Representation learning for cross-modality classification

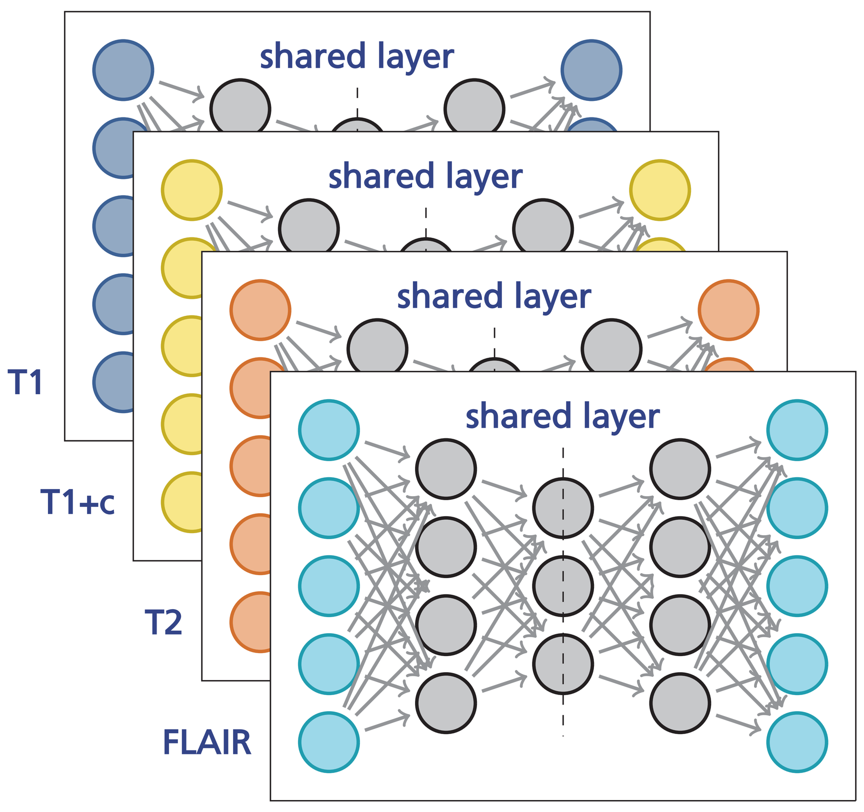
Gijs van Tulder and Marleen de Bruijne

Motivation

Differences in scanning parameters or modalities can complicate image analysis based on supervised classification. We present two representation learning approaches to learn representations that are similar across domains.

Model 1: Autoencoders

We train a separate autoencoder for each modality and add a similarity objective to minimise the representation difference across modalities.



The similarity objective compares the representation f_m for modality m with the mean over all modalities:

$$\mathcal{L}_{\text{sim},\,m} = \sum_{i=1}^{N} \left| f_m\left(\mathbf{x}_{m,i}\right) - \frac{1}{M} \sum_{m'=1}^{M} f_{m'}\left(\mathbf{x}_{m',i}\right) \right|$$

We add this to the standard learning objective of the autoencoders to get

$$\mathcal{L}_{\text{combined},\,m} = \alpha \mathcal{L}_{\text{sim},\,m} + (1-\alpha)\,\mathcal{L}_{\text{err},\,m}$$

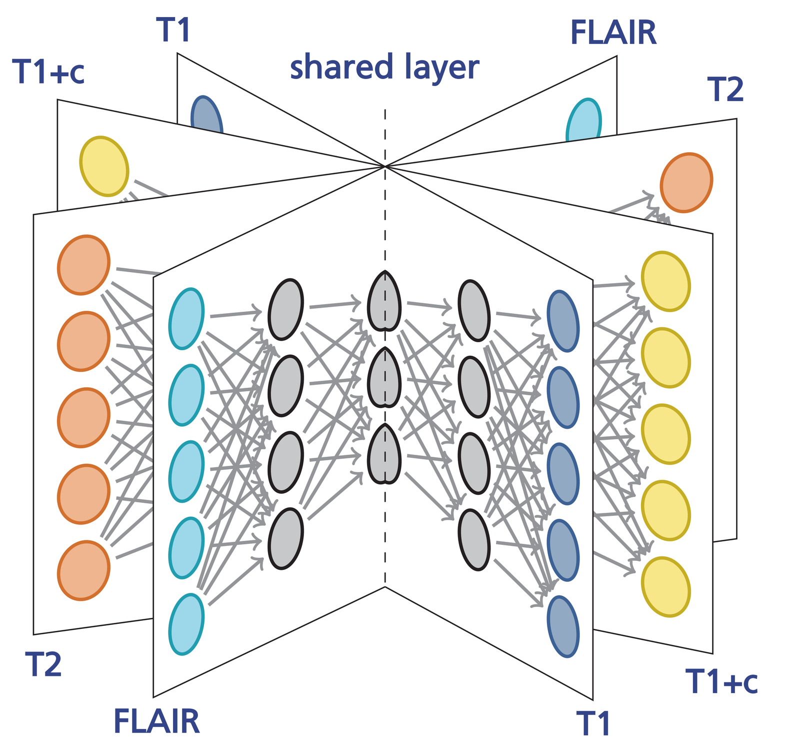
Conclusions

Learning a shared representation for data from different domains can improve classification accuracy.

