



# Introductory Overview Lecture CASA course (5/10/2018)

D. Freire-Obregón, PhD.









#### Plan for the 1st part of the CASA course

#### Week 0:

TS1: Introduction to CV & DL

#### Week 1:

- LS1: Our first NN. Different approaches.
- TS2: Fundamentals of ML & DL.

#### Week 2:

- LS2: Regularization is critical.
- TS3: Why regularization.









#### Plan for the 1st part of the CASA course



- LS3: DL on images.

TS4: DL on images.

Week 4:

- LS4: Convolutional NN.

TS5: Convolutional NN.









#### Plan for TS1

- Computer-Human vision overview
- Computer graphics vs. Computer Vision
- Last 20 years of papers (just the relevants)
- CV...ML or DL?
- Current state of the art
- Working with images? Feasible?

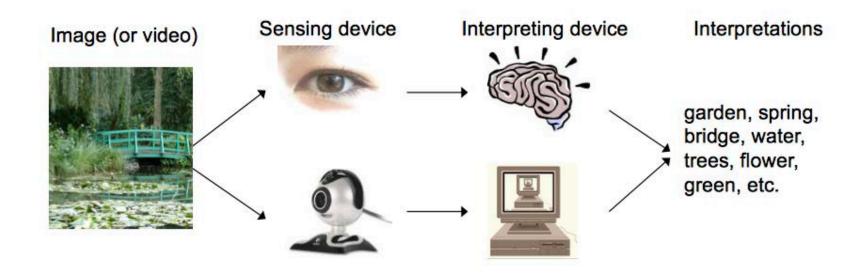












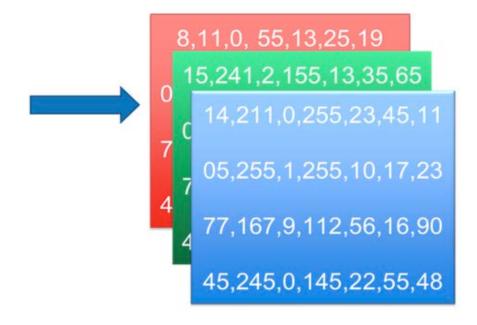








This is what the computer sees
An image is a big grid of numbers [0, 255]
e.g. 200x200x3 (3 channels RGB)



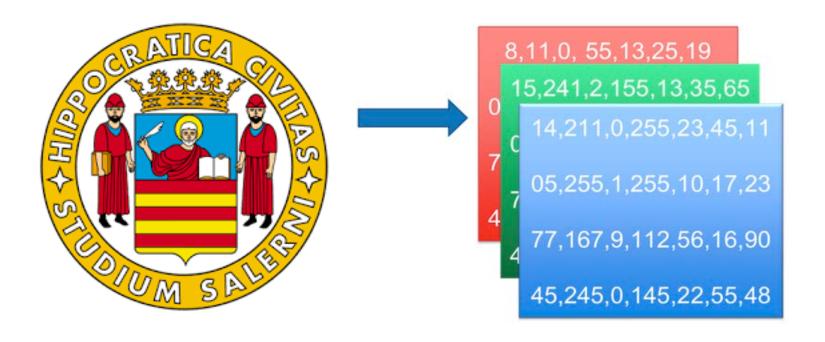








Challenge → Viewpoint variation











¿What is this? ¿A rock?¿A palm tree?













The perspective on what you see is relevant.

















Other challenges:

Illumination

ATICA CULTURE OF THE STATE OF T

**Deformation** 





Intraclass variations







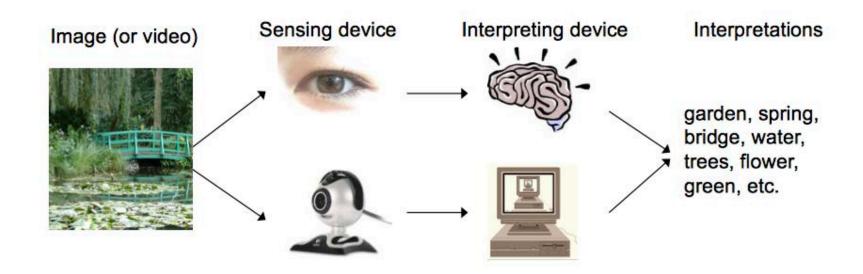
35017 Las Palmas de Gran Canaria e-mail: info@siani.es · www.siani.es







#### Where?











What do you see?

