addition, we explore the impact of the network depth augmentation with regard to  
 415 the performance of multimodal SR approach. The experiments shown in Figure 12 indicate  
 416 that the deeper structures do not lead to better results within the multimodal method.

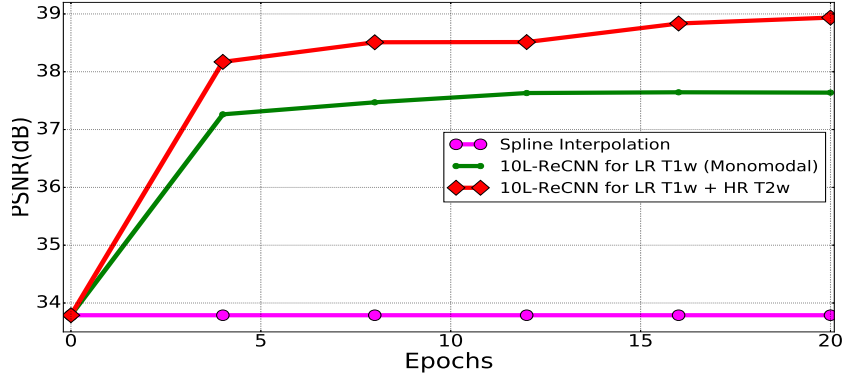
## 417 6. How Transferable are learnt Features?

418 Training a CNN from scratch requires a substantial amount of training data and may  
 419 take a long time. Moreover, to avoid overfitting, the training dataset has to reflect the  
 420 appearance variability of the images to reconstruct. In the context of brain MRI, part of  
 421 image variability comes from acquisition systems. Hence, we investigate the impact of such  
 422 image variability onto SR performance by evaluating transfer learning skills among different  
 423 datasets corresponding to the same imaging modality.

424 In order to characterize such generalization skills, we evaluate in which extent the selec-  
 425 tion of a given training dataset influences the reconstruction performance of the network.  
 426 To this end, we train from scratch two 20L-ReCNN networks separately for a 10-image  
 427 NAMIC T1-weighted dataset and a 10-image Kirby T1-weighted dataset, and we test the  
 428 trained models for the remaining 10-image NAMIC and Kirby T1-weighted datasets. The  
 429 considered case-study involves a scaling factor of  $(2 \times 2 \times 2)$ . For quantitative compari-  
 430 son, the PSNR and the structural similarity (SSIM) (the definition of SSIM can be found  
 431 in Wang et al. 2004) are used to evaluate the performance of each model in Table 2. For  
 432 benchmarking purposes, we also include a comparison with the following methods: cubic  
 433 spline interpolation, low-rank total variation (LRTV) (Shi et al. 2015), SRCNN3D (Pham  
 434 et al. 2017). The use of 20-layer CNN-based approaches for each training dataset can lead  
 435 to improvements over spline interpolation, LRTV method and SRCNN3D (with respect to  
 436 both PSNR and SSIM). Although, the gain is slightly lowered (e.g. PSNR: 0.55dB for test-  
 437 ing Kirby and 0.74dB for NAMIC, SSIM: 0.003 for Kirby and 0.0019 for NAMIC) when



(a) Multimodal experiments using Kirby 21 dataset for training and testing.



(b) Multimodal experiments using NAMIC dataset for training and testing.

Figure 10: Multimodality-guided SR experiments. The LR T1-weighted images are upscaled with isotropic scale factor  $\times 2$  using respectively monomodal network (10L-ReCNN for LR T1w), HR T2w multimodal network, HR Flair multimodal network and both HR Flair and T2w multimodal images.

438 using different training and testing dataset (i.e. different resolution), our proposed networks  
 439 obtain better results than compared methods.

440 For qualitative comparison, Figures 13 and 14 show the results of reconstructed 3D  
 441 images obtained from all the compared techniques. The zoom version of the reconstructions  
 442 20L-ReCNN shows sharpen edges and a grayscale intensity which are closest to the ground  
 443 truth. In addition, the HR reconstruction of the 20L-ReCNN model shows that its differences  
 444 from the ground truth are less than other methods (i.e. the contours of the residual image  
 445 of the 20L-ReCNN method are less visible than those of others). Hence, we can infer that  
 446 our proposed method better preserves contours, geometrical structures and better recovers  
 447 the image contrast compared with the other methods.

## 448 7. Application of Super-Resolution on Clinical Neonatal Data

449 In clinical routine, anisotropic LR MRI data are usually acquired in order to limit the  
 450 acquisition time due to patient comfort such as infant brain MRI scans (Makropoulos et al.,  
 451 2018), or in case of rapid emergency scans (Walter et al., 2003). These images usually have  
 452 a high in-plane resolution and a low through-plane resolution. Anisotropic data acquisition

