Human Baseline for Zero-Shot Transfer Learning of Good Dogs

William Boag, Shashank Srikant
CSAIL. MIT

SIGTBD 2019, Cambridge, MA

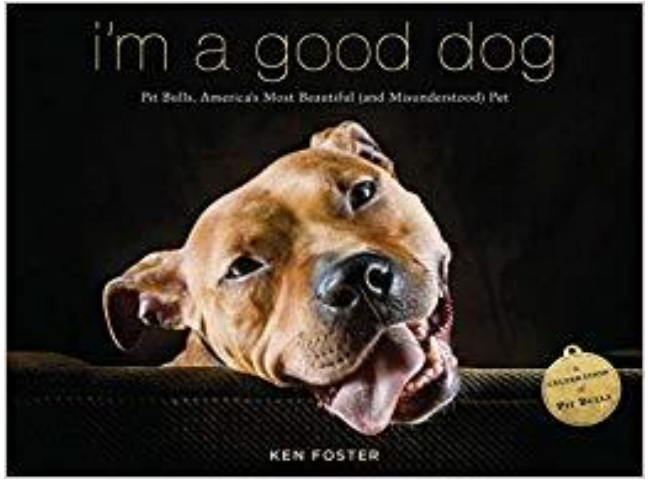
Thanks folks!

Kyungmi Lee, Lei Xu, and Frederic Koehler
RLE, CSAIL, Math
MIT

ML Across MIT

SIGTBD 2019

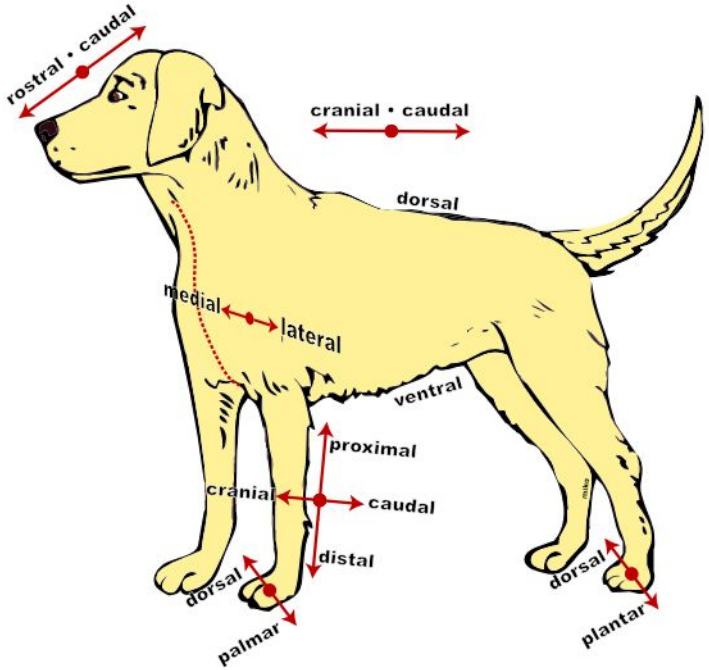
All dogs are good ...



… but, can’t meet them all, know them all

As a consequence

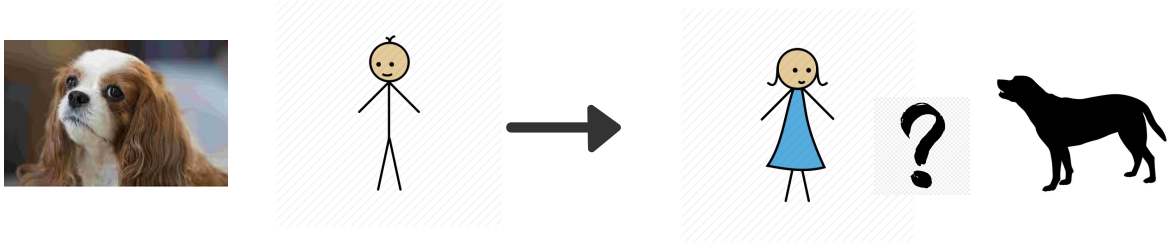
We rely on human descriptions mostly, when meeting new dogs



But how good are we at recognizing dogs from such descriptions?

Research question

- How good are we at discerning one dog (breed) from another, if we are provided descriptions of them by another human who has seen it.



