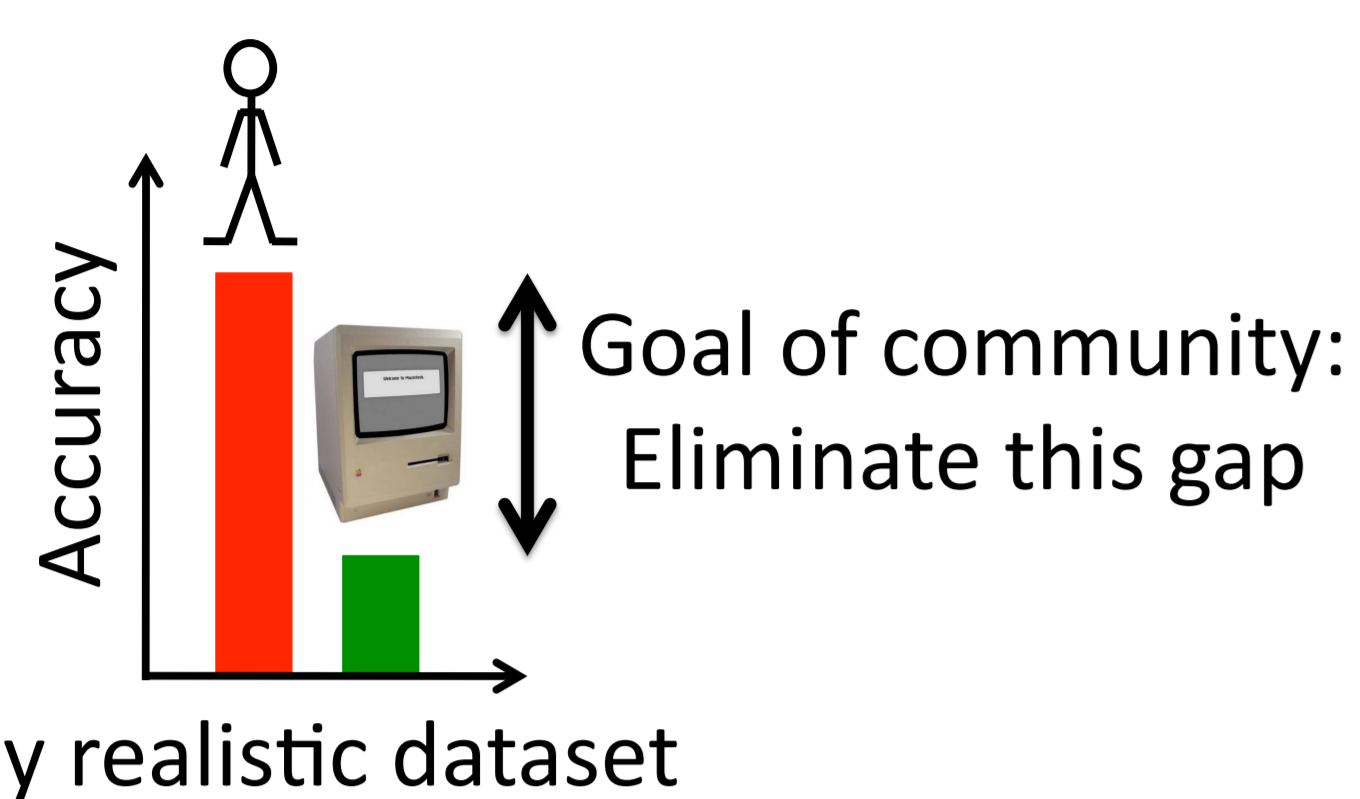# Recognizing Jumbled Images: The Role of Local and Global Information in Image Classification
Devi Parikh (TTIC)

