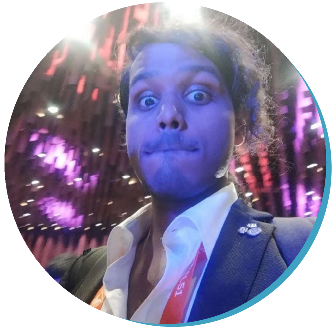
MICCAI 2025
SSL3D / FOMO60k Challenge

Puru Vaish (Mei)
The Latent Campus

UNIVERSITY OF TWENTE.

The Brain Stem of TheLatentCampus



Puru Vaish (Mei)[1]

Felix Meister[2]

Tobias Heimann[2]

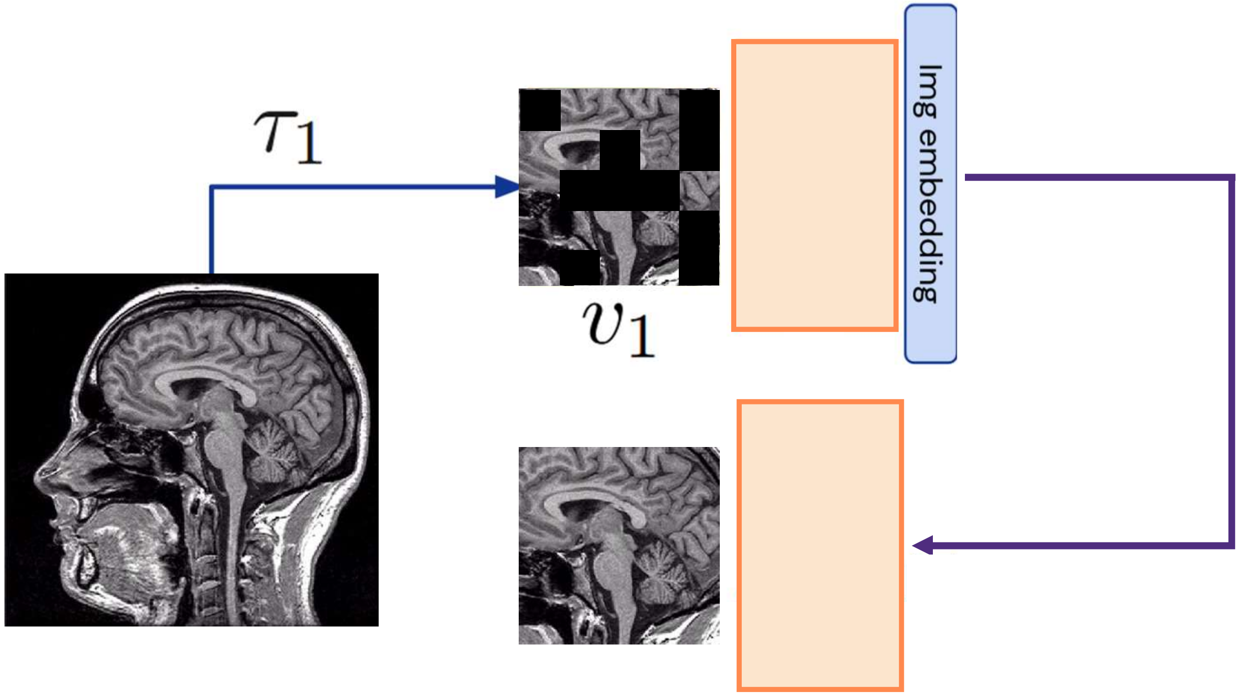
Christoph Brune[1]

Jelmer Wolterink[1]

1. Department of Applied Mathematics, Technical Medical Centre, University of Twente
2. Digital Technology and Innovation, Siemens Healthineers, Erlangen, Germany

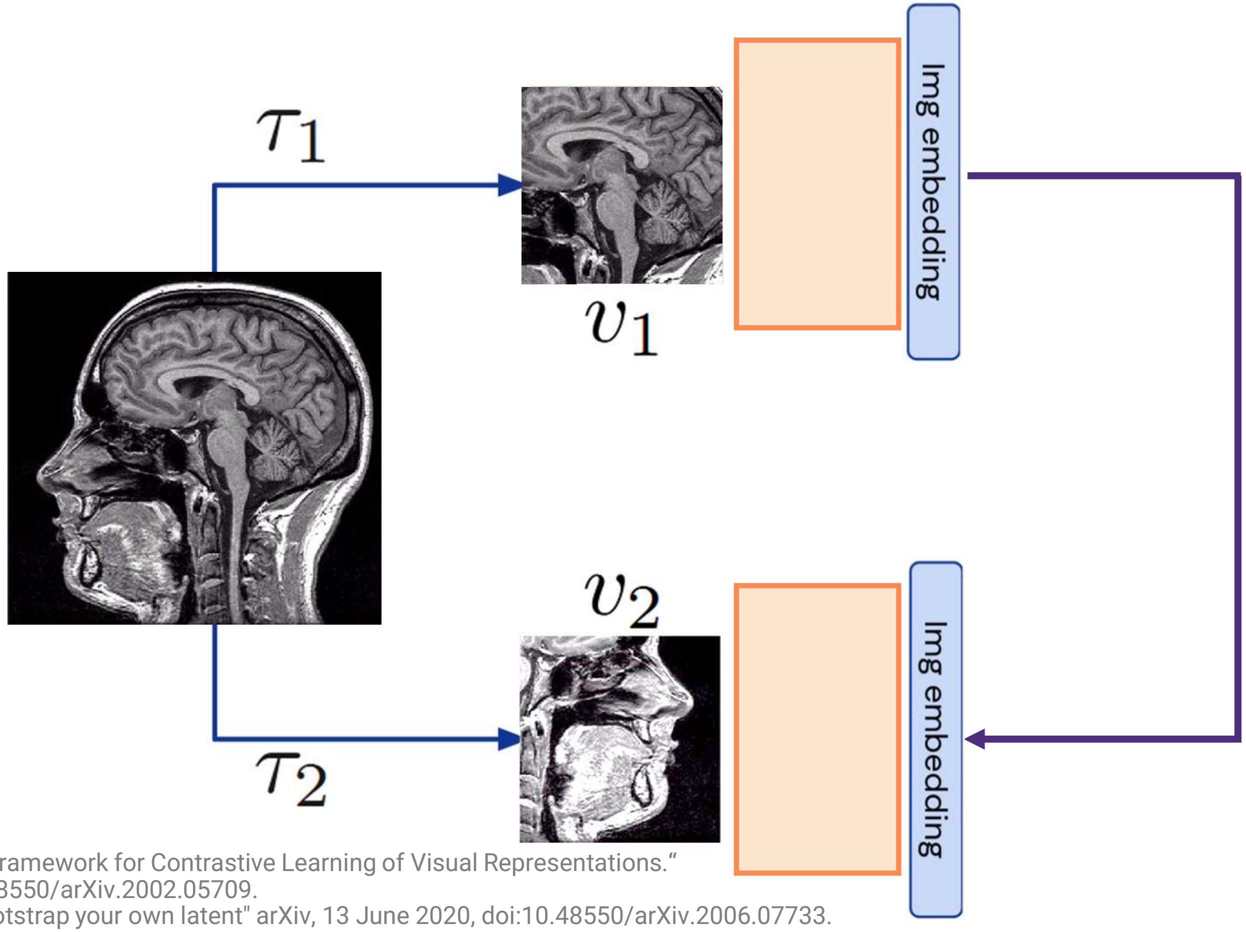
ICAI Innovation Center for Artificial Intelligence SIEMENS Healthineers NWO UNIVERSITY OF TWENTE.

State of Self-Supervised Learning



He, Kaiming, et al. "Masked Autoencoders Are Scalable Vision Learners." arXiv, 11 Nov. 2021, doi:10.48550/arXiv.2111.06377.
Huang, Zhicheng, et al. "Contrastive Masked Autoencoders are Stronger Vision Learners." arXiv, 27 July 2022, doi:10.1109/TPAMI.2023.3336525.

