# Nine is bigger than yours:

### Big datasets and computer vision

D.A. Forsyth, UIUC (and I'll omit the guilty)

#### Conclusion

- Collecting datasets is highly creative
  - rather than a nuisance activity
- There are no best practices for avoiding problems
  - apart from knowing precisely that your dataset represents problem well
- The dataset you collect can lock you into an ideology
  - which might get tricky

#### How do we assess different datasets?

- By what they are for
  - activity vs category
- By what they cover
  - many cases vs few
- By how well they represent the problem
  - in some special cases, it is easy to tell
  - what is the problem?
- By how big they are
  - easy!

#### Conclusion

- Collecting datasets is highly creative
  - rather than a nuisance activity
- There are no best practices for avoiding problems
  - apart from knowing precisely that your dataset represents problem well
- The dataset you collect can lock you into an ideology
  - which might get tricky

### Non-parametric regression



With a broad view of "match", "smooth", all classifiers fit into this story

### A=Image, B=Body pose

• Rosales+Sclaroff, 00; Shakhnarovich+Darrell, 03





## A=Image with hole, B=fill-in

Efros+Leung, 99; Hays+Efros 07



Input

Scene Matches

Output

### A=picture, B=location





#### Hays+Efros, 08

### A=picture, B=category

- Far too many to select one!
- Fergus et al 05; Fergus et al 04;Fei-Fei 06; Berg 05; Everingham et al Pascal Challenge reports 06, 07, 08;
  - etc etc etc etc etc

Table 1. Overall classification performance of the system, in various configurations, to 4289 control images and 565 test images. Configuration F is the primary configuration of the grouper, fixed before the experiment was run, which reports a nude present if either a girdle, a limb-segment girdle or a spine group is present, but not if a limb group is present. Other configurations represent various permutations of these reporting conditions; for example, configuration A reports a person present only if girdles are present. There are fewer than 15 cases, because some cases give exactly the same response.



#### Forsyth etal 96, 01

### A=motion window, B=words



Laptev Perez 2007; see also Laptev et al 08



### A=face image, B=name



President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters



Berg et al 04, 05; Guillaumin et al 08; Everingham et al 06; Ozkan et al 06; Zhao et al 08; Yagnik et al 07; lots of others

### A=picture, B=words

Wang et al 09

dog, pet, animal, Dogs! Dogs! Dogs!.



Find visually similar internet images



Cute, puppy, canine



Input Image



### Got the data, can't yet solve the problem!



A **dolphin** holds a **basketball** as it swims on its **back** .

A **dolphin** swimming upside down while holding a **basketball** 

A **dolphin** swims upside down holding a **basketball** between it 's **flippers**.

A **seal** floats on it 's back in the **water**, holding a **basketball**.

The **dolphin** on his **back** holds the **orange basketball** 

A=images, B=sentences

### Conclusion

- Collecting datasets is highly creative
  - rather than a nuisance activity
- There are no best practices for avoiding problems
  - apart from knowing precisely that your dataset represents problem well
- The dataset you collect can lock you into an ideology
  - which might get tricky

### Recognition datasets

#### • Collection strategies

- Web pix + fix
  - Flickr
  - Google image search
  - Microsoft image search
- Existing collections
  - Corel
- Photograph yourself
- Photograph isolated, then enrich

### Gotchas!

#### • Web pix+fix

- Bias (more later!)
- Might be few of the right kind (Sapp et al 08)

This difficulty probably exaggerated





A great hammer to hammer ... 400 x 378 - 67k - jpg www.drukhier.nl



386 x 385 - 7k - jpg rubayeet.wordpress.com



The Hammer is the most basic of all ...

