JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Best of Both Worlds: Multimodal Contrastive Learning with Tabular and Imaging Data

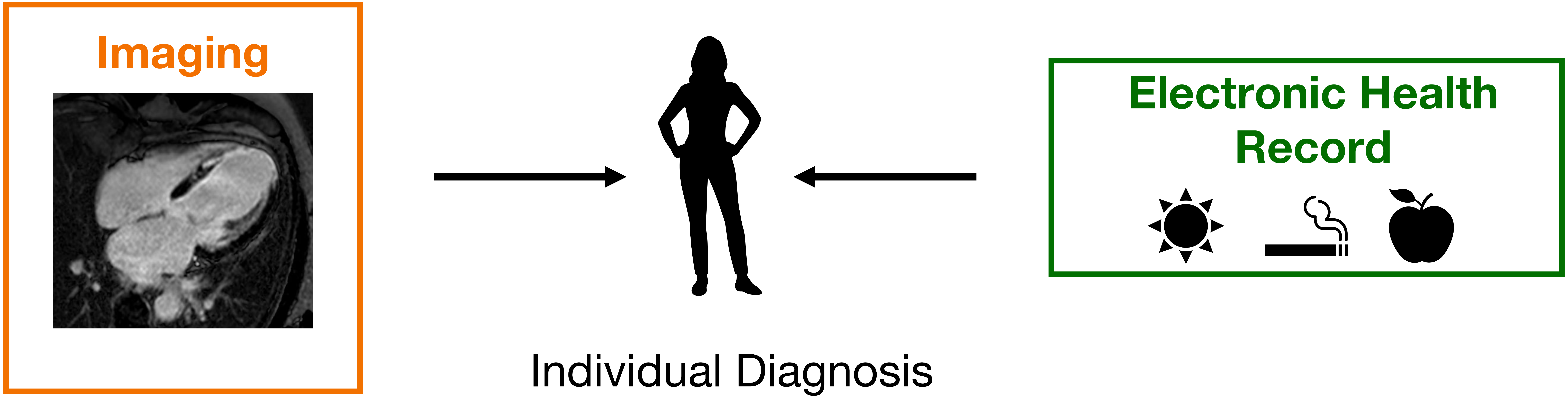
Paul Hager[1,2], Martin J Menten[1,2,3], Daniel Rückert[1,2,3]

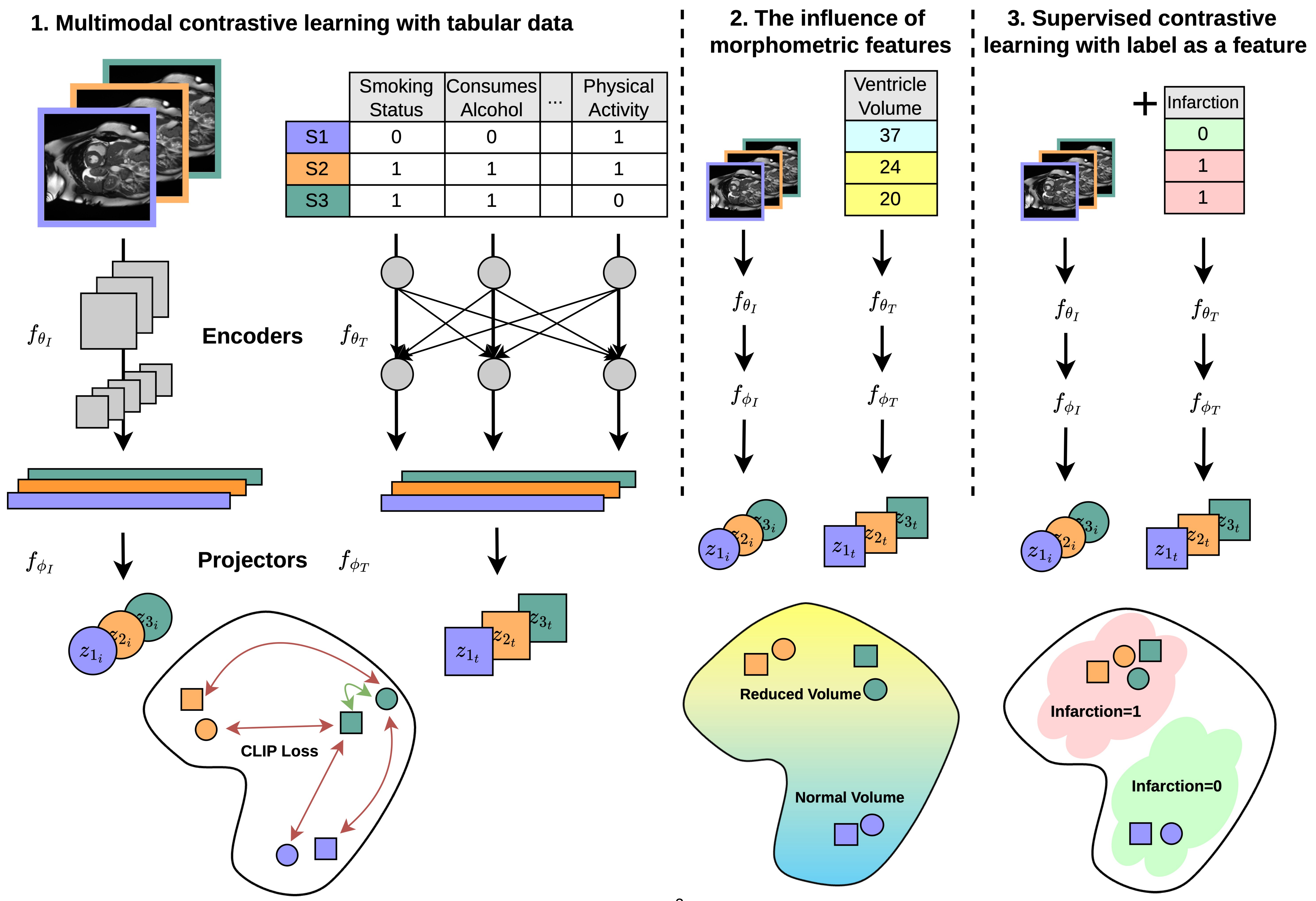
[1]Technical University of Munich, [2]Klinikum Rechts der Isar, [3]Imperial College London

THU-PM-317

Motivation

How can we learn multimodal from rich clinical data and infer unimodal using only images?




1. Multimodal contrastive learning with tabular data
Smoking Status
Consumes Alcohol
...
Physical Activity
S1
0
0
1
S2
1
1
1
S3
1
1
0
f_{θ_I}
Encoders
f_{θ_T}
f_{ϕ_I}
Projectors
f_{ϕ_T}
z_{1_i} z_{2_i} z_{3_i}
z_{1_t} z_{2_t} z_{3_t}
CLIP Loss
2. The influence of morphometric features
Ventricle Volume
37
24
20
Reduced Volume
Normal Volume
3. Supervised contrastive learning with label as a feature
+
Infarction
0
1
1
Infarction=1
Infarction=0

Setup - UK BioBank

biobank uk

- ~50k imaging subjects
- 1k+ tabular features
  - lifestyle, questionnaire, interview, physical measures, etc.

Setup - UK BioBank

- ~50k imaging subjects
  - cardiac MRI
- 1k+ tabular features
  - lifestyle, questionnaire, interview, physical measures, etc.
  - 117 with published cardiac effect

Setup - UK BioBank

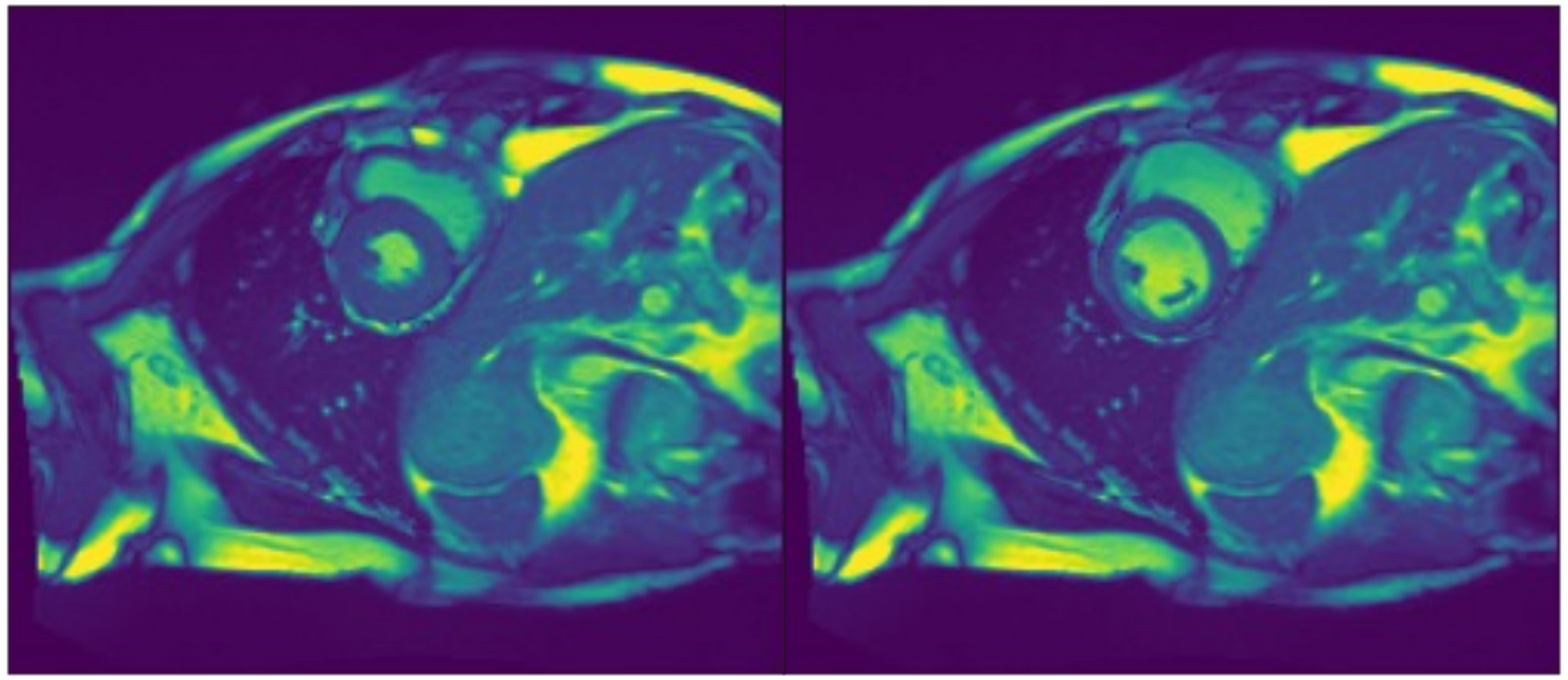
Smoker	Age	BP	BMI	Sex	Fitness	Alcohol
FALSE	62	150/90	29.2	Male	High	Moderate

- ~50k imaging subjects
  - cardiac MRI
- 1k+ tabular features
  - lifestyle, questionnaire, interview, physical measures, etc.
  - 117 with published cardiac effect

