List of References

Connor H. McCurley

Contents

1 Manifold/ Representation Learning

1.1 Classic Methods

van der Maaten et al. (2007) - *Dimensionality Reduction: A Comparative Review*
Summary:

Jindal and Kumar (2017) - *A Review on Dimensionality Reduction Techniques*
Summary:

Bengio et al. (2012) - *Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives*
Summary:

Tenenbaum et al. (2000) - *A Global Geometric Framework for Nonlinear Dimensionality Reduction*
Summary:

Roweis and Saul (2000) - *Nonlinear Dimensionality Reduction by Locally Linear Embedding*
Summary:

Saul and Roweis (2001) - *An introduction to locally linear embedding*
Summary:

Belkin and Niyogi (2003) - *Laplacian Eigenmaps for Dimensionality Reduction and Data Representation*
Summary:

Bishop et al. (1998) - *GTM: The Generative Topographic Mapping*
Summary:

Delaporte et al. (2008) - *An introduction to diffusion maps*
Summary:

Theodoridis and Koutroumbas (2008b) - *The Karhunen-Loeve Transform*
Summary:

Theodoridis and Koutroumbas (2008a) - *Kernel PCA*
Summary:

Tipping and Bishop (1999) - *Probabilistic Principal Component Analysis*
Summary:

Lawrence (2003) - *Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data*
Summary:

Lawrence (2005) - *Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models*
Summary:

Gorban and Zinovyev (2008) - *Elastic Maps and Nets for Approximating Principal Manifolds and Their Application to Microarray Data Visualization*
Summary:

Lee et al. (2016) - *Learning Representations from Multiple Manifolds*
Summary:

Kokiopoulou and Saad (2007) - *Orthogonal Neighborhood Preserving Projections: A Projection-Based Dimensionality Reduction Technique*
Summary:

Talmon et al. (2015) - *Manifold Learning for Latent Variable Inference in Dynamical Systems*
Summary:

Nickel and Kiela (2017) - *Poincaré Embeddings for Learning Hierarchical Representations*

Summary:

1.2 Competitive Hebbian Learning

Rumelhart and Zipser (1985) - *Feature Discovery by Competitive Learning*
Summary:

Kohonen (1990) - *The self-organizing map*
Summary:

The self-organizing map (SOM) creates spatially organized intrinsic representations of features. It belongs to the category of neural networks which use "competitive learning", or "self-organization". It is a sheet-like artificial neural network in which the cells become tuned to various input patterns through an unsupervised learning process. Only a neighborhood of cells give an active response to the current input sample. The spatial location or coordinates of cells in the network correspond to different modes of the input distribution. The self-organizing map is also a form of vector quantization (VQ). The purpose of VQ is to approximate a continuous probability density function $p(\boldsymbol{x})$ of input vectors $\boldsymbol{x}$ using a finite number of codebook vectors, $\boldsymbol{m}_i$, $i = 1, 2, \ldots, k$. After the "codebook" is chosen, the approximation of $\boldsymbol{x}$ involves finding the reference vector, $\boldsymbol{m}_c$ closest to $\boldsymbol{x}$. The "winning" codebook vector for sample $\boldsymbol{x}$ satisfies the following:

$$||\boldsymbol{x} - \boldsymbol{m}_c|| = \min_i ||\boldsymbol{x} - \boldsymbol{m}_i||$$

The algorithm operates by first initializing a spatial lattice of codebook elements (also called "units"), where each unit's representative is in $\boldsymbol{m}_i \in \mathbb{R}^D$ where D is the dimensionality of the input samples $\boldsymbol{x}$. The training process proceeds as follows. A random sample is selected and presented to the network and each unit determines its activation by computing dissimilarity. The unit who's codebook vector provides the smallest dissimilarity is referred to as the *winner*.

$$c(t) = \arg\min_i d(\boldsymbol{x}(t), \boldsymbol{m}_i(t))$$

Both the winning vector and all vectors within a neighborhood of the winner are updated toward the sample by

$$\boldsymbol{m}_i(t+1) = \boldsymbol{m}_i(t) + \alpha(t) \cdot h_{ci}(t) \cdot [\boldsymbol{x}(t) - \boldsymbol{m}_i(t)]$$

where $\alpha(t)$ is a learning rate which decreases over time and $h_{ci}(t)$ is a neighborhood function which is typically unimodal and symmetric around the location of the winner which monotonically decreases with increasing distance from the winner. A radial basis kernel is typically chosen for the neighborhood function as

$$h_{ci}(t) = \exp\left(-\frac{||\boldsymbol{r}_c - \boldsymbol{r}_i||^2}{2\sigma^2(t)}\right)$$

where the top expression represents the Euclidean distance between units c and i with $\boldsymbol{r}_i$ representing the 2-D location of unit i in the lattice. The neighborhood kernel's bandwidth is typically initialized to a value which covers a majority of the input space and decreases over time such that solely the winner is adapted toward the end of the training procedure.

The SOM essentially performs density estimation of high-dimensional data and represents it in a 2 or 3-D representation. At test time, the dissimilarity between each unit in the map and an input sample are computed. This dissimilarity can be used to effectively detect outliers, thus making the SOM a robust method which can provide confidence values for it's representation abilities.

In this paper, the SOM was applied to speech recognition, but made note of previous uses in robotics, control of diffusion processes, optimization problems, adaptive telecommunications, image compression, sentence understanding, and radar classification of sea-ice.

Rauber et al. (2002) - *The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data*

Summary: The Growing Hierarchical Self-organizing Map (GHSOM) is an extension of the classical SOM. It is an artificial neural network with a hierarchical architecture, composed of individually growing SOMs. Layer 0 is composed of a single neuron representing the mean of the training data. A global stopping criteria is developed as a fraction of the mean quantization error. This means that all units must represent their respective subsets of data an a MQE smaller than a fraction of the 0 layer mean quantization error. For all units not satisfying this criteria, more representation is required for that area of the feature space and additional units are added. After a particular number of training iterations, the quantization errors are computed and the unit with the highest error is selected as the *error unit*. The most dissimilar neighbor of the error unit is chosen is and a row/ column of nodes is injected between them. The growth process continues until a second stopping criteria is met. Any units still not satisfying the global criteria are deemed to need extra representation. Child map are initialized below these units and trained with the subset of data mapped to its parent node.

In conclusion, the GHSOM is a growing self-organizing map architecture which has the ability to grow itself until the feature space is adequately represented. For areas of the space needing a more specific level of granularity, a hierarchical structure is imposed to "fill-in" areas of high density.

