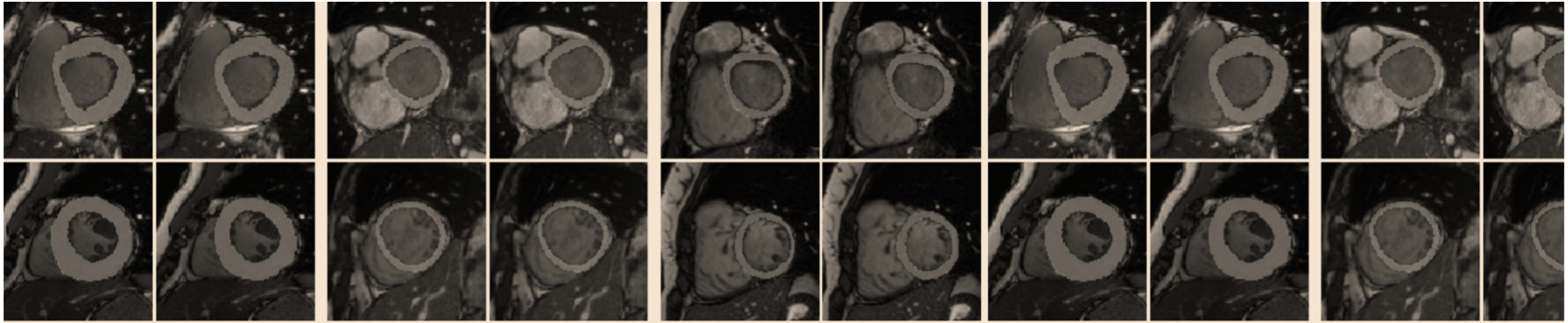
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Dr. Christian Baumgartner
Machine Learning for Medical Image Analysis Group

Cluster of Excellence: Machine Learning – New Perspectives for Science



Machine learning for medical image analysis and why clinicians are not using it

Quick Personal Introduction

- **Master of Science, ETH Zürich (2012)**
- **Doctoral Studies, King's College London @ St. Thomas Hospital (2016)**
  Title: "Manifold alignment for imaging and modelling respiratory motion" *Advisors:* Prof. Andrew King, Prof. Daniel Rueckert
- **Post-doc, Imperial College London (2016-2017)**
  Biomedical Image Analysis Group (BioMedIA)
  Advisor: Prof. Daniel Rueckert
- **Postdoctoral Fellow - ETH Fellowship, ETH Zürich (2017-2019)**
  Biomedical Image Computing Group
  Advisor: Prof. Ender Konukoglu
- **Senior Research and Development Engineer, PTC Vuforia (2020-2021)**
  Guest Researcher at ETH Zürich (2020-2021)

ETH zürich

KING'S College LONDON

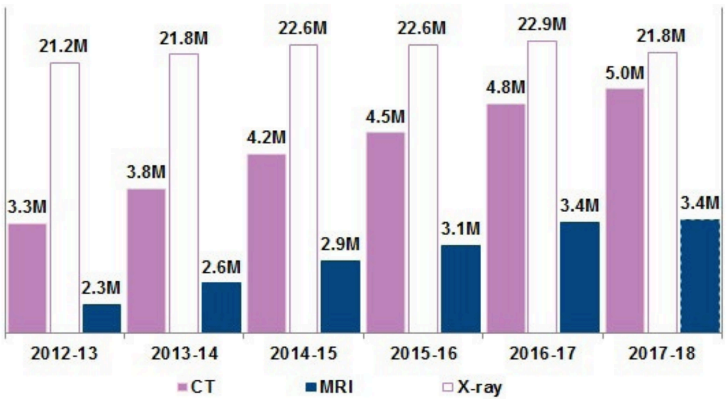
Imperial College London

ETH zürich

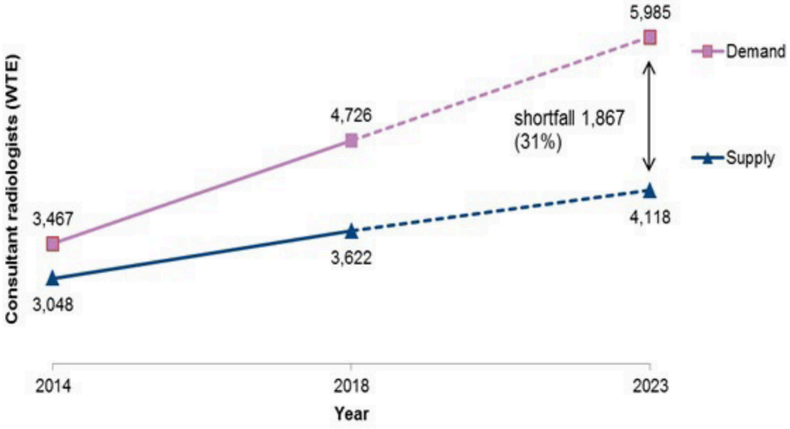
ptc / ETH zürich

Why do we need machine learning in medical image analysis?

→ **More efficient** processing of medical data



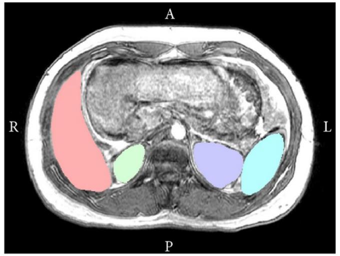
~31% increase in MR and CT images



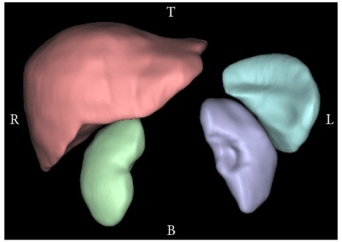
Only ~18% increase in radiologists

Why do we need machine learning in medical image analysis?

