



Universal and Foundational Models of Medical Imaging AI : Traits and Trends

S Kevin Zhou, Professor, PhD

University of Science and Technology of China
Fellow of AIMBE | IAMBE | IEEE | MICCAI | NAI
2025.11.27



中科大苏州

AI Hype



Dr. Geoffery Hinton

“We should stop training radiologists now. It’s just completely obvious that within **five years**, deep learning is going to do **better than** radiologists.”

-- New Yorker, 2016

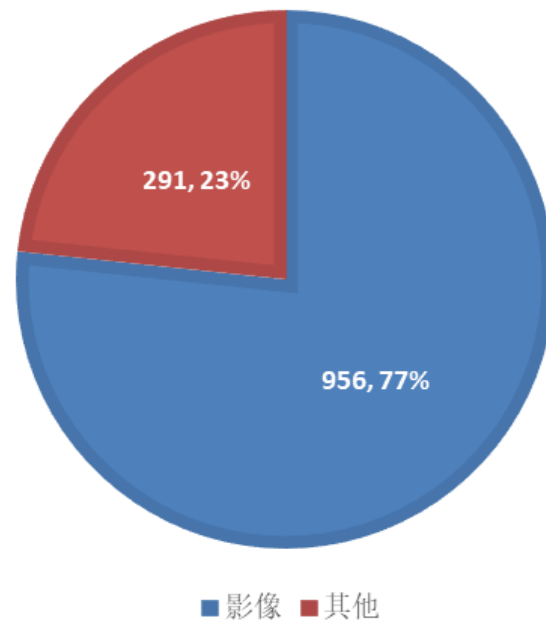


Dr. Curt Langlotz

“... radiologists who use AI will replace radiologists who don’t.”

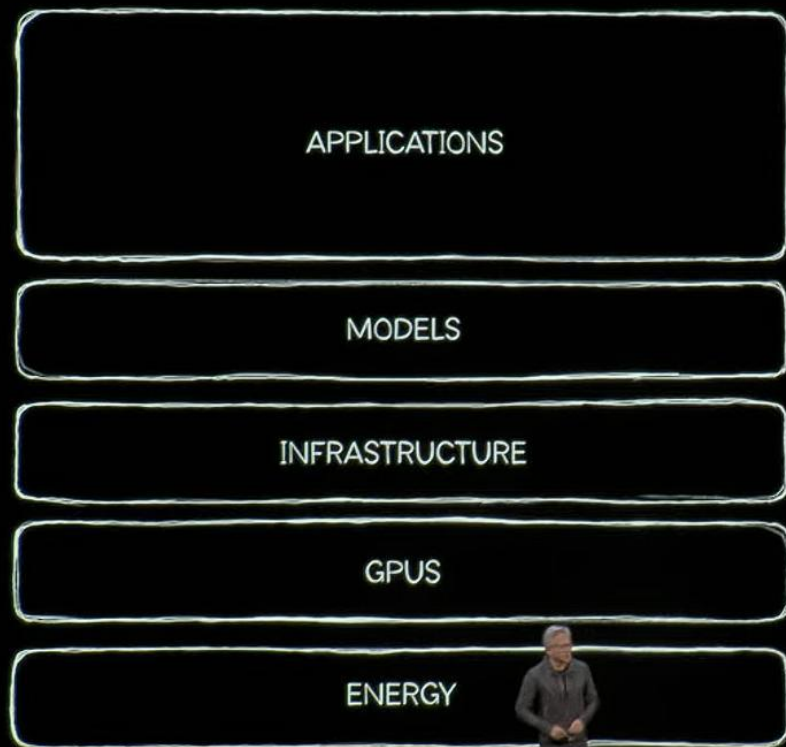
@RSNA, 2017

AI-Enabled Medical Devices



- <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>
- 1247 in total (till 2025.05.30)

AI - THE NEW INDUSTRIAL REVOLUTION



CHATBOTS



DIGITAL BIOLOGY



ROBOTAXI



ENTERPRISE
AI AGENTS



SCIENCE



ROBOTICS



MANUFACTURING



AI CODER

LLM

VLM

VLA

MMLLM

GPT

DM

GNN

MOE

SSM

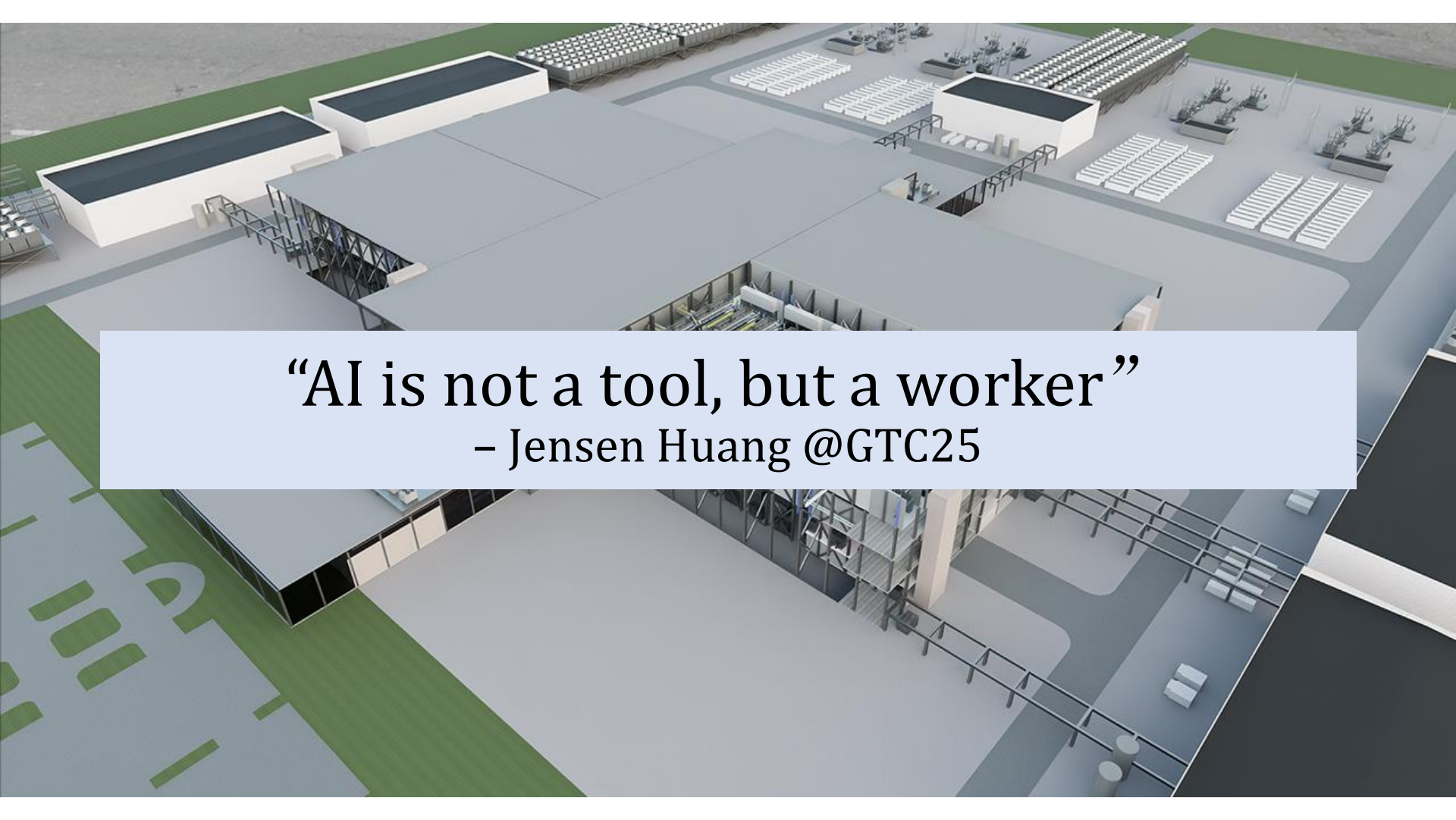
LBM

AI FACTORIES



AI Factory Era

“... generate a massive scale of tokens for all applications ...”



“AI is not a tool, but a worker”
– Jensen Huang @GTC25

Model complexity
Task complexity

AGENTIC AI
CODING ASSISTANT
CUSTOMER SERVICE
PATIENT CARE

GENERATIVE AI
DIGITAL MARKETING
CONTENT CREATION

PERCEPTION AI
SPEECH RECOGNITION
DEEP RECSYS
MEDICAL IMAGING

2012 ALEXNET



NVidia
GTC2025

Discriminative AI
Task modeling : $P(C|I)$

Model complexity
Task complexity

AGENTIC AI
CODING ASSISTANT
CUSTOMER SERVICE
PATIENT CARE

GENERATIVE AI
DIGITAL MARKETING
CONTENT CREATION

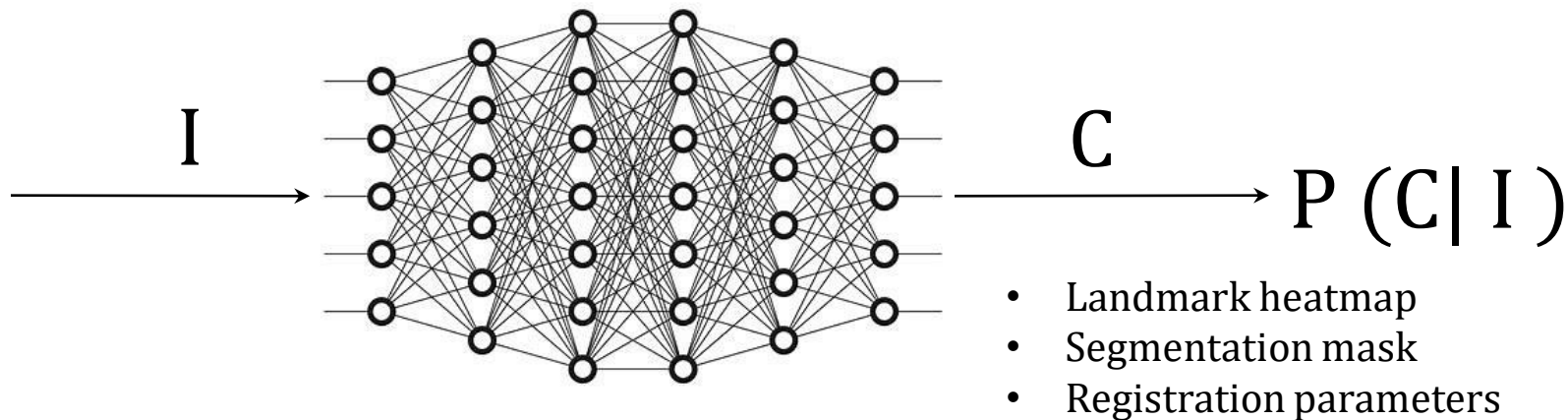
PERCEPTION AI
SPEECH RECOGNITION
DEEP RECSYS
MEDICAL IMAGING

2012 ALEXNET



NVidia
GTC2025

Discriminative AI



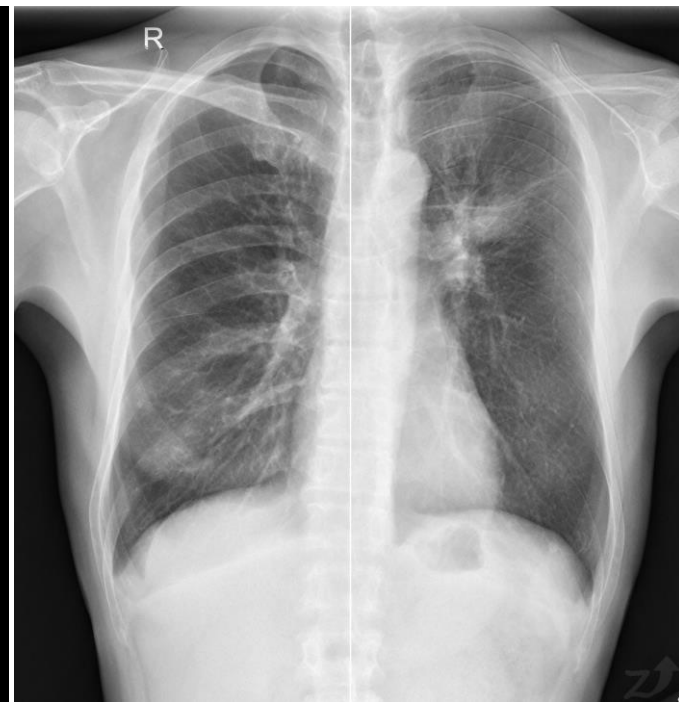
Rib analysis



CT rib unfolding

- Fracture detection rate up $\sim 10\%$
- Reading time down $\sim 50\%$

[Wu et al. CVPR'12]



Xray rib suppression

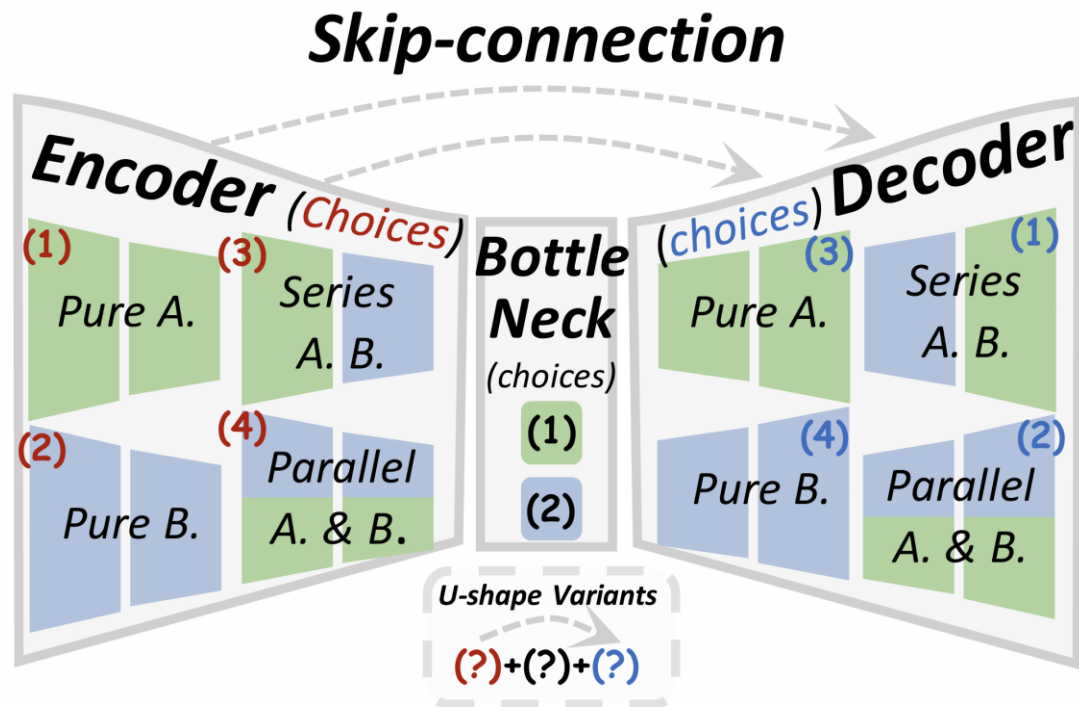
- FPR down $\sim 7\%$
- Reading time down $\sim 27\%$

[Li et al. TMI'20, Han et al. MIA'22]



U-Net

Medical Image
Segmentation



Support 🤖:

- ✓ Zero-shot evaluation
- ✓ Statistical Analysis
- ✓ Edge-aware Metric

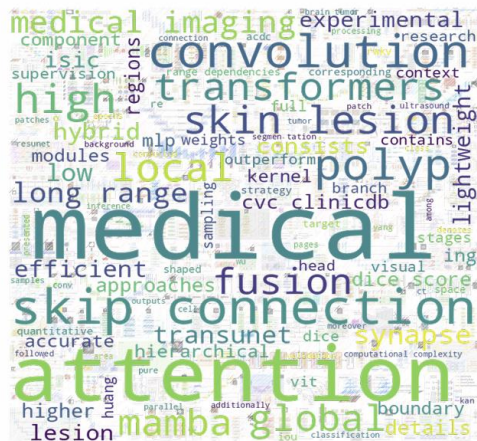
🤖 **Model
Advisor Agent**

🤖 **28 datasets**

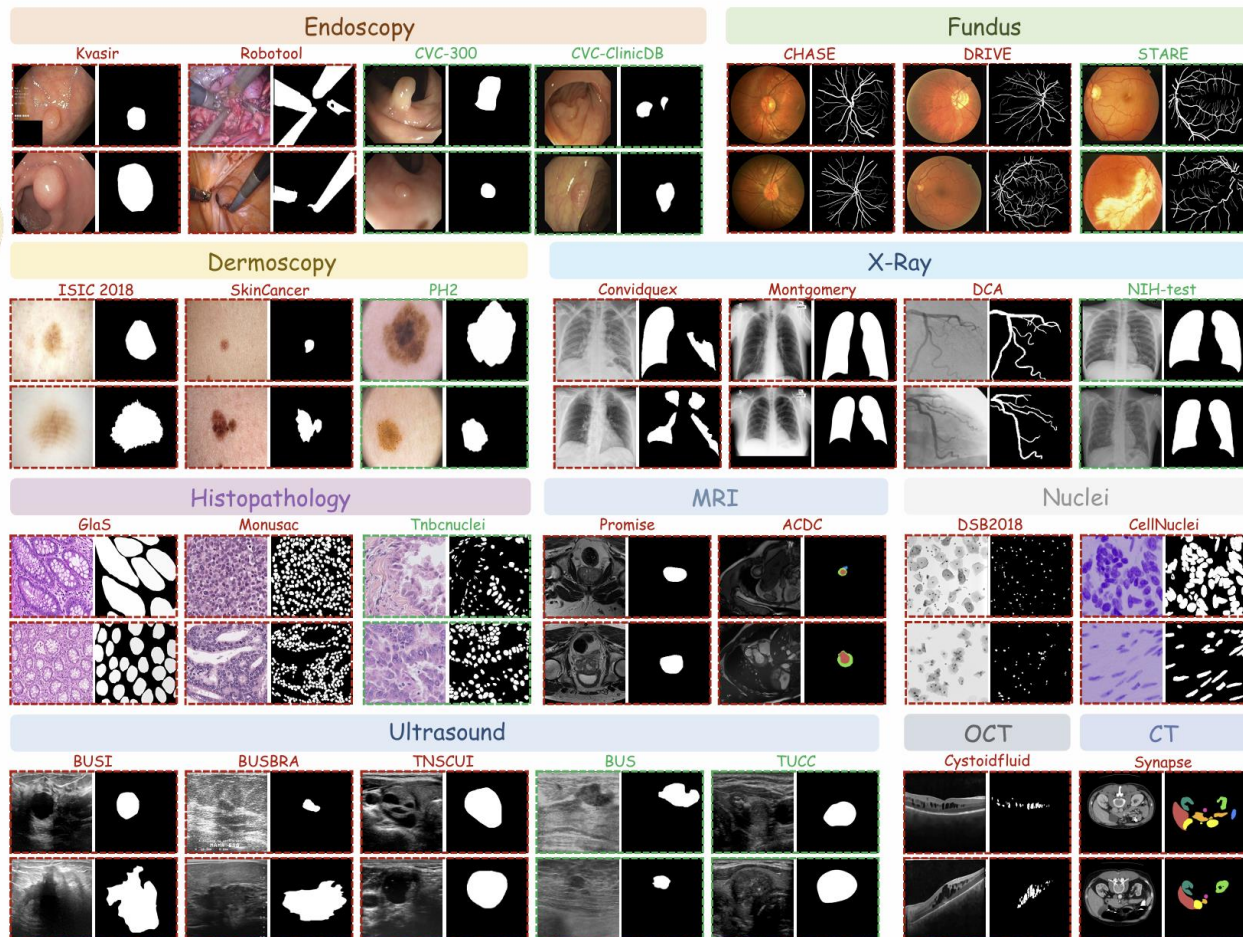
🤖 **100 networks**



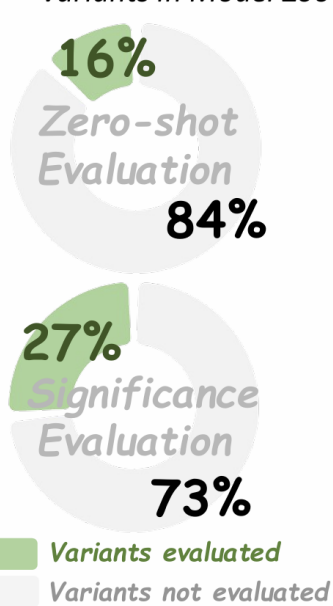
(A) Summary of **U-Bench**



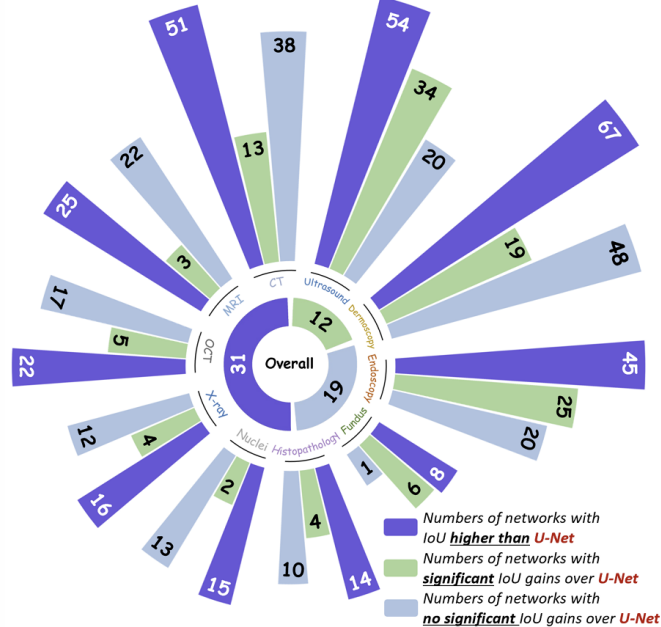
(B) Word cloud of 100 u-shape variants in Model Zoo



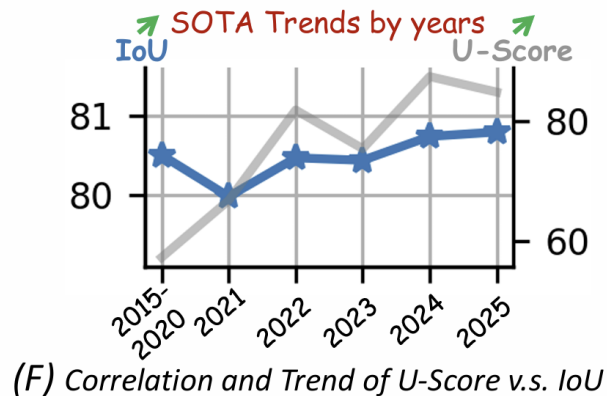
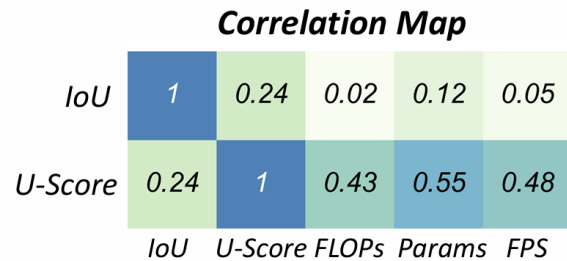
(C) Overview of the 28 datasets (**In-domain** / **Zero-shot**) across modalities in Data Zoo



(D) Few variants include zero-shot or significance evaluations



(E) Higher / Significance / no Significance than U-Net across different modalities



Generalization

Zero-shot Evaluation

Robustness

Statistical Significance Evaluation

Efficiency

Trade-off Metric: U-Score

Dataset Characteristics

foreground scale

Boundary sharpness

Shape complexity

Modality

Multi-Architecture

CNN

Mamba

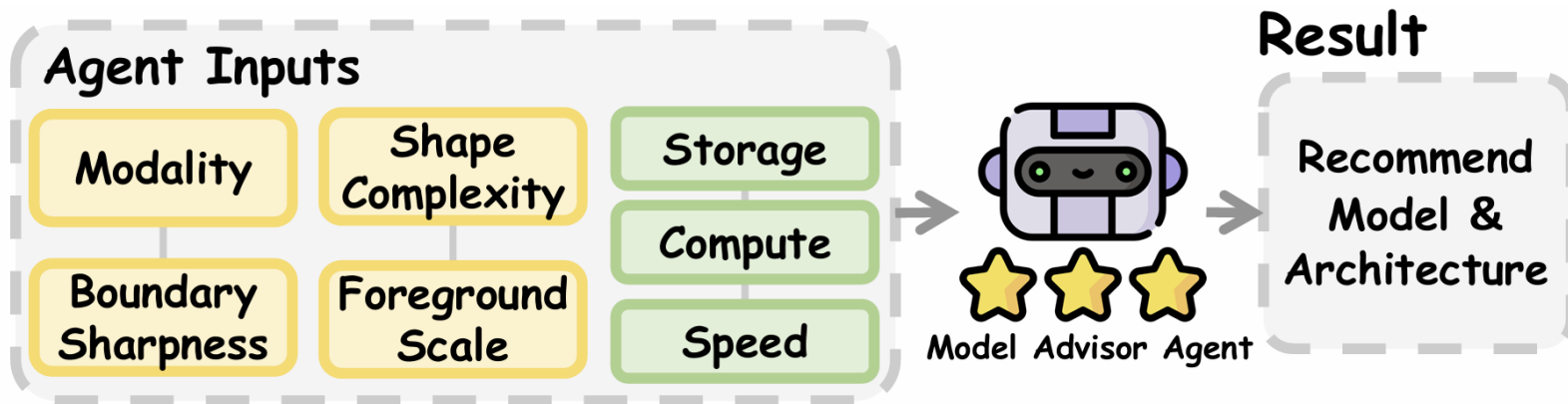
RWKV

Transformer

Hybrid

(G) The evaluation and analysis aspects covered in U-Bench.

