Carnegie Mellon University

Cleveland Clinic

Multimodal Representation Learning of Cardiovascular Magnetic Resonance Imaging

Jielin Qiu[*1], Peide Huang[*1], Makiya Nakashima[2], Jaehyun Lee[2], Jiacheng Zhu[1], Wilson Tang[2], Pohao Chen[2], Christopher Nguyen[2], Byung-Hak Kim[3], Debbie Kwon[2], Douglas Weber[2], Ding Zhao[1], David Chen[2]

[1] *Carnegie Mellon University,*

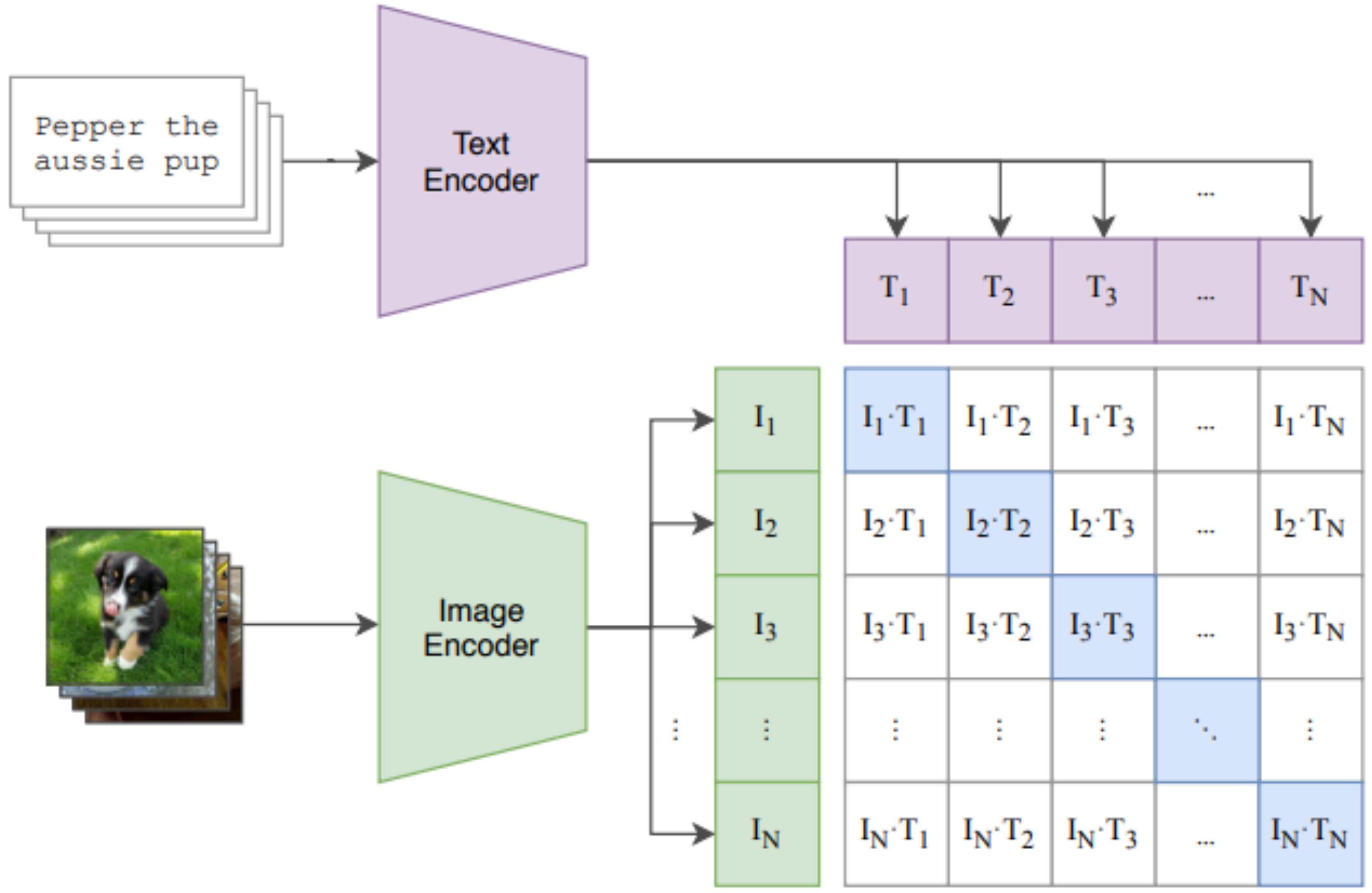
[2] *Heart Vascular and Thoracic Institute, Cleveland Clinic,*

[3] *CJ AI Center*

[*] *marked as equal contribution*

Motivation

- Contrastive image-text pre-training leverage the natural alignment between image and text pairs to provide co-supervision for each domain and achieved good performance on the natural image-text pairs collected from Internet.

