VQA Visual Question Answering Challenge

Yash Goyal
(Georgia Tech)

Aishwarya Agrawal
(Georgia Tech)

Dhruv Batra
(Georgia Tech / FAIR)

Devi Parikh
(Georgia Tech / FAIR)

Georgia Tech

facebook Artificial Intelligence Research

VQA Task

VQA Task

What is the mustache made of?

VQA Task

AI System

What is the mustache made of?

VQA Task

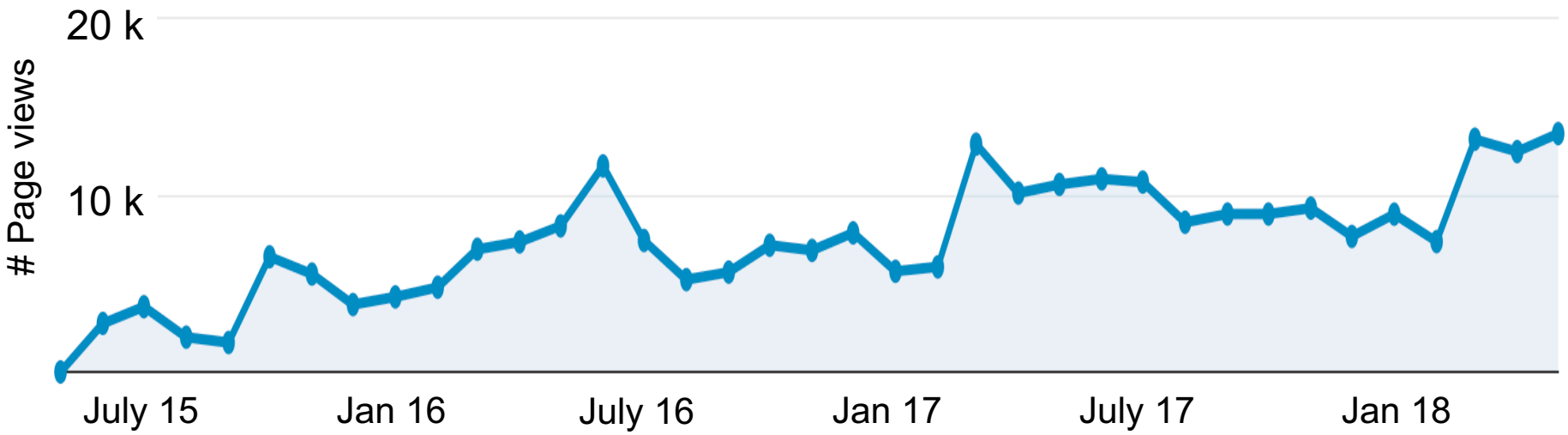
What is the mustache made of?

AI System

bananas

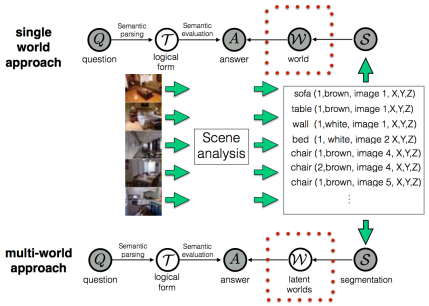
Interest in VQA

(http://www.visualqa.org/)

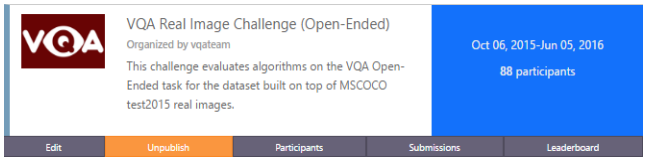


13k page views/month
during VQA Challenge 2018

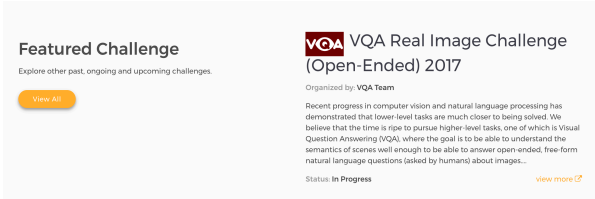
VQA Timeline



Geman et al., PNAS 14
Malinowski & Fritz, NIPS 14



1st VQA Challenge



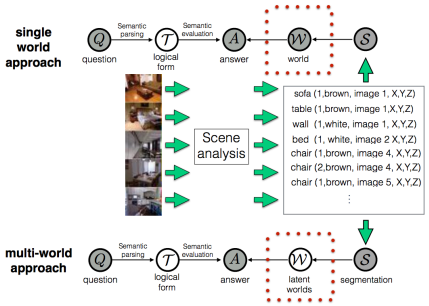
3rd VQA Challenge

2014 — 2015 — 2016 — 2017 — 2018

VQA v1.0 dataset released

VQA v2.0 released, 2nd VQA Challenge

VQA Timeline



Geman et al., PNAS 14
Malinowski & Fritz, NIPS 14

1st VQA Challenge

3rd VQA Challenge

2014 2015 2016 2017 2018

VQA v1.0 dataset released

VQA v2.0 released, 2nd VQA Challenge

VQA Timeline

Geman et al., PNAS 14

Malinowski & Fritz, NIPS 14

1st VQA Challenge

3rd VQA Challenge

2014 — **2015** — 2016 — 2017 — 2018

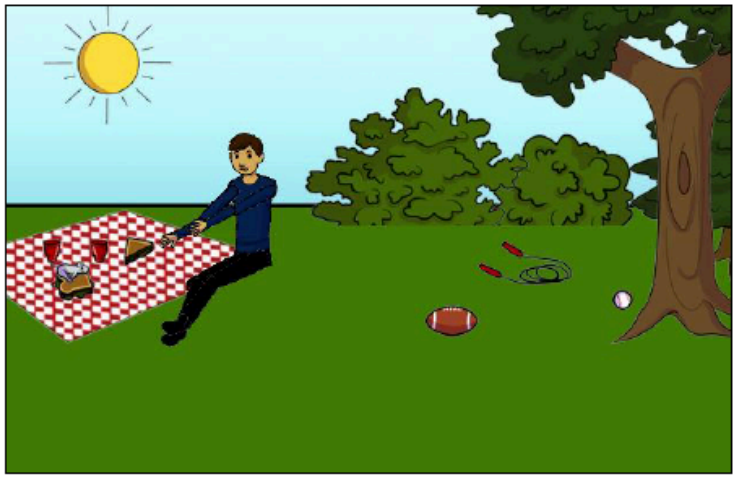
VQA v1.0 dataset released

VQA v2.0 released, 2nd VQA Challenge

VQA v1.0 Dataset

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

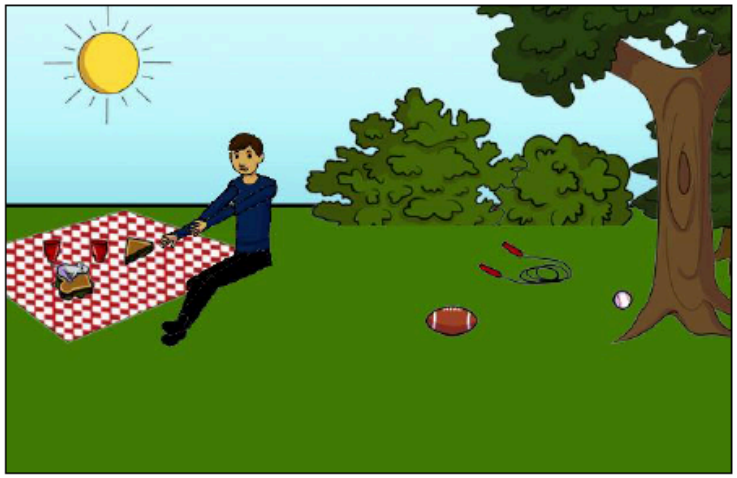
Does it appear to be rainy?
Does this person have 20/20 vision?

VQA v1.0 Dataset

About objects

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

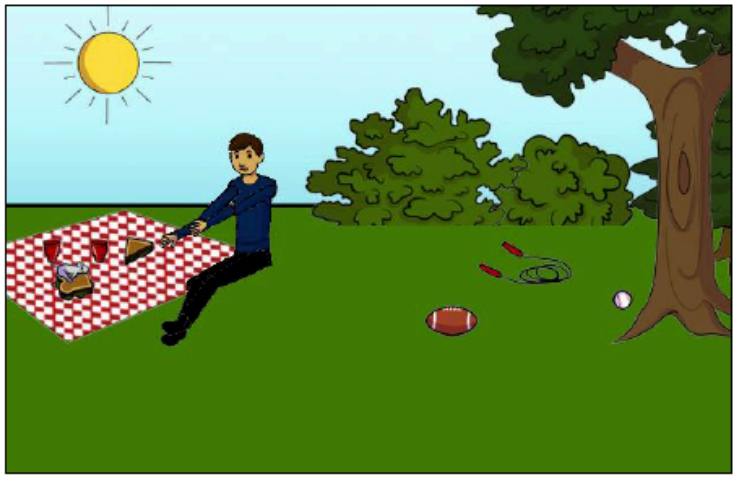
Does it appear to be rainy?
Does this person have 20/20 vision?

