Large Scale Visual Recognition Challenge (ILSVRC)

http://image-net.org/challenges/LSVRC/
Backpack
Flute

Strawberry

Traffic light

Backpack

Matchstick

Bathing cap

Sea lion

Racket
Large-scale recognition
Large-scale recognition

Need benchmark datasets
PASCAL VOC 2005-2012

20 object classes

Classification: person, motorcycle

Detection

22,591 images

Segmentation

Action: riding bicycle

Everingham, Van Gool, Williams, Winn and Zisserman.
Large Scale Visual Recognition Challenge (ILSVRC) 2010-2014

- 20 object classes: 22,591 images
- 200 object classes: 517,840 images
- 1000 object classes: 1,431,167 images

http://image-net.org/challenges/LSVRC/
Variety of object classes in ILSVRC

Olga Russakovsky, Jia Deng, Zhiheng Huang, Alex Berg, Li Fei-Fei
Detecting avocados to zucchinis: what have we done, and where are we going? ICCV 2013
http://image-net.org/challenges/LSVRC/2012/analysis
Variety of object classes in ILSVRC

ILSVRC detection

- birds
  - bird

- bottles
  - bottle

- cars
  - car

ILSVRC classification and localization

- flamingo
- cock
- ruffed grouse
- quail
- partridge
- pill bottle
- beer bottle
- wine bottle
- water bottle
- pop bottle
- race car
- wagon
- minivan
- jeep
- cab
Challenge procedure every year

1. **Training** and **validation** data released: images and annotations

2. **Test** data released: images only (annotations hidden)

3. Participants train their models on **train** & **validation** data

4. Submit text file with predictions on **test** images

5. We evaluate and release results, and run a workshop

Participation in ILSVRC over the years

Year	Number of entries
3 years: 2010-2012
ILSVRC 2010	40
ILSVRC 2011	60
ILSVRC 2012	81
Last year: 2013
ILSVRC 2013: 81 entries
Participation in ILSVRC over the years

- ILSVRC 2010: 81 entries
- ILSVRC 2011: 81 entries (sum of ILSVRC 2010 and 2011)
- ILSVRC 2013: 81 entries
- ILSVRC 2014: 123 entries

Summary:
- 3 years: 2010-2012
  - Number of entries: 60
- Last year: 2013
  - Number of entries: 81
- This year: 2014
  - Number of entries: 123
Experiment this year: open vs closed submissions

• Offered all teams an option:
  – Open = promise to reveal their method
  – Closed = participate without revealing the method

• Almost all teams chose to be “open” (31/36)
  – And 2 of the “closed” teams still presented spotlights and posters at the workshop
ILSVRC in detail: history and current state-of-the-art

ImageNet Large Scale Visual Recognition Challenge
Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander Berg, Li Fei-Fei

http://arxiv.org/abs/1409.0575

- Describes the construction of the ILSVRC datasets
- Highlights the most successful algorithms
- Provides statistical analysis of the results through ILSVRC2014
- Compares computer vision accuracy with human-level accuracy
Some questions for today

1. Are computers good at large-scale recognition?

2. Are all objects equally easy for computers?

3. Are we better than computers at recognition?
Some questions for today

1. Are computers good at large-scale recognition?

2. Are all objects equally easy for computers?

3. Are we better than computers at recognition?
ILSVRC image classification task

Steel drum
ILSVRC image classification task

Steel drum

Output:
- Scale
- T-shirt
- Steel drum
- Drumstick
- Mud turtle

Output:
- Scale
- T-shirt
- Giant panda
- Drumstick
- Mud turtle
ILSVRC image classification task

Steel drum

Output:  
- Scale  
- T-shirt  
- Steel drum  
- Drumstick  
- Mud turtle

Output:  
- Scale  
- T-shirt  
- Giant panda  
- Drumstick  
- Mud turtle

Error = \frac{1}{100,000} \sum_{\text{100,000 images}} 1[\text{incorrect on image } i]
### ILSVRC2014 classification results