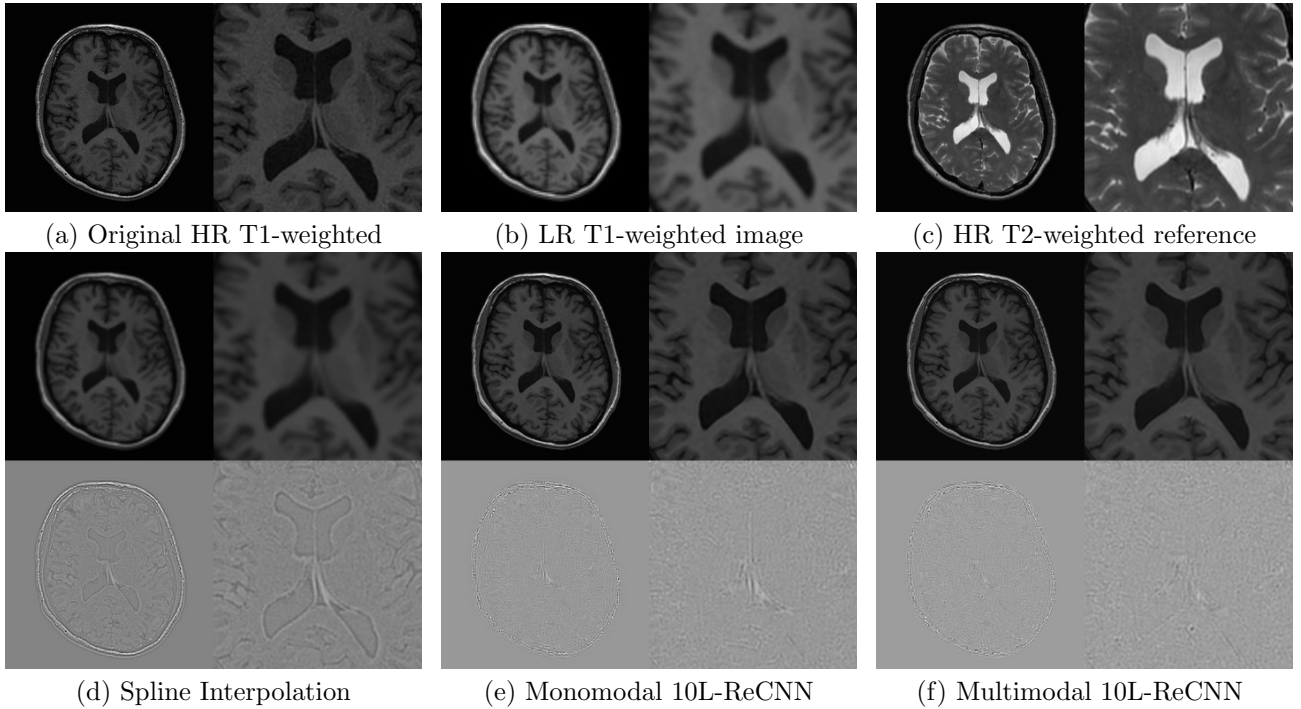


Figure 11: Illustration of the axial slices of monomodal and multimodal SR results (subject 01018, pathological case of testing set) with isotropic voxel upsampling using NAMIC. The LR T1-weighted image (b) with voxel size  $2 \times 2 \times 2\text{mm}^3$  is upsampled to size  $1 \times 1 \times 1\text{mm}^3$ . The monomodal network 10L-ReCNN is applied directly the LR T1-weighted image (b), whereas the multimodal network 10L-ReCNN uses the HR T2-weighted reference (c) to upscale LR T1-weighted image (b). The results of the monomodal and multimodal networks are shown in (e) and (f), respectively. The different between ground truth image and reconstruction results are at the bottom. Their zoom version are at the right.

453 severely limits the 3D exploitation of MRI data. Interpolation is commonly used to upsam-  
 454 pled these LR images to isotropic resolution. However, interpolated LR images may lead to  
 455 partial volume artifacts that may affect segmentation (Ballester et al., 2002). In this section,  
 456 we aim to use our single image SR method to enhance the resolution of such clinical data.

457 The idea is to apply our proposed convolutional neural networks-based SR method to  
 458 transfer the rich information available from high-resolution experimental dataset to lower-  
 459 quality image data. The procedure first uses CNNs to learn mappings between real HR  
 460 images and their corresponding simulated LR images with the same resolution of real data.  
 461 The LR data is generated by the observation model, which is decomposed into a linear  
 462 downsampling operator after a space-invariant blurring model as a Gaussian kernel with  
 463 the full-width-at-half-maximum (FWHM) equal to slice thickness (Greenspan, 2008). Once  
 464 models are learnt, these mappings enhance the LR resolution of unseen low quality images.

465 In order to verify the applicability of our CNN-based methods, we have used two neonatal  
 466 brain MRI dataset: the Developing Human Connectome Project (dHCP) (Hughes et al.,  
 467 2017) and clinical neonatal MRI data acquired in the neonatology service of Reims Hos-  
 468 pital (Multiphysics image-based Analysis for premature brAin development understanding

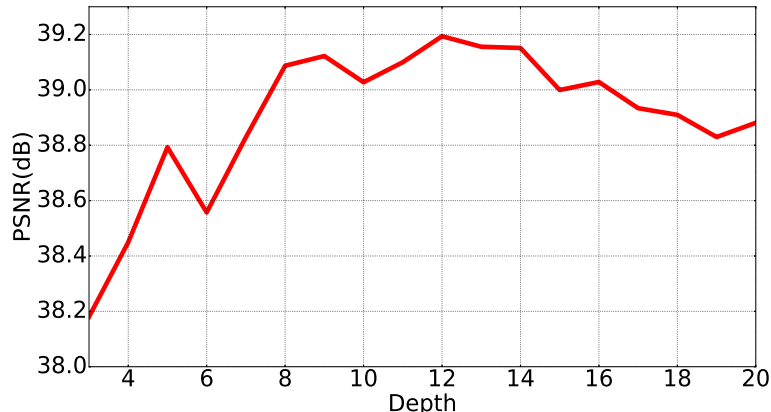


Figure 12: Depth vs. performance (multimodal SR using residual-learning networks with the same filter numbers  $n = 64$  and filter size  $f = 3$  over 20 training epochs using Adam optimization and tested with isotropic scale factor  $\times 2$  using NAMIC for training and testing).

Testing dataset	Spline Interpolation		LRTV		SRCNN3D		20L-ReCNN			
	PSNR	SSIM	PSNR	SSIM	Kirby (10 images)		Kirby (10 images)		NAMIC(10 images)	
					PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Kirby (5 images )	34.16	0.9402	35.08	0.9585	37.51	0.9735	38.93	0.9797	38.06	0.9767
Standard deviation	1.90	0.0111	2.09	0.0083	1.97	0.0053	1.87	0.0044	1.83	0.0045
Gain	-	-	0.92	0.0183	3.36	0.0333	4.77	0.0395	3.9	0.0365
NAMIC (10 images)	33.78	0.9388	34.34	0.9549	36.72	0.9694	37.73	0.9762	38.28	0.9781
Standard deviation	1.82	0.0071	1.79	0.0044	1.76	0.0035	1.81	0.0031	1.78	0.0029
Gain	-	-	0.56	0.0161	2.94	0.0306	3.95	0.0374	4.5	0.0393

Table 2: The results of PSNR/SSIM for isotropic scale factor  $\times 2$  with the gain between compared methods and the method of spline interpolation. One network 20L-ReCNN trained with 10 images of Kirby and one trained with NAMIC

469 - MAIA dataset). The HR images are T2-weighted MRIs of the dHCP and provided by  
 470 the Evelina Neonatal Imaging Centre, London, UK. Forty neonatal data were acquired on  
 471 a 3T Achieva scanner with repetition (TR) of 12 000 ms and echo times (TE) of 156ms,  
 472 respectively. The size of voxels is  $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ . In-vivo neonatal LR images has a  
 473 voxel size of about  $0.4464 \times 0.4464 \times 3 \text{ mm}^3$ .

