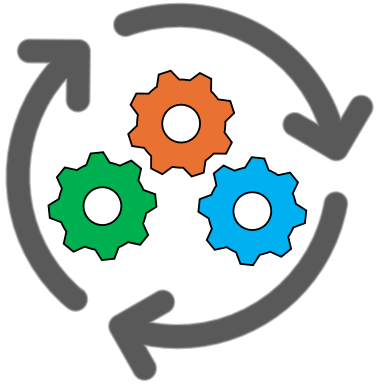


WACV 2025
TUCSON, ARIZONA • FEB 28 - MAR 4

Foundation X : Integrating Classification, Localization, and Segmentation through Lock-Release Pretraining Strategy for Chest X-ray Analysis

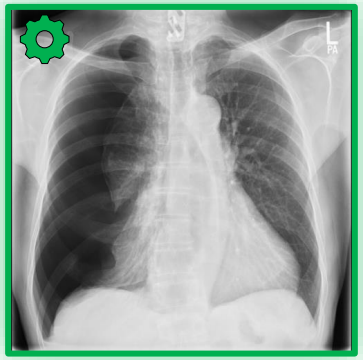
Nahid Ul Islam[1], DongAo Ma[1], Jiaxuan Pang[1], Shivasakthi Senthil Velan[1]
Michael B. Gotway[2], and Jianming Liang[1]

[1]Arizona State University [2]Mayo Clinic

Feng, S., Azzollini, D., Kim, J. S., Jin, C. K., Kim, E., Gordon, S., ... & Wilson, B. (2024). CANDID-PTX.

Feng, S., Azzollini, D., Kim, J. S., Jin, C. K., Kim, E., Gordon, S., ... & Wilson, B. (2024). CANDID-PTX.

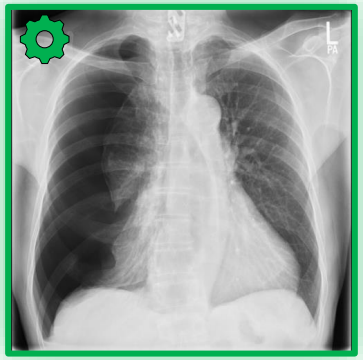
Classification Task



Pneumothorax: 1

Feng, S., Azzollini, D., Kim, J. S., Jin, C. K., Kim, E., Gordon, S., ... & Wilson, B. (2024). CANDID-PTX.

Classification Task



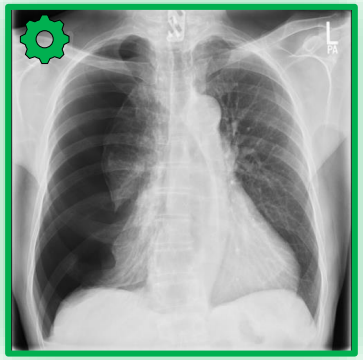
Pneumothorax: 1

Localization Task

Bounding Box

Feng, S., Azzollini, D., Kim, J. S., Jin, C. K., Kim, E., Gordon, S., ... & Wilson, B. (2024). CANDID-PTX.

Classification Task

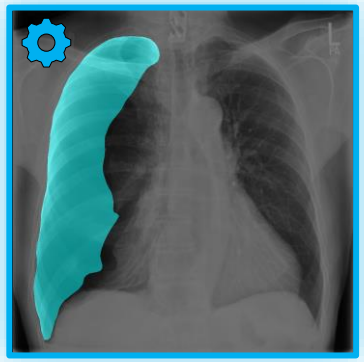


Pneumothorax: 1

Localization Task

Bounding Box

Segmentation Task



Mask

Feng, S., Azzollini, D., Kim, J. S., Jin, C. K., Kim, E., Gordon, S., ... & Wilson, B. (2024). CANDID-PTX.

Dataset Overview with Classification, Localization, and Segmentation Heads

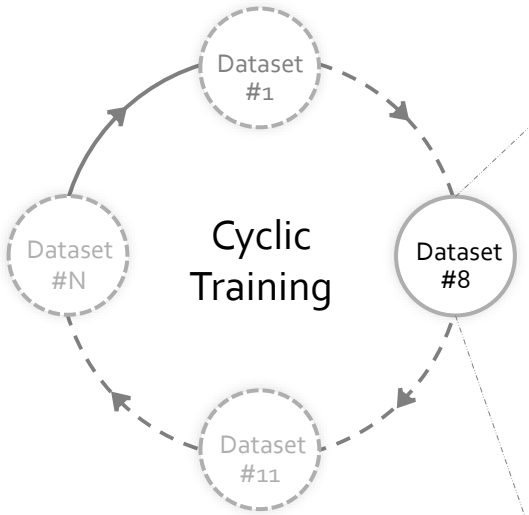
Dataset	Cls. Head	Loc. Head	Seg. Head
1. CheXpert	C_1	-	-
2. NIH ChestX-ray14	C_2	-	-
3. VinDr-CXR	C_3	-	-
4. NIH Shenzhen CXR	C_4	-	-
5. MIMIC-II	C_5	-	-
6. TBX11k	C_6	L_1	-
7. NODE21	C_7	L_2	-
8. CANDID-PTX	C_8	L_3	S_1
9. RSNA Pneumonia	C_9	L_4	-
10. ChestX-Det	C_{10}	L_5	S_2
11. SIIM-ACR	C_{11}	L_6	S_3

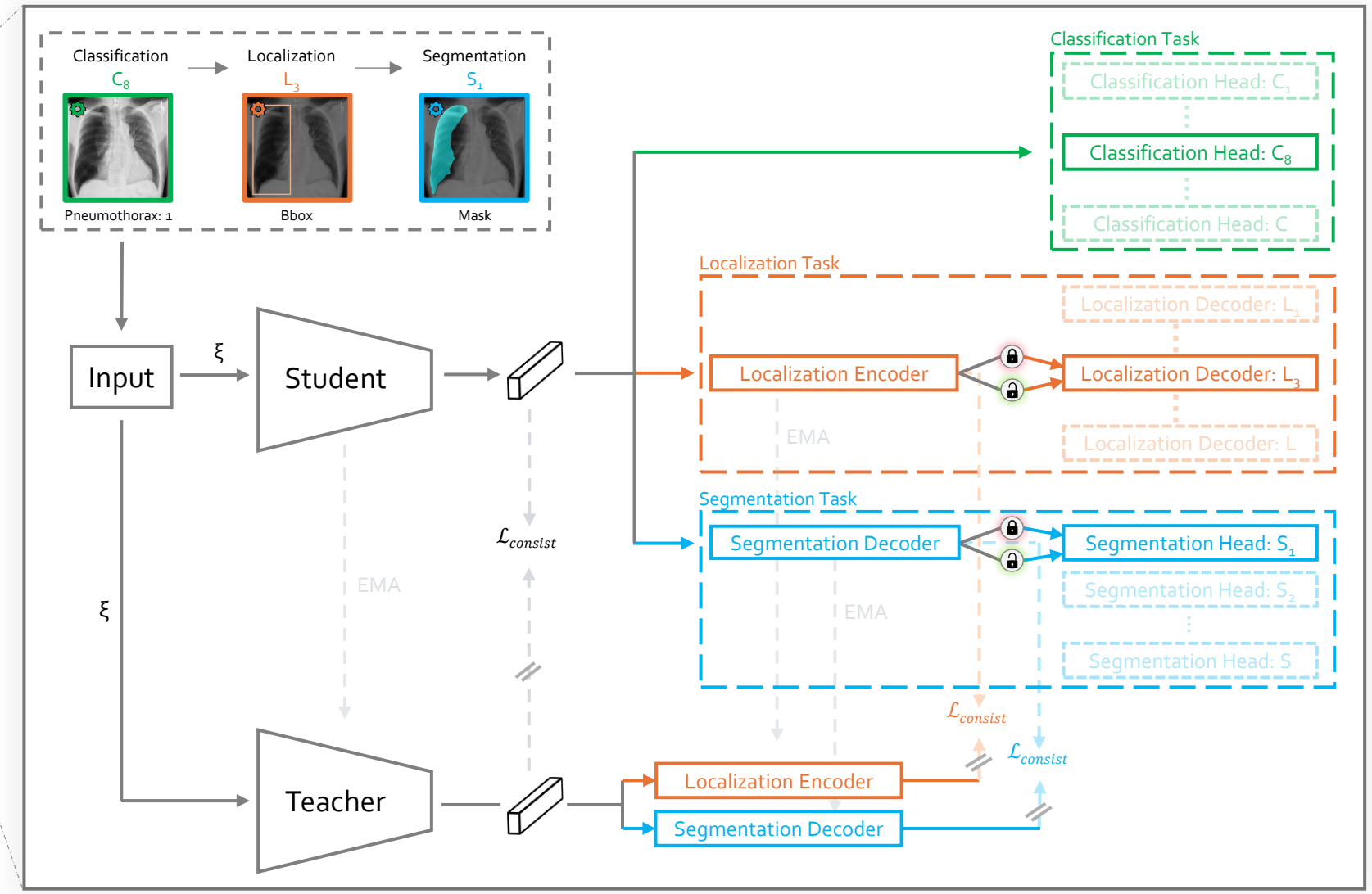
Dataset Overview with Classification, Localization, and Segmentation Heads

