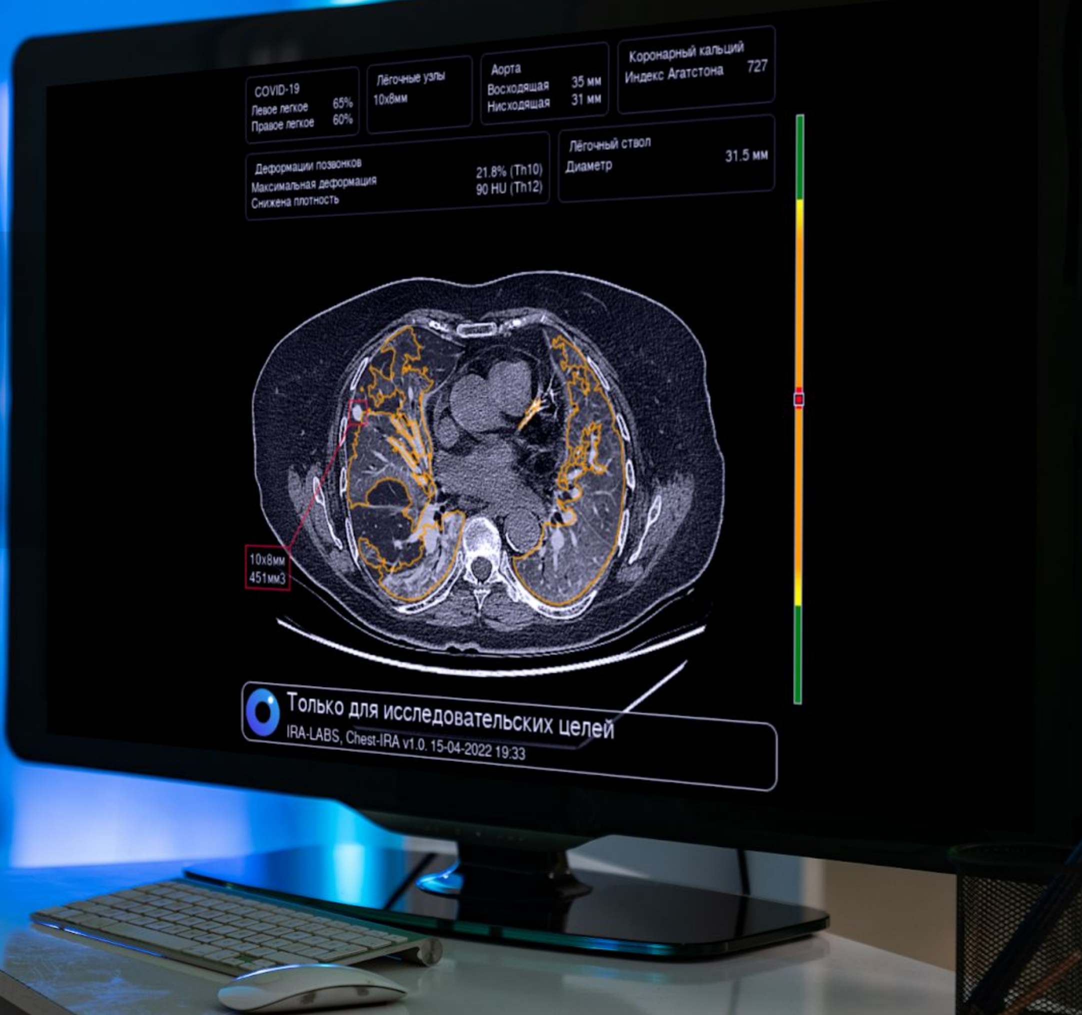
IRA LABS

Computer vision for medical imaging

research challenges & market demands

Mikhail Belyaev, PhD,
CEO at IRA Labs

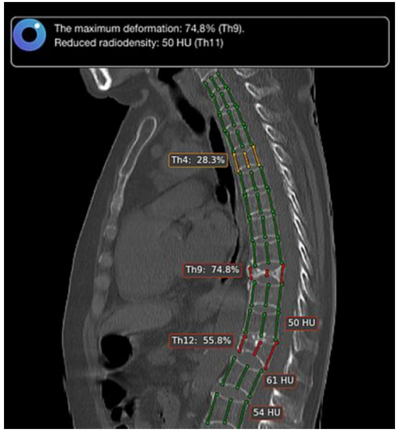
October 31, 2022

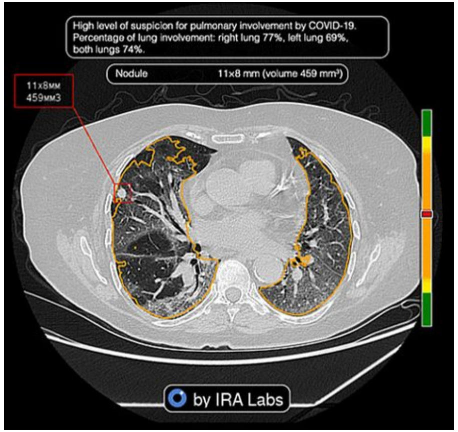


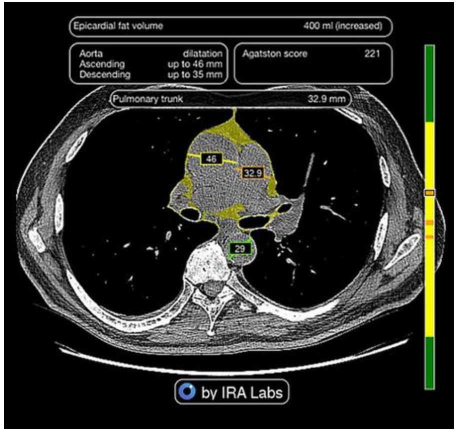
Medical Computer Vision

Why data science for medicine

- **Algorithmic challenges**: heterogeneous 3D medical images require new computer vision methods.
- **Social impact**: it's not about selling ads; it's about saving lives.
- **Innovations**: medical AI startups received [more funding than any other AI sector]() in 2019 ($4 bil), the amount tripled in 2021 ($12.4 bil)

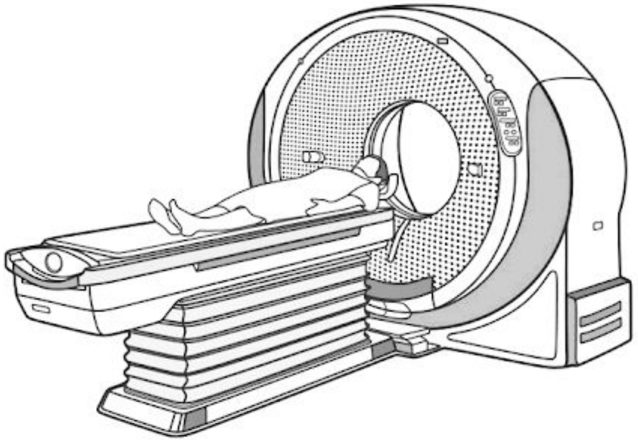


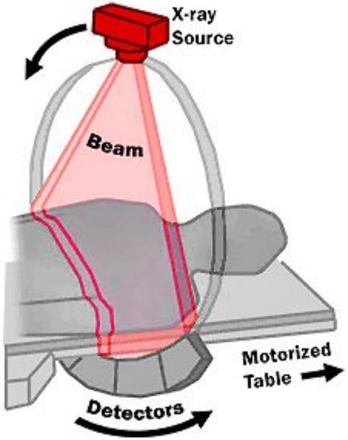


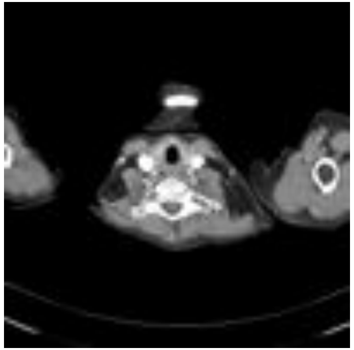


Medical Computer Vision

We work with 1-channel 3D images: Computed Tomography (CT) and Magnetic Resonance Imaging (MRI)





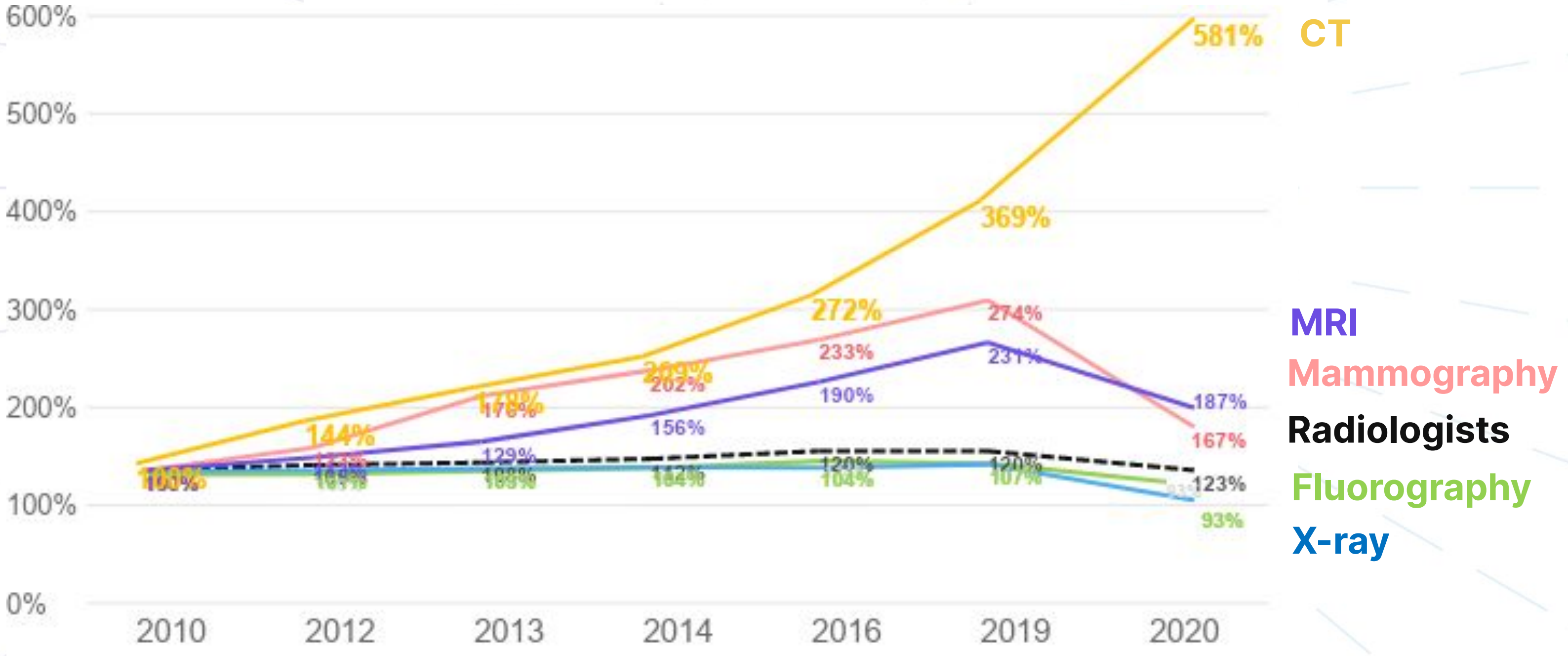


*Fast Lung Localization in Computed Tomography by a 1D Detection Network. **USBEREIT** 2021.*
Vladislav Proskurov, Anvar Kurmukov, Maxim Pisov, Mikhail Belyaev

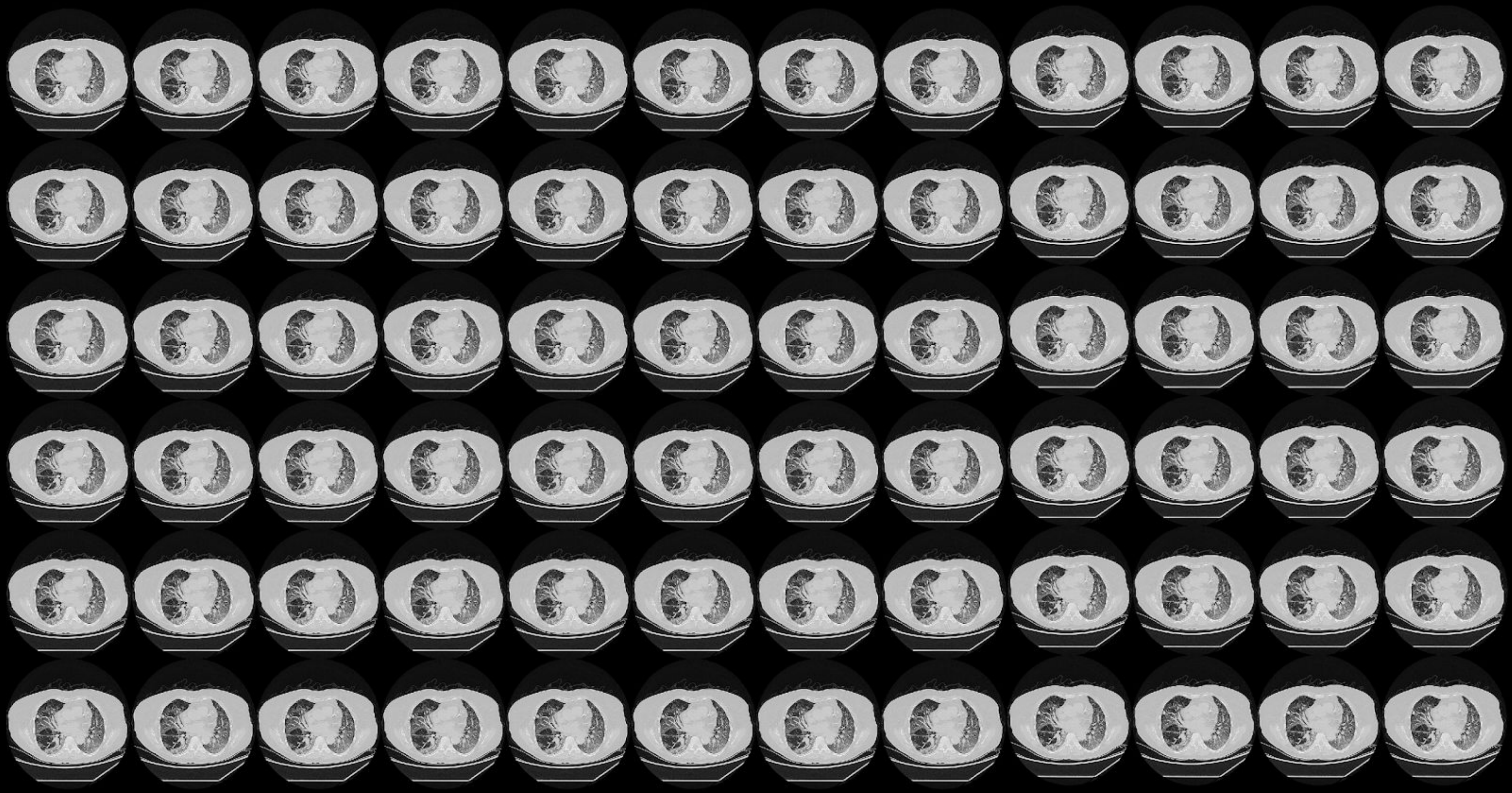
IRA LABS

The burden on radiologists is growing significantly faster than the number of radiologists

Dynamics of CT scans' and radiologists' increase in Russia



Radiologists need to analyze up to 50,000 similar images every day



Industrial view: AI in radiology

The claim is wrong, as we know in 2022. Moreover, many countries experiences shortage of radiologists despite the total number was increased. E.g. the UK expect 44% shortfall by 2025 [2]

Why?

