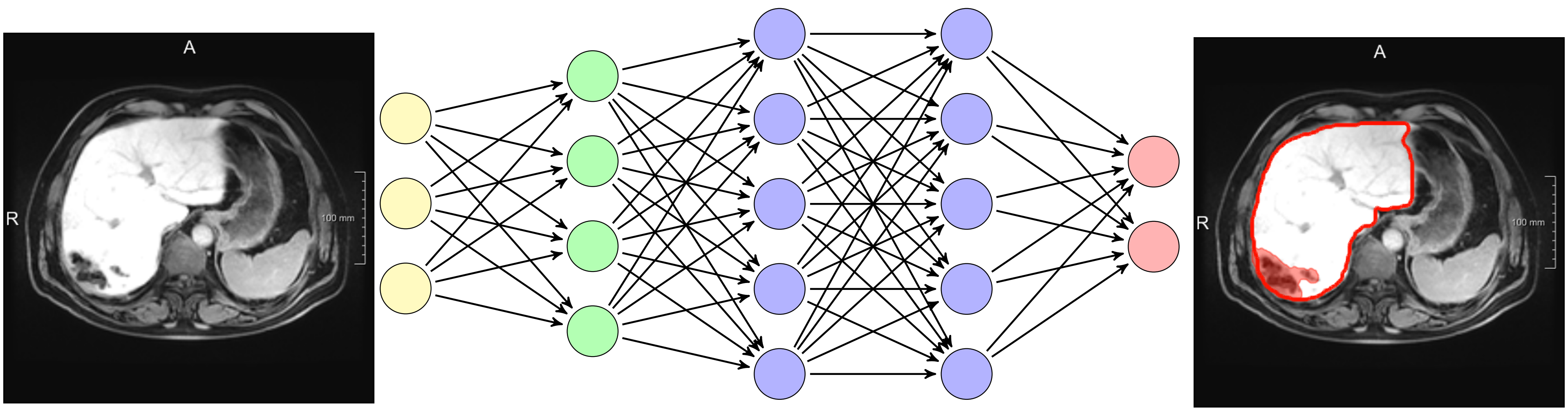
DEEP LEARNING ALGORITHMS FOR LIVER AND TUMOR SEGMENTATION

Grzegorz Chlebus, Hans Meine

SIRTOP



Medical Knowledge Through Research

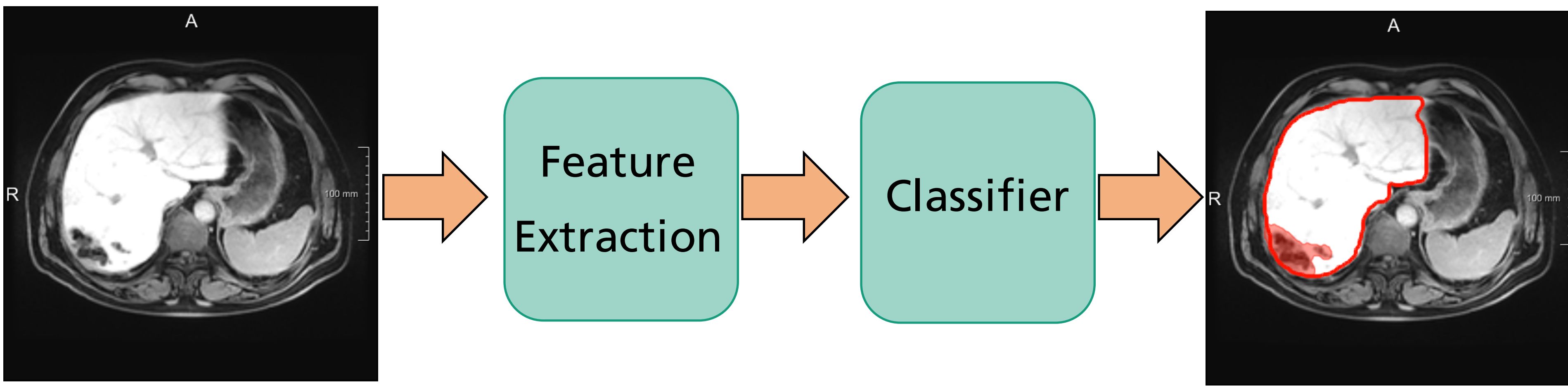
Städtisches Klinikum Dresden

Fraunhofer MEVIS

Deep learning is a major thing

Before 2013

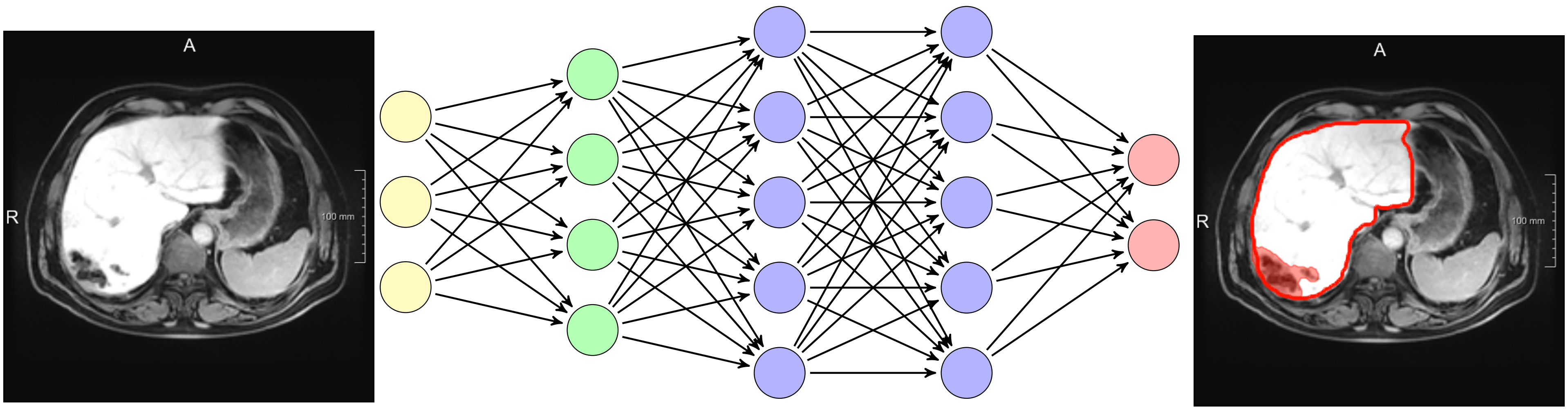
- We failed to design image analysis algorithms that perform better than humans.



