

# PROCEEDINGS OF SPIE

[SPIEDigitalLibrary.org/conference-proceedings-of-spie](https://SPIEDigitalLibrary.org/conference-proceedings-of-spie)

## Mutual information for unsupervised deep learning image registration

de Vos, Bob, van der Velden, Bas H. M., Sander, Jörg, Gilhuijs, Kenneth G. A., Staring, Marius, et al.

Bob D. de Vos, Bas H. M. van der Velden, Jörg Sander, Kenneth G. A. Gilhuijs, Marius Staring, Ivana Išgum, "Mutual information for unsupervised deep learning image registration," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 113130R (10 March 2020); doi: 10.1117/12.2549729

**SPIE.**

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

# Mutual Information for Unsupervised Deep Learning Image Registration

Bob D. de Vos<sup>a,b</sup>, Bas H. M. van der Velden<sup>b</sup>, Jörg Sander<sup>a,b</sup>, Kenneth G.A. Gilhuijs<sup>b</sup>,  
Marius Staring<sup>c</sup>, and Ivana Išgum<sup>a,b,d,e</sup>

<sup>a</sup>Department of Biomedical Engineering and Physics, Amsterdam University Medical Center,  
Amsterdam, The Netherlands

<sup>b</sup>Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>c</sup>Division of Image Processing, Leiden University Medical Center, Leiden, The Netherlands

<sup>d</sup>Department of Radiology and Nuclear Medicine, Amsterdam University Medical Center,  
Amsterdam, The Netherlands

<sup>e</sup>University of Amsterdam, Amsterdam, The Netherlands

## ABSTRACT

Current unsupervised deep learning-based image registration methods are trained with mean squares or normalized cross correlation as a similarity metric. These metrics are suitable for registration of images where a linear relation between image intensities exists. When such a relation is absent knowledge from conventional image registration literature suggests the use of mutual information. In this work we investigate whether mutual information can be used as a loss for unsupervised deep learning image registration by evaluating it on two datasets: breast dynamic contrast-enhanced MR and cardiac MR images. The results show that training with mutual information as a loss gives on par performance compared with conventional image registration in contrast enhanced images, and the results show that it is generally applicable since it has on par performance compared with normalized cross correlation in single-modality registration.

**Keywords:** Unsupervised machine learning, image registration, mutual information

## 1. INTRODUCTION

The majority of deep learning-based image registration methods comprise inference-based methods where the goal is one-pass image registration. In these methods a neural network, in particular a convolutional neural network (CNN), is trained that directly predicts the transformation parameters that align a pair of given input images. Such networks can be trained in a weakly-supervised manner using predetermined registration parameters or synthetically generated training examples;<sup>1,2</sup> or they can be trained in a fully unsupervised manner, using voxel-based similarity metrics,<sup>3,4</sup> similar to conventional image registration. Recent works proposing unsupervised deep learning-based image registration methods employ normalized cross correlation.<sup>3,4</sup> This metric is an excellent choice when a linear relation between voxel intensities exist, but it is suboptimal in case of a non-linear relation.

Ever since its introduction as a similarity metric, mutual information has developed into arguably the primary metric for voxel-based (multi-modal) image registration.<sup>5,6</sup> The benefit of mutual information is that it can model a probabilistic relation between voxel intensities. Hence, in addition to single-modality image registration, it is generally applicable to registration of medical images with non-linear relation between image intensities, e.g. for registration of images from different modalities or images with varying levels of contrast enhancement. Despite its popularity in conventional image registration, mutual information has not yet been applied to unsupervised deep learning-based image registration.

In this paper we investigate the application of mutual information with two popular network architectures used for unsupervised deep learning-based image registration: a U-Net-based architecture<sup>3</sup> for direct prediction of displacement vector fields (DVF) and a patch-based architecture (DIRNet)<sup>4</sup> for parameterized DVF prediction with B-splines. The method is evaluated on same-modality cardiac cine MRI and dynamic contrast-enhanced breast MRI having a nonlinear relation.

