Active Learning Methods on Graphs for Image, Video and Multispectral Datasets

NGA NURI grant # HM04762110003

PI: Dr. Andrea Bertozzi

Collaborator: Jeff Calder Univ. Minnesota

PhD students: Jason Brown, Kevin Miller

Undergraduates: Xoaquin Baca (Harvey Mudd) , Jack Mauro (LMU), Jason Setiadi (UMN) and Zhan Shi (UCLA)

Current project period: 5/28/2021 - 5/27/2022

Introduction

- **Supervised Learning** : Large amount of training (labeled) data
  - e.g. Convolutional neural network, support vector machine
- **Unsupervised Learning**: No training (labeled) data
  - e.g. Spectral clustering, k-means clustering, t-SNE embedding
- **Semi-supervised learning (SSL)**: Small amount of labeled data, but have unlabeled data as well
  - Graph-based SSL
    - Use similarity graph structure to aid in inferring classification of unlabeled data
    - Especially useful in low-label rate regime

Graph Construction

- Given inputs $Z = \{z_1, z_2, \ldots, z_n\}$, define a symmetric kernel function $k(z_i, z_j)$ between each pair of points
  - Larger $k(z_i, z_j)$ means higher similarity between the points
  - Define weight matrix, $W_{ij} = k(z_i, z_j)$
- For example, $$w_{ij} = exp(-d(z_i, z_j)^2/\tau)$$
  - **Metric (*d*)**: Euclidean, angular
  - **Scaling (*τ*)**:
    - Constant = Gaussian kernel
    - Distance with k^{th} nearest neighbor = Zelnik-Manor and Perona (ZMP)

Graph Construction: Degree and Graph Laplacian

- Degree of node z_i: $d_i = \Sigma_j w_{ij}$
  - Diagonal degree matrix: $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$

- Graph Laplacian: $L = D - W$
  - Common normalization: $L_s = D^{-1/2} L D^{-1/2}$ ("normalized graph Laplacian") - Hyperspectral Imagery
  - Positive semi-definite operator whose eigenvectors are useful for encoding clustering structure
    - E.g. Spectral Clustering

Accelerating the Diagonalization

Avoid costly computation of eigenvectors and eigenvalues on large, dense matrices.

- *k* nearest neighbor (kNN) graph
  - Only keep the weights of each node related to its *k* nearest neighbors ($k \ll n$)
  - Results in a **sparse** matrix
- Nyström Extension
  - **Low-rank approximation** of the dense weight matrix:
    - Constructs and stores only *k* columns ($k \ll n$)
  - Efficient computation of the eigenpairs of the graph Laplacian

Active Learning

Comparison of Goals:

- Semi-Supervised Learning : Accurate classification given current labeled data
- **Active Learning: "Optimally" select points to hand-label (classify) in order to improve underlying SSL classifier**

Active Learning Methods are usually more **explorative** or more **exploitative**

- **Exploration**: "Explore" the extent of the clustering structure of the underlying dataset
- **Exploitation:** "Exploit" the current level of knowledge about the classification of points in dataset

