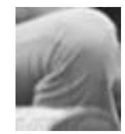
Atoms of Recognition in Human and Computer Vision

Efficient use of limited information: recognizing local configurations

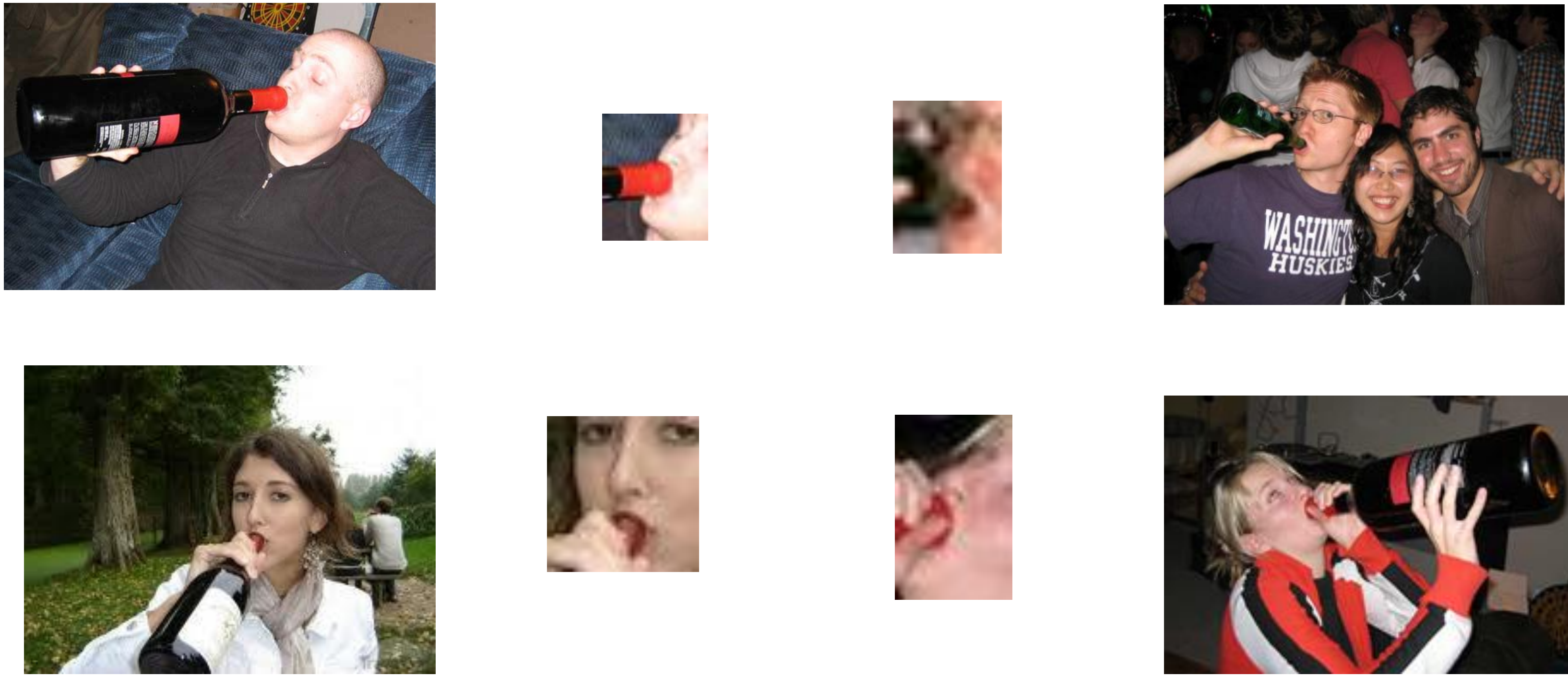




© Proceedings of the National Academy of Sciences. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Ullman, Shimon, Liav Assif, Ethan Fetaya, and Daniel Harari. "Atoms of recognition in human and computer vision. "Proceedings of the National Academy of Sciences 113, no. 10 (2016): 2744-2749.

Minimizing variability



© Proceedings of the National Academy of Sciences. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Ullman, Shimon, Liav Assif, Ethan Fetaya, and Daniel Harari. "Atoms of recognition in human and computer vision. "Proceedings of the National Academy of Sciences 113, no. 10 (2016): 2744-2749.

Useful for the interpretation of complex scenes

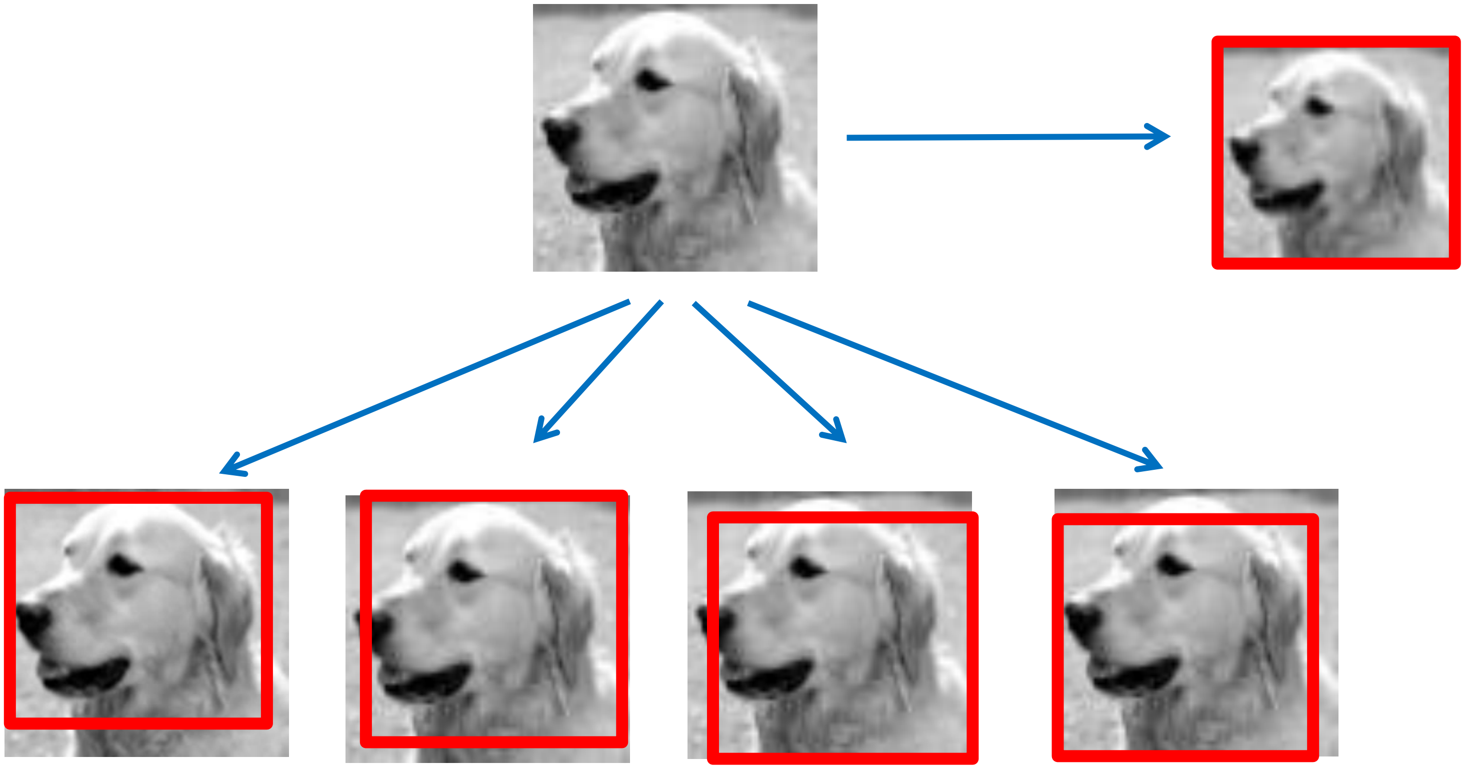
- Useful for dealing with complex scenes but challenging: non-redundant images

- Human studies
- Computational models
- Implications: representation for recognition, brain processing, CBMM

- Dan Harari
- Liav Assif
- Guy Ben-Yossef
- Eitan Fetaya
- Leyla Isik
- Yena Han

- ERC Advanced Grant ‘Digital Baby’

