

# سری کارگاه‌های یادگیری عمیق

یادگیری چند نمایی و کاربرد آن در کشف دارو



عباس مهربانیان

دانشجو ارشد هوش مصنوعی



# Deep Learning Workshops

Multi-view learning and its applications in drug discovery



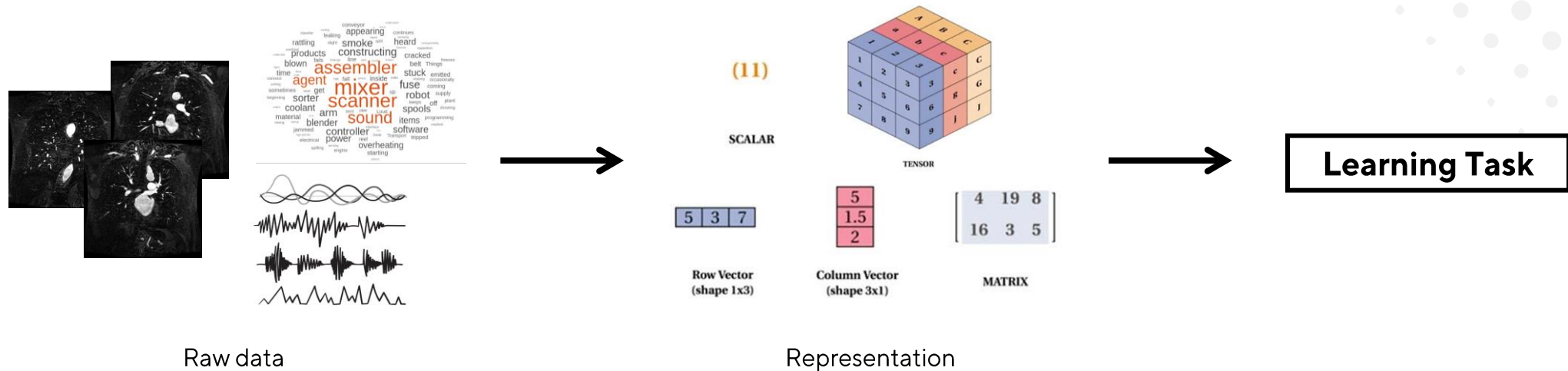
Abbas Mehrbaniyan

MSc. Artificial Intelligence



# Representation learning

- Representation of data matters!

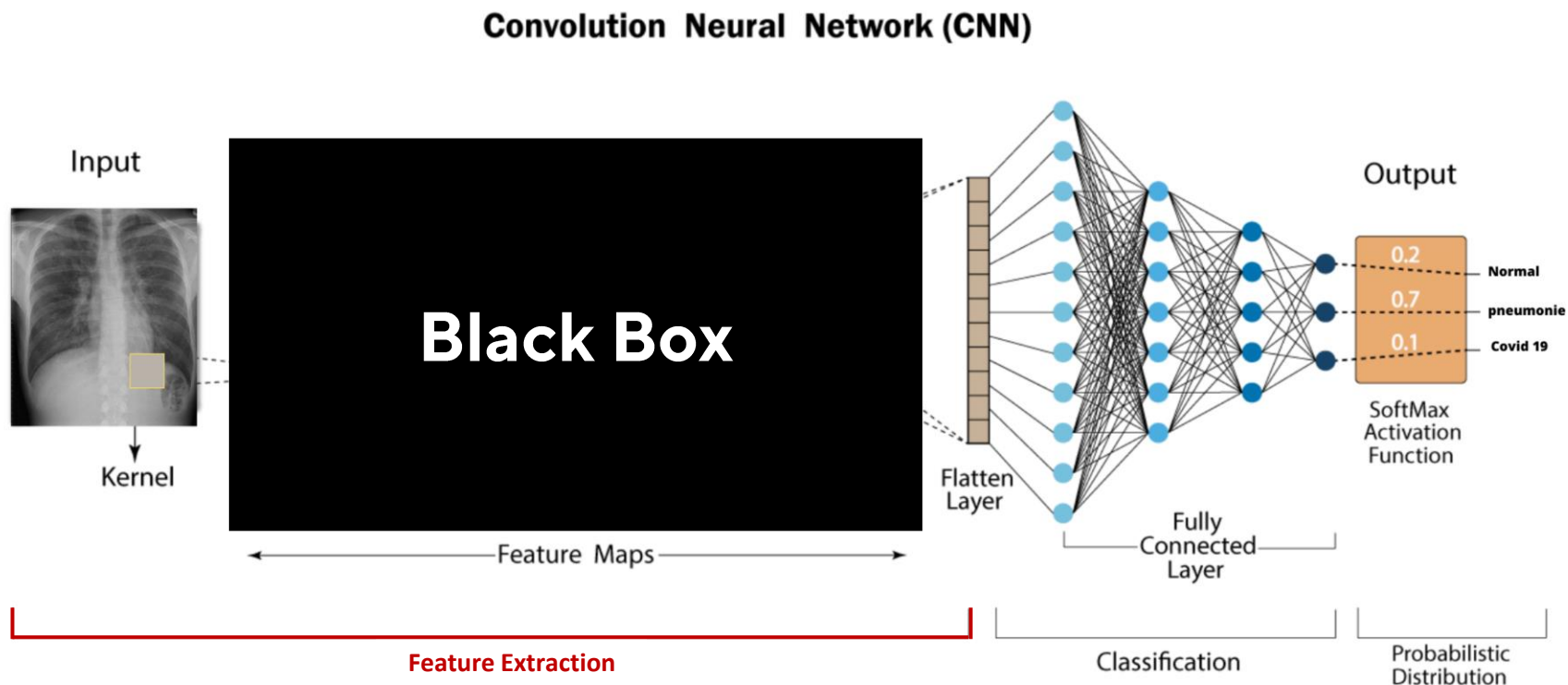


- Feature engineering had a **key role** in ML
  - Hand-crafted features (e.g., word co-occurrence, term frequency)



