

Feedback and Surround Modulated Boundary Detection

Arash Akbarinia · C. Alejandro Parraga

Received: 09/01/2017 / Accepted: date

Abstract Edges are key components of any visual scene to the extent that we can recognise objects merely by their silhouettes. The human visual system captures edge information through neurons in the visual cortex that are sensitive to both intensity discontinuities and particular orientations. The “classical approach” assumes that these cells are only responsive to the stimulus present within their receptive fields, however, recent studies demonstrate that surrounding regions and inter-areal feedback connections influence their responses significantly. In this work we propose a biologically-inspired edge detection model in which orientation selective neurons are represented through the first derivative of a Gaussian function resembling double-opponent cells in the primary visual cortex (V1). In our model we account for four kinds of receptive field surround, *i.e.* full, far, iso- and orthogonal-orientation, whose contributions are contrast-dependant. The output signal from V1 is pooled in its perpendicular direction by larger V2 neurons employing a contrast-variant centre-surround kernel. We further introduce a feedback connection from higher-level visual areas to the lower ones. The results of our model on three benchmark datasets show a big improvement compared to the current non-learning and biologically-inspired state-of-the-art algo-

rithms while being competitive to the learning-based methods.

Keywords boundary detection · surround modulation · biologically-inspired vision

1 Introduction

Our ability to recognise objects is completely entangled with our ability to perceive contours (Walther et al, 2011; Papari and Petkov, 2011). The primary and secondary visual cortices – *i.e.* V1 and V2 – play a crucial role in the process of detecting lines, edges, contours, and boundaries (Loffler, 2008), to such extent that an injury to these areas can impair a person’s ability to recognise objects (Zeki, 1993). Furthermore, edges (a form of image gradient sometimes also referred to as “boundaries” or “contours”) are indispensable components of computer vision algorithms in a wide range of applications, such as colour constancy (Van De Weijer et al, 2007), image segmentation (Arbelaez et al, 2011), document recognition (LeCun et al, 1998) and human detection (Dalal and Triggs, 2005).

Given their importance, many computational models have been proposed to detect edges – for a comprehensive review refer to (Papari and Petkov, 2011). In its earliest form a convolutional-based image gradient was proposed to capture local changes (Prewitt, 1970). Others attributed edges to zero-crossing points, therefore suggesting the Laplacian-of-Gaussian as a suitable operator (Marr and Hildreth, 1980). Previous algorithms were improved by incorporating non-maximum suppression and hysteresis thresholding (Canny, 1986). The greatest challenge faced by these classical methods is the distinction between authentic boundaries and undesired background textures. This issue was partially

This work was funded by the Spanish Secretary of Research and Innovation (TIN2013-41751-P and TIN2013-49982-EXP) and the CERCA Programme from the Generalitat de Catalunya.

Arash Akbarinia
Centre de Visió per Computador
Universitat Autònoma de Barcelona
Barcelona, Spain
E-mail: arash.akbarinia@cvc.uab.es

C. Alejandro Parraga

addressed by local smoothing techniques, such as bilateral filtering (Tomasi and Manduchi, 1998) and mean shift (Comaniciu and Meer, 2002). Thereafter, graph-based models emerged, *e.g.* (Felzenszwalb and Huttenlocher, 2004; Cour et al, 2005), allowing for closure to be taken into account. More recent frameworks extract relevant cues (*e.g.* brightness, colour, and texture) feeding them to machine learning, such as probabilistic boosting tree (Dollar et al, 2006), gradient descent (Arbelaez et al, 2011) and structured forest (Dollár and Zitnick, 2015). Currently, state-of-the-art algorithms (Kivinen et al, 2014; Bertasius et al, 2015a; Shen et al, 2015; Bertasius et al, 2015b; Xie and Tu, 2015) rely heavily on deep-learning techniques.

Despite their success, learning methods have their own set of challenges and drawbacks (Domingos, 2012): (a) their performance might be dataset dependant; (b) they are computationally demanding since for every single pixel a decision must be made (in both training and testing stages) on whether it corresponds to an edge or not; and (c) they require extremely large amounts of data for an effective training procedure. In addition to these, there is no biological or behavioural evidence that edge detection is the result of such a laboriously supervised learning process. On the contrary, biological systems compute edges in an unsupervised manner, starting from low-level features that are modulated by feedback from higher-level visual areas, *e.g.* those responsible for global shape (Loffler, 2008). In line with that, one of the earliest biologically-inspired edge detection models (MIRAGE) was proposed based on a combination of the half-wave rectified outputs of all spatial-frequency-channels that were used as primitives for spatial analysis (Watt and Morgan, 1985). Another influential model considered edges as maxima of the local energy obtained by convolving an image with a quadrature pair of one-dimensional filters (Morrone and Burr, 1988). Further, more sophisticated models which allow variably weighted sums of oriented filters were introduced by (Georgeson and Meese, 1997). All these biologically-inspired models were based on the physiology and generally tested against psychophysical data (such as, edge detection, discrimination or illusion).

More recently, a number of biologically-inspired edge detection algorithms have evolved computationally to be tested on large datasets of real-world images with promising results. A predictive coding and biased competition mechanism was proposed by (Spratling, 2013) to model the sparsity coding of neurons in V1. Importance of non-classical receptive fields was presented by (Wei et al, 2013) in a butterfly-shaped inhibition model operating at multiple spatial scales. Further improvement came from (Yang et al, 2013) who explored

imbalanced colour opponency in order to detect luminance boundaries. The same authors demonstrated employing the spatial sparseness constraint, typical to V1 neurons, helps to reserve desired fine boundaries while suppressing unwanted textures (Yang et al, 2015). Another improvement in contour detection originated from introducing multiple features to the classical centre-surround inhibition common to most cortical neurons (Yang et al, 2014). Accounting for feedback connections has also been beneficial, *e.g.* (Díaz-Pernas et al, 2014) extracted edges through oriented Gabor filters accompanied with top-down and region enhancement feedback layers.

In this article we extend our previous work (Akbarinia and Parraga, 2016) to propose a biologically-inspired edge detection model that incorporates recent knowledge of the physiological and psychophysical properties of our visual system. The proposed model is novel compared to other biologically-inspired methods in four main aspects: (i) we incorporate a more sophisticated set of cortical interactions which includes four types of surround, *i.e.* full, far, iso- and orthogonal-orientation; (ii) we account for contrast variation of surround modulation; (iii) we model V2 neurons that pool signals from V1 responses over a larger region corresponding to the centre and neighbouring spatial locations; and (iv) we consider a fast-conducting feedback connection from higher visual areas to the lower ones.

Fig. 1 illustrates the flowchart of our framework, which follows the functional structure of the human ventral pathway. Our processing starts in the retina, where the input image is convolved by single opponent cells and sent through the lateral geniculate nucleus (LGN) in the form of colour opponent channels (Shapley and Hawken, 2011). These channels are processed by double-opponent cells in V1 – known to be responsive to colour edges (Shapley and Hawken, 2011) – whose receptive field (RF) are modelled through the first derivative of a Gaussian function (Canny, 1986). To consider the RF surround: we define a short range circular (isotropic) region corresponding to full surround (Loffler, 2008), long range iso- and orthogonal-orientation surrounds along the primary and secondary axes of the RF (Field et al, 2014), and we model far surround via feedback connections to enhance the saliency of edge features. All these interactions are inversely dependant on the contrast of the RF centre (Shushruth et al, 2009). The output signal from V1 is pooled at V2 by a contrast-variant centre-surround mechanism applied orthogonally to the preferred direction of the V1 RF (Poirier and Wilson, 2006). Finally, to account for the impact of global shapes on local contours (Loffler, 2008), we feed the output of V2 layer back into V1.

