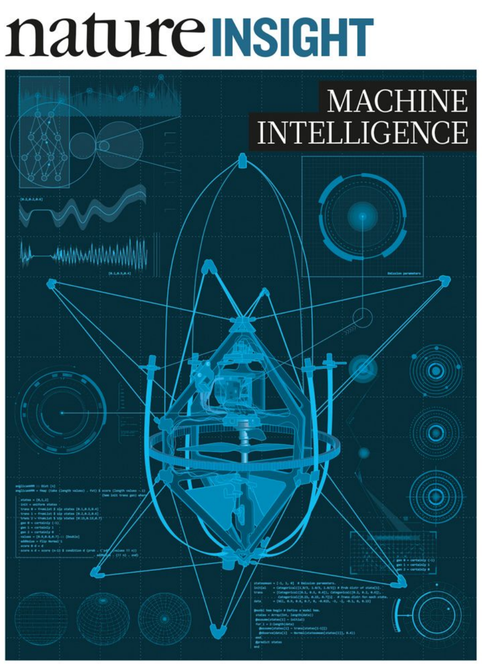
Medical Imaging and Deep Learning General Overview

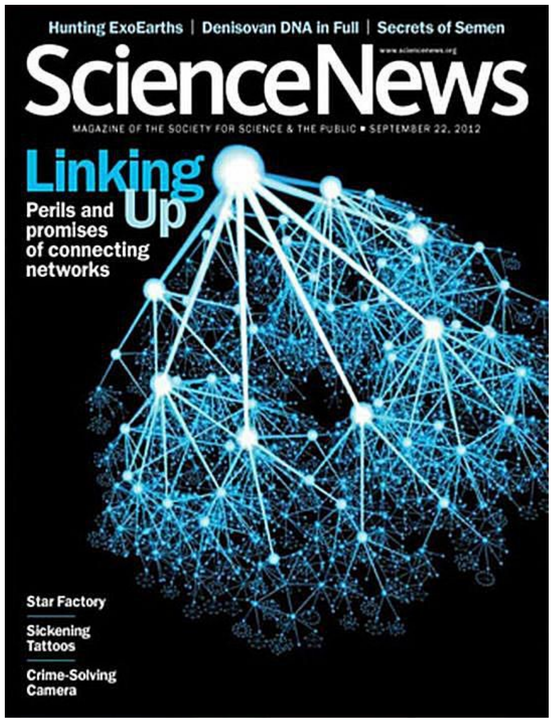
Maria Vakalopoulou
Mathematics and Informatics (MICS)
CentraleSupelec, University Paris-Saclay

Motivation

- **Artificial Intelligence for Health Care**

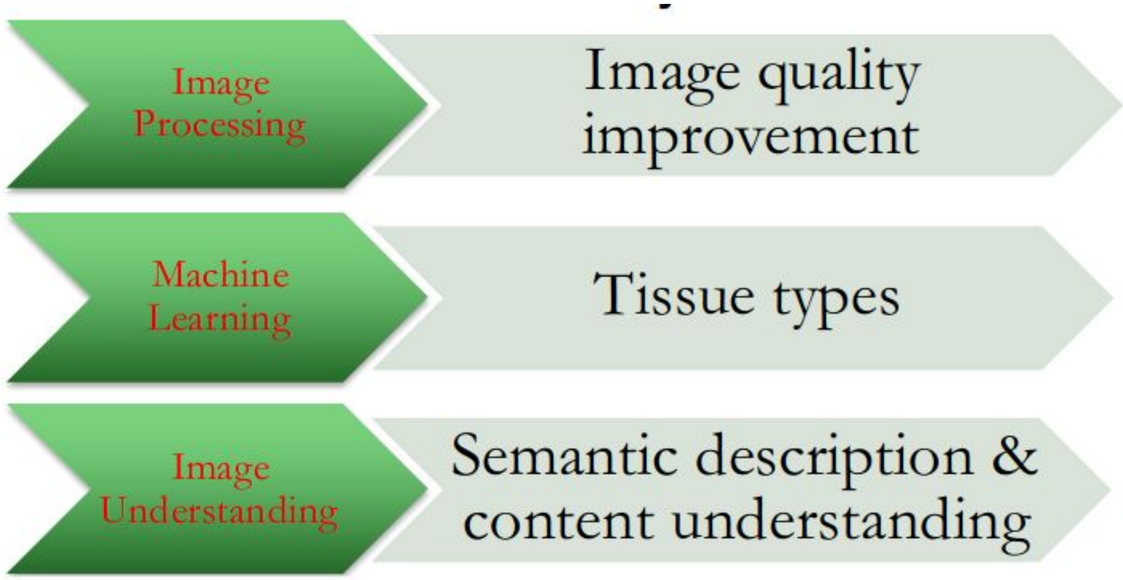
Geoff Hinton: “Medical schools should stop training radiologists now” (AI vs MD, New Yorker, April 2017)





Medical Imaging

- The most direct way to see inside the human (or animal) body is cut it open (i.e. surgery)
- We can see inside the human body in ways that are less invasive of (completely non-invasive)
- We can see metabolic/functional/molecular activities which are not visible to naked eye

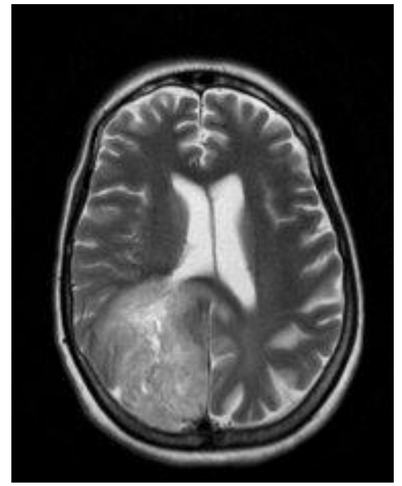


Medical Imaging

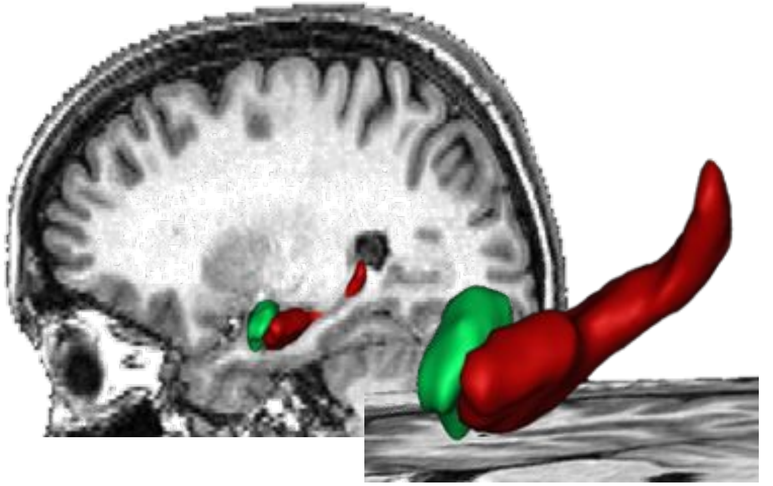
- **Medical Care (Clinical Practice)**
  - Detection of lesions/ anomalies
  - Screening
  - Quantification (extraction of biomarkers)
  - Follow up pathology
- **Research (Medical Image Computing)**
  - Automatize the clinical process
  - Prognosis e.g. evaluation of a specific treatment
  - Understand the alternation between specific diseases

65-year old female

Presented with
left-side weakness and
headache



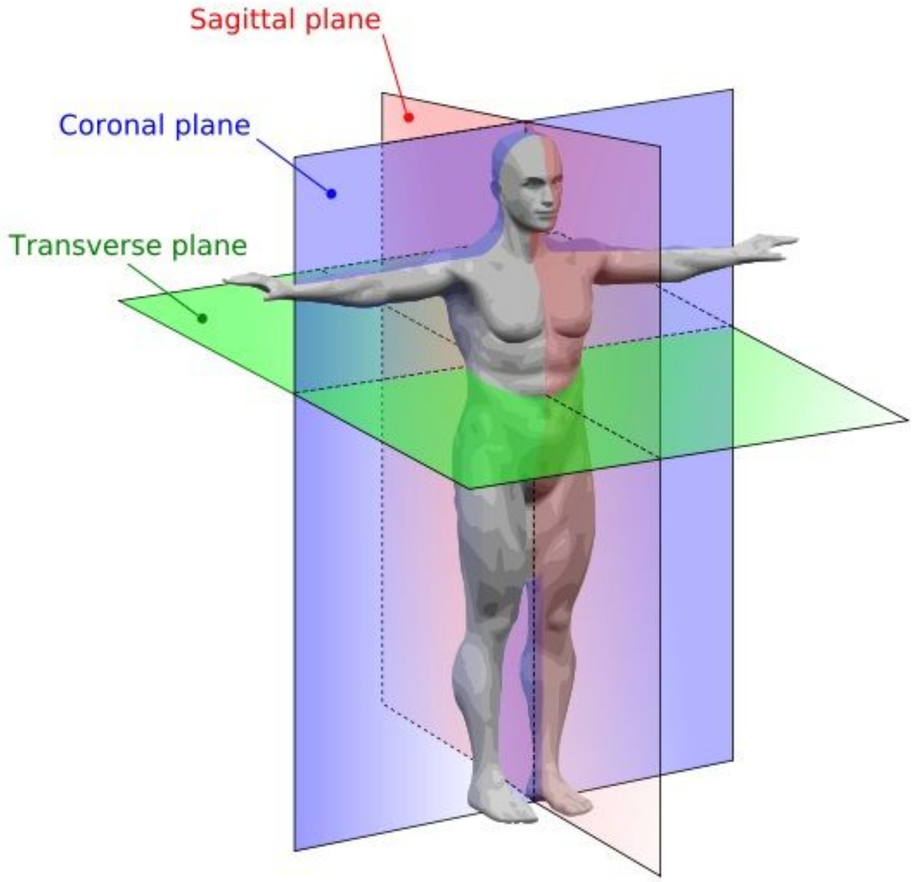
Brain tumor and Glioblastoma



Medical Imaging

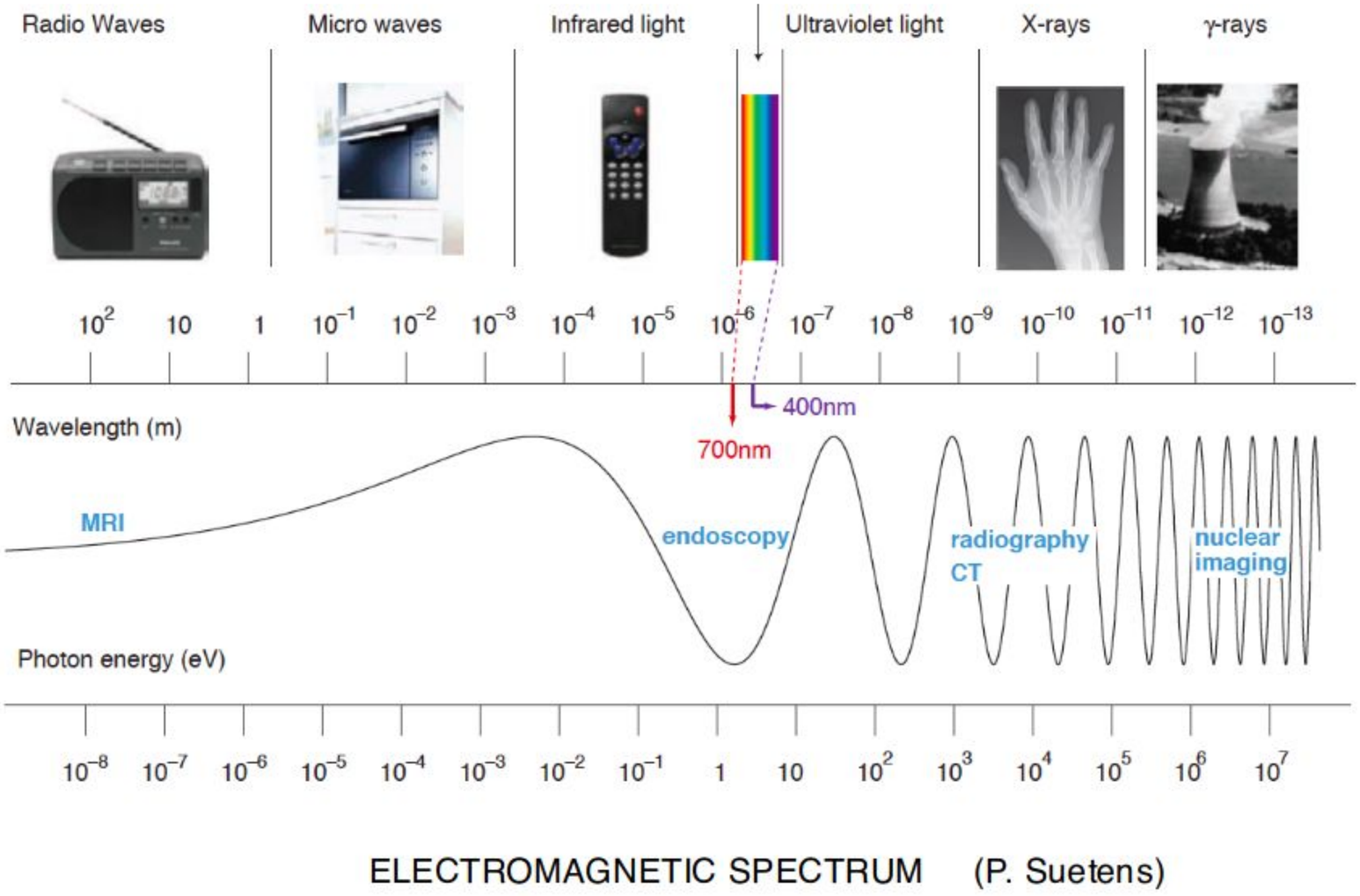
Non-invasive visualization of internal organs, tissues, etc:

- **Representations**
  - 2D signal f(x,y)
  - 3D signal f(x,y,z)
- **Major Modalities**
  - Projection X-ray (Radiography)
  - X-ray Computed Tomography (CT)
  - Nuclear Medicine (PET, …)
  - Ultrasound
  - Magnetic Resonance Imaging (MRI)
  - Mammography



Medical Imaging

Main Modalities



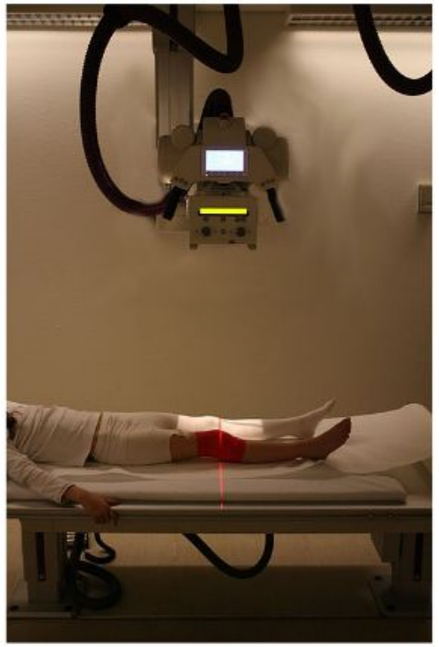
ELECTROMAGNETIC SPECTRUM (P. Suetens)

Medical Image Analysis

- Because of the rapid technical advances in medical imaging technology and the introduction of new clinical application, the field has become a **highly active research field**.
- Improvements in **image quality**, changing **clinical requirements**, advances in **computer hardware** and **algorithmic progress** (deep learning) in processing all have a direct impact on the state of the art in medical image analysis
- Medical images are often **multidimensional** (2D,3D,4D,nD) have a large dynamic range, are produced on different imaging modalities.
  - A high resolution MR image of the brain, for instance, may consist of more than 300 slices of 512x512 voxels each
  - The data, even a single image in the case of histopathology, can easily exceed some GBs

Main Modalities

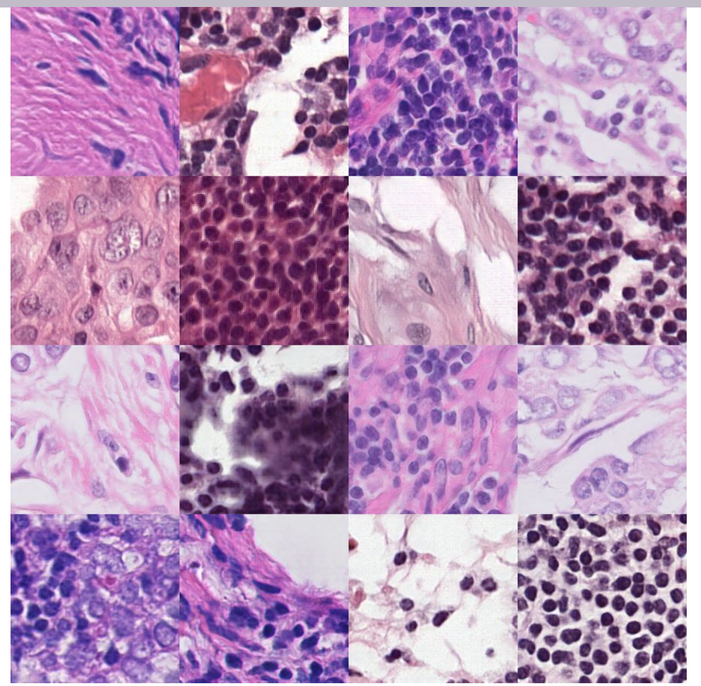
X-ray radiography

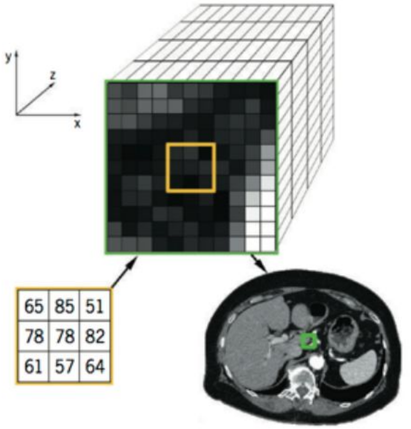


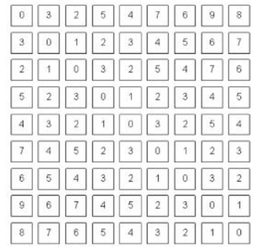
MRI

PET / CT

Ultrasound (echography)







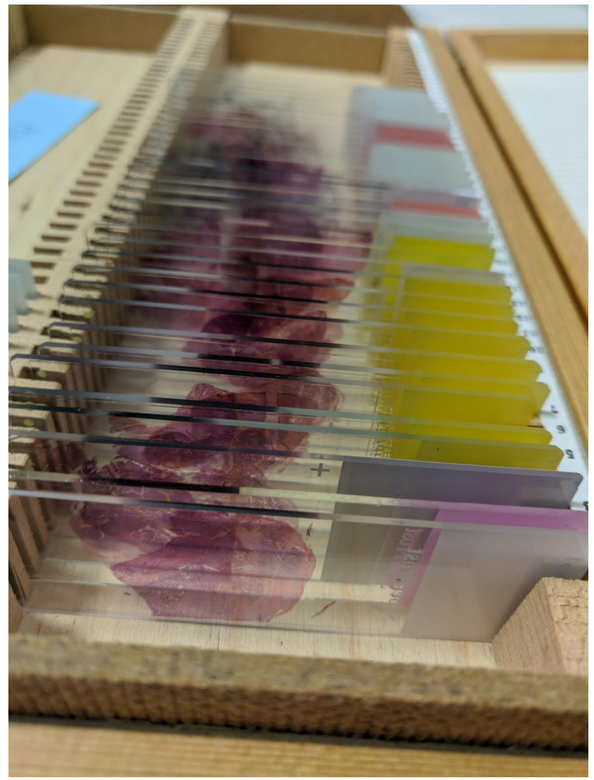
What computer sees!