VQA v1.0 Dataset

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Fine-grained recognition



Is this person expecting company?
What is just under the tree?

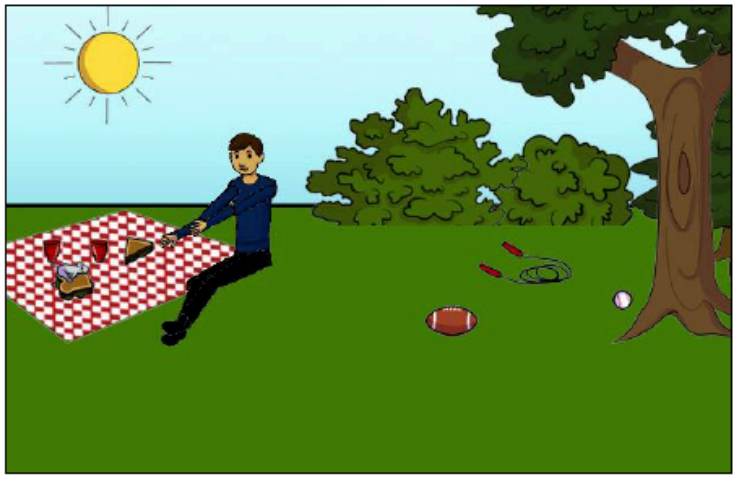
Does it appear to be rainy?
Does this person have 20/20 vision?

VQA v1.0 Dataset

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Counting



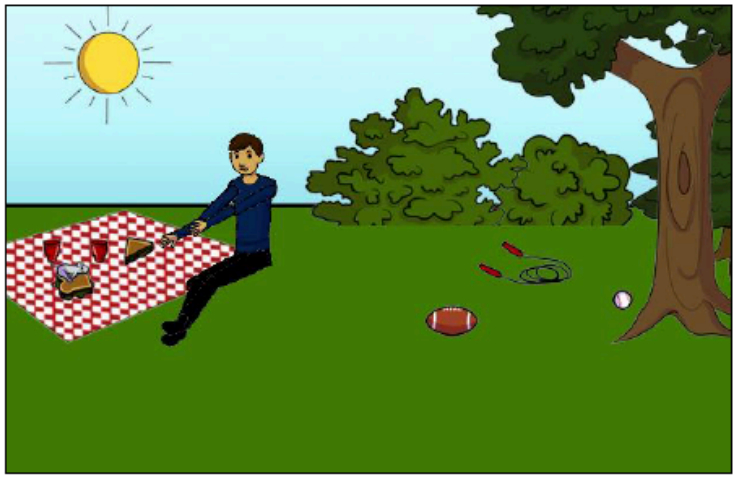
Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

VQA v1.0 Dataset

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

Common sense

VQA v1.0 Dataset Stats

- >200K images (COCO)

VQA v1.0 Dataset Stats

- >200K images (COCO)
- >600K questions (3 per image)

VQA v1.0 Dataset Stats

- >200K images (COCO)
- >600K questions (3 per image)
- ~8M answers (10 with image + 3 w/o image)

Accuracy Metric

$$\text{Acc}(ans) = \min\left\{\frac{\#\text{humans that said } ans}{3}, 1\right\}$$

1940. COCO_train2014_000000012015

Open-Ended/Multiple-Choice/Ground-Truth

Q: WHAT OBJECT IS THIS

Ground Truth Answers:

(1) television	(6) television
(2) tv	(7) television
(3) tv	(8) tv
(4) tv	(9) tv
(5) television	(10) television

Q: How old is this TV?

Ground Truth Answers:

(1) 20 years	(6) old
(2) 35	(7) 80 s
(3) old	(8) 30 years
(4) more than thirty years old	(9) 15 years
(5) old	(10) very old

Q: Is this TV upside-down?

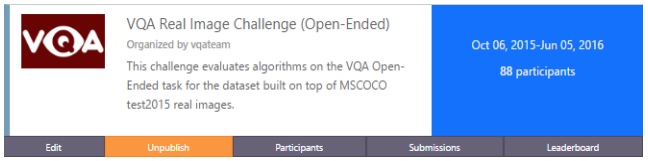
Ground Truth Answers:

(1) yes	(6) yes
(2) yes	(7) yes
(3) yes	(8) yes
(4) yes	(9) yes
(5) yes	(10) yes

VQA Timeline

Geman et al., PNAS 14

Malinowski & Fritz, NIPS 14



1st VQA Challenge

3rd VQA Challenge

2014

2015

2016

2017

2018

VQA v1.0 dataset released

VQA v2.0 released, 2nd VQA Challenge

VQA Challenge 2016
on www.codalab.org

VQA Real Image Challenge (Open-Ended)

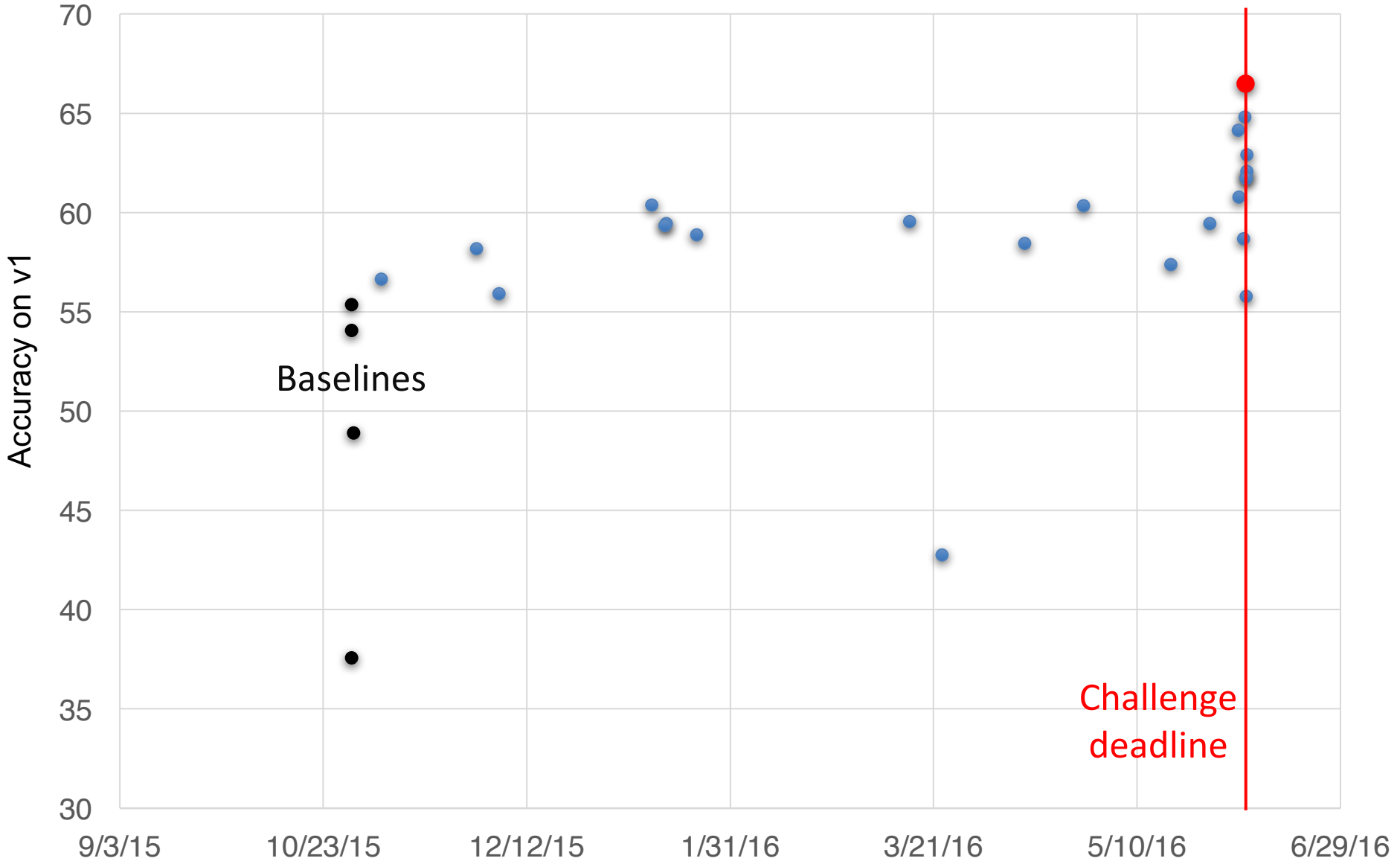
Organized by vqateam

This challenge evaluates algorithms on the VQA Open-Ended task for the dataset built on top of MSCOCO test2015 real images.