Zero shot setting

- Further, for the human providing the descriptions, is it enough to have seen just a few images of the breed (< 5)?

Related work - Boag (2018).

Good Dogs of Cambridge

Willie Boag
MIT
wboag@mit.edu

Abstract

*The city of Cambridge is home to over 105,000 residents, yet there are only barely registered 5,000 dogs. With more than a 20-to-1 ratio of people-to-dogs, it is now more important than ever that we use big data in order to automatically identify where to find the best dogs. Because Cambridge is really h*ckin big, it can be incredibly costly to go to a neighborhood that lacks sufficient good dogs, only to realize you need to go elsewhere if you wanna be part of a high-quality belly rub. In this work, we propose a method for identifying which parts of Cambridge can maximize one's time with some big ole puppers.*

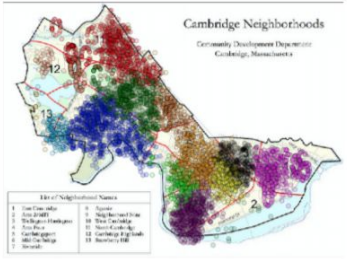
1. Background

We Rate Dogs is a Twitter account that rates pictures of dogs on a 1-10 scale. Since its inception in 2015, the account has gained over 2.2 million followers. Although most posts are of really good dogs, things were not always this way. The first dog ever rated, while still a very good pup, received a score of 8/10, as shown in Figure 1. This is not a slight against the nameless dog hero, but the low rating can certainly be understood as a reflection of the time. Nowadays, a rating like 8 is virtually unheard of. Nowadays, most ratings are either 12/10 or 13/10.

To date, @dog_rates has 6,762 posts, thus providing a rich dataset of thousands of gold truth annotations. By harnessing this data, we can model what would make my pet a really good boy. In this day and age, such an application of technology is essentially essential.

Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of relaxing on stair. 8/10 would pet

Figure 1: The first dog ever rated by We Rate Dogs.



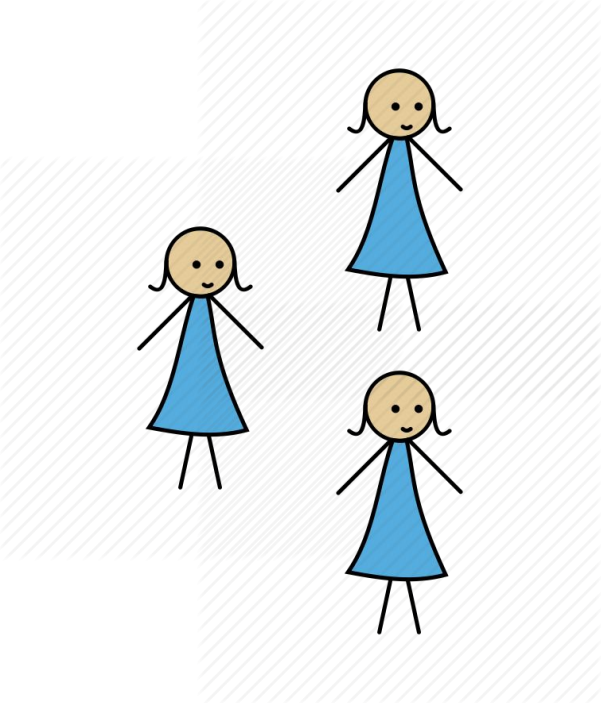
Related work - Karpathy (2015)

Niche research area.

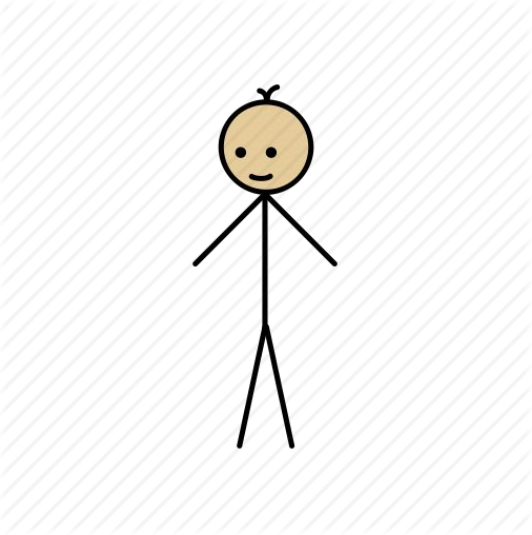
No explorers before this attempt.

