CONFORMAL INFERENCE FOR BIOMEDICAL IMAGE SEGMENTATION

SAMUEL DAVENPORT – UNIVERSITY OF CALIFORNIA SAN DIEGO

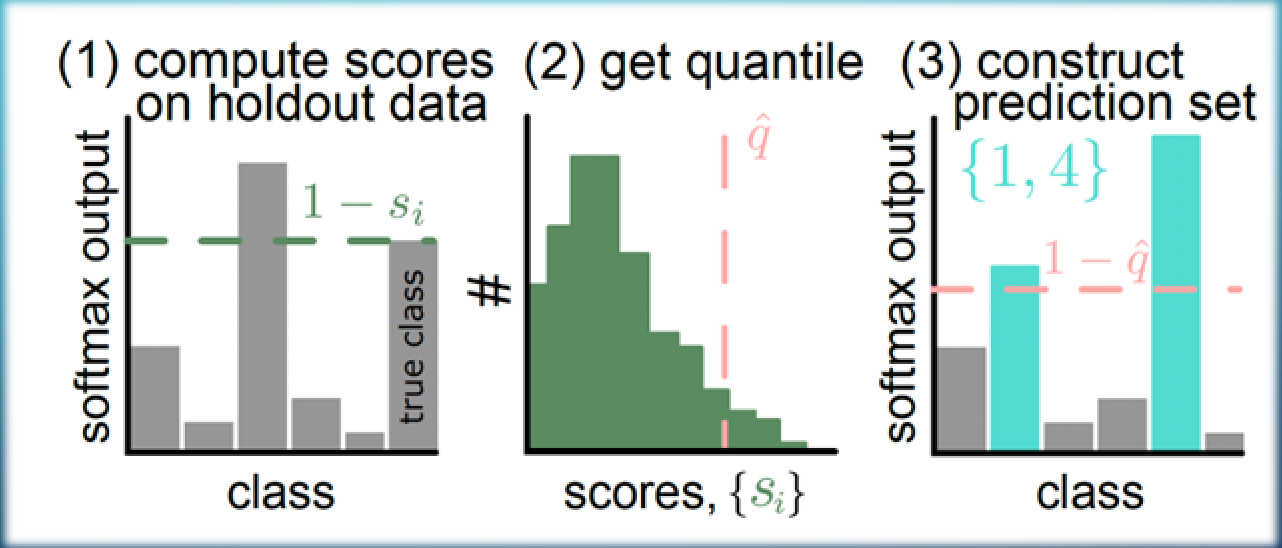
UNCERTAINTY QUANTIFICATION IN MACHINE LEARNING

- Machine learning models such as a neural networks as extremely useful
- Used to perform classification, segmentation, object detection, and regression.
- However they are often black-box models and do not provide uncertainty guarantees. They can be wrong and are often over-confident in their predictions.
- Mistakes, for instance in medical fields, can be very bad and have rather negative consequences.

INTRODUCTION TO CONFORMAL INFERENCE

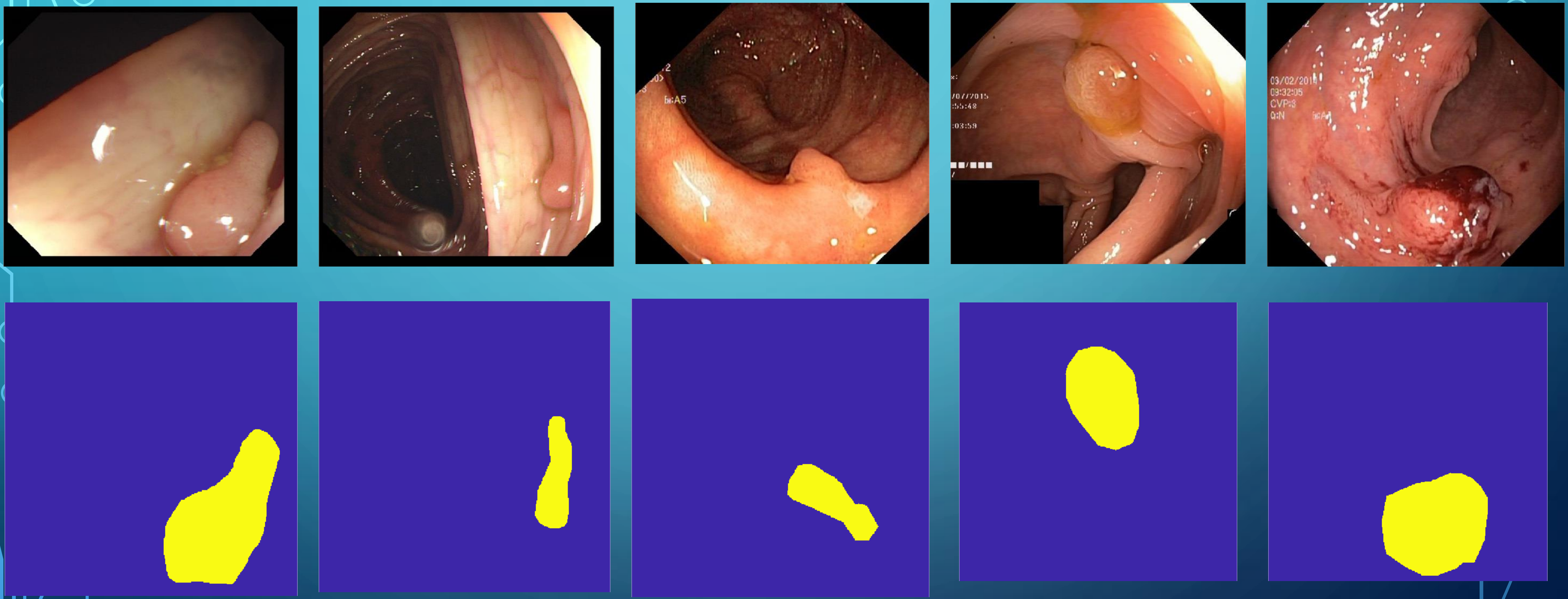
$$1 - \alpha \leq \mathbb{P}(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}}))$$

INTRODUCTION TO CONFORMAL INFERENCE

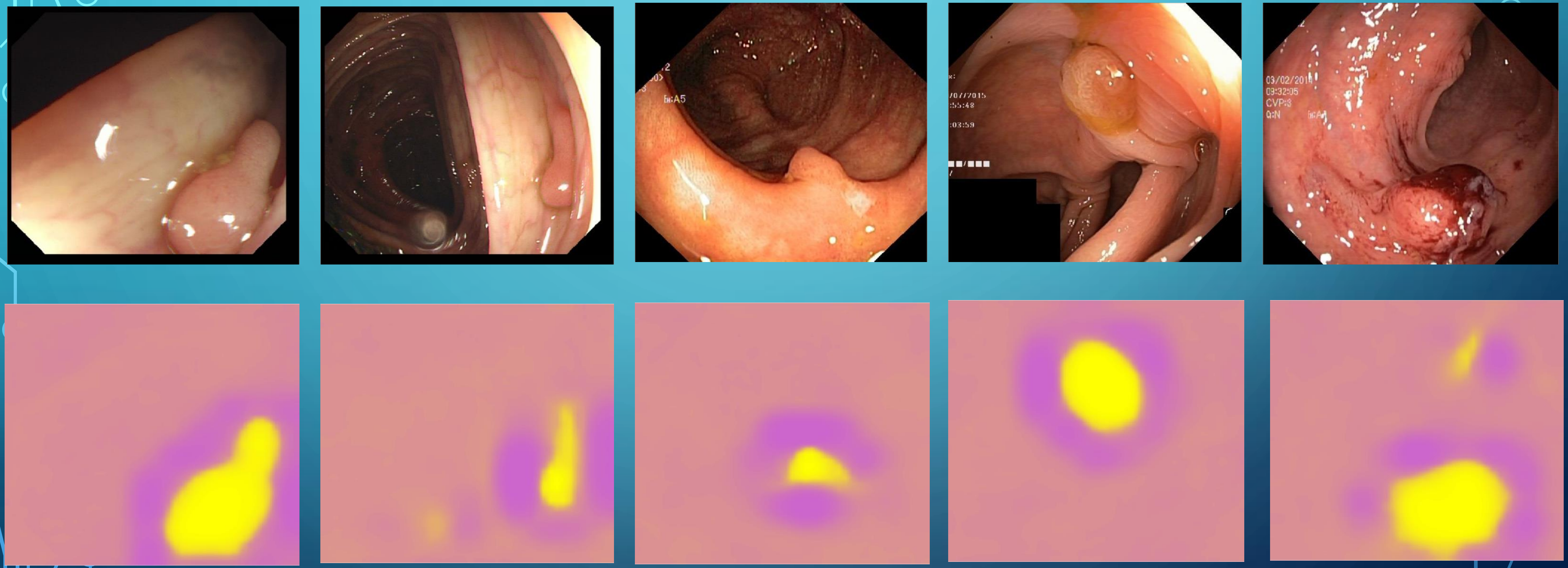


INTRODUCTION TO CONFORMAL INFERENCE

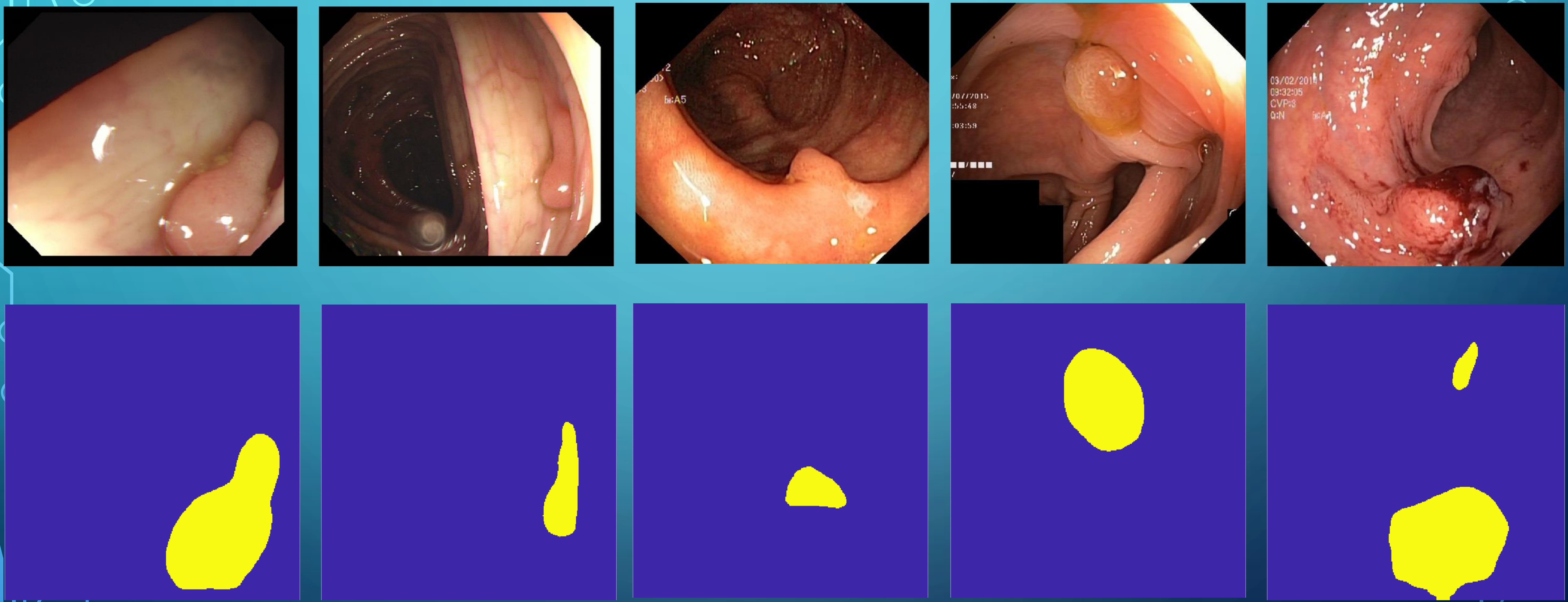
POLPYS SEGMENTATION



POLPYS SEGMENTATION



POLPYS SEGMENTATION



NOTATION

1. Let $\mathcal{V} \subset \mathbb{R}^m$ be a finite set which represents the pixels/voxels at which we observe imaging data.
2. Let $\mathcal{X} = \{g : \mathcal{V} \to \mathbb{R}\}$.
3. Let $\mathcal{Y} = \{g : \mathcal{V} \to \{0, 1\}\}$
4. Suppose that we observe a calibration dataset $(X_i, Y_i)_{i=1}^n$ of random images, where $X_i : \mathcal{V} \to \mathbb{R}$ represents the ith observed calibration image and $Y_i : \mathcal{V} \to \{0, 1\}$ outputs labels at each $v \in \mathcal{V}$ giving 1s at the true location of the objects in the image X_i that we wish to identify and 0s elsewhere.
5. Given a function $f : \mathcal{X} \to \mathcal{X}$, we shall write $f(X, v)$ to denote $f(X)(v)$ for all $v \in \mathcal{V}$.

