سری کارگاه‌های یادگیری عمیق

یادگیری چند نمایی و کاربرد آن در کشف دارو

پژوهشکده هوش مصنوعی
دانشگاه شیراز

دانشگاه شیراز

عباس مهربانیان

دانشجو ارشد هوش مصنوعی

Deep Learning Workshops

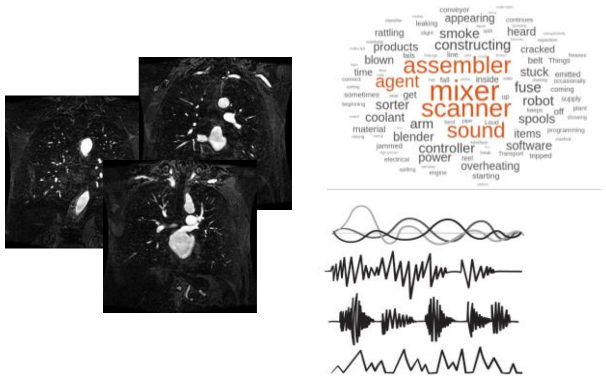
Multi-view learning and its applications in drug discovery

Abbas Mehrbaniyan

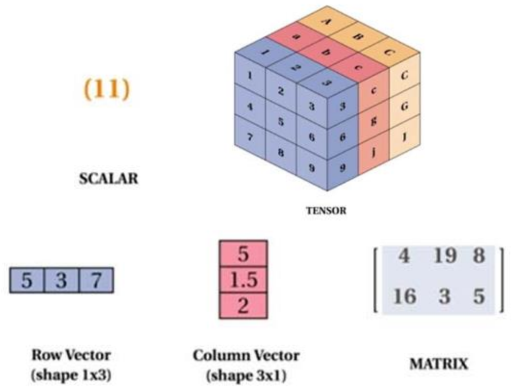
MSc. Artificial Intelligence

Representation learning

- Representation of data matters!



Raw data



Representation

Learning Task

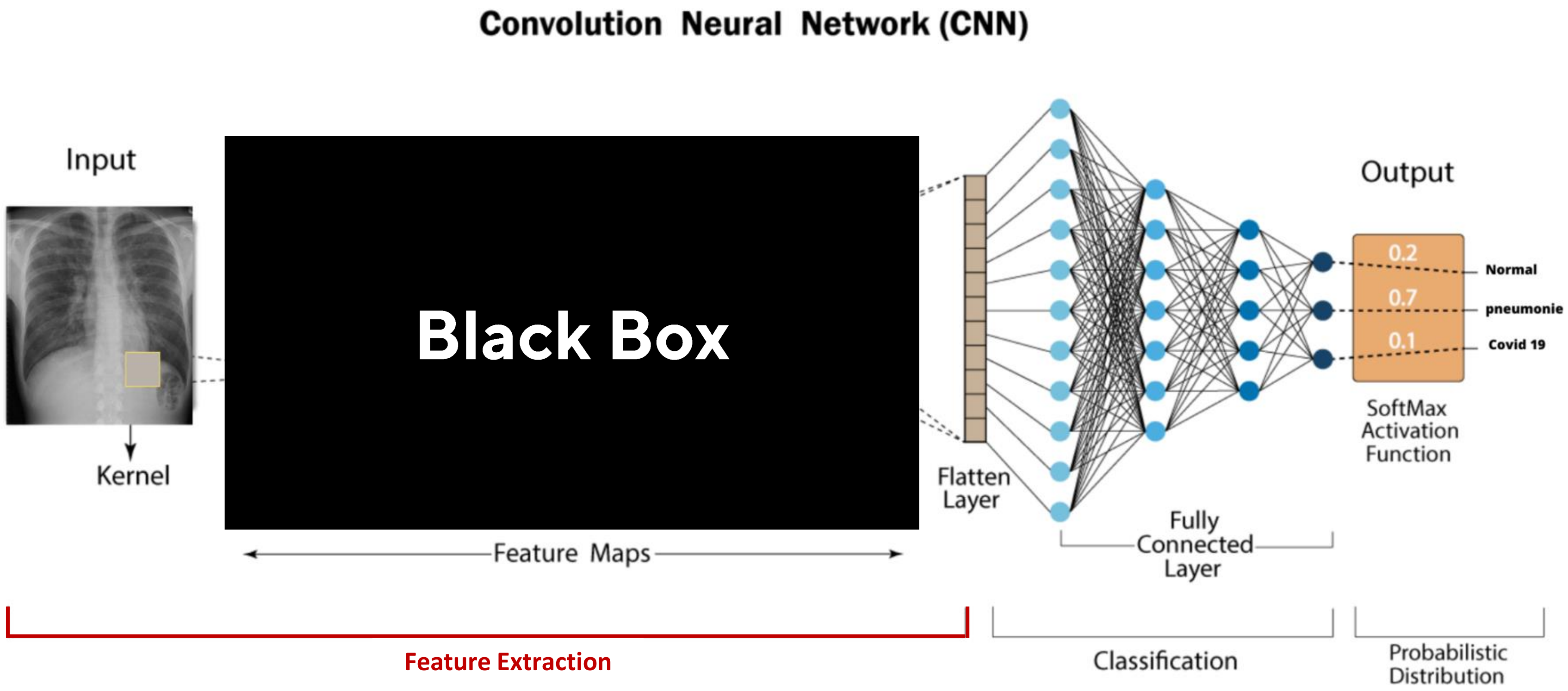
- Feature engineering had a key role in ML
  - Hand-crafted features (e.g., word co-occurrence, term frequency)

The child makes the dog happy.

$$x = \begin{array}{cl} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{array}{l} \text{dog} \\ \text{child} \\ \text{makes} \\ \text{happy} \end{array} \end{array}$$

Data driven features

- Deep neural network (DNN) as feature/representation learner:

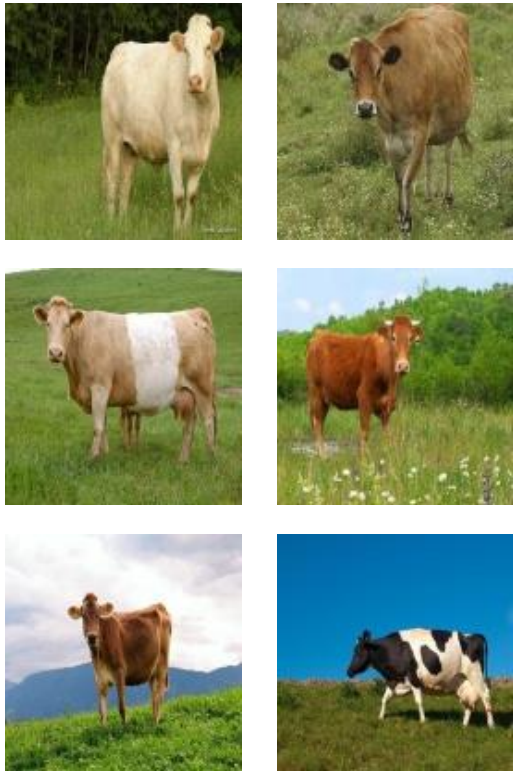


Data driven features

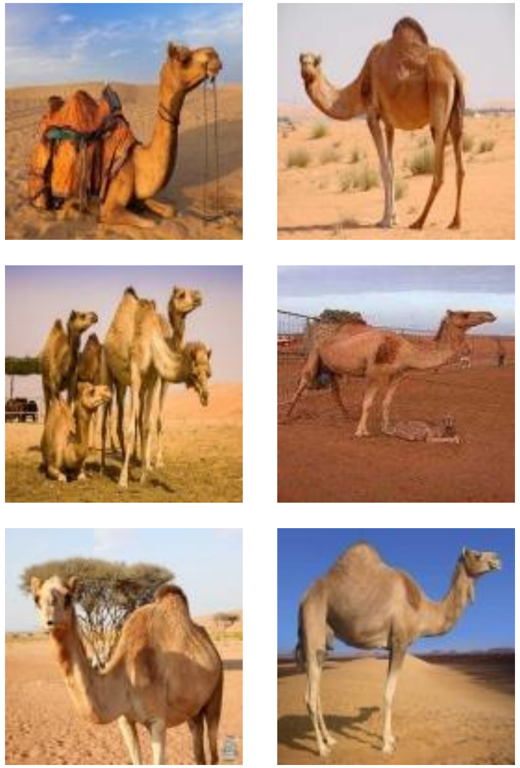
- Deep neural network (DNN) as feature/representation learner

Representation learning – Challenges

- **Training data for a DNN:**



Label: Cow



Label: Camel

Label: Polar Bear

Representation learning - Challenges

- **Testing model:**

✅ Predicted: Cow

❌ Predicted: Polar Bear

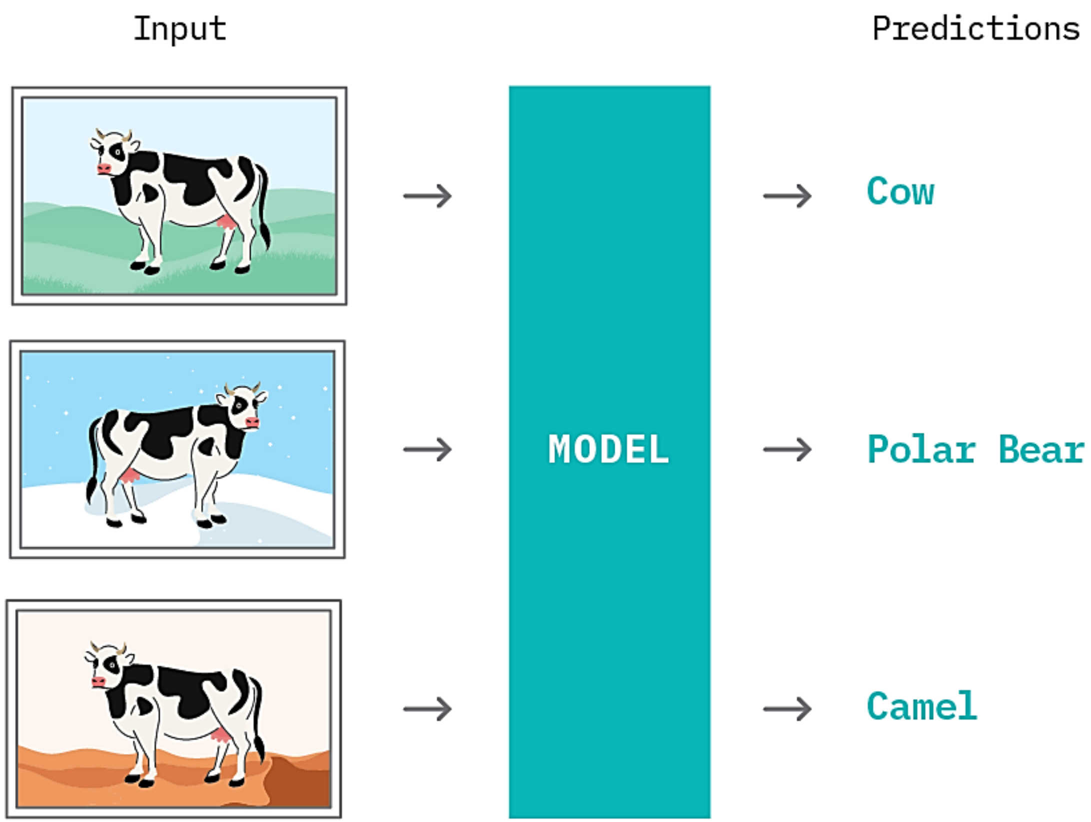
❌ Predicted: Camel

Representation learning - Challenges

- DNNs may not always find relevant representation

- **Challenges:**
  - Huge models, limited labels
  - Black-box nature of DNNs

- **What is a good representation?**



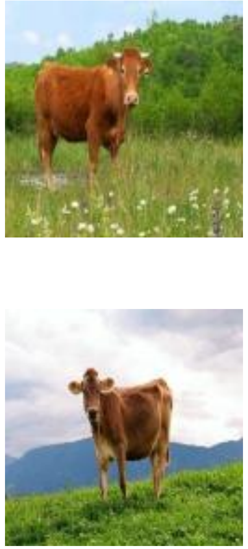
What is a good representation?

