Toward In-the-Field Canine Manifold Learning: Data Fusion for Evaluation of Potential Guide Dogs

Devon Martin
North Carolina State University
Raleigh, NC, USA
dmarti22@ncsu.edu

David Roberts
North Carolina State University
Raleigh, NC, USA
robertsd@csc.ncsu.edu

Alper Bozkurt
North Carolina State University
Raleigh, NC, USA
aybozkur@ncsu.edu

ABSTRACT

We seek to better classify canine behavior for guide dog training predictions. Dog temperament is a major factor in success rates and current training also has a blind spot when the puppies are with puppy raisers, who are lesser trained volunteers who socialize puppies up to 15 months old. We have used a custom designed smart collar to collect environmental and behavioral data from each puppy individually going through various parts of the guide dog training. We investigate long short-term memory networks (LSTMs), autoencoders (AE), and kernel principal component analysis (KPCA) as methods to identify canine behavior and use multi-sensor data fusion to find the best subset of sensors with the best at classifying temperament. Standard manifold learning experiments take place in controlled environments and translate poorly to real-world applications. This research aims to bridge this gap using guide dog In For Training (IFT) data, which is from a lesser controlled environment and use it to develop a broader data-pattern-to-behavior dictionary for future real-world canine studies.

CCS CONCEPTS

• **Computer systems organization** → Sensor networks; • **Computing methodologies** → **Learning latent representations**; *Kernel methods.*

KEYWORDS

manifold learning, LSTM, autoencoder, KPCA, canine monitoring

ACM Reference Format:
Devon Martin, David Roberts, and Alper Bozkurt. 2023. Toward In-the-Field Canine Manifold Learning: Data Fusion for Evaluation of Potential Guide Dogs. In *The Tenth International Conference on Animal-Computer Interaction (ACI '23), December 04–08, 2023, Raleigh, NC, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3637882.3637898

1 MOTIVATION

As of 2016, there were over 7.5 million non-institutionalized visually disabled people in the United States alone [1]. Guide dogs are used to help these people in everyday life, but not every dog is suitable for this mission critical role.

The cost to train a single guide dog for a visually impaired person is tens of thousands of dollars [6, 10]. This is exacerbated by the high dropout rate of guide dogs during training. Less than half of dogs in training will successfully become working guide dogs [11]. Even after graduation, there is a smaller but still impactful dropout rate until the dogs are retired after approximately 8.5 years [5]. Previous attempts to reduce dropout rate have been shown to be successful. For instance, approximately 30% of guide dogs were disqualified due to poor hip quality, but this was reduced to near-zero thanks to successful breeding programs by the late 1990s [21]. Secondly, The Seeing Eye passes back dogs that originally failed to undergo additional training [14]. Of the 40% of dogs that were passed back, 53% later graduated, upping the total graduation rate.

Nowadays, the biggest impediment to dog graduation is in the potential guide dog's temperament. Temperament is defined as behavioral tendencies that vary between individuals but are stable within a given individual. According to The Seeing Eye, behavioral reasons accounted for 21% of dogs being passed back, and the main behavioral reason to reject a guide dog is fearfulness, followed by easily distractedness and aggressiveness [14]. Rejection for behavioral reasons is also the worst reason for rejection, as these dogs are least likely to pass training upon further testing. Fear is a significant and reliable predictor of training failure in guide dogs [10]. Early traumatic events for puppies, like being attacked by other dogs or strangers, appear to have long-lasting effects that result in disqualification.

Dollion et al [10] investigated a serious blind spot in the dog raising process: volunteer raisers with variable environments. While breeds, genetics, and biological aspects can be well controlled with breeding programs and systematic medical evaluations, early developmental environments are not. Various behavioral differences were observed when dogs were returned, including owner/stranger/dog-related aggression or fear, non-social fear, and touch sensitivity. Some of the causes of these behavioral effects include owner experience, presence of teenagers, and whether other dogs were present in the household. Again, a miniature monitoring device would be useful for tracking these early events, especially since this development period falls outside facility auspices.