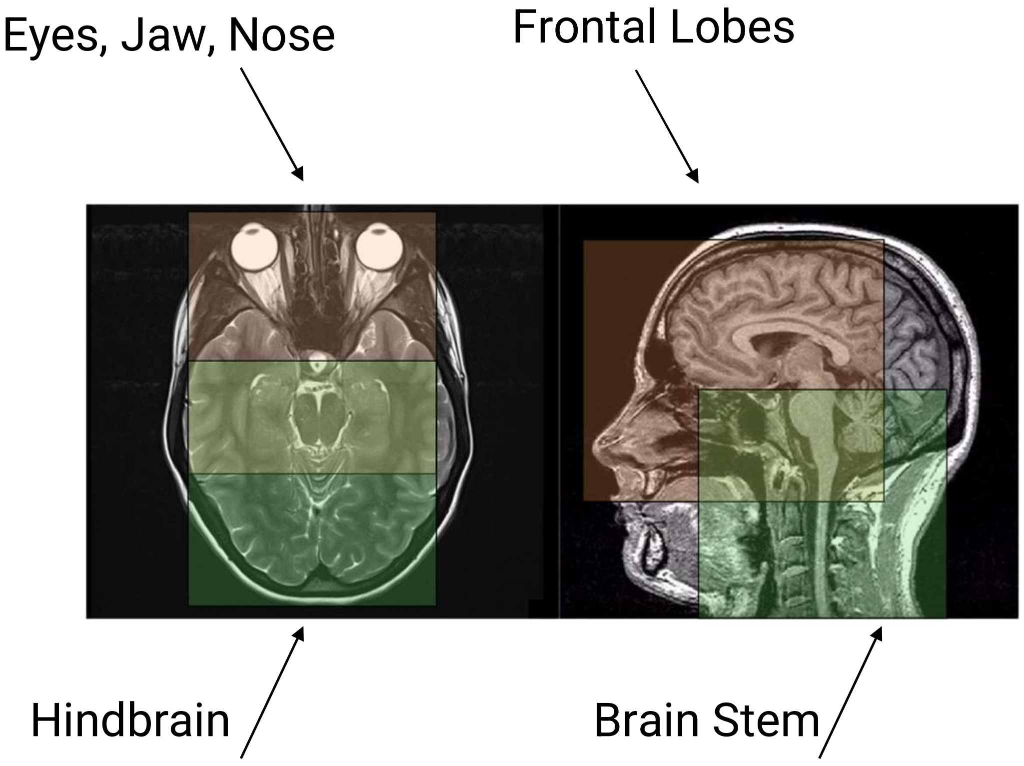
State of Self-Supervised Learning



Chen, Ting, et al. "A Simple Framework for Contrastive Learning of Visual Representations." arXiv, 13 Feb. 2020, doi:10.48550/arXiv.2002.05709.
Grill, Jean-Bastien, et al. "Bootstrap your own latent" arXiv, 13 June 2020, doi:10.48550/arXiv.2006.07733.

There is an Issue

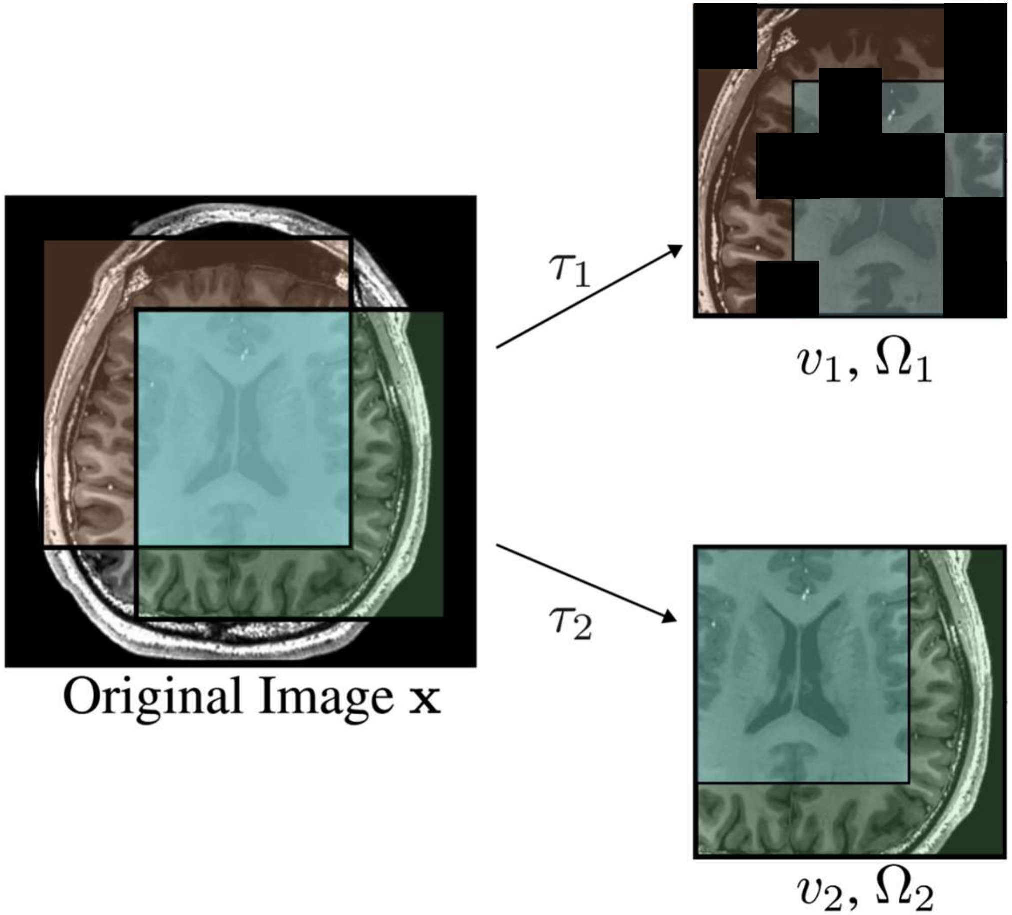
- Loosely correlated regions are enforced to be “similar”
- Poor representation learning and generalisation
- Need for alignment of features.



Consistent View

Given the original image get two volumes with an overlapping region.

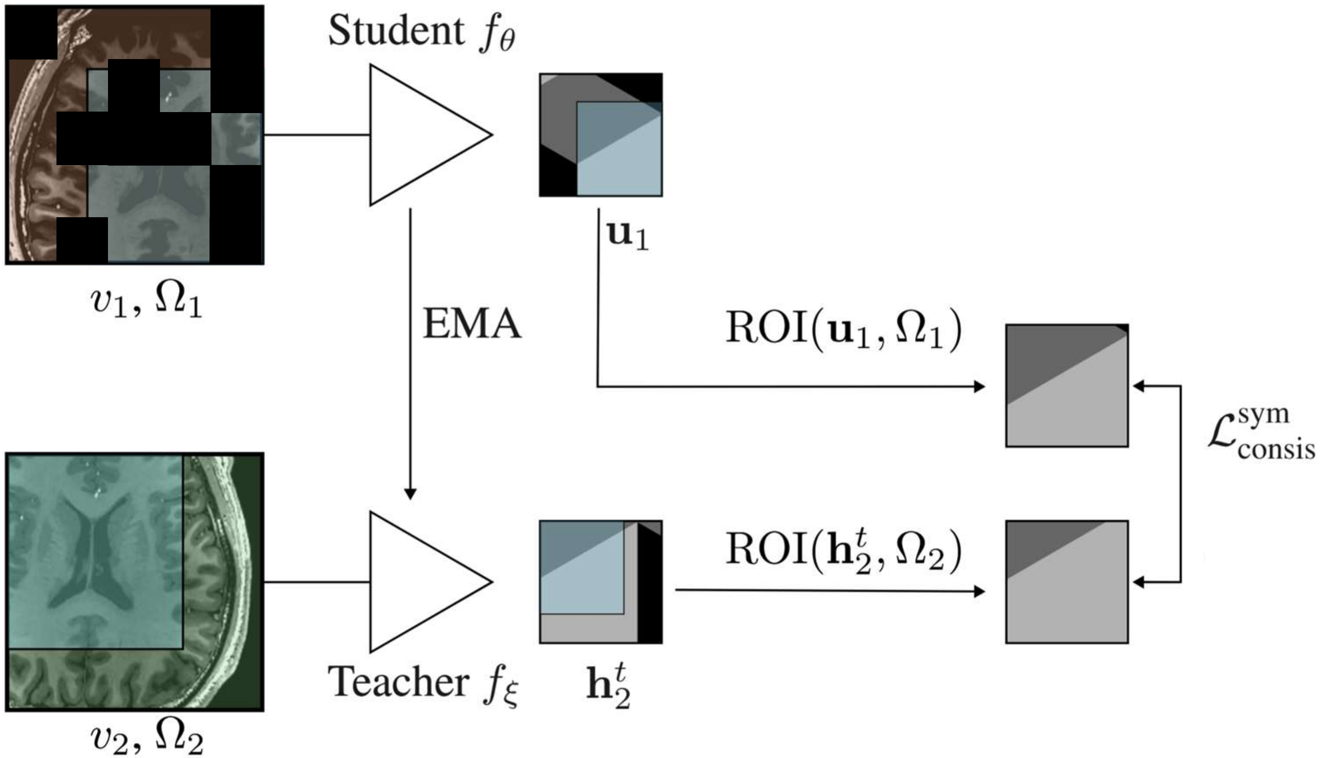
Keep track of where the overlap is.



