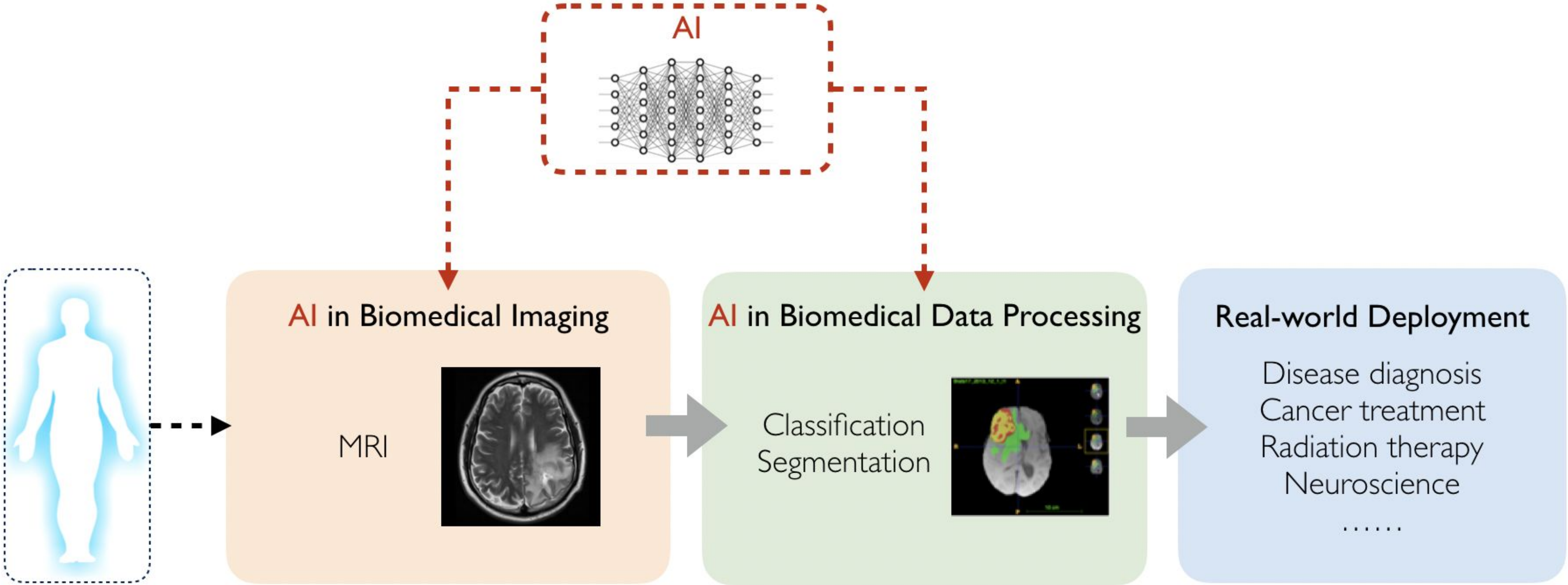
Lecture 5:
Medical Image Analysis

Announcements

- Next Thursday: Guest lecture on genomics and electronic health records (EHRs)
    - Reference paper 1
    - Reference paper 2
- For presenters next week:
    - Presentation slides due this Sunday 11:59 pm (only for the first-week presenters)
    - Please include your quiz questions AND answers in the last slide
- For all students:
    - Paper review due will be changed for next Wednesday due to guest lecture (more updates will be on Canvas)
    - Bonus credits for idea proposer
- Paper assignment and presentation schedule:
    - Already got most of confirmation!
    - Because of new drop/enroll, still in the finalizing …

In this class:

- Part I: AI in biomedical imaging
- Part II: AI in biomedical data processing



Today's agenda

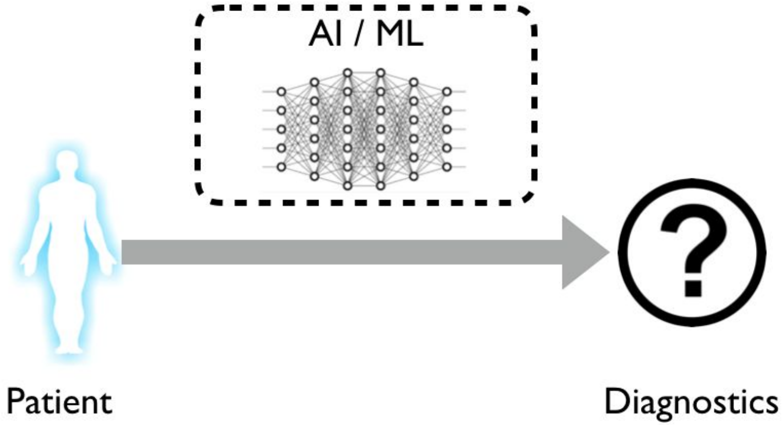
- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

Today's agenda

- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

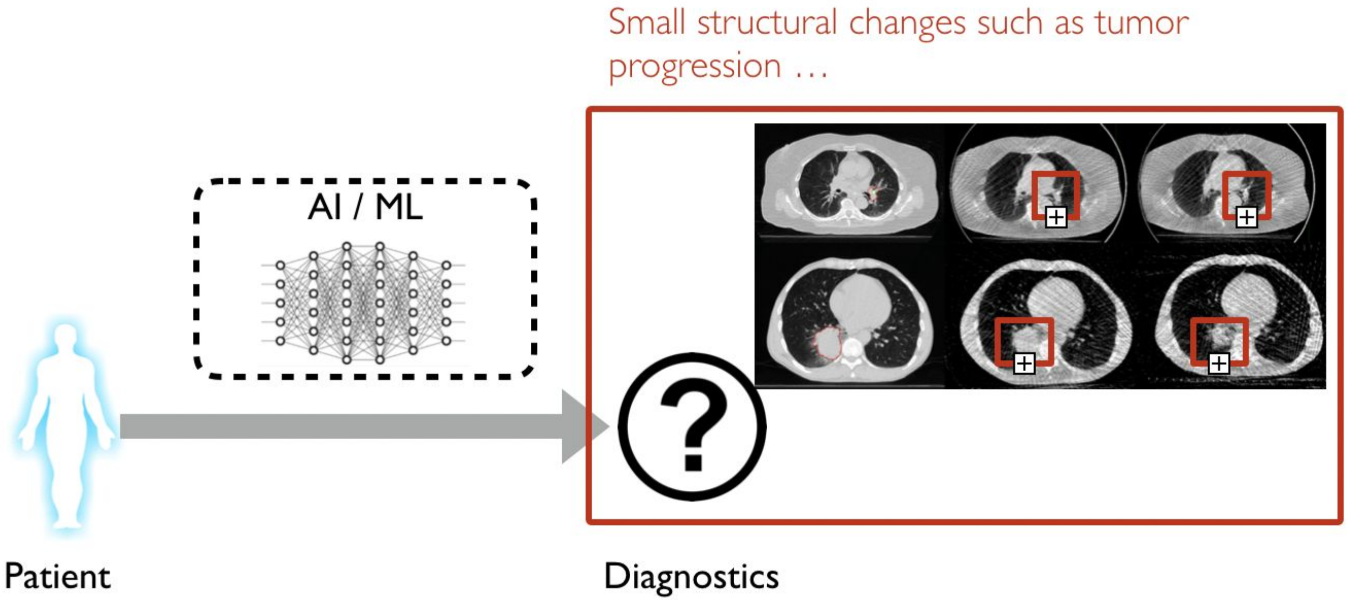
Challenges of Biomedical AI

- Reliability
- Generation
- Data efficiency
- Multi-modality



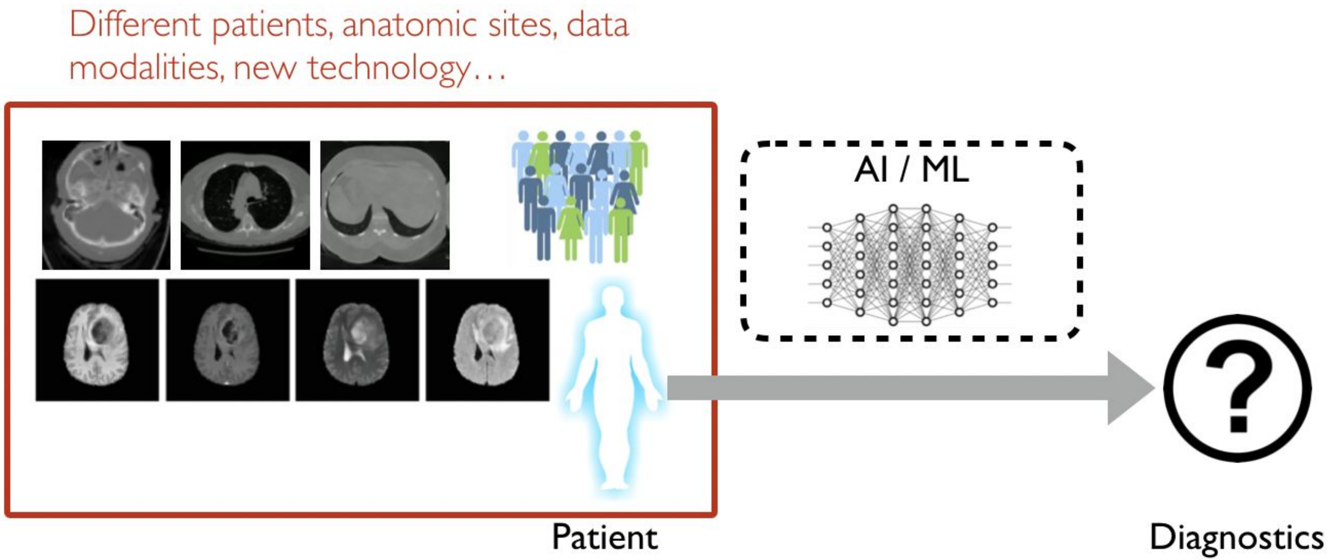
Challenges of Biomedical AI: Reliability

- AI can be insensitive to tumor shrinkage, lesion progression, etc.



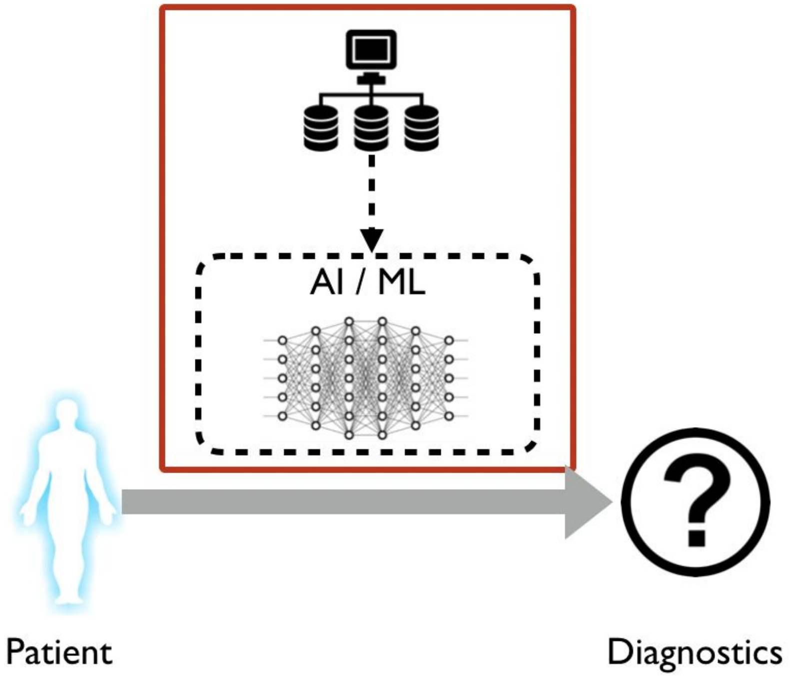
Challenges of Biomedical AI: Generalization

- AI model is limited when generalized to different input data



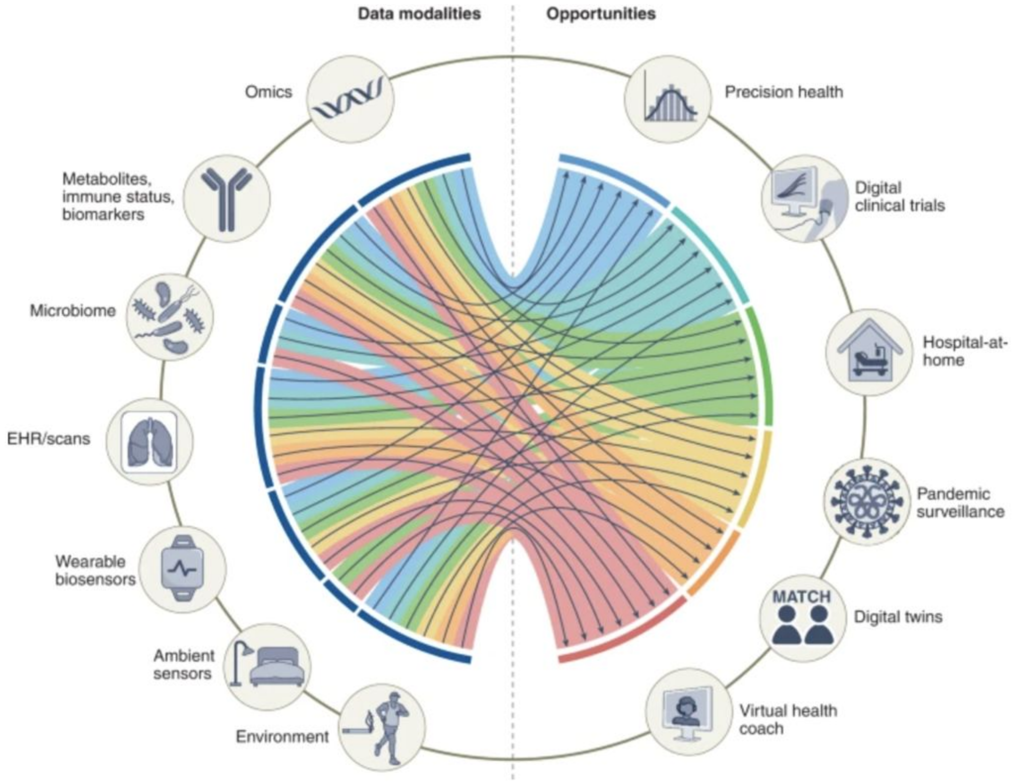
Challenges of Biomedical AI: Data efficiency

- Data collection and annotation can be a bottleneck for AI model development



Challenges of Biomedical AI: Multi-modality

- Biomedical data is inherently multi-modality



Today's agenda

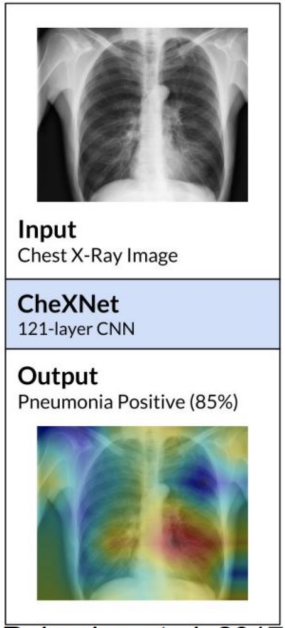
- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

*Some slides in this section are adapted from: [Stanford BIODS220: Artificial Intelligence in Healthcare]

Medical image classification

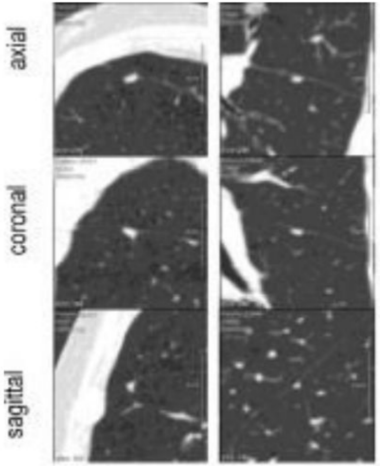
- Example tasks

Is an x-ray positive for pneumonia or not?



Rajpurkar et al. 2017

Is a CT scan of a lung nodule benign or not?

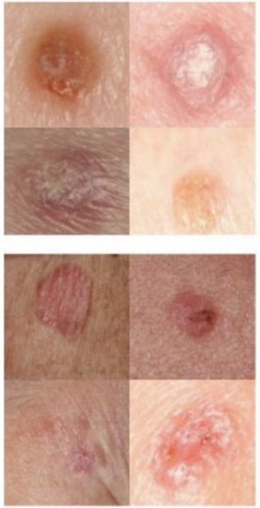


Ciompi et al. 2015

Is this moderate or worse diabetic retinopathy?

Gulshan et al. 2016

What types of skin lesions are these?



Esteva et al. 2017

Medical image classification

Traditional machine learning

Input → **Feature extractor** → **Machine learning model** → **Output**

(e.g., )

(e.g., color and texture histograms)

(e.g., support vector machines and random forests)

(e.g., presence or not or disease)

Medical image classification

Traditional machine learning

Input → **Feature extractor** → **Machine learning model** → **Output**

(e.g., )

(e.g., color and texture histograms)

(e.g., support vector machines and random forests)

(e.g., presence or not or disease)

Deep learning

Input → **Deep Learning Model** → **Output**

(e.g., )

(e.g., convolutional and recurrent neural networks)

(e.g., presence or not or disease)

Medical image classification

Traditional machine learning

Input → **Feature extractor** → **Machine learning model** → **Output**

(e.g., )

(e.g., color and texture histograms)

(e.g., support vector machines and random forests)

(e.g., presence or not or disease)

Deep learning

Directly learns what are useful (and better!) features from the training data

Input → **Deep Learning Model** → **Output**

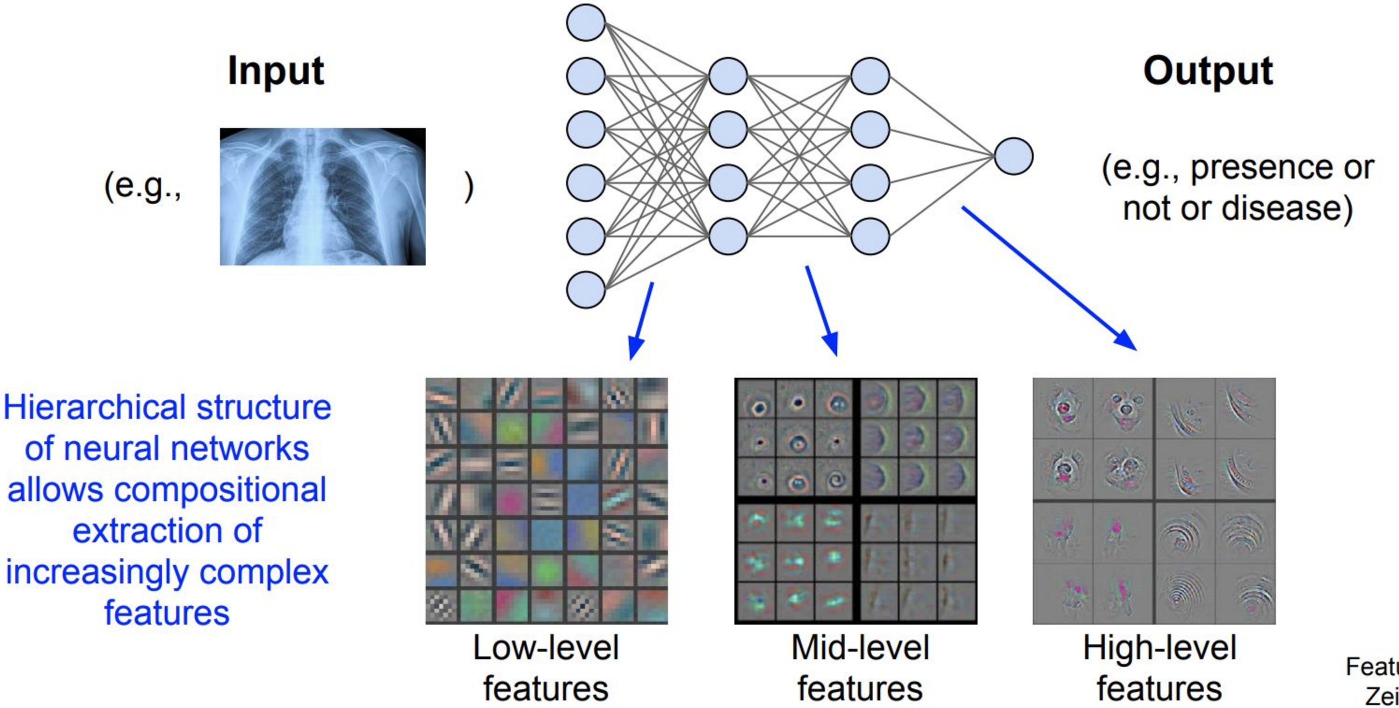
(e.g., )

(e.g., convolutional and recurrent neural networks)

(e.g., presence or not or disease)

Medical image classification

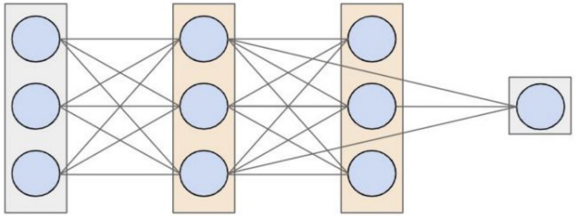
- Deep learning models perform feature extraction



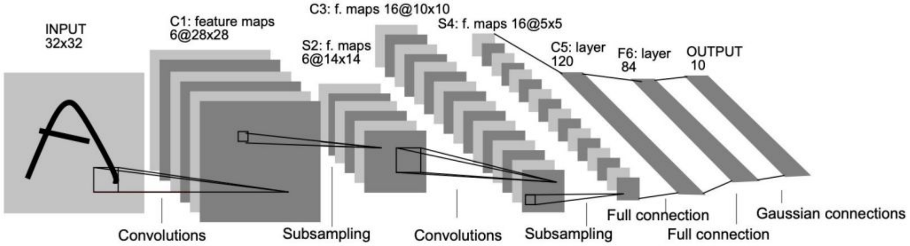
Feature visualizations from Zeiler and Fergus 2013

Medical image classification

- Different classes of deep neural networks

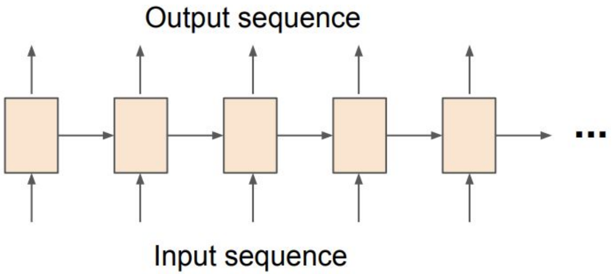


Fully connected neural networks
(linear layers, good for "feature vector" inputs)



Convolutional neural networks
(convolutional layers, good for image inputs)

Recurrent neural networks
(linear layers modeling recurrence relation across sequence, good for sequence inputs)



Medical image classification

- Regression task:
  - Predict a single real-valued output from input data

Model input: data vector $x = [x_1, x_2, ..., x_N]$ **Model output:** prediction (single number) $\hat{y}$

Example: predicting hospital length-of-stay from clinical variables in the electronic health record

$x =$ [age, weight, …, temperature, oxygen saturation] $\hat{y} =$ length-of-stay (days)

Example: predicting expression level of a target gene from the expression levels of N landmark genes

$x \in \mathcal{R}^N =$ expression levels of N landmark genes $\hat{y} =$ expression level of target gene

Medical image classification

- Classification task:
  - Predict a categorical output from input data

Model input: data vector $x = [x_1, x_2, ..., x_N]$

Model output: prediction of 1-of-K classes

$$\hat{y} \in \{1, ..., K\}$$

Example: predicting in-hospital mortality from clinical variables in the electronic health record

$x =$ [age, weight, …, temperature, oxygen saturation]

$\hat{y} \in \{0, 1\}$ for occurrence of in-hospital mortality

Medical image classification

- Different training loss functions

Regression Minimize squared difference between prediction output and target

$$L_{regression} = \frac{1}{M}\sum_i (\hat{y}^i - y^i)^2$$

Label is a continuous value.

Equivalent to the negative log of the probability of the correct ground truth class being predicted. Think about what the expression looks like when y_i = 1 vs. 0.

Binary Cross-Entropy

$$L_{BCE} = \frac{1}{M}\sum_i -(y_i log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i))$$

Label is binary in {0,1}. Prediction is a real number in (0,1) and is the probability of the label being 1. It is usually the output of a sigmoid operation after the final layer.

Softmax Negative log of the probability of the true class y_i, as with the BCE loss.

$$L_{Softmax} = \frac{1}{M}\sum_i -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

Label is 1 of K classes in {0, …, K}. Extension of binary cross-entropy loss to multiple classes. s_j corresponds to the score (e.g. output of final layer) for each class; the fraction in the log provides a normalized probability for each class.

SVM Incurs lowest loss of 0 (what we want) if the score for the true class y_i is greater than the score for each incorrect class j by a margin of 1

$$L_{SVM} = \frac{1}{M}\sum_i \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

Label is 1 of K classes in {0, …, K}. Same use case as softmax, but different way of encouraging the model to produce outputs that we “like”. In practice, softmax is more popular and provides a nice probabilistic interpretation.

Medical image classification

- Different evaluation metrics

Confusion matrix

		Prediction 0	Prediction 1
Ground Truth	0	TN	FP
	1	FN	TP

Accuracy: (TP + TN) / total

Sensitivity / Recall (true positive rate)**:**
TP / total positives

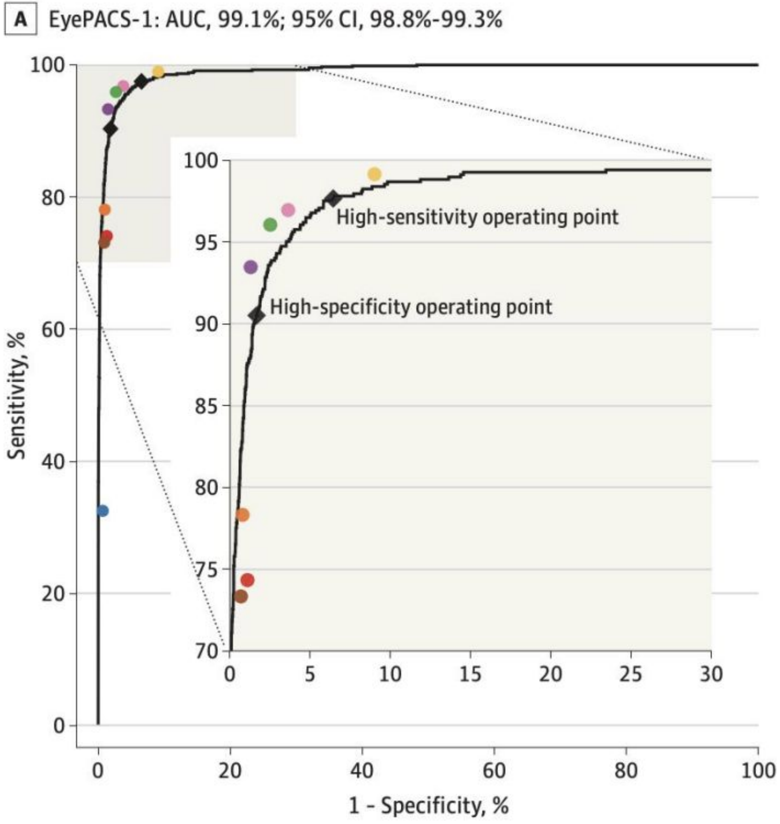
Specificity (true negative rate)**:**
TN / total negatives

Precision (positive predictive value)**:**
TP / total predicted positives

Negative predictive value:
TN / total predicted negatives

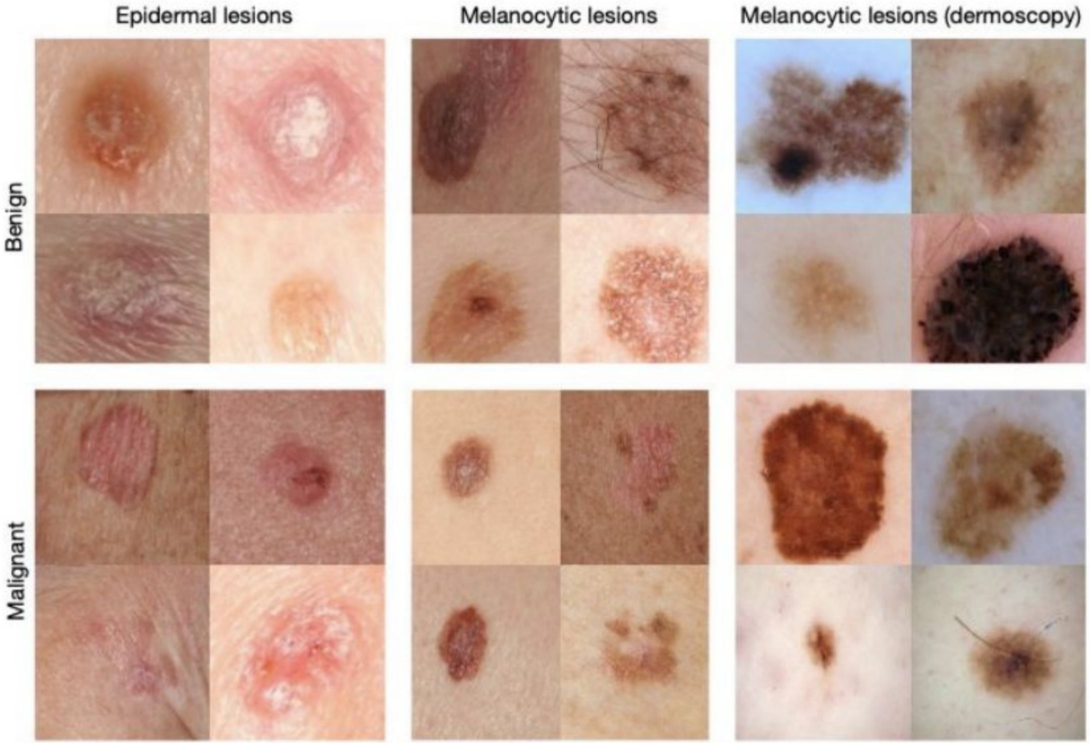
Medical image classification

- Receiver Operating Characteristic (ROC) curve:
  - Plots sensitivity and specificity (in fact, 1 - specificity) as prediction threshold is varied
  - Gives trade-off between sensitivity and specificity
  - Report summary statistic AUC (area under the curve)



Medical image classification

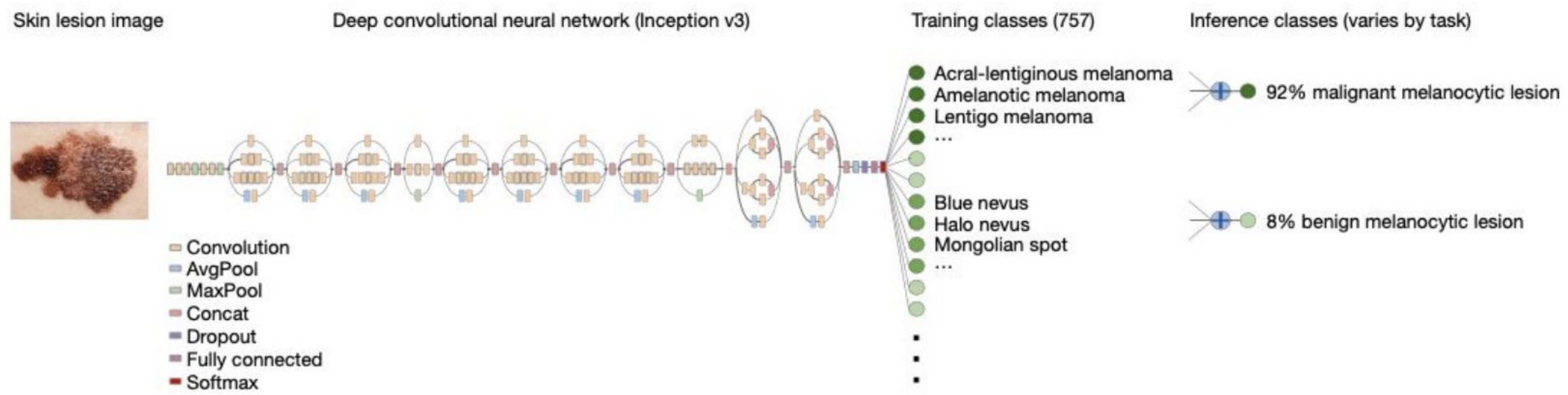
- Skin image classification
  - Two binary classification tasks: malignant vs. benign for epidermal lesions or melanocytic lesions



Esteva, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017.

Medical image classification

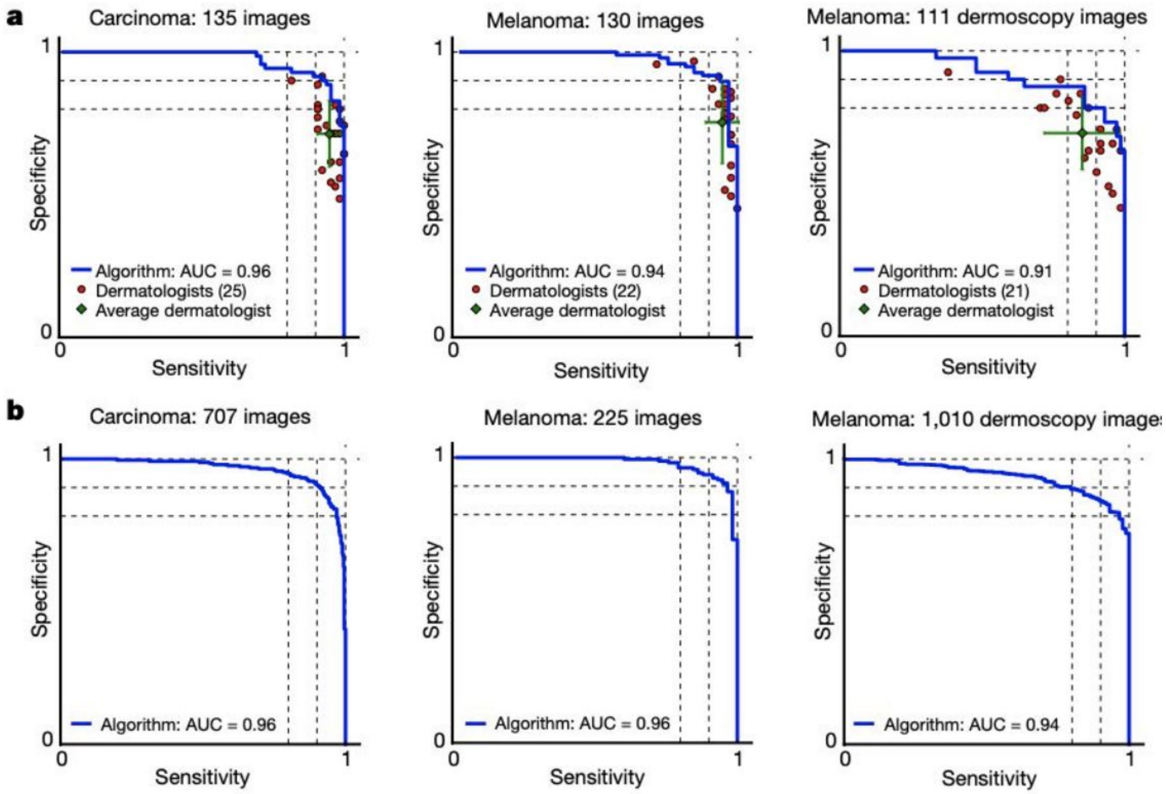
- Model structure:
  - Inception-v3 (GoogLeNet) CNN with ImageNet pre-training
- Dataset:
  - Fine-tuned on dataset of 129,450 lesions (from several sources) comprising 2,032 diseases



Esteva, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017.

Medical image classification

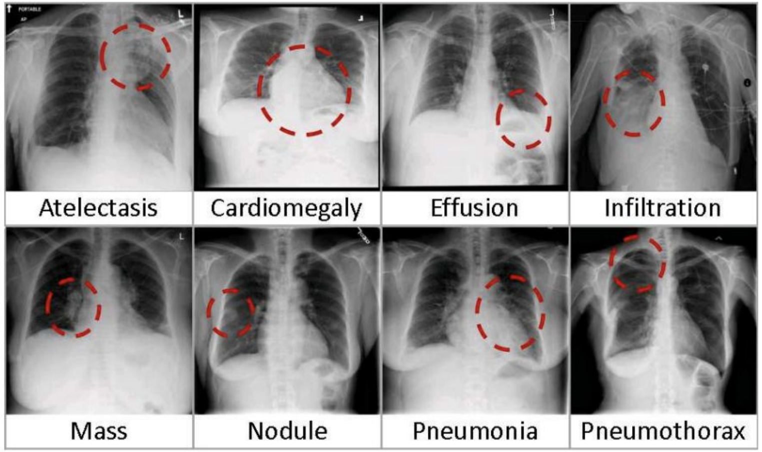
- Evaluation:
  - Deep learning algorithm vs. dermatologists



Esteva, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017.

Medical image classification

- New large-scale dataset of X-ray images:
  - 8 common thoracic diseases observed in chest X-rays
  - Labeling disease names by text mining from raw reports
  - ChestX-ray8 database is composed of 108,948 frontal-view X-ray images (from 32,717 patients)

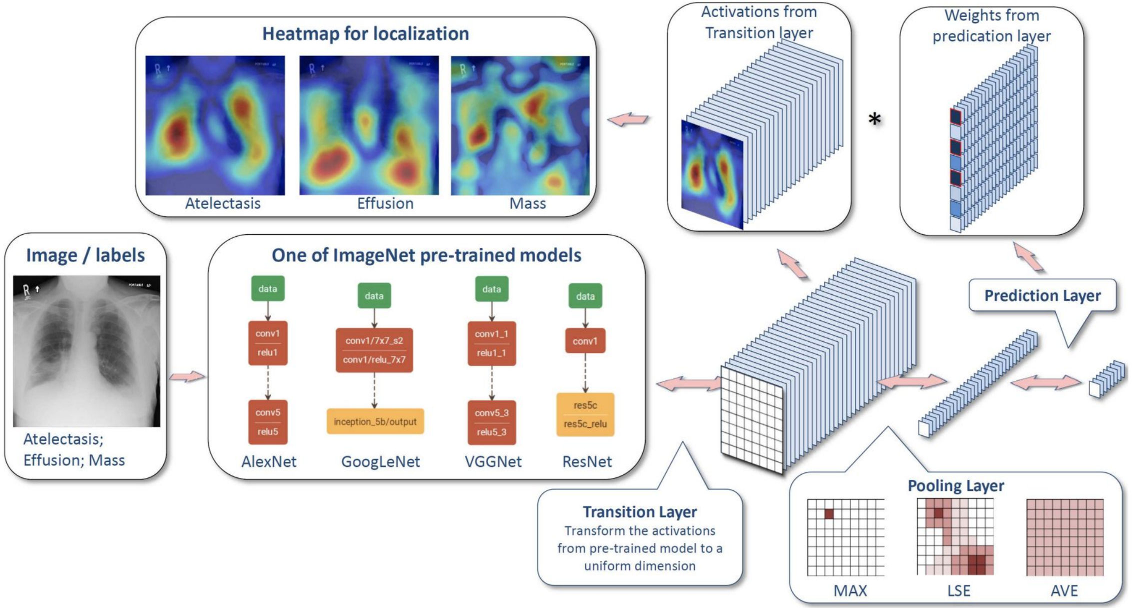


Item #	OpenI	Ov.	ChestX-ray8	Ov.
Report	2,435	-	108,948	-
Annotations	2,435	-	-	-
Atelectasis	315	122	5,789	3,286
Cardiomegaly	345	100	1,010	475
Effusion	153	94	6,331	4,017
Infiltration	60	45	10,317	4,698
Mass	15	4	6,046	3,432
Nodule	106	18	1,971	1,041
Pneumonia	40	15	1,062	703
Pneumothorax	22	11	2,793	1,403
Normal	1,379	0	84,312	0

Wang, et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. CVPR 2017.

Medical image classification

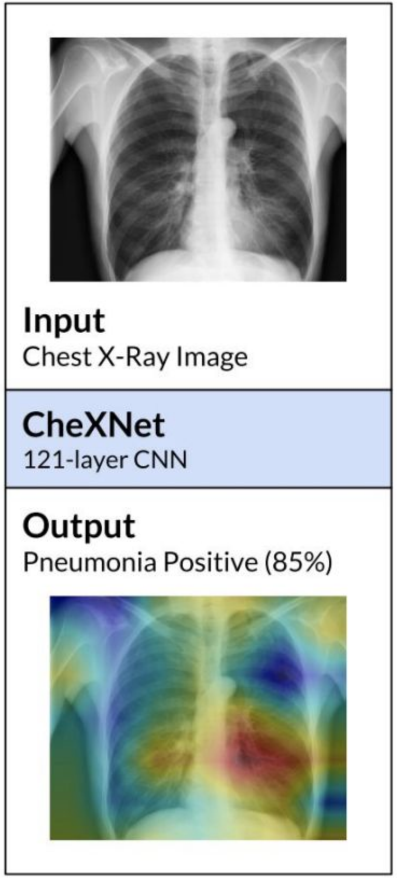
- Multi-label classification and weakly-supervised localization:
  - Benchmarks with different ImageNet pre-trained models



Wang, et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. CVPR 2017.

Medical image classification

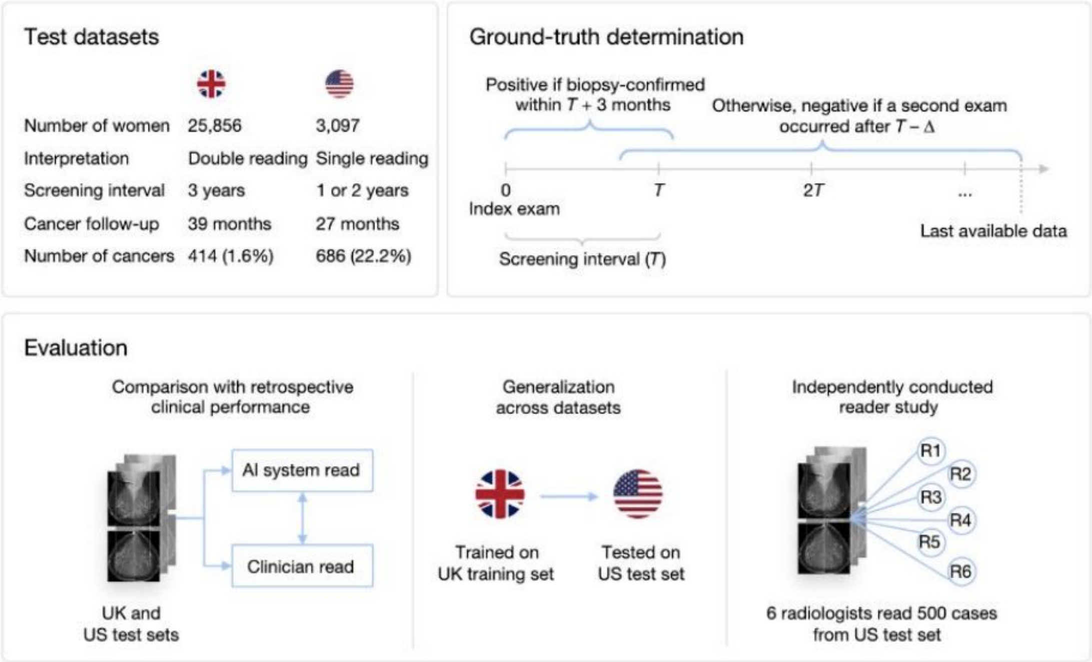
- X-ray image classification:
    - Binary classification of pneumonia presence in chest X-ray images
    - Used ChestX-ray14 dataset with over 100,000 frontal X-ray images with 14 diseases
    - Model structure: 121-layer DenseNet CNN
    - Compared algorithm performance with 4 radiologists



Rajpurkar et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. 2017.

Medical image classification

- Binary classification of breast cancer in mammograms
- Used an ensemble of models including ResNet
- International dataset and evaluation, across UK and US



McKinney et al. International evaluation of an AI system for breast cancer screening. Nature 2020.

Today's agenda

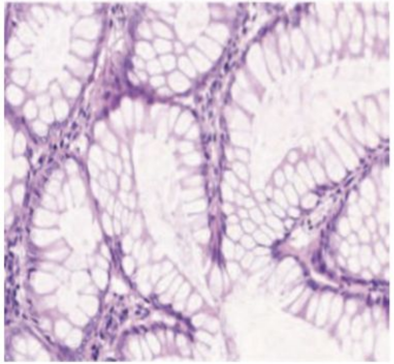
- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

**Some slides in this section are adapted from: [Stanford BIODS220: Artificial Intelligence in Healthcare]*

Medical image segmentation

- Different visual recognition tasks:

Classification

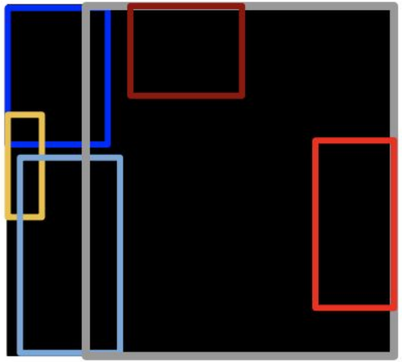


Output:
one category label for image (e.g., colorectal glands)

Semantic Segmentation

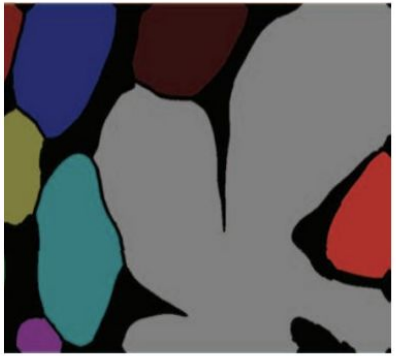
Output:
category label for each pixel in the image

Detection



Output:
Spatial bounding box for each **instance** of a category object in the image

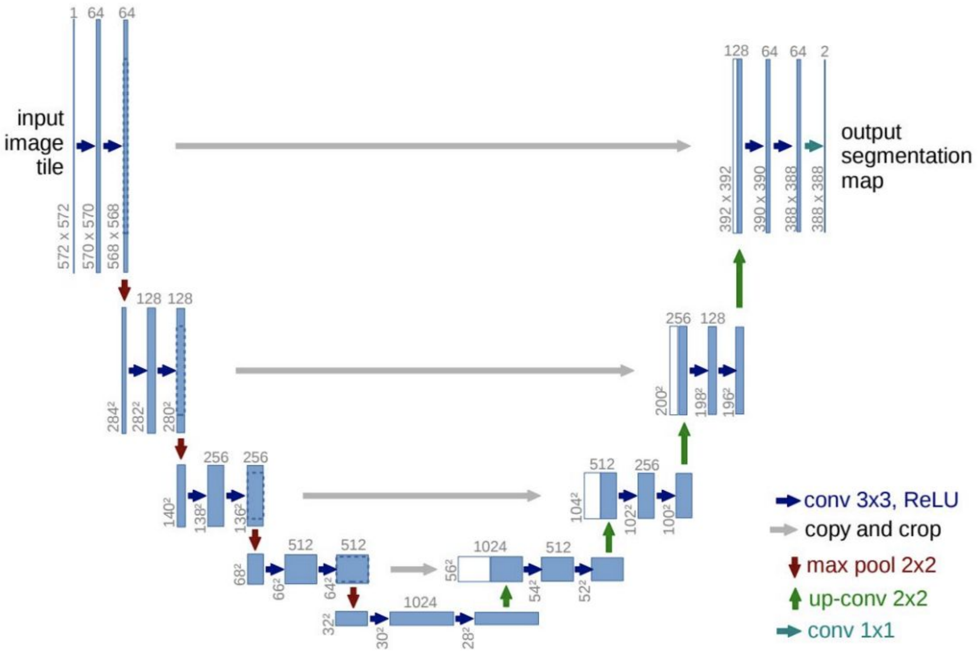
Instance Segmentation



Output:
Category label and instance label for each pixel in the image

Medical image segmentation

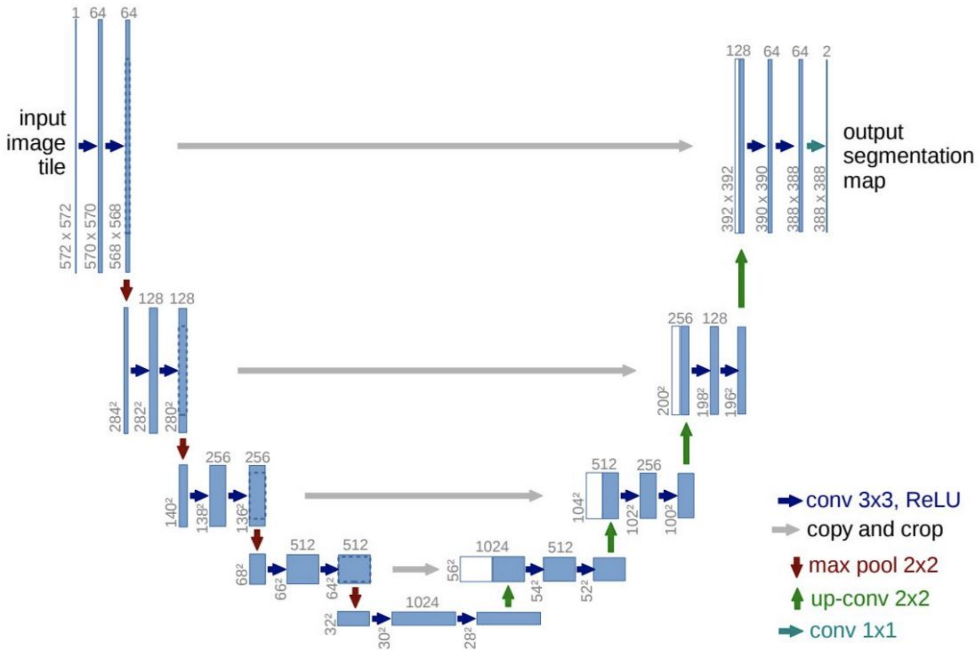
- U-Net model: 69000+ citations since 2015



Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015.

Medical image segmentation

- U-Net model: 69000+ citations since 2015

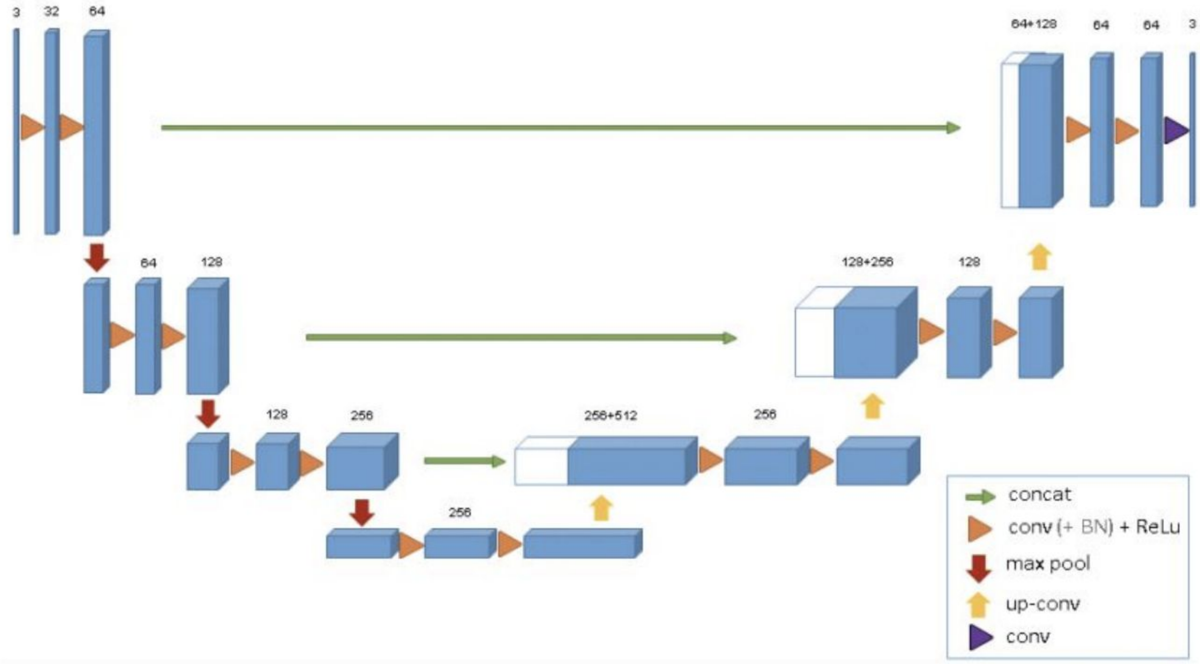


Skip-Connection:
Concatenate with same-resolution feature map during downsampling process to combine high-level information with low-level (local) information

Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015.

Medical image segmentation

- 3D U-Net model



Cicek et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. MICCAI 2016.

Medical image segmentation

- Intersection over Union (IoU) evaluation metric:

pixels included in both target and prediction maps

$$IoU = \frac{target \cap prediction}{target \cup prediction}$$

Total # pixels in the union of both masks

Medical image segmentation

- Dice coefficient evaluation metric:

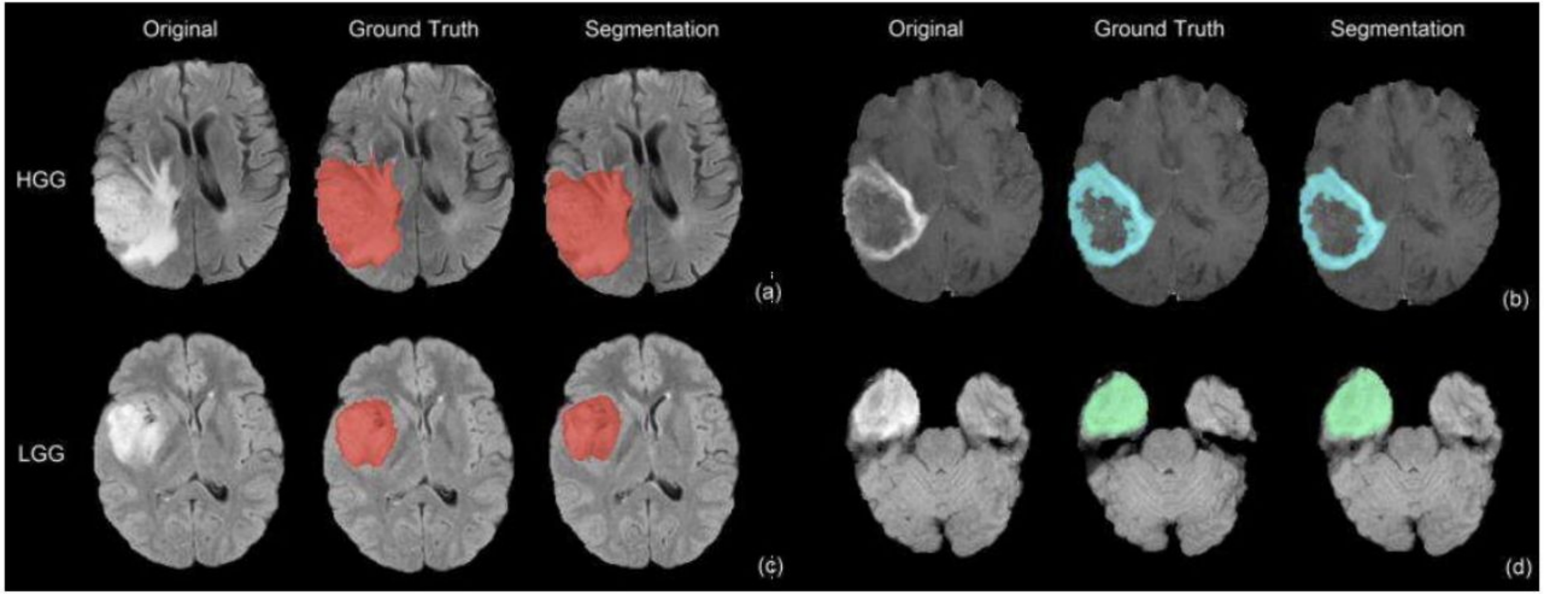
2 * intersection

$$\text{Dice Coefficient} = \frac{2 * (\text{target} \cap \text{prediction})}{\#\ \text{target mask pixels} + \#\ \text{prediction mask pixels}}$$

Sum of target mask size
+ prediction mask size

Medical image segmentation

- Brain Tumor Segmentation (BraTS) Challenge
  - Segment brain diffuse glioma and sub-regions in multi-contrast brain MR images



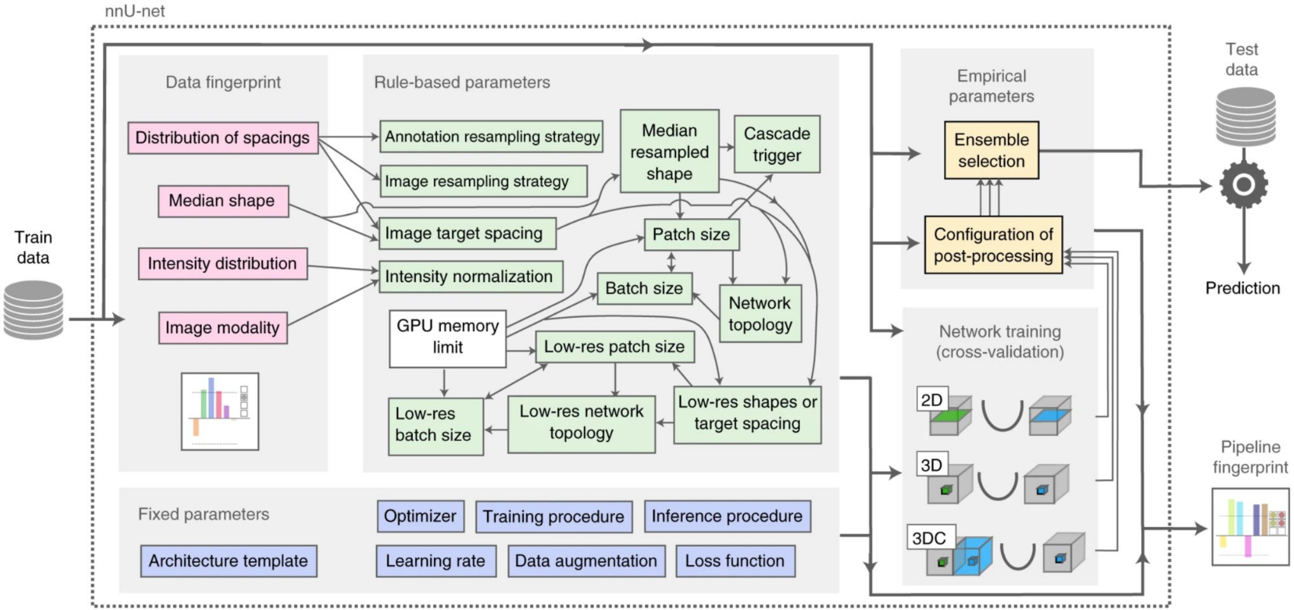
Medical image segmentation

- Brain Tumor Segmentation (BraTS) Challenge
  - Segment brain diffuse glioma and sub-regions in multi-contrast brain MR images
  - BraTS 2023 dataset: 5,880 MRI scans from 1,470 brain diffuse glioma patients
  - Hosted along with MICCAI every year

- BraTS2023 - Cluster of Challenges (Vancouver)- On-Going
- BraTS 2022 - Continuous Evaluation (Singapore) - On-Going
- BraTS 2021 (Strasbourg, France (Virtual)) - [proceedings coming soon]
- BraTS 2020 (Lima, Peru (Virtual)) - [proceedings: vol.1, vol.2]
- BraTS 2019 (Shenzhen, China) - [proceedings: vol.1, vol.2]
- BraTS 2018 (Granada, Spain) - [proceedings]
- BraTS 2017 (Quebec City, Canada) - [proceedings]
- BraTS 2016 (Athens, Greece) - [proceedings]
- BraTS 2015 (Munich, Germany) - [proceedings]
- BraTS 2014 (Boston, USA) - [proceedings]
- BraTS 2013 (Nagoya, Japan) - [proceedings]
- BraTS 2012 (Nice, France) - [proceedings]

Medical image segmentation

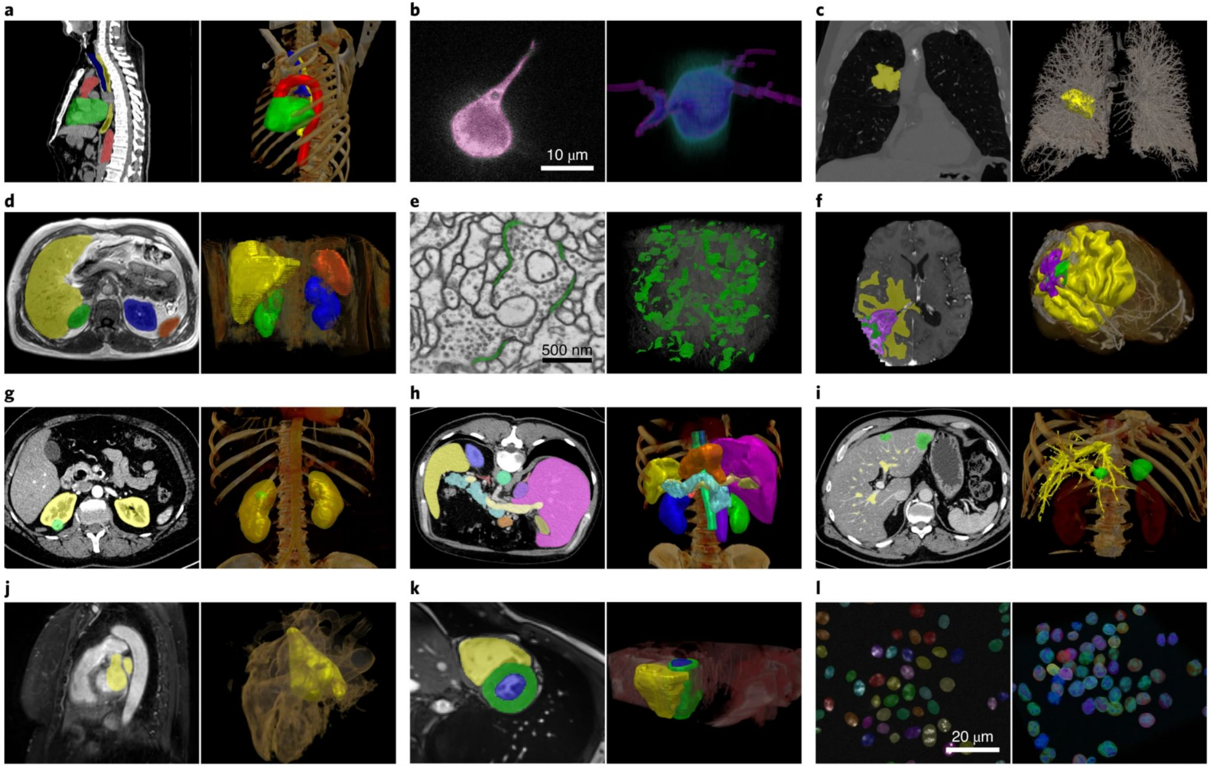
- nnU-Net:
  - A deep learning-based segmentation method that automatically configures itself, including preprocessing, network architecture, training and post-processing for any new task



Isensee et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 2020.

Medical image segmentation

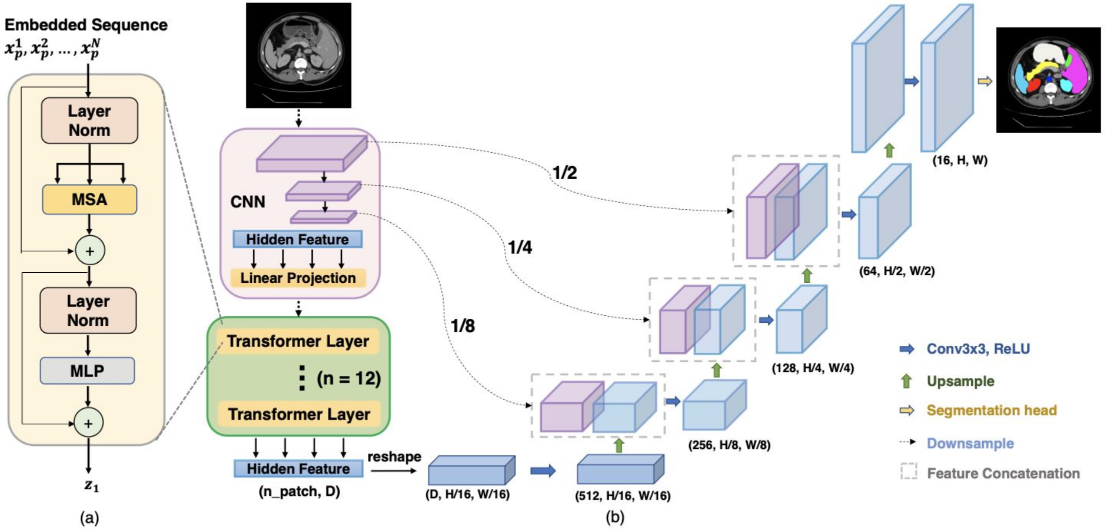
- nnU-Net:
  - SOTA on 23 public datasets used in international biomedical segmentation competitions



Isensee et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 2020.

Medical image segmentation

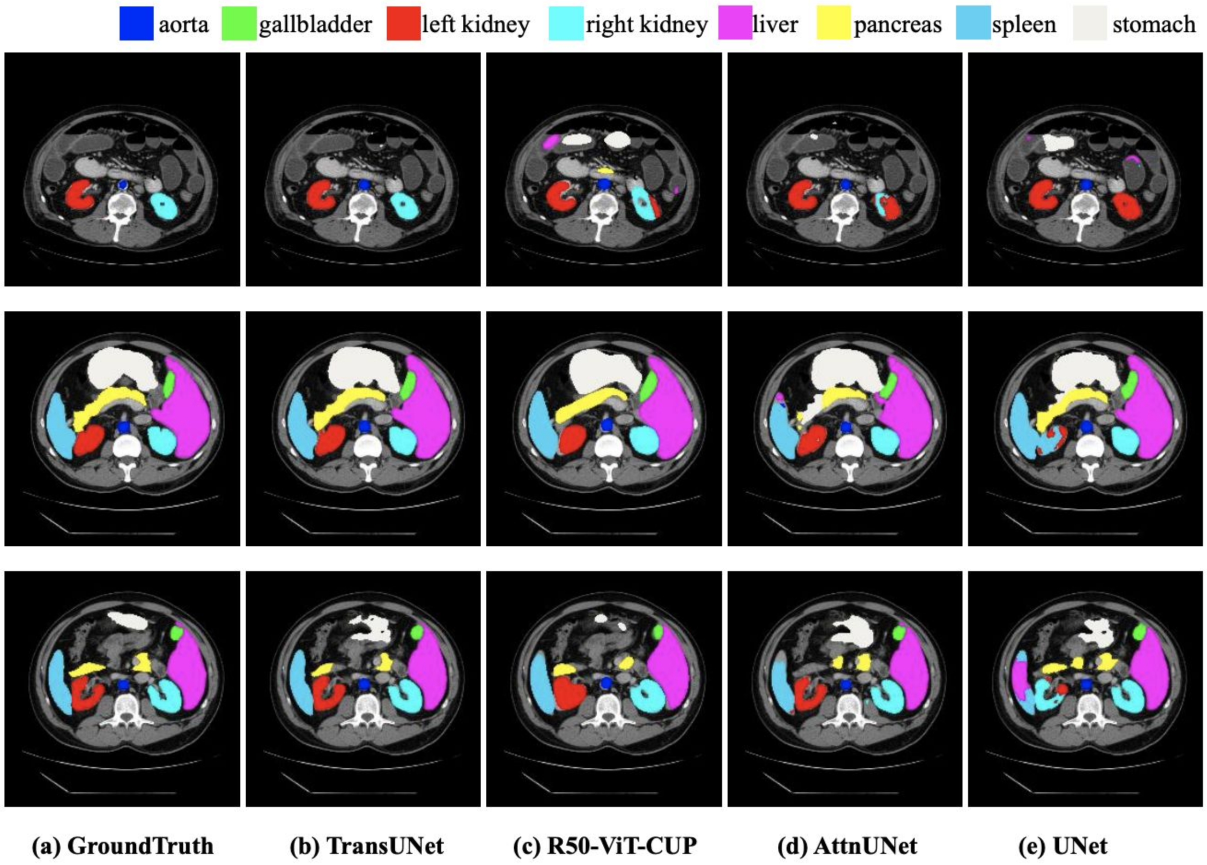
- TransUNet:
  - Transformer encodes tokenized image patches from CNN feature maps
  - Decoder upsamples encoded features which are then combined with the high-resolution CNN feature maps to enable precise localization



Chen et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv 2021.

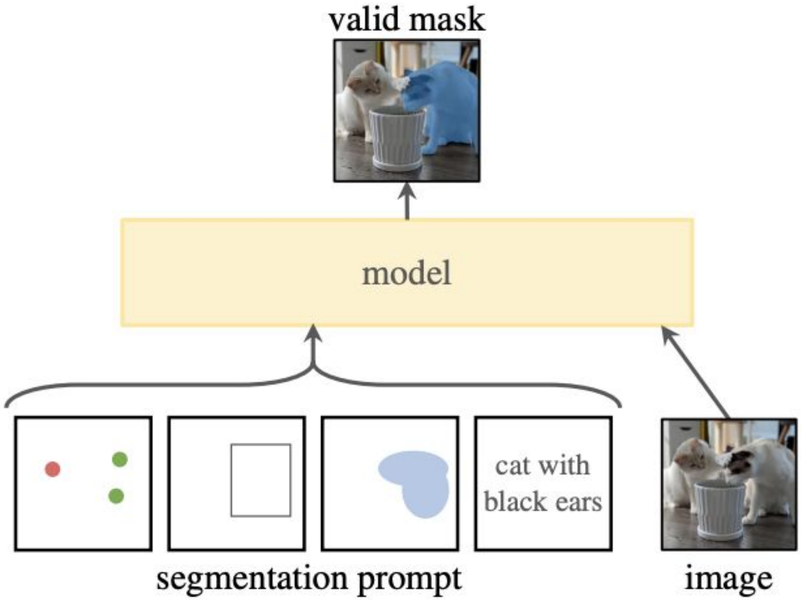
Medical image segmentation

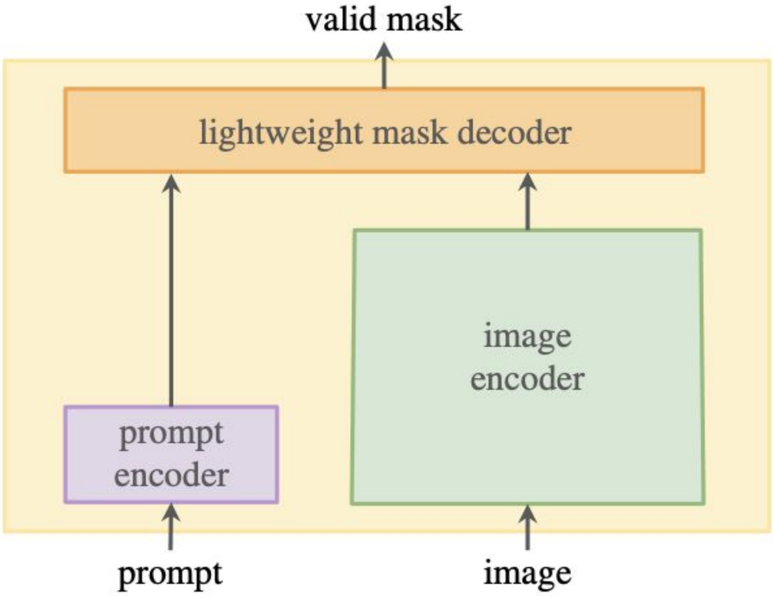
- TransUNet:



Medical image segmentation

- Vision foundation model for segmentation:
  - Segment Anything Model (SAM): enable zero-shot transfer to a range of tasks via prompt engineering





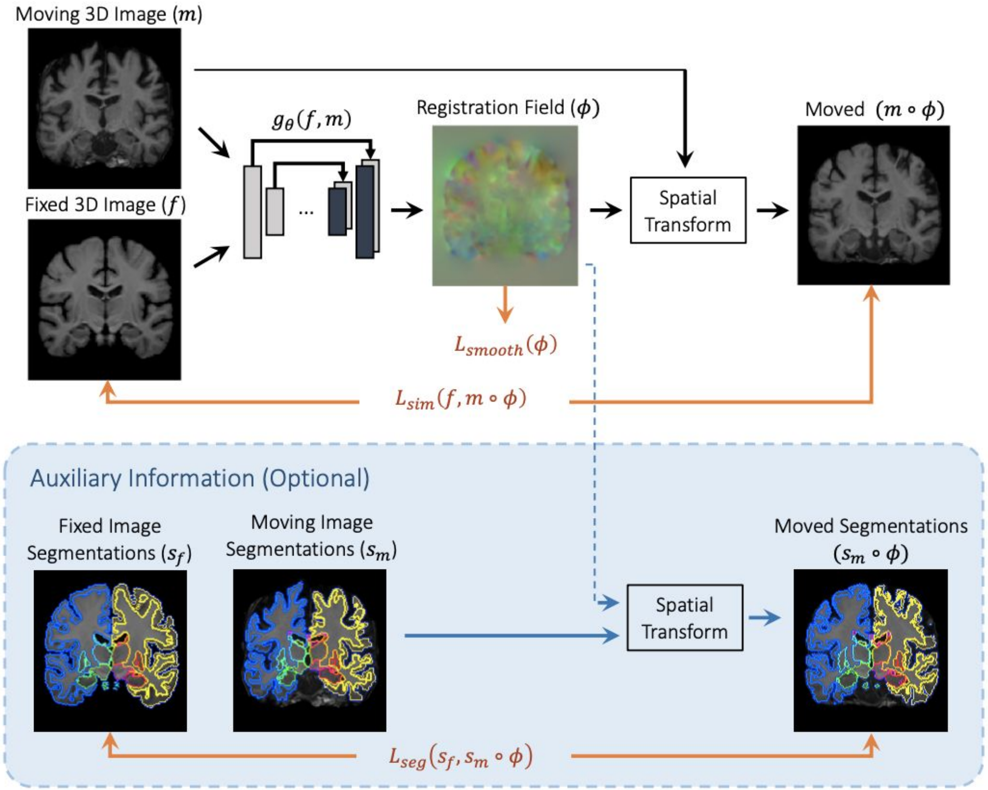
Kirillov et al. Segment Anything. arXiv 2023.

Today's agenda

- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

Medical image registration

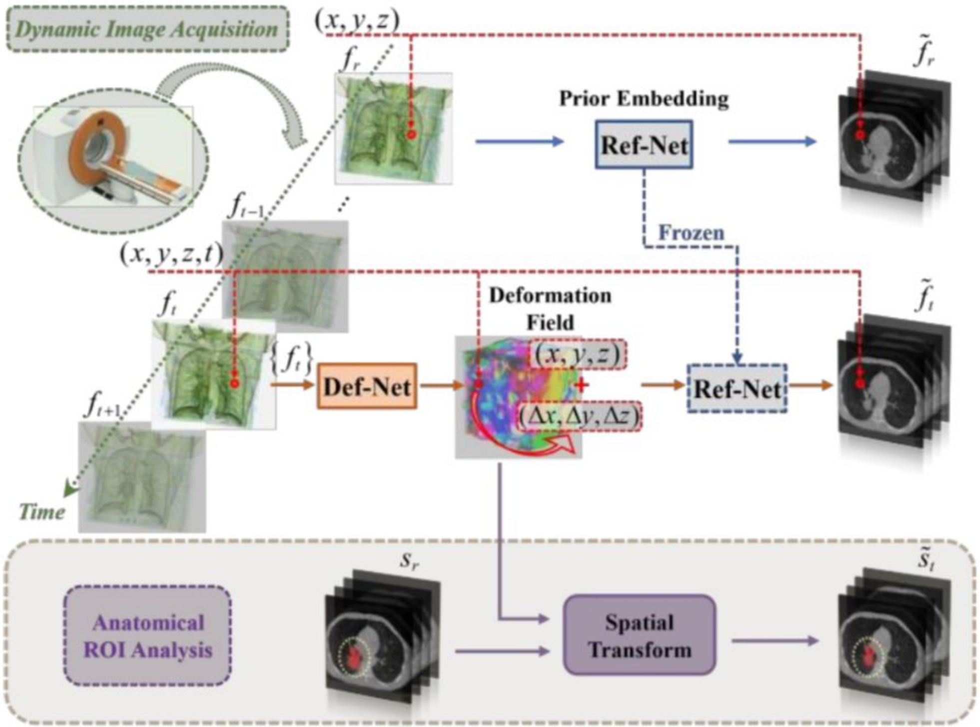
- VoxelMorph:
  - Formulate registration as a function that maps an input image pair to a deformation field that aligns these images
  - Parameterize the function via a convolutional neural network (CNN)



Balakrishnan et al. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. TMI 2018.

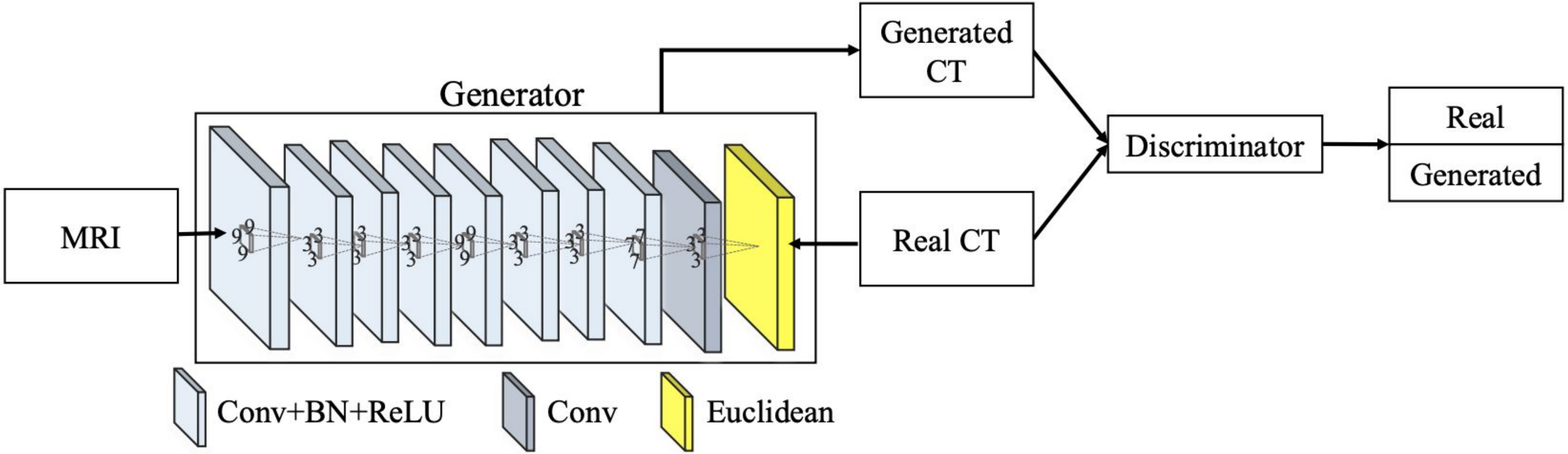
Medical image registration

- Implicit neural representation (INR) based deformable image registration



Medical image translation

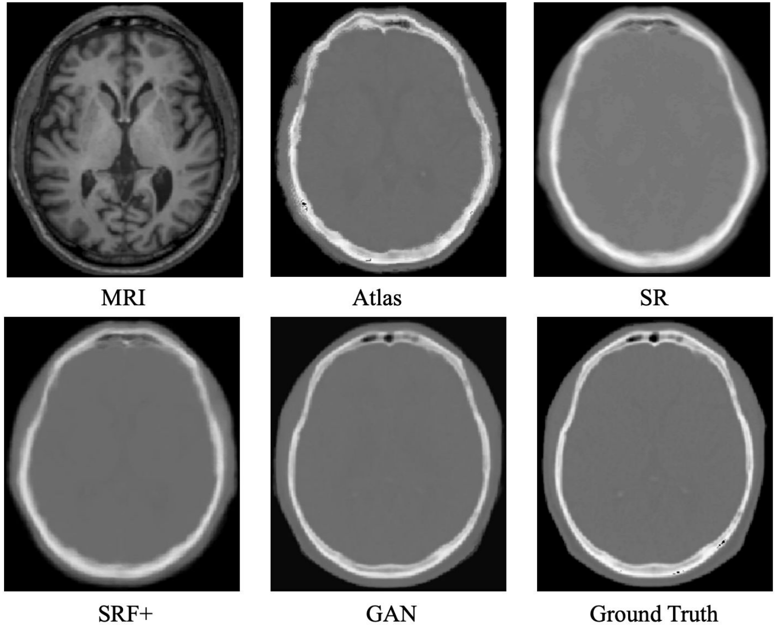
- GAN-based MRI-to-CT image translation

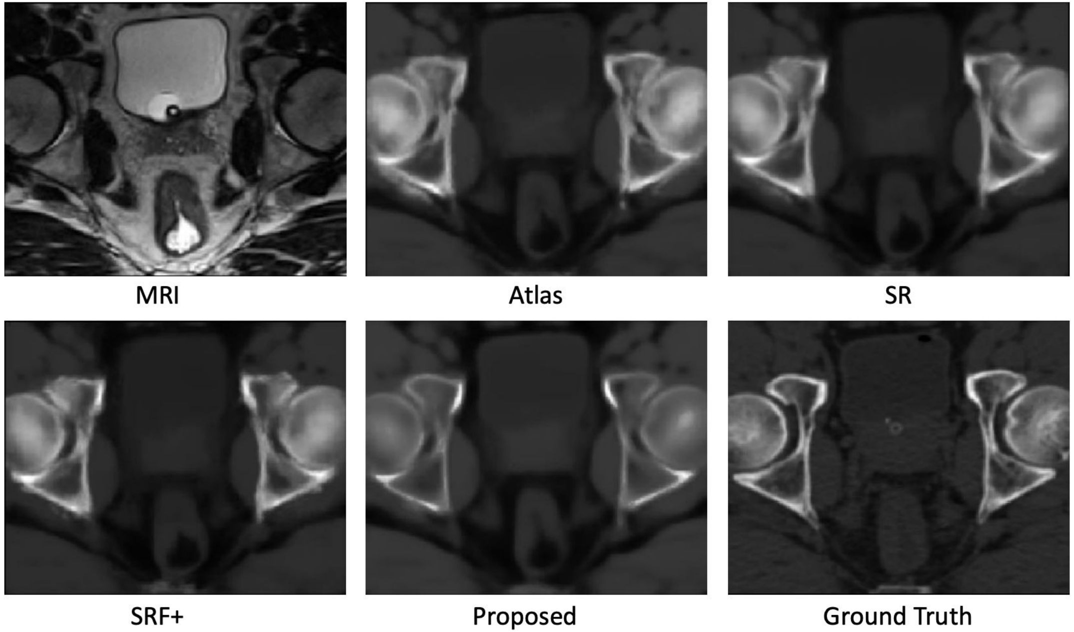


Nie et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. MICCAI 2017.

Medical image translation

- GAN-based MRI-to-CT image translation

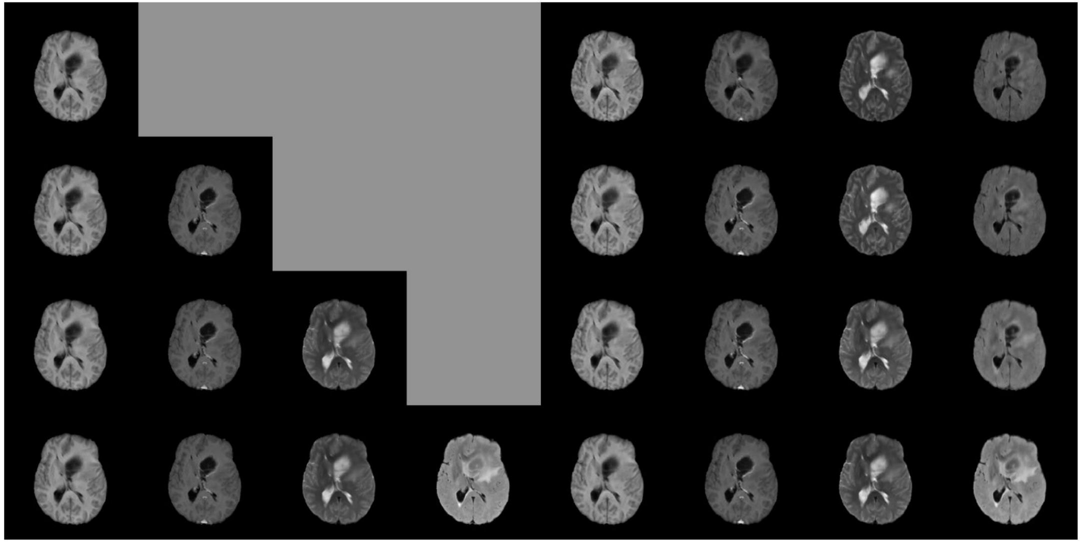


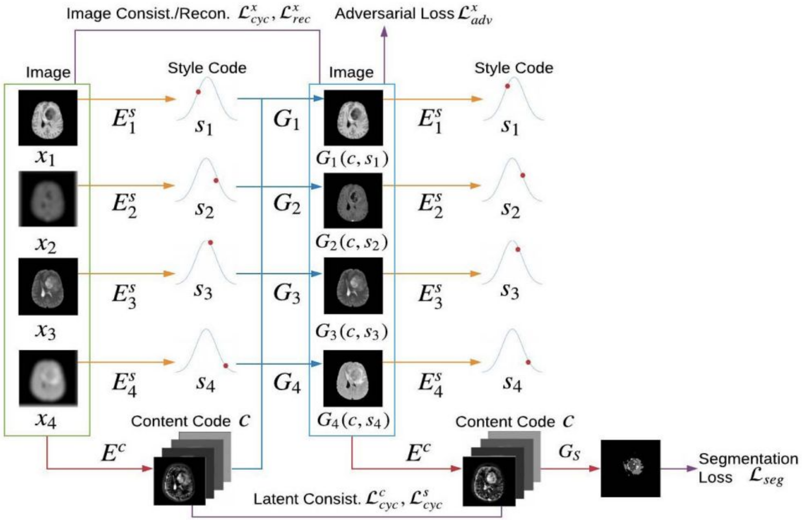


Nie et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. MICCAI 2017.

Medical image translation

- Multi-contrast MRI image translation with random missing modality

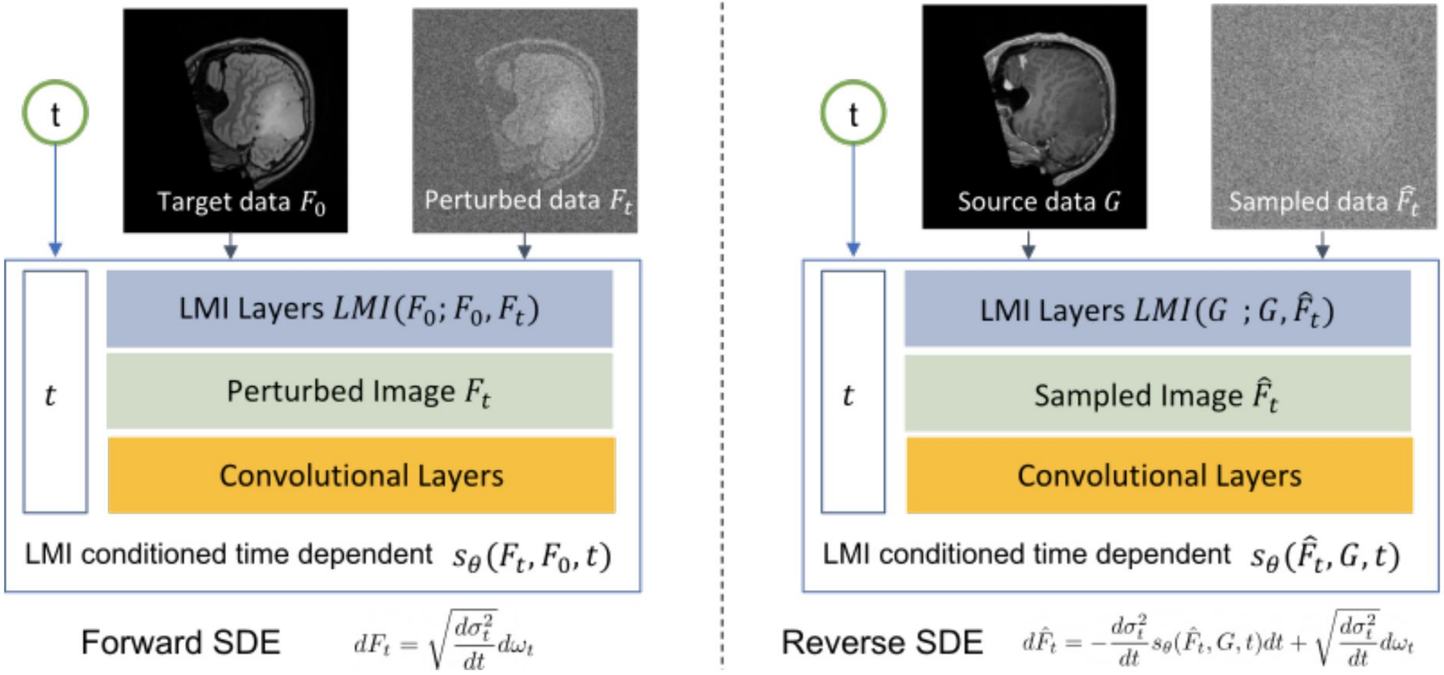




Shen et al. Multi-Domain Image Completion for Random Missing Input Data. TMI 2021.

Medical image translation

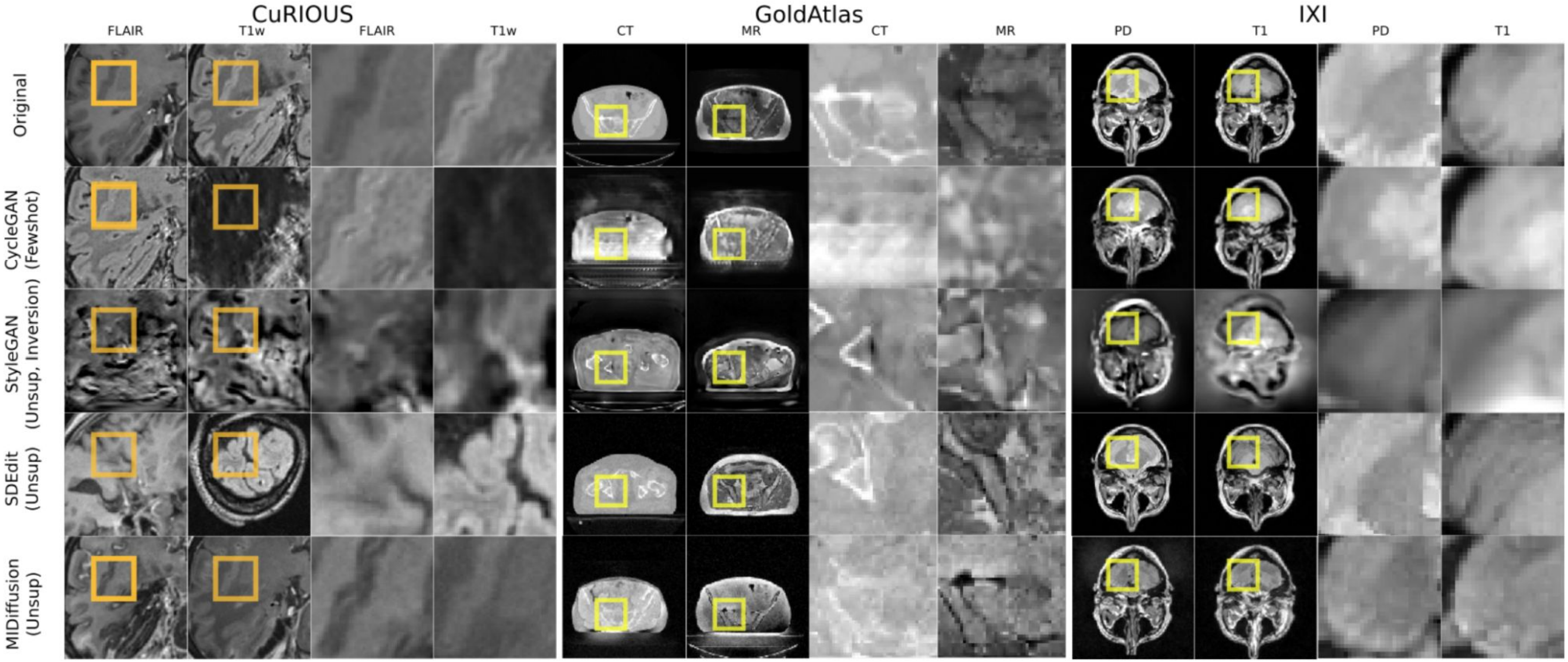
- Diffusion model-based medical image translation



Wang et al. Zero-shot-Learning Cross-Modality Data Translation Through Mutual Information Guided Stochastic Diffusion. arXiv 2023.

Medical image translation

- Diffusion model-based medical image translation



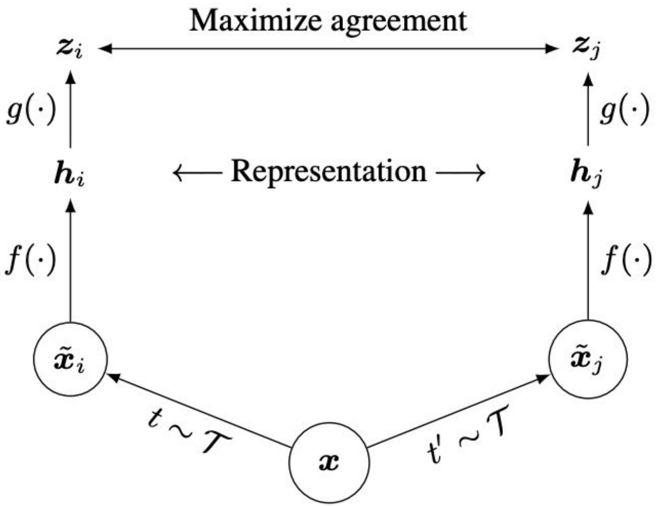
Today's agenda

- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

**Some slides in this section are adapted from: [Stanford BIODS220: Artificial Intelligence in Healthcare]*

Contrastive self-supervised learning

- SimCLR:
  - Different views of the same input should have more similar representation to each other than with a different input

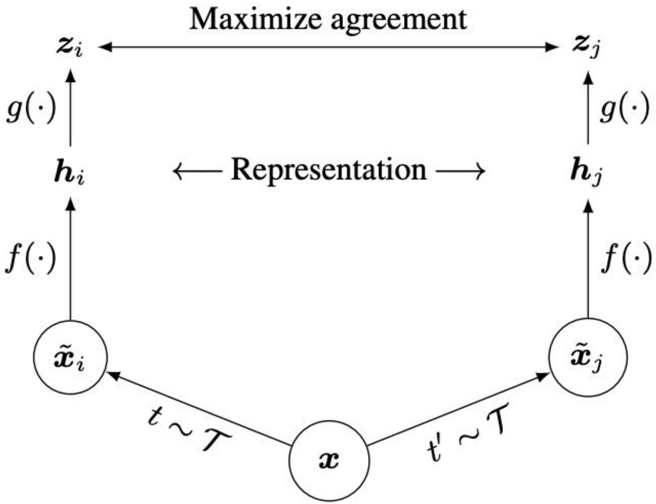


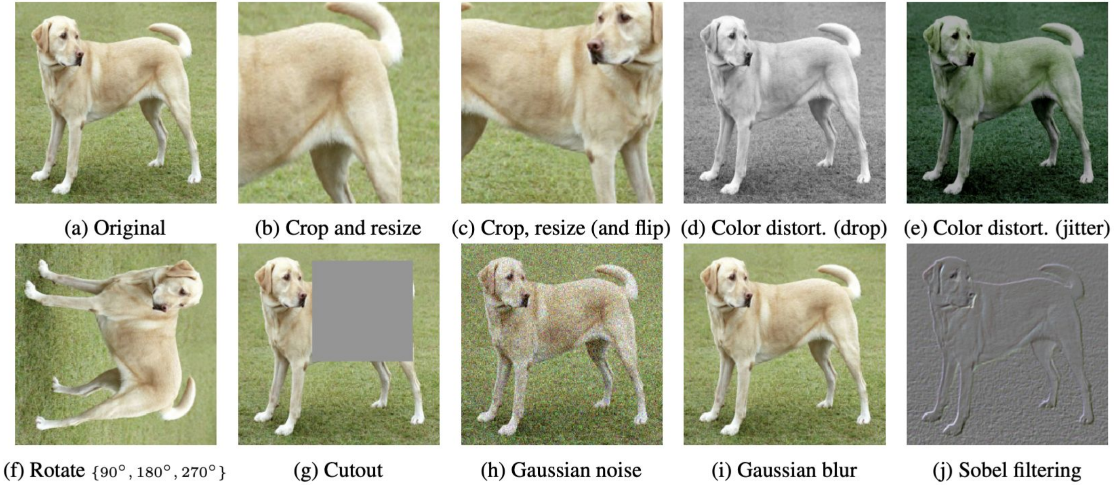
Chen et al. SimCLR - A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

Contrastive self-supervised learning

- Contrastive loss:

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

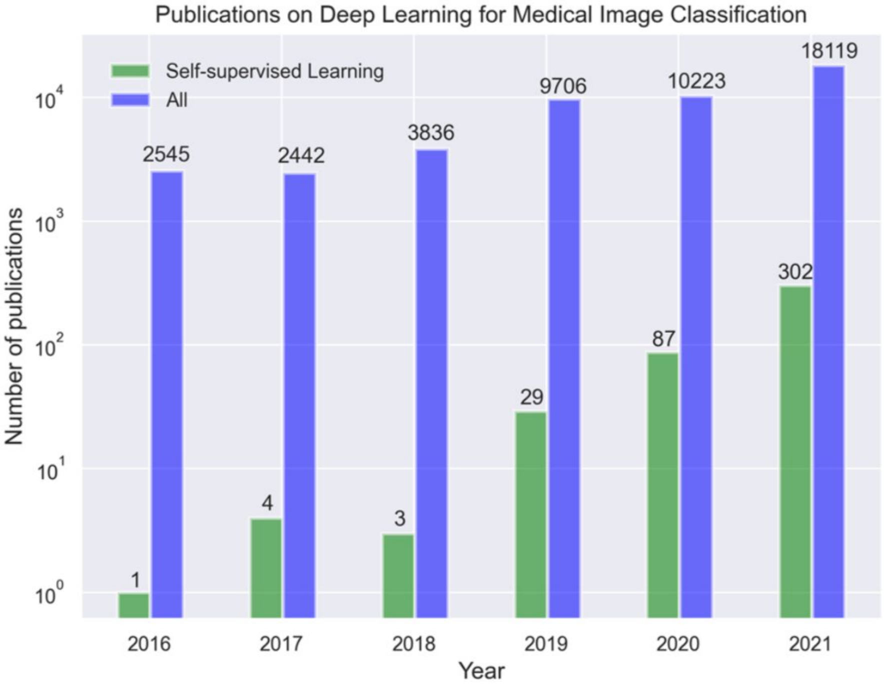




Chen et al. SimCLR - A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

Self-supervised learning for medical image analysis

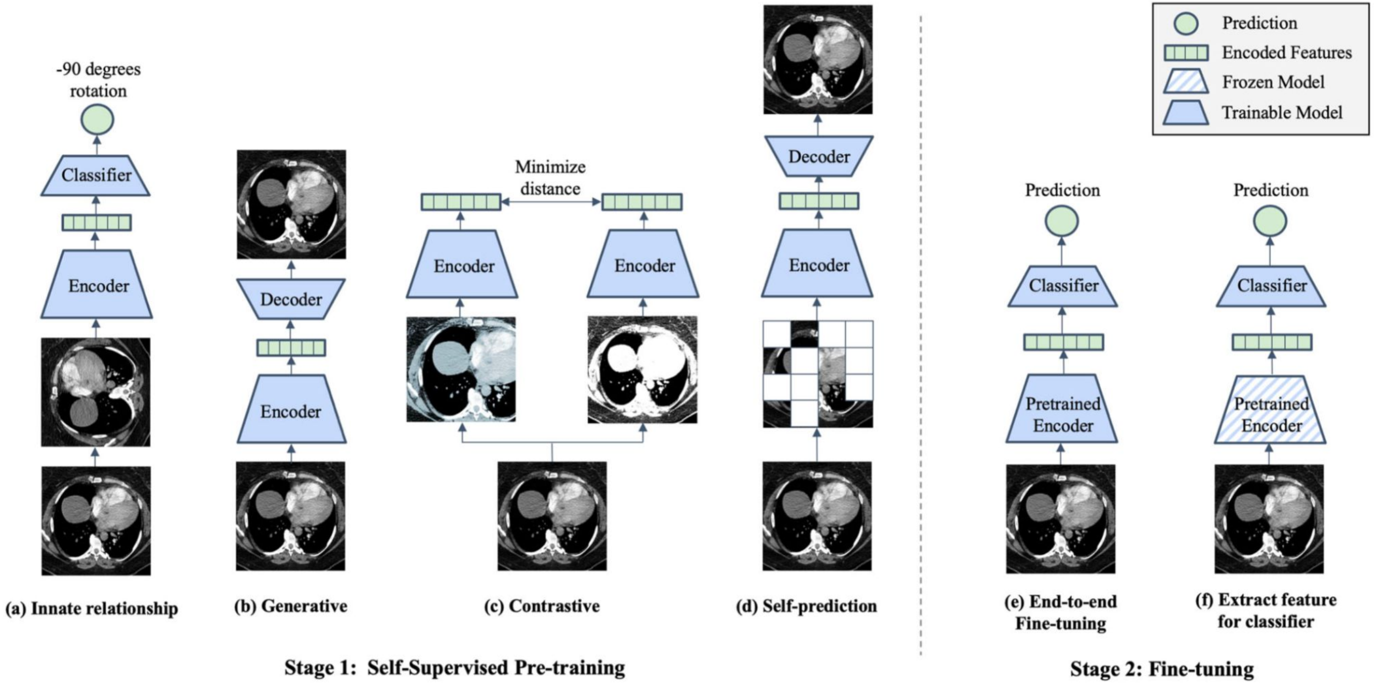
- Publications for 2016-2021:



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj digital med 2020.

Self-supervised learning for medical image analysis

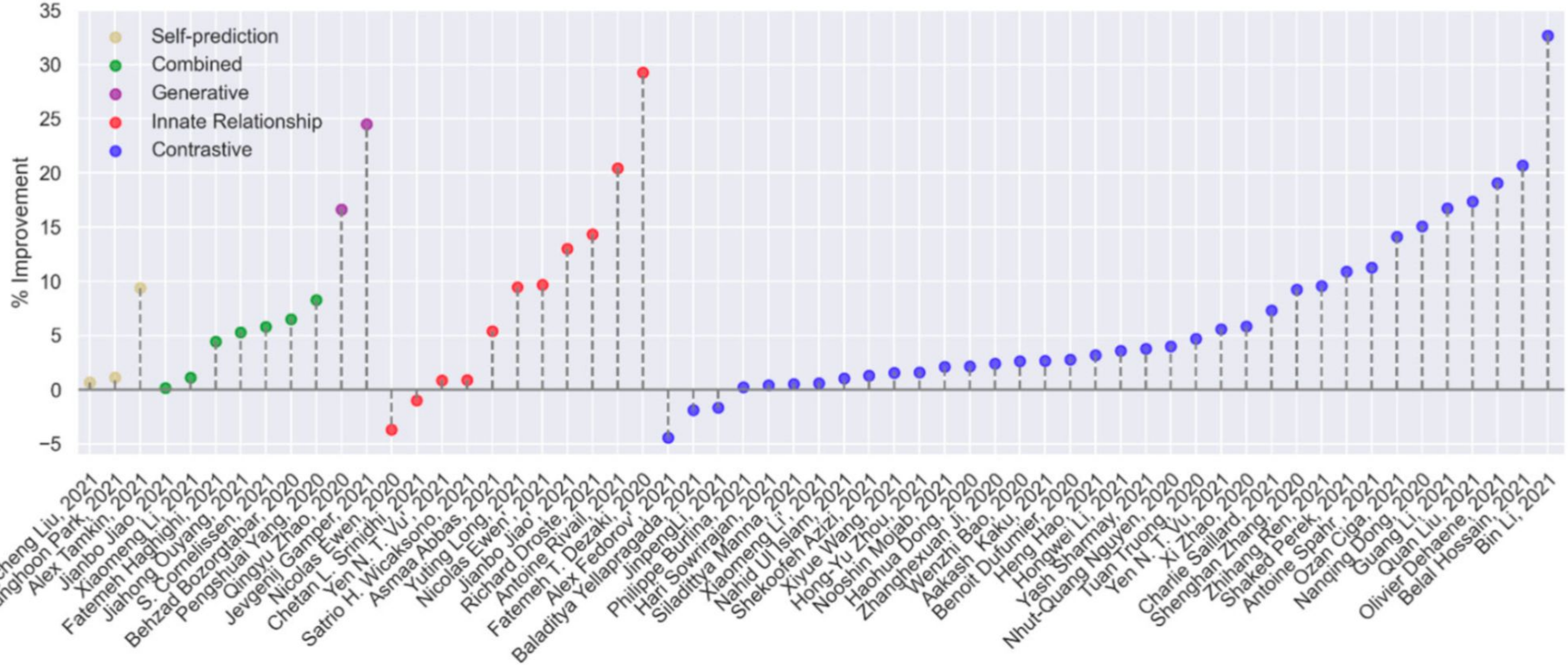
- Different self-supervised learning and fine-tuning strategies



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj digital med 2020.

Self-supervised learning for medical image analysis

- Relative difference in downstream task performance between self-supervised and non-self-supervised models

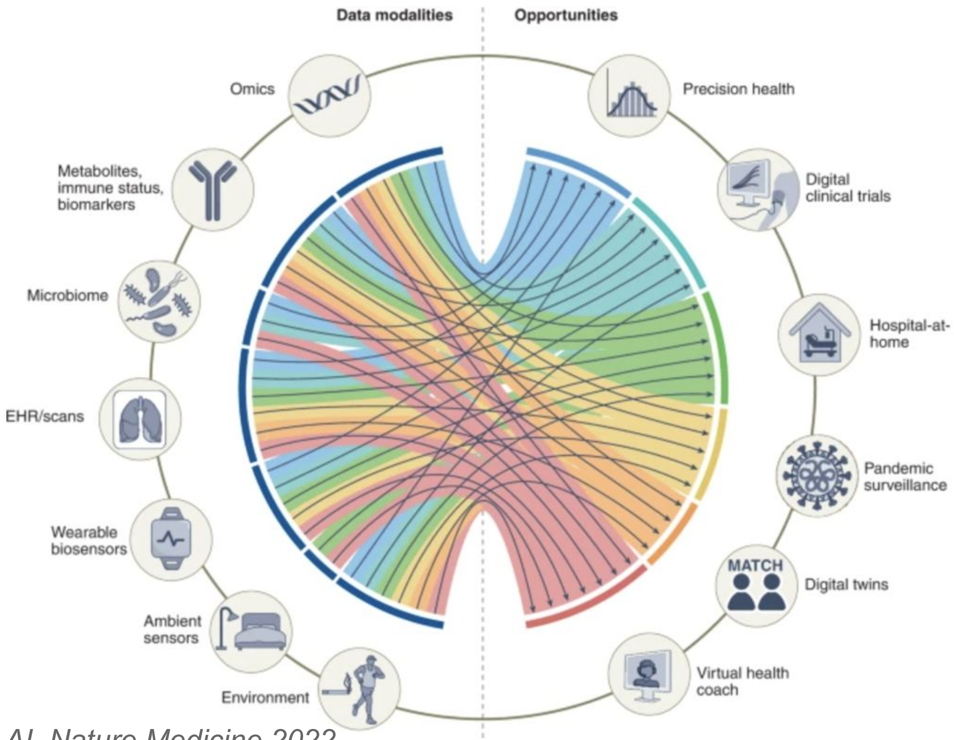


Today's agenda

- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

Multimodal learning

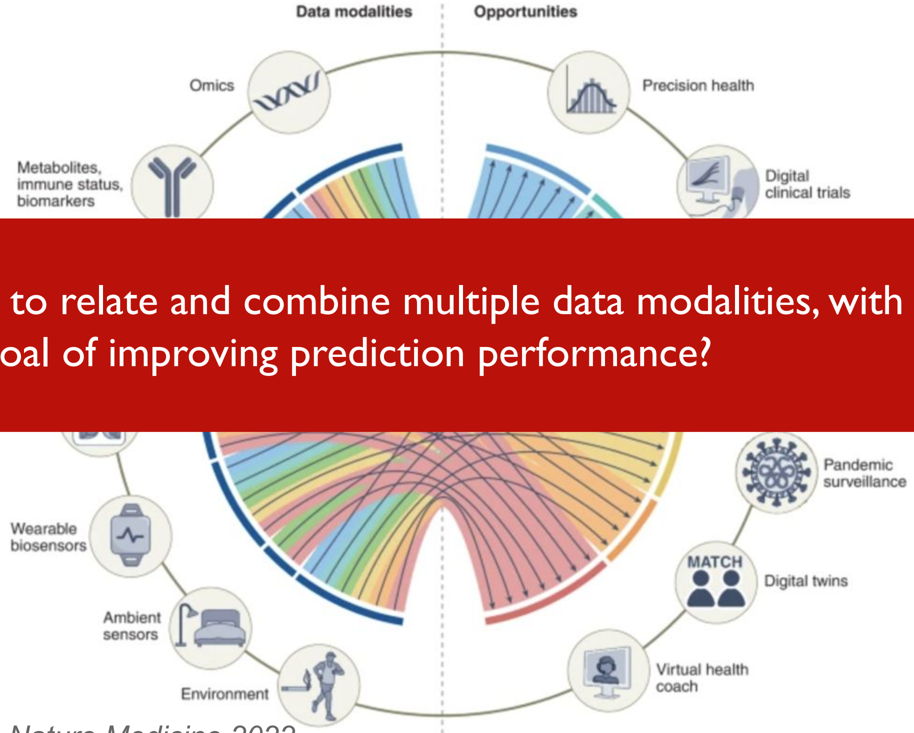
- Multimodal Biomedical AI can capture the complexity of human health and disease



Acosta, et al., Multimodal biomedical AI, Nature Medicine 2022.

Multimodal learning

- Multimodal Biomedical AI can capture the complexity of human health and disease

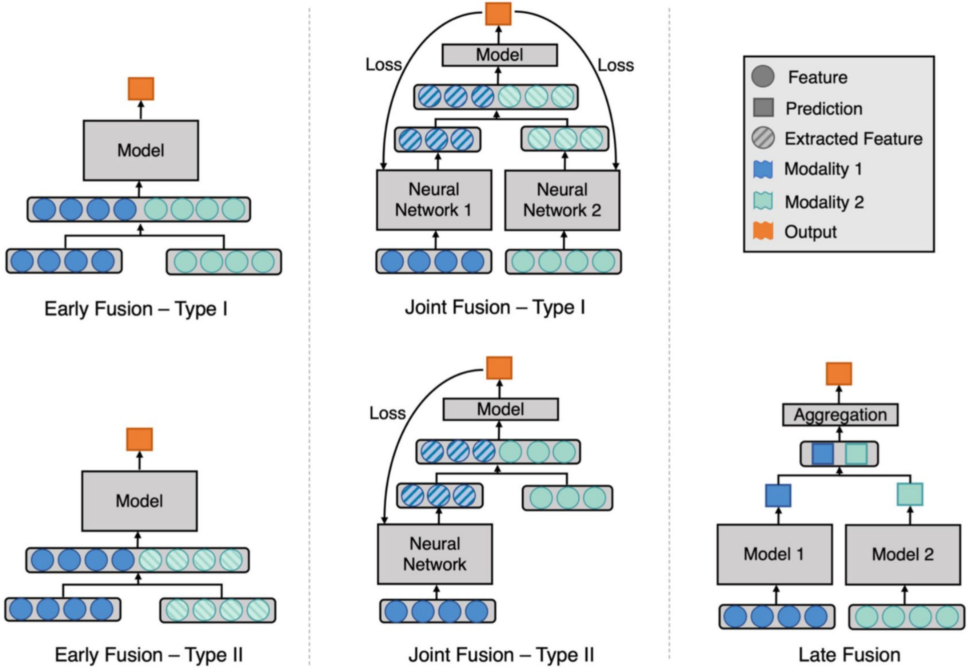


How to relate and combine multiple data modalities, with the goal of improving prediction performance?

Acosta, et al., Multimodal biomedical AI, Nature Medicine 2022.

Multimodal learning

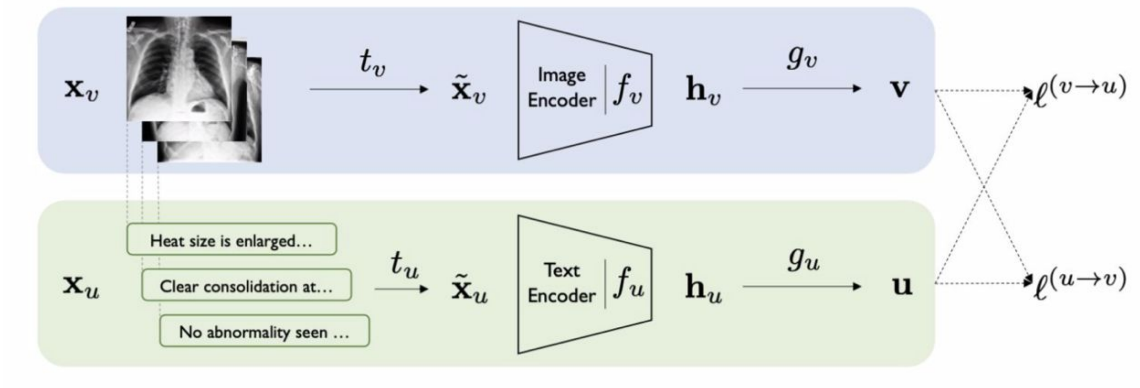
- Different multimodal fusion strategies



Huang et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. Npj digital med 2020.

Multimodal contrastive learning

- ConVIRT:
  - Multimodal contrastive pre-training with 217k paired X-ray image and radiology reports from MIMIC-CXR dataset

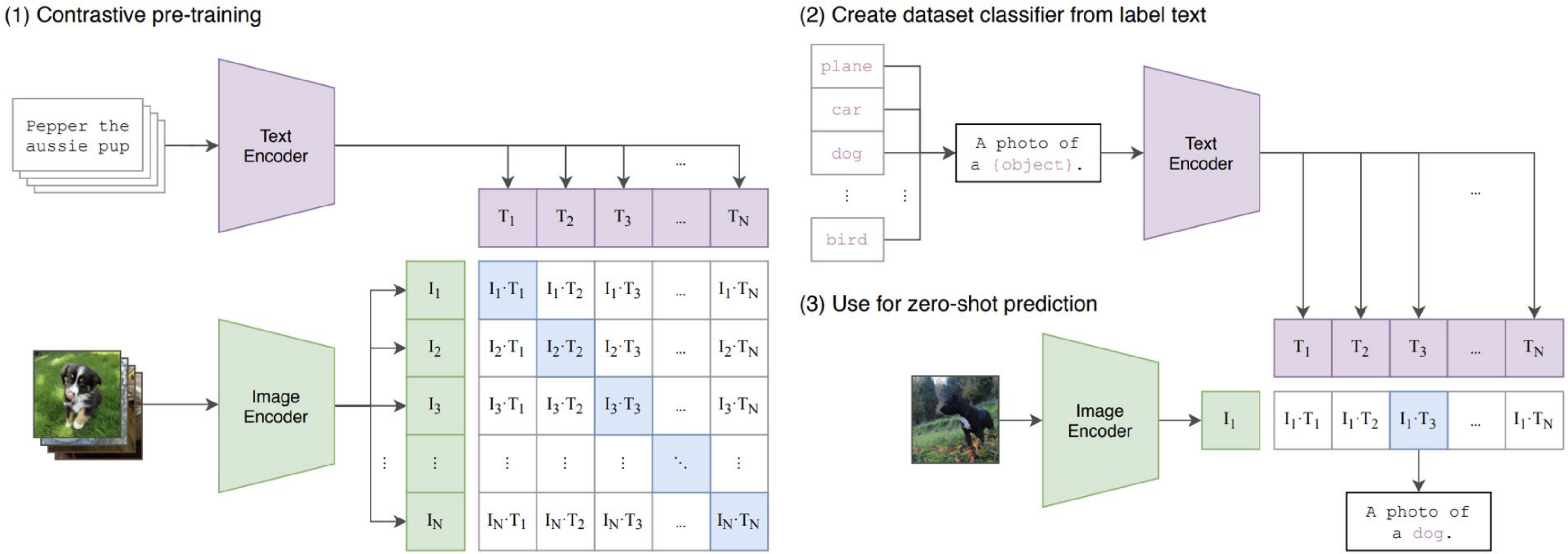


$$\ell_i^{(v\rightarrow u)} = -\log \frac{\exp(\langle \mathbf{v}_i, \mathbf{u}_i \rangle / \tau)}{\sum_{k=1}^{N} \exp(\langle \mathbf{v}_i, \mathbf{u}_k \rangle / \tau)} \qquad \langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^\top \mathbf{u} / \|\mathbf{v}\| \|\mathbf{u}\|$$

Zhang, et al., Contrastive learning of medical visual representations from paired images and text, arXiv 2020.

Multimodal contrastive learning

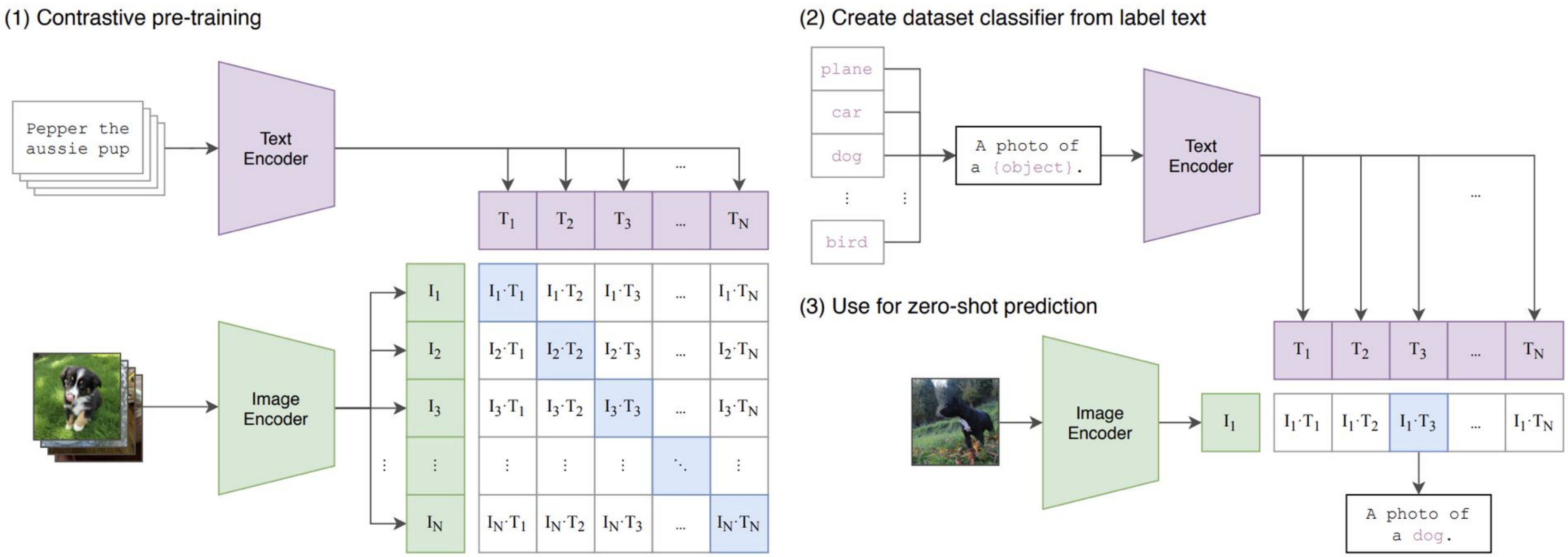
- CLIP:
    - Multimodal contrastive pre-training learns from images paired with raw text
    - Similar to ConVIRT, but now on very large dataset of 400 million image-text pairs



Radford, et al., Learning transferable visual models from natural language supervision, ICML 2021.

Multimodal contrastive learning

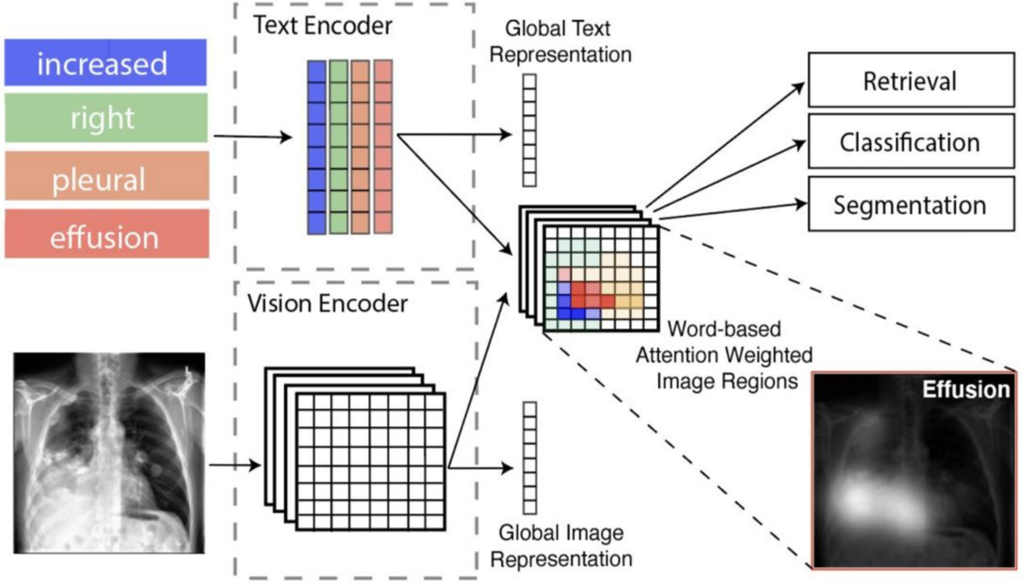
- Zero-shot classification:
  - Perform N-way classification without showing the model any paired examples of (input, class) for any of the N classes



Radford, et al., Learning transferable visual models from natural language supervision, ICML 2021.

Multimodal contrastive learning

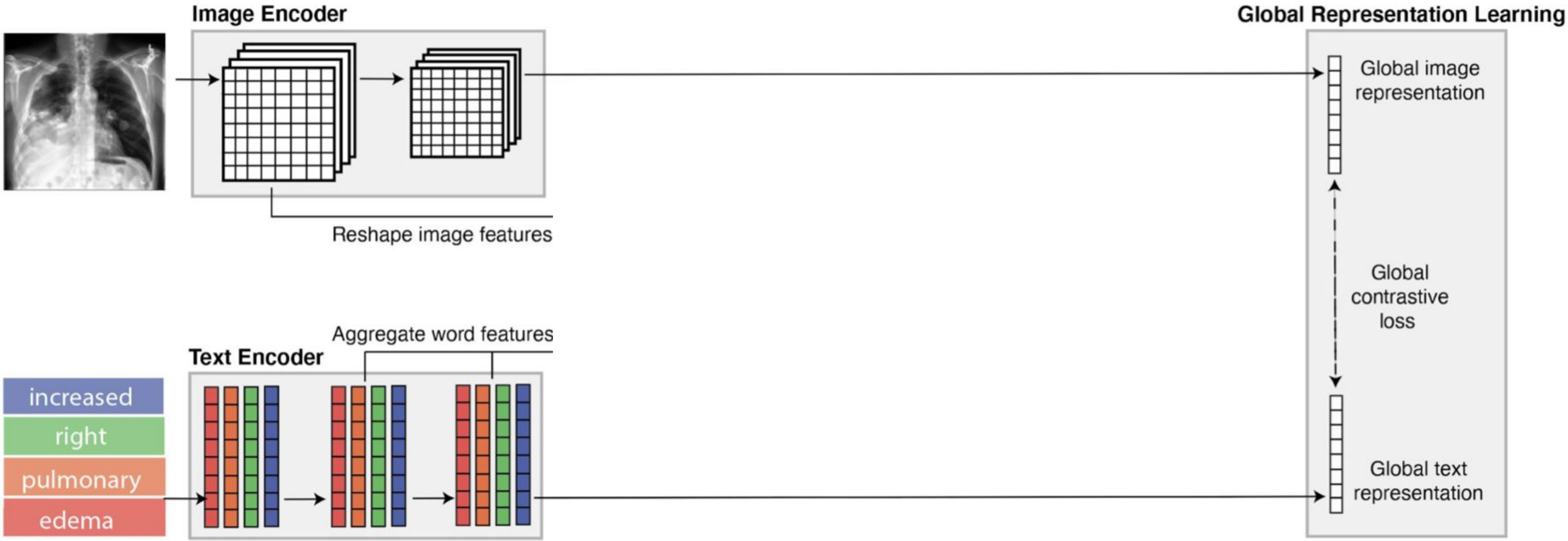
- GLoRIA:
  - Extension to ConVIRT: label-efficient representation learning model to learn global and localized features from medical images and radiology reports



Huang et al., GLoRIA: a multimodal global-local representation learning framework for label-efficient medical image recognition, ICCV 2021.

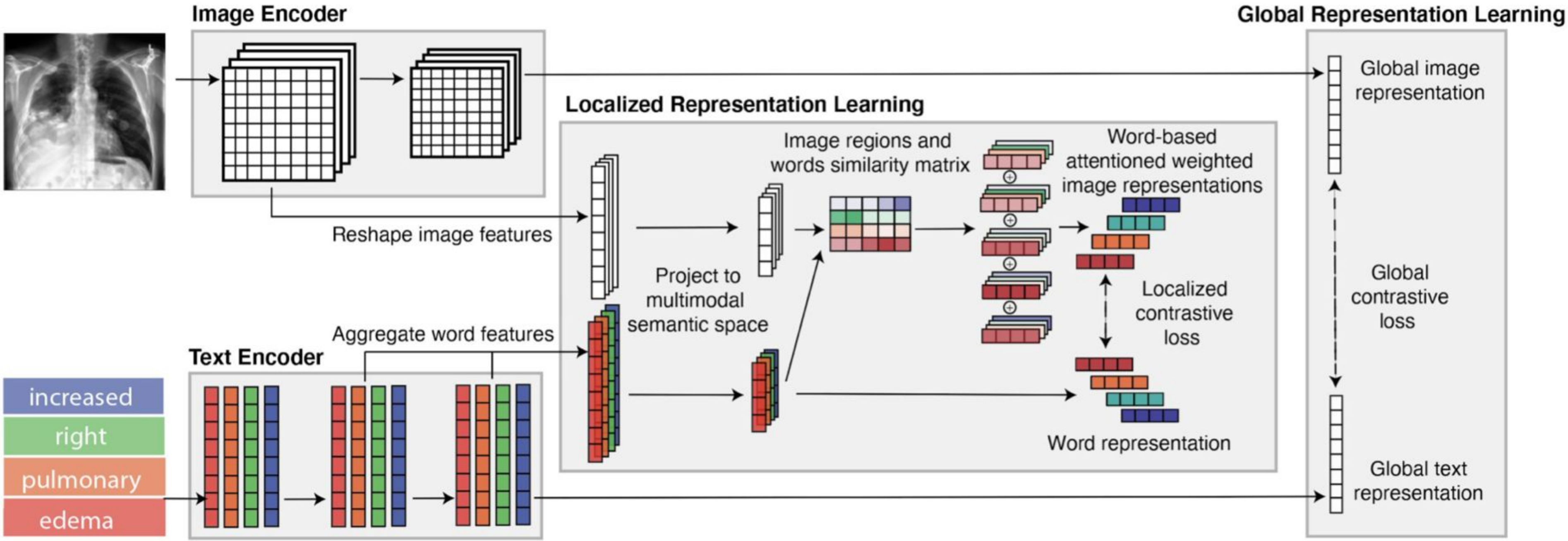
Multimodal contrastive learning

- GLoRIA:
  - Global representation learning with global contrastive loss



Multimodal contrastive learning

- GLoRIA:
  - Jointly train with a localized contrastive loss between words and attention-weighted regions of images

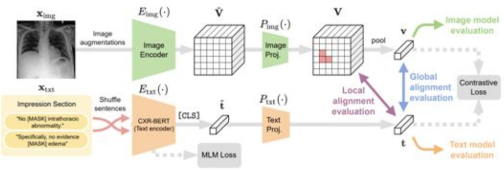


Medical multimodal learning

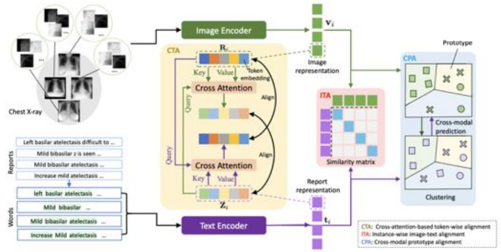
- Following up GLoRIA:

Cross-modal local alignment

Boecking et al. (ECCV 2022)

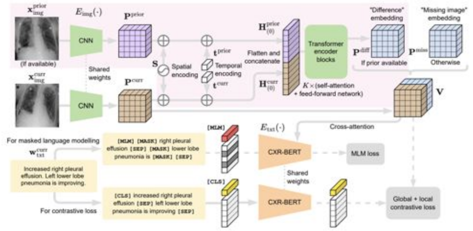


Wang et al. (NeurIPS 2022)

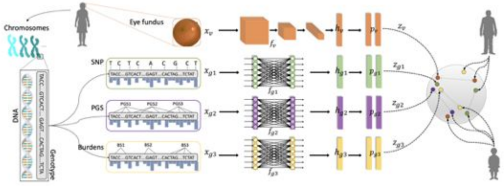


Extend to other data modality

Bannur et al. (arXiv 2023)

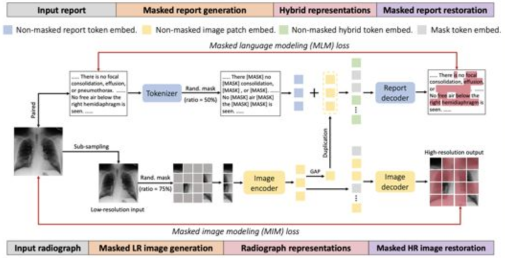


Taleb et al. (CVPR 2022)



Pre-training loss function design

Zhou et al. (ICLR 2023)



Today's agenda

- Challenges
- Medical image classification
- Medical image segmentation
- Other applications
- Self-supervised learning
- Multimodal learning

Next lecture

- Multimodal foundation models

Date	Lecture #	Topic	Papers	Instructor / Presenter
Tue 8/29	1	Introduction and course overview		Liyue Shen
Thu 8/31	2	Biomedical imaging with deep learning **[Fundamental]**		Liyue Shen
Tue 9/5	3	Implicit neural representation learning **[Advanced]**		Liyue Shen
Thu 9/7	4	Generative diffusion models **[Advanced]**		Liyue Shen
Tue 9/12	5	Medical image analysis **[Fundamental]**		Liyue Shen
Thu 9/14	6	Multimodal foundation models **[Advanced]**		Liyue Shen
Mon 9/18		Drop/add deadline for full term classes		
Tue 9/19	7	Implicit neural representation learning		
Thu 9/21	8	Implicit neural representation learning		
Tue 9/26	9	Implicit neural representation learning		
Thu 9/28	10	Implicit neural representation learning		
Tue 10/3	11	Generative diffusion models		
Thu 10/5	12	Generative diffusion models		
Tue 10/10	13	Generative diffusion models		
Thu 10/12	14	Generative diffusion models		
Tue 10/17		No class (fall study break)		
Thu 10/19	15	Self-supervised learning		
Tue 10/24	16	Self-supervised learning		
Thu 10/26	17	Multimodal learning		
Tue 10/31	18	Multimodal learning		
Thu 11/2	19	Transformer and LLM		
Tue 11/7	20	Transformer and LLM		