Previous wearable devices have been used in several applications of dog behavior monitoring [16, 17, 22–24], including sleep monitoring, temperament evaluations, and heart and breath rate monitoring. We presented a smart collar [8, 12, 19] as one of such a sensor system that collects various data continuously while being small and easy to implement. It records subject-of-interest information like accelerometry for movement and audio for barking as well

ACI '23, December 04–08, 2023, Raleigh, NC, USA

ACM ISBN 979-8-4007-1656-0/23/12
https://doi.org/10.1145/3637882.3637898

as immediate surrounding environmental information like temperature, humidity, and light levels. These devices are applicable for daily use and can help in both general behavioral monitoring as well as help bridge the uncertainty gap of volunteer raisers' various environments. In this paper, we focus on applying data collected from these devices during the In For Training (IFT) Evaluation protocol used by Guiding Eyes for the Blind for determining dog fitness before entering their official training program.

The IFT evaluation is conducted on 15-18 month old dogs following volunteer puppy raising. During these tests, temperament and adaptability to strangers, dogs, and new environments is evaluated. Only self-confident and composed dogs are deemed suitable for further training [2]. Over the last year (2022 and onward), IFT tests have been conducted with the smart collar [19]. The accompanying custom iOS app additionally allowed trainers to record the timing of controlled stimuli used as test events. This creates a time-accurate label that can be associated with the collar's data collection. The exploration of dog behavior captured in the data can help determine proper or improper reactions, adding objectivity to the test evaluations which helps qualify or disqualify a given dog earlier in the training process.

Because the smart collar system collects data from various sensors, we can employ data fusion to attain higher accuracy than using single sensors. There are many possible data fusion architectures, and optimal performance tends to be problem-specific [3, 7, 9, 15, 20, 25]. Therefore, one major goal of this paper is to find an optimal architecture to monitor canine behavior. Additionally, because data is collected as a time series, we expect temporal dependencies. While many time series models are possible such as ARIMA, they lack any fusion abilities. We instead use long short-term memory networks (LSTMs), which has shown strong predictability with sequential data [18].

Another goal of this study is to make future dog behavioral evaluations easier. We hypothesize that the categories of the Guiding Eyes for the Blind's IFT evaluation will not be truly representative of generic dog behavior. For instance, a dog may jump at any time during the evaluation; while this behavior is not categorized in the IFT, generic jumping behavior is a quantifiable and easily recognizable event that could be identified in other studies. For this reason, we want to include an unsupervised learning approach to categorize these dog-specific tendencies. Here we will be performing manifold learning, which is the categorization of common patterns of the sensor data [26]. Other approaches that could be used with these time series data include LSTMs and wavelet or matching pursuit-based classifications [28].

The specific goals we have include:

- Better canine behavior understanding
- Accounting for individual effects like personality on behavior
- Apply multi-sensor data fusion to canine understanding
- Begin setting a lexicon for in-field behavior interpretation

2 METHODOLOGY

We describe our approach to supervised and unsupervised manifold learning using LSTMs, AEs, and KPCA.

2.1 Manifold Learning

Manifold learning seeks to find a lower dimensional space from data [13, 26]. Even though a given datapoint when we account for all sensors has dimension at most eight, since we are incorporating sequences for pattern recognition, with n being sequence length, we actually have a much higher dimensionality of 8^n. This makes manifold learning more appropriate in this context.

2.2 Dataset Description

IFT sessions are conducted roughly monthly at GEB headquarters in New York. Typically, between 15 and 30 dogs are tested during each session, and each session lasts about ten minutes. During the session, a dog is presented with various stimuli that include a vet examination and a vacuum cleaner. The timing of these events are recorded by an on-site trainer in an iOS app custom developed for this task [19].

The smart collar system collects canine-specific data, acceleration and audio, and environmental conditions, consisting of ambient light, temperature, humidity, and pressure. Because of the various sensations occur over different time frames, the sensors have different sampling rates. In general, the IMU, audio, and light sensors are sampled at a much higher rate in the range of kilohertz than temperature, humidity, and pressure, which are on the order of a few hertz.

For preprocessing, we interpolated the data to account for differences in sensor sampling frequencies. Afterwards, we Z-normalized all the data.

2.3 Supervised Approches

For supervised learning, we use LSTMs, AEs, and KPCA.

