

Detecting Strange Objects via Visual Attributes

Babak Saleh ^{*}, Ahmed Elgammal ^{*}, Ali Farhadi ^{**}

^{*} Rutgers University ^{**} University of Washington

Abstract. We are capable of developing algorithms for detecting objects by learning a comprehensive object model based on their parts and visual attributes. These models usually fail to perform well on atypical images, however people are still able to recognize strange objects that are not completely following our model. In this work we focus on the question of what makes an object looks strange? We would like to investigate what is the main characteristics of an object category in terms of its visual attributes. Later we detect meaningful deviations from these expectations as atypical cases. We present interesting findings on the novel dataset of abnormal objects and show how we can improve object detectors by making assumptions on typicality of the objects.

Keywords: Visual Attributes, Object Recognition, Object Description, Anomaly Detection, Typicality Estimation.

1 Introduction

The variability between members of a category influences infants' category learning. 10-months-old infants can form a category structure and distinguish between category prototypes and atypical examples [4]. 14-months-olds use properties of objects to report atypical instances [2]. Wouldn't it be nice to have a recognition system that achieves exact same capability? In computer vision, there has been significant progress in forming the category structures. However, little attention has been paid to deviations from prototypical examples of categories. This paper is centered on modeling the typicalities from categories to be able to reason about abnormalities. Inspired by infant category learning, we propose to learn the structure of typical images using their attributes and then recognize abnormalities as special deviations from prototypical examples. Similar to infants' learning, we want to reason about abnormalities by only observing typical instances. Taxonomies of abnormalities are not known. This makes defining any fixed vocabulary for abnormalities unjustifiable. We believe that any reasoning about abnormalities should be based on understandings of normalities and should not require any observations about abnormal instances.

There has been recent interest in investigating what should be reported as an output of a recognition system [1]. When describing an image, humans tend not to mention the obvious (simple category memberships) instead to report what is worth mentioning about an image. We argue that abnormalities are among major components that form what is worth mentioning. We want to

form category structures in terms of common attributes in categories and reason about deviations from categories using attributes.

A diverse set of reasons may cause abnormality. An object can be abnormal due to the absence of typical attributes (a car without wheels) or the presence of atypical attributes (a car with wings). Also, abnormality can be caused by deviations from the extent by which an attribute varies inside a category (a furry dog). Furthermore, contextual irregularities and semantical peculiarities can also cause abnormalities such as an elephant in the room.

There are some works that focus on the context around the object to recognize atypicality [6, 7]. Context in images play an important role for object detectors and classifiers, however the notion of context is a broad concept that incorporates many elements (e.g. scene category, relative size and location of the objects). Choi et al [6] addressed the problem of detecting out-of-context objects by learning a hierarchy of objects and their supported context. They learn a tree model based on normal (expected) scene layouts in the SUN dataset [9], whose nodes are objects, and two nodes are connected if the parent can support its children in the images. They gathered 40 images that the context do not match the appearing objects. These objects do not follow the assumptions in the tree and are categorized as out-of-context objects.

Park et al [7] approached the contextual abnormality detection in the scene by reasoning about the following reasons rather than strange combination of detected objects in a given scene: 1) Abnormal relative location of detected objects, where each of the objects in an image looks normal by itself, but their relative position to each other is a case of abnormality (e.g. one car on top of another); 2) Abnormal relative size of objects. For example a huge car in the parking lot looks abnormal comparing to other parked cars. However at the end they leave this reason untouched. We proposed a framework for reasoning about abnormalities stemming from objects by themselves. They detect what is wrong about an object in terms of its visual attributes (e.g., an airplane without the jet engine), and can report missing or unexpected attributes in abnormal objects [8]. Isola et al [3] studied image memorability and trained a classifier to predict memorable images. Memorability is related to abnormality, since atypicality is one of the possible reasons of memorability.

What does studying abnormality in images tell us about object recognition? While being slower, humans seem to be able to recognize abnormalities and reason about category memberships of atypical instances without learning on any atypical instance [5]. Can state-of-the-art computer vision object categorization and detection algorithms generalize as well to atypical images? We argue that studying generalization to atypical images, without optimizing on them, provides insights on how a recognition algorithm might simulate human performance. In addition, there are various applications for developing an intelligent system that can detect abnormalities. Certain types of abnormality in images can be an indication of abnormal event.

In this work we provide more details of the novel dataset of abnormal objects that we published recently [8], We proposed models for recognition of atypical

objects and conducted more experiments on detecting abnormalities in images of this dataset.

2 Model and Experiments

For the purpose of our study, we used dataset of Abnormal Objects [8], which has 617 images of abnormal objects from six object categories. These images are result of image search for strange keywords like “atypical car”, “strange boat”, etc. This dataset comes with some human subject experiments conducted using Amazon Mechanical Turk as following. Given an image with a bounding box about the most salient object, subjects were asked several questions. We collected 10 responses for each image. First the subjects were asked whether the image seems normal or abnormal. If the subject decides that the image is abnormal the following questions were asked where multiple selections are allowed: 1) Which category best describes the object, from a list of the six categories in our database. 2) Whether abnormality is because of the object itself or its relation to the scene. 3) To rate the importance of each of these following attributes in affecting their decision that this object is abnormal (Color, Texture/Material, Shape/Part configuration, Object pose/viewing direction) 4) Also the subjects were asked to comment about context abnormality if it is the case.

