Confidence Contours: Uncertainty-Aware Annotation for Medical Semantic Segmentation

Andre Ye | UW URS ‘23

Mentor: Quanze (Jim) Chen; PI: Amy Zhang

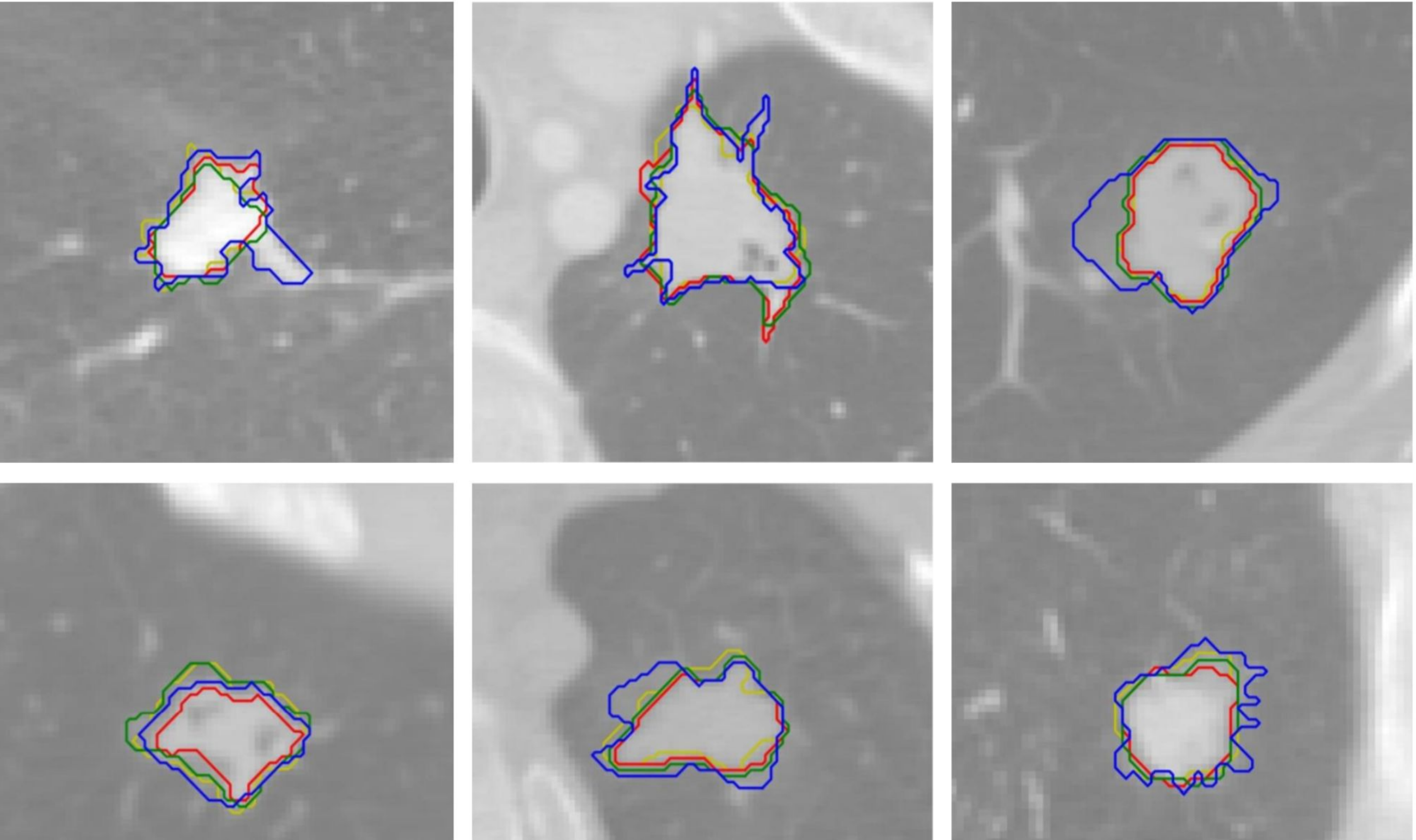
PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Social Futures Lab

What is ‘true’ about the ‘ground truth’?

Semantic segmentation models play an important role in medical imaging applications.

e.g. – segmenting lung nodules. Irregularly shaped or oversized nodules are a strong indicator for lung cancer.



nty.

'.

How to account for structural uncertainty in segmentation?

Existing uncertainty work is model-centric, producing ‘uncertainty of uncertainty’ (not designed for humans)

- Proposes modifications to the model
- Still training on standard maps
- Dominant paradigm in field

Uncertainty Estimates and Multi-Hypotheses Networks for Optical Flow

Eddy Ilg*, Özgün Çiçek*, Silvio Galesso*, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox

Deep Deterministic Uncertainty for Semantic Segmentation

Jishnu Mukhoti[1 2] Joost van Amersfoort[1] Philip H.S. Torr[2] Yarin Gal[1]

A Probabilistic U-Net for Segmentation of Ambiguous Images

Simon A. A. Kohl[1*2], Bernardino Romera-Paredes[1], Clemens Meyer[1], Jeffrey De Fauw[1], Joseph R. Ledsam[1], Klaus H. Maier-Hein[2], S. M. Ali Eslami[1], Danilo Jimenez Rezende[1], and Olaf Ronneberger[1]

