PaCX-MAE: Physiology-Augmented Chest X-Ray Masked Autoencoder

Yancheng Liu[1], Kenichi Maeda[1], Manan Pancholy[1]
Brown University

BROWN Computer Science

ICML International Conference On Machine Learning

Motivation

- Chest X-ray diagnosis often benefits from physiological context provided by ECGs and laboratory tests.
- Multimodal models can leverage physiological signals during training, but often assume that all modalities are available at inference.
- In practice, ECGs or laboratory measurements may be unavailable or delayed especially in acute clinical settings.

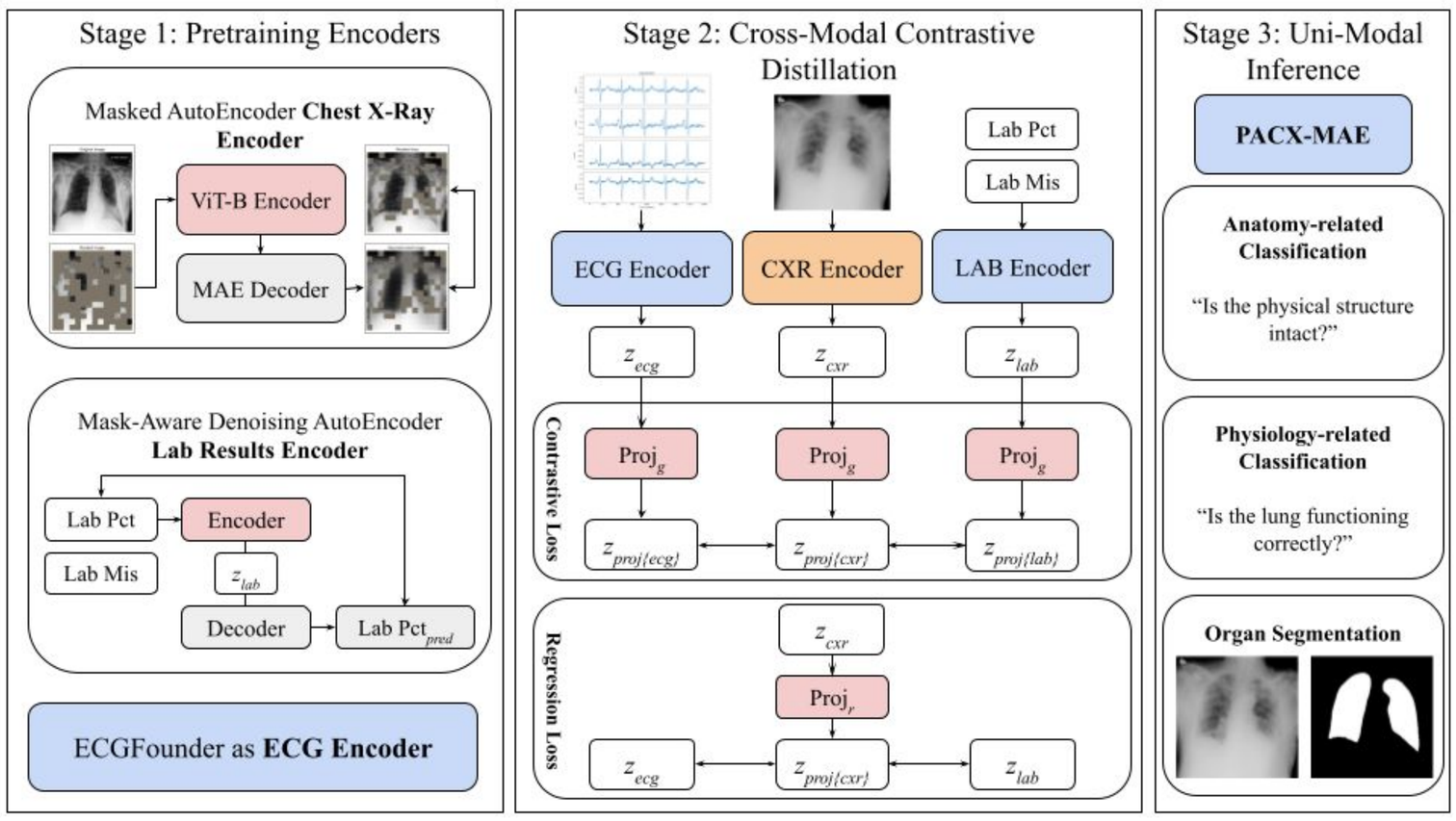
Can a CXR encoder learn physiological context from paired multimodal data while remaining image-only at deployment?

Our Contributions

We introduce **PaCX-MAE** (**P**hysiology **A**ugmented **C**hest **X**-Ray **M**asked **A**uto**e**ncoder), a framework that transfers physiological priors from ECG and laboratory data into a chest X-ray encoder while requiring only CXR input at inference.

We validate PaCX-MAE across **nine benchmarks**, showing that it:
- Outperforms standard unimodal MAE, particularly on physiology-rich tasks
- Preserves pixel-level fidelity for downstream segmentation
- Significantly improves label efficiency in low-data regimes

Method



Optimization status: red (trainable), orange (LoRA-adapted), and blue (frozen).

Stage 1: Unimodal Pretraining

CXR Encoder: ViT-B pretrained with MAE
Laboratory Encoder: Mask-aware denoising autoencoder
ECG Encoder: ECGFounder pretrained on 10M ECGs

Stage 2: Cross-Modal Distillation

$\mathcal{L}_{total} = \lambda_C \mathcal{L}_{contrastive} + \lambda_R \mathcal{L}_{regression}$

Contrastive Alignment: Align 3 modalities in a shared latent space using symmetric InfoNCE.
Latent Regression: Predict physiological embeddings from CXR features using cosine distance.