2.3.1 *LSTMs.*

LSTMs have shown tremendous performance in sequential data for their ability to track long-term data patterns. For this reason, we investigate the use of LSTMs for temperament classification. Previous hyperparameter searches for LSTMs [18] showed that the most important features of an LSTM are the learning rates followed by the hidden layer size. Other attributes account for a vast minority of performance in the general case. So we focus our attention on learning rate and hidden layer size when tuning our LSTM.

We are interested in understanding the relationship between different data subsets and preprocessing types; secondarily, we seek to optimize the hyperparameters for higher accuracy. With a total of three data types (IMU, audio, environmental sensors), there are eight possibilities. There are also two more states for bidirectionality. With four learning rates, three amounts of node sizes, and two numbers of layers, there are a total of 24 hyperparameter states. This produces 384 possible structures. To reduce the search size, we instead consider a sequential search strategy. For the search space search, we find an architecture, then the learning rate is optimized, and lastly, the relevant data fusion setups (only five) are optimized. This produces a simplified overall search schematic of 21 states.

2.3.2 *Autoencoders.*

The quintessential unsupervised learning method that has attained excellent results in terms of dimensional reduction is the autoencoder (AE). These neural net architectures have a set of encoding layers that get subsequently smaller until a small feature space, or latent representation, of only few traits is reached. Then, a set of decoding layers that get subsequently larger carries the latent space information back to the original size of the dataset. Loss is determined by the difference between the original data and the reconstructed data. In this way, only the most important characteristics of the data are learned.

However, vanilla AEs are not designed to handle sequential data. For this reason, we consider LSTM-AEs, where encoding and decoding layers are constructed out of layers of LSTM units. Since this setup uses LSTMs which were used with supervised learning with the IFT classifications, there is a potential bridge between the supervised and unsupervised algorithms.

2.3.3 Kernel PCA. Principle component analysis (PCA) is a commonly used method in multivariate analysis for identifying the most important linear combinations that describe a dataset. Kernel PCA expands upon this using the kernel trick, a method that maps all data into a higher-dimensional space where it is easier to cluster the data, while retaining computability with the use of a kernel matrix [26, 27]. In this case, we used a radial basis function as the kernel followed by a linear classifier, in this case Ridge Regression, to group the data. Unlike LSTMs or AEs, KPCA does not have a difficult setup process requiring extensive training; it is a more static method, with only one parameter, γ. This is a major advantage.

2.4 Unsupervised Methods

Previous methods have been designed with a known label from the IFT evaluations, but we would like to have these methods applicable for in-the-field testing in less controlled environments. Traditional pattern learning takes place in well-controlled environments with well-practiced actions, limiting real-world effects like personality differences, mistakes and corrections, and large signal noise. Despite this, certain prominent actions should be distinguishable.

We will be attempting several multivariate unsupervised learning methods to label certain consistent and repetitive behaviors observable within the data from the smart collars. Instead, we are interested in identifying patterns within sequential time segments of the data. For this purpose, we intend on modifying the LSTM, AE, and KPCA methods.

2.5 Unsupervised Interpretation and the Pre-Image Problem

The unsupervised learning methods are expected to discover a smaller, simpler dimensional variable space while maintaining discrimination between the classes. We are interested in input patterns because knowing these will add a great deal of interpretability to our findings. For interpretability of the latent space clusters, we are interested in the inverse problem, known as the pre-image problem [4, 29].

Let $\phi(x)$ be the mapping provided by an LSTM-AE, Kernel PCA, or other function estimation methods with $\chi \in \mathbb{R}^M$, and latent variable, $\Phi \in \mathbb{R}^N$ with $M >> N$. With this setup, our objective is to first select a $\phi^* \in \Phi$ that represents a high density latent region, and second, find the x whose mapping $\phi(x)$ gets as close as possible to ϕ^*. This comes to solving the inverse optimization problem:

$$x^* = \underset{x \in \chi}{\operatorname{argmin}} ||\phi(x) - \phi^*||^2.$$

3 EARLY RESULTS

We conducted a rudimentary experiment to ascertain the manifold learning process and present early LSTM and KPCA results.