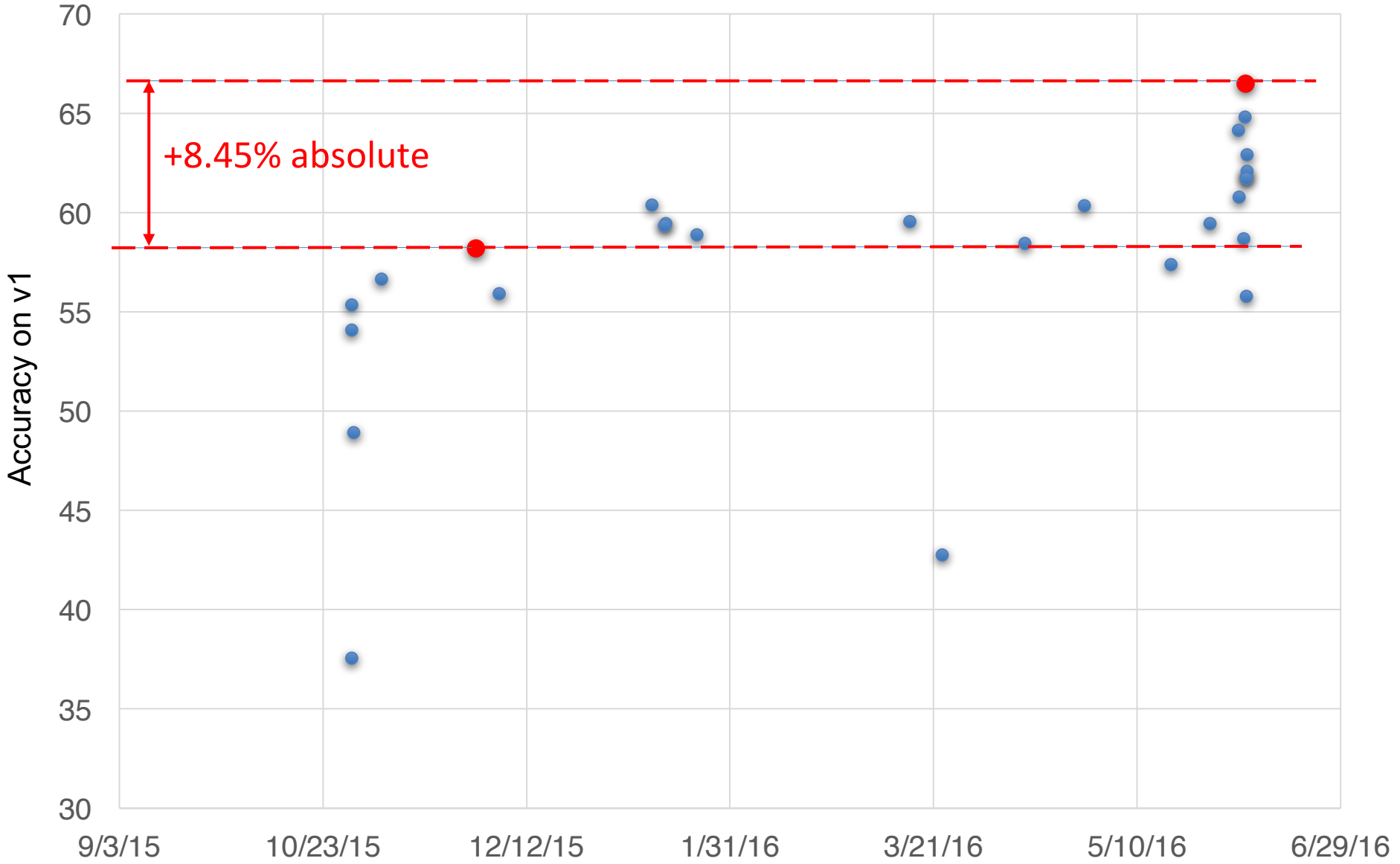
Oct 06, 2015-Jun 05, 2016

438 participant s

VQA Challenge 2016



VQA Challenge 2016



VQA Timeline

Geman et al., PNAS 14

Malinowski & Fritz, NIPS 14

1st VQA Challenge

3rd VQA Challenge

2014

2015

2016

2017

2018

VQA v1.0 dataset released

VQA v2.0 released, 2nd VQA Challenge

VQA v2.0 Dataset

Who is wearing glasses?

Similar images

man

Different answers

woman

VQA v2.0 Dataset Stats

- >200K images

VQA v2.0 Dataset Stats

- >200K images
- >1.1M questions

VQA v2.0 Dataset Stats

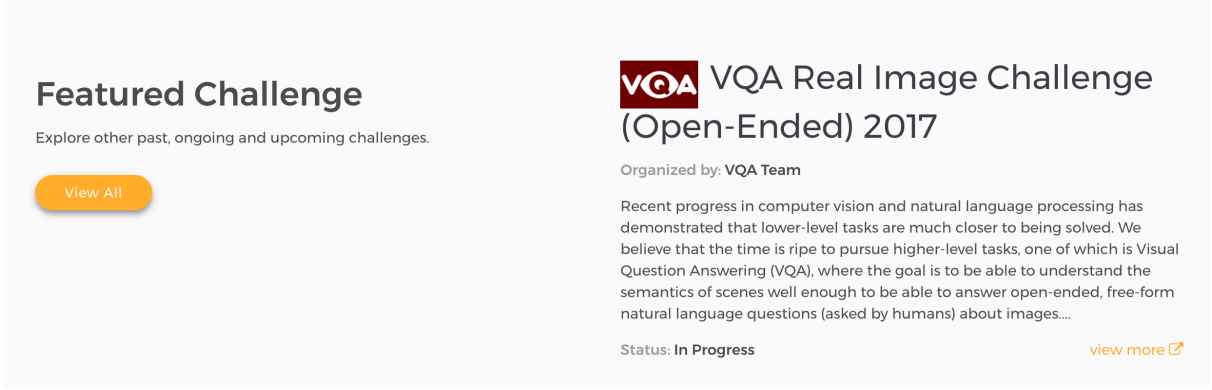
- >200K images
- >1.1M questions
- >11M answers

VQA v2.0 Dataset Stats

- >200K images
- >1.1M questions
- >11M answers

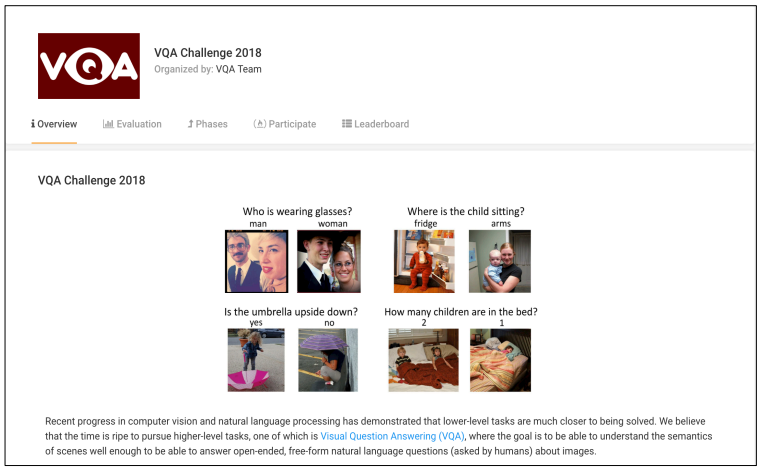
1.8 x VQA v1.0

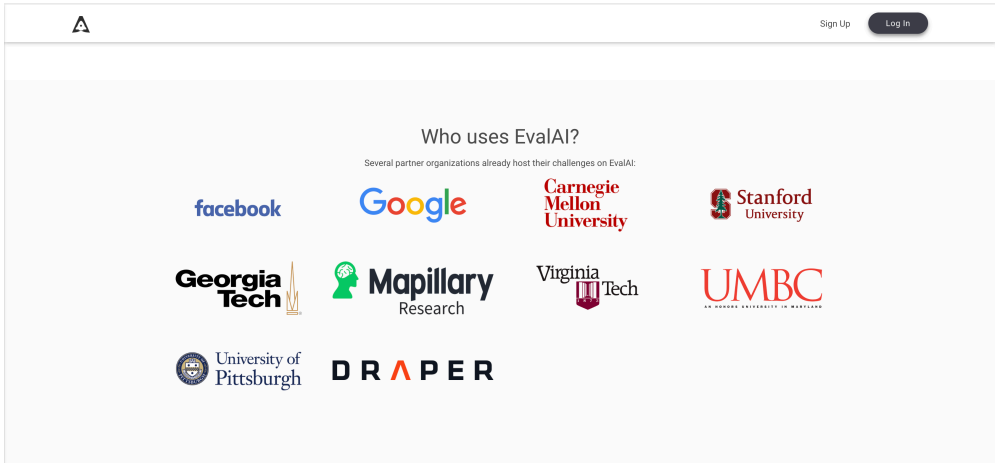
VQA Challenge 2017 on https://evalai.cloudcv.org/



EvalAI: Evaluating state of the art in AI

Open source platform that helps researchers to create, collaborate and participate in various AI challenges

