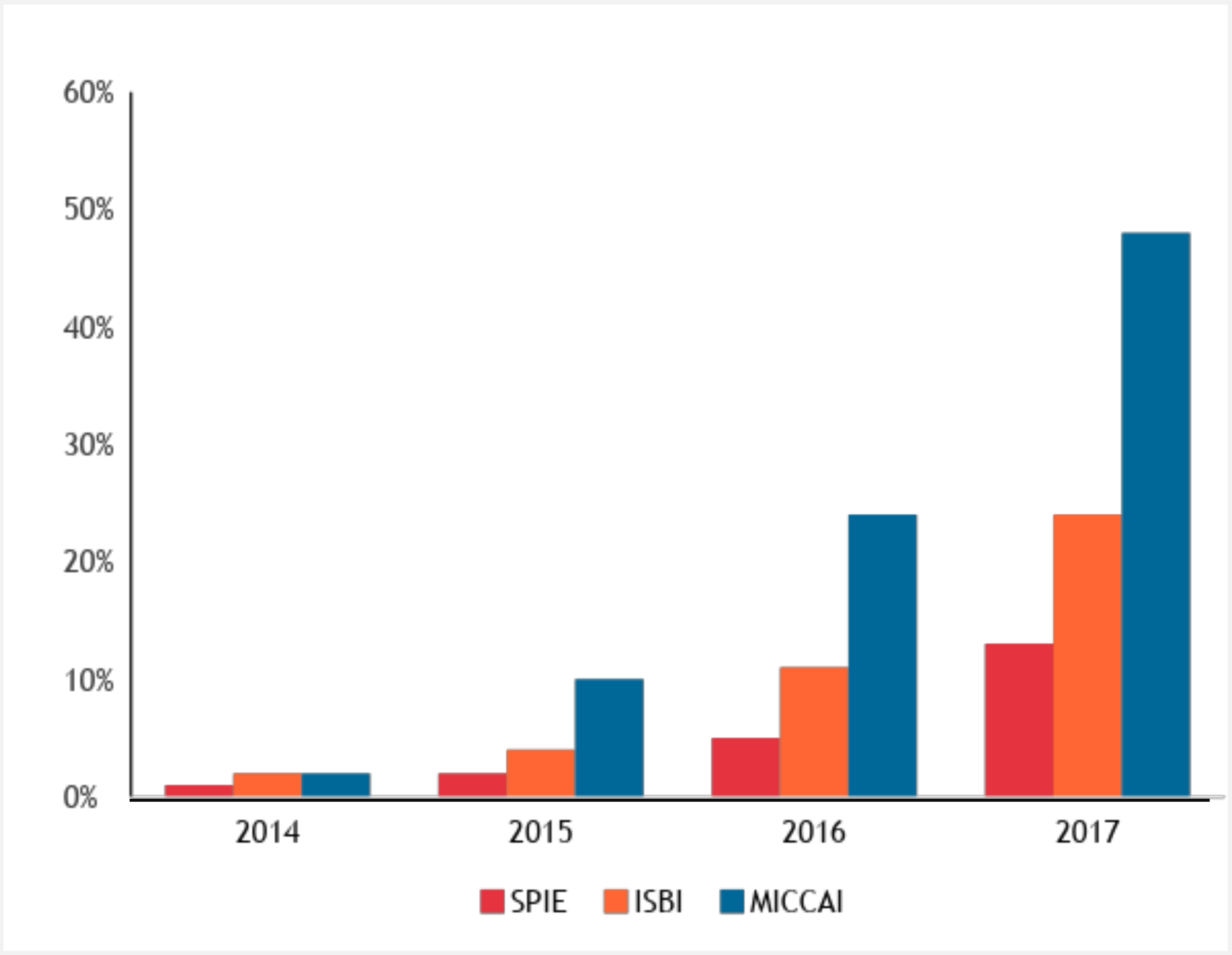
VOLUMETRIC MEDICAL IMAGE PROCESSING WITH DEEP LEARNING

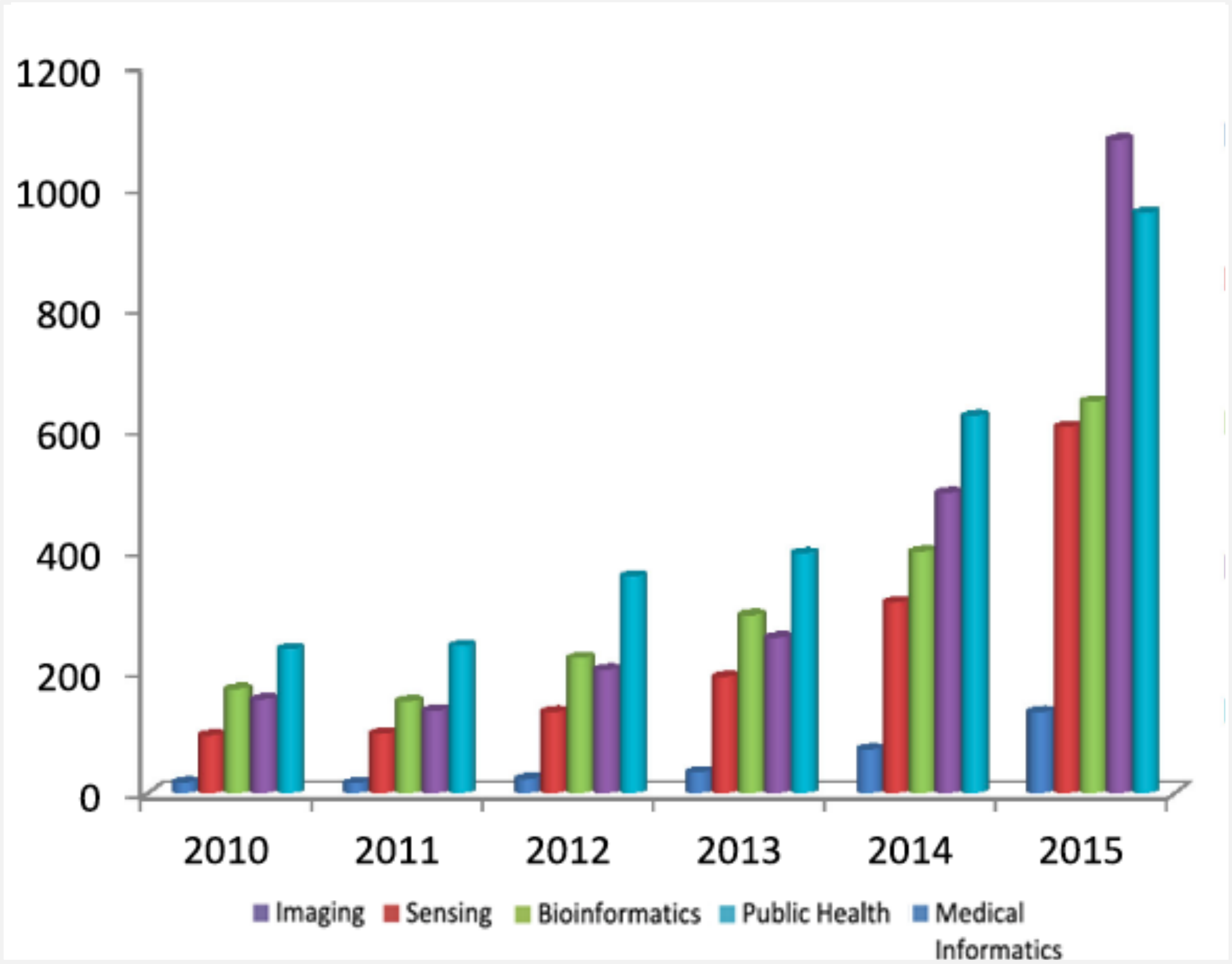
A presentation by Fausto Milletari, Ph.D. @faustomilletari

DEEP LEARNING MOMENTUM BUILDING

MEDICAL IMAGING PAPERS USING DEEP LEARNING



AI ACROSS HEALTHCARE ACADEMIC PUBS



Slide credits: NVIDIA

Deep Learning for Health Informatics - Daniele Ravi, et. al., IEEE Journal of Biomedical and Health Informatics, Vol. 21, No. 1, January 2017

DEEP LEARNING BEYOND RESEARCH

Increasing impact on patient lives and clinical practice

Another AI-powered device gets the FDA's blessing

In an ongoing effort to get more AI into healthcare, the FDA just approved the marketing of an algorithm that detects wrist fractures.

MIT technology review

TOM SIMONITE BUSINESS 02.28.18 07:00 AM

USING AI TO HELP STROKE VICTIMS WHEN 'TIME IS BRAIN'

Wired

Zebra Medical wins CE Mark for brain-bleed detection algorithm

MARCH 6, 2018 BY FINK DENSFORD — LEAVE A COMMENT

Mass+device

Google's AI Uses Retinal Images to Reveal Cardiovascular Risk Factors

Medscape

Ricki Lewis, PhD
February 28, 2018

JANUARY 25, 2017

Deep learning algorithm does as well as dermatologists in identifying skin cancer

In hopes of creating better access to medical care, Stanford researchers have trained an algorithm to diagnose skin cancer.

Stanford

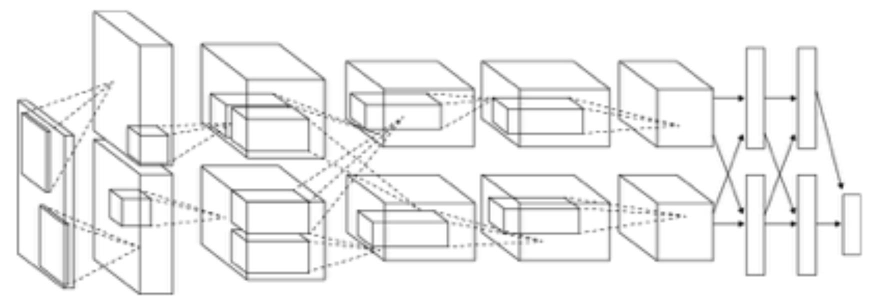
First FDA Approval For Clinical Cloud-Based Deep Learning In Healthcare

