Effect of Visual Attributes on Classification Accuracy of Zero-Shot Learning

Jacob Chandran

River Hill High School

Mentor: Dr. I-Jeng Wang

Johns Hopkins University Applied Physics Laboratory

Effect of Visual Attributes on Classification Accuracy of Zero Shot Learning

Table Contents

List of Figures

List of Tables

ABSTRACT

The problem of improving classification accuracy of the Zero Shot learning; where no training samples are available for some image classes is studied in this research. The method uses Semantic attributes which represent common properties across different categories to transfer knowledge between known and unknown categories. A Zero shot learning attribute classification method based on neural network features, semantic attribute mapping and support vector machines is implemented and a classification experiment is designed to test the effect of visual attributes on classification accuracy. To study the effect of visual attributes on classification accuracy of zero shot learning, eighty-one attributes are used in the experiment, of which 24 attributes correctly classified objects with the classification error between 17% - 46%. Based on the results of this experiment, visual attributes seem to outperform non-visual attributes.

I. **Introduction**

Image classification is the task of identifying objects from images. Standard object classification methods use labeled training images to classify objects. To achieve high classification accuracy, the system requires a very large manually labeled training dataset. Labeling this training set is costly and error prone. When no training samples are provided for the classes, classification of unseen classes is not possible. Humans can classify about 30,000 object classes and are capable of identifying unseen classes when provided with a high-level description. Zero-shot learning is the task of classifying unseen visual classes when no training samples are available for the classes. Zero shot learning can be accomplished using the attribute classification method. Attributes are the higher-level semantic description of the objects observable in images that have human designated name (e.g. the object is 'edible', 'furry' etc.). The semantic attributes shares information between classes, allowing the system to classify objects for which no training data exists. For example, the attribute 'striped' can be used to describe images of tigers, zebras and bees, thus transferring knowledge between classes. In the case of objects with no training labels, the attribute classification system can provide information about the unseen object using semantic attributes (e.g. ‘furry with four legs’ or ‘dog with spots’) and learn to recognize objects from description. Incorporating an attribute layer in the classification system integrates human intelligence into machine learning. Zero shot learning is useful in many situations where no labeled data is available as in the case of new products, the latest gadgets or new models of cars [Socher, R]. Zero shot learning is also useful in developing robotlearning methods. The set of attributes that describe an object (size, weight, etc.) can be predicted from visual features (texture, color, etc.) and these attributes can then be used to infer manipulative strategies of robots: grasping objects that are small, pushing heavy objects or rolling round objects etc. thus creating intelligent robots [Hermans, T.]. Leveraging the relationship between attribute and object class, which convey prior information about object categories can improve the classification accuracy of the zero shot learning systems. In this project, the question if choosing object categories with visual attributes improve classification accuracy of the zero shot learning system is investigated. This report is organized as follows. Related work on zero-shot learning is described in Section II. The hypothesis of the research project is explained in Section III. Section IV describes the formulation of the zero shot problem and the steps of the attribute classification system. Experiment to determine the effect of visual attributes is described in Section V. Results of the experiments are reported in Section VI and the report is concluded in Section VII.

II. **Prior Work**

Recent works in Zero-shot learning by Palatucci [Palatucci, et.al] show mapped functional magnetic resonance images (fMRI) of people thinking about certain words into a space of manually designed features. These people were able to predict semantic features for words which they had never seen before, which is called a zero-shot condition. The researcher Lambert [Lambert, et.al] constructed a set of binary attributes for image classes that conveyed various visual characteristics of animals and predicted attributes using direct and indirect strategies on animals with an attribute dataset. Their experiments showed that by using an attribute layer, it is possible to build a Zero Shot Learning System. The method in this research follows the work of Burlina [Burlina, et.al.,], where attribute classifiers were used with deep neural network visual image features and indirect attribute prediction. Indirect attribute classification is the process of describing an unknown class by drawing details from a similar known class. Instead of directly trying to classify the unknown class, the system returns a probability that the unknown class is each of the training classes. The most probable class is then used to describe the unknown object. For example, if the system had never seen a polar bear before, but had been trained for a brown bear, the traits of the brown bear would be used to describe the polar bear.

III. **Hypothesis**

In visual object classification tasks, human nameable visual attributes observable in images facilitate zero-shot learning by transferring information between objects. Identification of attributes based on human understanding, exploiting attribute-object relationship and discrimination of attributes are very important for improving the classification accuracy of zero shot learning. Certain attributes have a greater tangible and describable significance than others. Namely, visual attributes (e.g. Is it an animal?) which contain specific and observable characteristics, should be easier to test and classify than less visual attributes (e.g. Is it man made?) which do not necessarily have a set of concrete definitions. The hypothesis of this experiment is that visual attributes improve the classification accuracy of zero shot learning systems.

IV. **Visual Attribute Classification for Zero-Shot Learning**

In this section, the concepts of the zero shot learning and the steps involved in the implementation of the attribute classification system are described.