1. Motivation

Current State of Affairs

Accuracy

Goal of community: Eliminate this gap
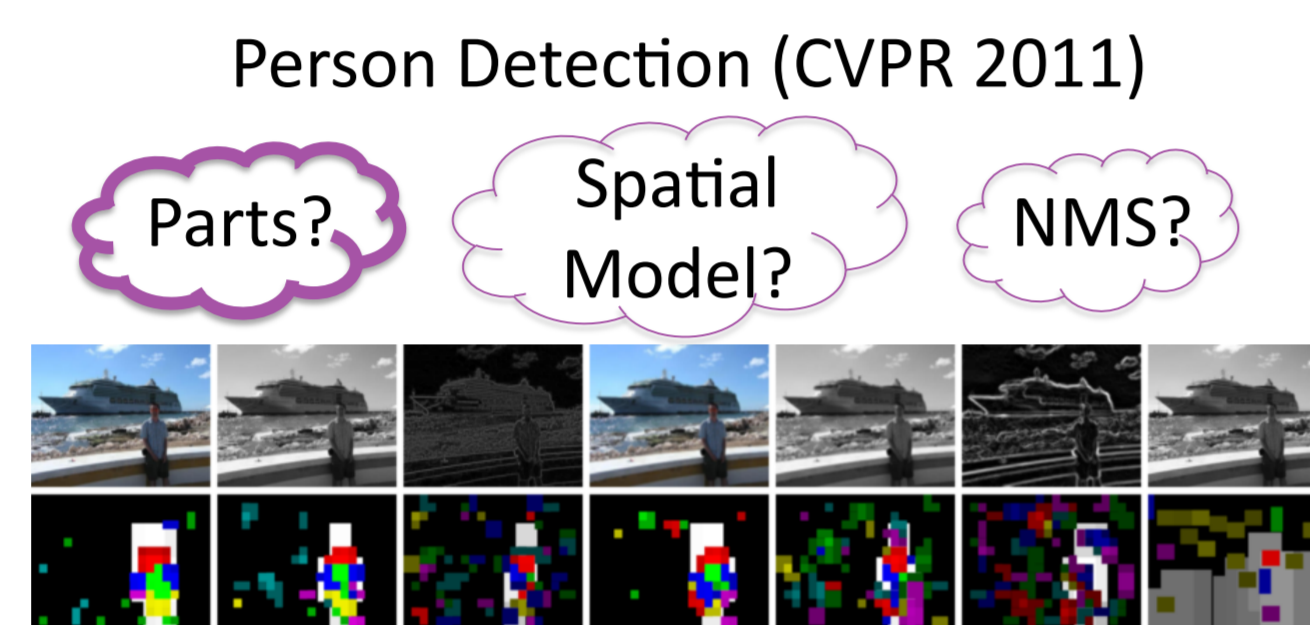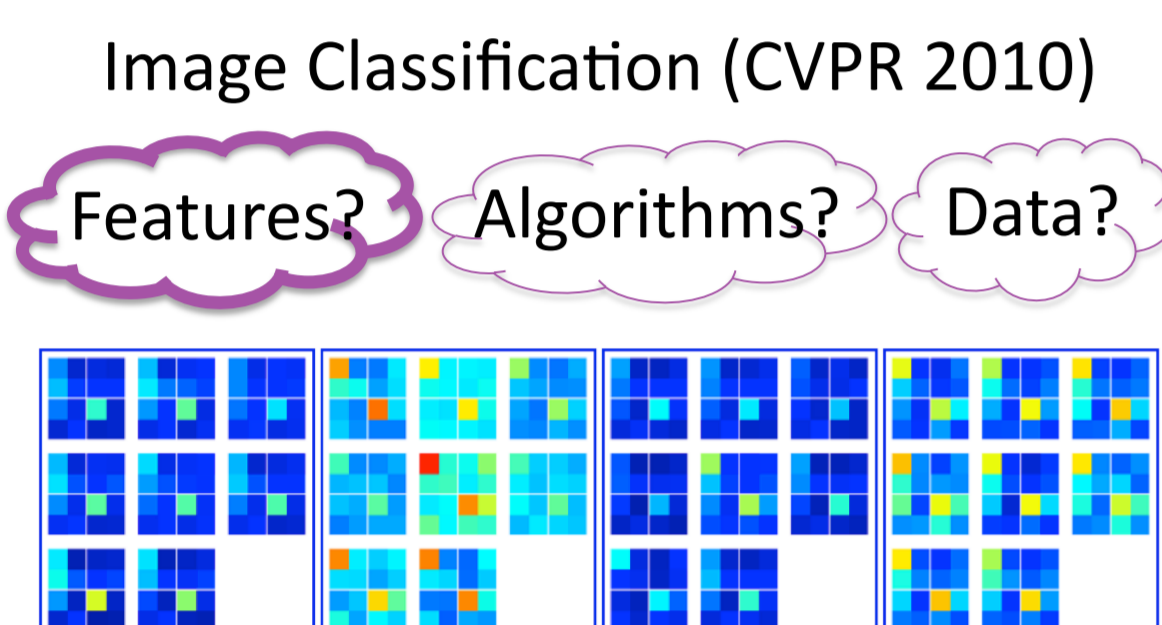
Any realistic dataset

Most Existing Efforts

✧ Focus on **how** to solve the problem:
  more features, algorithms, data…
✧ Often leave humans out of the picture
✧ Have made a lot of advancements
✧ BUT where to go next?