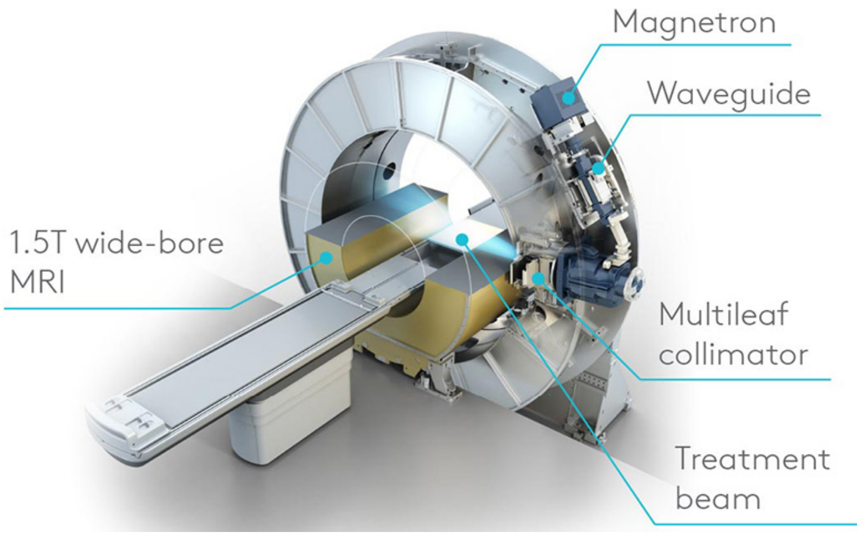
→ Personalisation of treatments, **novel workflows**



(a)

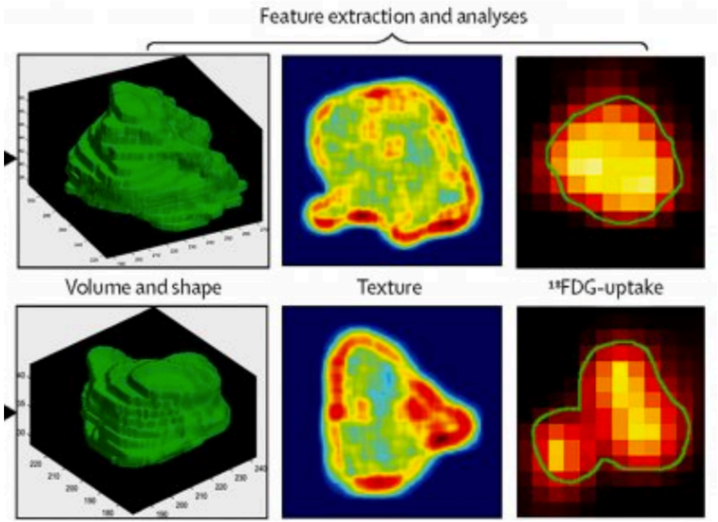


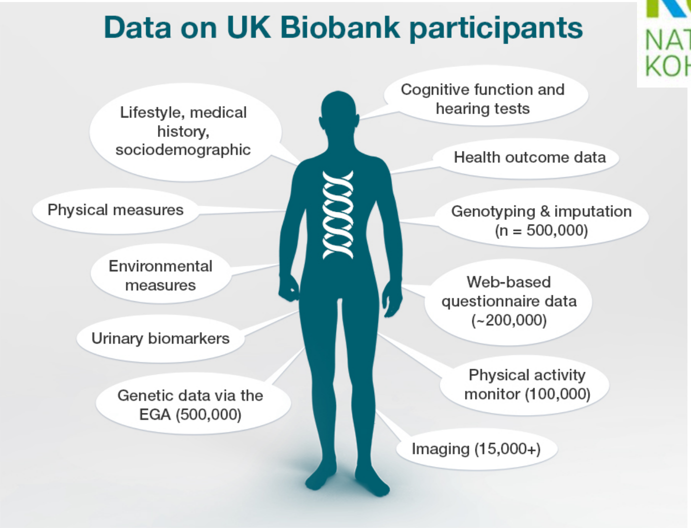
(b)



Why do we need machine learning in medical image analysis?

→ Discovery of **novel clinical biomarkers** from big medical data





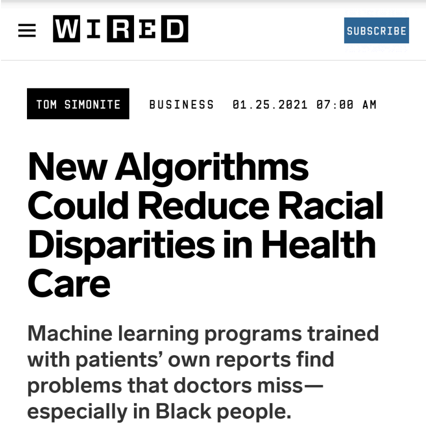
NATIONALE KOHORTE

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Why do we need machine learning in medical image analysis?

→ **Safer** medicine

- Large amount of misdiagnosis and unnecessary treatments
- In the US most people experience at least one diagnostic error in their life-time
- Physicians
  - are often overworked
  - use intuitive reasoning and "often cannot fully explain how a diagnostic hypothesis occurs to them"[1]
  - Are subject to biases (e.g. confirmation bias, self-serving bias, prejudice, ...)


WIRED

SUBSCRIBE

TOM SIMONITE BUSINESS 01.25.2021 07:00 AM

New Algorithms Could Reduce Racial Disparities in Health Care

Machine learning programs trained with patients' own reports find problems that doctors miss—especially in Black people.

[1]Brush, J. E., & Brophy, J. M. (2017). Sharing the Process of Diagnostic Decision Making. JAMA Internal Medicine, 177(9), 1245

Deep Learning on Medical Images

Human-level CMR image analysis with deep fully convolutional networks

Wenjia Bai[1*] Matthew Sinclair[1] Giacomo Tarroni[1] Ozan Oktay[1] Martin Rajchl[1] Ghislain Vaillant[1] Aaron M. Lee[2] Nay Aung[2] Elena Lukaschuk[3] Mihir M. Sanghvi[2] Filip Zemrak[2] Kenneth Fung[2] Jose Miguel Paiva[2] Valentina Carapella[3] Young Jin Kim[3] Hideaki Suzuki[4] Bernhard Kainz[1] Paul M. Matthews[4] Steffen E. Petersen[2] Stefan K. Piechnik[3] Stefan Neubauer[3] Ben Glocker[1] Daniel Rueckert[1]

Attaining human-level performance with atlas location autocontext for anatomical landmark detection in 3D CT data

Alison Q. O'Neil[1], Antanas Kascenas[1], Joseph Henry[1], Daniel Wyeth[1], Matthew Shepherd[1], Erin Beveridge[1], Lauren Clunie[1], Carrie Sansom[1], Evelina Šeduikytė[1] Keith Muir[2], and Ian Poole[1]

Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks

Matthew Sinclair, Christian F. Baumgartner, Jacqueline Matthew, Wenjia Bai, Juan Cerrolaza Martinez, Yuanwei Li, Sandra Smith, Caroline L. Knight, Bernhard Kainz, *Senior Member, IEEE,* Jo Hajnal, Andrew P. King, Daniel Rueckert, *Fellow, IEEE*

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Is the problem solved?

Geoff Hinton in 2016:

“People should stop training radiologists now. It’s just completely obvious that within 5 years deep learning is going to do better than radiologists.”