Towards safe deep learning: accurately quantifying biomarker uncertainty in neural network predictions

Zach Eaton-Rosen[1], Felix Bragman[1], Sotirios Bisdas[2,3], Sebastien Ourselin[4], and M. Jorge Cardoso[1,4]

Stochastic Segmentation Networks: ...elling Spatially Correlated Aleatoric Uncertainty

Miguel Monteiro, Imperial College London, mm6818@ic.ac.uk
Loïc Le Folgoc, Imperial College London, llefolgo@ic.ac.uk
Daniel Coelho de Castro, Imperial College London, dc315@ic.ac.uk
Nick Pawlowski, Imperial College London, np716@ic.ac.uk
Bernardo Marques, Imperial College London, bgmarque@ic.ac.uk
Konstantinos Kamnitsas, Imperial College London, kk2412@ic.ac.uk
Mark van der Wilk, Imperial College London, m.vdwilk@ic.ac.uk
Ben Glocker, Imperial College London, b.glocker@ic.ac.uk

Supervised Uncertainty Quantification for Segmentation with Multiple Annotations

Shi Hu[1], Daniel Worrall[1], Stefan Knegt[1], Bas Veeling[1], Henkjan Huisman[2], and Max Welling[1]

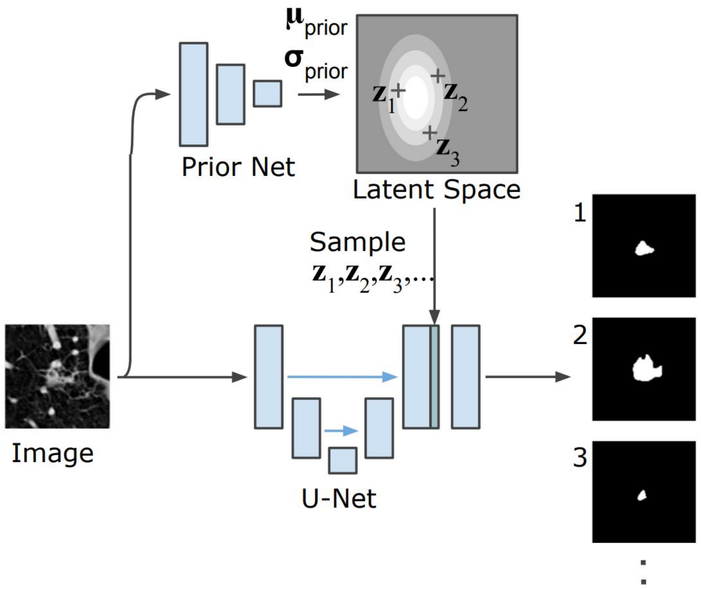
[1] University of Amsterdam
[2] Radboud University Medical Center

Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles

Balaji Lakshminarayanan Alexander Pritzel Charles Blundell
DeepMind
{balajiln,apritzel,cblundell}@google.com

Existing uncertainty work is model-centric, producing ‘uncertainty of uncertainty’ (not designed for humans)

Candidate Generation: infinite generation of possible segmentations

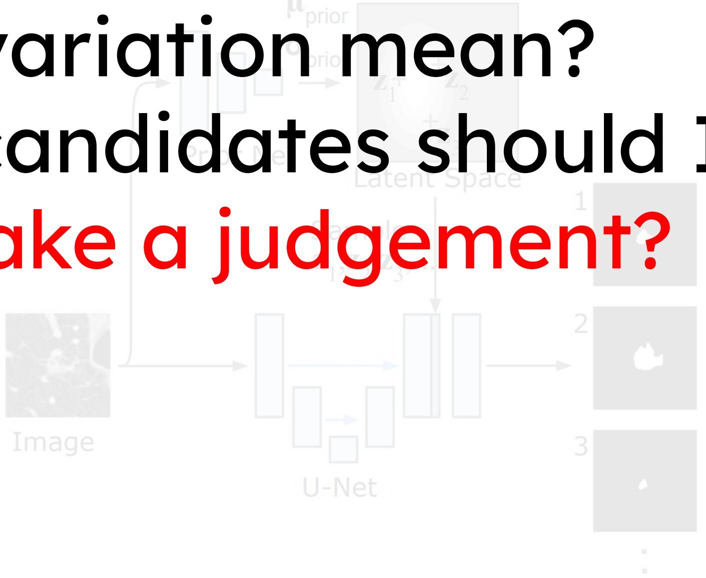


Kohl, “A Probabilistic U-Net for Segmentation of Ambiguous Images”

Existing uncertainty work is model-centric, producing 'uncertainty of uncertainty' (not designed for humans)

