Deep Learning based MRI contrast synthesis using full volume prediction

Jose V Manjon, Jose E Romero, Pierrick Coupé

Deep Learning based MRI contrast synthesis using full volume prediction

José V. Manjón[1], José E. Romero[1], Pierrick Coupe[2]

[1]Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politécnica de Valencia, Camino de Vera s/n, 46022, Spain

[2]CNRS, Univ. Bordeaux, Bordeaux INP, LaBRI, UMR5800, PICTURA Research Group, 351, cours de la Liberation F-33405 Talence cedex, France.

Abstract

In Magnetic Resonance Imaging (MRI), depending on the image acquisition settings, a large number of image types or contrasts can be generated showing complementary information of the same imaged subject. This multi-spectral information is highly beneficial since can improve MRI analysis tasks such as segmentation and registration, thanks to pattern ambiguity reduction. However, the acquisition of several contrasts is not always possible due to time limitations and patient comfort constraints. Contrast synthesis has emerged recently as an approximate solution to generate other image types different from those acquired originally. Most of the previously proposed methods for contrast synthesis are slice-based which result in intensity inconsistencies between neighbor slices when applied in 3D. We propose the use of a 3D convolutional neural network (CNN) capable of generating T2 and FLAIR images from a single anatomical T1 source volume. The proposed network is a 3D variant of the UNet that processes the whole volume at once breaking with the inconsistency in the resulting output volumes related to 2D slice or patch-based methods. Since working with a full volume at once has a huge memory demand we have introduced a spatial-to-depth and a reconstruction layer that allows working with the full volume but maintain the required network complexity to solve the problem. Our approach enhances the coherence in the synthesized volume while improving the accuracy thanks to the integrated three-dimensional context-awareness. Finally, the proposed method has been validated with a segmentation method, thus demonstrating its usefulness in a direct and relevant application.

Keywords: Image synthesis, Magnetic Resonance Imaging (MRI), Convolutional Neural Network (CNN), segmentation

1. Introduction

Magnetic Resonance Imaging (MRI) has gained importance over the years due to its capability of acquiring a diverse range of contrasts for the same underlying anatomy in the absence of tissue irradiation. Since its early years, neuroimaging has been of the main application of MRI thanks to the possibility of visualizing the brain soft tissues in a non-invasive manner and with high resolution level. The variety of MR image contrasts obtainable depending on the image acquisition settings provides complementary information of the same imaged subject. For example, T1 brain images clearly delimit white and grey matter tissues, whereas T2 images distinguish fluid from cortical tissue, and white matter lesions are more easily distinguishable in FLAIR contrasts. This increased availability of information leads to a wider perspective into diagnosis improving the understanding of the state of the living brain tissues by clinicians. In the same manner, automatic segmentation methods results are generally improved when receiving a multimodal input feeding.

Unfortunately, the acquisition of several contrasts is not always possible due to time limitations and patient comfort constraints. Thus, the ability to synthesize non-existent or corrupted contrasts from others successfully acquired has a potential value for enhancing MRI analysis. Despite not having yet achieved suitable results for direct clinical use, MRI synthesis is useful for the improvement of algorithm performance in relevant image analysis tasks such as registration and segmentation (Iglesias et al.,2013).

In neuroimaging, one of the first uses of contrast synthesis was the use of software packages designed for a specific input contrast different from the one available. In Iglesias et al. (2013) the authors generated T1 images from T2 images to be able to use the software Freesurfer in order to segment the whole brain.

Traditionally, two families of methods have been used to generate synthetic images: registration-based methods and intensity transformation-based methods. The first group originates from the work carried out in Miller et al. (1993), and consists generally in applying a geometric transformation to a target modality atlas to obtain the target modality image, having computed that transformation from registration of the source modality atlas to the source modality image. This approach assumes the relation between

images from different subjects is purely geometric, which can lead to significant errors in fine brain detail synthesis. Furthermore, registration-based synthesis methods perform poorly in the presence of anatomic abnormalities e.g. a tumour, as these are not present in the atlases used (Jog et al., 2017).

Intensity transformation-based methods overcome the previous limitations, not relying in a strictly geometric relation between subjects. This wide family embraces methods based in a variety of techniques, such as the sparse dictionary reconstruction Roy et al. (2013) or the image analogies Iglesias et al. (2013) approaches. However, these methods require large computational times and their application extent is somehow restricted.

During the last years, machine learning techniques have made their way in the medical image analysis field, achieving outstanding results. In MRI synthesis, non-linear regression methods represent the state of the art. These methods learn non-linear intensity mappings between pairs of image modalities. In Jog et al. (2017) the authors proposed REPLICA as a patch-based random forest method. The synthesized contrast depends on the average of the predicted outputs from independent random forest trees, which may produce a loss of high frequency detail. More recently, deep learning-based methods have been proposed to perform this task, Dar et al. (2019) used a conditional generative adversarial network to enforce the realism of the produced output. Unfortunately, although the produced outputs were highly realistic sometimes they hallucinate structures and showed strong inconsistencies between neighbor slices due to the 2D nature of their method. Nie et al. (2018), also used an adversarial network but this time using 2D patches and a cascade approach. On the other hand, Wei et al. (2019) were one of the first to use not only 3D inputs (3D patches) but also multiple input modalities with remarkable results. However, the need of using multiple input modalities limits its applicability in real world scenarios. Although all these methods showed promising results due to their ability to automatically learn structural features. Most of them are slice-based in which each slice of the volume is predicted independently (or with limited 3D support) that can lead to inconsistencies between neighbor slices. Moreover, the use of a limited spatial context may result in suboptimal results due to the lack of context-awareness.

To address the mentioned problem, we propose a 3D variant of the UNet network (Ronneberger et al., 2015) to synthesize T2 and FLAIR brain MRI contrasts from a single

anatomical T1- contrast. Our network processes the whole volume at once breaking with the inconsistency in the resulting output volumes related to 2D slice or patch-based methods. In that manner, it is enhancing the coherence in the synthesized volume and improving the accuracy via integrated three-dimensional context-awareness. This global approach contributes to yield visually and quantitatively more accurate output volumes through a maximization in the context information available and then use of a canonical intensity space.