Reality

As of October 2019 there are 28 medical AI solutions operating on image data that have been approved by the food and drug administration (FDA)

"The Lancet" meta-study

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu, MBChB † • Livia Faes, MD † • Aditya U Kale, MBChB • Siegfried K Wagner, BMBCh • Dun Jack Fu, PhD • Alice Bruynseels, MBChB • et al. Show all authors • Show footnotes

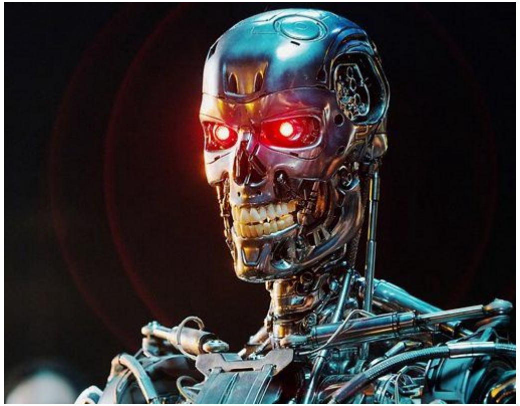
- Little evidence comparing performance of humans and machines
- Often not comparing performance in clinical environments
- **The true diagnostic power of AI remains uncertain**

Why isn't AI used more in clinical practice?

Technical challenges
Regulatory challenges

Human factors

Human factors



Physicians’ fears about AI:

- Interferes with ability to make independent diagnoses
- Hurts relationship with patients
- It’s a form of management control

- Med students who know the technology are less afraid of it
- Less than one-third of med students knows AI basics
- Med students get more AI information from media than lectures

➢ **Solution: Better education at student and professional level**

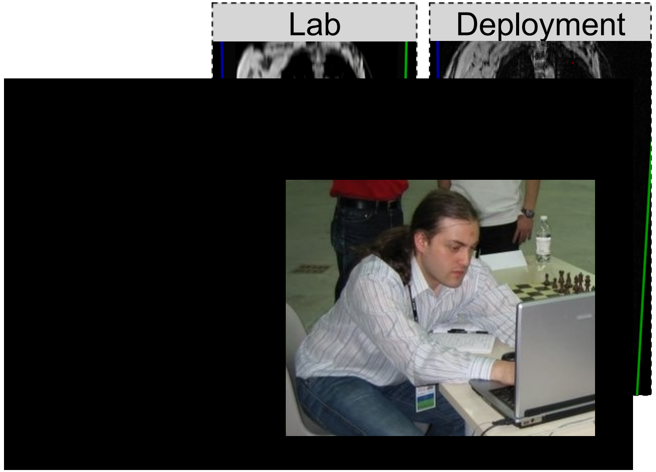
➢ **Better involvement of medical community in AI research**

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

Safi, Sabur, Thomas Thiessen, and Kurt JG Schmailzl. "Acceptance and resistance of new digital technologies in medicine: qualitative study." *JMIR research protocols*
Gallix, Benoit, and Jaron Chong. "Artificial intelligence in radiology: who’s afraid of the big bad wolf?." *European radiology* 29.4 (2019)

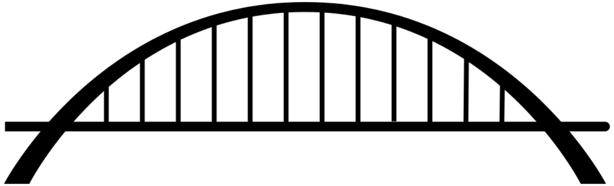
Technical challenges

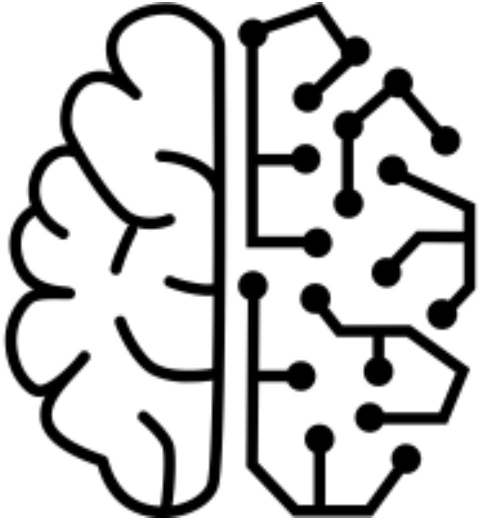
- Deep networks lack robustness + non-standardised medical image acquisition
- Annotated data scarce and expensive
- Medical images are often 3D or even 3D+t
- Human vs. AI approach, rather than human **with** AI



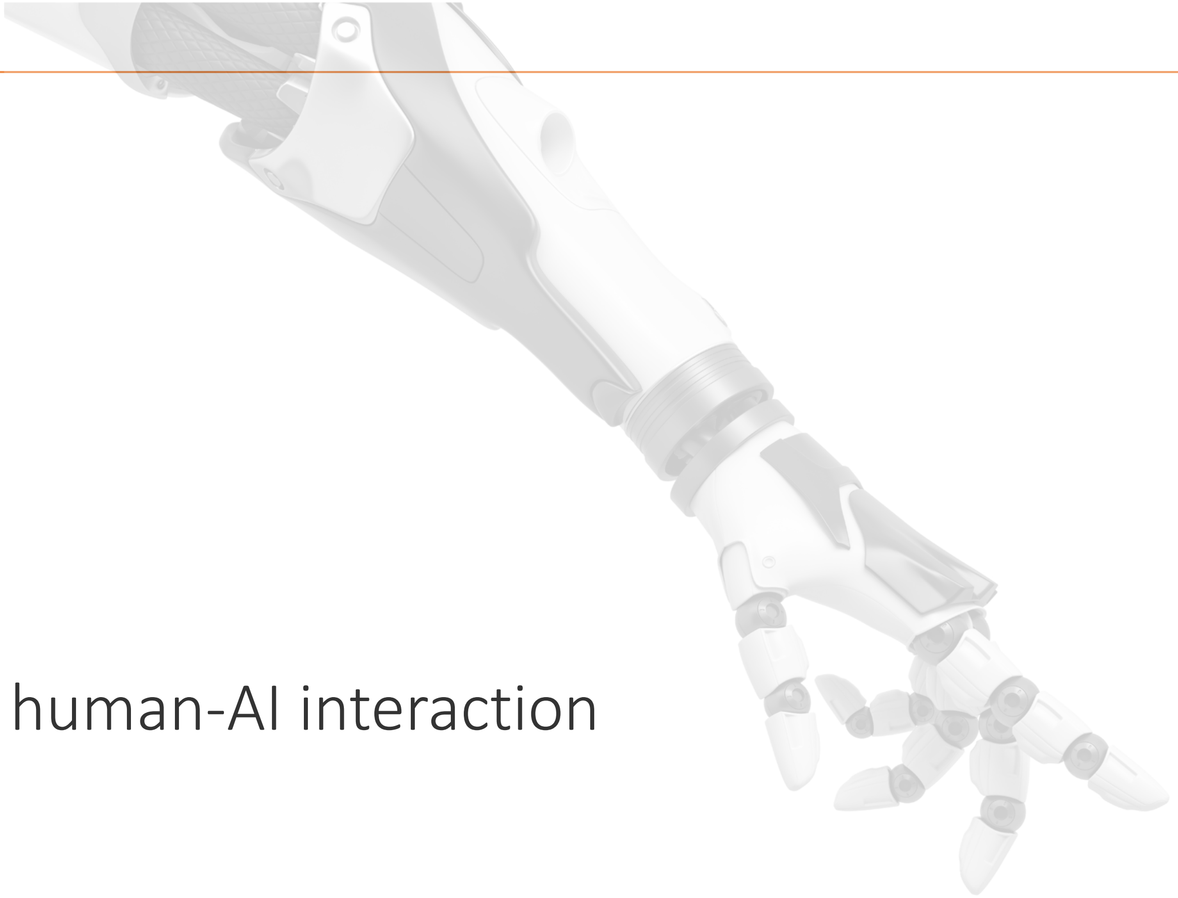
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