Comparison of Imaging Methods

	Chest	Abdomen	Head/Neck	Cardiovascular	Skeletal/muscular
CT	gold standard	Need contrast for excellency, widely used	Good for trauma	Gold standard	Gold standard
US	no use except heart or P.Effusion	Problems with gas	Poor	Poor	Elastography
Nuclear	Extensive use in heart and therapy in lung	CT or MRI is merged	PET	Perfusion	bone marrow
MRI	growing cardiac applications	Increased role of MRI	Gold standard	Will replace ct in near future	Excellent

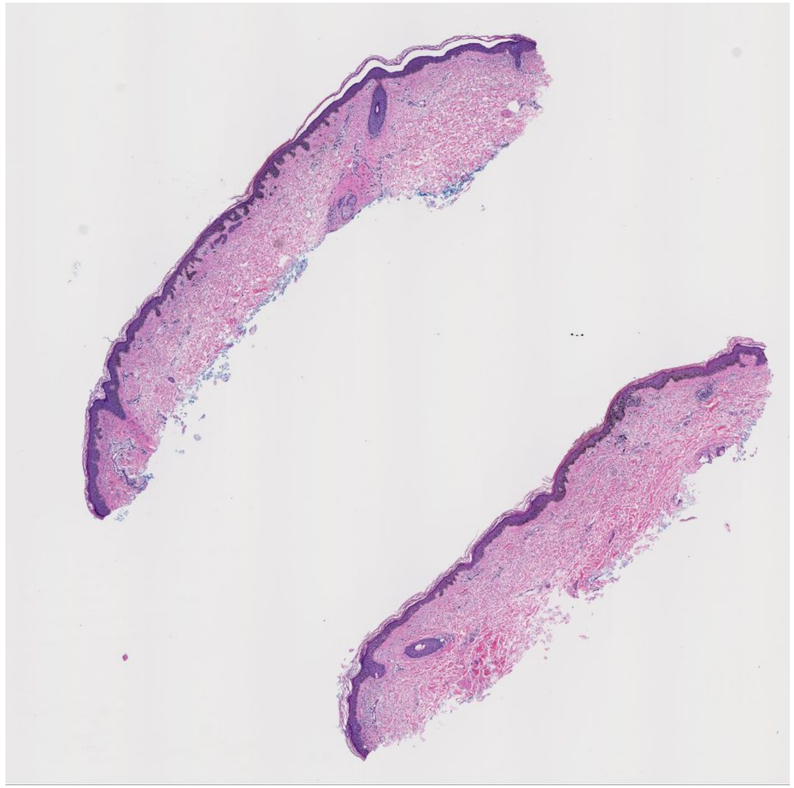
Digital Pathology (focus on Histopathology)

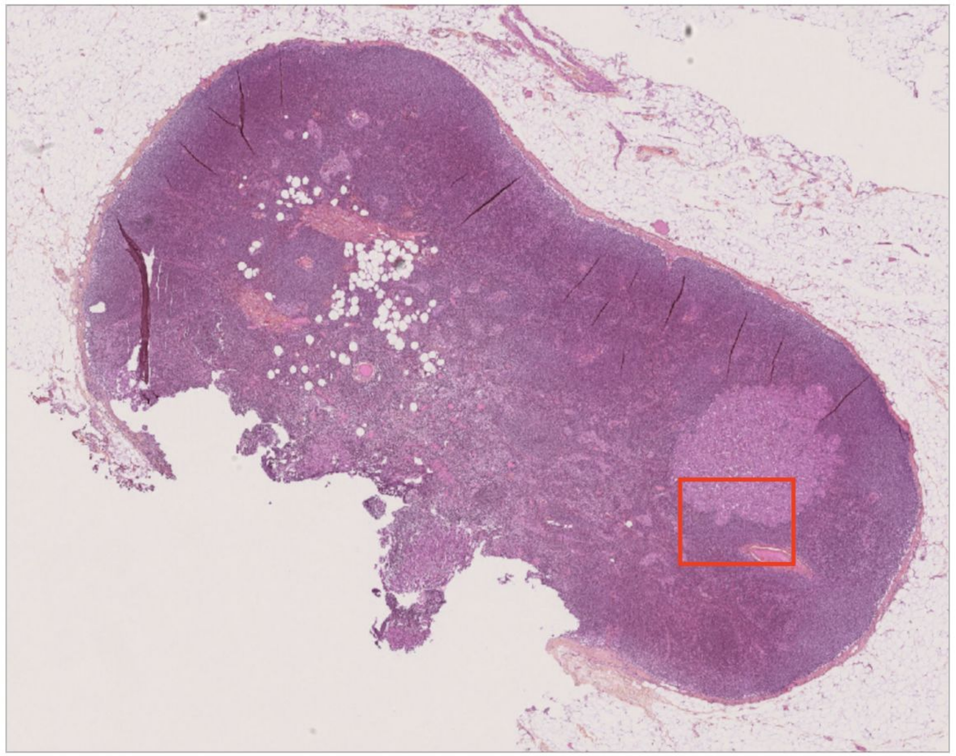
- Digital slides are created from glass slides using specialized scanning machines
- An integrated camera and a motorized stage to move the slide around while parts of the tissue are imaged.
- Two different types of scanning tile-based scanning and line-based scanning
  - Tile scanners capture square field-of-view images covering the entire tissue area on the slide
  - Line-scanners capture images of the tissue in long, uninterrupted stripes rather than tiles
- Software associated with the scanner stitch the tiles or lines together into a single, seamless image.

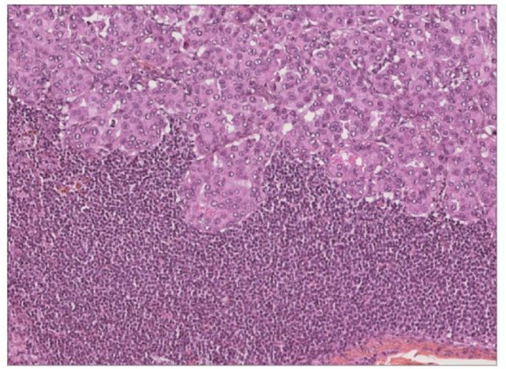


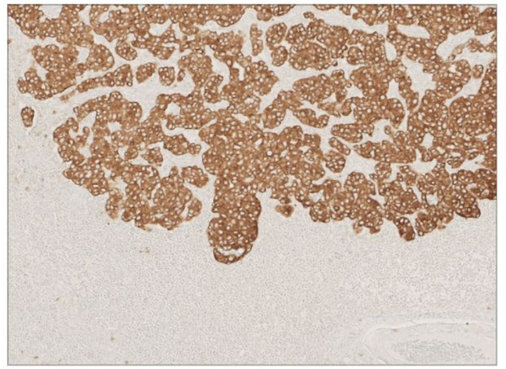
Digital Pathology (focus on Histopathology)

- Gigabytes of size images
- Millions of cells
- Cells are captured in their (usually tumor) microenvironment
- Different types of staining using different biomarkers









Commonly used software

Practical Info for processing medical imaging

- Software (with GUI) you can use for analysis of medical imaging
  - ITKSnap, Slicer, MTIK, ImageJ/Fiji, Mango,....

- Coding (self): ITK library is usually used
  - C/C++ and Python can be used to call libraries
  - SimpleITK with python is very commonly used

- Image Format
  - DICOM
  - Analyse (.img/hdr)
  - Nifti
  - ….

ITK Library

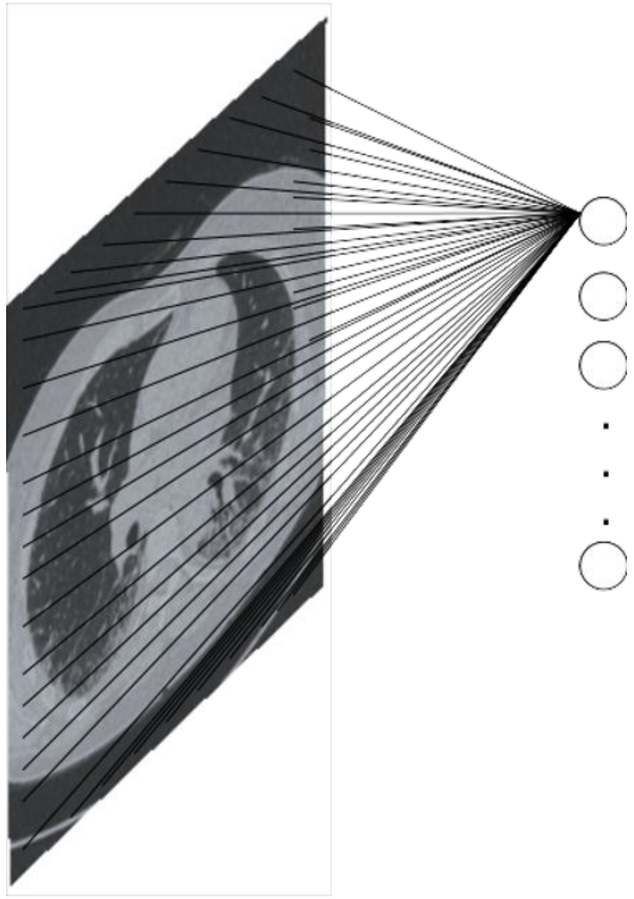
National Library of Medicine (of NIH) Insight Segmentation and Registration Toolkit (ITK)

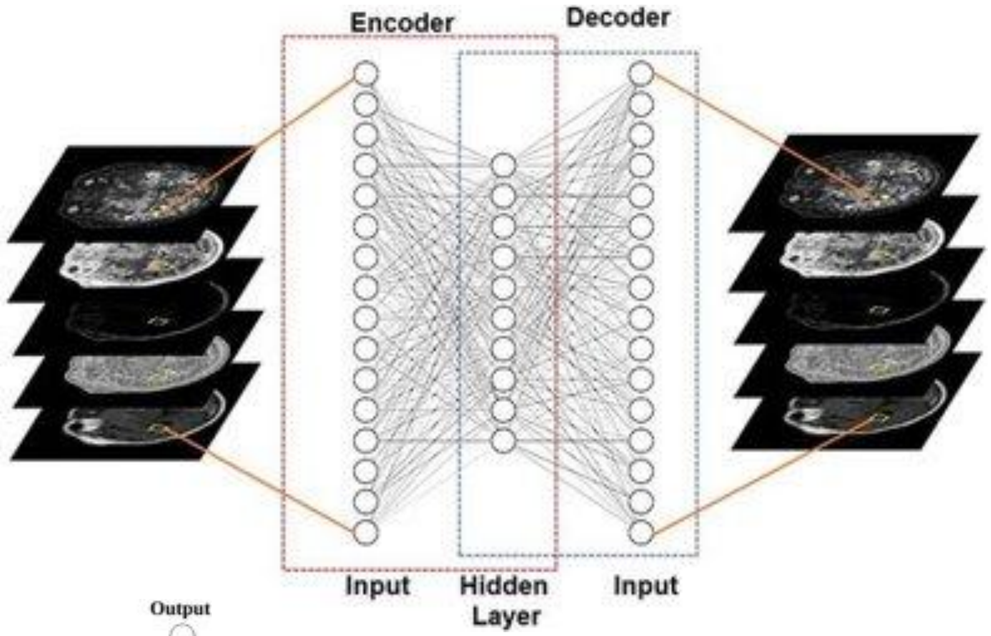
- ITK is an open-source, cross platform system that provides developers with an extensive suite of software tools for image analysis

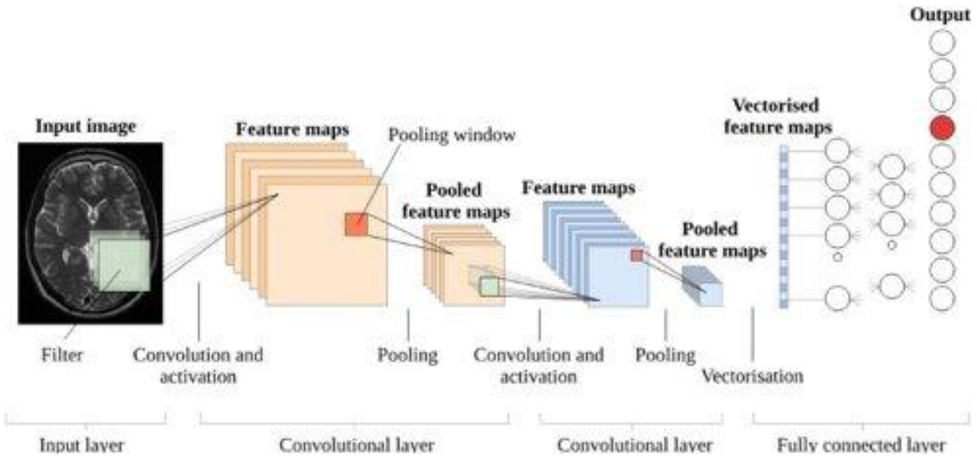
- Very very good documentation

- Tools for:
  - Image processing
  - Segmentation
  - Registration
  - ...

Medical Imaging and Deep Learning







Challenges

Artificial Intelligence for Health Care

- Limited annotated datasets
- Small samples of anomalies
- Variation in annotation even between doctors
- Variation in appearance due to different sensors, different parameters
- Bias in the dataset, the AI algorithms are built on
- No explanatory power. AI works well, but it is very difficult to explain why

Challenges

Artificial Intelligence for Health Care

- Limited annotated datasets
- Small samples of anomalies
- Variation in annotation even between doctors
- Variation in appearance due to different sensors, different parameters
- Bias in the dataset, the AI algorithms are built on
- No explanatory power. AI works well, but it is very difficult to explain why

[More and more publicly available datasets e.g. Luna, Brats, TCGA, Bowl 2017, ...]

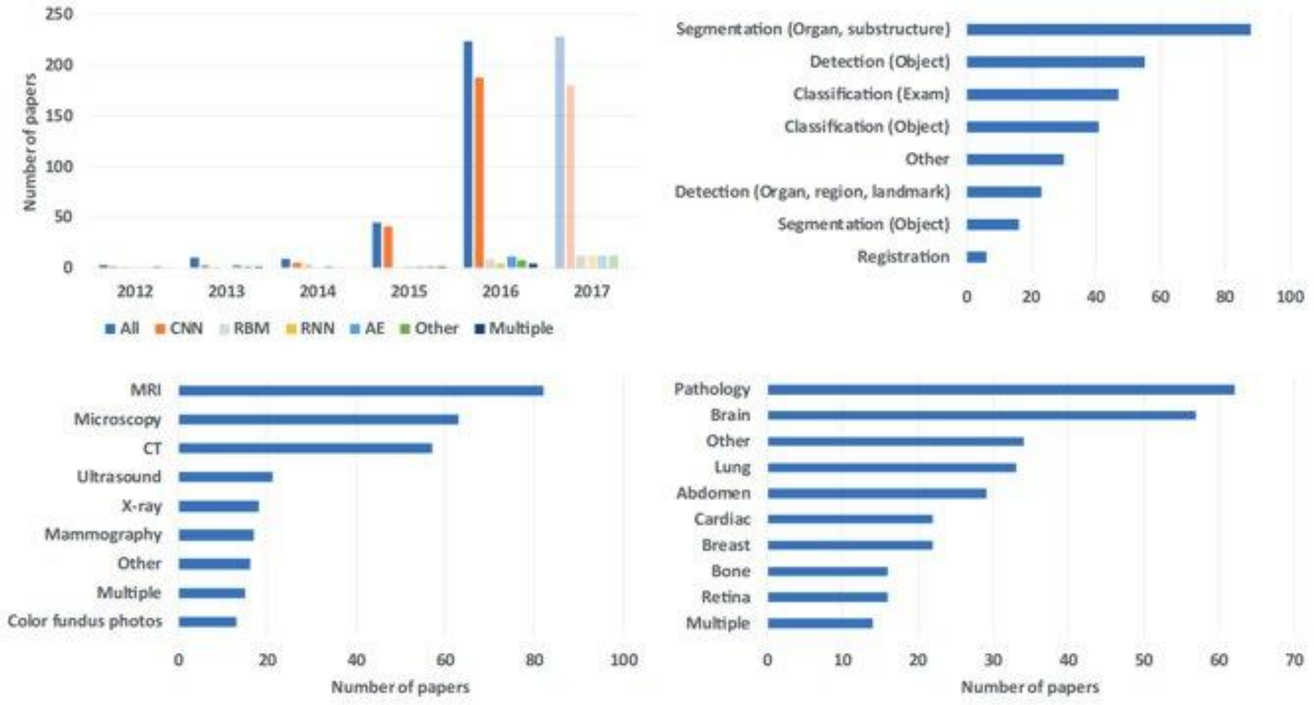
[Different models for different modalities, transfer learning, few shot learning, ...]

