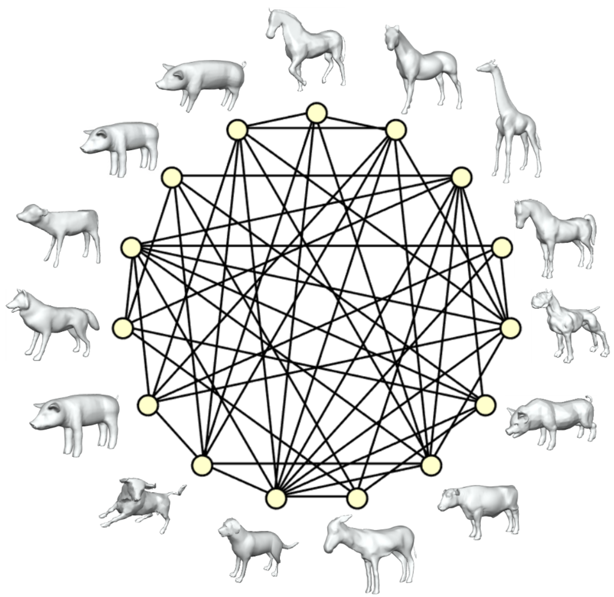
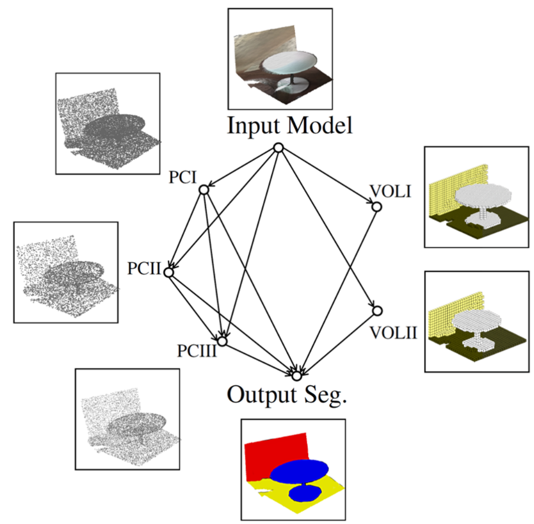
Guest Lecture on Cycle-Consistency and Synchronization

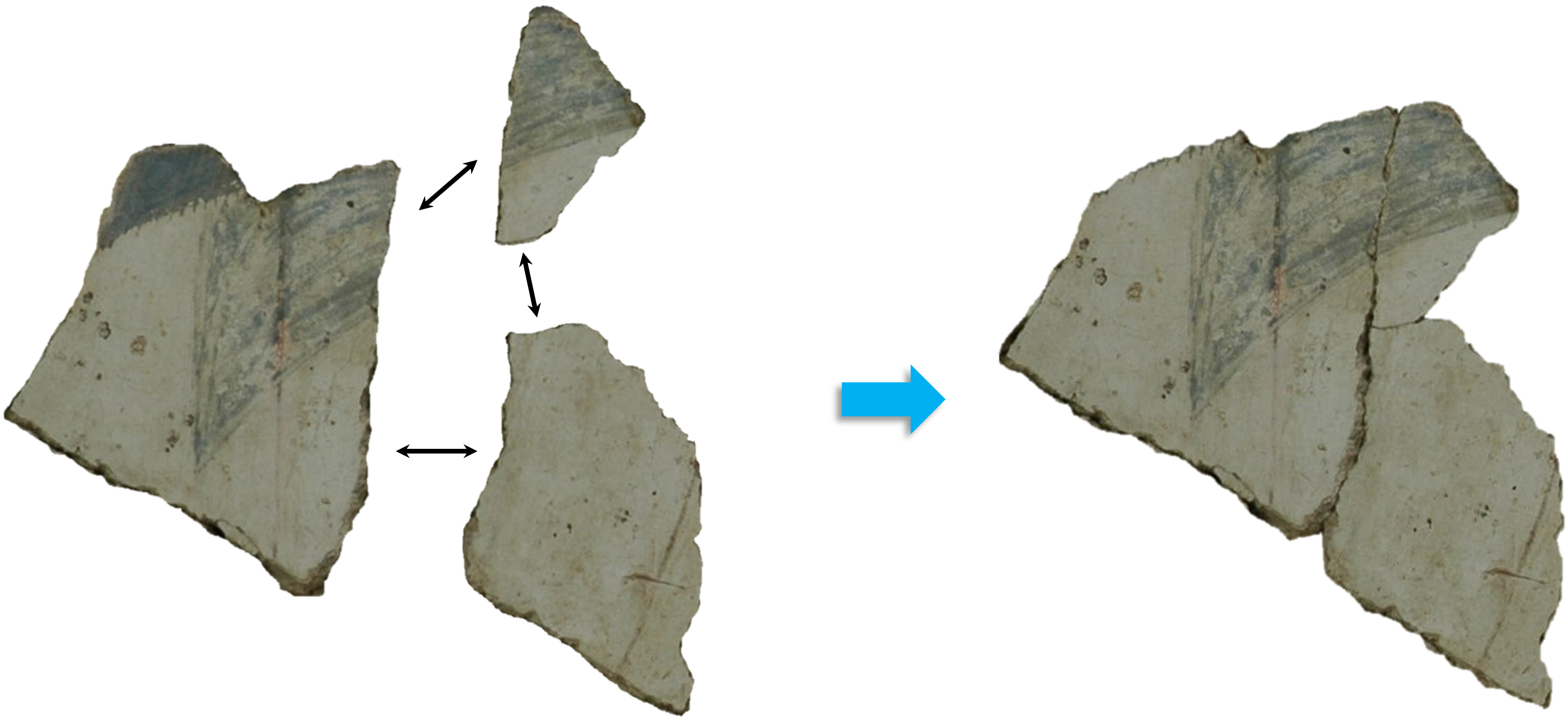




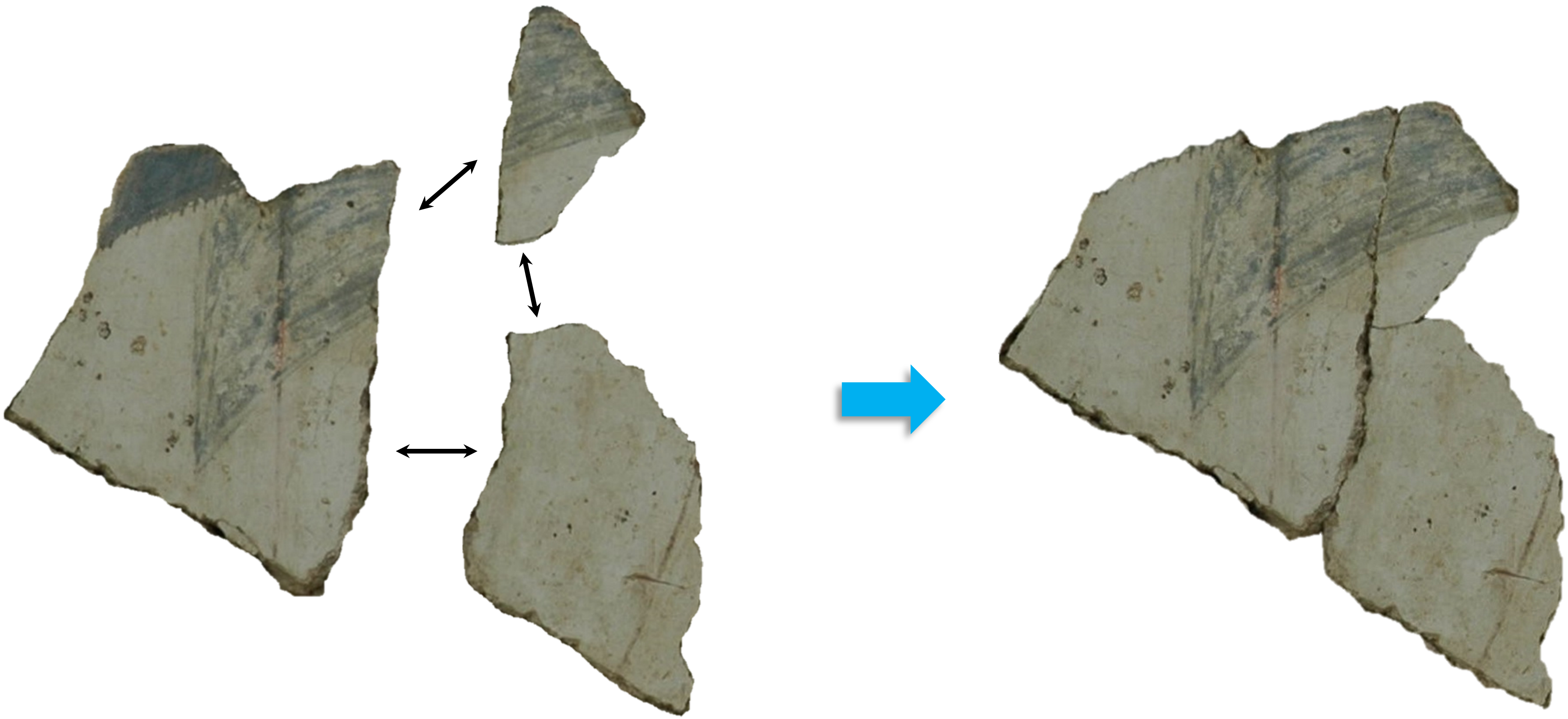
Motivations of Map Synchronization

Ambiguities in assembling pieces

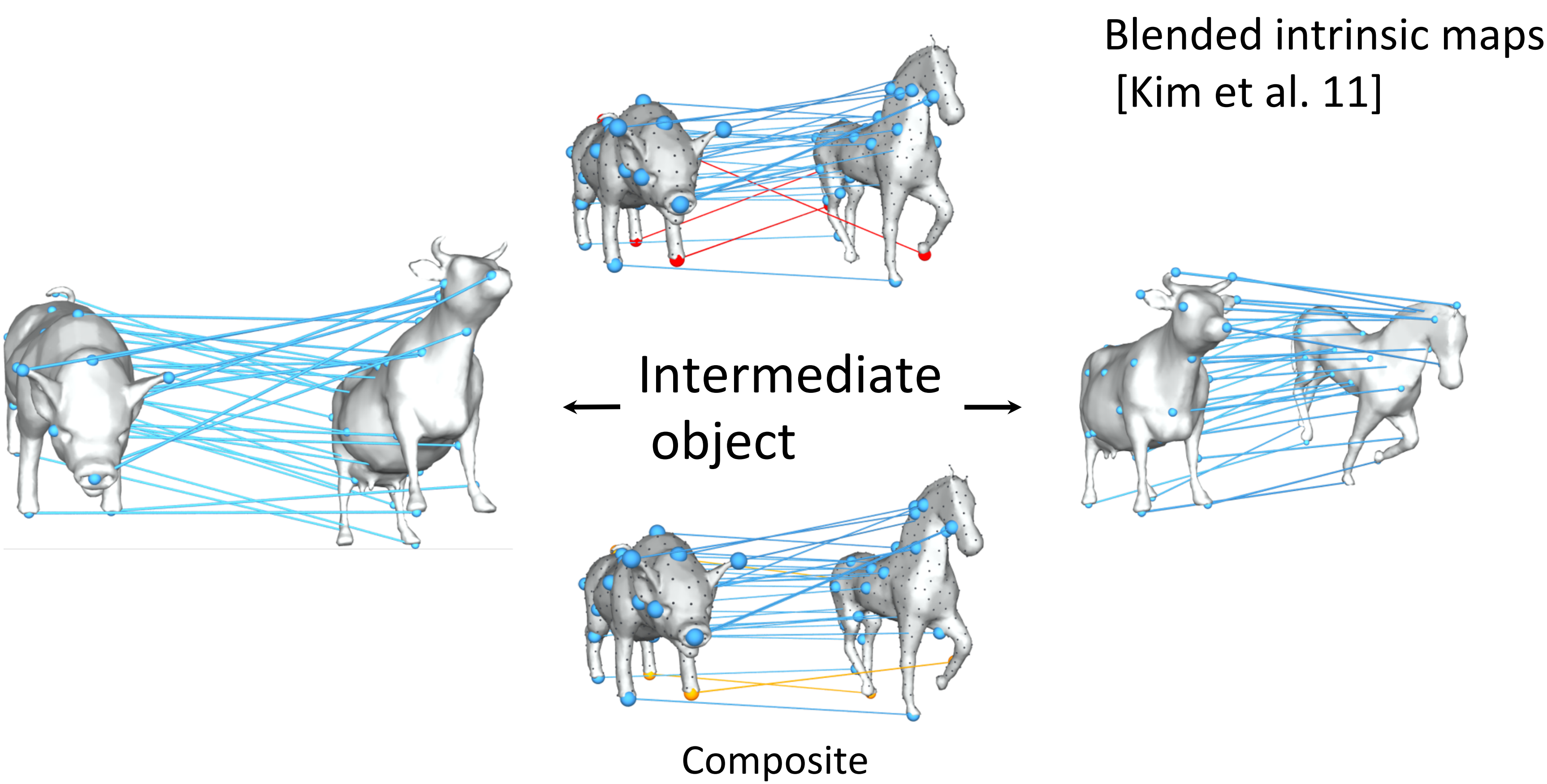
Resolving ambiguities by looking at additional pieces