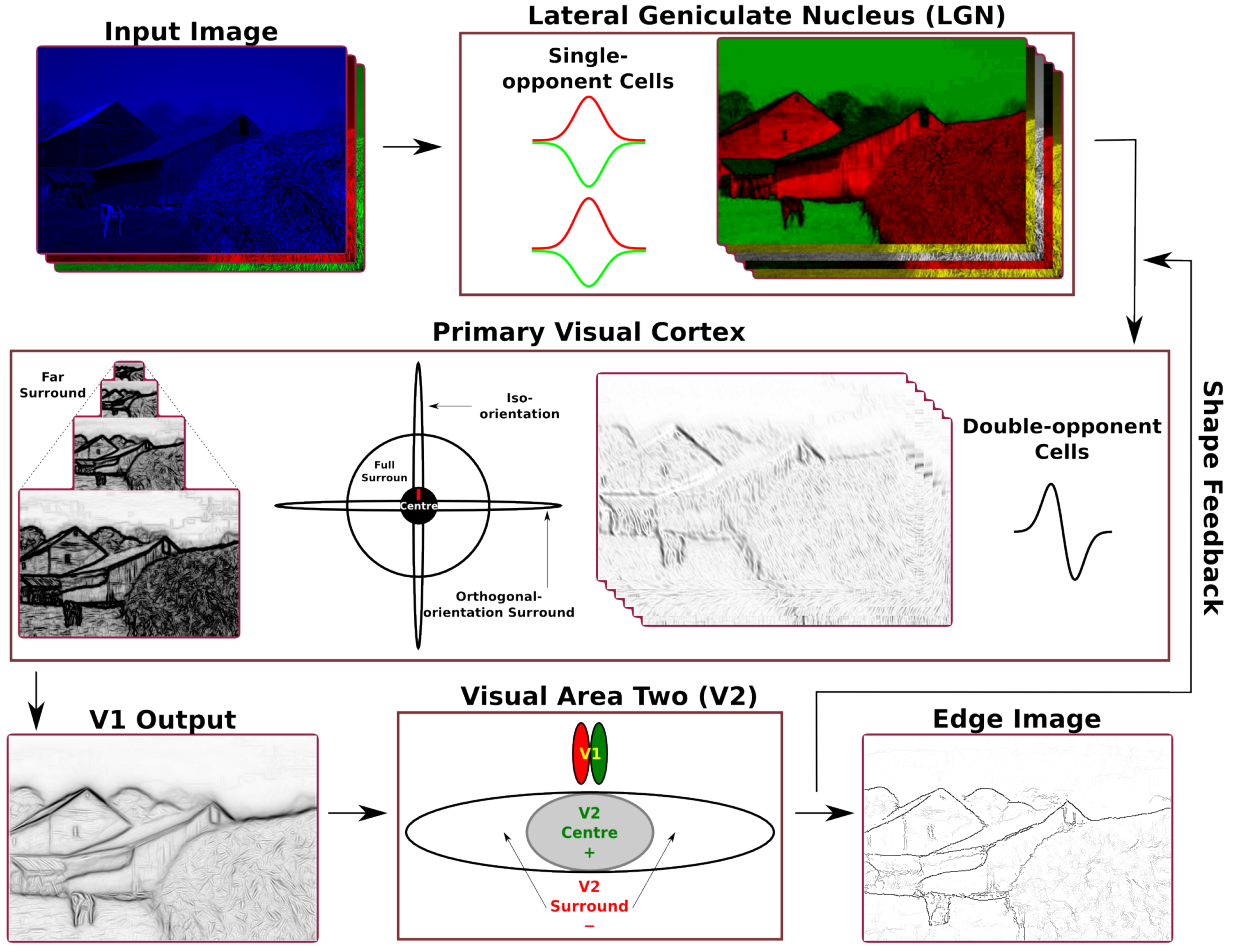


Fig. 1 The flowchart of our model. Balanced and imbalanced colour opponent channels are created in the retina and sent through the LGN. Orientation information is obtained in V1 by convolving the signal with a derivative of Gaussian at twelve different angles. We model four types of orientation-specific surround: full, far, iso- and orthogonal-orientation. In V2 the signal is further modified by input from surrounding areas in a directional orthogonal to that of the original RF. Shape feedback is sent to V1 as an extra channel.

2 Surround Modulation Edge Detection

2.1 Retina and lateral geniculate nucleus (LGN)

The retina is the starting point of visual processing in humans. Cone photoreceptor cells located at the back of the retina absorb photons at every spatial location. Their output is processed in an antagonistic manner by further layers of single-opponent cells (ganglion cells) and sent to the cortex through the LGN in the form of a luminance and two chromatically-opponent channels (Shapley and Hawken, 2011), usually modelled as

$$\begin{aligned} SO_{lu}(x, y) &= S_r(x, y) + S_g(x, y) + S_b(x, y), \\ SO_{rg}(x, y) &= \kappa_r S_r(x, y) - \kappa_g S_g(x, y), \\ SO_{yb}(x, y) &= \kappa_b S_b(x, y) - \kappa_{rg} \left(\frac{S_r(x, y) + S_g(x, y)}{2} \right), \end{aligned} \quad (1)$$

where SO represents the response of single-opponent cells, $\{lu, rg, yb\}$ denotes the luminance, red-green and yellow-blue opponent-channels, (x, y) are the spatial coordinates, and $\{r, g, b\}$ are the original red, green and blue cone signals. S is the spectral response function of each cone and can be approximated by a two dimensional Gaussian function as follows

$$S_h(x, y) = I_h(x, y) * G(x, y, \sigma), \quad (2)$$

where I is the input image, $h \in \{r, g, b\}$ is the index of colour channels, $*$ denotes the convolution operator and G is the circular two-dimensional Gaussian kernel, defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}, \quad (3)$$

with variance σ (set to 0.5 in this study as cells in LGN are substantially smaller than those of V1). This Gaus-

sian convolution is equivalent of a smoothing preprocessing stage in computer vision which has been demonstrated to play an important role in the successive edge detection (Papari and Petkov, 2011).

When the chromatically-opponent input to single-opponent cells in Eq. 1 is in equilibrium, parameter κ is equal to one for all channels. However, there is physiological evidence showing that some types of single-opponent cells combine chromatic information in an imbalanced fashion (Shapley and Hawken, 2011). The significance of these cells has also been shown practically in many computer vision algorithms, *e.g.* edge detection (Yang et al, 2013, 2015) and colour constancy (Gao et al, 2013; Parraga and Akbarinia, 2016). Following this insight, we included two imbalanced chromatic opponent-channels: $SO_{rg'}$ with $\kappa_g = 0.7$ and $SO_{yb'}$ with $\kappa_{rg} = 0.7$.

2.2 Primary visual cortex (V1)

Once the visual signal is preprocessed in the retina and the LGN, it is sent for further processing into the visual cortex. Early neurophysiological evidence established that the feedforward arrays coming from the LGN interact dynamically in the visual cortex, creating various gain control pools across all spatial orientations which can be modelled as “divisive normalisation” (Heeger, 1992). In this configuration, each cortical neuron computes a rectified combination of its inputs, followed by a normalisation where the neuron’s response is divided by the pooled activity of its neighbours. The overall effect of this gain normalisation is to both alter the contrast response of neural units, making them more responsive to boundaries, and to narrow their orientation bandwidths. Another mechanism contributing to orientation tuning stems from the long-range connections between neurons with similar orientation “collinear facilitation” (Malach et al, 1993). Both mechanisms are thought to enhance contour continuity, altering the effective orientation tuning of cells (Hansen and Neumann, 2008).

Although divisive normalisation and collinear facilitation are powerful mechanisms, recent studies have shown that they are likely to be oversimplifications, since stimuli outside of the classical receptive field of a cortical neuron can also modulate that neuron’s activity in various ways. The origin of this modulation is feedforward, feedback and lateral, stemming from previous connections, later connections and from neighbouring neurons in the visual pathway. However, it was not until the mid-1980s that the concept of non-classical (surround-modulated) receptive field became established

and characterised using circular or annular gratings of varying characteristics.

