

Whole Brain Segmentation and Labeling from CT Using Synthetic MR Images

Can Zhao^{1(✉)}, Aaron Carass¹, Junghoon Lee², Yufan He¹, and Jerry L. Prince¹

¹ Department of Electrical and Computer Engineering,
The Johns Hopkins University, Baltimore, MD 21218, USA
czhao20@jhu.edu

² Department of Radiation Oncology, The Johns Hopkins School of Medicine,
Baltimore, MD 21287, USA

Abstract. To achieve whole-brain segmentation—i.e., classifying tissues within and immediately around the brain as gray matter (GM), white matter (WM), and cerebrospinal fluid—magnetic resonance (MR) imaging is nearly always used. However, there are many clinical scenarios where computed tomography (CT) is the only modality that is acquired and yet whole brain segmentation (and labeling) is desired. This is a very challenging task, primarily because CT has poor soft tissue contrast; very few segmentation methods have been reported to date and there are no reports on automatic labeling. This paper presents a whole brain segmentation and labeling method for non-contrast CT images that first uses a fully convolutional network (FCN) to synthesize an MR image from a CT image and then uses the synthetic MR image in a standard pipeline for whole brain segmentation and labeling. The FCN was trained on image patches derived from ten co-registered MR and CT images and the segmentation and labeling method was tested on sixteen CT scans in which co-registered MR images are available for performance evaluation. Results show excellent MR image synthesis from CT images and improved soft tissue segmentation and labeling over a multi-atlas segmentation approach.

Keywords: Synthesis · MR · CT · Deep learning · CNN · FCN U-net · Segmentation

1 Introduction

Computed tomography (CT) imaging of the head has many clinical and scientific uses including visualization and assessment of head injuries, intracranial bleeding, aneurysms, tumors, headaches, and dizziness as well as for use in surgical planning. Yet due to the poor soft tissue contrast in CT images, magnetic resonance imaging (MRI) is almost exclusively used for localizing, characterizing, and labeling gray matter (GM) and white matter (WM) structures in the brain. Unfortunately, there are many scenarios in which only CT images are available—e.g., emergency situations, lack of an MR scanner, patient implants

or claustrophobia, and cost of obtaining an MR scan—and there is no approach to provide whole brain segmentation and labeling from these data.

There has been very limited work on GM/WM segmentation from CT images. A whole brain segmentation method for 4D contrast-enhanced CT based on a nonlinear support vector machine was recently published [12]. The authors point out that a key part of their method is the formation of a 3D image derived from all of the temporal acquisitions. The segmentation result is impressive, but it is not clear that their method will work on conventional 3D CT data. As well, their method only provides classification of GM, WM, and CSF and does not label the sub-cortical GM or cortical gyri. The authors of [12] provide an excellent summary of much of the previous work on GM/WM segmentation from non-contrast CT (cf. [6, 8, 10, 15]), and also point out the limitations of past approaches. It is clearly an area of investigation that deserves more research. In contrast to the situation in CT, GM/WM segmentation and labeling from MRI has been well studied and several excellent approaches exist (cf. [5, 9, 14, 18]). Thus, it is natural to wonder whether images that are synthesized from CT to look like MR images could be used for automatic segmentation and labeling; this is precisely what we propose.

Image synthesis methods provide intensity transformations between two image contrasts or modalities (cf. [1–3, 11, 17]). Previously reported image synthesis work has synthesized CT from MRI [1], T_2 -weighted (T_2 -w) from T_1 -weighted (T_1 -w) [3], and positron emission tomography (PET) from MRI [11]. In very recent work, Cao et al. [2] synthesized pelvic T_1 -w images from CT using a random forest and showed improvement in cross-modal registration. Some researchers have applied convolutional neural networks (CNNs) to synthesis (cf. [11]) yet Cao et al. [2] claimed that robust and accurate synthesis of MR from CT using a CNN is not feasible. We believe that because CNNs are resilient to intensity variations [4] and they can model highly nonlinear mappings, they are ideal for CT-to-MR synthesis. In fact, we demonstrate in this paper that such synthesis is indeed possible and that whole brain segmentation and labeling from these synthetic images is very effective.

