DISCRIMINATIVE MANIFOLD EMBEDDING WITH IMPRECISE, UNCERTAIN, AND
AMBIGUOUS DATA

By

CONNOR H. MCCURLEY

A ORAL QUALIFYING EXAM PROPOSAL PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

RGB image   Red, green, and blue channel image

LIST OF SYMBOLS AND NOMENCLATURE

RGB image   Red, green, and blue channel image

CHAPTER 1
INTRODUCTION

Target detection is a paramount area of research in the field of remote sensing which aims to locate an object or region of interest while suppressing unrelated objects and information. [1, 2]. Target detection can be formulated as a two-class classification problem where samples belonging to a class of interest are discriminated from a global background distribution [3]. The goal of target detection in remote sensing is to correctly classify every true positive instance (TP) in a given scene while indicating as few false alarms (FA) as possible (non-target samples predicted as targets). Most remote sensing target detection techniques in the literature are variations of constrained energy minimization or maximum likelihood with matched filters [1, 2]. The goal of learning in these scenarios is to discover prototypes which represent both the target and background classes that can be used to classify a sample at test. Traditional supervised learning approaches such as these require extensive amounts of highly precise groundtruth to guide algorithmic training. However, acquiring large quantities of accurately labeled training data can be expensive both in terms of time and resources, and in some cases, may even be infeasible to obtain. Consider the following hyperspectral target detection scenario described in [4] and [5]:

Hyperspectral (HS) sensors collect spatial and spectral information of a scene by receiving radiance data in hundreds of contiguous wavelengths [6]. Due to their inherent properties, HS cameras can provide a broad range of spectral information about the materials present in a scene, and are thus useful for detecting targets whose spectral signatures vary from the background. Such examples include airplanes on a tarmac or weeds in a cornfield. While HS provides nice properties for target detection, there are significant challenges when utilizing this modality. First, the spatial resolution of HS cameras is typically much lower than traditional digital cameras. As an example, some HS cameras have spatial resolution of $30m^2$. This implies that objects of interest in a scene, such as an airplane, will actually be contained in

a single pixel along with other materials, such as asphalt. When performing target detection/recognition on that pixel, the HS spectra will not be distinguishable as a single, pure material, but as a sub-pixel mixture where the actual materials present as well as their corresponding proportions are unknown. Second, assuming pure target pixels are available, accurate positioning at the desired resolution may not be. For example, when analyzing a scene from an airplane or satellite, it is necessary to denote the true locations of targets on the ground using a global positioning system (GPS). It is not uncommon, however, to experience GPS error in the order of meters. This implies that a halo of uncertainty potentially surrounds every target pixel in the hyperspectral image (HSI), thus making labeling on the pixel-level difficult.

This example demonstrates inherent infeasibility to obtain accurate sample-level labels due to sensor restrictions on both resolution and accuracy. Furthermore, label imprecision and ambiguity can often be presented from subjectivity between annotators. Many applications such as medical diagnosis and wildlife identification require domain experts to provide accurate data labels. However, there might not always be agreement between expert annotators and humans are prone to making mistakes. For example, when looking at computed tomography (CT) scans for malignant/benign tumors, many doctors would likely determine different pixel-level boundaries denoting a tumor, and in some cases, might even misclassify the detriment of the growth. Similarly, expert wildlife ecologists determining the identity of birds solely from their songs might be uncertain of a species due to corruptive background noise in the audio segment.

Finally, consider the scenarios shown in Figures 1-1 and 1-2. These figures show frames taken from the DSIAC MS-003-DB dataset which demonstrates mid-wave infrared (MWIR) video segments of moving military vehicles taken at approximately 30 frames per second (FPS). Many computer vision algorithms have already been developed to perform target detection using canonical bounding boxes (shown in green in Figure 1-1) [7]. However, drawing tight boxes around targets in each video frame is extremely tedious and time consuming. It would be beneficial if an annotator could provide a less-restrictive form of label, such as a relaxed