Which team is it? How do you know?











# Not because of the people...



















### Let's train considering the shirt



























#### Let's train considering the shirt















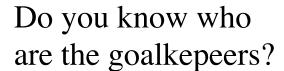






















#### Let's train considering the shirt































Again: 1-human segmentation 2-pattern recognition















































What about the Badge?? ©

Please...find one for me in the picture.











# Training data!



















#### Good luck with that



























#### Did they win?



How do you know that?

Tell me our best shot











#### Pattern Recognition:



#### Facial expression recognition:











e-mail: info@siani.es - www.siani.es







#### Google Street View























Not so easy...

















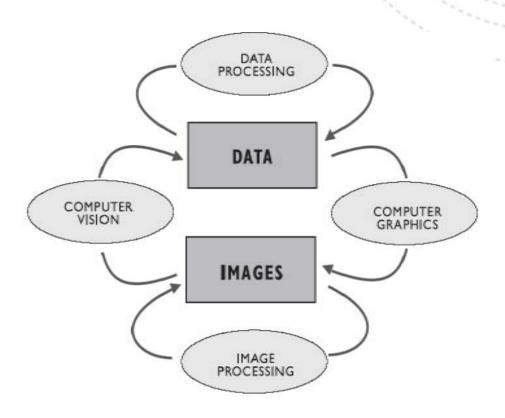






# Computer Graphics vs Computer Vision







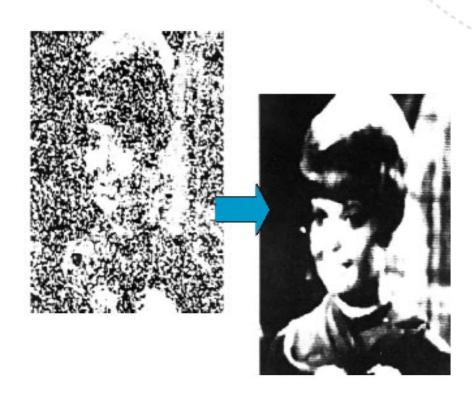
35017 Las Palmas de Gran Canaria e-mail: info@siani.es · www.siani.es







#### Noise removal





















#### Image enhancement



Original

Automatic Enhancement







Campus Universitario de Tafira 35017 Las Palmas de Gran Canaria e-mail: info@siani.es - www.siani.es







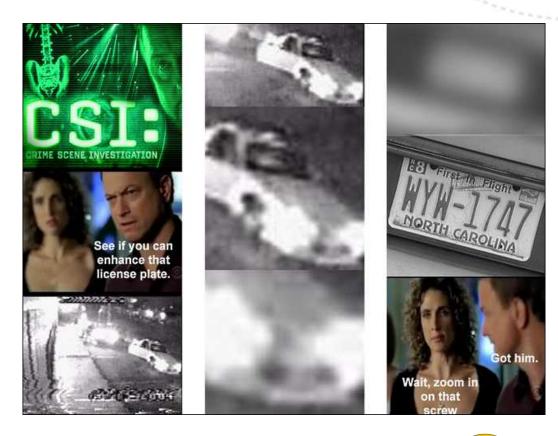






# Image enhancement

#### Not real





35017 Las Palmas de Gran Canaria e-mail: info@siani.es · www.siani.es







# Image restoration



























"Computer Vision describes the automatic deduction of the structure and the properties of a (possible dynamic) threedimensional world from either a single or multiple twodimensional images of the world"