> We should stop training radiologists now. It's just completely obvious that within five years, deep learning is going to do better than radiologists.
> **Geoff Hinton: On Radiology. November 24, 2016.**

1. https://www.youtube.com/watch?v=2HMPRXstSvQ
2. https://www.rcr.ac.uk/posts/new-rcr-census-shows-nhs-needs-nearly-2000-more-radiologists

Industrial view: AI in radiology

The claim is wrong, as we know in 2022. Moreover, many countries experiences shortage of radiologists despite the total number was increased. E.g. the UK expect 44% shortfall by 2025 [2]

Why?

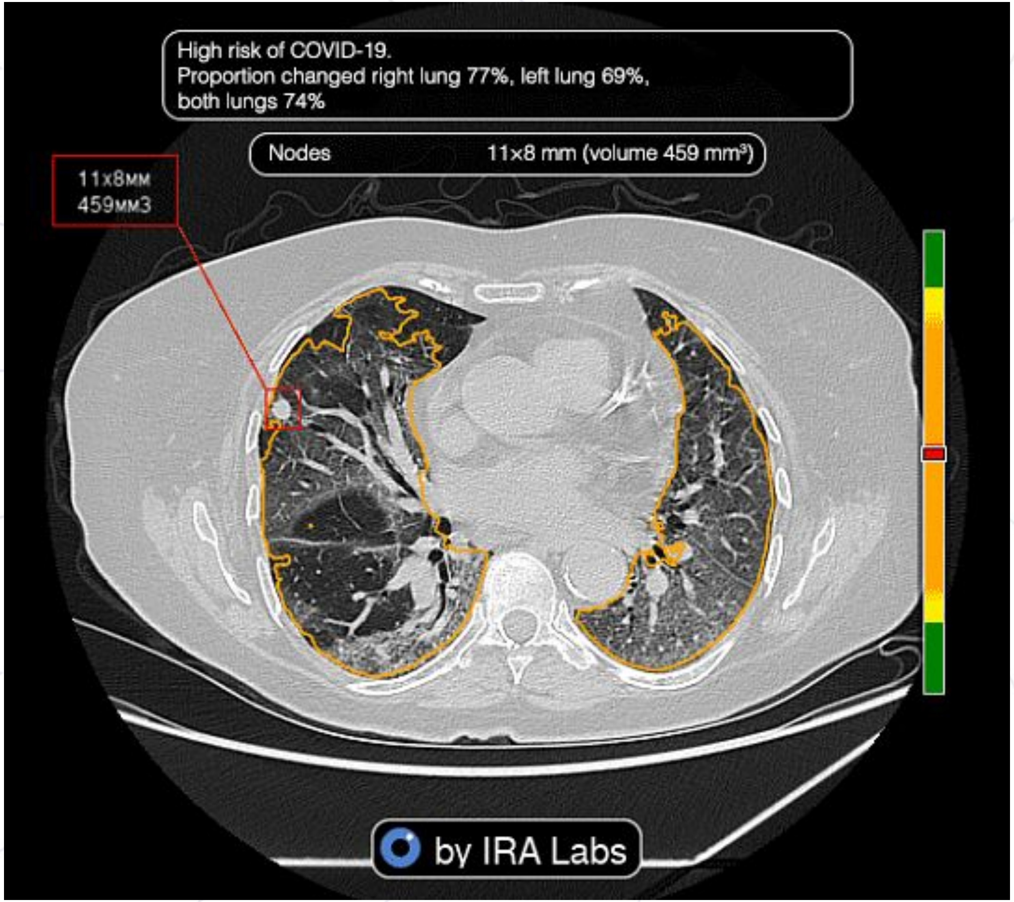
1. "Slow" industry: regulatory barriers, reimbursement challenges
2. Limited product functionality
3. Low accuracy of models in independent evaluations

We should stop training radiologists now. It's just completely obvious that within five years, deep learning is going to do better than radiologists.
Geoff Hinton: On Radiology. November 24, 2016.

Companies are usually focused on specific clinical use-cases with narrow, single-purpose algorithms... However, this specialist approach limits the utility of AI as radiologists require more clinically diverse AI tools that can support a wide range of clinical presentations and imaging findings
Signify Research White paper on Comprehensive AI solutions for medical imaging

1. *https://www.youtube.com/watch?v=2HMPRXstSvQ*
2. *https://www.rcr.ac.uk/posts/new-rcr-census-shows-nhs-needs-nearly-2000-more-radiologists*

AI automates routine tasks



Fast processing of CT images (up to 1 min — the best in the industry)

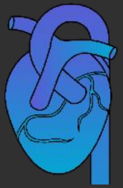
Convenient visualization of findings to reduce the time of their localization:

- The intuitive color scheme for coding of pathology's significance
- Navigation slider

Automatic quantification of complex biomarkers

Automatic generation of radiology reports

Our product is looking for *all* major pathologies



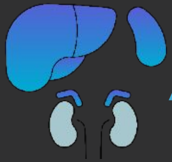
Cardio-vascular system

Coronary artery calcification

Aortic aneurysms and dilatations

Pulmonary artery dilation

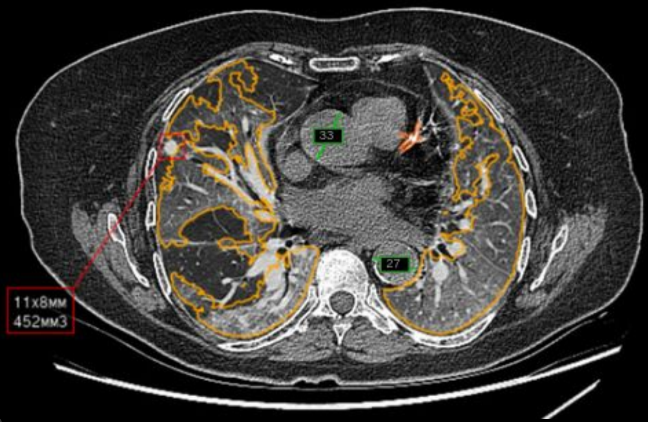
Pericardial fat



Abdomen

Adrenal masses

Fatty liver disease & liver lesions



More than

1 000 000 scans processed

75 seconds to process one scan

Respiratory system

COVID-19

Lung cancer

Emphysema

Bronchiectasis, honeycomb

Musculo-skeletal system

Osteoporosis

Vertebral fractures

Sarcopenia

Ribs fracture

Mosmed.ai: results

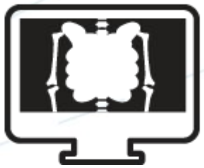
	IRA LABS	Nanox. Zebra	Rad logics	Botkin.Ai	Third Opinion	Celsus AI	Cvision Lab		Gamma med	SberMed AI	AiDiagno stic
COVID-19	0,97 🏆		0,85	0,96	0,97 🏆	0,96 🏆	0,97 🏆	0,82 🏆	0,92 🏆	N/A	
Lung cancer	0,932		0,88	0,81	N/A	0,88				N/A	0,88
Aortic aneurysm	0,997 🏆		N/A			N/A	0,92				N/A
Pulmonary hypertension	1,0 🏆					N/A					N/A
Coronary calcium	0,986 🏆	0,84				N/A	0,92				
COPD	0,989 🏆			0,68		N/A	0,86				
Epicardial fat	0,99						0,83				
Osteoporosis	0,995 🏆	0,91 🏆				N/A					
Adrenal mass	1.0										
Pleural effusion	0,999		N/A	N/A	0,92	N/A	0,89				
Ribs fractures	N/A										

Legend:
Numbers represent ROC AUC, ranging from 0.81 to 1;
participates in Mosmed.AI,
excluded from Mosmed.AI,
not entered Mosmed.AI,
🏆 - excellent quality.

IRA LABS

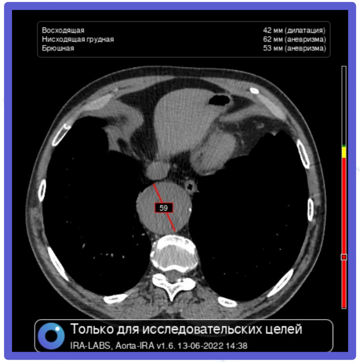
Cloud based storage

Workstation



Original
DICOM

Generated
DICOM



Prefilled report with
recommendations

Emphysematous bulla 3.2 cm. Round glass opacities in both lungs. Thoracic aortic aneurysm, up to 62 mm. **A consultation with a cardiovascular surgeon is required.**

IRA LABS

Medical Computer VIsion

Non medical images: ImageNet

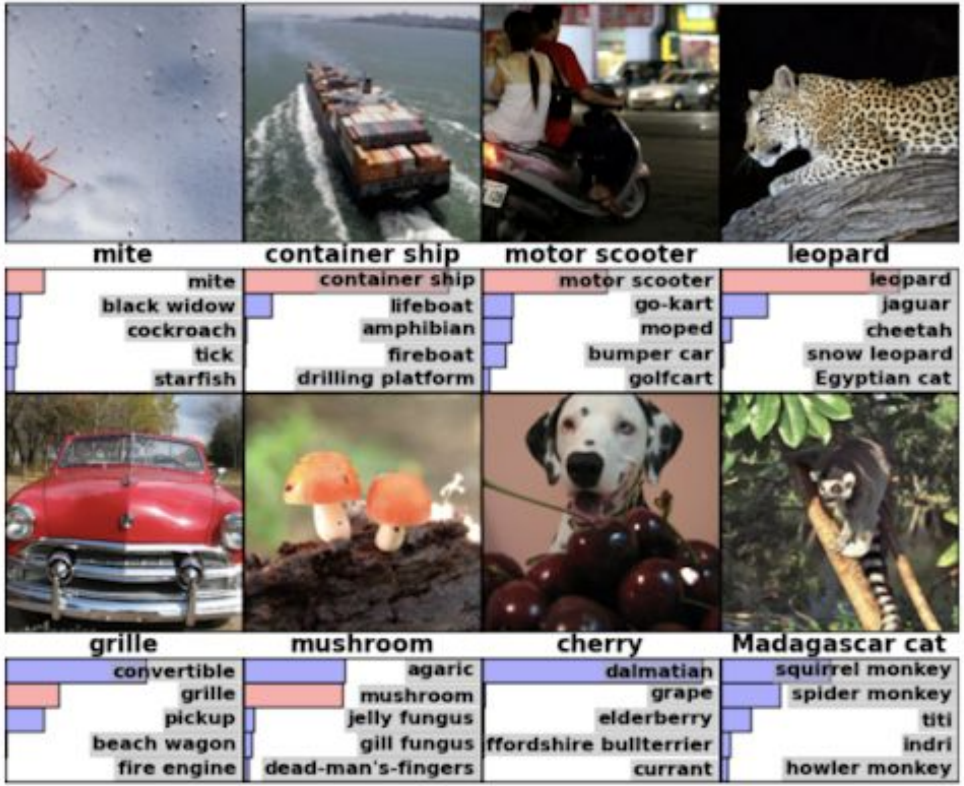
ImageNet 1k dataset:

- 1.2 millions of images
- 1000 of classes,

ImageNet full

- **14 millions** of images,
- >20k of classes

Typical image resolution ~256x256

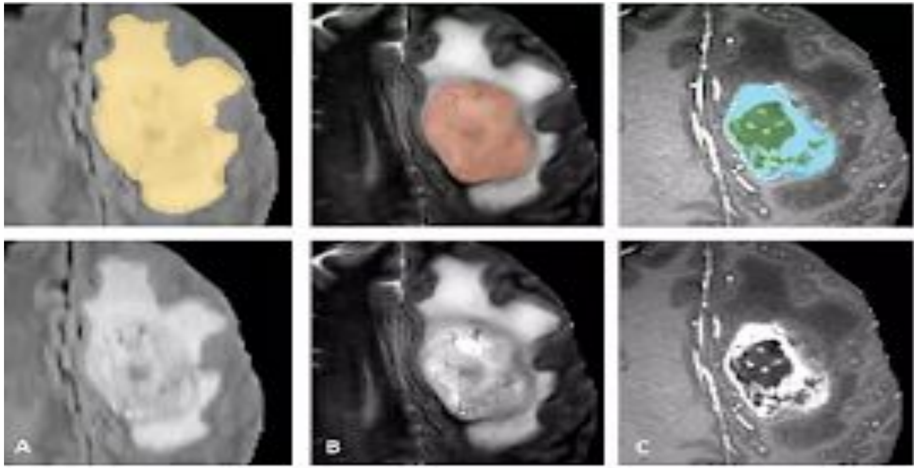


[ImageNet: A large-scale hierarchical image database]()

Medical imaging data

ImageNet

- Data set size: 14 millions of images,
- Image size: 256x256



A typical MRI/CT dataset

- Data set size: hundreds (up to a couple of thousands)
- Image size: 150x150x150 for 1mm resolution

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. and Lanczi, L., 2014. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, *34*(10), pp.1993-2024.

Medical CV challenges

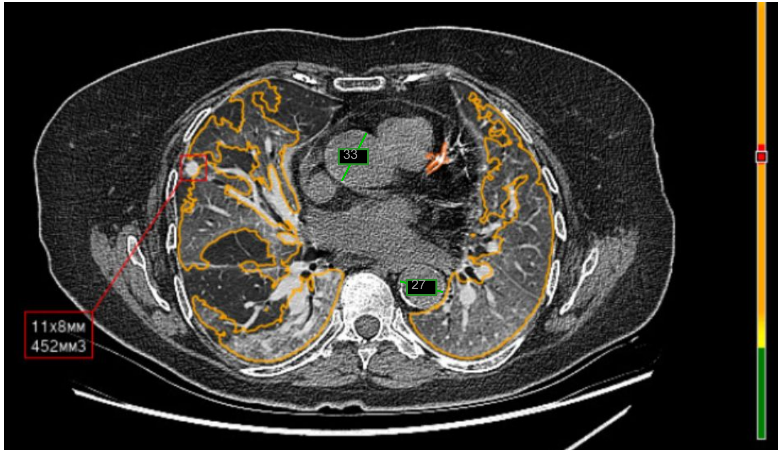
1. Datasets have extremely limited number of annotated cases

➔ Some privacy concerns frequently arise and prevent data publishing

➔ Detailed annotation of 3D images is time consuming

➔ It can hardly be outsourced to a broad community (e.g. via crowdsourcing)

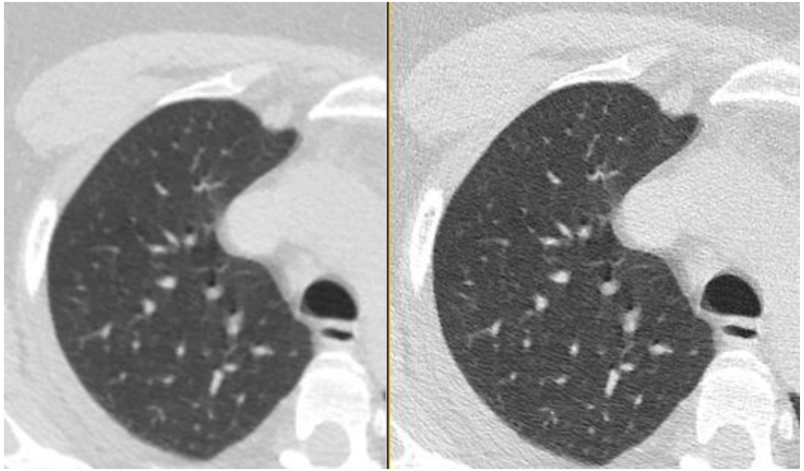
Impact on the industry:
- Low development speed
- Low product quality
- Lack of truly comprehensive solutions for CT & MRI.