2 Methods

Training and Testing Data. Twenty six patients had (T_1 -w) MR images acquired using a Siemens Magnetom Espree 1.5 T scanner (Siemens Medical Solutions, Malvern, PA) with geometric distortions corrected within the Siemens Syngo console workstation. The MR images were processed with N4 to remove any bias field and subsequently had their intensity scales adjusted to align their WM peaks. Contemporaneous CT images were obtained on a Philips Brilliance Big Bore scanner (Philips Medical Systems, Netherlands) under a routine clinical protocol for brain cancer patients treated with either stereotactic-body radiation therapy (SBRT) or radiosurgery (SRS). The CT images were resampled to have the same digital resolution as the MR images, which is $0.7 \times 0.7 \times 1$ mm. Then the MR images were rigidly registered to the CT images.

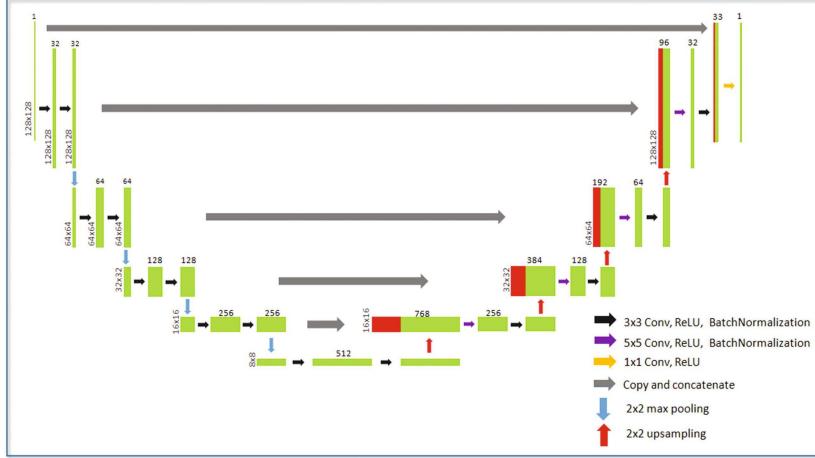


Fig. 1. Our modified U-net with four levels of contraction and expansion.

We use ten patient image pairs as training data for our CNN (see below). For each axial slice in the image domain, twenty-five 128×128 paired (CT and MR) image patches are extracted. The 128×128 patches can be thought of as subdividing the slice into a 5×5 grid with overlap between the image patches. These patch pairs are used to train an FCN based on a modified U-net [16] that will synthesize MR patches from CT patches. The synthetic MR patches are then used to construct an axial slice of the synthetic MR image. Our FCN, with 128×128 CT patches as input and 128×128 synthetic MR patches as output, is shown in Fig. 1.

FCN Algorithm for CT-to-MR Synthesis. The mapping between CT and MR is too nonlinear to be modeled accurately by the shallow features used in a random forest, which is why we explore a CNN based approach. As the mapping between CT and MR is dependent on anatomical structures, it makes intuitive sense that any CNN synthesis model should incorporate the ideas of semantic segmentation, for which fully convolutional networks (FCNs) were designed. Additionally, having already sacrificed some resolution in bringing the CT into alignment with MR, we want to be careful to not further degrade the image quality. Thus, we have selected as the basis of our FCN the U-net [16], which can achieve state-of-the-art performance for semantic segmentation and preserve high resolution information throughout the contraction-expansion layers of the network.

The encoder follows the typical architecture of a CNN. Each step contains two 3×3 convolutional layers, activated by a rectified linear unit (ReLU), and a 2×2 max pooling operation for downsampling. In the decoder, each step contains a 2×2 upsampling layer followed by a 5×5 convolutional layer and a

3×3 convolutional layer. The two convolutional layers are activated by ReLU. And the final layer is a 1×1 convolutional layer.

This FCN has four differences from the standard U-net.

Modification 1: the U-net decoder has two 3×3 layers, whereas we use one 5×5 layer and one 3×3 layer. We do this because the upsampling layer is simply repeating values in a 2×2 window. Thus, a 3×3 layer in the encoder can involve its eight connected neighbors, whereas a 3×3 layer after an upsampling layer only includes three-connected neighbors. By replacing this with a 5×5 layer, we can still involve all eight connected neighbors. There is a slight increase in the number of parameters to estimate, but the result has better accuracy.

Modification 2: CNN vision tasks benefit from increasing model depth; however, deeper models can have vanishing or exploding gradients [7]. In the original U-net, the decoder contains an upsampling layer, a convolutional layer, a layer merging it with high resolution representations, and another convolutional layer. Thus, the upsampled layer is convolved twice while the high resolution representation is convolved only once. We therefore exchange the order of the first convolutional layer and the merging layer so that both are convolved twice. With this change, we retain the same number of layers but our FCN can model greater non-linearity without introducing additional obstacles for back-propagation.