A. Formulation of the Zero Shot Problem and Processing Pipeline

Let X be a feature space, which represents the visual properties of an image, and Y, the set of class labels for which training images and an attribute representation A are available. A is a N_a x N_c attribute matrix where N_a is the number of attributes and N_c the number of classes. Attribute values in the matrix are {-1, 0, +1}. A value of +1 is given by the human annotator if the attribute agreed with the class (e.g., the class 'lion' 'Is it an animal?'). A value of -1 was given when the attribute didn't agree with the class (e.g., the class 'pizza' 'Is it an animal?') and 0 was given when the attribute value was subjective or ambiguous (e.g., the class 'rabbit' 'Is it tasty?'). Z is the set of class labels for classes where no training images are available. Given a set of N training images (x_1, l_1), … (x_n, l_n) included in the matrix X × Y, the zero shot problem aims to label samples from unseen classes in Z. This is done by constructing a classifier that applies image features in X to the Z classes, by transferring information between X and Z through the attribute representation A, only using labeled training images in X × Y.

The attribute classifier transfers the information from known class Y to Z using the attributes A of the known class Y. During training, a prediction model is created based on the known attributes of a set of classes using linear multiclass Support Vector Machines [Bishop]. During testing, data about attributes from the training classes are used to predict the attributes of unknown classes. The decisions for the test classes are based only on the attributes of the training classes because there is no data for the test classes.

Figure 1 Attribute based classification. A description of attributes such as furry, black, and striped, for the known class's dog, chimpanzee and tiger allows the transfer of knowledge between object categories. Learning visual attributes from these known classes with training samples helps identify unknown class objects that do not have any training images. Equipped with the attributes, the machine can now say something intelligent about the unknown class such as, stripped, black, white, and big to describe a zebra. What is most exciting about attributes is the fact that they provide a mode of communication between humans and machines.

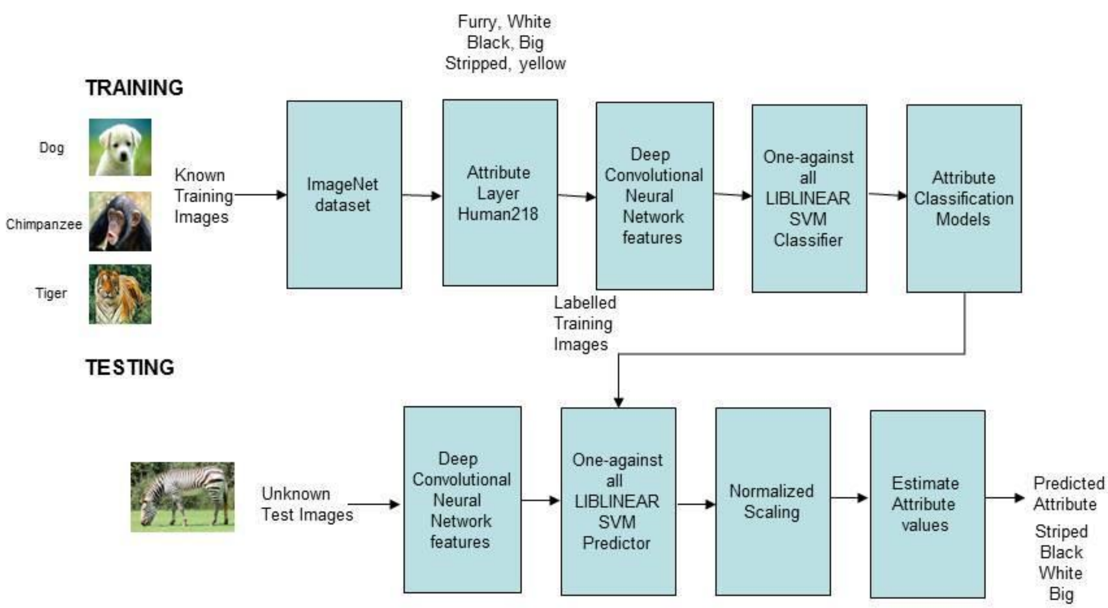

Figure 2. Processing pipeline of attribute classification method for Zero-Shot Learning

Figure 2 shows the processing pipeline of attribute classification, which consists of two phases: training and testing. In the training phase, an attribute classification model is generated and in the testing phase, the image class attributes are predicted using the model generated in the training phase. The steps of training are: first, identify training image classes from ImageNet dataset [Olga, R.], second, build an attribute layer using Human218 attribute set [Palatucci, M.] third, compute visual features using Deep Convolutional Neural Net OverFeat neural net architecture for ImageNet training images with class labels Y, and fourth, construct image classification models by training multi-class one against all linear support vector machines from training image features. The steps in testing are first, compute visual neural network features, and second, to use the model created in the training section to determine the probability that each training class is the testing class- a representation of the similarity between each training and test class. After these procedures, the training class with the highest probability, or greatest similarity is picked and used to predict attributes for the test class. In other words, in order to have information on the test class, which has no information; the attributes of a known class with the greatest similarity to the test class are used to predict the attributes of the unknown class.

