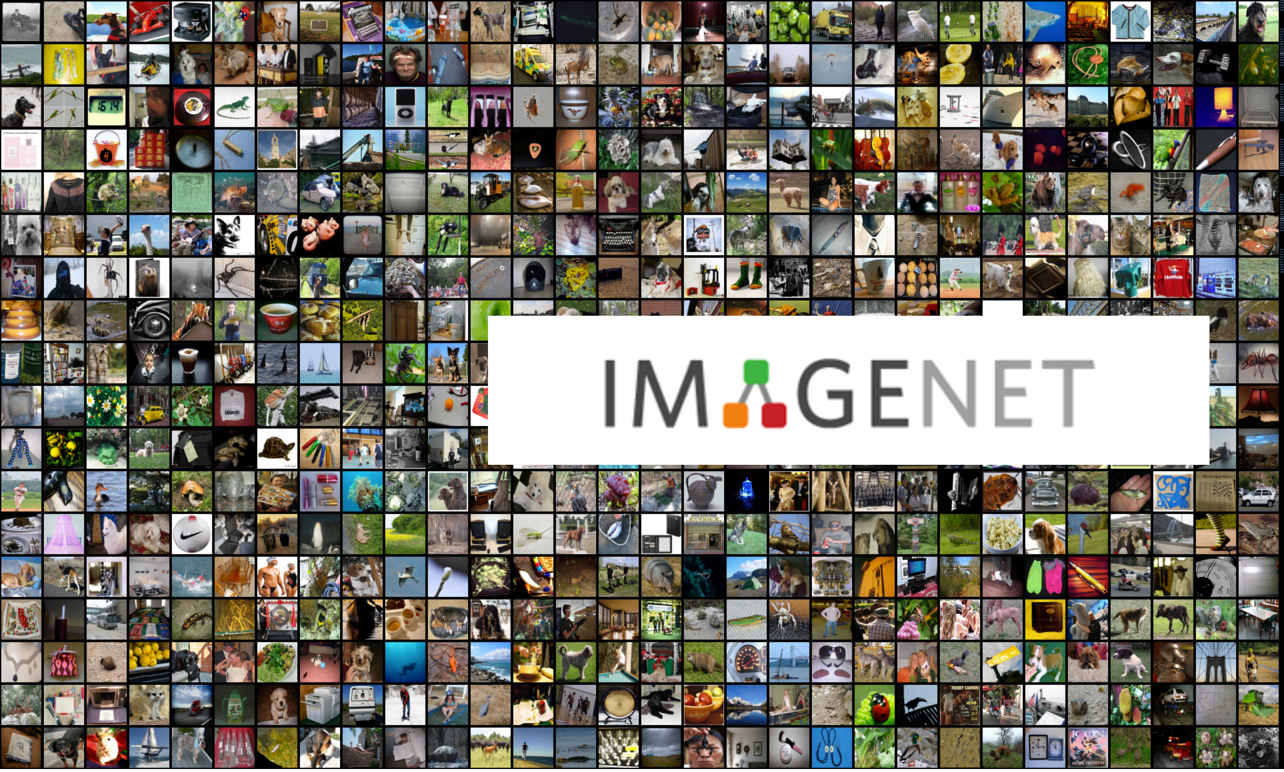
We provide pre-trained 3D models!

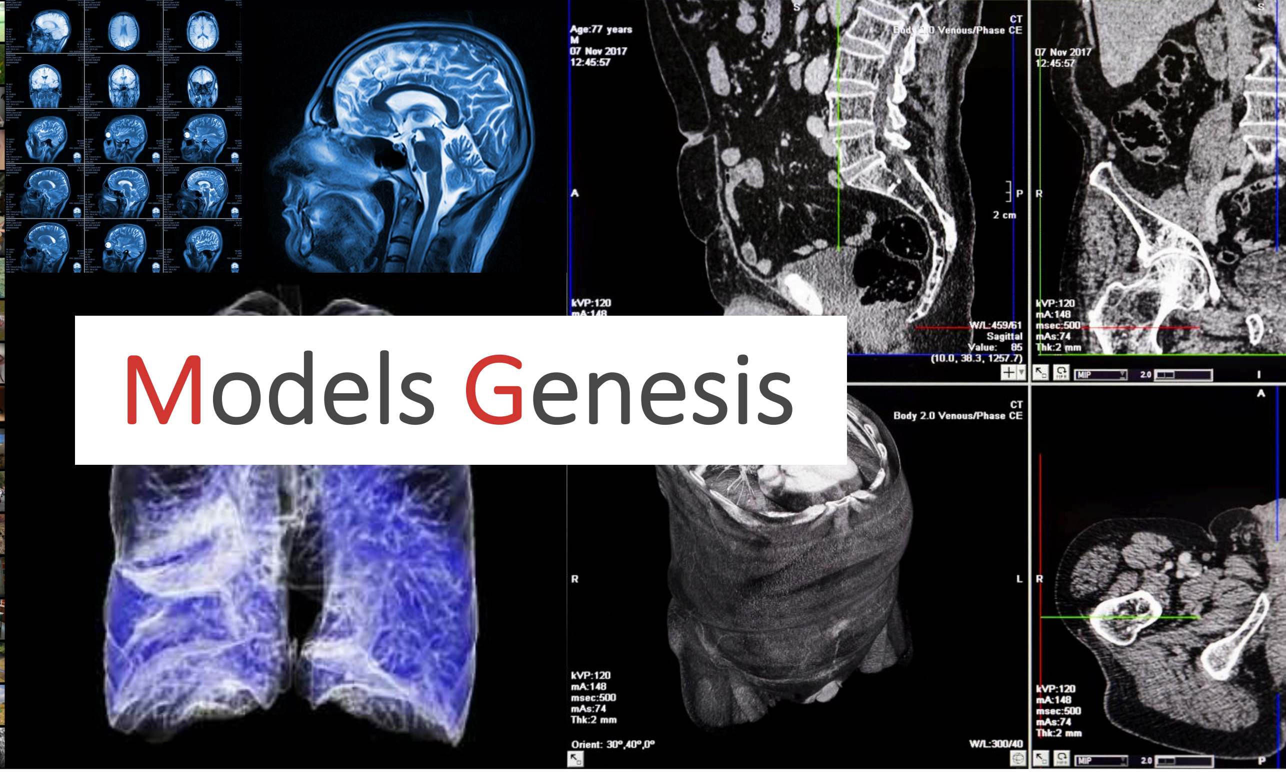
# Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis

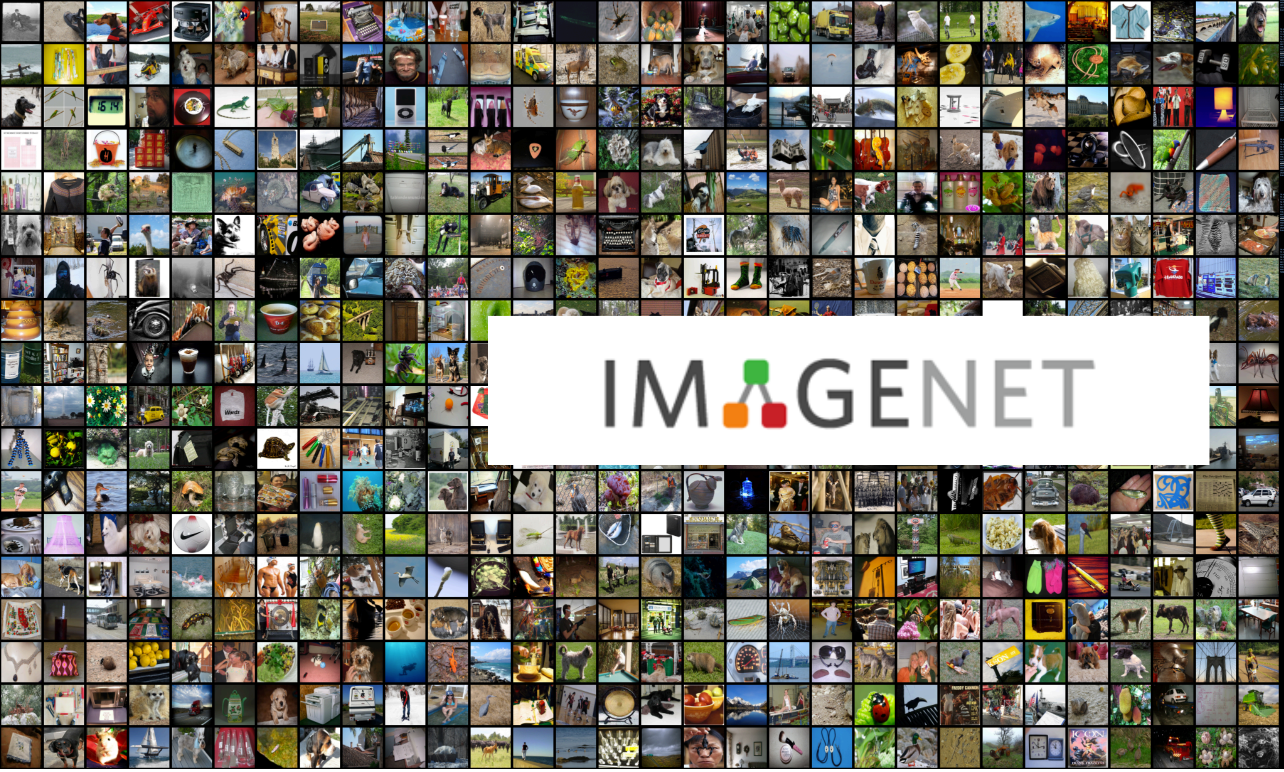
MICCAI 2019 Young Scientist Award

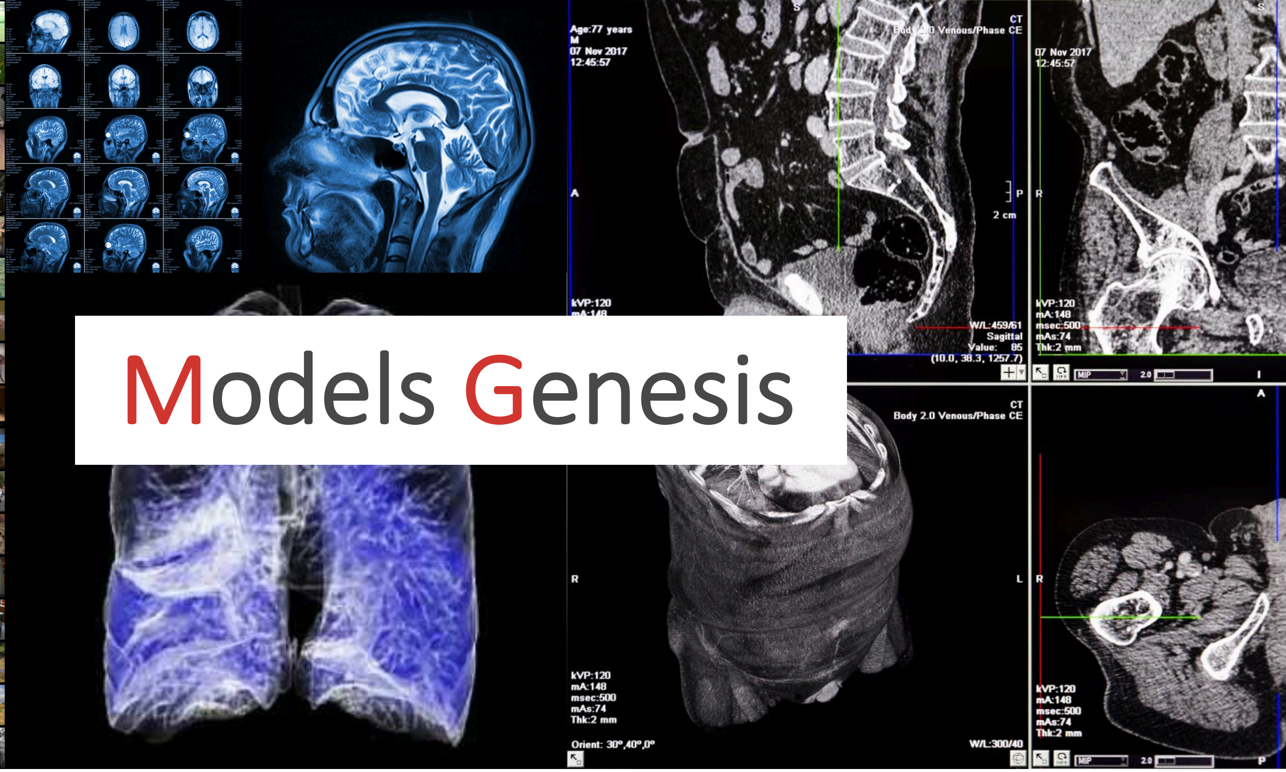
**Zongwei Zhou**[1], Vatsal Sodha[1], Md Mahfuzur Rahman Siddiquee[1], Ruibin Feng[1], Nima Tajbakhsh[1], Michael B. Gotway[2], and Jianming Liang[1]

