Simultaneous super-resolution and segmentation using a generative adversarial network: Application to neonatal brain MRI

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

SIMULTANEOUS SUPER-RESOLUTION AND SEGMENTATION USING A GENERATIVE ADVERSARIAL NETWORK: APPLICATION TO NEONATAL BRAIN MRI

C.-H. Pham[1], C. Tor-Díez[1], H. Meunier[2], N. Bednarek[2,3], R. Fablet[4], N. Passat[3], F. Rousseau[1]

[1] IMT Atlantique, LaTIM U1101 INSERM, UBL, Brest, France
[2] Service de médecine néonatale et réanimation pédiatrique, CHU de Reims, France
[3] Université de Reims Champagne-Ardenne, CReSTIC, Reims, France
[4] IMT Atlantique, Lab-STICC UMR CNRS 6285, Brest, France

ABSTRACT

The analysis of clinical neonatal brain MRI remains challenging due to low anisotropic resolution of the data. In most pipelines, images are first re-sampled using interpolation or single image super-resolution techniques and then segmented using (semi-)automated approaches. Image reconstruction and segmentation are then performed separately. In this paper, we propose an end-to-end generative adversarial network for simultaneous high-resolution reconstruction and segmentation of brain MRI data. This joint approach is first assessed on the simulated low-resolution images of the high-resolution neonatal dHCP dataset. Then, the learned model is used to enhance and segment real clinical low-resolution images. Results demonstrate the potential of our proposed method with respect to practical medical applications.

Index Terms— Super-resolution, segmentation, 3D generative adversarial networks, neonatal brain MRI.

1. INTRODUCTION

Long-term studies of the outcome of prematurely born infants have clearly documented that the majority of such infants may have significant motor, cognitive, and behavioral deficits. However, there is a limited understanding of the nature of the cerebral abnormality underlying these adverse neurologic outcomes. Magnetic Resonance Imaging (MRI) provides unique opportunities for in vivo investigation of the early developing human brain. However, the analysis of clinical neonatal brain MRI data remains challenging mainly due to low anisotropic image resolution. Improving morphological data processing such as image resolution enhancement and brain segmentation, is a key-point to provide robust morphometry analysis tools to the community.

One of the first key components of the processing pipeline of clinical MRI data is the upsampling image estimation. Super-resolution (SR) is a post-processing technique that aims at enhancing the resolution of an imaging system [1]. SR is a challenging inverse problem; in particular the estimation of texture and details remains difficult. Recently, supervised deep learning-based techniques have shown great improvement over model-based approaches. In this context, applying 3D convolutional neural networks (CNNs) yields promising results for MRI data [2, 3]. However, the use of ℓ_2-norm loss leads to smooth, unrealistic high resolution images [4, 5]. Generative adversarial networks (GANs) have thus been proposed to estimate textured and sharper images [5, 6].

Once the high resolution image reconstruction is performed, the implementation of an automatic segmentation robust approach is crucial for fine brain structure analysis [7]. Segmenting thin structures such as the neonatal cortical gray matter remains difficult and the segmentation step is always considered separately from image reconstruction.

In this article, we propose an end-to-end GAN-based approach which can generate both the perceptually super-resolved image and a cortical segmentation map from a single low-resolution (LR) image. The proposed approach called SegSRGAN is both assessed on simulated data and real clinical data.

2. METHOD

In this work, we focus on joint reconstruction and cortical segmentation of clinical LR neonatal brain MRI data. Section 2.1 describes the mathematical formulation of SR and Section 2.2 focuses on image segmentation. The proposed joint method is presented in Section 2.3, while technical details on the induced network architecture are provided in Section 2.4. Experiments on simulated and clinical MRI data are described and discussed in Section 3.

2.1. Formulation of single image super-resolution

The objective of single image SR is to estimate a high-resolution (HR) image $\mathbf{X} \in \mathbb{R}^m$ from one observed LR image $\mathbf{Y} \in \mathbb{R}^n$. SR problem can be formulated using the following linear observation model:

$$\mathbf{Y} = H_\downarrow B \mathbf{X} + N = \Theta \mathbf{X} + N \qquad (1)$$

where N is the additive noise, $B \in \mathbb{R}^{m \times m}$ is a blur matrix (depending on the point spread function), $H_\downarrow : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a downsampling decimation and $\Theta = H_\downarrow B \in \mathbb{R}^{n \times m} (m > n)$.