Our product is an exception as we analyze 11 organs at the same time

Narrow & heterogeneous domains

MRI & CT scanners have many settings and every clinical site uses its own unique combination

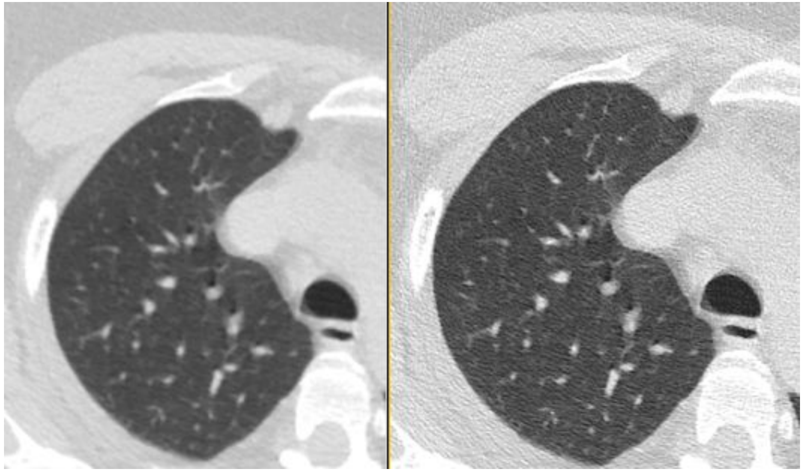


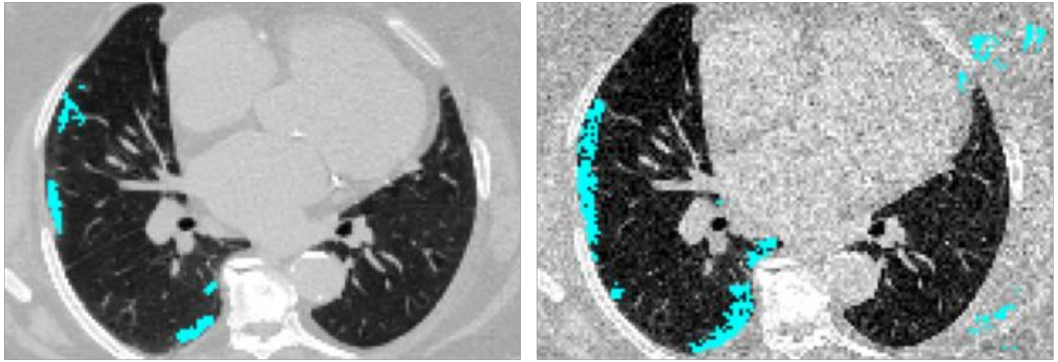
2 CTs of the same patient

5 MRI scans of different patients

Narrow & heterogeneous domains

MRI & CT scanners have many settings and every clinical site uses its own unique combination





Example of quality detioration: Predictions of COVID-19 lesions on two domains (the same patient).

A chest x-ray example

- ➜ Large datasets NIH & Mount Sinai Hospital more than 150 000 chest x-ray studies.
- ➜ Test dataset Indiana University Network for Patient Care, 4 000 cases
- ➜ Result: ROC AUC dropped from 0.93 to 0.81.

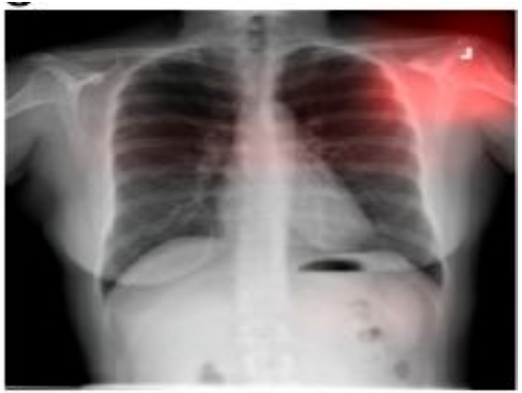
Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. J. Zech et al. PLoS One, 2018.

A chest x-ray example

- ➜ Large datasets NIH & Mount Sinai Hospital more than 150 000 chest x-ray studies.
- ➜ Test dataset Indiana University Network for Patient Care, 4 000 cases
- ➜ Result: ROC AUC dropped from 0.93 to 0.81.

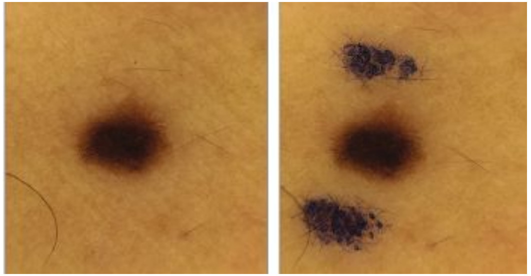
Clinical sites can be classified perfectly (accuracy 99.95%)!

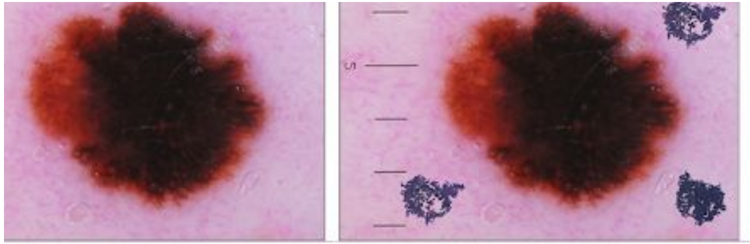
Here is the attention.



Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. J. Zech et al. PLoS One, 2018.

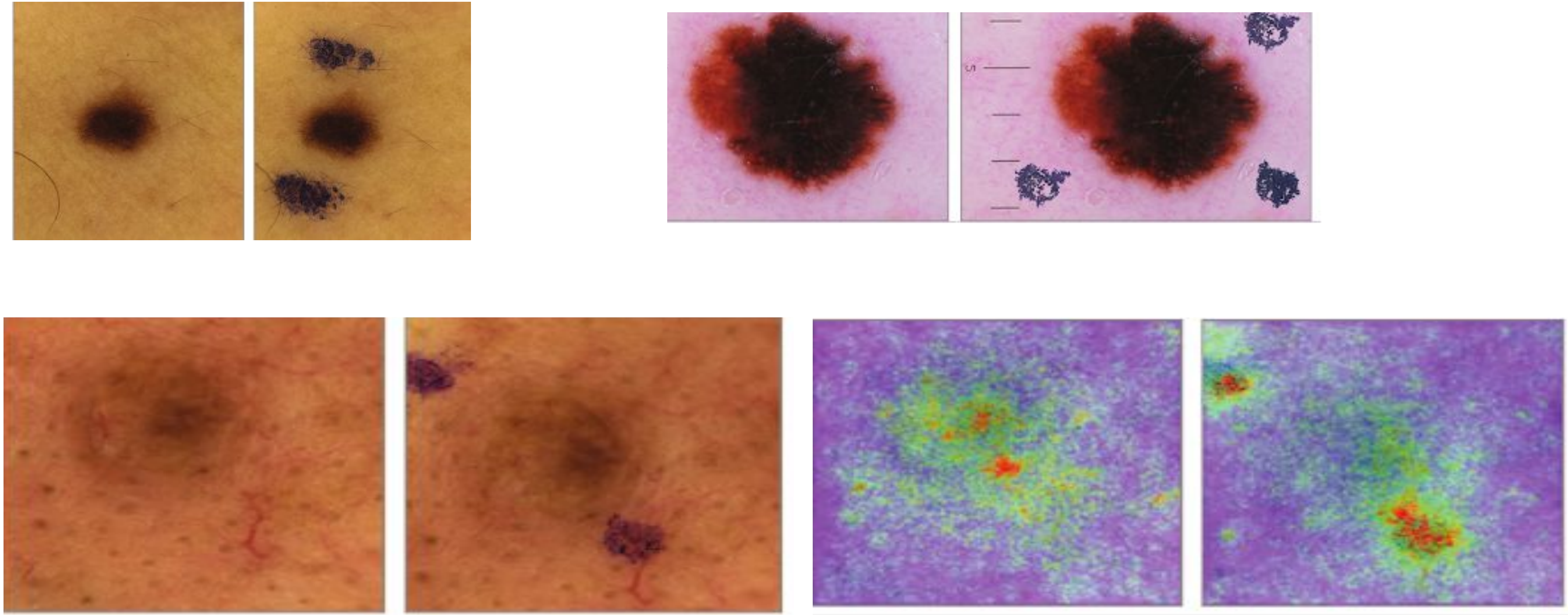
Another example: a CE mark product





Winkler J.K., et al, 2019. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA dermatology.

Another example: a CE mark product



Winkler J.K., et al, 2019. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA dermatology.

Medical CV challenges

2. Many narrow but heterogeneous domains.

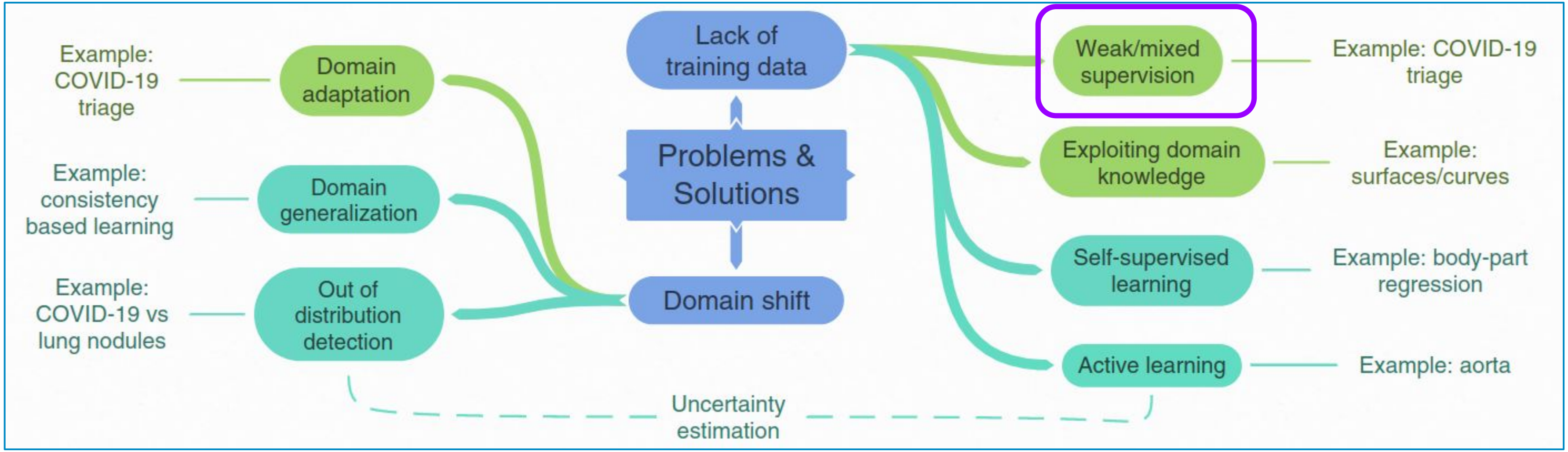
➜ In combination with (1): it’s hard to collect a representative data set.

➜ Possible changes not only in scanning parameters, but also in patient population (e.g. lung cancer AI broken by COVID-19)

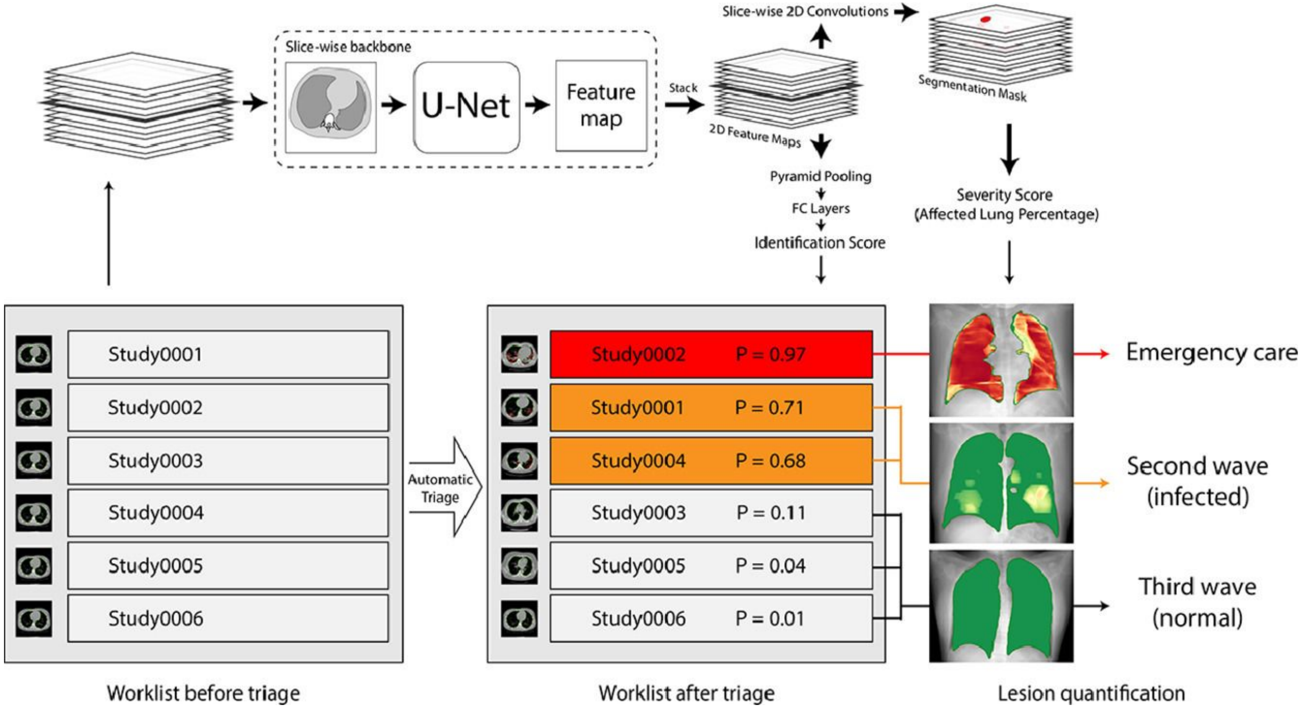
Impact on the industry:

- Lack of reproducibility.
- Additional regulatory barriers (though quite common for evidence-based medicine)

