

Understanding and Simulating Border Ownership Centered Segmentation

Tianlong Chen

tc@tichen.net

OpticArray Technologies, Inc.

Xuemei Cheng

OpticArray Technologies, Inc.

Thomas Tsao

OpticArray Technologies, Inc.

Article

Keywords: border ownership, category selectivity, global context awareness, disparity, disparity representation, channel representation, figure-ground organization, segmentation, contrast-defined objects, disparity-defined objects, illusory objects, active neurons, owner-side direction, Gestalt principles, distributedness, object border, surface filling-in

Posted Date: July 14th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-5793080/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on December 11th, 2025. See the published version at <https://doi.org/10.1038/s41598-025-18843-9>.

Understanding and Simulating Border Ownership Centered Segmentation

Tianlong Chen*, Xuemei Cheng and Thomas Tsao

Corresponding author: Tianlong Chen, tc@tichen.net, ORCID: 0009-0009-1397-9266

Affiliation: OpticArray Technologies, Inc. 1917 Dundee Road, Rockville, MD 20850 USA

Abstract

Border ownership is a critical component of biological vision systems, enabling object segmentation and figure-ground organization. Understanding border ownership-centered segmentation is essential for advancing vision neuroscience and computer vision. In our prior work, “Border Ownership, Category Selectivity, and Beyond” (Chen et al., 2022)[6], we introduced a channel-based representation for coding border ownership and category selectivity. Building on this foundation, this paper addresses the remaining challenges in border ownership-centered segmentation.

We explore binocular disparity representation, the generation of binocular border ownership for contrast- and disparity-defined objects, and border ownership generation for illusory and contour-defined objects, including the Rubin Face-Vase illusion. We also propose the creation of ‘object pointers’ through hypothetical active neurons to address the ‘surface filling-in’ process in neuroscience and generate ‘instance masks’ in computer vision.

Finally, synthesizing these findings, we present a comprehensive model for border ownership-centered figure-ground organization. This model integrates global context awareness, distributed processing, and multi-cue representations, bridging gaps between biological vision and computational applications.

Keywords: border ownership, category selectivity, global context awareness, disparity, disparity representation, channel representation, figure-ground organization, segmentation, contrast-defined objects, disparity-defined objects, illusory objects, active neurons, owner-side direction, Gestalt principles, distributedness, object border, surface filling-in.

Declarations

The authors declare no competing interests or funding.

1. Introduction

Object segmentation is a core task in computer vision, which has advanced significantly in recent years due to deep-learning techniques. However, current state-of-the-art methods in computer vision do not fully align with the concept of ‘figure-ground organization’—a neuroscientific term for object segmentation in biological vision—where border ownership plays a crucial role.

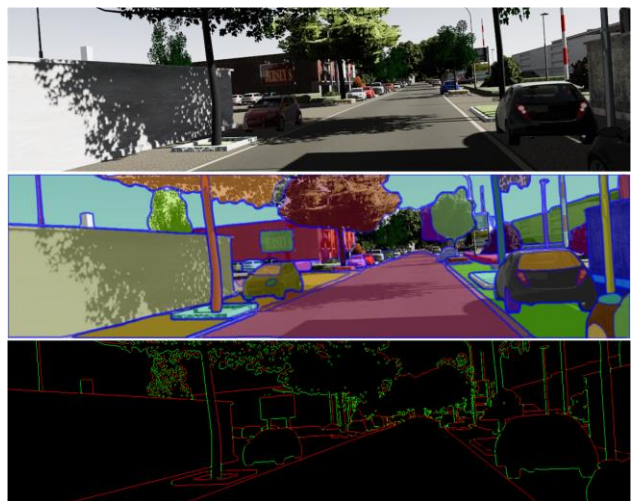


Figure 1 Object Segmentation with or without Border Ownership. **Top:** An RGB image from the Virtual Kitti 2 dataset (Cabon et al., 2020)[5] to be segmented. **Middle:** An instance segmentation map was generated by a state-of-the-art Segment-Anything (Kirillov et al., 2023)[24] with no border ownership. **Bottom:** A 2-channel border ownership map generated using the coding method from our prior work (Chen et al., 2022)[6] where **red** represents borders (occluding contours) with owners on the ‘below’ or ‘left’ sides, and **green** represents borders with owners on the ‘above’ or ‘right’ sides. A shared between two objects indicates depth ordering, where the border-owner object is positioned in front.

Border ownership refers to a vision phenomenon where a visual border (also known as ‘occluding contours’) between two regions in a scene or image is perceived as belonging to one of the regions, known as the ‘border

owner’. A region is then perceived as either a ‘figure’ or ‘background’, with the border carrying an orthogonal owner-side direction towards its ‘border owner’ region. At the border, the owner region is positioned in front of the other. Therefore, the ‘figure-ground organization’ in neuroscience is a border ownership-centered segmentation, where the ‘figure’ corresponds to what is referred to as an ‘object’ in computer vision.

