Training a watershed layer

L. Najman

Joint work with Rahul Chakwata, Aditya Challa,
Sravan Danda and B.S. Daya Sagar

30/09/2019 – MorphoNet Meeting

Previous work:
The Maximum Margin Partition

$$\arg\max_{M} \hat{\rho}(X_0, X_1, M) = \arg\max_{M} \left\{ \inf \left\{ \rho(X_0, \overline{M}), \rho(X_1, M) \right\} \right\}$$

$$\rho(X, Y) = \inf_{x \in X, y \in Y} \rho(x, y).$$

The distance can be Euclidean or ultrametric (pass value)

The watershed provides a solution to this problem

A. Challa, S. Danda, B. S. D. Sagar and L. Najman,
"Watersheds for Semi-Supervised Classification,"
IEEE Signal Processing Letters, vol. 26, no. 5, pp. 720-724, May 2019.

Main theme of this talk

Replacing the softmax classifier with a watershed

The Maximum Margin Partition

$$\arg\max_{M} \hat{\rho}(X_0, X_1, M) = \arg\max_{M} \left\{\inf \left\{\rho(X_0, \overline{M}), \rho(X_1, M)\right\}\right\}$$

$$\rho(X, Y) = \inf_{x \in X, y \in Y} \rho(x, y).$$

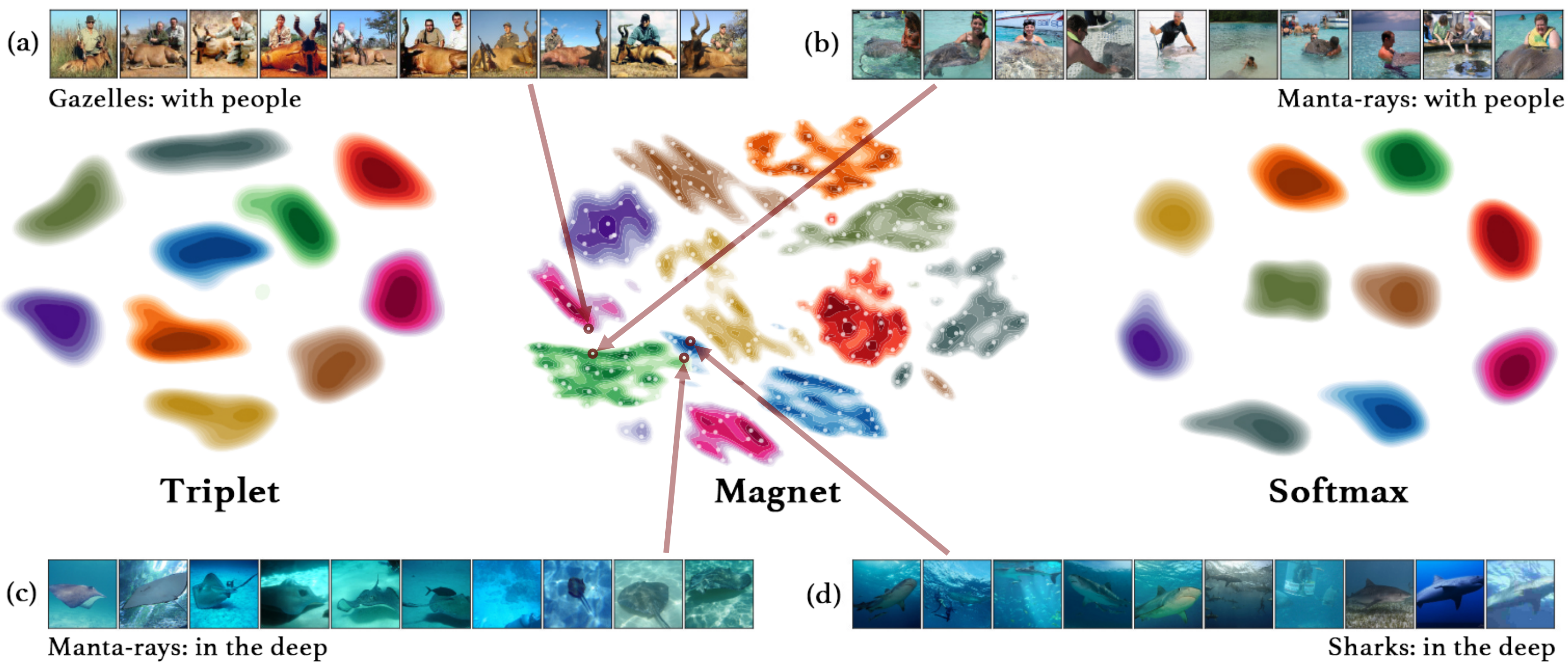
The distance can be Euclidean or ultrametric (pass value)

The watershed provides a solution to this problem

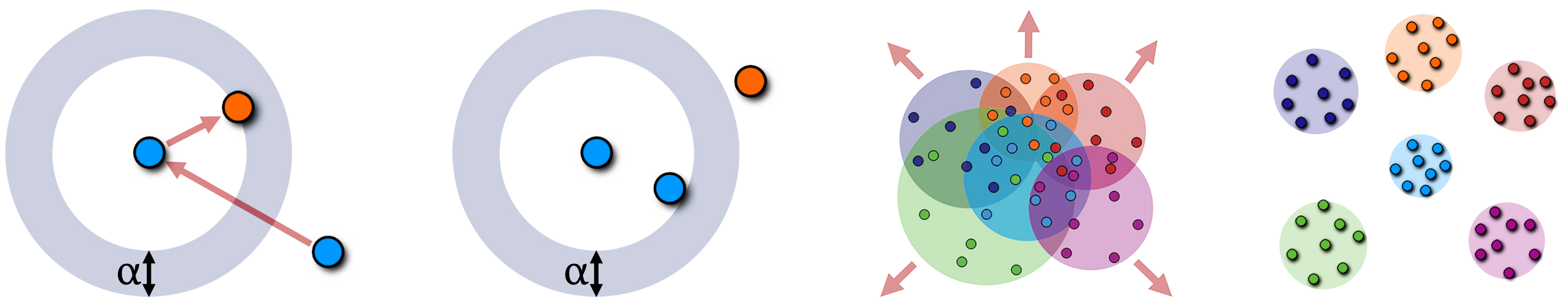
A. Challa, S. Danda, B. S. D. Sagar and L. Najman,
"Watersheds for Semi-Supervised Classification,"
IEEE Signal Processing Letters, vol. 26, no. 5, pp. 720-724, May 2019.

Magnet loss

METRIC LEARNING WITH ADAPTIVE DENSITY DISCRIMINATION - ICLR 2016



Magnet loss: intuition



(a) Triplet: before. (b) Triplet: after. (c) Magnet: before. (d) Magnet: after.

$$\mathscr{L}\left(\boldsymbol{\Theta}\right) = \frac{1}{N}\sum_{n=1}^{N}\left\{-\log\frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{r}_n-\boldsymbol{\mu}(\mathbf{r}_n)\|_2^2-\alpha}}{\sum_{c\neq C(\mathbf{r}_n)}\sum_{k=1}^{K}e^{-\frac{1}{2\sigma^2}\left\|\mathbf{r}_n-\boldsymbol{\mu}_k^c\right\|_2^2}}\right\}_+$$

Modifying magnet loss algorithm

- First possibility
  - Replace k-Means with iterated watershed
    - Sampriti Soor, Aditya Challa, Sravan Danda, B Daya Sagar, Laurent Najman. *Iterated Watersheds, A Connected Variation of K-Means for Clustering GIS Data*. IEEE Transactions on Emerging Topics in Computing, Institute of Electrical and Electronics Engineers, In press, 〈10.1109/TETC.2019.2910147〉
- Second possibility
  - Replace k-means with filtered watershed
    - No longer needs the “k” parameter

Replacing KNN with iterated watershed in magnet loss

SSL 6 into 2 class classification				
Without L2 norm		With L2 norm		
Train	Val	Train	Val	
99.91	98.33	99	97.66	
99.75	99.33	99.75	99.33	
99.58	97	98.66	97.33	
99.08	97	99.66	98.66	
99	98	99.83	97.33	
99.91	98.66	100	99.33	
99.52875	97.91375	99.48333333	98.27333333	Average

- Watershed performs better that the original magnet loss
- We can also show that we can take advantage of a variable number of cluster implementation (filtered watershed)

Watershed layer with a MST

- Using a MST / ultrametric is possible
- Unstable in practice
- Needs a regularization term
  - See B. Perret and G. Chierchia 2019 NeurIPS paper

