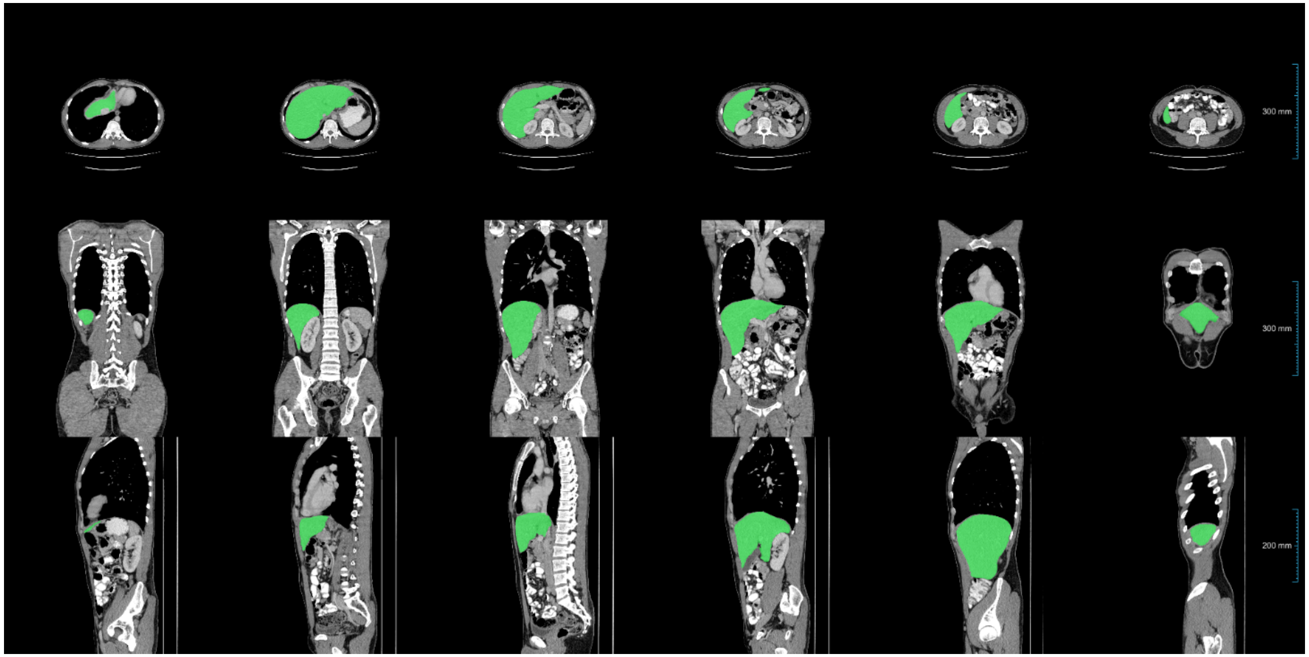
MIMICKING RADIOLOGISTS TO IMPROVE THE ROBUSTNESS OF DEEP-LEARNING BASED AUTOMATIC LIVER SEGMENTATION

Grzegorz Chlebus[1,2], Gabriel Efrain Humpire-Mamani[2], Andrea Schenk[1], Bram van Ginneken[2,1], Hans Meine[3,1]

[1]Fraunhofer Institute for Digital Medicine MEVIS, [2]Radboud University Medical Center, [3]University of Bremen



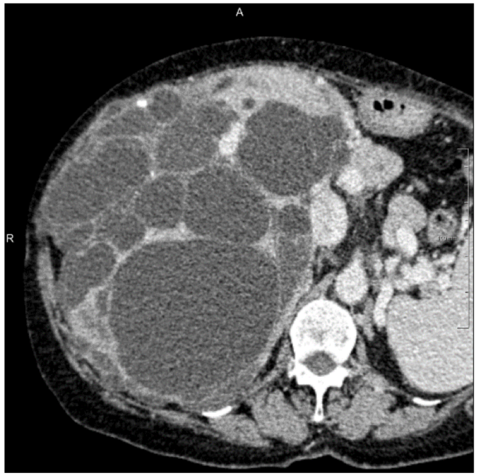
Medical Knowledge Through Research

Fraunhofer
MEVIS

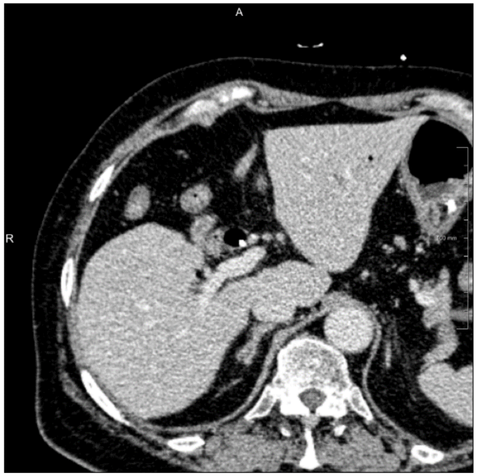
Motivation

- Develop segmentation algorithms working robustly on large, heterogeneous datasets

- Develop concepts for a big-scale error analysis to identify prevailing shortcomings of an algorithm



Polycystic liver



Resected liver

Data

<u>Training data</u>

- Yokohama City University Hospital, Japan
  - 80 cases
  - Contrast-enhanced CT, late phase
- Radboudumc, The Netherlands
  - 80 cases
  - Contrast-enhanced CT, late phase