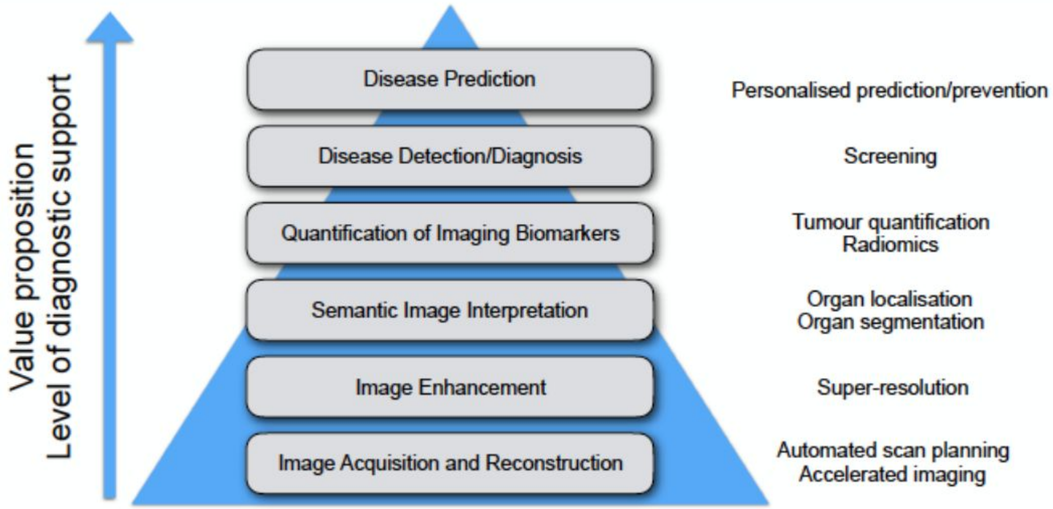
[Annotations from different medical experts, exploitation of more than one datasets.]

[Combine models with strong priors, analyse the decisions of the models]

Current Research

- A lot of research focuses currently on deep learning techniques and their use in medical imaging.





Arxiv: 1702.05747

Current Research

A lot of similarities and similar algorithmic choices with computer vision and deep learning with focus on:

- Limited amount of data
- Domain adaptation
- Specific metrics for evaluation and training
- Exploitation of the 3D nature of data
- Objects, organs that are very small in size and difficult to distinguish from the background
- Multimodality and mechanisms for proper fusion of data
- Self-supervision and transfer learning

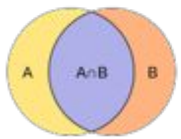
Semantic Segmentation

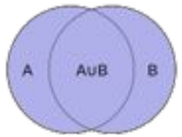
- **Evaluation metrics usually applied**
  - Most widely used measure for evaluating segmentation in medical imaging
  - It is calculated between the reference, and the precision
  - Dice Coefficient

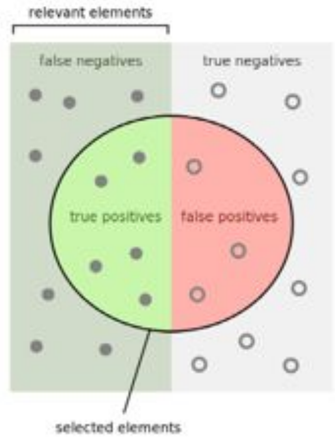
$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

- $|A| = TP + FN$
- $|B| = TP + FP$
- $|A \cap B| = TP$

$$DSC = \frac{2TP}{2TP + FP + FN} = F_1$$



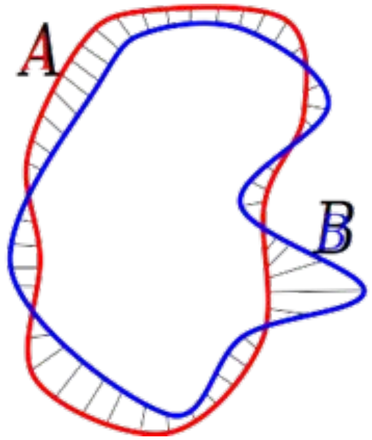




  - Hausdorff Distance

$$HD = \max\big(h(A,B), h(B,A)\big)$$

$$h(A,B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$



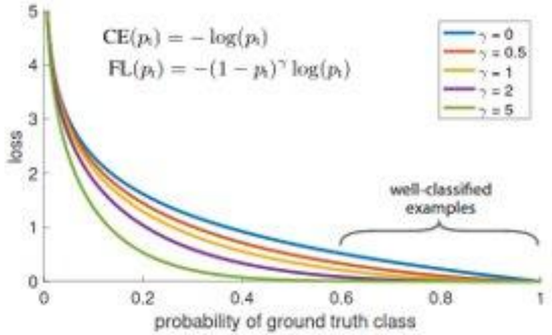
Semantic Segmentation

- **Loss functions**
  - Dice Loss [Milletari et al., 2016] $D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}$ $\frac{\partial D}{\partial p_j} = 2\left[\frac{g_j\left(\sum_i^N p_i^2 + \sum_i^N g_i^2\right) - 2p_j\left(\sum_i^N p_i g_i\right)}{\left(\sum_i^N p_i^2 + \sum_i^N g_i^2\right)^2}\right]$
  - Generalized Dice Loss [Sudre et al., 2017]

$$\mathrm{GDL} = 1 - 2\frac{\sum_{l=1}^{2} w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^{2} w_l \sum_n r_{ln} + p_{ln}}, \quad \frac{\partial \mathrm{GDL}}{\partial p_i} = -2\frac{(w_1^2 - w_2^2)\left[\sum_{n=1}^{N} p_n r_n - r_i \sum_{n=1}^{N}(p_n + r_n)\right] + N w_2 (w_1 + w_2)(1 - 2r_i)}{\left[(w_1 - w_2)\sum_{n=1}^{N}(p_n + r_n) + 2N w_2\right]^2}$$

  - Focal Loss [Li et al. 2018]
    Propose to reshape the loss function to down-weight easy examples and thus focus the training on hard negatives
    Binary Cross Entropy

$$\mathrm{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise.} \end{cases}$$



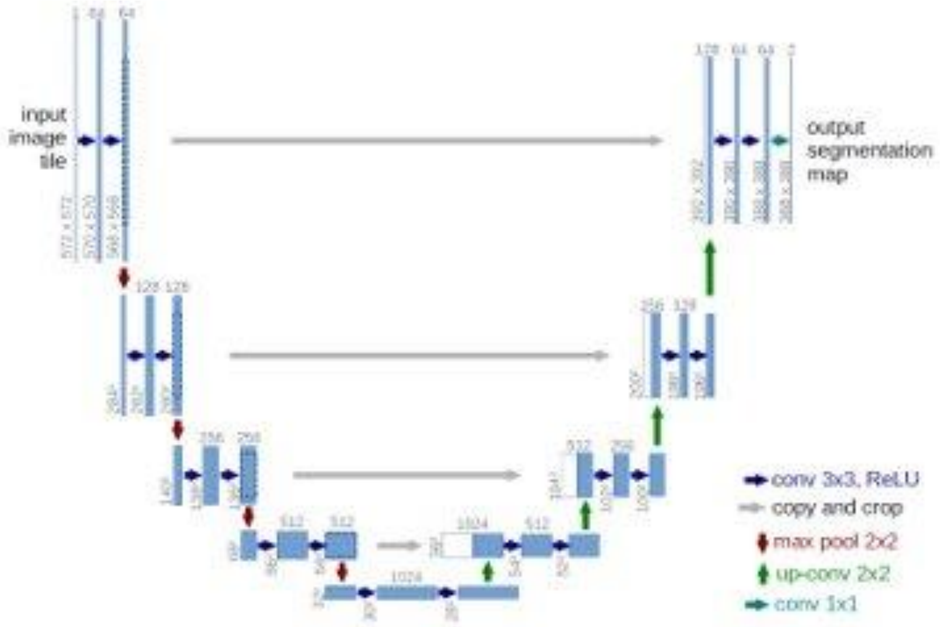
Focal Loss $\mathrm{FL}(p_t) = -(1-p_t)^\gamma \log(p_t).$
where $\gamma \geq 0$.
when an example is misclassified and p is small
the modulating factor is near 1 and the loss is unaffected. As p -> 1, the factor goes to 0 and the loss of the well-classified examples is down-weighted

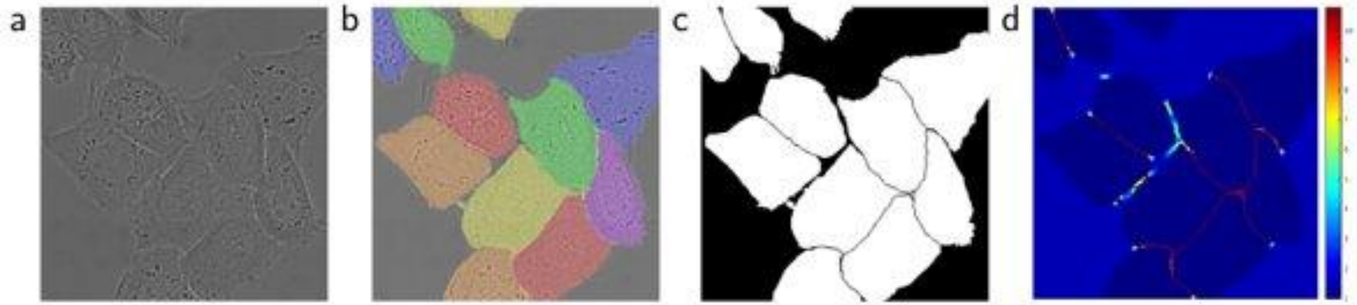
Semantic Segmentation

- **Ronneberger et al.** “U-Net: Convolutional Networks for BIomedical Image Segmentation” **In: MICCAI 2015**
  - Fully convolutional architecture
  - Depth of 5 layers with convolutions, max pooling, up convolution or deconvolution and skip connections
  - Loss Function:

$$E = \sum_{\mathbf{x} \in \Omega} w(\mathbf{x}) \log(p_{\ell(\mathbf{x})}(\mathbf{x}))$$

$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right)$$





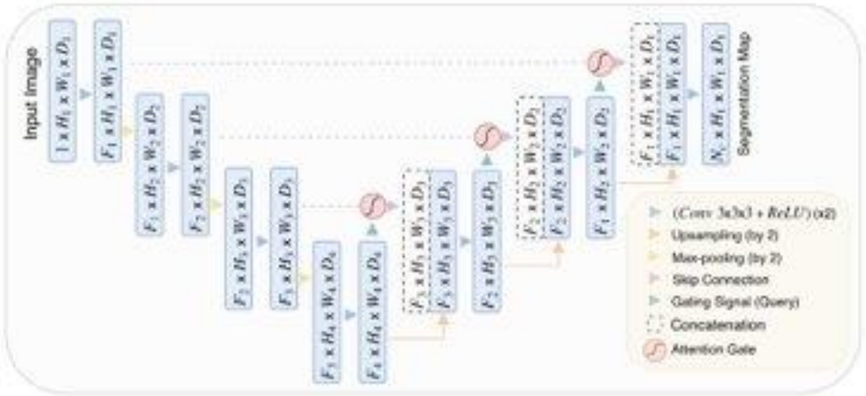
Semantic Segmentation

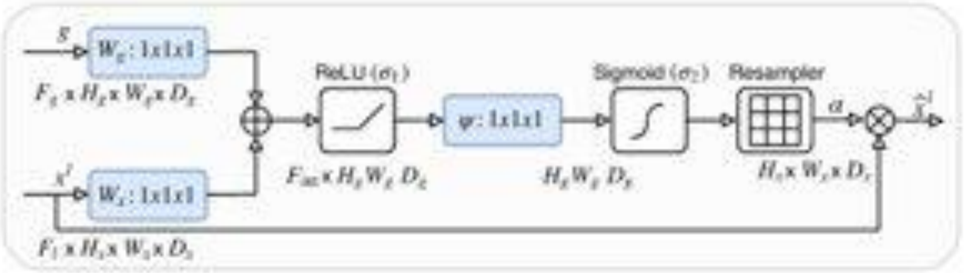
- **Oktay et al.** “Attention U-Net: Learning Where to Look for the Pancreas” **In: arXiv 2018**

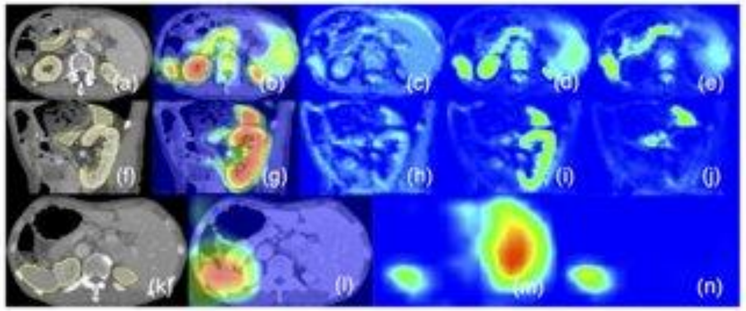
$$q_{att}^{l} = \psi^{T}\left(\sigma_1\left(W_x^{T} x_i^{l} + W_g^{T} g_i + b_g\right)\right) + b_\psi$$

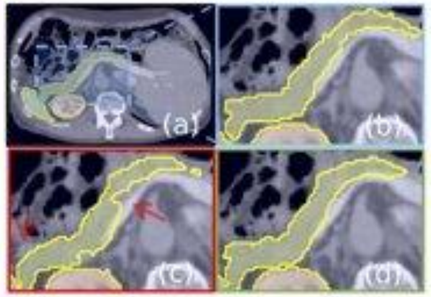
$$\alpha_i^{l} = \sigma_2\left(q_{att}^{l}(x_i^{l}, g_i; \Theta_{att})\right),$$

	Method	Dice Score	Precision	Recall	S2S Dist (mm)
BFT	U-Net [24]	0.690±0.132	0.680±0.109	0.733±0.190	6.389±3.900
	Attention U-Net	**0.712±0.110**	0.693±0.115	**0.751±0.149**	**5.251±2.551**
AFT	U-Net [24]	0.820±0.043	0.824±0.070	0.828±0.064	2.464±0.529
	Attention U-Net	**0.831±0.038**	0.825±0.073	**0.840±0.053**	**2.305±0.568**
SCR	U-Net [24]	0.815±0.068	0.815±0.105	0.826±0.062	2.576±1.180
	Attention U-Net	0.821±0.057	0.815±0.093	**0.835±0.057**	**2.333±0.856**









Semantic Segmentation

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

$$E = \sum_{\mathbf{x} \in \Omega} w(\mathbf{x}) \log(p_{\ell(\mathbf{x})}(\mathbf{x}))$$

(a) (b)

Fig. 3: **(a)** The confocal recording of our 3rd *Xenopus* kidney. **(b)** Resulting dense segmentation from the proposed 3D u-net with batch normalization.

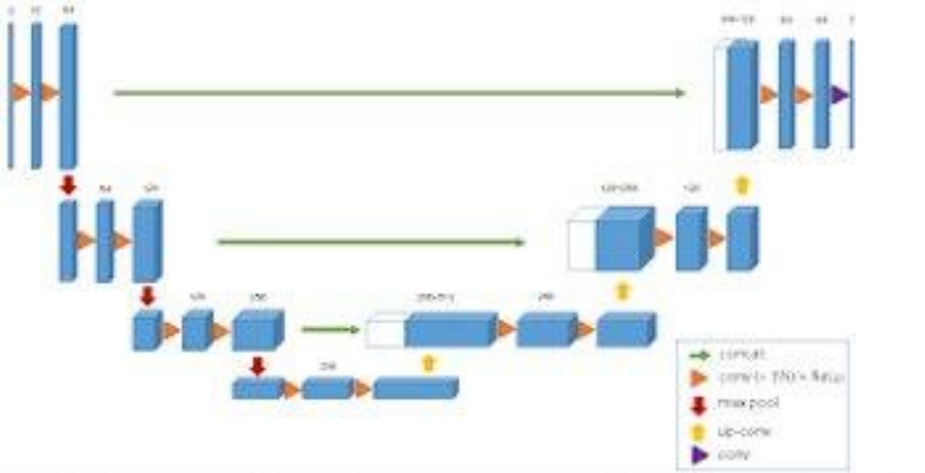


Table 1: Cross validation results for semi-automated segmentation (IoU)

test slices	3D w/o BN	3D with BN	2D with BN
subset 1	0.822	0.855	0.785
subset 2	0.857	0.871	0.820
subset 3	0.846	0.863	0.782
average	0.842	0.863	0.796

Table 2: Effect of # of slices for semi-automated segmentation (IoU)

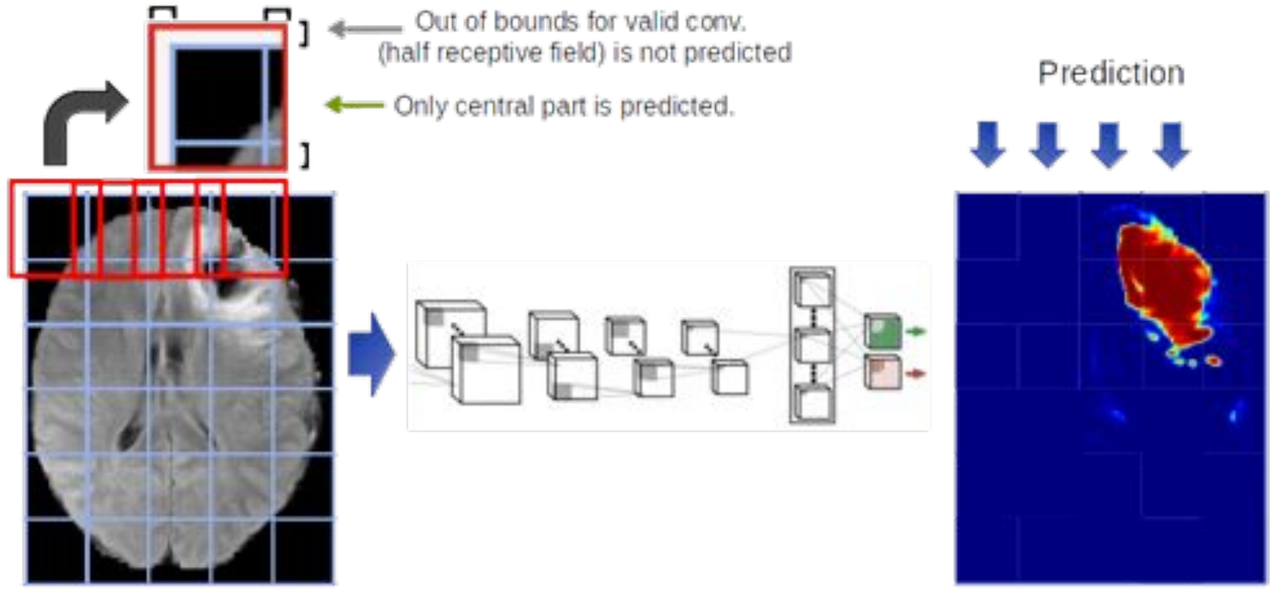
GT slices	GT voxels	IoU S1	IoU S2	IoU S3
1,1,1	2.5%	0.331	0.483	0.475
2,2,1	3.3%	0.676	0.579	0.738
3,3,2	5.7%	0.761	0.808	0.835
5,5,3	8.9%	0.856	0.849	0.872

