Modeling second-order boundary perception: A machine learning approach

Abbreviated Title: Modeling second-order boundary perception

Christopher DiMattina [1,*] and Curtis L. Baker, Jr.[2]

[1]Computational Perception Laboratory, Department of Psychology,
Florida Gulf Coast University, Fort Myers, FL, USA 33965-6565

[2]McGill Vision Research Unit, Department of Ophthalmology,
McGill University, Montreal, QC, Canada H3G1A4

***Corresponding Author:**
Christopher DiMattina
Whitaker Hall Room 215
Florida Gulf Coast University
10501 FGCU Blvd S.
Fort Myers, FL 33965-6565

Email: cdimattina@fgcu.edu
Web: http://itech.fgcu.edu/faculty/cdimattina/
Tel: (239)-590-1513

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: We thank the FGCU undergraduates in the Computational Perception Lab for help with data collection, Wilson S. Geisler for code implementing decision-variable correlation, Elizabeth Zavitz for micropattern texture stimuli, and Fred Kingdom for helpful discussions and comments. Supported by Canadian NSERC Grants OPG0001978 and RGPIN-2017-05292 to C.L.B. and a FGCU Professional Development Grant to C.D.

Abstract

Background: Visual pattern detection and discrimination are essential first steps for scene analysis. Numerous human psychophysical studies have modeled visual pattern detection and discrimination by estimating linear templates for classifying noisy stimuli defined by spatial variations in pixel intensities. However, such methods are poorly suited to understanding sensory processing mechanisms for complex visual stimuli such as second-order boundaries defined by spatial differences in contrast or texture.

Methodology / Principal Findings: We introduce a novel machine learning framework for modeling human perception of second-order visual stimuli, using image-computable hierarchical neural network models fit directly to psychophysical trial data. We apply this framework to modeling visual processing of boundaries defined by differences in the contrast of a carrier texture pattern, in two different psychophysical tasks: (1) boundary orientation identification, and (2) fine orientation discrimination. We utilize cross-validation analysis to optimize model hyper-parameters, and demonstrate that our models are able to accurately predict human performance on novel stimulus sets not used for fitting model parameters. We find that, like the ideal observer, human observers take a region-based approach to the orientation identification task, while taking an edge-based approach to the fine orientation discrimination task. We further investigate how observers integrate contrast modulation across orientation channels, by fitting psychophysical data with two models representing competing hypotheses, finding a preference for a model which combines multiple orientations at the earliest possible stage. We argue that this machine learning approach has much potential to advance the study of second-order visual processing, and we outline future steps towards generalizing our method to modeling visual segmentation of natural texture boundaries.

Conclusions / Significance: This study demonstrates that machine learning methodology can be fruitfully applied to psychophysical studies of second-order visual processing.

Author Summary

Many naturally occurring visual boundaries are defined by spatial differences in features other than luminance, for example by differences in texture or contrast. Quantitative models of such “second-order” boundary perception cannot be estimated using the standard regression techniques (known as “classification images”) commonly applied to "first-order", luminance-defined stimuli. Here we present a novel machine learning approach to modeling second-order boundary perception using hierarchical neural network models. In contrast to previous modeling of second-order boundary perception, we directly estimate network parameters using psychophysical trial data. We demonstrate that our method can reveal different spatial summation strategies that human observers utilize for different kinds of second-order boundary perception tasks, and can be used to compare competing hypotheses of how contrast modulation is integrated across orientation channels. We outline extensions of the methodology to other kinds of second-order boundaries, including those in natural images. This study demonstrates how machine learning techniques can be fruitfully applied to psychophysical experiments aimed at characterizing perception of complex naturalistic boundary stimuli.

Introduction

Many of the most common functions of sensory systems involve detection, identification or discrimination of particular stimuli. For example, a critically important ability of early visual processing is to segment images into distinct regions, which may correspond to cohesive surfaces or objects (Marr, 1982). To better understand the underlying mechanisms of visual surface segmentation, it would be desirable to fit biologically plausible models of surface segmentation to psychophysical data, and assess how well such models can account for independent datasets.

A widely used approach to modeling human psychophysics has been to assume that performance can potentially be understood in terms of how well a visual stimulus matches an internal "template", typically modeled as a linear spatial filter. This approach uses high-dimensional regression models fit to data from tasks in which a subject detects or classifies a sensory stimulus presented with superimposed white noise (Ahumada, 1996, 2002; Neri and Levi, 2006; Murray, 2011), to recover an estimate of the linear filter that best accounts for the data. This system-identification (SI) approach to psychophysics, often called “psychophysical reverse correlation” or “classification image analysis”, parallels neurophysiology studies that fit linear filter models to neural responses elicited by white noise stimuli (e.g. DeAngelis, Ohzawa, and Freeman 1993; Chichilnisky, 2001). Psychophysical SI methods have been applied in varied problem domains and yielded tremendous insight into perceptual mechanisms (Abbey and Eckstein, 2002; Eckstein, Shimozaki, and Abbey, 2002; Li, Klein, and Levi, 2004; Neri and Levi, 2006; Morgenstern and Elder, 2012; McIlhagga and May, 2012; Kurki, Saarinen, and Hyvarinen, 2014).

One limitation of most previous psychophysical system identification studies has been the use of stimuli that are typically defined by spatial or temporal variations in first-order statistics like luminance or color. However, many natural visual stimuli such as occlusion boundaries (Fig. 1a, top) are defined by spatial variations in higher-order image statistics, e.g. texture or contrast, rather than simple luminance differences (Schofield, 1999; Konishi et al., 2003; Johnson and Baker, 2004, 2005; Martin, Fowlkes, and Malik 2004; DiMattina, Fox, and Lewicki 2012). Such boundaries defined by texture or contrast (Fig. 1a, bottom) are commonly referred to as *second-order* boundaries or edges (Baker, 1999; Landy and Graham, 2004). These stimuli have been of particular interest because their detection cannot be explained by simple

linear filters, and consequently human performance cannot be readily modeled using conventional psychophysical SI methods. More generally, there is a need for a psychophysical modeling approach that can handle a broader range of stimuli and tasks.

Here we present a novel machine learning method to extend psychophysical SI to more complex naturalistic stimuli such as second-order boundaries. Our general approach is to fit image-computable hierarchical neural network models of second-order vision (Fig. 2) directly to psychophysical trial data, similar to recent studies using similar hierarchical network models to characterize nonlinear responses in sensory neurons (Willmore, Prenger, and Gallant, 2010; Vintch, Movshon, and Simoncelli, 2015; Harper et al. 2016; Rowenkamp and Sharpee, 2017). Our application of this framework to second-order boundary detection generalizes a popular class of biologically plausible psychophysical models of second-order vision known as "Filter-Rectify-Filter" (FRF) models (Landy and Graham, 2004). However, in contrast to previous general schemes for second-order vision (e.g. Malik and Perona, 1990; Bergen and Landy, 1991; Graham, Beck and Sutter, 1992) and quantitative models (Zavitz and Baker, 2013; Westrick and Landy, 2013), we take a machine learning approach in which we estimate the defining parameters of these models by direct fitting to psychophysical trial data. This permits us to make direct tests of model predictions on novel datasets in order to validate our models, and to evaluate competing models.

We focus on contrast-defined second-order boundaries (Fig. 1a, b) since they are well studied psychophysically (Dakin and Mareschal, 2000; Schofield and Georgeson, 1999, 2003; Ledgeway and Smith, 1994; Ellemberg, Allen, and Hess 2006), and there are neurons in the early mammalian visual cortex that give selective responses to contrast modulation stimuli (Mareschal and Baker, 1998; Tanaka and Ohzawa, 2006; Rosenberg et al, 2010; Li et al., 2014). We model performance on two tasks making use of contrast-defined boundaries: (1) identification of boundary orientation, and (2) boundary orientation discrimination. We find that subjects utilize spatial information in markedly different ways for these two tasks, using "region-based" processing for the orientation identification task and "edge-based" processing for the orientation discrimination task, in both cases consistent with an ideal observer.

We further demonstrate the power of this methodology for evaluating competing hypotheses regarding FRF model architecture (Kingdom, Prins, and Hayes, 2003; Motoyoshi and Nishida, 2004; Motoyoshi and Kingdom, 2007; Westrick and Landy 2013) by comparing the

ability of two different model architectures to fit human performance on a contrast edge detection task in which the carrier pattern contains elements with two orientations. Post-hoc Bayesian model comparison (Schwarz, 1978; Bishop, 2006) demonstrates a preference for a model in which carrier pattern orientation is integrated across channels by the second-stage filters. We argue that integration across orientation channels by second-stage filters may be optimal for task performance, and that this perceptual strategy is consistent with known visual neurophysiology.

Finally, we discuss potential applications of this methodology to future quantitative studies of first- and second-order cue integration (Allard and Faubert, 2007; Saarela and Landy, 2012; Hutchinson, Ledgeway, and Baker, 2016) and natural boundary detection (McDermott, 2004; DiMattina, Fox, and Lewicki, 2012). We suggest that the application of machine learning methodology similar to that presented here can potentially help to better quantify visual processing of complex visual stimuli like natural image region boundaries, complex shapes, and ecologically relevant objects.

Results

Observers performed two different psychophysical tasks making use of contrast-modulated boundary stimuli like those shown in Fig. 1. The first task was an *orientation-identification* task (Fig. 1c, left) in which observers determined in which of two oblique orientations (-/+ 45 deg. or L/R) a near-threshold contrast boundary was presented, with varying amounts of contrast modulation (Experiments 1 and 3). The second task was an orientation-discrimination task (Fig. 1c, right) in which the orientation of a supra-threshold contrast boundary was perturbed slightly from vertical by varying amounts (-/+ 6-7 deg.) and observers indicated the direction (Experiment 2). Trial-wise responses obtained from these experiments were used to fit and evaluate a model (Fig. 2) implementing a "Filter-Rectifty-Filter" (FRF) operation (Landy and Graham, 2004) on each stimulus image, with the goal of characterizing the nature of spatial summation for perception of contrast boundaries, as well as the shape of the second-stage filter non-linearity (Fig. 2, shown in blue). The model in Fig. 2 is a specific instance of a broader modeling framework for describing psychophysical data (METHODS) that naturally extends the linear classification image technique, much as in neuronal subunit models (Vintch et al. 2015; Harper et al. 2016; Rowenkamp and Sharpee, 2017) extend traditional linear receptive field modeling (DeAngelis et al. 1993). We describe the model in greater depth before presenting our main experimental results.

Modeling framework

The model used to fit psychophysical trial data (Fig. 2) was comprised of a bank of first-stage Gabor filters that compute local oriented energy in a manner similar to V1 complex cells (Adelson and Bergen, 1985). The normalized outputs of these first-stage filters were then down-sampled (details below) and fed into a pair of second-stage filters representing the two task alternatives (L/R). In contrast to the first-stage filters, these second-stage filters analyze the image globally and can therefore integrate variations in contrast across the boundary. Finally, the outputs of these two second-stage filters are compared with each other (via simple subtraction) to decide on a trial-by-trial basis if a given stimulus boundary appeared more likely to be tilted left or right (L/R). The output of the model is the probability $0 < P(R) < 1$ of the stimulus being

classified by the observer as R. The main goal of Experiments 1 and 2 was to learn the shapes of these second-stage filters employed in these tasks.

Since the images were 256x256 pixels, it is necessary to perform down-sampling between the first-stage and second-stage filters to reduce parameter space dimensionality. Previous vision modeling has considered various methods for spatial pooling of filter outputs (Morgenstern and Elder, 2012; Meese, 2010; Baker and Meese, 2014) and here we implemented three methods: (1) Average (AVG) pooling, (2) MAX (MAX) pooling (Goodfellow, Bengio, and Courville, 2016), and (3) Sub-sampling (SUB) pooling (Bergen and Landy, 1991). In Experiment 1 we considered all three pooling rules and 3 down-sampling sizes (22x22, 16x16 and 8x8) whereas in Experiments 2 and 3 we considered only AVG pooling and a single down-sampling size (Experiment 2: 22x22, Experiment 3: 12x12).

We implement the model's final classification (L/R) in three different ways: In the *deterministic* model (DET), if the output of the L filter is larger than that of the R filter (plus the bias term) so that P(R) > 0.5, the stimulus will be classified as R, otherwise it is classified as L. In the *noise-added* model (NSE), we add Gaussian noise to the second-stage filter outputs, with the noise variance optimized to match subject performance, as in previous work (Zavitz and Baker, 2013). Finally, in the *stochastic* model (STO), the model output P(R) specifies only the probability of classifying the stimulus as R or L. This introduces randomness to the model without needing to specify an additional free parameter.

A common approach to make the estimation problem more tractable is to utilize reasonable prior constraints on the model parameter space (Mineault, Barthelme, and Pack, 2009; Wu, David and Gallant, 2006; Park and Pillow, 2011). Here we use two kinds of Bayesian priors: (1) A ridge-regression prior (Hoerl and Kennard, 1970; Bishop 2006) to penalize large weights (*ridge*) and (2) A ridge-regression prior combined with an additional smoothness prior (Mineault et al., 2009) to penalize excessively "rough" second-stage filter weight profiles (*ridge + smooth*). We make use of the standard machine learning technique of *k*-fold cross-validation (Wu et al., 2006) to optimize the hyper-parameters defining the Bayesian priors. Performing *k*-fold cross-validation (Experiment 1-VAR: $k = 4$, Experiment 1-FIX: $k = 6$, Experiment 2: $k = 6$, Experiment 3: $k = 5$) we found that the hyper-parameter values $\lambda = 0.1 - 1$ (ridge regression prior) and $\omega = 1$ (smoothing prior) provided a good trade-off between fitting the training data and generalizing well to the validation data (Supplementary Figs. S1-S4). In all analyses

presented here, we used the hyper-parameter values $\lambda = 1, \omega = 1$ unless otherwise noted. Additional details of the model are given in METHODS.

Experiment 1: Boundary orientation identification

Stimuli and task

Observers were presented with contrast-defined boundaries with varying degrees of modulation depth (Fig. 1b), and performed forced-choice judgments of whether the boundary was oriented right- or left-oblique (45 degrees clockwise or counterclockwise), as shown in Fig. 1**c** (left). Boundaries were created by multiplying a high-spatial frequency carrier micro-pattern by a low spatial-frequency contrast-modulation envelope. By varying the depth of contrast modulation, we could manipulate the task difficulty. In one version of Experiment 1 (Experiment 1-VAR, N = 7 observers), contrast modulation depth was varied in a range mostly near threshold but also including some easy and chance levels. In another version (Experiment 1-FIX, N = 3 observers) the contrast modulation depth was held constant at levels near threshold.

Model fit to representative observers

Fig. 3a shows the results of fitting the model (16x16 down-sampling, AVG pooling) to data from one representative observer (AMA) in Experiment 1-VAR. Plots in the top panels show the second-stage filter weights averaged over N = 4 training sets (METHODS), and those in the bottom show the mean 1-D profiles. Black dashed lines show the 1-D profiles obtained by collapsing the 2-D profiles plotted above, whereas thick solid lines (red = *ridge*, green = *ridge + smooth*) show the mean 1-D profile obtained from over 30 bootstrapped resampled training sets. We see that these mean 1-D profiles are nearly indistinguishable. We find that the 1-D weight magnitudes lie well above zero, which generally lies well outside the 68% (+/- 1 SEM) and 95% (+/- 1.96 SEM) confidence intervals (Fig. 3, thin dashed lines show +/- 1 SEM). However, this is not the case for many of the individual 2-D weights (Supplementary Fig. S7).

We see from Fig. 3 that the average 1-D weight profiles are approximately constant as a function of distance from the diagonal, consistent with region-based processing. We also see

that the addition of the smoothing prior (*ridge + smooth*) results in weighting functions which are less "noisy" in appearance, consistent with previous results with classification images and receptive field modeling (Mineault et al., 2009; Park and Pillow, 2011). Furthermore, we found that choice of pooling rule did not affect the general shape of the weighting functions, as all three pooling rules yielded consistent 2-D and 1-D weight profiles (Supplementary Fig. S8).

Fig. 3b shows the model in Fig. 2 fit to data from a representative observer (JJF) in Experiment 1-FIX. In this version of the experiment, the contrast modulation depth was kept constant at approximately threshold level for each subject (Supplementary Table S1), similar to most previous work in the classification image literature (Abbey and Eckstein, 2002; Murray, 2011). Here we see very similar results as for the subject in Experiment 1-VAR, also observing evidence in favor of a region-based strategy for this task. As before, this was robust to pooling rule (Supplementary Fig. S9).

In order to determine whether subjects were employing the same spatial integration strategy as an ideal observer, we fit the model to stimulus category labels rather than observer responses. Fig. 3c shows the 2-D and 1-D profiles obtained for an ideal observer in Experiment 1-VAR, and we see that the ideal strategy is region-based. Similar results for ideal observer fit to stimulus category labels in Experiment 1-FIX are shown in Supplementary Fig. S10.

Edge-based vs. region-based analysis

Fig. 4 shows the 1-D averaged second-stage filter weights (as in Fig. 3) for all additional observers in Experiment 1-VAR (Fig. 4a) and Experiment 1-FIX (Fig. 4b). We see that the magnitudes of the 1-D weights are approximately constant as one moves from the diagonal to the boundary, consistent with region-based processing, and these results are consistent for SUB and MAX pooling rules. A plot of the normalized ideal observer 1-D weight profiles (30 bootstraps) is shown in Fig. 5a as a thick black line with 95% confidence intervals (+/- 1.96 SEM) for the ideal observer plotted as dashed black lines. We see that the normalized 1-D weight profiles for all observers (dashed color lines, mean over observers shown by thick red dashed line) generally lie within the 95% confidence interval of the ideal observer.

We further quantified the degree to which the 1-D profiles were localized near the diagonal by fitting a "rising/falling edge" function (METHODS) to the 1-D profile, where the

degree of localization to the boundary is captured by the free parameter ρ, with smaller values indicating greater localization to the edge (Fig. 6a). We see from our analysis of the population (Fig. 6c) that the 1-D weight profiles are not strongly localized to the diagonals.

The finding that the 1-D weighting functions do not seem to show a much stronger preference for the boundary region is somewhat counter-intuitive in light of the fact that previous neurophysiological investigations suggest contrast-modulated (CM) neurons in extra-striate cortex have Gabor-shaped envelope profiles (Li and Baker, 2012; Li et al., 2014). One possible explanation may be that the population of CM neurons has a preferred ratio of carrier:envelope spatial frequency of about 8:1. This was the ratio for the stimulus used in Experiment 1, since the micro-patterns were 32 pixels in size and the envelope was 256 pixels. It is possible that with our stimulus that CM neurons with smaller receptive fields localized to the boundary were not engaged by our stimulus, since for such neurons their preferred carrier patterns would have higher spatial frequencies (on average). Therefore, we ran an additional control version of Experiment 1-VAR on a population of N = 3 naïve observers (Experiment 1-HFC) where we reduced the micro-pattern size to 8 pixels (while increasing the density by a factor of 4), so that the carrier:envelope spatial frequency ratio was 32:1. This would potentially engage many more neurons whose receptive fields are localized to the boundary. However, we found no difference in spatial strategy in this case, with observers also tending to take a region-based strategy **(**Fig. 5b) in Experiment 1-HFC as well. Statistical tests revealed no difference between the median values of the localization parameter in Experiment 1-HFC and Experiment 1-VAR (Independent Samples Median Test (Fisher Exact), *ridge:* $p = 0.628$, *ridge + smooth:* $p = 0.628$, $N_1 = 14$, $N_2 = 6$).

<u>Second-stage filter non-linearity</u>

In Experiment 1, we found an expansive (>1) nonlinearity exponent α for the second-stage filters for all observers for both the *ridge* and *ridge + smooth* priors. For Experiment 1-VAR with 16x16 AVG sampling, we observed median values around 2 (*ridge*: median = 2.15, min = 1.70, max = 2.85; *ridge + smooth*: median = 1.95, min = 1.70, max = 3.35, N = 7), and for Experiment 1-FIX we obtained median values closer to 1.5 (*ridge*: median = 1.50, min = 1.40, max = 1.60; *ridge + smooth*: median = 1.60, min = 1.60, max = 1.90, N = 3). Similar results were obtained

for other down-sampling methods (Supplementary Tables S2, S3), and the ideal observer also learned an expansive non-linearity (16x16-AVG; Experiment 1-VAR: *ridge*: 4.00; *ridge + smooth*: 3.30; Experiment 1-FIX: *ridge*: 2.20; *ridge + smooth*: 2.30). Plots of the second-stage filters for observers AMA, JJF are shown in Supplementary Fig. S11.This finding is consistent with previous studies suggesting an accelerating non-linearity following the second-stage filters.