This Work

✧ Focus on determining **what** problems to solve
✧ Human debugging: use humans to identify research directions with potential

Image Classification (CVPR 2010)

Features? Algorithms? Data?

Person Detection (CVPR 2011)

Parts? Spatial Model? NMS?

2. Focus of this Work

Image Classification

Local Information? Bottle-neck? Global Information?

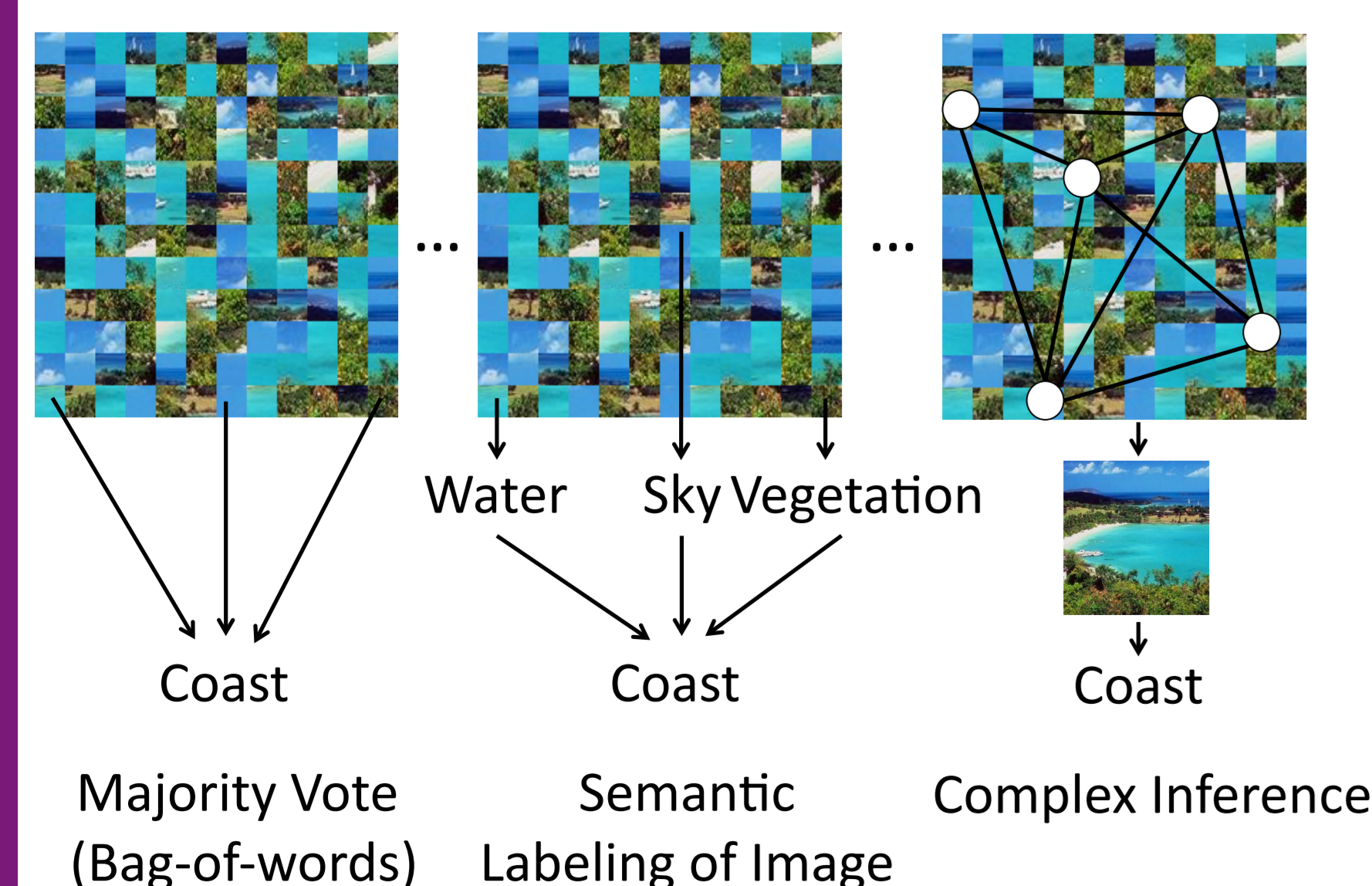Human and machine image classification with one (local) and not the other (global)

