Seeing through the Human Reporting Bias: Visual Classifiers from Noisy Human-Centric Labels

Ishan Misra[1] C. Lawrence Zitnick[3] Margaret Mitchell[2]
Ross Girshick[3]

[1]Carnegie Mellon University

[2] Microsoft Research

[3] Facebook AI Research

(experiments presented by An T. Nguyen)

Paper Review

Human Reporting Bias

- Captions, tags, keywords ...
- Report only salient/important objects.
- Cause visually biased classifiers.

Paper Review

Human Reporting Bias

- Captions, tags, keywords ...
- Report only salient/important objects.
- Cause visually biased classifiers.

This paper

- Model the bias as hidden variables.
- Improve classification of visual concepts.

Examples
(from MSCOCO dataset)

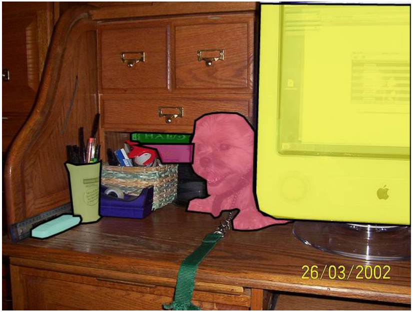
Captions:

- a small dog is on a wood desk
- a dog is sitting on a desk behind a computer.
- dog sitting on a desk next to a monitor
- a little dog with a leash laying on a desk behind a computer monitor.
- a dog sits on a desk behind a computer

Examples

Detection labels

- dog
- tv
- remote
- cup
- book book

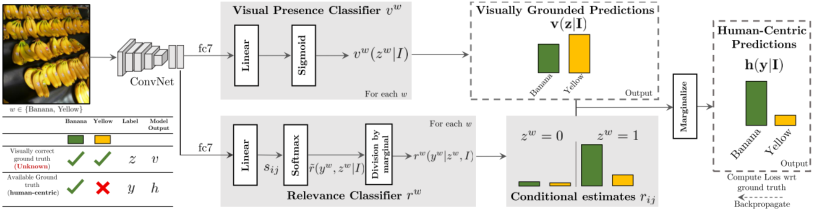


Two Classifiers

1. Visual Presence Classifier **v**
2. Relevance Classifier **r**

Two Classifiers

1. Visual Presence Classifier **v**
2. Relevance Classifier **r**



My experiments

Analyze the **relevance classifier r**

1. **r** with varying objects sizes, orientations
   (is **r** sensitive to sizes and orientations?)

My experiments

Analyze the **relevance classifier r**

1. **r** with varying objects sizes, orientations
   (is **r** sensitive to sizes and orientations?)
2. Evaluate the accuracy or **r**.
   (in detecting (un)reported objects)

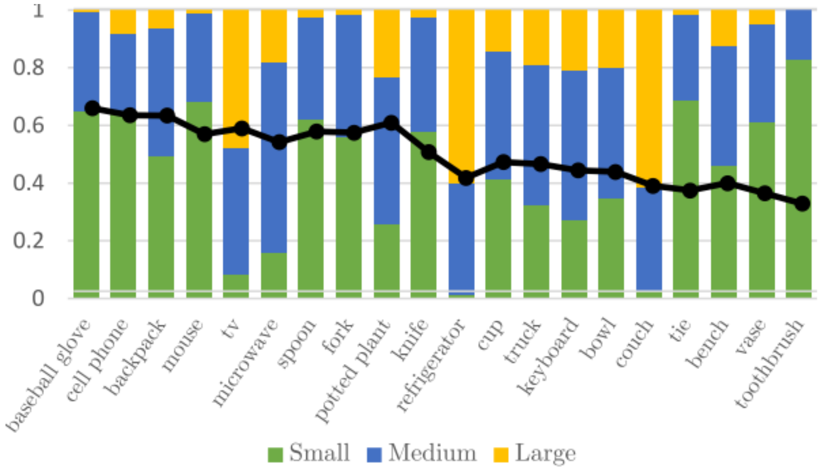
My experiments

Analyze the **relevance classifier r**

1. **r** with varying objects sizes, orientations
   (is **r** sensitive to sizes and orientations?)
2. Evaluate the accuracy or **r**.
   (in detecting (un)reported objects)
3. Evaluate the learned ‘representation’
   (as features in scene classification).

