RS-Net: Regression-Segmentation 3D CNN for Synthesis of Full Resolution Missing Brain MRI in the Presence of Pathologies

xxxxxxxx, and xxxxxxxx

***Abstract*—Accurate synthesis of a full 3D MR image containing tumours from available MRI (e.g. to replace an image that is currently unavailable or corrupted) would provide a clinician as well as downstream inference methods with important complementary information for disease analysis. In this paper, we present an end-to-end 3D convolution neural network that takes a set of acquired MR image sequences (e.g. T1, T2, T1ce) as input and concurrently performs (1) regression of the missing full resolution 3D MRI (e.g. FLAIR) and (2) segmentation of the tumour into subtypes (e.g. enhancement, core). The hypothesis is that this would focus the network to perform accurate synthesis in the area of the tumour. Experiments on the BraTS 2015 and 2017 datasets [1] show that: (1) the proposed method gives better performance than state-of-the art methods in terms of established global evaluation metrics (e.g. PSNR), (2) replacing real MR volumes with the synthesized MRI does not lead to significant degradation in tumour and substructure segmentation accuracy. The system further provides uncertainty estimates based on Monte Carlo (MC) dropout [2] for the synthesized volume at each voxel, permitting quantification of the system's confidence in the output at each location.**

***Index Terms*—Deep Learning, Image Synthesis, Brain MRI, Pathologies, Brain Tumour, Multiple Sclerosis**

I. Introduction

THE presence of a variety of different Magnetic Resonance (MR) sequences (e.g. T1, T2, Fluid Attenuated Inverse Recovery (FLAIR)) improves the analysis in the context of neurological diseases such as multiple sclerosis and brain cancers, because different sequences provide complementary information. In particular, the accuracy of detection and segmentation of lesions and tumours greatly increases should several sequences of MR be available [3], as different sequences assist in differentiating healthy tissues from focal pathologies. However, in real clinical practice, not all MR image sequences are always available for each patient for a variety of reasons, including cost or time constraints, or at times, images are available but not usable, for example due to corruption from noise or patient motion. As such, both clinical practice and automatic segmentation techniques would benefit greatly from the synthesis of one or more of the missing 3D MR image sequences based on the others provided [4]. However, synthesis of full 3D brain MR image is challenging especially in the presence of pathology as different MR sequences represent pathology in a different way.

Recently, modality synthesis has gained some attention from the medical image analysis community [5], [6], [7]. Several approaches have been explored, such as patch-based random forest [5] and sparse dictionary reconstruction [6]. Regression Ensembles with Patch Learning for Image Contrast Agreement (REPLICA) [5] was developed to synthesize T2-weighted MRI from T1-weighted MRI using the bagged ensemble of random forests based on nonlinear patch regression. Given the success of Convolutional Neural Networks (CNNs) [8] and Generative Adversarial Networks (GANs) [9] for image-to-image translation in the field of computer vision, several recent 2D CNN [10], [7] and 2D GANs [11] have been developed for modality synthesis in the context of medical imaging, showing promising results for synthesis of healthy subject MRI. A patch-based Location Sensitive Deep Network (LSDN) [7] was developed to combine intensity and spatial information for synthesizing T2 MRI from T1 MRI and vice versa. A 2D CNN model was developed to generate 2D synthesized images with missing input MRI [10]. Quantitative analysis showed superior performance over competing methods based on global image metrics (PSNR and SSIM). However, the performance of the method in the area of focal pathology was not examined.

In this paper, an end-to-end 3D CNN is developed that takes as input a set of acquired MRI sequences of patients with tumours and simultaneously performs (1) regression to generate a full resolution missing 3D MR modality and (2) segmentation of the brain tumour into subtypes. The hypothesis is that by performing regression and segmentation concurrently, the network should produce full-resolution, high quality 3D MR images, particularly the area of the tumour. The network is trained and tested on the MICCAI 2015 and 2017 BraTS datasets [1], as well as a large multi-site, multi-scanner, proprietary dataset of MS patient MRI. In the first set of experiments, the framework is evaluated against state-of-the-art synthesis methods [5], [7], [10] based on global image metrics used in previous work [10], where it is shown to give better performance compared all reported results. The second set of experiments evaluate the synthesis quality at pathological locations, by examining its performance on subsequent independent downstream tasks, namely tumour segmentation. Results show that real MR images can be swapped with the generated synthesized T1, T2, and FLAIR MR images with minimal loss in tumour segmentation performance. The network also quantifies the uncertainty of the regressed synthetic volumes through Monte Carlo dropout [2]. This permits the confidence in the synthesis results to be conveyed to radiologists and clinicians and to automatic downstream methods that would use the synthesized volumes as inputs. In the last set of experiments, we also evaluate the ability of

xxxxxxxxxxxxxxx.