2. Material and Methods

2.1. Datasets

2.1.1. Database I. T1/T2 dataset.

Paired and registered T1/T2 images in a standard space (MNI152) have been generated to train our deep learning-based method for T2 synthesis. For this purpose, the open access IXI dataset (https://brain-development.org/ixi-dataset/) has been used. This dataset contains images of 580 healthy subjects from several hospitals in London (UK). Both 1.5T and 3T T1 images were included in our training dataset. 3T images were acquired on a Philips Intera 3T scanner (TR = 9.6ms, TE = 4.6ms, flip angle= 8◦, volume size = 256 × 256 × 150 voxels, voxel dimensions = 0.94 × 0.94 × 1.2 mm^3). 1.5T images were acquired on a Philips Gyroscan 1.5T scanner (TR = 9.8ms, TE = 4.6ms, flip angle = 8◦, slice thickness = 1.2mm, volume size = 256 × 256 × 150, voxel dimensions = 0.94 × 0.94 × 1.2mm^3). The parameters of the 3T T2 images were (TR = 5725 ms, TE = 100 ms, flip angle= 90◦, volume size = 192 × 187 × 187 voxels, voxel dimensions = 0.94 × 0.94 × 1.2 mm^3) and for 1.5T were (TR = 8178 ms, TE = 100 ms, flip angle= 90◦, volume size = 192 × 187 × 187 voxels, voxel dimensions = 0.94 × 0.94 × 1.2 mm^3). Images from subjects presenting excessive artefacts were excluded from pair correlation-based selection, resulting in the final number of subjects used of 557. This selected set was then split into training, validation and testing sets (520, 20 and 17 subjects respectively).

2.1.2. Database II. T1/FLAIR dataset

Paired and registered T1/FLAIR images in a standard space (MNI152) have been generated to train our deep learning-based method for FLAIR synthesis. The lesionBrain (Coupe et al., 2018) pipeline database built from images uploaded to volBrain platform (Manjón and Coupé, 2016) by users with their express consent for use in research has

been used. Thus, it is a highly heterogeneous dataset: first, it includes both healthy subjects and subjects presenting different types of lesions (according to their intensity and location), and second, images have been acquired in a broad variety of hospitals and scanners. Initially, a total number of 5.463 subjects (T1/FLAIR pairs) were available. After an exhaustive manual quality-based selection process, a final number of 523 subjects were used. Those subjects either presenting excessive artefacts or presenting insufficient quality in terms of spatial resolution or that did not match the specified image modality, were discarded. This selected set was then split into training, validation and testing sets (493, 10 and 20 subjects respectively).

2.1.3. Segmentation database

To evaluate the usefulness of the image synthesis we used the MICCAI 2012 Multi-Atlas Labeling Challenge dataset. This dataset has 35 dense manually labeled cases (133 labels). We increased the number of cases by using also their left-right mirrored version resulting in a total number of cases of 70 subjects. We combined the 133 labels into 7 different tissue labels. These labels were: Cerebral Spinal Fluid (CSF), Cortical Grey Matter (CGM), cerebral White Matter (WM), SubCortical Grey Matter (SCGM), Cerebellar Grey Matter (CeGM), Cerebellar White Matter (CeWM) and BrainStem (BS). This dataset was divided into training, validation and testing sets (44, 6 and 20 respectively).

2.2. Preprocessing

Firstly, rigid registration between T1 and T2 images was estimated and concatenated with an affine registration of all the dataset to the MNI152 space using the ANTS software (Avants et al. ,2008). Afterwards, we performed an inhomogeneity correction of the registered images using N4 method (Tustison et al. ,2010). As we previously mentioned, our network learns a non-linear mapping between T1 and T2 or FLAIR images. However, mapping results are hindered if source and target spaces are not intensity-coherent themselves. Unlike other medical image modalities such as CT in which a standardized intensity scale (Hounsfield) exists, in MRI intensity distributions exhibit large variations depending on many technical factors such as magnetic field strength or acquisition parameters (TE, TR, flip angle, etc). Thus, arises the importance of intensity standardization of both input and output volumes to a reference canonical intensity space. Here we propose the use of histogram matching (Gonzales and Fittes, 1977) as a powerful

intensity normalization method. As reference, the non-linear MNI152 T1 and T2 templates were used for T1 and T2 volumes respectively. The preprocessing for the T1/FLAIR dataset was the same than for the T1/T2 dataset. For FLAIR volumes we used our own atlas generated by computing the median of 52 FLAIR volumes from database II non-linearly registered to the T1 MNI152 atlas and their horizontally flipped version. Finally, after this intensity mapping, we applied a mean-ratio normalization consisting in dividing the complete volume by its arithmetic mean. The segmentation database images (T1) were preprocessed in the same way than the other datasets with the exception of the histogram matching step.

2.3. MRI Synthesis Network

2.3.1. Architecture

The proposed method is a 3D UNet-like network (see figure 1), which learns non-linear mappings between the source volume (T1) and the target volume (T2 or FLAIR). However, the use of a full resolution input and output volume (182×218×182 voxels) imposes a limitation in the number of filters of the convolutional layers leading to suboptimal results (a maximum number of 8 filters could be used in our initial experiments). To deal with this limitation, we propose to decompose the input volume into 8 sub-volumes by decimating the original volume 8 times with one voxel shift in each 3D dimension. The resulting 8 sub-volumes were structured as an input tensor of 91×109×91x8 containing all the voxels of the original input volume. This channel-wise decomposition allows us to use a larger number of filters in the encoding and decoding paths as well as use all the information within the original source 3D volume. Particularly, the encoder path consists of a 3D convolutional layer (kernel size = 3×3×3 voxels) for each resolution level with ReLU activation and batch normalization layers followed by a strided convolution pooling layer (stride = 2). Symmetrically, an upsampling layer is used after each resolution level in the decoder path. The number of filters used begins with 40 filters in the first resolution level with an increment by a factor of 2 in the subsequent lower resolution levels in order to balance the loss of spatial resolution. In the same manner, the number of filters is decreased by a factor of 2 in the decoder path at subsequent crescent resolution levels.