Watershed layer by probabilistic inference

$$P(x \in \text{class } l) = \frac{\exp(-\rho(x, X_l))}{\sum_i \exp(-\rho(x, X_i))}$$

Fully supervised classification

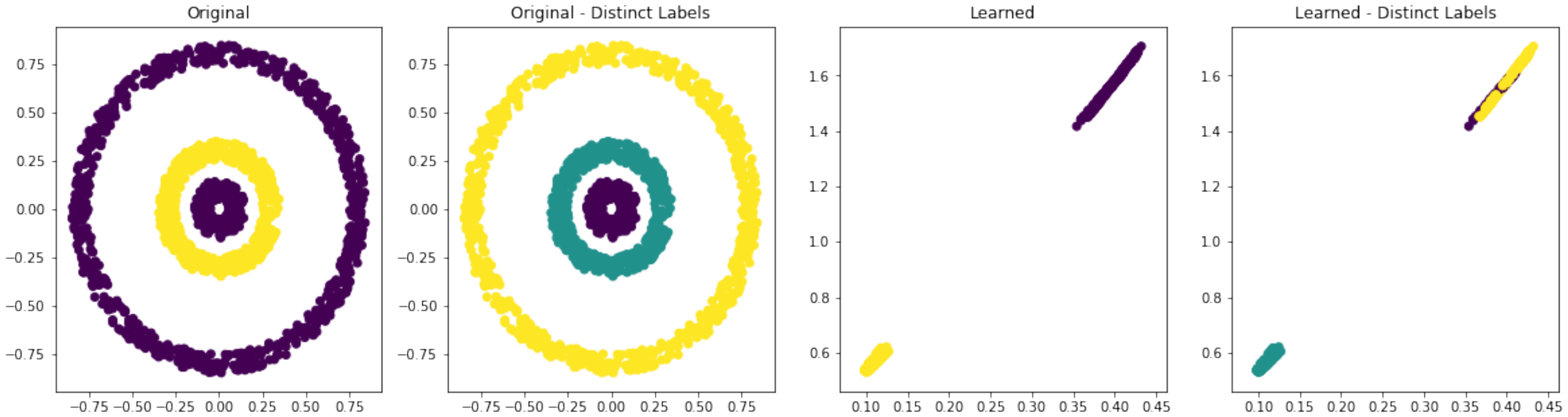
$$\text{Cost} = -\sum_{i,j} y_{i,j} log(P_{i,j})$$

Semi-supervised classification

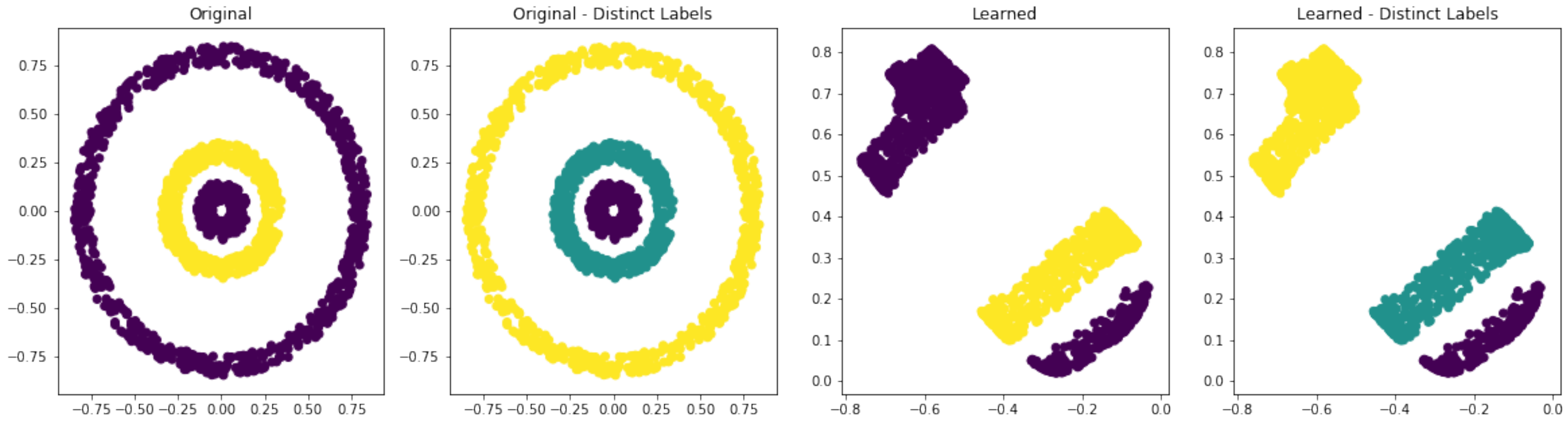
$$\text{Cost} = -\sum_{i \in \text{labelled}, j} y_{i,j} log(P_{i,j}) + \sum_{i \in \text{unlabelled}, j} P_{i,j} log(P_{i,j})$$