[1] Arizona State University [2] Mayo Clinic


IMAGENET


Models Genesis

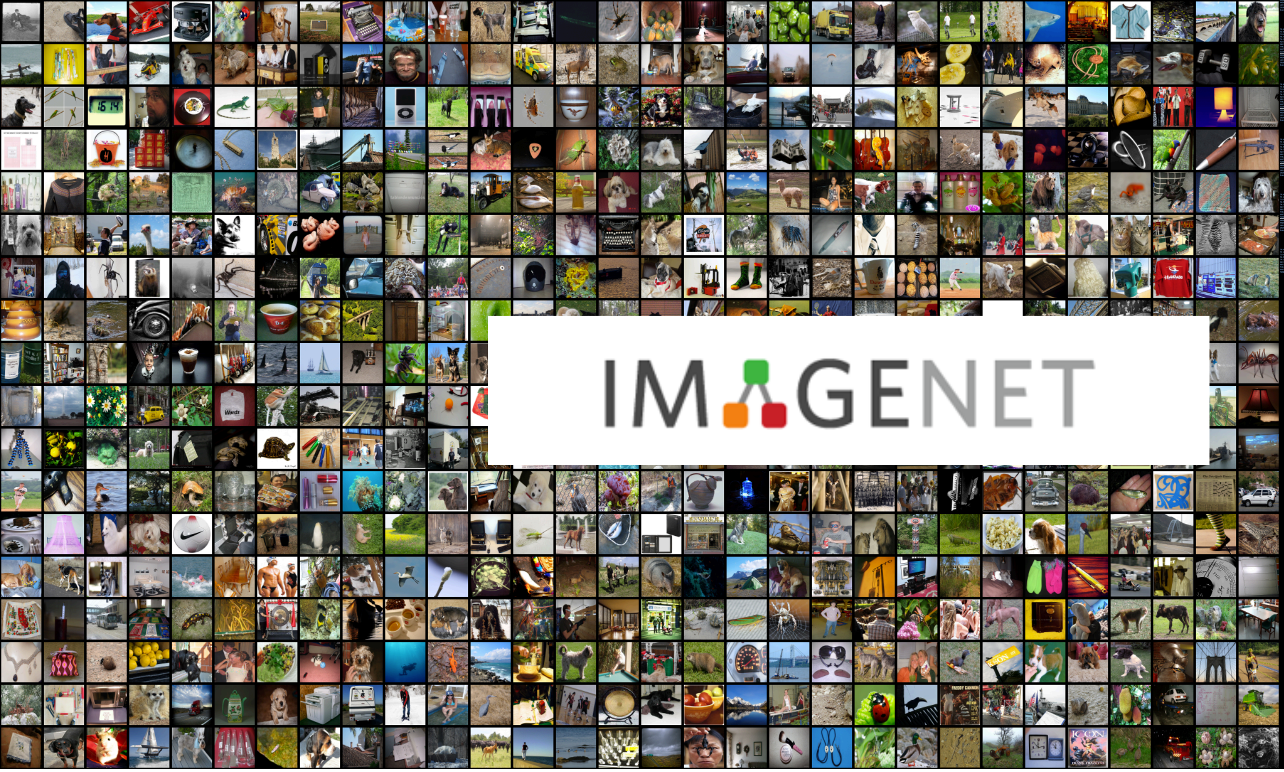


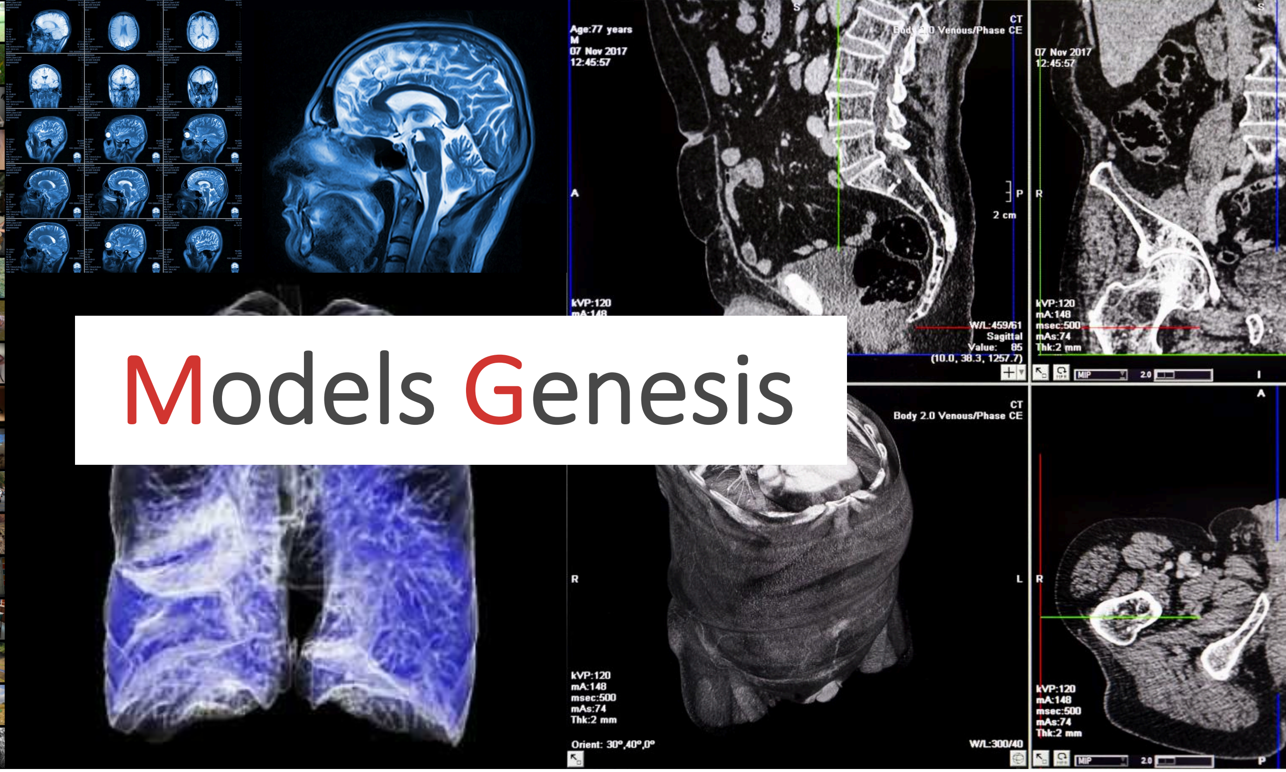




Natural images

Medical images

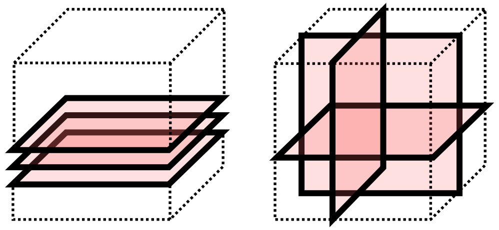
natural → medical < medical → medical



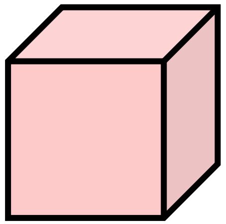



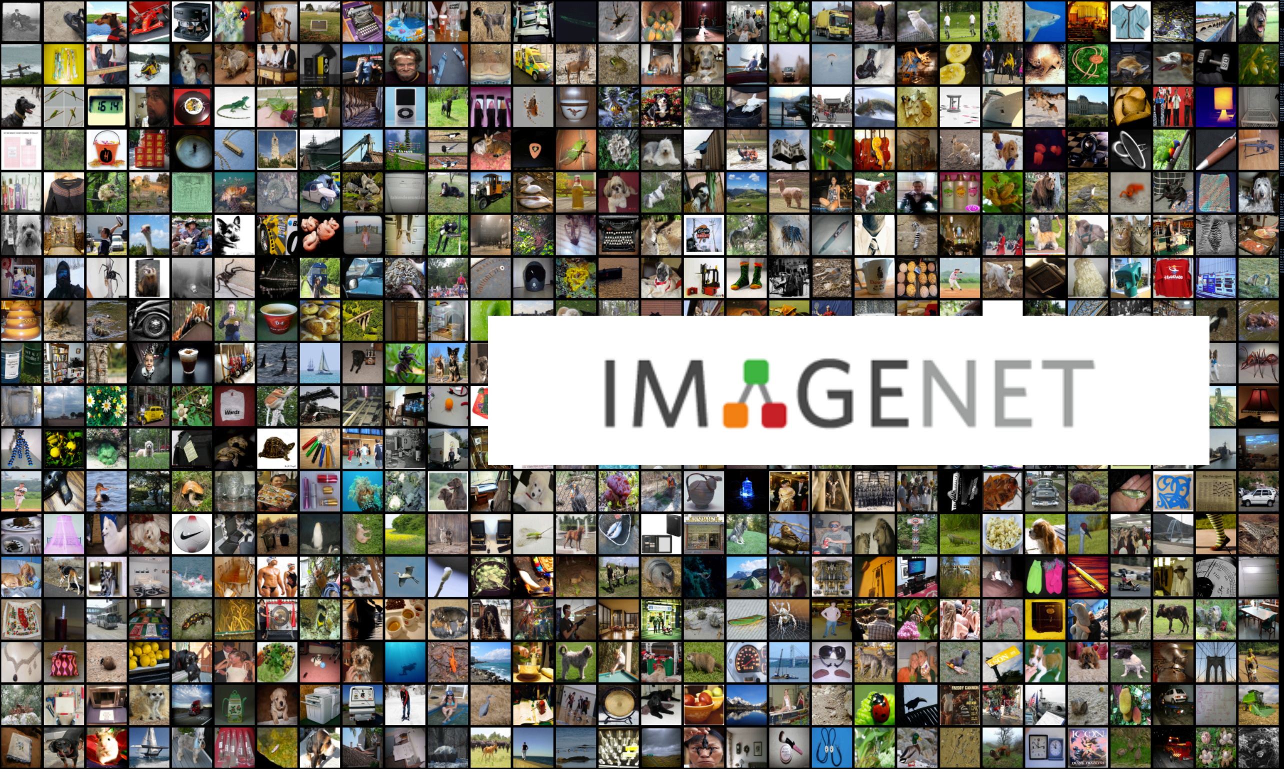

Natural images | Medical images

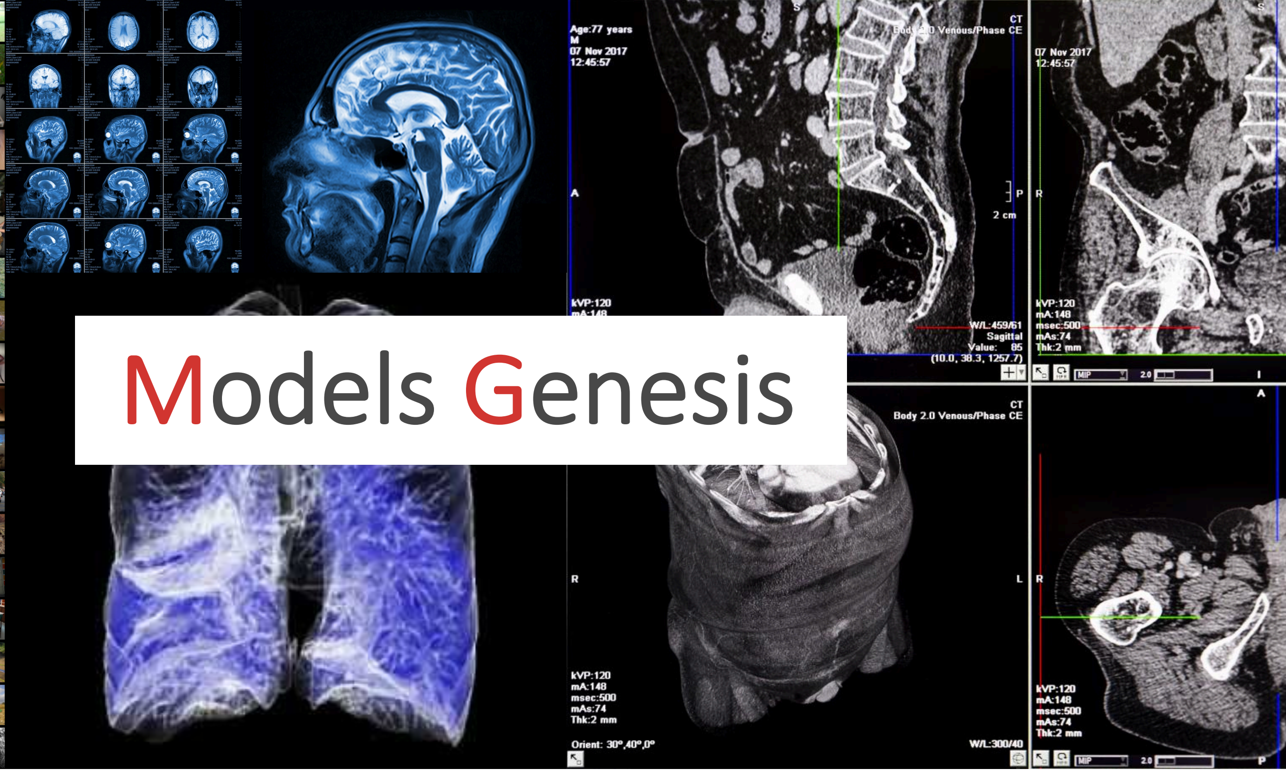
Formed in 2D | Formed in 3D



**3D imaging tasks should be solved in 3D**



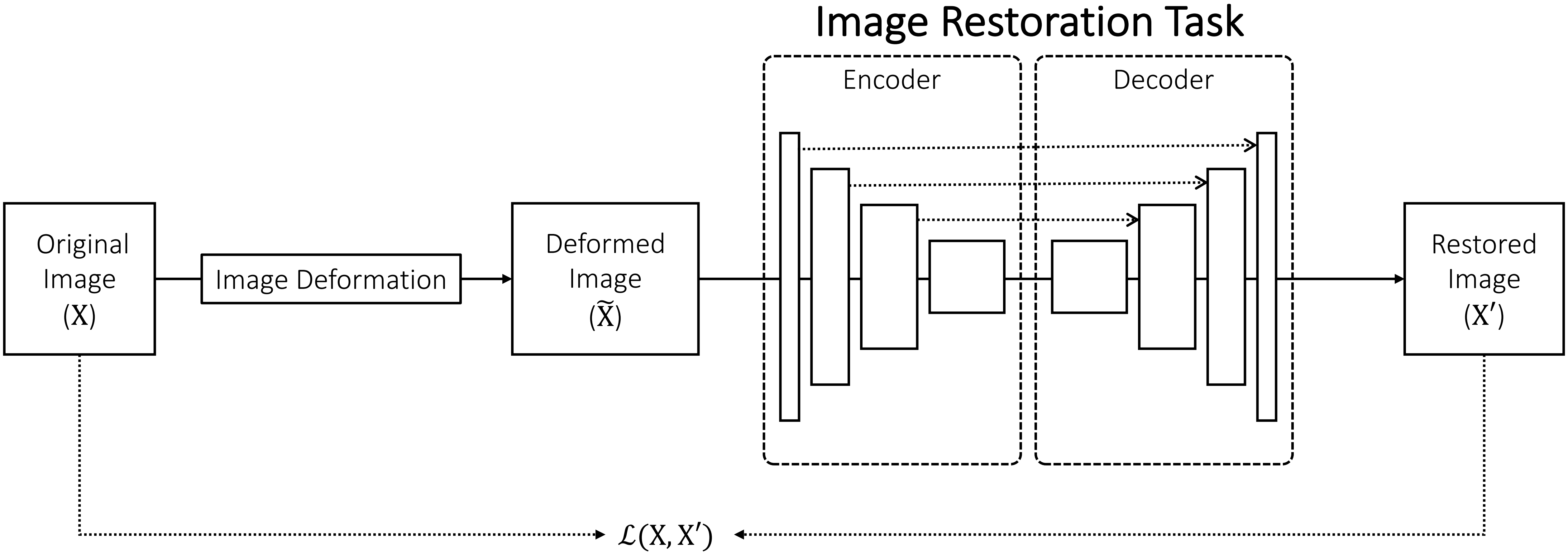


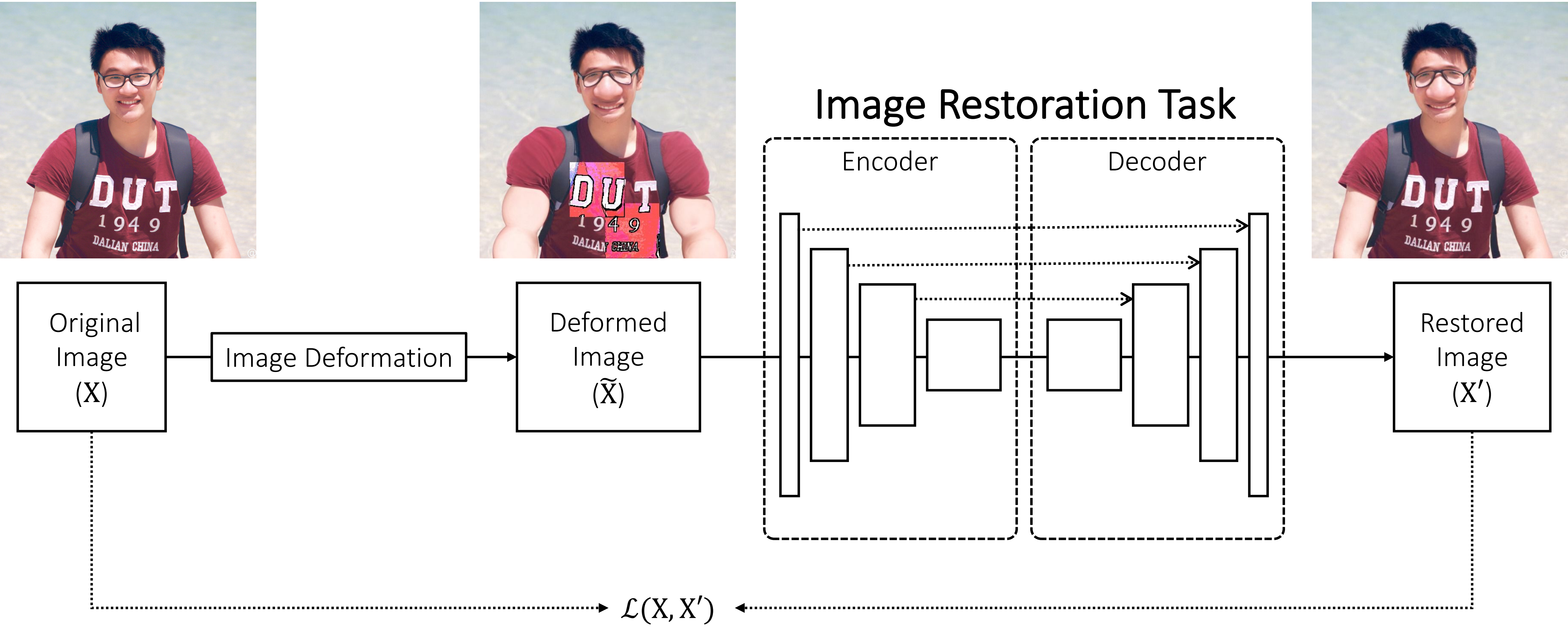

| | |
|---|---|
| Natural images | Medical images |
| Formed in 2D | Formed in 3D |
| >14,000,000 annotation | Zero annotation |