B. Visual Attributes

Nameable visual attributes are important features for object recognition in attribute classification [Parikh. D.]. In the standard classification system, the goal of a system is to correctly classify only known objects contained in a dataset. Attributes expand the scope of a system in that they allow the system to specifically name properties of both known and unknown objects, which aids in creating a more complete classification of an object. Attributes are more or less adjectives of human language, making them useful to compute meaningful description of unfamiliar objects. With zero shot learning, it is now possible to completely describe an unknown object which helps create a fairly accurate picture of the unknown object.

In this project, the semantic knowledge base human218 was used to create an attribute set [Palatucci, M.]. There were 218 attributes in the form of questions, and the questions were selected to include many different possible attributes of visual objects bundled into one question. For example, there are questions that judge the size, shape, surface properties, and typical usages of the objects such as “Is it man made?” or “Can you hold it?” etc.

The training and testing images for attribute classification are extracted from ImageNet database. The Image database “ImageNet” [Olga R.] is built on WordNet containing a set of synonyms (80,000 nouns) and has an average of 500-1000 clean images per object class without background clutter. The significance of this dataset is that it introduces new semantic relations for zero shot learning. ImageNet is uniquely linked to all nouns of WorldNet and the semantic relations of different words can be used to create new models in zero shot learning. Fifty-seven image classes were chosen and used with the Human218 attributes to form the attribute set A. The data set used for the project has 57 classes from ImageNet, the 218 questions from the Human218 attributes set and a total of 75,489 images [Burlina, M.]

A semantic attribute mapping matrix V of discrete values {-1, 0, +1} is created for each class. The value 1 corresponding to the class having the attribute, -1 not having the attribute and

0 for an ambiguous decision by a human annotator. The class-to-attribute mapping matrix has 57 classes with 218 attribute values. The attribute values are estimated during testing to predict the different qualities of unknown classes Z from the training information of known classes Y.

C. Visual Deep Neural Network Features

The Zero Shot learning method uses deep learning features from pre-trained Deep Convolutional Neural Network (DCNN) derived from the OverFeat neural net architecture [Sermanet, P.]. The OverFeat features are features used to represent an image for image classification and attribute detection. These features were found to perform well when applied to the classification of images trained with an image dataset and tested with another dataset [Razavian, A.S.]. In this project, 4096 features derived from ImageNet color images of size of 231 x 231 were used. The features were normalized to make classification invariant to changes in the image illumination.

D. Support Vector Machines

A Support Vector Machine (SVM) is a discriminative classifier defined by a separating hyperplane. Given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new test samples (Figure 3). LIBLINEAR is a linear Support Vector Machine classifier [Bishop, M, Boyd, S.] for multi-class classification (One vs. rest) and is faster for training and predicting large datasets [Fan, R]. In this research LIBLINEAR algorithm was used to classify image classes for selected attributes.

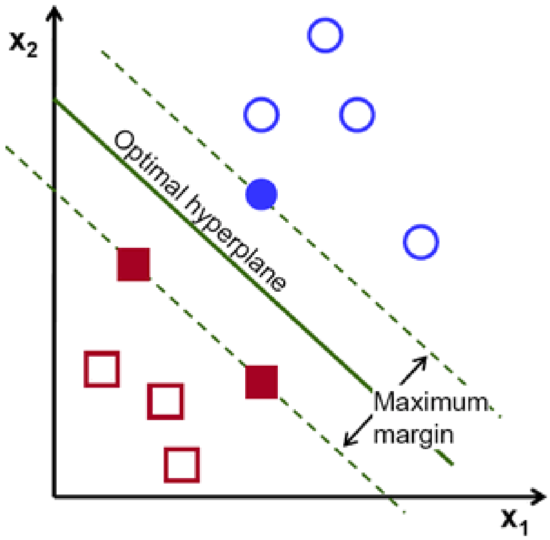

Figure 3 Support Vector Machines. The theory of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training samples. The optimal separating hyperplane maximizes the margin (distance between the data of two training classes). The training samples closest to the hyperplane are called Support Vectors.

E. Predicting Attributes

The predicted attribute values are estimated from the output of the one against all image classifier models trained using labeled image sets. The SVM classifier produces un-calibrated values that need to be modified in order to interpret the results using probability of the outcome. In order to understand the values, the values have to be rescaled to produce values from 0 to 100. To get the probability values, the output of the SVM Classifier is normalized using scaling [Platt, J. C.]. The probability values are then used to estimate the attribute value [-1, +1] of the test image classes. The output is thus a probability of class label Y with the input feature vector X and is used to estimate the attribute value {+1, -1}. This numerical value answers yes (1), no (-1), or unsure (0) for the attribute.