474 Figure 15 compares the qualitative results of HR reconstructions (spline interpolation,  
 475 NMU (Manjón et al., 2010b) and our proposed networks) of a LR image from MAIA dataset.  
 476 In this experiment, we do not have the ground truth of real LR data for calculating quanti-  
 477 tative metrics. The comparison reveals that the CNNs-based methods recover shaper images  
 478 and better defined boundaries than spline interpolation. For instance, the cerebrospinal fluid  
 479 (CSF) of the cerebellum of the proposed method, in Figure 15, is more visible than those  
 480 obtained with the compared methods. Our proposed technique reconstructs more curved  
 481 cortex and more accurate ventricles. These results tend to confirm qualitatively the efficacy  
 482 of our approach.



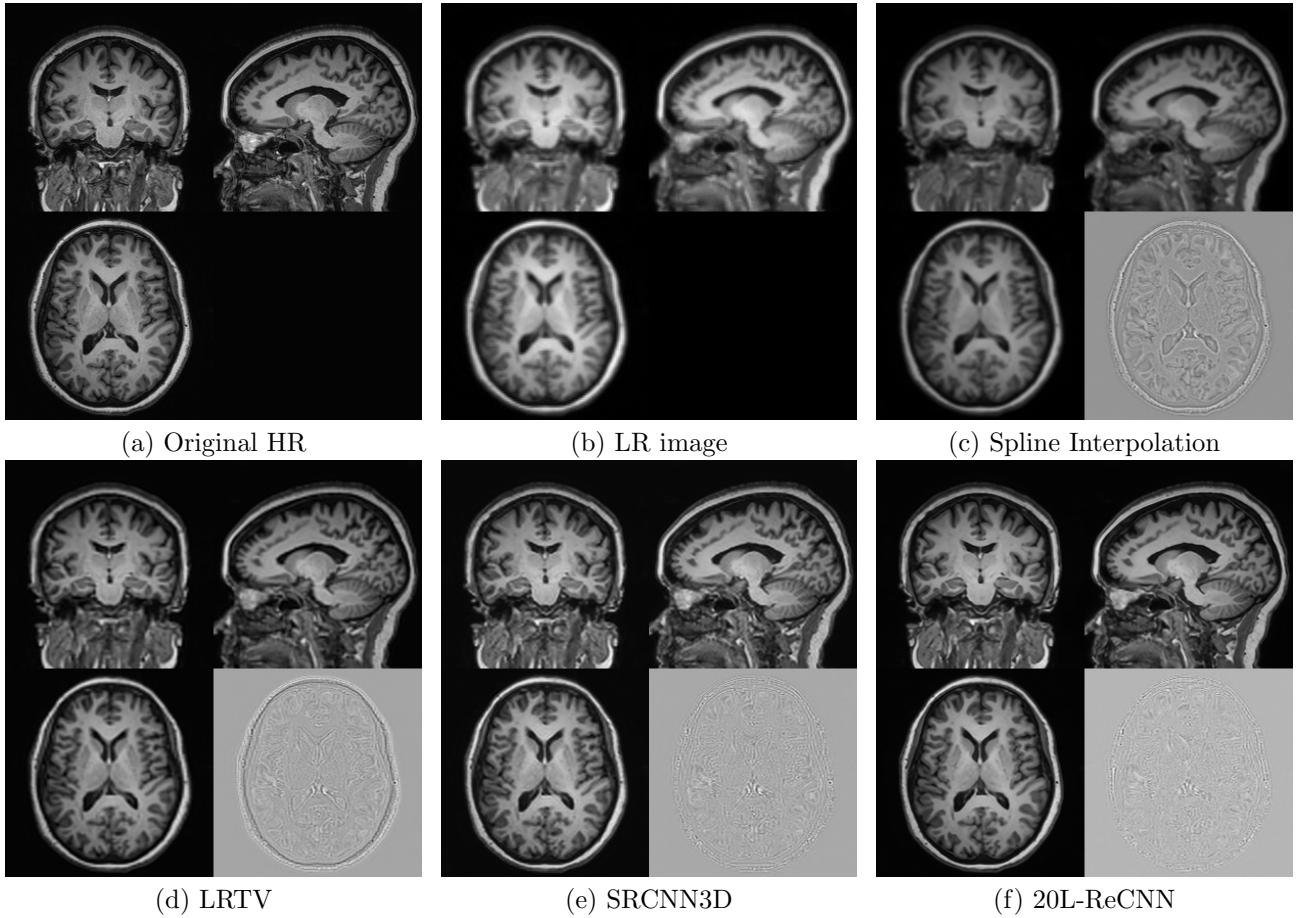


Figure 13: Illustration of SR results (subject KKI2009-02-MPRAGE, non-pathological case, in testing set of dataset Kirby 21) with isotropic voxel upsampling. LR data (b) with voxel size  $2 \times 2 \times 2.4\text{mm}^3$  is upsampled to size  $1 \times 1 \times 1.2\text{mm}^3$ . The difference between the ground truth image and the reconstruction results are in the right bottom corners. Both network SRCNN3D and network 20L-ReCNN are trained with the 10 last images of Kirby.

## 483 8. Discussion

484 This study investigates CNN-based models for 3D brain MR image SR. Based on a  
 485 comprehensive experimental evaluation, we would like to draw the following conclusions  
 486 and recommendations regarding the setup to be considered. We highlight that eight com-  
 487plementary factors may drive the reconstruction performance of CNN-based models. The  
 488 combination of 1) appropriate optimization with 2) weight initialization and 3) residual  
 489 learning is a key to exploit deeper networks with a faster and effective convergence. The  
 490 choice of an appropriate optimization method can lead to a PSNR improvement of (at least)  
 491 1dB. In this study, it has appeared that Adam method (Kingma and Ba 2015) provides  
 492 significantly better reconstruction results than other classical techniques such as SGD, and a  
 493 faster convergence. Moreover, weights initialization is a very important step. Indeed, some  
 494 approaches simply do not achieve convergence in the learning phase. This study has also