Candidate Generation: infinite generation of possible segmentations

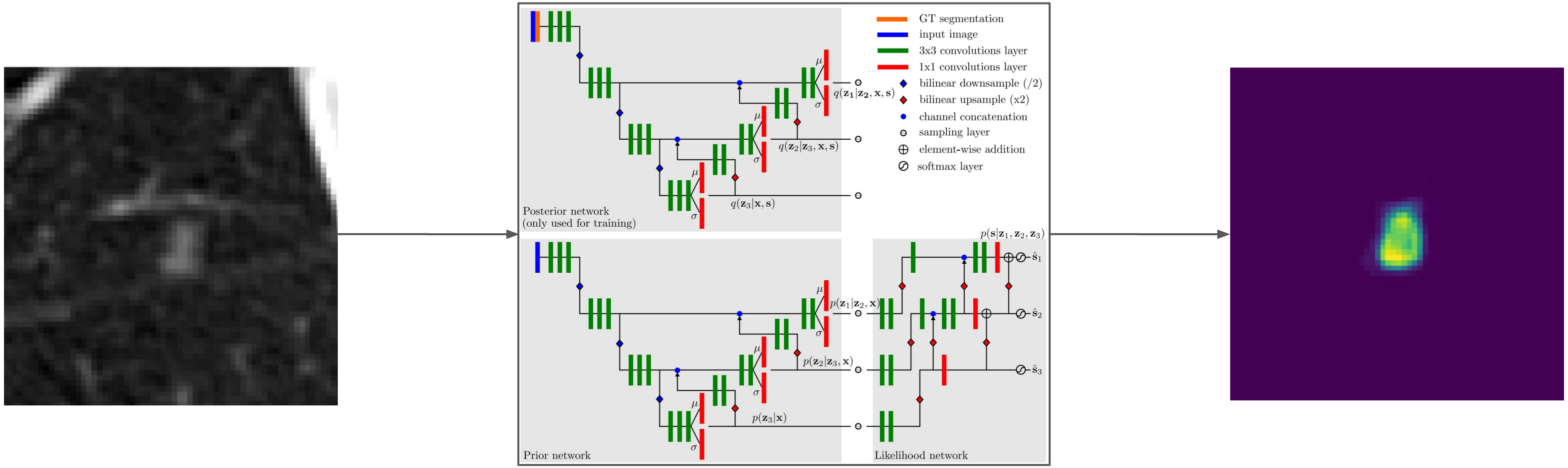
- Contingent on sampling strategy?
- What does variation mean?
- How many candidates should I consider?
- How do I make a judgement?



Kohl, "A Probabilistic U-Net for Segmentation of Ambiguous Images"

Existing uncertainty work is model-centric, producing ‘uncertainty of uncertainty’ (not designed for humans)

<u>Continuous Maps</u>: force non-discrete output

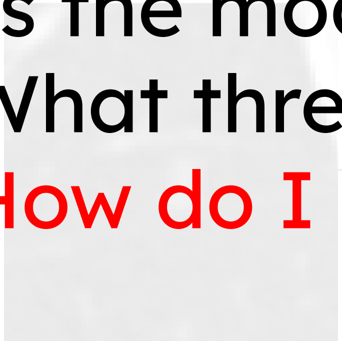


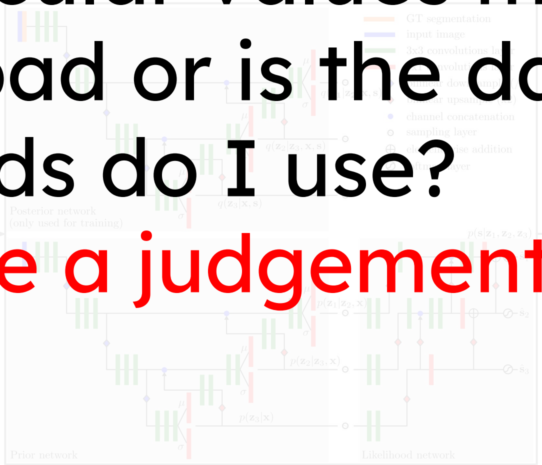
Kohl, “A Probabilistic U-Net for Segmentation of Ambiguous Images”

Existing uncertainty work is model-centric, producing ‘uncertainty of uncertainty’ (not designed for humans)

Continuous Maps: force non-discrete output

- What do particular values mean?
- Is the model bad or is the data hard?
- What thresholds do I use?
- How do I make a judgement?

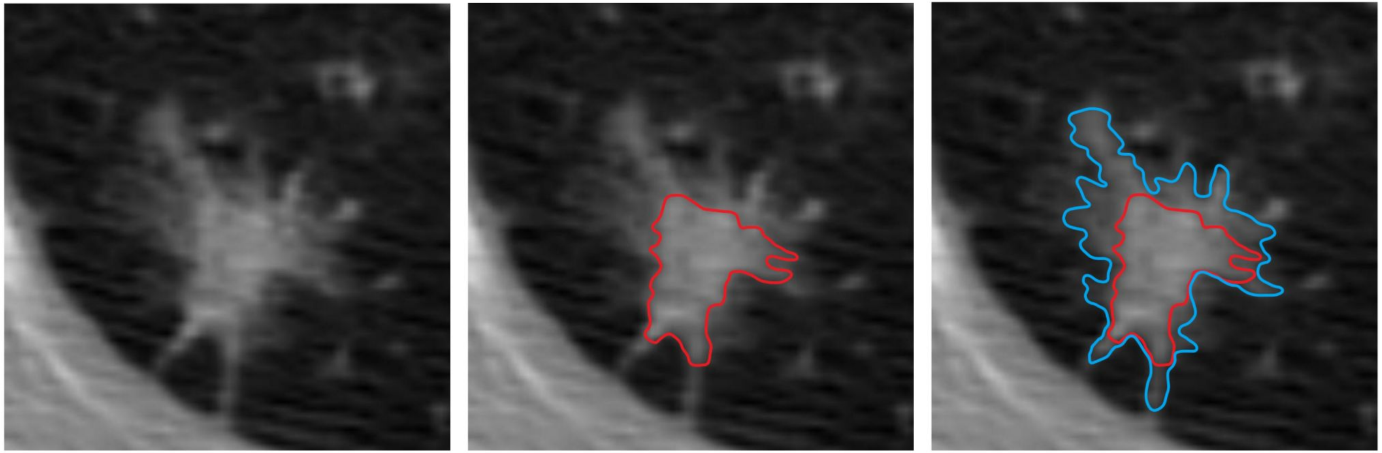




Kohl, “A Probabilistic U-Net for Segmentation of Ambiguous Images”

Model-centric approaches disconnect uncertainty from human judgement.