The GHSOM has been applied to the areas of finance, computer network traffic analysis, manufacturing and image analysis (Palomo 2017).

Chiang and Gader (1997) - *Hybrid fuzzy-neural systems in handwritten word recognition*

Summary:

Frigui and Gader (2009) - *Detection and Discrimination of Land Mines in Ground-Penetrating Radar Based on Edge Histogram Descriptors and a Possibilistic K-Nearest Neighbor Classifier*

Summary:

Fritzke (1994) - *A Growing Neural Gas Network Learns Topologies*

Summary: Abstract: An incremental network model is introduced which is able to learn the important topological relations in a given set of input vectors by means of a simple Hebb-like learning rule. In contrast to previous approaches like the "neural gas" method of Martinetz and Schulten (1991, 1994), this model has no parameters which change over time and is able to continue learning, adding units and connections, until a performance criterion has been met. Applications of the model include vector quantization, clustering, and interpolation.

In contrast to SOMs and "growing cell structures", which can project data onto non-linear subspaces which are chosen *a priori*, the GNG is able to adapt its topology to match that of the input data distribution. The growing process continues until a pre-defined level of quantization error has been reached.

The base algorithm is outlined in Palomo (2017), *Growing Hierarchical Neural Gas Self-Organizing Network*.

Palomo and Lopez-Rubio (2017) - *The Growing Hierarchical Neural Gas Self-Organizing Neural Network*

Summary:

Abstract: The growing neural gas (GNG) self-organizing neural network stands as one of the most successful examples of unsupervised learning of a graph of processing units. Despite its success, little attention has been devoted to its extension to a hierarchical model, unlike other models such as the self-organizing map, which has many hierarchical versions. Here, a hierarchical GNG is presented, which is designed to learn a tree of graphs. Moreover, the original GNG algorithm is improved by a distinction between a growth phase where more units are added until no significant improvement in the quantization error is obtained, and a convergence phase where no unit creation is allowed. This means that a principled mechanism is established to control the growth of the structure. Experiments are reported, which demonstrate the self-organization and hierarchy learning abilities of our approach and its performance for vector quantization applications. Experiments were performed in structure learning, color quantization, and video sequence clustering.

The aim of this method was to improve the adaptation ability of the Growing Hierarchical Self-Organizing Map proposed by Rauber (2002). This was to be done through the extension of the Growing Neural Gas, which disposes of the fixed lattice topology enforced by the SOM. Addtionally, the GNG learns a dynamic graph with variable numbers of neurons and connections. The graph represents the input data in a more plastic and flexible way than the fixed-topology map.

All clustering methods that learn a hierarchical structure have advantages even when used for non-hierarchical data. The learned hierarchical structure can be pruned at several levels, which yields alternative representations of the input data set at different levels of detail. This can be used to visualize a data set in coarser or more detailed way. For vector quantization applications, the different pruning levels correspond to smaller or larger codebooks, so that a balance can be attained between the size of the codebook and the quantization error within the same hierarchical structure.

The growing hierarchical neural gas (GHNG) model is defined as a tree of self-organizing graphs. Each graph is made of a variable number of neurons or processing units, so that its size can grow or shrink during learning. In addition, each graph is the child of a unit in the upper level, except for the top level (root) graph. The training procedure is described by the following:

Each graph begins with $H \geq 2$ units and one or more undirected connections between them. Both the units and connections can be created and destroyed during the learning process. It is also not necessary that the graph is connected. Let the training set be denoted as $\mathcal{S}$ with $\mathcal{S} \subset \mathbb{R}^D$, where D is the dimensionality of the input space. Each unit $i \in \{1, \ldots, H\}$ has an associated prototype $\boldsymbol{w}_i \in \mathbb{R}^D$ and an error variable $e_i \in \mathbb{R}$, $e_i \geq 0$. Each connection has an associated age, which is a nonnegative integer. The set of connections will be notetd as $A \subseteq \{1, \ldots, H\} \times \{1, \ldots, H\}$. The learning mechanism for the GHNG is based on the original GNG, but includes a novel procedure to control the growth of the graph. First, a growth phase is performed where the graph is allowed to enlarge until a condition is met, which indicates that further growing would provide no significant improvement in the quantization error. After that, a convergence phase is executed where no unit creation is allowed in order to carry out a fine tuning of the graph. the leraning algorithm is provided in the following steps.

1. Start with two units ($H = 2$) joined by a connection. Each prototype is initialized to a sample drawn at random from $\mathcal{S}$. The error variables are initialized to zero. The age of the connection is initialized to zero.

2. Draw a training sample $\boldsymbol{x}_t \in \mathbb{R}^D$ at random from $\mathcal{S}$.

3. Find the nearest unit q and second nearest unit s in terms of Euclidean distance

$$q = \underset{i \in \{1,\ldots,H\}}{\arg\min} ||\boldsymbol{w}_i(t) - \boldsymbol{x}(t)||$$
$$s = \underset{i \in \{1,\ldots,H\}-\{q\}}{\arg\min} ||\boldsymbol{w}_i(t) - \boldsymbol{x}(t)||$$

4. Increment the age of all edges departing from q

5. Update the winning unit's error variable, e_q

$$e_q(t+1) = e_q(t) + ||\boldsymbol{w}_q(t) - \boldsymbol{x}_t||$$

I believe the author's experimental approach did not take advantage of the method's strengths. The author's only demonstrated experiments in vector quantization, and used corresponding metrics. This method could be used to represent manifold topology of differing dimensionality. This could be useful in HSI imagery, for example where different environment patches require manifold representations of various dimensionality. Additionally, this could potentially be used to handle the sensor fusion problem with sensor loss/ drop-out.

