CVPR 2024

Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration

Tony C. W. Mok[1,2*] Zi Li[1,2*] Yunhao Bai[1] Jianpeng Zhang[1,2,4] Wei Liu[1,2]
Yan-Jie Zhou[1,2,4] Ke Yan[1,2] Dakai Jin[1] Yu Shi[3] Xiaoli Yin[3] Le Lu[1] Ling Zhang[1]

[1] DAMO Academy, Alibaba Group

[2] Hupan Lab, 310023, Hangzhou, China

[3] Shengjing Hospital of China Medical University, China

[4] College of Computer Science and Technology, Zhejiang University, China

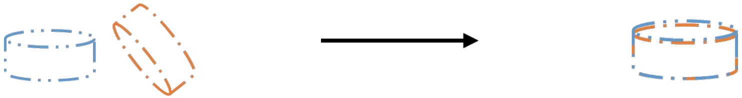
cwmokab@connect.ust.hk

Motivation

Photo by Leeloo The First on Pexels

Image Registration

Modified from Source: Ahmad Hammoudeh & Stéphane Dupont. (2023). Deep Learning in Medical Image Registration: Introduction and Survey. Qeios.



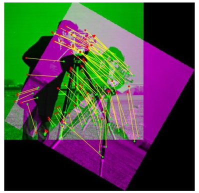
Modified from Source: Ahmad Hammoudeh & Stéphane Dupont. (2023). Deep Learning in Medical Image Registration: Introduction and Survey. Qeios.

- Align images of the same object (scene)
  - taken from different perspectives
  - at different times
  - in different conditions

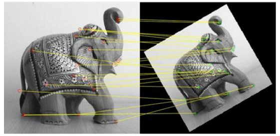
- But also transform different sets of data into one coordinate system

- Applications:
  - Image Fusion
  - Stereo Vision
  - Object Tracking
  - **Medical Image Analysis**

Modified from Source:
MathWorks. Image Registration.
Registering aerial photos using point mapping.
https://de.mathworks.com/discovery/image-registration.html [06/25/2024].



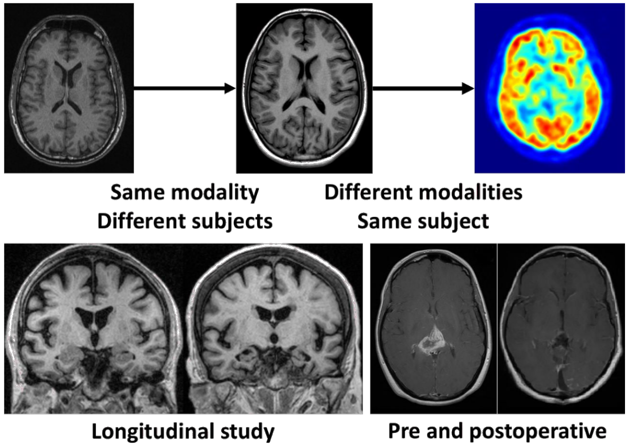
Modified from Source:
MathWorks. Image Registration. Interactively comparing feature-based, intensity-based, and nonrigid registration techniques using the Registration Estimator app.
https://de.mathworks.com/discovery/image-registration.html [06/25/2024].



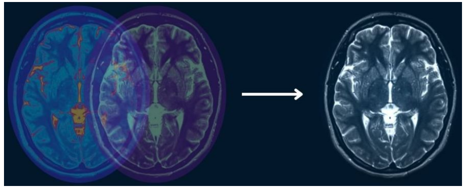
Modified from Source:
MathWorks. Image Registration.
Automatic registration using feature matching.
https://de.mathworks.com/discovery/image-registration.html
[06/25/2024].

Medical Image Registration

- Diagnostic settings
  - Combining information from multiple imaging modalities
- Studying disease progression (Longitudinal studies)
  - Monitoring changes in size, shape, position or image intensity over time
- Surgical planning & Image guided interventions or radiotherapy
  - Relating pre-operative images and surgical plans to the physical reality of the patient
- Post-operative evaluation
  - Relate preoperative and postoperative images after surgery
- Patient comparison or atlas construction
  - Relating one individual’s anatomy to another or to a standardized atlas



Pietro Gori. (2018). Introduction to medical image registration. Lecture Slides. Télécom Paris Tech



Nico Klingler. Image Registration and Its Applications. viso.ai. https://viso.ai/computer-vision/image-registration/ [/06/25/2024]