ICAI Innovation Center for Artificial Intelligence SIEMENS Healthineers NWO UNIVERSITY OF TWENTE.

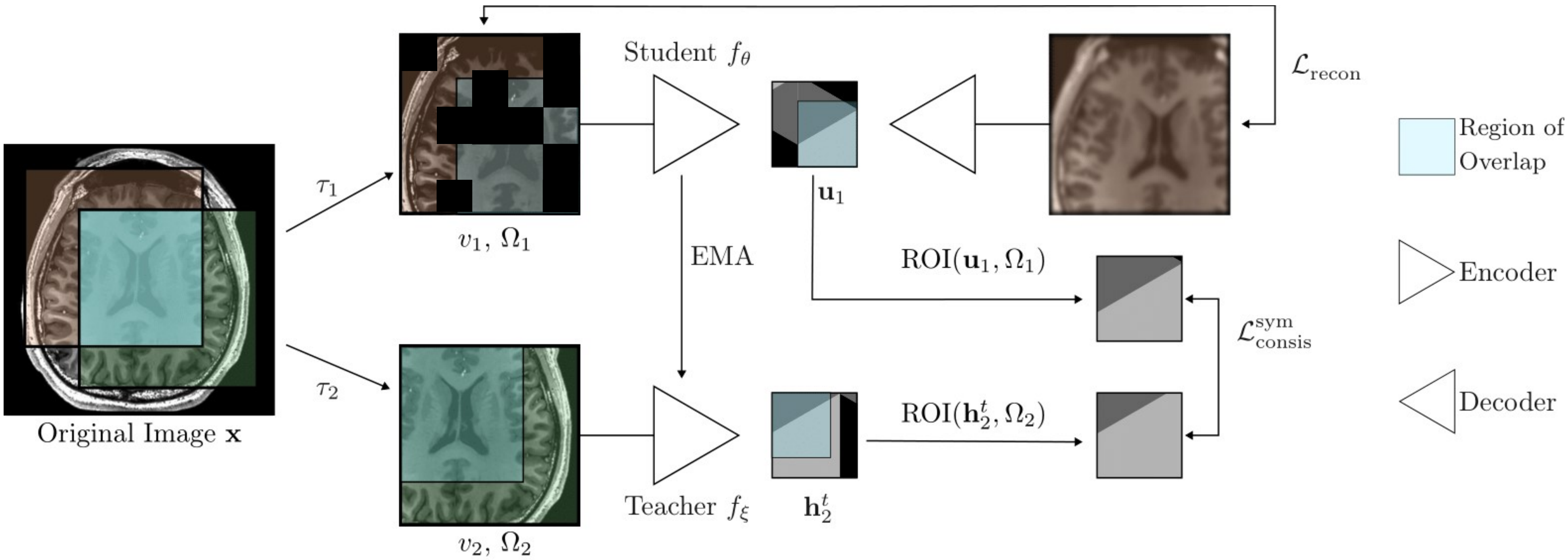
Alignment

Extract the features and using the saved bounding boxes to align the features.



ICAI Innovation Center for Artificial Intelligence SIEMENS Healthineers NWO UNIVERSITY OF TWENTE.

Consistent View + Alignment = CVA

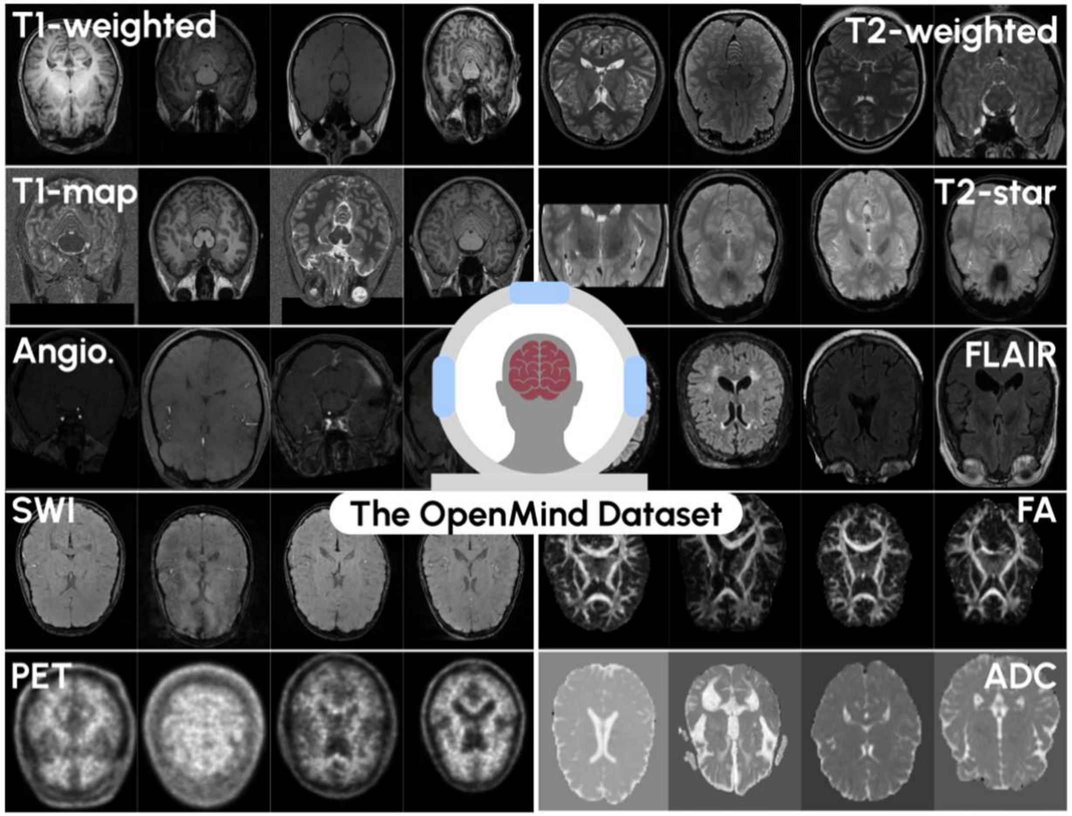


He, Kaiming, et al. "Masked Autoencoders Are Scalable Vision Learners." arXiv, 11 Nov. 2021, doi:10.48550/arXiv.2111.06377.
Chen, Xinlei and Kaiming He. "Exploring Simple Siamese Representation Learning." arXiv, 20 Nov. 2020, doi:10.48550/arXiv.2011.10566.

OpenMind Dataset

Dataset Overview

- Large publicly available head & neck MRI dataset (via OpenNeuro)
- Aggregates **800+ datasets**, released under **CC0/PDDL** licenses
- Standardized and cleaned: **114,570 3D volumes** from **34,191 patients**
- Covers 23 modalities MRI sequences (T1w, FLAIR, T2w, MD, FA, and more)
- Publicly hosted on **Hugging Face**



ICAI Innovation Center for Artificial Intelligence
SIEMENS Healthineers
NWO
UNIVERSITY OF TWENTE.

Two Stage Training

Stage	Architecture	Details	Duration Single L40
Pretraining - MAE - Huber Loss	ResEnc-L	1000 epochs	~7 days
	Primus-M		~9 days
Post-pretraining - Respective losses	ResEnc-L	250 epochs	~1.5 days
	Primus-M	150 epochs	~2 days

- Pros: 37 GPU days vs 112 GPU days over 14 ablations.
- Cons: Bias towards MAE representations.

Wald, Tassilo, et al. "An OpenMind for 3D medical vision self-supervised learning." arXiv, 22 Dec. 2024, doi:10.48550/arXiv.2412.17041.