Now we understand that *SO* channels arriving at the cortex are processed by a number of double-opponent cells in V1 that are responsive to boundaries (Shapley and Hawken, 2011), but also modulated by regions beyond their RF centres, with facilitation predominantly at low contrast and inhibition at high contrast (Kapadia et al, 1999; Shushruth et al, 2009; Angelucci and Shushruth, 2014). As a consequence of the above, we defined the receptive field of our orientation-tuned double-opponent cells *DO* as

$$DO_c(x, y, \theta) = CR_c(x, y, \theta) + \zeta_c^{-1}(x, y)SR_c(x, y, \theta) \quad (4)$$

where c is the index of *SO* channels, θ is the preferred orientation of the RF (set to twelve evenly distributed angles in the range of $[0, 2\pi)$ radians), CR and SR are the centre and surround responses respectively, and ζ is the contrast of the RF centre approximated by the local standard deviation of its constituent pixels. Double-opponent cells are typically modelled in biologically-inspired algorithms by Gabor filters, (Spratling, 2013; Yang et al, 2014; Díaz-Pernas et al, 2014), or the first derivative of a Gaussian function, (Yang et al, 2013, 2015). We settled for the later one originally proposed by (Canny, 1986), therefore, we defined the *DO* centre response, CR , as

$$CR(x, y, \theta) = SO * \left| \frac{\partial G(x, y, \sigma)}{\partial \theta} \right|, \quad (5)$$

where σ is the RF size (set to 1.5 corresponding to the typical RF size of foveally-connected neurons in V1 (Angelucci and Shushruth, 2014), *i.e.* 0.25° of visual angle that is equivalent to approximately 13 pixels when viewed from 100cm in a standard monitor).

2.2.1 Surround modulation

We defined the surround response, SR , as follows

$$SR(x, y, \theta) = LS(x, y, \theta) + IS(x, y, \theta) + OS(x, y, \theta) + FS(x, y, \theta), \quad (6)$$

where LS is full surround referring to the isotropic region around the RF; IS denotes iso-orientation surround along the RF preferred axis; OS is orthogonal-orientation surround in the direction perpendicular to the RF preferred axis; and FS denotes far surround.

The full surround is an isotropic region, *i.e.* stimulus occupying the entire surrounding region rather than isolated flanking lines (Loffler, 2008). Due to this, the full surround can be modelled as the average response of a circular window around the cell’s RF centre. This

surround is inhibitory when it shares the same orientation as the centre and strongly facilitatory when its orientation is perpendicular to the centre (Loffler, 2008). Thus, we defined the full surround LS as

$$LS(x, y, \theta) = \lambda \zeta^{-1}(x, y) \left(CR(x, y, \theta_{\perp}) * \mu \right) - \zeta(x, y) \left(CR(x, y, \theta) * \mu \right), \quad (7)$$

where $\theta_{\perp} = \theta + \frac{\pi}{2}$, μ is the circular average kernel and λ determines the strength of orthogonal facilitation in comparison to the iso-inhibition. The former facilitation is reported to be stronger than the later inhibition (Loffler, 2008), therefore λ must be larger than one.

The iso-orientation surround, IS , extends to a distance two to four times larger than the RF size (Field et al, 2014). Within this region elements parallel to the RF preferred orientation are facilitatory while orthogonal ones are inhibitory (Loffler, 2008; Field et al, 2014), therefore, we modelled IS as

$$IS(x, y, \theta) = \zeta^{-1}(x, y) \left(CR(x, y, \theta) * E(\sigma_x, \theta) \right) - \zeta(x, y) \left(CR(x, y, \theta_{\perp}) * E(\sigma_x, \theta) \right), \quad (8)$$

where E is an elliptical Gaussian function elongated in the direction θ , defined as

$$E(x, y, \sigma_x, \sigma_y, \theta) = e^{-(ax^2 - 2bxy + cy^2)}, \quad \text{with} \\ a = \frac{\cos^2 \theta}{2\sigma_x^2} + \frac{\sin^2 \theta}{2\sigma_y^2}, \quad b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2}, \\ c = \frac{\sin^2 \theta}{2\sigma_x^2} + \frac{\cos^2 \theta}{2\sigma_y^2}.$$

We set $\sigma_y = 0.1\sigma_x$ and $\sigma_x = 3\sigma$ corresponding to physiological measurements (Field et al, 2014).

The orthogonal-orientation surround, OS , projects to a distance half of the iso-orientation surround (Field et al, 2014). In the orthogonal-surround elements parallel to the RF preferred orientation are inhibitory while perpendicular ones are facilitatory (Loffler, 2008; Field et al, 2014), thus, we modelled OS as

$$OS(x, y, \theta) = \zeta^{-1}(x, y) \left(CR(x, y, \theta_{\perp}) * E(\sigma_x, \theta_{\perp}) \right) - \zeta(x, y) \left(CR(x, y, \theta) * E(\sigma_x, \theta_{\perp}) \right). \quad (9)$$

The far surround could extend to regions up to 12.5° of visual angle (Shushruth et al, 2009) which is approximately equivalent to 673 pixels when viewed from 100cm in a standard monitor. Consequently the feed-forward and horizontal connections in V1 that mediate interactions between the RF and its near surround are too slow to account for the fast onset of far surround. Due to this, it has been suggested that far surround is

operated through a different mechanism via inter-areal feedback connections (Shushruth et al, 2013; Angelucci and Shushruth, 2014). We speculate that parts of these inter-areal connections come from spatial scale layers in V1 (Hess, 2014), and assume their influence to be facilitatory when image elements in this region share the same orientation as the centre (Ichida et al, 2007). Therefore, we defined FS as

$$FS(x, y, \theta) = \zeta^{-1}(x, y) \sum_{s=2}^4 \frac{CR_s(x, y, \theta)}{s} \quad (10)$$

where s is the index of the corresponding spatial frequency scale. This processing is analogous to the multi-scale processing common to both visual sciences and computer vision, with the distinction that we account for both contrast and distance, since surround modulation has been reported to be stronger in the near than in the far regions (Angelucci and Shushruth, 2014).

2.3 Visual area two (V2)

Visual processing becomes more global along the brain's ventral pathway, where neurons in each consecutive area seem to pool information from increasingly larger spatial regions (i.e. exponentially larger receptive fields). This allow them to process increasingly complex image features, such as curved arcs, angles, and line intersections and eventually shapes and objects. The next interconnected adjacent area to V1 is V2, where many neurons have been reported to respond to curvatures and extended arcs (Wilson and Wilkinson, 2014). It has been proposed that RFs in area V2 extract curvature information by pooling signals from V1 using a centre-surround mechanism in the direction orthogonal to the V1 orientations (Wilson and Wilkinson, 2014; Poirier and Wilson, 2006). In order to model this, first, we defined the V1 response, $V1R$, as the most activated DO orientation. This operation is assumed to be realised by complex cells pooling the maximum value of DO cells (Thériault et al, 2015), modelled as

$$V1R_c(x, y) = \arg \max_{\theta \in [0, 2\pi)} (DO_c(x, y, \theta)). \quad (11)$$

The V2 RFs show similar contrast-variant surround modulation as those of V1 (Shushruth et al, 2009). Therefore, we modelled the V2 response, $V2R$, through a Difference-of-Gaussians (DoG) as

$$V2R_c(x, y) = V1R_{c,\theta}(x, y) * E(\sigma_x, \theta_{\perp}) - v_c(x, y) V1R_{c,\theta}(x, y) * E(5\sigma_x, \theta_{\perp}) \quad (12)$$

where v is the contrast of $V1R$ computed by its local standard deviation, the index θ at $V1R_{\theta}$ shows the

preferred orientation of that RF. Cortical RFs increase their diameters systematically by approximately a factor of three from lower to higher areas (Wilson and Wilkinson, 2014). Therefore, we set the size of V2 RF, σ_x , to three times the size of a V1 RF. In Eq. 12 surround is five times larger than the centre according to physiological findings (Shushruth et al, 2009).