Experiment setup

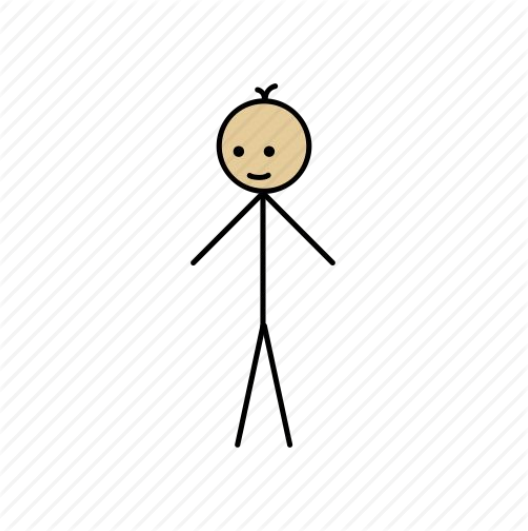
Group



Teacher



The Task



Teacher looks at few dog samples

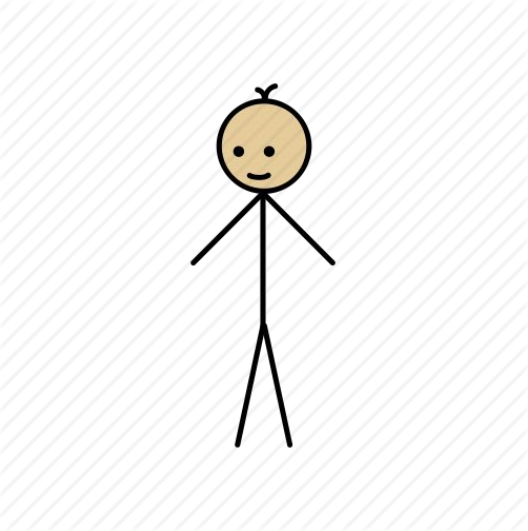
The Task

Train - Teacher Only

Affenpinscher

Afghan hound

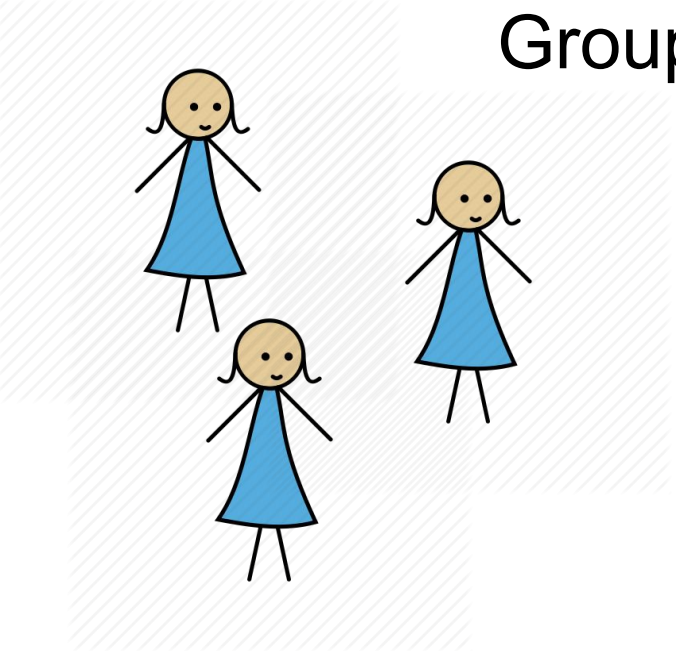
The Task



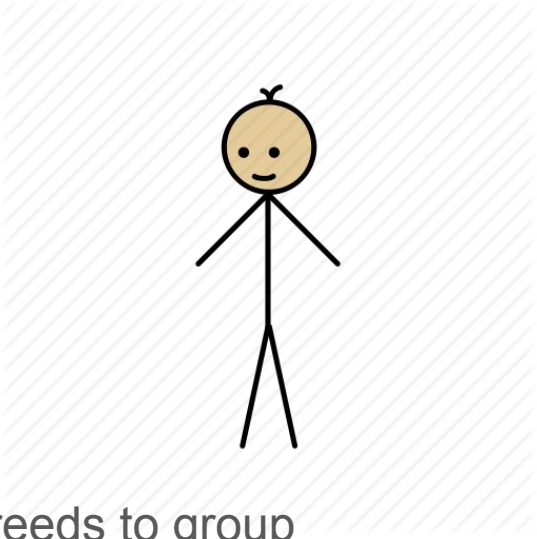
Teacher is given 30s to think about what s/he just saw