3.1 Manifold Learning Experiments

We performed a simple preliminary test to look for real-world manifolds using the smart collar. A user carried the device while performing exercises, including walking and jogging around a short route. Figure 1 shows two distinctly different manifolds collected from the IMU data space for both walking and jogging cycles. Lines are displayed to highlight the sequences of the datapoints, indicating rising patterns. Noticeably, the walking manifold is much smaller than the jogging manifold. These manifolds are dependent on collar orientation too, which adds to complexity.

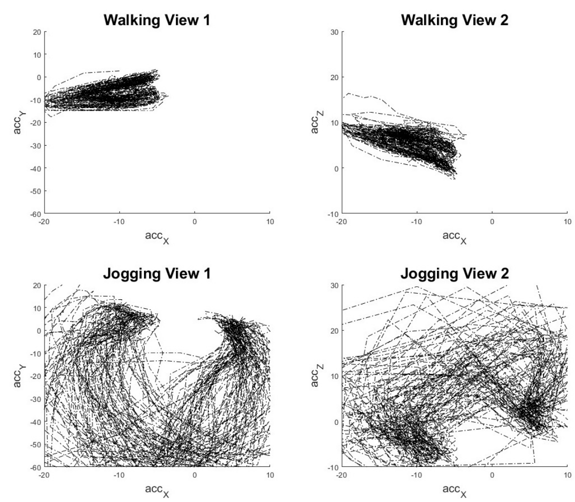

Figure 1: Manifold Against Action. Top figures are walking cycles and bottom figures are jogging cycles. Left figures show Y-axis against X-axis and right is an alternative view.

Clearly, different gaits are distinguishable in terms of manifolds in patterns using the smart collar device, at least in strongly controlled settings. However, the data from the IFT sessions is far less controlled and expected to be noisier. This is the challenge of creating a manifold dictionary.

Once we have a manifold dictionary set, we can examine new patterns of data and determine the distance from these manifolds. We project new data patterns onto the various manifolds and the one with the lowest reconstruction error is selected as the new behavior. We attempt the methodology detailed in [26] to create our own dictionary and evaluate the results using a similarity metric.

3.2 LSTM Setup and Training Results

Several LSTM models with various hyperparameters have been created and tested. Starting with a 50-state classification defined from the 50 IFT-defined state, overall, these models have performed at about 11% accuracy on the validation set after only a few epochs. This is vastly better than the 2% attained using a random model, but is still far from desirable. Currently, these models are built using only the acceleration score. A higher capacity model with four 256-node hidden layers followed by two dense layers shows similar performance.

By looking at only a smaller 12-state model focusing on the 12 most separable categorical classes, we can attain a more reasonable 35% validation accuracy (random would be 8%). The training curve (Figure 2) shows overfitting after only a few epochs with a flatlining validation curve, so there is room for LSTM design improvement. The confusion matrix (Figure 3) shows a desired diagonal.

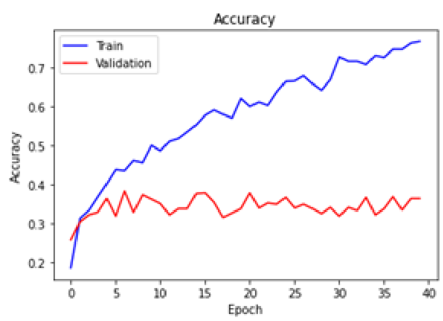

Figure 2: LSTM 12-State Training Curve

Figure 3: LSTM 12-State Confusion Matrix

3.3 KPCA Setup and Results

Using a 70/30 train test split, we use the Kernel PCA method. After optimizing the univariate gamma parameter to 0.004, this method attains a 42% accuracy on the 12-state system. The confusion matrix (Figure 4) has a desirable diagonal. The higher accuracy is surprising because this is a well-defined method that does not require training or much hyperparameter tuning like the LSTM. Being easier to setup and tune, and having higher accuracy, this method is clearly superior at the moment. We show the top two KPCA components in Figure 5, where we can see that the different states (represented by color) appear to be differentiated radially in elliptical curves.

Comparing KPCA to traditional PCA, we found that traditional PCA provided 20% accuracy on the 12-state system, suggesting that using kernels was indeed an important component.

