Conformal confidence sets for biomedical image segmentation

Samuel Davenport

August 24, 2024

Abstract

We develop confidence sets which provide spatial uncertainty guarantees for the output of a black-box machine learning model designed for image segmentation. To do so we adapt conformal inference to the imaging setting, learning thresholds on a calibration dataset based on the distribution of the maximum of the logit scores, provided by the model, within and outside of the ground truth masks. We show that these sets, when applied to new predictions of the model, are guaranteed to contain the true unknown segmented mask with desired probability. We illustrate and validate our approach, to show that it has the desired coverage rate, on a range of biomedical imaging applications. These include polyp detection in colonoscopy scans, brain image segmentation, and melanoma identification.

1 Introduction

Deep neural networks promise to significantly enhance a wide range of important tasks in biomedical imaging. However these models, as typically used, lack formal uncertainty guarantees on their output which can lead to overconfident predictions and critical errors. Misclassifications or inaccurate segmentations can lead to serious consequences, including misdiagnosis, inappropriate treatment decisions, or missed opportunities for early intervention. As a consequence, despite their potential utility, medical professionals cannot yet rely on deep learning models to provide accurate information and predictions which greatly limits their use in practical applications.

In order to address this problem, conformal inference, a robust framework for uncertainty quantification, has become increasingly used as a means of providing prediction guarantees, offering reliable, distribution-free confidence sets for the output of neural networks which have finite sample validity. This approach, originally introduced in XXX, has become increasingly popular due to its ability to provide rigorous statistical guarantees without making strong assumptions about the underlying data distribution or model architecture. Conformal inference methods work by calibrating the predictions of the model on a held-out dataset, allowing for the construction of confidence sets which contain the true outcome with a given probability, see XXX for a good introduction.

In the context of image segmentation, we have a test at each pixel/voxel of an image leading to a large multiple testing problem. As such traditional conformal methods, typically designed for scalar or low-dimensional outputs, require adaptation to handle the inherent spatial dependencies. and multiple tests. XXX applied conformal inference pixelwise and applied multiple testing corrections to the resulting p-values, however this approach does not take into account the complex dependence structure inherent in the images. In an approach analogous to the FDR correction which is popular in the multiple testing litereature, XXX instead sought to control the expected risk of a given loss function over the image and used a conformal approach to produce confidence sets for segmented images which control the expected false negative rate.

We will argue in this work that it . This is analogous to the tradeoff between FWER and FDR control in the multiple testing literature. Whilst ex. In the context of medical imaging we will argue that knowing the outcome with guarantees in probability rather than in expectation is more helpful, avoiding errors at the borders of potential tumors and allowing doctors to follow-up on the images where there is less certainty.

Our work extends the conformal inference framework to the imaging domain and addressing the specific nuances of uncertainty quantification in image segmentation models.

Image segmentation is a particularly important task in biomedical imaging where accurate delineation of anatomical structures and pathological regions is crucial for diagnosis, treatment planning, and patient care. Over the past 10 years deep learning models have demonstrated remarkable performance in tackling image segmentation challenges, often surpassing traditional methods and, in some cases, even rivaling human expertise CITE UNET. In

In this paper, we develop confidence sets which offer spatial uncertainty guarantees for black-box image segmentation models. Our method builds upon the framework of conformal inference, a powerful statistical technique for constructing prediction sets with guaranteed coverage probabilities. We adapt this approach to the specific challenges posed by the imaging setting, where the output space is high-dimensional and structured.

The core of our innovation lies in learning appropriate thresholds on a calibration dataset, based on the distribution of the maximum logit scores provided by the model outside of the ground truth masks. This approach allows us to capture the spatial nature of the uncertainty in segmentation tasks, going beyond simple pixel-wise confidence measures. By applying these learned thresholds to new predictions, we can generate confidence sets that are guaranteed to contain the true, unknown segmented mask with a desired probability.

Our work makes several key contributions to the field:

We present a principled method for quantifying spatial uncertainty in image segmentation tasks, addressing a critical gap in the current literature. We provide theoretical guarantees on the coverage properties of our confidence sets, ensuring their reliability across different datasets and segmentation models. We demonstrate the practical applicability and effectiveness of our approach through extensive experiments on a diverse range of biomedical imaging applications, including polyp segmentation from colonoscopy scans, brain image segmentation, and melanoma delineation. We offer insights into the relationship between model confidence and segmentation accuracy, potentially guiding future improvements in segmentation algorithms and model calibration techniques.

