Dataset and implementation details

- **ACDC (2D MRI) [4]**. We randomly selected 80 patients for training, 20 for validation, and 50 for testing. ED and ES image pairs are extracted from each sequence in a slice-by-slice manner from the longitudinal stacks. We center crop each slice pair to 128×128 w.r.t. myocardium centroid, yielding 751 image pairs for training, 200 for validation, and another 538 for testing.
- **CAMUS (2D Echo) [14]**. We resize each image pair to size 128×128 and randomly select 300 subjects for training, 100 subjects for validation, and 100 subjects for testing. This yields in total 600 image pairs for training, 200 pairs for validation, and 200 pairs for testing.
- **Private 3D Echo**. The private 3D echo dataset contains 99 cardiac ultrasound scans. ED and ES frames are manually identified and myocardium segmentation labels are provided for each sequence by experienced radiologists. Each 3D image is resized to $64 \times 64 \times 64$. We randomly select 60 3D pairs for training, 19 pairs for validation, and another 20 pairs for testing.
- **Implementation details**. All our experiments are conducted under the Pytorch framework and trained on NVIDIA V100/A5000 GPUs. The architecture of the variance estimator is implemented based on a U-Net. We use $\lambda = 0.01$ as the hyperparameter in Eq. 4. Both displacement and variance estimators are trained with learning rates 1×10^{-4} for 300 epochs.

Incorporating displacement uncertainty. To further demonstrate the versatility, we conducted a direct extension by simultaneously estimating heteroscedastic displacement uncertainty with the isotropic assumption. We add an additional layer in the displacement estimator to predict $\hat{\sigma}_z$, where $\hat{\sigma}_z(x) \in \mathbb{R}$ and the original prediction as displacement mean $\hat{\mu}_z$. We train our proposed displacement estimator using objective:

$$\mathcal{L}_\theta = \mathbb{E}_\Omega \left[\mathcal{T} \left[\left(\frac{I_f}{[\hat{\sigma}_I]} \right)^{2\gamma} \right] [I_f - \hat{I}_f]^2 + \alpha \left(\hat{\sigma}_z^2 - \log \hat{\sigma}_z^2 \right) + \lambda \|\nabla \hat{z}\|^2 \right]$$ derived from Eq. 1, with $\hat{z}$ sampled from distribution $\hat{z} \sim \mathcal{N}(\hat{\mu}_z, \hat{\sigma}_z^2 \mathbb{I})$ during training with reparameterization trick. We compare the quality of our predicted displacement along with its uncertainty estimate $\hat{\sigma}_z^2$ with vxm-diff [6].

We present our quantitative results in Table 5, illustrating the superiority of our formulation. We further present the qualitative visualization as shown in Fig. 4, demonstrating that our estimated heteroscedastic uncertainty $\hat{\sigma}_z^2$ accurately captures the randomness in the displacement prediction more accurately.

Additional private 3D Echo results. We present our qualitative result in Fig. 5 for registration accuracy and left Fig. 6 for noise heteroscedastic variance evaluation. We also quantitatively evaluate the result by repeating the sparsification error plot similar to Section 5.3. We observe that our predicted $\hat{\sigma}_I$ achieves a better error curve than β-NLL and NLL, which is consistent with our main results shown in Fig. 3.

Failure case. We present an example shown in Fig. 7 that all methods fail to match myocardium when it is considerably thin.

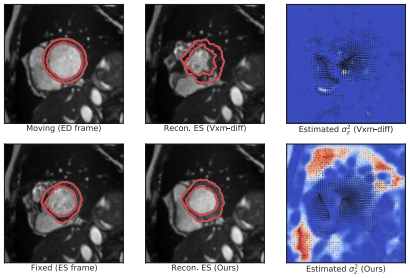

Fig. 4: Comparison of $\hat{\sigma}_z^2$ between our vxm-based framework and vxm-diff [6].

Table 5: Our method raises the upper bound on registration accuracy while providing useful displacement uncertainty estimates $\hat{\sigma}_z$.

	Uncertainty		ACDC [4]			CAMUS [14]		
	σ_z^2	σ_I^2	DSC ↑	HD ↓	ASD ↓	DSC ↑	HD ↓	ASD ↓
Vxm [3]	✗	✗	80.20	4.64	1.24	81.76	8.93	1.70
Vxm-diff [6]	✓	✗	76.19	5.75	**1.19**	76.74	10.76	1.88
Ours	✓	✗	79.80	4.74	1.22	81.47	8.67	1.69
Ours	✗	✓	**80.73**	**4.57**	1.21	**81.96**	8.80	1.66
Ours	✓	✓	79.87	4.62	1.20	81.91	**8.54**	**1.65**

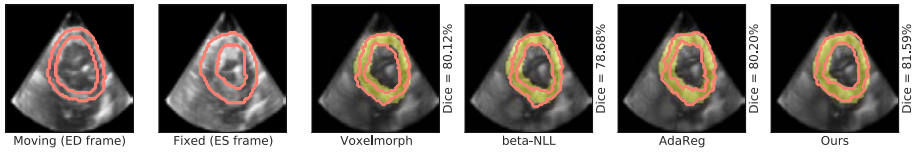

Fig. 5: Qualitative evaluation for our private 3D Echo dataset on voxelmorph architecture. We extract cross-sectional slices from the 3D volume for visualization. We overlay ground truth segmentation in yellow for comparison.

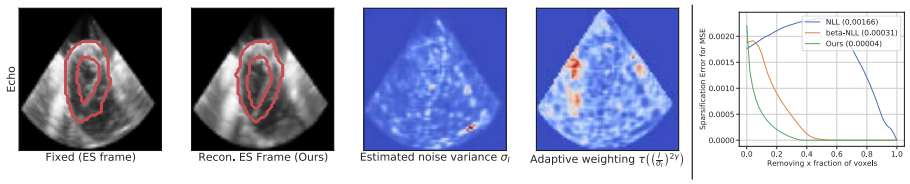

Fig. 6: Left: Estimated $\hat{\sigma}_I^2$ and the corresponding weighting map of our proposed framework under Voxelmorph architecture [3] using our private 3D Echo dataset. Right: Sparsification error plots of $\log \hat{\sigma}_I^2$ on our private 3D Echo dataset.

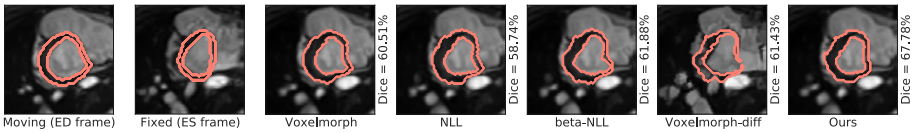

Fig. 7: An example of failure case on ACDC dataset.