2014 22nd International Conference on Pattern Recognition

Automatic Image Attribute Selection for Zero-shot Learning of Object Categories

Liangchen Liu*, Arnold Wiliem*, Shaokang Chen* and Brian C. Lovell*
*School of ITEE ,The University of Queensland, Australia
{l.liu9,a.wiliem}@uq.edu.au shaokangchenuq@gmail.com lovell@itee.uq.edu.au

***Abstract*— Recently the use of image attributes as image descriptors has drawn great attention. This is because the resulting descriptors extracted using these attributes are human understandable as well as machine readable. Although the image attributes are generally semantically meaningful, they may not be discriminative. As such, prior works often consider a discriminative learning approach that could discover discriminative attributes. Nevertheless, the resulting learned attributes could lose their semantic meaning. To that end, in the present work, we study two properties of attributes: discriminative power and reliability. We then propose a novel greedy algorithm called Discriminative and Reliable Attribute Learning (DRAL) which selects a subset of attributes which maximises an objective function incorporating the two properties. We compare our proposed system to the recent state-of-the-art approach, called Direct Attribute Prediction (DAP) for the zero-shot learning task on the Animal with Attributes (AwA) dataset. The results show that our proposed approach can achieve similar performance to this state-of-the-art approach while using a significantly smaller number of attributes.**

I. Introduction

Feature extraction is one of the prominent tasks in the image classification system pipeline. It serves as a transformation function mapping the images from their original high dimensional space to another space where the classification problem could be easier to solve. There are many works aimed to develop such a good transformation function [22]. For instance, the Scale Invariant Feature Transform (SIFT) [14] aims to extract features possessing invariant properties such as: location, scale, rotation and affine transformations. Another notable example is the Histogram Oriented Gradient (HOG) [7] which is basically a feature descriptor counting occurrences of specific gradient orientations in localized portions of an image. Despite their excellent performance reported for various vision applications [22], [7], these feature descriptors are very difficult to be interpreted by humans. Although each of their elements may have a relationship such as the gradient magnitude, they do not have a direct relationship to the high-level semantic concepts related to the problem domain [11], [13].

Image attributes can be described as inherent properties/characteristics of an image. For instance, a car image could have the following attributes: *is blue*, *has wheels*, *is metallic*. In this case, one could represent an image with a set of image attributes present in an image. Technically, each element in the descriptor defines the existence/absence of a specific image attribute. It can be detected by attribute detectors tested on the low-level features mentioned above. Attribute detector is basically a binary classifier trained beforehand. As such, one needs to construct different training sets for each attribute detector which could be expensive. To that end, one could use a crowd sourcing approach which could minimise the cost by using the Amazon Mechanical Turk (AMT) [1][17], [19]. Here, we can ask people on the internet to describe the images by words. Generally, the set of attributes found from this process is not necessarily discriminative for the vision task. This is due to the fact that it is difficult for humans to manually identify a set of discriminative attributes for a classification task which has a large number of categories.

The image attribute representation is successfully applied in various vision tasks such as face verification [9], complex event detection [15], human action recognition [26], visual knowledge extraction [6], and zero-shot learning [10]. Moreover, Parikh *et al.* proposed the notion of relative attributes such as *larger* or *more open space* which could be understood as an adjective comparing two images [18].

In this work, we focus on the attribute-based zero shot learning problem. Zero-shot learning [10] is the problem of object recognition when the testing categories do not have any training examples. However, humans can easily define the attribute representation for each test category without any training image due to the fact that each element of the attribute descriptor has semantic meaning. For instance, Lampert *et al.* showed that the image attribute descriptor could be used to address the *zero shot* learning [10]. They proposed two general frameworks of attributes-based zero-shot learning, Direct Attributes Prediction (DAP) and Indirect attributes prediction (IAP). However, in their work attribute discriminative power has not yet been considered.

Several works proposed approaches to automatically discover discriminative attributes [8], [20], [27]. These approaches are very similar to some feature selection works [5], [4]. Here, the attribute detectors are jointly learned with the image classifier in the max-margin framework. For instance, Farhadi *et al.* proposed to use random comparisons and within category prediction to learn the discriminative attributes respectively [8]. Nonetheless, the resulting set of attributes are not guaranteed to have semantic meaning; defeating one of the prime purposes of using attribute descriptors. Yu. *et al.* designed a category-level discriminative attribute learning algorithm according to category-separability and learnability [27] however, their method also cannot be used to describe images with concise semantic meaning due to the design of category-attribute matrix. There is also significant human effort required to build a new category-attribute matrix.