Images with Local But No Global Information: Jumbled Images

3. Goals

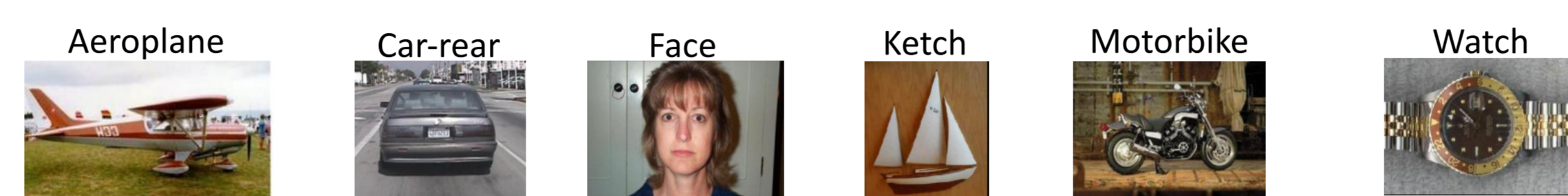1. Which functional model mimics how humans utilize local information alone?

Water   Sky Vegetation

Coast   Coast   Coast

Majority Vote (Bag-of-words)   Semantic Labeling of Image   Complex Inference

Hypothesis

Majority Vote

<scene> =?= <scene>
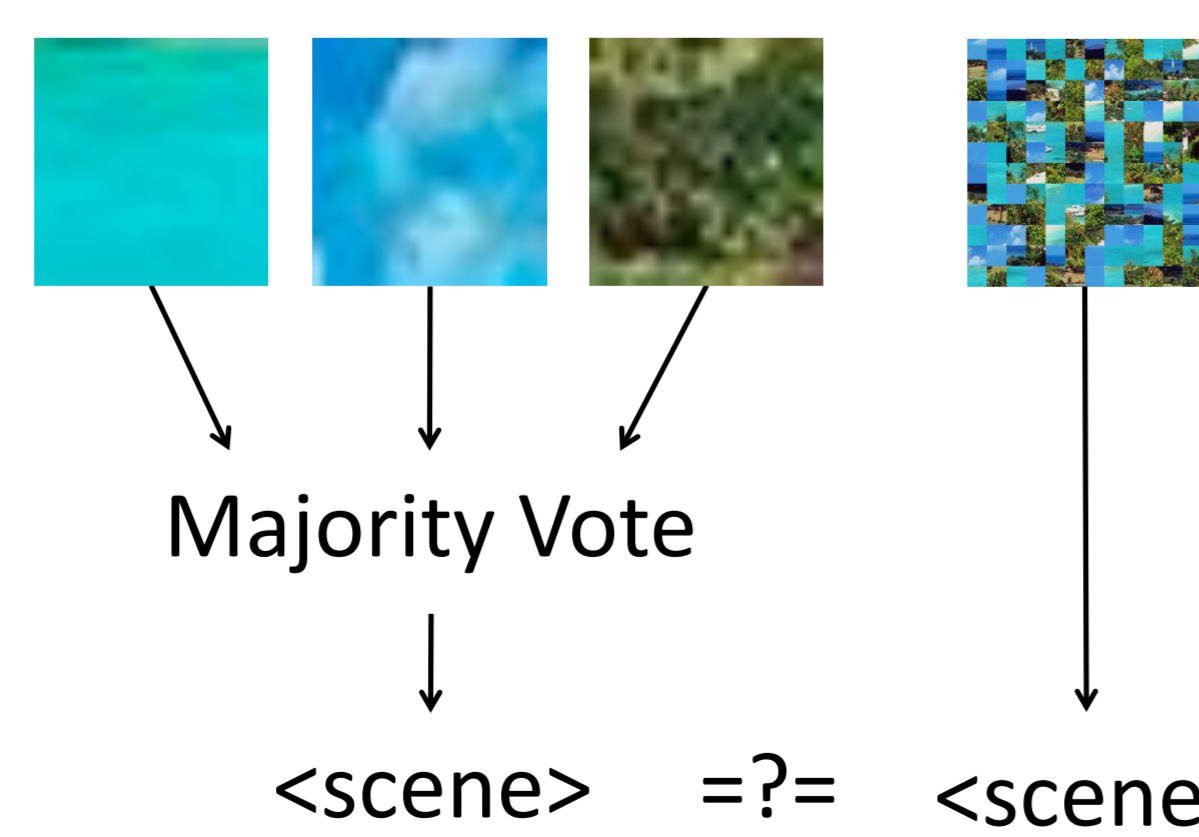
2. How does corresponding machine implementation compare?