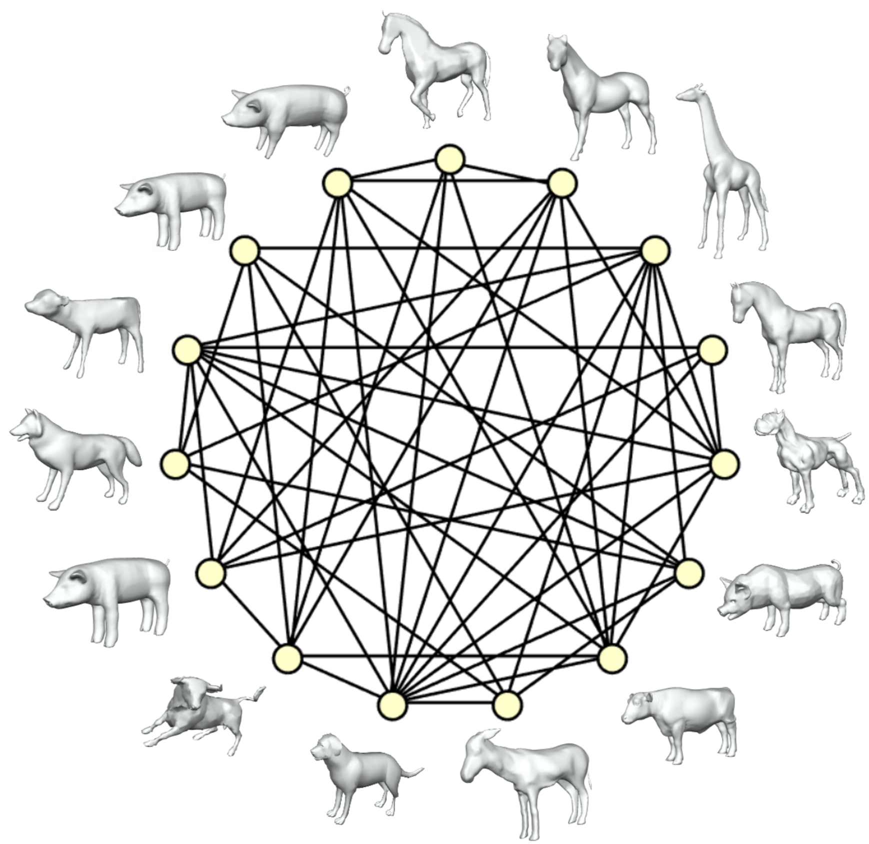
Resolving ambiguities by looking at additional pieces



Matching through intermediate objects --- map propagation

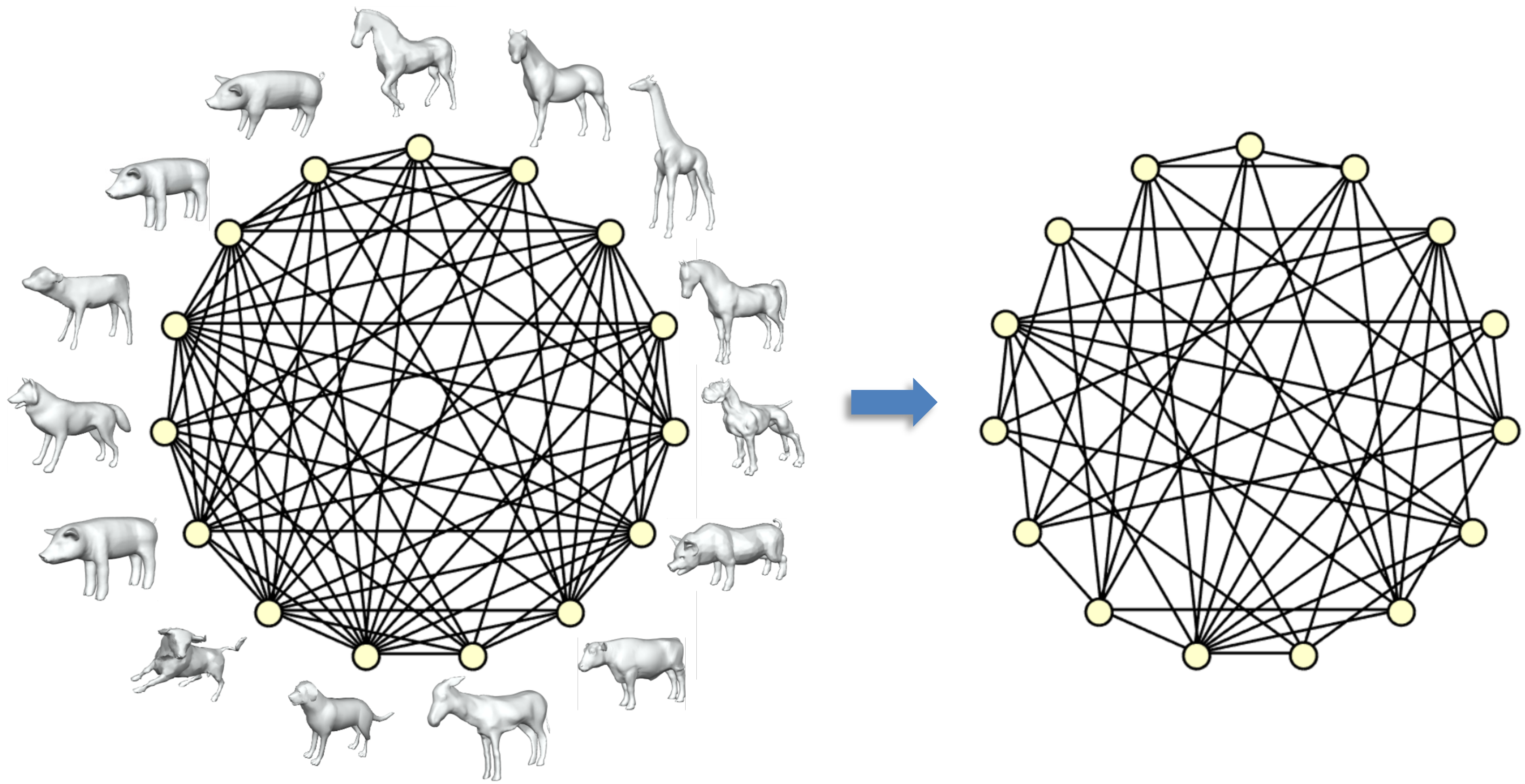


Pair-wise maps usually contain enough information



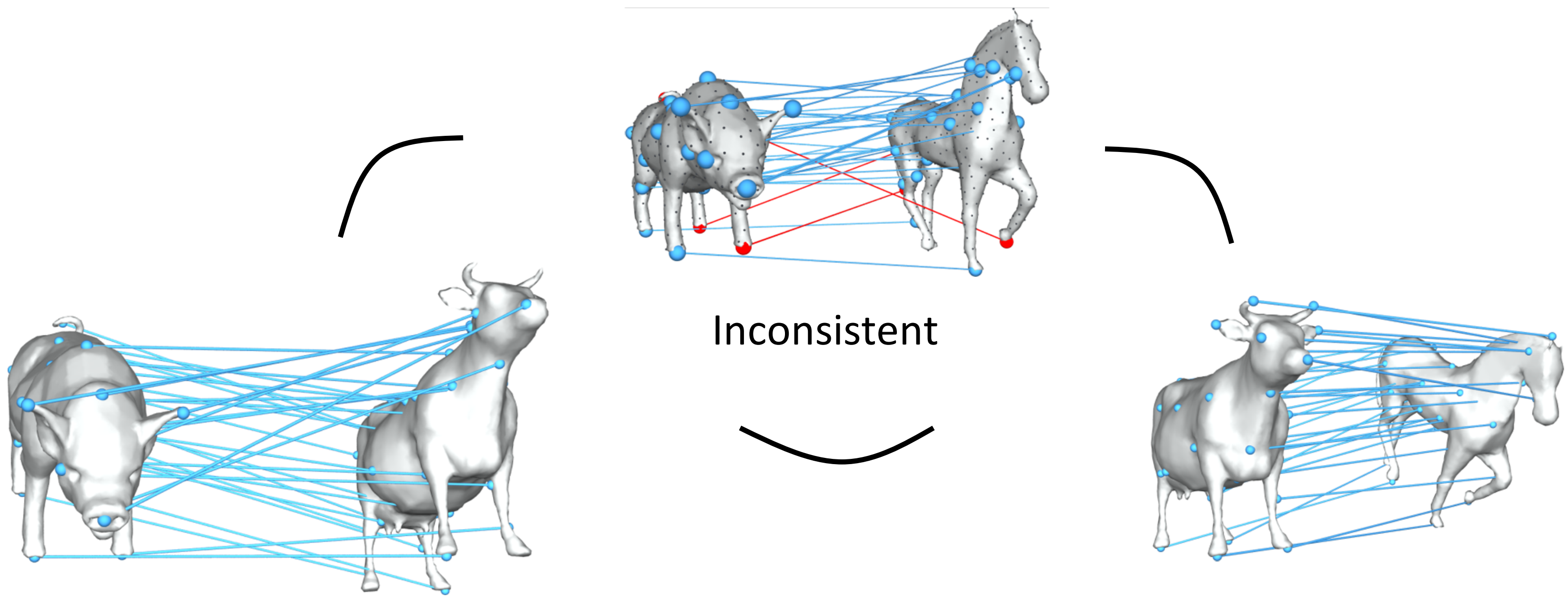
Network of approximately correct blended intrinsic maps

Map synchronization problem

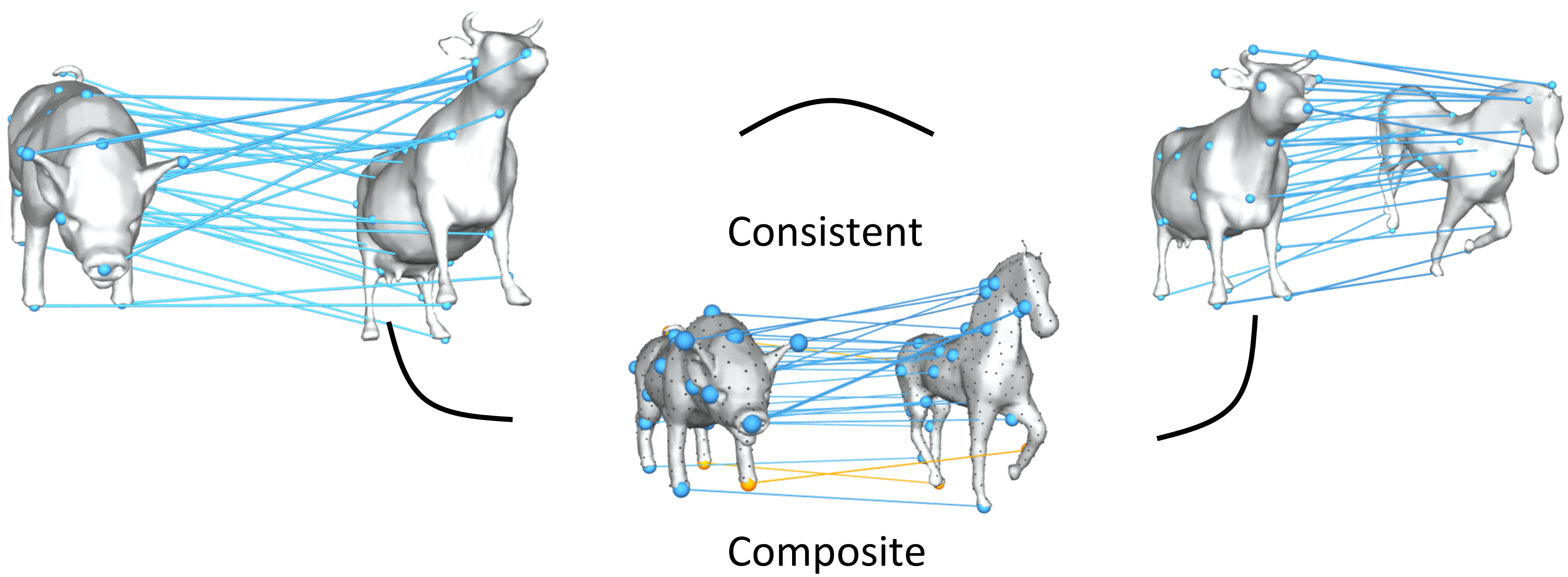


Identify correct maps among a (sparse) network of maps

A natural constraint on maps is that they should be consistent along cycles



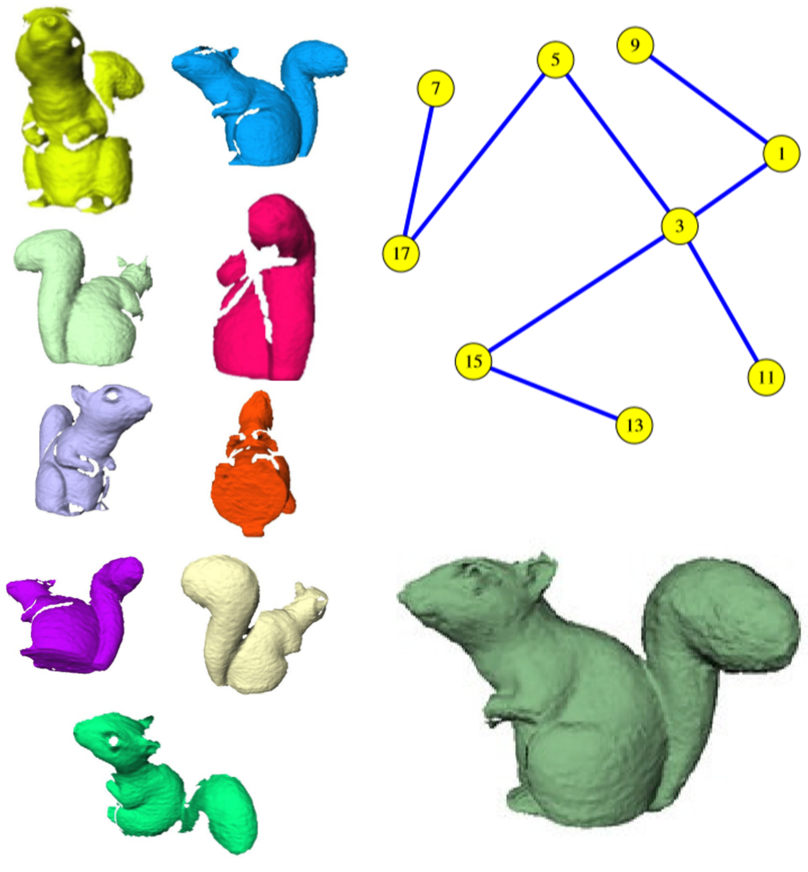
A natural constraint on maps is that they should be consistent along cycles