Technical Details

- Cosine Similarity Loss

$$\mathcal{L}_{\cos}(\mathbf{u}_1^{\Omega_1}, \mathbf{h}_2^{\Omega_2,t}) = 2 - 2\frac{\mathbf{u}_1^{\Omega_1} \cdot \mathbf{h}_2^{\Omega_2,t}}{\|\mathbf{u}_1^{\Omega_1}\|_2 \, \|\mathbf{h}_2^{\Omega_2,t}\|_2}.$$

- Symmetrisation

$$\mathcal{L}_{\cos}^{\mathrm{sym}} = \frac{1}{2}\mathcal{L}_{\cos}(\mathbf{u}_1^{\Omega_1}, \mathbf{h}_2^{\Omega_2,t}) + \frac{1}{2}\mathcal{L}_{\cos}(\mathbf{u}_2^{\Omega_2}, \mathbf{h}_1^{\Omega_1,t}).$$

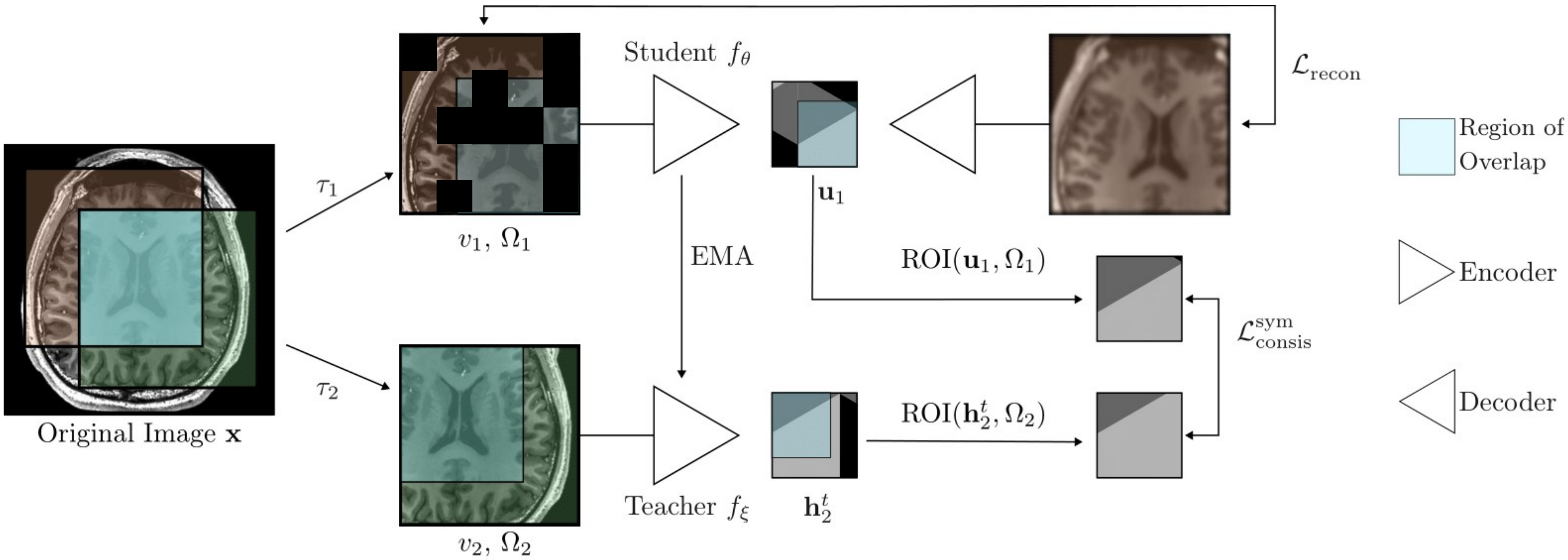
- Contrastive Signal

$$\mathcal{L}_{\text{NT-Xent}}(\mathbf{u}_1, \mathbf{h}_2^t) = -\log \frac{\exp\left(\mathrm{sim}(\mathbf{u}_1, \mathbf{h}_2^t)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp\left(\mathrm{sim}(\mathbf{u}_1, \mathbf{h}_k^t)/\tau\right)}.$$

- onemmiso preprocessed, Batch Size 4, Patch Size 160^3, no exclusion criteria
- Finetuning – at the FOMO60k talk.

Caron, Mathilde, et al. "Unsupervised Learning of Visual Features by Contrasting Cluster Assignments." arXiv, 17 June 2020, doi:10.48550/arXiv.2006.09882.
Huang, Zhicheng, et al. "Contrastive Masked Autoencoders are Stronger Vision Learners." arXiv, 27 July 2022, doi:10.1109/TPAMI.2023.3336525.
He, Kaiming, et al. "Masked Autoencoders Are Scalable Vision Learners." arXiv, 11 Nov. 2021, doi:10.48550/arXiv.2111.06377.
Chen, Xinlei and Kaiming He. "Exploring Simple Siamese Representation Learning." arXiv, 20 Nov. 2020, doi:10.48550/arXiv.2011.10566.

Consistent View + Alignment = CVA



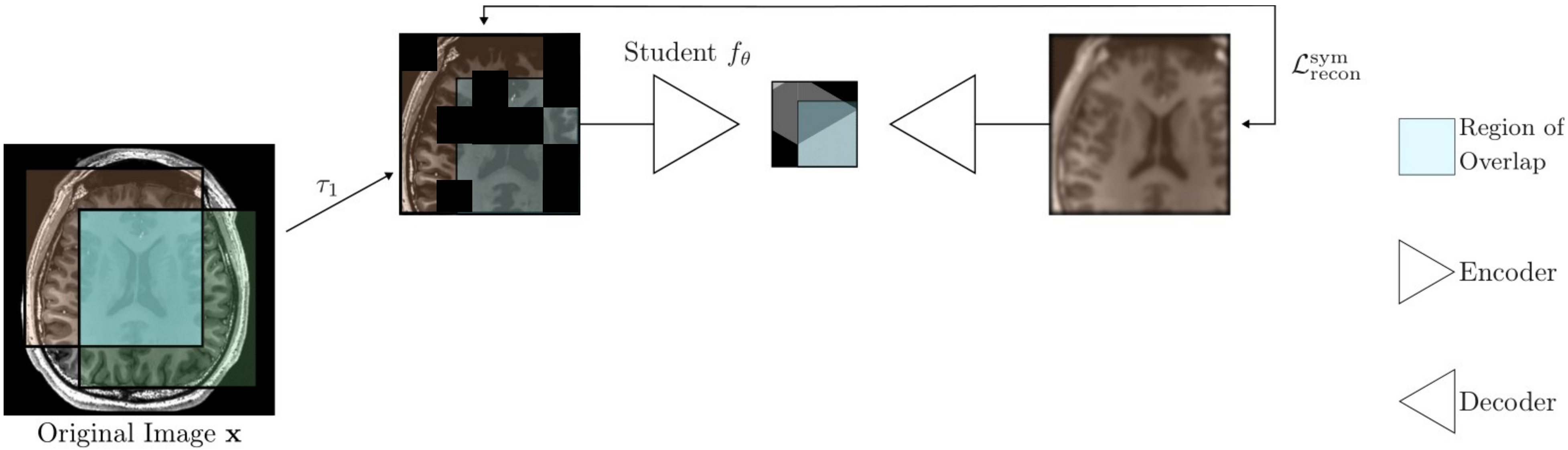
He, Kaiming, et al. "Masked Autoencoders Are Scalable Vision Learners." arXiv, 11 Nov. 2021, doi:10.48550/arXiv.2111.06377.
Chen, Xinlei and Kaiming He. "Exploring Simple Siamese Representation Learning." arXiv, 20 Nov. 2020, doi:10.48550/arXiv.2011.10566.