## 2. METHOD

### 2.1 Unsupervised Image Registration CNN

Training a CNN for unsupervised pair-wise image registration is similar to conventional image registration, because both use gradient descent-based optimization with a differentiable similarity metric. In both methodologies the aim is to align a *moving* images to *fixed* images. A *spatial transformation* maps points from the fixed to the moving image space, an *interpolator* resamples values mapped by the transform, and a *similarity metric* is used for optimization. However, unlike in conventional registration where transformation parameters are optimized, in DLIR the parameters of a CNN are optimized using all pairs of fixed and moving images from a training set. After optimization, the trained CNN can be used on unseen image pairs for image registration in one pass.

Popular CNN architectures for deep learning based image registration predict DVFs directly,<sup>1,3</sup> while other architectures use parameterized transformation models, like thin-plate splines,<sup>2</sup> or B-splines.<sup>4</sup> In our current work we employ two registration architectures: one outputting a DVFs directly using the design of<sup>3</sup> and an architecture outputting B-spline control points using the design of<sup>4</sup> (DIRNet). To mitigate folding, a bending energy penalty was added to the loss function, with a weight of 0.05.<sup>4,7</sup> The CNNs were trained using randomly sampled spatially linked image patches from fixed and moving image pairs. During training linear interpolation was used for resampling, and during final testing B-spline interpolation was used. Adam was employed for optimization using a learning rate of 0.001,  $\beta 1 = 0.9$ , and  $\beta 2 = 0.999$ .

### 2.2 Mutual information as a similarity metric

Mutual information originates from information theory. It is a measure of shared information between two variables.<sup>5</sup> In the case of image registration, these variables represent voxel intensities from a pair of images; the more information is shared between spatially corresponding voxels, the higher their mutual information. Mutual information relies on calculation of joint and marginal entropies, which relies on the calculation of probability distributions. To implement mutual information in a deep learning framework, we cannot use a histogram-based calculation of probability as this is non-differentiable. Hence, we rely on a Parzen window formulation using a Gaussian function.<sup>8</sup> By exploiting this, we are able to implement mutual information in a deep learning framework, where we make use of the automatic differentiation capabilities of PyTorch.<sup>9</sup> We use the following negative normalized version:

$$L_{NMI}(A, B) = -\frac{I(A, B)}{H(A) + H(B)}, \quad (1)$$

where  $I(A, B)$  is the mutual information of images A and B, and  $H(A)$  and  $H(B)$  their respective marginal entropies.

## 3. EXPERIMENTS

We evaluate mutual information for unsupervised deep learning image registration on intra-patient pair-wise image registration of dynamic contrast-enhanced breast MRI, the dataset has nonlinear voxel intensities caused by varying phases of contrast enhancement. To evaluate whether mutual information is generally applicable, we compare mutual information with normalized cross correlation on single-modality cardiac cine MRI from the ACDC challenge ???. For all experiments we used 64 bins and a Gaussian Parzen window with  $\sigma = 3 \cdot 10^{-3}$  to compute the loss  $L_{NMI}$ . The sigma was relatively small, because all image intensity values were rescaled and clamped to an interval of  $[0, 1)$  based on their 1<sup>st</sup> and 99<sup>th</sup> percentiles. The experiments were performed using Python 3.6 and PyTorch.<sup>9</sup>

### 3.1 Breast MRI

The dynamic contrast-enhanced breast MRI series were acquired on a 1.5 T MRI-scanner: one  $T_1$ -weighted image was made before contrast-injection and four after contrast-injection with intervals of 90 s. During scanning some patient motion may occur that can hamper correct assessment when performing quantitative analyses. Therefore, image registration is desired before quantitative analysis of breast MRI. The concentration of contrast agent per tissue-type changes dynamically over time, leading to non-linear differences in voxel-intensities among images. Thus, mutual information is ideally suited as a similarity metric for registration of these images.