1.1. What’s the Difference?

Figure 1 illustrates the difference between conventional segmentation methods that do not explicitly encode border ownership and border-ownership-centered segmentation. While both segment an image into ‘object’ or ‘background’ regions, the latter takes it further by using border ownership to identify which object at a shared border is positioned in front. Essentially, border ownership enables 3D segmentation (bottom image in Figure 1), whereas the current state-of-the-art remains in 2D (middle image in Figure 1). In biological vision, border ownership itself is integral to depth perception.

From a segmentation processing perspective, border ownership-centered segmentation begins with *contour-based* border ownership generation (identifying surrounding borders of objects with border ownership; bottom image in Figure 1) followed by ‘surface filling-in’ (determining the location and spatial coverage of object interiors enclosed by these borders). In this context, ‘surface’ is largely equivalent to ‘object’ in computer vision. In contrast, most state-of-the-art methods, which segment objects by direct pixel-wise labeling for their belonging in class and identity, are *region-based* (middle image in Figure 1), bypassing the use of explicit borders. Essentially, region-based methods perform ‘surface filling-in’ without borders.

Most modern segmentation methods (see survey papers by Hafiz et al., 2020[18]; Kirillov et al., 2018[25]; Elharrouss et al., 2021[14]; Sharma et al., 2022[39]; Gu et al., 2022[17]) integrate semantic classification and object detection to achieve segmentation, using class or object information as guidance. In contrast, border-ownership-centered segmentation typically involves limited, targeted classification via category selectivity (Chen et al., 2022[6]; see also Section 7.4). In biological vision, border ownership-centered segmentation precedes large-scale classification and recognition (Qiu et al., 2005[35]; Von der Heydt, 2023[43]); where category selectivity is rather a byproduct of the segmentation at the early vision stage.

The full border ownership segmentation output

(bottom of Figure 1) resembles panoptic segmentation (Kirillov et al., 2018[25]; Elharrouss et al., 2021[14]) in computer vision, which aims to segment everything visible in a scene. Meanwhile, category-selective segmentation (Figure 13) is more akin to instance segmentation (Hafiz et al., 2020[18]; Sharma et al., 2022[39]; Gu et al., 2022[17]) for selected categories.

1.2. Current Status in Brief

1.2.1 Border Ownership Coding

Contour-based, border ownership-centered segmentation in computer vision remains underdeveloped compared to *region-based* approaches. This disparity stems from the limited understanding of border ownership, a concept identified only recently by Zhou et al. (2000)[53], and the inherent complexity of replicating biological vision mechanisms. In contrast, region-based segmentation—both classical and deep learning-based—has benefited from over half a century of research, making it more mature and widely adopted.

Three major challenges hinder the development of border ownership-centered segmentation in computer vision. Two are related to the difficulties in coding or representing border ownership, while the third involves the challenge of performing ‘surface filling-in’ based on border ownership.

The greatest challenge lies in *representing border ownership*, where borders are treated as directed contours with orthogonal owner-side directions. A comparable analogy is the right-hand rule in physics, which links the directions of a magnetic field and moving charge: knowing one direction allows you to determine the other. In contrast to physics, object borders lack inherent directionality, and the orthogonal owner-side direction may lie on either side of the border.

The second major challenge involves the *generation of border ownership*. Williford et al., (2013)[50] provide a comprehensive review of border ownership and related topics from a neuroscience perspective, discussing three distinct models related to border ownership coding. All the models, explicitly or implicitly, assume that occluding contour generation and border ownership assignment are separate processes. Consequently, border ownership is assigned to occluding contours based on cues like T-junctions or through top-down guidance from higher-level vision hierarchies. These approaches face significant challenges in neuroscience (see discussions in Section 7), particularly in explaining the ‘short latency’ of border ownership generation and the presence of ‘global context awareness’, which appear contradictory. Moreover, the

complexity and unreliability of T-junctions coupled with the ambiguous logic behind top-down guidance for grouping contour segments, make these processes extremely challenging to simulate in computer vision.

The final challenge involves ‘*surface filling-in*’, which is the task of generating object regions from ‘contour-based’ border ownership within a ‘short latency’. This aspect has received less attention in computer vision due to the limited progress in addressing the first two challenges.

Our prior work (Chen et al., 2022)[6] proposed a practical and straightforward solution for border ownership coding in computer vision, effectively addressing the first two major challenges. This approach enables a plausible explanation for vision phenomena related to border ownership observed in neuroscience (see Section 7), resolving a key component of border ownership-centered segmentation. It also demonstrates that occluding contour generation and border ownership assignment can occur simultaneously without the need for T-junctions or top-down guidance.

1.2.2 Surface Filling-in

As Breitmeyer (2013)[3] suggested, a visual object can be defined as any “bounded segment of the visual field” that is distinguishable “by virtue of its contour and surface properties”. Breitmeyer further emphasized that “without surface qualia, there is no conscious object vision”.

Thus, contour-based border ownership generation represents only half the process required for complete segmentation. While it establishes a foundational structure by identifying object boundaries, it must be complemented by ‘surface filling-in’ to determine the interiors of objects—their location and spatial coverage—for complete segmentation. In neuroscience, this process is also referred to as ‘binding’ (Roelfsema, 2006)[36], similarly, Zhu et al. (2020)[54] describe the identity or control of ‘surface’ as the ‘object pointer’.