Pre-Challenge Ablation Ladder

- 4 Segmentation Datasets
  - Ischemic Stroke Lesions
  - Yale Brain Metastasis
  - BraTS: Post-Glioblastoma
  - Medical Segmentation Decathlon: Brain Tumour Segmentation
- 1 Classification Dataset
  - ABIDE 2
- Rankings should be interpreted with care

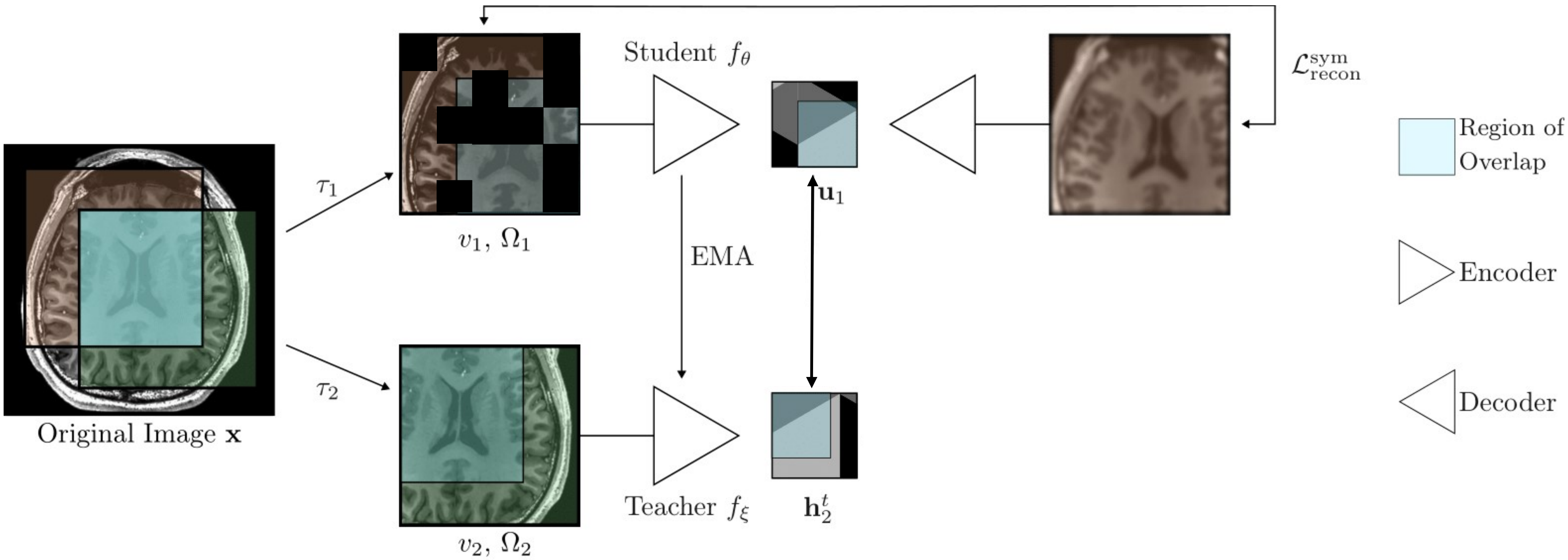
Maier-Hein, Lena, et al. "Why rankings of biomedical image analysis competitions should be interpreted with care." Nat. Commun. doi:10.1038/s41467-018-07619-7.

Ablation Ladder - MAE



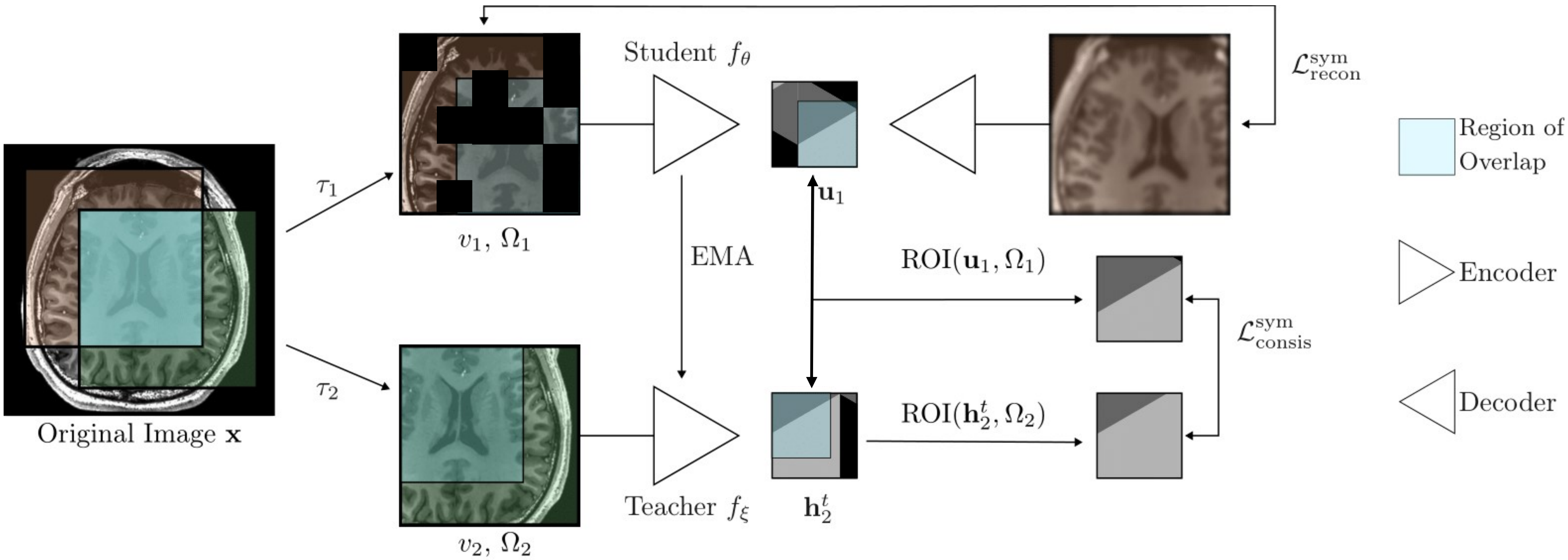
Track	Recon.	Consis.	Cont.	Avg Rank	Seg Rank	Cls Rank	Segmentation								Classification		
							ISL		YBM		GLI		MSD		ABD II		
							DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	Bal Acc.	AUROC	AP
ResEnc-L	MAE	✗	✗	5.22	5.00	5.67	78.87	76.66	61.21	68.68	69.83	75.02	72.22	76.49	57.33	60.18	58.89
Primus-M	MAE	✗	✗	6.31	6.13	6.67	77.18	74.98	52.70	59.01	65.82	72.58	71.41	75.44	54.80	58.75	58.26

Ablation Ladder – Contrastive MAE



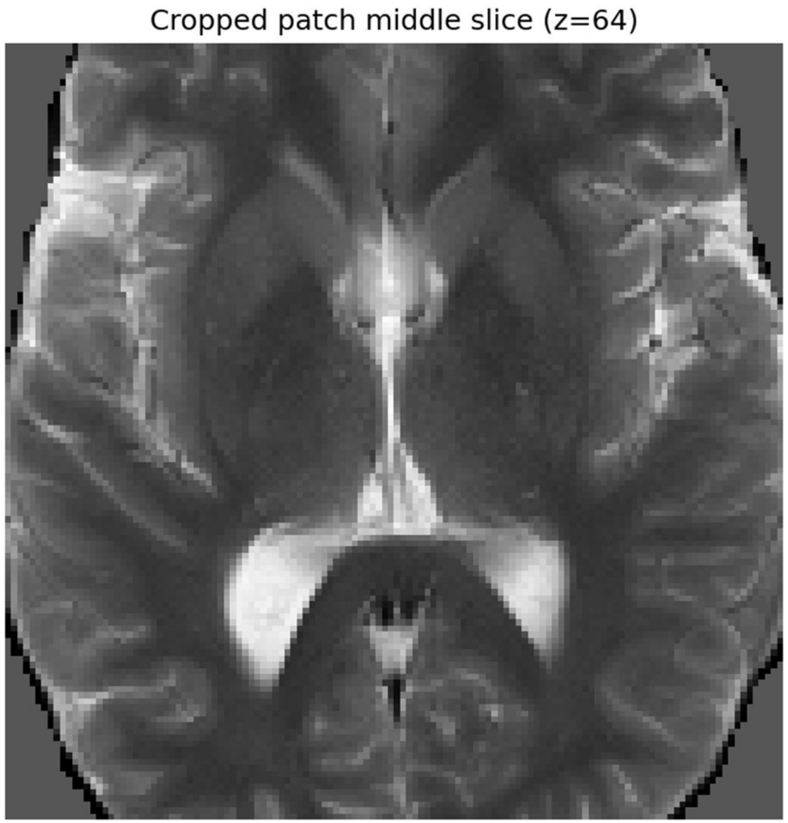
Track	Recon.	Consis.	Cont.	Avg Rank	Seg Rank	Cls Rank	Segmentation								Classification		
							ISL		YBM		GLI		MSD		ABD II		
							DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	Bal Acc.	AUROC	AP
ResEnc-L	MAE	✗	✓	2.53	3.13	**1.33**	80.05	78.18	62.31	70.37	69.82	74.84	72.80	76.70	61.09	64.93	62.67
Primus-M	MAE	✗	✓	**2.50**	3.25	**1.00**	77.18	75.36	54.87	61.74	65.82	72.65	71.78	75.85	58.55	62.42	61.60