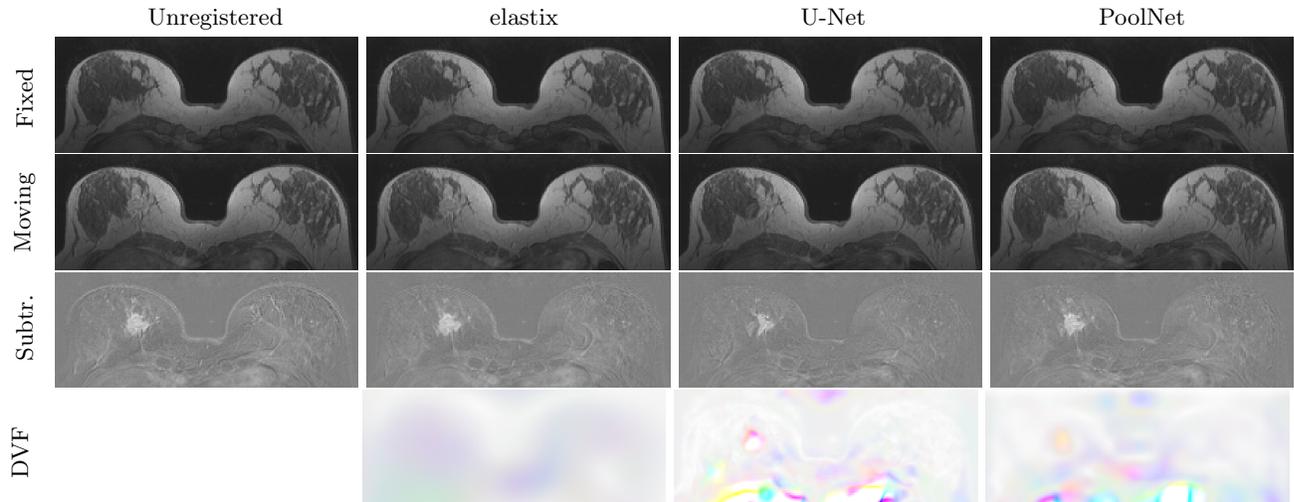


Figure 1: Registration results fixed image (top row) and the moving image (second row) for both elastix and PoolNet registrations. The U-Net registration shows an unrealistic deformation in the tumor, which is better visible in the subtraction images (third row). This is also visible in the corresponding DVFs (bottom row). The DVFs are RGBA color coded in the x, y, and z direction, and the scaled vector length as alpha channel. The fixed images in the top row are the same for each method, and are displayed for ease of comparison.

The 150 studies were split into 90 training, 10 validation, 50 hold-out test studies. A U-Net was trained having a downsampling and upsampling rate of  $2 \times 2 \times 2$ , and a PoolNet was trained having a grid-spacing of  $8 \times 8 \times 8$  voxels. Training was performed in 75 000 iterations showing mini-batches of 8 random permutations of patch-pairs of  $64 \times 64 \times 64$  voxels.

The deep learning-based image registration method was compared with conventional image registration. Conventional image registration used a registration schedule specifically optimized for breast MR registration using elastix<sup>10</sup> with normalized mutual information, 64 histogram bins, and 3 resolutions (3D smoothing pyramid of 4, 2, and 1). Each resolution was optimized in 1 000 iterations. Each iteration, a random coordinate image sampler sampled 4 096 points within a random sample region.

Registration performance was evaluated in a blinded study by an expert, using fixed and moving images from the test set having maximum deformation between them. Fixed images were those without contrast enhancement and moving images were the final contrast-enhanced images. The results of the three registration methods were shown in a random order to the expert. The expert was forced to make a ranking of preference by comparing warped images and subtraction images. The expert was free to adjust intensity window and level or inspect details in the images. There was a statistical difference among the registration methods ( $\chi^2 = 10.6$ ,  $P = 0.032$ , Table 1). When comparing between registration methods, it was shown that elastix significantly outperformed the U-Net registration (32 out of 50 times, 64%,  $\chi^2 = 1.48$ ,  $P = 0.048$ ). None of the other comparisons was significant.

Additionally, registration performance was quantitatively evaluated by computing histogram-based calcula-

Table 1: Contingency table showing the ranks of all registered images. PoolNet, conventional image registration (ConvIR), and U-Net were ranked resp. first, second, and third.