Setup - UK BioBank

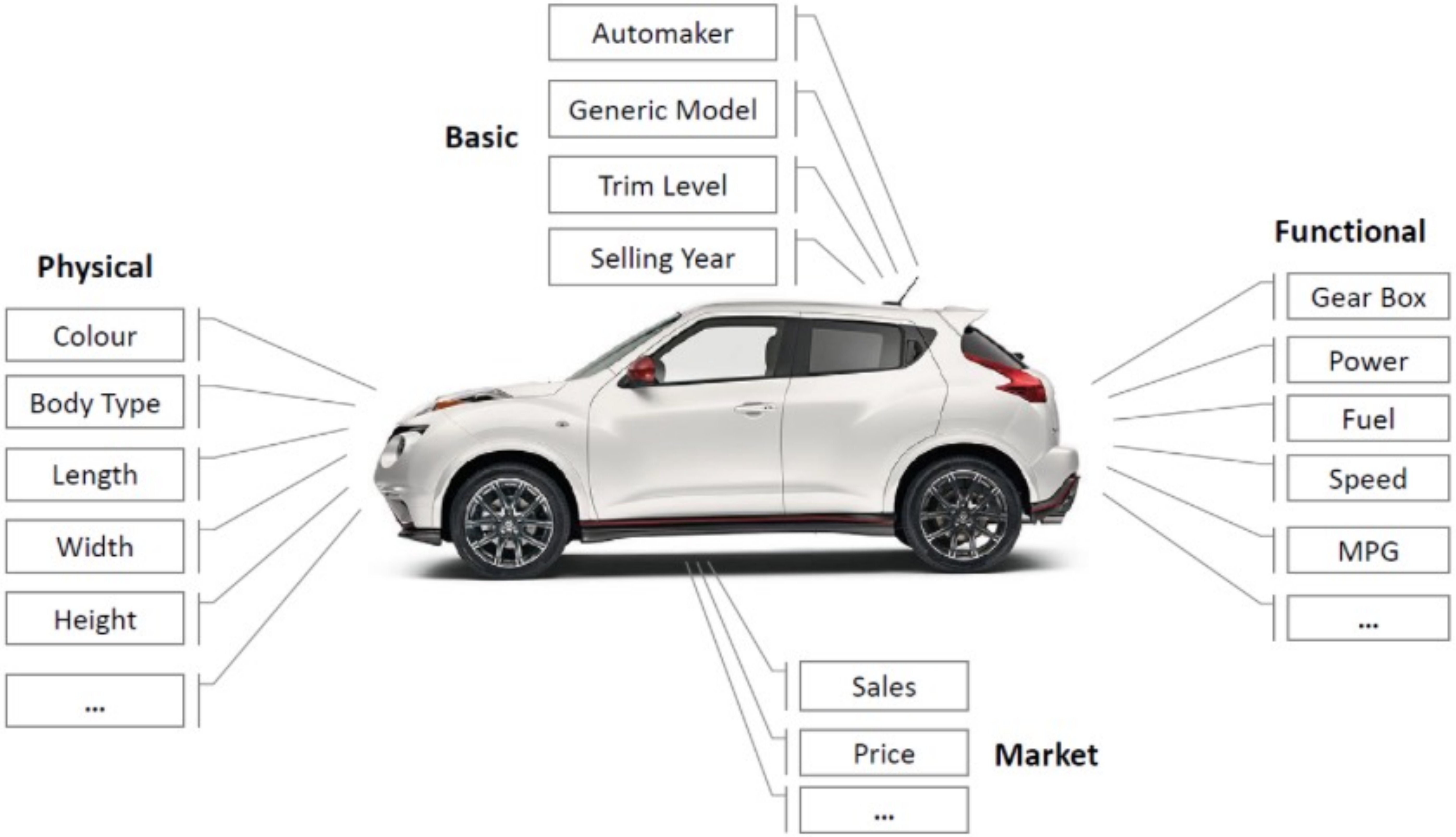
Smoker	Age	BP	BMI	Sex	Fitness	Alcohol
FALSE	62	150/90	29.2	Male	High	Moderate

- ~50k imaging subjects
    - cardiac MRI
- 1k+ tabular features
    - lifestyle, questionnaire, interview, physical measures, etc.
    - 117 with published cardiac effect



Targets:
Myocardial Infarction,
Coronary Artery Disease (CAD)

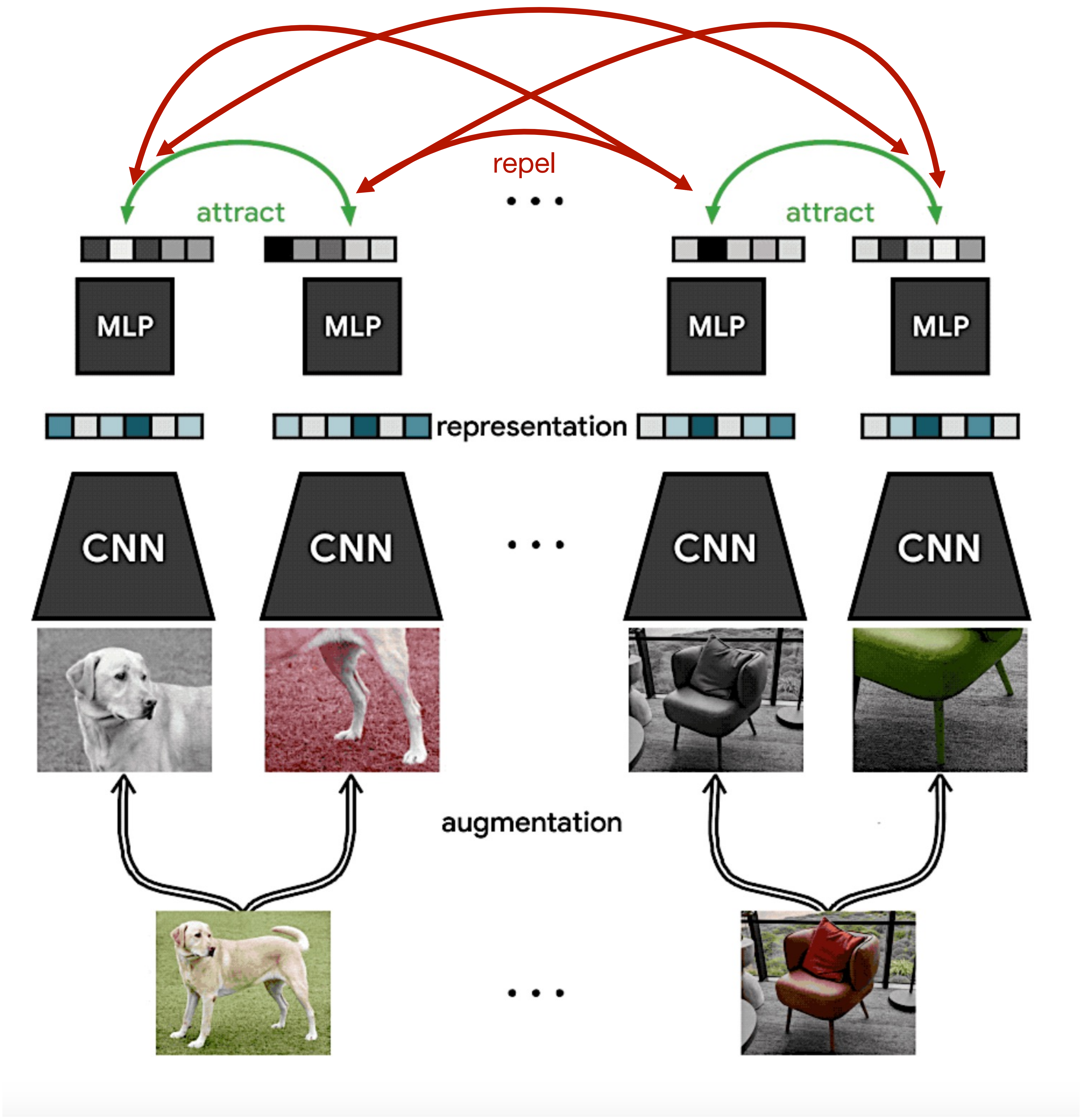
Setup - DVM Cars

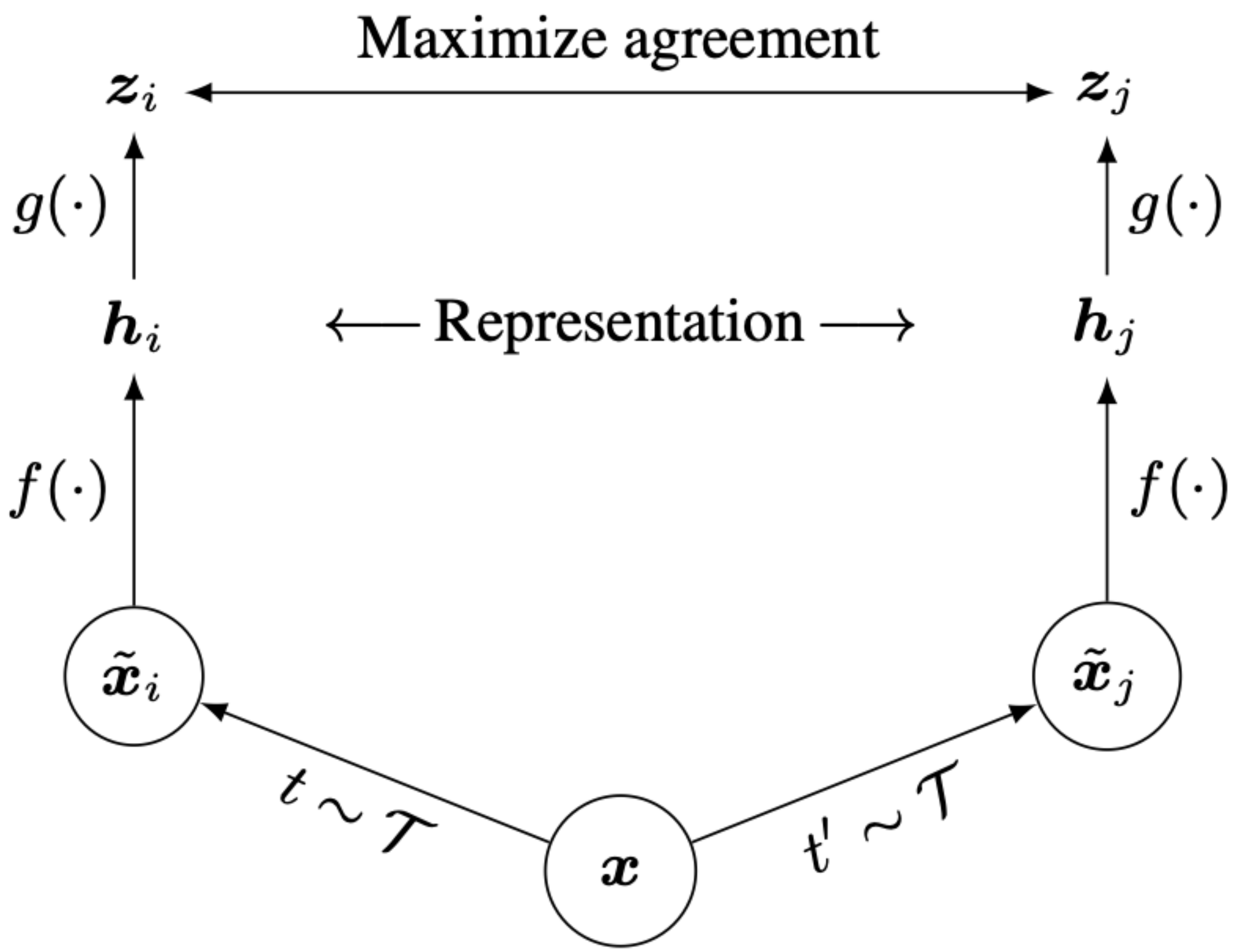


Target: **Car Model**

Huang, Jingmin, et al. "DVM-CAR: A large-scale automotive dataset for visual marketing research and applications." *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022.

