Week 8: Learning in Low Data Regime with Self-Supervised Learning on Graphs

Icare Sakr

École
polytechnique
fédérale
de Lausanne

Electrocortico-
graphy

Lesion

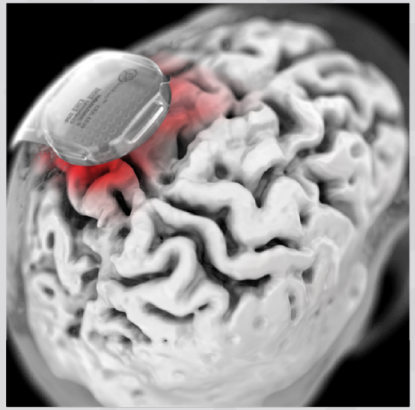
Sensing

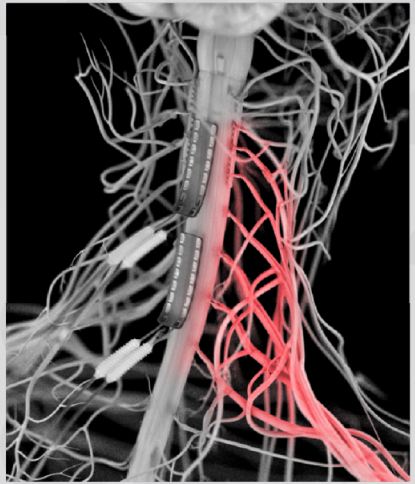
Stimulation

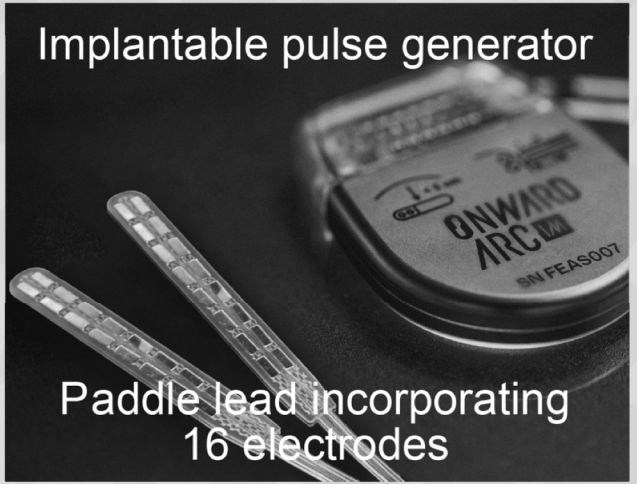
Targeted
epidural
electrical
stimulation

Selective
activation
of muscles

Cortical implant incorporating 64 channels







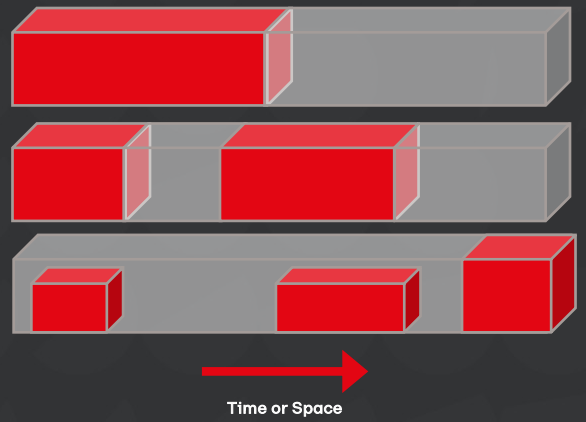
Training brain decoders

The big ____ fox ____ over the ____ dog

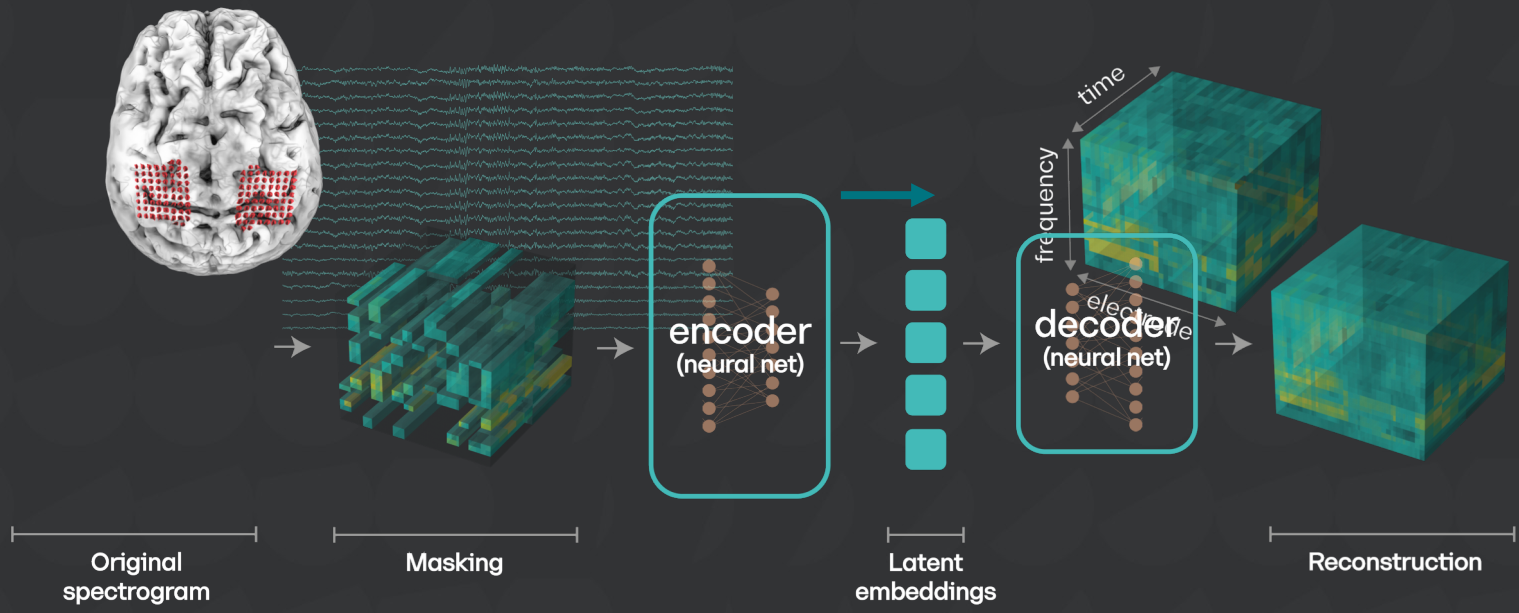
The big brown fox leaped over the lazy dog.

The big agile fox bounded over the playful dog.

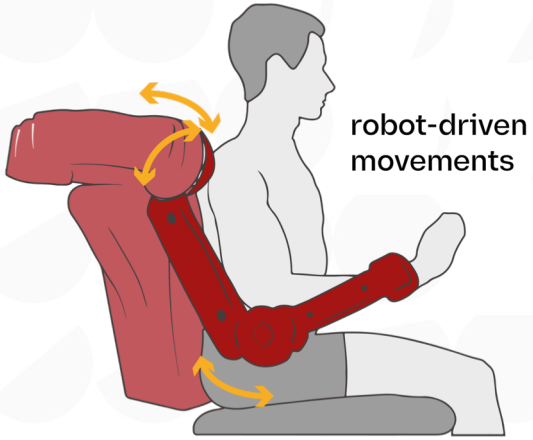
The big sly fox vaulted over the curious dog.



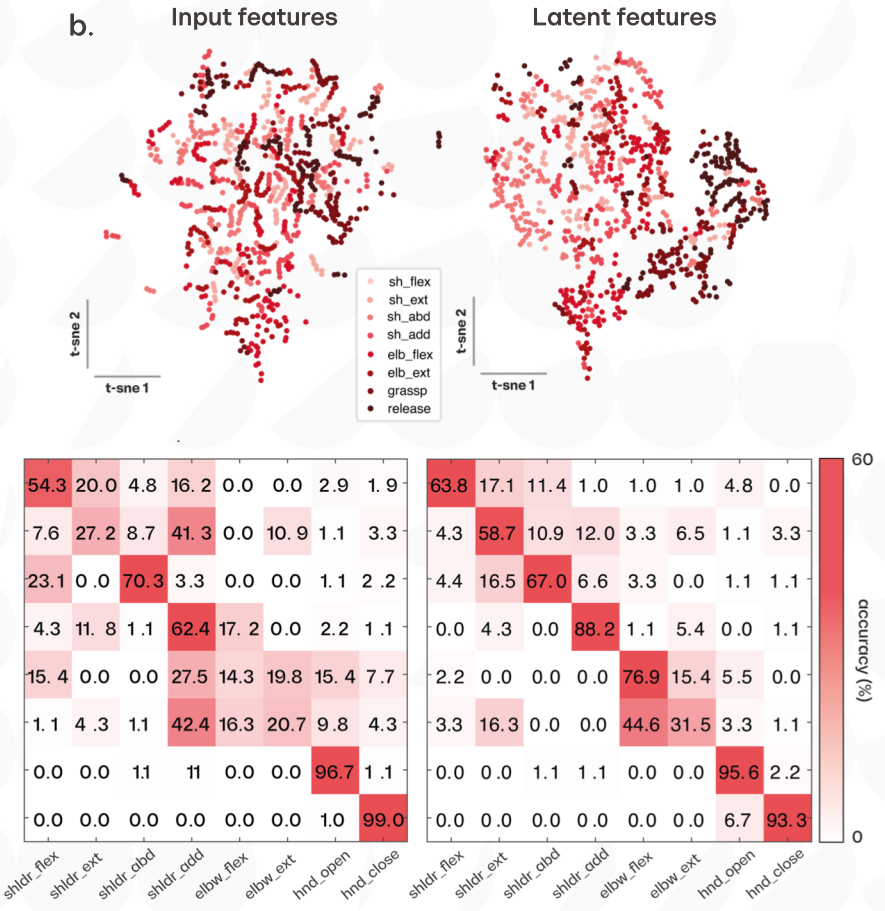
Self-supervised learning of brain signals



Self-supervised learning of brain signals



a. 8-states upper limb task



Learning in low data regime

- Despite the success of classical supervised learning methodologies they are limited by the availability of carefully annotated (labeled) datasets.
- The acquisition of these labels can be labor-intensive and time-consuming, especially in areas dealing with large-scale datasets (e.g. social networks), or that demand extensive domain knowledge (e.g. chemistry, medicine, neuroscience, …).
- Moreover, labels are often compressed (low-informative) descriptions of the data, as a result, machine learning models often lack generality and fail to cope with novel conditions, especially when training data is scarce.

Self-Supervised Learning (SSL)

- Self-supervised learning leverages readily available unlabeled data without the need of human-annotated labels.
- SSL learns informative and generic data representations that are useful across a variety of downstream tasks.
- Accumulates background knowledge or “common sense” about the data structure in a non task-specific manner
- SSL is behind the remarkable success of natural language processing models such as ChatGPT.

Self-Supervised Learning (SSL) – *pretext task*

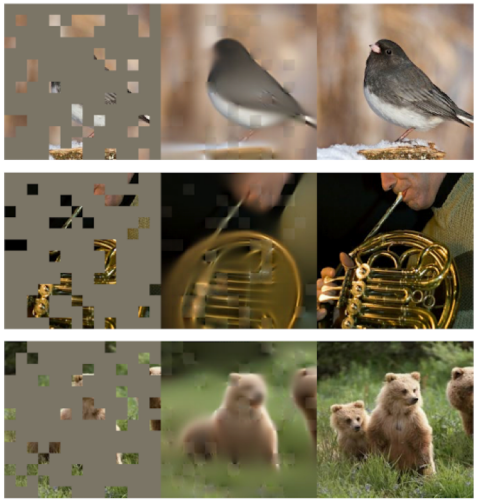
- Works by solving of a prior knowledge task (pretext task), in which the supervision signal is derived from the data itself, often leveraging the underlying structure of the data.
- Through this pretext task, the machine learning model learns meaningful data representations that are useful across a range of downstream tasks (e.g. text summarization, image classification)

The big ____ fox ____ over the ____ dog

The big brown fox leaped over the lazy dog.

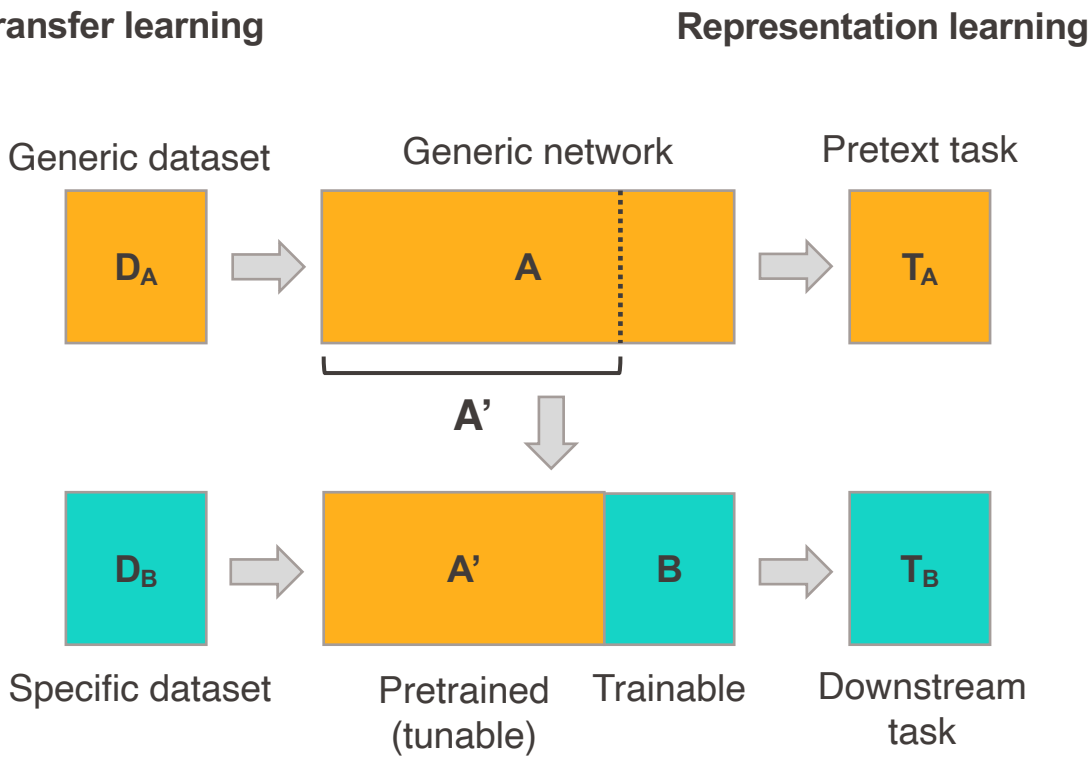
The big agile fox bounded over the playful dog.

The big sly fox vaulted over the curious dog.



Self-Supervised Learning (SSL) – *training schemes*

Icare Sakr

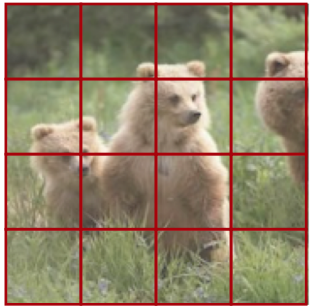


How can we apply SSL concepts to graphs

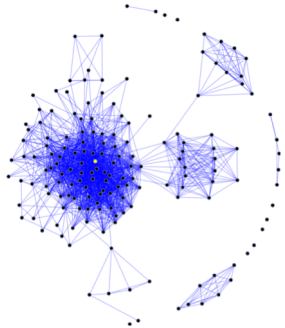
Challenges with graph structured data

- Graphs have highly irregular structures in non-Euclidean spaces (unlike images or text)
  - ➢ Need for methods that capture these irregular structures and relationships between nodes and edges
- Data samples (e.g. nodes) can be naturally linked with the topological structure of the graph

Goal: leverage the graph's inherent structure to learn node-level or graph-level representations, that are useful for downstream tasks (e.g. clustering, anomaly detection, classification,…)



regular-grid



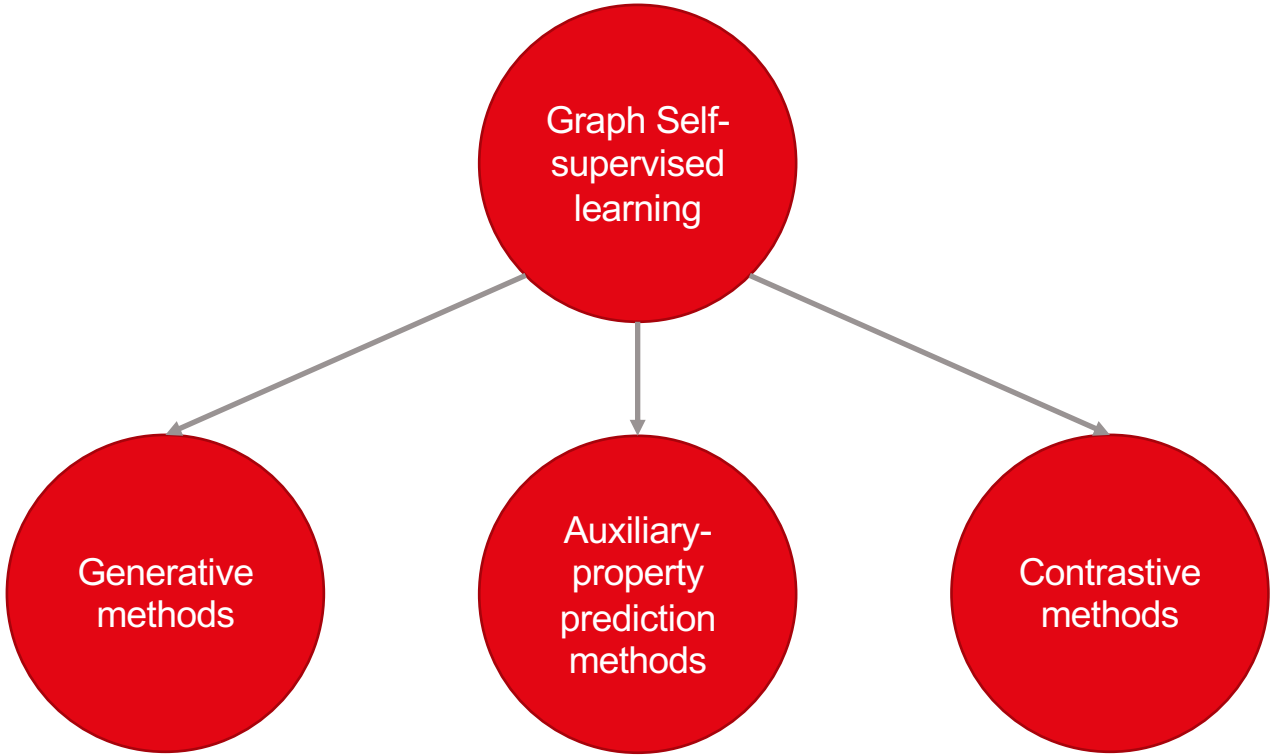
Social graph

Graph Self-Supervised Learning: A Survey

Yixin Liu, Ming Jin, Shirui Pan, Chuan Zhou, Yu Zheng, Feng Xia,
Philip S. Yu, Life Fellow, IEEE

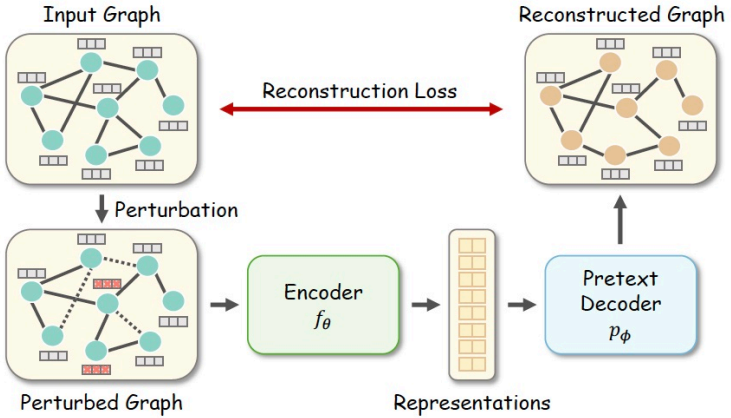
May 2022

Categorization of graph SSL methods



Generative (reconstructive) SSL objective

- Aim to reconstruct parts of the input data using the data itself or a perturbed version of it.
- Full graphs or subgraphs are used as input and a machine learning model learns by **feature generation** and/or **structure generation**.

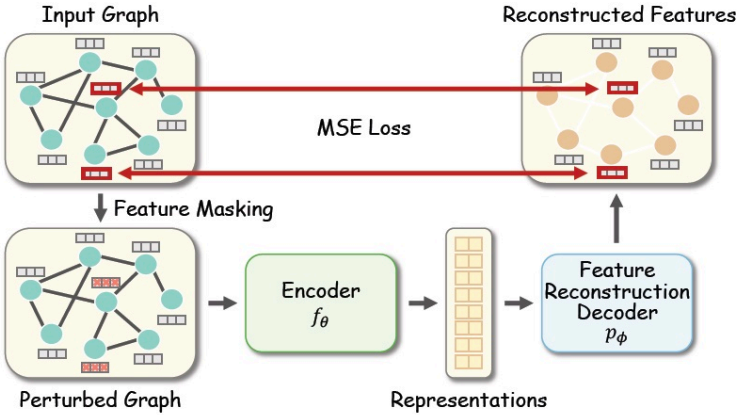


$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathcal{L}_{ssl}\Big(p_\phi\big(f_\theta(\tilde{\mathcal{G}})\big), \mathcal{G}\Big),$$

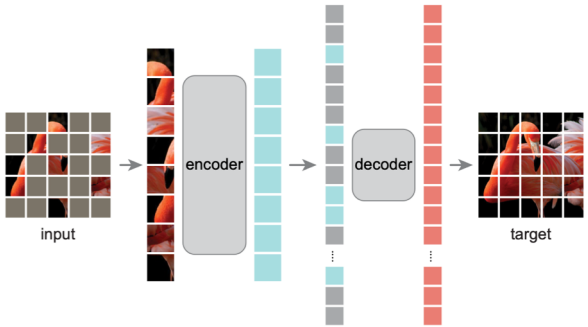
Generative (reconstructive) SSL objective

1. Feature generation

- Recover feature information from the original or perturbed graph.
- Node features, edge features, low dimensional feature matrix.
- Example: masked feature regression



Graph Completion

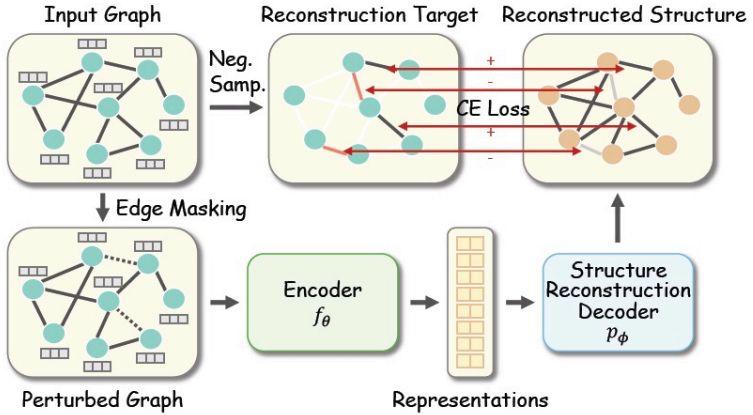


masked-autoencoders in computer vision

Generative (reconstructive) SSL objective

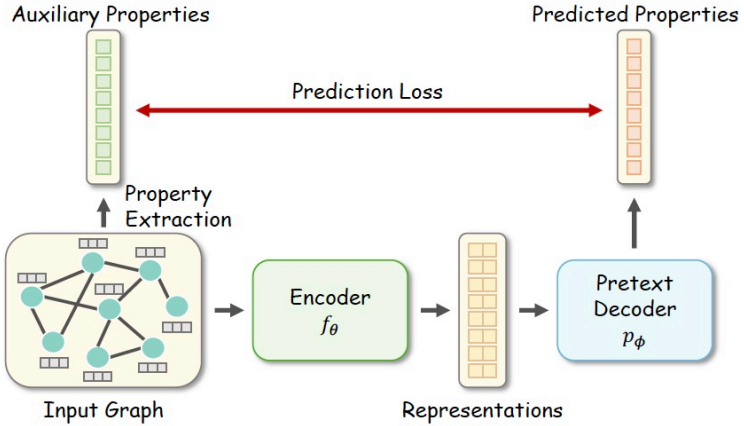
2. Structure generation

- Learn by predicting the structural (topological) information of graphs
- Typically, the objective is to reconstruct the adjacency matrix
- Capture more pair-level information since the structure generation focuses on edge generation
- Examples: Graph Autoencoder (GAE), Denoising link reconstruction



Auxiliary-properties prediction SSL objective

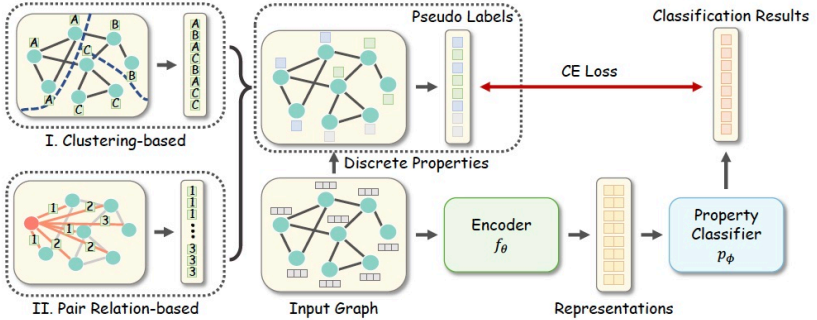
- Predict node, edge or graph level auxiliary properties (e.g. node degrees, path lengths, centrality, clustering index …)
- Enforces the model to learn representations that preserve the auxiliary variables information.
- Typically, multiple auxiliary variables can be used to enhance generality of the representations



Auxiliary-properties prediction SSL objective

1. **Auxiliary property classification**

- Predict discrete auxiliary properties
- Examples:
  - Node Clustering (predict the cluster of each node with predefined clusters)
  - Pair relation properties (e.g. closeness, centrality difference between two nodes)

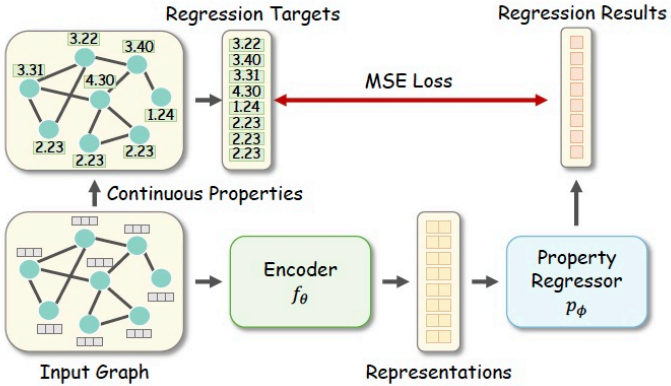


$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathcal{L}_{ce}\Big(p_\phi\big(f_\theta(\mathcal{G})\big), c\Big)$$

Auxiliary-properties prediction SSL objective

2. Auxiliary property regression

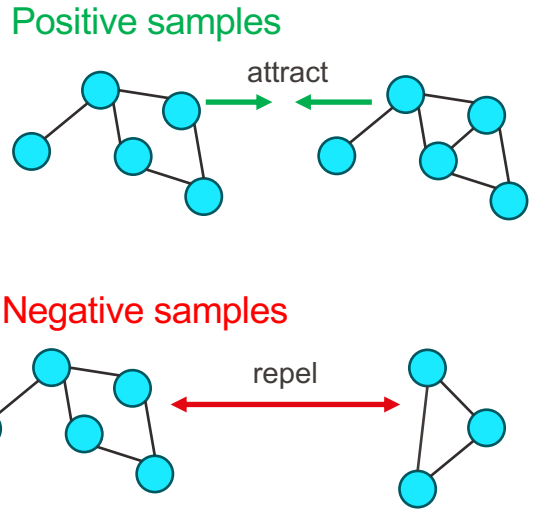
- Predict continuous auxiliary properties
- Examples:
  - Node-level properties (degree, local clustering coefficient)
  - Distance to all cluster centers (cluster center can be node with highest degree)



$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathcal{L}_{mse}\left(p_\phi\left(f_\theta(\mathcal{G})\right), c\right)$$

Contrastive SSL objective

- Maximize Mutual Information (MI) between instances with similar semantic information (positive samples)
- Minimize the MI between those with different semantic information (negative samples)
- How to define positive and negative samples?

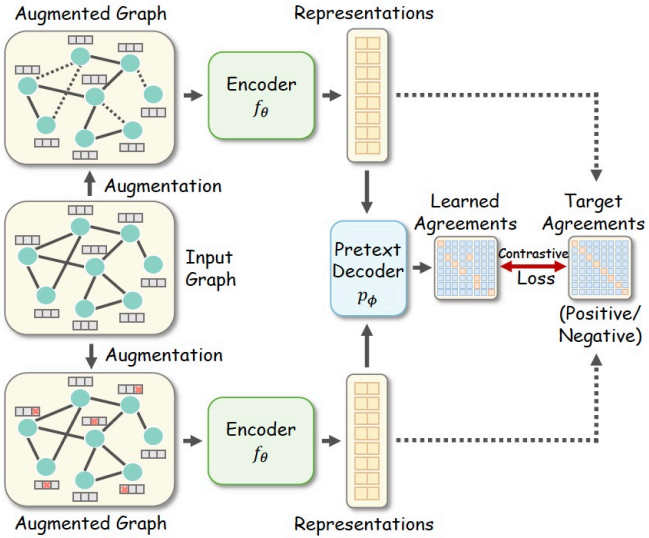


Contrastive SSL objective

You *et. al. ,* Graph Contrastive Learning with Augmentations (**GraphCL)**, 2021

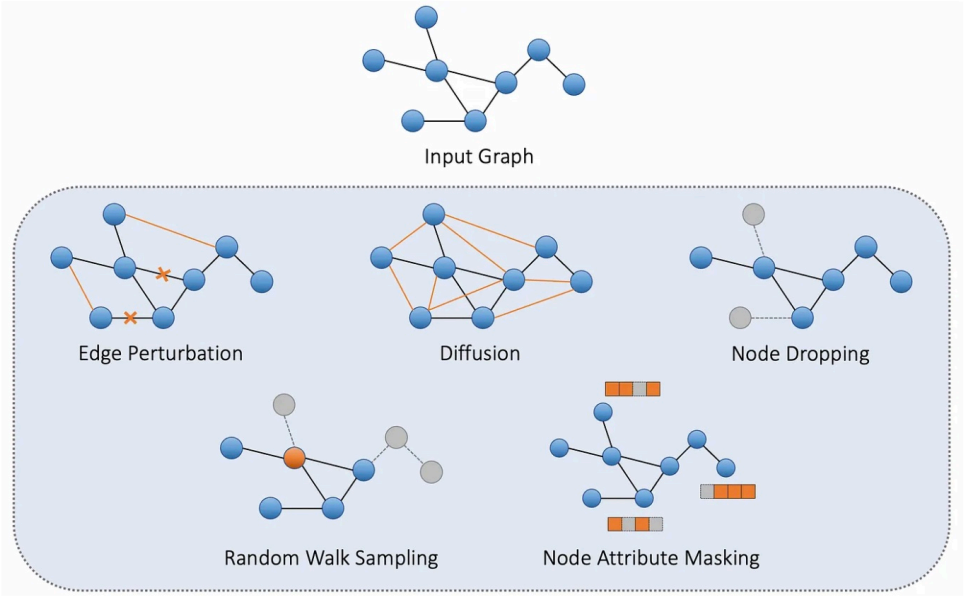
- Contrasts the representations on a graph level
- Positive samples: two augmented versions of the same input graph
- Negative samples: two different graphs from the dataset
- The task of the model is to maximize the cosine similarity between positive samples, while minimizing the cosine similarity between negative samples

$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathcal{L}_{con}\left(p_\phi\left(\tilde{\mathbf{g}}^{(1)}, \tilde{\mathbf{g}}^{(2)}\right)\right)$$



Contrastive SSL objective

Augmentations on graphs

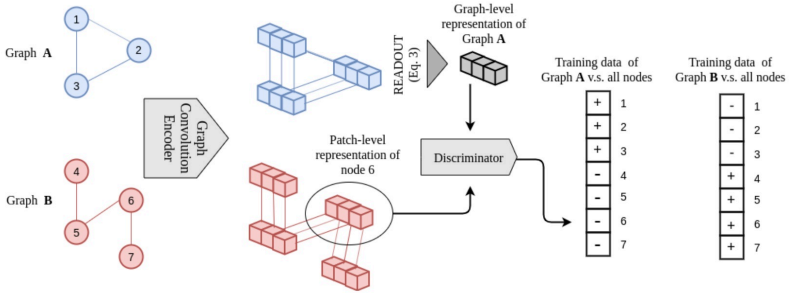


source: medium.com

Contrastive SSL objective

Sun *et. al.,* **InfoGraph:** Unsupervised and Semi-Supervised Graph-Level Representation Learning via Mutual Information Maximization, ICLR 2020

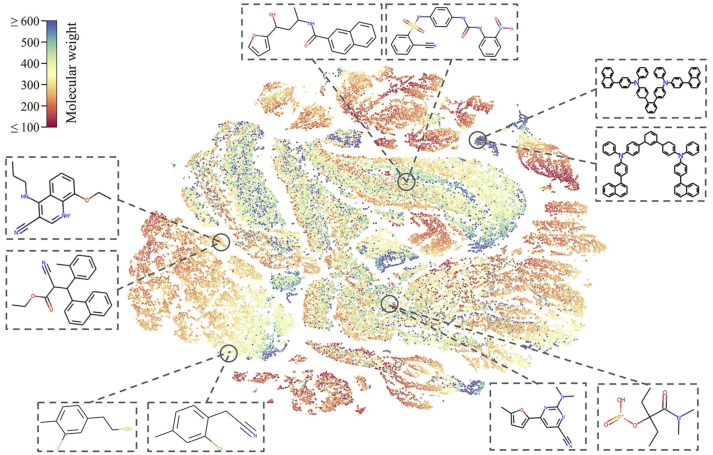
- Contrasts the between the representations of entire graphs and those of substructures (e.g. nodes)
- Positive samples: any graph with any of its nodes
- Negative samples: any graph with any of other graph's nodes
- The task of the model is to predict if the node embedding and the graph embedding originate from the same graph



Contrastive SSL objective

Icare Sakr

- With contrastive approaches, defining positive/negative sampling strategies can be challenging and sub-optimal.
- Example: slightly changing the structure of molecular graphs (e.g. removing an atom or a bond) would still lead to similar graph representations but might result in completely different behaviors



Some empirical results

- Some SSL methods such as GraphCL have already shown state-of-the-art performance especially when number of available labeled data is limited.

Dataset	NCI1	PROTEINS	DD	COLLAB	RDT-B	RDT-M5K	GITHUB	MNIST	CIFAR10
1% baseline	60.72±0.45	-	-	57.46±0.25	-	-	54.25±0.22	60.39±1.95	27.36±0.75
1% Aug.	60.49±0.46	-	-	58.40±0.97	-	-	56.36±0.42	67.43±0.36	27.39±0.44
1% GAE	61.63±0.84	-	-	63.20±0.67	-	-	59.44±0.44	57.58±2.07	21.09±0.53
1% Infomax	62.72±0.65	-	-	61.70±0.77	-	-	58.99±0.50	63.24±0.78	27.86±0.43
1% GraphCL	62.55±0.86	-	-	64.57±1.15	-	-	58.56±0.59	83.41±0.33	30.01±0.84
10% baseline	73.72±0.24	70.40±1.54	73.56±0.41	73.71±0.27	86.63±0.27	51.33±0.44	60.87±0.17	79.71±0.65	35.78±0.81
10% Aug.	73.59±0.32	70.29±0.64	74.30±0.81	74.19±0.13	87.74±0.39	52.01±0.20	60.91±0.32	83.99±2.19	34.24±2.62
10% GAE	74.36±0.24	70.51±0.17	74.54±0.68	75.09±0.19	87.69±0.40	53.58±0.13	63.89±0.52	86.67±0.93	36.35±1.04
10% Infomax	74.86±0.26	72.27±0.40	75.78±0.34	73.76±0.29	88.66±0.95	53.61±0.31	65.21±0.88	83.34±0.24	41.07±0.48
10% GraphCL	74.63±0.25	74.17±0.34	76.17±1.37	74.23±0.21	89.11±0.19	52.55±0.45	65.81±0.79	93.11±0.17	43.87±0.77

- The learned representations generalize well to unseen tasks and are more robust to noise and/or adversarial attacks.

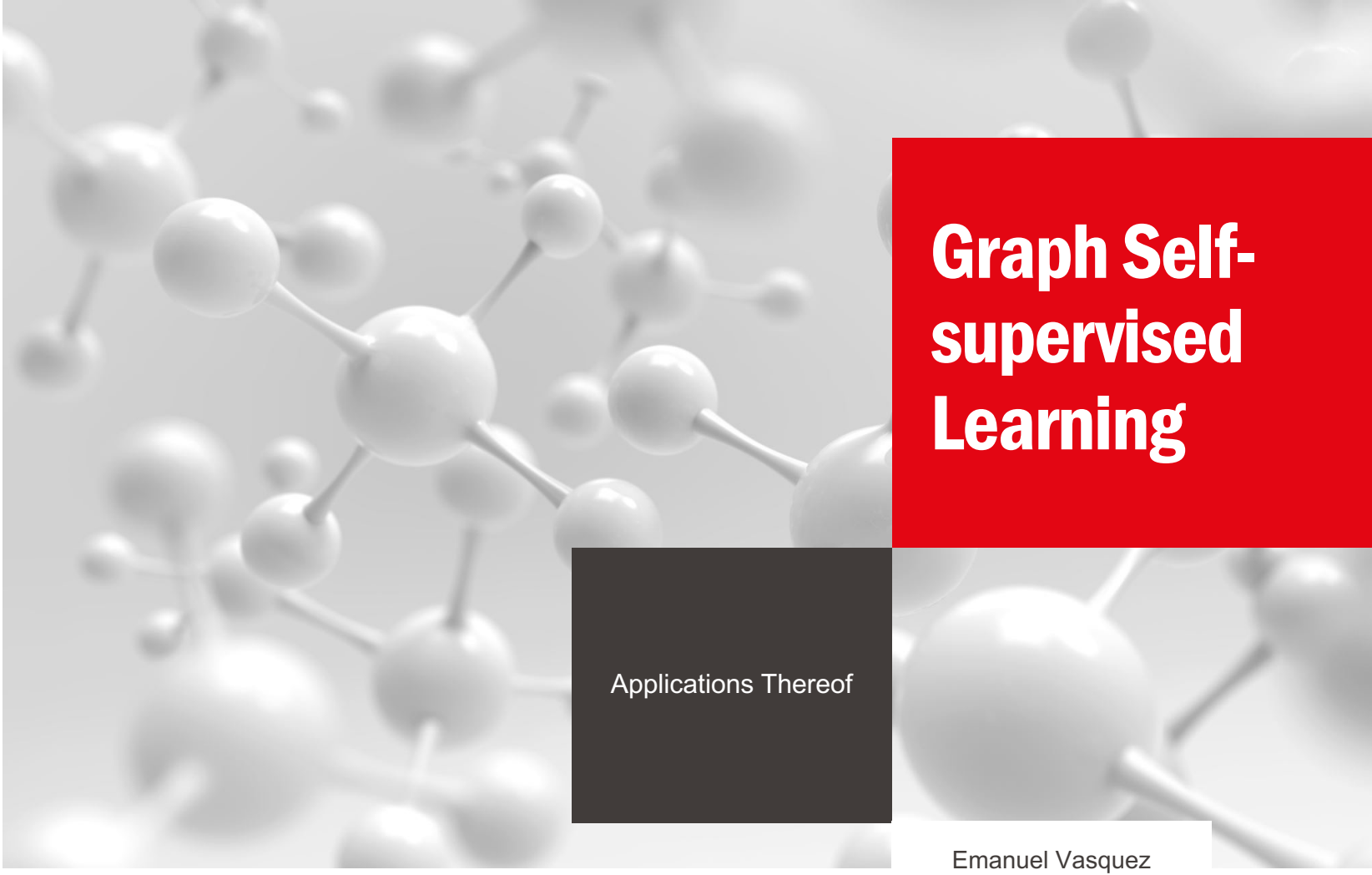
Outlook

- SSL in graphs has still not seen similar success compared to other fields such as natural language processing or computer vision (complicated to define good SSL learning schemes to infer meaningful semantic representations on graphs)

- Promising trends:
  - Hybrid SSL approaches combining multiple pretext tasks are emerging.
  - Domain-specific augmentations and scalable architectures are being developed.
  - SSL is increasingly adopted in real-world graph applications such as drug discovery and recommender systems.

Interesting references

- Graph Contrastive Learning with Augmentations - https://arxiv.org/pdf/2010.13902.pdf
- A Cookbook of Self-Supervised Learning - https://arxiv.org/pdf/2304.12210
- Graph Self-Supervised Learning: A Survey - https://arxiv.org/abs/2103.00111
- InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization - https://openreview.net/forum?id=r1lfF2NYvH



Graph Self-supervised Learning

Applications Thereof

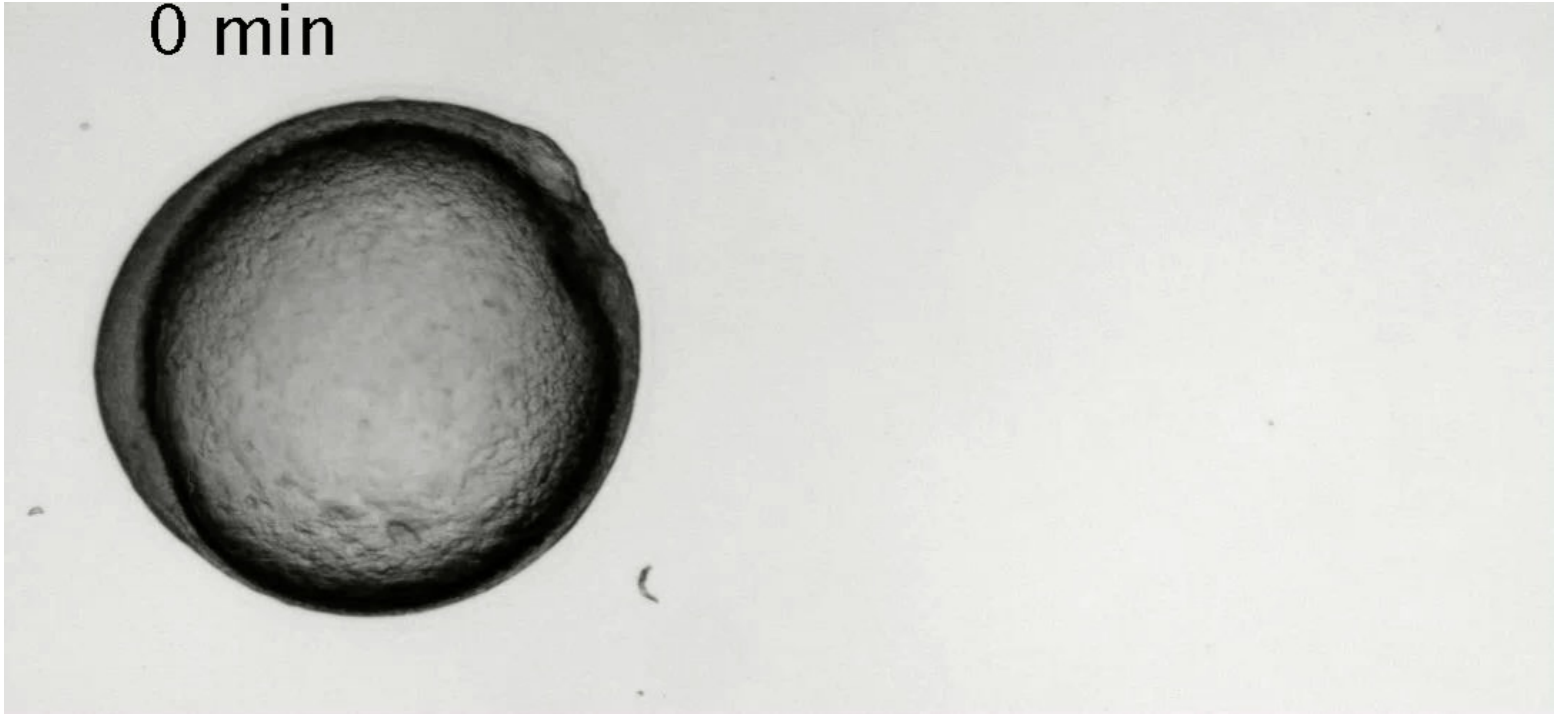
Emanuel Vasquez

Outline

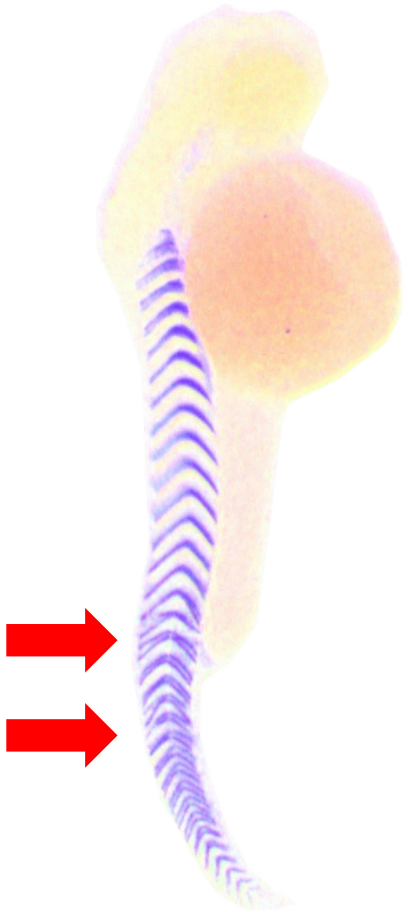
- Personal Background
- EEG Emotion Recognition
- Spatial Transcriptomics

Personal Background

Somitogenesis Oates Lab



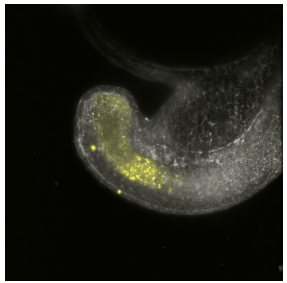
Oates Lab, Bio 221, somitogenesis movie



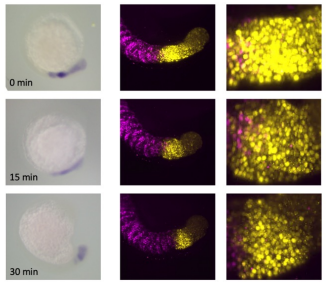
Periodic Defect Formation following Heat Shock in Zebrafish Embryos

- Emanuel Vasquez 1st Year PhD
- Current models do not explain how the PSM retains memory of heat shock
- Multimodal approaches for mechanistic explanation of defect formation

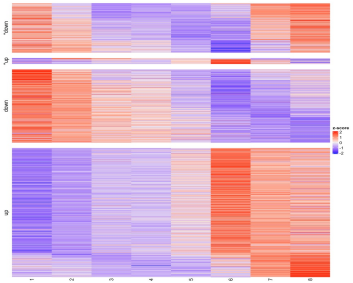
Live Imaging



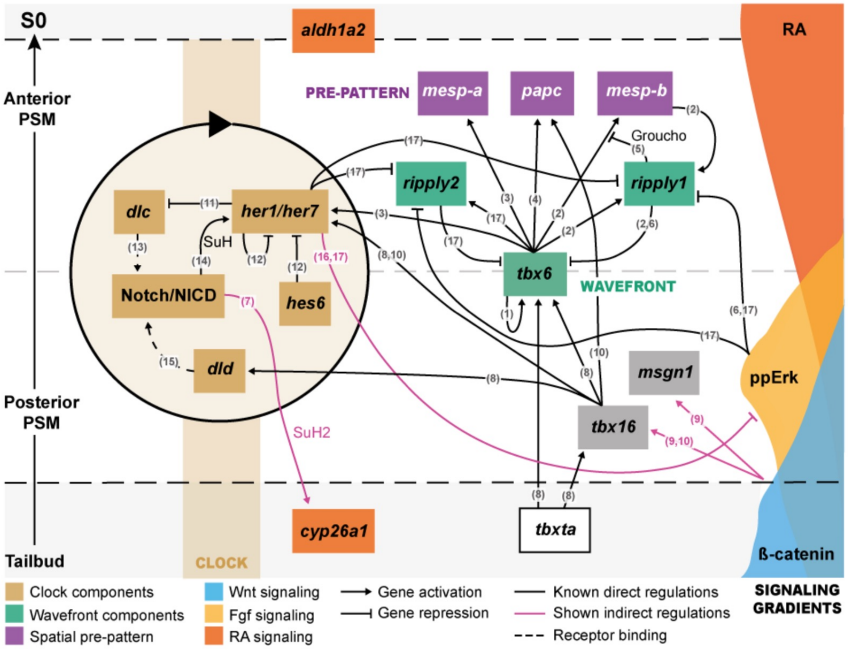
In Situ labeling

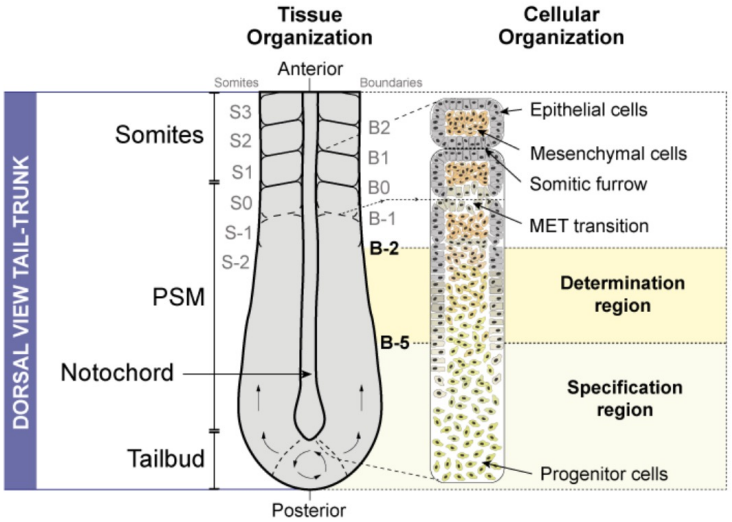


Transcriptomics

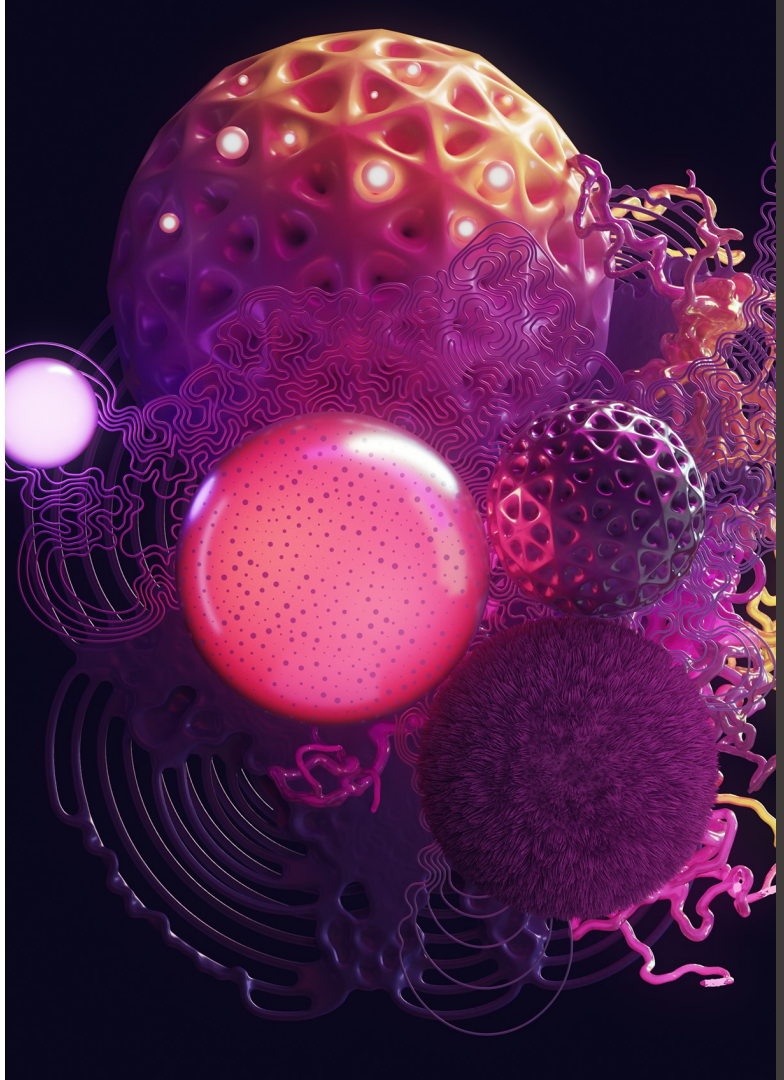


Biological Context





Loureiro, Cristina, Olivier F. Venzin, and Andrew C. Oates. "Generation of patterns in the paraxial mesoderm." *Current topics in developmental biology* 159 (2023): 372-405.



GMSS: Graph-Based Multi-Task Self-Supervised Learning for EEG Emotion Recognition

Li, Yang, et al.

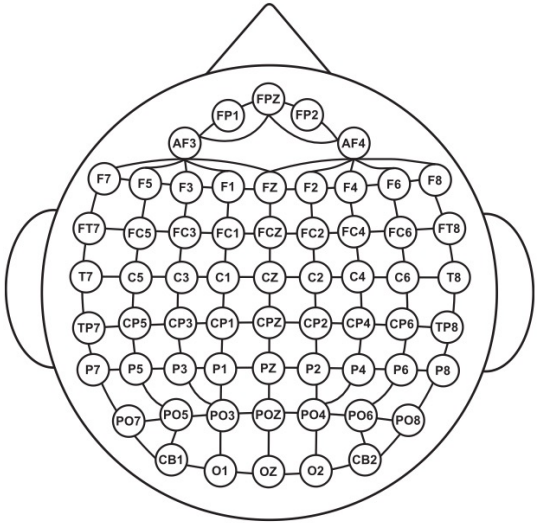
Emotion Classification from EEG

- Traditional ML
  - Handcrafted features
  - Expert input
- Current challenges
  - Generalization to new data
  - Use full EEG data for finer discrimination
  - Emotional label noise

- Three different self-supervised pretext tasks
  - Two graph-based jigsaw puzzle tasks
  - One contrastive learning task

Setup

- Spatial Jigsaw puzzle
  - Important electrodes to place as neighbors
- Frequency Jigsaw puzzle
  - Important frequency bands
- Contrastive Learning

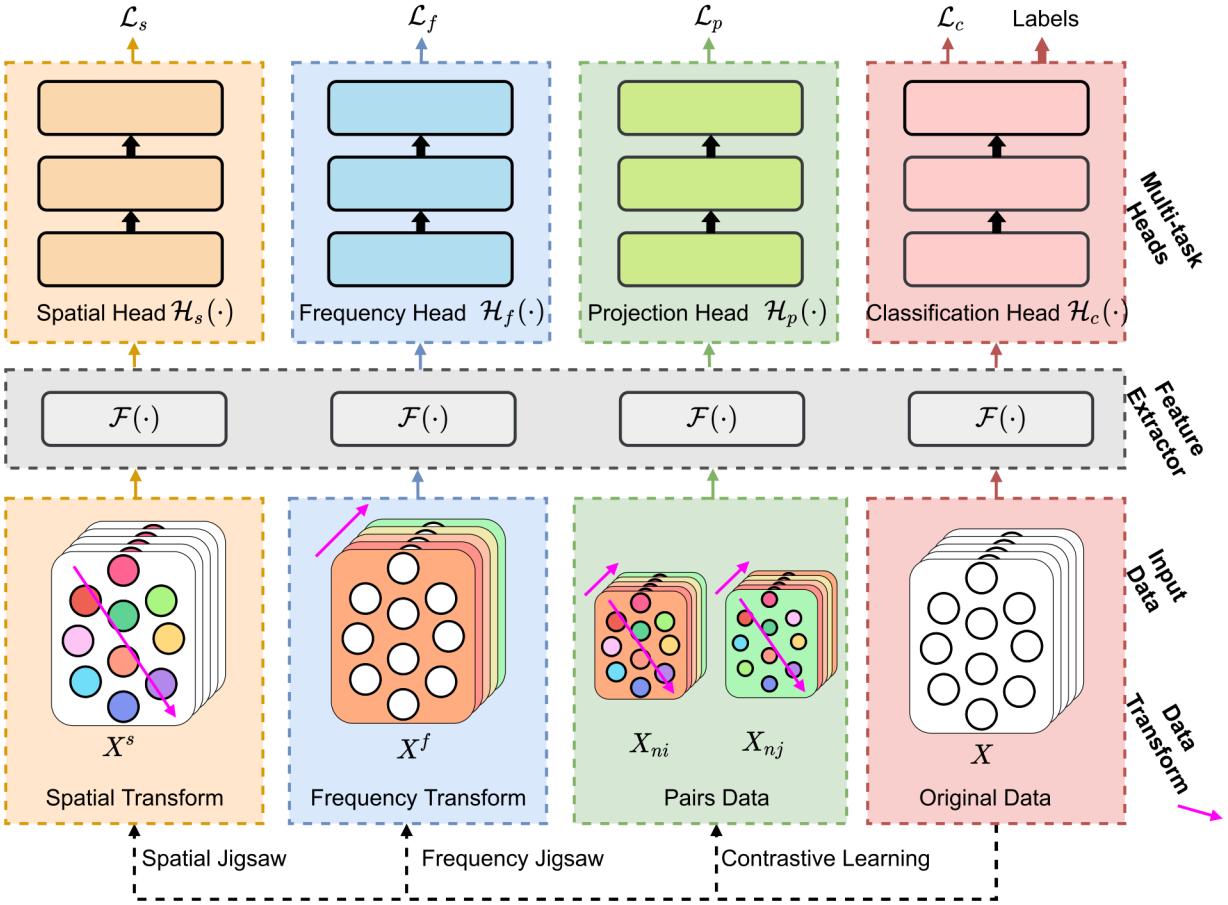


$\mathcal{G}(\mathcal{V}, \mathcal{E})$

V: EEG channels
E: Physical proximity

$X \in \mathbb{R}^{n \times d}$

n: channel number
d: frequency band

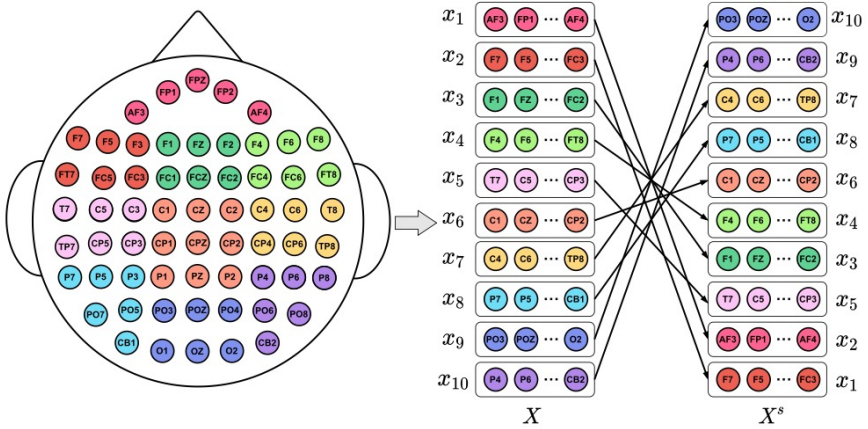

$\mathcal{L}_s$
$\mathcal{L}_f$
$\mathcal{L}_p$
$\mathcal{L}_c$
Labels
Spatial Head $\mathcal{H}_s(\cdot)$
Frequency Head $\mathcal{H}_f(\cdot)$
Projection Head $\mathcal{H}_p(\cdot)$
Classification Head $\mathcal{H}_c(\cdot)$
Multi-task Heads
$\mathcal{F}(\cdot)$
$\mathcal{F}(\cdot)$
$\mathcal{F}(\cdot)$
$\mathcal{F}(\cdot)$
Feature Extractor
X^s
X^f
X_{ni}
X_{nj}
X
Input Data
Spatial Transform
Frequency Transform
Pairs Data
Original Data
Data Transform
Spatial Jigsaw
Frequency Jigsaw
Contrastive Learning

Spatial Jigsaw Puzzle

- 10 blocks corresponding to brain regions
- Which spatial permutation corresponds to original data

$$\begin{cases} \hat{X}_1 &= (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{10}|y_1), \\ \hat{X}_2 &= (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_9|y_2), \\ &\vdots \\ \hat{X}_{10!} &= (\tilde{X}_{10}, \tilde{X}_9, \ldots, \tilde{X}_1|y_{10!}), \end{cases}$$

$$10! = 3628800$$



$$\mathcal{H}_s(\cdot)$$

Classification Head

$$(X^s, y^s) = R_{128}(X)$$

$$y^s \in \mathbb{Z}_+^{128}$$

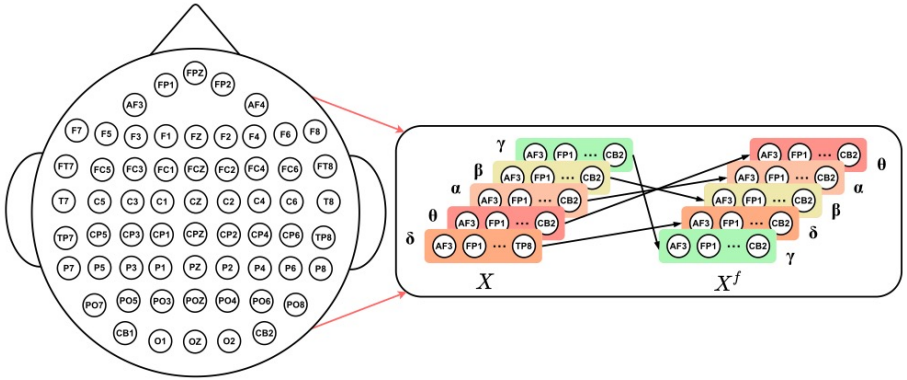
Maximize Hamming Distance

$$\mathcal{L}_s = -\sum_{i=1}^{N} \bar{y}_i^s log(\mathcal{H}_s(\mathcal{F}(X_i^s)))$$

Cross-Entropy

Frequency Jigsaw Puzzle

- 5 frequency bands are chosen
- Energy features extracted from EEG data
- Which permutation corresponds to original data



$$\delta \text{ (1-3 Hz)}, \theta \text{ (4-7 Hz)},$$
$$\alpha \text{ (8-13 Hz)}, \beta \text{ (14-30 Hz)}, \gamma \text{ (31-50 Hz)}$$

$$\mathcal{H}_f(\cdot)$$

Classification Head

$$(X^f, y^f) = R_{120}(X)$$

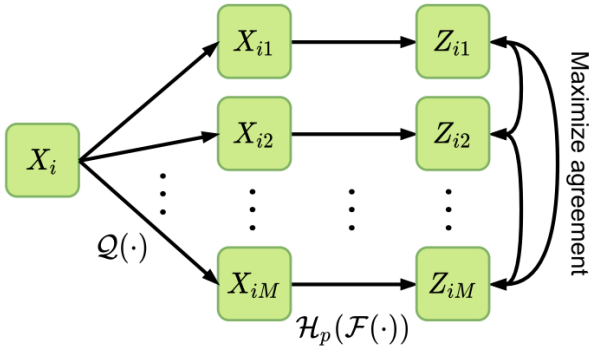
Maximize Hamming Distance

$$\mathcal{L}_f = -\sum_{j=1}^{N} \bar{y}_j^f log(\mathcal{H}_f(\mathcal{F}(X_j^f)))$$

Cross-Entropy

Contrastive Learning

- Both spatially and frequency permuted data
- Maximize self-similarity
- Decrease cross-similarity



EEG data $X_i, i \in \{1, 2, \ldots, N\}$

Augmented Data $\{X_{i1}, X_{i2}, \ldots, X_{iM}\} = \mathcal{Q}(X_i)$

$$\{X_{nm}; n \in \{1, 2, \ldots, N\}, m \in \{1, 2, \ldots, M\}\} = \mathcal{Q}(X_1) \cup \mathcal{Q}(X_2) \cup \cdots \cup \mathcal{Q}(X_N),$$

$$Z_{nm} = \mathcal{H}_p(\mathcal{F}(X_{nm}))$$

Projection Head

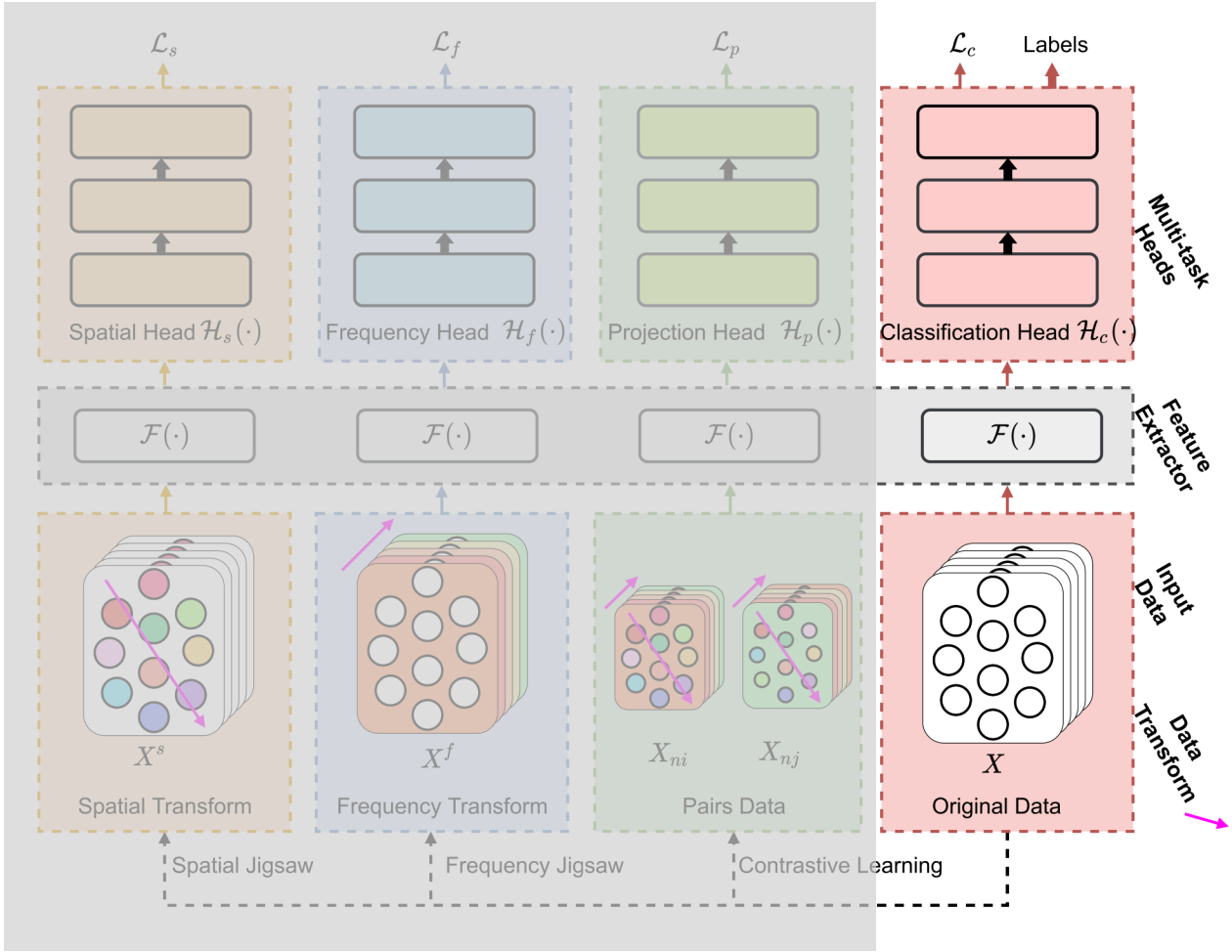
$$\mathcal{L}_p = \frac{1}{N} \sum_{n=1}^{N} \ell_n.$$

Mean positive pair similarity

$$\ell_n = -log \frac{g_+}{g_+ + g_-},$$

$$g_+ = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} exp(sim(Z_{ni}, Z_{nj})/\tau),$$

$$g_- = \sum_{o=1}^{M} \sum_{t=1}^{N} \sum_{w=1}^{M} exp(sim(Z_{no}, Z_{tw})/\tau), t \neq n,$$

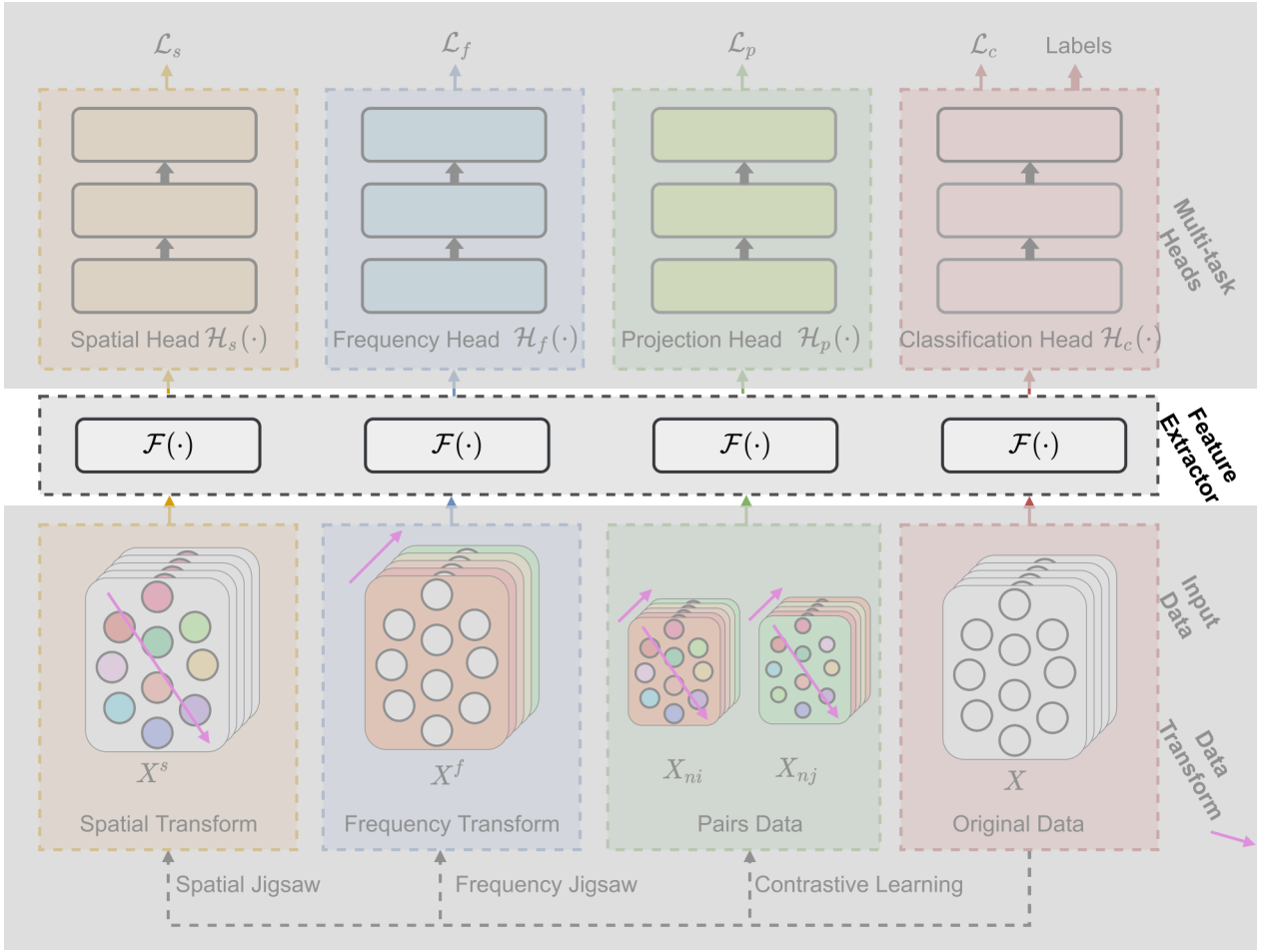


Supervised Mode

Labeled data training

Unsupervised Mode

Linear classifier



$$\mathcal{F}(X) = \sigma\left(\sum_{k=0}^{K-1} \beta_k T_k(\widetilde{L}) X\right)$$

Activation function

$\sigma(\cdot)$

Learning parameters

β_k

Chebyshev polynomial
K = 2

$T_k(\cdot)$

Scaled Laplacian

$\widetilde{L} = 2\ L/\lambda_{max} - I$

Performance Unsupervised

Three datasets

Subjects	15	15	30
Sessions	3	3	1
Emotions	3	6	4

Model	SEED		SEED-IV		MPED	
	dependent	independent	dependent	independent	dependent	independent
DeepCluster [47]	74.60/12.17*	59.01/17.65*	49.60/10.28*	44.54/09.88*	26.38/05.59*	23.25/04.86*
MoCo [49]	76.58/10.72*	58.26/15.05*	49.40/10.99*	46.19/10.04*	27.47/05.27*	23.86/04.66*
SwAV [48]	77.81/10.15*	58.65/16.66*	52.03/14.71*	49.28/10.44*	27.91/05.05*	23.50/04.81*
SimCLR [50]	81.79/11.15*	63.45/15.96*	52.47/11.57*	50.07/11.17*	29.53/05.36*	24.21/05.10*
SimSiam [51]	80.18/10.53*	63.95/11.95*	53.71/11.98*	51.24/12.47*	28.19/05.88*	24.31/04.61*
SSL-EEG [54]	83.32/09.20*	67.52/12.73*	63.59/19.82*	53.62/08.47*	25.22/04.25*	21.87/02.53*
SeqCLR [55]	82.91/08.97*	64.56/11.89*	63.13/15.41*	50.75/07.71*	30.47/06.07*	23.33/03.89*
GMSS	**89.18/09.74**	**76.04/11.91**	**65.61/17.33**	**62.13/08.33**	**34.81/06.88**	**26.97/05.01**

Classification Accuracy (mean/std)

Dependent: Within subject
Independent: Across subjects (Leave one subject out)

Supervised

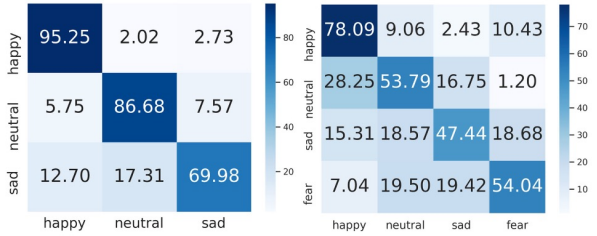
Model	SEED		SEED-IV		MPED	
	dependent	independent	dependent	independent	dependent	independent
SVM [61]	83.99/09.72	56.73/16.29	56.61/20.05	37.99/12.52	32.39/09.53	19.66/03.96
DGCNN [24]	90.40/08.49	79.95/09.02	69.88/16.29	52.82/09.23	32.37/06.08	25.12/04.20
DANN [37]	91.36/08.30	75.08/11.18	63.07/12.66	47.59/10.01	35.04/06.52	22.36/04.37
BiDANN [18]	92.38/07.04	83.28/09.60	70.29/12.63	65.59/10.39	37.71/06.04	25.86/04.92
A-LSTM [62]	88.61/10.16	72.18/10.85	69.50/15.65	55.03/09.28	38.99/07.53	24.06/04.58
BiHDM [25]	93.12/06.06	85.40/07.53	74.35/14.09	69.03/08.66	40.34/07.53	28.27/04.99
RGNN [20]	94.24/05.95	85.30/06.72	79.37/10.54	73.84/08.02	—	—
BiHDM w/o DA	91.07/08.21	81.55/09.74	72.22/14.69	67.47/08.22	38.55/07.22	27.43/04.96
RGNN w/o DA	—	81.92/09.35	—	71.65/09.34	—	—
GMSS	**96.48/04.63**	**86.52/06.22**	**86.37/11.45**	**73.48/07.41**	**40.16/06.08**	**28.49/04.42**

Model	SEED		SEED-IV		MPED	
	dependent	independent	dependent	independent	dependent	independent
DeepCluster [47]	74.60/12.17*	59.01/17.65*	49.60/10.28*	44.54/09.88*	26.38/05.59*	23.25/04.86*
MoCo [49]	76.58/10.72*	58.26/15.05*	49.40/10.99*	46.19/10.04*	27.47/05.27*	23.86/04.66*
SwAV [48]	77.81/10.15*	58.65/16.66*	52.03/14.71*	49.28/10.44*	27.91/05.05*	23.50/04.81*
SimCLR [50]	81.79/11.15*	63.45/15.96*	52.47/11.57*	50.07/11.17*	29.53/05.36*	24.21/05.10*
SimSiam [51]	80.18/10.53*	63.95/11.95*	53.71/11.98*	51.24/12.47*	28.19/05.88*	24.31/04.61*
SSL-EEG [54]	83.32/09.20*	67.52/12.73*	63.59/19.82*	53.62/08.47*	25.22/04.25*	21.87/02.53*
SeqCLR [55]	82.91/08.97*	64.56/11.89*	63.13/15.41*	50.75/07.71*	30.47/06.07*	23.33/03.89*
GMSS	**89.18/09.74**	**76.04/11.91**	**65.61/17.33**	**62.13/08.33**	**34.81/06.88**	**26.97/05.01**

Classification Accuracy (mean/std)

Dependent: Within subject
Independent: Across subjects (Leave one subject out)

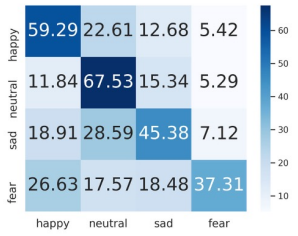
Confusion Matrix Unsupervised



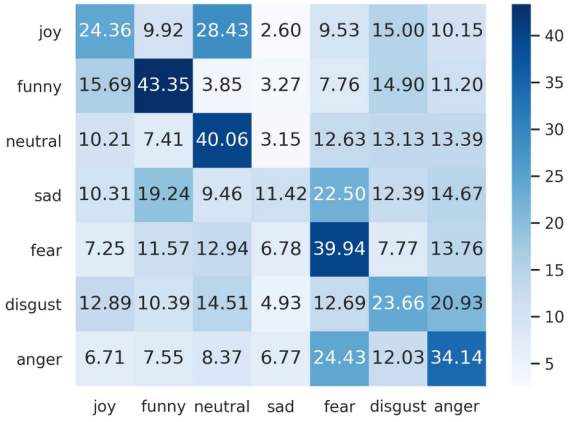
(a) SEED

(b) SEED-IV

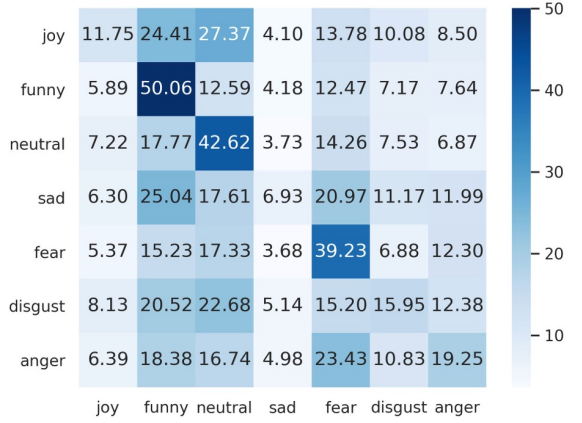
(d) SEED



(e) SEED-IV



(c) MPED



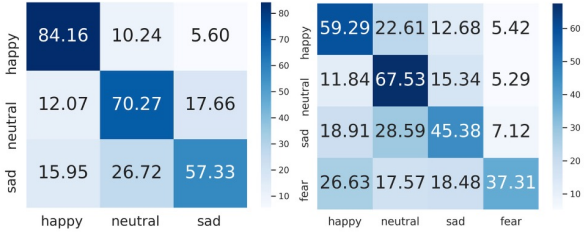
(f) MPED

(1) Subject-dependent experimental results

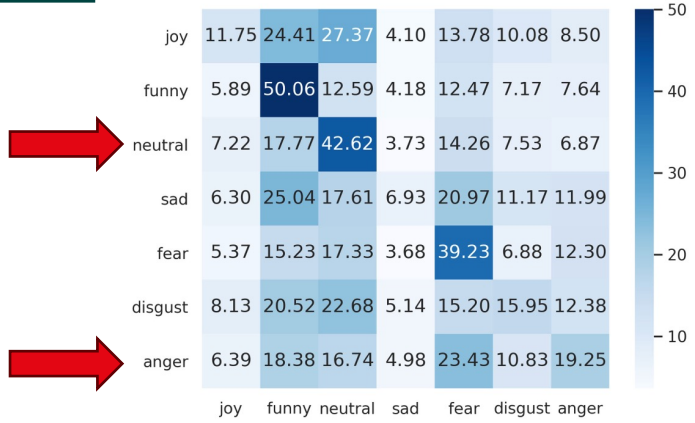
(2) Subject-independent experimental results

Confusion Matrix

Unsupervised



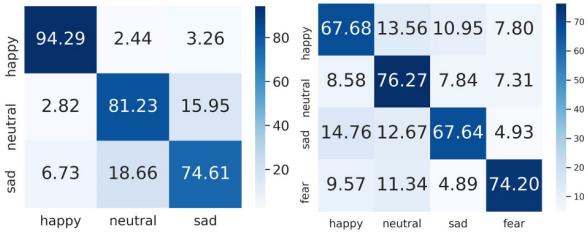
(d) SEED (e) SEED-IV



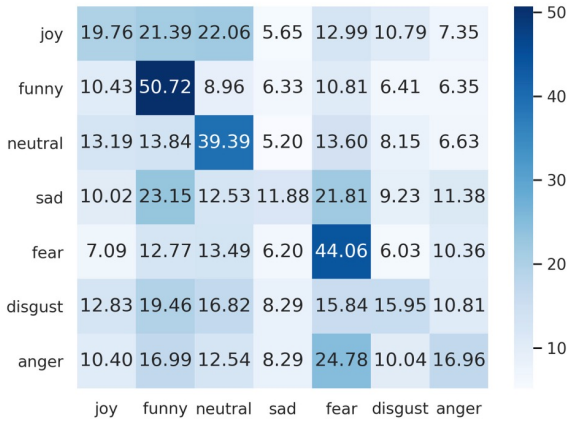
(f) MPED

(2) Subject-independent experimental results

Supervised



(d) SEED (e) SEED-IV



(f) MPED

(2) Subject-independent experimental results

Ablation Study

Ablation Models	SEED		SEED-IV		MPED	
	unsupervised	supervised	unsupervised	supervised	unsupervised	supervised
GMSS-S	**86.43/09.36**	**88.82/08.81**	**63.29/16.50**	79.01/16.13	31.91/06.09	**35.82/06.17**
GMSS-F	84.84/10.68	86.75/08.64	62.31/16.24	**79.56/14.31**	**33.32/06.45**	34.98/06.13
GMSS-C	84.14/10.65	85.92/09.78	59.77/17.64	77.68/15.02	32.28/06.07	35.59/05.99
GMSS-SF	**88.24/09.77**	**94.98/09.34**	**64.21/14.92**	**84.54/14.30**	**33.81/05.67**	37.78/05.95
GMSS-SC	86.81/10.37	93.94/09.57	62.66/17.47	83.42/11.83	32.42/06.42	**38.06/05.65**
GMSS-FC	86.35/10.15	92.93/08.29	62.83/17.29	83.83/12.49	33.65/06.66	37.11/05.97
GMSS	**89.18/09.74**	**96.48/04.63**	**65.61/17.33**	**86.37/11.45**	**34.81/06.88**	**40.16/06.08**

Self-supervised Learning

- Hybrid graph SSL
  - Multiple explicit pretext decoders
- Generation-based graph SSL
  - Data features and topology are permuted
- Contrast-based graph SSL
  - Pairwise data in contrastive learning

ResST: A graph self-supervised residual learning framework for domain identification and data integration of spatial transcriptomics

Huang, Jinjin, et al.

Spatial Transcriptomics

- Methods borrowed from single cell data
  - No spatial localization
- Other methods do not consider biological effects
  - Cell properties
  - Surrounding cells
  - Tissue microenvironment
- Methods integrate multiple data sets
  - Do not correct for batch effect

- Graph Neural Network
  - Non linear spatial relationships
- Marginal Disparity Discrepancy
  - Align latent embedding
  - Correcting batch effects

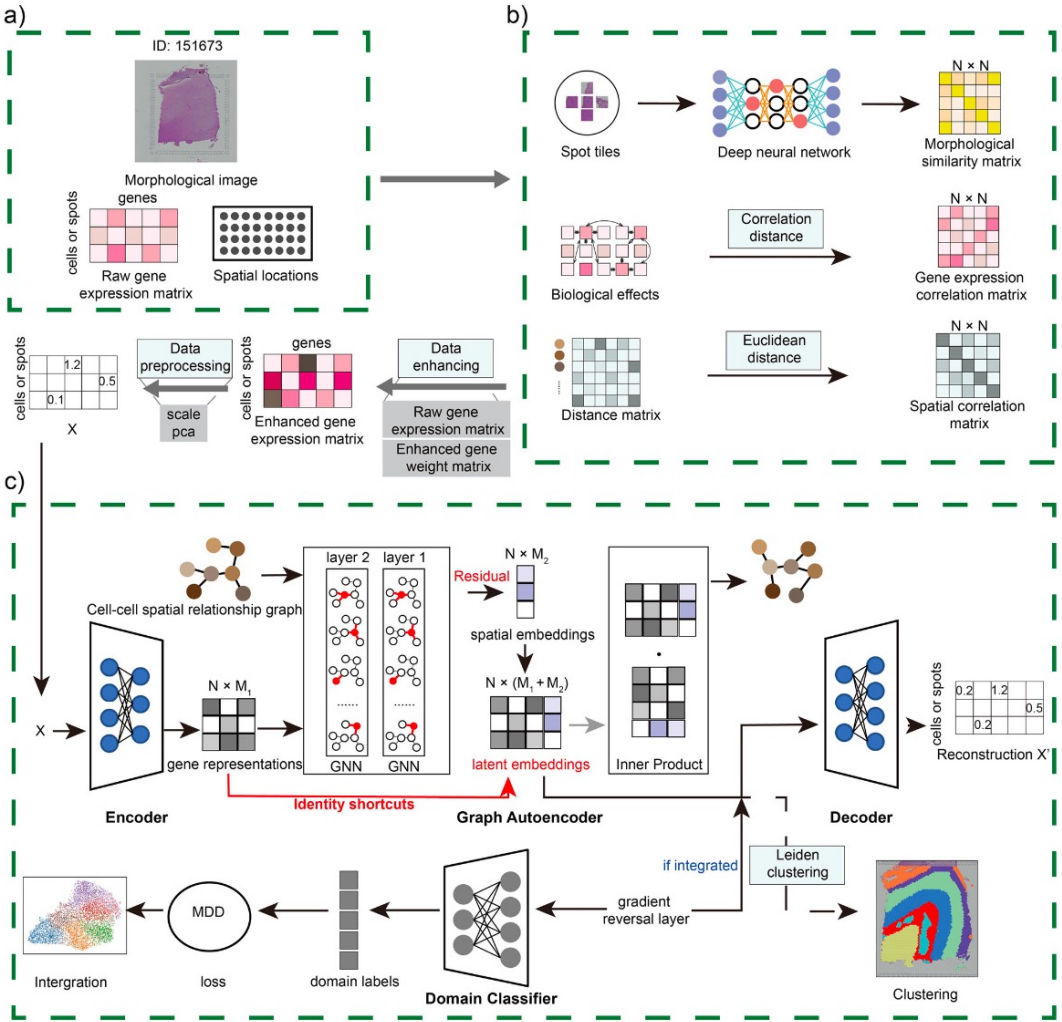

a)
ID: 151673
Morphological image
genes
cells or spots
Raw gene expression matrix
Spatial locations
b)
Spot tiles
Deep neural network
N × N
Morphological similarity matrix
Biological effects
Correlation distance
Gene expression correlation matrix
Distance matrix
Euclidean distance
Spatial correlation matrix
cells or spots
X
Data preprocessing
scale pca
genes
Enhanced gene expression matrix
Data enhancing
Raw gene expression matrix
Enhanced gene weight matrix
c)
Cell-cell spatial relationship graph
layer 2
layer 1
GNN
GNN
N × M1
gene representations
Residual
N × M2
spatial embeddings
N × (M1 + M2)
latent embeddings
Inner Product
Identity shortcuts
Encoder
Graph Autoencoder
Decoder
Reconstruction X'
if integrated
Leiden clustering
gradient reversal layer
Domain Classifier
domain labels
loss
MDD
Intergration
Clustering

Fig. 1

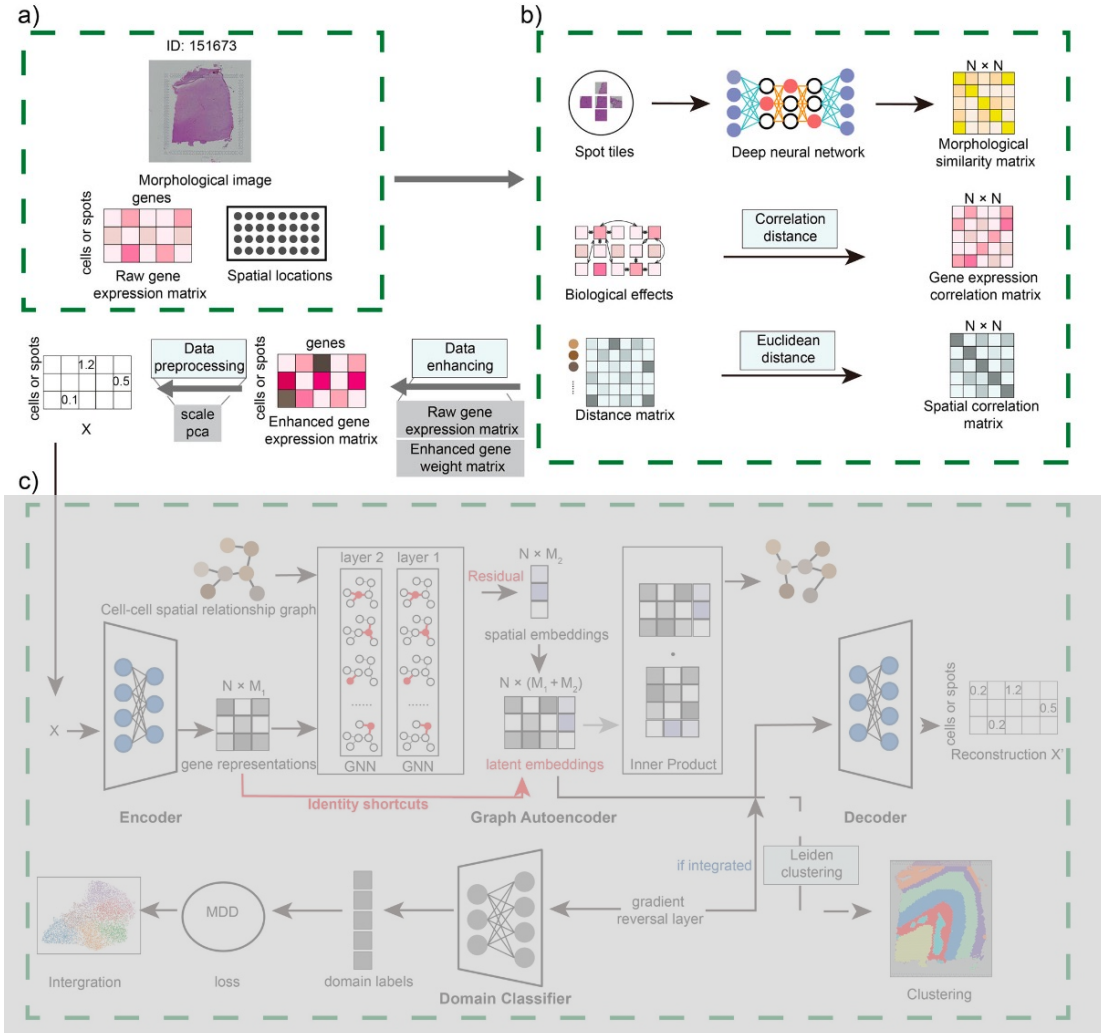

a)
ID: 151673
Morphological image
genes
cells or spots
Raw gene expression matrix
Spatial locations
b)
Spot tiles
Deep neural network
N × N
Morphological similarity matrix
Biological effects
Correlation distance
Gene expression correlation matrix
Distance matrix
Euclidean distance
Spatial correlation matrix
cells or spots
X
Data preprocessing
scale pca
genes
Enhanced gene expression matrix
Data enhancing
Raw gene expression matrix
Enhanced gene weight matrix
c)
Cell-cell spatial relationship graph
layer 2
layer 1
GNN
GNN
N × M1
gene representations
N × M2
Residual
spatial embeddings
N × (M1 + M2)
latent embeddings
Inner Product
Identity shortcuts
Encoder
Graph Autoencoder
Decoder
Reconstruction X'
if integrated
Leiden clustering
gradient reversal layer
Domain Classifier
domain labels
loss
MDD
Intergration
Clustering