Contributions The aim of the present work is to discover the set of semantic attributes which are also discriminative and reliable for the given classification task. To that end, we propose a discriminative selection algorithm which takes as input the image attributes discovered from the manual process via the AMT. There are two main advantages of using

[1] www.mturk.com

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.452

the proposed approach: (1) the feature dimensionality can be significantly reduced which simplifies the classification process and (2) the selected attributes can potentially improve the system performance due to the fact that the selection is based on the attribute discrimination power. The algorithm selects the subset of semantically meaningful attributes maximising two attribute properties: attribute discriminative power and attribute reliability. Attribute discriminative power is related to the property of the attribute descriptor to separate images of different categories, whilst, we relate the attribute reliability to the error produced during the attribute descriptor extraction process. We apply our method to the zero-shot learning problem investigated in [10]. We show that by applying our algorithm we can decrease the dimensionality of the attribute descriptor by 35% achieving better performance than the state-of-the-art approach proposed in [10].

We continue our paper as follows. Section II presents the proposed attribute properties. Then we describe the proposed algorithm in Section III. The experiment and results are discussed in Section IV. Finally the main findings and future direction are presented in Section V.

II. Property of attributes

Each element of an image attribute descriptor defines the existence/absence of an image property [10]. Generally, in an image classification task, each image is represented by the same set of image attributes. Let $\boldsymbol{z_i} \in \{0,1\}^B$ be the B dimensional attribute descriptor of image $\boldsymbol{I_i}$; the function $\Phi_b : \mathbb{R}^d \mapsto \{0,1\}$ be the b-th attribute detector. Each element in $\boldsymbol{z_i}$ is determined as:

$$z_{i,b} = \Phi_b(\boldsymbol{x_i}) \tag{1}$$

where $z_{i,b}$ is the b-th element of $\boldsymbol{z_i}$ and $\boldsymbol{x_i}$ is the set of features extracted from image $\boldsymbol{I_i}$.

In order to be successful in a classification task, one needs to ensure that the attribute descriptor sufficiently separates images from different categories. Nevertheless, as demonstrated by Farhadi *et al.* in [8], although a set of image attributes can effectively describe objects from different categories, it may not always be sufficient for distinguishing between different categories. This is due to the fact that most image attributes were generated by asking human to describe images. For instance, it is reasonable to describe a cat as a four legged animal. However this attribute is not useful to distinguish between cats and dogs as they are both four legged animals. In addition, it is almost impossible to manually identify the subset of discriminative attributes from a large pool of attributes for solving an image classification task with a large number of categories. Therefore, it is important to have an automatic system which is able to identify a subset of discriminative attributes for each application domain.

Another important aspect that should be considered to develop such a system is the fact that the attribute descriptor extraction process is not error free. This is because the attribute detectors $\{\Phi_b\}_{b=1}^B$ are essentially binary classifiers trained to minimise the classification generalisation error. It is preferable to have reliable attribute detectors which in turn could minimise the overall descriptor extraction error.

In the light of the above facts, we propose that there are intrinsically two aspects contributing to the performance of a classification system utilising image attribute descriptors: (1) attribute discriminative power and (2) attribute reliability. The former determines the separability between image categories and the latter determines the reliability of each attribute detector and also the semantic drift of the attribute classifier. Discriminative power has been explored in [27] to discover discriminative category-level attributes. Nevertheless, the discovered attributes resulting from this approach do not necessarily have semantic meaning.