Medical CV challenges - how to address



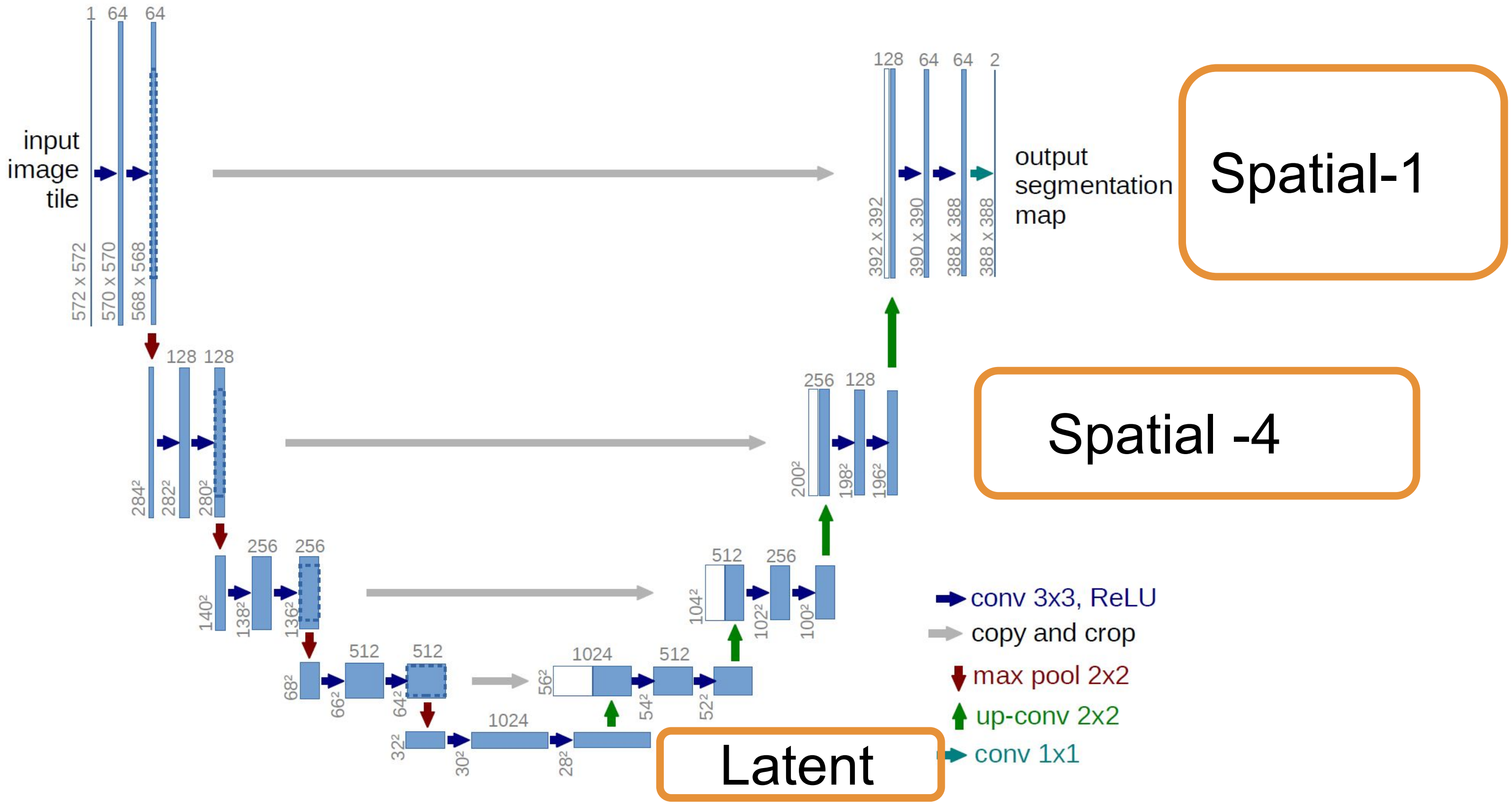
Multitask learning for patients' triage



Goal: combine weak image-wise & strong voxel-wise labels to improve patients' sorting (triage)

CT-based COVID-19 Triage: Deep Multitask Learning Improves Joint Identification and Severity Quantification 22 2021
M Goncharov, M Pisov, A Shevtsov, B Shirokikh, A Kurmukov, I Blokhin, ...
Medical image analysis 71, 102054

Multitask learning for patients' triage



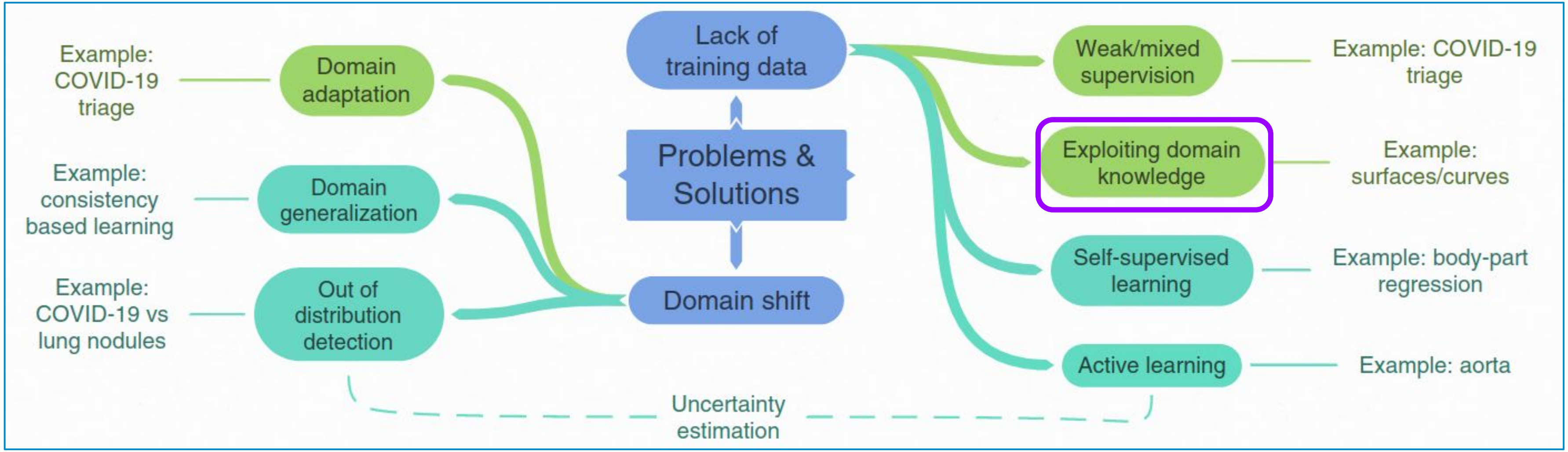
Multitask learning for patients' triage

	ROC-AUC (COVID-19 vs ·)			
	vs All others	vs Normal	vs Bac. Pneum.	vs Nodules
Multitask-Latent	$.79 \pm .06$	$.84 \pm .05$	$.73 \pm .06$	$.80 \pm .07$
Multitask-Spatial-4	$.89 \pm .03$	$.94 \pm .03$	$.83 \pm .05$	$.91 \pm .03$
Multitask-Spatial-1	$\mathbf{.93} \pm .01$	$\mathbf{.97} \pm .01$	$\mathbf{.87} \pm .01$	$\mathbf{.93} \pm .00$

Impact on the industry:

- We won competition with 15 other companies in 2020.
- Saved up to 30% of radiologists time in 100 hospitals.

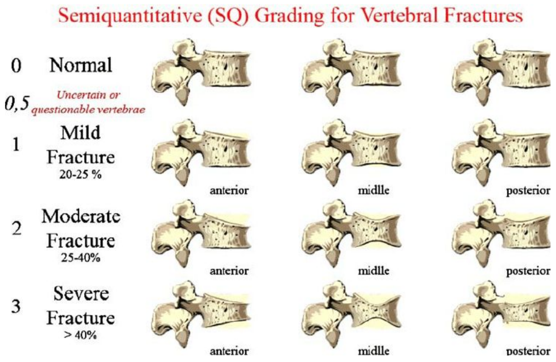
Medical CV challenges - how to address

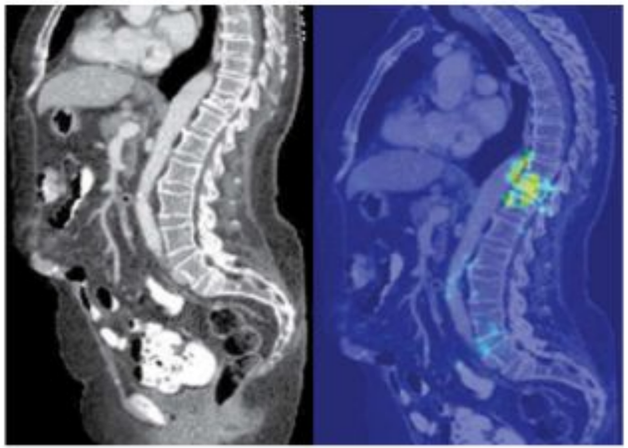


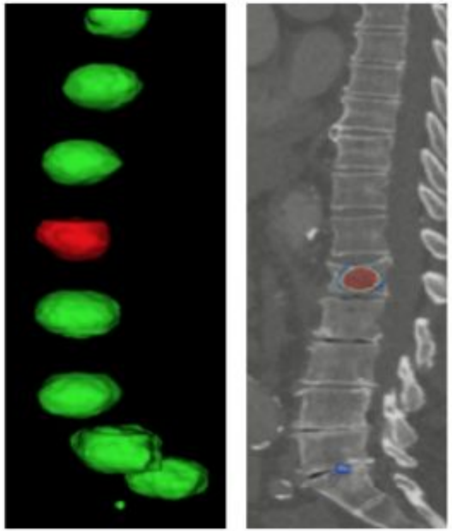
Vertebral fractures detection - methods

Classification-based

Segmentation - based





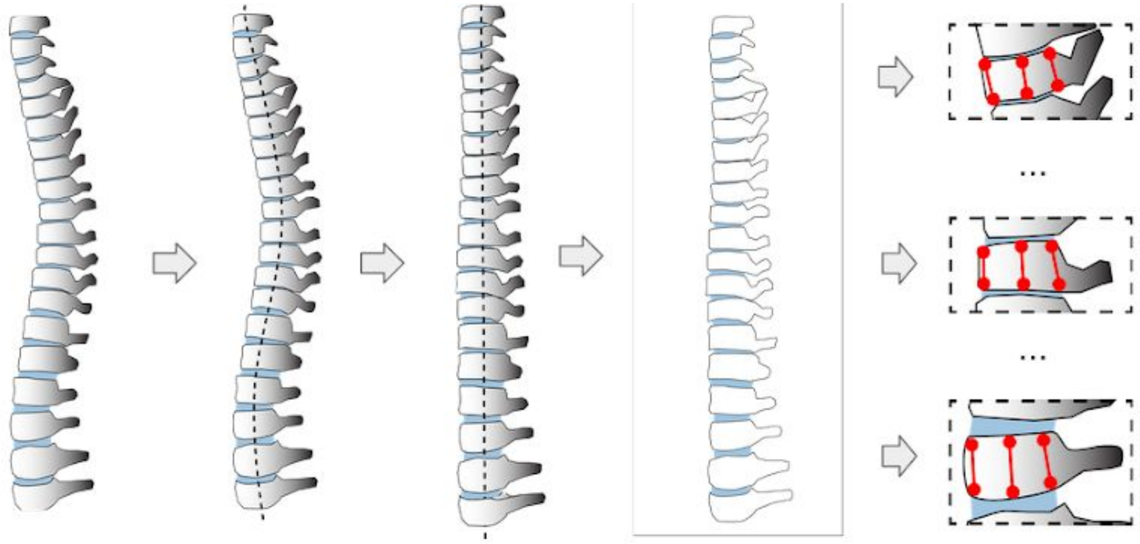


Impact on the industry:

- Lack of interpretability
- The next clinical step is unclear

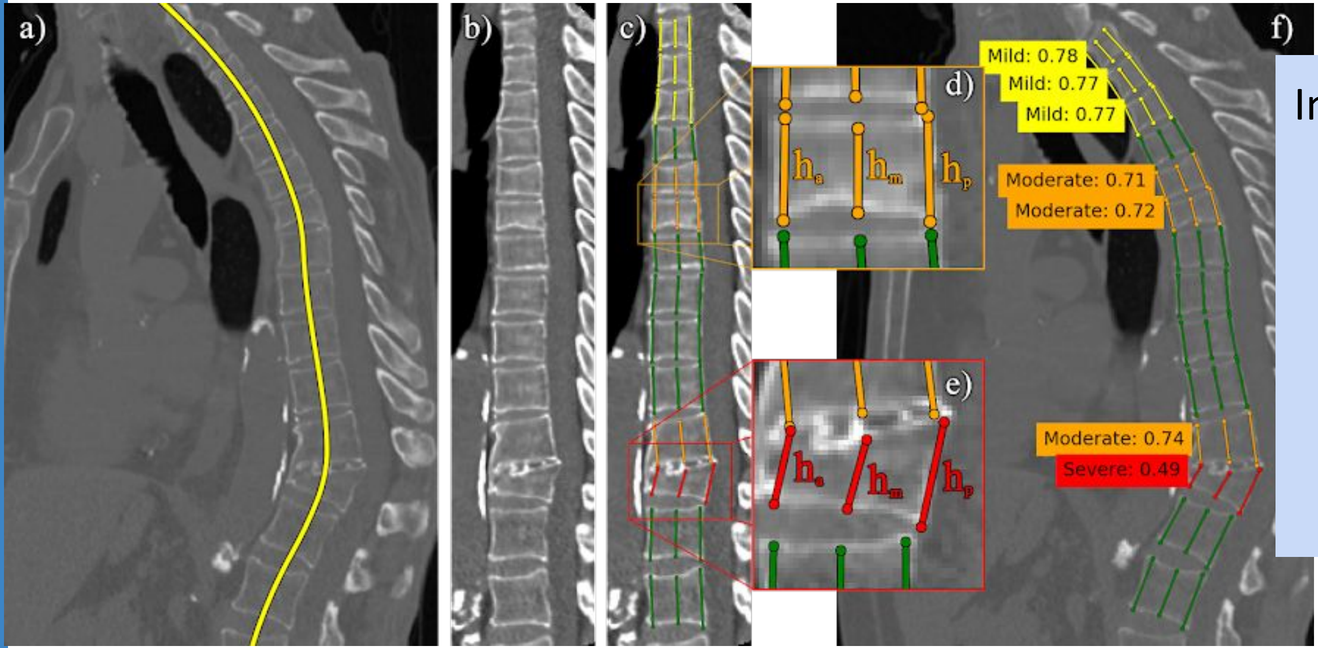
Vertebrae fractures detection

a. **Part 1: spine 'straightening' via curve estimation in 3D**
b. **Part 2: joint keypoint localization & vertebra detection in 2D**



Loss function: penalize errors for broken vertebrae more!

Vertebrae fractures detection



Impact on the industry:

- Easy-to-check outputs
- Automated clinical guidelines rather then segmentation / detection based approaches

Keypoints localization for joint vertebra detection and fracture severity quantification. ***M. Pisov, V. Kondratenko, A. Zakharov****. Medical Image Computing and Computer Assisted Interventions. 2020*

Vertebrae fractures detection - the product

IRA LABS

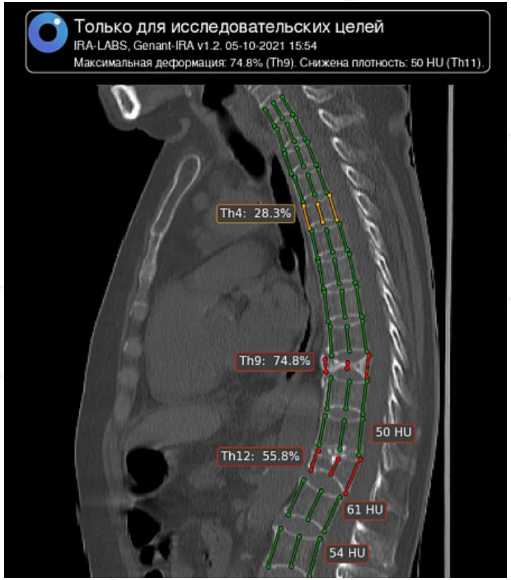
We compete with Zebra Medical Vision, a well funded Israeli startup (around $57 mil)