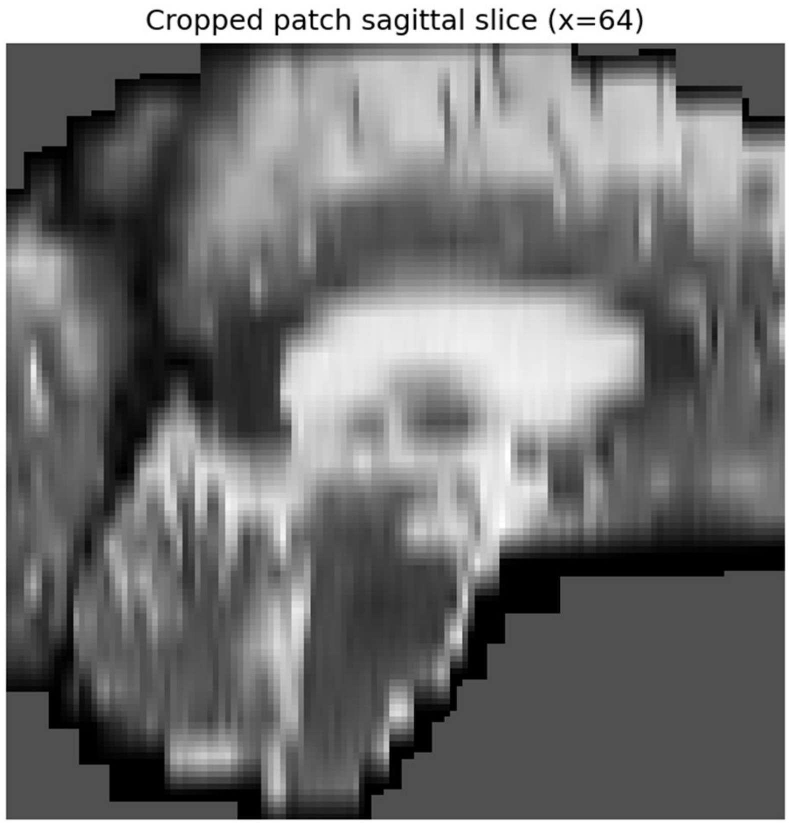
Ablation Ladder – CVA w/ Contrastive Signal



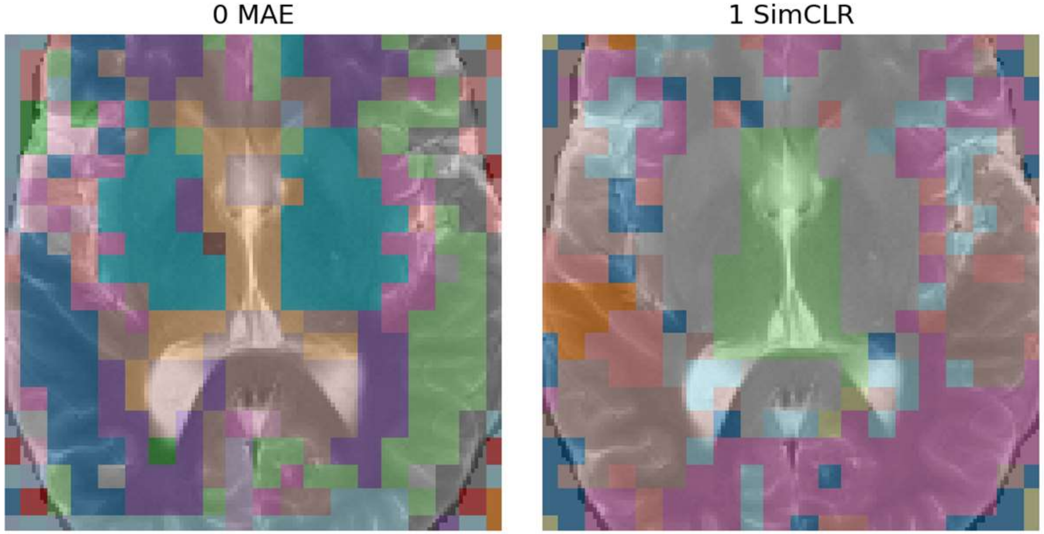
Track	Recon.	Consis.	Cont.	Avg Rank	Seg Rank	Cls Rank	Segmentation								Classification		
							ISL		YBM		GLI		MSD		ABD II		
							DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	Bal Acc.	AUROC	AP
ResEnc-L	MAE	CVA	✓	**2.47**	2.88	1.67	78.97	77.09	62.35	70.94	69.75	74.85	72.84	77.02	62.02	64.46	62.62
Primus-M	MAE	CVA	✓	**2.49**	**2.13**	3.21	77.33	75.14	54.78	61.60	66.48	73.27	71.86	76.10	56.13	59.10	59.31