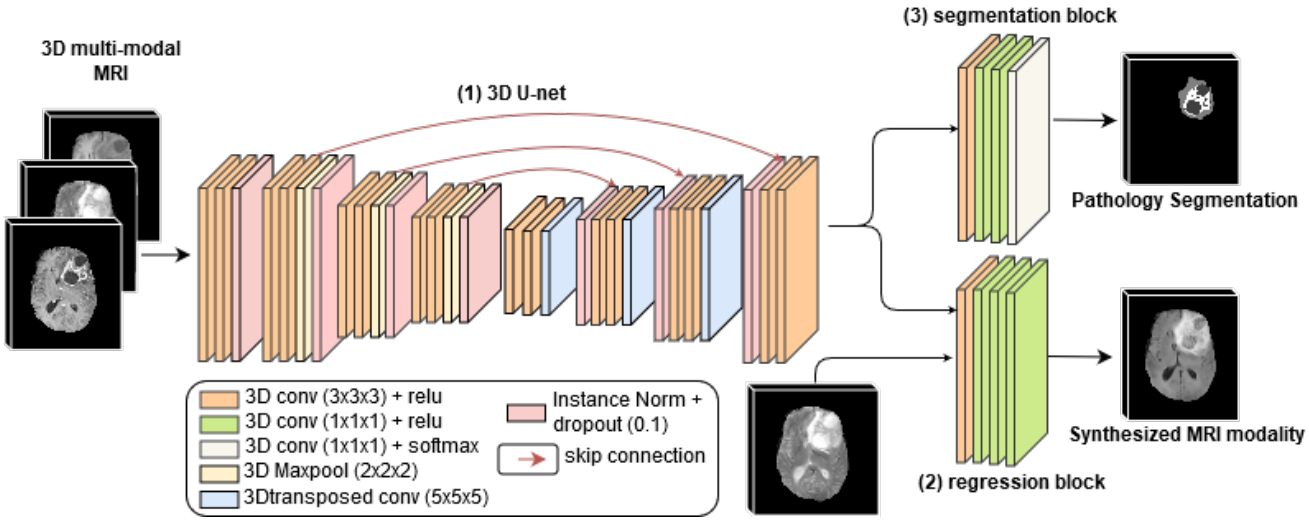

Fig. 1: Proposed Regression-Segmentation CNN architecture (RS-Net): (1) A 3D U-net, (2) Regression and (3) Segmentation convolution blocks. The model takes as input several full 3D MR image sequences, synthesizes the missing 3D MRI, while concurrently generating the multi-class segmentation of the tumour into sub-types.

RS-Net to synthesize missing modalities in case of Multiple Sclerosis (MS) patient MRIs. We evaluate the performance with a downstream MS T2 lesion segmentation/detection task. Results concur the findings reported for brain tumour segmentation task, and show that indeed missing modalities can be replaced by RS-Net synthesized modalities with minimal performance degradation.

II. Regression-Segmentation CNN Architecture

A flowchart of the proposed Regression-Segmentation CNN architecture (**RS-Net**) can be seen in Figure 1. The network consists of three main components: (1) a modified 3D U-net [12], (2) regression convolution block for synthesizing image sequence, and (3) segmentation convolution block for multi-class tumour segmentation. RS-Net takes as input full 3D volumes of all available sequences of a patient. The U-net generates an intermediate latent representation of the inputs which is provided to the regression and the segmentation convolution blocks. These then generate synthesis of the missing 3D MR image sequences and multi-class segmentation of tumours into sub-types, at the same resolution. The U-net learns latent representation which is common to both tumour segmentation and synthesis, with focus on high accuracy in the area containing tumour structures. In addition to the U-net output, the regression block is also provided with one of the input MRIs, which will provide necessary brain MR context to the regression block. The architecture details are now described.

The 3D U-net is similar to the one proposed in [12], with some modifications. The U-net consists of 4 resolution steps for both encoder and decoder paths. At the start, we use 2 consecutive 3D convolutions of size 3x3x3 with *k* filters, where *k* denotes the user-defined initial number of convolution filters. Each step in the encoder path consists of 2 3D convolutions of size 3x3x3 with $k * 2^n$ filters, where *n* denotes the U-net resolution step. This is followed by maxpooling of size 2x2x2. At the end of each encoder step, instance normalization [13] is applied, followed by dropout [14] with 0.1 probability. In the decoder path at each step, 3D transposed convolution of size 5x5x5 is applied, with 2x2x2 stride and $k * 2^n$ filters for the upsampling task. The output of the transposed convolution is concatenated with the corresponding output of the encoder path. This is, once again, followed by instance normalization and Dropout with 0.1 probability. Finally, 2 3D convolution of size 3x3x3 with $k * 2^n$ filters are applied. Rectified linear unit is chosen as a non-linearity function for every convolution layer.

Each of the segmentation and regression blocks contain 4 convolution layers. The first convolution layer is of size 3x3x3, and the rest are of size 1x1x1. The first three convolution layers have $k * 4$, $k * 2$ and k filters. In the regression block, the last layer has just 1 filter, while, for the segmentation block, there are C filters in the last layer, where C denotes the total number of classes for the segmentation task.

