Deep Learning for Multi-task Medical Image Segmentation in Multiple Modalities

Pim Moeskops[1,2](✉), Jelmer M. Wolterink[1], Bas H.M. van der Velden[1],
Kenneth G.A. Gilhuijs[1], Tim Leiner[3], Max A. Viergever[1], and Ivana Išgum[1]

[1] Image Sciences Institute, University Medical Center Utrecht,
Utrecht, The Netherlands
pim@isi.uu.nl

[2] Medical Image Analysis, Eindhoven University of Technology,
Eindhoven, The Netherlands

[3] Department of Radiology, University Medical Center Utrecht,
Utrecht, The Netherlands

Abstract. Automatic segmentation of medical images is an important task for many clinical applications. In practice, a wide range of anatomical structures are visualised using different imaging modalities. In this paper, we investigate whether a single convolutional neural network (CNN) can be trained to perform different segmentation tasks.

A single CNN is trained to segment six tissues in MR brain images, the pectoral muscle in MR breast images, and the coronary arteries in cardiac CTA. The CNN therefore learns to identify the imaging modality, the visualised anatomical structures, and the tissue classes.

For each of the three tasks (brain MRI, breast MRI and cardiac CTA), this combined training procedure resulted in a segmentation performance equivalent to that of a CNN trained specifically for that task, demonstrating the high capacity of CNN architectures. Hence, a single system could be used in clinical practice to automatically perform diverse segmentation tasks without task-specific training.

Keywords: Deep learning · Convolutional neural networks · Medical image segmentation · Brain MRI · Breast MRI · Cardiac CTA

1 Introduction

Automatic segmentation is an important task in medical images acquired with different modalities visualising a wide range of anatomical structures. A common approach to automatic segmentation is the use of supervised voxel classification, where a classifier is trained to assign a class label to each voxel. The classical approach to supervised classification is to train a classifier that discriminates between tissue classes based on a set of hand-crafted features. In contrast to this approach, convolutional neural networks (CNNs) automatically extract features

P. Moeskops and J.M. Wolterink—Equally contributed.

S. Ourselin et al. (Eds.): MICCAI 2016, Part II, LNCS 9901, pp. 478–486, 2016.
DOI: 10.1007/978-3-319-46723-8_55

that are optimised for the classification task at hand. CNNs have been successfully applied to medical image segmentation of e.g. knee cartilage [11], brain regions [1,10], the pancreas [12], and coronary artery calcifications [18]. Each of these studies employed CNNs, but problem-specific optimisations with respect to the network architecture were still performed and networks were only trained to perform one specific task.

CNNs have not only been used for processing of medical images, but also for natural images. CNN architectures designed for image classification in natural images [7] have shown great generalisability for divergent tasks such as image segmentation [13], object detection [3], and object localisation in medical image analysis [17]. Hence, CNN architectures may have the flexibility to be used for different tasks with limited modifications.

In this study, we first investigate the feasibility of using a single CNN *architecture* for different medical image segmentation tasks in different imaging modalities visualising different anatomical structures. Secondly, we investigate the feasibility of using a single *trained instance* of this CNN architecture for different segmentation tasks. Such a system would be able to perform multiple tasks in different modalities without problem-specific tailoring of the network architecture or hyperparameters. Hence, the network recognises the modality of the image, the anatomy visualised in the image, and the tissues of interest. We demonstrate this concept using three different and potentially adversarial medical image segmentation problems: segmentation of six brain tissues in brain MRI, pectoral muscle segmentation in breast MRI, and coronary artery segmentation in cardiac CT angiography (CTA).

2 Data

Brain MRI – 34 T_1-weighted MR brain images from the OASIS project [9] were acquired on a Siemens Vision 1.5 T scanner, as provided by the MICCAI challenge on multi-atlas labelling [8][1]. The images were acquired with voxel sizes of $1.0 \times 1.0 \times 1.25\,\text{mm}^3$ and resampled to isotropic voxel sizes of $1.0 \times 1.0 \times 1.0\,\text{mm}^3$. The images were manually segmented, in the coronal plane, into 134 classes that were, for the purpose of this paper, combined into six commonly used tissue classes: white matter, cortical grey matter, basal ganglia and thalami, ventricular cerebrospinal fluid, cerebellum, and brain stem.