Forbes

CAMBRIAN EXPLOSION

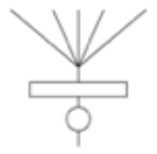
Of models and methods in Deep Learning

Convolutional Networks

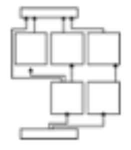


Encoder/Decoder

ReLu

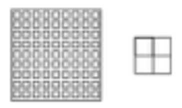


BatchNorm



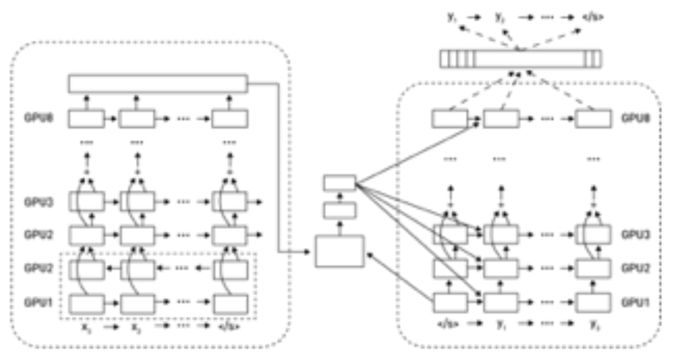
Concat

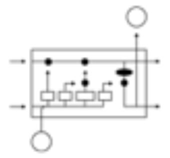
Dropout



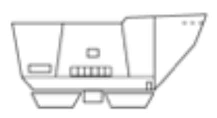
Pooling

Recurrent Networks

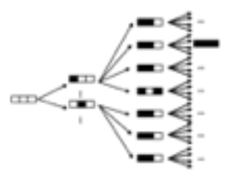




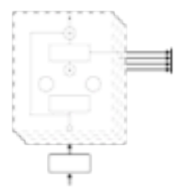
LSTM



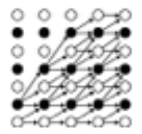
GRU



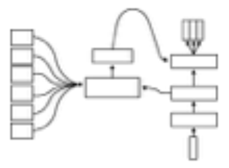
Beam Search



WaveNet

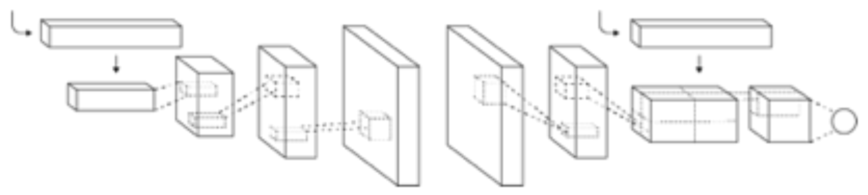


CTC

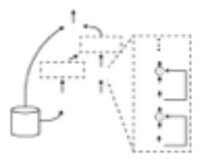


Attention

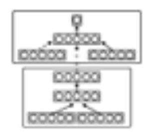
GAN



3D-GAN

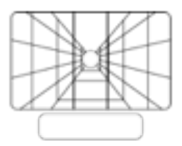


MedGAN



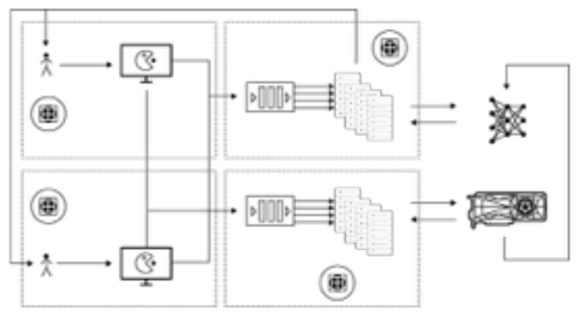
Conditional GAN

Coupled GAN

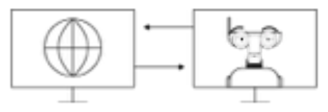


Speech Enhancement GAN

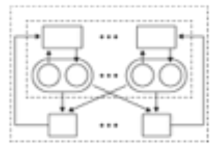
Reinforcement Learning



DQN

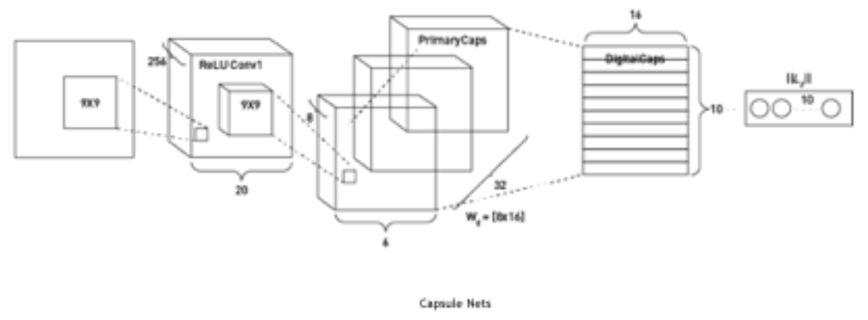


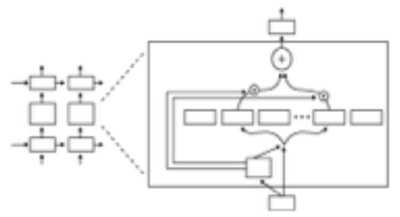
Simulation



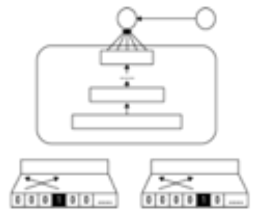
DDPG

New Species





Mixture of Experts



Neural Collaborative Filtering

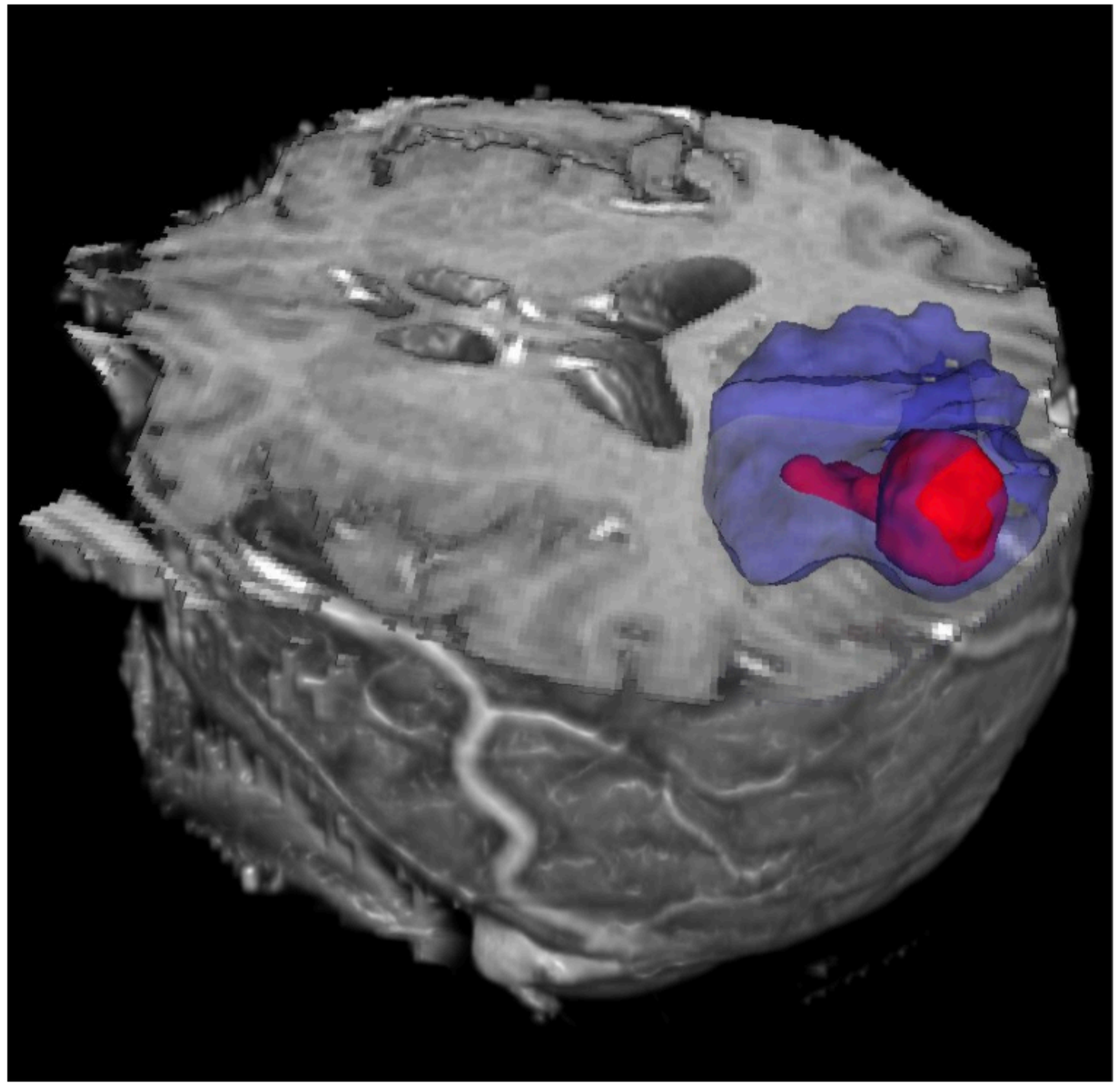
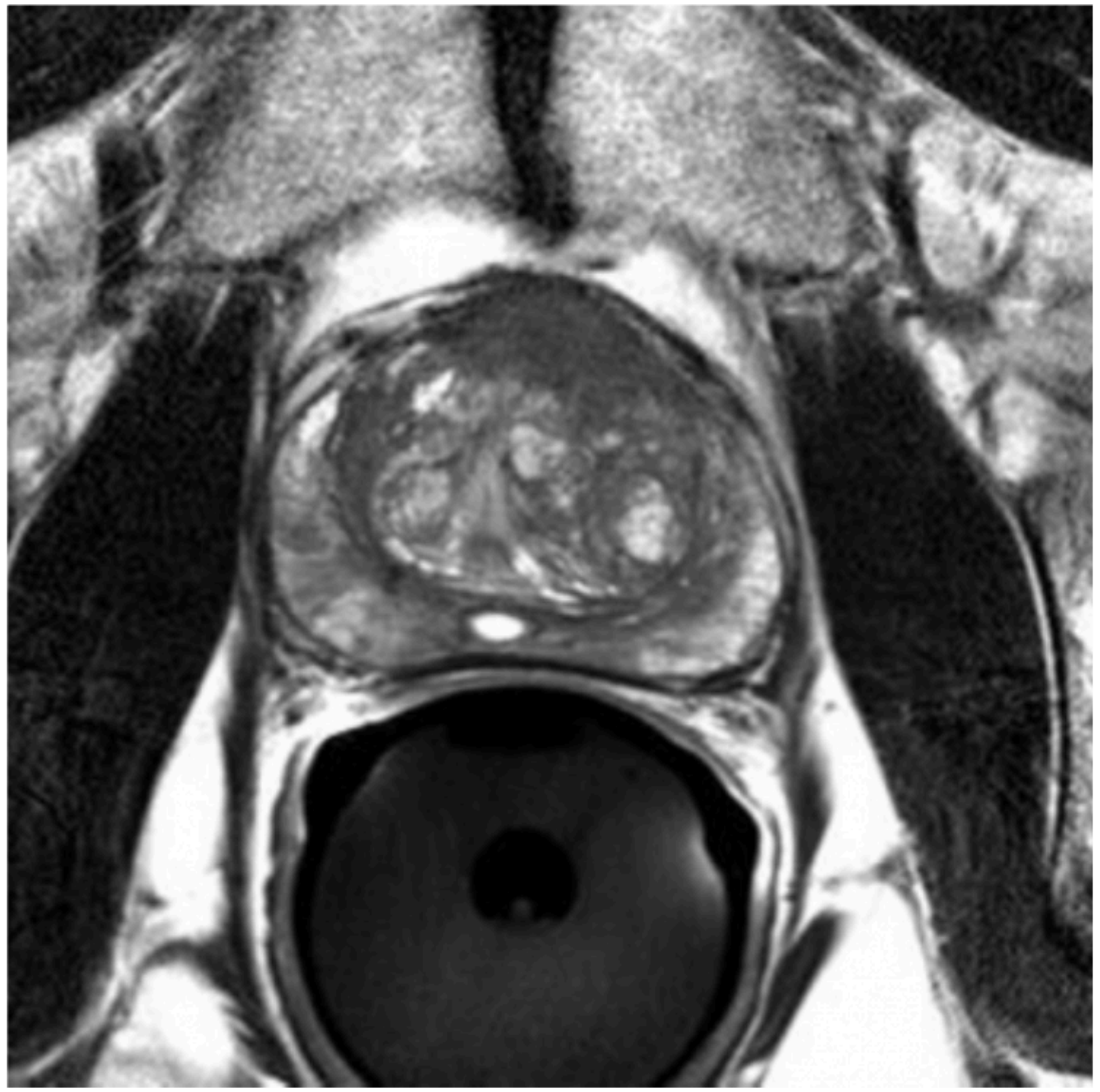
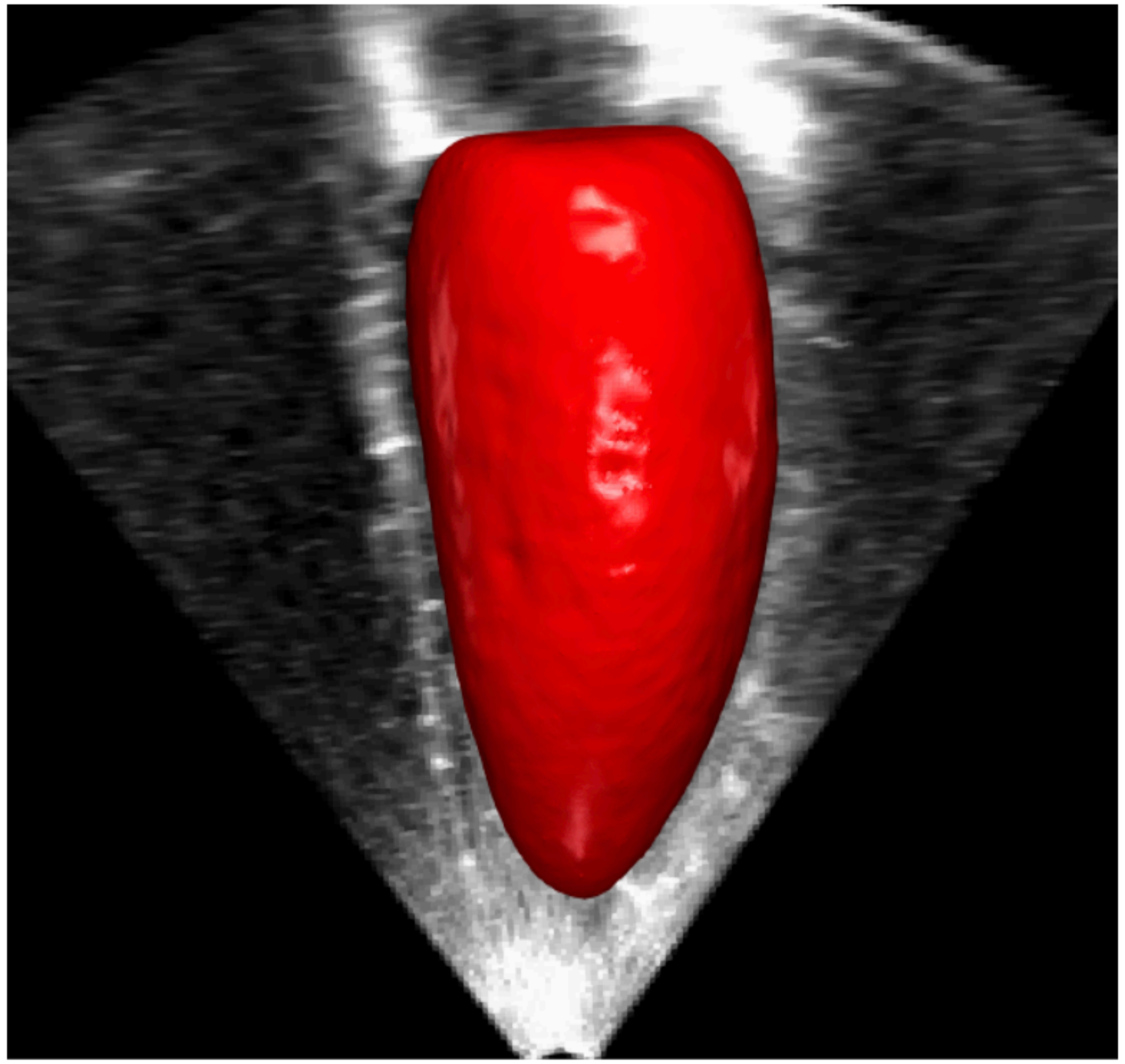
Block Sparse LSTM

