Seeking an Optimal Approach for Computer-aided Pulmonary Embolism Detection

Architectures · Initializations · Paradigms

What deep learning architectures, model initializations, and learning paradigms should be used for computer-aided pulmonary embolism detection?
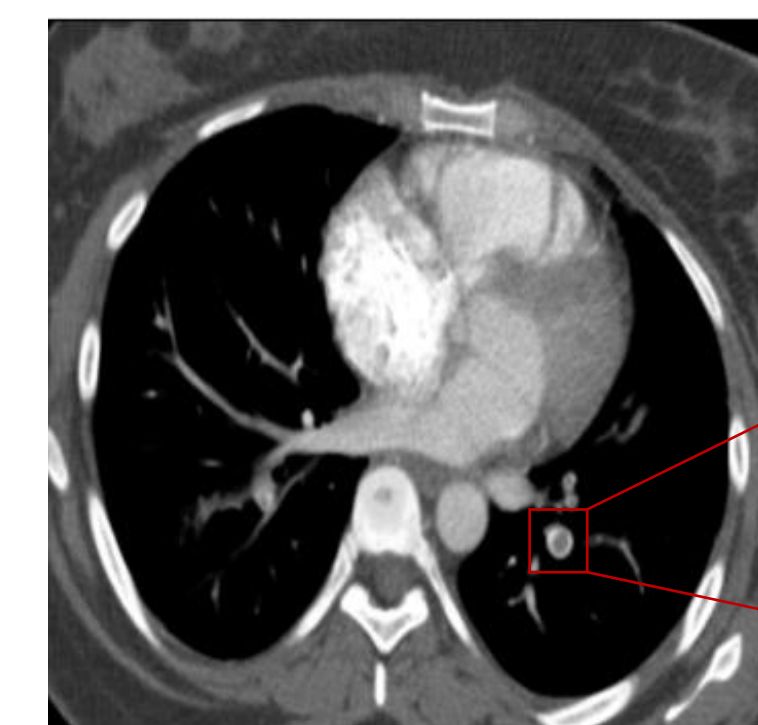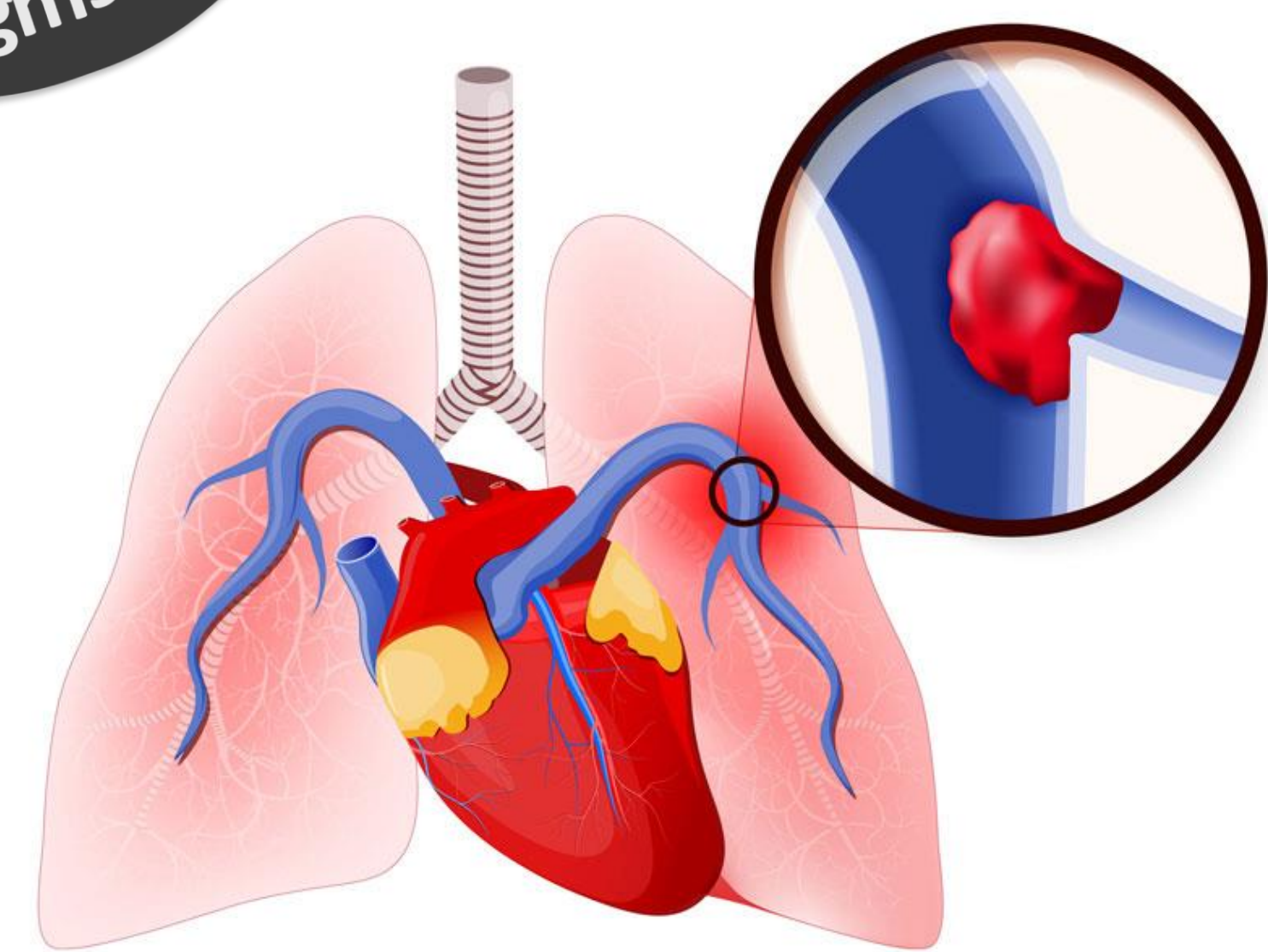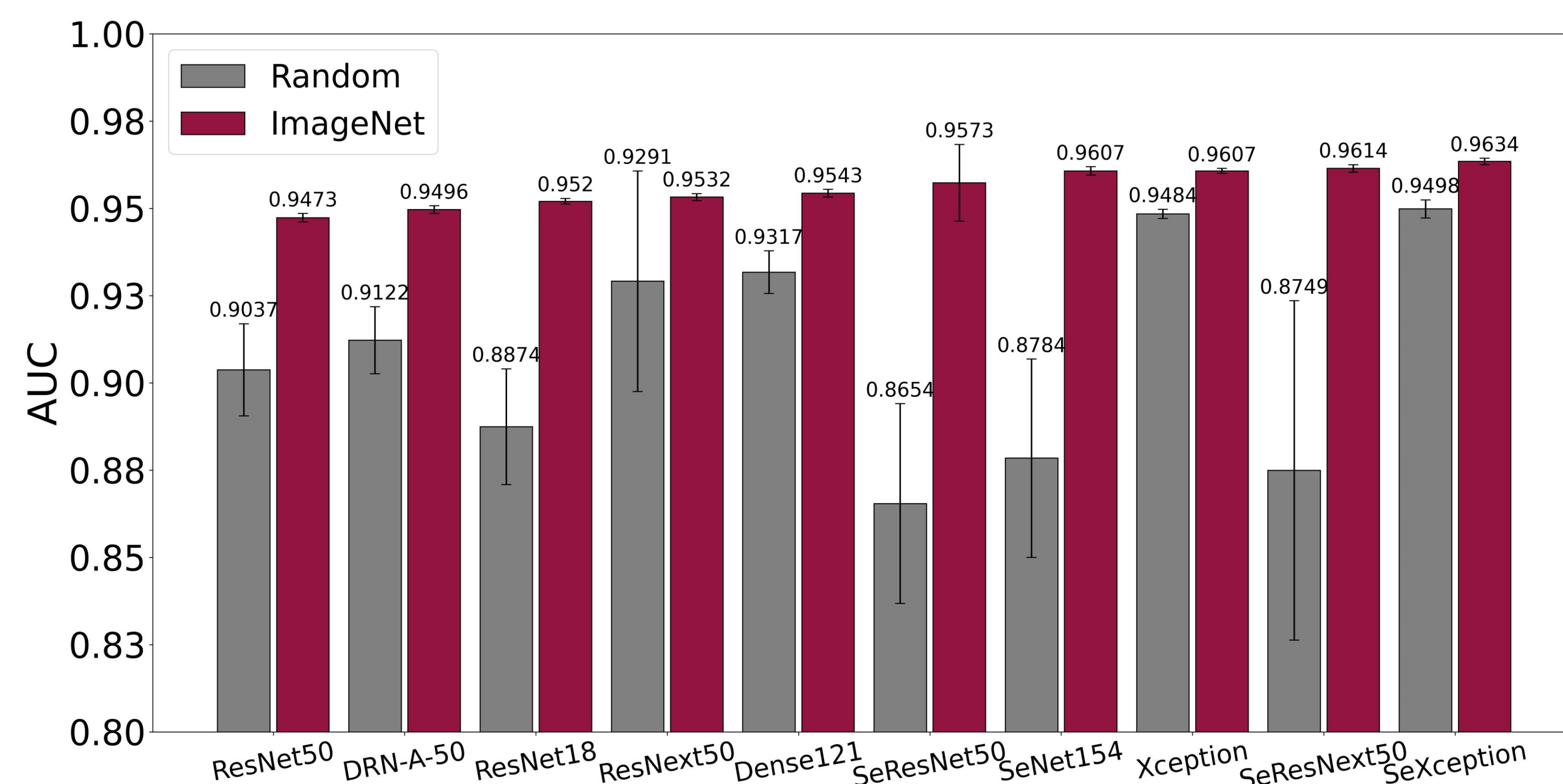
Slice-level Classification
- Training from scratch
- Fine-tuning
  - Supervised pretrained models
  - Self-supervised pretrained models

