CVPR2024 Oral: Correlation-aware Coarse-to-fine MLPs for Deformable Medical Image Registration

Mingyuan Meng[1,2], Dagan Feng[1], Lei Bi[2], and Jinman Kim[1]

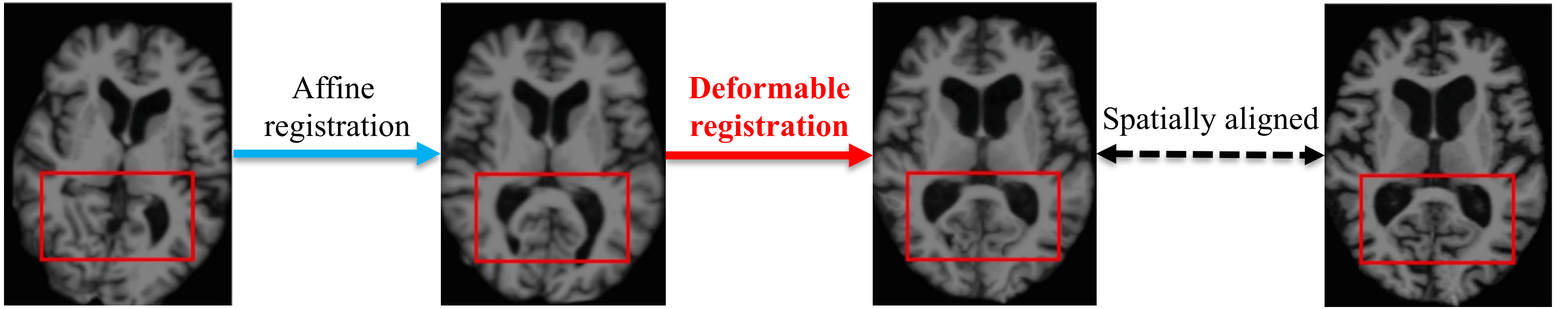
[1] School of Computer Science, The University of Sydney, Australia.

[2] Institute of Translational Medicine, Shanghai Jiao Tong University, China.

Background

– Medical image registration

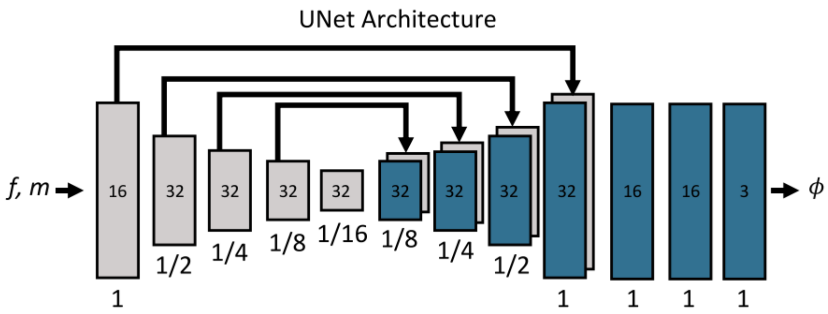
- Image registration finds a spatial transformation between a pair of fixed and moving images, through which the moving image can be warped to spatially align with the fixed image.
- Affine registration is firstly performed to eliminate the linear and large spatial misalignment between images. **Then, deformable registration is performed to reduce the local non-rigid deformations, which is the main research focus for medical image registration.**



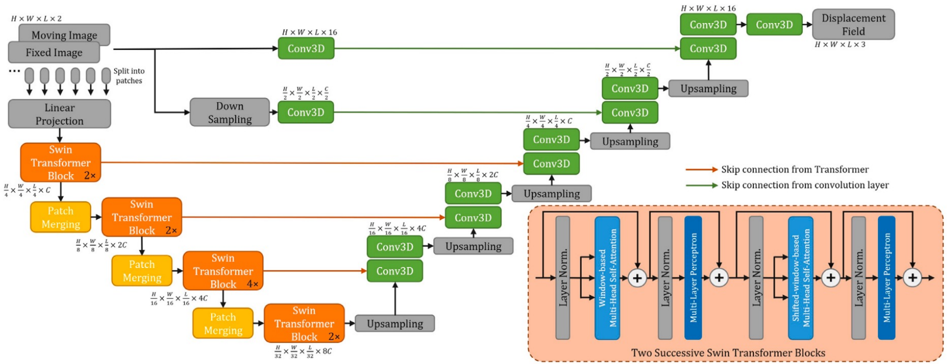
Moving image | Warped image after affine registration | Warped image after deformable registration | Fixed image

