

Unpaired Canonical Correlation Analysis

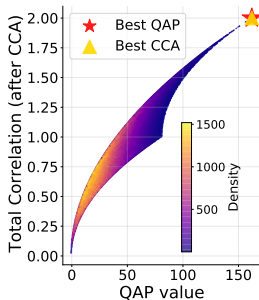
Nir Ben-Ari¹ Ronen Talmon² Uri Shaham¹
¹Bar-Ilan University ²Technion

NeurIPS 2026

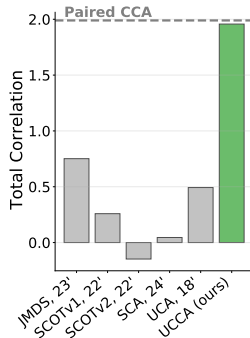


CCA Without Pairs?

- **CCA** finds maximally correlated linear projections of two views: a cornerstone of multiview learning.
- It **strictly relies on paired samples**, which are often expensive or impossible to obtain.
- Unpaired data, in contrast, is abundant.
- **Key theoretical insight:** unpaired CCA can be solved with a **linear-kernel QAP** (a).
- **UCCA** is the first method to learn maximally correlated projections **from unpaired data**, relying on this QAP–CCA connection (b).



(a)



(b)

Problem Setup

- Whitened views $X, Y \in \mathbb{R}^{n \times d}$, total correlation $\text{TC}(X, Y) = \text{tr}(X^\top Y)$.
- **CCA:** $(U^*, V^*) = \arg \max_{U^\top U = V^\top V = I} \text{TC}(XU, YV)$.
- **Unpaired:** in practice, the training views contain **no pairs at all**. For the theory, we model them as related by an unknown permutation P^* , and seek

$$\text{TC}(XU', P^*YV') = \text{TC}(XU^*, P^*YV^*).$$

- **Strategy:** find a **proxy pairing** P' and set $(U', V') = \text{CCA}(X, P'Y)$.

Note

We only need the proxy to give the *same correlation*: exact recovery $P' \approx P^*$ is not required.

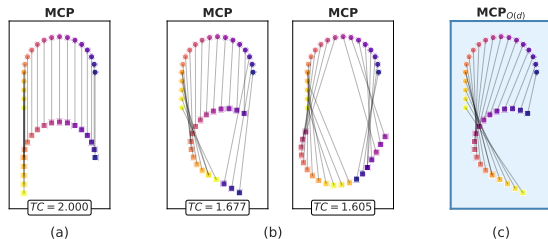
A Rotation-Invariant Proxy Pairing

Naive: the Maximum Correlation Pairing breaks under rotations of the input:

$$\text{MCP} = \arg \max_{P \in \mathcal{P}} \text{TC}(X, PY).$$

Fix: CCA picks the best orthogonal maps, so optimize over them too:

$$\text{MCP}_{O(d)} = \arg \max_{P \in \mathcal{P}} \max_{A, B \in O(d)} \text{TC}(XA, PYB).$$



(a) MCP works on aligned data, (b) but fails once the views are rotated. (c) $\text{MCP}_{O(d)}$ is rotation invariant and recovers the true pairing.

Takeaway

$\text{MCP}_{O(d)}$ defines a proxy pairing for Unpaired CCA

From $\text{MCP}_{O(d)}$ to the Quadratic Assignment Problem

KEY THEORETICAL INSIGHT

Unpaired CCA can be solved with a **linear-kernel QAP**

Our proxy (intractable): optimize over permutations *and* orthogonal matrices

$$\text{MCP}_{O(d)}(X, Y) = \arg \max_{P \in \mathcal{P}} \max_{A \in O(d), B \in O(d)} \text{TC}(XA, PYB).$$

Relaxation: replace $O(d)$ by the scaled Frobenius sphere

$$\text{MCP}_{S_F}(X, Y) = \arg \max_{P \in \mathcal{P}} \max_{A \in \sqrt{d} \mathbf{S}_F, B \in O(d)} \text{TC}(XA, PYB).$$

Theorem 1

$$\text{MCP}_{S_F}(X, Y) = \text{QAP}(XX^\top, YY^\top)$$

a classic problem with efficient approximate solvers

But does the relaxation change the pairing?

When Is the Relaxation Exact?