In computer vision, the outcomes of ‘surface filling-in’ are typically represented by ‘instance masks’, ‘semantic masks’, or ‘panoptic masks’ (Hafiz et al. 2020[18]; Kirillov et al., 2018[25]; Elharrouss et al., 2021[14]; Sharma et al., 2022[39]; Gu et al., 2022[17]), where each pixel is labeled according to instance or class identity. Meanwhile, most state-of-the-art segmentation methods perform ‘surface filling-in’ without explicitly relying on borders, using class and object information as guidance. In contrast, biological vision is thought to achieve classification and object recognition after segmentation, with contour-based border ownership likely serving as a

critical precursor to these subsequent processes.

In this paper, we propose a framework for ‘surface filling-in’ that is explicitly based on the generated border ownership (see Section 8).

1.2.3 Prior Works Worth Mentioning

Williford and von der Heydt (2013)[50] provide a comprehensive review of border ownership and related topics from a neuroscience perspective, discussing three distinct models related to border ownership coding. Von der Heydt (2023)[43] then offers an updated and extensive review of border ownership, which also explores the concept of ‘object pointer’. Both papers are foundational resources for understanding border ownership and offer essential context for this study.

Craft et al. (2007)[8] introduced a border ownership representation model from a neuroscience perspective, proposing a top-down guided approach driven by hypothetical high-level ‘Grouping cells’ (‘G cells’) that utilize local cues such as T-junction. In contrast, our prior work (Chen et al., 2022)[6] proposed a border ownership representation rule from a computer vision perspective, where border ownership is generated automatically by a trained deep neural network without relying on top-down guidance or T-junctions. While the mechanisms for generating border ownership differ, both approaches are nearly identical from a pure representation standpoint (excluding category selectivity). These distinct methodologies offer complementary insights into border ownership; see Section 7 for further discussion, and Section 2 for a recap of our prior work.

In computer vision, the Deep Occlusion Estimation (DOC) model by Wang et al. (2016)[46] is one of the most notable deep-learning approaches in this area. DOC employs a left-hand-rule coding method that combines an edge map with an edge-orientation map to represent border ownership. However, this approach does not closely align with neuroscience findings on border ownership properties (see Section 2 in this paper or Section 2.1 in our prior work, Chen et al., 2022[6]). As discussed in Section 1.2.1, the DOC method relies on directed contours to apply the left-hand rule, which is a limitation, making it more akin to category-selective segmentation than a comprehensive border ownership representation. In contrast, our coding method applies to any complex scenarios without such limitations, naturally integrating category selectivity along with border ownership in a manner consistent with neuroscience findings (see Section 2 for a ‘recap’ of our prior work and Section 7.4 on “Category Selectivity”).

The work by Arbeláez et al. (2011)[1] represented a state-of-the-art (non-deep learning) approach to contour detection and image segmentation before the advent of deep learning techniques. Their code and the Berkeley Segmentation Dataset (BSDS) have been integral to our research on border ownership. This work revealed unreliability and complexity of T-junction as cues for determining border ownership. Nonetheless, as a byproduct of this work, a set of sophisticated tools was developed for segmenting contours, which were later used to generate border ownership ground truth for our studies—a surprising and valuable outcome worth mentioning.

1.3. Organization of This Paper

In our prior work (Chen et al., 2022)[6], we addressed a key aspect of border ownership: its representation. For completeness, a recap of that work is provided in Section 2. This paper builds upon that foundation by investigating the remaining pieces of the puzzle surrounding border ownership-centered segmentation and exploring their potential implications for neuroscience and computer vision.

This exploration begins with binocular experiments (Section 3), covering binocular disparity representation (Section 3.1) and the generation of binocular border ownership with disparity on contrast- and disparity-defined objects (Section 3.2). We then extend this investigation to illusory objects (Section 4), analyze the Rubin Face-Vase illusion and its implications for border ownership assignment (Section 5), and examine contour-defined objects (Section 6).

Section 7 consolidates findings from Sections 3 through 6, providing a broader discussion on border ownership and its related topics. Section 8 introduces the process of ‘surface-filling-in’ based on border ownership, proposing a framework involving hypothetical ‘active neurons (pixels)’. In Section 9, these insights culminate in a comprehensive model for border ownership-centered ‘figure-ground organization’, integrating neuroscience-inspired concepts with computational efficiency. Section 10 acknowledges the current limitations of the work, and outlines directions for future research. Finally, Section 11 provides a summary, emphasizing the key contributions and broader implications of this study.

The primary objective of this paper is to establish a foundational understanding of border ownership-centered segmentation. While optimizing the border ownership coding method or enhancing the TcNet deep learning network remains valuable, they are not the central focus of this work. Instead, this paper highlights

how the current coding method aligns closely with neuroscientific findings on border ownership properties.

Throughout this paper, the 2-channel border ownership coding method is used consistently. Border ownership outputs are displayed using the same red-green-color coding scheme as illustrated in Figure 3 (c) of this paper or Fig. 1(c) of Chen et al. (2022)[6]. In this scheme, the ‘left’ or ‘bottom’ sides of a red-colored border segment represent the border owners, while the ‘right’ or ‘above’ sides of a green-colored border segment represent the border owners. These red and green segments are represented in separate channels.