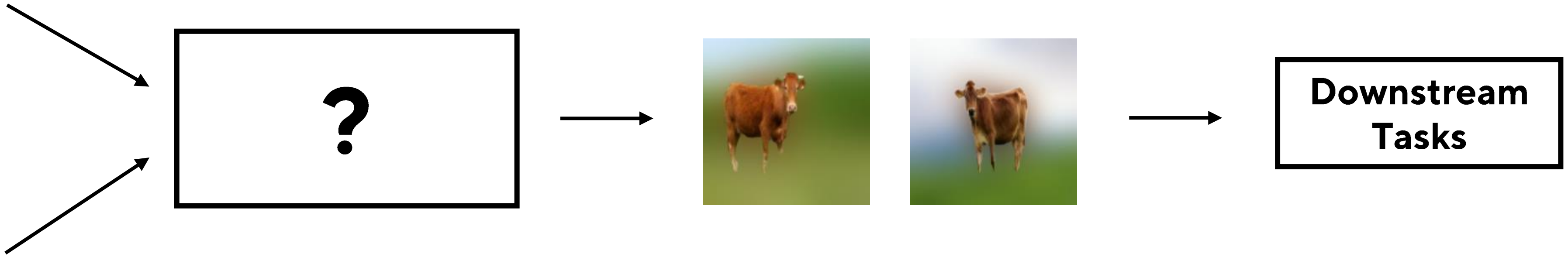
- **A good representation:**
  - ✓ Should be invariant across different scenes (views)
  - ✓ Should contain essential info, not the redundant info.



What is a good representation?

Here's an idea:

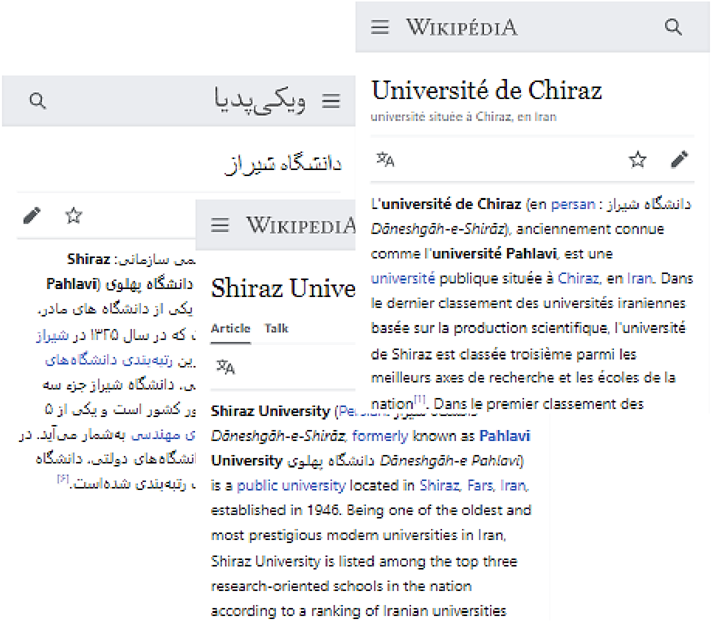




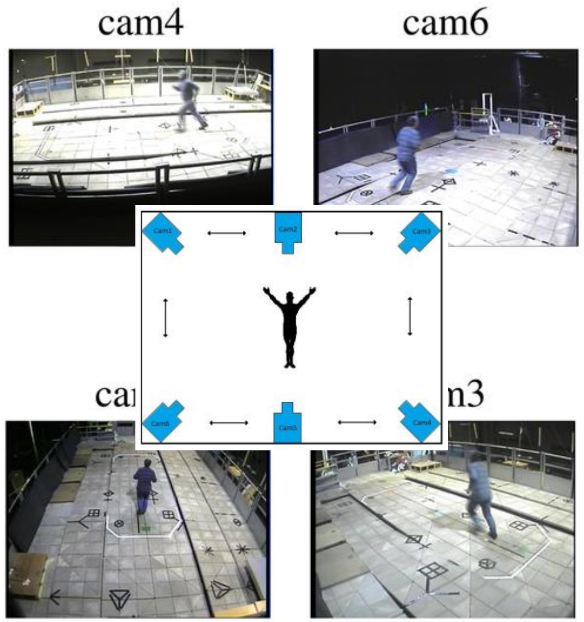
✅ **Multi-view representation learning model**

Multi-view data

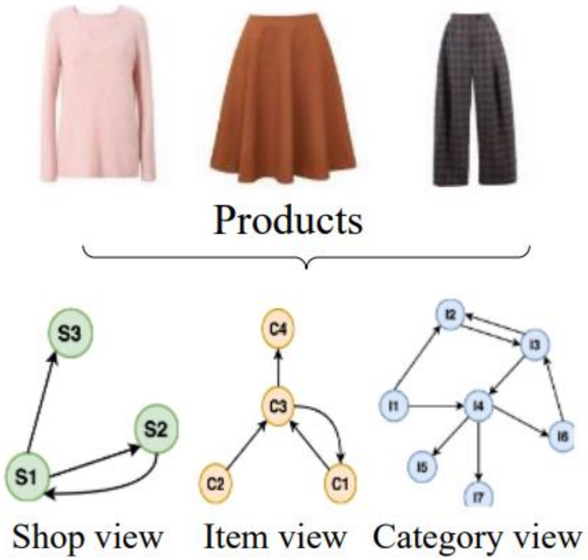
Natural multi-view data:



Multi language data



Data captured by multiple sensors (ex. Camera)



Multi-view shopping graphs

Multi-view data

Hand crafted multi-view data: **Augmentation**

Original

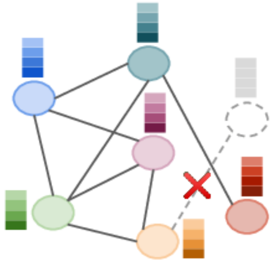
Rotate

Crop

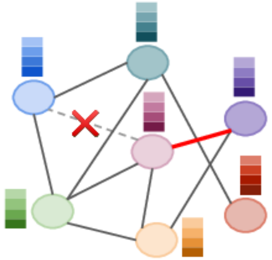
Crop & Rotate

Add Noise

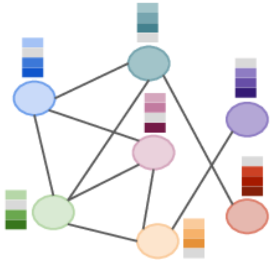
Original



Node Deletion



Edge Manipulation



Feature Manipulation

Sub Graph

Multi-view analysis - Formulation

Canonical Correlation Analysis (CCA)

- To understand the relationship between two sets of variable

$$(w_x^*, w_y^*) = \arg\max_{\mathrm{w}_x, \mathrm{w}_y} \mathrm{corr}(Xw_x, Yw_y)$$

$$\mathrm{corr}(Xw_x, Yw_y) = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x}\sqrt{\mathrm{w}_\mathrm{y}^\mathrm{T} \mathrm{C}_\mathrm{yy} \mathrm{w}_\mathrm{y}}}$$

Invariant to the scaling of w_x and w_y

- C_{xx} and C_{yy} are the covariance matrices of X and Y
- C_{xy} is the cross-covariance matrix between X and Y

Multi-view analysis - Formulation

Canonical Correlation Analysis (CCA)

- Constrained form:

$$\operatorname{corr}(Xw_x, Yw_y) = \underset{w_x, w_y}{\arg\max}\, w_x^T C_{xy} w_y$$
$$s.t.\;\; w_x^T C_{xx} w_x = I, \mathrm{w_y^T C_{yy} w_y} = I$$

- When the feature dimensionality is high, the covariance matrix C_{xx} (or $\mathrm{C_{yy}}$) is singular

$$C_{xx} = \frac{1}{N} XX^T + r_x I$$
$$C_{yy} = \frac{1}{N} YY^T + r_Y I$$

Multi-view analysis - Formulation

Canonical Correlation Analysis (CCA)

- **How to find w_x and w_y?**

1. Generalized eigenvalue decomposition problem:

$$\begin{bmatrix} \mathbf{0} & \Sigma_{xy} \\ \Sigma_{yx} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w_x} \\ \mathbf{w_y} \end{bmatrix} = \lambda \begin{bmatrix} \hat{\Sigma}_{xx} & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{w_x} \\ \mathbf{w_y} \end{bmatrix}$$

$$\Sigma_{xy} = \frac{1}{N} X Y^T$$
$$\Sigma_{yx} = \frac{1}{N} Y X^T$$

2. Preform singular value decomposition (SVD) on:

$$T = \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$$

$\mathrm{corr}(Xw_x, Yw_y) \rightarrow$ K largest singular values of T

Let W_x' and W_y' be the K largest left and right singular vectors of T

Multi-view analysis - Formulation

Canonical Correlation Analysis (CCA)

- **How to find w_x and w_y?**

2. Preform singular value decomposition (SVD) on:

$$T = \Sigma_{xx}^{-1/2}\Sigma_{xy}\Sigma_{yy}^{-1/2}$$

$\mathrm{corr}(Xw_x, Yw_y) \rightarrow$ K largest singular values of T

Let W_x' and W_y' be the K largest left and right singular vectors of T

Canonical matrices

$$W_x = \Sigma_{xx}^{-1/2}W_x' \qquad W_y = \Sigma_{yy}^{-1/2}W_y'$$

Canonical variables

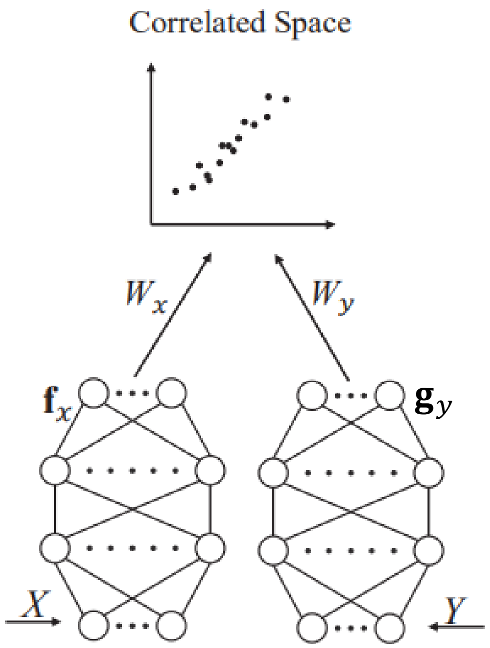
$$Z_x = W_x^T X \qquad Z_y = W_y^T Y$$

Multi-view analysis - Formulation

Deep Canonical Correlation Analysis (DCCA)

- Using DNNs as non-linear mappings.

$$\left(\theta_x^*, \theta_y^*\right) = \arg\max_{\theta_x, \theta_y} \operatorname{corr}\left(f(X; \theta_x), g\left(Y; \theta_y\right)\right)$$



- $f(\mathrm{X}; \theta_x)$ is a DNN that transforms X into a new representation, parameterized by θ_x
- $g\left(\mathrm{Y}; \theta_y\right)$ is another DNN that transforms Y into a new representation, parameterized by θ_y

Multi-view analysis - Formulation

Deep Canonical Correlation Analysis (DCCA)

$$T = \Sigma_{xx}^{-1/2}\Sigma_{xy}\Sigma_{yy}^{-1/2}$$

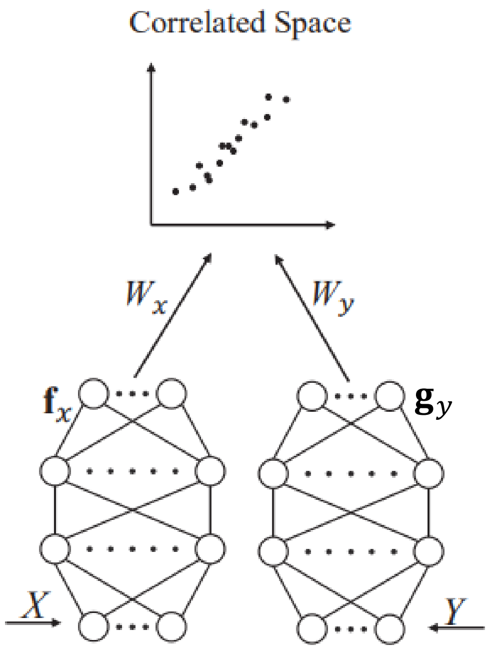
- $\mathrm{corr}\left(f(X;\theta_x), g(Y;\theta_y)\right) \rightarrow$ K largest singular values of T

$$T = \left(\frac{1}{N}f(X)f(X)^T + r_x I\right)^{-1/2}\left(\frac{1}{N}f(X)g(Y)^T\right)\left(\frac{1}{N}g(Y)g(Y)^T + r_y I\right)^{-1/2}$$

- CCA Loss function:

$$\underset{\theta_x,\theta_y,w_x,w_y}{\mathrm{maximize}} \sum_{k=1}^{K} \sigma_k(T)$$

$$s.t.\ \ w_x^T\left(\frac{1}{N}f(X)f(X)^T + r_x I\right)w_x = I, w_y^T\left(\frac{1}{N}g(Y)g(Y)^T + r_y I\right)w_y = I$$