Sun et al. (2017) - *Online growing neural gas for anomaly detection in changing surveillance scenes*
Summary:

Lopez-Rubio and Palomo (2011) - *Growing Hierarchical Probabilistic Self-Organizing Graphs*
Summary:

Palomo and Lopez-Rubio (2016) - *Learning Topologies with the Growing Neural Forest*
Summary:

1.3 Deep Learning

Goodfellow et al. (2016) - *Deep Learning*
Summary:

Haykin (2009) - *Neural networks and learning machines*
Summary:

Dai et al. (2017) - *dden Talents of the Variational Autoencoder*
Summary:

Rojas (1996) - *Associative Networks*
Summary:

2 Information Measures

Arandjelovic et al. (2005) - *Face recognition with image sets using manifold density divergence*
Summary:

Wang et al. (2012) - *ManifoldManifold Distance and its Application to Face Recognition With Image Sets*
Summary:

3 Manifold Regularization

Tsang and Kwok (2007) - *Large-Scale Sparsified Manifold Regularization*
Summary:

Ren et al. (2017) - *Unsupervised Classification of Polarimetirc SAR Image Via Improved Manifold Regularized Low-Rank Representation With Multiple Features*
Summary:

Belkin et al. (2006) - *Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples*
Summary:

Ratle et al. (2010) - *Semisupervised Neural Networks for Efficient Hyperspectral Image Classification*

Summary:

Li et al. (2015) - *Approximate Policy Iteration with Unsupervised Feature Learning based on Manifold Regularization*
Summary:

Meng and Zhan (2018) - *Zero-Shot Learning via Low-Rank-Representation Based Manifold Regularization*
Summary:

4 Multiple Instance Learning

4.1 Multiple Instance Concept Learning

Bocinsky (2019) - *Learning Multiple Target Concepts from Uncertain, Ambiguous Data Using the Adaptive Cosine Estimator and Spectral Match Filter*
Summary:

Jiao (2017) - *Target Concept Learning From Ambiguously Labeled Data*
Summary:

McCurley et al. (2019) - *Comparison of hand-held WEMI target detection algorithms*
Summary:

Bocinsky et al. (2019) - *Investigation of initialization strategies for the Multiple Instance Adaptive Cosine Estimator*
Summary:

Zare et al. (2015) - *Multiple instance dictionary learning for subsurface object detection using handheld EMI*
Summary:

Cook (2015) - *Task driven extended functions of multiple instances (TD-eFUMI)*
Summary:

Cook et al. (2016) - *Buried object detection using handheld WEMI with task-driven extended functions of multiple instances*
Summary:

Zare et al. (2016) - *Multiple Instance Hyperspectral Target Characterization*
Summary:

Jiao and Zare (2017) - *Multiple instance hybrid estimator for learning target signatures*
Summary:

Xiao et al. (2017) - *A Sphere-Description-Based Approach for Multiple-Instance Learning*
Summary:

Cheplygina et al. (2019) - *Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis*
Summary:

Li et al. (2017) - *Cross-validated smooth multi-instance learning*
Summary:

Cheplygina et al. (2016) - *Dissimilarity-Based Ensembles for Multiple Instance Learning*
Summary:

Wang et al. (2017) - *Incorporating Diversity and Informativeness in Multiple-Instance Active Learning*
Summary:

Hajimirsadeghi and Mori (2017) - *Multi-Instance Classification by Max-Margin Training of Cardinality-Based Markov Networks*
Summary:

Du et al. (2016) - *Multiple Instance Choquet integral for classifier fusion*
Summary:

Ilse et al. (2018) - *Attention-based Deep Multiple Instance Learning*
Summary:

Karem and Frigui (2016) - *Multiple Instance Learning with multiple positive and negative target concepts*
Summary:

Xiao et al. (2017) - *Multiple-Instance Ordinal Regression*
Summary:

Gao et al. (2017) - *C-WSL: Count-guided Weakly Supervised Localization*
Summary:

Li et al. (2017) - *Multi-View Multi-Instance Learning Based on Joint Sparse Representation and Multi-View Dictionary Learning*
Summary:

Cao et al. (2016) - *Weakly Supervised Vehicle Detection in Satellite Images via Multi-Instance Discriminative Learning*
Summary:

Dietterich et al. (1997) - *Solving the multiple instance problem with axis-parallel rectangles*
Summary:

Maron and Lozano-Pérez (1998) - *A Framework for Multiple-instance Learning*
Summary:

Maron and Ratan (1998) - *Multiple-Instance Learning for Natural Scene Classification*
Summary:

Carbonneau et al. (2016) - *Multiple Instance Learning: A Survey of Problem Characteristics and Applications*
Summary:

Zhang and Goldman (2002) - *EM-DD: An Improved Multiple-Instance Learning Technique*
Summary:

Zare and Jiao (2014) - *Extended Functions of Multiple Instances for target characterization*
Summary:

Jiao et al. (2018) - *Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection*
Summary:

4.2 Multiple Instance Classification

Cao et al. (2016) - *Weakly Supervised Vehicle Detection in Satellite Images via Multi-Instance Discriminative Learning*
Summary:

4.3 Multiple Instance Regression

Trabelsi and Frigui (2018) - *Fuzzy and Possibilistic Clustering for Multiple Instance Linear Regression*
Summary:

Ruiz et al. (2018) - *Multi-Instance Dynamic Ordinal Random Fields for Weakly Supervised Facial Behavior Analysis*
Summary:

-
Summary:

-
Summary:

-
Summary:

4.4 Applications

-
Summary:

5 Fusion

5.1 Classical Approaches

5.1.1 General Approach

Mohandes et al. (2018) - *Classifiers Combination Techniques: A Comprehensive Review*
Summary:

Ruta and Gabrys (2000) - *An Overview of Classifier Fusion Methods*
Summary:

"The objective of all decision support systems (DSS) is to create a model, which given a minimum amount of input data/information, is able to produce correct decisions." "the solution might be just to combine existing, well performing methods, hoping that better results will be achieved. Such fusion of information seems to be worth applying in terms of uncertainty reduction. Each of individual methods produces some errors, not mentioning that the input information might be corrupted and incomplete. However, different methods performing on different data should produce different errors, and assuming that all individual methods perform well, combination of such multiple experts should reduce overall classification error and as a consequence emphasize correct outputs." "Fusion of data/information can be carried out on three levels of abstraction closely connected with the flow of the classification process: data level fusion, feature level fusion, and classifier fusion" This paper focused on the later method of classifier fusion. This process can essentially be categorized into two eruditions. The first methods put emphasis on the classifier structure and do not do anything with the outputs until the combination process finds the best classifier or a selected group of classifiers. Then their outputs are taken as a final decision or used for further processing. The second category operates primarily on classifier outputs and can be further divided.
There are three possibles types of output labels generated by individual classifiers. Crisp labels provide the lowest amount of information for fusion, as no information about potential alternatives is available. Some additional information can be gleaned from labels in the form of class rankings. However, fusion methods operating on classifiers with soft/fuzzy outputs can be expected to produce the greatest improvement in classification performance. (Connor Note: This is valuable in terms of outlier rejection as well!). The following explains an overview of classifier fusion methods operating on single class labels, class rankings, and fuzzy measures, respectively.