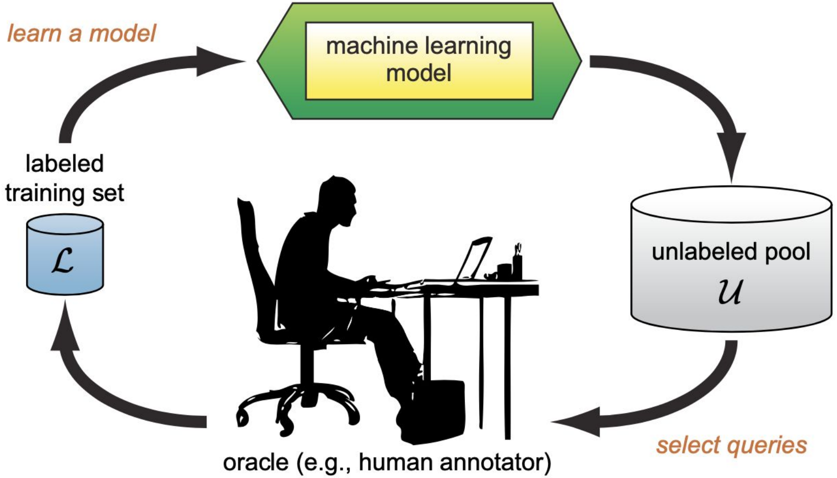
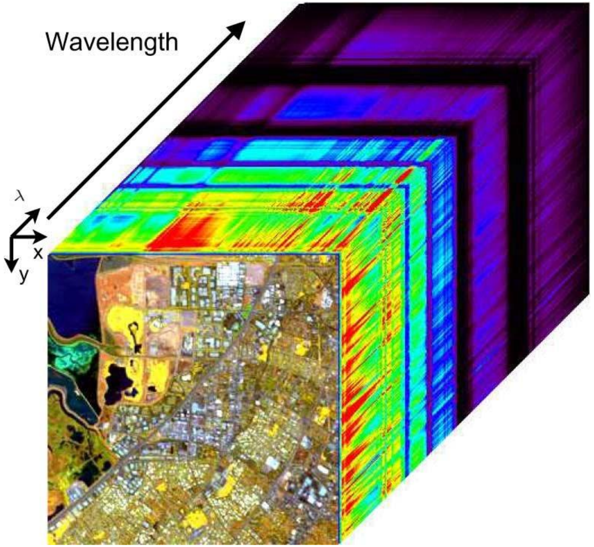


Image credit: Settles, Active Learning, 2013.

Application: Hyperspectral Imagery (HSI)



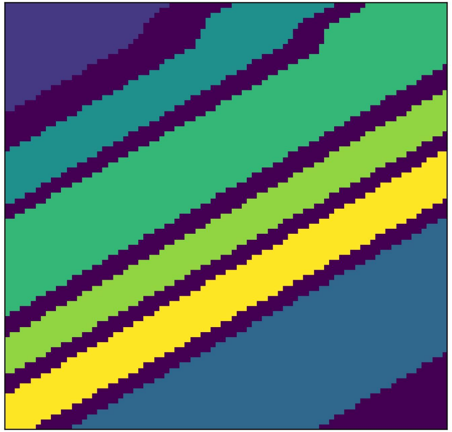
Hyperspectral images contain rich information about objects in the image, per the many wavelengths that are sampled when image is taken.

- Seek to classify the pixels into classes (e.g. water, dirt, grass, metal, etc)
- Noisy measurements, corrupted by weather and atmospheric effects

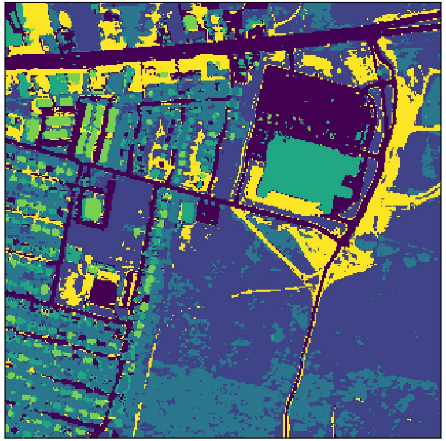
Apply active learning to incorporate human-in-the-loop to improve the accuracy of graph-based semi-supervised classification of pixels.

Image credit: Christophe, Emmanuel & Mailhes, Corinne & Duhamel, P. (2009)

HSI Datasets



Salinas-A: 6 classes of plant types



Urban: asphalt, grass, tree, roof, metal, and dirt

Graph Construction Details:

- 15-nearest neighbors graph
- Cosine similarity kernel, $k(x_i, x_j) = \langle x_i, x_j \rangle / ||x_i||\,||x_j||$

Active Learning with Graph-Based SSL

Given graph Laplacian L, define the energy $$\mathcal{J}(U;Y) := \frac{1}{2}\langle U, LU\rangle_F + \sum_{j\in\mathcal{L}} \ell(\mathbf{u}^j, \mathbf{y}^j)$$