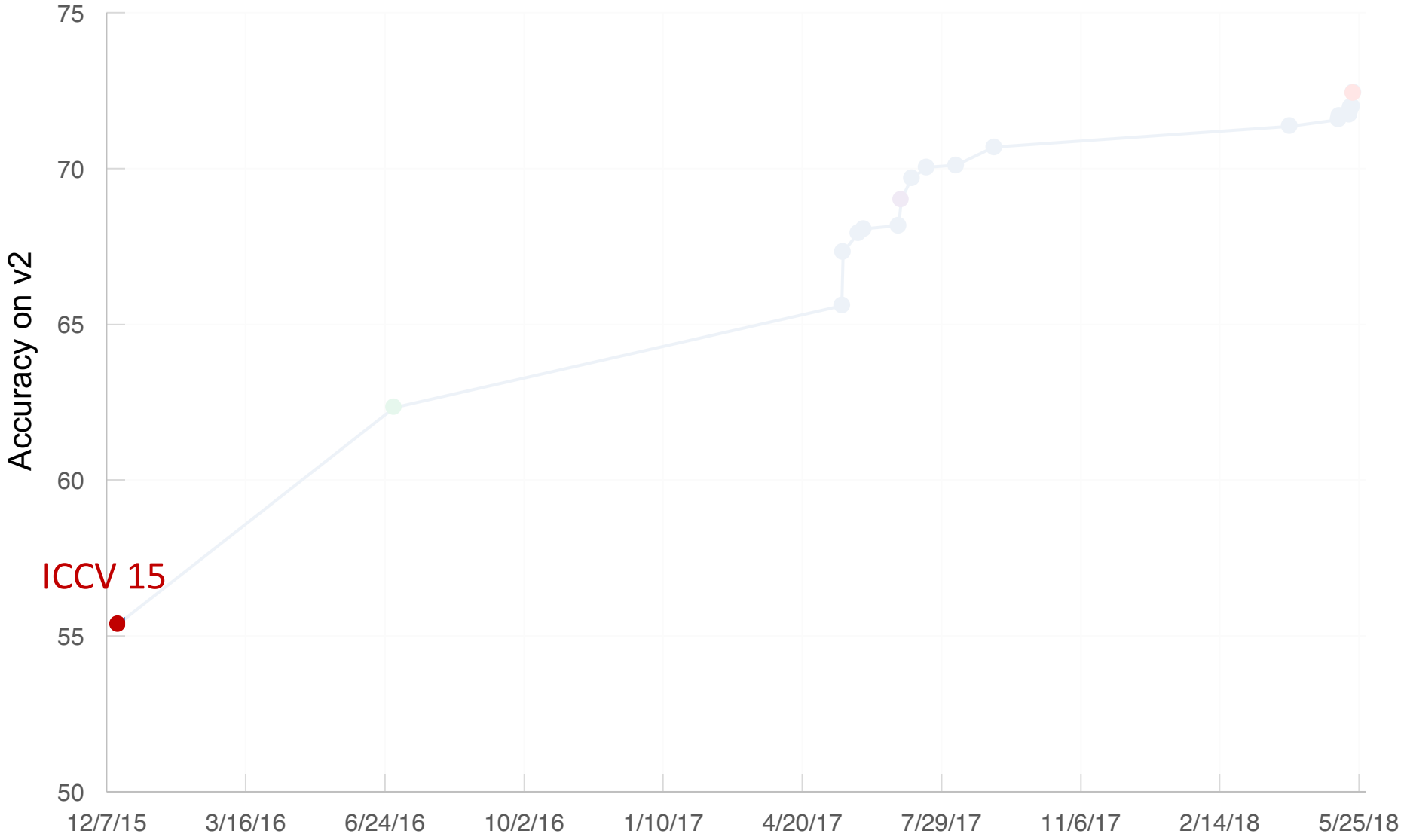




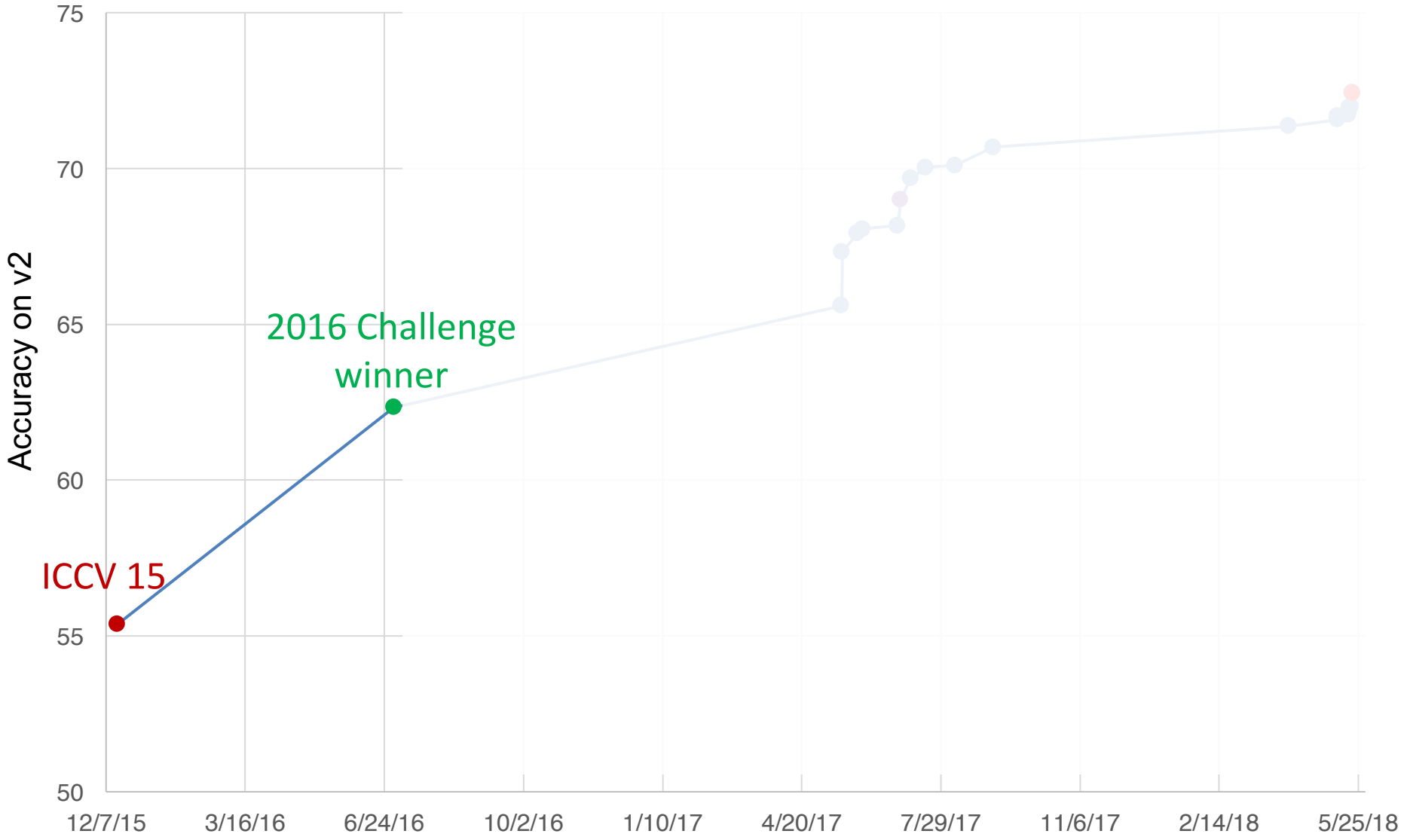
Contact: **deshraj@gatech.edu**

https://evalai.cloudcv.org/

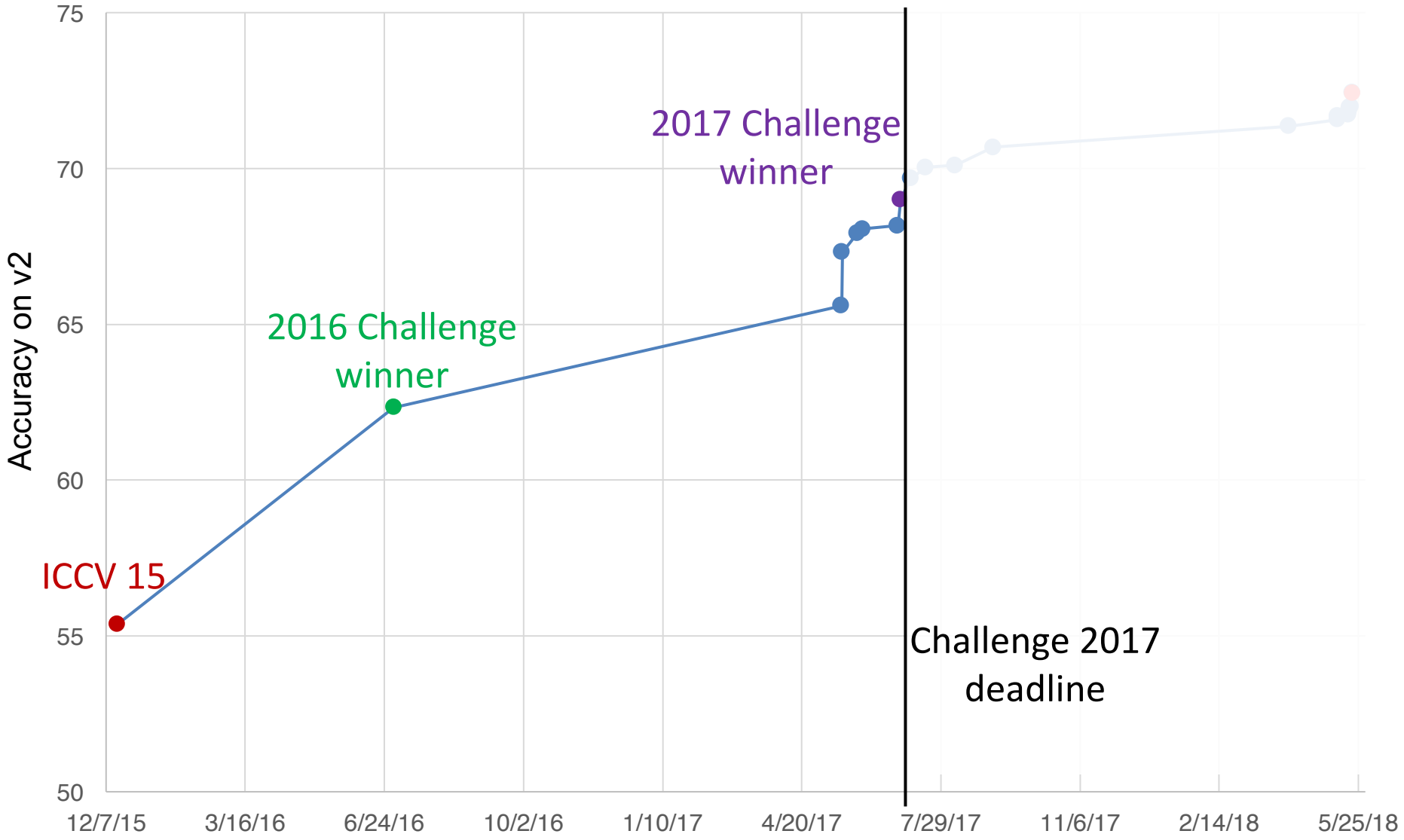
Benchmarking on VQA v2.0



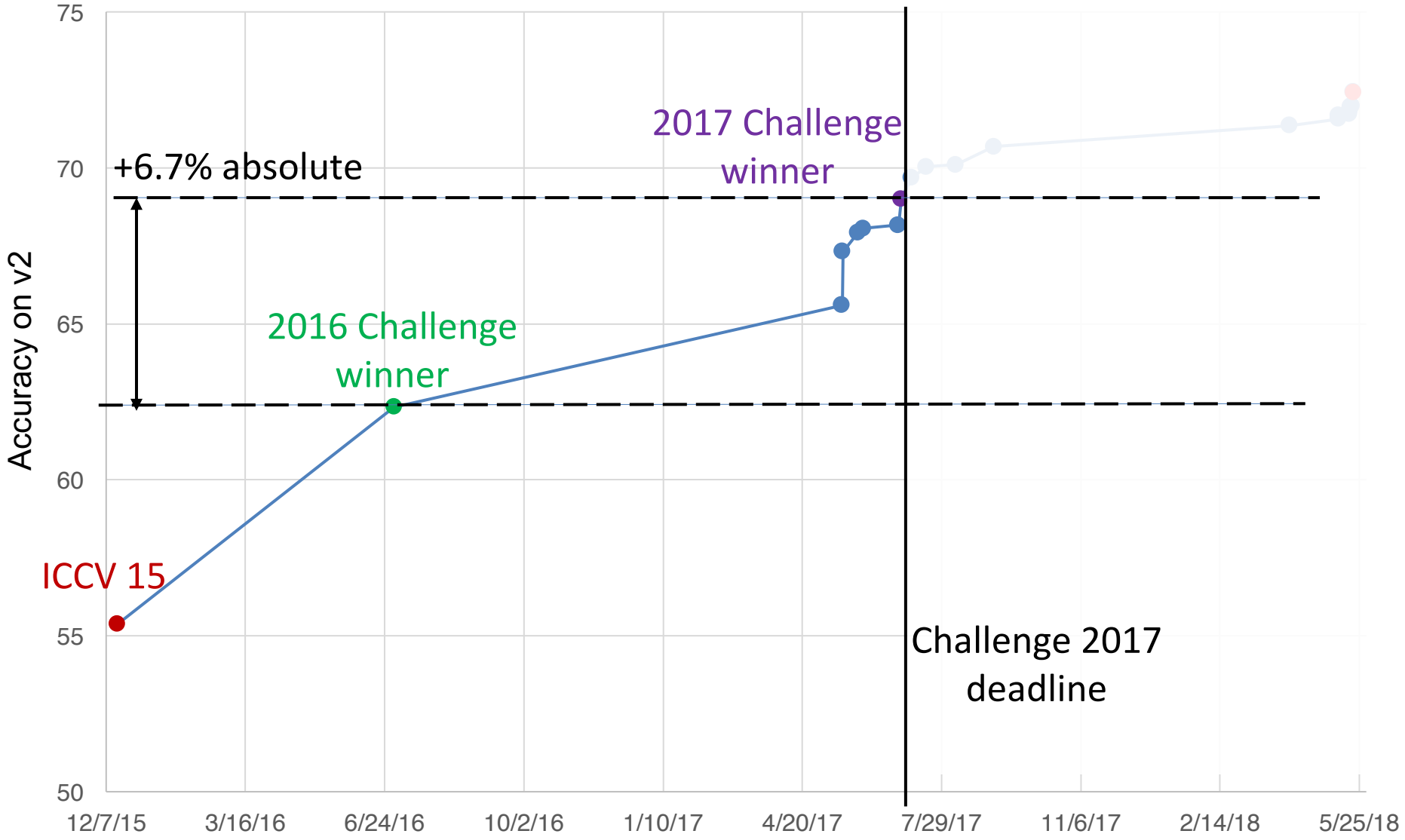
Benchmarking on VQA v2.0



VQA Challenge 2017



