COPER: Correlation-based Permutations for Multi-View Clustering

ICLR 2025

Ran Eisenberg* Jonathan Svirsky* Ofir Lindenbaum

Faculty of Engineering, Bar-Ilan University

IMVC Vision Conference

Multi-view Learning: Setting

A latent phenomenon of interest $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^d$

Multi-view Learning: Setting

A latent phenomenon of interest $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^d$

We observe two paired high-dimensional noisy views of the same samples:

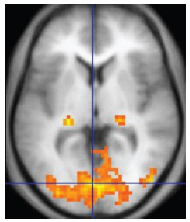
$$\boldsymbol{X} = f(\boldsymbol{\theta}, \boldsymbol{\psi}) \in \mathbb{R}^{p \times N}, \qquad \boldsymbol{Y} = g(\boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathbb{R}^{q \times N}.$$

Multi-view Learning: Setting

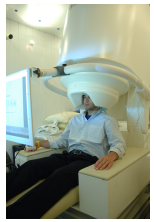
A latent phenomenon of interest $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^d$

We observe two paired high-dimensional noisy views of the same samples:

$$\boldsymbol{X} = f(\boldsymbol{\theta}, \boldsymbol{\psi}) \in \mathbb{R}^{p \times N}, \qquad \boldsymbol{Y} = g(\boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathbb{R}^{q \times N}.$$

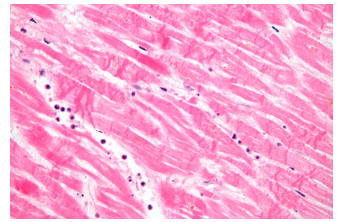


View $\boldsymbol{X}$: fMRI



View $\boldsymbol{Y}$: MEG/EEG

View $\boldsymbol{X}$: ultrasound



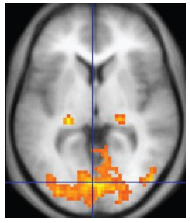
View $\boldsymbol{Y}$: pathology/omics

Multi-view Learning: Setting

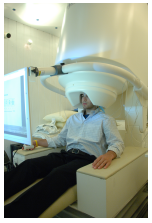
A latent phenomenon of interest $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^d$

We observe two paired high-dimensional noisy views of the same samples:

$$\boldsymbol{X} = f(\boldsymbol{\theta}, \boldsymbol{\psi}) \in \mathbb{R}^{p \times N}, \qquad \boldsymbol{Y} = g(\boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathbb{R}^{q \times N}.$$

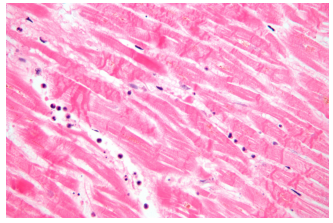


View $\boldsymbol{X}$: fMRI

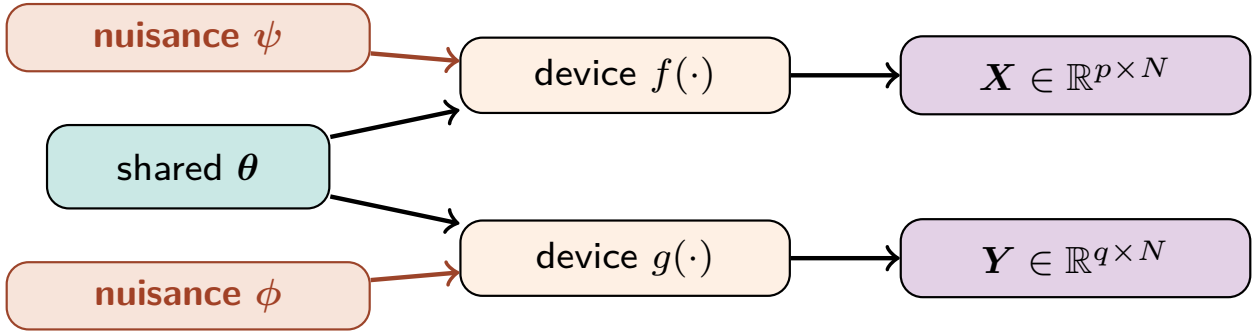


View $\boldsymbol{Y}$: MEG/EEG

View $\boldsymbol{X}$: ultrasound



View $\boldsymbol{Y}$: pathology/omics

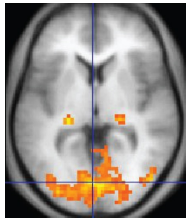


Multi-view Learning: Setting

A latent phenomenon of interest $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^d$

We observe two paired high-dimensional noisy views of the same samples:

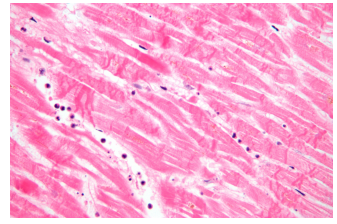
$$\boldsymbol{X} = f(\boldsymbol{\theta}, \boldsymbol{\psi}) \in \mathbb{R}^{p \times N}, \qquad \boldsymbol{Y} = g(\boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathbb{R}^{q \times N}.$$



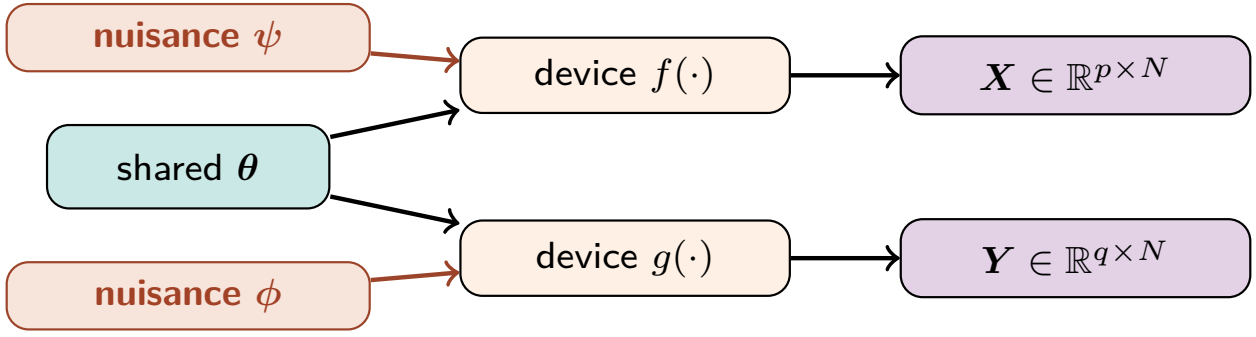
View $\boldsymbol{X}$: fMRI

View $\boldsymbol{Y}$: MEG/EEG

View $\boldsymbol{X}$: ultrasound



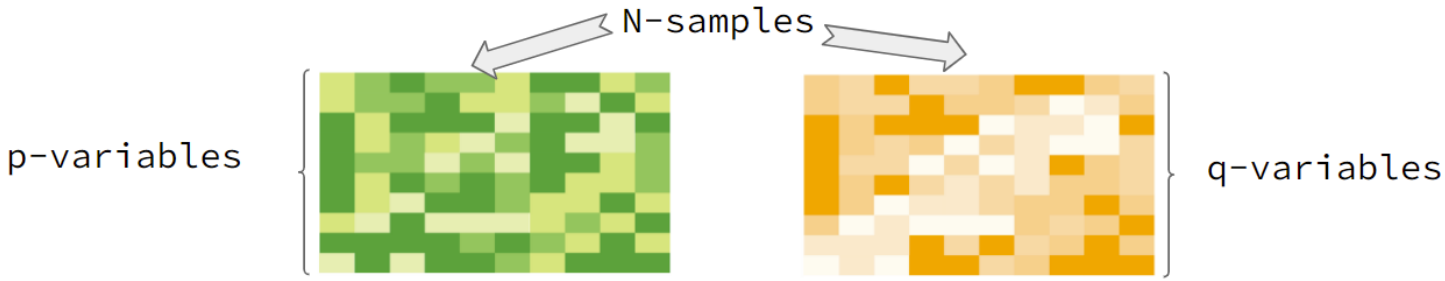
View $\boldsymbol{Y}$: pathology/omics



How can we recover the shared $\boldsymbol{\theta}$ for **clustering** while ignoring noise/nuisance $\boldsymbol{\psi}, \boldsymbol{\phi}$?

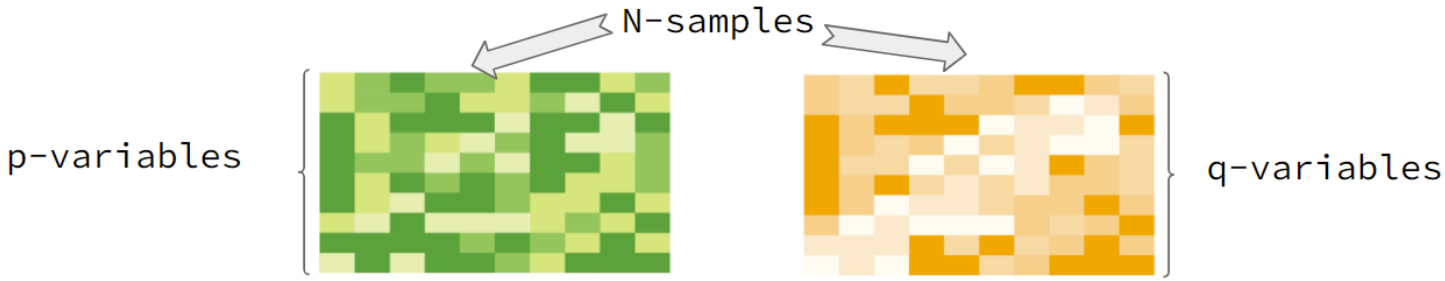
Canonical Correlation Analysis (CCA)- Hotelling 1936

Consider a coupled data $\boldsymbol{X} \in \mathbb{R}^{p \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{q \times N}$



Canonical Correlation Analysis (CCA)- Hotelling 1936

Consider a coupled data $\boldsymbol{X} \in \mathbb{R}^{p \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{q \times N}$

