Visual Recognition with Humans in the Loop

S. Branson et al.

Presented by Thomas Provan

Problem

How can we augment existing computer vision systems to make them practical to use?

Motivation

- Existing systems are not current good enough for practical purposes
- Systems are focused on tasks which humans are naturally good at

Motivation

(A) Easy for Humans

Chair? Airplane? …

(B) Hard for Humans

Finch? Bunting?…

(C) Easy for Humans

Yellow Belly? Blue Belly? …

S. Branson et al.

Human In The Loop

- Combine an object recognition system with a user interaction system
- Enables the combined system to accomplish practical results in more complicated settings
- Attempts to minimize human labor

Human In The Loop

The bird is a **Black-footed Albatross**

Is the belly white? **yes**
Are the eyes white? **yes**
The bird is a **Parakeet Auklet**

Is the beak cone-shaped? **yes**
Is the upper-tail brown? **yes**
Is the breast solid colored? **no**
Is the breast striped? **yes**
Is the throat white? **yes**
The bird is a **Henslow's Sparrow**

S. Branson et al.

Human In The Loop

Western Grebe

w/ vision:

Q #1: Is the throat white? **yes (Def.)**

w/o vision:

Q #1: Is the shape perching-like? **no (Def.)**

Rose-breasted Grosbeak

Only CV

Yellow-headed Blackbird

CV + *Q #1:* Is the crown black? **yes (Def.)**

Rose-breasted Grosbeak

S. Branson et al.

Basic Algorithm

Input Image (x)

Computer Vision

$p(c \mid x)$

Max Expected Information Gain

Question 1:
Is the belly black?

A: NO

u_1

$p(c \mid x, u_1)$

Max Expected Information Gain

Question 2:
Is the bill hooked?

A: YES

u_2

$p(c \mid x, u_1, u_2)$

S. Branson et al.

Basic Algorithm

Algorithm 1 Visual 20 Questions Game

1: $U^0 \leftarrow \emptyset$
2: **for** $t = 1$ to 20 **do**
3: $\quad j(t) = \max_k \mathrm{I}(c; u_k | x, U^{t-1})$
4: $\quad$ Ask user question $q_{j(t)}$, and $U^t \leftarrow U^{t-1} \cup u_{j(t)}$.
5: **end for**
6: Return class $c^* = \max_c p(c|x, U^t)$

S. Branson et al.

Basic Algorithm

Algorithm 1 Visual 20 Questions Game

1: $U^0 \leftarrow \emptyset$
2: **for** $t = 1$ to 20 **do**
3: $\quad j(t) = \max_k \mathrm{I}(c; u_k | x, U^{t-1})$ ← Chooses the next question to ask
4: $\quad$ Ask user question $q_{j(t)}$, and $U^t \leftarrow U^{t-1} \cup u_{j(t)}$.
5: **end for**
6: Return class $c^* = \max_c p(c|x, U^t)$

S. Branson et al.

Basic Algorithm

Select the next question that maximizes expected information gain

- Easy to compute if we can to estimate probabilities of the form:

$$p(c \mid x, u_1, u_2 \ldots u_t)$$

Object Class

Image

Sequence of user responses

S. Branson et al.

Basic Algorithm

$$p(c \mid x, u_1, u_2 \ldots u_t)$$

$$\approx p(u_1, u_2 \ldots u_t \mid c)\, p(c \mid x) / Z$$

Model of user responses

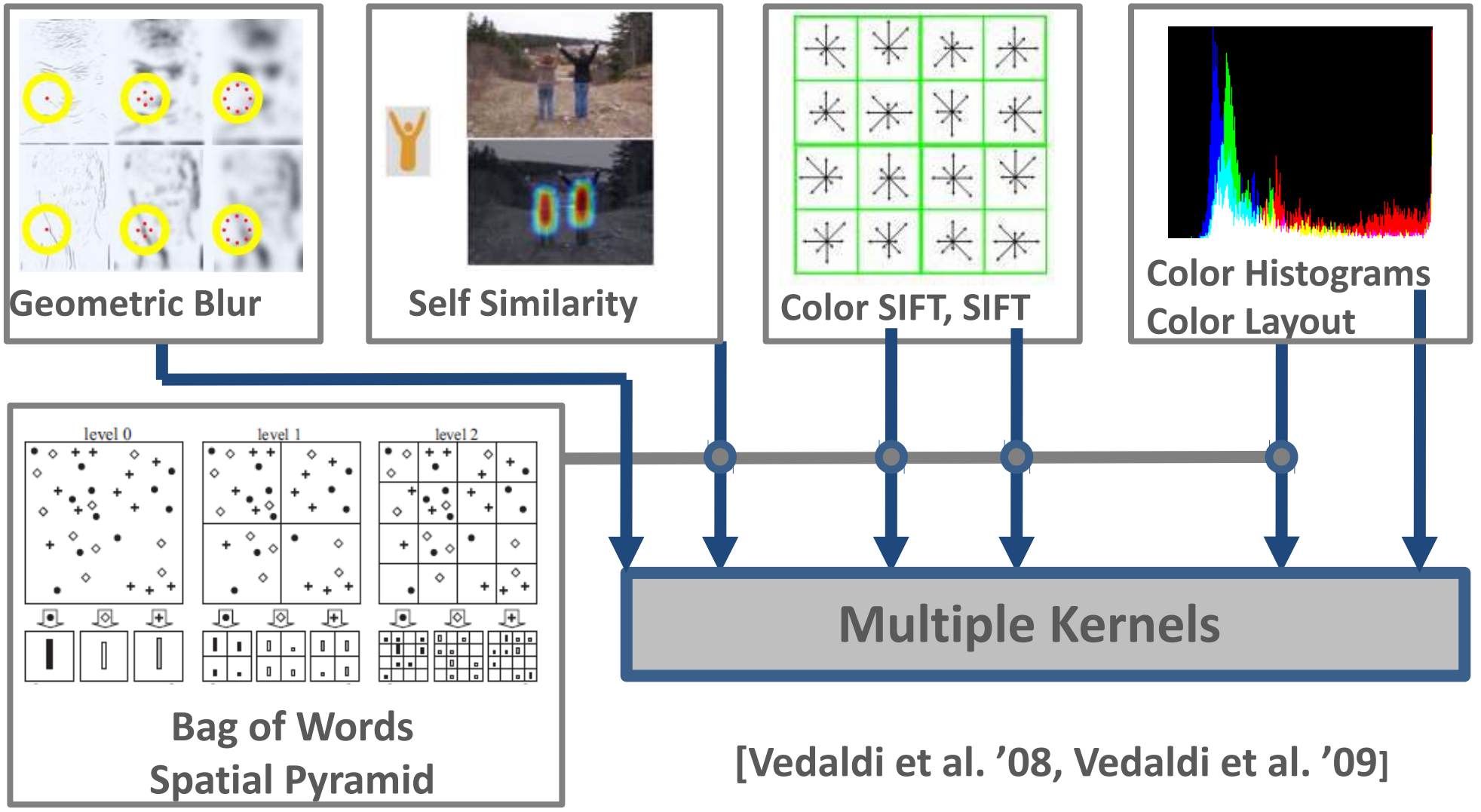
Computer vision estimate

Normalization factor

S. Branson et al.

Incorporating Computer Vision

- Used VLFeat and MKL code + color features



S. Branson et al.

Twenty Questions

You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.

What is the **color of the underparts** of the bird? 10/28

Select at least one. If the underparts aren't visible, make your best guess, then select "Guessing". If the color is a mixture of two colors, select both (e.g., for blue-green select blue and green). If the underparts have multiple regions or patterns with multiple colors, select all relevant colors (e.g., for yellow with black stripes, select yellow and black).

