Main Track | Machine learning for sciences

Learning Crossmodal Interaction Patterns via Attributed Bipartite Graphs for Single-Cell Omics

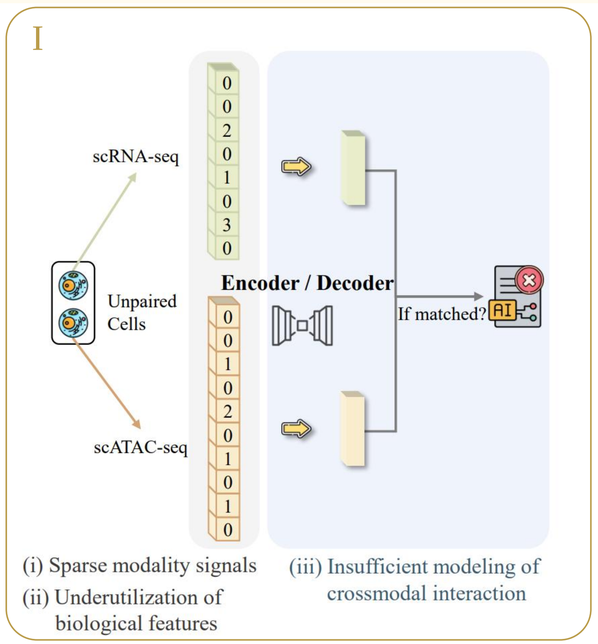
Xiaotang Wang[1], Xuanwei Lin[2], Yun Zhu[3], Hao Li[4], Yongqi Zhang[1,†]

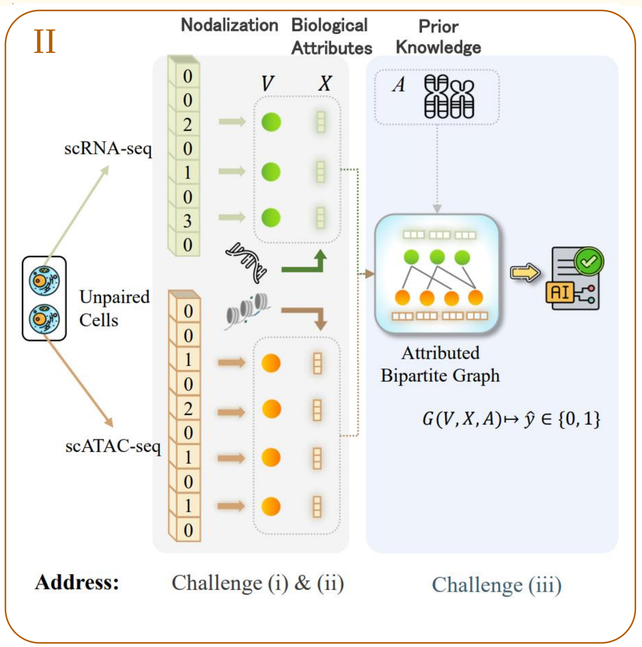
香港科技大学(广州) THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

FUZHOU UNIVERSITY 福州大学 1958

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

NEURAL INFORMATION PROCESSING SYSTEMS





I. Motivation: Challenges in Existing Frameworks

(i) Sparse modality signals.
The expression vectors of RNA and ATAC are sparse; thus, modeling them with dense vectors will introduce noise from unexpressed signals.

(ii) Underutilization of biological features.
Each RNA and ATAC signal carries rich biological attributes (e.g., statistical summaries, genomic annotations, and DNA sequences), which are not effectively incorporated by existing methods.

(iii) Insufficient modeling of crossmodal interaction.
Most current methods do not directly model the interaction between expressed ATAC and RNA signals, with limitations in understanding the underlying regulation.

II. Graph Construction: From Modality Expression to Attributed Bipartite Graphs

We introduce a graph-based perspective built upon the concept of **Attributed Bipartite Graphs (ABGs)**.

① Nodalizing Expressed Multimodal Signals into Bipartite Node Set.

$$\mathcal{V}_{\text{RNA}} = \{\texttt{RNA}_m \mid \mathbf{x}_{\text{RNA}}[m] \neq 0\}, \quad \mathcal{V}_{\text{ATAC}} = \{\texttt{ATAC}_n \mid \mathbf{x}_{\text{ATAC}}[n] \neq 0\}$$

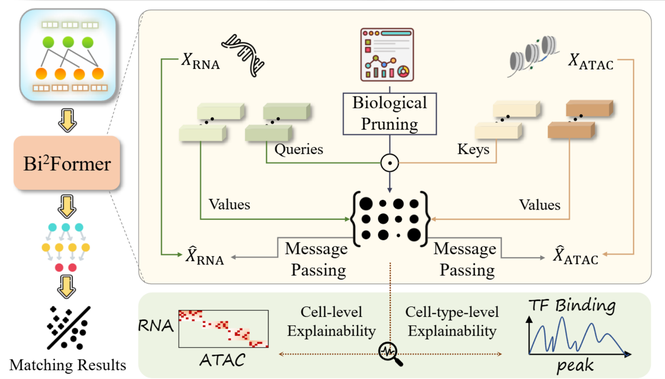
② Embedding Biological Attributes into Node Features.

$$X_v = \texttt{Concat}(\texttt{ID}(v), \texttt{Expr}(v), \texttt{BioAttr}(v))$$

③ Edge Design with Biological Prior Knowledge.

We introduce a chromosomal mask as adjacency matrix to constrain attention computation within chromosomally plausible regions. These prior-informed connections improve inductive bias and reduce noise from fully associations.

III. Bi^{2}Former: An ABG-based Interaction Learner (Biologically-driven Bipartite Graph Transformer)



Biologically-driven Crossmodal Attention.

$$Q = X_{\text{RNA}} W_Q, \quad K = X_{\text{ATAC}} W_K,$$
$$V_{\text{RNA}} = X_{\text{RNA}} W_{V_r}, \quad V_{\text{ATAC}} = X_{\text{ATAC}} W_{V_a},$$

Biological Pruning.

$$\tilde{\alpha} = \texttt{Threshold}(\sigma(\alpha), \tau) \in \mathbb{R}^{|\mathcal{V}_{\text{RNA}}| \times |\mathcal{V}_{\text{ATAC}}|}.$$

With a top-k pruning, aligns with the biological truth that each gene is typically regulated by a limited number of ATAC.

Message Passing.

$$X^{\text{cross}}_{\text{RNA}}[r] = \sum\nolimits_{a \in \mathcal{V}_{\text{ATAC}}} \tilde{\alpha}_{r,a} \cdot V_{\text{ATAC}}[a]$$
$$X^{\text{cross}}_{\text{ATAC}}[a] = \sum\nolimits_{r \in \mathcal{V}_{\text{RNA}}} \tilde{\alpha}_{r,a} \cdot V_{\text{RNA}}[r]$$

IV. Experiments: Main Results, Few-shot, Biological Interpretation, and Ablation

Crossmodal Matching

Dataset	ISSAAC-seq		10×PBMC		SHARE-seq		SNARE-seq		10×Multiome		Avg.	
Metric	ACC	ROC-AUC	ACC	ROC-AUC	ACC	ROC-AUC	ACC	ROC-AUC	ACC	ROC-AUC	ACC	AUC
MultiVI	66.21 ± 1.46	69.32 ± 1.07	60.93 ± 2.83	63.85 ± 1.96	64.42 ± 2.19	68.87 ± 1.03	56.76 ± 1.94	61.12 ± 1.18	69.35 ± 1.21	72.64 ± 1.36	63.53	67.16
CLUE	71.28 ± 1.24	75.01 ± 0.98	68.73 ± 1.67	72.26 ± 0.94	63.21 ± 2.08	67.96 ± 1.19	59.32 ± 1.59	63.17 ± 0.93	73.72 ± 0.97	76.92 ± 1.17	67.25	71.06
Cobolt	69.21 ± 2.51	73.69 ± 1.72	61.65 ± 3.05	66.74 ± 1.87	58.67 ± 3.14	61.74 ± 1.62	57.46 ± 2.03	60.91 ± 1.39	71.16 ± 1.44	74.15 ± 1.61	63.63	67.45
GLUE	74.28 ± 0.91	77.40 ± 0.92	72.51 ± 1.02	79.68 ± 0.71	66.89 ± 1.43	73.14 ± 1.17	64.47 ± 1.22	68.28 ± 1.21	76.93 ± 0.82	80.98 ± 0.92	71.01	75.90
scMoGNN	73.72 ± 0.96	78.58 ± 0.89	72.41 ± 1.37	80.76 ± 0.83	69.84 ± 1.81	74.39 ± 0.94	69.03 ± 1.22	72.32 ± 0.97	75.49 ± 1.31	80.04 ± 1.01	72.10	77.22
scMaui	71.64 ± 0.97	76.19 ± 0.83	63.19 ± 2.74	67.42 ± 1.52	65.93 ± 1.78	69.15 ± 0.96	58.42 ± 1.65	63.14 ± 0.95	75.07 ± 0.75	78.81 ± 1.13	66.85	70.94
MLP	67.39 ± 1.18	71.04 ± 0.79	62.25 ± 3.74	55.87 ± 2.06	58.97 ± 0.74	62.52 ± 0.57	54.74 ± 1.26	59.72 ± 1.01	70.44 ± 2.07	72.62 ± 1.98	62.76	64.35
GCNII	72.64 ± 1.29	77.32 ± 0.63	73.64 ± 0.98	79.60 ± 0.54	69.49 ± 1.13	74.01 ± 0.65	62.71 ± 1.07	67.93 ± 0.59	76.28 ± 1.13	81.06 ± 0.76	70.95	75.98
GraphSAGE	76.98 ± 0.61	82.37 ± 0.35	76.92 ± 1.32	81.52 ± 0.63	67.56 ± 1.24	70.93 ± 0.72	66.53 ± 0.83	70.56 ± 0.57	81.94 ± 1.89	85.79 ± 1.44	73.99	78.23
GT	73.42 ± 0.52	80.93 ± 0.31	78.04 ± 0.79	82.04 ± 0.47	72.17 ± 0.53	78.34 ± 0.36	68.74 ± 0.69	73.89 ± 0.42	80.12 ± 0.96	85.71 ± 0.61	74.50	80.18
Bi^{2}Former	**84.40 ± 0.48**	**89.24 ± 0.31**	**88.74 ± 0.36**	**92.37 ± 0.16**	**79.84 ± 0.29**	**84.96 ± 0.18**	**73.56 ± 0.37**	**77.30 ± 0.21**	**90.41 ± 0.24**	**93.41 ± 0.30**	**83.39**	**87.46**

