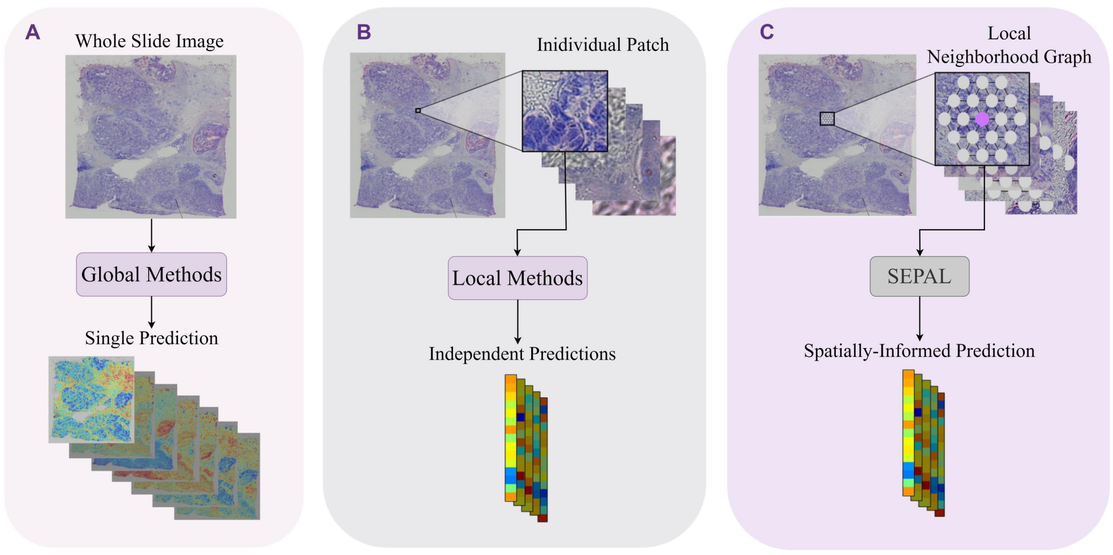
SEPAL: Spatial Gene Expression Prediction from Local Graphs

Gabriel Mejia, Paula Cárdenas*, Daniela Ruiz*, Angela Castillo, Pablo Arbeláez
Center For Research and Formation in Artificial Intelligence,
Universidad de los Andes, Bogotá, Colombia

