Networks for Joint Affine and Non-parametric Image Registration

Zhengyang Shen
UNC Chapel Hill
zyshen@cs.unc.edu

Xu Han
UNC Chapel Hill
xhs400@cs.unc.edu

Zhenlin Xu
UNC Chapel Hill
zhenlinx@cs.unc.edu

Marc Niethammer
UNC Chapel Hill
mn@cs.unc.edu

Affine- vSVF-Mapping (AVSM)

CVPR-2019

Presented by: Hengtao Guo
11/20/2019

Contents

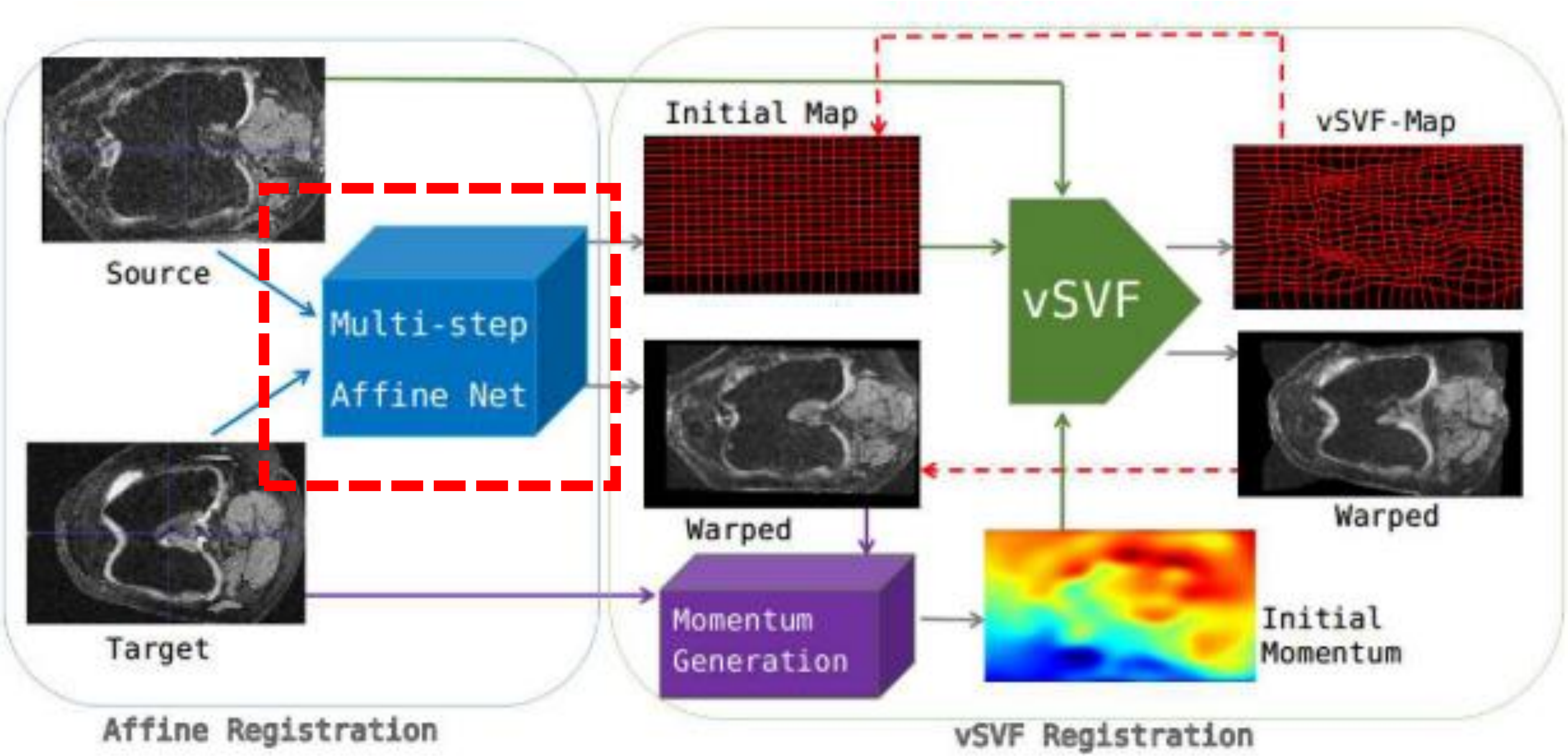
1. Task: longitudinal and cross-subject registrations on MR images (same-modality registration)
2. Joint affine and deformable image registration via separate networks

Motivations

1. Existing deep learning approaches to image registration exhibit multiple limitations:
   a) Assuming the images have already been pre-aligned, e.g., by rigid or affine registration.
   b) Limited by computational memory and hence either only work in 2D or resort to small patches in 3D
   c) do not explore iterative refinement