Weighted Mean Squared Error (W-MSE) loss is used for the synthesis task, and weighted Categorical Cross Entropy (W-CCE) loss for segmentation. Here, the weights are defined such that the weight increases whenever there are fewer voxels in a particular class.

$$\text{W-CCE}^i = -\sum_l w_l \sum_n y^i_{n,l} \log p^i_{n,l} \tag{1}$$

$$\text{W-MSE}^i = \sum_n w^i_n * (x^i_n - \hat{x}^i_n)^2 \tag{2}$$

$$w^i_n = w_l * y^i_n \qquad \text{where, } w_l = \left(\frac{\sum_{k=0}^{k=C} m_k}{m_l}\right) * r^{ep} + 1, \tag{3}$$

where, y^i_n, p^i_n, x^i_n, $\hat{x}^i_n$, and w^i_n denote true label, predicted label, true voxel values, predicted voxel value, and the weight for voxel n of volume i, respectively. w_l denotes the weight of class l. m_l is total number of voxels of l^{th} class in the

T2	REPLICA [5]	LSDN [7]	2D-CNN [10]	RS-Net (proposed)
SSMI	0.901 ±0.01	0.909 ±0.02	0.929 ±0.17	**0.934 ±0.02**
PSNR	28.62 ±1.69	30.12 ±1.62	30.96 ±1.85	**31.13 ±1.78**

FLAIR	REPLICA [5]	LSDN [7]	2D-CNN [10]	RS-Net (proposed)
SSMI	0.870 ±0.01	0.887 ±0.01	0.897 ±0.01	**0.900 ±0.01**
PSNR	28.32 ±1.38	29.68 ±1.56	30.32 ±1.61	**30.88 ±1.84**

TABLE I: Quantitative results (mean ± std) for T1-to-T2 (top) and T1-to-FLAIR (bottom) synthesis based on PSNR and SSIM. Higher values indicate better performance. Absolute highest performing results seen in bold.

training dataset. w_l are decayed over each epoch ep with a rate of $r \in [0, 1]$. It should be noted that w_l converges to 1 as ep becomes large. The final loss function for the network, L^i, (for volume i) is a weighted combination of both of these loss functions:

$$L^i = \lambda_1(\text{W-MSE}^i) + \lambda_2(\text{W-CCE}^i). \quad (4)$$

Given the challenges associated with regressing a synthesized volume, errors are bound to exist. As such, deterministic outputs present dangers to subsequent clinical decisions as well as to downstream automatic methods that make use of the results. In this work, the network output is augmented with uncertainty estimates based on Monte Carlo dropout [2]. During testing, N Monte Carlo (MC) samples of the output are acquired by passing each set of input volumes N times through the network to predict N different synthesized output MR volumes with probability of randomly dropping any neuron of the network equal to the dropout rate. Uncertainty in the synthesized volume, during testing, is estimated based on the variance of the MC samples at every voxel.

III. Experiments and Results

We now evaluate the performance of the RS-Net using two sets of experiments. In the first set of experiments, we compare the quality of the synthesized volume generated by RS-Net against other methods [10], [5], [7] using PSNR and SSIM on 2015 MICCAI BraTS dataset [1]. In the second set of experiments, we evaluate the quality of the synthesized volumes in a downstream task of tumor segmentation on 2017 MICCAI BraTS datasets [1].

RS-Net uses 4 initial convolutional filters and 4 steps for U-net encoder and decoder paths.This results in a network with a total of 674455 learnable parameters. Values of λ_1 and λ_2 in the loss function (Eq. 4), to combine CCE and MSE, were fixed to 1.0 and 0.1 respectively based on experimentation evidence. The networks were trained on a NVIDIA Titan Xp GPU for 240 epochs. Approximate training time was 3 days. The networks were trained with batch size of 1, using Adam optimizer [15] with the following hyperparameters: learning rate $= 0.0002$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-08}$. During testing time, a total of 20 samples of the output were generated to estimate the uncertainty in the synthesized volumes.

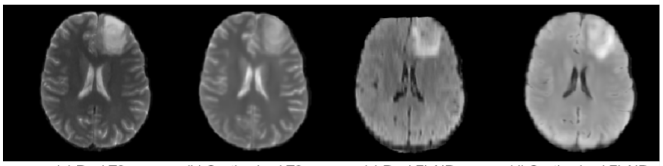

Fig. 2: Example slice from synthetic MR volumes generated by the proposed RS-Net on BraTS 2015 dataset for T1-to-T2 and T1-to-FLAIR synthesis.

A. Comparison of RS-Net synthesis results against other methods

In order to compare the quality of the synthesized volumes produced by RS-Net against other state-of-the-art methods, namely REPLICA [5], LSDN [7], and 2D CNN [10], we train two different RS-Nets for T2 and FLAIR synthesis from T1 MRI, as done by Chartsias et al. [10]. We use the evaluation metrics, SSIM [16] and PSNR, defined in [10], to evaluate the quality of the synthesized volumes.

Given a ground-truth volume X and its corresponding synthesized volume $\hat{X}$, SSIM is computed as

$$SSIM(X, \hat{X}) = \frac{(2\mu_X\mu_{\hat{X}} + c_1)(2\sigma_{X\hat{X}} + c_2)}{(\mu_X^2 + \mu_{\hat{X}}^2 + c_1)(\sigma_X^2 + \sigma_{\hat{X}}^2 + c_1)} \quad (5)$$

where μ_X and σ_X^2 are mean and variance of volume X and $\sigma_{X\hat{X}}$ is the covariance between X and $\hat{X}$.

PSNR is computed as

$$PSNR(X, \hat{X}) = 10\log_{10}(\frac{MAX_I^2}{MSE}) \quad (6)$$

where MAX_I is the maximum intensity of the volume and MSE is the mean squared error between volumes X and $\hat{X}$.

