Impact of PCA-based preprocessing and different CNN structures on deformable registration of sonograms

Christian Schmidt

Westfälische Hochschule
University of Applied
Sciences
Neidenburger Strasse 43
45897 Gelsenkirchen
Germany

christian.schmidt
@w-hs.de

Heinrich Martin Overhoff

Westfälische Hochschule
University of Applied
Sciences
Neidenburger Strasse 43
45897 Gelsenkirchen
Germany

heinrich-martin.overhoff
@w-hs.de

ABSTRACT

Central venous catheters (CVC) are commonly inserted into the large veins of the neck, e.g. the internal jugular vein (IJV). CVC insertion may cause serious complications like misplacement into an artery or perforation of cervical vessels. Placing a CVC under sonographic guidance is an appropriate method to reduce such adverse events, if anatomical landmarks like venous and arterial vessels can be detected reliably. This task shall be solved by registration of patient individual images vs. an anatomically labelled reference image. In this work, a linear, affine transformation is performed on cervical sonograms, followed by a non-linear transformation to achieve a more precise registration. Voxelmorph (VM), a learning-based library for deformable image registration using a convolutional neural network (CNN) with U-Net structure was used for non-linear transformation. The impact of principal component analysis (PCA)-based pre-denoising of patient individual images, as well as the impact of modified net structures with differing complexities on registration results were examined visually and quantitatively, the latter using metrics for deformation and image similarity. Using the PCA-approximated cervical sonograms resulted in decreased mean deformation lengths between 18% and 66% compared to their original image counterparts, depending on net structure. In addition, reducing the number of convolutional layers led to improved image similarity with PCA images, while worsening in original images. Despite a large reduction of network parameters, no overall decrease in registration quality was observed, leading to the conclusion that the original net structure is oversized for the task at hand.

Keywords
Medical image registration, deformable registration, sonograms, Voxelmorph, CNN

1 INTRODUCTION

Placement of a central venous catheter (CVC) is a procedure that carries risk for multiple complications, e.g., arterial puncture of the common carotid artery has an occurrence rate of $6\% - 9\%$ [Bee03]. This work aims to further improve the ultrasound guided CVC placement into the internal jugular vein (IJV) (Fig. 1), by detecting the IJV and indicating a needle target position in a manually acquired patient individual ultrasound image. The needle target position in such an image is to be de-

termined by automated, computer-based analysis. It is assumed that an optimal needle target position is defined in a reference image. The task at hand is to map the optimal needle target position onto patient individual images. In order to realize this mapping, the patient individual images are to be registered vs. the reference image.

Overfitting occurs when a model learns the training data well, but does not generalize the acquired information to new data. This is an issue in medical machine learning applications, since these datasets are usually small compared to the complexity of deep neural network structures, or due to low signal-to-noise ratio in the data.

The hypothesis of this work is: An improvement of the signal-to-noise ratio in image data and a systematic reduction of network size can yield improved registration results with overall less deformation, and thus a more regular registration field. In this work, principal com-
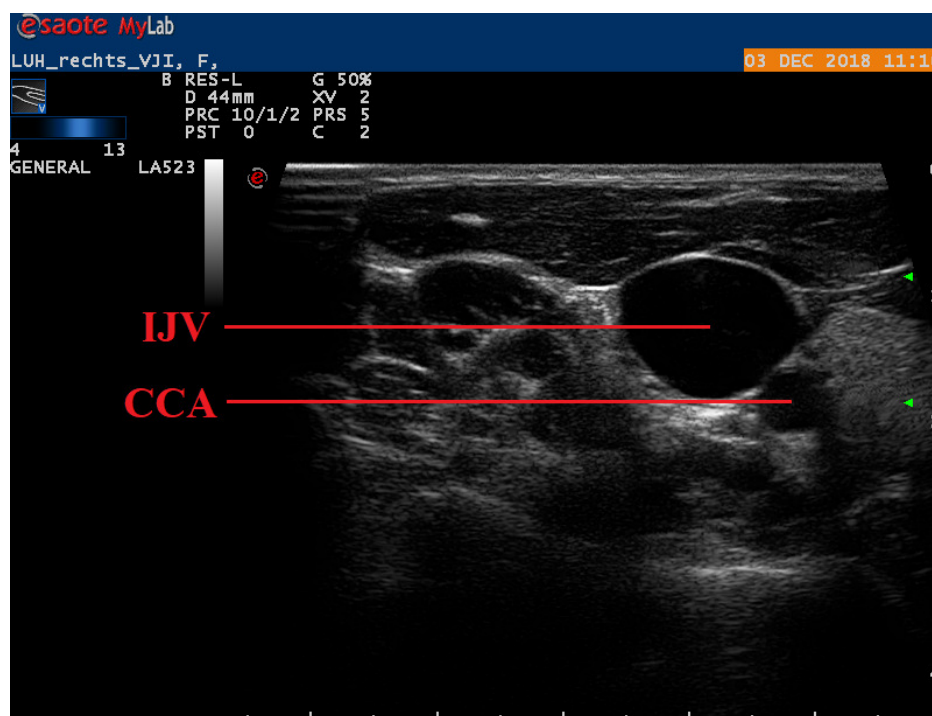
Figure 1: **Example image of an original cervical sonogram. Internal jugular vein (IJV) and common carotid artery (CCA) are labelled.**