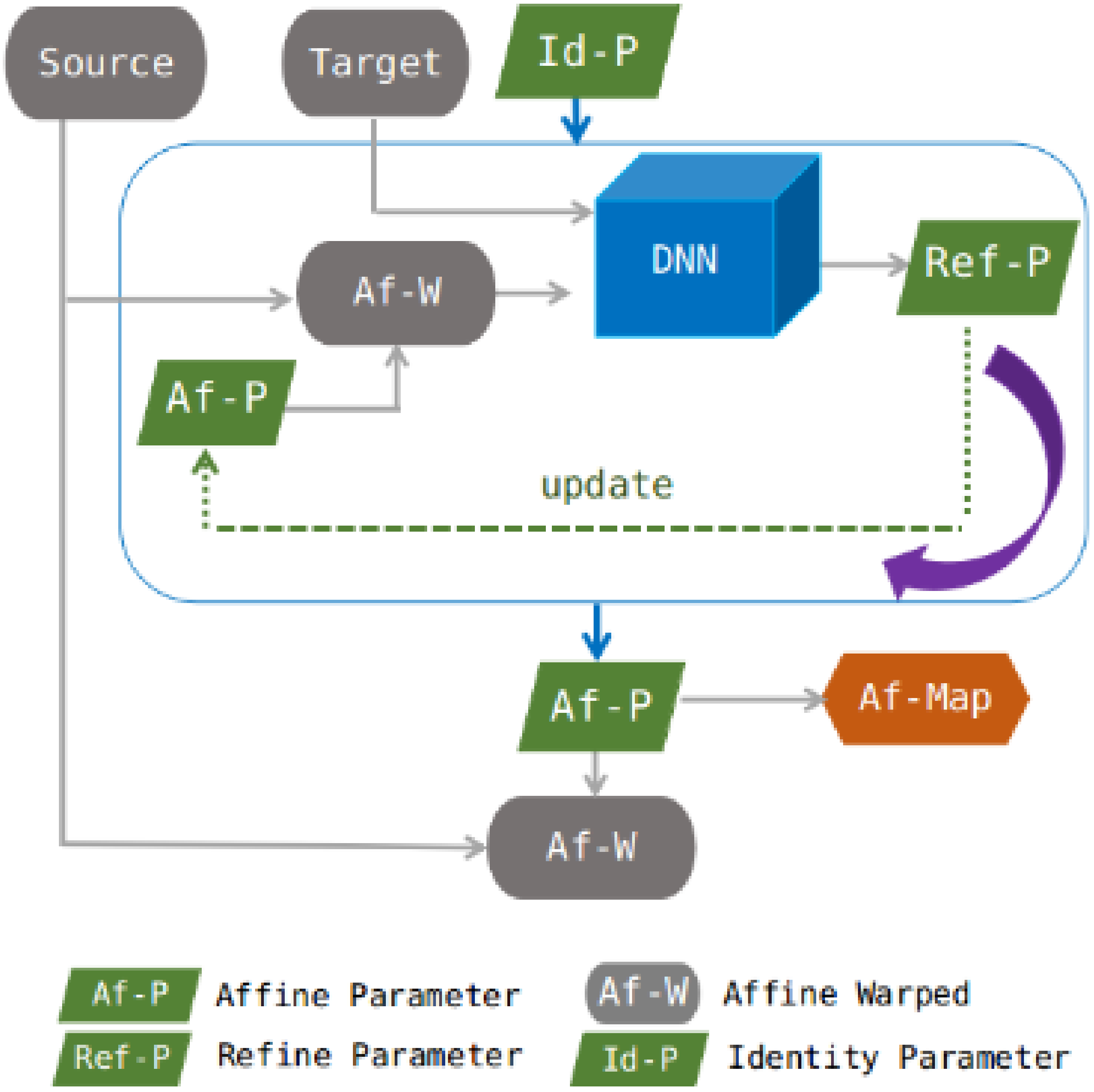
Contributions

1. A novel vector momentum-parameterized stationary velocity field registration model (vSVF)
2. An end-to-end registration method, merging affine and vSVF registration into a single framework.
3. A multi-step approach for the affine and the vSVF reg- istration components
4. An inverse consistency loss both for the affine and the vSVF registration components


Initial Map
vSVF-Map
Source
Multi-step
Affine Net
vSVF
Warped
Warped
Target
Momentum
Generation
Initial
Momentum
Affine Registration
vSVF Registration

Affine Component

1. To account for large, global displacements or rotations
2. This network can be iteratively trained and used, under specific strategies.
3. Trained independently from the vSVF component



Affine Component Loss Functions

1. **Image similarity loss: Primary loss of registration**
2. Localized Normalized Cross Correlation (LNCC)
3. Ns refers to the number of sliding windows with cubic size s×s×s

$$\kappa_s(x,y) = \frac{1}{N_s}\sum_j \frac{\sum_{i\in\zeta_j^s}(x_i-\bar{x}_j)(y_i-\bar{y}_j)}{\sqrt{\sum_{i\in\zeta_j^s}(x_i-\bar{x}_j)^2\sum_{i\in\zeta_j^s}(y_i-\bar{y}_j)^2}}$$

$$\begin{aligned} &L_{a\text{-}sim}(I_0, I_1, \Gamma) = \sum_i \omega_i \kappa_{s_i}(I_0 \circ \Phi_a^{-1}, I_1), \\ &\text{s.t.} \quad \Phi_a^{-1}(x,\Gamma) = Ax + b \ \text{ and } \sum \omega_i = 1, w_i \geq 0. \end{aligned} \tag{3}$$

Affine Component Loss Functions

1. **Regularization Loss: Force the affine to be smooth at the begining**
2. Penalizes deviations of the composed affine transform from the identity
3. Ns refers to the number of sliding windows with cubic size s×s×s
4. where $\|\cdot\|_F$ denotes the Frobenius norm and λar ≥ 0 is an epoch-dependent weight factor designed to be large at the beginning of the training to constrain large deformations and then gradually decaying to zero.

$$L_{a\text{-}reg}(\Gamma) = \lambda_{ar}(||A - I||_F^2 + ||b||_2^2),$$

Affine Component Loss Functions

1. **Symmetry Loss: Inverse the source and target, force the transformation to be inverse as well**
2. Encourage that the transformation computed from source to target im- age is the inverse of the transformation computed from the target to the source image

$$(i.e., A^{ts}(Ax + b) + b^{ts} = x)$$

$$L_{a\text{-}sym}(\Gamma, \Gamma^{ts}) = \lambda_{as}(||A^{ts}A - I||_F^2 + ||A^{ts}b + b^{ts}||_2^2), \quad (5)$$