Research area -> Computers perceive, process and understand visual data





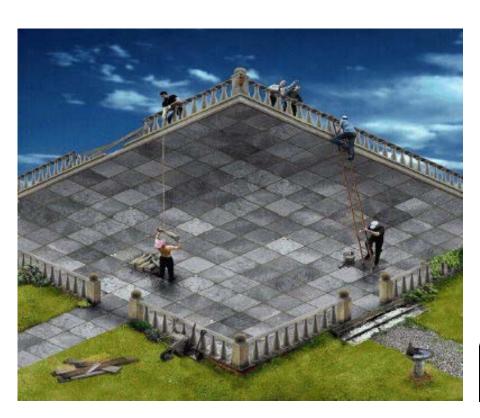




# Do we need to copy the human vision approach?



Is it reliable enough? Perception





















Humans → See in 3D Machines → Only in 2D

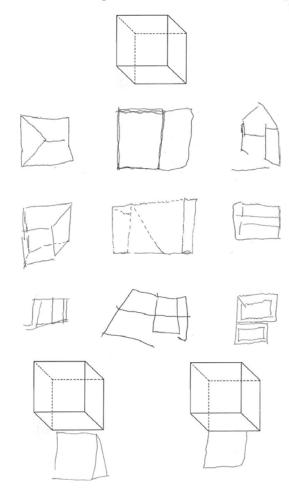




























#### **Innovations**



Hardware

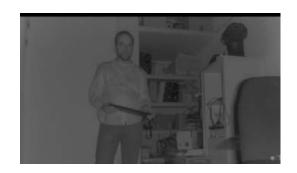
Kinect, 2010

https://www.youtube.com/watch?v=

StQyUdeux0o

















## The last 20 years: Normalized cut



https://www.youtube.com/watch?v=eUhvKEC3YTc



35017 Las Palmas de Gran Canaria e-mail: info@siani.es · www.siani.es







## The last 20 years: SIFT & Object recog.









Image is CC BY-SA 2.0



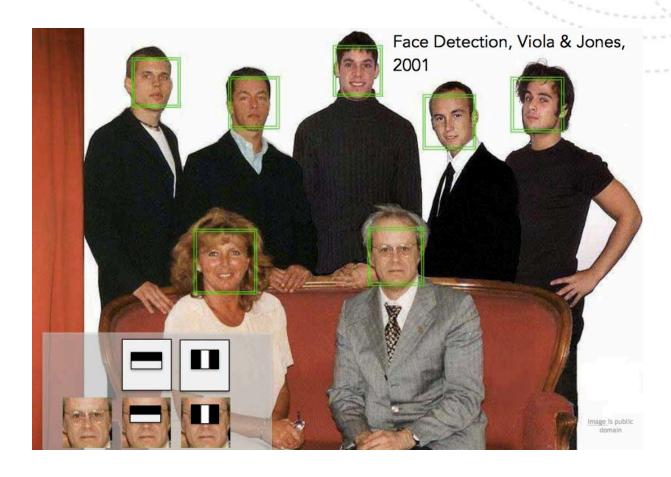
e-mail: info@siani.es - www.siani.es







## The last 20 years: Face Detector





e-mail: info@siani.es - www.siani.es





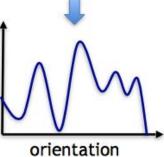




## The last 20 years: HoG

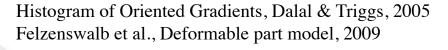
Image is CC0 1.0 public domain













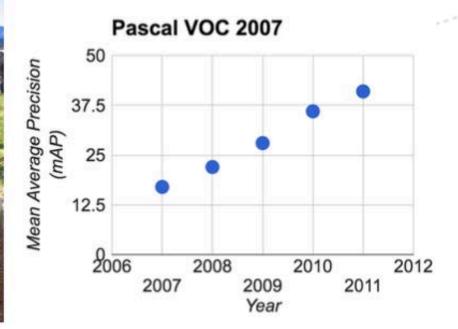




## The last 20 years: Competitions



This image is licensed under CC BY-SA 2.0; changes made





















## The last 20 years: Imagenet

22K categories and 14M images e.g. Animals:

- Bird
- Fish
- Mammal
- Invertebrate

#### Plants:

- Tree
- Flower

Food

Person

. . .