Table 3: Cross validation results for fully-automated segmentation (IoU)

test volume	3D w/o BN	3D with BN	2D with BN
1	0.655	0.761	0.619
2	0.734	0.798	0.698
3	0.779	0.554	0.325
average	0.723	0.704	0.547

Semantic Segmentation

- **Tiling for 3D networks**



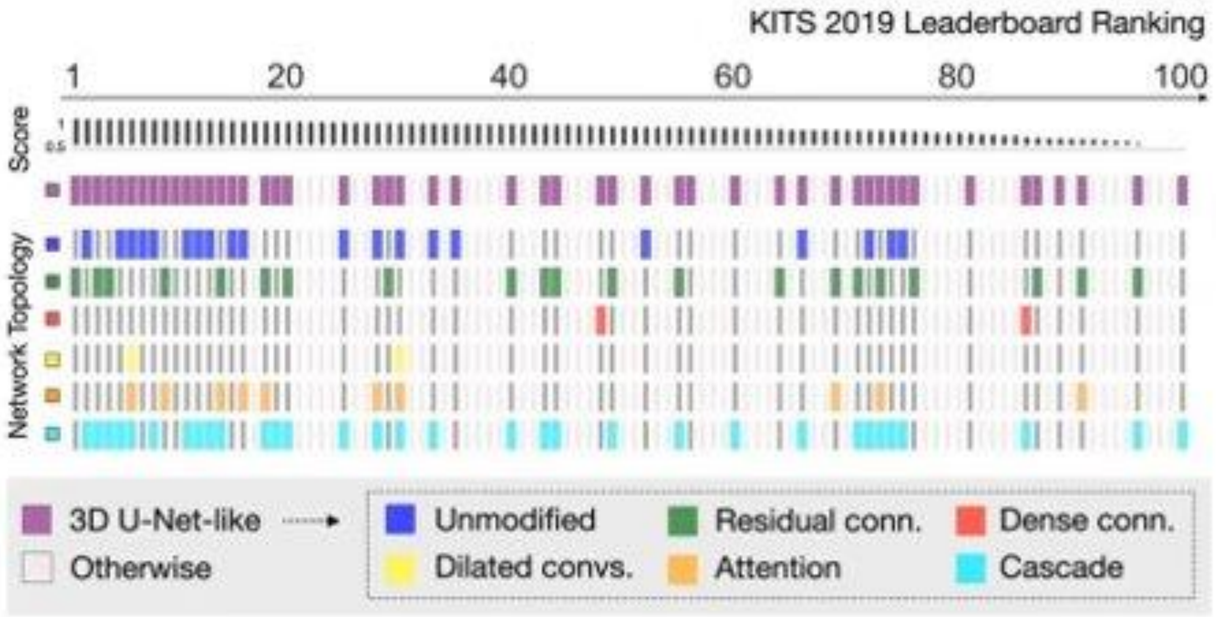
Bigger input tiles (or segments):

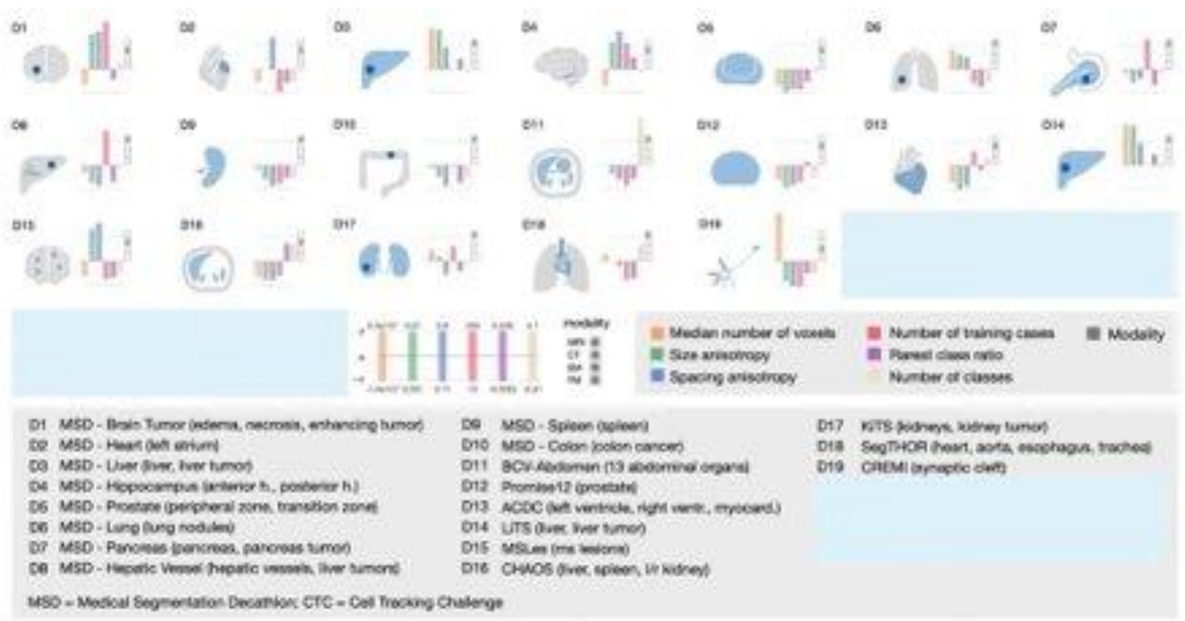
- Require more memory (FMs expand)
- But fewer redundant computations (less overlap)

Tile size during testing can be different from training

Semantic Segmentation

- **Isensee et al.** “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation” **In: Nature Methods 2021**





Semantic Segmentation

- **Isensee et al.** “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation” **In: Nature Methods 2021**

Overall observations

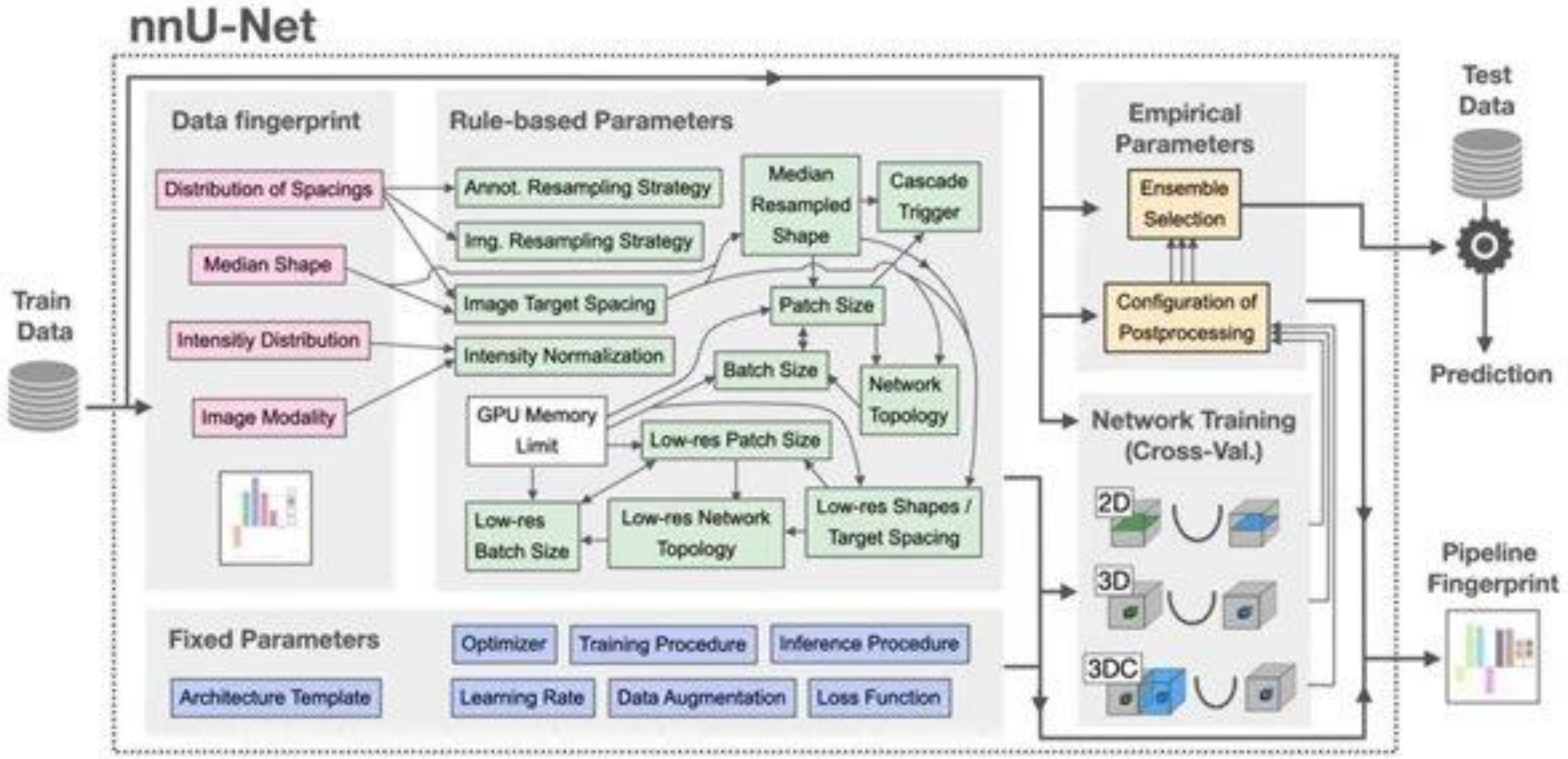
- Time consuming trial and error process
- Success depends on experience of the researcher
- Needs to be repeated on every dataset

nnU-Net: Proposed recipe

- **Fixed Parameters:** Collect design decisions that do not require adaptation between datasets and identify a robust common configuration
- **Rule-based Parameters**: For as many of the remaining decisions as possible formulate explicit dependencies between specific dataset properties ("dataset fingerprint") and design choices ("pipeline fingerpring") in the form of heuristic rules to allow for almost instant adaptation on application.
- **Empirical Parameters:** Learn only the remaining decisions empirically from the data

Semantic Segmentation

- **Isensee et al.** “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation” **In: Nature Methods 2021**



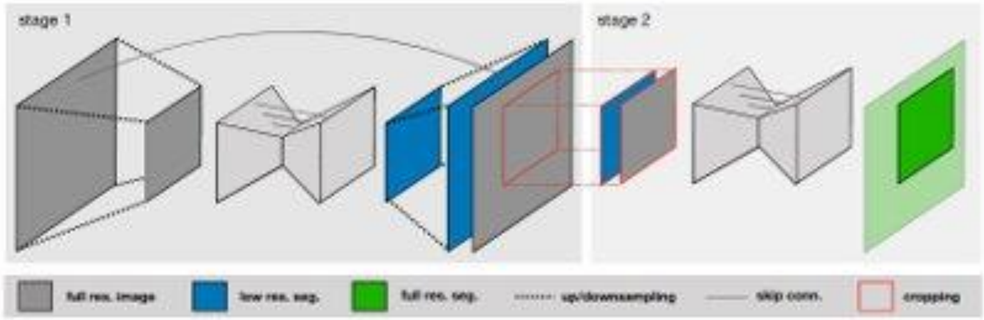
Semantic Segmentation

- **Isensee et al.** “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation” **In: Nature Methods 2021**

$$\mathcal{L}_{total} = \mathcal{L}_{dice} + \mathcal{L}_{CE}$$

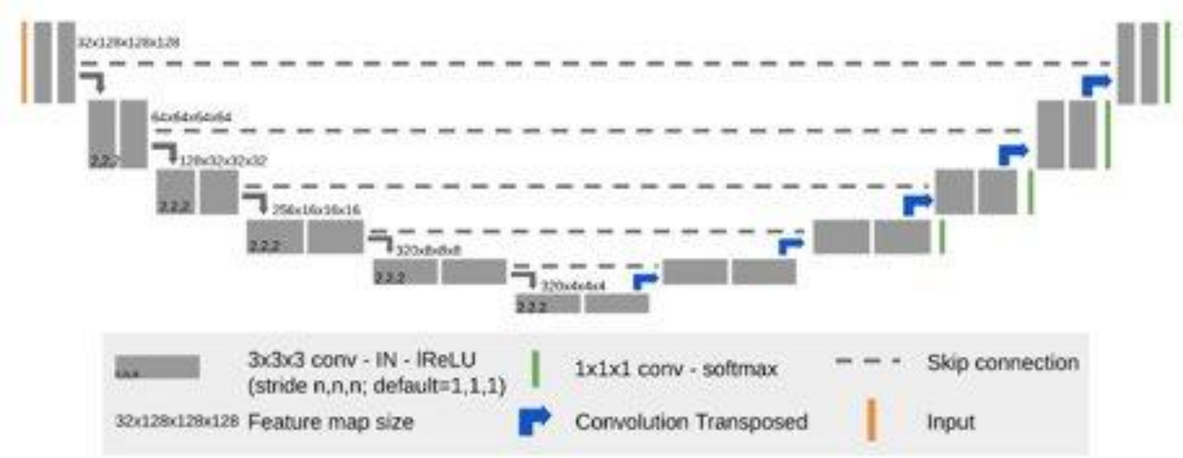
		2D U-Net	3D U-Net	3D U-Net lowres
BrainTumour	median patient shape	169x138	138x169x138	-
	input patch size	192x160	128x128x128	-
	batch size	89	2	-
	num pool per axis	5, 5	5, 5, 5	-
Heart	median patient shape	320x232	115x320x232	58x160x116
	input patch size	320x256	80x192x128	64x160x128
	batch size	33	2	2
	num pool per axis	6, 6	4, 5, 5	4, 5, 5
Liver	median patient shape	512x512	482x512x512	121x128x128
	input patch size	512x512	128x128x128	128x128x128
	batch size	10	2	2
	num pool per axis	6, 6	5, 5, 5	5, 5, 5
Hippocampus	median patient shape	50x35	36x50x35	-
	input patch size	56x40	40x56x40	-
	batch size	366	9	-
	num pool per axis	3, 3	3, 3, 3	-
Prostate	median patient shape	320x319	20x320x319	-
	input patch size	320x320	20x192x192	-
	batch size	26	4	-
	num pool per axis	6, 6	2, 5, 5	-
Lung	median patient shape	512x512	252x512x512	126x256x256
	input patch size	512x512	112x128x128	112x128x128
	batch size	10	2	2
	num pool per axis	6, 6	4, 5, 5	4, 5, 5
Pancreas	median patient shape	512x512	96x512x512	96x256x256
	input patch size	512x512	96x160x128	96x160x128
	batch size	10	2	2
	num pool per axis	6, 6	4, 5, 5	4, 5, 5

label	BrainTumour 1	BrainTumour 2	BrainTumour 3	Heart 1	Liver 1	Liver 2	Hippoc. 1	Hippoc. 2	Prostate 1	Prostate 2	Lung 1	Pancreas 1	Pancreas 2
2D U-Net	78.60	58.65	77.42	91.36	94.37	53.94	88.52	86.70	61.98	84.31	52.68	74.70	35.41
3D U-Net	**80.71**	**62.22**	**79.07**	92.45	94.11	61.74	**89.87**	**88.20**	60.77	83.73	55.87	77.69	42.69
3D U-Net stage1 only (U-Net Cascade)	-	-	-	90.63	94.69	47.01	-	-	-	-	65.33	79.45	49.65
3D U-Net (U-Net Cascade)	-	-	-	92.40	95.38	58.49	-	-	-	-	**66.85**	**79.30**	**52.12**
ensemble 2D U-Net+ 3D U-Net	80.79	61.72	79.16	**92.70**	94.30	60.24	89.78	88.09	**63.78**	**85.31**	55.96	78.26	40.46
ensemble 2D U-Net+ 3D U-Net (U-Net Cascade)	-	-	-	92.64	95.31	60.09	-	-	-	-	61.18	78.79	45.46
ensemble 3D U-Net+ 3D U-Net (U-Net Cascade)	-	-	-	92.63	**95.43**	**61.82**	-	-	-	-	65.16	79.70	49.14
test set	67.71	**47.73**	**68.16**	**92.77**	**95.24**	**73.71**	**90.37**	**88.95**	**75.81**	**89.59**	**69.20**	**79.53**	**52.27**



Semantic Segmentation

- **Isensee et al.** “nnU-Net for Brain Tumor Segmentation” **In: MICCAI Brainlesion 2020**



- Additional Modifications:
  - Region-based training [R]: training on the validated classes
  - Post processing especially on the enhancing tumor (remove predictions smaller than a threshold)
  - Big batch size
  - More data augmentation [DA]
    - Increase probability for rotation, scaling
    - Brightness augmentation
    - Elastic deformations

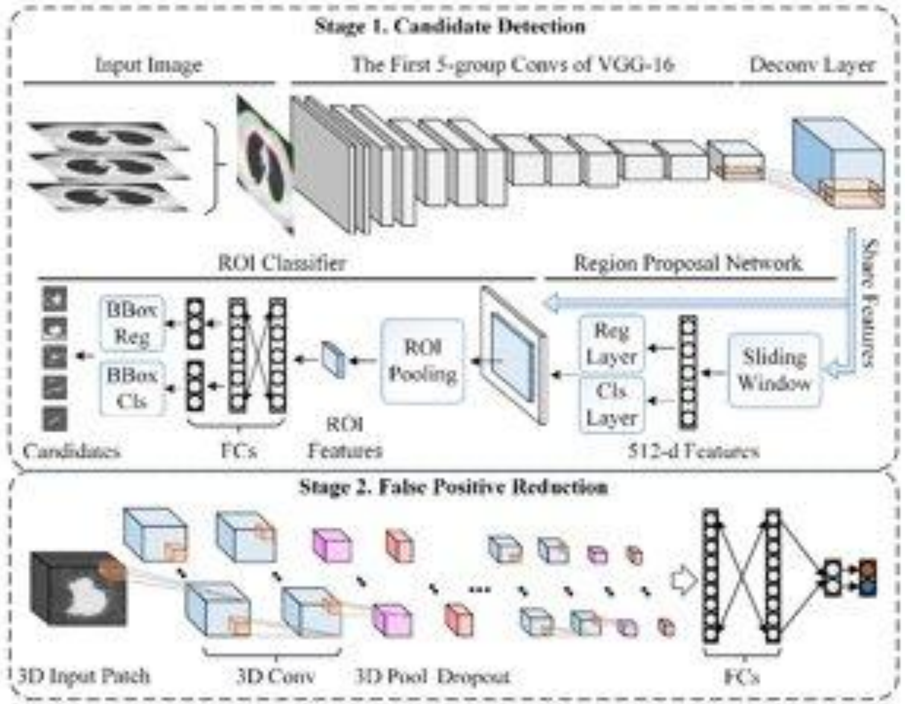
Semantic Segmentation

- **Isensee et al.** “nnU-Net for Brain Tumor Segmentation” **In: MICCAI Brainlesion 2020**
  - Additional Modifications:
    - Batch normalization [BN] (instead of instance normalization)
    - Batch dice [BD] Calculate the dice for the entire mini-batch and not per sample. It helps with the cases with very small number of segmentations

Model	Training set				Validation set			
	rank based on mean Dice		BraTS ranking		rank based on mean Dice		BraTS ranking	
	value	rank	value	rank	value	rank	value	rank
BL	86.55	8	0.3763	8	84.18	6	0.4079	8
BL*	86.09	9	0.3767	9	83.76	9	0.4236	9
BL*+R	86.73	5	0.3393	5	84.13	7	0.4005	7
BL*+R+DA	87.07	**1**	0.3243	3	84.73	5	0.3647	5
BL*+R+DA+BN	86.82	3	0.3377	4	85.20	3	0.3577	4
BL*+R+DA+BD	86.79	4	0.3231	2	84.12	8	0.3726	6
BL*+R+DA+BN+BD	86.87	2	0.3226	**1**	85.11	4	0.3487	3*
BL*+R+DA*+BN	86.68	6	0.3521	6	85.58	**1**	0.3125	**1***
BL*+R+DA*+BN+BD	86.64	7	0.3595	7	85.29	2	0.3437	2*

Object Detection

- **Ding et al.** “Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks” **In: MICCAI 2017**
  - Lung Nodule Detection (Luna):
    - A big number of false positives for nodule detection due to very small size.
    - 2 stage process!
    - Faster R-CNN to detect the boxes of possible nodules
    - 3D network to decide if the detection is good or not!