Proposition 1. With $M_P = X^\top P Y$, $\text{MCP}_{O(d)}$ is an optimization over P alone:

$$\text{MCP}_{O(d)} = \arg \max_{P \in \mathcal{P}} \|M_P\|_* = \arg \max_{P \in \mathcal{P}} \sum_{i=1}^d \sigma_i(M_P).$$

Let $P_O = \text{MCP}_{O(d)}$ and $P \neq P_O$:

By definition: the full sum

$$\sum_{i=1}^d \sigma_i(M_{P_O}) > \sum_{i=1}^d \sigma_i(M_P)$$

Assumption 1: every partial sum

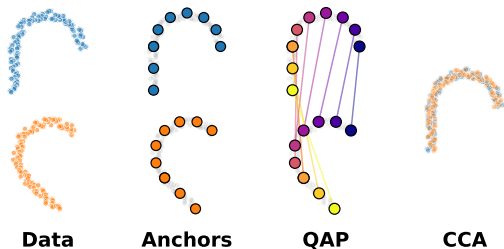
$$\sum_{i=1}^m \sigma_i(M_{P_O}) \geq \sum_{i=1}^m \sigma_i(M_P) \\ \text{for every } m = 1, \dots, d \text{ (weak majorization)}$$

Mild: holds exactly for isometries $Y = XQ$, and w.p. $\rightarrow 1$ as $n \rightarrow \infty$ vs. random permutations.

Theorem 2

Under Assumption 1, $\text{MCP}_{O(d)} = \text{MCP}_{S_F}$.
...and it holds in practice: verified empirically on all evaluated datasets.

UCCA: The Algorithm



Algorithm 1: Unpaired CCA

Input: Unpaired data X, Y

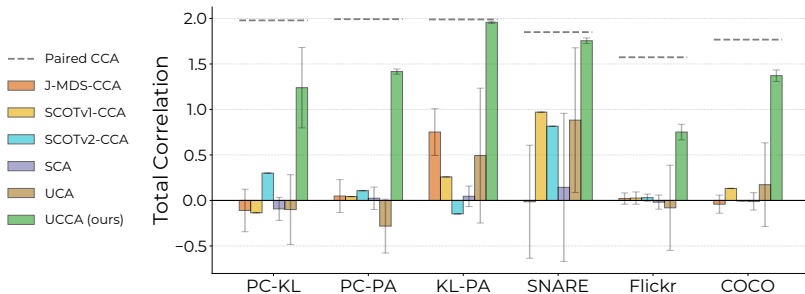
Output: Projections U, V

- 1 $Z_X, Z_Y \leftarrow \text{Whiten}(X, Y);$
 - 2 $A_X, A_Y \leftarrow \text{Anchors}(Z_X, Z_Y);$
 - 3 $P' \leftarrow \text{QAP}(A_X A_X^\top, A_Y A_Y^\top);$
 - 4 $U, V \leftarrow \text{CCA}(A_X, P' A_Y);$
 - 5 **return** U, V
-

- The QAP is solved on a few **K-Means anchors** per view, so it stays cheap.
- Repeated over many clusterings: matched anchors serve as **pseudo-pairs** for a standard CCA.

Recovering the True Correlation

Training views come from **disjoint halves** of each dataset, so no pair is ever seen. We report the total correlation of the hidden true pairs on held-out data.



Takeaway

UCCA is the only method that consistently recovers the true correlation

Unsupervised Classification

Accuracy of a kNN classifier fit on one view's embedding and evaluated on the other. This enables classifying an unlabeled view without any paired data.

Method	PC-KL	PC-PA	KL-PA	SNARE	COCO
J-MDS-CCA	0.083	0.087	<u>0.211</u>	0.247	0.237
SCOTv1-CCA	0.074	<u>0.121</u>	0.102	<u>0.650</u>	0.141
SCOTv2-CCA	<u>0.148</u>	0.090	0.113	0.646	0.164
SCA	0.097	0.083	0.100	0.312	0.161
UCA	0.088	0.100	0.159	0.378	0.132
UCCA (ours)	0.292	0.315	0.564	0.849	<u>0.232</u>
Paired CCA	0.562	0.516	0.629	0.887	0.232

Conclusion

- **Theory:** Unpaired CCA reduces to a linear-kernel QAP: the tractable MCP_{S_F} coincides with the rotation-invariant $\text{MCP}_{O(d)}$ under a mild majorization assumption.
- **Practice:** UCCA recovers canonical correlations from strictly unpaired data, significantly outperforming unpaired alignment baselines.
- **More in the paper:** UCCA is efficient (QAP runs only on anchors) and keeps improving with more unpaired samples.



shaham-lab.github.io/UCCA_page
github.com/shaham-lab/UCCA

Thank you!