V. Experiment to determine the effect of visual attribute on classification accuracy

An experiment using the attribute classification system was conducted to determine the effect of visual attributes on classification accuracy. Figure 4 shows the steps of the experiment. Twenty test classes were chosen randomly from 57 image classes. For each attribute, ten training classes not in the selected test classes were chosen. The ten train classes were equally split between classes with an attribute value of +1 and -1. The value +1 indicated that the attribute satisfied visual characteristics of the image class while -1 did not. Since the attribute values of the test classes were unknown, the train classes were chosen to reflect the +1 and -1 values. The SVM model was used to predict the test classes. For each test class, OverFeat features are extracted and normalized. The normalized features are input to the 10-1 class SVM Model generated in training. The predicted values from SVM Model were then scaled to obtain probability estimates between the range 0 and 1.0. For each attribute, there were about 20 Test Classes x 10 Train Classes x Number of Images in each test class probability estimates.

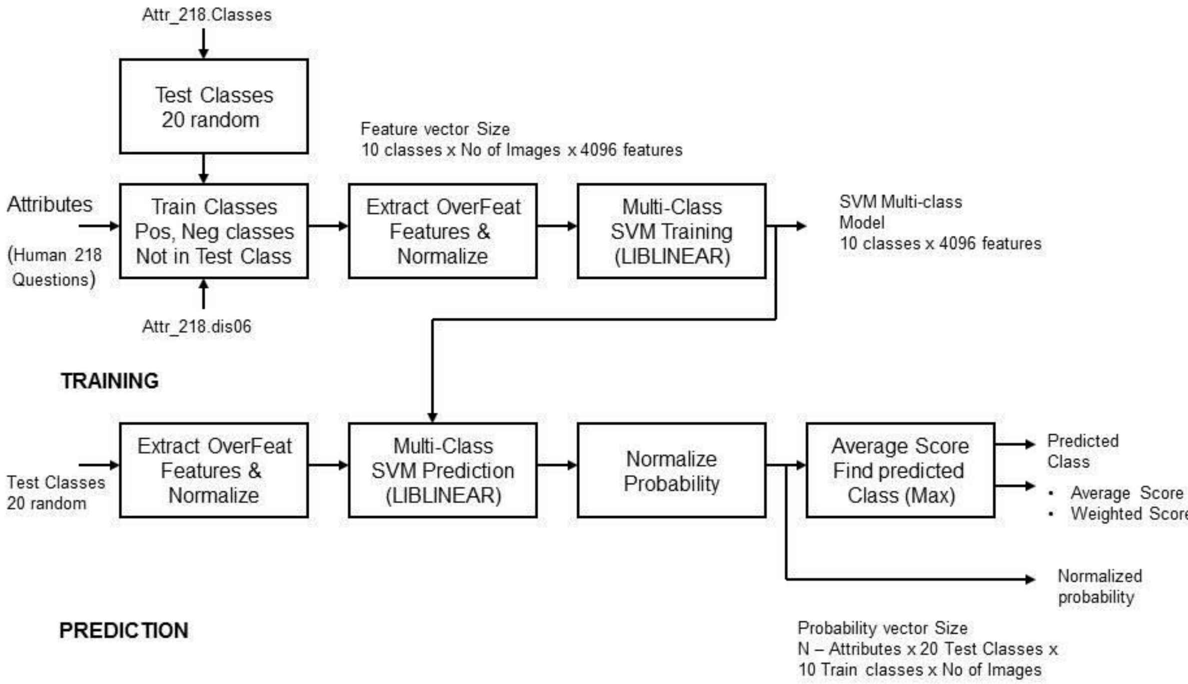

Figure 4. Experimental processing steps to determine the effect of visual attribute on classification accuracy

For the attribute question and for each test class, the true attribute value was taken from the human218 attribute matrix (e.g., Is it an Animal, test Class Comb, attribute value -1). Then, all the scaled probability estimates for the N images in the test class were collected. The probability estimates of the first 5 values were weighted by +1 and the last 5 by a weight of -1. These weights corresponded to the choice of 5 positive and 5 negative valued training classes. The probability estimate values were then summed to find a single value for each image in the set. Finally, a threshold of +0.6 and -0.6 was applied to the probability estimate values. The probability estimate values above +0.6 were assigned an attribute value of 1. The probability estimate values below -0.6 were assigned an attribute value -1 and those between -0.6 and +0.6 were assigned a zero value. The estimated attribute values obtained from the probability estimates were then compared with the attribute values of the test classes to determine if the classification of the attribute was correct and error in classification is computed.