<table>
<thead>
<tr>
<th>Team Name</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogLeNet</td>
<td>6.7</td>
</tr>
<tr>
<td>VGG</td>
<td>7.3</td>
</tr>
<tr>
<td>MSRA Visual computing</td>
<td>8.1</td>
</tr>
<tr>
<td>Andrew Howard</td>
<td>8.1</td>
</tr>
<tr>
<td>DeeperVision</td>
<td>9.5</td>
</tr>
<tr>
<td>NUS-BST</td>
<td>9.8</td>
</tr>
<tr>
<td>TTIC_ECP – Epitomic Vision</td>
<td>10.2</td>
</tr>
<tr>
<td>XYZ</td>
<td>11.2</td>
</tr>
</tbody>
</table>

**GoogLeNet:**
Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Drago Anguelov, Dumitru Erhan, Andrew Rabinovich
Google

**VGG:**
Karen Simonyan, Andrew Zisserman
University of Oxford

---

ILSVRC over the years

1.7x reduction in classification error since last year

4.2x reduction in classification error since 2010

http://arxiv.org/abs/1409.0575
What changed in ILSVRC classification?

Year 2010

NEC-UIUC

Dense grid descriptor: HOG, LBP

Coding: local coordinate, super-vector

Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

SuperVision

Year 2014

GoogLeNet

VGG

MSRA

[Simonyan arxiv 2014]  [He arxiv 2014]  [Szegedy arxiv 2014]

[Dense grid descriptor: HOG, LBP

Coding: local coordinate, super-vector

Pooling, SPM

Linear SVM

[Krizhevsky NIPS 2012]  [Lin CVPR 2011]  [Krizhevsky NIPS 2012]  [Szegedy arxiv 2014]  [Simonyan arxiv 2014]  [He arxiv 2014]
What changed in ILSVRC2014 classification?

1. Networks became deeper

Inception module  [Going deeper with convolutions. Szegedy et al. 2014]

More layers but with smaller kernels (3x3 convolution, 2x2 pooling)

What changed in ILSVRC2014 classification?

2. Fully connected layers were (sometimes) removed

Network in Network. Min Lin, Qiang Chen and Shicheng Yan. ICLR 2014
Also used in GoogLeNet and others

What changed in ILSVRC2014 classification?

3. Almost all successful systems used

- Extensive data augmentation
- Multiscale training across more scales
- Network fusion

And, most importantly, ...
What changed in ILSVRC2014 classification?

3. Almost all successful systems used

- Extensive data augmentation
- Multiscale training across more scales
- Network fusion

And, most importantly, ...

Caffe!

Are the winning classification systems really significantly more accurate?

<table>
<thead>
<tr>
<th>Year</th>
<th>Team name</th>
<th>Error (percent)</th>
<th>99.9% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>VGG</td>
<td>7.32</td>
<td>7.05 - 7.60</td>
</tr>
<tr>
<td>2014</td>
<td>MSRA</td>
<td>8.06</td>
<td>7.78 - 8.34</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Clarifai</td>
<td>11.20</td>
<td>10.87 - 11.53</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>SuperVision</td>
<td>15.32</td>
<td>14.94 - 15.69</td>
</tr>
</tbody>
</table>

http://arxiv.org/abs/1409.0575
ILSVRC classification + localization task

Steel drum
ILSVRC classification + localization task

Steel drum

Output

Persian cat
Picket fence
Steeldrum
Folding chair
Loud speaker
ILSVRC classification + localization task

Steel drum

Output

Output (bad localization)

Output (bad classification)
ILSVRC classification + localization task

\[
\text{Error} = \frac{1}{100,000} \sum_{\text{100,000 images}} 1[\text{incorrect on image } i]
\]
# ILSVRC2014 localization results

<table>
<thead>
<tr>
<th>Team Name</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG</td>
<td>25.3</td>
</tr>
<tr>
<td>GoogLeNet</td>
<td>26.4</td>
</tr>
<tr>
<td>Adobe-UIUC</td>
<td>30.1</td>
</tr>
<tr>
<td>SYSU_Vision</td>
<td>31.9</td>
</tr>
<tr>
<td>MIL</td>
<td>33.7</td>
</tr>
<tr>
<td>MSRA Visual computing</td>
<td>35.5</td>
</tr>
<tr>
<td>Trimps-Soushen</td>
<td>42.2</td>
</tr>
<tr>
<td>ORANGE-BUPT</td>
<td>42.7</td>
</tr>
</tbody>
</table>