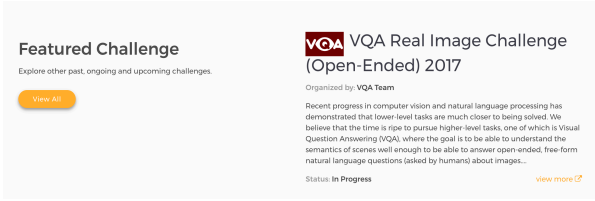
VQA Challenge 2017



VQA Timeline

Geman et al., PNAS 14
Malinowski & Fritz, NIPS 14

1st VQA Challenge



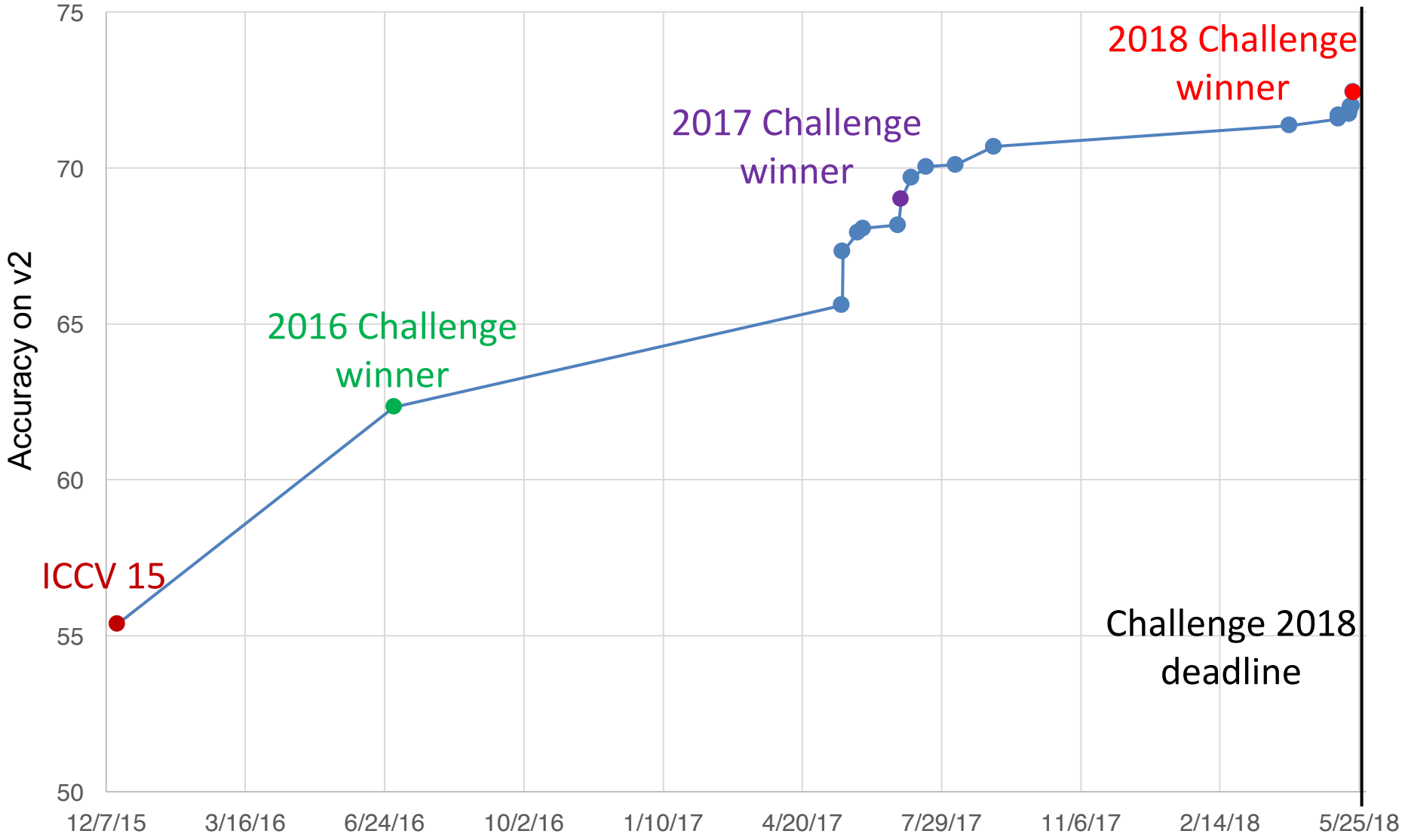
3rd VQA Challenge

2014 — 2015 — 2016 — 2017 — 2018

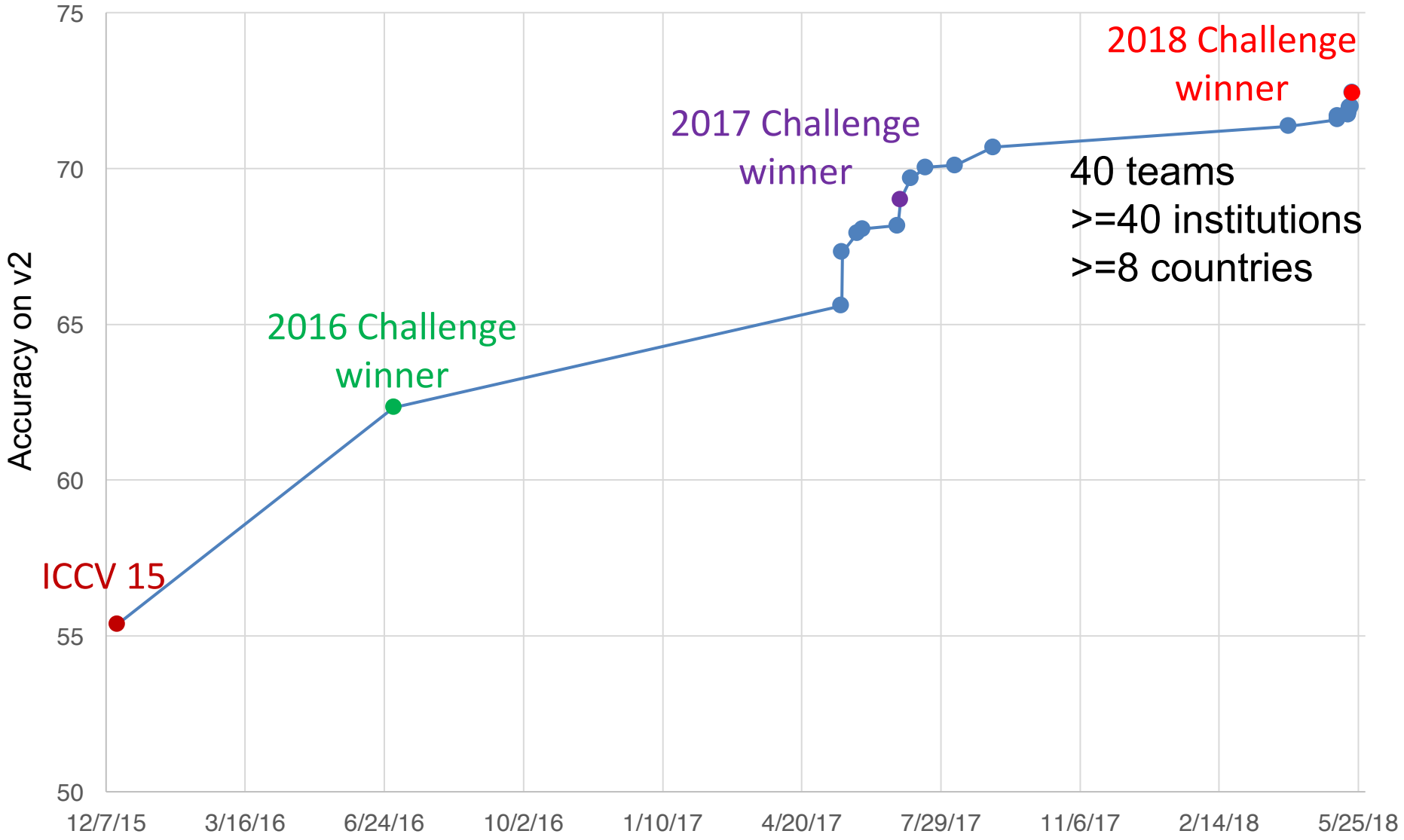