2. Monocular Border Ownership

This section recaps our prior work (Chen et al., 2022)[6], providing an overview of border ownership and category-selectivity coding, as well as the encoder-decoder neural network, TcNet, which can be trained to generate border ownership maps with (optional) category-selective maps.

In Section 2.1 of our prior paper, we introduced a channel-based representation for coding border ownership that can be automatically generated from a trained encoder-decoder neural network, TcNet. The representation is grounded on two key neuroscience findings on border ownership selective neurons (Von der Heydt, 2015)[42]: “Border ownership preference is a *fixed property* of neurons” and “each neuron has a *fixed preference for location of one side or the other*”.

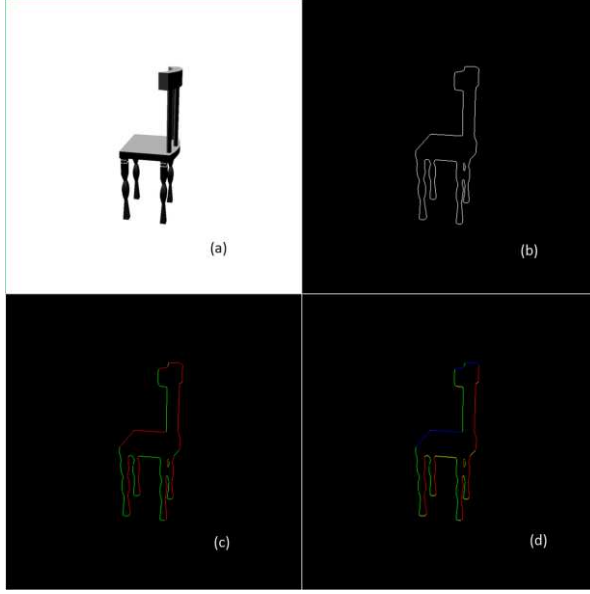


Figure 3 Border Ownership Coding (adapted from Fig. 1 in Chen et al., 2022[6]). (a) an example chair image from the FlyingChairs dataset (Aubry et al., 2014)[2]; (b) occluding contour map of the chair; (c) 2-channel border ownership coding: in the 1st channel, red-colored contours indicate border owners on the ‘left’ or ‘below’ sides; while in 2nd channel, green-colored contours indicate border owners on the ‘right’ or ‘above’ sides. (d) 4-channel border ownership coding: red, green, blue, and yellow-colored contours are placed in 1st, 2nd, 3rd, and 4th channels respectively, representing border owners on the ‘left’, ‘right’, ‘below’, and ‘above’ sides.

To represent border ownership of occluding contours, object contours are divided into relatively straight-line segments (occluding segments), these segments are then placed into different channels based on their border owner sides, following the ‘*opposite channel rule*’: *occluding segments with opposite border owner sides are put into different channels*. The rule aligns with the neuroscience findings (Von der Heydt, 2015)[42] that “each neuron has a fixed preference for location of one side or the other”. Figure 3 illustrates this concept: Figure 3(a) shows an example image; Figure 3(b) displays its occluding contour map; Figure 3 (c) and (d) demonstrate 2-channel and 4-channel border ownership coding, respectively. Theoretically, this border-ownership coding representation can be extended to 8, 16, or more channels using similar approaches.

In addition to border ownership, we (Chen et al., 2022)[6] demonstrated a channel-based representation for object category-selectivity coding: *all occluding contours for each category are put into one separate channel*, referred to as *category-selective channels* or *category channels*. Border ownership specific to individual object categories can then be derived by taking the inner product of border ownership channels with the object category channels (Figure 3).

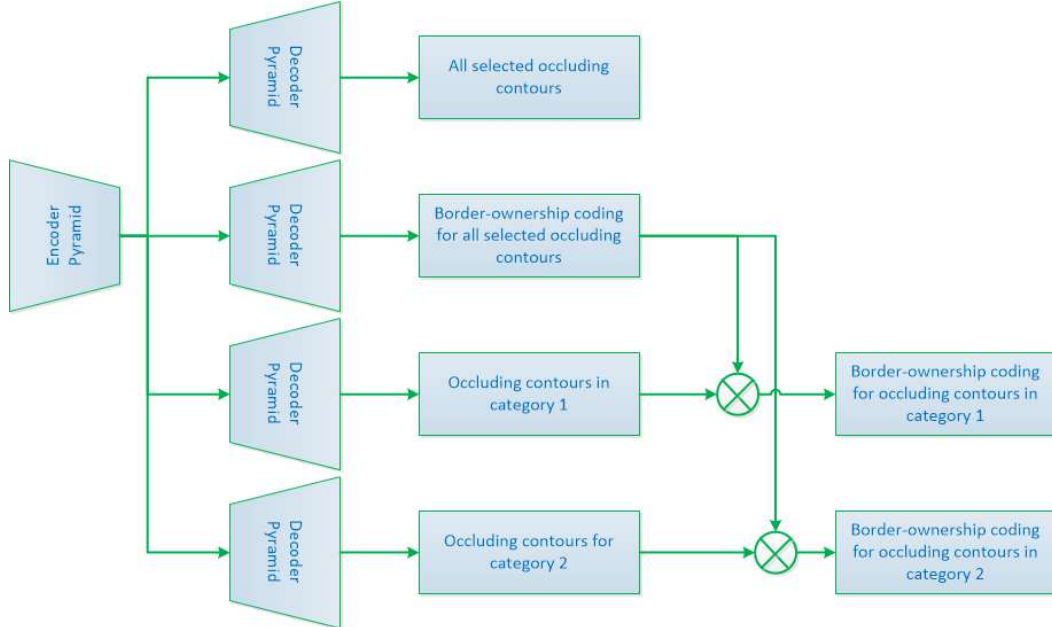


Figure 2 TcNet Multi-Branch Architecture for Multi-Category-Selective Border Ownership (adapted from Fig. 3 in Chen et al., 2022[6]). In this example, a single Encoder Pyramid is shared across multiple Decoder Pyramid branches for feature extraction. Each of the four Decoder Pyramid branches, from top to bottom, is used to process all contours, border ownership, and two selected categories, respectively. The border ownership for the two selected categories is computed by taking the inner product (denoted by the circled ‘X’ in the diagram) of the associated category channels and border ownership channels.

We also showcased an example encoder-decoder neural network, TcNet, in Figure 3 (identical to Fig. 3 in Chen et al., 2022[6]). This network can be trained to generate border ownership maps along with (optional) category-selective maps.

All the training and evaluation examples in Chen et al. (2022)[6] used monocular input. We will rename the TcNet in Chen et al. (2022)[6] as TcNet01 to distinguish it from the TcNet02 discussed in this paper, which processes binocular input. However, *the methods for border ownership and category-selectivity coding remain consistent across both versions.*

3. Binocular Vision and Border Ownership

By default, a healthy human vision system uses two eyes, making binocular vision more significant than monocular vision. The slightly different views from each eye, known as binocular disparities, provide one of the primary cues for depth perception in the biological vision system. Neuroscience studies by Qiu et al. (2005)[35] suggest that the binocular disparity-selective neurons may play a role in border ownership generation and contribute to the formation of a ‘figure-ground organization’, though the exact role of binocular disparity in border ownership generation remains unclear.

Before delving into binocular border ownership generation, for which we have a candidate border ownership representation as demonstrated in our prior work (Chen et al., 2022)[6], we will first examine how binocular disparity could be represented or encoded within this process.

3.1. Binocular Disparity

In computer vision, binocular disparity is defined as the difference in the image location of an object point in the left and right eyes (or cameras). According to the Epipolar geometry of binocular vision (Figure 4), the (horizontal) binocular disparity of P is inversely proportional to its depth (Eq. 3.1) when the image planes of both eyes are parallel. If the images are not aligned in this configuration, a process called ‘rectification’ is applied to make them parallel, ensuring accurate disparity estimation. Configurations beyond this setup is beyond the scope of this paper. Furthermore, the binocular configuration in Figure 4 can be interpreted as two observations of P by a monocular eye under horizontal motion parallax.

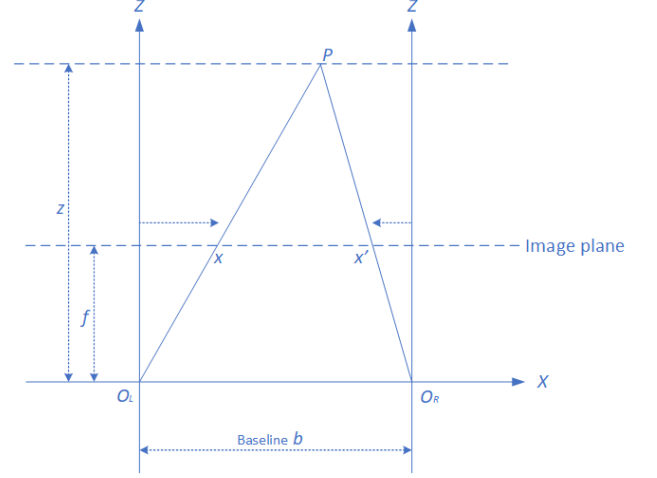


Figure 4 Epipolar Geometry on Binocular Vision. An object point P is observed by the left and right eyes (or cameras) O_L and O_R , separated by a distance (called ‘baseline’) b . Together, P , O_L , and O_R define the Epipolar plane. The image plane is positioned at a distance f away from the baseline and perpendicular to the paper. A 3-dimensional reference system is used with the x -axis pointing to the right, the y -axis pointing to the paper and pointing to the paper, and the z -axis pointing up, where z (the distance between P and the baseline) is referred to as the depth of point P , i.e. $z = \text{Depth}$. Assuming both eyes share the same focal length f , the x , and x' coordinates represent the locations of P appearing in the left and right reference system, respectively. Then the (horizontal) disparity is inversely proportional to the depth, as described by (Eq. 3.1), where $k = f*b$ is a known constant for fixed cameras (eyes) and their location.

$$\text{Disparity} = x - x' = \frac{f*b}{z} = \frac{k}{\text{Depth}} \quad (\text{Eq. 3.1})$$

Theoretically, (Eq. 3.1) represents the relationship between depth and disparity, it assumes that each pixel in one binocular image can find its corresponding pixel in the other. In practice, this is not always the case. As a result, the disparity map derived from (Eq. 3.1) is specific to each eye (or camera), meaning it is eye-oriented (or camera).