Few-shot Settings

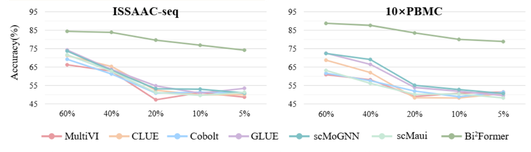


Figure 3: Results for crossmodal matching task with different training sizes.

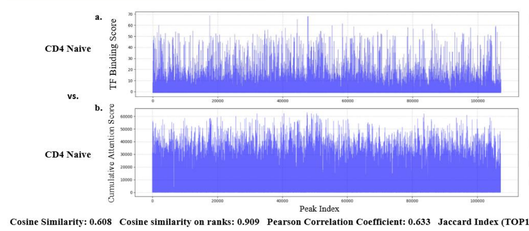
Cross-cell-type Generalization

Dataset	ISSAAC-seq	10×PBMC	SHARE-seq	SNARE-seq
MultiVI	47.32 ± 1.39	54.72 ± 1.68	52.81 ± 1.51	51.83 ± 2.15
CLUE	61.28 ± 1.16	61.41 ± 1.31	57.31 ± 1.32	56.45 ± 1.82
Cobolt	57.84 ± 2.01	58.67 ± 1.96	55.27 ± 2.47	48.63 ± 2.51
scMaui	62.43 ± 1.84	58.54 ± 1.79	54.64 ± 2.05	52.49 ± 1.74
MLP	65.72 ± 1.13	61.84 ± 1.69	58.74 ± 1.12	53.67 ± 1.33
GCNII	72.18 ± 1.29	72.52 ± 0.78	67.37 ± 0.96	61.38 ± 1.27
GraphSAGE	74.15 ± 0.93	76.31 ± 0.65	67.02 ± 1.31	64.56 ± 0.92
GT	71.92 ± 0.82	76.43 ± 0.81	69.94 ± 0.95	67.85 ± 0.81
Bi^{2}Former	**82.74 ± 0.74**	**84.96 ± 0.49**	**78.07 ± 0.61**	**71.28 ± 0.32**

Ablation Study

Methods	Avg.Δ
Bi^{2}Former	-
w/o ID	↓ 3.78
w/o Attribute	↓ 2.93
w/o BP	↓ 1.43
w/o Edge	↓ 4.28
w/o Attribute, BP, and Edge	↓ 11.52

Biological Interpretation



Left: Comparison between TF binding scores and cumulative attention scores within the same cell type (CD4).

Right: Case study of RNA–ATAC attention matrices for a representative single cell under different threshold values τ.