<u>Evaluation data</u>

- Radboudumc, The Netherlands
  - 826 cases
  - Various protocols

Baseline model

- 3D u-net [1]
- 4 771 826 parameters

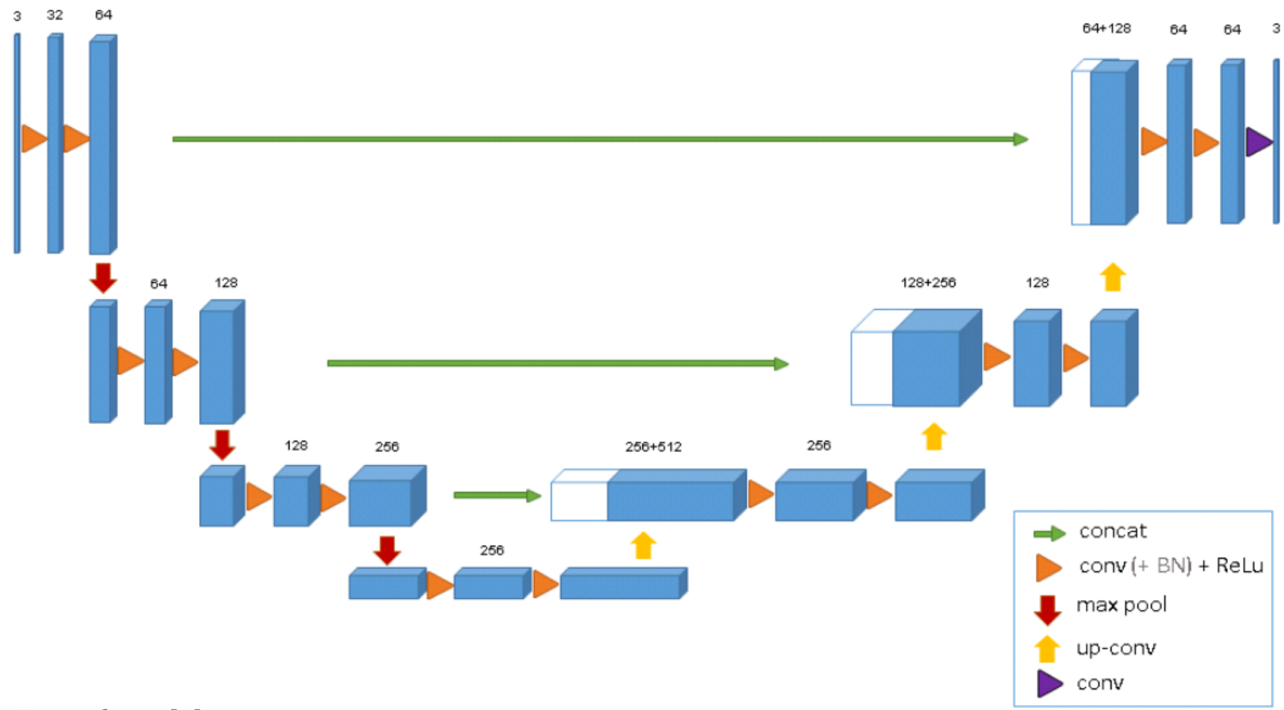


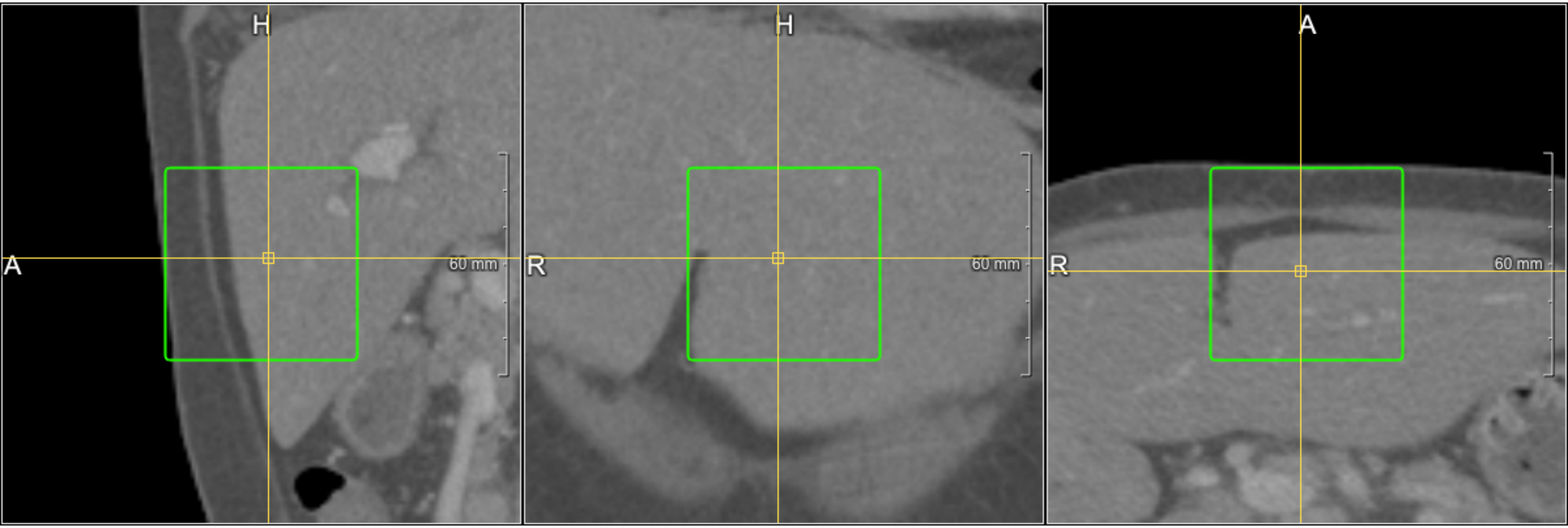
Figure from [1]

[1] Çiçek, Özgün, et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation." International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2016.

Baseline model

Receptive field:

- 92x92x92 mm^3

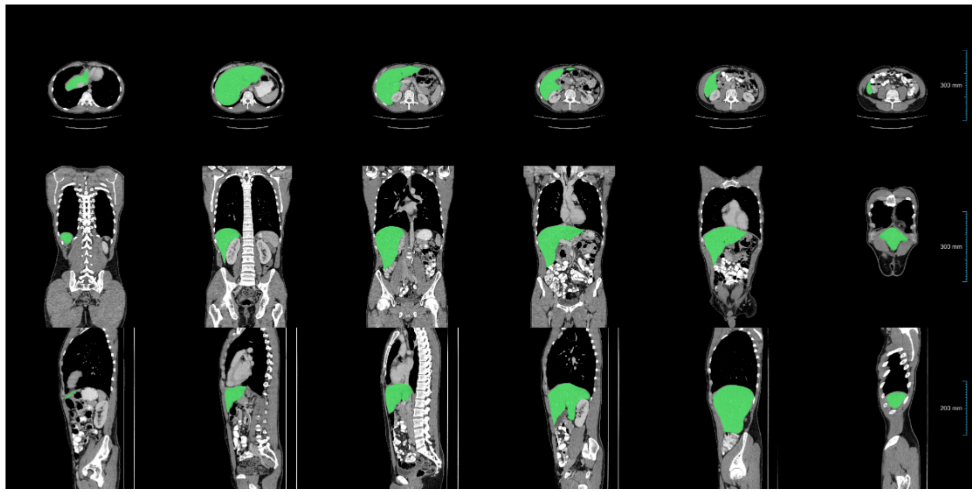


Input image patch in sagittal, coronal and axial orientation required by the baseline model to segment the region within the green box.