We need to represent uncertainty explicitly with a data-centric approach. Introducing **Confidence Contours**



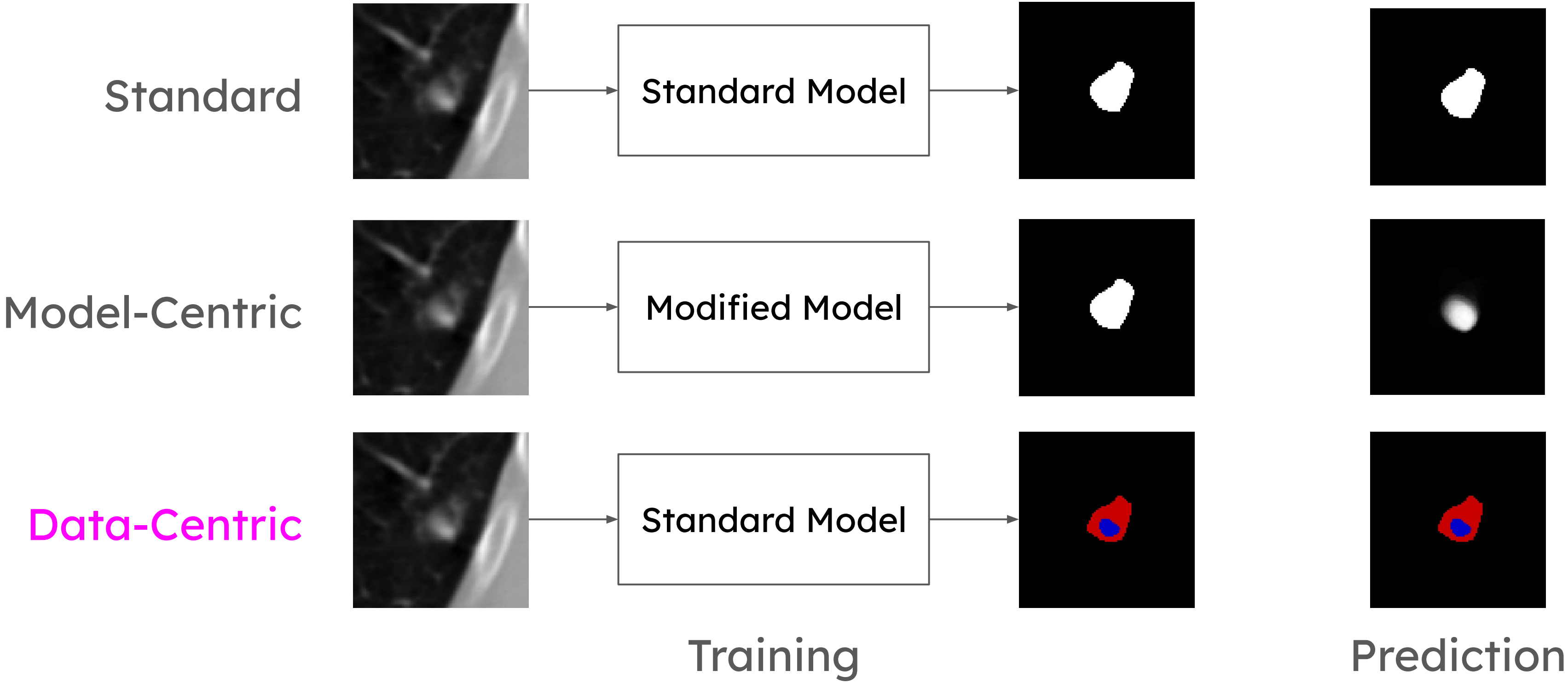
Step 1

Draw min

Step 2

Draw max

Training models on Confidence Contours requires no architectural modifications, unlike other methods