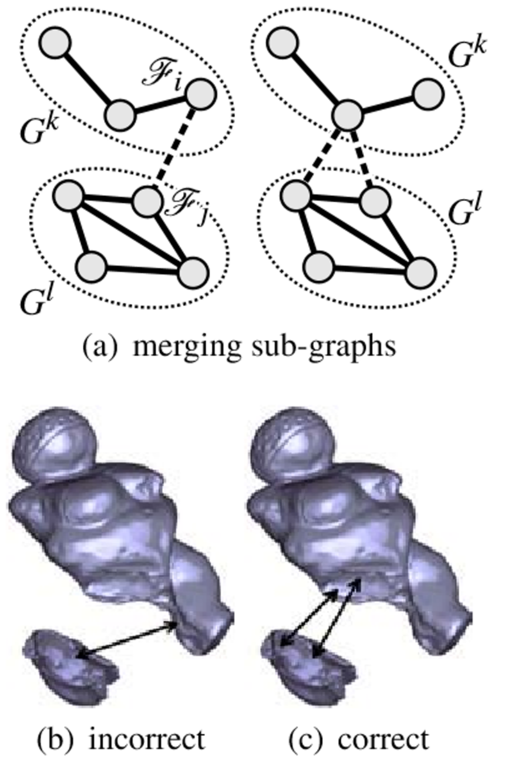
Literature on utilizing the cycle-consistency constraint

- Spanning tree optimization [Huber et al. 01, Huang et al. 06, Cho et al. 08, Crandel et al. 11, Huang et al. 12]

Greedy algorithm for spanning tree computation



[Huber and Hebert 02]



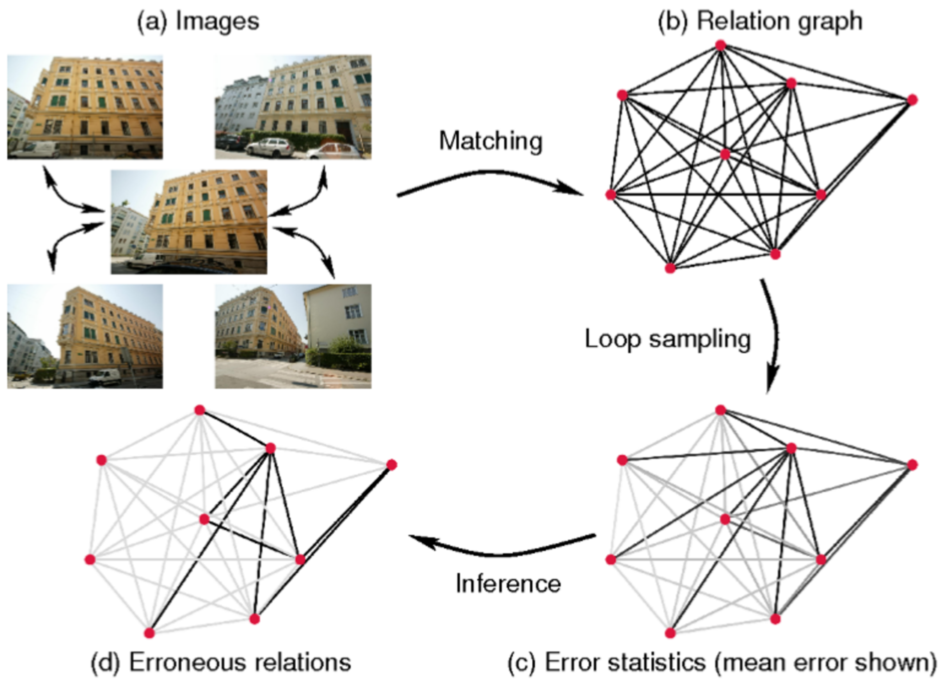
[Huang et al. 06]

Literature on utilizing the cycle-consistency constraint

- Spanning tree optimization [Huber et al. 01, Huang et al. 06, Cho et al. 08, Crandel et al. 11, Huang et al. 12]

- Sampling inconsistent cycles [Zach et al. 10, Nyugen et al. 11, Zhou et al. 15]

Linear programming formulation [Zach et al. 10]



min $\sum_{e} \rho_e x_e + \sum_{L} \rho_L x_L$

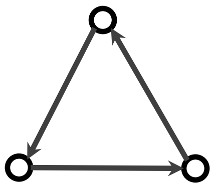
s.t. $x_L \geq x_e \; \forall e \in L$

$x_L \leq \sum_{e \in L} x_e$

$x_L \in [0, 1]$

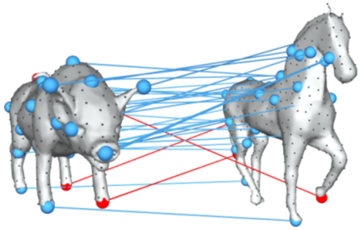
$x_e \in [0, 1]$

Compressive sensing view of map synchronization



Cycle-consistency

Compressible



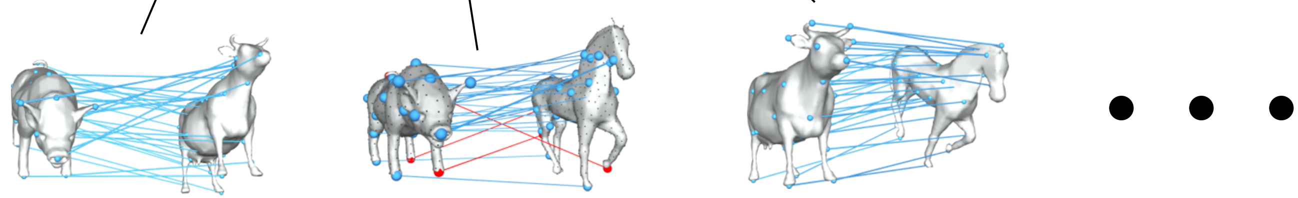
Input maps

Noisy observations

Map synchronization as constrained matrix optimization

[HG13]

$$X = \begin{bmatrix} I_m & X_{12} & \cdots & X_{1n} \\ X_{21} & I_m & \cdots & \vdots \\ \vdots & \vdots & \ddots & X_{n-1,n} \\ X_{n1} & \cdots & X_{n,n-1} & I_m \end{bmatrix}$$

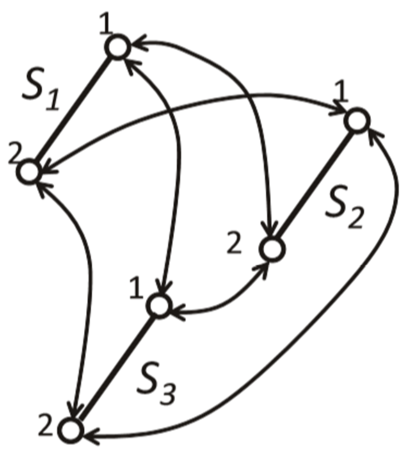


Noisy measurements of matrix blocks

The equivalence among cycle-consistency, low-rankness, and SDP

[HG13]

- The following three statements are equivalent:
  - The maps are cycle-consistent
  - X is low-rank and the rank equals to #points per surface
  - X is positive semidefinite



$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \succeq 0$$

Example: permutation synchronization

[HG13]

Objective function:

$$\text{minimize} \sum_{(i,j)\in\mathcal{G}} \|X_{ij}^{\text{input}} - X_{ij}\|_1$$

Observation graph

Constraints:

$$X \succeq 0$$

cycle-consistency

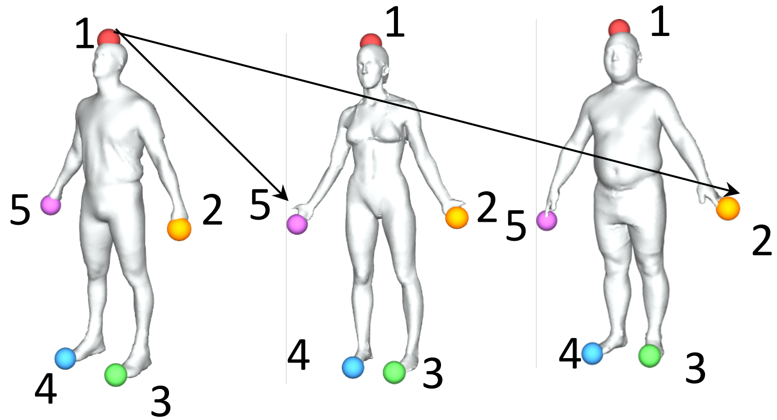
$$\boxed{\begin{array}{l} X_{ii} = I_m, \quad 1 \le i \le n \\ X_{ij}\mathbf{1} = \mathbf{1}, X_{ij}^T\mathbf{1} = \mathbf{1}, \quad 1 \le i < j \le n \\ 0 \le X \le 1 \end{array}}$$

mapping constraint

Deterministic guarantee

- *Theorem[HG13]: Given noisy input maps, permutation synchronization recovers the underlying maps if*

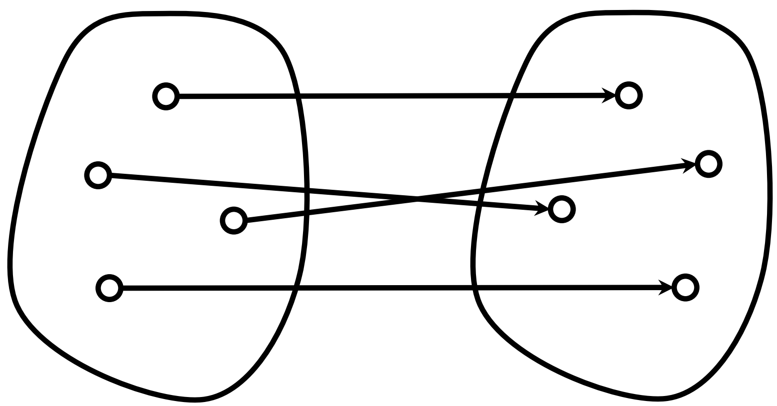
$$\#\text{incorrect corres. of each point} < \frac{\lambda_2(G)}{4}$$



Optimality when the object graph G is a clique

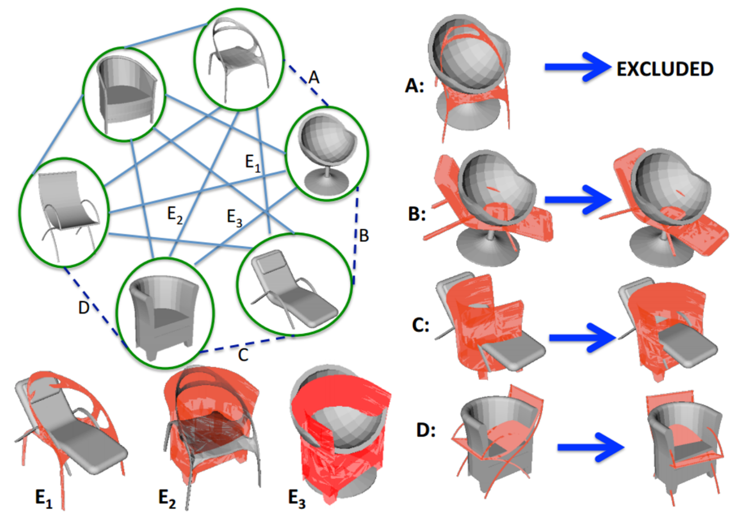
[HG13]

- 25% incorrect correspondences
- Worst-case scenario
  - Two clusters of objects of equal size
  - Wrong correspondences between objects of different clusters only (50%)



Justification of maximizing $\lambda_2(G)$ for map graph construction

Imageweb [Heath et al 10]



Fuzzy correspondences on shapes [Kim et al 12]

Randomized setting

[CGH14]

- Generalized Erdős–Rényi model:
  - p_{obs}: the probability that two objects connect
  - p_{true}: the probability that a pair-wise map is correct
  - Incorrect maps are random permutations

- Theorem [CGH14]: *The underlying permutations can be recovered w.h.p if*

$$p_{\mathrm{true}} \geq c\frac{\log^2(mn)}{\sqrt{np_{\mathrm{obs}}}}$$

Optimality when m is a constant

- Exact recovery condition:

$$p_{\mathrm{true}} > c\frac{\log^2(n)}{\sqrt{np_{\mathrm{obs}}}}$$

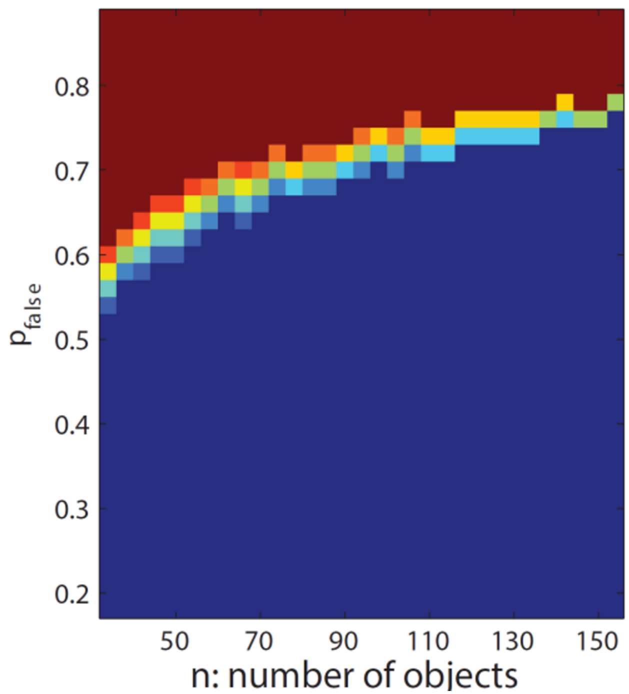
- Information theoretic limits [Chen et al 15]:

No method works if $p_{\mathrm{true}} \le c_1 \dfrac{1}{\sqrt{np_{\mathrm{obs}}}}$

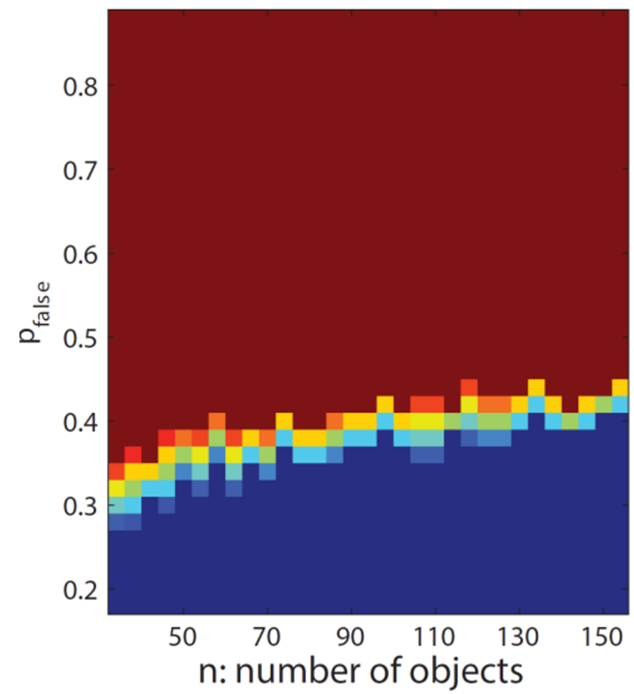
Comparison to a generic low-rank matrix recovery method

[CGH14]

Permutation synchronization



RPCA [Candes et al. 09]



Phase transitions in empirical success probability (p_{obs} =1)

Noise distribution when perturbing permutations

[CGH14]

- RPCA can handle dense corruption if the perturbations exhibit random sign pattern, yet

$$E_{\mathcal{P}_m}\big(\mathrm{sgn}\big(X_{ij} - I_m\big)\big) = -I_m + \frac{1}{m}\mathbf{1}\mathbf{1}^T$$

- The map constraints incur a quotient space defined by

$$\mathcal{K} = \{Z : |Z \in \mathbb{R}^{m\times m},\ Z\mathbf{1} = 0,\ Z^T\mathbf{1} = 0\}$$

- The expectation under this quotient space

$$E_{\mathcal{P}_m/\mathcal{K}}\big(\mathrm{sgn}\big(X_{ij} - I_m\big)\big) = 0$$

Partial point-based map synchronization

[CGH14]

Step I: Spectral method:

m <= #dominant eigenvalues of X^{input} after trimming

Step II:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \sum_{(i,j)\in\mathcal{G}} \langle \lambda \mathbf{1}\mathbf{1}^T - 2X_{ij}^{\text{input}}, X_{ij} \rangle \\ \text{subject to} \quad & X_{ii} = I_{m_i}, \quad 1 \le i \le n \\ & 0 \le X \le 1 \\ & \begin{bmatrix} m & \mathbf{1}^T \\ \mathbf{1} & X \end{bmatrix} \succeq 0 \end{aligned}$$

Size of the universe

Exact recovery condition

[CGH14]

- Randomized model: n objects, universe size m
  - Each object contains a fraction p_{set} of m elements
  - Each pair is observed w.p. p_{obs}
  - Each observed is randomly corrupted w.p. $1 - p_{true}$

- Theorem. When $\lambda \in [\frac{1}{m}, \frac{1}{\sqrt{p_{obs}}}]$, the underlying maps can be recovered with high probability if

$$p_{true} \geq c_2 \frac{\log^2(mn)}{p_{\text{set}}^2 \sqrt{n p_{\text{obs}}}}$$

Spectral Map Synchronization

Intuition

$$X^{\text{observation}} = X^{\text{ground-truth}} + X^{\text{noise}}$$

David-Kham theorem:

$$\|U_m(X^{obs}) - U_m(X^{gt})\| \leq \frac{\|X^{noise}\|}{\lambda_m(X^{gt}) - \lambda_{m+1}(X^{gt})}$$

Algorithm

[Pachauri et al 13, Shen et al 16]

- Step I: Leading eigen-vector computation
  - Power method, which can be done very efficiently

- Step II: Rounding via linear assignment
  - Hungarian algorithm

Theoretical Analysis

- Deterministic setting
  - A constant fraction of noise [Huang et al. 19]
  - 1/8 for clique graphs (a gap from SDP formulations)

- Randomized setting [Bajaj et al. 18]

$$p \geq O\left(\sqrt{\frac{\log(n)}{nq}}\right)$$

Fraction of correct maps

Sampling density

Non-Convex Optimization

Translation Synchronization

- Pair-wise differences along a graph [Huang et al. 17]
- Convex optimization

$$\text{minimize} \sum_{(i,j)\in\mathcal{E}} |t_{ij} - (x_i - x_j)|, \quad \text{subject to} \quad \sum_{i=1}^{n} x_i = 0$$

- Truncated least squares

$$\{x_i^{(k)}\} = \underset{\{x_i\}}{\text{argmin}} \sum_{(i,j)\in\mathcal{E}} w_{ij}|t_{ij} - (x_i - x_j)|^2, \quad \text{subject to} \quad \sum_{i=1}^{n} \sqrt{d_i} x_i = 0, \quad d_i := \sum_{j\in\mathcal{N}(i)} w_{ij}$$

$$w_{ij} = Id(|t_{ij} - (x_i^{(k-1)} - x_j^{(k-1)})| < \delta_k)$$

Exact recovery condition

- Deterministic
  - A constant fraction of noise (1/6 for clique graphs)
  - 2/3 of the optimal ratio
- Randomized

$$t_{ij} = \begin{cases} x_i^{gt} - x_j^{gt} + U[-\sigma, \sigma] & \text{with probability } p \\ x_i^{gt} - x_j^{gt} + U[-a, b] & \text{with probability } 1-p \end{cases}$$

Exact recovery if $p > c/\sqrt{\log(n)},$

Summary of low-rank based techniques

$$X^{\text{observation}} = X^{\text{ground-truth}} + X^{\text{noise}}$$

Recovery if **In some reduced space**

$$\text{spectral-gap}(X^{\text{ground-truth}}) \geq c\|X^{\text{noise}}\|$$

The constant depends on the optimization techniques being used

Many (non-convex) techniques require further understanding!

Learning Transformation Synchronization

[With X. Huang, Z. Liang, X. Zhou, X. Yao, L. Guibas]

Hand-crafted objective function

[HG13]

Objective function:

$$\text{minimize} \sum_{(i,j)\in\mathcal{G}} \|X_{ij}^{\text{input}} - X_{ij}\|_1$$

Observation graph

Constraints:

$$X \succeq 0$$

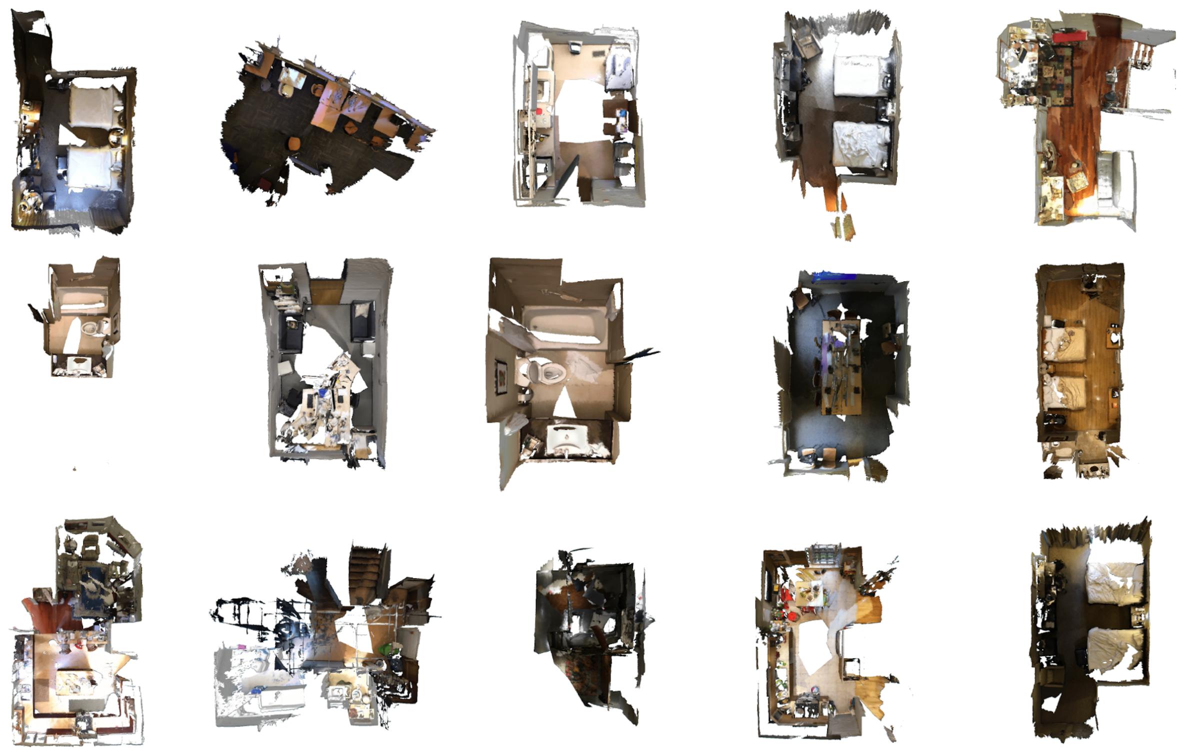
cycle-consistency

$$\boxed{\begin{array}{l} X_{ii} = I_m, \quad 1 \leq i \leq n \\ X_{ij}\mathbf{1} = \mathbf{1}, X_{ij}^T\mathbf{1} = \mathbf{1}, \quad 1 \leq i < j \leq n \\ 0 \leq X \leq 1 \end{array}}$$

mapping constraint

3D scene reconstruction from depth scans

[Dai et al. 17]



- Similar noise sources
  - Scanning noise, frame rate, and symmetry structures

Reweighted least square synchronization

Rotation:

$$\underset{R_i \in SO(3), 1 \le i \le n}{\text{minimize}} \sum_{(i,j)\in\mathcal{E}} w_{ij} \|R_{ij}R_i - R_j\|_{\mathcal{F}}^2$$

Solved by the first 3 eigenvectors of a Connection Laplacian

$$L_{ij} := \begin{cases} \sum\limits_{j\in\mathcal{N}(i)} w_{ij} I_3 & i = j \\ -w_{ij}R_{ij}^T & (i,j)\in\mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

Translation

:

$$\underset{\boldsymbol{t}_i, 1 \le i \le n}{\text{minimize}} \sum_{(i,j)\in\mathcal{E}} w_{ij} \|R_{ij}\boldsymbol{t}_i + \boldsymbol{t}_{ij} - \boldsymbol{t}_j\|^2$$