FURTHER NOTATION

1. Let $s : \mathcal{X} \to \mathcal{X}$ be a score function such that given an image pair $(X, Y) \in \mathcal{X} \times \mathcal{Y}$, $s(X)$ is a score image in which $s(X, v)$ is intended to be higher at the $v \in \mathcal{V}$ for which $Y(v) = 1$.
2. The score function can for instance be the logit scores obtained from a deep neural network image segmentation method to the image X.
3. Given $X \in \mathcal{X}$, let $\hat{M}(X) \in \mathcal{Y}$ be the predicted mask.
4. Let $\mathcal{P}(\mathcal{V})$ be the set of subsets of $\mathcal{V}$.

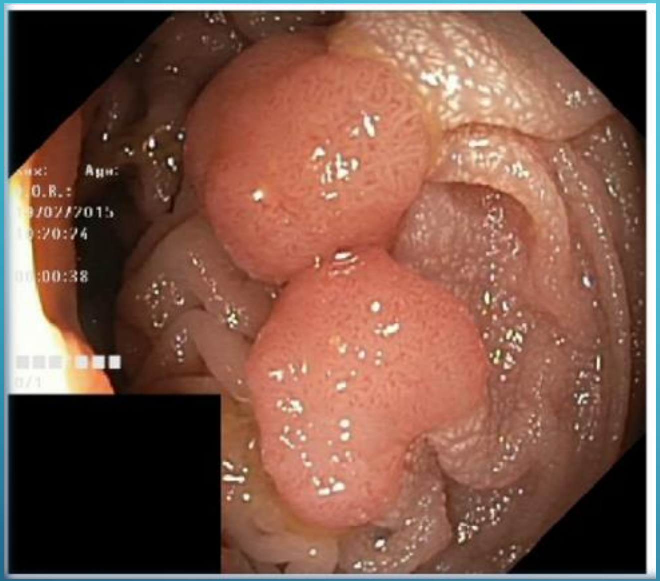
CONFIDENCE SETS

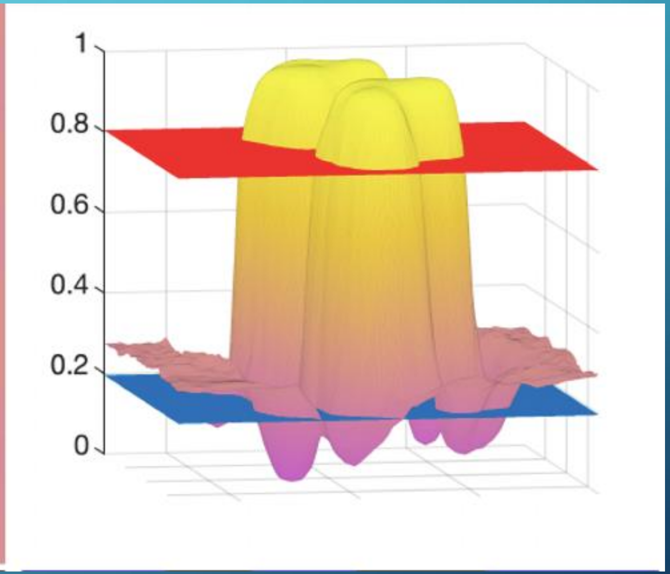
In what follows we will use the calibration dataset to construct a confidence functions $I, O : \mathcal{X} \to \mathcal{P}(\mathcal{V})$ such that for a new image pair $(X, Y) \sim \mathcal{D}$, given error rates $\alpha_1, \alpha_2 \in (0, 1)$ we have

$$\mathbb{P}\left(I(X) \subseteq \{v \in \mathcal{V} : Y(v) = 1\}\right) \geq 1 - \alpha_1, \tag{1}$$

$$\text{and } \mathbb{P}\left(\{v \in \mathcal{V} : Y(v) = 1\} \subseteq O(X)\right) \geq 1 - \alpha_2. \tag{2}$$

THRESHOLDING BASED ON THE NEURAL NETWORK SCORES





ASSUMPTIONS FOR VALID INFERENCE

Assumption 1. Given a new random image pair, (X_{n+1}, Y_{n+1}), suppose that $(X_i, Y_i)_{i=1}^{n+1}$ is an exchangeable sequence of random image pairs in the sense that

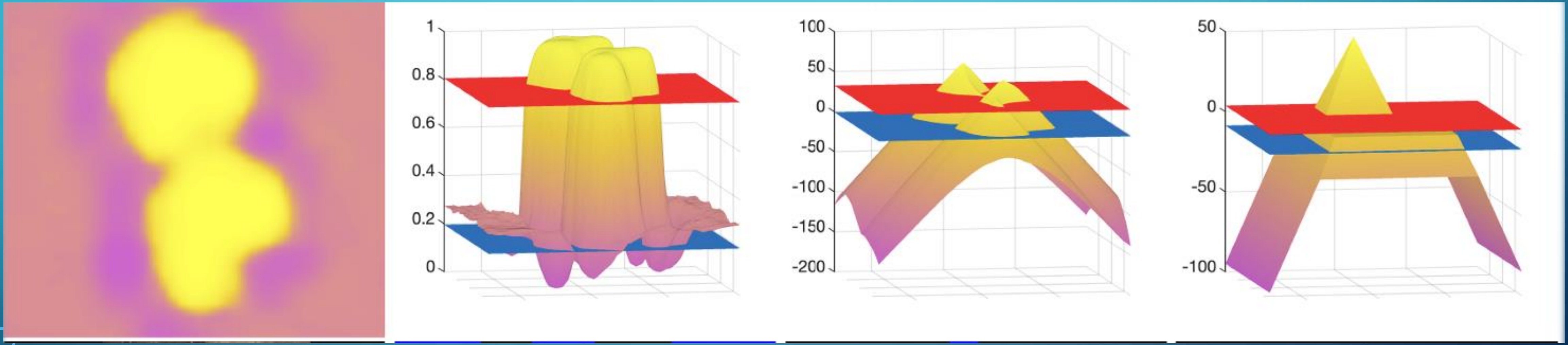
$$\{(X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})\} =_d \{(X_{\sigma(1)}, Y_{\sigma(1)}), \ldots, (X_{\sigma(n+1)}, Y_{\sigma(n+1)})\}$$

for any permutation $\sigma \in S_{n+1}$. Here $=_d$ denotes equality in distribution and S_{n+1} is the group of permutations of the integers $\{1, \ldots, n+1\}$.

Assumption 2. (Independence of scores) $(X_i, Y_i)_{i=1}^{n+1}$ is independent of the functions s, f_O, f_I.

SCORE TRANSFORMATIONS

$$f_I, f_O : \mathcal{X} \to \mathcal{X}$$



MARGINAL INNER SET

$$\tau_i = \max_{v \in \mathcal{V}: Y_i(v)=0} f_I(s(X_i), v)$$

Theorem 2.1. *(Marginal inner set) Under Assumptions 1 and 2, given* $\alpha_1 \in (0,1)$, *let*

$$\lambda_I(\alpha_1) = \inf\left\{\lambda : \frac{1}{n}\sum_{i=1}^{n} 1\left[\tau_i \leq \lambda\right] \geq \frac{\lceil(1-\alpha_1)(n+1)\rceil}{n}\right\},$$

and define $I(X) = \{v \in \mathcal{V} : f_I(s(X), v) > \lambda_I(\alpha_2)\}$. *Then,*

$$\mathbb{P}\left(I(X_{n+1}) \subseteq \{v \in \mathcal{V} : Y_{n+1}(v) = 1\}\right) \geq 1 - \alpha_1.$$

MARGINAL OUTER SET

$$\gamma_i = \max_{v \in \mathcal{V}: Y_i(v)=1} -f_O(s(X_i), v)$$

Theorem 2.2. *(Marginal outer set) Under Assumptions 1 and 2, given $\alpha_2 \in (0, 1)$, let*

$$\lambda_O(\alpha_2) = \inf\left\{\lambda : \frac{1}{n}\sum_{i=1}^{n} 1\left[\gamma_i \leq \lambda\right] \geq \frac{\lceil(1-\alpha_2)(n+1)\rceil}{n}\right\},$$

and define $O(X) = \{v \in \mathcal{V} : f_O(-s(X), v) \leq \lambda_O(\alpha_2)\}$. Then,

$$\mathbb{P}\left(\{v \in \mathcal{V} : Y_{n+1}(v) = 1\} \subseteq O(X_{n+1})\right) \geq 1 - \alpha_2.$$

DISTANCE TRANSFORMED SCORES

Distance transformation:

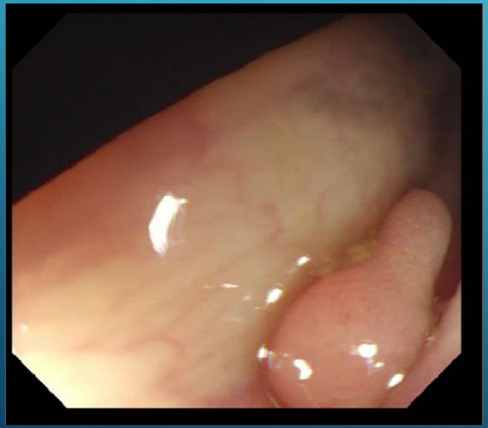
$$d_\rho(\mathcal{A}, v) = \mathbf{sign}(\mathcal{A}, v) \min\{\rho(v, e) : e \in E(\mathcal{A})\},$$

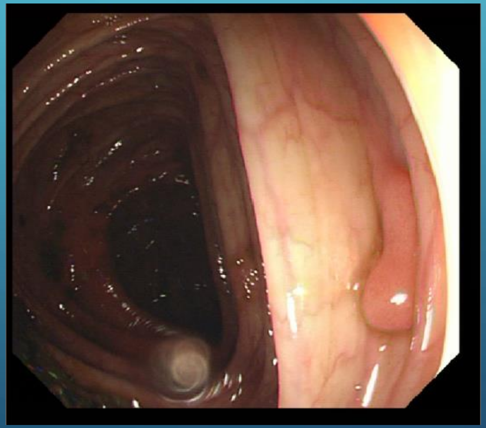
Distance transformed scores:

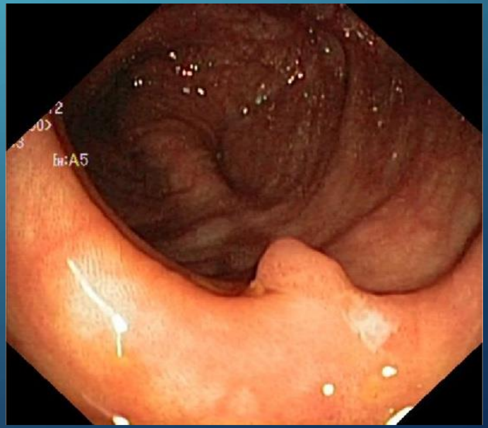
$$f_I(s(X), v) = f_O(s(X), v) = d_\rho(\hat{M}(X), v)$$

