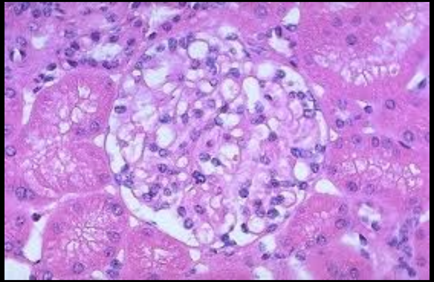
Glomeruli Segmentation in Whole-Slide Images: Is Better Local Performance Always Better?

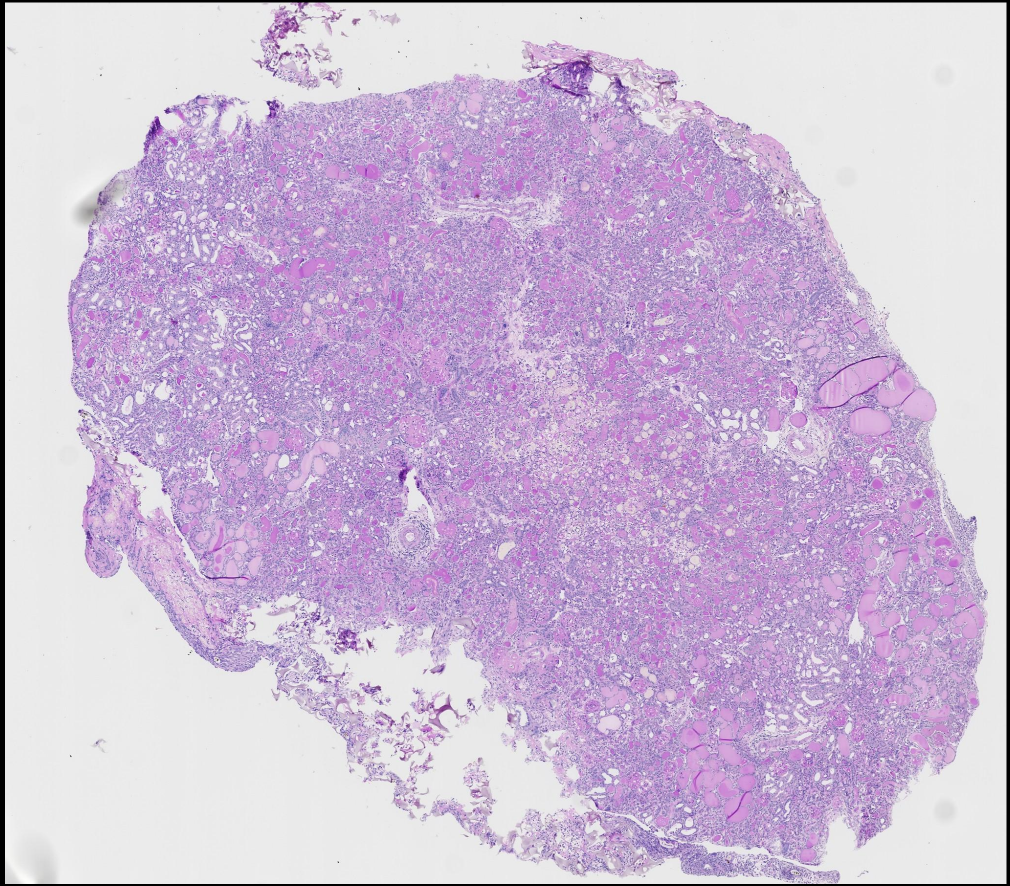
Adrian Galdran, MSC Research Fellow
CVC Barcelona, Spain | Universitat Pompeu Fabra, Barcelona, Spain

With: M. Sánchez , H. Sánchez , C. Pérez de Arenaza ,
D. Ribalta,N. Arrarte , O. Cámara