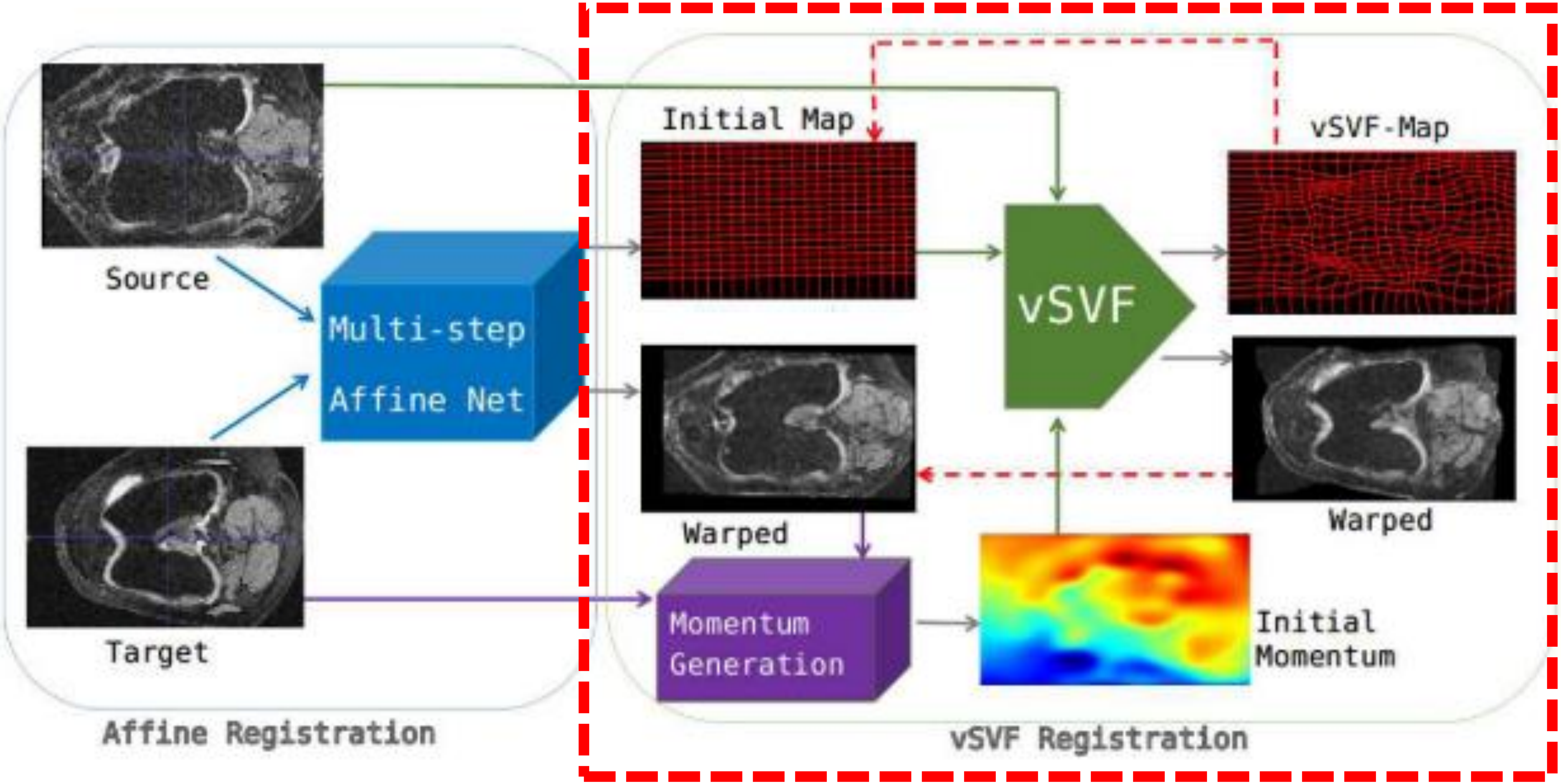
where $\lambda_{as} \geq 0$ is a chosen constant.