**VGG:**
Karen Simonyan, Andrew Zisserman
University of Oxford

**GoogLeNet:**
Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Drago Anguelov, Dumitru Erhan, Andrew Rabinovich
Google

ILSVRC object detection task

Fully annotated 200 object classes across 120,000 images

 Allows evaluation of generic object detection in cluttered scenes at scale

Modeled after PASCAL VOC
ILSVRC object detection data
ILSVRC object detection task

All instances of all target object classes expected to be localized on all test images

Evaluation modeled after PASCAL VOC:

- Algorithm outputs a list of bounding box detections with confidences
- A detection is considered correct if IOU with ground truth > threshold
- Evaluated by average precision per object class
- Winners of challenge is the team that wins the most object categories

ILSVRC2014 object detection approach #1

R-CNN: Regions with CNN features

1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

Rich feature hierarchies for accurate object detection
Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik
http://arxiv.org/abs/1311.2524

ILSVRC2014 object detection approach #2

SPP-net: Spatial Pyramid Pooling

Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition
Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

ILSVRC detection since 2013

1.9x increase in object detection average precision in one year

http://arxiv.org/abs/1409.0575
ILSVRC detection since 2013

1.9x increase in object detection average precision in one year

~3% due to more data
~18% due to better methods

http://arxiv.org/abs/1409.0575
Some questions for today

1. Are computers good at large-scale recognition?

2. Are all objects equally easy for computers?

3. Are we better than computers at recognition?
Some questions for today

1. Are computers good at large-scale recognition?  
   Yes!  CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?

3. Are we better than computers at recognition?
Some questions for today

1. Are computers good at large-scale recognition?
   Yes! CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?

3. Are we better than computers at recognition?
Easiest classification categories

- red fox (100)
- hen-of-the-woods (100)
- ibex (100)
- goldfinch (100)
- flat-coated retriever (100)
- tiger (100)
- hamster (100)
- porcupine (100)
- stingray (100)
- Blenheim spaniel (100)

... and 111 more categories with 100% accuracy!

(Highest accuracy in percent achieved by any method in ILSVRC2012-ILSVRC2014)

http://arxiv.org/abs/1409.0575
Hardest classification categories

muzzle (71)  hatchet (68)  water bottle (68)  velvet (68)  loupe (66)
hook (66)  spotlight (66)  ladle (65)  restaurant (64)  letter opener (59)

(Highest accuracy in percent achieved by any method in ILSVRC2012-ILSVRC2014)

http://arxiv.org/abs/1409.0575
Easiest localization categories

Leonberg (100)  ruffe grouse (100)  ruddy turnstone (100)  giant schnauzer (99)  
Maltese dog (99)  Japanese spaniel (99)  Tibetan mastiff (99)  hare (99)  African hunting dog (99)  

(Highest accuracy in percent achieved by any method in ILSVRC2012-ILSVRC2014) 

http://arxiv.org/abs/1409.0575
Hardest localization categories

horizontal bar (41) flagpole (38) hook (37) lakeside (36) letter opener (36)
spotlight (35) wing (35) ladle (28) pole (27) space bar (23)

(Highest accuracy in percent achieved by any method in ILSVRC2012-ILSVRC2014)

http://arxiv.org/abs/1409.0575
Easiest object detection categories

1. butterfly (93)
2. dog (84)
3. volleyball (83)
4. rabbit (83)
5. frog (82)
6. basketball (80)
7. snowplow (80)
8. bird (78)
9. tiger (77)
10. zebra (77)

(Highest average precision in percent achieved by any method in ILSVRC2013 and ILSVRC2014)

http://arxiv.org/abs/1409.0575
Hardest object detection categories

- lamp (15)
- flute (15)
- horizontal bar (14)
- spatula (13)
- nail (13)
- ski (12)
- microphone (11)
- rubber eraser (10)
- ladle (9)
- backpack (8)