VI. RESULTS

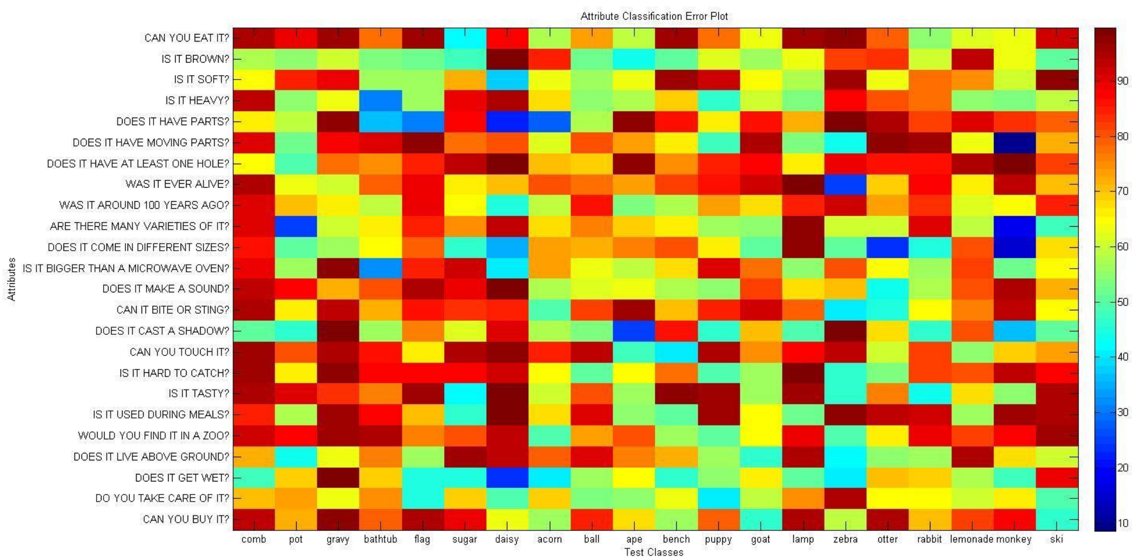

Figure 5 Attribute Classification Errors

Figure 5 shows the plot of the percentage error of attribute vs. test class. The x axis corresponds to 20 test classes (Table 1) and the y axis corresponds to the correct attribute for lowest classification error. Of the 81 attributes used in the experiment, 24 attributes correctly the underlying objects. Figure 6 and Table 2 show the 24 attributes, the correctly classified objects and the classification error. Classification error varied from 17% to 46 %. Based on the results of this experiment, visual attributes seem to outperform non visual attributes. Attributes like, “Is it brown” (43.44%) or “Does it cast a shadow” (25%) consistently produced less error than classes like “Can you buy it” or “is it tasty”. One reason that visual attributes may produce less error than nonvisual attributes is that they are directly verifiable based on an image of the object, which is how the system is trained. An attribute such as “can you buy it” is relative and can be affected by environmental conditions not actively observable in an image. In addition, most visual attributes contain fewer conditions than nonvisual attributes, which require much more information than an image can provide in order to verify certain qualities. In conclusion, when trying to accurately describe an object, the visual and directly observable characteristics are those that can be most accurately predicted rather than those that have specific cases based on environmental and outer conditions not observable from a simple image. Also, those attributes that depend on the least amount of conditions offer less room for error in classification. In order to reduce error, attribute questions should be as direct as possible, not leaving any room for possible misinterpretation or variability.