Fei Fei Lee

Series C - Zebra Medical Vision

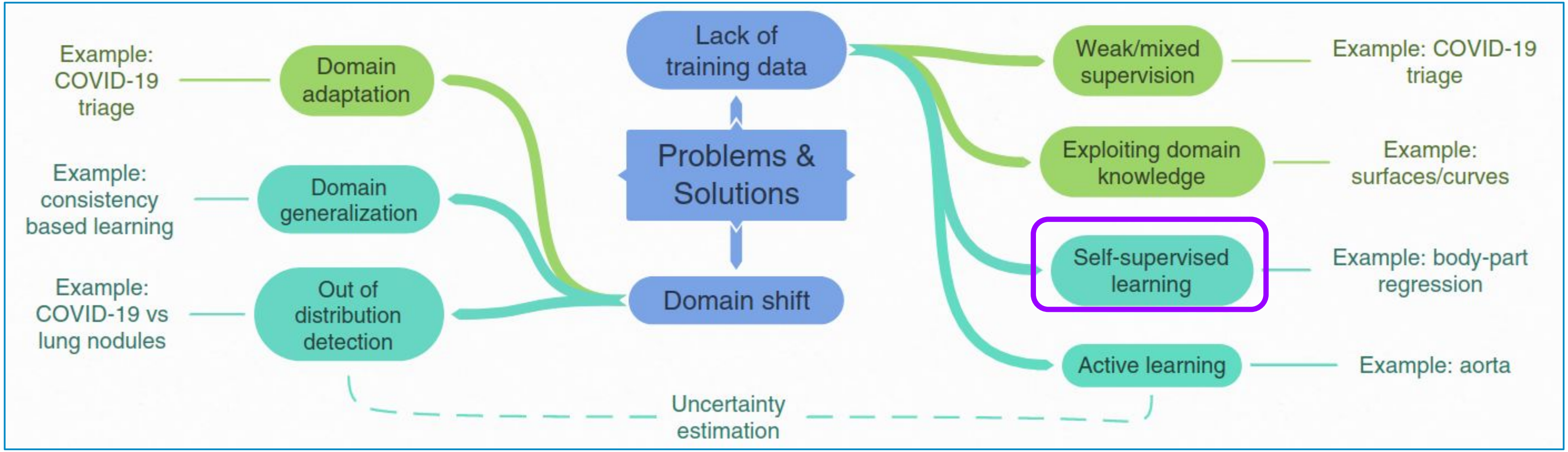
Impact on the industry:

- 2.5 better metrics (closer to ideal classification)
- 2 times faster
- Up to 50% better clinical feedback



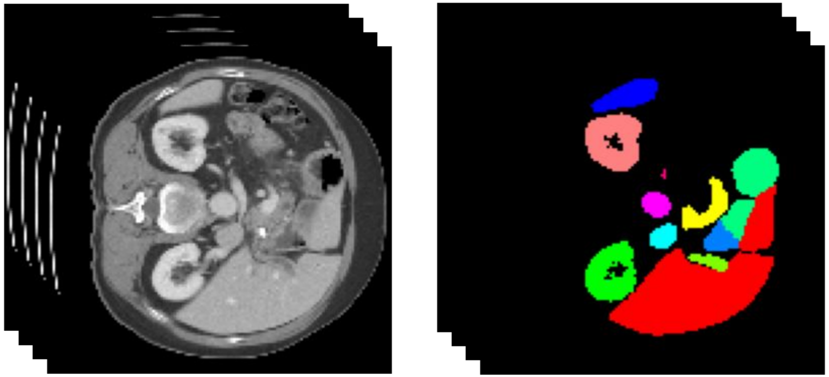
Genant-IRA

Medical CV challenges - how to address

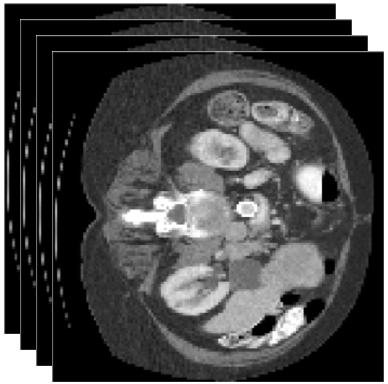


Self-supervised learning motivation

~10^2-10^3
Labeled
3D images

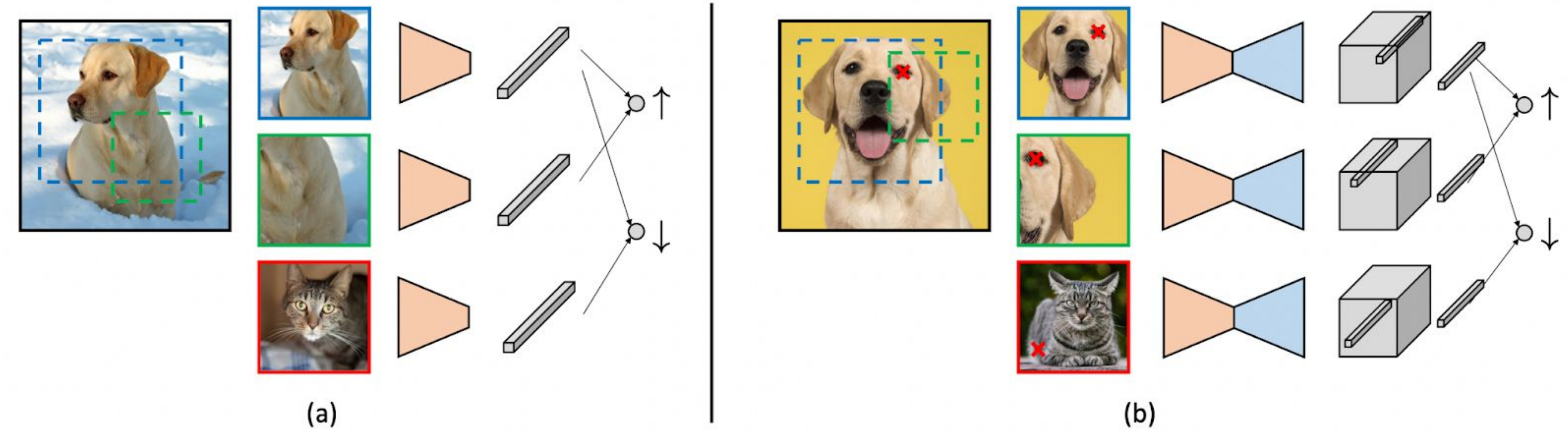


~10^6
Unlabeled
3D images



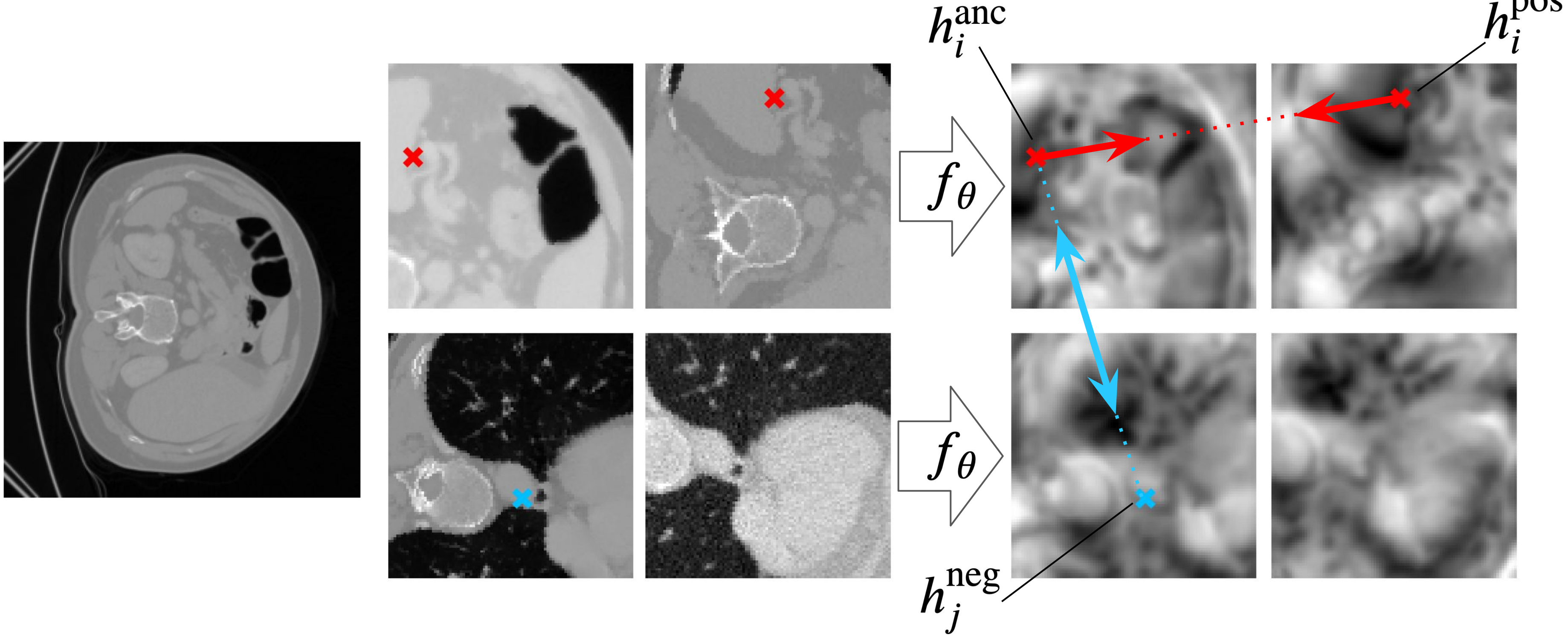
We want to exploit unlabeled images to train better algorithms

Dense vs Global Representations



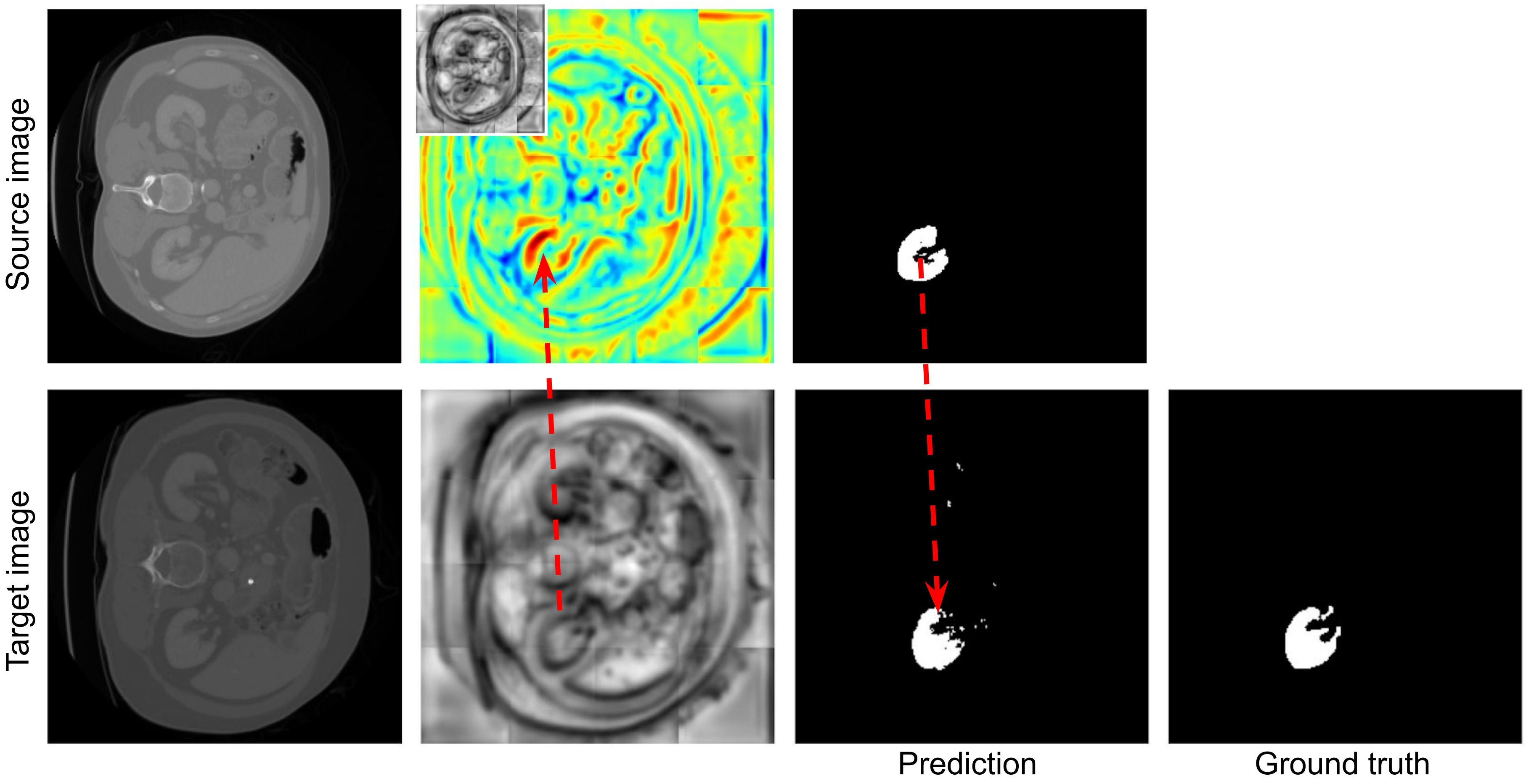
Pinheiro, Pedro O., et al. "Unsupervised Learning of Dense Visual Representations." arXiv, 11 Nov. 2020, doi:10.48550/arXiv.2011.05499.

Contrastive learning of in(equi)variant voxel-level features



$$-\sum_i \log \frac{\exp(\cos(h_i^{\text{anc}}, h_i^{\text{pos}})/\tau)}{\exp(\cos(h_i^{\text{anc}}, h_i^{\text{pos}})/\tau) + \sum_j \exp(\cos(h_i^{\text{anc}}, h_j^{\text{neg}})/\tau)} \to \min_\theta$$

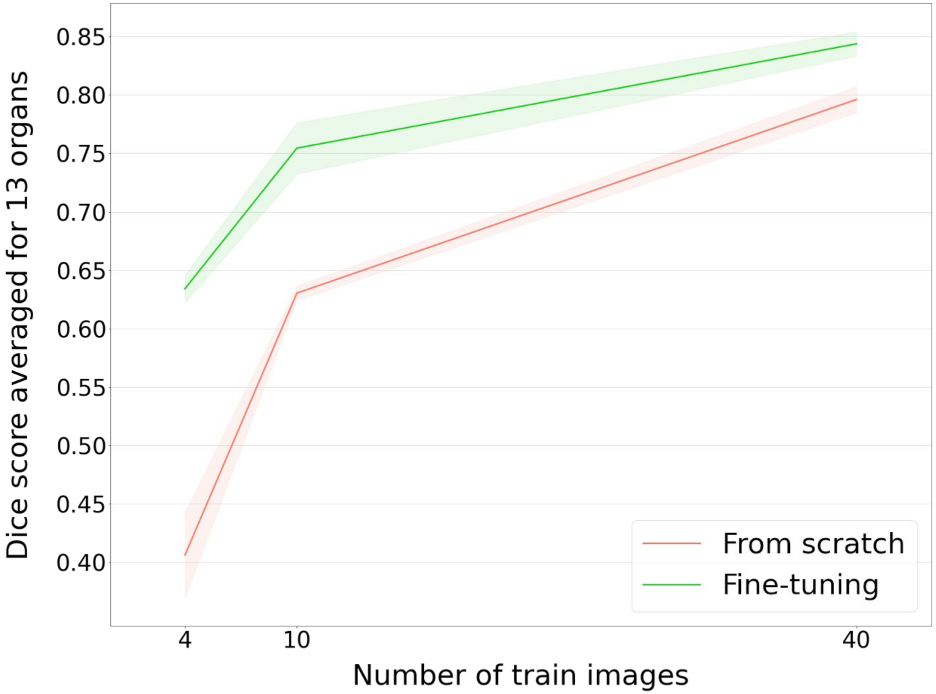
One-shot segmentation based on similarity search



