xMINT: A Multimodal Integration Transformer for Xenium Gene Imputation

Xiaohui Jiang; Yuxia Xie; Jichun Xie

AccMLBio @ ICML 2024

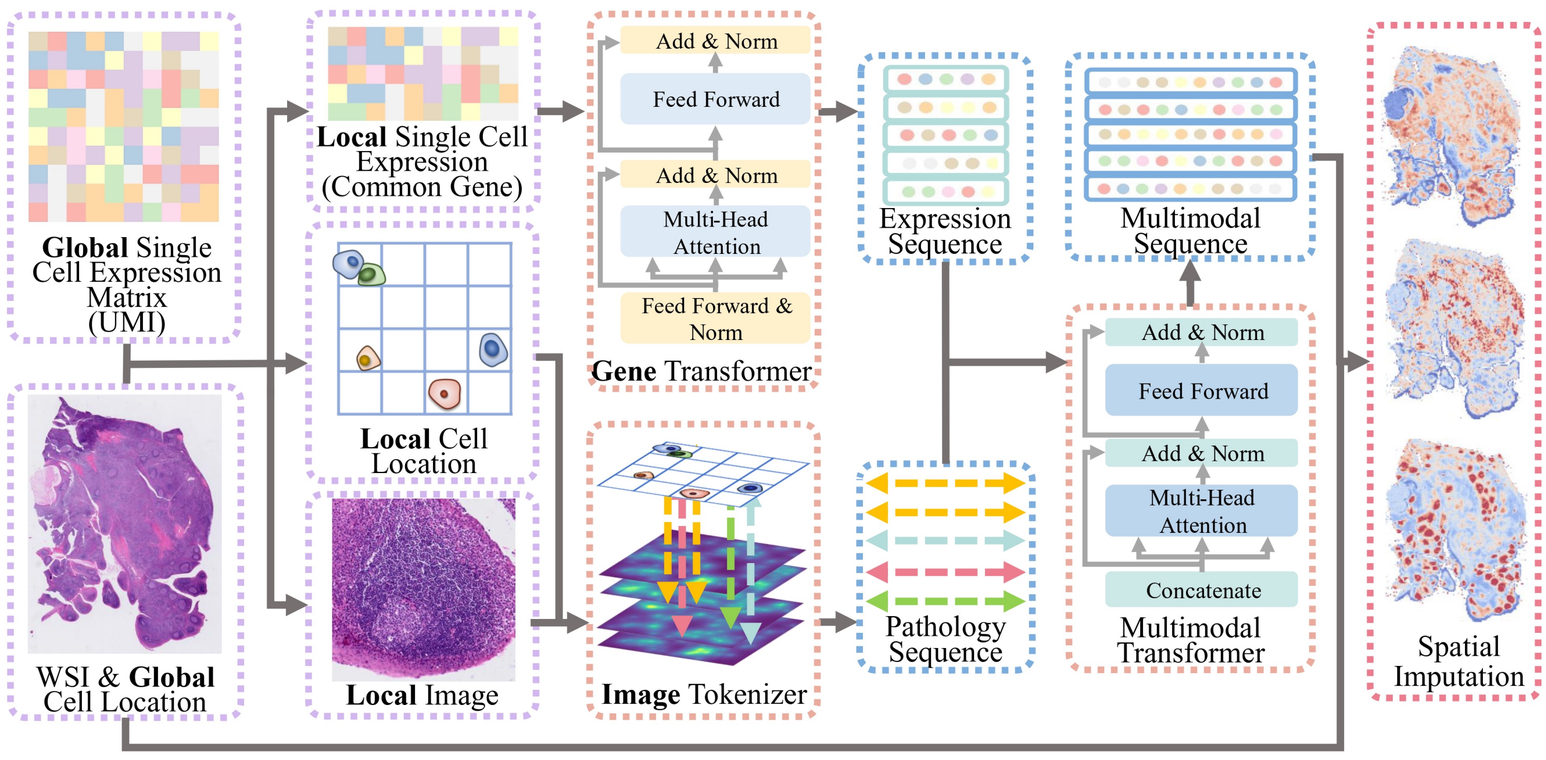
Background

- **Gene Sequencing Limitations:** Xenium technology sequences only around 500 genes, limiting comprehensive analysis.
- **Technological Discrepancies:** Integration of Xenium data with scRNA-seq is problematic due to differences in sequencing technologies.
- **Data Underutilization:** Current methods do not leverage available pathology images, which can provide critical insights into cell types and states.

Contribution

- **Multimodal Integration:** xMINT uses a Transformer-based model to integrate gene expression data with high-resolution pathology images, enhancing gene imputation accuracy.
- **Enhanced Accuracy:** Achieves better imputation accuracy than existing scRNA-seq-based methods by utilizing additional visual information from pathology images.