Deep learning is a major thing

2013 – now

- We found that deep learning works well for image understanding tasks thanks to faster computers and better training algorithms.



Deep neural network

DL algorithms can surpass expert performance

JAMA | Original Investigation

Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer

Babak Ehteshami Bejnordi, MS; Mitko Veta, PhD; Paul Johannes van Diest, MD, PhD; Bram van Ginneken, PhD; Nico Karssemeijer, PhD; Geert Litjens, PhD; Jeroen A. W. M. van der Laak, PhD; and the CAMELYON16 Consortium

IMPORTANCE Application of deep learning algorithms to whole-slide pathology images can potentially improve diagnostic accuracy and efficiency.

OBJECTIVE Assess the performance of automated deep learning algorithms at detecting metastases in hematoxylin and eosin–stained tissue sections of lymph nodes of women with breast cancer and compare it with pathologists' diagnoses in a diagnostic setting.

DESIGN, SETTING, AND PARTICIPANTS Researcher challenge competition (CAMELYON16) to develop automated solutions for detecting lymph node metastases (November 2015-November 2016). A training data set of whole-slide images from 2 centers in the Netherlands with (n = 110) and without (n = 160) nodal metastases verified by immunohistochemical staining were provided to challenge participants to build algorithms. Algorithm performance was evaluated in an independent test set of 129 whole-slide images (49 with and 80 without metastases). The same test set of corresponding glass slides was also evaluated by a panel of 11 pathologists with time constraint (WTC) from the Netherlands to ascertain likelihood of nodal metastases for each slide in a flexible 2-hour session, simulating routine pathology workflow, and by 1 pathologist without time constraint (WOTC).

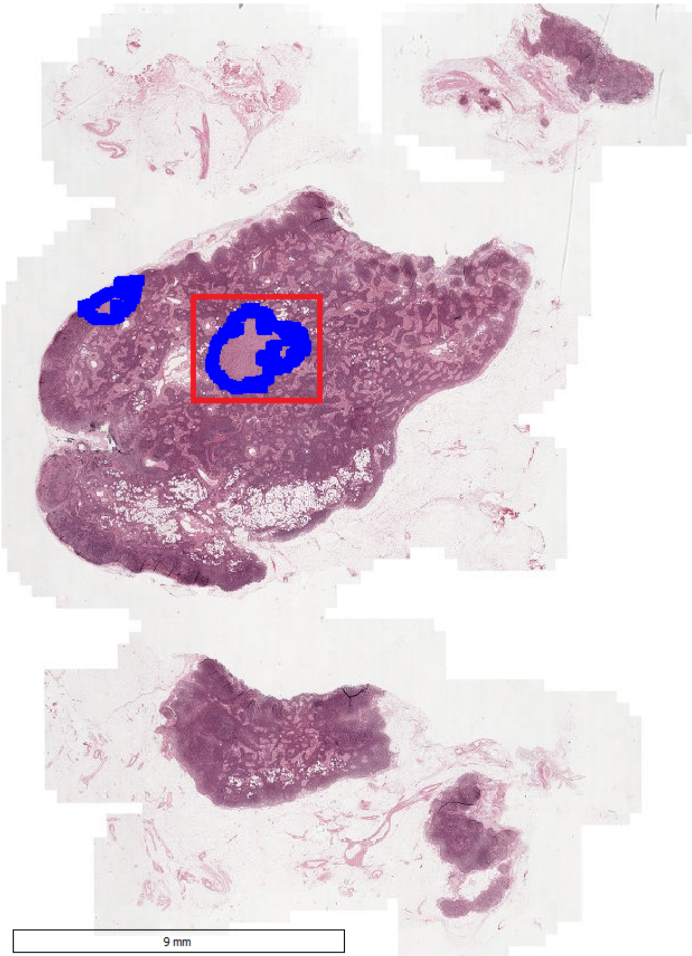
Editorial page 2184

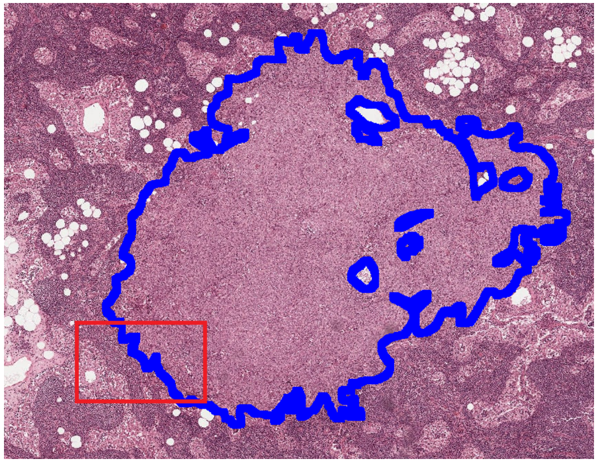
Related articles page 2211 and page 2250

Supplemental content

CME Quiz at jamanetwork.com/learning and CME Questions page 2252

DL algorithms can surpass expert performance





Conclusions

In the setting of a challenge competition, some deep learning algorithms achieved better diagnostic performance than a panel of 11 pathologists participating in a simulation exercise designed to mimic routine pathology workflow; algorithm performance was comparable with an expert pathologist interpreting slides without time constraints. Whether this approach has clinical utility will require evaluation in a clinical setting.

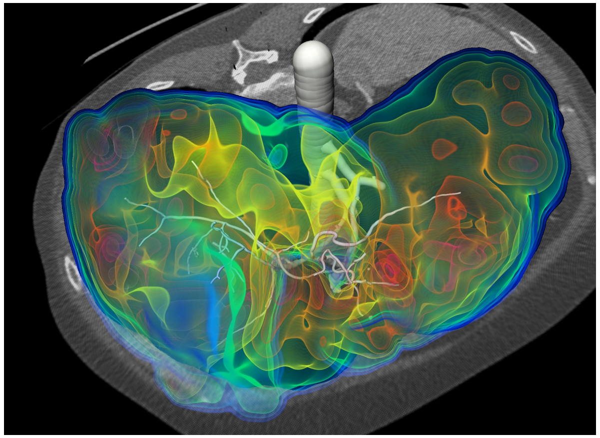
What medical tasks can be solved by DL?

