Visual Psychophysics for Making Face Recognition Algorithms More Explainable

Brandon RichardWebster[1], So Yon Kwon[2], Christopher Clarizio[1], Samuel E. Anthony[2,3], Walter J. Scheirer[1]

[1]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

[2]Perceptive Automata, Inc.

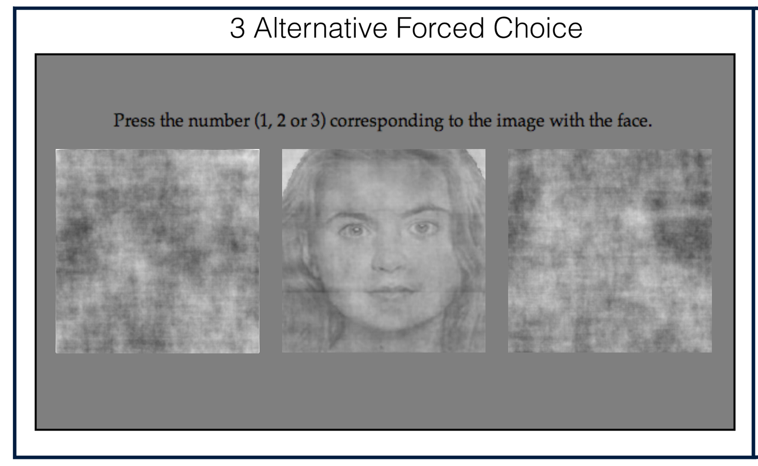
[3]Department of Psychology, Harvard University, Cambridge, MA 02138, USA

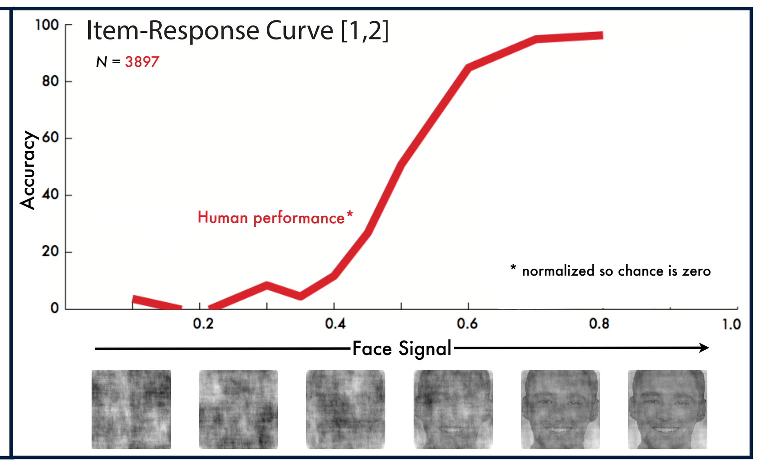
Motivation

In computer vision, procedures for understanding how a target model perceives a face have largely been in the form of dataset evaluation for recognition tasks where summary statistics are used to measure progress. While aggregate performance has continued to improve, understanding individual causes of failure has been difficult, as it is not always clear why a particular face fails to be recognized, or why an impostor is recognized by an algorithm. Importantly, other fields studying vision have addressed this via the use of visual psychophysics: the controlled manipulation of stimuli and careful study of the responses they evoke in a subject. We suggest that visual psychophysics is a viable methodology for making face recognition algorithms more explainable.