Rank	elastix	U-Net	PoolNet
First	17	13	20
Second	18	12	20
Third	15	25	10

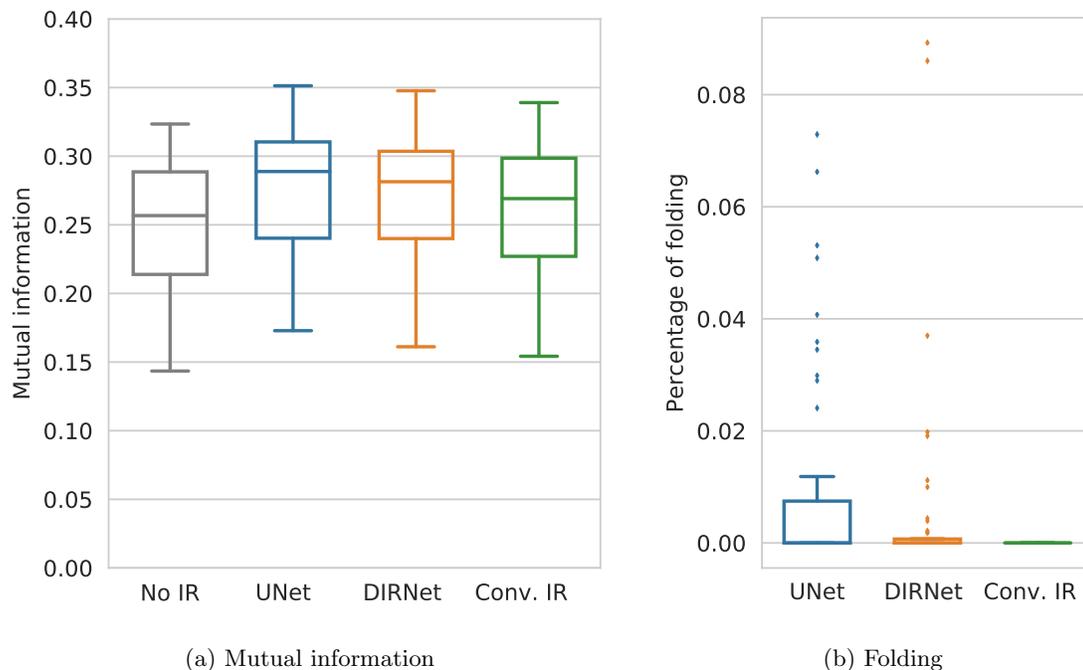


Figure 2: Left: Mutual information before (No IR) and after image registration with U-Net, PoolNet and conventional image registration (Conv IR). Right: Amount of folding for registration with U-Net, PoolNet and conventional image registration (Conv IR).

tion of normalized mutual information after registration, and by evaluating the Jacobian determinant of the DVFs (Figure 2). Highest mutual information was shown for U-Net and lowest for conventional image registration. Amount of image folding was quantified by determining the volume percentage of negative Jacobian determinants. Folding was found in 23 (out of 50) registrations for registration with U-Net, in 14 registration for PoolNet, and no folding had occurred in conventional image registration.

### 3.2 Cardiac MRI

Cardiac cine MRI was used from the Automated Cardiac Diagnosis Challenge (ACDC).<sup>11</sup> The dataset contains short-axis reconstructions of highly anisotropic volumetric images. The images have varying spatial and temporal resolution. All images visualize the heart during a single heart beat. The images have reference labels of the time-points with maximum deformation: the end-diastolic and end-systolic time-points. The challenge provides annotations for the left ventricle blood pool (LVb), the left ventricle myocardium (LVm), and the right ventricle (RV).

To evaluate whether mutual information is generally applicable, we compared mutual information with normalized cross correlation on this single-modality non-contrast-enhanced dataset.<sup>4</sup> Since the dataset is highly anisotropic, we adjusted the U-Net and PoolNet designs to account for this. Both networks were implemented in 3D, but downsampling was not applied through-plane; downsampling and upsampling factors for the U-Net were set to  $(2 \times 2 \times 1)$ , and grid spacing of the PoolNet was set to  $(8 \times 8 \times 1)$  voxels.