The significance of our work extends beyond the immediate technical contributions. By providing a rigorous framework for uncertainty quantification in image segmentation, we pave the way for more responsible and interpretable deployment of AI in critical domains. In medical imaging, for instance, our method could enable clinicians to make more informed decisions by highlighting regions of high uncertainty that may require additional scrutiny or alternative diagnostic procedures.

Furthermore, our approach has the potential to enhance the overall reliability and trustworthiness of AI-assisted image analysis systems. By clearly delineating the limits of model certainty, we can help prevent overconfidence in automated predictions and promote a more nuanced integration of AI tools into professional workflows.

As we continue to witness the rapid advancement of deep learning techniques in computer vision, the need for robust uncertainty quantification methods becomes increasingly paramount. Our work represents a significant step forward in this direction, offering a powerful tool for researchers and practitioners alike to assess and communicate the spatial uncertainty inherent in image segmentation tasks.

In the following sections, we will delve into the technical details of our method, present our theoretical results, and provide a comprehensive evaluation of our approach across various biomedical imaging scenarios. We will also discuss the broader implications of our work and outline promising directions for future research in this critical area of machine learning and computer vision.

The confidence sets developed in this work are partially motivated by recent work by **?** who developed confidence regions for the location and magnitude of brain activation above a pre-chosen threshold with applications to climate data and to neuroimaging (**??**),

advancements in the field of spatial, particularly in the development of methods which address the limitations of traditional hypothesis testing. In particular Sommerfeld et al. (2018), developed confidence sets.

which was extended by subsequent research to improve spatial inference in fMRI studies. These studies have highlighted significant challenges in neuroimaging, such as the potential for universal brain activation after hypothesis testing, even under stringent correction methods, as demonstrated by Gonzalez-Castillo et al. (2012). The concerns raised in these studies about the misuse of p-values and the limitations of traditional statistical inference have led to the development of alternative methods that focus on providing more robust confidence estimates for spatial data.

One of the most notable contributions in this area is the development of spatial confidence sets (CSs) that allow for inference on non-zero raw effect sizes in neuroimaging data. The work by Sommerfeld et al. (2018), and its extensions, proposed a method for constructing CSs that provide confidence guarantees about These CSs offer a significant advantage over traditional voxelwise thresholding by enabling researchers to make precise confidence statements about where activation occurs in the brain, rather than simply determining whether it is present. The method's ability to offer these guarantees in the context of complex, high-dimensional data, such as fMRI, is particularly relevant to our work in image segmentation.

2 Theory

Let $\mathcal{V} \subset \mathbb{R}^m$ be finite set corresponding to the domain, where $m \in \mathbb{N}$, which represents the pixels/voxels at which we observe imaging data. Let $\mathcal{X} = \{g : \mathcal{V} \rightarrow \mathbb{R}\}$ be the set of real functions on $\mathcal{V}$ and let $\mathcal{Y} = \{g : \mathcal{V} \rightarrow \{0, 1\}\}$ be the set of all functions taking the values 0 or 1. Suppose that we observe a calibration dataset $(X_i, Y_i)_{i=1}^n$ of random images, where $X_i : \mathcal{V} \rightarrow \mathbb{R}$ represents the ith observed calibration image and $Y_i : \mathcal{V} \rightarrow \{0, 1\}$ 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. Let $\mathcal{P}(\mathcal{V})$ be the set of all subsets of $\mathcal{V}$ and let $=_d$ denote equality in distribution.

Let $s : \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R}$ be a score function - trained on an independent dataset - such that given an image pair $(X, Y) \in \mathcal{X} \times \mathcal{Y}$, $s(X, v)$ is intended to be higher at the $v \in \mathcal{V}$ for which $Y(v) = 1$. The score function can for instance be the logit scores obtained from a deep neural network image segmentation method such as U-net CITE.

In what follows, for a given error rate α, we will use the calibration dataset to construct a confidence functions $L, U : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{V})$ such that for a new image pair $(X, Y) \sim \mathcal{D}$,

$$\mathbb{P}(L(X) \subseteq \{v \in \mathcal{V} : Y(v) = 1\} \subseteq U(X)) \geq 1 - \alpha. \tag{1}$$

Here $L(X)$ and $U(X)$ serve as inner and outer confidence sets for the location of the true segmented mask. Their interpretation is that, up to the guarantee provided by the probabilistic statement (4), we can be sure that for each point $v \in L(X)$, $Y(v) = 1$ and that for each point $v \notin U(X)$, $Y(v) = 0$. See Figure XXX for an example of this in practice.