Fig. 1

Data Enhancement

- Gene expression $\mathbf{GE} \in \mathbb{R}^{n \times d}$
  - Filtered / reduced PCA

$$\mathbf{GW}(s_i, s_j) = \mathbf{GW}_{ij} = \frac{\left(\mathbf{GE}_i - \overline{\mathbf{GE}_i}\right) \cdot \left(\mathbf{GE}_j - \overline{\mathbf{GE}_j}\right)}{||\mathbf{GE}_i - \overline{\mathbf{GE}_i}||_2 ||\mathbf{GE}_j - \overline{\mathbf{GE}_j}||_2}$$

- Morphological similarity
  - CNN on segmented histological image + PCA

$$\mathbf{MW}(s_i, s_j) = \mathbf{MW}_{ij} = \frac{\mathbf{M}_i \cdot \mathbf{M}_j}{||\mathbf{M}_i||_2 ||\mathbf{M}_j||_2}$$

- Spatial correlation
  - Euclidian distance

$$\mathbf{SW}_{ij} = \begin{cases} 1, \mathbf{SW}_{ij} < r \\ 0, \mathbf{SW}_{ij} \geq r \end{cases}$$

$$\mathbf{EGW}(s_i, s_j) = \mathbf{EGW}_{ij} = \mathbf{SW}_{ij} \cdot \mathbf{MW}_{ij} \cdot \mathbf{GW}_{ij}$$

$$\mathbf{GE}'_i = \mathbf{GE}_i + \frac{\sum_{j=1}^{m} \mathbf{GE}_j \cdot \mathbf{EGW}_{ij}}{m}$$

GE'

- Log scaled
- Normalized
  - Unit Variance
  - Zero mean

Graph Construction

- K nearest neighbor
  - Euclidian distance ranked
  - K = 12

  - V: Cells or spots
  - A: Adjacency matrix
  - D: Degree matrix

$$G = (V, A)$$

$$A' = D^{-1/2} \cdot A \cdot D^{-1/2}$$

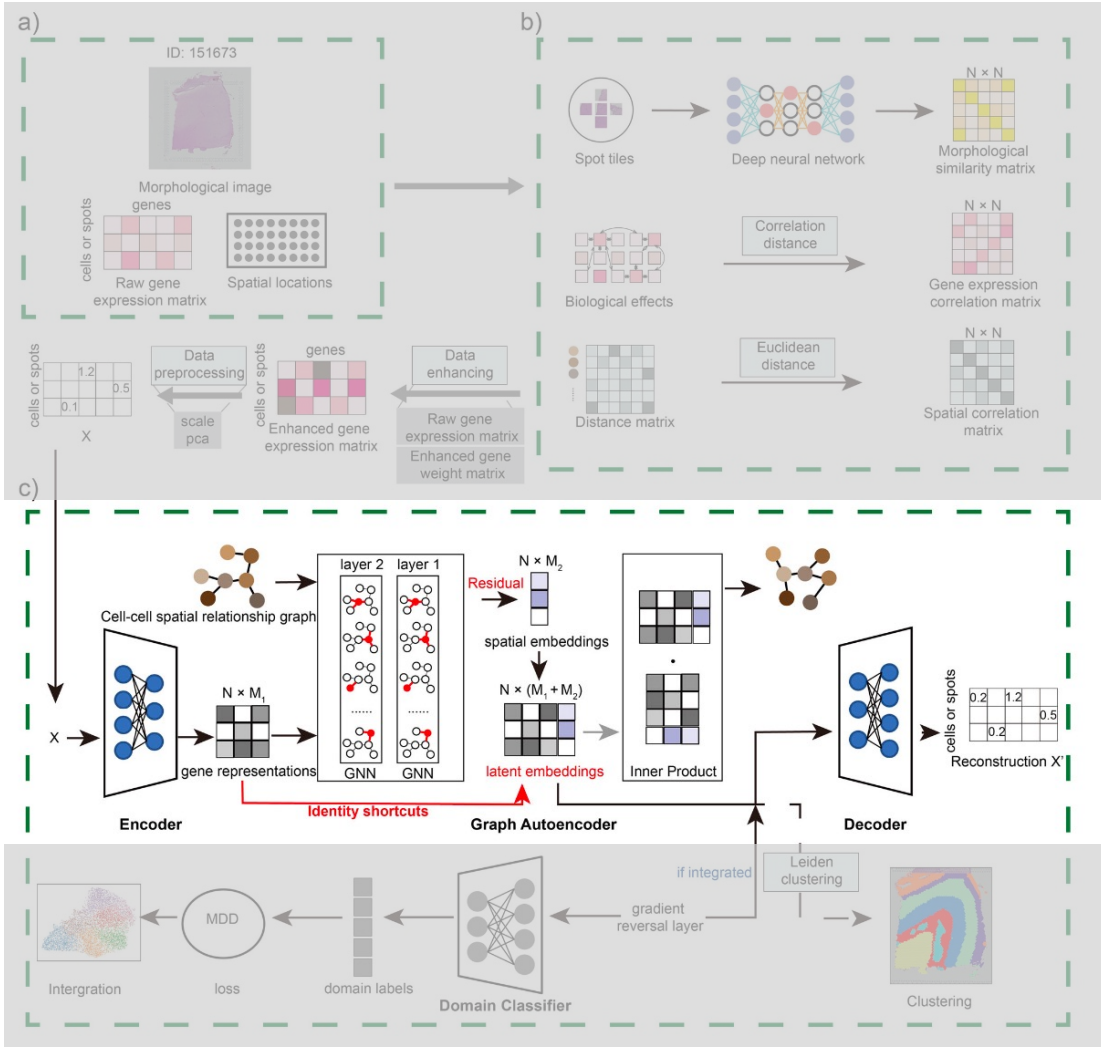
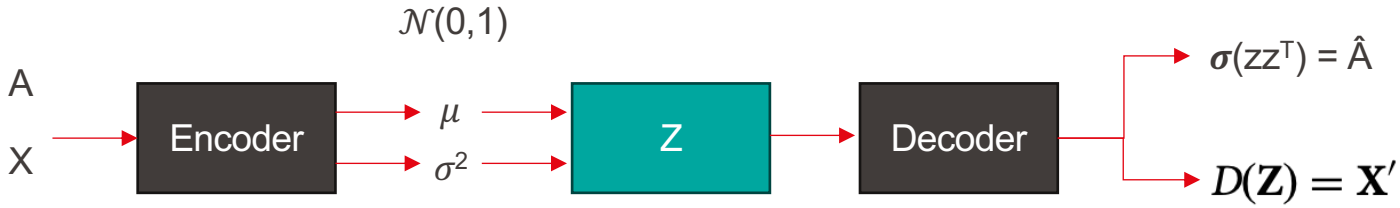
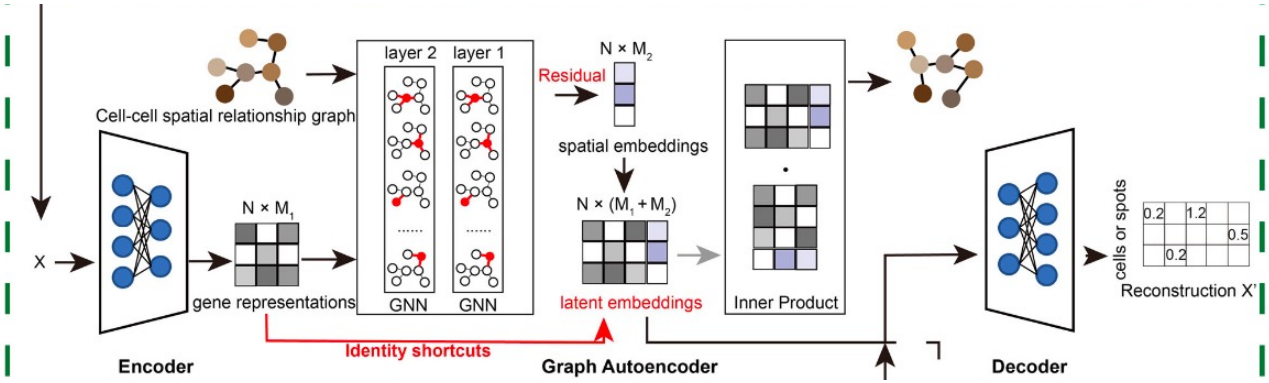

a)
ID: 151673
Morphological image
genes
cells or spots
Raw gene expression matrix
Spatial locations
b)
Spot tiles
Deep neural network
N × N
Morphological similarity matrix
Biological effects
Correlation distance
Gene expression correlation matrix
Distance matrix
Euclidean distance
Spatial correlation matrix
cells or spots
X
Data preprocessing
scale
pca
Enhanced gene expression matrix
Data enhancing
Raw gene expression matrix
Enhanced gene weight matrix
c)
Cell-cell spatial relationship graph
X
N × M1
gene representations
layer 2
layer 1
GNN
GNN
N × M2
Residual
spatial embeddings
N × (M1 + M2)
latent embeddings
Inner Product
Identity shortcuts
Encoder
Graph Autoencoder
Decoder
Reconstruction X'
MDD
Intergration
loss
domain labels
Domain Classifier
gradient reversal layer
if integrated
Leiden clustering
Clustering

Fig. 1

Dimension Reduction

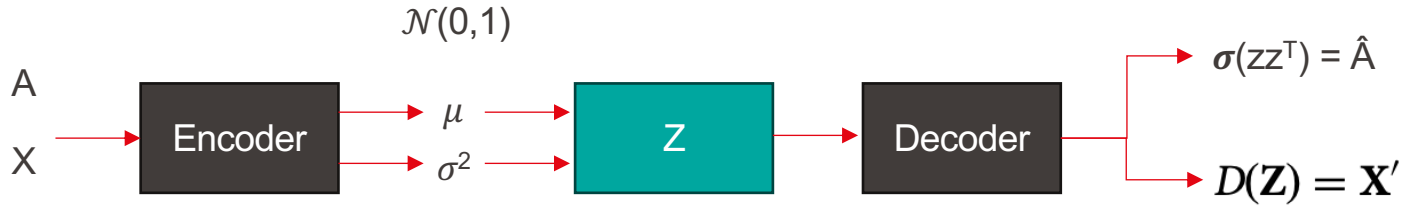
- VGAE: Variational Graph Autoencoder





Dimension Reduction

- VGAE: Variational Graph Autoencoder



$$L = \frac{1}{n}\sum_{i=1}^{n} ||\mathbf{X}_i - \mathbf{X'}_i||^2 + KL[q(\mathbf{Z}_g|\mathbf{A}, \mathbf{Z}_e)||p(\mathbf{Z}_g)]$$

Kullback–Leibler divergence

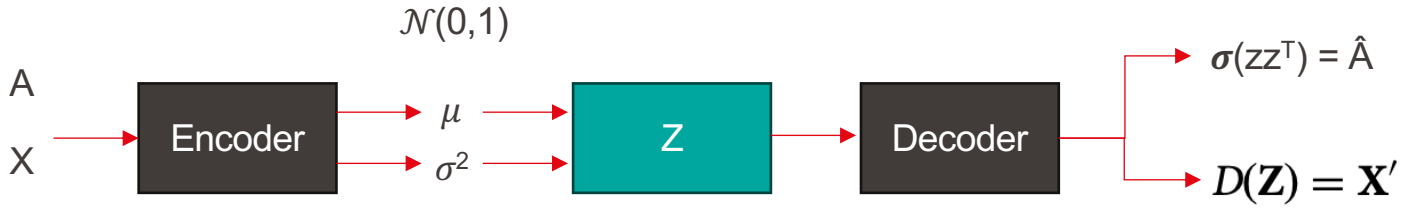
$$q\left(\mathbf{Z_g}|\mathbf{A}, \mathbf{Z_e}\right) = \prod_{i=1}^{n} q\left(\mathbf{z}_i|\mathbf{A}, \mathbf{Z_e}\right)$$

$$p(\mathbf{Z}_g) = \prod_i \mathcal{N}(\mathbf{z}_i|0, I)$$

$$q\left(\boldsymbol{z}_i|\mathbf{A}, \mathbf{Z_e}\right) = \mathcal{N}\left(\mathbf{z}_i|\mu_i, diag\left(\sigma_i^2\right)\right)$$

Dimension Reduction

- VGAE: Variational Graph Autoencoder



$$\mathbf{z}_i = GNN_\mu(\mathbf{A}, \mathbf{Z_e})$$
$$\log \sigma = GNN_\sigma(\mathbf{A}, \mathbf{Z_e})$$

First layer: $W_{0\sigma} = W_{0\mu}$ Dim. reduction

$$GNN(\mathbf{Z_e}, \mathbf{A}) = \mathbf{A}' ReLU\left(\mathbf{A}'\mathbf{Z_e}\mathbf{W_o}\right)\mathbf{W_1}$$

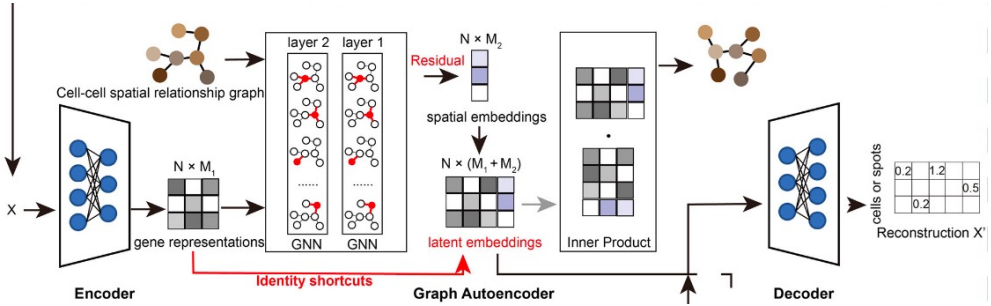
$$\mathbf{A}' = \mathbf{D}^{-1/2} \cdot \mathbf{A} \cdot \mathbf{D}^{-1/2}$$