APPLICATION TO POLYPS DATA

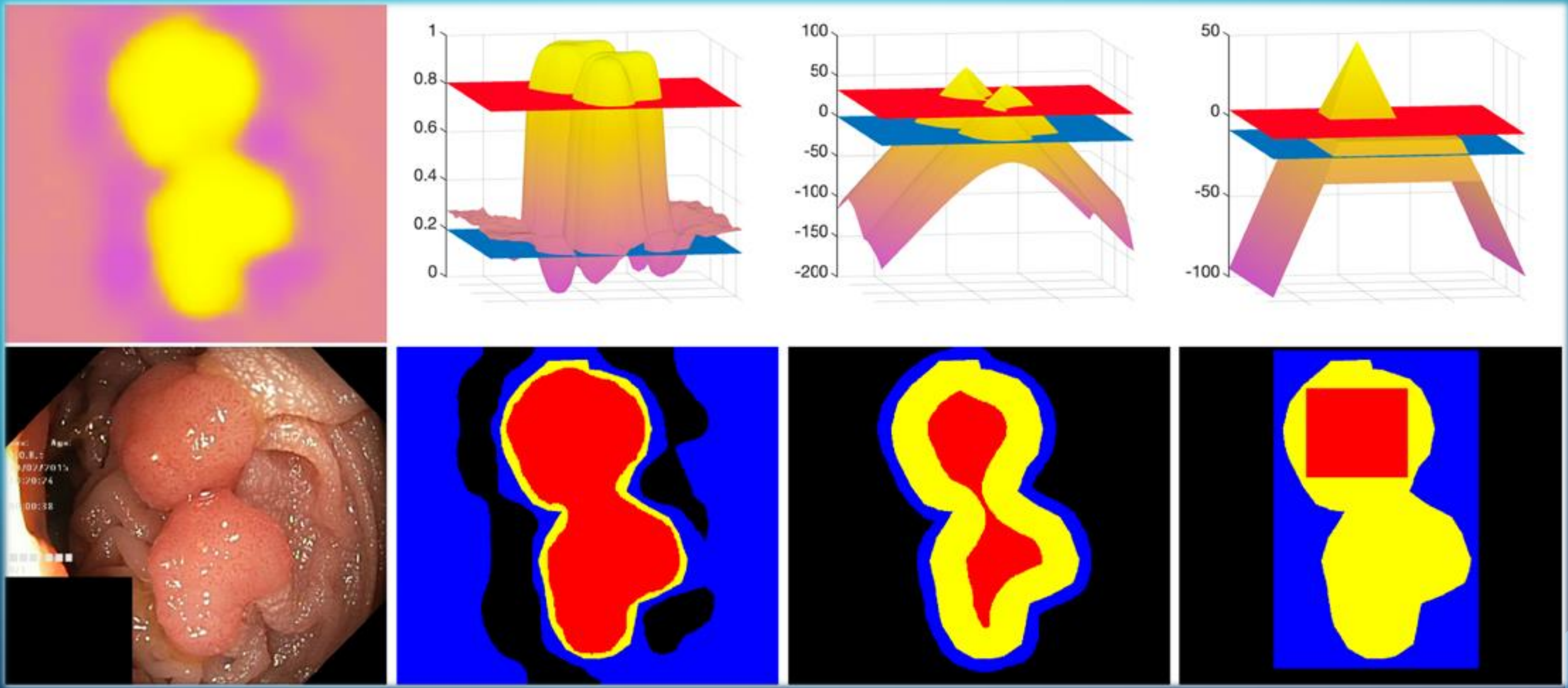
- We have 1798 polyps images (from different subjects)
- We divide these into a learning dataset of 298 and use the rest for inference
- Other existing approaches use the untransformed scores instead of learning the best approach.