Moreover, (Eq. 3.1) becomes unreliable at large depths where the disparity measurement errors equal or exceed the disparity itself. As illustrated in the 5th image in Figure 5, the disparity ‘fades’ into white noise as depth increases. This phenomenon has been observed in other computer-processed disparity studies, such as the ‘Depth Anything’ by Yang et al. (2024)[52], where objects at large distances become noisy in disparity maps. Such effects are especially pronounced when nearby and distant objects have significant depth differences, with large disparities overshadowing small disparities.

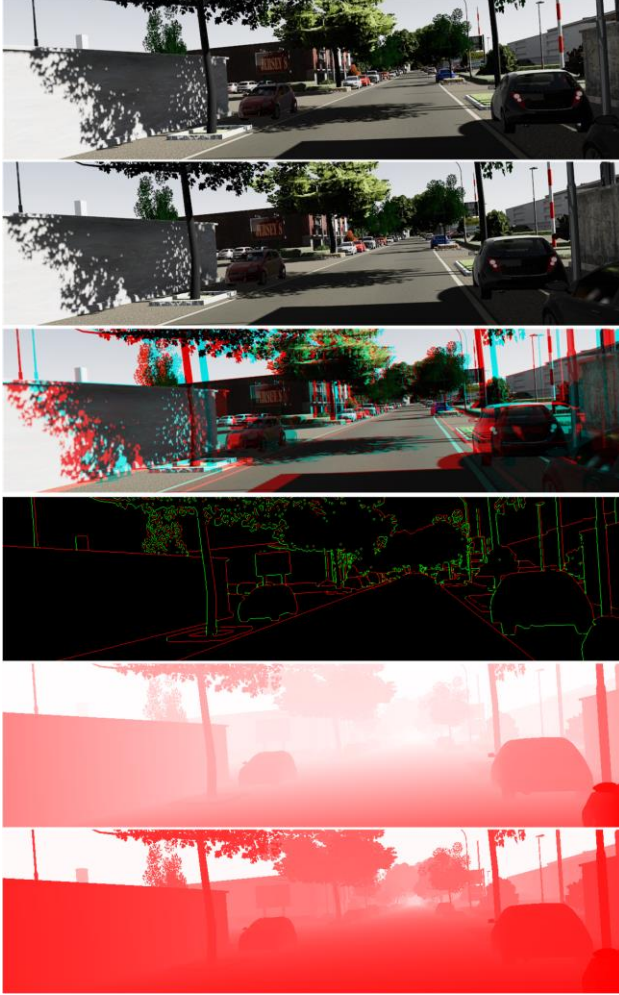


Figure 5 Binocular Vision Examples (image from Virtual Kitti 2 dataset, Cabon et al., 2020[5]). **The first two** images are the stereo left and right images, respectively; **the 3rd** image is a stereo anaglyph combining the two; **the 4th** image is a 2-channel border ownership coding map when viewed from the left camera; **the 5th** image is the ‘Near’ (Epipolar) disparity map where the greater depth appear lighter in red; **the 6th** image is the ‘Far’ (logarithmic) disparity map, also using red for depth coding, preserving depth order across distances.

In our daily experience, our eyes can seamlessly perceive depth across short and long distances. This suggests that the disparity-depth relationship in (Eq. 3.1) is insufficient. Vision neuroscience observations may offer additional insights into representing this relationship.

In neuroscience, binocular disparity is classified into ‘absolute disparity’ and ‘relative disparity’. According to Cumming et al. (1999)[9], ‘absolute disparity’ measures the angular difference between the projections of a single point on the two retinas, as illustrated in Figure 4. In contrast, the ‘relative disparity’ measures the angular

difference between their respective absolute disparities. The disparity in (Eq. 3.1) corresponds to an absolute disparity with fixation at infinity. In addition, Poggio (1995)[32] suggested the presence of ‘Far’ and ‘Near’ neurons in the V1 visual cortex area, selectively tuned to these disparities, though Prince et al. (2002)[33] found no distinct neuron classes for ‘Far’ and ‘Near’ disparities. Nevertheless, ‘Far’ and ‘Near’ disparities exist, regardless of whether they are generated by different neurons or not.

It appears that limited research has been conducted on disparity at large distances. For example, Palmisano et al. (2010)[30] explored disparities up to a depth of 248 meters, which barely covers the range depicted in Figure 5. This gap in understanding motivates alternative representations.