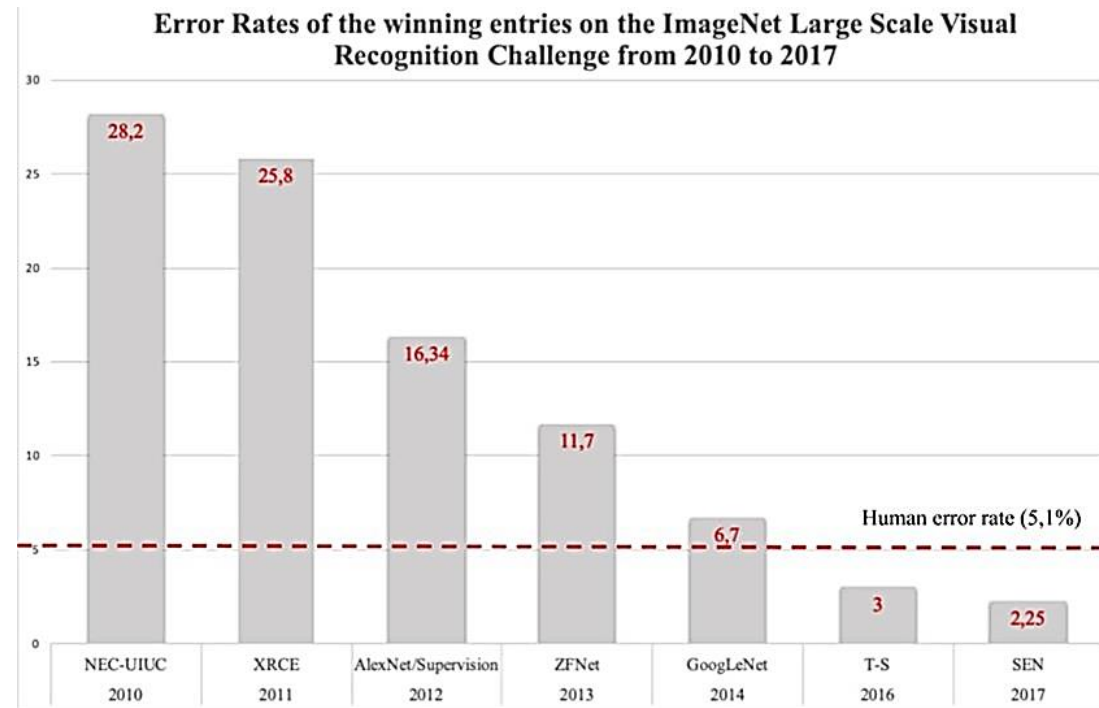
# Data driven features

- 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



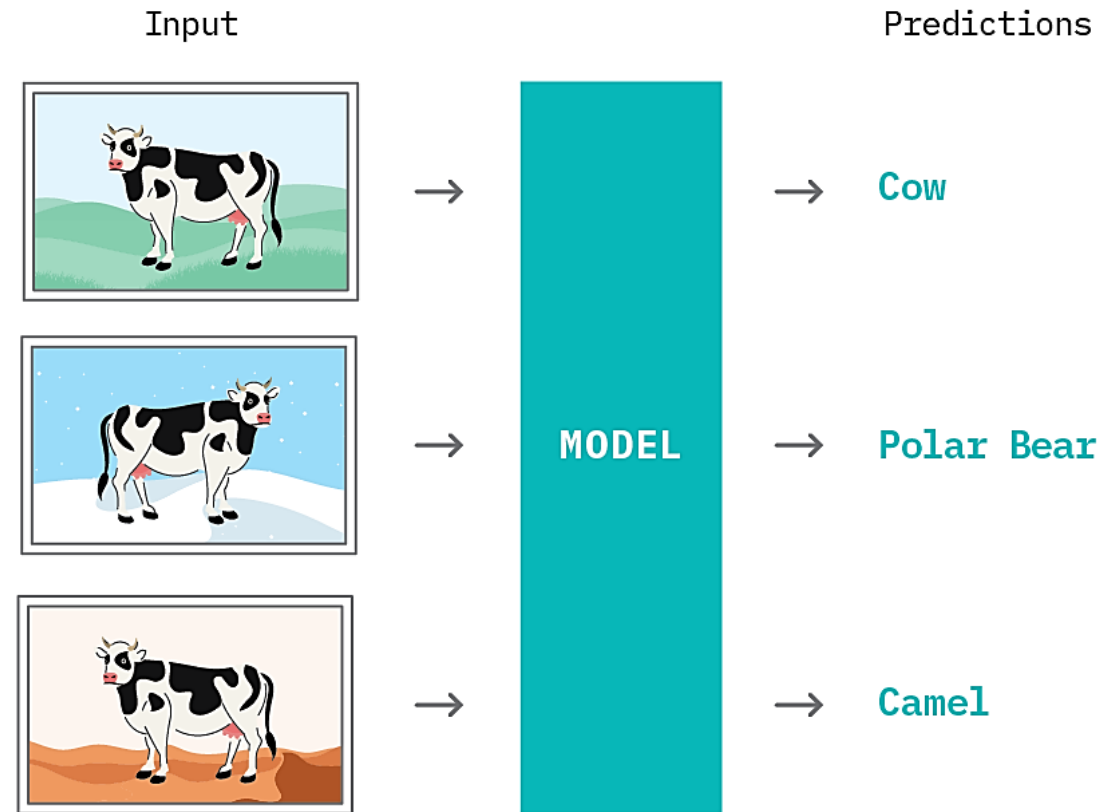
✗ Predicted: Polar Bear



✗ 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?**

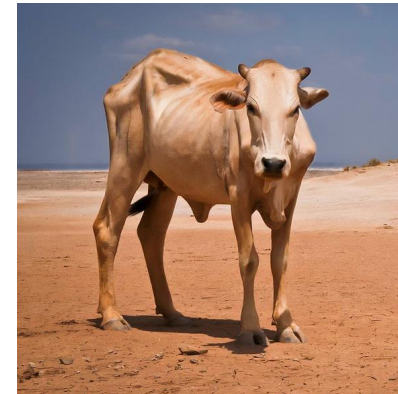




# What is a good representation?

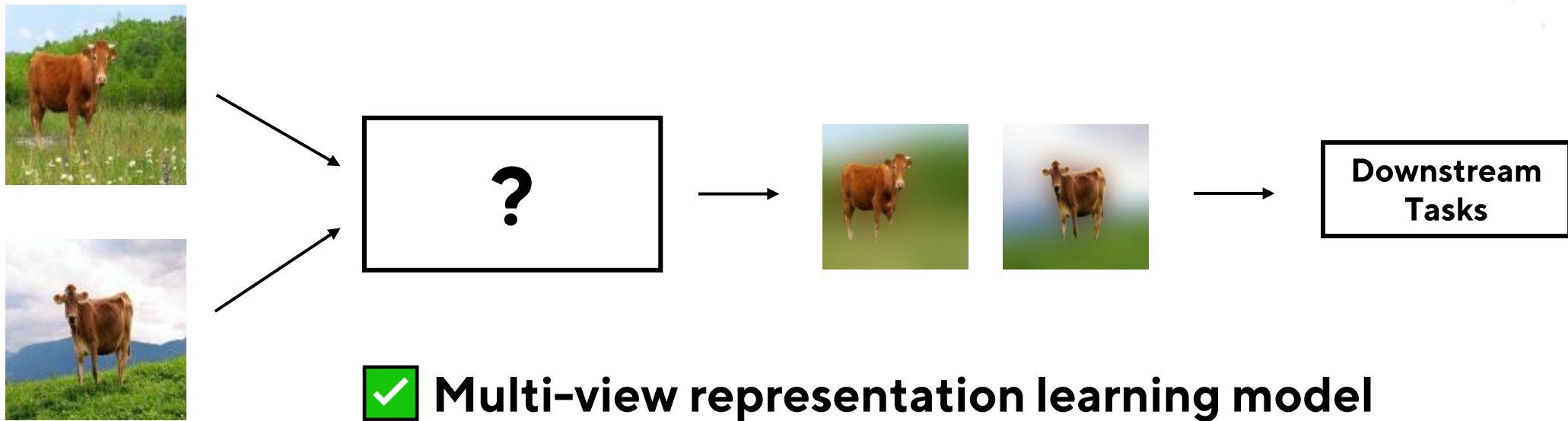
- **A good representation:**

- ✓ 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 data

Natural multi-view data:

WIKIPÉDIA

دانشگاه شیراز

Shiraz Unive

Shiraz University (Pahlavi: دانشگاه شیراز, formerly known as Pahlavi University) is a public university located in Shiraz, Fars, Iran, established in 1946. Being one of the oldest and most prestigious