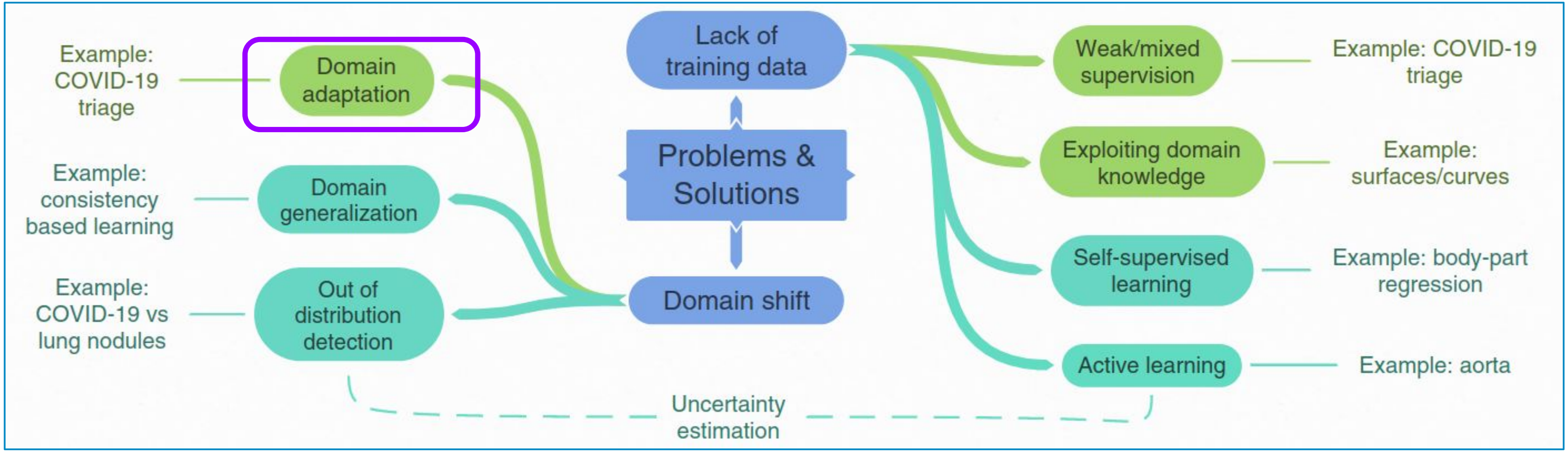
Fine-tuning vs from scratch

Impact on the industry:

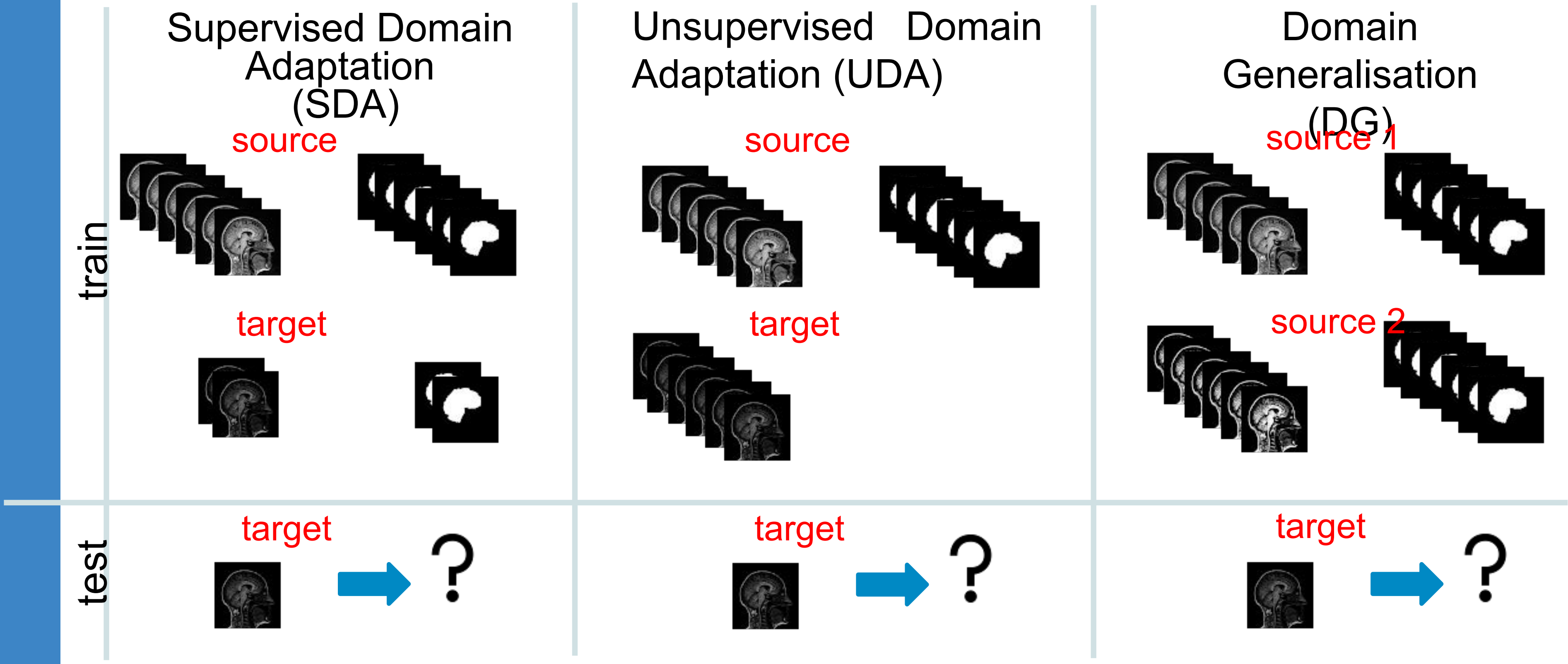
- To be estimated
- Much faster MVP development is expected



Medical CV challenges - how to address



Domain Adaptation: some possible set-ups



Anatomy of Domain Shift Impact on U-Net Layers in MRI Segmentation.
Ivan Zakazov, Boris Shirokikh, Alexey Chernyavskiy and Mikhail Belyaev. MICCAI-2021

The Data & The Metric

- We work with a publicly available dataset CC-359:
  - ➜ T1-weighted volumetric brain MRI scans; ground truth for brain segmentation
  - ➜ 3 vendors
  - ➜ Magnetic field strength of either 1.5T or 3T
  - ➜ In total: 6 domains (60 or 59 scans in each)
- Metric: the surface Dice (to capture how good the edges of the brain are segmented)

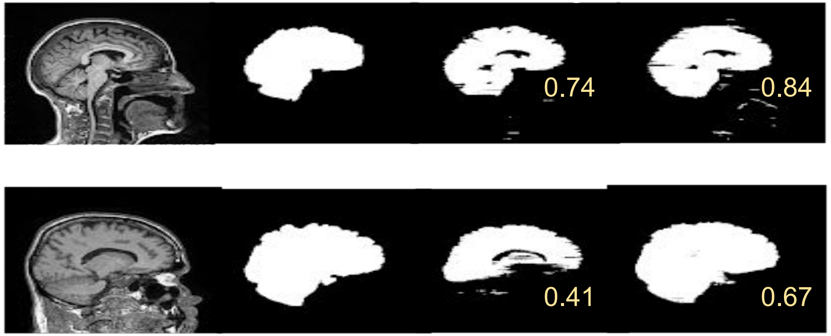


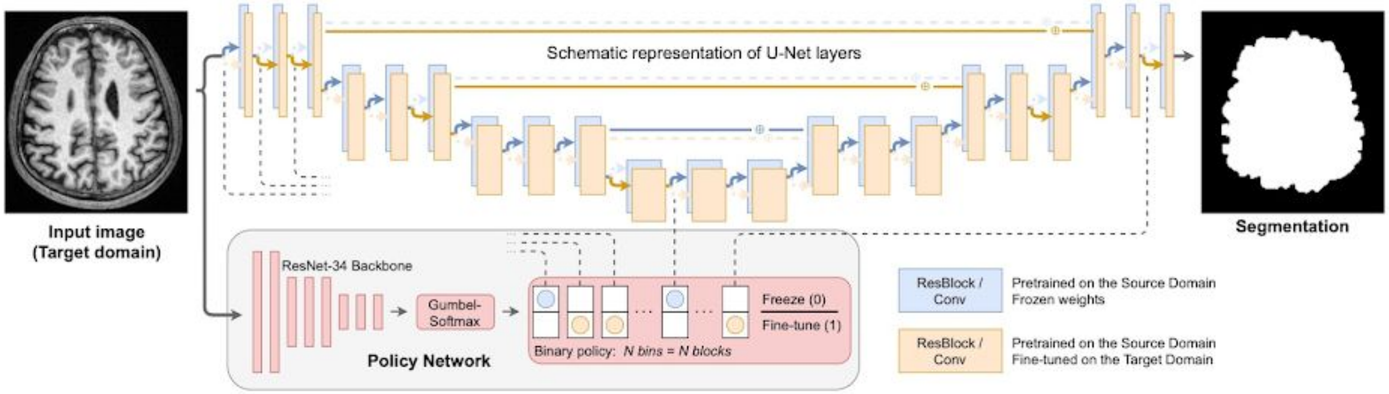
Fig. 1: *brain segmentation visualization. Columns: a) slices b) the ground truth segmentations c) baseline (no DA) d) transferring results (DA applied; fine-tuning the first layers on 1/36 of a scan). The figures: surface Dice Scores*

Anatomy of Domain Shift Impact on U-Net Layers in MRI Segmentation.
Ivan Zakazov, Boris Shirokikh, Alexey Chernyavskiy and Mikhail Belyaev. MICCAI-2021

Supervised Domain Adaptation: SpotTUnet

Case-specific regularization introduced, to account for extreme data scarcity cases:

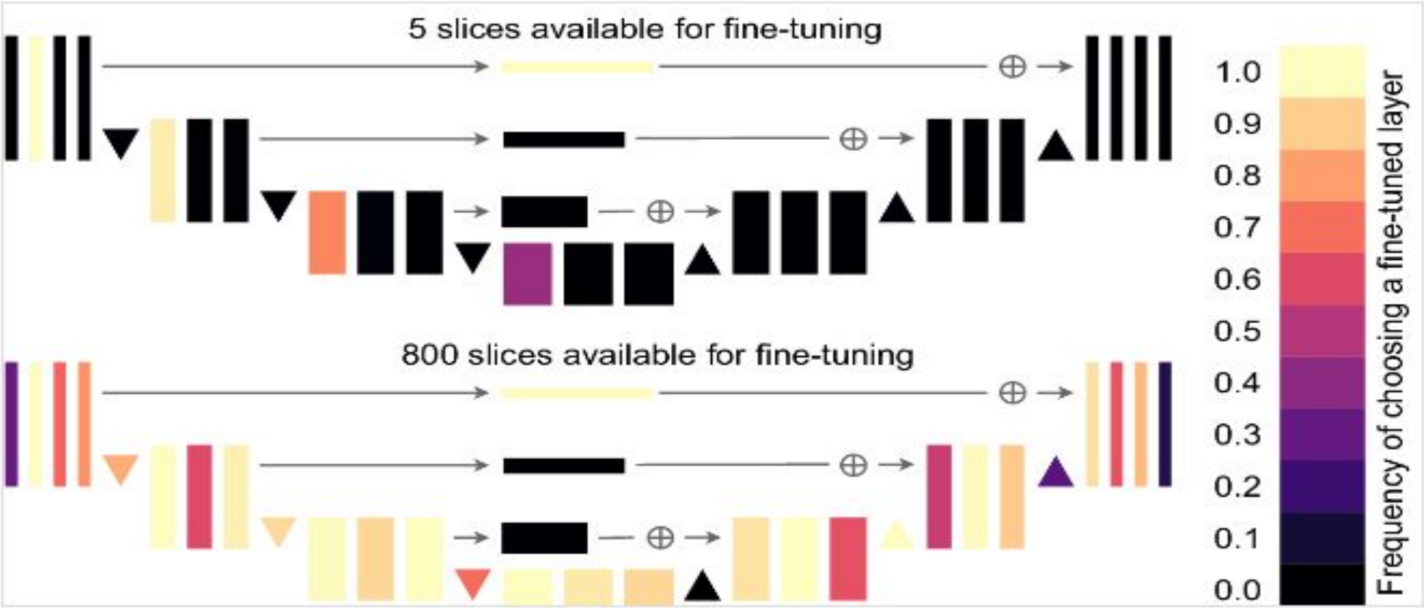
$$\mathcal{L} = \mathcal{L}_{segm} + \lambda \sum_{l=1}^{N} (1 - I_l(x))$$



SpotTune adapted for MRI segmentation (SpottU-net)

Anatomy of Domain Shift Impact on U-Net Layers in MRI Segmentation.
Ivan Zakazov, Boris Shirokikh, Alexey Chernyavskiy and Mikhail Belyaev. MICCAI-2021

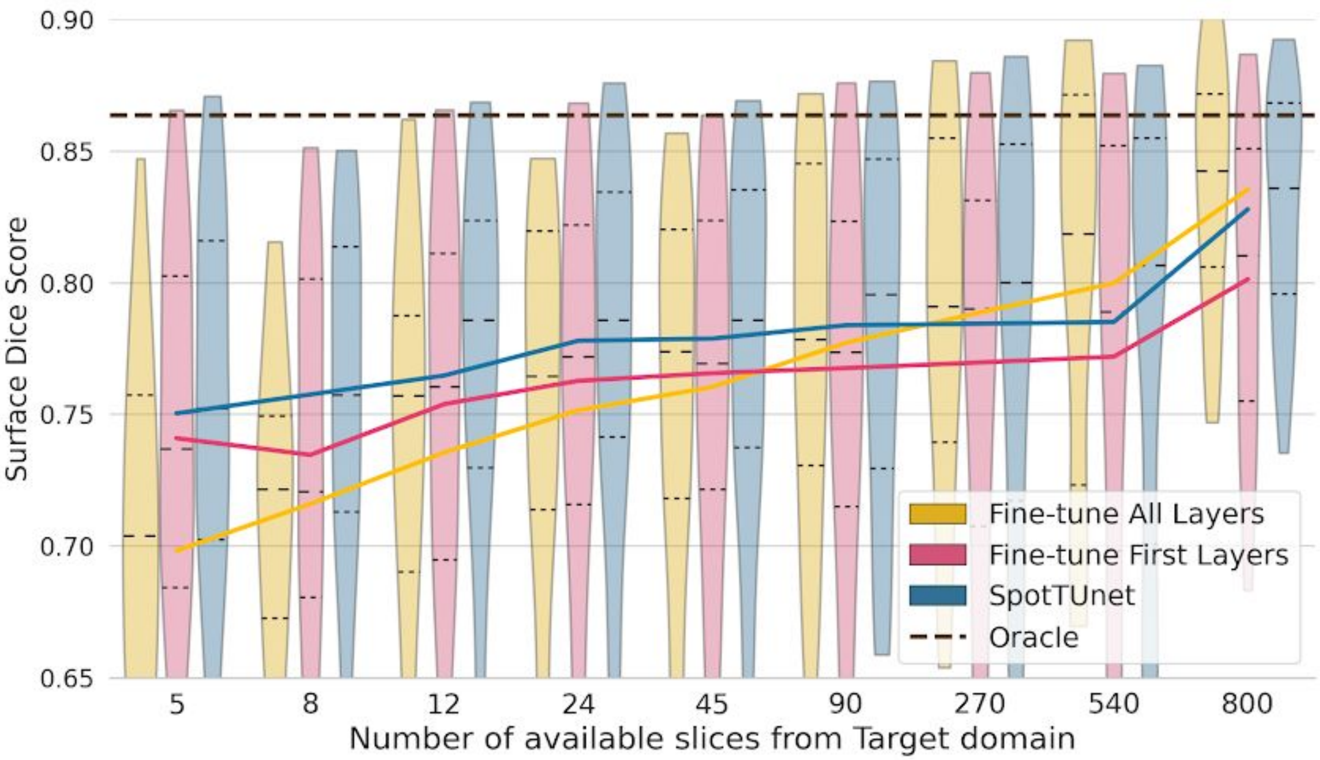
Results



Fine-tuning statistics

Anatomy of Domain Shift Impact on U-Net Layers in MRI Segmentation.
Ivan Zakazov, Boris Shirokikh, Alexey Chernyavskiy and Mikhail Belyaev. MICCAI-2021