In order to compare our results to those in the paper [10], experiments were performed on the 2015 MICCAI BraTS training dataset [1]. This dataset consists of High-Grade Glioma (HGG) and Low-Grade Glioma (LGG) cases. 54 LGG cases were acquired with T1, T2, T1ce, and FLAIR. Four tumour sub-classes were defined. Volumes are skull-stripped, co-registered, and interpolated to $1mm^3$ voxel dimension. Each volume is of size 240 x 240 x 155. We follow the same pre-processing steps followed in [10], where we normalize each volume by dividing by the volume's average intensity. Following [10], we perform 5-fold cross validation on the dataset (LGG cases). Here, for each cross-validation fold, the dataset is divided into three sets, namely, training, validation,

	T1	T2	FLAIR	T1ce	DE	DT	DC
Real	✓	✓	✓	✓	**68.2 ±31.0**	**87.9 ±09.8**	**75.7 ±23.1**
T1 Synthesis	⊙	✓	✓	✓	67.6 ±31.2	87.9 ±09.8	75.5 ±23.1
T2 Synthesis	✓	⊙	✓	✓	66.3 ±32.1	87.3 ±11.4	75.6 ±23.6
FLAIR Synthesis	✓	✓	⊙	✓	66.8 ±31.8	83.6 ±10.7	73.1 ±24.7
T1ce Synthesis	✓	✓	✓	⊙	24.8 ±20.2	87.3 ±10.0	54.0 ±19.9

TABLE II: Comparison of multi-class brain tumour segmentation based on S-Net on the BraTS 2017 Validation dataset. The results using all 4 real MRI volumes are compared against replacing 1 real MRI volume with a synthesized MRI volume produced by RS-Net. Notation: Real MR volume (✓), and synthesized MR volume using RS-Net (⊙). Quantitative segmentation results based on Dice coefficients (mean ± std) for: enhancing tumor (DE), whole tumor (DT), and tumor core (DC). Higher values indicate better performance.

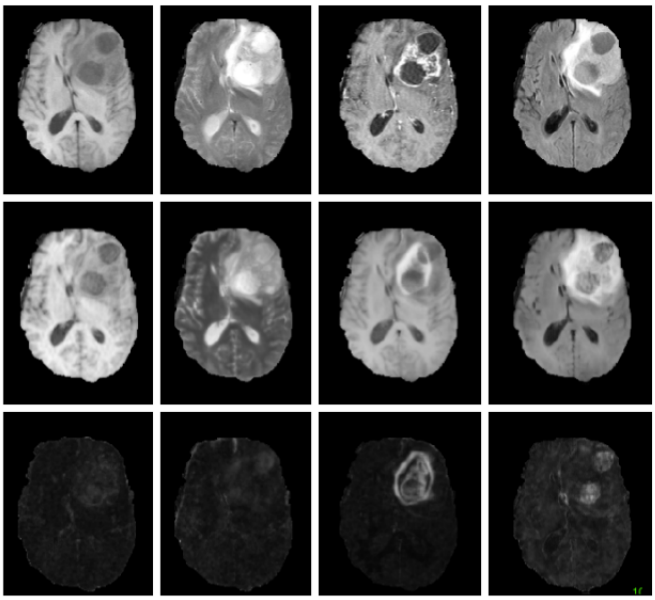

Fig. 3: Example slice from synthetic MR volumes generated using the proposed RS-Net along with its associated uncertainties. Real MRI (Row 1); synthesized volumes (Row 2) and its associated uncertainty (Row 3) produced as mean and variance across 20 MC dropout samples. Columns from left to right: T1, T2, T1ce, and FLAIR. Notice that uncertainties are highest where predicted tumour enhancements in T1ce are incorrect.

and testing. Each set consists of 42, 6, and 6 volumes respectively.

Quantitative comparison of all different methods is given in Table I. It should be noted that we didn't reproduce the results for other methods and instead report them as listed in [10]. Results indicate that RS-Net performs slightly better than other methods based on the global metrics of PSNR and SSIM, for both T1-to-T2 and T1-to-FLAIR synthesis. The results also show the advantage of using the proposed 3D CNN over 2D CNN. An example showing qualitative results based on RS-Net for both T2 and FLAIR synthesis on a testing volume is shown in Figure 2. Note that the resulting MR images are visually similar to the real images, particularly in the area of the tumour.

B. Evaluation of RS-Net synthesis results on downstream tumour segmentation task

The metrics used in the previous section can be useful in assessing global synthesis quality, but in the context of volumes with pathological structures such as lesions or tumours synthesis quality assessment should focus on the pathological areas. To this end, we quantitatively evaluate the synthesis performance based on their effect on downstream method, tumour segmentation and tumour sub-class segmentation. To this end, we train a new segmentation CNN, for the specific task of multi-class tumor segmentation (referred to as **S-Net**). This network is similar to the RS-Net but modified such that the synthesis convolution block is removed. S-Net is trained using all 4 real MR volumes with weighted CCE as the loss function. To evaluate the quality of the synthesized volume, one of the real MR volumes is swapped with the synthesized one and the segmentation accuracy is measured. Note that we do not retrain the S-Net with the synthesized volume. This allows us to measure quality of the synthesized volumes in comparison to the real volumes.