modern universities in Iran, Shiraz University is listed among the top three research-oriented schools in the nation according to a ranking of Iranian universities

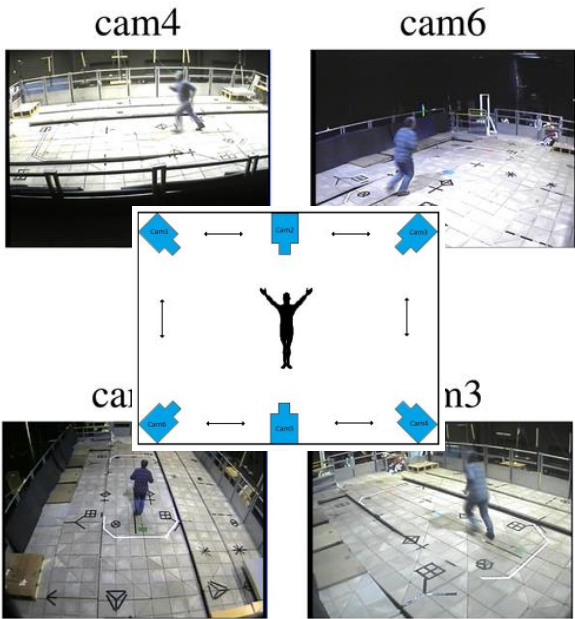
WIKIPÉDIA

دانشگاه شیراز

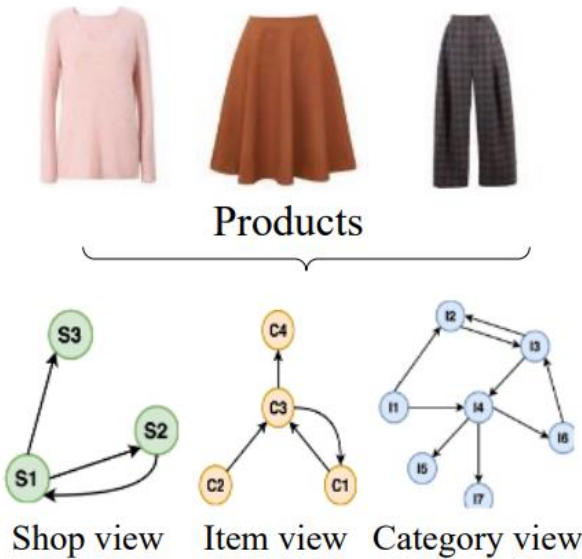
Shiraz Unive

Shiraz University (Pahlavi: دانشگاه شیراز, formerly known as Pahlavi University) is a public university located in Shiraz, Fars, Iran, established in 1946. Being one of the oldest and most prestigious modern universities in Iran, Shiraz University is listed among the top three research-oriented schools in the nation according to a ranking of Iranian universities

Multi language data



Data captured by multiple sensors (ex. Camera)