My Research





EBERHARD KARLS UNIVERSITÄT TÜBINGEN

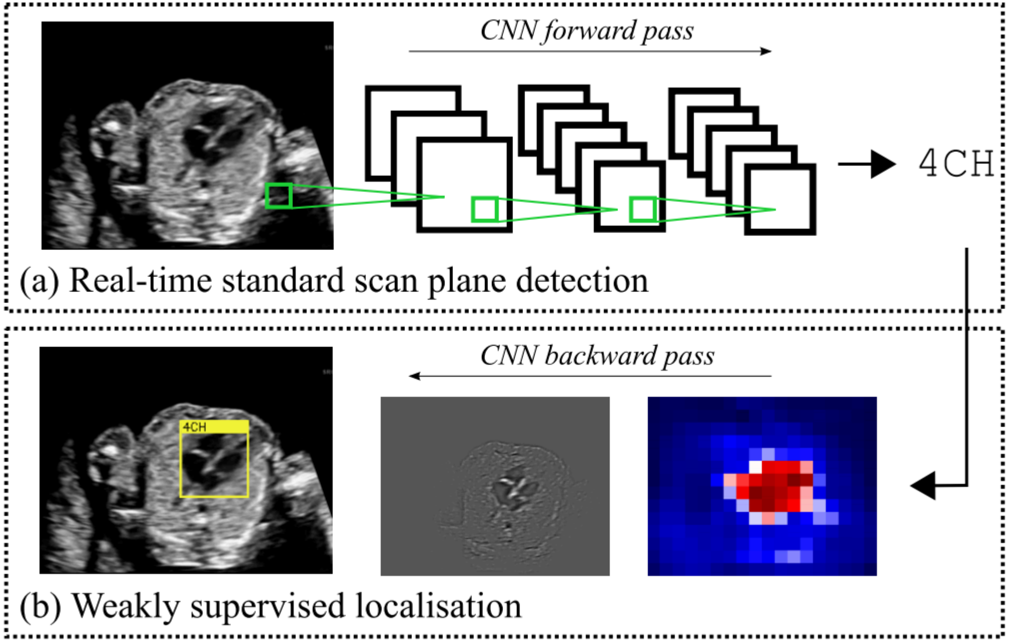


Technologies for better human-AI interaction

Interpretability of fetal scan plane detection network

Problem: Identify correct standard view planes in 20 weeks old fetus

Scan plane classification and associated saliency-map explanations



Real-Time Detection and Localisation of Fetal Standard Scan Planes in 2D Freehand Ultrasound

Christian F. Baumgartner, Konstantinos Kamnitsas, Jacqueline Matthew, Tara P. Fletcher, Sandra Smith, Lisa M. Koch, Bernhard Kainz and Daniel Rueckert

Video Demonstration

-

Overview of proposed method

Interpreting a classifier vs. understanding a class

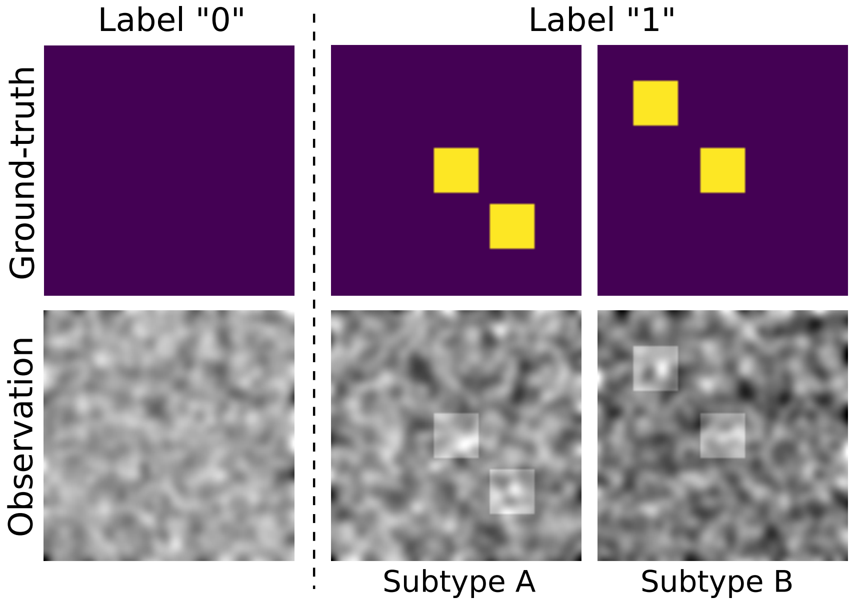
- Saliency maps show us which pixels a classifier is paying attention to
- Saliency maps **do not** show us all pixels characterising a class

➢ **Take away:** We should not use saliency maps for weakly supervised localisation

➢ In this work we investigated an alternative way of identifying *all* pixels belonging to a class

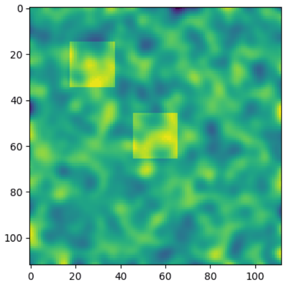
Baumgartner et al., *Visual Attributions using Wasserstein GANs*, Proc. CVPR (2018)

