

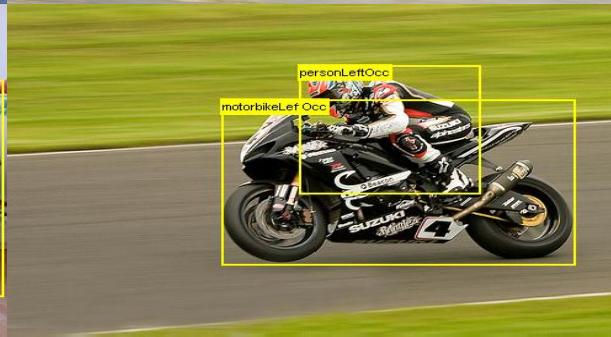
The Role of Typicality in Object Classification

Babak Saleh

Outline

- Motivation
- Related Work
- Problem Definition
- Estimating Typicality of objects/images
 - Data Collection
 - Computational Models
 - Abnormal Object Detection
 - Taxonomy of Abnormalities in Images

Motivation



Motivation



Motivation



Despite human vision, CNN models categorize abnormal objects with a high confidence.

It looks weird!!



It looks weird!!
However it looks like a Car and a Boat



It looks weird!!
However it looks like a Car and a Boat
Boat with a wheel??
Car with a bow?



Motivation

- State-of-the-art (CNN based models) fail drastically to recognize abnormal objects.

Method	Top-1 error (%)			Top-5 error (%)		
	Train	Test-T	Test-A	Train	Test-T	Test-A
AlexNet [Krizhevsky <i>et al.</i> , 2012]	38.1	49.5	74.96	15.32	24.01	47.07
OverFeat [Sermanet <i>et al.</i> , 2013]	35.1	45.36	75.62	14.2	22.27	46.73
Caffe [Jia <i>et al.</i> , 2014]	39.4	51.88	77.12	16.6	24.74	46.86
VGG-16 [Simonyan and Zisserman, 2015]	30.9	44.04	77.82	15.3	26.31	47.49
VGG-19 [Simonyan and Zisserman, 2015]	30.5	43.72	76.35	15.2	26.85	45.99

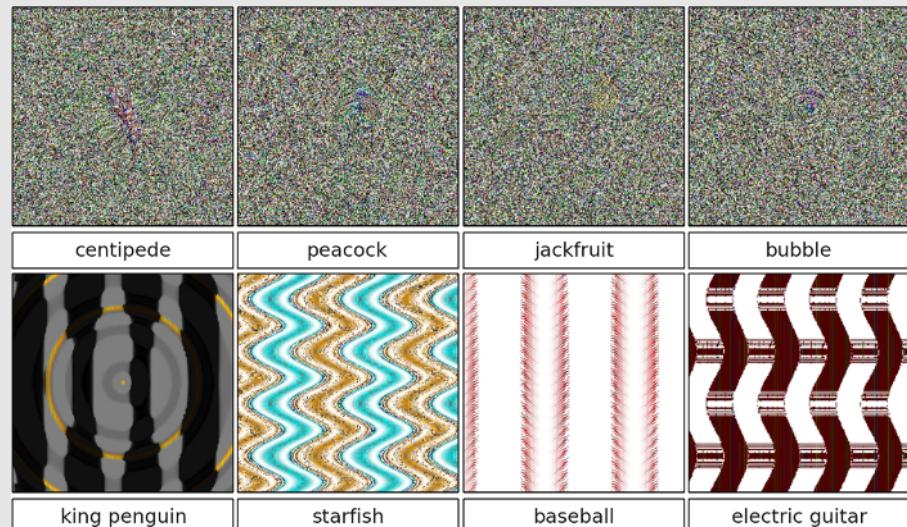


Related Work

- Analysis of failure modes for CNN

- [Goodfellow et al.'15]

$$\begin{array}{ccc} \text{panda} & + .007 \times & \text{panda} \\ x & \text{sign}(\nabla_x J(\theta, x, y)) & = \\ \text{"panda"} & \text{"nematode"} & \text{"gibbon"} \\ 57.7\% \text{ confidence} & 8.2\% \text{ confidence} & 99.3 \% \text{ confidence} \end{array}$$



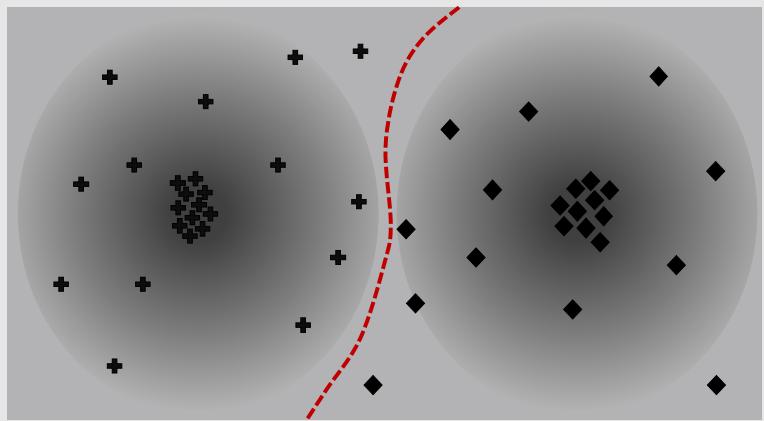
- [Nguyen et al.'15]

Related Work

Training strategies for learning:

- Technical specifications
 - Learning rate
 - Batch size
- Curriculum learning
 - Order of the training samples are important.
 - Better classifiers are trained with seeing better examples first

- Discriminative Models
 - Boundary examples
- Generative Models
 - Center of density
- Hybrid Models
 - Using generative hints in training discriminative models. Best of both worlds.
 - Not all abnormal samples are boundary examples.



Problem Definition

- Improving generalization capacity of discriminative classifiers using signals from a generative model that focuses on representativeness of samples.
- In this work, we take most normal/typical examples as most representative samples.

What makes an object abnormal?

- What is abnormal object/image?
 - Not a well-agreed answer.
- What makes an object/image look strange?
 - Shape, parts, texture.
- Why abnormality happens?
 - Unexpected context, relative size, strange location.



Data Collection

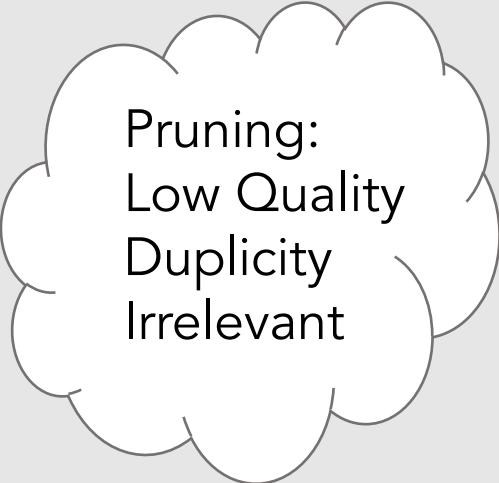
Abnormal
Weird
Strange
Atypical

Object
class
+names
from
Pascal

Weird Car, Strange Sofa, etc.