A. Attribute discriminative power

Attribute discriminative power governs how well a set of image attributes separate images from different categories. The attribute discriminative power, Δ can be defined as:

$$\Delta = \sum_i \sum_j \|\boldsymbol{z_i} - \boldsymbol{z_j}\|_H \quad \boldsymbol{z_i} \in c, \boldsymbol{z_j} \notin c \tag{2}$$

where $\boldsymbol{z_i}$ and $\boldsymbol{z_j}$ are the attribute descriptors of the i-th and j-th images which belong to different categories, respectively; $\|\cdot\|_H$ is the hamming distance. The above equation can be easily extended to the zero-shot learning where only category-level attributes are available:

$$\Delta = \sum_i \sum_j \|\boldsymbol{h_i} - \boldsymbol{h_j}\|_H \quad i \neq j \tag{3}$$

where $\boldsymbol{h_i}, \boldsymbol{h_j} \in \{0,1\}^B$ are the category-level attribute descriptor. Intuitively, when the attribute discriminative power Δ is maximised, the margin between pair-wise categories will be maximised in the attribute feature space. This will lead to high category separability.

B. Attribute reliability

We define the attribute reliability, Ω which measures the reliability of a set of attribute detectors as:

$$\Omega = \sum_{b=1}^{B} \omega_b \tag{4}$$

where ω_b is the reliability score of the b-th attribute detector. The individual reliability score ω_b is related to the generalisation error of the attribute detector Φ_b. Indeed it is difficult to determine the generalisation error of a classifier [3], [24]. One possible alternative is to define $\hat{\omega_b}$ which is the approximation of ω_b. Thus, the approximated attribute reliability, $\hat{\Omega}$, is defined as

$$\hat{\Omega} = \sum_{b=1}^{B} \hat{\omega_b} \tag{5}$$

.

In the present work we determine $\hat{\omega_b}$ by first constructing the Receiver Operating Characteristic (ROC) curve of the attribute detector Φ_b and computing the Area Under the Curve (AUC). We further perform non-linear normalisation using a sigmoid function in order to increase the contrast between the reliable and non-reliable attributes. Therefore, we define $\hat{\omega_b}$ as:

$$\hat{\omega_b} = \frac{1}{1 + e^{-\beta(\mathrm{AUC_b} - \gamma)}} \tag{6}$$

where AUC_b is the AUC of the attribute detector Φ_b; β, γ are the normalisation parameters. We determine both the AUC and the normalisation parameters from a cross-validation set. It is noteworthy to mention the attribute reliability relies on two factors: (1) the generalisation error of the attribute detector and (2) the semantic drift caused from the noise in the attribute detector training process. The semantic drift happens when an attribute detector accidentally learns a concept different from the initial intention [8]. For instance, when we use car and non-car images as positive and negative samples in order to learn *has wheel* attribute, the corresponding attribute detector may learn *is metallic* concept as the most discriminative feature to differentiate cars with non-car images. Our proposed approximation of the attribute reliability $\hat{\Omega}$ captures the former factor. Nevertheless, it is still difficult to measure the degree of the attribute detector semantic drift.

III. Discriminative and Reliable Attribute Learning

A. Prior Work

In this part, we will briefly introduce the Discriminative Attribute Prediction (DAP) method proposed in [10]. The DAP uses the Bayes rule to model the relationships between attribute descriptor $\boldsymbol{z_i}$ and low level feature representation $\boldsymbol{x_i}$ of an image as well as $\boldsymbol{z_i}$ and the unseen test category label v. The attribute descriptor $\boldsymbol{z_i^y}$ for a seen training category y can be represented as a vector $\left[z_{i,1}^y, ..., z_{i,B}^y\right]^T$, the Bayes posteriori probability for a test category v given an input $\boldsymbol{x_i}$ can be defined as Eqn. 7 :

$$P\left(v|\boldsymbol{x_i}\right) = P\left(v\right)\prod_{b=1}^{B}\frac{P\left(z_{i,b}^v|\boldsymbol{x_i}\right)}{P\left(z_{i,b}^v\right)} \tag{7}$$

where $P(v)$ is the prior of the test category v, $P(z_{i,b}^v)$ denotes the attribute prior, $P(z_{i,b}^v|\boldsymbol{x_i})$ is the image-attribute probability output of the attribute detector ϕ_b. The authors assume identical test category prior and then ignore $P(v)$ effectively. They also use empirical means $P(z_{i,b}^v) = \frac{1}{K}\sum_{k=1}^{K} I(z_{i,b}^k = z_{i,b}^v)$ for all the training categories, where $I(\cdot)$ is the indicator function that gives value one when the condition is met, zero otherwise; and K is the number of training categories. Finally, the best output category from all test categories $v_1, ..., v_q$ is assigned to a test sample $\boldsymbol{x_i}$ according to the (maximum a posteriori probability) MAP prediction as Eqn. 8:

$$f\left(\boldsymbol{x_i}\right) = \underset{q=1,...,Q}{\arg\max}\ P\left(v|x_i\right) = \underset{q=1,...,Q}{\arg\max}\prod_{b=1}^{B}\frac{P\left(z_{i,b}^{v_p}|\boldsymbol{x_i}\right)}{P\left(z_{i,b}^{v_p}\right)} \tag{8}$$

B. Discriminative and Reliable Attribute Selection

Given a pool of image attributes U, the goal of the present work is to mine the set of attributes which have high discriminative power as well as reliability. To that end, we define our objective function $\mathrm{J}(\cdot)$ as:

$$\mathrm{J}\left(\{U, \{\boldsymbol{h_i}\}_{i=1}^{C}, \{\Phi_b\}_{b=1}^{B}\}\right) = \underset{S\in U}{argmax}\left(\alpha\hat{\Omega}_S + (1-\alpha)\Delta_S\right) \tag{9}$$

where $U = \{1 \cdots B\}$ is the set of all image attributes; $S \in U$ is the selected subset of image attributes; Δ_S and $\hat{\Omega}_S$ are the selected attribute discriminative power and attribute reliability, respectively; $\{\Phi_b\}_{b=1}^{B}$ is the set of attribute detectors; C is the number of categories; α is the mixing parameter which determines the importance between attribute discriminative power and reliability.

We note that the optimisation problem presented in Eqn. 9 is NP-hard as it involves optimisation in binary space [16]. This means that the problem cannot be solved by any traditional optimisation algorithm such as gradient descent algorithms. As such, we propose a greedy algorithm wherein for each step, it chooses the attribute that maximises the objective. We call this algorithm Discriminative and Reliable Attribute Learning (DRAL).

The goal of the DRAL algorithm is to select a subset of attributes S so that it maximises $\mathrm{J}(\cdot)$. The algorithm is presented in Algorithm 1. The algorithm optimises the function $\mathrm{J}(\cdot)$ by optimising a single attribute at a time. Let us suppose that we want to optimise the k-th attribute in S. This can be done by converting Eqn. 9 into:

$$\begin{aligned}\mathrm{J}\left(\{U, \{\boldsymbol{h_i}\}_{i=1}^{C}, \{\Phi_b\}_{b=1}^{B}\}\right) = \\ \underset{k\in U}{argmax}\left(\alpha\hat{\omega}_k + (1-\alpha)\sum_i\sum_j\|h_{i,k} - h_{j,k}\|_H + \right. \\ \left.\alpha\sum_{b\neq k}\hat{\omega}_b + (1-\alpha)\sum_i\sum_j\sum_{b\neq k}\|h_{i,b} - h_{j,b}\|_H\right)\end{aligned} \tag{10}$$

which then can be further simplified into:

$$\begin{aligned}\mathrm{J}\left(\{U, \{\boldsymbol{h_i}\}_{i=1}^{C}, \{\Phi_b\}_{b=1}^{B}\}\right) = \\ \underset{k\in U}{argmax}\left(\alpha\hat{\omega}_k + (1-\alpha)\sum_i\sum_j\|h_{i,k} - h_{j,k}\|_H + C\right)\end{aligned} \tag{11}$$

where $C = \alpha\sum_{b\neq k}\hat{\omega}_b + (1-\alpha)\sum_i\sum_j\sum_{b\neq k}\|h_{i,b} - h_{j,b}\|_H$; $h_{i,k}$ is the k-th element of the category-level attribute descriptor i. To solve the above equation, the proposed algorithm chooses k from U which optimises the above function. Here the k attribute is not included in the set S

Before optimising the objective function with respect to $k \in U$, we would need to choose $l \in S$ which will be replaced by k. In this case, l needs to be the attribute that is most unreliable and non-discriminative. This means we need to solve the following problem:

$$\begin{aligned}\mathrm{J}\left(\{U, \{\boldsymbol{h_i}\}_{i=1}^{C}, \{\Phi_b\}_{b=1}^{B}\}\right) = \\ \underset{l\in S}{argmin}\left(\alpha\hat{\omega}_l + (1-\alpha)\sum_i\sum_j\|h_{i,l} - h_{j,l}\|_H + C\right)\end{aligned} \tag{12}$$

where $C = \alpha\sum_{b\neq l}\hat{\omega}_b + (1-\alpha)\sum_i\sum_j\sum_{b\neq l}\|h_{i,b} - h_{j,b}\|_H$. The above equation can be addressed by choosing the attribute from the selected subset S which minimises the function.