e-mail: info@siani.es - www.siani.es







# The last 20 years: The Image Class. Challenge

1,000 object classes 1,431,167 images











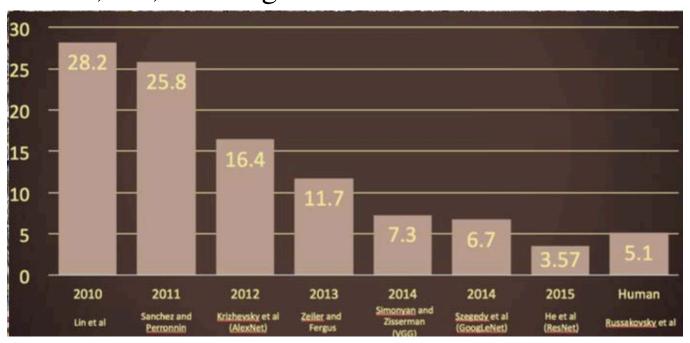






# The last 20 years: The Image Class. Challenge

1,000 object classes 1,431,167 images





35017 Las Palmas de Gran Canaria e-mail: info@siani.es · www.siani.es







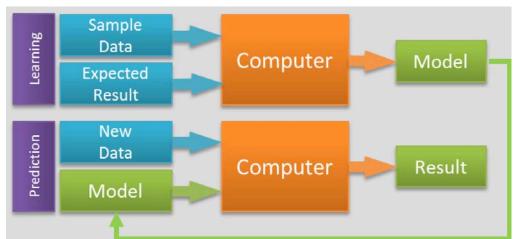
## INGENIERIA COMPUTACIONAL

#### CV & ML

Traditional modeling - CV (feat.ext.)+ML(class.):



#### Deep learning:





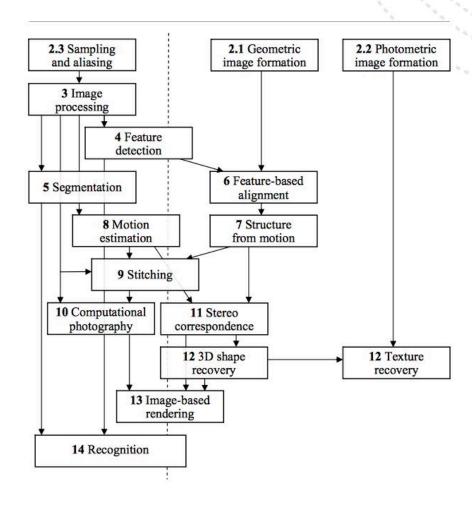
e-mail: info@siani.es - www.siani.es







#### Structure of a CV book











#### Structure of a DL book

**Fundamentals:** 

Brands of ML

Evaluation

Data processing

Overfitting & underfitting

DL in practice:

**CNN** 

**RNN** 

Generative models

Features??



e-mail: info@siani.es - www.siani.es







#### Where is DL?

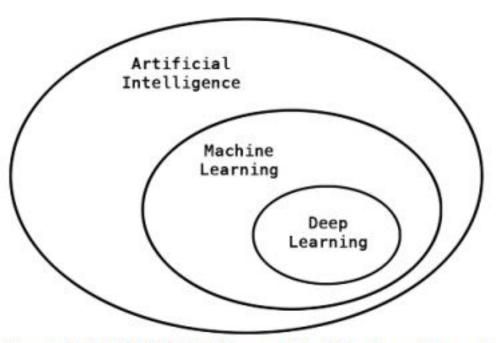


Figure 1.1 Artificial Intelligence, Machine Learning and Deep Learning



e-mail: info@siani.es - www.siani.es







## Why DL? Why now?

- 1. Hardware:
  - 1. GPUs...NVIDIA rules!
  - 2. TPUs...10xGPUs.
- 2. Data
  - 1. A lot places with datasets available
- 3. Algorithms:
  - 1. Better activation functions
  - 2. Better optimization techniques
  - 3. Better weight-initialization methods
- 4. New wave of investment:
  - 1.  $2011 \rightarrow 19M \dots 2014 \rightarrow 394M \$$









#### Where are we now?

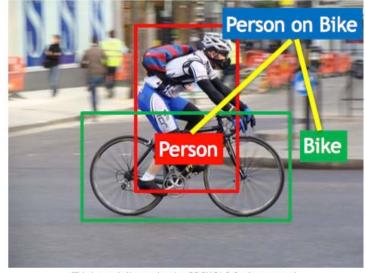


Object detection Action classification Image captioning

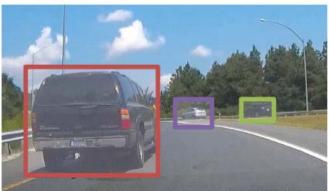


This image is licensed under CC BY-SA 2.0; changes made

Person



This image is licensed under CC BY-SA 3.0; changes made



This image is licensed under CC BY-NC-SA 2.0; changes made









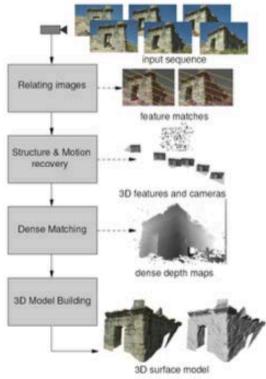


#### Vision as measurement device



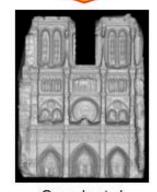






Pollefeys et al.