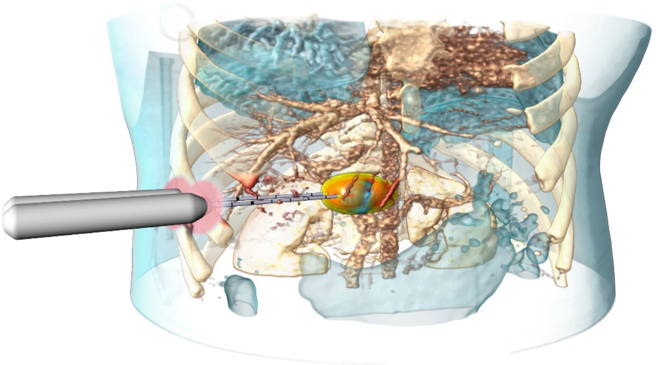
- **Detection** Is it present or not?
- **Classification** What type of thing is it?
- **Segmentation** How big is it, what shape does it have?
- **Prediction** What are the chances that this patient will get cancer in X years from now?
- **Recommendation** Which therapy option would be the best for this patient?

Automatic liver and tumor segmentation

Motivation

- Automate/improve the planning process of liver interventions
  - SIRT planning
    - Basis for tumor load computation
    - Required for dose computation
- Manual or semi-automatic segmentation
  - Tedious and time consuming
  - Inter-observer variability





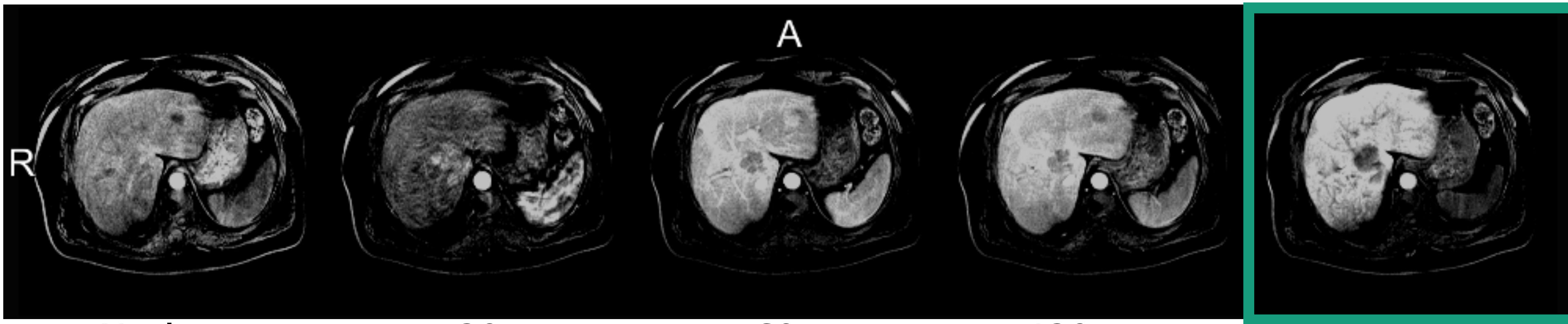
CT Data

- LiTS Challenge dataset
  - 131 CT scans with reference segmentations of liver and tumors
  - ~0.8 mm in-plane resolution
  - ~1.5 mm slice thickness

- Liver surgery planning dataset
  - 179 CT scans with reference segmentations of liver
  - ~0.6 mm in plane-resolution
  - ~0.8 mm slice thickness

MRI Data

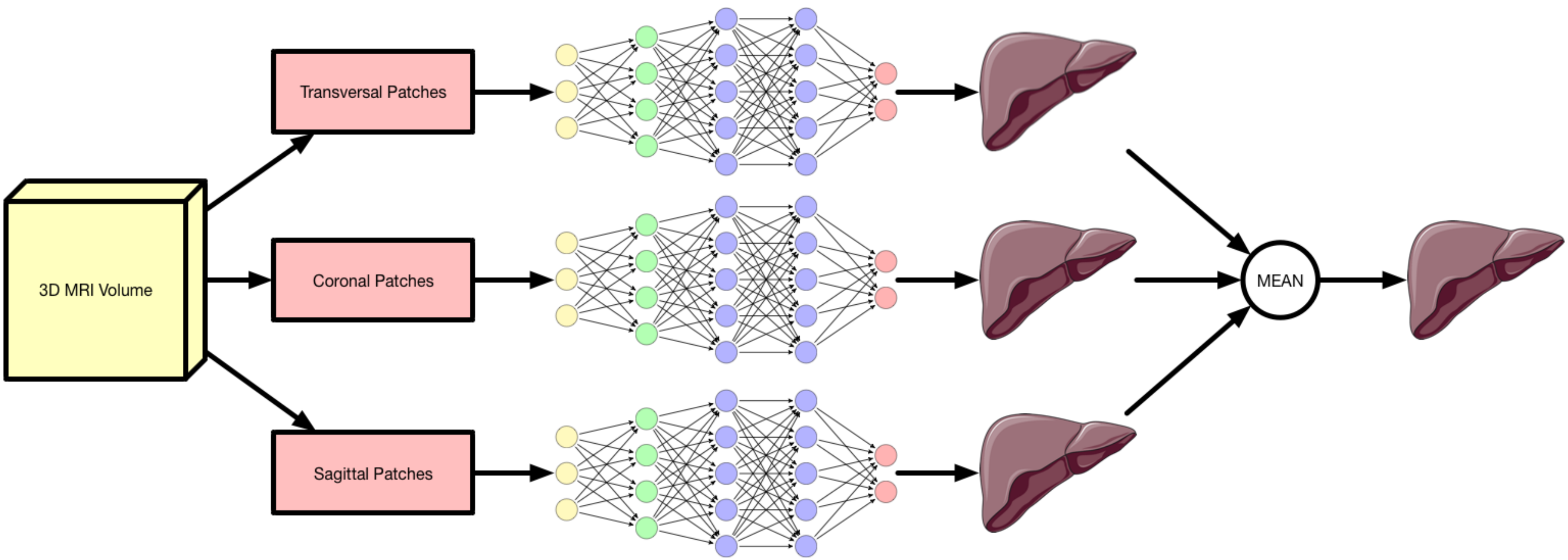
- SIRTOP dataset
  - 90 DCE-MRI scans with reference liver and tumor segmentations
  - Acquired at Städtisches Klinikum Dresden, Germany
  - 0.74-1.76 mm in-plane resolution
  - 2-5 mm slice thickness