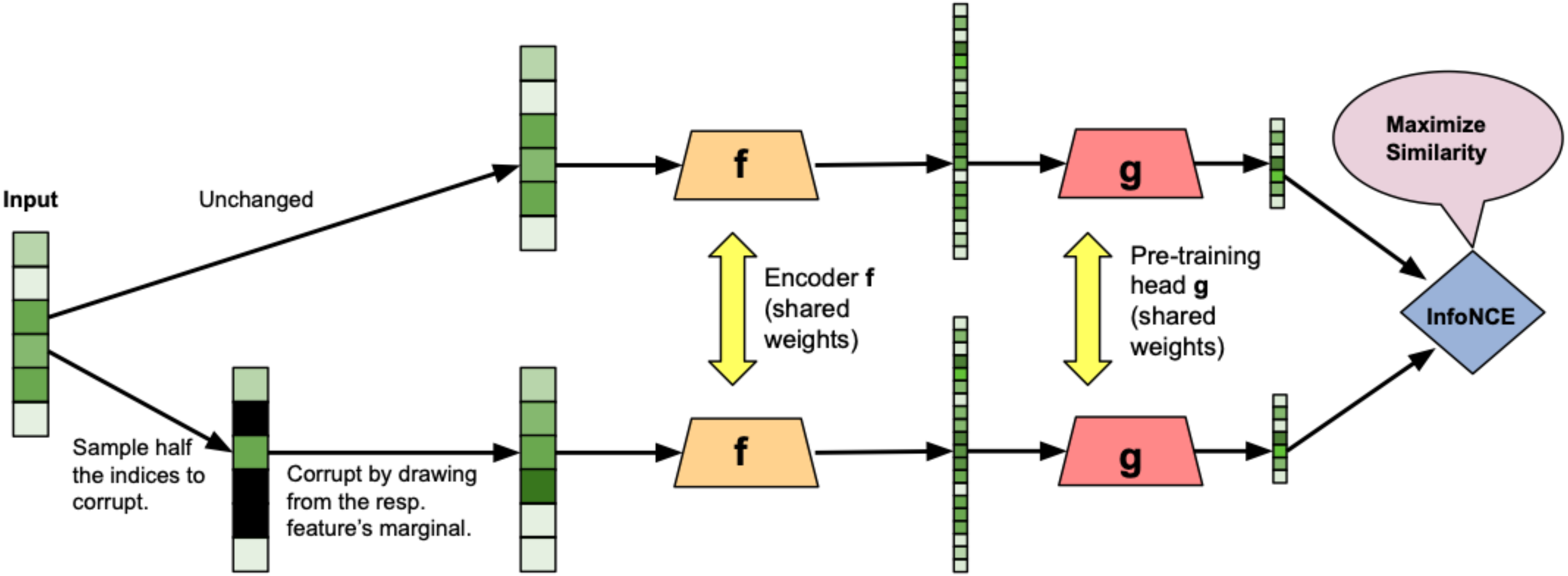
Contrastive Learning - SimCLR





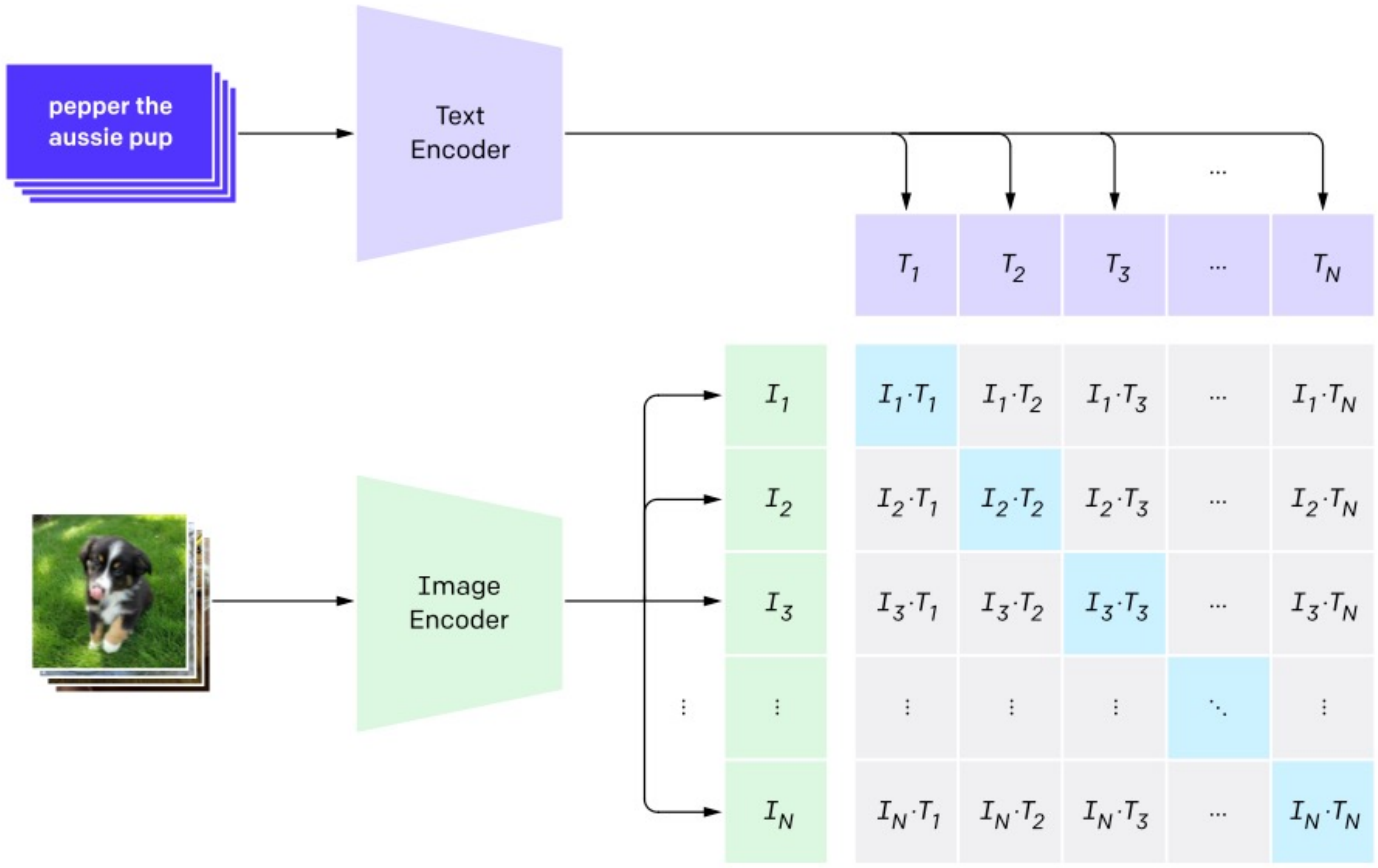
Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. *37th International Conference on Machine Learning, ICML 2020, PartF168147-3*, 1575–1585. https://doi.org/10.48550/arxiv.2002.05709

Contrastive Learning - SCARF



Bahri, Dara, et al. "Scarf: Self-Supervised Contrastive Learning using Random Feature Corruption." International Conference on Learning Representations.

Multimodal Contrastive Learning

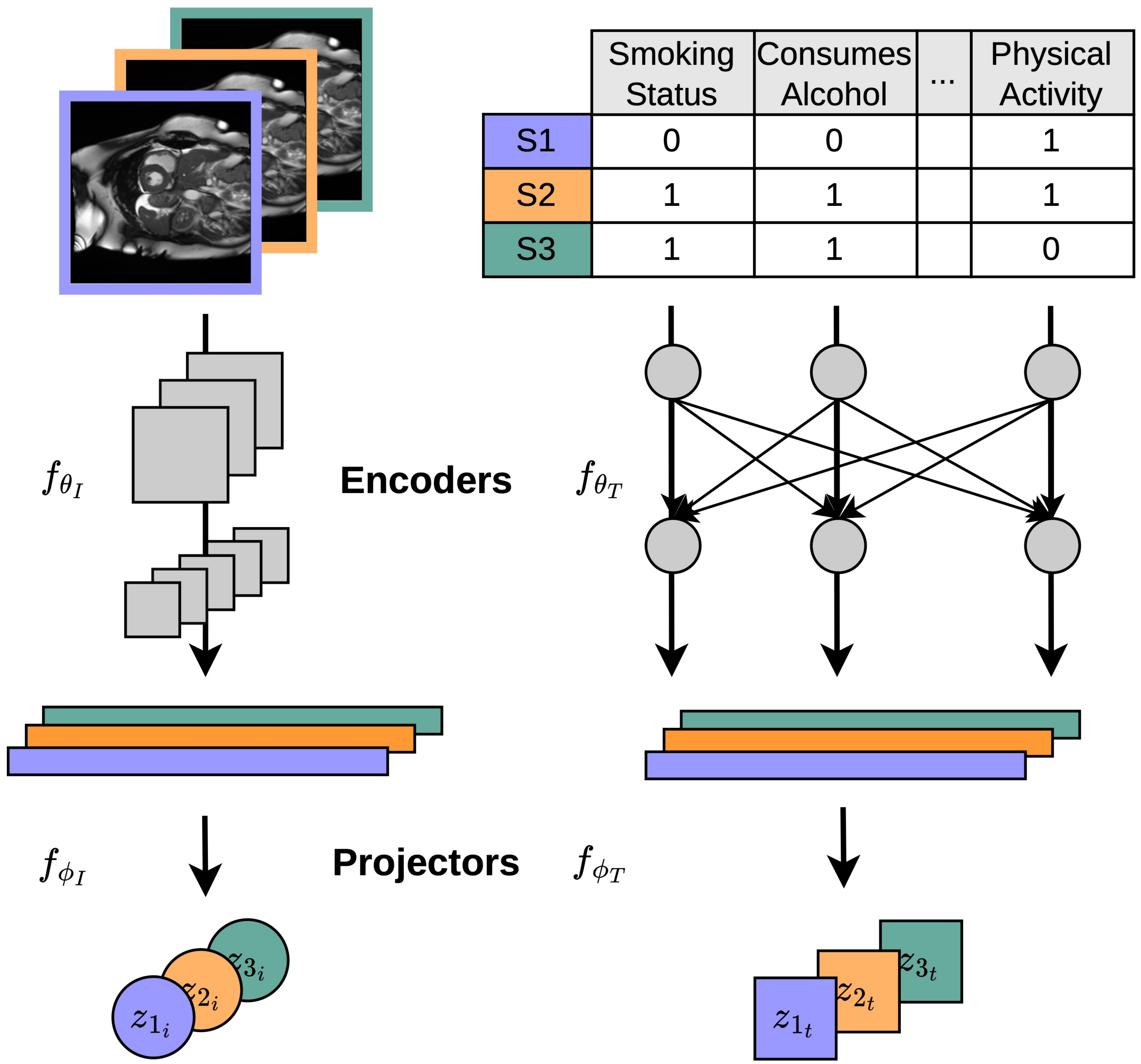


CLIP

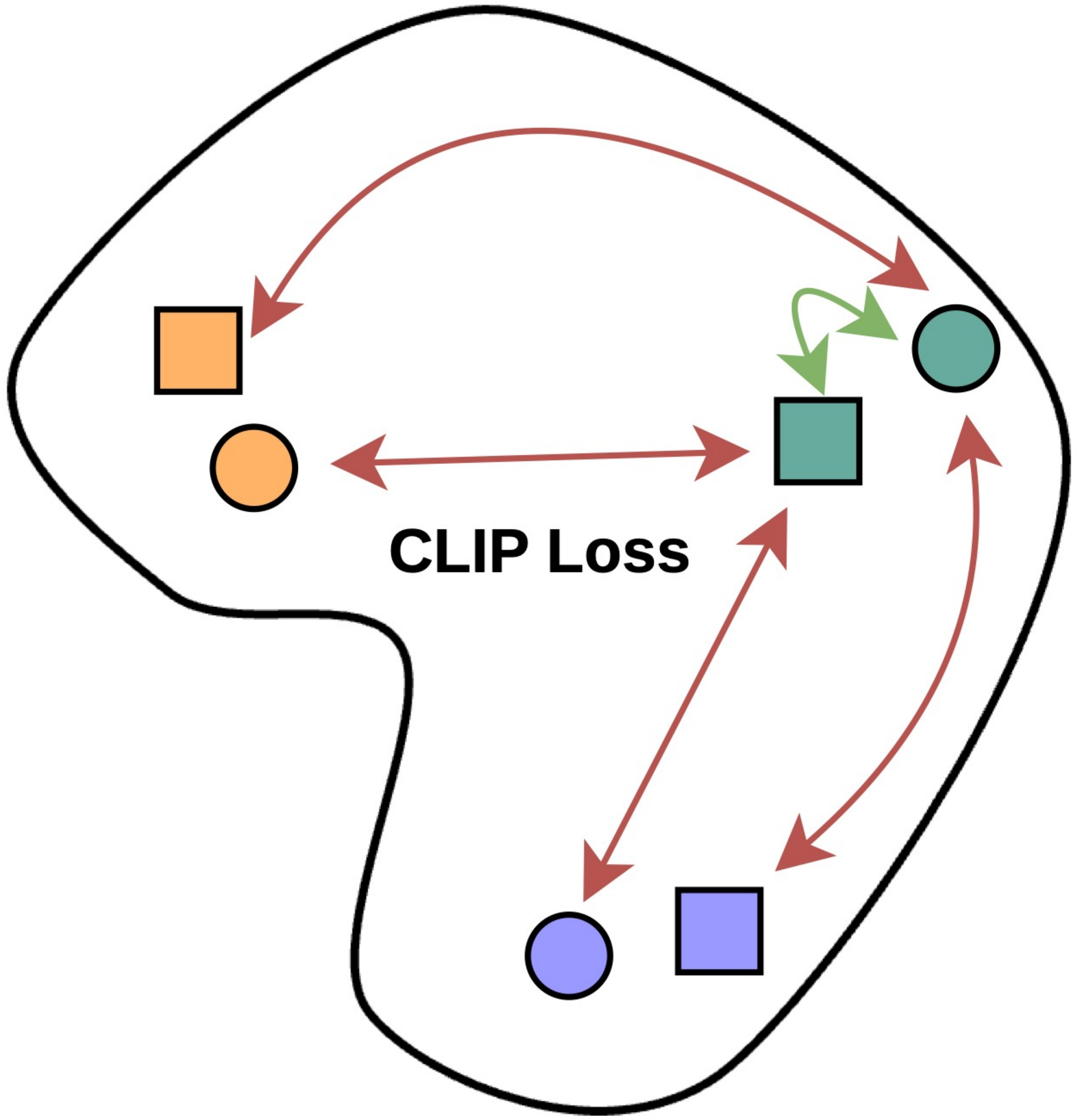
Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.