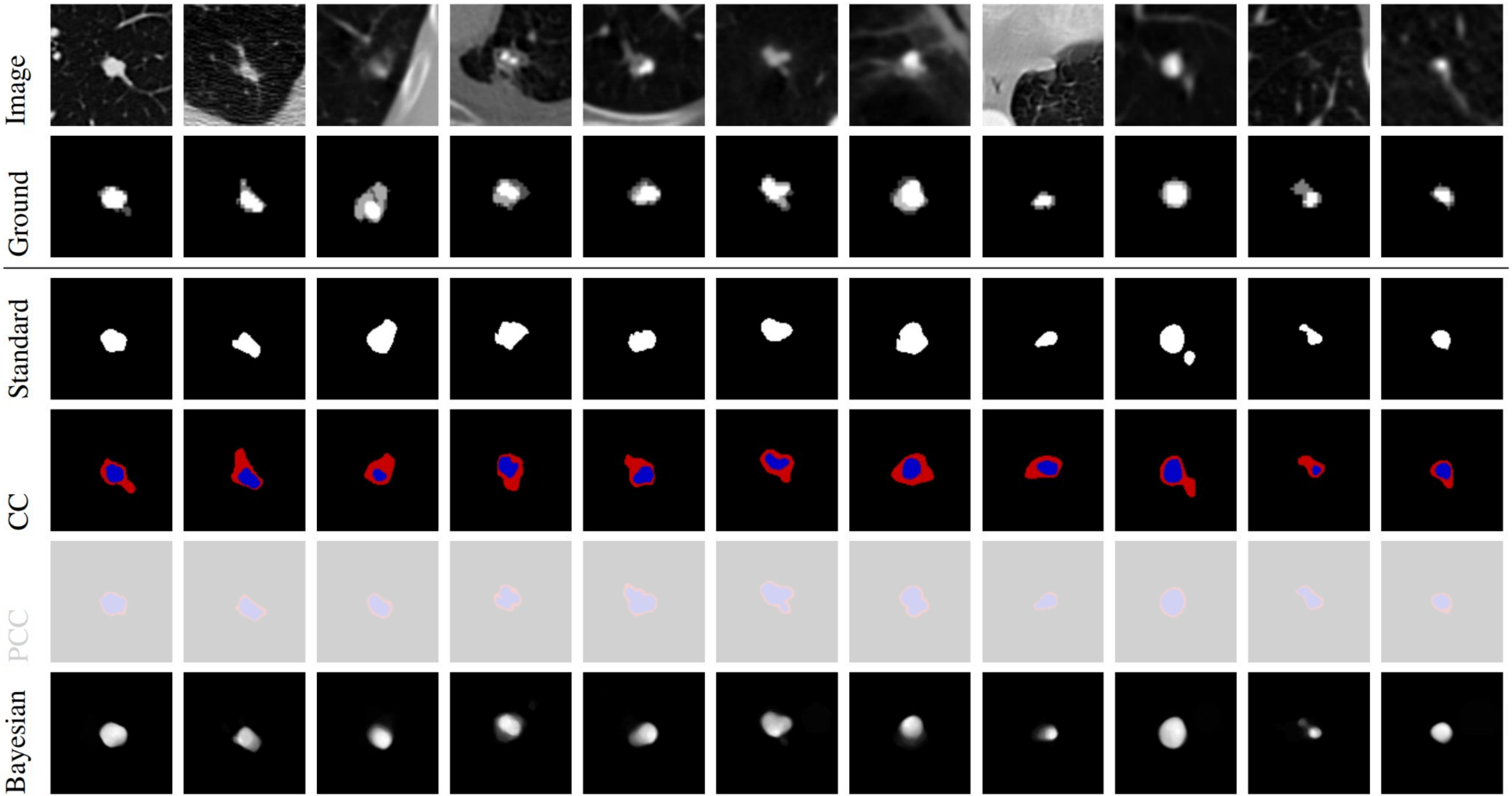
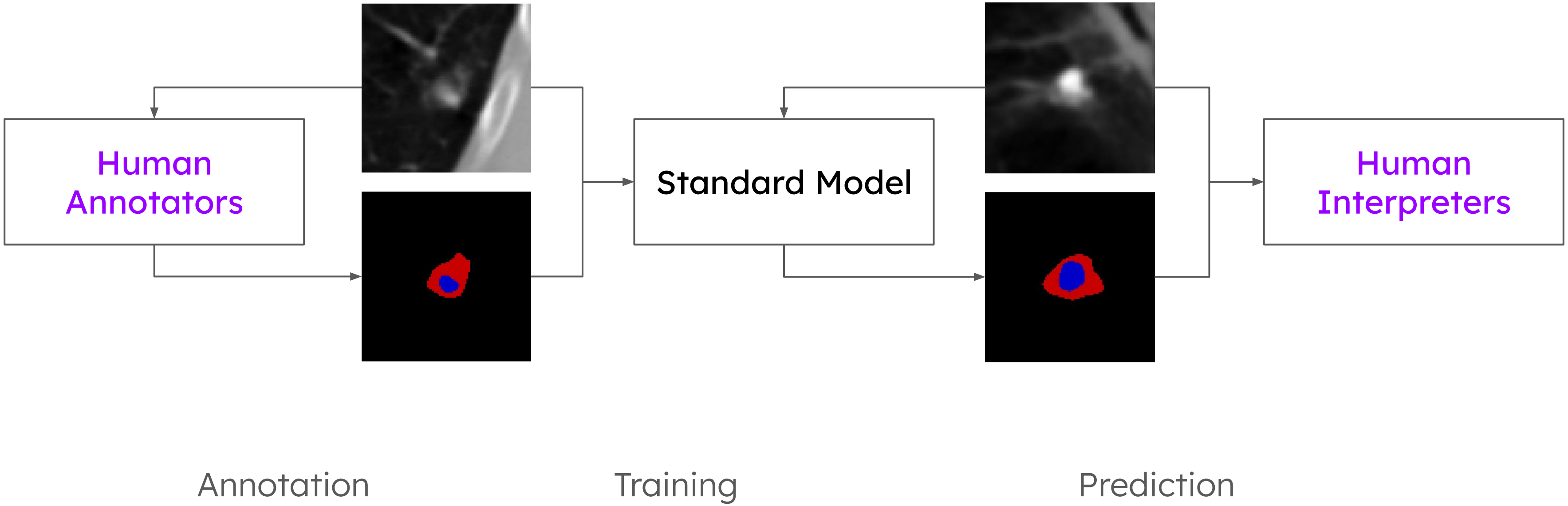
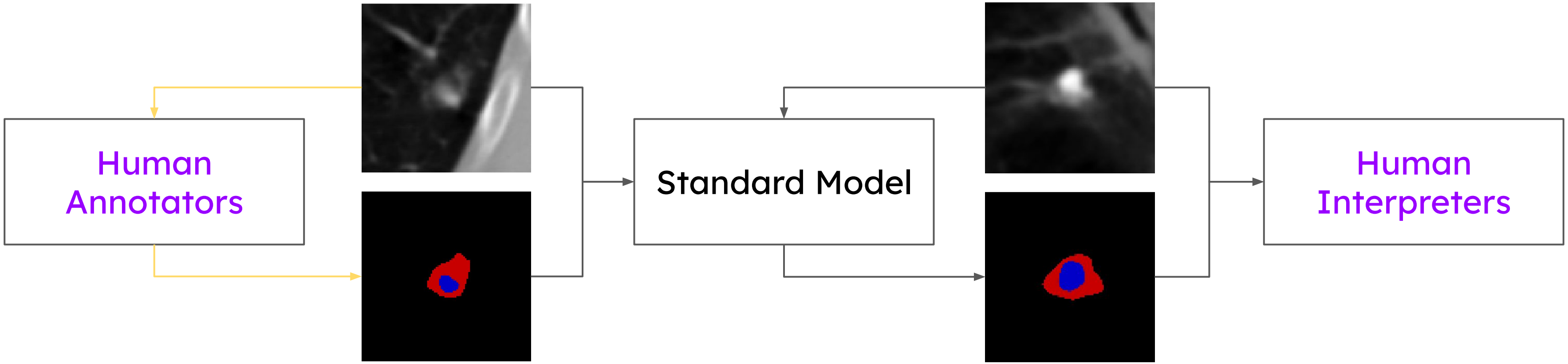