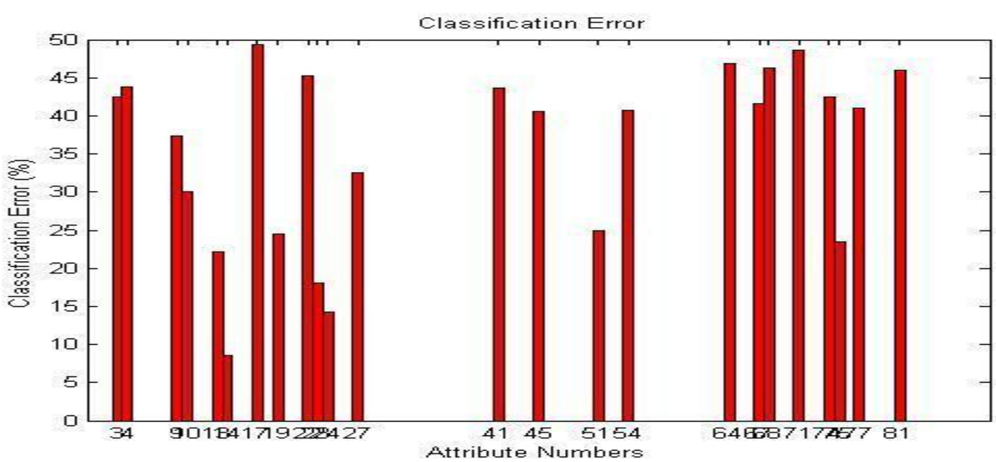

Figure 6 Attribute Numbers vs. Classification Error

No	Attribute	Class	Error %
3	CAN YOU EAT IT?	sugar	42.523
4	IS IT BROWN?	ape	43.844
9	IS IT SOFT?	daisy	37.423
10	IS IT HEAVY?	bathtub	30.034
13	DOES IT HAVE PARTS?	daisy	22.209
14	DOES IT HAVE MOVING PARTS?	monkey	8.5893
17	DOES IT HAVE AT LEAST ONE HOLE?	pot	49.374
19	WAS IT EVER ALIVE?	zebra	24.559
22	WAS IT AROUND 100 YEARS AGO?	daisy	45.276
23	ARE THERE MANY VARIETIES OF IT?	monkey	17.994
24	DOES IT COME IN DIFFERENT SIZES?	monkey	14.295
27	IS IT BIGGER THAN A MICROWAVE OVEN?	bathtub	32.466
41	DOES IT MAKE A SOUND?	otter	43.697
45	CAN IT BITE OR STING?	zebra	40.638
51	DOES IT CAST A SHADOW?	ape	25.00
54	CAN YOU TOUCH IT?	bench	40.738
64	IS IT HARD TO CATCH?	zebra	46.811
67	IS IT TASTY?	sugar	41.622
68	IS IT USED DURING MEALS?	sugar	46.306
71	WOULD YOU FIND IT IN A ZOO?	zebra	48.575
74	DOES IT LIVE ABOVE GROUND?	zebra	42.469
75	DOES IT GET WET?	daisy	23.497
77	DO YOU TAKE CARE OF IT?	puppy	41.053
81	CAN YOU BUY IT	goat	46.006

Table 1. Attribute, identified classes and error in classification

	Test Classes	Comb	Pot	Gravy	Bathtub	Flag	sugar	daisy	acorn	ball	ape	bench	puppy	goat	lamb	zebra	otter
No	Attribute																
1	IS IT AN ANIMAL?	89	56	90	57	76	73	58	40	67	99	59	93	92	95	93	52
2	IS IT MANMADE?	91	84	59	73	91	55	96	95	63	64	75	75	75	97	47	36
3	CAN YOU EAT IT?	95	89	97	78	97	43	87	58	74	60	96	78	63	96	98	79
4	IS IT BROWN?	57	55	61	53	52	48	99	85	53	44	51	62	56	63	82	83
5	DOES IT HAVE A FRONT AND A BACK?	73	78	97	41	74	76	87	90	91	49	53	68	80	48	99	91
6	DOES IT HAVE A FLAT / STRAIGHT TOP?	50	78	64	69	47	77	97	96	97	44	54	61	46	54	65	38
7	DOES IT HAVE CORNERS?	75	59	93	67	93	60	94	95	73	63	97	73	46	82	7	17
8	IS IT HAIRY?	57	41	71	75	94	67	54	47	57	99	53	99	96	84	92	94
9	IS IT SOFT?	64	84	88	56	56	72	37	64	56	63	96	92	65	58	96	64
10	IS IT HEAVY?	94	55	64	30	57	88	95	67	54	57	69	47	60	53	87	81
11	CAN IT BEND?	62	66	76	72	54	75	67	51	74	75	49	88	70	67	52	38
12	CAN IT BREAK?	85	49	82	60	37	66	52	46	52	90	57	82	75	64	42	73
13	DOES IT HAVE PARTS?	67	59	98	36	31	88	22	28	57	98	86	66	87	72	100	95
14	DOES IT HAVE MOVING PARTS?	91	52	88	90	98	78	80	62	81	73	66	47	95	53	43	98
15	DOES IT CONTAIN SOMETHING ELSE?	50	56	83	53	71	57	45	90	61	56	53	85	72	70	91	53
16	DOES IT HAVE INTERNAL STRUCTURE?	60	87	96	65	81	93	58	44	86	68	43	90	77	65	99	85
17	DOES IT HAVE AT LEAST ONE HOLE?	64	49	77	75	85	94	98	70	70	98	75	84	88	66	89	86
18	IS IT ALIVE?	90	61	77	75	88	75	99	39	71	80	63	76	68	98	57	56
19	WAS IT EVER ALIVE?	94	64	61	79	89	67	70	80	78	73	82	86	92	99	25	69
20	IS IT MANUFACTURED?	96	94	67	92	92	72	24	77	89	61	92	73	63	99	7	26
21	WAS IT INVENTED?	91	61	59	73	91	55	96	95	52	64	75	75	75	97	47	36
22	WAS IT AROUND 100 YEARS AGO?	91	70	66	60	89	65	45	60	87	54	57	73	68	85	92	73
23	ARE THERE MANY VARIETIES OF IT?	91	25	61	67	85	75	94	68	76	69	67	56	55	98	60	60
24	DOES IT COME IN DIFFERENT SIZES?	86	51	56	65	79	46	35	74	72	76	80	66	50	98	51	23
25	DOES IT GROW?	90	66	81	85	82	84	85	89	75	70	70	78	87	99	63	56
26	IS IT BIGGER THAN A LOAF OF BREAD?	74	51	82	84	52	80	62	94	58	84	91	53	89	53	99	94
27	IS IT BIGGER THAN A MICROWAVE OVEN?	89	57	97	32	84	92	40	73	63	59	68	90	78	55	80	65
28	DOES IT HAVE A TAIL?	88	61	92	54	87	77	70	48	69	54	57	86	94	95	93	97
29	DOES IT HAVE LEGS?	88	62	92	83	96	79	97	53	79	98	83	89	95	79	93	98
30	DOES IT HAVE FOUR LEGS?	89	56	90	58	76	73	58	40	67	81	56	93	92	95	93	52