Since the use of the same channel-wise decomposition in the output might result in local inconsistencies in the synthesized volume, a final reconstruction block was added

consisting in a trilinear interpolation layer followed by a 3D convolution layer (with 8 filters) plus a ReLU and batch normalization layers. Finally, a last 3D convolution layer with a single filter of size 1 × 1 × 1 was used to produce the output. Thus, the proposed network has a total number of 9.135.865 parameters, of which 9.131.769 are trainable.

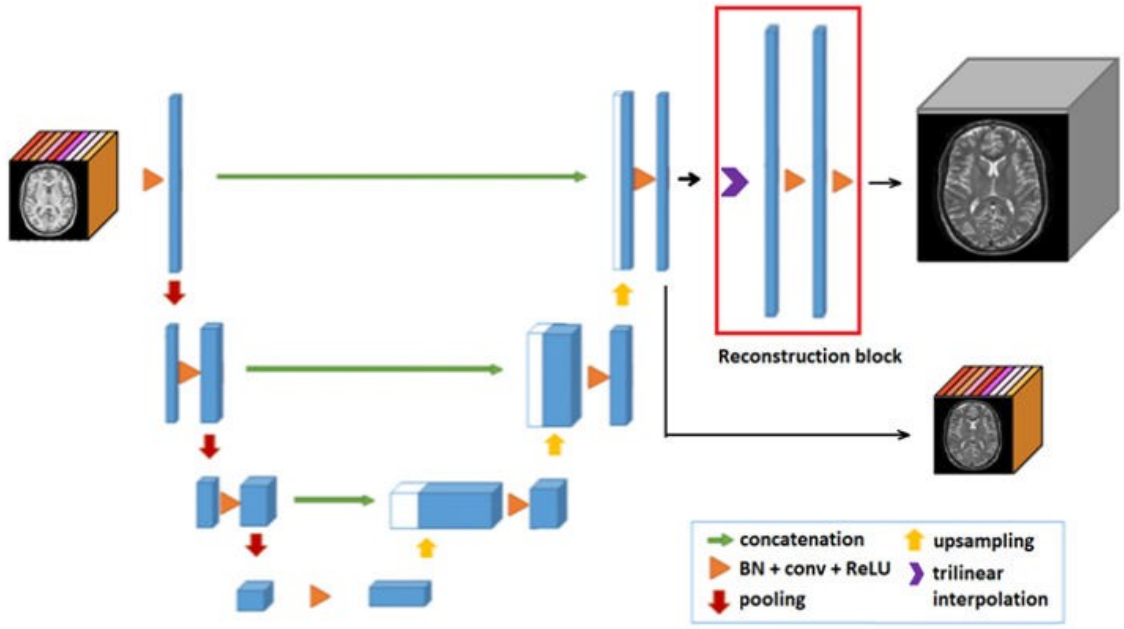

Figure 1: Proposed 3D network for MRI contrast synthesis scheme.

To Train the proposed network we used a deep supervision approach (Lee et al., 2015; Dou et al., 2017). The first output is the low-resolution channel-wise output and the second is the full resolution volume after the reconstruction block (see figure 1).

For all convolution operations, zero padding is used in such manner that the image is completely covered by the filter and specified stride. On the other hand, in reference to downsampling, strided convolution layers were chosen over classical pooling because of the advantage it provides having trainable parameters as opposed to the fixed operation the second performs. In this manner, the network can learn to summarize the data, thus improving its accuracy (Ayachi et al., 2018).

2.4. Training and evaluation

All the experiments were performed using Tensorflow 1.12.0 and Keras 2.2.4 in a NDIVIA GeForce GTX 1080 Ti GPU with 11GB of memory. To train the network we used an Adam optimizer (Kingma and Ba, 2014) (learning rate = 0.001) for both T2 and FLAIR synthesis and a mean squared error (MSE) loss for the first and a sum of the MSE and the mean absolute error (MAE) loss for the second, for being the ones yielding better results.

A data generator function was used to feed the network during 200 epochs (50 cases per epoch). We trained the network providing the 8 T1 sub-volumes as input data and the complete target volumes (T2 or FLAIR) as well as the 8 sub-volume decomposition as target data. We used data augmentation within the generator consisting in left-right flipping both the input and output randomly, thus taking advantage of the pseudo-symmetry of the brain.

In order to measure the performance of the method based on the quality of the results, three different metrics providing information about the similarity between the synthesized and the target volumes from different perspectives were used. In addition to classic peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), we include the Pearson correlation coefficient as a powerful metric to measure the structural information shared between ground-truth and predicted brain volumes independently from intensity. We consider it a more robust evaluation metric for our purpose, since MRI synthesis is about generating the correct contrasts between tissues independently of the intensity level. Evaluation metrics were performed both on full-head and skull-stripped volumes, as the main target areas of interest to synthesize accurately are the brain tissues.

2.5. Validation on Segmentation

As mentioned above, one of the most common applications of MRI synthesis is segmentation. Therefore, we validated our method by demonstrating the improvement it introduces in the accuracy of a volumetric 3D UNet segmentation network. The method segments the brain volume into the 7 brain tissues already described.

We performed 3 experiments to highlight the impact of synthetic data multimodal input feeding to the segmentation method. The results for these experiments will allow us to quantify the impact of including synthetic T2 and FLAIR volumes in the segmentation process. The different inputs received by the segmentation method in each of the 3 experiments are as follows:

- Monomodal: the segmentation network was trained and tested with a single original T1 input volume.
- Bimodal: the segmentation network was trained and tested with a dual input: the original T1 volume and its corresponding synthesized T2 volume.
- Trimodal: the segmentation network was trained and tested with a triple input: the original T1 volume and its corresponding synthesized T2 and FLAIR volumes.