Affine Component Loss Functions

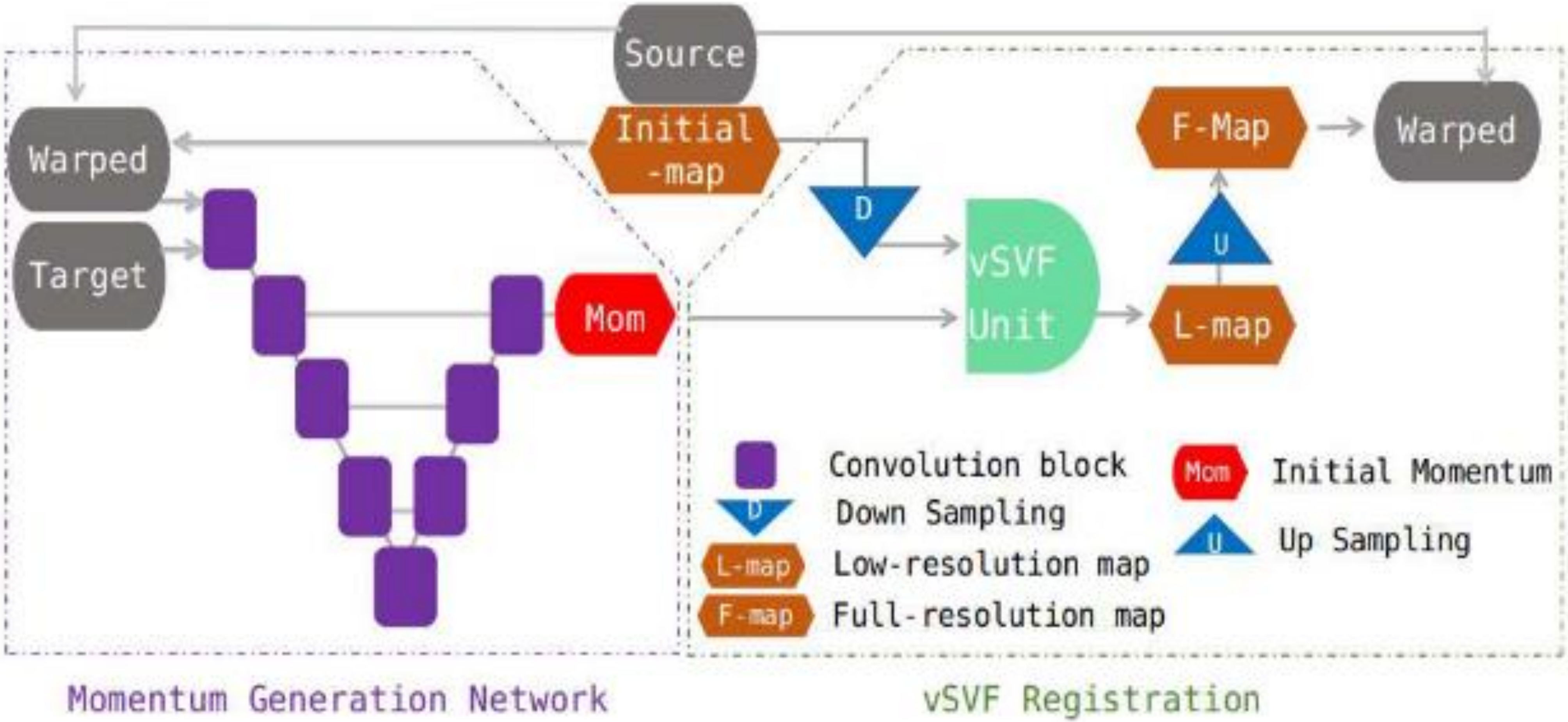
The **complete loss** $\mathcal{L}_a(I_0, I_1, \Gamma, \Gamma^{ts})$ is then:

$$\begin{aligned}\mathcal{L}_a(I_0, I_1, \Gamma, \Gamma^{ts}) =& \ell_a(I_0, I_1, \Gamma) + \ell_a(I_1, I_0, \Gamma^{ts}) \\ &+ L_{a\text{-}sym}(\Gamma, \Gamma^{ts}),\end{aligned}$$

where $\ell_a(I_0, I_1, \Gamma) = L_{a\text{-}sim}(I_0, I_1, \Gamma) + L_{a\text{-}reg}(\Gamma)$.

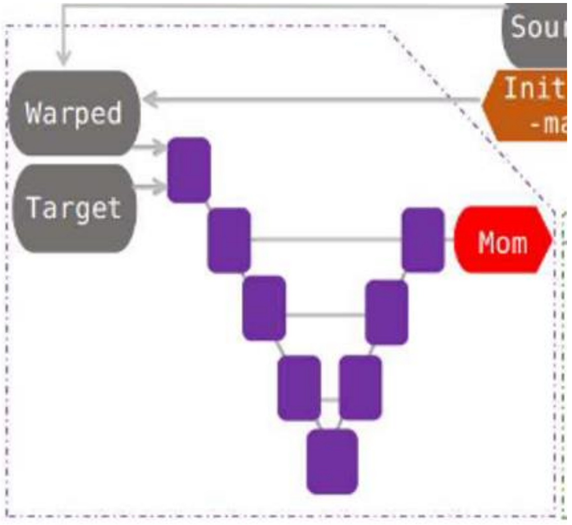

Source
Target
Multi-step
Affine Net
Affine Registration
Initial Map
vSVF
vSVF-Map
Warped
Warped
Momentum
Generation
Initial
Momentum
vSVF Registration

Vector momentum-parameterized stationary velocity field (vSVF) Component

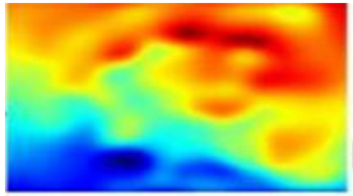


Momentum Generation Network

1. **Input:** Affine-warped source image and target image pair
2. **Output:** Low-resolution momentum
3. **Structure:** 4-level U-net structure



Momentum Generation Network



Initial Momentum

Momentum Generation Network

1. Three loss terms, similar to the affine component:
2. **Similarity loss:**

$$L_{v\text{-}sim}(I_0, I_1, \Phi^{-1})$$

3. **Regularization loss:**

$$L_{v\text{-}reg}(m_0) = \lambda_{vr}||v||_L^2 = \lambda_{vr}\langle m_0, v_0\rangle,$$

where $v_0 = (L^\dagger L)^{-1} m_0$. We implement $(L^\dagger L)^{-1}$ as a convolution with a multi-Gaussian kernel [21].

