

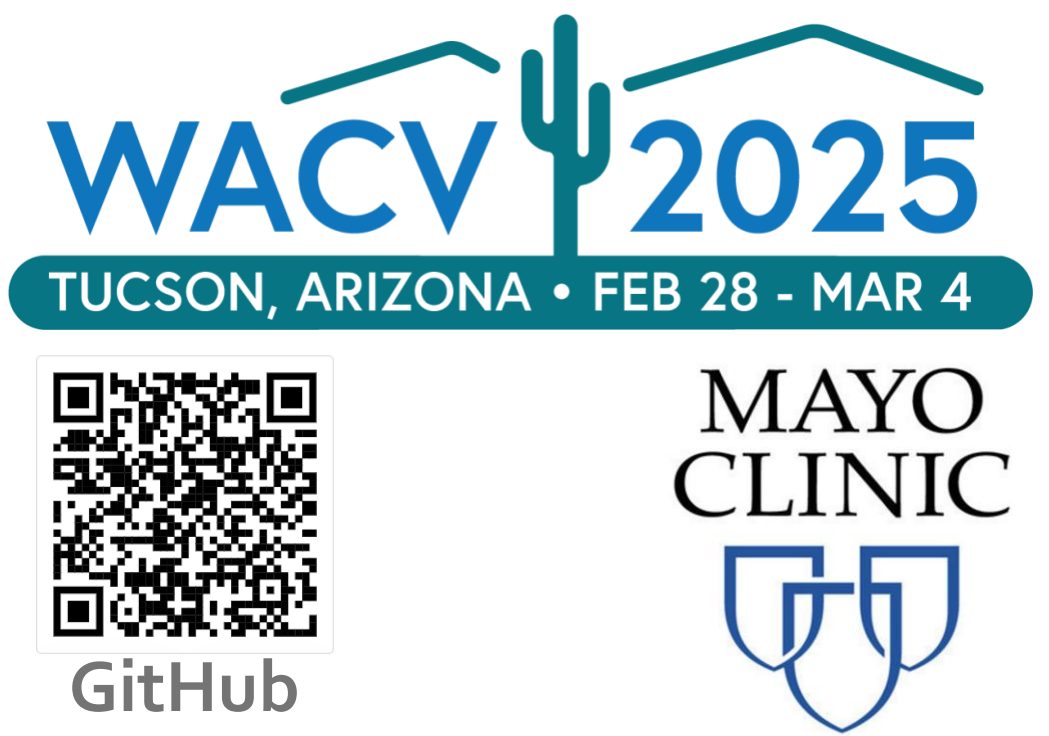
**ASU** College of  
**Health Solutions**  
Arizona State University

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

Nahid Ul Islam<sup>1</sup>, DongAo Ma<sup>1</sup>, Jiaxuan Pang<sup>1</sup>, Shivasakthi Senthil Velan<sup>1</sup>