Multimodal Contrastive Learning

1. Multimodal contrastive learning with tabular data



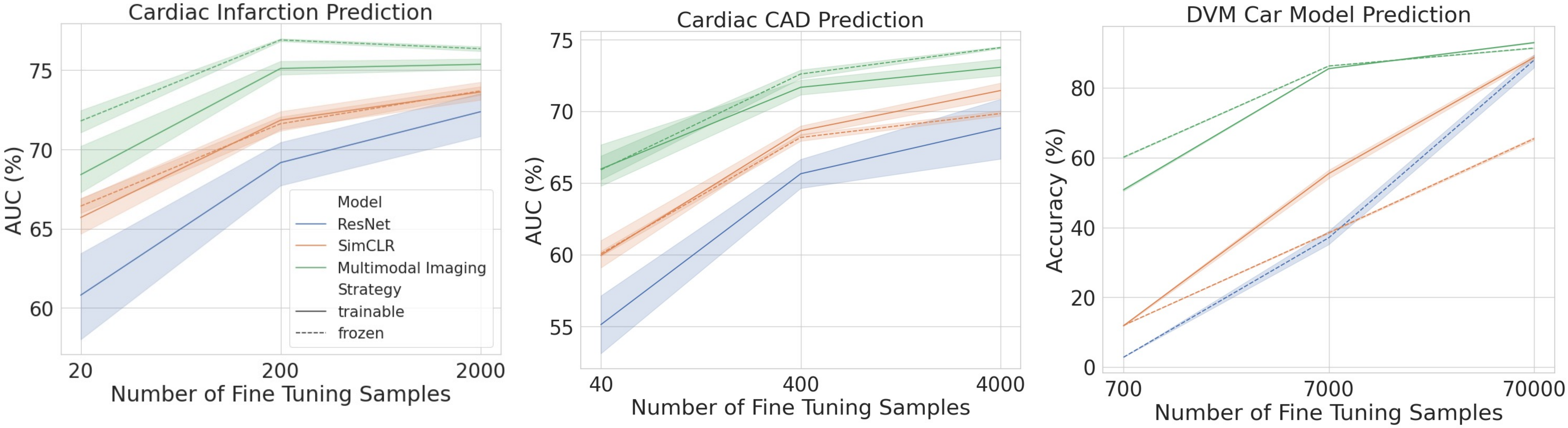
	Smoking Status	Consumes Alcohol	...	Physical Activity
S1	0	0		1
S2	1	1		1
S3	1	1		0



Multimodal Pretraining Improves Unimodal Prediction

Model	AUC (%) Frozen / Infarction	AUC (%) Trainable / Infarction	AUC (%) Frozen / CAD	AUC (%) Trainable / CAD	Top-1 Accuracy (%) Frozen / DVM	Top-1 Accuracy (%) Trainable / DVM
Supervised ResNet50	72.37±1.80	72.37±1.80	68.84±2.54	68.84±2.54	87.97±2.20	87.97±2.20
SimCLR	73.69±0.36	73.62±0.70	69.86±0.21	71.46±0.71	65.48±0.48	88.76±0.81
BYOL	69.18±0.43	70.69±2.09	66.91±0.19	70.66±0.22	59.73±0.28	89.18±0.90
SimSiam	71.72±0.18	72.31±0.26	67.79±0.12	70.13±0.35	22.11±2.83	87.43±0.88
BarlowTwins	66.06±1.11	71.35±1.23	62.90±0.23	69.63±0.58	52.57±0.08	85.47±0.82
Multimodal Imaging	**76.35±0.19**	**75.37±0.43**	**74.45±0.09**	**73.08±0.75**	**91.43±0.13**	**93.00±0.18**

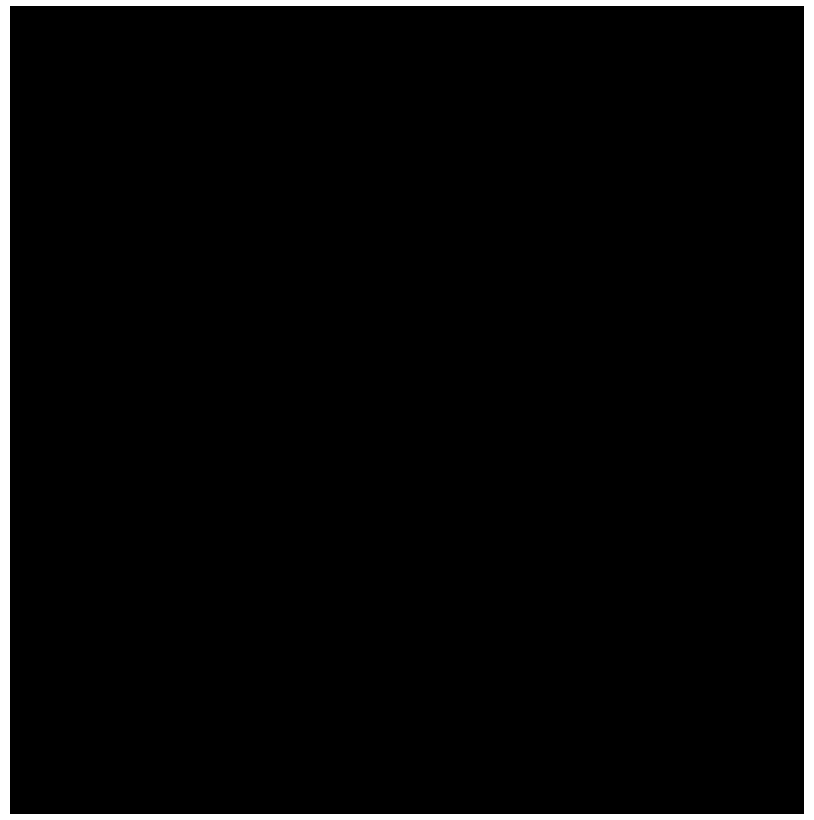
Multimodal Pretraining Is Beneficial in Low-Data Regimes



Integrated Gradients and Explainability

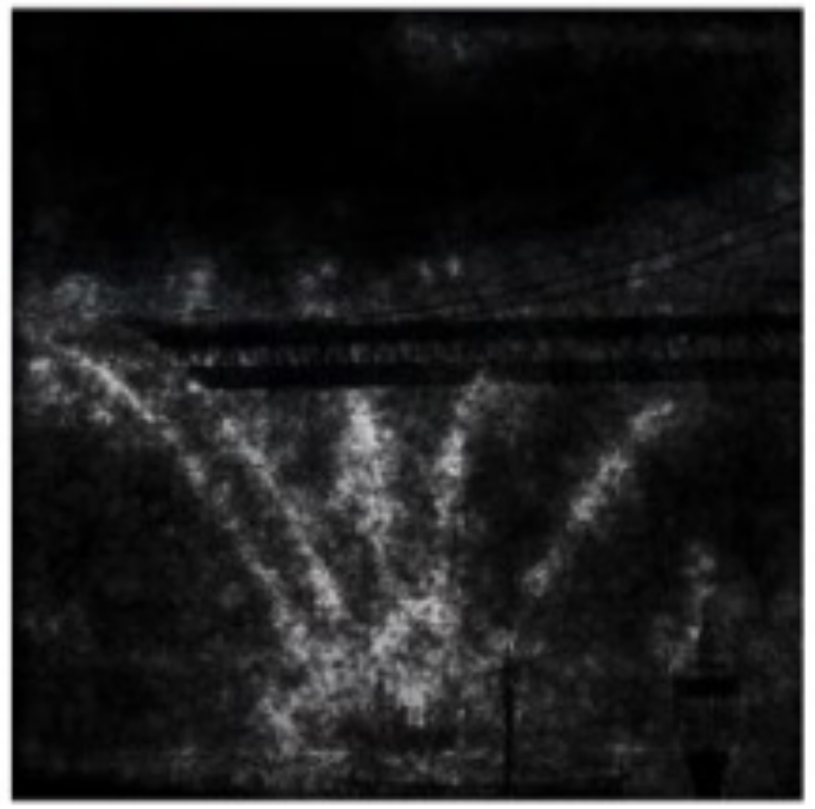
$$IG(input, base) ::= (input - base) * \int_{0-1} \nabla F(\alpha * input + (1-\alpha) * base)\, d\alpha$$

Baseline Image



Original image

Integrated Gradients



Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. "Axiomatic attribution for deep networks." *International conference on machine learning*. PMLR, 2017.