Breast MRI – 34 T_1-weighted MR breast images were acquired on a Siemens Magnetom 1.5 T scanner with a dedicated double breast array coil [16]. The images were acquired with in-plane voxel sizes between 1.21 and 1.35 mm and slice thicknesses between 1.35 and 1.69 mm. All images were resampled to isotropic voxel sizes corresponding to their in-plane voxel size. The pectoral muscle was manually segmented in the axial plane by contour drawing.

[1] https://masi.vuse.vanderbilt.edu/workshop2012.

Cardiac CTA – Ten cardiac CTA scans were acquired on a 256-detector row Philips Brilliance iCT scanner using 120 kVp and 200–300 mAs, with ECG-triggering and contrast enhancement. The reconstructed images had between 0.4 and 0.5 mm in-plane voxel sizes and 0.45/0.90 mm slice spacing/thickness. All images were resampled to isotropic 0.45 × 0.45 × 0.45 mm^3 voxel size. To set a manual reference standard, a human observer traversed the scan in the craniocaudal direction and painted voxels in the main coronary arteries and their branches in the axial plane.

3 Method

All voxels in the images were labelled by a CNN using seven different training experiments (Fig. 1).

3.1 CNN Architecture

For each voxel, three orthogonal (axial, sagittal, and coronal) patches of 51 × 51 voxels centred at the target voxel were extracted. For each of these three patches, features were determined using a deep stack of convolution layers. Each convolution layer contained 32 small (3 × 3 voxels) convolution kernels for a total of 25 convolution layers [14]. To prevent over- or undersegmentation of structures due to translational invariance, no subsampling layers were used. To reduce the number of trainable parameters in the network and hence the risk of over-fitting, the same stack of convolutional layers was used for the axial, sagittal and coronal patches.

The output of the convolution layers were 32 features for each of the three orthogonal input patches, hence, 96 features in total. These features were input to two subsequent fully connected layers, each with 192 nodes. The second fully

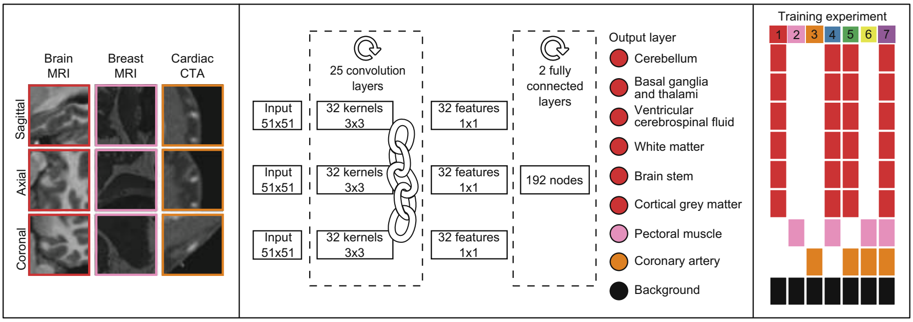

Fig. 1. Example 51 × 51 triplanar input patches (*left*). CNN architecture with 25 shared convolution layers, 2 fully connected layers and an output layer with at most 9 classes, including a background class common among tasks (*centre*). Output classes included in each training experiment (*right*).

connected layer was connected to a softmax classification layer. Depending on the tasks of the network, this layer contained 2, 3, 7, 8 or 9 output nodes. The fully connected layers were implemented as 1 × 1 voxel convolutions, to allow fast processing of arbitrarily sized images. Exponential linear units [2] were used for all non-linear activation functions. Batch normalisation [5] was used on all layers and dropout [15] was used on the fully connected layers.

3.2 Training Experiments

The same model was trained for each combination of the three tasks. In total seven training experiments were performed (Fig. 1, right): three networks were trained to perform one task (Experiments 1–3), three networks were trained to perform two tasks (Experiments 4–6), and one network was trained to perform three tasks (Experiment 7). The number of output nodes in the CNN was modified accordingly. In each experiment, background classes of the target tasks were merged into one class.

Each CNN was trained using mini-batch learning. A mini-batch contained 210 samples, equally balanced over the tasks of the network. For each task, the training samples were randomly drawn from all training images, balanced over the task-specific classes. All voxels with image intensity > 0 were considered samples. The network parameters were optimized using Adam stochastic optimisation [6] with categorical cross-entropy as the cost-function.