Goesele et al.





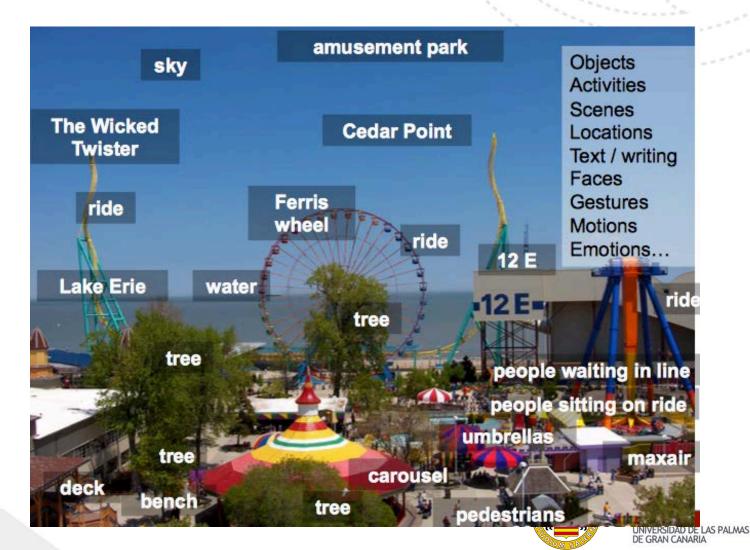


e-mail: info@siani.es · www.siani.es





## Vision as a source of semantic information





e-mail: info@siani.es · www.siani.es



## Data everywhere



























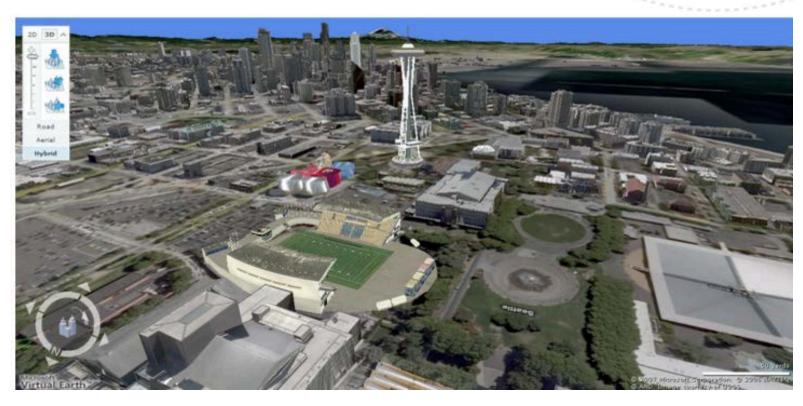






## 3D urban modeling























#### Cameras: face detection





















#### Cameras: smile detection















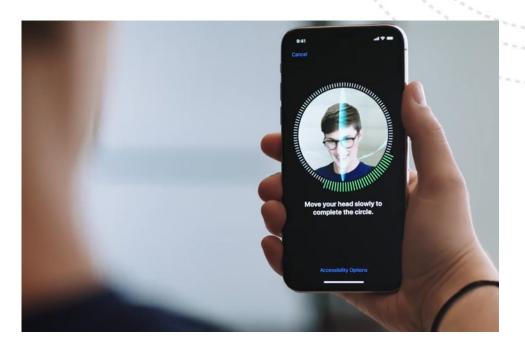








## Face recognition



https://www.youtube.com/watch?v=FhbMLmsCax0

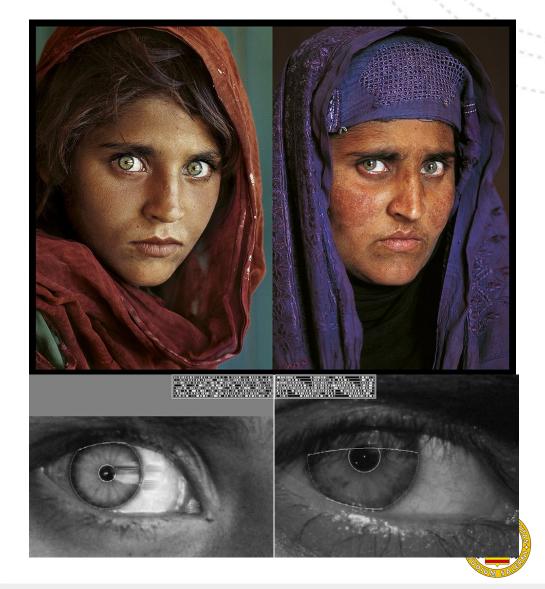








### **Biometrics**



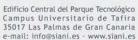
















### Optical character recognition (OCR)





Digit recognition, AT&T labs



License plate readers

http://en.wikipedia.org/wiki/Automatic\_number\_plate\_recognition









## 



## Toys & robots





















## Mobile visual search: Google Goggles

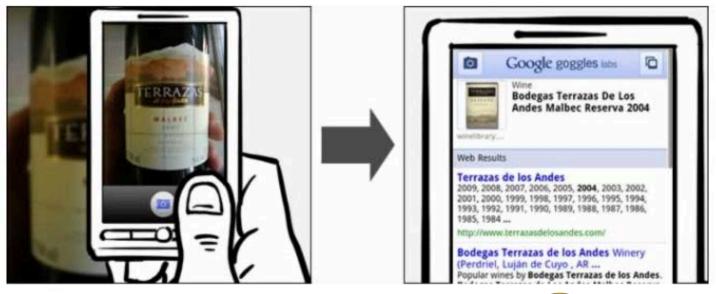


#### Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.









35017 Las Palmas de Gran Canaria e-mail: info@siani.es · www.siani.es







## Mobile visual search: iPhone apps





Query Images















































#### Where are we now?



#### https://www.youtube.com/watch?v=OoC8oH0CLGc

Transfer control: Human and machine.

Two cameras in the driver: face and back.

2 Deep NN: Raw segmentation + Object detection