2.4 Feedback connections

In the primate visual system there are generally massive feedback connections from higher visual areas into lower ones (Angelucci and Shushruth, 2014). For instance, the majority of the LGN inputs are feedback connections from other areas of the brain, in particular the visual cortex. The functional role of this cortical feedback in visual processing is still poorly understood, although new evidence shows that these projections are organised into parallel streams and their effects include tune-sharpening, gain-modulation and various adjustments to behavioural demands (Briggs and Martin, 2014).

In our model we accounted for only a fraction of the feedback from V2 to V1 corresponding to the well established fact that global shape influences local contours (Loffler, 2008). We simulated this global shape by averaging the V2 outputs of all channels and sending it as feedback to V1. This feedback is processed as all other inputs to V1. The final edge map is created as a sum of all V2 output channels

$$\text{edge}(x, y) = \sum_c V2R_c(x, y), \quad \text{with} \quad (13)$$

$$c \in \{lu, rg, yb, rg', yb', \text{feedback}\}.$$

3 Experiments and results

We tested our model – termed *Surround-modulation Edge Detection (SED)* – on three benchmark datasets¹: (i) the Berkeley Segmentation Dataset and Benchmark (BSDS) (Martin et al, 2001; Arbelaez et al, 2011), (ii) the Multi-cue Boundary Detection Dataset (MBDD) (Mély et al, 2016), and (iii) the Contour Image Database (CID) (Grigorescu et al, 2003). Each image of all three datasets is supplemented with a ground truth that is created from manually-drawn edges by number of human subjects. We evaluated our algorithm in the standard precision-recall curve based on its harmonic mean (referred to as F-measure) on three criteria: optimal scale for the entire dataset (ODS) or per image (OIS)

¹ The source code and all the experimental materials are available at <https://github.com/ArashAkbarinia/BoundaryDetection>.

and average precision (AP). Naturally, ODS is the most representative of all to measure the performance since it uses a fixed threshold for all images in the dataset (Arbelaez et al, 2011). The results we report in this paper were obtained with a fixed set of parameters (see details in Section 2) for all datasets much in the same way as the human visual system.

3.1 Berkeley Segmentation Dataset and Benchmark (BSDS)

The BSDS (Martin et al, 2001; Arbelaez et al, 2011) contains two sets of colour images BSDS300 (100 test images and 200 training images) and BSDS500 (200 test images). This dataset contains a wide range of natural and man-made objects. Size of each image is 481×321 pixels. Arguably BSDS is considered as the benchmark dataset for boundary detection in the field of computer vision.

Table 1 compares the results of our model to several other state-of-the-art algorithms that have also reported their results on the BSDS dataset. From this table we can observe that in BSDS500 our model improves the ODS of methods driven by low-level and biological features by 4%. This improvement is 3% in BSDS300. It must be noted that deep-learning methods often use BSDS300 as the training set and therefore they do not report their results on this fragment of BSDS.

In order to study the robustness of different algorithms to achromatic images, we conducted a further experiment on the grey-scale version of BSDS images. The results of this experiment for our model along with five other algorithms driven by low-level features and one learning method (whose source code were publicly available) are presented in Table 2. We can observe similar patterns as chromatic images: our proposal enhances ODS by 3% in both BSDS300 and BSDS500. It is worth highlighting that performance of the learning-based gPb significantly drops in grey-scale images. Our model obtains a 4% higher ODS comparing to gPb in both BSDS300 and BSDS500.

3.2 Multi-cue Boundary Detection Dataset (MBDD)

The MBDD (Mély et al, 2016) composed of short binocular video sequences of natural scenes. This dataset contains challenging scenes for boundary detection by framing a few dominant objects in each shot under a variety of appearances. Size of each image is 1280×720 pixels. The dataset contains 100 scenes and offers two sets of hand-annotations: one for *object boundaries* and another for *lower-level edges*.

			BSDS300			BSDS500		
Method			ODS	OIS	AP	ODS	OIS	AP
Biological	Low-level features	Human	0.79	0.79	–	0.80	0.80	–
		Canny (Canny, 1986)	0.58	0.62	0.58	0.60	0.63	0.58
		Mean Shift (Comaniciu and Meer, 2002)	0.63	0.66	0.54	0.64	0.68	0.56
		Felz-Hutt (Felzenszwalb and Huttenlocher, 2004)	0.58	0.62	0.53	0.61	0.64	0.56
		Normalised Cuts (Cour et al, 2005)	0.62	0.66	0.43	0.64	0.68	0.45
		PC/BC (Spratling, 2013)	0.61	–	–	–	–	–
		CO (Yang et al, 2013)	0.64	0.66	0.65	0.64	0.68	0.64
		MCI (Yang et al, 2014)	0.62	–	–	0.64	–	–
		dPREEN (Díaz-Pernas et al, 2014)	0.65	–	–	–	–	–
		SCO (Yang et al, 2015)	0.66	0.68	0.70	0.67	0.71	0.71
Deep-learning	Machine-learning	BEL (Dollar et al, 2006)	0.65	–	–	0.61	–	–
		gPb (Arbelaez et al, 2011)	0.70	0.72	0.66	0.71	0.74	0.65
		DeepNets (Kivinen et al, 2014)	–	–	–	0.74	0.76	0.76
		DeepEdge (Bertasius et al, 2015a)	–	–	–	0.75	0.75	0.80
		DeepContour (Shen et al, 2015)	–	–	–	0.76	0.77	0.80
		HFL (Bertasius et al, 2015b)	–	–	–	0.77	0.79	0.80
		HED (Xie and Tu, 2015)	–	–	–	0.78	0.80	0.83
		SED (Proposed)	0.69	0.71	0.71	0.71	0.74	0.74

Table 1 Results of several edge detection algorithms on the BSDS300 and BSDS500 (Martin et al, 2001; Arbelaez et al, 2011).

			BSDS300			BSDS500		
Method			ODS	OIS	AP	ODS	OIS	AP
Canny	(Canny, 1986)		0.58	0.62	0.53	0.60	0.63	0.54
PC/BC	(Spratling, 2013)		0.61	0.63	0.40	0.64	0.65	0.41
CO	(Yang et al, 2013)		0.60	0.63	0.60	0.61	0.64	0.61
MCI	(Yang et al, 2014)		0.62	0.64	0.55	0.64	0.66	0.56
SCO	(Yang et al, 2015)		0.62	0.64	0.64	0.63	0.67	0.66
gPb	(Arbelaez et al, 2011)		0.61	0.64	0.60	0.63	0.66	0.62
SED	(Proposed)		0.65	0.67	0.68	0.67	0.70	0.70

Table 2 Results of several edge detection algorithms on the grey-scale images of BSDS300 and BSDS500 (Martin et al, 2001; Arbelaez et al, 2011).

We have reported the results of our model along with five algorithms driven by low-level features and one learning method for both types of annotations in Table 3. In comparison to the non-learning methods, we can observe that the proposed model improves the ODS by 3% in case of *object boundaries*. With respect to the *lower-level edges* this improvement is 2%. Our results are also better than the learning-based gPb for both types of annotations.

We believe that the *object boundaries* annotation is more relevant to the problem we are addressing in this paper since the *low-level edges* annotation contains many uninformative line segments from small objects (e.g. leaves and grass). This is evident in an exemplary sample illustrated in Fig. 2.

Similar to BSDS, in order to study the role of colour on each algorithm, we performed an experiment on the grey-scale images of MBDD. Table 4 shows the results of this experiment. SED still performs better than other algorithms driven by low level features (1% ODS im-



Fig. 2 Comparison of *object boundaries* and *low-level edges* annotations of MBDD (Mély et al, 2016).

provement in both types of annotations). In comparison to the learning-based gPb we obtain the same ODS in *object boundaries*, however SED performs slightly better in *low-level edges*. A surprising detail emerges when the results of CO or SCO for colour images is compared to the grey-scale ones; both algorithms perform slightly better in absence of colour (see Tables 3 and 4). This suggests imbalanced *colour opponency* require more careful implementation. We speculate this might also be the reason that our improvement in the grey-scale images of MBDD falls to minimal. This issue can be addressed in future studies.