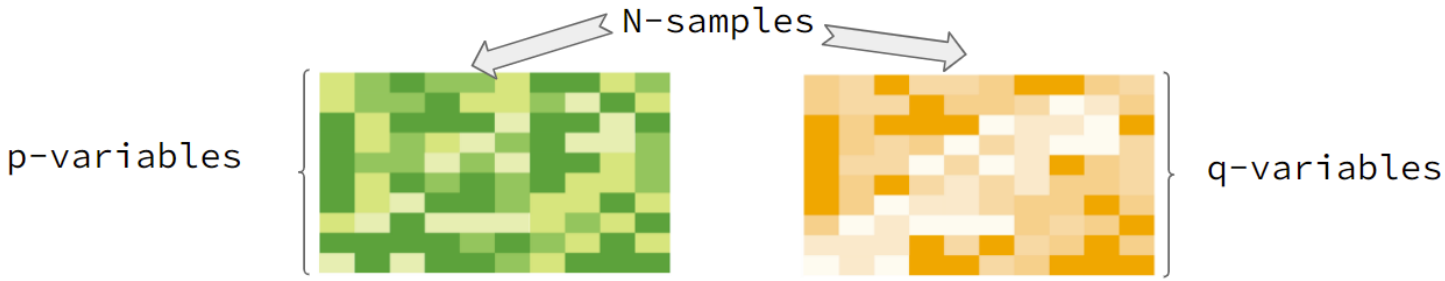


CCA seeks $\boldsymbol{a} \in \mathbb{R}^p, \boldsymbol{b} \in \mathbb{R}^q,$ such that $, \boldsymbol{u} = \boldsymbol{a}^T \boldsymbol{X}, \boldsymbol{v} = \boldsymbol{b}^T \boldsymbol{Y}$, will maximize the sample correlations between the *canonical variates*, i.e.

$$\max_{\boldsymbol{a}, \boldsymbol{b} \neq 0} \quad \rho(\boldsymbol{a}^T \boldsymbol{X}, \boldsymbol{b}^T \boldsymbol{Y}) = \frac{\boldsymbol{a}^T \boldsymbol{X} \boldsymbol{Y}^T \boldsymbol{b}}{\|\boldsymbol{a}^T \boldsymbol{X}\|_2 \|\boldsymbol{b}^T \boldsymbol{Y}\|_2}.$$

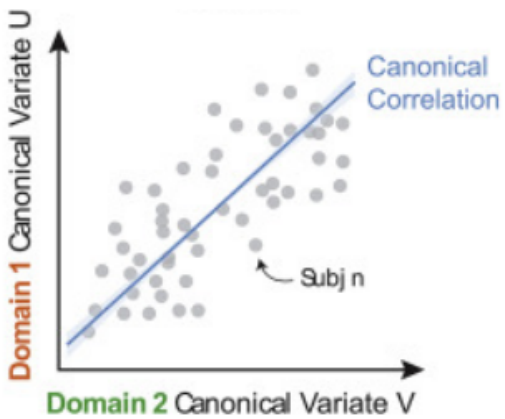
Canonical Correlation Analysis (CCA)- Hotelling 1936

Consider a coupled data $\boldsymbol{X} \in \mathbb{R}^{p \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{q \times N}$



CCA seeks $\boldsymbol{a} \in \mathbb{R}^p, \boldsymbol{b} \in \mathbb{R}^q,$ such that $, \boldsymbol{u} = \boldsymbol{a}^T \boldsymbol{X}, \boldsymbol{v} = \boldsymbol{b}^T \boldsymbol{Y}$, will maximize the sample correlations between the *canonical variates*, i.e.

$$\max_{\boldsymbol{a}, \boldsymbol{b} \neq 0} \quad \rho(\boldsymbol{a}^T \boldsymbol{X}, \boldsymbol{b}^T \boldsymbol{Y}) = \frac{\boldsymbol{a}^T \boldsymbol{X} \boldsymbol{Y}^T \boldsymbol{b}}{\|\boldsymbol{a}^T \boldsymbol{X}\|_2 \|\boldsymbol{b}^T \boldsymbol{Y}\|_2}.$$



CCA classical solution intuition

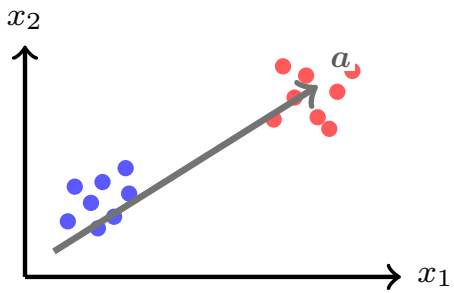
CCA finds directions that make paired projections agree, while exposing the shared cluster structure:

$$\boldsymbol{C}_x^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_y^{-1}\boldsymbol{C}_{yx}\boldsymbol{a} = \rho^2\boldsymbol{a}, \qquad u = \boldsymbol{a}^T\boldsymbol{X}, \quad v = \boldsymbol{b}^T\boldsymbol{Y}.$$

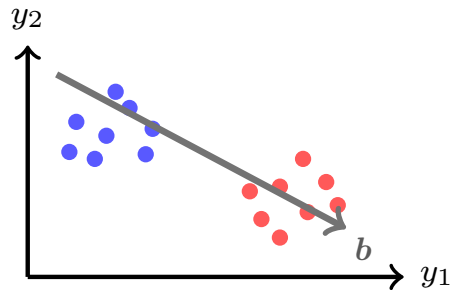
CCA classical solution intuition

CCA finds directions that make paired projections agree, while exposing the shared cluster structure:

$$\boldsymbol{C}_x^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_y^{-1}\boldsymbol{C}_{yx}\boldsymbol{a} = \rho^2\boldsymbol{a}, \qquad u = \boldsymbol{a}^T\boldsymbol{X}, \quad v = \boldsymbol{b}^T\boldsymbol{Y}.$$



view $\boldsymbol{X}$

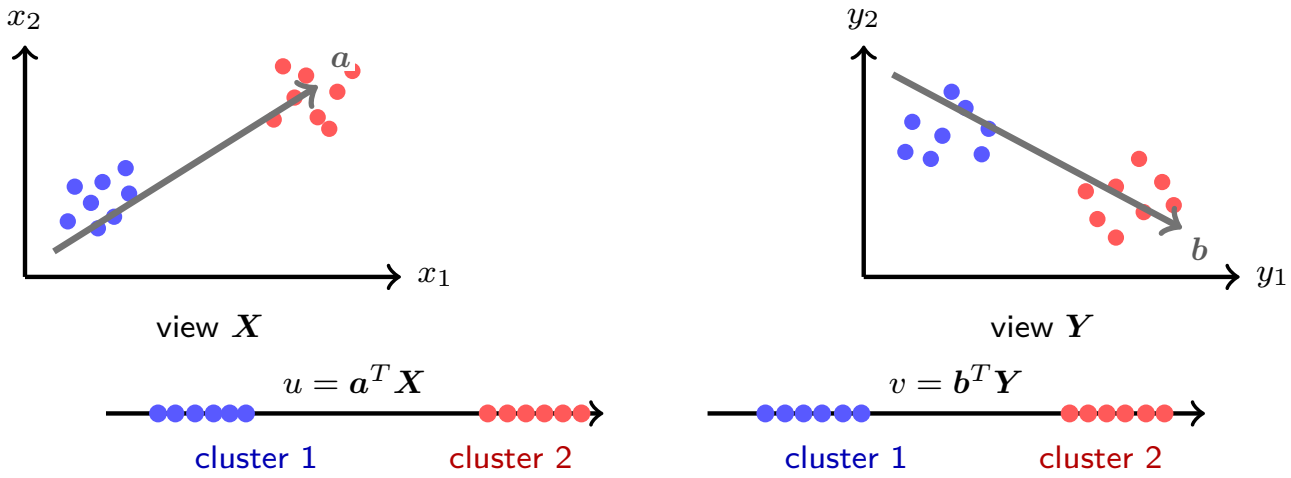


view $\boldsymbol{Y}$

CCA classical solution intuition

CCA finds directions that make paired projections agree, while exposing the shared cluster structure:

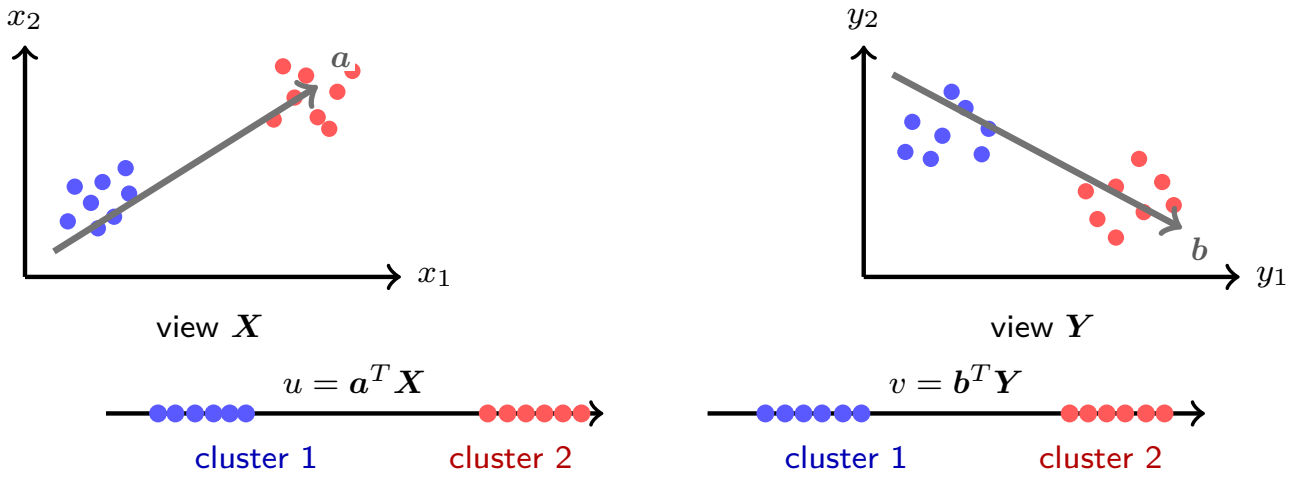
$$\boldsymbol{C}_x^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_y^{-1}\boldsymbol{C}_{yx}\boldsymbol{a} = \rho^2\boldsymbol{a}, \qquad u = \boldsymbol{a}^T\boldsymbol{X}, \quad v = \boldsymbol{b}^T\boldsymbol{Y}.$$



CCA classical solution intuition

CCA finds directions that make paired projections agree, while exposing the shared cluster structure:

$$\boldsymbol{C}_x^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_y^{-1}\boldsymbol{C}_{yx}\boldsymbol{a} = \rho^2\boldsymbol{a}, \qquad u = \boldsymbol{a}^T\boldsymbol{X}, \quad v = \boldsymbol{b}^T\boldsymbol{Y}.$$



CCA aligns paired samples and yields **cluster-separating** canonical coordinates in both views.