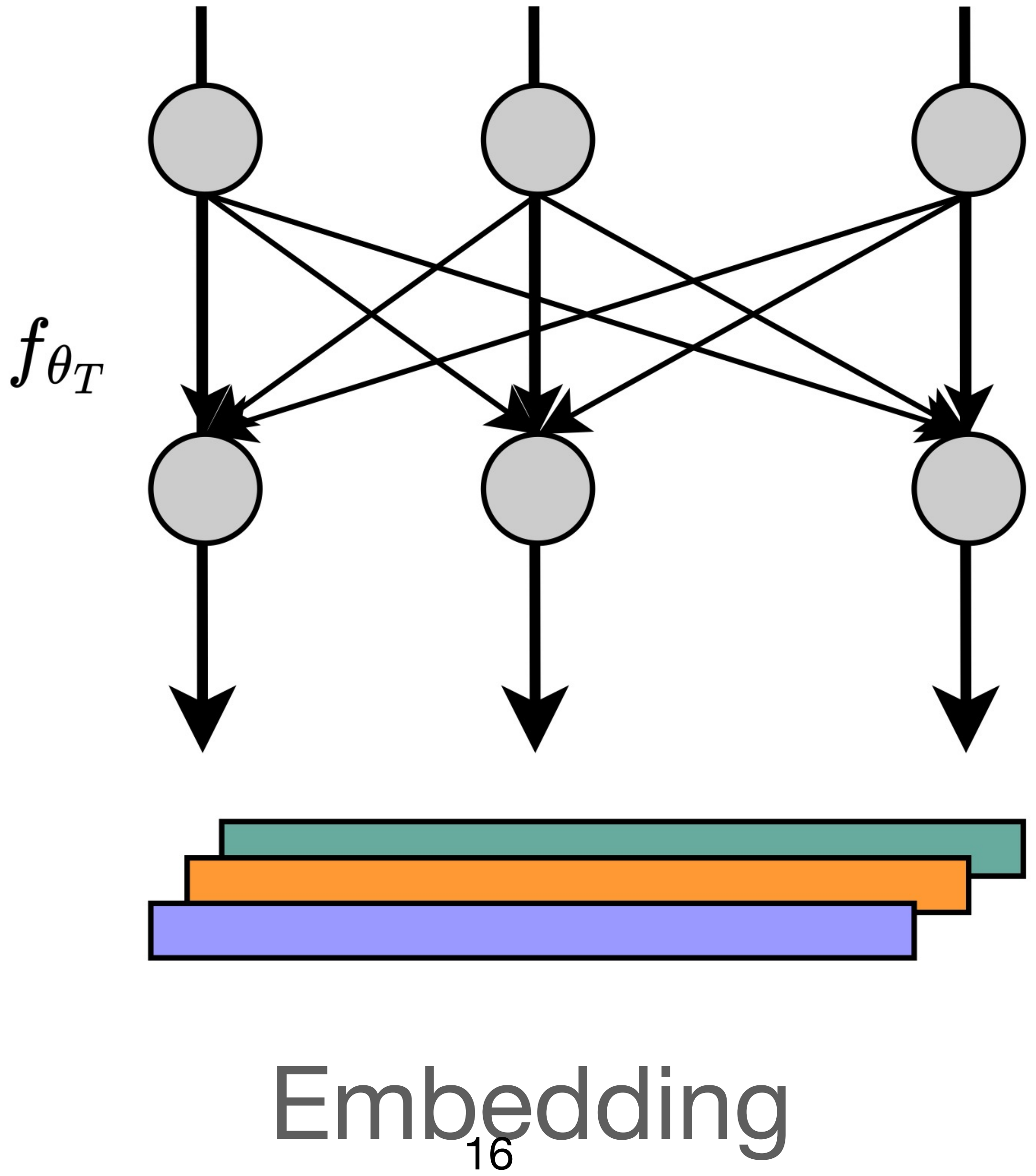
Integrated Gradients and Explainability

Baseline

Smoking Status	Consumes Alcohol	...	Physical Activity
0	0		0

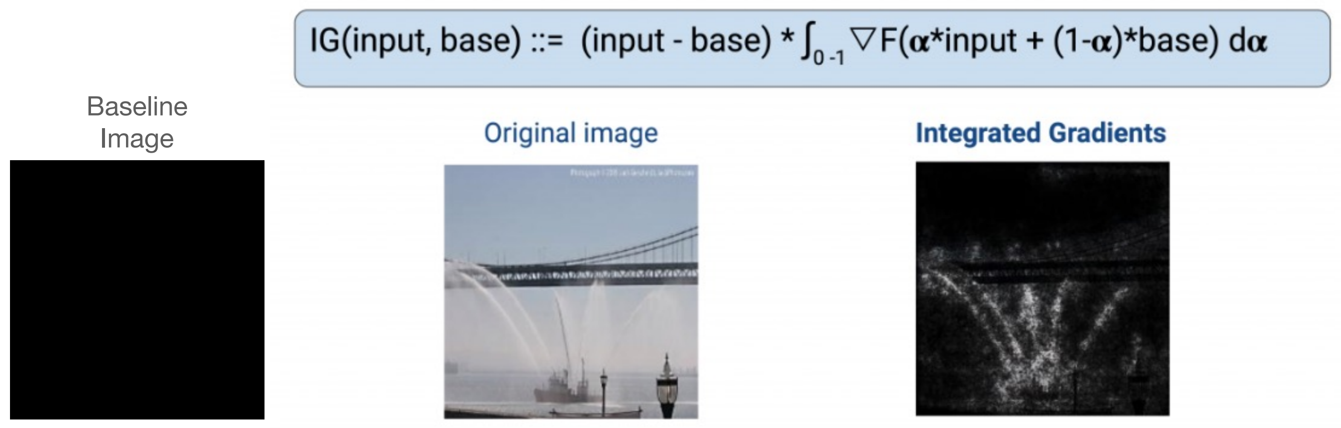
Original Tabular Entry

	Smoking Status	Consumes Alcohol	...	Physical Activity
S1	0	0		1
S2	1	1		1
S3	1	1		0

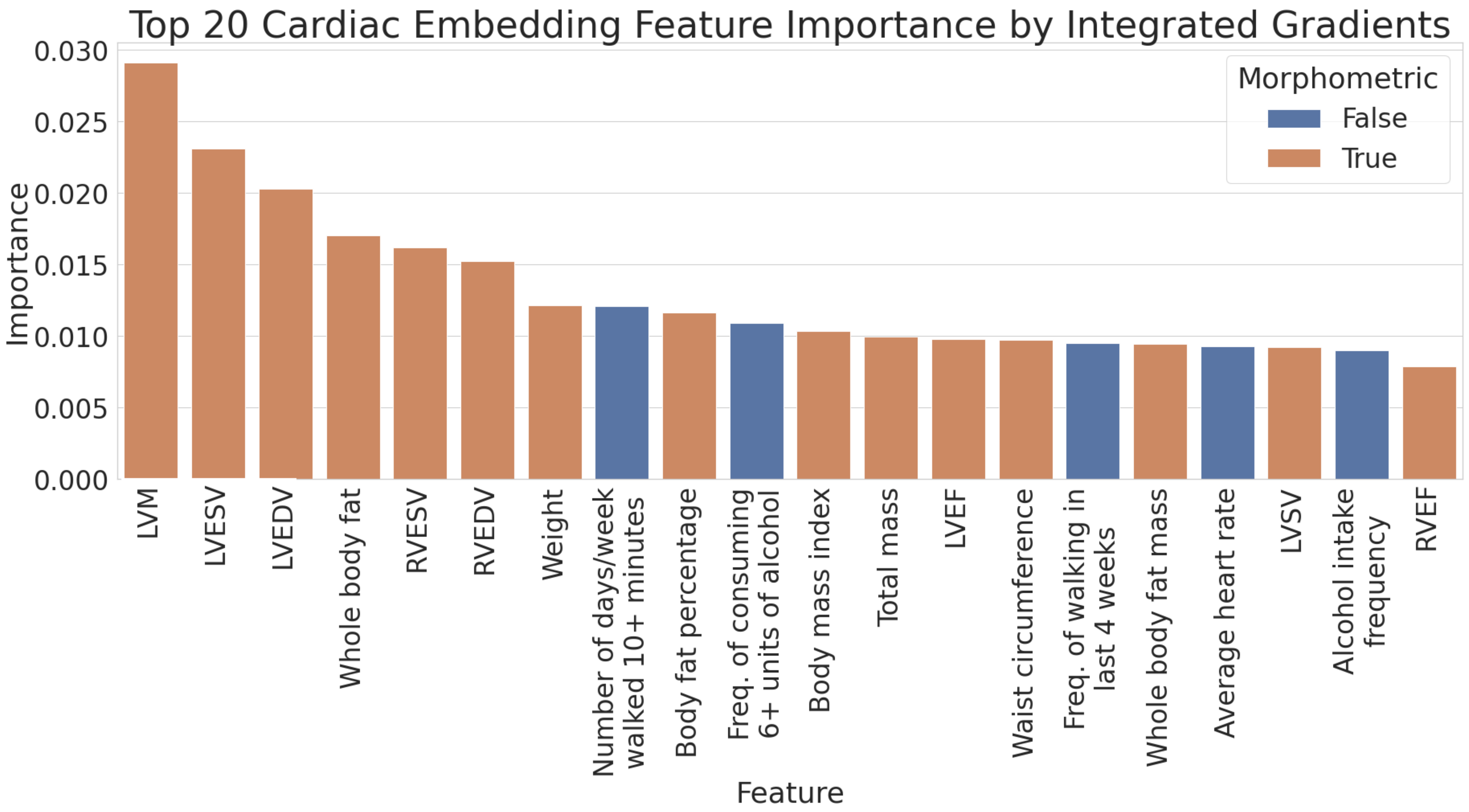


Integrated Gradients

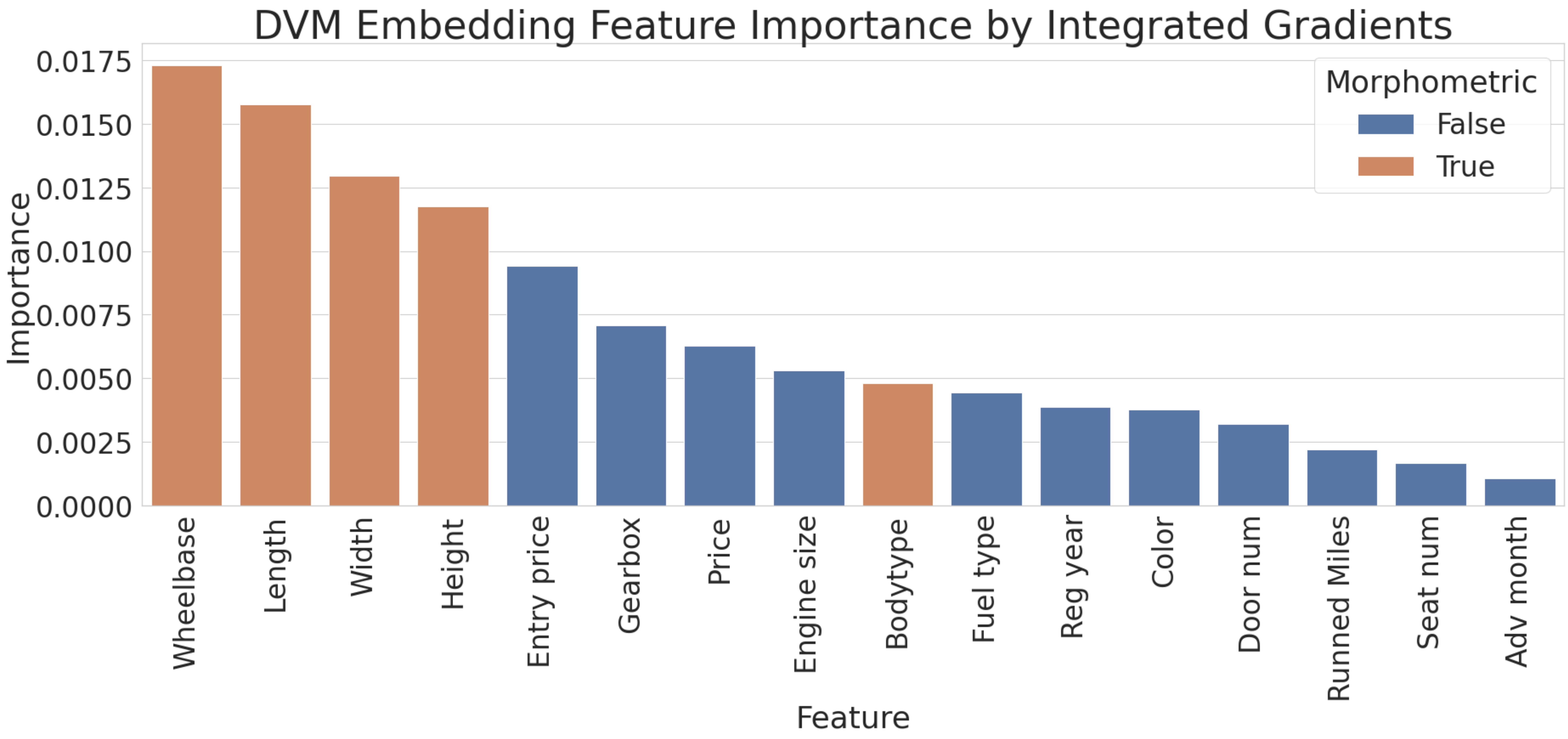
Smoking Status	Consumes Alcohol	...	Physical Activity
0.145	0.678		-0.365