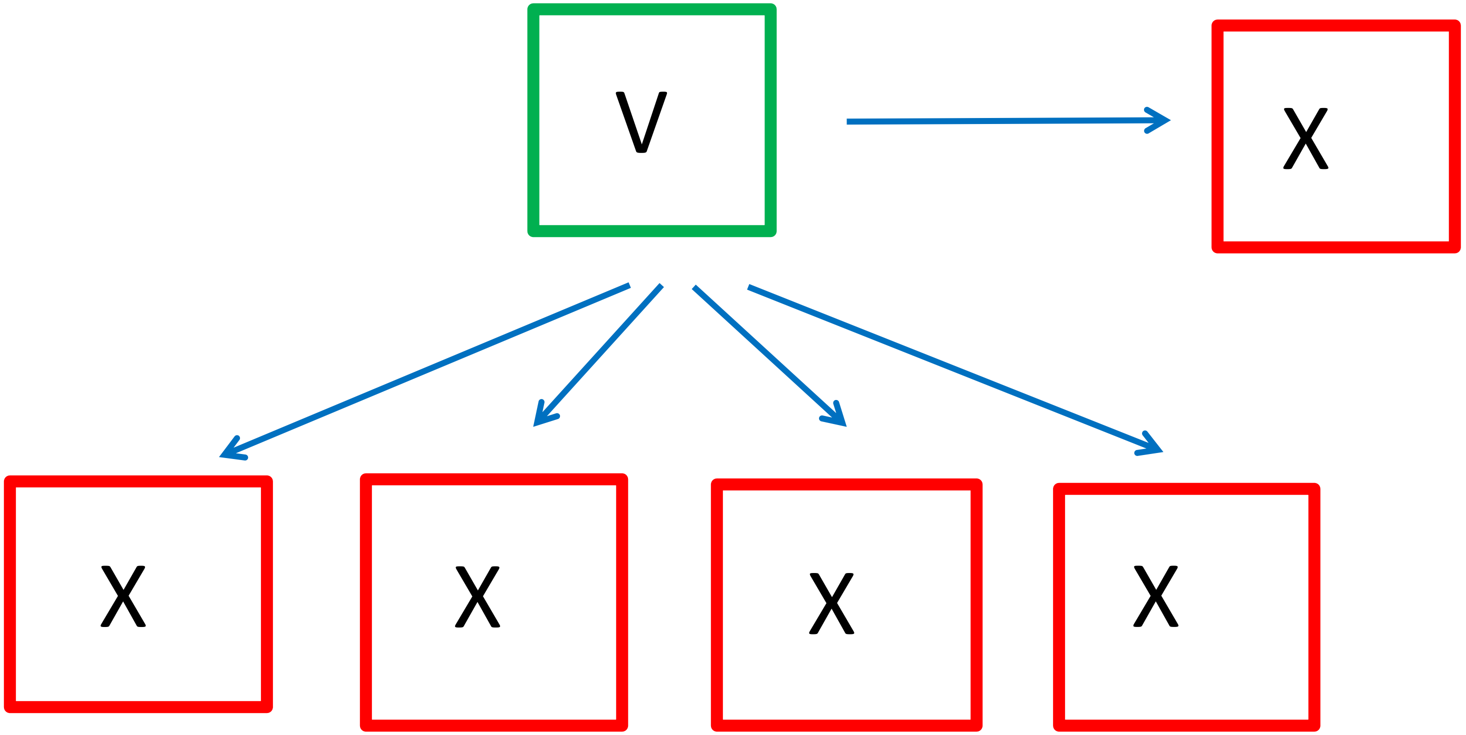
Searching for Minimal Images



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Over 15,000 subjects, laboratory controls

‘MIRC’ (Minimal Recognizable Configuration): all 5 descendants are unrecognizable



Sharp transition

Pairs

Parent – MIRC,
Child – 'sub-MIRC'

Example:

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

0.93

0.03

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

© Source Unknown. All rights reserved.This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