Google™
YAHOO!
bing™



Pruning:
Low Quality
Duplicity
Irrelevant



Human Subject Experiment



amazon®
mechanical turk
beta

Abnormal Image Dataset



Human Subject Exp. I

- The goal of the first round of our human subject experiment is manifold:
 1. Verifying that our image collection is abnormal.
 2. Abnormality rating for objects as a graded concept.
 3. Delving into reasons that cause abnormality in objects.
 4. Rich set of annotation in terms of object categories and abnormality reasons.

1. Do the objects in the image seem normal or abnormal (i.e. strange) to you?: (select only ONE of the following)

- Normal
- Abnormal

2. Which of the following categories best describe the object? (select all that apply):

- Aeroplane
- Boat
- Car
- Chair
- Motorbike
- Sofa
- None of them

Showing HIT 1 of 1000

[Next HIT](#)

[Cancel](#)

[Next](#)

3. Why do you think the object in the image is abnormal? (select all that apply)

- Object by itself
- Object in relation to the other scene parameter

If you selected "Object by itself" for abnormality in question 3: (If this part doesn't apply to this image, please select "Doesn't apply")

4. Please rate importance of each of these following attributes in affecting your decision that this object is Abnormal. You are allowed to assign the same rating to more than one attribute.

Attribute name	Normal(rate 1)	somewhat Abnormal or Strange (rate 2)	Abnormal or Unusual (rate 3)	Doesn't apply
Color	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shape (including Object size, parts configuration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Texture / Material	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Object pose / Viewing direction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Showing HIT 1 of 1000

[Next HIT](#)

[Cancel](#)

[Next](#)

If the abnormality comes from "Object in relation to the scene parameter" (in question 3) or in case that you think abnormality is a result of something that is missing in questions above:

5. Please describe why you think this image is Abnormal by characterizing the Abnormality aspects:

some of Abnormality aspects are:

- 1) Geometrical relationships with other objects in the scene. (for example: a train on top of a car)
- 2) Semantically inappropriate. (for example: an elephant inside the living room)
- 3) Size of the objects with respect to other objects on the scene



Showing HIT 1 of 1000

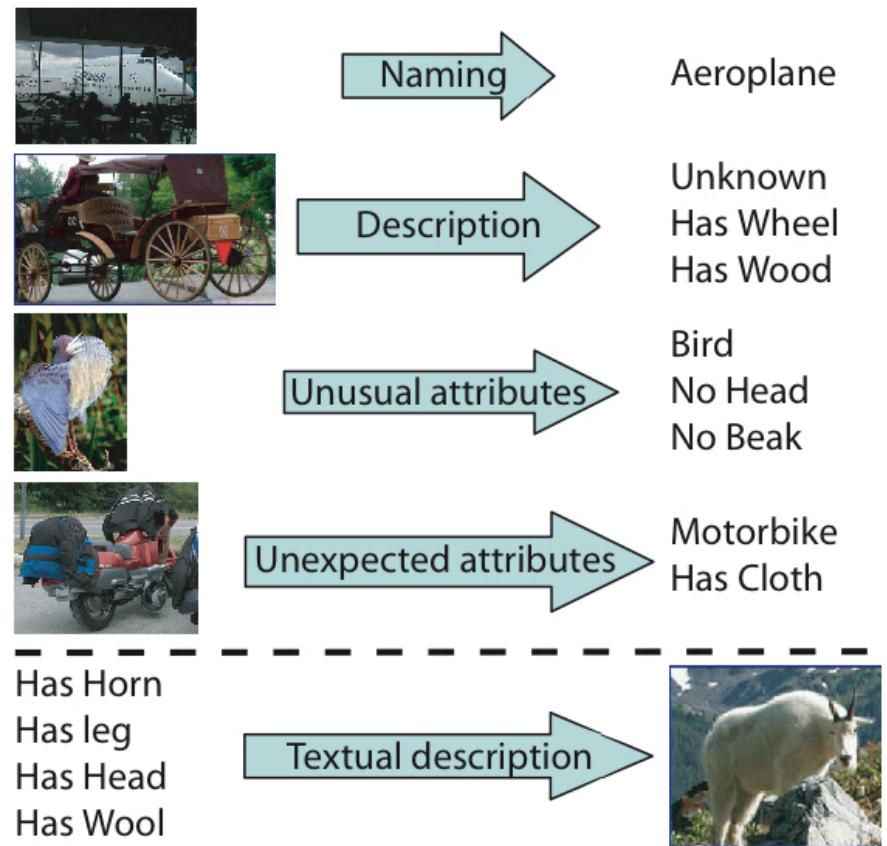
Abnormal Object Detection

- We should NOT learn on top of abnormal samples
 - It is not a two way classification problem.
 - Imbalanced data
 - Uncontrolled variations in the visual space
 - Human-like vision
- We just learn the model based on normal samples and define:

“Abnormality is Meaningful Deviation from Normality”

Visual Attribute

- Semantically meaningful visual characteristics. [Farhadi et al.'09]
- Pre-defined list of attributes, usually require expert knowledge.
- Visual classifiers for each attribute, trained on low-level visual features (HoG, SIFT, etc.)
- Shareable between object categories. Normal and Abnormal objects.



Abnormal Object Detection

Graphical Model for Normality (N)
C: object **C**ategory, **A:** visual **A**ttributes