Literature Review

- Vision backbones for deformable registration: From CNN to Transformer
  - CNN backbone: VoxelMorph, Diffeomorphic VoxelMorph
  - Transformer backbone: TransMorph, Swin-VoxelMorph, TransMatch



CNN-based VoxelMorph



Transformer-based TransMorph

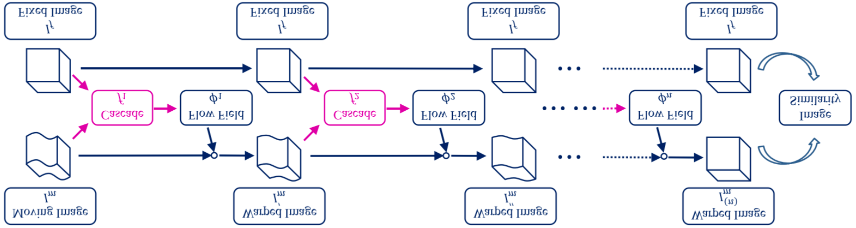
Literature Review

CVPR JUNE 17-21, 2024 SEATTLE, WA

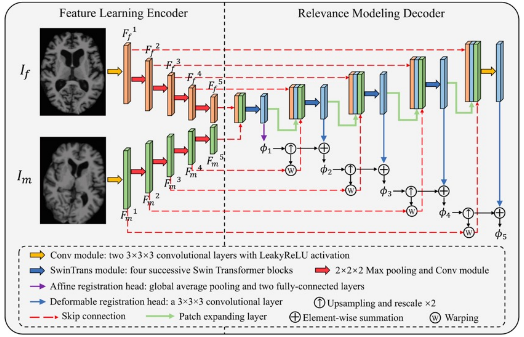
THE UNIVERSITY OF SYDNEY

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

- Architectures for deformable registration: From direct registration to progressive (coarse-to-fine) registration
  - Progressive registration architectures perform **multiple steps of coarse-to-fine registration**.
  - Iterative coarse-to-fine methods use cascaded networks or run a single network with multiple iterations to perform the multiple registration steps, such as RCN, LapIRN, ULAE-net.
  - Non-iterative coarse-to-fine methods perform multiple registration steps by running a single pyramid network for a single iteration, such as NICE-Net, NICE-Trans, ModeT.



Iterative coarse-to-fine method (RCN)



Non-iterative coarse-to-fine method (NICE-Tra...

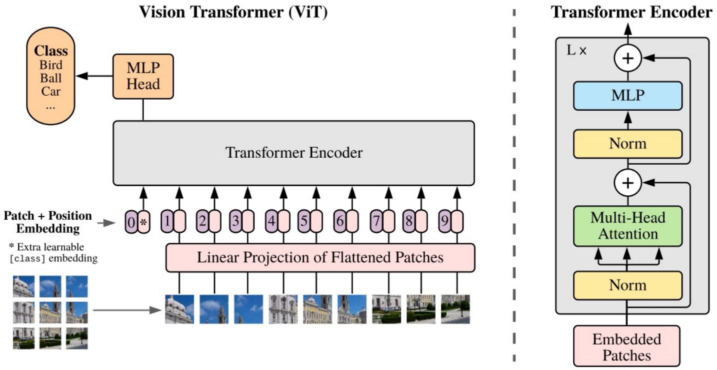
Literature Review

CVPR JUNE 17-21, 2024 SEATTLE, WA

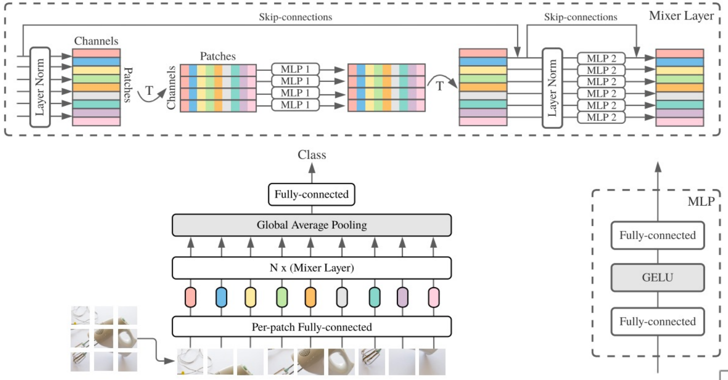
THE UNIVERSITY OF SYDNEY

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

- Vision backbones based on Multi-layer Perceptrons (MLPs)
- MLPs capture long-range dependence without relying on self-attention, showing advantages over transformers on computation and memory consumption.
- MLPs can process high-resolution image feature maps to capture fine-grained long-range dependence at full resolution, which is crucial for medical image dense prediction (Ref: *arXiv:2311.16707, Full-resolution MLPs Empower Medical Dense Prediction*)



