Rebuttal Supplementary Material for Reviewer tnYZ of Submission 272: STOFM: a Multi-scale Foundation Model for Spatial Transcriptomics

Table Rebuttal-Reviewer-tnYZ.1: Experimental details. The number of categories for multi-class classification in Sections 4.2 and 4.3 of the paper are shown. We filtered out categories with fewer cells than 1% of the total number of cells.

Task	Data	Number of categories	Number of categories after filtering
Tissue region semantic segmentation (Sec. 4.2)	Embryo	18	16
Tissue region semantic segmentation (Sec. 4.2)	DLPFC	6	6
Cell type annotation (Sec. 4.3)	Brain1	25	21
Cell type annotation (Sec. 4.3)	Brain2	8	6

Table Rebuttal-Reviewer-tnYZ.2: Ablation study of data volume. We use 12.5% and 50% of the SToCorpus-88M for multi-scale ST representation learning pre-training, and compare them with model pre-trained on the full dataset. The results show that a larger pre-training dataset can improve model performance. We attribute this to the **increased diversity** of the data. We will add more experiments and discussions on the model scaling law in the subsequent revised version of the paper.

Data volume	Embryo2 F1	EmbryoCross F1
12.5%	0.758	0.423
50%	0.782	0.450
100%	0.801	0.459

Table Rebuttal-Reviewer-tnYZ.3: Ablation study of $\mathcal{L}_{\text{PDR}}$ and spatial distance matrix. The results demonstrate that removing the spatial distance matrix significantly decreases model performance.

Model	Embryo2 F1	EmbryoCross F1
w/o $\mathcal{L}_{\text{PDR}}$	0.749	0.437
w/o spatial distance matrix	0.721	0.413
STOFM	0.801	0.459

Table Rebuttal-Reviewer-tnYZ.4: Cell type annotation on scRNA-seq data. We follow the settings of LangCell [4] and select the PBMC10k dataset containing 8 cell types for the cell type annotation task. We compare the cell encoder of STOFM with Geneformer [5] used for initializing the cell encoder and scBERT [6] as a baseline. The results show that compared to the Geneformer used for initializing the cell encoder, the performance of the trained cell encoder on scRNA-seq tasks does not decrease significantly, proving that catastrophic forgetting is not severe.

Model	Accuracy	F1
scBERT	0.975	0.905
Geneformer	0.978	0.957
STOFM-CellEncoder	0.980	0.944

Refs:

[1] Uni-Mol: a universal 3D molecular representation learning framework

[2] scGPT: toward building a foundation model for single-cell multi-omics using generative AI

[3] Should You Mask 15% in Masked Language Modeling?

[4] LangCell: language-cell pre-training for cell identity understanding

[5] Transfer learning enables predictions in network biology

[6] scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data