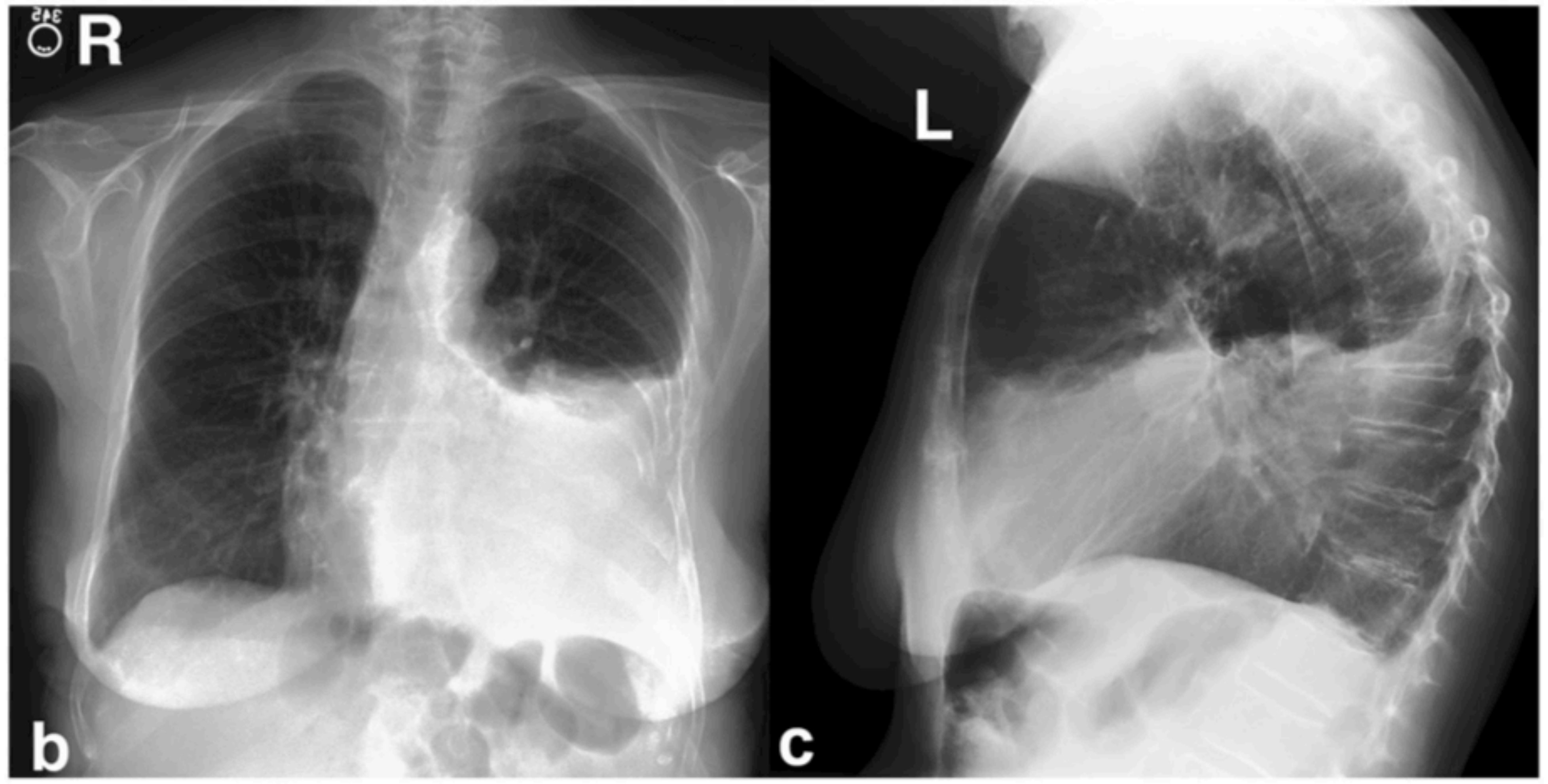


CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples

Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.

Motivation

- Contrastive image-text pre-training is crucial for clinical imaging applications, given the lack of explicit labels in healthcare.



chest radiographs (frontal and lateral)

EXAMINATION: CHEST (PA AND LAT)

INDICATION: ___ year old woman with ?pleural effusion // ?pleural effusion

TECHNIQUE: Chest PA and lateral

COMPARISON: ___

FINDINGS:

Cardiac size cannot be evaluated. Large left pleural effusion is new. Small right effusion is new. The upper lungs are clear. Right lower lobe opacities are better seen in prior CT. There is no pneumothorax. There are mild degenerative changes in the thoracic spine

IMPRESSION:

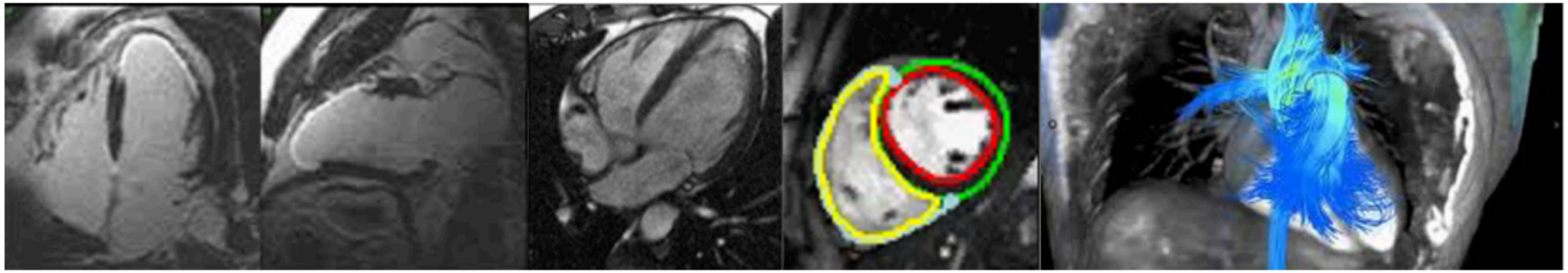
Large left pleural effusion

radiology report

Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data 6, 317 (2019). https://doi.org/10.1038/s41597-019-0322-0

What Is Cardiac Magnetic Resonance Imaging (CMR)?

- CMR allows to visualize the 3D cardiac anatomy and function in an unlimited number of views.
- CMR studies are able to visualize the morphology, motion, tissue characteristics, and even tissue perfusion within a single study.
- Each type of image has different characteristics, which make them sensitive to different pathophysiologies.
- The associated radiology report incorporates findings that describe both individual images and findings that synthesize from multiple image types and views.