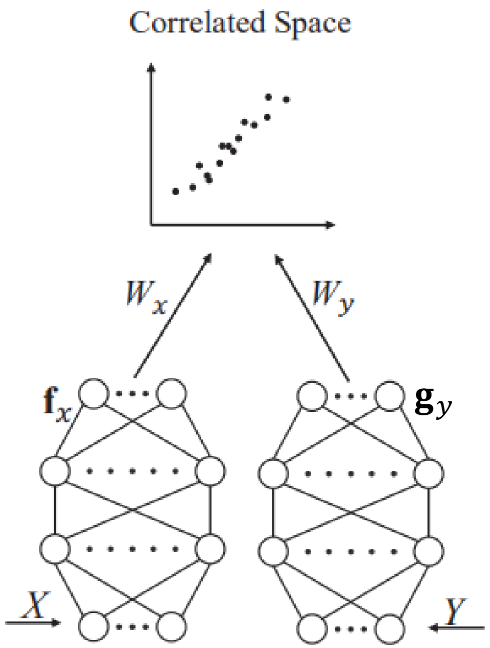
Multi-view analysis - Formulation

Deep Canonical Correlation Analysis (DCCA)

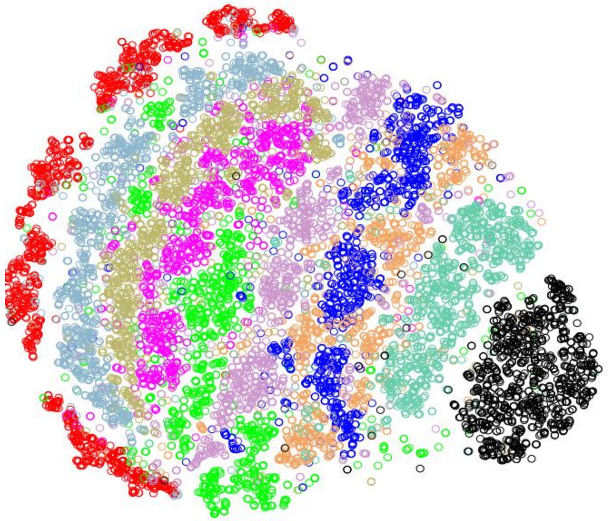
- If K = hidden dim, CCA Loss function:

$$\underset{\theta_x,\theta_y,w_x,w_y}{\text{maximize}}\ Tr(TT')^{1/2}$$

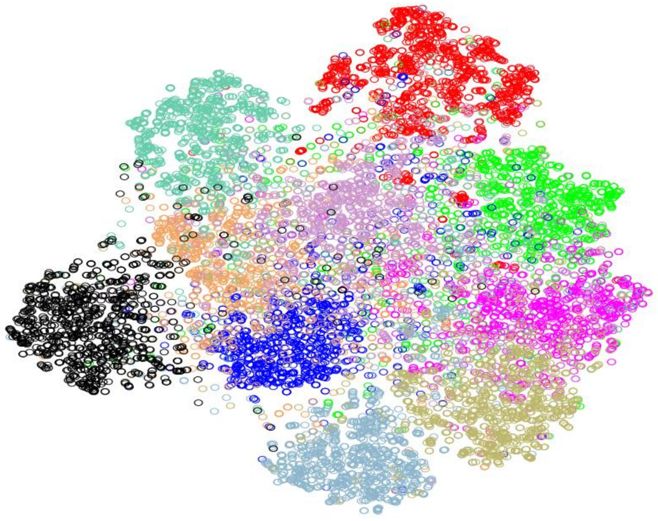
$$s.t.\ \ w_x^T\left(\frac{1}{N}f(X)f(X)^T + r_x I\right)w_x = I, w_y^T\left(\frac{1}{N}g(Y)g(Y)^T + r_y I\right)w_y = I$$



