A Systematic Benchmarking Analysis of Transfer Learning for Medical Image Analysis

Mohammad Reza Hosseinzadeh Taher[1], Fatemeh Haghighi[1], Ruibin Feng[2], Michael B. Gotway[3], and Jianming Liang[1]

[1] Arizona State University [2] Stanford University [3] Mayo Clinic

Transfer Learning

Knowledge Transfer

+ Improving the performance of the target task
+ Accelerating the target model convergence
+ Reducing human annotation-efforts

Thorax Diseases Classification

Atelectasis | Pneumonia | Pneumothorax

Skin Diseases Classification

Epidermal | Melanocytic | Dermoscopy

Diabetic Retinopathy Classification

Proliferative | Severe | Moderate

Fine-grained Pre-training?

Self-supervised Pre-training?

Supervised IMAGENET Pre-training

Medical Pre-training?

Domain-adaptive Pre-training?

How do newly-developed pre-training techniques work for medical image analysis?

iNaturalist

vs.

IMAGENET

IMAGENET

iNaturalist

Partridge

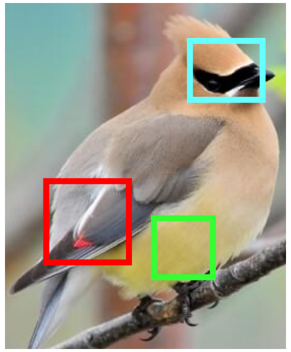
Magpie