- N x n_c matrix U (# pixels by # classes)
    - **u^j** = j^th row of matrix U
- **y^j** = "one-hot" encoding of classification of pixel j
- Loss functions:
    - Squared-Error $\ell(\mathbf{s},\mathbf{t}) = \frac{1}{2\gamma^2}\|\mathbf{s}-\mathbf{t}\|_2^2$ (will refer to as Multiclass Gaussian Regression, MGR)
    - Cross-Entropy (CE) $\ell(\mathbf{s},\mathbf{t}) = \sum_{c=1}^{n_c} s_c \ln t_c$

Look-Ahead Model

- "Hypothetical" model, *if were to label pixel k according to class* ***y^k.***

$$\mathcal{J}^{+k,\mathbf{y}^k}(U;Y,\mathbf{y}^k) := \frac{1}{2}\langle U, LU\rangle_F + \sum_{j\in\mathcal{L}} \ell(\mathbf{u}^j, \mathbf{y}^j) + \ell(\mathbf{u}^k, \mathbf{y}^k)$$

Active Learning with Graph-Based SSL

Given graph Laplacian L, define the energy $$\mathcal{J}(U;Y) := \frac{1}{2}\langle U, LU\rangle_F + \sum_{j\in\mathcal{L}} \ell(\mathbf{u}^j, \mathbf{y}^j)$$

- N x n_c matrix U (# pixels by # classes)
  - **u^j** = j^th row of matrix U
- **y^j** = "one-hot" encoding of classification of pixel j
- Loss functions:
  - Squared-Error $\ell(\mathbf{s}, \mathbf{t}) = \frac{1}{2\gamma^2}\|\mathbf{s} - \mathbf{t}\|_2^2$ (will refer to as Multiclass Gaussian Regression, MGR)
  - Cross-Entropy (CE) $\ell(\mathbf{s}, \mathbf{t}) = \sum_{c=1}^{n_c} s_c \ln t_c$

Look-Ahead Model

- "Hypothetical" model, *if were to label pixel k according to class* ***y^k.***

$$\mathcal{J}^{+k,\mathbf{y}^k}(U;Y,\mathbf{y}^k) := \frac{1}{2}\langle U, LU\rangle_F + \sum_{j\in\mathcal{L}} \ell(\mathbf{u}^j, \mathbf{y}^j) + \ell(\mathbf{u}^k, \mathbf{y}^k)$$

How does labeling point k help our overall classification model?

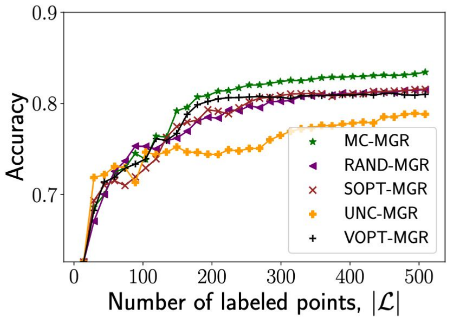
Active Learning Acquisition Functions

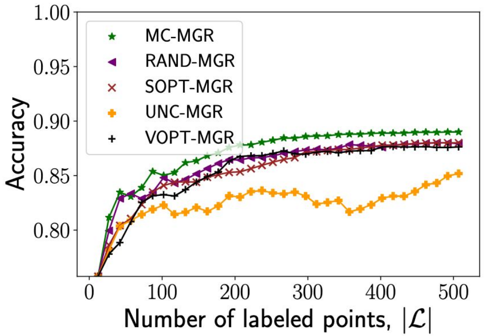
Acquisition function: Active learning criterion "function" that quantifies the utility of labeling an unlabeled point k

- **Random Sampling** : Select points uniformly at random
- **Uncertainty Sampling** : Select point that current classifier is "most uncertain" about
- **Variance Minimization** : Select point that will decrease the variance of the current classifier the most
    - **VOpt**: Minimize Trace of covariance matrix of Gaussian distribution of associated graph-based classifier
    - **SOpt**: "Sigma Optimality", variant of VOpt
- **Model Change** : Select point that will change the current classifier "the most"
    - Use look-ahead model with hypothetical "pseudo-label" to calculate

