embedded VISION SUMMIT 2018

Bad Data Bad Network

Mike Schmit

May 22, 2018

Agenda

- Introduction to image classification
- Example training problems
  - Overfitting and underfitting
- Statistical results on the Test dataset
- Guidelines for Dataset construction
- Gaining intuition on how computer vision works
  - Sample Images, issues and corrective actions
- Conclusions and further work
- Resources

AMD

Image classification

- Image classification

(this talk is about classification)

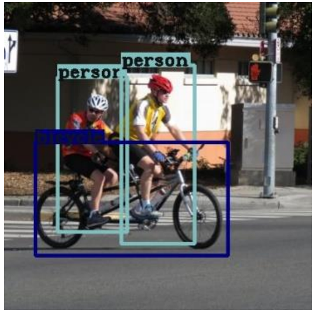
corn	
candy store	97.3%
pill bottle	0.4%
packet	0.3%
corn	0.2%
grocery store	0.1%

Manually labelled ground truth

Network generated confidence

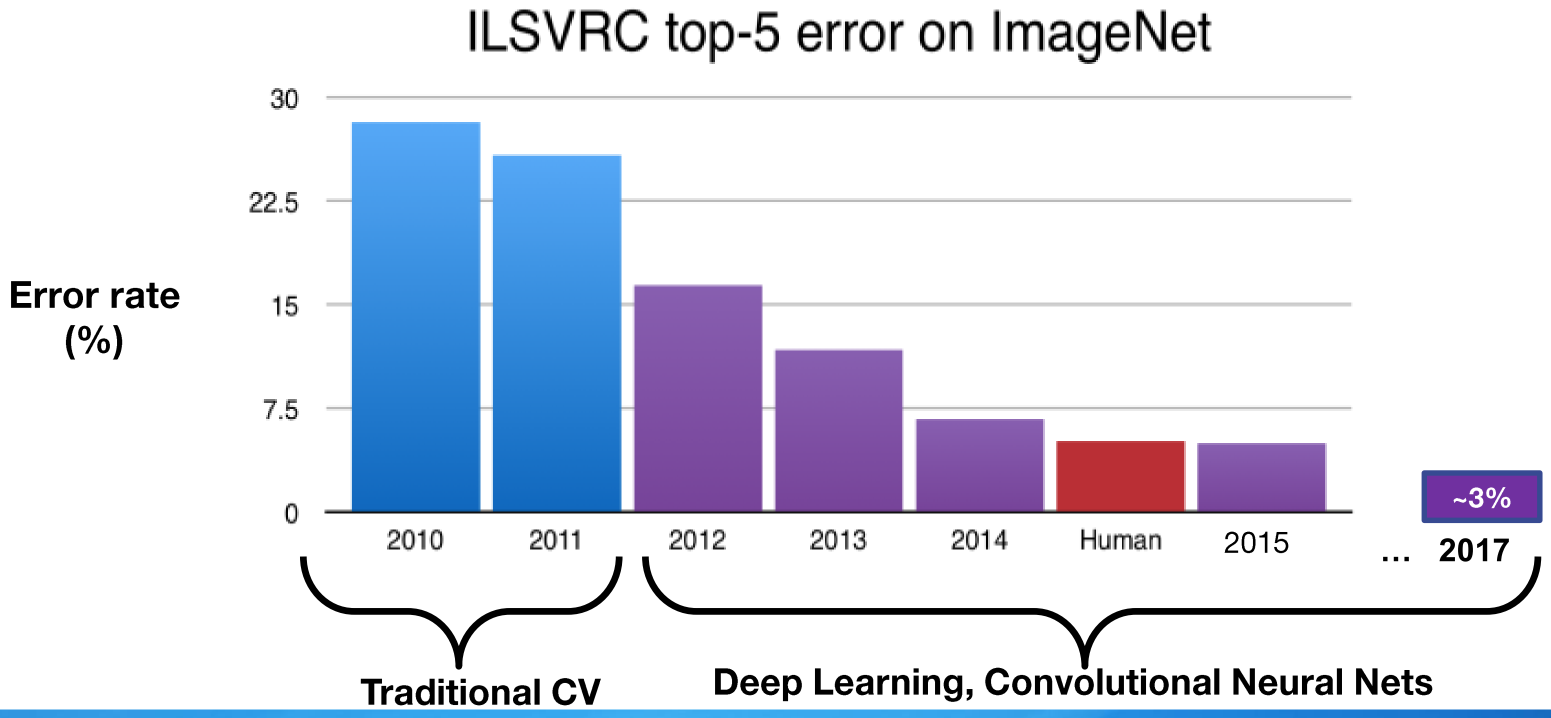
Top 5 of 1000 object classes

- Object detection is different

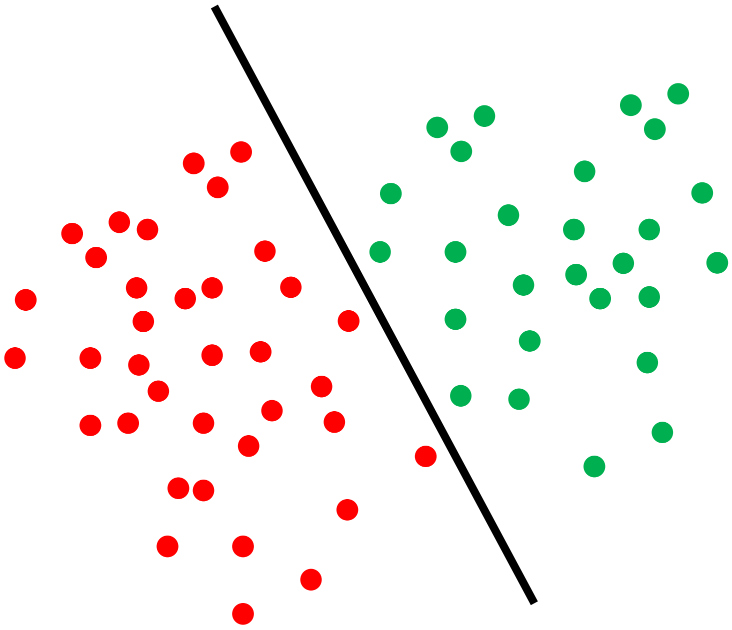


ImageNet Large Scale Visual recognition Challenge

IMAGENET 15M images, 1000 categories

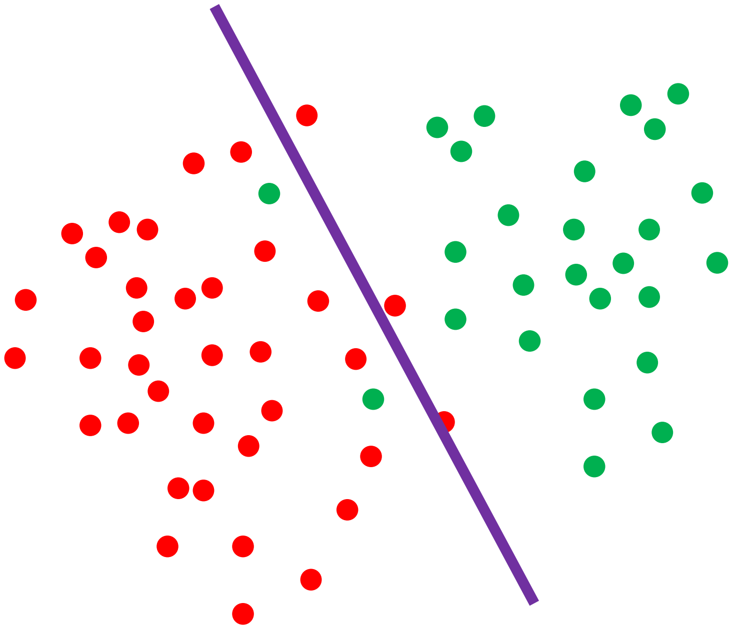