In all of the 3 experiments, the network was trained during 400 epochs and subsequently evaluated. The classic Dice Coefficient was used to measure the similarity between predicted and ground-truth segmentations. It was computed as the average of the values for each of the 7 tissues individually.

3. Results and Discussion

3.1. Histogram matching intensity normalization impact

In our experiments, we tested the impact of histogram matching intensity normalization of the dataset in the resulting synthesized volumes. The network showed a clear improvement when the images were intensity normalized using histogram matching, as can be seen visually in Figure 2. The synthesized images using this intensity normalization method were able to better infer fine brain details. This may be explained by the fact that the intensity standardization to a common canonical space makes it easier for the network to learn the source- to-target mapping, and thus it can deepen into high frequency information.

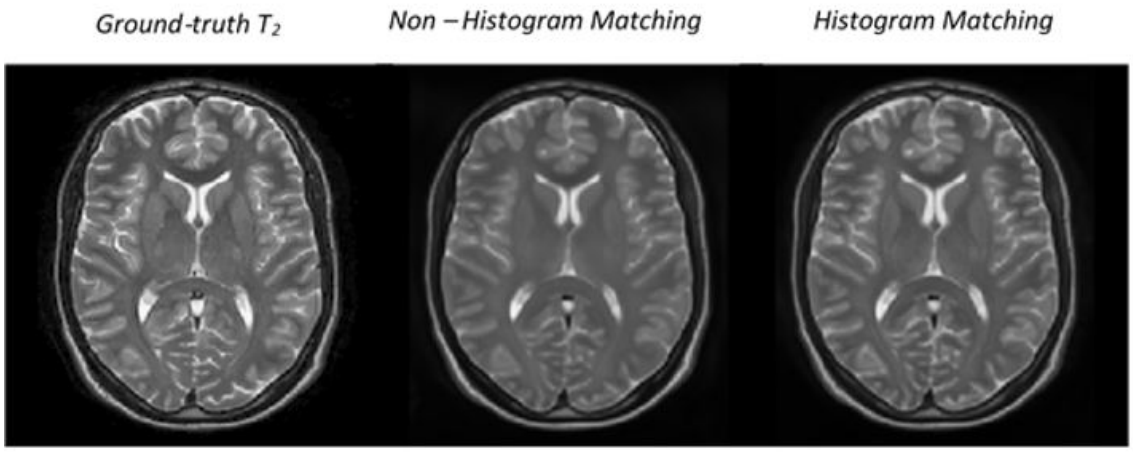

Figure 2: Target T2 contrasts (left) and synthesized images for non-normalized input (center) and histogram matching normalized input (right).

3.2. 3D context impact

To evaluate the importance of a complete 3D context awareness, we compared the proposed 3D method with its corresponding 2D version. To do it, we simply interchanged the 3D convolutions by 2D convolutions and used a standard 2D UNet, removing the channel-wise decomposition and reconstruction as well as the deep supervision. To make a fair comparison, we increased the number of filters to 70 (instead of 40) to have a similar number of learnable parameters.

In Table 1 the results of this comparison are shown. As can be noticed, the 2D approach obtained worse results for all metrics (both full-head and skull- stripped) and the resulting images showed the expected inter-slice artefacts due to a lack of context in the third dimension, as can be appreciated in Figure 3. Thus, this demonstrates the importance of an integrated three-dimensional context awareness for the brain MRI synthesis.