Model Advisor Agent



Discriminative AI
Task modeling : $P(C|I)$

VS

Generative AI
Domain modeling: $P(I)$ or $P(I|C)$
 $P(C|I) \propto P(I|C) P(C)$

Model complexity
Task complexity

AGENTIC AI
CODING ASSISTANT
CUSTOMER SERVICE
PATIENT CARE



2012 ALEXNET

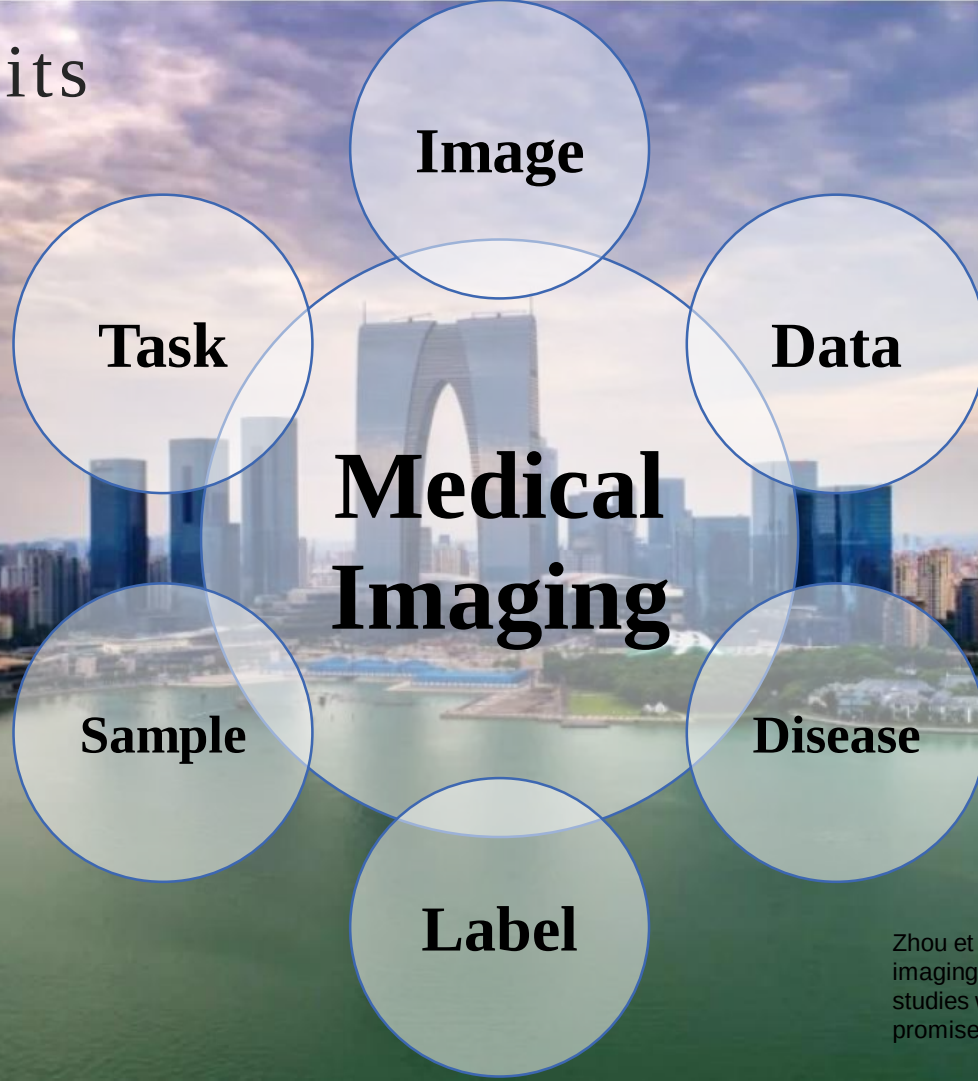


PERCEPTION AI
SPEECH RECOGNITION
DEEP RECSYS
MEDICAL IMAGING



NVidia
GTC2025

Imaging Traits



Zhou et al. A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises. Proceedings of IEEE. 2021.

Xray



MRA



MRI



PET



CT

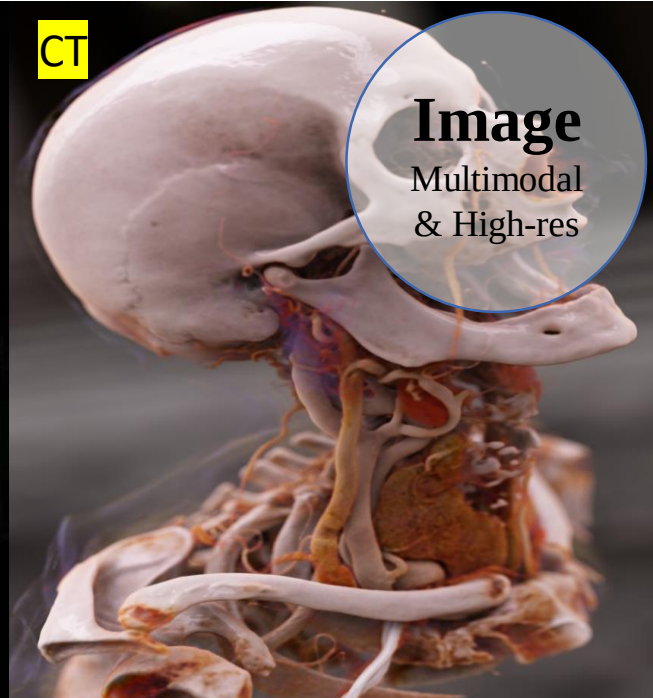


Image
Multimodal
& High-res

US





Data
non-
standard &
isolated

 Gamuts

Breast Imaging

Cardiac Radiology

Chest Radiology

Computed Tomography

Diagnostic Radiology

Gastrointestinal Radiology

Genitourinary Radiology

Head and Neck Radiology

Magnetic Resonance Imaging

Musculoskeletal Radiology

Neuroradiology

Obstetric/Gynecologic Radiology

Oncologic Imaging

Other

Pediatric Radiology

Ultrasound

Vascular Imaging

GAMUT OF THE DAY

Lymphangiectasia





CAUSED BY

- Infection
- Stewart-Treves syndrome
- Tuberculosis
- Metastasis to lymph node
- plus 12 more

Disease
Long-tail
& bursting

The Long Tail

Number of results

← more generic more specific →

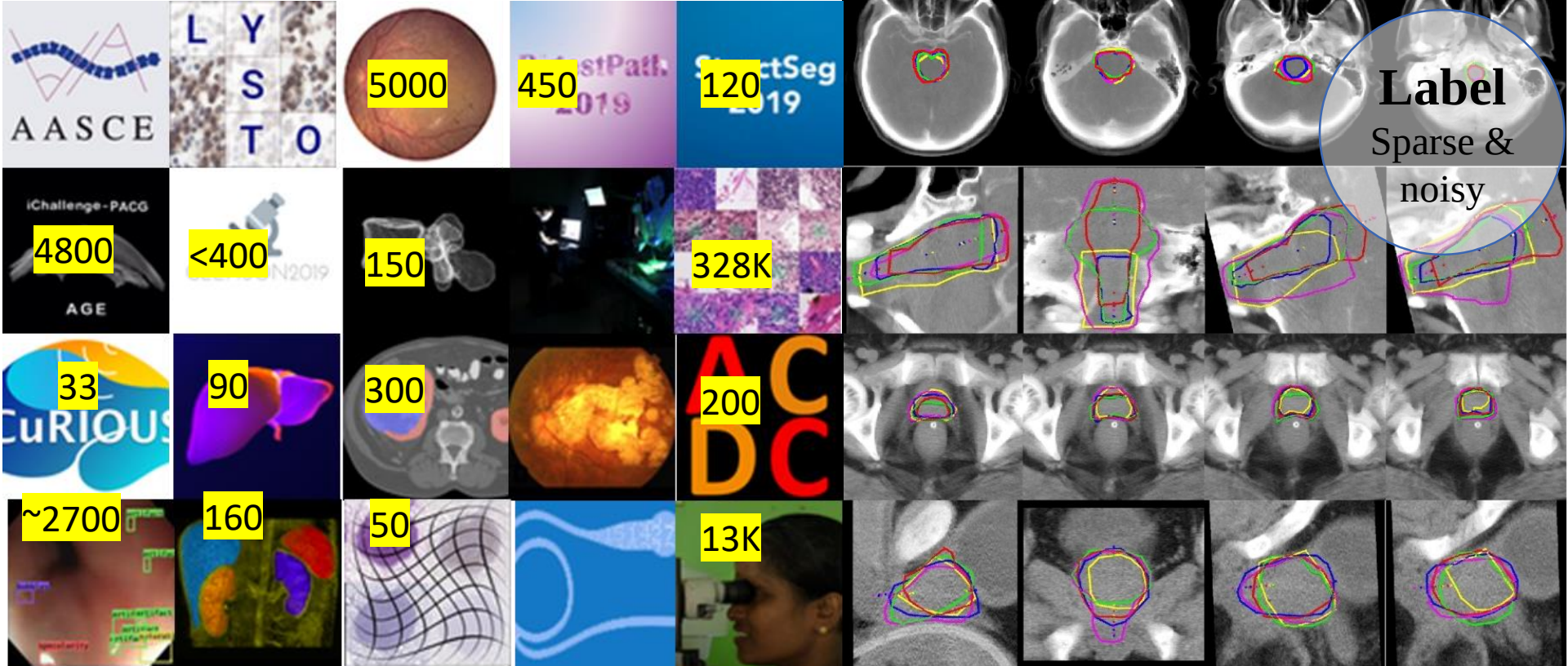
a small amount of diseases

a large amount of data

a large amount of diseases

16,912 diseases & imaging findings

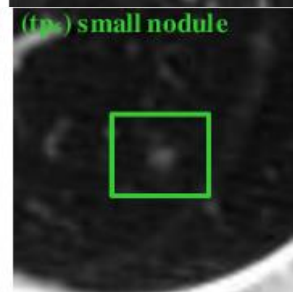
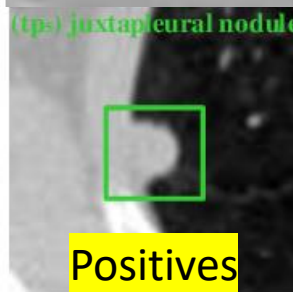
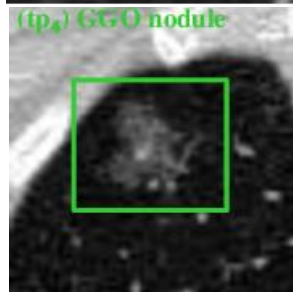
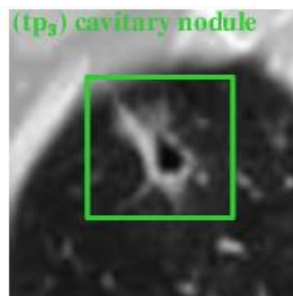
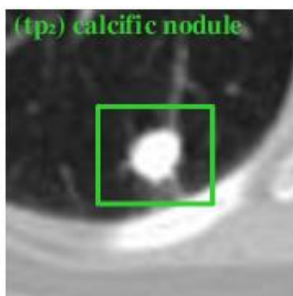
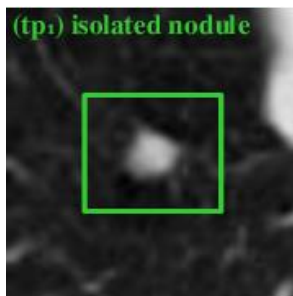
a small amount of data



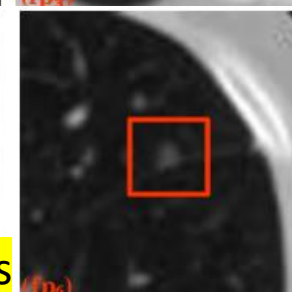
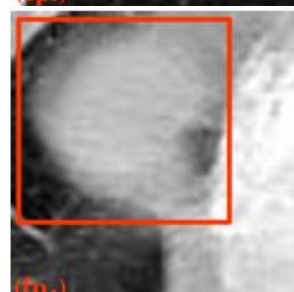
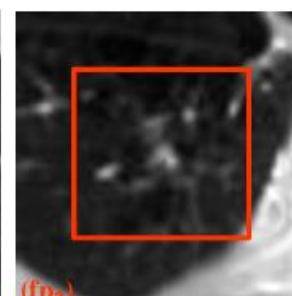
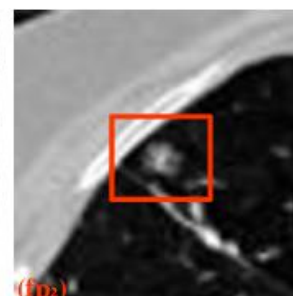
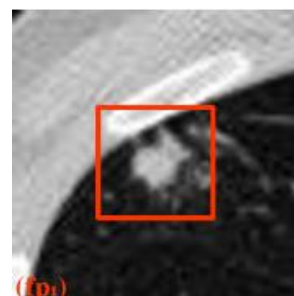
- RSNA2019 Intracranial Hemorrhage Detection: 25,000+ brain CTs
- Stanford: CheXpert: 224,316 chest radiographs of 65,240 patients
- NIH Chest X-ray dataset: 112,120 X-ray images from 30,805 patients
- Deep lesion: 32,000 images from 4400+ unique patients
- CT Pevic 1K, CT Spine 1K, CT Film 20K

- ~ 15% of reports don't accurately describe images
- 31–37% discordant interpretations
- NLP traditional “expert systems”

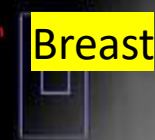
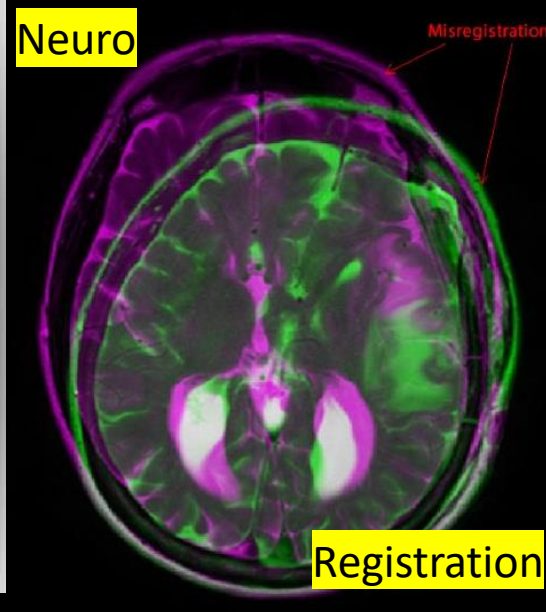
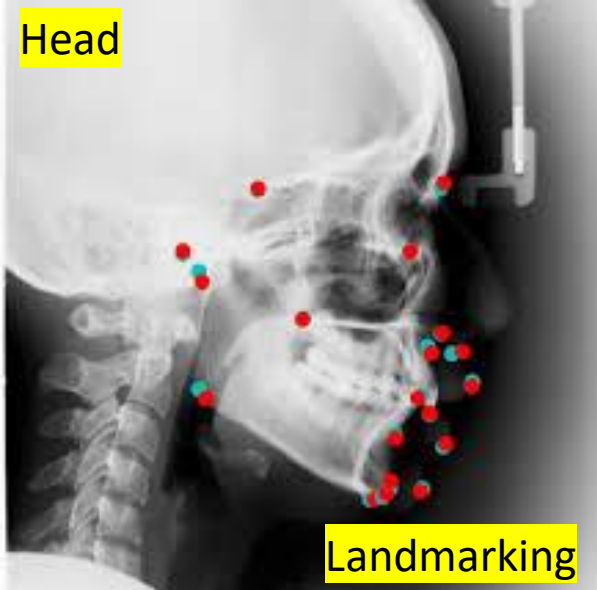
Sample
Hetero-
geneous &
imbalanced



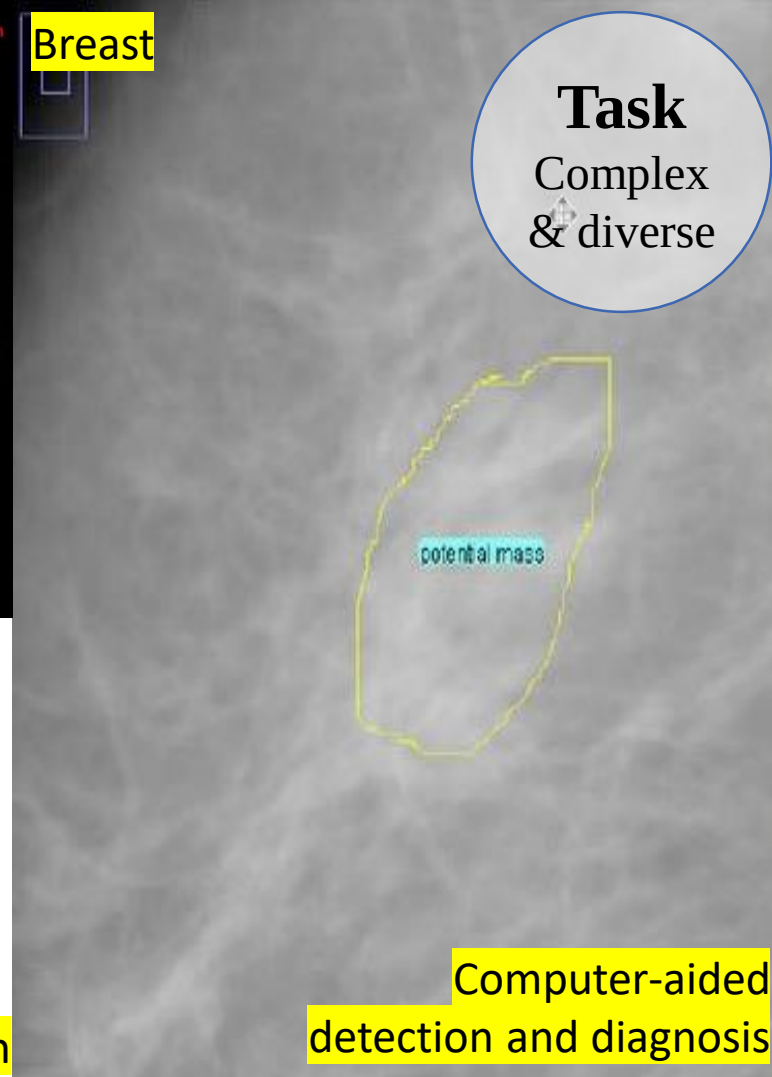
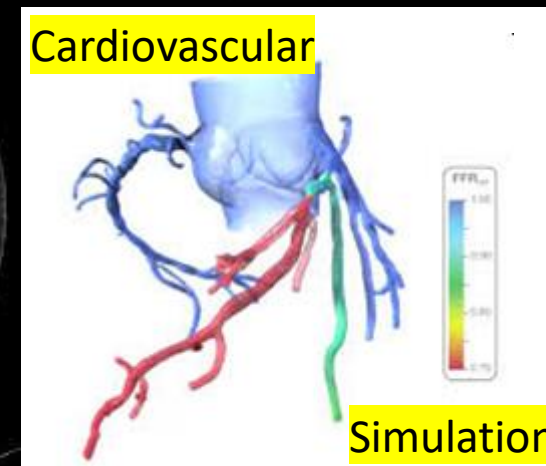
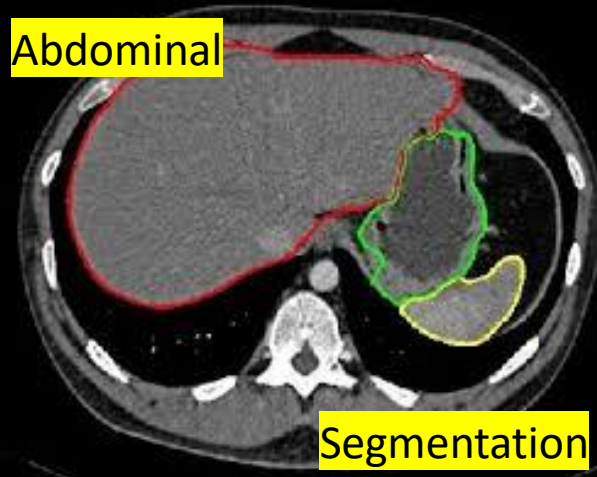
Positives