A Simple Example: Overfitting and Underfitting



Example of a set of data points where a simple line partitions the data

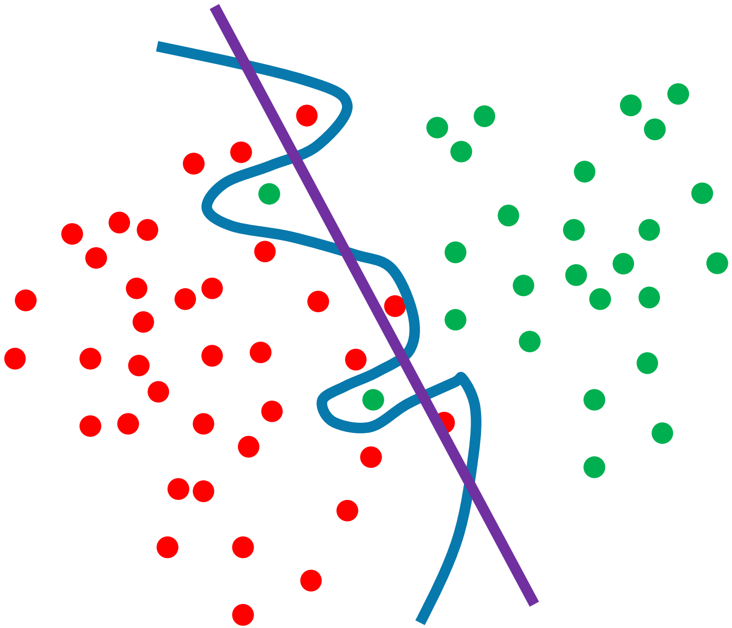
Example: Overfitting and Underfitting



Underfitting

The same line misses a couple of points on each side

Example: Overfitting and Underfitting

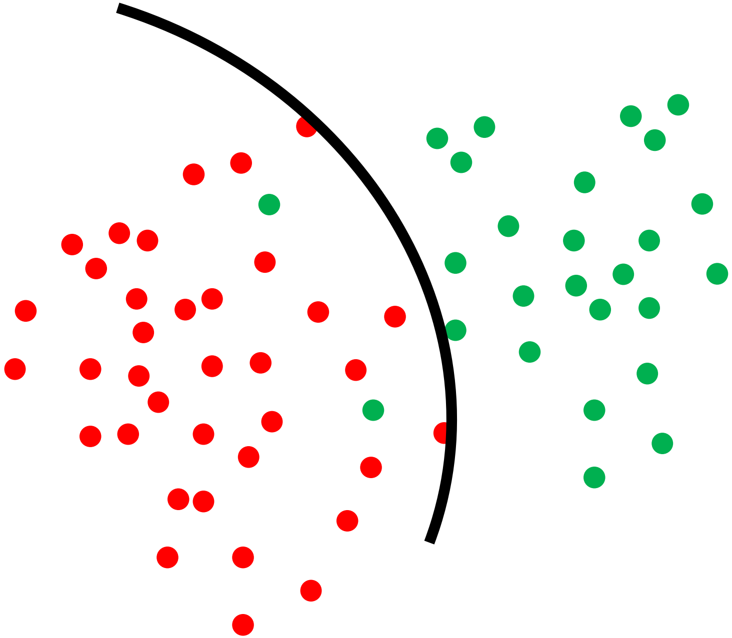


Underfitting

Overfitting

This is like memorizing images rather than finding features

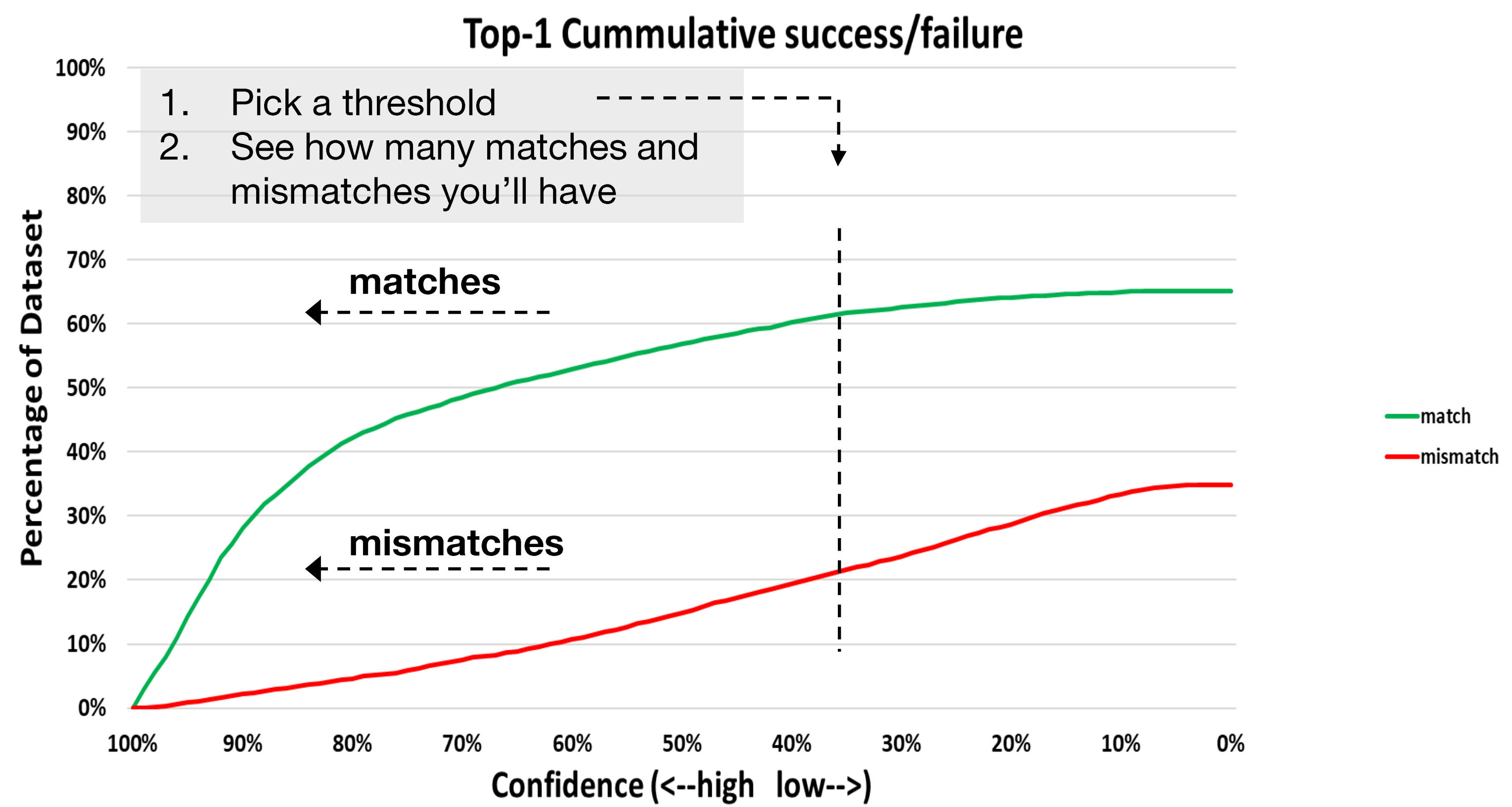