bounding box (shown in blue in Figure 1-1 and bottom left in Figure 1-2) or as a small subset of target pixels such as single dot or scribble as shown in Figure 1-2. Labeling burden could be reduced even further if a single frame could be labeled at a high level as "including" or "excluding" target pixels, as shown in Figure 1-3.
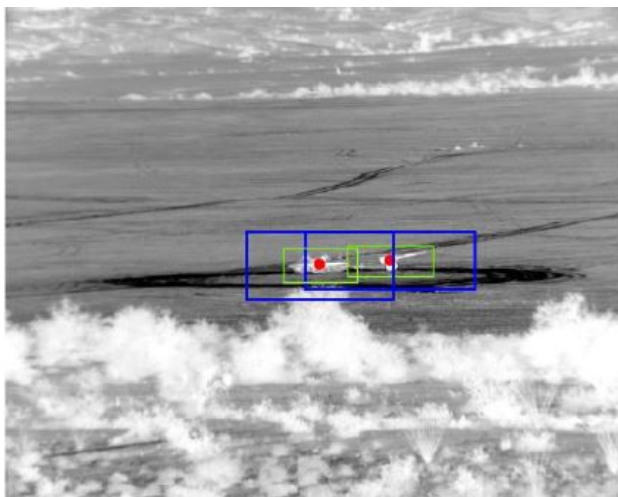
Figure 1-1: A sample frame from the DSIAC MS-003-DB MWIR dataset. Two targets are shown with canonical bounding boxes (green) and relaxed bounding boxes (blue). Red dots represent the centers of the target objects.

Techniques which can address these forms of label ambiguity while achieving comparable or better target detection than standard supervised methods can greatly ease the burdens associated with many remote sensing labeling tasks and allow for the application of pattern recognition techniques which would otherwise be infeasible.

Learning from uncertain, imprecise and ambiguous data has been an active area of research since the late 1990s and is known as multiple instance learning (MIL) [5]. Supervised learning assumes that each training sample is paired with a corresponding classification label. In multiple instance learning, however, the label of each sample is not necessarily known. Instead, MIL approaches learn from groups of data points called *bags*, and each bag concept is paired with a label [8]. Under the two-class classification scenario the bags are labeled as
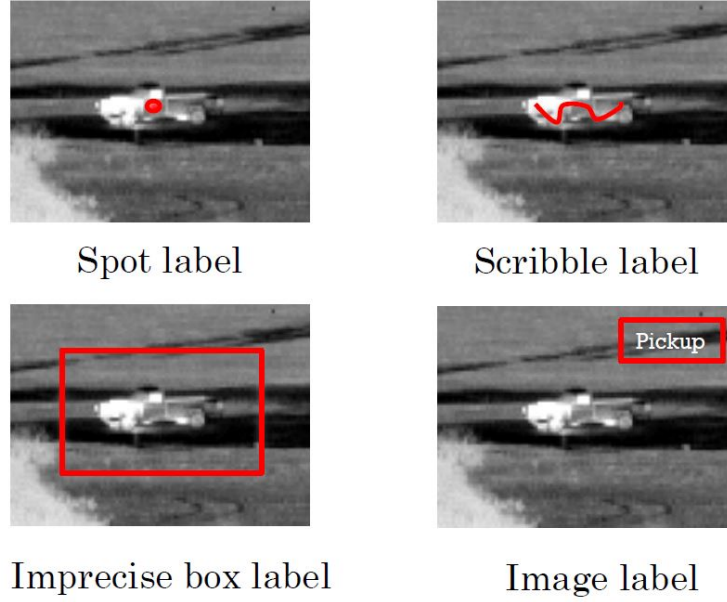
Figure 1-2: Examples of weakly-labeled infrared imagery. The images demonstrate various forms of weak groundtruth around a pickup truck taken with a mid-wave infrared camera. The images show spot, scribble, imprecise bounding box and image-level labels, respectively.

Figure 1-3: Example of image-level labels for binary target detection. Image (a) is denoted to contain pixels belonging to the target class somewhere within the image, while image (b) clearly contains samples solely from the background distribution.

negative if all data points (or *instances*) are known to belong to the background class (not the class of interest). While the actual number of positive and negative instances may be unknown, bags are labeled *positive* if *at least one* instance is known to belong to the target class (also called a "true positive") [3] or to contain a proportion of true target. The goal of learning under the MIL framework is to train a model which can classify a bag as positive or negative (bag-level classification) or to predict the class labels of individual instances (instance-level classification). Consider again the example shown in Figure 1-3. The figure labeled as "Target"

could be considered as a positive bag because, while the image is not accompanied with pixel-level labels, it is known that a true target pixel exists somewhere within the set. Additionally, the image denoted as "Background" could be considered as a negative bag, since it does not contain any pure target pixels. Given data of this type, multiple instance learning could be used to discover target and/or background class representatives which could be used for class discrimination, or a classifier could be trained to label and unseen test image as containing or excluding target pixels (bag-level classification) or to label each individual pixel as belonging to the target class (instance-level classification). As with this example, the MIL problem formulation fits many remote sensing scenarios and is thus an important area of investigation [4].

Multiple instance learning approaches in the literature can be broadly generalized into two categories: learning a concept(s) which effectively describes the positive and/or negative classes, or training a classifier to discriminate bags or individual instances. Existing approaches in the literature have two primary limitations: (1) MIL methods assume that target instances are already separable in the feature space or that a signature(s) can be learned which represents the target class well, while poorly representing the entirety of the negative class, regardless of potential distribution overlap [9]. (2) Current methods are unable to address learning complications associated with high-dimensionality.