FULLY CONVOLUTIONAL NEURAL NETWORKS

As computing power increases and GPU memory constraints disappear, fully convolutional methods capable of processing high resolution 3D images appear as viable alternatives delivering even better results than older approaches.

VOLUMETRIC IMAGE SEGMENTATION

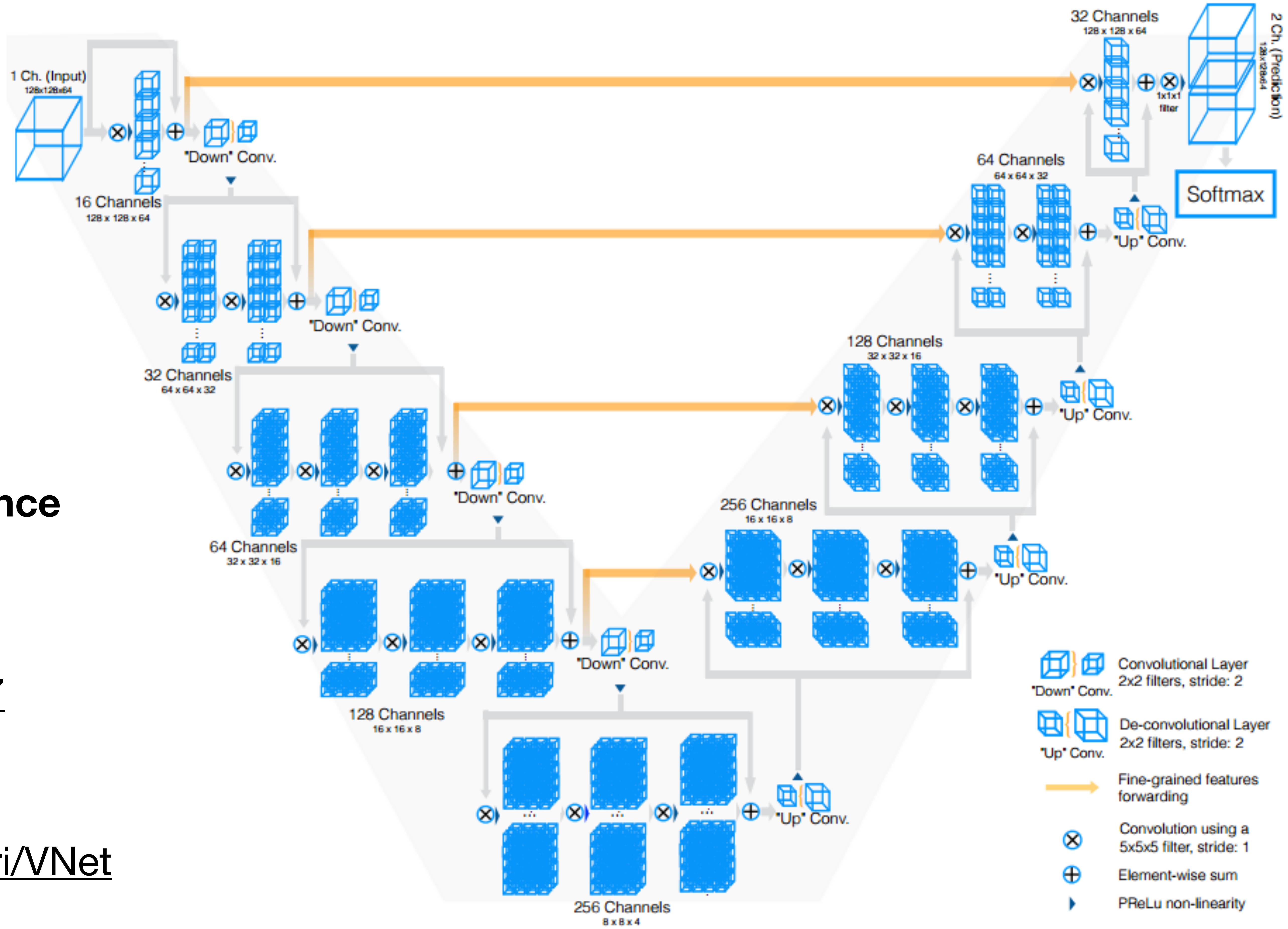
With fully convolutional neural networks (FCNNs)

Many medical modalities are 3D. This is the case for MRI, CT, PET, Ultrasound, etc. Tomographic medical data introduces new computational challenges but also new opportunities in medical image analysis

V-NET FOR 3D SEGMENTATION (1/7)

Paving the road for FCNN-based volumetric segmentation



More than 3188 citations in 4 years
Most cited paper of whole 3DV conference
Most cited paper of Nassir Navab lab

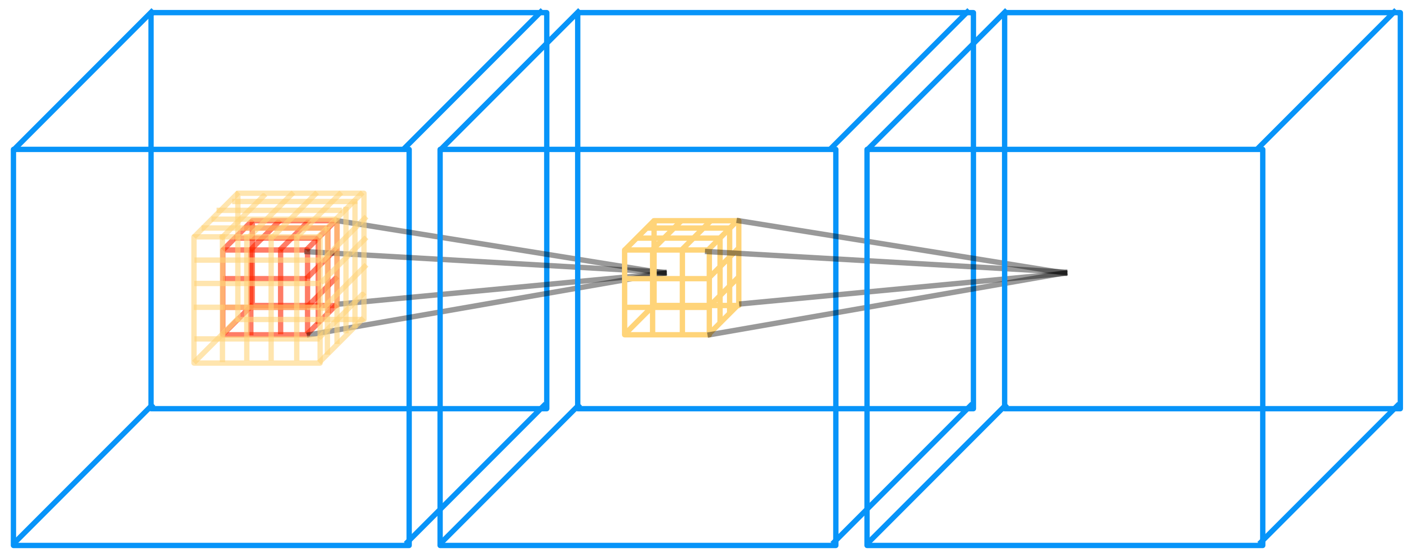
https://arxiv.org/abs/1606.04797

https://github.com/faustomilletari/VNet

Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on* (pp. 565-571). IEEE.
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

V-NET FOR 3D SEGMENTATION (2/7)

Paving the road for FCNN-based volumetric segmentation



- Introduces **dice loss**. Solves class imbalance problem
- V-Net uses 3D convolutions with a 3D receptive field
- Total V-Net receptive field is very high and exceeds input size
- Use padding for convolutions: prediction has same size as input and we don't use tiling
- More GPU memory translates to better performances by simple increasing input resolution
- Short skip connections for training speed. Long skip connection for fine grained details

Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on* (pp. 565-571). IEEE.