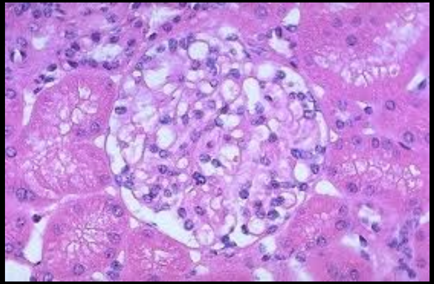
Contents

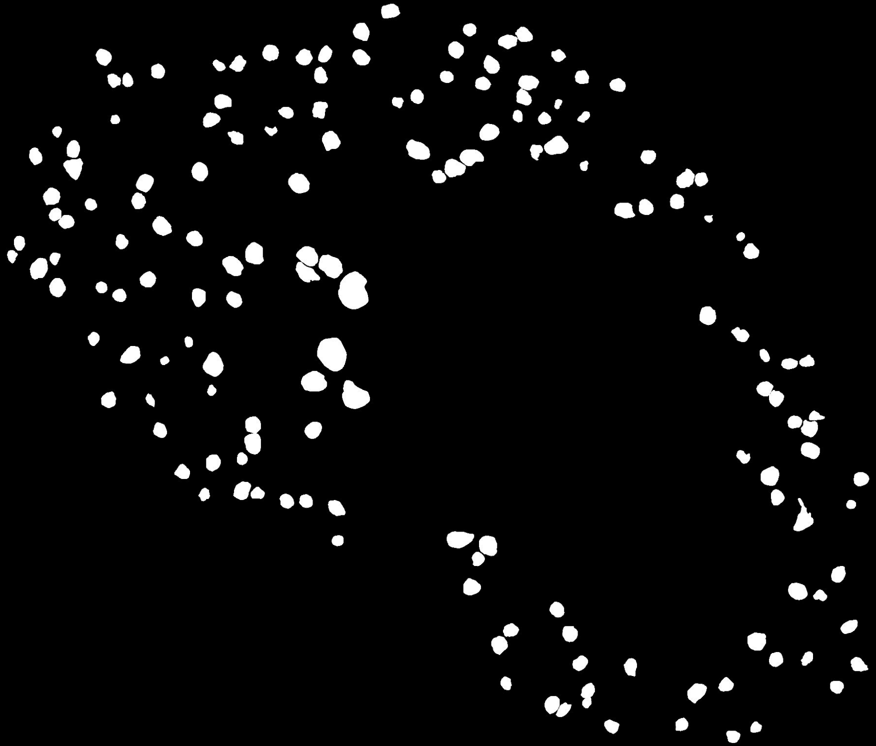
1. Introduction: The KPIs Challenge



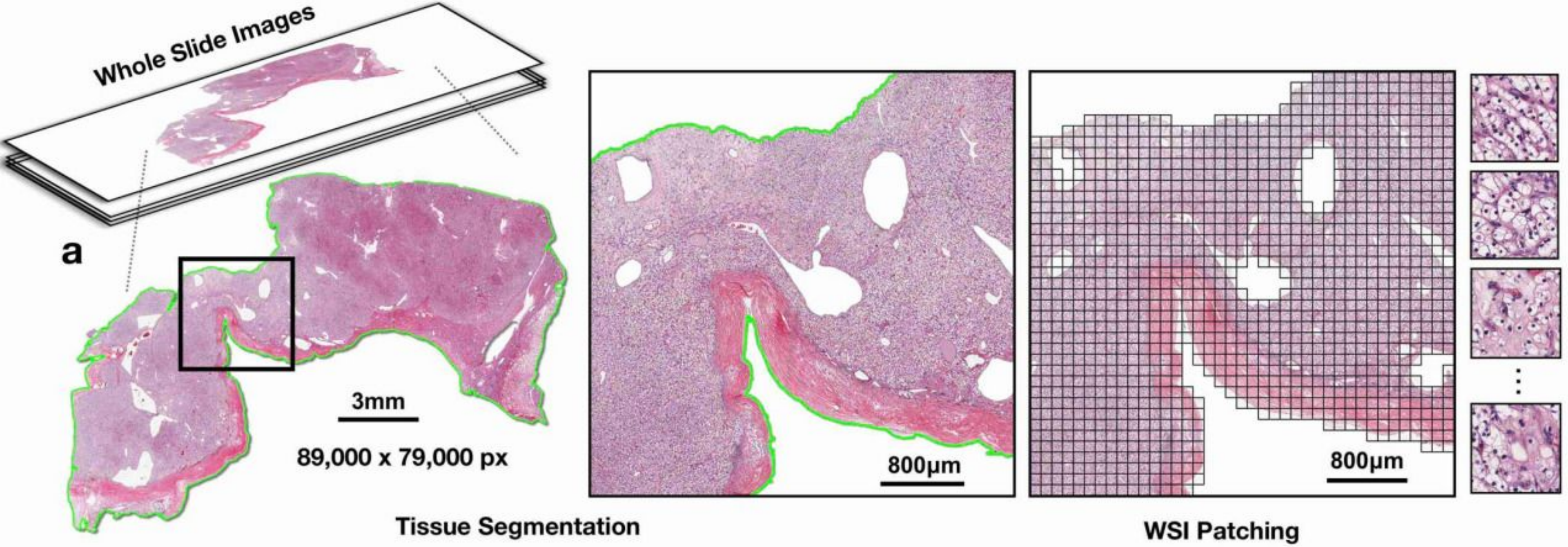


1. Introduction: The KPIs Challenge





1. Introduction: The KPIs Challenge



Credit: https://pixelscientia.com/articles/from-patches-to-slides/

2. Segmentation Models, Local to Global

Model 1 = FPN/MiT-B2 vs Model 2 = UNet++ / ResNeSt

- **Optimizer:** We used the Nadam optimizer [3], which combines the benefits of the Nesterov accelerated gradient and Adam optimizers, offering faster convergence and improved stability during training.
- **Learning Rate:** The learning rate was set to 1×10^{-4}, a value chosen after initial hyperparameter tuning to balance the speed of convergence with model performance.
- **Batch Size:** A batch size of 8 was used to manage the memory requirements while ensuring efficient gradient computation.
- **Image size:** The input images were resized to 1024×1024 pixels for ***model 1*** and 512×512 pixels for ***model 2***.
- **Number of Epochs:** Both models were trained for 60 epochs. This number was chosen to ensure sufficient training time while monitoring the validation loss and the dice value to avoid overfitting.