Dataset	Cls. Head	Loc. Head	Seg. Head
1. CheXpert	C_1	-	-
2. NIH ChestX-ray14	C_2	-	-
3. VinDr-CXR	C_3	-	-
4. NIH Shenzhen CXR	C_4	-	-
5. MIMIC-II	C_5	-	-
6. TBX11k	C_6	L_1	-
7. NODE21	C_7	L_2	-
8. CANDID-PTX	C_8	L_3	S_1
9. RSNA Pneumonia	C_9	L_4	-
10. ChestX-Det	C_{10}	L_5	S_2
11. SIIM-ACR	C_{11}	L_6	S_3

Foundation X

Example: 8. CANDID-PTX





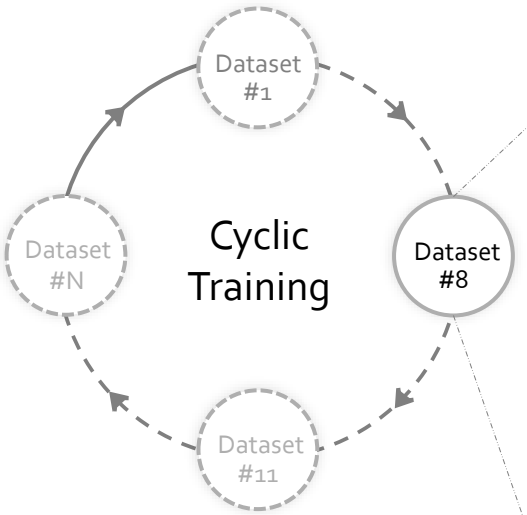
- Stop gradient
- Data augmentation
- Classification branch
- Localization branch
- Segmentation branch
- During training, all left-side components are **locked** (frozen)
- During training, all left-side components are **released** (trainable)

Ma, D., Pang, J., Gotway, M. B., & Liang, J. (2023). Foundation Ark: Accruing and Reusing Knowledge for Superior and Robust Performance. *MICCAI, 651–662*. Springer.

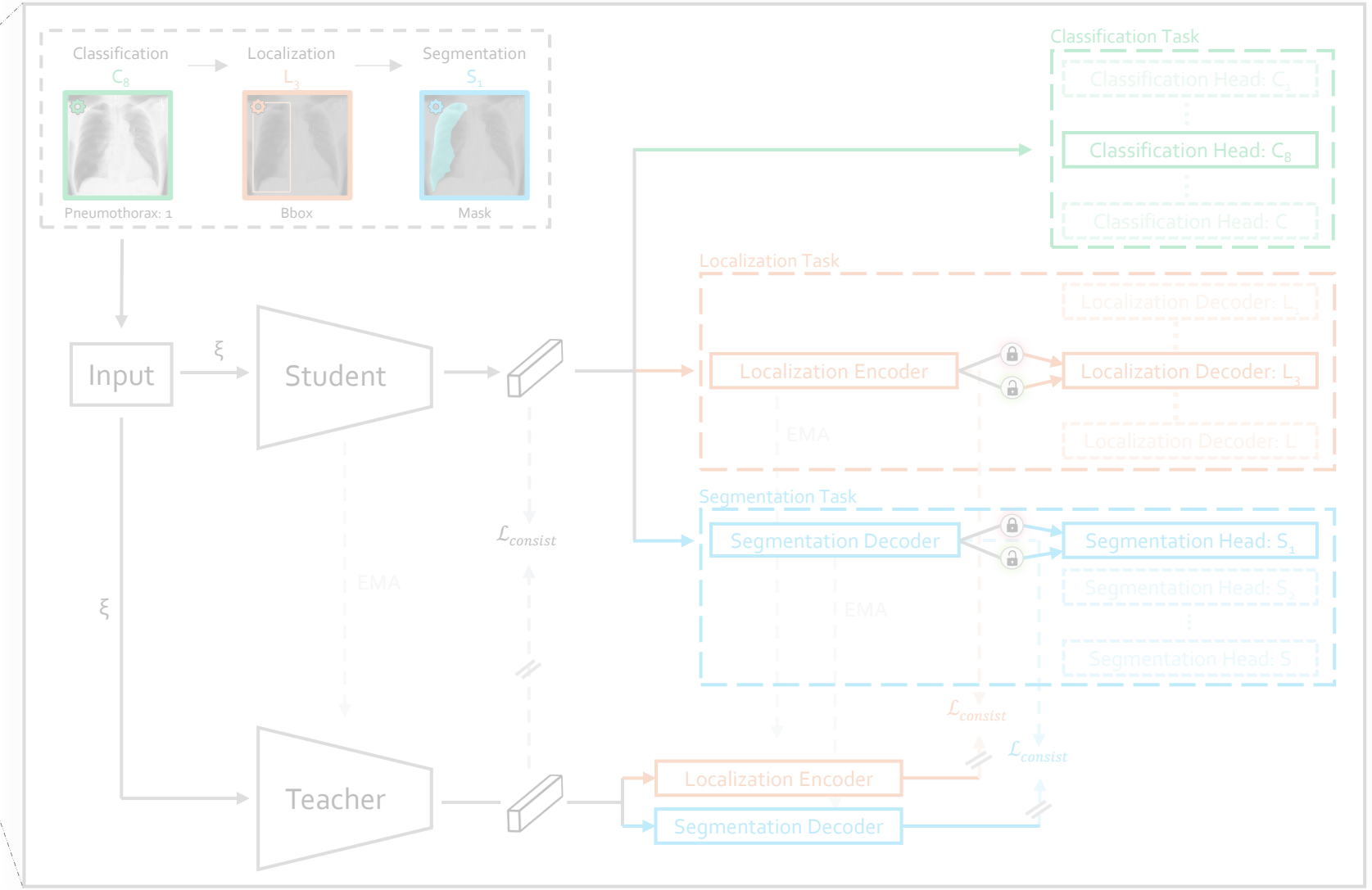
Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L. M., & Shum, H. Y. (2022). DINO: DETR with improved denoising anchor boxes for end-to-end object detection.

Xiao, T., Liu, Y., Zhou, B., Jiang, Y., & Sun, J. (2018). Unified perceptual parsing for scene understanding. *ECCV*, 418–434.

Foundation X



Example: 8. CANDID-PTX



Stop gradient
Data augmentation
Classification branch
Localization branch
Segmentation branch
During training, all left-side components are **locked** (frozen)
During training, all left-side components are **released** (trainable)

Ma, D., Pang, J., Gotway, M. B., & Liang, J. (2023). Foundation Ark: Accruing and Reusing Knowledge for Superior and Robust Performance. *MICCAI, 651–662*. Springer.

Foundation X

Lock-Release Pretraining Strategy

	Epoch #	Data Size	Backbone	Cls.Head	Loc.Enc	Loc.Dec	Seg.Dec	Seg.Head	Mode	Training Task
Cycle 1	1	Full	Trainable	Trainable	-	-	-	-	Release	Classification
	2	Half	Frozen	-	Frozen	Trainable	-	-	Lock	Localization
	3	Full	Trainable	-	Trainable	Trainable	-	-	Release	Localization
	4	Half	Frozen	-	-	-	Frozen	Trainable	Lock	Segmentation
	5	Full	Trainable	-	-	-	Trainable	Trainable	Release	Segmentation

Foundation X

Lock-Release Pretraining Strategy

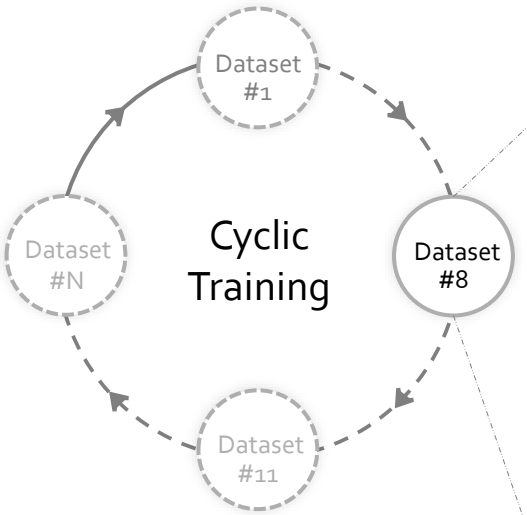
	Epoch #	Data Size	Backbone	Cls.Head	Loc.Enc	Loc.Dec	Seg.Dec	Seg.Head	Mode	Training Task
Cycle 1	1	Full	Trainable	Trainable	-	-	-	-	Release	Classification
	2	**Half**	**Frozen**	-	**Frozen**	**Trainable**	-	-	**Lock**	Localization
	3	Full	Trainable	-	Trainable	Trainable	-	-	Release	Localization
	4	**Half**	**Frozen**	-	-	-	**Frozen**	**Trainable**	**Lock**	Segmentation
	5	Full	Trainable	-	-	-	Trainable	Trainable	Release	Segmentation