V-NET FOR 3D SEGMENTATION (3/7)

Details and lessons learned from experience with 3D FCNN

- Dice loss has two valid formulations leading to the same result but different gradient behavior [1]

$$D = \boxed{\frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}} = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i} \quad s.t.\ p_i \in \{0,1\} \wedge g_i \in \{0,1\}$$

$$\frac{\partial D_1}{\partial p_j} = \frac{\partial}{\partial p_j}\left(\frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}\right) = 2\left[\frac{g_j\left(\sum_i^N \boxed{p_i^2} + \sum_i^N \boxed{g_i^2}\right) - \boxed{2p_j}\sum_i^N p_i g_i}{\left(\sum_i^N \boxed{p_i^2} + \sum_i^N \boxed{g_i^2}\right)^2}\right]$$

$$\frac{\partial D_2}{\partial p_j} = \frac{\partial}{\partial p_j}\left(\frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i}\right) = 2\left[\frac{g_j\left(\sum_i^N p_i + \sum_i^N g_i\right) - \sum_i^N p_i g_i}{\left(\sum_i^N p_i + \sum_i^N g_i\right)^2}\right]$$

- Concatenation of long skip-connection can be replaced by sums. Better result reported in [2]
- BatchNorm is not a proper normalization technique for V-Net (small batch size) use group norm [3]
- Dice loss can be extended to multi-class Dice loss but there are multiple possible formulations [4]
- Variations of V-Net architecture with dense connections or even different shapes deliver slightly better results

[1] Milletari, Fausto. *Hough Voting Strategies for Segmentation, Detection and Tracking.* Diss. Universität München, 2018.
[2] Drozdzal, Michal, et al. "The importance of skip connections in biomedical image segmentation." *Deep Learning and Data Labeling for Medical Applications*. Springer, Cham, 2016. 179-187.
[3] Wu, Yuxin, and Kaiming He. "Group normalization." *arXiv preprint arXiv:1803.08494* (2018).
[4] Gibson, Eli, et al. "NiftyNet: a deep-learning platform for medical imaging." *arXiv preprint arXiv:1709.03485* (2017).

V-NET FOR 3D SEGMENTATION (4/7)

Details and lessons learned from experience with 3D FCNN

- Dice loss has two valid formulations leading to the same result but different gradient behavior [1]

$$D = \boxed{\frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}} = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i} \quad s.t.\ p_i \in \{0,1\} \wedge g_i \in \{0,1\}$$

$$\left.\frac{\partial D_1}{\partial p_j}\right|_{P=G} = 2\left[\frac{p_j\left(\sum_i^N p_i^2 + \sum_i^N p_i^2\right) - 2p_j\sum_i^N p_i^2}{\left(\sum_i^N p_i^2 + \sum_i^N p_i^2\right)^2}\right] = 2\left[\frac{2p_jK - 2p_jK}{(2K)^2}\right] = 0.$$

$$\left.\frac{\partial D_2}{\partial p_j}\right|_{P=G} = 2\left[\frac{p_j\left(\sum_i^N p_i + \sum_i^N p_i\right) - \sum_i^N p_i^2}{\left(\sum_i^N p_i + \sum_i^N p_i\right)^2}\right] = \frac{(2p_j - 1)}{2K} \neq 0$$

- Concatenation of long skip-connection can be replaced by sums. Better result reported in [2]
- BatchNorm is not a proper normalization technique for V-Net (small batch size) use group norm [3]
- Dice loss can be extended to multi-class Dice loss but there are multiple possible formulations [4]
- Variations of V-Net architecture with dense connections or even different shapes deliver slightly better results

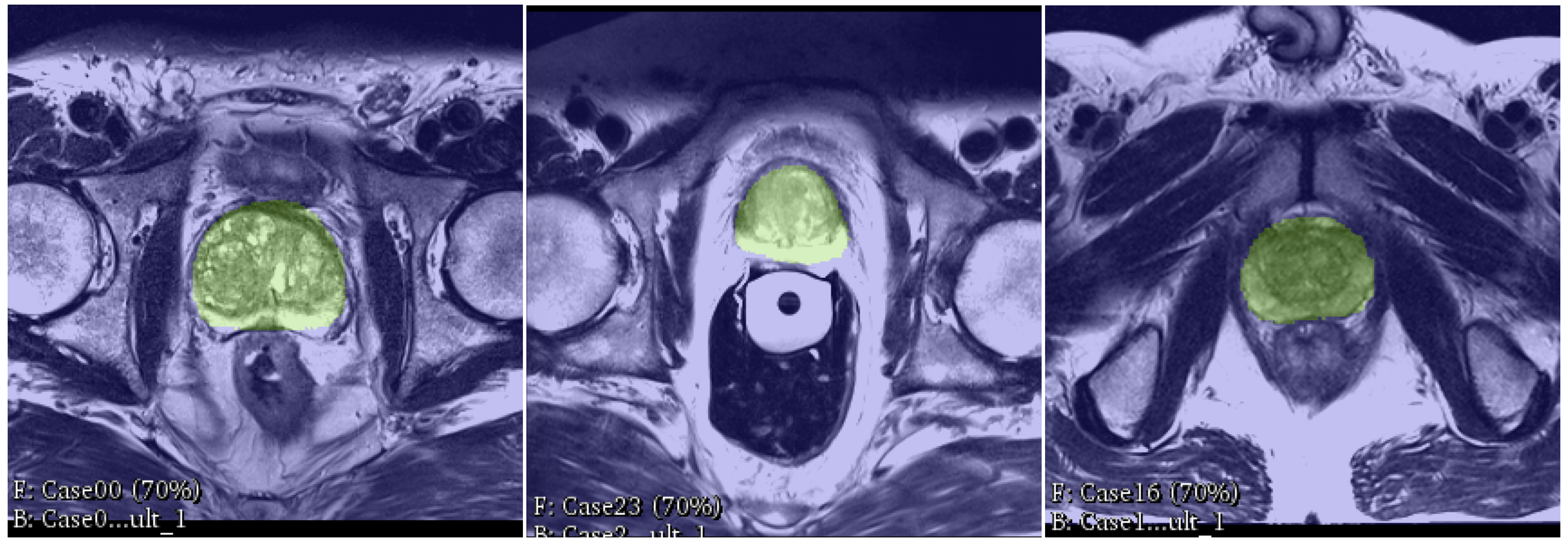
[1] Milletari, Fausto. *Hough Voting Strategies for Segmentation, Detection and Tracking.* Diss. Universität München, 2018.
[2] Drozdzal, Michal, et al. "The importance of skip connections in biomedical image segmentation." *Deep Learning and Data Labeling for Medical Applications*. Springer, Cham, 2016. 179-187.
[3] Wu, Yuxin, and Kaiming He. "Group normalization." *arXiv preprint arXiv:1803.08494* (2018).
[4] Gibson, Eli, et al. "NiftyNet: a deep-learning platform for medical imaging." *arXiv preprint arXiv:1709.03485* (2017).

V-NET FOR 3D SEGMENTATION (5/7)

Experimental evaluation

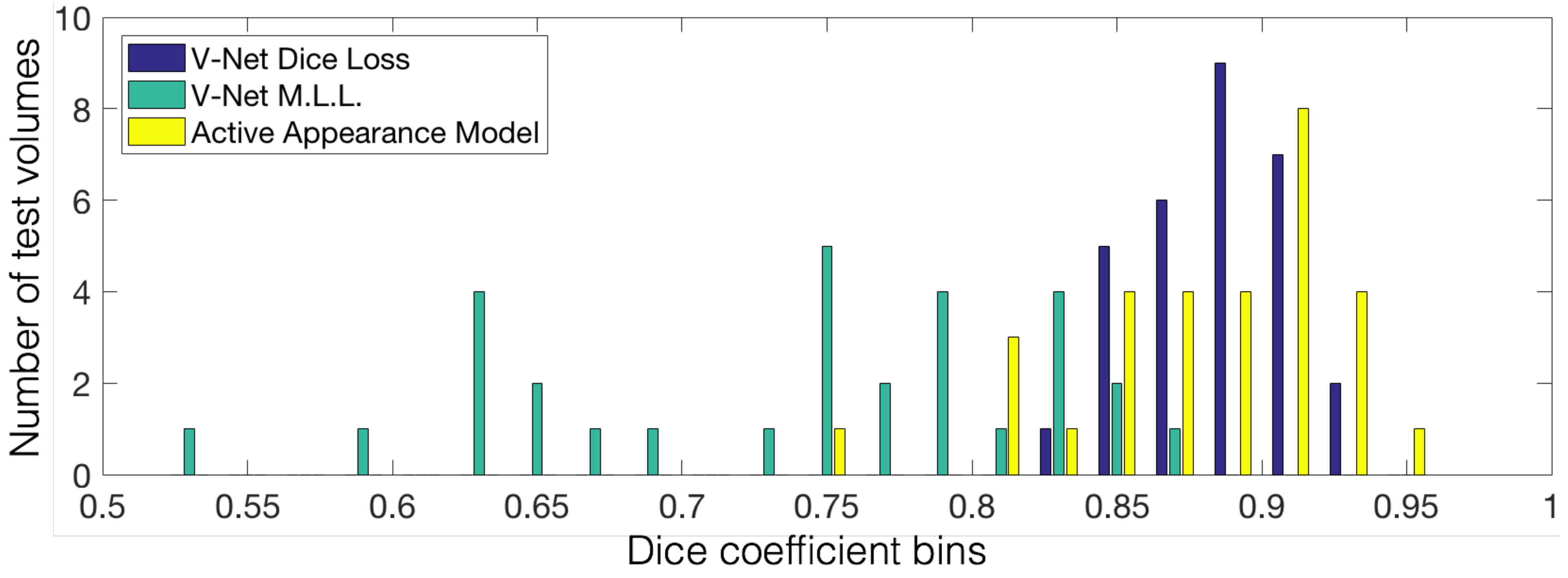
Prostate segmentation

- Promise 2012 MRI dataset
- Limited training data
- Augmentation with deformations, translations and histogram matching



Axial

Algorithm	Avg. Dice	Avg. Hausdorff dist	Score Promise	Speed
V-Net + Dice-based loss	0.869 ± 0.033	5.71 ± 1.20 mm	82.39	1 sec.
V-Net + mult. logistic loss	0.739 ± 0.088	10.55 ± 5.38 mm	63.30	1 sec.
Imorphics [161]	0.879 ± 0.044	5.935 ± 2.14 mm	84.36	8 min.
ScrAutoProstate	0.874 ± 0.036	5.58 ± 1.49 mm	83.49	1 sec.
SBIA	0.835 ± 0.055	7.73 ± 2.68 mm	78.33	–
Grislies	0.834 ± 0.082	7.90 ± 3.82 mm	77.55	7 min.