White	Black	Grey	Buff	Brown
Rufous	Red	Pink	Orange	Yellow
Green	Olive	Blue	Purple	Shiny / Iridescent

Go Back | Guessing | Probably | Definitely

S. Branson et al.

Birds 200 Dataset

- 200 classes, 6000+ images, 288 binary attributes
- Task: Classify the particular bird species

Black-footed Albatross

Groove-Billed Ani

Parakeet Auklet

Field Sparrow

Vesper Sparrow

Arctic Tern

Forster's Tern

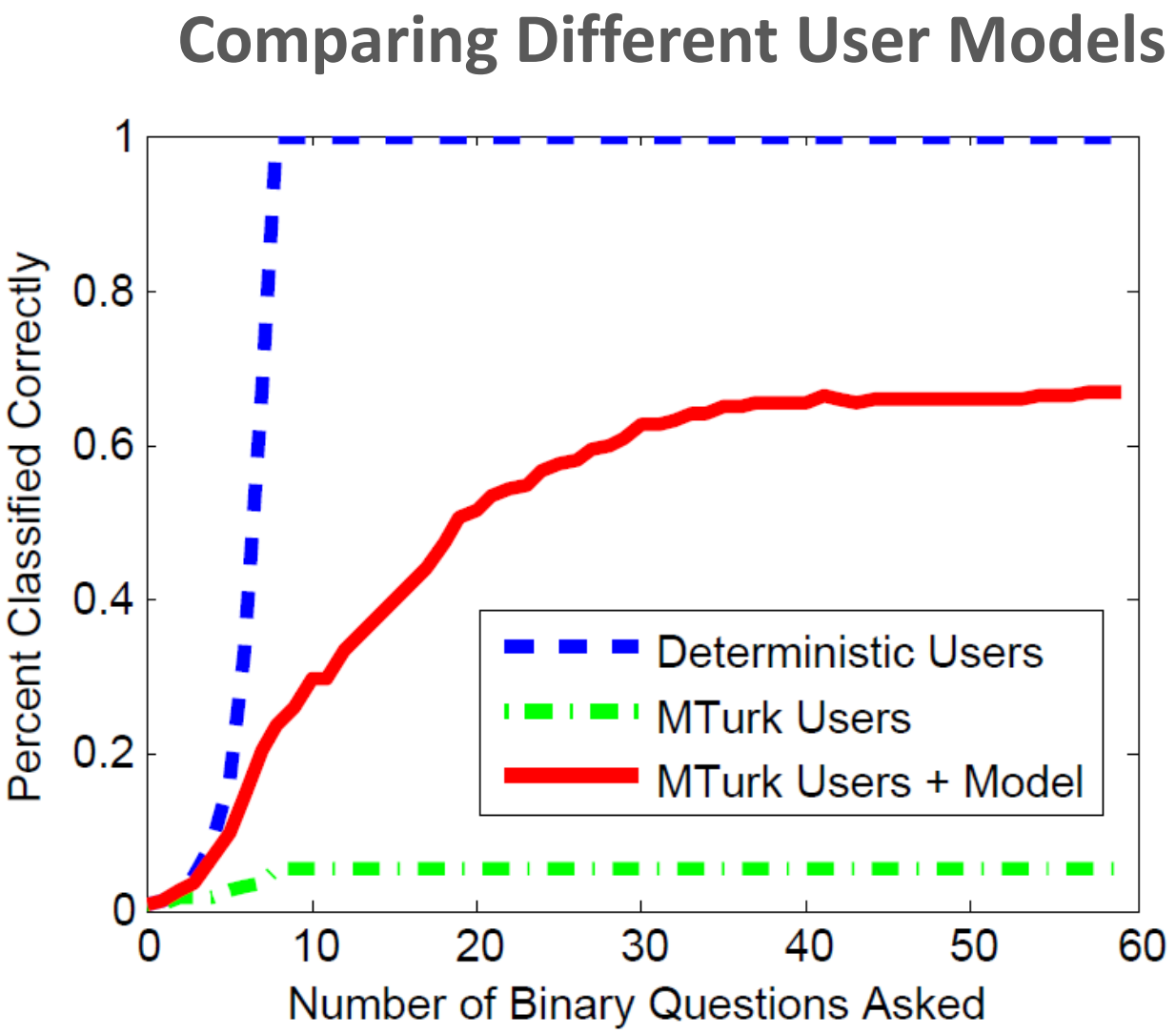
Common Tern

Baird's Sparrow

Henslow's Sparrow

S. Branson et al.

Results: Without Computer Vision



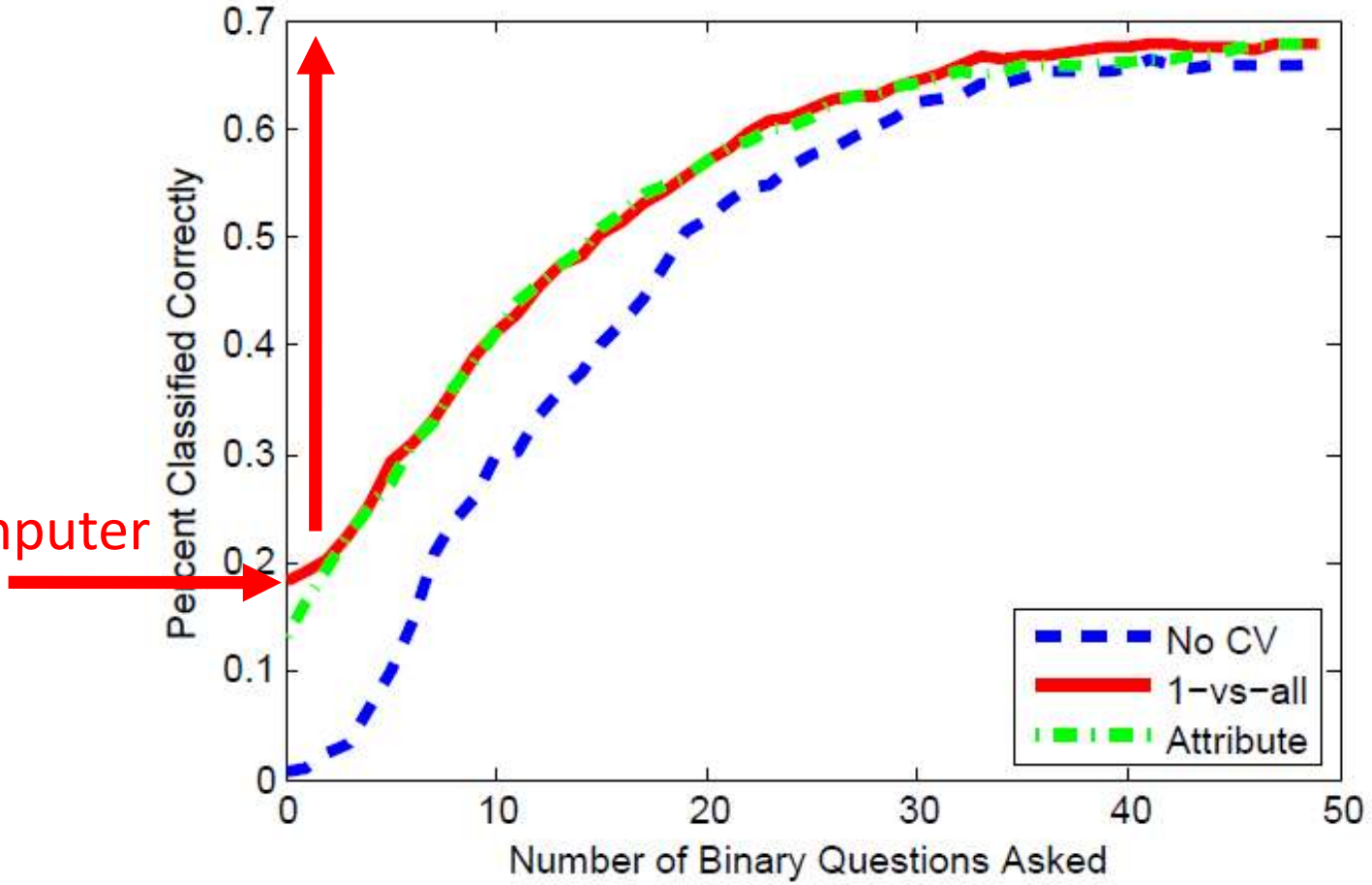
S. Branson et al.