These two detrimental limitations can potentially be resolved with the application of *manifold learning*, also commonly referred to as *dimensionality reduction* (DR), *feature embedding*, *geometric machine learning* or *representation learning* in the literature. The goal of manifold learning can be posed as discovering intrinsic (often lower-dimensional) features from the data which meet an overarching objective, such as: preserving variance, finding compressed representations of data, maintaining global or local structure or promoting discriminability in the embedded space [10, 11, 12, 13]. Manifold learning and dimensionality reduction have been studied extensively in the literature and have been used for visualization, classification, redundancy removal, compression and data management, improving computational tractability and

13

efficiency, and arguably the most important application, combating the curse of dimensionality [14, 15, 16, 17, 12, 18, 19, 20, 11].

Dimensionality plays a significant role in determining class separability. Enough features should be incorporated as to adequately describe a class of interest, yet too many features may introduce redundancy, and thus be detrimental to the learning process. The Curse of Dimensionality states that the number of samples needed to characterize the space spanned by an entity grows exponentially with dimensionality [21, 22]. This fact is overwhelming in the context of remote sensing, as it is often both difficult to acquire large quantities of labeled data and there are often only few samples available to describe the target class (such as in air- or space-born HSI target detection). Additionally, dissimilarity metrics often break down in high dimensional spaces, making application of traditional classifiers difficult. *Therefore, it is intuitive that approaches should be developed which can combat the Curse of Dimensionality while also addressing the problems associated with uncertain, imprecise, and ambiguously labeled training data.*

The underlying assumption in using manifold learning for discrimination is that opposing classes either reside on separate manifolds or different regions of a joint, intrinsic manifold, where samples of the same class are close and samples from disparate classes are metrically far. Essentially, the governing class distributions can be described by hyper-surfaces which follow constraints on properties such as continuity and smoothness [23]. The goal of learning in this scenario is to discover embedding functions from the input feature space to a lower-dimensional embedding space where the transformed feature representations allow for improved class discrimination.

Methods for representation learning have recently gained in popularity because they typically result in high levels of classification accuracy. Some of these methods learn features in a supervised manner to obtain more discriminative representations. However, this learning cannot be done directly in MIL because of the uncertainty on the labels. Thus, *adaptation of discriminative feature learning methods would be beneficial to MIL* [9], yet this area of research

has scarcely been explored. The first true experimentation was performed in [24]. In this work, Sun et al. showed that Principal Component Analysis (PCA) failed to incorporate bag-level label information and thus provided poor separation between positive and negative bags. Additionally, Linear Discriminant Analysis (LDA) was used to project bags into a latent space which maximized between-bag separation, while minimizing within-bag dissimilarity. However, LDA often mixed the latent bag representations due to the uncertainty of negative sample distributions in the positive bags. To address these issues, Sun proposed Multiple Instance Dimensionality Reduction (MIDR) which optimized an objective through gradient descent to discover sparse, orthogonal projection vectors into the latent space. Their approach relied on fitting a distribution of negative instances and applying maximum likelihood estimation. This approach was later extended in [25] in attempt to improve sparsity. Both of these methods rely on fitting a distribution of negative instances and determining positive instances as samples which have low likelihood of belonging to the distribution. However, these approaches fail when the the target and background instances are similar. It would, therefore, be beneficial to first transform the instance representations into more discriminable forms. Most existing approaches in the literature extend LDA to distinguish between positive and negative bags [26, 25]. These methods typically rely on costly optimization procedures to maximize an objective for orthogonal and sparse projection vectors. The work in [27] investigated metric learning on sets of data (where each bag was a set) to learn an appropriate similarity metric to compare bags. They do not go as far as to discriminate positive instances within the positive bags, nor do they propose positive target concepts. In fact, virtually all MIL DR methods in the literature are applied, solely, to predict bag-level labels. While bag-level label prediction is useful in many remote sensing applications, such as region of interest (ROI) proposal for anomaly detection, they limit the the ability of learners to classify on the pixel or sample level. Xu et al. proposed an importance term to weight samples believed to be true target exemplars, as those instances are more important in determining the bag-level label [27]. However, to the author's knowledge, no work has been done to investigate instance-level discrimination through

manifold learning/ dimensionality reduction. *With this in mind, the goal of this work is, using only bag-level labels, to find discriminative instance embeddings which allow for accurate sample-level classification.*