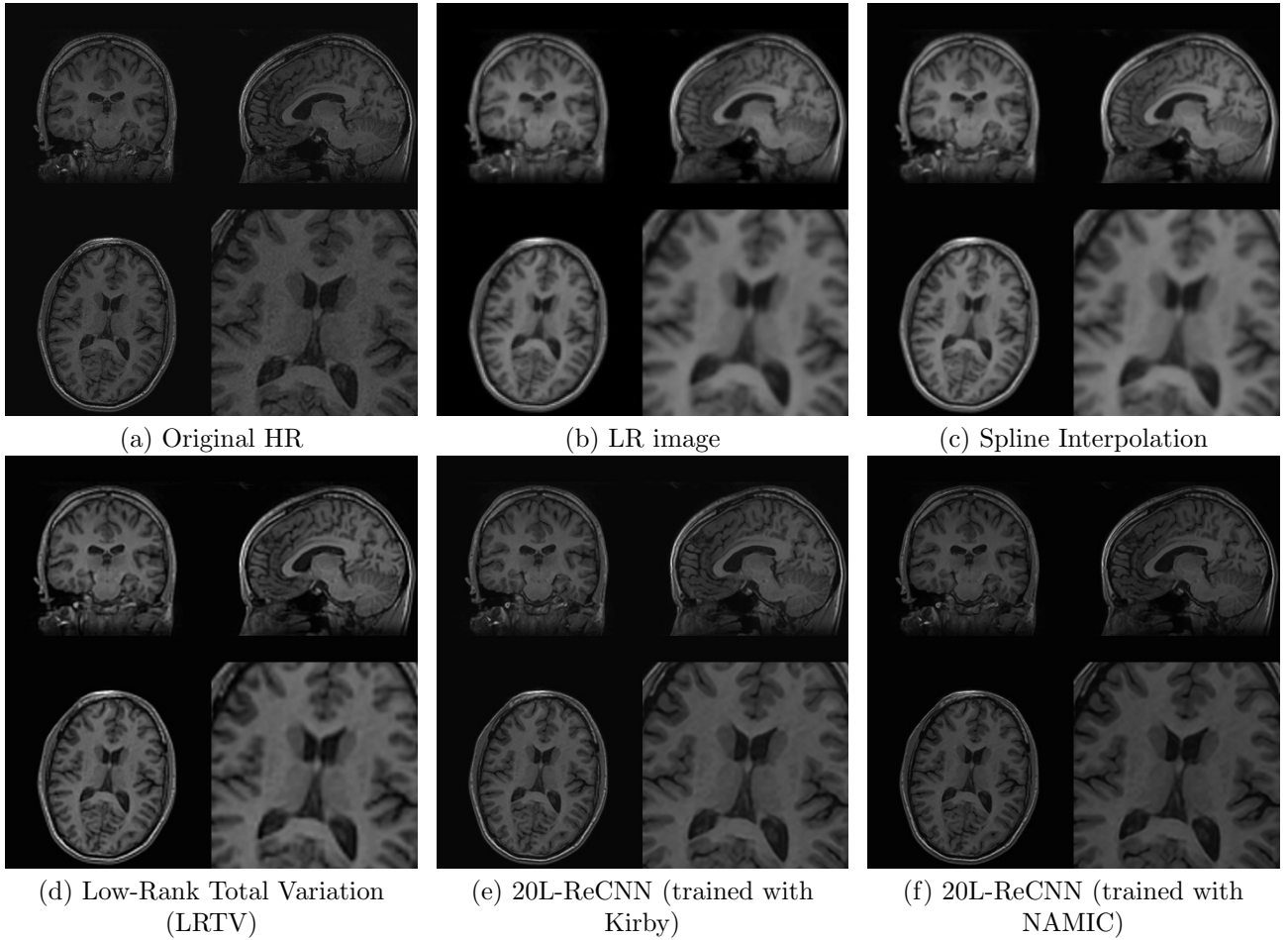


Figure 14: Illustration of SR results (subject 01011-t1w, pathological case, in testing set of dataset NAMIC) with isotropic voxel upsampling. LR data (b) with voxel size  $2 \times 2 \times 2\text{mm}^3$  is upsampled to size  $1 \times 1 \times 1\text{mm}^3$ . The zoom versions of the axial slices are in the right bottom corners.

495 shown that residual modeling for single image SR is a straightforward technique to improve  
 496 the reconstruction performances (+0.4dB) without requiring major changes in the network  
 497 architecture. Appropriate weight initialization methods described by [He et al. \(2015\)](#); [Glorot and Bengio \(2010\)](#)  
 498 allow us to build deeper residual-learning networks. From our point  
 499 of view, these three aspects of SR algorithm are the first to require special attention for the  
 500 implementation of a SR technique based on CNN.

501 Overall, we show that better performance can be achieved by learning 4) a deeper fully 3D  
 502 convolution neural network, 5) exploring more filters and 6) increasing filter size. In addition,  
 503 using 7) larger training patch size and 8) augmentation of training subject leads to increase  
 504 the performance of the networks. The adjustment of these 5 elements provides a similar  
 505 improvement (about 0.5dB). Although it seems natural to implement the deepest possible  
 506 network, this parameter is not always the key to obtaining a better estimate of a high-  
 507 resolution image. Our study shows that, depending on the type of input data (monomodal  
 508 or multimodal), network depth is not necessarily the main parameter leading to better