Native 20s 60s 120s 15 min

Segmentation Pipeline

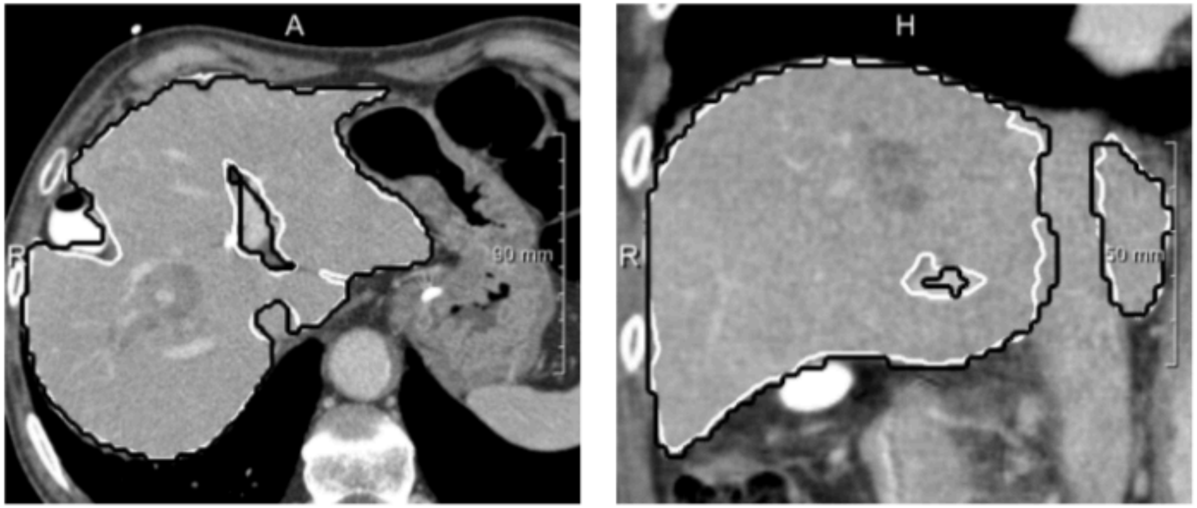
- OrthoMean [1]



[1] Prasoon A et al., "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network", MICCAI 2013.

Results: CT Liver Segmentation

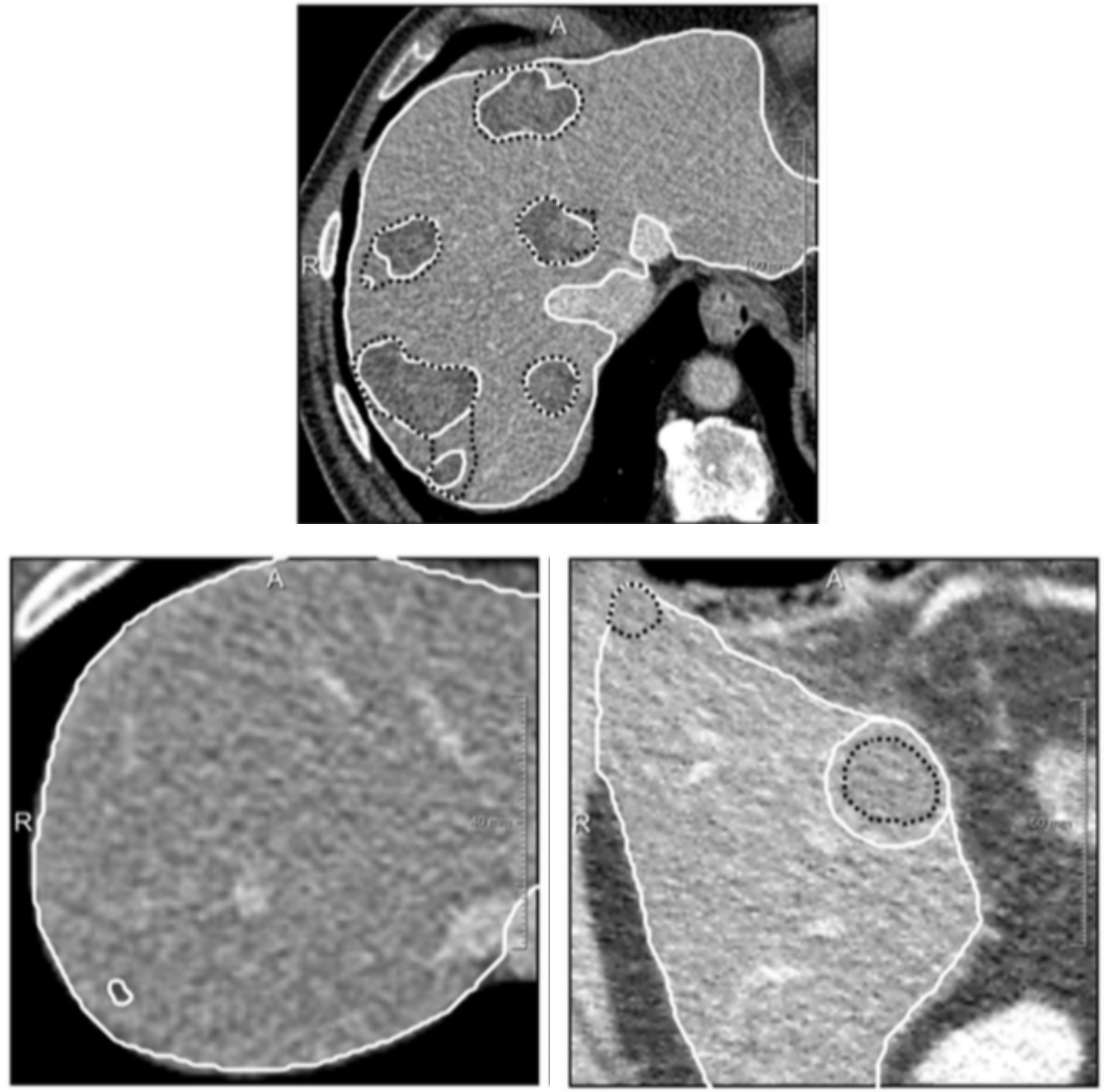
- 40 test cases
- Automatic method: 79 points according to MICCAI score [1]
- Trained human performance (no radiological expert): 75 points



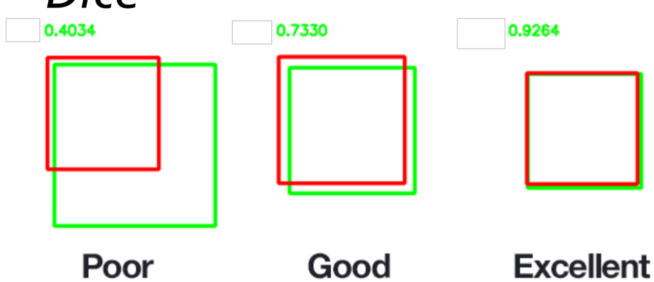
[1] Heimann T et al., "Comparison and Evaluation of Methods for Liver Segmentation from CT Datasets", IEEE TMI 2009.