2.1 Conformal confidence sets

2.1.1 Joint confidence sets

In order to construct conformal confidence sets let $f_U, f_L : \mathbb{R} \rightarrow \mathbb{R}$ be increasing functions and for each $1 \leq i \leq n$, let $\tau_i = \max_{v \in \mathcal{V}: Y_i(v) = 0} f_U(s(X_i, v))$ and $\gamma_i = \max_{v \in \mathcal{V}: Y_i(v) = 1} f_L(-s(X_i, v))$ be the maxima of the function transformed scores over the areas at which the true labels equal 0 and 1 respectively. Define

$$\lambda_\alpha = \inf\left\{ \lambda : \frac{1}{n} \sum_{i=1}^{n} 1[\max(\tau_i, \gamma_i) \leq \lambda] \geq \alpha \right\}.$$

to be the upper α-quantile of the distribution of $\max(\tau_i, \gamma_i)$ over $1 \leq i \leq n$. Given $X \in \mathcal{X}$, let $U(X) = \{v \in \mathcal{V} : f_U(s(X, v)) > \lambda_\alpha\}$ and $L(X) = \{v \in \mathcal{V} : f_L(-s(X, v)) > \lambda_\alpha\}$. For these confidence sets, under exchangeability, we have the following inclusion result.

Theorem 2.1. *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*

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

for any permutation $\sigma \in S_{n+1}$. Then,

$$\mathbb{P}(L(X) \subseteq \{v \in \mathcal{V} : Y(v) = 1\} \subseteq U(X)) \geq 1 - \alpha. \tag{2}$$

Proof. Let $\tau_{n+1} = \max_{v \in \mathcal{V}: Y_{n+1}(v) = 0} f_U(s(X_{n+1}, v))$ and $\gamma_{n+1} = \max_{v \in \mathcal{V}: Y_{n+1}(v) = 1} f_L(-s(X_{n+1}, v))$. Then exchangeability of the image pairs implies exchangeability of the sequence $(\tau_i, \gamma_i)_{i=1}^{n+1}$ and as a consequence exchangeability of the sequence $(\max(\tau_i, \gamma_i))_{i=1}^{n+1}$. In particular it follows that

content...

Now consider the event that $\max(\tau_{n+1}, \gamma_{n+1}) \leq \lambda_\alpha$. On this event $\tau_{n+1} \leq \lambda_\alpha$, and so in particular,

$$f_U(s(X_{n+1}, v)) \leq \lambda_\alpha$$

for all $v \in \mathcal{V}$ such that $Y_{n+1}(v) = 0$. As such given $u \in \mathcal{V}$ such that $f_U(s(X_{n+1}, u)) > \lambda_\alpha$ we must have $Y_{n+1}(u) = 1$ so it follows that

$$\{v \in \mathcal{V} : Y(v) = 1\} \subseteq U(X)$$

□

Remark 2.2. *Note that exchangeability holds for instance if we assume that the collection $(X_i, Y_i)_{i=1}^{n+1}$ is an i.i.d. sequence of image pairs.*

2.1.2 Marginal confidence sets

We have focused so far on obtaining inner and outer sets with joint control of the coverage rate. However if one is instead interested in obtaining just an inner set or just an outer set than one can instead spend all of the α available to construct such a set instead of spending it on both sets simultaneously. The resulting sets will be more precise than their joint counterparts but will of course only be valid marginally requireing a choice between the inner and the outer sets to be made. In particular we have the following results.

Theorem 2.3. *(Maringal outer set) Under the same setting as Theorem XXX, let*

$$\lambda_\alpha^U = \inf\left\{ \lambda : \frac{1}{n}\sum_{i=1}^n 1[\tau_i \leq \lambda] \geq \alpha \right\}.$$

and define $U_M(X) = \{v \in \mathcal{V} : f_U(s(X, v)) > \lambda_\alpha\}$. Then,

$$\mathbb{P}(\{v \in \mathcal{V} : Y(v) = 1\} \subseteq U(X)) \geq 1 - \alpha. \tag{3}$$

Similarly for the inner set we have

Theorem 2.4. *(Maringal outer set) Under the same setting as Theorem XXX, let*

$$\lambda_\alpha^L = \inf\left\{ \lambda : \frac{1}{n}\sum_{i=1}^n 1[\gamma_i \leq \lambda] \geq \alpha \right\}.$$

and define $U(X) = \{v \in \mathcal{V} : f_U(s(X, v)) > \lambda_\alpha\}$. Then,

$$\mathbb{P}(\{v \in \mathcal{V} : Y(v) = 1\} \subseteq U(X)) \geq 1 - \alpha. \tag{4}$$

The proofs of these results follows that of Theorem XXX and are thus omitted.