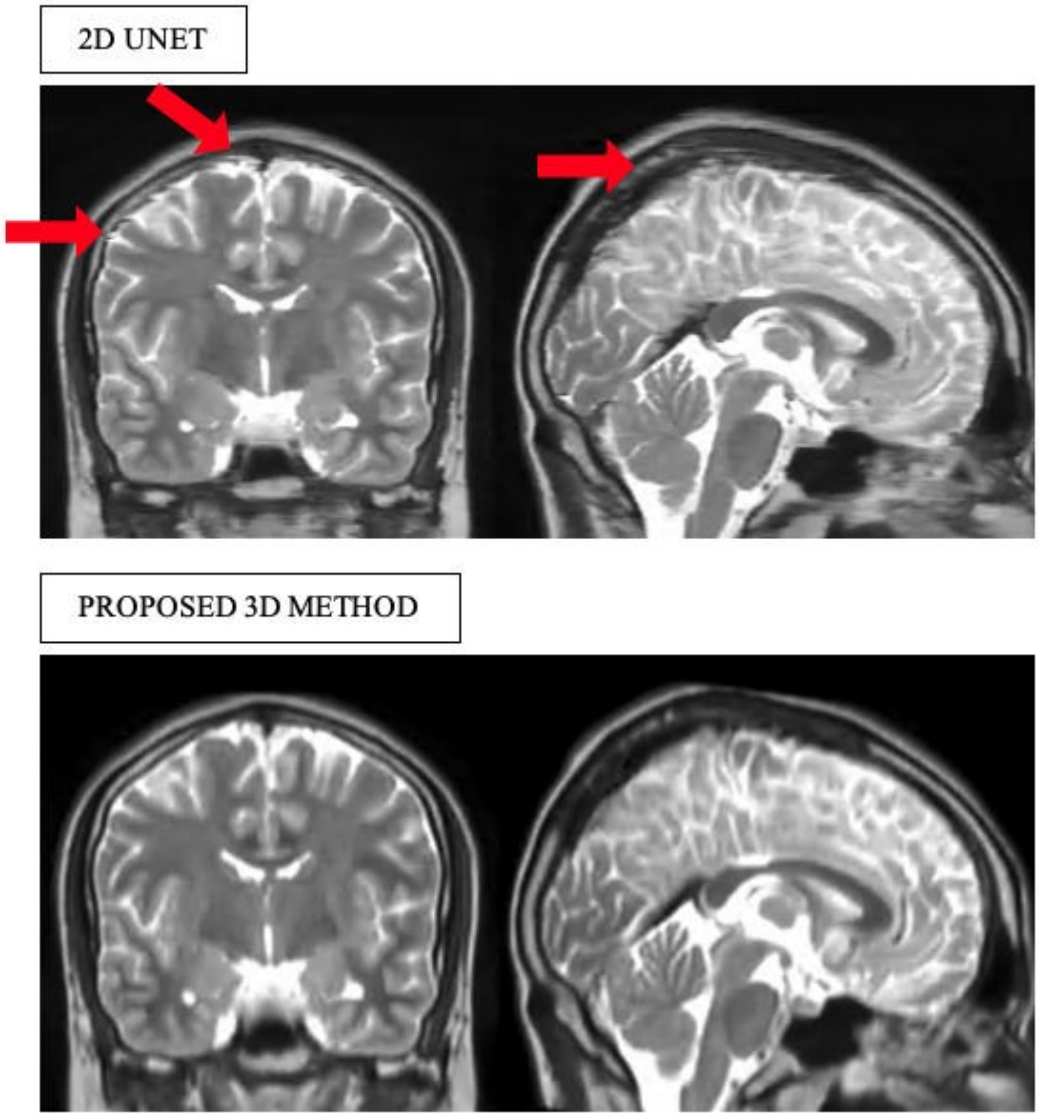

Figure 3: T2 contrasts generated by the 2D UNet (upper row) and by the proposed method (lower row). The inter-slice artefacts generated by the 2D UNet can be appreciated, due to the lack of a 3D context (red arrows).

3.3. Proposed architecture impact

Similarly, to evaluate the impact of the proposed 3D architecture, we compared it with a plain 3D UNet (no channel-wise decomposition and reconstruction) with the maximum number of filters able to fit into memory (nf=8). As shown in Table 1, the 3D plain UNet

obtained better results than its 2D counterpart (highlighting the importance of 3D context information) but worse than the proposed method, most probably due to the lower number of filters (nf = 8 vs nf = 40 with the proposed architecture). Visually, this performance difference between both networks can be observed in the resulting output volumes. As can be appreciated in Figure 4, different from the proposed method, the 3D UNet is not capable of generating the high frequency details in the images due to its restricted number of filters. Those fine details are fundamental in brain MRI, as they constitute important anatomical structures such as blood vessels. Therefore, the proposed network architecture allowed to considerably increase the number of filters, consequently achieving an important improvement in the results, mainly in relation to high frequency content.

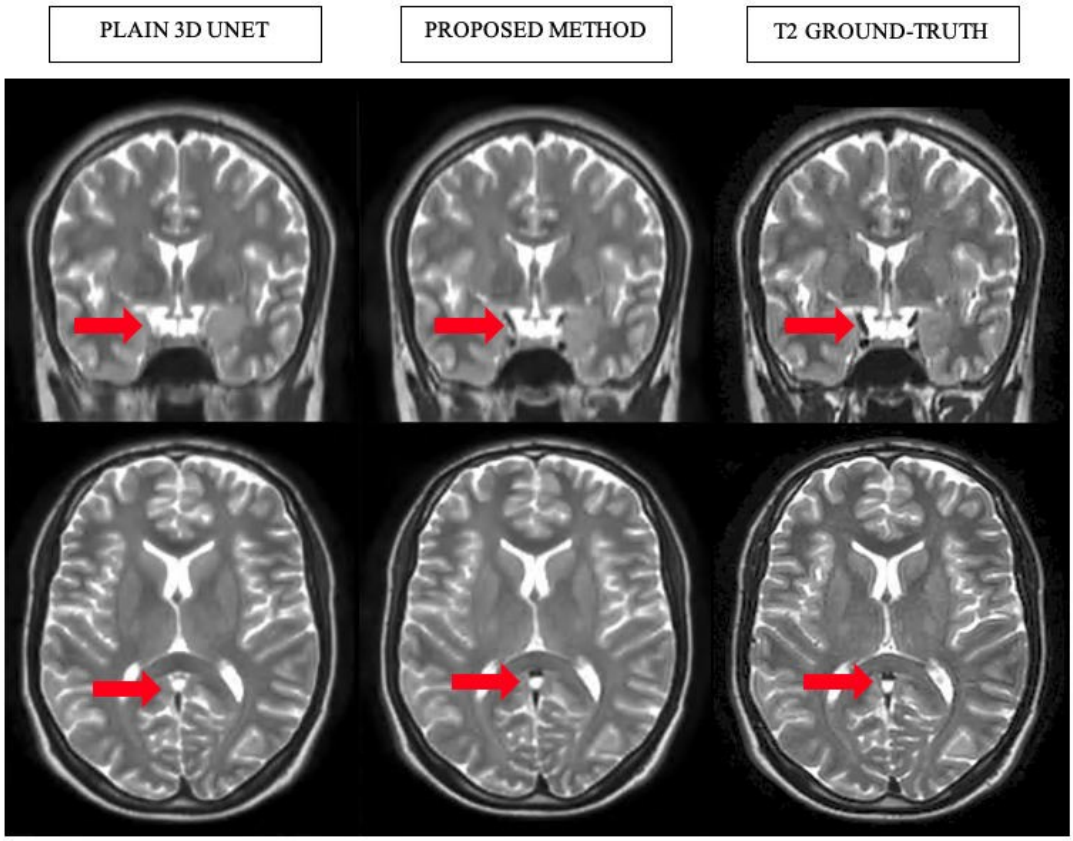

Figure 4: T2 contrasts generated by the 3D UNet (upper row) and by the proposed method (lower row). The 3D UNet inability of reconstructing fine details can be appreciated (red arrows).

	Proposed method		2D UNet		Plain 3D UNet	
	Full-head	*Skull-stripped*	*Full-head*	*Skull-stripped*	*Full-head*	*Skull-stripped*
CC	**0.974**	**0.994**	0.970	0.991	0.968	0.992
PSNR	**26.057**	**32.333**	22.881	26.881	25.203	31.096
SSIM	**0.844**	**0.963**	0.784	0.954	0.815	0.955

Table 1: Test results for the comparative among the proposed network, the 2D UNet and the plain 3D UNet. Best results in bold (CC: Correlation coefficient, PSNR: Peak Signal To Noise Ratio, SSIM: Structural Similarity Image Metric).