31	DOES IT HAVE FEET?	89	56	90	57	76	73	58	40	67	99	56	93	92	95	93	52
32	DOES IT HAVE PAWS?	91	47	86	66	79	71	60	37	69	25	71	93	56	95	45	74
33	DOES IT HAVE CLAWS?	94	67	94	68	81	87	77	73	76	86	60	89	57	97	67	44
34	DOES IT HAVE HOOVES?	95	67	97	83	87	92	93	72	79	66	55	73	97	95	94	73
35	DOES IT HAVE A FACE?	89	56	90	57	76	73	58	40	67	99	59	93	92	95	93	52
36	DOES IT HAVE A BACKBONE?	53	62	92	83	96	79	97	53	79	98	57	89	95	79	93	98
37	DOES IT HAVE EARS?	89	56	90	57	76	73	58	40	67	99	59	93	92	95	93	52
38	DOES IT COME FROM A PLANT?	69	80	75	44	41	74	53	55	41	73	57	50	78	71	99	83
39	DOES IT HAVE SOME SORT OF NOSE?	94	62	94	74	90	84	77	47	81	95	71	86	90	97	90	86
40	DOES IT CONTAIN LIQUID?	96	58	34	72	95	71	75	38	77	98	83	92	73	96	57	25
41	DOES IT MAKE A SOUND?	93	88	72	80	95	89	99	58	62	64	58	54	81	67	71	44
42	CAN IT RUN?	89	56	90	57	76	73	58	40	67	99	59	93	92	95	93	52
43	IS IT FAST?	89	80	90	80	87	90	98	43	54	61	81	50	55	97	59	46
44	CAN IT JUMP?	94	76	96	71	87	84	84	53	61	91	65	86	92	97	50	47
45	CAN IT BITE OR STING?	95	66	93	73	87	83	85	50	81	96	71	84	91	80	41	45
46	IS IT WILD?	98	80	97	79	95	94	37	62	89	65	73	79	51	98	39	69
47	IS IT WARM BLOODED?	89	56	90	57	76	73	58	40	67	99	59	93	92	95	93	52
48	IS IT A MAMMAL?	92	60	90	74	95	83	85	42	76	89	70	87	94	97	91	92
49	IS IT CONSCIOUS?	95	66	93	73	87	83	85	50	81	96	71	84	91	97	84	77
50	DOES IT HAVE FEELINGS?	96	69	96	79	93	86	83	42	83	67	75	84	93	98	52	56
51	DOES IT CAST A SHADOW?	51	46	99	56	76	62	91	57	54	25	86	46	70	49	99	68
52	DO YOU SEE IT DAILY?	98	69	67	63	83	75	42	75	70	61	81	62	70	99	17	90
53	IS IT HELPFUL?	94	53	53	71	46	71	57	54	58	73	98	55	52	99	83	68
54	CAN YOU TOUCH IT?	96	81	95	86	66	95	97	85	93	47	41	94	74	87	93	60
55	CAN YOU HOLD IT?	83	61	96	93	93	92	86	76	86	61	60	84	58	75	69	64
56	CAN YOU HOLD IT IN ONE HAND?	95	82	72	39	95	83	94	75	75	56	65	54	71	72	56	62
57	DO YOU HOLD IT TO USE IT?	59	55	29	81	53	62	70	75	64	73	93	57	74	56	98	96
58	CAN YOU USE IT?	96	73	51	69	96	86	47	48	82	67	59	86	52	99	18	86
59	CAN YOU USE IT UP?	66	50	98	59	73	60	94	65	57	62	90	51	64	56	97	44
60	CAN YOU PICK IT UP?	87	88	98	63	96	83	98	93	93	52	60	91	61	74	75	72