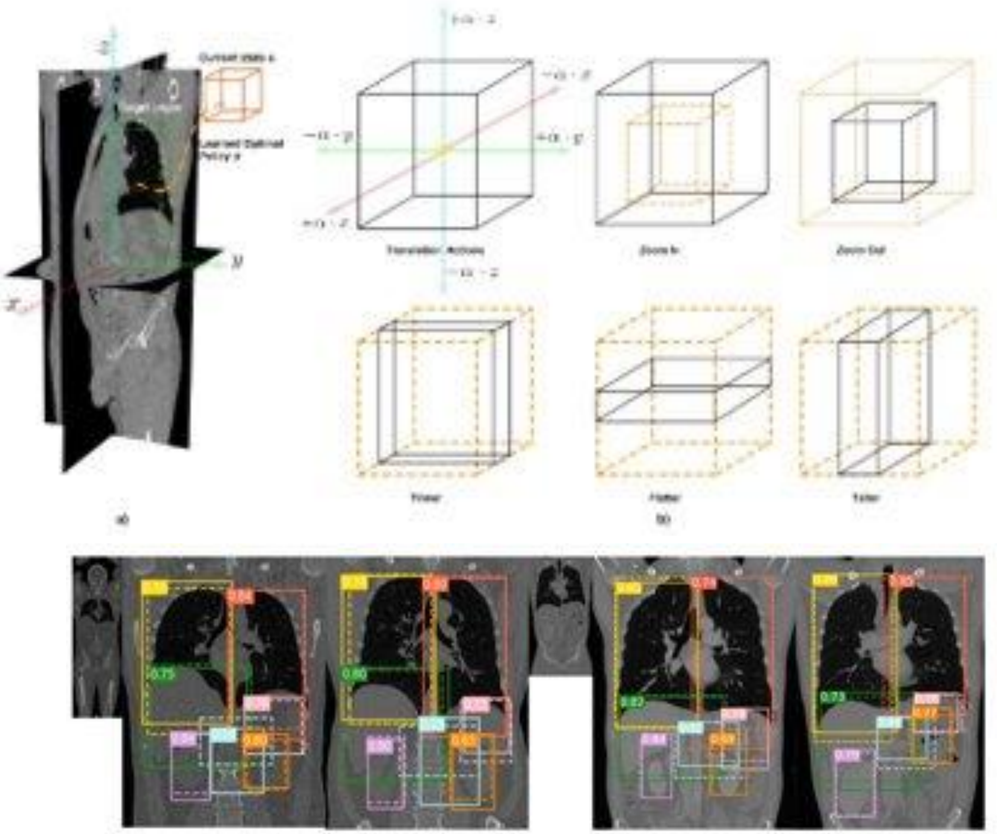
Object Detection

- **Navarro et al.** “Deep reinforcement learning for organ localization in CT” **In: MIDL 2020**
  - The RL environment and action space is the 3D CT scan.
  - The action space consists of 6 actions for translation, 2 for scaling the whole box and 3 to scale the box in each of the three directions.
  - Based on Q-learning algorithm

$$Q^*(s,a) = \max_{\pi} \mathbb{E}[R_t \mid s_t = s, a_t = a, \pi]$$

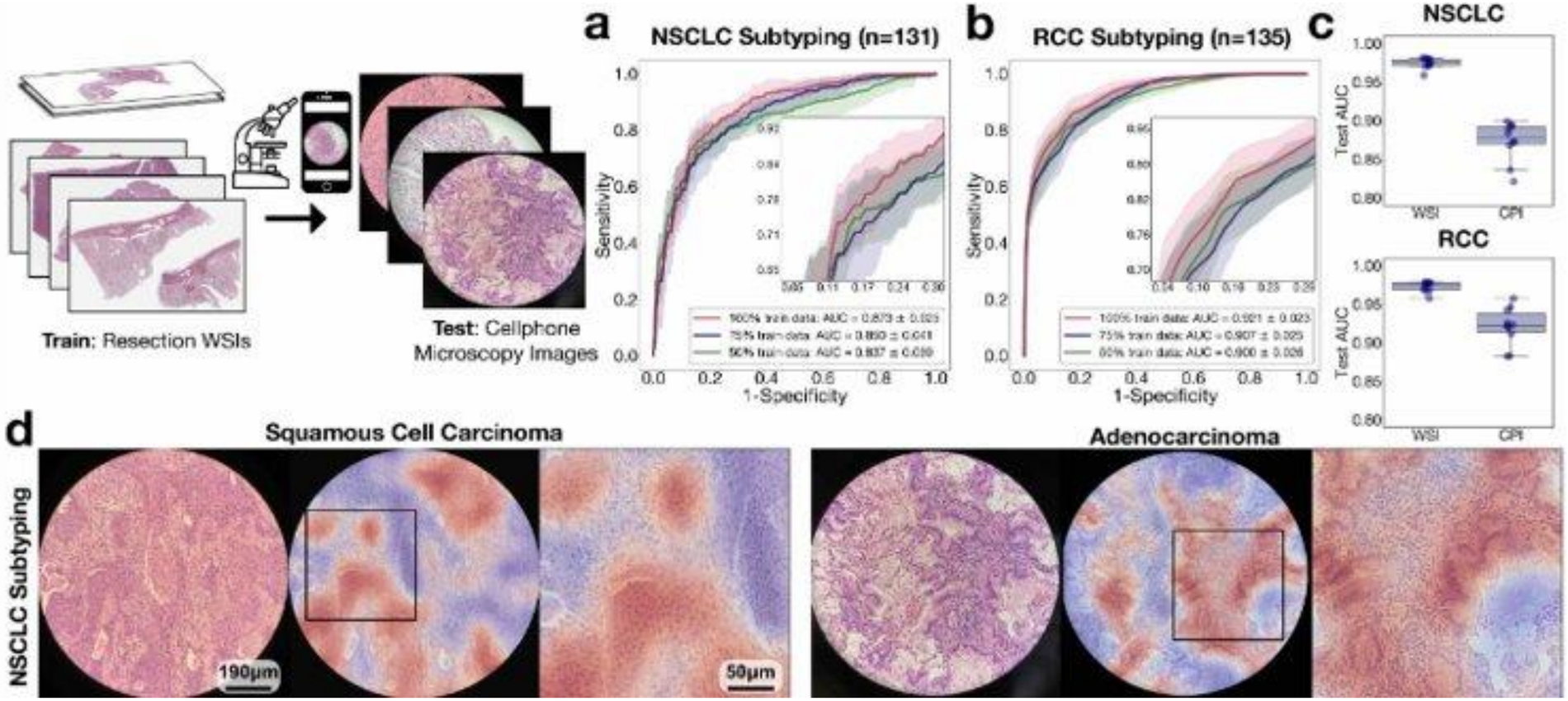
  - R_t: expected reward, π policy

	Avg IoU	Wall dist [mm]	Centroid dist [mm]
Right Lung	0.77	3.46 ± 5.28	6.06 ± 10.25
Left Lung	0.73	4.91 ± 7.38	10.32 ± 17.09
Right Kidney	0.60	2.96 ± 2.91	5.69 ± 5.67
Left Kidney	0.57	4.06 ± 4.98	7.52 ± 9.02
Liver	0.80	2.41 ± 0.70	3.36 ± 1.34
Spleen	0.60	5.25 ± 7.23	9.20 ± 12.03
Pancreas	0.32	12.26 ± 13.60	20.79 ± 20.38
Global	0.63	5.04 ± 6.01	8.99 ± 10.82
Median	0.60	2.25	3.65



Classification

- **Lu et al.** “Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images” **In: Nature Biomedical Engineering 2021**



Classification

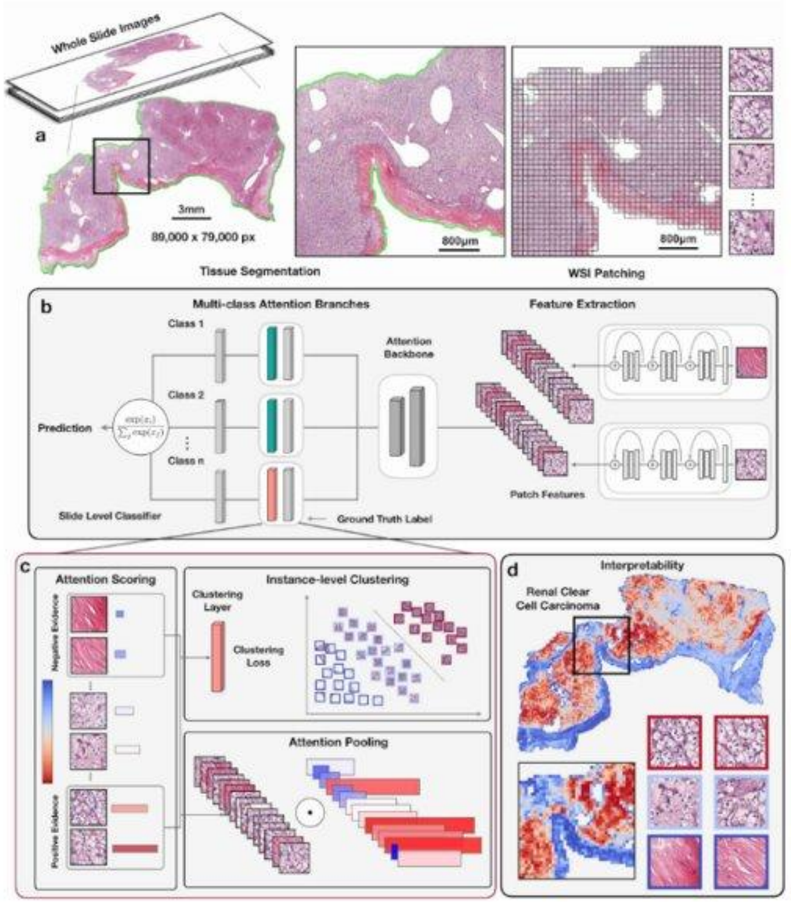
- **Lu et al.** “Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images” **In: Nature Biomedical Engineering 2021**

Clustering-constrained attention multiple instance learning (CLAM)

- Built on top of multi instance learning setup
- Use of different aggregation functions (e.g. mean operator, generalizec mean, log-sum-exp, the quantile function, Noisy-OR, Nosiy-And)
- Use of a pretrained ResNet50 network per patch.
- Reduce of the dimensionality from 1024-dimensional patch to 512-dimentionsional patch
- Attention module using two layers V_a and U_a

$$a_{k,m} = \frac{\exp\left\{\mathbf{W}_{a,m}\left(\tanh\left(\mathbf{V}_a\mathbf{h}_k^\top\right)\odot \mathrm{sigm}\left(\mathbf{U}_a\mathbf{h}_k^\top\right)\right)\right\}}{\sum_{j=1}^{N}\exp\left\{\mathbf{W}_{a,m}\left(\tanh\left(\mathbf{V}_a\mathbf{h}_j^\top\right)\odot \mathrm{sigm}\left(\mathbf{U}_a\mathbf{h}_j^\top\right)\right)\right\}}$$

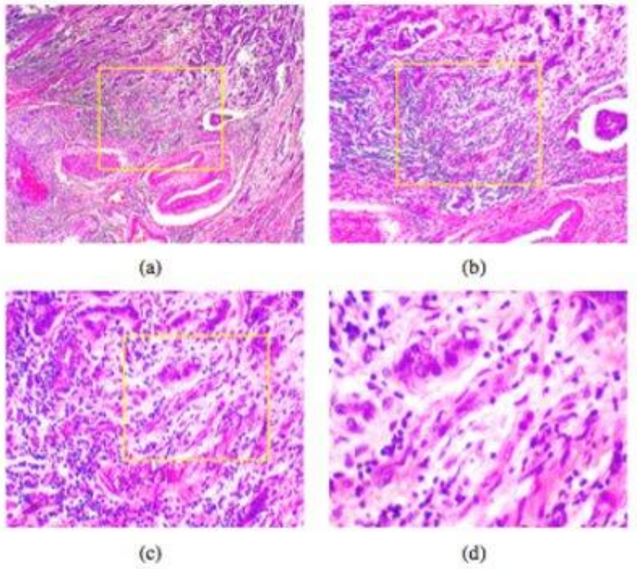
- Two different tasks to solve:
  - Instance-level clustering using a fully connected layer with 512 hidden units
  - Smooth top1 SVM loss (ICLR 2018)



Classification

- **Sahasrabudhe et al.** “Self-Supervised Nuclei Segmentation in Histopathological Images Using Attention” **In: MICCAI 2020**

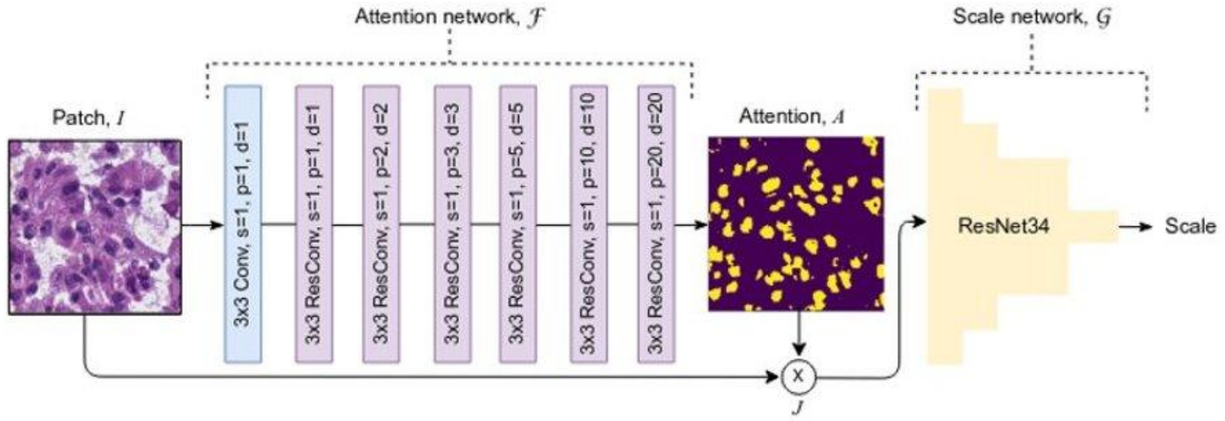
  Main assumption: Given a patch extracted from a WSI viewed at a certain magnification, the level of magnification can be ascertained by looking at the size and texture of the nuclei in the patch.
  - Train a scale-sensitive network which specifically learns discriminative features for correct scale classification.



A slide of breast malignant tumor (stained with HE) seen in different magnification factors: (a) 40X, (b) 100X, (c) 200X, and (d) 400X.