386 x 385 - 9k - jpg homerepair.about.com hammer

300 x 400 - 15k - jpg bombmatt.wordpress.com www.edspresso.com



If I had a hammer 490 x 433 - 7k - gif 600 x 710 - 172k - jpg uzar.wordpress.com





Hammer 400 x 340 - 20k - jpg www.bbc.co.uk



307 x 307 - 15k - png

[ More from

upload.wikimedia.org

commons.wikimedia.org



RIP HAMMER HAM1 350 x 350 - 13k - jpg www.ancinterproducts.com



Hammer OS Certified Systems Engineer 300 x 450 - 23k - jpg hammeros.wordpress.com



Ultimate Geeks Multi Tool Hammer 382 x 351 - 39k - jpg nexus404.com



360 x 360 - 9k www.germesonline.com



1600 x 1200 - 66k - jpg library.thinkquest.org



... hammer beer bottle opener. 450 x 381 - 9k - jpg www.geekologie.com



shingler's hammers 300 x 300 - 9k - jpg



Claw Hammer 360 x 360 - 7k - jpg www.daviddarling.info www.lakewoodconferences.com More from www.lakewoodconferences.com



Machinists Hammer With Fibre Glass ... 360 x 360 - 7k zhukegiang.en.alibaba.com

We didn't really use

hammers much ... 500 x 362 - 29k - jpg ocw.mit.edu



... the Hammer. 620 x 344 - 47k - jpg www.didntyouhear.com





Large Chocolate Hammer: 504 x 262 - 20k - jpg www.creativechocolatesofvt.com

Hammer toe 400 x 320 - 11k - jpg www.mdconsult.com





### Gotchas!

#### • Existing collections

- mainly stock photo's like Corel
- Massive bias issues with corel
  - one can predict CD number from color histogram rather accurately (Chappelle et al, 99)
- Photograph yourself
  - hard work

### Gotchas!

#### • Enriching

- Use a probabilistic "model" to
  - enrich background
  - vary foreground
- DANGER
  - strong unnatural high frequencies at blend
  - unnatural illumination relations
  - no surface texture distortion

#### • Random

• Example: aspect and symmetry



Sapp Saxena Ng, 08 AAAI

### Recognition datasets

#### • Taxonomy strategies

- Choose some categories (Fei-Fei 04; Griffin 07; Everingham 06)
- Wordnet (Deng 09)
- Other?
- Labelling strategies
  - query image search, check responses (Fei-Fei 04; Griffin 07; Everingham 06)
  - tagging by volunteers
    - benevolent people (Antonio's mom) (Russell 08)
    - game players (von Ahn 04)
  - tagging by paid annotators (Yao 07; Sorokin 08)
  - active learning (Berg, 06; Li, 06; Wang 08; Collins 08)

#### Bias (which isn't always bad)

- If all the faces on the web are politicians
  - one needs only to be good at politicians to be good at the web
- If no users can tell an ape from a monkey
  - you might not have to either
- If people really only want to search videos for "kissing"
  - then you don't need a general activity recognition strategy

#### Selection Bias

Should not be perjorative

#### • Frequencies in the dataset may not reflect the application

- Because of what gets collected
  - eg. pictures from the web are selected not like a camera on head
  - eg. "Profession" labelling for faces in news pictures
- Because of what gets labelled
  - P(labelledlX) is not uniform
  - eg obscure but important objects in complex clutter
  - eg pedestrians in crowds
- Because the labels are often wrong
- Because the loss has unfamiliar properties
  - eg. people doing wierd things in pedestrian datasets

s X=data Y=labels X\_i = unlabelled examples (X\_j, Y\_j)=labelled examples

#### Collection issues

- Collected data is not a fair sample of X
  - X\_i, X\_j not like X
- Images on the web are "selected"



#### Loeff et al, 06

- Iconography seems to be a big effect here
  - visual "modes" of representation
    - see Berg+Berg, next talk for which one is good
  - we might not see them all
  - cf Google image search with Flickr