We approach the problem of atypicality detection by learning a model for typical objects and find meaningful deviations from this model as strange objects. As is depicted in figure 1, we can assume there is a boundary between normal object classes. As we move away from the center of each of these classes, we will have more atypical samples. This quasi model emphasis that some objects can be atypical samples of more than one object category. In order to model the typical objects we have tired following approaches.

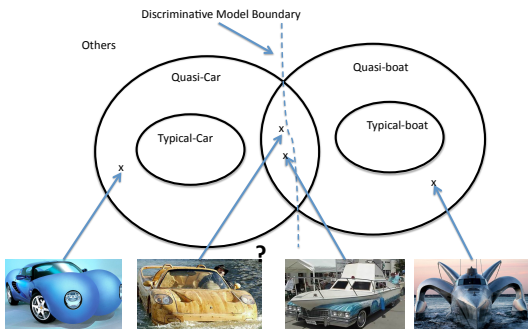


Fig. 1. There is a clear boundary between object categories while we cannot find a discriminative boundary between typical and atypical samples of one class. Normal samples of cars build a dense cluster while there are many reason that can make a car look strange. The variety of abnormality reasons results in a space of objects that are not typical and fall into a quasi space of object categories. This figure shows how a quasi car can be a quasi boat as well.

For modeling typicality we need to learn generative models in terms of the conditional class densities $p(x|T, c_k)$. We use an attribute space for that purpose, i.e. we need to model $p(A_1(x), \dots, A_M(x)|T, c_k)$, where M is the number of attributes. We investigated several models of typicality, which we will summarize in this section.

Naive Bayes' Model: In this approach we model the density

$$p(A_1(x), \dots, A_M(x)|T, c_k) = \prod_i p(A_i(x)|T, c_k)$$

where we use a Gaussian model for each attribute density:

$$p(A_i(x)|T, c_k) \sim \mathcal{N}(\mu_i^k, \sigma_i^{k,2}).$$

Non-parametric Model: In this approach we model each conditional class density using kernel density estimation, i.e., we achieve an estimate of the density in the form $\hat{p}(A_1(x), \dots, A_M(x)|T, c_k) = \frac{1}{N} \sum_{j=1}^N \prod_{i=1}^M g(A_i(x) - A_i(x_j))$, where $g(\cdot)$ is a kernel function and $\{x_j\}$ are training images of class k . Here we use the kernel product, which is typically used to approximate multivariate densities.

Modeling typicality manifold: In this approach we hypothesize that typical images lie on a low-dimensional manifold in the attribute space. We explicitly model that typicality manifold for each category and compute deviation from abnormality by modeling the distance of a test image to that manifold. Given a test image we find its nearest neighbor from the training data of a given category and then compute the perpendicular distance to the tangent space of the manifold at that point. This can be achieved by projecting the test image to a local subspace for the manifold patch around the nearest neighbor point. There are two probability models for the distance to the manifold that we investigated: 1) a global Gaussian model for the whole manifold, 2) a local Gaussian model at each patch of the manifold. There are two parameters for this model, the patch size, k and the local subspace dimensionality d .

Manifold-based density model: This approach is similar to the Naive Bayes' Model, however instead of computing the densities $p(A_i(x)|T, c_k)$ globally, these densities are computed locally for patch of the typicality manifold. The rationale is each part of the typicality manifold is expected to have different distribution.

One-class SVM: One-class SVM is typically used for estimating regions of high density. Given typical examples for each class in the attribute space, one-class svm is used to estimate a boundary of volume of high density, which is then can be used to detect deviation.

References

1. A. Farhadi, S. M. M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. A. Forsyth. Every picture tells a story: Generating sentences from images. 2010.
2. Ann E.; Samuelson Larissa K.; Trejo Erika; Worzalla Samantha L.; Peltan Jessica R.; Oakes Lisa M. Horst, Jessica S.; Ellis. *Developmental Science*, v12.
3. Phillip Isola, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva. What makes a photograph memorable? *PAMI*, 2013.

4. Plunkett K. Mather E. *Cognition*.
5. Mark T. Keane Michael W. Eysenck. *Cognitive psychology: a student's handbook*. Psychology press, 2005.
6. Alan S. Willsky Myung Jin Choi, Antonio Torralba. Context models and out-of-context objects. In *To appear in Pattern Recognition Letters*, 2012.
7. Sangdon Park, Wonsik Kim, and Kyoung Mu Lee. Abnormal object detection by canonical scene-based contextual model. In *ECCV*, 2012.
8. Babak Saleh, Ali Farhadi, and Ahmed Elgammal. Object-centric anomaly detection by attribute-based reasoning. In *CVPR*, pages 787–794, 2013.
9. Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, pages 3485–3492, 2010.