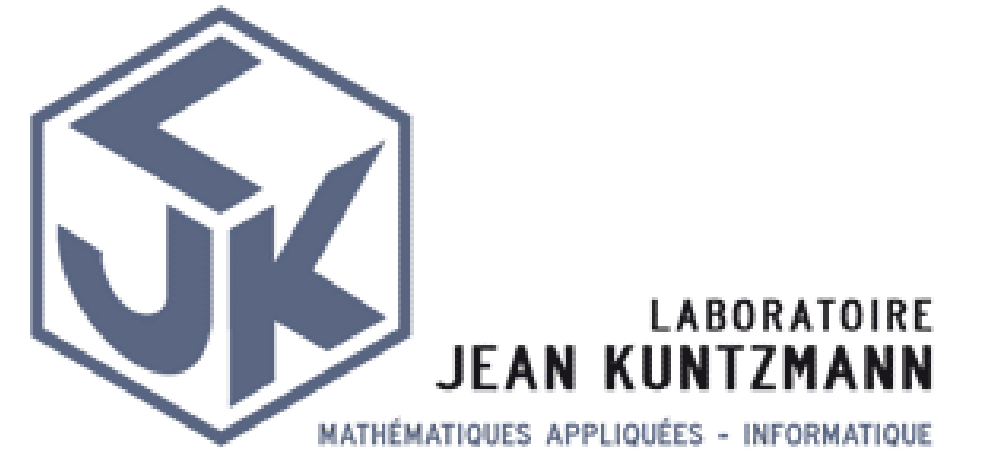


Semi-Supervised Dimensionality Reduction Using Pairwise Equivalence Constraints

Hakan Cevikalp, Jakob Verbeek, Frédéric Jurie, Alexander Kläser
LEAR Team, INRIA Rhône-Alpes, Laboratoire Jean Kuntzmann, France



Summary

- In classification problems of high dimensional data lack of labeled data can be problematic, and explicit labeling is expensive.
- We propose **CLPP**: a dimension reduction method that exploits pairwise class information (easy labeling of pairs: same class or not).
- Our method is based on a neighborhood graph to reflect the intuition that nearby examples tend to have the same label.
- Classification and clustering performance is improved, more so than unsupervised dimension reduction followed by metric learning.

Semi-supervised dimensionality reduction

- As input for the semi-supervised problem we have:
 - data points $X = \{x_1, x_2, \dots, x_n\}$,
 - pairs of the same class $S = \{(i, j) \mid x_i \text{ and } x_j \text{ belong to the same class}\}$,
 - pairs of the different classes $D = \{(i, j) \mid x_i \text{ and } x_j \text{ belong to different classes}\}$.
- High dimensional data is often embedded on a low dimensional manifold, therefore we construct neighborhood graph to capture *local* data structure.
- We modify the Locality Preserving Projection (LPP) method [1] to take into account the class information in the form of pairwise constraints.

Locality Preserving Projections

- LPP is an unsupervised dimension reduction method that is based on a neighborhood graph over a set of points X . A pair of points x_i and x_j are connected by an edge if they are nearest neighbors in the input space; the associated edge weight $W_{ij} \in [0, 1]$ depends on their distance.
- The objective of LPP is to find a linear projections of the inputs $y = a^\top x$ that minimize

$$E(a) = \frac{1}{2} \sum_{i,j} (y_i - y_j)^2 W_{ij} \quad (1)$$

under the constraint that the projected data has unit weighted variance: $\sum_i y_i^2 \left(\sum_j W_{ij} \right) = 1$.

- The optimal projection vectors, respecting mutual orthogonality, are found by calculating the eigenvectors with minimum eigenvalue of the equation