Reportability with varying size

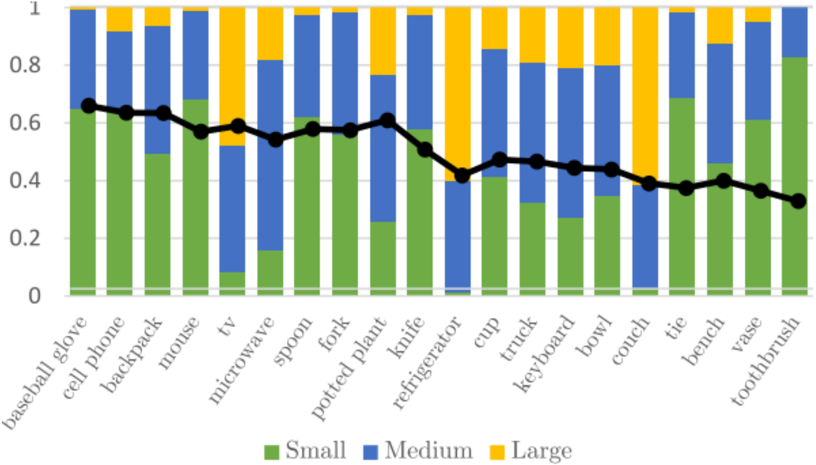
(image from the paper, black line = prob. of not reporting)



- Small size correlates with not reported.

Reportability with varying size

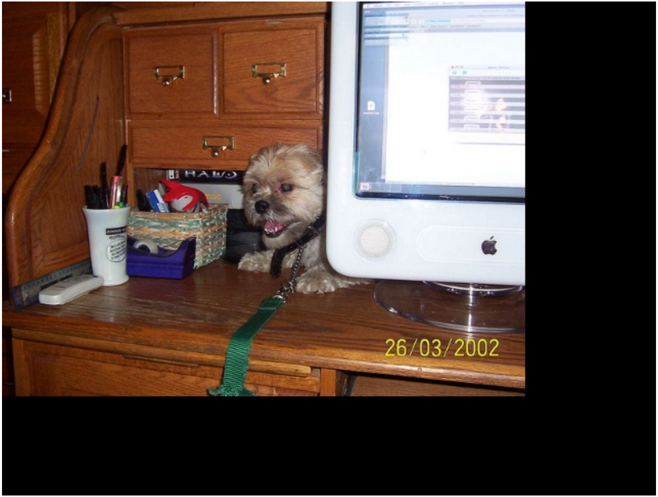
(image from the paper, black line = prob. of not reporting)



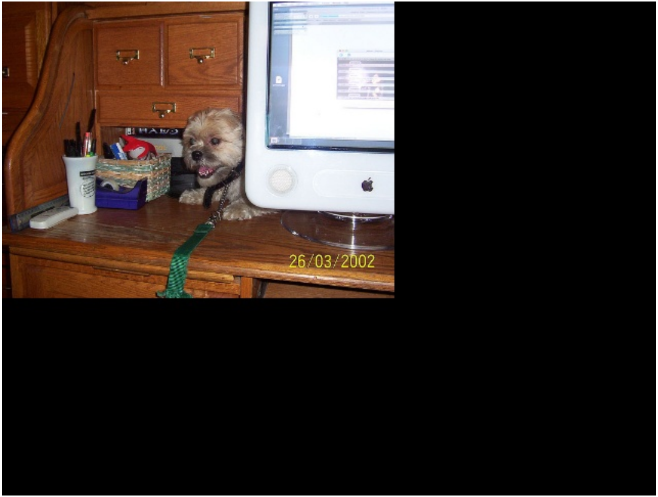
- Small size correlates with not reported.
- **Question: Does r capture this?**