VQA v1.0 dataset released

VQA v2.0 released, 2nd VQA Challenge

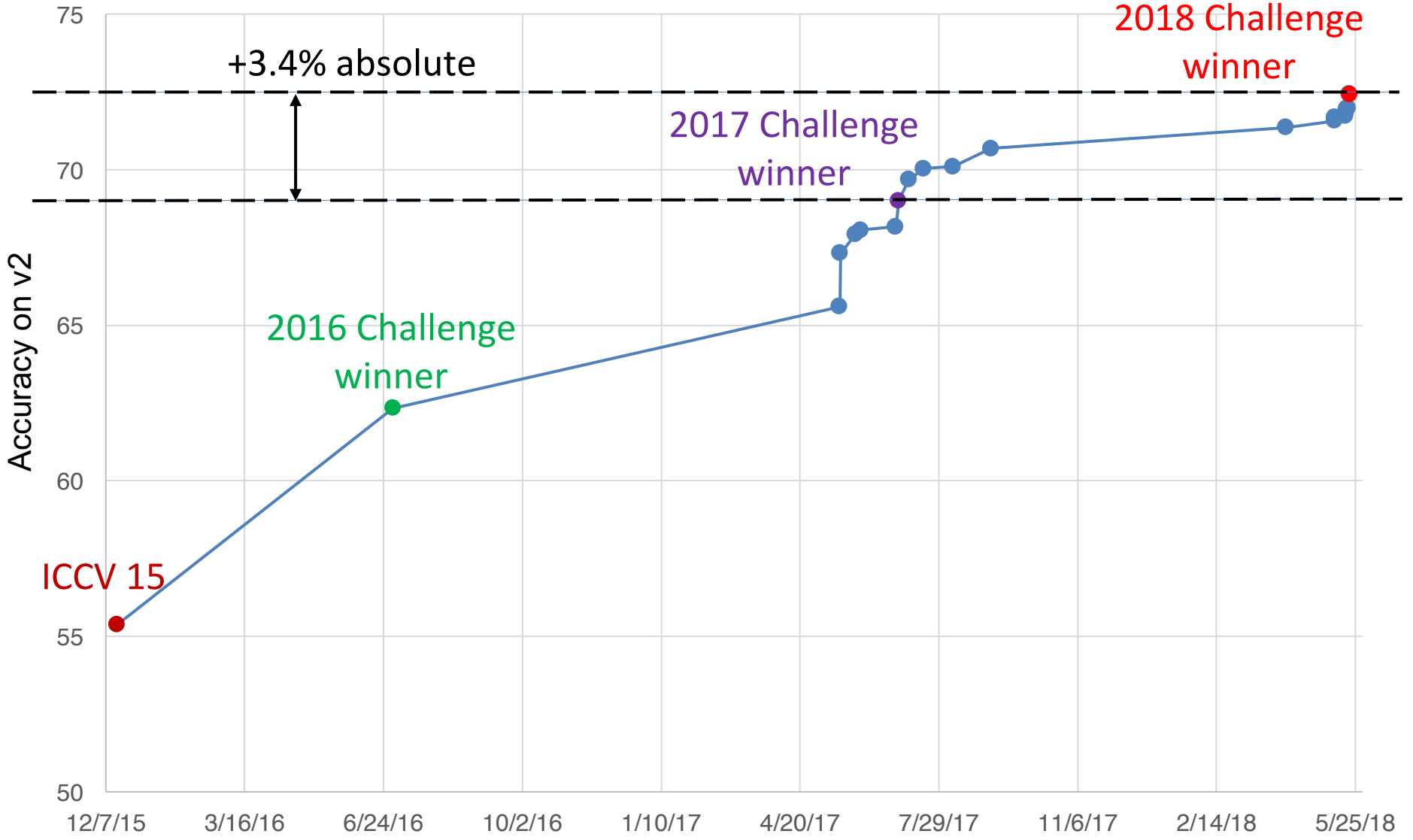
VQA Challenge 2018



VQA Challenge 2018



VQA Challenge 2018

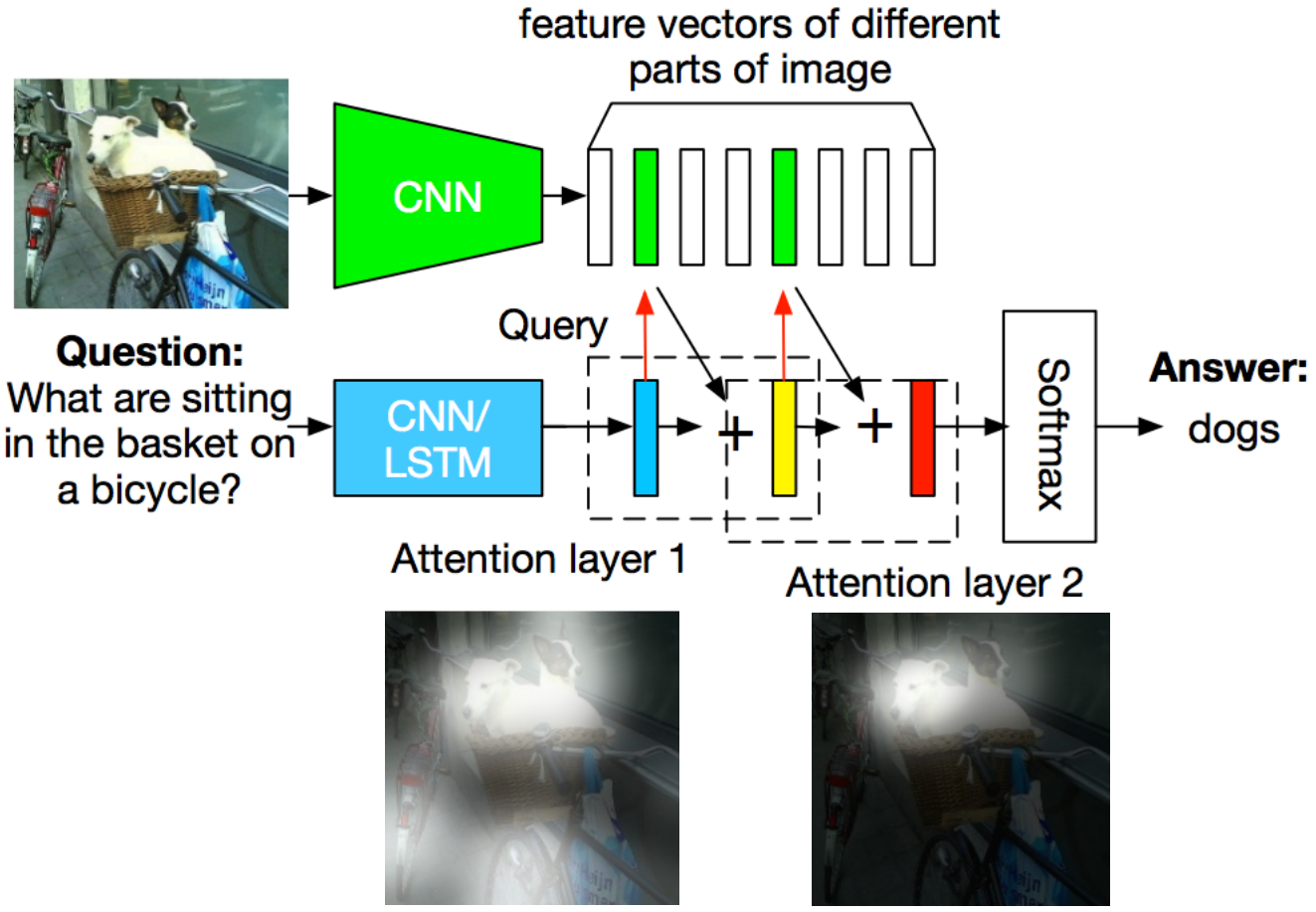


What has the community gained?

What has the community gained?

- Attention Networks