Foundation X

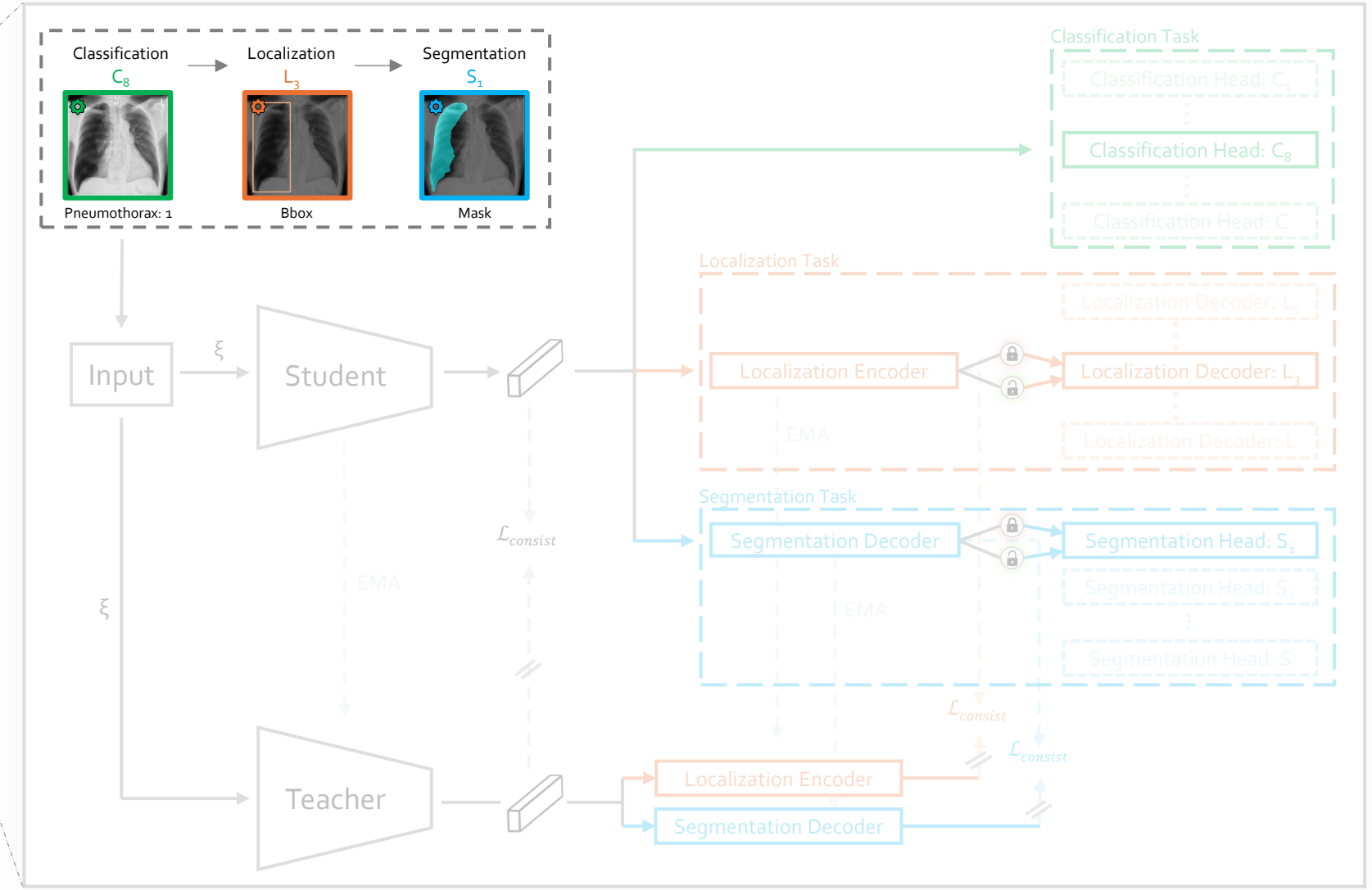
Lock-Release Pretraining Strategy

	Epoch #	Data Size	Backbone	Cls.Head	Loc.Enc	Loc.Dec	Seg.Dec	Seg.Head	Mode	Training Task
Cycle 1	1	**Full**	**Trainable**	**Trainable**	-	-	-	-	**Release**	**Classification**
	2	Half	Frozen	-	Frozen	Trainable	-	-	Lock	**Localization**
	3	**Full**	**Trainable**	-	**Trainable**	**Trainable**	-	-	**Release**	**Localization**
	4	Half	Frozen	-	-	-	Frozen	Trainable	Lock	**Segmentation**
	5	**Full**	**Trainable**	-	-	-	**Trainable**	**Trainable**	**Release**	**Segmentation**