3.4. Comparison with state of the art

T1-T2 synthesis

In order to compare our method performance in the task of synthesizing T2 images given T1 input images, two representative state-of-the-art methods were used. The first method is REPLICA (Jog et al., 2017), a supervised random forest image synthesis approach that learns a non-linear regression to predict intensities of alternate tissue contrasts given specific input tissue contrasts. We trained REPLICA on the IXI dataset using the authors' code. In this method, different regressors perform feature analysis in a range of resolution levels, and eventually the individual outputs are averaged to yield the final output contrast. This functioning leads to certain blurriness in the resulting images and an important deficit in generating high resolution details as seen in Figure 5. In addition, since each voxel in the output contrast is predicted from 2D image patches in the source contrast, the built-up volume from the individually synthesized images tends to neglect coherence in the third dimension.

The second method under comparison was pGAN (Dar et al., 2019), a deep learning-based approach for multi-contrast MRI synthesis based on conditional generative adversarial networks. Test images were kindly processed by the authors and they sent us the results. The adversarial training strategy is able to generate more perceptually realistic images and even replicate the noise present in the source image. In spite of this, significant artefacts can be appreciated in the generated volumes (see Figure 5), such as bright areas confusing the background with cerebrospinal fluid. Even though the images are perceptually the most realistic, the generation of these artefacts results in low values for the evaluation metrics (Table 2). This realism has as a drawback that the structural information in the synthesized volume is not completely true to reality in the ground-truth

volume hallucinating some details. This may be due to a lack of a complete 3-dimensional context awareness of the volume.

Quantitative results of the comparison are shown in Table 2. As can be noticed, the proposed method outperformed the compared methods for both full-head and skull-stripped regions. However, we want to clarify that pGAN method was trained by the authors using a different dataset so SSIM and PSNR values cannot be exactly compared (we included these metrics for orientation purposes). Fortunately, correlation is an intensity range independent measure that is a less biased metric to compare structural information allowing a direct method comparison.

	Proposed method		REPLICA		pGAN	
	Full-head	*Skull-stripped*	*Full-head*	*Skull-stripped*	*Full-head*	*Skull-stripped*
CC	**0.974**	**0.994**	0.881	0.906	0.842	0.909
PSNR	**26.057**	**32.333**	23.964	27.756	18.147	24.570
SSIM	**0.844**	**0.963**	0.694	0.903	0.602	0.893

Table 2: Test results for our proposed network, REPLICA and pGAN methods. Best results in bold (CC: Correlation coefficient, PSNR: Peak Signal To Noise Ratio, SSIM: Structural Similarity Image Metric).

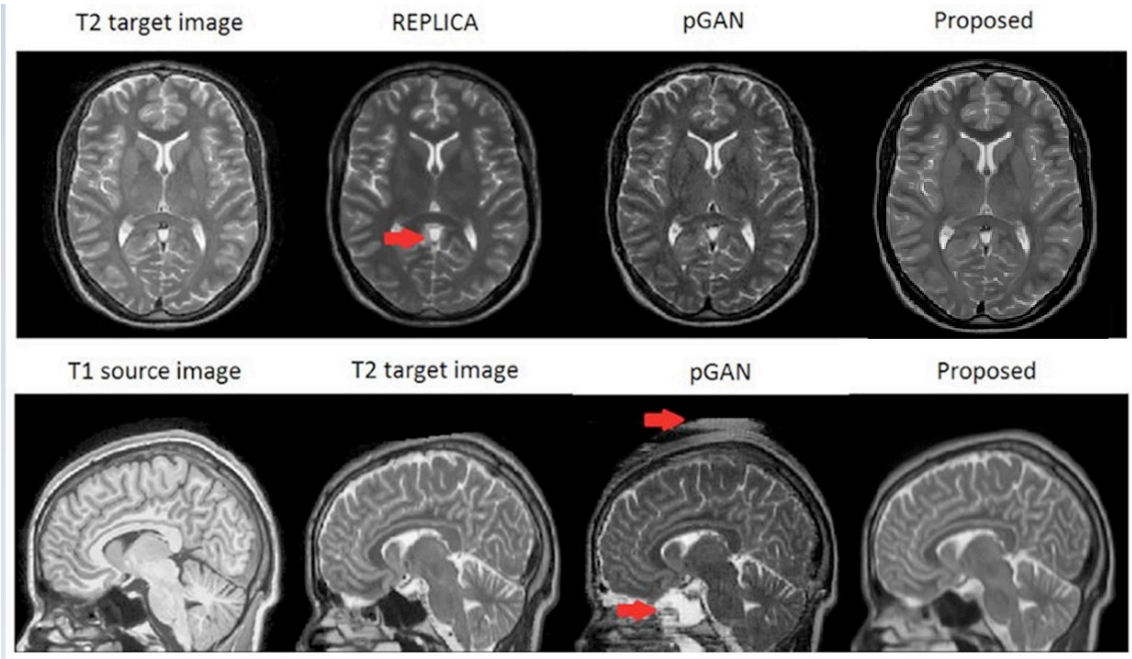

Figure 5: Upper row: Target T2 contrasts (left), synthesized images with REPLICA, the proposed method and pGAN. Bottom row: T1 source image, T2 target image, and images synthesized by pGAN and the proposed method. Note that REPLICA is not able to recover dark areas in the center of the brain and pGAN introduces slice dependent artefacts (red arrows).

Another aspect that is worth to mention is that results of the proposed method were very intensity coherent, showing intensity distributions matching the reference MNI152 T2 template (see Figure 5). Also, a relevant aspect is the fact that the proposed method resulting volumes did not exhibit strong artefacts unlike REPLICA and pGAN.

From an efficiency point of view, the proposed method takes around 0.5 seconds to predict a 3D T2 volume in the MNI space, while REPLICA takes around 15 minutes and pGAN around 10 seconds.