Tough negatives



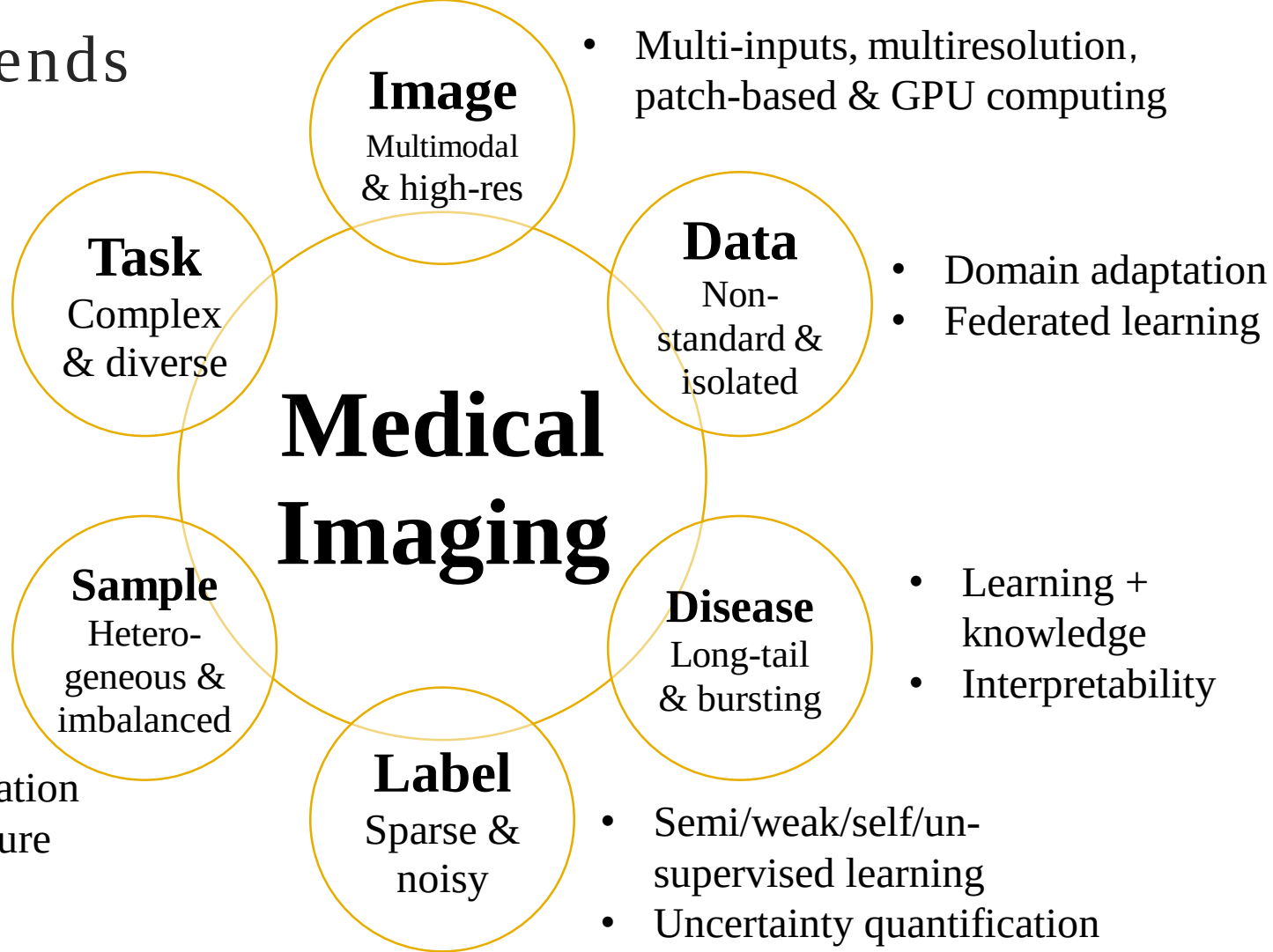
Task
Complex
& diverse

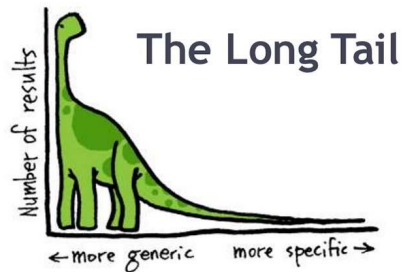
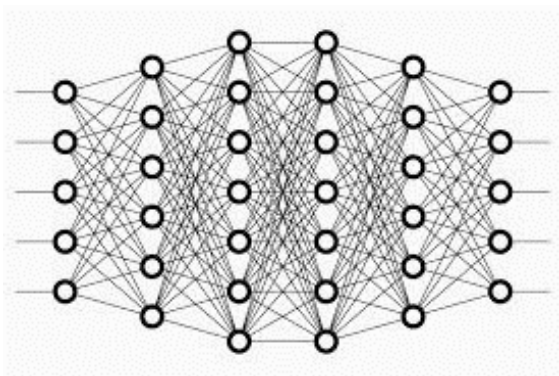


Traits & Trends

- Universal representation learning (URL)

- Adversarial generation
- Network architecture



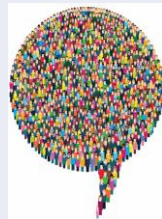
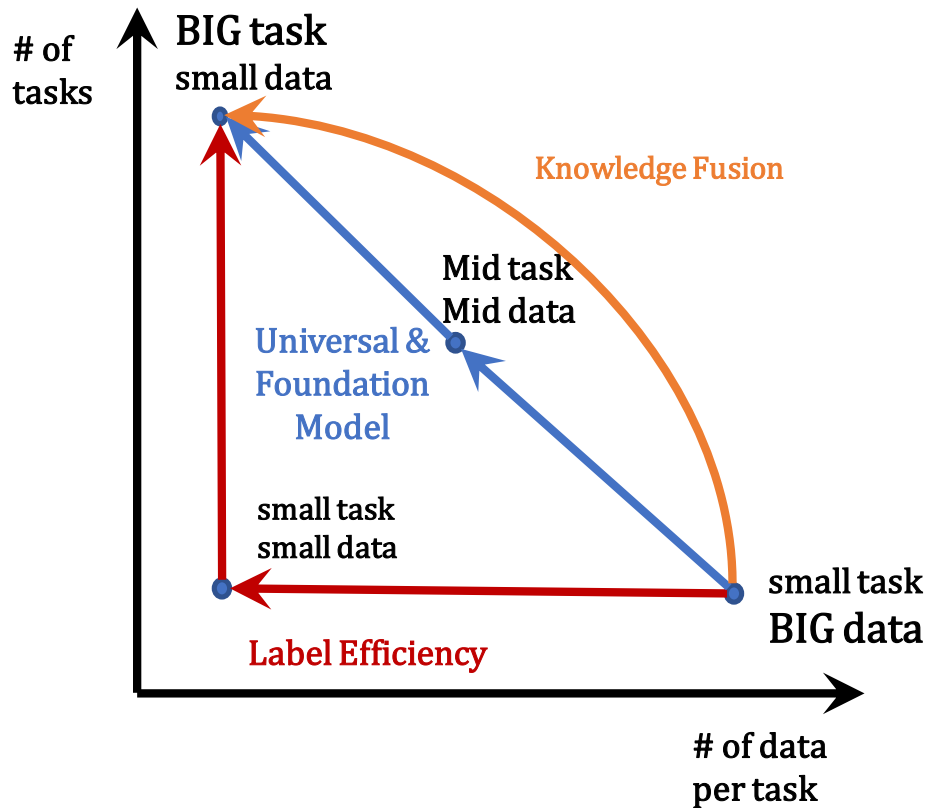


SMALL task
BIG data
“Perception AI”



BIG task
 SMALL data
“Generative AI”

Three technical paths



Label Efficiency

Universal &
Foundation Model



Knowledge
is power

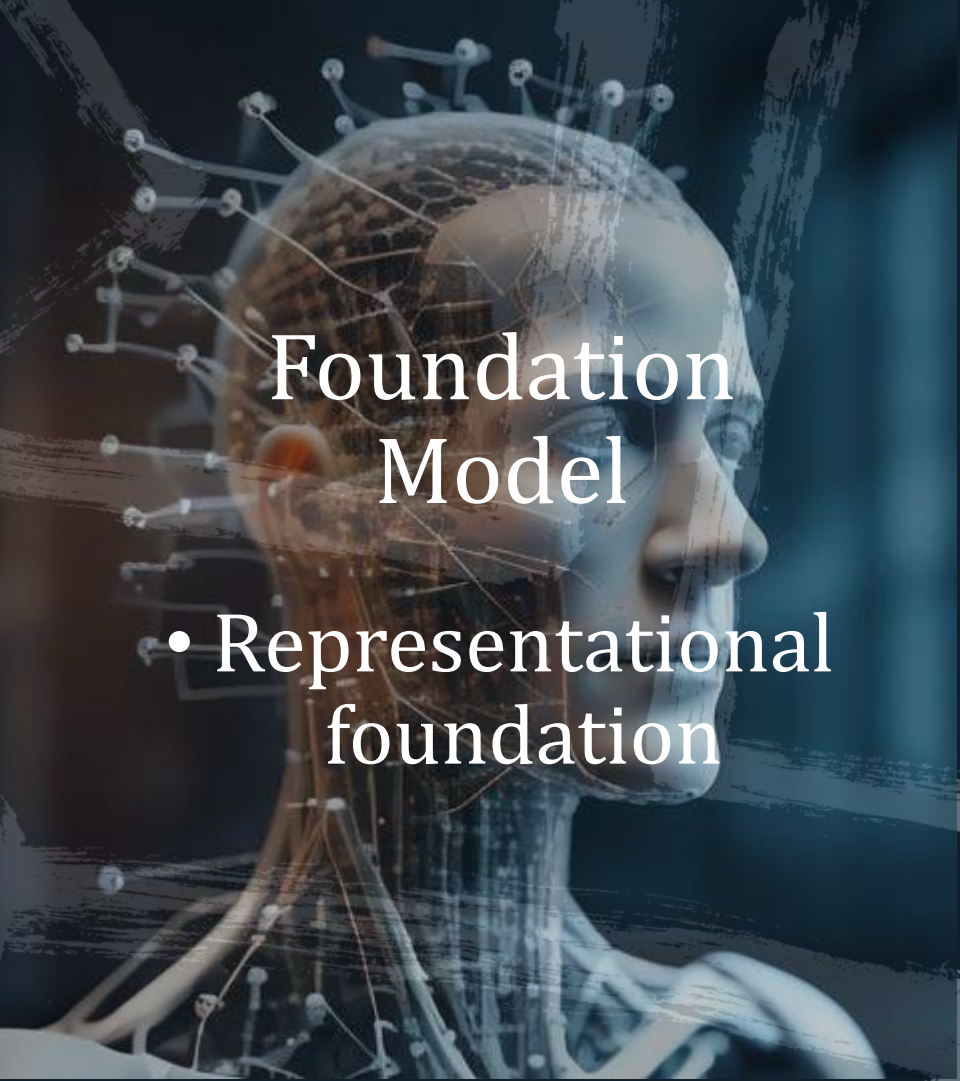


Knowledge Fusion



Universal Model

- Task
universality



Foundation Model

- Representational
foundation

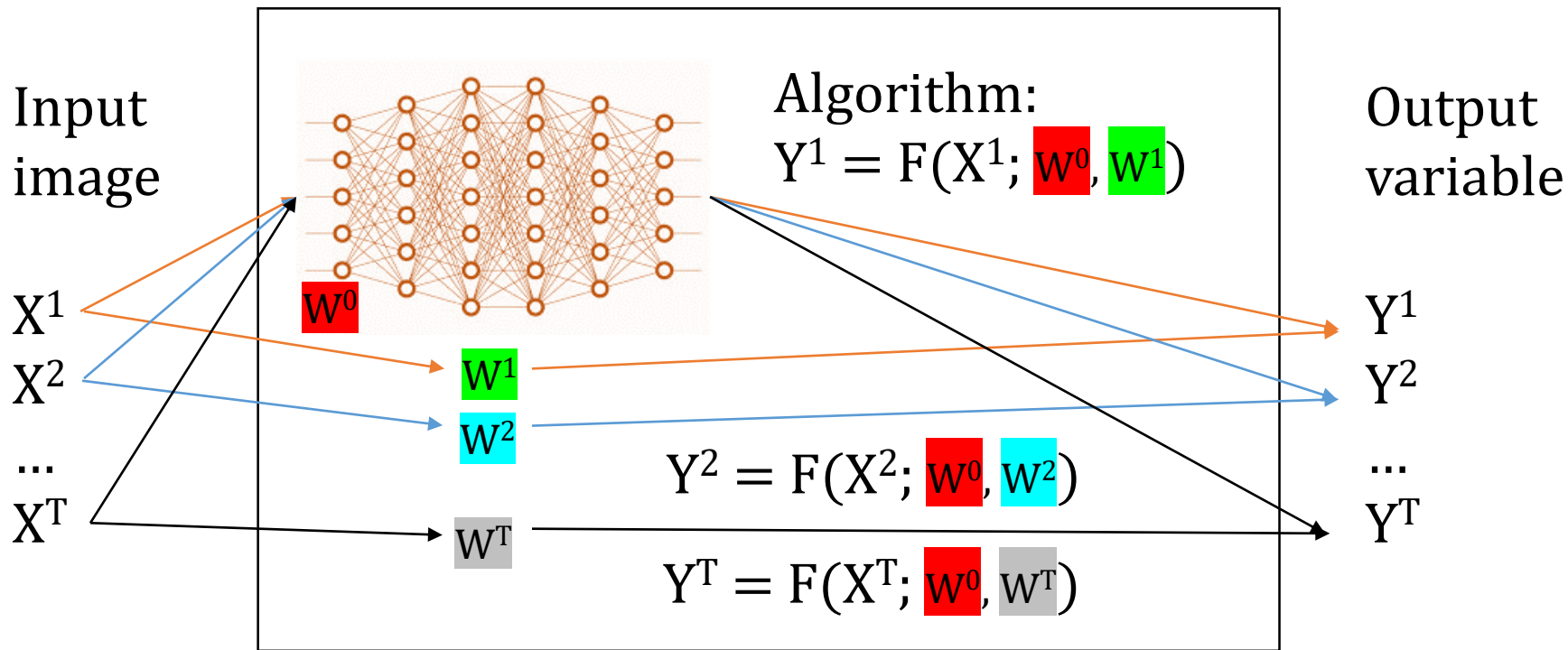
Universal model (UM)

Learn once, solve multiple tasks

- Integrate multiple tasks
- Fit multi-domain multicenter data
- Model “commonality + individuality”

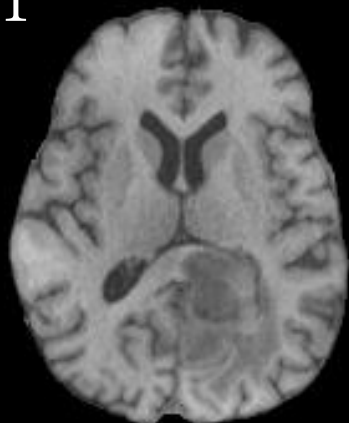


Deep universal learning

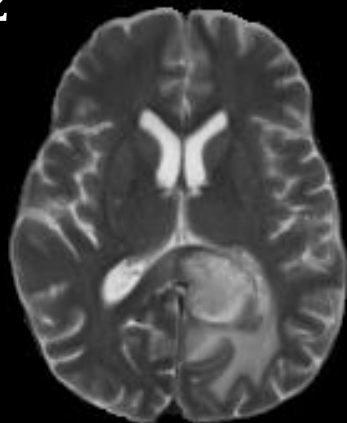


Universal MRI synthesis

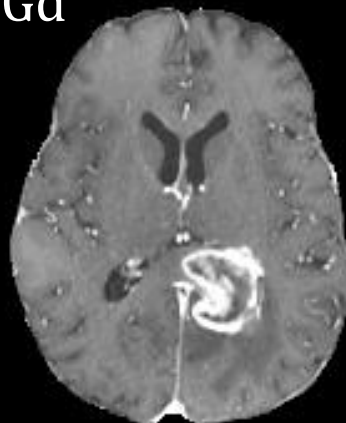
T1



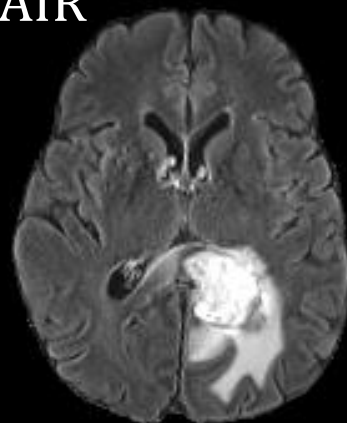
T2



T1Gd

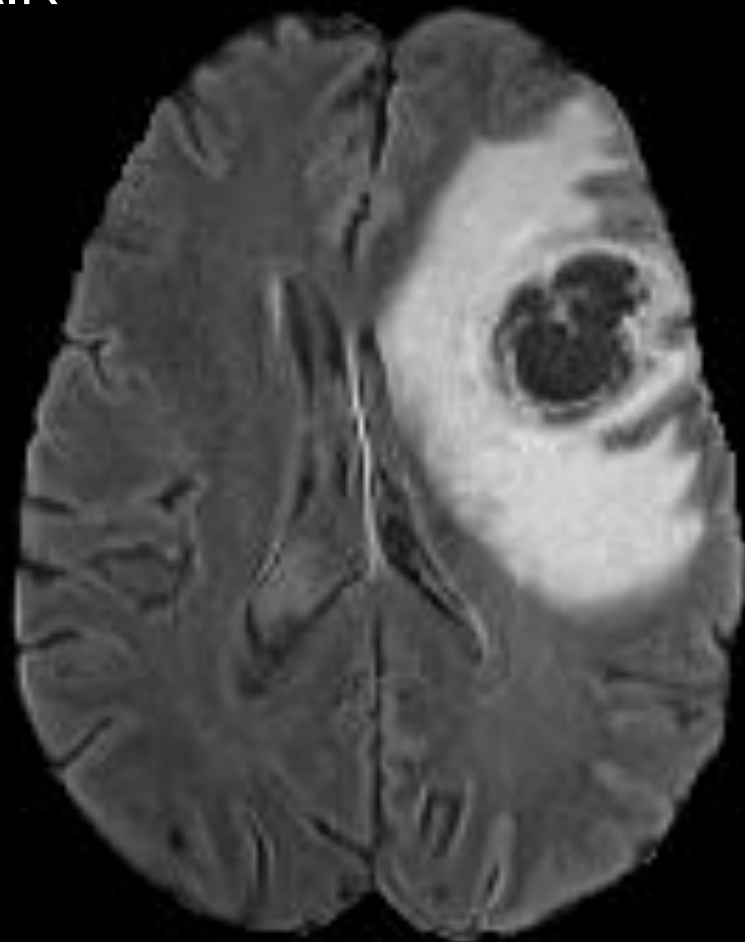


FLAIR



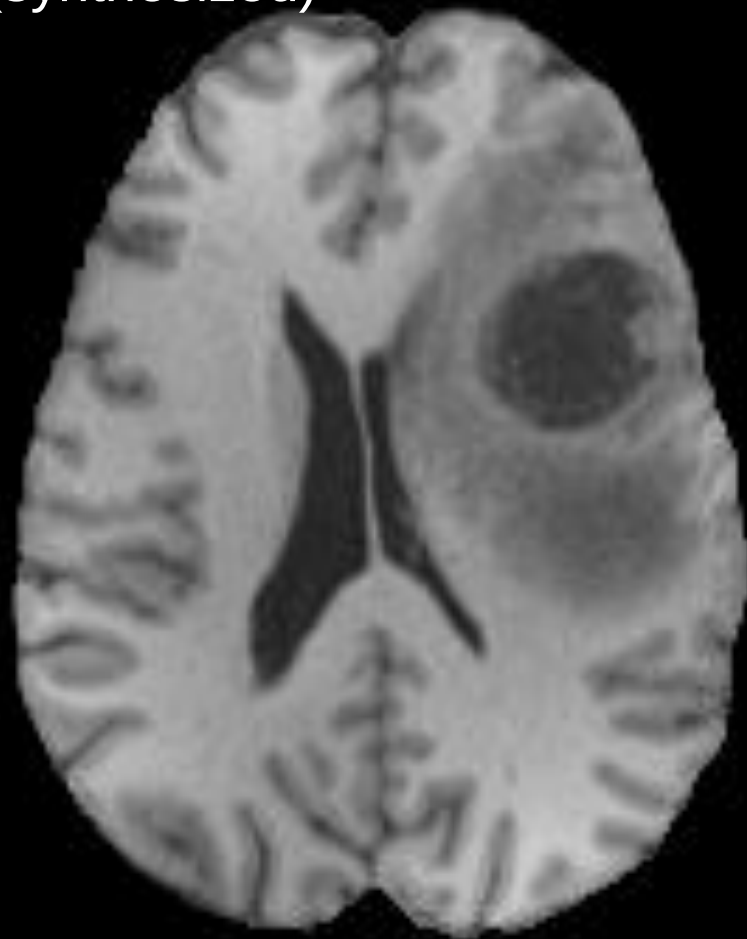
FLAIR -> T1

FLAIR



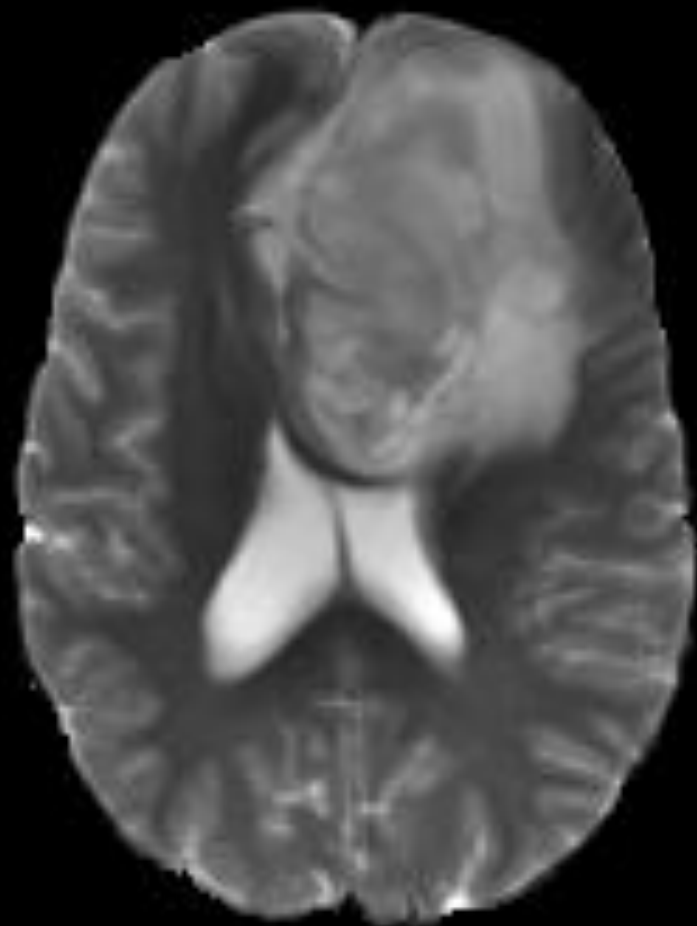
FLAIR -> T1

T1 (synthesized)



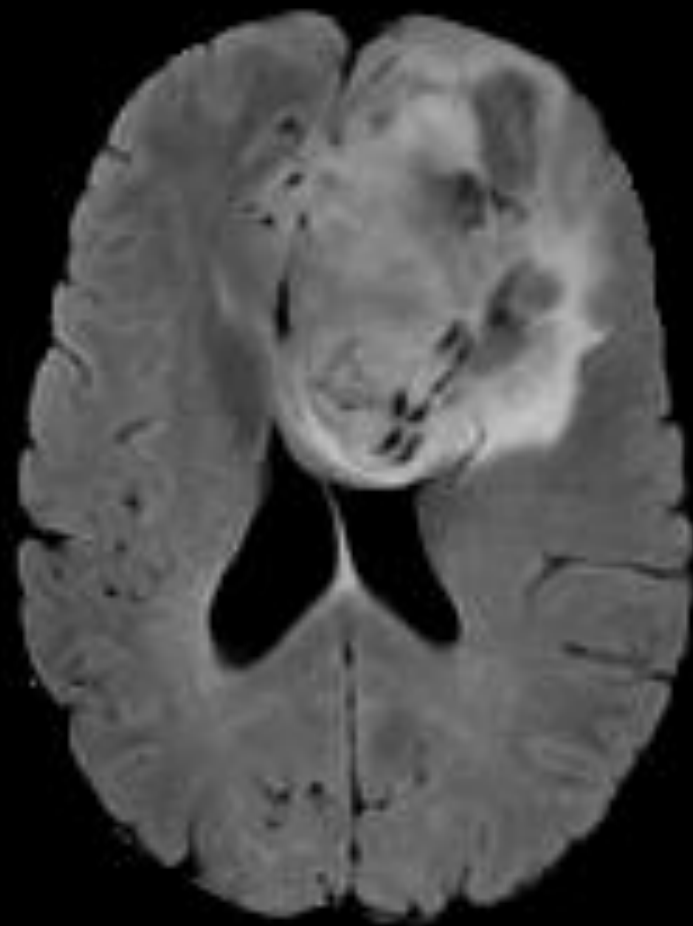
T2+FLAIR -> T1

T2



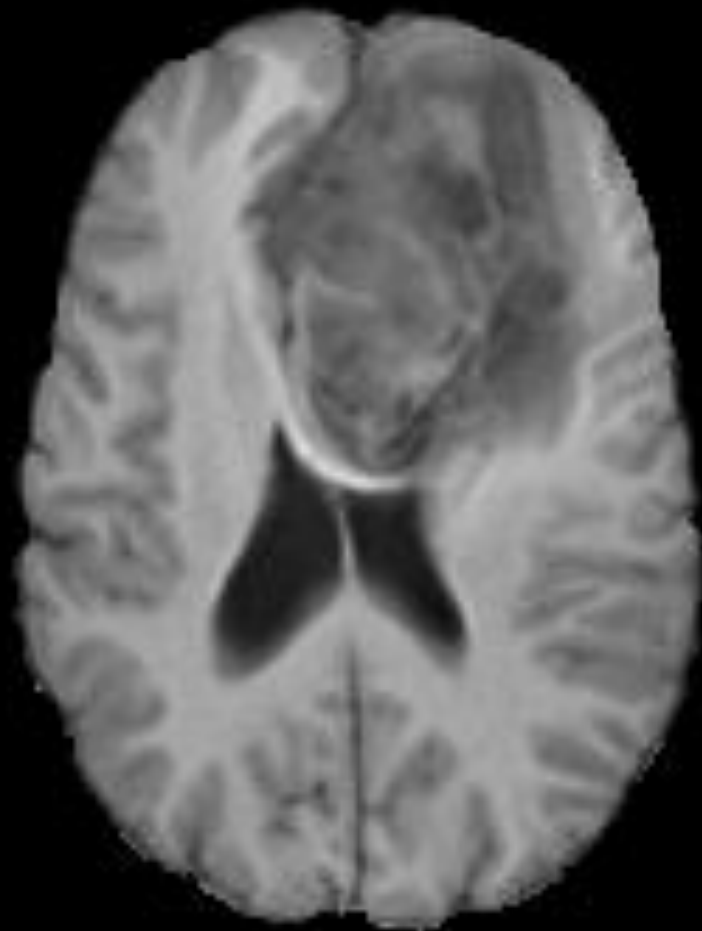
T2+FLAIR -> T1

+ FLAIR



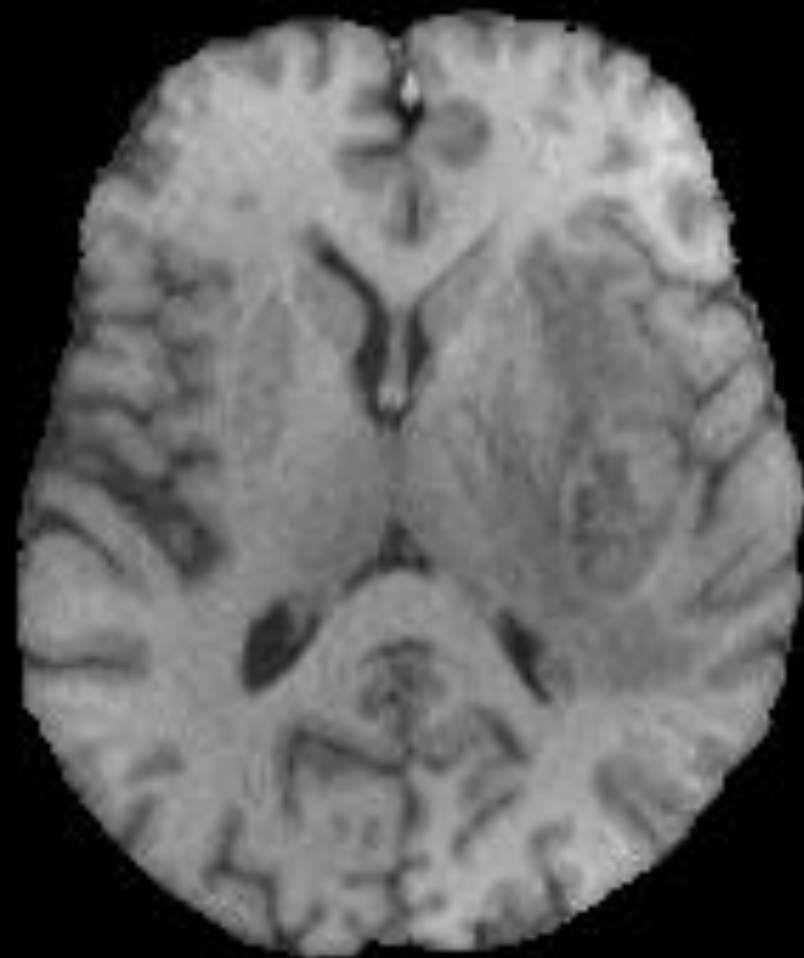
T2+FLAIR -> T1

T1 (synthesized)



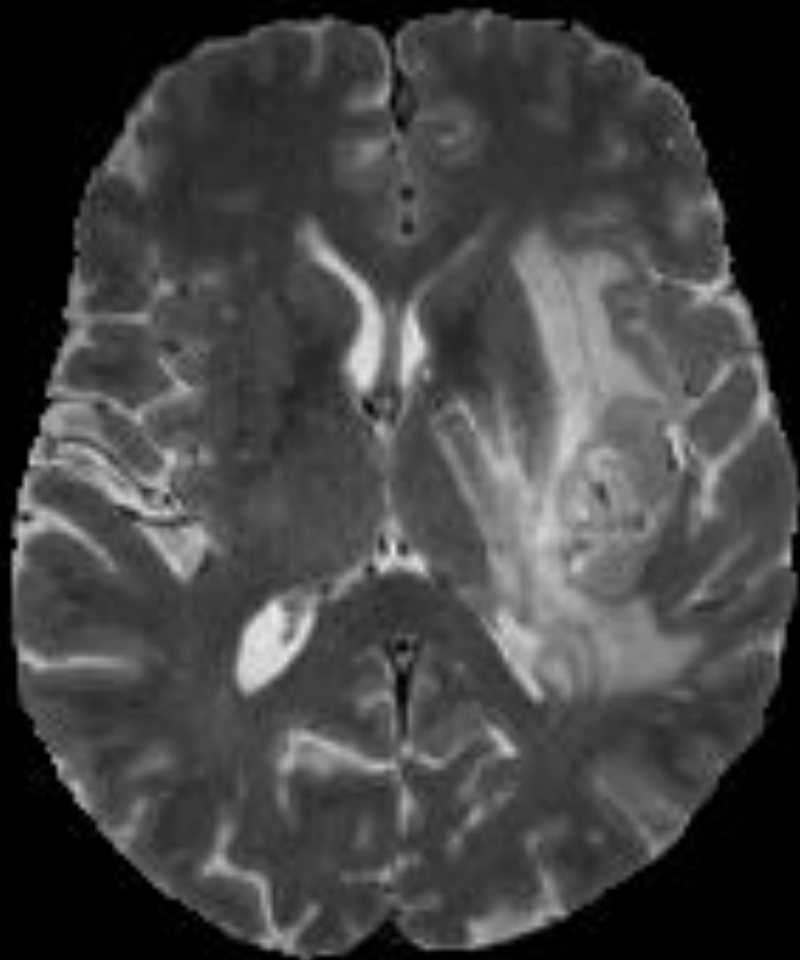
T1+T2+T1Gd -> FLAIR

T1



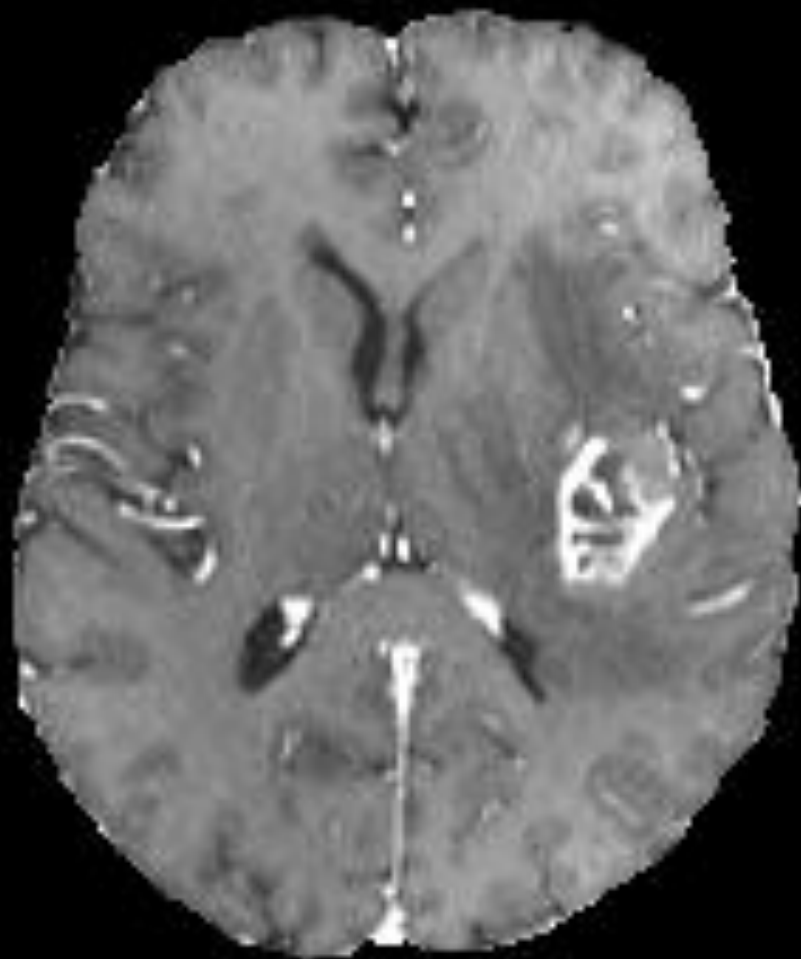
T1+T2+T1Gd -> FLAIR

+ T2



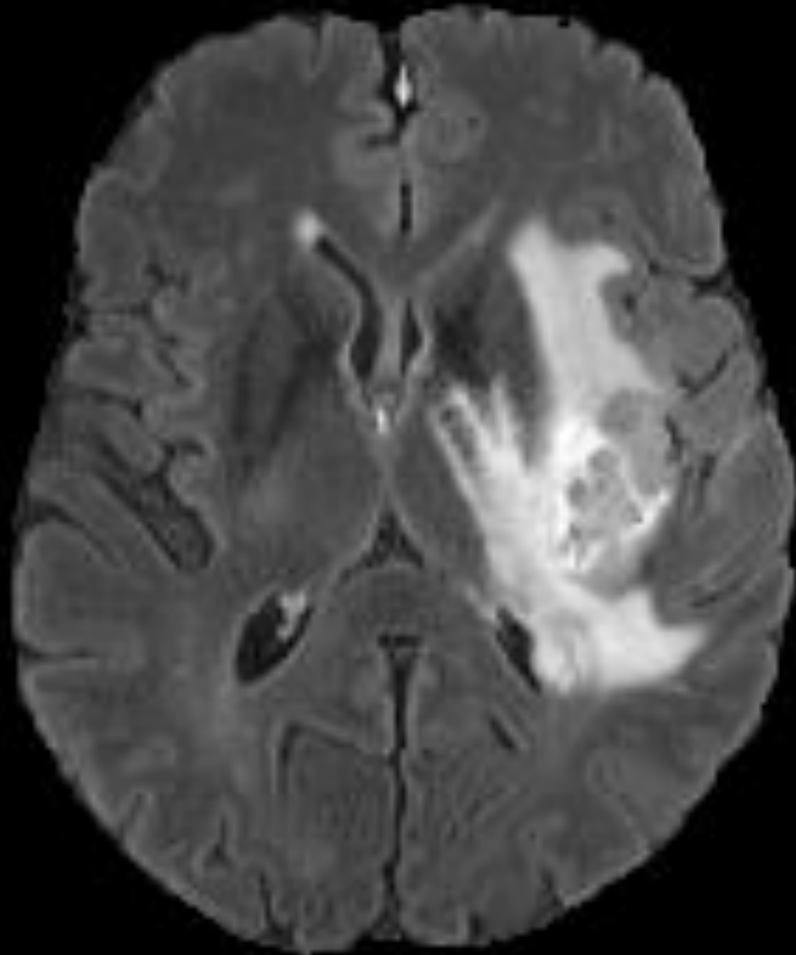
T1+T2+T1Gd -> FLAIR

+ T1Gd

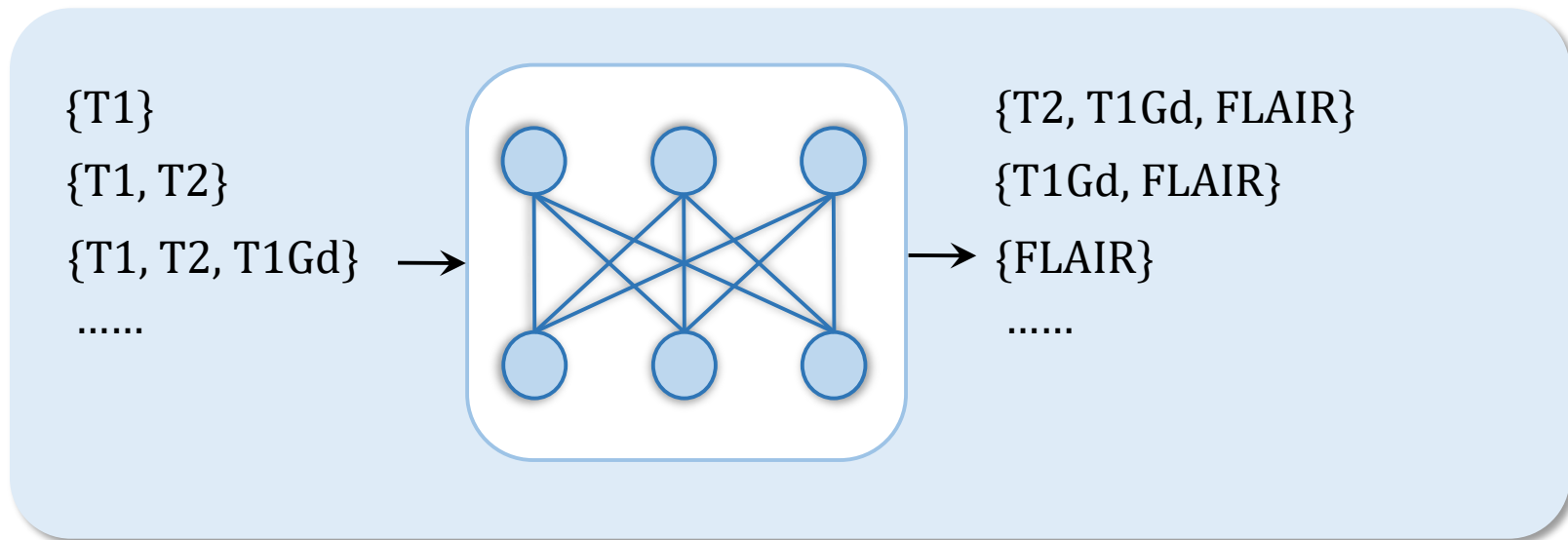


T1+T2+T1Gd -> FLAIR

FLAIR (synthesized)



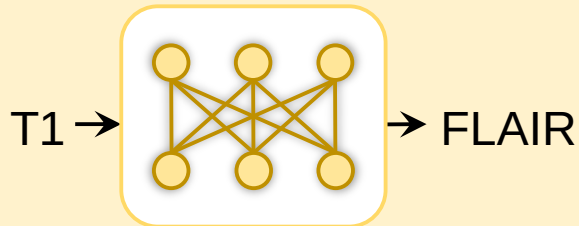
Universal MRI synthesis



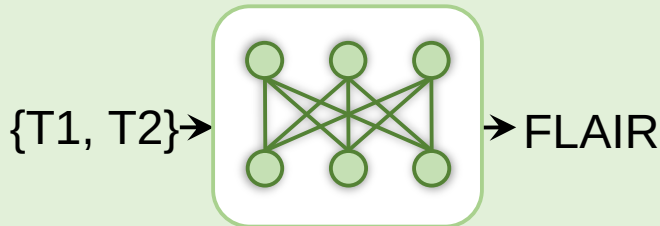
One model deals with all synthesis settings
($12+12+4=28$ settings for 4 modalities)

Synthesis categories

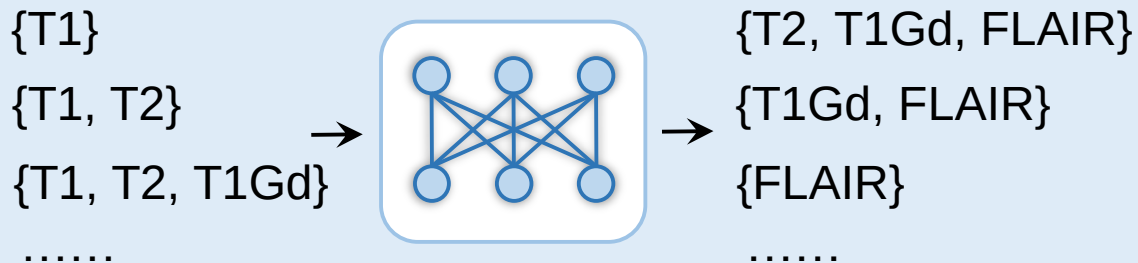
One-to-one synthesis



Many-to-one synthesis



Unified synthesis



Two challenges



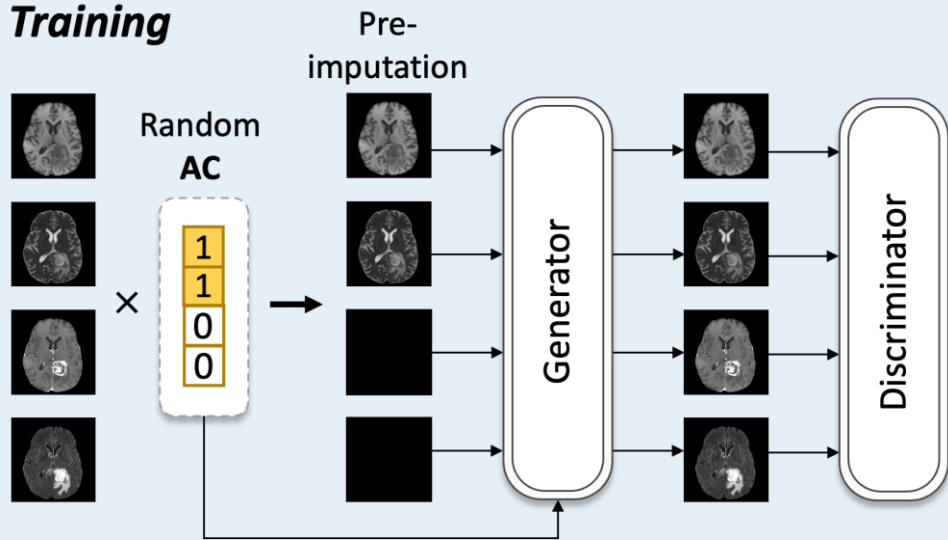
How to fully exploit the commonality and discrepancy information contained in multiple available modalities?



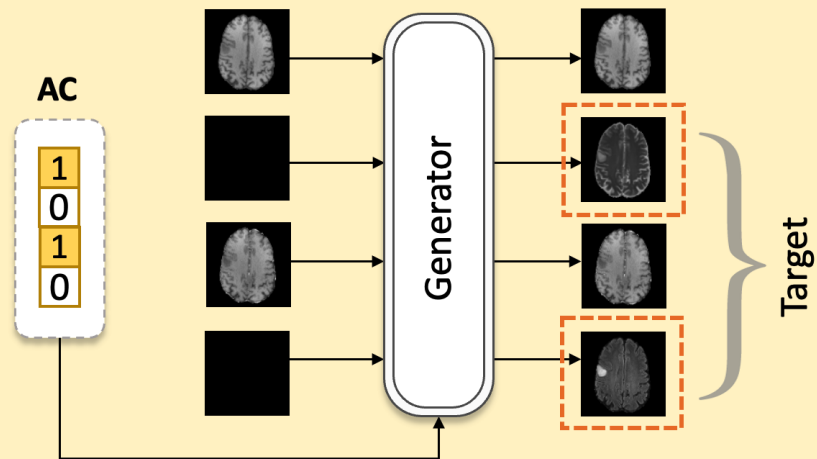
How to combine information from a varying number of available modalities?

Pipeline

Training

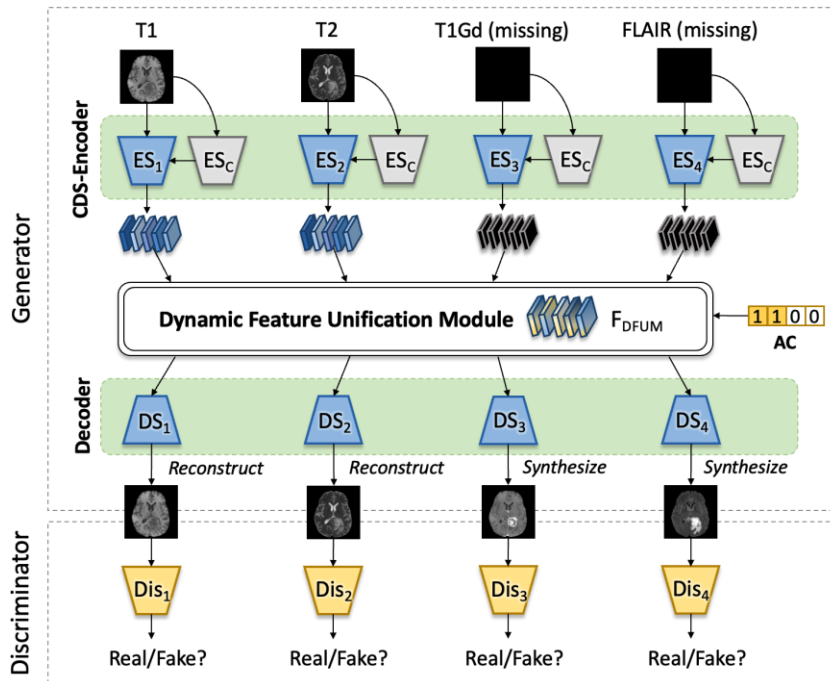


Inference



* **AC:** Availability Condition

Universal MRI synthesis framework



ES_C: Common encoding stream
for modality-invariant information extraction

ES_i: Modality-Specific encoding stream for
modality-specific information extraction

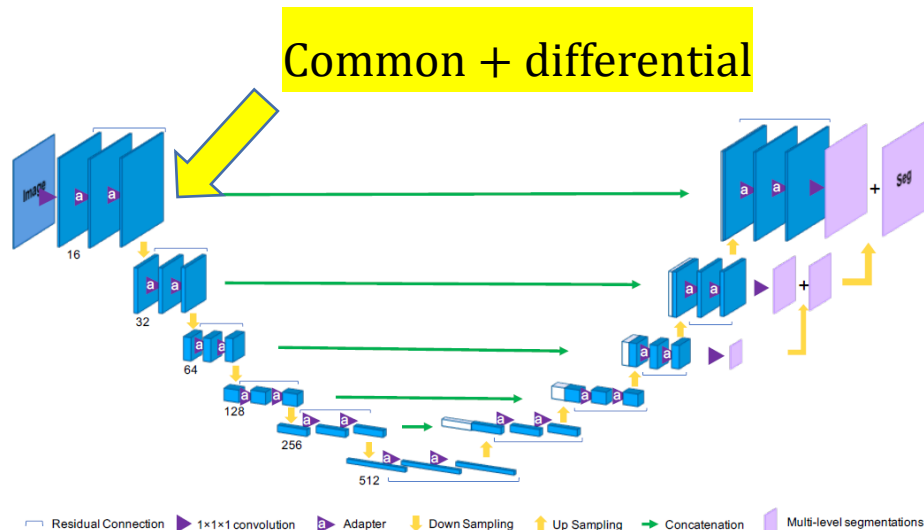
DS_i: Modality-specific decoding stream for
synthesizing different modalities

Results

PSNR (dB)	One to One		Two to One		Unified Synthesis	
	T1 -> T2	FLAIR -> T1	T1+FLAIR -> T2	T2+FLAIR -> T1	T1+T1Gd+T2 -> FLAIR	T2 +T1Gd+FLAIR -> T1
pGAN	27.10	26.91	-	-	-	-
Pix2Pix	26.94	26.86	-	-	-	-
Hi-Net	-	-	27.14	27.25	-	-
MM-Synthesis	-	-	-	-	25.87	30.60
MM-GAN	-	-	-	-	26.64	31.06
UMMIS (ours)	27.78	27.38	28.68	29.47	27.39	31.94

- One universal network beats any SOTA single-task network

Universal multi-domain segmentation

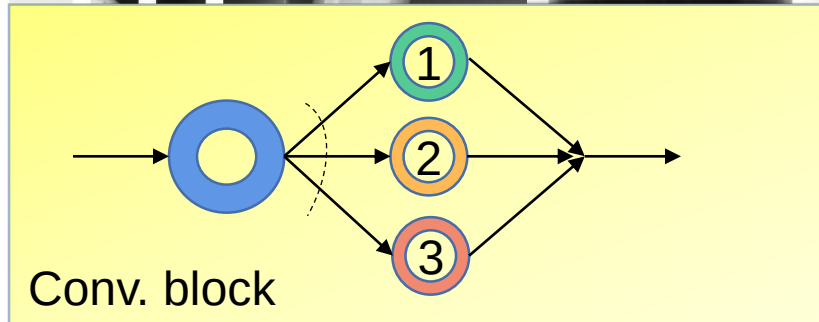
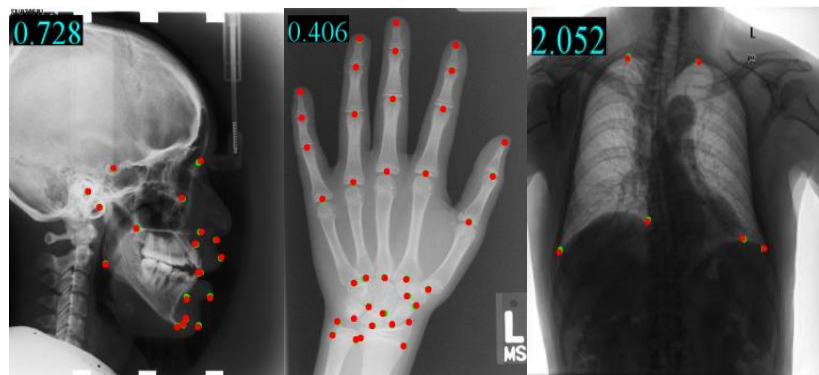


- 6 different organs (MRI/CT)
- similar performances to individual networks with only **1% parameters**
- **Great transferrability**

Dice%	MR LAtri um	CT Liver	MR Hippo	MR Prosta te	CT Pancr e.	CT Spleen
Indep. (127M)	93.3	95.0	89.6	58.4	78.8	92.4
Shared (31M)	92.7	93.4	89.3	68.4	57.6	90.7
Ours (1.7M)	92.0	93.5	89.3	68.5	62.1	91.6

Universal landmark detection

1 Model



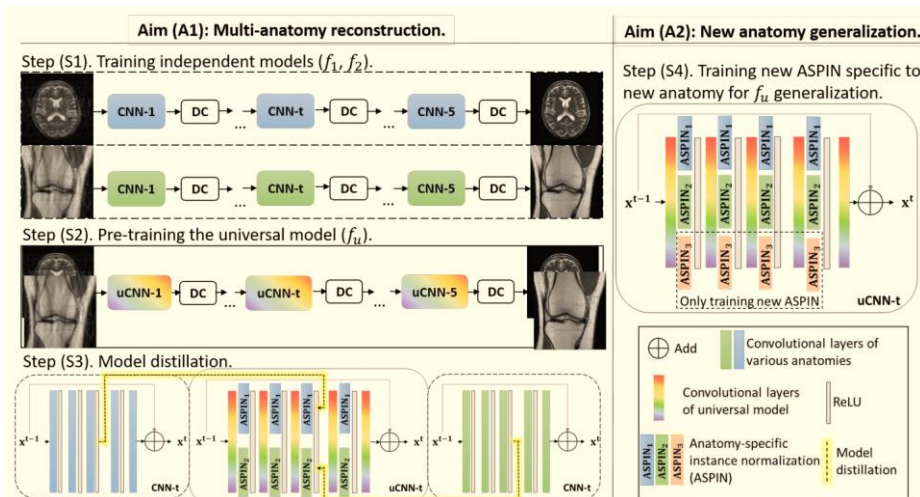
1 Model						
	Head		Hand		Chest	
	MRE (mm)	SDR% (4mm)	MRE (mm)	SDR% (4mm)	MRE (px)	SDR% (6px)
SOTA	1.67	89.9	0.66	99.3		
UNet	12.5	73.7	6.14	92.5	5.61	82.3
GU2 Net	1.54	94.9	0.84	99.4	5.57	82.7
DATR	1.47	95.6	0.86	100	4.30	77.4

- Beats any SOTA single-task network
- With less parameters: ~10%
- Transformer model is even better

Zhu et al. You only learn once: Universal anatomical landmark detection. MICCAI 2021.

Zhu et al. DATR: Domain-adaptive transformer for multi-domain landmark detection. ArXiv 2203.06433.

Universal undersampled MRI reconstruction



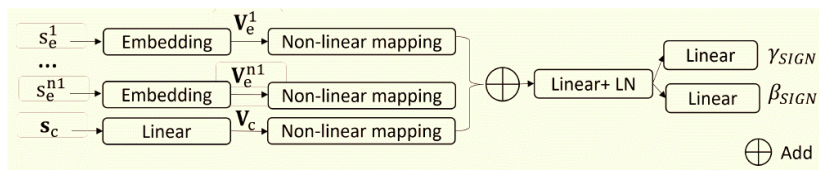
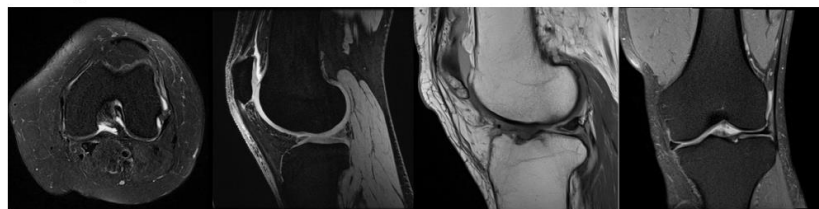
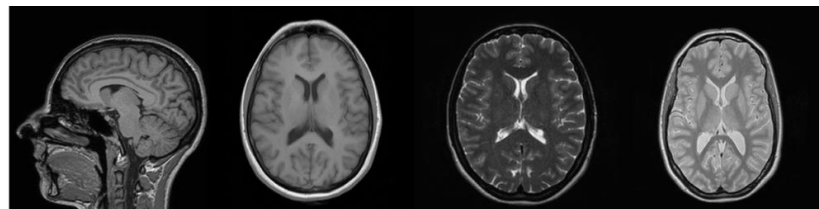
$$\text{ASPIN}_a(h) = \gamma_a \left(\frac{h - \mu(h)}{\sigma(h)} \right) + \beta_a, \quad a = 1, 2, \dots, A.$$

Instance Normalization (IN)

	Universal		Adaptation		
PSNR	Brain (311K)	Knee (274K)	Cardio (25K)	Abdomen (1.3K)	Prostate (904)
Zero Fill	31.60	30.45	29.20	33.52	26.28
Indep.	40.85	35.46	37.05	40.81	30.91
Shared	40.78	35.43	35.18	41.15	30.30
Ours	43.16 +2.31	36.39 +0.93	37.30 +0.25	42.76 +1.61	31.61 +0.70

- Beats any SOTA single-task network
- Less parameters
- Good adaptability with extra 0.1% parameters

Undersampled MRI reconstruction



PSNR(dB)	Brain (4x)	Knee (4x)	Brain (6x)	Knee (6x)
Zero Fill	29.92	30.04	22.03	24.63
D5C5	39.29	35.34	36.03	33.23
D5CS+SIGN	41.70 +2.41	37.18 +1.74	38.16 +2.13	34.66 +1.43
OUCR	40.29	36.48	36.93	34.16
OUCR+SIGN	41.64 +1.35	37.06 +0.58	37.96 +1.03	34.47 +0.31

Medical imaging universal model publications

1. C Huang, et al. 3D U2Net: A 3D **universal** U-net for multi-domain medical image segmentation. MICCAI, 2019.
2. H Li, H Han, and S Kevin Zhou. Bounding maps for **universal** lesion detection. MICCAI, 2020.
3. H Li, et al. Conditional training with bounding map for **universal** lesion detection. MICCAI, 2021.
4. H Zhu, Q Yao, L Xiao, and S Kevin Zhou. You only learn once: **Universal** anatomical landmark detection. MICCAI, 2021.
5. H Zhu, et al. Learning to localize cross-anatomy landmarks in X-ray images with a **universal** model. BME Frontiers, 2022.
6. Zhang et al. **Unified** multi-modal image synthesis for missing modality imputation. IEEE TMI, 2024.
7. X Liu, J Wang, F Liu, and S Kevin Zhou. **Universal** undersampled MRI reconstruction. MICCAI 2021. (oral)
8. H Li, L Chen, H Han, and S Kevin Zhou. SATr: Slice attention with transformer for **universal** lesion detection. MICCAI, Singapore, 2022.
9. X Liu, et al. Undersampled MRI reconstruction with side information guided normalization. MICCAI 2022.
10. H Zhu, Q Quan, Q Yao, Z Liu, and S Kevin Zhou. UOD: **Universal** one-shot detection of anatomical landmarks. MICCAI 2023.
11. P Zhao, H Li, R Jin, and S Kevin Zhou. DiffULD: Diffusive **universal** lesion detection. MICCAI, 2023.

Foundation model (FM)

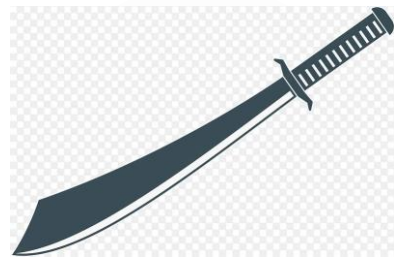
Learn, then transfer to multiple tasks.

- ① Image Feature FM
- ② Vision Language FM
- ③ Image Generation FM
- ④ Image Segmentation FM





universal



foundational



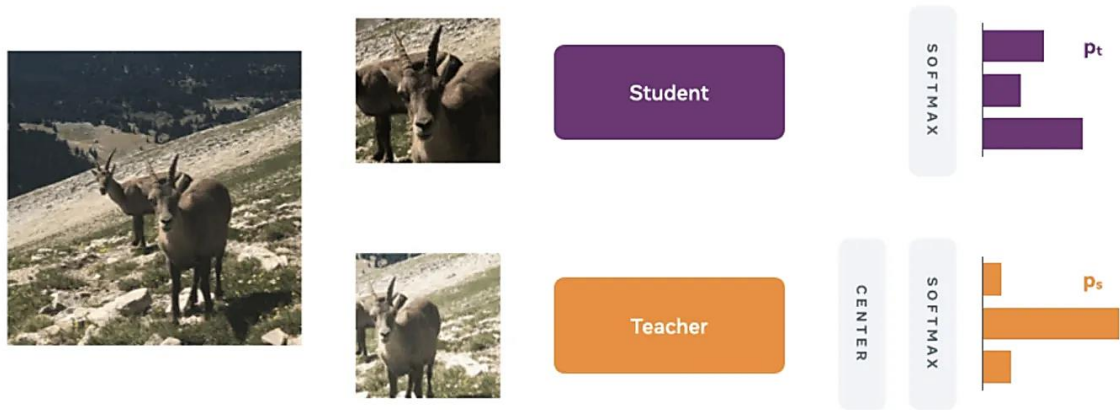
Medial Imaging Foundation Model

- ① Medical Image Feature FM
- ② Medical Vision Language FM
- ③ Medical Image Generation FM
- ④ Medical Image Segmentation FM



① Image Feature FM

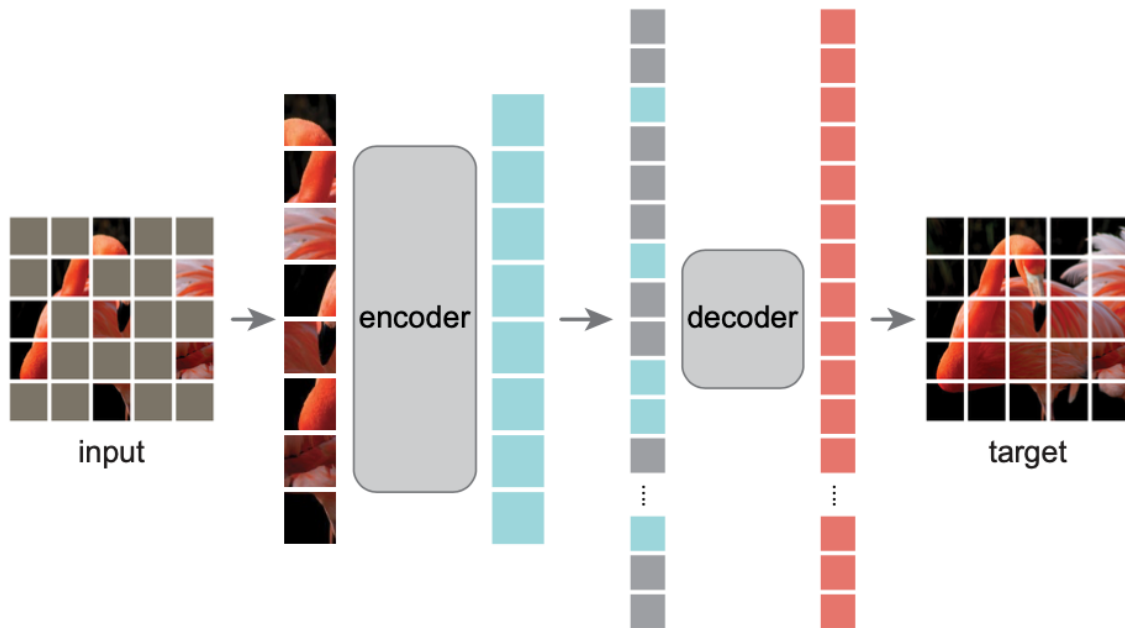
DINO (self-Distillation with NO labels)



- Feature consistency under deformation
- **Implicit semantic understanding at a global level**
- Effective for **image-level** classification tasks after fine-tuning

① Image Feature FM

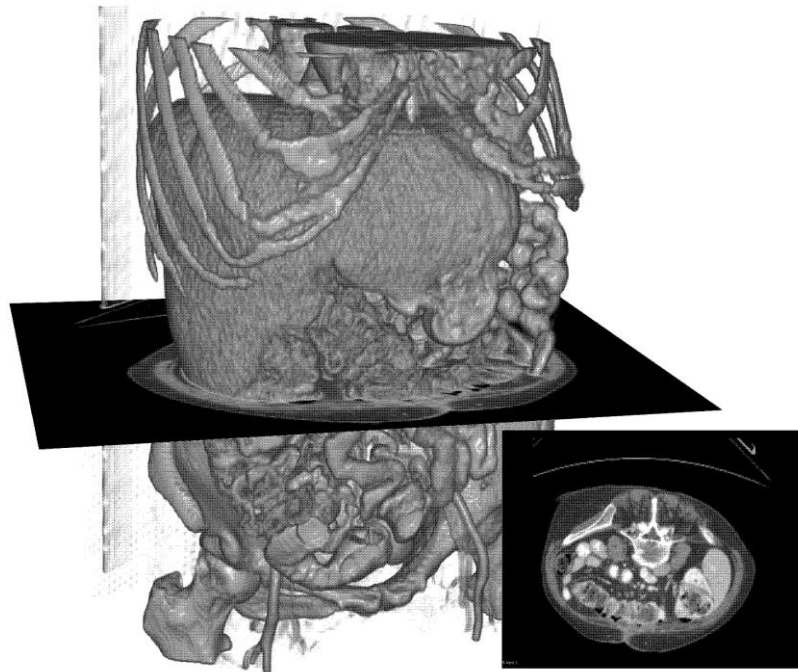
MAE (Masked Autoencoder)



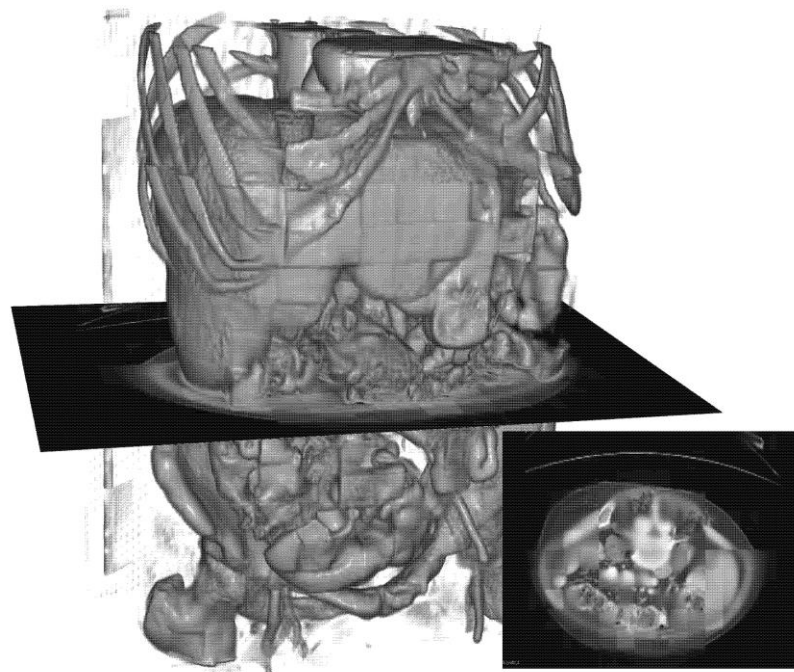
- Inferring the whole from local parts
- **Implicit semantic understanding at a local level**
- Effective for **pixel-level** tasks (e.g. segmentation, object detection) after fine-tuning

HySpark recon. results

Image: 221,184d vec --> 663,552d vec
GPT-4: 32,768 tokens --> a token (768d vec)

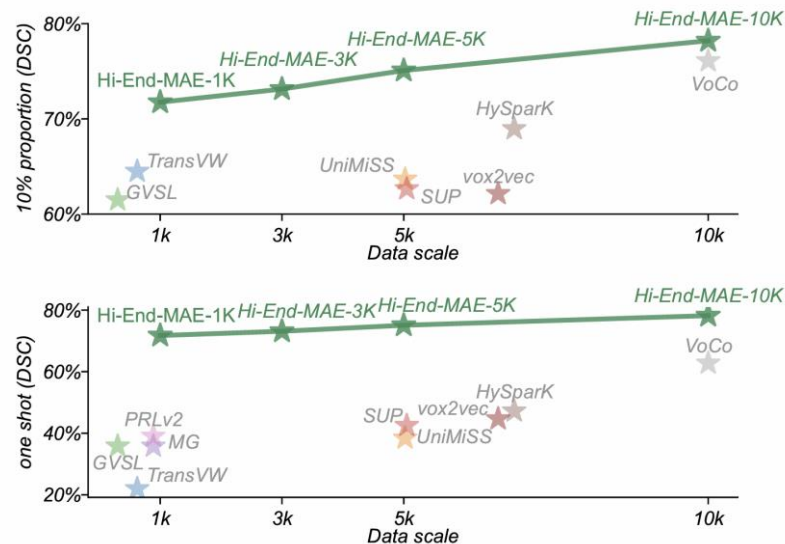
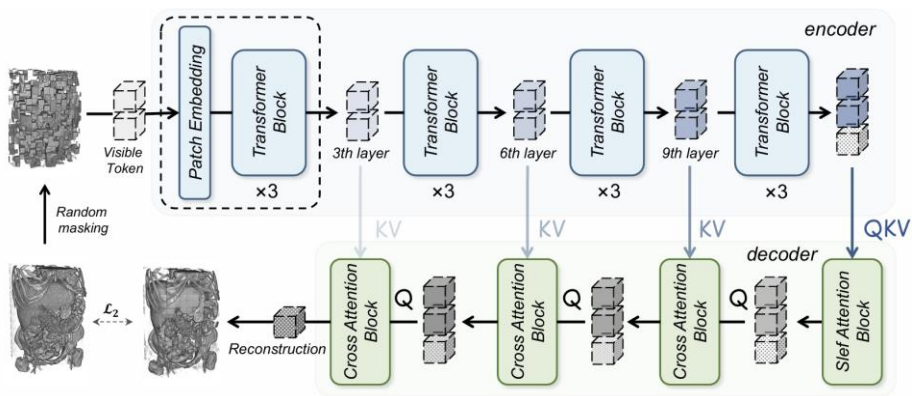


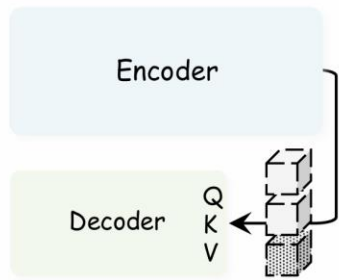
Original



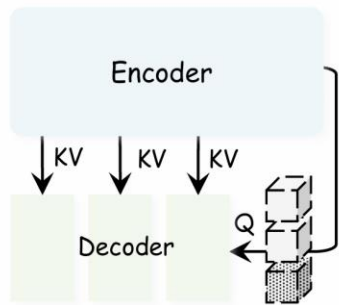
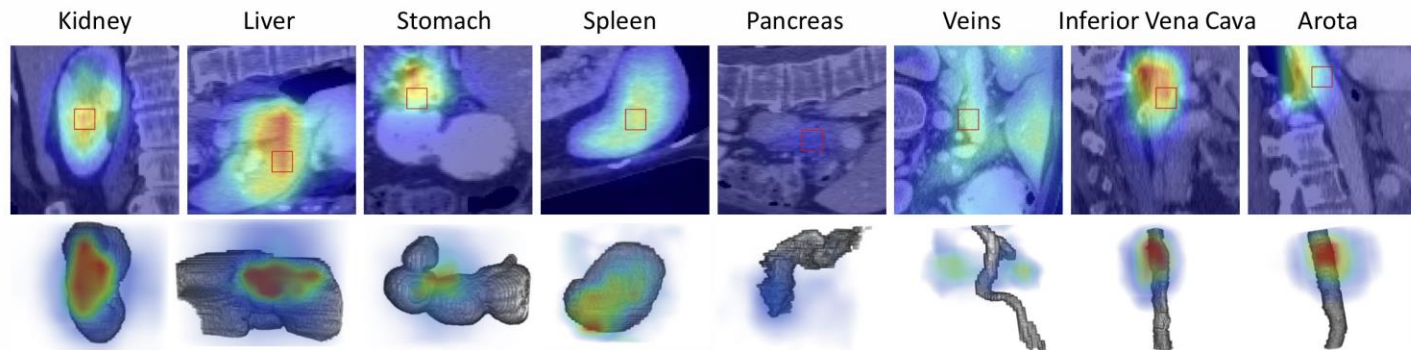
Reconstructed

Hi-End-MAE: Hierarchical encoder-driven MAE

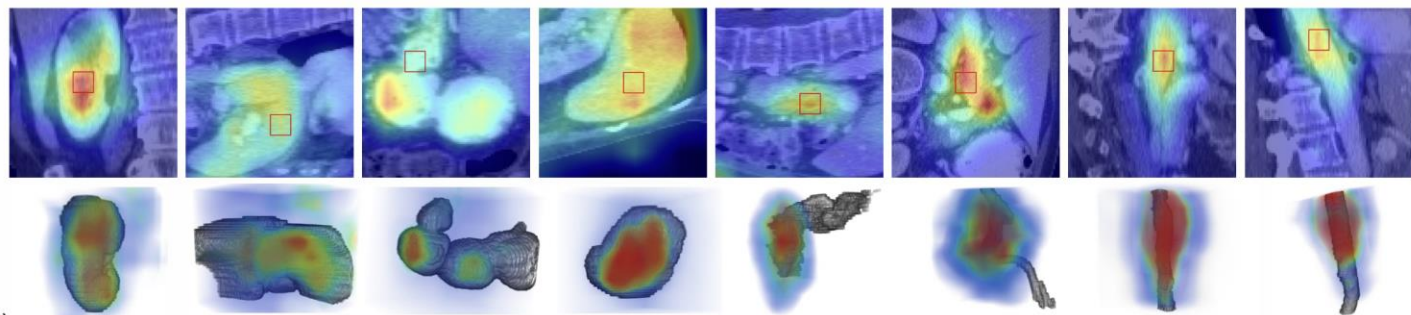




(a) **Decoder-driven** reconstruction

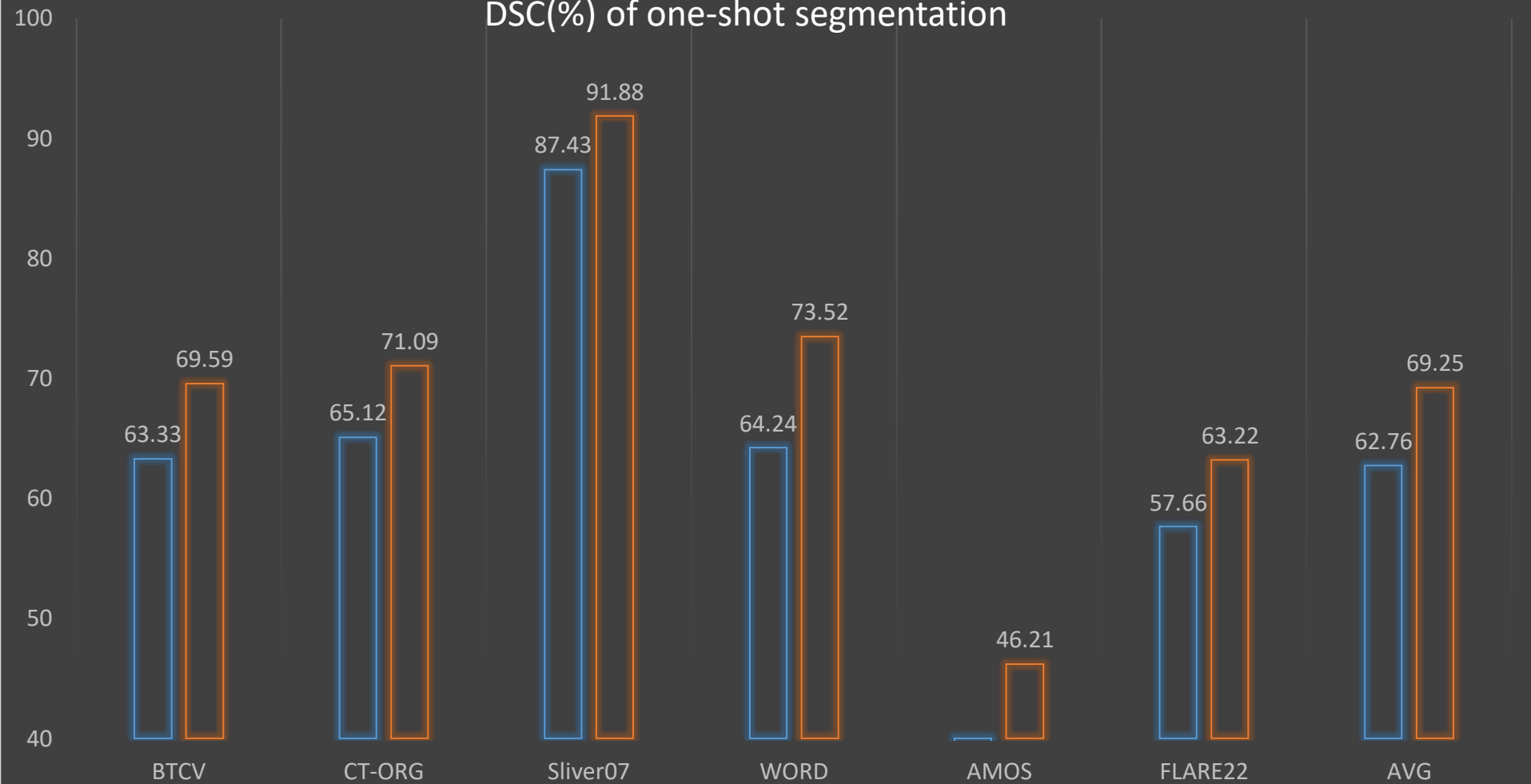


(b) **Encoder-driven** reconstruction (ours)



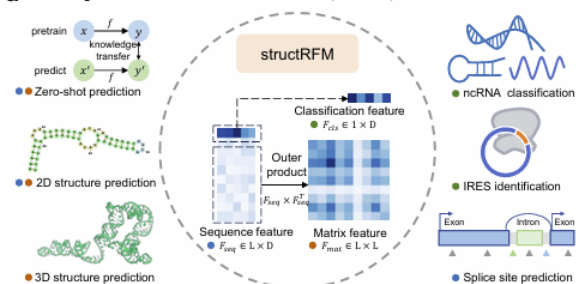
VoCo Hi-End-MAE

DSC(%) of one-shot segmentation

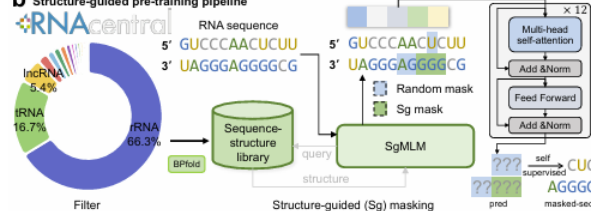


Structure-guided RNA FM [Nature major rev.]

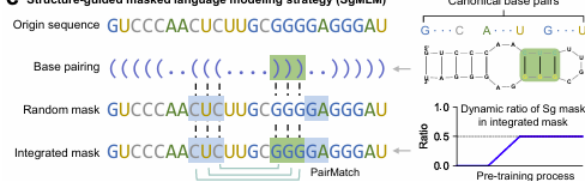
a Structure-guided RNA foundation model for zero-shot, structure, and function inference



b Structure-guided pre-training pipeline



c Structure-guided masked language modeling strategy (SgMLM)



Task	Dataset	N	Metric	Method	Mean
Zero-shot homology classification	Rfam	24,523	OR ↓	RNAErnie	0.156
				MP-RNA	0.234
				UTR-LM	0.202
				RNA-FM	0.071
				structRFM	0.097
	ArchiveII600	3,611	OR ↓	RNAErnie	0.516
				MP-RNA	0.546
				UTR-LM	0.630
				RNA-FM	0.381
				structRFM	0.483
Zero-shot structure prediction	PDB	116	F1 ↑	structRFM	0.436
Secondary structure prediction	ArchiveII600	3,611	F1 ↑	RNAErnie	0.873
				RNA-FM	0.847
				structRFM	0.873
				RNAErnie	0.608
				RNA-FM	0.639
				structRFM	0.641
Tertiary structure prediction	CASP15-natural	8	RMSD (Å) ↓	trRosettaRNA	18.4
				AlphaFold3	15.6
				structRFM	14.8
	RNA Puzzles	20	RMSD (Å) ↓	trRosettaRNA	7.28
				AlphaFold3	8.11
				structRFM	6.82
	CASP15	12	F1 ↑	trRosettaRNA	0.625
				structRFM	0.739
				RNA Puzzles	0.646
	RNA Puzzles	20	F1 ↑	trRosettaRNA	0.646
				structRFM	0.665
Splice site prediction	SpliceAI	16,505	Accuracy ↑	RNA-FM	0.350
				structRFM	0.416
IRES identification	IRES	1,164	F1 ↑	DeepIRES	0.284
				IRESfinder	0.475
				structRFM	0.708
ncRNA classification	nRC	2,600	F1 ↑	RNAErnie	0.964
				RNA-FM	0.972
				structRFM	0.973

7 tasks, 12 experiments, 10/12 rank #1

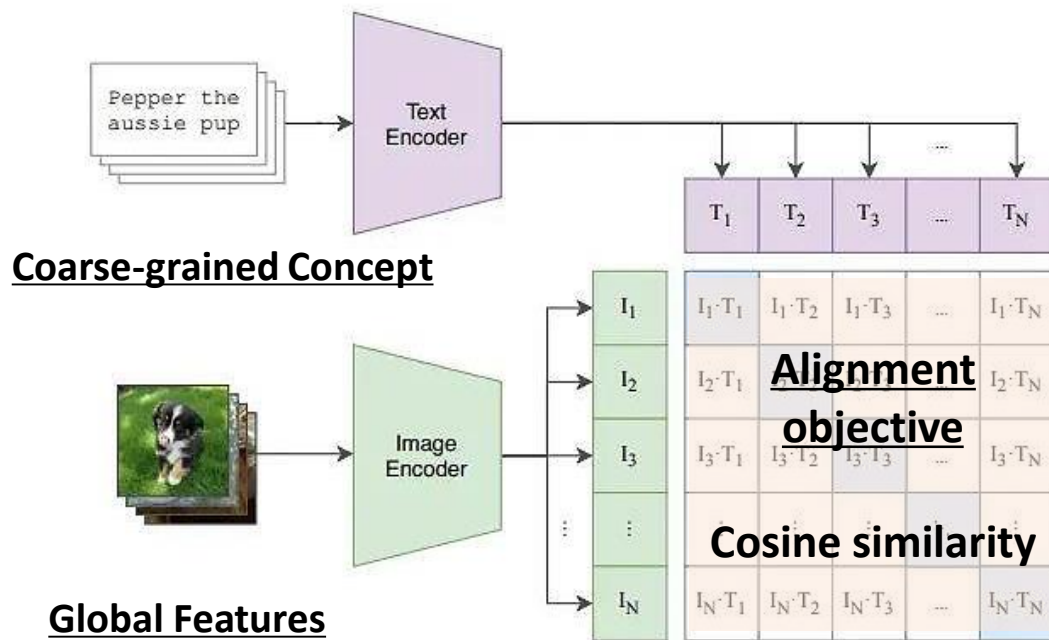
Medial Imaging Foundation Model

- ① Medical Image Feature FM
- ② Medical Vision Language FM
- ③ Medical Image Generation FM
- ④ Medical Image Segmentation FM



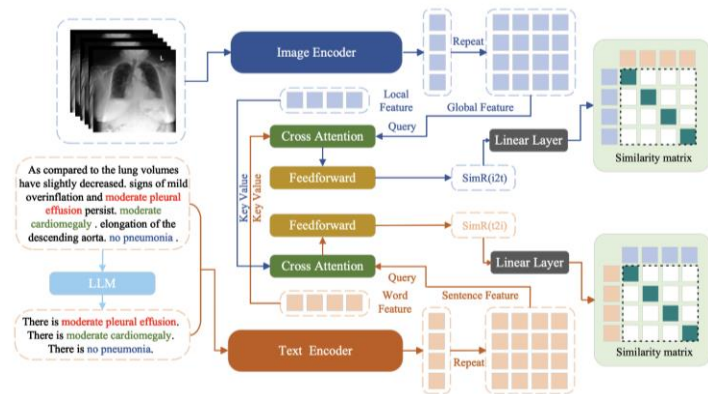
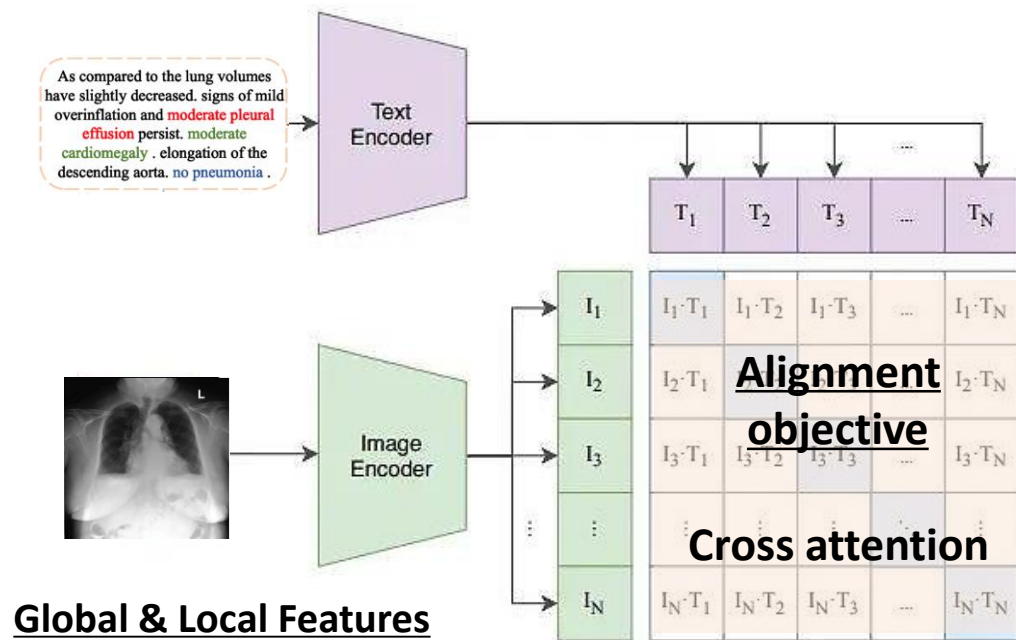
② Vision-Language FM

CLIP (Contrastive Language-Image Pre-Training)



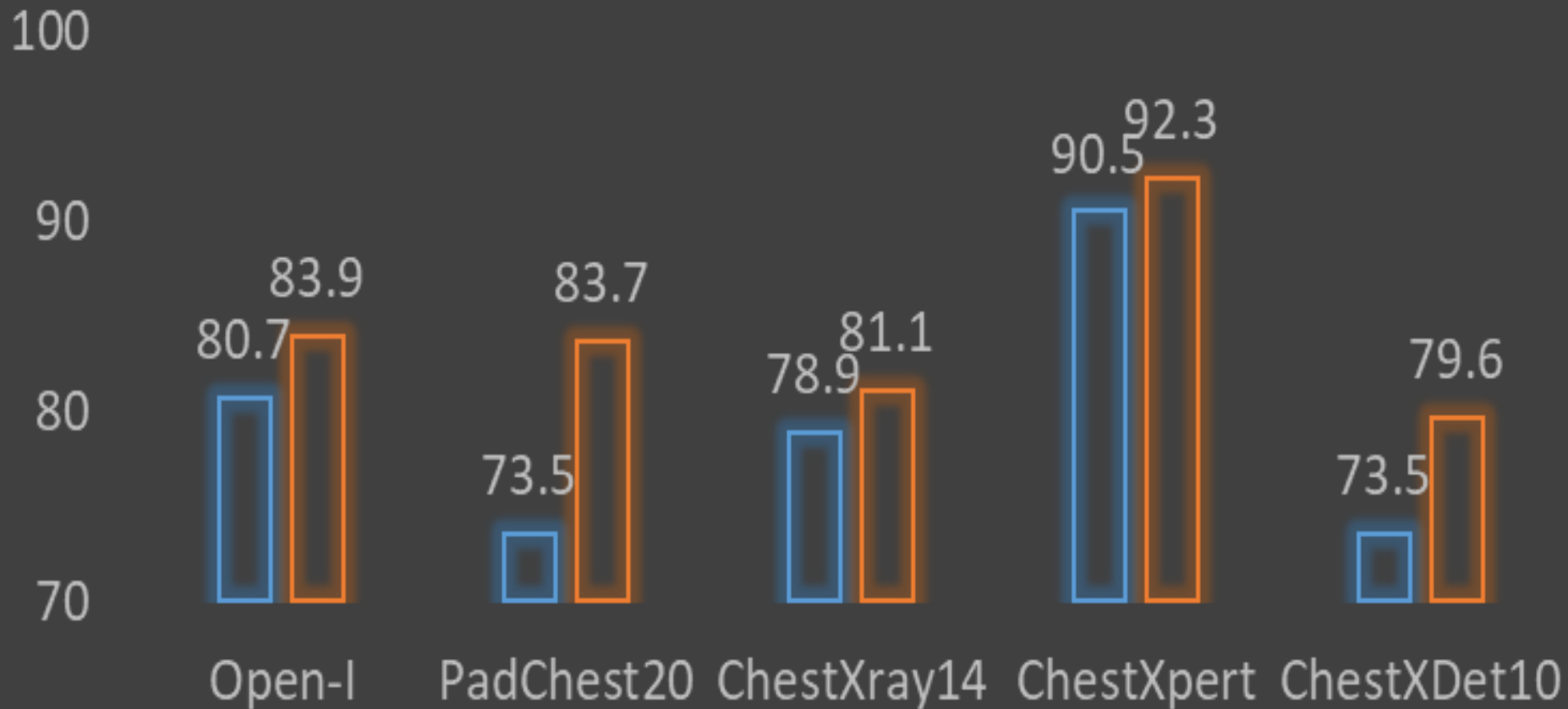
- Use image-text pairs to pre-train encoders for **explicit cross-modal semantic alignment**
- Effective for zero-/few-shot image classification, open-set tasks, etc.
- Problem when applied to medical image:
Granularity

② Medical Vision Language FM CARZero

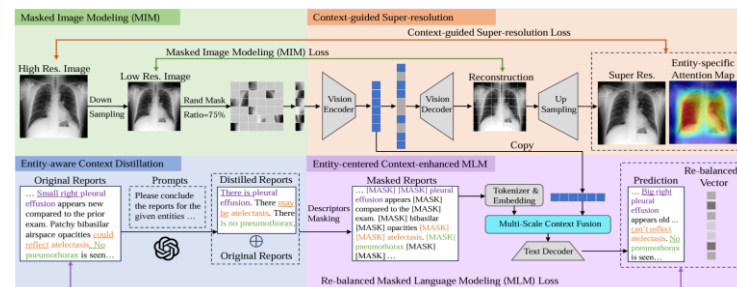
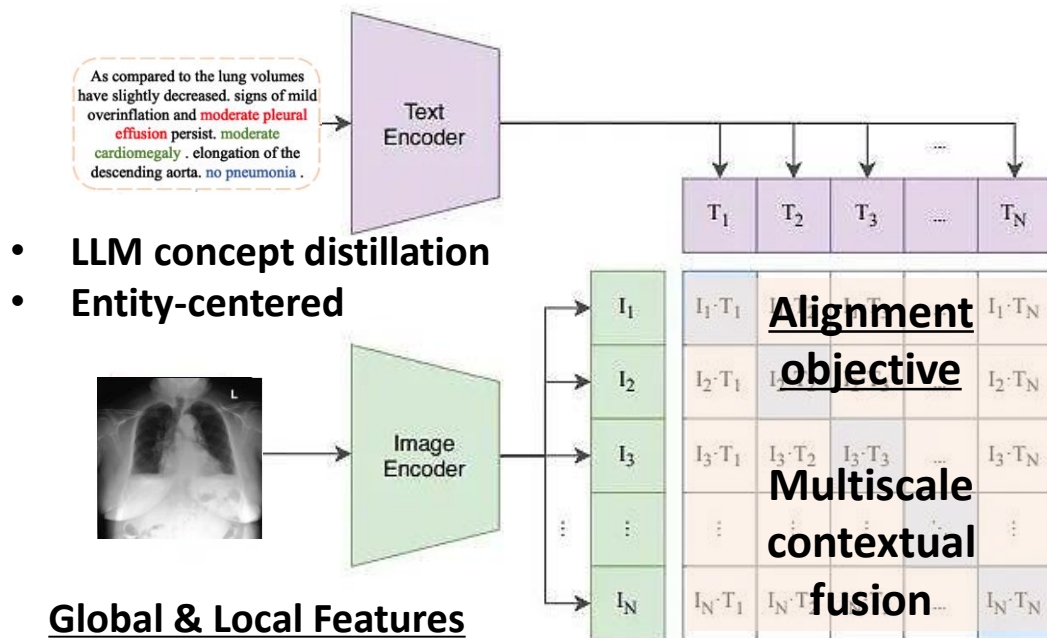


Zero-shot disease classification

□ SOTA □ Ours

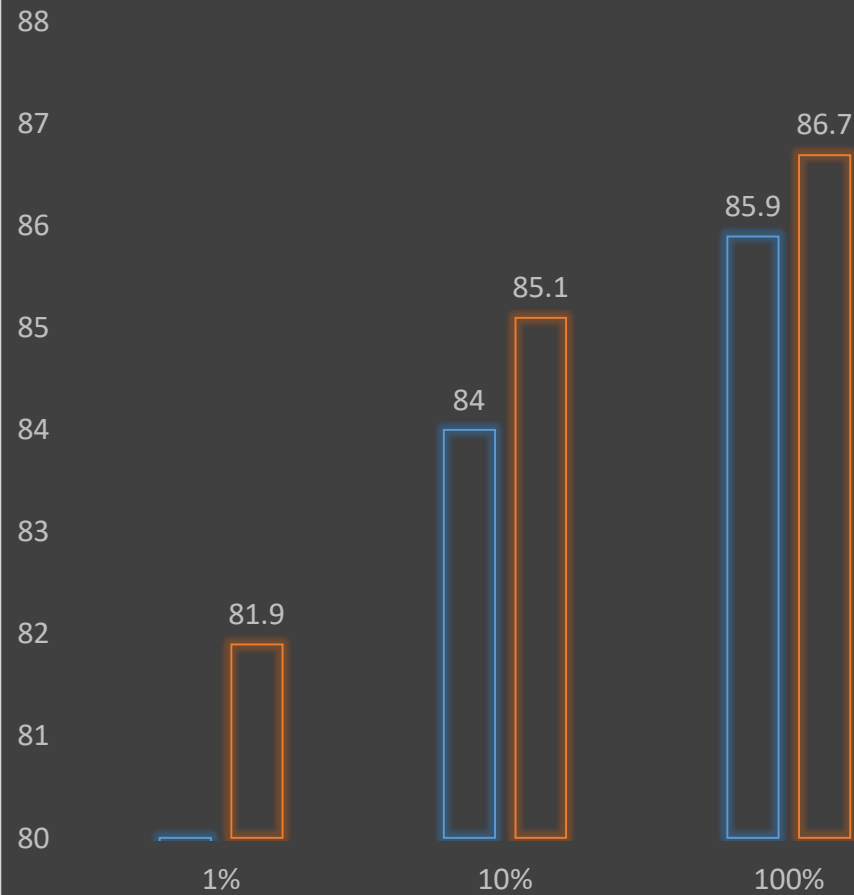


② Medical Vision Language FM ECAMP



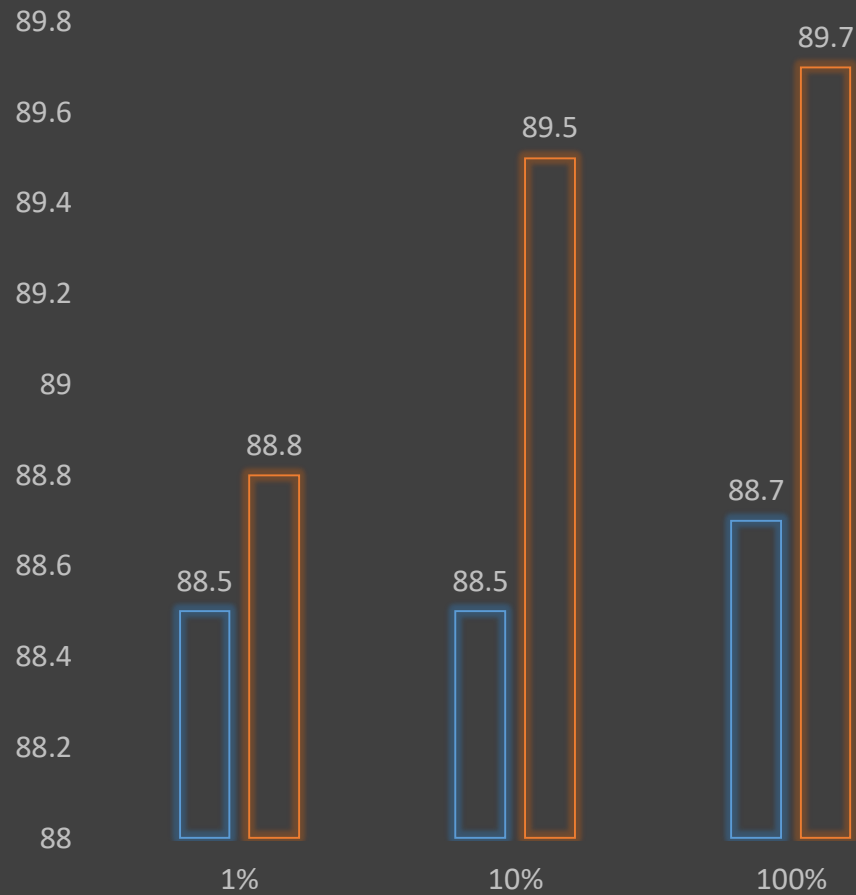
ChestX-Ray14

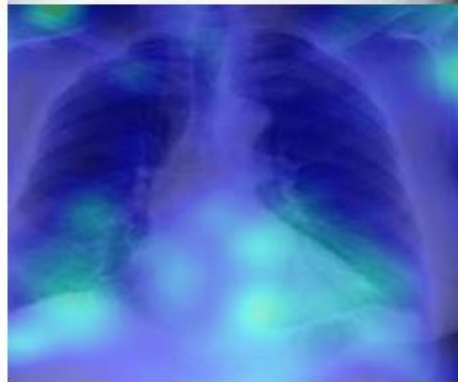
□ SOTA □ Ours



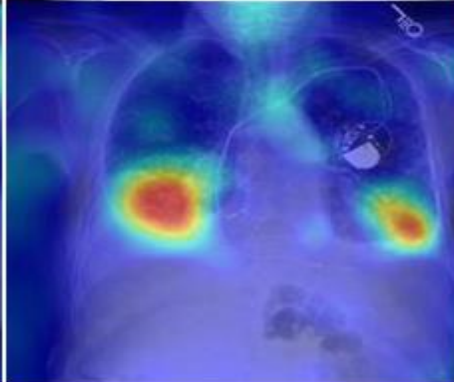
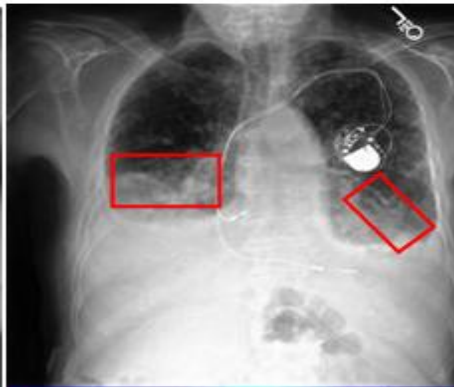
CheXpert