Algorithm 1 The proposed greedy algorithm for solving Eqn. 9. The final result is S which is the most discriminative and reliable attribute set selected from U; N is the number of attributes (*i.e.* $N = |S|$)

Require: $\{U, \{\boldsymbol{h_i}\}_{i=1}^{C}, \{\Phi_b\}_{b=1}^{B}\}, N$
1: $S \leftarrow$ randomly select N number of attributes from U
2: **repeat**
3: $\quad l \in S \leftarrow$ Solve Eqn. 12
4: $\quad S = S - \{l\}$
5: $\quad k \in U \leftarrow$ Solve Eqn. 11
6: $\quad S = S \cup \{k\}$
7: **until** S does not change

Given a subset S, the algorithm will alternate between solving Eqn. 12 and Eqn. 11. It stops when the member of subset S does not change any further.

There are several design choices on how S is initialised. However, from our empirical analysis, initialising S by randomly selecting attributes from U always gives quick convergence. Therefore, we will use random selection to initialise S. The full algorithm is presented in Algorithm 1. We will later show in the experiment that by doing this procedure, the algorithm monotonically increases the objective function and thus convergence can be reached.

Another way to solve Eqn. 9 is by considering a group of attributes instead of individual attribute. We call this approach as group selection approach. Unlike the proposed approach, in the group selection approach, at one instance, we would like to select a group of attributes that will optimise Eqn. 9. Nevertheless, from our observation, in this setting, the solution can always be reduced to the single attribute selection presented in Eqn. 11 and Eqn. 12. This entails the group selection would give virtually the same results as the proposed approach.

IV. Experiment Evaluation

In this section, the variants of the proposed approach are evaluated and compared. Then the best performing system will be contrasted to the state-of-the-art method named Direct Attribute Prediction (DAP) [10]. We note that we use the same classifier as DAP for all variants. The difference is that the DAP uses the whole set of attribute pool. We consider the zero-shot learning problem applied in the Animal with Attribute dataset (AwA) [10].

A. Dataset and Experiment settings

The AwA dataset contains 35,474 images of 50 animal categories with 85 attribute labels. It has two types of labels for each image: the attribute label and category label. Category label indicates the animal category to which the image belongs. Attribute label represents the presence/absence of an attribute in an image. Therefore, each image is represented by 85 dimensional attribute descriptor. We note that in this dataset, all the images in a same category have same attribute representation. We follow the experiment protocol and the settings used in [10] for the zero-shot learning problem. In particular, the categories are divided into two disjoint sets: 40 categories for training and 10 categories for testing. In this way, there is no training image given for the 10 categories in the test set. However, the manually labelled category-level attributes for each test category are given.

For the low-level feature used to train the attribute detectors and detect the attributes, we use the same extracted features as in [10] such as: HSV colour histogram, SIFT [14], rgSIFT [23], PHOG [2], SURF [1] and local self-similarity [21]. All the features are combined using the Multiple Kernel Learning. We also use the kernels provided from the author, to make our results comparable to the previous works. In addition, we also use the same parameters to train the attribute detectors and repeat the experiment 5 times.

The proposed DRAL algorithm has three parameters: β and γ which are used for Eqn. 6 and the mixing coefficient α. The values of all parameters are selected from the cross-validation set. From our empirical analysis we found that $\gamma = \frac{1}{B}\sum_b \hat{\omega}_b$ to be a good value. In addition β is determined from range $[0, 100]$.