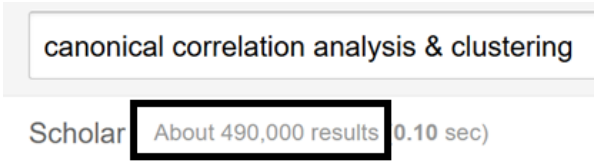
CCA Extensions

- **2002** Kernel CCA – *Bach and Jordan*.
- **2013** Deep CCA – *Andrew et al.*
- **2016** Non-parametric CCA – *Michaeli et al.*
- **2021** Deep tensor CCA – *Wong et al.*
- **2021** Deep probabilistic CCA – *Karami et al.*
- **2022** ℓ_0-DCCA – *Lindenbaum et al.*

CCA Extensions

- **2002** Kernel CCA – *Bach and Jordan*.
- **2013** Deep CCA – *Andrew et al.*
- **2016** Non-parametric CCA – *Michaeli et al.*
- **2021** Deep tensor CCA – *Wong et al.*
- **2021** Deep probabilistic CCA – *Karami et al.*
- **2022** ℓ_0-DCCA – *Lindenbaum et al.*

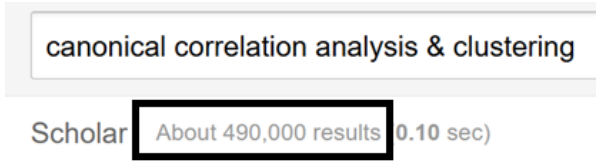
CCA is useful for clustering: correlated views reveal shared cluster structure.

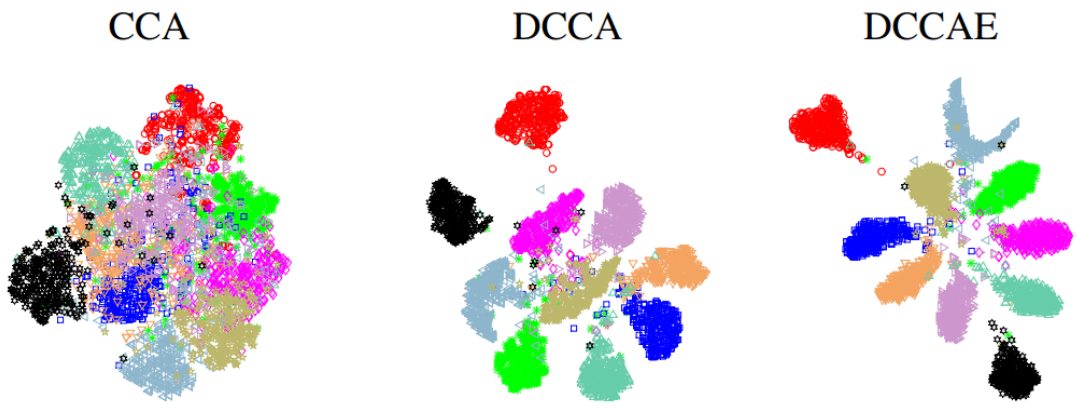


CCA Extensions

- **2002** Kernel CCA – *Bach and Jordan*.
- **2013** Deep CCA – *Andrew et al.*
- **2016** Non-parametric CCA – *Michaeli et al.*
- **2021** Deep tensor CCA – *Wong et al.*
- **2021** Deep probabilistic CCA – *Karami et al.*
- **2022** ℓ_0-DCCA – *Lindenbaum et al.*

CCA is useful for clustering: correlated views reveal shared cluster structure.

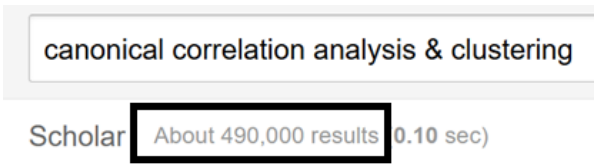


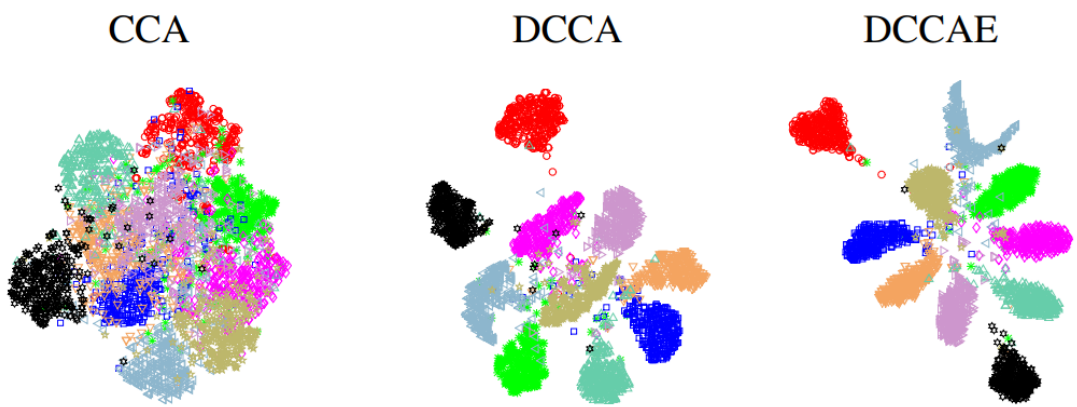


CCA Extensions

- **2002** Kernel CCA – *Bach and Jordan*.
- **2013** Deep CCA – *Andrew et al.*
- **2016** Non-parametric CCA – *Michaeli et al.*
- **2021** Deep tensor CCA – *Wong et al.*
- **2021** Deep probabilistic CCA – *Karami et al.*
- **2022** ℓ_0-DCCA – *Lindenbaum et al.*

CCA is useful for clustering: correlated views reveal shared cluster structure.





Question: can we enhance cluster separation?

Main idea: Permute samples within clusters

Within each pseudo-label cluster: keep $\boldsymbol{x}_i$, swap its pair in $\boldsymbol{Y}$.

$$(\boldsymbol{x}_i, \boldsymbol{y}_i) \quad \longrightarrow \quad (\boldsymbol{x}_i, \boldsymbol{y}_{\pi(i)}), \qquad \hat{c}_{\pi(i)} = \hat{c}_i.$$

Main idea: Permute samples within clusters

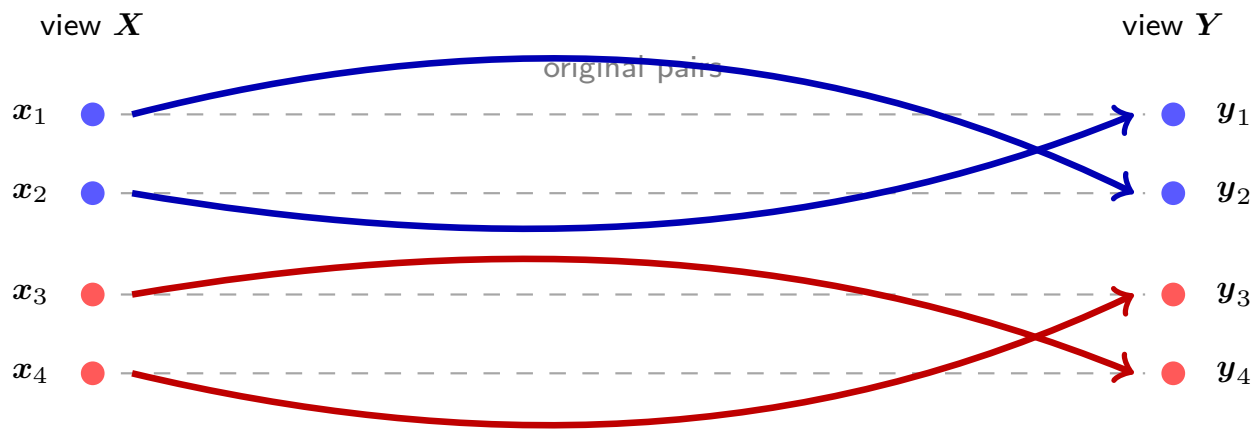
Within each pseudo-label cluster: keep $\boldsymbol{x}_i$, swap its pair in $\boldsymbol{Y}$.

$$(\boldsymbol{x}_i, \boldsymbol{y}_i) \quad \longrightarrow \quad (\boldsymbol{x}_i, \boldsymbol{y}_{\pi(i)}), \qquad \hat{c}_{\pi(i)} = \hat{c}_i.$$

Main idea: Permute samples within clusters

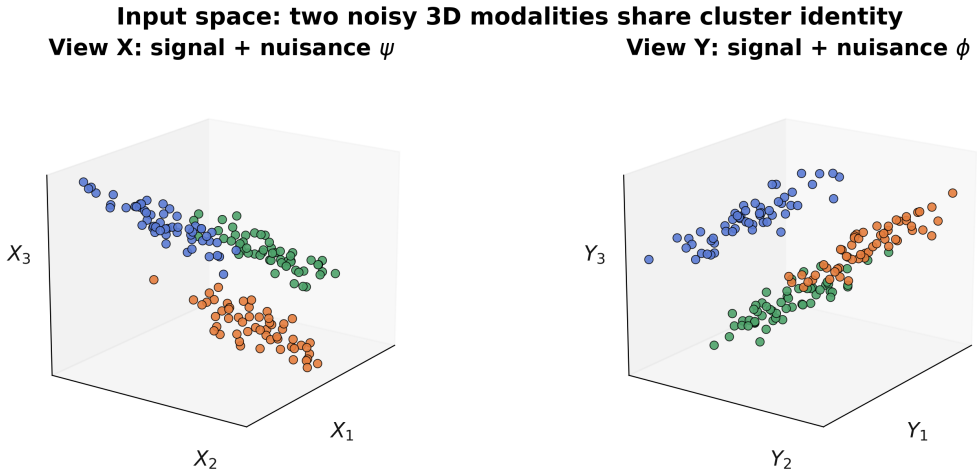
Within each pseudo-label cluster: keep $\boldsymbol{x}_i$, swap its pair in $\boldsymbol{Y}$.