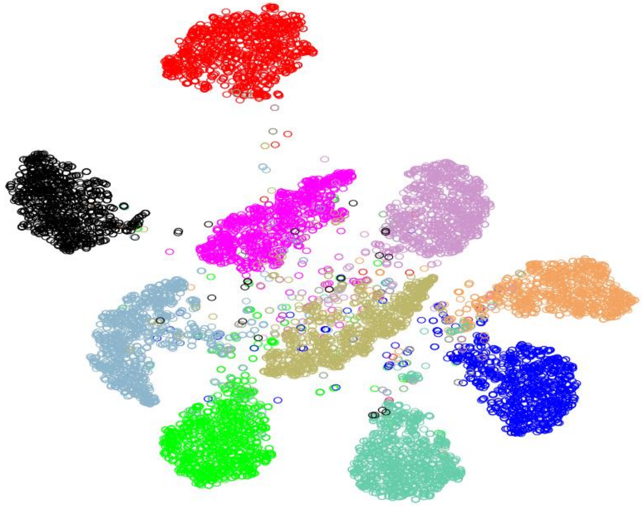
CCA vs. DCCA



t-SNE embedding



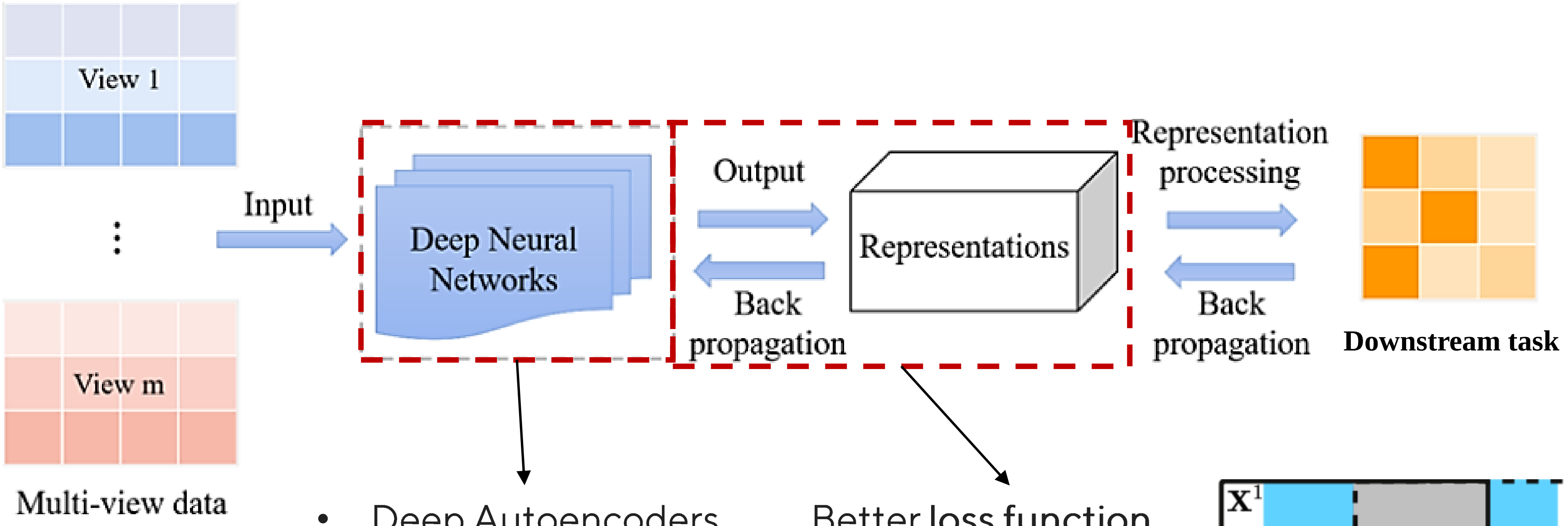
CCA



DCCA

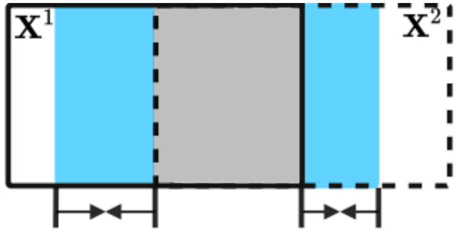
Deep multi-view learning

A general framework architecture:



- Deep Autoencoders
- GNN
  Transformer
- ...

Better loss function,
Better multi-view consistency!



Deep multi-view learning - DCCAE

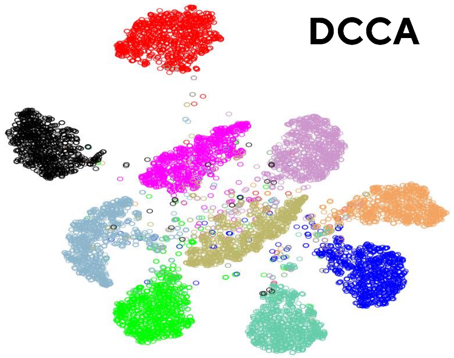
- Autoencoders are widely used in many applications!

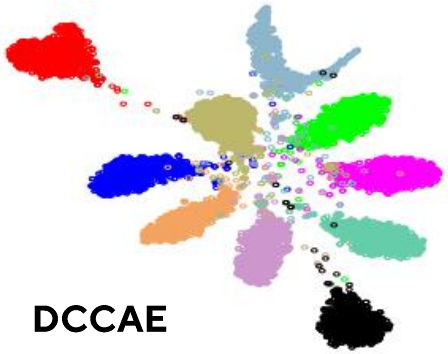
$$L_{\text{recon\_X}} = \|X - \mathrm{D_x}(\mathrm{E_x}(X;\theta_x);\phi_x)\|^2$$

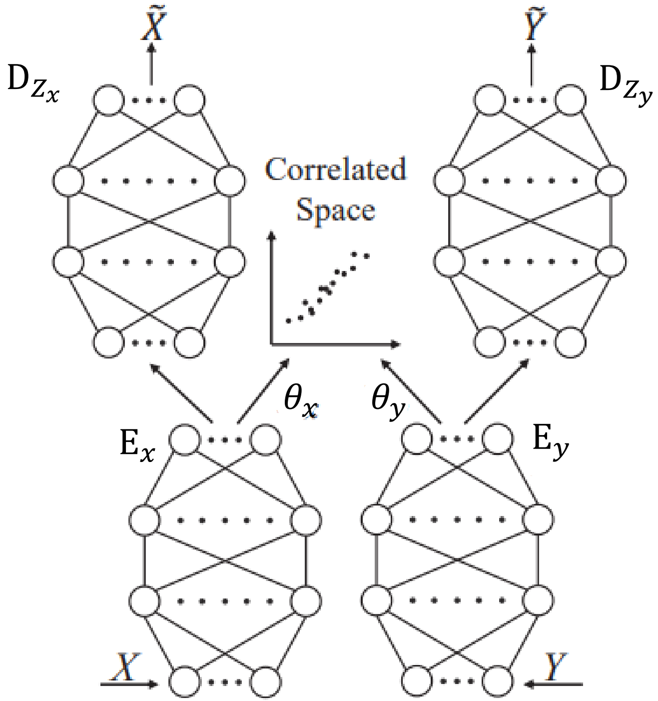
$$L_{\text{recon\_Y}} = \|Y - \mathrm{D_y}(\mathrm{E}_y(Y;\theta_y);\phi_y)\|^2$$

Minimize:

$$\mathcal{L}(\theta_x,\theta_y,\phi_x,\phi_y) = -\mathrm{corr}\left(\mathrm{E}(X;\theta_x),\mathrm{E}(Y;\theta_y)\right) + \lambda(L_{\text{recon\_Y}} + L_{\text{recon\_x}})$$