ponent analysis (PCA) is used for noise reduction, and Voxelmorph, a U-Net-based convolutional neural network (CNN), is used as a reference network structure. To evaluate the hypothesis, three models for both image types are parameterized for a fine registration of affinely pre-registered ultrasound images of the human neck. The main tasks of this work are:

- reduce the size of original ultrasound images to a region of interest (ROI) that contains mainly the IJV. This shall be done by feature-based image segmentation. Subsequently, apply affine pre-registration to the ROI images. Because only few clinical images are available, and those have varying anatomical structures and image contrast, this procedure shall make the deformable image registration less error-prone.

- Perform a PCA on the image data set and approximate it by linear combination of the most relevant principal components.

- Train neural networks with three different structures and different number of free parameters to register image pairs (non-linear, deformable transformation). For each net structure, train two versions, one for original images, and one for PCA-approximated images.

- Quantitatively analyze the impact of the number of net parameters and the image type on the registration

result, using evaluation metrics for deformation and image similarity.

2 RELATED WORK

Many different methods for deep learning-based medical image registration have been proposed in the past. For rigid transformations in particular, deep reinforcement learning (RL) techniques [Mni15] have gained some popularity. [Lia17] proposed a RL strategy for rigid 3D-3D registrations in computer tomography images, which is based on finding the optimal sequence of motion actions (rotations and translations) for image alignment. Since RL networks are constrained to low dimensionality of outputs, they have been used almost exclusively for rigid registrations, since those can be expressed by a small number of transformation parameters. With the rise of networks, which can directly estimate deformation vector fields (DVF) and are not constrained to rigid transformations, RL-based methods fell out of favor in recent years [FuY20].