$$(\boldsymbol{x}_i, \boldsymbol{y}_i) \quad \longrightarrow \quad (\boldsymbol{x}_i, \boldsymbol{y}_{\pi(i)}), \qquad \hat{c}_{\pi(i)} = \hat{c}_i.$$

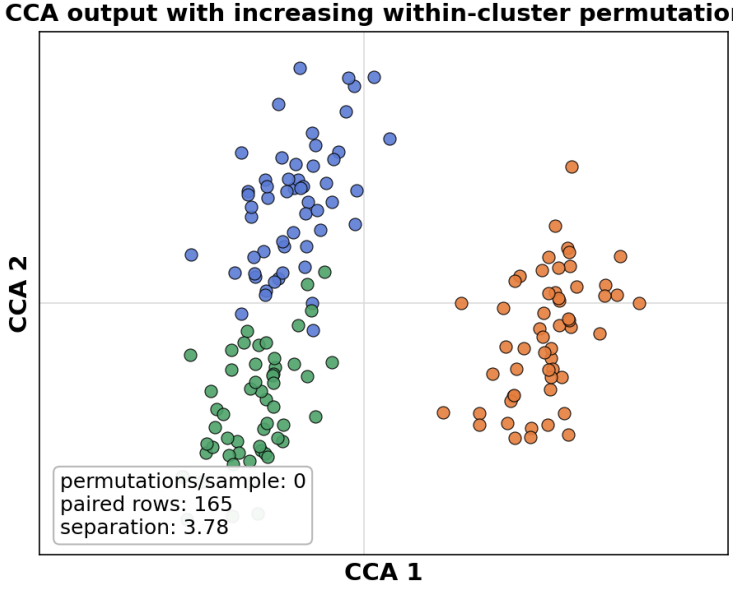


Pairs change
CCA on the new matrices

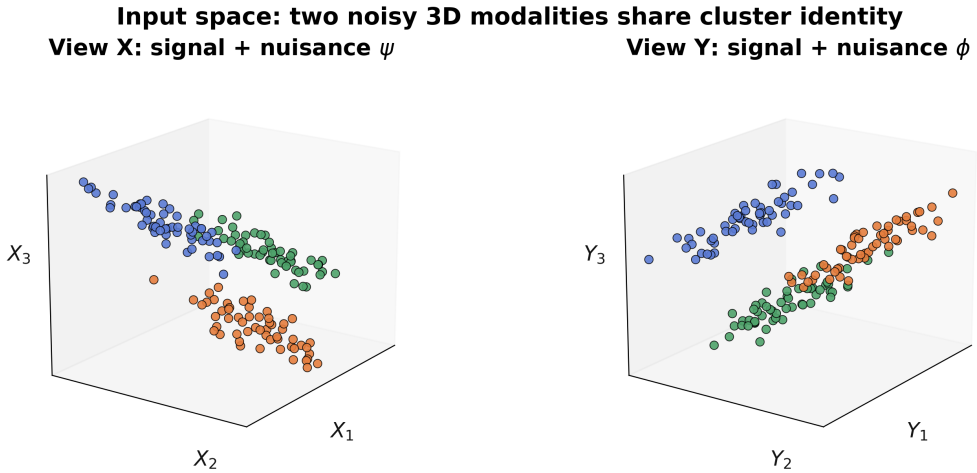
Synthetic permutation visualization



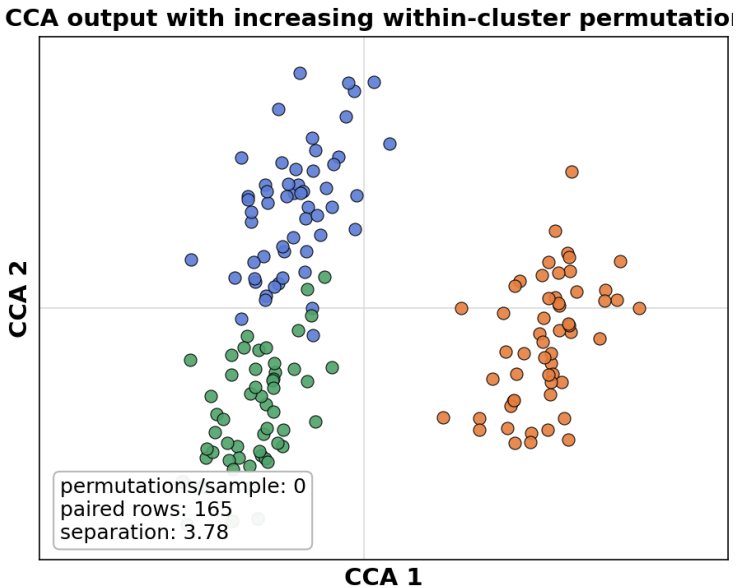
Input: two noisy 3D views share labels but have separate nuisance.



Synthetic permutation visualization

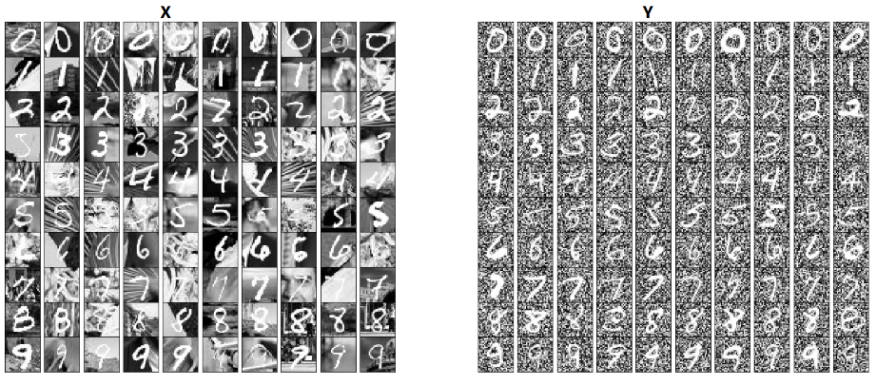


Input: two noisy 3D views share labels but have separate nuisance.



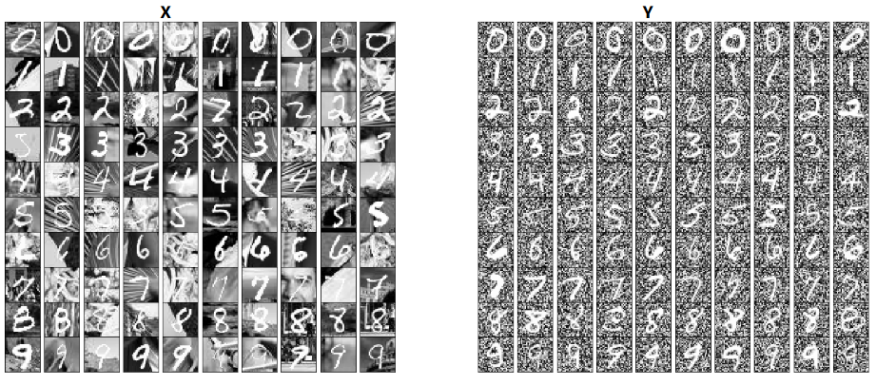
Result: within-cluster permutations add valid CCA pairs and improve separation.

Within cluster permutation: data and effect

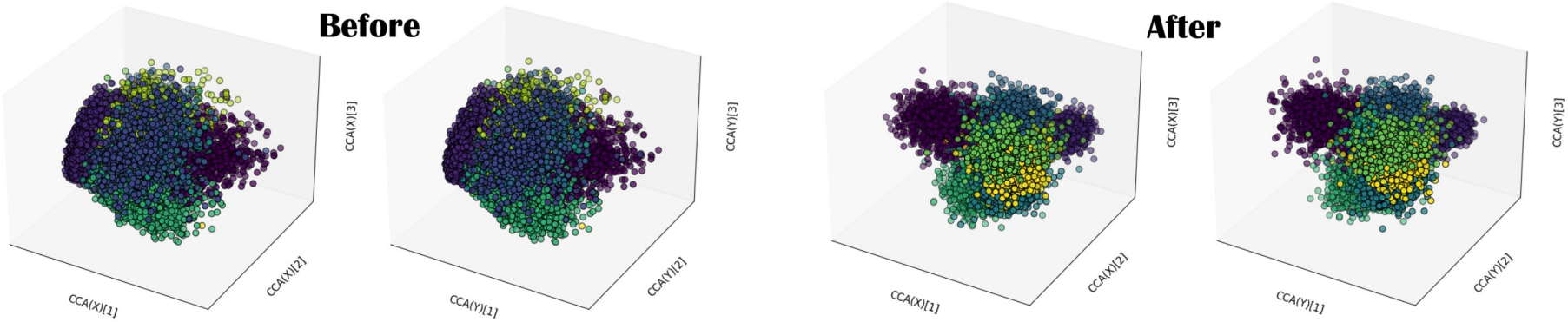


Same digit identity, different corruptions.

Within cluster permutation: data and effect



Same digit identity, different corruptions.



Permutations add valid pairs and sharpen separation.

Theory: CCA links to Linear Discriminant Analysis (LDA)

Let the shared variable be the class label, $\boldsymbol{\theta}_i = \ell_i$.

Linear Discriminant Analysis (LDA) maximizes between-class scatter relative to within-class scatter:

$$\max_{\boldsymbol{w} \neq 0} \frac{\boldsymbol{w}^T \boldsymbol{C}_b \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{C}_w \boldsymbol{w}} \quad \leftarrow \textbf{between vs. within class variance}$$

$$\boldsymbol{C}_b = \sum_{k=1}^{K} n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \quad \text{(between-class)}$$

$$\boldsymbol{C}_w = \sum_{k=1}^{K} \sum_{i:\ell_i = k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \quad \text{(within-class)}.$$

Theory: CCA links to Linear Discriminant Analysis (LDA)

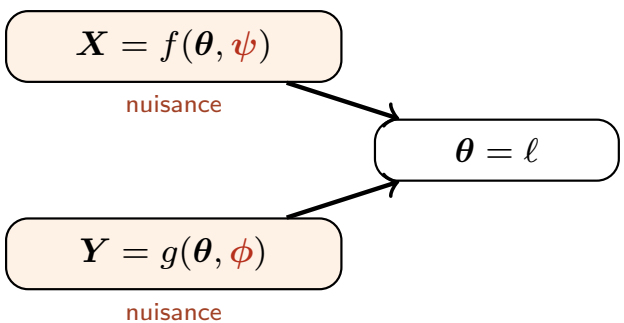
Let the shared variable be the class label, $\boldsymbol{\theta}_i = \ell_i$.

