

# Appendices

---



---

## Appendix A. Equal Area Circle Maximises IoU with Ellipse - Proof

**Proof.**

**Proposition.** *Let  $E$  be a fixed ellipse with semi-axes  $a > b$  and area  $A_E = \pi ab$ . Let  $C$  be a concentric circle with radius  $R$  and area  $A_C = \pi R^2$ . The Intersection over Union  $IoU(C, E)$  is globally maximised when  $A_C = A_E$ .*

We observe that for  $r \geq a$  we have:

$$IoU = \frac{\pi ab}{\pi R^2}, \frac{dIoU}{dR} = -\frac{ab}{R^3}$$

which is monotonically decreasing. For  $R \leq b$  we have:

$$IoU = \frac{\pi R^2}{\pi ab}, \frac{dIoU}{dR} = \frac{2R}{ab}$$

which is monotonically increasing, as such no maxima exist in these regions and we reduce our investigation to only  $b < R < a$ . We can express the IoU in polar coordinates as:

$$IoU(R) = \frac{I(R)}{U(R)} = \frac{\frac{1}{2} \int_0^{2\pi} \min(R^2, r_E^2(\theta)) d\theta}{\frac{1}{2} \int_0^{2\pi} \max(R^2, r_E^2(\theta)) d\theta}$$

where  $r_E(\theta)$  is the distance from the origin to the ellipse. We define the region where  $R^2 \leq r_E^2(\theta)$  as:

$$\Omega_R = \{\theta : R \leq r_E(\theta)\}, \Omega_R^c = \{\theta : R > r_E(\theta)\}$$

We can also rewrite the IoU in terms of the intersection only as:

$$IoU(R) = \frac{I(R)}{\pi R^2 + \pi ab - I(R)}$$

with the derivative:

$$\frac{dIoU}{dR} = \frac{I'(R)(\pi R^2 + \pi ab) - 2\pi RI(R)}{(\pi R^2 + \pi ab - I(R))^2}.$$

we are interested in the critical points which occur when:

$$\boxed{I'(R)(R^2 + ab) = 2RI(R)}$$

Meaning we are interested in evaluating the intersection and its derivative in terms of  $R$ . Our intersection term can be expressed as:

$$I(R) = \frac{1}{2} \int_0^{2\pi} R^2 \mathbf{1}_{\Omega_R}(\theta) d\theta + \frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R^c}(\theta) d\theta$$

where  $\mathbf{1}_A(\theta)$  is an indicator function which takes the value 1 if  $\theta \in A$  and 0 otherwise. The first term of the intersection is trivially:

$$\frac{1}{2} \int_0^{2\pi} R^2 \mathbf{1}_{\Omega_R}(\theta) d\theta = \frac{1}{2} R^2 |\Omega_R|$$

where  $|\Omega_R|$  is the angular length of  $\Omega_R$  i.e. the size of the set of angles in  $\Omega_R$ . The second term can be expressed as:

$$\frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R^c}(\theta) d\theta = \frac{1}{2} \left[ \int_0^{2\pi} r_E^2(\theta) d\theta - \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R}(\theta) d\theta \right] = \frac{1}{2} \left[ 2\pi ab - \int_{\Omega_R} r_E^2(\theta) d\theta \right]$$

The intersection term becomes:

$$\boxed{I(R) = \frac{1}{2} R^2 |\Omega_R| + \pi ab - \frac{1}{2} \int_{\Omega_R} r_E^2(\theta) d\theta}$$

We write the intersection as

$$I(R) = \frac{1}{2} \int_{\Omega_R} R^2 d\theta + \frac{1}{2} \int_{\Omega_R^c} r_E^2(\theta) d\theta,$$

We differentiate the first term using Leibniz' rule for a parameter-dependent domain:

$$\frac{d}{dR} \left( \frac{1}{2} \int_{\Omega_R} R^2 d\theta \right) = \frac{1}{2} \left( 2R |\Omega_R| + R^2 \frac{d}{dR} |\Omega_R| \right).$$

We write

$$I(R) = \frac{1}{2} \int_0^{2\pi} R^2 \mathbf{1}_{\Omega_R}(\theta) d\theta + \frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R^c}(\theta) d\theta.$$