Results: CT Liver Tumor Segmentation

- 30 test cases
- Automatic Method
  - 0.58 Dice per case
  - 0.69 Dice per tumor
- MTRA performance
  - 0.7 Dice per case
  - 0.72 Dice per tumor



Dice



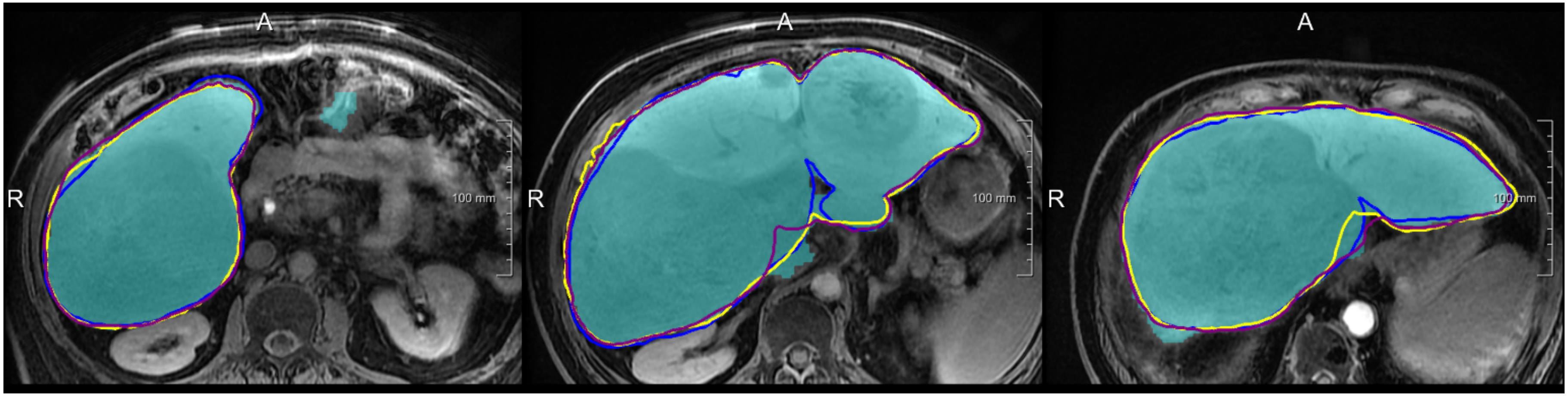
Results: CT Liver Tumor Segmentation

				Lesion								
#	User	Entries	Date of Last Entry	Dice per case ▲	Dice global ▲	VOE ▲	RVD ▲	ASSD ▲	MSD ▲	RMSD ▲	Precision at 50% overlap ▲	Recall at 50% overlap ▲
1	leHealth	20	08/04/17	0.7020 (1)	0.7940 (5)	0.394 (11)	5.921 (18)	1.189 (12)	6.682 (5)	1.726 (8)	0.156 (14)	0.437 (3)
2	hchen	12	08/04/17	0.6860 (2)	0.8290 (1)	0.356 (3)	5.164 (17)	1.073 (5)	6.055 (1)	1.562 (2)	0.409 (4)	0.408 (4)
3	hans.meine	7	07/30/17	0.6760 (3)	0.7960 (4)	0.383 (10)	0.464 (12)	1.143 (8)	7.322 (12)	1.728 (9)	0.496 (2)	0.397 (5)

- State-of-the-art results
- 3rd place at MICCA round of the LiTS challenge
  - 28 teams

Results: MRI Liver Segmentation

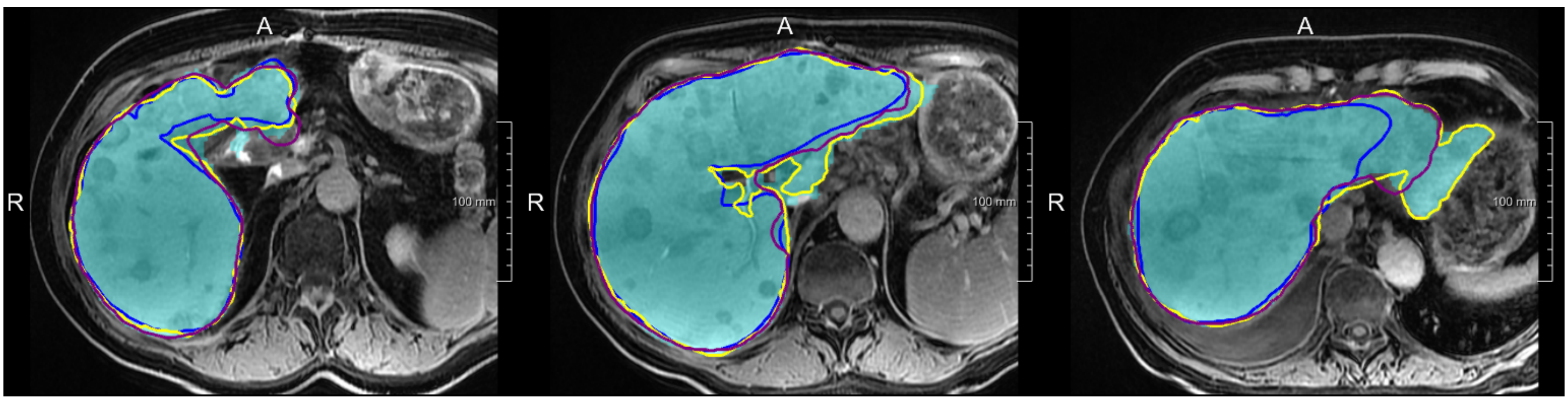
- 28 test cases
- Automatic method: 0.95 Dice [1]
- Human performance: 0.94-0.95 Dice
  - 1 radiologist
  - 2 residents



[1] Chlebus G et al., "Automatic Liver and Tumor Segmentation in Late-Phase MRI Using Fully Convolutional Neural Networks", CURAC 2018.