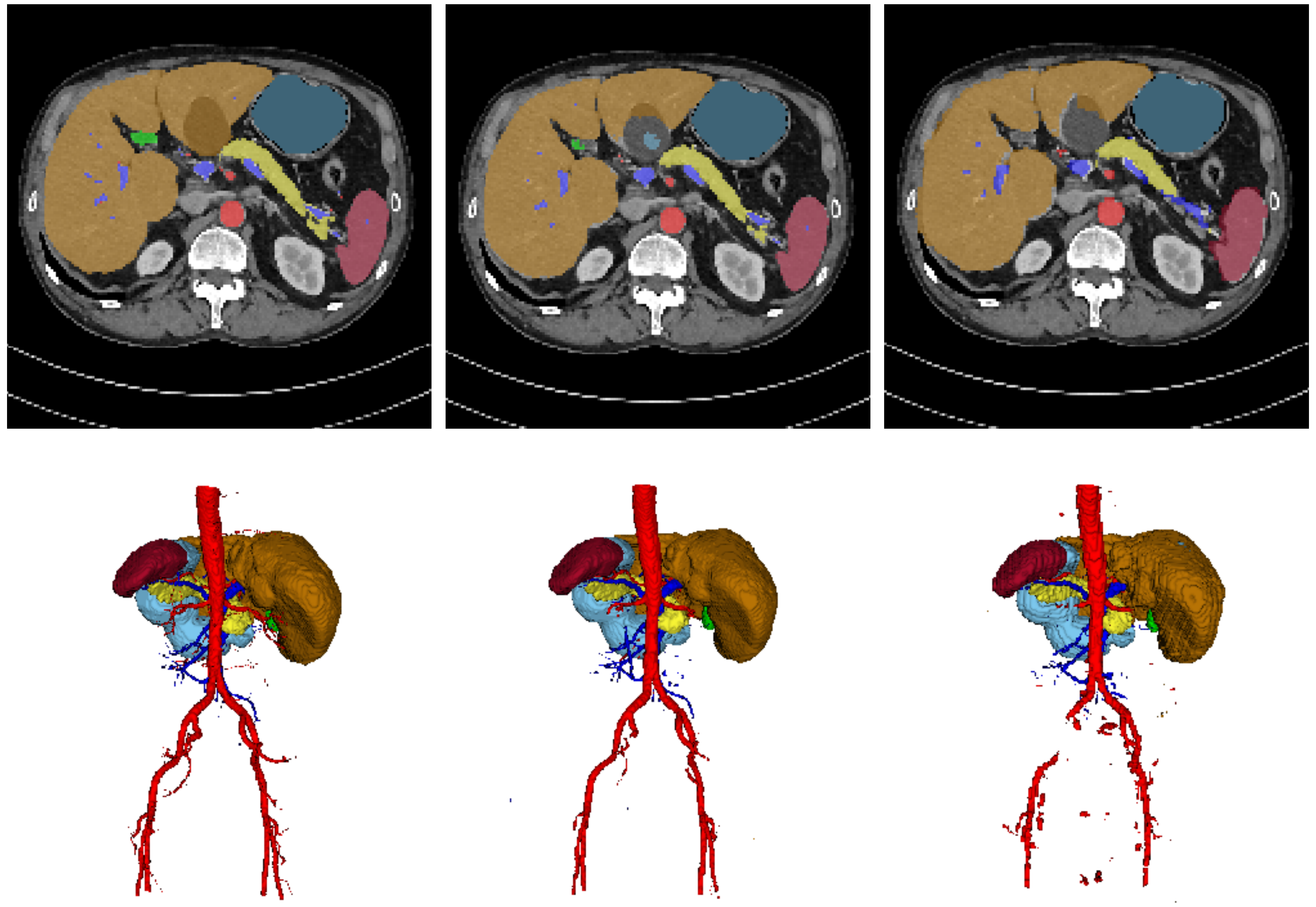


V-NET FOR 3D SEGMENTATION (6/7)

Experimental evaluation

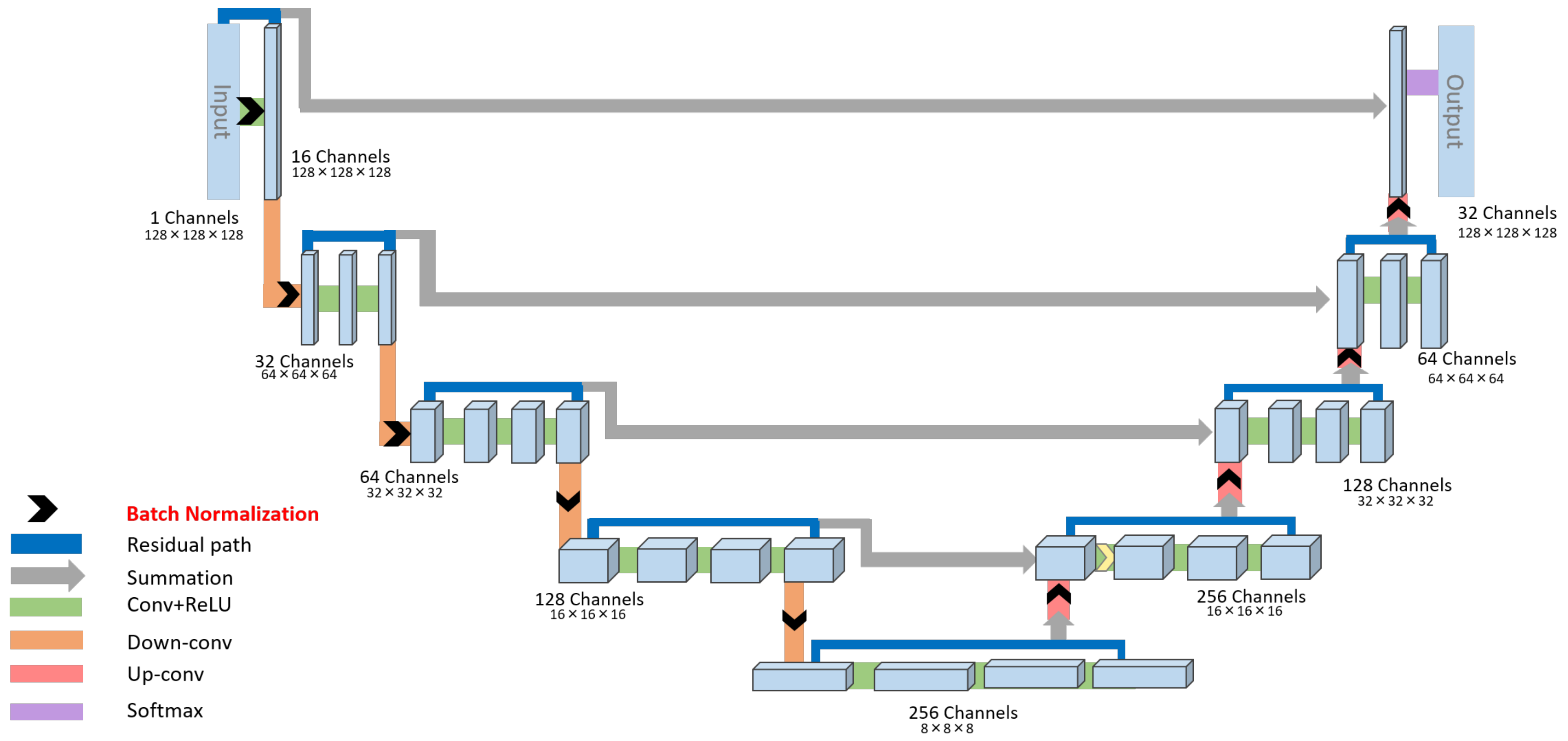
Multi-class abdominal CT segmentation

- Train on closed dataset (340 cases)
- Test on Public + closed dataset (167 cases)
- Improved version of V-Net with normalisation
- We expect to improve results with group norm



(a) Ground truth (b) V-Net 2.0 (c) 3D U-net

Dice (%)	train/test	artery	vein	liver	spleen	stomach	gall.	pancreas	Avg.
V-Net 2.0	340/37	84.7	76.9	96.2	97.0	94.5	77.5	82.8	87.1
[Unseen test]	none/130	-	-	86.5	85.6	-	74.0	66.6	78.2
Gibson et al. [2]	72 (8-CV)	-	-	92	-	83	-	66	80.3
Zhou et al. [11][a]	228/12	73.8	22.4	93.7	86.8	62.4	59.6	56.1	65.0
Hu et al. [4]	140 (CV)	-	-	96.0	94.2	-	-	-	95.1

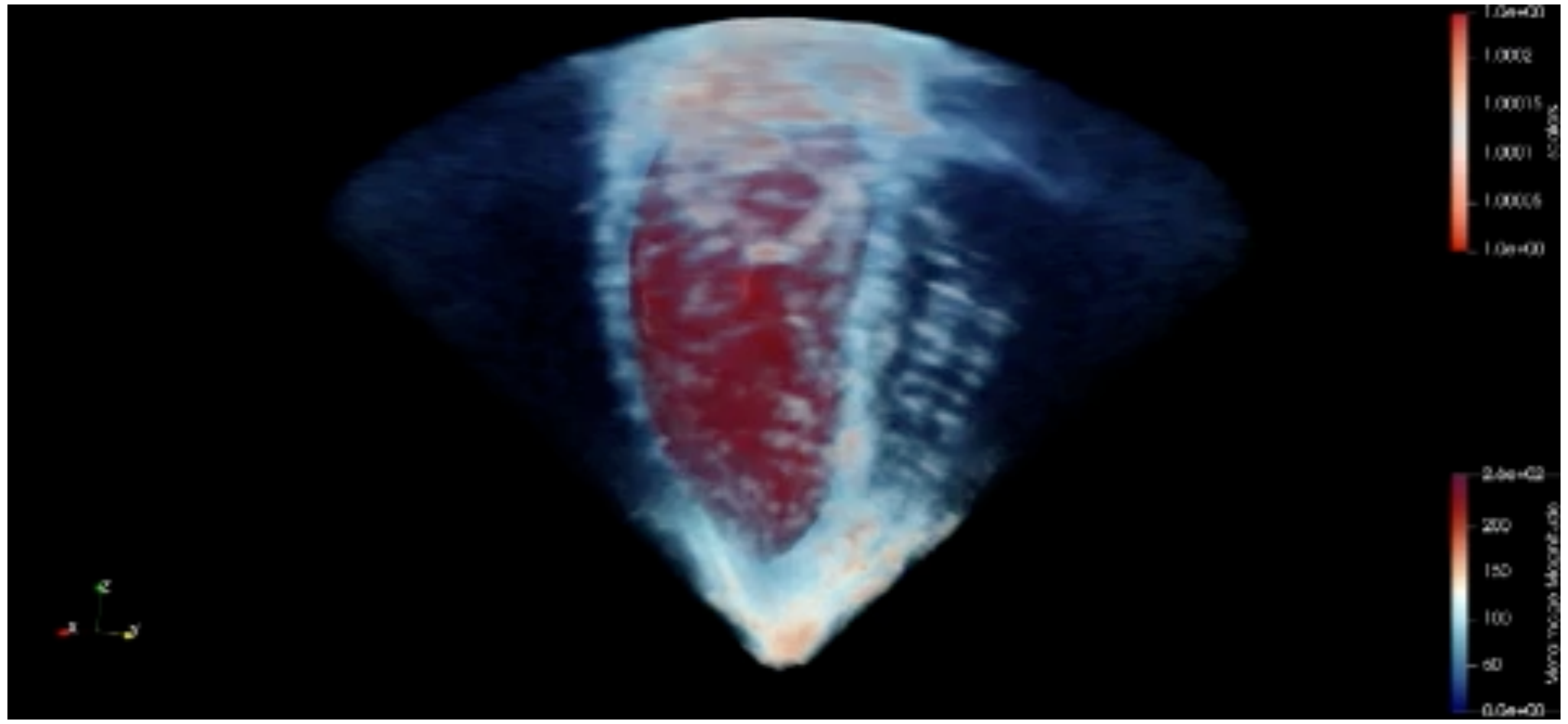


V-NET FOR 3D SEGMENTATION (7/7)