Michael B. Gotway<sup>2</sup>, and Jianming Liang<sup>1</sup>

<sup>1</sup>Arizona State University <sup>2</sup>Mayo Clinic



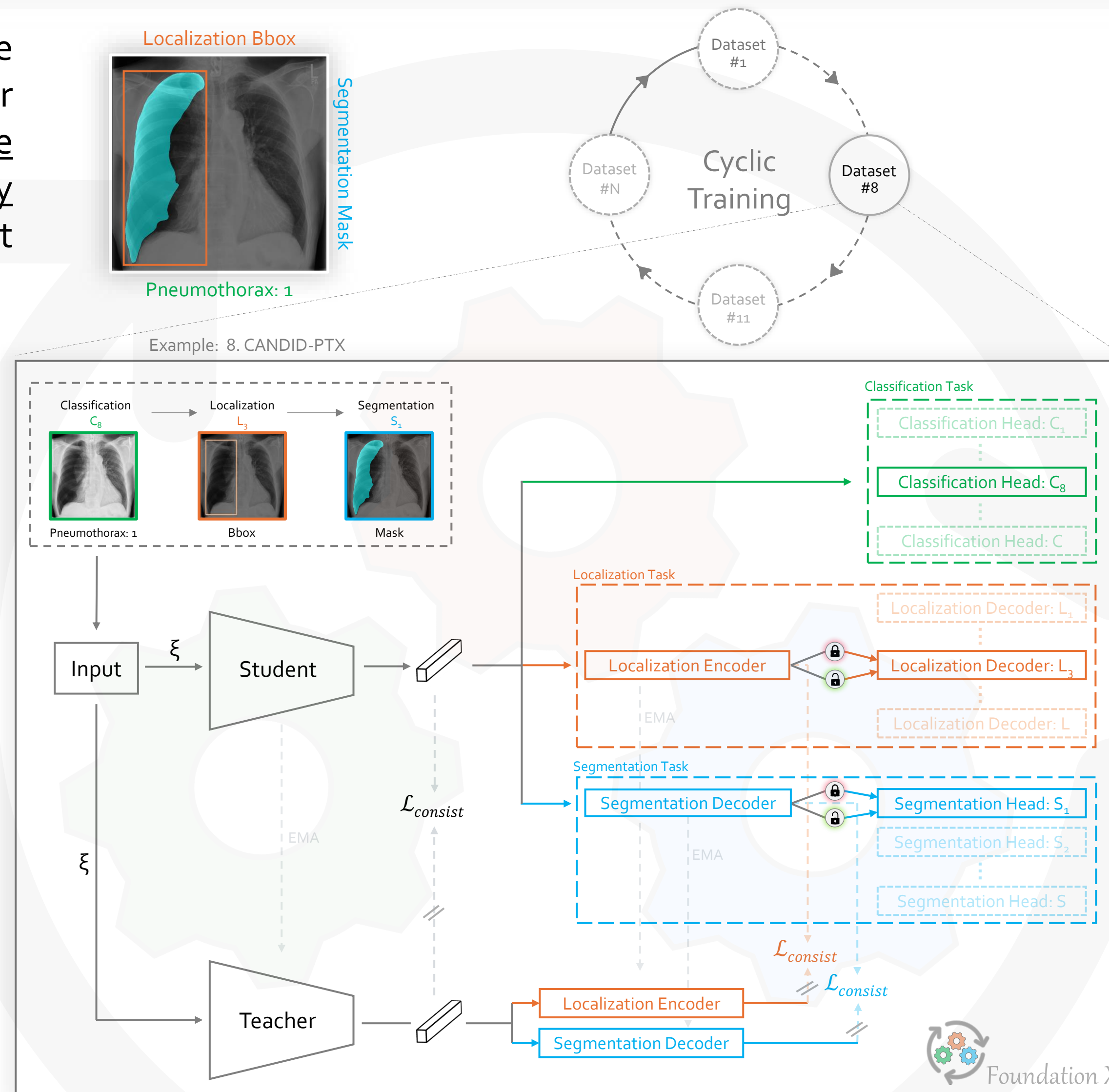
Robust diagnostic models are crucial for improving medical imaging outcomes, but diverse annotations across tasks like classification, localization, and segmentation pose challenges for integration. **Foundation X** introduces a unified framework leveraging our proposed Lock-Release pretraining strategy to combine multi-task annotations effectively. By training on 11 chest X-ray datasets, Foundation X achieves superior multi-task performance, demonstrating significant advancements in cross-dataset learning for classification, localization, and segmentation.

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

Dataset	Baseline	Baseline	Foundation X		Dataset	Ark	POPAR	Foundation X
VinDr-CXR	Loc.	Seg.	Loc.	Seg.				
Heart	80.17	95.82	<b>88.41</b> $\uparrow_{8.24}$	<b>96.15</b> $\uparrow_{0.33}$	JSRT-Heart	94.62	94.64	<b>95.42</b> $\uparrow_{0.78}$
Left Lung	90.72	97.46	<b>95.58</b> $\uparrow_{4.86}$	<b>97.57</b> $\uparrow_{0.11}$	JSRT-Lung	97.48	97.71	<b>98.04</b> $\uparrow_{0.33}$
Right Lung	92.42	98.03	<b>96.78</b> $\uparrow_{4.36}$	<b>98.13</b> $\uparrow_{0.10}$	JSRT-Clavicle	90.05	90.18	<b>91.17</b> $\uparrow_{0.99}$
					NIH Montgomery	97.68	97.78	<b>98.29</b> $\uparrow_{0.51}$
					VinDr-RibCXR	63.96	61.17	<b>71.12</b> $\uparrow_{7.16}$