Results: MRI Liver Segmentation

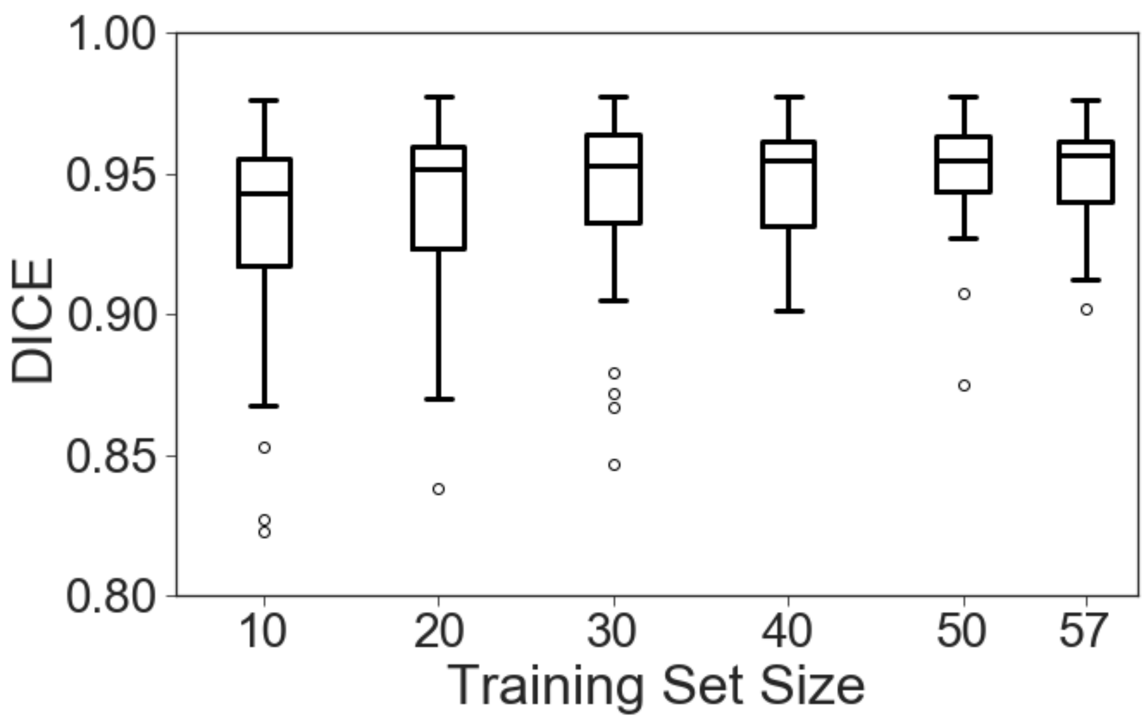
- 28 test cases
- Automatic method: 0.95 Dice [1]
- Human performance: 0.94-0.95 Dice
  - 1 radiologist
  - 2 residents



[1] Chlebus G et al., “Automatic Liver and Tumor Segmentation in Late-Phase MRI Using Fully Convolutional Neural Networks”, CURAC 2018.