Linear system:

$$\boldsymbol{t}^\star = L^+ \boldsymbol{b}$$

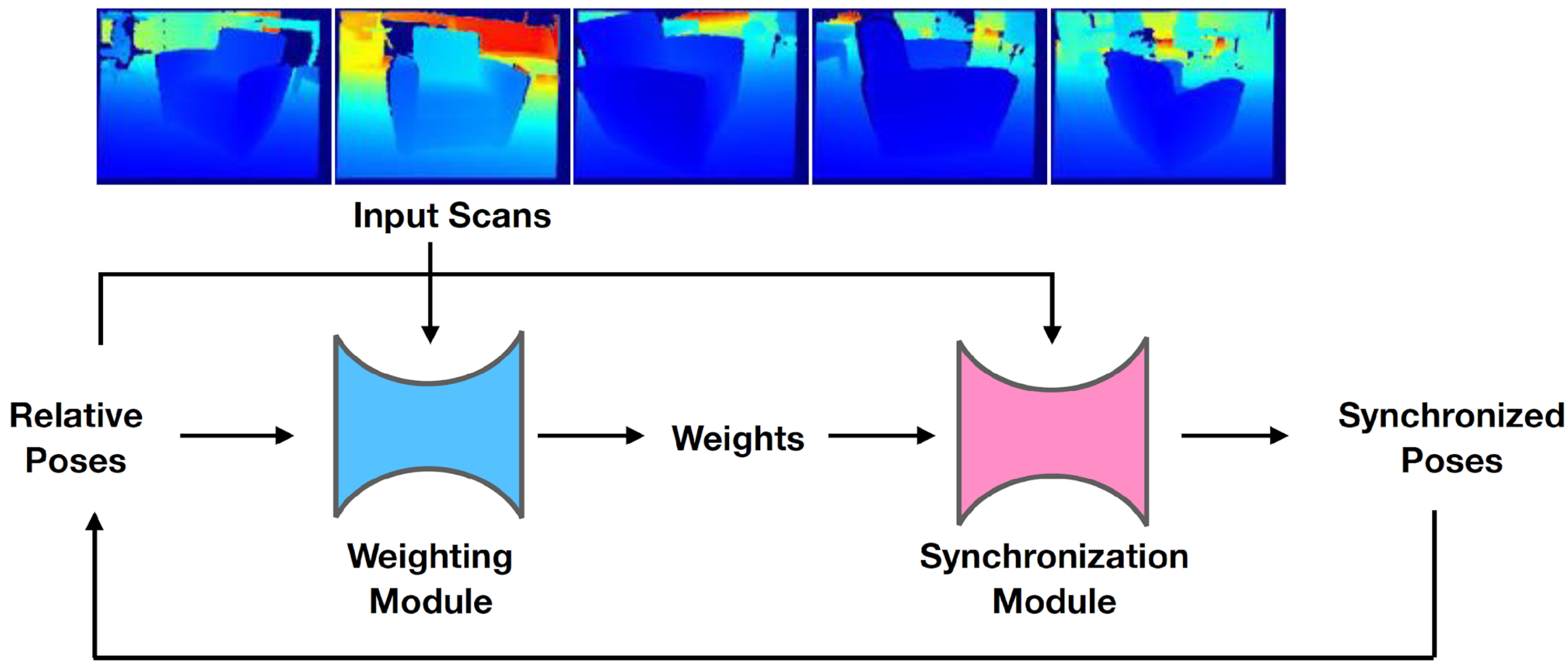
Where

$$\boldsymbol{b}_i := -\sum_{j\in\mathcal{N}(i)} w_{ij} R_{ij}^T \boldsymbol{t}_{ij}$$

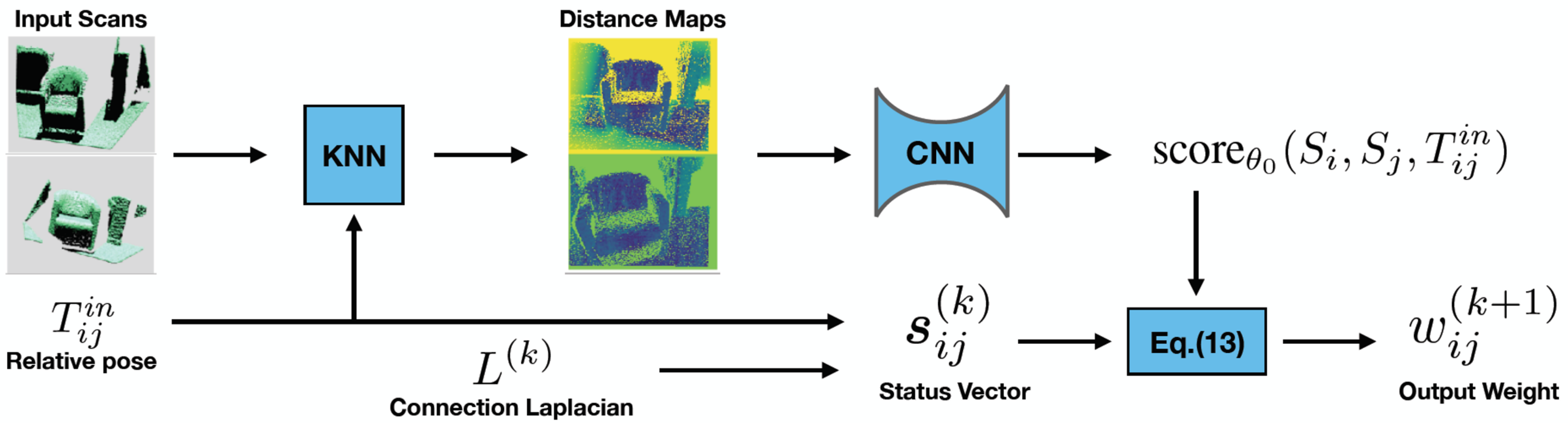
Robust recovery under a constant fraction of adversarial noise if

$$w_{ij} = \rho(\|R_{ij}R_i^{(k)} - R_j^{(k)}\|) \quad \text{where} \quad \rho(x) = \frac{\epsilon^2}{\epsilon^2 + x^2}$$

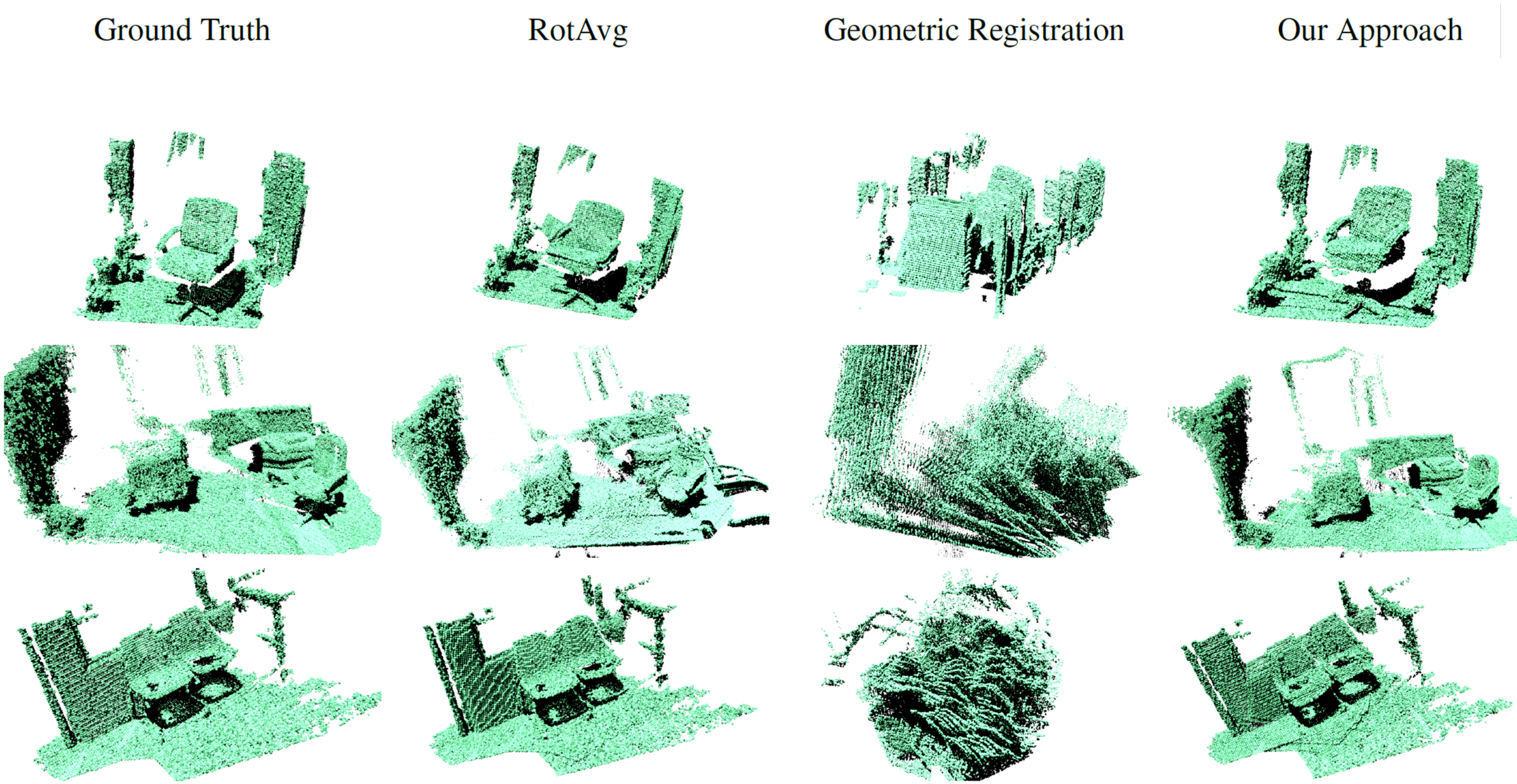
Network design



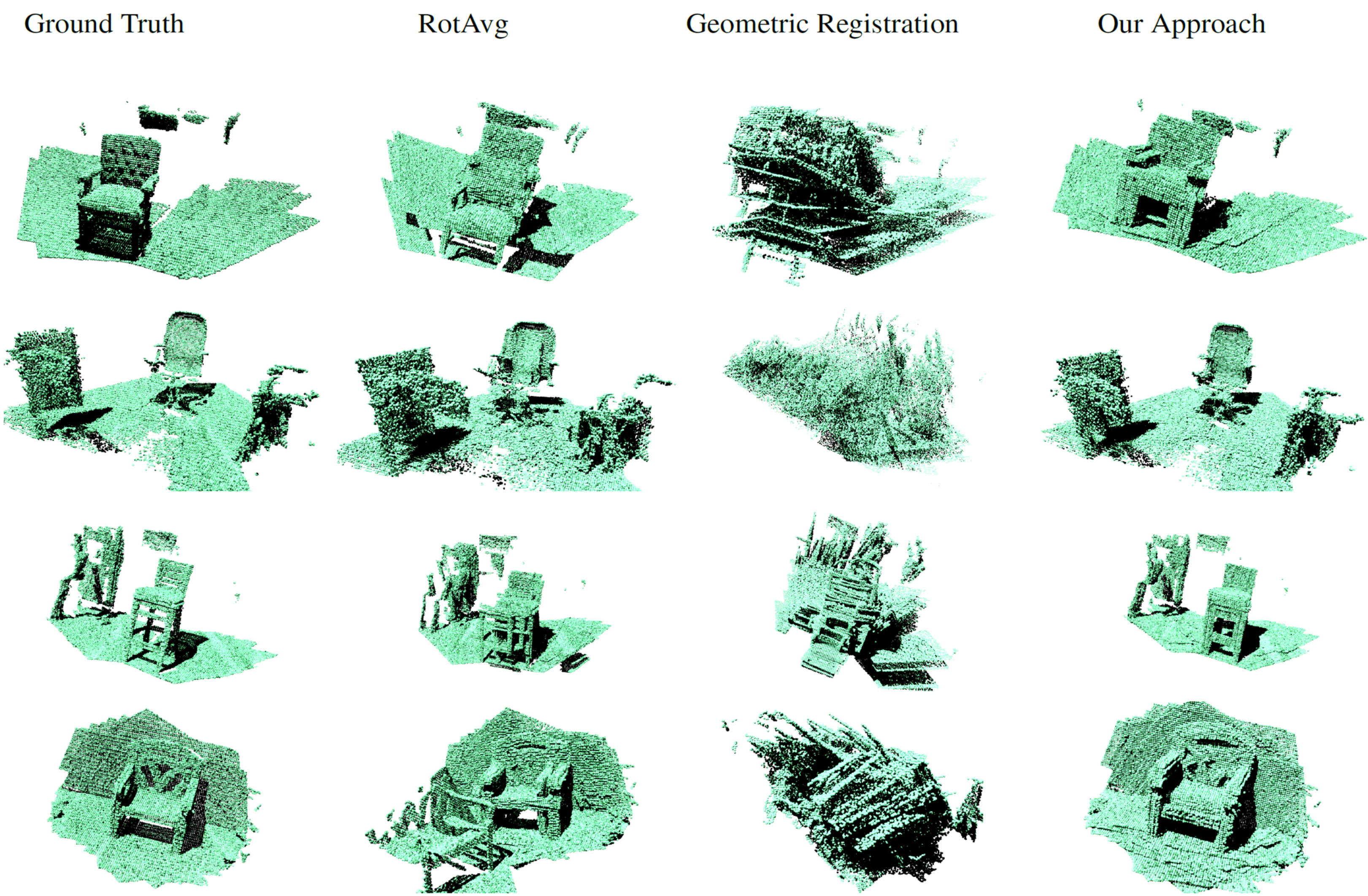
Weighting module



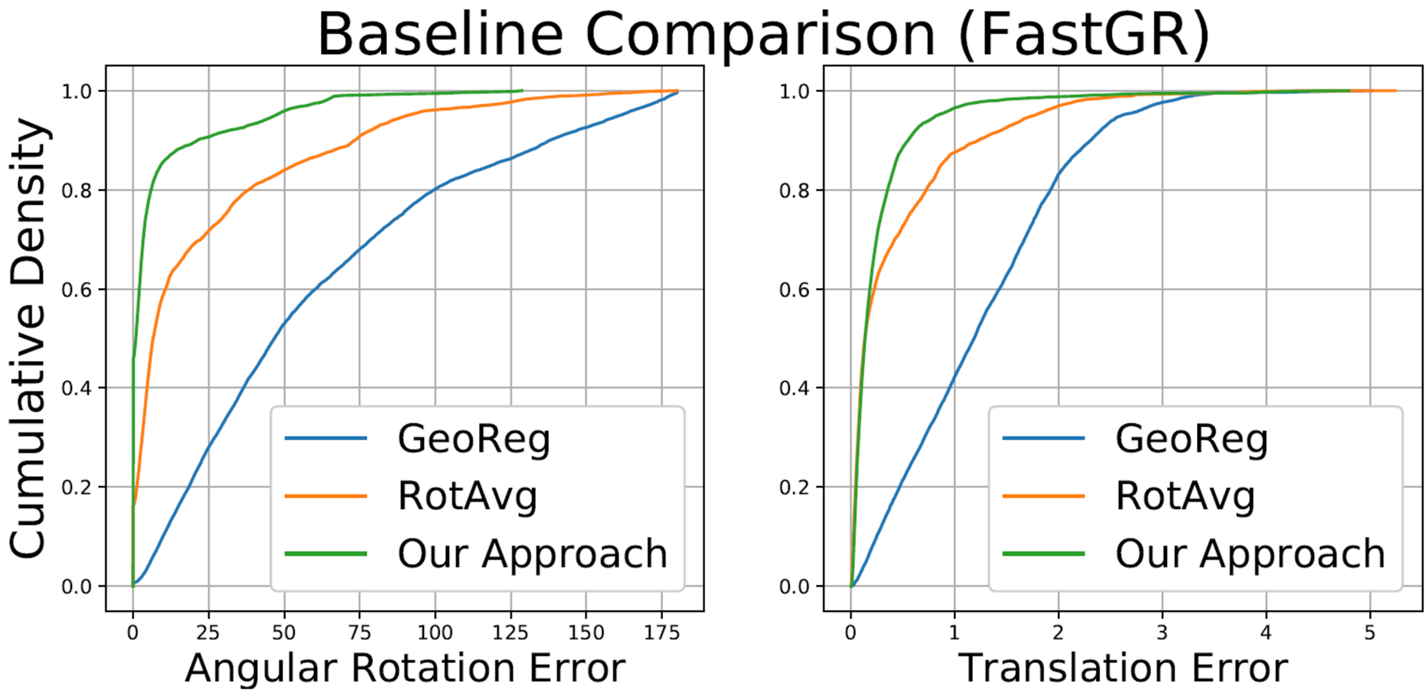
Qualitative results



Qualitative results



Quantitative results



Redwood dataset

Further reading (a partial list)

- Uncertainty quantification, Rotation/transformation synchronization, and lower bounds

1. T. Birdal, U. Simsekli. Probabilistic Permutation Synchronization using the Riemannian Structure of the Birkhoff Polytope. CVPR 2019.