Motivation: toy classification problem

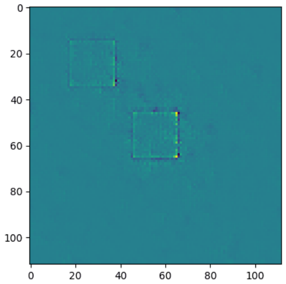


Result of Guided-Backpropagation

Input Image:



Saliency map:

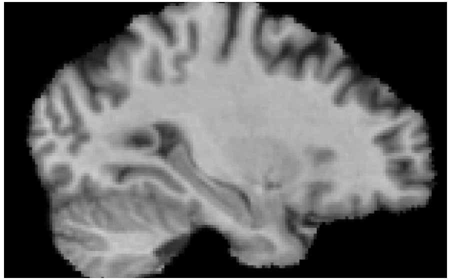


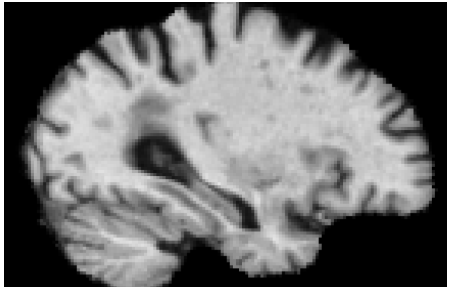
Classifier focuses on minimum features required for classification!

Alternative method for feature attribution not involving classifiers

Assume our image data are samples from two probability distributions:

$$p(x|c=0) \qquad\qquad p(x|c=1)$$



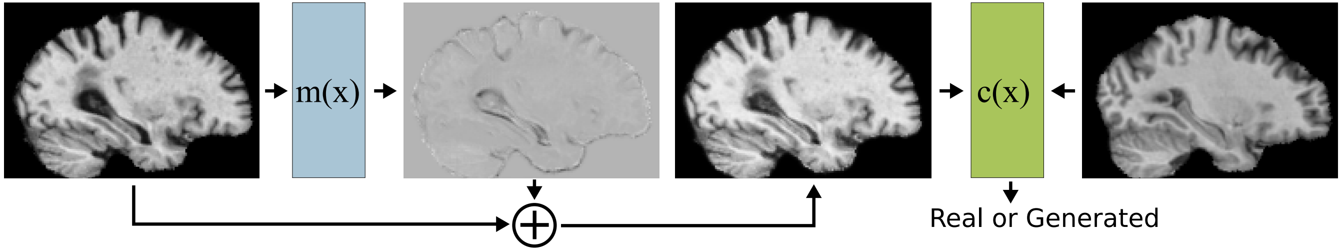


c=0 (mild cognitive impairment)
c=1 (Alzheimer's disease)

Can we find a function *m(x)* that when added to a healthy image, will make it look like a diseased image?

Finding the class-related pixels using Wasserstein GANs

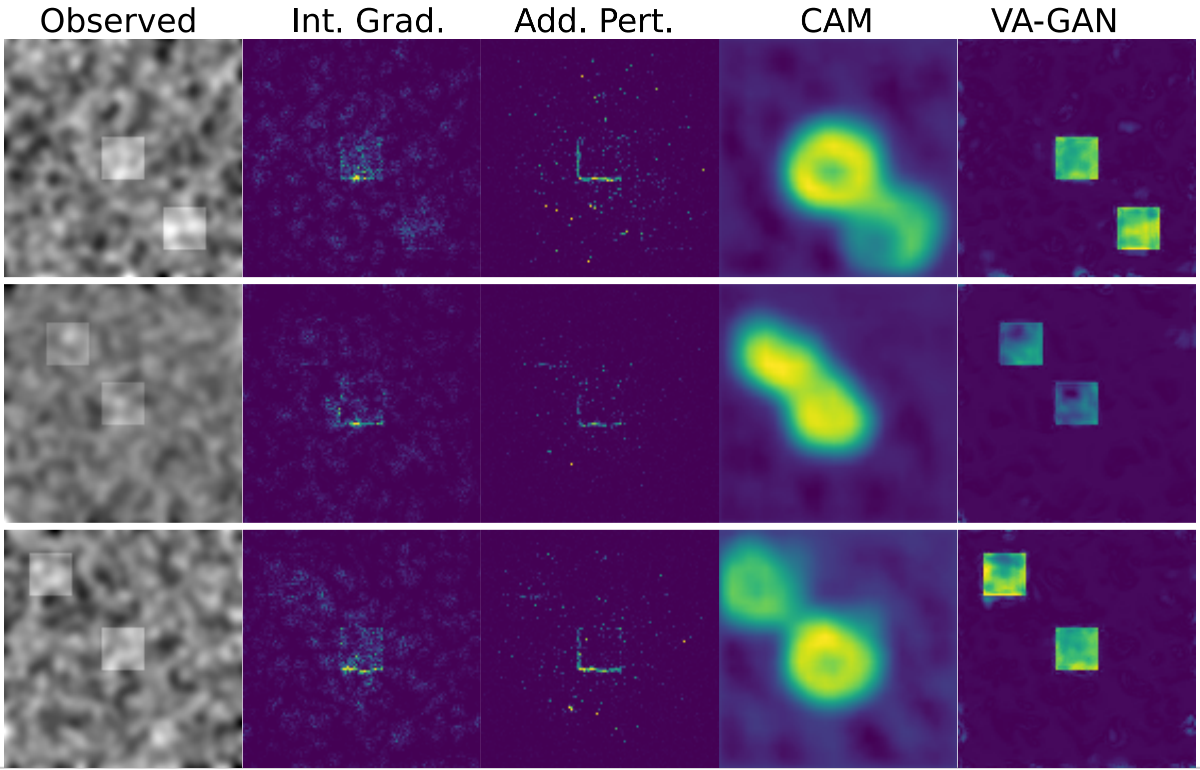
WGANs are a flavour of GANs that can be shown to minimize the Wasserstein-1 or Earth Movers Distance between the probability distributions.