Experiments were performed using a 4-fold cross-validation setup using the 100 training images. The network of each fold was trained in 50 000 iterations of mini-batches consisting of patch pairs with a size  $(128 \times 128 \times 6)$  voxels, regardless of voxel spacing.

Boxplots in Figure 4 show amount of folding and label propagation results. Folding is much more apparent in U-Net based architectures, especially when normalized cross correlation is employed as a similarity metric. Lowest amount of folding was seen in a PoolNet architecture using mutual information. Figure 3 shows registrations with corresponding DVFs which are exemplary for all experiments: the DVFs of U-Net trained with mutual information show less smooth DVFs than the other approaches.

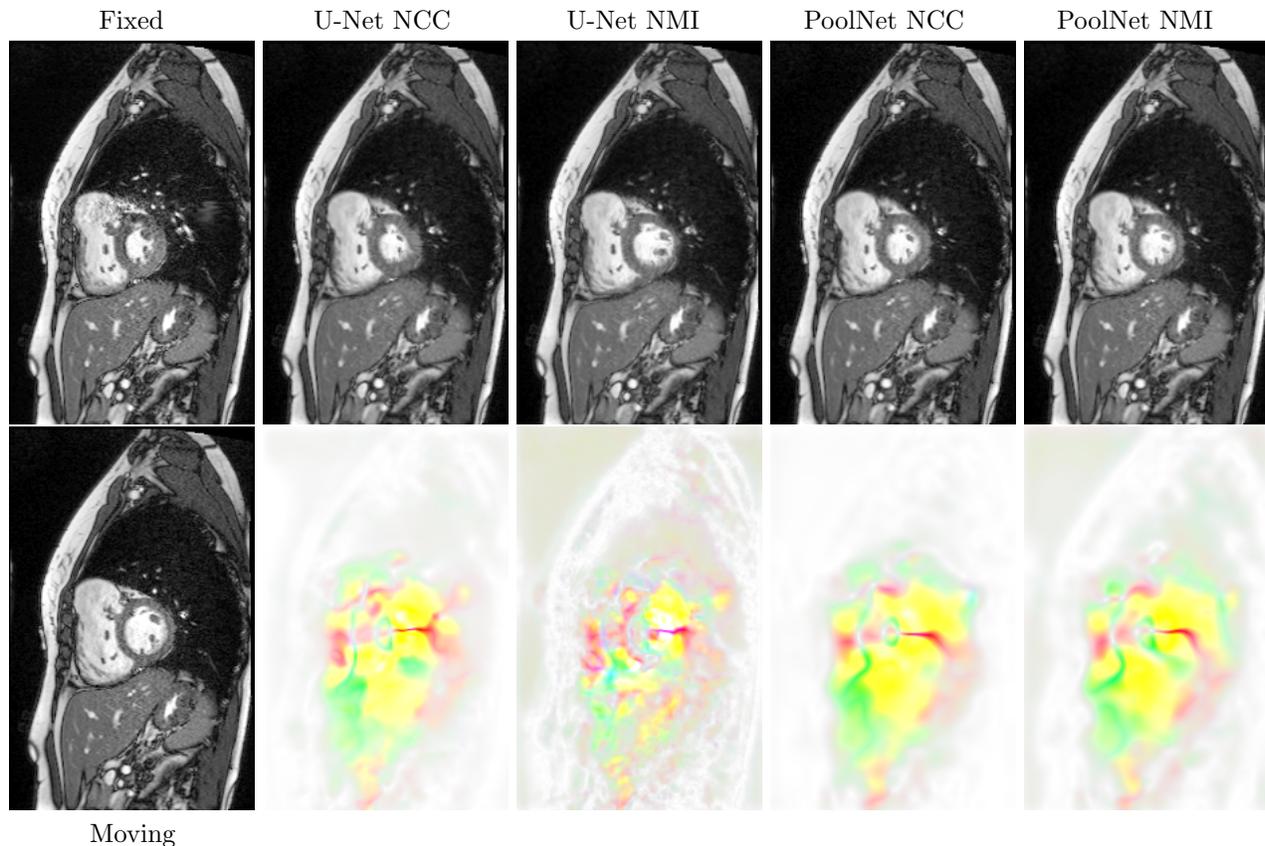


Figure 3: Registration results of cardiac MR. Fixed and moving images are image slices taken from volumes at end-systole and end-diastole. The top row show warped images. The bottom row shows the corresponding color coded DVFs.