Linear Discriminant Analysis (LDA) maximizes between-class scatter relative to within-class scatter:

$$\max_{\boldsymbol{w} \neq 0} \frac{\boldsymbol{w}^T \boldsymbol{C}_b \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{C}_w \boldsymbol{w}} \quad \leftarrow \textbf{between vs. within class variance}$$

$$\boldsymbol{C}_b = \sum_{k=1}^{K} n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \quad \text{(between-class)}$$

$$\boldsymbol{C}_w = \sum_{k=1}^{K} \sum_{i:\ell_i = k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \quad \text{(within-class)}.$$



Theory: CCA links to Linear Discriminant Analysis (LDA)

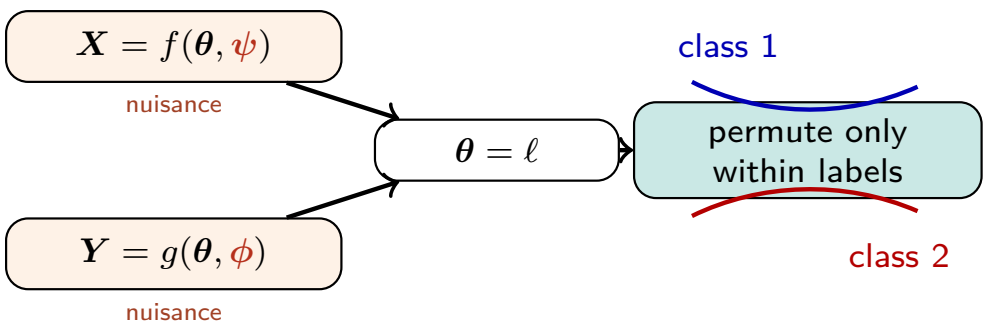
Let the shared variable be the class label, $\boldsymbol{\theta}_i = \ell_i$.

Linear Discriminant Analysis (LDA) maximizes between-class scatter relative to within-class scatter:

$$\max_{\boldsymbol{w} \neq 0} \frac{\boldsymbol{w}^T \boldsymbol{C}_b \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{C}_w \boldsymbol{w}} \quad \leftarrow \textbf{between vs. within class variance}$$

$$\boldsymbol{C}_b = \sum_{k=1}^{K} n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \quad \text{(between-class)}$$

$$\boldsymbol{C}_w = \sum_{k=1}^{K} \sum_{i:\ell_i = k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \quad \text{(within-class)}.$$



Theory: CCA links to Linear Discriminant Analysis (LDA)

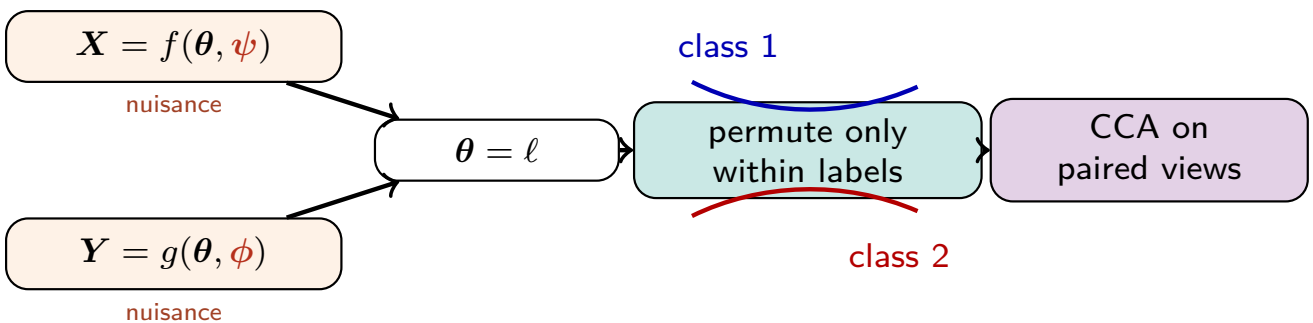
Let the shared variable be the class label, $\boldsymbol{\theta}_i = \ell_i$.

Linear Discriminant Analysis (LDA) maximizes between-class scatter relative to within-class scatter:

$$\max_{\boldsymbol{w} \neq 0} \frac{\boldsymbol{w}^T \boldsymbol{C}_b \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{C}_w \boldsymbol{w}} \quad \leftarrow \textbf{between vs. within class variance}$$

$$\boldsymbol{C}_b = \sum_{k=1}^{K} n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \quad \text{(between-class)}$$

$$\boldsymbol{C}_w = \sum_{k=1}^{K} \sum_{i:\ell_i = k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \quad \text{(within-class)}.$$



Theory: CCA links to Linear Discriminant Analysis (LDA)

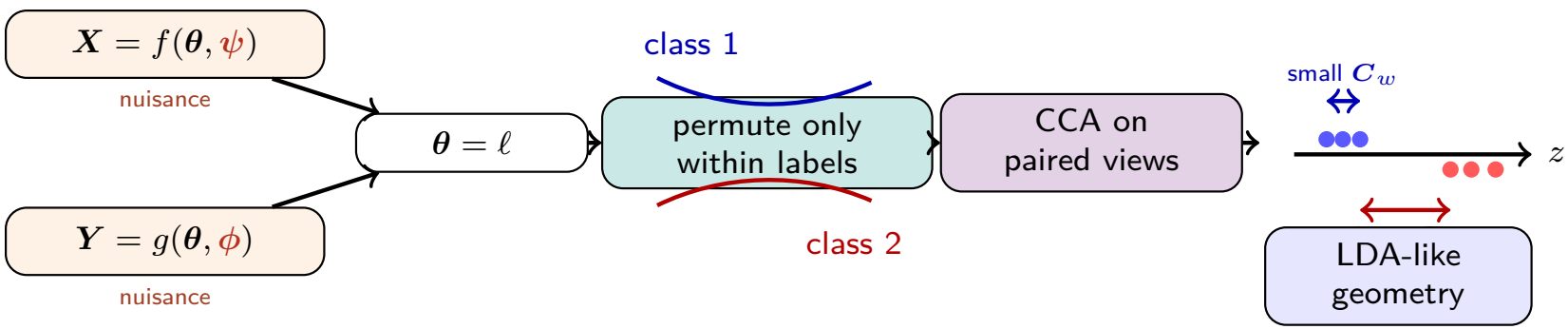
Let the shared variable be the class label, $\boldsymbol{\theta}_i = \ell_i$.

Linear Discriminant Analysis (LDA) maximizes between-class scatter relative to within-class scatter:

$$\max_{\boldsymbol{w} \neq 0} \frac{\boldsymbol{w}^T \boldsymbol{C}_b \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{C}_w \boldsymbol{w}} \quad \leftarrow \textbf{between vs. within class variance}$$

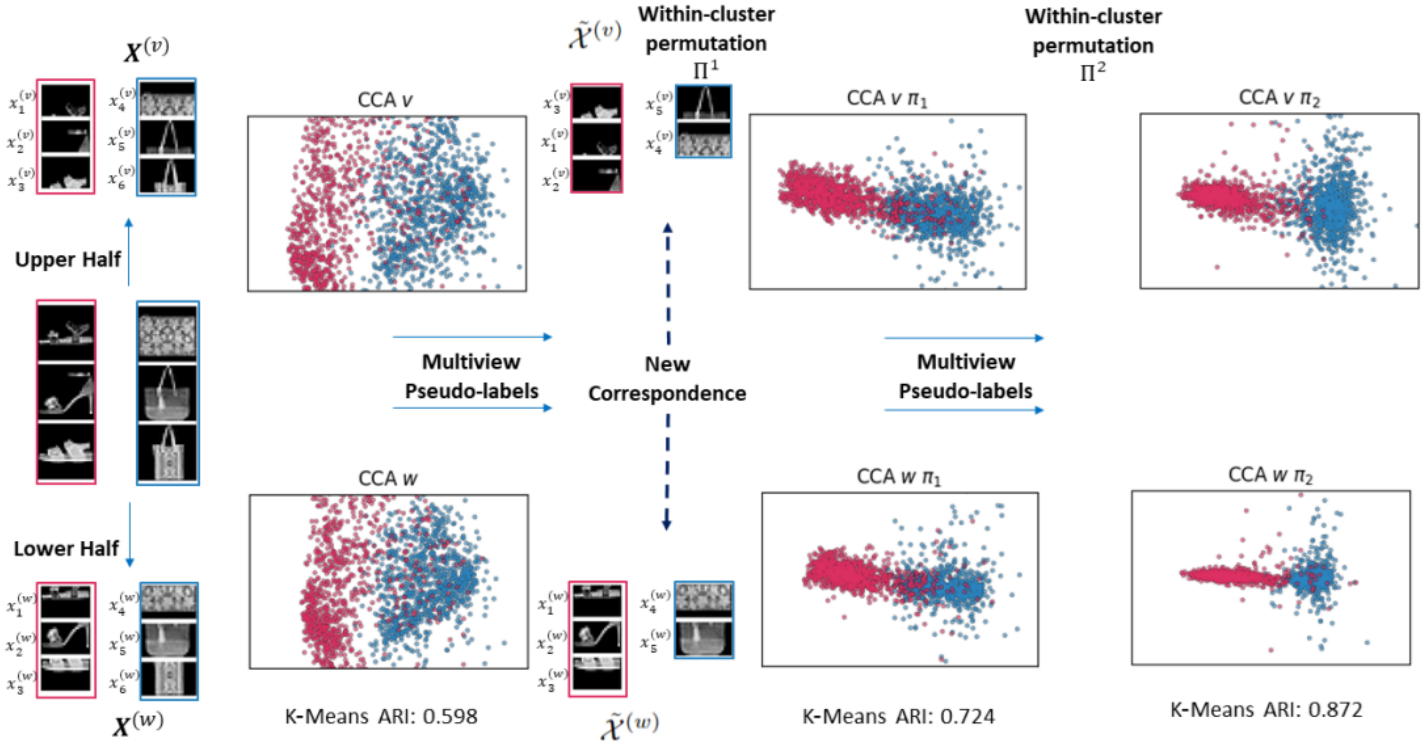
$$\boldsymbol{C}_b = \sum_{k=1}^{K} n_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \quad \text{(between-class)}$$

$$\boldsymbol{C}_w = \sum_{k=1}^{K} \sum_{i:\ell_i = k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \quad \text{(within-class)}.$$

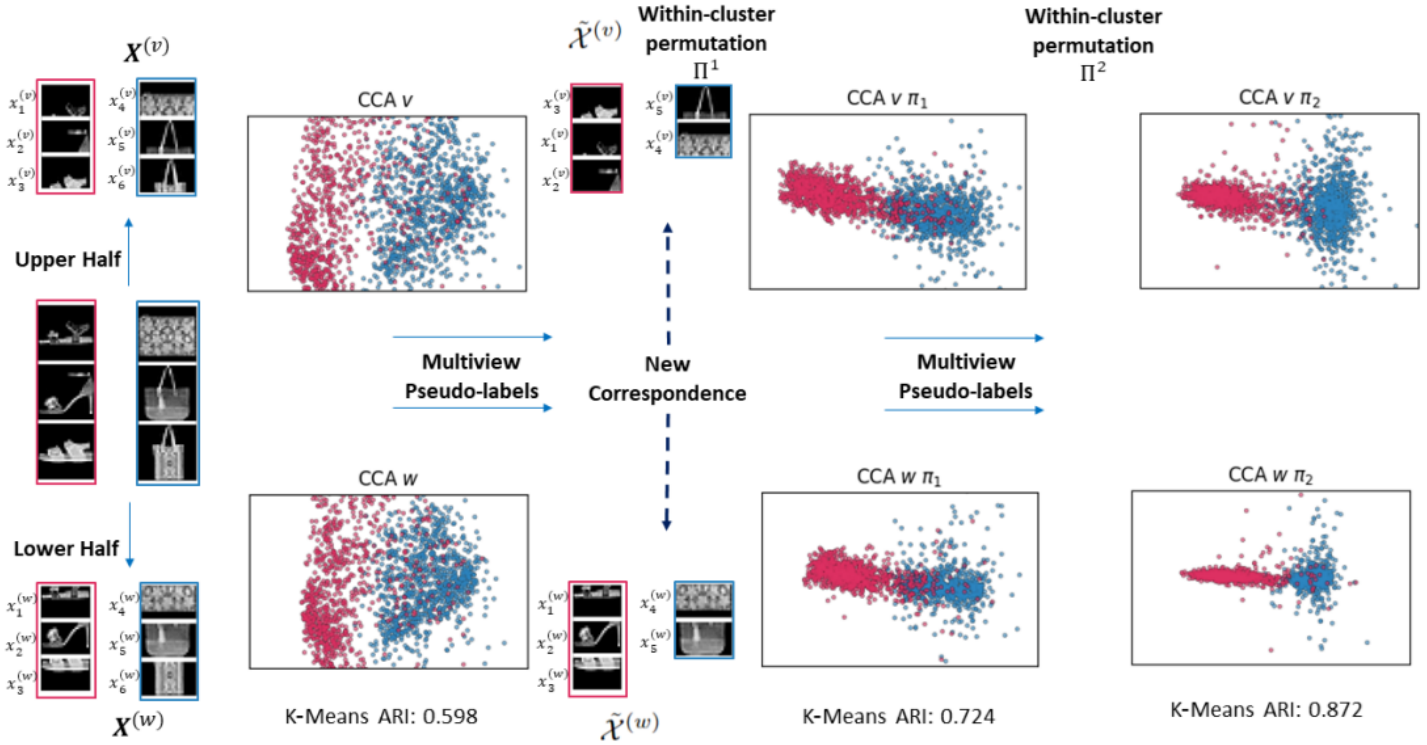


Label-preserving pairs make CCA favor LDA-like separation.

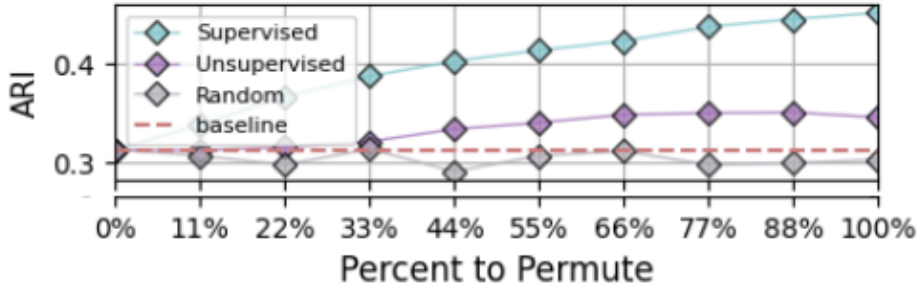
Illustration



Illustration



More valid permutations enhance cluster separation.



Theory: What pseudo-label noise costs

Clean label-preserving pairs give the LDA-like CCA operator; wrong pseudo-labels add noise.

$$\underbrace{\boldsymbol{A} = \boldsymbol{C}_w^{-1}\boldsymbol{C}_b}_{\text{ideal LDA/CCA operator}} \qquad \underbrace{\widehat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{D}}_{\text{pseudo-label errors add perturbation}}$$

C_w: **within-class covariance**; C_b: **between-cluster covariance**

Theory: What pseudo-label noise costs

Clean label-preserving pairs give the LDA-like CCA operator; wrong pseudo-labels add noise.

$$\underbrace{\boldsymbol{A} = \boldsymbol{C}_w^{-1}\boldsymbol{C}_b}_{\text{ideal LDA/CCA operator}} \qquad \underbrace{\widehat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{D}}_{\text{pseudo-label errors add perturbation}}$$

$\boldsymbol{C}_w$: **within-class covariance**; $\boldsymbol{C}_b$: **between-cluster covariance**

$|\widehat{\lambda}_i - \lambda_i| \leq \|\boldsymbol{D}\|_2 \quad \Rightarrow \quad$ smaller $\|\boldsymbol{D}\|_2$ means more stable CCA directions.

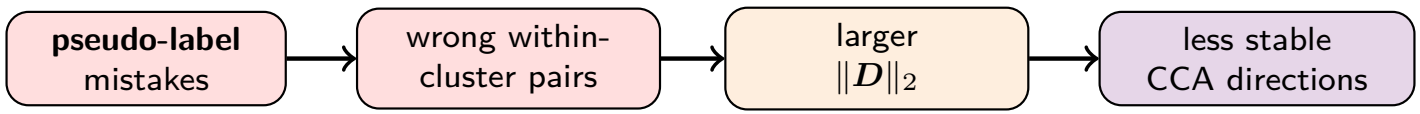
Theory: What pseudo-label noise costs

Clean label-preserving pairs give the LDA-like CCA operator; wrong pseudo-labels add noise.

$$\underbrace{\boldsymbol{A} = \boldsymbol{C}_w^{-1}\boldsymbol{C}_b}_{\text{ideal LDA/CCA operator}} \qquad \underbrace{\widehat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{D}}_{\text{pseudo-label errors add perturbation}}$$

$\boldsymbol{C}_w$: **within-class covariance**; $\boldsymbol{C}_b$: **between-cluster covariance**

$|\widehat{\lambda}_i - \lambda_i| \leq \|\boldsymbol{D}\|_2 \quad \Rightarrow \quad$ smaller $\|\boldsymbol{D}\|_2$ means more stable CCA directions.



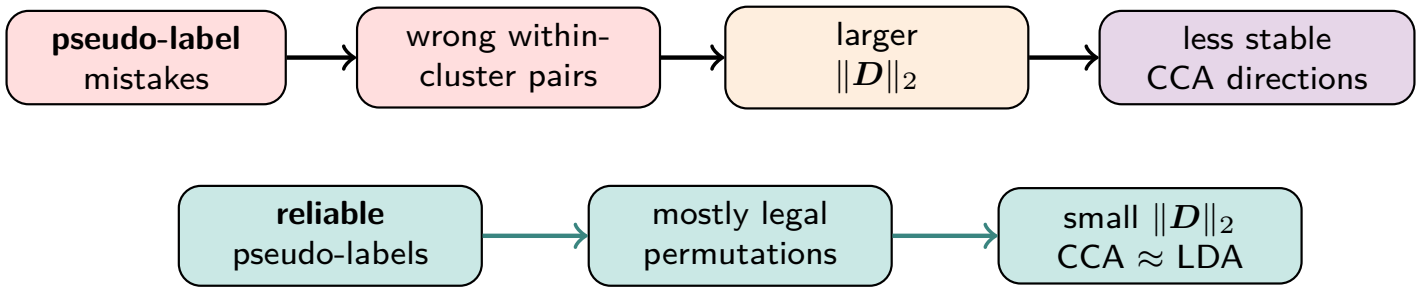
Theory: What pseudo-label noise costs

Clean label-preserving pairs give the LDA-like CCA operator; wrong pseudo-labels add noise.

$$\underbrace{\boldsymbol{A} = \boldsymbol{C}_w^{-1}\boldsymbol{C}_b}_{\text{ideal LDA/CCA operator}} \qquad \underbrace{\widehat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{D}}_{\text{pseudo-label errors add perturbation}}$$

$\boldsymbol{C}_w$: **within-class covariance**; $\boldsymbol{C}_b$: **between-cluster covariance**

$|\widehat{\lambda}_i - \lambda_i| \leq \|\boldsymbol{D}\|_2 \quad \Rightarrow \quad$ smaller $\|\boldsymbol{D}\|_2$ means more stable CCA directions.