Experimental evaluation

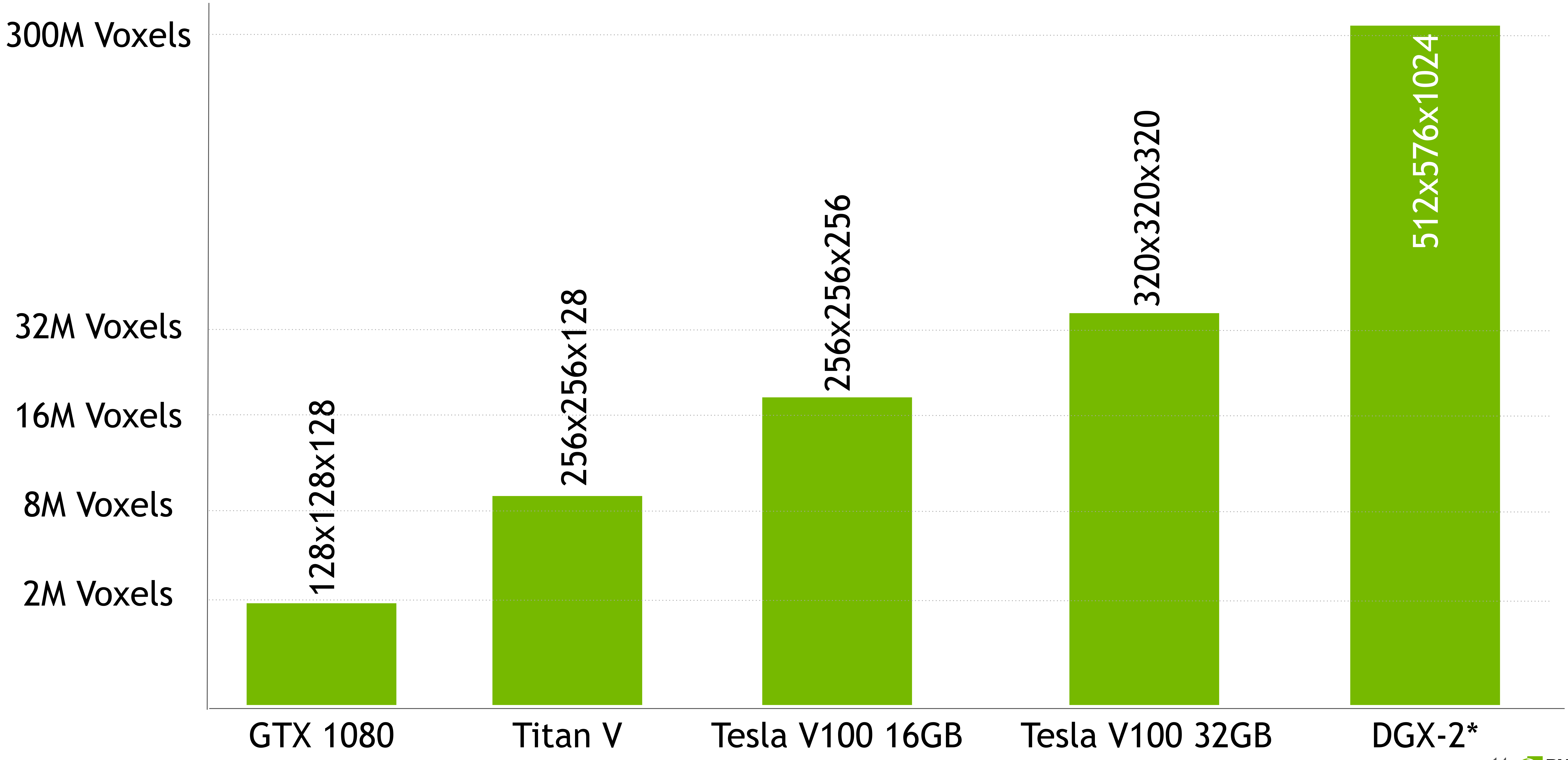
Left ventricle segmentation in 3D ultrasound

- Train on public dataset CETUS challenge
- Test on CETUS dataset
- Improved version of V-Net with 8 times more voxels (double resolution in each dimension)
- Increasing resolution yields current best challenge results
- Method available through TOMAAT.cloud



	End Diastole (ED)				End Systole (ES)			
	MV-FCN [18]	ACNN [19]	BEAS [20]	HD-VNet (ours)	MV-FCN [18]	ACNN [19]	BEAS [20]	HD-VNet (ours)
Mean dist (mm)	1.98±1.03	1.89±0.51	1.81±0.59	**1.64±0.57**	2.83±1.89	2.09±0.77	1.98±0.66	**1.75±0.79**
Hausdorff dist (mm)	11.94±9.46	6.96±1.75	**6.31±1.69**	6.32±2.11	12.45±10.69	7.75±2.65	6.95±2.14	**6.46±2.59**
Dice overlap (%)	0.906±0.026	.912±.023	.875±.046	**.917±.029**	.872±.050	.873±.051	.875±.046	**.892±.050**
EF (Pearson corr)	0.885	**0.913**	0.911	0.907				
EF (Bias+LOA) (ml)	2.74±12.01	1.78±10.09	1.7±5.18	**0.88±6.11**				

IMAGE RESOLUTION IMPROVES PERFORMANCE



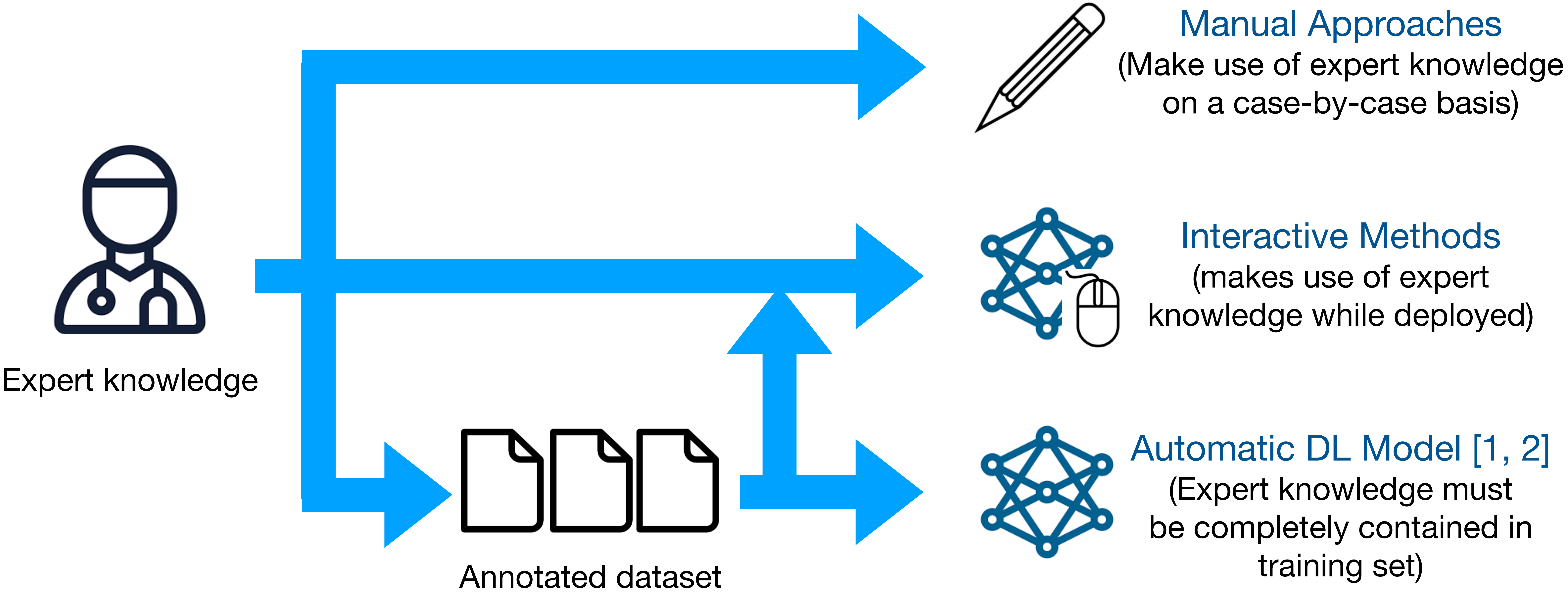
*Projected Performance based on DGX-1 (with Tesla V100). Workload: V-Net training | batch size 1. Projections Subject to change.

INTERACTIVE CNNs

Artificial intelligence applied to medicine can benefit from human expertise and interaction at test time. In this way radiologist can complete their tasks relying on deep learning models that work in synergy with them accelerating their work and boosting their confidence.

HUMAN KNOWLEDGE AND EXPERTISE

Different ways of including human knowledge in AI models



INTERACTIVE METHODS

Shallow Vs. Deep interactive methods

“Shallow” interactive methods

- Techniques such as graph cuts [1], random walks [2] and others. Work via scribbles
- Require a pretty high amount of interactions and can be time consuming
- Implemented in commercial software, and currently used in a few tasks

“Deep” interactive methods

- Use bounding boxes, clicks, scribbles and other interaction methods [3, 4]
- Applied mainly to tasks such as mass annotation of images for computer vision problems
- Require just few interactions which are supplied intuitively, yield 10-fold speed up [5]

[1] Y. Boykov and G. Funka-Lea, “Graph cuts and efficient nd image segmentation,” *International journal of computer vision*, vol. 70, no. 2, pp. 109–131, 2006.
[2] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *De- partmental Papers (CIS)*, p. 107, 2000.
[3] N. Xu, B. Price, S. Cohen, J. Yang, and T. S. Huang, “Deep interactive object selection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 373–381
[4] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, “Deep ex- treme cut: From extreme points to object segmentation,” *arXiv preprint arXiv:1711.09081*, 2017.
[5] T. L. Kline, P. Korfiatis, J. Erickson, “Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys,” *Journal of digital imaging*, vol. 30, no. 4, pp. 442–448, 2017.

INTERACTIVE NEURAL NETWORKS

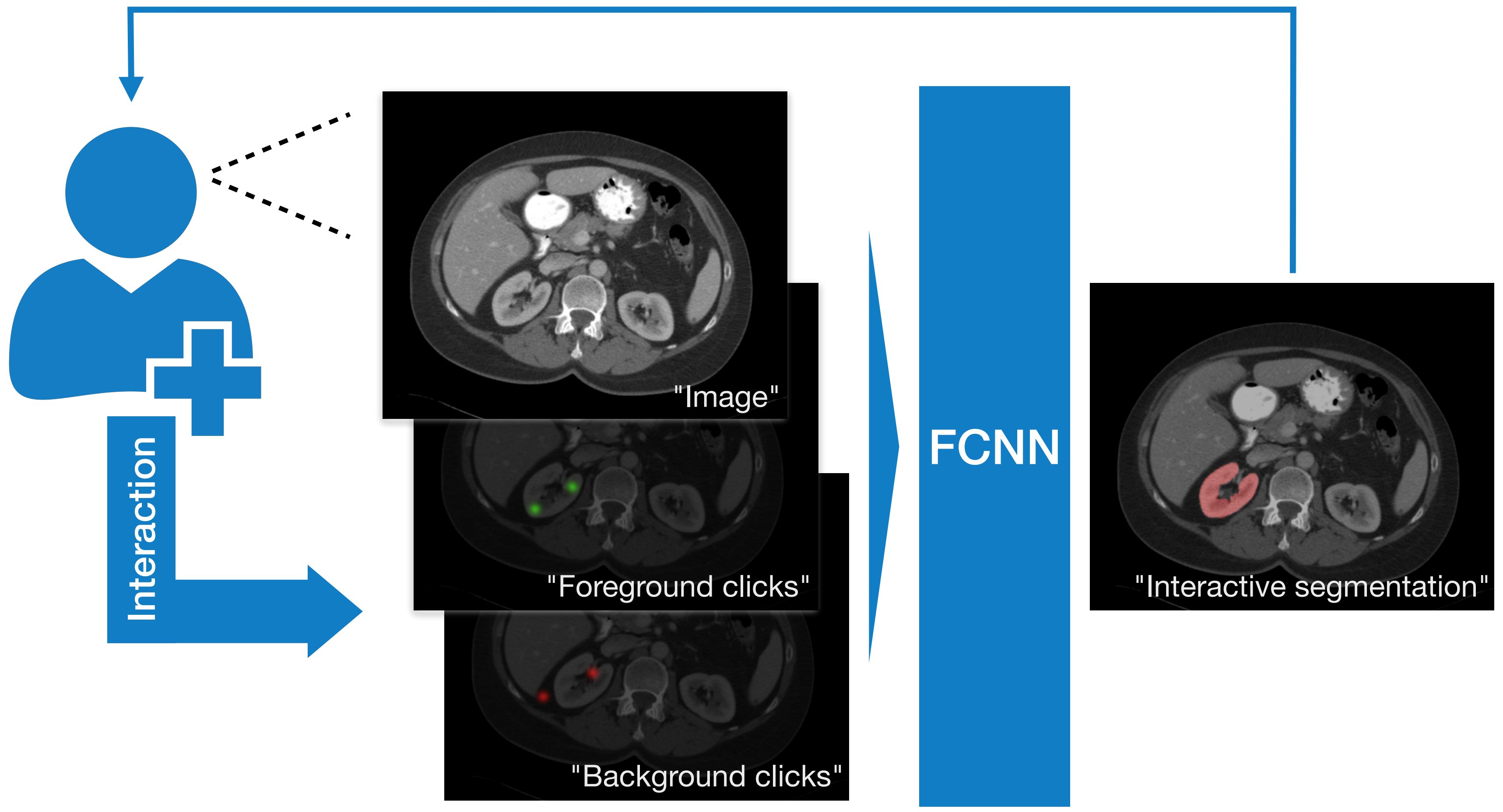
Building synergy between humans and machines