Analysis of model accuracy

Psychometric function agreement

We constructed psychometric functions of proportion correct on the orientation identification task as a function of the modulation depth across the boundary for observers in Experiment 1-VAR, and three representatives are shown in Fig. 7 (AVG down-sampling). We see that the model predictions on the test datasets (which were not used to train models or optimize hyper-parameters), generally lie within or near the 95% binomial proportion confidence intervals (+/- 1.96 SEM) of subject performance (Fig. 7, blue dashed lines). However, for the deterministic (DET) model (leftmost column), we observed a tendency of the model to over-predict performance at larger contrast modulations for many observers. Implementing the model stochastically (STO, center column) or adding noise (see METHODS) to the deterministic model (NSE, right column) improved the agreement between observer and model performance. Similar results as those shown in Fig. 7 were observed with MAX down-sampling. For the deterministic model with SUB down-sampling, we see a reduced tendency to over-predict performance (Supplementary Fig. S12), perhaps because the sub-sampling strategy integrates information over a much smaller number of first-stage filters. The over-prediction of the (deterministic) AVG/MAX models and better performance of the SUB model (with appropriate sampling rate – see Supplementary Fig. S13) would be consistent with the idea that human observers are sub-optimally pooling the local contrast information.

Overall performance agreement

Fig. 8 shows observer and model performance in Experiment 1-VAR for each individual test fold on every observer (7 observers, 4 test folds each = 28 total folds). For the deterministic model with AVG pooling (AVG-DET, Fig. 8a, top left), we found that there was a tendency to over-predict subject performance, with the majority (*ridge*: 18/28; *ridge + smooth*: 23/28) of test folds having significantly different (binomial proportion difference test, 95% CI) proportions of correct responses (Fig. 8b, top row). Comparison of subject and model performance (difference = subject - model) across subjects (sign-rank test) reveled a significant median difference (*ridge*: median = -0.06, $p = 0.038$*; ridge + smooth*: median = -0.08, $p = 0.002$), as did comparison across all test folds (*ridge*: median = -0.06, $p < 0.001$*; ridge + smooth*: median = -0.08, $p < 0.001$, N = 28). Despite the systematic over-prediction of subject performance (by about 6-8%), we still observe significant correlations between predictions and actual performance over both subjects (*ridge*: $r = 0.983$, $p < 0.001$; *ridge + smooth*: $r = 0.972$, $p < 0.001$, N = 7) and folds (*ridge*: $r = 0.872$, $p < 0.001$; *ridge + smooth*: $r = 0.806$, $p < 0.001$, N = 28).

By contrast, we found that when the classification was implemented stochastically we observed much better agreement (AVG-STO: Fig. 8a, top right), with relatively few (*ridge*: 4/28; *ridge + smooth*: 6/28) test folds having significantly different proportions (Fig. 8b, second row). Comparison of subject and model performance (sign-rank test) failed to reveal a significant median difference across subjects (*ridge*: median = 0.02, $p = 0.318$; *ridge + smooth*: median = 0.03, $p = 0.209$, N = 7), although we did observe a slight (but significant) under-prediction of subject performance across folds (*ridge*: median = 0.02, $p = 0.034$; *ridge + smooth*: median = 0.02, $p = 0.014$, N = 28). As before, strong correlations between model and subject performance are observed across subjects (*ridge*: $r = 0.996$, $p < 0.001$; *ridge + smooth*: $r = 0.997$, $p < 0.001$, N = 7) and folds (*ridge*: $r = 0.890$, $p < 0.001$; *ridge + smooth*: $r = 0.911$, $p < 0.001$, N = 28).

Similar to implementing the classification stochastically, adding noise (Supplementary Table S4) to the deterministic model (AVG-NSE) was also effective at increasing agreement with observer performance. We find few folds (*ridge*: 2/28; *ridge + smooth*: 3/28) having significantly different proportions of correct responses, and no significant differences across subjects (*ridge*: median = 0.0, $p = 1.0$; *ridge + smooth*: median = 0.0, $p = 0.71$, N = 7) or folds (*ridge*: median = 0.0, $p = 0.851$; *ridge + smooth*: median = 0.0, $p = 0.476$, N = 28).

We see in Fig. 8 that in the deterministic case, performance was more accurate for the SUB down-sampling strategy (SUB-DET) than for either AVG or MAX down-sampling. We find many fewer folds (*ridge*: 10/28; *ridge + smooth*: 8/28) having significantly different proportions of correct responses, and no significant differences across subjects (*ridge*: median = 0.01, $p = 0.383$; *ridge + smooth*: median = -0.01, $p = 0.710$, N = 7) or folds (*ridge*: median = 0.02, $p = 0.075$; *ridge + smooth*: median = -0.01, $p = 0.629$, N = 28).

Similar results were obtained for Experiment 1-FIX, shown in Supplementary Fig. S14. We found over-prediction of performance for the DET model with AVG down-sampling for nearly all folds (*ridge*: 16/18; *ridge + smooth*: 18/18), with a significant difference (sign-rank test) across folds (*ridge*: median = -0.08, $p < 0\ .001$; *ridge + smooth*: median = -0.1, $p < 0.001$, N = 18). Applying the model stochastically (AVG-STO) somewhat reduced this discrepancy, leading to somewhat fewer folds differing with subject performance (*ridge*: 12/18; *ridge + smooth*: 13/18), but now we saw an under-prediction (by about 7%) of performance across folds (*ridge*: median = 0.07, $p < 0.001$; *ridge + smooth*: median = 0.07, $p < 0.001$, N = 18). Introducing and optimizing an internal noise parameter β for each subject (METHODS, Supplementary Table S5), we were able to drastically enhance model performance (Supplementary Fig. S14a, b: AVG-SOP). In this case we found few if any folds exhibited significant disagreement between model and subject (*ridge*: 0/18; *ridge + smooth*: 2/18) with no significant difference across folds (*ridge*: median = 0.00, $p = 0\ .949$; *ridge + smooth*: median = -0.01, $p < 0.383$, N = 18).

As with Experiment 1-VAR, adding noise optimized for the validation data (AVG-NSE) also yielded strong agreement between model and subject performance, with relatively few folds (*ridge:* 1/18; *ridge + smooth:* 3/18) showing significant difference, and no significant difference across folds (*ridge*: median = 0.00, $p = 0.318$; *ridge + smooth*: median = -0.01, $p = 0.150$, N = 18). Similarly, in deterministic mode we found somewhat better performance for SUB down-sampling (SUB-DET) than AVG or MAX down-sampling, with many fewer folds showing significant disagreement (*ridge*: 8/18; *ridge + smooth*: 9/18), and smaller discrepancies across folds (*ridge*: median = 0.04, $p = 0.043$; *ridge + smooth*: median = 0.02, $p = 1.000$). In all cases of Experiment 1-FIX, we did not observe significant linear correlations between fold and model performance, most likely due to the fact that there was little variability in performance (range restriction) since we attempted to hold task performance approximately constant near threshold.

For both Experiment 1-VAR and Experiment 1-FIX, we obtained similar results with MAX pooling.

Effects of down-sampling rate

We analyzed the effects of the down-sampling rate on model accuracy for Experiment 1-FIX for the AVG, MAX and SUB down-sampling methods. We find that model performance is relatively independent of down-sampling rate for the AVG and MAX cases, over-predicting subject performance in all cases for the deterministic model (Supplementary Fig. S13). In contrast, we find that it is necessary to optimize the sampling rate with the deterministic SUB model – we observed major under-prediction of subject performance for 8x8, with much better agreement for 16x16 and 22x22 (Supplementary Fig. S13). This may be because higher sampling rates integrate over enough first-stage filters to compensate for stimulus-to-stimulus pattern variability and integrate contrast modulations as well as the human observers. We return to this point in the DISCUSSION.

Double-pass analysis

For N = 5 observers in Experiment 1-VAR, we performed a double-pass analysis (Neri and Levi, 2006; Neri, 2010; Hasan, Joosten, and Neri, 2012) in which observers were tested on the same set of stimuli twice. Testing twice on the same set of stimuli provides an upper bound on the performance of any model by comparing model-observer agreement on the double-pass set to observer-observer agreement on passes 1 and 2. A fold of N = 1000 stimuli was used as the double-pass set, and the model was trained on the remaining N = 3000 stimuli. We see in Fig. 9 that the DET-AVG or DET-MAX models agree with the observer on the two passes as well as the observer agrees with him/her-self on the two passes. For DET-AVG, we find across subjects for both priors a small and statistically insignificant (rank-sum test) median difference (*ridge*: median = 0.022, $p = 0.063$; *ridge* + *smooth*: median = 0.009, $p = 0.313$, N = 5), with most or all of the (binomial proportion difference) confidence intervals containing zero (*ridge*: 4/5; *ridge* + *smooth*: 5/5). Similarly, for DET-MAX we have small median differences (*ridge*: median = 0.018, $p = 0.125$; *ridge* + *smooth*: median = 0.02, $p = 0.313$, N = 5) with most confidence

intervals (*ridge*: 4/5; *ridge + smooth*: 4/5) containing zero. In contrast, for the DET-SUB (Fig. 9a**,** bottom panels), we find somewhat larger differences (*ridge*: median = 0.052, $p = 0.063$; *ridge + smooth*: median = 0.024, p = 0.063, N = 5), although they fail to reach statistical significance. In the DET-SUB case we find for the *ridge* prior 0/5 confidence intervals containing zero, and 3/5 for *ridge + smooth*.

We see in Fig. 9b that implementing the decision with AVG down-sampling stochastically (AVG-STO) or adding noise (AVG-NSE) somewhat reduces the agreement between the model and the subject. For AVG-STO, we find larger differences than AVG-DET, although these also fail to reach significance (*ridge*: median = 0.075, $p = 0.063$; *ridge + smooth*: median = 0.088, $p = 0.063$, N = 5), with none of the confidence intervals containing zero. Similarly, for AVG-NSE we also find somewhat larger differences (*ridge:* median = 0.067, $p = 0.063$; *ridge + smooth:* median = 0.071, $p = 0.063$, N = 5), with only 1/5 confidence intervals containing zero (both priors).

Decision variable correlation analysis

Although comparing the percentage correct between a model and observer and finding strong agreement is consistent with the interpretation that the model is accurately modeling observer strategies, it is not fully compelling. It may be the case that while the model and observer both choose category A or B the same proportion of times, there is no correlation between observer and model responses on a trial-by-trial basis. This situation could possibly arise in cases where the observer and model are making use of different stimulus features for their decisions. Therefore, to more carefully quantify model performance, we implemented a recently developed analysis known as *decision variable correlation* (DVC) analysis which estimates the correlation between a model's decision variable and the observer's hypothesized decision variable using trial-by-trial responses (Sebastian and Geisler, 2018). This method extends standard signal detection theory by hypothesizing that the joint distribution of model and observer decision variables for each stimulus category (A = left-oblique/L, B = right-oblique/R) can be modeled as bi-variate Gaussian distributions having correlation parameters ρ_A, ρ_B which characterizes the correlation between model and observer decision variables. Given the proportions of times the

observer and model agree and disagree in their decisions for each stimulus category allows one to estimate ρ_A, ρ_B using a simple maximum likelihood procedure (Sebastian and Geisler, 2018).

Previous work (and our own investigations) reveal that DVC analysis is most accurate (as measured by bootstrapped estimates of SEM) for large numbers of trials at a constant stimulus level sufficiently near threshold that consistent proportions of the four logical possibilities (model says A/B, observer says a/b) are obtained (Sebastian and Geisler, 2018). Therefore, we applied DVC analysis to Experiment 1-FIX, where we had thousands of trials (N = 6000) of both observer and model responses at a single stimulus level, enabling us to get several reliable estimates of DVC from each of the three observers (CJD, MAK, JJF). In this analysis, for each observer we trained the model on N = 5000 trials and computed the DVC using the remaining N = 1000 trials, for N = 6 choices of hold-out fold. Since one obtains independent estimates of the the DVC using both trials where the stimulus is L as well as R, we obtained 12 estimates of DVC from each observer**.** Fig. 10 shows the DVC measured from each subject (blue = CJD, magenta = MAK, black = JJF) for each fold for both stimulus alternatives. Thin lines denote +/- 1 SEM, which was estimated using N=200 bootstrapped re-samplings. We see from the plots that these estimates of DVC are overwhelmingly positive and significantly (> 1.96 SEM) larger than zero. Therefore, it seems that not only are the models and observers in agreement about overall proportions of each category, but are strongly correlated in their decisions on a trial-by-trial basis, consistent with the interpretation that the models are using the same processing strategy as the human observers.

Applying our analysis to the 16x16 models, we find that for AVG down-sampling, 32/36 (*ridge*) and 33/36 (*ridge + smooth*) estimates of DVC are significantly larger than zero, with DVC mean +/- SD of 0.235 +/- 0.094 (*ridge*) and 0.254 +/- 0.095 for (*ridge + smooth*). We see similar results for MAX down-sampling, where 33/36 (*ridge*) and 35/36 (*ridge + smooth*) estimates of DVC are significantly greater than zero, with DVC mean +/- SD of 0.245 +/- 0.094 (*ridge*) and 0.263 +/- 0.086 (*ridge + smooth*). Interestingly, we find weaker correlations for SUB down-sampling with only 20/36 (*ridge*) and 24/36 (*ridge + smooth*) estimates of DVC significantly greater than zero, and DVC means of mean +/- SD of 0.147 +/- 0.082 (*ridge*) and 0.164 +/- 0.085 (*ridge + smooth*). Pairwise comparisons of median DVC differences revealed significant differences between AVG and SUB (*ridge*: difference = 0.096, $p < 0.001$; *ridge + smooth*: difference = 0.104, $p < 0.001$) and MAX and SUB (*ridge*: difference = 0.094, $p < 0.001$;

ridge + smooth: difference = 0.102, $p < 0.001$). However, we observed less significant or non-significant differences between AVG and MAX median DVC (*ridge*: difference = -0.008, $p = 0.041$; *ridge + smooth*: difference = -0.002, $p = 0.131$). We observed no significant differences (paired comparsion test) between the estimates of DVC attained from L and R trials for both *ridge* (AVG: median = 0.005, $p = 0.396$; MAX: median = -0.003, $p = 0.500$; SUB: median = 0.015, $p = 0.744$, N = 18) and *ridge + smooth* (AVG: median = -0.02, $p = 0.679$; MAX: median = 0.005, $p = 0.327$; SUB: median = 0.005, $p = 0.586$, N = 18) priors.

One can also apply DVC analysis to determine how well an observer's performance at time 1 correlates with that same observer's performance on the same stimulus set at time 2 (Sebastian and Geisler, 2018), extending standard double-pass methods used to estimate internal noise (Neri, 2010; Hasan, Joosten and Neri, 2012). This allows one to put the values of the model-observer DVC in context, since the observer-observer DVC provides us with a theoretical upper limit on model-observer correlation. Performing DVC analysis for each observer tested in the double-pass analysis (Experiment 1-VAR, 5 observers) for our model (16x16 AVG downsampling) at four different near-threshold contrast modulation levels revealed a median observer-observer DVCs of 0.454 (both priors). Unfortunately, since for each level we only had 100 trials, and the variance of the DVC estimates can be quite large with small numbers of trials (Sebastian and Geisler, 2018), relatively few of the individual observer-observer DVC values (Supplementary Fig. S15) were significantly greater than zero (17/40, both priors). However, across all observers and levels, the median observer-observer DVC (0.454) was significantly greater than zero (sign-rank test: $p < 0.001$, N = 40). For both Pass 1 and Pass 2, the median observer-model DVC (*ridge:* Pass 1 median = 0.194, Pass 2 median = 0.141; *ridge + smooth*: Pass 1 median = 0.223, Pass 2 median = 0.235) were somewhat smaller, but also significantly greater than zero (sign-rank test: $p < 0.001$, N = 40). Taking the pass with the larger DVC as representative of model-observer DVC, we obtained a somewhat larger median (*ridge*: 0.317; *ridge + smooth*: 0.343). Comparing the median observer-observer DVC with observer-model DVC on each pass (pairwise-comparison test) revealed a significant difference for each individual pass (*ridge*: Pass 1 median difference = 0.157, $p = 0.002$, Pass 2 median difference = 0.234, $p = 0.001$; *ridge + smooth*: Pass 1 median difference = 0.161, $p = 0.001$, Pass 2 median difference = 0.235, $p = 0.003$). However, when the observer-model DVC was taken to be the pass yielding the larger DVC, we did not observe a significant difference between observer-

observer and observer-model DVC (*ridge*: median difference = 0.032, $p = 0.361$; *ridge + smooth*: median difference = 0.021, $p = 0.687$). Overall, our DVC analysis reveals that although there are definitely strong and significant positive correlations between the observer and model on a trial-by trial basis, the model is somewhat less accurate at predicting decisions on a trial-by-trial basis than using the observer's own previous responses to the same stimuli.

Ideal-observer analysis

In order to compare the spatial strategies and task performance of human observers to those of an ideal observer, we trained our model on the true stimulus category labels rather than observer classifications of these stimuli. We find that the ideal observer models make use of a region-based strategy similar to the human observers (Fig. 3). Comparing subject performance to that of the deterministic (16x16 AVG-DET) ideal observer in Experiment 1-VAR (Supplementary Fig. 16) revealed that across N = 6 observers analyzed, most confidence intervals (binomial proportion difference test, 95% confidence interval) did not contain zero (*ridge*: 6/24; *ridge + smooth:* 3/24), with those containing zero coming mainly from two observers (AMA, SRS). In contrast, we found that for the stochastic ideal observer (AVG-STO) many more confidence intervals contained zero (*ridge*: 11/24; *ridge + smooth*: 10/24), suggesting a better agreement with observer performance. Similar results were observed for the N = 3 observers tested in Experiment 1-FIX (Supplementary Fig. S16). Comparing human to deterministic ideal observer (AVG-DET) performance, we also found that most confidence intervals did not contain zero (*ridge*: 3/18; *ridge + smooth*: 4/18), while for the stochastic ideal observer most did contain zero (*ridge:* 14/18; *ridge + smooth*: 15/18). On the whole, this is similar to our findings for the models fitted to observer data – namely a general tendency for the AVG + MAX models operating deterministically to over-predict performance, while these models operating stochastically seem to agree reasonably well.

Correlations between observer performance and model parameters

Although there was a robust finding that the non-linearity is expansive, there was weak evidence across the set of individuals in Experiment 1-VAR that the magnitude of the exponent α

explained individual variations in performance (16x16 AVG, Pearson; *ridge*: r = -0.458, p = 0.302; *ridge + smooth:* r = -0.787, p = 0.036, N = 7). The significant result for *ridge + smooth* case was due to the presence of an outlier, and computing these correlations using Spearman's rho revealed no significant effect (16x16 AVG, Spearman's rho; *ridge: rho* = -0.357, p = 0.444; *ridge + smooth*: *rho* = -0.214, p = 0.662). Results were similar for MAX and SUB models (Supplementary Table S7).

By contrast, we obtained compelling evidence that individual variability in observer performance is highly correlated with the norms of the second-stage filters (Supplementary Fig. S17). Correlating the sum of the norms of the filters with overall performance across the set of N = 7 observers yielded strong, significant positive correlations (16x16 AVG, *ridge*: r = 0.970, p < 0.001; *ridge + smooth*: r = 0.993, p < 0.001), with similar results for MAX, SUB models (Supplementary Table S8). Consistent with this, we found that the ideal observer, had a larger filter norm (*ridge*: 5.67; *ridge + smooth*: 5.37) than the human observers (*ridge*: median = 4.30, min = 2.36, max = 4.59; *ridge + smooth:* median = 4.04, min = 2.27, max = 4.34, N = 7), as revealed by a sign-rank test (*ridge:* p = 0.018; *ridge + smooth:* p = 0.018).

Experiment 2: Boundary orientation discrimination

Application of our modeling methodology to the orientation identification task reveals that observers are utilizing region-based processing. In order to see if our modeling method could reveal situations where observers utilize edge-based processing, we considered a fine orientation discrimination task (Fig. 1c, *right*). In this task, second-order boundary orientation was perturbed slightly (JND at 32% contrast – see METHODS) to the left or right of vertical and observers had to determine which direction it was perturbed. In one variant of this task (Experiment 2-VAR), contrast modulation was varied in 9 logarithmic steps from 16% to 64% centered at 32% (plus zero modulation) to make the task easier and harder. In a second variant (Experiment 2-FIX), contrast modulation was held fixed at 32%. In both variants of Experiment 2, the perturbation of orientation was fixed near JND for each observer at 32% contrast (CJD: 7 deg., JJF: 6 deg., VHB: 7 deg.). In our analysis, we used a higher down-sampling rate (22 x 22) to better reveal fine spatial details, and only considered AVG pooling, as Experiment 1 demonstrated that the spatial strategies revealed by our model are robust to pooling rule.

Fig. 11 shows the second-stage filter weights from representative observers tested in Experiment 2-VAR (Fig. 11a) and Experiment 2-FIX (Fig. 11b). We see that in contrast to Experiment 1, the observers employ an edge-based strategy, assigning substantially more weight to regions near the vertical boundary. Bootstrapped significance plots for individual weights are shown Supplementary Fig. S18. Supplementary Fig. S19 shows consistent results from other observers in these experiments. The edge-based strategy employed by these observers is very similar to that obtained from an ideal observer (Supplementary Fig. S20). As with Experiment 1, we find a reasonably strong agreement between model predictions and observed performance on sets of novel data not used for model parameter estimation (Supplementary Figs. S21, S22), and find an expansive second-stage filter non-linearity for Experiment 2 (Supplementary Table S9). Taken together with Experiment 1, Experiment 2 shows that our modeling methodology can reveal different spatial strategies that observers may employ for different kinds of psychophysical tasks.