Quality of Representation

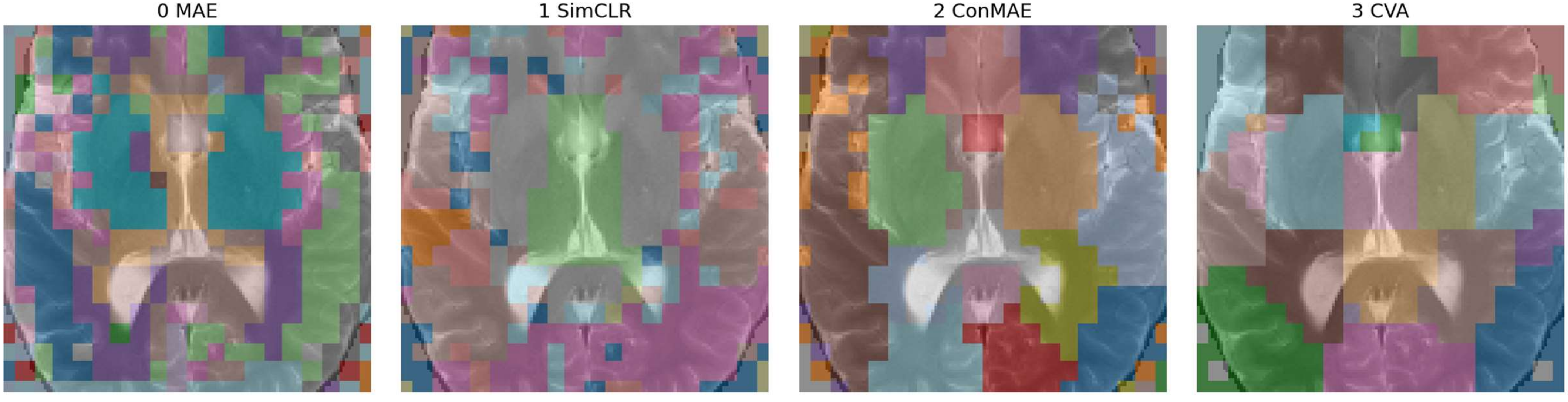




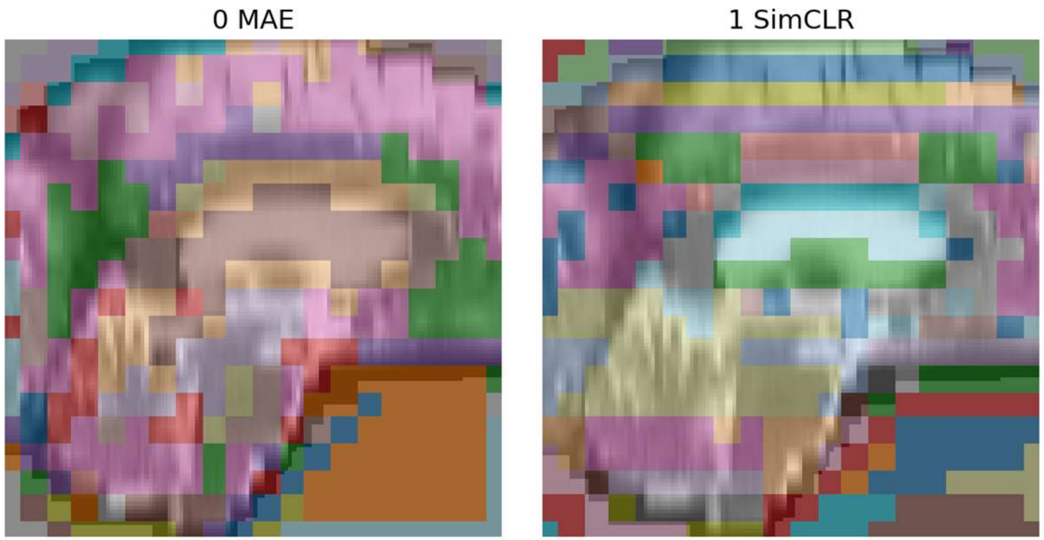
k-Means Clustering



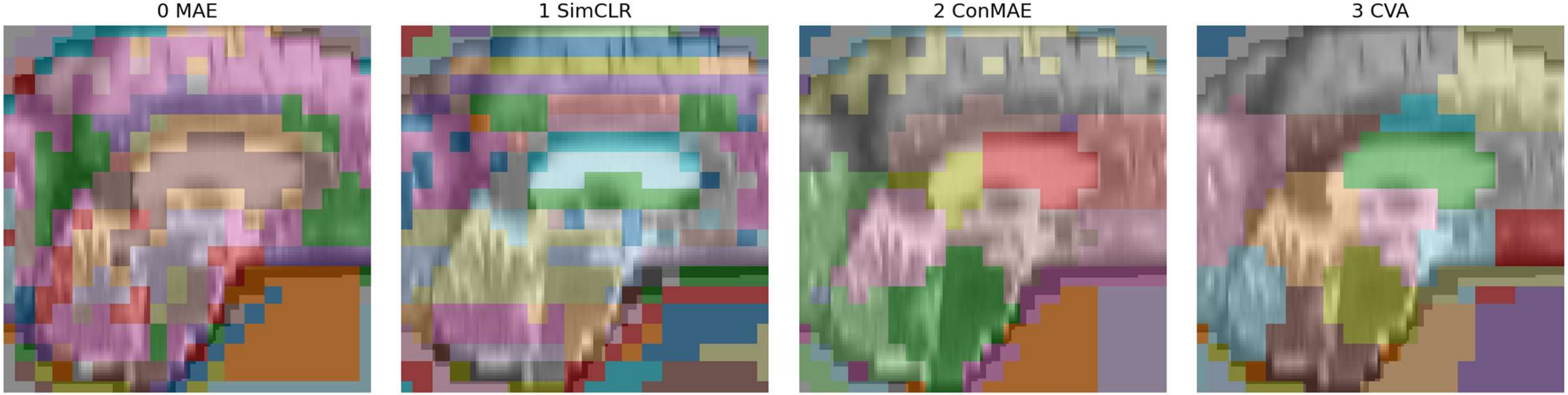
k-Means Clustering



k-Means Clustering



k-Means Clustering



FOMO60K - Finetuning

- Classification
  - Baseline Codebase provided by the Challenge
    - 128 patch size
    - 8 batch accumulation

- Brain Age Regression
  - Baseline Codebase provided by the Challenge
    - 128 patch size
    - 16 batch accumulation

- Segmentation
  - 50 epochs, 100 iterations of nnUNet training from pretrained checkpoint.
    - 5 fold cross validation.
    - Best model sent for final testing.

Flexible Framework makes the World Work

Even more ablations!

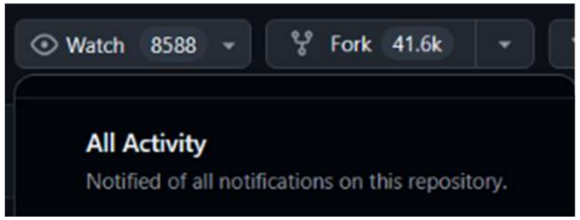
- Different types of consistency:
  - Effect of contrastive signal
  - Contrastive
  - Gram Matrix
- Ranking Interpretations

Code Publicly Available

- Extension of nnssl and compatible with nnunet_adaption

Pretrained Weights

- Soon on Zenodo (watch on GitHub for updates)



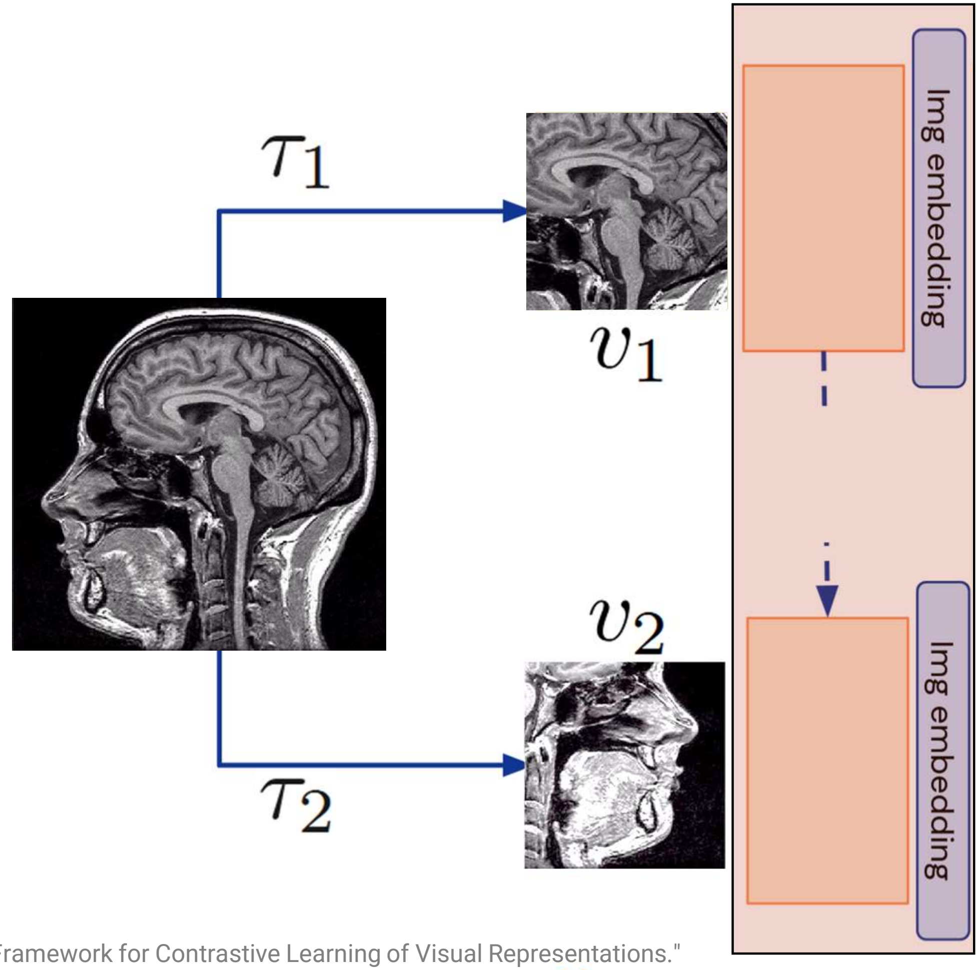
ICAI Innovation Center for Artificial Intelligence
SIEMENS Healthineers
NWO
UNIVERSITY OF TWENTE.

Being Consistent Helps! Alignment Helps!

- These models are our babies:
  - We need to be consistent in our loss terms.
  - We need to align them to our domain.
  - We need to be patient and let them learn.