Inspired by logarithmic depth representations (Elgen et al., 2014)[13], we propose a logarithmic form for (Eq. 3.1), shown in (Eq. 3.2), where $C=\log(k)$ is a constant. The 6th image in Figure 5 is an example of (Eq. 3.2) converted from the depth map of the 5th image.

$$\log(\text{Disparity}) = C - \log(\text{Depth}) \quad (\text{Eq. 3.2})$$

Visual inspection and evaluation of disparity maps in Figure 5 suggest that the 5th image excels at presenting short-distance disparity, while the 6th image excels at long-distance disparity. The logarithmic form preserves the depth order of all objects and backgrounds at all distances and *flattens the magnitude difference between small and large disparities (i.e., long and short depth)*, maintaining good disparity representation even at large depths.

Combining (Eq. 3.1) and (Eq. 3.2), a candidate representation for a disparity-depth relationship could be expressed as the following function:

$$G(\text{Disparity}) = G\{F_n(\text{Depth}), F_f(\text{Depth})\} \quad (\text{Eq. 3.3})$$

where

$$F_n(\text{Depth}) = k/\text{Depth} \quad (\text{Eq. 3.4})$$

$$F_f(\text{Depth}) = C - \log(\text{Depth}) \quad (\text{Eq. 3.5})$$

and G denotes a function combining the ‘Near’ and ‘Far’ disparities, $F_n(\text{Depth})$ represents ‘Near’ (Epipolar) disparity while $F_f(\text{Depth})$ represents ‘Far’ (logarithmic) disparity. Neuroscience studies suggest that these disparities F_n and F_f , classified as ‘absolute disparities’, are likely generated in the V1 visual cortex (Cumming et al., 1999)[9].

Building on the channel representation of border ownership from our prior paper (Chen et al., 2022)[6], we

propose a ‘*layered disparity representation*’, where ‘*Far*’, ‘*Near*’, and other disparities are organized into separate disparity channels. For example, the disparity maps in the 5th and 6th images in Figure 5 are organized into two separate channels. Neuroscience findings (Poggio, 1995[32]; Prince et al., 2002[33]) suggest the existence of additional types of disparities beyond ‘*Near*’ and ‘*Far*’, and this ‘*layered disparity representation*’ of different disparities offers greater flexibility for the combination function G in (Eq. 3.3), *potentially allowing the inclusion of other disparities or optical flow derived from motion parallax in future extensions.*

According to the neuroscience findings by Qiu et al. (2005)[35], binocular border ownership is generated in the V2 visual cortex area with the involvement of disparity-selective neurons. The next section will evaluate the usability of ‘*Far*’ and ‘*Near*’ disparities in the generation of binocular border ownership.

It is worth mentioning that if we rearrange (Eq. 3.1) and (Eq. 3.2) respectively as (Eq. 3.1’) and (Eq. 3.2’) below, we would note that the *Disparity* and *Depth* in the equations are equivalent in position. We will discuss this equivalency in Section 7.6.

$$Disparity * Depth = k \quad (\text{Eq. 3.1}')$$

$$\log(Disparity) + \log(Depth) = C \quad (\text{Eq. 3.2}')$$

3.2. Binocular Border Ownership

In our prior paper (Chen et al., 2022)[6], we introduced a channel-based representation for border ownership and category selectivity, along with an encoder-decoder neural network, TcNet, designed for monocular image input. In this study, we extend the same border ownership channel representation, with optional category-selectivity, to binocular images, using a similar TcNet architecture.

To ensure a fair comparison between the border ownership generation results from monocular and binocular inputs, we adopt a version of TcNet closely aligned with that used in Chen et al. (2022)[6], making only minimal modification to the original network. Specifically, TcNet01 processes monocular input, while TcNet02 handles binocular inputs, as described in Section 2.

It is important to note that improving the performance of TcNet is not the primary objective of this paper.

3.2.1 Neural Network Setup

According to the neuroscience study by Qiu et al. (2005)[35], disparity-selective neurons are involved in binocular border ownership generation. Building on the different disparities discussed in Section 3.1, we designed three neural network configurations (Figure 6) to test the varying levels of disparity involvement in binocular border ownership generation.

Binocular Configurations:

1. First configuration (Figure 6(a)): Binocular images are stacked as input into TcNet02, with the expectation that the TcNet02 will implicitly learn the knowledge of disparities and generate border ownership maps directly.
2. Second configuration (Figure 6(b)): In this two-step process, one or both ‘*Far/Near*’ disparities are explicitly generated as intermediate results by TcNet02. These intermediate disparities are then combined with binocular images and fed into TcNet02CM, which outputs the border ownership maps.
3. Third configuration (Figure 6(c)): Similar to the second configuration, but in this case, the ‘*Far*’ and ‘*Near*’ disparities are generated separately by two separately trained TcNet02 networks.

Monocular Configurations for Comparison:

Depth can also be perceived monocularly under the assumption that monocular depth can be represented by a disparity map. For comparison, we include results from monocular configurations (Figure 7), where Figure 7(a) mirrors Figure 6(a), and Figure 7(b) mirrors Figure 6(b).

Design Choices

The primary difference between the monocular networks (TcNet01/TcNet01CM) in Figure 7 and the binocular networks (TcNet02/TcNet02CM) in Figure 6 lies in the number of inputs. Minimal modifications were intentionally made to ensure comparable results. Additionally, this design reflects the perspective that the monocular process is a special case of the binocular processing, albeit with a few exceptions where monocular approaches may fail.