Experiment 3: Integrating contrast modulation across orientation channels

In Experiments 1 and 2, we considered a fixed FRF model architecture and focused on characterizing the spatial strategies observers employ in different second-order vision tasks. However, one of the most powerful applications of this methodology is fitting models embodying different assumptions about FRF model architecture, and comparing their ability to account for psychophysical data. Indeed, a major focus of much of the second-order vision literature has been to design experiments whose express goal is to test competing hypotheses of FRF architecture (Motoyoshi and Kingdom, 2003; Nishida & Motoyoshi, 2004; Kingdom, Prins & Hayes, 2003; Motoyoshi and Kingdom, 2007; Westrick and Landy 2013).

We applied our method to ask the question how second-stage filters which detect contrast modulation integrate orientation modulations across multiple channels. Indeed, in many computational studies of second-order processing, the FRF models applied to account for the data make the assumption that each second-stage filter only analyzes a single orientation/spatial frequency channel (Wilson 1993; Kingdom, Keeble & Moulden 1995; Landy & Oruc, 2002; Zavtiz & Baker 2013). However, a number of psychophysical studies have suggested that second-stage filters may actually integrate contrast modulation across orientation and spatial

frequency (Motoyoshi and Kingdom, 2003; Nishida and Motoyoshi, 2004; Kingdom, Prins, & Hayes, 2003). Furthermore, many non-linear responses seen in visual neurons are believed to involve interactions between neurons with different orientation preferences (Tanaka & Okazawa, 2009; Wilmore et al., 2010; Rowenkamp & Sharpee, 2017).

To address this question systematically, we performed an orientation identification task as in Experiment 1, although in this experiment the carrier pattern contained two orthogonal orientations (Fig. 1a, bottom), which are processed by two different first-order orientation channels. In this analysis, since we were not concerned spatial integration strategy, we used a coarser down-sampling rate (12 x 12) and only considered AVG pooling. We considered two models of how information is integrated by the second-stage filters. Model 1 (Fig. 12a, top), which we refer to as the "*non-opponent*" model, assumes that each channel is analyzed by its own pair of second-stage filters (L/R oblique). Model 2 (Fig. 12a, bottom), which we refer to as the "*orientation-opponent*" model, assumes that each second-stage filter receives inputs from both first-order channels. Model 1 and model 2 actually have the exact same number of parameters, making the application of standard Bayesian model comparison methods like the Bayes Information Criterion (Schwarz, 1978; Bishop 2006) quite simple. Given a set of models indexed by an integer *i*, the Bayes Information Criterion (BIC) approximates the asymptotic *a posteriori* probability of the *i*-th model by

$$BIC_i = \ln p_i(D) - \frac{N_i}{2} \ln n, \qquad \text{(R.1)}$$

where $\ln p_i(D)$ is the log-likelihood of the data, N_i is the number of free parameters in the model, and *n* is the total number of observations. Therefore, to compare models 1 and 2 we simply compute $\Delta_{21} = BIC_2 - BIC_1$, preferring model 2 if $\Delta_{21} > 0$ and model 1 otherwise. When two models have the same number of parameters, the complexity-penalizing terms of (R.1) cancel, so that $\Delta_{21} = \ln p_2(D) - \ln p_1(D)$. That is, when the models are of equal complexity, we simply prefer the model with higher log-likelihood. An identical argument holds when using the Akaike Information Criterion (AIC; Akaike, 1974), another popular criterion for behavioral model comparison (van den Berg, Awh, and Ma, 2014). Therefore, for comparison of models having equal complexity, the BIC and AIC are equivalent.

As in Experiments 1 and 2, we find that the models are in reasonably good agreement with subject performance on the task (Supplementary Fig. S24). Likewise, we found that the fitted models also had an expansive second-order nonlinearity (Supplementary Table S10). We

see in Fig. 12b that across all three observers that the BIC shows a consistent preference for the orientation-opponent model (Model 2). Ideal observer analysis also revealed a preference for the orientation-opponent model fit to stimulus category labels (BIC_1 = -1774.7, BIC_2 = -1535.5).

Standard Bayesian model comparison methods like the BIC make use of all available data (Kass and Rafferty, 1995; Bishop 2006), as opposed to fitting a model on a sub-set of data and testing it on a different set as one does in a cross-validation approach (Wu et al., 2006). However, we also wanted to compare the predictive abilities of the two models by evaluating their agreement with observer performance on data not used for training. Therefore we performed a bootstrap analysis where we fit both models (with hyper-parameter optimized for each model, see Supplementary Figs. 23, 25) on a randomly selected sample of 4000 stimuli and tested on the remaining 1000 (50 bootstraps). We see in Fig. 12c that there is a consistent preference across observers for Model 2 (magenta line). Performing a 3 x 2 factorial ANOVA (3 observers, 2 models, 50 observations each) on the test set likelihoods revealed significant main effects of both observer ($F_{2,294}$ = 201.085, p < 0.001) and model ($F_{1,294}$ = 39.811), but no significant interaction ($F_{2,294}$ = 3.001, p = 0.051).

Discussion

Overview

Second-order stimuli represent a challenge for computational models of visual processing, as they do not contain spatial variations in stimulus intensity and cannot be detected by simple linear filtering operations (Baker, 1999; Baker and Mareschal, 2001). Consequently, standard linear 'classification image' techniques (Ahumada, 1996; Murray, 2011) are not readily applicable to studies of second-order visual processing. The present study demonstrates that second-order vision can be fruitfully studied using the same statistical machine learning framework commonly applied to first-order vision (Neri and Levi, 2006; Murray, 2011) and sensory neurophysiology (Wu et al., 2006; Willmore et al. 2010; Vintch et al. 2015; Harper et al. 2016; Antolic et al. 2016).

In this study, we define a simple hierarchical neural network implementing a Filter-Rectify-Filter (FRF) model of second-order visual processing (Bovik, Clark and Geisler, 1990; Malik and Perona, 1990; Bergen and Landy, 1991; Sutter, Beck, and Graham, 1989; Landy and Graham, 2004) and estimate its parameters directly from psychophysical trial data. We show that our FRF model can accurately model human performance on second-order vision tasks by comparing observer and model performance on novel test data not used for training the model. We then show that one can use this modeling approach to address several biologically well-motivated questions about second-order visual processing. In particular, we test whether observers segment near-threshold contrast-defined boundaries using an edge-based or region-based processing in an orientation-identification task (Experiment 1), finding support for region-based segmentation. To the best of our knowledge, our experiment is the first time the issue of edge/region-based processing has been addressed for contrast-defined boundaries as it has for other second-order boundaries (Nothdurft, 1985; Gurnsey and Laundy 1992; Wolfson and Landy, 1998). We also demonstrate that our method can reveal experimental situations where observers might employ an edge-based spatial integration strategy (Experiment 2). Finally, we show that our general methodology can be used to adjudicate between competing models of FRF model architecture using Bayesian model comparison methods (Experiment 3), and suggest that our approach can be used to generate critical stimuli optimized to distinguish between the many

competing FRF models in current the literature. Here we discuss the limitations of the present study, relate it to previous work on computational models of vision, and outline directions for future research.

Limitations of the current study

One critical limitation to psychophysical system identification studies is the relatively limited amount of data that can be collected in an experiment relative to the number of model parameters one needs to estimate. Although one can alleviate the demand for vast amounts of data using reasonable prior constraints (Knobloch and Maloney, 2008a, b; Mineault et al., 2009), it is often necessary to fit simplified models which omit many known biological details. Here we discuss several of the modeling simplifications used in the current study which may need to be addressed when extending this methodology to more complicated second-order stimuli and tasks.

First-stage filters

In FRF models like those considered here, the initial stage of filtering and rectification is meant to model computations believed to be taking place in visual area V1 (Mareschal and Baker, 2001). In the present study, we modeled the first-stage filters using a standard energy model of V1 complex cells with normalization (Adelson and Bergen, 1985; Heeger 1992). Although for our stimuli a front-end comprised entirely of complex cells was perfectly adequate, for many texture segmentation tasks it may be necessary to model simple cells as well. For instance, it is well known that one can segment second-order images using cues of carrier contrast phase (Kingdom, Hayes, and Field, 2001) and polarity (Rentschler, Hubner, and Caelli, 1988; Malik and Perona, 1990; Hansen & Hess, 2006) , which is not possible using only complex cells, since their responses are invariant to phase and polarity (Mechler and Ringach, 2002). Therefore, an extension to our set of first-stage filters which would make the model appropriate for more generic texture segmentation tasks should include models of simple cells (both even and odd phase) to permit polarity-based segmentation. Likewise, it may also be desirable to include center-surround “difference-of-Gaussian” (DOG) filters as well, since recent studies have suggested that non-oriented filters are present in area V1 (Ringach, Shapley and Hawken, 2002),

and these kinds of filters may potentially be serve as a mechanism for polarity-based segmentation (Motoyoshi and Kingdom, 2007).

In Experiments 1 and 2, we implemented only a single spatial frequency and orientation channel matched to the carrier, and in Experiment 3 we only considered a single spatial frequency channel and two orientation channels. This simplification neglects the possible contributions of off-orientation and off-spatial frequency channel to task performance. Indeed, it has been shown previously that for some second-order vision tasks like detecting modulations of orientation or spatial frequency that off-center channels may be the ones most useful for task performance (Prins and Kingdom, 2002). Likewise, several other studies have challenged the notion that contrast-modulation tuned mechanisms are narrowly tuned for a single first-order channel, with many studies suggesting broad pooling over carrier spatial frequency (Prins and Kingdom 2002; Graham, Sutter and Venkatesan 1993; Westrick and Landy, 2013). Therefore, it is of future interest to incorporate a more realistic complement of first-stage orientation and spatial frequency channels (as in Fig. 13) to investigate the possible contributions of off-center channels to our tasks.

Another simplification in the design of our first-stage filters was the fact that we did not include many well-known processing nonlinearities present in visual neurons like cross-orientation suppression (Bishop, Coombs, and Henry, 1973; Morrone, Burr, & Maffei, 1982; Kneieim and van Essen, 1991), co-linear facilitation (Kapadia, Westheimer, and Gilbert, 2000) or surround suppression (Walker, Ohzawa, Freeman 1999; Tanka and Ohzawa, 2009). These processing nonlinearities may play a role in detecting second-order boundaries (Kastner, Nothdurft and Pigarev, 1997; Nothdurft, Gallant, and van Essen, 1999, 2000; Tanka and Ohzawa, 2009; Willmore et al., 2010; Hallum and Movshon, 2014), and some of these nonlinearities have been incorporated previously into various FRF models of texture segmentation. Therefore, it is of interest for future work to incorporate additional nonlinearities into the first-stage filters.

Second-stage filters

The main focus of our study was demonstrating that one could learn the properties (shapes and non-linearities) of a set of second-stage filters which detect second-order visual features like

contrast modulation. For our particular task of contrast-boundary detection, the simplest possible model involves two second-stage filters representing the two potential stimulus categories, much like the model shown in Fig. 2. However, there is no reason that our model needs to contain only two second-stage filters. For psychophysical tasks involving more complex second-order stimuli defined by multiple cues, it may be necessary to have several second-stage filters whose responses are combined to obtain a perceptual decision. Indeed, natural boundaries typically contain both first-order and second-order information (Schofield, 2000; Martin, Fowlkes and Malik, 2004; Johnson and Baker, 2004, 2005), all which are integrated in perceptual tasks (Rivest and Cavanaugh, 1996; Landy and Komija 2001; McDermott, 2004; Johnson, Prins and Baker, 2007; Saarela and Landy, 2012; DiMattina, Fox and Lewicki, 2012). Furthermore, previous psychophysical studies strongly suggest that there is unlikely to be a single general-purpose FRF mechanism for all second-order boundaries, but rather a number of different FRF mechanisms specialized for different kinds of second-order boundaries (Kingdom, Prins and Hayes, 2003; Motoyoshi and Nishida, 2004; Motoyoshi and Kingdom, 2007). Future work with natural boundaries should incorporate many second-stage filters (Fig. 13) which can learn to specialize for the various kinds of second-order information present at a boundary, with Bayesian model selection techniques (Schwarz, 1978; Bishop, 2006) employed to determine the appropriate level of network complexity.

An additional limitation of our model (and many other FRF models) is the assumption that there is only a single stage of second-order filtering, and that the outputs of the second-stage filters are combined to generate the perceptual decision. However, recent neurophysiological and fMRI studies have suggested that texture selectivity may develop gradually as one moves up the visual hierarchy from V1 to V2 to V4 (Freeman et al. 2013; Goda et. al. 2014; Okazawa, Tajima, and Komatsu, 2015, 2016; Ziemba et al. 2016; Kohler et al. 2016). This gradual development of selectivity, in particular the increase in selectivity between areas V2 and V4 (Okazawa et al., 2016), suggests that the idea of a single layer of ‘second-stage’ filters may represent an over-simplification of the underlying biology. Interestingly, in various studies using convolutional neural networks to solve image recognition tasks, one often sees texture-sensitive units develop gradually over several processing layers (Yamins et al. 2014; Kahligh and Kriegeskorte, 2014; Zeiler and Fergus, 2014; Yamins and DiCarlo, 2016). Therefore, adding of another layer of

processing to the model shown in Fig. 2 would permit us to learn more complex interactions between second-stage filters.

Comparison with previous work

Despite the numerous simplifications made in the present study, it represents a substantial methodological advance over many previous efforts to model second-order visual perception. Although previous studies have developed image-computable FRF models, these models have generally fixed not only the model architecture, but also the shapes of the second-stage filters. However, the appropriate shape of the second-stage filters for a given task is in fact an empirical question. The present study demonstrates how the second-stage filter shapes can be learned directly from experimental data in order to reveal the spatial integration strategies used in different tasks (Experiments 1 and 2).

Another major difference with previous work is that in most previous studies FRF model free parameters are determined by fitting threshold data (e.g. Zavitz and Baker, 2013). Since thresholds are typically determined at only a few levels of a 1-dimensional independent variable, this gives a small number of data points for curve-fitting and hence greatly limits the number of free FRF model parameters one can estimate from psychophysical data. In our work, we directly estimate hundreds of model parameters by fitting stimulus-response data from thousands of psychophysical trials. This enabled us to characterize the shapes of the second-stage filters used in our task, rather than simply assuming their shapes as in past work. Although the present study focused on learning second-order filter shapes (Experiments 1 and 2), if one is able to employ a low-dimensional parameterization of the filter shapes (Lewi et al. 2009b; DiMattina and Zhang, 2011) one can drastically increase the number of second-stage filters and fit much more architecturally complex FRF models like the model shown in Fig. 13. We return to this point below in greater detail.

In addition to extending the FRF modeling literature, our work also complements and extends previous classification image studies, which have generally been applied in the context of first-order vision (Neri and Levi, 2006; Murray, 2011). Although classification images yield image-computable predictive models of observer responses, relatively few studies validate these models by testing their predictive performance on a set of stimuli not used for model estimation

(Neri, 2010). We tested model performance on novel stimuli not used for model estimation using both standard measures like percent correct, as well as more advanced methods like decision variable correlation (Sebastian and Geisler, 2018). We followed previous classification image studies in performing double-pass analysis (Neri, 2010; Hasan, Joosten and Neri 2012), but extended this analysis by using decision variable correlation in addition to measures of proportion agreement used in previous work.

Finally, in contrast to classification image studies, here we systematically address the issue of model comparison. Psychophysical system identification can be formulated in the framework of estimating a generalized linear model (Knobloch and Maloney, 2008a,b; Mineualt et al., 2009), which is unique up to choice of link function. Therefore, issues of model comparison (beyond coefficient shrinkage) do not naturally arise in this context. In contrast, the FRF models employed in studies of second-order vision vary greatly in their architectural details (number of filtering layers, nonlinearities, connectivity, etc…), and hence model comparison is of great interest in this literature. We compare two competing FRF architectures making different assumptions about how the second-stage filters integrate across orientation channels, finding a preference for a model with early integration by the second-stage filters, consistent with previous psychophyscial work (Kingdom, Prins and Hayes, 2003; Prins and Kingdom, 2003; Motoyoshi and Nishida, 2004).

Edge versus region-based segmentation

Although the main focus of the current study is methodological, we applied our methodology to the biologically interesting question of whether or not observers take a region-based approach or an edge-based approach to segmenting a contrast boundary (Experiment 1). Although this issue has been addressed in previous psychophysical work for boundaries defined by orientation (Nothdurft 1985; Bergen and Landy, 1991; Wolfson and Landy, 1998) or complex textures (Frisby, 1980; Gurnsey and Laundry, 1992), to the best of our knowledge it has not been addressed for contrast-defined boundaries. This question is of great interest in light of the fact that numerous studies have revealed possible neurophysiological mechanisms for contrast-boundary detection (Mareschal and Baker 1998; Rosenberg and Issa, 2010; Li and Baker 2012; Li et al. 2014; Gharat and Baker, 2016). This body of work has suggested a population of

contrast-modulation (CM) tuned neurons in extra-striate cortex which exhibits band-pass tuning to both first-order carrier frequency as well as second-order modulation envelope frequency. Such CM tuning behavior is consistent with a Gabor-shaped second-stage filter applied to the outputs of a bank of first-stage filters selectively tuned to the carrier pattern (Baker, 1999).

If CM neurons like those described in the ventral visual pathway were the sole mechanism responsible for psychophysical detection of CM boundaries, one might expect the 1-D weighting profiles to have a Gabor profile, placing greater weight on regions near the boundary. In our investigations, we find that this was not the case and that all areas of the image are weighted equally for the orientation-identification task (Experiment 1, Fig. 3). Furthermore, in additional control manipulation where we increased the carrier frequency in order to make sure CM neurons with smaller RFs would be optimally engaged (since over the population the preferred carrier : envelope ratio is 8:1), we still did not see any change in processing strategy (Experiment 1-HFC, Fig. 5c). These findings are consistent with at least two distinct possibilities for the underlying neural mechanisms. The first possibility is that the task is performed using CM neurons (in V2 or beyond) with large receptive fields (> 4 dva) encompassing the entire stimulus. Such neurons would prefer a carrier frequency much higher than their envelope frequency, which is not at all implausible since previous work in primate V2 has shown CM neurons can have preferred carrier:envelope ratios of up to 40:1 (Li et al., 2014). The second possibility is that the task is performed not only with the population of CM neurons, but also by neurons sensitive to texture (‘texture-units’), with the responses of such units being modulated by texture contrast. Previous physiological studies of the ventral visual stream have demonstrated the presence of texture-tuned units (Hanazawa and Komatsu, 2001; Okazawa et al., 2015, 2016; Ziemba et al., 2016), and such units have also been implied to exist by computational studies of higher-order image statistics (Karlkin and Lewicki, 2003, 2009; Caudieu and Olshausen 2012) and studies of deep neural networks (Yasmins et al., 2014; Khalig-Razavi and Kriegeskorte, 2014; Zeiler and Fergus, 2014).

The general issue of whether observers employ region-based or edge-based processing of second-order boundaries has been investigated in numerous studies and remains a question of great interest. Clearly, the fact that we can readily distinguish non-juxtaposed textures implies that there must be mechanisms for texture representation that operate in the absence of a boundary. However, in some cases texture segmentation can certainly benefit from boundary

information, as in the textural Craik-Cornsweet illusion (Nothdurft, 1985). The examples in the literature where there is the strongest evidence for a contribution of the boundary comes from studies of orientation-defined textures (Nothdurft, 1985; Kingdom and Keeble, 1996; Wolfson and Landy 1998). In one of these studies (Wolfson and Landy, 1998), one can observe that direct juxtaposition of textures having different orientations gives rise to a novel visual cue of a junction or corner, and it is known that there are neurons in the ventral visual stream which are selective for cues (Hegde and Van Essen, 2000; Anzai, Peng and Van Essen, 2007).

However, other work on texture segmentation has shown a minimal contribution of information on the boundary. A study using natural image patches in a task of discriminating occlusions from textures found that covering up the boundary had little effect on performance (DiMattina, Fox and Lewicki, 2012). However, performance on natural occlusion edge and junction detection tasks can be seriously degraded when textural information, which is by its nature defined over a region, is removed (McDermott 2004; DiMattina, Fox and Lewicki, 2012). Similarly, work on segmenting complex textures defined by letters shows little effect of boundary information (Gurnsey and Laundry, 1992). Even for orientation-defined textures in many cases observers are taking a region-based strategy, for instance when texture difference is defined by the variance rather than the mean micropattern orientation (Wolfson and Landy, 1998).

In general, the answer to this question is most likely highly dependent on the particular stimulus and the nature of the psychophysical task. In our view, the question of “edge” versus “region”-based processing is somewhat of a false dichotomy, since clearly for a very large stimulus extending into the far periphery there will be reduced weighting at the largest eccentricities, for the entirely trivial reason that visual acuity falls off sharply outside of the fovea. The real question of interest is the exact shape of the fall-off in weight as a function of stimulus eccentricity, and our method is well suited to address this issue for many different kinds of second-order boundaries, much like classification images can be used to address this for first-order boundaries.