Visual Psychophysics

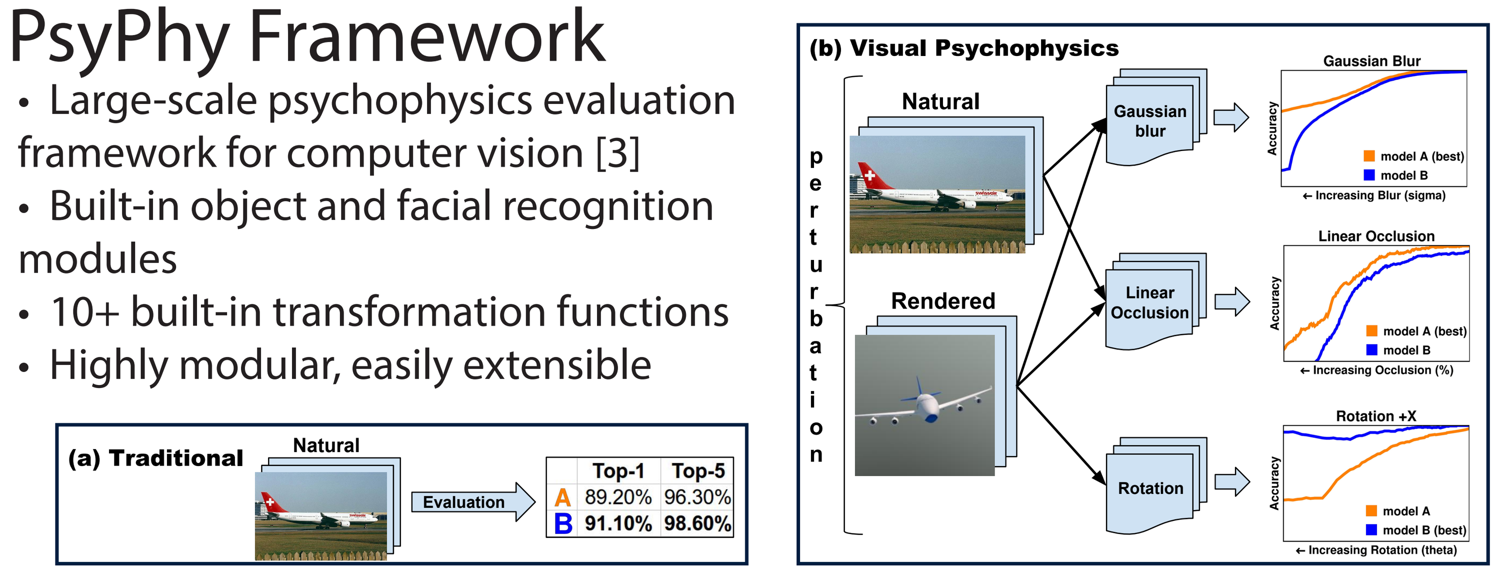
- A question is asked or a task presented to a subject
- Subject gives response
- Response, response time, and other characteristics are recorded
- Subject repeats task at different stimulus levels
- Multiple subjects responses are aggregated together
- Represents the relationship between the population's response and stimuli
- Precise point of failure in visual perception can be identified





PsyPhy Framework

- Large-scale psychophysics evaluation framework for computer vision [3]
- Built-in object and facial recognition modules
- 10+ built-in transformation functions
- Highly modular, easily extensible

