Describing Objects via Attribute Detection

Kylie McCarty and **Abdullah Jamal**
University of Central Florida

Abstract

Traditional classification tasks seek to name the object present in an image after learning from a large amount of labeled examples for each object class. Our task is to essentially shift the goal of recognition from naming the object to describing it's attributes. Doing this opens the opportunity for systems that can identify object classes never seen before using high level descriptions. In theory, this will allow us to decrease the number of necessary training examples since not all object classes will need to be trained on. We propose two simple, baseline methods in order to explore this idea: SVM and fine-tuned CNN.

1 Introduction

Traditional classification tasks seek to name the object present in an image after learning from a large amount of labeled examples for each object class. Although humans can also learn very well from examples, humans also have the capability to detect classes with no visual examples when provided with a high-level description. In 1987, Biederman estimated that humans distinguish between at least 30,000 visual object categories [1]. He also showed that the natural arrangement and co-occurrence of objects in scenes strongly influences how easy it is to detect objects [2]. With this in mind, we explore the possibility of a system that can produce these high level descriptions that have enabled humans to discern between thousands of objects without requiring millions of training examples.

Our task is to essentially shift the goal of recognition from naming the object to describing it's attributes. Attributes can range from simple concepts such as color and shape to slightly more complex concepts like the type of material and texture and extend to high level concepts such as arm, face, door, and window. Our goal is to determine whether attribute detection can be generalizable across classes and if it is successful enough to become a viable method in object classification.

1.1 Benefits

The benefits of using this technique for object classification include:

1. It is possible to identify an object despite never having trained on that object class since attributes are generalizable across classes.
2. Reduce the number of manually labeled training examples required to train a network to identify objects
3. Can offer further, more descriptive information about the identified object.
4. Even if the classifier fails to identify the object, can still offer some valuable information about the object.

1.2 Challenges

Unfortunately, some of the same aspects that make this technique so promising also make the task especially difficult:

1. Attributes must be generalizable across classes; There is significant variance within positive samples for a single attribute.

2. Some attributes are also highly conceptual and difficult to distinguish visually such as domestic and fast.

See **Figure 1.**



Figure 1: Demonstrates some of the more difficult attributes

2 Related Works

Learning-based methods for recognizing objects has been well researched and made huge improvements over the past decade. However, only more recently have researchers have began looking into attribute based recognition. Q. Chen et al. [3] applied a deep model to detect clothing attributes. V. Escorcia et al. [4] and B. Gong et al. [5] also explore the use of deep networks in attribute detection. Farhadi et al. [6] also performed attribute detection and also chose to model cross-category generalization of the attributes, similar to our approach.

3 Method

For this project we chose two distinct data sets. Our first data set (a-Yahoo / a-Pascal 2008) was collected for the paper "Describing Objects by Their Attributes." The data set contains images labeled for 64 attributes (ex. mouth, metal, furniture leg, vegetation) and contains 20 object classes in the training set and 12 object classes in the testing set. The object classes in the train and test sets are disjoint in order for us to determine how recognizable attributes are across varying object classes. Our second data set is Animals with Attributes (AWA). The AWA data set is annotated with 85 attributes (ex. hairless, horns, ocean, bipedal) with 40 animal classes in the training set and 10 animal classes in the testing set. The test and train sets of this data set are also disjoint.

We also chose to compare two different methods to detect and predict these attributes; the first method is SVM and the second is a fine-tuned CNN.

3.1 SVM

In order to utilize SVM's to do attribute detection, meaningful features must be chosen as input to the SVM. We chose to extract the last fully connected layer of VGG-16/VGG-19 as the feature vectors for our images. The feature vectors for the training set are fed into the SVM and the SVM is trained to do binary classification for that attribute: present or not present. The feature vectors for the testing set are then fed into the SVM to predict the presence of the attribute. This process must be repeated for all 64/85 attributes. See **Figure 2.**

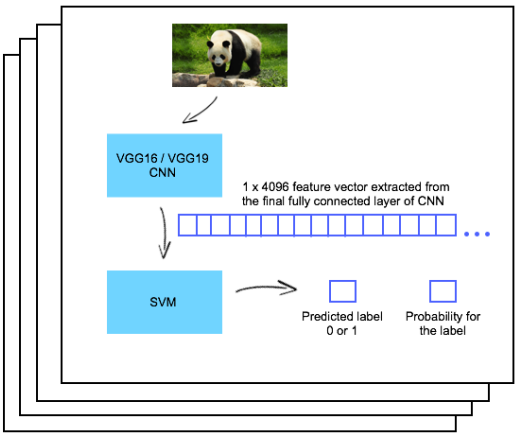


Figure 2: SVM Model

3.2 Fine-tuned CNN

The second method we explored was the use of CNN's. We chose to fine-tune an existing CNN (VGG-16). Fine-tuning involved:

1. removing the final two layers of the network
2. adding a new fully connected layer and a weighted sigmoid cross-entropy loss layer
3. altering the input data layers to take multiple labels and weights per image

See **Figure 3.**

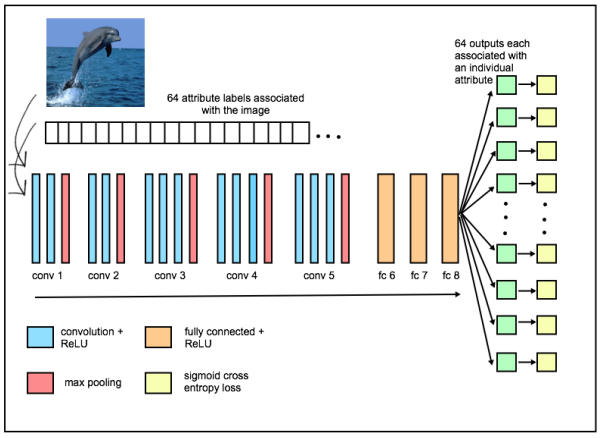


Figure 3: Fine-tuned CNN Model

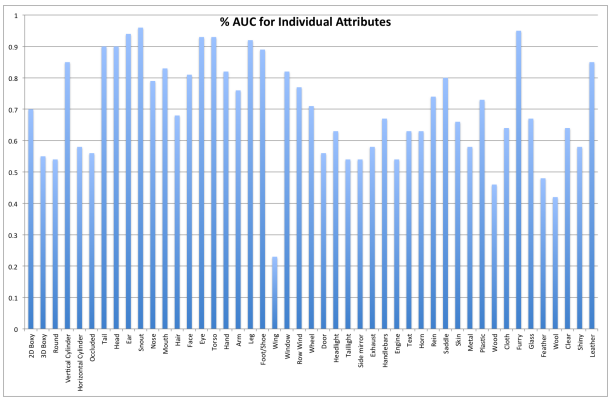


Figure 4: % AUC for different attributes in a-Pascal/a-Yahoo data set

	Accuracy	AUC	mAP
Method 1	0.91	0.70	0.33
Method 2	0.89	0.64	0.55

Figure 5: Side by side comparison of our 2 methods on the a-Pascal / a-Yahoo data shown in 3 different performance metrics

Approaches	AWA Dataset	a-Yahoo Dataset
IAP[1]	74.0 – 79.2	-
ALE[2]	65.7	-
HAP[3]	74.0 – 79.1	58.2
CSHAP[3]	74.3 – 79.4	58.2
DAP[1]	72.8 – 78.9	77.4
KDICA[4]	84.4	84.7
Method 1		**70.1**
Method 2	-	**65.0**

Figure 6: Average Attribute Prediction Accuracy in % AUC

4 Results

Our goal was to determine if attributes are generalizable across classes and to build efficient, precise techniques to detect these attributes. Our results show that there is definitely an ability to develop attribute detectors that can detect attributes in classes they have not seen before, though there is clear room for improvement. **Figure 4** shows that some attributes were much easier for our proposed methods to learn than others. In order for these attributes to go on to be used for classification, it is essential for successful % AUC for all of the attributes. **Figure 5** compares our two methods against each other using a couple different metrics and **Figure 6** compares our method to some other methods. All of the results in **Figure 6** are reported in Gong et al[5]. Going forward, it would be ideal to develop a data set better groomed for our exact task. The next phase of the project would be attempting object classification from these detected attributes.

Acknowledgements

I would like to thank Dr. Gong, Dr. Lobo, and Dr. Shah for their guidance throughout this process. I would also like to thank the National Science Foundation and the Center for Research in Computer Vision at UCF for providing the opportunity for this REU.

References

1. I. Biederman. Recognition by components - a theory of human image understanding. Psychological Review 94(2), 115147. 1987.

2. I. Biederman, R.J. Mezzanotte, and J.C. Rabinowitz. Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology 14, 143177. 1982.

3. Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan. Deep domain adaption for descrbing people based on fine-grained clothing attributes. In CVPR, 2015.

4. V. Escorcia, J. C. Niebles, and B. Ghanem. On the relationship between visual attributes and convolutional networks. In CVPR, 2015.

5. C. Gan, B. Gong, and T. Yang. Learning At-

tributes Equals Multi-Source Domain Generalization. In CVPR, 2016.

6. A. Farhadi, and I. Endres, D. Hoiem, D. Forsyth. Describing Objects by their Attributes. University of Illinois at Urbana-Champaign. 2008.

7. C.H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In CVPR, 2009.

8. Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid. Label-embedding for attribute-based classification. In CVPR, 2013.

9. S.W. Choi, C. H. Lee, and A. Rostamizadeh. Algorithms for learning kernels based on centered alignment. The Journal of Machine Learning Research, 13(1):795-828, 2012.