Methods operating on classifiers:
Dynamic Classifier Selection (DCS) methods replect the tendency to extract a single best classifier instead of mixing many different classifiers, by attempting to determing the single classifier which is most likely to produce the correct classification label for an input sample. Only the output of the selected classifier is taken as a final decision. The classifier selection process includes a partitioning of the input samples. A classifier is is for each partition is selected locally. All DCS methods rely on strong training data and by choosing only locally best classifier. **They potentially lose some useful information from other well-performing classifiers.** Classifiers and their combination functions are typically organized in parallel and simultaneously and separately get their outputs as in input for a combination function. A more reasonable approach, however, is **to organize all classifiers into groups and to apply different fusion methods for each group.** A very important factor for the success of this method is the diversity of classifier types, training data, and methods involved. **Any classification improvement may only be achieved if the total information uncertainty is reduced.** This in turn depends on the diversity of information supporting different classification methods. **The same goal can be achieved by reduction of errors produced by individual classifiers.** *Hierarchical Mixture of Experts* (HME) is an example of a fusion method whose strength comes from classifier's structure. It is a supervised learning method based on the *divide-and-conquer* principle. It is organized as a tree-like structure of leaves. Each leaf represents an individual expert in the network, each of which tries to solve a local supervised learning problem. The outputs of the elements of the same node are partitioned and combined by the gating network and the total output of the node is given as a convex combination. The expert networks are trained to increase the posterior probability according to Bayes rule. A number of learning algorithms can be applied to tune the mixture model. *Expectation-Maximization* (EM) is often used to learn the model parameters. *The HME technique does not seem to be applicable to large-dimensional datasets.*
Fusing Single Class Labels: Classifiers producing crisp, single-class labels (SCL) provide the least amount of useful information for the combination process. The two most common techniques for fusing SCL classifiers are *Generalized Voting* and *Knowledge-Behavior Space* methods.
Voting Methods:
Voting strategies can be applied to a multiple classifier system assuming that each classifier gives a single class label as as output and no training data are available. While there are many methods for combining these labels, they all lead to the following generalized voting definition. Let the output of the classifiers form the decision vector $\boldsymbol{d} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_n]^T$ where $\boldsymbol{d}_i \in \{c_1, c_2, \ldots, c_m, r\}$, c_i denotes the class label of the i-th class and r the rejection of assigning the input sample to any classes. The binary characteristic function is defined as follows: (Have not input math)
Class Ranking Based Techniques:
There are two primary methods for fusion of class rankings. *Class set reduction* (CSR) attempts to reduce the number of eligible classes by compromising between minimizing the class set size and maximizing the likelihood of inclusion in the true class. This is typically performed through the *intersection or union of neighborhoods.* The second popular CSR method is *Class Set Reordering* (CSRR) which tries to improve the overall rank of the true class through techniques such as the *Highest Rank Method*, *Borda Count*, or *Logistic Regression.*
Soft-Label Classifier Fusion:
Soft labels are outputs in the range $[0, 1]$ and are typically referred to as *fuzzy measures*, which cover all known measures of evidence: probability, possibility, necessity, belief, and plausibility. Each of these measures are used to describe different dimensions of information uncertainty. This class of fusion attempts to reduce the level of uncertainty by maximizing suitable measures of evidence. Common methods for this type of fusion include: Bayesian, Fuzzy Integrals, Dempster-Shaffer Combination, Fuzzy Templates, Product of Experts, and Artificial Neural Networks. *Bayesian* methods can be applied under the condition that the outputs of the classifier are expressed as posterior probabilities. Typical methods of Bayesian fusion include Bayes Average and Bayes Belief Integration. *Fuzzy Integrals* aim at searching for the maximal agreement between the real possibilities relating to objective evidence and the expectation, g, which defines the level of importance of a subset of sources. The concept of fuzzy integrals arises from the λ-fuzzy measure, g, developed by Sugeno. Common methods for Fuzzy Integration include the Sugeno Fuzzy Integral, Choquet Fuzzy Integral, and Webster Fuzzy Integral. *Product of Experts* combines different probabilistic models of the same data by performing a weighted average of individual probability distributions.

Tulyakov et al. (2008) - *Review of Classifier Combination Methods*
Summary:

Hackett and Shah (1990) - *Multi-sensor fusion: a perspective*
Summary:

Multi-Sensor fusion deals with the combination of complementary and sometimes competing sensor data into a reliable estimate of the environment to achieve an output which is better than the modalities, individually. Multi-sensor fusion has been used in target recognition, autonomous robot navigation, automatic manufacturing, scene segmentation, sensor modeling, and object recognition. *Sensor fusion combines the outputs from two or more devises that retrieve a particular property of the environment.* Each sensor's measurements are, in general, imprecise and contain errors and uncertainties, so the consensus of multiple sensors measuring the same property can reduce uncertainty and reduce measurement ambiguity. Every sensor modality is sensitive to a different property of the environment; it is necessary to use multiple sensors in order to address these sensitivities. *Sensor fusion deals with the selection of a proper model for each sensor, and identification of an appropriate fusion method.* There are several methods for combining multiple data sources. A few are: deciding, guiding, averaging, Bayesian statistics, and integration. Deciding is the use of a particular data source during a certain time of the fusion process, usually based on some confidence measure. Averaging is the weighted combination of several data sources. This type of fusion ensures all sensors contribute to the fusion process, but not all to the same degree. Guiding is the use of one or more sensors to focus the attention of another sensor on some part of the scene. Integration is the delegation of carious sensors to particular tasks, thus eliminating redundancy in sensor measurements. The most simple method of fusion uses raw data of the same property obtained by multiple sensors of the same type. Multi-sensor integration is the use of several sensors in a sequential manner.
Data from different sensors must be put into equivalent forms to allow for fusion. In order for data from multiple sources to be fused, there must be some method to relate data points from one sensor with corresponding data points from the other sensors. The *registered* data points allow for easy gathering of sensor information about one particular point in the scene.
Fusion methods can be broadly classified into two categories, *direct* and *indirect*. Direct fusion combines raw sensor measurements while indirect methods transform the sensor data to be fused.