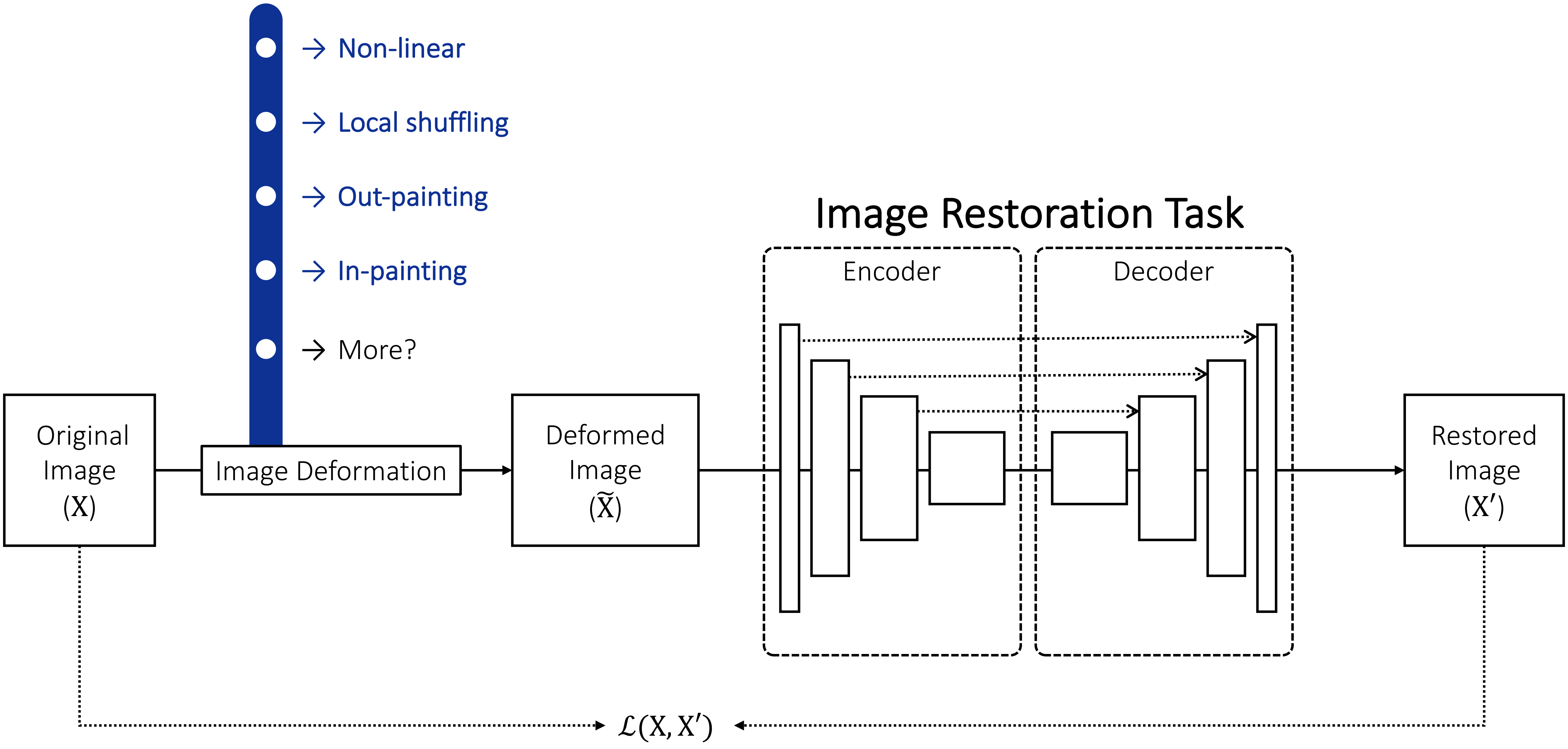
**ImageNet demands huge amount of annotation efforts, but Models Genesis are pre-trained with self-supervision.**

We design it as a simple image restoration task, through which, the model can learn representation directly from image data itself.



We design it as a simple image restoration task, through which, the model can learn representation directly from image data itself.





→ Non-linear
→ Local shuffling
→ Out-painting
→ In-painting
→ More?
Image Restoration Task
Encoder
Decoder
Original Image (X)
Image Deformation
Deformed Image ($\tilde{X}$)
Restored Image (X′)
$\mathcal{L}(X, X')$

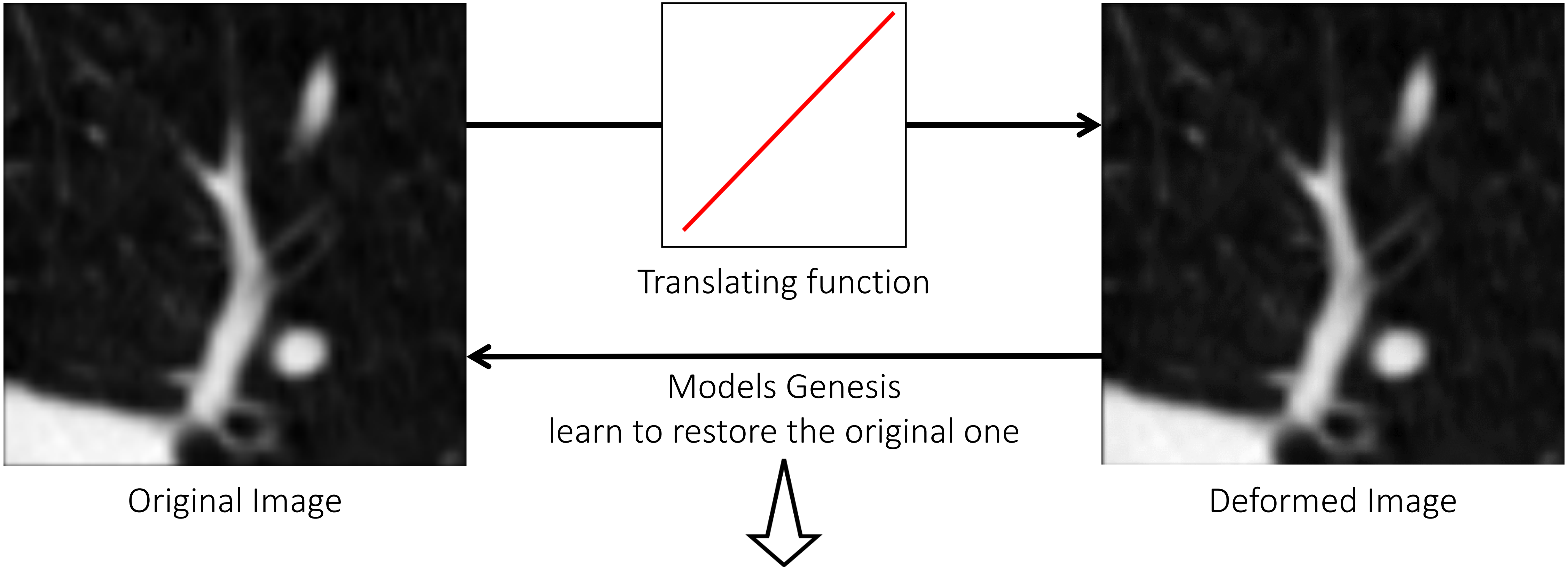
# I. Non-linear transformation

**CT scan itself naturally comes with the *pixel-wise* annotation**

| Substance | | Hounsfield units (HU) |
|---|---|---|
| Air | | -1000 |
| Fat | | -120 to -90 |
| Water | | 0 |
| Bone | Cancellous | +300 to +400 |
| | Cortical | +1800 to +1900 |
| Parenchyma | Lung | -700 to −600 |
| | Kidney | +20 to +45 |
| | Liver | +54 to +66 |
| | Lymph nodes | +10 to +20 |
| | Muscle | +35 to +55 |

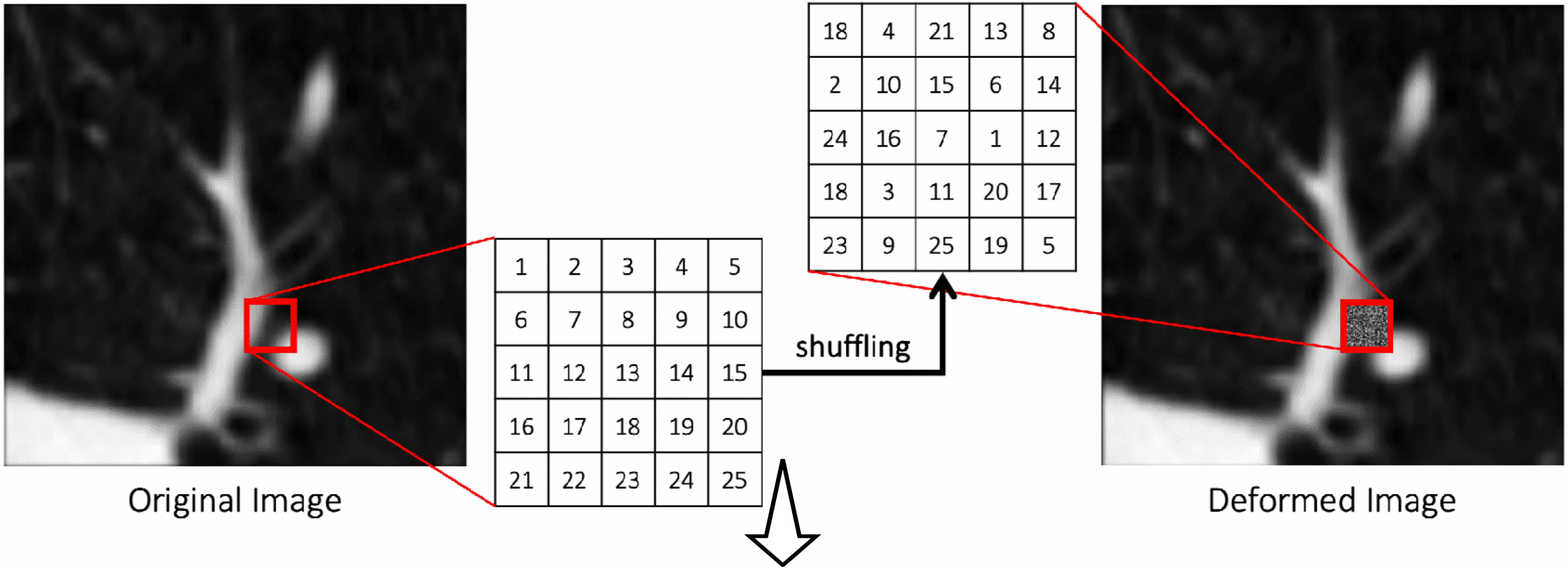
Source from en.wikipedia.org/wiki/Hounsfield_scale