Coarse-grained dataset

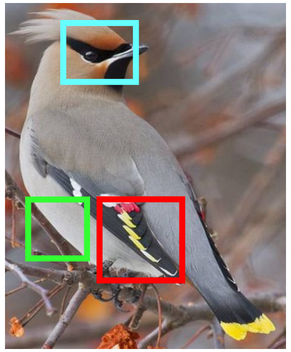
⬇

Coarse-grained features

Cedar Waxwing



Bohemian Waxwing

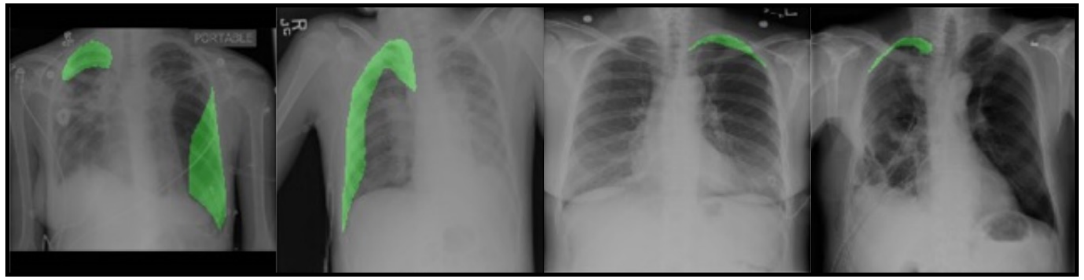


Fine-grained dataset

⬇

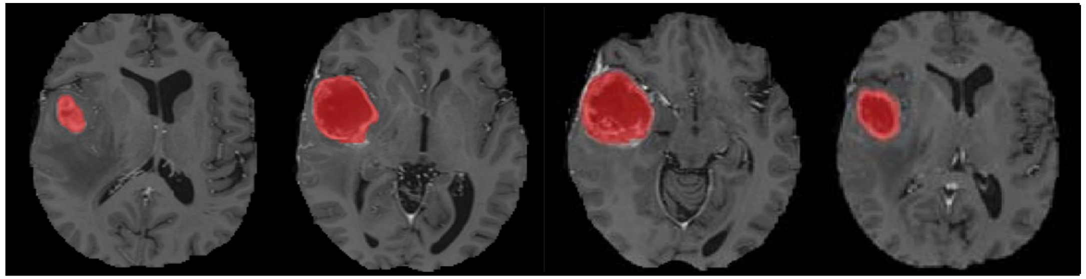
Fine-grained features

Fine-grained Tasks Demand Fine-grained Features

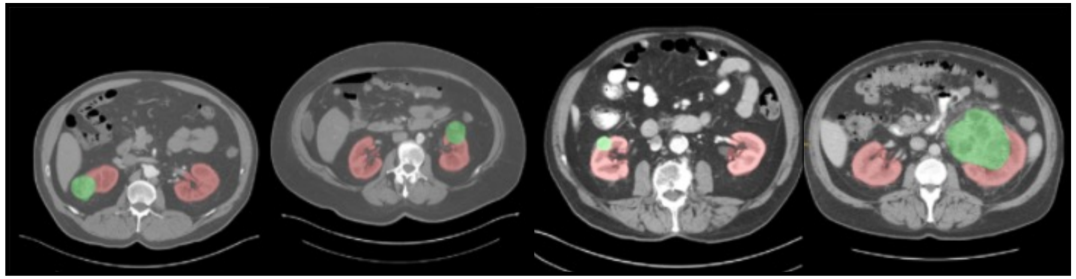


Pneumothorax segmentation

Blood vessel segmentation



Brain tumor segmentation

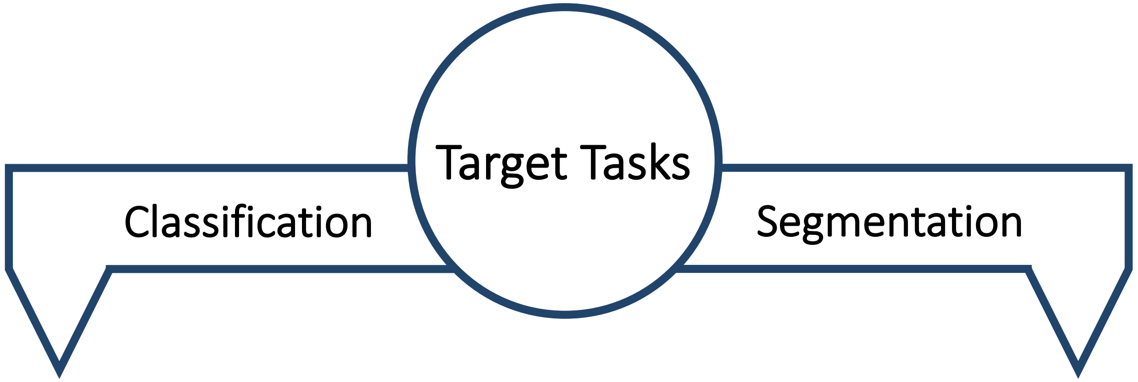


Kidney/lesion segmentation

1

What advantages can supervised iNat2021 models offer for medical imaging in comparison with supervised ImageNet models?

Benchmarking the Transferability of Supervised iNat2021 Model



Fourteen thorax diseases classification (X-ray)

Tuberculosis Detection (X-ray)

Five thorax diseases classification (X-ray)

Pulmonary Embolism Detection (CT)

Pneumothorax Segmentation (X-ray)

Lung Segmentation (X-ray)

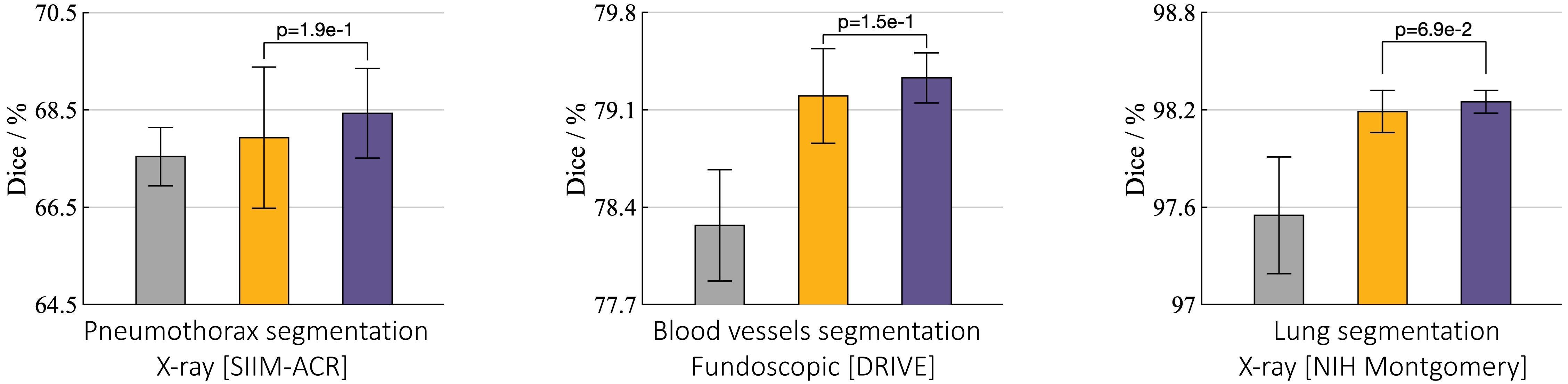
Blood vessel classification (Fundus)