A popular approach that solves SR problem defines the matrix Θ^{-1} as the combination of a restoration operator $F \in \mathbb{R}^{m \times m}$ and an upscaling interpolation operator $S^\uparrow : \mathbb{R}^n \rightarrow \mathbb{R}^m$ computing the interpolated LR image $\mathbf{Z} \in \mathbb{R}^m$ ($\mathbf{Z} = S^\uparrow \mathbf{Y}$). In the context of supervised learning, given a set of HR images $\mathbf{X}_i$ and their corresponding LR images $\mathbf{Y}_i$, the restoration operator F can be estimated by minimizing the following loss function: $\widehat{F} = \arg\min_F \sum_i \|\mathbf{X}_i - F(\mathbf{Z}_i)\|_2^2$.

However, it is known that the use of ℓ_2-norm may lead to oversmoothing high resolution images. In order to provide realistic HR images, perceptual loss function [4] have been used in a GAN [5]. This is a paradigm shift since it is no longer a question of minimizing only the reconstruction error but of estimating a realistic image, i.e. a high resolution image that corresponds to the observation model with realistic texture details.

A perceptual loss can be formulated as the weighted sum of the content loss (based, e.g., on pixel-wise mean squared error loss) and an adversarial loss component. In GAN-based approaches, the purpose is to train a generating network G that estimates for a given LR input image a corresponding HR image. The goal of the discriminator network D is to classify real and simulated HR images.

2.2. Formulation of image segmentation

In this work, image segmentation is viewed as a supervised regression problem: $\mathbf{S_X} = R(\mathbf{X})$ where R denotes a non-linear mapping from the upscaled image $\mathbf{X}$ to the segmentation map $\mathbf{S_X}$. Similarly to the SR problem, assuming that we have a set of images $\mathbf{X}_i$ and corresponding segmentation maps $\mathbf{S}_{\mathbf{X}_i}$, a general approach for solving this segmentation problem is to find the mapping R by minimizing the following loss function: $\widehat{R} = \arg\min_R \sum_i \|\mathbf{S}_{\mathbf{X}_i} - R(\mathbf{X}_i)\|_2^2$. Unlike the SR problem, the use of ℓ_2-norm is less critical as it is expected to estimate smooth segmentation maps.

2.3. Joint mapping by generative adversarial networks

We propose the use of a GAN-based approach to estimate jointly a HR image and its corresponding segmentation map from one LR image. To this end, a convolution-based generator network G takes as input an interpolated LR image $\mathbf{Z}$

and computes a HR image $\widehat{\mathbf{X}}$ and a segmentation map $\widehat{\mathbf{S_X}}$ by minimizing the following reconstruction loss:

$$\mathcal{L}_{rec} = \min_G \sum_i \rho\left((\mathbf{X}, \mathbf{S_X})_i - G(\mathbf{Z}_i)\right) \qquad (2)$$

where $(\mathbf{X}, \mathbf{S_X})_i$ are concatenated along the feature channel. In this work, we use a robust loss as Charbonnier loss [8]: $\rho(x) = \sqrt{x^2 + \nu^2}$ where ν is set to 10^{-3}.

The discriminator network D attempts to distinguish the real data $(\mathbf{X}, \mathbf{S_X})$ and the generated ones $G(\mathbf{Z})$. The game between the generator G and the discriminator D is usually modeled with a minimax objective. However, using such loss function, GAN may be unstable or can suffer from mode collapse during training. Thus, in this work, we propose to use Wasserstein GAN loss described in [9]:

$$\mathcal{L}_{adv} = \min_G \max_D \mathbb{E}_{\mathbf{X} \sim \mathbb{P}_\mathbf{X}, \mathbf{S_X} \sim \mathbb{P}_{\mathbf{S_X}}}[D((\mathbf{X}, \mathbf{S_X}))]-$$
$$\mathbb{E}_{\mathbf{Z} \sim \mathbb{P}_\mathbf{Z}}[D(G(\mathbf{Z}))] + \lambda_{gp} \mathbb{E}_{\widehat{\mathbf{XS}}}[(\| (\nabla_{\widehat{\mathbf{XS}}} D(\widehat{\mathbf{XS}}) \|_2 - 1)^2]$$
$$(3)$$