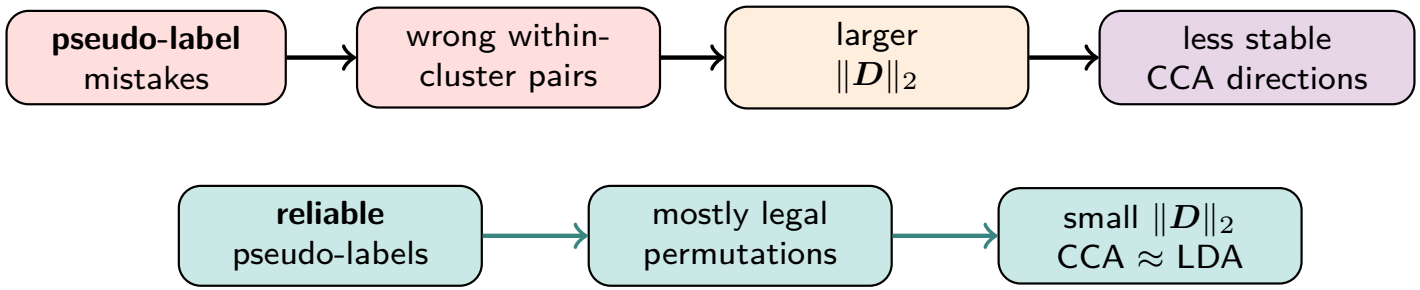
Theory: What pseudo-label noise costs

Clean label-preserving pairs give the LDA-like CCA operator; wrong pseudo-labels add noise.

$$\underbrace{\boldsymbol{A} = \boldsymbol{C}_w^{-1}\boldsymbol{C}_b}_{\text{ideal LDA/CCA operator}} \qquad \underbrace{\widehat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{D}}_{\text{pseudo-label errors add perturbation}}$$

$\boldsymbol{C}_w$: **within-class covariance**; $\boldsymbol{C}_b$: **between-cluster covariance**

$|\widehat{\lambda}_i - \lambda_i| \leq \|\boldsymbol{D}\|_2 \quad \Rightarrow \quad$ smaller $\|\boldsymbol{D}\|_2$ means more stable CCA directions.

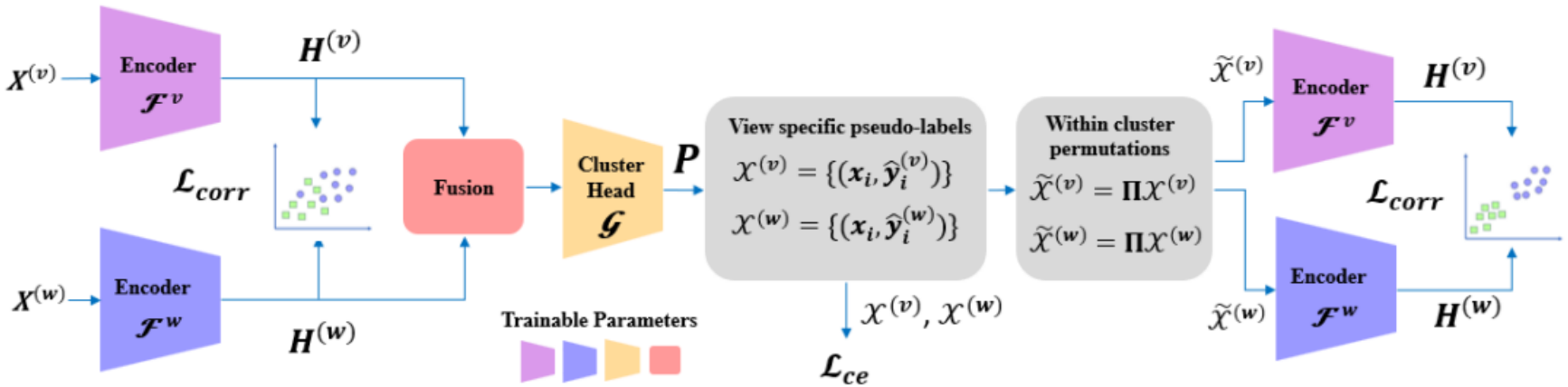


Message

The theorem says **pseudo-label quality controls** how far the learned CCA representation drifts from the ideal supervised LDA representation.

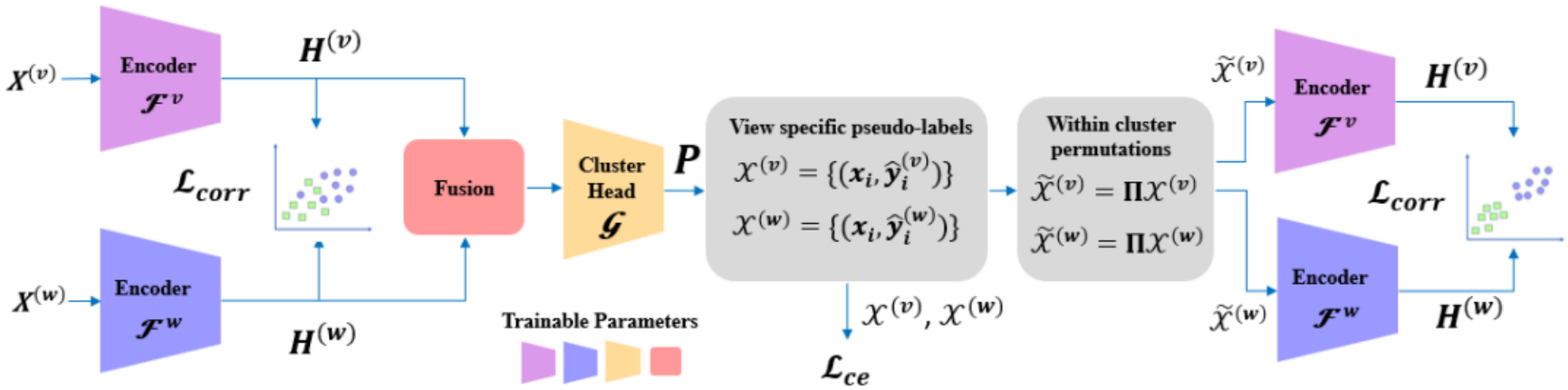
COPER: COrrelation-based PERmutations

COPER is trained end-to-end: within-cluster permutations are sampled during training and fed to the CCA objective.



COPER: COrrelation-based PERmutations

COPER is trained end-to-end: within-cluster permutations are sampled during training and fed to the CCA objective.

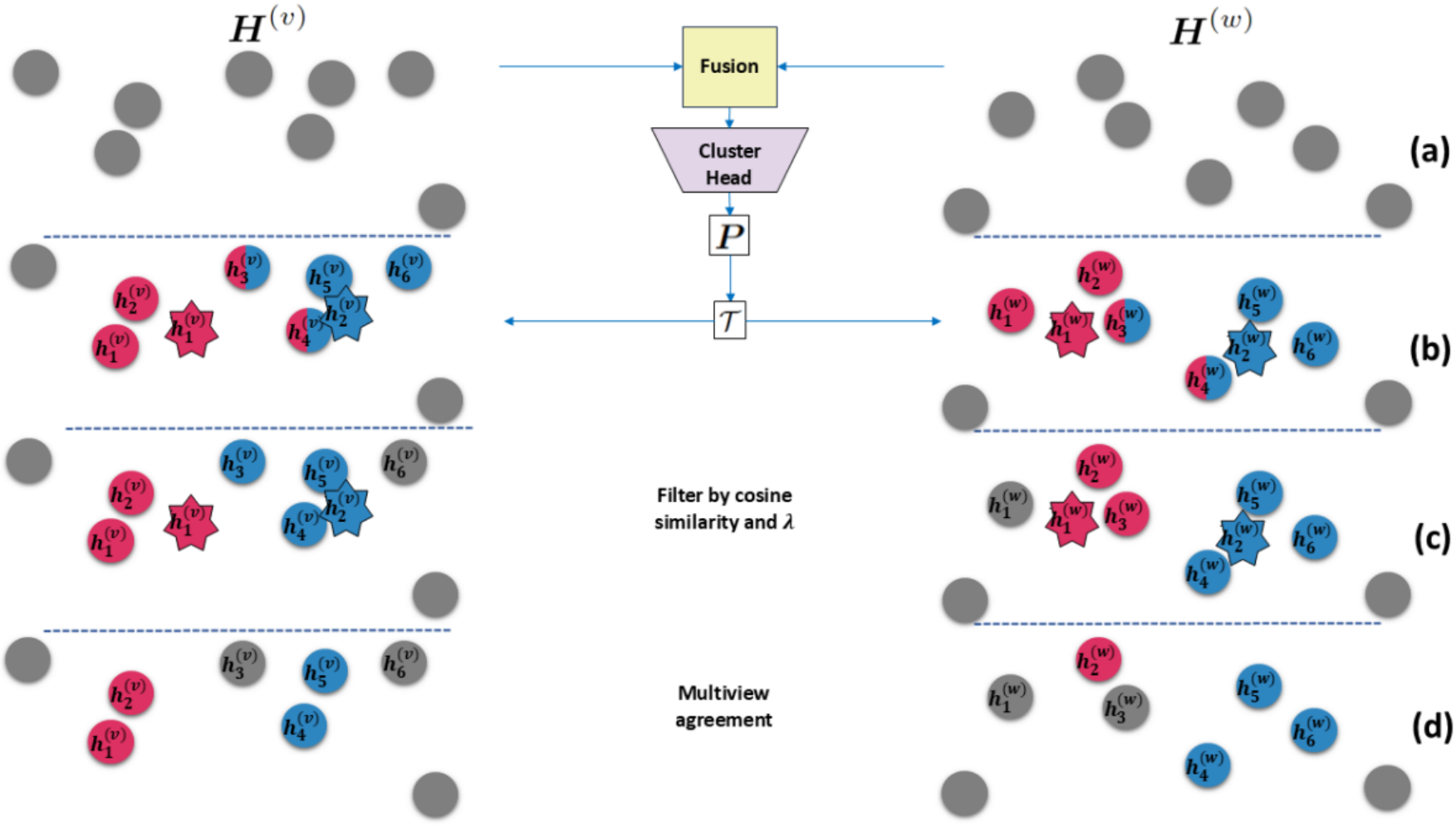


Core mechanism

Reliable pseudo-labels define valid within-cluster permutations; CCA then learns sharper clusters.

Creating Reliable Labels

Four steps for reliable pseudo-labels.