Result I: Pre-trained model on iNat2021 is better suited for segmentation tasks

- ✓ Finer data granularity of iNat2021 yields a more fine-grained visual feature space
  - ✓ Capturing essential pixel-level cues for medical segmentation tasks

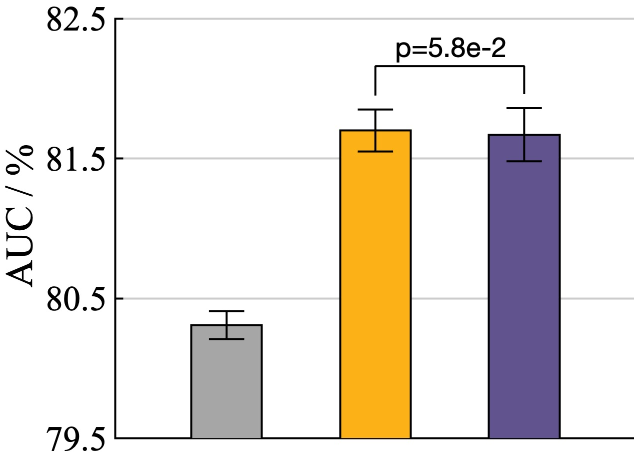
Random initialization
ImageNet
iNat2021*

**The most recent large-scale fine-grained datasets*



Result I: Pre-trained model on ImageNet is superior on classification tasks

- ✓ Coarser data granularity yields high-level semantic features.



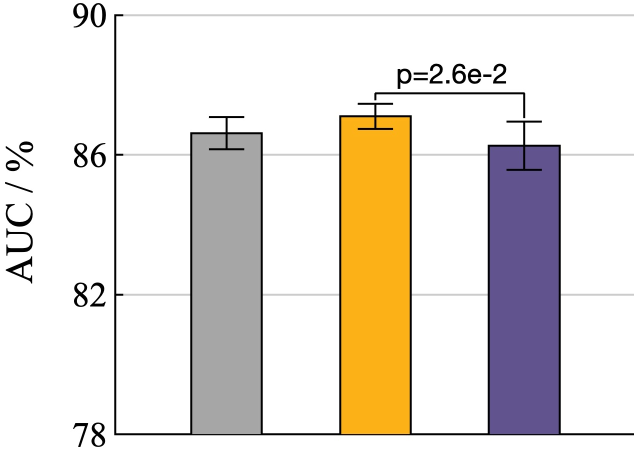
Fourteen thorax diseases classification
X-ray [NIH ChestX-Ray14]