Zhang (2010) - *Multi-source remote sensing data fusion: Status and trends*
Summary:

5.1.2 Hierarchical Mixture of Experts

Jordan and Jacobs (1993) - *Hierarchical mixtures of experts and the EM algorithm*
Summary:

Yuksel et al. (2012) - *Twenty Years of Mixture of Experts*
Summary:

Beyer et al. (2009) - *Heterogeneous mixture-of-experts for fusion of locally valid knowledge-based submodels*
Summary:

Shazeer et al. (2017) - *Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer*
Summary:

5.1.3 Choquet Integral

Du (2017) - *Multiple Instance Choquet Integral For MultiResolution Sensor Fusion*
Summary:

Ryan E. Smith (2017) - *Aggregation of Choquet integrals in GPR and EMI for handheld platform-based explosive hazard detection*
Summary:

Smith et al. (2017) - *Genetic programming based Choquet integral for multi-source fusion*
Summary:

Du and Zare (2019) - *Multiple Instance Choquet Integral Classifier Fusion and Regression for Remote Sensing Applications*
Summary:

Anderson et al. (2017) - *Binary fuzzy measures and Choquet integration for multi-source fusion*
Summary:

Du and Zare (2018) - *Multi-Resolution Multi-Modal Sensor Fusion For Remote Sensing Data With Label Uncertainty*
Summary:

Gader et al. (2004) - *Multi-sensor and algorithm fusion with the Choquet integral: applications to landmine detection*
Summary:

5.1.4 Deep Learning

L.Jian et al. (2019) - *A Symmetric Encoder-Decoder with Residual Block for Infrared and Visible Image Fusion*
Summary:

5.1.5 Graph-Based

Vivar et al. (2019) - *Multi-modal Graph Fusion for Inductive Disease Classification in Incomplete Datasets*
Summary:

5.2 Co-registration

Dawn et al. (2010) - *Remote Sensing Image Registration Techniques: A Survey*
Summary:

Brigot et al. (2016) - *Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images*
Summary:

Zitov and Flusser (2003) - *Image registration methods: a survey*
Summary:

5.2.1 Geocoding

5.2.2 Similarity Measures

5.2.3 Transformation, Interpolation, Re-sampling

5.2.4 Conflation

5.3 Multi-resolution Fusion

5.4 Fusion of Mixed Data Types

Butenuth et al. (2007) - *Integration of heterogeneous geospatial data in a federated database*
Summary:

Guo (2019) - *Latent Variable Algorithms for Multimodal Learning and Sensor Fusion*

Summary:

Zhang et al. (2019) - *Fusion of Heterogeneous Earth Observation Data for the Classification of Local Climate Zones*

Summary:

5.5 Unsorted

Shen et al. (2016) - *An Integrated Framework for the SpatioTemporalSpectral Fusion of Remote Sensing Images*

Summary:

-

Summary:

6 Outlier/ Adversarial Detection

7 Army

Hall et al. (2018) - *Probabilistic Object Detection: Definition and Evaluation*

*Summary:*A probabilistic object detection metric (PDQ - Probability-based Detection Quality) was proposed, thus defining the new task of defining probabilistic object detection metrics. The ability of deep CNNs to quantify both *epistemic* and *aleatoric uncertainty* is paramount for deployment safety-critical applications. PDQ aims to measure the accuracy of an image object detector in terms of its label uncertainty and spatial quality. This is achieved through two steps. First, a detector must reliably quantify its *semantic uncertainty* by providing full probability distributions over known classes for each detection. Next, the detectors must quantify spatial uncertainty by reporting *probabilistic bounding boxes*, where the box corners are modeled as normally distributed. A loss function was constructed to consider both label and spatial quality when providing a final detection measure. The primary benefit of this method is that it provides a measure for the level of uncertainty in a detection.

Is it possible to replace the probabilistic metric with a possibilistic one? Could this be more effective at handling outlying cases?

Mahalanobis and McIntosh (2019) - *A comparison of target detection algorithms using DSIAC ATR algorithm development data set*

*Summary:*The authors provided an initial characterization of detection performance on the DSIAC dataset using the *Faster R-CNN* algorithm and *Quadratic Correlation Filter (QCF)*. Performance was evaluated on two datasets, “easy” and “difficult”, where the difficulty was determined by number of pixels on target and local contrast. Under difficult conditions, the Faster R-CNN algorithm achieved noteworthy performance, detecting as much as 80% of the targets at a low false alarm rate of 0.01 FA/Square degree. The dataset was limited by a lack of background diversity.

Tanner and Mahalanobis (2019) - *Fundamentals of Target Classification Using Deep Learning*

*Summary:*A shallow CNN was utilized for ATR on the DSIAC MWIR dataset. The goal of the study was to determine the range of optimal thresholds which would optimally separate the target and clutter class distributions defined by the CNN predictions (output of softmax), as well as determine an upper bound on the number of training images required for optimizing performance. The shallow CNN (5 layers) and a Difference of Gaussians (DoG), which finds regions of high intensity on dark backgrounds were used to detect and classify targets. The CNN could correctly classify 96% of targets as targets and as few as 4% of clutter as targets. It was found that the DoG detector failed when the targets were small (long range) or if the overall image was bright (infrared taken during the daytime). It was also determined that guessing the bright pixels were at the center of the targets was a bad assumption. (The brightest part of a target is not necessarily at its center.)

Li2 - *Collaborative sparse priors for multi-view ATR*
Summary:

Kokiopoulou and Frossard (2010) - *Graph-based classification of multiple observation sets*
Summary:

8 Segmentation

Caselles et al. (1997) - *Geodesic Active Contours*
Summary:

Álvarez et al. (2010) - *Morphological Snakes*
*Summary:*The authors introduce a morphological approach to curve evolution. Snakes or curves iteratively solve partial differential equations (PDEs). By doing so, the shape of the snake deforms to minimize the internal and external energies along its boundary. The internal component keeps the curve smooth, while the external component attaches the curve to image structures such as edges, lines, etc. Curve evolution is one of the most widely used image segmentation/ object tracking algorithms. The main contribution of the paper is a new morphological approach to the solution of the PDE associated with snake model evolution. They approach the solution using only inf-sup operators which has the main benefit of providing simpler level sets (0 outside the contours and 1 inside).

Márquez-Neila et al. (2014) - *A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces*
Summary:

References

L. Álvarez, L. Baumela, P. Henríquez, and P. Márquez-Neila. Morphological snakes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2197–2202, June 2010. doi: 10.1109/CVPR.2010.5539900.

D. T. Anderson, M. A. Islam, R. King, N. H. Younan, J. R. Fairley, S. Howington, F. Petry, P. Elmore, and A. Zare. Binary fuzzy measures and choquet integration for multi-source fusion. In *2017 International Conference on Military Technologies (ICMT)*, pages 676–681, May 2017. doi: 10.1109/MILTECHS.2017.7988843.

O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 581–588 vol. 1, June 2005. doi: 10.1109/CVPR.2005.151.

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1162/089976603321780317.

M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.*, 7:2399–2434, December 2006. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1248547.1248632`.

Y. Bengio, A. C. Courville, and P. Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR*, abs/1206.5538, 2012. URL `http://arxiv.org/abs/1206.5538`.