Foundation X

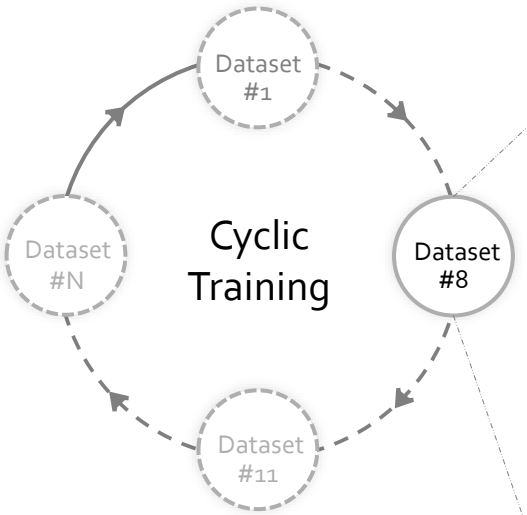


Example: 8. CANDID-PTX

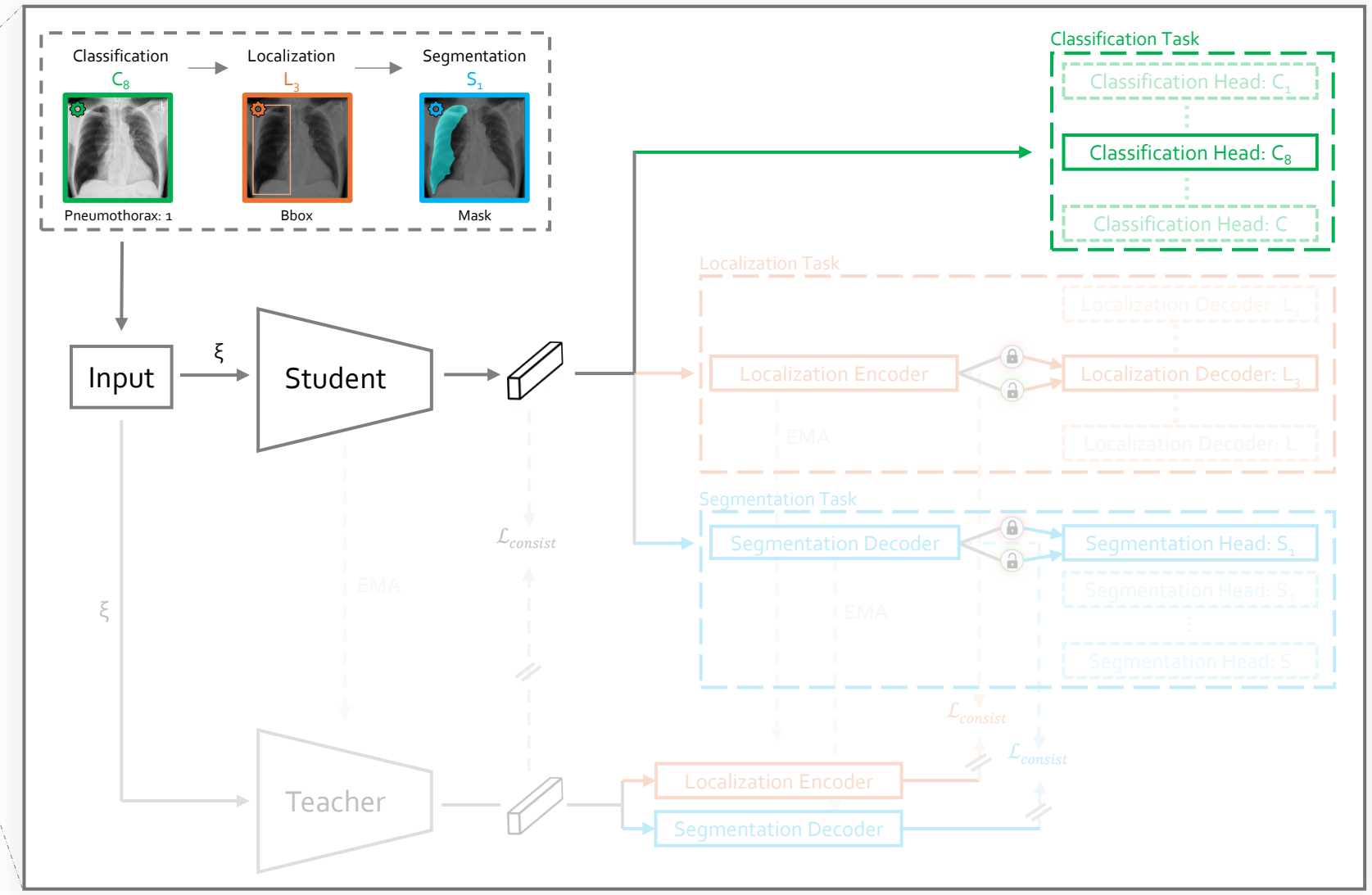


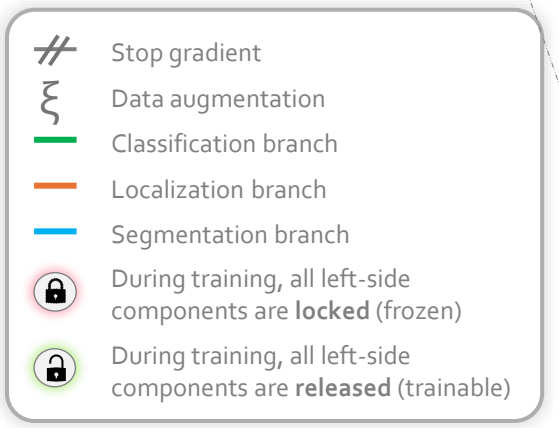
Feng, S., Azzollini, D., Kim, J. S., Jin, C. K., Kim, E., Gordon, S., ... & Wilson, B. (2024). CANDID-PTX.