Method



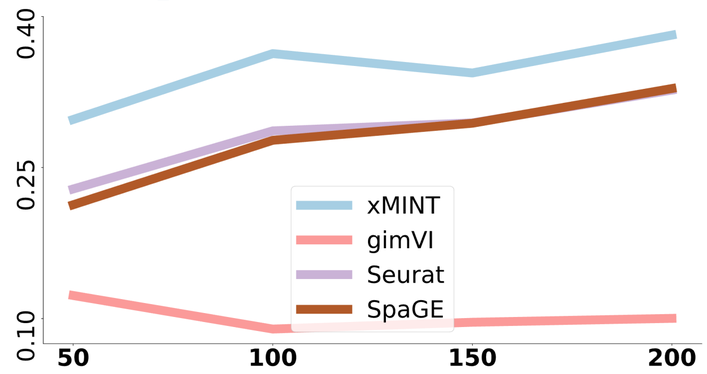
- **Initial Division**: Whole Slide Image is divided into local regions.
- **Sequence Generation**: Within each local region, sequences are generated using Morphology features & RNA sequencing values.
- **Concatenation**: The pathology sequence and expression sequence are concatenated.
- **Multimodal Transformer**: The concatenated sequences are input into a multimodal transformer.
- **Output**: The multimodal transformer outputs imputed gene sequences.

Parameter Number & Computation Time

f	Exp.	Head	Layer	Parameters	Time (s/batch)
256	4	4	4	16,799,760	1.03
256	8	8	8	53,564,432	1.09
128	4	4	4	4,338,064	0.56
128	8	8	8	13,545,360	0.60

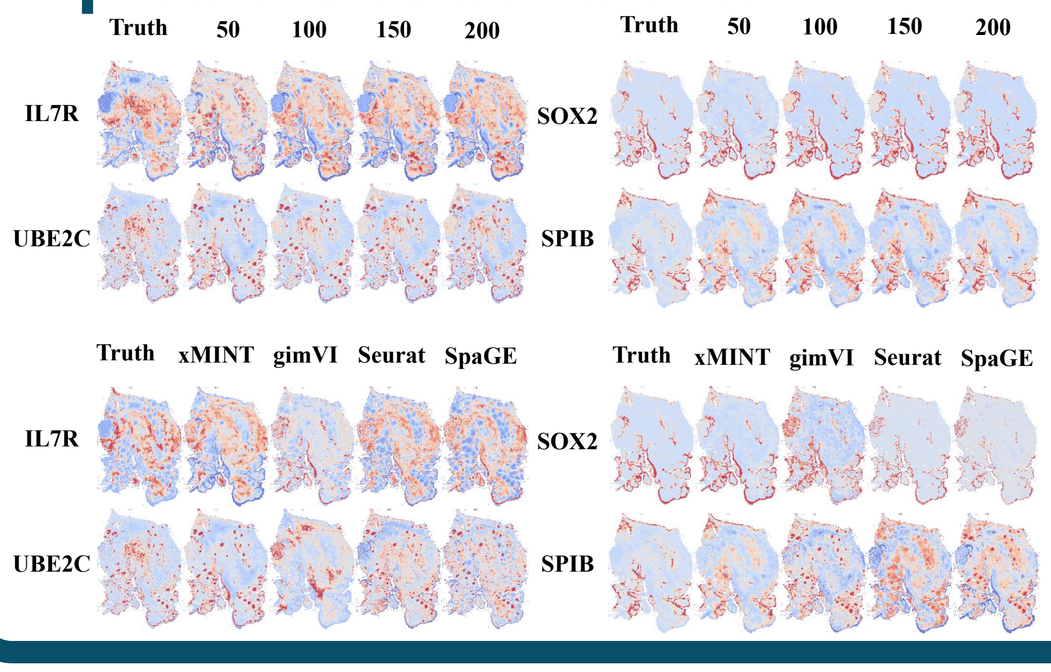
- **Exp.**: Forward expansion factor.
- **Head**: Number of attention heads.
- **Layer**: Number of transformer layers.
- **f**: Embedding dimension.

Comparison

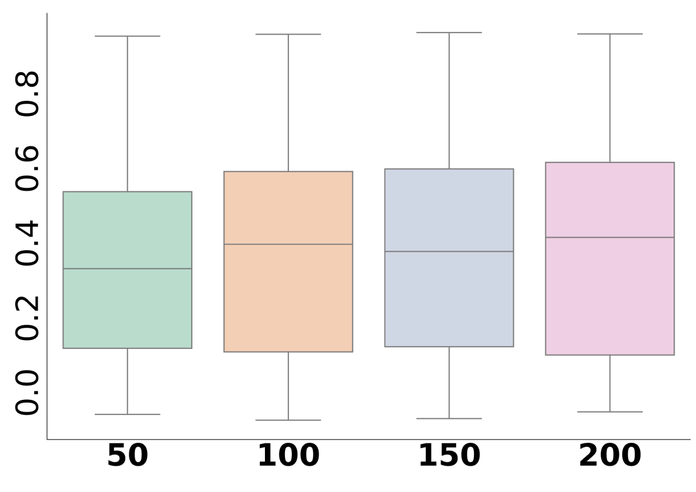


Median spearman correlation, with 50, 100, 150, 200 shared genes.

Spatial Distribution



Robustness



Spearman correlation distribution, with 50, 100, 150, 200 shared genes.

Github

OpenReview