Example: Overfitting and Underfitting



This misses a couple of outliers
but is a good general fit

Proper fitting

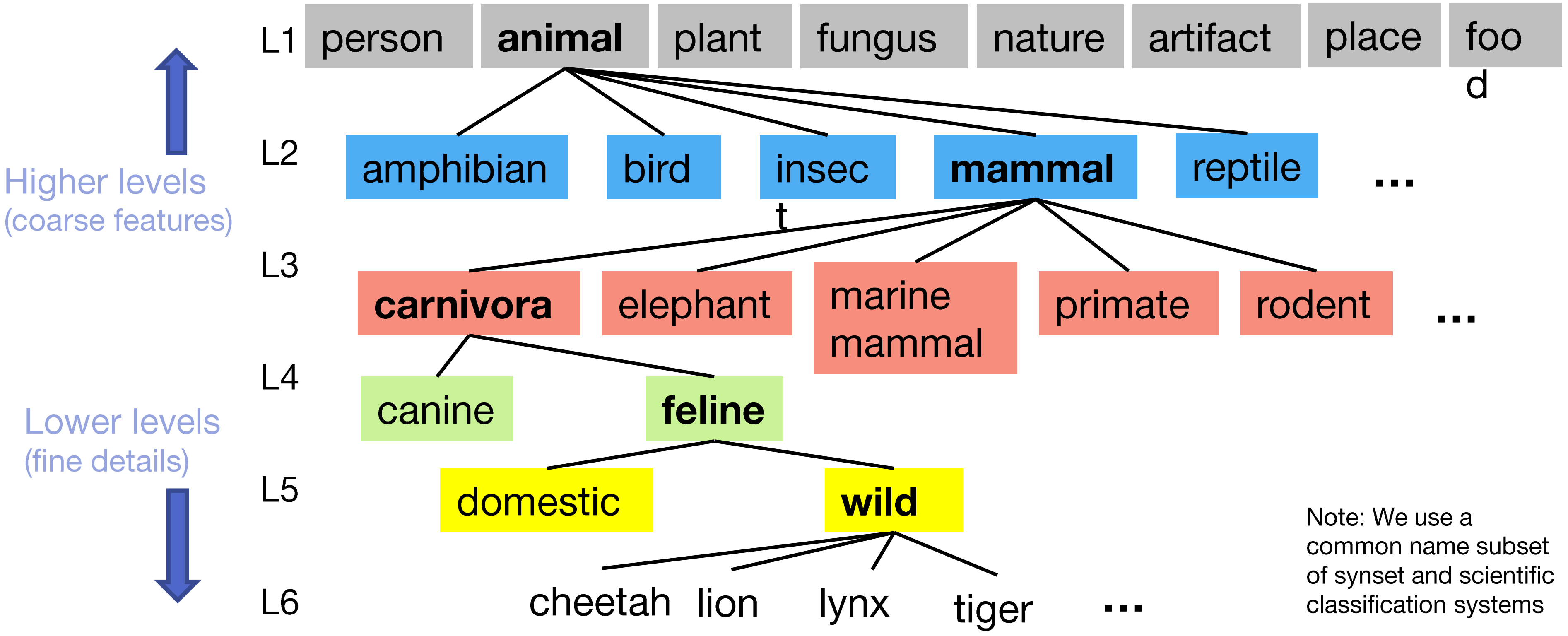
Match / Mismatch vs reported confidence



Setting your threshold for what the neural net confidence should be accepted as a match

Test dataset Inception V4

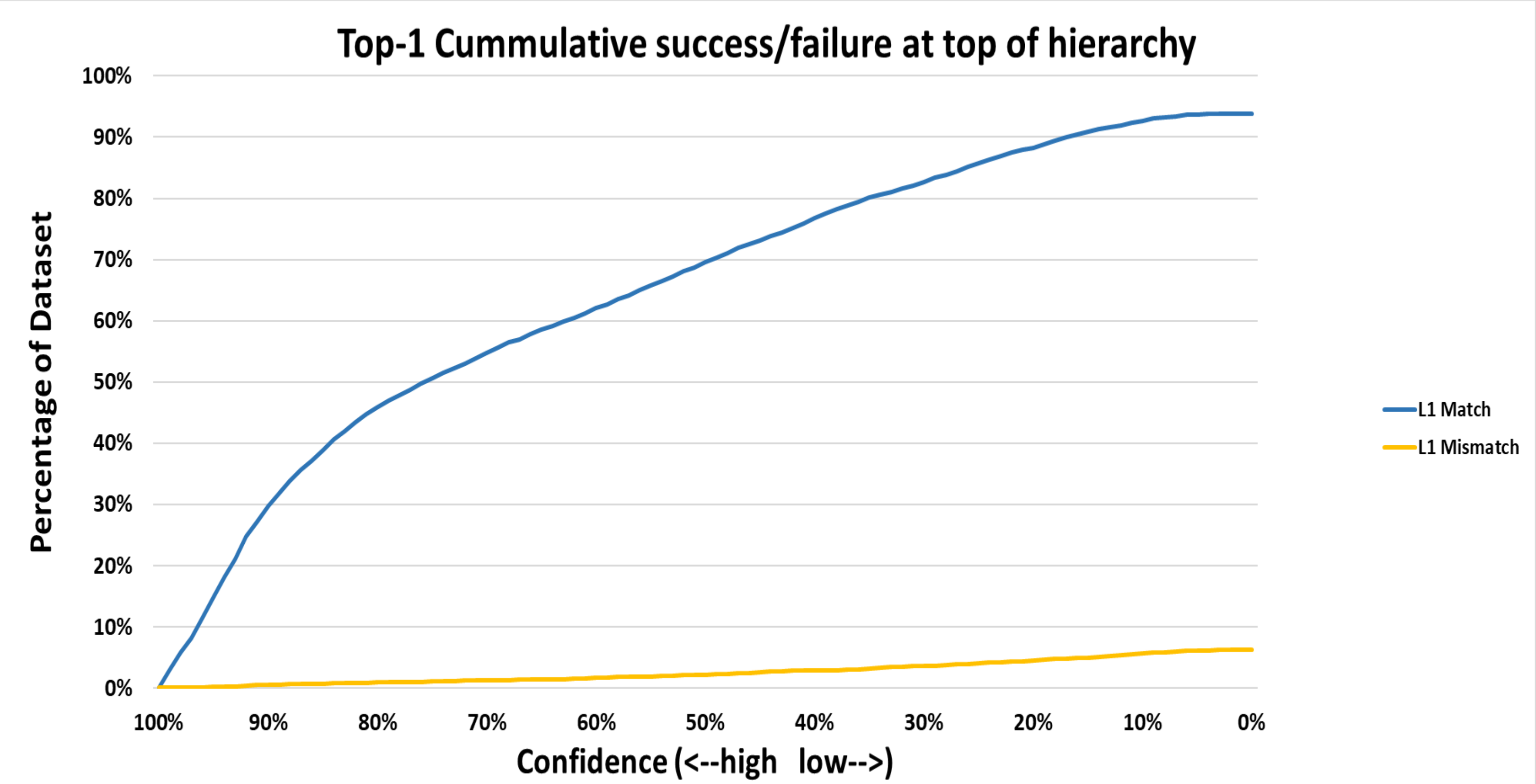
Introducing matching within a hierarchy



Note: We use a common name subset of synset and scientific classification systems

