Human and computer face detection under occlusion

Samuel E. Anthony[1], Walter Scheirer[2], David Cox[2] & Ken Nakayama[1]

[1]Department of Psychology, [2]Department of Molecular and Cell Biology

Harvard University

VISION SCIENCES LABORATORY
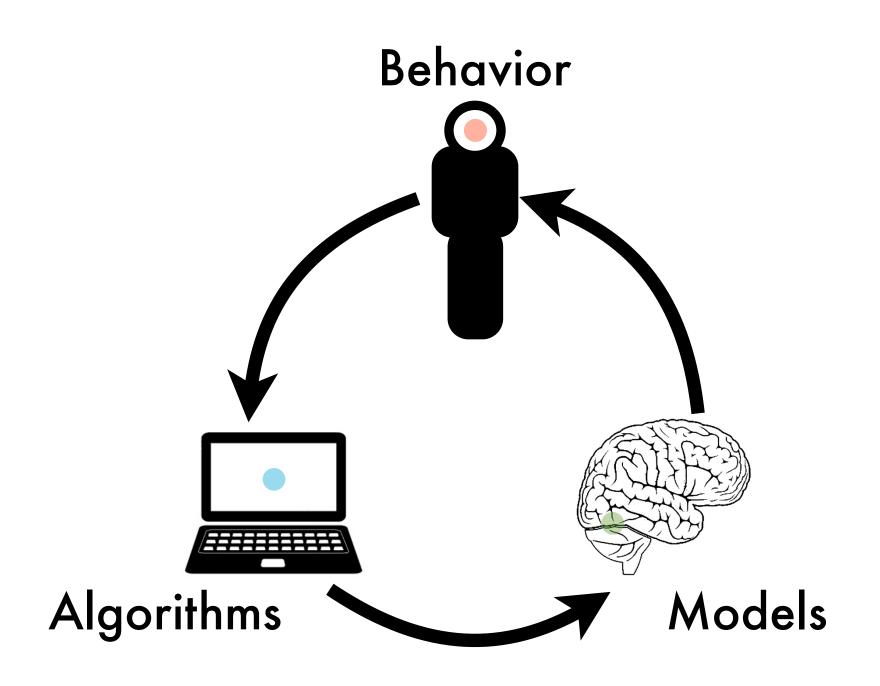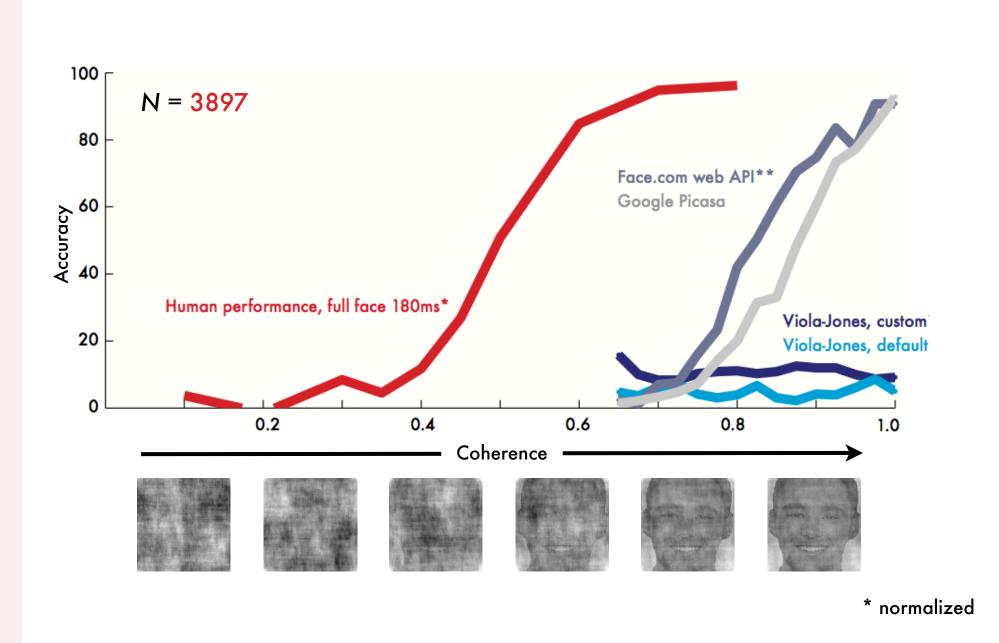
Motivation