Results: Training Data Size

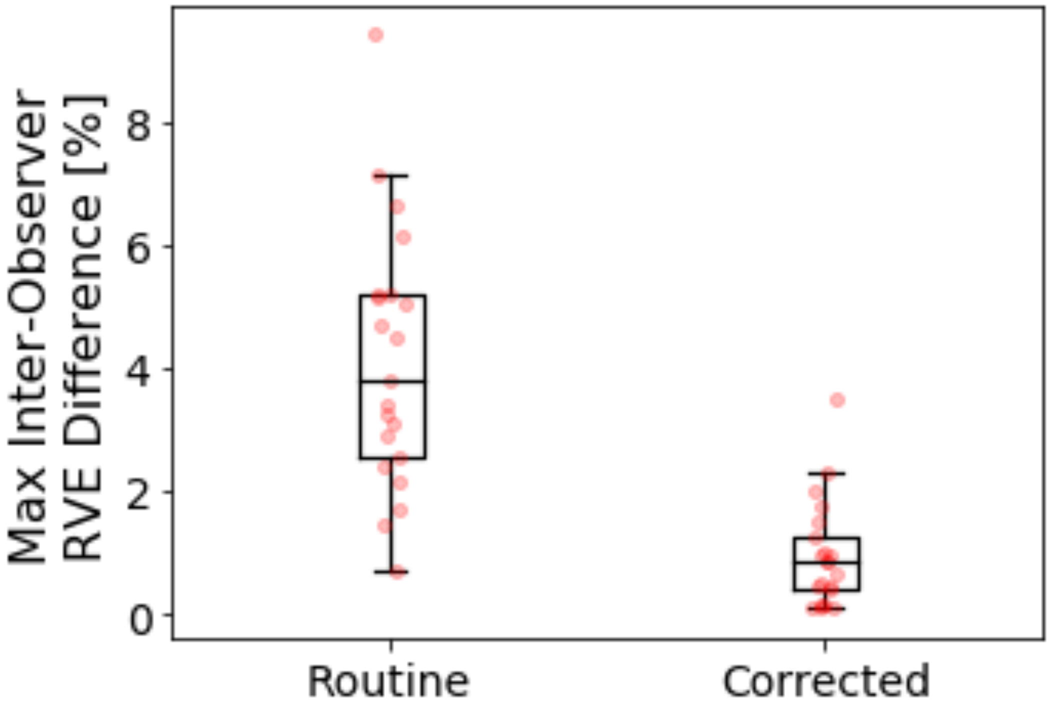
- Liver segmentation quality in MRI



Städtisches Klinikum Dresden Fraunhofer MEVIS

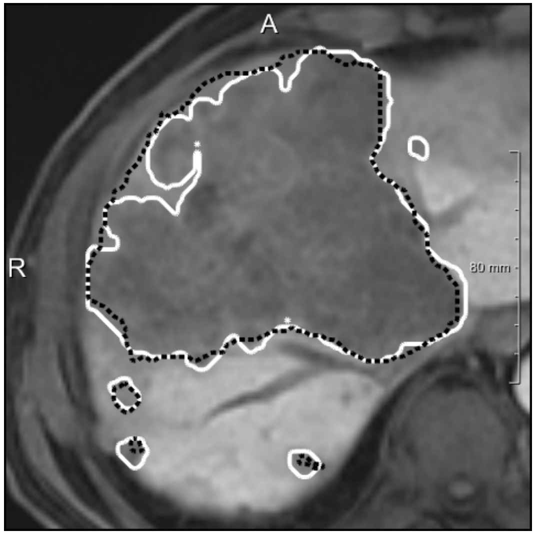
Inter-observer variability

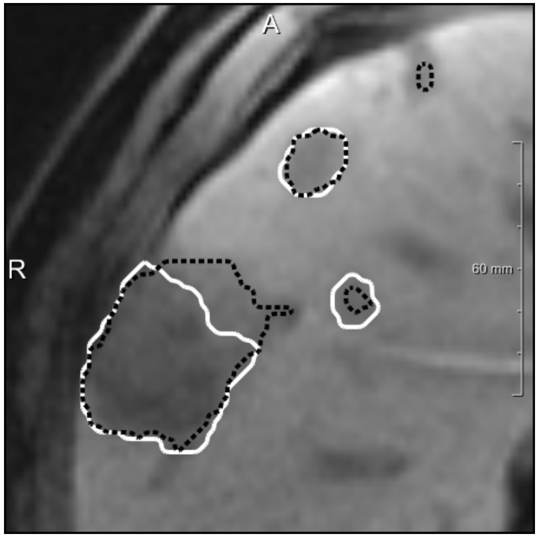
- Routine vs. corrected liver segmentations
- Ca 35% of slices were corrected (3 observers)
- Average 5 min per case correction time

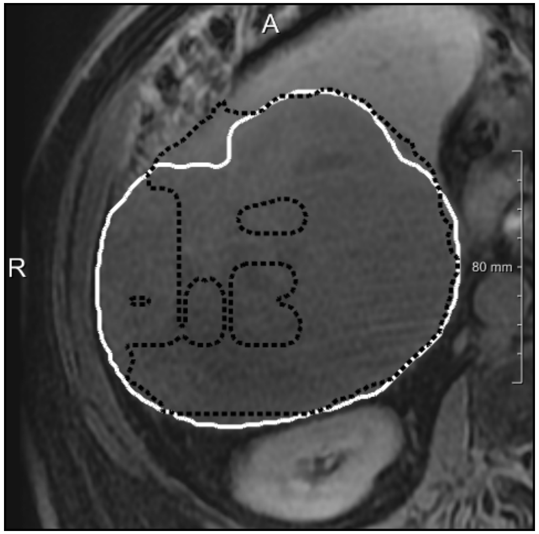


Results: MRI Liver Tumor Segmentation

- 20 test cases
- Automatic method: 0.65 Dice [1]
- Human performance: 0.90-0.93 Dice [2]







[1] Chlebus G et al., "Automatic Liver and Tumor Segmentation in Late-Phase MRI Using Fully Convolutional Neural Networks", CURAC 2018.

[2] Budjan J et al., "Semi-automatic Volumetric Measurement of Treatment Response in Hepatocellular Carcinoma after TACE", 2016.

Städtisches Klinikum Dresden

Fraunhofer MEVIS

Summary

- Deep learning algorithms are very successful at image analysis tasks
- Deep learning methods can help radiologist to perform their work faster and more accurate
- Liver segmentation quality of our automatic method was comparable to that of human segmentations
- Tumor segmentation is a more difficult task than liver segmentation
- Acquiring more training data has a positive impact on the model performance
- Future work
  - More extensive validation

Thank you for your attention ☺

Questions?

Automatic liver and tumor segmentation

Motivation

Reduce inter-observer variability

- RECIST 1.1 study by Bellomi et al. [1]
  - 100 radiologists
  - 3 cases

<u>Conclusion</u>

Age and expertise of the radiologist remain the most critical factors.