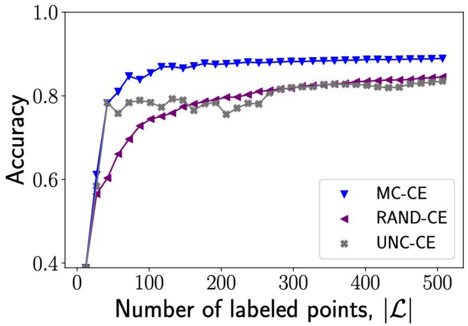
Results in HSI Application

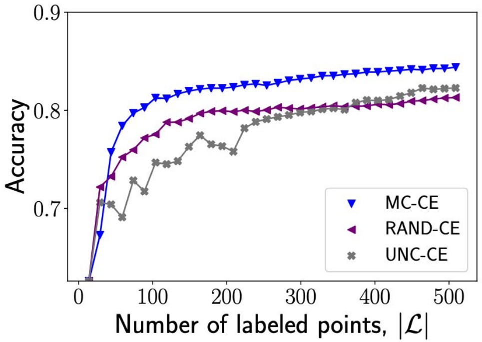
Multiclass Gaussian Regression





Cross-Entropy





Salinas-A (~7K total points)

Urban (~90K total points)

Synthetic Aperture Radar (SAR) Images

- Finer resolution images than standard radar
  - Mimic large antenna properties with multiple measurements from smaller antenna
  - Present in moving objects such as aircraft/spacecraft as well as drones
- Useful for Automatic Target Recognition (ATR) problems

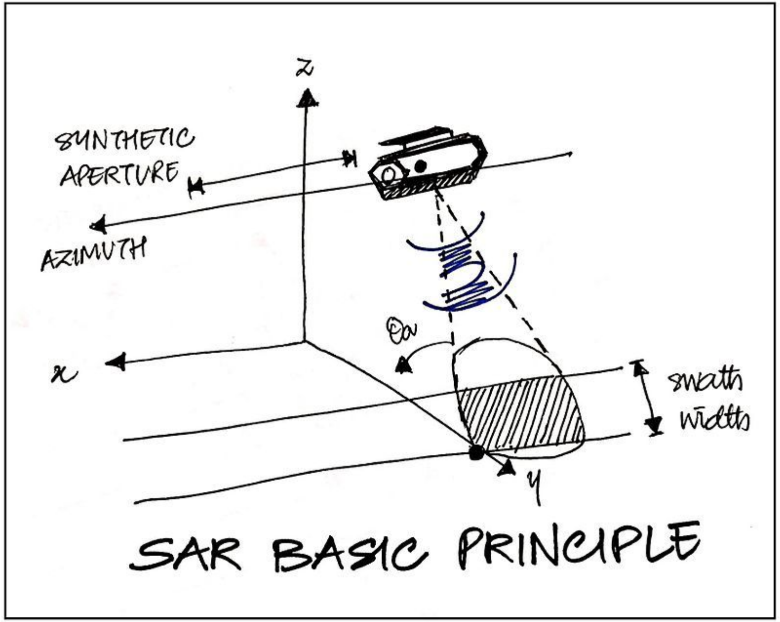


Image credit: https://commons.wikimedia.org/wiki/File:Synthetic_Aperture_Radar.jpg

MSTAR Dataset

- Collection of SAR images from 1995-1997.
  - 10 distinct types of ground vehicles such as tanks and trucks.
  - 6,874 images of size 88 x 88
- Magnitude and phase data