Modification 3: Every convolution loses border pixels; thus, the border of the predicted patch may not be as reliable as the center. The standard U-net crops each patch after each convolutional layer so that the predicted patch is smaller than the input patch. Our FCN keeps the boundary pixels instead of cropping them. However, when reconstructing a slice we use only the central 90×90 region of the image patches (except at the boundaries of the image, where we retain the side of the patch that touches the boundary).

Modification 4: U-net was used for solving segmentation, while synthesis is a regression task. That is, U-net only needed labels to distinguish edges, while we need to predict intensity values. Thus the batch normalization layers which are throughout U-net are a concern; there is no effect on image contrast but absolute intensity values are lost and CT numbers have a physical meaning. In order to include this information, we merge the original CT patches before the last convolutional layer. Also, U-net used softmax to activate the last layer for segmentation, while we use ReLU for regression.

Automatic Whole-Brain Segmentation and Labeling. We use MALP-EM [9] to provide whole-brain segmentation and labeling from the synthetic MR images. Since the synthetic MR images are naturally registered with the CT images, the result is a segmentation and labeling of the CT images. MALP-EM uses an atlas cohort of 30 subjects having both MR images and labels from the OASIS database [13]. These atlases are deformably registered to the target and the labels are combined using joint label fusion (JLF) [18]. Finally, these labels are adjusted using an intensity based EM method to provide additional robustness to pathology, especially traumatic brain injury. We used the code that has been made freely available by the original authors of the method.

3 Experiments and Results

Image Synthesis. Our FCN was trained on 45,575 128×128 image patch pairs derived from ten of the co-registered MR and CT images. It took two days to train and 1 min to synthesize one MR image from the input CT on a NVIDIA GPU GTX1070SC. Figures 2(a)–(c) show an example input CT image, the resulting synthetic $T1-w$, and the ground truth $T1-w$, respectively.

Experiment 1: MALP-EM Segmentation. We applied MALP-EM on both synthetic and ground truth $T1-w$ images. Figure 2(e) shows the segmentation result from the synthetic $T1-w$ in Fig. 2(b), while Fig. 2(f) shows the result from the ground truth $T1-w$ in Fig. 2(c). There are differences between the two results, but this is the first result showing such a detailed labeling of CT brain images. We compute Dice coefficients between segmentation results obtained using synthetic $T1-w$ and those obtained using the true $T1-w$. Here are mean Dice coefficients for a few brain structures. For hippocampus, they are 0.62 (right) and 0.59 (left); for precentral gyrus, they are 0.52 (right) and 0.55 (left); for postcentral gyrus, they are 0.51 (right) and 0.52 (left); and for caudate, they are