4 Experiments and Results

The data for brain MRI, breast MRI and cardiac CTA were split into 14/20, 14/20 and 6/4 training/test images, respectively. Four results were obtained for each task: one with a network trained for only that task, two with networks trained for that task and an additional task, and one with a network trained for all tasks together. Each network was trained with 25000 mini-batches per task.

No post-processing steps other than probability thresholding for evaluation purposes were performed. The results are presented on the full test set. In brain MRI, the voxel class labels were determined by the highest class activation. The performance was evaluated per brain tissue type, using the Dice coefficient between the manual and automatic segmentations. In breast MRI and cardiac CTA, precision-recall curve analysis was performed to identify the optimal operating point, defined, for each experiment, as the highest Dice coefficient over the whole test set. The thresholds at this optimal operating point were then applied to all images.

Figure 2 shows the results of the described quantitative analysis, performed at intervals of 1000 mini-batches per task. As the networks learned, the obtained Dice coefficients increased and the stability of the results improved. For each segmentation task, the learning curves were similar for all experiments. Nevertheless, slight differences were visible between the obtained learning curves. To assess whether these differences were systematic or caused by the stochastic

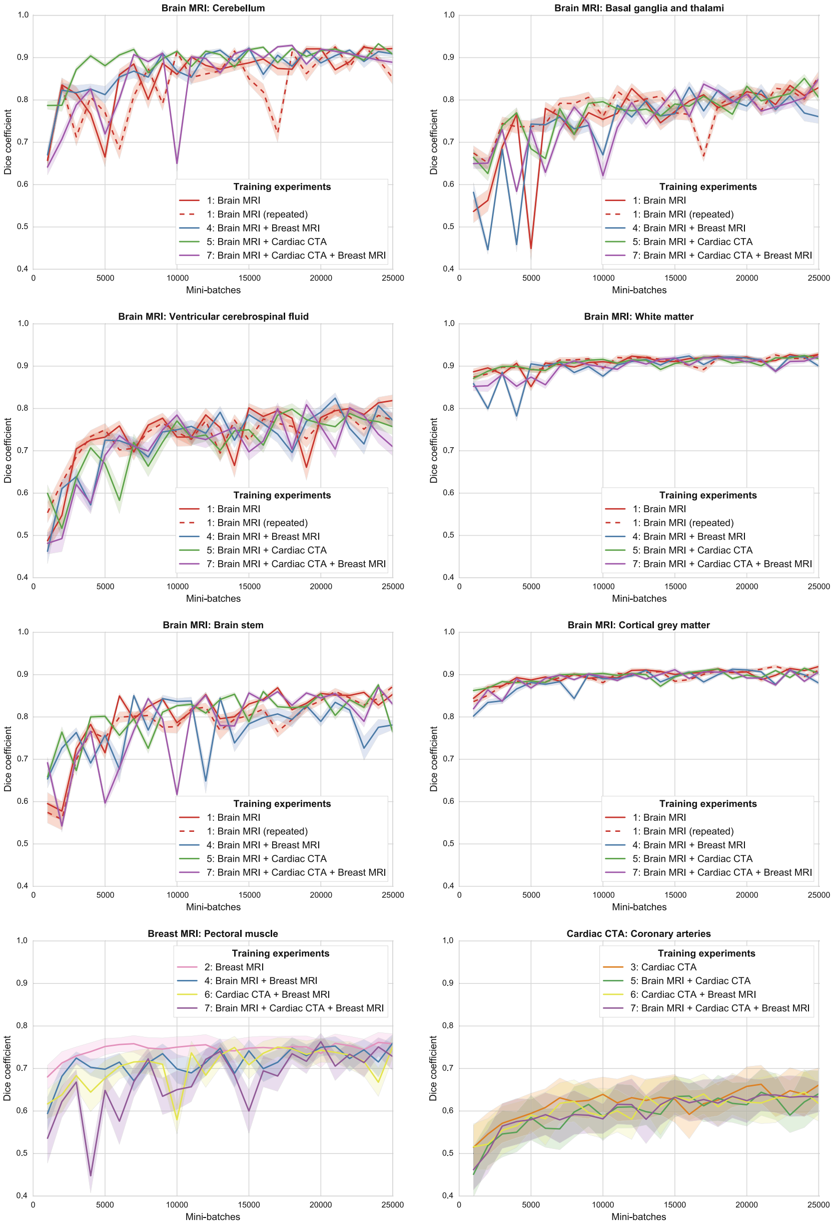

Fig. 2. Learning curves showing Dice coefficients for tissue segmentation in brain MRI (*top three rows*), breast MRI (*bottom left*), and cardiac CTA (*bottom right*), reported at 1000 mini-batch intervals for experiments including that task. The line colours correspond to the training experiments in Fig. 1.