Method		Object Boundaries			Low-level Edges		
		ODS	OIS	AP	ODS	OIS	AP
Canny	(Canny, 1986)	0.61	0.65	0.54	0.75	0.78	0.76
PC/BC	(Spratling, 2013)	0.69	0.70	0.43	0.80	0.81	0.70
CO	(Yang et al, 2013)	0.64	0.67	0.66	0.77	0.80	0.83
MCI	(Yang et al, 2014)	0.69	0.70	0.70	0.77	0.77	0.66
SCO	(Yang et al, 2015)	0.68	0.71	0.72	0.79	0.82	0.86
gPb	(Arbelaez et al, 2011)	0.71	0.72	0.70	0.78	0.81	0.82
SED	(Proposed)	0.72	0.74	0.77	0.81	0.83	0.86

Table 3 Results of several edge detection algorithms on the MBDD (Mély et al, 2016), for two ground truth annotations of *object boundaries* and *low level edges*.

Method		Object Boundaries			Low-level Edges		
		ODS	OIS	AP	ODS	OIS	AP
Canny	(Canny, 1986)	0.60	0.65	0.53	0.74	0.78	0.76
PC/BC	(Spratling, 2013)	0.68	0.69	0.43	0.79	0.82	0.69
CO	(Yang et al, 2013)	0.65	0.67	0.67	0.77	0.80	0.83
MCI	(Yang et al, 2014)	0.69	0.70	0.67	0.73	0.73	0.59
SCO	(Yang et al, 2015)	0.69	0.71	0.73	0.79	0.82	0.83
gPb	(Arbelaez et al, 2011)	0.70	0.71	0.71	0.78	0.81	0.82
SED	(Proposed)	0.70	0.71	0.74	0.80	0.82	0.86

Table 4 Results of several edge detection algorithms on the grey-scale images of MBDD (Mély et al, 2016), for two ground truth annotations of *object boundaries* and *low level edges*.

3.3 Contour Image Database (CID)

The CID (Grigorescu et al, 2003) contains 40 grey-scale images of natural scenes and animal wildlife. Size of each image is 512×512 pixels. Table 5 compares the results of SED to five algorithms driven by low-level features and one learning method on this dataset. We can observe that SED exceeds other methods by 5% ODS improvement. It is worth highlighting that the learning-based gPb scores 7% lower than the proposed model.

Method	CID		
	ODS	OIS	AP
Canny (Canny, 1986)	0.56	0.64	0.57
PC/BC (Spratling, 2013)	0.58	0.62	0.42
CO (Yang et al, 2013)	0.55	0.63	0.57
MCI (Yang et al, 2014)	0.60	0.63	0.53
SCO (Yang et al, 2015)	0.58	0.64	0.61
gPb (Arbelaez et al, 2011)	0.57	0.61	0.54
SED (Proposed)	0.65	0.69	0.68

Table 5 Results of seven edge detection algorithms on the CID dataset (Grigorescu et al, 2003).

3.4 Component analysis

In our algorithm we have modelled different areas and aspects of the visual cortex. In order to investigate

the contribution of each component of our model, we conducted four additional experiments on the BSDS dataset:

- **Gaussian Derivative** – In this scenario, we accounted neither for the surround modulation in V1, nor for the V2 pooling and feedback. Essentially only convolving the single-opponent cells with the first derivative of Gaussian function similar to CO (Yang et al, 2013).
- **Only V1 Surround** – In this case, we excluded V2 pooling and feedback. We only included full, far, iso- and orthogonal-orientation surround modulation for V1 RFs.
- **No V2 Feedback** – In this scenario, we excluded the shape feedback sent from area V2 to V1, *i.e.* $c \in \{lu, rg, yb, rg', yb'\}$ in Eq. 13.
- **No Far surround** – In this case, we did not account for far surround modulation, *i.e.* $FS = 0$ in Eq. 6.

The precision-recall curves of these experiments for BSDS300 and BSDS500 are shown in Fig. 4. Edge outputs of different components of our algorithm along with the full model on a few exemplary images are illustrated in Fig. 5.

4 Discussion

Quantitative results of conducted experiments on three benchmark datasets (*i.e.* BSDS, MBDD and CID), for

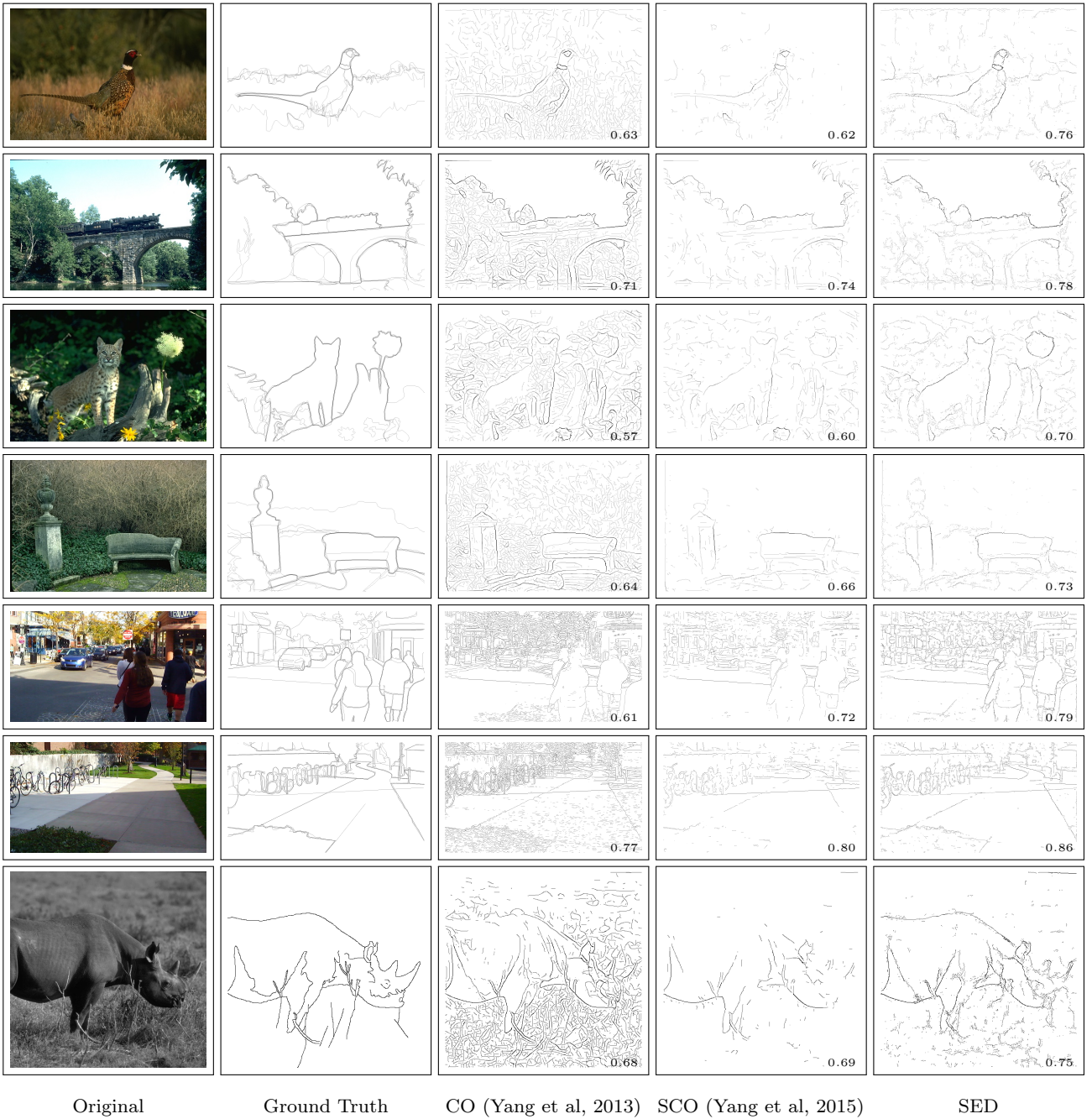


Fig. 3 Qualitative results of three biologically-inspired methods. The F-measure is indicated on the right bottom corner. The first two rows belong to BSDS300, the third and forth to BSDS500, the sixth and seventh to MBDD, and the last row to CID.

coloured and grey-scale images, demonstrate that SED systematically improves state-of-the-art for the category of non-learning based models. Our proposed model outperforms other methods driven by low-level features and biologically-inspired algorithms in all three criteria of ODS, OIS and AP (see Tables 1–5 in Section 3). This enhancement is also qualitatively pronounced in exemplary results shown in Fig. 3. On the one hand, our model shows greater robustness in textural areas

in comparison to CO, on the other hand, thanks to its surround modulation, SED performs better at detecting continuous lines compared to SCO. For instance, in the first row of the Fig. 3, it is evident that CO is strongly troubled with the textural information originating from the background vegetation, however SED successfully suppresses a large amount of those. At the same time, it is apparent that SCO blends the contours of the present bird with the straws, however SED

correctly extracts the boundaries of the bird from the grassland. We can observe similar patterns in the rest of the pictures of the Fig. 3.