1) Dataset and Pre-processing:: The 2017 MICCAI BraTS [1] datasets were used for all the experiments in this section. The BraTS training dataset was used to train the networks. This dataset is comprised of 210 HGG and 75 LGG patients with T1, T1 post contrast (T1ce), T2, and FLAIR MRI for each patient, along with expert tumor labels for each of 3 classes: edema, necrotic/non-enhancing core, and enhancing tumor core. 228 volumes were randomly selected for training the network and another remaining 57 for network validation. A separate BraTS 2017 validation dataset, held out during training, was used to test the synthesis and segmentation performance. This dataset contains 46 patient multi-channel MRI (with no labels provided). The BraTS challenge provided pre-processed volumes that were skull-stripped, co-aligned, and resampled to 1 mm^3 voxel volume. The intensities were additionally rescaled using mean subtraction, divided by the standard deviation, and rescaled from 0 to 1 and were cropped to 184 x 200 x 152. For this context, the additional complementary input presented to the regression block (see Figure 1(3)) for T1, T2, T1ce, and FLAIR sequences were T1ce, FLAIR, T1, and T2 respectively. This was chosen as T1ce is the gadolinium enhanced version of T1, and FLAIR is the fluid attenuated version of T2.

	T1	T2	FLAIR	T1ce	DE	DT	DC
Real	✓	✓	✓	✓	**68.2 ±31.0**	**87.9 ±09.8**	**75.7 ±23.1**
T1 Synthesis	⊙	✓	✓	✓	67.6 ±31.2	87.9 ±09.8	75.5 ±23.1
	●	✓	✓	✓	67.5 ±31.3	87.8 ±09.9	75.3 ±23.3
T2 Synthesis	✓	⊙	✓	✓	66.3 ±32.1	87.3 ±11.4	75.6 ±23.6
	✓	●	✓	✓	66.1 ±32.0	87.2 ±11.9	75.4 ±23.8
FLAIR Synthesis	✓	✓	⊙	✓	66.8 ±31.8	83.6 ±10.7	73.1 ±24.7
	✓	✓	●	✓	62.9 ±33.3	81.3 ±17.4	71.5 ±25.8
T1ce Synthesis	✓	✓	✓	⊙	24.8 ±20.2	87.3 ±10.0	54.0 ±19.9
	✓	✓	✓	●	24.1 ±22.1	85.9 ±11.0	53.9 ±23.4

TABLE III: Comparison of multi-class brain tumour segmentation results based on S-Net on the BraTS 2017 Validation dataset, where each real MR input volume is replaced by its corresponding synthesized MR volume generated by either RS-Net or R-Net in a leave-one-out fashion. Notation: Real MR volume (✓), synthesized MR volume using RS-Net (⊙), and R-Net (●). Quantitative segmentation results based on Dice coefficients (mean ± std) for: enhancing tumor (DE), whole tumor (DT), and tumor core (DC). Higher values indicate better performance.

2) Qualitative Evaluation:: Synthesis MR volumes produced in a leave-one-out approach by 4 different RS-Nets such that three real MR sequences are used to synthesize the fourth (see Figure 3). The results indicate that the network is able to produce high-quality, high-resolution, 3D synthesized MR volumes, particularly for T1 and T2 sequences, and even for FLAIR. As T1ce shows enhancement within the tumour based on injection of a contrast agent, it was not expected to be easily synthesized from other sequences and error resulted. However, the system indicates locations where the network is uncertain about the regressed output. Qualitative results indicate that errors within the tumour enhancement have associated relatively high uncertainties. This suggests that these uncertainties can be communicated to a clinician or radiologist to indicate trustworthy regions of the synthesized images, and that automatic downstream methods using the synthesized volumes can focus computations on the areas of high confidence, which should be explored in future work.

3) Replacing real with synthetic MRI Volumes:: In Table II, we compare the tumour segmentation using S-Net in two different testing scenarios, (i) all 4 real MR volumes are provided as input and (ii) 1 real MR volume is replaced with synthesized MR volume for each sequence generated by RS-Net, in turn. We train 4 different RS-Nets to synthesize 4 MR image sequences, where 3 real sequences are presented as input to RS-Net to synthesize the fourth. The synthesized MR volume, along with the 3 real corresponding MR volumes, were then presented to the S-Net previously trained on all four real MRIs. This will allow us to measure quality of the synthesized volume in comparison to the real volume. The resulting labels for BraTS 2017 validation set were uploaded to the BraTS Challenge server, where quantitative segmentation results were provided based on the Dice coefficients for: whole tumor, enhancing tumor, and tumor core. These results (Table II) indicate that by swapping out real MR volumes with the synthesized T1 or T2 MR volumes generated by the RS-Net leads to comparable brain tumour segmentation performance based on all three reported Dice metrics. For the slightly harder problem of FLAIR synthesis, results indicate a small degradation in tumour segmentation performance for all three Dice metrics. T1ce synthesis results in no loss of whole tumour segmentation performance, but, as predicted, led to a significant reduction in performance in terms of enhancement and necrotic core. This was expected as T1ce is a challenging MRI to synthesize due to its reliance on a contrast agent, which is not used by any other MR sequences.