2. T. Birdal, U. Simsekli, M. Eken, S. Ilic. Bayesian Pose Graph Optimization via Bingham Distributions and Tempered Geodesic MCMC. In NIPS 2018.

3. A. Perry, J. Weed, A. S. Bandeira, P. Rigollet, A. Singer, "The sample complexity of multi-reference alignment". SIAM Journal on Mathematics of Data Science

4. O. Özyeşil, N. Sharon, A. Singer, ``Synchronization over Cartan motion groups via contraction", SIAM Journal on Applied Algebra and Geometry, 2 (2), pp. 207-241 (2018)

5. A. S. Bandeira, N. Boumal, A. Singer, ``Tightness of the maximum likelihood semidefinite relaxation for angular synchronization", Mathematical Programming, series A, 163 (1):145-167 (2017).

6. A. Singer, H.-T. Wu, ``Spectral Convergence of the Connection Laplacian from Random Samples", Information and Inference: A Journal of the IMA, 6 (1):58-123 (2017).

Further reading (a partial list)

- Uncertainty quantification, Rotation/transformation synchronization, and lower bounds

7. K. N. Chaudhury, Y. Khoo, A. Singer, ``Global registration of multiple point clouds using semidefinite programming”, SIAM Journal on Optimization, 25 (1), pp. 468-501 (2015).

8. N. Boumal, A. Singer, P.-A. Absil and V. D. Blondel, ``Cramér-Rao bounds for synchronization of rotations”, Information and Inference: A Journal of the IMA, 3 (1), pp. 1--39 (2014).

9. A. Singer, ``Angular Synchronization by Eigenvectors and Semidefinite Programming'', Applied and Computational Harmonic Analysis, 30 (1), pp. 20-36 (2011).

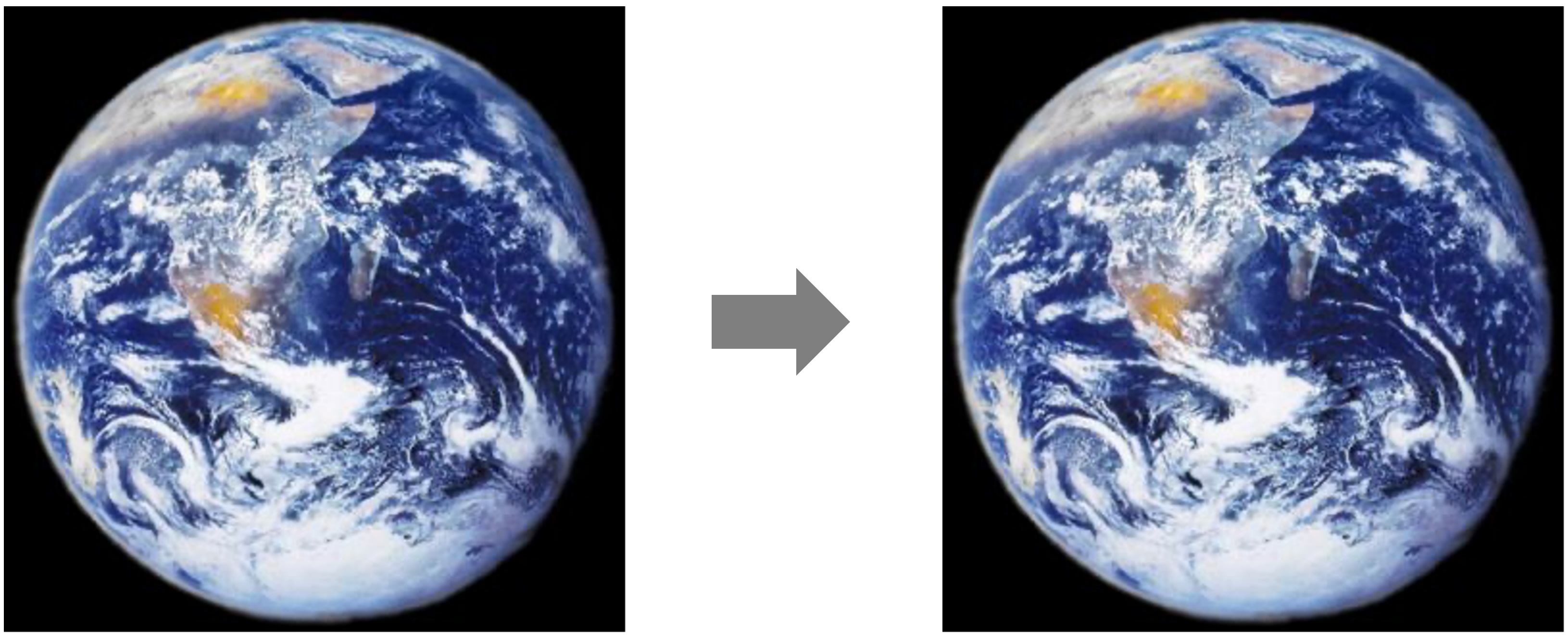
10. SE-Sync: A Certifiably Correct Algorithm for Synchronization over the Special Euclidean Group David M. Rosen, Luca Carlone, Afonso S. Bandeira, and John J. Leonard. (2018)

11. Robust synchronization in SO (3) and SE (3) via low-rank and sparse matrix decomposition. Federica Arrigoni, Beatrice Rossi, Pasqualina Fragneto, Andrea Fusiello. Computer Vision and Image Understanding. 174. pp. 95-113 (2018)

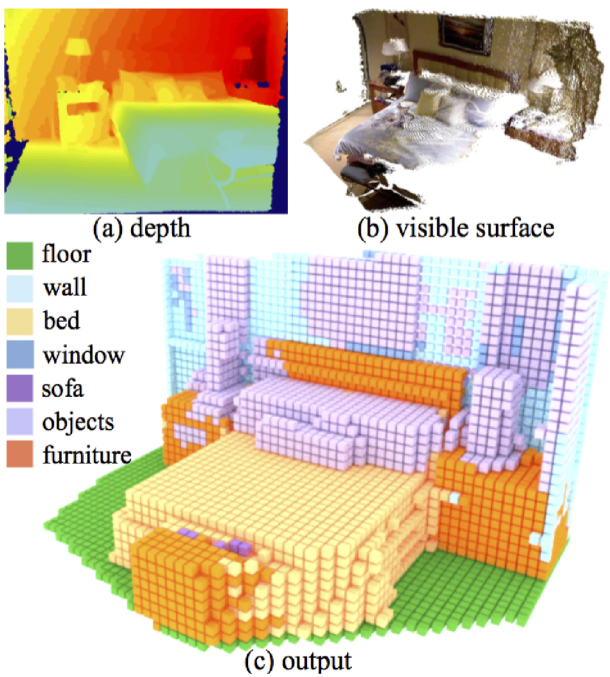
Neural networks as maps

Neural networks are maps

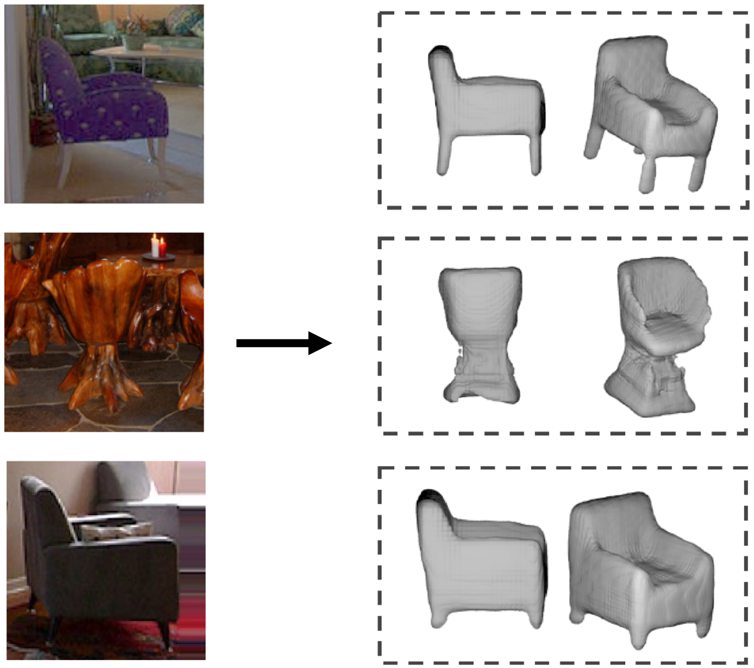
- Approximate any function given sufficient data



Monocular reconstruction



Semantic scene completion [Song et al. 17]



MarrNet [Wu et al. 17]

Space of images → Space of 3D models

Image Captioning

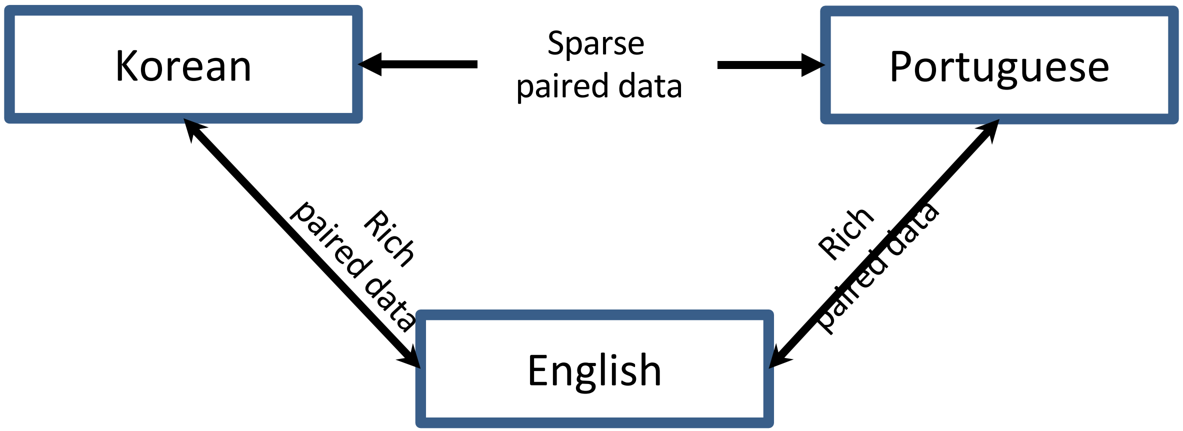
Space of images → Space of natural language descriptions

Joint Learning of Neural Networks

Advantage I: Leverage more training data

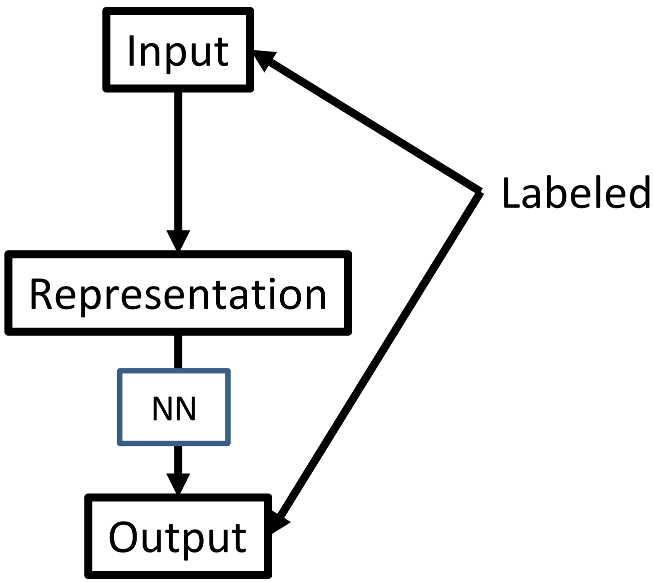
A toy example

[Johnson et al. 16]