Differentiating the first term gives

$$\frac{d}{dR} \left( \frac{1}{2} \int_0^{2\pi} R^2 \mathbf{1}_{\Omega_R}(\theta) d\theta \right) = R|\Omega_R| + \frac{1}{2} R^2 \frac{d}{dR} |\Omega_R|.$$

For the second term, we use

$$\mathbf{1}_{\Omega_R^c} = 1 - \mathbf{1}_{\Omega_R},$$

so that

$$\frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R^c}(\theta) d\theta = \frac{1}{2} \int_0^{2\pi} r_E^2(\theta) d\theta - \frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R}(\theta) d\theta.$$

The first term is constant in  $R$ , equal to  $\pi ab$ , hence its derivative is zero. For the second term, since  $\Omega_R$  depends on  $R$  only through a moving boundary, we apply the standard shape derivative identity:

$$\frac{d}{dR} \int_{\Omega_R} f(\theta) d\theta = \int_{\partial\Omega_R} f(\theta) V_R(\theta) dS,$$

where  $V_R(\theta)$  is the normal velocity of the boundary  $\partial\Omega_R$ .

Applying this with  $f(\theta) = \frac{1}{2} r_E^2(\theta)$  gives

$$\frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R}(\theta) d\theta = \frac{1}{2} \int_{\partial\Omega_R} r_E^2(\theta) V_R(\theta) dS.$$

On the boundary  $\partial\Omega_R$ , we have the defining relation

$$r_E(\theta) = R,$$

hence we have a constant:

$$r_E^2(\theta) = R^2 \quad \text{on } \partial\Omega_R.$$

Therefore,

$$\frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R}(\theta) d\theta = \frac{1}{2} R^2 \int_{\partial\Omega_R} V_R(\theta) dS.$$

$$\frac{d}{dR} \left( \frac{1}{2} \int_0^{2\pi} r_E^2(\theta) \mathbf{1}_{\Omega_R}(\theta) d\theta \right) = \frac{1}{2} R^2 \frac{d}{dR} |\Omega_R|.$$

Therefore overall,

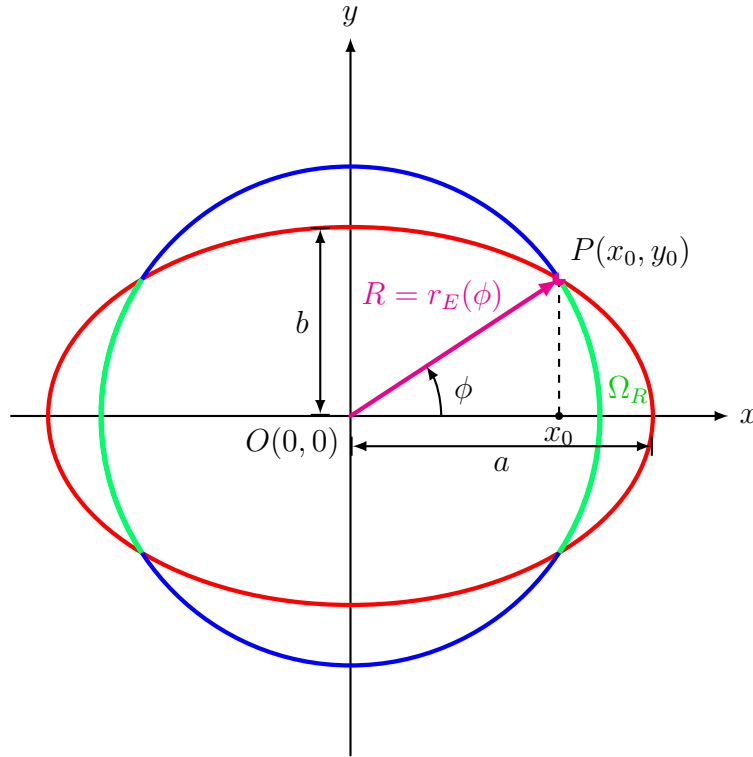
$$\frac{d}{dR} \left( \frac{1}{2} \int_{\Omega_R^c} r_E^2(\theta) d\theta \right) = -\frac{1}{2} R^2 \frac{d}{dR} |\Omega_R|.$$