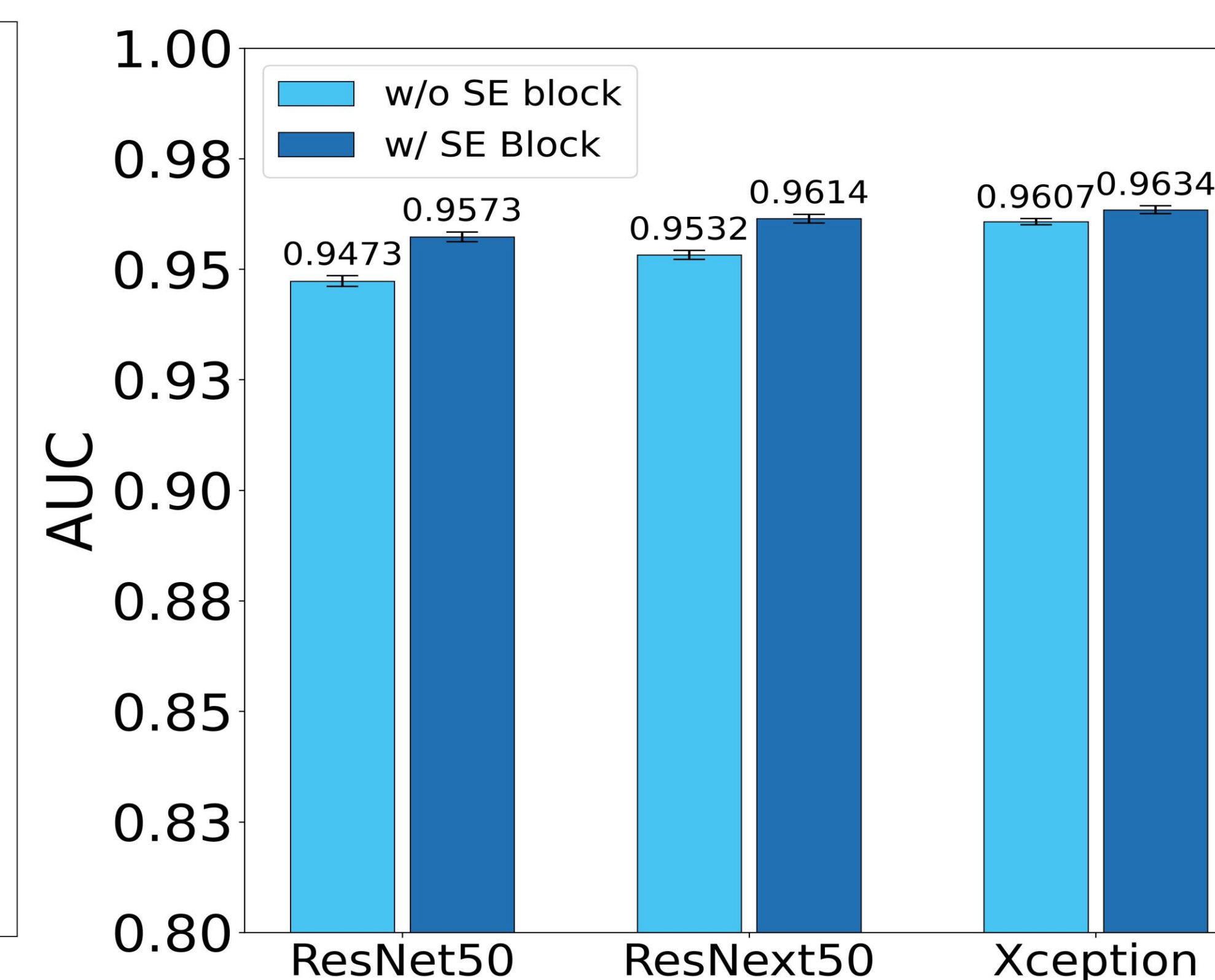Organize for each exam

Feature vector
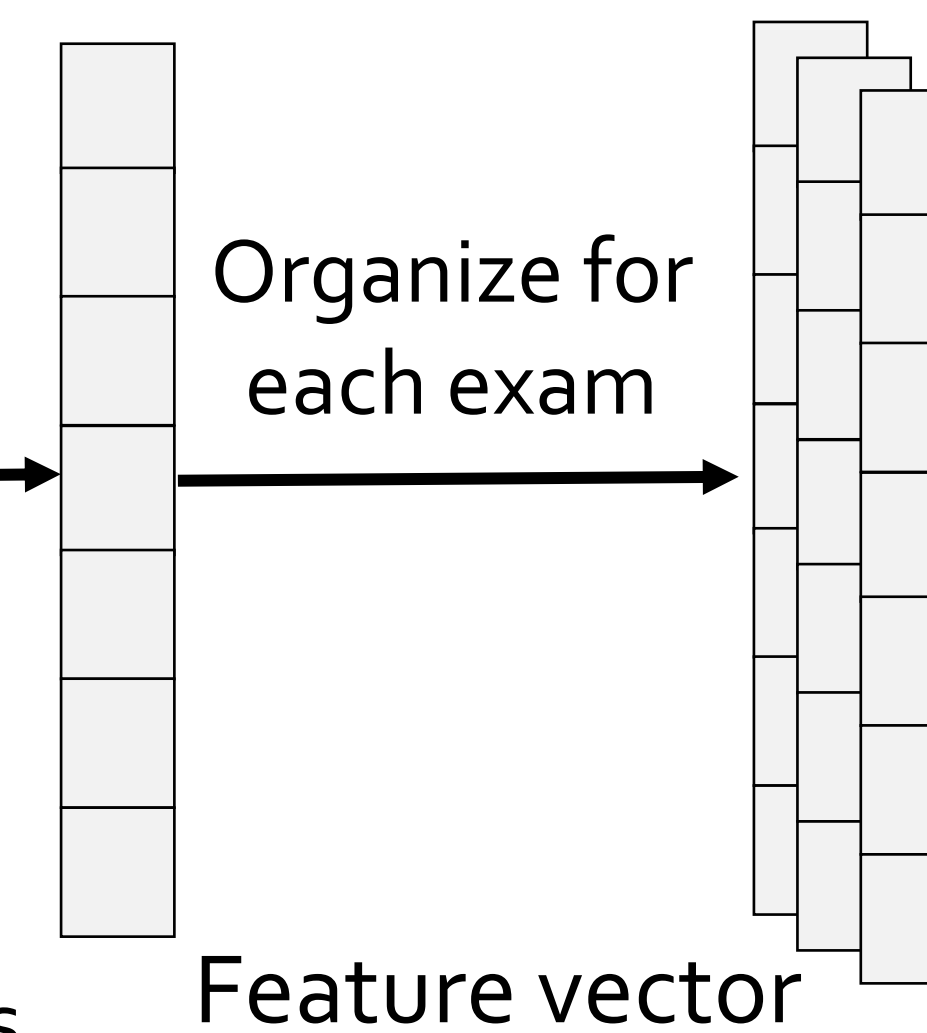
Exam-level Classification
- Bidirectional gated recurrent unit (GRU)
- Attention-based deep multiple instance learning (MIL)