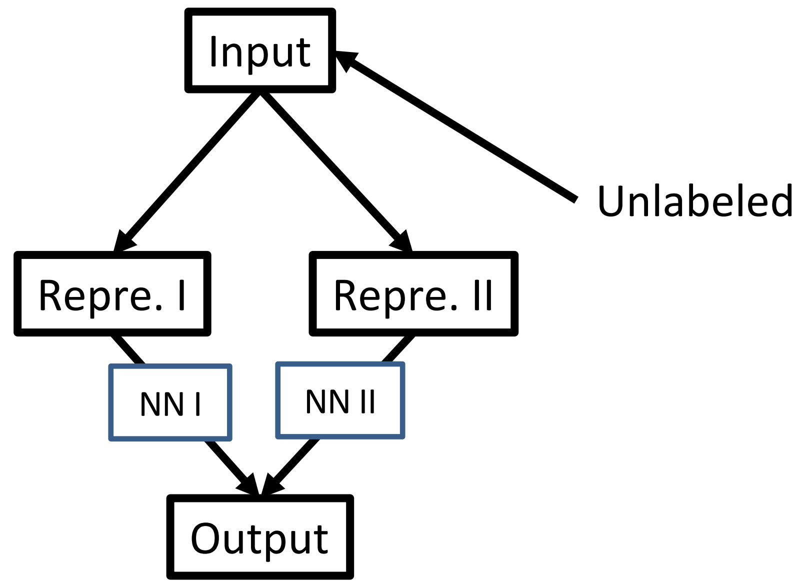


Advantage II: Leverage Unlabeled Data

A toy example

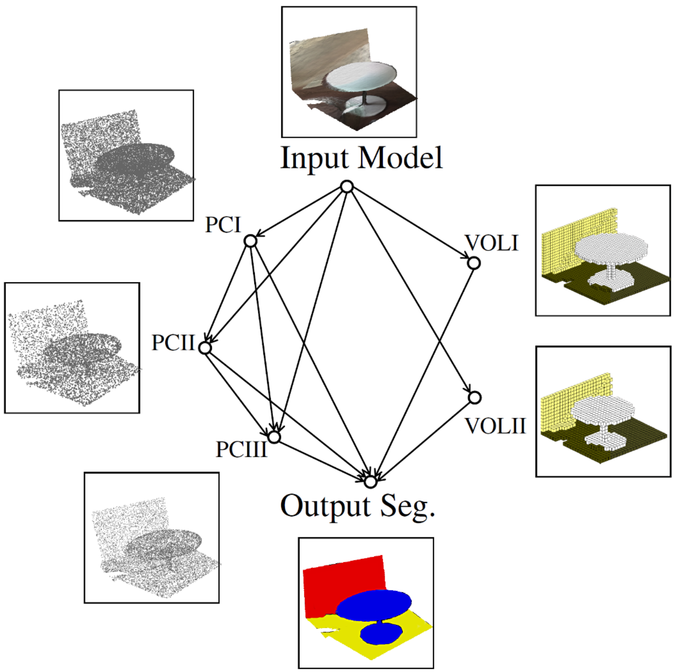


Standard setting:



Joint setting:

Limitations of low-rank approaches



$$X = \begin{bmatrix} I_m & X_{12} & \cdots & X_{1n} \\ X_{21} & I_m & \cdots & \vdots \\ \vdots & \vdots & \ddots & X_{n-1,n} \\ X_{n1} & \cdots & X_{n,n-1} & I_m \end{bmatrix}$$

Matrix representations

Undirected maps

Neural networks

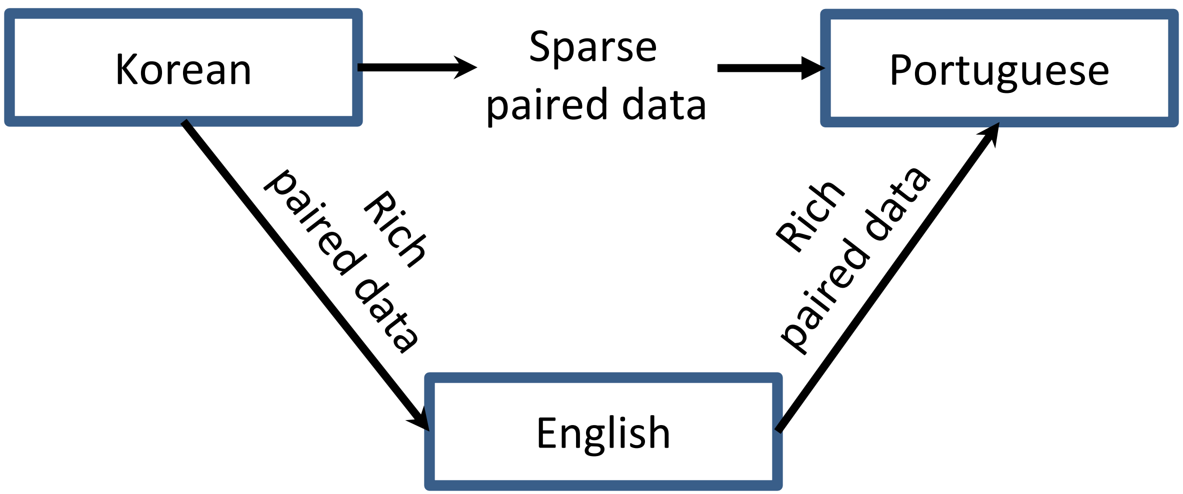
Directed maps

Path-invariant map networks

[Zhang, Liang, Wu, Zhou, H, CVPR’ 2019 oral]

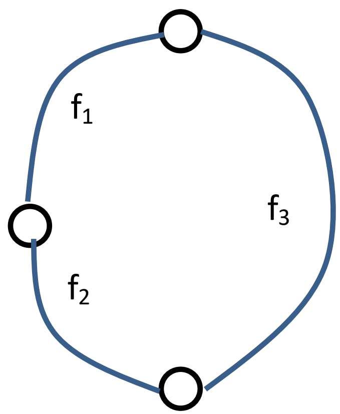
Multi-lingual translation

[Johnson et al. 16]



Abstraction

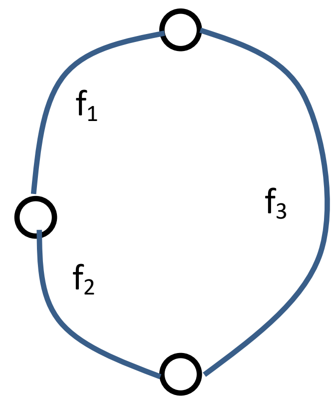
[Zhang et al. CVPR 19]



$f_3 = f_2 f_1$

Path-invariance

$\neq$

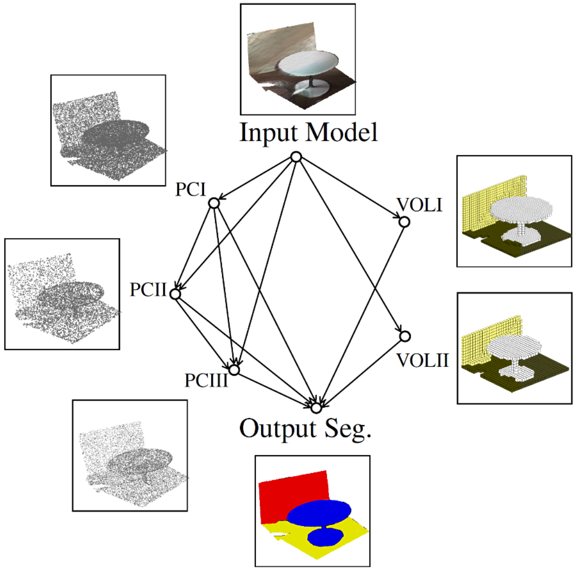


$f_3 f_2 f_1 = Id$

Cycle-consistency

Path-invariance

[Zhang et al. CVPR 19]



Definition 3. *Let $\mathcal{G}_{\text{path}}(u, v)$ collect all paths in $\mathcal{G}$ that connect u to v. We define the set of all possible path pairs of $\mathcal{G}$ as*

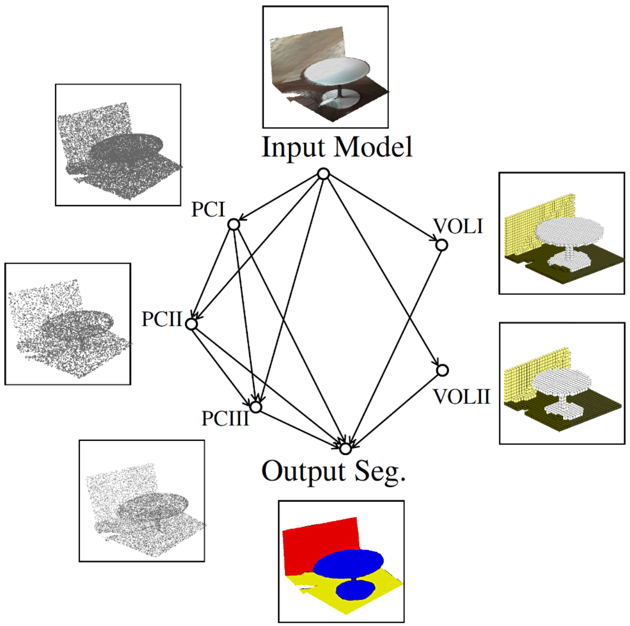
$$\mathcal{G}_{\text{pair}} = \bigcup_{u,v \in \mathcal{V}} \{(p, q) | p, q \in \mathcal{G}_{\text{path}}(u, v)\}.$$

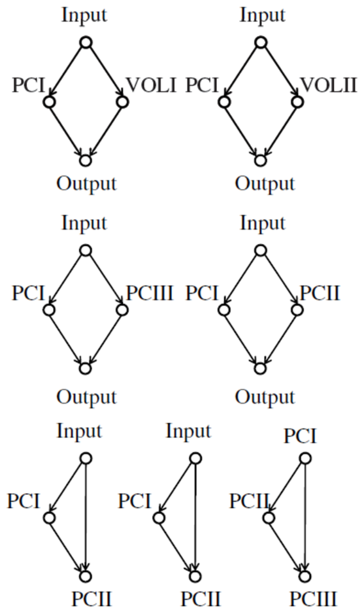
We say $\mathcal{F}$ is path-invariant if

$$f_p = f_q, \qquad \forall (p, q) \in \mathcal{G}_{\text{pair}}.$$

Path-invariance basis

[Zhang et al. CVPR 19]

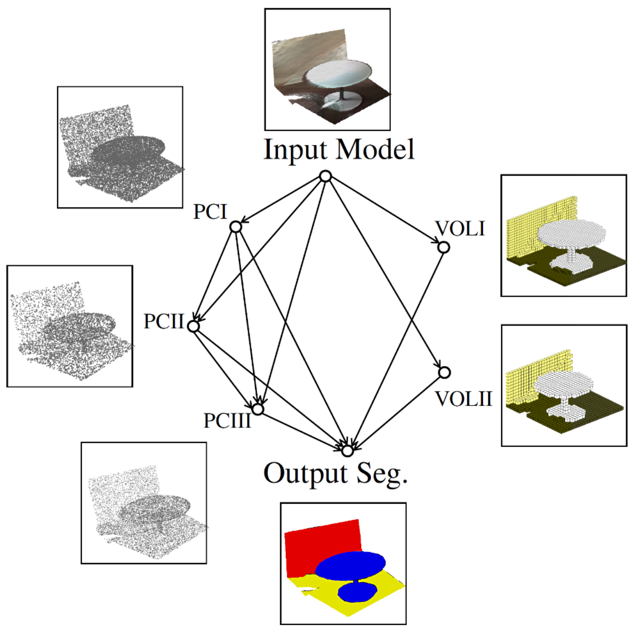


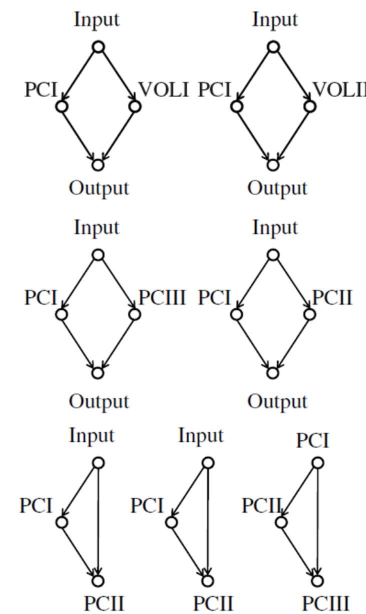


Can induce the path-invariance property of the entire graph

Path-invariance provides a regularization for training neural networks

[Zhang et al. CVPR 19]