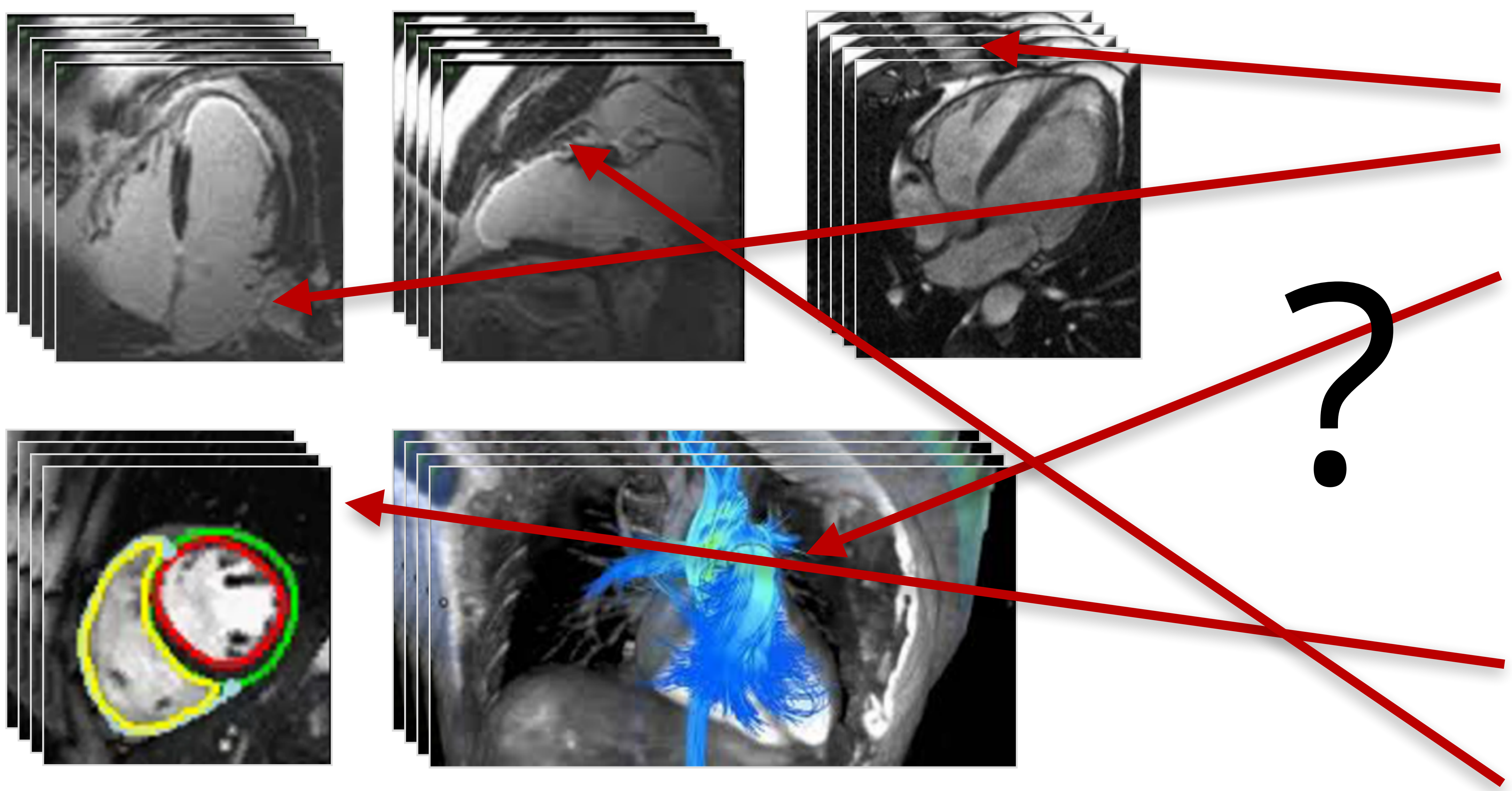


Examples of Cardiac magnetic resonance (CMR) images

Carnegie Mellon University Cleveland Clinic

Challenges in Contrastive Pre-training for CMR

- Weak alignment between hundreds of images (with different views and types) and a clinical report
  - Information Synthesize from both a single frame and motion from a series of frames
  - Multiple co-morbidities of the patients
- Far less data available compared to other popular clinical modalities



Impression:
1. There are no definite findings to suggest prior ischemic damage or an infiltrative process. There is mild patchy increased signal on delayed imaging in the mid inferior septum at the RV insertion point, which is a non-specific finding, and suggestive of mild interstitial fibrosis.
2. The left ventricle is dilated with concentric hypertrophy and moderately reduced systolic function, EF 39%. There are no segmental wall motion abnormalities. Quantitative values are as noted above.
3. The right ventricle is dilated with moderately reduced systolic function, EF 38%.
4. Normal aortic, mitral, and tricuspid valve function.
5. Mildly dilated aortic root measuring 4.0-cm. Mildly prominent ascending thoracic aorta measuring 3.9-cm.
6. Mildly dilated main pulmonary artery, suggestive of pulmonary hypertension.