Siamese network example

Contrastive loss (no watershed here)

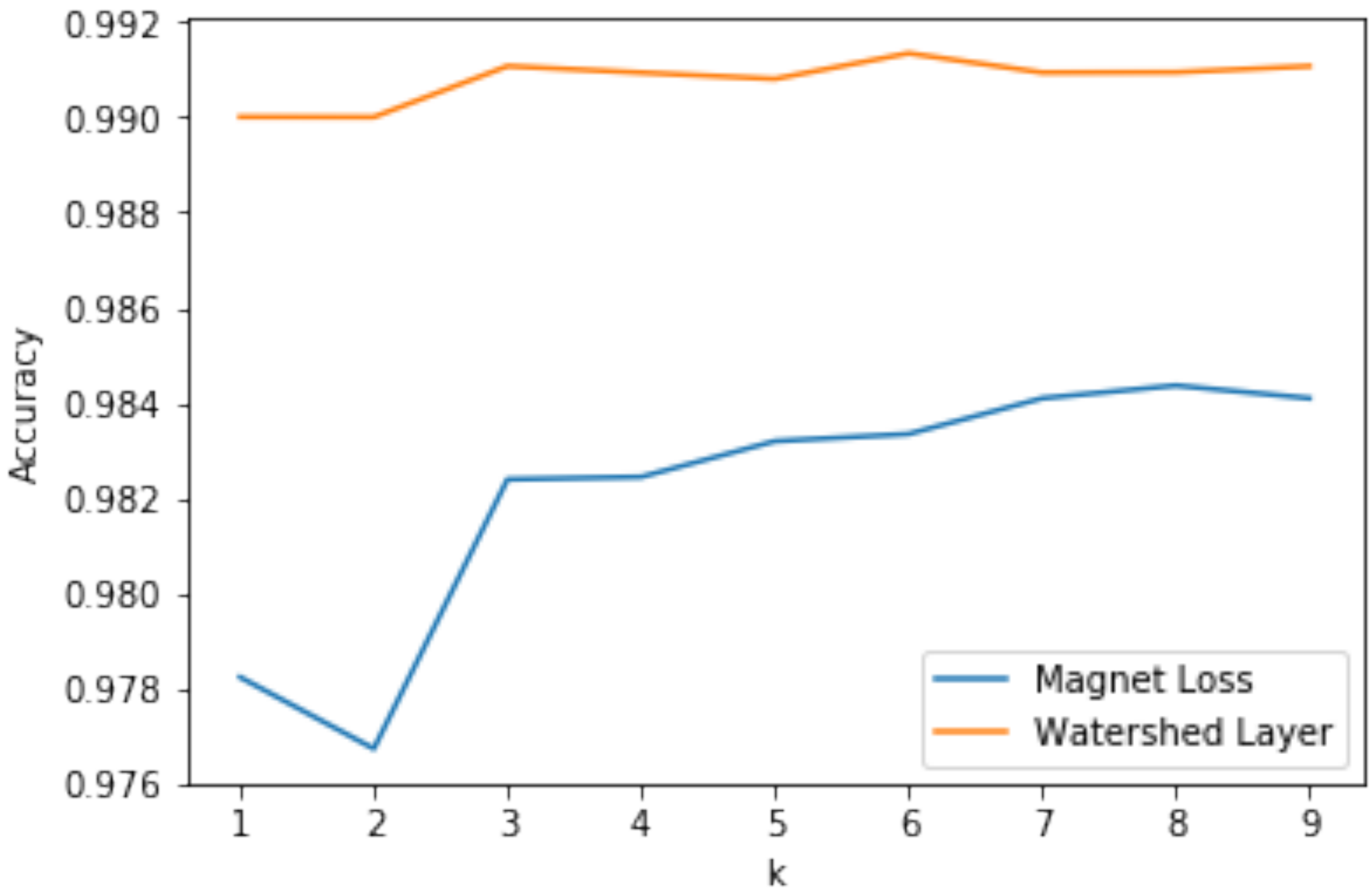


Watershed example

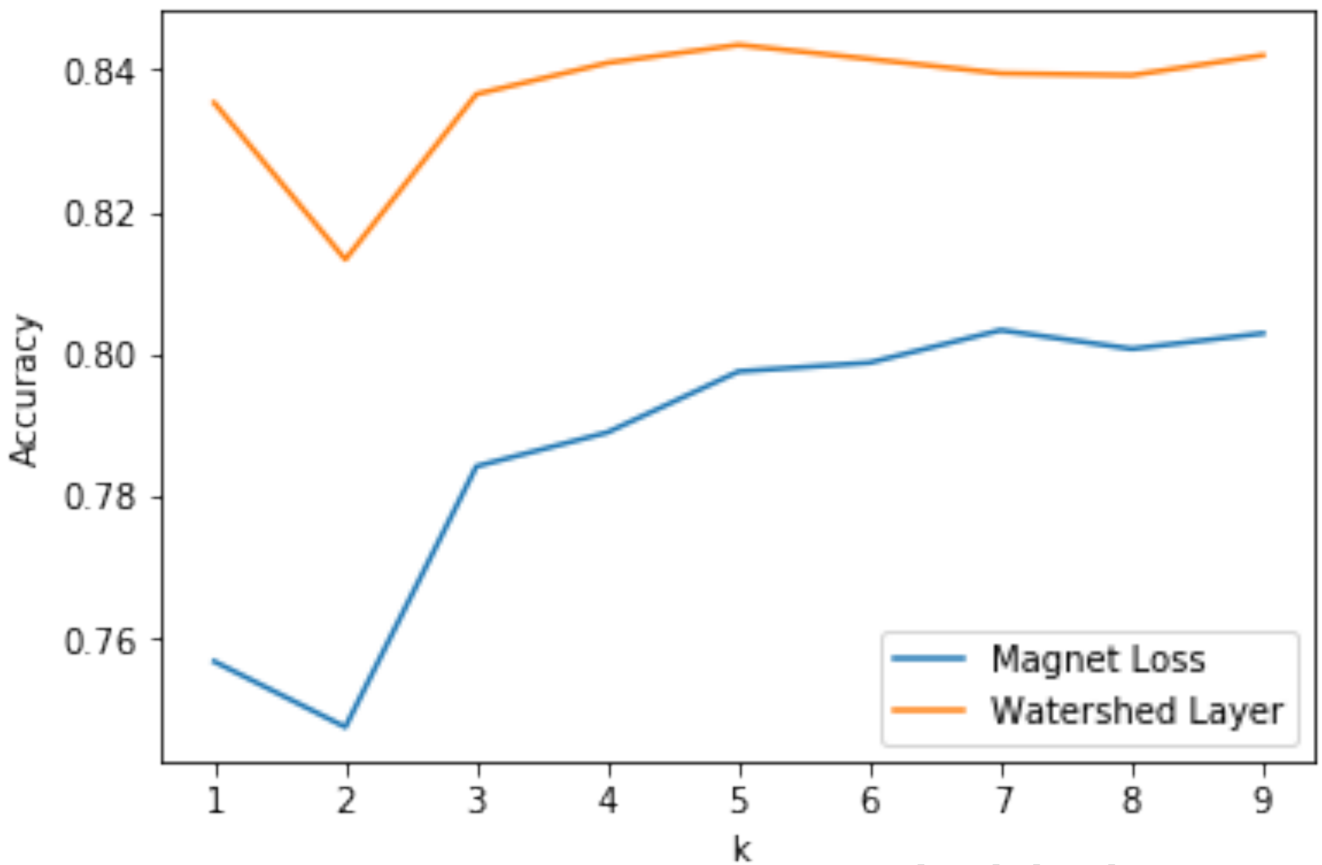


Watershed preserves the structure of the data while contrastive loss removes it

MNIST Dataset: KNN accuracy



Preserving the hierarchical structure



MNIST is splited in two classes [0-4], [5-9]
KNN accuracy for the original 10 classes

Perspectives

- Batch mode
- Compactness properties of the clusters
- Hierarchical properties
  - Practically
  - Theoretically
- Regularization term
  - Connected filters

- Using watershed for regression
  - First results quite promising