Aishwarya Agrawal (Georgia Tech)

Yash Goyal (Georgia Tech)

Aishwarya Agrawal (Georgia Tech)
Outline

Overview of Task and Dataset

Overview of Challenge

Winner Announcements

Analysis of Results
Outline

Overview of Task and Dataset

Overview of Challenge

Winner Announcements

Analysis of Results
Outline

Overview of Task and Dataset

Overview of Challenge

Winner Announcements

Analysis of Results
Outline

Overview of Task and Dataset

Overview of Challenge

Winner Announcements

Analysis of Results
VQA Task
VQA Task

What is the mustache made of?
What is the mustache made of?
VQA Task

What is the mustache made of?

AI System

bananas

What is the mustache made of?
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?

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

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

Fine-grained recognition
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

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 v2.0 Dataset

Who is wearing glasses?
- man
- woman

Where is the child sitting?
- fridge
- arms

Is the umbrella upside down?
- yes
- no

How many children are in the bed?
- 2
- 1
Who is wearing glasses?

Similar images

man

Different answers

woman

VQA v1.0

New in VQA v2.0
VQA v2.0 Dataset Stats

• >200K images

• >1.1M questions

• >11M answers

1.8 x VQA v1.0
Accuracy Metric

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

Overview of Task and Dataset

Overview of Challenge

Winner Announcements

Analysis of Results
VQA Challenge on
https://evalai.cloudcv.org/

Featured Challenge
Explore other past, ongoing and upcoming challenges.

VQA Challenge 2018
Organized by: VQA Team
Recent progress in computer vision and natural language processing has demonstrated that lower-level tasks are much closer to being solved. We believe that the time is ripe to pursue higher-level tasks, one of which is Visual Question Answering (VQA), where the goal is to be able to understand the semantics of scenes well enough to be able to answer open-ended, free-form natural language questions (asked by humans) about images....
Status: In Progress

View All
## Dataset splits

<table>
<thead>
<tr>
<th></th>
<th>Images</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>80K</td>
<td>443K</td>
<td>4.4M</td>
</tr>
</tbody>
</table>

Dataset size is approximate
## Dataset splits

<table>
<thead>
<tr>
<th></th>
<th>Images</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td>80K</td>
<td>443K</td>
<td>4.4M</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>40K</td>
<td>214K</td>
<td>2.1M</td>
</tr>
</tbody>
</table>

*Dataset size is approximate*
## Dataset splits

<table>
<thead>
<tr>
<th></th>
<th>Images</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td>80K</td>
<td>443K</td>
<td>4.4M</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>40K</td>
<td>214K</td>
<td>2.1M</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td>80K</td>
<td>447K</td>
<td></td>
</tr>
</tbody>
</table>

Dataset size is approximate
Test Dataset

- 4 splits of approximately equal size

- **Test-dev** *(development)*
  - Debugging and Validation.

- **Test-standard** *(publications)*
  - Used to score entries for the Public Leaderboard.

- **Test-challenge** *(competitions)*
  - Used to rank challenge participants.

- **Test-reserve** *(check overfitting)*
  - Used to estimate overfitting. Scores on this set are never released.

Slide adapted from: MSCOCO Detection/Segmentation Challenge, ICCV 2015
Challenge Stats

- 40 teams
- >=40 institutions*
- >=8 countries*

*Statistics based on teams that have replied
Challenge Runner-Ups

Joint Runner-Up Team 1

SNU-BI

Jin-Hwa Kim (Seoul National University)
Jaehyun Jun (Seoul National University)
Byoung-Tak Zhang (Seoul National University & Surromind Robotics)

Challenge Accuracy: 71.69
Challenge Runner-Ups

Joint Runner-Up Team 2

HDU-UCAS-USYD

Zhou Yu (Hangzhou Dianzi University, China)
Jun Yu (Hangzhou Dianzi University, China)
Chenchao Xiang (Hangzhou Dianzi University, China)
Liang Wang (Hangzhou Dianzi University, China)
Dalu Guo (The University of Sydney, Australia)
Qingming Huang (University of Chinese Academy of Sciences)
Jianping Fan (Hangzhou Dianzi University, China)
Dacheng Tao (The University of Sydney, Australia)

Challenge Accuracy: 71.91
Challenge Winner

FAIRIA*
Yu Jiang† (Facebook AI Research)
Vivek Natarajan† (Facebook AI Research)
Xinlei Chen† (Facebook AI Research)
Marcus Rohrbach (Facebook AI Research)
Dhruv Batra (Facebook AI Research & Georgia Tech)
Devi Parikh (Facebook AI Research & Georgia Tech)

Challenge Accuracy: **72.41**

† equal contribution
Outline

Overview of Task and Dataset

Overview of Challenge

Winner Announcements

Analysis of Results
Challenge Results
Challenge Results
Challenge Results

+3.4% absolute
Statistical Significance

- Bootstrap samples 5000 times
- @ 95% confidence
Easy vs. Difficult Questions
Easy vs. Difficult Questions

Percentage of questions correctly answered by teams

Number of top 10 teams

0/10 1/10 2/10 3/10 4/10 5/10 6/10 7/10 8/10 9/10 10/10
82.5% of questions can be answered by at least 1 method!
Easy vs. Difficult Questions

Percentage of questions correctly answered by teams:

For Easy Questions, most of the top 10 teams achieved 10/10.