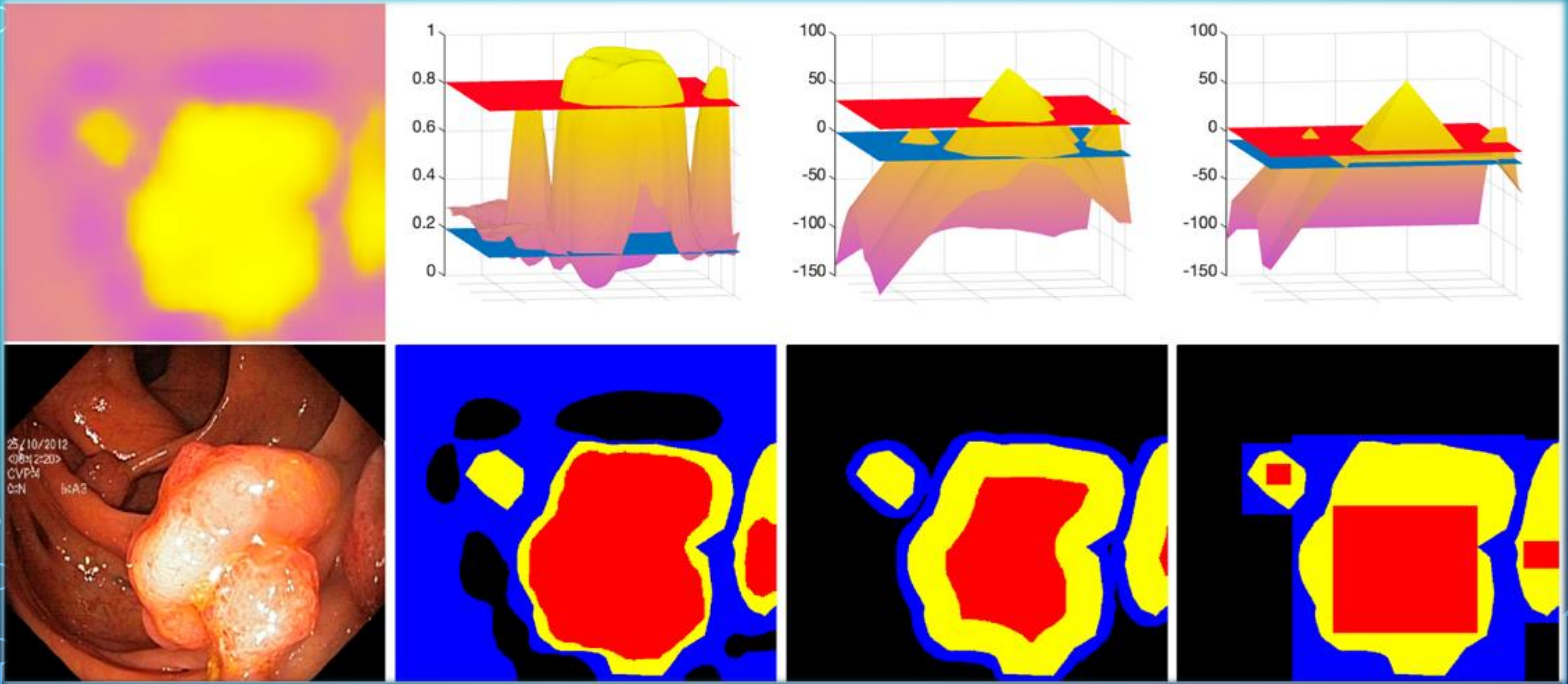




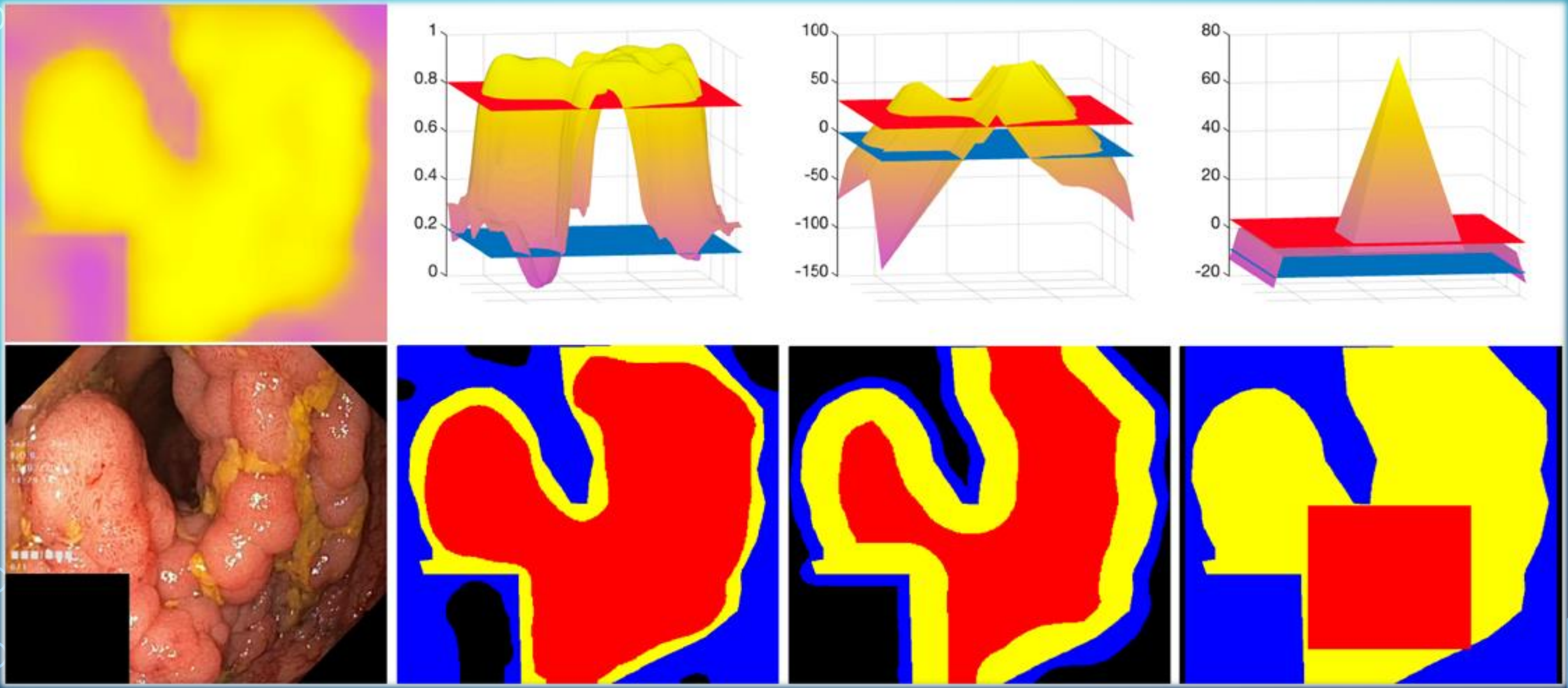
STUDYING THE LEARNING DATA



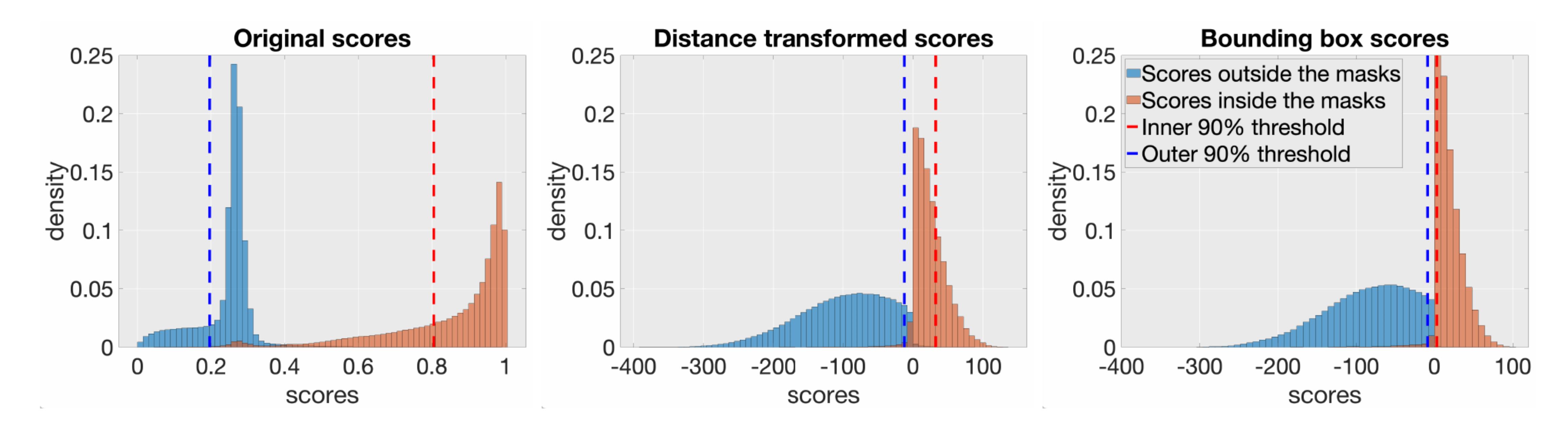
STUDYING THE LEARNING DATA



STUDYING THE LEARNING DATA



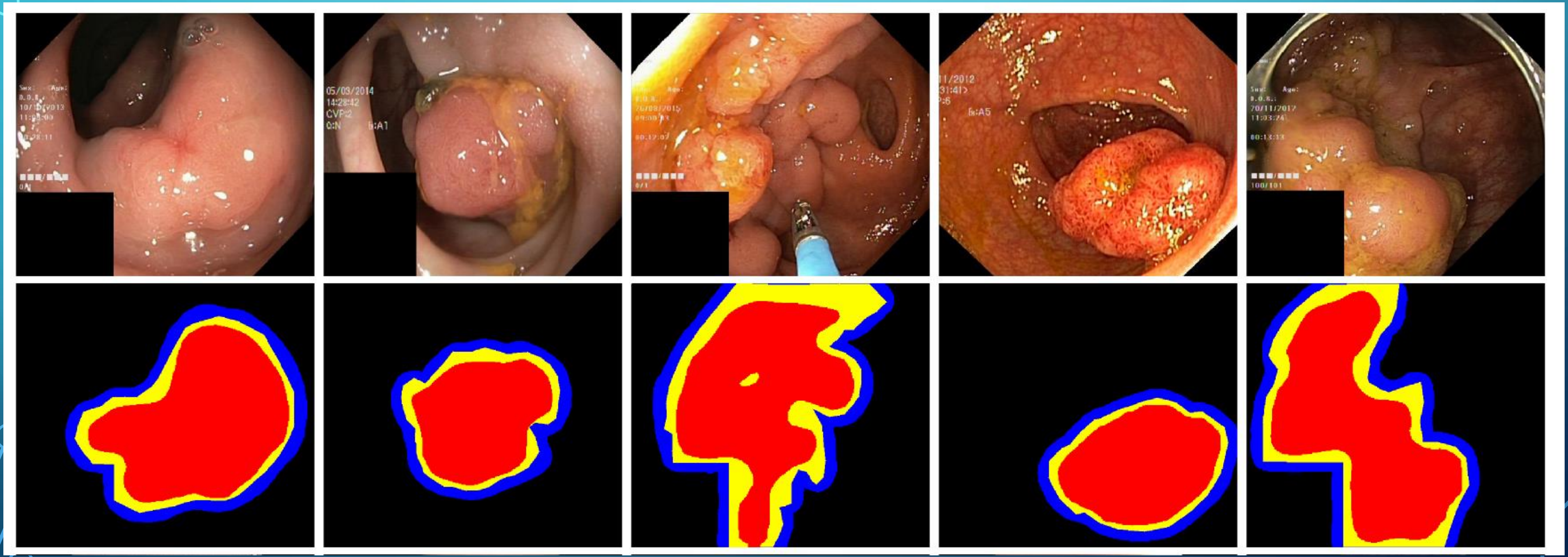
HISTOGRAMS OF THE SCORES ON THE LEARNING DATA



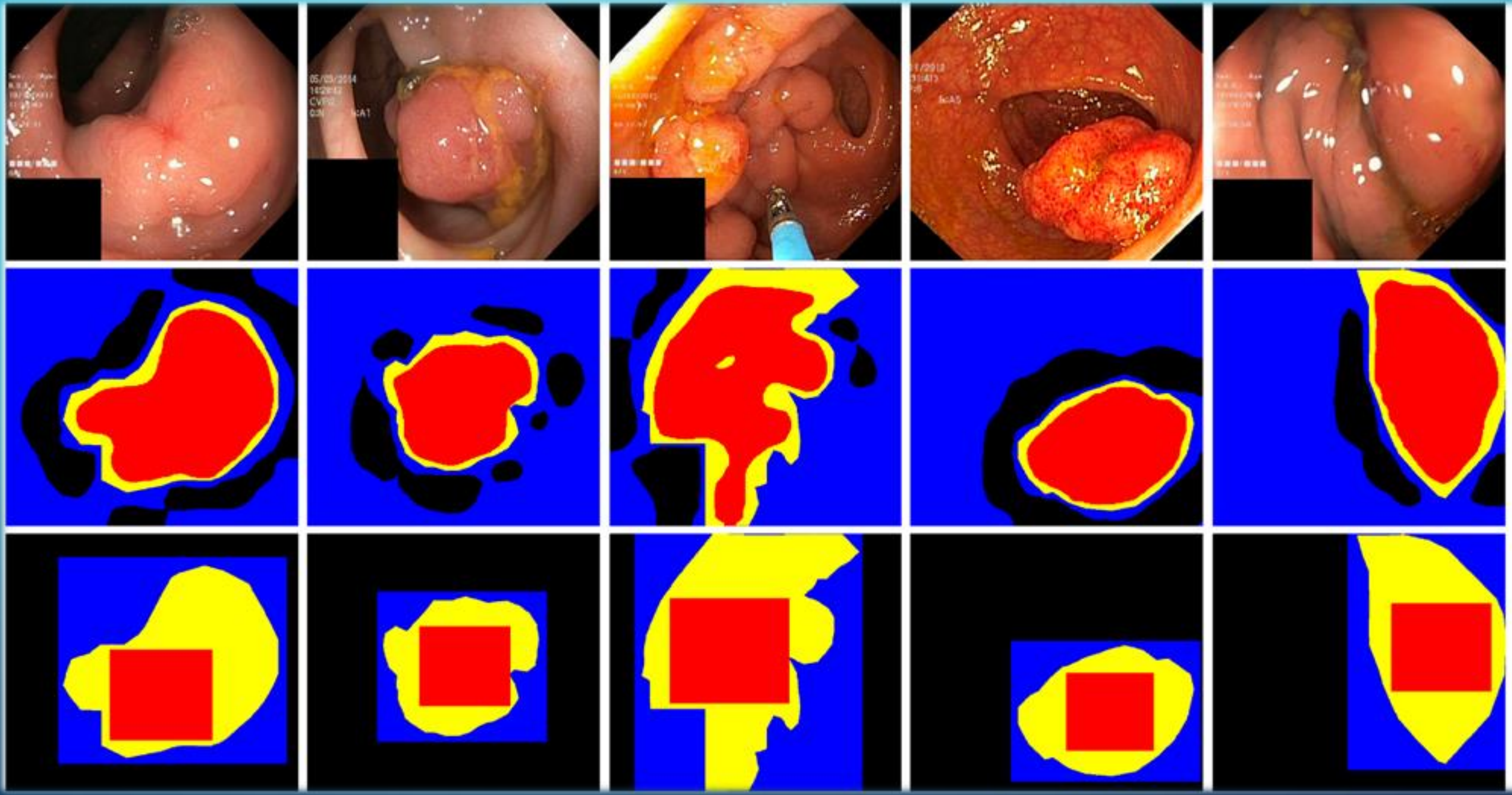
PREPARING FOR THE CALIBRATION/VALIDATION

- From the learning data we can see that mixing the original and distance based score functions appears to the best combination.
- For our results we can then divide the remaining 1500 images into a calibration set of 1000 and a validation set of 500 images. And visualize the results.