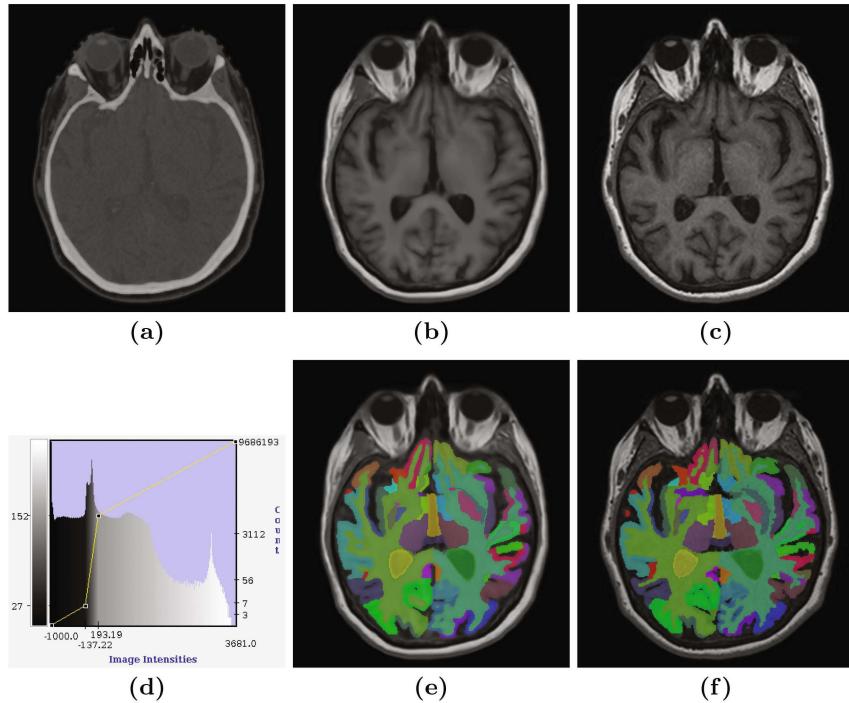


Fig. 2. For one subject, we show the (a) input CT image, the (b) output synthetic $T1-w$, and the (c) ground truth $T1-w$ image. (d) is the dynamic range of (a). Shown in (e) and (f) are the MALP-EM segmentations of the synthetic and ground truth $T1-w$ images, respectively.

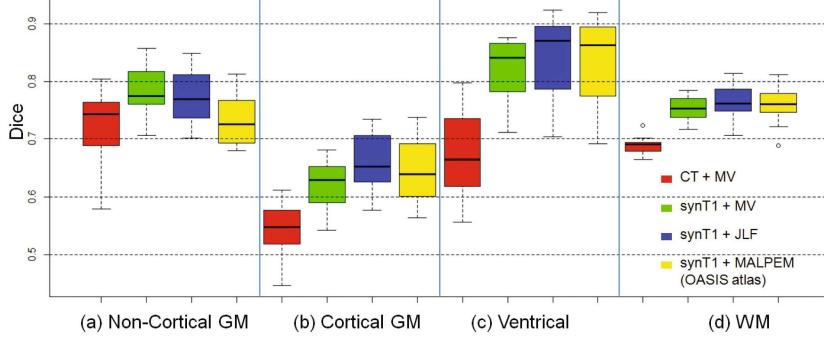


Fig. 3. With MALP-EM processing of the ground truth $T1-w$ as the reference, we compute the Dice coefficient between multi-atlas segmentations using either the subject CT images with MV label fusion (red), or synthetic $T1-w$ with MV or JLF, as the label fusion, and MALP-EM (yellow). Note that MALP-EM uses the OASIS atlas with manually delineated labels, while the other three use the remaining 15 images with MALP-EM computed labels from the true $T1-w$ images as atlases. (Color figure online)

0.70 (right) and 0.73 (left). After merging the labels, box plots of the Dice coefficients for four labels: non-cortical GM, cortical GM, ventricles, and WM, are shown in Fig. 3 (yellow).

Experiment 2: Comparison to Direct Multi-atlas Segmentation. To demonstrate the benefits of our approach, we carried out a set of algorithm comparisons. Ideally, we would like to evaluate how well our CT images could be labeled directly from the OASIS atlases; but there are no CT data associated with OASIS. Instead, we used the 16 subjects (which do not overlap the 10 subjects used to train our FCN) in a set of leave-one-out experiments and let the MALP-EM labels act as the “ground truth”. For each of the 16 subjects, we used the remaining 15 (having $T1-w$ and MALP-EM labels) as atlases. To mimic the desired experiment, we first carried out multi-modal registration from each of the 15 $T1-w$ atlases to the target CT using mutual information (MI) as the registration cost metric. Because this is a multi-modal registration task, JLF is not available to combine labels, so we used majority voting (MV) instead. We next computed a synthetic $T1-w$ image from the target CT image and registered each atlas to this target using mean squared error (MSE) as the registration metric. To provide a richer comparison, we combined these 15 labels using both MV and JLF.

Given these three leave-one-out results, we computed Dice coefficient on four labels: non-cortical GM, cortical GM, ventricles, and WM. The results are shown in Fig. 3 (using the red, green, and blue graphics). We can see that use of the synthetic $T1-w$ is significantly better than using the original CT images whether labels are combined with either MV or JLF. JLF seems to provide somewhat better performance.

4 Discussion and Conclusion

The synthetic images that we achieve with FCN are quite good visually as demonstrated by the single (typical) example shown here (Fig. 2(b)), visually much better than those shown in Cao et al. [2] (their Fig. 7). This speaks very well to the potential of the FCN architecture for estimating synthetic cross-modality images. Besides whole-brain segmentation and labeling, there are a host of other potential applications for these images.