### Iconographic phenomena



Jing+Baluja, 08

#### Google "rooms"



... virtual tour > room photos 644 x 446 - 39k - jpg www.mandalaybay.com



Bed Room Sets 599 x 402 - 33k - jpg www.chiphi-pi.org



16 Creative and Sexy Art Hotel Rooms ... 468 x 354 - 111k - jpg weburbanist.com More from weburbanist.com



Rooms > 450 x 300 - 25k - jpg www.radisson.com More from www.radisson.com



Bookcase Secret Room Door 468 x 391 - 98k - jpg



The large room known today as the ... 350 x 353 - 48k - jpg

weburbanist.com www.royalacademy.org.uk



To reserve a room call 212-596-1200 ... 640 x 480 - 93k - jpg www.columbiaclub.org



Now let's see some amazing rooms. 450 x 300 - 19k - jpg freshome.com



Room for physicallychallenged 600 x 395 - 244k - jpg www.hotelnikkohanoi.com.vn



basement family room 450 x 325 - 48k - jpg www.thisoldhouse.com



Handicap Room 300 x 301 - 22k - jpg intl-house.howardhotels.com



Spacious Guest Room 450 x 300 - 29k - jpg www.radisson.com

twin beds and ... 370 x 486 - 40k - jpg www.inisrael.com



Rooms may also include This bright room on the 2nd floor of ... 1728 x 1152 - 283k - jpg biosphere.ec.gc.ca



These twenty rooms ... 468 x 352 - 97k - jpg weburbanist.com



Texas' enormous locker room facility ... 530 x 343 - 34k - ipa



Two Queen Room 450 x 300 - 26k - jpg



trent room The Trent Room was first ...



Image of changing room



Tour the USC Marshall Capture Room www.countryinns.com 346 x 450 - 54k - ipg 450 x 388 - 75k - ipg 637 x 481 - 160k - ipg



large drawing room in two room suite 737 x 551 - 70k - ipa

### Flickr "rooms"















New living roon by flowers & machine 🗐 29 comments 👷 1

notes

Tagged with art, home, vint Taken on December 3, 200 December 4, 2007

See more of flowers & photos, or visit her profile.

South Side of M (Fisheye) by joeys 🗐 6 comments 🛛 👷 9 fa notes Tagged with pets, cat, close Taken on January 16, 2008 16, 2008 See more of joeysplar visit his profile.





1

















🗐 147 comments 🔹 notes Tagged with nikkor50mmf1 argbacktoworktoday, samer Taken on January 3, 2009, 5, 2009

### Selection for labelling

- P(labelledlX) is not uniform
  - or P(Xllabelled) is not the same as P(Xlnot labelled)

#### X\_j are not like X\_i

#### • There are models

- problem sometimes called dataset shift, see (Quinonero-Candela 09)
- can be addressed with, say, large unlabelled datasets

#### • Important effect

- can make high capacity classifiers generalize better than low capacity
- (maybe) be very cautious about linear SVM's

#### Labels that are wrong

#### • Fact of life

- Can fix when there are many instances
  - consistency (Zhao et al 08)
  - smoothing (Berg, 06; Li, 06; Wang 08; Collins 08)
- Might be able to fix with hierarchy+generalization
  - we should never mix up "cat"'s and "truck"'s

#### Nasties

#### • Unfamiliar loss:

- eg Pedestrians
  - We may not run down people who behave strangely
  - i.e application demands a loss guarantee that is absolute in form
    - not relative to distribution

#### • Collective overfitting:

• we should not repeatedly compare features, methods on the same test set

### Conclusion

- Collecting datasets is highly creative
  - rather than a nuisance activity
- There are no best practices for avoiding problems
  - apart from knowing precisely that your dataset represents problem well

You can't get away from bias by saying you must know your problem well before you collect

- The dataset you collect can lock you into an ideology
  - which might get tricky

Obtain dataset

Build features

Mess around with classifiers, probability, etc

Produce representation

### Big questions



• What should we say about visual data?