And hence the total derivative of the intersection is given by,

$$I'(R) = \frac{1}{2} \left( 2R|\Omega_R| + R^2 \frac{d}{dR} |\Omega_R| \right) - \frac{1}{2} R^2 \frac{d}{dR} |\Omega_R|,$$

which simplifies to

$$\boxed{I'(R) = R|\Omega_R|}.$$



The total angle  $|\Omega_R|$  is defined by the angle of the intersection point  $R = r_E(\phi)$  for some  $\phi$ . We can calculate this by simply equating the polar

equation of the ellipse to  $R$ :

$$r_E^2(\phi) = \frac{a^2 b^2}{a^2 \sin^2 \phi + b^2 \cos^2 \phi} = R^2$$

solving for  $\phi$  we get:

$$\sin^2 \phi = \frac{b^2(a^2 - R^2)}{R^2(a^2 - b^2)}$$

or as a function of  $R$ :

$$\phi(R) = \arccos \left( \frac{a}{R} \sqrt{\frac{R^2 - b^2}{a^2 - b^2}} \right)$$

This is the solution for the upper right quadrant, so the total angular length is:

$$|\Omega_R| = 4\phi(R) = 4 \arccos \left( \frac{a}{R} \sqrt{\frac{R^2 - b^2}{a^2 - b^2}} \right)$$

Our expression for the critical points of the IoU can now be re-expressed:

$$I'(R)(R^2 + ab) = 2RI(R) \implies R|\Omega_R|(R^2 + ab) = 2R \left( \frac{1}{2}R^2|\Omega_R| + \pi ab - \frac{1}{2} \int_{\Omega_R} r_E^2(\theta) d\theta \right)$$

Which simplifies to:

$$ab(2\pi - |\Omega_R|) = \int_{\Omega_R} r_E^2(\theta) d\theta$$

This can be rearranged using  $\int_0^{2\pi} r_E^2(\theta) d\theta = 2\pi ab$  to:

$$ab(2\pi - |\Omega_R|) = 2\pi ab - \int_{\Omega_R^c} r_E^2(\theta) d\theta$$

This then cancels down to:

$$\int_{\Omega_R^c} r_E^2(\theta) d\theta = ab|\Omega_R|$$

Returning to our interrogation of  $|\Omega_R|$  earlier and using the symmetry of the problem to rewrite the integral we can write this as:

$$4 \int_{\phi}^{\pi/2} r_E^2(\theta) d\theta = ab(4\phi) \implies \int_{\phi}^{\pi/2} r_E^2(\theta) d\theta = ab\phi$$

We evaluate this integral for the parametric form of the ellipse:

$$\int_{\phi}^{\pi/2} r_E^2(\theta) d\theta = \int_{\phi}^{\pi/2} \frac{a^2 b^2}{a^2 \sin^2 \theta + b^2 \cos^2 \theta} d\theta = \int_{\tan \phi}^{\infty} \frac{a^2 b^2}{a^2 u^2 + b^2} du$$

where we divide numerator and denominator by  $\cos^2 \theta$  and substituting  $u = \tan \theta$ ,  $du = \sec^2 \theta d\theta$ . This evaluates to:

$$\int_{\tan \phi}^{\infty} \frac{a^2 b^2}{a^2 u^2 + b^2} du = ab \arctan\left(\frac{au}{b}\right) \Big|_{\tan \phi}^{\infty} = ab \left( \frac{\pi}{2} - \arctan\left(\frac{a \tan \phi}{b}\right) \right)$$

using  $\arctan(x) + \arctan(1/x) = \pi/2$  for  $x > 0$ , we can rearrange this to get:

$$\int_{\phi}^{\pi/2} r_E^2(\theta) d\theta = ab \arctan\left(\frac{b}{a \tan \phi}\right)$$

Solving this for  $\phi$  we get:

$$ab \arctan\left(\frac{b}{a \tan \phi}\right) = ab\phi \implies \boxed{\tan^2 \phi = \frac{b}{a}}$$