(Highest average precision in percent achieved by any method in ILSVRC2013 and ILSVRC2014)

http://arxiv.org/abs/1409.0575
Smaller objects not necessarily harder

- Each dot is an object class
- X-axis: average fraction of image area occupied by an instance of that class on the validation set
- Y-axis: highest accuracy achieved by any method in ILSVRC2012-ILSVRC2014

http://arxiv.org/abs/1409.0575
Smaller objects not necessarily harder

- Each dot is an object class
- X-axis: average fraction of image area occupied by an instance of that class on the validation set
- Y-axis: highest accuracy achieved by any method in ILSVRC2012-ILSVRC2014

http://arxiv.org/abs/1409.0575
Manually annotated object class properties

Amount of Texture
- Screwdriver
- Hatchet
- Ladybug
- Honeycomb

Color Distinctiveness
- Coffee mug
- Cleaver
- Bagel
- Red Wine

Shape Distinctiveness
- Jigsaw Puzzle
- Foreland
- Lipstick
- Bell

Real-world Size
- Orange
- Mask
- Parachute
- Airliner

Olga Russakovsky, Jia Deng, Zhiheng Huang, Alex Berg, Li Fei-Fei
Detecting avocados to zucchinis: what have we done, and where are we going? ICCV 2013
Textured objects are easier

Image classification

Classification accuracy

None Low Medium High

Amount of Texture

http://arxiv.org/abs/1409.0575
Textured objects are easier

http://arxiv.org/abs/1409.0575
Some questions for today

1. Are computers good at large-scale recognition?  
   Yes!  
   CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?

3. Are we better than computers at recognition?
Some questions for today

1. Are computers good at large-scale recognition?
   
   **Yes!** CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?
   
   **No.** Thin and untextured objects are still hard for computers.

3. Are we better than computers at recognition?
Some questions for today

1. Are computers good at large-scale recognition?
   Yes! CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?
   No. Thin and untextured objects are still hard for computers.

3. Are we better than computers at recognition?
What is human accuracy on ILSVRC2014 classification?

But data is manually annotated, isn’t human accuracy 100%?

Current crowdsourcing annotation interface

Is this a badger? Yes or No

Very different from

Which one of 1000 classes is this?

http://arxiv.org/abs/1409.0575
New web-based annotation interface with 1000 object classes


http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

- Compared expert human annotators with winning GoogLeNet entry

<table>
<thead>
<tr>
<th></th>
<th>Annotator 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of images</td>
<td>1500</td>
</tr>
<tr>
<td>GoogLeNet classification error</td>
<td>6.8%</td>
</tr>
<tr>
<td>Human classification error</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

- Annotator 1 achieved better accuracy than GoogLeNet by 1.7% ($p = 0.022$)
- Task required *significant* amount of training for humans

http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

- Compared expert human annotators with winning GoogLeNet entry

<table>
<thead>
<tr>
<th></th>
<th>Annotator 1</th>
<th>Annotator 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of images</td>
<td>1500</td>
<td>258</td>
</tr>
<tr>
<td>GoogLeNet classification error</td>
<td>6.8%</td>
<td>5.8%</td>
</tr>
<tr>
<td>Human classification error</td>
<td>5.1%</td>
<td>12.0%</td>
</tr>
</tbody>
</table>

- Annotator 1 achieved better accuracy than GoogLeNet by 1.7% ($p = 0.022$)
- Task required *significant* amount of training for humans

http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th></th>
<th>GoogLeNet correct</th>
<th>GoogLeNet wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human correct</strong></td>
<td>1352/1500</td>
<td>72/1500</td>
</tr>
<tr>
<td><strong>Human wrong</strong></td>
<td>46/1500</td>
<td>30/1500</td>
</tr>
</tbody>
</table>

http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th></th>
<th>GoogLeNet correct</th>
<th>GoogLeNet wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human correct</td>
<td>1352/1500</td>
<td>72/1500</td>
</tr>
<tr>
<td></td>
<td><img src="Image" alt="Smiley Face" /></td>
<td>- Objects very small or thin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Abstract representations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Image filters</td>
</tr>
<tr>
<td>Human wrong</td>
<td>46/1500</td>
<td>30/1500</td>
</tr>
</tbody>
</table>

http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th>Human correct</th>
<th>GoogLeNet correct</th>
<th>GoogLeNet wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>1352/1500</td>
<td></td>
<td>72/1500</td>
</tr>
<tr>
<td>46/1500</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