where $\widehat{\mathbf{XS}}$ is the interpolation of the true data and the generated one as $(1-\epsilon)(\mathbf{X}, \mathbf{S_X}) + \epsilon G(\mathbf{Z})$, $\epsilon \sim \mathcal{U}[0,1]$ (uniform distribution). λ_{gp} and ∇ denote the gradient penalty coefficient and gradient operator, respectively. The images $\mathbf{X}$, $\mathbf{S_X}$ and $\mathbf{Z}$ are extracted randomly from the data distributions of HR images $\mathbb{P}_\mathbf{X}$, HR segmentation maps $\mathbb{P}_{\mathbf{S_X}}$ and LR images $\mathbb{P}_\mathbf{Z}$. The terms $D((\mathbf{X}, \mathbf{S_X}))$, $D(G(\mathbf{Z}))$ and $D(\widehat{\mathbf{XS}})$ are the responses of the discriminator with respect to the real data, the generated data and the interpolated data, respectively. The full objective function is expressed as: $\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_{adv}\mathcal{L}_{adv}$ where λ_{adv} is a trade-off parameter between reconstruction loss and adversarial loss.

2.4. Architecture of generator and discriminator networks

The generator network is a convolution-based network with residual blocks. Let C_j^i-S^k be a block consisting of the following layers: a convolution layer of j filters of size i^3 with stride of k, an instance normalization layer (InsNorm) [10] and a rectified linear unit (ReLU). R_k denotes a residual block as Conv-InsNorm-ReLU-Conv-InsNorm that contains 3^3 convolution layers with k filters. U_k denotes layers as Upsampling-Conv-InsNorm-ReLU layer with k filters of 3^3 and stride of 1. After the last layer, we apply a sigmoid activation for the channel of segmentation map and an element-wise sum of the intensity channel and the interpolated LR image. The generator architecture is: C_{16}^7-S^1, C_{32}^3-S^2, C_{64}^3-S^2, R_{64}, R_{64}, R_{64}, R_{64}, R_{64}, R_{64}, U_{32}, U_{16}, C_2^7-S^1.

The discriminator network contains five convolutional layers with an increasing number of filter kernels, increasing by a factor of 2 from 32 to 512 kernels. Let C_k be a block consisting of the following layers: a convolution layer of k filters of size 4^3 with stride of 2 and a Leaky ReLU with a

negative slope of 0.01. The last layer C_1^2 is a 2^3 convolution filter with stride of 1. No activation layer is used after the last layer. The discriminator consists of C_{32}, C_{64}, C_{128}, C_{256}, C_{512}, C_1^2.

3. EXPERIMENTS AND RESULTS

3.1. Datasets and network training

To assess the ability to reconstruct HR volume and segment the cerebral cortex , we applied the proposed method on T2-weighted (T2w) MR images of the developing Human Connectome Project[1] (dHCP). 40 T2w images were acquired using a 3T Achieva scanner with a $0.5 \times 0.5 \times 0.5$ mm^3 resolution with TR = 12 000 ms TE = 156 ms, respectively. 30 images were used for training networks, whereas the other 10 were used as testing images. As in [1], LR images were generated by using a Gaussian blur with the full-width-at-half-maximum (FWHM) set to slice thickness before a downsampling step to obtain a $0.5 \times 0.5 \times 1.5$ mm^3 resolution.

We have also applied the proposed method onto clinical neonatal MRI data acquired in the neonatology service of Reims Hospital. These LR images have a resolution of $0.446 \times 0.446 \times 3$ mm^3. 40 HR images of the dataset dHCP were filtered and downsampled as in [1] in order to generate LR images with a same resolution as clinical data. The network was trained using 40 pairs of simulated data and then applied to real LR images for visual evaluations. All data had bias correction and for network training, they were normalized between 0 and 1. No subjects nor image patches appear twice in the different subsets.