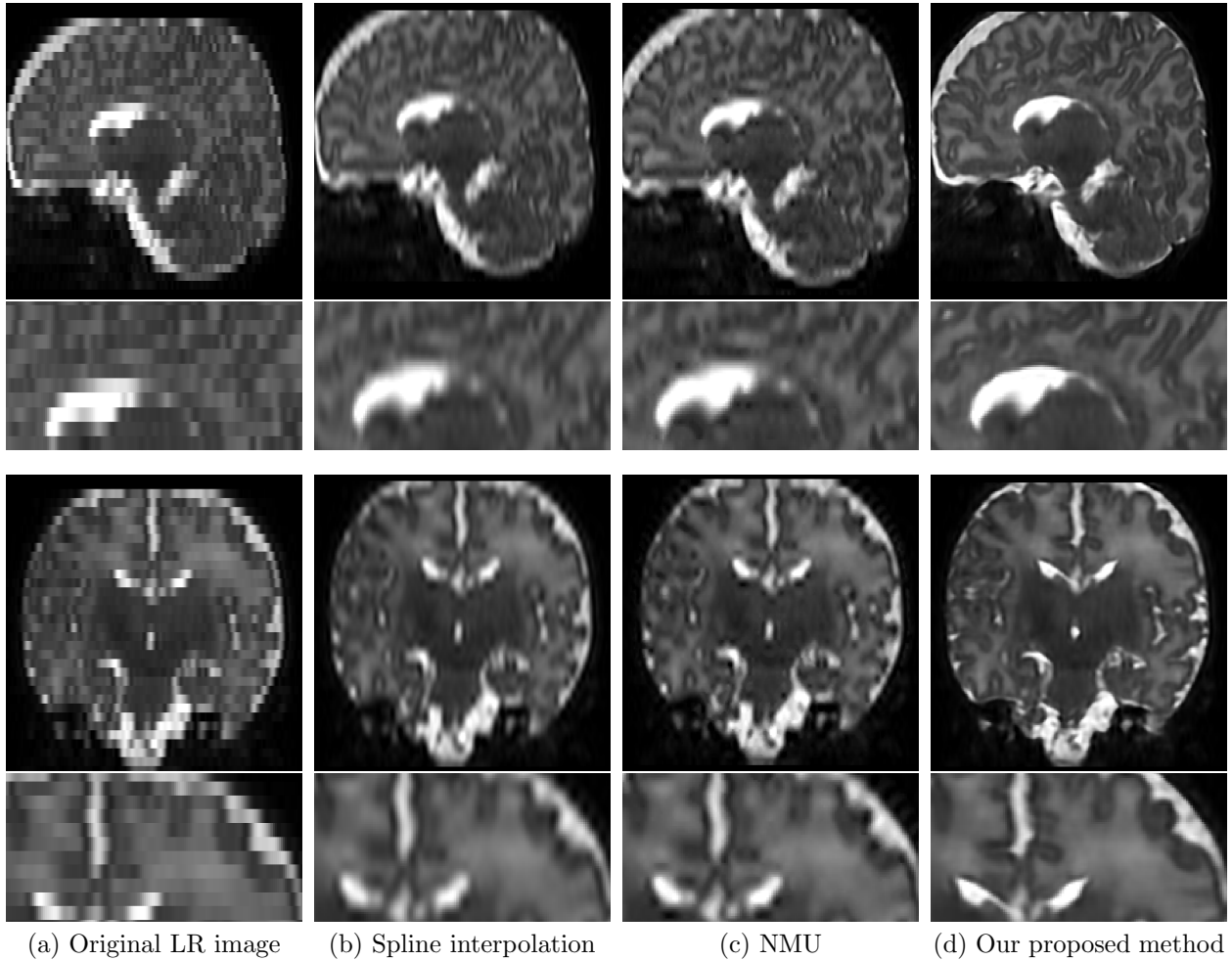


Figure 15: Illustration of SR results on a clinical data of MAIA dataset with isotropic voxel upsampling. Original data with voxel size of about  $0.4464 \times 0.4464 \times 3$  is resampled to size  $0.5 \times 0.5 \times 0.5$  mm<sup>3</sup>. Networks are trained with the dHCP dataset. The first, second, third and last rows presents the sagittal slices, the zoom versions of the sagittal slices, the coronal slices and the zoom versions of the coronal slices, respectively.

509 image reconstruction. In addition, it is important to take into account the time of the  
 510 learning phase as well as the maximum memory available in the GPU in order to choose the  
 511 best architecture of the network. For instance, for the monomodal SR case based on the  
 512 simulations of Kirby dataset, we suggest using 20-layer networks with 64 small filters with  
 513 size of  $3^3$  regarding 10 training subjects of size  $25^3$  to achieve practicable results.

514 In CNN-based approaches, the upscaling operation can be performed by using transposed  
 515 convolution (so-called fractionally strided convolutional) layers as proposed by [Oktay et al.](#)  
 516 [\(2016\)](#); [Dong et al. \(2016b\)](#) or sub-pixel layers ([Shi et al. 2016](#)). However, the trained weights  
 517 of these networks are totally applied for a specified scale factor. This is a limiting aspect of  
 518 CNN-based SR for MR data since a fixed upscaling factor is not appropriate in this context.  
 519 In this study, we have presented a multiscale CNN-based SR method for single 3D brain  
 520 MRI that is capable of learning multiple scales by training multiple isotropic scale factors