61	CAN YOU CONTROL IT?	84	63	80	74	73	55	39	61	59	58	53	57	67	95	46	85
62	CAN YOU SIT ON IT?	70	46	88	75	77	52	97	85	58	44	91	41	52	90	57	52
63	CAN IT BE WASHED?	48	84	78	66	89	75	71	89	58	37	49	77	54	91	87	43
64	IS IT HARD TO CATCH?	96	66	97	87	88	88	91	64	50	65	77	49	56	98	47	53
65	CAN IT BE EASILY MOVED?	73	38	34	32	57	43	37	95	88	47	39	81	37	75	30	21
66	DOES IT GO IN YOUR MOUTH?	84	87	99	87	75	56	84	58	70	63	95	78	93	94	100	41
67	IS IT TASTY?	95	90	84	76	97	42	98	61	81	56	97	96	57	96	46	77
68	IS IT USED DURING MEALS?	85	57	97	88	70	46	99	68	90	54	51	96	65	53	98	94
69	IS IT USUALLY OUTSIDE?	82	56	86	82	49	63	77	50	43	98	51	42	69	56	90	65
70	WOULD YOU FIND IT ON A FARM?	80	68	43	69	64	74	93	64	72	68	63	58	84	70	14	49
71	WOULD YOU FIND IT IN A ZOO?	92	88	96	95	76	81	93	49	73	80	57	51	65	89	49	66
72	WOULD YOU FIND IT IN A RESTAURANT?	51	71	94	82	62	55	32	47	66	55	59	94	53	61	49	94
73	WOULD YOU FIND IT IN A HOUSE?	58	89	70	82	55	85	69	68	58	93	53	55	47	68	95	57
74	DOES IT LIVE ABOVE GROUND?	72	43	64	76	56	97	94	79	91	76	72	56	46	95	42	55
75	DOES IT GET WET?	48	69	99	68	44	44	23	40	57	64	46	55	66	51	41	70
76	CAN IT LIVE OUT OF WATER?	85	39	36	66	93	80	99	95	90	100	98	93	97	73	100	48
77	DO YOU TAKE CARE OF IT?	71	74	64	74	45	70	49	70	53	55	64	41	60	75	94	65
78	DOES IT MAKE YOU HAPPY?	47	66	50	62	51	57	87	90	59	85	85	68	86	50	88	67
79	DO YOU LOVE IT?	60	47	42	48	72	57	76	84	58	45	82	79	87	61	90	47
80	WOULD YOU MISS IT IF IT WERE GONE?	57	62	57	80	48	50	82	73	67	57	99	52	72	59	95	36
81	CAN YOU BUY IT?	94	72	98	79	94	88	64	57	84	68	56	80	46	95	59	94

Table 2 Classification error for 81 attributes and 20 test classes

VII. CONCLUSION

Zero shot learning is an attribute classification method based on neural network features, semantic attribute mapping and support vector machines. In this experiment, the indirect attribute prediction method was used in order to classify unknown objects and note which attributes produced the least amount of error. The results confirmed the hypothesis that visual attributes produce less error than non visual attributes. In further research, in order to decrease the amount of training data and the amount of questions used to classify unknown objects, systems should be trained to identify visual attributes which will decrease the error of classification, and will also aid in more effective classification of unknown objects.

VIII. Work Cited

Bishop, Christopher M. 2007. ‘Pattern Recognition and Machine Learning’. Chapter 7. Sparse Kernel Machines, pages 325-339, Published by Springer, 2007.

Boyd, S., and Vandenberghe, V., “Convex Optimization,” Chapter 8.6. Classification, pages 422 – 431. Cambridge University Press, 2004

Burlina, Philippe M., Schmidt, Aurora C., Wang, I-Jeng Wang “Zero Shot Deep Learning from Semantic Attributes”. IEEE International Conference on Machine Learning and Applications, Dec 2015

Fan, R., Chang, K., Hsieh, C., Wang, X., and Lin, C., “LIBLINEAR: A library for large linear classification,” *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

Farhadi, A., Endres, I., Hoiem, D. and Forsyth, D. Describing objects by their attributes. In Conference on Computer Vision and Pattern Recognition, 2009.

Hermans, T., Rehg, J.M, Bobick, A., "Affordance prediction via learned object attributes"International Conference on Robotics and Automation: Workshop on Semantic Perception, Mapping, and Exploration, 2011.

Lampert, C.H., Nickisch, H., and Harmeling, S., “Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer,” Procedings IEEE Conference Computer Vision and Pattern Recognition (CVPR), 2009

Lampert, C.H., Nickisch, H., and Harmeling, S., Attribute-Based Classification for Zero-Shot Visual Object Categorization. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No.3, March 2014

Olga R., and Li Fei-Fei, “Attribute learning in large-scale datasets,” in Trends and Topics in Computer Vision, pp. 1–14. Springer, 2012.

Platt, J. C., “Platt Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods”, Advances in large margin classifier, Vol. 10, no 3, pp. 61-74, 1999

Palatucci, M., Pomerleau, D., Hinton, G., Mitchell, T., “Zero-Shot Learning with Semantic Output Codes”. *Neural Information Processing Systems (NIPS*), Dec, 2009

Parikh. D., and Grauman, K., “Interactively Building a Discriminative Vocabulary of Nameable Attributes,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1681-1688, 2011.

Razavian, A. H., Azizpour, H., Sullivan, J., and Carlsson, S. "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition", *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops* pp. 512-519, 2014

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., and Fergus, R., “Overfeat: Integrated recognition, localization and detection using convolutional networks”.,arXiv preprint arXiv:1312.6229, 2013

Socher, R., Ganjoo, M., Manning, C.D., Ng, A. J., “Zero-Shot Learning Through Cross-Modal Transfer”, Advances in Neural Information Processing Systems 26 (NIPS 2013)

IX. Acknowledgements

I would like to thank Dr. I-Jeng Wang my mentor for his technical input and help through the project. I thank Dr. Burlina for his technical advice and help. I thank Ms. Schmidt for help in set up and helping me with MATLAB questions. I thank my intern-mentor teacher Ms. Sasser and River Hill High School for permission to do research at APL