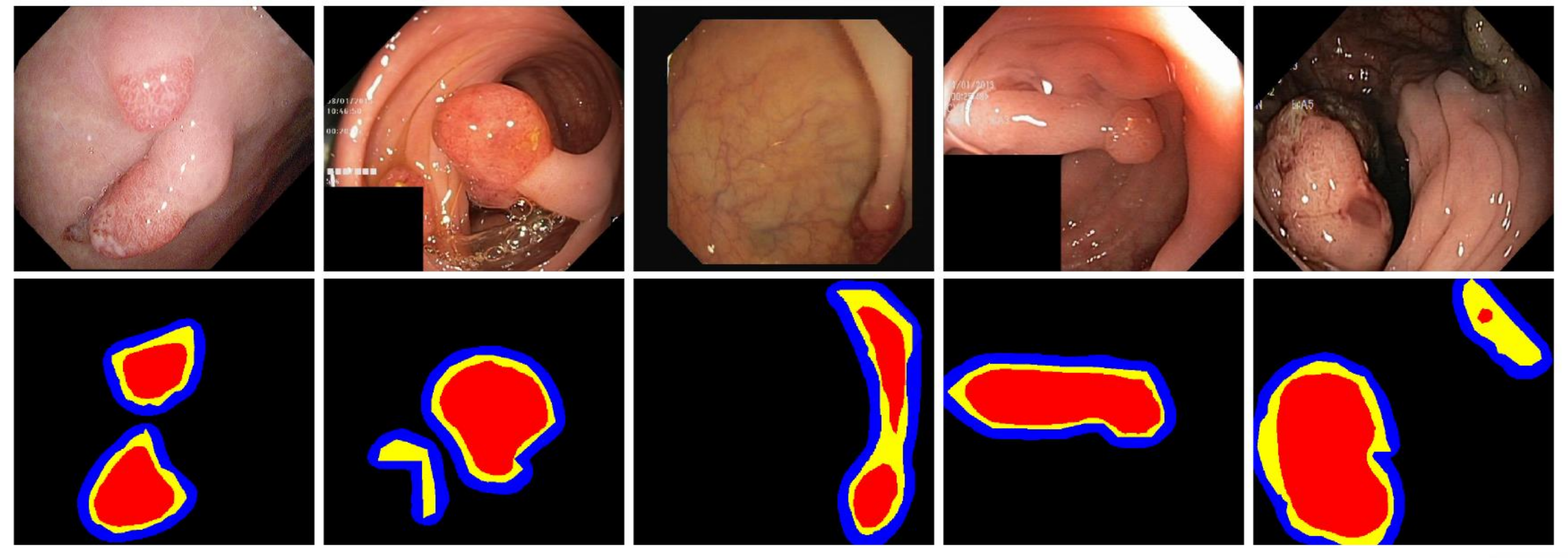
CONFIDENCE SETS APPLICATION



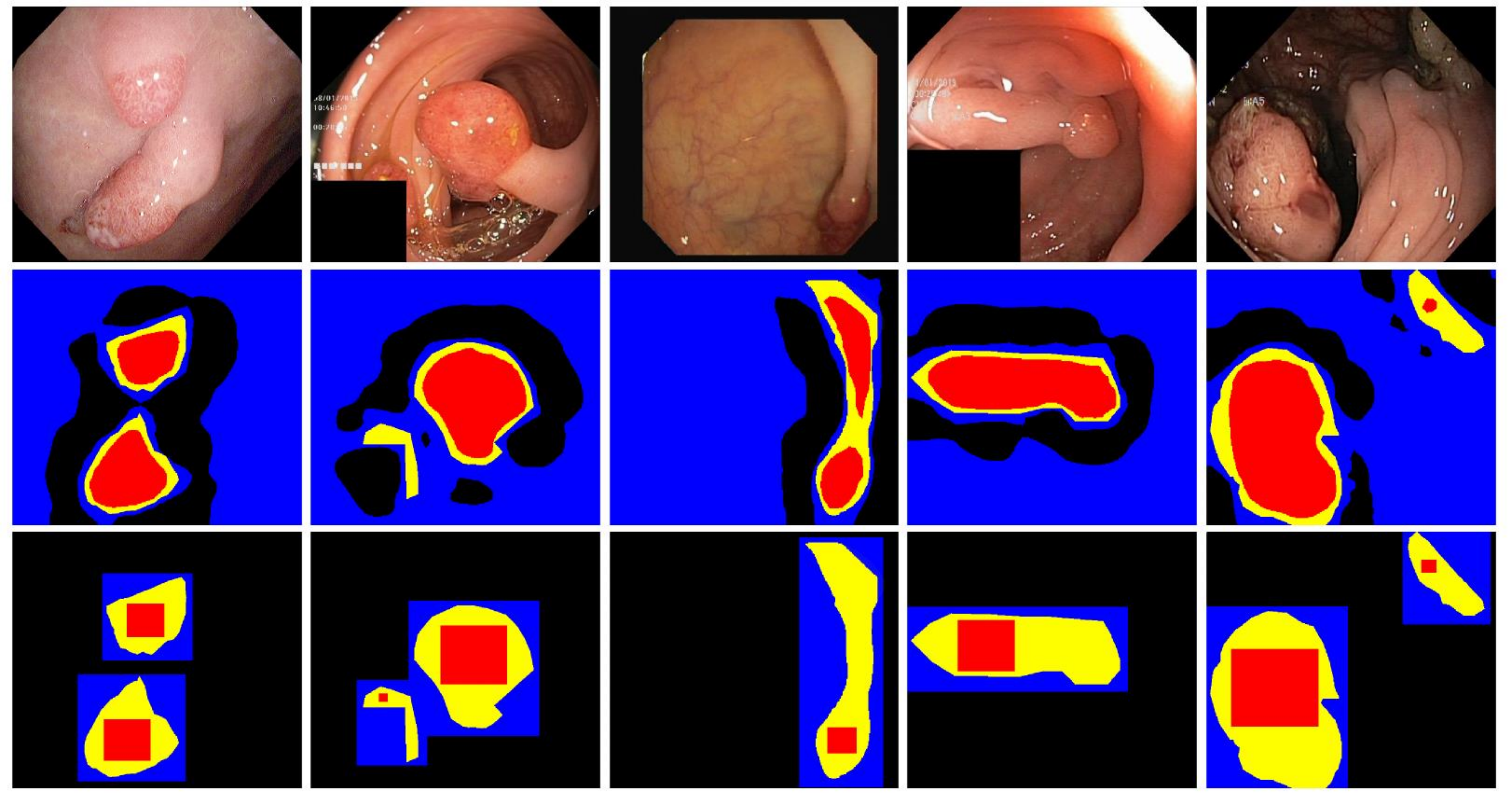
CONFIDENCE SETS APPLICATION

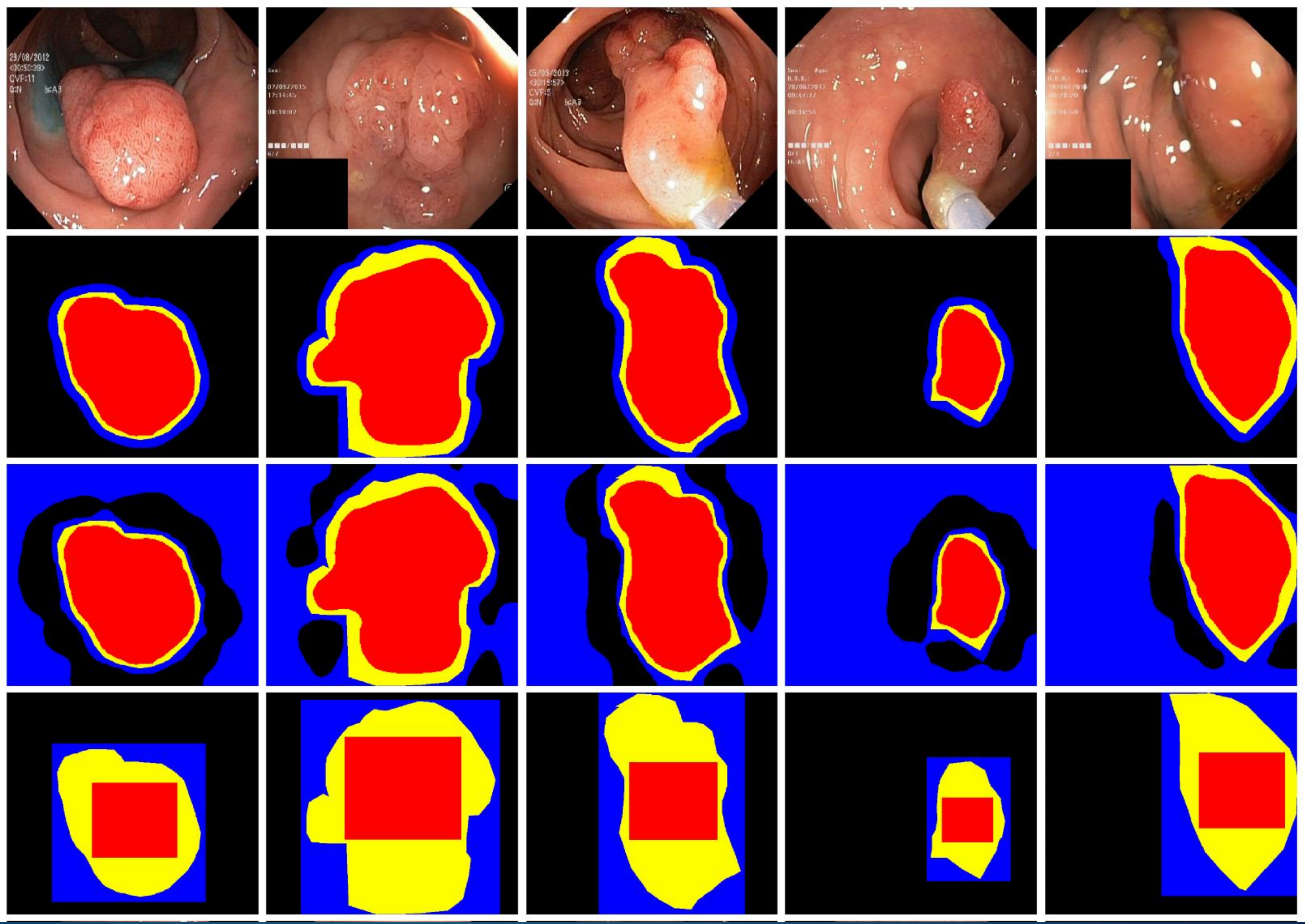


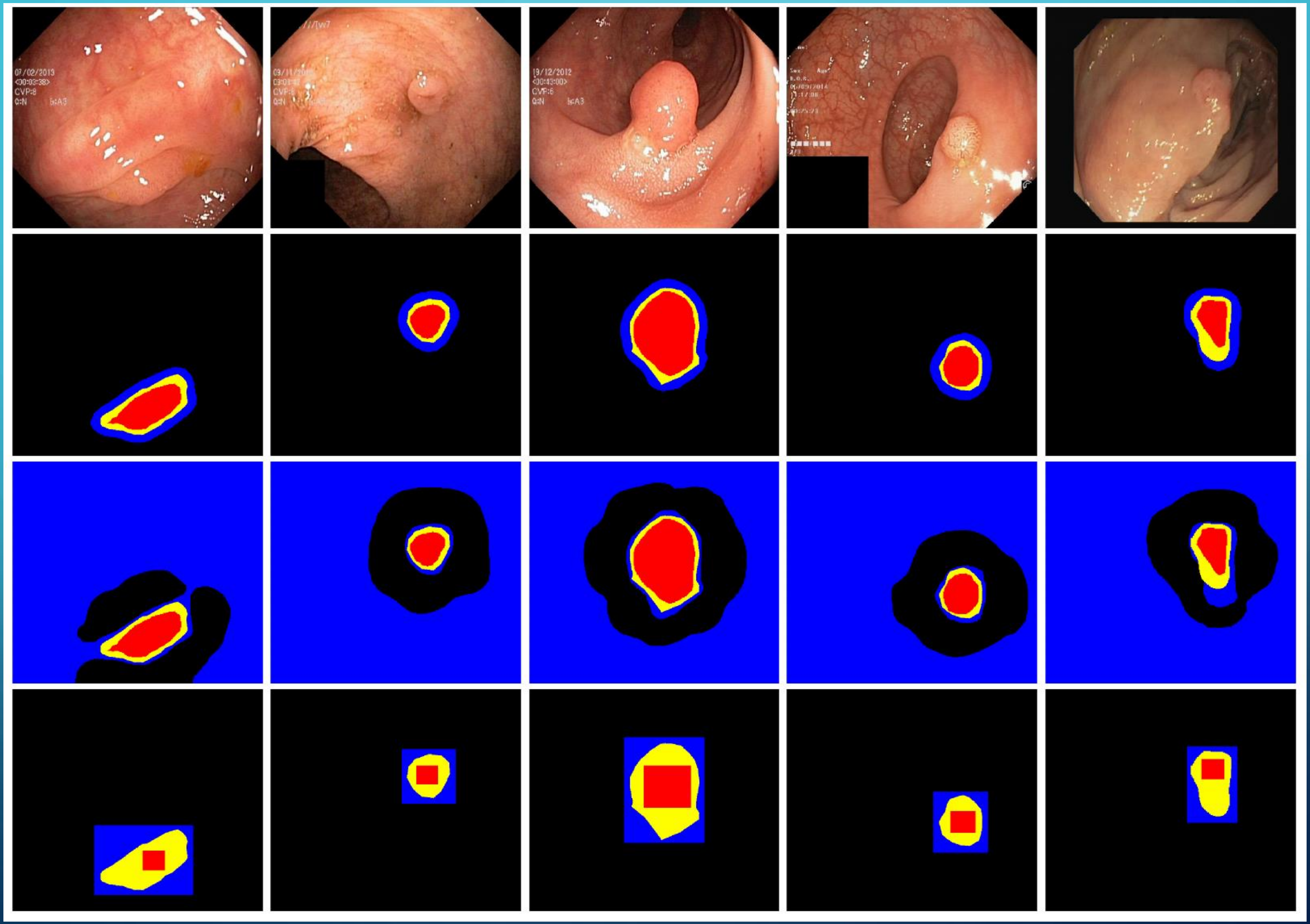
CONFIDENCE SETS APPLICATION



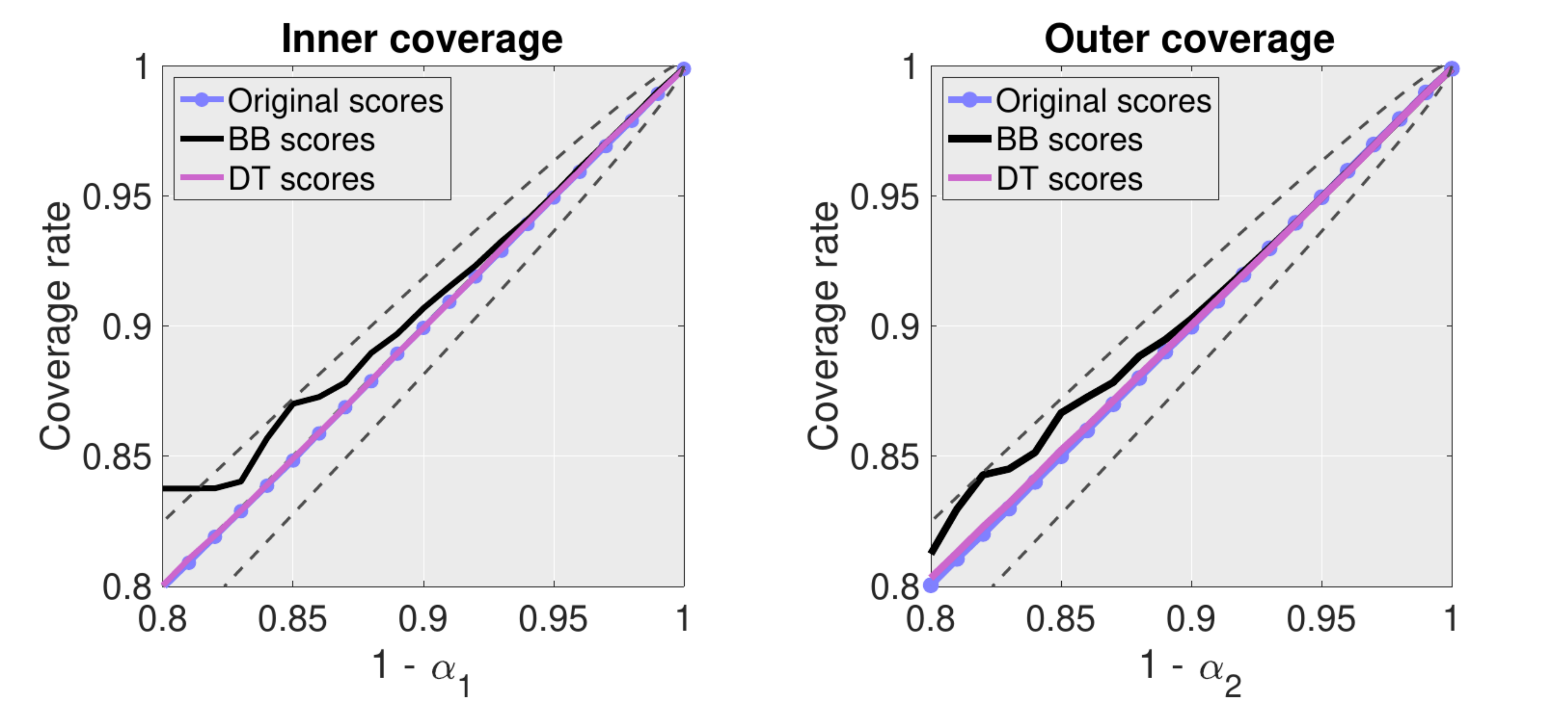
CONFIDENCE SETS APPLICATION



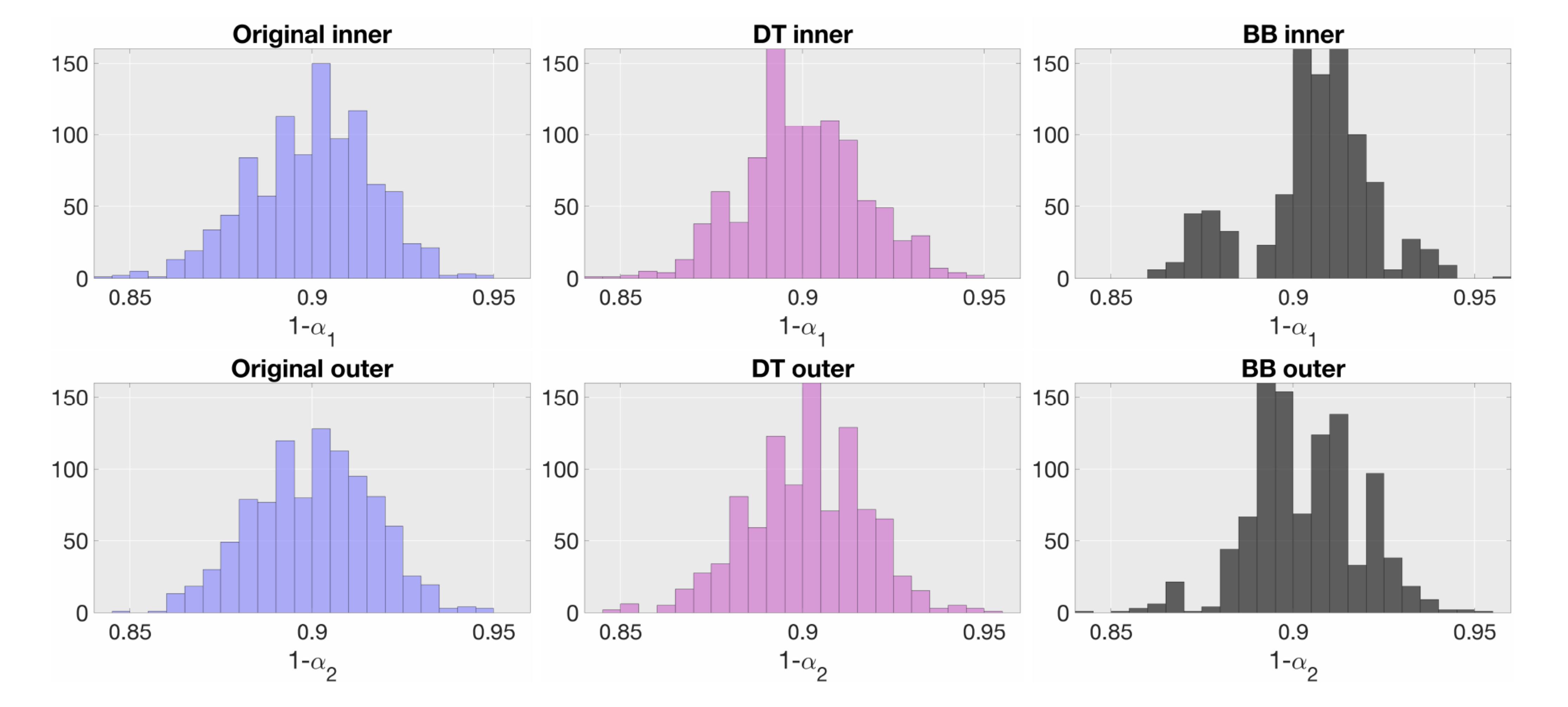




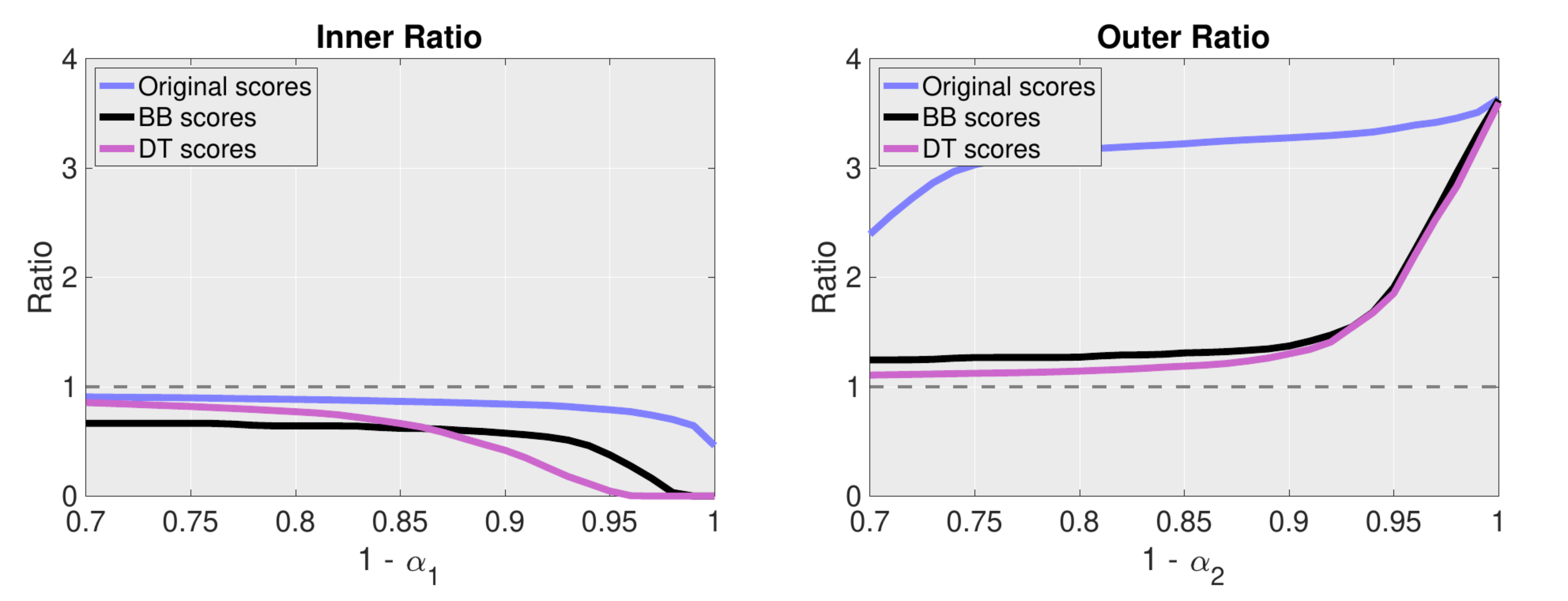
VERIFYING THE COVERAGE RATE



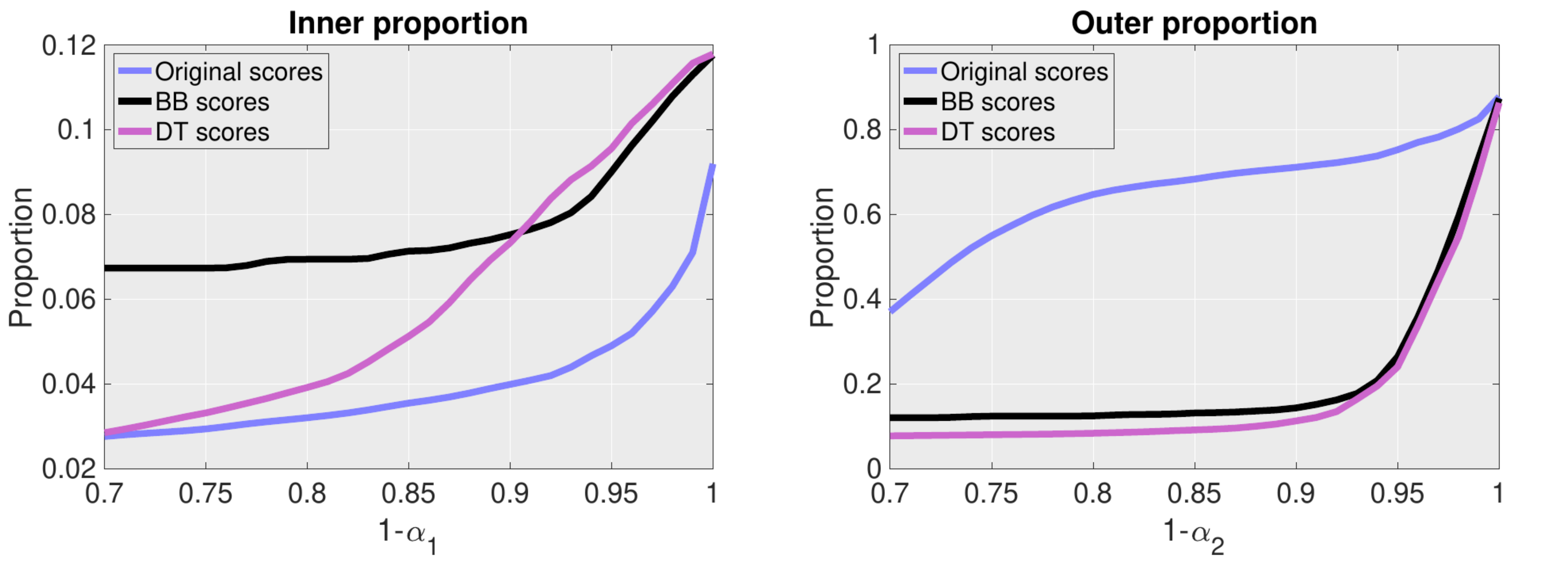
HISTOGRAM OF THE COVERAGE AT 90%



UNDERSTANDING THE EFFICIENCY



DETERMINING THE PROPORTION CAPTURED



CONCLUSIONS

- Conformal inference provides uncertainty guarantees for neural networks
- Transforming the scores (with transformations chosen on a learning dataset can lead to a big boost in performance).
- Conformal confidence sets provide strong guarantees whilst allowing for meaningful inference.
- Preprint available: Davenport, Samuel. "Conformal confidence sets for biomedical image segmentation." *arXiv preprint arXiv:2410.03406* (2024).