Image
Ground
Standard
CC
PCC
Bayesian

Confidence Contours recenters the humans at both sides of the uncertainty modeling pipeline



Human annotators directly mark uncertainty in the image with minimally more effort

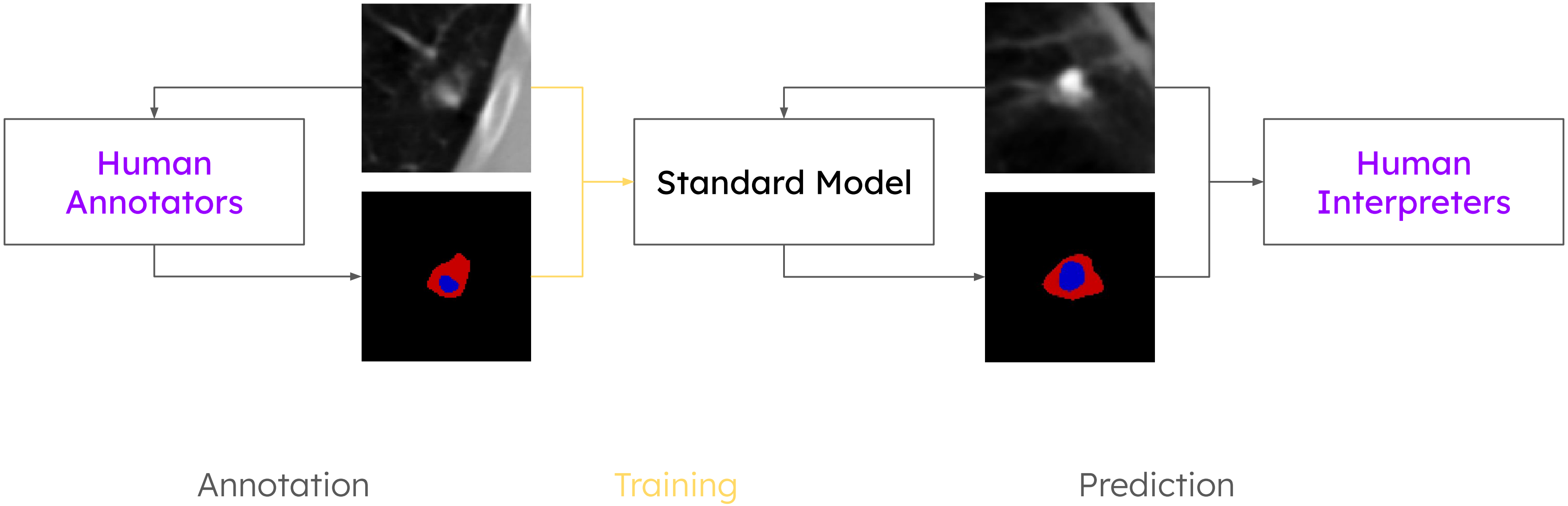


Annotation

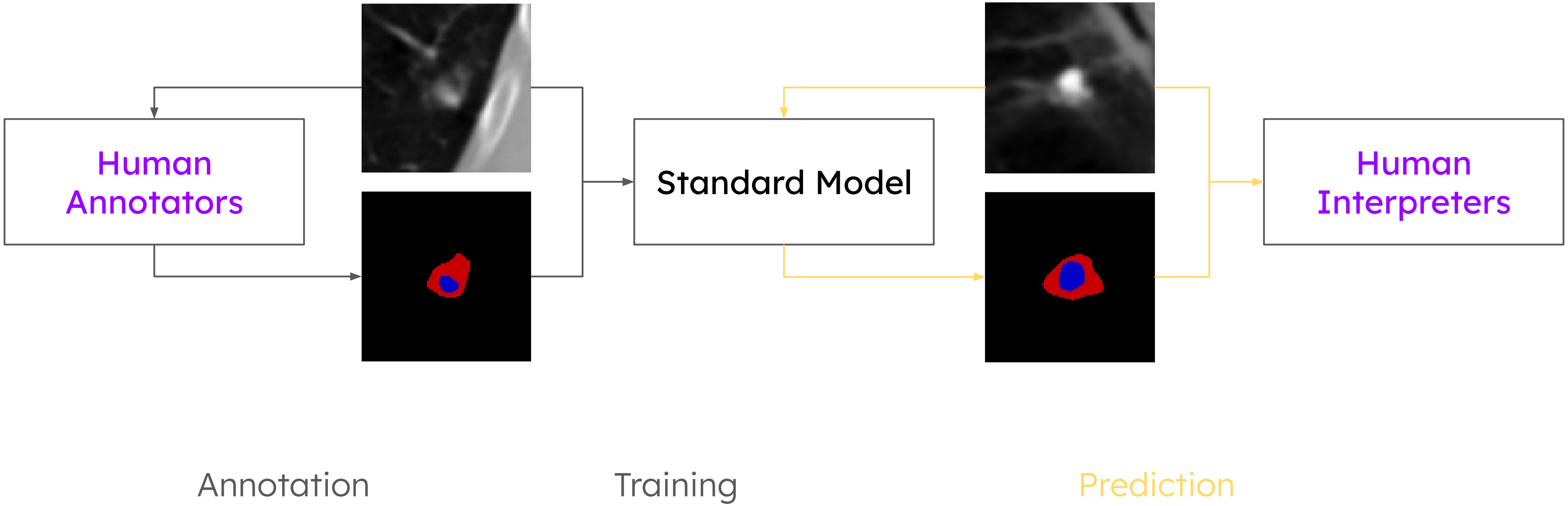
Training

Prediction

Models are simply trained by predicting two rather than one segmentation maps; no bells & whistles needed



All uncertainty information directly corresponds to human annotations. No black-box uncertainty inferences!



What we designate the ‘ground truth’ shapes downstream tasks and can be strategically designed

Massive thanks🙏

Jim Chen and Amy Zhang

Confidence Contours: Uncertainty-Aware Annotation for Medical Semantic Segmentation

Andre Ye[1] , **Quanze Chen**[1] and **Amy Zhang**[1]