Dr. Ulas Bagci. (2017). Medical Image Computing. Lecture 15. Lecture Slides. UCF

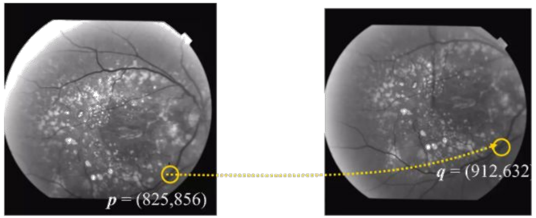
Background

Photo by Engin Akyurt on Pexels

Medical Image Registration - Taxonomy

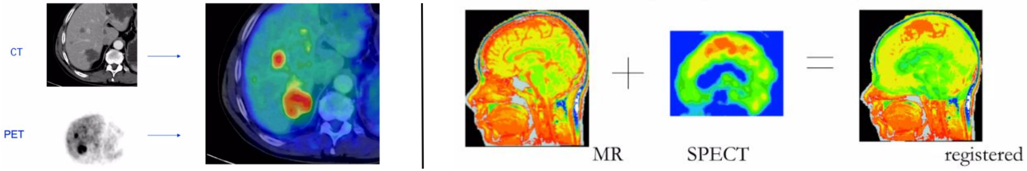
- Dimensionality
  - 2D - 2D, **3D - 3D**, 2D - 3D
- Transformation
  - Rigid, Affine, **Deformable**
- Modalities
  - Mono-modal, **multi-modal**
- Naming Convention
  - Moving image, fixed image
- Subject
  - **Intra-subject**, inter-subject, atlas
- Domain
  - **Local**, global
- Object
  - Whole-body, **organ**, ...

Registration is an alignment problem



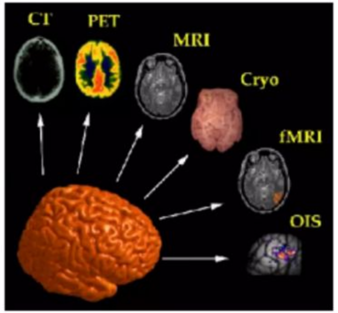
Medical Image Registration - Definition

- **Spatial Transform** that maps points from one image to corresponding points in another image
- Matching two images so that corresponding coordinate points in the two images **correspond to the same physical region** of the scene being imaged
- Also referred to as:
  - image fusion
  - superimposition
  - matching
  - merge



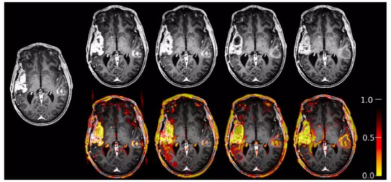
Dr. Ulas Bagci. (2017). Medical Image Computing. Lecture 15. Lecture Slides. UCF

Mono- vs Multi-Modality



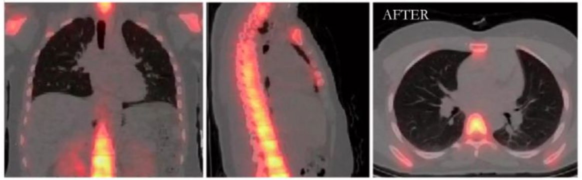
Mono-modality:

- A series of same modality images (CT/CT, MR/MR, Mammogram pairs, ...)
- Images may be acquired weeks or months apart (or taken from different viewpoints)
- Aligning images in order to detect subtle changes in intensity or shape



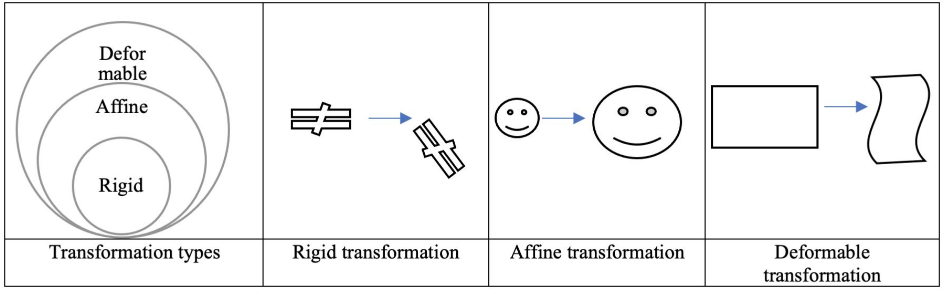
Multi-modality:

- Complementary anatomic and functional information from multiple modalities can be obtained for the precise diagnosis and treatment