$$X L X^\top a = \lambda X G X^\top a, \quad (2)$$

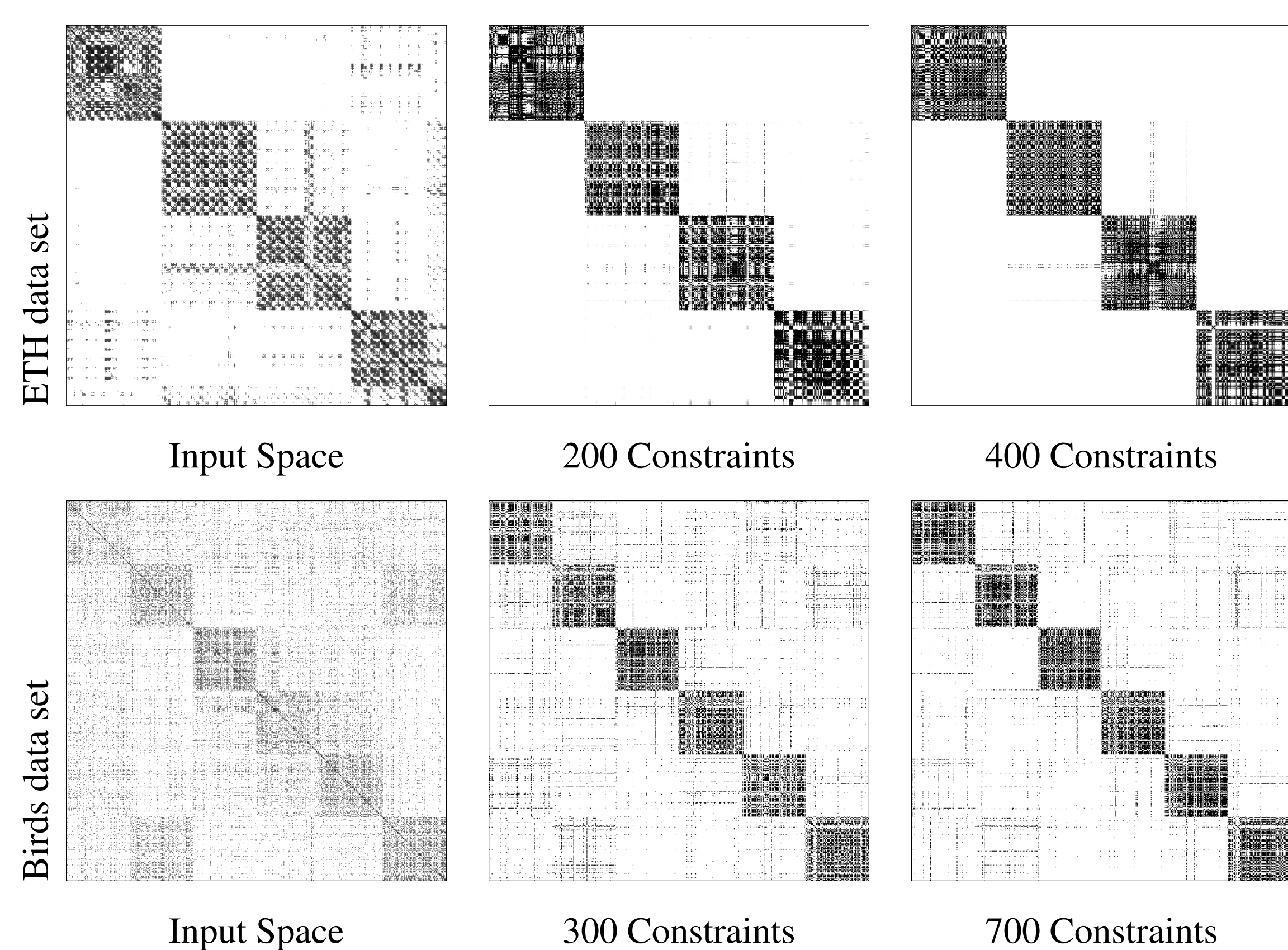
where $L = G - W$ is the graph Laplacian, and G is a diagonal matrix with $G_{ii} = \sum_j W_{ij}$.

CLPP: Integrating Equivalence Constraints

- We assume points in sufficiently small neighborhood tend to have the same label.
- Pairwise constraints (same class or not) modify local connectivity: same class pairs induce long-range edges, and edges between neighbors of different classes get deleted.
- We construct a graph over the data points as follows:
 1. **Construct neighborhood graph.** Compute nearest neighbors using distance metric d (Chi-square in our experiments). Set edge weight for k nearest neighbors to $W_{ij} = \exp(-d(x_i, x_j)/t)$.
 2. **Integrate similarity information.** If $(i, j) \in S$ then set $W_{ij} = 1$. If x_i and x_j have common neighbors, then for each common neighbor x_k set $W_{ik} = W_{jk} = 1$.
 3. **Integrate dissimilarity information.** If $(i, j) \in D$ and $W_{ij} > 0$ then set $W_{ij} = -1$. If x_i and x_j have common neighbor x_k , then set $W_{ik} = 0$ if $W_{ik} < W_{jk}$, or $W_{jk} = 0$ if $W_{ik} > W_{jk}$.
- The objective function of our method is (using $i \sim j$ to denote that x_i and x_j are neighbors):

$$E(a) = \frac{1}{2} \left(\sum_{(i,j) \in S} (y_i - y_j)^2 - \sum_{(i,j) \in D} (y_i - y_j)^2 + \sum_{\text{other pairs}} (y_i - y_j)^2 W_{ij} \right). \quad (3)$$

- As for LPP projections $y = a^\top x$ are found by solving for the minimum eigenvalue solutions of a generalized eigenvector equation.
- Non-linear data representations can be obtained by solving for the y_i directly, without requiring $y_i = a^\top x_i$, in this case the eigenvalue problem scales with the number of data points rather than with their dimensionality.