A limitation of our evaluation is our lack of manual brain labels in a CT dataset, as it would be interesting to compare our approach with a top multi-atlas segmentation algorithm that would use only CT data. The fact that our method appears to perform worse than the straight multi-atlas results in Fig. 3 is because the MALP-EM result is using manually delineated OASIS labels to estimate automatically generated MALP-EM labels, whereas the other two approaches are estimating MALP-EM labels from MALP-EM atlases. In the future, a more thorough evaluation including a quantitative comparison with Cao et al. [2] is warranted.

Recent research on contrast-enhanced 4D CT brain segmentation achieves slightly higher mean Dice than ours, with 0.81 and 0.79 for WM and GM [12], compared to ours as 0.77 and 0.76. However, because their data was 4D CT, its combined 3D image probably has lower noise and also enables them to use temporal features which we do not have. Furthermore, theirs was a contrast CT study while ours is a non-contrast study.

In summary, we have used a modified U-net to synthesize *T1-w* images from CT, and then directly segmented the synthetic *T1-w* using either MALP-EM or a multi-atlas label fusion scheme. Our results show that using synthetic MR can significantly improve the segmentation over using the CT image directly. This is the first paper to provide GM anatomical labels on a CT neuroimage. Also, despite previous assertions that CT-to-MR synthesis is impossible from CNNs, we show that it is not only possible but it can be done with sufficient quality to open up new clinical and scientific opportunities in neuroimaging.

Acknowledgments. This work was supported by NIH/NIBIB under grant R01 EB017743.

References

1. Burgos, N., Cardoso, M.J., Thielemans, K., Modat, M., Pedemonte, S., Dickson, J., Barnes, A., Ahmed, R., Mahoney, C.J., Schott, J.M., et al.: Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies. *IEEE Trans. Med. Imag.* **33**(12), 2332–2341 (2014)
2. Cao, X., Yang, J., Gao, Y., Guo, Y., Wu, G., Shen, D.: Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. *Med. Image Anal.* (2017, in press)
3. Chen, M., Carass, A., Jog, A., Lee, J., Roy, S., Prince, J.L.: Cross contrast multi-channel image registration using image synthesis for MR brain images. *Med. Image Anal.* **36**, 2–14 (2017)

4. Dodge, S., Karam, L.: Understanding how image quality affects deep neural networks. In: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6. IEEE (2016)
5. Fischl, B.: Freesurfer. *Neuroimage* **62**(2), 774–781 (2012)
6. Gupta, V., Ambrosius, W., Qian, G., Blazejewska, A., Kazmierski, R., Urbanik, A., Nowinski, W.L.: Automatic segmentation of cerebrospinal fluid, white and gray matter in unenhanced computed tomography images. *Acad. Radiol.* **17**(11), 1350–1358 (2010)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Hu, Q., Qian, G., Aziz, A., Nowinski, W.L.: Segmentation of brain from computed tomography head images. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005, pp. 3375–3378. IEEE (2006)
9. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
10. Kemmling, A., Wersching, H., Berger, K., Knecht, S., Groden, C., Nölte, I.: Decomposing the hounsfield unit. *Clin. Neuroradiol.* **22**(1), 79–91 (2012)
11. Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S.: Deep learning based imaging data completion for improved brain disease diagnosis. In: Golland, P., Hata, N., Barillot, C., Horngger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 305–312. Springer, Cham (2014). doi:[10.1007/978-3-319-10443-0_39](https://doi.org/10.1007/978-3-319-10443-0_39)
12. Manniesing, R., Oei, M.T., Oostveen, L.J., Melendez, J., Smit, E.J., Platel, B., Sánchez, C.I., Meijer, F.J., Prokop, M., van Ginneken, B.: White matter and gray matter segmentation in 4D computed tomography. *Sci. Rep.* **7** (2017)
13. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**(9), 1498–1507 (2007)
14. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Işgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans. Med. Imag.* **35**(5), 1252–1261 (2016)
15. Ng, C.R., Than, J.C.M., Noor, N.M., Rijal, O.M.: Preliminary brain region segmentation using FCM and graph cut for CT scan images. In: 2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS), pp. 52–56. IEEE (2015)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Horngger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
17. Roy, S., Wang, W.T., Carass, A., Prince, J.L., Butman, J.A., Pham, D.L.: PET attenuation correction using synthetic CT from ultrashort echo-time MR imaging. *J. Nuclear Med.* **55**(12), 2071–2077 (2014)
18. Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. *IEEE Trans. Patt. Anal. Mach. Intell.* **35**(3), 611–623 (2013)