The 3D network was trained over 200 epochs with batch size of 16, using Adam method with learning rate of 0.0001 and updates the discriminator 5 times before training the generator as in [9]. The parameters λ_{adv} and λ_{gp} were set to 0.001 and 10 respectively. The training patch size is 64^3. At test time, the whole HR image and segmentation volume were reconstructed by the weighted predictions of patches. A thresholding at 0.5 has been performed to obtain binary segmentation maps.

3.2. Results

Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) have been used to evaluate the performance of SR reconstructions. Table 1 provides a summary of quantitative evaluations for the following methods: cubic spline interpolation, a 20-layers CNN-based SR approach (20L-SRReCNN) [11] and our proposed SegSRGAN. It can be seen that 20L-SRReCNN provides highest PSNRs as in [4, 5] since this approach minimizes a ℓ_2-norm-based loss. However, while the two CNN-based approaches (20L-SRReCNN and SegSRGAN) lead qualitatively to similar realistic results

---

[1]http://www.developingconnectome.org

Table 1. Quantitative evaluation of SR methods on dHCP dataset.

	Interpolation	20L-SRReCNN [11]	SegSRGAN
PSNR	30.70	35.84	31.75
SSIM	0.9492	0.9739	0.9624

Table 2. Quantitative evaluation of segmentation methods on dHCP dataset.

	IMAPA [12]	DrawEM [13]	SegSRGAN
Dice (mean)	0.788	0.818	0.886
Dice (std. dev.)	0.061	0.014	0.011

on dHCP dataset (see Fig. 1), the proposed approach provides the best reconstructed HR image on clinical data with better contrast on cortical gray matter (see Fig. 2).

The Dice index is used to evaluate the cortical segmentation maps obtained by the following state-of-the-art methods: iterative multi-atlas patch-based approach (IMAPA) [12], DrawEM [13] and the proposed SegSRGAN. As in a typical clinical setting, the three methods have been applied on interpolated images. Table 2 shows that quantitatively the proposed approach lead to the best cortical segmentation results with significant improvement with respect to the two other methods. Moreover, as mentioned in [12], the use of IMAPA applied on original HR dHCP images leads to a mean DICE of 0.887 (std. dev. of 0.011) that is very similar to the results obtained with SegSRGAN (applied on interpolated images). Results on real LR data (see Fig. 2) confirm the potential of the proposed approach for fine analysis of clinical neonatal brain MRI.

4. CONCLUSION

In this paper, we have presented a simultaneous super-resolution and segmentation method for 3D brain MR images using a generative adversarial network. Our experiments on both simulated and clinical data have shown that better performance can be achieved by this joint approach compared to state-of-the-art techniques, opening up new perspectives in the processing of clinical LR neonatal brain MRI data.

5. REFERENCES

[1] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, 2008.

[2] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain MRI super-resolution using deep 3D convolutional networks," in *IEEE ISBI*, 2017, pp. 197–200.

[3] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, "Brain MRI super resolution using 3D deep densely connected neural networks," in *IEEE ISBI*, 2018, pp. 739–742.

(a) Original dHCP HR   (b) Spline interpolation   (c) 20L-SRReCNN [11]   (d) Proposed approach