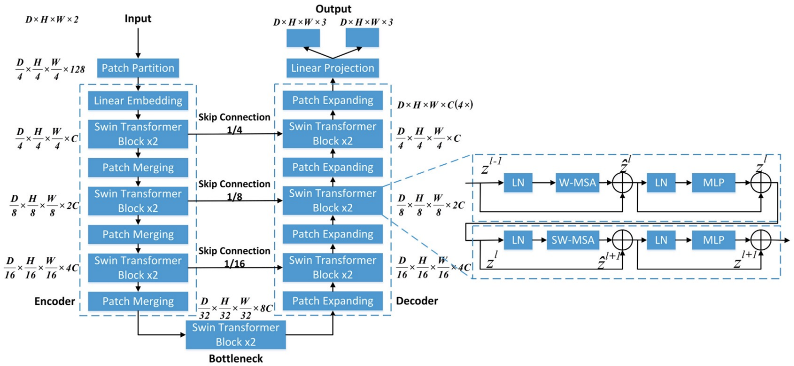
Vision Transformer (ViT)



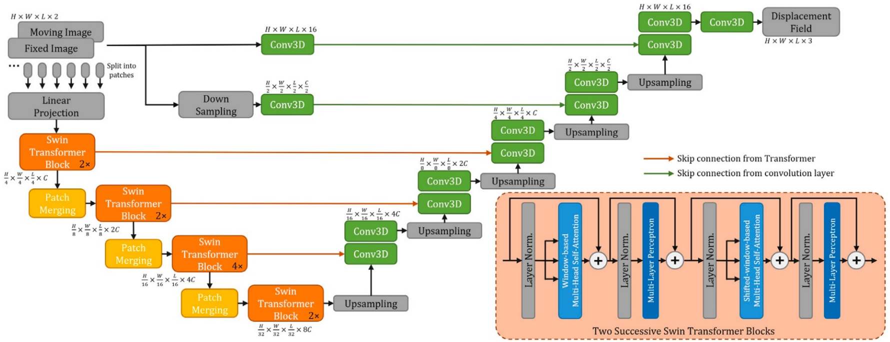
MLP-Mixer

Limitations and Motivation

- Transformer cannot capture fine-grained long-range dependence at the full image resolution. This limits the registration performance as deformable registration necessitates precise dense correspondence between each image pixel.
- MLPs enable the feasibility of modeling fine-grained long-range dependence at full resolution.



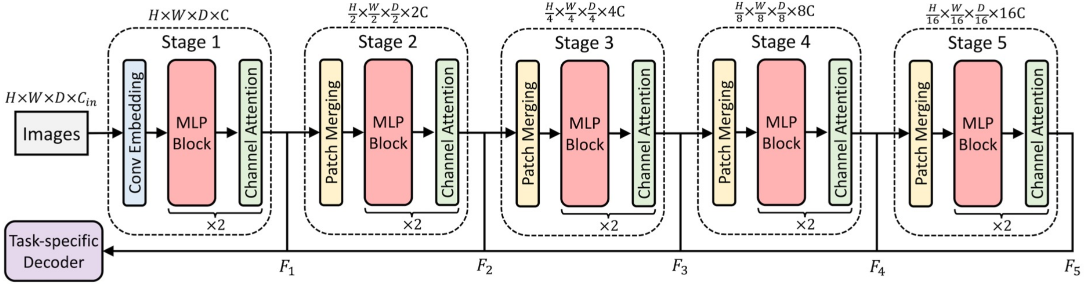
Swin-VoxelMorph (Pure transformer)



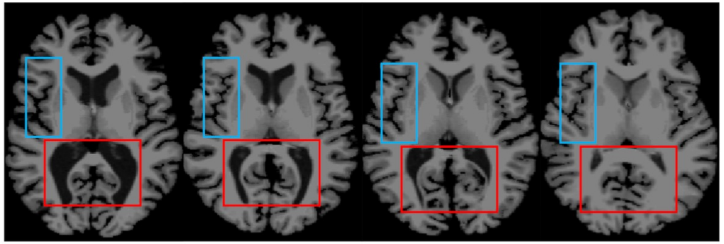
TransMorph (Hybrid CNN-transformer)

Limitations and Motivation

- MLPs have not been extensively explored for image registration and lack the consideration of inductive bias crucial for medical image registration tasks.
- Existing MLP-based models tend to simply stack MLP blocks as the feature-extraction encoder and have not been optimized in the state-of-the-art coarse-to-fine architecture for progressive medical image registration.
- Existing MLP blocks (i) tend to mix spatial information globally and are insensitive for local-range dependence, and (ii) do not explicitly model the local correlations between features.