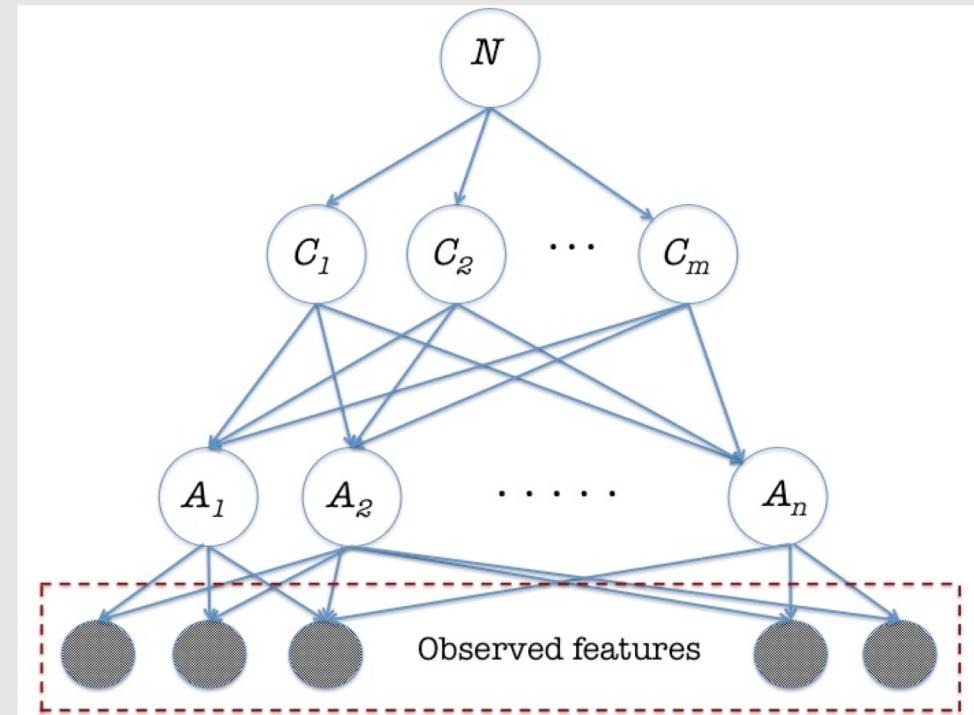
$$P(\neg N|A) = 1 - P(N|A)$$

$$P(N|A) = P(A|N) * P(N)/P(A)$$

$$P(A|N) = \sum_j P(A|C_j, N)P(C_j|N)$$

$$P(A|N) = \sum_j \prod_{i=1}^k P(A_i|C_j)P(C_j|N)$$

$$P(A_i|C_j) \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$$



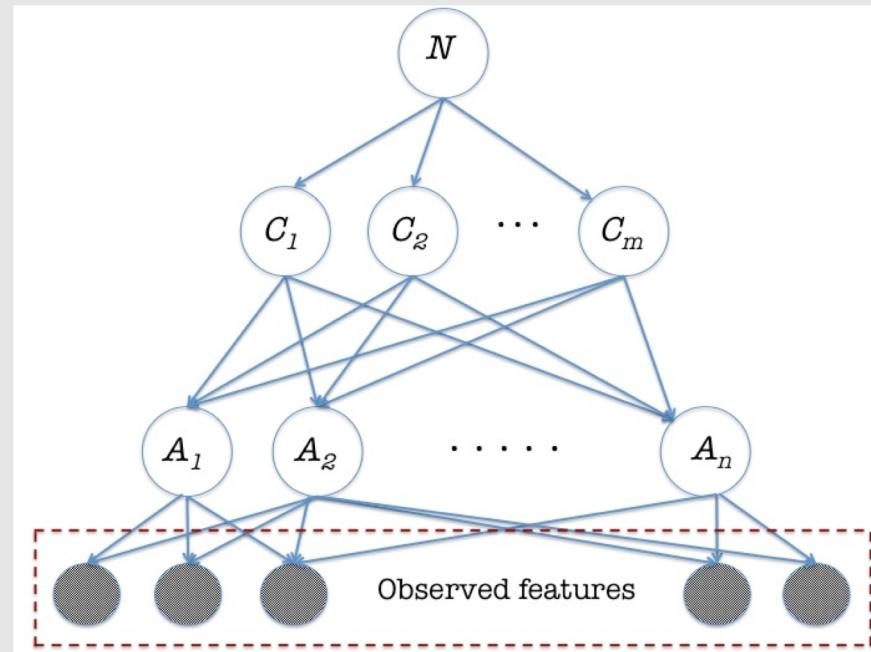
Abnormal Object Detection

Surprise Score for Abnormality

- The amount of information produced by an attribute is different from one category to another.
 - We measure it via information theory

$$I(A_i = a | C_j, N) = -\log P(A_i = a | C_j, N)$$

- Attribute classifiers are noisy!
 - Measure reliability of attributes as its accuracy over validation set.



Abnormal Object Detection

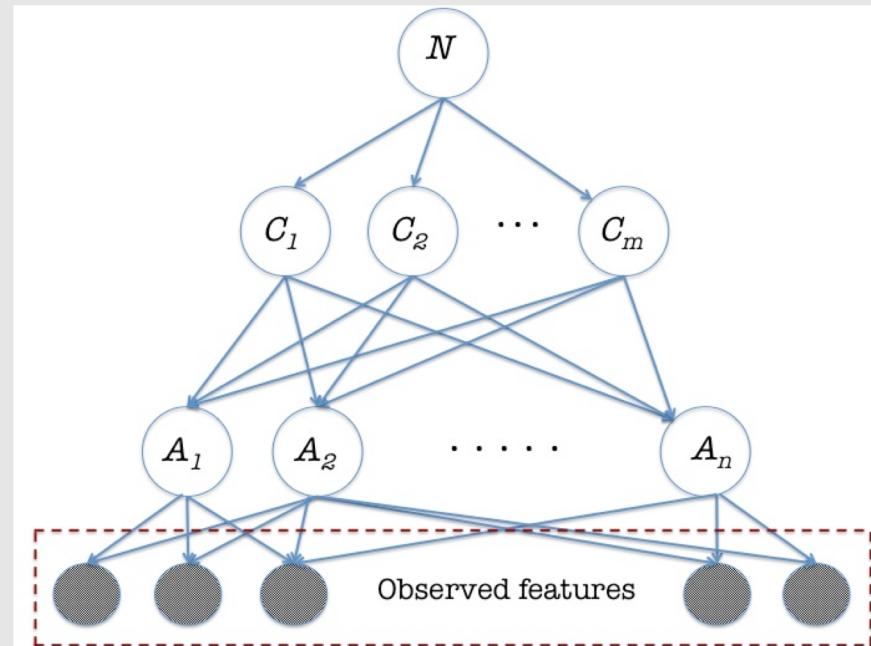
Surprise Score for Abnormality

- Importance of visual attributes is dependent on the object category.
- Compute relevance of an attribute over training set once.

$$\text{relevance}(A_i|C_j) = 1/H(A_i|C_j)$$

- Surprise score is aggregation over all attributes and object categories.

$$\text{surprise}_{(A_i|C_j)}(a) = \text{reliability}(A_i) * I(A_i = a|C_j) * \text{relevance}(A_i|C_j)$$



Abnormality beyond Objects

- Comprehensive understanding of what can cause abnormality? Going beyond abnormal objects
- However, the taxonomy of possible reasons for abnormality in images is not defined
- Still, this is Not a two-way classification
- We use Information theoretic approaches, because it mimics human reasoning about these images

Human Subject Exp. II

- Investigating human thinking about abnormality
- Proposing a taxonomy of reasons that can make an image look abnormal.