- Objects very small or thin
- Abstract representations
- Image filters

<table>
<thead>
<tr>
<th>reel</th>
<th>hatchet</th>
<th>sidewinder</th>
</tr>
</thead>
<tbody>
<tr>
<td>stethoscope</td>
<td>vase</td>
<td>maze, labyrinth</td>
</tr>
<tr>
<td>whistle</td>
<td>pitcher, ewer</td>
<td>gar, garfish</td>
</tr>
<tr>
<td>ice lolly, lolly</td>
<td>coffeepot</td>
<td>valley, vale</td>
</tr>
<tr>
<td>hair spray</td>
<td>mask</td>
<td>hammerhead</td>
</tr>
<tr>
<td>maypole</td>
<td>cup</td>
<td>sea snake</td>
</tr>
</tbody>
</table>

http://arxiv.org/abs/1409.0575
# Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th>Human correct</th>
<th>GoogLeNet correct</th>
<th>GoogLeNet wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>1352/1500</td>
<td></td>
<td>72/1500</td>
</tr>
<tr>
<td></td>
<td>• Objects very small or thin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Abstract representations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Image filters</td>
<td></td>
</tr>
<tr>
<td>Human wrong</td>
<td>46/1500</td>
<td>30/1500</td>
</tr>
<tr>
<td></td>
<td>• Fine-grained recognition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Class unawareness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Insufficient training data</td>
<td></td>
</tr>
</tbody>
</table>

### Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th>GoogLeNet <strong>correct</strong></th>
<th>GoogLeNet <strong>wrong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1352/1500</td>
<td>72/1500</td>
</tr>
<tr>
<td>Human <strong>correct</strong></td>
<td></td>
</tr>
<tr>
<td>1352/1500</td>
<td>Objects very small or thin</td>
</tr>
<tr>
<td></td>
<td>Abstract representations</td>
</tr>
<tr>
<td></td>
<td>Image filters</td>
</tr>
<tr>
<td>Human <strong>wrong</strong></td>
<td></td>
</tr>
<tr>
<td>46/1500</td>
<td>Fine-grained recognition</td>
</tr>
<tr>
<td></td>
<td>Class unawareness</td>
</tr>
<tr>
<td></td>
<td>Insufficient training data</td>
</tr>
</tbody>
</table>

- GoogLeNet correct: 1352/1500
- GoogLeNet wrong: 72/1500
- Human correct: 1352/1500
- Human wrong: 46/1500

http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th>Human correct</th>
<th>GoogleLeNet correct</th>
<th>1352/1500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human wrong</td>
<td>GoogleLeNet wrong</td>
<td>72/1500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Objects very small or thin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abstract representations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Image filters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fine-grained recognition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class unawareness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Insufficient training data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>46/1500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30/1500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple objects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incorrect annotations</td>
<td></td>
</tr>
</tbody>
</table>

http://arxiv.org/abs/1409.0575
Human vs computer accuracy on ILSVRC2014 classification

<table>
<thead>
<tr>
<th>Human correct</th>
<th>GoogLeNet correct</th>
<th>GoogLeNet wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>1352/1500</td>
<td>72/1500</td>
<td>Objects very small or thin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract representations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Image filters</td>
</tr>
<tr>
<td>46/1500</td>
<td>30/1500</td>
<td>Multiple objects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incorrect annotations</td>
</tr>
</tbody>
</table>

- Fine-grained recognition
- Class unawareness
- Insufficient training data

Objects very small or thin
Abstract representations
Image filters
Multiple objects
Incorrect annotations

http://arxiv.org/abs/1409.0575
Some questions for today

1. Are computers good at large-scale recognition?
   
   Yes!  
   
   CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?
   
   No.  
   
   Thin and untextured objects are still hard for computers.

3. Are we better than computers at recognition?
Some questions for today

1. Are computers good at large-scale recognition?
   
   **Yes!** CNNs are getting deeper, accuracy is getting better.

2. Are all objects equally easy for computers?
   
   **No.** Thin and untextured objects are still hard for computers.

3. Are we better than computers at recognition?
   
   **Not always...** We are worse than computers at large-scale fine-grained classification.
So have we solved computer vision?
So have we solved computer vision?

*RCNN output:*
So have we solved computer vision?

- Male, brown hair
- TV facing away from camera
- Wooden table
- Tall male, wearing pants
- Black backpack
So have we solved computer vision?
So have we solved computer vision?

Not quite yet
So have we solved computer vision?

Not quite yet

But you should still read our paper:

ImageNet Large Scale Visual Recognition Challenge

http://arxiv.org/abs/1409.0575