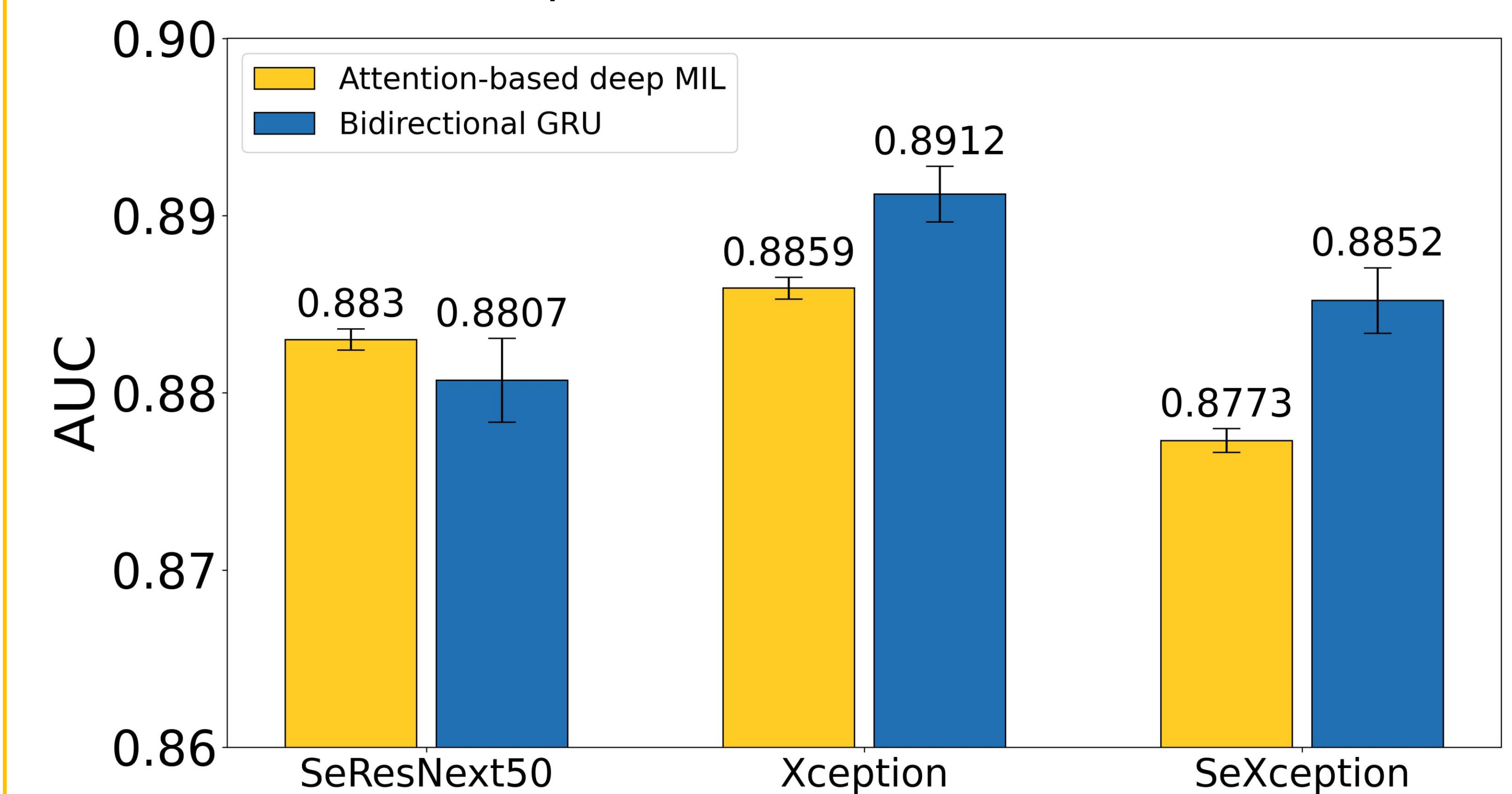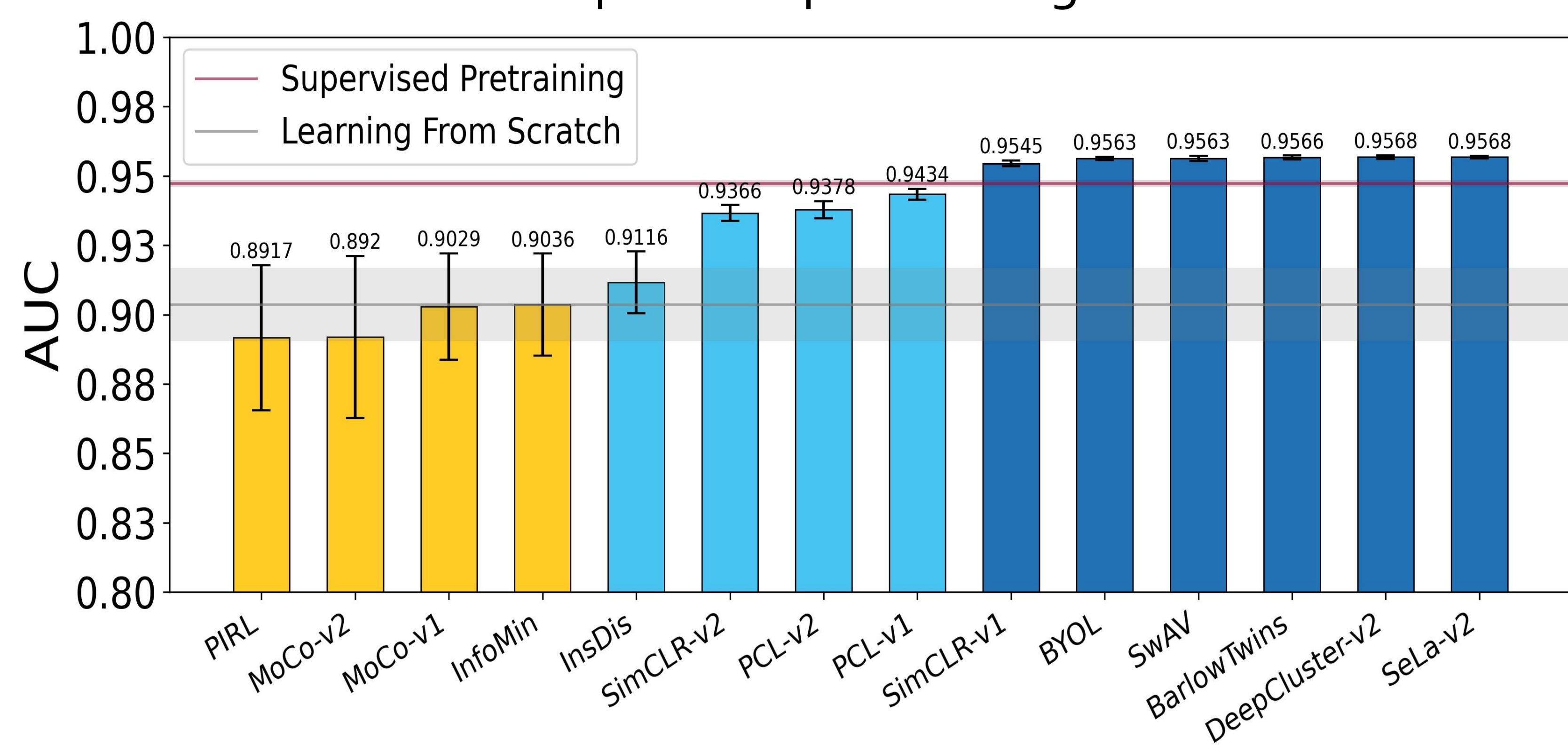## Transfer learning consistently improves performance across the 10 different CNN architectures