Random initialization

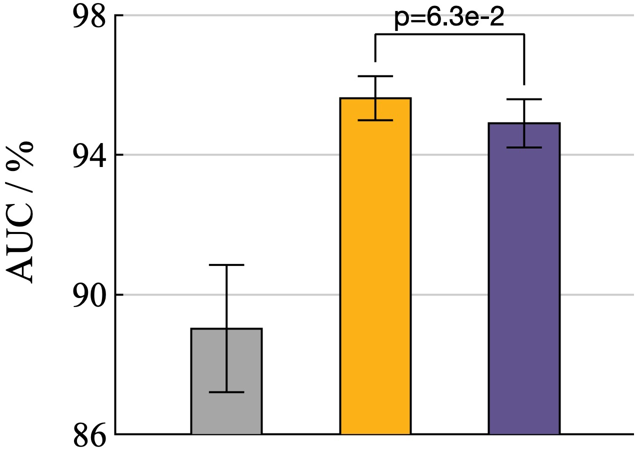
ImageNet

iNat2021*

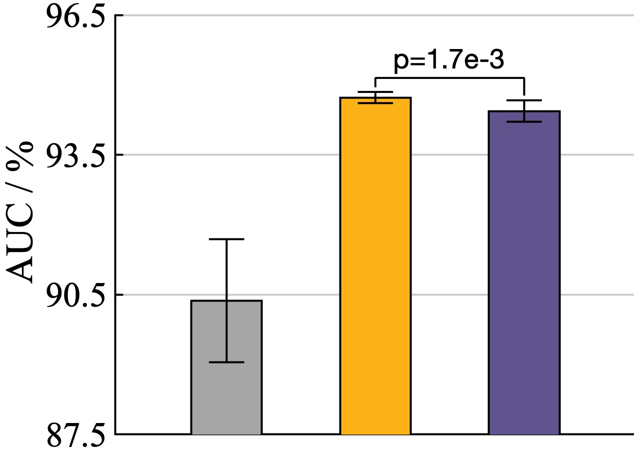
**The most recent large-scale fine-grained datasets*



Five thorax diseases classification
X-ray [CheXpert]



Tuberculosis Detection
X-ray [NIH Shenzhen CXR]



Pulmonary Embolism Detection
CT [RSNA PE Detection]

Q 1

What advantages can supervised iNat2021 models offer for medical imaging in comparison with supervised ImageNet models?

Pre-trained models on fine-grained data are better suited for segmentation tasks, while pre-trained models on coarse-grained data prevail on classification tasks.

Self-supervised ImageNet models

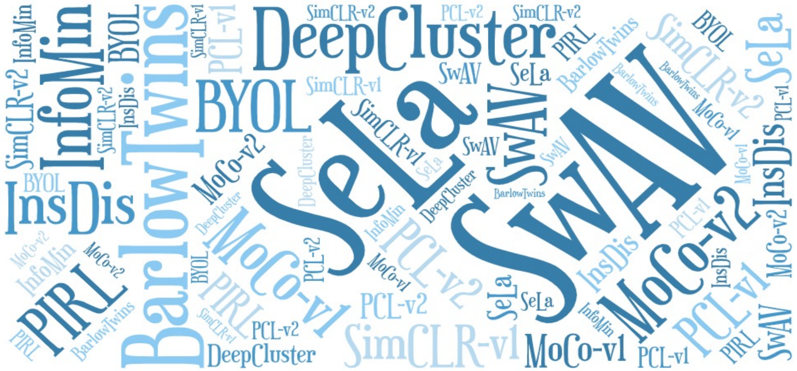
VS.

Supervised ImageNet models

Supervised Learning

Self-supervised Learning

← IMAGENET →



High-level semantics

⬇

Domain-specific features

Low/mid-level semantics

⬇

More generalizable features

A. Islam, et al. "A Broad Study on the Transferability of Visual Representations with Contrastive Learning". ICCV. 2021.

2

How generalizable are the self-supervised ImageNet models to medical imaging in comparison with supervised ImageNet models?

Benchmarking the Transferability of 14 Self-supervised Methods

1. Contrastive learning based on instance discrimination
   InsDis, MoCo-v1, MoCo-v2, SimCLR-v1, SimCLR-v2, and BYOL
2. Contrastive learning based on JigSaw shuffling
   PIRL
3. Clustering
   DeepCluster-v2 and SeLa-v2
4. Clustering bridging Contrastive learning
   PCL-v1, PCL-v2, and SwAV
5. Mutual information reduction
   InfoMin
6. Redundancy reduction
   Barlow Twins

Result II: Top self-supervised ImageNet models outperform supervised ImageNet model

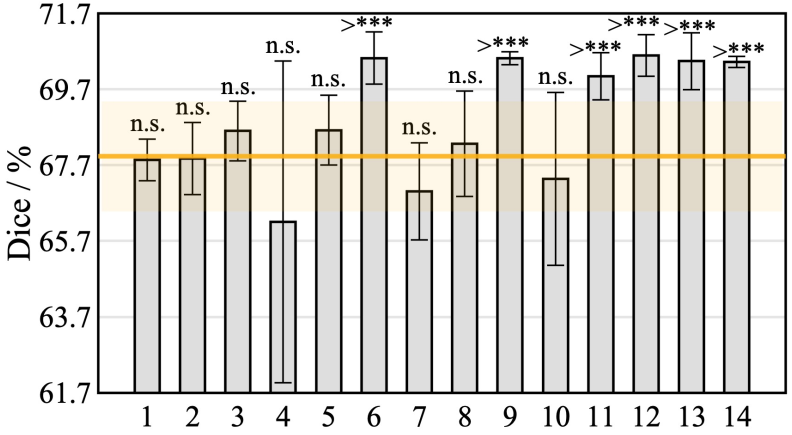
✓ Self-supervised models attend to larger image regions, helping in higher transferability to segmentation tasks.