Carnegie Mellon University Cleveland Clinic

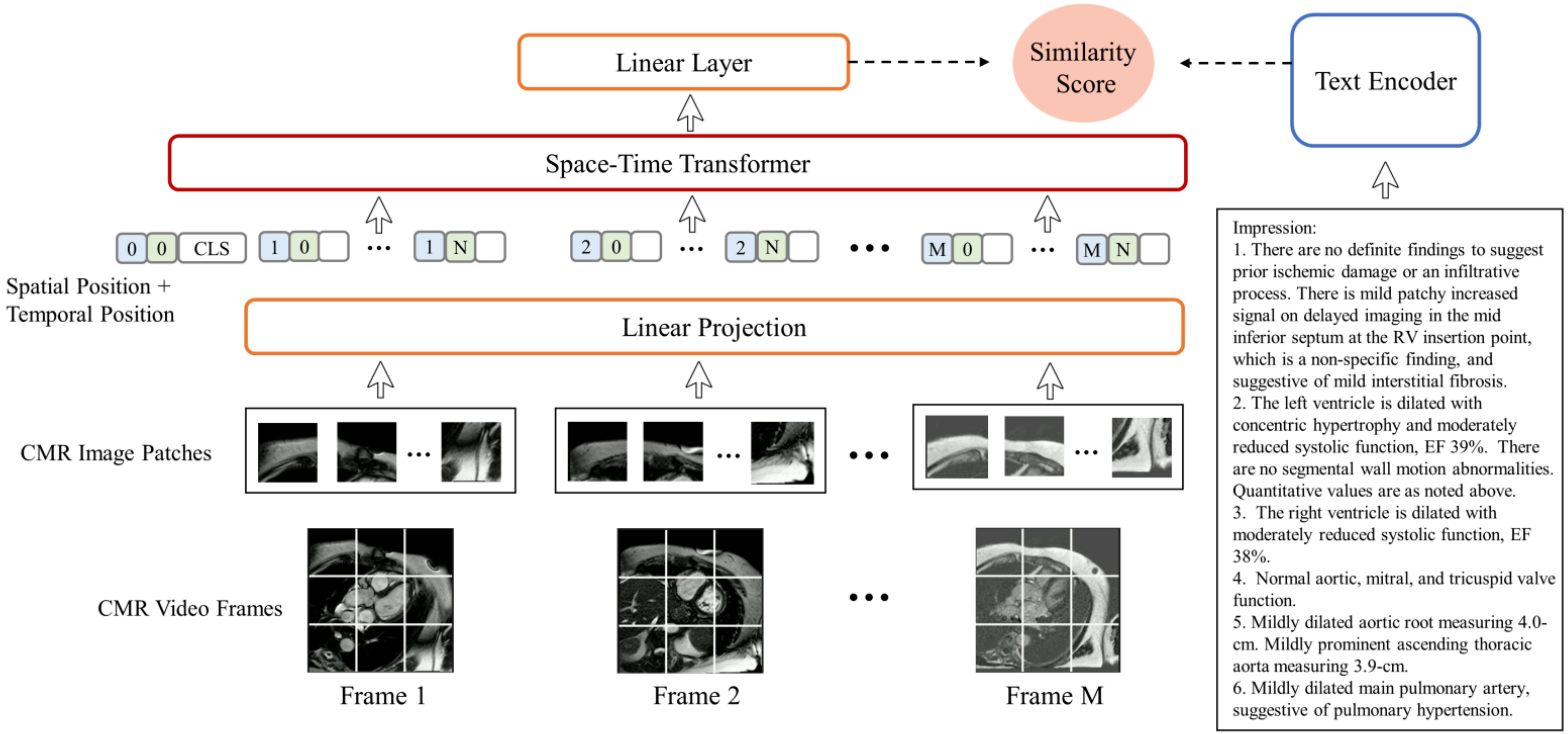
Contributions of Our Work

CMRformer
A multimodal learning framework that addresses the weak alignment problem of CMR

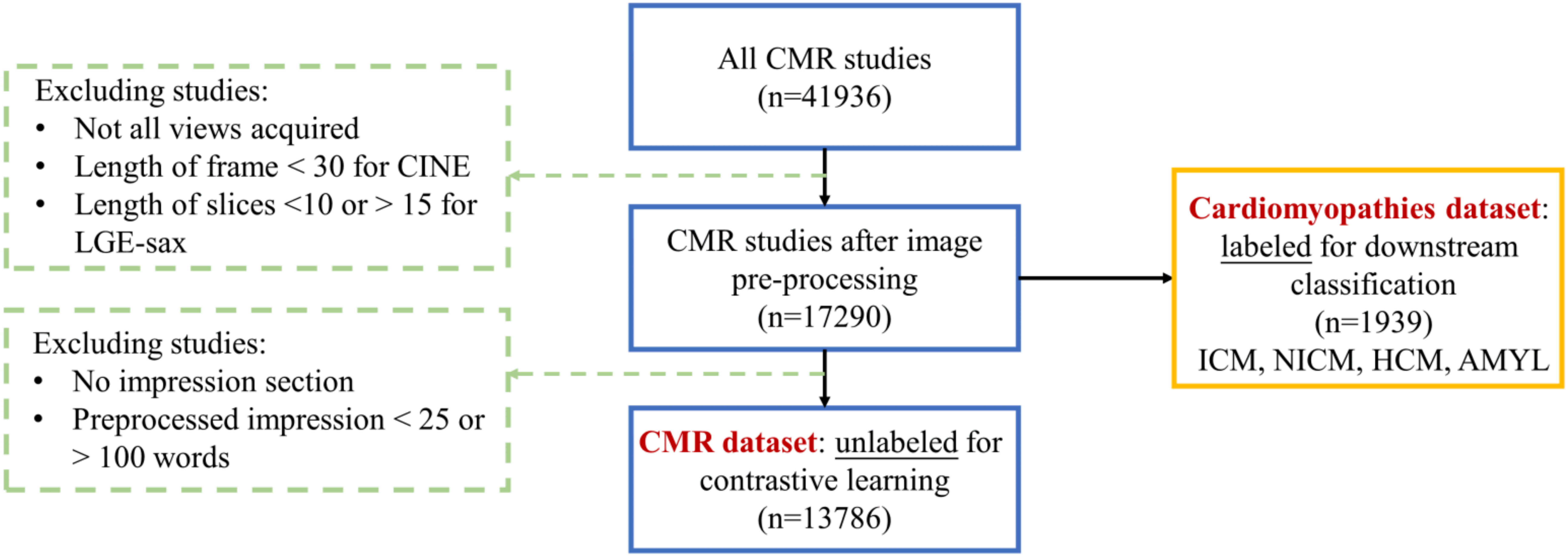
CMR dataset
A comprehensive dataset consisting of 13,786 studies derived from actual clinical cases

Cardiomyopathies dataset
An expert-labeled dataset of 1939 studies for the diagnosis of various cardiomyopathies

Overall Architecture



Data Prepossessing of the Dataset



Carnegie Mellon University Cleveland Clinic

Statistics and Comparison with existing CMR datasets

Table 1: Statistics of length of impression sections from text reports.

Text Length	Count	Percentage
20-30	728	5.3%
30-40	2445	17.7%
40-50	2926	21.2%
50-60	2666	19.3%
60-70	2078	15.1%
70-80	1425	10.3%
80-90	929	6.7%
90-100	589	4.3%

Table 2: Statistics of the number of images of each study.

Number of Images	Count	Percentage
0-200	61	0.4%
200-300	24	0.2%
300-400	1166	8.5%
400-500	12054	87.4%
500-600	126	0.9%
600-700	79	0.6%
700-800	162	1.2%
> 800	114	0.8%

Table 3: Comparison with existing CMR datasets.