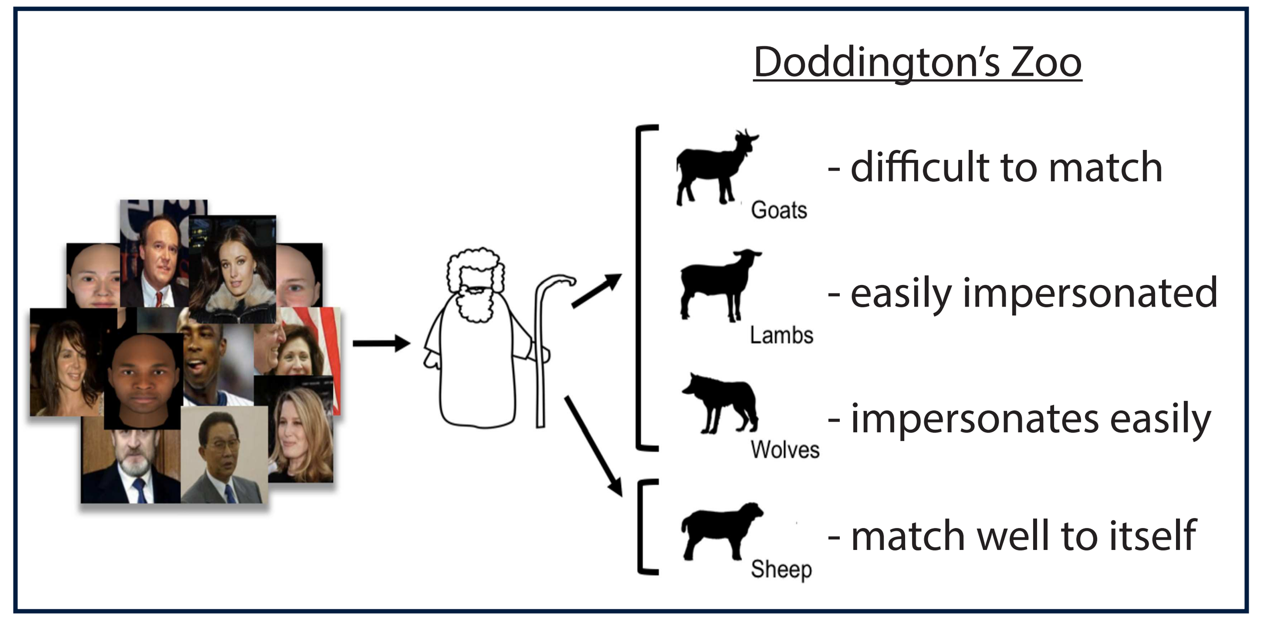


https://github.com/0xBrandon/psyphy-public

Procedures

1. "Herding"

In the process of herding, a shepherd function determines from a set of identities which are 2nd order sheep within the context of Doddington's Zoo [4], ensuring that the transformation was the reason for a decrease in performance.



The procedure to remove goats, lambs, and wolves:

$H(\Upsilon, I)$: a "herding" function to isolate Doddington et al.'s sheep from the goats, lambs, and wolves

Input: Υ, a "shepherd" function for a face recognition algorithm
Input: I, a set of input identities from a dataset
1: $S \leftarrow \Upsilon(I, I)$
2: $S \leftarrow \frac{(S+S^\intercal)}{2}$
3: $t_h \leftarrow$ optimize loss function λ with TPE
4: $I_h \leftarrow \lambda(S, t_h)$
Output: t_h, the optimal threshold to produce I_h
Output: I_h, the "sheep" identities isolated by the optimal threshold t_h

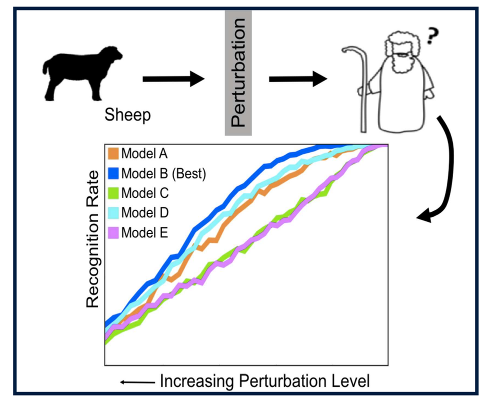
2. "Shepherding"

A shepherd function is wrapper function to the face recognition algorithm to be evaluated. A definition of a shepherd function:

$\Upsilon_f(I_p, I_g)$: a "shepherd" function that produces a similarity matrix for the face recognition function f

Input: f, a face recognition function that produces a feature representation
Input: I_p, a set of probe identities
Input: I_g, a set of gallery identities
1: $R_p \leftarrow i \in I_p : f(i)$
2: $R_g \leftarrow i \in I_g : f(i)$
3: $S \leftarrow r_p \in R_p, r_g \in R_g : \mathrm{dist}(r_p, r_g)$
4: $S \leftarrow \mathrm{normalize}(S)$
Output: S, the similarity matrix

3. Item-response curve generation



Use a transformation function (e.g., Gaussian blur, occlusion, noise) to perturb the input sheep stimuli, and measure the match rate as the response.

To generate an item-response curve:

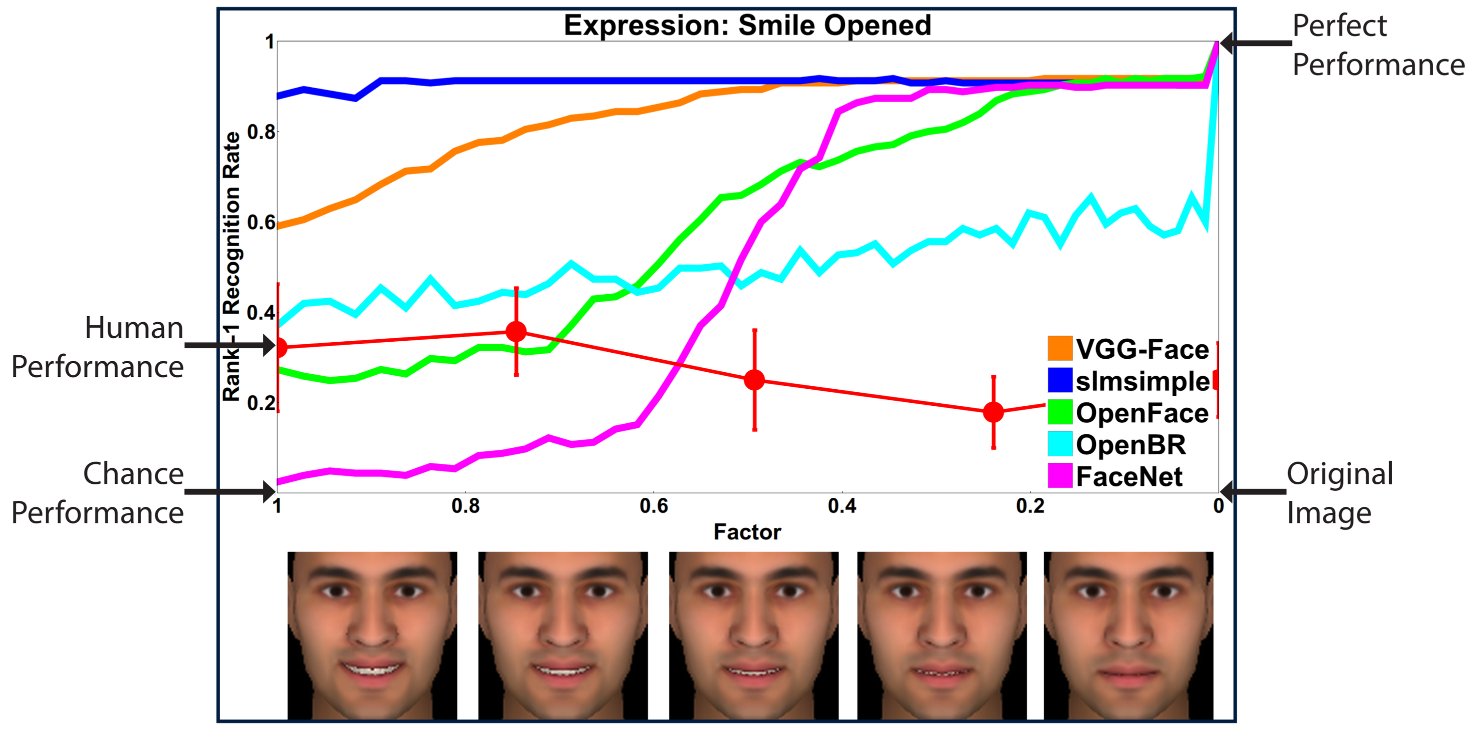
$\Phi_T(\Upsilon, I_h, t_h, \delta)$: an item-response point generation function for any image transformation function $T(i, \delta)$

Input: Υ, a "shepherd" function for a facial recognition model
Input: I_h, the "sheep" identities for the found threshold t_h
Input: t_h, the optimal threshold to produce I_h
Input: δ, the stimulus level
1: $I'_h \leftarrow i \in I_h : T(i, \delta)$
2: $S \leftarrow \Upsilon(I'_h, I_h)$
3: $M \leftarrow S \geq t_h$
4: $\alpha \leftarrow \frac{|M \wedge \mathcal{I}|}{|I_h|}$
Output: $\{s, \alpha\}$, an x, y coordinate pair (stimulus level, match rate)

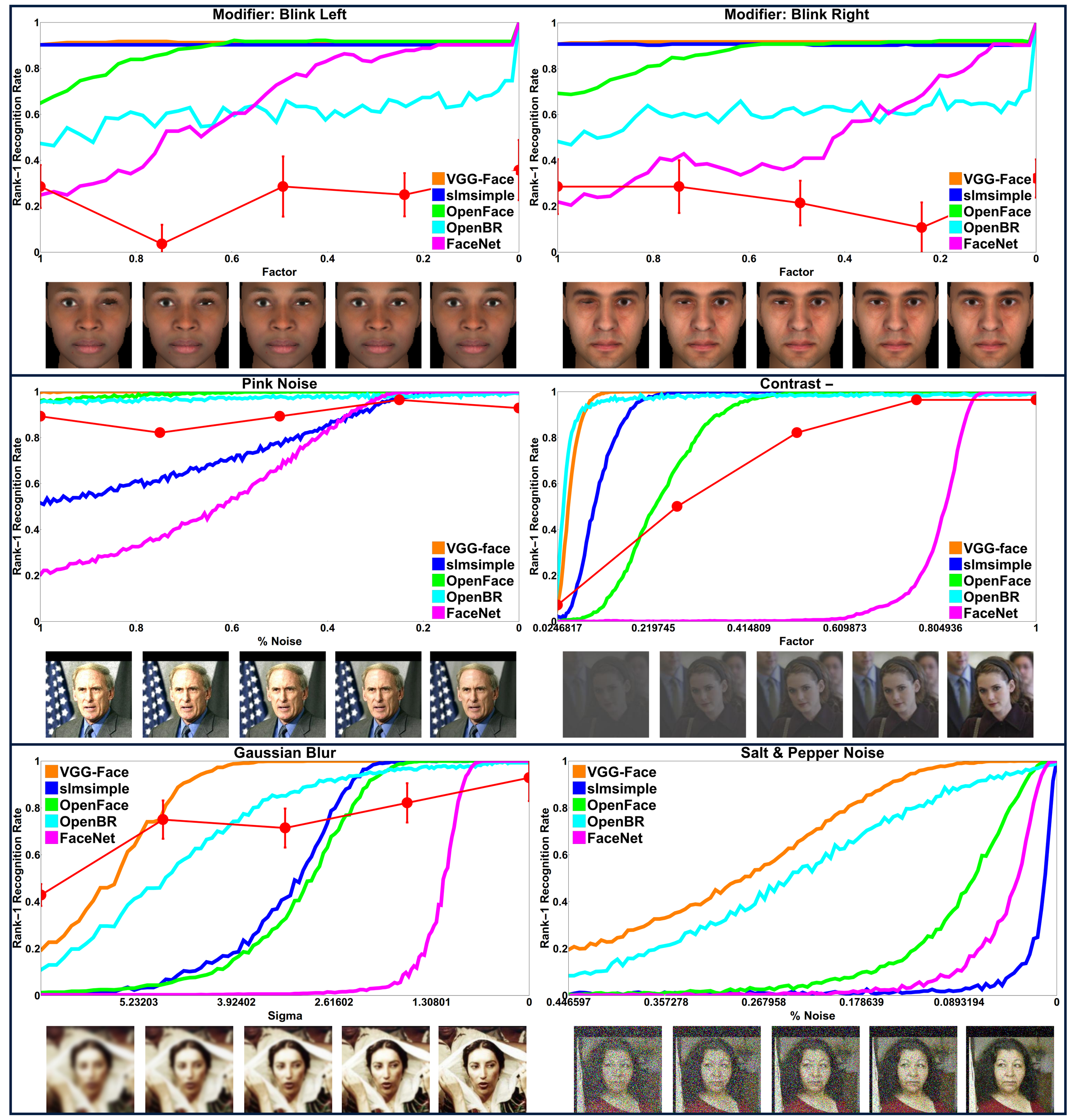
Experiments

We generated item-response curves using data from the LFW dataset and rendered FaceGen models. The images at the bottom of each curve show how the perturbations increase from right to left, starting with no perturbation.

	Identities	Sheep	Transformations	Levels	Images	Comparisons
LFW	1000	1000	14	200	~5.5 mil	~13.6 bil
FaceGen	220	206	22	50	~400,000	~17.5 bil



Increasing Perturbation



Interesting Observations

- OpenBR handcrafted features are not the worst performing approach
- "slmsimple", a shallow network with random weights, performs on par with VGG-Face for FaceGen experiments
- FaceNet and OpenFace exhibit radically different behavior despite both reporting to be implementations of Google's FaceNet [5]

References:

[1] Embretson, S.E., Reise, S.P.: Item response theory for psychologists. Lawrence Erlbaum Associates, Inc. (2000).
[2] Scheirer, W.J., Anthony, S.E., Nakayama, K., Cox, D.D.: Perceptual annotation: Measuring human vision to improve computer vision. IEEE T-PAMI 36(8) (2014) 1679–1686.
[3] RichardWebster, B., Anthony, S., Scheirer, W.: Psyphy: A psychophysics driven evaluation framework for visual recognition. IEEE T-PAMI (2018).
[4] Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.: Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. Technical report, NIST (1998).
[5] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. IEEE CVPR. (2015).

Acknowledgments:

Funding was provided under IARPA contract #D16PC00002, NSF DGE #1313583, and NSF SBIR Award #IIP-1738479. Hardware support was generously provided by the NVIDIA Corporation, and made available by the National Science Foundation (NSF) through grant #CNS-1629914.