#### 4. DISCUSSION AND CONCLUSION

We presented a method for unsupervised deep learning image registration using mutual information as a similarity metric. The results show that CNNs trained with mutual information perform on par with a conventional method specifically designed for breast MRI registration. Additionally, we have shown general applicability of mutual information for single-modality deep learning registration of cardiac MRI.

Mutual information can be used to train any CNN architecture, but there are some noticeable differences in performance. The experiments with registration of breast MR images show that the highest mutual information was obtained using a CNN for direct DVF prediction. This is likely a consequence of finer granularity of the DVF, resulting in more precise alignment. High mutual information, however, was not related to the expert's preference. On the contrary, the results from B-spline based CNN and conventional image registration were preferred more often, although this was not significant. Fine granularity is not always favorable in image registration, since it might result in unrealistic deformations and increased folding (e.g. Figure 1). The expert's preference for B-spline registration is likely a consequence of the inherent smoothness of the DVFs generated with B-splines.

All CNNs trained for Cardiac MRI registration showed high performance in terms of Dice coefficient and Hausdorff distance.<sup>4</sup> Furthermore, CNNs trained with mutual information generated DVFs with negligible folding, especially when compared to CNNs trained with normalized cross correlation. This is likely a consequence of mutual information being relatively insensitive to small intensity differences between images because of its probabilistic nature. While cardiac MR images did not contain contrast agents, there are differences in image intensity among timepoints, especially in the right ventricle. In these cases, normalized cross correlation might