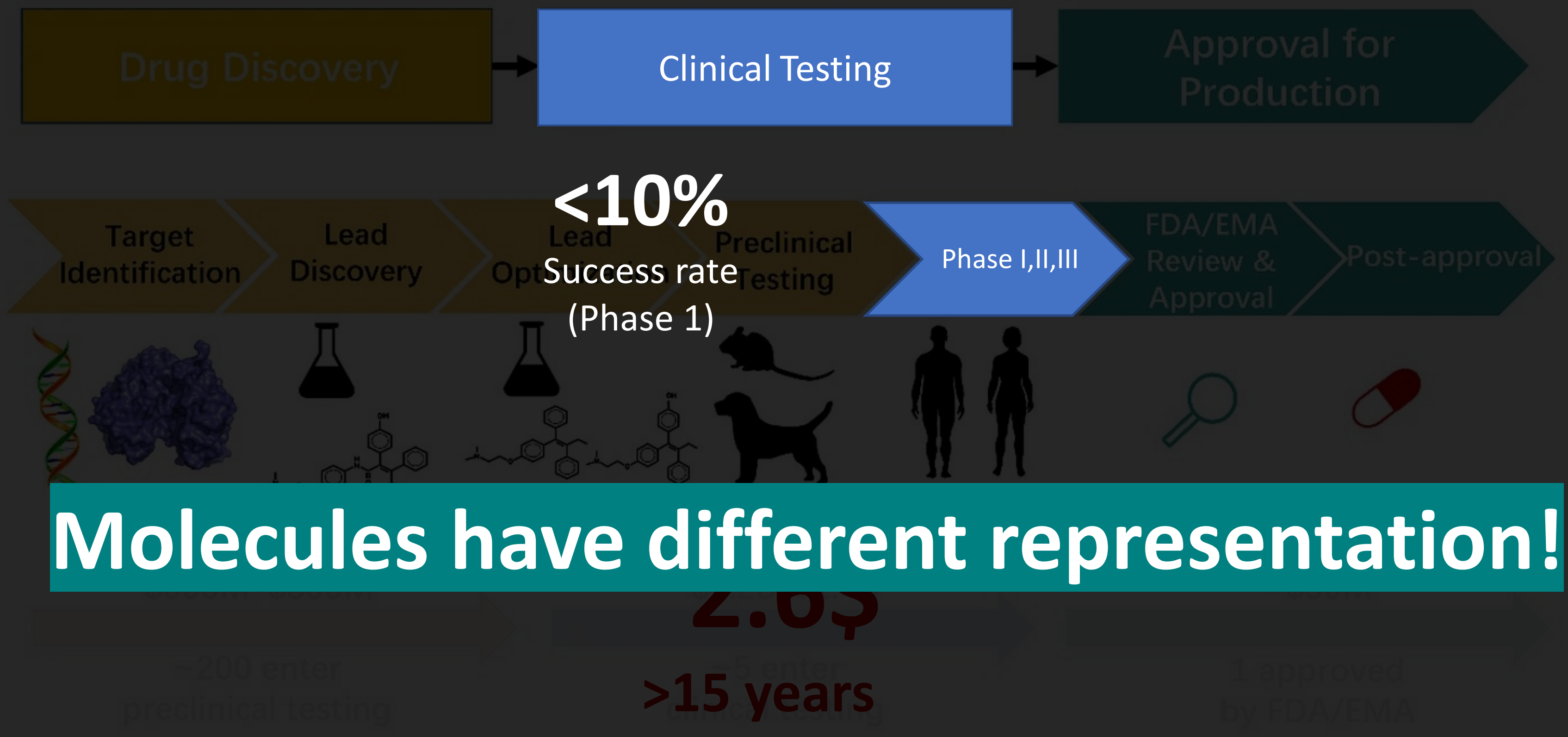






Deep multi-view learning in drug discovery

Drug discovery is an expensive, time-consuming process, with low success rates



Deep multi-view learning in drug discovery

Multi-view data of molecules/compounds

(Textual)

(a) Sequence-based

SMILES	CN1C=NC2=C1C(=O)N(C(=O)N2C)C
InChI	1S/C8H10N4O2/c1-10-4-9-6-...3H3
SELFIES	[C][N][C][=Branch1][C][=O][C][=C]...[N][=O]
Morgan	[000000...000000001001000...000]
MACCS	[000000...11100010101111111110]

IUPAC 1,3,7-trimethylpurine-2,6-dione

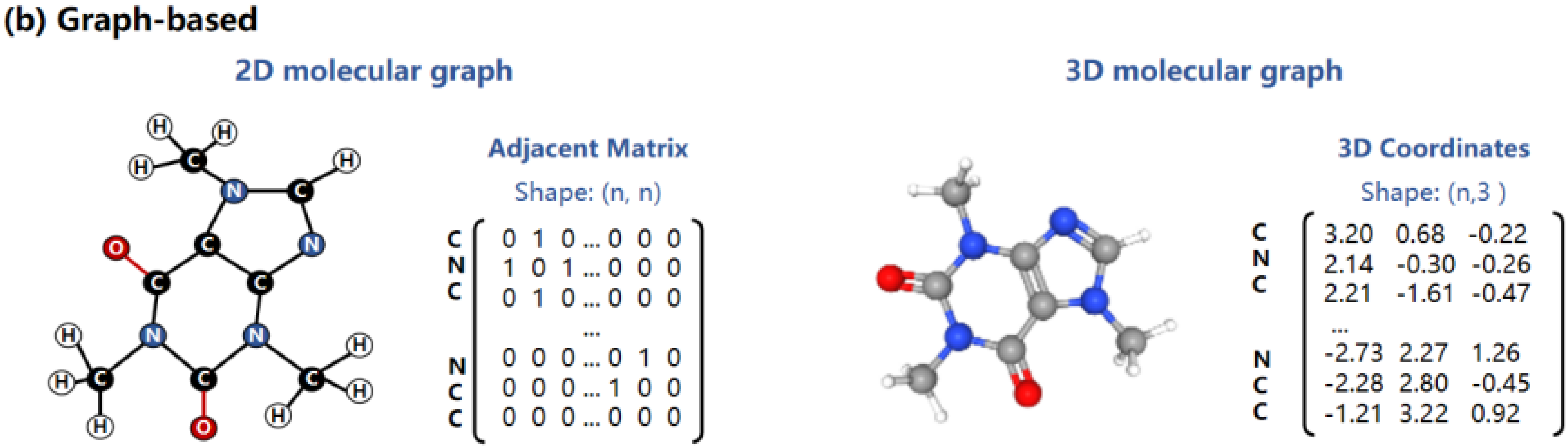
caption

Caffeine is a trimethylxanthine in which the three methyl groups are located at positions 1, 3, and 7. A purine alkaloid that occurs naturally in tea and coffee.