HIT Preview



After looking at this image tell us if you think this image looks typical to you or not by selecting one of the following options:

- Nothing abnormal about this image
- Strange contour (boundary) for typical objects of that category
- Other:

- *Nothing abnormal about this image
- *Strange contour (boundary) for typical objects of that category
- *Object in the shape of another object
- *Object is not complete
- *object is not complete
- *missing specific part(s)
- *misplaced part(s)
- *weird-shaped part(s)
- *unexpected parts, extra parts
- *strange texture
- *strange color
- *strange pattern
- *atypical material
- *something wrong with the shape of the object, but I cannot easily name it
- *atypical combination of categories
- *strange body posture

Showing HIT 1 of 1000

[Next HIT](#)

[Cancel](#)

[Next](#)

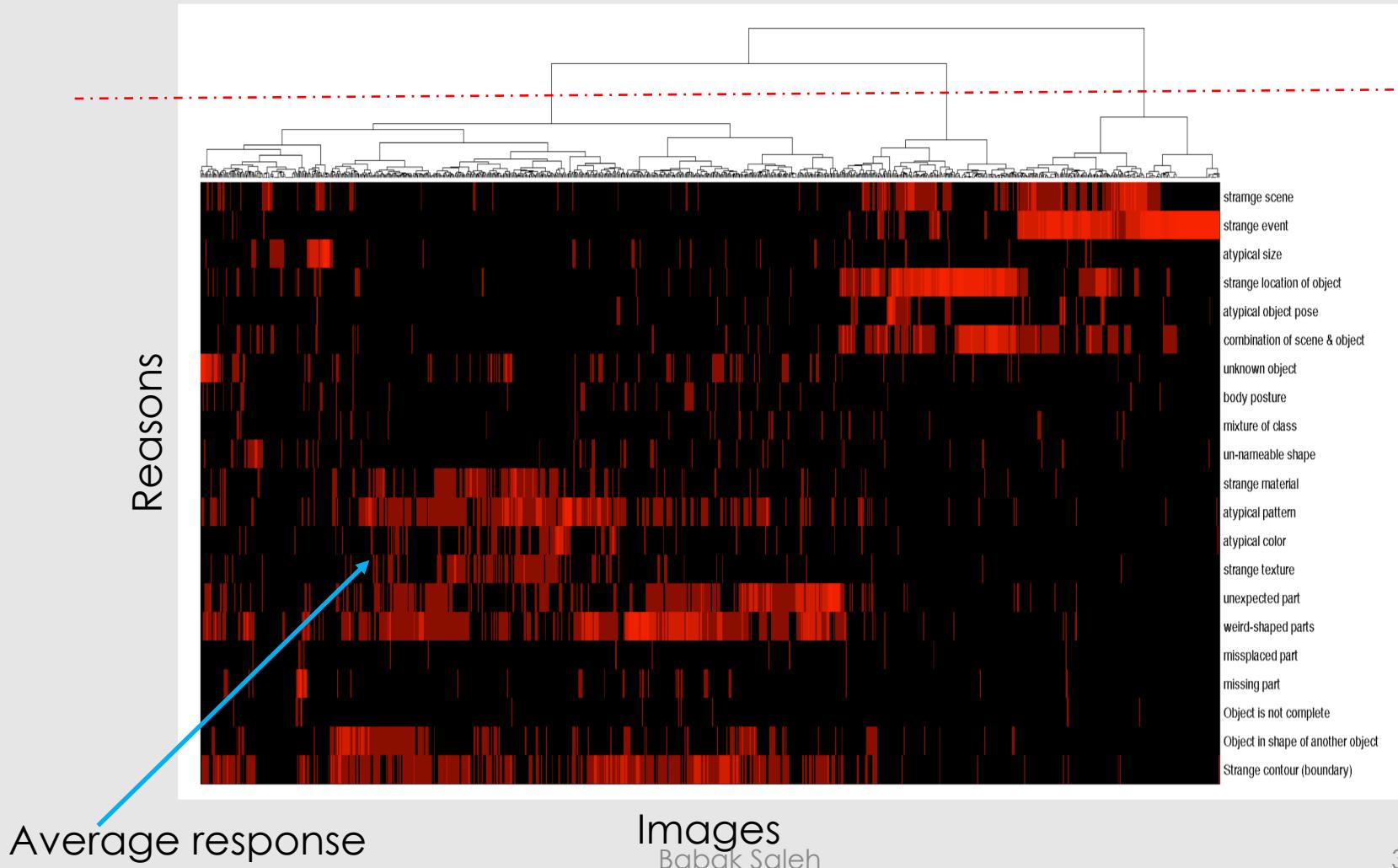
HIT Preview

- Atypical material
- Something wrong with the shape of the object, but I cannot easily name it
- Atypical combination of categories
- Strange body posture
- This image is strange because I do not know what is this object
- Some objects typically do not belong to the scene
- Some objects are in atypical pose (e.g. upside down)
- Strange location for some objects in the scene
- Atypical size for some objects relative to the scene
- Strange event (accident, fire, etc.)
- Strange scene
- Weird but for a reason not in this list
- Weird for the following reason:

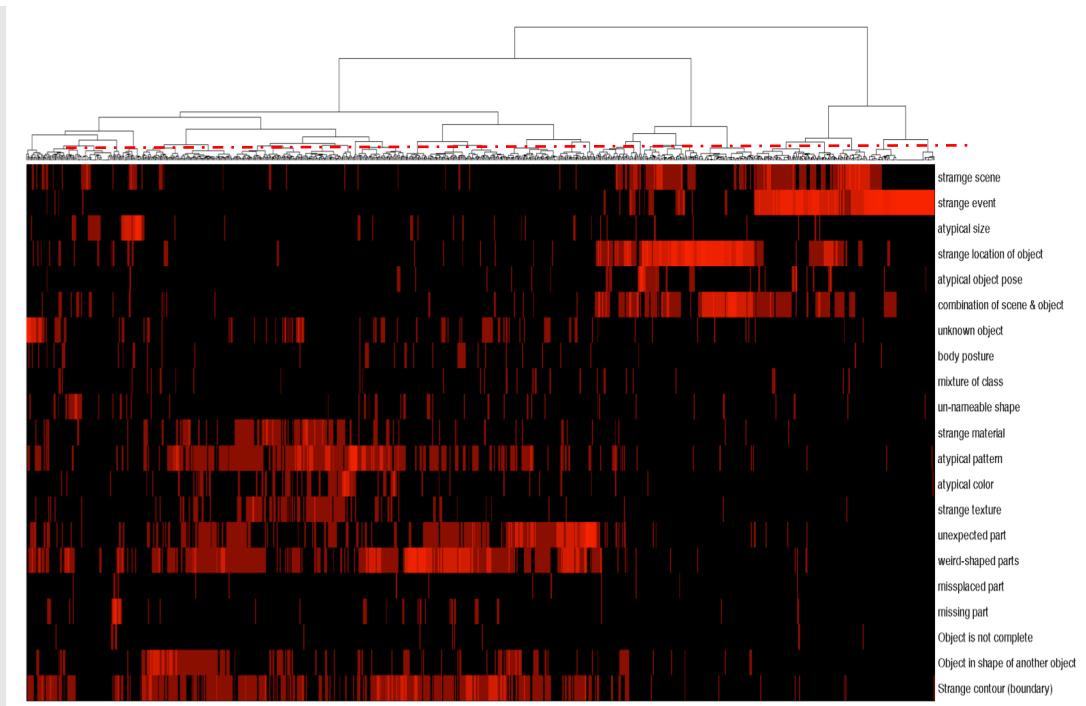
- *image is strange because I do not know what is this object
- *some objects typically do not belong to the scene
- *some objects are in atypical pose (e.g. upside down)
- *strange location for some objects
- *atypical size for some objects relative to the scene
- *strange event (accident, fire, etc.)
- *strange scene
- *weird but for a reason not in this list