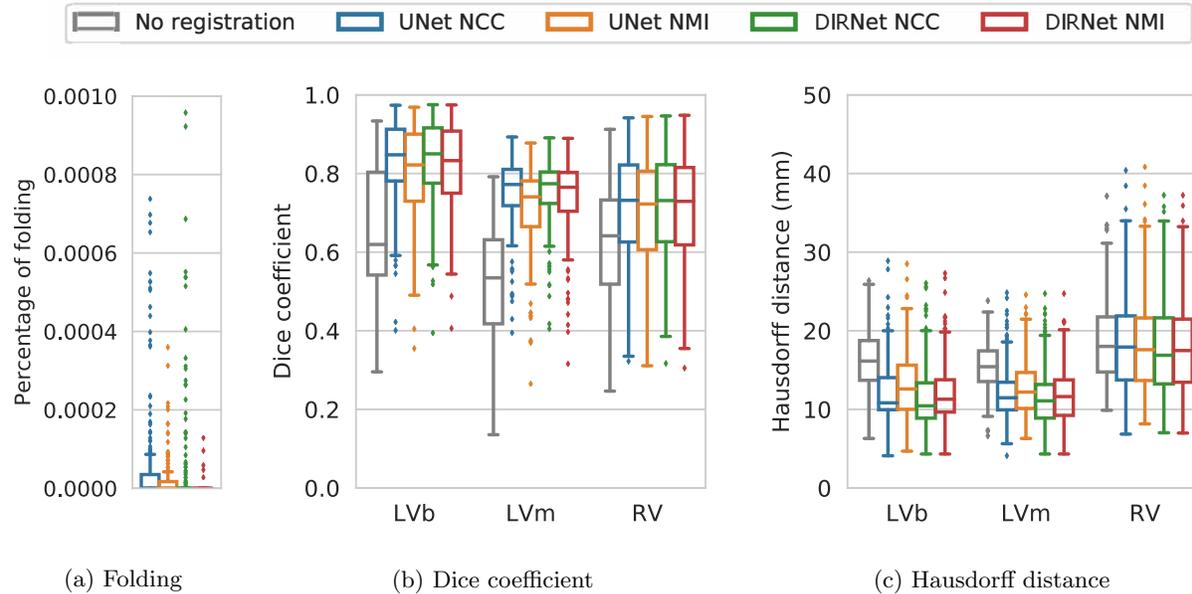


Figure 4: Boxplots showing registration results of U-Net-based architecture and pooling-based architectures for registration of cardiac MR data. Left: amount of folding (i.e. singularities) expressed in percentages of total volume. Middle: Dice coefficients obtained by label propagation of end-diastolic and end-systolic volumes. Right: Hausdorff distances obtained by label propagation of end-diastolic and end-systolic volumes. For label propagation three structures were individually analyzed: left ventricle bloodpool (LVb), left ventricle myocardium (LVm), and the right ventricle (RV). Label propagation results are similar for all methods, but image folding is less apparent for mutual information based networks.

be too stringent and force voxel-precise alignment thereby inadvertently generating implausible deformations. While, the employed bending energy penalty enforced smooth DVF topology and mitigated image folding, implausible deformations caused by normalized cross correlation might be fully prevented by using a diffeomorphic approach.<sup>12</sup>

We have shown that mutual information is suitable for unsupervised deep learning image registration regardless of employed CNN architecture. It is likely that mutual information can be used to train CNNs for other (multi-modality) registration applications that were previously handled by conventional image registration, or by normalized cross correlation-driven deep learning image registration.

## ACKNOWLEDGMENTS

This work is part of the research program Deep Learning for Medical Image Analysis with program number P15-26, which is partially financed by the Netherlands Organization for Scientific Research (NWO) and PIE Medical Imaging.

## REFERENCES

- [1] Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B., Išgum, I., and Staring, M., “Nonrigid image registration using multi-scale 3D convolutional neural networks,” in *[Medical Image Computing and Computer Assisted Intervention MICCAI 2017]*, 232–239, Springer International Publishing, Cham (2017).
- [2] Eppenhof, K. A. J., Lafarge, M. W., Moeskops, P., Veta, M., and Pluim, J. P. W., “Deformable image registration using convolutional neural networks,” in *[Medical Imaging 2018: Image Processing]*, Angelini, E. D. and Landman, B. A., eds., **10574**, 192 – 197, International Society for Optics and Photonics, SPIE (2018).

- [3] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V., “Voxelmorph: A learning framework for deformable medical image registration,” *IEEE Transactions on Medical Imaging* **38**, 1788–1800 (Aug 2019).
- [4] de Vos, B., Berendsen, F., Viergever, M., Sokooti, H., Staring, M., and Igum, I., “A deep learning framework for unsupervised affine and deformable image registration,” *Medical Image Analysis* **52**, 128 – 143 (2019).
- [5] Pluim, J. P., Maintz, J. A., and Viergever, M. A., “Mutual-information-based registration of medical images: a survey,” *IEEE transactions on medical imaging* **22**(8), 986–1004 (2003).
- [6] Sotiras, A., Davatzikos, C., and Paragios, N., “Deformable medical image registration: A survey,” *IEEE Transactions on Medical Imaging* **32**, 1153–1190 (July 2013).
- [7] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J., “Nonrigid registration using free-form deformations: application to breast mr images,” *IEEE Transactions on Medical Imaging* **18**, 712–721 (Aug 1999).
- [8] Wells, W., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R., “Multi-modal volume registration by maximization of mutual information,” *Medical image analysis* **1**(1), 35–51 (1996).
- [9] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., “Automatic differentiation in pytorch,” (2017).
- [10] Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. W., “elastix: A toolbox for intensity-based medical image registration,” *IEEE Transactions on Medical Imaging* **29**, 196–205 (Jan 2010).
- [11] Bernard, O., Lalonde, A., Zotti, C., Cervenansky, F., Yang, X., et al., “Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?,” *IEEE Transactions on Medical Imaging* **37**, 2514–2525 (Nov 2018).
- [12] Dalca, A., Balakrishnan, G., Guttag, J., and Sabuncu, M., “Unsupervised learning for fast probabilistic diffeomorphic registration,” in [*Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*], 729–738, Springer International Publishing, Cham (2018).