Dr. Ulas Bagci. (2017). Medical Image Computing. Lecture 15. Lecture Slides. UCF

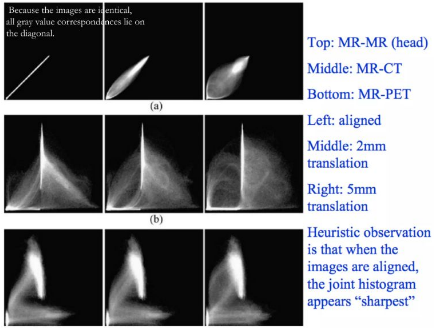
Transformations



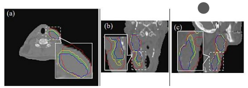
- Rigid Registration
  - Preserves distances between every pair of points
  - Rotations and translations
  - No compensation for motion or patient position
- Affine
  - Preserve parallelism and lines but no constraints on the preservation of distances
  - Rotations, translations, skew and scaling
- Deformable Registration
  - Free-form mapping (do not preserve the rigidity or affinity constraints)
  - Crucial when structures have changed position or shape between or during scans due to voluntary or physiological motion or imperfect scanning protocols

Nico Klingler. Image Registration and Its Applications. viso.ai.
https://viso.ai/computer-vision/image-registration/ [/06/25/2024]

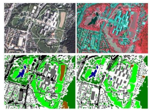
Similarity Criteria



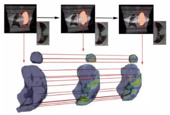
- Pixel Based
  - Alignment based on comparing pixel values
  - Simple and effective for high similarities
  - High processing complexity



- Contour Based
  - Based on shapes and outlines
  - Robust to changes in lighting & intensities
  - Depends on challenging edge detection



- Point Mapping
  - Identifying & matching key points/landmarks
  - Effective for reliable & distinctive landmarks
  - Can be difficult in featureless or homogeneous regions



- Feature Based
  - Utilizes extracted, distinctive features (e.g edges)
  - Robust to changes in scale, rotation, and lighting
  - Can struggle with images lacking distinct features

- Intensity Based
  - Alignment based on similarity of intensity values
  - Utilizes the entire image information
  - Less effective when there are intensity differences

- Information Theory (Mutual Information) Based
  - Based on maximizing MI of statistical dependency between intensities
  - Does not require images to have same intensities
  - May require careful parameter tuning

- Deep Learning Based
  - Learn transformation directly from image pairs
  - Learn complex transformations and handle distortions
  - Large amount of data required + computationally heavy

- Optical Flow
  - Estimates the motion of objects between consecutive frames or images by computing the apparent flow of intensity patterns

Nico Klingler. Image Registration and Its Applications. viso.ai. https://viso.ai/computer-vision/image-registration/ [/06/25/2024]

Deep Learning Based - Registration

Inherently difficult for multi-modal registration !

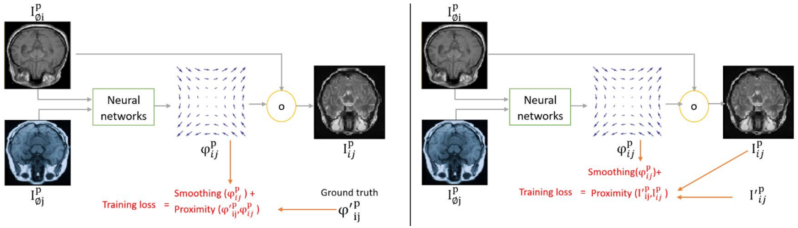
Supervised

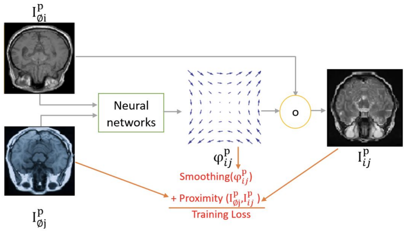
- Input images are fed into NN
- NN produces registration field (RF)
- RF is applied to the moving image
- Ground truth registration field or resulting wrapped image are used for supervision

Unsupervised

- Input images are fed into NN
- NN produces registration field (RF)
- RF is applied to the moving image
- Instead of relying on GT alignment, minimize cost function between fixed image and wrapped moving image

Ahmad Hammoudeh & Stéphane Dupont. (2023). Deep Learning in Medical Image Registration: Introduction and Survey. Qeios.





Contribution

Photo by Rodolfo Quirós on Pexels

Contribution

- Analyze and expose the limitations of self-similarity-based feature descriptors and mutual information-based methods in multi-modality registration
- Propose a novel **self-supervised** structural image representation learning paradigm dedicated to learning expressive **deep structural image representations** (DSIRs) without the need for anatomical delineations or perfectly aligned training image pair for supervision.
- Introduce the **Deep Neighbour Self-similarity** (DNS), which can capture long-range and complex structural information from medical images addressing the ambiguity in classical feature descriptors and similarity metrics
- Propose a novel **contrastive learning strategy** with non-linear intensity transformation, **maximizing the discriminability** of the feature representation across anatomical positions with homogeneous and heterogeneous intensity distribution
- Reduces the multimodal registration problem to a **monomodal one**, in which existing well-established monomodal registration algorithms can be applied