Second layer: $W_{1\sigma} \neq W_{1\mu}$

Dimension Reduction

- X: Gene expression vector
  - n: number of spots
  - Top 200 principal components

$$\mathbf{X} \in \mathbb{R}^{n \times 200}$$



- Multi layer fully connected encoder
  - H_e: 20

$$E(\mathbf{X}) = \mathbf{Z_e}$$

$$\mathbf{Z_e} \in \mathbb{R}^{n \times H_e}$$

- Two layer GNN
  - H_g: 8

$$GNN(\mathbf{Z_e}) = \mathbf{Z_g}$$

$$\mathbf{Z_g} \in \mathbb{R}^{n \times H_g}$$

Residual: $Z \in R^{n \times (H\_e + H\_g)}$

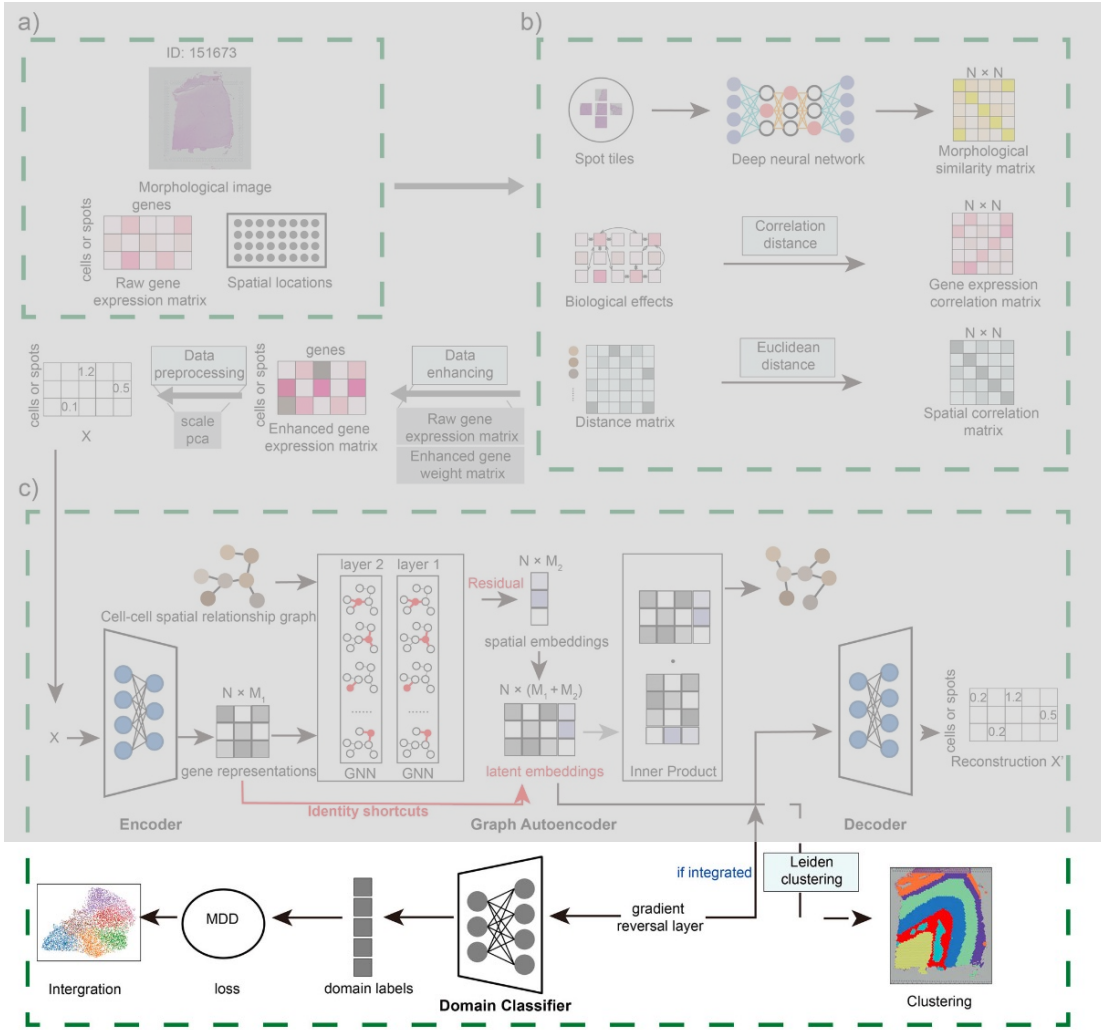

a)
ID: 151673
Morphological image
genes
cells or spots
Raw gene expression matrix
Spatial locations
b)
Spot tiles
Deep neural network
N × N
Morphological similarity matrix
Biological effects
Correlation distance
Gene expression correlation matrix
Distance matrix
Euclidean distance
Spatial correlation matrix
cells or spots
X
Data preprocessing
scale pca
Enhanced gene expression matrix
Data enhancing
Raw gene expression matrix
Enhanced gene weight matrix
c)
Cell-cell spatial relationship graph
layer 2
layer 1
N × M2
Residual
spatial embeddings
N × (M1 + M2)
N × M1
gene representations
GNN
latent embeddings
Inner Product
Identity shortcuts
Encoder
Graph Autoencoder
Decoder
Reconstruction X'
if integrated
Leiden clustering
gradient reversal layer
MDD
Intergration
loss
domain labels
Domain Classifier
Clustering

Fig. 1

Integration of Data Types

- Vertical integration
  - Same samples, different analytes
  - Align based on H&E images
  - Construct G
- Horizontal integration
  - Same analyte, different samples
  - MDD
- Batch effect
  - MDD
- MDD
  - Loss function for dissimilarity
  - Domain classifier attempts to differentiate between samples (y_i)

$$L = -\frac{1}{n}\sum_{i=1}^{n} y_i \cdot \log\left(p\left(y_i\right)\right) + \left(1 - y_i\right) \cdot \log\left(1 - p\left(y_i\right)\right)$$

Binary Cross-entropy

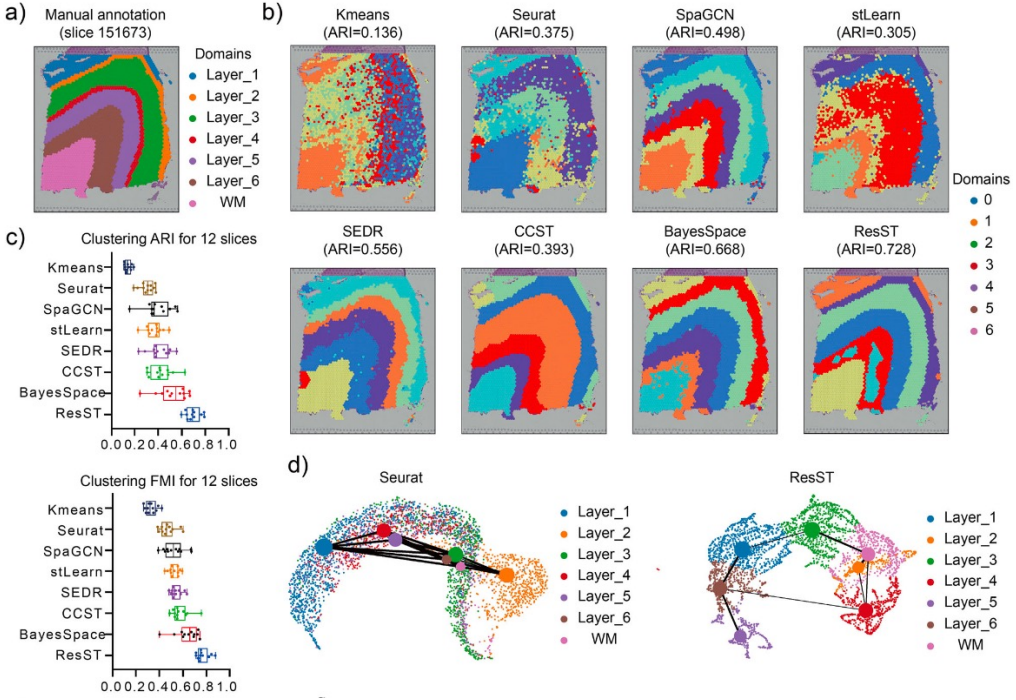
Training

- Pre-Training
  - X is augmented with Noise
- Training
  - Iterative loss minimization
    - Reconstruction error
    - KL divergence
    - MDD : Adversarial loss
  - K means clustering on embeddings
  - Calinski-Halabasz Index
    - Measure of cluster separation
  - CH score used to select resolution

Claims

- Captures finer details
- Integrates multiple datasets / sections
- Corrects for batch effect

Performance

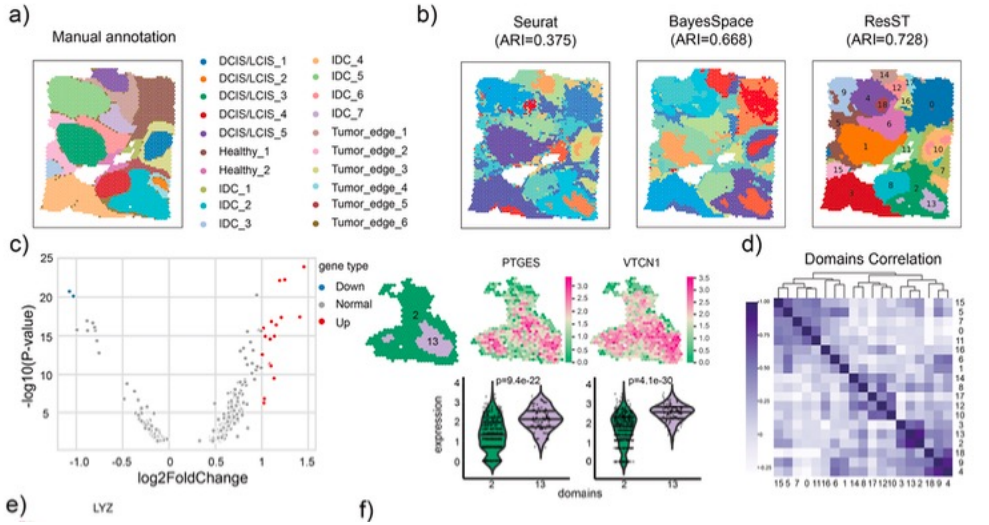


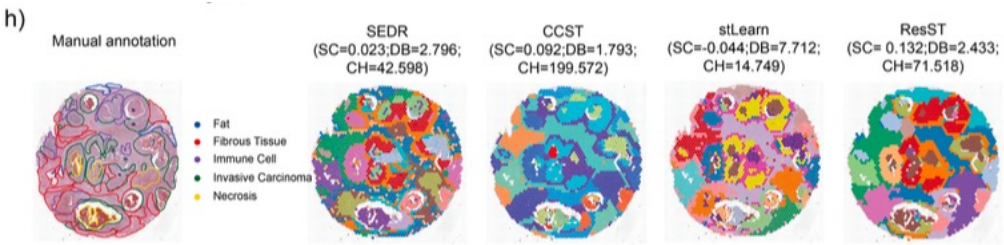
Human dorsolateral prefrontal cortex

Adjusted Rand Index
Folkes-Mallows Index

Fig. 2

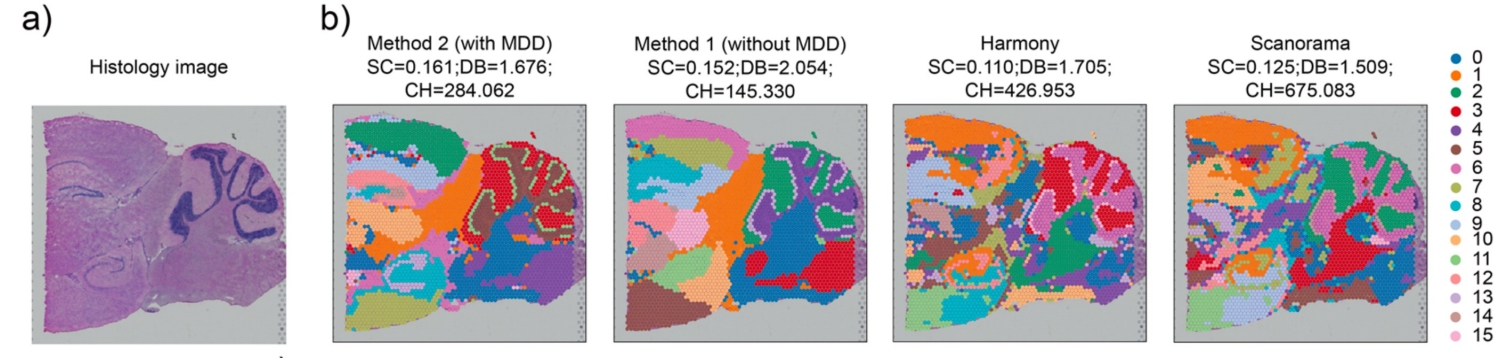
Finer details

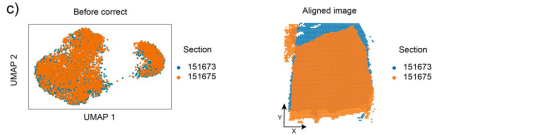


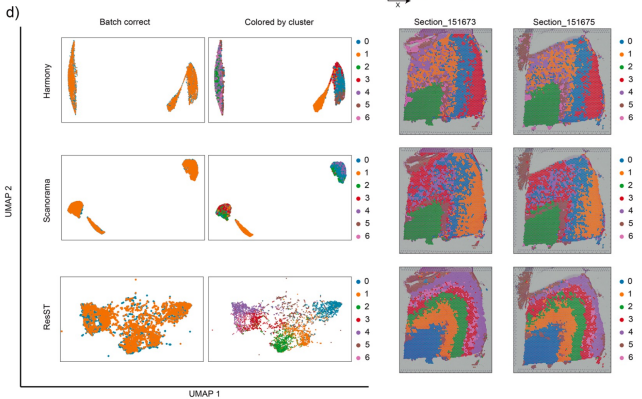


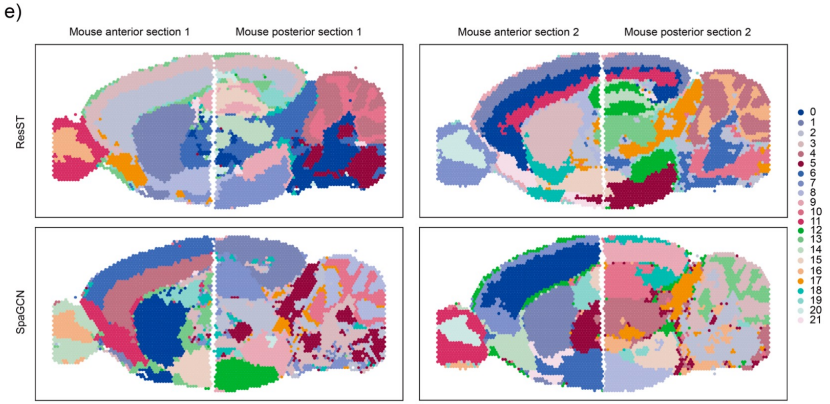
Identification of subcluster

MDD


a)
Histology image
b)
Method 2 (with MDD)
SC=0.161;DB=1.676;
CH=284.062
Method 1 (without MDD)
SC=0.152;DB=2.054;
CH=145.330
Harmony
SC=0.110;DB=1.705;
CH=426.953
Scanorama
SC=0.125;DB=1.509;
CH=675.083
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15


c)
Before correct
Aligned image
UMAP 2
UMAP 1
Section
151673
151675


d)
Batch correct
Colored by cluster
Section_151673
Section_151675
Harmony
Scanorama
ResST
UMAP 2
UMAP 1
0 1 2 3 4 5 6


e)
Mouse anterior section 1
Mouse posterior section 1
Mouse anterior section 2
Mouse posterior section 2
ResST
SpaGCN
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Claims

- Captures finer details
  - Relative to manual annotation
  - Blind spot of expert perception of data
  - Limited by ST technique
    - Lower resolution may perform similar to other methods
- Integrates multiple datasets / sections
- Corrects for batch effect
  - MDD effect not quantitatively shown

Self-supervised Learning

- Hybrid-graph SSL
  - MDD is used for data integration
  - Three loss functions are used when training

- Generation-based graph SSL
  - Noisy data (X) is used in pre-training
  - Perturbation of features