4. **Symmetric loss:**

$$L_{v\text{-}sym}(\Phi^{-1}, (\Phi^{ts})^{-1}) = \lambda_{vs}||\Phi^{-1} \circ (\Phi^{ts})^{-1} - id||_2^2,$$

Momentum Generation Network

The **complete loss** $\mathcal{L}_v(I_0, I_1, \Phi^{-1}, (\Phi^{ts})^{-1}, m_0, m_0^{ts})$ for vSVF registration with one step is as follows:

$$\begin{aligned}\mathcal{L}_v(I_0, I_1, \Phi^{-1}, (\Phi^{ts})^{-1}, m_0, m_0^{ts}) &= \ell_v(I_0, I_1, \Phi^{-1}, m_0) \\ &+ \ell_v(I_1, I_0, (\Phi^{ts})^{-1}, m_0^{ts}) \\ &+ L_{v\text{-}sym}(\Phi^{-1}, (\Phi^{ts})^{-1}),\end{aligned} \tag{11}$$

where:

$$\ell_v(I_0, I_1, \Phi^{-1}, m_0) = L_{v\text{-}sim}(I_0, I_1, \Phi^{-1}) + L_{v\text{-}reg}(m_0).$$

For the vSVF model with T steps, the complete loss is:

$$\begin{aligned}&\sum_{\tau=1}^{T} \mathcal{L}_v(I_0, I_1, \Phi_{(\tau)}^{-1}, \Phi_{(\tau)}^{ts\ -1}, m_{0(\tau)}, m_{0(\tau)}^{ts}) \quad \text{s.t.} \\ &\Phi_{(\tau)}^{-1}(x, 0) = \Phi_{(\tau-1)}^{-1}(x, 1), \\ &(\Phi_{(\tau)}^{ts})^{-1}(x, 0) = (\Phi_{(\tau-1)}^{ts})^{-1}(x, 1).\end{aligned} \tag{12}$$

Experiments

Dataset: The Osteoarthritis Initiative (OAI) dataset consists of 176 manually labeled magnetic resonance (MR) images from 88 patients (2 longitudinal scans per patient) and 22,950 unlabeled MR images from 2,444 patients. Labels are available for femoral and tibial cartilage. All images are of size $384 \times 384 \times 160$, where each voxel is of size $0.36 \times 0.36 \times 0.7mm^3$. We normalize the intensities of each image such that the 0.1th percentile and the 99.9th percentile are mapped to $0, 1$ and clamp values that are smaller to 0 and larger to 1 to avoid outliers. All images are down-sampled to size $192 \times 192 \times 80$.

Training Multi-step Affine Net

1. Difficult to train the multi-step affine network from scratch
2. Train a single-step network first and use its parameters to initialize the multi-step network.
3. For longitudinal registration, we train with a three-step affine network, but use a seven-step network during testing.
4. For cross- subject registration we train with a five-step network and test with a seven-step one.
5. The affine symmetry factor λas is set to 10.

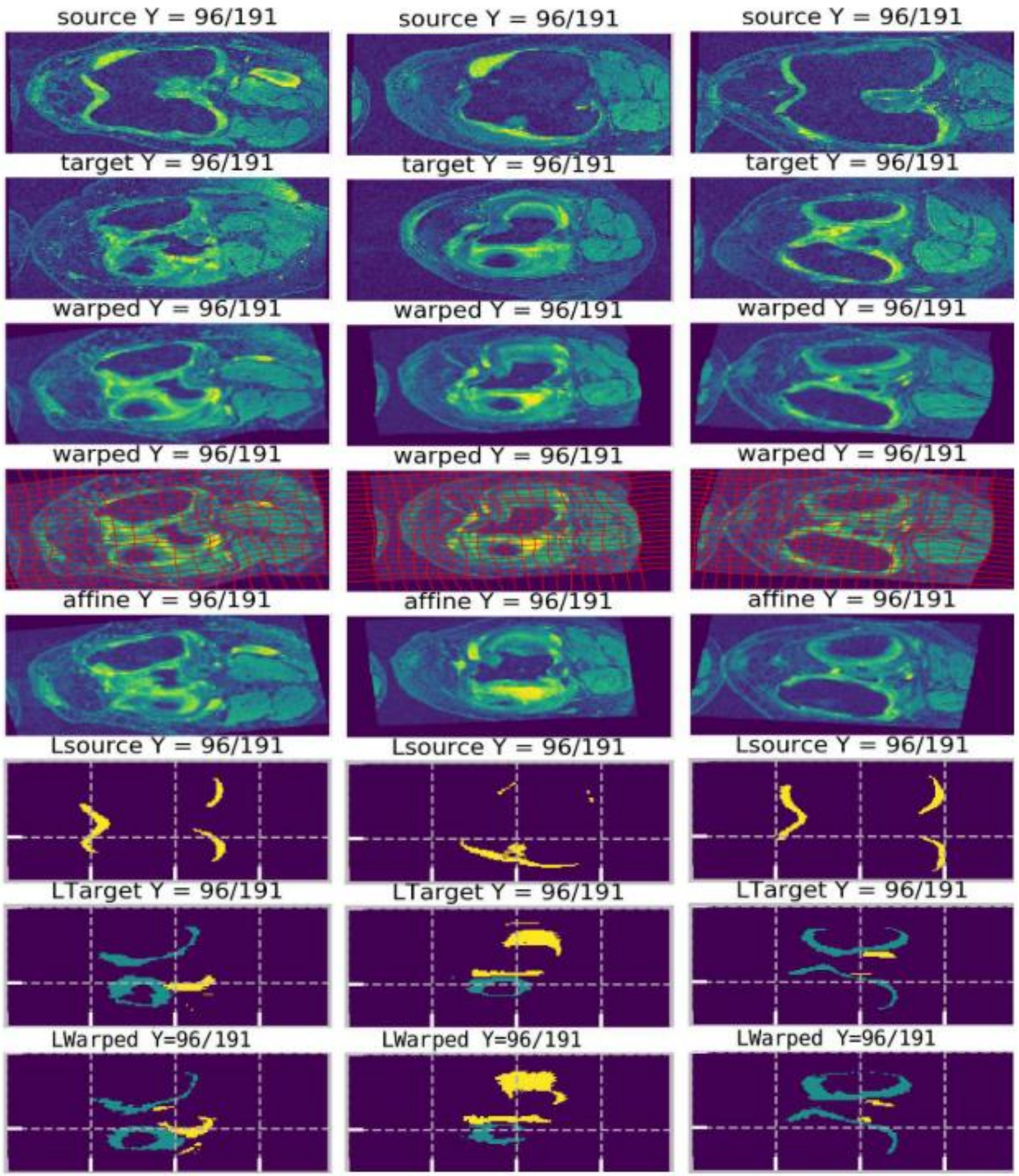
Training Momentum Generation Network

1. Use 10 time-steps and a multi-Gaussian kernel with standard deviations {0.05, 0.1, 0.15, 0.2, 0.25} and corresponding weights {0.067, 0.133, 0.2, 0.267, 0.333}
2. Train with two steps for both longitudinal and cross-subject registrations