Multi-view shopping graphs



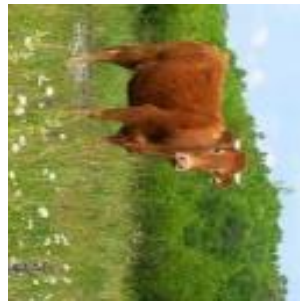
# Multi-view data

Hand crafted multi-view data:

**Augmentation**



Original



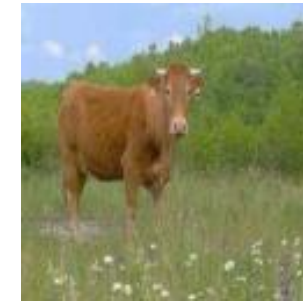
Rotate



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

$$(w_x^*, w_y^*) = \arg \max_{w_x, w_y} \text{corr}(Xw_x, Yw_y)$$

$$\text{corr}(Xw_x, Yw_y) = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x} \sqrt{w_y^T C_{yy} w_y}}$$

→ 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:

$$\begin{aligned} \text{corr}(Xw_x, Yw_y) &= \arg \max_{w_x, w_y} w_x^T C_{xy} w_y \\ \text{s.t. } w_x^T C_{xx} w_x &= I, w_y^T C_{yy} w_y = I \end{aligned}$$

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

$$\begin{aligned} C_{xx} &= \frac{1}{N} XX^T + r_x I \\ C_{yy} &= \frac{1}{N} YY^T + r_y I \end{aligned}$$

# Multi-view analysis - Formulation

## Canonical Correlation Analysis (CCA)

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

1. Generalized eigenvalue decomposition problem:

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

$$\Sigma_{xy} = \frac{1}{N}XY^T$$
$$\Sigma_{yx} = \frac{1}{N}YX^T$$

2. Perform singular value decomposition (SVD) on:

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

$\text{corr}(Xw_x, Yw_y) \rightarrow K$  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. Perform singular value decomposition (SVD) on:

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

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

Let  $W'_x$  and  $W'_y$  be the  $K$  largest left and right singular vectors of  $T$

### Canonical matrices

$$W_x = \Sigma_{xx}^{-1/2} W'_x \quad W_y = \Sigma_{yy}^{-1/2} W'_y$$

### Canonical variables

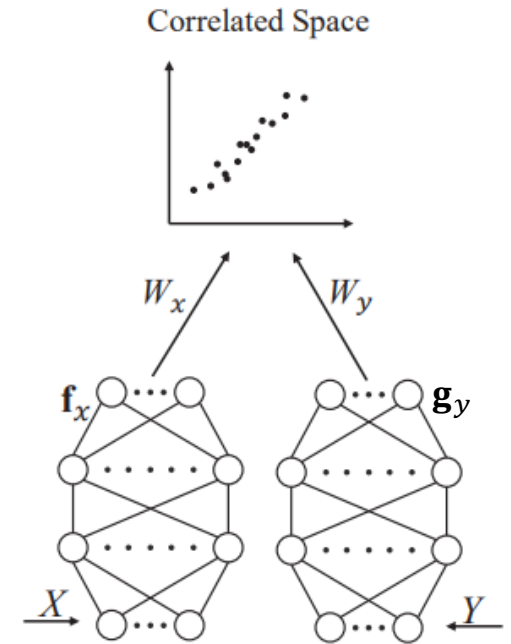
$$Z_x = W_x^T X \quad Z_y = W_y^T Y$$

# Multi-view analysis - Formulation

## Deep Canonical Correlation Analysis (DCCA)

- Using DNNs as non-linear mappings.

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



- $f(X; \theta_x)$  is a DNN that transforms  $X$  into a new representation, parameterized by  $\theta_x$
- $g(Y; \theta_y)$  is another DNN that transforms  $Y$  into a new representation, parameterized by  $\theta_y$

# Multi-view analysis - Formulation

## Deep Canonical Correlation Analysis (DCCA)

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

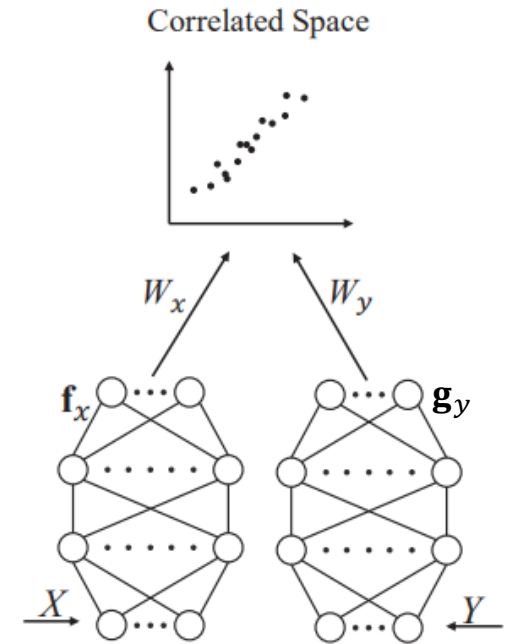
- $\text{corr}(f(X; \theta_x), g(Y; \theta_y)) \rightarrow K$  largest singular values of  $T$

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

- CCA Loss function:

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

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





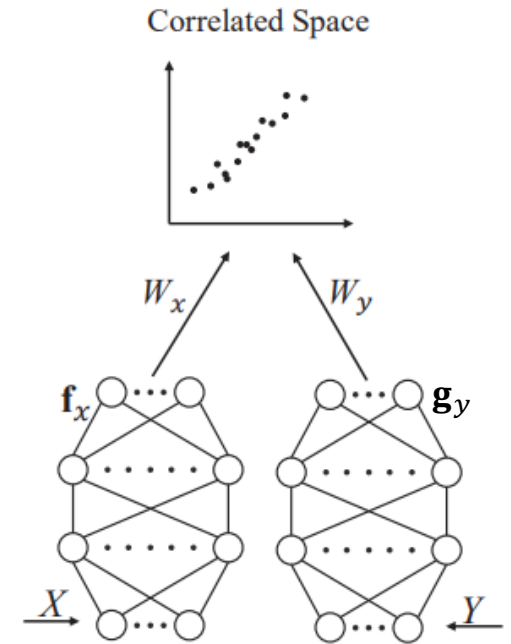
# Multi-view analysis - Formulation

## Deep Canonical Correlation Analysis (DCCA)

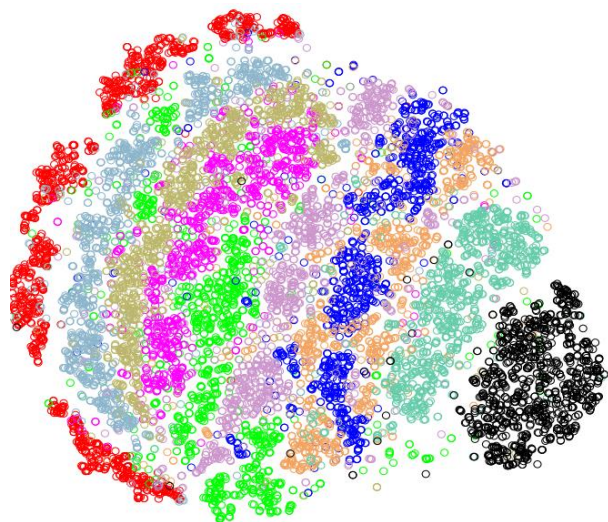
- If  $K$  = hidden dim, CCA Loss function:

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

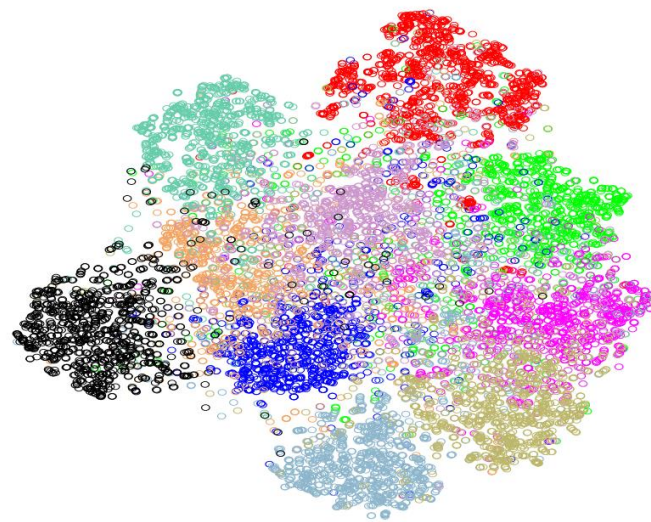
$$\text{s.t. } w_x^T \left( \frac{1}{N} f(X)f(X)^T + r_x I \right) w_x = I, w_y^T \left( \frac{1}{N} f(Y)f(Y)^T + r_y I \right) w_y = I$$