Now we can use the definition of  $\phi$  as the angle where  $R = r_E(\phi)$ :

$$R^2 = \frac{a^2 b^2}{a^2 \sin^2 \phi + b^2 \cos^2 \phi} = \frac{a^2 b^2 (1 + \tan^2 \phi)}{a^2 \tan^2 \phi + b^2}$$

substituting in our value for  $\tan^2(\phi)$  we get:

$$R^2 = \frac{ab^2(a+b)}{b(a+b)} = ab$$

and finally:

$$\boxed{R = \sqrt{ab}}$$

This is a maximum of the IoU in the region  $b < R < a$ . We began by showing that there are no maxima in the regions outside of this. Since  $\tan^2 \phi$  is monotone increasing on  $(0, \pi/2)$  we have only one solution to  $\tan^2 \phi = \frac{b}{a}$  and therefore  $R = \sqrt{ab}$  is the unique maximum. ■

## Appendix B. Properties of Output Probability and Entropy Maps

To show that our output probability map is constrained in our weak annotation case we model our annotation set as being generated by:

$$P(X \in y_f | X \in gt_f) = \rho(R), P(X \in y_f^c | X \in gt_f) = k(R)$$

Where  $X$  is some pixel,  $y_f$  is the foreground annotation set,  $gt_f$  is the ground truth foreground set.  $\rho(R)$  is the annotation recall and  $k(R)$  is the mislabelling chance, which is given by  $k(R) = \rho(R)(1 - p(R))$  for precision  $p(R)$ . These are all functions of the disk radius  $R$ . This describes our weak annotation scenario in which only a subset of the ground truth is labelled and we incur some small annotation error from the cases where disks are inaccurate. Our foreground label probability for a given pixel  $X$  is then given by:

$$P(y_f | X) = P(y_f | gt_f, X)P(gt_f | X) + P(y_f^c | gt_f, X)P(gt_f^c | X)$$

We assume that the labelling noise is conditionally independent of the image, i.e.:

$$P(y_f | gt_f, X) = P(y_f | gt_f)$$

Our expression then reduces to:

$$P(y_f | X) = \rho(R)P(gt_f | X) + k(R)(1 - P(gt_f | X))$$

and further to:

$$P(y_f|X) = k(R) + (\rho(R) - k(R))P(gt_f|X) \quad (\text{B.1})$$

$$= \rho(R) - \rho(R)p(R)(1 - P(gt|X)) \quad (\text{B.2})$$

From the definition of  $k(R)$  we can see that it takes values of at most  $\rho(R)$ , so the first line gives a lower bound of  $k(R)$  on  $P(y_f|X)$ . The second line similarly gives an upper bound of  $\rho(R)$  as the final term is always  $\geq 0$ . Hence we can say:

$$k(R) \leq P(y_f|X) \leq \rho(R)$$

This demonstrates that in the weak labelling scenario the output probabilities are fundamentally constrained to  $(k(R), \rho(R))$ . Given that our Bayesian network is an approximator of the foreground probability distribution, we can describe it as learning the function:

$$f_\theta(X) = P(y_f|X) + \epsilon_\theta(X)$$

with some error term  $\epsilon_{\theta(X)}$  and network parameters  $\theta$ . Making the assumption that we have chosen a good approximator we consider that our error is less than some constant  $\delta$

$$|\epsilon_\theta(X)| \leq \delta$$

Our bound then becomes:

$$k(R) - \delta \leq P(y_f|X) \leq \rho(R) + \delta$$

We make the assumption that our network is a good approximator and that  $\delta$  is small. Empirically we find that the disk annotations have recall  $\rho(R) \in (0.05, 0.5)$  and error  $k(R) \in (0.01, 0.10)$ , so the range of our output probabilities is substantially constrained.

To show that the entropy is a monotonic transformation of  $P(y_f|X)$  if  $P(y_f|X) \leq 0.5$  we use the definition of the entropy:

$$H(P) = -P(\log P) - (1 - P)\log(1 - P)$$

Where  $P$  is  $P(y_f|X)$ . This expression is increasing for  $P < 0.5$ , so as long as our recall  $\rho(R)$  is less than 0.5 and our error  $\delta$  is small, the entropy map will behave as a monotone transformation of the probability which preserves the ordering of the pixels.