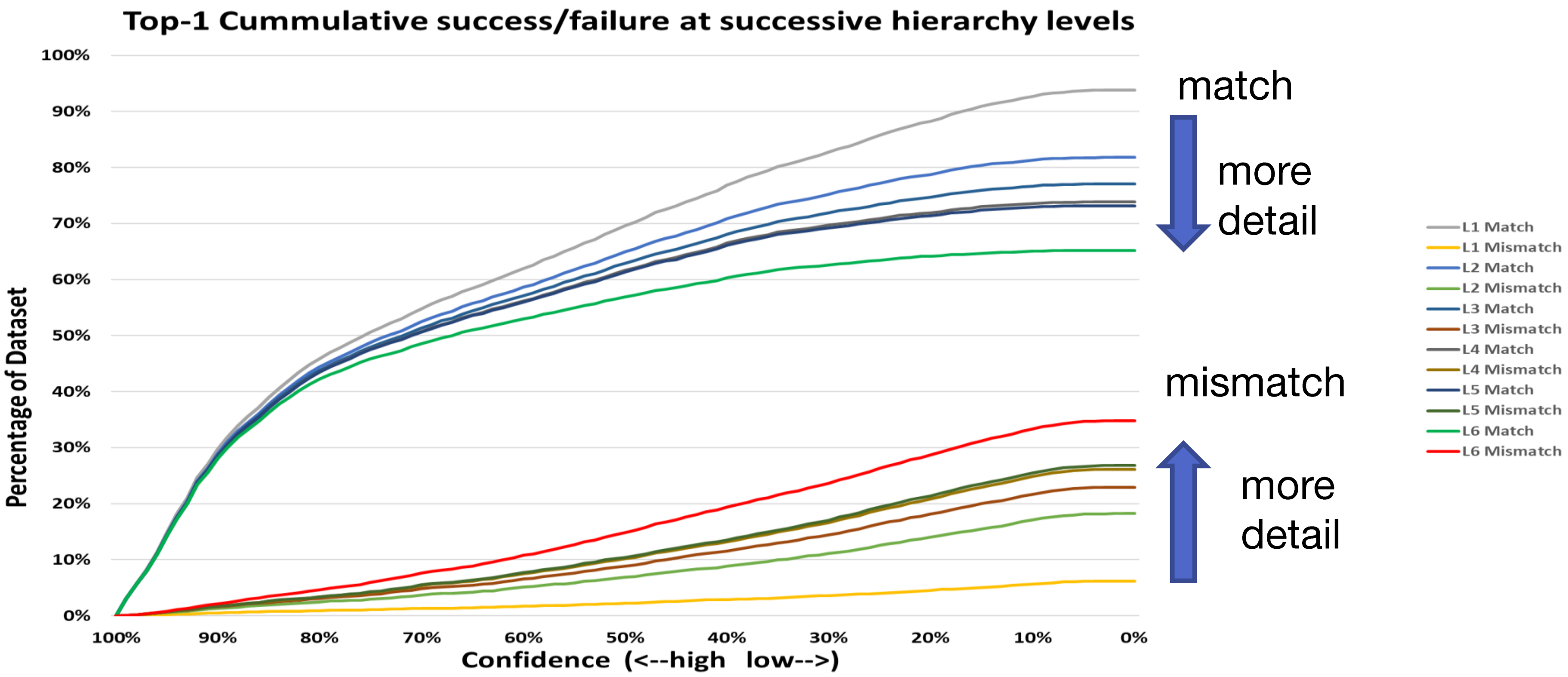
AMD

Match / Mismatch at different hierarchical levels



L1 = highest hierarchy level; such as person, animal, plant…

Example


Top-1 Cummulative success/failure at successive hierarchy levels
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percentage of Dataset
100% 90% 80% 70% 60% 50% 40% 30% 20% 10% 0%
Confidence (<--high low-->)
match
more detail
mismatch
more detail
L1 Match
L1 Mismatch
L2 Match
L2 Mismatch
L3 Match
L3 Mismatch
L4 Match
L4 Mismatch
L5 Match
L5 Mismatch
L6 Match
L6 Mismatch

The Test dataset

Classification results*

Model	top 1	2nd	3rd	4th	5th	top 5
Inception V4	70.1%	11.2%	4.8%	2.5%	1.7%	90.3%
Resnet-50	57.2%	12.2%	6.3%	3.7%	2.3%	81.6%
Resnet-101	53.5%	12.5%	7.2%	4.0%	2.8%	79.9%
Resnet-152	62.2%	12.7%	5.4%	3.0%	2.0%	85.1%
VGG-19	50.8%	12.7%	6.2%	4.2%	3.0%	76.8%
VGG-16	51.7%	12.3%	6.7%	4.0%	2.4%	77.1%

Labeled Images	7741
Negative tests (not labeled)	1871
Total Images	9612
Total classes	757 (of 1000)

* Results not statistically comparable to ILSVRC due to a different mix of images and some intentionally difficult images

Choosing or Constructing your Dataset(s)

- Match your dataset to your actual use cases
- Know your dataset (beware of label synonyms)
  - ImageNet has several errors like this
- Have plenty of training examples of isolated, iconic images of each object class – **but not exclusively**
- Clearly show any defining object characteristics (such as a rattle on a rattlesnake) – **but not exclusively**
  - If you don’t have these examples the neural net can’t learn them