Glance reading classification: 3D CNN

High level planning decisions:

- Transfer control or stop vehicle
- Decision fussion algorithm
  - Combines:
  - Risk factor at the external environment
  - Driver state (paying attention?)









## Vision in supermarkets



















#### Vision based interaction





Microsoft Kinect



Assistive technologies









#### Working with images

Method	Accuracy
Best from [10]	$77.8 \pm 1.3$
Best from [23]	$79.3 \pm 0.0$
Proposed using single crop	$85.9 \pm 1.4$
Proposed using over-sample	$86.8 \pm 1.4$

Table 2. Gender estimation results on the Adience benchmark. Listed are the mean accuracy  $\pm$  standard error over all age categories. Best results are marked in bold.

Method	Exact	1-off
Best from [10]	$45.1 \pm 2.6$	$79.5 \pm 1.4$
Proposed using single crop	$49.5 \pm 4.4$	$84.6 \pm 1.7$
Proposed using over-sample	$50.7 \pm 5.1$	$84.7 \pm 2.2$

Table 3. Age estimation results on the Adience benchmark. Listed are the mean accuracy  $\pm$  standard error over all age categories. Best results are marked in bold.

1-off means that the correct decision is the nearest neighbor to the current decision.



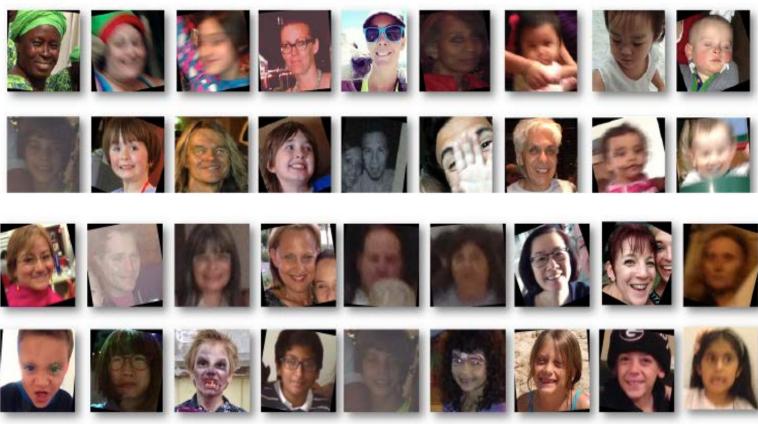






## Working with images is hard





















#### **Credits**



These slides are partially based on:

Niebles and Fei-Fei. Introduction to computer vision. Stanford AI Lab