Results: With Computer Vision

Users drive performance: 19% → 68%

Just Computer Vision

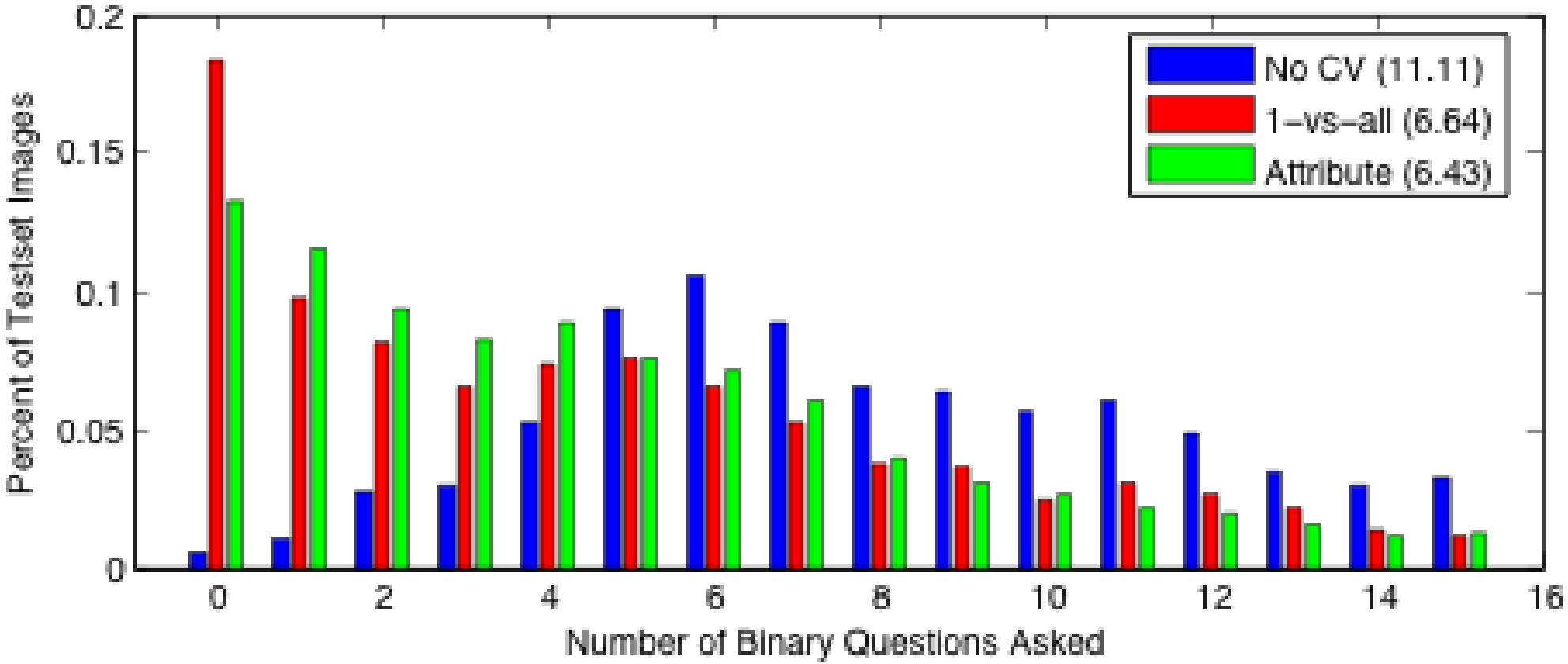
19%



S. Branson et al.

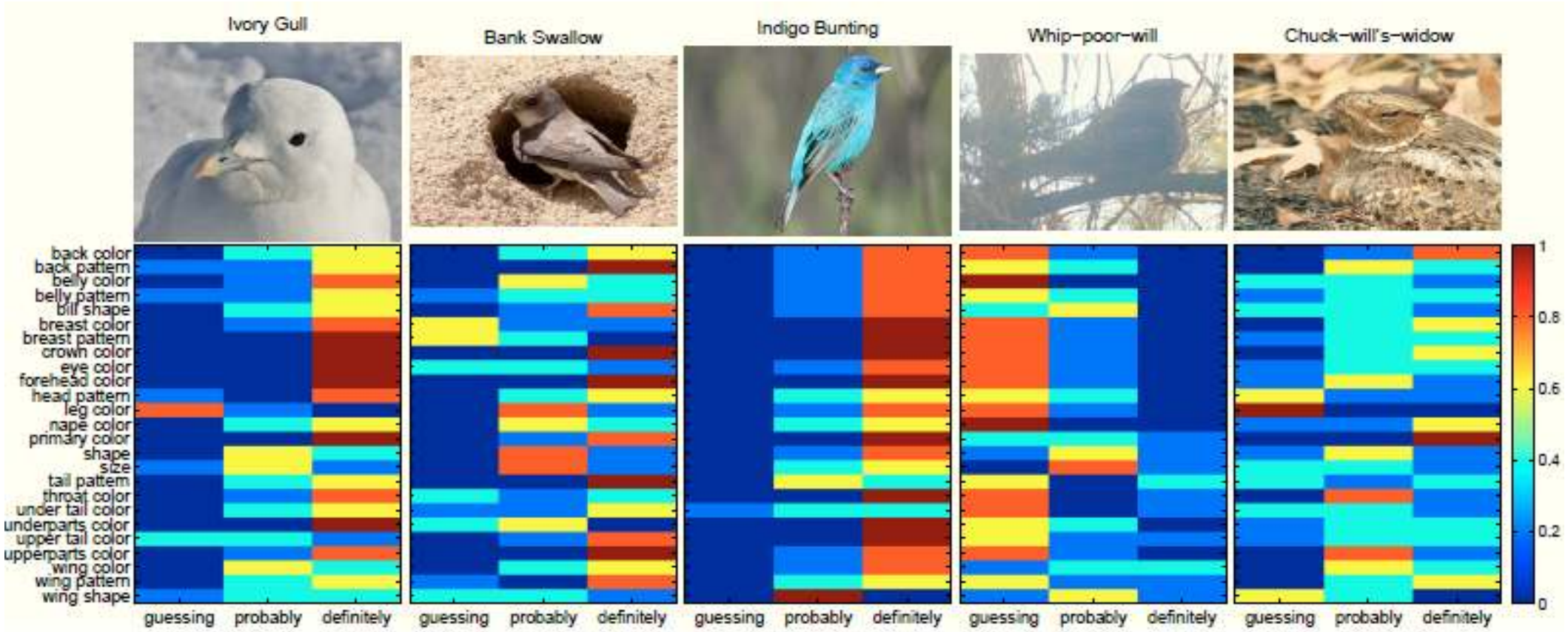
Results: With Computer Vision

Computer Vision Reduces Manual Labor: 11.1 →6.5 questions



S. Branson et al.

Results



S. Branson et al.

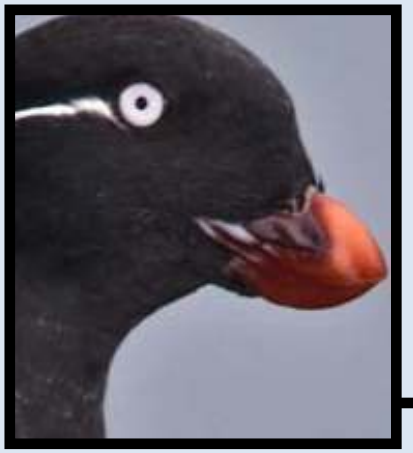
Recognition is Not Always Successful

Acadian Flycatcher

Unlimited questions

Least Flycatcher

Parakeet Auklet



Is the belly multi-colored?

yes (Def.)

Least Auklet

S. Branson et al.

Future Work

- Different recognition systems
- Different interaction systems
- Tailored systems specifically for this framework