Foundation X

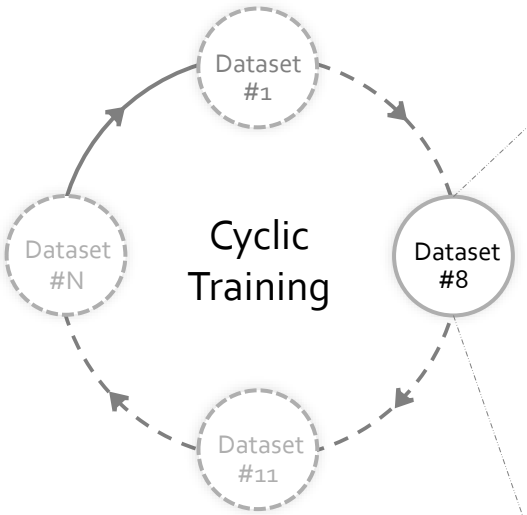


Example: 8. CANDID-PTX

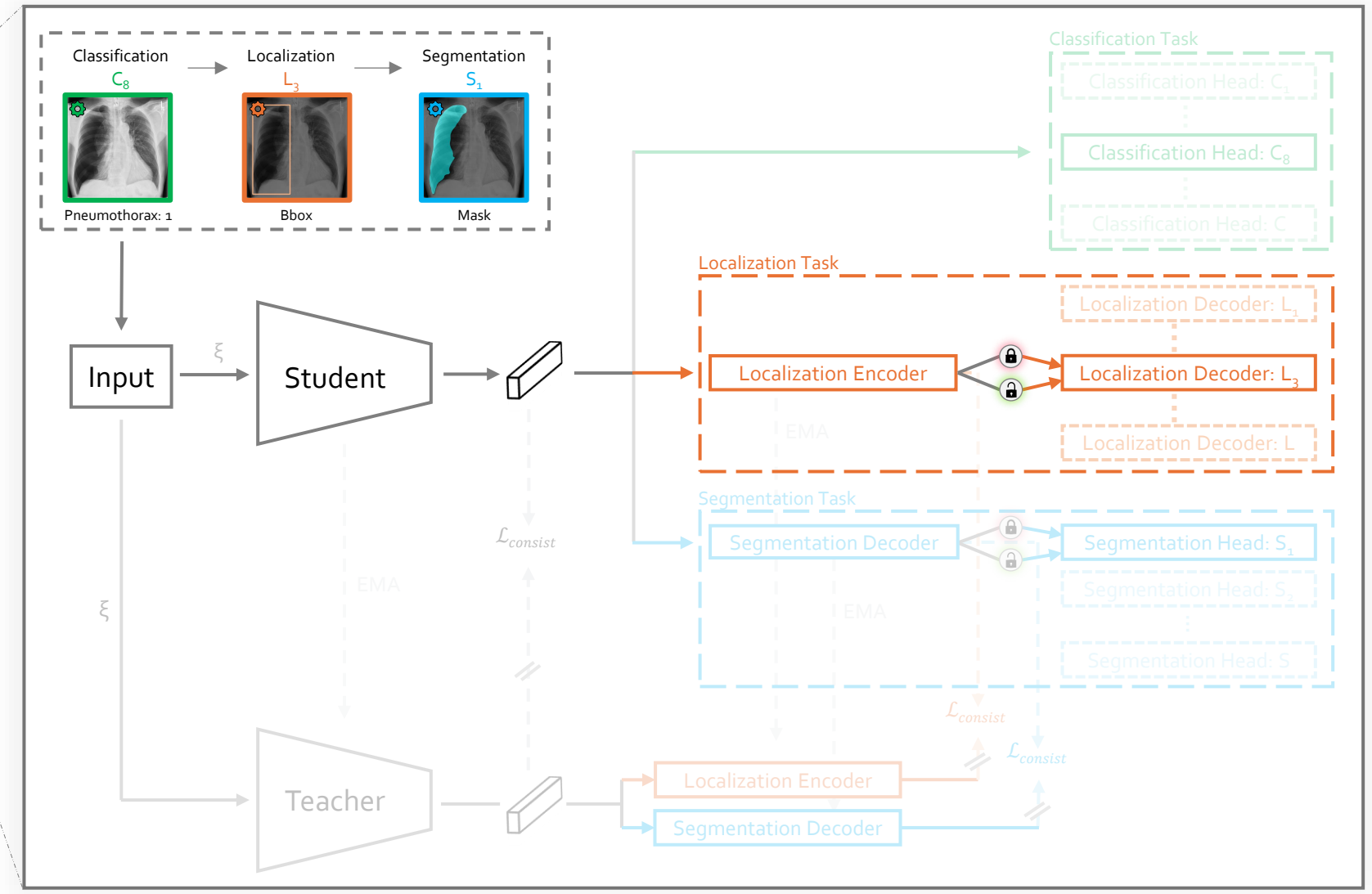


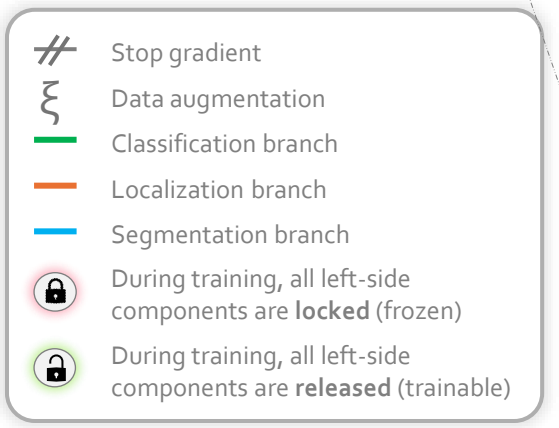


Foundation X

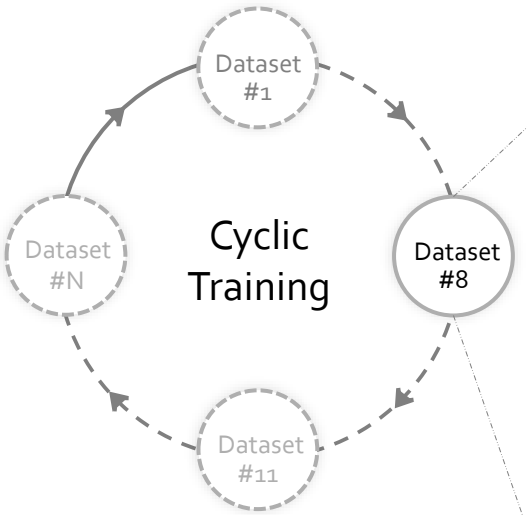


Example: 8. CANDID-PTX

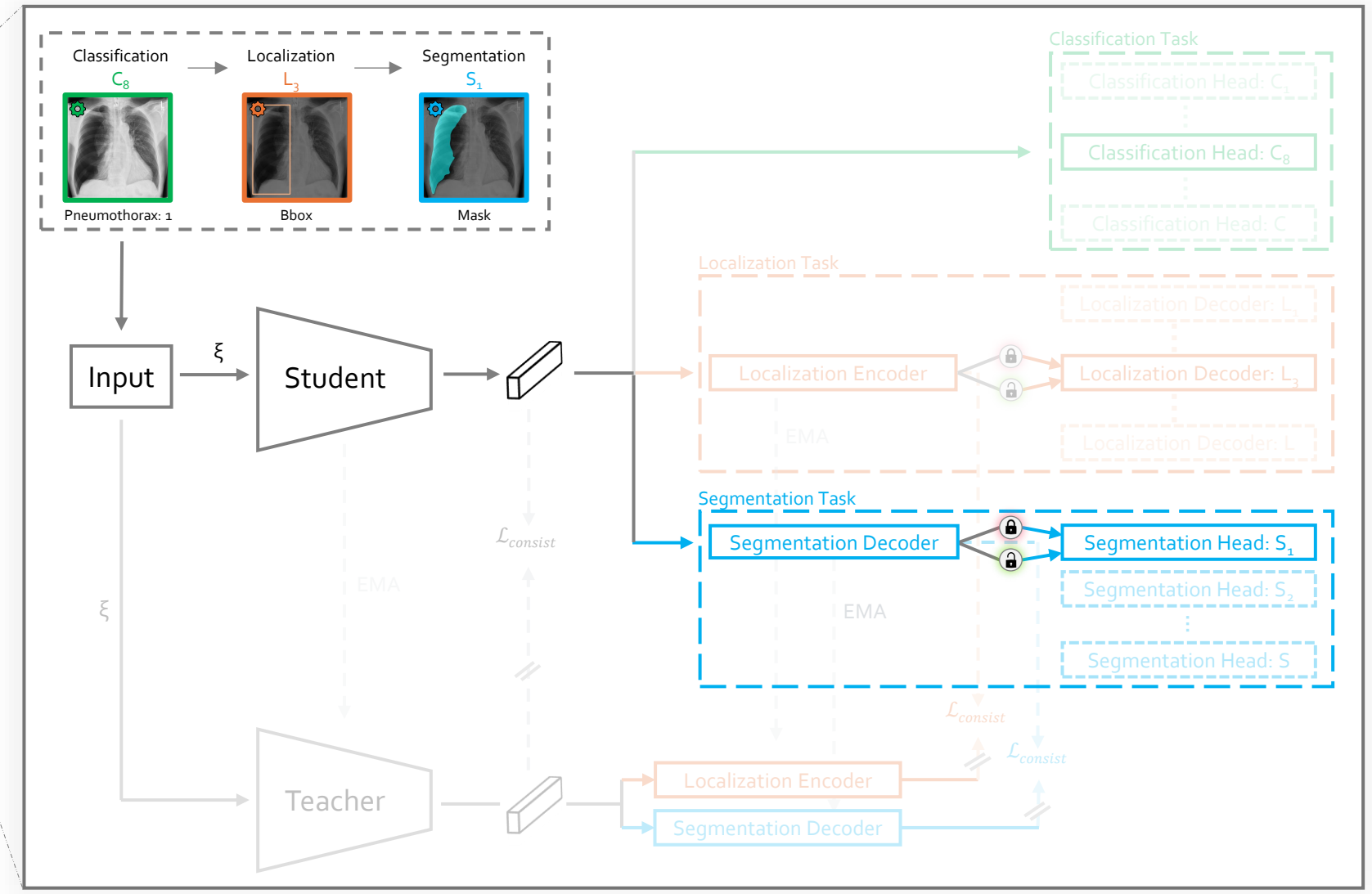


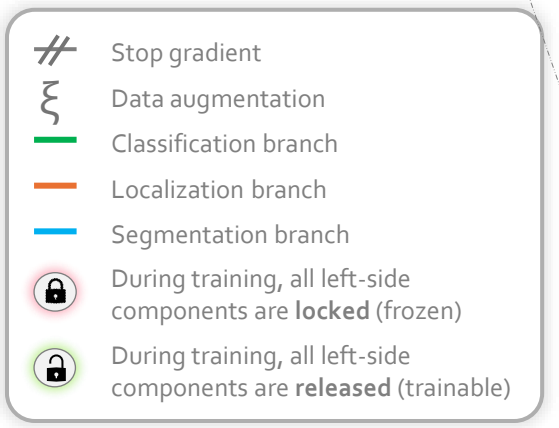


Foundation X

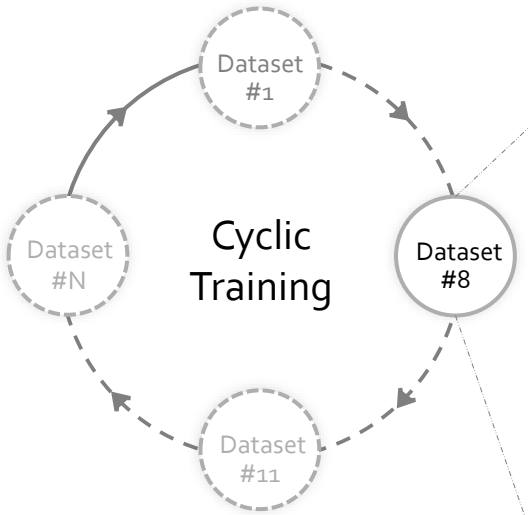


Example: 8. CANDID-PTX

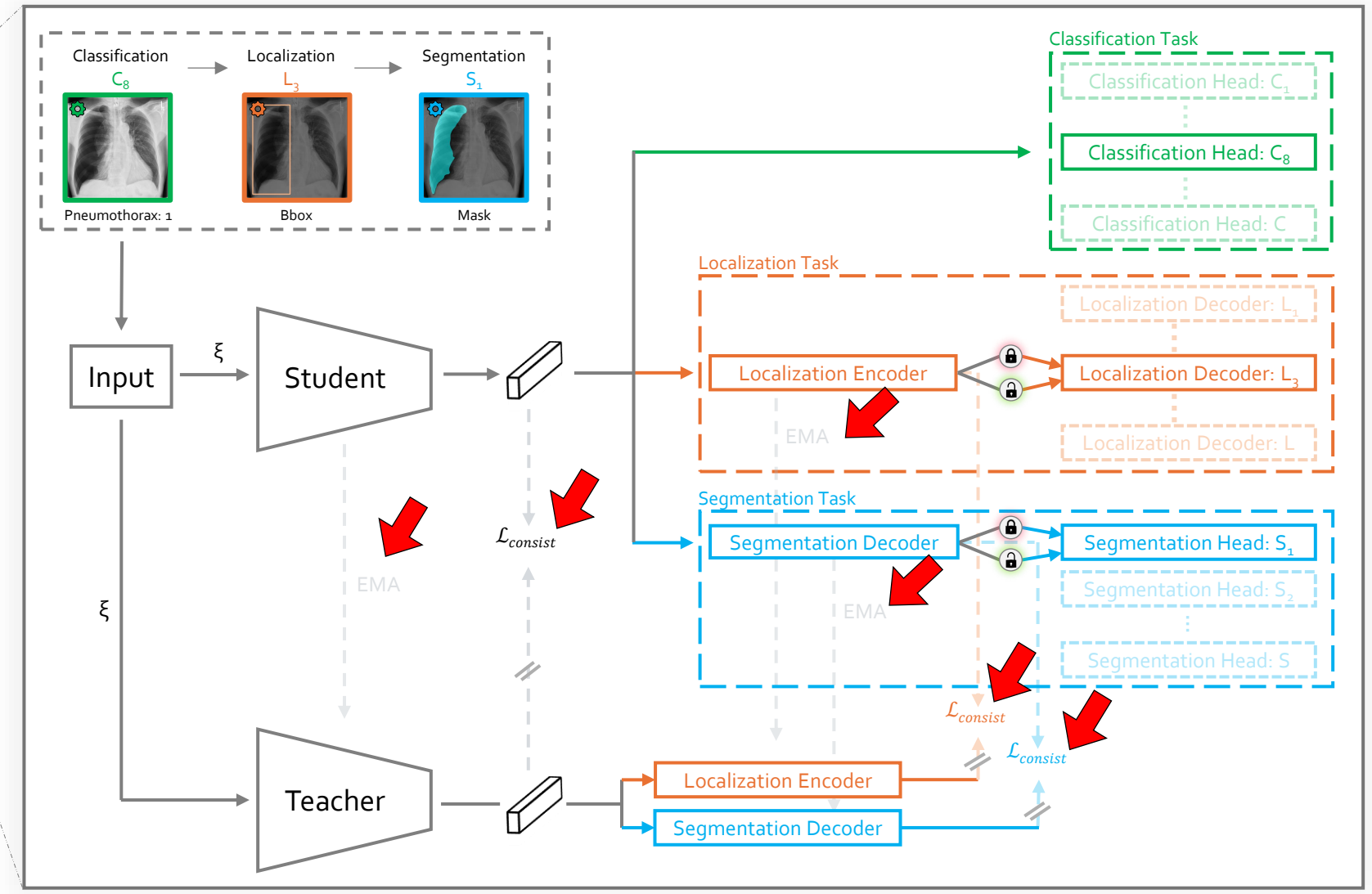




Foundation X

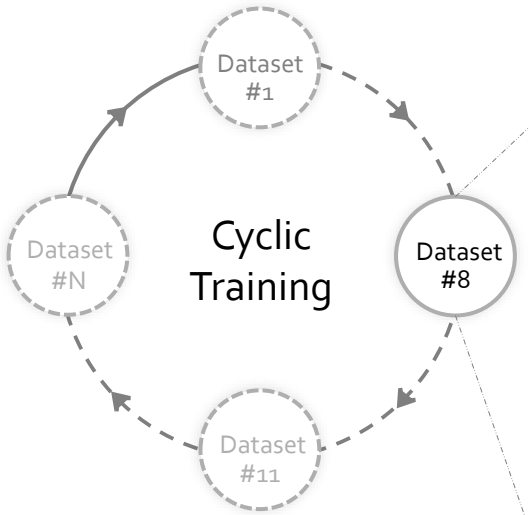


Example: 8. CANDID-PTX

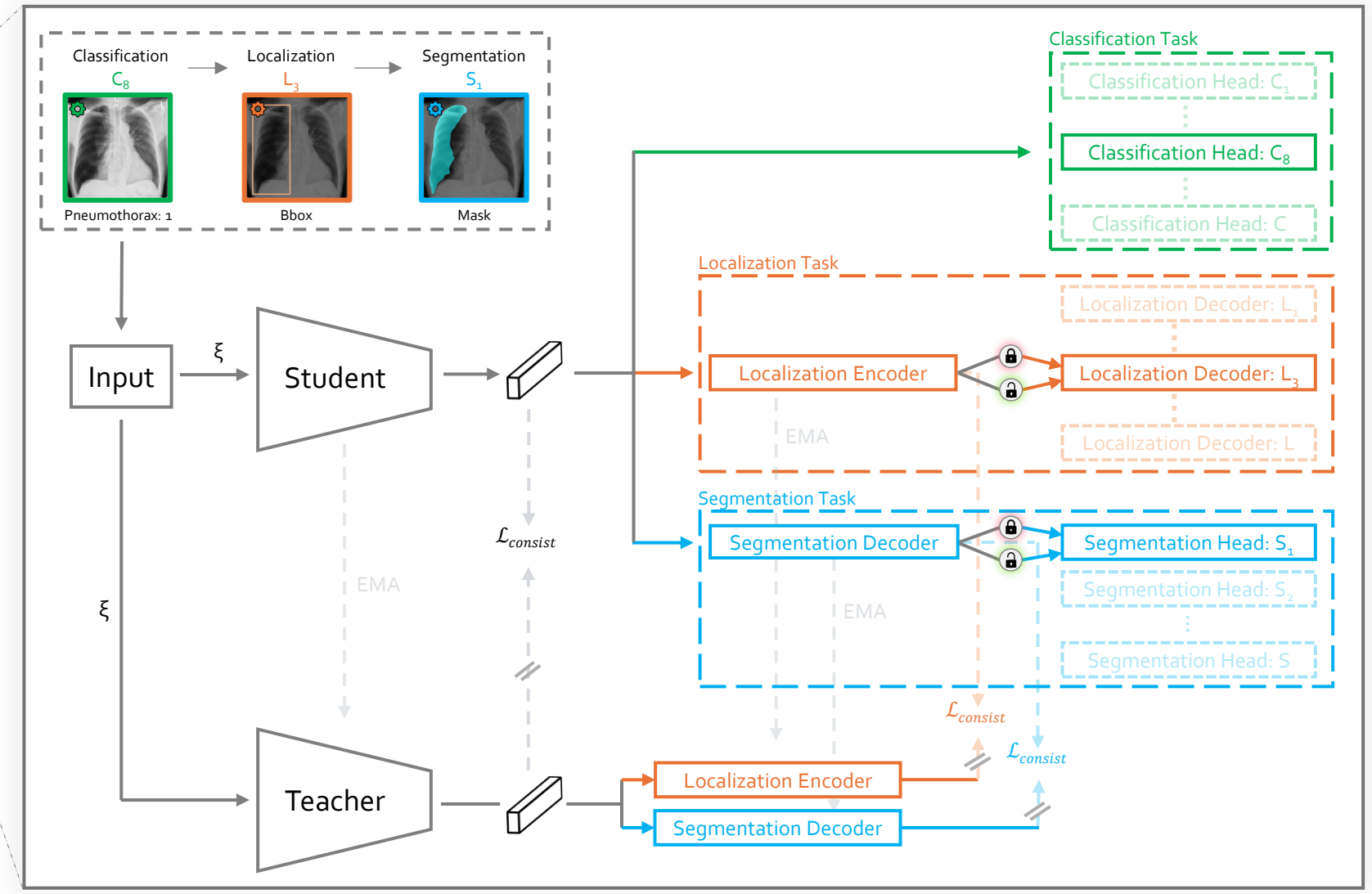


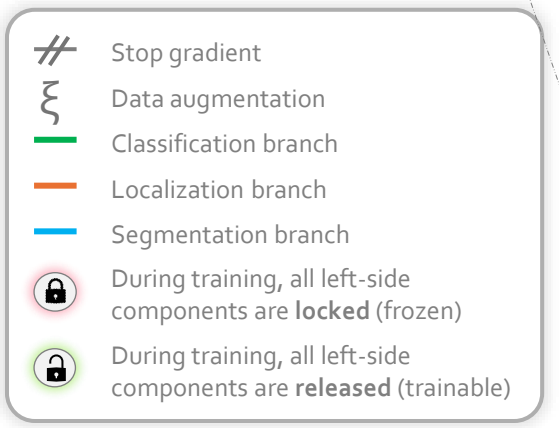
- Stop gradient
- Data augmentation
- Classification branch
- Localization branch
- Segmentation branch
- During training, all left-side components are **locked** (frozen)
- During training, all left-side components are **released** (trainable)

Foundation X



Example: 8. CANDID-PTX





Foundation X enhances performance when jointly trained for localization and segmentation and excels during finetuning

Dataset VinDr-CXR	Baseline Loc.	Baseline Seg.	Foundation X	
			Loc.	Seg.
Heart	80.17	95.82	**88.41** ↑ 8.24	**96.15** ↑ 0.33
Left Lung	90.72	97.46	**95.58** ↑ 4.86	**97.57** ↑ 0.11