Evaluation

<u>Screenshot-based</u>

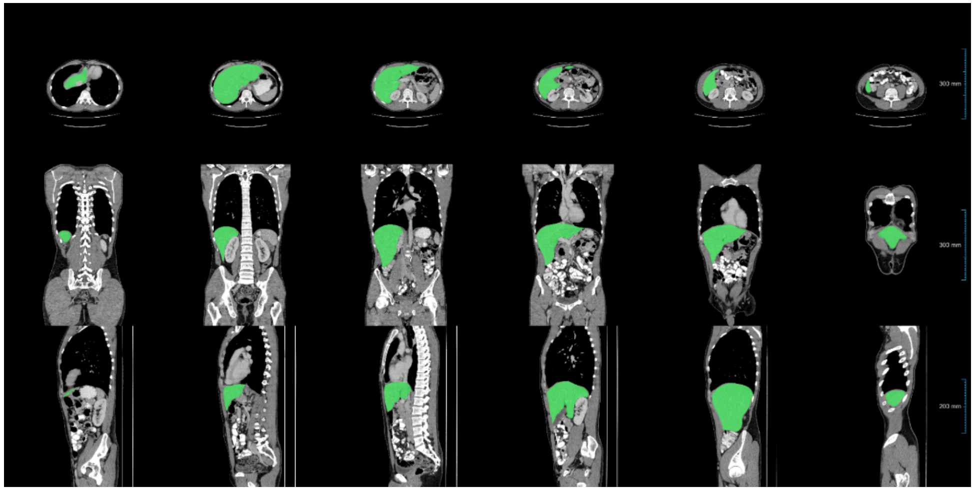
- Visual inspection by a human expert
- Scoring speed ~500 cases / hour



Evaluation

Screenshot-based

- Visual inspection by a human expert
- Scoring speed ~500 cases / hour

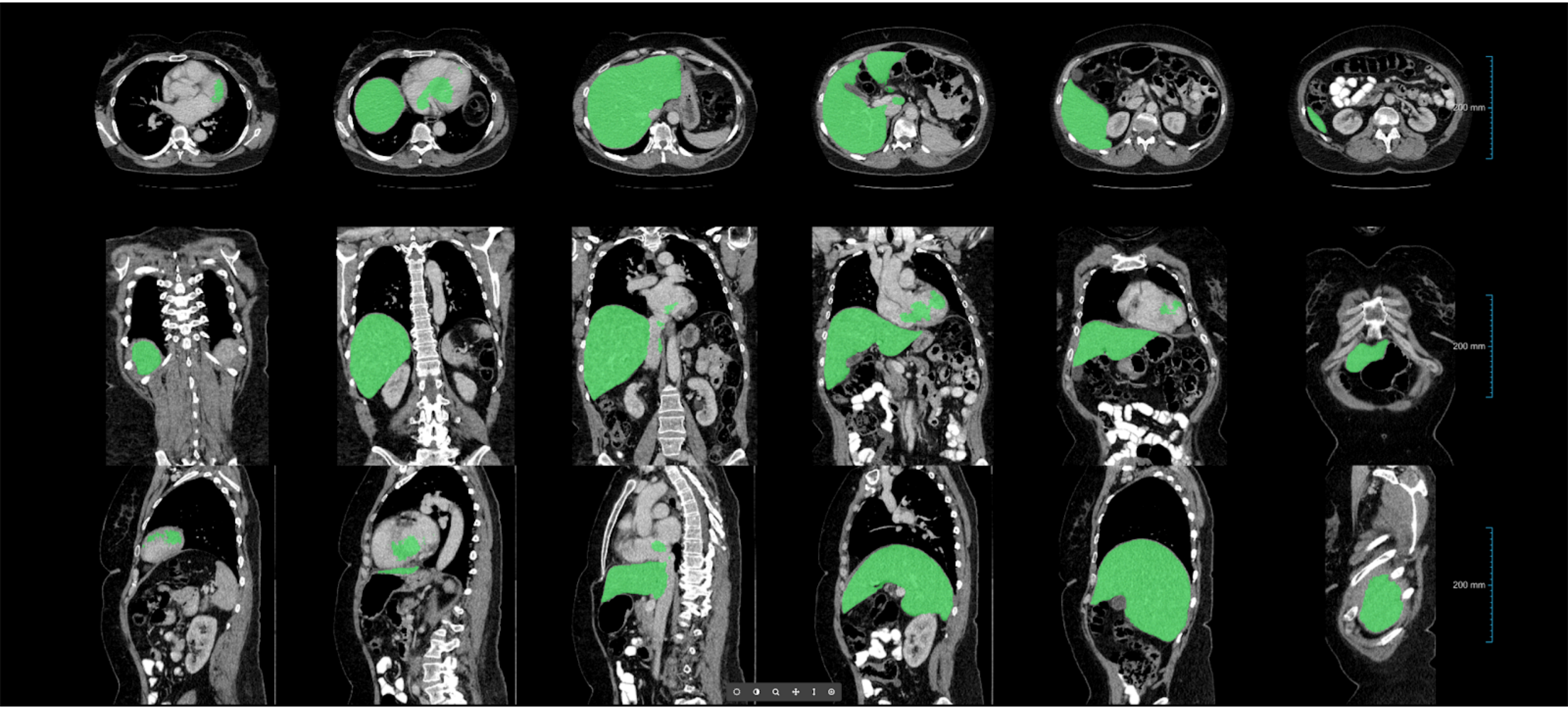


Baseline model results

- 1 - no corrections required – 516 cases (62%)
- 2 - minor corrections required – 123 cases (15%)
- 3 - major corrections required – 187 cases (23%)