Deep multi-view learning in drug discovery

Multi-view data of molecules/compounds

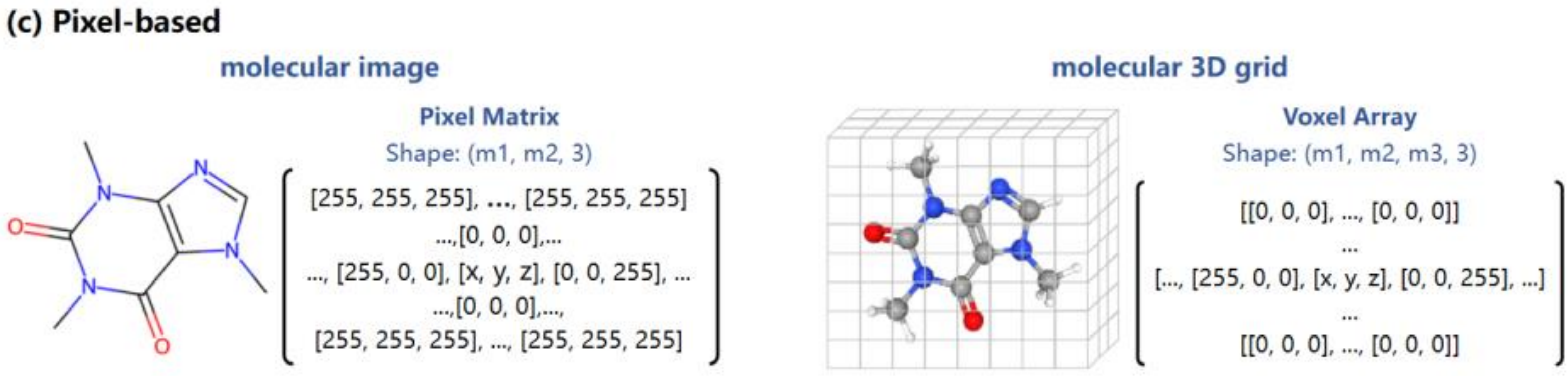
(Graph)



Deep multi-view learning in drug discovery

Multi-view data of molecules/compounds

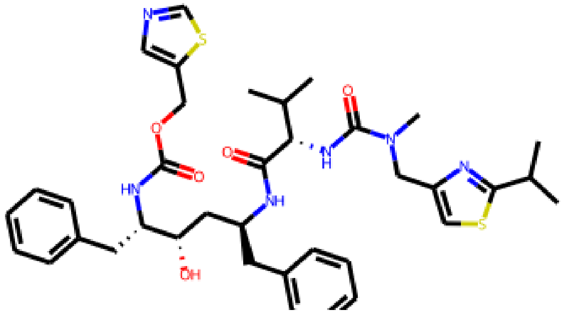
(Image)



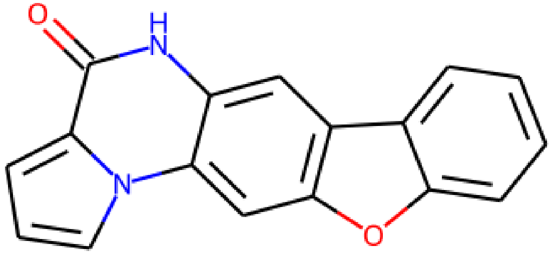
Deep multi-view learning in drug discovery

Dual-view Molecular Pre-training (DVMP)

ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023

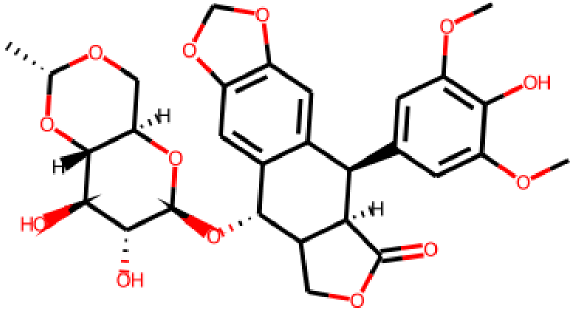


Different molecular representations describe molecules from different aspects!



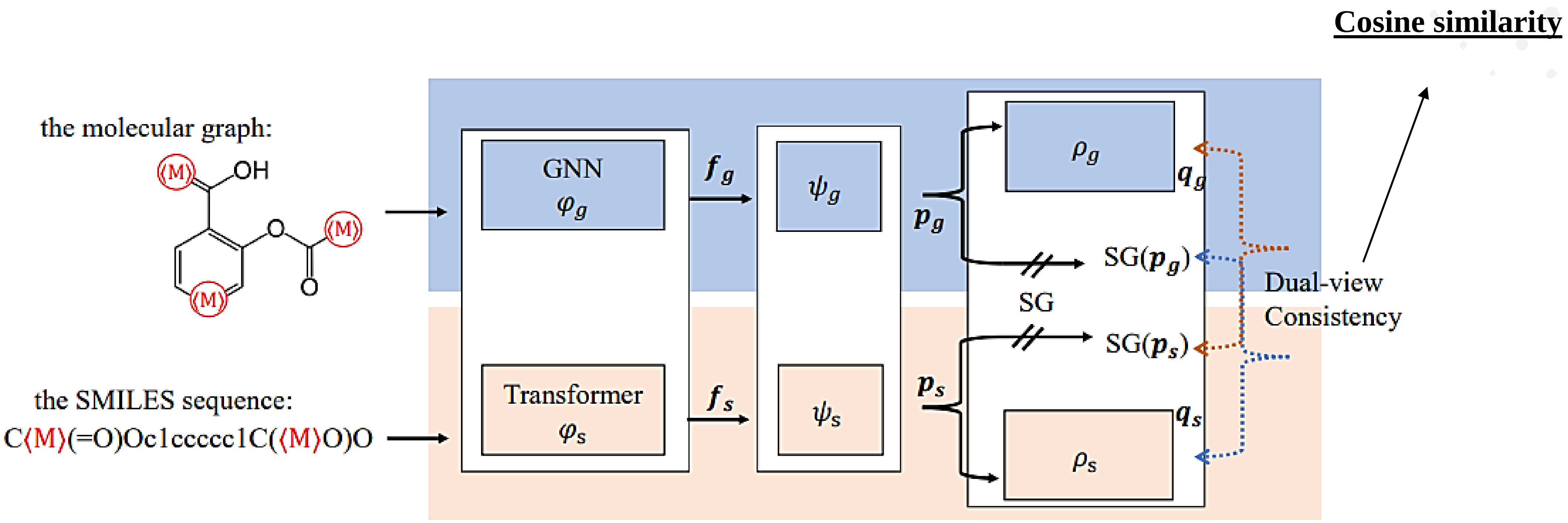
Transformer ❎ GNN ✅

✅ DVMP Succeeds!



Deep multi-view learning in drug discovery

Dual-view Molecular Pre-training (DVMP)



Thank you!

Feel free to ask any question!

Presentation materials

Abbas Mehrbaniyan

@mehrbanian

abbas.mrbn@gmail.com

پژوهشکده هوش مصنوعی
دانشگاه شیراز

دانشگاه شیراز