Networks, which directly estimate the DVF can be classified into supervised and unsupervised methods. Supervised networks require ground truth transformations (either DVF, in case of deformable registration, or rigid transformation parameters). Ground truth transformations can be obtained by artificially de-aligning images with random rotations and translations [Sal19, Epp18] or by using traditional, non-learning methods to register image pairs and use the resulting DVFs as ground truth
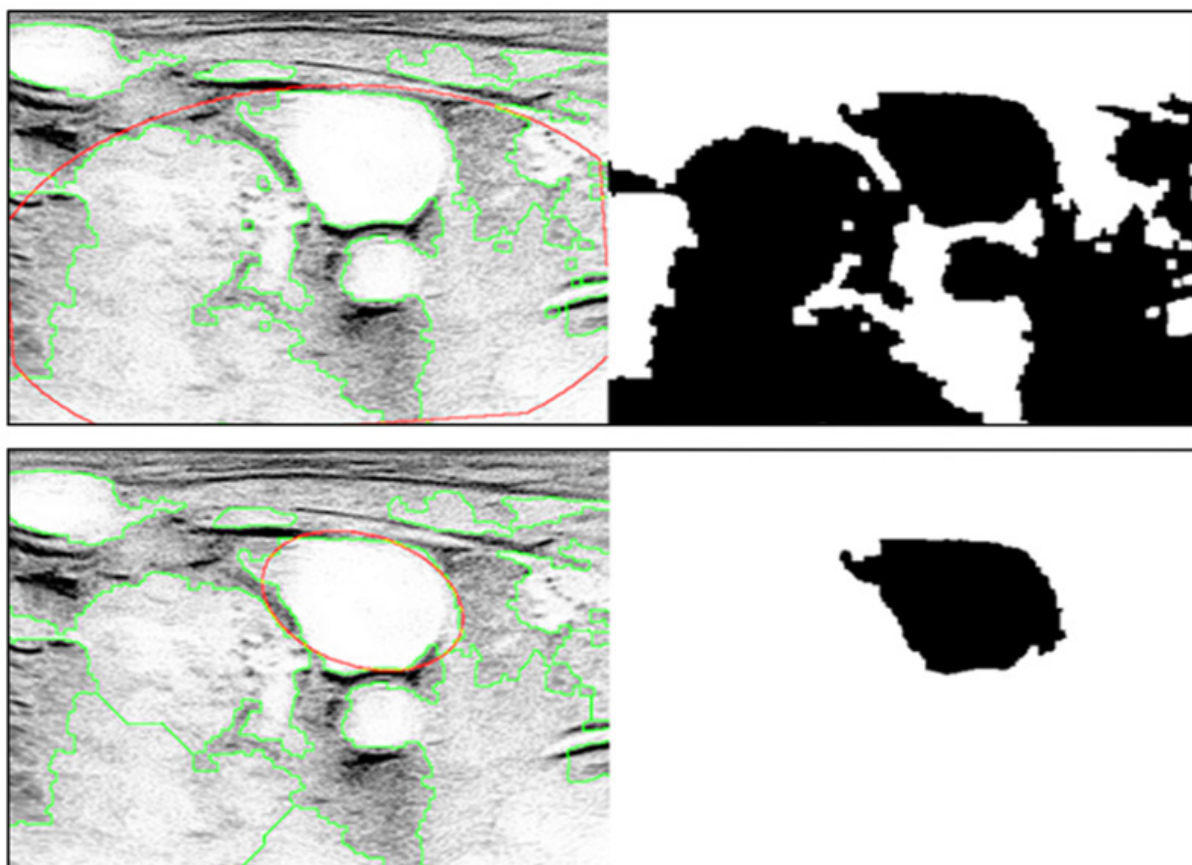
Figure 2: *n* **largest objects after binarization are delineated with green contours, the object identified as correct is marked with a red ellipse (left). Top row shows objects before, bottom row objects after watershed-transformation. Corresponding binary images after thresholding with** g_{thresh} **are shown on the right. Sonograms are dispayed with inverted grayscale pixel values for better visibility.**
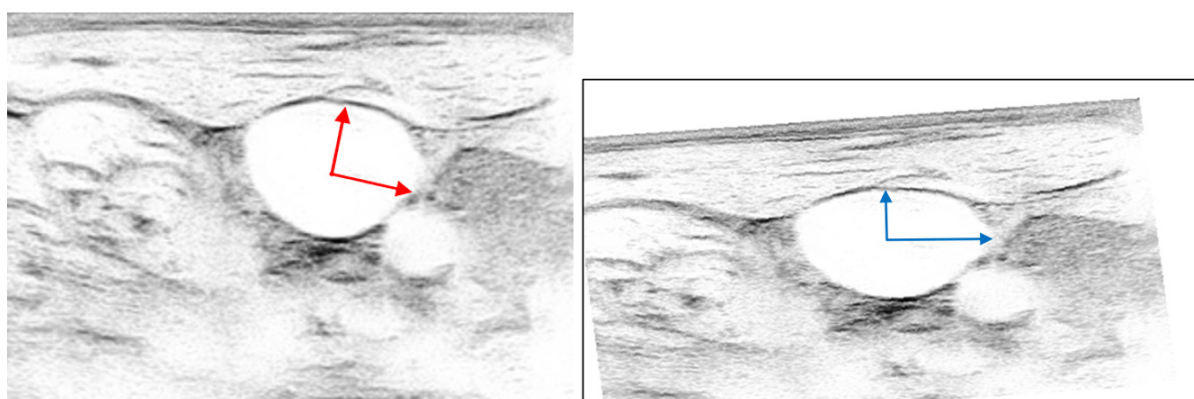

Figure 3: **Example sonogram of the IJV before (left) and after (right) affine transformation. Ellipse parameters resulting from the preceding feature-based segmentation, are used to translate (x_{OR}), rotate ($-\varphi$) and scale (s_x, s_y) the images to coarsely pre-register them for the subsequent deformable deformation by the CNN.**

q	1	3	5	7	9	11	13	15	17	19
$cEVR$	0.12	0.31	0.46	0.55	0.62	0.67	0.71	0.75	0.77	0.79

Table 1: **Cumulative explained variance ratio** $cEVR$ **by number of first** q **principal components.**
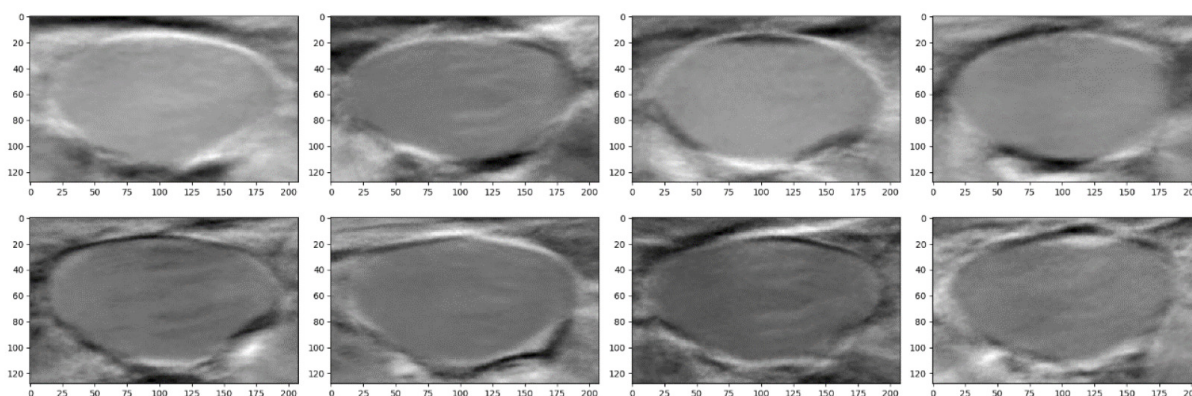
Figure 4: **First $q = 1 \ldots 8$ principal component images $\mathbf{G}(\mathbf{y}_1)$ through $\mathbf{G}(\mathbf{y}_8)$ from performing a PCA on the ultrasound image data set (top row $\mathbf{G}(\mathbf{y}_1) \ldots \mathbf{G}(\mathbf{y}_4)$, bottom row $\mathbf{G}(\mathbf{y}_5) \ldots \mathbf{G}(\mathbf{y}_8)$).**