With separate autoencoders for each domain, an additional similarity objective can be used to learn similar representations across domains.
An axial neural network combines the modalities in a single model, but still requires the similarity term.

Model 2: Axial neural network

We combine all modalities in a single model. There are different weights for each modality, but all modalities lead to the same shared representation.



At the shared layer, the incoming values from all M modalities are averaged into a shared representation:

$$\frac{1}{M} \sum_{m'=1}^{M} f_{m'}\left(\mathbf{x}_{m',i}\right)$$

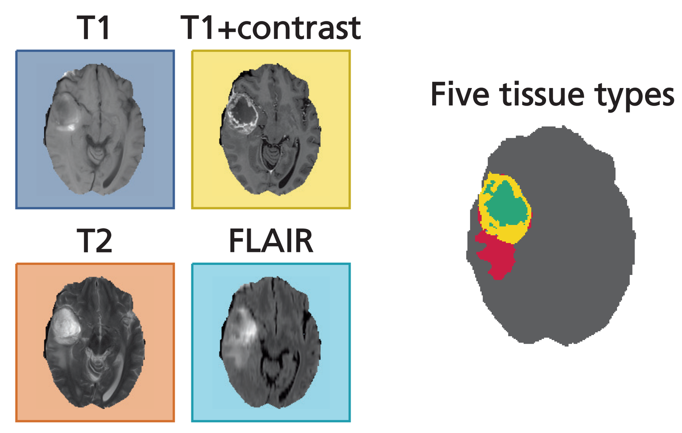
The learning objective includes an additional similarity term:

$$\mathcal{L}_{\text{sim}} = \sum_{i=1}^{N} \sum_{m=1}^{M} \left| f_m\left(\mathbf{x}_{m,i}\right) - \frac{1}{M} \sum_{m'=1}^{M} f_{m'}\left(\mathbf{x}_{m',i}\right) \right|$$

$$\mathcal{L}_{\text{combined}} = \alpha \mathcal{L}_{\text{sim}} + (1-\alpha)\,\mathcal{L}_{\text{err}}$$

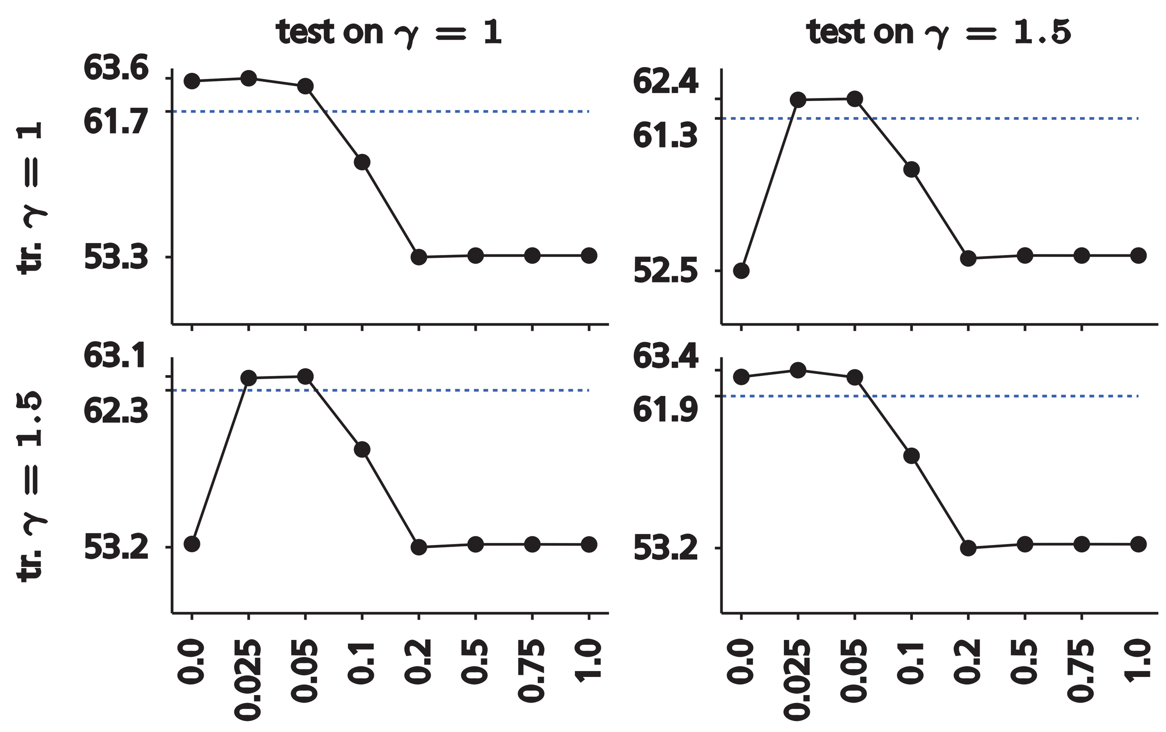
Experiments

We used images from the BRATS 2013 brain tumor segmentation dataset.