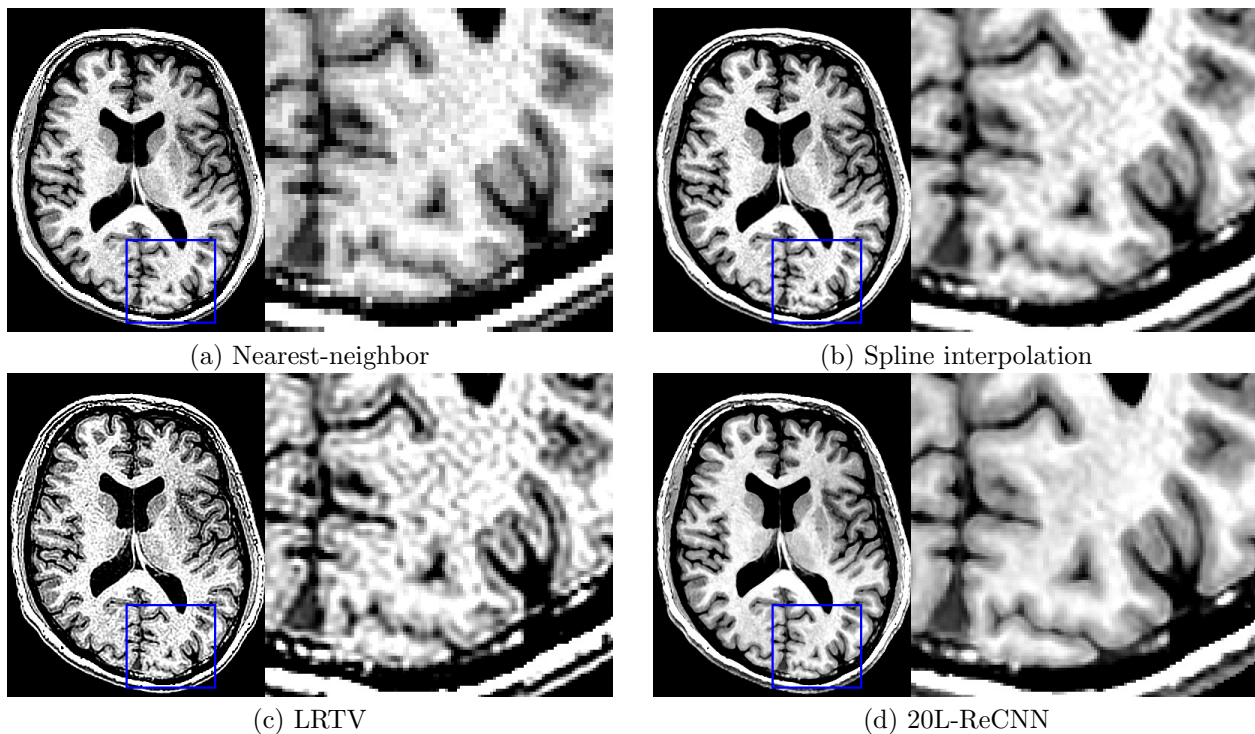


Figure 16: Illustration of SR results (subject 01018-t1w in testing set of dataset NAMIC) with isotropic voxel upsampling. Original data with voxel size of  $1 \times 1 \times 1 \text{ mm}^3$  is upsampled to size  $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ . 20L-ReCNN is trained with the NAMIC dataset.

521 due to an independent upsampling technique such as spline interpolation. Handling multiple  
 522 scales is related to multi-task learning. The lack of flexibility of learnt network architectures  
 523 raises an open issue motivating further studies: how can we build a network that can deal  
 524 with a set of observation models (i.e. multiple scales, arbitrary point spread functions, non  
 525 uniform sampling, etc.)? For instance, when applying SR techniques in a realistic setting,  
 526 the choice of the PSF is indeed a key element for SR methods and it depends on the type  
 527 of MRI sequence. More specifically, the shape of the PSF depends on the trajectory in the  
 528 k-space (Cartesian, radial, spiral). Making the network independent from the PSF model  
 529 (i.e. doing blind SR) would be a major step for its use in routine protocol. Further research  
 530 directions could focus on making more flexible CNN-based SR methods for greater use of  
 531 these techniques in human brain mapping studies.

532 Evaluation of SR techniques is carried out on simulated LR images. However, one po-  
 533 tential use of SR techniques would be to improve the resolution of isotropic data acquired in  
 534 clinical routine. Figure 16 shows upsampling results on isotropic T1-weighted MR images  
 535 (the resolution was increased from  $1 \times 1 \times 1 \text{ mm}^3$  to  $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ ). In this experi-  
 536 ment, the applied network has been trained to increase image resolution from  $2 \times 2 \times 2 \text{ mm}^3$   
 537 to  $1 \times 1 \times 1 \text{ mm}^3$ . Although quantitative results cannot be computed, visual inspection  
 538 of reconstructed upsampled images tend to show the potential of this SR method. Thus,  
 539 features learnt at a lower scale ( $2 \text{ mm}$  in this experiment) may be used to compute very

540 high-resolution images that could be involved in fine studies of thin brain structures such  
541 as the cortex. Further work is required to investigate this aspect and more particularly the  
542 link with self-similarity based approaches.

543 In this article, we have proposed a multimodal method for brain MRI SR using CNNs  
544 where a HR reference image of the same subject can drive the reconstruction process of the  
545 LR image. By concatenating these HR and LR images, the reconstruction of the LR one  
546 can be enhanced by exploiting the multimodality feature of MR data. As shown in previous  
547 works (Rousseau et al. 2010a; Manjón et al. 2010a), the use of HR reference can lead to  
548 significant improvements of the reconstruction process. However, unlike the monomodal  
549 setup, a deeper network does not lead to better performance within the experiments on  
550 NAMIC dataset. Experiments from our study show that future work is needed to understand  
551 the relationship between network depth and the quality of HR image estimation.

552 Moreover, we have experimentally investigated the performances of CNN for generalizing  
553 on a different dataset (i.e. how a learnt network can be used in another context). More  
554 specifically, our study illustrates how knowledge learnt from one MR dataset is transferred  
555 to another one (different acquisition protocol and different scales). We have used Kirby  
556 and NAMIC datasets for this purpose. Although a slight decrease in performance can be  
557 observed, CNN-based approach can still achieve better performances than existing methods.  
558 These results tend to demonstrate the potential applications of CNN-based techniques for  
559 MRI SR. Further investigations are required to fully assess the possibilities of transfer learn-  
560 ing in medical imaging context, and the contributions of fine-tuning technique (Tajbakhsh  
561 et al. 2016).

562 Finally, future research directions for CNN-based SR techniques could focus on other  
563 elements of the network architecture or the learning procedure. For instance, batch normal-  
564 ization (BN) step has been proposed by Ioffe and Szegedy (2015). The purpose of a BN layer  
565 is to normalize the data through the entire network, rather than just performing normaliza-  
566 tion once in the beginning. Although BN has been shown to improve classification accuracy  
567 and decrease training time (Ioffe and Szegedy 2015), we attempt to include BN layers into  
568 CNN for image SR but they do not lead to performance increase. Similar observations have  
569 been made in a recent SR challenge (Timofte et al. 2017). Moreover, whereas the classical  
570 MSE-based loss attempts to recover the smooth component, perceptual losses (Ledig et al.  
571 2017; Johnson et al. 2016; Zhao et al. 2017) are proposed for natural image SR to better  
572 reconstruct fine details and edges. Thus, adding this type of layer (BN or residual block) or  
573 defining new loss functions may be beneficial for MRI SR and may provide new directions  
574 for research.

575 In this study, we have investigated the impact of adding data (about 3 200 patches  
576 per added subject of Kirby dataset) on SR performances through PSNR computation. It  
577 appeared that using more subjects slightly improves the reconstruction results in this exper-  
578 imental setting. However, further work could focus on SR-specific data augmentation by  
579 rotation and flipping, which is usually used in many works (Kim et al. 2016a; Timofte et al.  
580 2016), for improving algorithm generalization.

581 The evaluation on synthetic low-resolution images remains a limited approach to accu-  
582 rately quantify the performance of developed algorithms in real-world situations. However,

583 acquiring low-resolution (LR) and high-resolution (HR) MRI data with the exact same con-  
584 trast is very challenging. The development of such a database is beyond the scope of this  
585 work. There is currently no available dataset of pairs of real HR and LR images. This is  
586 the reason why all the works on SR perform quantitative analysis on synthetic data.