# I. Non-linear transformation




**Learning organ appearance and intensity distribution**

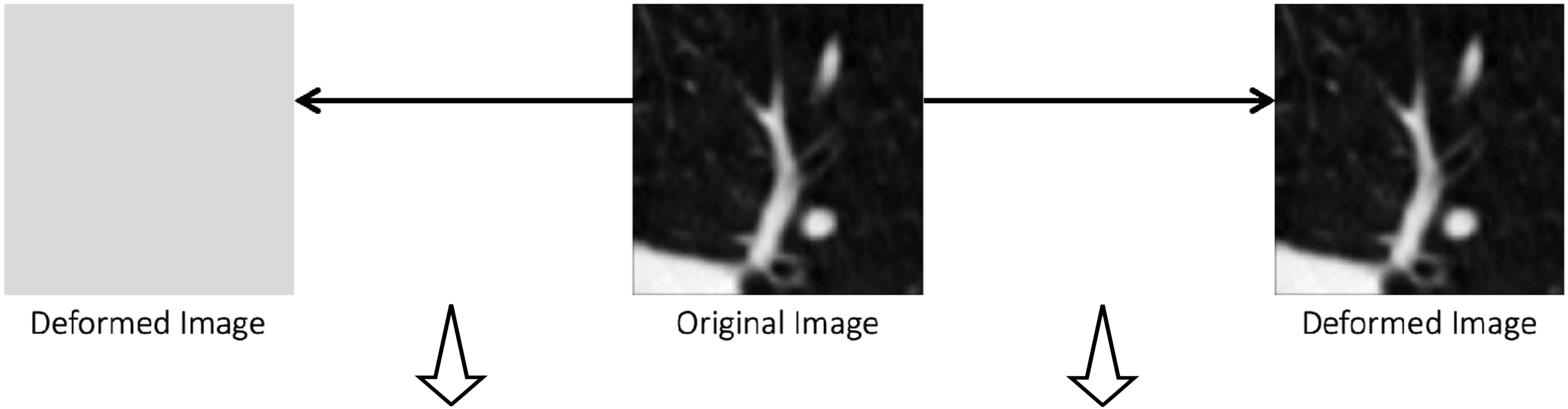
# II. Local pixel shuffling




Learning organ texture and local boundaries

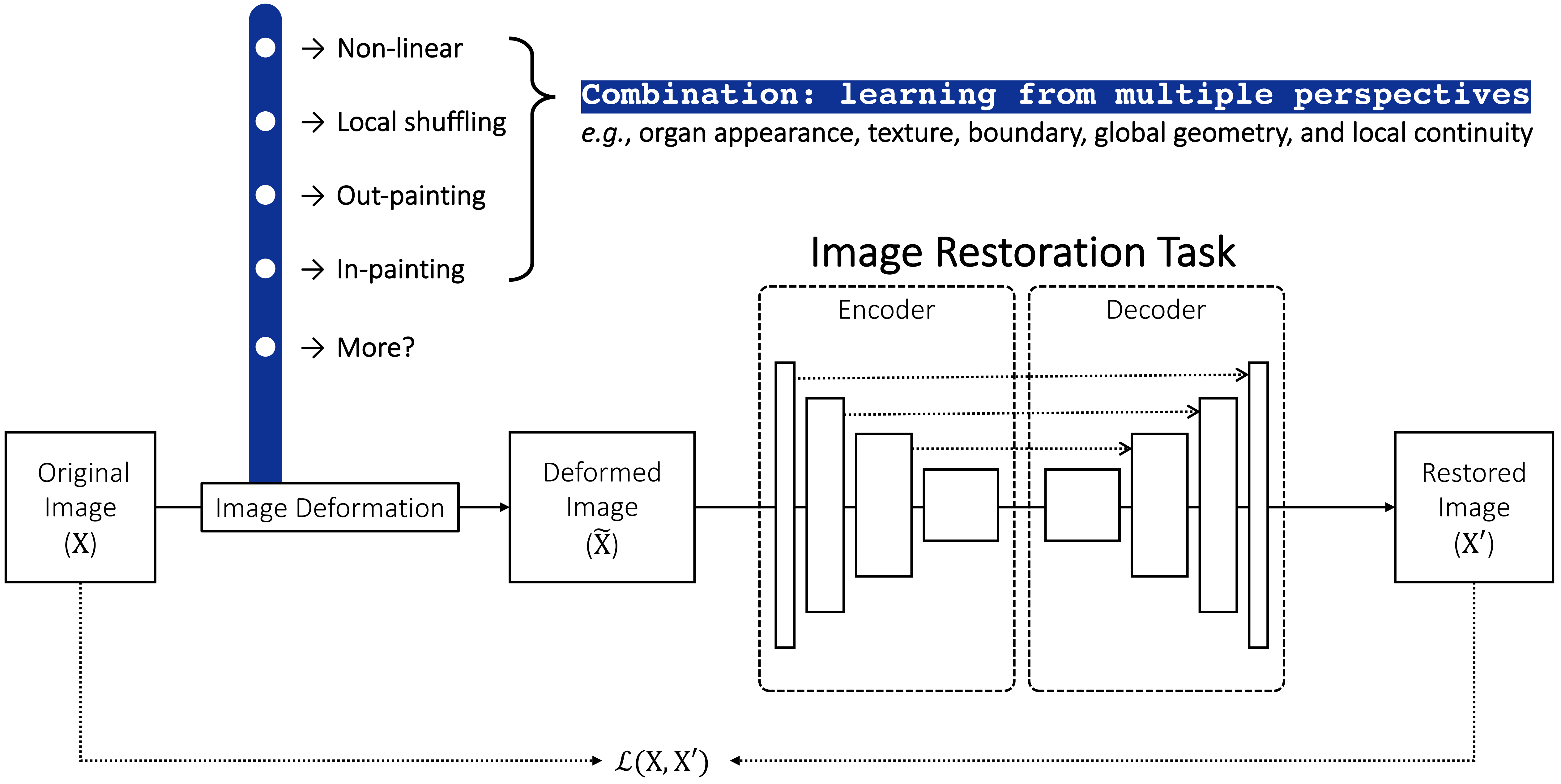
## III. Out-painting

## IV. In-painting

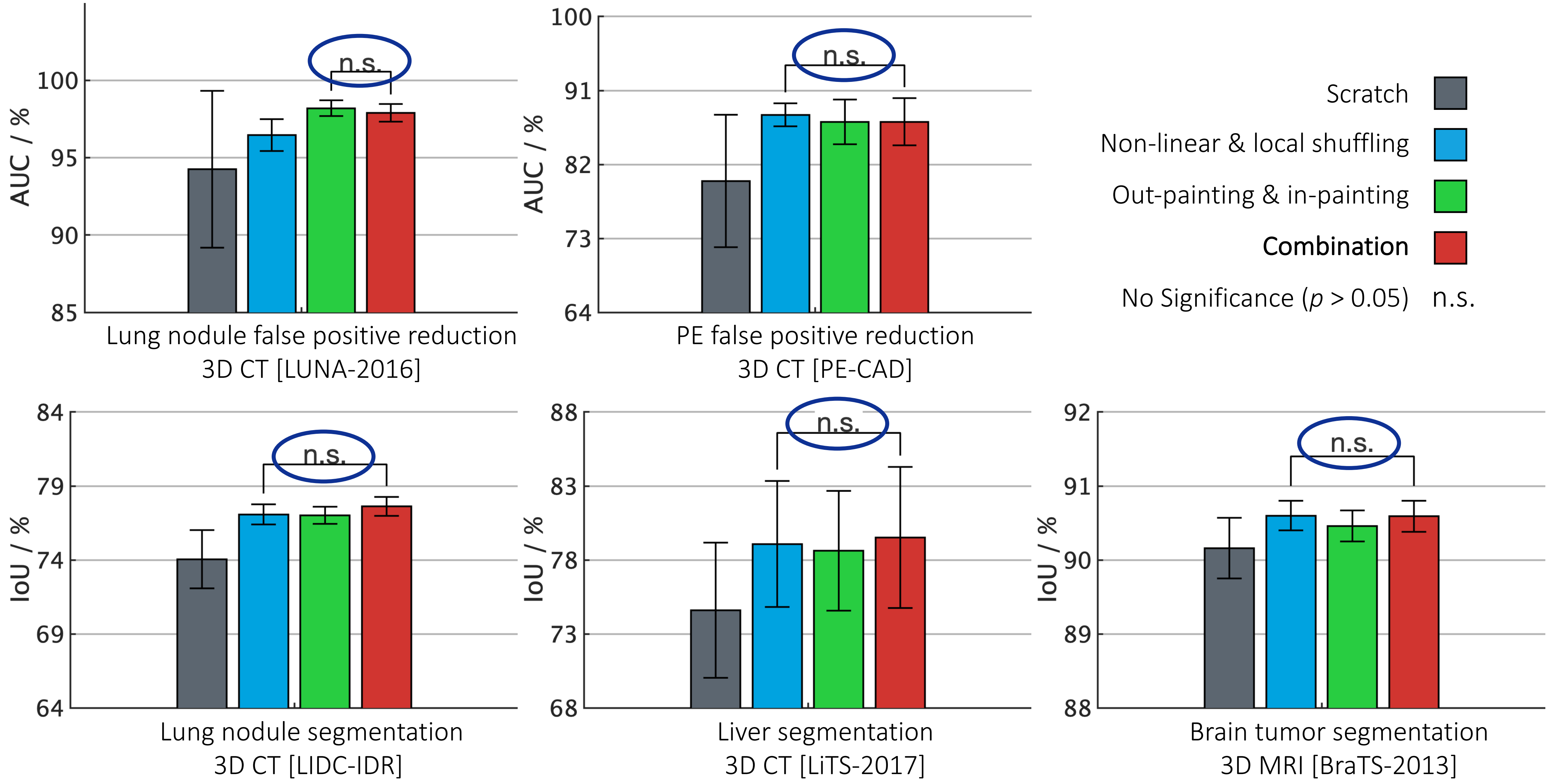



Learning organ spatial layout and global geometry

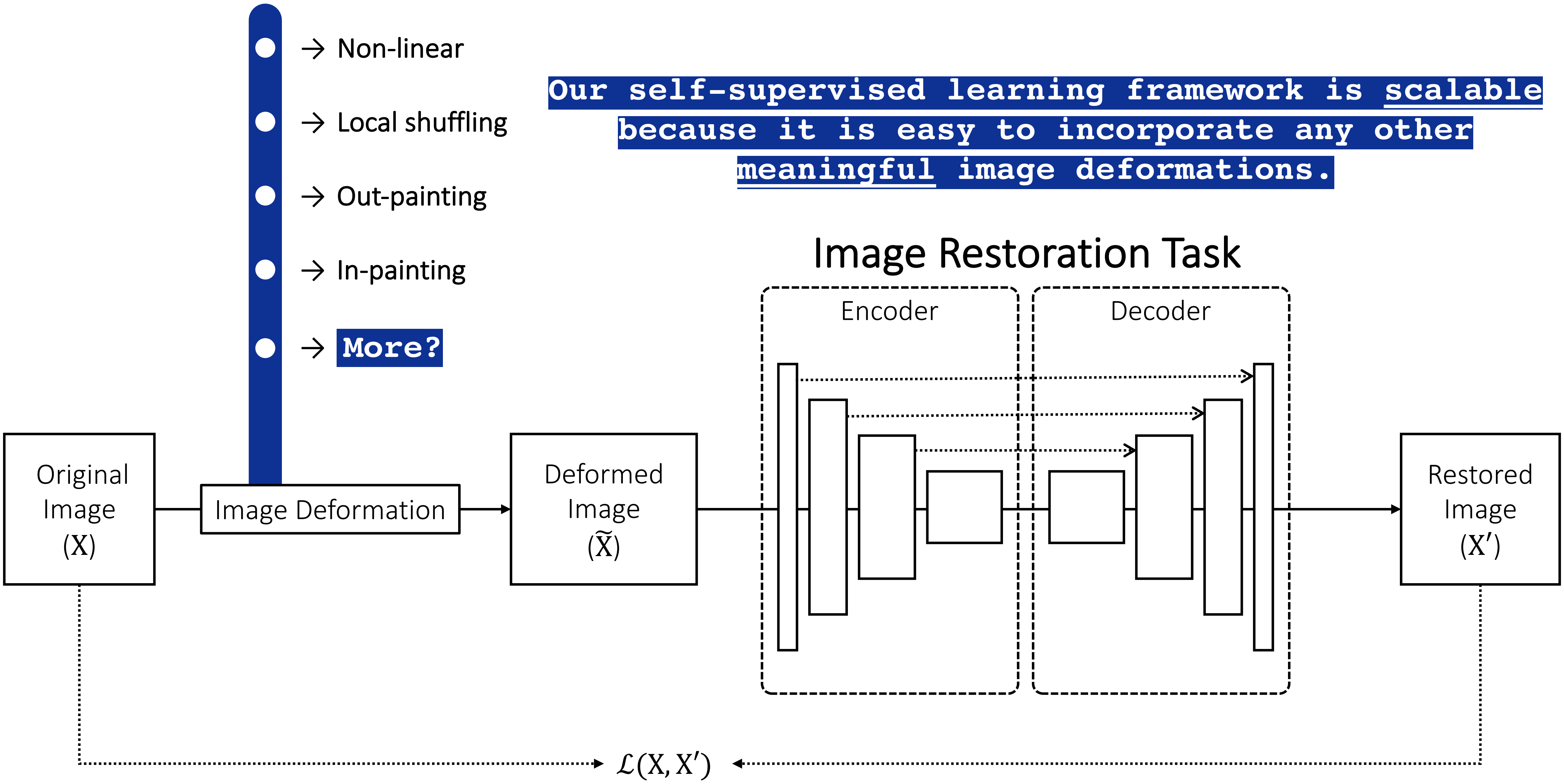
Learning local continuities


→ Non-linear
→ Local shuffling
→ Out-painting
→ In-painting
→ More?
Combination: learning from multiple perspectives
*e.g.*, organ appearance, texture, boundary, global geometry, and local continuity
Image Restoration Task
Encoder
Decoder
Original Image (X)
Image Deformation
Deformed Image ($\tilde{X}$)
Restored Image (X′)
$\mathcal{L}(X, X')$

# Learning from multiple perspectives leads to more `robust` models across all target tasks

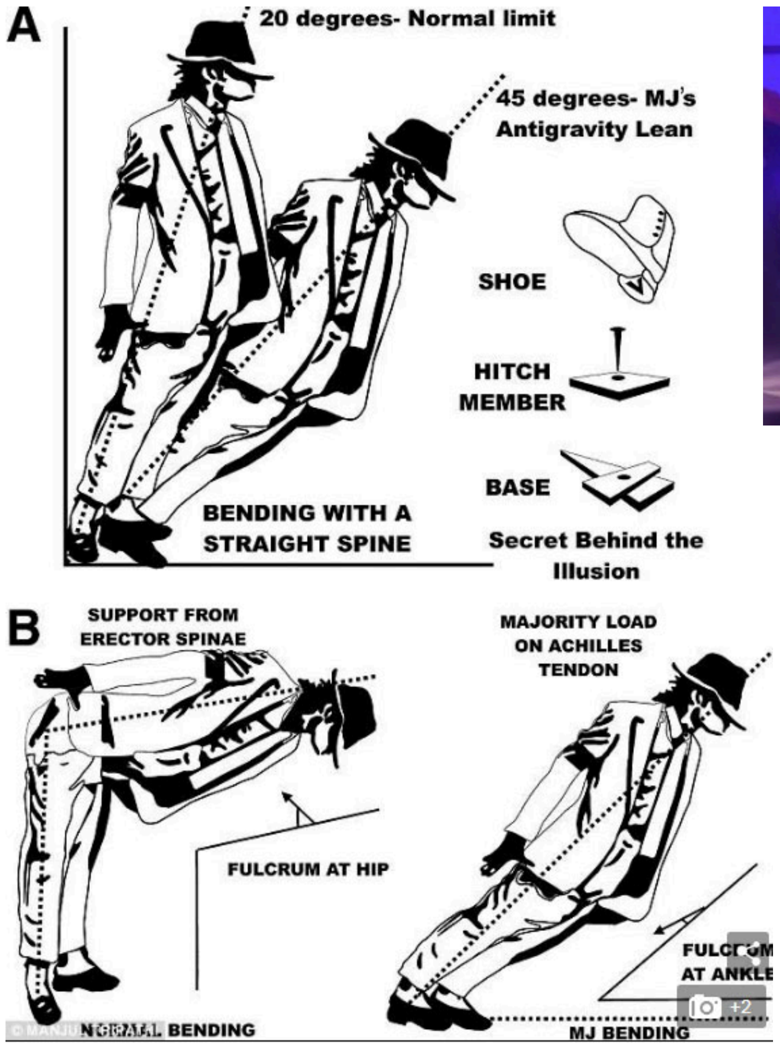


**Our self-supervised learning framework is scalable because it is easy to incorporate any other meaningful image deformations.**

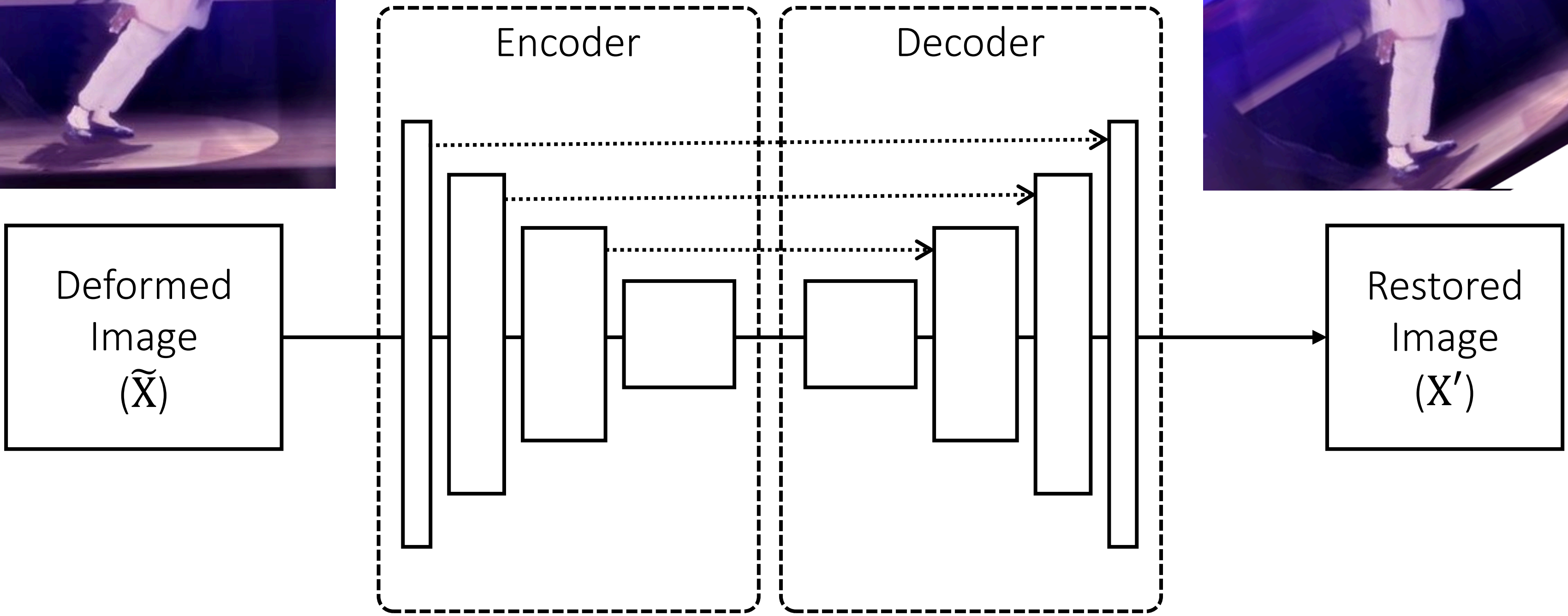


To restore the original (usual) image,
the model must first notice the unusual from the inputs,
so we innovate a collection of deformations.

FYI ☺



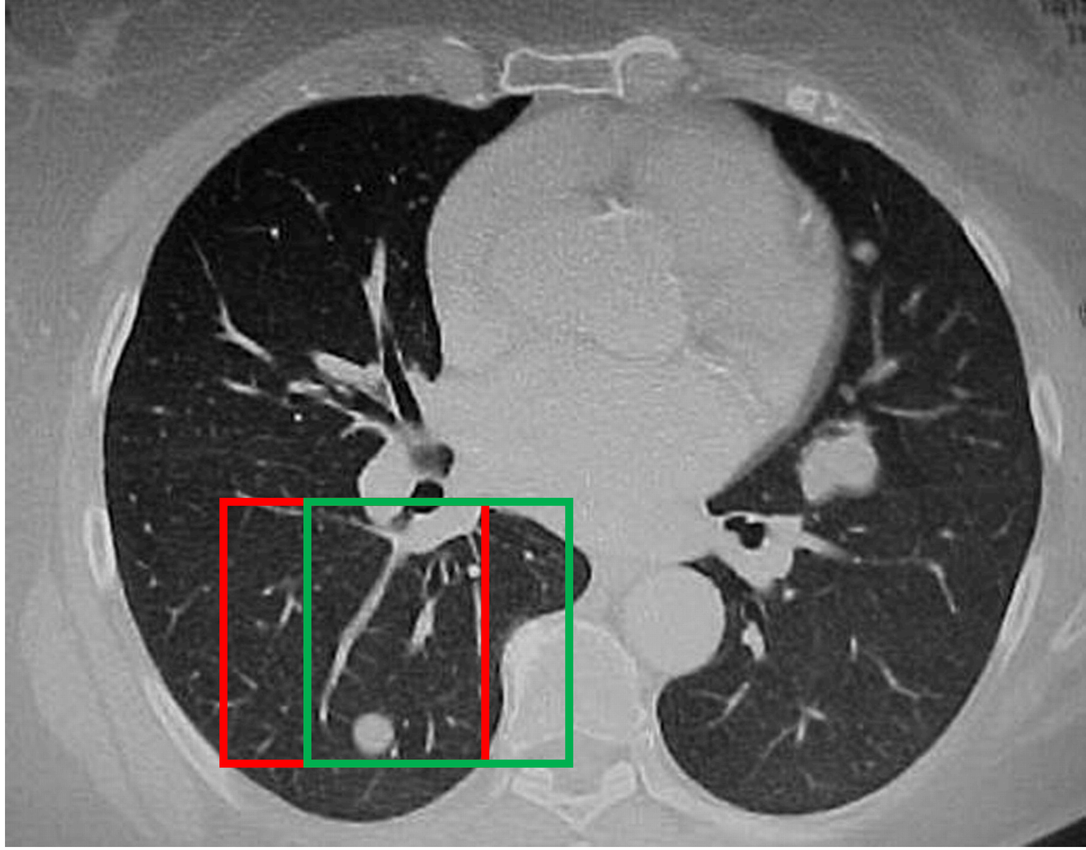

## Image Restoration Task

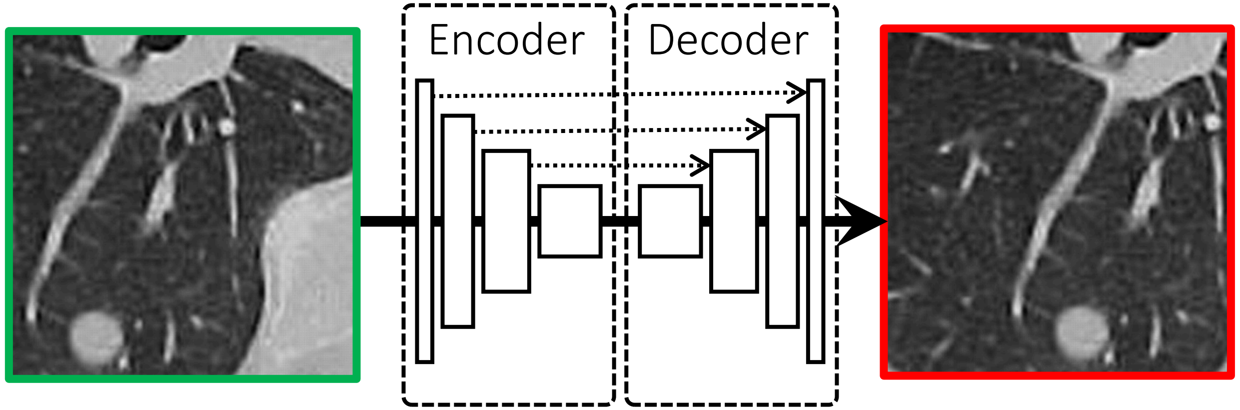



# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

| Example | Image deformation | Data augmentation |
|---|---|---|
| Translation | ✗ | ✓ |