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

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

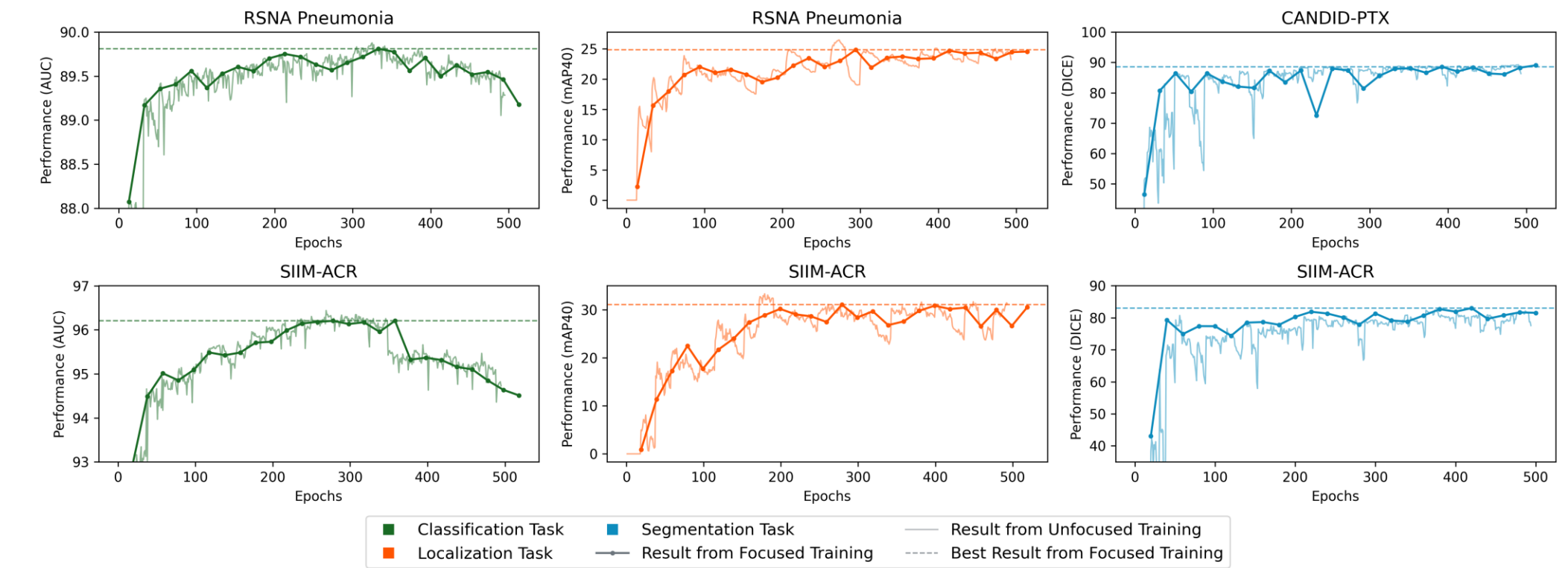


## How can we integrate classification, localization, and segmentation tasks into a single model to enhance performance and generalization in Chest X-ray image analysis?

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

JSRT-Clavicle Training Samples	Ark	POPAR	Foundation X
24	86.32	86.14	<b>88.81</b> $\uparrow_{2.49}$
20	84.87	86.27	<b>88.23</b> $\uparrow_{1.96}$
15	84.73	83.23	<b>86.65</b> $\uparrow_{1.92}$
12	80.82	81.46	<b>85.89</b> $\uparrow_{4.43}$
6	82.71	79.03	<b>83.03</b> $\uparrow_{0.32}$
3	74.98	70.68	<b>78.18</b> $\uparrow_{3.20}$

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



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