Showing HIT 1 of 1000

- For all images, average responses of 10 users
- bottom-up agglomerative clustering
- Ward's minimum variance criteria as the Linkage function
- Three main branches on top



Main Category	Detailed Reasons in Amazon Mechanical Turk Experiment
Scene-centric	Strange event happening in the scene(21); Strange scene(20)
Contextual	Atypical object size(19); Strange location of the object(18); Atypical object pose(17); Weird combination of objects and scene(16)
Object-centric	Unexpected part(7); Weird shaped part(6); Misplaced part(5); Missing part(4); Body posture(14); Mixture of object classes(13); Un-nameable shape(12); Object is not complete(3); Unknown object(15); Object in the shape of another object(2); Atypical pattern(10), Weird color(9), Strange material(11), Weird texture(8); Strange object contour(1)



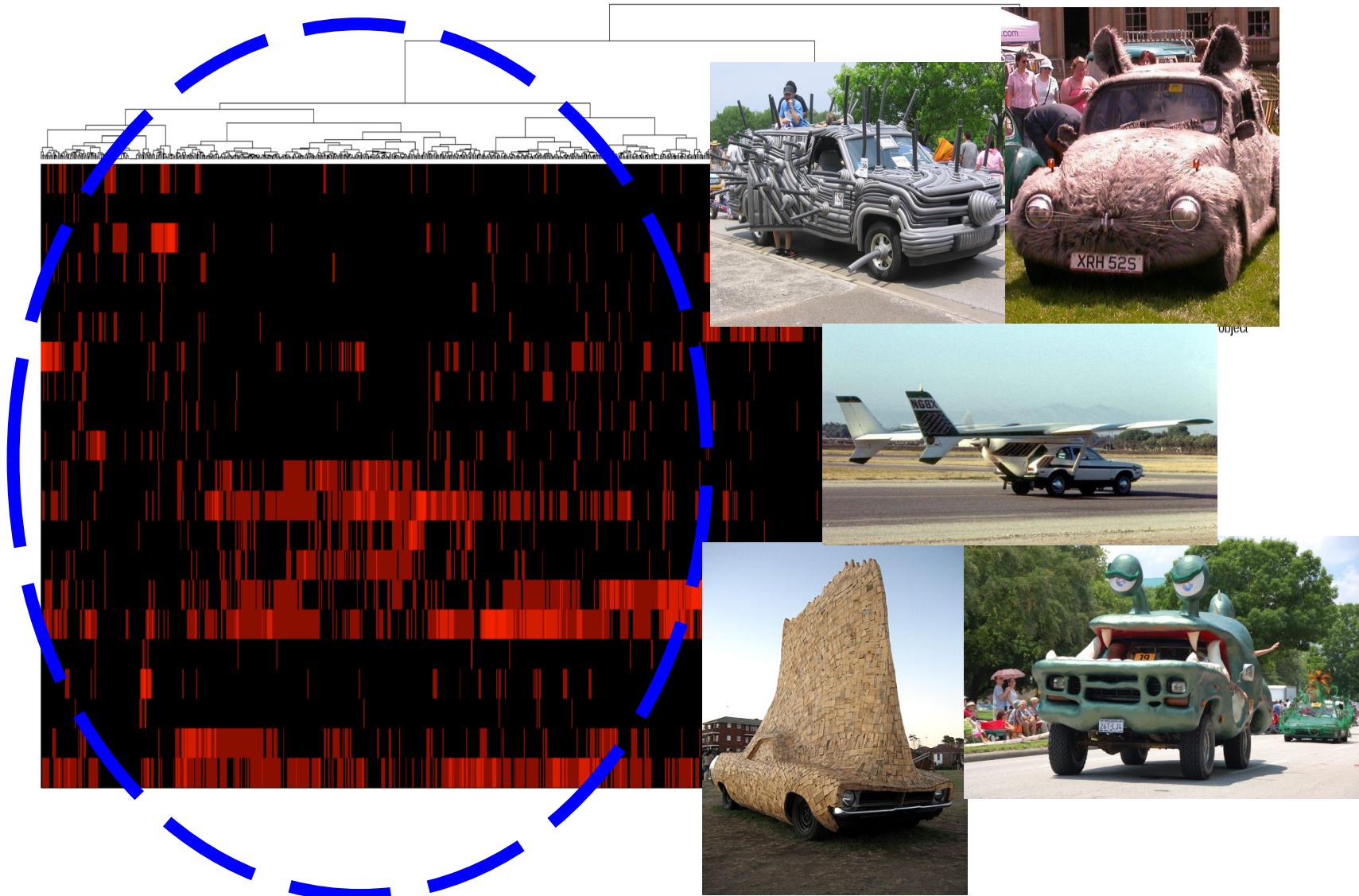
Scene-centric Abnormality



Contextual Abnormality



Object-centric Abnormality



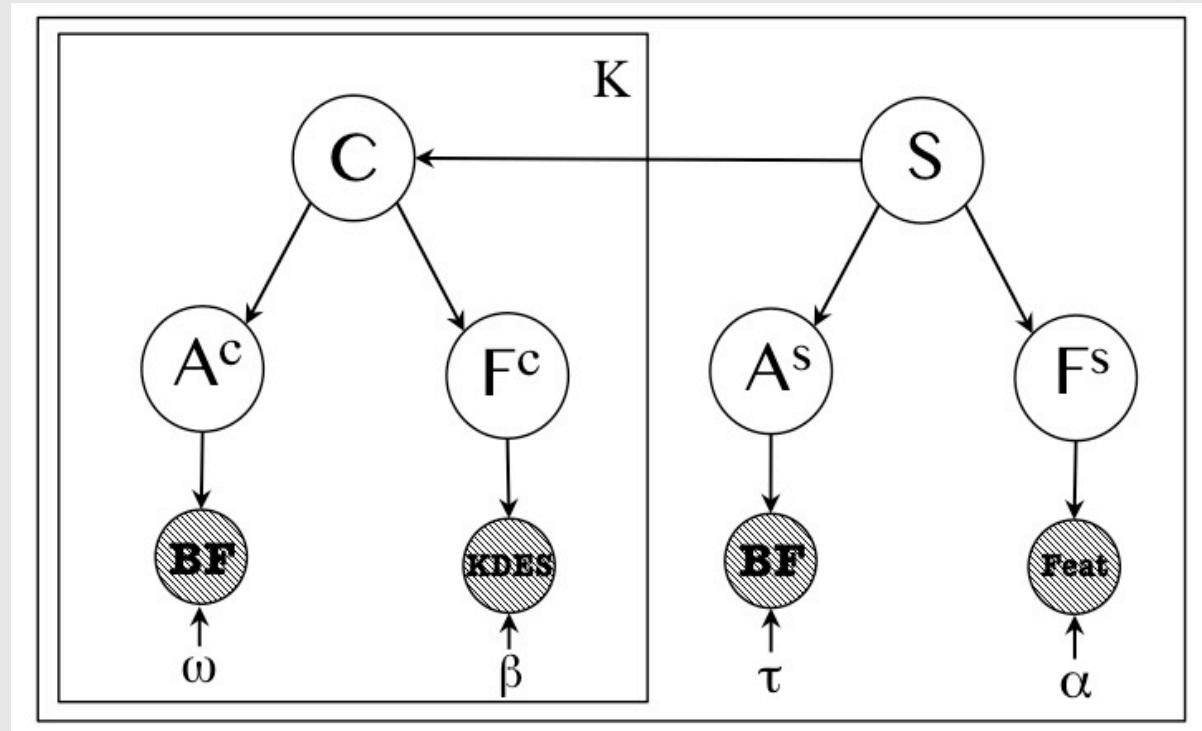
- Following the inferred taxonomy of abnormality, we build three computational models to measure how surprising is an image:
 - Object-centric surprise score
 - Scene-centric surprise score
 - Contextual surprise score