(e) Original dHCP segmentation   (f) IMAPA [12]   (g) DrawEM [13]   (h) Proposed approach
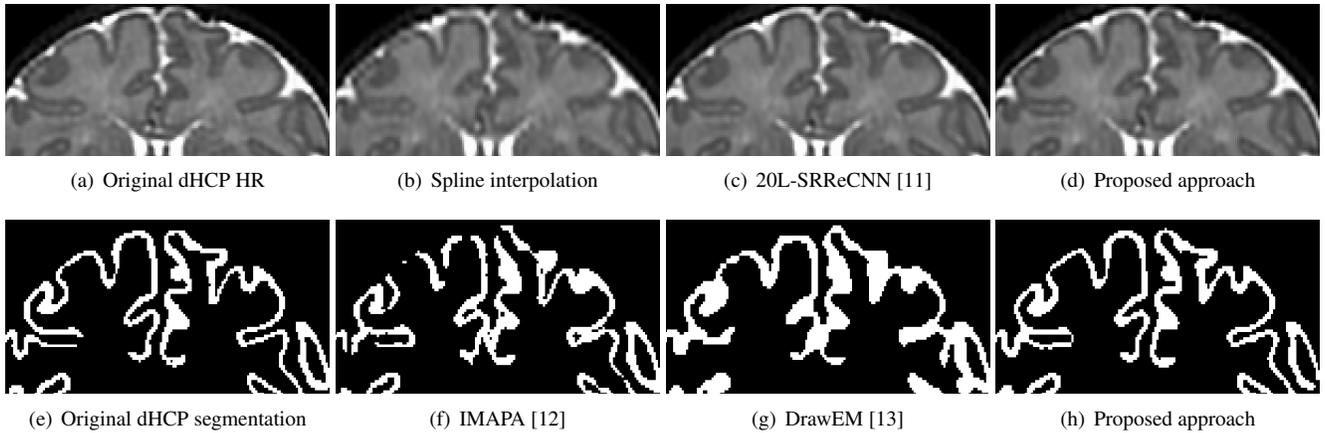
Fig. 1. SR and segmentation results for one dHCP subject: (a) original HR image; (b–d) SR reconstruction of the LR image generated from (a); (e) segmentation ground-truth of (a); (f,g) segmentation of (b); (h) HR segmentation from the LR image using the joint SegSR-GAN method.
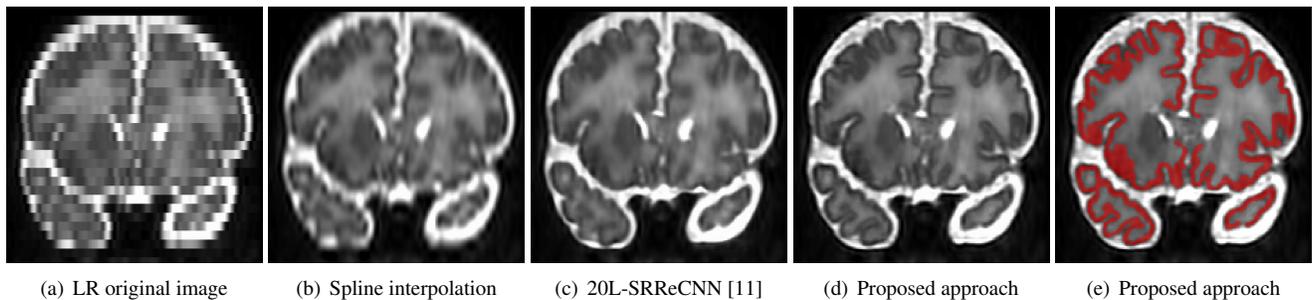

(a) LR original image   (b) Spline interpolation   (c) 20L-SRReCNN [11]   (d) Proposed approach   (e) Proposed approach

Fig. 2. Reconstruction (b–d) and segmentation results (e) on a real LR neonatal brain image (a) with voxel size of $0.446 \times 0.446 \times 3 \text{ mm}^3$, re-sampled to $0.5 \times 0.5 \times 0.5 \text{ mm}^3$.

[4] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016, pp. 694–711.

[5] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE CVPR*, 2017, pp. 105–114.

[6] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," in *MICCAI*, 2018, pp. 91–99.

[7] A. Makropoulos, S. J. Counsell, and D. Rueckert, "A review on automatic fetal and neonatal brain MRI segmentation," *NeuroImage*, vol. 170, pp. 231–248, 2017.

[8] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[9] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *NIPS*, 2017, pp. 5769–5779.

[10] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *IEEE CVPR*, 2017, pp. 4105–4113.

[11] C.-H. Pham, R. Fablet, and F. Rousseau, "Multi-scale brain MRI super-resolution using deep 3D convolutional networks," Preprint hal-01635455, 2017.

[12] C. Tor-Díez, N. Passat, I. Bloch, S. Faisan, N. Bednarek, and F. Rousseau, "An iterative multi-atlas patch-based approach for cortex segmentation from neonatal MRI," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 73–82, 2018.

[13] A. Makropoulos et al., "Automatic whole brain MRI segmentation of the developing neonatal brain," *IEEE Transactions on Medical Imaging*, vol. 33, no. 9, pp. 1818–1831, 2014.