Source	Studies	Image Types	Labels
ACDC	150	Cine	segmentation
DSB-CC	1,140	Cine	end-systolic and end-diastolic volumes
STACOM	<200	varies (mostly Cine)	varies (mostly segmentation)
Ours	13,786/1,939	Cine, LGE	radiology reports/cardiomyopathy diagnosis

CMR Sequences as Video Input

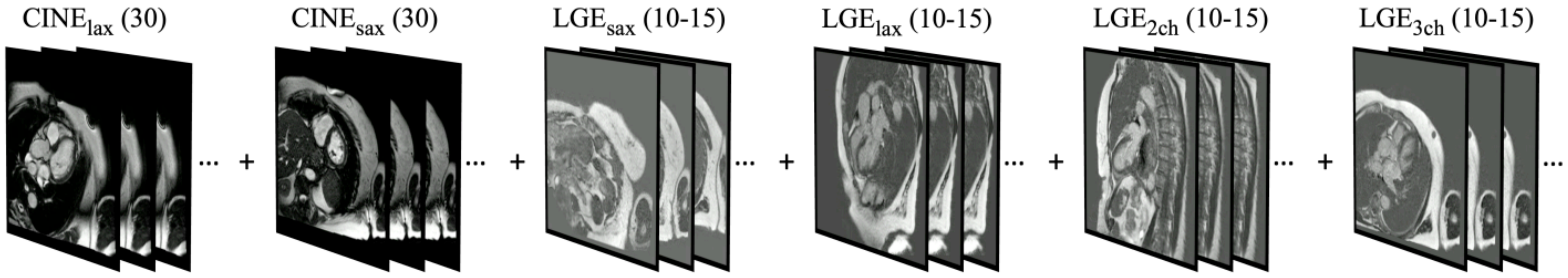


Figure 4: Example of CMR image sequences constructed by $CINE_{\text{lax-sax}} + LGE_{\text{lax-sax-2ch-3ch}}$, where $(\cdot)$ represents the number of images of each type-view combination. For $CINE_{\text{lax-sax}}$, each frame represents the time dimension. For LGE_{sax}, each frame corresponds to the depth dimension, and for $LGE_{\text{lax-2ch-3ch}}$, each image is duplicated to be consistent with LGE_{sax}.

Experimental Results for Retrieval Tasks

- Learned representations showed better performance than zero-shot results
- More types/views contributed better performance
- Increasing the number of CMR images resulted in better performance

Table 4: Experimental results for retrieval experiments. (·) represents the number of input frames. Zero-Shot evaluation was done using $\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$.

Method	Text-to-Video Retrieval R@5	R@10	R@50	Video-to-Text Retrieval R@5	R@10	R@50	RSUM
Zero-shot (16)	0.3	0.4	1.8	0.2	0.4	1.5	4.6
CINE_{sax} (8)	9.4	15.0	38.6	9.2	14.9	39.1	126.2
$\text{CINE}_{\text{lax-sax}}$ (8)	13.9	21.1	45.2	13.3	19.9	44.3	157.7
$\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (8)	14.1	22.3	50.3	14.2	22.3	50.8	174.1
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax}}$ (16)	16.4	23.9	54.0	15.4	23.9	54.6	188.1
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (1)	6.3	9.7	27.0	6.3	9.6	27.6	86.7
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (4)	14.5	21.8	46.7	14.0	21.8	45.3	164.0
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (8)	14.8	23.7	51.0	14.4	23.4	51.1	178.5
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (16)	17.9	25.9	53.1	17.3	26.0	54.1	194.3
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (32)	17.7	26.5	55.3	17.8	26.1	56.2	199.8
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (64)	**18.5**	**28.1**	**56.3**	**18.1**	**27.5**	**56.4**	**204.8**

Experimental Results for Cardiomyopathies Classification Task

- We observed a correlation between the linear probing and the retrieval performance,
  - CMRformer learned valuable CMR representations that are transferrable to downstream tasks.

Table 5: Linear probing results on the Cardiomyopathies dataset for downstream disease classification task.

Model	NICM			ICM		
	Acc	AUC	F1	Acc	AUC	F1
Zero-shot	0.69	0.69	0.71	0.77	0.62	0.41
SimCLR	0.71	0.71	0.74	0.75	0.62	0.40
CINE_{sax} (8)	0.75	0.75	0.77	0.79	0.71	0.55
$\text{CINE}_{\text{lax-sax}}$ (8)	0.80	0.80	0.83	**0.84**	0.76	0.64
$\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (8)	0.81	0.81	0.82	**0.84**	**0.79**	**0.67**
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (16)	0.82	0.82	**0.84**	0.84	0.77	0.65
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (64)	**0.84**	**0.84**	**0.85**	**0.84**	**0.79**	**0.67**

Model	AMYL			HCM		
	Acc	AUC	F1	Acc	AUC	F1
Zero-shot	0.93	0.70	0.43	0.90	0.79	0.66
SimCLR	0.93	0.69	0.43	0.94	0.84	0.78
CINE_{sax} (8)	0.93	0.75	0.50	0.96	0.93	0.87
$\text{CINE}_{\text{lax-sax}}$ (8)	0.96	0.81	0.66	0.98	0.97	0.94
$\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (8)	0.96	0.84	0.70	0.98	0.98	0.94
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (16)	0.95	0.80	0.63	**0.99**	**0.99**	**0.97**
$\text{CINE}_{\text{lax-sax}}$ + $\text{LGE}_{\text{lax-sax-2ch-3ch}}$ (64)	**0.97**	**0.86**	**0.76**	**0.99**	**0.99**	**0.97**

Carnegie Mellon University Cleveland Clinic

Experimental Results for Classification Task on ACDC Dataset

- Our model are able to generalize to the public ACDC dataset.
- The visual embeddings obtained from our CMRformer are more categorically separated.

Model	Acc	AUC	F1
Zero-shot	0.30	0.59	0.30
SimCLR	0.42	0.82	0.42
Ours ($\mathrm{CINE_{sax}}$) (8)	**0.70**	**0.95**	**0.70**

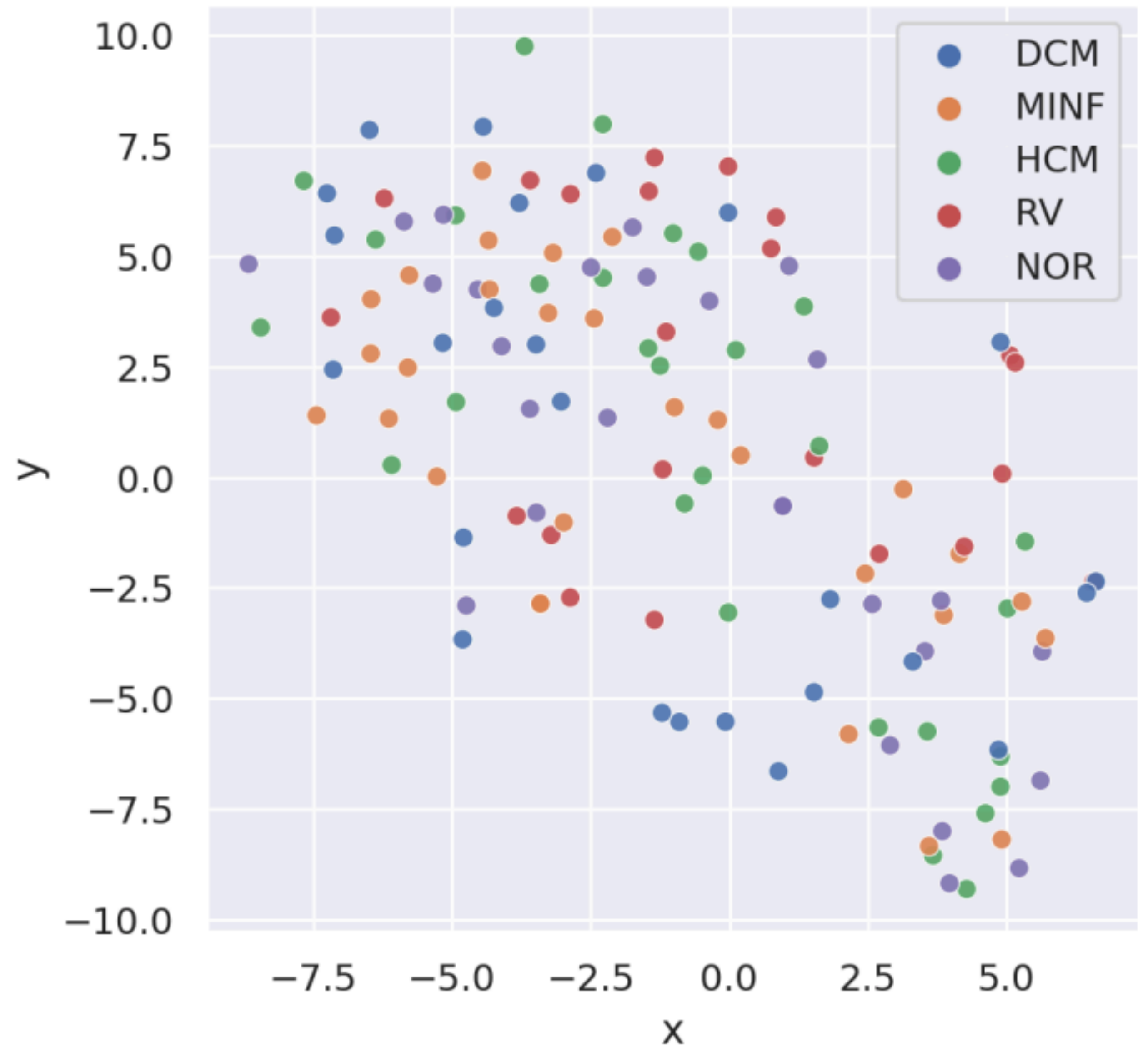


Figure 5: t-SNE visualization of zero-shot visual embeddings on the ACDC dataset.

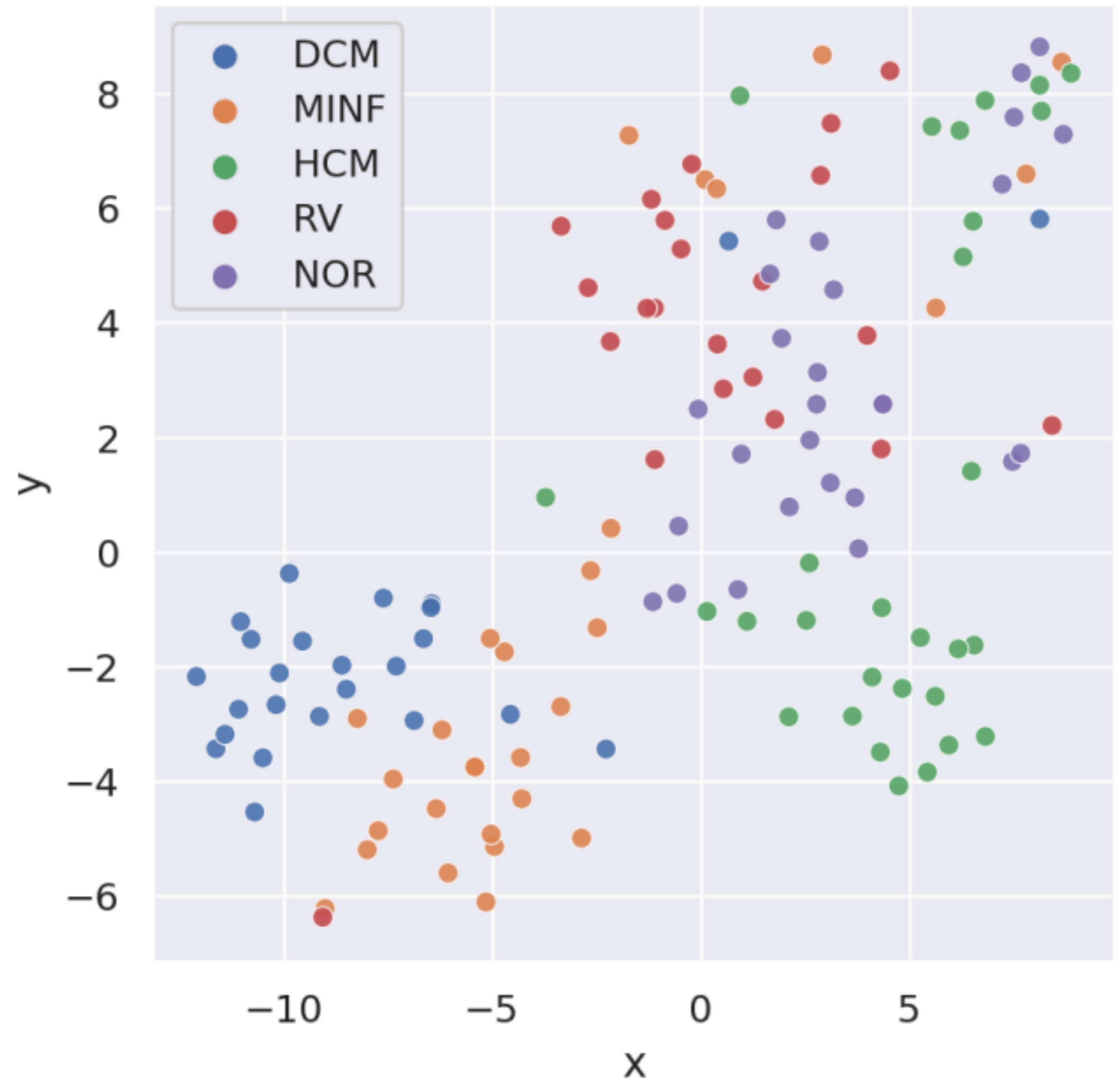


Figure 6: t-SNE visualization of learned visual embedding by CMRformer on ACDC.

Take-aways

- Vision-language contrastive learning for Cardiovascular Magnetic Resonance (CMR) is challenging due to the weak alignment between the images and text.
- We proposed **the first multimodal vision-language contrastive learning framework** that enables the acquisition of **CMR representations** accompanied by cardiologist's reports.
- We collected **a large, single-site CMR dataset consisting of 13,786 studies** derived from actual clinical cases. We also collected and labeled **a Cardiomyopathies dataset, with 1,939 studies** for the downstream disease classification task.
- We conducted extensive experiments to investigate **the retrieval performance of various types and views of CMR images**.
- We utilized the visual embeddings acquired from the visual encoder in the CMRformer and showed the **generalizability** of our trained model for **downstream image classification tasks**.