Baseline model evaluation: error analysis

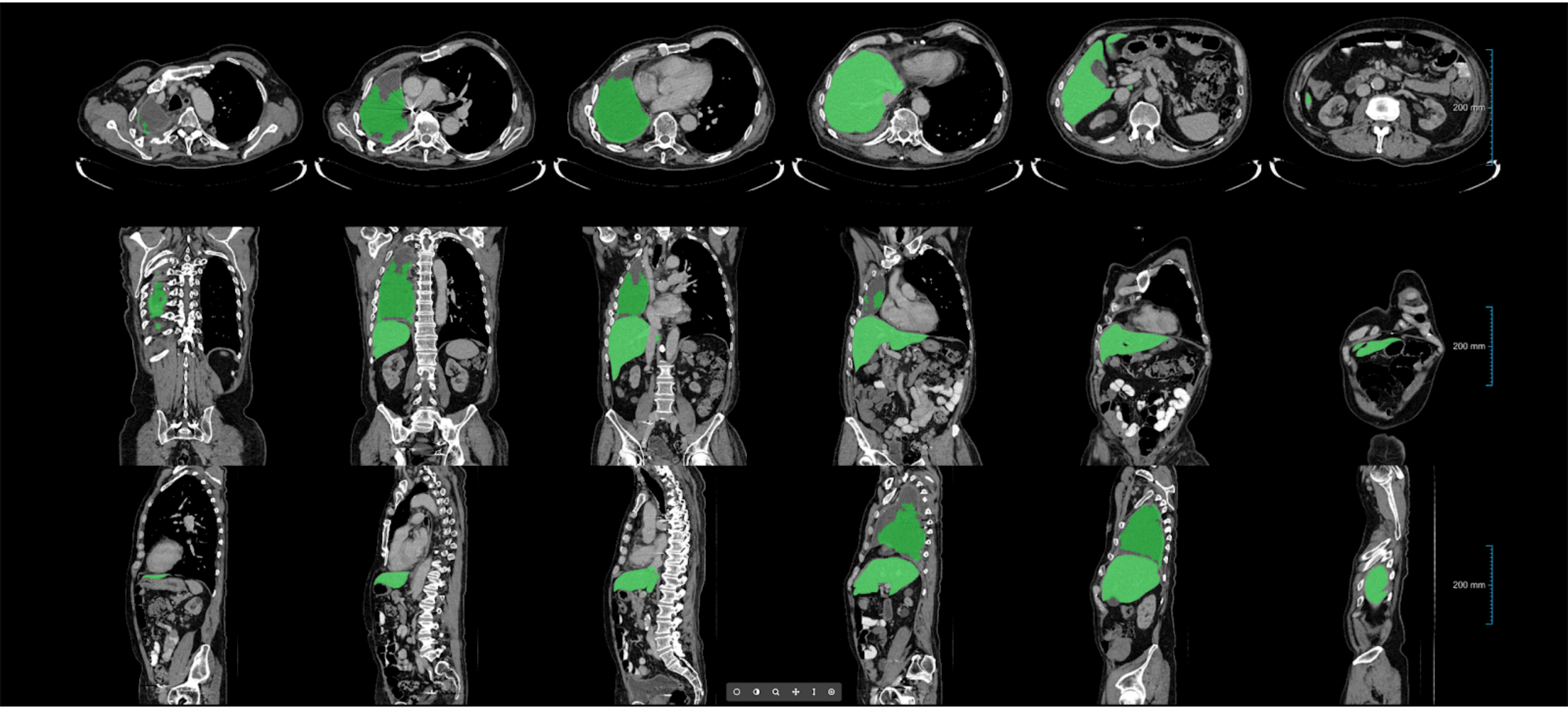
- Leakage into surrounding structures



Fraunhofer MEVIS

Baseline model evaluation: error analysis

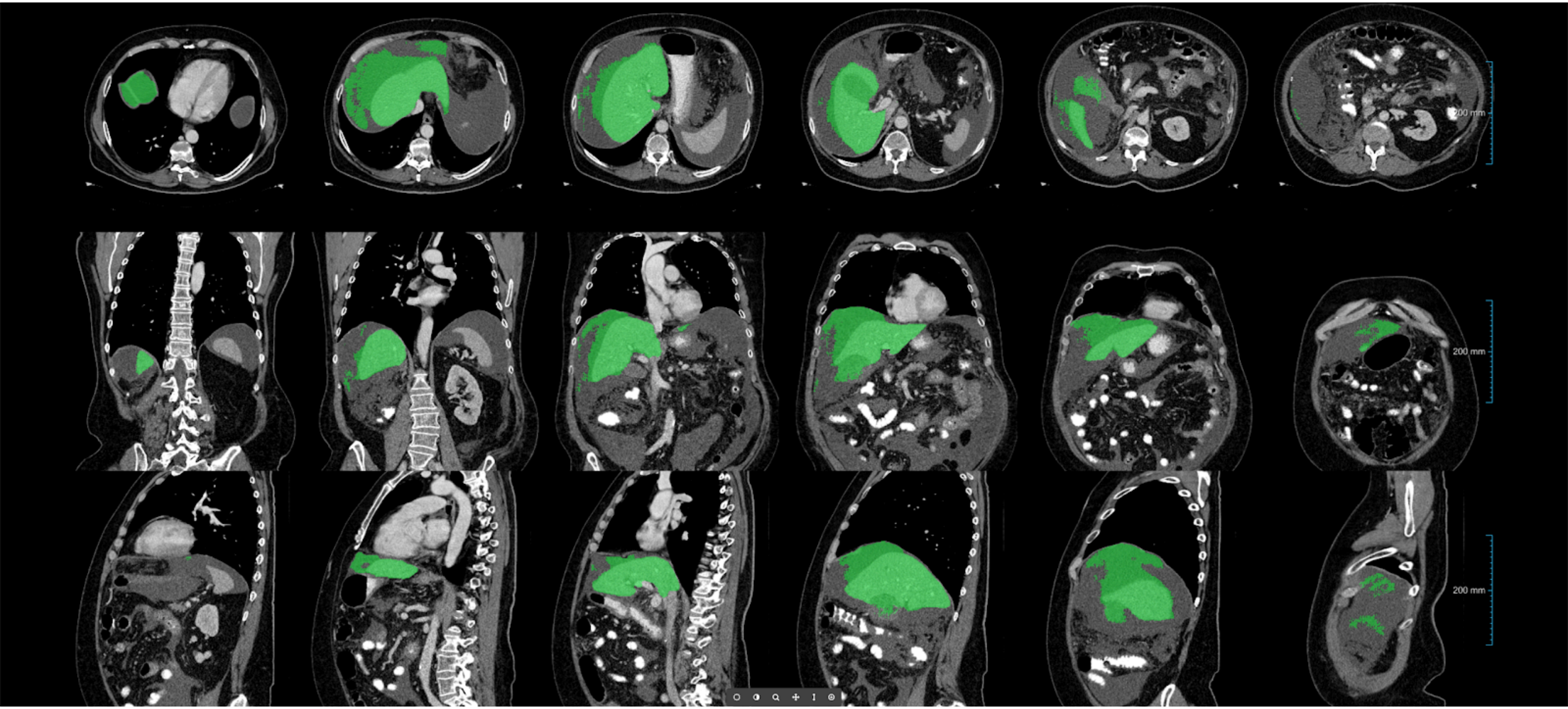
- Segmentation of surrounding pathologies