J. Beyer, K. Heesche, W. Hauptmann, and C. Otte. Heterogeneous mixture-of-experts for fusion of locally valid knowledge-based submodels. 01 2009.

C. Bishop, M. Svensn, and C. K. I. Williams. Gtm: The generative topographic mapping. 10:215–234, January 1998. URL `https://www.microsoft.com/en-us/research/publication/gtm-the-generative-topographic-mapping/`.

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.

J. Bocinsky, C. H. McCurley, D. Shats, and A. Zare. Investigation of initialization strategies for the multiple instance adaptive cosine estimator. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV, 110120N*, volume 11012 of *Proc.SPIE*, May 2019. doi: 10.1117/12.2519463.

G. Brigot, E. Colin-Koeniguer, A. Plyer, and F. Janez. Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7):2923–2939, July 2016. ISSN 1939-1404. doi: 10.1109/JSTARS.2016.2578362.

M. Butenuth, G. v. Gsseln, M. Tiedge, C. Heipke, U. Lipeck, and M. Sester. Integration of heterogeneous geospatial data in a federated database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62:328–346, 10 2007. doi: 10.1016/j.isprsjprs.2007.04.003.

L. Cao, F. Luo, L. Chen, S. Yihan, H. Wang, C. Wang, and R. Ji. Weakly supervised vehicle detection in satellite images via multi-instance discriminative learning. *Pattern Recognition*, 64, 12 2016. doi: 10.1016/j.patcog.2016.10.033.

M. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon. Multiple instance learning: A survey of problem characteristics and applications. *CoRR*, abs/1612.03365, 2016. URL `http://arxiv.org/abs/1612.03365`.

V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, Feb 1997. ISSN 1573-1405. doi: 10.1023/A:1007979827043. URL `https://doi.org/10.1023/A:1007979827043`.

V. Cheplygina, D. M. J. Tax, and M. Loog. Dissimilarity-based ensembles for multiple instance learning. *IEEE Transactions on Neural Networks and Learning Systems*, 27(6):1379–1391, June 2016. ISSN 2162-237X. doi: 10.1109/TNNLS.2015.2424254.

V. Cheplygina, M. Bruijne, and J. P. W. Pluim. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis*, 54:280 – 296, 2019. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2019.03.009.

J. Chiang and P. D. Gader. Hybrid fuzzy-neural systems in handwritten word recognition. *IEEE Transactions on Fuzzy Systems*, 5(4):497–510, Nov 1997. ISSN 1063-6706. doi: 10.1109/91.649901.

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

M. Cook, A. Zare, and D. K. C. Ho. Buried object detection using handheld wemi with task-driven extended functions of multiple instances. In *Proc. SPIE 9823, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXI, 98230A*, volume 9823 of *Proc. SPIE*, pages 9823 – 9823 – 9, Apr. 2016. doi: 10.1117/12.2223349.

B. Dai, Y. Wang, J. Aston, G. Hua, and D. Wipf. Hidden talents of the variational autoencoder, 2017.

S. Dawn, V. Saxena, and B. Sharma. Remote sensing image registration techniques: A survey. In A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, and J. Meunier, editors, *Image and Signal Processing*, pages 103–112, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-13681-8.

J. Delaporte, B. M. Herbst, W. Hereman, and S. Van der Walt. An introduction to diffusion maps. 2008.

T. G. Dietterich, R. H. Lathrop, and T. Lozano-Prez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1):31 – 71, 1997. ISSN 0004-3702. doi: https://doi.org/10.1016/S0004-3702(96)00034-3.

X. Du. *Multiple Instance Choquet Integral For MultiResolution Sensor Fusion.* PhD thesis, Univ. of Missouri, Columbia, MO, Dec. 2017.

X. Du and A. Zare. Multi-resolution multi-modal sensor fusion for remote sensing data with label uncertainty. *CoRR*, abs/1805.00930, 2018. URL `http://arxiv.org/abs/1805.00930`.

X. Du and A. Zare. Multiple instance choquet integral classifier fusion and regression for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5):2741–2753, May 2019. ISSN 0196-2892. doi: 10.1109/TGRS.2018.2876687.

X. Du, A. Zare, J. M. Keller, and D. T. Anderson. Multiple instance choquet integral for classifier fusion. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 1054–1061, July 2016. doi: 10.1109/CEC.2016.7743905.

H. Frigui and P. Gader. Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic k-nearest neighbor classifier. *IEEE Transactions on Fuzzy Systems*, 17(1), Feb 2009. ISSN 1063-6706. doi: 10.1109/TFUZZ.2008.2005249.

B. Fritzke. A growing neural gas network learns topologies. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, NIPS'94, pages 625–632, Cambridge, MA, USA, 1994. MIT Press. URL `http://dl.acm.org/citation.cfm?id=2998687.2998765`.

P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, and A. Zare. Multi-sensor and algorithm fusion with the choquet integral: applications to landmine detection. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages 1605–1608 vol.3, Sep. 2004. doi: 10.1109/IGARSS.2004.1370635.

M. Gao, A. Li, R. Yu, V. I. Morariu, and L. S. Davis. C-WSL: count-guided weakly supervised localization. *CoRR*, abs/1711.05282, 2017. URL `http://arxiv.org/abs/1711.05282`.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

A. N. Gorban and A. Y. Zinovyev. Elastic maps and nets for approximating principal manifolds and their application to microarray data visualization. In A. N. Gorban, B. Kégl, D. C. Wunsch, and A. Y. Zinovyev, editors, *Principal Manifolds for Data Visualization and Dimension Reduction*, pages 96–130, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-73750-6.

L. Guo. Latent variable algorithms for multimodal learning and sensor fusion. *CoRR*, abs/1904.10450, 2019. URL `http://arxiv.org/abs/1904.10450`.

J. K. Hackett and M. Shah. Multi-sensor fusion: a perspective. In *Proceedings., IEEE International Conference on Robotics and Automation*, pages 1324–1330 vol.2, May 1990. doi: 10.1109/ROBOT.1990.126184.

H. Hajimirsadeghi and G. Mori. Multi-instance classification by max-margin training of cardinality-based markov networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1839–1852, Sep. 2017. ISSN 0162-8828. doi: 10.1109/TPAMI.2016.2613865.

D. Hall, F. Dayoub, J. Skinner, P. Corke, G. Carneiro, and N. Sünderhauf. Probability-based detection quality (PDQ): A probabilistic approach to detection evaluation. *CoRR*, abs/1811.10800, 2018. URL `http://arxiv.org/abs/1811.10800`.