Modeling Normal Images

S: Scene category
C: Object category

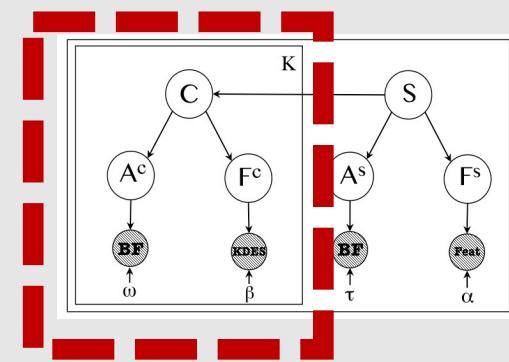
$F^S F^C$: Vis. Feat for
Scene/Object
categorization

$A^S A^C$: Vis.
Attributes for
Scene/Object



This model is able to find three
main reasons of abnormality

Object-centric Surprise



$$\sum_k P(O_k) * \left(\sum_i I(A_i^o | O_k) * \Upsilon(A_i^o) * \Omega(A_i^o, O_k) \right)$$

Object category

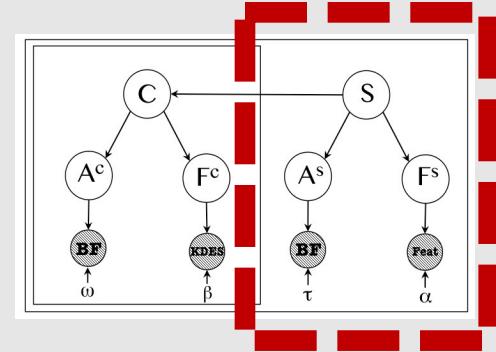
$$-\log(P(A_i^o | O_k))$$

Amount of surprise by observing i-th attribute for the given object category

Attribute reliability

Importance of i-th attribute for the given object category

Scene-centric Surprise



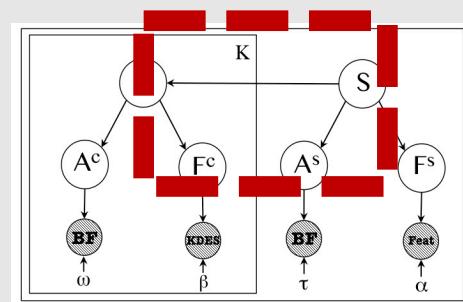
$$\sum_i I(A_i^s | S_j) \Upsilon(A_i^s) \Omega(A_i^s, S_j)$$

Amount of information by observing i-th attribute for the given scene category

Attribute reliability

Importance of i-th attribute for the given scene category

Contextual Surprise



$$\sum_k \sum_j \Lambda(O_k) [\hat{I}(O_k|S_j) + I(L_k|O_k)]$$

Relative size of the object
Modeled by Gamma Dist.

Amount of information
that location of the
object reveals

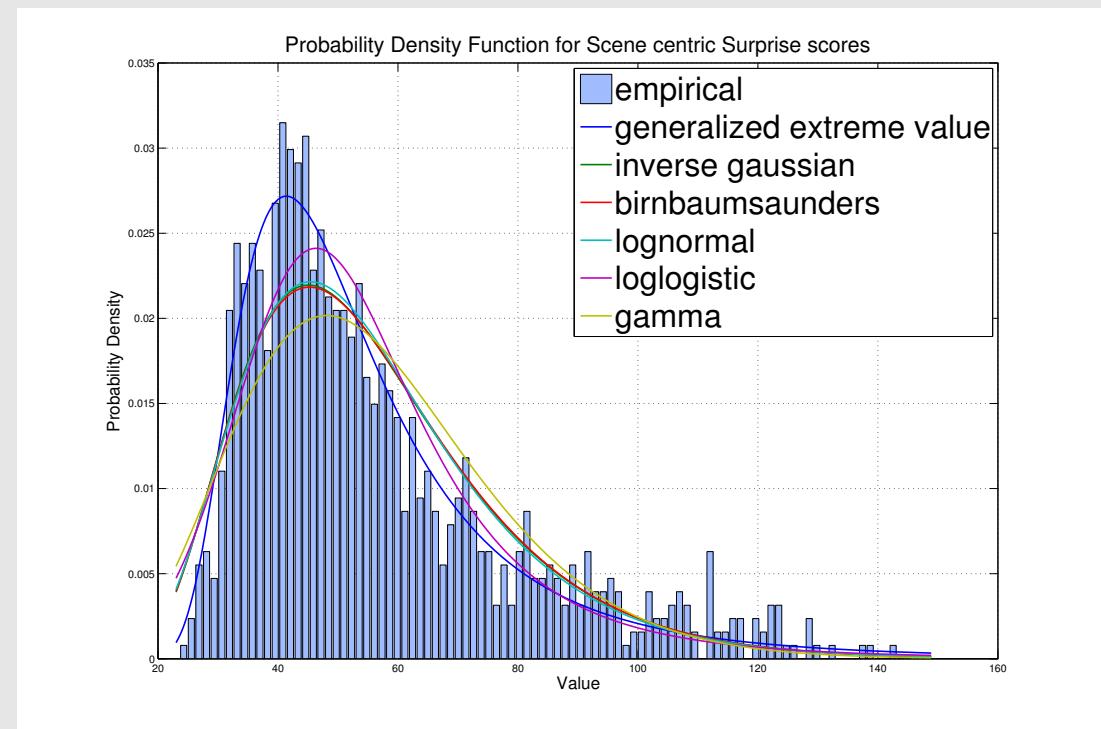
$$\hat{I}(O_k|S_j) = P(S_j)P(O_k)I(O_k|S_j)$$