Results

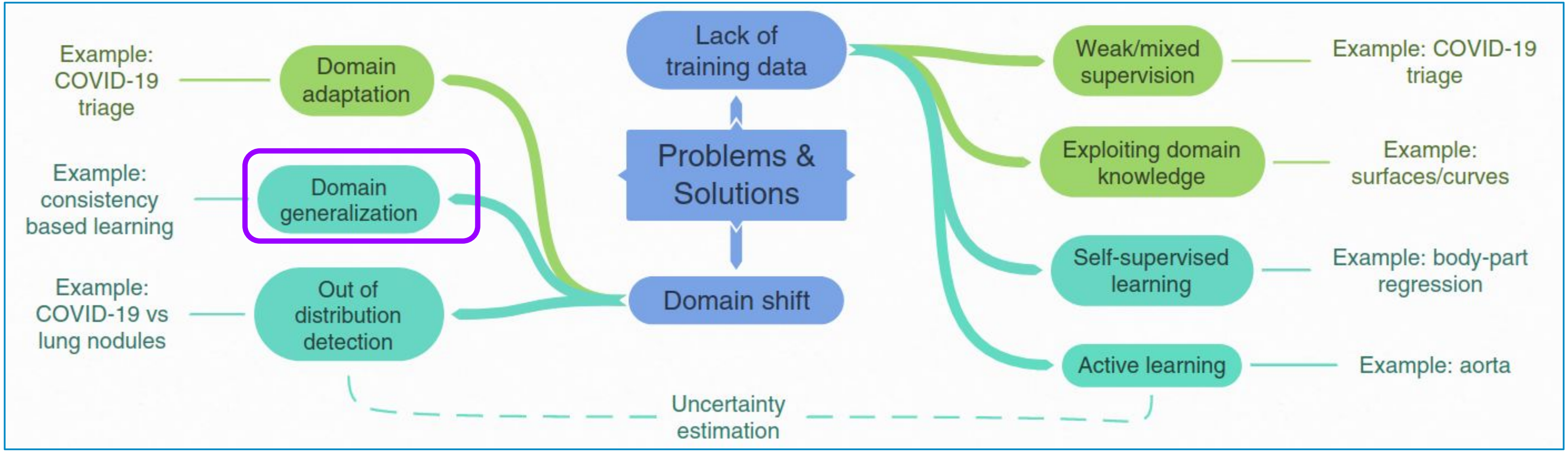


Fine-tuning approaches comparison

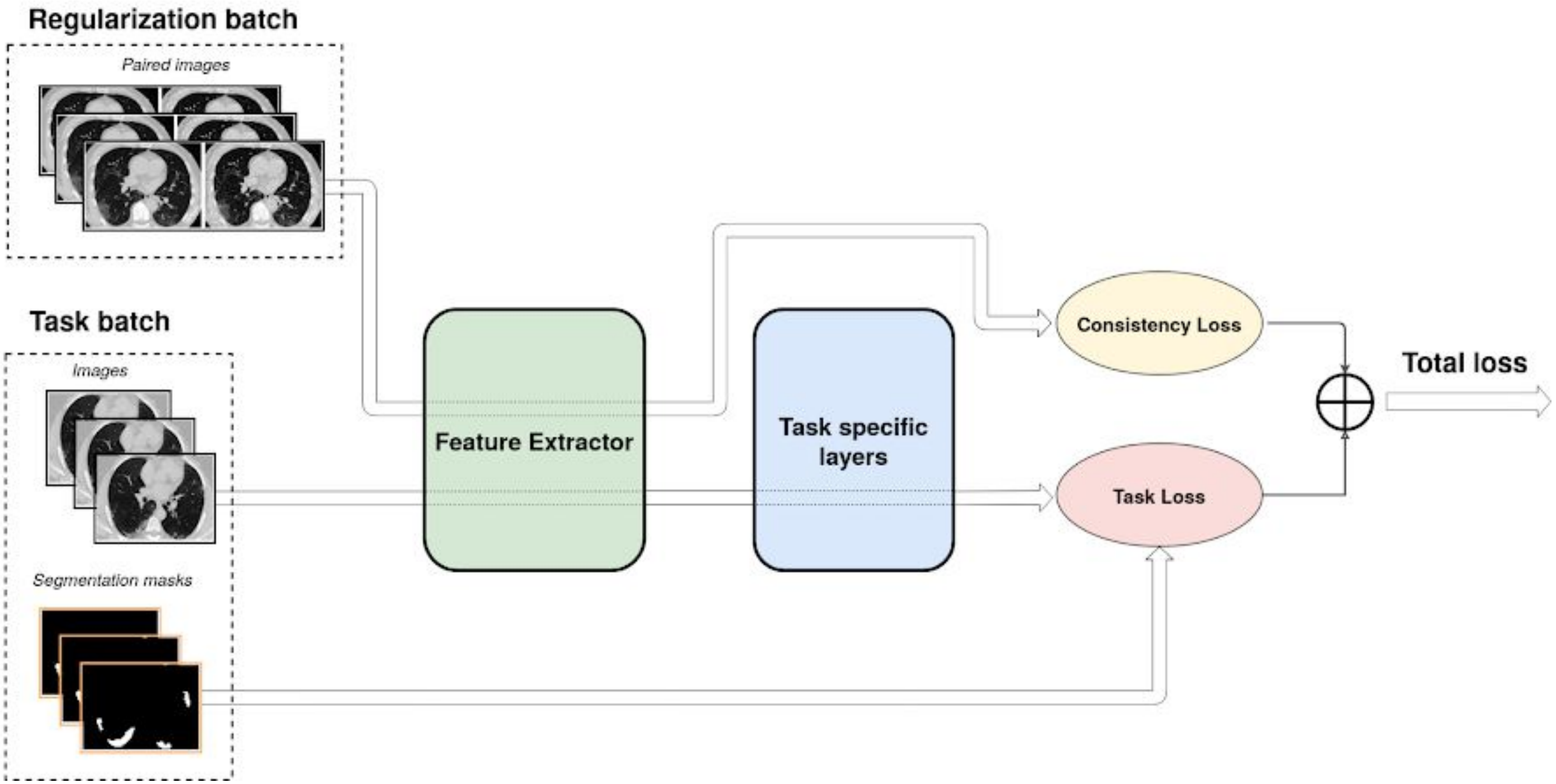
Impact on the industry:

We usually needs domain generalization (see below)