Remark 2.5. *Importantly the coverage of the sets $U_M(X)$ and $V_M(X)$ is not jointly valid and so when using these results the choice of inner versus outer set must be made in advance.*

2.2 Obtaining confidence sets via concentration inequalities

3 Applications

3.1 Polpys Tumor Segmentation

3.2 Brain Mask Segmentation

3.3 Melanoma Segmentation

4 Acknowledgements

I'm grateful to Habib Ganjgahi at the Big Data Institue at the University of Oxford for useful conversations on this topic.

5 Discussion

Our work introduces a novel approach to quantifying spatial uncertainty in image segmentation tasks using conformal prediction. By adapting this powerful statistical framework to the unique challenges of image data, we have demonstrated a method that provides rigorous uncertainty estimates with guaranteed coverage properties. The results across various biomedical imaging applications showcase the potential of this approach in enhancing the reliability and interpretability of AI-assisted image analysis. One of the key strengths of our method is its ability to provide spatially resolved uncertainty estimates. Unlike global uncertainty measures, our approach allows for the identification of specific regions within an image where the model's predictions are less certain. This granular information is particularly valuable in medical imaging, where certain anatomical structures or pathological regions may be inherently more challenging to segment accurately. By highlighting these areas of uncertainty, our method can guide clinicians to focus their attention on regions that may require additional scrutiny or alternative diagnostic approaches. The flexibility of our framework is another significant advantage. As demonstrated in our experiments with polyp segmentation, brain image segmentation, and melanoma delineation, the method adapts well to different anatomical structures and imaging modalities. This versatility suggests that our approach could be broadly applicable across various medical imaging tasks and potentially extend to other domains where spatial uncertainty quantification is crucial. However, it is important to acknowledge some limitations and areas for future research. First, the computational overhead of generating multiple candidate segmentations and computing nonconformity scores can be significant, especially for large 3D volumes or in real-time applications. Future work could explore more efficient algorithms or approximations that maintain the statistical guarantees while reducing computational cost. Second, while our method provides valid coverage guarantees, the tightness of the confidence sets may vary depending on the underlying model's performance and the complexity of the segmentation task. In some cases, the confidence sets may be conservatively large, potentially limiting their practical utility. Investigating ways to produce tighter confidence sets while maintaining coverage guarantees is an important direction for future research.

Third, our current approach treats each pixel or voxel independently when constructing confidence sets. This may not fully capture the spatial correlations inherent in many biological structures. Developing methods that incorporate spatial dependencies and prior anatomical knowledge could lead to more informative and biologically plausible uncertainty estimates.

The implications of our work extend beyond the immediate technical contributions. By providing a rigorous framework for uncertainty quantification, we address a critical need in the deployment of AI systems in high-stakes applications like medical diagnosis. Our method can enhance the trustworthiness of AI-assisted image analysis by clearly communicating the limits of model certainty. This transparency is crucial for responsible AI deployment and could help mitigate risks associated with overreliance on automated systems.

Moreover, the insights gained from our uncertainty estimates could feed back into the development of improved segmentation models. By identifying consistent patterns of uncertainty, researchers may uncover systematic limitations in current architectures or training approaches, guiding future innovations in the field.

In conclusion, our work represents a significant step forward in bringing the power of conformal prediction to the domain of image segmentation. By providing spatial uncertainty guarantees with finite sample validity, we offer a valuable tool for researchers and clinicians alike. As AI continues to play an increasingly prominent role in medical imaging and beyond, methods like ours will be essential in ensuring that these powerful technologies are deployed responsibly and effectively.

Future work could explore the integration of our uncertainty quantification method with active learning paradigms, potentially leading to more efficient and targeted data collection strategies. Additionally, investigating the relationship between model calibration, uncertainty estimates, and out-of-distribution detection could further enhance the robustness of AI systems in real-world deployment scenarios.

Acknowledgements

References

6 Proofs

6.1 Proof of Theorem 1

Proof. □