Visualization of affinity matrices obtained from the ETH dataset (top) and Birds dataset (bottom).



Datasets ETH-80 (top row) and Birds (the second and third rows, 2 illustrative images per category).

Experimental Evaluation

- Datasets:
 - Categories *Apple, Car, Cow, Cup* from ETH-80 [2], 10–14 images per category, flat background.
 - Birds dataset [3], 6 categories, 100 images per category, highly cluttered background.
- Bag-of-features representation, patches sampled on a dense regular grid over the image
 - Each patch represented using 128 dimensional SIFT descriptor [4],
 - SIFT descriptors quantized using k-means into 500 (ETH) or 2000 (Birds) clusters,
 - image represented by 500 (2000) dimensional cluster frequency histogram.

Experimental setup

Comparison of our CLPP with, separate dimension reduction (PCA or LPP) + metric learning [5].

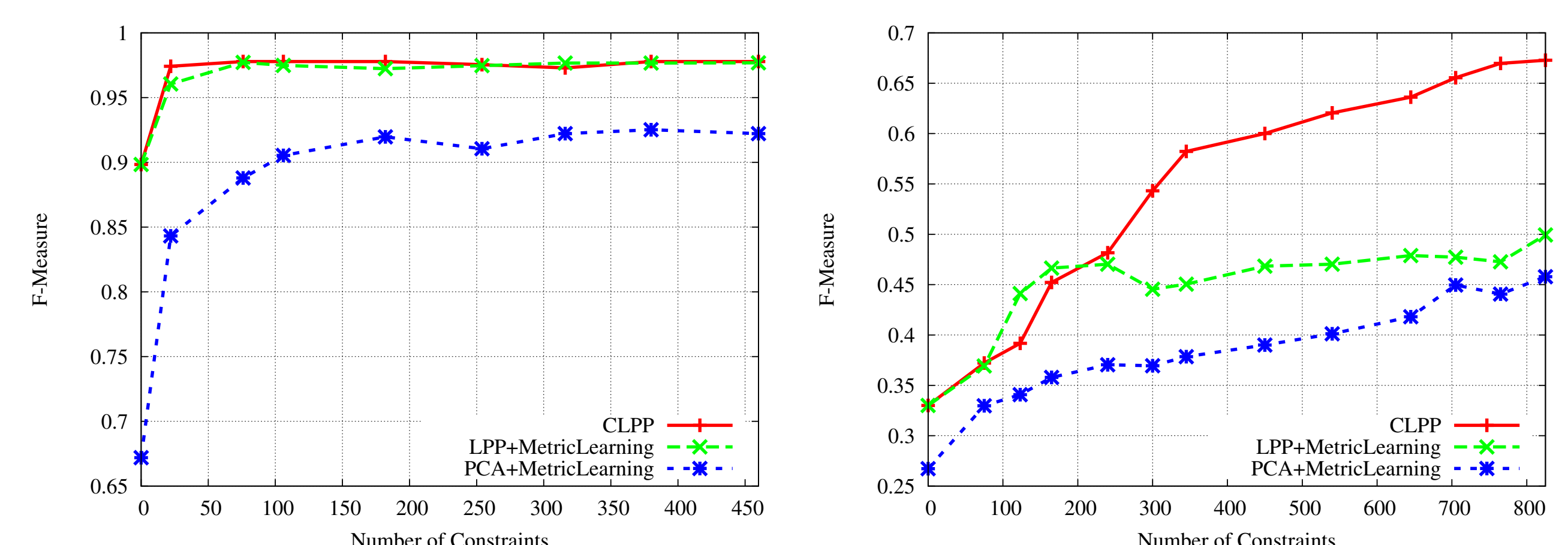
- Pairwise constraints taken from uniformly randomly selected pairs.
- Perform semi-supervised dimensionality reduction with varying number of pairwise constraints.
- Perform k-means clustering on projected data.
- Calculate pairwise F-measure from *pairwise* precision and recall (cross-validated).