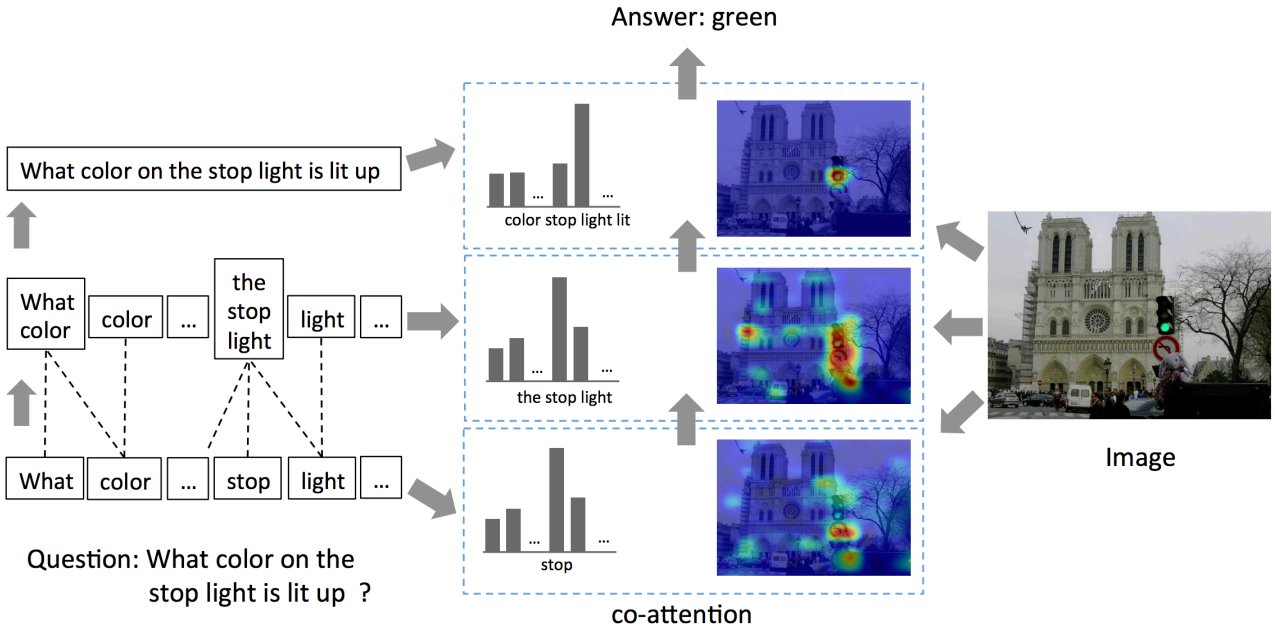
Stacked Attention Networks.
Yang et al., CVPR 16



What has the community gained?

- Attention Networks
- Hierarchical Co-attention

Hierarchical Question-Image Co-Attention
Lu et al., NIPS 16

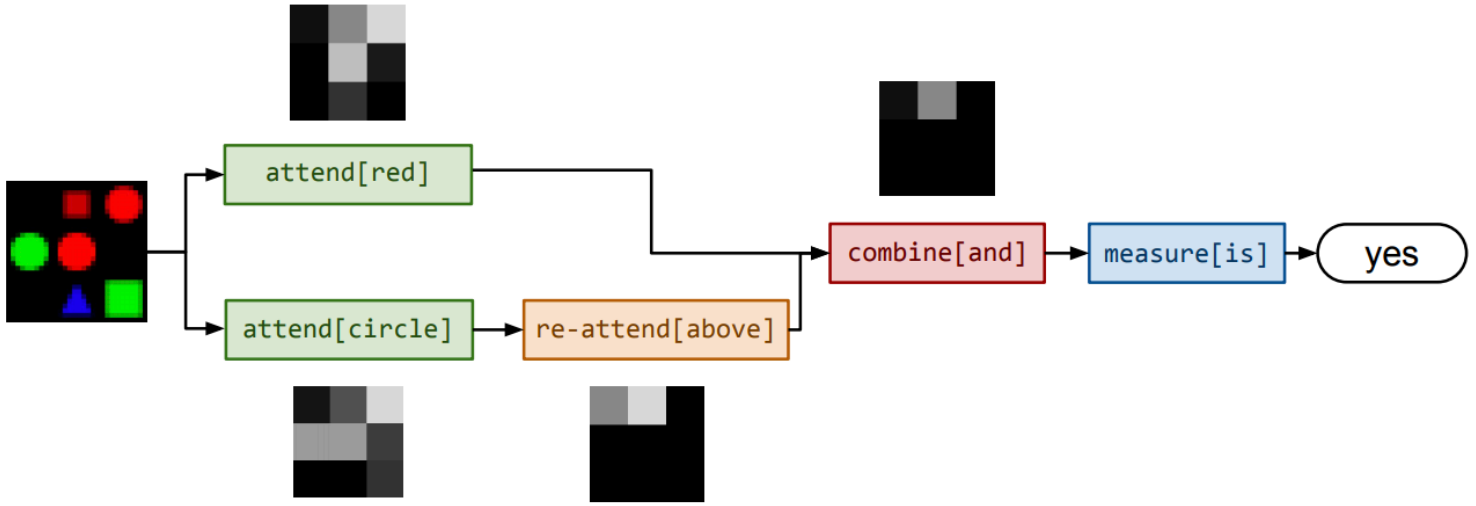


What has the community gained?