For Difficult Questions, fewer teams achieved 10/10, with many teams scoring lower.
Easy vs. Difficult Questions

Percentage of questions correctly answered by teams

Number of top 10 teams

0/10 1/10 2/10 3/10 4/10 5/10 6/10 7/10 8/10 9/10 10/10

2016 2017 2018
Difficult Questions with Rare Answers
Difficult Questions with Rare Answers

- What is the name of ...
- What is the number on ...
- What is written on the ...
- What does the sign ...
- What time is it?
- What kind of ...
- What type of ...
- Why is the ...
Easy vs. Difficult Questions
Easy vs. Difficult Questions

Difficult Questions with Frequent Answers

Easy Questions
Answer Type Analyses

- SNU_BI performs the best for “number” questions
Results on “number” questions
Answer Type Analyses

• SNU_BI performs the best for “number” questions

• No team statistically significantly better than the winner team for “yes/no” and “other”
Are models sensitive to subtle changes in images?

Who is wearing glasses?

Similar images

Different answers

man

woman
Are models sensitive to subtle changes in images?

- Are predictions different for complementary images?
- Are predictions accurate for complementary images?
Are predictions **different** for complementary images?
Are predictions **accurate** for complementary images?
Are predictions accurate for complementary images?
Are models driven by priors?

Only consider those questions whose answers are not popular (given the question type) in training

- 1-Prior: Test answers are not the top-1 most common in training
- 2-Prior: Test answer are not the top-2 most common in training

Agrawal et al., CVPR 2018
Are models driven by priors?

5-6% drop
Are models driven by priors?

15-16% drop
Are models driven by priors?
Improvement from 2017 challenge

- 1-Prior: Best performance improved by 3.8%
- 2-Prior: Best performance improved by 3.3%
Are models compositional?

Only consider those questions which are compositionally novel:

- QA pair is not seen in training
- Constituting concepts seen in training

Agrawal et al., Arxiv 2018
Are models compositional?

**Training**

Q: What color is the plate?
A: Green

Q: What color are stop lights?
A: Red

**Testing**

Q: What color is the stop light?
A: Green

Q: What is the color of the plate?
A: Red
Are models compositional?

12-13% drop
Are models compositional?

2017 winner

56.5%
Are models compositional?
Average answer recall

- New accuracy metric proposed in Kafle and Kannan, ICCV 17
  - Also known as “Normalized accuracy”

- Method:
  - Computes accuracy for each unique answer
  - Take the mean over all unique answers

- Rewards models which perform well for rare answers
Average answer recall

- FAR-A
- HDU-USYD
- SNU-BI
- casia.CV Lab
- MIL-UT
- ut-swk
- graph-attention-msm
- DCD-ZJU
- Vqabyte
- fs
- UTS_YZD
- Adelaide-Tenerex
- VQA-ReasonTensor
- UPMC-LIP6
- wwerbale
- caption_vqa
- ovqa
- nagzero
- CFM-UESTC
- VQA_NTU
- yudf2010
- mmab612
- TsinghuaCVLab
- CIST-VQA
- VLC_Southampton
- HeiVQA
- University of Guelph MLRG
- NTU_ROSE_USTC
- tf-explain
- zhi-smile
- VQA-Machine+
- xle
- Vandaan
- HACKERS
- AE-VQA
- dandelion
- ghost
- VQA-Learning
- vqa-suchow
- HAIBIN
- windLBL
- VQA_San
- vqateam_mcb_benchmark
- akshay_isical
Progress in VQA

Accuracy on v2

ICCV 15

X-axis: Dates from 12/7/15 to 5/25/18
Y-axis: Accuracy on v2
Progress in VQA

Accuracy on v2

2016 Challenge winner

ICCV 15
Progress in VQA

Accuracy on v2

2016 Challenge winner

+7.0% absolute

ICCV 15

5/25/18

2/14/18

11/6/17

7/29/17

4/20/17

1/10/17

10/2/16

6/24/16

3/16/16

12/7/15

55

60

65

70

75

5.0

10.0

15.0

20.0

25.0

30.0

35.0

40.0

45.0

50.0

55.0

60.0

65.0

70.0

75.0
Progress in VQA

Accuracy on v2

ICCV 15

2016 Challenge winner

2017 Challenge winner

Challenge 2017 deadline
Progress in VQA

ICCV 15

2016 Challenge winner

2017 Challenge winner

Challenge 2017 deadline

+6.7% absolute

Accuracy on v2
Progress in VQA

Accuracy on v2

ICCV 15

2016 Challenge winner

2017 Challenge winner

2018 Challenge winner

Challenge 2018 deadline
Progress in VQA

+3.4% absolute

2016 Challenge winner

2017 Challenge winner

2018 Challenge winner

Accuracy on v2

ICCV 15

12/7/15  3/16/16  6/24/16  10/2/16  1/10/17  4/20/17  7/29/17  11/6/17  2/14/18  5/25/18
Visual Dialog Challenge 2018

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

[visualdialog.org/challenge/2018](visualdialog.org/challenge/2018)

**VisDial v1.0**

- ~130k images (COCO)
- 10-round dialog / image
- ~1.3 million QA pairs

**Evaluation**
- Automatic metrics
- Human annotations
Thanks!

Questions?