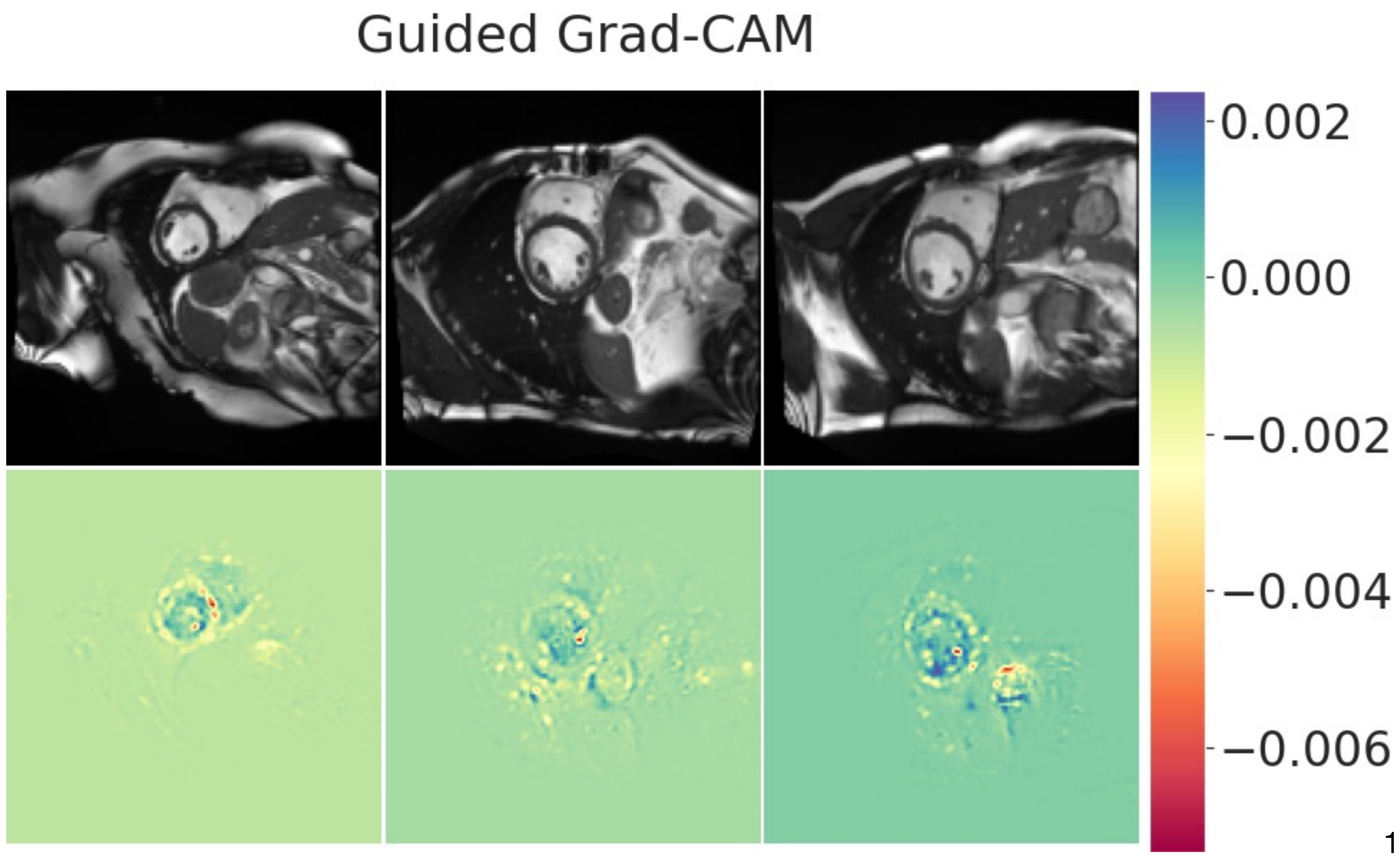
Morphometrics Features Improve Embedding Quality

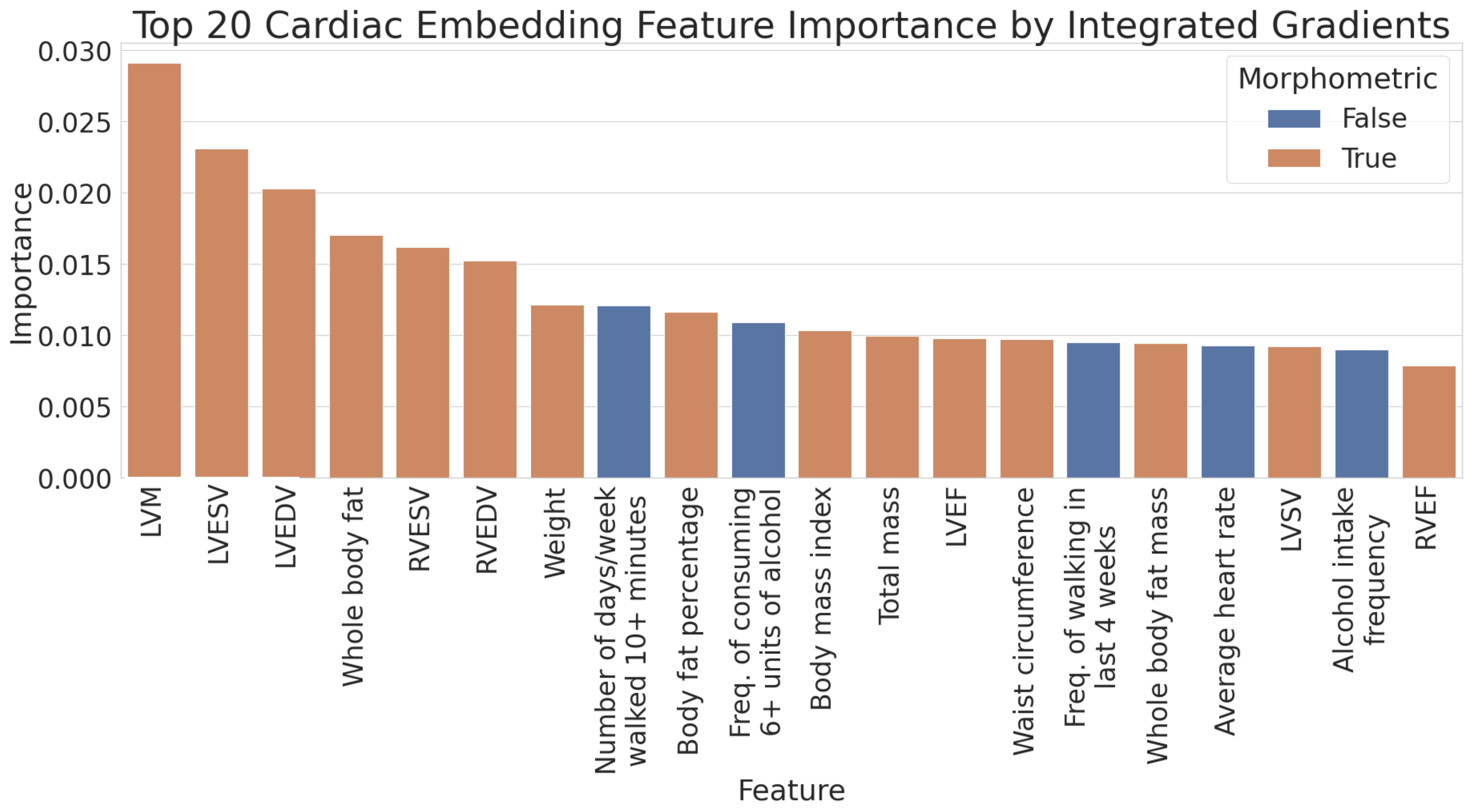


Morphometrics Features Improve Embedding Quality



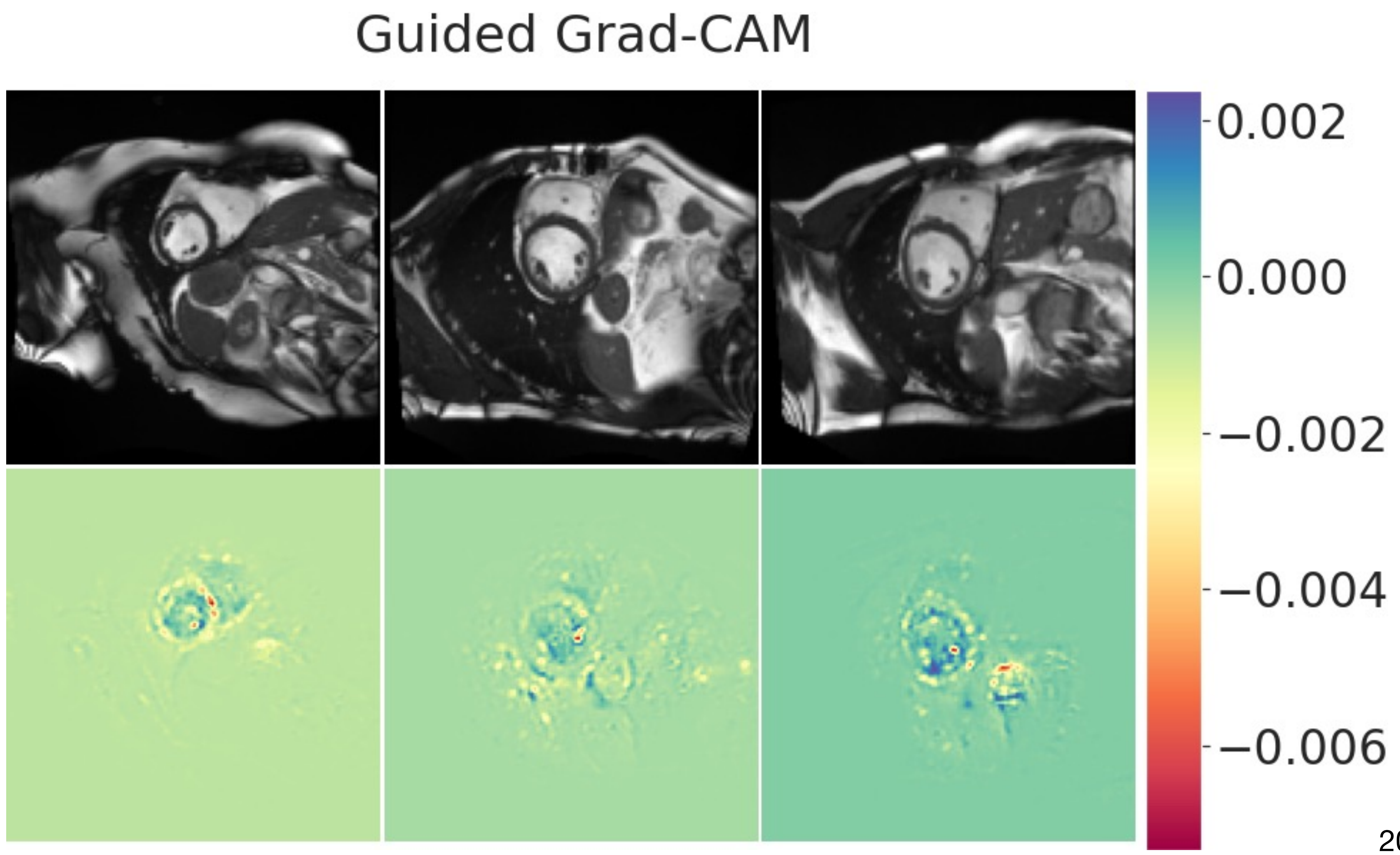
Morphometrics Features Improve Embedding Quality