To address these points, I propose the following. During this project, techniques will be explored for use in instance-level classification given uncertain and imprecise groundtruth. These methods will be developed as universal approaches for discriminative manifold/feature representation learning and dimensionality reduction and will be evaluated on a variety of sensor modalities, including: mid-wave IR, visible, hyperspectral and multispectral imagery, LiDAR and more. *The aim of this project is to develop dimensionality reduction methods which promote class discriminability and are simultaneously capable of addressing uncertainty and imprecision in training data groundtruth.* The motivating idea is to facilitate instance/concept proposition by increasing instance discriminability under the constraints of multiple instance learning. Roughly, the following research questions will be addressed during the scope of this project:

1. Supervised and semi-supervised manifold learning have proven effective at discovering low-dimensional data representations which provide adequate class separation in the latent space. However, only a handful of manifold learning procedures consider data that is weakly or ambiguously labeled. To address this gap in the literature, a method for weakly-supervised manifold learning will be developed. How does this method of manifold construction compare to state-of-the-art manifold learning techniques as well as alternative MI dimensionality reduction methodologies for instance-level label prediction?

2. Metric embedding has been shown to promote class separability through dimensionality reduction. Without instance-level labels, a method for ranking loss embedding will be developed in conjunction with Objective 1 to improve class separation of individual instances. Additionally, a procedure to select the most influential examples for training will be developed.

3. Do the proposed methods facilitate concept learning/selection? Using alternative state-of-the-art MIL approaches, are the selected target instances/concepts more discriminable with the proposed methods than without? How do the proposed methods compare to the alternatives in terms of representation dimensionality, computational complexity, and promotion of discriminability?

Experiments will be conducted on both synthetic data and real applications such as target detection, scene understanding and semantic segmentation in remote sensing imagery. Datasets will include the DSIAC MS-003-DB Algorithm Development Database, MUUFL Gulfport, Faces in the Wild (LFW) and benchmark MIL classification datasets [28, 29, 30, 31, 32]. Initial results demonstrate the aptitude of the proposed approaches and suggest further development and evaluation of these methods.

CHAPTER 2
BACKGROUND

This chapter provides a literature review on Manifold Learning, including classic approaches, supervised and semi-supervised methods and uses of manifolds for functional regularization. A review of the Multiple Instance Learning framework for learning from weak and ambiguous annotations is provided. Additionally, this chapter reviews the existing literature on metric embedding, focusing heavily on the utilization of contrastive and triplet-based loss evaluation. Reviews describe basic terminology and definitions. Foundational approaches are elaborated and advances are addressed.

2.1 Multiple Instance Learning

Multiple Instance Learning (MIL) was originally proposed in [33] as a method to handle inherent observation difficulties associated with drug activity prediction. This problem, among many others, fits well into the framework of MIL where training labels are associated with sets of data points, called *bags* instead of each individual data points, or *instances*. Under the *standard MIL assumption*, a bag is given a "positive" label if it is known that *at least one* sample in the set represents pure or partial target. Alternatively, a bag is labeled as "negative" if does not contain any positive instances [9]. Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{d \times N}$ be training data where d is the dimensionality of an instance, $\boldsymbol{x}_n$, and N is the total number of training instances. The data is grouped into K *bags*, $\boldsymbol{B} = \{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_K\}$, with associated binary bag-level labels, $\mathcal{L} = \{L_1, \ldots, L_K\}$ where

$$L_k = \begin{cases} 1, & \exists \boldsymbol{x}_{kn} \in \boldsymbol{B}_k^+ \ni l_{kn} = 1 \\ -1, & l_{kn} = 0 \quad \forall \boldsymbol{x}_{kn} \in \boldsymbol{B}_k^- \end{cases} \tag{2-1}$$

and $\boldsymbol{x}_{kn}$ denotes the nth instance in positive bag $\boldsymbol{B}_k^+$ or negative bag $\boldsymbol{B}_k^-$ [3] and l_{kn} denotes the instance-level label on instance $\boldsymbol{x}_{kn}$. Figure 2-1 demonstrates the concept of MIL bags. The objective of learning under MIL is, given only bag-level label information, to fit a model