Right Lung	92.42	98.03	**96.78** ↑ 4.36	**98.13** ↑ 0.10

Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., Tran, D. Q., Nguyen, D. B., & Vu, V. (2022). VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. *Scientific Data, 9*(1), 429.

Gaggion, N., Mosquera, C., Mansilla, L., et al. (2024). CheXmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images. *Scientific Data, 11*(1), 511.

Foundation X enhances performance when jointly trained for localization and segmentation and excels during finetuning

Dataset (Segmentation)	**Ark**	**POPAR**	**Foundation X**
JSRT-Heart	94.62	94.64	**95.42** ↑0.78
JSRT-Lung	97.48	97.71	**98.04** ↑0.33
JSRT-Clavicle	90.05	90.18	**91.17** ↑0.99
NIH Montgomery	97.68	97.78	**98.29** ↑0.51
VinDr-RibCXR	63.96	61.17	**71.12** ↑7.16

Nguyen, H. C., Le, T. T., Pham, H. H., & Nguyen, H. Q. (2021VinDr-RibCXR: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest X-rays. *arXiv:2107.01327*.

Gordienko, Y., Gang, P., Hui, J., et al. (2019). Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer*, 13*, 638–647. Springer.

Jaeger, S., Candemir, S., Antani, S., et al. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg, 4*(6), 475.