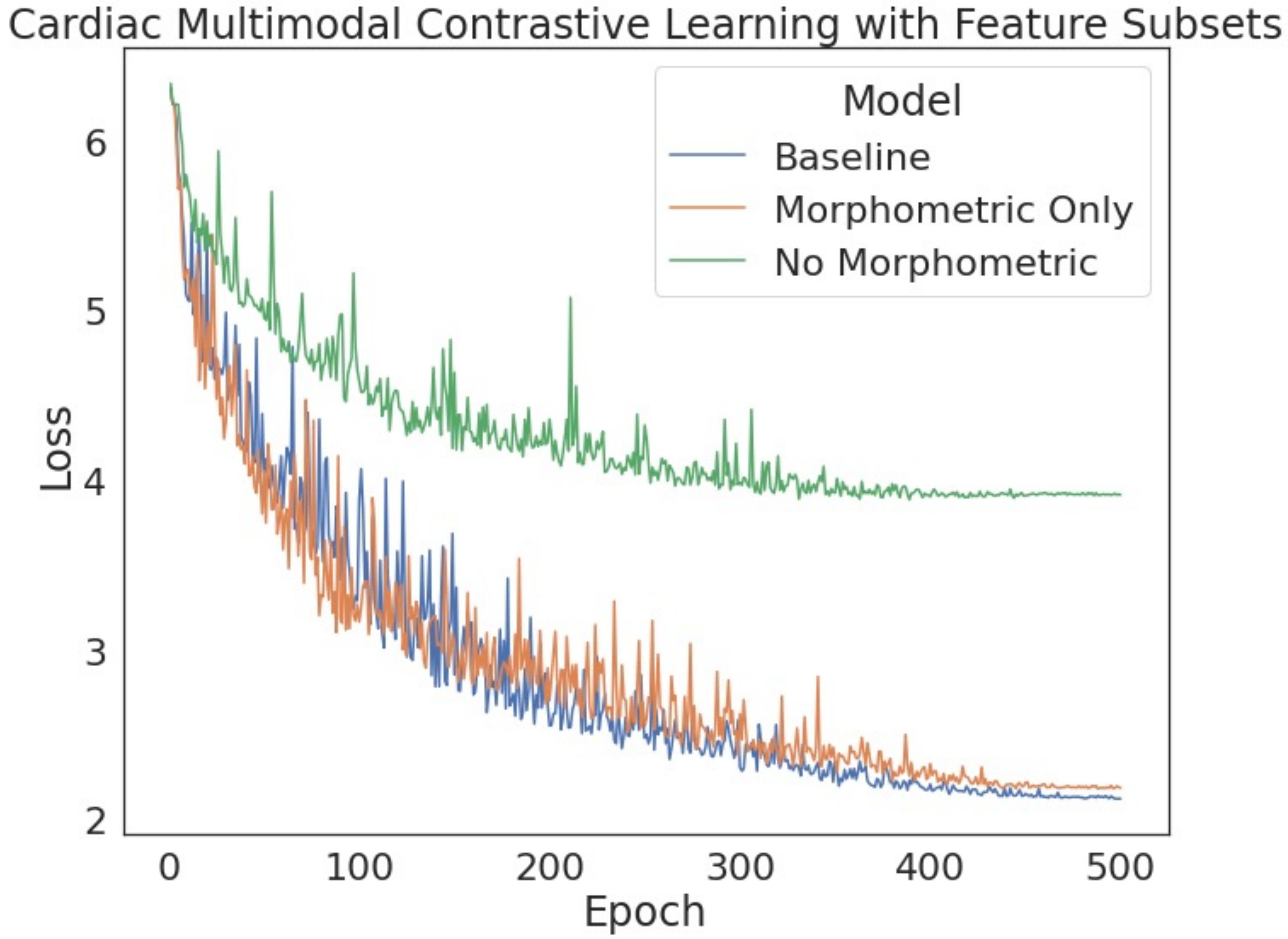




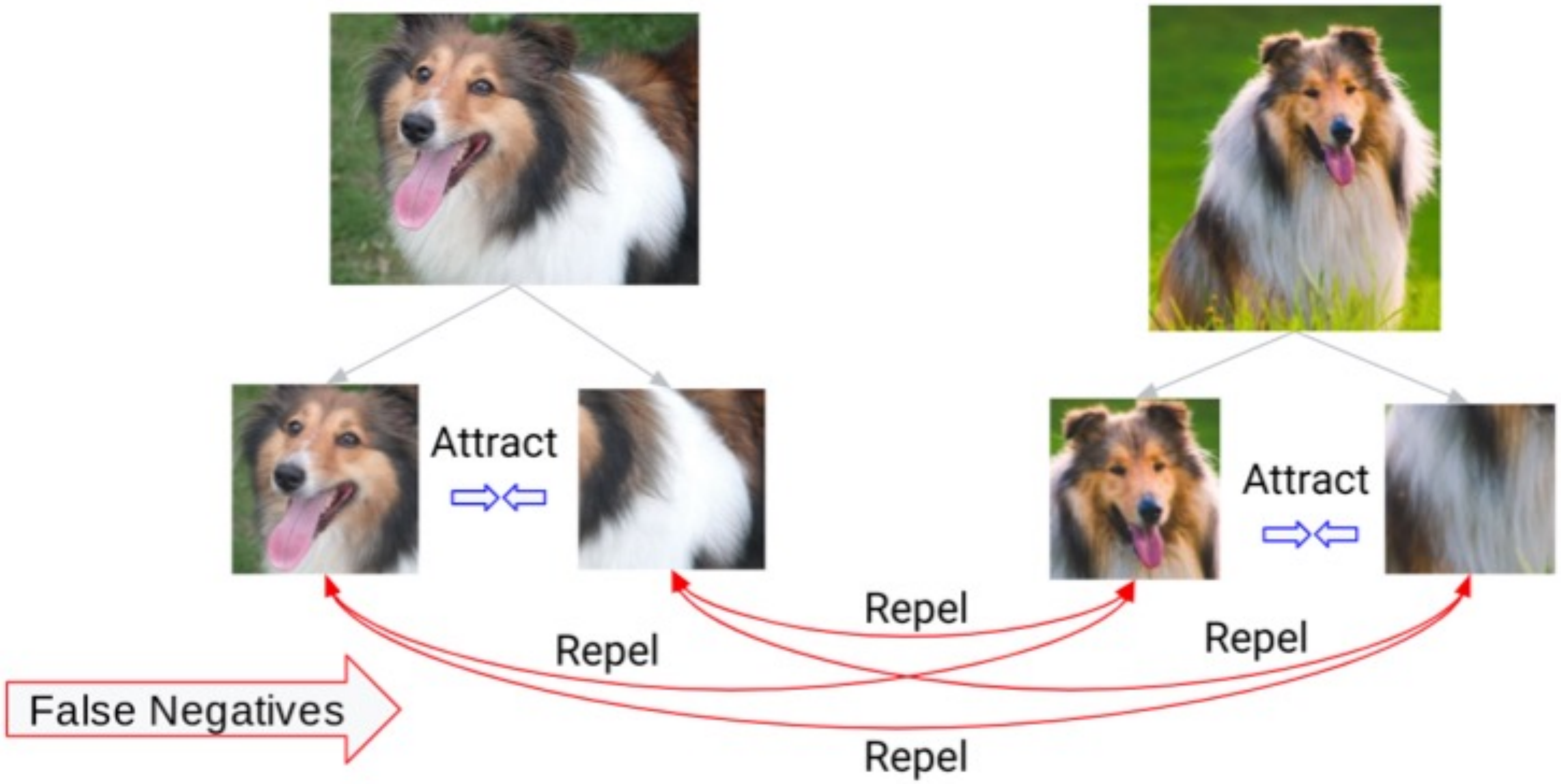
Morphometrics Features Improve Embedding Quality

Experiment	Tabular Features	Importance Percentage (%)	AUC (%) Infarction	AUC (%) CAD	Tabular Features	Importance Percentage (%)	Top-1 Accuracy (%) DVM
MM Imaging Baseline	117	100.0	**76.35±0.19**	**74.45±0.09**	16	100.0	91.43±0.13
Morphometric Features	24	47.0	75.22±0.30	73.71±0.09	5	56.4	**92.33±0.05**
Non-Morphometric Features	93	53.0	75.46±0.19	72.18±0.25	11	43.6	89.14±0.24



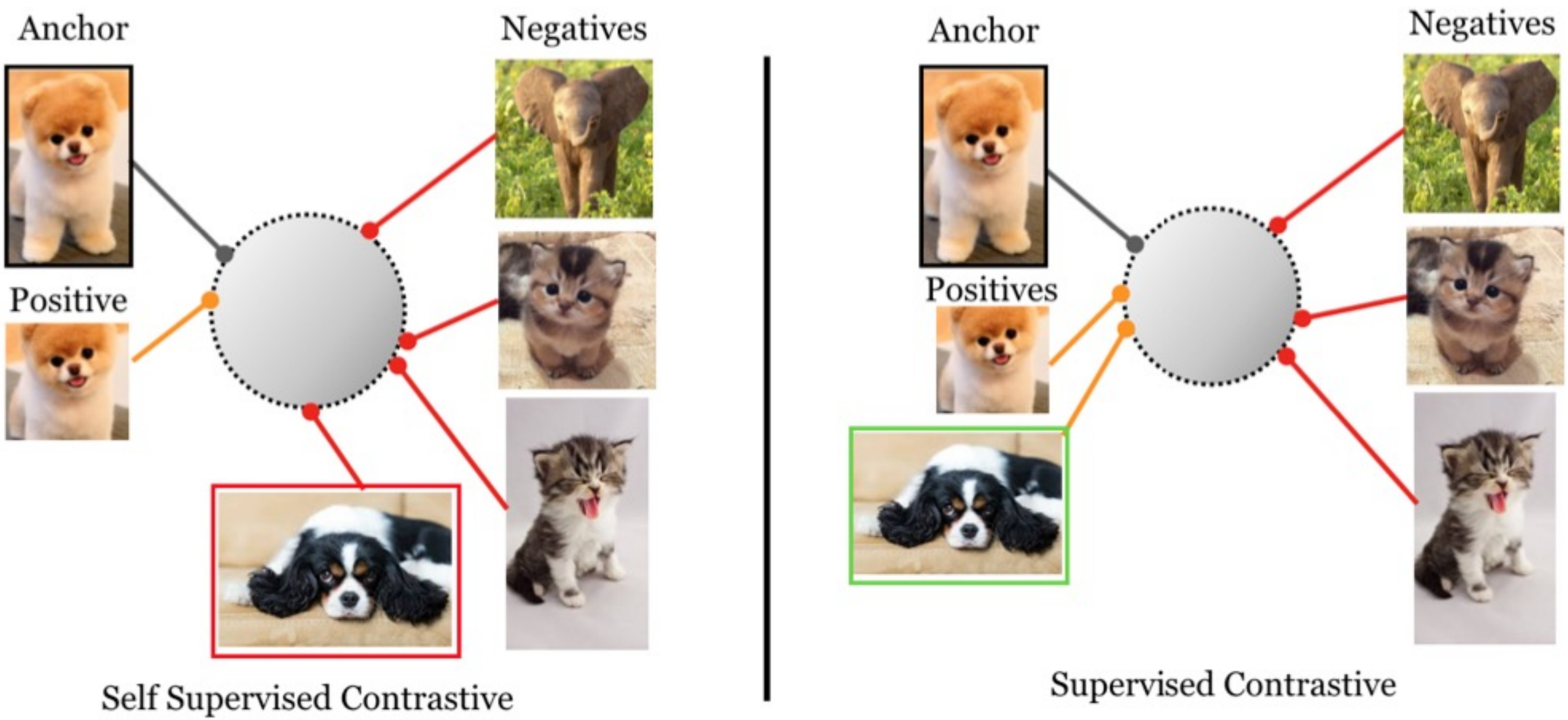


Supervised Contrastive Learning - FN Elimination



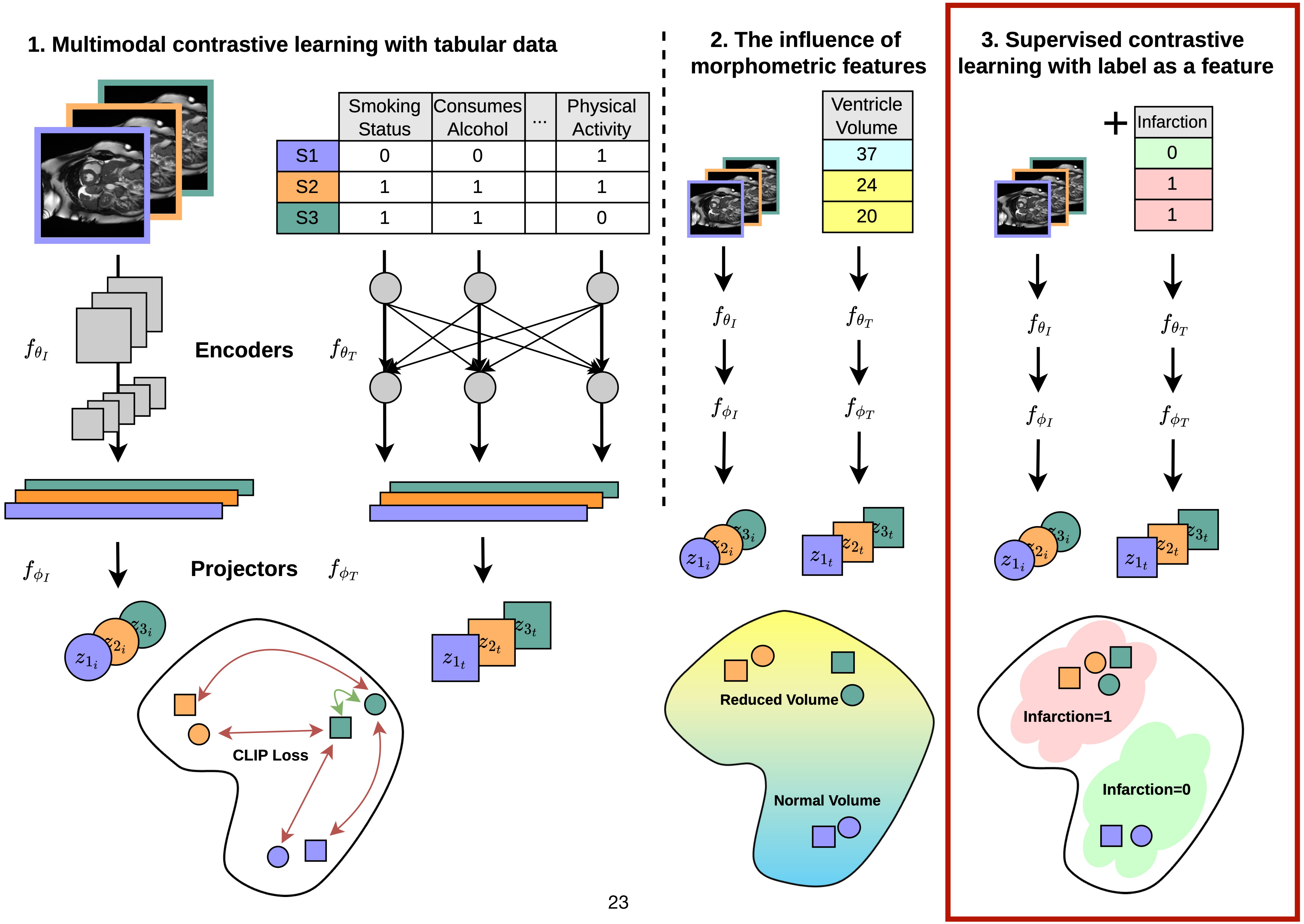
Huynh, Tri, et al. "Boosting contrastive self-supervised learning with false negative cancellation." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2022.