- **Loss Function:** The Dice loss function was used to handle the class imbalance in the dataset and improve segmentation performance by focusing on overlap between predicted and true segmentations. This was added to the Cross-Entropy loss, as this combination is known to provide benefits when dealing with overfitting and miscalibration

Other details available on our github repository `github.com/agaldran/kpis`.

3. Experimental Analysis

Table 1: Five-fold cross-validation performance (Dice similarity score) for patch-wise segmentation, with mean and standard deviation across folds.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	$\mu \pm \sigma$
Model 1	**92.37**	**90.54**	**91.47**	**89.65**	**92.50**	**91.31 ± 1.09**
Model 2	**91.96**	**89.75**	**89.94**	**87.95**	**91.87**	**90.29 ± 1.67**

Table 2: Test Set Performance on the hidden test set for task 1 (patch segmentation, dice score) and task 2 (WSI segmentation, F1 score for instance-segmentation. Final competition rank in parenthesis.

	Task 1	Task 2 - F1	Task 2 - Dice
Model 1	**94.28 (2nd/24)**	**35.01 (14th/15)**	**81.94 (11th/15)**
Model 2	**93.23 (8th/24)**	**86.81 (6th/15)**	**92.74 (5th/15)**

4. Conclusions

Do not prefer by default higher performance at patch level

Always use supplementary instance-wise metrics

Dramatic differences might be worth further research

The END

Thanks!