Foundation X excels in few-shot learning and shows strong performance across training samples

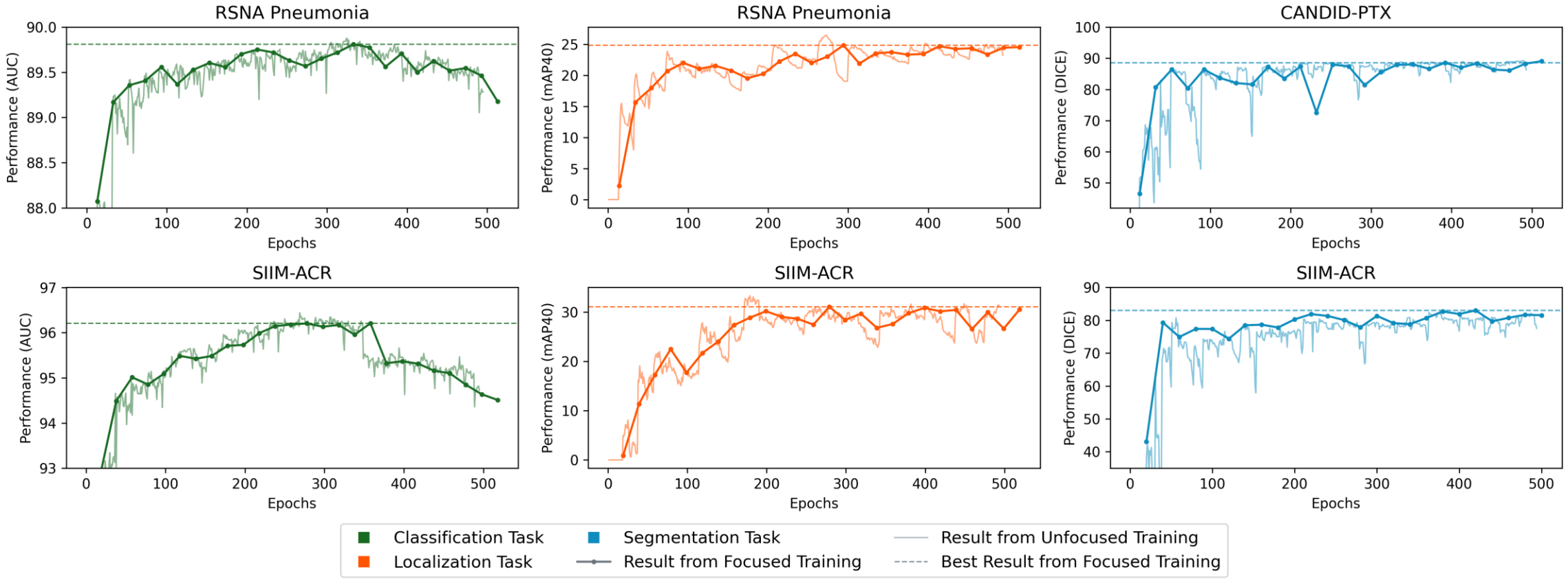
JSRT–Clavicle (Segmentation) Training Samples	Ark	POPAR	Foundation X
24	86.32	86.14	**88.81** ↑ 2.49
20	84.87	86.27	**88.23** ↑ 1.96
15	84.73	83.23	**86.65** ↑ 1.92
12	80.82	81.46	**85.89** ↑ 4.43
6	82.71	79.03	**83.03** ↑ 0.32
3	74.98	70.68	**78.18** ↑ 3.20

Gordienko, Y., Gang, P., Hui, J., et al. (2019). Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer, *13*, 638–647. Springer.

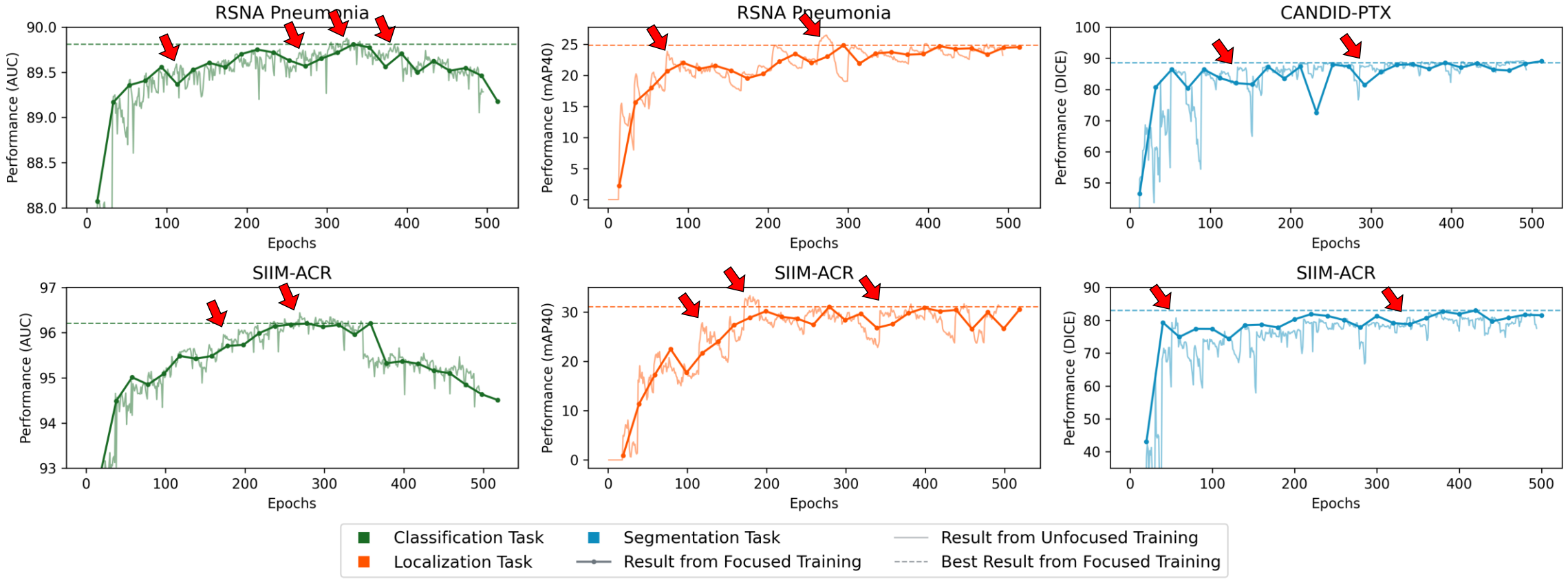
Foundation X maximizes performance improvements by utilizing all available annotations for classification, localization, and segmentation

Dataset	Baseline Cls.	Baseline Loc.	Baseline Seg.	Foundation X		
				Cls.	Loc.	Seg.
1. CheXpert	90.03	-	-	**90.64** ↑0.61	-	-
2. NIH ChestX-ray14	83.05	-	-	**83.35** ↑0.30	-	-
3. VinDr-CXR	95.07	-	-	**95.85** ↑0.78	-	-
4. NIH Shenzhen CXR	98.99	-	-	**99.64** ↑0.65	-	-
5. MIMIC-II	79.12	-	-	78.94 ↓0.18	-	-
6. TBX11K	99.89	78.08	-	**99.95** ↑0.06	**81.80** ↑6.72	-
7. NODE21	99.35	37.78	-	**99.68** ↑0.33	**46.57** ↑8.79	-
8. CANDID-PTX	72.61	50.51	86.36	**73.86** ↑7.25	**54.14** ↑3.63	**89.81** ↑3.45
9. RSNA Pneumonia	88.87	20.83	-	**89.88** ↑1.01	**27.44** ↑6.61	-
10. ChestX-Det	88.17	38.12	79.33	**89.89** ↑1.72	**43.98** ↑5.86	79.17 ↓0.16
11. SIIM-ACR	95.01	28.56	81.92	**96.44** ↑1.43	**34.59** ↑6.03	**83.65** ↑1.73