0.79

0.0

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Original images

Plane

Ship

House fly

Bald eagle

Horse

Bike

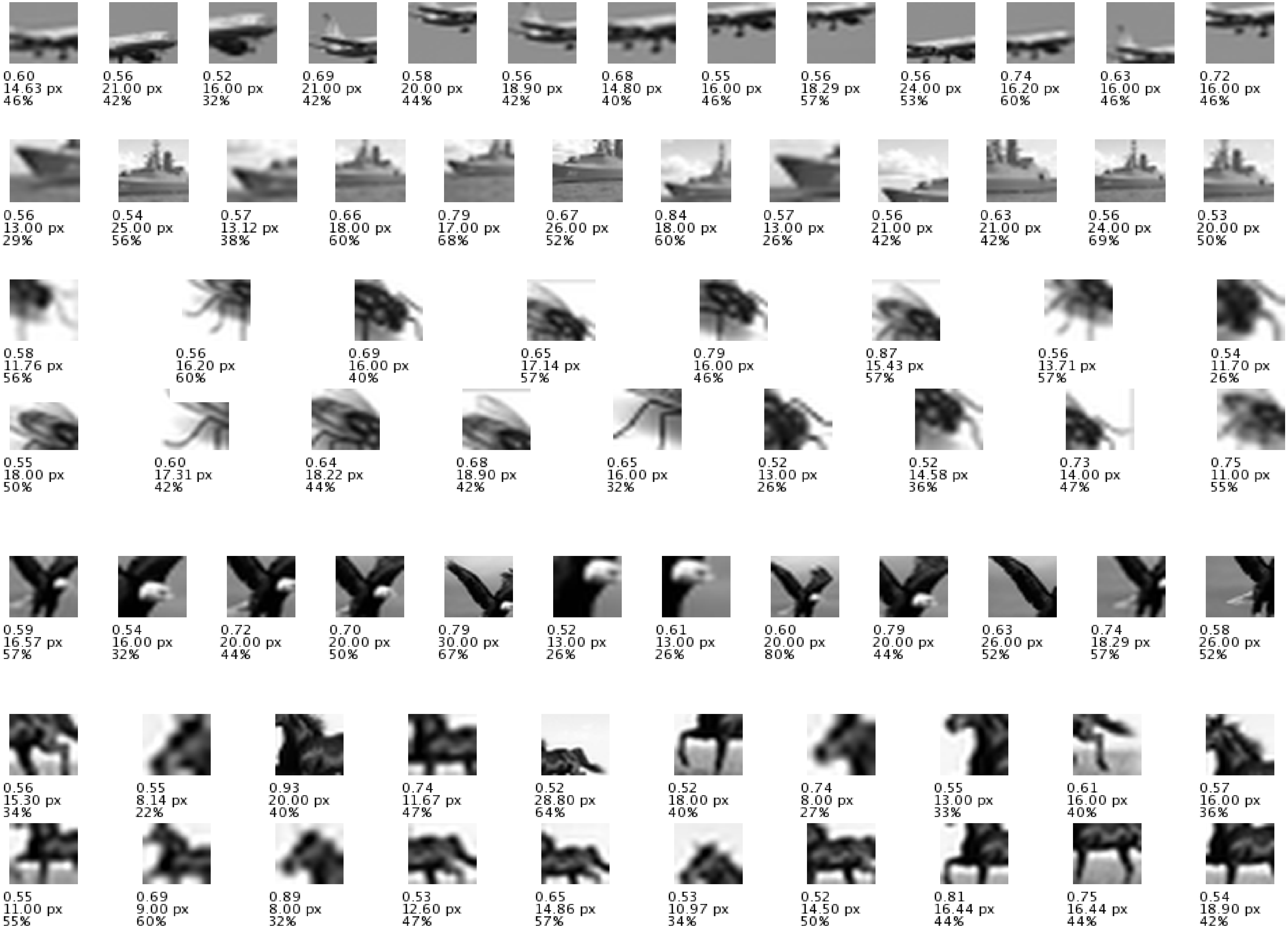
Car door

Human eye

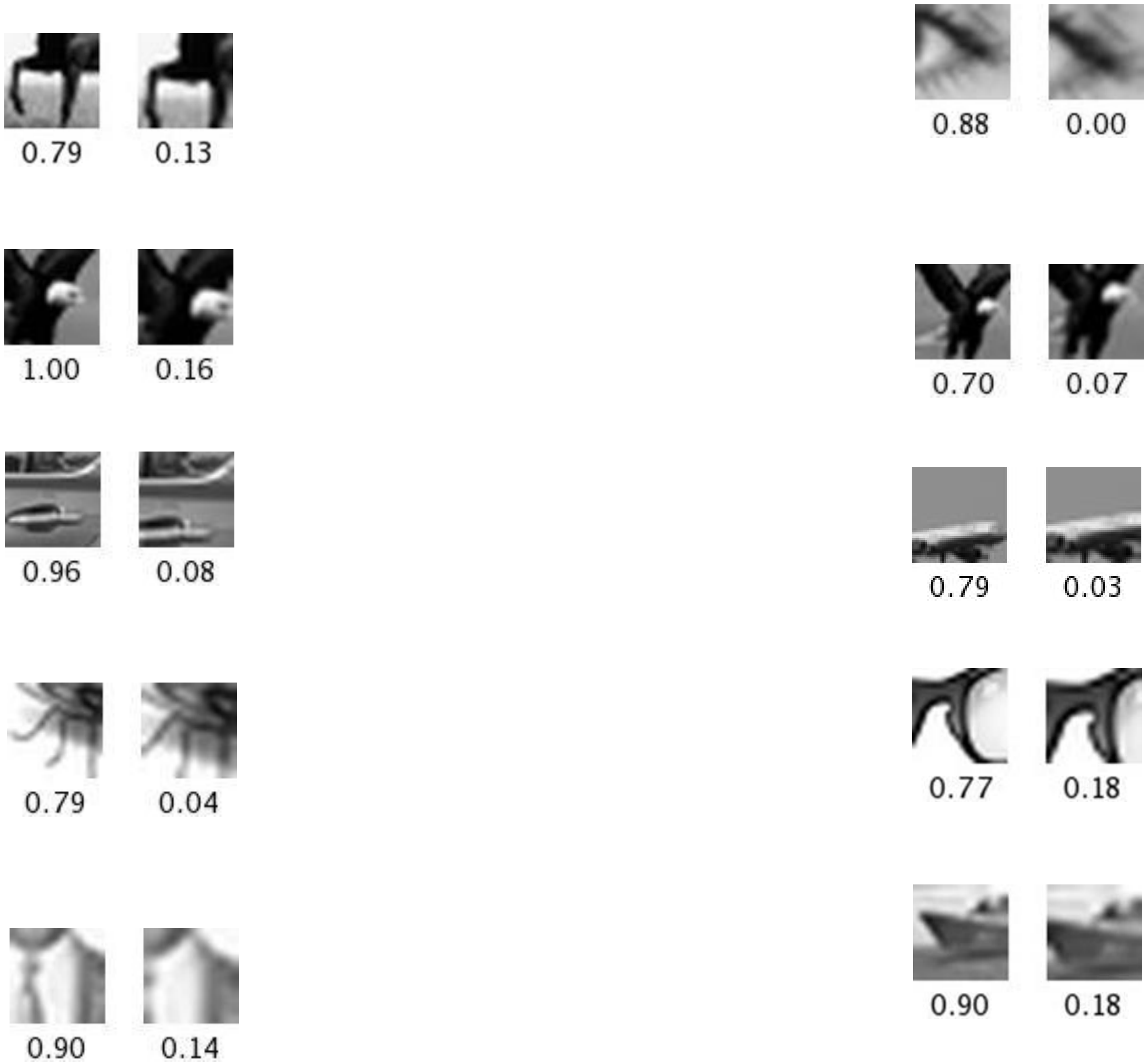
Eyeglasses

Suit&tie

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.



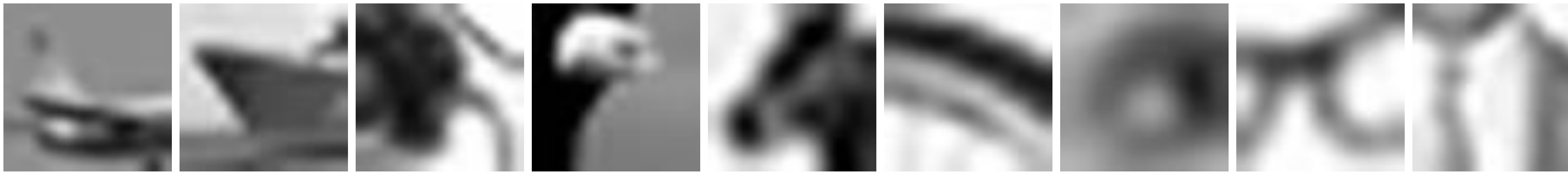
© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Visual Elements

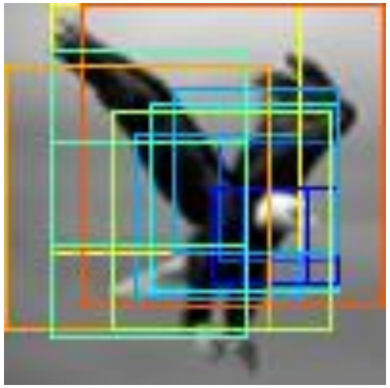
14 13 12 13 7 8 6 6 10



24 26 19 30 29 25 20 18 21

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Cover



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Average 16.9 / class
Highly redundant
Each MIRC is non-redundant

- Sensitive tool to compare representations

- Differences between MIRCs and sub-MIRCs to infer visual features

- Recognition features not captured by human feed-forward models and computer vision representations

Testing computational models

- Training of object images,
- Testing on minimal images
- MIRCs and sub-MIRCs









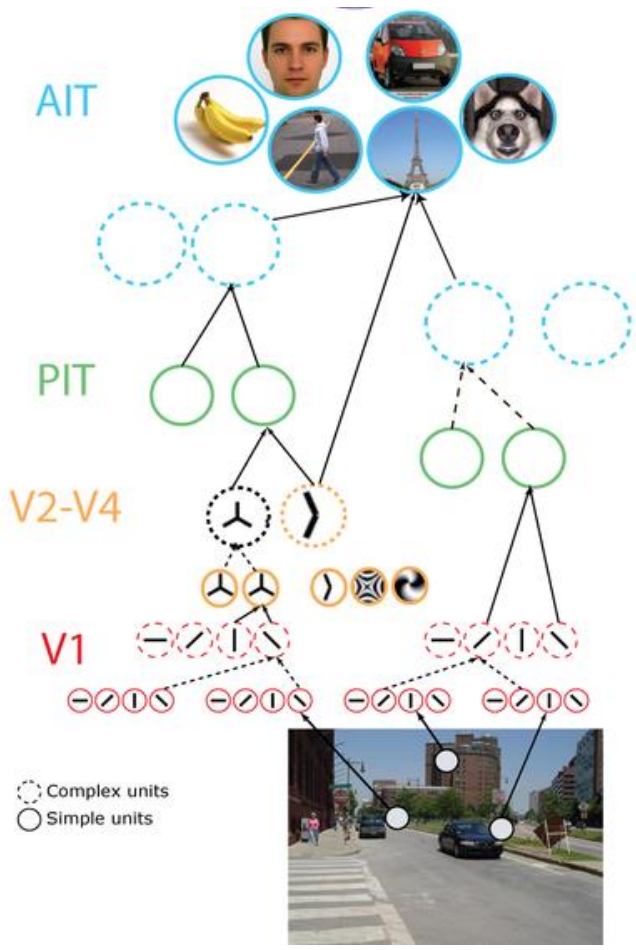



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Testing on:

- DPM Deformable Parts Models
- Bag-of-Words / VLAD (vector of locally aggregated descriptors)
- R-CNN (Deep Convolutional Neural Network)   Malik
- Hmax -- model of recognition in the cortex
-
- Consistent winners of standardized recognition competitions
- (PASCAL, ImageNet)

