Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation

Mohsen Ghafoorian*[1,2,3], Alireza Mehrtash* [2,4], Tina Kapur [2], Nico Karssemeijer [1], Elena Marchiori [3], Mehran Pesteie [4], Charles R. G. Guttmann [2], Frank-Erik de Leeuw [5], Clare M. Tempany [2], Bram van Ginneken [1], Andriy Fedorov [2], Purang Abolmaesumi [4], Bram Platel [1], William M. Wells III[2]

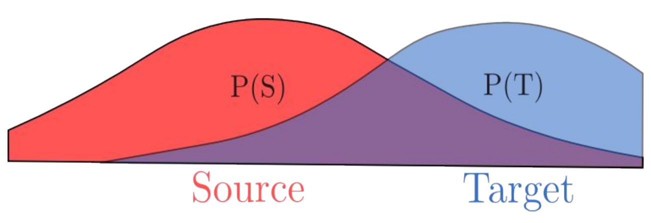
* equal contribution

[1] Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands,
[2] Radiology Department, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA,
[3] Institute for Computing and Information Sciences, Radboud University, Nijmegen, the Netherlands,
[4] Robotics and Control Laboratory, University of British Columbia, Vancouver, BC, Canada,
[5] Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands

Introduction

Motivations

- Variations in MRI acquisition protocols result in different appearances of normal and diseased tissue in the images.
- Convolutional neural networks (CNNs) are typically sensitive to the variations in imaging protocols. Therefore networks trained on data acquired with one MRI protocol do not perform satisfactorily on data acquired with different protocols.
- This limits the use of models trained with large annotated legacy datasets on a new dataset with a different domain which is a recurring situation in clinical settings.
- In this study, we aim to answer the following central questions regarding domain adaptation: Given a fitted legacy model,
  a. *How much data from the new domain is required for a decent adaptation of the original network?*
  b. *What portion of the pre-trained model parameters should be retrained given a certain number of the new domain training samples?*

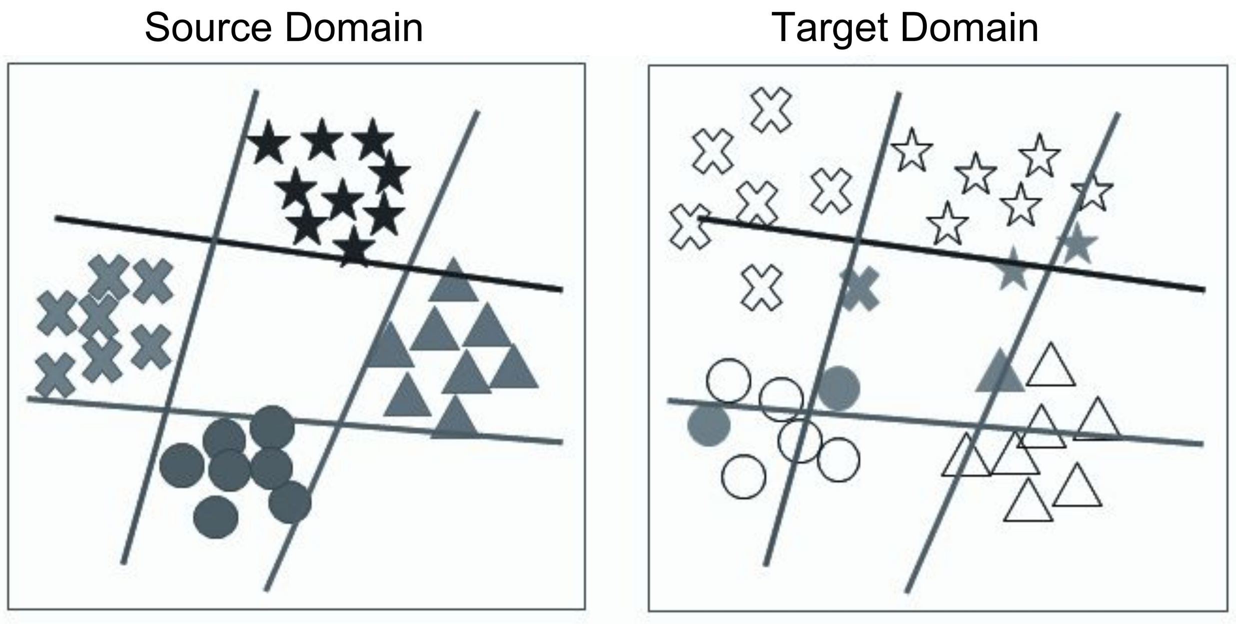


Contributions

- We trained a CNN on legacy MR images of brain and evaluated the performance of the domain-adapted network on the same task with images from a different domain.
- We then compared the performance of the model to the surrogate scenarios where either the same trained network is used or a new network is trained from scratch on the new dataset. The domain-adapted network tuned only by two training examples achieved a Dice score of 0.63, substantially outperforming a similar network trained on the same set of examples from scratch.