Experiment: varying sizes

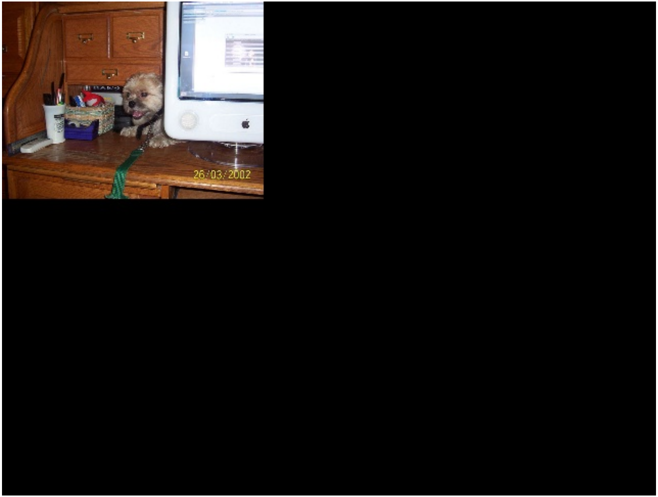
Experiment: varying sizes



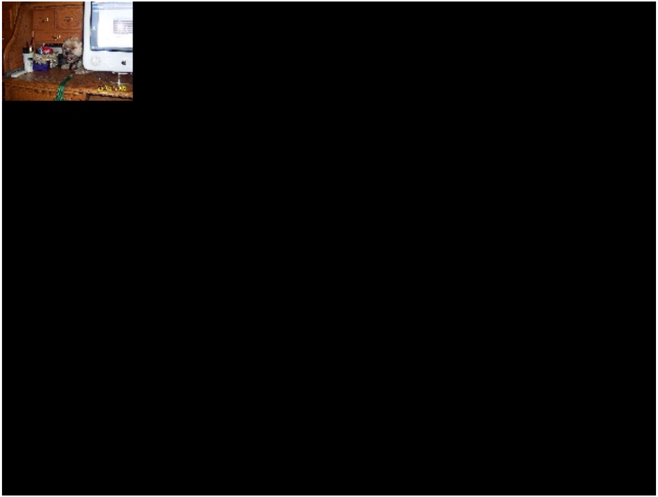
Experiment: varying sizes



Experiment: varying sizes

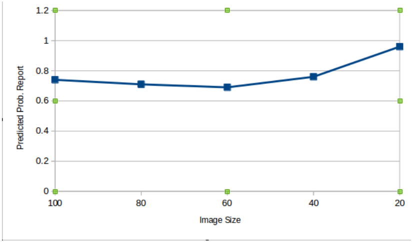


Experiment: varying sizes



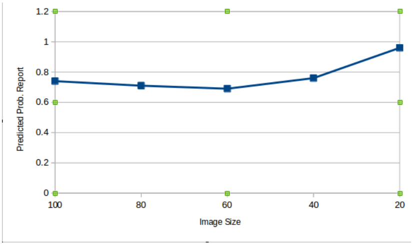
Experiment: varying sizes

(Average over 1000 images in test set and
over common objects: glove, phone, backpack, ...)



Experiment: varying sizes

(Average over 1000 images in test set and
over common objects: glove, phone, backpack, ...)

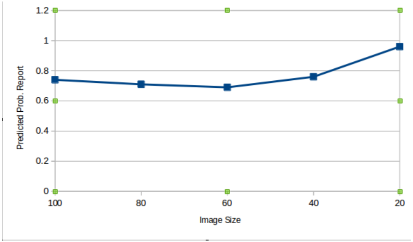


Observations:

1. (Almost) same from 100% to 60%
2. But increase from 60% to 20%

Experiment: varying sizes

(Average over 1000 images in test set and
over common objects: glove, phone, backpack, ...)



Observations:

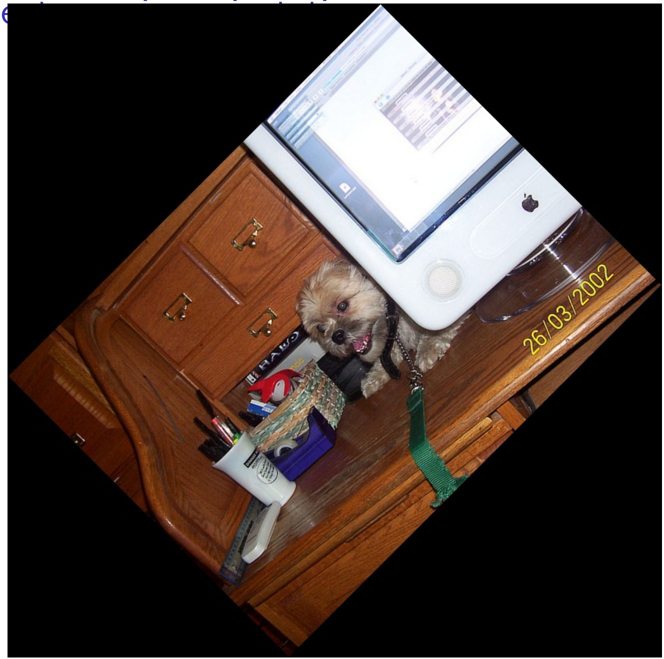
1. (Almost) same from 100% to 60%
2. But increase from 60% to 20%