In the current paper on second-order boundaries, in Experiment 1 every part of the image is useful for performing the task, and hence the ideal observer will integrate information from all image regions. In Experiment 2, the only useful information for performing the task is near the boundary and so an ideal observer will be indifferent to the stimulus elsewhere. In terms of

possible mechanisms, we might expect CM neurons with larger receptive fields to be engaged in Experiment 1, whereas the most useful neurons in Experiment 2 would be CM neurons with smaller receptive fields. This is because CM neurons with large (small) receptive fields would be the ones whose responses best distinguish the two stimulus categories in Experiment 1 (2). It is of great interest for future work to better understand the potential neural bases of the observer strategies employed in different edge perception tasks.

Integrating over first-order channels

A common ingredient to the wide variety of FRF models proposed to account for psychophysical data is the analysis of a set of first-stage filters by one or more second-stage filters defined on a much larger spatial scale. However, one aspect in which models differ is whether or not the second stage filters analyze first stage filters at a single orientation/spatial frequency, or whether there is integration across first-order channels. In one 'standard' model of FRF processing, each second-stage filter only analyzes the outputs of a single first-stage orientation/spatial frequency channel (Malik and Perona, 1990; Wilson, 1993; Landy and Oruc, 2002). However, a number of psychophysical results have called this model into question. One study using a sub-threshold summation paradigm find support for a model where there is linear integration across orientation for detecting contrast modulations (Motoyoshi and Nishida, 2004). Another study making use of energy-frequency analysis reveals the existence of orientation-opponent mechanisms in second-order vision (Motoyoshi and Kingdom, 2003). Finally, other work has suggested at least partly independent mechanisms for detecting second-order modulations defined by contrast and spatial/frequency orientation, with the suggestion that contrast modulations may be pooled across first-order channels (Kingdom, Prins and Hayes 2003). In the present paper, in Experiment 3 we find support for a model that integrates across orientation channels, consistent with previous psychophysical studies, and physiological studies demonstrating interaction of information across orientations in visual neurons (Kastner, Nothdurft and Pigarev, 1997; Nothdurft et al., 1999, 2000; Tanka and Ohzawa, 2009; Willmore et al., 2010).

Future directions

Adaptive stimulus generation for FRF model estimation and comparison

Our study demonstrates that formulating second-order vision as a machine learning problem can quickly give rise to models of great complexity which must be estimated using limited amounts of psychophysical data. This is very much analogous to the problem faced in sensory neurophysiology, where complex models of nonlinear sensory neuron responses must be estimated from limited spike data (Wu et al., 2006). In recent years, a number of investigators have proposed computational methods for efficient data collection in neurophysiology and psychophysics experiments making use of adaptive stimulus generation (Watson and Pelli, 1983; Kontsevich and Tyler, 1999; Lewi, Butera and Paninski, 2009a; DiMattina and Zhang, 2011, 2013; Lesmes et al. 2010; DiMattina, 2015; Pillow and Park, 2016). These methods can drastically reduce the number of observations needed to estimate a model to a desired level of accuracy by focusing observations where they are most useful for parameter estimation, as measured by various mathematical criteria like the mutual information (Lindley, 1956; Paninski, 2005). Although in the current experiment we utilized non-adaptive procedures, for future work it is of great interest to apply adaptive stimulus generation to estimate our FRF models. In fact, given the difficult machine learning problem which one faces when modeling psychophysical data with high-dimensional stimuli (Mineault et al., 2009), adaptive stimulus generation may in fact be necessary for accurate estimation of complex FRF models containing a large number of second-stage filters (Fig. 13).

Another application of adaptive stimulus generation methods which may be immensely useful in future extensions of this technique is the ability to adaptively optimize stimuli for model comparison. The current literature in second-order vision includes a large number of variations on the basic FRF model which have been proposed to account for new findings. What would be helpful is a systematic method for performing experiments which quantitatively compare the predictions of competing models. A growing body of literature in brain and cognitive sciences has addressed this problem by proposing algorithms for adaptively generating stimuli for purposes of model comparison (Cavagnaro et al. 2010; DiMattina and Zhang, 2011,

2013; Myung, Cavagnaro, and Pitt, 2013). Such methods have been shown in both experimental and computational work to greatly increase the efficiency of experiments for distinguishing competing models in a variety of problem domains (Wang and Simoncelli, 2008; Cavagnaro, Pitt and Myung, 2011; Cavagnaro et al., 2013; DiMattina, 2016). Given the wide variety of competing ideas about FRF model architecture which have been proposed in the literature, second-order vision certainly seems like a problem domain which would be well suited to the application of these model comparison methods.

Application to natural boundaries

Pure texture boundaries, let alone contrast boundaries, are very rare in the natural visual world. The vast majority of naturally occurring region boundaries (for instance, those defined by occlusions) contain a wide variety of both first-order and second-order cues, all of which are integrated in visual perception (Rivest and Cavanaugh, 1996; Landy and Komija 2001; McDermott, 2004; Johnson, Prins and Baker, 2007; Saarela and Landy, 2012; DiMattina, Fox and Lewicki, 2012). Therefore, it is our ultimate goal to develop a quantitative model which can predict observer performance on tasks like those considered here but using arbitrary natural boundaries.

Several challenges arise in attempting to define and estimate an FRF model which can predict segmentation performance for arbitrary natural boundaries. One major challenge is that unlike luminance and color, 'texture' is an ill-defined concept which can include differences in a number great number of perceptual attributes, including orientation, spatial frequency, contrast, and density (Landy and Graham, 2004; Rosenholtz 2014). Therefore, it is likely that a large number of second-stage filters may be necessary to detect all of the various possible combinations of the features that define natural texture boundaries. Assuming that the first-order filters span several spatial scales and orientations, and we have many second-stage units, we end up having to estimate a very large number of connections between the first-stage filters and second-stage filters from a limited amount of psychophysical data. Concretely, suppose that we are using 128x128 images analyzed at spatial scales 4 spatial scales (4, 8, 16, 32 pixels) with 8 orientations (steps of 22.5 degrees). This gives us 32 banks of first-stage filters. If each first-stage filter bank is down-sampled to produce a 10x10 image, we end up with a total of 3200

first-stage filter outputs. Conservatively assuming only 20 pairs of second-stage filters (assuming symmetry), we still end up with 64,000 parameters in our model. Such a model is far to complex to be tractable, since the upper limit of psychophysical observations which a single observer can provide is on the order of 10,000 trials.

If we can make the simplifying assumption that each filter bank is read out using a region-based strategy, as suggested by our results in Experiment 1, then can simplify the problem somewhat and characterize the outputs of each first-stage filter bank by four numbers. We simply measure the average activity in the upper left and lower-right diagonal regions (for detecting left-oblique boundaries) and that in the upper right and lower-left diagonal regions (for detecting right-oblique boundaries). Therefore, our first layer now has 32*4 = 128 first-stage filter outputs. Therefore, 20 pairs of second-stage filters translate to a much more reasonable 2560 parameters, which is on the order of complexity of the models estimated here. Application of a Bayesian prior (Bishop, 2006; Knoblauch and Maloney, 2008a,b; Mineault et al. 2009) incorporating additional assumptions like sparse connections between the first-stage and second-stage filters can make estimation of this model tractable with 10,000 or fewer psychophysical observations, especially if we incorporate adaptive stimulus generation methods (e.g. DiMattina, 2015) to use our trials as efficiently as possible

Conclusions

Statistical machine learning has in recent years become a powerful and widely applied methodology in vision science and computational neuroscience. Our work demonstrates that the same machine learning methodology used to characterize neural coding (Wu et al., 2006; Willmore et al., 2010; Kriegskorte, 2015; Yamins and DiCarlo, 2016; Antolik et al., 2016) and first-order visual processing (Mineault et al. 2009; Murray, 2011) can be fruitfully applied to extend psychophysical system identification approaches to complex visual stimuli, including second-order boundaries. We anticipate that future applications of this methodology may lead to a better understanding of how multiple first- and second-order cues are combined by human observers to detect natural boundaries, and hopefully provide insights into neural and computational mechanisms of image region segmentation.

Methods

Experimental Paradigms

Stimuli and general task

We created contrast-modulated texture boundaries by multiplying a low spatial frequency contrast modulation envelope by a high spatial frequency carrier pattern (Fig. 1a, b). The carrier patterns were produced by adding odd-phase Gabor micropatterns at random locations in a large image. By varying the properties of the Gabor micropatterns like their orientation, spatial frequency, and placement density, we can create a wide variety of carrier textures. To produce contrast-modulated stimuli, carrier textures were multiplied by a circular raise cosine window modulated by a half-disc oriented either right oblique (clockwise) or left-oblique (counter-clockwise). The root-mean-square (RMS) contrast of the carrier patterns was set to 14%, which is 6 dB above previously observed average carrier detection thresholds (Arsenault, Yonessi and Baker, 2011). Examples of right-oblique boundaries with varying carrier mircopattern densities and depths of contrast modulation are shown in Fig. 1b. These micropattern stimuli have been used in previous studies (Zavitz and Baker, 2013, 2014) and permit us to produce a wide variety of naturalistic texture patterns amenable to parametric variation.

In all experiments, observers performed a single-interval 2AFC task to determine if a briefly displayed (100 msec) boundary stimulus was contrast-modulated in a right-oblique (clockwise) or left-oblique (counter-clockwise) orientation. Visual feedback was provided after each trial to maximize observer performance. Data was collected over multiple (4-5) sessions conducted over several days, with each session lasting from 30 minutes to 1 hour. The initial session was a training session dedicated to learning the orientation identification task (+/- 45 degree) by performing a short version (300 trials) at various micro-pattern densities (256, 1024, 2048 patterns or 512, 1024, 2048 patterns). Consistent with previous studies on second-order edge detection (Zavitz and Baker, 2013, 2014), we found in preliminary experiments that performance on the task was best when the carrier texture pattern had a high density of micropatterns. Therefore, we used relatively high micropattern densities (2048 per 256x256

image for Experiments 1,2 and 512 for Experiment 3), higher than in previous studies (Zavitz and Baker, 2013), with the goal of optimizing observer performance.

Psychophysical observers

Observers were author CD and N = 7 members of his research group who were initially naïve to the purposes of the experiments. For Experiments 1 and 2, we also recruited N = 5 additional observers from two Psychology classes (EXP-3202, PSB-4002) at Florida Gulf Coast University (FGCU). All observers provided written informed consent and all procedures were approved beforehand by the IRB at FGCU (Protocol 2014-01) and McGill University (Protocol MED-B-96-279), in accordance with the Declaration of Helsinki.

Displays and experimental control

At FGCU, all experiments were run in one of two windowless vision psychophysics chambers with no light sources other than the computer. Chamber 1 contained a factory-calibrated (gamma=1.0) Cambridge Research Systems (CRS) Display++ LCD having a 120Hz refresh rate and 100 cd/m^2 gray level. Chamber 2 contained a gamma-corrected (gamma = 1.0) SONY FD-500 Trinitron CRT monitor (75 Hz, gray level = 59.2 cd/m^2) driven by CRS Bits# stimulus processor. Viewing distances were 134 cm in Chamber 1 and 132 cm in Chamber 2 (using a chin-rest for observer comfort), so that the stimulus image subtended approximately 4 dva for 256 x 256 stimuli. At McGill University, data was collected from author CJD for Experiments 2 and 3. Stimuli were displayed on a SONY Multiscan G400 CRT with hardware gamma correction (gamma = 1.0) and 53 cd/m^2 gray level. Viewing distance was 125 cm so that the stimulus subtended 4 dva, and experiments were conducted in a laboratory room with normal lighting. We observed remarkable consistency in our results despite differences in monitor characteristics and room lighting conditions, so all data was pooled for analysis. Experiments were controlled using custom-authored software written in the MATLAB® programming language (MathWorks®, Inc.), making use of routines from the PsychToolbox® Version 3.0.12 (Kleiner, Brainard & Pelli, 2007).

Experiment 1: Orientation identification

The stimuli for Experiment 1 were contrast-modulated discs shown in Fig. 1b. Stimuli were 256 x 256 pixels in size with a carrier pattern comprised of 2048 randomly located, vertically oriented (0 degrees) 32-pixel Gabor micro-pattern stimuli. This carrier was modulated by an obliquely oriented half-disc, graduated by a cosine taper in order to avoid introduction of first-order artifacts. In this task, the two stimulus alternatives had very different boundary orientations, with the right/left-oblique boundaries being oriented +/- 45 deg. with respect to vertical.

In the main version of Experiment 1, for N = 7 observers contrast modulation depths were varied 9 logarithmically spaced levels from 8% to 32%, centered at 16%, plus 0% modulation (Experiment 1-VAR). This set of levels did a good job of focusing trials at or near threshold for most observers, while still including some relatively easy trials as well as catch trials with 0% modulation. After an initial training session, these observers performed N = 4000 trials over the course of multiple sessions spaced over several days. For one observer (author CJD), 2975 patterns were used in this session. For an additional N = 3 observers (CJD and naïve observers MAK and JJF), contrast modulation depths were fixed for each observer at a value slightly above that subject's threshold (Experiment 1-FIX), with the goal of obtaining near-threshold task performance of approximately 80% correct (Supplementary Fig. S13). This is more typical of most classification image literature (Murray, 2011), which often fixes stimulus intensities at or near task performance thresholds. These observers performed N = 6000 trials over multiple sessions spaced over several days. To speed up experiments, carrier patterns were pre-computed and stored at 8 bit precision on the experimental computer. A unique set of pre-computed carrier patterns was used for each stimulus paradigm, so that the observer never saw the same carrier pattern more than once.

Experiment 1-HFC: Effects of carrier-to-envelope spatial frequency ratio

As an additional control for Experiment 1, we modified Experiment 1-VAR so that the carrier pattern had a much higher spatial frequency. Experiment 1-HFC was run on three naïve observers (LJW, JMM, VHB) tested at variable contrasts (same levels as Experiment 2-VAR).

We used 256 x 256 images with carrier texture comprised of vertical (0 degree) 8-pixel micro-patterns at a higher density (8912 patterns). The purpose of running this variation was to test whether or not changing the ratio of envelope to carrier spatial frequency (by using smaller micro-patterns) would affect the perceptual strategy used to solve the task. This was motivated by recent physiological findings (Li et al., 2014) demonstrating that the population of second-order neurons in V2 have an optimal carrier-to-envelope spatial frequency ratio of about 8:1. Therefore, using a higher spatial frequency carrier would potentially activate more second-order neurons whose receptive fields lie near the boundary, potentially leading to edge-based processing of our stimuli.

Experiment 2: Fine orientation discrimination

In order to show the general applicability of our methodology to tasks where observers may employ an edge-based strategy, we ran a fine orientation discrimination task where observers had to determine whether a clearly visible contrast-modulated boundary (same carrier parameters as in Experiment 1) was tilted slightly to the left or to the right of vertical. Experiment 2 is complementary to Experiment 1 since in Experiment 1 the orientation difference is very large (+/- 45 degrees from vertical), while the contrast modulation is near detection threshold, whereas in Experiment 2 the contrast modulation is well above detection threshold, but the orientation difference is very small. This paradigm was run on N = 2 observers, one of whom (JJF) was naïve to the purpose of the experiment. In a preliminary experiment at 32% contrast, orientation discrimination JNDs (approximately 75% performance) were determined for each observer (JJF: 6 deg., CD: 7 deg.). Then, in subsequent sessions, observers performed the fine orientation discrimination at this JND for a set of stimuli whose contrast modulation was varied from 16% to 64% (center = 32%). Observers performed N = 6000 trials of the task over multiple sessions. We refer to this experiment as Experiment 2-VAR. For two naïve observers (VHB, JJF) we also ran another variant of Experiment 2 (Experiment 2-FIX) where the contrast modulation was fixed at 32% and the orientation difference was fixed at JND for each observer.

Experiment 3: Orientation identification with multi-species carrier

Experiment 3 was the same as Experiment 1-VAR except using a multi-species carrier containing two orientations as in Fig. 1a (bottom). For this experiment we elongated the micropatterns (2:1 aspect ratio instead of 1:1 as in Experiments 1, 2) and reduced the density (512 micropatterns instead of 2048) in order to create a stronger percept of oriented structure.

Modeling approach

Modeling framework

Standard psychophysical SI methods assume that the decision variable arises from the dot product of the sensory stimulus with a set of linear weights (Ahumada, 2002; Murray, 2011), essentially modeling the entire chain of brain processes involved in stimulus-processing and decision-making by a single-layer perceptron (Rosenblatt, 1962). To permit the development of more complex models, we propose an approach to perceptual modeling which we refer to as *psychophysical neural network modeling*. In this approach, a decision variable is modeled as a function of the stimulus-dependent activities of a set of psychophysical subunits (hidden units), each of which may be sensitive to different aspects of the stimulus or observe activity in different sensory channels. This idea closely parallels recent work modeling neural responses by combining the outputs of multiple sub-units selective for different stimulus features (Willmore, Prenger and Gallant, 2010; Vintch, Simoncelli and Movshon, 2015; Harper et al. 2016; Rowenkamp and Sharpee, 2017).

Given a psychophysical classification task for a stimulus $\mathbf{x} \in \mathbb{R}^d$ with two behavioral alternatives ($b = 1, b = -1$), one may define a generic psychophysical model as

$$p(b = 1|\gamma, \boldsymbol{\theta}, \mathbf{x}) = \frac{\gamma}{2} + (1 - \gamma)\sigma\big(u(\mathbf{x}, \boldsymbol{\theta})\big), \tag{M.1}$$

$$p(b = -1) = 1 - p(b = 1), \tag{M.2}$$

where σ is a psychometric function, $u(\mathbf{x}, \boldsymbol{\theta})$ denotes the stimulus-dependent input to the psychometric function whose parameters $\boldsymbol{\theta}$ characterize the sensory processing model, and γ denotes the lapse rate. The stimulus inputs $\mathbf{x}$ may be either the stimulus parameterization itself,

or a statically transformed version of the stimulus parameterization ($\mathbf{x} = \phi(\mathbf{I})$). We define a *psychophysical neural network model* by making the additional assumption that the input $u(\mathbf{x}, \boldsymbol{\theta})$ to the psychometric function is some function of the activities $s_1, \cdots, s_K$ of a set of K psychophysical subunits, i.e.

$$u(\mathbf{x}, \boldsymbol{\theta}) = F(s_1, \cdots, s_K, \mathbf{v}), \tag{M.3}$$

where $\mathbf{v}$ is a set of parameters characterizing the cue combination rule. Perhaps the simplest combination rule is linear combination, i.e.

$$F(s_1, \cdots, s_K, \mathbf{v}) = v_0 + \sum_{i=1}^{K} v_i\, s_i, \tag{M.4}$$

although other rules are certainly possible. We assume that the subunit activities are a non-linear function of the projection of the stimulus $\mathbf{x}$ onto some filter $\mathbf{w}_i$,

$$s_i = h\left(\mathbf{w}_i^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}\right), \tag{M.5}$$

where $\boldsymbol{\alpha}$ parameterizes the shape of the nonlinearity. Combining Eq. (1), (4) and (5) yields a final expression for a psychophysical subunit model with linear cue combination

$$p(b = 1|\gamma, \boldsymbol{\theta}, \mathbf{x}) = \frac{\gamma}{2} + (1 - \gamma)\sigma\left(v_0 + \sum_{i=1}^{K} v_i\, h\left(\mathbf{w}_i^{\mathrm{T}}\mathbf{x}, \alpha\right)\right). \tag{M.6}$$

We may simplify notation by combining all model parameters and the lapse rate into the single vector $\boldsymbol{\eta}$.

The model in (M.6) can be easily reduced to various standard models used in machine learning. For instance, one recovers standard binary logistic regression (Bishop, 2006) in the special case of $\gamma = 0$ and a single subunit whose nonlinearity is a simple linear pass-through ($h(u) = u$) with fixed output weight $v_1 = 1$, so that (M.6) reduces to

$$p(b = 1|\boldsymbol{\theta}, \mathbf{x}) = \sigma\left(v_0 + \mathbf{w}_i^{\mathrm{T}}\mathbf{x}\right). \tag{M.7}$$

Likewise, with $\gamma = 0$ and the sigmoidal subunit non-linearity $h(u) = 1/(1 + e^{-u})$, this model (6) becomes a standard three layer connectionist neural network (Rumelhart, Hinton, & McClelland, 1986; Bishop, 2006).

A model of the form (M.6) arises naturally where there are two subunits representing the two stimulus alternatives whose activities are compared to make a decision. Subtracting the subunit activities $s_1 - s_2$ and adding a bias term b and Gaussian internal noise $\varepsilon \sim N(0, \rho^2)$ models the proportion of times alternative 1 is chosen (Zavitz and Baker, 2013). One obtains the expression

$$p(b = 1) = \Phi\left(\frac{\mu}{\rho}\right), \tag{M.8}$$

where Φ is the cumulative unit normal distribution and $\mu = s_1 - s_2 + b$. Using the approximation $\Phi(u) \approx \sigma(ku)$, where σ is the standard sigmoid (good fits are obtained for $k \approx 1.6 - 1.7$), we can re-write (M.8) as

$$p(b = 1) = \sigma\big(\beta(s_1 - s_2 + b)\big), \tag{M.9}$$

where $\beta = \frac{k}{\rho}$ is a term inversely proportional to the noise which also absorbs k. A model containing this term β optimized to match subject performance was implemented in one condition (SOP, or "stochastic optimized") of Experiment 1-FIX (Supplementary Fig. S14).