Medical CV challenges - how to address

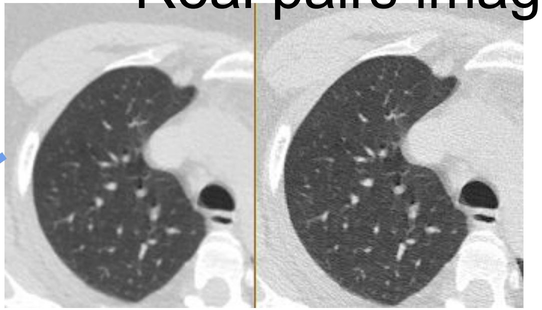


Consistency learning: idea

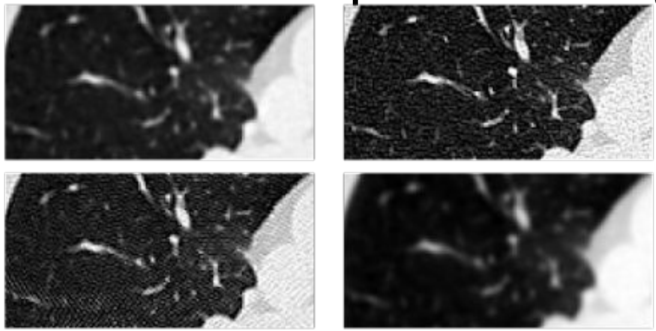


Simulating domain differences

Real pairs images



Simulated pairs of images

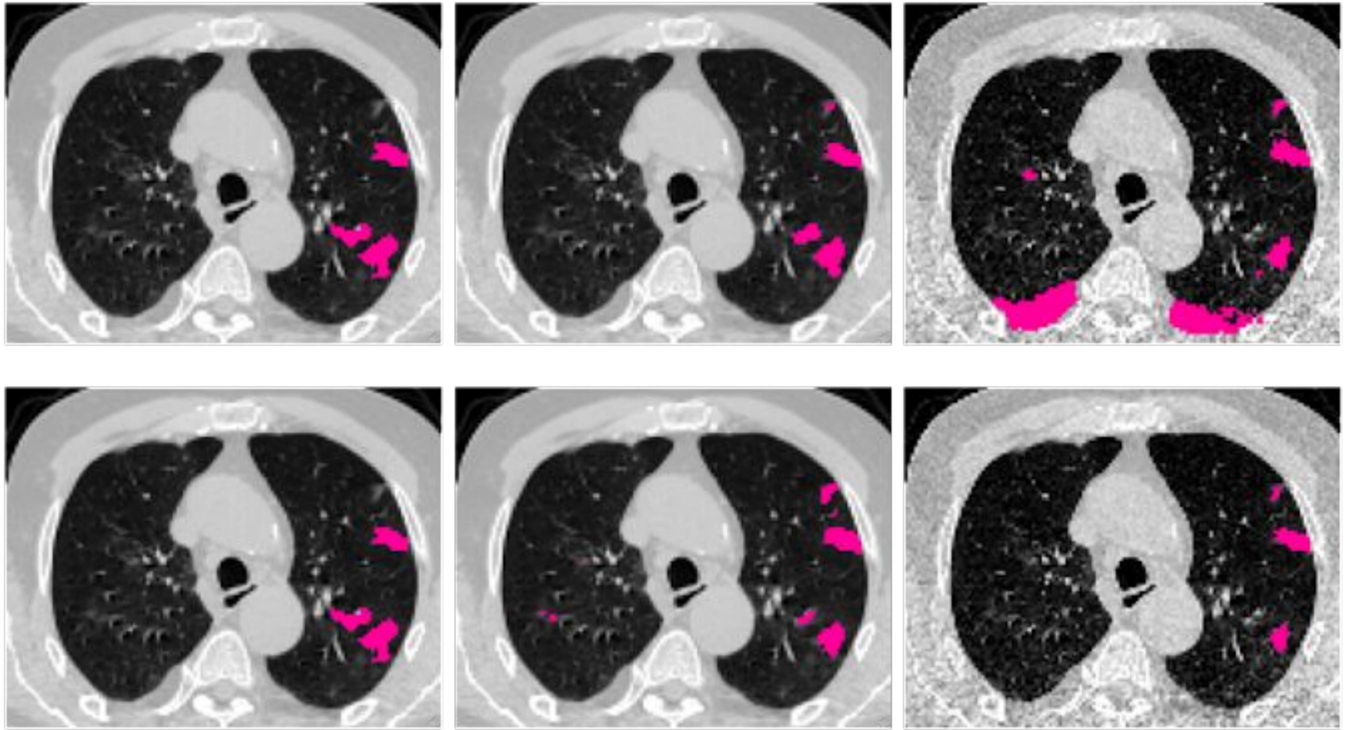


Results: segmentation

Ground truth | FC07 domain | FC07 domain

Base model

Consistency-based DA model

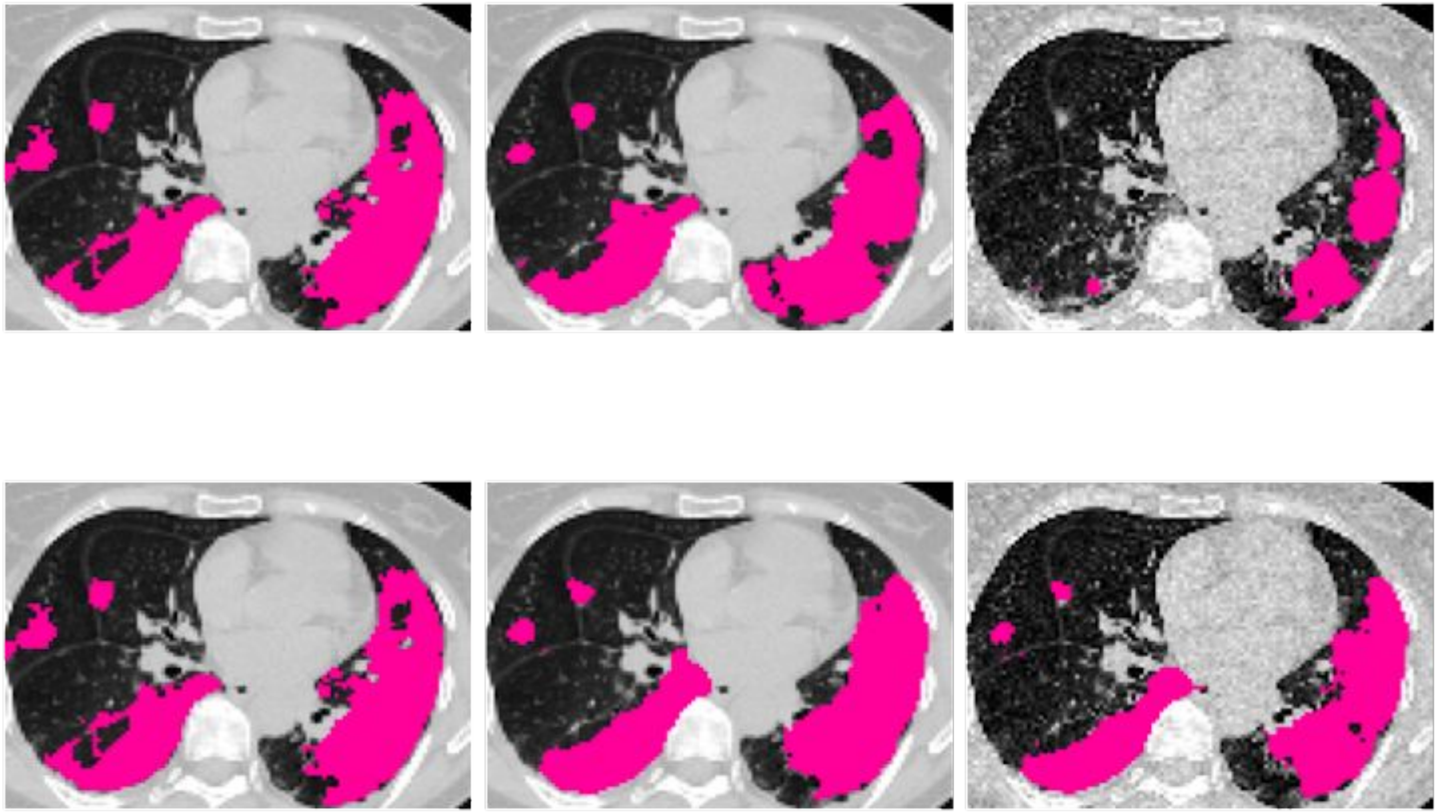


Results: segmentation

Ground truth | FC07 domain | FC07 domain

Base model

Consistency-based DA model

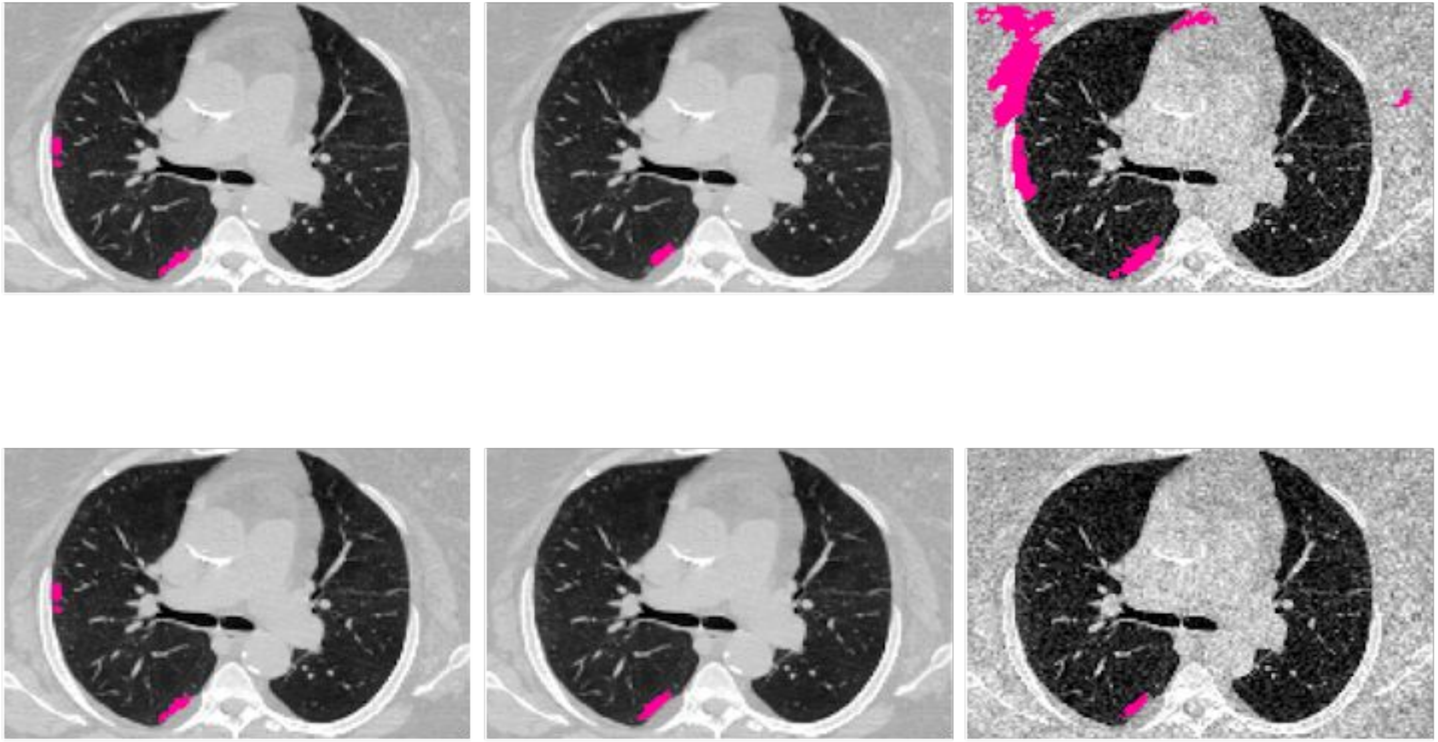


Results: segmentation

Ground truth | FC07 domain | FC07 domain

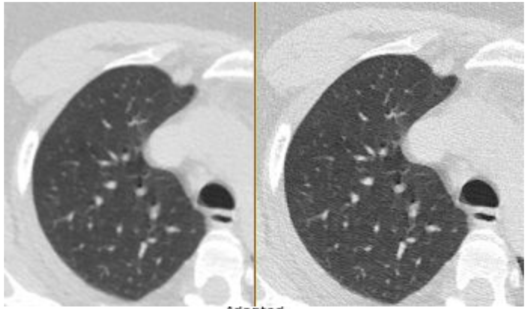
Base model

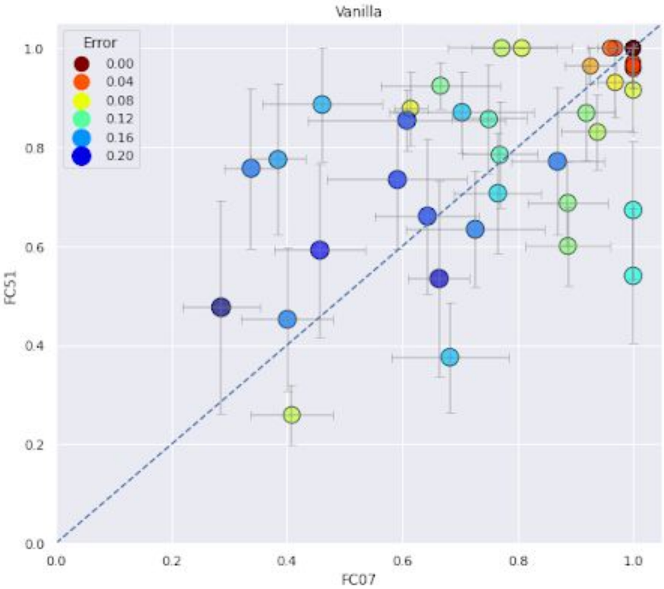
Consistency-based DA model

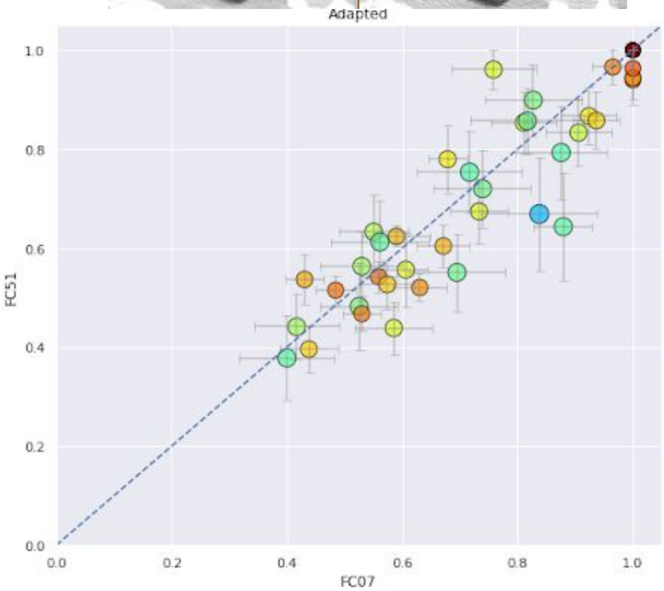


Classification: paired images

We assume that classification (COVID vs others) should stable for different image styles /domains for paired images



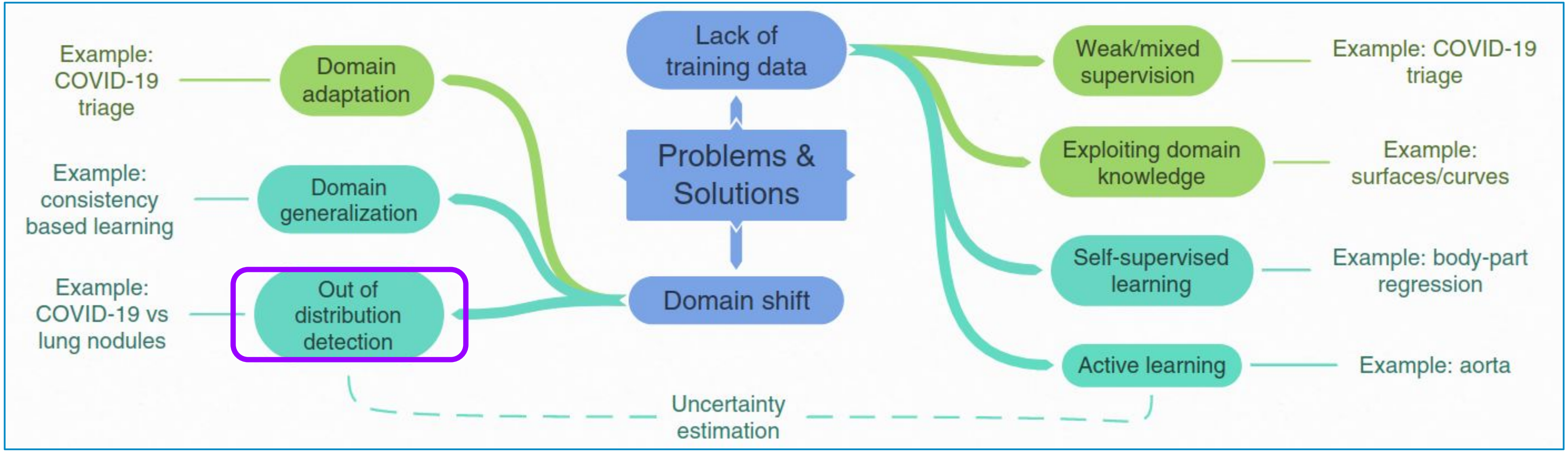




Impact on the industry:

- Up to 3x better results for specific scanning protocols (~10% of the customers)

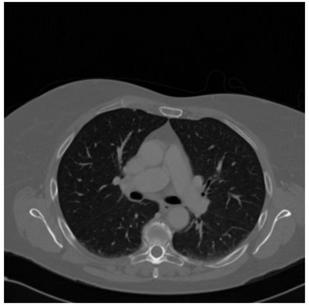
Medical CV challenges - how to address



Mahalanobis distance over the *image histogram features* vector

Step 1

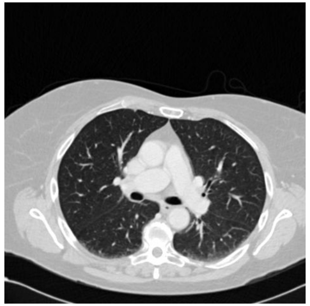
Original image



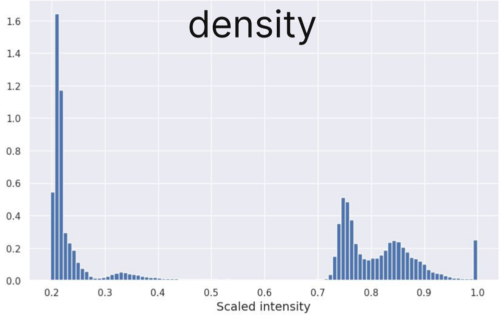
Standard preprocessing:

- HU clipping
- Scale to [0, 1]
- Zoom to 1^3 mm

Preprocessed image



m-bin histogram with normed density



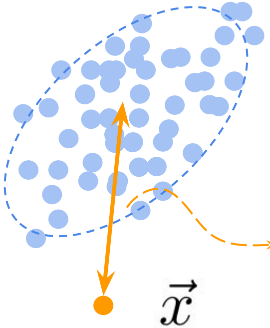
$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

m=150

In-distribution (ID) data (available on training)

Test sample (might be OOD or ID)

Step 1



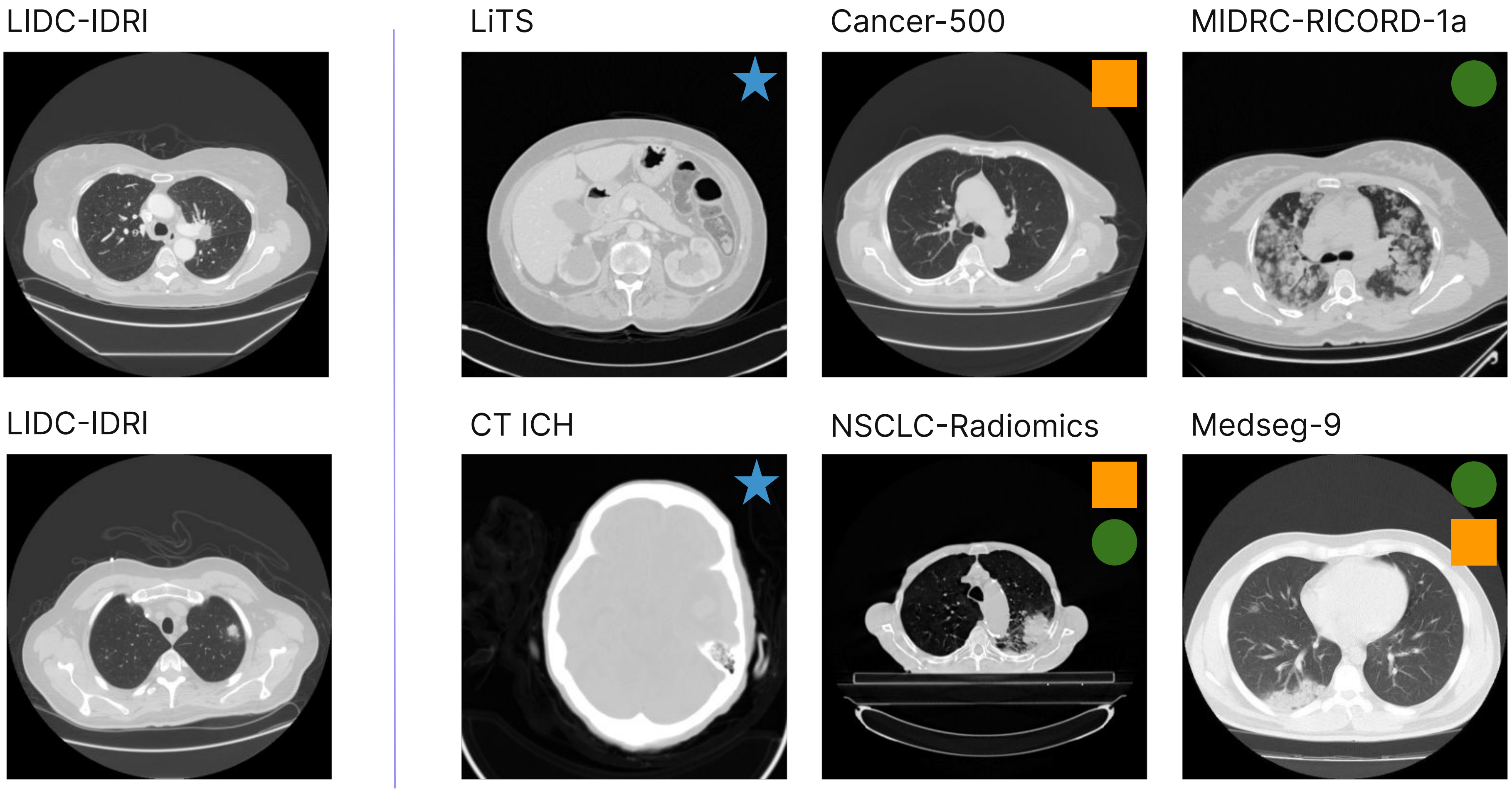
$\vec{x}$

Step 2

- mean $\vec{\mu}$
- covariance matrix $\mathbf{\Sigma}$

OOD score $\vec{x}$) = $\sqrt{(\vec{x}-\vec{\mu})^T \, \Sigma^{-1} \, (\vec{x}-\vec{\mu})}$

AIRI



- another body-part
- acquisition protocol
- patient population

Deep Learning-based methods

ID dataset	OOD dataset	Baseline	MC-Dropout	Deep Ensembles	GODIN	SVD	IHF (ours)
LIDC (CT)	CT-ICH	.646	.842	.955	.408	.999	**1.00**
	LiTS	.407	.584	.783	.277	.883	**.994**
	Cancer500	.505	.547	.516	.536	.592	**.994**
	NSCLC	.686	.915	.770	.515	.778	**.992**
	MIDRC	.797	.889	.441	.773	.702	**.996**
	Medseg9	.674	.815	.512	.813	.894	**.999**
VS-SEG (MRI)	CC359	.363	.632	.150	.112	.979	**1.00**
	EGD filtered	.364	.465	.192	.104	.933	**1.00**
	CrossMoDA LDN	.468	.441	.467	.475	.156	**.984**
	CrossMoDA ETZ	.755	.797	.101	.736	**1.00**	**1.00**

Table 2. AUROC in the OOD detection tasks.

Today's MOOD challenge results (MICCAI)

CitAI	1
NeuroML	2
HiWJ	3

Out of domain detection