CVPR 2024

Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration

Tony C. W. Mok[1,2*] Zi Li[1,2*] Yunhao Bai[1] Jianpeng Zhang[1,2,4] Wei Liu[1,2]
Yan-Jie Zhou[1,2,4] Ke Yan[1,2] Dakai Jin[1] Yu Shi[3] Xiaoli Yin[3] Le Lu[1] Ling Zhang[1]

[1] DAMO Academy, Alibaba Group

[2] Hupan Lab, 310023, Hangzhou, China

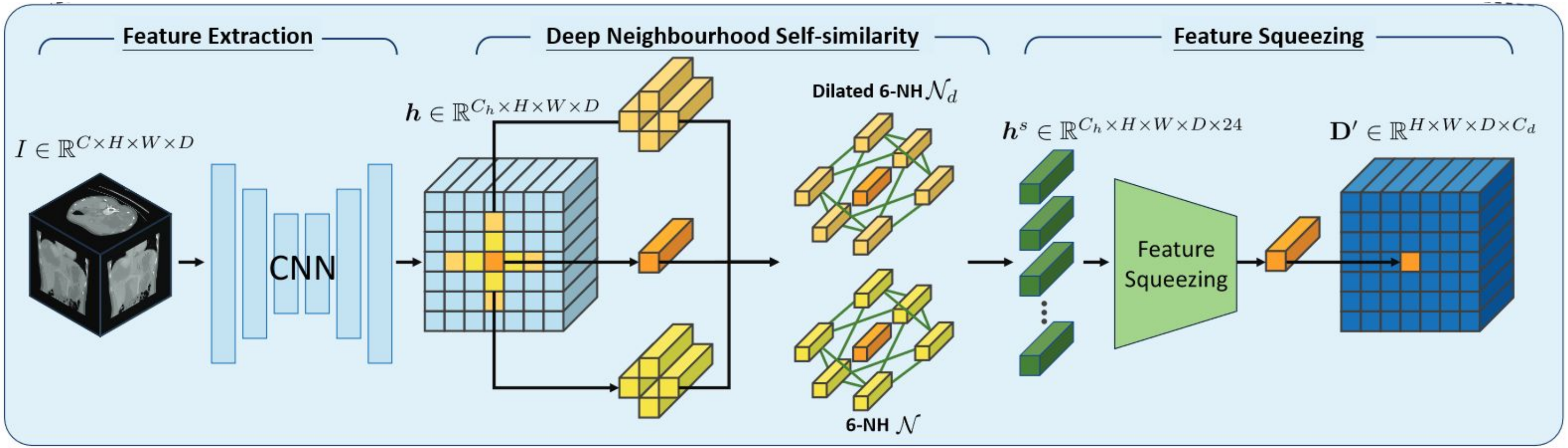
[3] Shengjing Hospital of China Medical University, China

[4] College of Computer Science and Technology, Zhejiang University, China

cwmokab@connect.ust.hk

Method

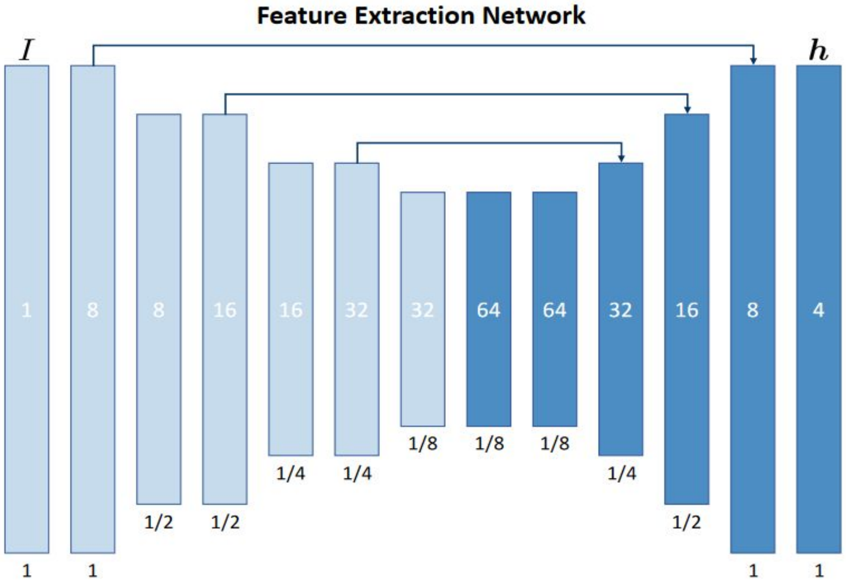
Modality-Agnostic Deep Structural Representation Network