Additionally, the affine regularization factor λ_{ar} is *epoch-dependent* during training and defined as:

$$\lambda_{ar} := \frac{C_{ar} K_{ar}}{K_{ar} + e^{n/K_{ar}}}, \tag{13}$$

Results


source Y = 96/191
target Y = 96/191
warped Y = 96/191
warped Y = 96/191
affine Y = 96/191
Lsource Y = 96/191
LTarget Y = 96/191
LWarped Y=96/191

Method	Longitudinal		Cross-subject		
	Dice	Folds	Dice	Folds	Time (s)
affine-NiftyReg	75.07 (6.21)	0	30.43 (12.11)	0	45
affine-opt	**78.61** (4.48)	0	34.49 (18.07)	0	8
affine-net (7-step)	77.75 (4.77)	0	**44.58** (7.74)	0	0.20
Demons	**83.43** (2.64)	10.7 [0.56]	63.47 (9.52)	19.0 [0.56]	114
SyN	83.13 (2.67)	0	65.71 (15.01)	0	1330
NiftyReg-NMI	83.17 (2.76)	0	59.65 (7.62)	0	143
NiftyReg-LNCC	83.35 (2.70)	0	67.92 (5.24)	203.3 [35.19]	270
vSVF-opt	82.99 (2.68)	0	67.35 (9.73)	0	79
VoxelMorph(w/o aff)	71.25 (9.54)	2.72 [1.57]	46.06 (14.94)	83.0 [18.13]	0.12
VoxelMorph(with aff)	82.54 (2.78)	5.85 [0.59]	66.08 (5.13)	39.0 [3.31]	0.31
AVSM (2-step)	82.60 (2.73)	0	67.59 (4.47)	5.5 [0.39]	0.62
AVSM (3-step)	82.67 (2.74)	3.4 [0.12]	**68.40** (4.35)	14.3 [1.07]	0.83

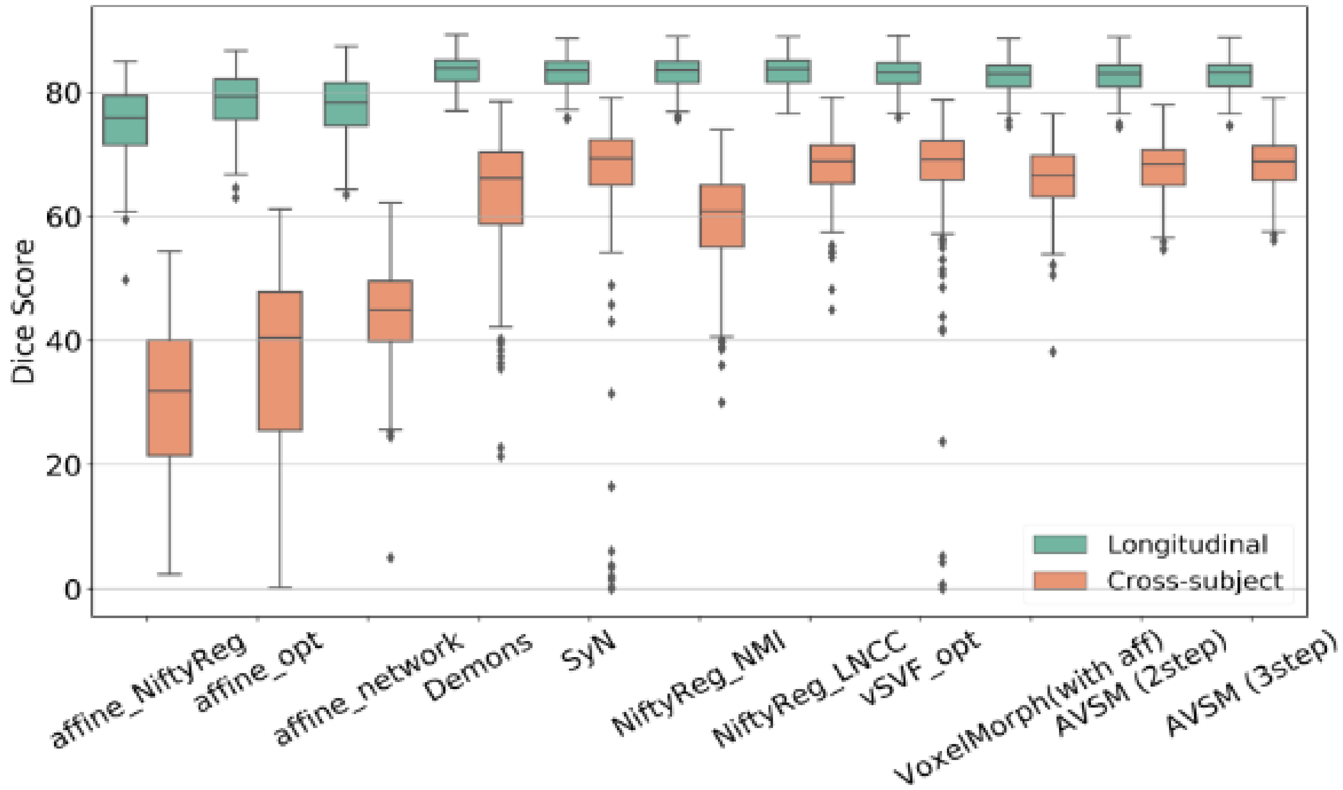


Figure 5. Box-plots of the performance of the different registration methods for longitudinal registration (green) and cross-subject registration (orange). Both AVSM and NiftyReg (LNCC) show high performance and small variance.