Previous Work

* normalized

Questions

Phase-scrambled noise is not a plausible real-world degradation of face images. Would the gap persist with more ecologically valid degradations?
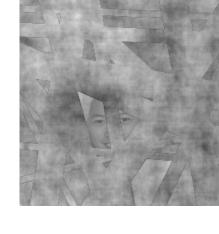
If the gap between humans and computers persists, is it possible to learn anything about why the humans are superior?
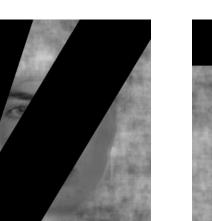
Behavioral Testing

Images were generated with parametric occluders. Occlusion is well known as a difficult challenge for face detection algorithms.

Failed detections under occlusion in Google Street View

Five types of generated occluded stimuli

Complex Noise Simple Perpendicular Simoncelli

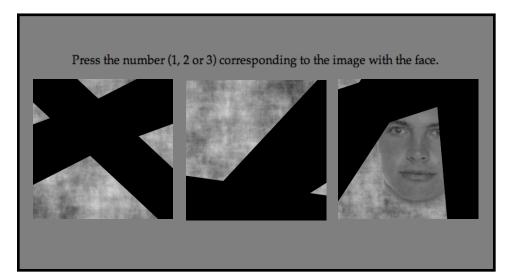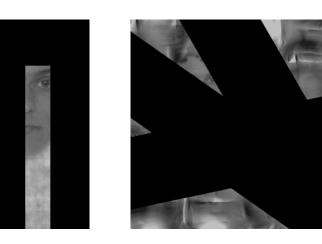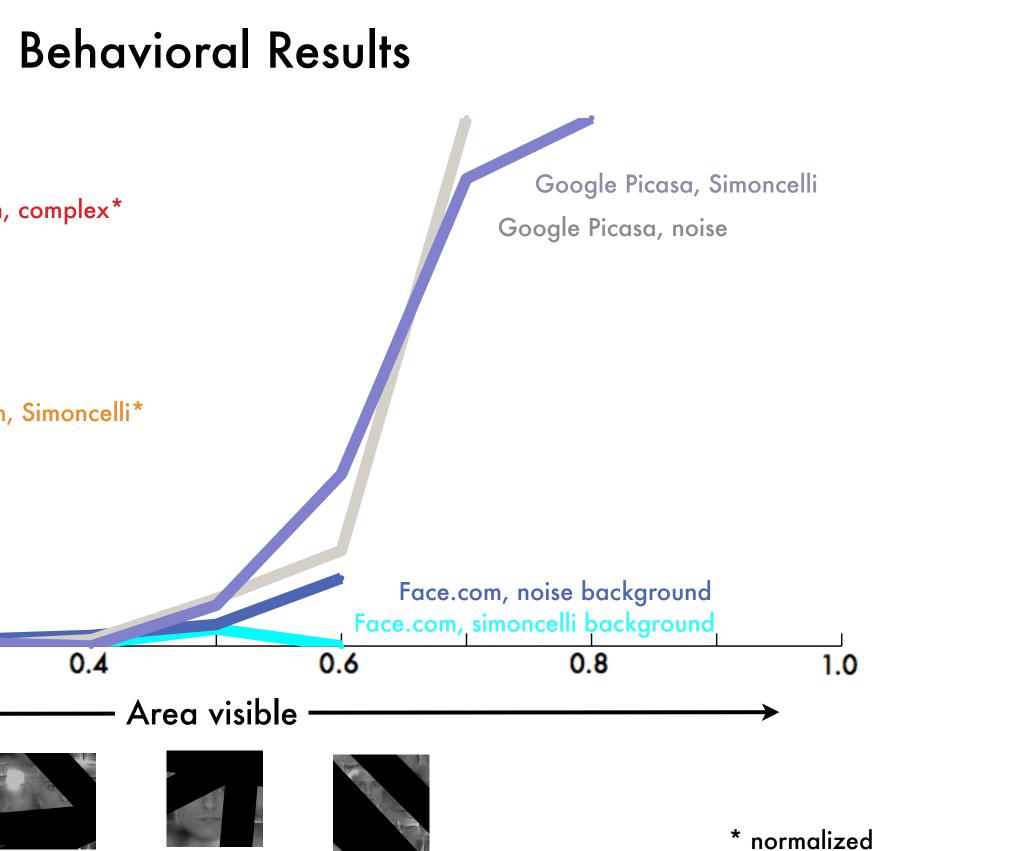