4) Effectiveness of combined Regression-Segmentation task:: RS-Net has two output streams for synthesis and segmentation tasks. To check how RS-Net performs in comparison to a network which is trained only for the task of synthesis, we train a new network (**R-Net**) which is similar to RS-Net but modified such that the segmentation block is removed as well as the additional input to the regression block, and training is based only on weighted MSE. R-Net was trained for the synthesis of all 4 MR image sequences separately, in a leave-one-out approach, and tested for tumor segmentation using S-Net on the BraTS validation dataset exactly as described above. From Table III, we can observe that R-Net performs comparably to RS-Net, when T1 and T2 are synthesized but shows a small degradation in performance for FLAIR and T1ce synthesis on all three Dice metrics. This shows that performing synthesis and segmentation together allows the network to focus more on tumour part, and in turn gives better quality of the synthesized volume, especially for FLAIR and T1ce.

5) Performance of Segmentation part of RS-Net:: One of the advantages of the RS-Net is that, in addition to MRI synthesis, it also provides tumour segmentation labels. In this section, we will analyze this segmentation part of RS-Net (Figure 1 (2)). Table IV indicates that the segmentation performance based on RS-Net directly is lower than the results based on using all 4 real MR volumes in S-Net, but is generally lower in comparison to the segmentation results when synthesized MR volumes generated by RS-Net is used in place of a real MR volumes. This trend is consistent across all MR image sequences for all three Dice metrics, except for FLAIR where the enhancing and core tumour Dice is

	T1	T2	FLAIR	T1ce	DE	DT	DC
Real	✓	✓	✓	✓	**68.2 ±31.0**	**87.9 ±09.8**	**75.7 ±23.1**
T1 Synthesis	⊙	✓	✓	✓	67.6 ±31.2	87.9 ±09.8	75.5 ±23.1
	×	✓	✓	✓	66.4 ±33.0	85.2 ±15.3	71.0 ±27.4
T2 Synthesis	✓	⊙	✓	✓	66.3 ±32.1	87.3 ±11.4	75.6 ±23.6
	✓	×	✓	✓	66.5 ±32.3	87.0 ±10.6	71.1 ±28.4
FLAIR Synthesis	✓	✓	⊙	✓	66.8 ±31.8	83.6 ±10.7	73.1 ±24.7
	✓	✓	×	✓	69.0 ±31.0	81.7 ±15.1	72.4 ±28.8
T1ce Synthesis	✓	✓	✓	⊙	24.8 ±20.2	87.3 ±10.0	54.0 ±19.9
	✓	✓	✓	×	23.1 ±19.8	86.5 ±10.8	52.0 ±20.8

TABLE IV: Comparison of multi-class brain tumour segmentation results based on S-Net against the results generated directly from the segmentation module of RS-Net for the BraTS 2017 Validation dataset. Notation: Real MR volume (✓), synthesized MR volume using RS-Net (⊙), and segmentation output of RS-Net without MR volume (×). Quantitative segmentation results based on Dice coefficients (mean ± std): enhancing tumor (DE), whole tumor (DT), and tumor core (DC). Higher values indicate better performance.

higher for segmentation directly from the RS-Net over the segmentation results from S-Net with a synthesized input (for unknown reasons).

C. Evaluation of RS-Net synthesis results for MS patient MRIs

MS is a chronic, inflammatory demyelinating disease of the central nervous system with presently no known cure. The presence of lesions in MRI is one of the hallmarks of MS. As a result, MRI has been used for diagnosis and to monitor disease progression and treatment efficacy. Similar to brain tumours, segmentation of T2 lesion, which is useful for staging MS patients, requires availability of multiple MR sequences like FLAIR, T2, T2, PDw etc. In particular FLAIR or T2 MR images are routinely used for visualization and segmentation of T2 lesion as they appear hyperintense in FLAIR/T2 images. In this section, we validate the usefulness of RS-Net by synthesizing FLAIR or T2 images from other modalities available, and check its effectiveness by evaluating it on a downstream T2 lesion segmentation/detection task.

We train two different RS-Net to synthesize FLAIR and T2 MR sequence from the other available MR sequence (T1,T2,PDw for FLAIR synthesis and T1,FLAIR,PDw for T2 synthesis). We train a S-Net on all 4 real MR sequences and at test time replace one of them (FLAIR or T2) with the synthesized one. This allows us to measure quality of the synthesized volumes in comparison to the real volumes. Similar to Sec.III-B, we compare this against R-Net.

The method was evaluated on a proprietary, multi-site, multi-scanner, clinical trial dataset of 1064 Relapsing-Remitting MS (RRMS) patients, scanned annually over a 24-month period. T1, T2, FLAIR, and PDW MRI sequences were acquired at a 1mm x 1mm x 3mm resolution and pre-processed with brain extraction, N3 bias field inhomogeneity correction, Nyul image intensity normalization, and registration to the MNI-space. Ground truth T2 lesion segmentation masks were provided with the data. These were obtained using a proprietary approach where the result of an automated segmentation method was manually corrected by expert human annotators. All networks (RS-Net/R-Net/S-Net) were trained on 65% of the subjects, with 17.5% held out for validation and 17.5% for testing.