AMD

Choosing or Constructing your Dataset(s)

- After the traditional training and validation
  - Include realistic positive test cases as well as **negative** test cases
    - You want to know how your software handles unknown objects – will it confidently think it found a similar object or will it confidently give a strange result?
  - For many uses, not every image your software “sees” will have a corresponding trained-for image class. Training with a few classes of “blank” background cases can prevent some of the wildly wrong results…because it knows the backgrounds classes

Choosing or Constructing your Dataset(s)

- Manually inspect images that mismatched the top-1 AND have high confidence for clues to what is going wrong (could be a poor dataset, could be bad labeling, etc.)
- Some common errors
  - Too many training images, in a class, have the same background
    - Grass alone gets detected as soccer balls, dogs, golf balls
  - Objects commonly found together get confused for each other
    - Ping-pong paddles (not trained for and tested for) get detected as “ping-pong balls”
    - Arrows get detected as “bows”

CD players and Records

LP record (not in list)	
CD player	95.6%
tape player	0.7%
stove	0.5%
potter's wheel	0.4%
wok	0.2%

Objects commonly seen together get mistaken for each other

Need to have training examples of objects alone as well

Lions

Lion	
warthog	81.4%
wombat	1.5%
hog	1.5%
hippo	1.4%
water buffalo	0.6%

35% of lions detected in top 1
54% in top 5
(many in odd poses and/or occluded)

Probably no training in odd poses

Lawn mowers

lawn mower	
go-kart	41.5%
lawn mower	16.5%
Amphibious vehicle	2.3%
harvester	1.5%
thresher	0.01%

sweatshirt	
lawn mower	55.3%
croquet ball	4.7%
chainsaw	4.3%
wheelbarrow	2.8%
shovel	2.3%

Go-kart and lawn mower are similar

Overfitting to the person mowing

Stage actors, dancers

Stage	
balance beam	77.7 %
horizontal bar	5.9%
pole	1.2%
parallel bars	1.2%
unicycle	0.5%

Numerous examples of people in poses commonly associated with particular objects mistaken for the object itself, especially gymnastics

Need to have training examples of objects (balance beam) alone as well

Mittens

mitten	
sleeping bag	88.8%
bulletproof vest	0.8%
ski mask	0.4%
mitten	0.3%
meerkat	0.3%

mitten	
sleeping bag	86.5%
diaper	1.9%
Band Aid	0.7%
sweatshirt	0.6%
ski mask	0.6%

Probably due to texture matching to sleeping bags

AMD

Tigers

tiger	
lynx	55.6%
tiger cat	6.9%
wallaby	4.4%
tabby cat	3.5%
Egyptian cat	2.4%

tiger (cubs in nursery)	
tray	22.8%
scabbard	3.4%
photocopier	3.0%
printer	3.0%
rifle	2.1%

A mystery (21 of 22 other tigers found)

Tigers under glass not found

Socks

Sock	
diaper	51.3%
brassiere	7.4%
bib	6.8%
nipple	6.5%
handkerchief	3.8%

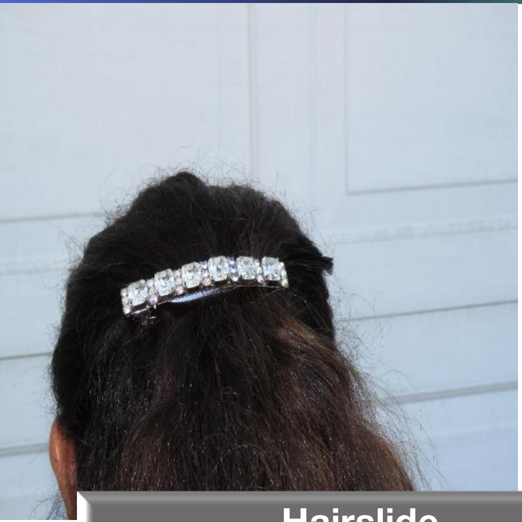
Sock	
mink	14.9%
Mexican hairless (dog)	5.8%
weasel	5.5%
paper towel	2.4%
Great Dane	2.4%

Probably not enough training for socks not on feet

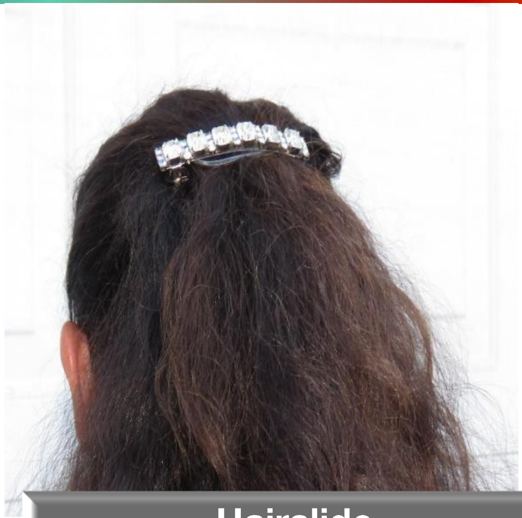
Socks on feet always found

AMD

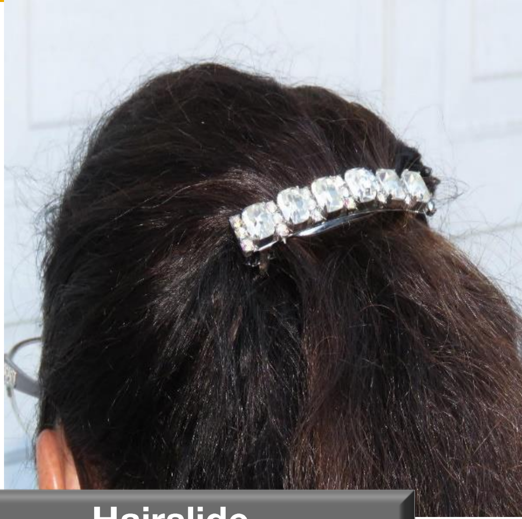
Minor changes in camera angle & zoom



Hairslide	
bearskin (busby)	99.3%
wig	0.04%
Sloth bear	0.04%
chimpanzee	0.03%
hairslide	0.03%