Hmax Model

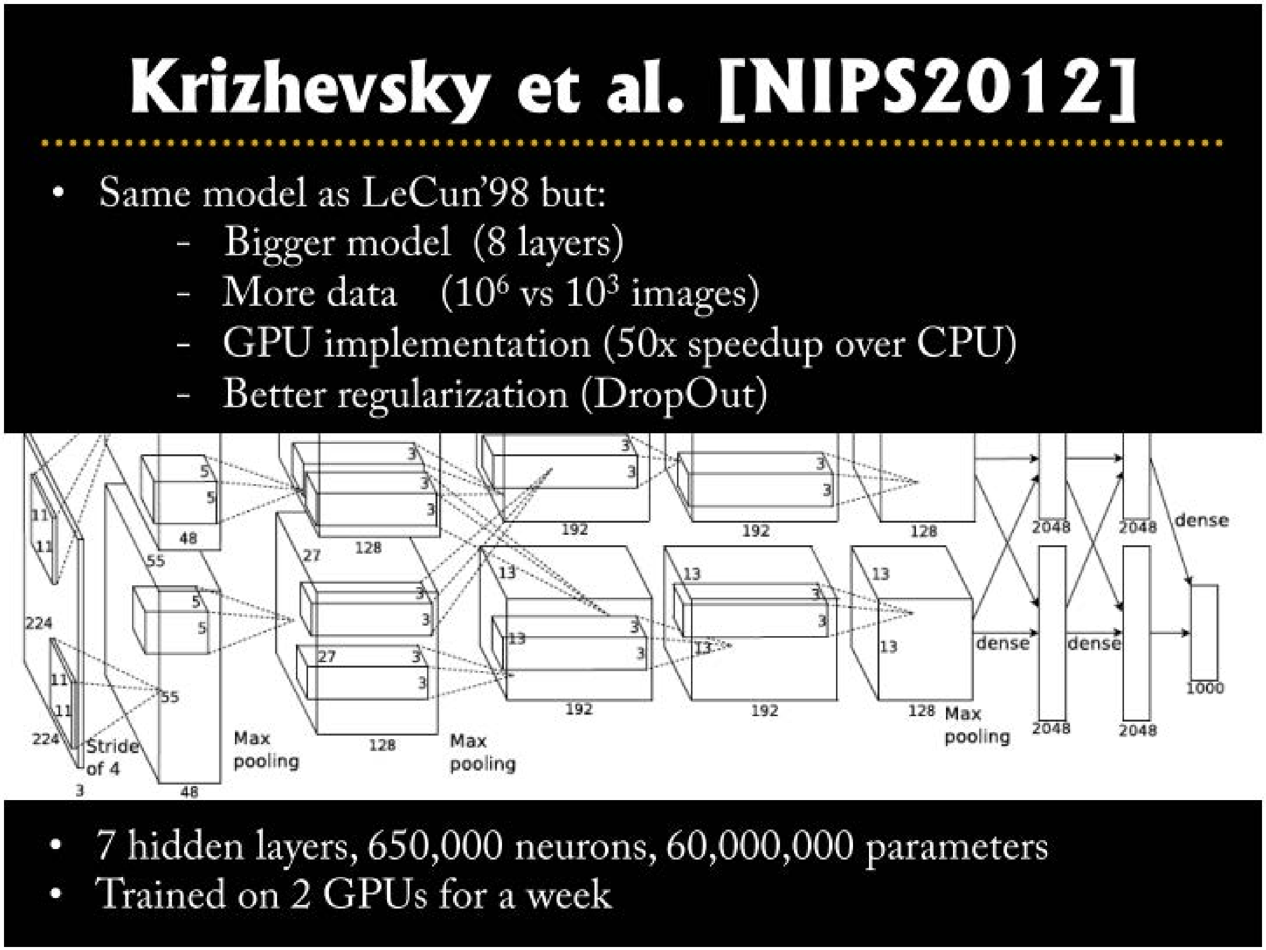


© Scholarpedia. License BY-NC-SA

Riesenhuber and Poggio, 1999

Serre et al 2007

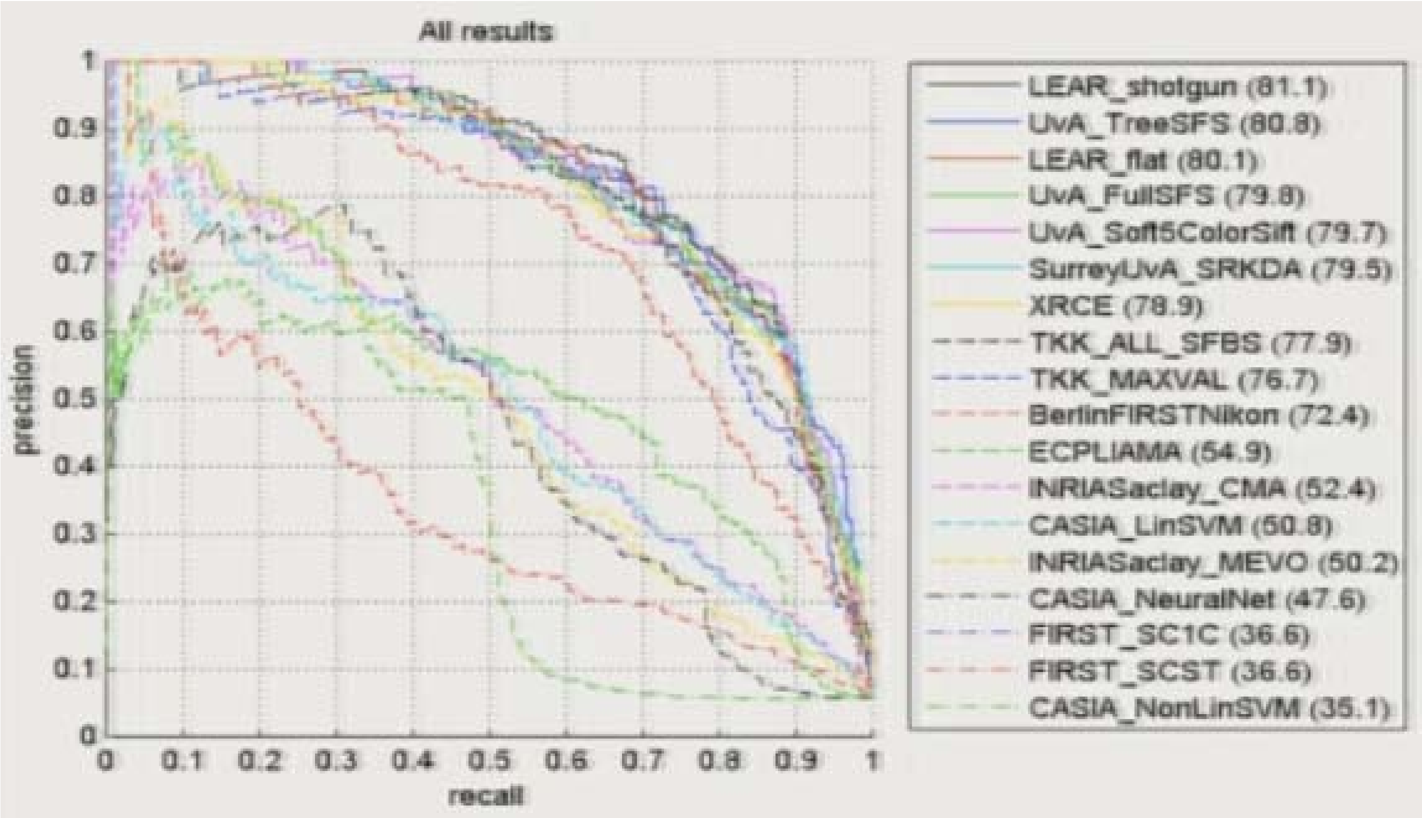
Deep Network Models

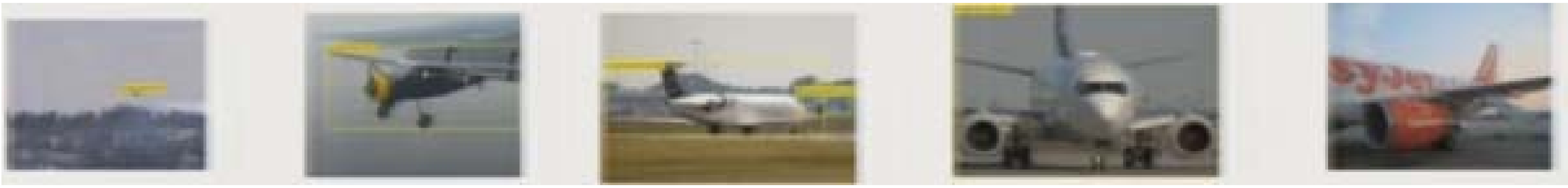


© Neural Information Processing Systems. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Source: Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105. 2012.