Classification

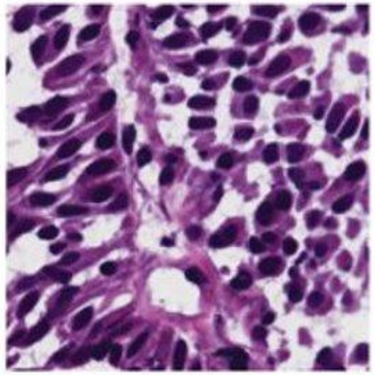
- **Sahasrabudhe et al.** “Self-Supervised Nuclei Segmentation in Histopathological Images Using Attention” **In: MICCAI 2020**



- Our network is composed by two different components:
  - The attention network F [Input: I, Output A]
    - Based on dilated filters
    - $J = A \odot I$, multiplication of the “attended” image and the initial/ patch
  - The scale network G [Input: J, Output: Scale]
    - ResNet34

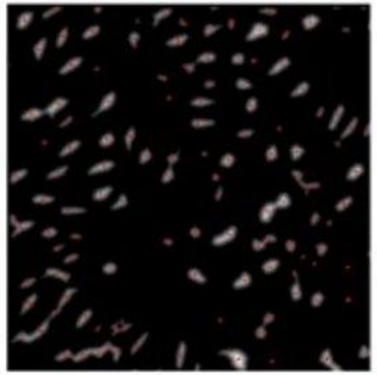
Classification

- **Sahasrabudhe et al.** “Self-Supervised Nuclei Segmentation in Histopathological Images Using Attention” **In: MICCAI 2020**
  - Instance segmentation of the nuclei applying a simple post-processing:
    - Morphological Operations
    - Distance Transform to detect the centers of the bloods
    - Watershed algorithm for the instance segmentation

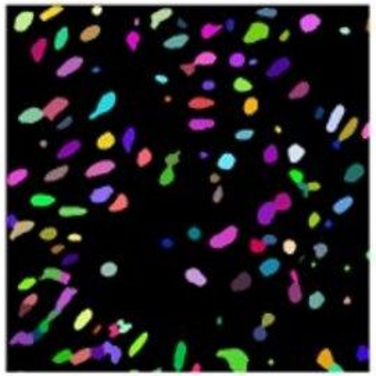


Initial Patch

Attention map



Distance Transform

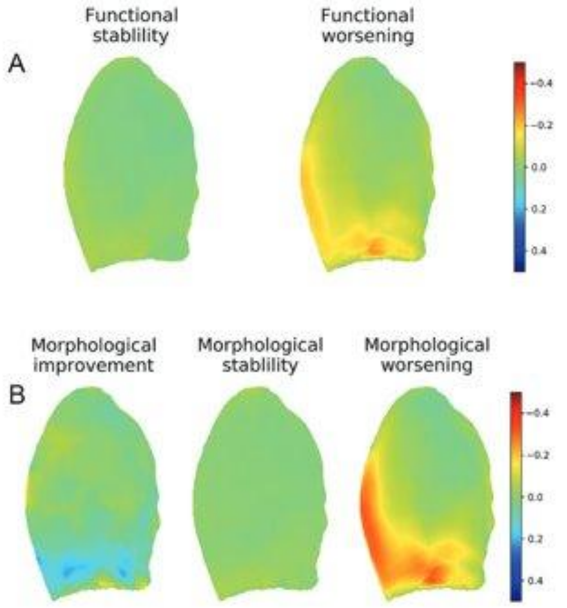


Instance Segmentation

Test dataset	Method	AJI [13]	AHD	ADC
MoNuSeg test	CNN2 [13]†	0.3482	8.6924	0.6928
	CNN3 [13]†	0.5083	7.6615	**0.7623**
	Best Supervised [13]†	**0.691**	-	-
	CellProfiler [13]	0.1232	9.2771	0.5974
	Fiji [13]	0.2733	8.9507	0.6493
	$M_{\neg sparse}$	0.0312	13.1415	0.2283
	$M_{\neg smooth}$	0.1929	8.8166	0.4789
	$M_{\neg WSI}$	0.3025	8.2853	0.6209
	$M_{\neg equiv}$	0.4938	8.0091	0.7136
	$M_{proposed}$	**0.5354**	**7.7502**	**0.7477**
TNBC[15]	U-Net[7]†	0.514	-	0.681
	SegNet+WS[7]†	0.559	-	**0.758**
	HoverNet[7]†	**0.590**	-	0.749
	CellProfiler	0.2080	-	0.4157
	$M_{proposed}$	0.2656	-	0.5139
CoNSeP[7]	SegNet[7]†	0.194	-	**0.796**
	U-Net[7]†	**0.482**	-	0.724
	CellProfiler[7]	0.202	-	0.434
	QuPath[7]	0.249	-	0.588
	$M_{proposed}$	0.1980	-	0.587

Image Registration

- The goal of registration is to define a transformation ***T*** which projects one image (source) to the other (target).
  - Two main categories
    - Rigid methods
    - Non-rigid (Deformable methods)
  - A lot of applications
    - Atlas based segmentation
    - Fusion of different modalities
    - Monitoring of diseases
    - Intra and inter patient



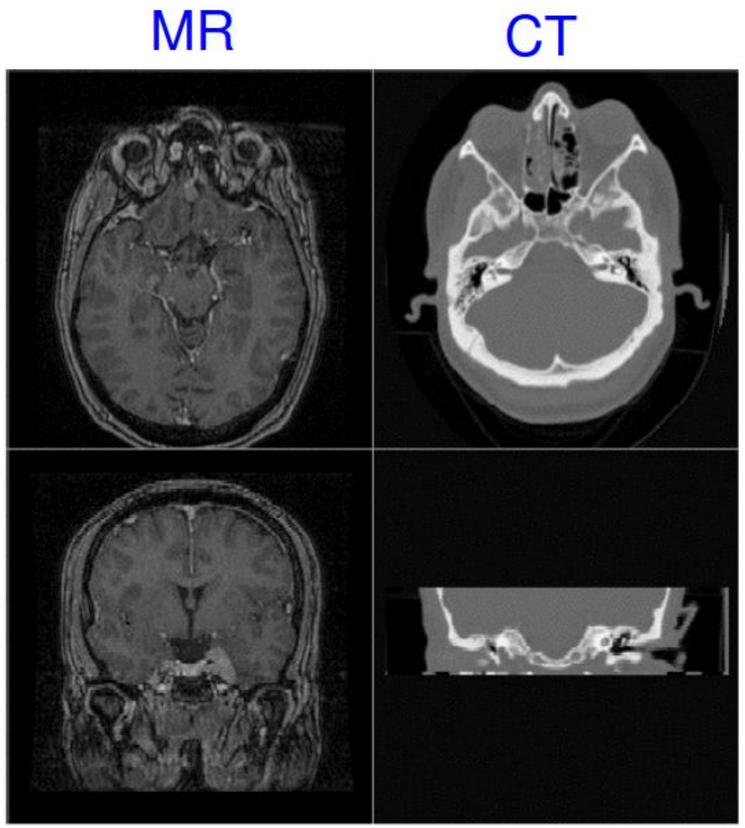


Image Registration

- **Christodoulidis et al.** “Linear and Deformable Image Registration with 3D Convolutional Neural Networks” **In: RAMBO MICCAI 2018**
  - Non-rigid, deformable procedure, the observed signals are associated through a non-linear dense transformation, or a spatially varying model.
  - Three main components
    - 3D Spatial Transformer
    - 3D CNN Architecture
    - Loss Function

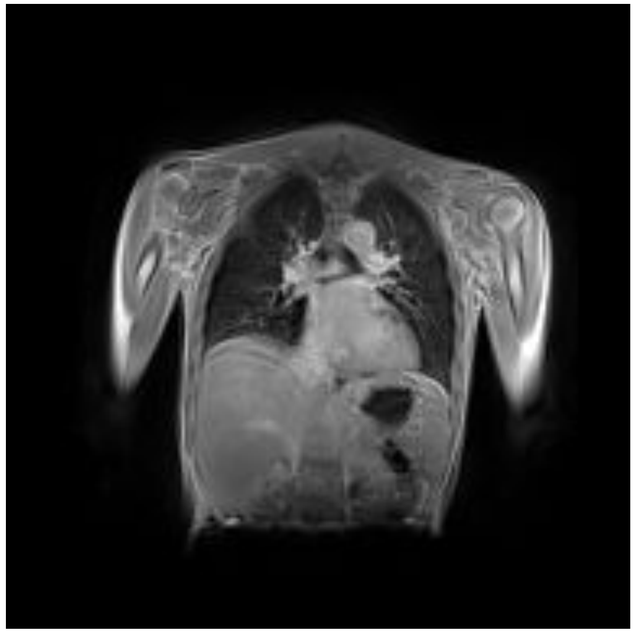


Image Registration

- **Christodoulidis et al.** “Linear and Deformable Image Registration with 3D Convolutional Neural Networks” **In: RAMBO MICCAI 2018**
  - 3D Spatial transformer
    - A differentiable module which applies a spatial transformation to an image. Backward trilinear interpolation sampling

$$D(\mathbf{p}) = \sum_{\mathbf{q}} S(\mathbf{q}) \prod_{d} \max\left(0, 1 - |[G(\mathbf{p})]_d - \mathbf{q}_d|\right),$$

  - 3D CNN Architecture
    - Based on an encoder-decoder framework, using dilated convolutional kernels. Output of the decoder is the spatial gradient along each axis ∇G. The grid is calculated by:

$$G(\mathbf{p}) = \int_{-\infty}^{\mathbf{p}} \nabla G \, \mathrm{d}\mathbf{p}$$

  - Loss Function

$$\text{Loss} = \|R - \mathcal{W}_G(S)\|^2 + \alpha \|A - A_I\|_1 + \beta \|\Phi - \Phi_I\|_1,$$

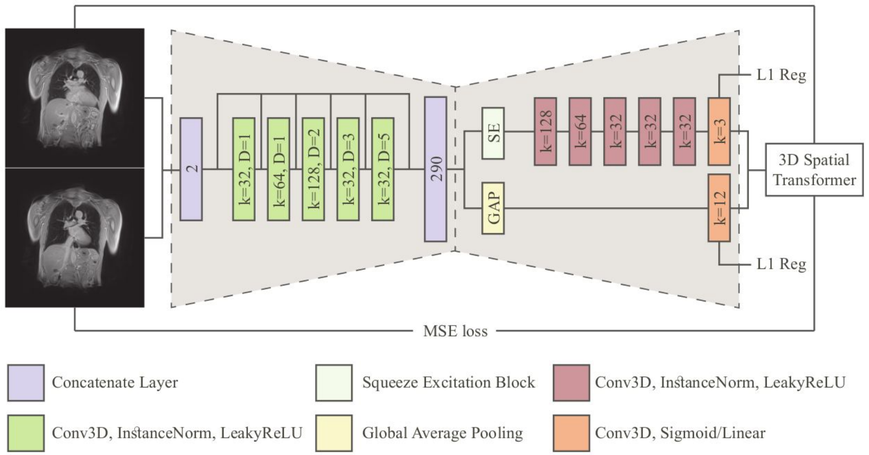
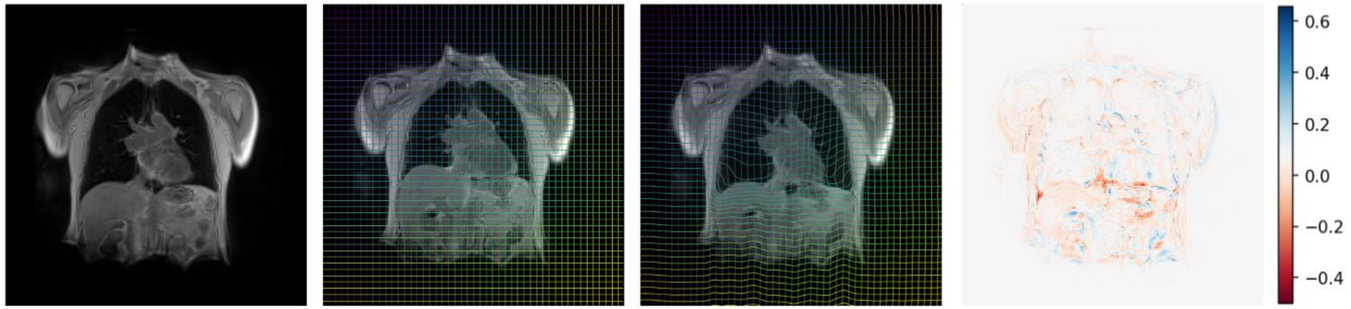


Image Registration

- **Christodoulidis et al.** "Linear and Deformable Image Registration with 3D Convolutional Neural Networks" **In: RAMBO MICCAI 2018**

Method	*dx*	*dy*	*dz*	*ds*
Inter-observer	1.664	2.545	1.555	3.905
Deformable with NCC, DWM, and MI [7]	1.855	3.169	2.229	4.699
Proposed w/o Affine	2.014	2.947	1.858	4.569
Proposed	**1.793**	**2.904**	**1.822**	**4.358**

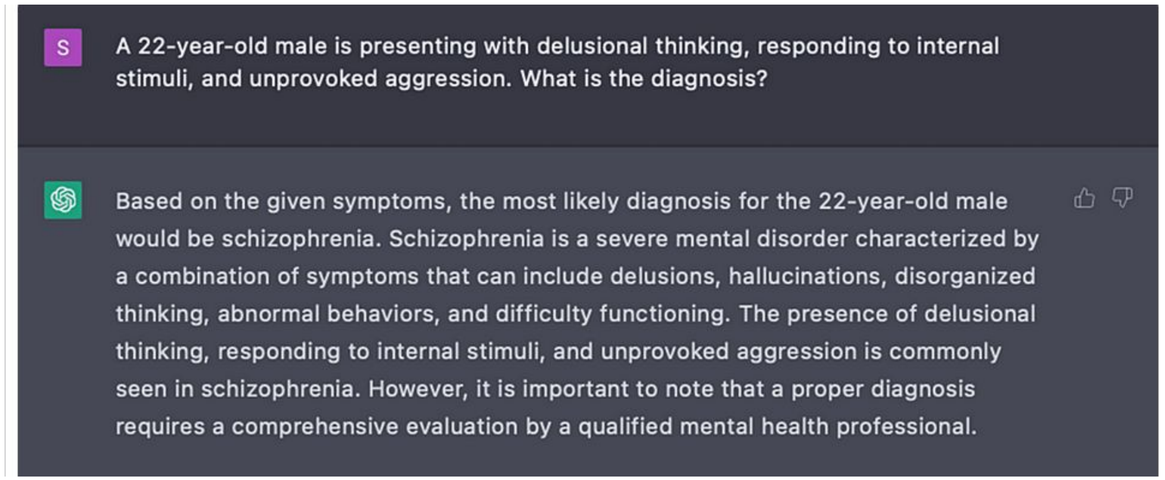
Method	Inhale-Exhale	All Combinations	Time/subject (s)
Unregistered	75.62±10.89	57.22±12.90	–
Deformable with NCC [7]	84.25±6.89	76.10±7.92	∼1 (GPU)
Deformable with DWM [7]	88.63±4.67	75.92±8.81	∼2 (GPU)
Deformable with MI [7]	88.86±5.13	76.33±8.74	∼2 (GPU)
Deformable with all above [7]	88.81±5.85	78.71±8.56	∼2 (GPU)
SyN [2]	83.86±6.04	–	∼2500 (CPU)
Proposed w/o Affine	91.28±2.47	81.75±7.88	∼0.5 (GPU)
Proposed	**91.48±2.33**	**82.34±7.68**	∼0.5 (GPU)

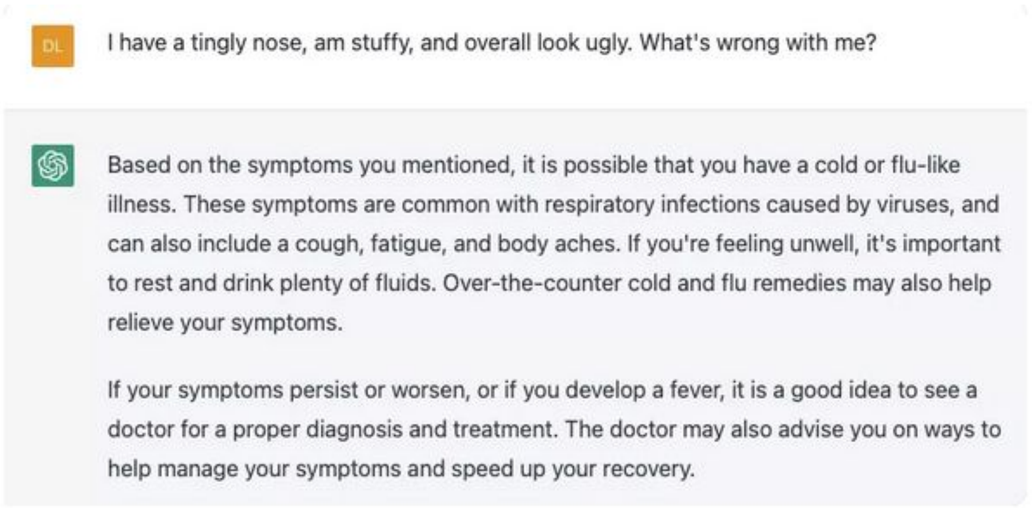


(a) Reference image (b) Moving image (c) Deformed image (d) Difference

Towards Large Deep Learning Models

- The community is working on building huge and powerful models for different tasks with currently impressive results on:
  - Large Language Models inferring relationships between words within the text
    - ChatGPT: 176-billion parameters trained in 570GB data from text, books, websites, and more





Galido P, Butala S, Chakerian M, et al. (April 26, 2023) A Case Study Demonstrating Applications of ChatGPT in the Clinical Management of Treatment-Resistant Schizophrenia . Cureus 15(4): e38166. doi:10.7759/cureus.38166

https://www.verywellhealth.com/chatgpt-in-healthcare-7107800

https://openai.com/blog/chatgpt

Towards Large Deep Learning Models

- The community is working on building huge and powerful models for different tasks with currently impressive results on:
  - Imaging tasks and image segmentation
    - SAM: trained on 1B masks and from 11M licensed and privacy-preserving image

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y. and Dollár, P., 2023. Segment anything. arXiv preprint arXiv:2304.02643.

Towards Large Deep Learning Models

- The community is working on building huge and powerful models for different tasks with currently impressive results on:
  - Image generation from text
    - DALL-E: 12-billion parameters trained in 250 million text-images pairs
    - GLIDE: 3.5-billion parameter text-conditional diffusion, and another 1.5 billion parameter text-conditional upsampling diffusion model.

"a boat in the canals of venice"

"a painting of a fox in the style of starry night"

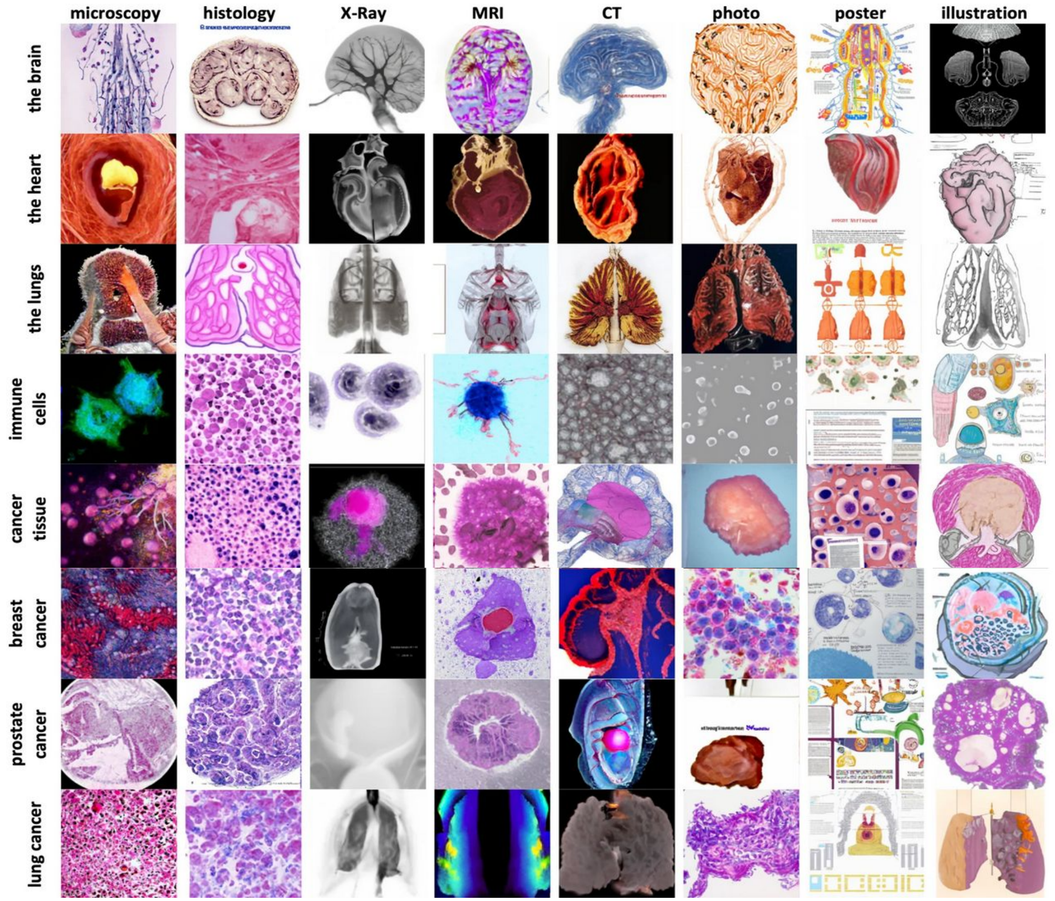
"a red cube on top of a blue cube"

"a stained glass window of a panda eating bamboo"

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I. and Chen, M., 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv preprint arXiv:2112.10741.