L. Ericsson, et al. CVPR. 2021.

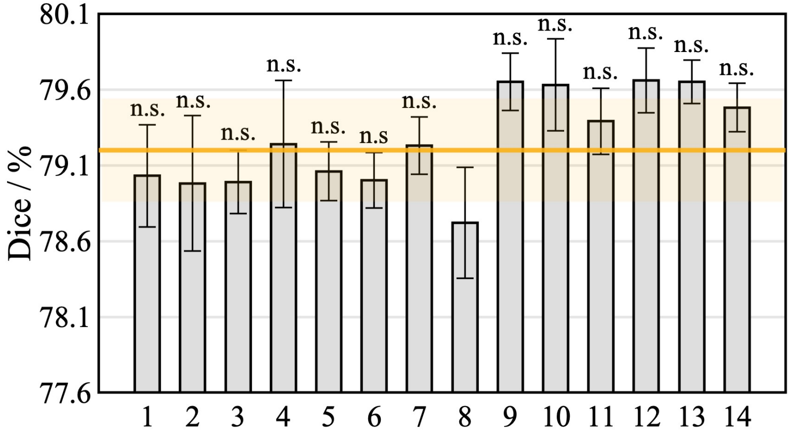
1: InsDis
2: MoCo-v1
3: PCL-v1
4: PIRL
5: PCL-v2
6: SimCLR-v1
7: MoCo-v2
8: SimCLR-v2
9: Sela-v2
10: InfoMin
11: BYOL
12: DeepCluster-v2
13: SwAV
14: Barlow Twins

Baseline (supervised)

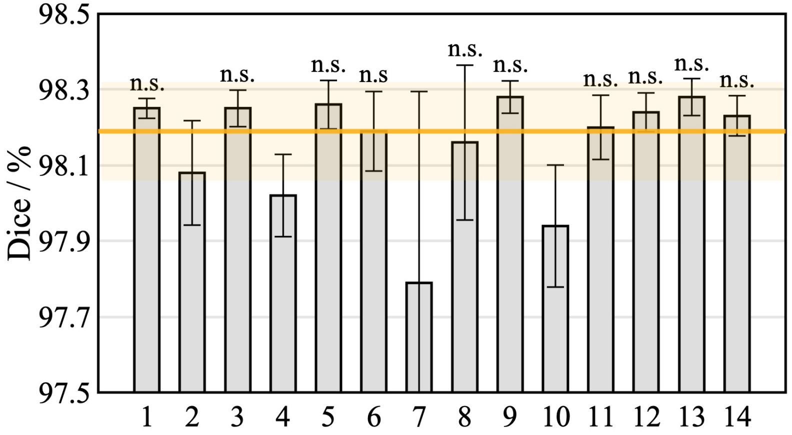
n.s. No Significance
** p<0.01
* p<0.05
>*** p<0.0001



Pneumothorax segmentation
X-ray [SIIM-ACR]

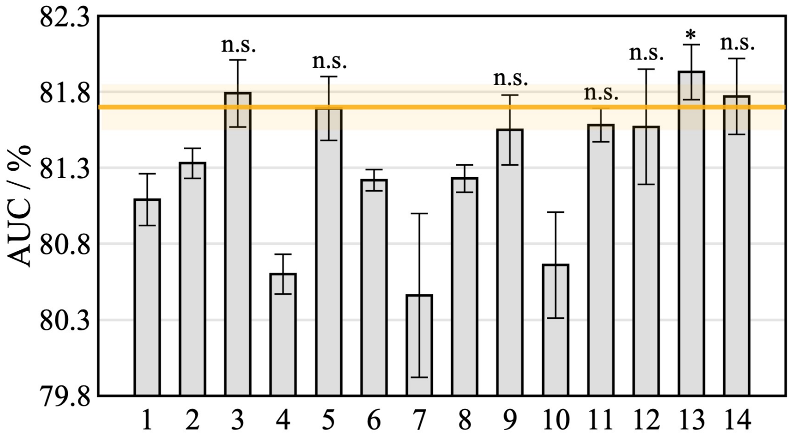


Blood vessels segmentation
Fundoscopic [DRIVE]