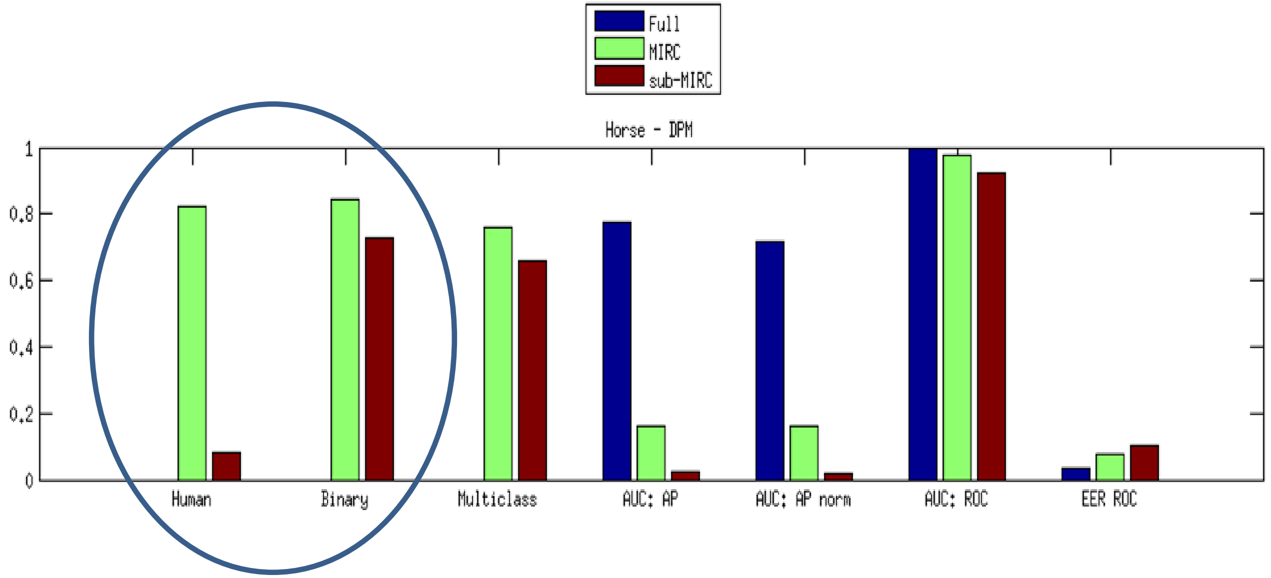
'Pascal Challenge' Airplanes



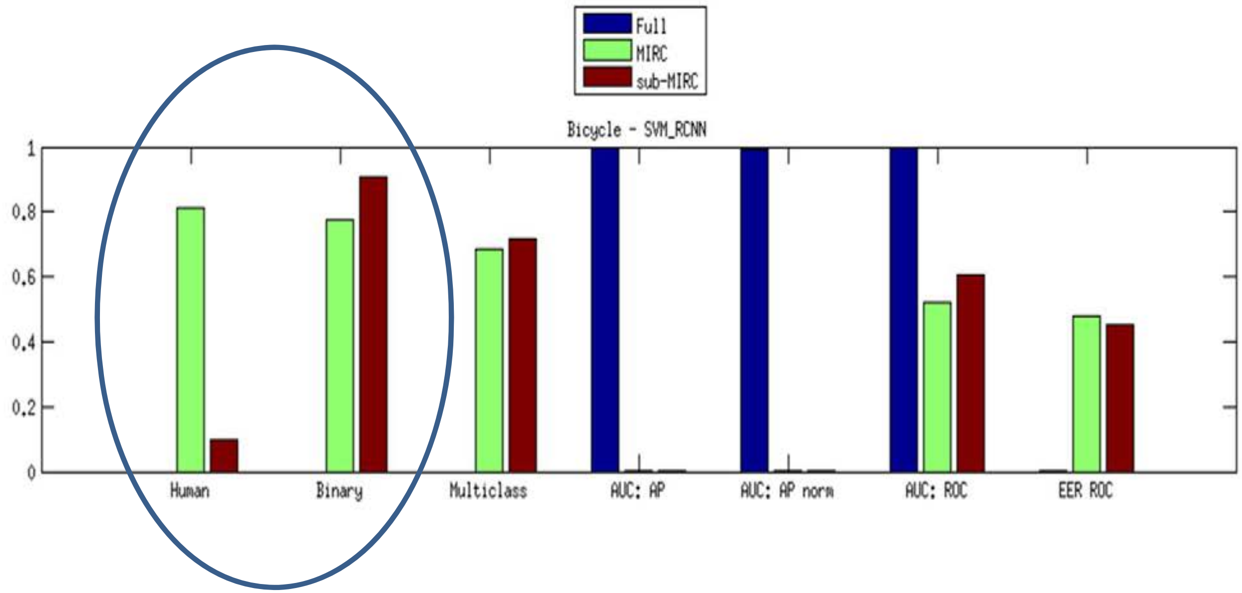


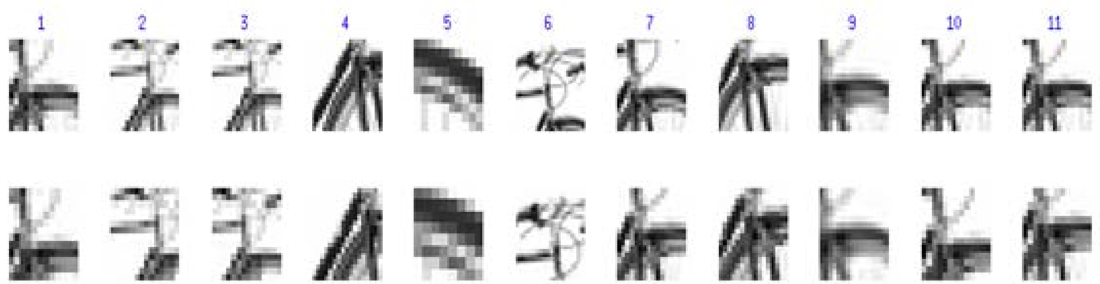
© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

The recognition gap is not reproduced



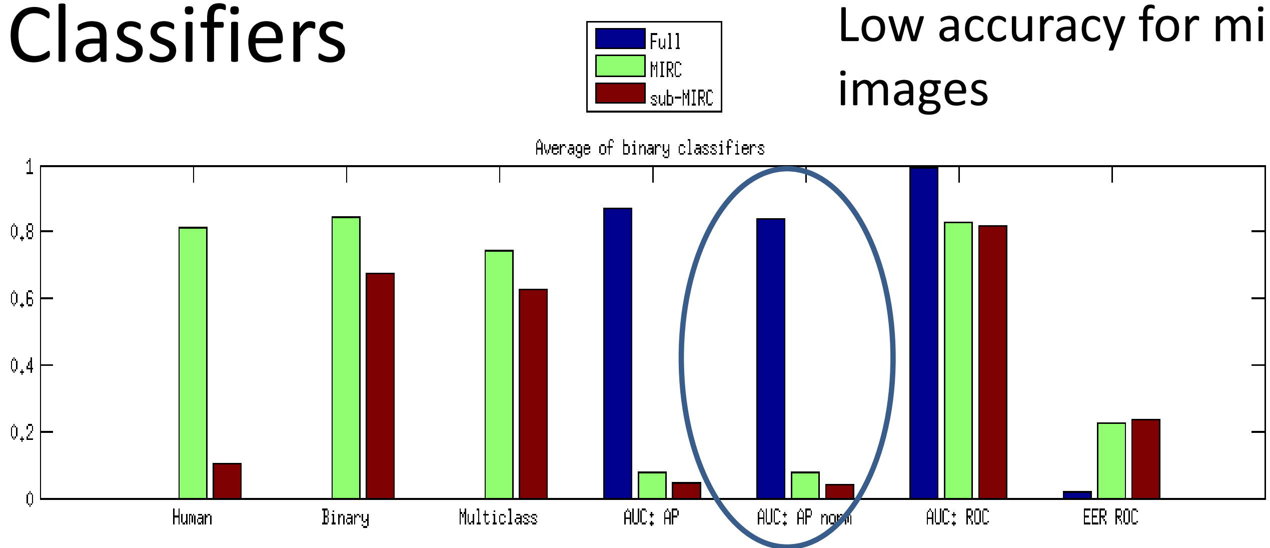
R-CNN ‘Deep-net’ Recognition Model





All Classifiers

Low accuracy for minimal images



Statistical significance

	p
H0: Binary gap equal to multi-class. H1: otherwise	0.2443
H0: Human gap equal to Binary. H1: Human gap greater than Binary	1.3290e-04
H0: Human gap equal to multi-class. H1: Human gap greater than multi-class	4.5180e-06
H0: Binary MIRC recall equal to sub-MIRC. H1: otherwise	0.1711
H0: Human MIRC recall equal to sub-MIRC. H1: Human MIRC greater than sub-MIRC	1.7502e-12
H0: AP full object equal to MIRC. H1: AP full object greather than MIRC	5.3537e-05
H0: AP full object equal to sub-MIRC. H1: AP full object greather than sub-MIRC	1.2785e-06

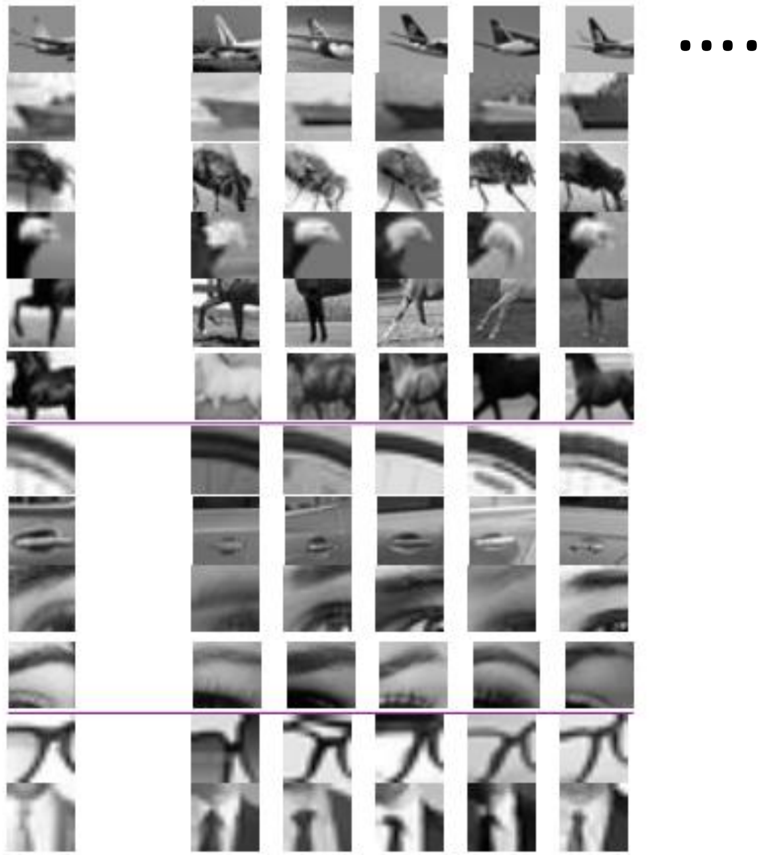
Data

	Full	MIRC	sub-MIRC
Human recall	NaN	0.81	0.10
Binary recall	NaN	0.84	0.67
Multiclass recall	NaN	0.74	0.63
AUC: AP	0.87	0.07	0.05
AUC: AP (normalized)	0.84	0.07	0.04
AUC: ROC	0.99	0.83	0.82
EER ROC	0.02	0.23	0.23
Gap (MIRC recall 50%-90%) - Binary: max , mean	NaN	0.38	0.27
Gap (MIRC recall 50%-90%) - Multi: max , mean	NaN	0.27	0.20
Num of negatives patches	245970.93	414795.17	414795.17
# normalized negative images	2259.50	1428.78	1428.78
Num of positives	28.20	10.00	15.20
Gap s.d. (human, Binary, multi-class)	0.05	0.27	0.18

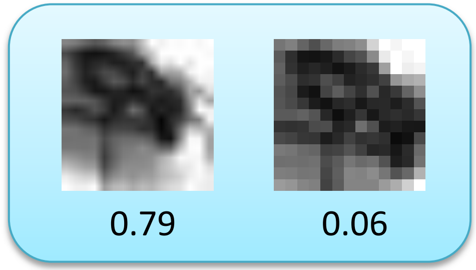
- Recognition of minimal images does not emerge by training any of the existing models tested.