[1]University of Washington
andreye@uw.edu, cqz@cs.washington.edu, axz@cs.uw.edu

Abstract

Medical image segmentation modeling is a high-stakes task where understanding of uncertainty is crucial for addressing visual ambiguity. Prior work has developed segmentation models utilizing probabilistic or generative mechanisms to infer uncertainty from labels where annotators draw a singular boundary. However, as these annotations cannot represent an individual annotator's uncertainty, models trained on them produce uncertainty maps that are difficult to interpret. We propose a novel

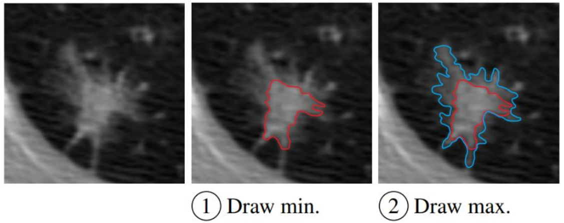


Figure 1: The two steps of the process for producing Confidence Contours annotations, demonstrated on a sample from LIDC.

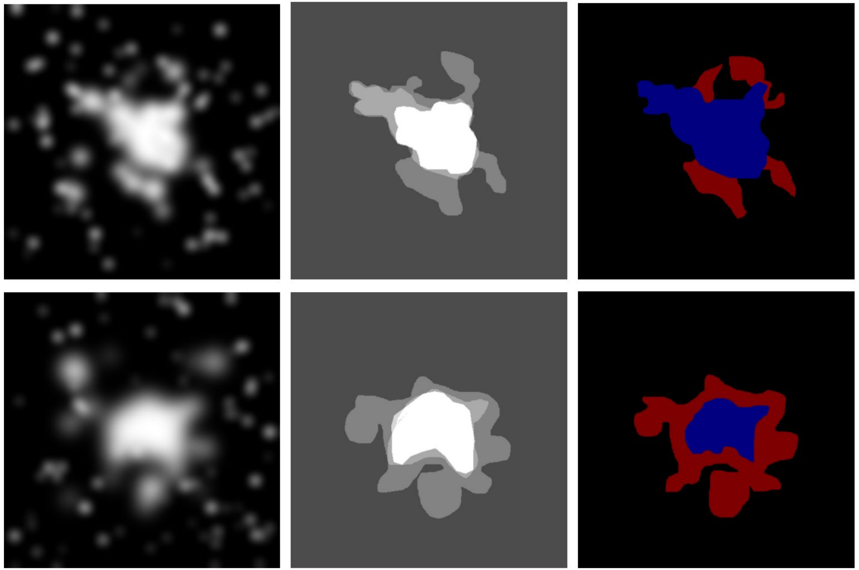
Thank you!