Domain Adaptation

- A domain D can be expressed by a feature space χ and a marginal probability distribution $P(X)$, where $X=\{x_1,\ldots,x_n\} \in \mathcal{X}$ [1].
- A supervised learning task on a specific domain $D=\{\mathcal{X}, P(X)\}$, consists of a pair of a label space Y and an objective predictive function $f(\cdot)$ (denoted by $T=\{Y, f(\cdot)\}$).
- The objective function $f(\cdot)$ can be learned from the training data, which consists of pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in Y$.
- After the training process, the learned model denoted by $\mathcal{F}(\cdot)$ is used to predict the label for a new instance x.
- Given a source domain DS with a learning task T_S and a target domain D_T with learning task T_T, transfer learning is defined as the process of improving the learning of the target predictive function $f_T(\cdot)$ in D_T using the information in D_S and TS, where $D_S \neq D_T$, or $T_S \neq T_T$.
- We denote $\mathcal{F}_{ST}(\cdot)$ as the predictive model initially trained on the source domain D_S, and domain-adapted to the target domain D_T.



A visualization of the effectiveness of transfer learning for domain adaptation. Solid shapes represent labeled samples. The trained models on a source domain might not be as effective on a target domain with a slightly different data distribution. A small number of labeled samples can be used to fine-tune the model.

Materials and Methods

Task: Segmentation of white matter hyperintensities on the RUNDMC [2] dataset

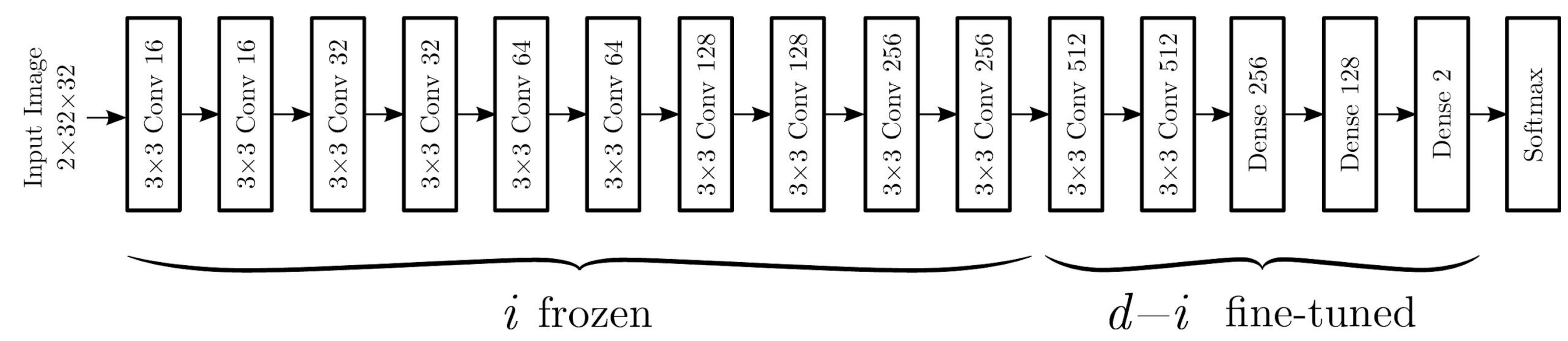
- Source domain: T1:1.0×1.0×1.0 mm, FLAIR: 1×1.2×6 mm
- Target domain: T1: 1.0×1.0×1.0 mm, FLAIR: 1×1.2×3 mm

Number of patients for the domain adaptation experiments:

	Source Domain			Target Domain		
set	train	validation	test	Train	Validation	Test
size	200	300	50	100	26	33

Training Procedure

Adam update rule, mini-batch size: 128, binary cross entropy loss, He initialization method, batch normalization, L2 regularization with decay factor: 0.0001, decaying learning rate, early stopping criterion, patch-based training, fully convolutional segmentation.