Filter-rectify filter model

A psychophysical network model for detecting whether a second-order edge is oriented in a left-oblique or right-oblique direction is illustrated in Fig. 2. In this model, the inputs $\mathbf{x}$ are the down-sampled outputs of a bank of first-stage energy filters resembling V1 complex cells applied to the original image $\mathbf{I}$, a transformation which we represent as $\mathbf{x} = \phi(\mathbf{I})$. There are two second-stage filters which serve as the hidden units. One subunit (L) learns a set of filter weights (blue lines) to detect a left-oblique edges from the pattern of activity in the first-stage subunits, while the other subunit (R) learns a set of filter weights to detect a right-oblique edge. The filter outputs are passed through a rectified power-law nonlinearity $h(u) = |u|^{\alpha}$ (blue curve) to generate the subunit activities s_L, s_R. These subunit activities are combined assuming fixed output weights $v_R = 1, v_L = -1$, a bias term v_0 is added, and this sum is multiplied by a noise parameter β to obtain $u = \beta(s_R - s_L + v_0)$. The fixed output weights are a consequence of our choice of a power-law non-linearity, since for all inputs $\mathbf{x}$, filters $\mathbf{w}$ and output weights v, the term $v|\mathbf{w}^{\mathrm{T}}\mathbf{x}|^{\alpha}$ is identical for all $v, \mathbf{w}$ satisfying $v\|\mathbf{w}\|^{\alpha} = C$, where C is some constant and $\|\mathbf{w}\|$ denotes the norm of $\mathbf{w}$. The general issue of parameter identification for machine learning using neural network models is discussed in depth elsewhere (Ran and Hu, 2017; DiMattina and Zhang, 2010). Many variants on this basic FRF model architecture shown in Fig. 2 are possible, and fitting multiple models to a given dataset permit us to investigate questions about FRF model architecture (Motoyoshi and Nishida, 2004; Motoyoshi and Kingdom, 2007; Westrick and Landy, 2013).

Model estimation and regularization

Given psychophysical stimulus-response data $D = \{\mathbf{x}_i, b_i\}_{i=1}^{n}$, and prior constraints $p_0(\boldsymbol{\eta})$ (possibly uninformative) on the model parameters, the machine learning problem is to find the maximum a posteriori (MAP) estimate of $\boldsymbol{\eta}$, given by

$$\hat{\boldsymbol{\eta}} = \underset{\boldsymbol{\eta}}{\operatorname{argmax}} \ln p(\boldsymbol{\eta}|D) = \ln p(D|\boldsymbol{\eta}) + \ln p_0(\boldsymbol{\eta}). \tag{M.10}$$

In this framework, the prior constraints $p_0(\boldsymbol{\eta})$ are often used to prevent over-fitting (Bishop, 2006), which is often a serious problem in classification image studies due to the relatively large number of parameters (hundreds or thousands) which need to be estimated from a relatively small number (generally < 10K) of psychophysical trials. Various techniques have been applied to help reduce the problems of noise and over-fitting, including Bayesian priors (Mineault, Barthelme and Pack, 2009), radial averaging (Abbey and Eckstein, 2002) or using simplified low-dimensional stimuli (Li, Levi and Klein, 2004). In our analyses, we considered two different Bayesian methods for regularization of the second-stage filter weights. We applied a Gaussian "ridge-regression" prior (Bishop, 2006) to each of the second-stage filters weight vectors $\mathbf{w}_i$ (Fig. 2, blue lines) which encourages small weights to prevent fitting noise. The prior is given by

$$\ln p_0(w_i) = -\frac{1}{2\lambda}\sum_{j=1}^{d} w_{ij}^2, \tag{M.11}$$

where w_{ij} denotes the *j*-th element of the filter $\mathbf{w}_i$ and the hyper-parameter λ controls the scale. We see from (M.11) that smaller values of λ impose a stronger size penalty, and larger values impose a weaker constraint. We optimized λ using *k*-fold cross validation (Bishop, 2006), finding the range of values $\lambda = 0.1 - 1$ to provide a good trade-off between fitting to the training data and generalizing to novel data in the test set in Experiment 1. (Supplementary Figs. S1, S2). Since Experiment 3 involved fitting a novel model architecture than that used in Experiments 1 and 2, this parameter was re-optimized for Experiment 3 using the same procedure (Supplementary Fig. S23).

We also considered the effects of an additional prior $q_0(\mathbf{w}_i)$ which encourages smooth second-stage filter weights (Wilmore and Smyth, 2003). Such a prior is constructed by constructing a matrix $\mathbf{S_2}$ whose rows are the 2-D Laplace operator

$$\mathbf{L}_2 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{M.12}$$

centered at every point in a $p \times p$ matrix of zeros and unwrapped into a vector of length $d = p^2$, where p is the down-sampled image dimensionality. This smoothness prior is specified by

$$\ln q_0(\mathbf{w}_i) = -\frac{1}{2\omega}\mathbf{w}^{\mathrm{T}}(\mathbf{S}_2^{\mathrm{T}}\mathbf{S}_2)\mathbf{w}, \tag{M.13}$$

with hyper-parameter ω controlling the extent to which roughness is penalized. As with the ridge regression prior (M.11), smaller values of ω impose a stronger constraint than larger values. We optimized ω using k-fold cross validation and found for the 2D model with $\lambda = 1$ there was a slight improvement in generalization when we also include the smoothness prior (M.13), with the value of ω =1 providing a good tradeoff between fitting the data and generalizing to novel test data in Experiment 1 (Supplementary Figs. S3, S4). We found that adding this additional smoothness constraint reduced fitting to the test set, but improved generalization of the model to novel data (Supplementary Figs. S5, S6).

Numerical optimization

Second-stage filter weights $\mathbf{w}_i$ and the exponent α of the subunit nonlinearity were estimated using alternating coordinate ascent in a manner similar to the EM algorithm (Dempster, Laird and Rubin, 1970). Weights were optimized one filter $\mathbf{w}_i$ at a time while all other filters $\mathbf{w}_{j\neq i}$ and the power law exponent α were fixed. To optimize the second-stage filter weights we made use of a quasi-Newton optimization method (l-BFGS algorithm implemented in MATLAB®), which can be much more effective than standard stochastic gradient ascent (Bishop, 2006). To estimate the power law exponent α, we fixed all of the second-stage filters and performed a 1-D line search over the range [0.1,4]. Unfortunately, the optimizations for $\mathbf{w}_i$ and α do not have the convenient property of log-concavity (guaranteed by a negative definite Hessian), which guarantees no local maxima are possible (Paninski, Pillow and Lewi, 2007). However, in practice we usually found robust and consistent estimates of model parameters across optimization runs starting from several starting points. One kind of "local maximum" which the search was occasionally susceptible to was learning only a single non-zero filter. Therefore, as an additional precaution, after the first round of optimization (of at least 4 total rounds depending on

the experiment) we would re- initialize the search at a point in parameter space where the smaller filter (in magnitude) was set to be mirror-symmetric to the current value of the larger one. We would then allow the optimization to proceed as usual for several more rounds before termination. This procedure had the effect of eliminating any problems with local minima.

Preliminary investigations also included estimation of the lapse rate γ, but always yielded $\gamma = 0$, so this parameter was not included in our final model. Previous work has shown that the best estimates of lapse rates are attained when a large number of easy trials are included (Prins, 2012). This was not the case in our study, nor in the general classification image literature, which tends to use near-threshold stimuli and does not typically estimate lapse rates (Murray, 2011).

Analyses of fitted models

We were interested in using our modeling approach to determine what sorts of spatial strategies observers use to detect or discriminate contrast-modulated edges. Therefore, we visualized the second-order weights in a variety of ways.

In Experiment 1, in order to visualize the relationship between weight magnitudes and distance from the diagonal, we averaged the second-order weights along each of the diagonals (left-oblique units) and anti-diagonals (right-oblique units) to obtain a 1-D profile of average weight magnitude as a function of distance from the contrast modulated edges. In addition to performing this analysis for model fits to k = 4, 5 or 6 single training sets, we obtained an average profile by fitting our model to N = 30 random re-samplings of the dataset, and used this to estimate confidence intervals for the 1-D profiles. In order to quantify the extent to which the average 1-D profiles were characteristic of an edge-based or region-based approach, we fit the following "rising/falling edge" functions to the normalized averaged 1-D profiles

$$f(x) = \pm A \tanh(\beta x) \exp\left(-\frac{x^2}{\sigma^2}\right), \tag{M.14}$$

with x denoting the distance from the diagonal and free parameters A, β, σ, with σ determining the extent to which the second-stage filter weights are localized towards the diagonal. Examples of these functions with various values of σ are given in Fig. 6a.

In Experiment 2, where observers performed a fine orientation discrimination task, we increased the sampling density of the first-stage filter outputs in order to obtain better spatial resolution. Therefore we only down-sampled to 22 x 22. Since the two stimulus alternatives were

both nearly vertical, we analyzed the extent to which observers employed a boundary-based versus region based strategy by simply averaging the spatial weighting profiles along the vertical dimension.

In Experiment 3, we were not concerned with spatial strategy but rather with how information is integrated across multiple orientation channels. The "standard" FRF model architecture used in many previous studies assumes that each second-stage filter only receives first-order inputs having a single orientation (Malik and Perona 1990; Graham, 1991; Zavitz and Baker, 2013). However, recent psychophysical work has called this assumption into question by demonstrating that observers integrate luminance contrast and contrast polarity over multiple orientations (Motoyoshi and Nishida, 2004; Motoyoshi and Kingdom, 2007). We sought to demonstrate that our modeling framework could be elegantly applied to compare competing hypotheses of FRF architecture by fitting competing models to the same dataset and evaluating their goodness of fit using standard model comparison techniques like the Bayesian (BIC) or Akaike (AIC) Information criteria (Akaike, 1974; Schwarz, 1978; Bishop, 2006). These model comparison techniques reward goodness-of-fit, as measured by the data likelihood, while penalizing model complexity, as quantified by the number of free parameters. For the two models we fit to our data (Fig. 12), the number of free parameters was identical, so the AIC/BIC both simply reduce to the data likelihood.

References

Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of vision*, *2*(1), 5-5.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA A*, *2*(2), 284-299.

Ahumada Jr, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, *25*(suppl), 2-2.

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, *2*(1), 8-8.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Allard, R., & Faubert, J. (2007). Double dissociation between first-and second-order processing. *Vision Research*, *47*(9), 1129-1141.

Antolík, J., Hofer, S. B., Bednar, J. A., & Mrsic-Flogel, T. D. (2016). Model Constrained by Visual Hierarchy Improves Prediction of Neural Responses to Natural Scenes. *PLoS computational biology*, *12*(6), e1004927.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature neuroscience*, *10*(10), 1313.

Baker, C. L. (1999). Central neural mechanisms for detecting second-order motion. *Current opinion in neurobiology*, *9*(4), 461-466.

Baker, C. L., & Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. In *Progress in brain research* (Vol. 134, pp. 171-191). Elsevier.

Baker, D. H., & Meese, T. S. (2014). Measuring the spatial extent of texture pooling using reverse correlation. *Vision research*, *97*, 52-58.

Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In: M.S. Landy and J.A. Movhshon (Eds.) *Computational models of visual processing* (253-271). Cambridge: MIT Press.

Bishop, C. M. (2006). Pattern recognition and machine learning: Springer New York.

Bishop, P. O., Coombs, J. S., & Henry, G. H. (1973). Receptive fields of simple cells in the cat striate cortex. *The Journal of physiology*, *231*(1), 31-60.

Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE transactions on pattern analysis and machine intelligence*, *12*(1), 55-73.

Cadieu, C. F., & Olshausen, B. A. (2012). Learning intermediate-level representations of form and motion from natural movies. *Neural computation*, *24*(4), 827-866.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural computation*, *22*(4), 887-905.

Cavagnaro, D. R., Pitt, M. A., & Myung, J. I. (2011). Model discrimination through adaptive experimentation. *Psychonomic bulletin & review*, *18*(1), 204-210.

Cavagnaro, D. R., Gonzalez, R., Myung, J. I., & Pitt, M. A. (2013). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management science*, *59*(2), 358-375.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, *12*(2), 199-213.

Dakin, S. C., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision research*, *40*(3), 311-329.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, *69*(4), 1118-1135.

DiMattina, C. (2015). Fast adaptive estimation of multidimensional psychometric functions. *Journal of Vision*, *15*(9), 5-5.

DiMattina, C. (2016). Comparing models of contrast gain using psychophysical experiments. *Journal of vision*, *16*(9), 1-1.

DiMattina, C., Fox, S. A., & Lewicki, M. S. (2012). Detecting natural occlusion boundaries using local cues. *Journal of vision*, *12*(13), 15-15.

DiMattina, C., & Zhang, K. (2011). Active data collection for efficient estimation and comparison of nonlinear neural models. *Neural computation*, *23*(9), 2242-2288.

DiMattina, C., & Zhang, K. (2013). Adaptive stimulus optimization for sensory systems neuroscience. *Frontiers in neural circuits*, *7*, 101.

Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, *2*(1), 3-3.

Ellemberg, D., Allen, H. A., & Hess, R. F. (2006). Second-order spatial frequency and orientation channels in human vision. *Vision research*, *46*(17), 2798-2803.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature neuroscience*, *16*(7), 974.

Gharat, A., & Baker, C. L. (2017). Nonlinear Y-like receptive fields in the early visual cortex: An intermediate stage for building cue-invariant receptive fields from subcortical Y cells. *Journal of Neuroscience*, *37*(4), 998-1013.

Goda, N., Tachibana, A., Okazawa, G., & Komatsu, H. (2014). Representation of the material properties of objects in the visual cortex of nonhuman primates. *Journal of Neuroscience*, *34*(7), 2660-2673.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge: MIT press.

Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision research*, *32*(4), 719-743.

Graham, N., Sutter, A., & Venkatesan, C. (1993). Spatial-frequency-and orientation-selectivity of simple and complex channels in region segregation. *Vision Research*, *33*(14), 1893-1911.

Gurnsey, R., & Laundry, D. S. (1992). Texture discrimination with and without abrupt texture gradients. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *46*(2), 306.

Hallum, L. E., & Movshon, J. A. (2014). Surround suppression supports second-order feature encoding by macaque V1 and V2 neurons. *Vision research*, *104*, 24-35.

Hanazawa, A., & Komatsu, H. (2001). Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *Journal of Neuroscience*, *21*(12), 4490-4497.

Hansen, B. C., & Hess, R. F. (2006). The role of spatial phase in texture segmentation and contour integration. *Journal of Vision*, *6*(5), 5-5.

Harper, N. S., Schoppe, O., Willmore, B. D., Cui, Z., Schnupp, J. W., & King, A. J. (2016). Network receptive field modeling reveals extensive integration and multi-feature selectivity in auditory cortical neurons. *PLoS computational biology*, *12*(11), e1005113.

Hasan, B. A. S., Joosten, E., & Neri, P. (2012). Estimation of internal noise using double passes: Does it matter how the second pass is delivered?. *Vision research*, *69*, 1-9.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual neuroscience*, *9*(2), 181-197.

Hegdé, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience*, *20*(5), RC61-RC61.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Hutchinson, C. V., Ledgeway, T., & Baker, C. L. (2016). Phase-dependent interactions in visual cortex to combinations of first-and second-order stimuli. *Journal of Neuroscience*, *36*(49), 12328-12337.

Johnson, A. P., & Baker, C. L. (2004). First-and second-order information in natural images: a filter-based approach to image statistics. *JOSA A*, *21*(6), 913-925.

Johnson, A. P., & Baker, C. L. (2005). Spatiochromatic statistics of natural scenes: first-and second-order information and their correlational structure. *JOSA A*, *22*(10), 2050-2059.

Johnson, A. P., Prins, N., & Baker Jr, C. L. (2007). Ecologically valid combinations of first-and second-order surface markings facilitate texture discrimination. *Vision research*, *47*(17), 2281-2290.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of neurophysiology*, *84*(4), 2048-2062.

Karklin, Y., & Lewicki, M. S. (2003). Learning higher-order structures in natural images. *Network: Computation in Neural Systems*, *14*(3), 483-499.

Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, *457*(7225), 83.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773-795.

Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision research*, *37*(4), 371-376.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, *10*(11), e1003915.

Kingdom, F.A.A., Hayes, A., & Field, D. J. (2001). Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision research*, *41*(5), 585-598.

Kingdom, F.A.A., & Keeble, D. R. (1996). A linear systems approach to the detection of both abrupt and smooth spatial variations in orientation-defined textures. *Vision research*, *36*(3), 409-420.

Kingdom, F.A.A., Keeble, D., & Moulden, B. (1995). Sensitivity to orientation modulation in micropattern-based textures. *Vision research*, *35*(1), 79-91.

Kingdom, F.A.A., Prins, N., & Hayes, A. (2003). Mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism. *Visual neuroscience*, *20*(1), 65-76.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961-980.

Knoblauch, K., & Maloney, L. (2008a). Classification images estimated by generalized additive models. *Journal of Vision*, *8*(6), 344-344.

Knoblauch, K., & Maloney, L. T. (2008b). Estimating classification images with generalized linear and additive models. *Journal of Vision*, *8*(16), 10-10.

Kohler, P. J., Clarke, A., Yakovleva, A., Liu, Y., & Norcia, A. M. (2016). Representation of maximally regular textures in human visual cortex. *Journal of Neuroscience*, *36*(3), 714-729.

Konishi, S., Yuille, A. L., Coughlan, J. M., & Zhu, S. C. (2003). Statistical edge detection: Learning and evaluating edge cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(1), 57-74.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision research*, *39*(16), 2729-2737.

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, *1*, 417-446.

Kurki, I., Saarinen, J., & Hyvärinen, A. (2014). Investigating shape perception by classification images. *Journal of vision*, *14*(12), 24-24.

Landy, M. S., & Graham, N. (2004). Visual Perception of Texture. In: L.M. Chalupa and J.S. Werner (Eds.) *The visual neurosciences* (1106-1118). Cambridge: MIT Press.
Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *JOSA A*, *18*(9), 2307-2320.

Landy, M. S., & Oruç, I. (2002). Properties of second-order spatial frequency channels. *Vision research*, *42*(19), 2311-2329.

Ledgeway, T., & Smith, A. T. (1994). Evidence for separate motion-detecting mechanisms for first-and second-order motion in human vision. *Vision research*, *34*(20), 2727-2740.

Lesmes, L. A., Lu, Z. L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of vision*, *10*(3), 17-17.

Lewi, J., Butera, R., & Paninski, L. (2009a). Sequential optimal design of neurophysiology experiments. *Neural Computation*, *21*(3), 619-687.

Lewi, J., Butera, R., Schneider, D. M., Woolley, S., & Paninski, L. (2009b). Designing neurophysiology experiments to optimally constrain receptive field models along parametric submanifolds. In *Advances in Neural Information Processing Systems* (pp. 945-952).

Li, G., & Baker, C. L. (2012). Functional organization of envelope-responsive neurons in early visual cortex: Organization of carrier tuning properties. *Journal of Neuroscience*, *32*(22), 7538-7549.

Li, G., Yao, Z., Wang, Z., Yuan, N., Talebi, V., Tan, J., ... & Baker, C. L. (2014). Form-cue invariant second-order neuronal responses to contrast modulation in primate area V2. *Journal of Neuroscience*, *34*(36), 12081-12092.

Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. *Nature neuroscience*, *7*(2), 178.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 986-1005.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *JOSA A*, *7*(5), 923-932.

Mareschal, I., & Baker Jr, C. L. (1998). A cortical locus for the processing of contrast-defined contours. *Nature neuroscience*, *1*(2), 150.

Marr, D. (1982). *Vision.* WH Freeman and Company.

Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, *26*(5), 530-549.

Meese, T. S. (2010). Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of vision*, *10*(8), 14-14.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception*, *33*(9), 1101-1127.

McIlhagga, W. H., & May, K. A. (2012). Optimal edge filters explain human blur detection. *Journal of vision*, *12*(10), 9-9.

Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. *Vision research*, *42*(8), 1017-1033.

Mineault, P. J., Barthelme, S., & Pack, C. C. (2009). Improved classification images with sparse priors in a smooth basis. *Journal of Vision*, *9*(10), 17-17.

Morgenstern, Y., & Elder, J. H. (2012). Local visual energy mechanisms revealed by detection of global patterns. *Journal of Neuroscience*, *32*(11), 3679-3696.

Morrone, M. C., Burr, D. C., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc. R. Soc. Lond. B*, *216*(1204), 335-354.

Motoyoshi, I. Kingdom, FAA (2003). Orientation opponency in human vision revealed by energy-frequency analysis. *Vision Research*, *43*(21), 2197-2205.