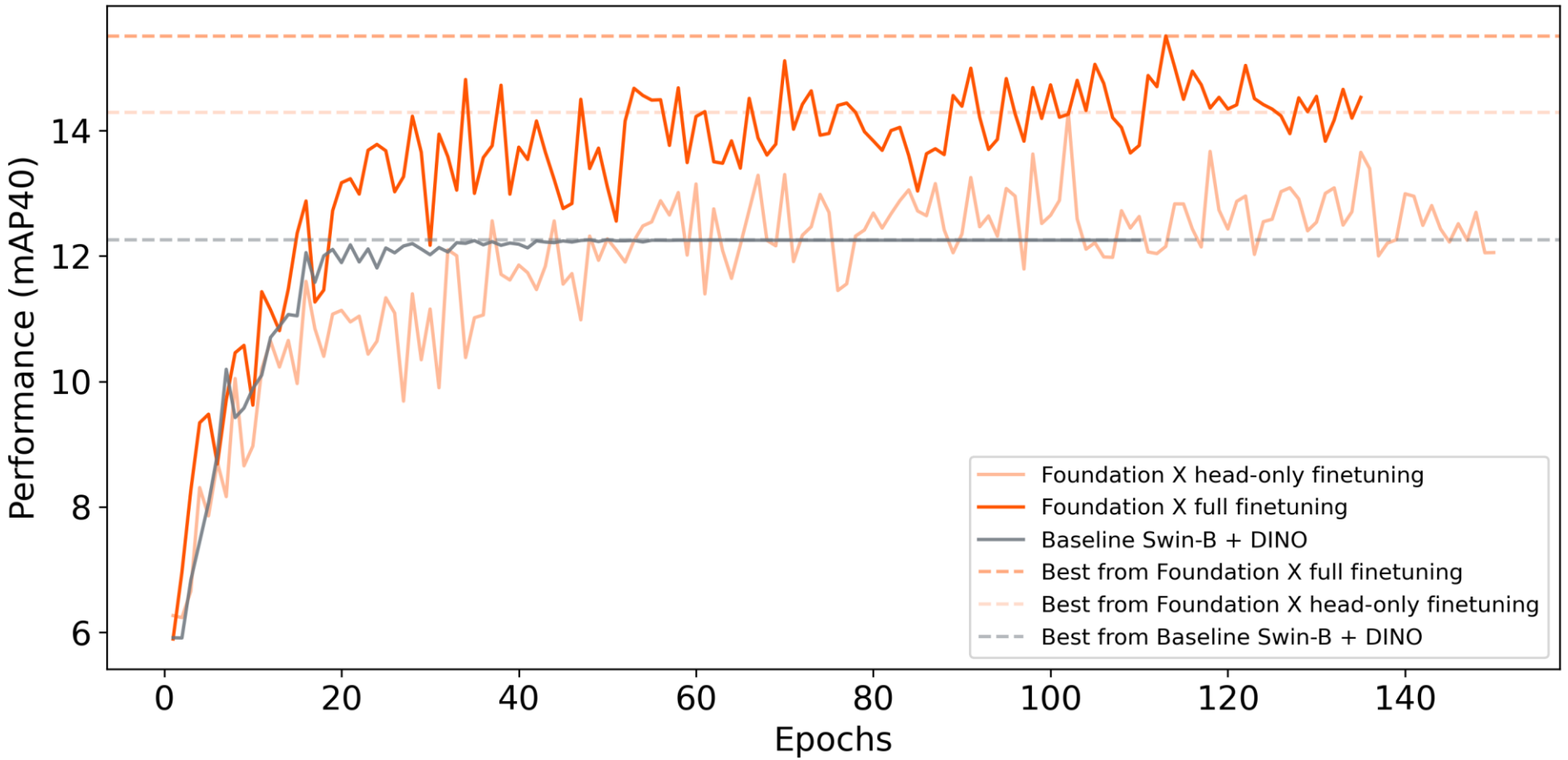
Foundation X maximizes performance with cross-dataset and cross-task learning



Foundation X maximizes performance with cross-dataset and cross-task learning

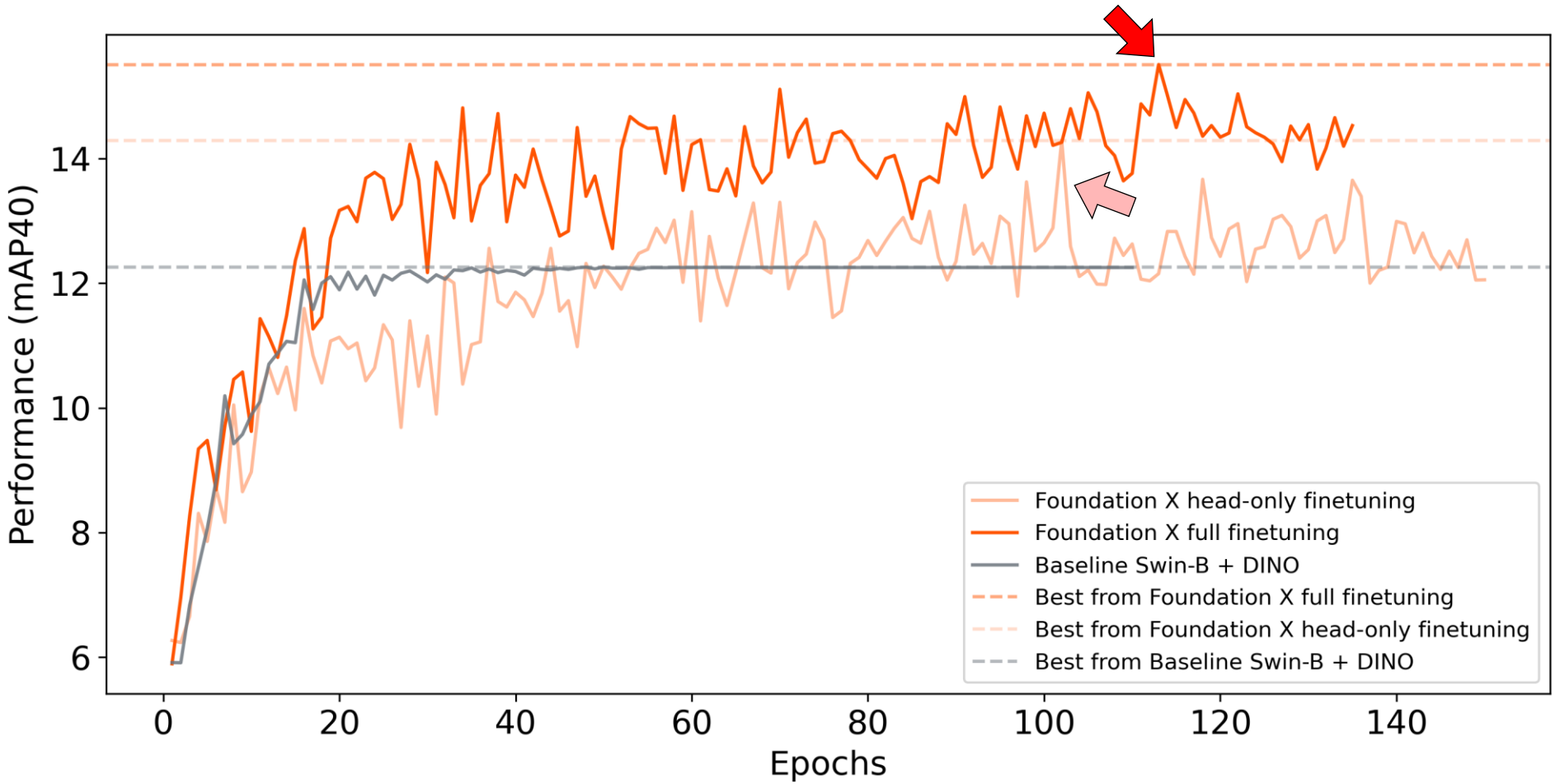


Foundation X full finetuning outperforms head-only finetuning and baseline models



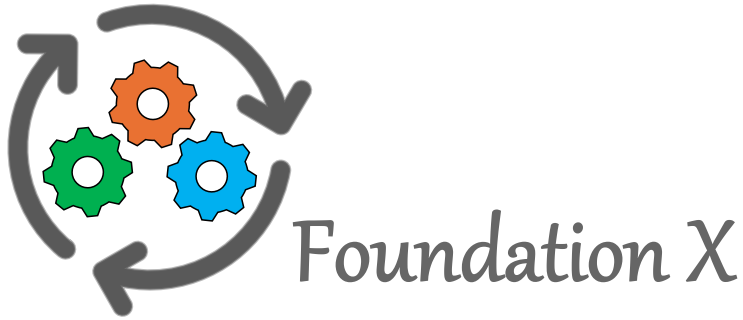
Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., Tran, D. Q., Nguyen, D. B., & Vu, V. (2022). VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. *Scientific Data, 9*(1), 429.

Foundation X full finetuning outperforms head-only finetuning and baseline models



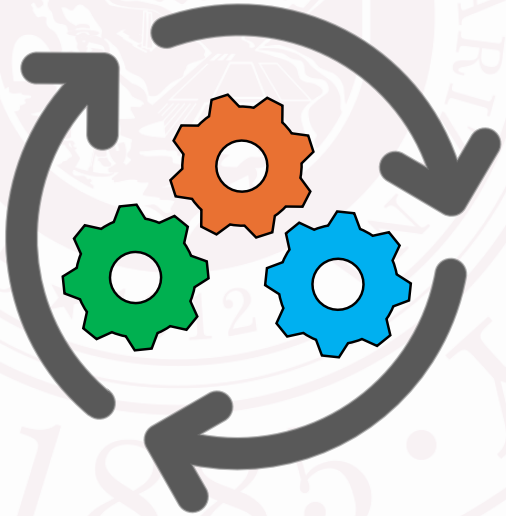
Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., Tran, D. Q., Nguyen, D. B., & Vu, V. (2022). VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. *Scientific Data, 9*(1), 429.

Conclusions



- Introduced Foundation X, an integrated model for classification, localization, and segmentation tasks for Chest X-ray images.
- We proposed a Lock-Release pretraining strategy to enhance the cyclic learning from multiple datasets, preventing task overfitting and ensuring balanced learning across tasks and datasets.
- Provided comprehensive experimental results to demonstrate Foundation X’s improved performance and generalization ability.

Foundation X: Integrating Classification, Localization, and Segmentation through Lock-Release Pretraining Strategy for Chest X-ray Analysis



WACV 2025
TUCSON, ARIZONA • FEB 28 - MAR 4

GitHub