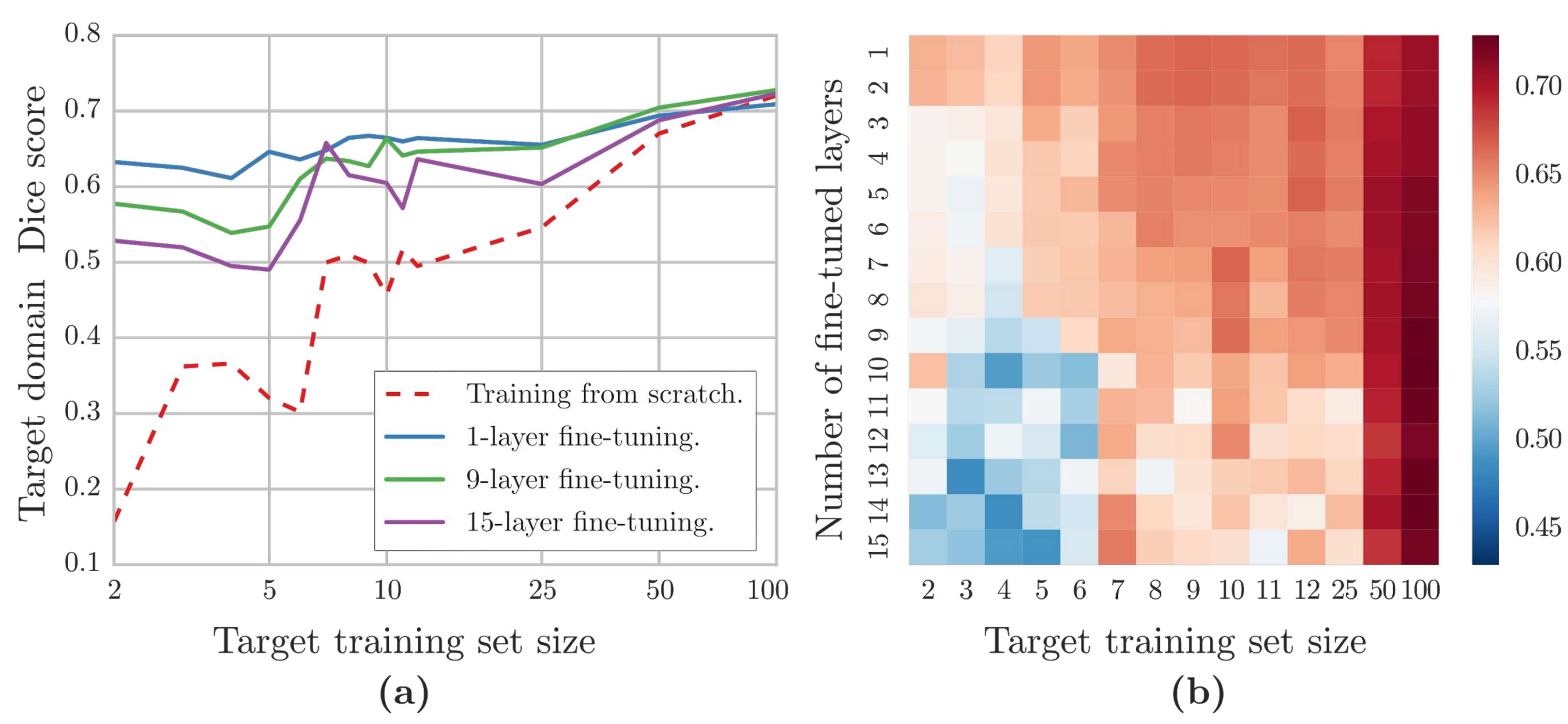


Experiments:

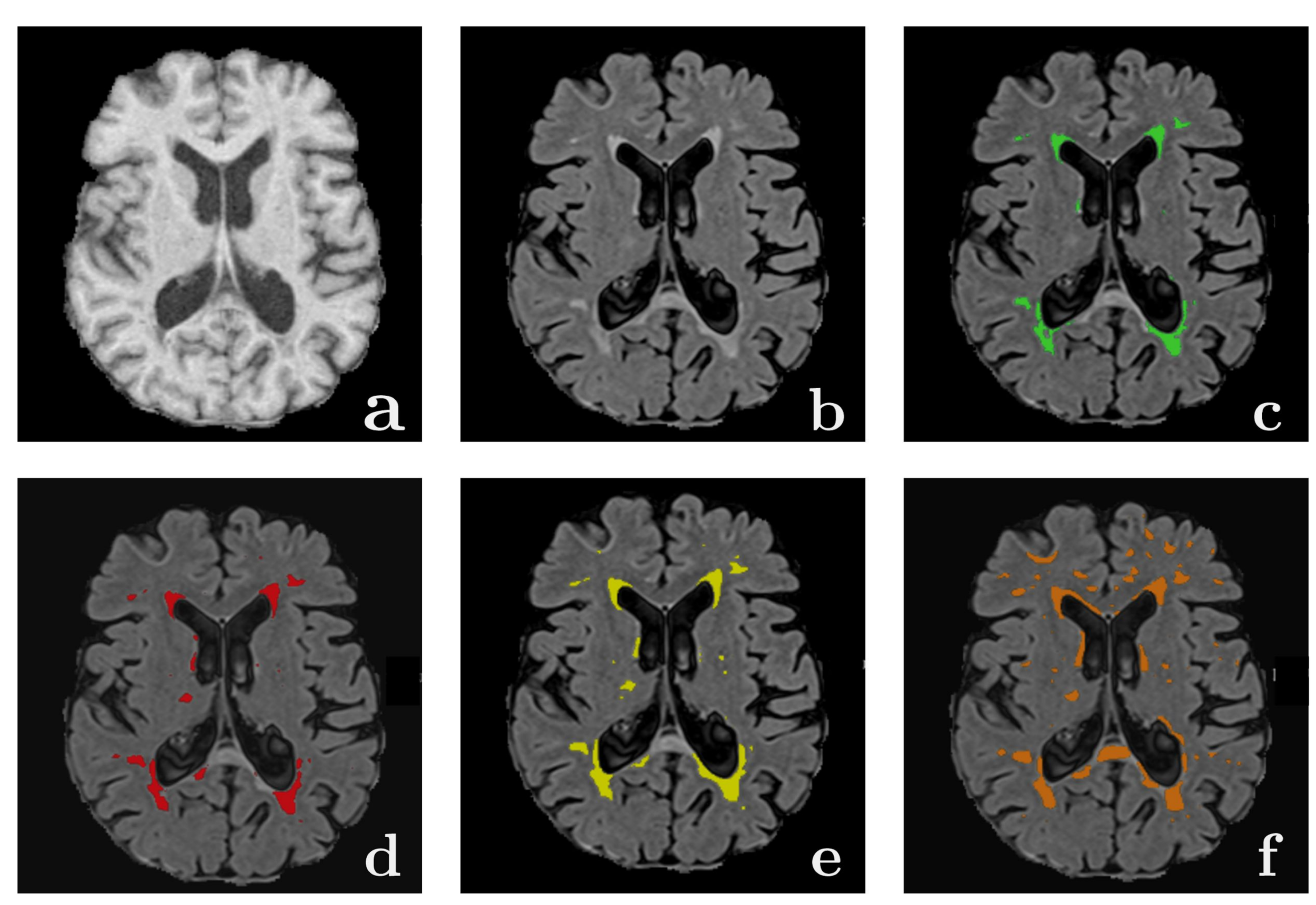
1. Source model ($f_S(.)$) directly on the target domain D_T
2. Target model ($f_T(.)$) on the target domain D_T
3. Fine-tuned source model ($f_{ST}(.)$) on target domain D_T
   - Different number of layers fine-tuned
   - Different number of labeled samples on the target domains

Results and Conclusions

- $f_S(\cdot)$ on 200 D_S cases: 0.76 Dice
- $f_S(\cdot)$ on D_T: 0.005 Dice
- The domain-adapted network tuned only by two training examples achieved a Dice score of 0.63 substantially outperforming a similar network trained on the same set of examples from scratch.



(a) The comparison of Dice scores on the target domain with and without transfer learning. A logarithmic scale is used on the x axis.
(b) Given a deep CNN with $d = 15$ layers, transfer learning was performed by freezing the initial layers and fine-tuning the last $d - i$ layers.



Examples of the brain WMH MRI segmentations. (a) Axial T1-weighted image. (b) FLAIR image. (c-f) FLAIR images with WMH segmented labels: (c) Reference (green) WMH. (d) WMH (red) from a domain adapted model $\mathcal{F}_{ST}(\cdot)$ fine-tuned on five target training samples. (e) WMH (yellow) from model trained from scratch $\mathcal{F}_T(\cdot)$ on 100 target training samples. (f) WMH (orange) from model trained from scratch $\mathcal{F}_T(\cdot)$ on 5 target training samples.

Discussion and Conclusions

- Using a small set of labeled images is sufficient for a decent performance on the target domain
- Fine-tune only the last few layers, when only a small set of labeled data is available
- Once a larger set of labeled data is available on the target domain, one can safely tine-tune more shallower layer.
- Fine-tuning the shallowest representations is rarely useful given their domain-independent characteristics.

References and Acknowledgements

[1] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering* 22.10 (2010): 1345-1359.
[2] van Norden, Anouk GW, et al. "Causes and consequences of cerebral small vessel disease. The RUN DMC study: a prospective cohort study. Study rationale and protocol." BMC neurology 11.1 (2011): 29.

Research reported in this publication was supported by NIH Grant No. P41EB015898, Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canadian Institutes of Health Research (CIHR) and a VIDI innovational grant from the Netherlands Organisation for Scientific Research (NWO, grant 016.126.351)