4. Datasets

Indoor Scene Recognition (ISR)

Bathroom Bedroom Dining room Gym Kitchen Living room Movie theater Stair case

Caltech Object Recognition (CAL)

Aeroplane Car-rear Face Ketch Motorbike Watch

Outdoor Scene Recognition (OSR)

5. Human Studies

Categorize an image:

Examples of scene categories:

Coast Highway Mountain Street
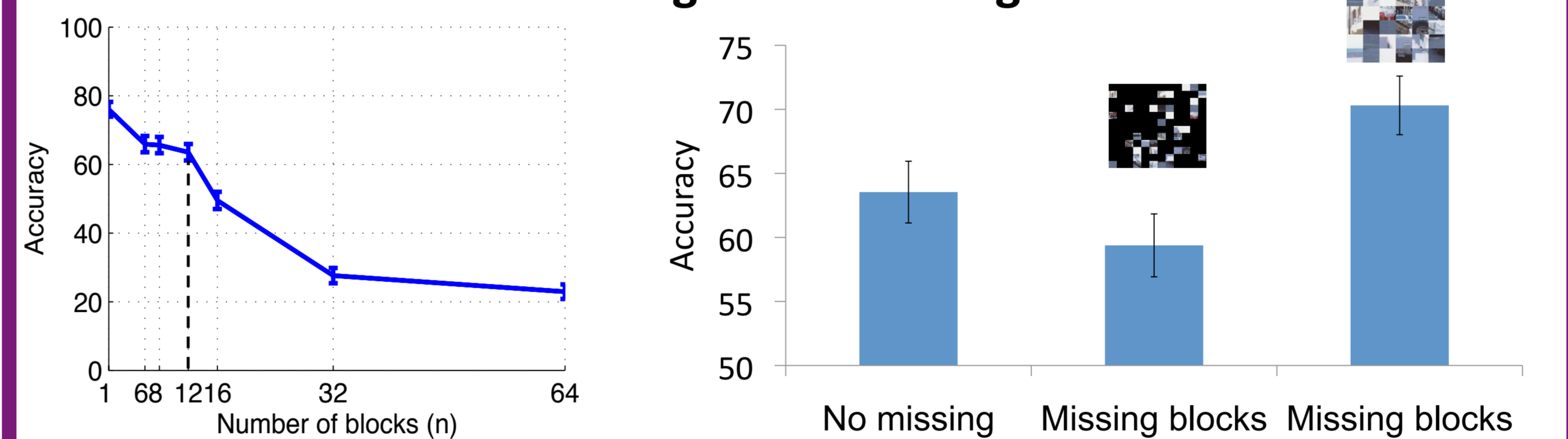
Forest Inside city Open country Tall buildings

What is the scene category depicted in the image below?

amazon mechanical turk
10 subjects per instance

○ Coast   ○ Highway   ○ Mountain   ○ Street
○ Forest ○ Inside city ○ Open country ○ Tall buildings

6. Method: Jumbled Images

Creating Jumbled Images

Accuracy

Number of blocks (n)

No missing blocks   Missing blocks visible   Missing blocks not visible
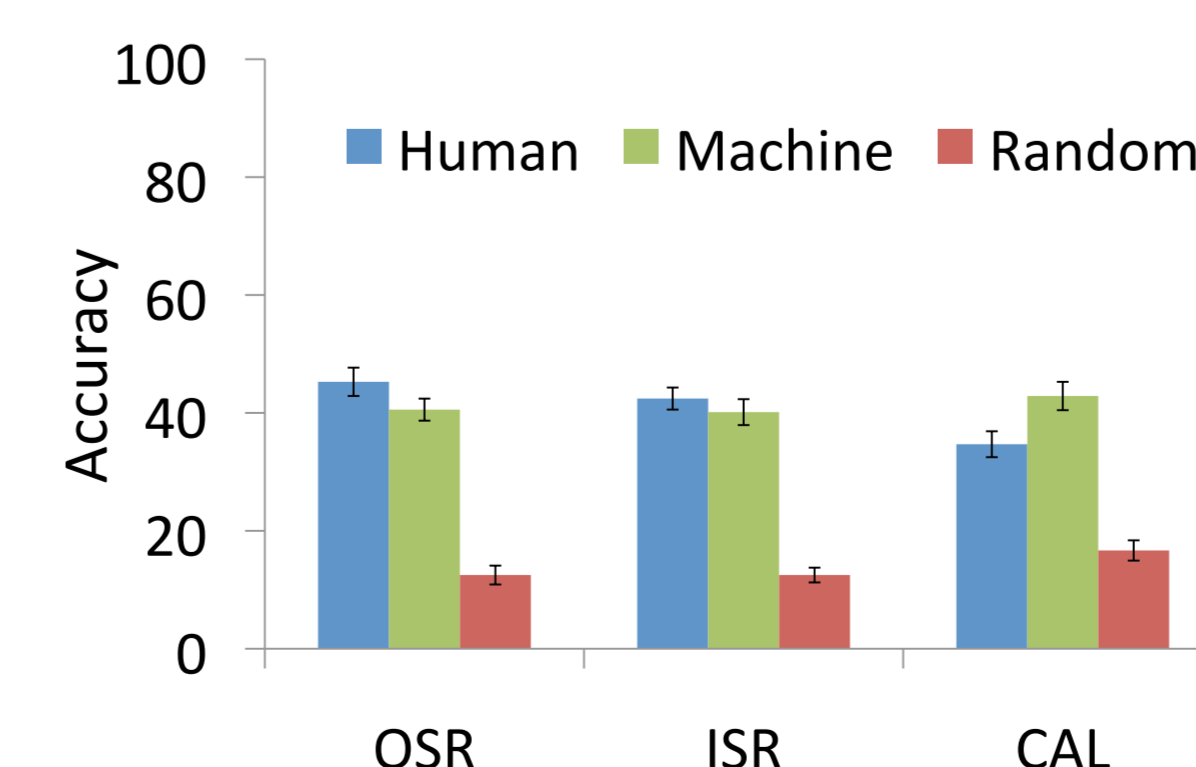
Intact   Jumbled