T2-FLAIR synthesis

When we compare the performance of the proposed method on both T2 and FLAIR synthesis tasks, we observe better results for T2 synthesis (see Table 3). The reason underlying this difference probably stems from the distinct nature of the two databases used to train the method. While the IXI T1/T2 dataset is fairly homogeneous, including only healthy subjects and acquired in only two different scanners, the FLAIR dataset is much more heterogeneous. The latter presents several variability sources: images from subjects both healthy and with diverse lesions are included, and moreover these were acquired in multiple different scanners around the world. Thus, the FLAIR synthesis case reflects more accurately the problems introduced by the inherent MRI heterogeneity due not only to the pathological variability, but also to the acquisition process, which hinders automatic MRI analysis tasks. Despite having achieved slightly inferior results compared to the T2 synthesis network, the FLAIR synthesis method has been trained on a more diverse dataset and thus it should present a higher generalization capability.

	T1/T2		T1/FLAIR	
	Full-head	*Skull-stripped*	*Full-head*	*Skull-stripped*
CC	0.9740	0.9938	0.9634	0.9908
PSNR	26.0575	32.3327	22.3147	28.1550
SSIM	0.8445	0.9634	0.7203	0.9239

Table 3: Test results for our proposed method for the tasks of T2 and FLAIR synthesis (CC: Correlation coefficient, PSNR: Peak Signal To Noise Ratio, SSIM: Structural Similarity Image Metric).

3.5. Validation on segmentation

As can be appreciated in Table 4, the Dice coefficient value increases as the number of input modalities to the network increases. In Figure 6 it can also be observed visually the higher accuracy of the tri-modal segmentation compared to the mono-modal, especially in thin tissue areas, being the first more detailed and closer to the ground-truth segmentation.

The introduction of synthesized T2 and FLAIR images from a source T1 case as additional inputs to the segmentation method, provides complementary information and different perspectives of the same underlying tissues. As expected, segmentation results are benefited by this increasing in the amount of available information in the form of multiple input modalities. Image synthesis is thus a means to take advantage of the wealth of information MRI offers through its broad variety of contrasts, being demonstrated the usefulness of the proposed T2/FLAIR images synthesis method in the improvement of 3D brain segmentation accuracy.

	Dice Coefficient							
	CSF	*CGM*	*WM*	*SCGM*	*CeGM*	*CeWM*	*BS*	*Average*
Monomodal (T1)	0.868	0.943	0.967	0.919	0.953	0.941	0.950	0.9348
Bimodal (T1, T2)	0.872	0.946	**0.970**	**0.922**	0.954	0.946	0.951	0.9374
Trimodal (T1, T2, FLAIR)	**0.883**	**0.949**	0.968	0.921	**0.956**	**0.947**	**0.953**	**0.9395**

Table 4: Test results for UNet segmentations for monomodal, bimodal and trimodal inputs. Best results in bold.

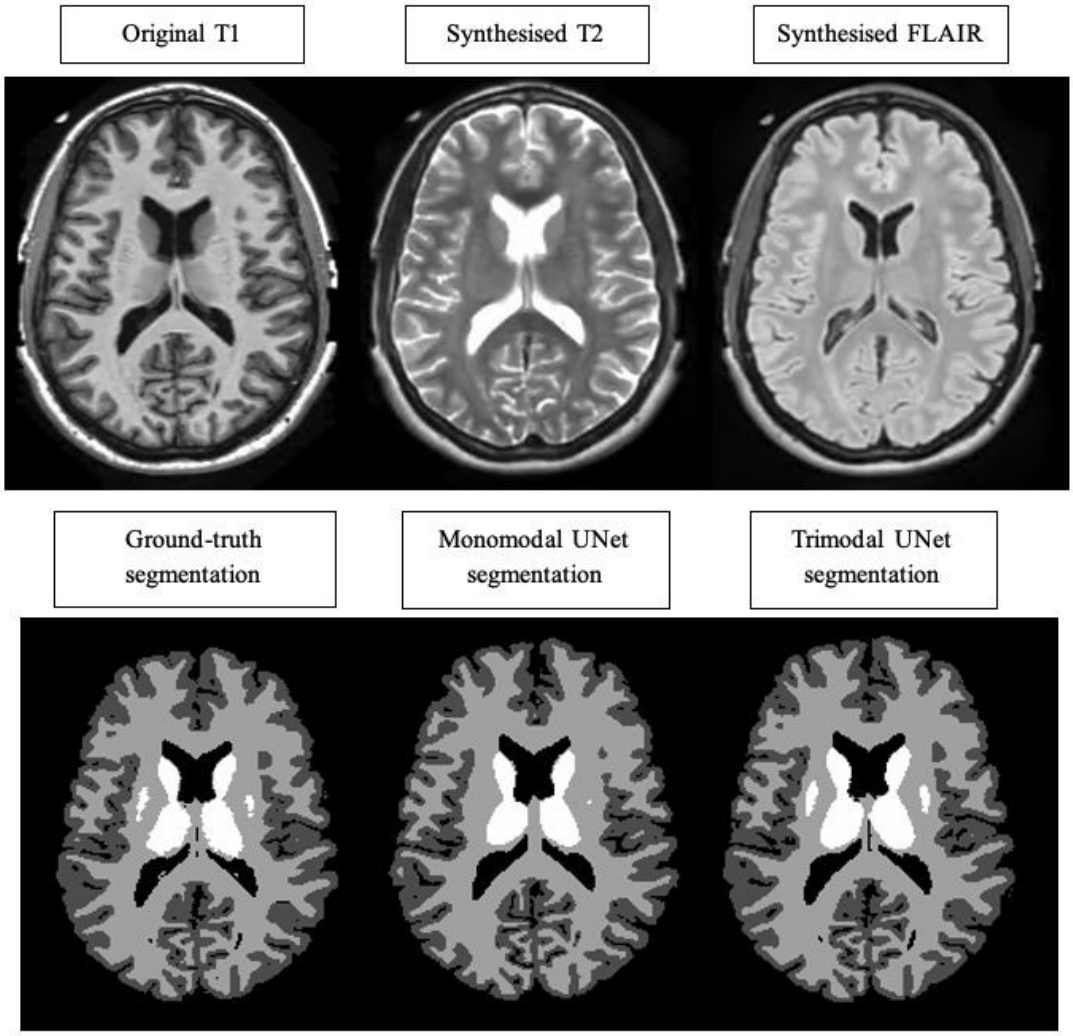

Figure 6: Upper row: Original T1 image, and T2 and FLAIR images generated by the proposed method. Bottom row: ground truth segmentation and UNet segmentations for mono and tri-modal inputs. Note the higher precision of the tri-modal segmentation, which can be clearly appreciated in finer areas.