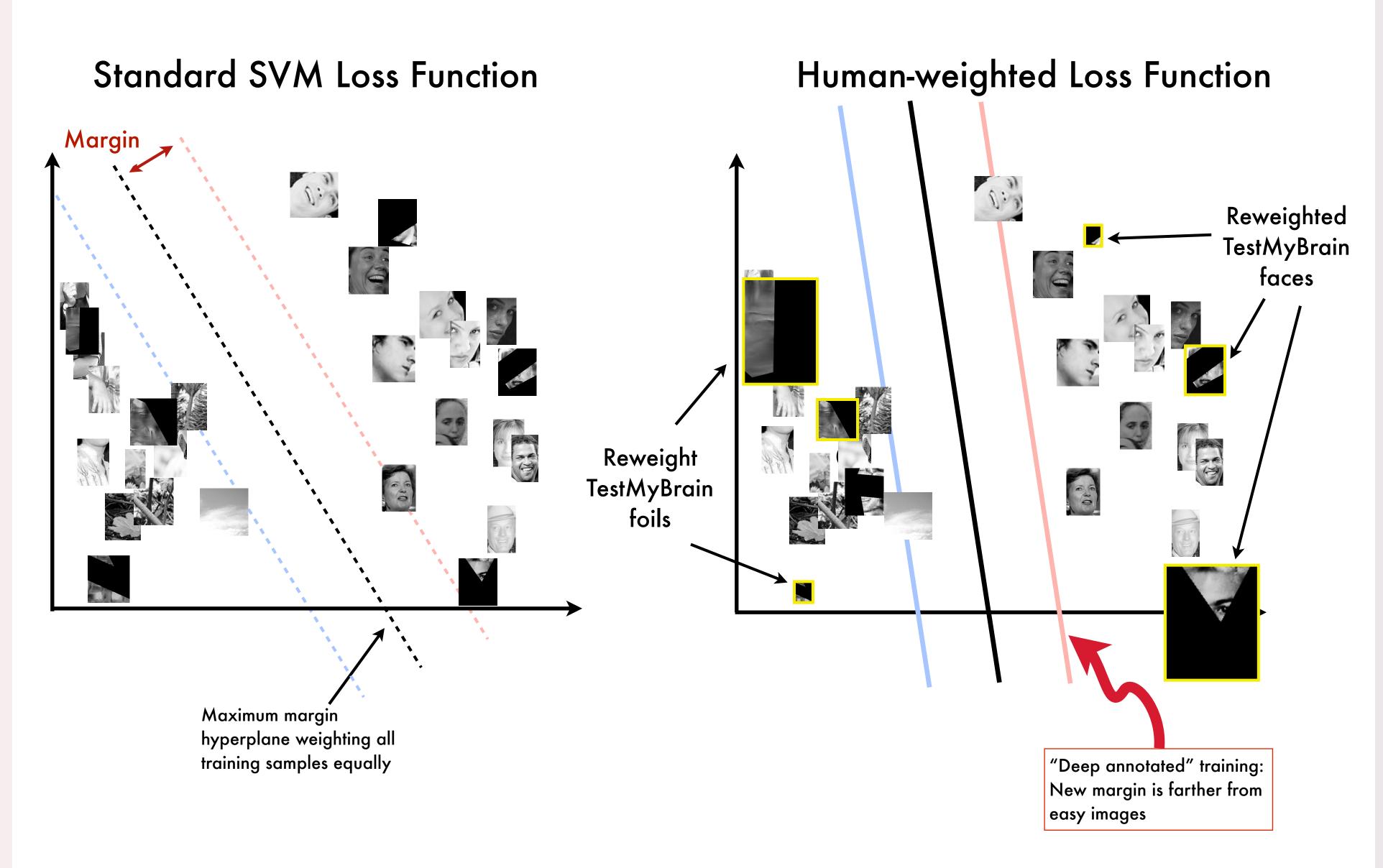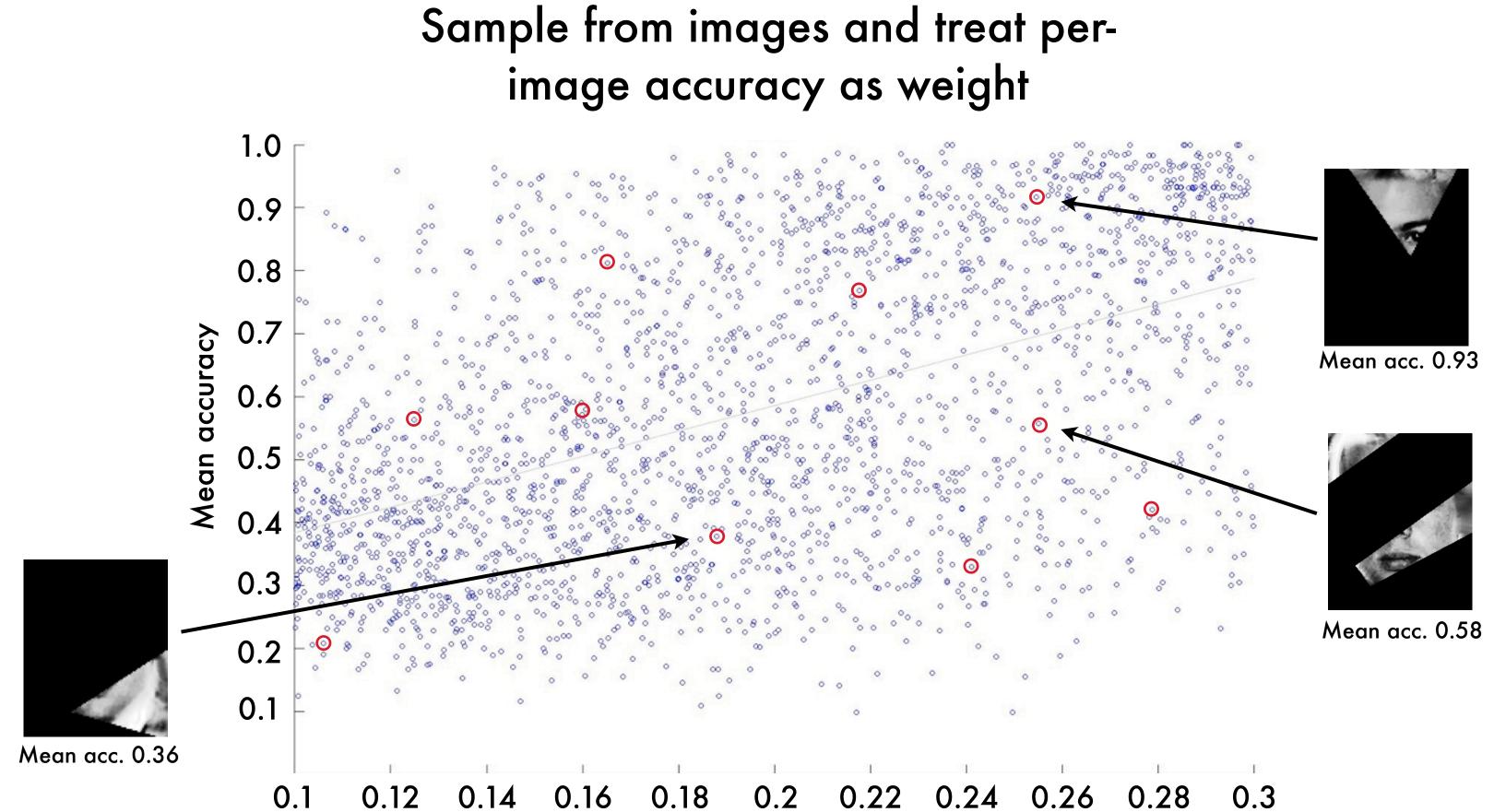All stimuli have noise backgrounds; Simoncelli condition has 2nd order statistic-matched Portilla-Simoncelli textures.
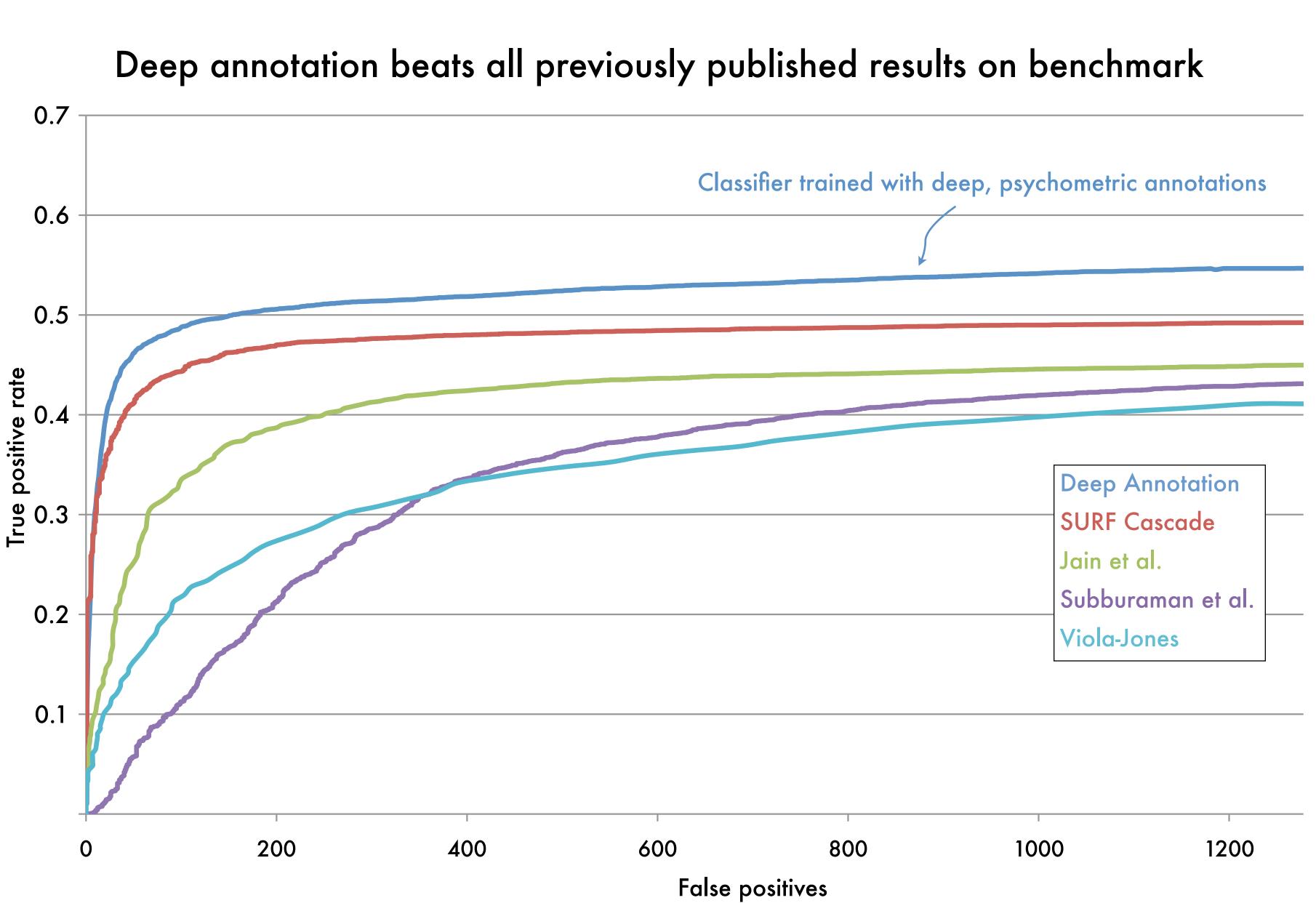
Task