Supervised Contrastive Learning



Khosla, Prannay, et al. "Supervised contrastive learning." *Advances in neural information processing systems* 33 (2020): 18661-18673.

Appending the Label as a Tabular Feature



Appending the Label as a Tabular Feature

Contrastive	Label Used	Model	3% Positive AUC (%) Infarction	6% Positive AUC (%) CAD
✓		Multimodal Imaging Baseline	76.35 ± 0.19	**74.45 ± 0.09**
	✓	Supervised ResNet50	72.37 ± 1.80	68.84 ± 2.54
✓	✓	Label as a Feature (LaaF)	**76.60 ± 0.42**	73.76 ± 0.31
✓	✓	FN Elimination	75.38 ± 0.06	72.45 ± 0.09
✓	✓	FN Elimination + LaaF	75.30 ± 0.05	72.39 ± 0.08
✓	✓	SupCon	—	—
✓	✓	SupCon + LaaF	—	—

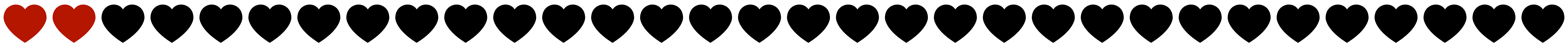
Appending the Label as a Tabular Feature

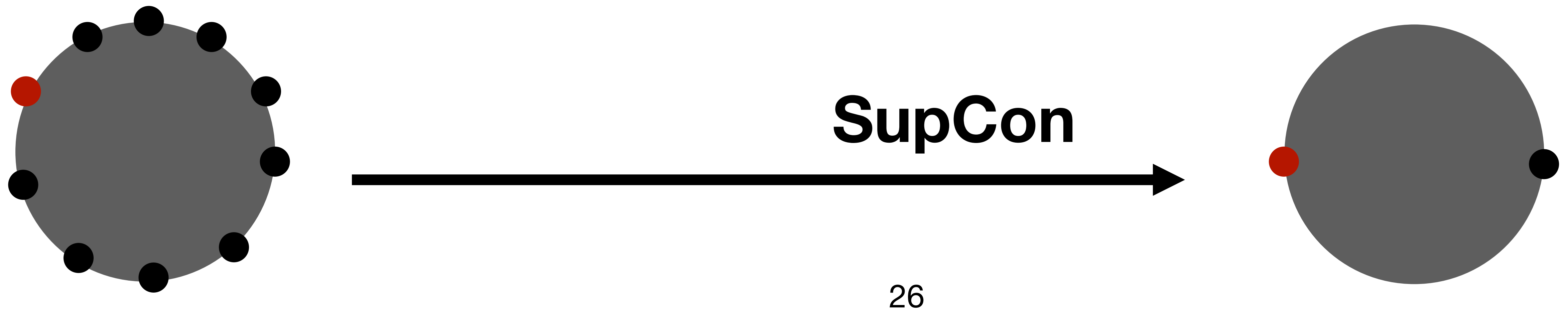
Contrastive	Label Used	Model	3% Positive AUC (%) Infarction	6% Positive AUC (%) CAD
✓		Multimodal Imaging Baseline	76.35±0.19	**74.45±0.09**
	✓	Supervised ResNet50	72.37±1.80	68.84±2.54
✓	✓	Label as a Feature (LaaF)	**76.60±0.42**	73.76±0.31
✓	✓	FN Elimination	75.38±0.06	72.45±0.09
✓	✓	FN Elimination + LaaF	75.30±0.05	72.39±0.08
✓	✓	SupCon	—	—
✓	✓	SupCon + LaaF	—	—

FN Elimination

Appending the Label as a Tabular Feature

Contrastive	Label Used	Model	AUC (%) Infarction (3% Positive)	AUC (%) CAD (6% Positive)
✓		Multimodal Imaging Baseline	76.35±0.19	**74.45±0.09**
	✓	Supervised ResNet50	72.37±1.80	68.84±2.54
✓	✓	Label as a Feature (LaaF)	**76.60±0.42**	73.76±0.31
✓	✓	FN Elimination	75.38±0.06	72.45±0.09
✓	✓	FN Elimination + LaaF	75.30±0.05	72.39±0.08
✓	✓	SupCon	—	—
✓	✓	SupCon + LaaF	—	—



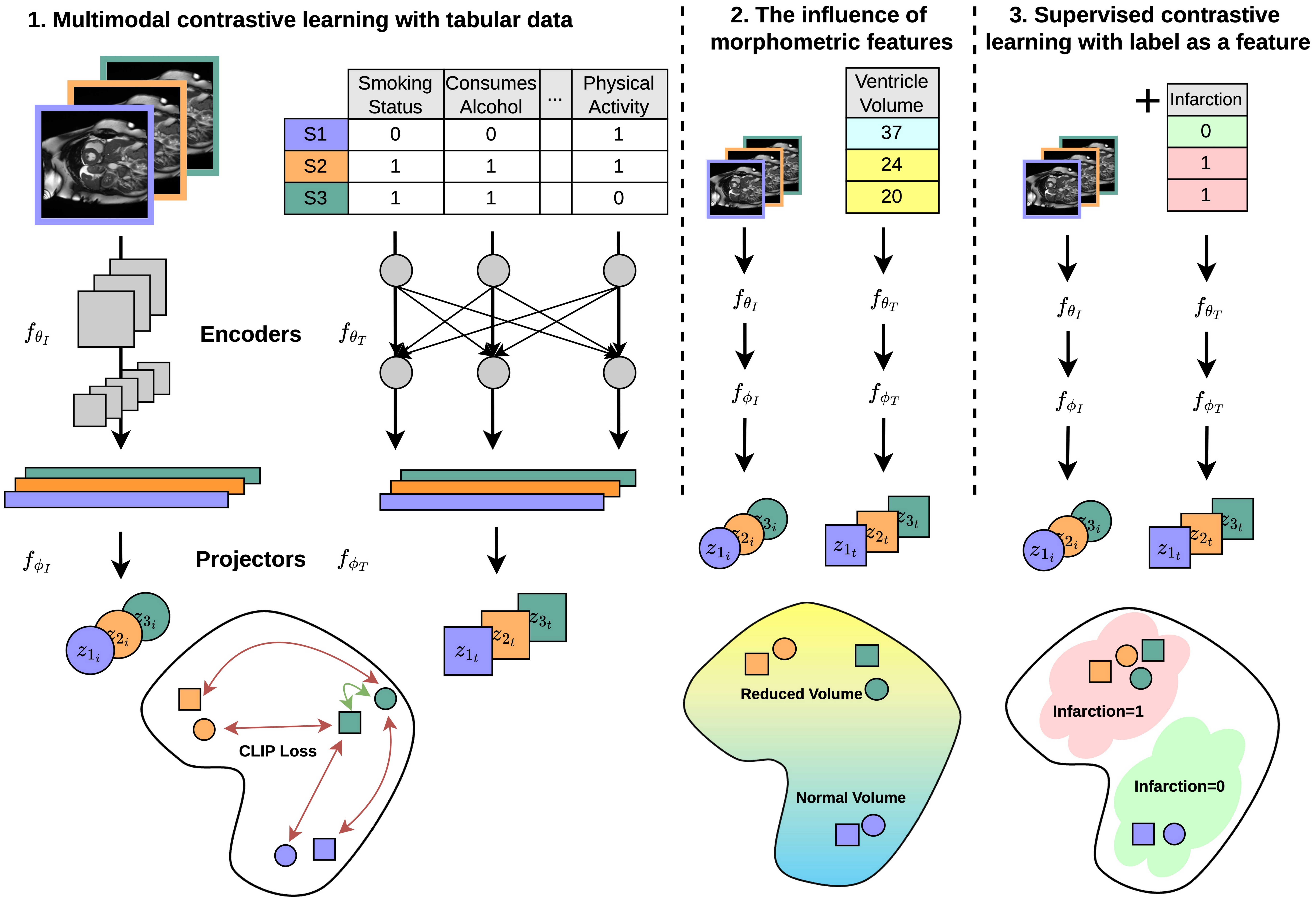


Appending the Label as a Tabular Feature

Contrastive	Label Used	Model	AUC (%) Infarction	AUC (%) CAD	Top-1 Accuracy (%) DVM
✓		Multimodal Imaging Baseline	76.35±0.19	**74.45±0.09**	91.43±0.13
	✓	Supervised ResNet50	72.37±1.80	68.84±2.54	87.97±2.20
✓	✓	Label as a Feature (LaaF)	**76.60±0.42**	73.76±0.31	93.56±0.08
✓	✓	FN Elimination	75.38±0.06	72.45±0.09	92.39±0.18
✓	✓	FN Elimination + LaaF	75.30±0.05	72.39±0.08	94.07±0.05
✓	✓	SupCon	—	—	93.82±0.11
✓	✓	SupCon + LaaF	—	—	**94.40±0.04**

Appending the Label as a Tabular Feature

Model	Top-1 Acc. (%) DVM (100%)	Top-1 Acc. (%) DVM (10%)	Top-1 Acc. (%) DVM (1%)
Multimodal Baseline	91.43±0.13	86.30±0.08	60.18±0.21
Supervised ResNet50	87.97±2.20	30.69±14.02	2.84±0.00
Label-as-a-Feature (LaaF)	93.56±0.08	89.87±0.03	**67.50±0.10**
FN Elim.	92.39±0.18	87.61±0.07	63.95±0.14
FN Elim. + LaaF	<u>94.07±0.05</u>	<u>89.99±0.05</u>	63.37±0.70
SupCon	93.82±0.11	89.75±0.08	63.29±0.33
SupCon + LaaF	**94.40±0.04**	**90.37±0.05**	<u>64.01±0.77</u>



1. Multimodal contrastive learning with tabular data
Smoking Status
Consumes Alcohol
...
Physical Activity
S1
0
0
1
S2
1
1
1
S3
1
1
0
f_{θ_I}
Encoders
f_{θ_T}
f_{ϕ_I}
Projectors
f_{ϕ_T}
z_{1_i} z_{2_i} z_{3_i}
z_{1_t} z_{2_t} z_{3_t}
CLIP Loss
2. The influence of morphometric features
Ventricle Volume
37
24
20
Reduced Volume
Normal Volume
3. Supervised contrastive learning with label as a feature
+
Infarction
0
1
1
Infarction=1
Infarction=0