S. S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition, 2009.

M. Ilse, Jakub M. Tomczak, and M. Welling. Attention-based deep multiple instance learning. *CoRR*, abs/1802.04712, 2018. URL `http://arxiv.org/abs/1802.04712`.

C. Jiao. *Target Concept Learning From Ambiguously Labeled Data*. PhD thesis, Univ. of Missouri, Columbia, MO, Dec. 2017.

C. Jiao and A. Zare. Multiple instance hybrid estimator for learning target signatures. In *2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, pages 988–991, July 2017. doi: 10.1109/IGARSS.2017.8127120.

Changzhe Jiao, Chao Chen, Ronald G. McGarvey, Stephanie Bohlman, Licheng Jiao, and Alina Zare. Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:235 – 250, 2018. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2018.08.012.

P. Jindal and D. Kumar. A review on dimensionality reduction techniques. *International Journal of Computer Applications*, 173, 09 2017.

M. I. Jordan and R. A. Jacobs. Hierarchical mixtures of experts and the em algorithm. In *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, volume 2, pages 1339–1344 vol.2, Oct 1993.

A. Karem and H. Frigui. Multiple instance learning with multiple positive and negative target concepts. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 474–479, Dec 2016. doi: 10.1109/ICPR.2016.7899679.

T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, Sep. 1990. ISSN 0018-9219. doi: 10.1109/5.58325.

E. Kokiopoulou and P. Frossard. Graph-based classification of multiple observation sets. *Pattern Recognition*, 43(12):3988 – 3997, 2010. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2010.07.016.

E. Kokiopoulou and Y. Saad. Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12): 2143–2156, Dec 2007. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.1131.

N. Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *J. Mach. Learn. Res.*, 6:1783–1816, dec 2005. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1046920.1194904`.

N. D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, NIPS'03, pages 329–336, Cambridge, MA, USA, 2003. MIT Press. URL `http://dl.acm.org/citation.cfm?id=2981345.2981387`.

C. Lee, A. Elgammal, and M. Torki. Learning representations from multiple manifolds. *Pattern Recogn.*, 50(C):74–87, February 2016. ISSN 0031-3203. doi: 10.1016/j.patcog.2015.08.024. URL `http://dx.doi.org/10.1016/j.patcog.2015.08.024`.

B. Li, C. Yuan, W. Xiong, W. Hu, H. Peng, X. Ding, and S. Maybank. Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2554–2560, Dec 2017. ISSN 0162-8828. doi: 10.1109/TPAMI.2017.2669303.

D. Li, L. Zhu, W. Bao, F. Cheng, Y. Ren, and D. Huang. Cross-validated smooth multi-instance learning. pages 1321–1325, 05 2017. doi: 10.1109/IJCNN.2017.7966005.

H. Li, D. Liu, and D. Wang. Approximate policy iteration with unsupervised feature learning based on manifold regularization. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, July 2015.

L.Jian, X. Yang, Z. Liu, G. Jeon, M. Gao, and D. Chisholm. A symmetric encoder-decoder with residual block for infrared and visible image fusion. *ArXiv*, abs/1905.11447, 2019.

E. Lopez-Rubio and E. J. Palomo. Growing hierarchical probabilistic self-organizing graphs. *IEEE Transactions on Neural Networks*, 22(7):997–1008, July 2011. ISSN 1045-9227. doi: 10.1109/TNN.2011.2138159.

A. Mahalanobis and B. McIntosh. A comparison of target detection algorithms using dsiac atr algorithm development data set. Proc.SPIE, Apr. 2019.

O. Maron and T. Lozano-Pérez. A framework for multiple-instance learning. In *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*, NIPS '97, pages 570–576, Cambridge, MA, USA, 1998. MIT Press. ISBN 0-262-10076-2.

O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *Proceedings of the Fifteenth International Conference on Machine Learning*, ICML '98, pages 341–349, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc. ISBN 1-55860-556-8.

P. Márquez-Neila, L. Baumela, and L. Álvarez. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):2–17, Jan 2014. ISSN 0162-8828.

C. H. McCurley, J. Bocinsky, and A. Zare. Comparison of hand-held wemi target detection algorithms. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV, 110120U*, volume 11012 of *Proc.SPIE*, May 2019. doi: 10.1117/12.2519454.

M. Meng and X. Zhan. Zero-shot learning via low-rank-representation based manifold regularization. *IEEE Signal Processing Letters*, 25(9):1379–1383, Sep. 2018. ISSN 1070-9908. doi: 10.1109/LSP.2018.2857201.

M. Mohandes, M. Deriche, and S. O. Aliyu. Classifiers combination techniques: A comprehensive review. *IEEE Access*, 6:19626–19639, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2813079.

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

E. J. Palomo and E. Lopez-Rubio. Learning topologies with the growing neural forest. *International Journal of Neural Systems*, 26(04):1650019, 2016. doi: 10.1142/S0129065716500192. URL `https://doi.org/10.1142/S0129065716500192`. PMID: 27121995.

E. J. Palomo and E. Lopez-Rubio. The growing hierarchical neural gas self-organizing neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 28(9):2000–2009, Sep. 2017. ISSN 2162-237X. doi: 10.1109/TNNLS.2016.2570124.

F. Ratle, G. Camps-Valls, and J. Weston. Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2271–2282, May 2010. ISSN 0196-2892. doi: 10.1109/TGRS.2009.2037898.

A. Rauber, D. Merkl, and M. Dittenbach. The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. *IEEE Transactions on Neural Networks*, 13(6):1331–1341, Nov 2002. ISSN 1045-9227. doi: 10.1109/TNN.2002.804221.

B. Ren, B. Hou, J. Zhao, and L. Jiao. Unsupervised classification of polarimetirc sar image via improved manifold regularized low-rank representation with multiple features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2):580–595, Feb 2017. ISSN 1939-1404. doi: 10.1109/JSTARS.2016.2573380.

R. Rojas. Associative networks. In *Neural Networks - A Systematic Introduction*, chapter 12, pages 311–336. Springer-Verlag, Berlin, New-York, 1st edition, 1996.

S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 (5500):2323–2326, 2000. ISSN 0036-8075. doi: 10.1126/science.290.5500.2323. URL `https://science.sciencemag.org/content/290/5500/2323`.

A. Ruiz, O. Rudovic, X. Binefa, and M. Pantic. Multi-instance dynamic ordinal random fields for weakly supervised facial behavior analysis. *IEEE Transactions on Image Processing*, 27(8):3969–3982, Aug 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2830189.