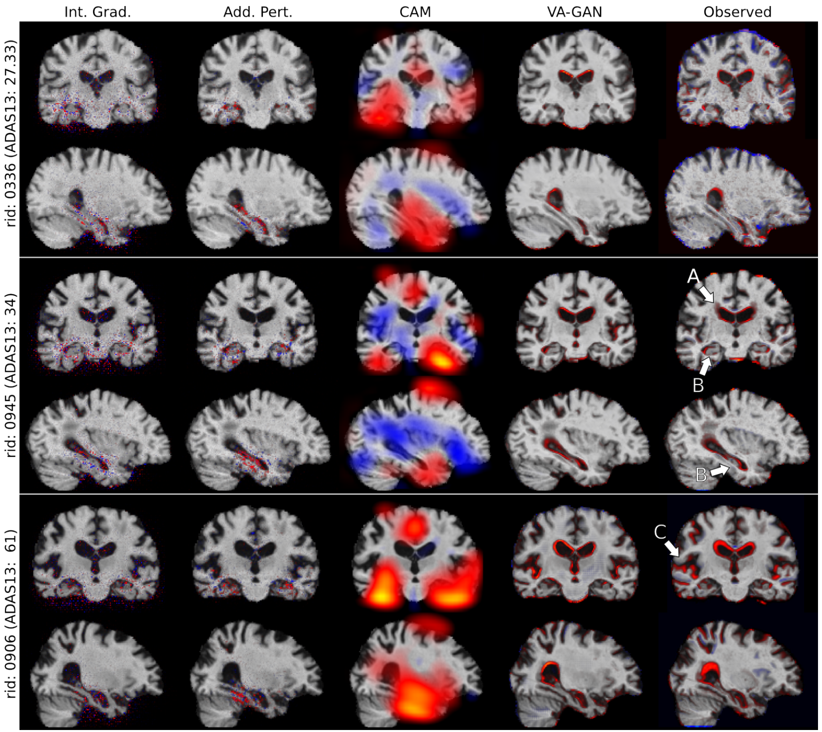
$$\mathcal{L}_{GAN}(\theta,\phi) = \mathbb{E}_{x\sim p_d(x|c=0)}[c_\phi(x)] - \mathbb{E}_{x\sim p_d(x|c=1)}[c_\phi(x+m_\theta(x))]$$

$$\theta^* = \arg\min_\theta \max_\phi \mathcal{L}_{GAN}(\theta,\phi) + \lambda\mathcal{L}_{reg}(\phi),$$

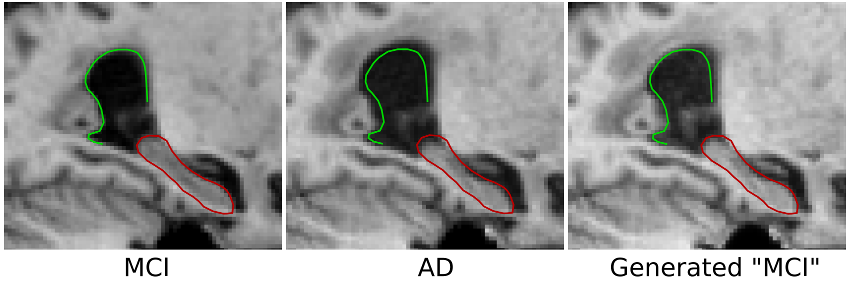
Experiments on toy data



Experiments on Alzheimer Brain Data

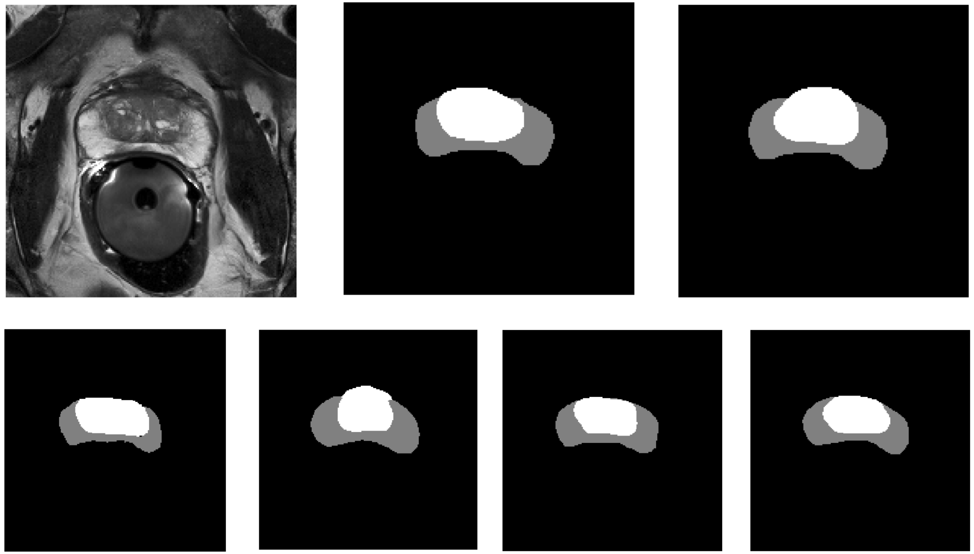


Close-up of Hippocampus



Inter-operator uncertainty quantification

Segmentation of prostate transitional and peripheral zones

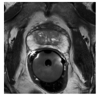


Baumgartner et al., *PHiSeg: Capturing uncertainty in medical image segmentation*, Proc MICCAI (2019)

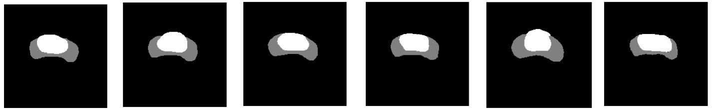
We would like to model the distribution

$$p(s|x)$$

x

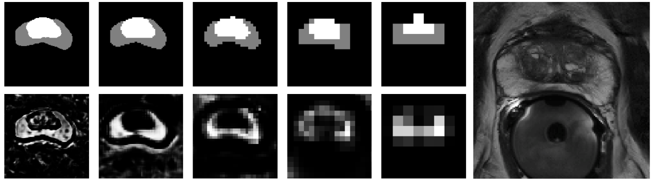


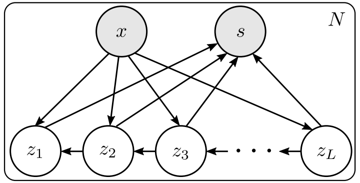
s



$\cdots$

We postulate a hierarchical generative process for the images





$$p(s|x) = \int p(s|z_1, \ldots, z_L)p(z_1|z_2, x) \cdots p(z_{L-1}|z_L, x)p(z_L|x)dz_1 \cdots dz_L$$

Variational Inference in our conditional hierarchical model

$$p(s|x) = \int p(s|z_1, \ldots, z_L)\underline{p(z_1|z_2, x) \cdots p(z_{L-1}|z_L, x)p(z_L|x)}dz_1 \cdots dz_L$$

Use q(z|s,x) to approximate *p(z|x)*

$$\log p(\mathbf{s}|\mathbf{x}) = \mathcal{L}(\mathbf{s}|\mathbf{x}) + \mathrm{KL}(q(\mathbf{z}|\mathbf{s}, \mathbf{x})||p(\mathbf{z}|\mathbf{s}, \mathbf{x}))$$

Evidence lower bound (ELBO)