User Study

- Recruited 45 students to annotate 600 images across 2 datasets
    - LIDC: Lung Image Dataset Consortium (Pulmonary Nodule Segmentation)
    - FoggyBlob: synthetic dataset simulating structural uncertainty
- Each image annotated with 3 standard and 3 Confidence Contours
- Two groups to counteract learning bias

One CC can represent multiple standard annotations

- Significant reductions in underflow and overflow
- Significant reductions in disagreement between annotations



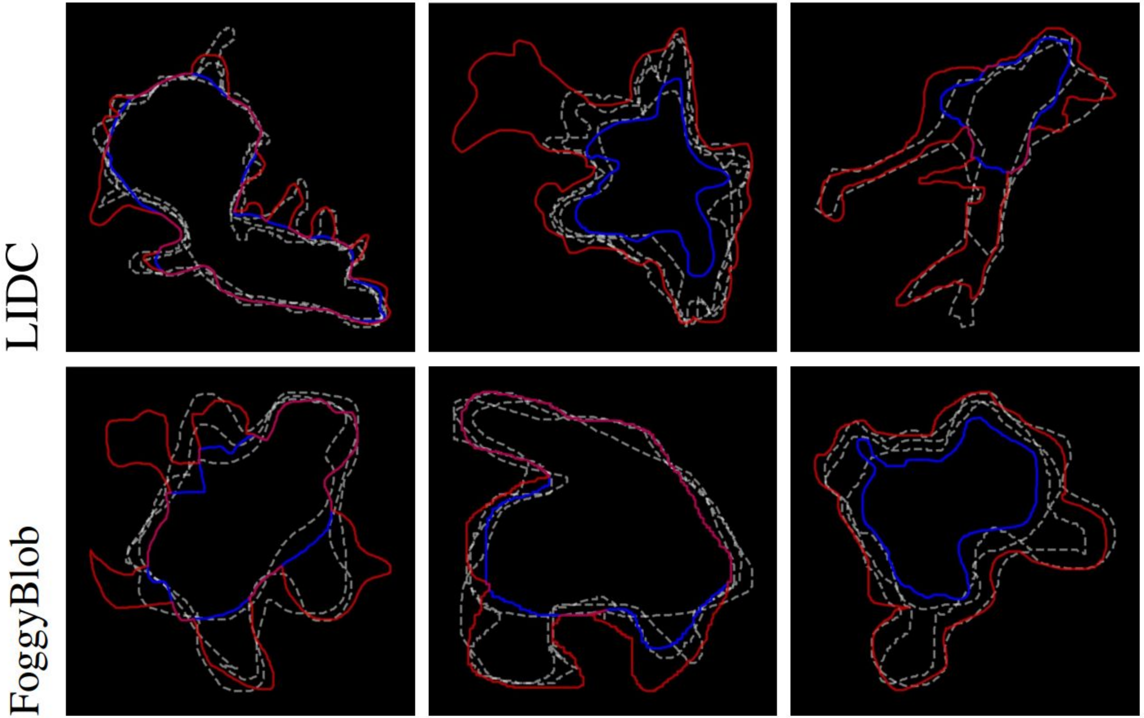
Original image Composited standard Single CC

Annotators find CCs more demanding, but not by much

Dimension	LIDC		FoggyBlob	
	Singular	CC	Singular	CC
Mental Demand	3.7	*4.9	3.3	*4.6
Physical Demand	2.7	3.3	3.9	3.7
Temporal Demand	4.2	*4.9	5.0	5.5
Performance	6.9	6.9	6.8	6.9
Effort	4.8	*5.7	5.0	5.1
Frustration	3.0	*4.2	2.7	*4.0

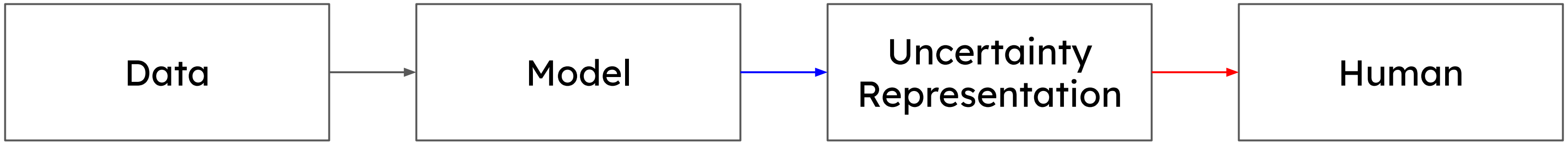
Table 3: Average annotator responses across six dimensions and two datasets on the experience annotating using the singular and the CC methods, evaluated on a 10 point scale (1="very low", 10="very high"). * indicates a statistically significant relationship, measured with a relative t-test by annotator.

CC annotations give positive information to more pixels, ‘expanding the ground truth’



Red: max
Blue: min
White: standard

Designing uncertainty representations with humans in mind



Uncertainty representations should clearly correspond to interactions between the model and the data

The correspondence should be clear to the human.