Fraunhofer
MEVIS

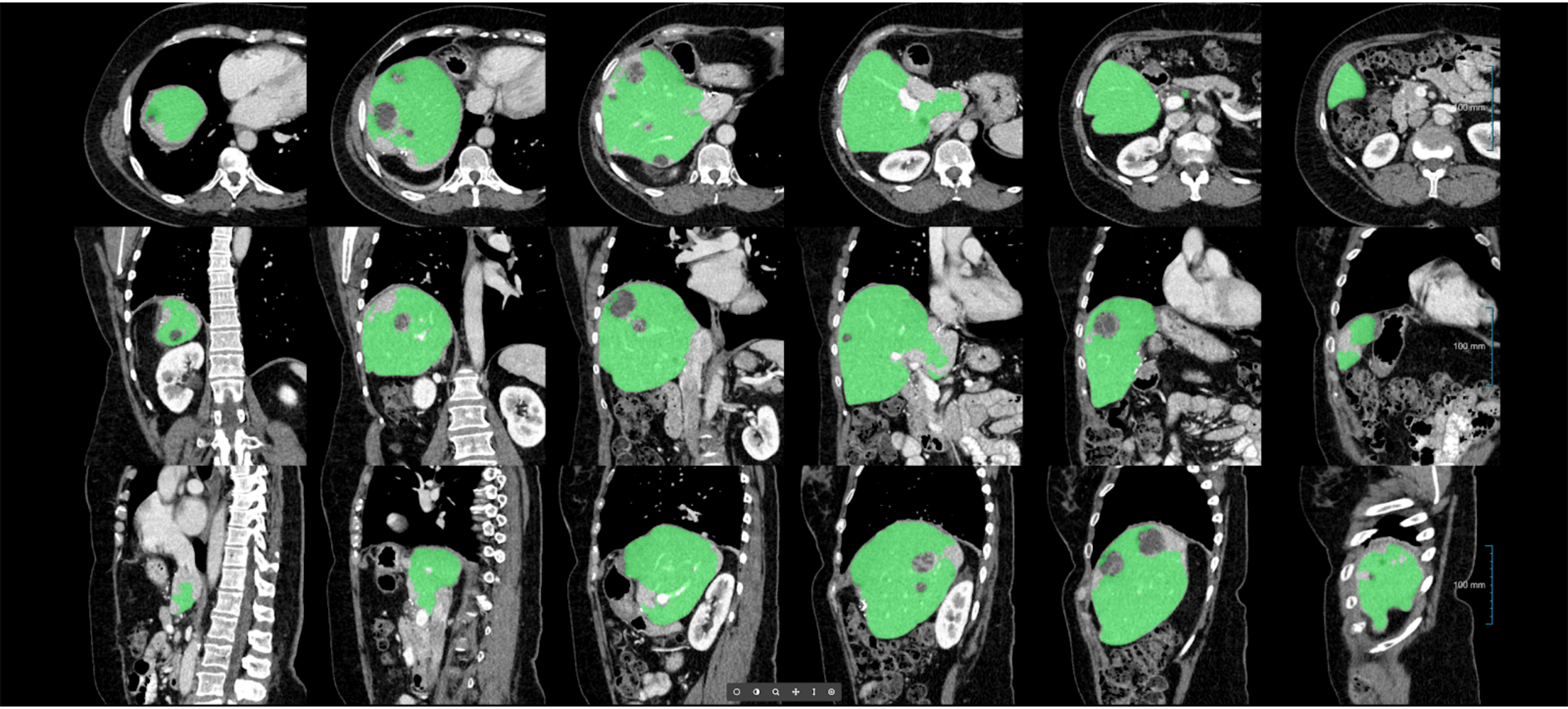
Baseline model evaluation: error analysis

- Segmentation of surrounding pathologies



Baseline model evaluation: error analysis

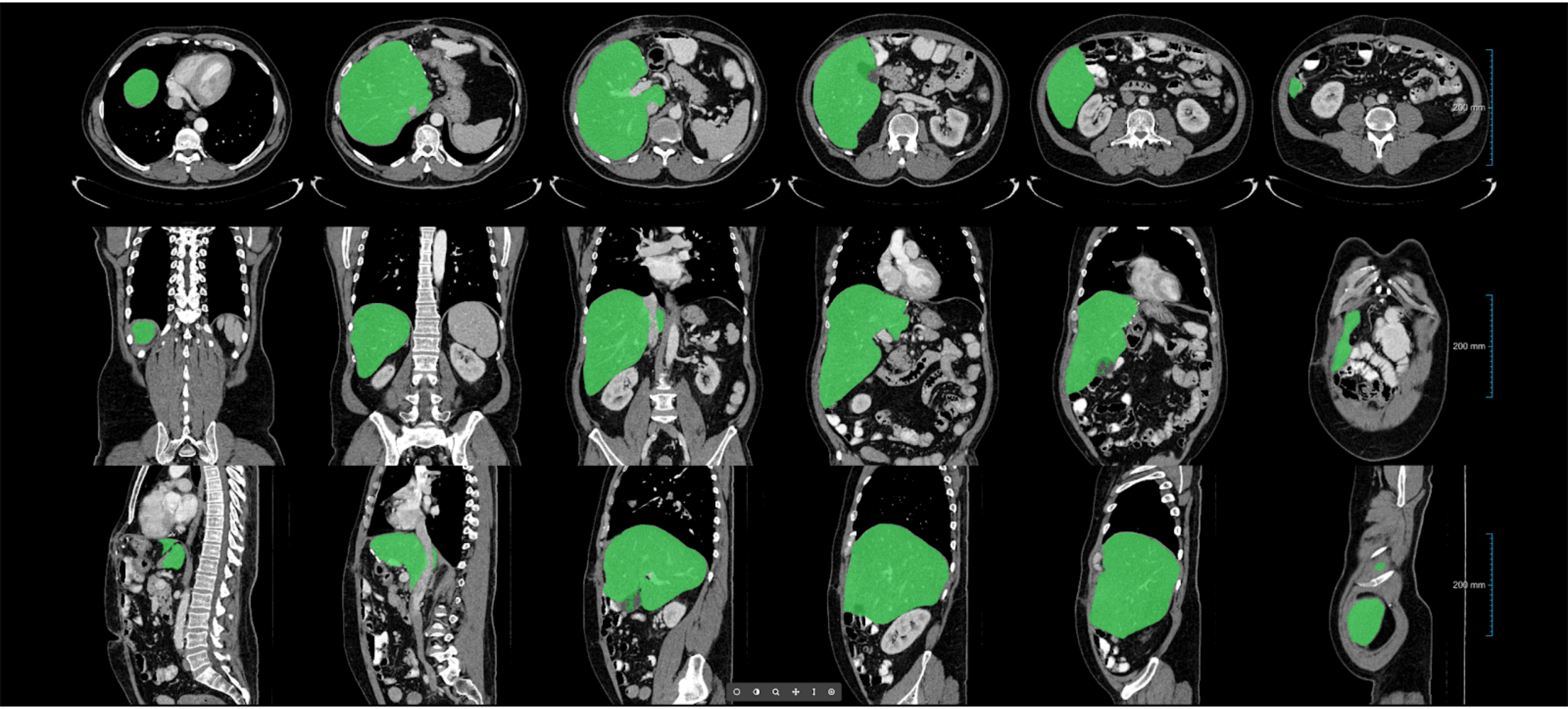
- Hypodense lesions excluded from the liver mask