The mixing coefficient α determines the importance of the attribute properties (*i.e.* attribute discriminative power and attribute reliability). We search α with range $[0.1...0.9]$ and we find that $\alpha = 0.9$ to perform best. Intuitively, we should put more important into the attribute discriminative power when there are a large number of categories. This can be explained from the fact that large number of categories require longer binary code to sufficiently separate them. However, as mentioned, there are only 10 categories in the test set, thus, we need to put more importance toward the attribute reliability.

B. Experimental Results

For the first evaluation, we compare five variants of the proposed system: (1) DRAL using only the attribute discriminative power information (*i.e.* $\alpha = 0$), denoted DRAL (discriminative); (2) DRAL using only the attribute reliability information (*i.e.* $\alpha = 1.0$), DRAL (reliability); (3) the proposed DRAL using both attribute properties, denoted DRAL (both); (4) semi-random selection and (5) random selection. The semi-random selection approach uses the DRAL algorithm without solving Eqn. 12. Instead the approach randomly selects $l \in S$. Whilst the random selection approach randomly selects S from U.

We first present the empirical study of the study the proposed algorithm's convergence. Fig. 1 shows the plot of the objective function score presented in Eqn. 9 for each variant of DRAL in every loop. Note that for the case of random selection, the attribute set S is randomly selected for every loop. This result suggests that when using both attribute properties, the proposed algorithm achieves the highest convergence rate (*i.e.* after iteration 20). Moreover, the other approaches are not able to maximise the objective function. The semi-random selection variant requires a more iterations to converge. This shows that our strategy requires both attribute properties in order to maximise the objective function.

It is noteworthy to mention that the algorithm did not converge when using only the attribute discriminative power property. On closer examination, we found that the system picked many unreliable attributes generating discriminative attribute descriptors that sufficiently separates the 10 test categories. This generated large errors during the attribute descriptor extraction on each test image which led to a large classification error.

In the second evaluation, we compared the performance of all variants in the test set. To this end, we varied the number of selected attribute N from 35 to 75. Fig. 2 presents the results. The proposed DRAL algorithm using both attribute

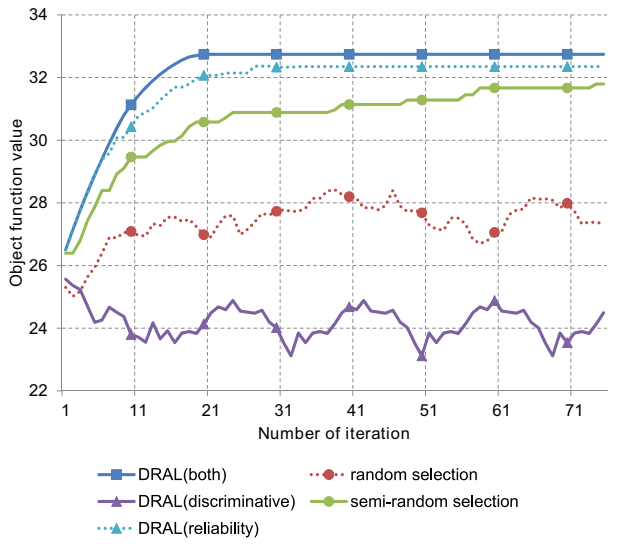


Fig. 1. The plot of objective function(Eqn. (9) for each variant of the proposed approach

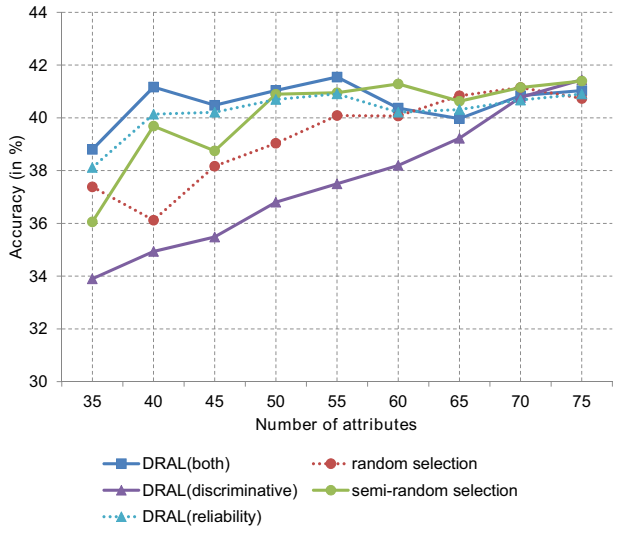


Fig. 2. Comparison of the proposed approach variants when the number of selected attributes varies from 35 to 75, The best performance (41.5% in accuracy outperforms that of DAP) appears at the point when 55 selected attributes used .