Motoyoshi, I., & Kingdom, F.A.A. (2007). Differential roles of contrast polarity reveal two streams of second-order visual processing. *Vision Research*, *47*(15), 2047-2054.

Motoyoshi, I., & Nishida, S. Y. (2004). Cross-orientation summation in texture segregation. *Vision Research*, *44*(22), 2567-2576.

Murray, R. F. (2011). Classification images: A review. *Journal of vision*, *11*(5), 2-2.

Myung, J. I., Cavagnaro, D. R., & Pitt, M. A. (2013). A tutorial on adaptive design optimization. *Journal of mathematical psychology*, *57*(3-4), 53-67.

Neri, P. (2010). How inherently noisy is human sensory processing?. *Psychonomic Bulletin & Review*, *17*(6), 802-808.

Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision research*, *46*(16), 2465-2474.

Nothdurft, H. C. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision research*, *25*(12), 1957-1968.

Nothdurft, H. C., Gallant, J. L., & van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: correlates of "popout" under anesthesia. *Visual neuroscience*, *16*(1), 15-34.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (2000). Response profiles to texture border patterns in area V1. *Visual neuroscience*, *17*(3), 421-436.

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences*, *112*(4), E351-E360.

Okazawa, G., Tajima, S., & Komatsu, H. (2016). Gradual development of visual texture-selective properties between macaque areas V2 and V4. *Cerebral Cortex*, *27*(10), 4867-4880.

Paninski, L. (2005). Asymptotic theory of information-theoretic experimental design. *Neural Computation*, *17*(7), 1480-1507.

Park, M., & Pillow, J.W. (2011). "Receptive field inference with localized priors." *PLoS computational biology* 7.10 (2011): e1002219.

Pillow, J. W., & Park, M. (2016). Adaptive Bayesian methods for closed-loop neurophysiology. *Closed Loop Neuroscience (AE Hady, ed.), Elsevier*.

Prins, N., & Kingdom, FAA (2002). Orientation-and frequencymodulated textures at low depths of modulation are processed by off-orientation and off-frequency texture mechanisms. *Vision Research*, *42*, 705-713.

Rentschler, I., Hübner, M., & Caelli, T. (1988). On the discrimination of compound Gabor signals and textures. *Vision Research*, *28*(2), 279-291.

Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque V1: diversity and laminar dependence. *Journal of Neuroscience*, *22*(13), 5639-5651.

Rivest, J., & Cabanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision research*, *36*(1), 53-66.

Rosenberg, A., Husson, T. R., & Issa, N. P. (2010). Subcortical representation of non-Fourier image features. *Journal of Neuroscience*, *30*(6), 1985-1993.

Rosenholtz, R. (2014). Texture perception. *Oxford handbook of perceptual organization*, 167-186.

Rowekamp, R. J., & Sharpee, T. O. (2017). Cross-orientation suppression in visual area V2. *Nature communications*, *8*, 15739.

Saarela, T. P., & Landy, M. S. (2012). Combination of texture and color cues in visual segmentation. *Vision research*, *58*, 59-67.

Sebastian, S., & Geisler, W. S. (2018). Decision-variable correlation. *Journal of vision*, *18*(4), 3-3.

Schofield, A. J. (2000). What does second-order vision see in an image?. *Perception*, *29*(9), 1071-1086.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research*, *39*(16), 2697-2716.

Schofield, A. J., & Georgeson, M. A. (2003). Sensitivity to contrast modulation: The spatial frequency dependence of second-order vision. *Vision research*, *43*(3), 243-259.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Sebastian, S., & Geisler, W. S. (2018). Decision-variable correlation. *Journal of vision*, *18*(4), 3-3.

Sutter, A., Beck, J., & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception & Psychophysics*, *46*(4), 312-332.

Tanaka, H., & Ohzawa, I. (2009). Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features. *Journal of neurophysiology*, *101*(3), 1444-1462.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological review*, *121*(1), 124.

Vintch, B., Movshon, J. A., & Simoncelli, E. P. (2015). A convolutional subunit model for neuronal responses in macaque V1. *Journal of Neuroscience*, *35*(44), 14829-14841.

Walker, G. A., Ohzawa, I., & Freeman, R. D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *Journal of Neuroscience*, *19*(23), 10536-10553.

Wang, Z., & Simoncelli, E. P. (2008). Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, *8*(12), 8-8.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*, *33*(2), 113-120.

Westrick, Z. M., & Landy, M. S. (2013). Pooling of first-order inputs in second-order vision. *Vision research*, *91*, 108-117.

Willmore, B. D., Prenger, R. J., & Gallant, J. L. (2010). Neural representation of natural images in visual area V2. *Journal of Neuroscience*, *30*(6), 2102-2114.

Wilson, H. R. (1993). Nonlinear processes in visual pattern discrimination. *Proceedings of the National Academy of Sciences*, *90*(21), 9785-9790.

Wolfson, S. S., & Landy, M. S. (1998). Examining edge-and region-based texture analysis mechanisms. *Vision Research*, *38*(3), 439-446.

Wu, M. C. K., David, S. V., & Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.*, *29*, 477-505.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619-8624.

Zavitz, E., & Baker, C. L. (2013). Texture sparseness, but not local phase structure, impairs second-order segmentation. *Vision research*, *91*, 45-55.

Zavitz, E., & Baker, C. L. (2014). Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues. *Journal of vision*, *14*(4), 14-14.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer.

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016). Selectivity and tolerance for visual texture in macaque V2. *Proceedings of the National Academy of Sciences*, *113*(22), E3140-E3149.

Figure Legends

Figure 1: Examples of second-order boundary stimuli, with varying texture density and modulation depth. **(a)** *Top:* A natural occlusion edge having similar luminance on both sides, but clearly visible due to differences in texture. *Bottom:* A contrast-defined edge with identical mean luminance in each region, but different contrasts of the texture elements. In this example the contrast modulation envelope has a left-oblique (-45 deg. w.r.t vertical) orientation. **(b)** Examples of contrast-modulated micropattern stimuli, for three densities and three modulation depths, all having a right-oblique (+45 deg.) boundary orientation. **(c)** Schematic illustration of the two psychophysical tasks (*orientation identification*: Experiments 1 and 3, *orientation discrimination*: Experiment 2).

Figure 2: Neural network implementing a FRF (Filter-Rectify-Filter) arrangement for texture segmentation. Stimulus image **I** is filtered by a bank of first-stage energy filters, resembling V1 complex cells, whose down-sampled responses provide inputs **x** to two second-stage filters. An optimization algorithm adjusts connection weights $\mathbf{w}_L$, $\mathbf{w}_R$ (blue lines), to produce second-stage filters which are selective for left-oblique (L) or right-oblique (R) contrast-defined boundaries, to produce responses consistent with psychophysical data. The output of each second-stage filter is passed through a non-linearity $h(u) = |u|^{\alpha}$ (blue curve, inset) whose shape parameter α is also estimated from the psychophysical data. Fixed output weights $v_R = +1$, $v_L = -1$ lead to the decision variable $u = s_R - s_L + v_0$, which is input to a sigmoid function to determine the probability of the subject classifying a given edge as being right-oblique.

Figure 3: Second stage filter weights for two representative subjects in Experiment 1. **(a)** Second-stage filter weights learned by the model for representative observer AMA in Experiment 1-VAR for two different priors (left: *ridge*; right: *ridge + smooth*) with 16x16 AVG down-sampling. Top panels show the 2-D filter weights (averaged over 4 training folds) and bottom panels show these 2-D weights collapsed into a 1-D profile (black dashed lines) by averaging along the matrix diagonals (left-oblique subunit) or anti-diagonals (right-oblique subunit). Thick lines (red: *ridge*; green: *ridge + smooth*) denote average over 30 resampled bootstrapped training sets, and thin lines dashed lines show +/- 1 SEM.. **(b)** Same as (a) but for

observer JJF in Experiment 1-FIX. **(c)** Ideal observer for Experiment 1-VAR. Organization as in (a), (b) except here thick black line denotes average over 4 training folds and thin dashed lines show fits to individual folds.

Figure 4: 1-D profiles of the second-stage filter weights for all individual observers in Experiment 1. **(a)** 1-D profiles for observers in Experiment1-VAR (colored lines) plotted with the average across observers (thick black lines) for three sampling rules and both priors. **(b)** Same as (a) but for Experiment 1-FIX.

Figure 5: Normalized 1-D profiles of the average second-stage filter weights (30 bootstrapped re-samples) for the ideal observer (thick black lines).Thin black dashed lines indicate +/- 1.96 SEM, and colored lines indicate normalized individual profiles. **(a)** Experiment 1-VAR (top) and Experiment 1-FIX (bottom). **(b)** Experiment 1-HFC.

Figure 6: Rising edge functions fit to 1-D second-stage filter weight profiles. **(a)** Rising edge functions for various values of the localization parameter. We see for (log-transformed) values greater than or equal to 1, this function exhibits nearly uniform weighting with distance.
(b) Rising edge functions fit to 1-D weight profiles from one observer in Experiment 1-VAR (AMA). **(c)** Histograms of localization parameter values for all observers in Experiment 1-VAR.

Figure 7: Psychometric functions for three representative observers in Experiment 1-VAR. Blue lines denote observer performance on the test set (stimuli not used for model estimation) as a function of modulation depth, together with 95% binomial proportion confidence intervals (+/- 1.96 SEM, blue dashed lines). Red and green lines denote model predictions for test stimuli as a function of contrast modulation depth for both priors (red: *ridge*; green: *ridge + smooth*). Left column shows model operating deterministically (DET), middle column shows model operating stochastically (STO), and right column shows deterministic model with noise optimized to fit observer performance (NSE).

Figure 8: Analysis of model accuracy for Experiment 1-VAR. (**a**) Plots of observer performance (horizontal axis) and model performance (vertical axis) averaged across observers (left sub-

panels, circles, N = 7) or test folds (right sub-panels, solid symbols, N = 28). Observer and model performance was compared on a set of novel test stimuli (N = 500) not used for model estimation or hyper-parameter optimization. *Top left:* Deterministic model with AVG down-sampling (AVG-DET). Correlation coefficients are color coded for each choice of Bayesian prior (red: *ridge*; green: *ridge + smooth*). *Top right:* Stochastic model with AVG down-sampling (AVG-STO). *Bottom left:* Deterministic model with noise added to optimize fit with observer performance (AVG-NSE). *Bottom right:* Deterministic model with SUB down-sampling (SUB-DET). (**b**) Difference between observer and model performance (observer - model) for all for each individual test fold (4 folds per subject) for all models shown in (a). Lines show 95% confidence intervals of the difference (binomial proportion difference test). Colors indicate different choices of Bayesian prior (red: *ridge*; green: *ridge + smooth*).

Figure 9: Double-pass analysis for N = 5 observers in Experiment 1-VAR. (**a**) Models operating in deterministic (DET) mode for three different down-sampling rules. *Right panels:* Observer-observer (horizontal axis) and observer-model (vertical axis) proportion agreement on the double-pass set. Thin lines denote 95% binomial proportion confidence intervals. *Left panels*: Difference between observer-observer and observer-model proportion agreement for all observers. Lines denote 95% binomial proportion difference confidence intervals. (**b**) Model with AVG down-sampling operating in stochastic mode (AVG-STO) and deterministic mode with added noise (AVG-NSE). Organization is the same as (a).

Figure 10: Decision variable correlation analysis. Decision variable correlation (DVC) computed from the L trials (ordinate) and R trials (abscissa) for three observers (CJD = blue, MAK = magenta, JJF = black) in Experiment 1-FIX for both priors and all choices of pooling rule. Thin lines denote +/-1 SEM (200 bootstrap re-sampling).

Figure 11: Second-stage filter weights for two representative observers from Experiment 2. **(a)** Observer CJD in Experiment 2-VAR. *Top:* 2-D profiles averaged over k = 4 training folds. *Bottom:* 1-D profiles showing mean weight magnitude (thick lines), averaged over 30

bootstrapped re-samplings. Thin dashed lines show +/- 1 SEM. **(b)** Observer JJF in Experiment 2-FIX.

Figure 12: Model comparison for Experiment 3. **(a)** Two competing models of FRF architecture fit to psychophysical trial data in Experiment 3. *Top:* Model 1 assumes that each first-stage orientation channel is analyzed by its own pair of (L/R) second-stage filters. *Bottom:* Model 2 assumes that each (L/R) second-stage filter integrates over both first-stage channels.
(**b**) Bayesian model comparison making use of all data reveals consistent preference for Model 2 for all three observers (color lines and symbols). (**c**) Bootstrapping analysis where model likelihood is evaluated on novel data not used for model training also reveals a preference for Model 2.

Figure 13: Hypothetical extension of our modeling framework with additional hidden layers of second-stage filters and a more complete implementation of the first-stage filters. Such a model can potentially account for human performance on boundary perception tasks with complex stimuli with both first- and second-order cues, like natural occlusion boundaries.

Figure 1

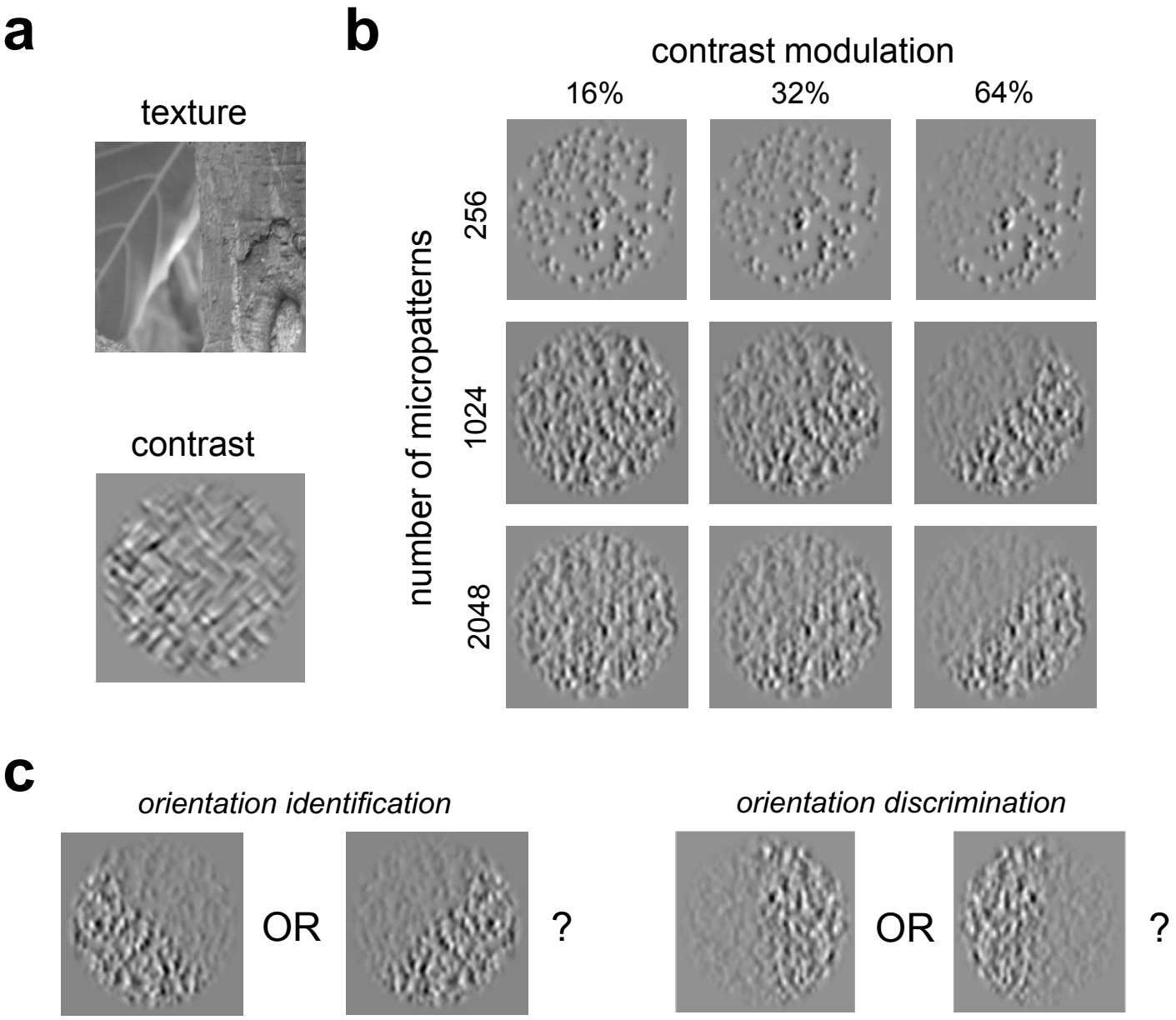

a
texture
contrast
b
contrast modulation
16%
32%
64%
number of micropatterns
256
1024
2048
c
orientation identification
OR
?
orientation discrimination
OR
?

Figure 2

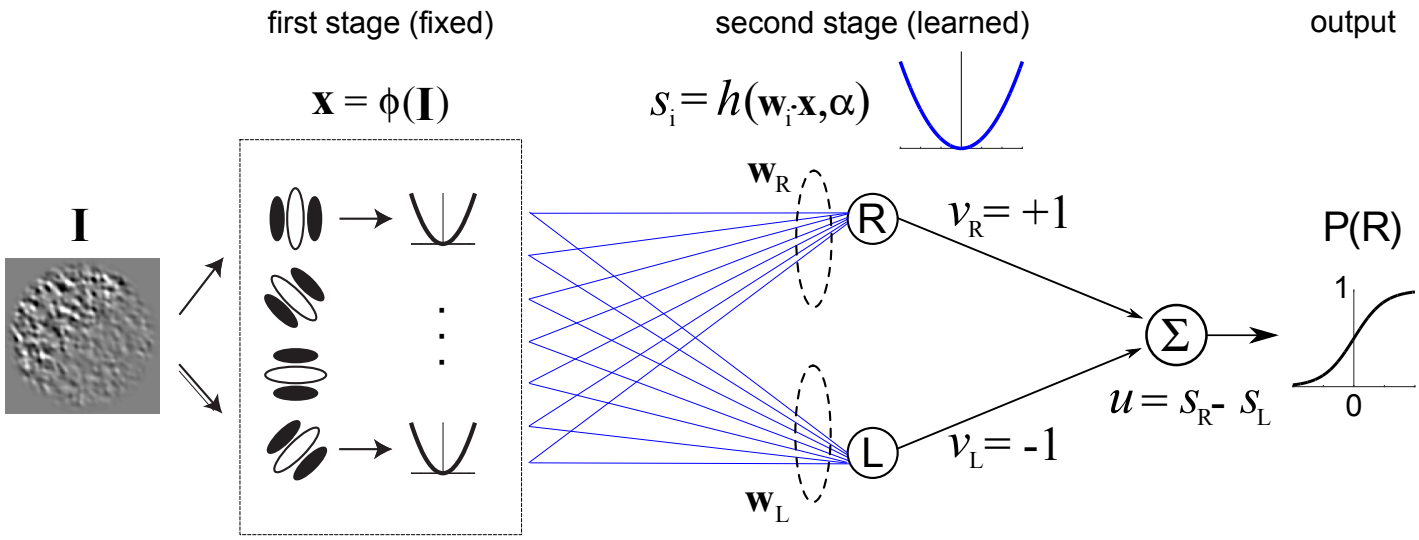

first stage (fixed)
second stage (learned)
output
$\mathbf{x} = \phi(\mathbf{I})$
$s_i = h(\mathbf{w}_i \cdot \mathbf{x}, \alpha)$
$\mathbf{I}$
$\mathbf{w}_R$
$\mathbf{w}_L$
R
L
$v_R = +1$
$v_L = -1$
Σ
$u = s_R - s_L$
P(R)
1
0