3 Alternative Forced-Choice

Subjects on the TestMyBrain website completed 200 trials of 3-AFC spot-the-face. All conditions besides Simoncelli were behaviorally indistinguishable

TESTMYBRAIN
Test yourself at www.testmybrain.org!

Behavioral Results

* normalized

Deep Annotation

Standard SVM Loss Function

Margin

Maximum margin hyperplane weighting all training samples equally

Human-weighted Loss Function

Reweighted TestMyBrain faces

Reweight TestMyBrain foils

"Deep annotated" training: New margin is farther from easy images

Sample from images and treat per-image accuracy as weight

Mean acc. 0.93

Mean acc. 0.58

Mean acc. 0.36

Mean acc. 0.36

Accuracy is only one choice for psychometric measure; we have had success with RT, and there are many other candidates
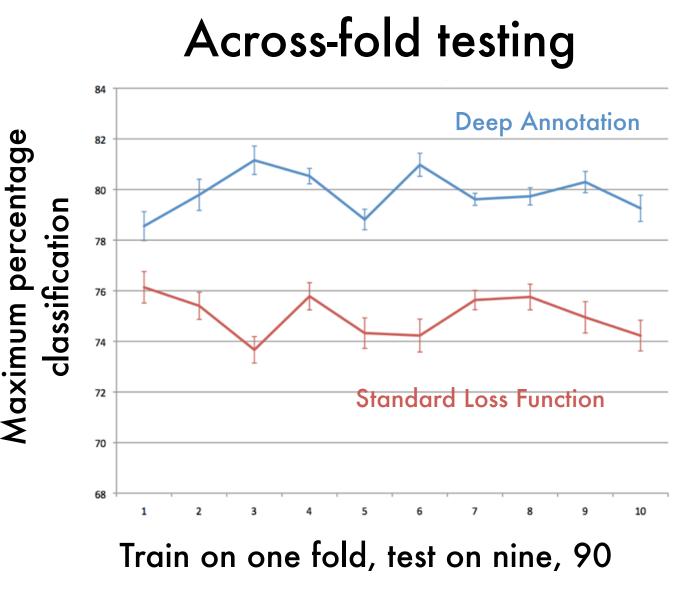
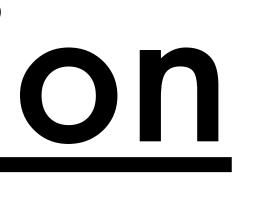Algorithm includes a preprocessing stage where ultra-low-threshold viola-jones delivers candidate regions to speed up analysis

Results

Deep annotation beats all previously published results on benchmark

Classifier trained with deep, psychometric annotations

Deep Annotation
SURF Cascade
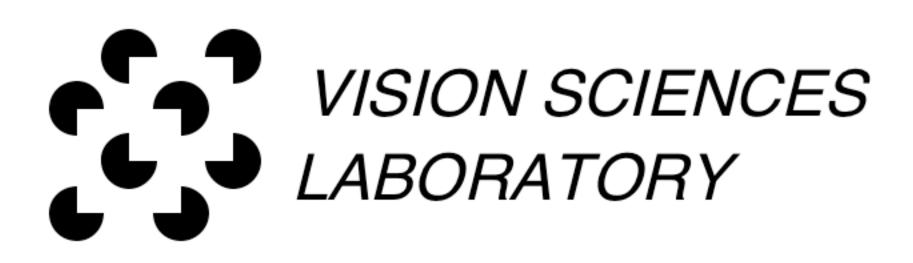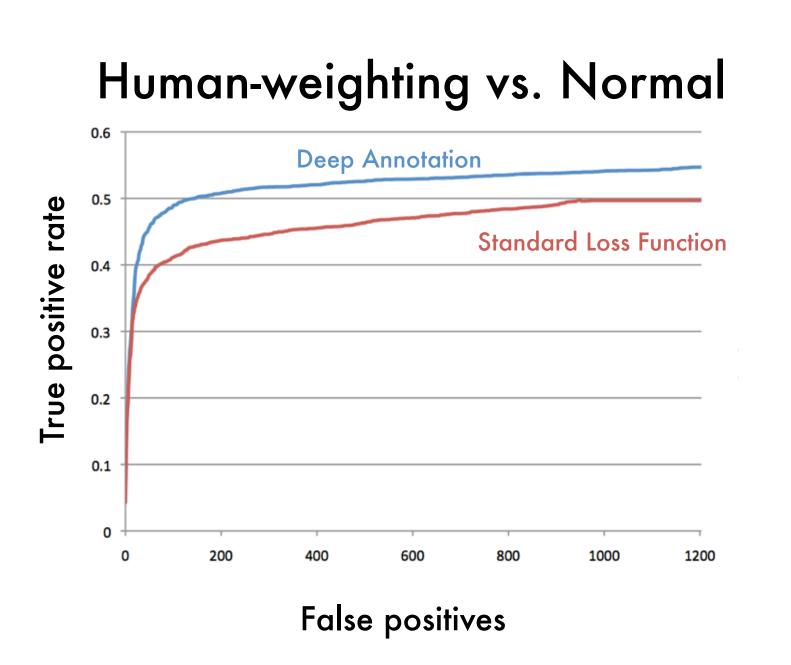Jain et al.
Subburaman et al.
Viola-Jones

Models were tested against FDDB benchmark

• 2845 images with 5171 hand-annotated faces • 10 "folds" trained and tested independently • Standard scoring function for model comparison • Challenging faces including many with occlusion
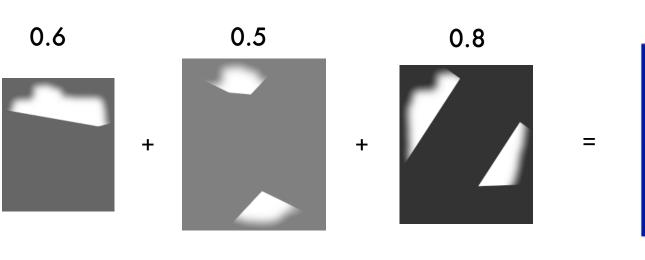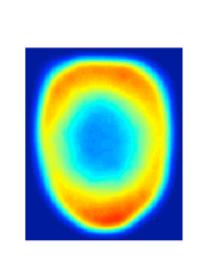
Across-fold testing

Deep Annotation

Standard Loss Function

Train on one fold, test on nine, 90

Human-weighting vs. Normal

Deep Annotation

Standard Loss Function

Closing the Loop

Generate heatmaps from weighted sum of face image masks

0.6 0.5 0.8

Human
Compared to noise, Simoncelli heatmap is less intense at face edges; subjects are worse at images without central face visible

Simoncelli Human-weighted

Noise Standard
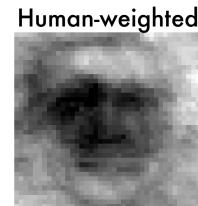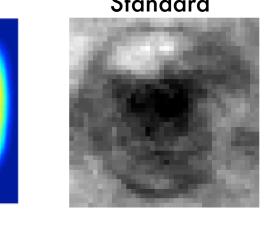
Computer
Compared to normally trained SVM, mean image of best detections for human-weighted loss has more internal features

Conclusions

Deep annotation can dramatically improve algorithm performance.

The use of psychometric data in classifier training can provide improved generalization and state-of-the-art performance.

Future Directions

Apply deep annotation to everything

Quantify model "humanness"

References

1. Heisele, B. et al. (2007) *IJCV* 2. Portilla J. & Simoncelli E.P. (2000) *IJCV* 3. Jones, MJ & Viola, P (2001) *Work. on Stat. and Comp. Th. of Vis.* 4. Jain, V & Learned-Miller, E (2010) *Tech. Rep. UM-CS-2010-009*