for training [Sen18]. Lack of medical datasets with known ground truth DVFs led to a rising demand for unsupervised networks. With the introduction of spatial transformer networks [Jad15], calculating image similarity losses was made possible during training. This is achieved by warping the estimated DVF with the input image and comparing the resulting image with the reference image. These networks do not require supervision by ground truth annotations and in addition to image similarity loss, employ a regularization loss term, to ensure smooth and anatomically plausible transformations [Zha18, Bal19].

Looking at image modalities, DL-based registration of ultrasound (US)-images only played a subordinate role in recent research, despite the high prevalence of sonography in clinical practice. A review of current publications in medical image registration [Bov20] found, that US-images were only used in about 5% of research papers with the topic of DL-based medical image registration, while magnetic resonance imaging (MRI) (52%) and computer tomography (CT) (19%) dominated the field. This is mainly due to higher availability of public MR training datasets and the fact, that the majority of papers examined registrations of brain images, which are predominantly recorded in MR and CT scans. US-images were most often used in multimodal registration tasks [HuY18, Yan18], in which, e.g., pre-procedural MR scans were aligned with intraprocedural US-images.

3 PROPOSED SOLUTION

Segmentation and affine pre-registration

Firstly, the original ultrasound image is being cropped and the interface of the ultrasound machine is removed. To segment the image into foreground and background, it is binarized using a binarization threshold value g_{thresh}. Subsequently, everything but the n largest objects are removed, to filter out structures that are too small to reasonably be considered. A watershed

transform is implemented (Fig. 2), since as g_{thresh} increases, the increasingly forming clusters need to be separated, in order to be detected as individual entities. Choosing the correct IJV-object out of the remaining n, is done via the y-coordinate of the object's center and the distance between the common carotid artery (CCA) and the respective object. The object center's y-coordinates can be utilized because the distal location of the IJV is roughly the same for all subjects. Additionally, since CCA and IJV are in close anatomical proximity, all objects outside of a certain distance to the CCA can be excluded.

To obtain a smoother contour and an object which is geometrically parameterizable, the correctly identified object is approximated with an ellipse. The ellipse parameters major and minor axis length (a, b), and major axis rotation angle vs. the x-axis (φ) are extracted from the ellipse approximation. These parameters are subsequently used in the affine transformation (preregistration).

This affine transformation between object IJV (O) and reference IJV (R) (Fig. 3) was performed as follows: At first, the image is translated by $\mathbf{x}_{OR}$, which aligns the center of the approximated ellipse with the image center. This is also the origin of the new reference coordinate system. Secondly, a rotation of $-\varphi$ degrees is applied to the image. This alignment of the center and rotation angle of the object IJV and the new reference coordinate system can be described as a 2D rigid body transform. Subsequently, image coordinate axes are scaled using the scaling factors s_x and s_y. Values for $s_x = \frac{a_R}{a_O}$ and $s_y = \frac{b_R}{b_O}$ are used to match the major and minor axes lengths of object vs. reference IJV ellipses. Finally, a rectangle of size 208×128 pixels around the object center is cropped, to only leave relevant parts of the image for later use as training data.

Principal Component Analysis

PCA is used to reduce dimensionality in the ultrasound image data set. For this purpose, $p = 81$ pre-registered

ultrasound images of the human neck (transversal plane) from 14 different subjects (five to six images per subject) of size 208×128 pixels are investigated.