Legend: Random, ImageNet

Architecture	Random	ImageNet
ResNet50	0.9037	0.9473
DRN-A-50	0.9122	0.9496
ResNet18	0.8874	0.952
ResNext50	0.9291	0.9532
Dense121	0.9317	0.9543
SeResNet50	0.8654	0.9573
SeNet154	0.8784	0.9607
Xception	0.9484	0.9607
SeResNext50	0.8749	0.9614
SeXception	0.9498	0.9634

AUC (y-axis)

Squeeze & excitation blocks enhance CNN performance

Legend: w/o SE block, w/ SE Block

Architecture	w/o SE block	w/ SE Block
ResNet50	0.9473	0.9573
ResNext50	0.9532	0.9614
Xception	0.9607	0.9634

Bidirectional GRU marginally outperforms attention-based deep MIL at the exam level

Legend: Attention-based deep MIL, Bidirectional GRU

Architecture	Attention-based deep MIL	Bidirectional GRU
SeResNext50	0.883	0.8807
Xception	0.8859	0.8912
SeXception	0.8773	0.8852

Self-supervised pre-training overtakes (fully) supervised pre-training

Legend: Supervised Pretraining, Learning From Scratch

Method	AUC
PIRL	0.8917
MoCo-v2	0.892
MoCo-v1	0.9029
InfoMin	0.9036
InsDis	0.9116
SimCLR-v2	0.9366
PCL-v2	0.9378
PCL-v1	0.9434
SimCLR-v1	0.9545
BYOL	0.9563
SwAV	0.9563
BarlowTwins	0.9566
DeepCluster-v2	0.9568
SeLa-v2	0.9568

Vision transformers underperform CNNs

Legend: Best CNN performance; Random + 512 ImageSize; ImageNet21K + 224 ImageSize; ImageNet21K + 512 ImageSize; ImageNet21K + 576 ImageSize