Figure 3

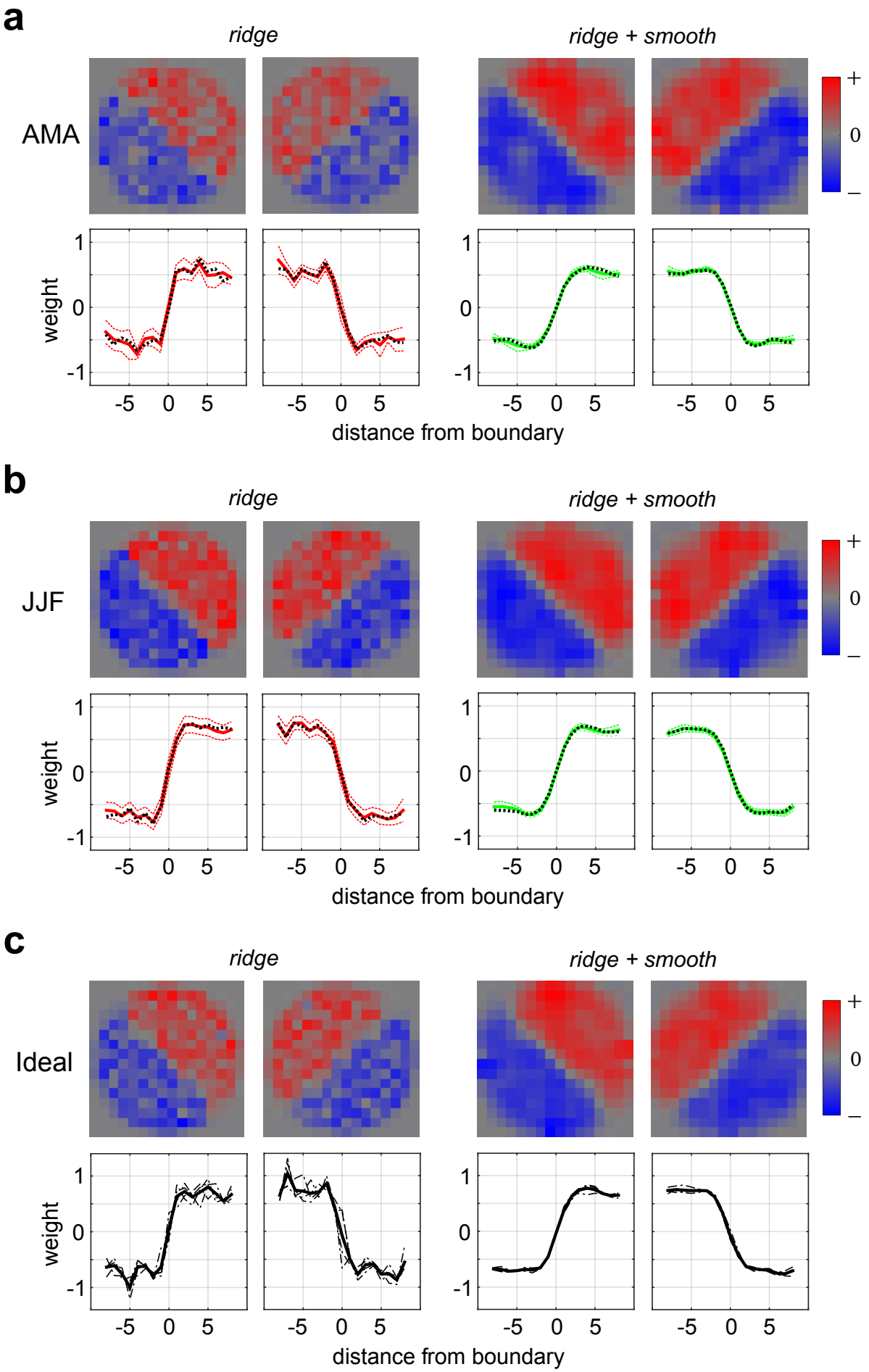

a
ridge
ridge + smooth
AMA
+
0
–
weight
1
0
-1
-5
0
5
distance from boundary
b
ridge
ridge + smooth
JJF
+
0
–
weight
1
0
-1
-5
0
5
distance from boundary
c
ridge
ridge + smooth
Ideal
+
0
–
weight
1
0
-1
-5
0
5
distance from boundary

Figure 4

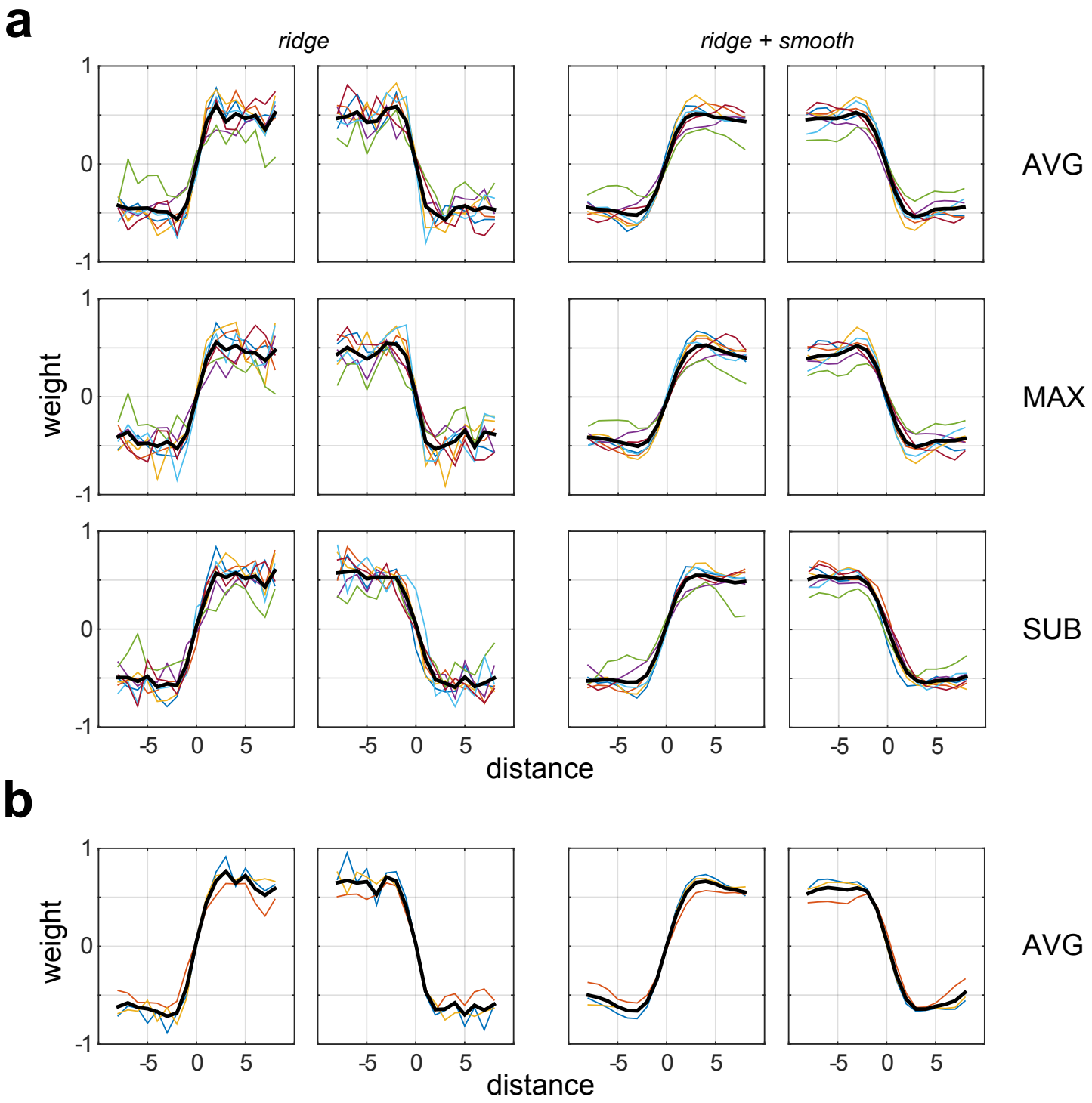

a
ridge
ridge + smooth
AVG
MAX
SUB
weight
distance
1
0
-1
-5
0
5
b
AVG
weight
distance

Figure 5

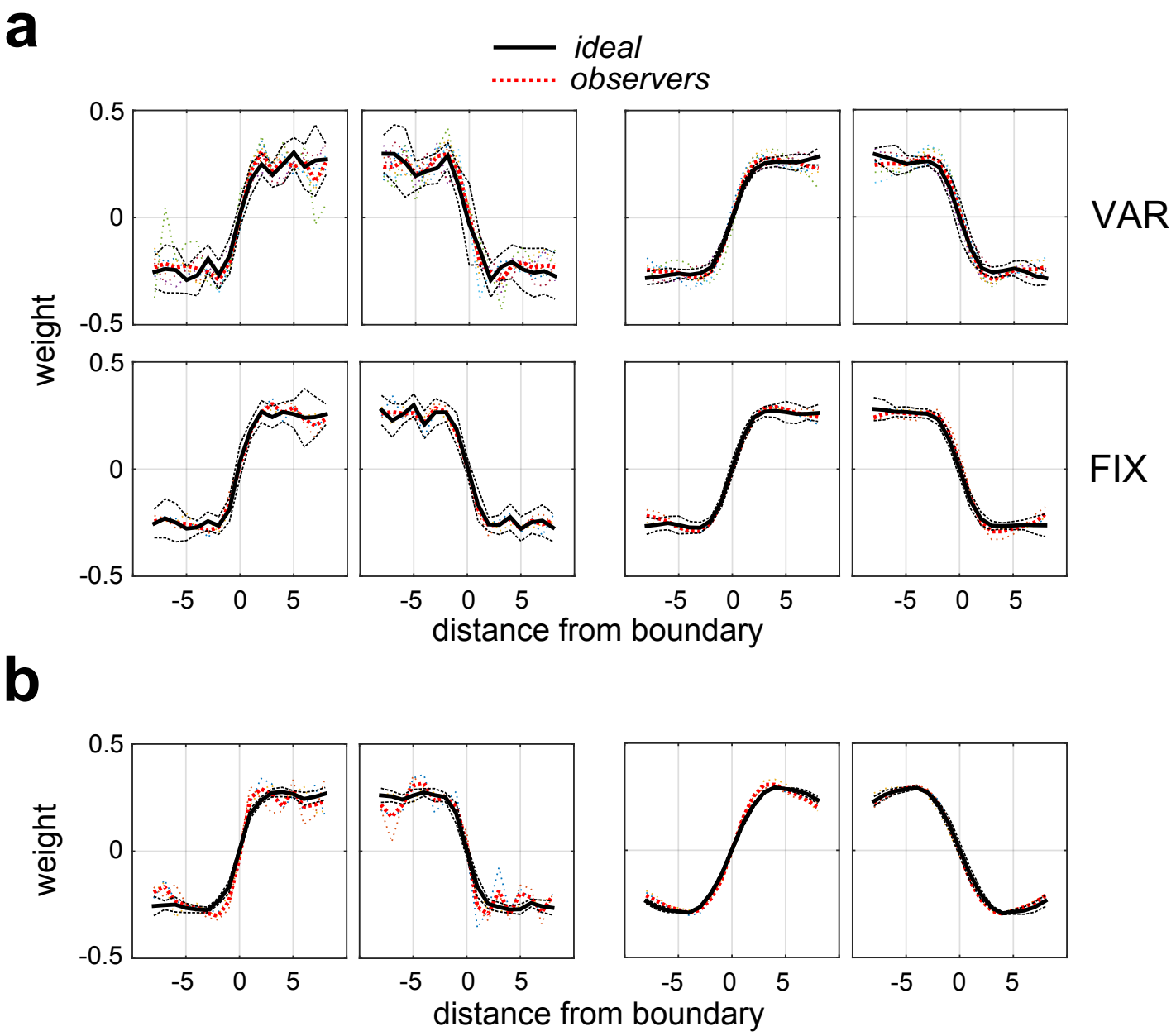

a
ideal
observers
0.5
0
-0.5
weight
VAR
FIX
-5
0
5
distance from boundary
b
0.5
0
-0.5
weight
-5
0
5
distance from boundary

Figure 6

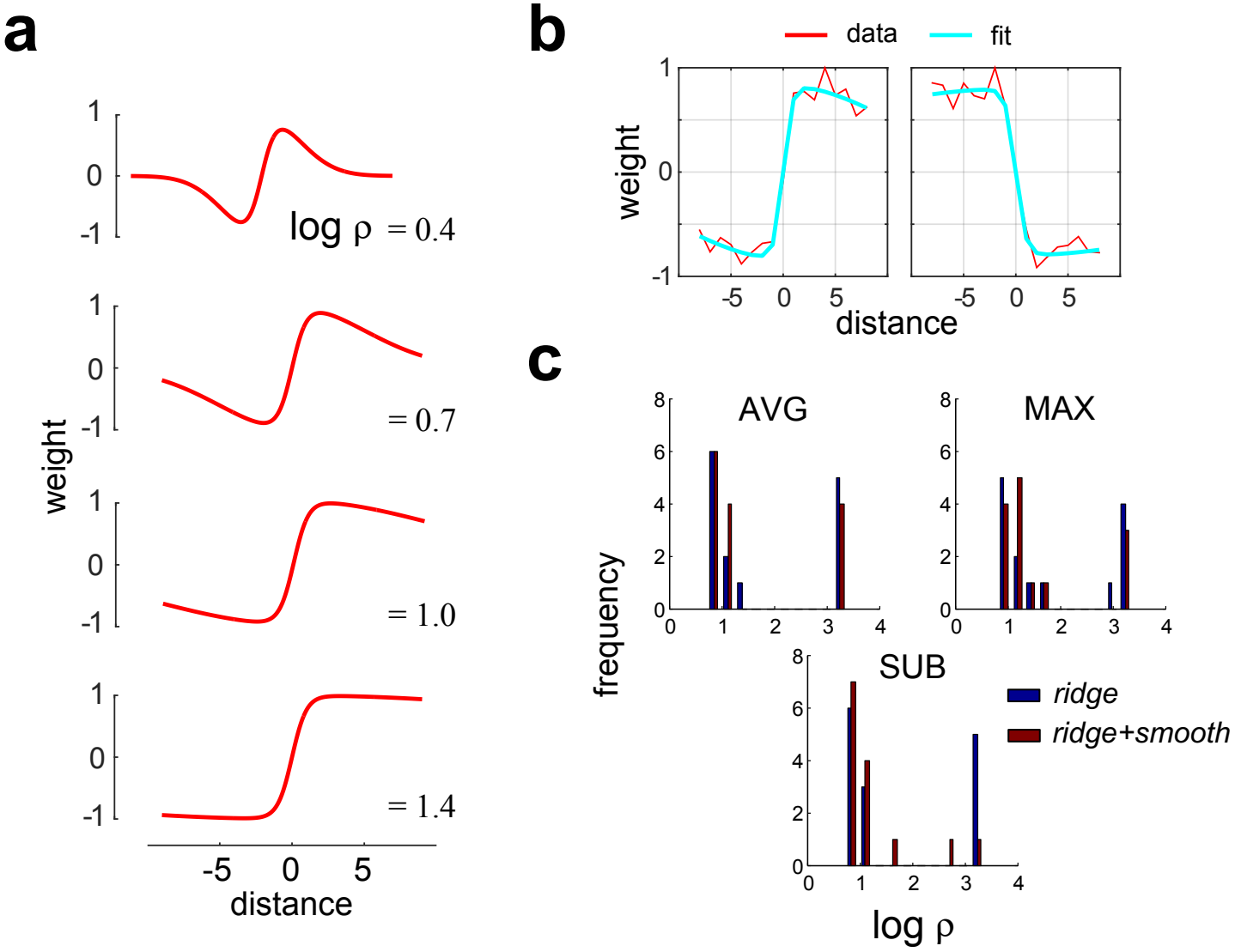

a
log ρ = 0.4
= 0.7
= 1.0
= 1.4
weight
distance
b
data
fit
weight
distance
c
AVG
MAX
SUB
frequency
ridge
ridge+smooth
log ρ

Figure 7

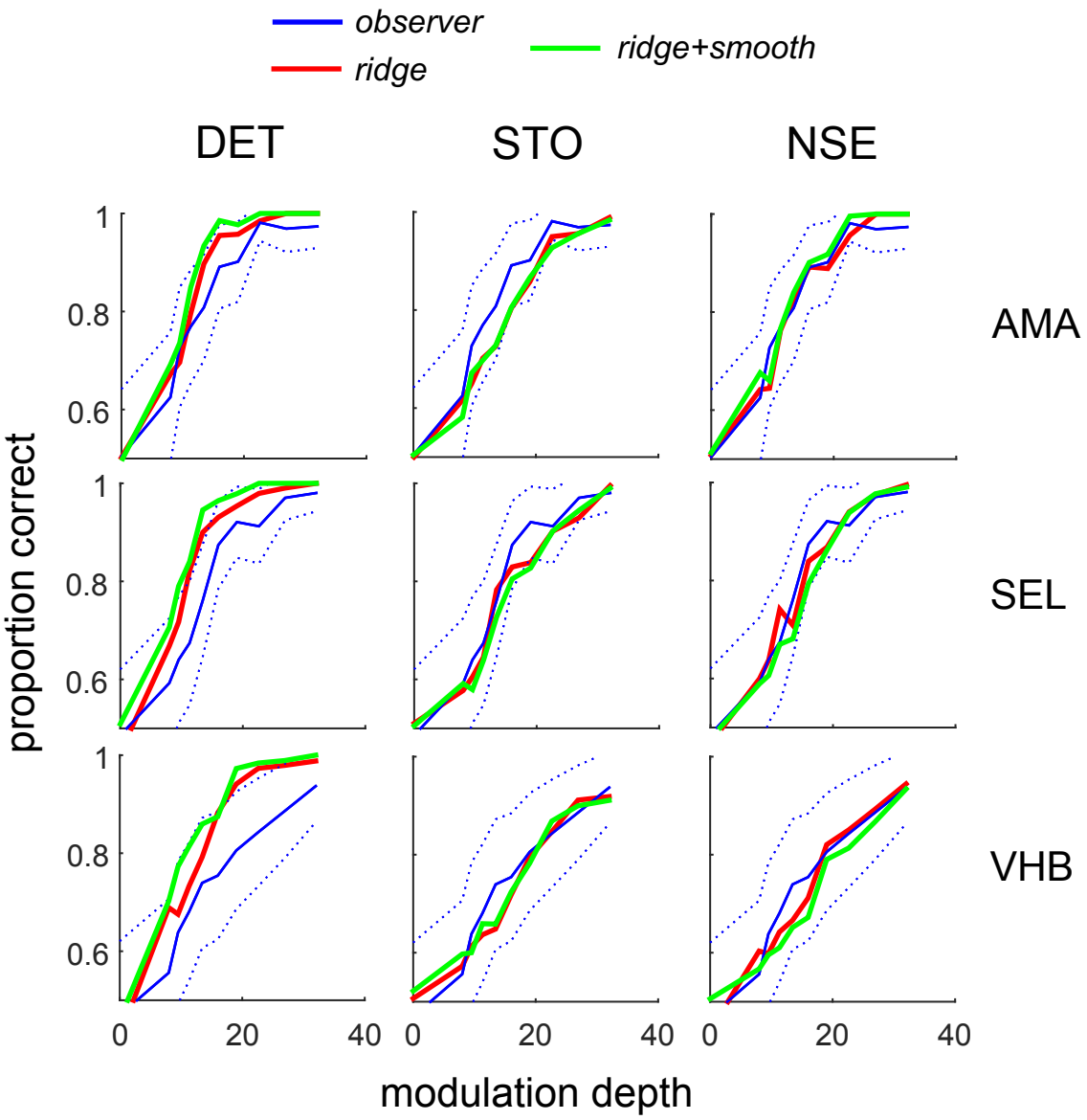

observer
ridge
ridge+smooth
DET
STO
NSE
AMA
SEL
VHB
proportion correct
modulation depth
1
0.8
0.6
0
20
40

Figure 8

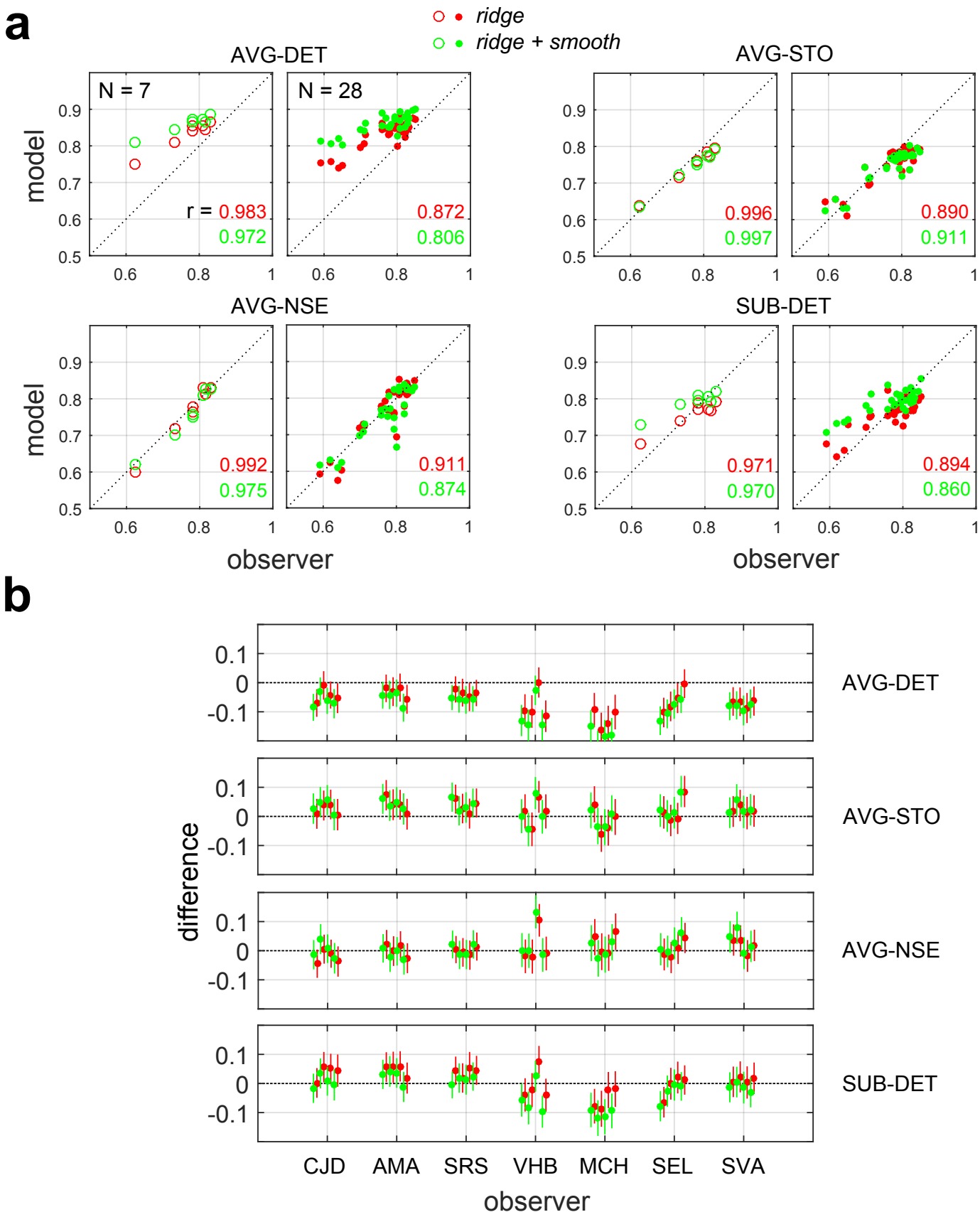

a
ridge
ridge + smooth
AVG-DET
AVG-STO
AVG-NSE
SUB-DET
N = 7
N = 28
r =
0.983
0.972
0.872
0.806
0.996
0.997
0.890
0.911
0.992
0.975
0.911
0.874
0.971
0.970
0.894
0.860
model
observer
b
difference
AVG-DET
AVG-STO
AVG-NSE
SUB-DET
CJD
AMA
SRS
VHB
MCH
SEL
SVA
observer