1. Feature Extraction

- A 4-level encoder-decoder structure with skip connection

- Take an input image and output a feature map with same size

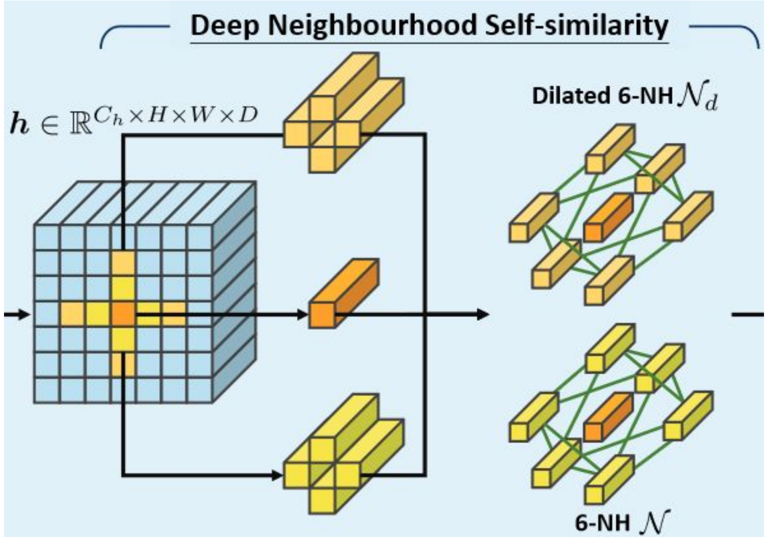


2. Deep Neighbourhood Self-similarity (DNS)

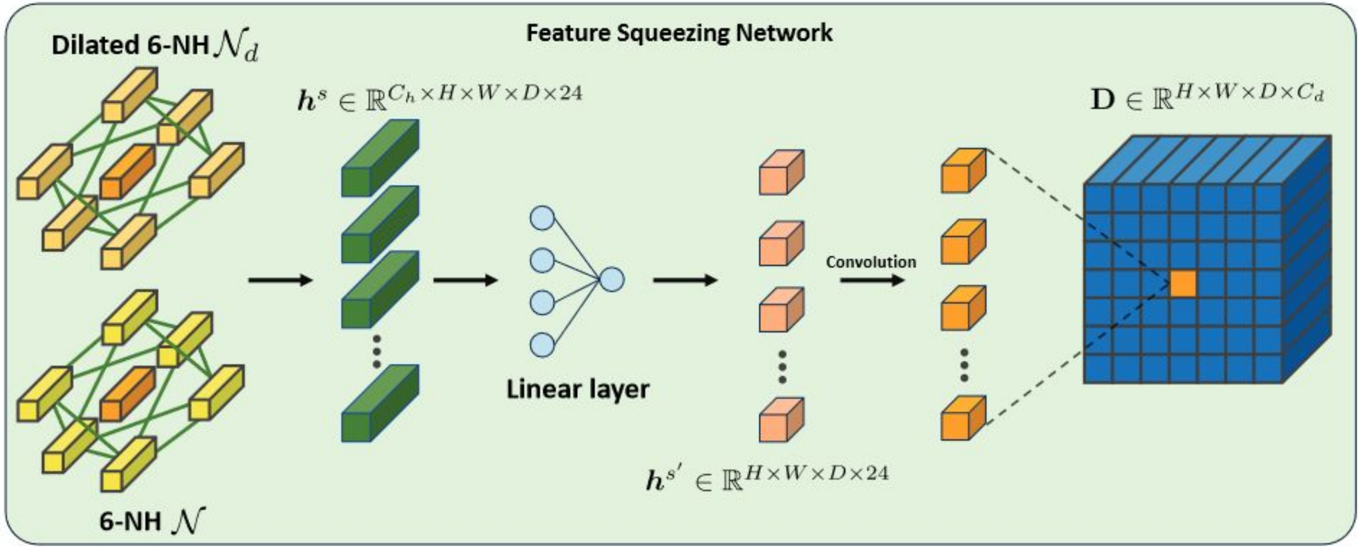
$$\mathbf{S}(\boldsymbol{h},x,y)=\exp\left(-\sum_{y'\in\mathcal{N}(x)}\frac{(\boldsymbol{h}(y)-\boldsymbol{h}(y'))^2}{\sigma^2}\right), y'\neq y,$$