Table C.1: Table of optimal per-image threshold Dice scores for CPM17, PanNuke and Lizard datasets for both Bayes annotation approaches for optimal choice of  $r$ .

| <b>Dataset</b> | <b>Method</b> | <b>lp</b>   |             |             |             |             |
|----------------|---------------|-------------|-------------|-------------|-------------|-------------|
|                |               | <b>1.00</b> | <b>0.75</b> | <b>0.50</b> | <b>0.25</b> | <b>0.10</b> |
| CPM17          | Bayes (C)     | 0.80        | 0.78        | 0.79        | 0.79        | 0.74        |
|                | Bayes (R)     | 0.81        | 0.80        | 0.79        | 0.76        | 0.70        |
| PanNuke        | Bayes (C)     | 0.67        | 0.67        | 0.68        | 0.67        | 0.63        |
|                | Bayes (R)     | 0.71        | 0.70        | 0.68        | 0.66        | 0.60        |
| Lizard         | Bayes (C)     | 0.64        | 0.63        | 0.64        | 0.62        | 0.54        |
|                | Bayes (R)     | 0.65        | 0.64        | 0.62        | 0.59        | 0.52        |

Table C.2: Table of pixel-wise Average Precision for CPM17, PanNuke and Lizard datasets for both Bayes annotation approaches for optimal choice of  $r$ .

| <b>Dataset</b> | <b>Method</b> | <b>lp</b>   |             |             |             |             |
|----------------|---------------|-------------|-------------|-------------|-------------|-------------|
|                |               | <b>1.00</b> | <b>0.75</b> | <b>0.50</b> | <b>0.25</b> | <b>0.10</b> |
| CPM17          | Bayes (C)     | 0.85        | 0.86        | 0.87        | 0.87        | 0.79        |
|                | Bayes (R)     | 0.88        | 0.87        | 0.86        | 0.82        | 0.74        |
| PanNuke        | Bayes (C)     | 0.71        | 0.72        | 0.72        | 0.71        | 0.65        |
|                | Bayes (R)     | 0.75        | 0.74        | 0.72        | 0.68        | 0.61        |
| Lizard         | Bayes (C)     | 0.66        | 0.66        | 0.67        | 0.63        | 0.48        |
|                | Bayes (R)     | 0.66        | 0.66        | 0.67        | 0.63        | 0.48        |

## Appendix C. Separability and Upper-bound Performance of Bayesian Approach

The optimal per-image Dice scores in Tab. C.1 illustrate the same trend in terms of dataset performance as our prior analysis at the 0.5 threshold, indicating that the difference in performance across datasets is not an artefact introduced by a poor choice of threshold value. What is notable is that the performance at lower label percentages is significantly improved for the PanNuke and CPM17 datasets, for 10% of labels there is a performance improvement of around 50% when thresholding optimally compared to thresholding at 0.5. Although the performance still degrades significantly from 100% labels down to 10% we see a dramatically smaller loss in performance for both 10% and 25% labels. This indicates that good performance is still achievable in these very low labelling regimes but that a new threshold choice needs to be made, this supports our analysis of the behavior of the output probability and entropy maps in Sec. Appendix B which predicts exactly this kind of outcome.

Notably the performance on the Lizard dataset is not dramatically improved under optimal thresholding, indicating that the poor performance on this dataset is due to something more fundamental. This supports our hypothesis that the network performs poorly on the smaller instances, since morphologically the nuclei are similar and in some cases identical to those in the PanNuke and CPM17 datasets. More experimentation would be required to test this hypothesis.

The Average Precision scores in Tab. C.2 tell a very similar story to the optimal Dice scores, with the CPM17 and PanNuke datasets achieving good scores and these being largely maintained even down to 10% annotations. These scores indicate that the pixels in the output entropy maps are well-ordered, again supporting our analysis in Sec. Appendix B that the entropy map has approximately the same ordering as the probability map. Once again the AP scores for the Lizard dataset are the worst of the three datasets, reinforcing that the Bayesian method is much less suited to this dataset.