4. Conclusions

In this paper we presented a deep learning-based approach to address the task of brain MRI contrast synthesis and we demonstrated its performance in the T1 to T2 and T1 to FLAIR contrast synthesis cases. The proposed method is based on a memory efficient 3D variant of the well-known UNet architecture allowing the whole brain volume processing at once while giving the proposed network enough complexity to deal with the synthesis task. This has been done thanks to the use of spatial-to-depth decomposition and a final reconstruction layer combined with a deep-supervision based training.

Our proposed method has been shown to outperform two relevant slice/patch-based state-of-the-art methods, REPLICA and pGAN, in terms of accuracy and efficiency. The full-volume nature of the proposed method allowed to obtain consistent predictions with no slice-based artefacts yielding improved results by maximizing the 3D context information.

More specifically, remarkable improvement in the results accuracy was obtained by using an additional intensity normalization via histogram matching. We believe the use of this normalization increases the generalization capability of the proposed method via mapping into a canonical intensity space and can potentially reduce the domain adaptation problem when using images from different scanners or with different acquisition settings.

The utility of the proposed synthesis method has been demonstrated through its application to a brain volume segmentation problem. It has been shown the improvement introduced in the performance of a 3D segmentation UNet when including as additional input the corresponding T2 and FLAIR contrasts generated by the proposed network. This complementary information contribution has been found to be useful for the crucial task of automatic segmentation of the brain (probably due to a reduction in the tissue pattern ambiguity).

The proposed method achieved relevant performance in the task of generating T2 and FLAIR contrasts from a T1 source contrast, and these particular synthesized images have been proved valuable for improving brain segmentation results. However, generated images are still a bit blurred. This probably is related to the used loss function (based on L1 norm). In the future we will explore other metrics such as perceptual or texture losses and then use of a weighted version of adversarial losses to force pattern consistency among image modalities.

Finally, the proposed architecture might potentially be used for generating different target contrasts from different acquired source 3D images, with the possibility of making an impact on another relevant task's performance such as registration. Further experimentation would be required for alternative potential applications.

Acknowledgements

This research was supported by the Spanish DPI2017-87743-R grant from the Ministerio de Economia, Industria y Competitividad of Spain. This study has been also carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the Investments for the future Program IdEx Bordeaux (ANR-10-IDEX-03-02, HL-MRI Project) and Cluster of excellence CPU and TRAIL (HR-DTI ANR-10-LABX-57). The authors gratefully acknowledge the support of NVIDIA Corporation with their donation of the TITAN X GPU used in this research.

References

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis, 12 , 26–41.

Ayachi, R., Afif, M., Said, Y., & Atri, M. (2018). Strided convolution instead of max pooling for memory efficiency of convolutional neural networks. In International conference on the Sciences of Electronics, Technologies of In- formation and Telecommunications (pp. 234–243). Springer.

Coupé, P., Tourdias, T. , Linck, P., Romero, J. E. and Manjón J. V.. LesionBrain: An Online Tool for White Matter Lesion Segmentation. International Workshop on Patch-based Techniques in Medical Imaging (Patch-MI). Springer International Publishing, pp. 95-103. Springer, Cham, 2018.

Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., & C¸ ukur, T. (2019). Image synthesis in multi-contrast mri with conditional generative adversarial networks. IEEE transactions on medical imaging , 38 , 2375–2388.

Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., & Heng, P.-A. (2017). 3d deeply supervised network for automated segmentation of volumetric medical images. Medical image analysis, 41, 40–54.

Gonzales, R., & Fittes, B. (1977). Gray-level transformations for interactive image enhancement. Mechanism and Machine Theory , 12 , 111–122.

Iglesias, J. E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., & Fischl, B. (2013). Is synthesizing mri contrast useful for inter-modality analysis? In International Conference on Medical Image Computing and Computer- Assisted Intervention (pp. 631–638). Springer.

Jog, A., Carass, A., Roy, S., Pham, D. L., & Prince, J. L. (2017). Random forest regression for magnetic resonance image synthesis. Medical image analysis ,35 , 475–488.

Jog, A., Hoopes, A., Greve, D. N., Van Leemput, K., & Fis- chl, B. (2019). Psacnn: Pulse sequence adaptive fast whole brain segmentation. NeuroImage, 199 , 553-569.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2015). Deeply-supervised nets. In Artificial intelligence and statistics, 562–570.

Manjón, J. V. and Coupe, P. (2016). volbrain: an online mri brain volumetry system. Frontiers in neuroinformatics, 10, 30.

Miller, M. I., Christensen, G. E., Amit, Y., & Grenander, U. (1993). Mathe- matical textbook of deformable neuroanatomies. Proceedings of the National Academy of Sciences, 90, 11944–11948.

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering ,65 ,2720–2730.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, 234–241.

Roy, S., Carass, A., & Prince, J. L. (2013). Magnetic resonance image example- based contrast synthesis. IEEE transactions on medical imaging, 32, 2348– 2363.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: improved n3 bias correction. IEEE transactions on medical imaging, 29, 1310–1320.

Wen Wei, Emilie Poirion, Benedetta Bodini, Stanley Durrleman, Olivier Colliot, Bruno Stankoff, and Nicholas Ayache. 2019. Fluid-attenuated inversion recovery MRI synthesis from multisequence MRI using three-dimensional fully convolutional networks for multiple sclerosis. Journal of Medical Imaging, 6(1): 014005.