which can classify bags as either being positive or negative (bag-level classification) or to predict the class labels of individual instances (instance-level classification).
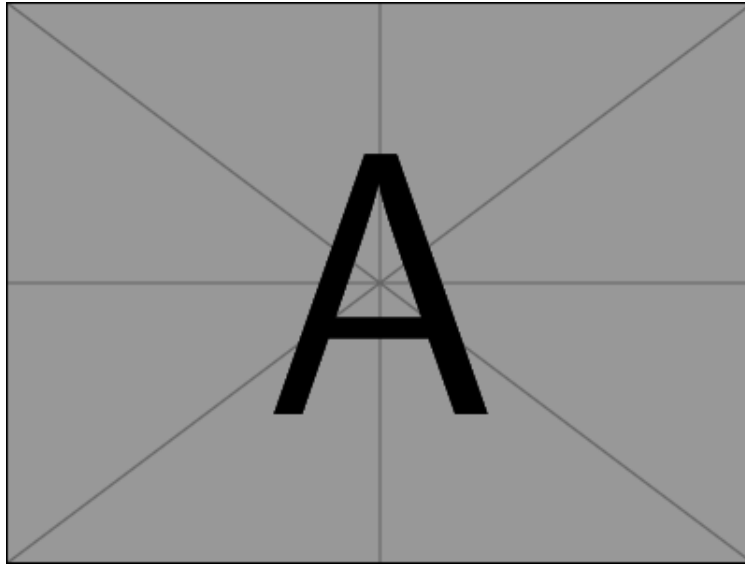

Figure 2-1: Placeholder for examples of positive and negative bag concepts

2.1.1 Multiple Instance Space Paradigm

Multiple Instance Learning via Embedded Instance Selection

2.1.2 Multiple Instance Concept Learning

2.1.3 Multiple Instance Classification

2.1.4 Multiple Instance Boosting

2.2 Manifold Learning

Real-world remote sensing data such as hyperspectral imagery, ground-penetrating radar scans and sonar signals are naturally represented by high-dimensional feature vectors. However, in order to handle such real-world data adequately, its dimensionality usually needs to be reduced [10, 23]. The problem considered in this work is discovering feature representations that promote class discriminability for target or anomaly detection. This is typically achieved in one of two ways. First, features can be projected into a high-dimensional space (such as a Kernel Hilbert Space) using a kernel function. The second option, which is the focus of this

work, is to transform the data into a new (often lower-dimensional) coordinate system which optimizes feature representations for discrimination [34].

The application of *dimensionality reduction* (DR) has proven useful in myriad applications in the literature, such as: visualization of high-dimensional data, classification, redundancy removal, compression and data management, improving computational tractability and efficiency, and reducing the effects of the Curse of Dimensionality [14, 15, 16, 17, 12, 18, 19, 20, 11]. In classification of object entities, it is often assumed that classes can be described by an *intrinsic* subset of representative features which demonstrate geometrical structure [35]. These structures are called intrinsic *manifolds*, and they represent the generating distributions of class objects exactly by the number of degrees of freedom in a dataset [13, 23]. Consider the example shown in Figure 2-2. This classic example demonstrated in [13] shows samples from a pose-estimation dataset **CITE**. While each individual image is represented by a vector of features (pixel intensities in this case) in $\mathbb{R}^{4096}$, the dataset only exhibits three degrees of freedom: 1 light variation parameter and 2 rotation angles. Thus, it is intuitive that the dataset lies on a smooth, intrinsic submanifold spanning three dimensions which inherently capture the degrees of freedom in the data.
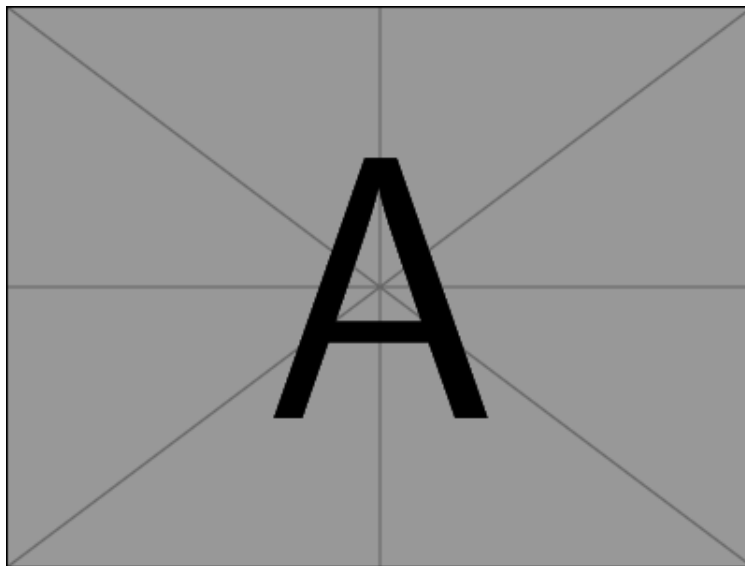

Figure 2-2: Placeholder for example of high D data lying on a low-dimensional sub-manifold.