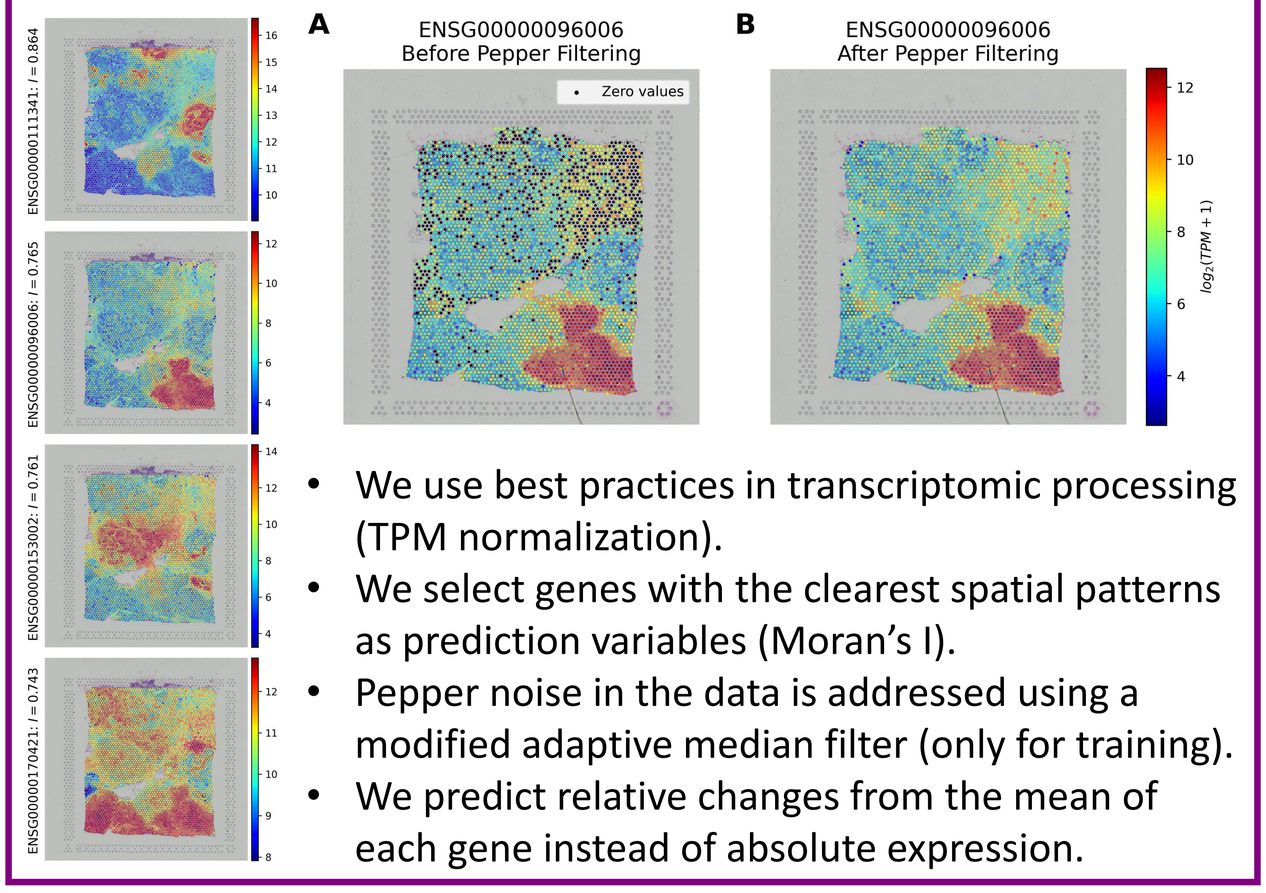
ICCV23 PARIS

1. Introduction



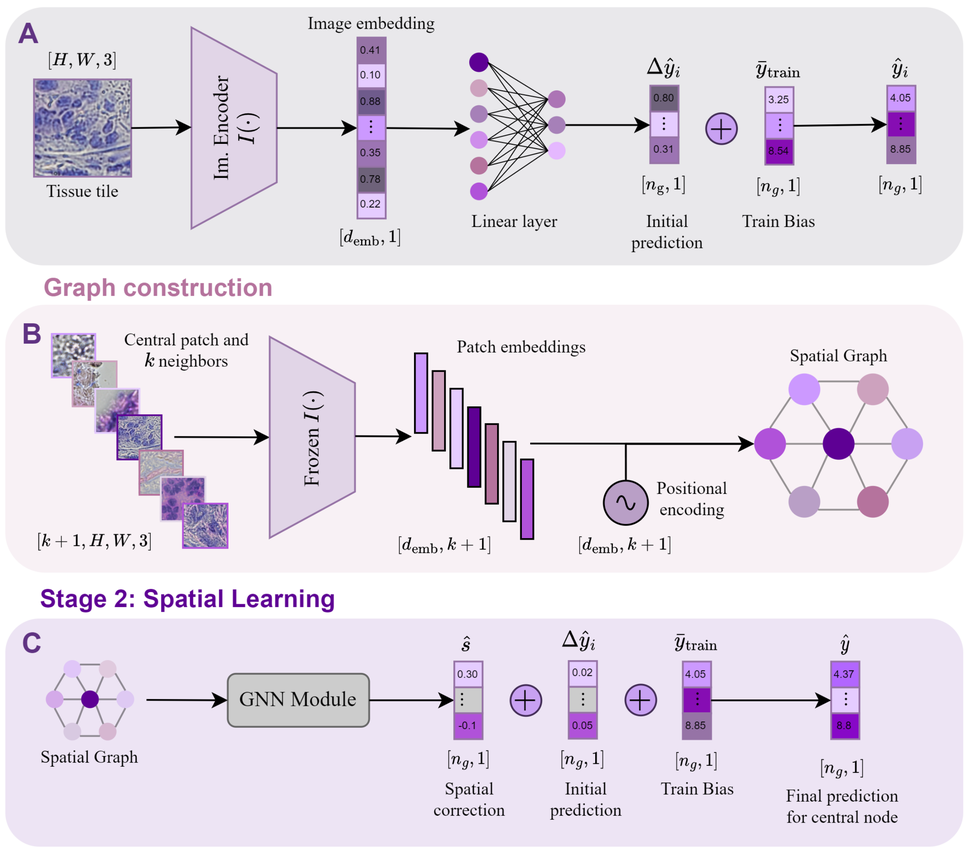
- We use spatial transcriptomics data to predict gene expression vectors from histology images.
- Global methods overfit due to lack of data and local methods cannot include visual context.
- We leverage upon both global and local analysis using spatial neighbors to inform the prediction.

2. Dataset Construction



- We use best practices in transcriptomic processing (TPM normalization).
- We select genes with the clearest spatial patterns as prediction variables (Moran's I).
- Pepper noise in the data is addressed using a modified adaptive median filter (only for training).
- We predict relative changes from the mean of each gene instead of absolute expression.

3. SEPAL



- The local learning stage trains an image encoder to predict gene expression.
- Graphs are constructed with neighbors' features.
- In the spatial learning stage, a graph neural network module integrates information from of neighboring patches and predicts a spatial correction.