Exemplified MLP-based model

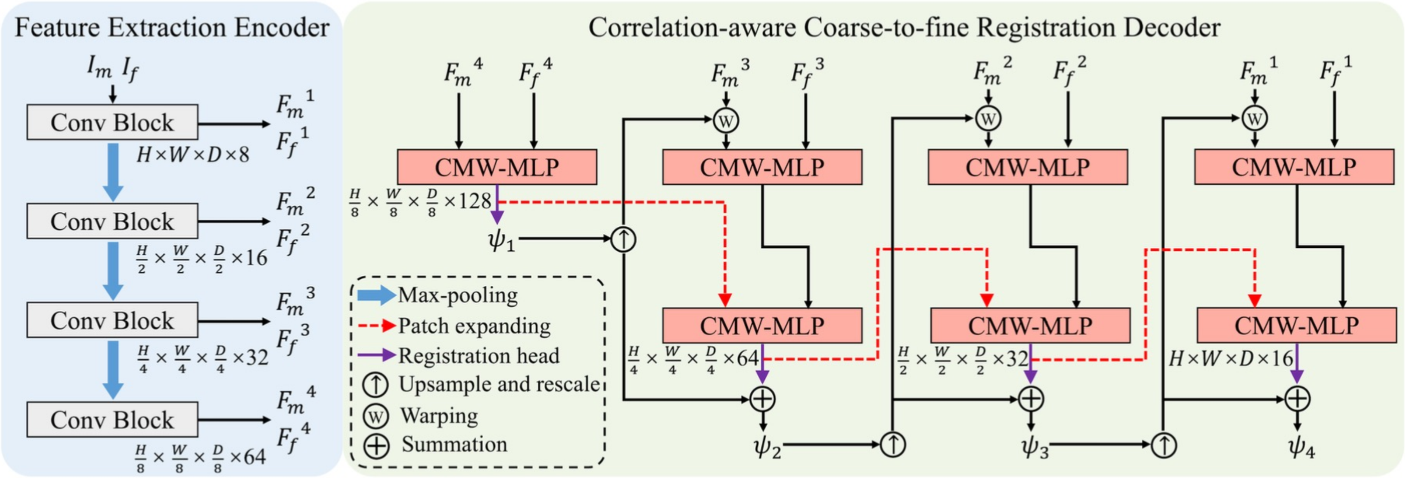


Examples of large (red) and small (blue) local deformations in medical images.

Method

CorrMLP: a correlation-aware coarse-to-fine MLP-based network for deformable registration

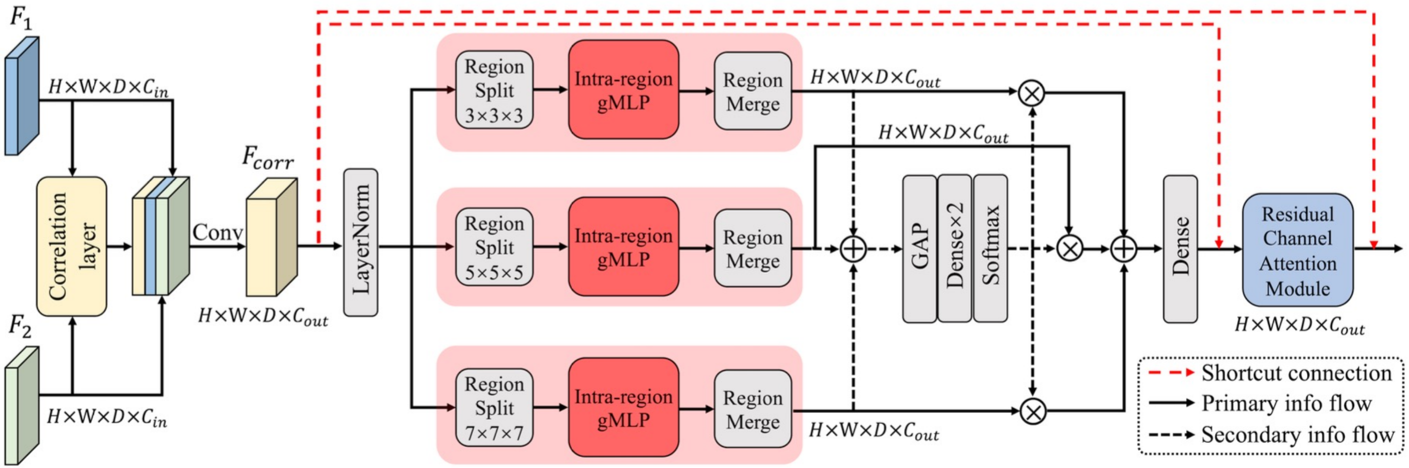
- CNN-based hierarchical encoder to extract two feature pyramids
- Correlation-aware coarse-to-fine registration decoder based on our CMW-MLP blocks
- A novel correlation-aware coarse-to-fine registration architecture that considers both image-level and step-level correlations