Figure 4: KPCA Confusion Matrix

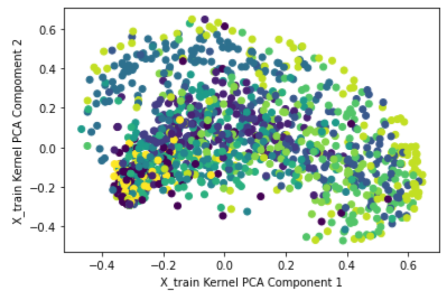

Figure 5: KPCA Components

4 FUTURE STEPS

We proposed three methods for supervised and unsupervised manifold learning. We have reached a good starting place for the LSTM and KPCA methods, but we still need to design the AE. Further hyperparameter tuning will be needed for the AE and LSTM. We suspect that convolutional layers will produce better results for the LSTM. And lastly, we will be adding the data from the other sensors to select the most useful subset for behavioral prediction.

5 CONCLUSION

I have presented several methodologies with which to evaluate a multi-sensor data fusion analysis approach for canine behavior determination. LSTMs and AEs are promising for manifold learning, and KPCA has shown surprisingly good accuracy despite its simplicity. We hope 1) to attain better accuracy with these other sensors and 2) that one methodology will prove superior to the others. As we apply these methods toward unsupervised approaches, our methods will help with in-the-field data interpretation.

ACKNOWLEDGMENTS

A special thanks to Guiding Eyes for the Blind staff for collecting the IFT data, in particular: Jane Russenberger, Madeline Zimmermann, Tyler Jacobs, Gerald Brenninkmeyer and Helen West. We thank IBM for their support and especially Lorraine Trapani. We also thank financial support from NSF through multiple grants (CCSS-1554367, IIS-2037328 and ECC-1160483) and IBM Faculty Awards.

REFERENCES

[1] 2016. National Federation of the Blind. (2016). https://www.nfb.org/resources/blindness-statistics

[2] 2023. (2023). https://www.guidingeyes.org/journey-of-a-guide-dog/in-for-training/

[3] R. Antony. 2009. Data Management Support to Tactical Data Fusion. In *Handbook of Multisensor Data Fusion Theory and Practice*, M. Liggins, D. Hall, and J. Llinas (Eds.). CRC Press, 619–653.

[4] O. Arif, P. Vela, and W. Daley. 2010. Pre-image Problem in Manifold Learning and Dimensional Reduction Methods. *Ninth International Conference on Machine Learning and Applications* (2010).

[5] L. Asher, N. Harvey, M. Green, and G. England. 2017. Application of Survival Analysis and Multistate Modeling to Understand Animal Behavior: Examples from Guide Dogs. *frontiers in Veterinary Science* (2017).

[6] G. Berns, A. Brooks, M. Spivak, and K. Levy. 2017. Functional MRI in Awake Dogs Predicts Suitability for Assistance Work. *Sci Rep* 7, 43704 (2017).

[7] Eva Borràs, Joan Ferré, Ricard Boqué, Montserrat Mestres, Laura Aceña, and Olga Busto. 2015. Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Analytica Chimica Acta* 891 (2015), 1–14. https://doi.org/10.1016/j.aca.2015.04.042

[8] Z. Cleghern, E. Williams, S. Mealin, M. Foster, T. Holder, A. Bozkurt, and D. Roberts. 2019. An IoT and Analytics Platform for Characterizing Adolescent Dogs' Suitability for Guide Work. *Proceedings of the Sixth International Conference on Animal-Computer Interaction* 1 (2019), 1–6.

[9] B.V. Dasarathy. 1997. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE 85.1* (1997), 24–38. https://doi.org/10.1109/5.554206

[10] N. Dollion, A. Paulus, N. Champagne, N. St-Pierre, E. St-Pierre, M. Trudel, and P. Plusquellec. 2019. Fear/Reactivity in working dogs: An analysis of 37 years of behavioral data from the Mira Foundation's future service dogs. *Applied Animal Behavior Science* 221 (2019).

[11] D. Duffy and J. Serpell. 2012. Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science* 138 (2012), 99–109.