PCA is a statistical method, which is used for projecting a p-dimensional data set into a q-dimensional sub-set ($q < p$), while preserving characteristic data variability [Jol16, Kon17]. A data set consists of p variables x_i, $1 \leq i \leq p$, with n observations. Each variable x_i has a mean μ_i and a variance σ_i^2 calculated over its n observations. The sum over the variance of all variables is the total variance

$$\sigma_{\text{total}}^2 = \sum_{i=1}^{p} \sigma_i^2$$

Observations x_{mi} of variable x_i, $1 \leq m \leq n$, are noted as a $n \times 1$ vector $\mathbf{x}_i$. With the observed mean for vector $\mathbf{x}_i$ being μ, the PCA is calculated over centered observations $\mathbf{X}_i = \mathbf{x}_i - \mu$. The eigenvalues λ_j of the co-variance matrix are indexed in descending order, they represent variances and fulfill

$$\sigma_{\text{total}}^2 = \sum_{j=1}^{p} \lambda_j$$

The PCA yields new variables $\mathbf{y}_j$, the so-called principal components (PCs). PCs are linear combinations of centered observations and have an identical co-variance matrix, i.e., the eigenvalues λ_j are the variances of variables $\mathbf{y}_j$. The total variance σ_{total}^2 is identical for PCs, original data, and centered observations, but is distributed differently among the variables. Goal of the PCA is to explain a major part of the total variance with a small number of variables q, the cumulative explained variance ratio ($cEVR$) is given by

$$cEVR = \frac{\sum\limits_{j=1}^{q} \lambda_j}{\sigma_{\text{total}}^2}$$

The first q of all p new variables $\mathbf{y}_1 \dots \mathbf{y}_q$ determine the data sub-set, such that

$$\mathbf{x}_i \approx \tilde{\mathbf{x}}_i = \sum_{j=1}^{q} \beta_{ij}\mathbf{y}_j + \mu$$

To perform the PCA, pre-registered ultrasound images of the human neck $\mathbf{G}_i$ are reshaped into column vectors $\mathbf{x}(\mathbf{G}_i)$ with $n = 26624$ observations each. After dimensionality reduction, the above-described image vectors can be reorganized as images $\mathbf{G}(\tilde{\mathbf{x}}_i) \approx \mathbf{G}$. The first q PCs (Fig. 4) are used to approximate the original dataset. In the upcoming sections, $q = 8$ is used, which accounts for about 58% of the data's variance (Table 1) while reducing 90% of dimensionality (from $208 \times 128 \times 81$ to $208 \times 128 \times 8$).

Voxelmorph and variation of net structures

Voxelmorph (VM) [Bal19], a learning-based library for deformable image registration, which uses a U-Net-based [Ron15] net structure, is used to perform the deformable sonogram registrations. An atlas-based registration approach is used in this work; thus, an image pair consists of a varying patient individual image (moving image m) and a reference image (fixed image f).

Voxelmorph uses a two-part loss function

$$J = \mathscr{L}_{\text{sim}}(f, m \circ \phi) + \gamma \mathscr{L}_{\text{smooth}}(\phi),$$

which consists of a similarity term $\mathscr{L}_{\text{sim}}(f, m \circ \phi)$ and a deformation term $\mathscr{L}_{\text{smooth}}(\phi)$. The loss function J penalizes differences in grayscale values as well as deformations and is minimized by learning optimal convolutional kernels (filters). When registering an image pair, the network yields pixel-wise displacement vectors

$$\mathbf{u} = \begin{bmatrix} u_x \\ u_y \end{bmatrix} = \begin{bmatrix} x_f \\ y_f \end{bmatrix} - \begin{bmatrix} x_m \\ y_m \end{bmatrix}, l = \|\mathbf{u}\|$$

between moving image m and fixed image f. The registration field

$$\phi = Id + \mathbf{u}$$

is formed by adding $\mathbf{u}$ to the identity transform. The resulting registration field ϕ generates a moved image $m \circ \phi$, which is similar to f.

By employing a regularization term, Voxelmorph encourages smooth, diffeomorphic deformations, i.e. deformations which are anatomically reasonable. γ serves as the regularization parameter, in this work we used $\gamma = 0.001$. As γ increases, deformation becomes more costly and the resulting deformation field, therefore, becomes more regular (smooth) and vice versa. The resulting registration field ϕ is applied to the moving image m by a spatial transformer function, to obtain the moved image $m \circ \phi$ (m warped by ϕ).

To examine the effects of reducing the number of free parameters in the CNN by cutting down its size, three net structures are introduced:

- The "full" structure proposed in the original VM paper, consisting of four encoder and seven decoder convolutional layers with 16 or 32 filters (convolutional kernels) each (16, 32, 32, 32 | 32, 32, 32, 32, 32, 16, 16). This net structure contains about 110,000 parameters.