Method

CMW-MLP: correlation-aware multi-window MLP block

- Take two sets of feature maps and explore the potential correspondence between them.
- A 3D correlation layer to calculate the local correlations between two feature maps, followed by a multi-window MLP module to capture correlation-aware multi-range dependence to handle both large and small local deformations.



Summary of Technical Contributions

- The technical contributions are three-fold:

- We investigate in leveraging MLPs for deformable medical image registration and propose the CorrMLP, to the best of our knowledge, which is the first MLP-based coarse-to-fine registration network.

- We propose the CMW-MLP block, an MLP block specifically optimized for deformable registration to capture correlation-aware multi-range dependence.

- We propose a novel correlation-aware coarse-to-fine registration architecture that considers both image-level and step-level correlations to provide enriched contextual information to guide each registration step.

Experimental Setup

We evaluated our CorrMLP with two well-benchmarked deformable image registration tasks (3D inter-patient brain image registration and 4D intra-patient cardiac image registration), involving seven public medical datasets:

- Inter-patient brain image registration
  - Training: 2,656 MRI images acquired from ADNI, ABIDE, ADHD, and IXI datasets
  - Testing: Mindboggle and Buckner datasets

- Intra-patient cardiac image registration
  - Training: 100 cine-MRI images from the official training set of ACDC dataset
  - Testing: 50 cine-MRI images from the official testing set of ACDC dataset

Results

Table 1 and Table 2 present the quantitative comparison for brain and cardiac image registration:

- CorrMLP achieved significantly higher DSCs without sacrificing transformation smoothness.
- The runtime of CorrMLP is similar to existing deep registration methods, allowing real-time registration with GPUs (<0.5s for one image pair).

Method		Mindboggle dataset		Buckner dataset		Runtime	
		DSC ↑	NJD (%) ↓	DSC ↑	NJD (%) ↓	CPU (s)	GPU (s)
Before registration		0.347*	/	0.406*	/	/	/
SyN [17]	Traditional	0.534*	1.956	0.567*	1.874	3427	/
NiftyReg [18]	Traditional	0.569*	2.364	0.611*	2.175	159	/
VoxelMorph [7]	CNN, direct	0.552*	2.532	0.589*	2.220	2.84	0.23
Swin-VoxelMorph [13]	Transformer, direct	0.566*	2.254	0.605*	2.016	5.67	0.52
TransMorph [12]	Transformer, direct	0.571*	2.400	0.608*	2.183	3.68	0.35
TransMatch [15]	Transformer, direct	0.578*	2.036	0.622*	1.995	3.06	0.28
LapIRN [9]	CNN, coarse-to-fine	0.605*	2.164	0.632*	2.112	4.97	0.46
ULAE-net [35]	CNN, coarse-to-fine	0.610*	2.000	0.640*	1.940	5.37	0.51
Dual-PRNet++ [32]	CNN, coarse-to-fine	0.608*	2.424	0.636*	2.195	4.61	0.44
SDHNet [36]	CNN, coarse-to-fine	0.598*	1.872	0.634*	1.843	3.24	0.26
NICE-Net [11]	CNN, coarse-to-fine	0.618*	2.043	0.643*	1.963	3.55	0.32
NICE-Trans [22]	Transformer, coarse-to-fine	0.625*	2.324	0.649*	2.277	4.02	0.37
CorrMLP (Ours)	MLP, coarse-to-fine	**0.642**	**1.821**	**0.661**	**1.788**	5.48	0.49

Table 1: Quantitative comparison for brain image registration. The best results in each dataset are in bold. ↑: the higher is better. ↓: the lower is better. *: P<0.05, in comparison to CorrMLP.