□ SOTA □ Ours

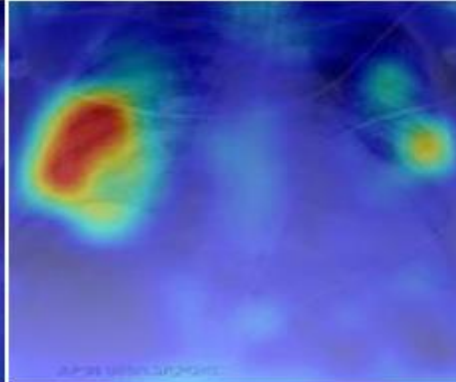
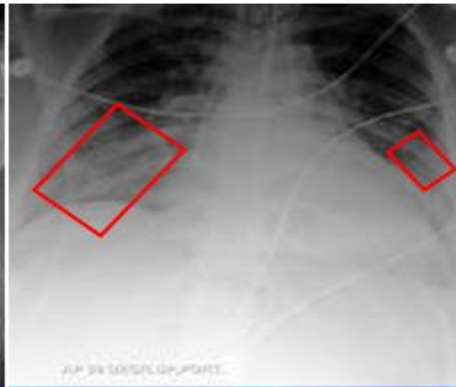




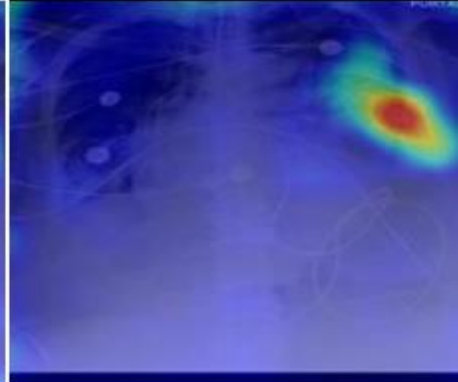
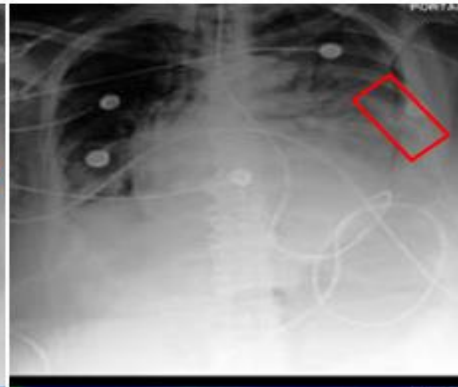
**There is no
disease.**



**There is bilateral
atelectasis.**



**There is
pulmonary
edema.**

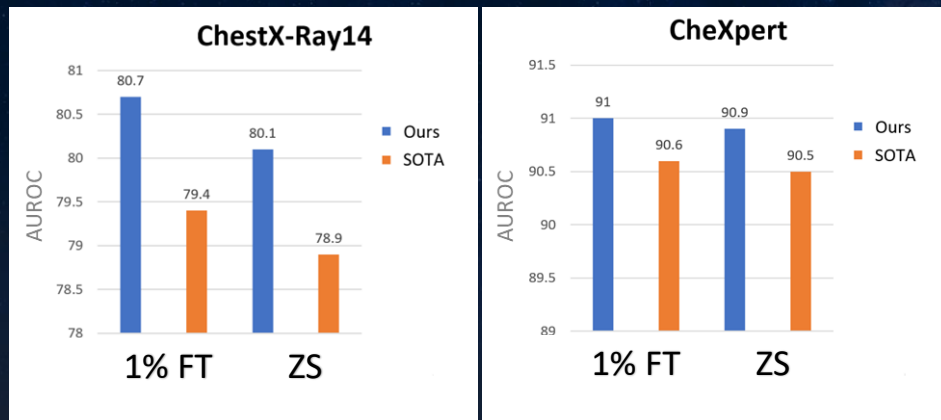
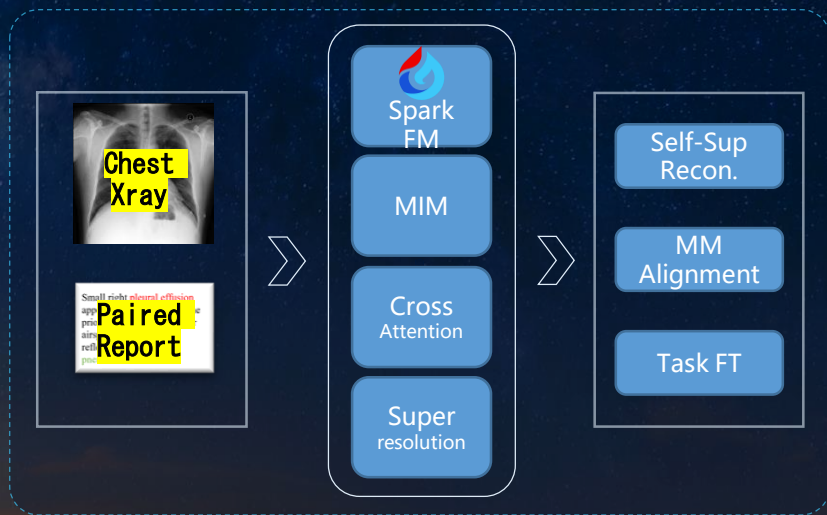


**There is
moderate
effusion.**

Multimodal Medical Foundation Model

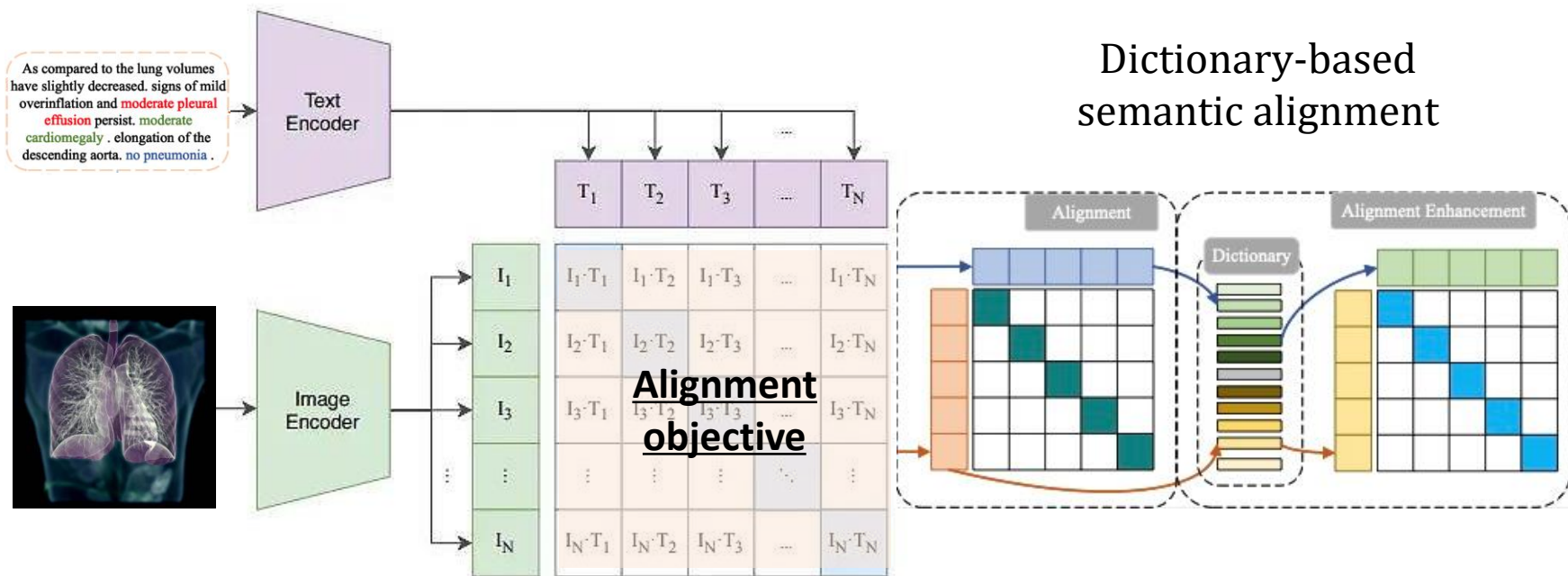
- Trained from ~1M <chest-xray, report> pairs

- Leading-edge accuracy on 1% fine-tuning (FT) and zero-shot(ZS) disease classification



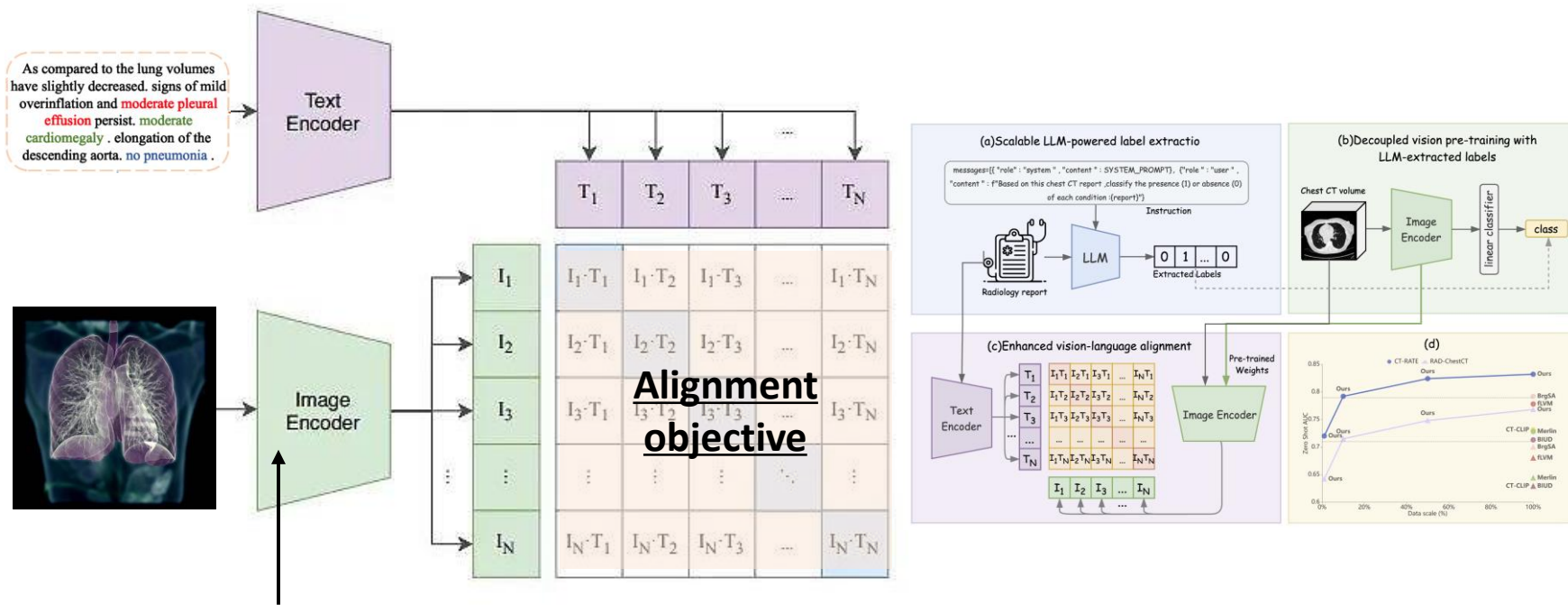
② Medical Vision Language FM

BrgSA: Bridged semantic alignment



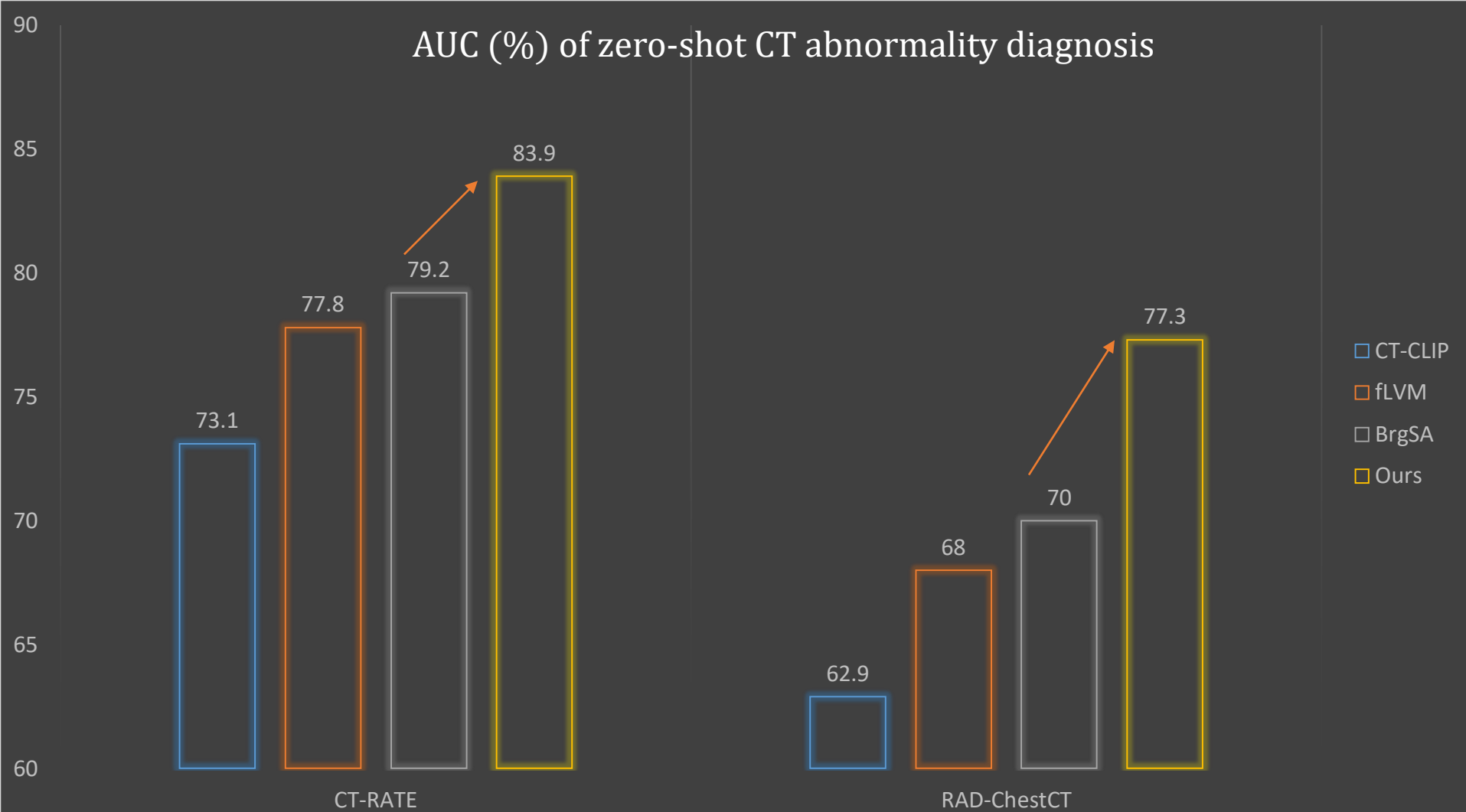
② Medical Vision Language FM

LLM-Pseudo-Label

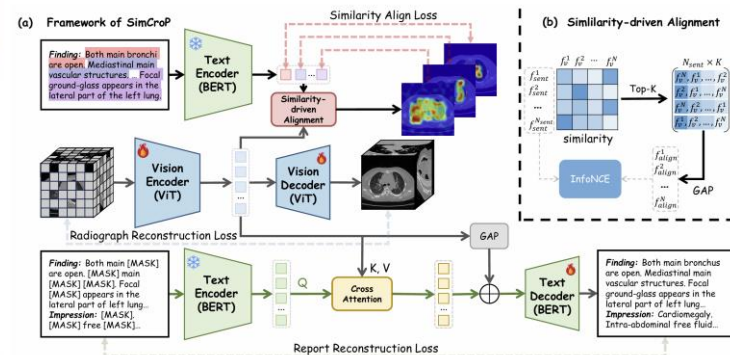
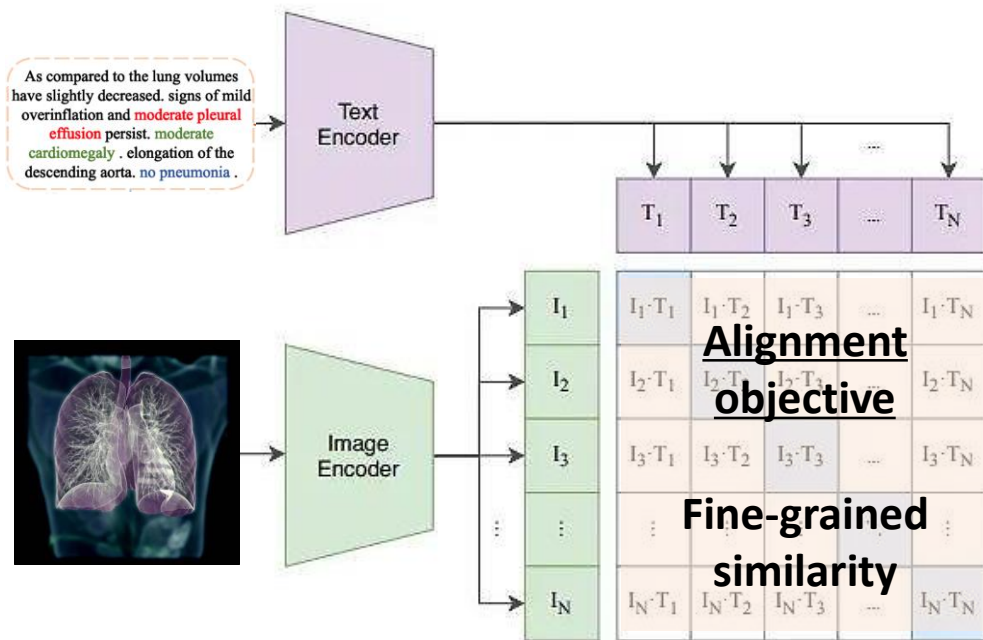


Pretrain abnormality classifier using LLM-based pseudo-labels

AUC (%) of zero-shot CT abnormality diagnosis

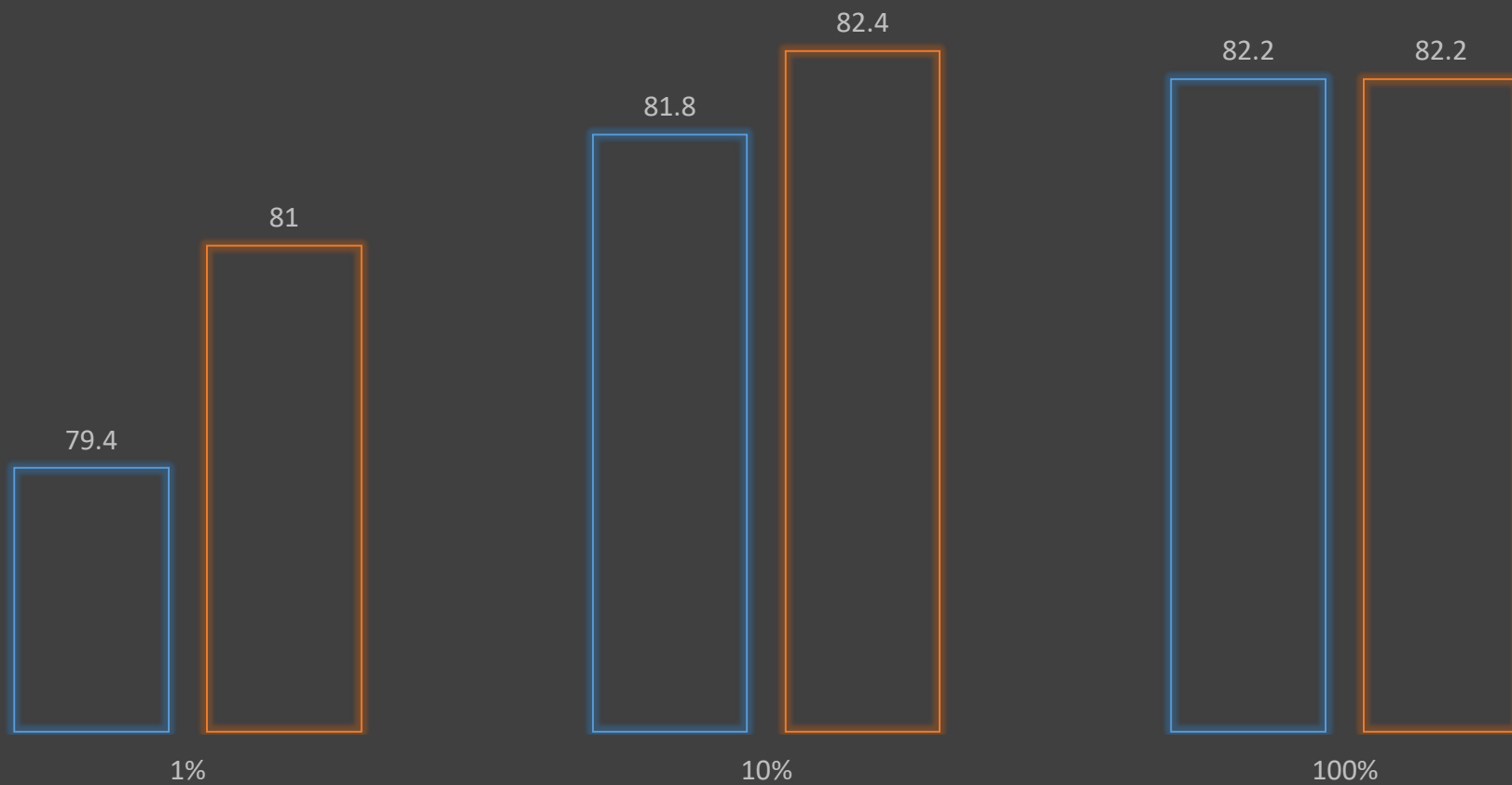


② Medical Vision Language FM SimCroP



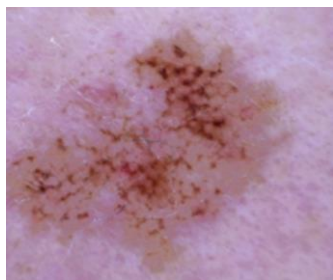
CT-Rate

□ SOTA □ Ours



② Medical Vision-Language FM

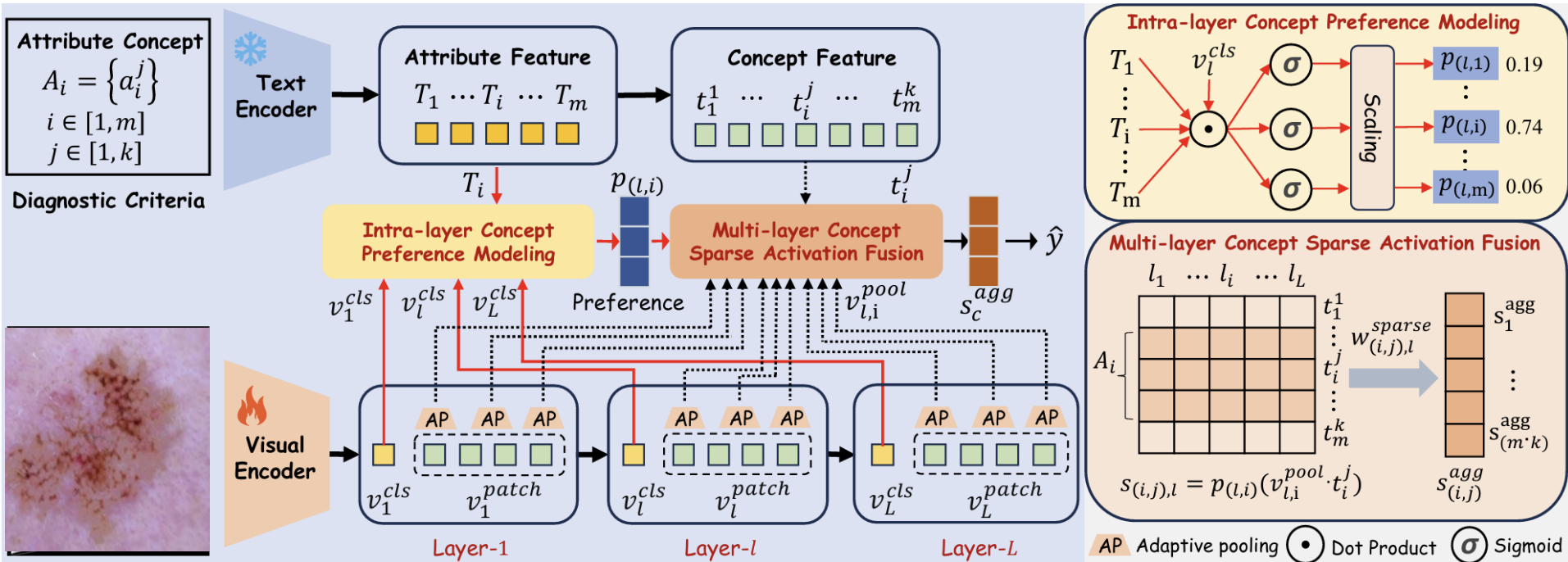
Concept bottleneck model (CBM)