Lung segmentation
X-ray [NIH Montgomery]

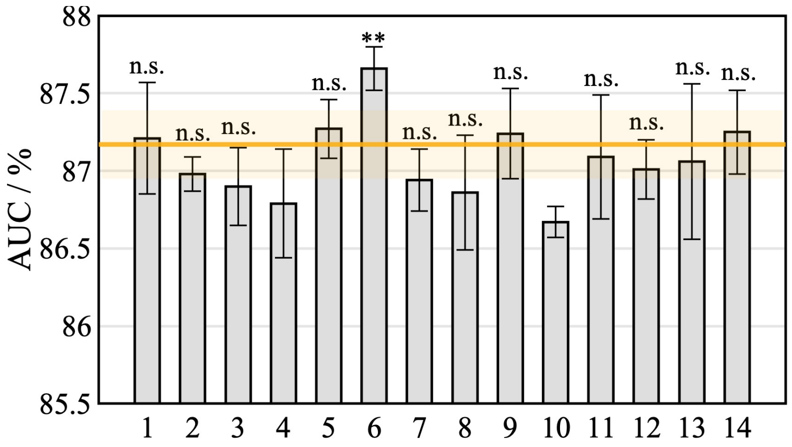
Result II: Top self-supervised ImageNet models outperform supervised ImageNet model



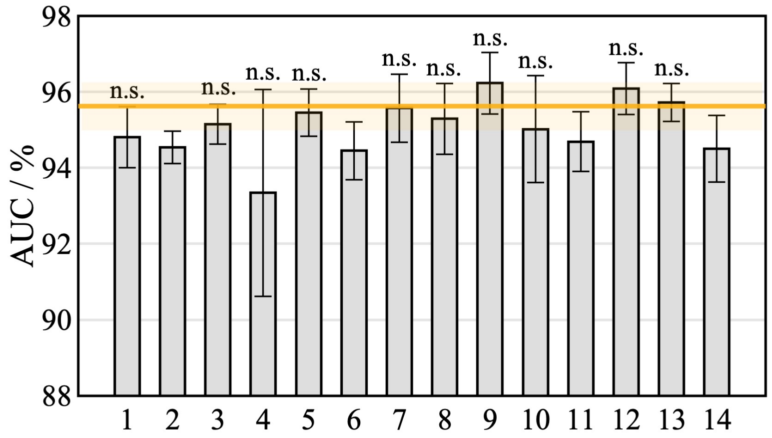
Fourteen thorax diseases classification
X-ray [NIH ChestX-Ray14]

1: InsDis
2: MoCo-v1
3: PCL-v1
4: PIRL
5: PCL-v2
6: SimCLR-v1
7: MoCo-v2
8: SimCLR-v2
9: Sela-v2
10: InfoMin
11: BYOL
12: DeepCluster-v2
13: SwAV
14: Barlow Twins
Baseline (supervised)

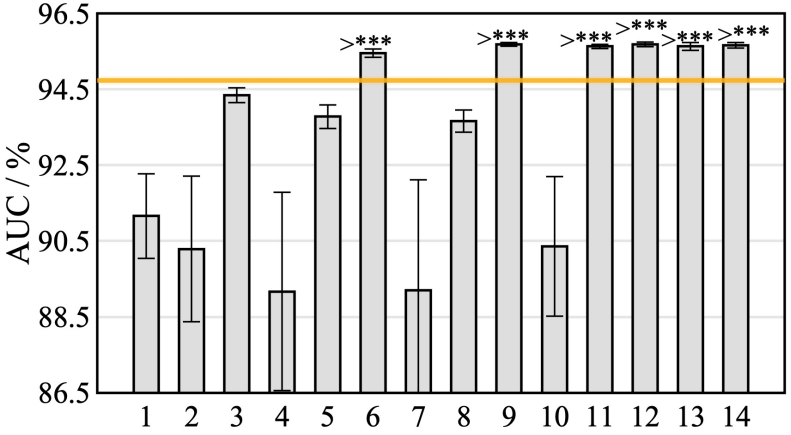
n.s. No Significance ** p<0.01 * p<0.05 >*** p<0.0001



Five thorax diseases classification
X-ray [CheXpert]



Tuberculosis detection
X-ray [NIH Shenzhen CXR]



Pulmonary embolism detection
CT [RSNA PE Detection]

2

How generalizable are the self-supervised ImageNet models to medical imaging in comparison with supervised ImageNet models?