Hairslide	
bearskin (busby)	57.3%
wig	13.1%
Affenpinscher	5.1%
Tibetan terroir	xx%
Sloth bear	1.74%



Hairslide	
hairslide	80.0%
bearskin (busby)	2.8%
wig	1.7%
Kimono	0.5%
hairspray	0.3%

Does the network learn the major features?

Diamondback Rattlesnake	
rock python	45.6%
night snake	12.3%
Indian cobra	6.2%
sidewinder	4.3%
sand viper	3.6%

Diamondback Rattlesnake	
sidewinder	39.6%
Rattlesnake	14.5%
rock python	13.6%
sand viper	8.0%
night snake	7.6%

More clear training data needed?

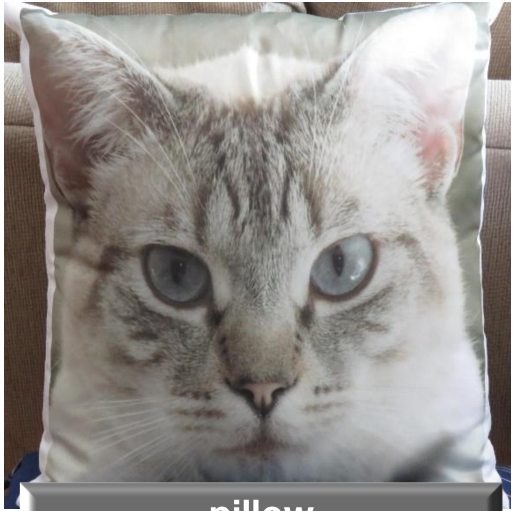
Pillow

pillow	
pillow	69.2%
sleeping bag	1.1%
quilt	1.0%
studio couch	0.6%
packet	0.4%

Cats or pillows?

pillow	
Siamese cat	87.4%
Egyptian cat	0.8%
hamper	0.1%
Tabby cat	0.1%
Plastic bag	0.1%

pillow	
Egyptian cat	62.1%
tabby cat	15.1%
lynx	6.0%
Tiger cat	2.8%
Siamese cat	0.7%



pillow	
Egyptian cat	48.5%
tabby cat	39.0%
tiger cat	[illegible]
lynx	[illegible]
Siamese cat	0.1%

Real cats also fooled by these

Ships

Navy warship (destroyer)	
aircraft carrier	90.6%
drilling platform	0.6%
submarine	0.3%
container ship	0.1%
fireboat	0.1%

Destroyers not in the training dataset… the primary differentiating feature of an aircraft carrier is missed – flat top

Oranges

orange	
abacus	87.7%
croquet ball	0.7%
apple	0.5%
solar collector	0.2%
ping-pong ball	0.2%

Not enough training for oranges at a grocery store

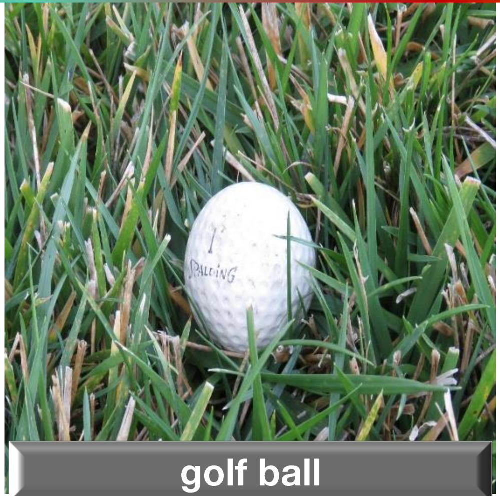
Strange mailbox

mailbox	
street sign	84.4%
mailbox	3.5%
pole	2.7%
seashore	0.7%
lakeside	0.4%

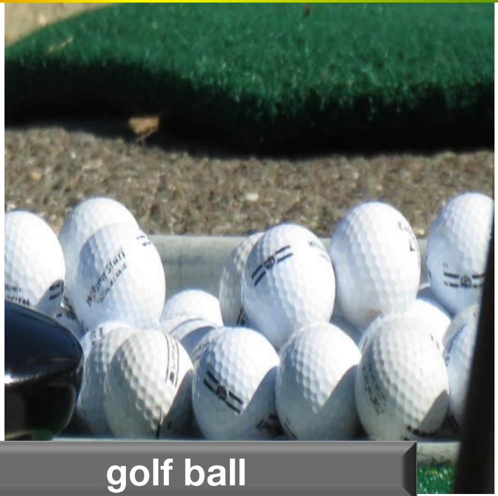
AMD

Golf balls

golf ball	
golf ball	96.9%
bassoon	0.1%
tarantula	0.0%
warthog	0.0%
tennis ball	0.0%



golf ball	
golf ball	96.1%
ping-pong ball	0.7%
baseball	0.2%
tennis ball	0.1%
croquet ball	0.1%



golf ball	
golf ball	97.7%
baseball	0.2%
ping-pong ball	0.1%
croquet ball	0.0%
tennis ball	0.0%