Fraunhofer
MEVIS

Baseline model evaluation: error analysis

- Gallbladder partly included in the liver mask



Fraunhofer
MEVIS

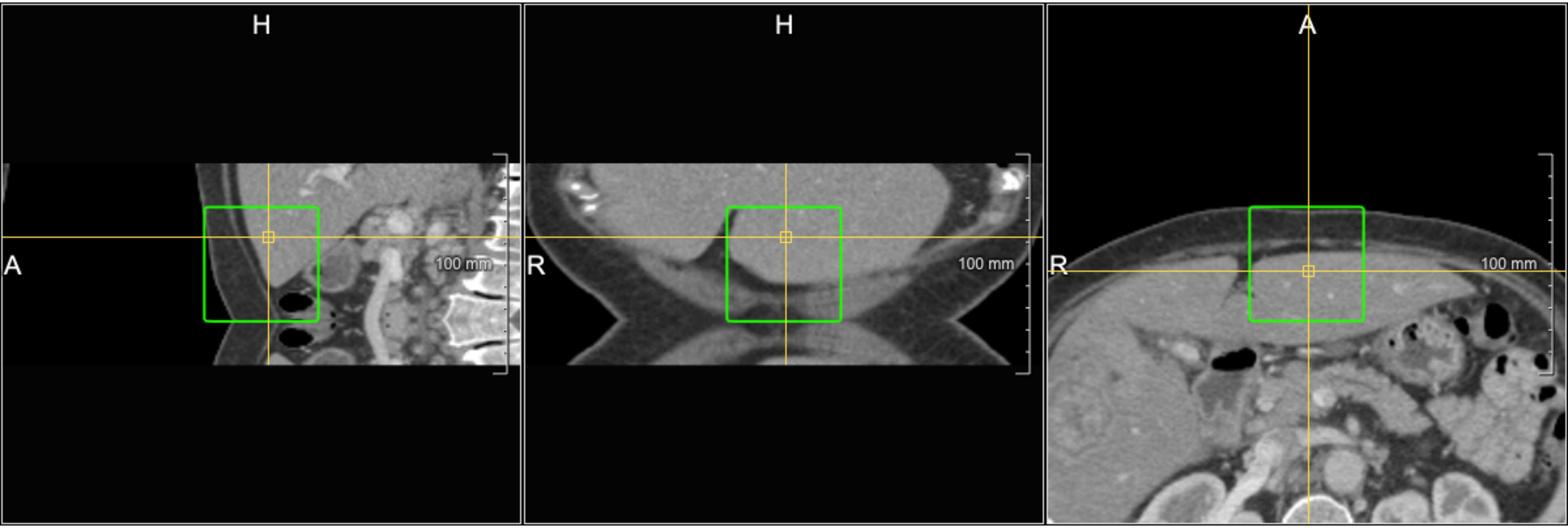
Anisotropic-net

<u>Intuition</u>

Radiologists segmenting organs typically consider a couple of neighboring slices while taking into account the whole in-plane context

3D CNN with an anisotropic receptive field (4 985 906 parameters)

- 188 x 188 x 66 mm^3

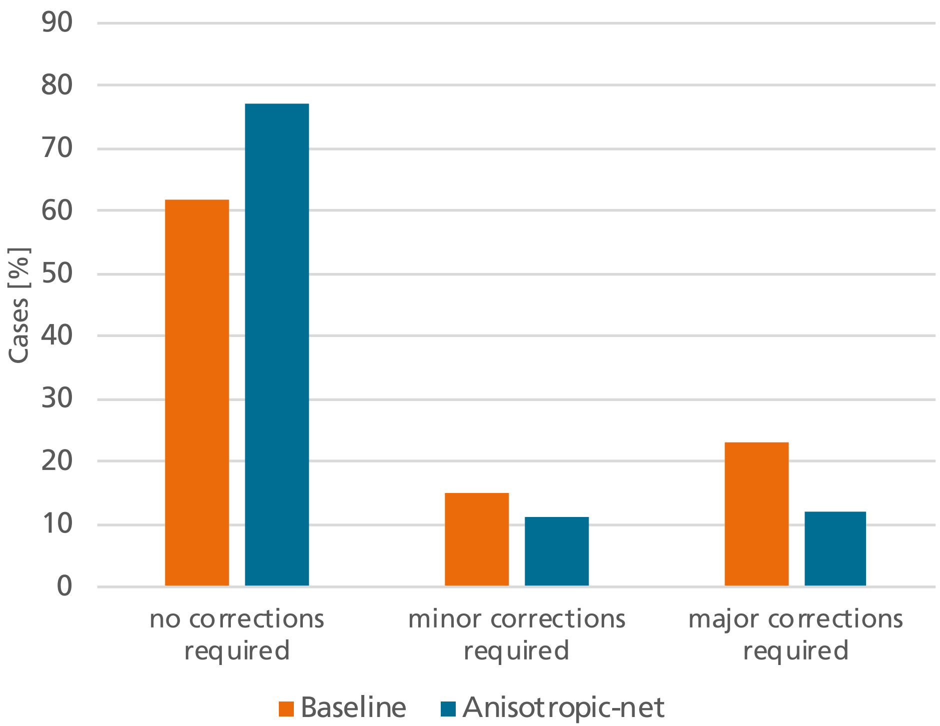


Fraunhofer MEVIS

Evaluation results

Anisotropic-net

- 1 - no corrections required – 634 cases (77%, change +15%)
- 2 - minor corrections required – 93 cases (11%, change -4%)
- 3 - major corrections required – 99 cases (12%, change -11%)