Surprise rooted in the co-occurrence
of object categories and scene

Abnormality Scores

$$\operatorname{argmax}_{o,s,c} (\phi_o(\text{Surprise}_o), \phi_s(\text{Surprise}_s), \phi_c(\text{Surprise}_c))$$

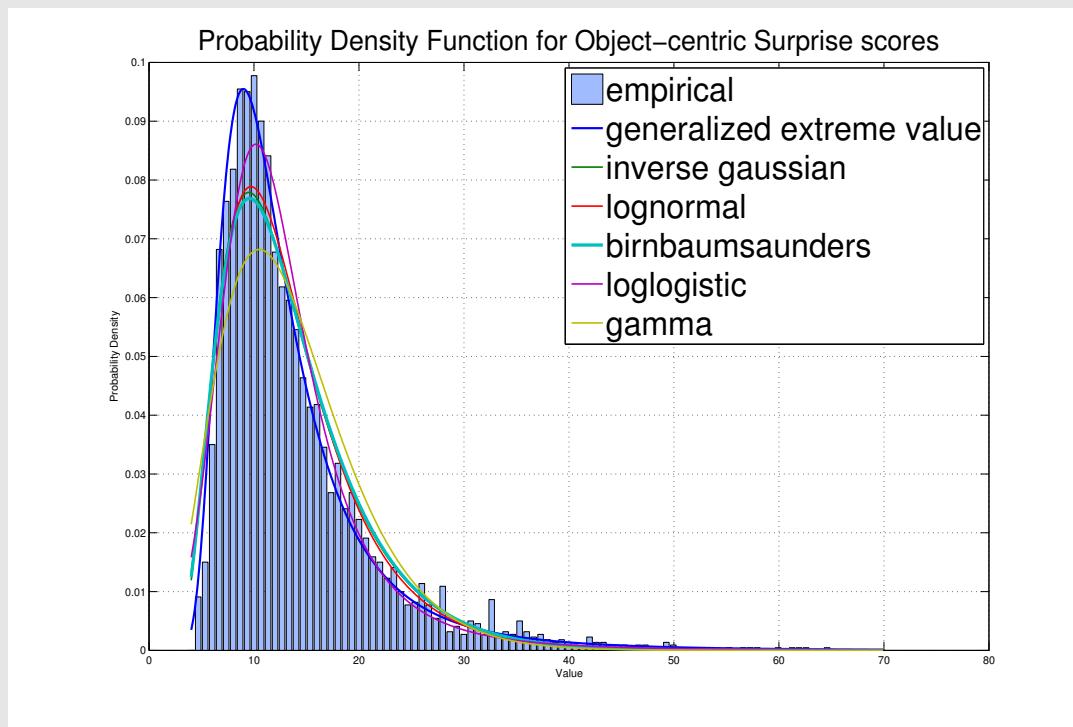
- $\phi_o(\cdot), \phi_s(\cdot), \phi_c(\cdot)$ are Inverse Gaussian CDF for Object, Scene, and context surprise.
- Taking the maximum among the complimentary probability values for these reasons, indicates the most influential reason behind the atypicality of the image.



Surprise scores(scene centric) for normal images

Abnormality Scores

- Raw surprise score are not comparable, and cannot be used to judge abnormality of an image.
- However, raw surprise scores follow exponential distributions when we plot the histogram of values for normal images.



Surprise scores (object centric) for normal images

Dataset

- Abnormal images:
 1. Continues abnormality score [0,1]
 2. Main reason of abnormality [cluster index in the inferred dendrogram]
 3. Rating of fine-grained reasons of abnormality.
- Normal images:
 1. Standard dataset for image categorization (PASCAL)

Experimental Setting

- Object Representation:
 - 64 Visual attributes.
 - Low-level features for categorization:
 - Kernel Descriptors [Bo et al, NIPS 2010]
- Scene Representation:
 - 102 Visual attributes [Patterson et al, CVPR 2012]
 - Scene categorization [Juneja et al, CVPR 2013]

Abnormality Detection – Prior Art

Experiment Number	Method	Accuracy
I	Object-centric baseline (Saleh, Farhadi, and Elgammal 2013) Our Model - Object-centric	0.9125 0.9311
II	Context-centric baseline (Park, Kim, and Lee 2012) Our Model - Context-centric	0.8518 0.8943

- Prior art in object-centric abnormality detection suffer from poor object classifiers (attribute-based). Our model also benefits from a richer set of visual features used for learning attribute classifiers.
- Similar to our work, Park et al. built their models only based on normal images. However, our reasoning based on information theoretic terms outperforms their CRF-based model of supporting context and objects.

Abnormality Detection - Baselines

Experiment Number	Method	Accuracy
III	One Class SVM - based on Attributes	0.5361
	Two Class SVM - based on Attributes	0.7855
	One class SVM - based on Deep features (fc6)	0.5969
	Two class SVM - based on Deep features (fc6)	0.8524
IV	Our Model - No Object-centric score	0.8004
	Our Model - No Context-centric score	0.8863
	Our Model - No Scene-centric score	0.8635
	Our Model - All three reasons	0.8914

- Our model is learned using only normal images, but outperforms two-way classifiers. Either trained via attributes or vanilla deep features.
- We relate this to the complexity of visual world, which makes it hard to find support vectors. Also, training SVM classifiers suffer from imbalanced data very much.

Abnormality Detection–Ablation Exp.