## Appendix D. Consideration of Foundation Model approaches

For completeness we include in Table Tab. D.3 an evaluation of the masks produced by a selection of promptable foundation models when conditioned on the same set of point annotations on the three datasets. These models

have risen to be a dominant paradigm in computer vision, providing highly effective pretrained backbones for a variety of downstream tasks and imaging domains. The foundation models chosen were selected for their varying levels of specialisation in segmentation of H&E stained histopathology images.

Table D.3: Table comparing Dice score of outputs of point-prompted foundation models under different proportions of annotations. (\*) Patho-SAM contains all three of these datasets within its training data so this data contamination must be taken into consideration.

| Dataset | Method     | lp          |             |             |             |             |
|---------|------------|-------------|-------------|-------------|-------------|-------------|
|         |            | 1.00        | 0.75        | 0.50        | 0.25        | 0.10        |
| CPM17   | Patho-SAM* | <b>0.82</b> | <b>0.68</b> | <b>0.51</b> | 0.29        | 0.11        |
|         | Micro-SAM  | 0.74        | 0.66        | 0.56        | 0.36        | 0.19        |
|         | MedSAM2    | 0.50        | 0.51        | 0.49        | <b>0.38</b> | <b>0.19</b> |
| PanNuke | Patho-SAM* | <b>0.82</b> | <b>0.68</b> | <b>0.51</b> | 0.29        | 0.13        |
|         | Micro-SAM  | 0.57        | 0.54        | 0.49        | <b>0.40</b> | <b>0.25</b> |
|         | MedSAM2    | 0.46        | 0.46        | 0.45        | 0.38        | 0.24        |
| Lizard  | Patho-SAM* | <b>0.85</b> | <b>0.73</b> | <b>0.57</b> | 0.35        | 0.16        |
|         | Micro-SAM  | 0.54        | 0.50        | 0.45        | <b>0.40</b> | <b>0.41</b> |
|         | MedSAM2    | 0.20        | 0.25        | 0.29        | 0.3         | 0.3         |

MedSAM2 is a well established model in the space of foundation models but pathology imaging was not a significant proportion of its training set, Micro-SAM is a highly effective microscopy foundation model trained on a wide range of microscopy datasets with varying modalities and annotations but not trained specifically on any histopathology imaging such as those in the CPM17, PanNuke and Lizard datasets. Finally Patho-SAM is a highly specific histopathology foundation model trained on multiple histopathology datasets and designed to perform well when given point prompts.

Unsurprisingly, the foundation models perform increasingly well the more specialised their training data is to the target pathology images. It is important to note that such models have been trained on exactly this data, so a comparison of their results to ours does not make sense. Patho-SAM performs similarly well to a fully supervised UNet++ shown in Table 3. Across all models the DSC of the prior diminishes as we remove annotations as we would expect as the point prompts are required for the models to output a segmentation in that area.

While these models are attractive as methodological starting points given

their remarkable out-of-the-box performance, they cannot be relied upon in general as there are cases where appropriate models are not guaranteed to exist (such as for novel or rare imaging domains or for tasks which target non-standard objects). Moreover, many of these models are intended as annotation aids [1] or starting points for fine-tuning rather than generic medical image segmenters, which can make their direct application to weakly supervised problems awkward since not all models will have been trained to perform well under the specific weakly supervised circumstance being considered.

Although the metrics indicate that promptable foundation models could serve as high-quality prior generators under favourable circumstances, these methods are conceptually orthogonal to weakly supervised learning frameworks, which are designed to facilitate automatic prediction based on sparse annotations alone and not as interactive annotation aids. We contrast this with our Bayesian approach, which is not reliant on the existence of a domain-appropriate backbone, operates directly on post-hoc point annotations, and is also able to perform robustly even under reduced quantities of annotation making it more directly suited to the weakly-supervised setting.

## References

- [1] J. Ma, Z. Yang, S. Kim, B. Chen, M. Baharoon, A. Fallahpour, R. Asakereh, H. Lyu, B. Wang, Medsam2: Segment anything in 3d medical images and videos (2025). arXiv:2504.03600.  
URL <https://arxiv.org/abs/2504.03600>