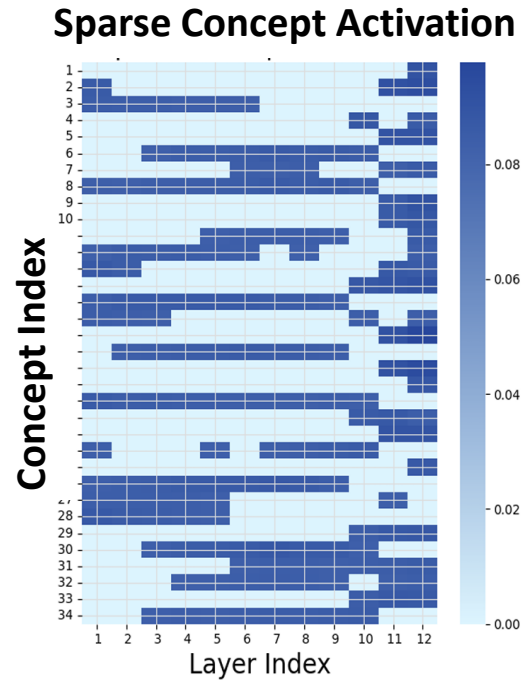
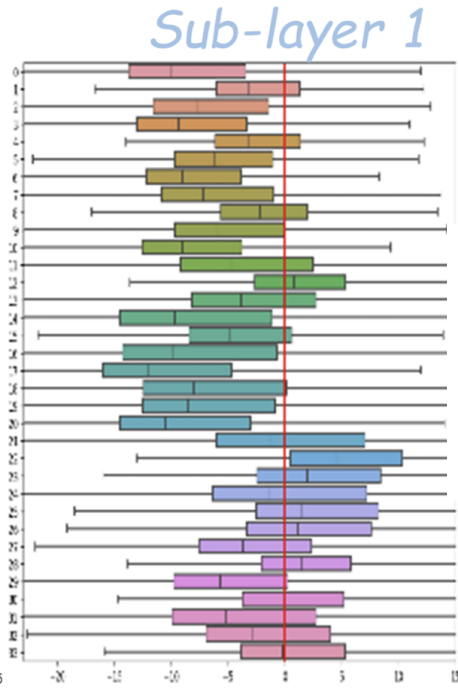
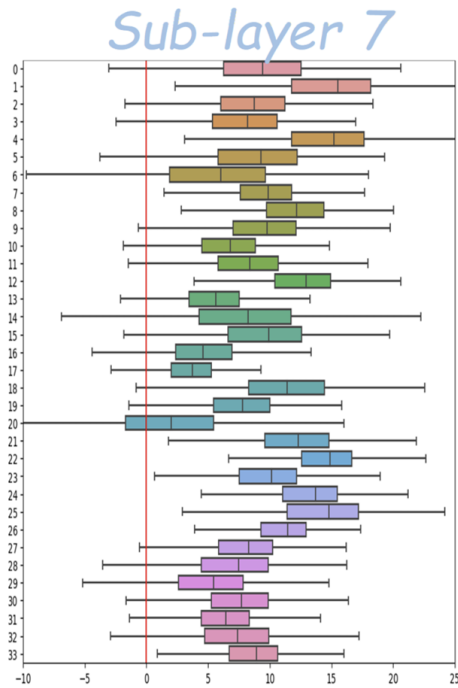
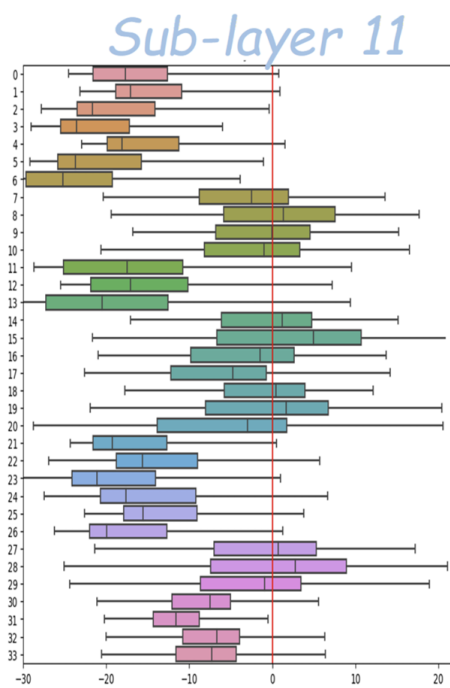


Attributes	Concepts
'color':	['highly variable, often with multiple colors (black, brown, red, white, blue)', 'uniformly tan, brown, or black', 'translucent, pearly white, sometimes with blue, brown, or black areas', 'red, pink, or brown, often with a scale', 'light brown to black', 'pink brown or red', 'red, purple, or blue']
'shape':	['irregular', 'round', 'round to irregular', 'variable']
'border':	['often blurry and irregular', 'sharp and well-defined', 'rolled edges, often indistinct']
'dermoscopic patterns':	['atypical pigment network, irregular streaks, blue-whitish veil, irregular', 'regular pigment network, symmetric dots and globules', 'arborizing vessels, leaf-like areas, blue-gray avoid nests', 'strawberry pattern, glomerular vessels, scale', 'cerebriform pattern, milia-like cysts, comedo-like openings', 'central white patch, peripheral pigment network', 'depends on type (e.g., cherry angiomas have red lacunae; spider angiomas have a central red dot with radiating legs)']
'texture':	['a raised or ulcerated surface', 'smooth', 'smooth, possibly with telangiectasias', 'rough, scaly', 'warty or greasy surface', 'firm, may dimple when pinched']
'symmetry':	['asymmetrical', 'symmetrical', 'can be symmetrical or asymmetrical depending on type']
'elevation':	['flat to raised', 'raised with possible central ulceration', 'slightly raised', 'slightly raised maybe thick']

② Medical Vision-Language FM

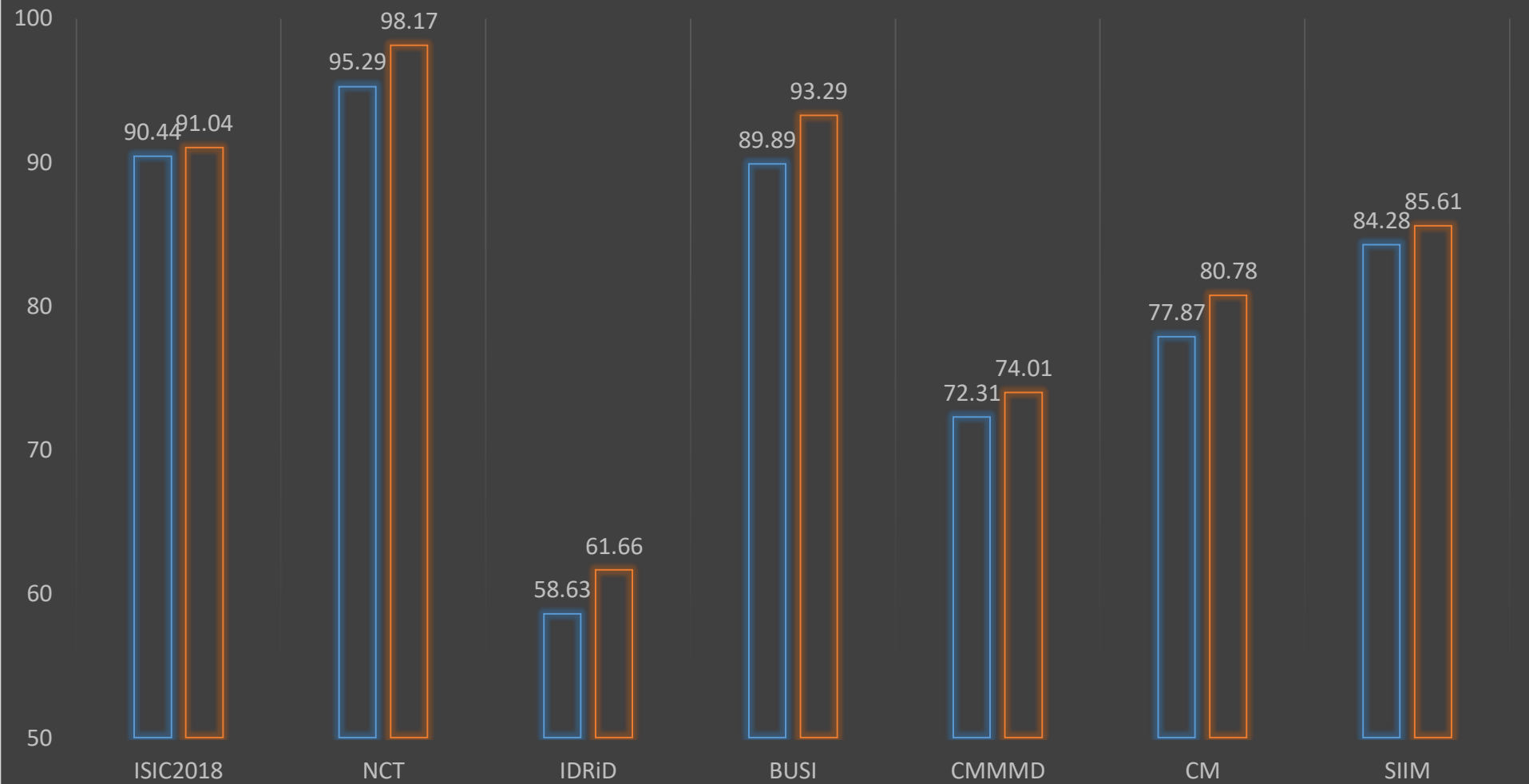
Multi-layer visual preference-enhanced CBM (MVP-CBM)





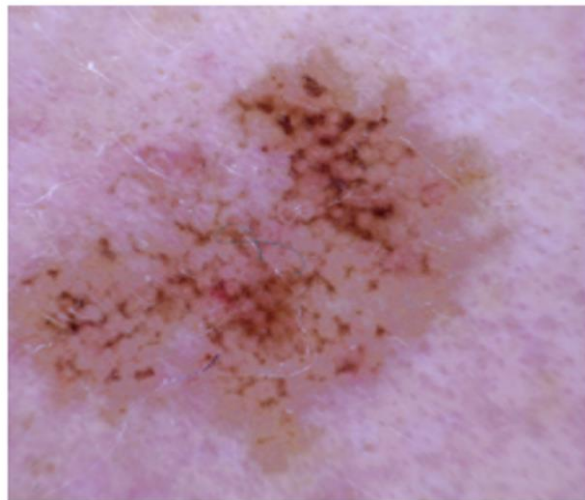
Concept Activation Values

ExpLid (sota) MVP-CBM (ours)



MVP-CBM

(a) ISIC-2018

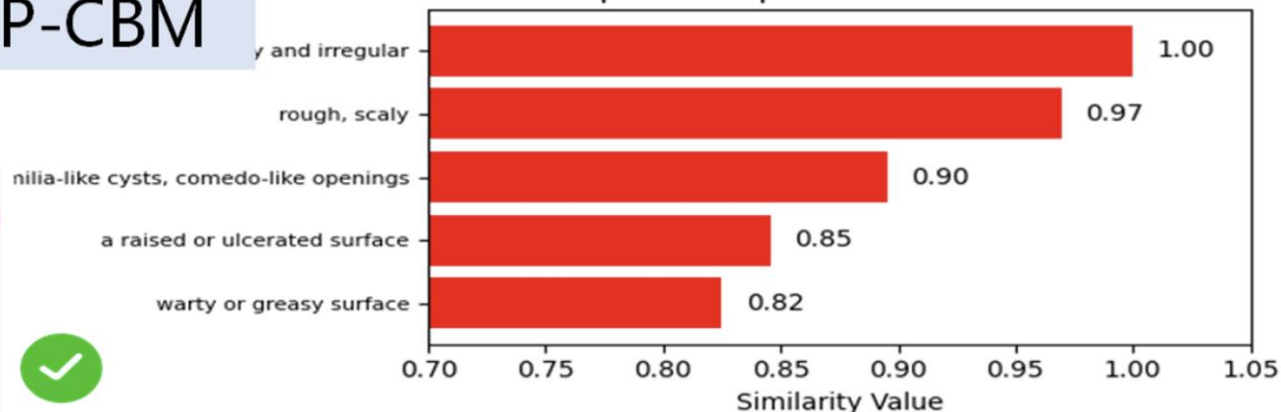


Melanoma

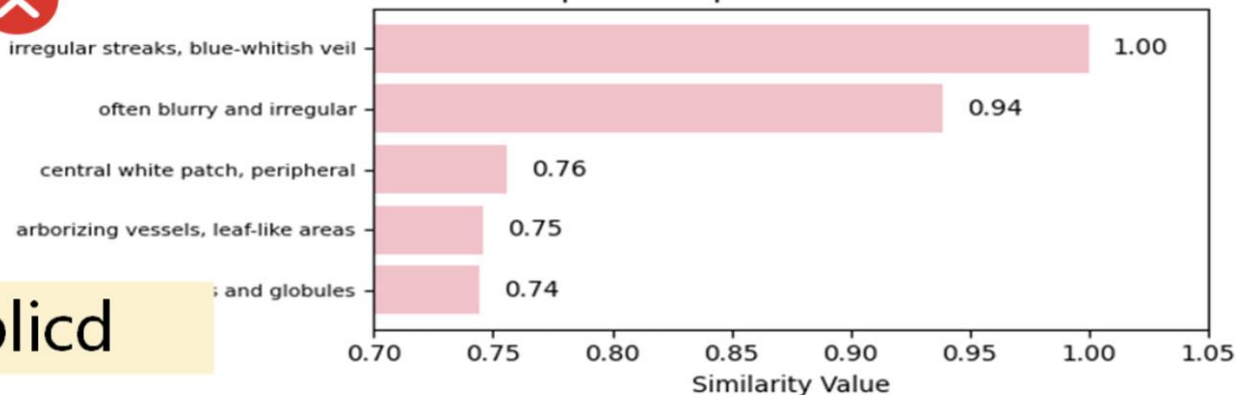
Explicd



Top 5 Concept Contribution Values



Top 5 Concept Contribution Values



Medial Imaging Foundation Model

- ① Medical Image Feature FM
- ② Medical Vision Language FM
- ③ Medical Image Generation FM
- ④ Medical Image Segmentation FM



MI-UFM: Challenges and Trends

Challenges

- Data curation
- Label error
- Hallucination
- Computing (training/infer.)
- Clinical applications

Trends

- More modalities [image, report, WSI, gene, omics]
- More data [ten millions]
- More tasks [diagnosis, prognosis]
- Generative + Discriminative AI
- Agent (multiple FMs)
- Model compression
- Trustworthy, causal, interpretable
- Ethics & safety

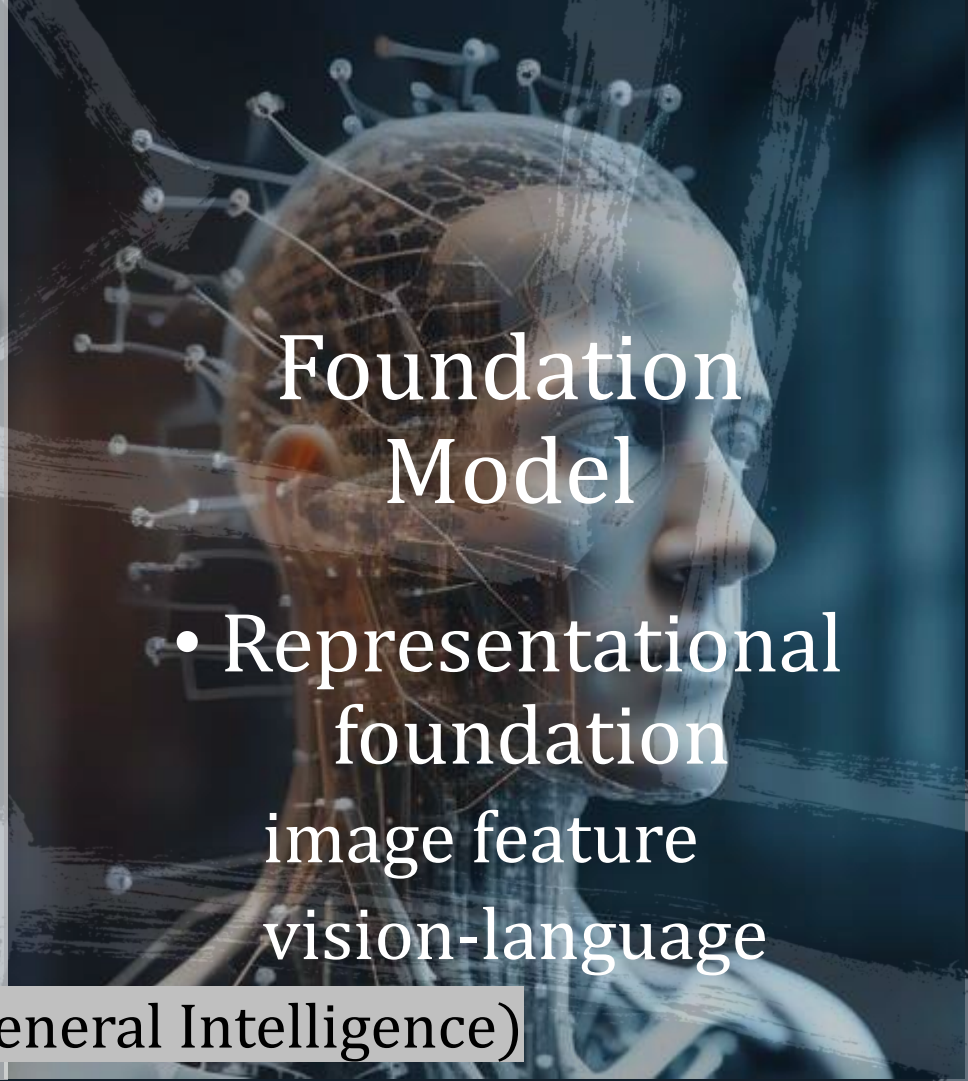


Universal Model

- Task universality

commonality + individuality

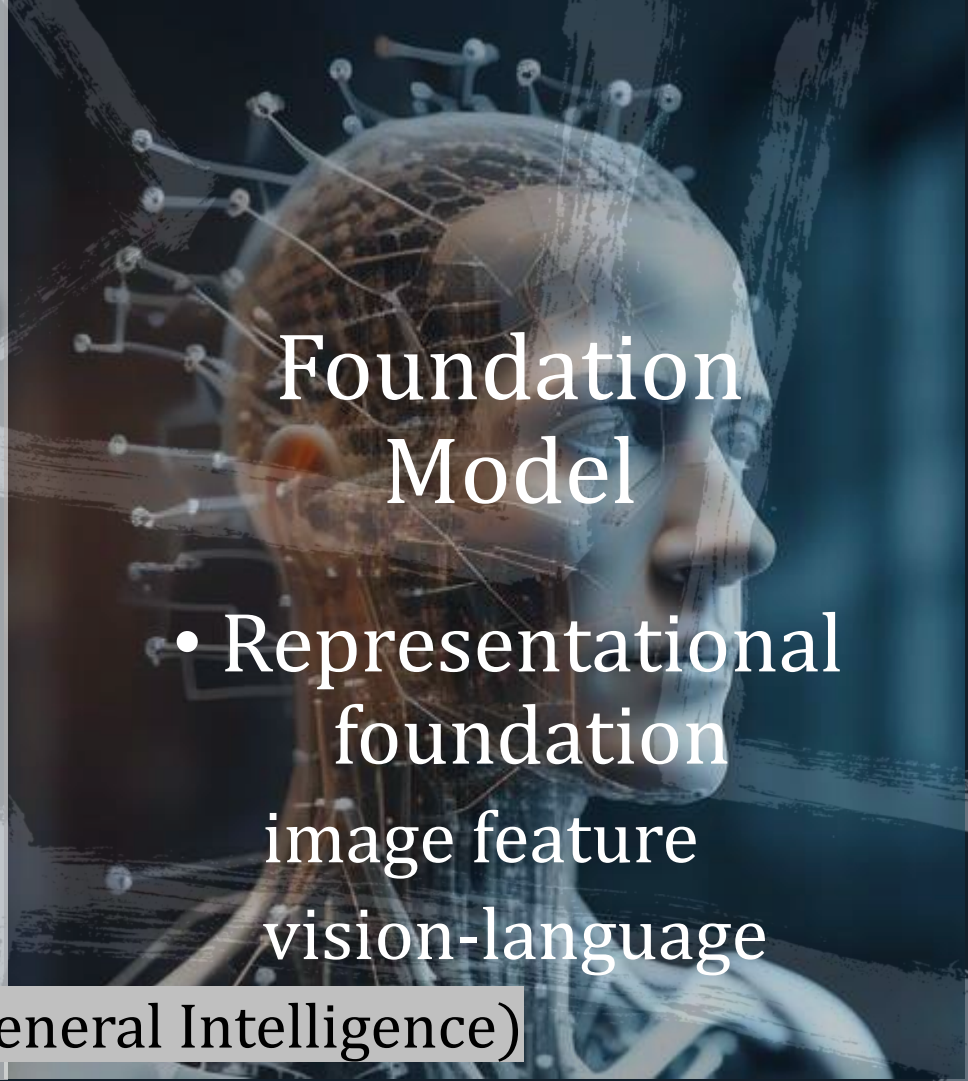
- 
- # Universal Model
- Task universality
- commonality + individuality



Foundation Model

- Representational foundation
image feature
vision-language

General Intelligence)

- 
- # Foundation Model
- Representational foundation
image feature
vision-language
- General Intelligence)

Goal: AGI (Artificial General Intelligence)

Hinton's Q: When will AI replace radiologist?



the
decoder

DE

AI and society

May 15, 2025

Geoffrey Hinton's wildly overconfident AI prediction failed—now it's a lesson in humility

Factors

- Data (privacy, assets)
- Algorithm (alignment, reasoning)
- Compute or electricity (enough?)
- Regulation (?)





我希望是…一万年