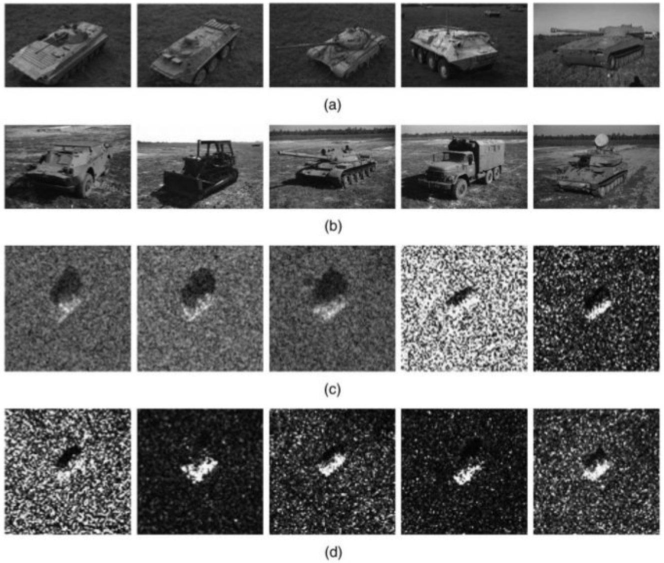


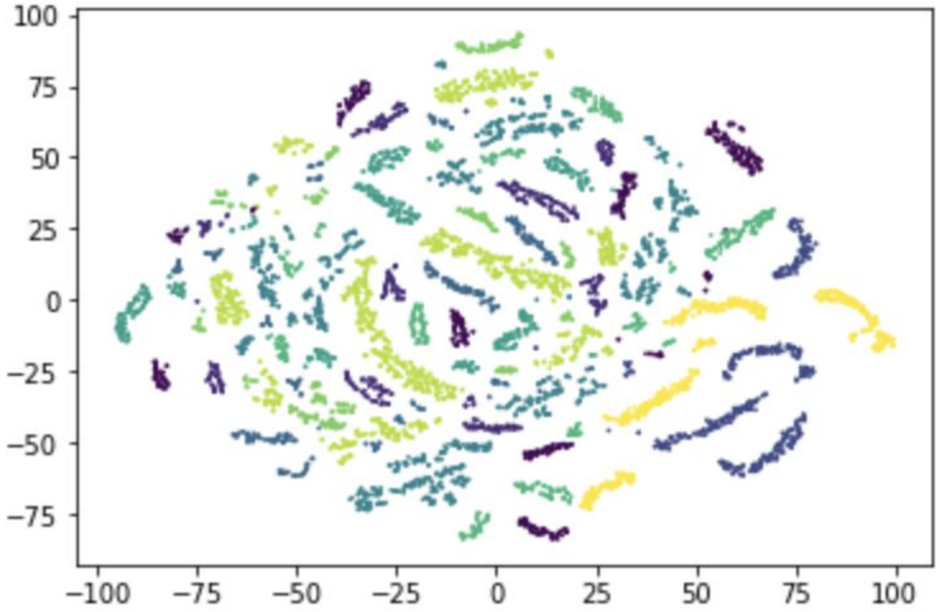
Fig. 2 MSTAR database. (a) and (b) Visible light images for BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7, T62, ZIL131, and ZSU23/4. (c) and (d) Corresponding SAR images for 10 targets measured at azimuth angle of 45 deg.

Image credit: Perumal, Vasuki (2013).