Variational Inference in our conditional hierarchical model

We can use a variational approximation *q(z|s,x)* to approximate *p(z|x).*

$$\ln p(s|x) \geq \mathcal{L}(s|x)$$

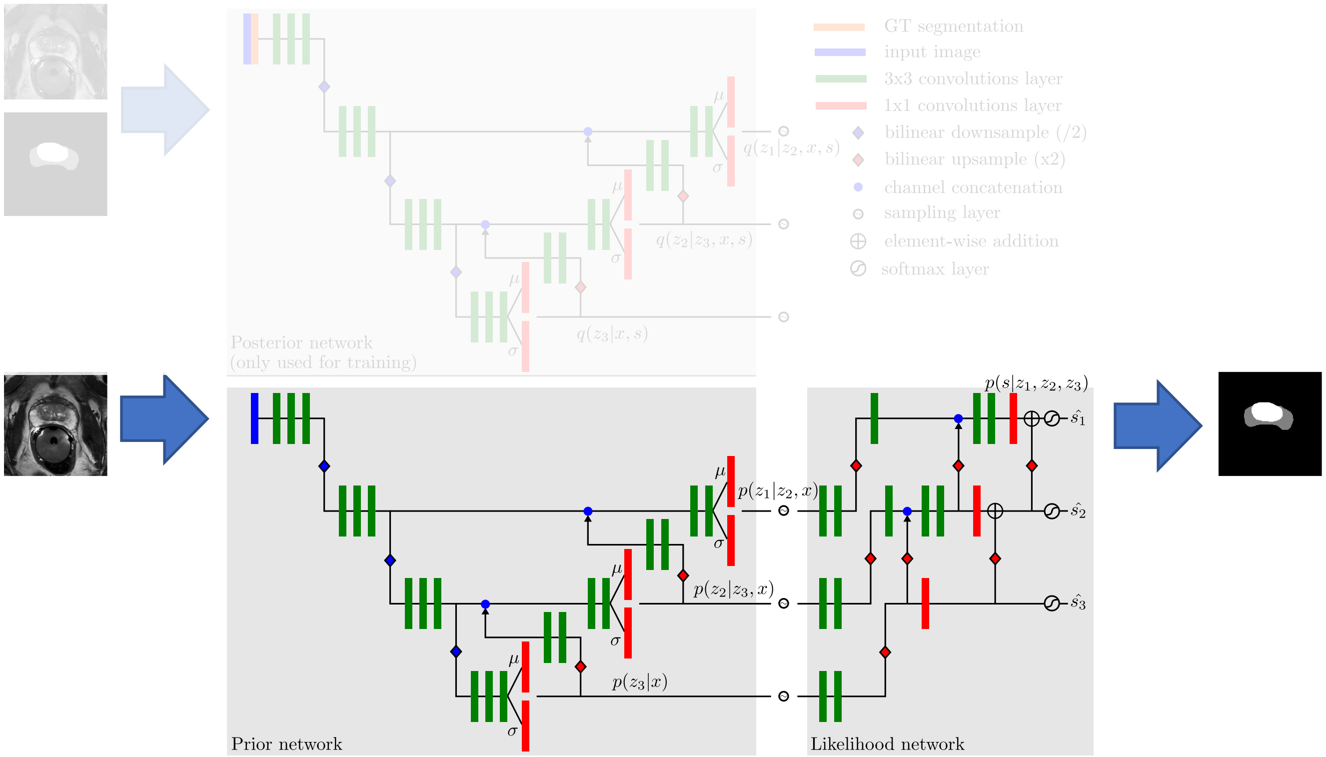
$$\begin{aligned}\mathcal{L} =& \mathrm{E}_{q(z_1,\ldots,z_L|x,s)}\left[\log p(s|z_1,\ldots,z_L)\right] - \alpha_L \mathrm{KL}\left[q(z_L|s,x)||p(z_L|x)\right] \\ & - \sum_{\ell=1}^{L-1} \alpha_\ell \mathrm{E}_{q(z_{\ell+1}|s,x)}\left[\mathrm{KL}\left[q(z_\ell|z_{\ell+1},s,x)||p(z_\ell|z_{\ell+1},x)\right]\right],\end{aligned}$$

$$p(z_\ell|z_{\ell+1},x) = \mathcal{N}\left(z_\ell|\phi_\ell^{(\mu)}(z_{\ell+1},x),\phi_\ell^{(\sigma)}(z_{\ell+1},x)\right)$$

$$q(z_\ell|z_{\ell+1},x,s) = \mathcal{N}\left(z_\ell|\theta_\ell^{(\mu)}(z_{\ell+1},s,x),\theta_\ell^{(\sigma)}(z_{\ell+1},s,x)\right)$$

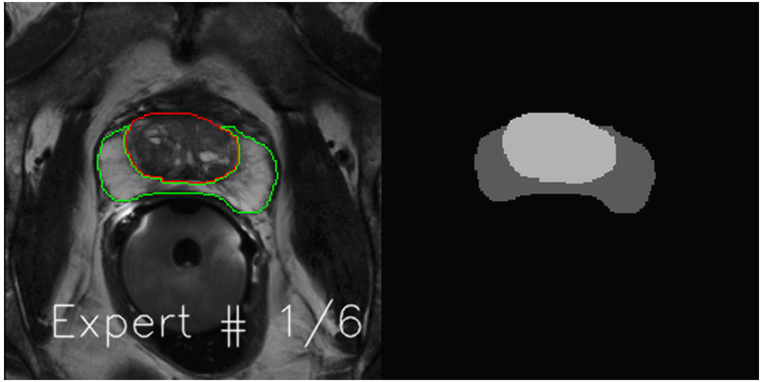
$$p(s|z) = \mathrm{Cat}(s|\pi(z))$$

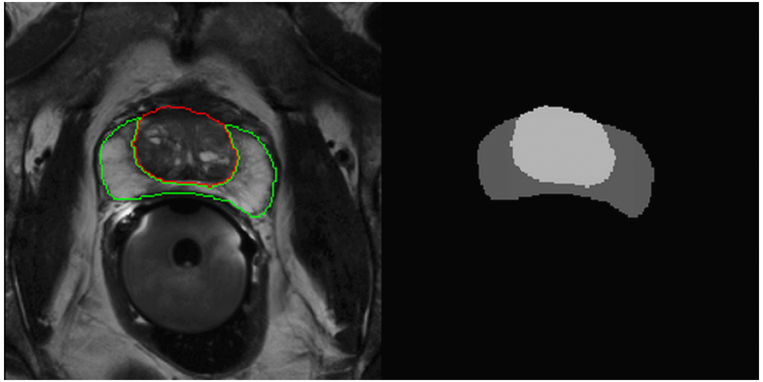
Implementing the hierarchical model with neural networks