The goal of manifold Learning is then to discover embedding functions which take data from the input feature space and transform it into a lower-dimensional (ideally intrinsic) coordinate system (also called a *latent space* in the literature) which captures the "useful" properties of the data, while enforcing constraints such as smoothness (the transformation function should not produce sporadic images), continuity (no discontinuous points on the hyper-surface), topological ordering (neighbors in the input space should also be neighbors in the embedded space) or class separability (samples from the same class should fall metrically close to each other in the embedded space and disparate classes should be distinctly far) [34].

This dissertation focuses on investigating the use of manifold learning to increase instance discriminability in the latent space, where labels are solely provided at the bag-level. While there is an expansive literature in unsupervised manifold learning methods, this document will pay special attention to both strictly- and semi-supervised methods, since they are typically adaptations of unsupervised approaches, as well as manifold learning under the MIL framework.

2.2.1 Definition and General Notation

Most studies perform classification or regression after applying unsupervised dimensionality reduction. However, it has been shown that there are advantages of learning the low-dimensional representations and classification/regression models simultaneously [36, 37]. Considering classification as the main goal of dimensionality reduction, this section provides a summary of the current literature in the area.

Given a data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{D \times N}$ where N is the total number of samples and D is the dimensionality of the input feature space, general dimensionality reduction seeks to find a representation $\boldsymbol{U} \in \mathbb{R}^{K \times N}$ with $K \ll D$ that enhances the between-class separation while preserving the intrinsic geometric structure of the data[34]. In other words, it is assumed that the data lie on a smooth manifold $\mathcal{X}$, which is the image of some parameter domain $\mathcal{U} \subset \mathbb{R}^K$ under a smooth mapping $\Psi : \mathcal{U} \to \mathbb{R}^D$. The goal of manifold learning is to discover an inverse mapping to the low-dimensional pre-image coordinates $\boldsymbol{u}_n \in \boldsymbol{U}$ corresponding to

points $x_n \in X$. The data matrices $X = [x_1, \ldots, x_N]$ and $U = [u_1, \ldots, u_N]$ are of size $D \times N$ and $K \times N$, respectively. Traditional classification methods can operate on these *latent* data representations of reduced-dimensionality. The primary difference between traditional, unsupervised manifold learning and supervised approaches is that, in supervised manifold learning, data matrix X is accompanied with a corresponding label vector $y = [y_1, \ldots, y_N]$ indicating the corresponding class labels of each sample in X.

Manifold learning methods can be subdivided into a wide taxonomy of approaches, with *linear* and nonlinear at the root. Nonlinear approaches can be further divided into purely global methods and approaches that capture global structure solely from local information, local. We begin with a review of popular linear manifold learning techniques before moving into the realm of nonlinear approaches. Base, unsupervised methods are reviewed along with corresponding supervised and semi-supervised adaptations.

2.2.2 Linear Manifold Learning

A review of linear manifold learning approaches is provided. Linear approaches are advantageous over nonlinear because they allow for out-of-sample extensions. In other words, linear transformation matrices are learned which can be easily applied on data not included in the training set. However, linear approaches are limited in their abilities to capture irregular data surfaces **CITE**. Principal Component Analysis (PCA), Multi-dimensional Scaling (MDS) and Fisher's Linear Discriminant Analysis (LDA) are reviewed. General approaches are discussed and supervised as well as nonlinear extensions are elaborated. Special focus is given to (LDA), as it is the only inherently-supervised technique out of the included approaches.

2.2.2.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is arguably the most popular (and best-studied) technique for dimensionality reduction and manifold learning. It attempts to learn an orthogonal projection of the input data into a lower-dimensional space, known as the principal subspace, such that the variance of the projected data is maximized [36].

Unsupervised PCA.

Supervised PCA.

Kernel PCA.

2.2.2.2   Multi-Dimensional Scaling (MDS)

Unsupervised MDS.

Supervised MDS.

Kernel MDS.

2.2.2.3   Fisher's Linear Discriminant Analysis (LDA)

Linear LDA.

Kernel LDA (KDA).

2.2.2.4   Locality Preserving Projection (LPP)

2.2.2.5   Graph-based Methods

Nonlinear dimensionality reduction methods typically rely on the use of adjacency graphs. These graphs represent data structure pooled from local neighborhoods of samples. An overview of computational graphs, as well as the two most prominent methods for graph construction in manifold learning are presented. Overview and terminology

K - **Nearest Neighbor Graph**.

ϵ - **Neighborhood Graph**.

Geodesic Distance Approximation.

2.3   Metric Embedding

The concepts of "near" and "far" are very powerful and useful utilities in everyday life. They classify the relationship between two "primitives" as being similar or dissimilar, as well as the degree of compatibility [13]. As an example, a medical doctor might consider a machine learning researcher and a software engineer as being similar (near), because they both perform research for computer applications. However, the same researcher and engineer would likely consider their jobs as being very disparate (far) based on the details of their work. In order to capture this abstraction of distance, a *metric space* is defined as a mathematical construction of this vague generality.