t-SNE Embedding of MSTAR for visualization

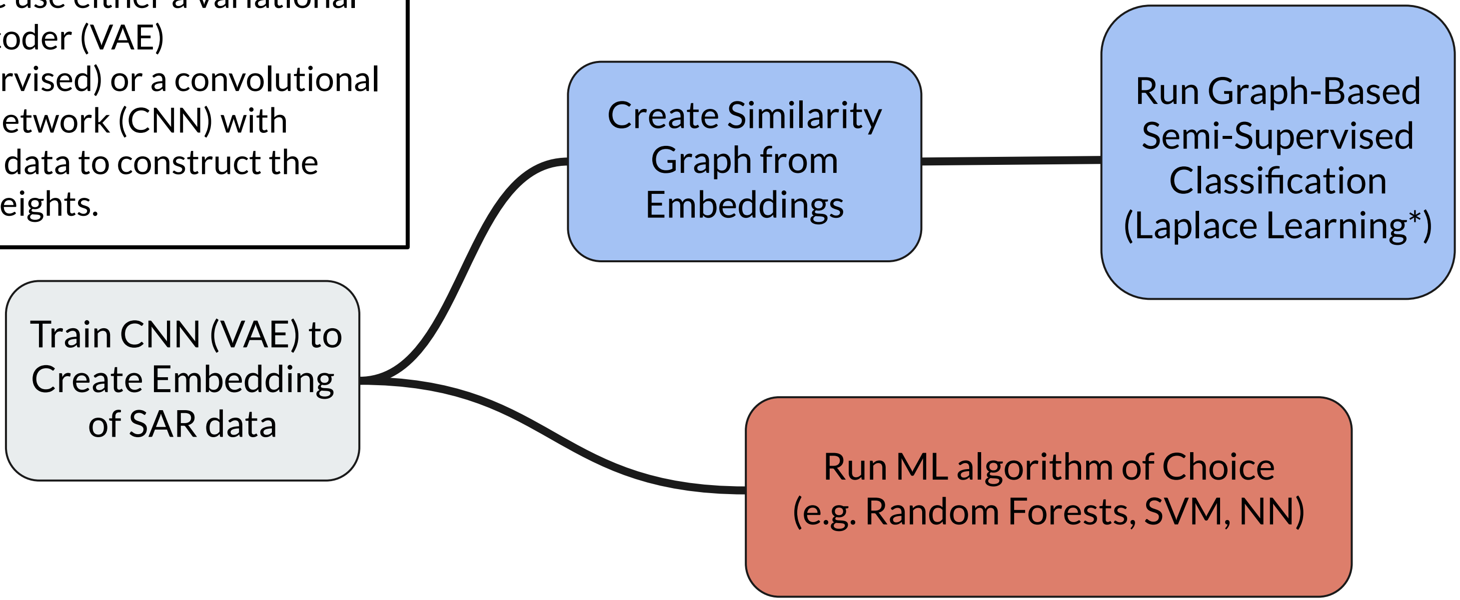
t-Distributed Stochastic Neighbor Embedding (t-SNE) is an unsupervised nonlinear embedding.   Analogous to PCA but preserving only small pairwise distances

- Each color represents a different class:
    - 2s1 gun
    - zsu23-4 gun
    - bmp2 tank
    - t62 tank
    - t72 tank
    - brdm2 truck
    - zill31 truck
    - btr60 transport
    - btr70 transport
    - bulldozer
- Seemingly “natural” clustering structure with minimal overlap
    - **Great candidate for graph-based learning!**



MSTAR Machine Learning Pipeline

Here we use either a variational autoencoder (VAE) (unsupervised) or a convolutional neural network (CNN) with training data to construct the graph weights.

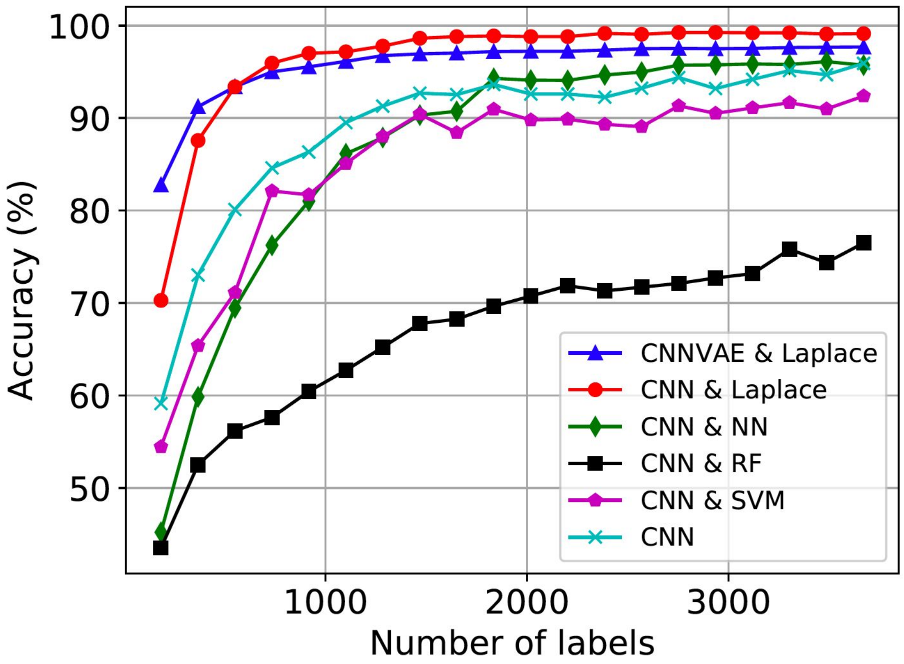


*Laplace Learning (AKA Label Propagation) of Zhu, Gharamani, and Lafferty (2003).