Figure 9

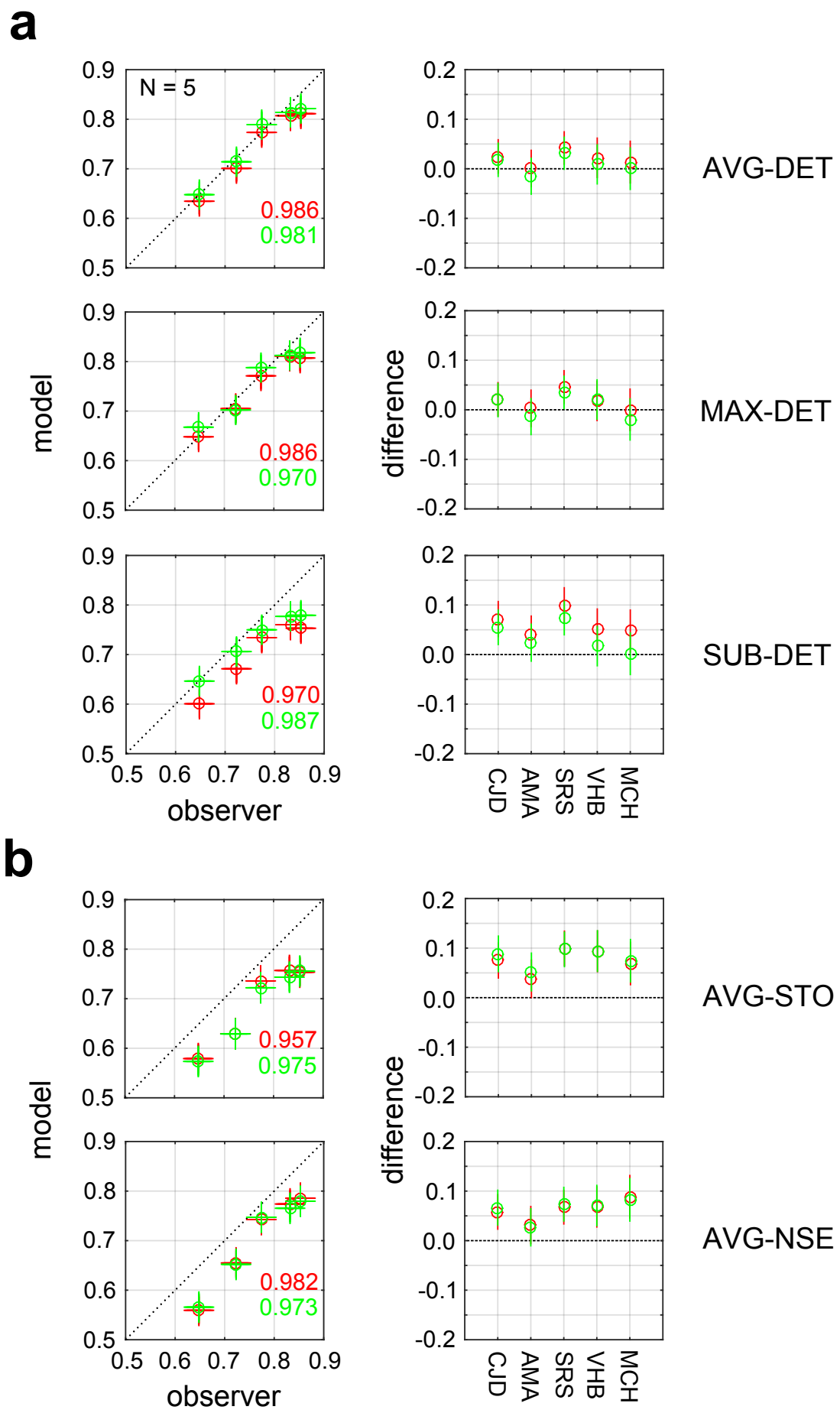

a
N = 5
0.986
0.981
0.986
0.970
0.970
0.987
model
observer
difference
AVG-DET
MAX-DET
SUB-DET
CJD
AMA
SRS
VHB
MCH
b
0.957
0.975
0.982
0.973
AVG-STO
AVG-NSE

Figure 10

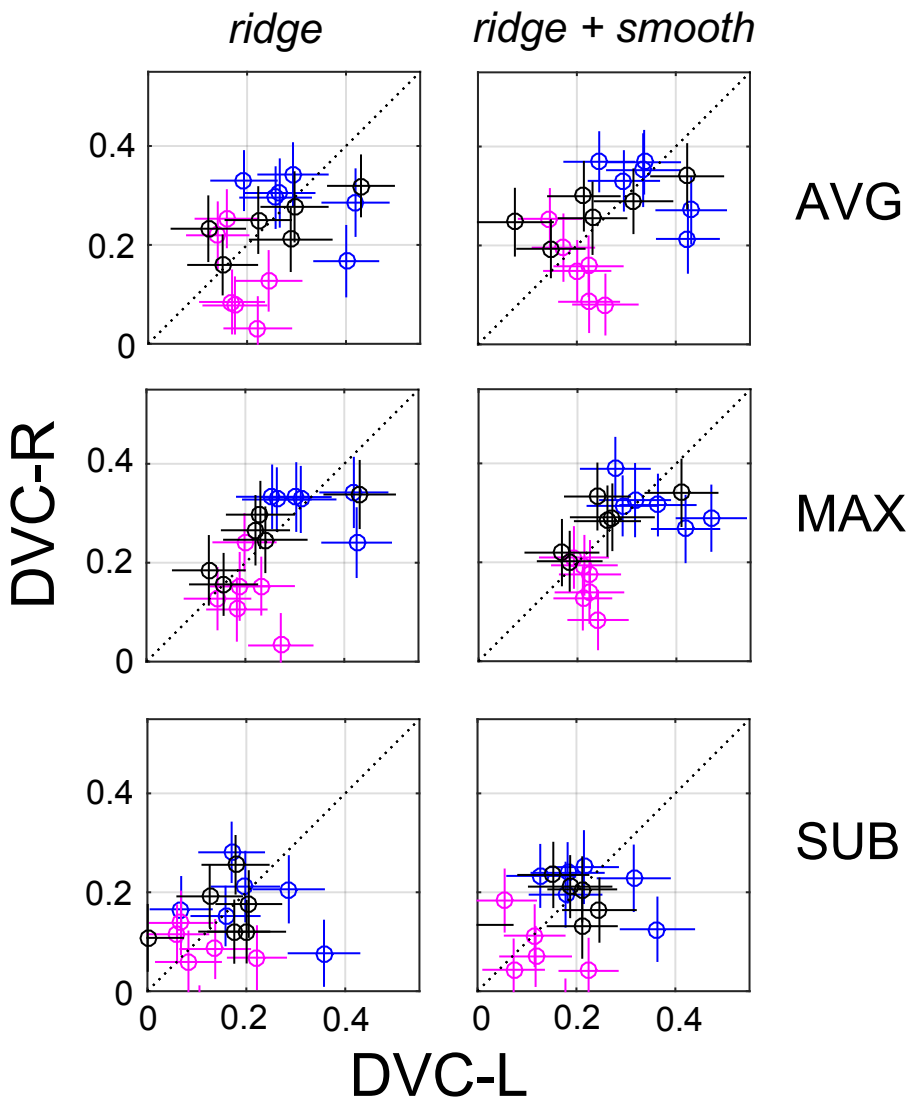


Figure 11

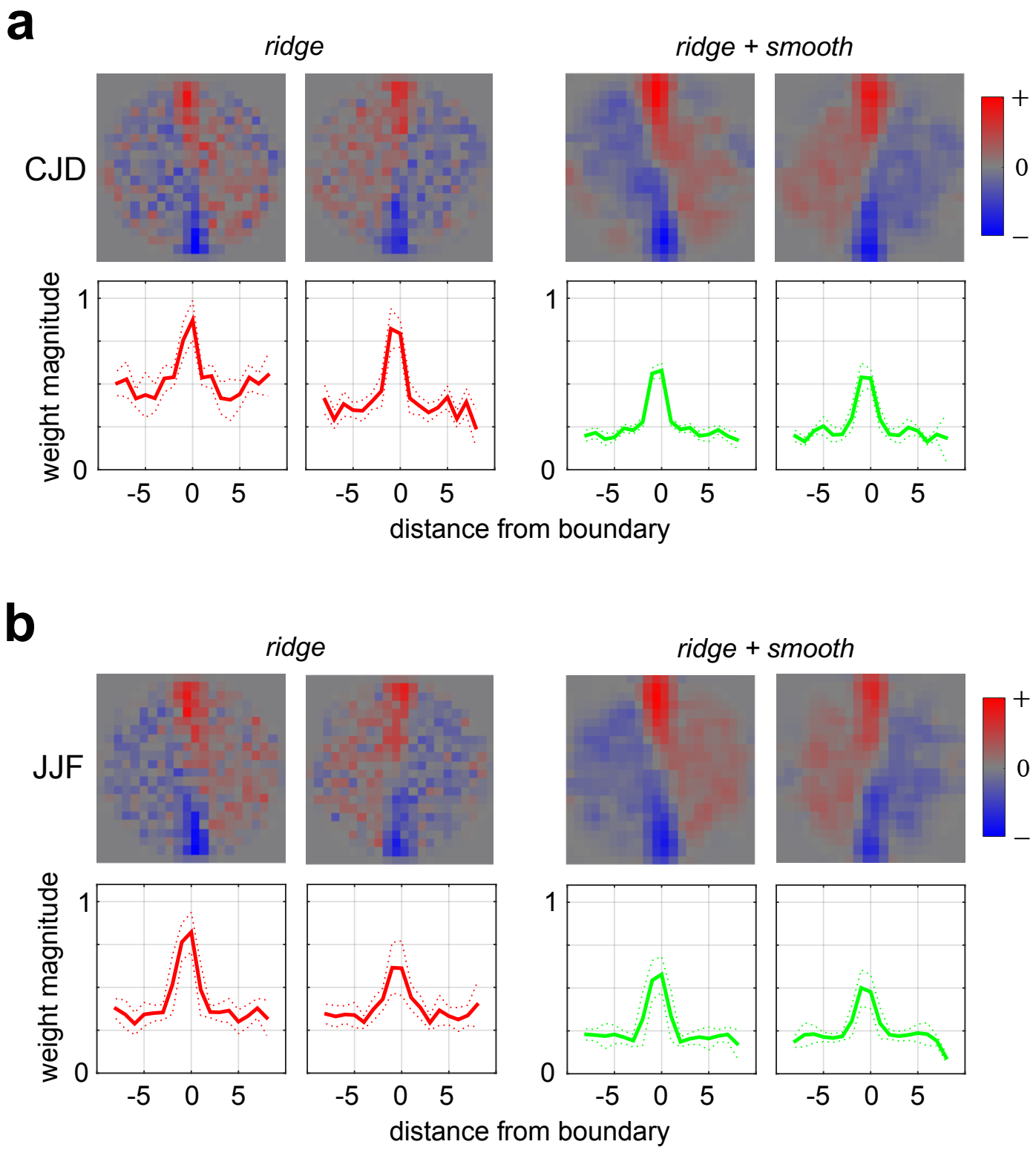

a
ridge
ridge + smooth
CJD
+
0
–
weight magnitude
1
0
-5
0
5
distance from boundary
b
ridge
ridge + smooth
JJF
+
0
–
weight magnitude
1
0
-5
0
5
distance from boundary

Figure 12

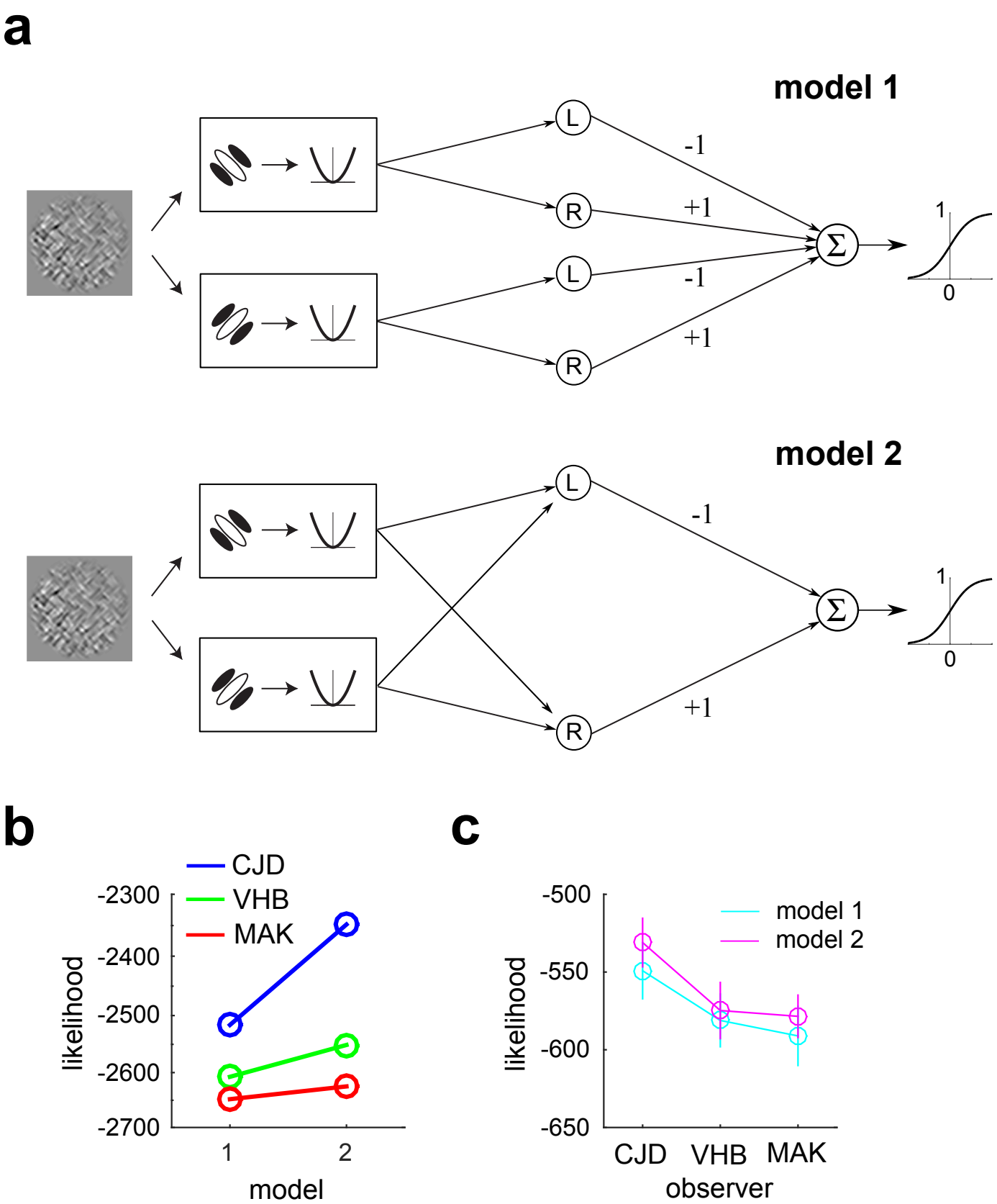
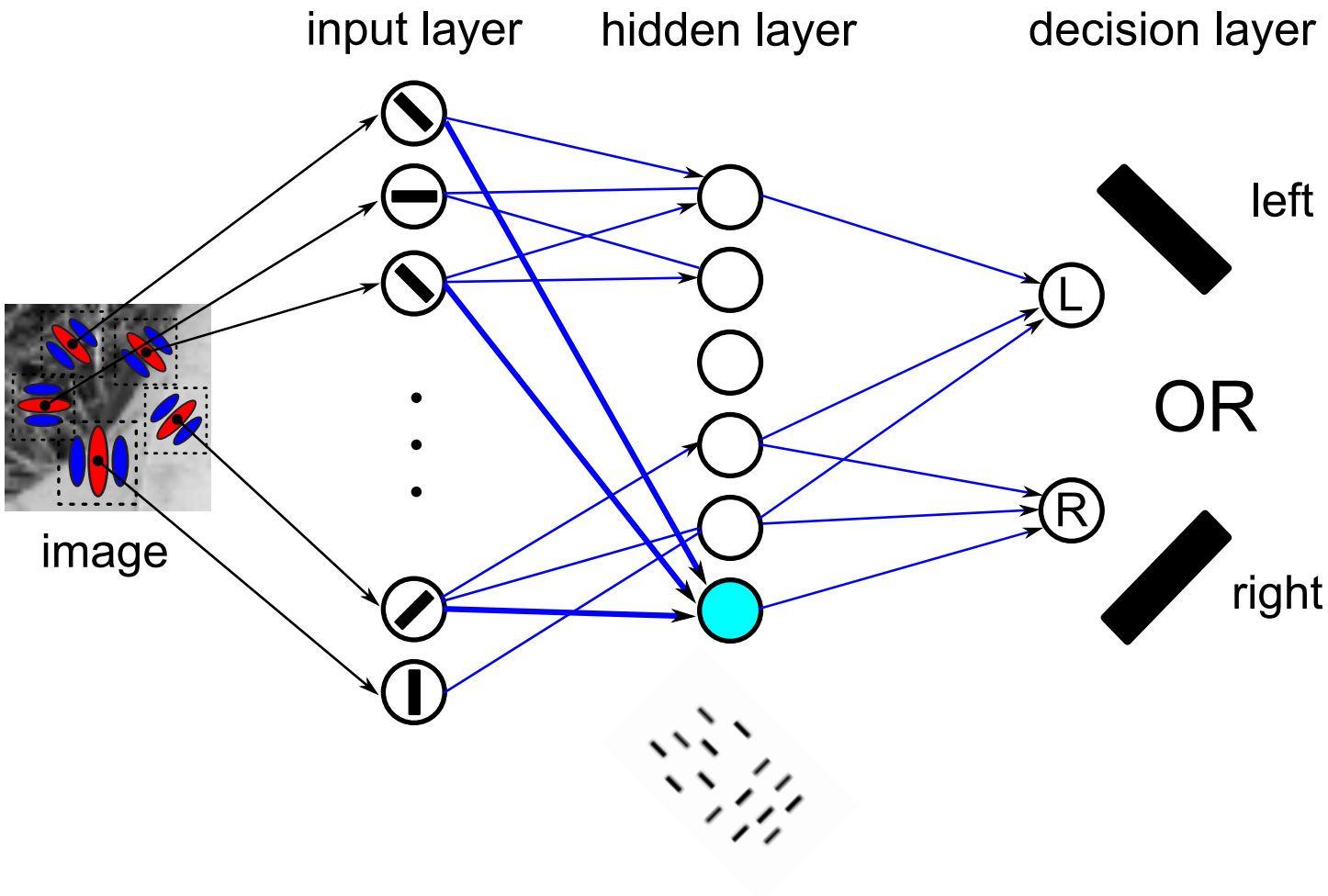

a
model 1
L
R
L
R
-1
+1
-1
+1
Σ
1
0
model 2
L
R
-1
+1
Σ
1
0
b
CJD
VHB
MAK
-2300
-2400
-2500
-2600
-2700
likelihood
1
2
model
c
model 1
model 2
-500
-550
-600
-650
likelihood
CJD
VHB
MAK
observer

Figure 13


input layer
hidden layer
decision layer
left
L
OR
R
right
image