properties generally perform better than the other variants. The variant achieves slightly better performance than the original DAP when only 55 attributes were selected (*i.e.* 35% less). Moreover, we can reduce this to 40 with a price of slight performance loss (41.2%). This suggests that proposed algorithm is able to select the most discriminative attribute set from the 85 attributes provided in the dataset. Table I presents further detailed results when the number of attributes was set to 55. These results are consistent with the convergence evaluation presented before.

TABLE I. ZERO-SHOT MULTI-CLASS CLASSIFICATION ACCURACY ON 10 NOVEL ANIMALS CATEGORIES SELECTING 55 ATTRIBUTES.

Methods	Accuracy (in %)
DRAL (both)	**41.5**
DRAL (reliability)	40.9
DRAL (discriminative)	37.5
Semi-random selection	40.6
Random selection	40.2
original DAP [10]	41.4

C. Comparative analysis to DAP

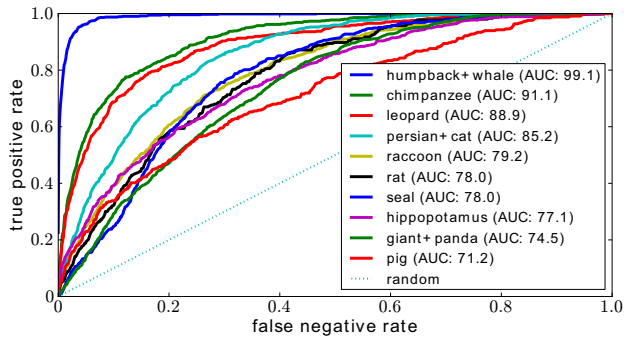
In this evaluation, we use the best performing system previously found (*i.e.* DRAL(both)). Fig. 3 present the comparison between the DAP and the DRAL ROC curves. This further validates the efficacy of the proposed system. The AUCs of the system in most categories are better than those of the DAP. That suggests that the automatic selection of discriminative and reliable attributes does indeed notably improve the performance over the DAP in most test categories. In particular, it significantly outperforms DAP in *leopard*, *persian+cat*, *chimpanzee* and *seal*. However, we note that there are still two categories performing worse namely the *pig* and *hippopotamus* categories.

V. MAIN FINDINGS AND FUTURE DIRECTION

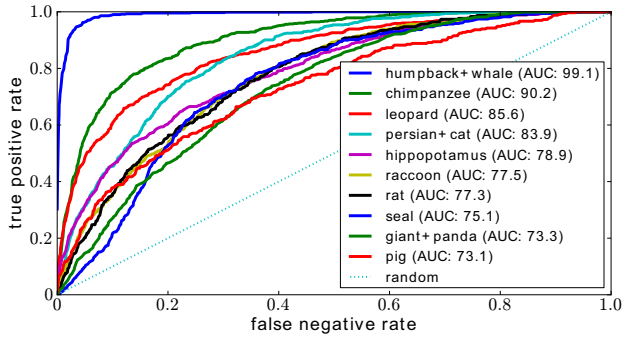
Image attributes offer a convenient way of bringing semantic concepts into machine-readable image representation. Although these image attributes are generally semantically meaningful, they are not necessarily discriminative. This means, there is no guarantee for image classification systems using this approach to achieve good performance. To that end, in the present work we study two properties of image attributes: attribute discriminative power and attribute reliability. The attribute discriminative power is related to the property of a set of image attributes to separate images of different categories. Whilst, the attribute reliability is related to the error produced during the attribute descriptor extraction process. We propose a greedy algorithm, here denoted Discriminative Reliable Attribute Learning (DRAL), to select a subset of attributes maximising an objective function that incorporates the two properties. Given a pool of image attributes, the algorithm first selects the image attribute minimising the objective function from the selected set. Then, it replaces the image attribute from the pool with the one maximising the objective function. The process iterates until the selected set does not change.