(Possible) explanation:

1. **r** is not sensitive to size. (it predicts based on other features)
2. Objects too small → not recognized → default to reported

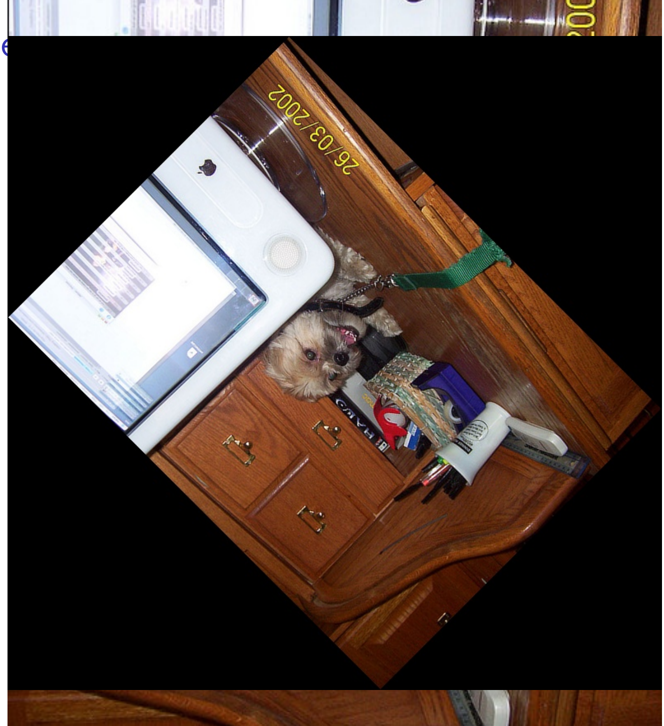
Experiment: varying orientations

Experime



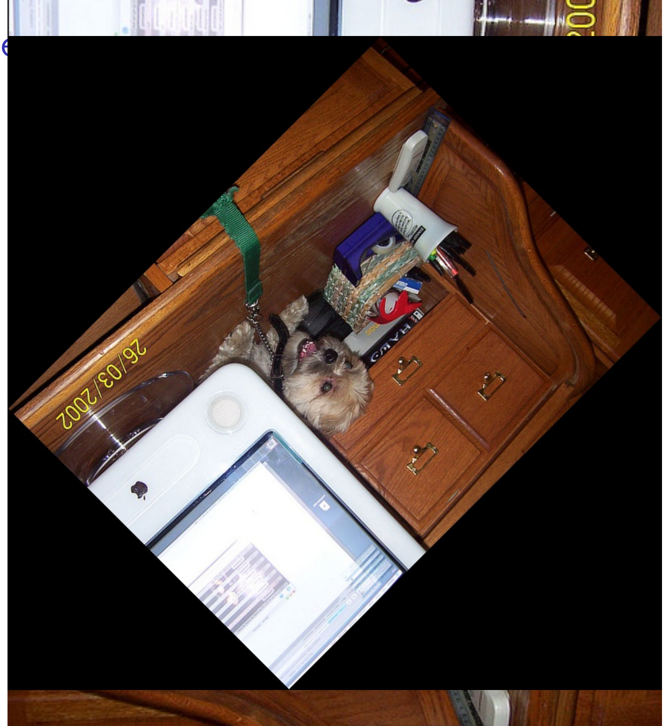
Experime

Experime

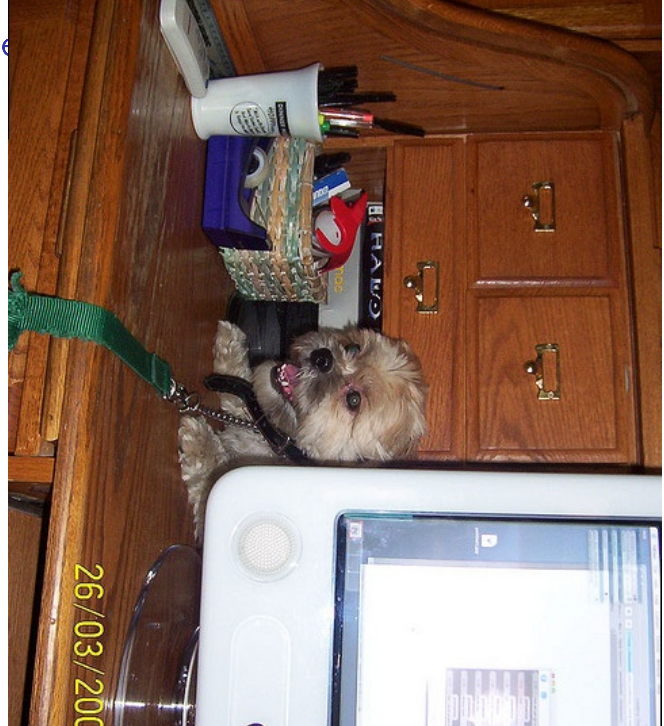


Experime

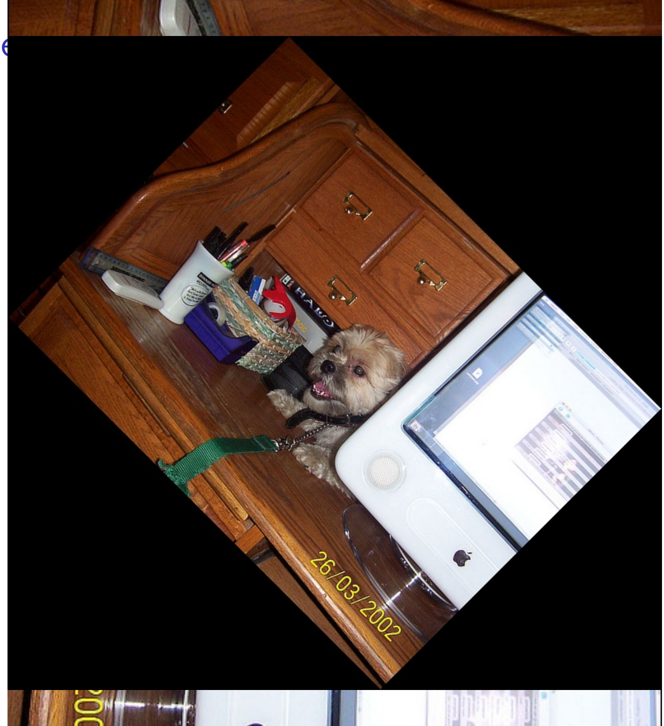
Experime



Experime

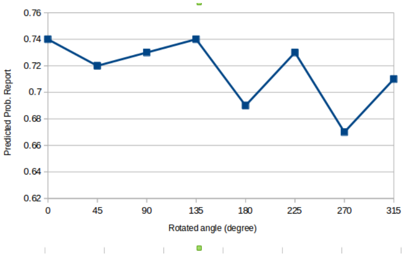


Experime



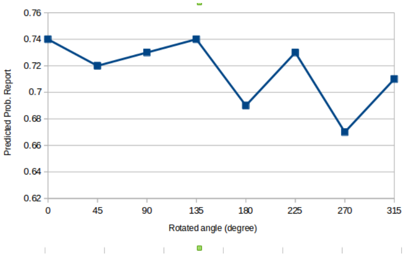
Experiment: varying orientations

(Average over 1000 images in test set and