4. Main Results

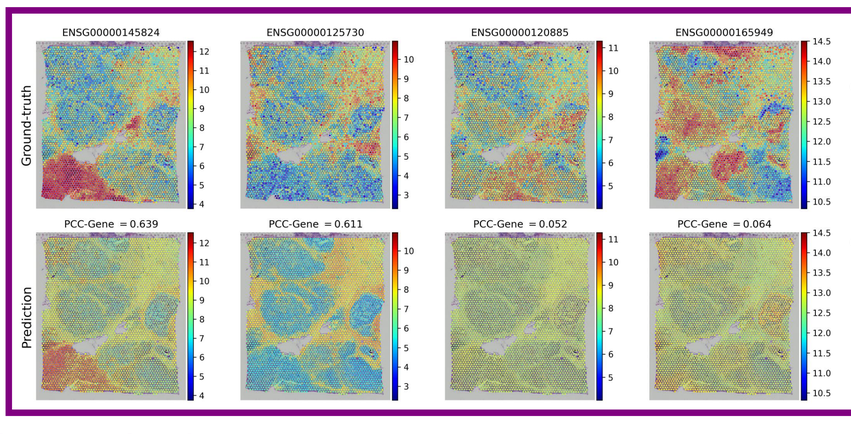
	Method	Local: STNet [1]	Local: EGN [2]	Local: EGGN [3]	Global: HisToGene [4]	Hybrid: SEPAL	Hybrid: SEPAL*
Visium	MAE	0.654	0.659	0.645	0.665	**0.630**	0.636
	MSE	0.762	0.772	0.736	0.784	**0.708**	0.717
	PCC-Gene	0.300	0.314	0.313	0.199	**0.383**	0.353
	R2-Gene	0.053	0.038	0.070	0.024	**0.106**	0.091
	PCC-Patch	0.924	0.922	0.926	0.921	**0.928**	0.927
	R2-Patch	0.843	0.841	0.846	0.839	**0.853**	0.851
STNet dataset	MAE	0.560	0.520	0.550	0.529	**0.519**	0.527
	MSE	0.537	0.480	0.549	0.493	**0.478**	0.489
	PCC-Gene	0.030	**0.064**	0.011	-0.007	-0.004	0.002
	R2-Gene	-0.165	-0.037	-0.228	-0.066	**-0.028**	-0.052
	PCC-Patch	0.910	**0.911**	0.908	**0.911**	**0.911**	**0.911**
	R2-Patch	0.779	0.806	0.780	0.799	**0.809**	0.802

- SEPAL obtains state-of-the-art results in two breast cancer datasets.
- Our method outperforms both local and global models.

Method	ViT	ViT+Δ	ViT+Δ+S7	SEPAL
MAE	0.655	0.638	0.648	**0.630**
MSE	0.760	0.725	0.737	**0.708**
PCC-Gene	0.282	0.347	0.339	**0.383**
R2-Gene	0.053	0.086	0.065	**0.106**
PCC-Patch	0.924	0.927	0.925	**0.928**
R2-Patch	0.843	0.849	0.847	**0.853**

- Just predicting relative differences already gives state-of-the-art performance.
- Adding visual context by increasing patch size is not as effective as our graph processing approach.

5.Qualitative Results



- Easiest genes show significant correlation but oversmoothed output.
- The most difficult genes are predicted to be approximately constant.

6. Conclusions

- Prediction of deltas instead of absolute values of expression improves performance. This result is probably derived from the fact that the model does not have to learn priors and can focus on the physiological variation in the data.
- Using graphs to include spatial context improves over local prediction and outperforms increasing patch scale.
- SEPAL bridges the gap between local and global methods leveraging small sample sizes while also being able to include spatial context.

7. Acknowledgements

Gabriel Mejia acknowledges the support of a UniAndes-DeepMind Scholarship 2022. This work was supported by Azure sponsorship credits granted by Microsoft's AI for Good Research Lab. We thank Andres Hernandez for his valuable help during the conception of the project.

8. References

[1] Bryan He, Ludvig Bergenstrahle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Ake Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4:827–834, 6 2020.

[2] Yan Yang, Md Zakir Hossain, Eric A Stone, and Shafin Rahman. Exemplar guided deep neural network for spatial transcriptomics analysis of gene expression prediction. *arXiv*, 10 2022.

[3] Yan Yang, Zakir Hossain, Eric Stone, and Shafin Rahman. Spatial transcriptomics analysis of gene expression prediction using exemplar guided graph neural network. *bioRxiv*, page 2023.03.30.534914, 3 2023.

[4] Minxing Pang, Kenong Su, and Mingyao Li. Leverag ing information in spatial transcriptomics to predict superresolution gene expression from histology images in tumors. *bioRxiv*, page 2021.11.28.470212, 11 2021.

9. Contact