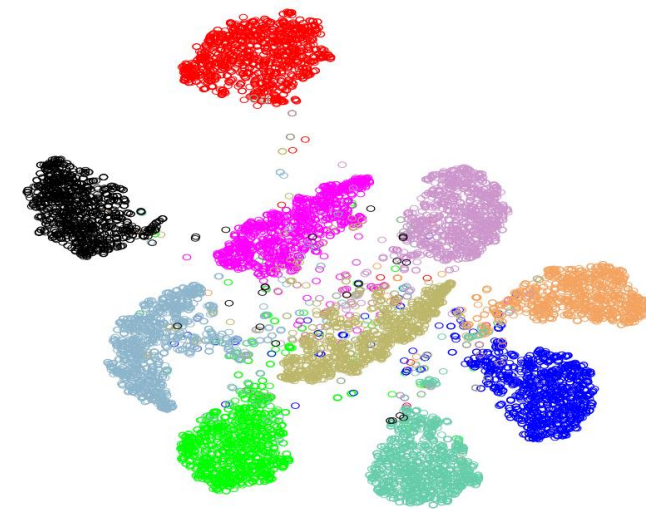
# CCA vs. DCCA



t-SNE embedding



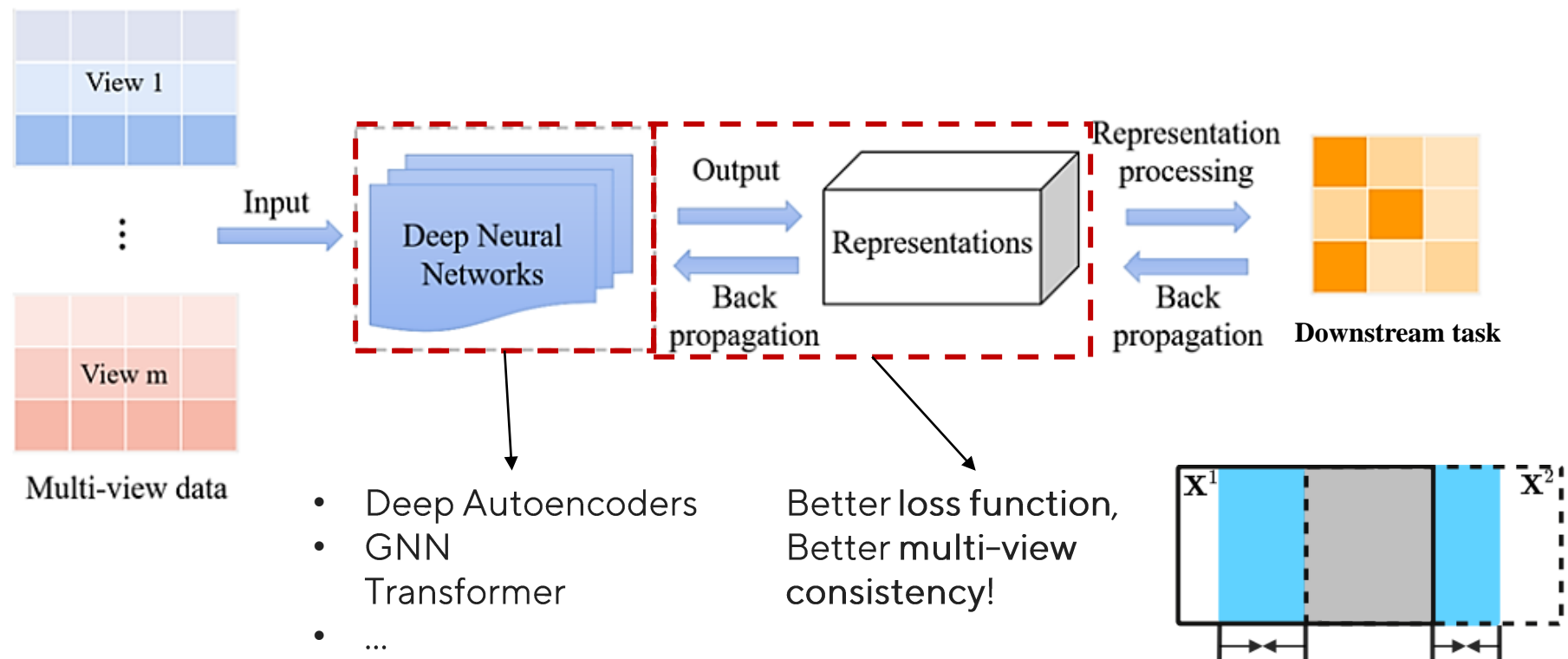
CCA



DCCA

# Deep multi-view learning

A general framework architecture:



# Deep multi-view learning - DCCAE

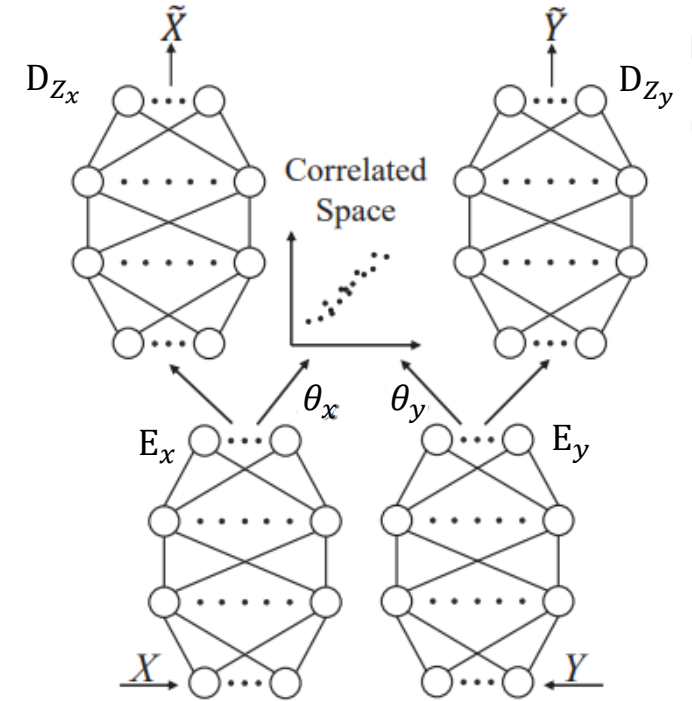
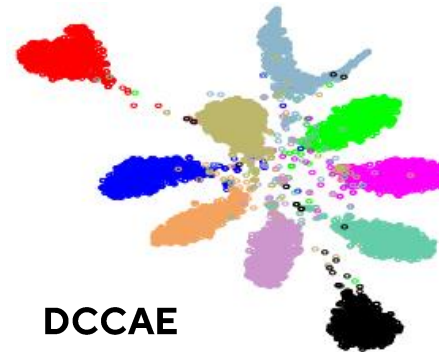
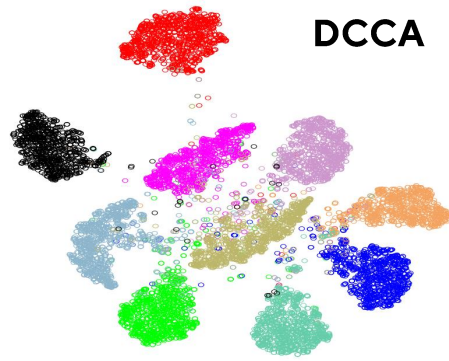
- Autoencoders are widely used in many applications!

$$L_{\text{recon}_X} = \|X - D(E(X; \theta_x); \phi_x)\|^2$$

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

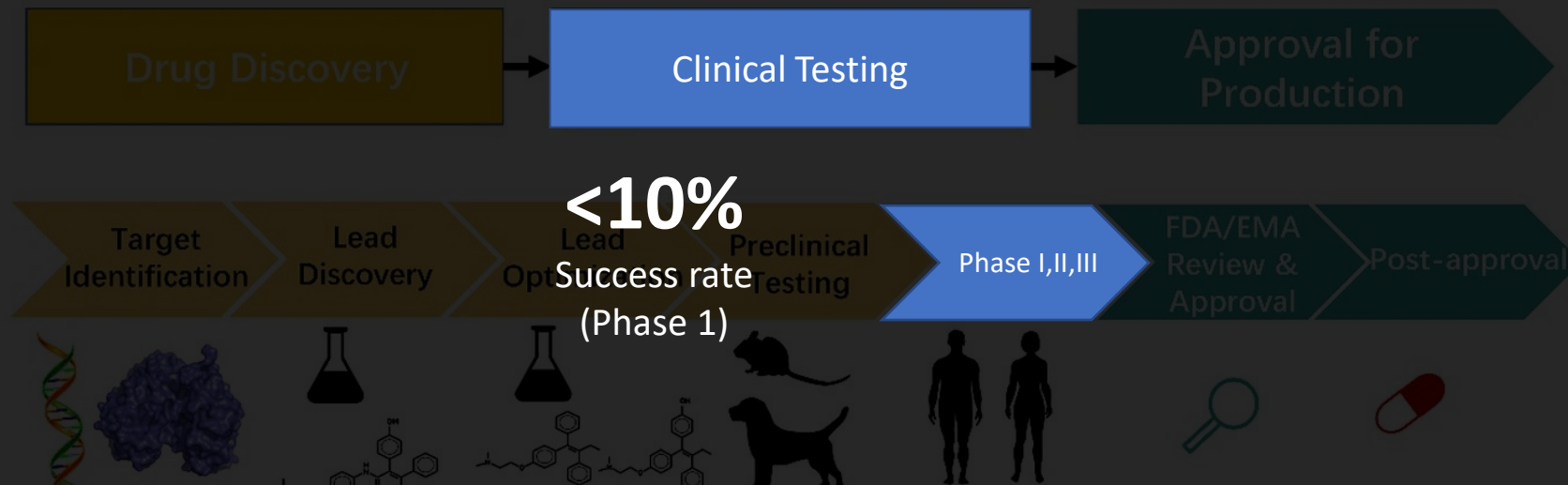
Minimize:

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



# Deep multi-view learning in drug discovery

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



**Molecules have different representation!**

~200 enter  
preclinical testing

2.0\$  
5 enter  
>15 years

1 approved  
by FDA/EMA



# Deep multi-view learning in drug discovery

Multi-view data of molecules/compounds

(Textual)

## (a) Sequence-based

SMILES	<chem>CN1C=NC2=C1C(=O)N(C(=O)N2C)C</chem>
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...1110001010111111110]

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

### caption

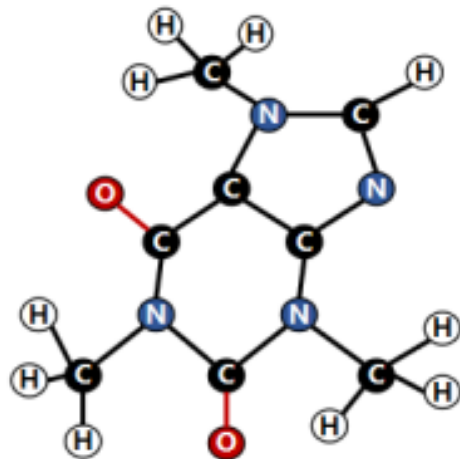
Caffeine is a trimethylxanthine in which the three methyl groups are located at positions 1, 3, and 7. A purine alkaloid that occurs naturally in tea and coffee.

# Deep multi-view learning in drug discovery

Multi-view data of molecules/compounds  
(Graph)

## (b) Graph-based

2D molecular graph

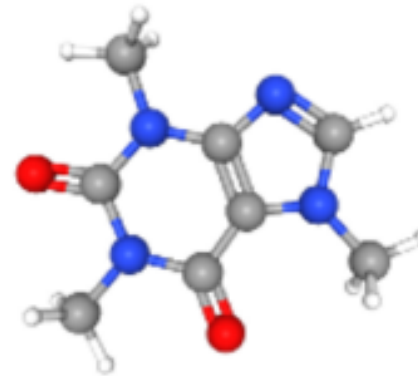


Adjacent Matrix

Shape: (n, n)

$$\begin{matrix} \text{C} \\ \text{N} \\ \text{C} \\ \text{C} \\ \text{N} \\ \text{C} \\ \text{C} \end{matrix} \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ & & & \dots & & & \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

3D molecular graph



3D Coordinates

Shape: (n, 3)

$$\begin{matrix} \text{C} \\ \text{N} \\ \text{C} \\ \text{C} \\ \text{N} \\ \text{C} \\ \text{C} \end{matrix} \begin{pmatrix} 3.20 & 0.68 & -0.22 \\ 2.14 & -0.30 & -0.26 \\ 2.21 & -1.61 & -0.47 \\ \dots & \dots & \dots \\ -2.73 & 2.27 & 1.26 \\ -2.28 & 2.80 & -0.45 \\ -1.21 & 3.22 & 0.92 \end{pmatrix}$$

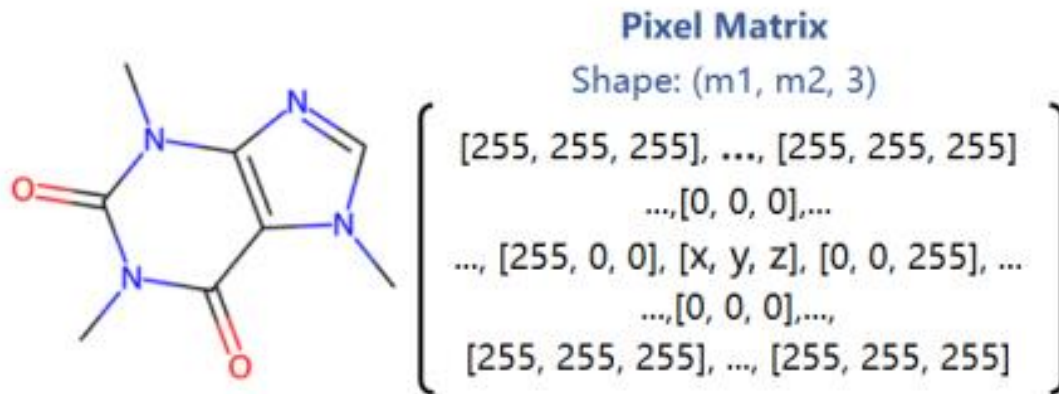
# Deep multi-view learning in drug discovery