Since the downstream outcome of interest is the accurate detection of T2 lesions, we evaluate the performance of networks based on lesion-level True Positive Rate (TPR) and False Detection Rate (FDR). To obtain lesion-level detections from the voxel-based segmentations, a connected component analysis is performed to group lesion voxels together in an 18-connected neighbourhood. A true positive (TP) lesion is detected when the segmentation, including its 18-connected neighbourhood, overlaps with at least three, or more than 50%, of the ground truth lesion voxels. Insufficient overlap results in a false negative (FN), and candidate lesions of 3 or more voxels that do not overlap with a ground truth lesion are counted as false positives (FP). The TPR ($= \frac{TP}{TP+FN}$) and FDR ($= \frac{FP}{FP+TP}$) are then calculated at the lesion level and are used to plot receiver operating characteristic (ROC) curves. Given that MS lesions vary greatly in size, the system performance is evaluated on lesions grouped into three size bins: small (3-10 vox), medium (11-50 vox), and large (51+ vox).

Quantitative evaluation (ROC curve of TPRvsFDR) of RS-Net against R-Net for FLAIR and T2 synthesis by replacing real MR sequence with synthesized MR sequence in S-Net is given in Fig:4 and Fig:5. From these figures we can see that RS-Net performs better compared to R-Net for all lesions. This also holds true for all individual lesion size ROC curves. Value of TPR at 0.2 FDR (the clinical operating point of interest) is given in Table:**??**. From this table, we can see that RS-Net synthesized MR sequences (FLAIR or T2) consistently gives better performance compared to R-Net synthesized MR sequence for all lesion size. This shows that performing synthesis and segmentation together gives better performance compared to only synthesizing the missing MR sequences.

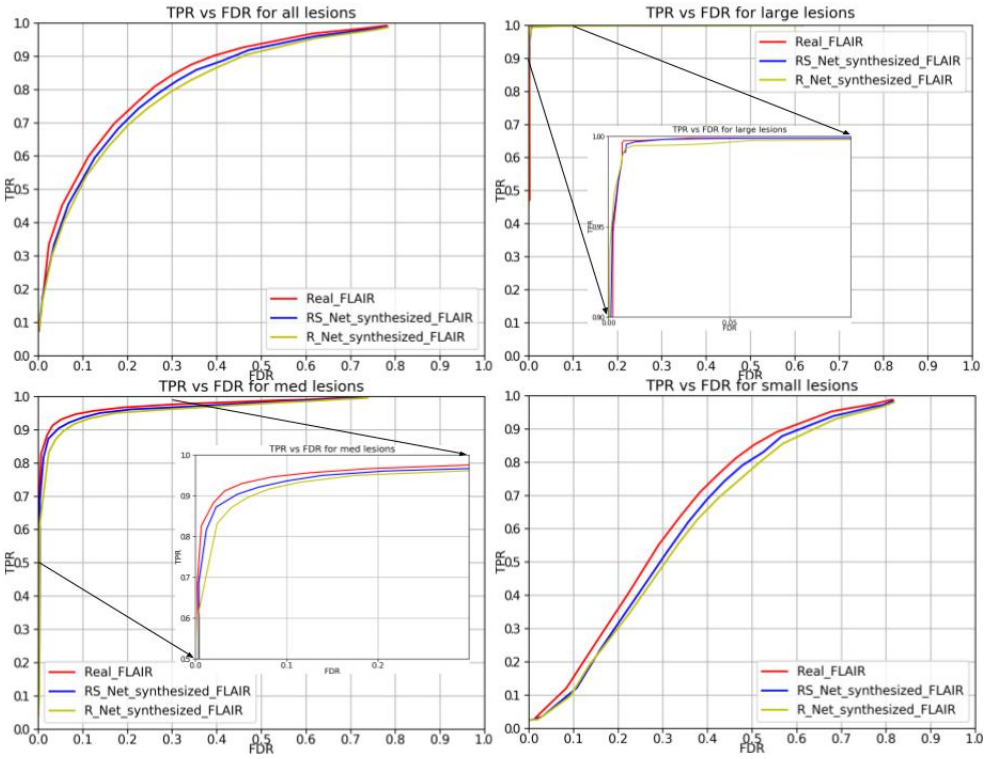

Fig. 4: Comparison of T2 lesion detection results based on S-Net (Red) for FLAIR synthesis, where FLAIR MR input image is replaced by its corresponding synthesized MR volume generated by either RS-Net (Blue) or R-Net (Yellow). Here, Receiver-operating characteristic (ROC) curves are plotted, illustrating TPR (true positive rate) vs. FDR (false detectionrate) across all lesions (Top Left), large lesions (Top Right), medium lesions (Bottom Left) and small lesions (Bottom Right).