The Case for Graph Learning on MSTAR Data

- With CNN trained on 5%, 10%, 15%, … of training data, report the testing accuracies of various ML algorithms.
  - Provides “upper bound” on hoped for capability of unsupervised representations
- CNN-VAE representations trained on all of training data, **but without any label information**.

Graph Learning appears to be superior at using these learned representations!

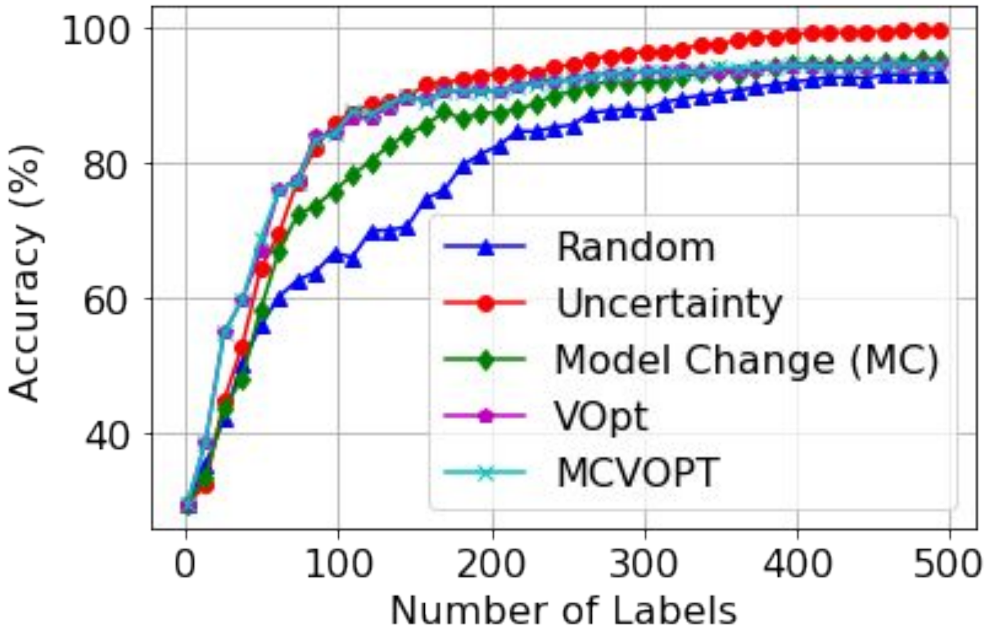


Active Learning on MSTAR

Given effectiveness of graph learning in low-label regime, we apply active learning to further improve the performance.

- Use CNN-VAE representations (i.e. no labeled data required for representation learning)
- Starting with *only 1 labeled point per class*, select 500 labeled points sequentially via the following acquisition functions:
    - Random sampling
    - Uncertainty
    - VOpt
    - Model Change (MC)
    - MCVOPT : A novel combination of VOpt and Model Change acquisition functions

Active Learning with Graph Learning on MSTAR



Here we use only the CNN-VAE for graph construction **without** any labels

Top performing: **Uncertainty Sampling**

- Related to geometry of dataset with many, distinct small clusters
- Exploration **and** Exploitation as a result

Conclusion

- Active learning in conjunction with graph-based learning is effective and efficient way to improve semi-supervised learning
- The natural clusters in MSTAR dataset ideal for graph learning
  - Clustering structure allows even simple acquisition functions (i.e. Uncertainty Sampling) to perform well
- Code available on GitHub
  - HSI experiments (https://github.com/millerk22/model-change-paper/)
  - MSTAR experiments (https://github.com/jwcalder/MSTAR-Active-Learning/)

References 1

[1] A. Bertozzi and A. Flenner, "Diffuse interface models on graphs for classification of high dimensional data," SIAM Review, vol. 58, pp. 293-328, 01 2016.

[2] B. Settles, “Active learning literature survey,” 2010.

[3] J. Calder, B. Cook, M. Thorpe, and D. Slepčev, “Poisson learning: Graph based semi-supervised learning at very low label rates,” in 37th International Conference on Machine Learning, ICML 2020, H. Daume and A. Singh, Eds. International Machine Learning Society (IMLS), 2020, pp. 1283–1293.