587 In order to demonstrate the potential of SR methods for enhancing the quality of clinical  
588 LR images, we have presented a practical application: image quality transfer from high-  
589 resolution experimental dataset to real anisotropic low-resolution images. Our CNN-based  
590 SR method shows clear improvements over interpolation, which is the standard technique  
591 to enhance image quality from visualisation by a radiologist. SR method is therefore a  
592 highly relevant alternative to interpolation. Future work on the support of SR techniques  
593 for applications in brain segmentation could be investigated to evaluate the performance of  
594 these methods. Besides, the best way to evaluate the performances of SR methods is to  
595 gather datasets where pairs of real HR and LR images are available.

## 596 9. Acknowledgement

597 The research leading to these results has received funding from the ANR MAIA project,  
598 grant ANR-15-CE23-0009 of the French National Research Agency, INSERM, Institut Mines  
599 Télécom Atlantique (Chaire “Imagerie médicale en thérapie interventionnelle”) and the  
600 American Memorial Hospital Foundation. We also gratefully acknowledge the support of  
601 NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

## 602 References

- 603 Ballester, M. A. G., Zisserman, A. P., Brady, M., 2002. Estimation of the partial volume effect in MRI.  
604 *Medical Image Analysis* 6 (4), 389–405.
- 605 Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult.  
606 *IEEE Transactions on Neural Networks* 5 (2), 157–166.
- 607 Chen, Y., Xie, Y., Zhou, Z., Shi, F., Christodoulou, A. G., Li, D., 2018. Brain MRI super resolution using  
608 3D deep densely connected neural networks. In: 2018 IEEE 15th International Symposium on Biomedical  
609 Imaging (ISBI 2018). IEEE, pp. 739–742.
- 610 Dong, C., Loy, C. C., He, K., Tang, X., 2016a. Image super-resolution using deep convolutional networks.  
611 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2), 295–307.
- 612 Dong, C., Loy, C. C., Tang, X., 2016b. Accelerating the super-resolution convolutional neural network. In:  
613 *European Conference on Computer Vision*. Springer, pp. 391–407.
- 614 Fogtman, M., Seshamani, S., Kroenke, C., Cheng, X., Chapman, T., Wilm, J., Rousseau, F., Studholme,  
615 C., 2014. A unified approach to diffusion direction sensitive slice registration and 3D DTI reconstruction  
616 from moving fetal brain anatomy. *IEEE Transactions on Medical Imaging* 33 (2), 272–289.
- 617 Gholipour, A., Estroff, J. A., Warfield, S. K., 2010. Robust super-resolution volume reconstruction from slice  
618 acquisitions: application to fetal brain MRI. *IEEE Transactions on Medical Imaging* 29 (10), 1739–1758.
- 619 Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks.  
620 In: *Proceedings of the thirteenth International Conference on Artificial Intelligence and Statistics*. pp.  
621 249–256.
- 622 Greenspan, H., 2008. Super-resolution in medical imaging. *The Computer Journal* 52 (1), 43–63.
- 623 He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance  
624 on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision*.  
625 pp. 1026–1034.

- 626 He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of  
627 the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.
- 628 Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks.  
629 In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708.
- 630 Hughes, E., Grande, L. C., Murgasova, M., Hutter, J., Price, A., Gomes, A. S., Allsop, J., Steinweg,  
631 J., Tusor, N., Wurie, J., et al., 2017. The developing human connectome: announcing the first release of  
632 open access neonatal brain imaging. 23rd Annual Meeting of the Organization for Human Brain Mapping,  
633 25–29.
- 634 Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal  
635 covariate shift. In: International Conference on Machine Learning. pp. 448–456.
- 636 Jain, S., Sima, D. M., Nezhad, F. S., Hangel, G., Bogner, W., Williams, S., Van Huffel, S., Maes, F.,  
637 Smeets, D., 2017. Patch-based super-resolution of MR spectroscopic images: Application to multiple  
638 sclerosis. *Frontiers in neuroscience* 11.
- 639 Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., Smith, S. M., 2012. *Fsl. Neuroimage*  
640 62 (2), 782–790.
- 641 Jia, Y., Gholipour, A., He, Z., Warfield, S. K., 2017. A new sparse representation framework for recon-  
642 struction of an isotropic high spatial resolution MR volume from orthogonal anisotropic resolution scans.  
643 *IEEE Transactions on Medical Imaging* 36 (5), 1182–1193.
- 644 Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.,  
645 2014. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM  
646 International Conference on Multimedia. ACM, pp. 675–678.
- 647 Jog, A., Carass, A., Prince, J. L., 2016. Self super-resolution for magnetic resonance images. In: International  
648 Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 553–560.
- 649 Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution.  
650 In: European Conference on Computer Vision. Springer, pp. 694–711.
- 651 Kainz, B., Steinberger, M., Wein, W., Kuklisova-Murgasova, M., Malamateniou, C., Keraudren, K., Torsney-  
652 Weir, T., Rutherford, M., Aljabar, P., Hajnal, J. V., et al., 2015. Fast volume reconstruction from motion  
653 corrupted stacks of 2D slices. *IEEE Transactions on Medical Imaging* 34 (9), 1901–1913.
- 654 Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D.,  
655 Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion  
656 segmentation. *Medical Image Analysis* 36, 61–78.
- 657 Kim, J., Lee, J. K., Lee, K. M., 2016a. Accurate image super-resolution using very deep convolutional  
658 networks. In: in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.
- 659 Kim, J., Lee, J. K., Lee, K. M., 2016b. Deeply-recursive convolutional network for image super-resolution.  
660 In: in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.
- 661 Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on  
662 Learning Representations.
- 663 Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural  
664 networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.
- 665 Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A., Bogovic, J. A., Hua,  
666 J., Chen, M., Jarso, S., et al., 2011. Multi-parametric neuroimaging reproducibility: a 3T resource study.  
667 *Neuroimage* 54 (4), 2854–2866.
- 668 LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recogni-  
669 tion. *Proceedings of the IEEE* 86 (11), 2278–2324.
- 670 Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz,  
671 J., Wang, Z., et al., 2017. Photo-realistic single image super-resolution using a generative adversarial  
672 network. In: in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.
- 673 Lim, B., Son, S., Kim, H., Nah, S., Lee, K. M., 2017. Enhanced deep residual networks for single image super-  
674 resolution. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.  
675 Vol. 1. p. 3.
- 676 Makropoulos, A., Robinson, E. C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S. J., Steinweg,

677 J., Vecchiato, K., Passerat-Palmbach, J., et al., 2018. The developing human connectome project: A  
678 minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage* 173, 88–112.