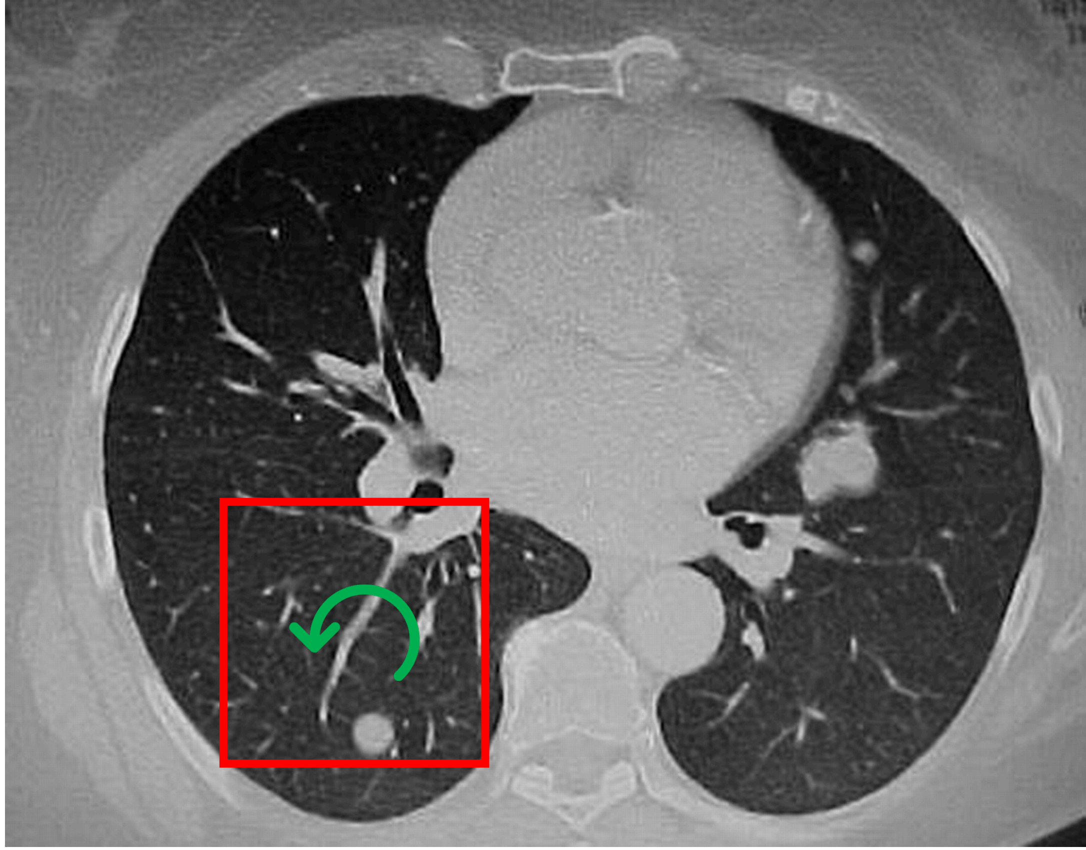


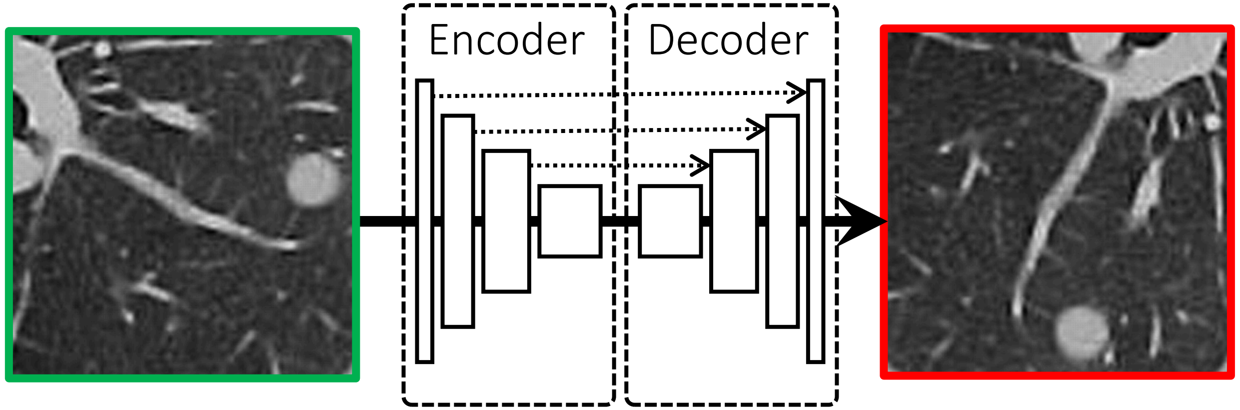


# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

| Example | Image deformation | Data augmentation |
| --- | --- | --- |
| Translation | ✗ | ✓ |
| Rotation | ✗ | ✓ |

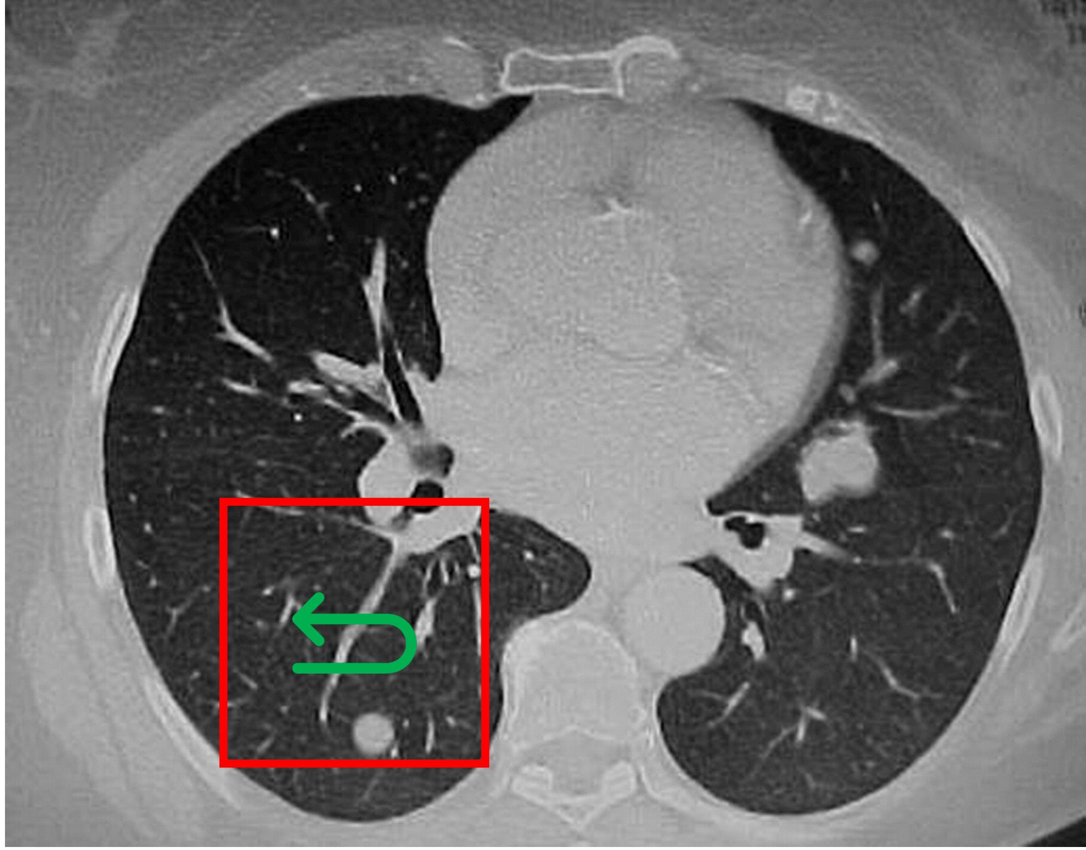


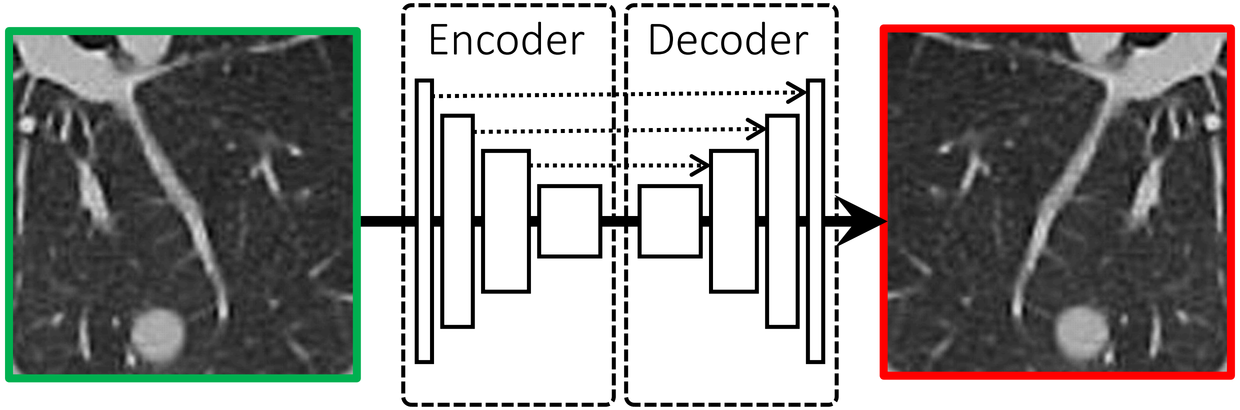


# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

| Example | Image deformation | Data augmentation |
|---|---|---|
| Translation | ✗ | ✓ |
| Rotation | ✗ | ✓ |
| Flipping | ✗ | ✓ |

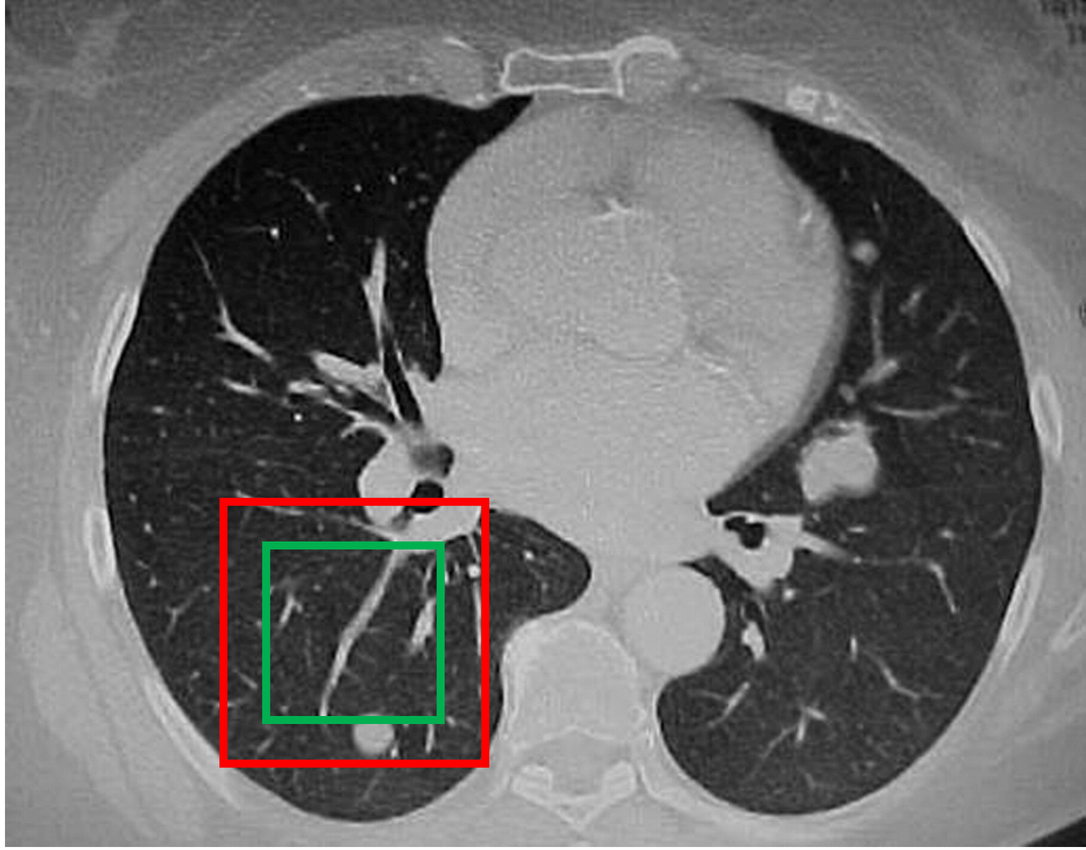


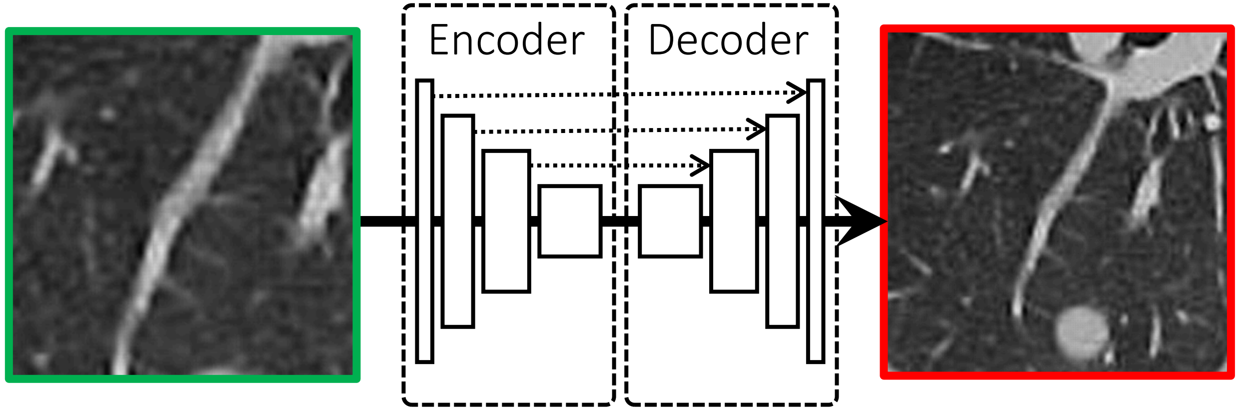


# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

| Example | Image deformation | Data augmentation |
|---|---|---|
| Translation | ✗ | ✓ |
| Rotation | ✗ | ✓ |
| Flipping | ✗ | ✓ |
| Scaling | ✗ | ✓ |