Our improvements over state-of-the-art originates from the combination of different components of proposed model. The precision-recall curves in Fig. 4 shows that excluding surround modulation and the V2 module all together drops the ODS F-measure to 0.63 (black curves in both BSDS300 and BSDS500). This is in line with the results of CO which is essentially the same as our model in the absence of both V1 surround modulation and the V2 module. Including surround modulation (*i.e.* full, far, iso- and orthogonal-orientation regions) contributes to a significant enhancement of results by boosting the F-measure to 0.66 and 0.67 in BSDS300 and BSDS500 respectively (pink curves). This clearly shows that surrounding regions play a crucial role in the process of edge detection in agreement with previous psychophysical findings (Loffler, 2008). Importance of surround modulation has been even shown in convolutional neural networks, for example considering the contextual information from neighbouring pixels was reported to facilitate detection of objects' occluded boundaries (Fu et al, 2016). Qualitative comparison of the second and third columns of Fig. 5 suggests that although V1 surround modulation does not contribute to texture suppression, it strengthens continues contours (we have marked a few examples by the red and blue ovals, for instance the exterior borders of bricks in the last row are more continuous in the “*Only V1 Surround*” column in comparison to the “*Gaussian Derivative*” one, at the same time the intermedial borders are correctly suppressed in “*Only V1 Surround*” as a result of accounting for iso- and orthogonal-orientation surround modulation).

Comparison of “*Only V1 Surround*” and “*No V2 Feedback*” pictures in Fig. 5 reveals that the V2 module strongly assist the process of eliminating textural and noisy patches. This is consistent with physiological findings that suggest texture statistics might be encoded in V2 (Landy, 2014; Freeman et al, 2013). The robustness of our model to noisy scenes and undesired background textures could be explained by the fact that V2 RFs are large and therefore suppress small discontinuities across neighbouring pixels. Although V2 centre-surround suppression is beneficial in general with 1% (BSDS300) and 2% (BSDS500) improvements in F-measures (the red curves versus the pink ones in Fig. 4), it causes occasional over-smoothing and consequently in high recalls the precisions of the pink curves are higher than the red ones. We postulate that this problem can partially be addressed by accounting for a mechanism similar to the visual cortex where suppression can turn to facilitation

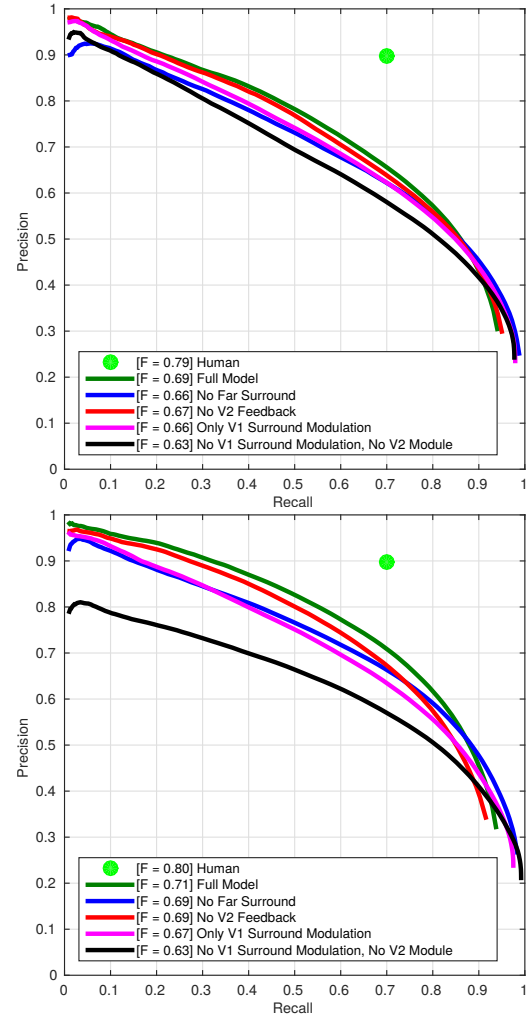


Fig. 4 Precision-recall curves of difference components of our model on the BSDS300 (top) and BSD500 (bottom). In the legends the ODS F-measures are indicated.

at low contrast levels (Angelucci and Shushruth, 2014). Modelling this phenomenon is onerous since the threshold between suppression and facilitation is cell specific and there is no universal contrast level or surround stimulus size that triggers facilitation across the entire cell population (Angelucci and Shushruth, 2014). Furthermore, neural recordings of macaque demonstrates that the activation level of V2 neurons are higher when exposed to naturalistic texture in comparison to spectrally matched noise (Freeman et al, 2013). This feature was not present among V1 neurons. This indicates that a more complex V2 model is required to treat noise and texture distinctively.

Excluding the global shape feedback from our model lowers the ODS F-measure by 2% (compare the green and red curves in Fig. 4). It is difficult to qualitatively appreciate the influence of this feedback connection in Fig. 5, however a closer comparison of the green ovals