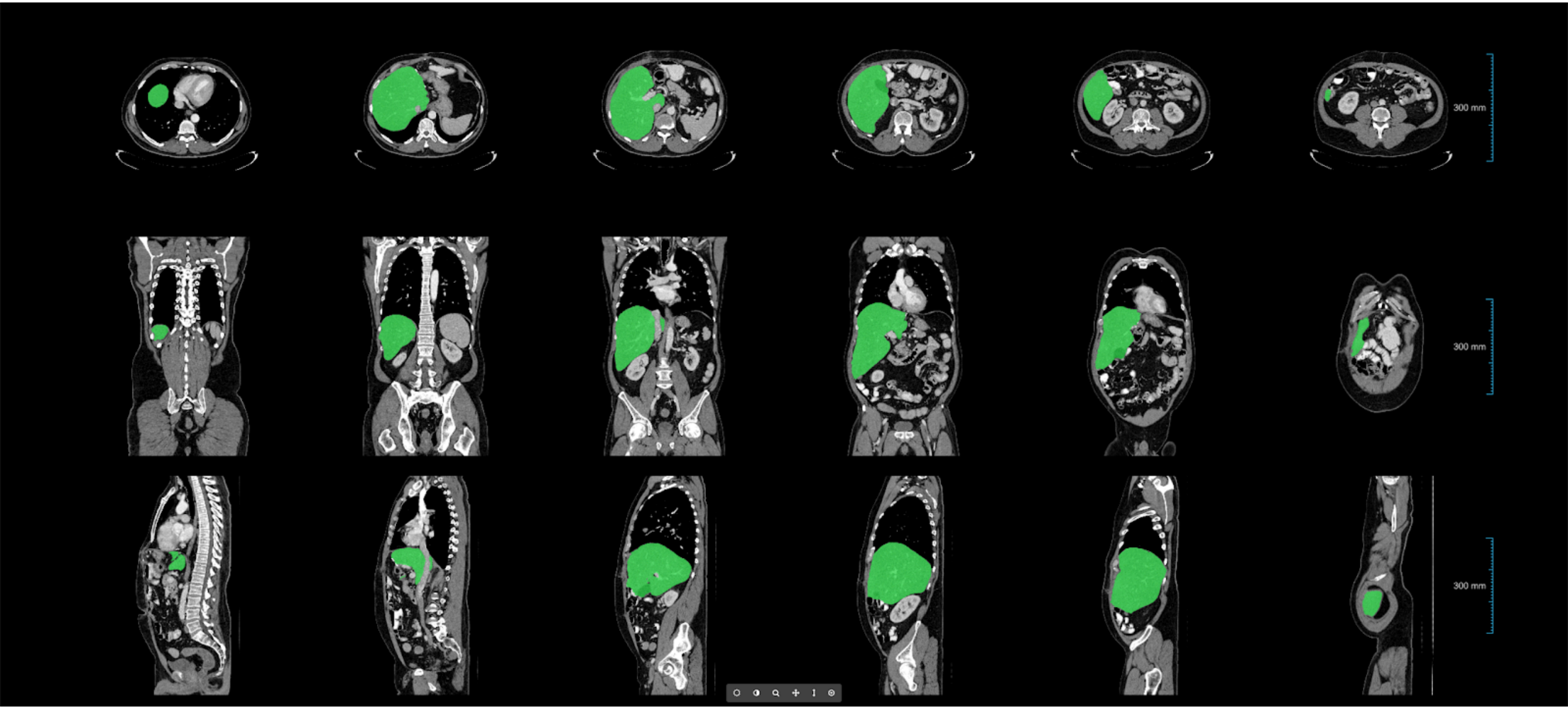
Evaluation results

3D u-net score	Anisotropic-net score	Case count
2	1	87
3	1	94
3	2	22
1	1	453
2	2	23
3	3	71
1	2	48
1	3	15
2	3	13

- 1 - no corrections required
- 2 - minor corrections required
- 3 - major corrections required