Let A denote the set of image pairs assigned to the same k -means cluster, and let B denote the set of image pairs that contain the same object category. With $|A|$ denoting the cardinality of A :

$$\text{Precision} = \frac{|A \cap B|}{|A|}, \quad \text{Recall} = \frac{|A \cap B|}{|B|}, \quad \text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Experimental Results

- Graph-based methods out-perform PCA + metric learning.
- On Birds dataset, CLPP significantly outperforms LPP + metric learning, most discriminative information lost during unsupervised dimension reduction, due to class-irrelevant clutter.
- Affinity matrices W clearly show increased separability.
- Applications to user-guided image segmentation in paper.



F-measure as a function of number of constraints for held-out test data; ETH (left) and Birds (right).

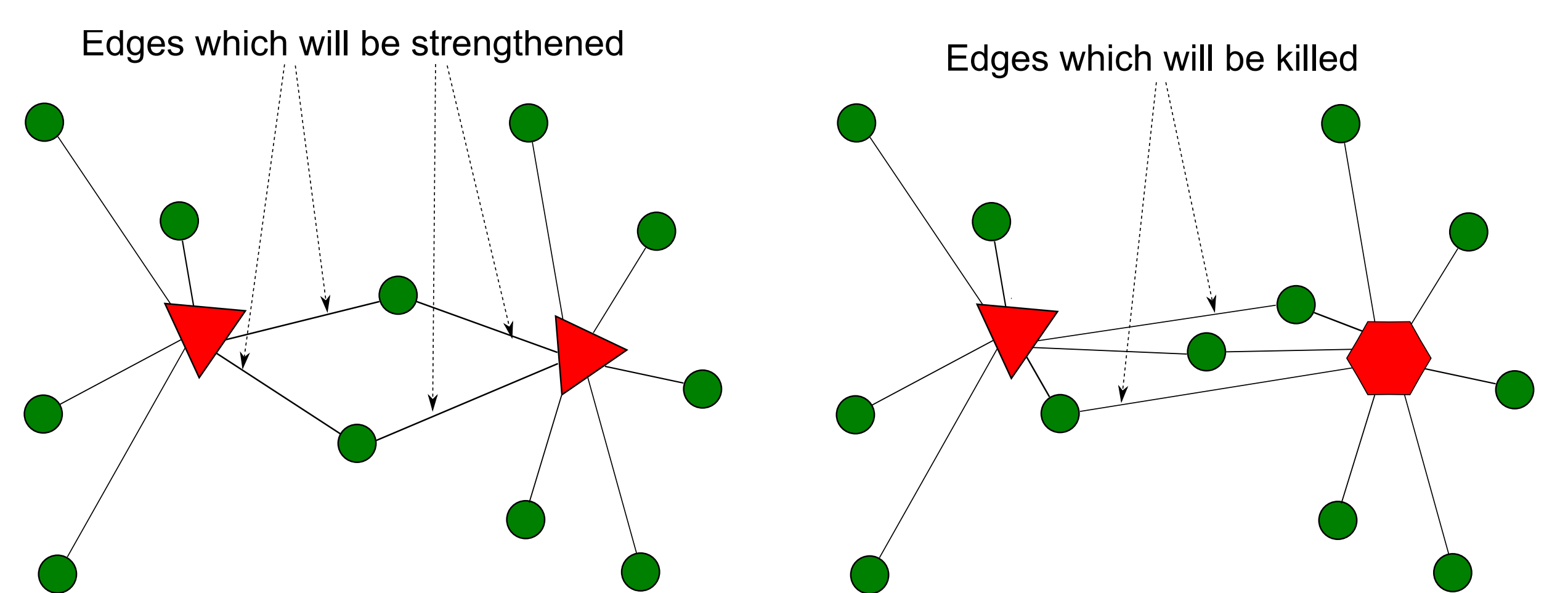


Illustration of how pairwise equivalence constraints modify the the local connectivity in the neighborhood graph in case of equivalence (left) and inequivalence (right).

References

- [1] X. He and P. Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems*, volume 16, pages 153–160. MIT Press, Cambridge, MA, USA, 2004.
- [2] B. Leibe and B. Schiele. Interleaved object categorization and segmentation. In *British Machine Vision Conference (BMVC)*, 2003.
- [3] S. Lazebnik, C. Schmid, and J. Ponce. A maximum entropy framework for part-based texture and object recognition. In *International Conference on Computer Vision (ICCV)*, 2005.
- [4] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [5] I. Tsang and J. Kwok. Distance metric learning with kernels. In *International Conference on Artificial Neural Networks*, 2003.