- Representations used by existing models do not capture differences that human recognition is sensitive to

Test 2: Train on Patches

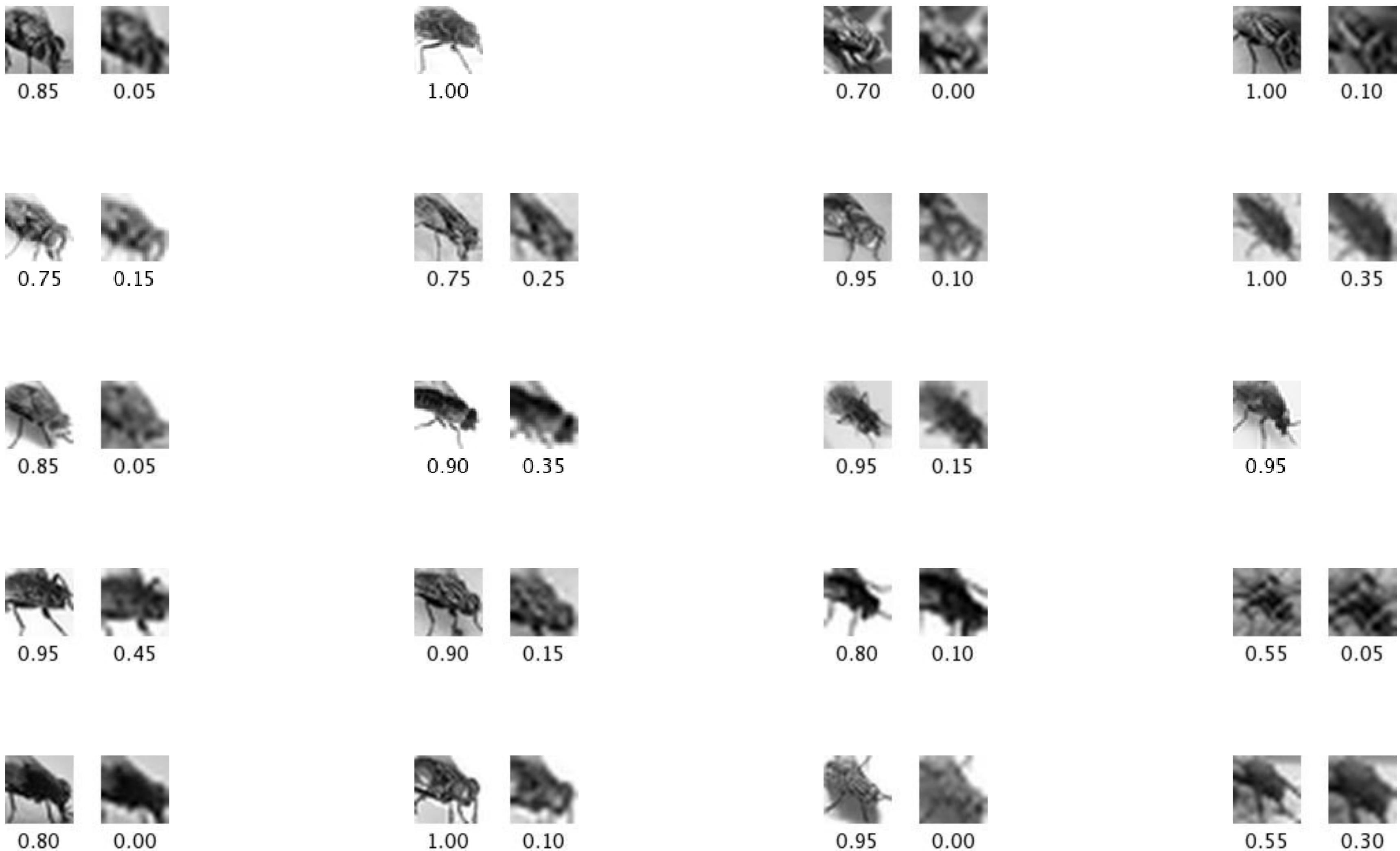


© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

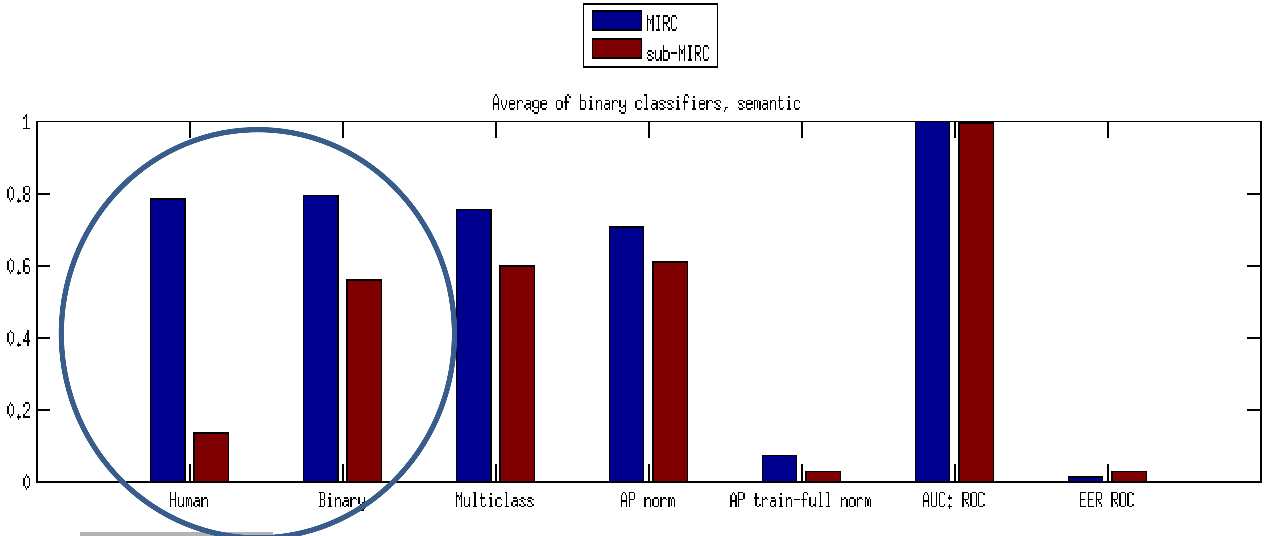


30 ‘siblings’ for each MIRC and sub-MIRC

Example siblings



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.



Statistical significance

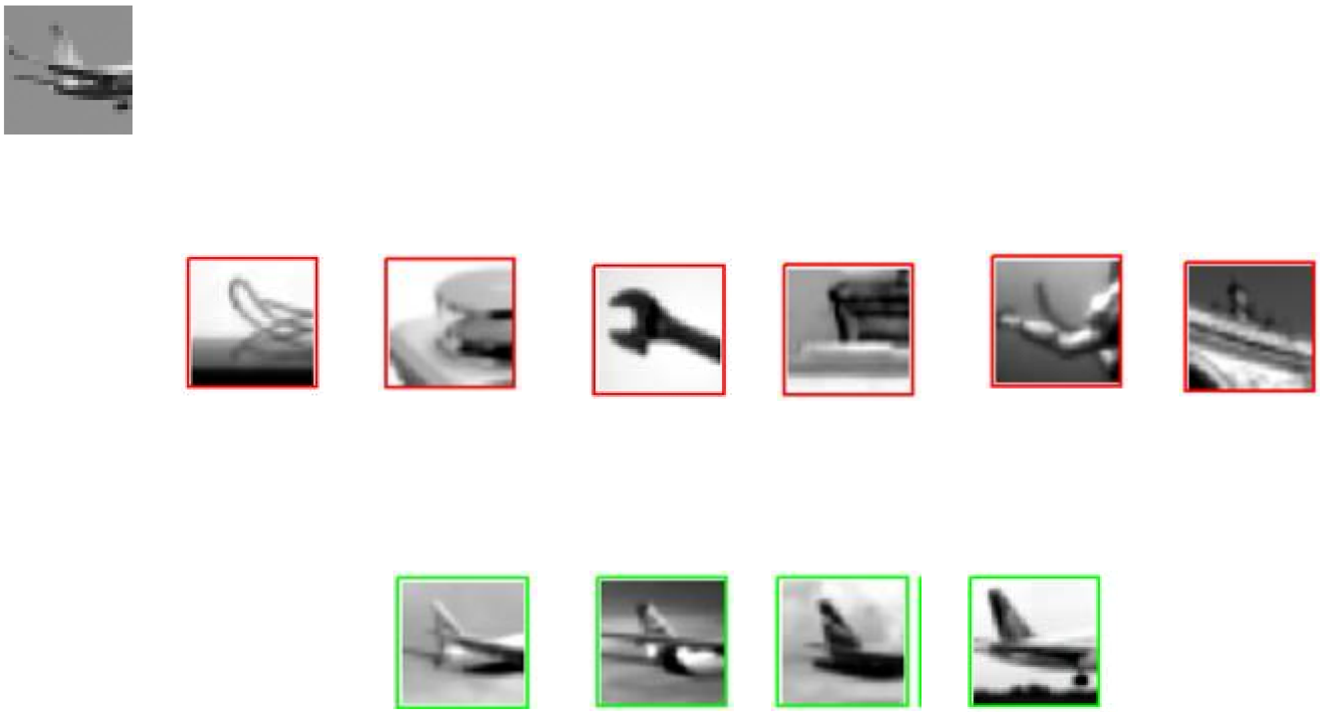
	p
H0: Binary gap equal to multi-class. H1: otherwise	0.0284
H0: Human gap equal to Binary. H1: Human gap greater than Binary	5.2497e-05
H0: Human gap equal to multi-class. H1: Human gap greater than multi-class	1.9858e-06
H0: Binary MIRC recall equal to sub-MIRC. H1: otherwise	0.0016
H0: Human MIRC recall equal to sub-MIRC. H1: Human MIRC greater than sub-MIRC	6.9954e-10