Definition 2.3.1. Metric Space A *metric space* is an ordered pair $(\mathcal{X}, \mathcal{D})$ where $\mathcal{X}$ is a set and $\mathcal{D}$ is a metric on $\mathcal{X}$, or $\mathcal{D} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{X}$, the following holds:

1. Non-negativity: $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) \geq 0$

2. Identity: $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff \boldsymbol{x} = \boldsymbol{y}$

3. Symmetry: $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{D}(\boldsymbol{y}, \boldsymbol{x})$

4. Triangle Inequality: $\mathcal{D}(\boldsymbol{x}, \boldsymbol{z}) \leq \mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) + \mathcal{D}(\boldsymbol{y}, \boldsymbol{z})$

The goal of *metric embedding learning* it to learn a function $f_\theta(\boldsymbol{x}) : \mathbb{R}^D \to \mathbb{R}^K$ which maps semantically similar points from the data input feature space of $\mathbb{R}^D$ onto *metrically close* points in $\mathbb{R}^K$. Similarly, f_θ should map semantically different points in $\mathbb{R}^D$ onto metrically distant points in $\mathbb{R}^K$. The function f_θ is parameterized by θ and can be anything ranging from a linear transformation to a complex non-linear mapping as in the case of deep artificial neural networks [38]. Let $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) : \mathbb{R}^K \times \mathbb{R}^K \to \mathbb{R}$ be a metric function measuring dissimilarity in the embedded space. For succinctness, $\mathcal{D}_{i,j} = \mathcal{D}(f_\theta(\boldsymbol{x_i}), f_\theta(\boldsymbol{x_j}))$ defines the dissimilarity between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, after being embedded.

2.3.1 Ranking Loss

2.3.1.1 Pairwise Loss

2.3.1.2 Contrastive Loss

Definition of Contrastive Loss:

2.3.1.3 Triplet Loss

Definition of Triplet Loss:

Triplet loss was extended in [39] to simultaneously optimize against N negative classes.

2.3.1.4 Large-Margin K-Nearest Neighbors (LMNN)

2.3.1.5 FaceNet

FaceNet is a convolutional neural network which learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity [40].

$$||f(x_i^a) - f(x_i^p)||_2^2 + \alpha < ||f(x_i^a) - f(x_i^n)||_2^2, \quad \forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} \qquad (2\text{-}2)$$

$$\mathcal{L} = ||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + \alpha \qquad (2\text{-}3)$$

2.3.1.6   Siamese Neural Networks

2.3.2   Manifold Regularization

2.3.3   Multiple Instance Metric Learning

CHAPTER 3
PROBLEM DESCRIPTION

CHAPTER 4
EXPERIMENTAL DESIGN

CHAPTER 5
PRELIMINARY WORK

CHAPTER 6
FUTURE TASKS

CHAPTER 7
CONCLUSIONS

REFERENCES

[1] X. Geng, L. Ji, and Y. Zhao, "The basic equation for target detection in remote sensing," 2017.

[2] B. Chaudhuri and S. Parui, "Target detection: Remote sensing techniques for defence applications," *Defence Science Journal*, vol. 45, pp. 285–291, 04 1995.

[3] A. Zare, C. Jiao, and T. Glenn, "Discriminative multiple instance hyperspectral target characterization," *IEEE Trans. Pattern Anal. Mach. Inteli.*, vol. 40, no. 10, pp. 2342–2354, Oct. 2018.

[4] X. Du, "Multiple instance choquet integral for multiresolution sensor fusion," Ph.D. dissertation, Univ. of Missouri, Columbia, MO, Dec. 2017.

[5] J. Bocinsky, "Learning multiple target concepts from uncertain, ambiguous data using the adaptive cosine estimator and spectral match filter," Master's thesis, Univ. of Florida, Gainesville, FL, May 2019.

[6] A. Zare, "Hyperspectral endmember detection and band selection using bayesian methods," Ph.D. dissertation, Univ. of Florida, Gainesville, FL, 2008.

[7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018. [Online]. Available: http://arxiv.org/abs/1804.02767

[8] M. Cook, "Task driven extended functions of multiple instances (td-efumi)," Master's thesis, Univ. of Missouri, Columbia, MO, 2015.

[9] M. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *CoRR*, vol. abs/1612.03365, 2016. [Online]. Available: http://arxiv.org/abs/1612.03365

[10] L. van der Maaten, E. Postma, and H. Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research - JMLR*, vol. 10, 01 2007.