Relevance in radiology

- Medical tasks often done offline by expert radiologists
- No real-time constraint, but must yield speedup
- Interactivity helps with “black box” problem
- Interactive methods crucial for continuous learning

Our proposal

- Standard network architecture augmented to be interactive
- Simple interaction method, for example based on clicks
- Resulting algorithm must reduce time *and* improve results



Proposed interactive CNN method

https://arxiv.org/abs/1903.08205

INTERACTIVE NEURAL NETWORKS

Adding interactivity to well known network architectures

Convolutional neural network

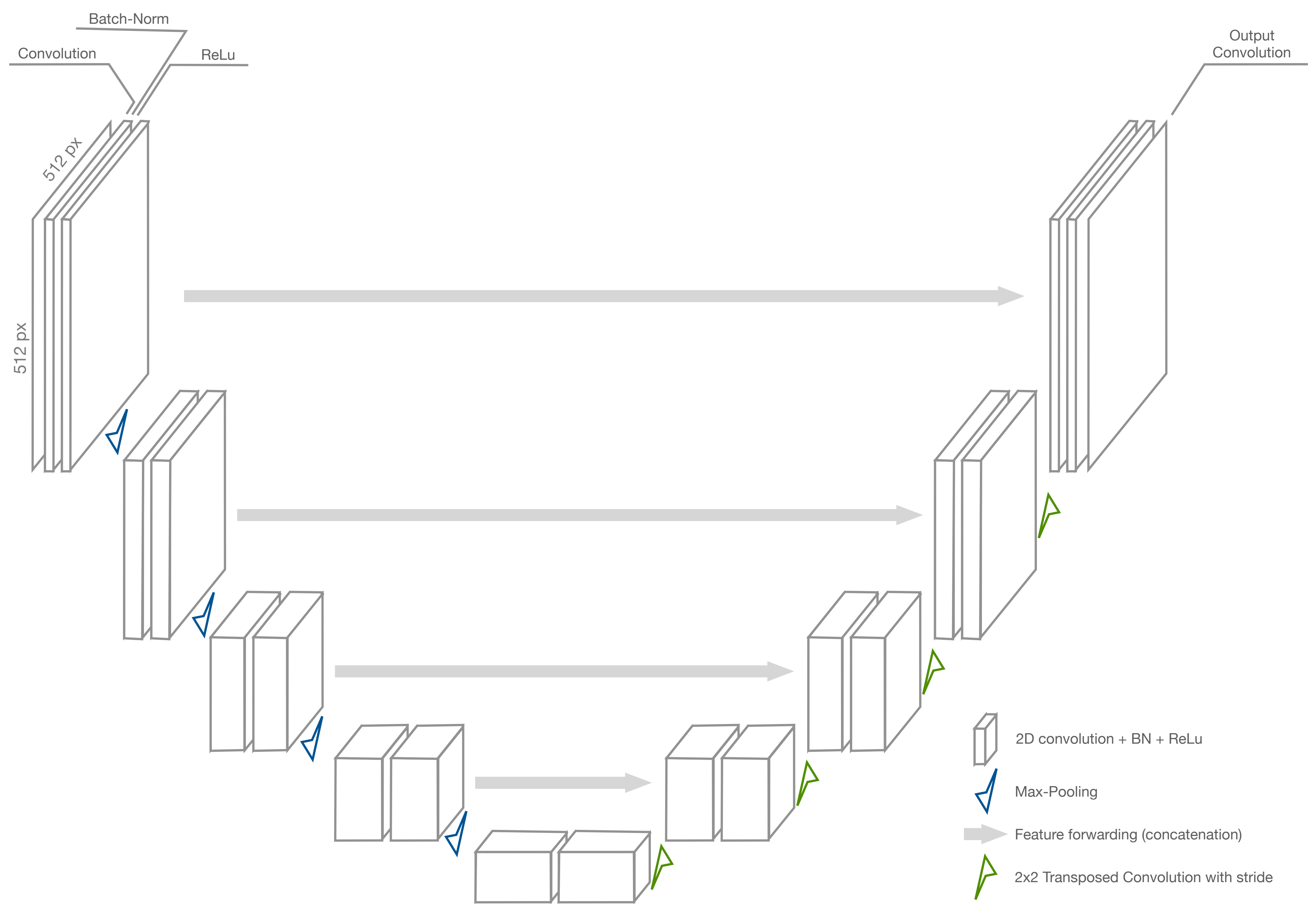
- Similar to U-Net, uses batch norm
- Optimised with ADAM and Dice loss

User interaction: mouse clicks

- Left click - attracts foreground (FG)
- Right click - attracts background (BG)

Additional input channels to model interactions

- Images are grayscale (CT images) or multi-channel
- Two additional channels are supplied to capture FG and BG
- Clicks are modelled via Gaussian centred at click location



Sakinis, Tomas, et al. "Interactive segmentation of medical images through fully convolutional neural networks." *arXiv preprint arXiv:1903.08205* (2019).

TRAINING AN INTERACTIVE METHOD

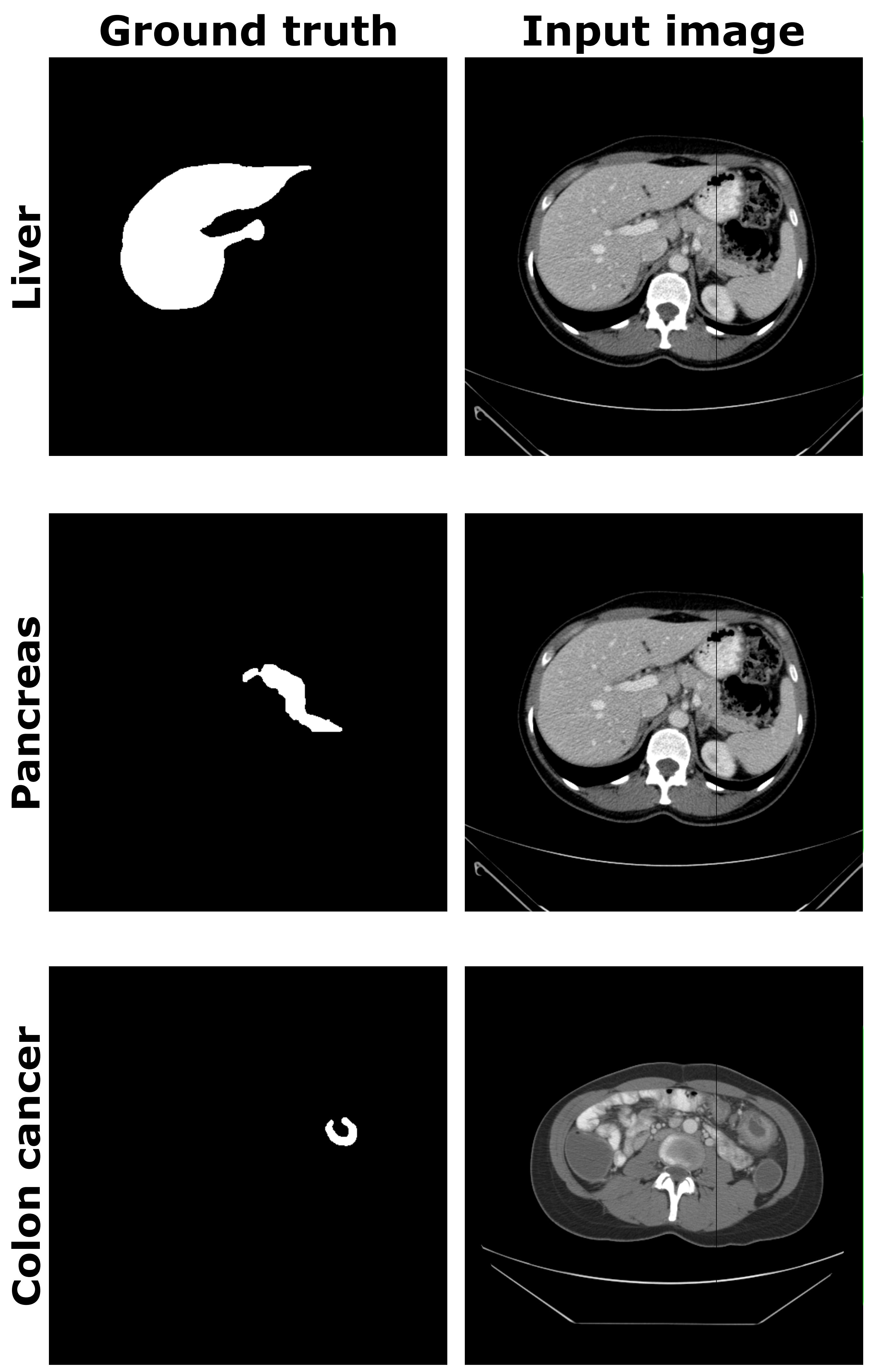
Teaching networks to interpret user interactions

Interactive segmentation dataset

- We convert multi-class datasets to binary
- Each image occurs multiple times, with a **different label** each time
- This means that the same training image will be associated to multiple, disjoint labels
- How do we tell the network which organ we are interested in segmenting? Via interaction signal!

Simulating user interaction

- During training, we simulate user behaviour during interaction
- Our simulator places clicks on the image, and the network is asked to learn the correct segmentation given inputs and clicks
- Multiple clicks can be placed on images, both for FG and BG



Sakinis, Tomas, et al. "Interactive segmentation of medical images through fully convolutional neural networks." *arXiv preprint arXiv:1903.08205* (2019).

SIMULATING HUMAN BEHAVIOUR

Mimicking user interactions during training

User interaction during training

- What constitutes realistic user interaction?
- Humans tend to correct biggest mistakes first
- Humans tend to click around the centre of the wrong area
- Our simulator acts accordingly

Error maps and click probabilities

- We run N interaction (inference) iterations before supplying each batch for training
- We define an error map to be difference between GT and prediction (init prediction is zero)
- For each round [1 … N] we select at random, with decreasing probabilities, the areas of biggest mistakes and we add a click at random with decreasing probabilities from the area's centroid.

Sakinis, Tomas, et al. "Interactive segmentation of medical images through fully convolutional neural networks." *arXiv preprint arXiv:1903.08205* (2019).