ADVANTAGE OF THE DISTANCE TRANSFORMED SCORES

Using distance transformed scores ensure that as the predicted masks improve the confidence sets improve – not true for using the original scores

Theorem 2.8. *For each $v \in \mathcal{V}$, let $f_O(s(X), v) = d_\rho(\hat{M}(X), v)$ and define $O(X)$ as in Section $\boxed{2.2}$. Suppose that $H_\rho(\hat{M}(X_i), Y_i) \leq k$, some $k \in \mathbb{R}$, for all $i \in J$, for some $J \subseteq \{1, \ldots, n\}$ such that $\frac{|J|}{n} > 1 - \alpha_2$. Then $H_\rho(\hat{M}(X_{n+1}), O(X_{n+1})) \leq k$. In particular if $H_\rho(\hat{M}(X_{n+1}), Y_{n+1}) \leq k$, then it follows that $H_\rho(O(X_{n+1}), Y_{n+1}) \leq 2k$.*

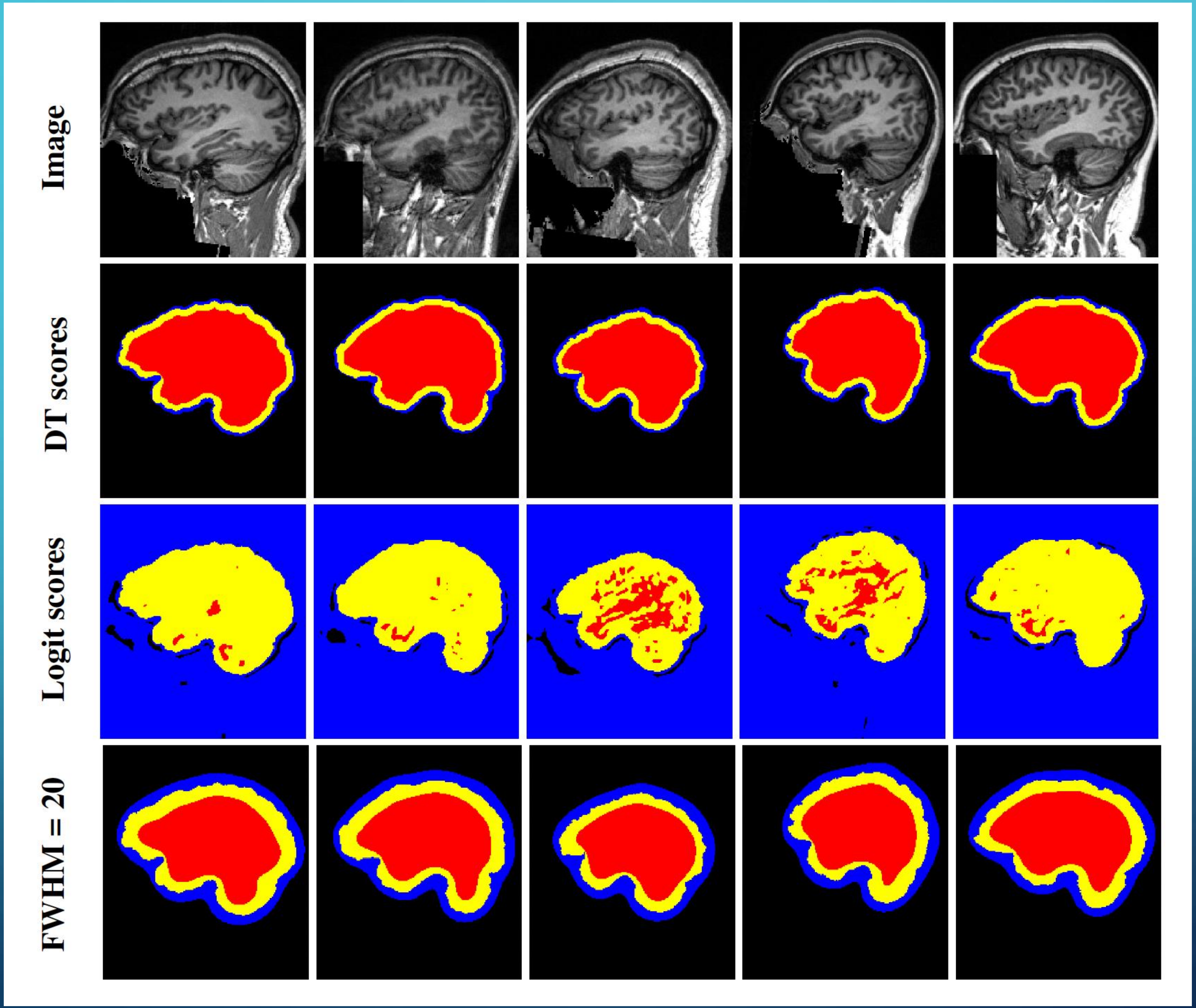
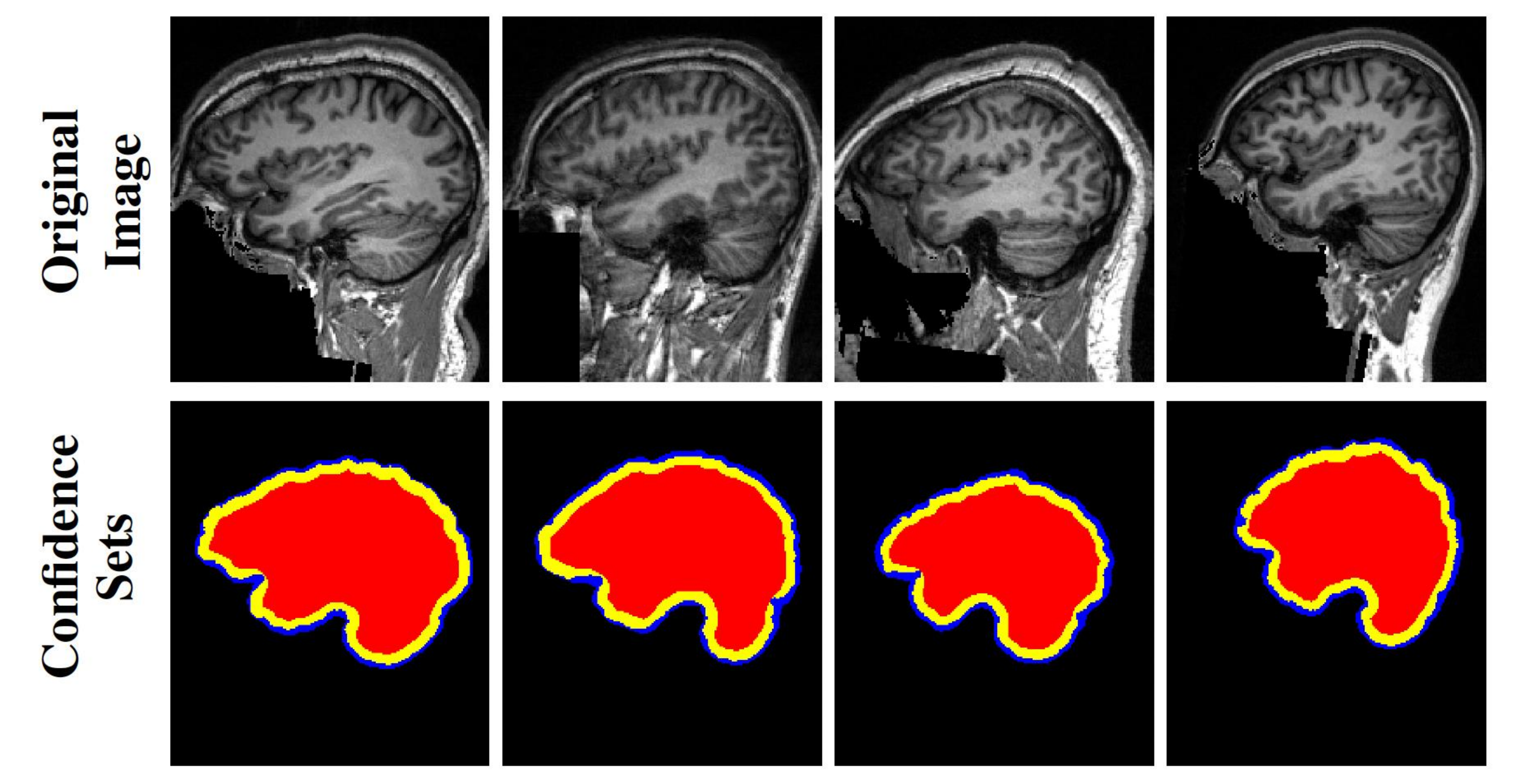