Evaluation

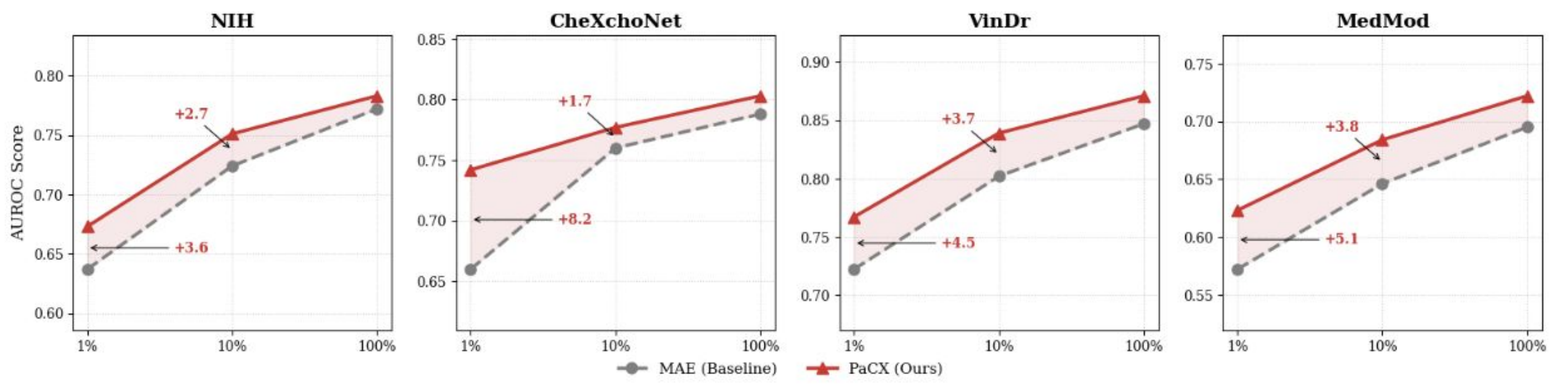
Clinical Transfer

Dataset	Metric	ImageNet	MAE	PaCX	Δ (pp)
TB[1]	AUROC	0.887	0.899	**0.910**	+1.1
	F1	0.818	0.814	**0.846**	+3.2
CheXchoNet[1]	AUROC	0.728	0.788	**0.803**	+1.5
	F1	0.147	0.215	**0.266**	+5.1
ChestX6[2]	AUROC	0.983	0.988	**0.989**	+0.1
	F1	0.876	0.905	**0.906**	+0.1
VinDr-CXR[3]	AUROC	0.751	0.847	**0.871**	+2.4
	F1	0.097	0.191	**0.256**	+6.5
NIH-14[3]	AUROC	0.721	0.772	**0.783**	+1.1
	F1	0.048	0.113	**0.115**	+0.2
MedMod[3]	AUROC	0.612	0.695	**0.722**	+2.7
	F1	0.091	0.231	**0.253**	+2.2
COVID-QU-Ex[4]	IoU	0.894	**0.942**	**0.942**	0.0
	Dice	0.943	**0.970**	**0.970**	0.0
QaTa-COV19[4]	IoU	0.622	**0.726**	0.715	–1.1
	Dice	0.766	**0.841**	0.833	–0.8
CXL-Seg[4]	IoU	0.984	**0.996**	**0.996**	0.0
	Dice	0.992	**0.998**	**0.998**	0.0

[1] binary, [2] multiclass, [3] multilabel, [4] segmentation.

PaCX retains the strong anatomical representation learned by MAE, while **improving classification on physiology**-dense benchmarks.

Data Efficiency

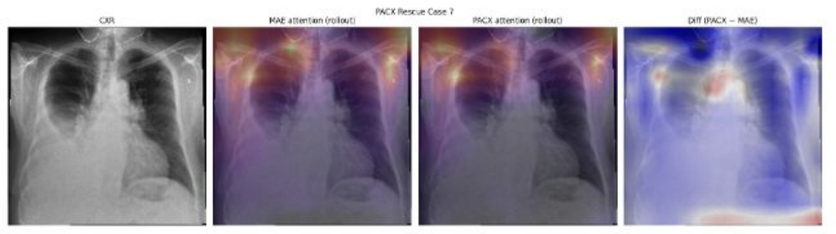


Zero-shot Alignment

Metric	ECG Targets ImNet	ECG Targets MAE	ECG Targets PaCX	Lab Targets ImNet	Lab Targets MAE	Lab Targets PaCX
Cos Sim	0.143	0.204	**0.229**	0.187	0.239	**0.252**
R@5	1.51%	5.17%	**5.60%**	1.51%	3.66%	**3.66%**

PaCX **successfully transfers** ECG and laboratory structure into the visual representation.

Attention Shift



PaCX shifts attention from body structures **toward clinically relevant soft-tissue** regions.

Component Analysis

Dataset	Metric	Modality Ablation ECG	Lab	PaCX	Loss Ablation Cont	Reg	PaCX
CheXchoNet	AUC	0.801	0.795	**0.803**	0.799	0.789	**0.803**
	F1	**0.296**	0.275	0.266	**0.273**	0.227	0.266
MedMod	AUC	0.717	0.721	**0.722**	**0.722**	0.673	**0.722**
	F1	0.243	0.245	**0.253**	**0.258**	0.131	0.253
VinDr	AUC	0.871	**0.875**	0.871	0.866	0.843	**0.871**
	F1	0.233	0.248	**0.256**	0.241	0.130	**0.256**

The **full PaCX** configuration yields the **most consistent** performance across tasks.

Scan here to see the full paper: