

Save the Whales with your data-scientist' superpowers

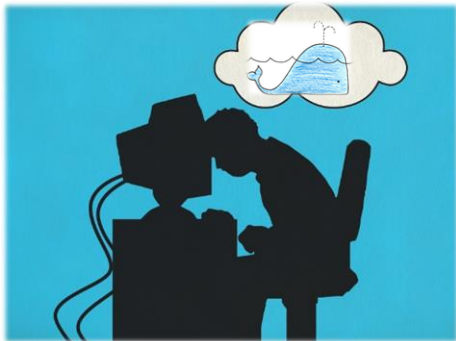
Amine EL HELOU,
amine.elhelou@mathworks.fr



kaggle™

Right Whales are in danger...

- $N \leq 500$ Right Whales left in the oceans
- Actually, whales are photographed during aerial surveys & then manually matched to a photo-identification catalog
- **Extremely time consuming & requires special training**



✓ Real-time recognition would allow researchers to spontaneously save whales as they struggle to free themselves from fishing gears



<https://www.kaggle.com/c/noaa-right-whale-recognition>

-
- Organize ▾ Include in library ▾ Share with ▾ Slide show Burn New folder
- Public
- Amine El Helout
Computer - Shortcut
Desktop
Downloads
Mathworks
Recent Places
- Libraries
- Documents
Music
Pictures
Videos
- Computer
- OSDisk (K)
- ethelhou (\\CENTRAL-fr\\home) (H)
Presentation (\\CENTRAL-FR\\Foum
Foundation (\\CENTRAL-FR) (L)
Marketing (\\CENTRAL-FR) (M)
SalesService (\\CENTRAL-FR) (N)
- | Name | Type | Size | Dimensions |
|------|------------|----------|-------------|
| w_9 | JPEG image | 448 KB | 2773 x 1848 |
| w_9 | JPEG image | 586 KB | 3038 x 2024 |
| w_4 | JPEG image | 414 KB | 2897 x 1931 |
| w_5 | JPEG image | 758 KB | 3718 x 2478 |
| w_9 | JPEG image | 503 KB | 3069 x 2046 |
| w_10 | JPEG image | 1,004 KB | 4094 x 1383 |
| w_11 | JPEG image | 523 KB | 3669 x 2446 |
| w_12 | JPEG image | 604 KB | 3114 x 2076 |
| w_13 | JPEG image | 226 KB | 2406 x 1504 |
| w_14 | JPEG image | 618 KB | 3238 x 2139 |
| w_17 | JPEG image | 860 KB | 3423 x 2282 |
| w_19 | JPEG image | 477 KB | 3012 x 2008 |
| w_20 | JPEG image | 1,187 KB | 4934 x 3089 |
| w_23 | JPEG image | 666 KB | 2894 x 2000 |
| w_24 | JPEG image | 752 KB | 3235 x 2240 |
| w_25 | JPEG image | 314 KB | 2564 x 1710 |
| w_26 | JPEG image | 397 KB | 2889 x 1903 |
| w_27 | JPEG image | 655 KB | 3535 x 2257 |
| w_29 | JPEG image | 448 KB | 3456 x 2204 |
| w_30 | JPEG image | 1,992 KB | 5117 x 3411 |
| w_31 | JPEG image | 642 KB | 2710 x 1807 |
- 6,925 items



whale_95370

N = 417

	1	2	3
	whaleID	ImgCount	FileList
1	whale_95370	47	7x1 cell
2	whale_38681	43	3x1 cell
3	whale_28892	33	3x1 cell
4	whale_90957	30	0x1 cell
5	whale_36851	30	0x1 cell
6	whale_24458	29	9x1 cell
7	whale_85464	28	8x1 cell
8	whale_51195	28	8x1 cell
9	whale_65586	27	7x1 cell
10	whale_52749	26	6x1 cell
11	whale_89615	24	4x1 cell
12	whale_78280		
13	whale_34656		
...
445	whale_08729	1	6656.jpg
446	whale_05140	1	5252.jpg
447	whale_04480	1	3576.jpg
448	whale_04480	1	1067.jpg