Self-supervised ImageNet models learn holistic features more effectively than supervised ImageNet models.

Domain-adapted models

VS.

Supervised ImageNet models

Natural Images

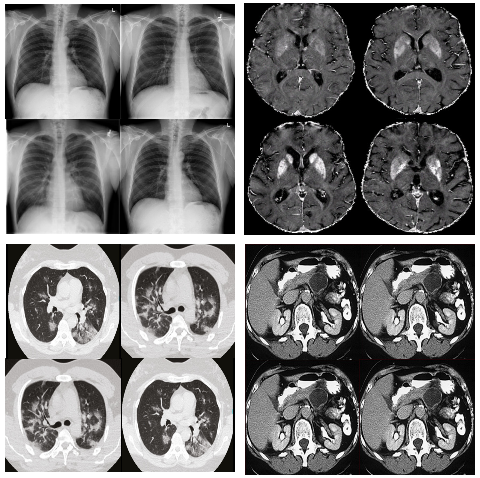
Colorful

Different objects

Lower resolution

Domain shift

Medical Images



Grayscale

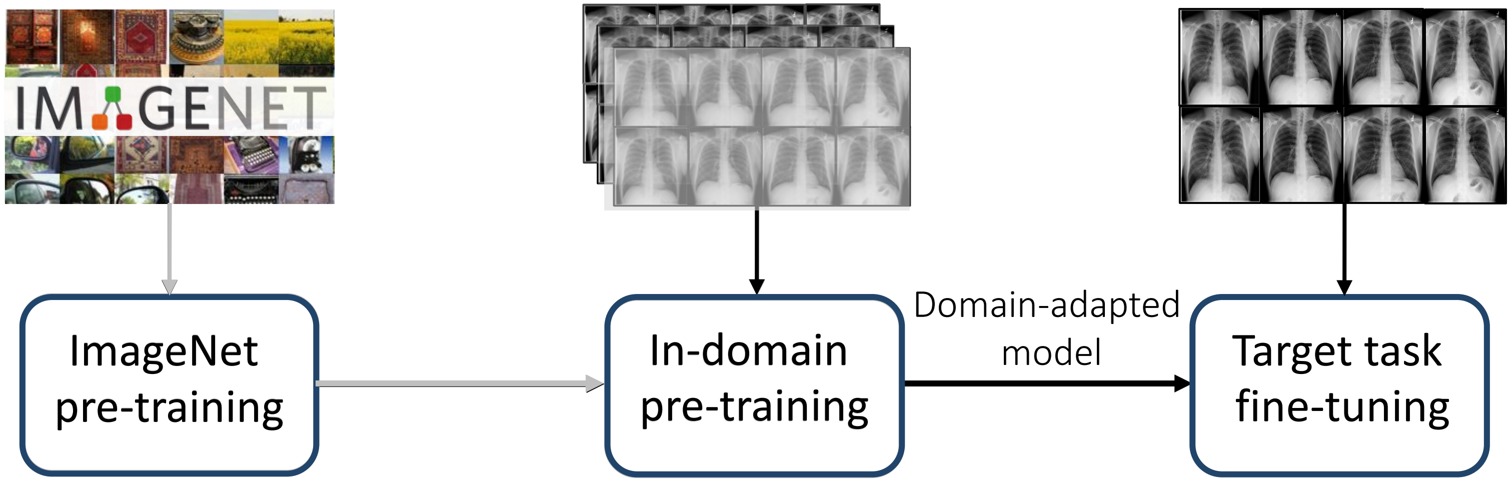
Consistent anatomical structures

Higher resolution

3

Can moderately sized
medical image datasets
help bridge
the domain gap between
natural and medical
images?

Domain-adaptive Pre-training



Result III: Continual pre-training bridges the domain gap between natural and medical images

Task \ Initialization	Scratch	ImageNet	ChestX-ray14	CheXpert	ImageNet→ ChestX-ray14	ImageNet→ CheXpert
14 thorax diseases classification	80.31±0.10	81.70±0.15	---	**81.99±0.08**	---	82.25±0.18
5 thorax diseases classification	86.60±0.17	87.10±0.36	**87.40±0.26**	---	87.09±0.44	---
Tuberculosis detection	89.03±1.82	95.62±0.63	96.32±0.65	**97.07±0.95**	98.47±0.26	97.33±0.26
Pneumothorax segmentation	67.54±0.60	67.93±1.45	68.92±0.98	**69.30±0.50**	69.52±0.38	69.36±0.49
Lung segmentation	97.55±0.36	98.19±0.13	98.18±0.06	**98.25±0.04**	98.27±0.03	98.31±0.05

*When pre-training and target tasks are the same, transfer learning is not applicable, denoted by "---".