D. E. Rumelhart and D. Zipser. Feature discovery by competitive learning. *Cognitive Science*, 9(1):75–112, 1985. doi: 10.1207/s15516709cog0901_5. URL `https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog0901_5`.

D. Ruta and B. Gabrys. An overview of classifier fusion methods. *Computing and Information Systems*, 7: 1–10, 01 2000.

John E. Ball Alina Zare Brendan Alvey Ryan E. Smith, Derek T. Anderson. Aggregation of choquet integrals in gpr and emi for handheld platform-based explosive hazard detection. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII, 1018217*, volume 10182, May 2017. doi: 10.1117/12.2263005. URL `https://doi.org/10.1117/12.2263005`.

L. K. Saul and S. T. Roweis. An introduction to locally linear embedding. *Journal of Machine Learning Research*, 7, 01 2001.

N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. V. Le, G. E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *CoRR*, abs/1701.06538, 2017. URL `http://arxiv.org/abs/1701.06538`.

H. Shen, X. Meng, and L. Zhang. An integrated framework for the spatiotemporalspectral fusion of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7135–7148, Dec 2016. ISSN 0196-2892. doi: 10.1109/TGRS.2016.2596290.

R. E. Smith, D. T. Anderson, A. Zare, J. E. Ball, B. Smock, J. R. Fairley, and S. E. Howington. Genetic programming based choquet integral for multi-source fusion. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, July 2017. doi: 10.1109/FUZZ-IEEE.2017.8015481.

Q. Sun, H. Liu, and T. Harada. Online growing neural gas for anomaly detection in changing surveillance scenes. *Pattern Recognition*, 64:187 – 201, 2017. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2016.09.016. URL `http://www.sciencedirect.com/science/article/pii/S0031320316302771`.

R. Talmon, S. Mallat, H. Zaveri, and R. R. Coifman. Manifold learning for latent variable inference in dynamical systems. *IEEE Transactions on Signal Processing*, 63(15):3843–3856, Aug 2015. ISSN 1053-587X. doi: 10.1109/TSP.2015.2432731.

I. L. Tanner and A. Mahalanobis. Fundamentals of target classification using deep learning. Proc.SPIE, Apr. 2019.

J. B. Tenenbaum, V. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. ISSN 0036-8075. doi: 10.1126/science.290.5500.2319. URL `https://science.sciencemag.org/content/290/5500/2319`.

S. Theodoridis and K. Koutroumbas. Kernel pca. In *Pattern Recognition, Fourth Edition*, chapter 6, pages 351–353. Academic Press, Inc., Orlando, FL, USA, 4th edition, 2008a. ISBN 1597492728, 9781597492720.

S. Theodoridis and K. Koutroumbas. The karhunen-loeve transform. In *Pattern Recognition, Fourth Edition*, chapter 6, pages 326–334. Academic Press, Inc., Orlando, FL, USA, 4th edition, 2008b. ISBN 1597492728, 9781597492720.

M. E. Tipping and C. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 21(3):611–622, January 1999. URL `https://www.microsoft.com/en-us/research/publication/probabilistic-principal-component-analysis/`.

M. Trabelsi and H. Frigui. Fuzzy and possibilistic clustering for multiple instance linear regression. pages 1–7, 07 2018. doi: 10.1109/FUZZ-IEEE.2018.8491540.

I. W. Tsang and J. T. Kwok. Large-scale sparsified manifold regularization. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1401–1408. MIT Press, 2007. URL `http://papers.nips.cc/paper/3005-large-scale-sparsified-manifold-regularization.pdf`.

S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann. *Review of Classifier Combination Methods*, pages 361–386. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-76280-5. doi: 10.1007/978-3-540-76280-5_14. URL `https://doi.org/10.1007/978-3-540-76280-5_14`.

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

G. Vivar, H. Burwinkel, A. Kazi, A Zwergal, N. Navab, and S. Ahmadi. Multi-modal graph fusion for inductive disease classification in incomplete datasets. *CoRR*, abs/1905.03053, 2019. URL `http://arxiv.org/abs/1905.03053`.

R. Wang, S. Shan, X. Chen, Q. Dai, and W. Gao. Manifoldmanifold distance and its application to face recognition with image sets. *IEEE Transactions on Image Processing*, 21(10):4466–4479, Oct 2012. ISSN 1057-7149. doi: 10.1109/TIP.2012.2206039.

R. Wang, X. Wang, S. Kwong, and C. Xu. Incorporating diversity and informativeness in multiple-instance active learning. *IEEE Transactions on Fuzzy Systems*, 25(6):1460–1475, Dec 2017. ISSN 1063-6706. doi: 10.1109/TFUZZ.2017.2717803.

Y. Xiao, B. Liu, and Z. Hao. Multiple-instance ordinal regression. *IEEE Transactions on Neural Networks and Learning Systems*, PP:1–16, 11 2017. doi: 10.1109/TNNLS.2017.2766164.

Y. Xiao, B. Liu, and Z. Hao. A sphere-description-based approach for multiple-instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):242–257, Feb 2017. ISSN 0162-8828. doi: 10.1109/TPAMI.2016.2539952.

S. E. Yuksel, J. N. Wilson, and P. D. Gader. Twenty years of mixture of experts. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8):1177–1193, Aug 2012. ISSN 2162-237X. doi: 10.1109/TNNLS.2012.2200299.

A. Zare and C. Jiao. Extended functions of multiple instances for target characterization. In *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, June 2014. doi: 10.1109/WHISPERS.2014.8077525.

A. Zare, M. Cook, B. Alvey, and D. K. Ho. Multiple instance dictionary learning for subsurface object detection using handheld emi. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX, 94540G*, Proc. SPIE, May 2015. doi: 10.1117/12.2179177. URL `https://doi.org/10.1117/12.2179177`.

A. Zare, C. Jiao, and T. C. Glenn. Multiple instance hyperspectral target characterization. *CoRR*, abs/1606.06354, 2016. URL `http://arxiv.org/abs/1606.06354`.

G. Zhang, P. Ghamisi, and X. Zhu. Fusion of heterogeneous earth observation data for the classification of local climate zones. *ArXiv*, abs/1905.12305, 2019.

J. Zhang. Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion*, 1:5–24, 03 2010. doi: 10.1080/19479830903561035.

Q. Zhang and S. A. Goldman. Em-dd: An improved multiple-instance learning technique. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 1073–1080. MIT Press, 2002.

B. Zitov and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977 – 1000, 2003. ISSN 0262-8856. doi: https://doi.org/10.1016/S0262-8856(03)00137-9.