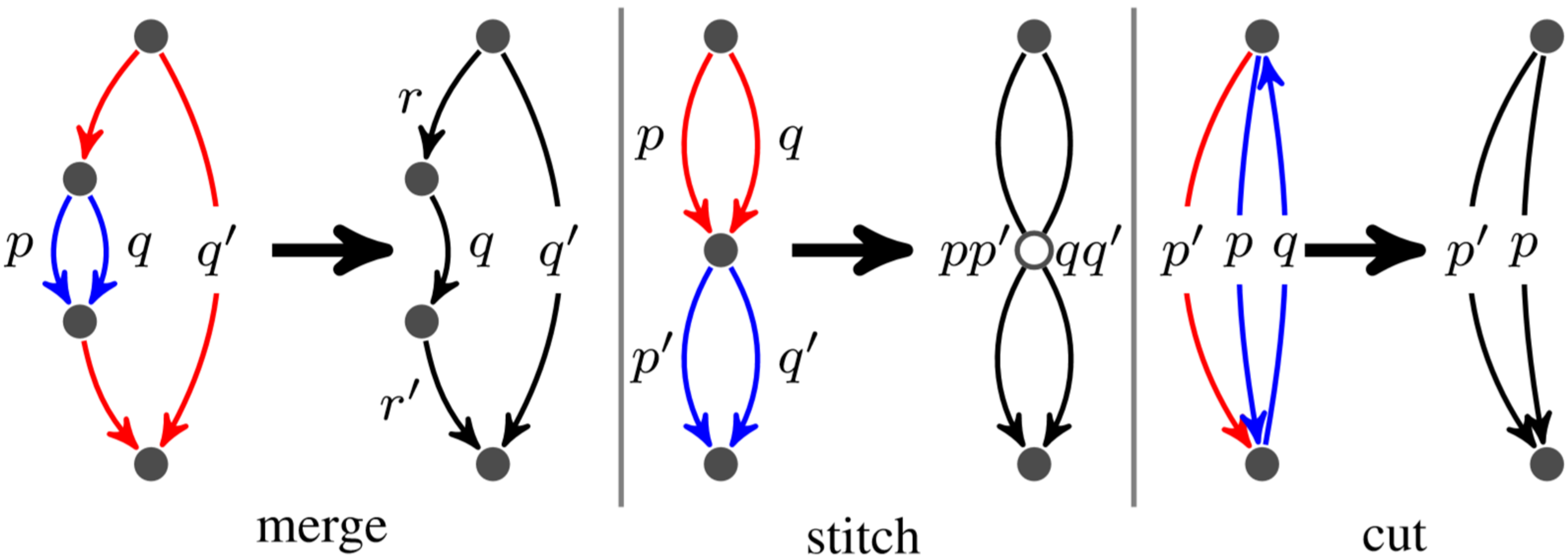


$$\min_{\Theta} \sum_{(i,j)\in\overline{\mathcal{E}}} l_{ij}(f^{\Theta}_{\widehat{v_i v_j}}) + \lambda \sum_{(p,q)\in\mathcal{B}} \underset{v\sim P_{p_t}}{E} d_{\mathcal{D}_{p_t}}(f_p^{\Theta}(v), f_q^{\Theta}(v))$$

Supervised loss

Unsupervised loss

Induction operations



Primitive operations that preserve the path-invariance property

Main result

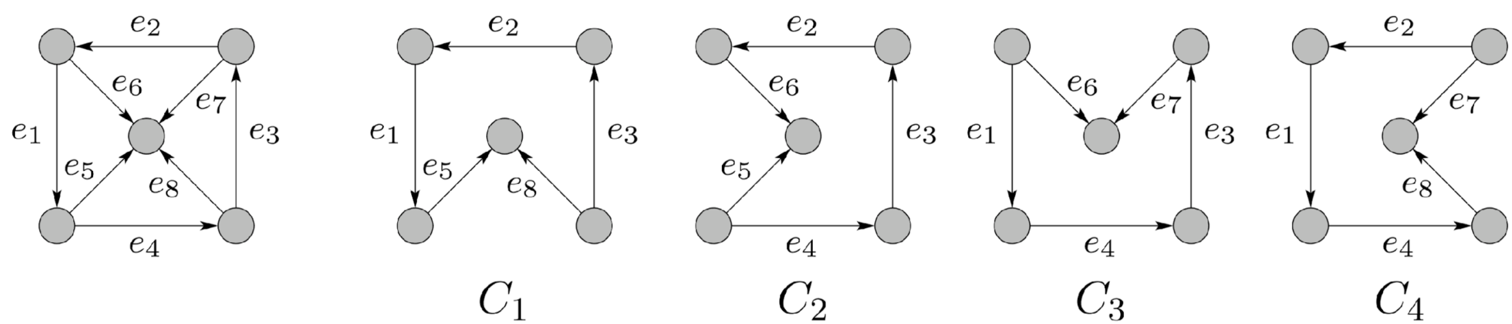
[Zhang et al. CVPR 19]

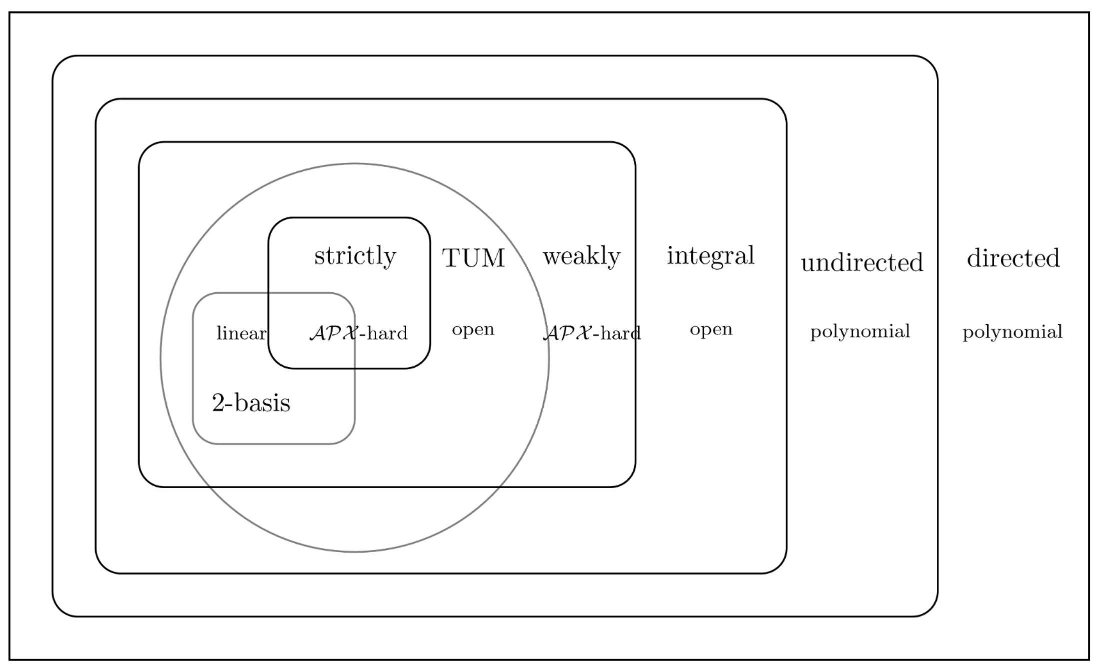
- Theorem: *Given a directed graph with n vertices and m edges, there exists a path-invariance basis with size at most O(nm)*

- Main idea for the proof
  - A directed graph is a directed acyclic graph (DAG) of strongly connected components
  - Use a vertex order to construct a path-invariance basis for DAG

Connection to cycle-basis

[Kavitha et al. 09]





Cycle-consistency basis

[Guibas, H., Liang, 19]

- Defined on undirected graphs
- Operations: merge and stitch
- Minimum size of a cycle-consistency basis
  – #edges - #vertices + 1

- Conjecture I:
  – Computing the minimum path-invariance basis of a given graph is NP-hard
- Conjecture II:
  – Testing a collection of cycles (or path pairs) is a cycle-consistency basis (or path-invariance basis) is also NP-hard

Three advantages over randomly sampling path-pairs

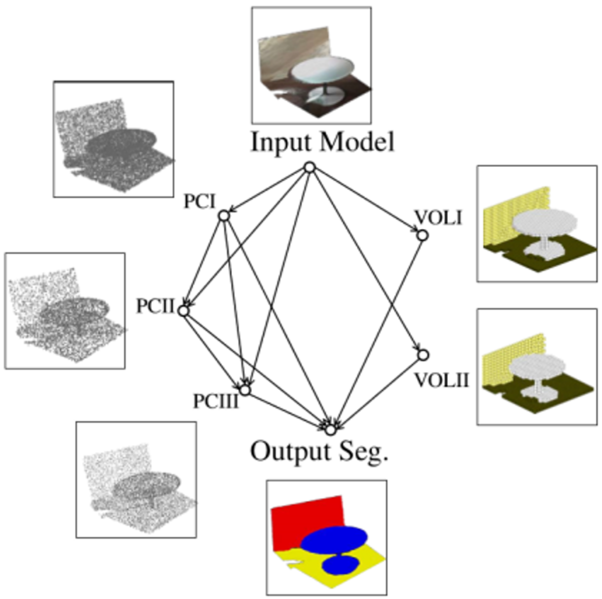
[Zhang et al. CVPR 19]

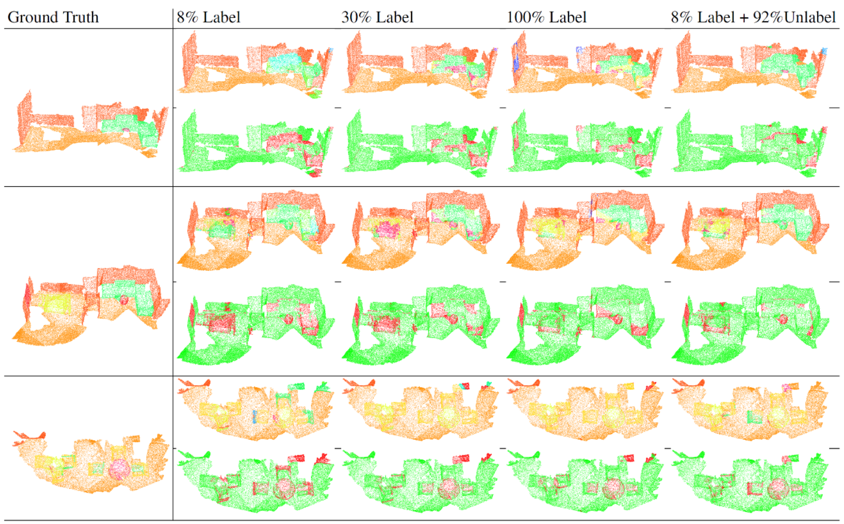
- One may need to sample many (exponentially number of) path pairs to ensure the path-invariance property
  - Many long path pairs

- There is a cost of implementing one path pair

- Convergence of stochastic algorithms

Semantic segmentation on ScanNet

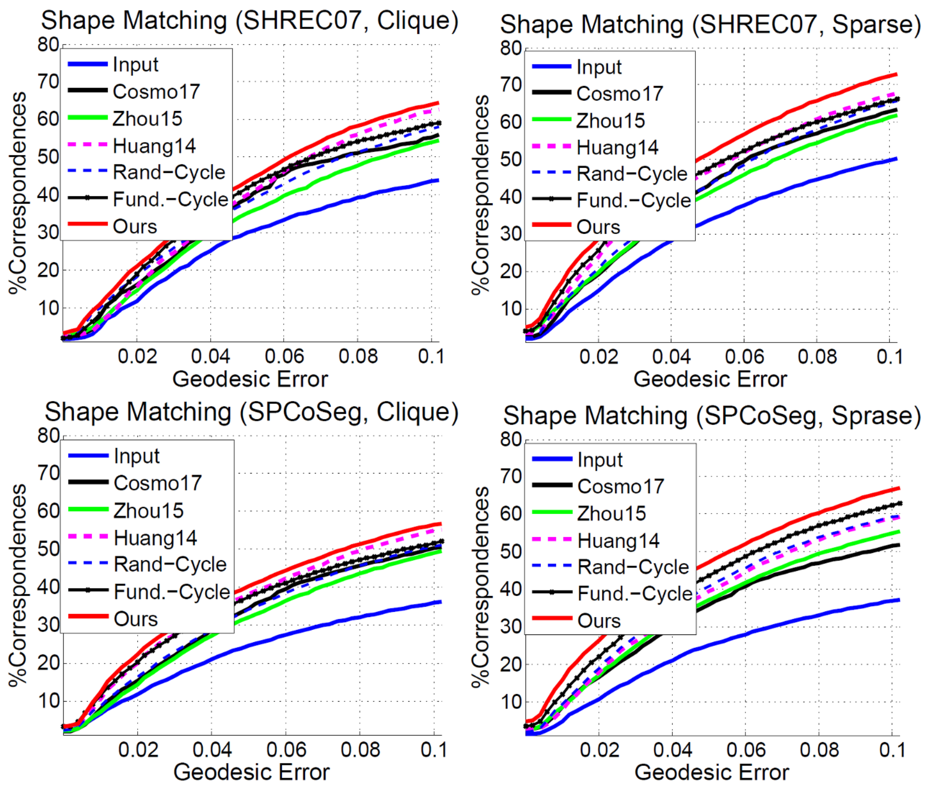




	PCI	PCII	PCIII	VOLI	VOLII
100% Label (Isolated)	84.2	83.3	83.4	81.9	81.5
8% Label (Isolated)	79.2	78.3	78.4	78.7	77.4
8% Label + 92%Unlabel (Joint)	81.7	81.7	81.4	81.1	78.7
30% Label (Isolated)	80.8	81.9	81.2	80.3	79.5

8% labeled + 92% unlabeled $\approx$ 30% labeled

Comparisons on computing object correspondences



Better than low-rank based techniques on sparse graphs

Future directions

- Joint optimization of maps and map consistency
- End-to-end learning for synchronization
- Analysis of self-supervision constraints (path-invariance and cycle-consistency) from the perspective of optimization

Further reading (a partial list)

- Joint learning of neural networks

1. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV 2017.

2. Tinghui Zhou, Philipp Krähenbühl, Mathieu Aubry, Qixing Huang, Alexei A. Efros. Learning Dense Correspondence via 3D-guided Cycle Consistency. CVPR 2016.

3. Amir R. Zamir, Alexander Sax, Teresa Yeo, Oguzhan Kar, Nikhil Cheerla, Rohan Suri, Zhangjie Cao, Jitendra Malik, Leonidas Guibas. Robust Learning Through Cross-Task Consistency. CVPR 2020.