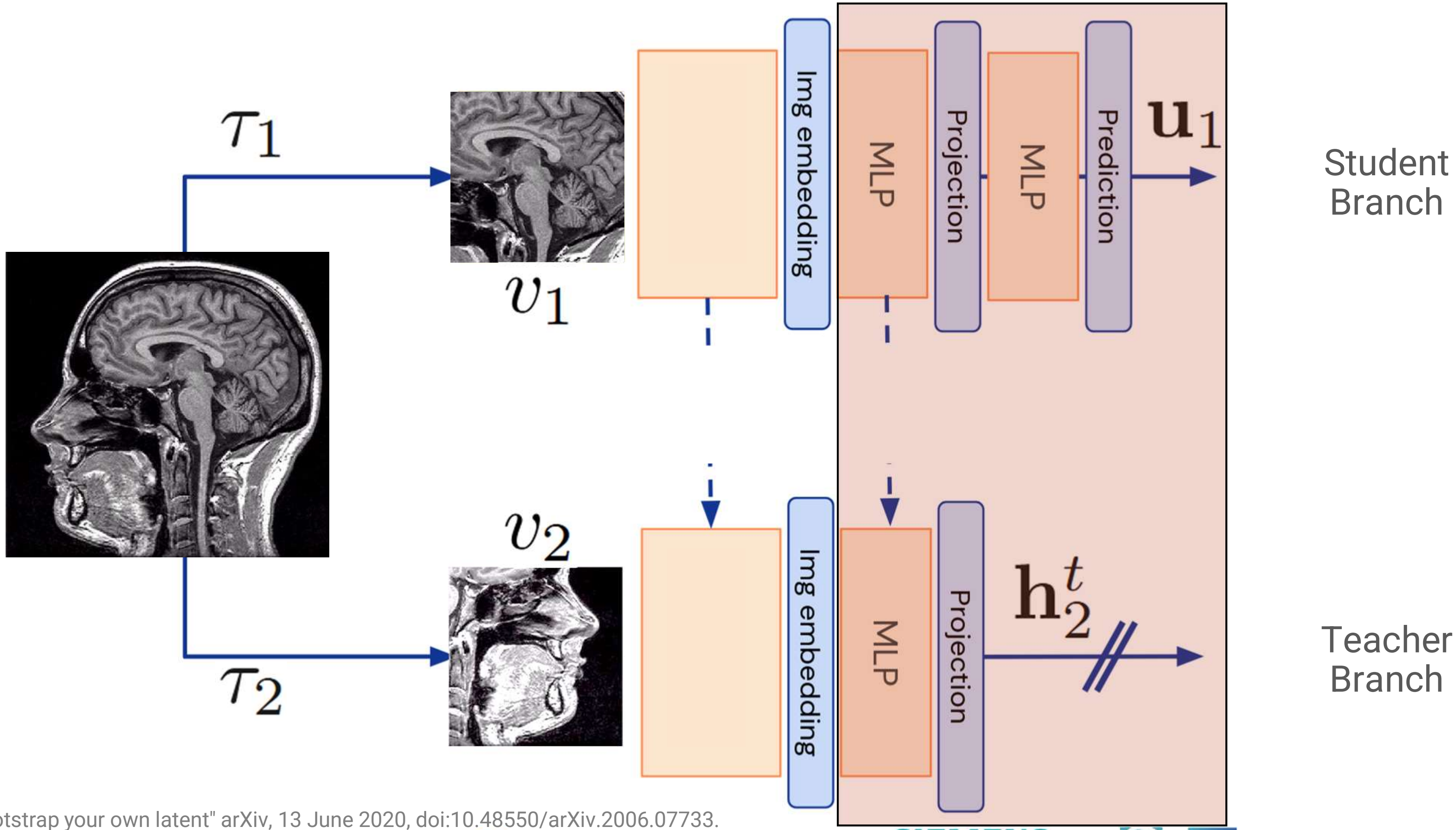
ICAI Innovation Center for Artificial Intelligence SIEMENS Healthineers NWO UNIVERSITY OF TWENTE.

State of Self-Supervised Learning



Chen, Ting, et al. "A Simple Framework for Contrastive Learning of Visual Representations." arXiv, 13 Feb. 2020, doi:10.48550/arXiv.2002.05709.

State of Self-Supervised Learning



Grill, Jean-Bastien, et al. "Bootstrap your own latent" arXiv, 13 June 2020, doi:10.48550/arXiv.2006.07733.

Quick Ablations

Track	Recon.	Consis.	Cont.	Avg Rank	Seg Rank	Cls Rank	Segmentation								Classification		
							ISL		YBM		GLI		MSD		ABD II		
							DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	Bal Acc.	AUROC	AP
ResEnc-L	AE	✗	✗	6.08	6.63	5.00	77.34	75.57	60.92	69.44	68.38	73.41	72.66	76.64	57.30	60.61	60.03
	MAE	✗	✗	5.22	5.00	5.67	78.87	76.66	61.21	68.68	69.83	75.02	72.22	76.49	57.33	60.18	58.89
	MAE	CVA	✗	3.83	4.25	3.00	77.98	76.23	62.10	70.97	69.15	74.45	72.84	77.15	60.14	63.69	62.00
	MAE	C-CVA	✗	5.14	4.38	6.67	78.58	76.81	62.27	70.41	69.55	74.71	72.72	76.89	56.43	59.64	59.06
	MAE	✗	✓	2.53	3.13	**1.33**	80.05	78.18	62.31	70.37	69.82	74.84	72.80	76.70	61.09	64.93	62.67
	MAE	CVA	✓	**2.47**	2.88	1.67	78.97	77.09	62.35	70.94	69.75	74.85	72.84	77.02	62.02	64.46	62.62
	MAE	C-CVA	✓	2.72	**1.75**	4.67	79.65	77.90	62.43	70.30	69.94	75.18	72.86	77.24	57.17	62.48	61.60
				Range			2.70	2.61	2.02	2.28	1.67	1.83	0.64	0.74	5.60	5.28	3.78
Primus-M	AE	✗	✗	5.03	6.38	2.33	76.05	73.35	51.92	58.43	63.35	69.93	71.44	75.90	56.09	61.79	60.51
	MAE	✗	✗	6.31	6.13	6.67	77.18	74.98	52.70	59.01	65.82	72.58	71.41	75.44	54.80	58.75	58.26
	MAE	CVA	✗	4.64	4.63	4.67	77.18	75.00	53.58	59.95	65.96	72.73	71.56	75.54	55.83	59.17	58.38
	MAE	C-CVA	✗	2.97	2.63	3.67	77.40	75.01	53.42	59.36	67.21	73.99	71.82	76.14	55.84	59.25	58.84
	MAE	✗	✓	**2.50**	3.25	**1.00**	77.18	75.36	54.87	61.74	65.82	72.65	71.78	75.85	58.55	62.42	61.60
	MAE	CVA	✓	**2.49**	**2.13**	3.21	77.33	75.14	54.78	61.60	66.48	73.27	71.86	76.10	56.13	59.10	59.31
	MAE	C-CVA	✓	4.03	2.88	6.33	78.07	75.77	54.44	60.92	66.20	72.91	71.66	75.61	55.55	58.85	57.69
				Range			2.01	2.42	2.95	3.31	3.86	4.07	0.45	0.69	3.75	3.67	3.90

ICAI Innovation Center for Artificial Intelligence SIEMENS Healthineers NWO UNIVERSITY OF TWENTE.

Ablations – Dynamic Range

Track	Recon.	Consis.	Cont.	Avg Rank	Seg Rank	Cls Rank	Segmentation								Classification		
							ISL		YBM		GLI		MSD		ABD II		
							DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	Bal Acc.	AUROC	AP
ResEnc-L	AE	✗	✗	6.08	6.63	5.00	77.34	75.57	60.92	69.44	68.38	73.41	72.66	76.64	57.30	60.61	60.03
	MAE	✗	✗	5.22	5.00	5.67	78.87	76.66	61.21	68.68	69.83	75.02	72.22	76.49	57.33	60.18	58.89
	MAE	CVA	✗	3.83	4.25	3.00	77.98	76.23	62.10	70.97	69.15	74.45	72.84	77.15	60.14	63.69	62.00
	MAE	C-CVA	✗	5.14	4.38	6.67	78.58	76.81	62.27	70.41	69.55	74.71	72.72	76.89	56.43	59.64	59.06
	MAE	✗	✓	2.53	3.13	**1.33**	80.05	78.18	62.31	70.37	69.82	74.84	72.80	76.70	61.09	64.93	62.67
	MAE	CVA	✓	**2.47**	2.88	1.67	78.97	77.09	62.35	70.94	69.75	74.85	72.84	77.02	62.02	64.46	62.62
	MAE	C-CVA	✓	2.72	**1.75**	4.67	79.65	77.90	62.43	70.30	69.94	75.18	72.86	77.24	57.17	62.48	61.60
				Range			2.70	2.61	2.02	2.28	1.67	1.83	0.64	0.74	5.60	5.28	3.78
Primus-M	AE	✗	✗	5.03	6.38	2.33	76.05	73.35	51.92	58.43	63.35	69.93	71.44	75.90	56.09	61.79	60.51
	MAE	✗	✗	6.31	6.13	6.67	77.18	74.98	52.70	59.01	65.82	72.58	71.41	75.44	54.80	58.75	58.26
	MAE	CVA	✗	4.64	4.63	4.67	77.18	75.00	53.58	59.95	65.96	72.73	71.56	75.54	55.83	59.17	58.38
	MAE	C-CVA	✗	2.97	2.63	3.67	77.40	75.01	53.42	59.36	67.21	73.99	71.82	76.14	55.84	59.25	58.84
	MAE	✗	✓	**2.50**	3.25	**1.00**	77.18	75.36	54.87	61.74	65.82	72.65	71.78	75.85	58.55	62.42	61.60
	MAE	CVA	✓	**2.49**	**2.13**	3.21	77.33	75.14	54.78	61.60	66.48	73.27	71.86	76.10	56.13	59.10	59.31
	MAE	C-CVA	✓	4.03	2.88	6.33	78.07	75.77	54.44	60.92	66.20	72.91	71.66	75.61	55.55	58.85	57.69
				Range			2.01	2.42	2.95	3.31	3.86	4.07	0.45	0.69	3.75	3.67	3.90