[12] M. Foster T. Holder D. Roberts A. Bozkurt E. Williams, Z. Cleghern. 2020. A Smart Collar for Assessment of Activity Levels and Environmental Conditions for Guide Dogs. *IEEE 42nd Annual International Conference EMBC* (2020), 4628–4631. https://doi.org/10.1109/EMBC44109.2020.9175814

[13] A. Elgammal and C. Lee. 2011. Human Motion Analysis Applications of Manifold Learning. In *Manifold Learning Theory and Applications*. Taylor & Francis Group, Chapter 11, 277–321.

[14] I. Ennik, A. Liinamo, E. Leighton, and J. Arendonk. 2006. Suitability for field service in 4 breeds of guide dogs. *Journal of Veterinary Behavior* 1 (2006), 67–74.

[15] J. Esteban, A. Starr, R. Willetts, P. Hannah, and P. Bryanston-Cross. 2005. A Review of data fusion models and architectures: towards engineering guidelines. *Neural Comput & Applic* 14 (2005), 273–281.

[16] M. Foster, J. Wang, E. Williams, D. Roberts, and A. Bozkurt. 2020. Inertial Measurement Based Heart and Respiration Rate Estimation of Dogs During Sleep for Welfare Monitoring. *Proceedings of the Seventh International Conference on Animal-Computer Interaction* (2020), 1–6.

[17] M. Foster, J. Wang, E. Williams, D. Roberts, and A. Bozkurt. 2021. ECG and Respiration Signal Reconstruction from an IMU at Various Orientations during Rest or Sleep for Dog Welfare Monitoring. *Proceedings of the Eight International Conference on Animal-Computer Interaction* 7 (2021), 1–9.

[18] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. 2017. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems* 28, 10 (2017), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

[19] T. Holder, E. Williams, D. Martin, A. Kligerman, E. Summers, Z. Cleghern, J. Dieffenderfer, J. Russenberger, D. Roberts, and A. Bozkurt. 2021. From Ideation to Deployment: A Narrative Case Study of Citizen Science Supported Wearables for Raising Guide Dogs. *Proceedings of the Eight International Conference on Animal-Computer Interaction* 6 (2021), 1–13.

[20] J. Mendes Jr., M. Vieira, M. Pires, and S. Stevan Jr. 2016. Sensor Fusion and Smart Sensor in Sports and Biomedical Applications. *Sensors (Basel)* (2016). https://doi.org/10.3390/s16101569

[21] E. Leighton. 1997. Genetics of canine hip dysplasia. *Journal of the American Veterinary Medical Association* 210, 10 (1997), 1474.

[22] S. Mealin, Z. Cleghern, M. Foster, A. Bozkurt, and D. Roberts. 2019. Predicting Guide Dog Temperament Evaluation Outcomes Using Raw ECG Signals. *Proceedings of the Sixth International Conference on Animal-Computer Interaction* (2019), 1–12.

[23] S. Mealin, M. Foster, Z. Cleghern, A. Bozkurt, and D. Roberts. 2020. Using Inertial Measurement Unit Data for Objective Evaluations of Potential Guide Dogs. *Proceedings of the Seventh International Conference on Animal-Computer Interaction* 2 (2020), 1–11.

[24] S. Mealin, M. Foster, K. Walker, S. Yushak, B. Sherman, A. Bozkurt, and D. Roberts. 2017. Creating an Evaluation System for Future Guide Dogs: A Case Study of Designing for Both Human and Canine Needs. *Proceedings of the Fourth International Conference on Animal-Computer Interaction* (2017), 1–6.

[25] H. Mitchell. 2007. Sensors. In *Multi-Sensor Data Fusion An Introduction*. Springer.

[26] C. O'Reilly, K. Moessner, and M. Nati. 2017. Univariate and Multivariate Time Series Manifold Learning. *Knowledge-Based Systems* 133 (2017), 1–16.

[27] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. 1998. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation* 10, 5 (07 1998), 1299–1319. https://doi.org/10.1162/089976698300017467 arXiv:https://direct.mit.edu/neco/article-pdf/10/5/1299/813905/089976698300017467.pdf

[28] Y. Wang. 2007. Seismic time-frequency spectral decomposition by matching pursuit. *Geophysics* 72, 1 (2007).

[29] J. Weston, B. Scholkopf, and G. Bakir. 2003. Learning to Find Pre-Images. *NIPS* (2003).