Compute 12 pair-wise distances between different neighbour location

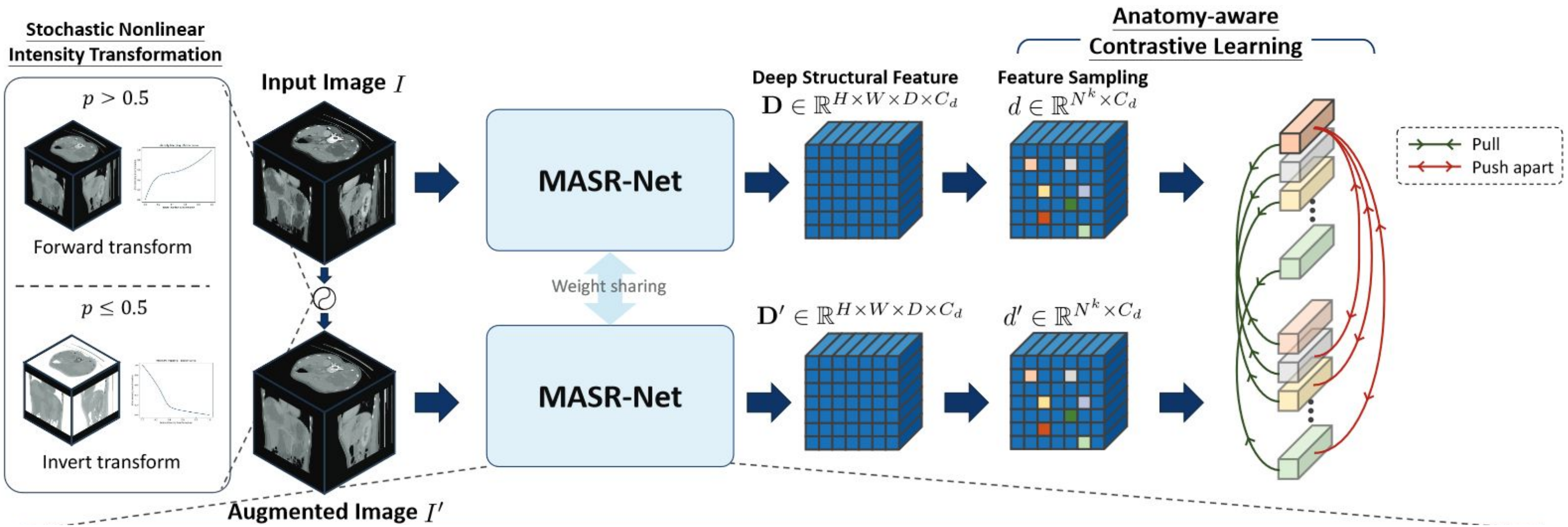
Output: 5D DNS feature map $\boldsymbol{h}^s \in \mathbb{R}^{C_h\times H\times W\times D\times 24}$