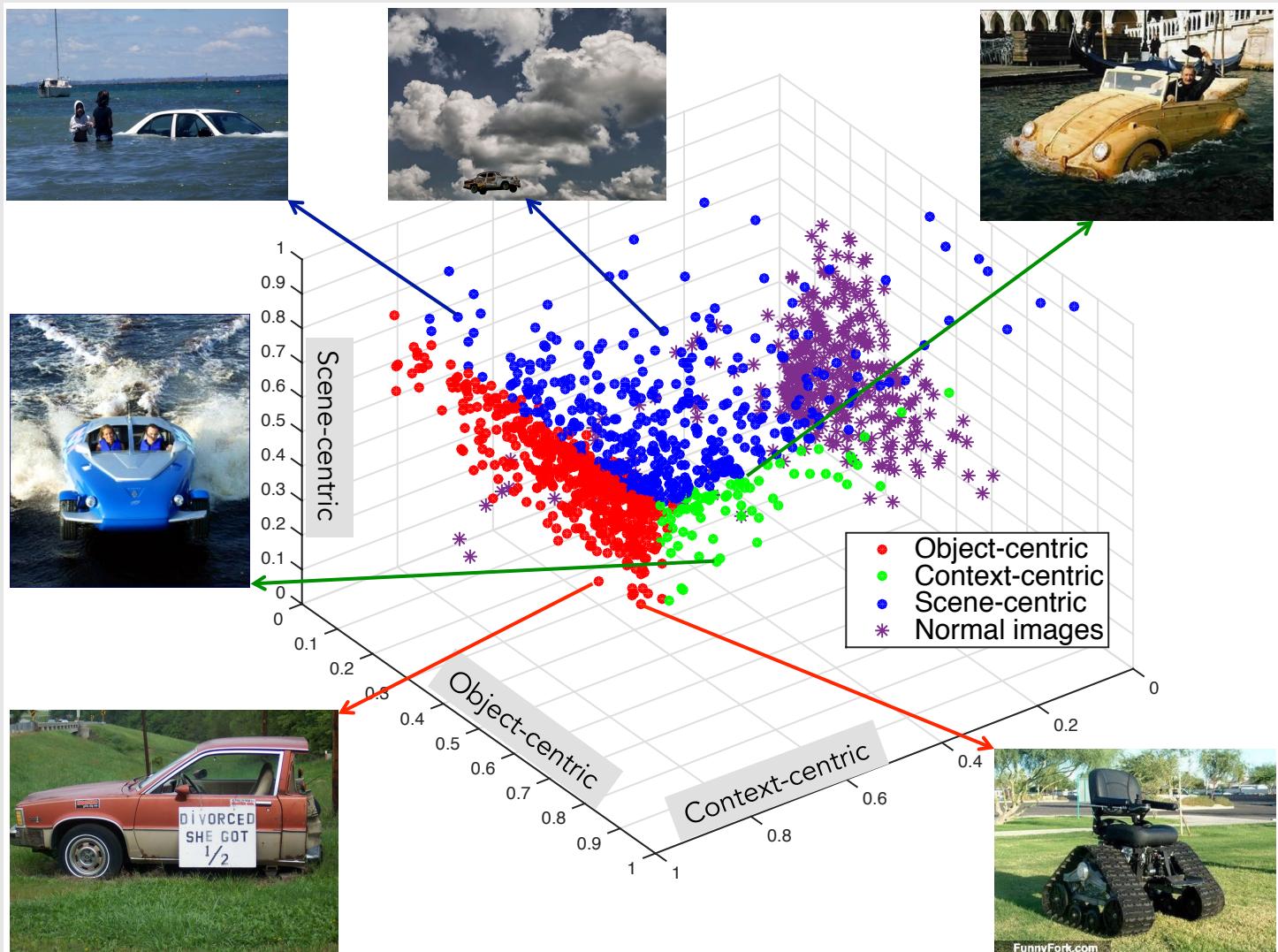
Experiment Number	Method	Accuracy
IV	Our Model - No Object-centric score	0.8004
	Our Model - No Context-centric score	0.8863
	Our Model - No Scene-centric score	0.8635
	Our Model - All three reasons	0.8914

- This ablation experiment shows the importance of each component of our final mode, where we held out one reason at a time.
- Clearly, the object-centric is the most important component of our performance in abnormality recognition in images.
- We relate this to the fact that majority of images in our dataset are atypical due to having abnormal object(s).

Abnormality Ranking

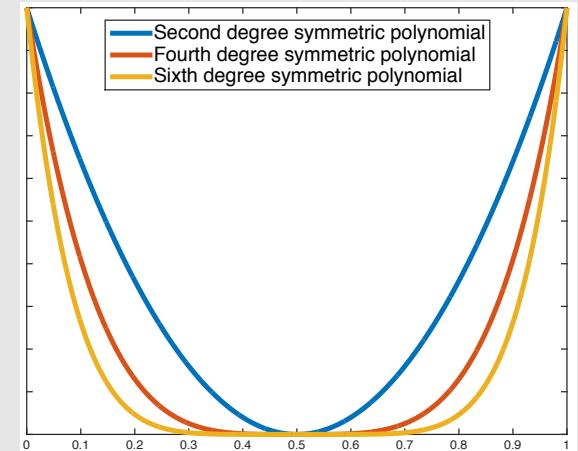


Abnormality in Perspective



Approach - Weighting Samples

- We propose a new training strategy for CNN:
 - Same training data (ImageNet), and same optimization (SGD)
 - Not all training examples are equally important
 - Modeling the importance of each training image based on how typical/normal it looks like
 - Weighting training sample during training as a function of their typicality.
 - At inference/test time, no need to measure the typicality of image
- Even-degree Polynomials
 - To emphasize on both normal and abnormal samples
 - Suitable for hybrid model, where typicality is as important as atypicality (discriminative)



Alternative Weighting function

- Internal scores of typicality
 - Entropy of the output of the layers of CNN resembles human's opinions about the typicality [Lake et al'15]
 - We used these scores to re-train the same CNN
 - This approach predicts the typicality of test images as well

Experiments

Weighting Function used in Fine-Tuning	Mean Accuracy (%)			
	Test Atypical		Test Typical	
	Epoch 1	Epoch 10	Epoch 1	Epoch 10
No weight	56.39	65.18	78.15	83.51
Random	57.15	66.45	73.60	83.84
Typicality	64.53	68.58	69.22	79.90
Atypicality	66.61	70.65	75.82	84.07
Cls-Typ	67.25	70.84	77	81.88
Cls-Atyp	63.26	68.46	76.96	83.40
Log-Typ	64.38	68.28	78.80	83.67
Log Cls-Atyp	64.21	67.80	76.13	83.24
Memorability	64.69	68.33	76.31	83.96
Poly Deg-2	59.13	69.49	80.03	84.42
Poly Deg-4	60.22	71.52	77.74	83.45
Poly Deg-6	60.86	70.31	77.66	84.22
In-Probability	65.97	69.53	80.71	85.82
In-Entropy	60.54	68.05	79.44	82.29
In-Prob + Atyp	62.94	68.21	75.82	83.09

Summary

1. There is a big gap between humans and computer vision, when it comes to atypical images.
2. Similar to human learning, we built algorithms that understand atypical images, without seeing them before.
3. We used information theory and visual attributes to show that computational reasoning correlates with human reasoning about abnormality.
4. Proposed a taxonomy for abnormality that has three main branches: Object, Context and Scene.
5. Incorporating typicality estimation in learning discriminative object classifiers improves the performance on both normal and abnormal objects.

Thank you!



Ahmed Elgammal
Rutgers



Jacob Feldman
Rutgers



Ali Farhadi
Apple
Babak S. Saleh
UW



Kostas Bekris
Amazon
Rutgers



Devi Parikh
Facebook AI Research
Georgia Tech
54