Result III: Continual pre-training bridges the domain gap between natural and medical images

Task \ Initialization	Scratch	ImageNet	ChestX-ray14	CheXpert	ImageNet→ ChestX-ray14	ImageNet→ CheXpert
14 thorax diseases classification	80.31±0.10	81.70±0.15	---	81.99±0.08	---	**82.25±0.18**
5 thorax diseases classification	86.60±0.17	87.10±0.36	**87.40±0.26**	---	87.09±0.44	---
Tuberculosis detection	89.03±1.82	95.62±0.63	96.32±0.65	97.07±0.95	**98.47±0.26**	97.33±0.26
Pneumothorax segmentation	67.54±0.60	67.93±1.45	68.92±0.98	69.30±0.50	**69.52±0.38**	69.36±0.49
Lung segmentation	97.55±0.36	98.19±0.13	98.18±0.06	98.25±0.04	98.27±0.03	**98.31±0.05**

*When pre-training and target tasks are the same, transfer learning is not applicable, denoted by "---".

Result III: Continual pre-training bridges the domain gap between natural and medical images

Task \ Initialization	Scratch	ImageNet	ChestX-ray14	CheXpert	ImageNet→ ChestX-ray14	ImageNet→ CheXpert
14 thorax diseases classification	80.31±0.10	81.70±0.15	---	81.99±0.08	---	**82.25±0.18**
5 thorax diseases classification	86.60±0.17	87.10±0.36	**87.40±0.26**	---	87.09±0.44	---
Tuberculosis detection	89.03±1.82	95.62±0.63	96.32±0.65	97.07±0.95	**98.47±0.26**	97.33±0.26
Pneumothorax segmentation	67.54±0.60	67.93±1.45	68.92±0.98	69.30±0.50	**69.52±0.38**	69.36±0.49
Lung segmentation	97.55±0.36	98.19±0.13	98.18±0.06	98.25±0.04	98.27±0.03	**98.31±0.05**

*When pre-training and target tasks are the same, transfer learning is not applicable, denoted by “---”.

*The best methods are bolded while the others are highlighted in blue if they achieve equivalent performance compared with the best one (i.e., p > 0.05).

3

Can moderately sized medical image datasets help bridge the domain gap between natural and medical images?

Continual pre-training bridges the domain gap between natural and medical images.

Benchmarking Transfer Learning for Medical Imaging

- ✓ What truly matters for the segmentation tasks is fine-grained representation
- ✓ Self-supervised ImageNet models learn holistic features more effectively than supervised ImageNet models
- ✓ Continual pre-training can bridge the domain gap between natural and medical images

A Systematic Benchmarking Analysis of Transfer Learning for Medical Image Analysis

Try it for yourself

Code, data, and models
are available online

github.com/JLiangLab/BenchmarkTransferLearning

References

For figures in slide 2, credit to:

1. Wang, X., et al.; ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2097-2106.
2. Esteva, A., et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
3. Papadopoulos, A., et al. An interpretable multiple-instance approach for the detection of referable diabetic retinopathy in fundus images. Sci Rep 11, 14326 (2021).

For figures in slide 6, credit to:

1. Zhao, Q., et al; Pneumothorax Detection and Localization in X-Ray Images Given Richer Annotation Information. SIIM Conference on Machine Intelligence in Medical Imaging. 2018.
2. Zhan, T. et al. Brain Tumor Segmentation in Multi-modality MRIs Using Multiple Classifier System and Spatial Constraint. 2015 3rd International Conference on Computer, Information and Application (2015): 18-21.
3. Santini,G.; AI in Medical Imaging: The Kidney Tumor Segmentation Challenge. 2019
4. Tavakoli, M., et al. Unsupervised automated retinal vessel segmentation based on Radon line detector and morphological reconstruction. IET Image Processing 15 (7), pp. 1484-1498, 2021.