Challenges in the Medical Domain

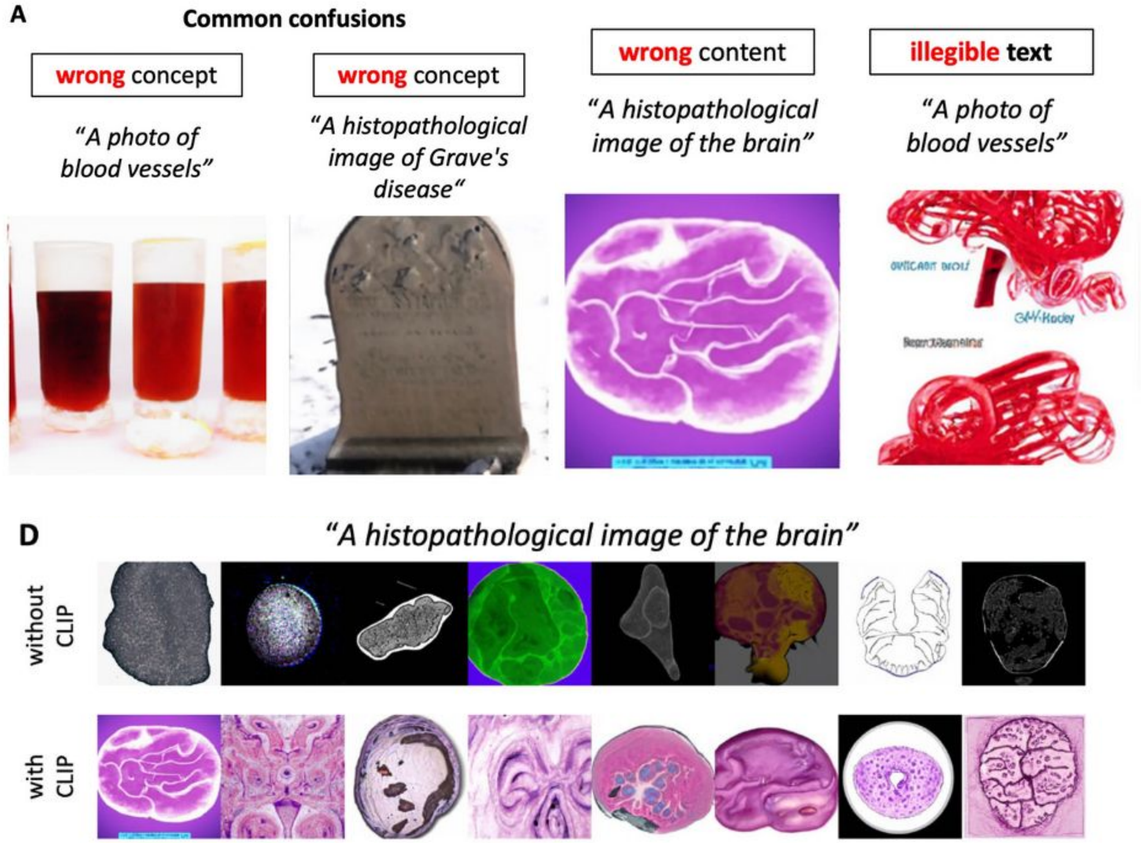
- GLIDE model: a diffusion model for the problem of text-conditional synthesis and compare two different guidance strategies: CLIP guidance and classifier-free guidance. *[December 2021]*
- Dataset: 250 million text-image from the internet incorporating Conceptual Captions, the text-image pairs from Wikipedia and a subset of YFCC100M.



Kather, J.N., Ghaffari Laleh, N., Foersch, S. *et al.* Medical domain knowledge in domain-agnostic generative AI. *npj Digit. Med.* **5,** 90 (2022)

Challenges in their use on the Medical Domain

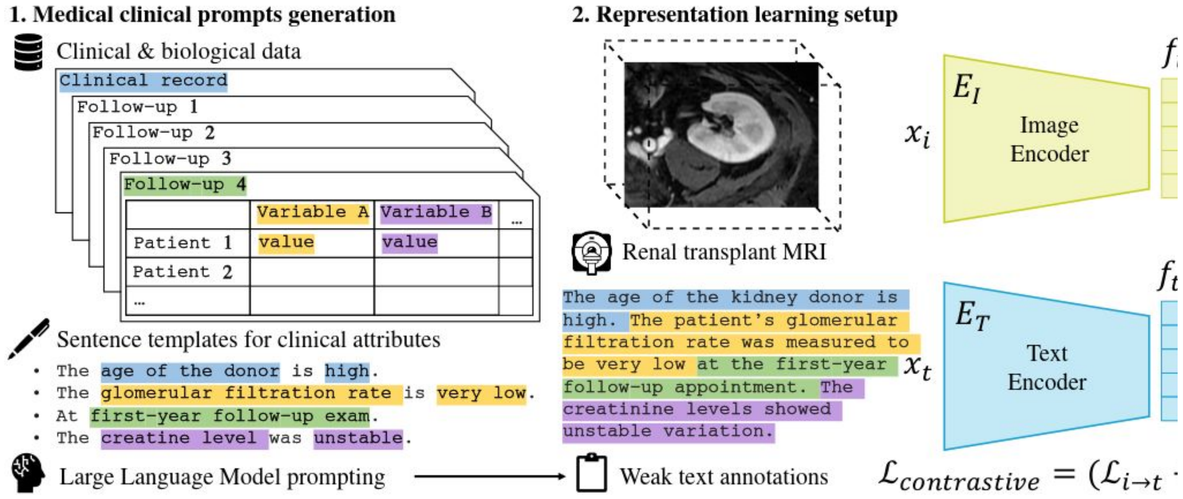
- These models do not generalize well on medical data and particular clinical questions
  - GLIDE performance evaluation:
    - high score for: histopathology, scientific posters and scientific illustration
    - poor score for: X-rays, MRI and CT
  - Similar models could be used for exploring biological mechanisms:
    - *A histology image of a patient who benefits from immunotherapy*
    - *An MRI image of a patient who should be treated with a statin*



Kather, J.N., Ghaffari Laleh, N., Foersch, S. *et al.* Medical domain knowledge in domain-agnostic generative AI. *npj Digit. Med.* **5,** 90 (2022)

Integrate LLMs on medical applications

- Explore pretrained models and their use on multimodal learning and integration of clinical attributes into text.
- Use contrastive learning to match medical imaging with text, enhancing explainability and boosting performance on prognosis of graft kidney transplant.

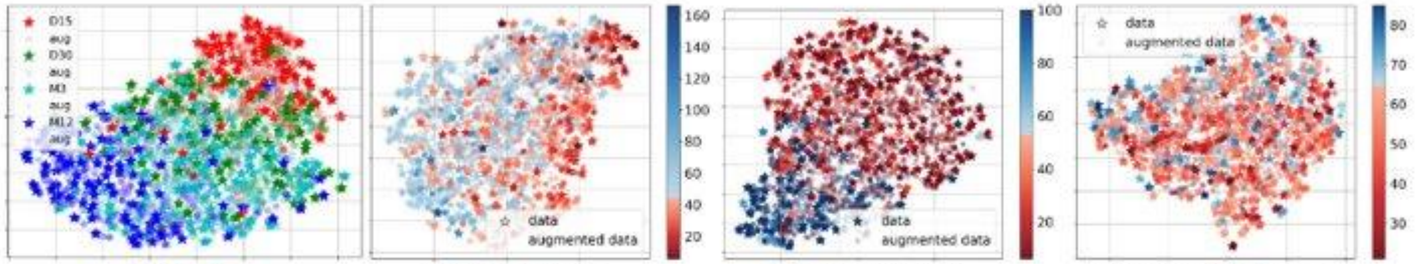


Method	Weak annotations GFR	Exam	Creat	D.A.	2 years AUC	2 years F1	3 years AUC	3 years F1	4 years AUC	4 years F1	Mean AUC	Mean F1
CLIP weights					62.6	73.7	52.5	78.1	51.3	54.6	55.5	68.8
CosEmbLoss	✓				76.2	86.4	77.8	70.6	67.0	77.3	73.6	78.1
CosEmbLoss				✓	75.5	81.1	75.6	68.8	66.1	78.1	72.4	76.0
CosEmbLoss++	✓			✓	84.4	88.9	82.5	**86.4**	73.9	85.7	80.3	87.0
CosEmbLoss++	✓		✓		81.6	87.8	71.3	85.1	71.3	90.2	74.7	87.7
CosEmbLoss++	✓		✓	✓	78.2	87.0	75.0	83.3	74.8	87.0	76.0	85.8
CosEmbLoss++	✓	✓	✓	✓	75.5	85.7	62.0	69.8	63.5	80.9	67.0	78.8
MEDIMP	✓				56.5	83.3	51.9	79.1	49.6	**90.2**	52.6	84.2
MEDIMP	✓	✓			81.0	**89.4**	81.9	80.0	74.8	84.4	79.2	84.6
MEDIMP	✓	✓		✓	76.9	73.2	**86.3**	85.7	74.8	**90.2**	79.3	83.0
MEDIMP	✓	✓	✓		72.8	86.4	71.9	81.0	71.3	71.8	72.0	79.7
MEDIMP	✓	✓	✓	✓	**85.0**	**89.4**	84.4	83.7	**75.7**	**90.2**	**81.7**	**87.8**

Milecki, L., Kalogeiton, V., Bodard, S., Anglicheau, D., Correas, J. M., Timsit, M. O., & Vakalopoulou, M. (2023). MEDIMP: Medical Images and Prompts for renal transplant representation learning. MIDL 2023

Integrate LLMs on medical applications

- This approach provides an elegant way to incorporate clinical or biological information into the learning process of feature extraction of medical imaging data.



(1) Exam (2) GFR (mL.min^{-1}) (3) Creat (µmol.L^{-1}) (4) Donor's Age (year)

Method	Weak annotations				2 years		3 years		4 years		Mean	
	GFR	Exam	Creat	D.A.	AUC	F1	AUC	F1	AUC	F1	AUC	F1
CLIP weights					62.6	73.7	52.5	78.1	51.3	54.6	55.5	68.8
CosEmbLoss	✓				76.2	86.4	77.8	70.6	67.0	77.3	73.6	78.1
CosEmbLoss				✓	75.5	81.1	75.6	68.8	66.1	78.1	72.4	76.0
CosEmbLoss++	✓			✓	84.4	88.9	82.5	**86.4**	73.9	85.7	80.3	87.0
CosEmbLoss++	✓		✓		81.6	87.8	71.3	85.1	71.3	90.2	74.7	87.7
CosEmbLoss++	✓		✓	✓	78.2	87.0	75.0	83.3	74.8	87.0	76.0	85.8
CosEmbLoss++	✓	✓	✓	✓	75.5	85.7	62.0	69.8	63.5	80.9	67.0	78.8
MEDIMP	✓				56.5	83.3	51.9	79.1	49.6	**90.2**	52.6	84.2
MEDIMP	✓	✓			81.0	**89.4**	81.9	80.0	74.8	84.4	79.2	84.6
MEDIMP	✓	✓		✓	76.9	73.2	**86.3**	85.7	74.8	**90.2**	79.3	83.0
MEDIMP	✓	✓	✓		72.8	86.4	71.9	81.0	71.3	71.8	72.0	79.7
MEDIMP	✓	✓	✓	✓	**85.0**	**89.4**	84.4	83.7	**75.7**	**90.2**	**81.7**	**87.8**

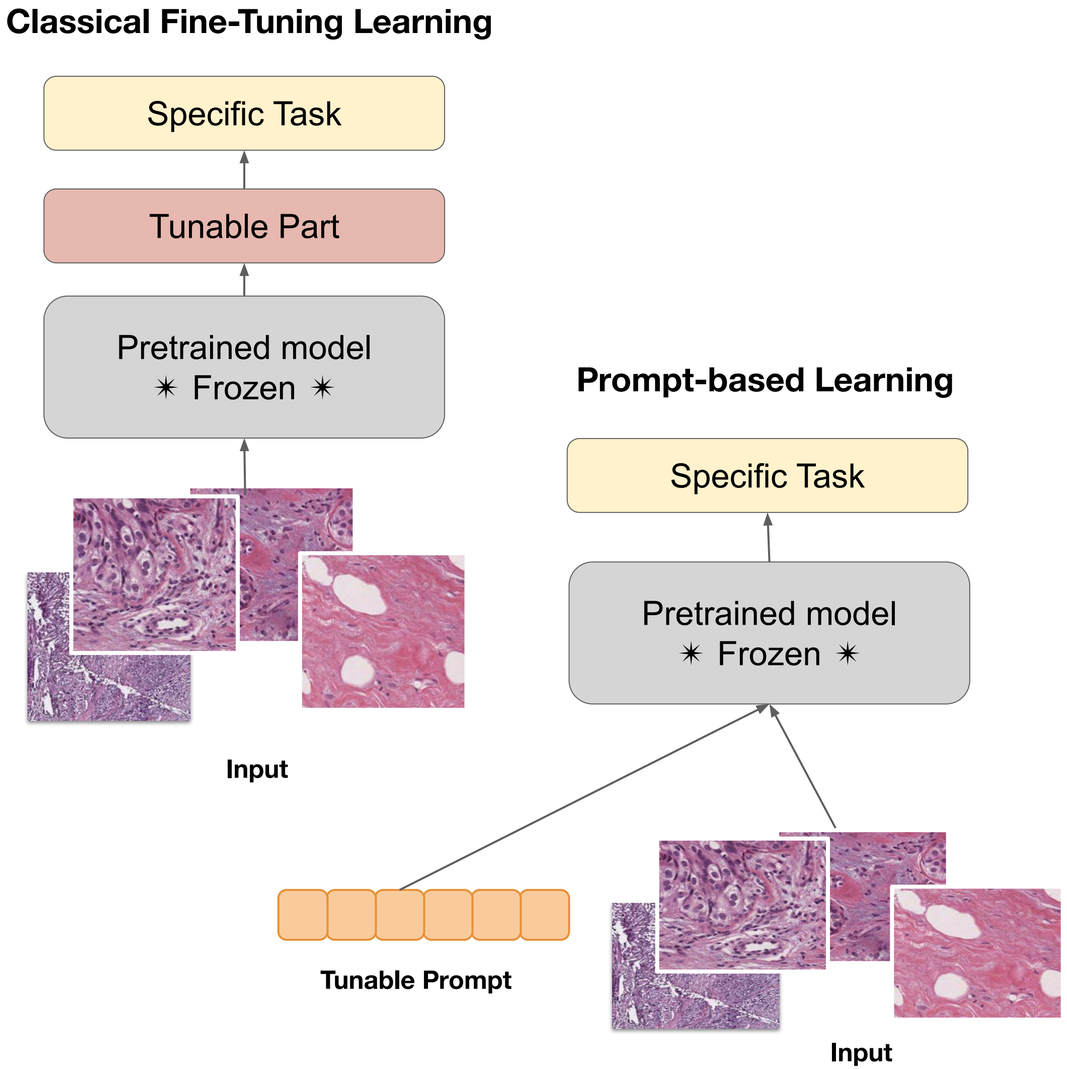
- MEDIMP augmentation perform better than other sentence augmentations

Method	2 years		3 years		4 years		Mean	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
MEDIMP	**85.0**	**89.4**	**84.4**	**83.7**	75.7	**90.2**	**81.7**	**87.8**
Manual	74.2	76.2	80.6	62.1	**80.0**	76.9	78.3	71.7
T5 WebNLG	74.8	85.7	78.8	83.3	74.8	85.7	76.1	84.9

Milecki, L., Kalogeiton, V., Bodard, S., Anglicheau, D., Correas, J. M., Timsit, M. O., & Vakalopoulou, M. (2023). MEDIMP: Medical Images and Prompts for renal transplant representation learning. MIDL 2023

Prompt-based learning

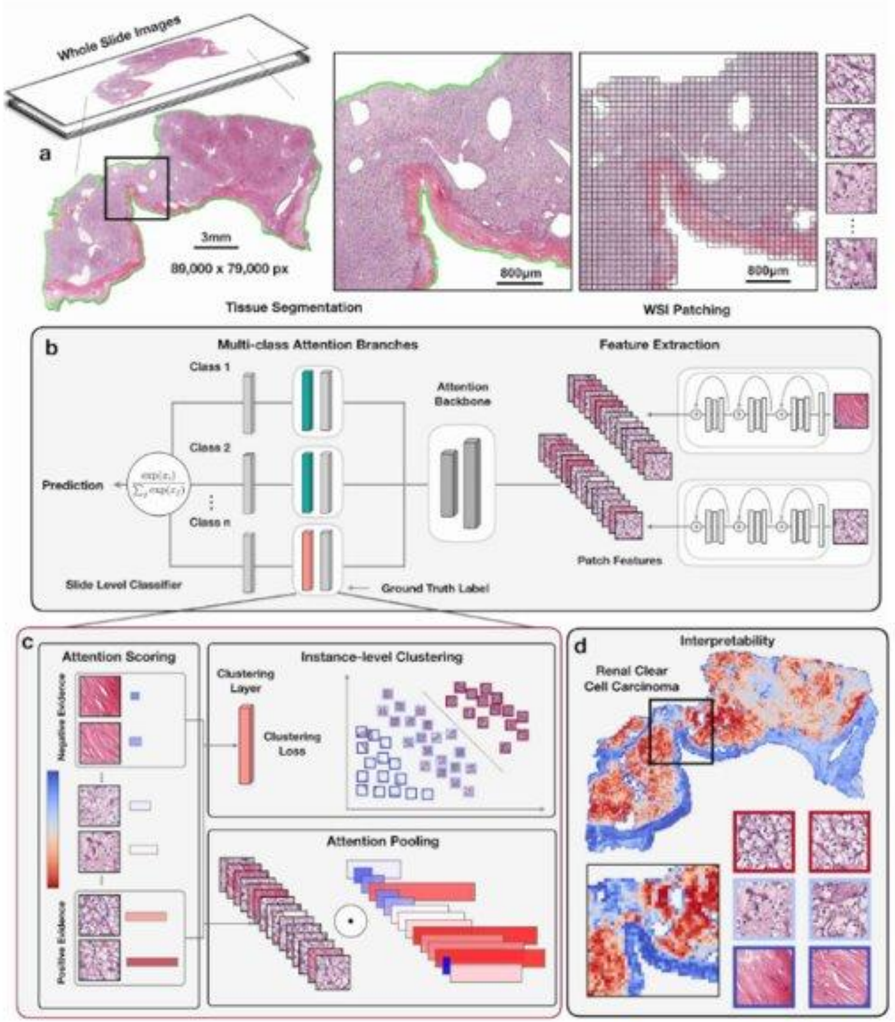
- Instead of traditional approaches that use pre-trained models and fine-tune them on specific tasks, prompts can be used.
- Prompt-based learning or prompting, started from the NLP community one year ago as an instruction by the user for the model to execute or complete.
- Several advantages with respect to traditional methods
  - Data-efficient *[also achieved with few-shot and zero-shot learning]*
  - Parameter-efficient *[also achieved with tuning]*



Classical WSI-level MIL methods

- Whole slide image (WSI) classification is a critical task in computational pathology.
- Multi-instance learning (MIL) methods is currently widely used in computationally pathology
- Clustering-constrained attention multiple instance learning (CLAM)
  - Use of a pre-trained ResNet50 network per patch.