679 Manjón, J. V., Coupé, P., Buades, A., Collins, D. L., Robles, M., 2010a. MRI super-resolution using self-  
680 similarity and image priors. *Journal of Biomedical Imaging* 2010, 17.

681 Manjón, J. V., Coupé, P., Buades, A., Fonov, V., Collins, D. L., Robles, M., 2010b. Non-local MRI upsam-  
682 pling. *Medical Image Analysis* 14 (6), 784–792.

683 Milanfar, P., 2010. *Super-Resolution Imaging*. CRC press.

684 Nesterov, Y., 1983. A method of solving a convex programming problem with convergence rate  $o(1/k^2)$ . In:  
685 *Soviet Mathematics Doklady*. Vol. 27. pp. 372–376.

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

690 Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. *ICML*  
691 (3) 28, 1310–1318.

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

695 Pham, C.-H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., Rousseau, F., 2019. Simulta-  
696 neous super-resolution and segmentation using a generative adversarial network: Application to neonatal  
697 brain MRI. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE.

698 Poot, D. H., Jeurissen, B., Bastiaansen, Y., Veraart, J., Van Hecke, W., Parizel, P. M., Sijbers, J., 2013.  
699 Super-resolution for multislice diffusion tensor imaging. *Magnetic Resonance in Medicine* 69 (1), 103–113.

700 Ramos-Llordén, G., Arnold, J., Van Steenkiste, G., Jeurissen, B., Vanhevel, F., Van Audekerke, J., Ver-  
701 hoye, M., Sijbers, J., 2017. A unified maximum likelihood framework for simultaneous motion and  $t_{1}$   
702 estimation in quantitative mr  $t_{1}$  mapping. *IEEE Transactions on Medical Imaging* 36 (2), 433–446.

703 Rousseau, F., Initiative, A. D. N., et al., 2010a. A non-local approach for image super-resolution using  
704 intermodality priors. *Medical Image Analysis* 14 (4), 594–605.

705 Rousseau, F., Kim, K., Studholme, C., Koob, M., Dietemann, J.-L., 2010b. On super-resolution for fetal  
706 brain MRI. In: *International Conference on Medical Image Computing and Computer-Assisted Interven-*  
707 *tion*. Springer, pp. 355–362.

708 Rueda, A., Malpica, N., Romero, E., 2013. Single-image super-resolution of brain MR images using over-  
709 complete dictionaries. *Medical Image Analysis* 17 (1), 113–132.

710 Scherrer, B., Gholipour, A., Warfield, S. K., 2012. Super-resolution reconstruction to increase the spatial  
711 resolution of diffusion weighted images from orthogonal anisotropic acquisitions. *Medical Image Analysis*  
712 16 (7), 1465–1476.

713 Shi, F., Cheng, J., Wang, L., Yap, P., Shen, D., 2015. LRTV: MR image super-resolution with low-rank and  
714 total variation regularizations. *IEEE Transactions on Medical Imaging* 34 (12), 2459–2466.

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

718 Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. In:  
719 *International Conference on Learning Representations*.

720 Steenkiste, G., Jeurissen, B., Veraart, J., Den Dekker, A. J., Parizel, P. M., Poot, D. H., Sijbers, J., 2016.  
721 Super-resolution reconstruction of diffusion parameters from diffusion-weighted images with different slice  
722 orientations. *Magnetic Resonance in Medicine* 75 (1), 181–195.

723 Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., Liang, J., 2016.  
724 Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Transactions*  
725 *on Medical Imaging* 35 (5), 1299–1312.

726 Tieleman, T., Hinton, G., 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent  
727 magnitude. *COURSERA: Neural networks for machine learning* 4 (2), 26–31.

- 728 Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., Zhang, L., et al., 2017. NTIRE 2017 challenge on  
729 single image super-resolution: Methods and results. In: The IEEE Conference on Computer Vision and  
730 Pattern Recognition (CVPR) Workshops.
- 731 Timofte, R., De Smet, V., Van Gool, L., 2013. Anchored neighborhood regression for fast example-based  
732 super-resolution. In: Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, pp. 1920–  
733 1927.
- 734 Timofte, R., De Smet, V., Van Gool, L., 2014. A+: Adjusted anchored neighborhood regression for fast  
735 super-resolution. In: Computer Vision–ACCV 2014. Springer, pp. 111–126.
- 736 Timofte, R., Rothe, R., Van Gool, L., 2016. Seven ways to improve example-based single image super  
737 resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp.  
738 1865–1873.
- 739 Tor-Díez, C., Pham, C.-H., Meunier, H., Faisan, S., Bloch, I., Bednarek, N., Passat, N., Rousseau, F.,  
740 2019. Evaluation of cortical segmentation pipelines on clinical neonatal MRI data. In: 41st International  
741 Engineering in Medicine and Biology Conference (EMBC 2019).
- 742 Van Steenkiste, G., Poot, D. H., Jeurissen, B., Den Dekker, A. J., Vanhevel, F., Parizel, P. M., Sijbers,  
743 J., 2017. Super-resolution T1 estimation: Quantitative high resolution T1 mapping from a set of low  
744 resolution T1-weighted images with different slice orientations. *Magnetic Resonance in Medicine* 77 (5),  
745 1818–1830.
- 746 Walter, C., Kruessell, M., Gindele, A., Brochhagen, H., Gossmann, A., Landwehr, P., 2003. Imaging of renal  
747 lesions: evaluation of fast MRI and helical CT. *The British Journal of Radiology* 76 (910), 696–703.
- 748 Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility  
749 to structural similarity. *IEEE Transactions on Image Processing* 13 (4), 600–612.
- 750 Zeyde, R., Elad, M., Protter, M., 2012. On single image scale-up using sparse-representations. In: *Curves  
751 and Surfaces*. Springer, pp. 711–730.
- 752 Zhao, C., Carass, A., Dewey, B. E., Prince, J. L., 2018. Self super-resolution for magnetic resonance images  
753 using deep networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).  
754 IEEE, pp. 365–368.
- 755 Zhao, H., Gallo, O., Frosio, I., Kautz, J., 2017. Loss functions for neural networks for image processing.  
756 *IEEE Transactions on Computational Imaging* 2017.