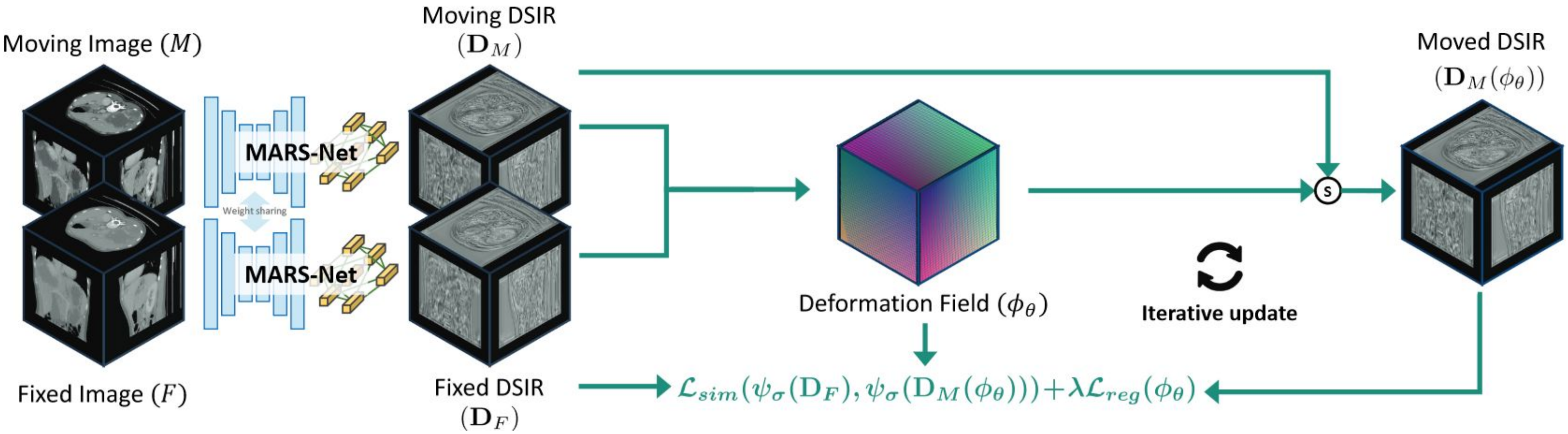
3. Feature Squeezing



Anatomy-aware Contrastive Learning



Multimodal image registration using DNS



Experiments

Liver Multiphase CT

Method	Metric	Pre-contrast ← Venous & Arterial					Arterial ← Venous & Pre-contrast				
		Tumour		Organ		$\%\|J_\phi\| < 0 \downarrow$	Tumour		Organ		$\%\|J_\phi\| < 0 \downarrow$
		DSC ↑	HD95 ↓	DSC ↑	HD95 ↓		DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	
Initial	–	75.51 ± 20.64	4.12 ± 4.00	88.03 ± 8.77	3.80 ± 3.31	–	77.84 ± 19.86	3.61 ± 3.19	90.15 ± 7.41	3.15 ± 2.68	–
ANTs	MI	76.52 ± 20.16	3.97 ± 3.15	87.40 ± 9.75	3.88 ± 2.91	0.00 ± 0.00	80.18 ± 18.47	3.42 ± 2.73	90.07 ± 8.04	3.22 ± 2.80	0.00 ± 0.00
NiftyReg	MI	77.10 ± 17.00	3.54 ± 2.54	90.87 ± 6.57	2.98 ± 2.10	0.35 ± 1.20	79.22 ± 16.54	3.18 ± 2.07	92.39 ± 5.61	2.53 ± 1.79	0.29 ± 1.18
DEEDs	MIND	79.65 ± 15.30	3.14 ± 1.98	93.90 ± 3.47	2.09 ± 1.50	0.23 ± 0.72	80.57 ± 15.23	3.42 ± 2.06	94.80 ± 2.75	**1.84** ± 1.04	0.12 ± 0.13
VM	NMI	75.78 ± 17.35	3.62 ± 2.69	91.70 ± 4.43	2.81 ± 1.89	0.00 ± 0.00	76.62 ± 18.19	3.23 ± 2.30	92.69 ± 3.81	2.45 ± 1.63	0.00 ± 0.00
VM	MIND	75.16 ± 17.07	3.94 ± 5.51	91.70 ± 4.92	2.77 ± 2.05	0.00 ± 0.00	75.18 ± 17.71	3.63 ± 5.11	92.15 ± 4.39	2.55 ± 1.65	0.00 ± 0.00
LapIRN	NMI	78.80 ± 15.37	3.33 ± 2.80	93.53 ± 3.75	2.35 ± 1.52	0.01 ± 0.02	80.17 ± 15.38	2.87 ± 2.00	94.48 ± 3.72	1.97 ± 1.53	0.02 ± 0.16
LapIRN	MIND	77.50 ± 16.71	3.96 ± 3.28	93.32 ± 4.41	2.46 ± 1.81	0.00 ± 0.01	79.49 ± 15.00	3.19 ± 2.44	94.29 ± 4.19	2.07 ± 1.77	0.01 ± 0.04
LapIRN (ours)	DNS	79.72 ± 14.44	3.06 ± 2.23	94.07 ± 3.36	**2.08** ± 1.45	0.00 ± 0.01	80.66 ± 14.56	**2.71** ± 1.85	94.73 ± 3.09	1.84 ± 1.25	0.02 ± 0.18
IO	MIND	76.27 ± 16.44	3.66 ± 2.74	92.54 ± 3.41	2.63 ± 1.32	0.08 ± 0.37	76.91 ± 16.14	3.54 ± 2.23	92.74 ± 3.45	2.70 ± 1.31	0.12 ± 0.51
IO (ours)	DNS	**80.43** ± 13.72	**2.94** ± 2.23	**94.26** ± 3.32	2.10 ± 1.50	0.03 ± 0.20	**81.07** ± 13.83	2.74 ± 1.80	**94.89** ± 2.92	1.85 ± 1.22	0.06 ± 0.35