Multi-view data of molecules/compounds

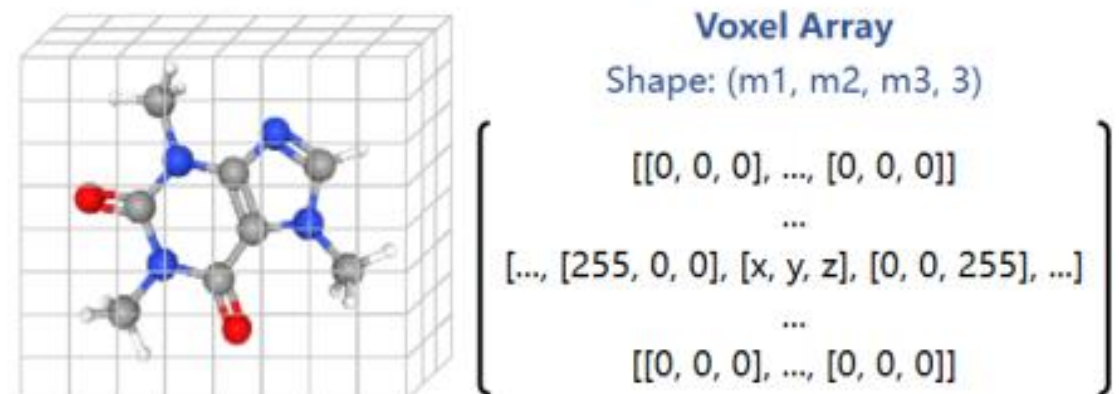
(Image)

## (c) Pixel-based

molecular image



molecular 3D grid



# Deep multi-view learning in drug discovery

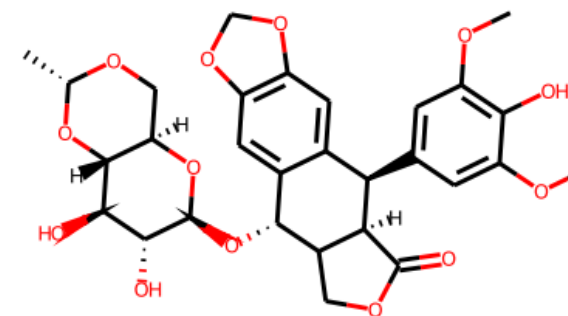
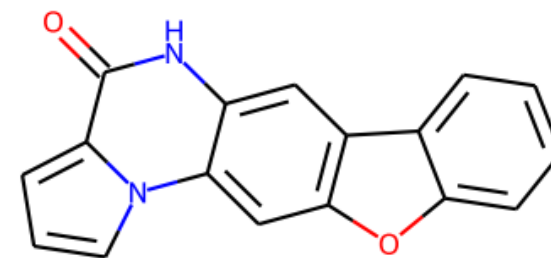
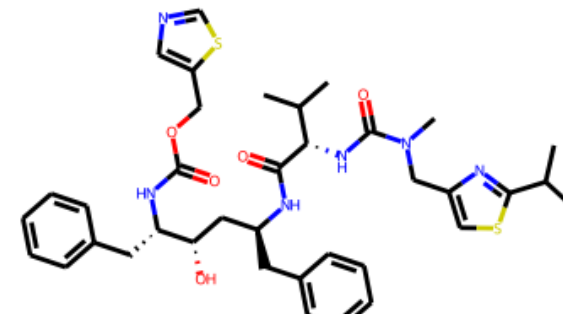
## Dual-view Molecular Pre-training (DVMP)

ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023

Different molecular representations describe molecules from different aspects!

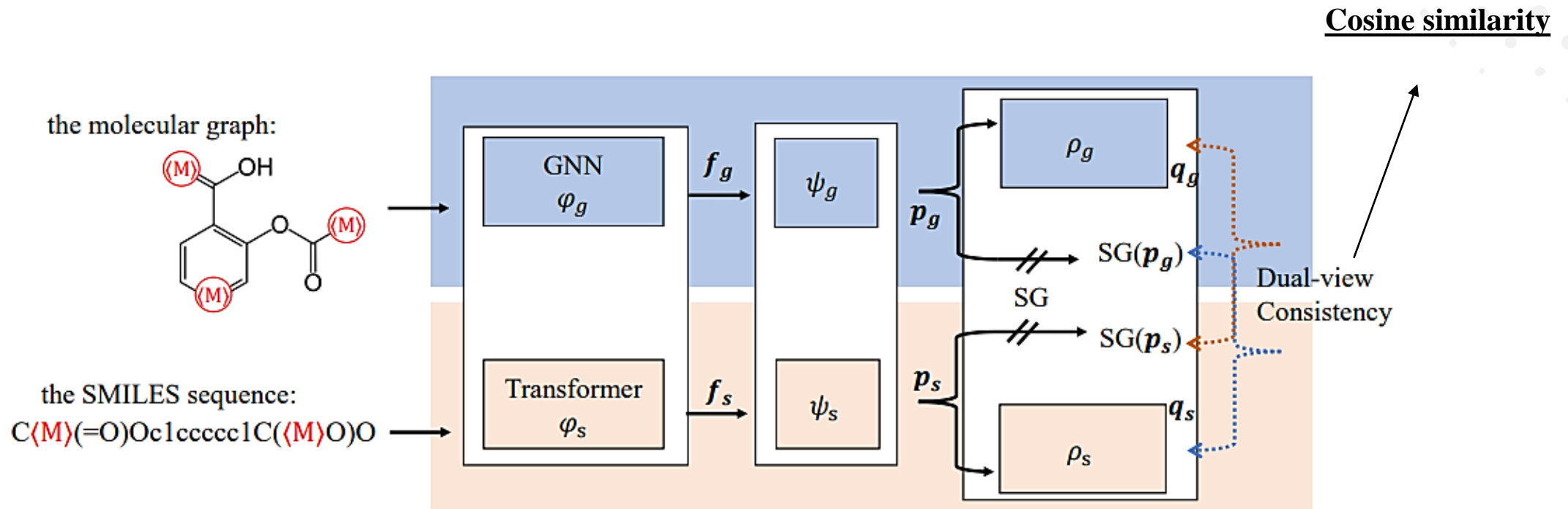
Transformer ❌ GNN ✅

✅ 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