More golf balls

golf ball	
killer whale	30.3%
grey whale	9.1%
red-breasted merganser	6.7%
speedboat	3.5%
pelican	2.7%

The winner: Safety pins (perfectly clean iconic images)

Safety pin	
Safety pin	99.980%

Safety pin	
Safety pin	100.000000%

Future work

- The industry needs better metrics than the simplistic top-1 and top-5. For example:
  - Scoring based on the confidence percentage
  - Penalizing wildly incorrect predictions that are far out in the hierarchy level (wrong dog breed is less of an error than predicting a marine mammal on a golf course)
- We need better tools for creating and cleaning large datasets to minimize bad data
  - Large datasets have enabled vast improvements, now we need to enable extremely high quality datasets as well

Conclusion

- Deep neural networks for image classification are continuing to advance, but can be no better than your dataset and classification system allows
- Careful construction of your dataset for your use case can help eliminate the wildly wrong results that a human would not make

Resources

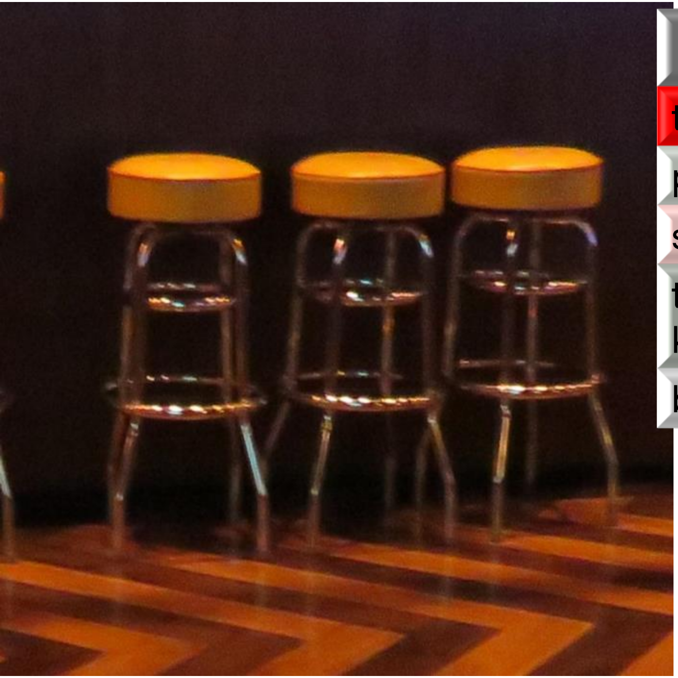
- Embedded Vision Alliance links
  - https://www.embedded-vision.com/industry-analysis/technical-articles/are-neural-networks-future-machine-vision
- AMD links
  - https://gpuopen.com/compute-product/miopen/
- Industry references:
  - ImageNet: http://www.image-net.org/
  - DNN Overview: https://arxiv.org/pdf/1703.09039.pdf
  - Inception v4: https://arxiv.org/pdf/1602.07261v2.pdf
  - Labeling Images: http://cs.stanford.edu/people/karpathy/ilsvrc/

- Visit the AMD demo **(booth 800)** in the showcase to see demos of the tools we developed to analyze datasets

Backup Slides

Soccer balls

Some images with objects not in the training set



Barstool (not in list)	
thimble	18.4%
pill bottle	8.6%
saltshaker	4.0%
typewriter keyboard	2.4%
beer bottle	1.9%

Barstool (not in list)	
nail	72.5%
safety pin	2.9%
hook	2.8%
screw	1.5%
plate rack	1.4%

Pool Table and air hockey

Pool table	
pool table	99.99%

Air hockey (not in list)	
Pool table	56.9%
lifeboat	8.6%
harvester	4.0%
scale	1.2%
fireboat	0.9%

First-time snake hunter

rattlesnake	
sea snake	22.1%
garfish	16.9%
eel	11.4%
water snake	6.6%
Indian cobra	6.2%

Roman ruins

Coliseum (not in list)	
palace	56.6%
monastery	13.2%
castle	3.4%
bell cote	2.3%
triumphal arch	1.8%

Famous statues

Statue of David (not in list)	
pedestal	73.6%
bikini	13.9%
fountain	0.7%
palace	0.6%
bannister	0.5%

Crash Helmet (changes in pose)

Crash Helmet	
crash helmet	83.3%
computer mouse	12.1%
waffle iron	0.3%

Crash Helmet	
jeweler's loupe	75.1%
waffle iron	3.3%
crash helmet	2.9%

Crash Helmet	
computer mouse	47.7%
waffle iron	12.1%
crash helmet	10.4%

More Examples: Real “background” Photos


grass


sidewalk


asphalt road


parking lot


parking lot

	grass		sidewalk		asphalt road		parking lot		parking lot	
Inception V4	golf ball	24%	ladle	3%	sandbar	4%	velvet	13%	velvet	6%
Resnet-50	Rottweiler	6%	missile	9%	hockey puck	94%	artic fox	7%	paper towel	33%
Resnet-101	Rottweiler	4%	hockey puck	6%	hockey puck	99%	electrical switch	7%	paper towel	74%
Resnet-152	poodle	6%	dirigible	50%	hockey puck	99%	toilet tissue	23%	bath towel	67%
VGG-16	hay	25%	leatherback turtle	20%	hockey puck	8%	paper towel	10%	bath towel	33%
VGG-19	hay	22%	missile	20%	ski	6%	bath towel	71%	bath towel	27%

These are not adversarial images, they are unedited natural photos (except for cropping)