- The "reduced" structure. Two encoder and decoder layers are removed for the second configuration (16, 32 | 32, 32, 32, 16 16), resulting in about 53,000 parameters, a reduction of 52% compared to the full net.
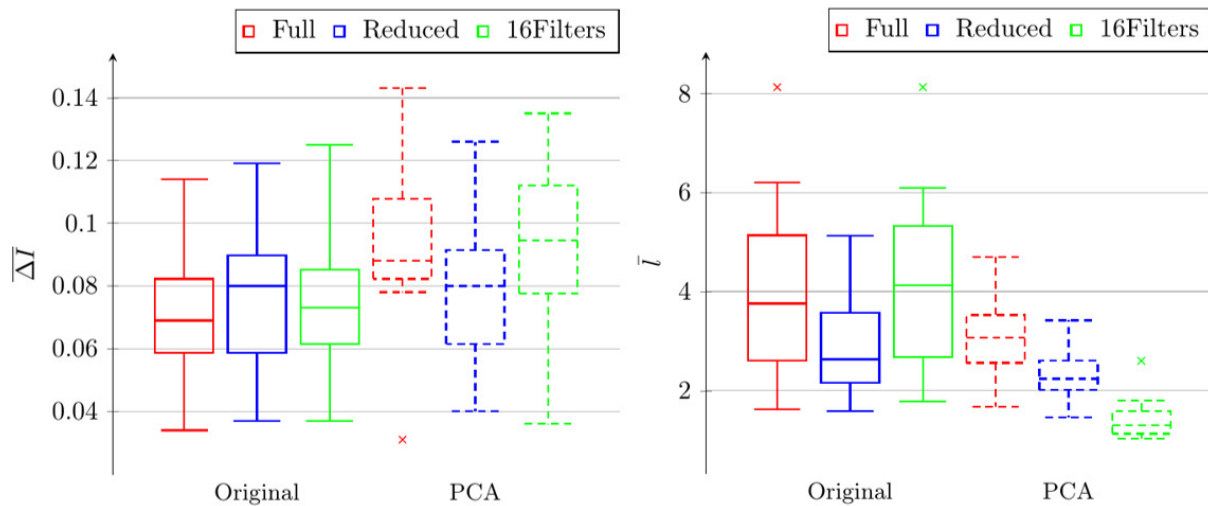
Figure 5: **Results of registrations with different net structures, using original and PCA-approximated images. Mean of differences of absolute grayscale intensities $\overline{\Delta I}$ is shown on the left, mean deformation vector length $\bar{l}$ on the right.**

- The "16 filters" structure contains all eleven convolutional layers, with 16 filters in each layer (16, 16, 16, 16 | 16, 16, 16, 16, 16, 16, 16), resulting in about 33,000 parameters, 70% less than the full net.

For each net structure, two versions are trained:

- Original image data vs. the reference image, i.e., $\mathbf{G}(\mathbf{x}_i) = \mathbf{G}(\tilde{\mathbf{x}}_i(q = 81))$ vs. $\mathbf{G}(\mathbf{x}_{(\mathrm{Ref})}) = \mathbf{G}(\tilde{\mathbf{x}}_{(\mathrm{Ref})}(q = 81))$

- PCA-approximated image set vs. the reference image, i.e., $\mathbf{G}(\mathbf{x}_i) = \mathbf{G}(\tilde{\mathbf{x}}_i(q = 8))$ vs. $\mathbf{G}(\mathbf{x}_{(\mathrm{Ref})}) = \mathbf{G}(\tilde{\mathbf{x}}_{(\mathrm{Ref})}(q = 81))$

In the upcoming results section however, PCA images are visually evaluated against the PCA-approximation $\mathbf{G}(\tilde{\mathbf{x}}_{(\mathrm{Ref})}(q = 8))$ of the reference image to account for the difference in brightness and contrast between original and PCA images.

Quantitative analysis

To quantitatively analyze the properties and quality of performed registrations, two evaluation metrics are introduced:

- Let $0 \leq I \leq 1$ be the normalized image grayscale intensities, and $\Delta I = I(f) - I(m \circ \phi)$ the pixel-wise differences between intensities of fixed image f and moved image $m \circ \phi$. Therefore, $-1 \leq \Delta I \leq 1$ holds. We define $\overline{\Delta I}$ as the mean of absolute grayscale intensity differences ΔI.

- Mean deformation vector lengths $\bar{l}$ of the registration field ϕ, where l measures the pixel-wise deformation (in pixels) that is applied to the moving image m.

We use $\overline{\Delta I}$ and $\bar{l}$ analogously to the similarity term $\mathscr{L}_{\mathrm{sim}}(f, m \circ \phi)$ and the deformation term $\mathscr{L}_{\mathrm{smooth}}(\phi)$ of the Voxelmorph loss function. Since the region around the IJV's contour is of primary importance in this work, $\overline{\Delta I}$ and $\bar{l}$ are only evaluated in a belt-like along the IJV contour.

In addition, significances α of metric differences between the above mentioned net and image pair variants are determined with a two-tailed, paired t-test. We used a 70/30 split between training and test data, training the net's parameters on a NVIDIA GeForce RTX 2060 GPU takes 2.5 to 3 minutes, depending on net structure. Registering a single image pair takes 1 to 2 seconds.