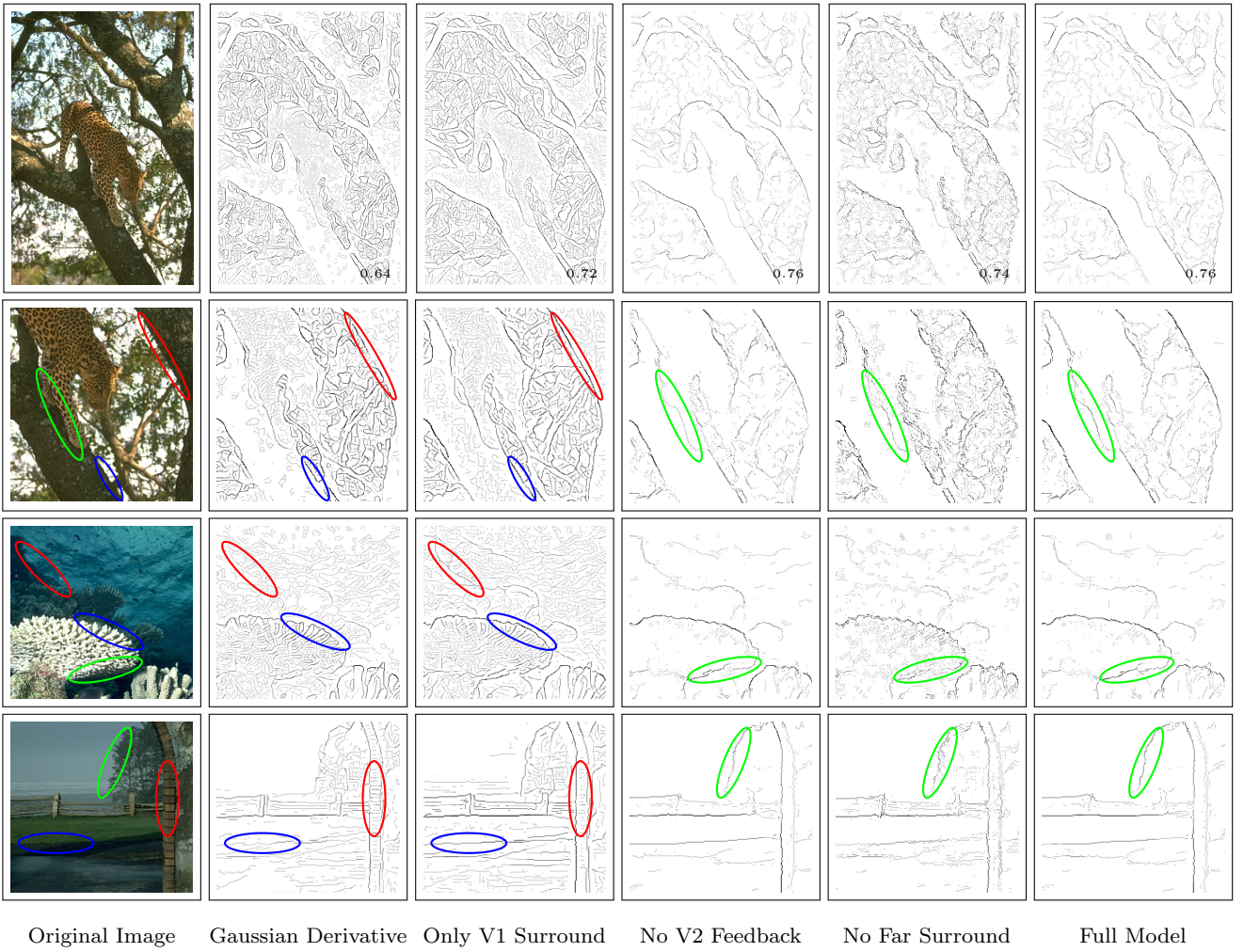


Fig. 5 Evaluation of the different components of SED. The images show the result of our full model on one exemplary image along with the four experiments we conducted. F-measures are on the right bottom corner of images.

in the “No V2 Feedback” and “Full Model” columns suggests that shape feedback re-enforce the true edges (the intensity of pixels along edges are higher in “Full Model” in comparison to their corresponding pairs in “No V2 Feedback”). This is in line with previously stabilised neurophysiological findings that show one of the functional roles of feedback connections is amplification and focus of neuronal activities in subsequent lower areas (Hupe et al, 1998). Correspondingly, in computer vision, iterative subroutines that are analogous to the role of feedback connections have been shown beneficial in recovering occluded boundaries by employing regional cues as criteria to eliminate weak edges while forming larger areas iteratively (Hoiem et al, 2007).

The precision-recall curves in Fig. 4 shows that excluding far surround modulation reduces the ODS F-measure to 0.66 and 0.69 in BSDS300 and BSDS500 respectively (blue curves), which still is better than other non-learning state-of-the-art algorithms. A qualitative

comparison of “No Far Surround” and “Full Model” results in Fig. 5 reveals that far surround appears to contribute in enhancing continuous edges while suppressing broken ones (e.g. the contours marked with green ovals in “No Far Surround” contain more abrupt alternate right and left turns in comparison to the “Full Model”, at the same time “No Far Surround” contains larger number of line fragments). Quantitatively, we observe a similar issue in far surround modulation to the V2 surround modulation: in high recalls “No Far Surround” has a higher precision than “Full Model” (blue versus green curves in Fig. 4). Resolving this is a subject for further investigation.

Computational Complexity

In principle, our model ought to be computationally very low cost since its building blocks are simple Gaus-

sian convolutions. With this in mind, we reported the average computational time of seven algorithms on the BSDS500 in Table 6 and to our surprise, the Matlab implementation of SED is rather slow. After a careful analysis of the different components of our model, we discovered that the *imfilter* function of Matlab is substantially slower when an image is convolved with an oriented elliptical Gaussian kernel across right angles. This is presumably due to the fact that *imfilter* is optimised for separable two-dimensional kernels and behaves significantly slower for non-separable ones. This turned out to be an important issue for our V2 RF surround modulation which uses a kernel of size 157×157 pixels computed for twelve orientations. Since OpenCV *filter2D* does not suffer from this problem, the C++ implementation of our model offers real-time processing. It is worth mentioning that we did not take advantage of any GPU programming in the C++ implementation. We believe our model can greatly benefit from the GPU parallel architecture due to the fact that its basic units are matrix operations.

Method	Time (s)
Canny (Canny, 1986)	0.54
PC/BC (Spratling, 2013)	$\gg 1800$
CO (Yang et al, 2013)	0.73
MCI (Yang et al, 2014)	21.00
SCO (Yang et al, 2015)	2.27
gPb (Arbelaez et al, 2011)	7.88
SED (Proposed)	7.45 (0.60)*

Table 6 Average computational time (in seconds) of seven edge detection algorithms driven by low-level features on the BSDS500 under the Matlab framework with Intel(R) Xeon(R) CPU E5-1620 v2 @ 3.70GHz. *C++ Implementation of our algorithm.

5 Conclusion

In this paper, we present a biologically-inspired edge detection model grounded on physiological findings of the visual cortex and psychophysical knowledge of visual perception. Our main contributions can be summarised as follows: (i) modelling a contrast-dependant surround modulation of V1 receptive fields by accounting for full, far, iso- and orthogonal-orientation surround; (ii) introducing a V2 centre-surround mechanism to pool V1 signals in their perpendicular orientations; and (iii) accounting for shape feedback connections from V2 to V1. We quantitatively compared the proposed model to current state-of-the-art algorithms on three benchmark datasets (on both colour and grey scale images) and

our results show a significant improvement compared to the other non-learning and biologically-inspired models while being competitive to the learning ones. Detailed analysis of different components of our model suggest that V1 surround modulation strengthen edges and continues lines while V2 module contributes to the suppression of undesired textural elements.

It must be acknowledged that within the proposed framework we have managed to only model a portion of what is known about surround modulation. The entire mechanism is considerably more complex. For instance, in our formulation, we simplified the interaction between surrounding regions by treating them as individual entities. However, from psychophysical studies it is evident that the non-linear interactions between surround and central regions are part of a multiplex mechanism determined by a number of factors (Spillmann et al, 2015), such as, (a) the configurations of both inducers and targets (Loffler, 2008), and (b) the spatial gap between target and surround (Tzvetanov and Dresch, 2002). Consequently, in a more comprehensive model these short and long interactions must be unified under one regime. Furthermore, it is recognised that the interaction across different areas of the cortex is a of dynamic nature and higher-level inputs can facilitate lower-level routines (Spillmann et al, 2015). Therefore, our model can benefit from a more entangled formulation of bottom-up and top-down processes. For instance, it is well established that perception of shape is significantly influenced by points where multiple edges meet, *e.g.* corners (Koenderink and Van Doorn, 1982). A solution to this border ownership dilemma can be sought by persisting the initial responses of neurons and utilising them in the later figure-ground discrimination (O’Herron and von der Heydt, 2011). At the same time, in future works, we should model the complex shape processing occurring in V4 (*e.g.* by concentric summation of V2 signals (Wilson and Wilkinson, 2014)), and consider its feedback connections to V1 and V2 with respect to phenomena such as filling-in surfaces and grouping.

Biologically-inspired solutions such as the one presented here make two contributions, one technological (advancing the state-of-the-art) and the other scientific (understanding the relationship between the human visual system and the visual environment). The more we learn about the properties of the human visual system the better we can explain visual behaviour. Within current limitations (both in knowledge and resources) we have tried to keep our modelling decisions as close as possible to what we know about the physiology, in three main respects: (a) our architecture reflects the low-level features that are common to mammalian cortical ar-

chitecture and emerged after millions of years of evolution (*i.e.* are not ad-hoc or dataset-dependant); (b) our model parameters are the same in all experiments, which is a feature of how the human visual system operates; and (c) our paradigm does not include supervised learning from large datasets, which is also a feature of how biological systems operate (their low-level features learning tends to be largely unsupervised).