[4] J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, pp. 888–905, 2000.

[5] M. Ji and J. Han, “A variance minimization criterion to active learning on graphs,” in Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, ser. Proceedings of Machine Learning Research, N. D. Lawrence and M. Girolami, Eds., vol. 22. La Palma, Canary Islands: PMLR, 21–23 Apr 2012, pp. 556–564. [Online]. Available: http://proceedings.mlr.press/v22/ji12.html

References 2

[6] Z. Meng, E. Merkurjev, A. Koniges, and A. L. Bertozzi, “Hyperspectral image classification using graph clustering methods,” Image Processing On Line, vol. 7, pp. 218–245, 2017.

[7] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, AAAI Press, 2003, pp. 912-919.

[8] K. Miller, H. Li, and A. L. Bertozzi, “Efficient graph-based active learning with probit likelihood via Gaussian approximations,” arXiv preprint arXiv:2007.11126, 2020.

[9] Y. Ma, R. Garnett, and J. Schneider, “Σ-optimality for active learning on Gaussian random fields,” in Advances in Neural Information Processing Systems, vol. 26, 2013.

[10] Z. M. Boyd, E. Bae, X.-C. Tai, and A. L. Bertozzi, “Simplified energy landscape for modularity using total variation,” SIAM Journal on Applied Mathematics, vol. 78, no. 5.

[11] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G.Percus, “Multiclass data segmentation using diffuse interface methods on graphs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1600–1613, 2014.

MSTAR, HSI, and SAR References 1

Y. Sun, Z. Liu, S. Todorovic, and J. Li. Adaptive boosting for sar automatic target recognition. IEEE Transactions on Aerospace and Electronic Systems, 43(1):112–125, 2007.

C. Coman et al. A deep learning sar target classification experiment on mstar dataset. In 2018 19th international radar symposium (IRS), pages 1–6. IEEE, 2018.

M. A. Koets and R. L. Moses. Feature extraction using attributed scattering center models on sar imagery. In Algorithms for Synthetic Aperture Radar Imagery VI, volume 3721, pages 104–115. International Society for Optics and Photonics, 1999.

S. Wagner, K. Barth, and S. Brüggenwirth. A deep learning sar atr system using regularization and prioritized classes. In 2017 IEEE Radar Conference (RadarConf), pages 0772–0777. IEEE, 2017.

H. Wang, S. Chen, F. Xu, and Y.-Q. Jin. Application of deep-learning algorithms to mstar data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pages 3743–3745. IEEE, 2015.

E. R. Keydel, S. W. Lee, and J. T. Moore. Mstar extended operating conditions: A tutorial. In Algorithms for Synthetic Aperture Radar Imagery III, volume 2757, pages 228–242. International Society for Optics and Photonics,1996.

MSTAR, HSI, and SAR References 2

L. C. Potter and R. L. Moses. Attributed scattering centers for sar atr. IEEE Transactions on image processing,6(1):79–91, 1997.

A. M. Raynal, J. Miller, E. Bishop, V. Horndt, and A. Doerry. Shadow probability of detection and false alarm formedian-filtered sar imagery. Sandia National Laboratories Report, 4877, 2014.

Q. Zhao and J. C. Principe. Support vector machines for sar automatic target recognition. IEEE Transactions on Aerospace and Electronic Systems, 37(2):643–654, 2001.

S. Chen, H. Wang, F. Xu, and Y.-Q. Jin. Target classification using the deep convolutional networks for sar images. IEEE transactions on geoscience and remote sensing, 54(8):4806–4817, 2016.

MSTAR, HSI, and SAR References 3

Christophe, Emmanuel & Mailhes, Corinne & Duhamel, P. (2009). Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-D wavelet coding. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society. 17. 2334-46. 10.1109/TIP.2008.2005824.

Perumal, Vasuki. Automatic target classification of manmade objects in synthetic aperture radar images using Gabor wavelet and neural network. Journal of Applied Remote Sensing. Volume 7. 073592-1. 10.1117/1.JRS.7.073592, 2013.