Liver Multiphase CT

Method	Metric	Venous ← Arterial & Pre-contrast					Average Score across Three Tasks				
		Tumour		Organ		$\%\lvert J_\phi\rvert <0 \downarrow$					
		DSC ↑	HD95 ↓	DSC ↑	HD95 ↓		DSC ↑	HD95 ↓	$\%\lvert J_\phi\rvert <0 \downarrow$	T_{Test}	# Param
Initial	–	78.10 ± 19.85	3.59 ± 3.07	88.96 ± 7.49	3.50 ± 2.41	–	83.10 ± 12.60	3.63 ± 2.97	–	–	–
ANTs	MI	81.14 ± 17.14	3.37 ± 2.69	89.20 ± 8.28	3.52 ± 2.46	0.00 ± 0.00	84.09 ± 12.84	3.56 ± 2.60	0.00 ± 0.00	250.05 ± 367.42*	–
NiftyReg	MI	78.53 ± 17.05	3.32 ± 2.28	91.19 ± 6.80	2.90 ± 2.06	0.32 ± 1.19	84.88 ± 9.92	3.07 ± 1.98	0.32 ± 1.19	79.17 ± 30.77*	–
DEEDs	MIND	81.28 ± 14.65	3.32 ± 1.87	**94.58** ± 2.44	**2.11** ± 1.14	0.11 ± 0.14	87.46 ± 8.97	2.65 ± 1.72	0.15 ± 0.43	47.92 ± 12.77*	–
VM	NMI	77.20 ± 17.33	3.28 ± 2.09	92.65 ± 3.31	2.60 ± 1.43	0.00 ± 0.00	84.44 ± 9.47	3.00 ± 1.83	0.00 ± 0.00	0.14 ± 0.01	1.14M
VM	MIND	75.94 ± 17.18	3.34 ± 1.54	92.25 ± 4.07	2.63 ± 1.54	0.00 ± 0.00	83.73 ± 9.33	3.14 ± 2.57	0.00 ± 0.00	0.16 ± 0.01	1.14M
LapIRN	NMI	81.37 ± 14.05	2.86 ± 2.12	94.22 ± 3.46	2.24 ± 1.57	0.00 ± 0.00	87.10 ± 8.89	2.61 ± 1.91	0.00 ± 0.01	0.19 ± 0.01	1.59M
LapIRN	MIND	81.03 ± 14.56	3.09 ± 2.61	93.99 ± 3.98	2.33 ± 1.82	0.00 ± 0.00	86.61 ± 9.81	2.85 ± 2.29	0.00 ± 0.01	0.18 ± 0.01	1.59M
LapIRN (ours)	DNS	**81.62** ± 13.42	2.81 ± 1.99	94.37 ± 3.03	2.08 ± 1.41	0.00 ± 0.00	87.53 ± 8.65	2.43 ± 1.70	0.01 ± 0.01	0.18 ± 0.01	1.59M
IO	MIND	77.44 ± 16.26	3.48 ± 2.37	93.02 ± 3.30	2.56 ± 1.42	0.01 ± 0.02	84.82 ± 9.83	3.10 ± 1.90	0.07 ± 0.30	4.75 ± 0.44	-
IO (ours)	DNS	81.51 ± 13.67	**2.79** ± 1.94	94.53 ± 3.01	2.11 ± 1.48	0.00 ± 0.01	**87.78** ± 8.41	**2.42** ± 1.70	0.03 ± 0.19	5.05 ± 0.47	-

Abdomen MR-CT & Brain MR T1w-T2w

Method	Metric	Abdomen MR ← CT		Brain MR T1w ↔ T2w		
		DSC ↑	T_{Test}	$\text{DSC}_{\text{T1}\leftarrow\text{T2}}$ ↑	$\text{DSC}_{\text{T2}\leftarrow\text{T1}}$ ↑	T_{Test}
Initial	–	37.32 ± 17.23	–	53.90 ± 0.70	53.90 ± 0.70	–
DEEDs	MIND	83.53 ± 8.58	165.21 ± 25.14*	61.47 ± 0.96	60.70 ± 0.83	14.45 ± 0.54
ANTs	MI	39.43 ± 17.67	38.54 ± 3.51*	61.00 ± 1.00	52.80 ± 1.10	18.46 ± 1.33
NiftyReg	MI	48.60 ± 31.60	37.88 ± 10.95*	63.90 ± 1.10	61.90 ± 0.70	34.75 ± 3.04
IO (ours)	DNS	85.61 ± 5.95	4.52 ± 0.50	62.66 ± 0.73	61.91 ± 0.66	5.84 ± 0.54

Ablation Studies

Methods	DSC ↑	HD95 ↓
Initial	81.77 ± 16.96	3.96 ± 3.66
Backbone network feature (Random initialization)	75.38 ± 21.03	5.90 ± 5.46
+Deep Neighbourhood Self-similarity	86.50 ± 13.43 (+11.12)	2.79 ± 2.30 (-3.11)
+Nonlinear Intensity & Contrastive Learning	87.13 ± 12.06 (+0.63)	2.57 ± 1.95 (-0.22)
+Gaussian Smoothing	87.35 ± 12.14 (+0.22)	2.52 ± 1.95 (-0.05)

Conclusion

- Proposes a deep structural image representation learning method for multi-modal medical image registration

- Leverages deep neighbourhood self-similarity to learn highly discriminative, contrast invariance structural representations

- Anatomy-aware contrastive learning to further enhance the expressiveness and discriminability of the structural representation, reducing the ambiguity in matching anatomical correspondence

Thank you for your attention!

Tray22, Lord Of GIF on tenor
https://tenor.com/de/view/lord-of-the-rings-applause-gif-24208468