Method	Af-Reg	Af-Multi	Af-Sym	Af-MK	vSVF	vSVF-MK	vSVF-Multi	vSVF-Sym	Longitudinal	Better?	Cross-subject	Better?
I									-		-	
II	✓								55.41	✓	28.68	✓
III	✓	✓							64.78	✓	36.31	✓
IV	✓	✓	✓						68.87	✓	37.54	✓
V	✓	✓	✓	✓					77.75	✓	44.58	✓
VI	✓	✓	✓		✓				80.71	✓	59.21	✓
VII	✓	✓	✓	✓	✓	✓			81.64	✓	64.56	✓
VIII	✓	✓	✓	✓	✓	✓	✓		82.81	✓	69.08	✓
IV	✓	✓	✓	✓	✓	✓	✓	✓	82.67	✗	68.40	✗

Table 2. Ablation study of AVSM using different combinations of methods. **Af-** and **vSVF-** separately refer to the affine and to the vSVF related methods; *Reg* refers to adding epoch-dependent regularization; *Multi* refers to multi-step training and testing; *Sym* refers to adding the symmetric loss; *MK* refers to using mk-LNCC as similarity measure (default NCC). Except for the last approach which uses *vSVF-Sym* (last row) and encourages symmetric vSVF solutions, all other approaches result in performance improvements.

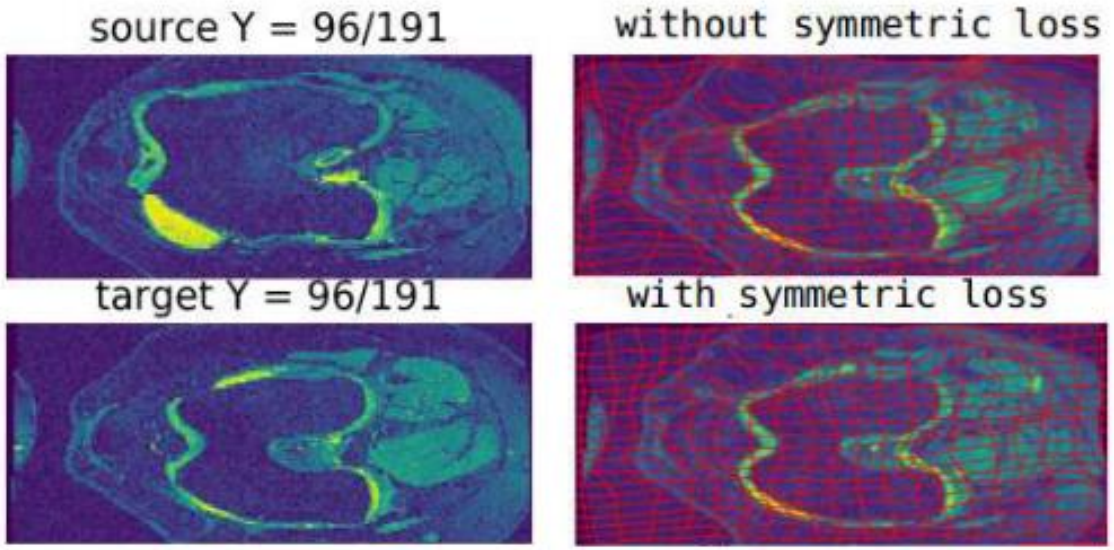


Figure 6. Illustration of symmetric loss for AVSM. The left column shows the source and target images. The right column shows the warped image from a network trained with and without symmetric loss. The deformation with symmetric loss is smoother.