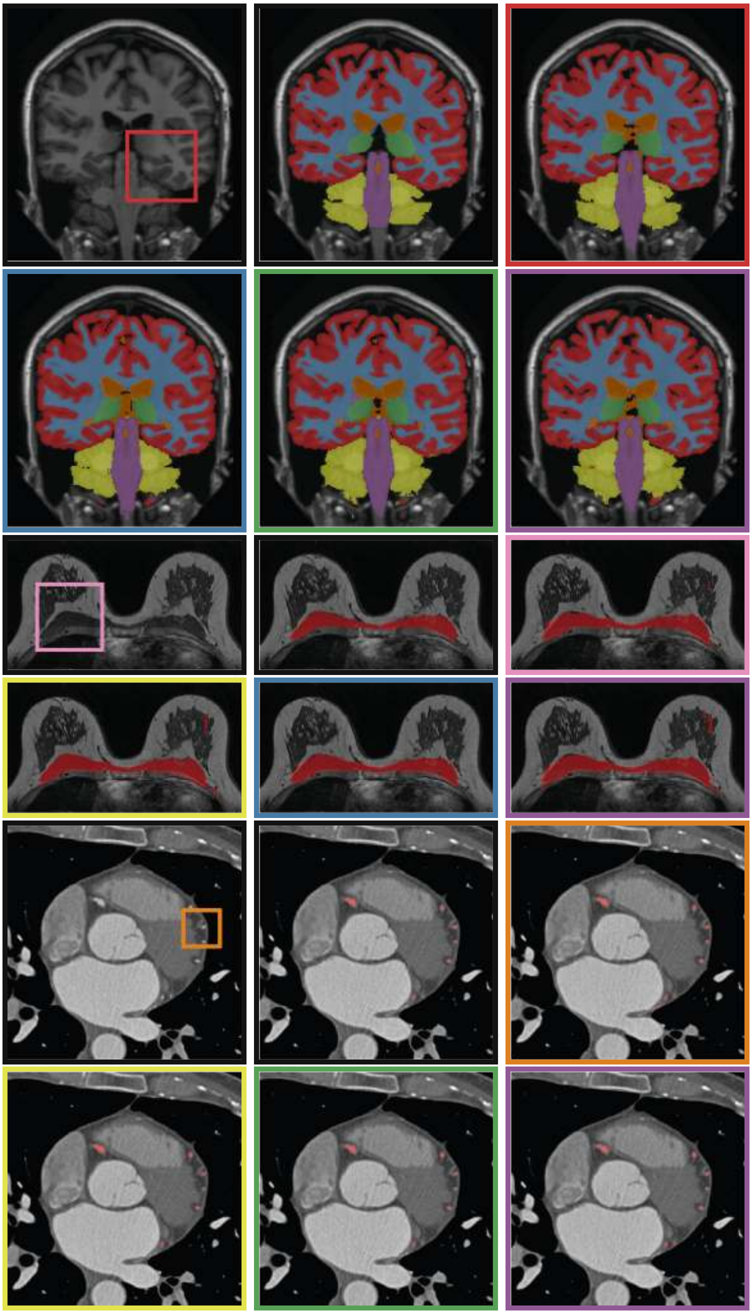

Fig. 3. Example segmentations for (*top to bottom*) brain MRI, breast MRI, and cardiac CTA. Shown for each task: (*left to right, first row*) image with an input patch as shown in Fig. 1, reference standard, segmentation by task-specific training, (*left to right, second row*) two segmentations by networks with an additional task, segmentation by a network combining all tasks. The coloured borders correspond to the training experiments in Figs. 1 and 2.

nature of CNN training, the training experiment using only brain MR data (Experiment 1) was repeated (dashed line in Fig. 2), showing similar inter-experiment variation. Figure 3 shows a visual comparison of results obtained for the three different tasks. For all three tasks, all four networks were able to accurately segment the target tissues.