4 RESULTS

Post-registration mean of absolute intensity differences $\overline{\Delta I}$ for all net structures and image types are shown in Fig. 5. With original images, an increase in the mean $\overline{\Delta I}$ of around 12% can be observed, when using the reduced net instead of the full net (mean $\overline{\Delta I}$: 0.078 vs. 0.070, $\alpha = 0.027$). When registering PCA-approximated images however, a 17% decrease was measured when using the reduced over the full net structure (mean $\overline{\Delta I}$: 0.079 vs. 0.094, $\alpha = 0.007$). Comparing the full net structure to their respective 16 filters version showed no significant change in mean $\overline{\Delta I}$.

Looking at mean deformation vector lengths $\bar{l}$ (Fig. 5), networks trained with PCA-approximations showed decreases in mean $\bar{l}$ vs. their original image counterpart of 24% for the full, 18% for the reduced and 66% for the 16 filters net structure. In addition, registrations with PCA-approximated images display the expected smoothing and noise reducing properties, removing unwanted artifacts from the vessel lumen (Fig. 6).
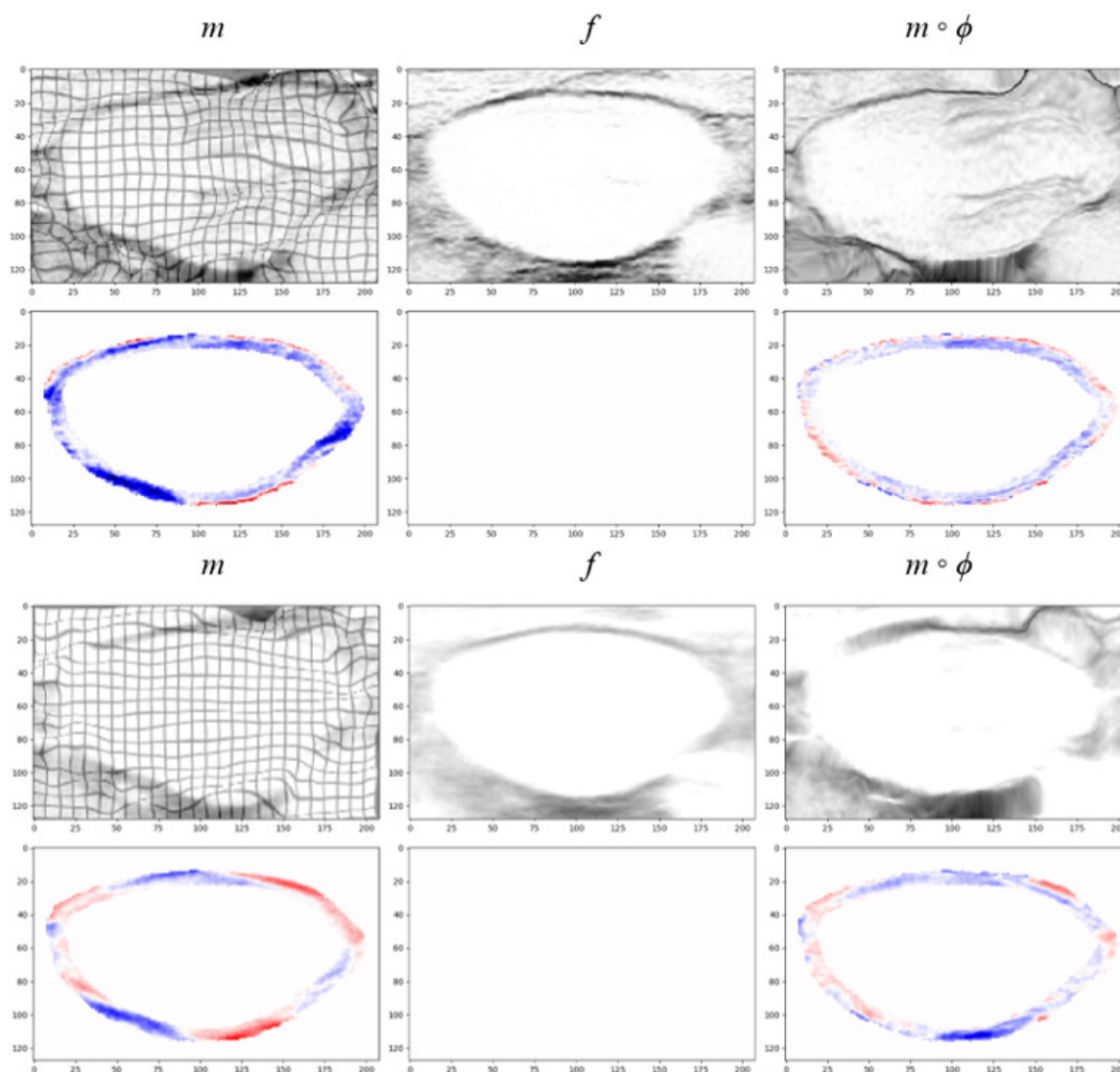
Figure 6: **Example registration results of original image (top) and PCA-approximation (bottom), using the full net structure. The registration field ϕ is warped with a regular square grid and superimposed over the moving image m, to show the extent and direction of local deformation which is applied to individual image parts. In addition, values of the pixel-wise pre- and post-registration grayscale difference ΔI are displayed color-coded, to illustrate the effect of registration on image similarities (colors ranging from dark red for $\Delta I = 1$ to dark blue for $\Delta I = -1$).**