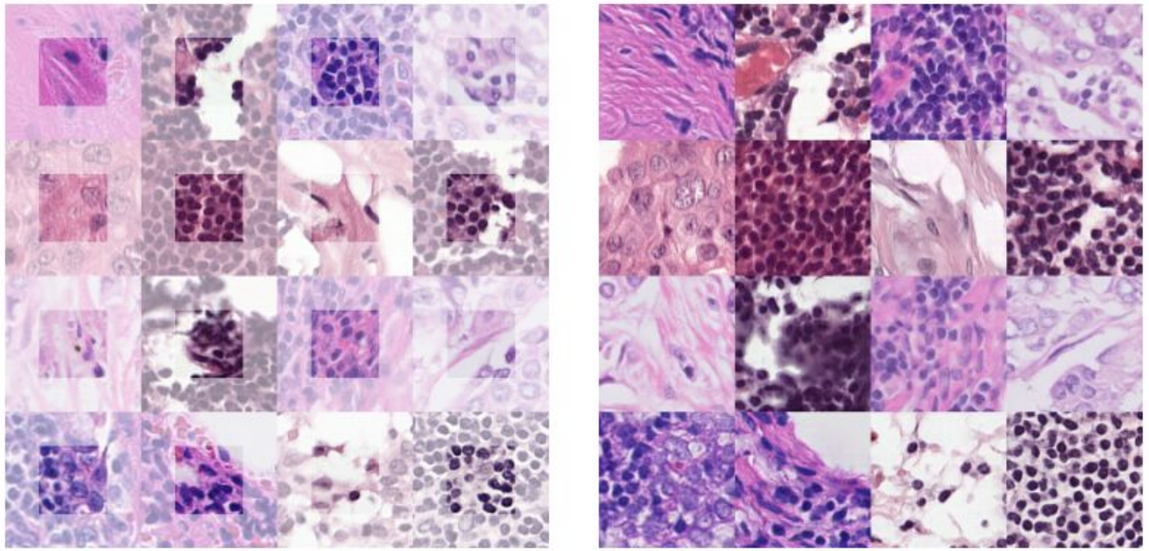
- However, the representation of the patch are *task and data agnostic*



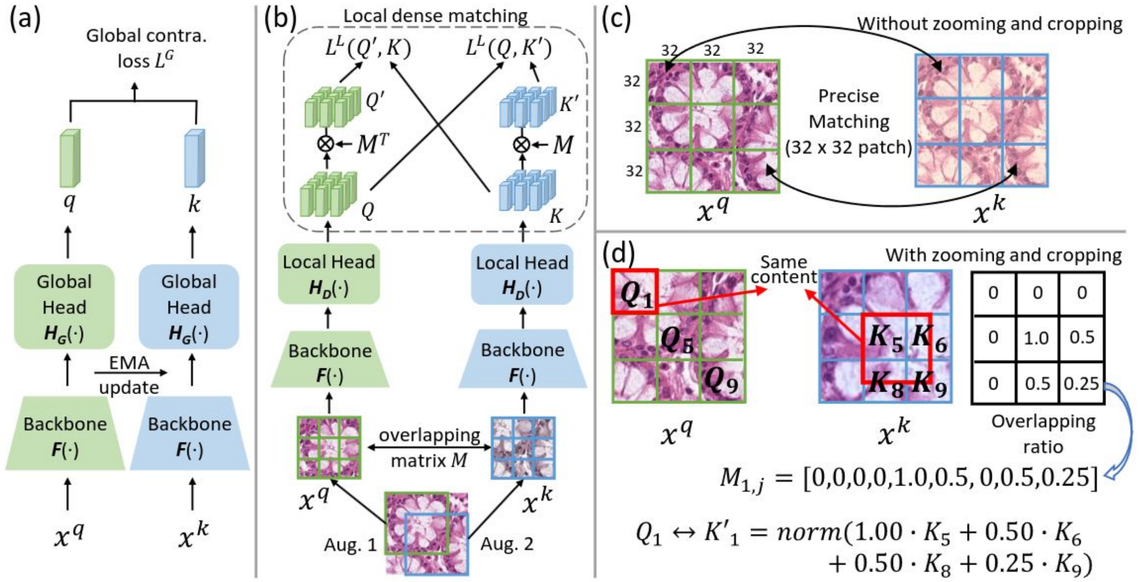
Lu, et al. "Data Efficient and Weakly Supervised Computationally Pathology on Whole Slide Images."Nature Biomedical Engineering, 2021.

Data-Specific Representations

- Existing MIL methods do not fine-tune their features together with the classification task due to GPU memory limitations.
- Huge large models are not easy to be trained on histopathology data, due to model complexity and limited data - *current works use mainly self-supervised tasks for the pretraining of such models* -



Boyd, J., Liashuha, M., Deutsch, E., Paragios, N., Christodoulidis, S., & Vakalopoulou, M. (2021). Self-supervised representation learning using visual field expansion on digital pathology. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 639-647).



Zhang, J., Kapse, S., Ma, K., Prasanna, P., Vakalopoulou, M., Saltz, J., & Samaras, D. (2023). Precise Location Matching Improves Dense Contrastive Learning in Digital Pathology. IPMI 2023.

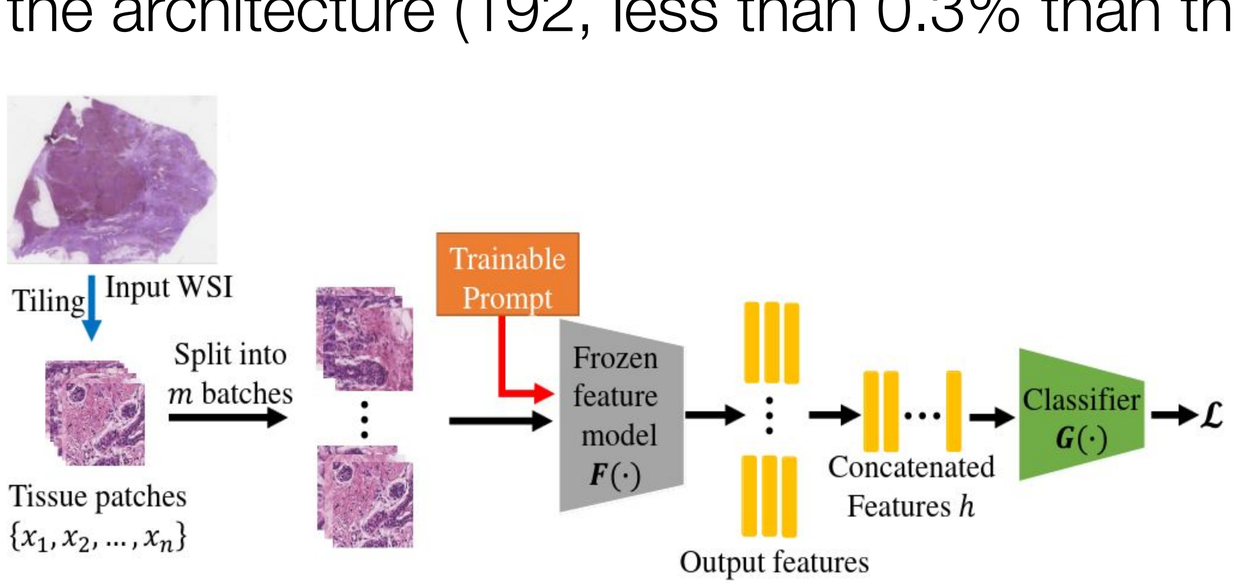
Task-specific Prompt for MIL Schemes

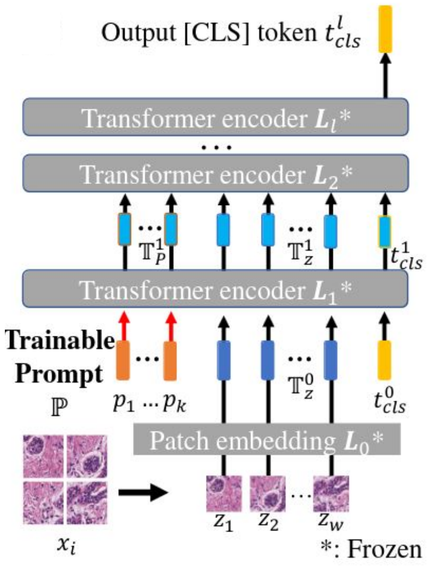
- Existing MIL methods do not fine-tune their features together with the classification task due to GPU memory limitations.
- Huge large models are not easy to be trained on histopathology data, due to model complexity and limited data - *current works use mainly self-supervised tasks for the pretraining of such models* -
- Prompt-MIL: Calibration of the pre-trained on task-specific information using only a small fraction of tunable parameters rather than conventional fine-tuning.
  - Train only the prompt and downstream network without re-training the large backbone
  - Better performances on limited labeled data - ideal for computational pathology

Zhang, J., Kapse, S., Ma, K., Prasanna, P., Saltz, J., Vakalopoulou, M., & Samaras, D. (2023). Prompt-MIL: Boosting Multi-Instance Learning Schemes via Task-specific Prompt Tuning. arXiv preprint arXiv:2303.12214.

Task-specific Prompt for MIL Schemes

- Prompt-MIL framework
  - a frozen feature model F(.)
  - a classifier to perform the task (G(.))
  - a trainable prompt P(.)
- The number of the parameters is negligible with respect to the total amount of parameters of the architecture (192, less than 0.3% than the total parameters).





Zhang, J., Kapse, S., Ma, K., Prasanna, P., Saltz, J., Vakalopoulou, M., & Samaras, D. (2023). Prompt-MIL: Boosting Multi-Instance Learning Schemes via Task-specific Prompt Tuning. arXiv preprint arXiv:2303.12214.

Experimental Setting - Datasets

- Experiments in 3 public available WSI H&E datasets
  - Subtyping in TCGA-BRCA: 1034 diagnostic digital slides of two breast cancer subtyping: invasive ductal carcinoma (IDC) and invasive lobutal carcinoma (ILC)
  - Classification in TCGA-CRC: 430 diagnostic digital slides of colorectal cancer for classification in chromosomal instability (CIN) or genome stable (GS)
  - Classification in BRIGHT: 503 diagnostic slides for breast tissue for classification in non-cancerous, pre-cancerous and cancerous sides

- Splitting of the datasets in training/ validation/ testing using either the official splitting or using the schemes presented in bibliography

Zhang, J., Kapse, S., Ma, K., Prasanna, P., Saltz, J., Vakalopoulou, M., & Samaras, D. (2023). Prompt-MIL: Boosting Multi-Instance Learning Schemes via Task-specific Prompt Tuning. arXiv preprint arXiv:2303.12214.

Experimental Setting - Quantitative

- Our prompt-MIL outperform in all our experiments conventional MIL approaches as well as full fine-tuning using also much less parameters and GPU memory

Dataset	TCGA-BRCA		TCGA-CRC		BRIGHT		Num. of
Metric	Accuracy	AUROC	Accuracy	AUROC	Accuracy	AUROC	Parameters
Conventional MIL	92.10	96.65	73.02	69.24	62.08	80.96	70k
Full fine-tuning	88.14	93.78	74.53	56.63	56.13	75.87	5.6M
Prompt-MIL (ours)	**93.47**	**96.89**	**75.47**	**75.45**	**64.58**	**81.31**	70k+192

- Increase performance when using SSL-data specific representations trained on TCGA Pan-cancer dataset

Dataset	TCGA-BRCA		BRIGHT	
Metric	Accuracy	AUROC	Accuracy	AUROC
ViT-small [27]	91.75	97.03	54.17	76.76
ViT-small w/ Prompt-MIL	**92.78**	**97.53**	**57.50**	**78.29**

Zhang, J., Kapse, S., Ma, K., Prasanna, P., Saltz, J., Vakalopoulou, M., & Samaras, D. (2023). Prompt-MIL: Boosting Multi-Instance Learning Schemes via Task-specific Prompt Tuning. arXiv preprint arXiv:2303.12214.

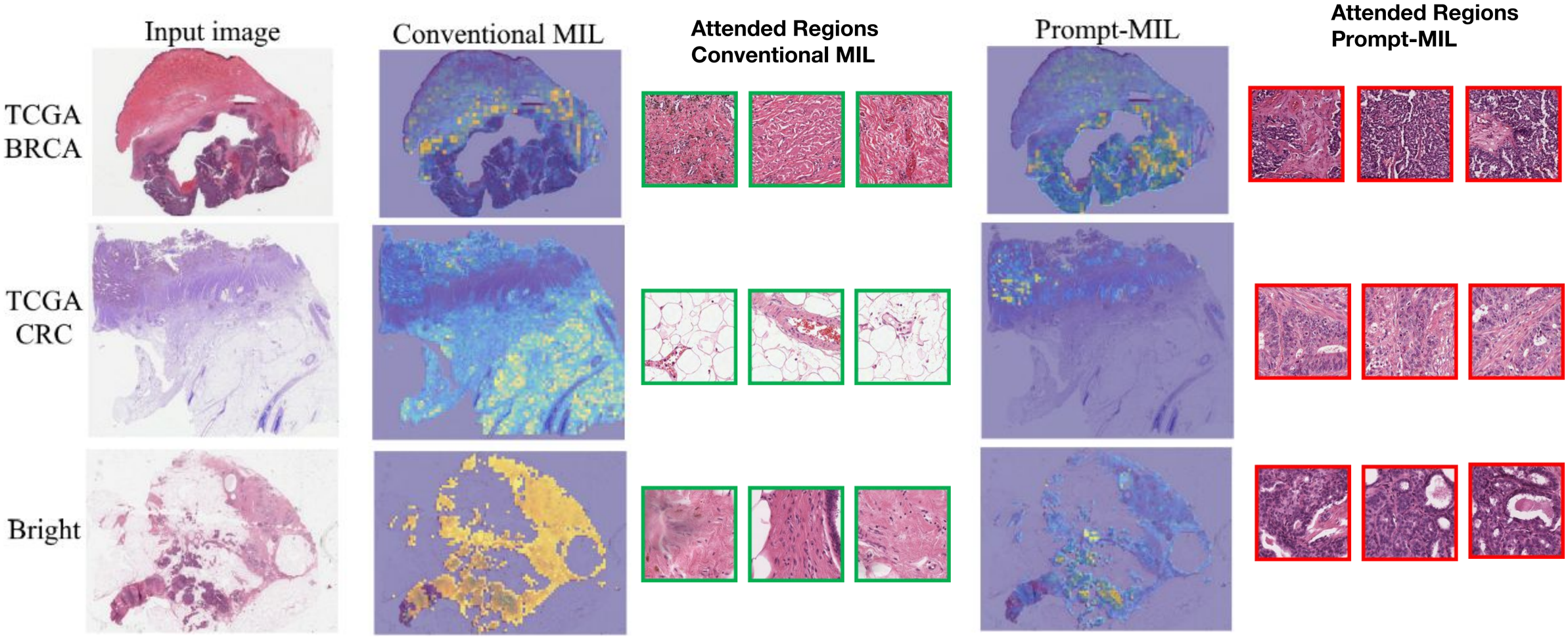
Experimental Setting - Quantitative

- Increase on performance when adapted on different MIL schemes

MIL scheme	Method	TCGA-CRC		BRIGHT		Num. of Parameters
		Accuracy	AUROC	Accuracy	AUROC	
DSMIL	Conventional MIL	73.02	69.24	62.08	80.96	64k
	Fine-tuning L_{12}	69.81	70.72	61.25	80.10	509k
	Prompt-MIL (ours)	**75.47**	**75.45**	**64.58**	**81.31**	64k+192
ABMIL	Conventional MIL	74.10	68.56	61.25	**80.35**	25k
	Fine-tuning L_{12}	70.37	**70.78**	**62.50**	78.92	470k
	Prompt-MIL (ours)	**75.87**	**70.10**	**62.50**	79.30	25k+192
CLAM	Conventional MIL	75.87	77.50	62.08	82.97	59k
	Fine-tuning L_{12}	74.07	71.40	63.33	81.32	504k
	Prompt-MIL (ours)	**76.19**	**80.84**	**64.17**	**84.31**	59k+192

Zhang, J., Kapse, S., Ma, K., Prasanna, P., Saltz, J., Vakalopoulou, M., & Samaras, D. (2023). Prompt-MIL: Boosting Multi-Instance Learning Schemes via Task-specific Prompt Tuning. arXiv preprint arXiv:2303.12214.

Experimental Setting - Qualitative



Zhang, J., Kapse, S., Ma, K., Prasanna, P., Saltz, J., Vakalopoulou, M., & Samaras, D. (2023). Prompt-MIL: Boosting Multi-Instance Learning Schemes via Task-specific Prompt Tuning. arXiv preprint arXiv:2303.12214.

Conclusions

- Deep Learning will play a key role the next years in Precision Medicine
- Very active research area
- A number of challenges should be still addressed both from the engineering and medical society
- The collection of data is very important and define the use and selection of the algorithms
- Interpretability is very important for medical problems
- Fusion of models will be very important for clinical endpoints