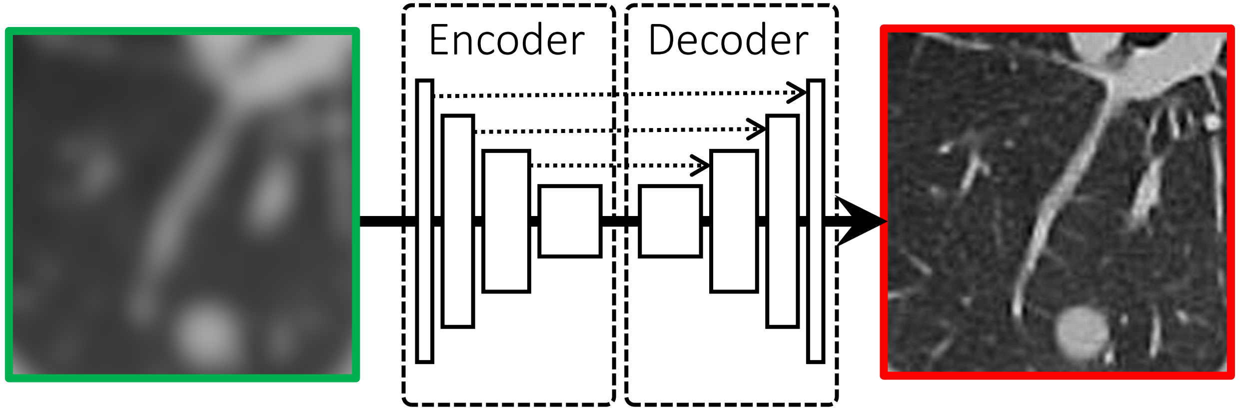




# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

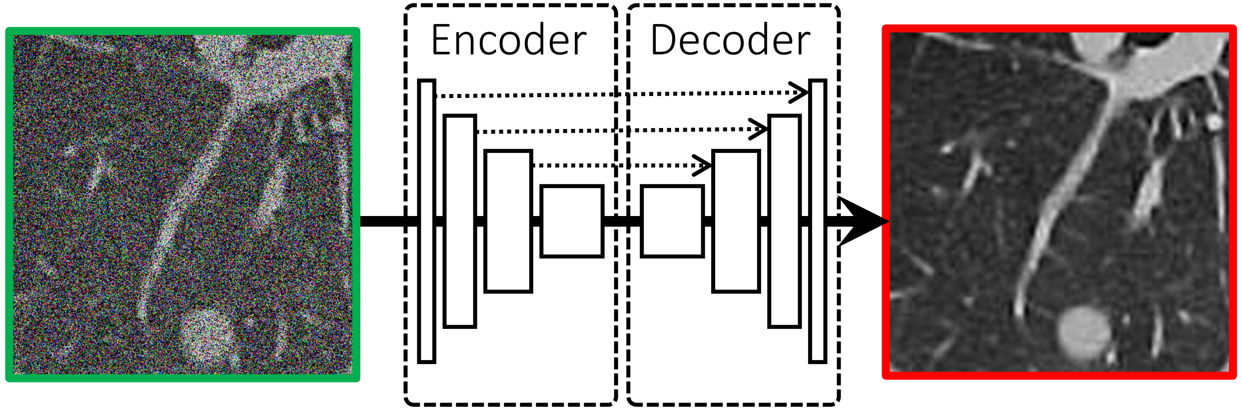
| Example | Image deformation | Data augmentation |
|---|---|---|
| Translation | ✗ | ✓ |
| Rotation | ✗ | ✓ |
| Flipping | ✗ | ✓ |
| Scaling | ✗ | ✓ |
| Blur | ✓ | ✓ |



# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

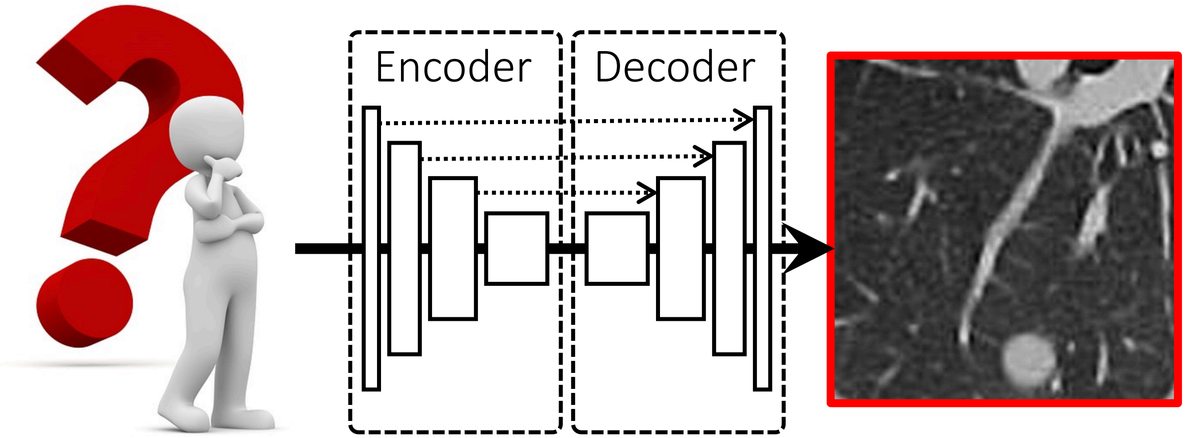
| Example | Image deformation | Data augmentation |
|---|---|---|
| Translation | ✗ | ✓ |
| Rotation | ✗ | ✓ |
| Flipping | ✗ | ✓ |
| Scaling | ✗ | ✓ |
| Blur | ✓ | ✓ |
| Noise | ✓ | ✓ |



# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

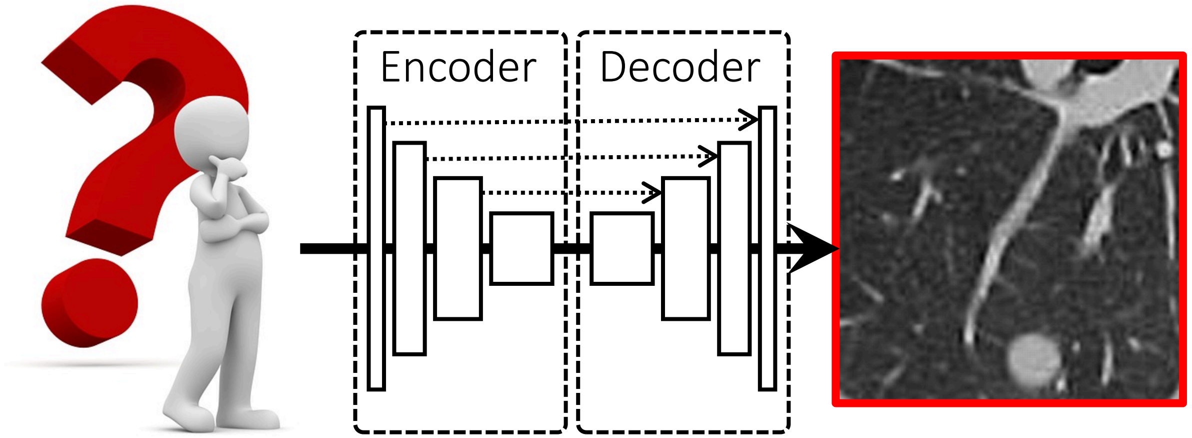
| Example | Image deformation | Data augmentation |
| --- | --- | --- |
| Translation | ✗ | ✓ |
| Rotation | ✗ | ✓ |
| Flipping | ✗ | ✓ |
| Scaling | ✗ | ✓ |
| Blur | ✓ | ✓ |
| Noise | ✓ | ✓ |
| … | ? | ? |

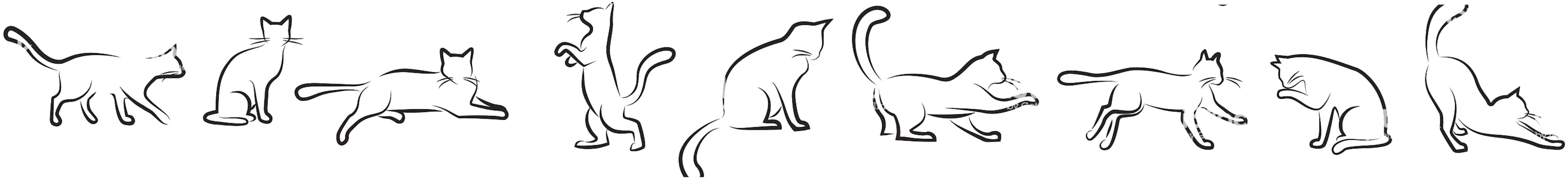


# Image deformation vs. data augmentation?

- https://github.com/albu/albumentations

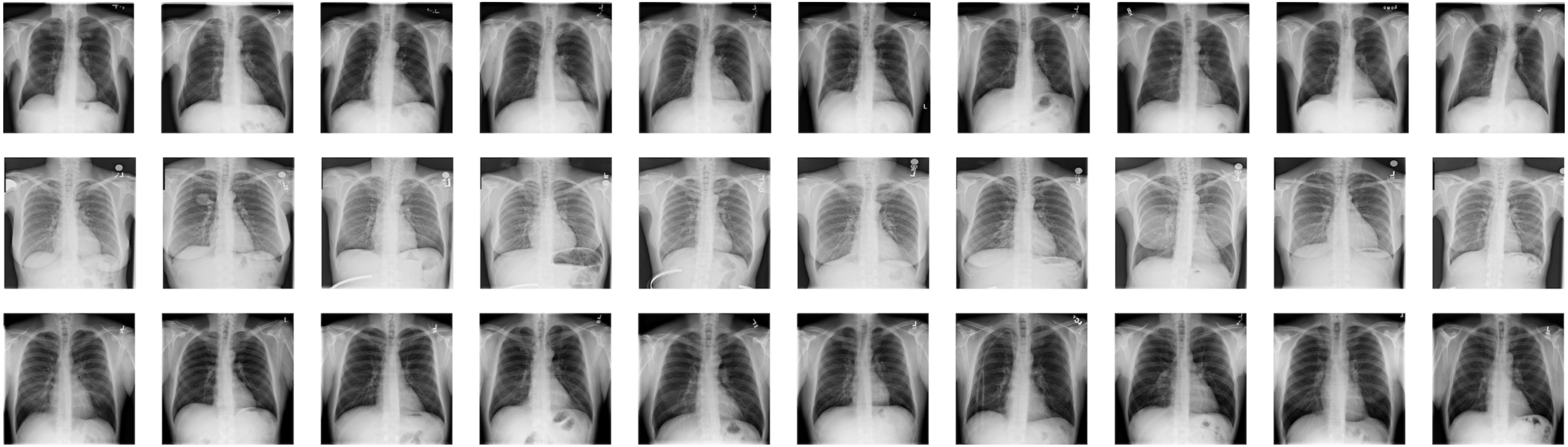
**To incorporate other meaningful image deformations into our framework, the deformation should belong to pixel-level transform, rather than spatial-level transform.**





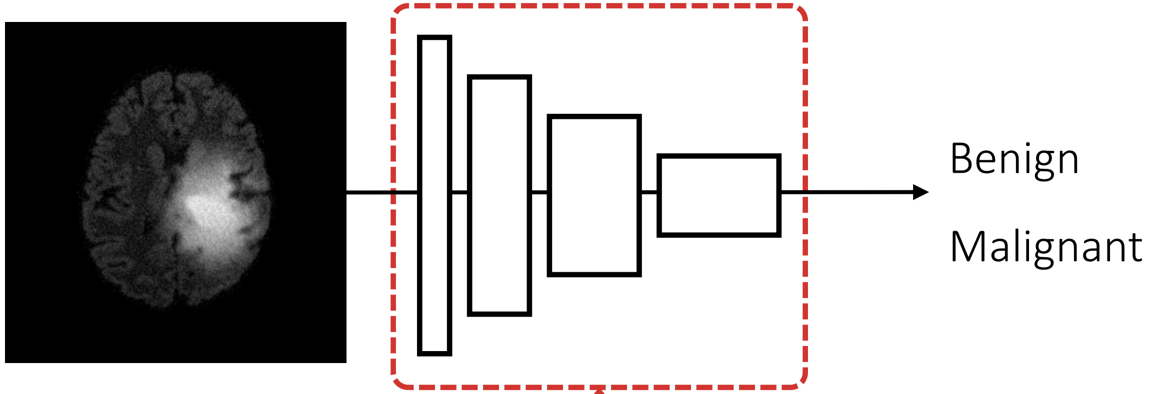
# Why are the proposed image deformations in your paper effective?

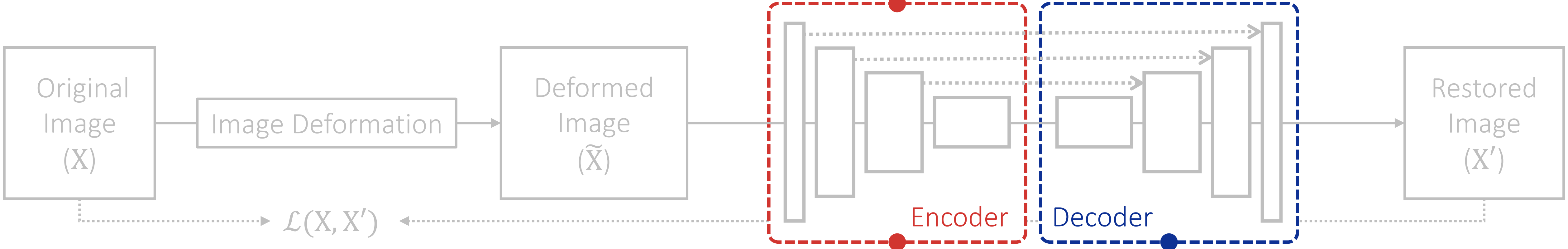
**Medical images contain similar anatomy. The sophisticated yet recurrent anatomy offers consistent patterns for self-supervised learning to discover common representation of a particular body part.**



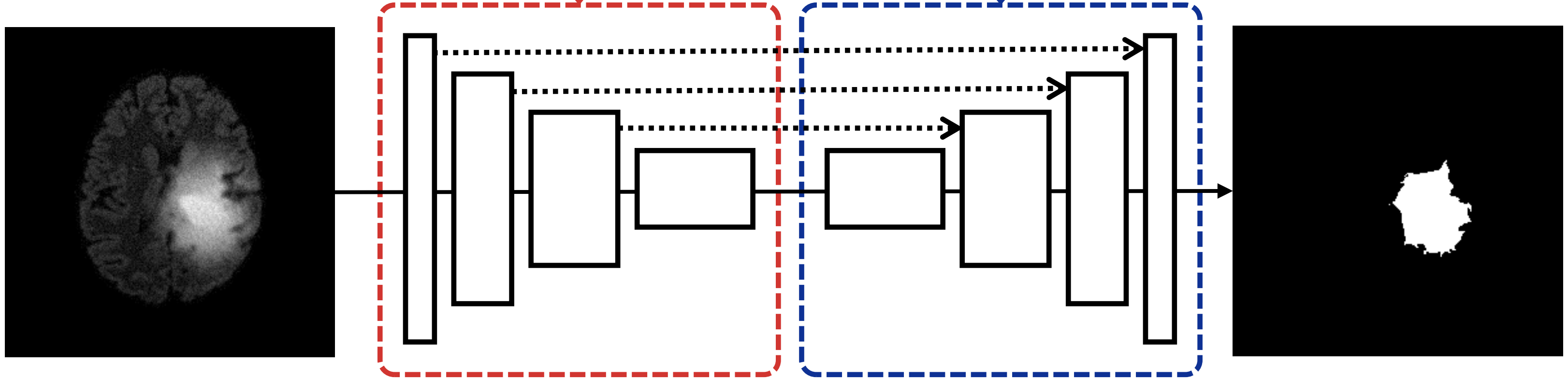
Once pre-trained,

the **encoder** could be used
for target classification tasks
*e.g.*, brain tumor classification;



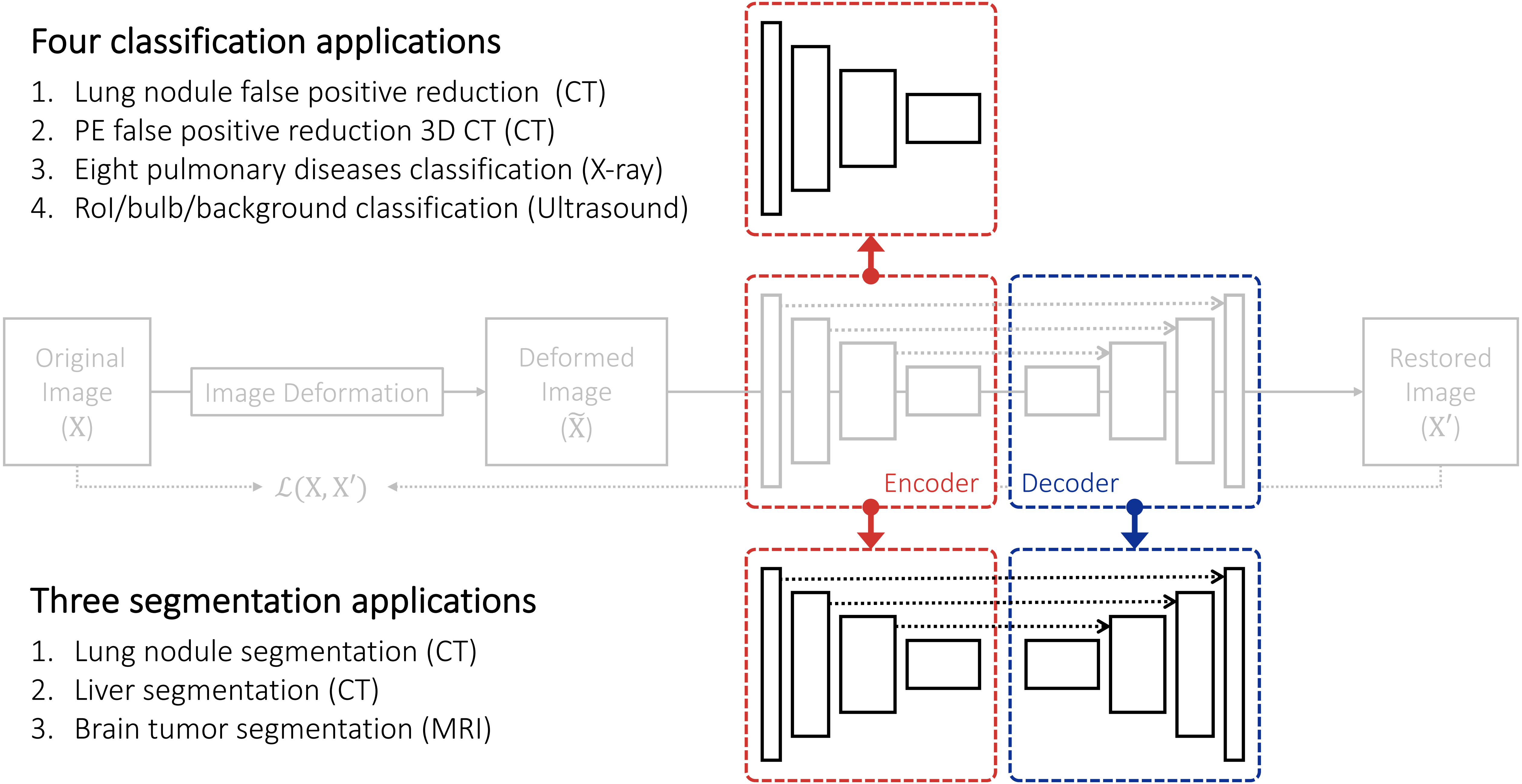




the **encoder-decoder** could be used
for target segmentation tasks
*e.g.*, brain tumor segmentation