The Task

Group

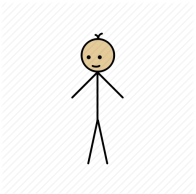


Teacher



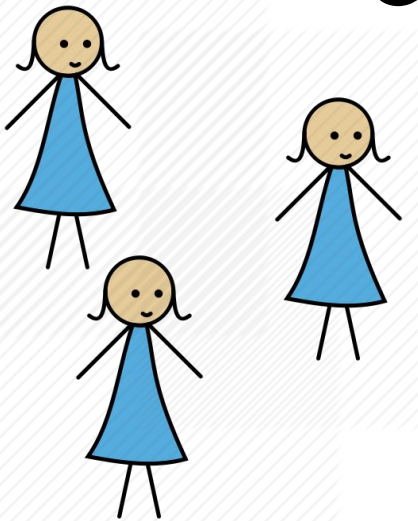
Teacher verbally describes the two breeds to group

The Task



Teacher asked to shut up

Group

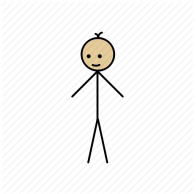


☐ Affenpinscher

☐ Afghan hound

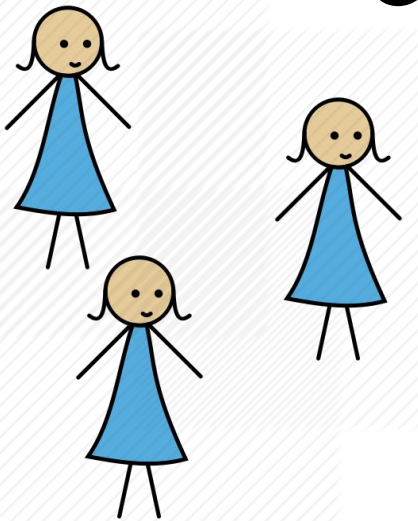
Group shown a random unseen image of one of the two breeds

The Task



Teacher asked to shut up

Group



☐ Affenpinscher

☐ Afghan hound

Group discusses and guesses one of the two breeds

Remember

Why zero shot?

Teacher just sees a few images of each of the doggas.

Why transfer learning?

Teacher has learnt a model and is now trying to transfer it to group.

What was hooman accuracy on this task?
Guesses?

Results

	Accuracy
Zero shot transfer learning (this work)	**87.5%**
Zero shot image classification (Karpathy et al's work on ImageNet)	**85 to 87%**

Discussion - Classification strategies

Each team had its own strategy to aggregate results from multiple classifiers.

Ensemble did not help much

Some chose to visually encode descriptions; others wrote down what they heard.

Discussion - Perils of correlation based predictors

Curly-coated retriever

Bouvier des Flandres

Discussion - Perils of correlation based predictors

- ☐ Curly-coated retriever
- ☐ Bouvier des Flandres

Discussion - The neuroscience of ZS transfer learning

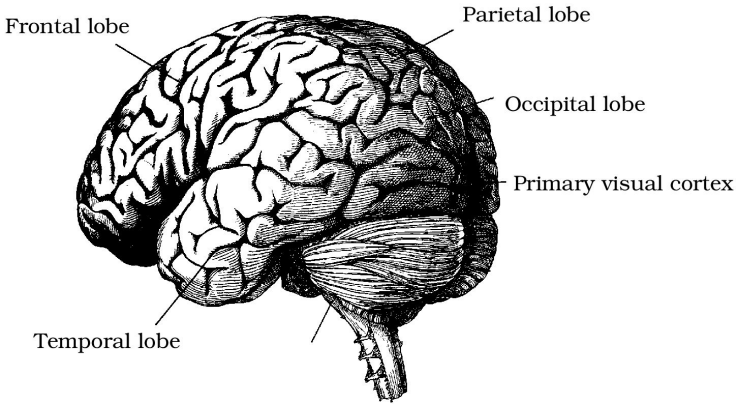
What did we observe.