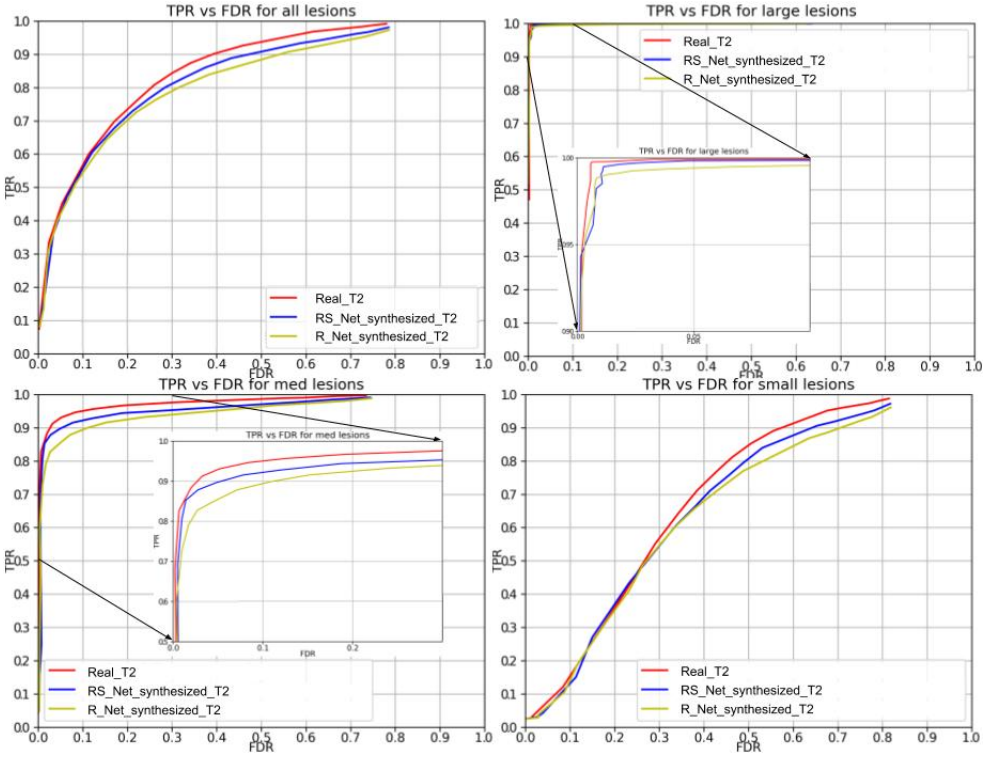

Fig. 5: Comparison of T2 lesion detection results based on S-Net (Red) for T2 synthesis, where T2 MR input image is replaced by its corresponding synthesized MR volume generated by either RS-Net (Blue) or R-Net (Yellow). Here, Receiver-operating characteristic (ROC) curves are plotted, illustrating TPR (true positive rate) vs. FDR (false detectionrate) across all lesions (Top Left), large lesions (Top Right), medium lesions (Bottom Left) and small lesions (Bottom Right).

	FLAIR synthesis				T2 synthesis			
	All	**Large**	**Med.**	**Small**	**All**	**Large**	**Med.**	**Small**
All Real sequences (4)	0.740	0.999	0.970	0.360	0.740	0.999	0.970	0.360
3 Real + 1 R-Net synthesized sequences	0.695	0.998	0.952	0.300	0.705	0.990	0.925	0.350
3 Real + 1 RS-Net synthesized sequences	0.715	0.999	0.960	0.315	0.720	0.998	0.945	0.365

TABLE V: Comparison of TPR at 0.2 FDR for different lesions sizes for RS-Net synthesized and R-Net synthesized MR sequences (FLAIR and T2) against Real sequences.

IV. Conclusions

In this paper, a full resolution 3D end-to-end CNN was developed for the task of MR volume synthesis in the presence of brain tumours. The network was trained for the concurrent tasks of synthesizing a missing MRI sequence and tumour sub-tissue segmentation. Experimental results on BraTS 2015 challenge dataset indicated that the proposed method outperforms all previous methods in terms of traditional evaluation metrics like PSNR and SSIM. The quality of the synthesized images was further evaluated by assessing their effects on the performance in independent tumour segmentation experiments. Experiments on the BraTS 2017 challenge dataset indicated that multi-task learning helps in synthesizing high quality volumes over synthesis alone particularly in more challenging contexts (i.e. FLAIR and T1ce). Evaluation on downstream segmentation/detection task for brain tumour / Multiple Sclesions patient indicated that real MRIs can be replaced with synthesized T1, T2, and FLAIR volumes with minimum degradation in segmentation accuracy, whereas synthesizing T1ce is still too challenging. However, uncertainty measure based on Monte Carlo dropout was shown to be helpful in communicating the confidence in the synthesis results, which will be essential for their adoption by clinicians and downstream automatic methods. The code for the proposed method is available here: https://github.com/RagMeh11/RS-Net.

References

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[2] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.

[3] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "Hemis: Hetero-modal image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 469–477.

[4] G. Van Tulder and M. de Bruijne, "Why does synthesized data improve multi-sequence classification?" in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 531–538.

[5] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Medical image analysis*, vol. 35, pp. 475–488, 2017.

[6] S. Roy, A. Carass, and J. Prince, "A compressed sensing approach for mr tissue contrast synthesis," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 371–383.

[7] H. Van Nguyen, K. Zhou, and R. Vemulapalli, "Cross-domain synthesis of medical images using efficient location-sensitive deep network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 677–684.

[8] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *European conference on computer vision*. Springer, 2016, pp. 649–666.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[10] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris, "Multimodal mr synthesis via modality-invariant latent representation," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 803–814, 2017.

[11] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep mr to ct synthesis using unpaired data," in *International workshop on simulation and synthesis in medical imaging*. Springer, 2017, pp. 14–23.

[12] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[13] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.