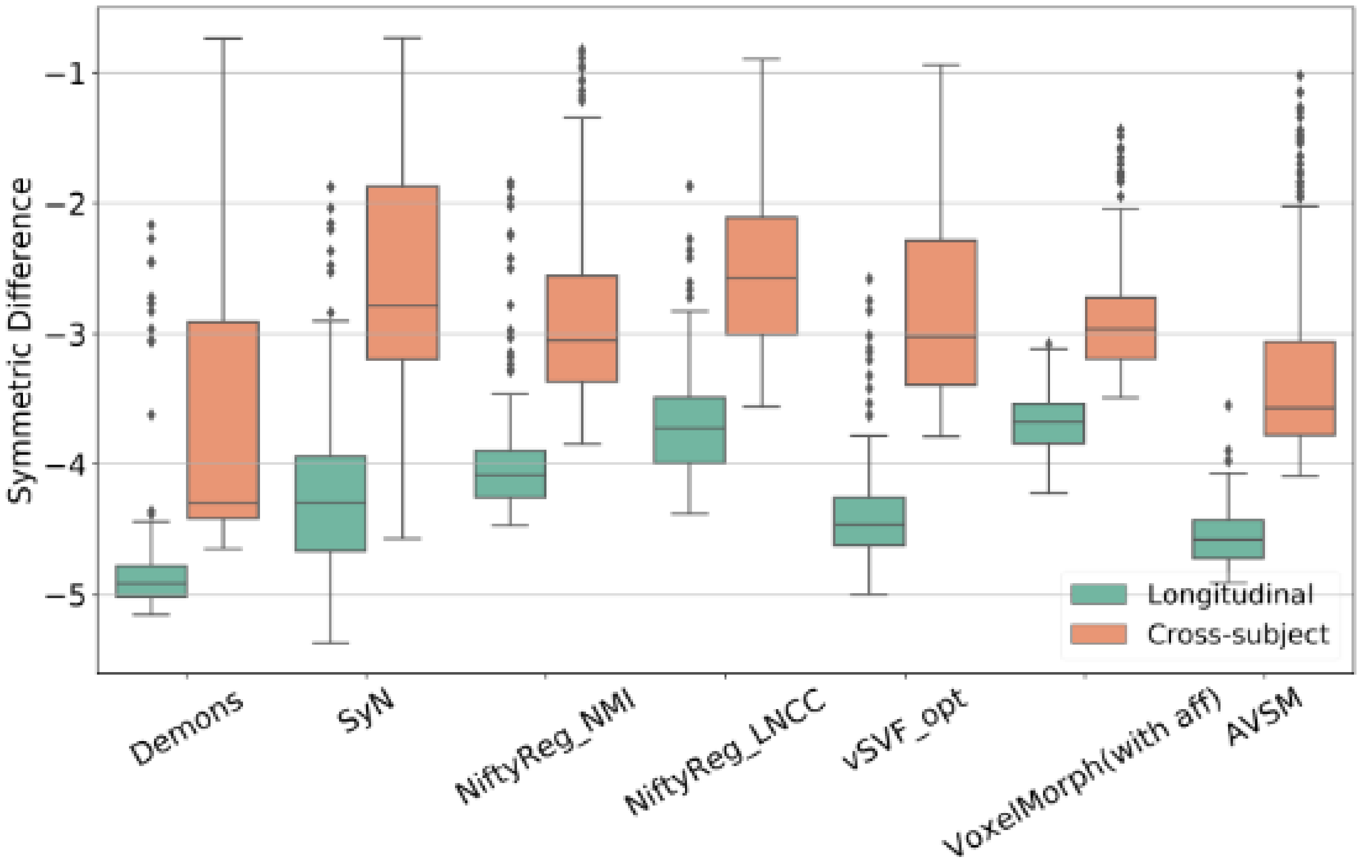


Figure 7. Box-plots of the symmetry evaluation (the lower the better) of different registration methods for longitudinal registration (green) and cross-subject registration (orange). AVSM (tested with two-step vSVF) shows good results.

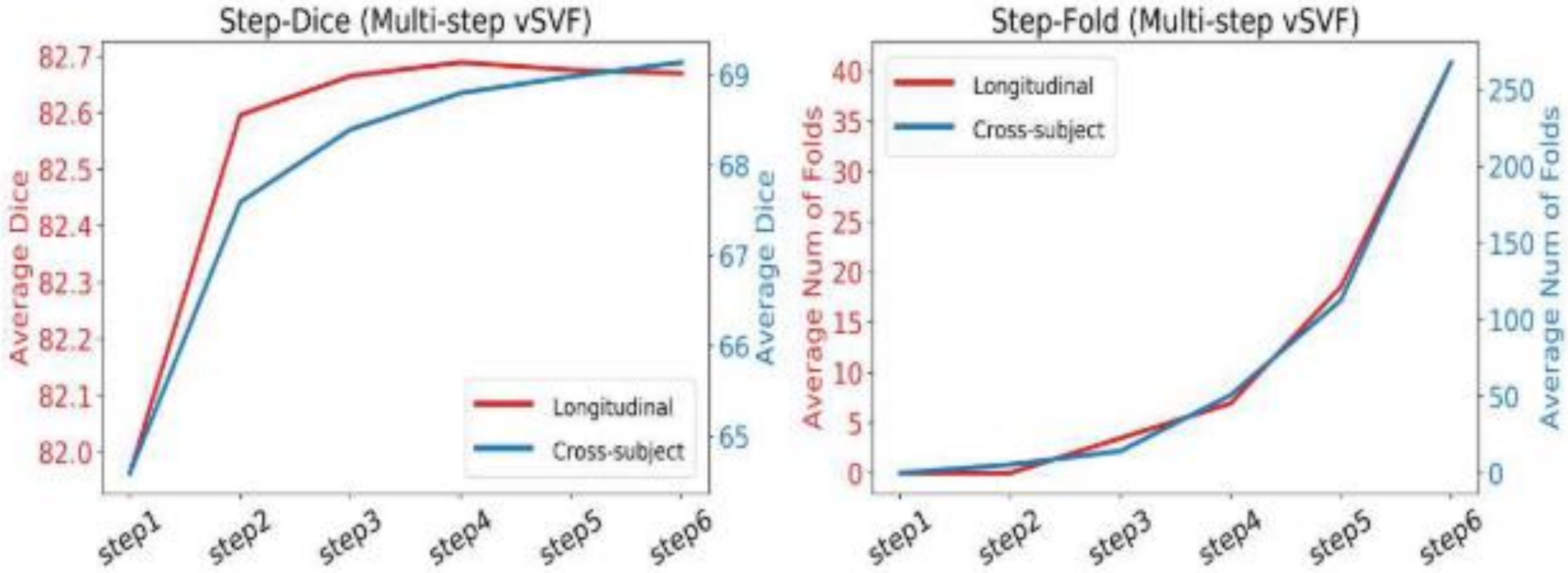


Figure 8. Multi-step vSVF registration results for two-step vSVF training. Performance increases with steps (left), but the number of folds also increases (right).

Conclusions

1) Introduced an end-to-end 3D image registration approach (AVSM) consisting of a multi-step affine network and a deformable registration network using a momentum-based SVF algorithm.
2) AVSM outputs a transformation map which includes an affine pre-registration and a vSVF non- parametric deformation in a single forward pass.
3) Future work will focus on also learning regularizers and evaluations on other registration tasks, e.g. in the brain and the lung.