Acknowledgements We would like to thank Dr. Xavier Otazu, members of the NeuroBiT Group at the Centre de Visió per Computador, Dr. Thorsten Hansen, and anonymous reviewers of the manuscript for their thoughtful insights and comments.

References

- Akbarinia A, Parraga CA (2016) Biologically plausible boundary detection. In: Proceedings of the British Machine Vision Conference (BMVC), BMVA Press
- Angelucci A, Shushruth S (2014) Beyond the classical receptive field: Surround modulation in primary visual cortex. Werner JS Chalupa L(Eds) The new visual neurosciences pp 425–444
- Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(5):898–916
- Bertasius G, Shi J, Torresani L (2015a) Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4380–4389
- Bertasius G, Shi J, Torresani L (2015b) High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In: Proceedings of the IEEE International Conference on Computer Vision, pp 504–512
- Briggs F, Martin U (2014) Functional properties of cortical feedback to the primate lateral geniculate nucleus. Werner JS Chalupa L(Eds) The new visual neurosciences pp 315–322
- Canny J (1986) A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6):679–698
- Comaniciu D, Meer P (2002) Mean shift: A robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(5):603–619
- Cour T, Benezit F, Shi J (2005) Spectral segmentation with multiscale graph decomposition. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, IEEE, vol 2, pp 1124–1131
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, IEEE, vol 1, pp 886–893
- Díaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D (2014) Double recurrent interaction v1–v2–v4 based neural architecture for color natural scene boundary detection and surface perception. Applied Soft Computing 21:250–264
- Dollár P, Zitnick CL (2015) Fast edge detection using structured forests. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37(8):1558–1570
- Dollar P, Tu Z, Belongie S (2006) Supervised learning of edges and object boundaries. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, IEEE, vol 2, pp 1964–1971
- Domingos P (2012) A few useful things to know about machine learning. Communications of the ACM 55(10):78–87
- Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. International Journal of Computer Vision 59(2):167–181
- Field DJ, Golden JR, Hayes A (2014) Contour integration and the association field. Werner JS Chalupa L(Eds) The new visual neurosciences pp 627–638
- Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. Nature neuroscience 16(7):974–981
- Fu H, Wang C, Tao D, Black MJ (2016) Occlusion boundary detection via deep exploration of context. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 241–250
- Gao S, Yang K, Li C, Li Y (2013) A color constancy model with double-opponency mechanisms. In: Proceedings of the IEEE International Conference on Computer Vision, pp 929–936
- Georgeson MA, Meese TS (1997) Perception of stationary plaids: The role of spatial filters in edge analysis. Vision research 37(23):3255–3271
- Grigorescu C, Petkov N, Westenberg MA (2003) Contour detection based on nonclassical receptive field inhibition. Image Processing, IEEE Transactions on 12(7):729–739
- Hansen T, Neumann H (2008) A recurrent model of contour integration in primary visual cortex. Journal of Vision 8(8):8–8
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Visual neuroscience 9(02):181–197

- Hess RF (2014) Spatial scale in visual processing. Werner JS Chalupa L(Eds) *The new visual neurosciences* pp 595–615
- Hoiem D, Stein AN, Efros AA, Hebert M (2007) Recovering occlusion boundaries from a single image. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, IEEE, pp 1–8
- Hupe J, James A, Payne B, Lomber S, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature* 394(6695):784–787
- Ichida JM, Schwabe L, Bressloff PC, Angelucci A (2007) Response facilitation from the suppressive receptive field surround of macaque v1 neurons. *Journal of Neurophysiology* 98(4):2168–2181
- Kapadia MK, Westheimer G, Gilbert CD (1999) Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proceedings of the National Academy of Sciences* 96(21):12,073–12,078
- Kivinen JJ, Williams CK, Heess N, Technologies D (2014) Visual boundary prediction: A deep neural prediction network and quality dissection. In: *AISTATS*, vol 1, p 9
- Koenderink JJ, Van Doorn AJ (1982) The shape of smooth objects and the way contours end. *Perception* 11(2):129–137
- Landy MS (2014) Texture analysis and perception. Werner JS Chalupa L(Eds) *The new visual neurosciences* pp 639–652
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324
- Loffler G (2008) Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research* 48(20):2106–2127
- Malach R, Amir Y, Harel M, Grinvald A (1993) Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proceedings of the National Academy of Sciences* 90(22):10,469–10,473
- Marr D, Hildreth E (1980) Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences* 207(1167):187–217
- Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, IEEE, vol 2, pp 416–423
- Mély DA, Kim J, McGill M, Guo Y, Serre T (2016) A systematic comparison between visual cues for boundary detection. *Vision research* 120:93–107
- Morrone MC, Burr D (1988) Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London B: Biological Sciences* 235(1280):221–245
- O’Herron P, von der Heydt R (2011) Representation of object continuity in the visual cortex. *Journal of vision* 11(2):12–12
- Papari G, Petkov N (2011) Edge and line oriented contour detection: State of the art. *Image and Vision Computing* 29(2):79–103
- Parraga CA, Akbarinia A (2016) Colour constancy as a product of dynamic centre-surround adaptation. *Journal of Vision* 16(12):214–214
- Poirier F, Wilson HR (2006) A biologically plausible model of human radial frequency perception. *Vision research* 46(15):2443–2455
- Prewitt JM (1970) Object enhancement and extraction. *Picture processing and Psychopictorics* 10(1):15–19
- Shapley R, Hawken MJ (2011) Color in the cortex: single-and double-opponent cells. *Vision research* 51(7):701–717
- Shen W, Wang X, Wang Y, Bai X, Zhang Z (2015) Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 3982–3991
- Shushruth S, Ichida JM, Levitt JB, Angelucci A (2009) Comparison of spatial summation properties of neurons in macaque v1 and v2. *Journal of Neurophysiology* 102(4):2069–2083
- Shushruth S, Nurminen L, Bijanzadeh M, Ichida JM, Vanni S, Angelucci A (2013) Different orientation tuning of near-and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *The Journal of Neuroscience* 33(1):106–119
- Spillmann L, Dresch-Langley B, Tseng Ch (2015) Beyond the classical receptive field: the effect of contextual stimuli. *Journal of vision* 15(9):1–23
- Spratling MW (2013) Image segmentation using a sparse coding model of cortical area v1. *Image Processing, IEEE Transactions on* 22(4):1631–1643
- Thériault C, Thome N, Cord M (2015) Cortical networks of visual recognition. *Biologically Inspired Computer Vision: Fundamentals and Applications*
- Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: *Computer Vision, 1998. Sixth International Conference on*, IEEE, pp 839–846
- Tzvetanov T, Dresch B (2002) Short-and long-range effects in line contrast integration. *Vision research* 42(22):2493–2498
- Van De Weijer J, Gevers T, Gijzenij A (2007) Edge-based color constancy. *IEEE Transactions on image*

- processing 16(9):2207–2214
- Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L (2011) Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences* 108(23):9661–9666
- Watt R, Morgan M (1985) A theory of the primitive spatial code in human vision. *Vision research* 25(11):1661–1674
- Wei H, Lang B, Zuo Q (2013) Contour detection model with multi-scale integration based on non-classical receptive field. *Neurocomputing* 103:247–262
- Wilson H, Wilkinson F (2014) Configural pooling in the ventral pathway. Werner JS Chalupa L(Eds) *The new visual neurosciences* pp 617–626
- Xie S, Tu Z (2015) Holistically-nested edge detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 1395–1403
- Yang K, Gao S, Li C, Li Y (2013) Efficient color boundary detection with color-opponent mechanisms. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2810–2817
- Yang KF, Li CY, Li YJ (2014) Multifeature-based surround inhibition improves contour detection in natural images. *Image Processing, IEEE Transactions on* 23(12):5020–5032
- Yang KF, Gao SB, Guo CF, Li CY, Li YJ (2015) Boundary detection using double-opponency and spatial sparseness constraint. *Image Processing, IEEE Transactions on* 24(8):2565–2578
- Zeki S (1993) *A Vision of the Brain*. Oxford Univ Press