Data

	MIRC	sub-MIRC
Human recall	0.78	0.13
Binary recall	0.79	0.56
Multiclass recall	0.75	0.60
AUC: AP (normalized)	0.70	0.61
AUC: train-full AP (normalized)	0.07	0.02
AUC: ROC	1.00	0.99
EER ROC	0.01	0.02
Gap (MIRC recall 50%-90%) - Binary: max , mean	0.32	0.23
Gap (MIRC recall 50%-90%) - Multi: max , mean	0.22	0.17
# normalized (positive,negatives) patches	10.00	500000.00
Num of positives	21.90	21.90
Gap s.d. (human, human)	0.08	0.08
Gap s.d. (Binary, multi-class)	0.16	0.10

- No sharp gap between MIRCs and sub-MITCs, humans’ is much larger
- Limited recognition
  - (60% accuracy at 75% recall)
- Humans are much better
- Less false detections, different false detections

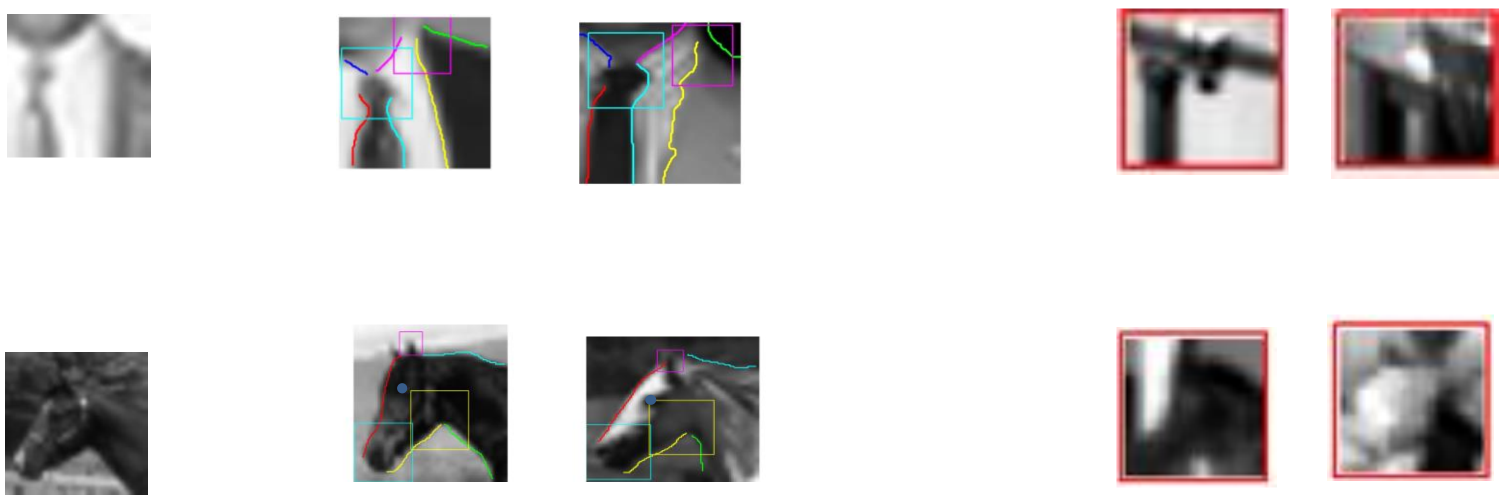
Example false detections (DPM)



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Internal Interpretation

also for Validation



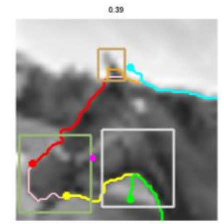
© Wiley. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Ben-Yosef, Guy, Liav Assif, Daniel Harari, and Shimon Ullman. "A model for full local image interpretation." In CogSci. 2015.

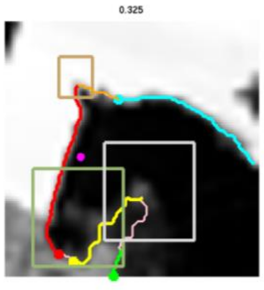
internal interpretations, produced automatically by a model
Cannot be produced by existing feed-forward models

Interpretation Features

- Shape of an extended contour   (grouping)
- ‘Visual words’, texture inside a region (local segmentation)
- Small features at a specific location

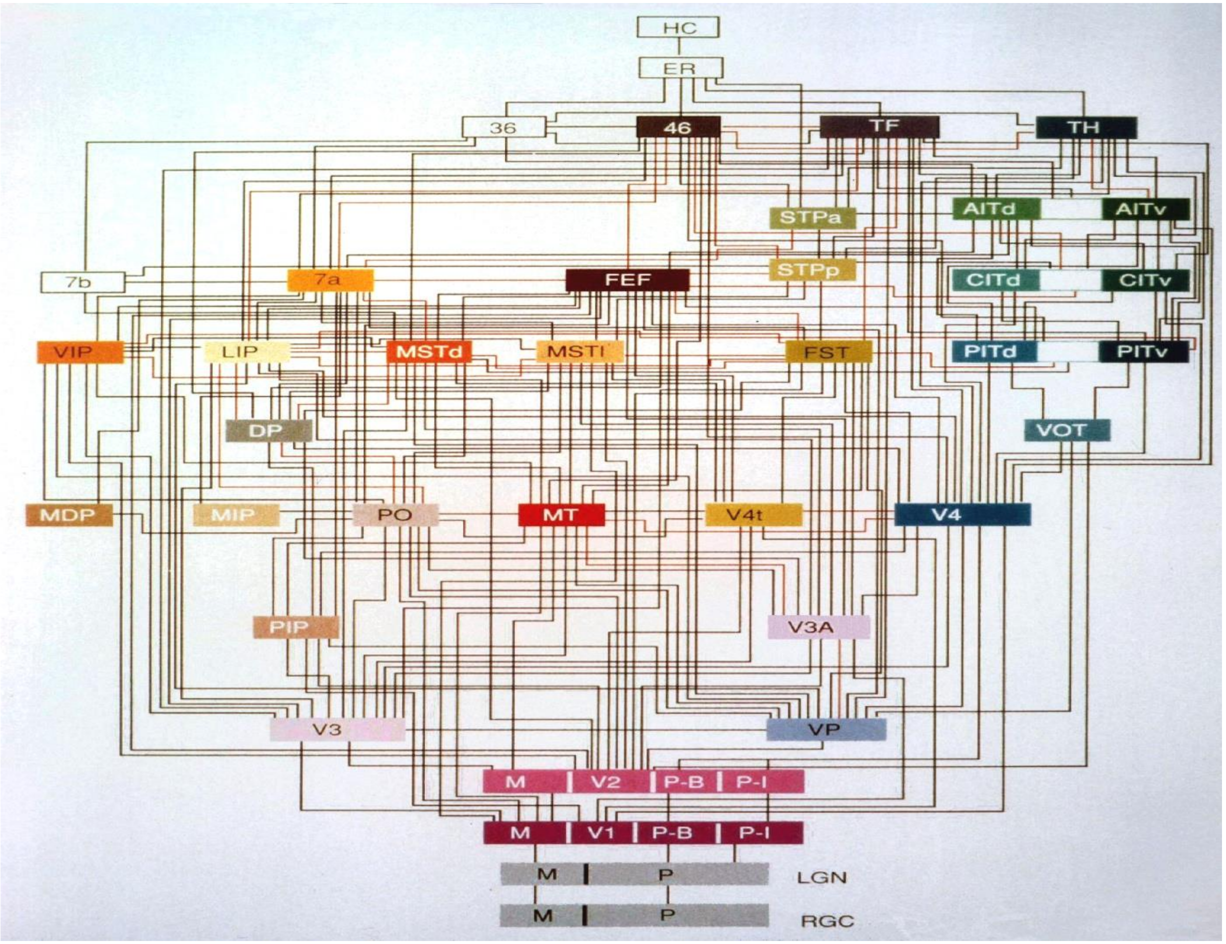
Likely Top-Down: complex and class-specific





© Wiley. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Ben-Yosef, Guy, Liav Assif, Daniel Harari, and Shimon Ullman. "A model for full local image interpretation." In CogSci. 2015.

The Visual Hierarchy



© Oxford University Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Felleman, Daniel J., and David C. Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." Cerebral cortex 1, no. 1 (1991): 1-47.

