Highlights

- Collect human annotations of image difficulty for PASCAL VOC 2012
- Analyze what image properties can predict visual search difficulty
- Train regression model to predict difficulty scores
- Our predictor generalizes well to new classes
- Applications:
- Weakly supervised object localization (8% improvement)
- Semi-supervised object classification (1% improvement)
- Code and data: http://image-difficulty.herokuapp.com

Visual search difficulty = human response time

Answer by YES or NO







- 11540 images \times 2 questions \times 3 annotators = 69K annotations • Human agreement:
- One-vs-all \rightarrow compare time of an annotator to mean of all annotators
- Kendall au = 0.56 ightarrow 78% image pairs ranked in the same order
- Visual search difficulty can be consistently measured

Correlation between image properties and difficulty

Image property	Kendall τ
Number of objects	0.32
Mean area covered by objects	-0.28
Non-centeredness	0.29
Number of different classes	0.33
Number of truncated objects	0.22
Number of occluded objects	0.26
Number of difficult objects	0.20
Combine all properties with ν -SVR	0.36



Difficult = many instances of various classes scattered all over the image

Examples of lost and found objects by annotators

lost bottle



found car

90 found (not annotated in PASCAL VOC 2012) and 40 lost objects

Class	Human	[Chatfield, BMVC'14]	Class	Human	[Chatfield, BMVC'14]
	time	mAP		time	mAP
bird	3.08	92.5%	bicycle	3.41	90.4%
cat	3.13	91.9%	boat	3.44	89.6%
aeroplane	3.15	95.3%	car	3.46	91.5%
dog	3.20	89.7%	bus	3.50	81.9%
horse	3.24	92.2%	sofa	3.54	68.0%
sheep	3.24	82.9%	bottle	3.55	54.4%
COW	3.28	76.3%	tv monitor	3.57	74.4%
motorbike	3.35	86.9%	dining table	3.57	74.9%
train	3.36	95.5%	chair	3.58	64.1%
person	3.39	95.2%	potted plant	3.64	60.7%

Automatically estimating difficulty

• Baselines:

2	
-	
01	
ea	





raducu.ionescu@gmail.com

How hard can it be? Estimating the difficulty of visual search in an image

Radu Tudor Ionescu¹, Bogdan Alexe^{1,4}, Marius Leordeanu³, Marius Popescu¹, Dim P. Papadopoulos², Vittorio Ferrari²

¹University of Bucharest, ²University of Edinburgh, ³Institute of Mathematics of the Romanian Academy, ⁴Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy

Human versus machine at the class level

• IOP 6 most difficult are the same, 3 in top 5 most easy are the same

• Based on CNN features and linear regression with ν -SVR Regress from whole image to human difficulty score

Kendall τ
0.002
0.052
0.106
0.238
0.240
0.271
0.299
0.472
-

• Our model is even better than the image properties













Weakly supervised object localization



• Evaluation: CorLoc (IoU > 0.5) on PASCAL VOC 2007 trainval

Plain MIL: 34.4%

- All images in every iteration
- 9 iterations (no objectness, no fancy stuff)

Easy-to-Hard MIL: 42.8%

- Images ranked by difficulty and split into 3 batches
- Window classifier progressively updated from easy batch to hard batch
- 3 batches \times 3 iterations