OSR   ISR   CAL

Dataset Filtering:

✧ Semantic inconsistencies
✧ Human inconsistencies
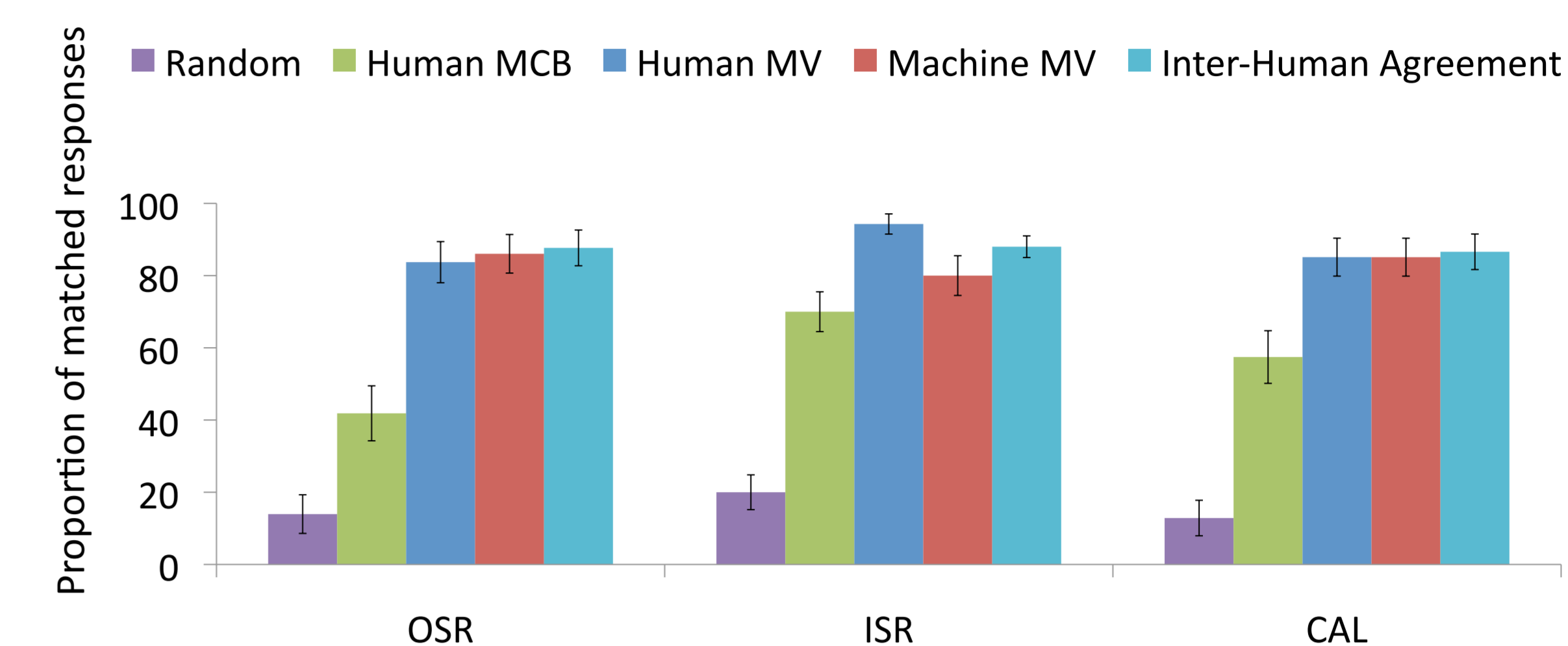✧ OSR: 35/8, ISR: 55/15, CAL: 42/5