Two stages in recognition

Class-specific validation
and interpretation
Internal and external

Entire image

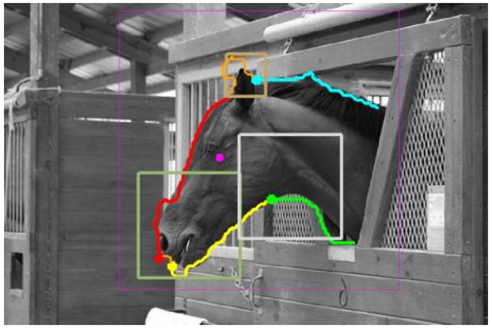
MIRCs level

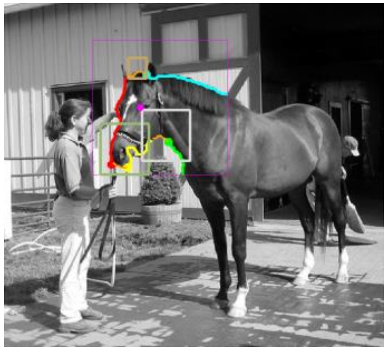
Generic
classification

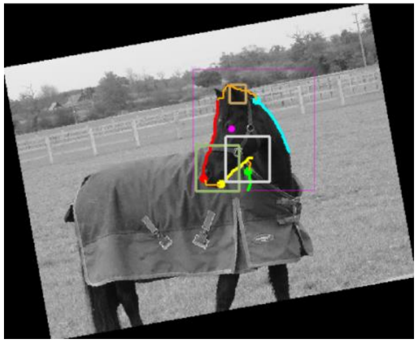
Image

© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

BU – TD Interpretation
Examples from a Model







MEG studies of minimal images

- On the role of top-down processing

- Leyla Isik
- Yena Han

MEG: Decoding Image Identity and Category (MIRCs vs. hard-negatives

Minimal images as a sensitive tool

- Eagle:

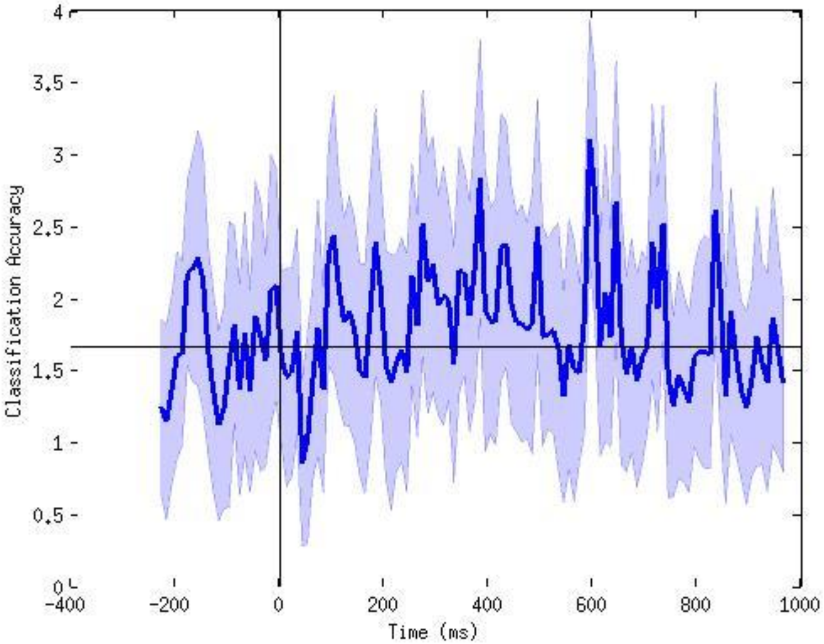
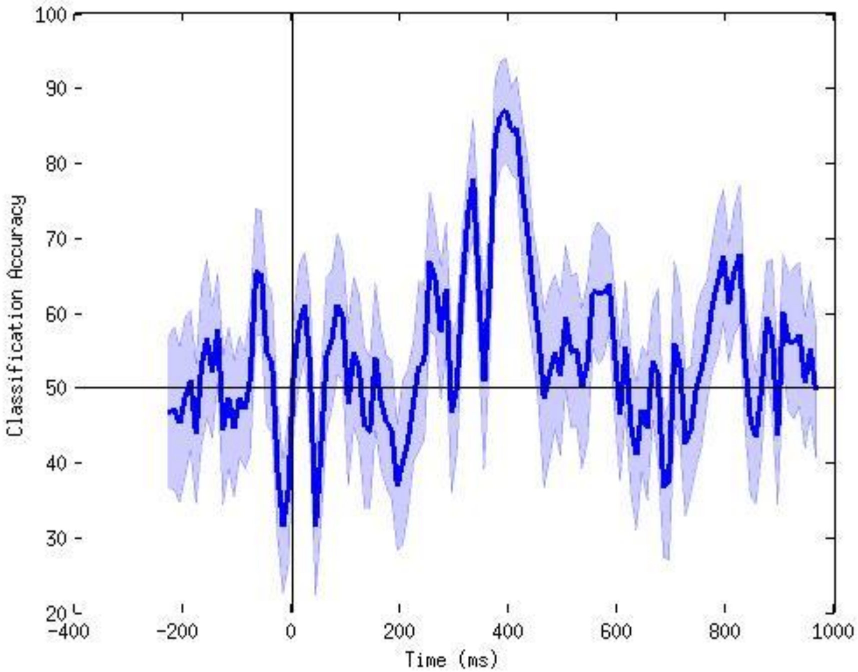


Image Identity :
60 (30 class + 30 non-class) ways classification



Class/non-class classification

Decoding Image Identity and Category (class/non-class)

- Bike

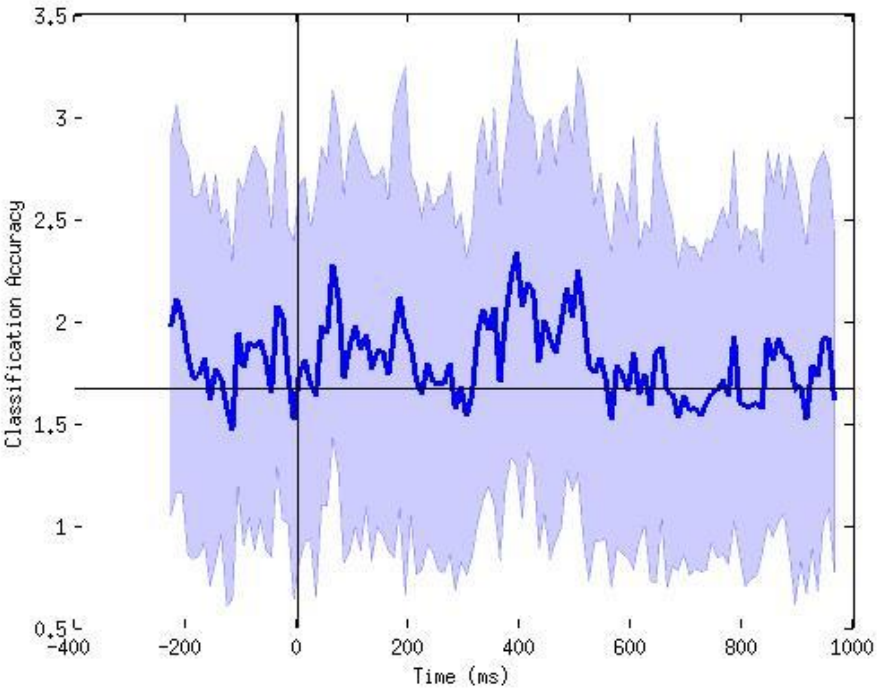
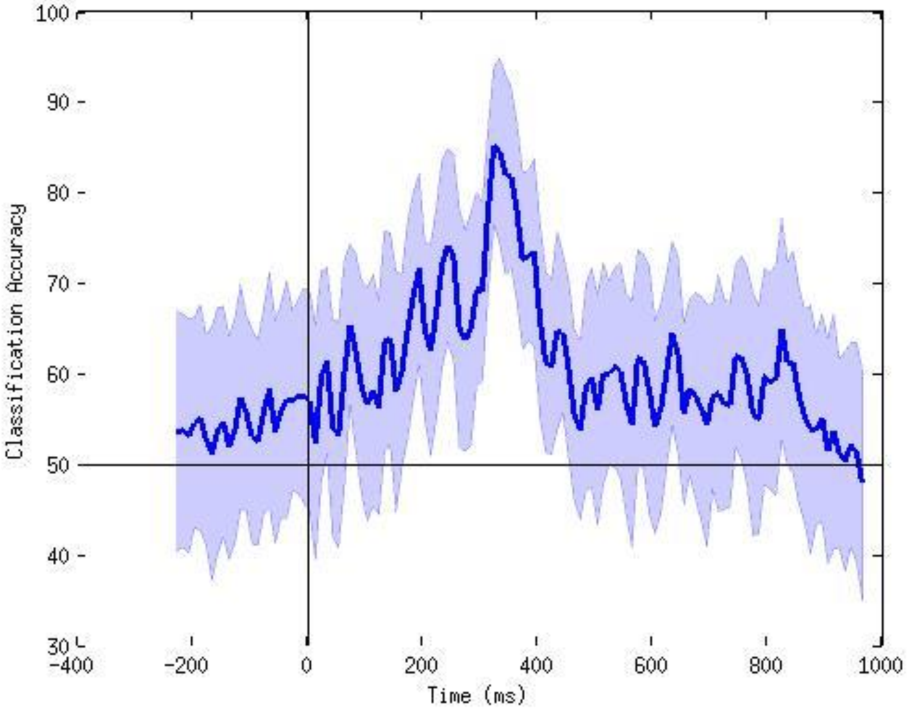


Image Identity :
60 (30 class + 30 non-class) ways classification



Class/non-class classification

General summary comments on bottom-up and top-down

- Initial feed-forward sweep, DNN type can be very useful

- Triggering computations that depend on complex and class-specific properties and relations, top-down routines

- More related to recurrent, recursive computation, working memory, RNN, LSTM, sequential processes,

- Innate structures are more complex, and contain more information about the world

Summary

- Studies of recognition have focused on the first stage only
- Model: minimal images followed by interpretation and validation
- Features and representations used
- Neural circuits involved, top-down processing
- Full image interpretation
- Actions,   agents interactions

MIT OpenCourseWare
https://ocw.mit.edu

Resource: Brains, Minds and Machines Summer Course
Tomaso Poggio and Gabriel Kreiman

The following may not correspond to a particular course on MIT OpenCourseWare, but has been provided by the author as an individual learning resource.

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.