## Four classification applications

1. Lung nodule false positive reduction  (CT)
2. PE false positive reduction 3D CT (CT)
3. Eight pulmonary diseases classification (X-ray)
4. RoI/bulb/background classification (Ultrasound)




## Three segmentation applications

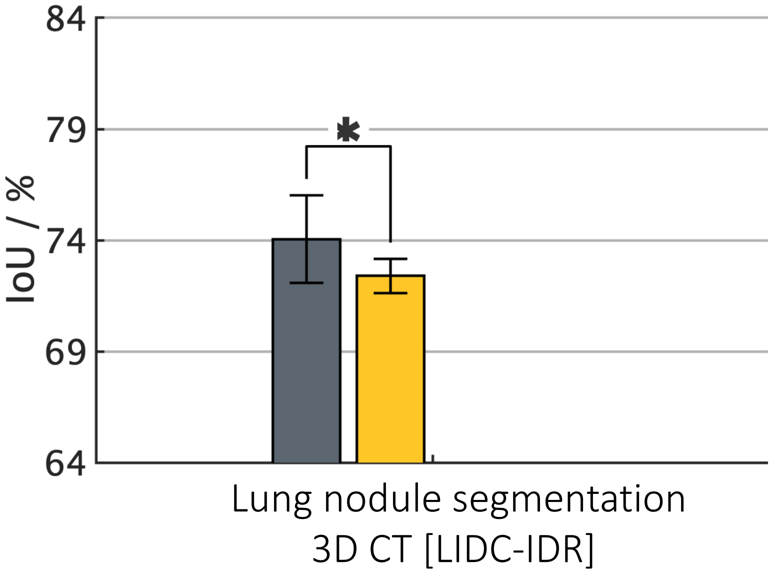
1. Lung nodule segmentation (CT)
2. Liver segmentation (CT)
3. Brain tumor segmentation (MRI)

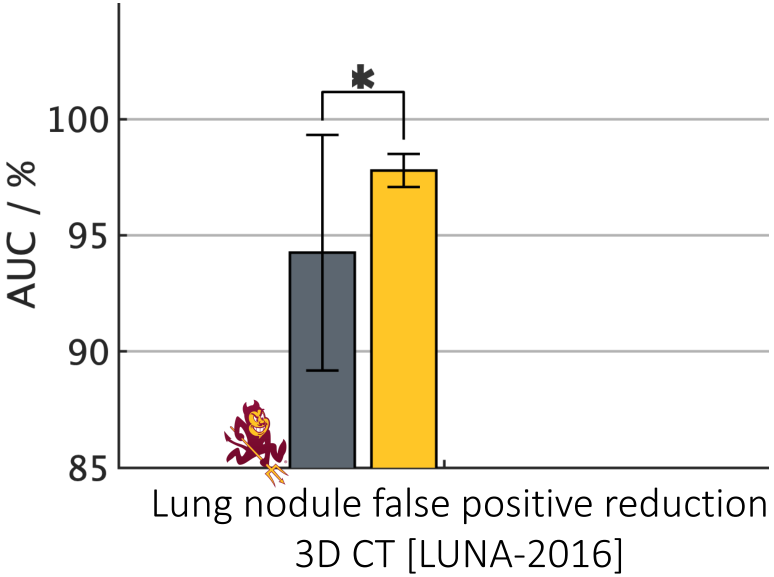
## Devils in 3D Models

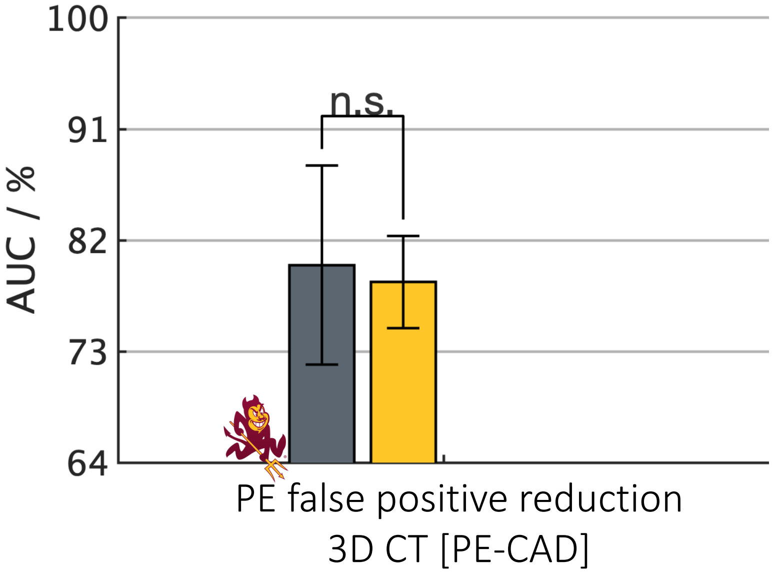
: Learning from scratch *simply* in 3D may not necessarily yield performance better than fine-tuning from ImageNet in 2D

Scratch 2D
Scratch 3D
ImageNet
Genesis X-ray 2D
Genesis CT 2D
Genesis CT 3D

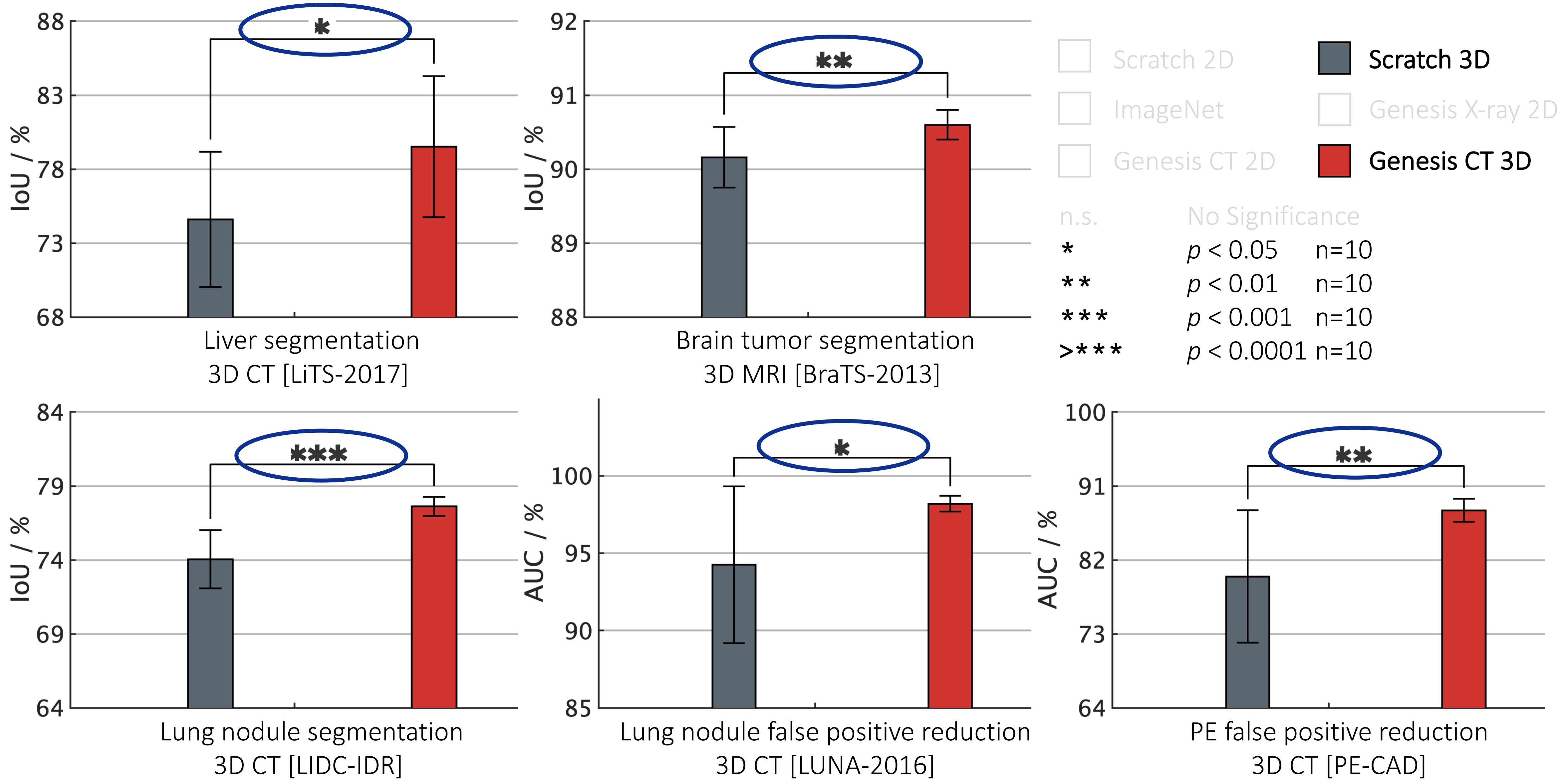
| | | |
|---|---|---|
| n.s. | No Significance | |
| * | $p < 0.05$ | n=10 |
| ** | $p < 0.01$ | n=10 |
| *** | $p < 0.001$ | n=10 |
| >*** | $p < 0.0001$ | n=10 |



# Result I: Models Genesis outperform 3D models trained from scratch



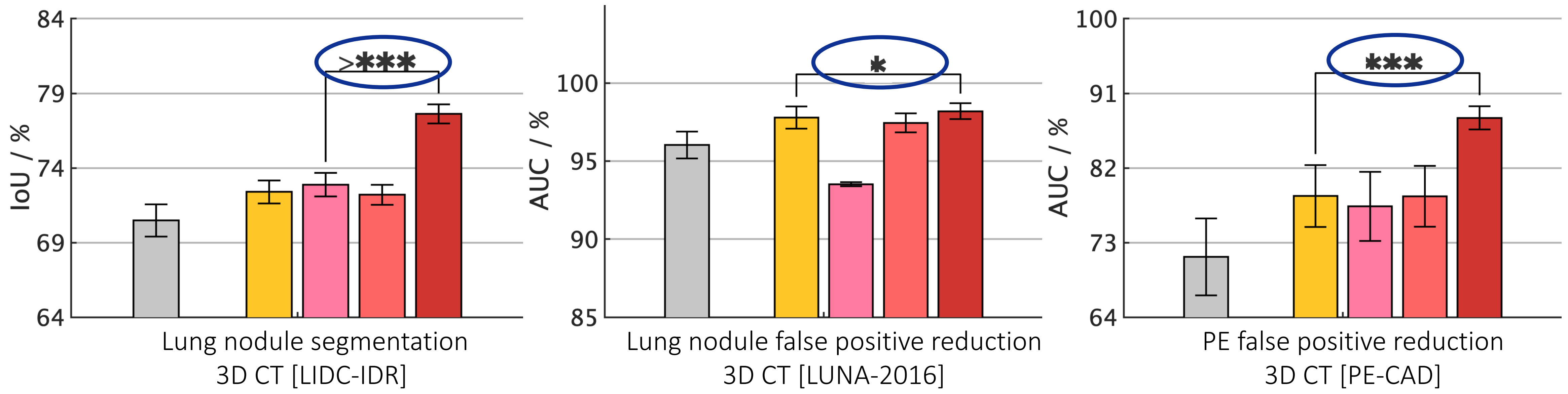
# Result II: Models Genesis consistently outperform any 2D approaches

including

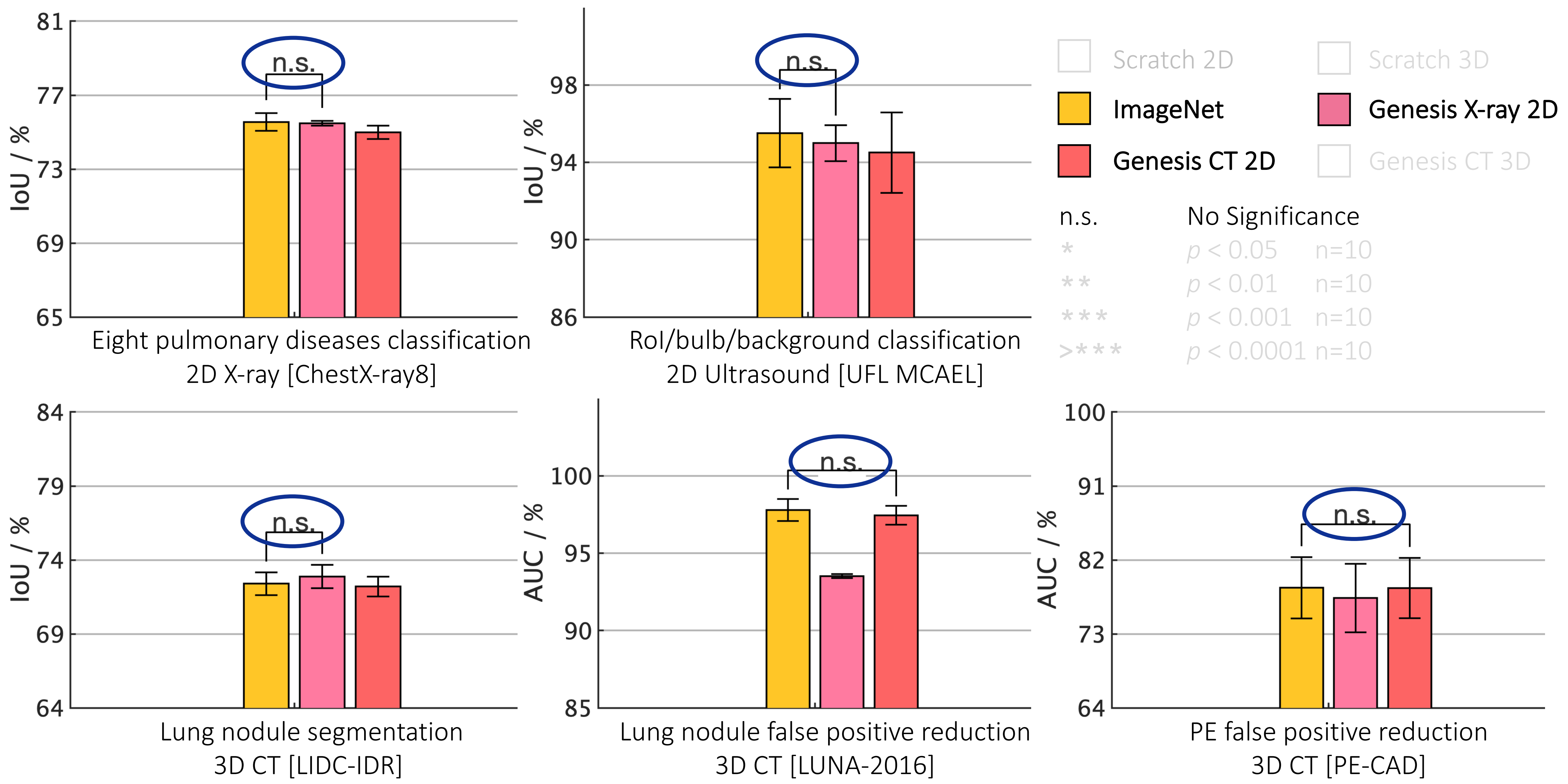
1. ImageNet (state-of-the-art)
2. Models Genesis 2D (degraded)
   - Genesis X-ray 2D: pre-trained on NIH X-ray dataset
   - Genesis CT 2D: pre-trained on LUNA-2016 dataset

Scratch 2D
Scratch 3D
ImageNet
Genesis X-ray 2D
Genesis CT 2D
Genesis CT 3D

| | | |
|---|---|---|
| n.s. | No Significance | |
| * | $p < 0.05$ | n=10 |
| ** | $p < 0.01$ | n=10 |
| *** | $p < 0.001$ | n=10 |
| >*** | $p < 0.0001$ | n=10 |



# Result III: Models Genesis 2D (self-supervised) ≈ ImageNet (supervised)



# Medical ImageNet?


IMAGENET

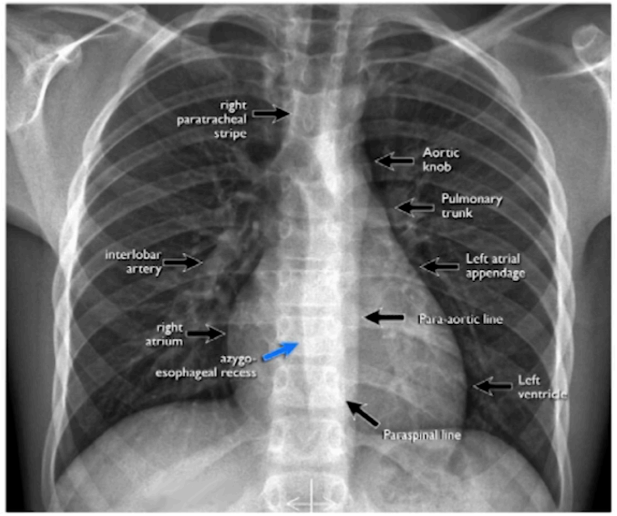
14,197,122 images, 21841 synsets indexed

Explore Download Challenges Publications Updates About

Not logged in. Login I Signup

**ImageNet** is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.
Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.