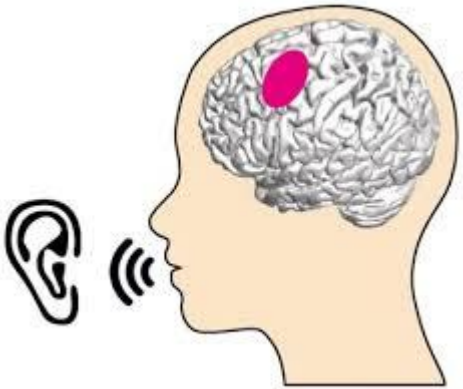
No team managed a 100% accuracy. Why?

Discussion - The neuroscience of ZS transfer learning

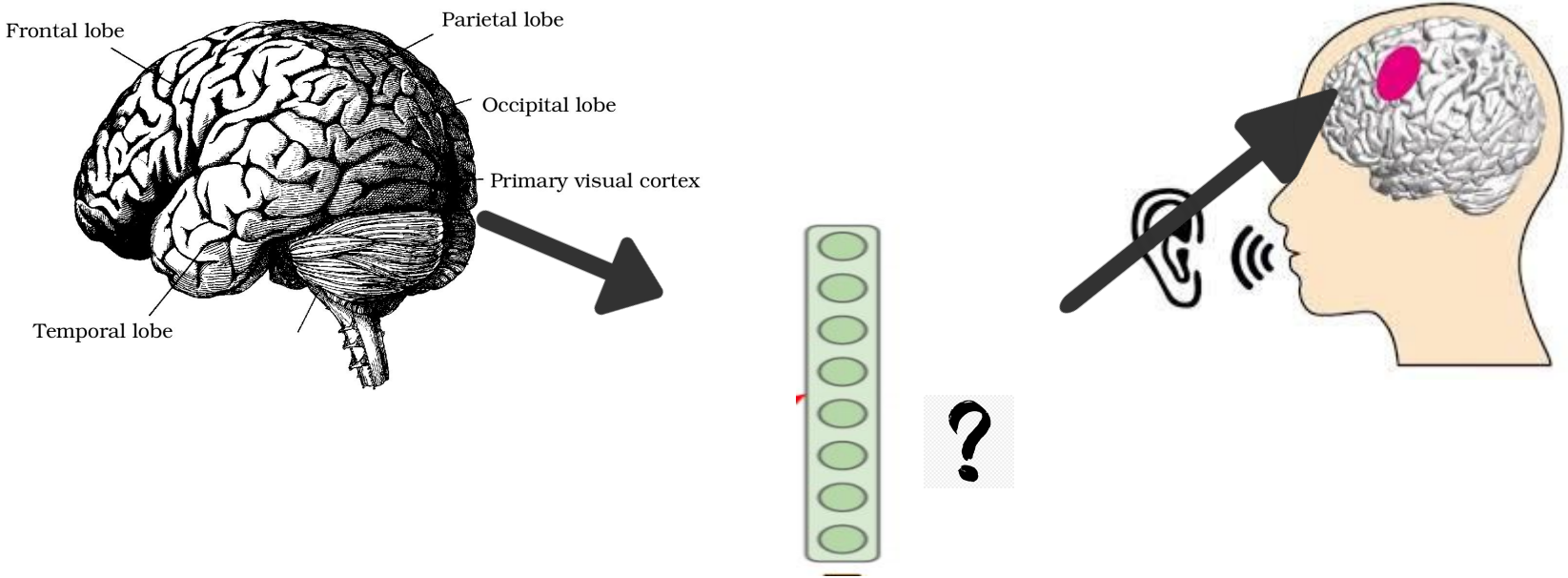
No team managed a 100% accuracy. Why?

One possible explanation





Discussion - The neuroscience of ZS transfer learning



Conclusion

All dogs are good. And will continue being so.

Humans are good at zero-shot learning of good dogs.

Opens up a bunch of research questions in this nascent field of DogAI