Figure 7: **Illustration of negative original image features being transferred to PCA-approximations. Reverberation artifacts of original image of subject A (left) appear in the PCA-approximation of subject B (right), even though no such artifacts are present in the original image of subject B (middle).**

5 CONCLUSION

Despite a reduction in net parameters of up to 70% compared to the originally proposed full net and reducing the mean deformation vector lengths $\bar{l}$ by 18% - 66%, no overall reduction in registration quality was measurable in the downscaled net structures. Specifically, for the combination of reduced net structure with PCA-approximated images, a significant decrease of $\bar{l}$ ($\bar{l} = 2.32$ vs. 2.85, $\alpha = 0.045$) vs. original images was observed, while $\overline{\Delta I}$ remained nearly unchanged ($\overline{\Delta I} = 0.079$ vs. 0.078). This confirms the hypothesis described in the introduction section, and leads to the conclusion that the full net structure is unnecessarily oversized for the problem at hand.

The net structure can be reduced in size to diminish problems like overfitting, while also running up to 15% faster during training compared to the full net structure. In case of images which contain similar, regularly shaped structures, it is recommended to pre-process them with the proposed PCA procedure and employ reduced net structures, to reduce mean deformations and yield more regular registration fields. Since PCA is based on variances, it is highly sensitive to outliers. Thus, noisy images (outliers) in the original data set negatively affect the quality of the principal components, which then in turn affect the approximated PCA images (Fig. 7).

6 ACKNOWLEDGMENTS

7 REFERENCES

[Bal19] Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging.

[Bee03] Beer, F. Preventing complications of central venous catheterization. The New England journal of medicine 348 (2003), 2684-6, author reply 2684.

[Bov20] Boveiri, HR., Khayami, R., Javidan, R., Mehdizadeh, A. (2020). Medical Image Registration Using Deep Neural Networks: A Comprehensive Review.

[Epp18] Eppenhof, K., Lafarge, M., Moeskops, P., Veta, M., Pluim, J., (2018). Deformable image registration using convolutional neural networks. 27. 10.1117/12.2292443.

[FuY20] Fu, Y., Lei, Y., Wang, T., Curran, WJ., Liu, T., Yang, X. Deep learning in medical image registration: a review. Phys Med Biol. 2020;65(20):20TR01. Published 2020 Oct 22

[HuY18] Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E. et al. Weakly-supervised convolutional neural networks for multimodal image registration. Med Image Anal. 2018 Oct;49:1-13. doi: 10.1016/j.media.2018.07.002.

[Kon17] Kong, X., Hu, C., Duan, Z. Principal Component Analysis Networks and Algorithms. 1st. Springer Publishing Company, Incorporated, 2017.

[Jad15] Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K. (2015). Spatial Transformer Networks. Advances in Neural Information Processing Systems 28 (NIPS 2015).

[Jol16] Jolliffe, I., Cadima, J. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2016).

[Lia17] Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., Comaniciu, D. (2017). An Artificial Agent for Robust Image Registration. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1).

[Mni15] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu AA, Veness J, Bellemare MG, Graves A. et al. Human-level control through deep reinforcement learning. Nature. 2015 Feb 26;518(7540):529-33. doi: 10.1038/nature14236. PMID: 25719670.

[Ron15] Ronneberger, O., Fischer, P., Brox. T. U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS 9351 (2015), pp. 234-241.

[Sal19] Salehi, M., Khan S, Erdogmus D, Gholipour A. Real-Time Deep Pose Estimation With Geodesic Loss for Image-to-Template Rigid Registration. IEEE Trans Med Imaging. 2019;38(2):470-481. doi:10.1109/TMI.2018.2866442

[Sen18] Sentker, T., Madesta, F., Werner, R. (2018). GDL-FIRE4D: Deep Learning-Based Fast 4D CT Image Registration.10.1007/978-3-030-00928-1.

[Yan18] Yan, P., Xu, S., Rastinehad, A., Wood, B. (2018). Adversarial Image Registration with Application for MR and TRUS Image Fusion.

[Zha18] Zhang, Jun. (2018). Inverse-Consistent Deep Networks for Unsupervised Deformable Image Registration.