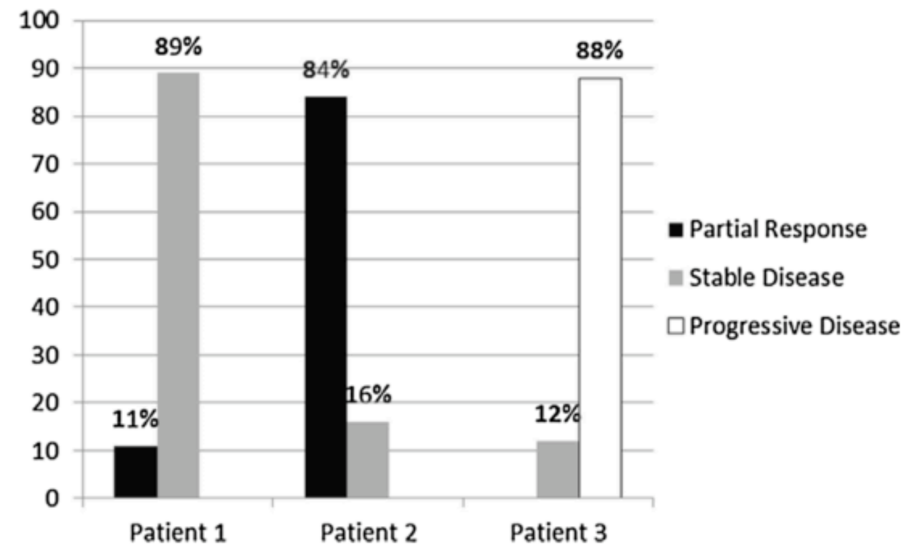


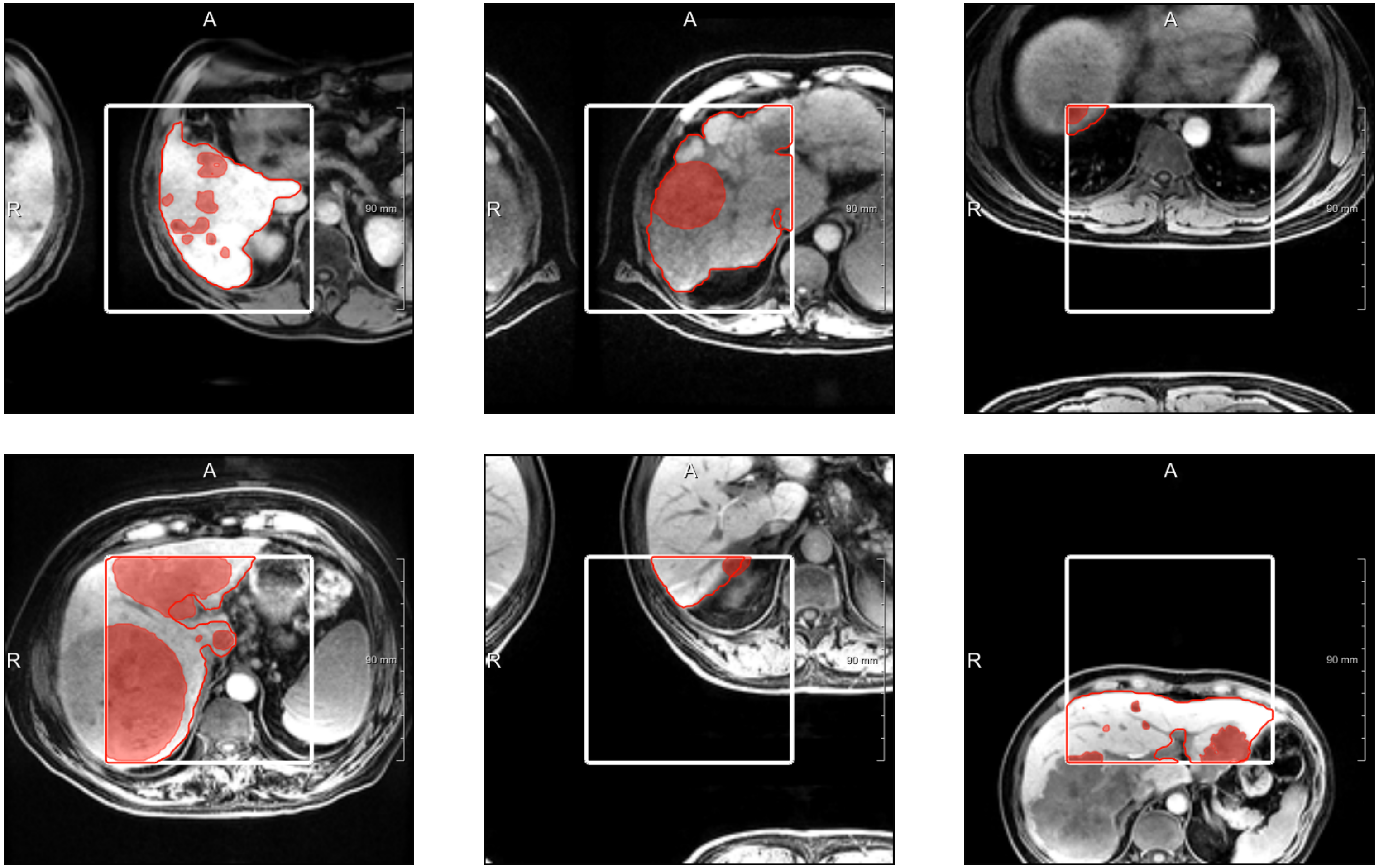
Fig. 2. Percentage of raters for each RECIST classification response by case report

[1] Bellomi M. et al. "Evaluation of inter-observer variability according to RECIST 1.1 and its influence on response classification in CT measurement of liver metastases" 2017.

Städtisches Klinikum Dresden

Fraunhofer MEVIS

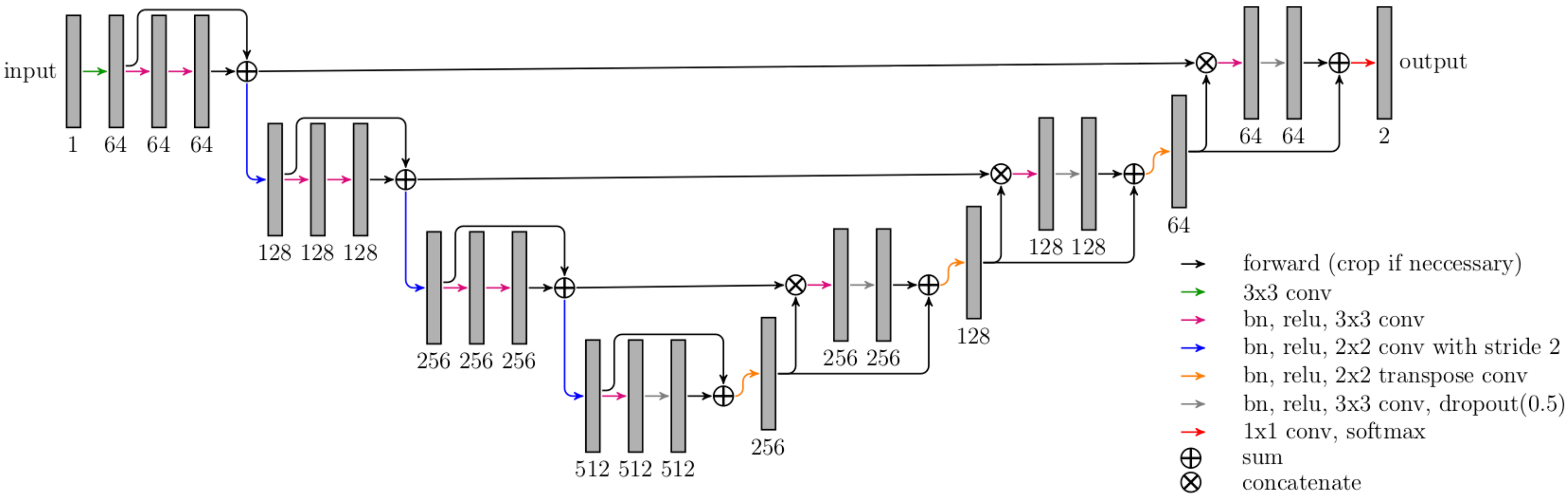
What does the neural network see?



Training

```
1. Training images with reference labels REF
2. Initialize neural network NN parameters randomly
3. DO
4.    Apply NN to a batch of training images → OUTPUT
5.    Compute the difference between OUTPUT and REF → LOSS
6.    Compute LOSS derivatives w.r.t. NN parameters → GRADIENTS
7.    Apply GRADIENTS to update NN parameters
8. UNTIL convergence
```

Neural network architecture



- U-net like [1]
- 4 resolution levels
- 9M trainable parameters
- Receptive field 94x94 voxels
- 3x3 convolution kernels
- Short skip connections [2]
- Batch normalization
- Spatial dropout

[1] Ronneberger O et al., "Convolutional networks for biomedical image segmentation", MICCAI 2015.
[2] Drozdzal M et al., "The importance of skip connections in biomedical image segmentation", 2016.