Results

Nonlinear methods

Avg. ACC gain: +5.7

Method	METABRIC	Reuters	Caltech101-20	VOC	Caltech5V-7	RBGD	MNIST-USPS	CCV	MSRVC1	Scene15
					ACC					
DSMVC	40.60±3.8	46.37±4.4	39.33±2.4	57.82±5.0	79.24±9.5	39.77±3.6	70.06±10.3	17.90±1.2	60.71±15.2	34.30±2.9
CVCL	42.66±6.2	45.06±8.0	33.50±1.4	36.88±3.1	78.58±5.0	31.04±1.8	99.38±0.1	26.23±1.9	77.90±12.3	40.16±1.8
ICMVC	32.12±1.16	38.18±0.78	26.54±0.52	36.94±1.29	60.97±2.37	32.97±1.23	99.29±0.08	20.88±1.21	66.28±6.11	**41.26±0.89**
RMCNC	32.55±1.30	37.61±1.50	36.73±1.38	39.66±1.48	56.77±2.98	33.22±1.85	76.37±4.13	21.46±0.69	32.38±1.35	40.03±0.98
OPMC	41.25±0.40	47.50±0.65	48.25±1.95	62.86±1.78	**88.51±0.57**	41.96±1.40	72.31±0.24	26.91±0.44	88.00±1.10	40.97±1.69
MVCAN	48.32±1.28	45.44±2.81	45.56±1.67	17.85±0.35	75.63±0.14	21.93±0.49	85.57±1.14	19.29±0.80	42.95±1.14	38.51±1.85
AE	38.92±2.5	43.35±4.0	40.39±2.3	36.66±2.69	54.48±3.2	32.16±1.6	33.83±3.6	15.58±0.6	56.76±3.8	33.59±2.3
DCCA-AE	45.39±2.8	43.18±2.8	48.18±3.4	46.53±2.7	58.37±3.9	32.59±1.3	92.19±2.4	18.28±0.9	59.62±1.1	33.70±2.0
ℓ_0-DCCA	41.64±2.8	34.44±3.69	39.24±2.04	44.46±1.18	47.46±2.38	28.10±1.74	99.54±0.09	25.41±0.83	42.57±2.60	38.67±1.88
COPER	**49.13±3.2**	**53.15±3.3**	**54.83±3.9**	**64.65±2.6**	82.81±5.1	**49.80±0.8**	**99.88±0.0**	**28.06±1.1**	**92.57±0.8**	40.68±1.6

Results

Nonlinear methods

Avg. ACC gain: +5.7

Method	METABRIC	Reuters	Caltech101-20	VOC	Caltech5V-7	RBGD	MNIST-USPS	CCV	MSRVC1	Scene15
					ACC					
DSMVC	40.60±3.8	46.37±4.4	39.33±2.4	57.82±5.0	79.24±9.5	39.77±3.6	70.06±10.3	17.90±1.2	60.71±15.2	34.30±2.9
CVCL	42.66±6.2	45.06±8.0	33.50±1.4	36.88±3.1	78.58±5.0	31.04±1.8	99.38±0.1	26.23±1.9	77.90±12.3	40.16±1.8
ICMVC	32.12±1.16	38.18±0.78	26.54±0.52	36.94±1.29	60.97±2.37	32.97±1.23	99.29±0.08	20.88±1.21	66.28±6.11	**41.26±0.89**
RMCNC	32.55±1.30	37.61±1.50	36.73±1.38	39.66±1.48	56.77±2.98	33.22±1.85	76.37±4.13	21.46±0.69	32.38±1.35	40.03±0.98
OPMC	41.25±0.40	47.50±0.65	48.25±1.95	62.86±1.78	**88.51±0.57**	41.96±1.40	72.31±0.24	26.91±0.44	88.00±1.10	40.97±1.69
MVCAN	48.32±1.28	45.44±2.81	45.56±1.67	17.85±0.35	75.63±0.14	21.93±0.49	85.57±1.14	19.29±0.80	42.95±1.14	38.51±1.85
AE	38.92±2.5	43.35±4.0	40.39±2.3	36.66±2.69	54.48±3.2	32.16±1.6	33.83±3.6	15.58±0.6	56.76±3.8	33.59±2.3
DCCA-AE	45.39±2.8	43.18±2.8	48.18±3.4	46.53±2.7	58.37±3.9	32.59±1.3	92.19±2.4	18.28±0.9	59.62±1.1	33.70±2.0
ℓ_0-DCCA	41.64±2.8	34.44±3.69	39.24±2.04	44.46±1.18	47.46±2.38	28.10±1.74	99.54±0.09	25.41±0.83	42.57±2.60	38.67±1.88
COPER	**49.13±3.2**	**53.15±3.3**	**54.83±3.9**	**64.65±2.6**	82.81±5.1	**49.80±0.8**	**99.88±0.0**	**28.06±1.1**	**92.57±0.8**	40.68±1.6

Linear methods

Avg. ACC gain: +1.4

Method	METABRIC	Reuters	Caltech101-20	VOC	Caltech5V-7	RBGD	MNIST-USPS	CCV	MSRVC1	Scene15
					ACC					
Raw	35.18±2.9	37.16±5.9	42.98±3.1	52.41±7.3	73.54±6.3	43.15±1.8	69.78±6.3	16.25±0.7	74.00±7.3	37.91±0.9
PCA	37.72±1.4	42.56±2.9	41.47±3.4	43.66±4.9	74.26±6.3	42.64±1.9	68.14±3.9	16.17±0.5	74.14±6.1	37.53±1.2
CCA	40.05±2.1	42.85±2.5	42.34±3.1	53.1±4.1	75.96±5.2	41.47±2.2	80.15±5.0	16.16±0.7	73.29±7.3	37.59±1.3
CCA w perm	**40.6**±1.6	**43.16**±1.7	**43.08**±3.8	**53.52**±4.4	**77.44**±5.6	**44.29**±2.1	**84.99**±6.3	**16.39**±0.6	**75.00**±4.1	**38.25**±1.2

Results

Nonlinear methods

Avg. ACC gain: +5.7

Method	METABRIC	Reuters	Caltech101-20	VOC	Caltech5V-7	RBGD	MNIST-USPS	CCV	MSRVC1	Scene15
					ACC					
DSMVC	40.60±3.8	46.37±4.4	39.33±2.4	57.82±5.0	79.24±9.5	39.77±3.6	70.06±10.3	17.90±1.2	60.71±15.2	34.30±2.9
CVCL	42.66±6.2	45.06±8.0	33.50±1.4	36.88±3.1	78.58±5.0	31.04±1.8	99.38±0.1	26.23±1.9	77.90±12.3	40.16±1.8
ICMVC	32.12±1.16	38.18±0.78	26.54±0.52	36.94±1.29	60.97±2.37	32.97±1.23	99.29±0.08	20.88±1.21	66.28±6.11	**41.26±0.89**
RMCNC	32.55±1.30	37.61±1.50	36.73±1.38	39.66±1.48	56.77±2.98	33.22±1.85	76.37±4.13	21.46±0.69	32.38±1.35	40.03±0.98
OPMC	41.25±0.40	47.50±0.65	48.25±1.95	62.86±1.78	**88.51±0.57**	41.96±1.40	72.31±0.24	26.91±0.44	88.00±1.10	40.97±1.69
MVCAN	48.32±1.28	45.44±2.81	45.56±1.67	17.85±0.35	75.63±0.14	21.93±0.49	85.57±1.14	19.29±0.80	42.95±1.14	38.51±1.85
AE	38.92±2.5	43.35±4.0	40.39±2.3	36.66±2.69	54.48±3.2	32.16±1.6	33.83±3.6	15.58±0.6	56.76±3.8	33.59±2.3
DCCA-AE	45.39±2.8	43.18±2.8	48.18±3.4	46.53±2.7	58.37±3.9	32.59±1.3	92.19±2.4	18.28±0.9	59.62±1.1	33.70±2.0
ℓ_0-DCCA	41.64±2.8	34.44±3.69	39.24±2.04	44.46±1.18	47.46±2.38	28.10±1.74	99.54±0.09	25.41±0.83	42.57±2.60	38.67±1.88
COPER	**49.13±3.2**	**53.15±3.3**	**54.83±3.9**	**64.65±2.6**	82.81±5.1	**49.80±0.8**	**99.88±0.0**	**28.06±1.1**	**92.57±0.8**	40.68±1.6

Linear methods

Avg. ACC gain: +1.4

Method	METABRIC	Reuters	Caltech101-20	VOC	Caltech5V-7	RBGD	MNIST-USPS	CCV	MSRVC1	Scene15
					ACC					
Raw	35.18±2.9	37.16±5.9	42.98±3.1	52.41±7.3	73.54±6.3	43.15±1.8	69.78±6.3	16.25±0.7	74.00±7.3	37.91±0.9
PCA	37.72±1.4	42.56±2.9	41.47±3.4	43.66±4.9	74.26±6.3	42.64±1.9	68.14±3.9	16.17±0.5	74.14±6.1	37.53±1.2
CCA	40.05±2.1	42.85±2.5	42.34±3.1	53.1±4.1	75.96±5.2	41.47±2.2	80.15±5.0	16.16±0.7	73.29±7.3	37.59±1.3
CCA w perm	**40.6**±1.6	**43.16**±1.7	**43.08**±3.8	**53.52**±4.4	**77.44**±5.6	**44.29**±2.1	**84.99**±6.3	**16.39**±0.6	**75.00**±4.1	**38.25**±1.2

Permutation gains: +7.2 ACC, +9.8 ARI, +7.0 NMI

Model	ACC	ARI	NMI
COPER with linear $\mathcal{F}$	48.00±4.6	22.22±5.0	32.26±5.3
COPER w/o $\mathcal{L}_{corr}$	79.71±3.3	64.46±4.4	70.76±4.1
COPER w/o permutations	85.33±7.5	74.80±10.2	79.87±7.2
COPER w/o multi-view agree.	86.29±7.7	76.10±10.6	80.81±7.6
COPER	92.57±0.8	84.57±1.4	86.91±1.3

Conclusion

- **Reliable pseudo-labels** create valid cross-view permutations.
- Permuted CCA gives **LDA-like separation** and improves multi-view clustering.
- **COPER** is a leading end-to-end multi-view / multimodal clustering model.

GitHub code