Anisotropic-net evaluation: error analysis

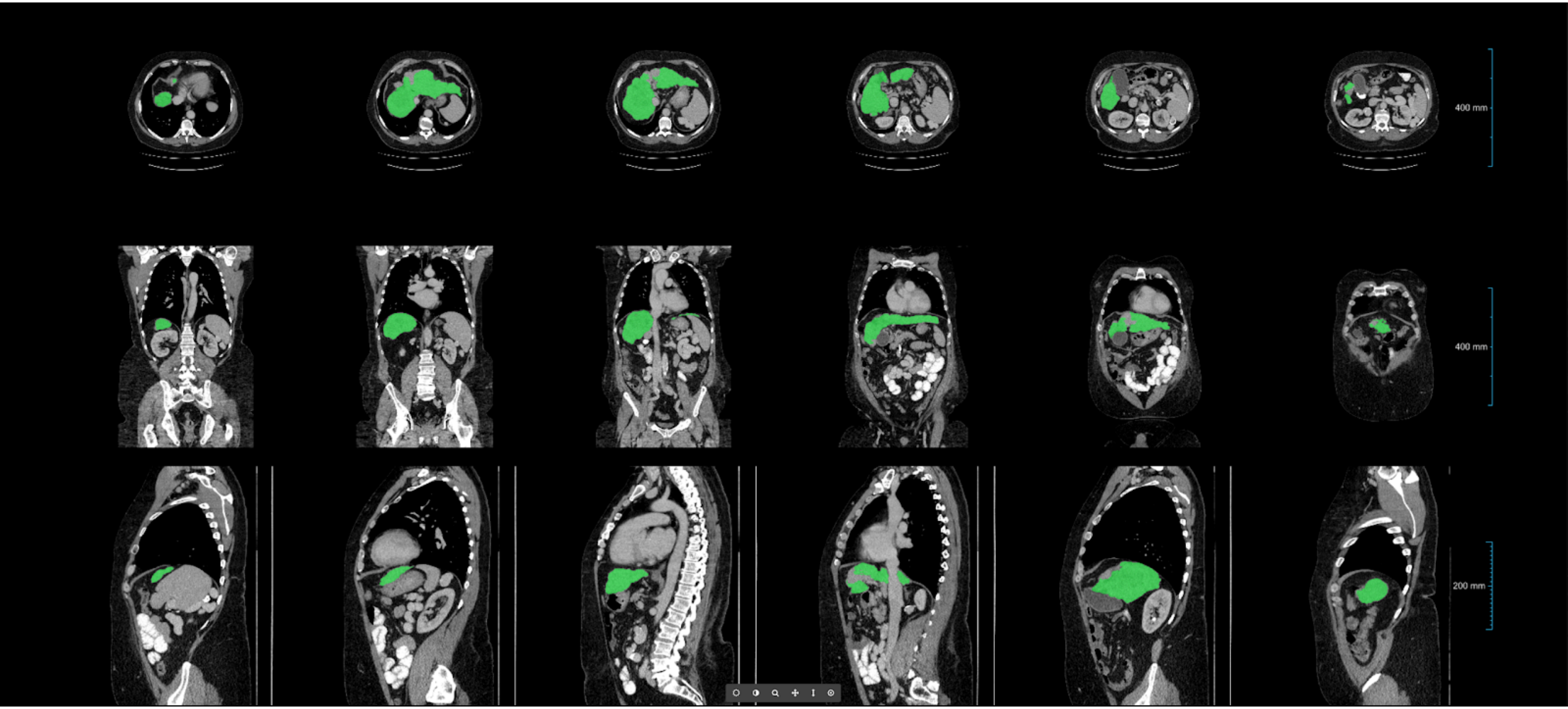
- Leakage into the gallbladder



Fraunhofer
MEVIS

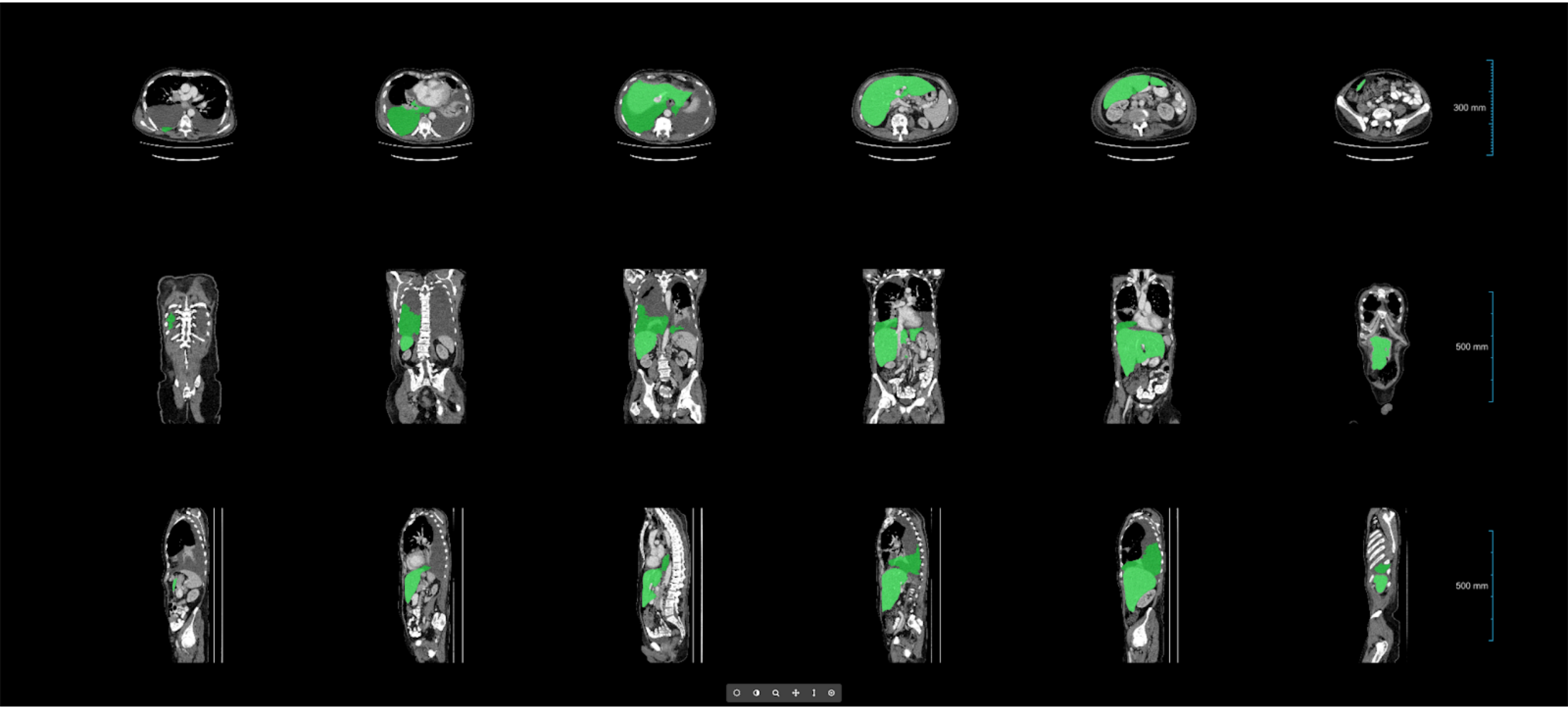
Anisotropic-net evaluation: error analysis

- Underestimation of the liver parenchyma



Anisotropic-net evaluation: error analysis

- Segmentation of surrounding pathologies



Summary

Conclusions

- 3D anisotropic-net produces less segmentations requiring corrections than the baseline model (3D u-net)
- Screenshot-based evaluation is an efficient way to score large amount of cases

Future work

- More extensive validation of the 3D anisotropic-net
- Active-learning to address remaining segmentation errors