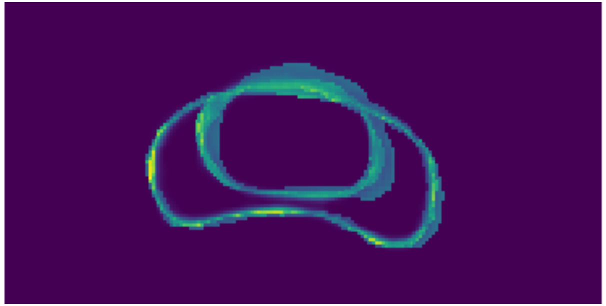


UNIVERSITÄT TÜBINGEN
EBERHARD KARLS

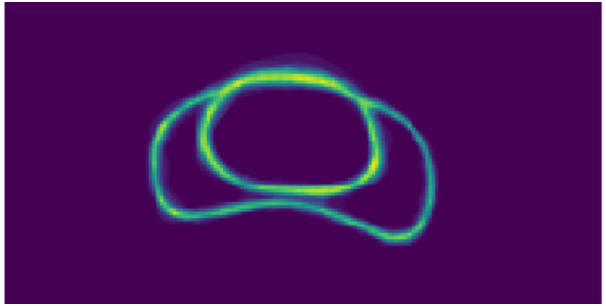
Learned uncertainty can be displayed as samples or as uncertainty map







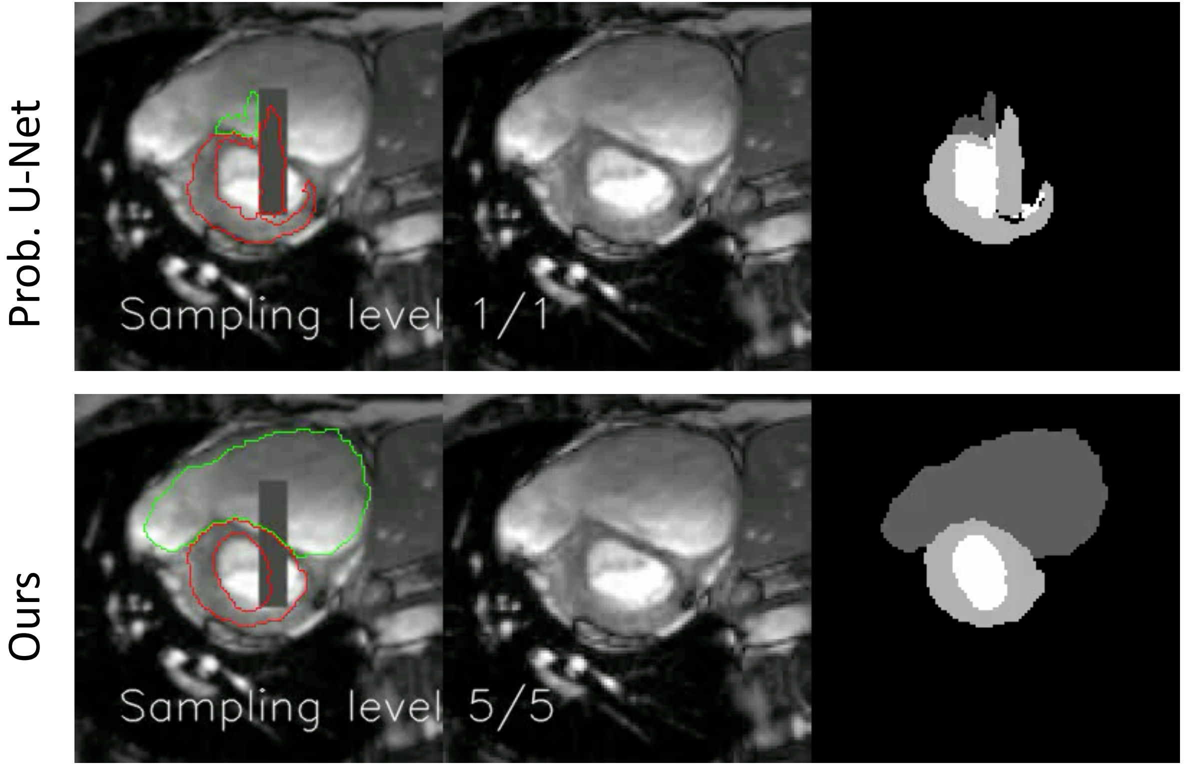
Annotator variance



Predicted variance

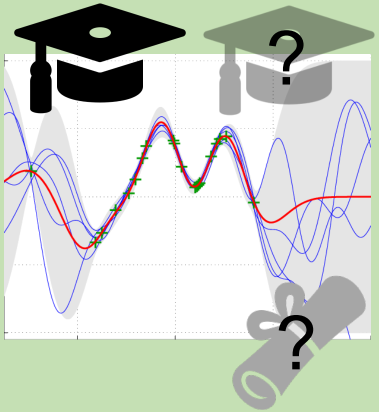
Baumgartner et al., *PHiSeg: Capturing uncertainty in medical image segmentation*, Proc MICCAI (2019)

Intriguing Robustness properties



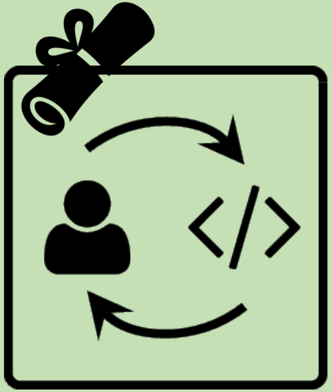
Future directions

Methodological advances to enable human-AI collaboration



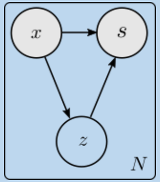
Uncertainty Estimation
Robustness

Interpretable
ML



Human-in-the-loop
Systems

Generative Modelling



Probabilistic
Inference

Summary

- We are seeing the beginning of an AI revolution in medical imaging
- This revolution will open many exciting avenues for improving patient outcomes
- However, the initial, superficial success of deep learning is misleading
- There remain challenging methodological problems to be solved, especially involving the human-AI interface
- Real research progress will happen in collaborations between clinicians and AI researchers

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Thank you for your attention!

? PhD in robust ML for Medical Image Analysis (joint with M. Hein)

? PhD or Post-doc in interpretable ML for Medical Image Analysis

? Master student uncertainty quantification in neuroimaging (joint with AIRAmed artificial intelligence in radiology)

christian.baumgartner@uni-tuebingen.de