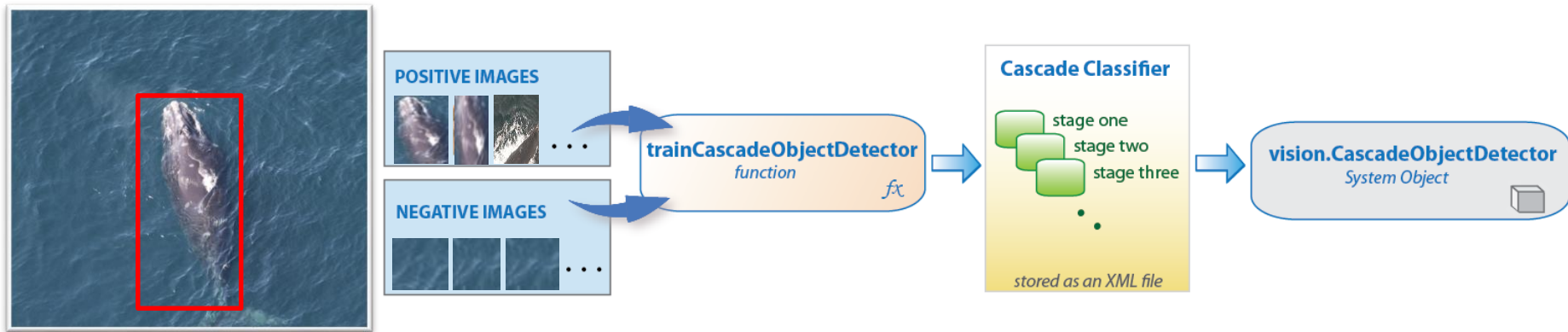


A computer vision + ML problem... ?

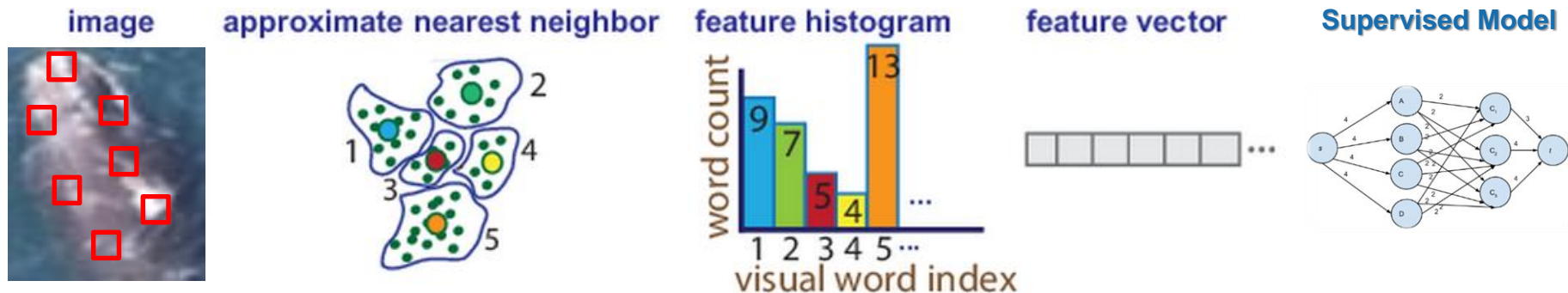
Proposed starter codes: a 2 step approach



1. Whale detection from images using a (*semi-automated*) Vision model



2. Supervised classification using features extracted from detected whale



A computer vision + ML problem... ?

Obvious challenges

- Image quality variability
- Image rotations → correction ?
- Manual labelling part → operator dependent
- Model (& Cascade detector) implementation...?
 - One vs. All (train 1 model for each specimen vs. all rest)
 - Multi-class
 - Non-homogenous response distributions



How am I scored... ?



- Evaluation criteria == Multi-class Logarithmic Loss
 - For each image: submit a set of predicted probabilities (one for every whale)

Image	whale_00195	whale_00442	whale_02411	whale_02608	whale_02839	whale_03103	whale_03227	whale_03623	whale_03731	whale_03935
w_1947.jpg	1	0	0	0	0	0	0	0	0	0
w_11096.jpg	1	0	0	0	0	0	0	0	0	0
w_10973.jpg	1	0	0	0	0	0	0	0	0	0
w_10442.jpg	1	0	0	0	0	0	0	0	0	0
w_10606.jpg	1	0	0	0	0	0	0	0	0	0
w_11167.jpg	1	0	0	0	0	0	0	0	0	0
w_1464.jpg	1	0	0	0	0	0	0	0	0	0
w_5072.jpg	1	0	0	0	0	0	0	0	0	0
w_5492.jpg	1	0	0	0	0	0	0	0	0	0
w_10124.jpg	1	0	0	0	0	0	0	0	0	0
w_9503.jpg	1	0	0	0	0	0	0	0	0	0
w_11267.jpg	1	0	0	0	0	0	0	0	0	0
w_1590.jpg	1	0	0	0	0	0	0	0	0	0
w_10175.jpg	1	0	0	0	0	0	0	0	0	0
w_10175.jpg	1	0	0	0	0	0	0	0	0	0

sample_submission.csv

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

Submission Limits

You may submit a maximum of 5 entries per day.
You may select up to 2 final submissions for judging.

- N: number of test images
- M: number of whale labels
- y_{ij} : 0/1 if observation i belongs to whale j
- p_{ij} : predicted probability

So join the fight...



- ✓ Prizes:
 - 5000/3000/2000 \$



- ✓ Deadline: 7th of January 2016

1	0	7	teams
1	2	7	players
5	5	2	entries



NOAA
FISHERIES



New England
Aquarium



MathWorks®

kaggle™