over common objects: glove, phone, backpack, ...)



Experiment: varying orientations

(Average over 1000 images in test set and
over common objects: glove, phone, backpack, ...)



- **r** sensitive to orientations.
- Unusual rotation $\rightarrow$ not recognized $\rightarrow$ ...

Accuracy of **r**

(Surprisingly not reported by the paper)

For each concept:

- Negative instances: object present but no captions mentioned.
- Positive instances: object present and captions mentioned.

Accuracy of $\mathbf{r}$

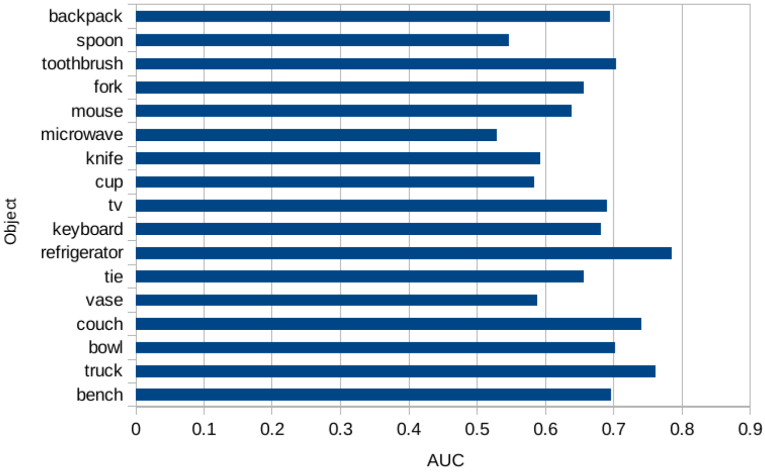
(Surprisingly not reported by the paper)