Patch	Random + 512	ImageNet21K + 224	ImageNet21K + 512	ImageNet21K + 576
32x32	0.8212	0.8456	0.8847	
16x16	0.8385	0.8826	0.9065	0.9179
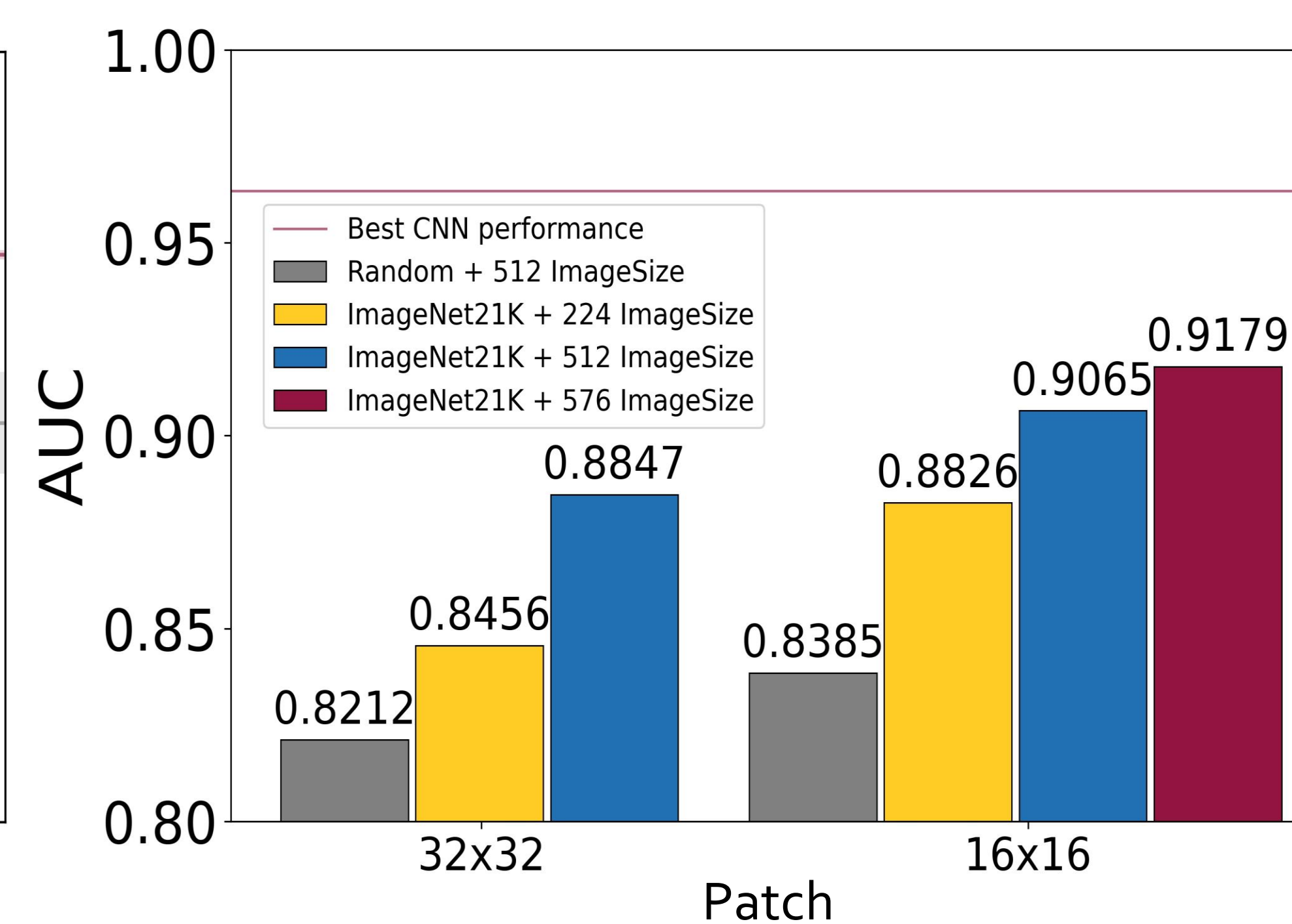
Attention maps highlight the potential PE