Karpathy, Andrej & Li, Fei Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions, CVPR, 2015

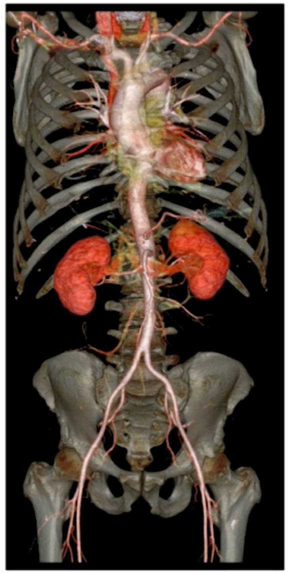



http://www.radiologyassistant.nl/

"Medical ImageNet"*

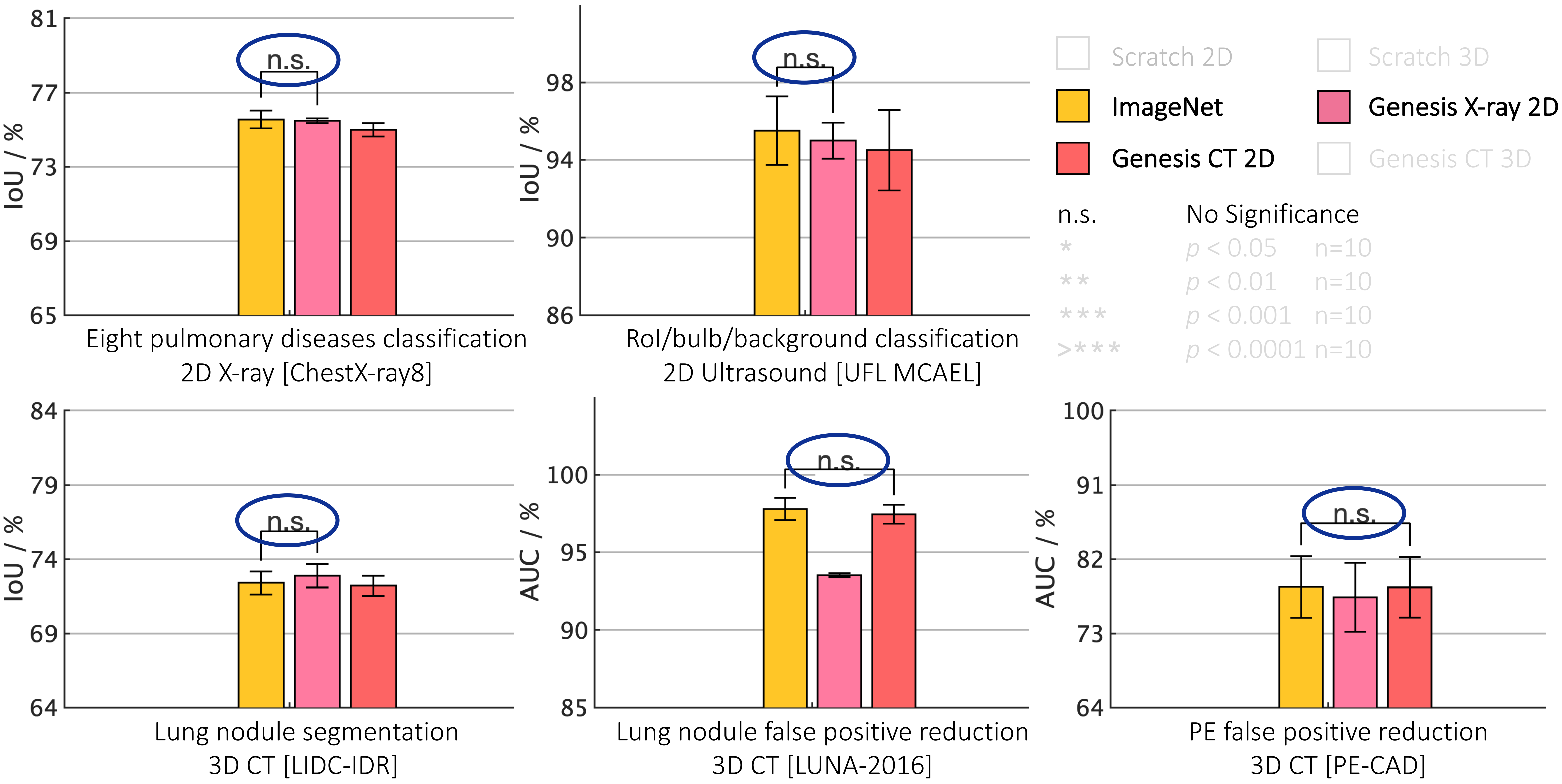
**A cloud-based, petabyte-scale, searchable, repository of diagnostic imaging studies for developing intelligent image analysis systems**

*Thanks to Fei Fei Li



Stanford | MEDICINE

# Result III: Models Genesis 2D (self-supervised) ≈ ImageNet (supervised)



| Family | | ImageNet | | Places205 | |
|---|---|---|---|---|---|
| | | Prev. | Ours | Prev. | Ours |
| A | Rotation[11] | 38.7 | **55.4** | 35.1 | **48.0** |
| R | Exemplar[8] | 31.5 | 46.0 | - | 42.7 |
| R | Rel. Patch Loc.[8] | 36.2 | 51.4 | - | 45.3 |
| A | Jigsaw[34, 51] | 34.7 | 44.6 | 35.5 | 42.2 |
| V | CC+vgg-Jigsaw++[36] | 37.3 | - | 37.5 | - |
| A | Counting[35] | 34.3 | - | 36.3 | - |
| A | Split-Brain[51] | 35.4 | - | 34.1 | - |
| V | DeepClustering[3] | **41.0** | - | **39.8** | - |
| R | CPC[37] | 48.7$^{\dagger}$ | - | - | - |
| R | Supervised RevNet50 | 74.8 | 74.4 | - | 58.9 |
| R | Supervised ResNet50 v2 | 76.0 | 75.8 | - | 61.6 |
| V | Supervised VGG19 | 72.7 | 75.0 | 58.9 | 61.5 |

$^{\dagger}$ *marks results reported in unpublished manuscripts.*

Kolesnikov et al. "Revisiting Self-Supervised Visual Representation Learning." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019). https://github.com/google/revisiting-self-supervised

| Method | Pretext Tasks | Classification | Detection | Segmentation |
|---|---|---|---|---|
| **ImageNet Labels** [8] | — | 79.9 | 56.8 | 48.0 |
| Random(Scratch) [8] | — | 57.0 | 44.5 | 30.1 |
| ContextEncoder [19] | Generation | 56.5 | 44.5 | 29.7 |
| BiGAN [122] | Generation | 60.1 | 46.9 | 35.2 |
| ColorfulColorization [18] | Generation | 65.9 | 46.9 | 35.6 |
| SplitBrain [42] | Generation | 67.1 | 46.7 | 36.0 |
| RankVideo [38] | Context | 63.1 | 47.2 | $35.4^{\dagger}$ |
| PredictNoise [46] | Context | 65.3 | 49.4 | $37.1^{\dagger}$ |
| JigsawPuzzle [20] | Context | 67.6 | 53.2 | 37.6 |
| ContextPrediction [41] | Context | 65.3 | 51.1 | — |
| Learning2Count [130] | Context | 67.7 | 51.4 | 36.6 |
| **DeepClustering** [44] | **Context** | **73.7** | **55.4** | **45.1** |
| WatchingVideo [81] | Free Semantic Label | 61.0 | 52.2 | — |
| CrossDomain [30] | Free Semantic Label | 68.0 | 52.6 | — |
| AmbientSound [154] | Cross Modal | 61.3 | — | — |
| TiedToEgoMotion [95] | Cross Modal | — | 41.7 | — |
| EgoMotion [94] | Cross Modal | 54.2 | 43.9 | — |

Jing, et al. “Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey." arXiv preprint arXiv:1902.06162 (2019).
https://arxiv.org/pdf/1902.06162.pdf

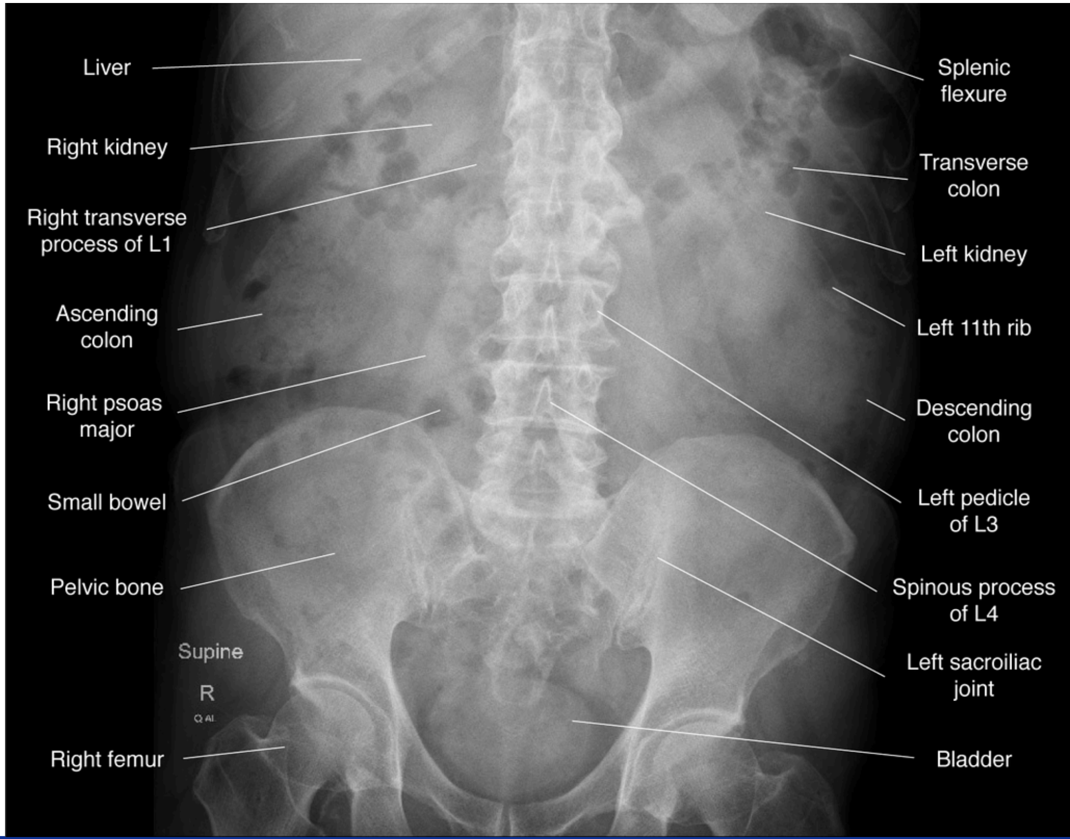
# Medical ImageNet?

**Models Genesis are not designed to replace such a large, strongly annotated dataset for medical image analysis like ImageNet for computer vision, but rather helping create one.**

**Open Question**

Models Medical ImageNet > Models Genesis?

1. Millions of systematic annotated medical images
2. Disease/organ class imbalance
3. Pixel/voxel utilization rate
4. Availability of medical images



# Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis

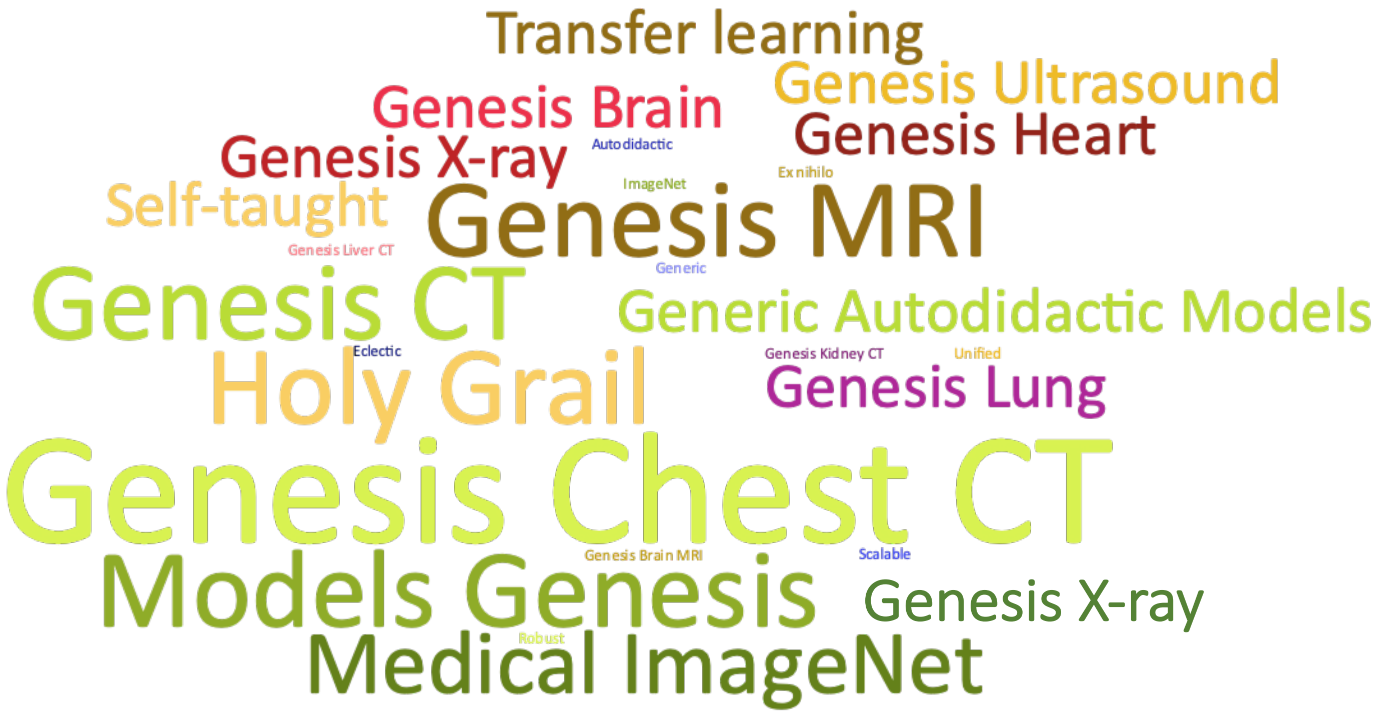
We offer a set of powerful pre-trained 3D models, concluding that

1. Models Genesis outperform 3D models trained from scratch
2. Models Genesis consistently outperform any 2D approaches
3. Models Genesis (2D) offer performances equivalent to supervised pre-trained models

Genesis Chest CT

Genesis X-ray

# Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis



# Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis

**Acknowledgement**

**Try it for yourself**

Code, data, and models

are available online

github.com/MrGiovanni/ModelsGenesis