Confusion between tasks was very low. For the network trained with three tasks, the median percentage of voxels per scan labelled with a class alien to the target (e.g. cortical grey matter identified in breast MR) was $<0.0005\,\%$ for all tasks.

5 Discussion and Conclusions

The results demonstrate that a single CNN architecture can be used to train CNNs able to obtain accurate results in images from different modalities, visualising different anatomy. Moreover, it is possible to train a single CNN instance that can not only segment multiple tissue classes in a single modality visualising a single anatomical structure, but also multiple classes over multiple modalities visualising multiple anatomical structures.

In all experiments, a fixed CNN architecture with triplanar orthogonal input patches was used. We have strived to utilise recent advances in deep learning such as batch normalisation [5], Adam stochastic optimisation [6], exponential linear units [2], and very deep networks with small convolution kernels [14]. Furthermore, the implementation of fully connected layers as 1×1 convolution layers and the omission of downsampling layers allowed fast processing of whole images compared with more time-consuming patch-based scanning [10–12,18]. The ability of the CNN to adapt to different tasks suggests that small architectural changes are unlikely to have a large effect on the performance. Volumetric 3D input patches might result in increased performance, but would require a high computational load due to the increased size of the network parameter space.

The results for brain segmentation are comparable with previously published results [10]. Due to differences in image acquisition and patient population, the obtained results for pectoral muscle segmentation and coronary artery extraction cannot be directly compared to results reported in other studies. Nevertheless, these results appear to be in line with previously published studies [4,19]. No post-processing other than probability thresholding for evaluation purposes was applied. The output probabilities may be further processed, or directly used as input for further analysis, depending on the application.

Including multiple tasks in the training procedure resulted in a segmentation performance equivalent to that of a network trained specifically for the task (Fig. 2). Similarities between the tasks, e.g. presence of the pectoral muscle in both breast MR and cardiac CTA, or similar appearance of brain and breast tissue in T_1-weighted MRI, led to very limited confusion. In future work, we will further investigate the capacity of the current architecture with more data and segmentation tasks, and investigate to what extent the representations within the CNN are shared between tasks.

References

1. de Brébisson, A., Montana, G.: Deep neural networks for anatomical brain segmentation. In: CVPR Bioimage Computing Workshop (2015)
2. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (ELUs). In: ICLR (2016)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 142–158 (2016)
4. Gubern-Mérida, A., Kallenberg, M., Martí, R., Karssemeijer, N.: Segmentation of the pectoral muscle in breast MRI using atlas-based approaches. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 371–378. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33418-4_46
5. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
6. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2015)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
8. Landman, B.A., Ribbens, A., Lucas, B., Davatzikos, C., Avants, B., Ledig, C., Ma, D., Rueckert, D., Vandermeulen, D., Maes, F., et al.: MICCAI 2012 Workshop on Multi-atlas Labeling. CreateSpace Independent Publishing Platform (2012). https://www.amazon.com/MICCAI-2012-Workshop-Multi-Atlas-Labeling/dp/1479126187
9. 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)
10. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Išgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans. Med. Imaging **35**(5), 1252–1261 (2016)
11. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 246–253. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40763-5_31
12. Roth, H.R., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E.B., Summers, R.M.: DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 556–564. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24553-9_68
13. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. (2016)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
15. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
16. van der Velden, B.H., Dmitriev, I., Loo, C.E., Pijnappel, R.M., Gilhuijs, K.G.: Association between parenchymal enhancement of the contralateral breast in dynamic contrast-enhanced MR imaging and outcome of patients with unilateral invasive breast cancer. Radiology **276**(3), 675–685 (2015)

17. de Vos, B., Wolterink, J., de Jong, P., Viergever, M., Išgum, I.: 2D image classification for 3D anatomy localization; employing deep convolutional neural networks. In: SPIE Medical Imaging, p. 97841Y (2016)
18. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 589–596. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24553-9_72
19. Zheng, Y., Loziczonek, M., Georgescu, B., Zhou, S.K., Vega-Higuera, F., Comaniciu, D.: Machine learning based vesselness measurement for coronary artery segmentation in cardiac CT volumes. In: SPIE Medical Imaging, p. 79621K (2011)