Image
DT scores
Logit scores
FWHM = 20


Original Image
Confidence Sets

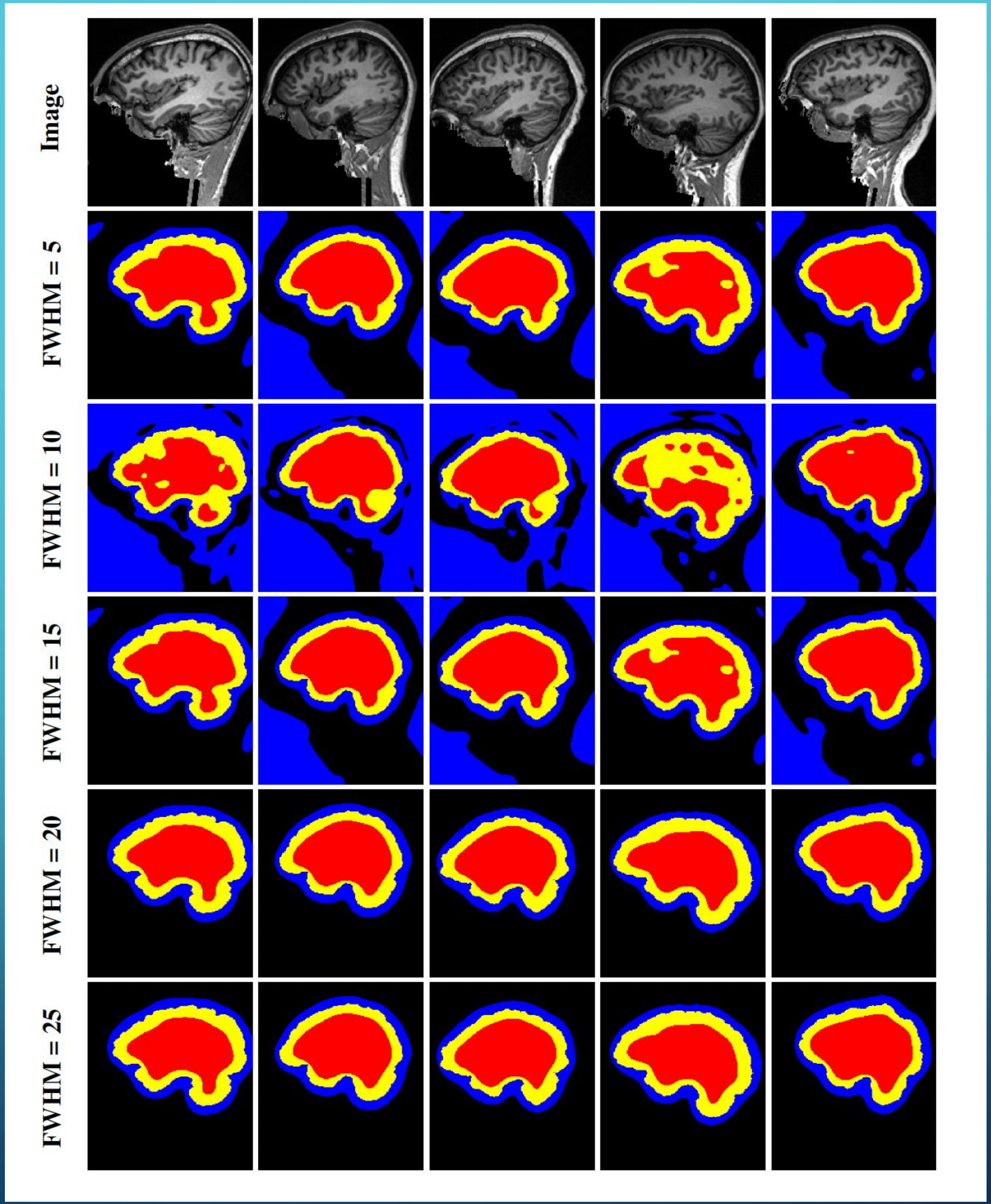
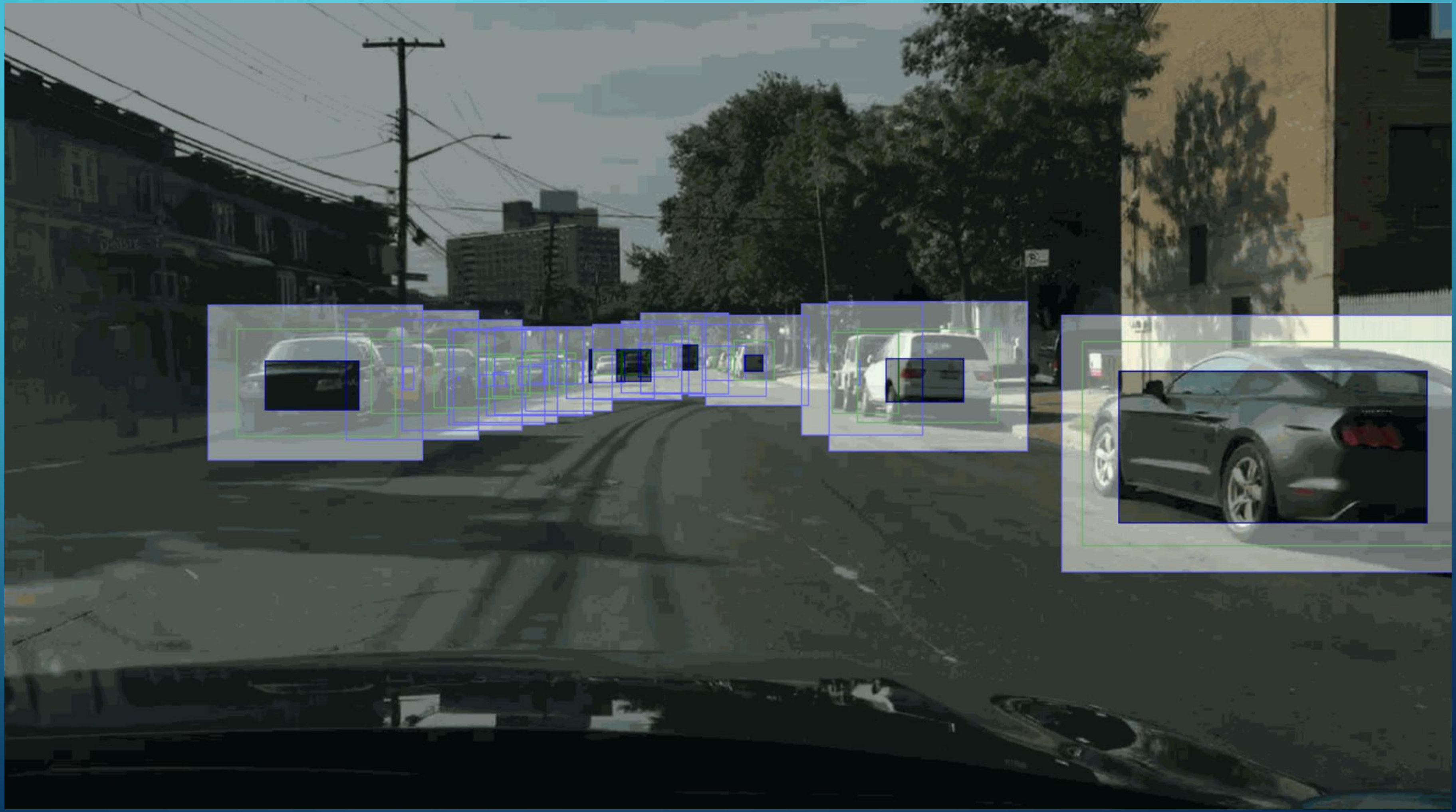


Image
FWHM = 5
FWHM = 10
FWHM = 15
FWHM = 20
FWHM = 25

EXISTING APPROACHES FOCUS ON BOUNDING BOXES

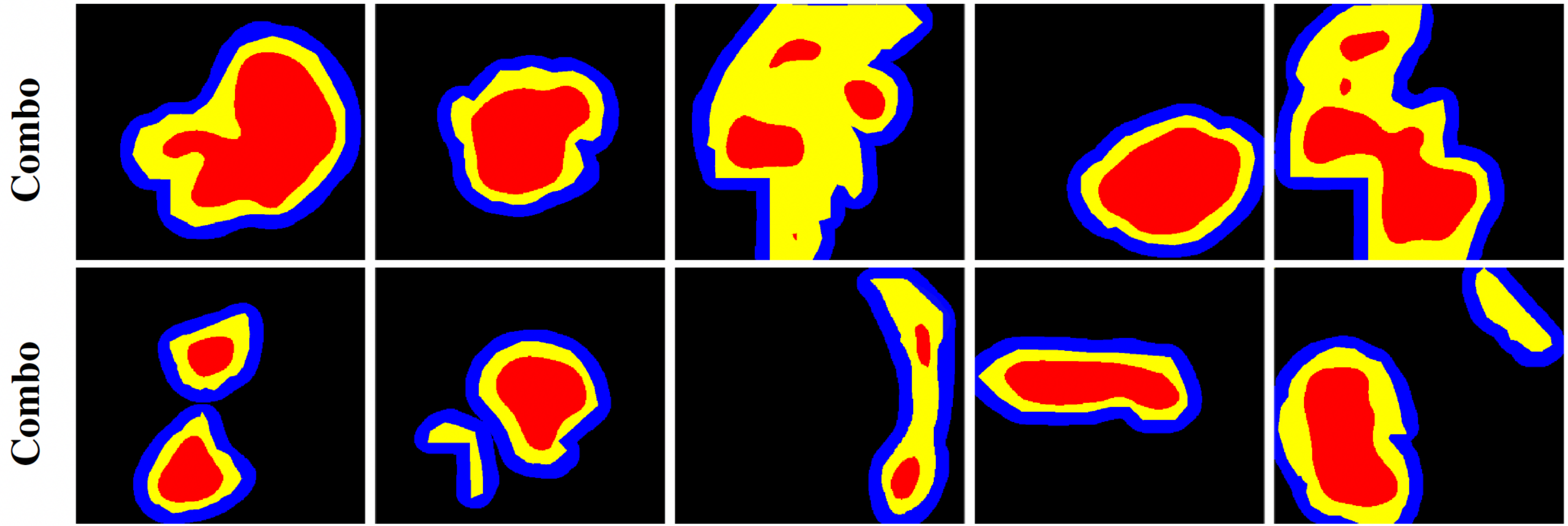


JOINT INFERENCE

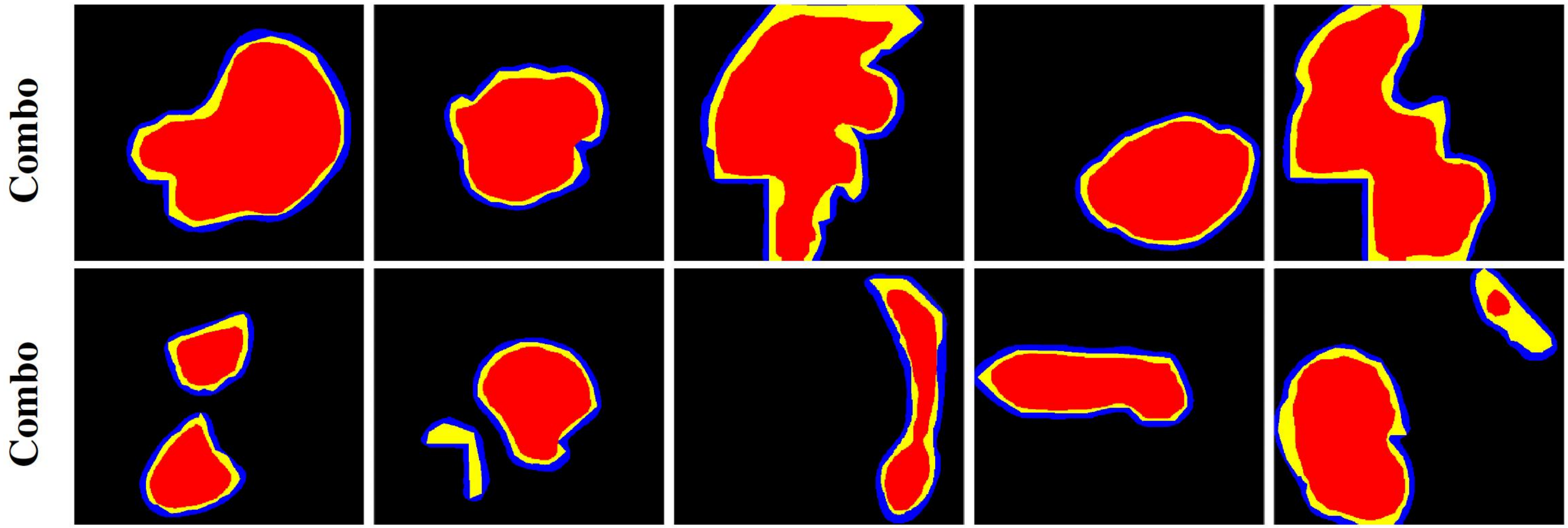
Corollary 2.5. *(Joint from marginal) Assume Assumptions* $\boxed{1}$ *and* $\boxed{2}$ *hold and given* $\alpha \in (0,1)$ *and* $\alpha_1, \alpha_2 \in (0,1)$ *such that* $\alpha_1 + \alpha_2 \leq \alpha$*, define* $I(X)$ *and* $O(X)$ *as in Theorems* $\boxed{2.1}$ *and* $\boxed{2.2}$*. Then*

$$\mathbb{P}\left(I(X_{n+1}) \subseteq \{v \in \mathcal{V} : Y_{n+1}(v) = 1\} \subseteq O(X_{n+1})\right) \geq \frac{\lceil (1-\alpha)(n+1) \rceil}{n}. \tag{5}$$

A.6.5 Joint 90% confidence regions



A.6.6 Marginal 80 % Confidence regions



A.6.7 Marginal 95 % Confidence regions