Method	ACDC		Runtime	
	DSC ↑	NJD (%) ↓	CPU (s)	GPU (s)
Before registration	0.590*	/	/	/
VoxelMorph [7]	0.754*	0.440	0.36	0.02
Swin-VoxelMorph [13]	0.763*	0.412	0.91	0.08
TransMorph [12]	0.768*	0.492	0.59	0.05
TransMatch [15]	0.770*	0.425	0.55	0.04
MAXIM [29]	0.785*	0.437	1.82	0.17
MAXIM×3 [29]	0.788*	0.716	5.45	0.51
LapIRN [9]	0.790*	0.454	0.77	0.06
ULAE-net [35]	0.792*	0.447	0.86	0.07
Dual-PRNet++ [32]	0.777*	0.479	0.75	0.06
SDHNet [36]	0.789*	0.395	0.45	0.03
NICE-Net [11]	0.785*	0.443	0.49	0.04
NICE-Trans [22]	0.799*	0.473	0.64	0.05
CorrMLP (Ours)	**0.810**	**0.389**	0.83	0.07

Table 2: Quantitative comparison for cardiac image registration. The best results are in bold. ↑: the higher is better. ↓: the lower is better. *: P<0.05, in comparison to CorrMLP.

Ablation Analysis

Table 3 presents an analysis on architecture designs:

- By using MLP block in Unet-style architecture, our baseline MLPMorph has outperformed VoxelMorph and TransMorph by a large margin, demonstrating the superiority of MLPs on deformable image registration: MLPs can capture fine-grained long-range dependence at high-resolution features, which is crucial for finding precise dense correspondence.
- By employing MLP blocks in our correlation-aware coarse-to-fine architecture, CorrMLP outperformed MLPMorph by a large margin. Moreover, separately removing either image- or step-level correlation information degraded the registration performance.

Method	Mindboggle	Buckner	ACDC
VoxelMorph [7]	0.552	0.589	0.754
TransMorph [12]	0.571	0.608	0.768
MLPMorph (Ours)	0.604	0.632	0.780
No correlation	0.628	0.650	0.800
Only image-level correlation	0.637	0.657	0.806
Only step-level correlation	0.634	0.655	0.805
CorrMLP (Ours)	**0.642**	**0.661**	**0.810**

Table 3: DSC results of the ablation study on architecture designs. The best results are in bold.

Ablation Analysis

Table 4 presents an analysis on MLP blocks:

- Replacing the CMW-MLP block with five different existing MLP blocks all resulted in lower DSCs, showing the effectiveness of our CMW-MLP block.
- Even when the correlation layer was removed, the MW-MLP block still outperformed the five existing MLP blocks, implying that our multi-window MLP design is beneficial for deformable registration.
- Removing MLP branches degraded the registration performance; No further improvement by adding an extra MLP branch with $9\times9\times9$ window size. This suggests that a $7\times7\times7$ MLP branch has been sufficient to capture large deformations, while the $3\times3\times3$ and $5\times5\times5$ MLP branches are crucial to capture subtle deformations.

MLP block	Mindboggle	Buckner	ACDC
S^2-MLP [26]	0.621	0.644	0.794
sMLP [27]	0.622	0.645	0.794
Hire-MLP [28]	0.620	0.643	0.793
Swin-MLP [20]	0.624	0.646	0.797
Multi-axis gated MLP [29]	0.625	0.647	0.798
MW-MLP (Ours)	0.628	0.650	0.800
No $3\times3\times3$ MLP branch	0.639	0.657	0.808
No $5\times5\times5$ MLP branch	0.635	0.654	0.805
No $7\times7\times7$ MLP branch	0.637	0.655	0.806
CMW-MLP (Ours)	**0.642**	**0.661**	**0.810**

Table 4: DSC results of the ablation study on MLP blocks. The best results are in bold.

Conclusion

- In this study, we have shown the effectiveness of MLPs for deformable medical image registration by developing the first MLP-based coarse-to-fine registration network (CorrMLP).

- Our study suggests that MLP could be a superior alternative to popular transformers for its advantage on modeling fine-grained long-range dependence at full resolution.

- We suggest that the proposed CMW-MLP block could serve as a general block applying to various network architectures for image registration tasks to leverage its capability to capture correlation-aware multi-range dependence among features.

Thank You

Code: https://github.com/MungoMeng/Registration-CorrMLP

This work was supported in part by Australian Research Council (ARC) under Grant DP200103748.

THE UNIVERSITY OF SYDNEY