We empirically showed that the algorithm converges and was able to optimise the objective function. We contrasted our proposed approach with the state-of-the-art approach, denoted DAP for the zero shot learning problem in the Animal with Attribute dataset. The results demonstrated that with significantly less number of attributes the proposed approach achieved a comparable performance to the DAP approach.

There are many extensions and feasible enhancements can be explored in the future. For instance, we could use a better approximation to measure the attribute reliability property that considers both the detector performance as well as the semantic drift. Another interesting future direction is to find the smallest set of attributes by adding an additional regularisation term in the objective function. We can also explore some novel applications for the proposed strategy such as super resolution [12], 3D reconstruction [28], [29] or anomaly detection in



(a) DRAL



(b) DAP

Fig. 3. Comparison of the Performance between the proposed method DRAL and DAP ROC-curves and AUC value for the ten test classes

surveillance systems [25]. Here we can use attributes of low resolution image as the query to collect the high resolution images which have similar parts to that, then use the patches of the high resolution images as sources to approximate the patches of low resolution image and reconstruct the high resolution images.

ACKNOWLEDGEMENTS

This research was partly funded by Sullivan Nicolaides Pathology, Australia and the Australian Research Council (ARC) Linkage Projects Grant LP130100230.

REFERENCES

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
[2] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the ACM international conference on Image and video retrieval (CVIR)*, pages 401–408, 2007.
[3] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.
[4] X. Chang, F. Nie, Y. Yang, and H. Huang. A convex formulation for semi-supervised multi-label feature selection. In *Proc. AAAI*, 2014.
[5] X. Chang, H. Shen, S. Wang, J. Liu, and X. Li. Semi-supervised feature analysis for multimedia annotation by mining label correlation. In *PAKDD*, 2014.
[6] X. Chen, A. Shrivastava, and A. Gupta. Neil: Extracting visual knowledge from web data. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013.
[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
[8] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *CVPR*, 2009.
[9] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009.
[10] C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99:1, 2013.
[11] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Advances in neural information processing systems*, 2010.
[12] L. Liu, W. Li, S. Tang, and W. Gong. A novel separating strategy for face hallucination. In *IEEE International Conference on Image Processing (ICIP)*, 2012.
[13] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282, 2007.
[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
[15] Z. Ma, Y. Yang, Z. Xu, S. Yan, N. Sebe, and A. G. Hauptmann. Complex event detection via multi-source video attributes. In *CVPR*, 2013.
[16] B. Manthey and R. Reischuk. The intractability of computing the hamming distance. *Theoretical Computer Science*, 337(13):331 – 346, 2005.
[17] D. Parikh and K. Grauman. Interactively building a discriminative vocabulary of nameable attributes. In *CVPR*, 2011.
[18] D. Parikh and K. Grauman. Relative attributes. In *ICCV*, 2011.
[19] A. Parkash and D. Parikh. Attributes for classifier feedback. In *ECCV*. 2012.
[20] M. Rastegari, A. Farhadi, and D. Forsyth. Attribute discovery via predictable discriminative binary codes. In *European Conference on Computer Vision (ECCV)*. 2012.
[21] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *CVPR*, 2007.
[22] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280, 2008.
[23] K. E. Van De Sande, T. Gevers, and C. G. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.
[24] V. Vapnik and O. Chapelle. Bounds on error expectation for support vector machines. *Neural Computation*, 12(9):2013–2036, 2000.
[25] A. Wiliem, V. Madasu, W. Boles, and P. Yarlagadda. A suspicious behaviour detection using a context space model for smart surveillance systems. *Computer Vision and Image Understanding*, 116(2):194–209, 2012.
[26] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. J. Guibas, and L. Fei-Fei. Action recognition by learning bases of action attributes and parts. In *ICCV*, 2011.
[27] F. Yu, L. Cao, R. Feris, J. Smith, and S.-F. Chang. Designing category-level attributes for discriminative visual recognition. In *CVPR*, 2013.
[28] Y. Zhu, D. Huang, F. De La Torre, and S. Lucey. Complex non-rigid motion 3d reconstruction by union of subspaces. In *CVPR*, 2014.
[29] Y. Zhu and S. Lucey. Convolutional sparse coding for trajectory reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2013.