For each concept:

- Negative instances: object present but no captions mentioned.
- Positive instances: object present and captions mentioned.
- Metric: AUC of $\mathbf{r}$ prediction.

Accuracy of **r**

(over all images in test set)



Evaluate the learned ‘representation’

- **r** outputs 4 ‘probabilities’ for each concept.

Evaluate the learned 'representation'

- **r** outputs 4 'probabilities' for each concept.
- 1000 concepts $\rightarrow$ 4000-dim vector.

Evaluate the learned ‘representation’

- **r** outputs 4 ‘probabilities’ for each concept.
- 1000 concepts $\rightarrow$ 4000-dim vector.
- Assumed to be features for scene classification.

Evaluate the learned 'representation'

- **r** outputs 4 'probabilities' for each concept.
- 1000 concepts $\rightarrow$ 4000-dim vector.
- Assumed to be features for scene classification.
- Pretext task = predict human reporting bias.

Evaluate the learned 'representation'

- **r** outputs 4 'probabilities' for each concept.
- 1000 concepts $\rightarrow$ 4000-dim vector.
- Assumed to be features for scene classification.
- Pretext task $=$ predict human reporting bias.
- Same data as in Assignment 2.

Evaluate the learned 'representation'

(LinearSVM, no finetuning, test set 2)

Features	Accuracy(%)
HumanBias	58.24
Alex	81.36
HumanBias + Alex	82.62
ResNet	87.12
HumanBias + ResNet	87.73

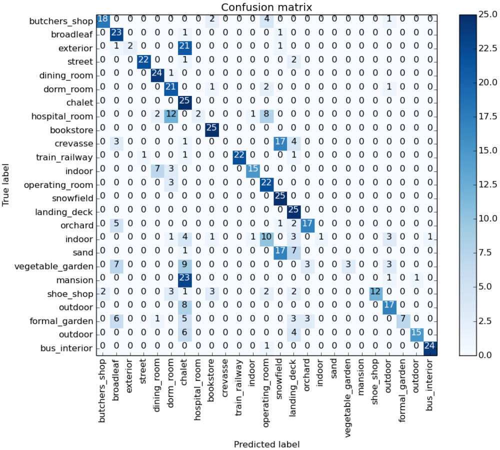
Evaluate the learned 'representation'

(LinearSVM, no finetuning, test set 2)

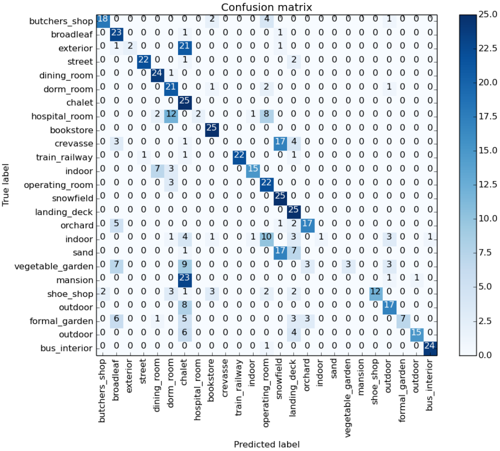
Features	Accuracy(%)
HumanBias	58.24
Alex	81.36
HumanBias + Alex	82.62
ResNet	87.12
HumanBias + ResNet	87.73

- Features are informative
- Complementary to Alex & ResNet

Confusion Matrices for Human Bias



Confusion Matrices for Human Bias



- Less distinctive, mix categories.
- e.g. exterior,mansion,chalet.

Summary

- **r** classifies into reported/unreported by human.
- Sensitive to orientations, not to scale.
- Good performance in AUC.
- Learn informative features.

Summary

- **r** classifies into reported/unreported by human.
- Sensitive to orientations, not to scale.
- Good performance in AUC.
- Learn informative features.

Questions?