[11] Y. Bengio, A. C. Courville, and P. Vincent, "Unsupervised feature learning and deep learning: A review and new perspectives," *CoRR*, vol. abs/1206.5538, 2012. [Online]. Available: http://arxiv.org/abs/1206.5538

[12] X. Geng, D. Zhan, and Z. Zhou, "Supervised nonlinear dimensionality reduction for visualization and classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 6, pp. 1098–1107, Dec 2005.

[13] N. Thorstensen, "Manifold learning and applications to shape and image processing," Ph.D. dissertation, Ecole Nationale des Ponts et Chaussees, Paris, France, Nov. 2009.

[14] C. Bishop, M. Svensn, and C. K. I. Williams, "Gtm: The generative topographic mapping," vol. 10, pp. 215–234, January 1998. [Online]. Available: https://www.microsoft.com/en-us/research/publication/gtm-the-generative-topographic-mapping/

[15] M. Nickel and D. Kiela, "Poincaré embeddings for learning hierarchical representations," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 6338–6347.

[16] R. Talmon, S. Mallat, H. Zaveri, and R. R. Coifman, "Manifold learning for latent variable inference in dynamical systems," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3843–3856, Aug 2015.

[17] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000. [Online]. Available: https://science.sciencemag.org/content/290/5500/2319

[18] E. J. Palomo and E. Lopez-Rubio, "The growing hierarchical neural gas self-organizing neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 2000–2009, Sep. 2017.

[19] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[20] B. Kgl, D. Wunsch, and A. Zinovyev, *Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58*, 01 2008.

[21] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[22] S. Theodoridis and K. Koutroumbas, "Kernel pca," in *Pattern Recognition, Fourth Edition*, 4th ed. Orlando, FL, USA: Academic Press, Inc., 2008, ch. 6, pp. 351–353.

[23] M. Belkin and P. Niyogi, "Semi-supervised learning on riemannian manifolds," *Machine Learning*, vol. 56, no. 1, pp. 209–239, Jul 2004.

[24] Y.-Y. Sun, M. K. Ng, and Z.-H. Shou, "Multi-instance dimensionality reduction," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, ser. AAAI'10. AAAI Press, 2010, pp. 587–592. [Online]. Available: http://dl.acm.org/citation.cfm?id=2898607.2898702

[25] H. Zhu, L.-Z. liao, and M. K. Ng, "Multi-instance dimensionality reduction via sparsity and orthogonality," *Neural Comput.*, vol. 30, no. 12, pp. 3281–3308, dec 2018. [Online]. Available: https://doi.org/10.1162/neco_a_01140

[26] J. Chai, X. Ding, H. Chen, and T. Li, "Multiple-instance discriminant analysis," *Pattern Recognition*, vol. 47, no. 7, pp. 2517 – 2531, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320314000387

[27] Y. Xu, W. Ping, and A. T. Campbell.

[28] "DSIAC MS-003-DB Algorithm Development Database," https://www.dsiac.org/resources/research-materials/cds-dvds-databases-digital-files/atr-algorithm-development-image.

[29] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "Muufl gulfport hyperspectral and lidar airborne data set," University of Florida, Gainesville, FL, Tech. Rep. 570, October 2013.

[30] X. Du and A. Zare, "Technical report: Scene label ground truth map for muufl gulfport data set," University of Florida, Gainesville, FL, Tech. Rep. 417, April 2017. [Online]. Available: http://ufdc.ufl.edu/IR00009711/00001

[31] T. Glenn, A. Zare, P. Gader, and D. Dranishnikov, "Bullwinkle: Scoring code for sub-pixel targets." [Online]. Available: https://github.com/GatorSense/MUUFLGulfport/

[32] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[33] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Prez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1, pp. 31 – 71, 1997.

[34] E. Vural and C. Guillemot, "A study of the classification of low-dimensional data with supervised manifold learning," *CoRR*, vol. abs/1507.05880, 2018. [Online]. Available: http://arxiv.org/abs/1507.05880

[35] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Dec. 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1248547.1248632

[36] G. Chao, Y. Luo, and W. Ding, "Recent advances in supervised dimension reduction: A survey," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 341–358, 2019.

[37] I. Rish, G. Grabarnik, G. A. Cecchi, F. Pereira, and G. J. Gordon, "Closed-form supervised dimensionality reduction with generalized linear models," 01 2008, pp. 832–839.

[38] A. H. an Lucas Beyer and B. Leibe, "In defense of the triplet loss for person re-identification," *CoRR*, vol. abs/1703.07737, 2017. [Online]. Available: http://arxiv.org/abs/1703.07737

[39] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 1857–1865.

[40] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CoRR*, vol. abs/1503.03832, 2015. [Online]. Available: http://arxiv.org/abs/1503.03832