7. Method: Isolated Blocks

✧ 58k human responses

✧ Machine implementation:
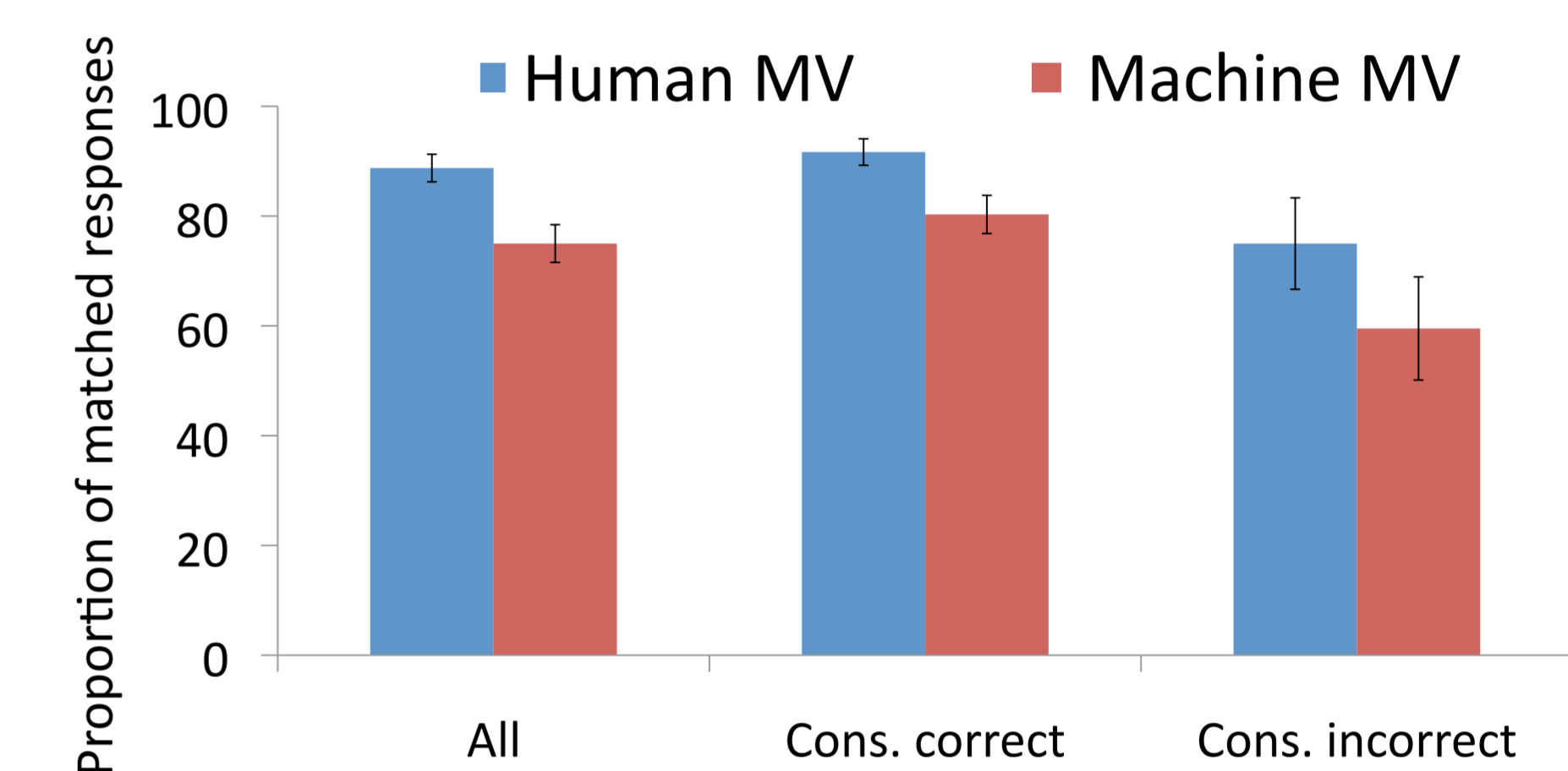 120D descriptor, 500 words
  RBG, HSV, Filter-bank response