AN APPROACH GUIDED BY CLICKS

The more you click, the better results you get

Network learns "foregroundness"

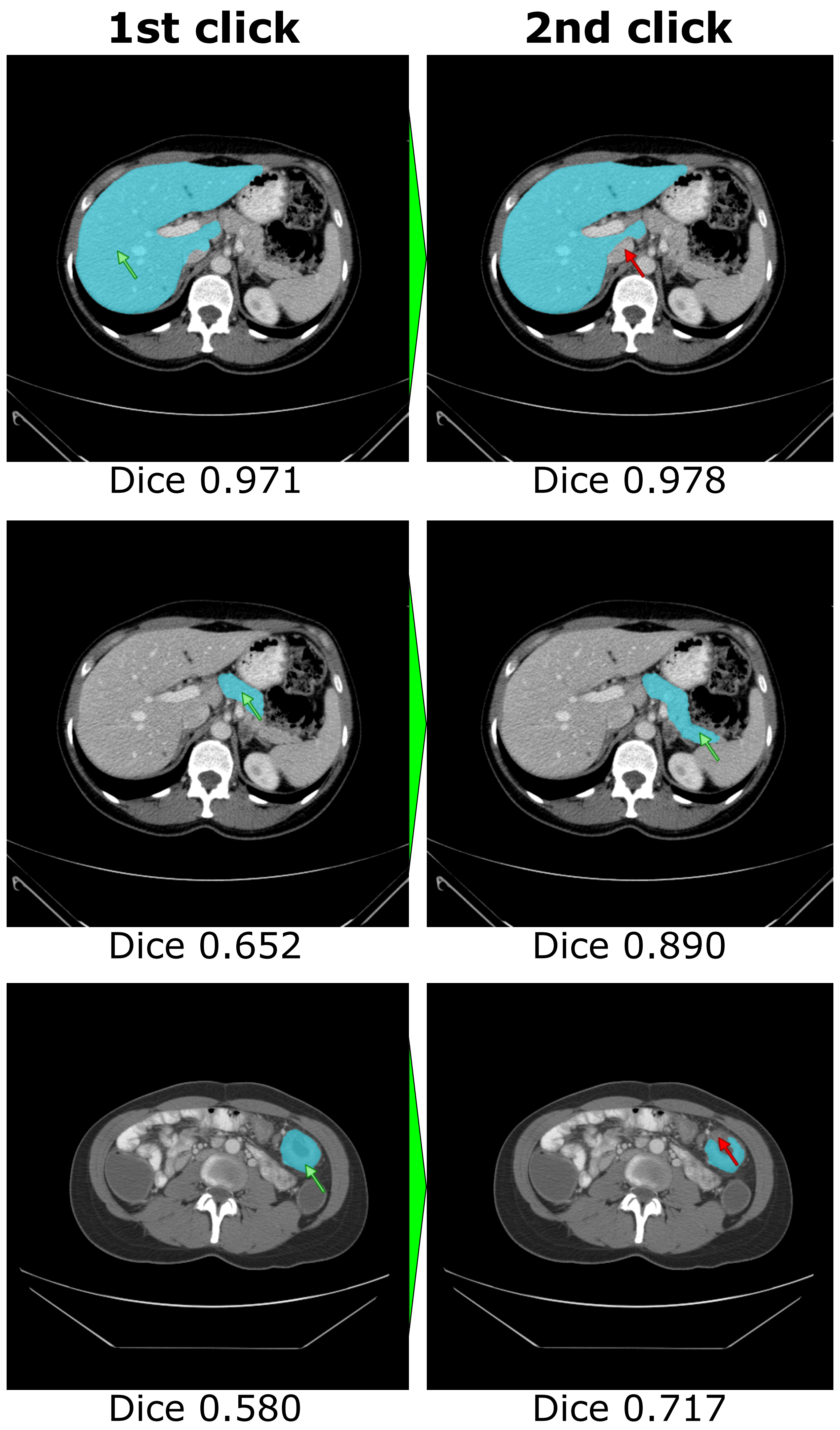
- It learns that a foreground click should attract foreground
- And that a background click should do the contrary
- It does not really learn specific object. It learns to help the user

More clicks, better results

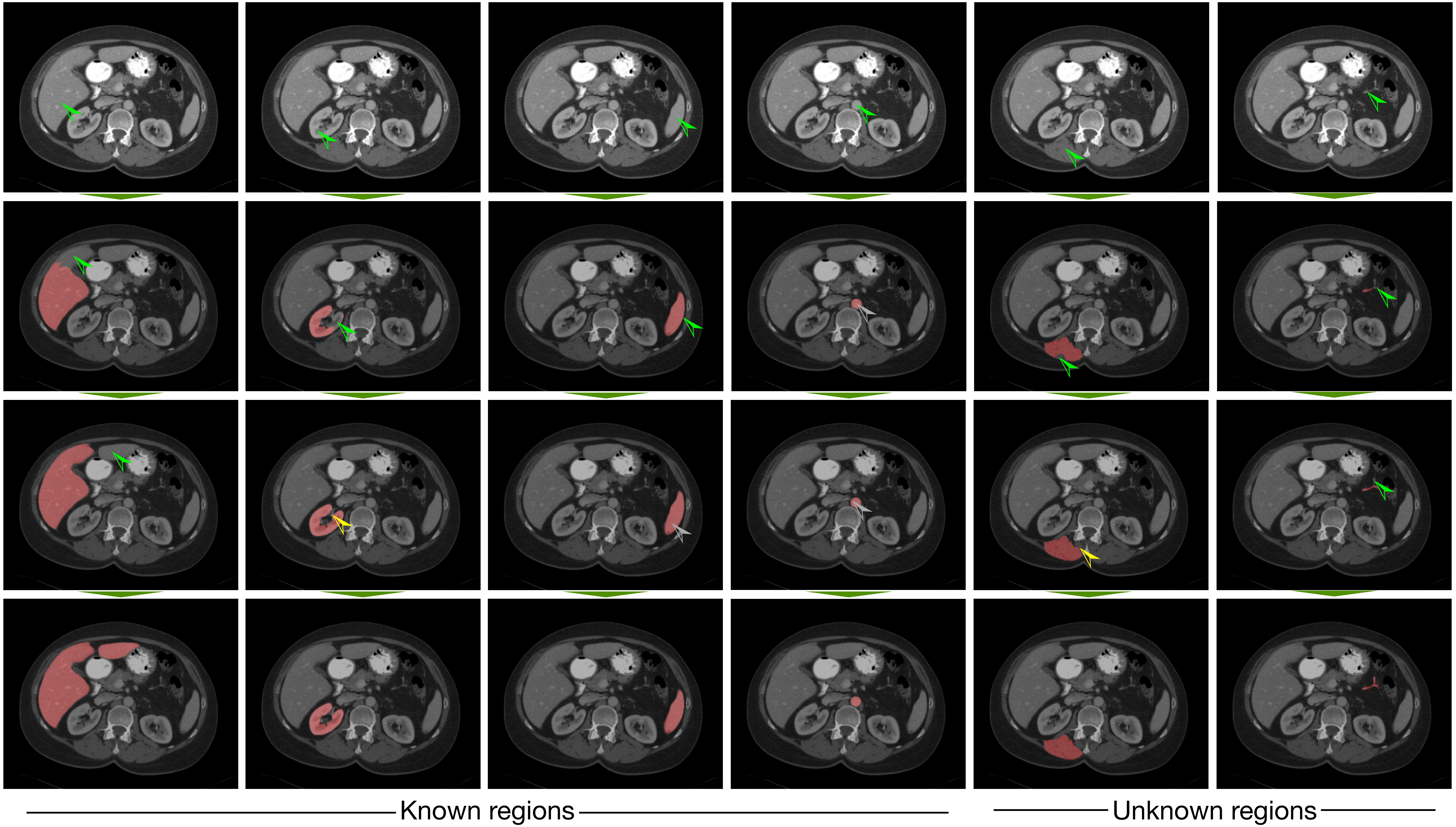
- Method is well behaved
- The more you click the better the results

Can segment stuff you have never seen before

- You trained on spleen, pancreas and liver?
- No problems! You can segment also tumors, and other things that were never seen before!
- It will require more clicks to get a nice result though



Sakinis, Tomas, et al. "Interactive segmentation of medical images through fully convolutional neural networks." *arXiv preprint arXiv:1903.08205* (2019).


Known regions
Unknown regions

EXPERIMENTAL EVALUATION

Multiple datasets, multiple tasks, previously unseen classes

Quantitative results

- On previously "seen" classes (on test set from same dataset)
- On previously "unseen" datasets
- On previously "unseen" classes (new anatomy!)

AVERAGE RESULTS ON THE "BCV" TEST DATASET. THE COLUMNS "HDD" AND "MAD" RESPECTIVELY CONTAIN THE AVERAGE HAUSDORFF AND MEAN ABSOLUTE DISTANCES (IN MM.) ACHIEVED BY EACH ALGORITHM.

Approach	Dice	HDD	MAD
Ours	**0.94023**	**7.3693**	**0.25351**
"nnUNet"	0.88104	17.2583	1.3894
"ResNet101KL"	0.84968	18.468	1.4504
DLTK (U-Net) [39]	0.81535	62.8724	1.8613
Auto Context [28]	0.7824	26.0951	1.9362
DeepSeg (FCN) [29]	0.76728	27.3397	2.9587

PERFORMANCE ACHIEVED BY OUR ALGORITHM ON THE MSD DATASET IN COMPARISON WITH STATE-OF-THE-ART (SOA) RESULTS.

Anatomy	SoA Dice	Our Dices				
Interactions	–	1	2	5	10	15
Spleen	0.96 [38]	0.89	0.93	0.95	**0.96**	**0.97**
Colon cancer	0.56 [38]	**0.64**	**0.72**	**0.81**	**0.87**	**0.89**

REGION-WISE RESULTS OBTAINED BY OUR ALGORITHM IN COMPARISON WITH THE CURRENT BEST RESULT REPORTED ON THE WEBSITE OF THE BCV CHALLENGE AS OF THE 19TH OF MARCH 2019.

Method	Metric	Spln.	R-Kid.	L-Kid.	G.Blad	Esoph.	Liver	Stom.	Aorta	IVC	P/S V.	Pancr.	R-AG	L-AG
Ours	Dice	**0.978**	**0.966**	**0.968**	**0.923**	**0.935**	**0.980**	**0.962**	**0.967**	**0.954**	**0.908**	**0.922**	**0.876**	**0.882**
	HDD	**4.785**	**5.848**	–	–	**4.303**	**16.582**	15.**244**	**3.367**	**7.297**	**7.289**	**10.180**	**3.539**	**5.512**
	MAD	**0.177**	**0.239**	–	–	**0.200**	**0.356**	**0.343**	**0.205**	**0.249**	**0.2064**	**0.351**	**0.215**	**0.234**
BCV Best	Dice	0.968	0.921	0.957	0.783	0.834	0.975	0.920	0.925	0.870	0.836	0.830	0.788	0.781
	HDD	6.489	10.88	–	–	10.31	28.50	30.38	19.98	19.74	23.41	17.74	6.58	7.41
	MAD	0.370	0.831	0.470	–	1.000	0.993	1.862	1.289	1.588	0.866	1.166	0.591	0.687

Sakinis, Tomas, et al. "Interactive segmentation of medical images through fully convolutional neural networks." *arXiv preprint arXiv:1903.08205* (2019).

DEEPGROW DEMO

Segmentation on seen and unseen classes

Sakinis, Tomas, et al. "Interactive segmentation of medical images through fully convolutional neural networks." *arXiv preprint arXiv:1903.08205* (2019).

BUILDING USEFUL AI

Not all deep learning methods are created equal. In fact, most of current research done by leading universities will never find practical application. AI is a game-changing technology for the whole medical field, but it needs to be tailored to its users, the scenario where it is deployed and the true clinical needs of every application.

“AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't”

–Curtis Langlotz

USEFUL AI IS AI THAT CAN BE USED

Don’t replace people. Make their work easier instead

Radiologist are trained professionals

- They have expertise and in general they know how to interpret medical images
- We should aim at assisting them, without trying to automatise everything for them
- Assist existing workflows, enable workflows that would require superhuman powers
  - Faster segmentation 20x faster using assisted methods.
  - Deep learning based solution to time consuming algorithms (Eg. registration) 100x faster with automated methods
  - Mining whole PubMed database for information at superhuman speed via AI

Keep experts in the loop

- Allow interaction instead of end-to-end automation
- Interactive methods are useful for tasks done offline (radiology)
- Their capabilities reach beyond those of automatic methods
- Easier to be accepted in existing workflows, keep user in charge of their work

CONCLUSIONS

Methods in ML/DL for radiology

Accuracy is important, lives depend on it

- Need to implement strategies to reduce errors!
- Some strategies I have worked on are: Hough voting, PCA, and better network architectures
- But interactivity is so far the most effective strategy

Keep experts in the loop

- Creating methods that are able to do “everything” is not something people want
- Claiming full-automation corresponds very often to over-promising
- Black boxes are way more transparent when you can play with them and see how they respond

Architectures are standard, improvement is in how methods can be used in practice

- Architectures such as U-Net, V-Net, ResNet, are standard
- Techniques such as convolution, de-convolution, BatchNorm, etc are standard
- Innovation can be created by making these methods useful to their users (radiologist)
- Start from the problem and find a solution (keeping the user experience and use case in mind)

Thank You!

– Questions?