Synthetic transformations

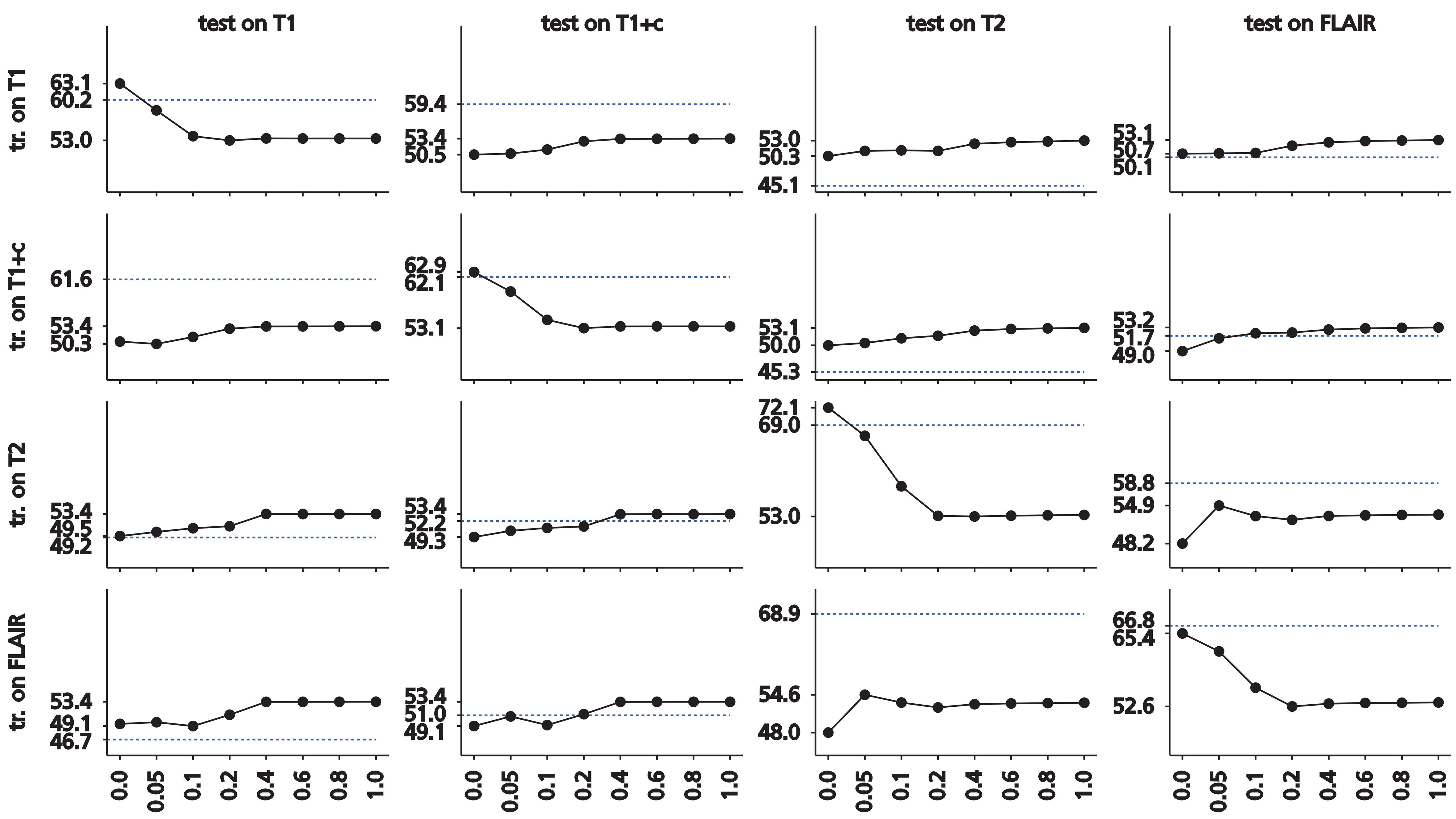
We ran experiments with synthetic multi-modal data by transforming the T1+contrast intensities x with $f(x) = x^{\gamma}$. When training and testing data is different, including the similarity term improves the classification accuracy.



Classification accuracy with axial neural network representations across synthetic modalities.

Training and testing on different modalities

Cross-domain classification is difficult, adding the similarity term might help.



Random forest classification accuracy when using representations learned by axial neural networks, for different weights α for the similarity term. (Dashed-blue lines show results with PCA features.)

Biomedical Imaging Group Rotterdam, the Netherlands
www.bigr.nl · g.vantulder@erasmusmc.nl

MICCAI MCV Workshop 2016
October 2016, Athens, Greece

This research is financed by the Netherlands Organization for Scientific Research (NWO).

Erasmus MC
University Medical Center Rotterdam