### Taxonomy/Category problems

• A choice of taxonomy is a profound commitment

- which may enhance/distort future research
- Examples:
  - Activity
  - Recognition

#### Activity taxonomy = k classes

#### • Research strategy

- Rack up a bunch of activity categories, and discriminate
- Signal representation
  - body probably doesn't matter
- Unsatisfying, because
  - how many categories are enough?
  - can one movement have two categories?
  - what are the categories?
    - the verb argument (probably) fails
      - if there are few movement, many goal verbs
    - introspection suggests too few words

AnswerPhone, GetOutCar, Handshake, Kiss, Hugperson, SitDown, SitUp, StandUp

Goal achieved by body movement

Body movement

### Activity taxonomy=map goals, intentions

#### • Attractive, because

- Activity composes freely into complex structures
- Most human activities cause changes of state, meet goals
  - similar movements will meet different goals
  - different movements can meet the same goal

#### • Research goals

- We should probably be trying to "recognize" things
  - whose names we do not know, affected by
    - nearby objects
    - observer, observation context
  - for which we have seen no examples

#### • Signal representation

• probably do need to segment the body

### Taxonomy/Category problems

• A choice of taxonomy is a profound commitment

- which may enhance/distort future research
- Examples:
  - Activity
  - Recognition

### Object recognition = k class classification

- current data sets ok,
  - improve coverage
- research agenda:
  - more features, better classifiers:
  - perhaps category hierarchies for statistical leverage (tying)

#### Are these monkeys?





pider Monkey, Spider Monkey OMFG MONKEY Profile ... NIPS2 470 x 324 - 29k - jpg 444 x 398 - 40k - jpg animals.nationalgeographic.com www.bestweekever.tv More from More from nimals.nationalgeographic.com www.bestweekever.tv



Vampire Monkey 350 x 500 - 32k - jpg paranormal.about.com



... monkeys for ...





The Monkey Cage 300 x 306 - 35k - jpg www.themonkeycage.org



... be monkey ... 300 x 350 - 29k - jpg my.opera.com



... monkey's interests ... 378 x 470 - 85k - jpg www.schwimmerlegal.com



"You will be a monkey. 358 x 480 - 38k - jpg kulxp.blogspot.com



... monkey and I am 342 x 324 - 17k - jpg www.azcazandco.com



Monkey 353 x 408 - 423k - bmp www.graphicshunt.com



The Monkey Park 400 x 402 - 24k - jpg www.lysator.liu.se



Monkey cloning follow up ... 450 x 316 - 17k - jpg blog.bioethics.net



So here's one of my monkeys 400 x 300 - 13k - jpg www.gamespot.com



monkeys ... 400 x 310 - 85k - jpg ioaquinvargas.com



MONKEY TEETH 308 x 311 - 18k - jpg repairstemcell.wordpress.com



The Blow Monkey is Spider Monkey Picture, Spider Monkey ... 500 x 500 - 30k - jpg 800 x 600 - 75k - jpg



a..... monkey! mammal monkey 525 x 525 - 99k - jpg www.uberreview.com animals.nationalgeographic.com www.sodahead.com



WTF Monkey



Monkey



374 x 300 - 23k - jpg 512 x 768 - 344k - jpg 787 x 1024 - 131k - jpg www.myspace.com www.exzooberance.com runrigging.blogspot.com

### Object recognition = describing what objects are like

#### • current datasets

- are largely of the wrong form
  - and no declarative data about objects

#### • research agenda

- learning by reading
- sensible responses to objects of unknown category
- within class variance has semantics
- architectures, representations, semantics

#### Attributes



#### Farhadi et al 09; Lampert 09

Collect attribute labels
5 e5 = \$600

### Conclusion

- Collecting datasets is highly creative
  - rather than a nuisance activity
- The dataset you collect can lock you into an ideology
  - which might get tricky
- There are no best practices for avoiding problems
  - apart from knowing precisely that your dataset represents problem well