Human   Machine   Random

OSR   ISR   CAL

Mediocre classification of individual isolated blocks

8. Results

Majority vote is an accurate functional model

Random   Human MCB   Human MV   Machine MV   Inter-Human Agreement

Proportion of matched responses

OSR   ISR   CAL

Majority vote model also predicts human mistakes

Human MV   Machine MV

Proportion of matched responses

All   Cons. correct   Cons. incorrect

Machine accuracy on jumbled images is comparable to humans

Accuracy

OSR   ISR   CAL
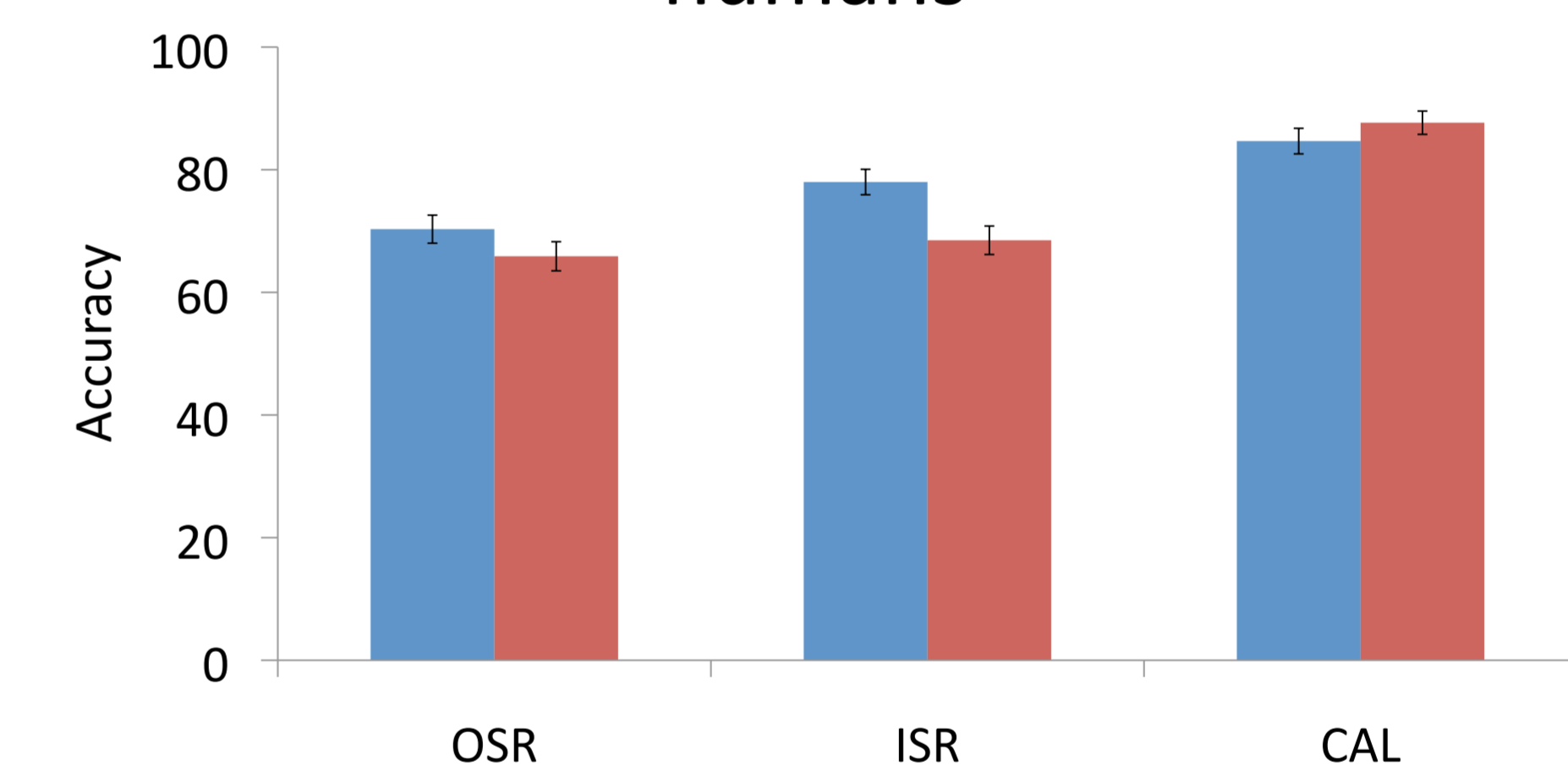
9. Conclusions

✧ Human debugging

✧ Bottleneck: global information
✧ Simple majority vote model mimics how humans classify jumbled images
✧ Existing implementations leverage local information effectively
✧ More advancements required to model global information well e.g. study low-res images