- Attention Networks
- Compositionality

- Hierarchical Co-attention

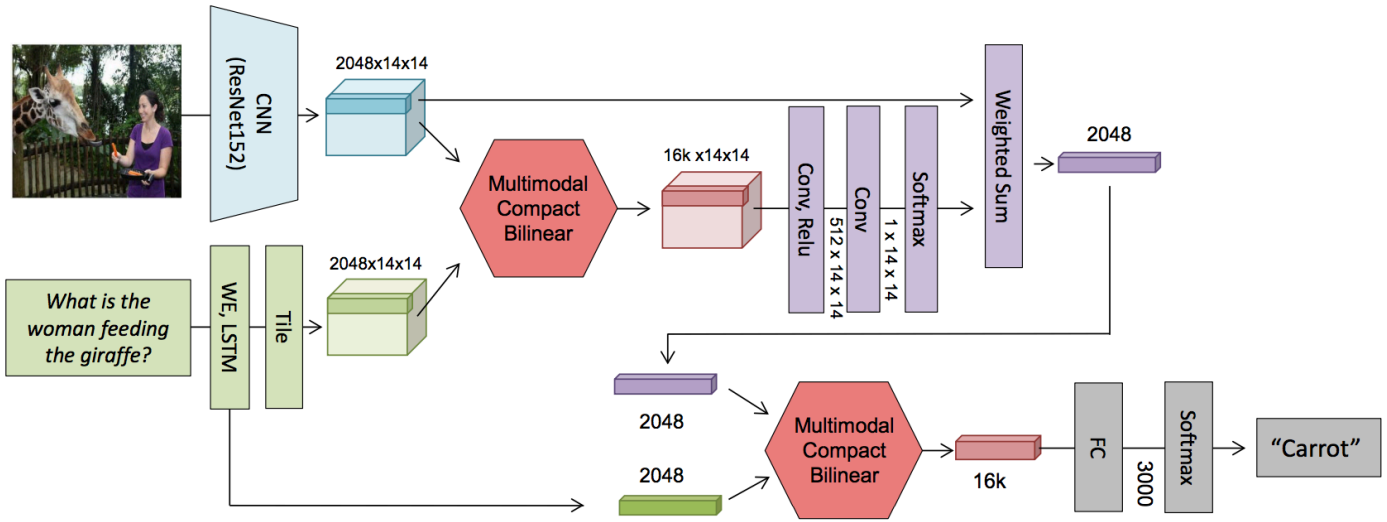
Neural Module Networks
Andreas et al., CVPR 16



What has the community gained?

- Attention Networks
- Compositionality
- Hierarchical Co-attention
- Multi-modal Pooling

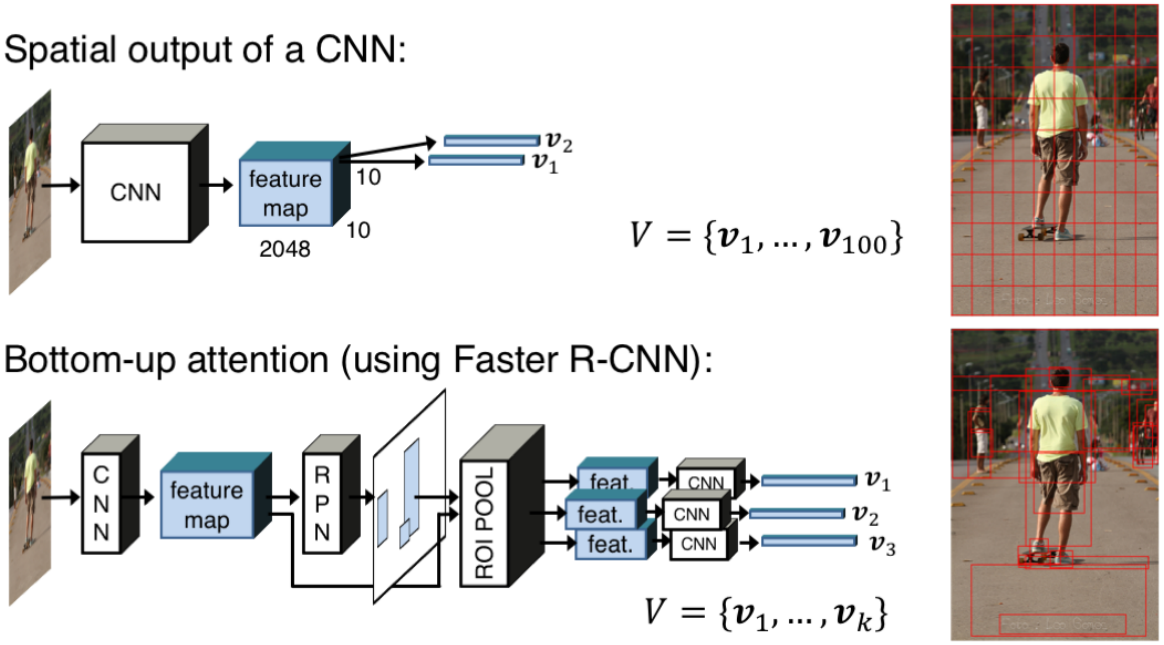
Multimodal Compact Bilinear Pooling
Fukui et al., EMNLP 16



What has the community gained?

- Attention Networks
- Compositionality
- Bottom-up Top-down attention
- Hierarchical Co-attention
- Multi-modal Pooling

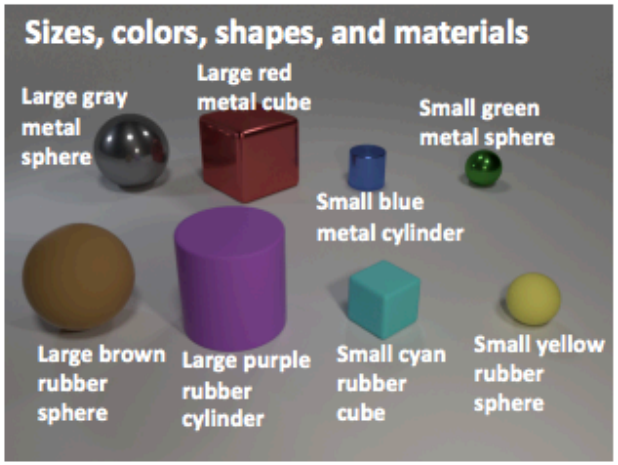
Bottom-Up and Top-Down Attention
Anderson et al., CVPR 18

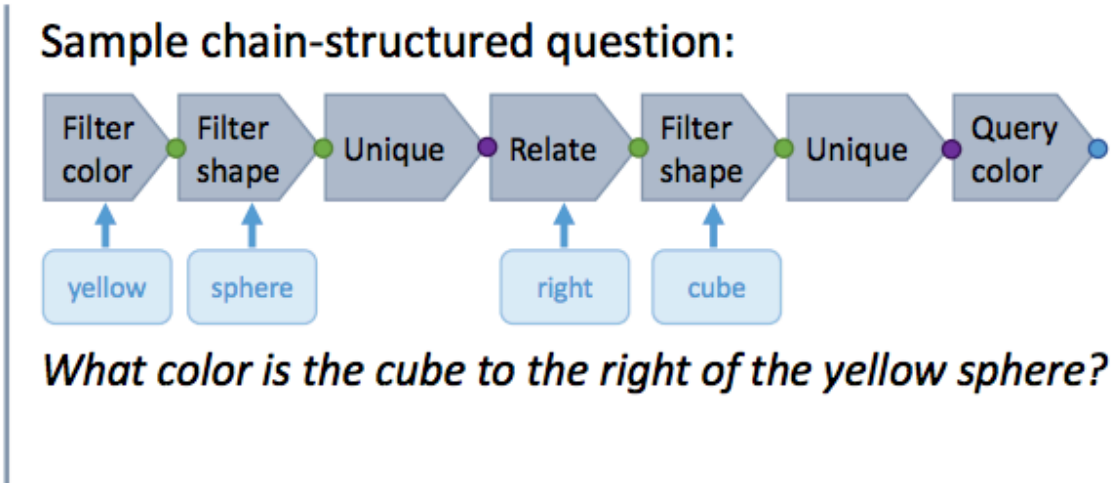


What has the community gained?

- Attention Networks
- Compositionality
- Bottom-up Top-down attention
- Hierarchical Co-attention
- Multi-modal Pooling
- Visual Reasoning

CLEVR
Johnson et al., CVPR 17





Visual Dialog Challenge 2018

VisDial v1.0

- ~130k images (COCO)
- 10-round dialog / image
- ~1.3 million QA pairs
- Evaluation
  - Automatic metrics
  - Human annotations

- Deadline: **mid-August, 2018**
- Results: **September 8th, 2018** at ECCV 2018

visualdialog.org/challenge/2018

Aishwarya Agrawal
Georgia Tech

Stanislaw Antol
Traptic, Inc.

Jiasen Lu
Georgia Tech

Tejas Khot
CMU

Peng Zhang
Amazon

Akrit Mohapatra
Virginia Tech

Douglas Summers-Stay
Army Research Lab

Meg Mitchell
Google Research

Larry Zitnick
FAIR

Dhruv Batra
Georgia Tech / FAIR

Devi Parikh
Georgia Tech / FAIR

Sponsors


NSF


DEPARTMENT OF THE NAVY
USN
ONR
Science & Technology


ARL


SCIENCE TO SHAPE THE FUTURE OF THE ARMY
ARMY RESEARCH OFFICE
1951


Google
Faculty Research Awards


amazon
web services™


Windows Azure


NVIDIA


Google
SUMMER
of CODE
2015


facebook


Google


THE PAUL G. ALLEN
FAMILY FOUNDATION


ICTAS
INSTITUTE for CRITICAL TECHNOLOGY
and APPLIED SCIENCE Virginia Tech

Thanks!

Questions?

Email: visualqa@gmail.com