Notably, for simplicity, the same neural network (TcNet) was used for both disparity and border ownership generation, which is convenient but not strictly necessary.

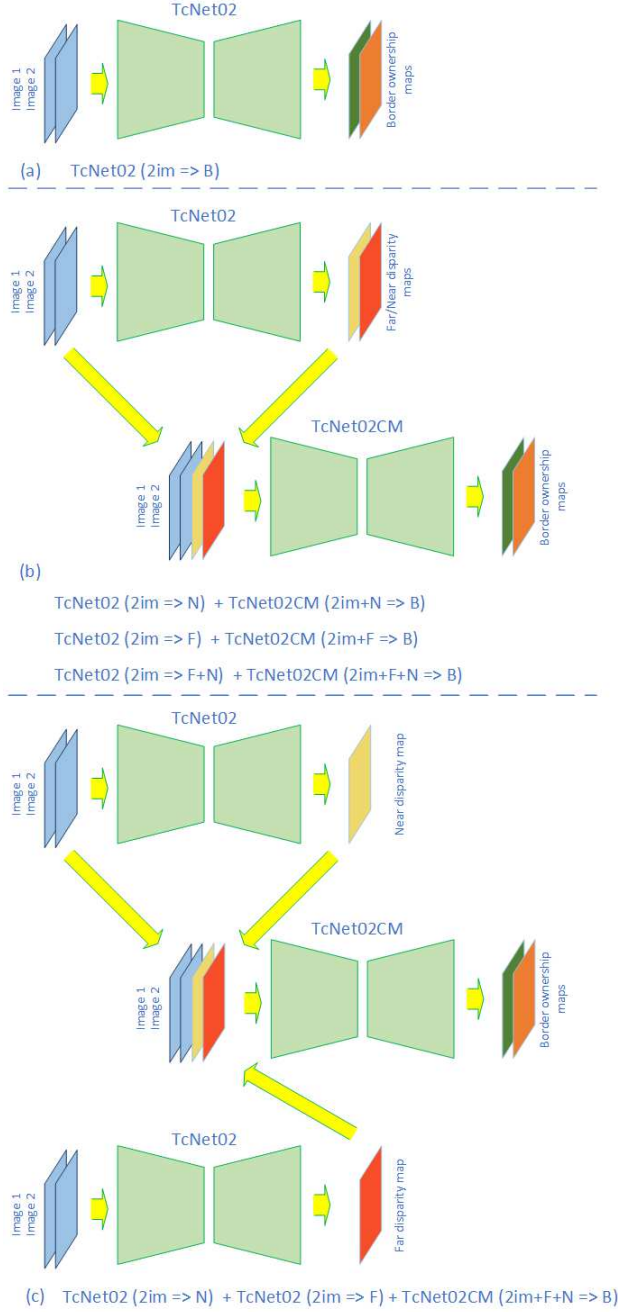


Figure 6 Three Neural Network Configurations for Binocular Border Ownership Generation. **(a)**: In this architecture, binocular images are stacked as input into TcNet02, expecting the network to implicitly learn the knowledge of disparities and generate border ownership maps. **(b)**: A two-step process in this architecture where one or both ‘Far/Near’ disparities are explicitly generated as intermediate results by TcNet02. These intermediate disparities are then combined with binocular images as input to TcNet02CM, which outputs border ownership maps. **(c)** is similar to **(b)** except the ‘Far’ and ‘Near’ disparities are generated separately by two independently trained TcNet02

network.

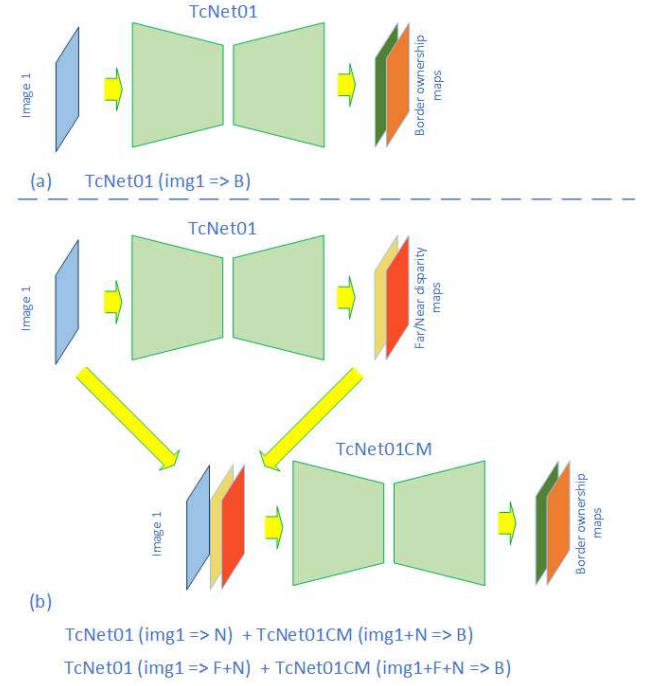


Figure 7 Neural Network Configuration for Monocular Border Ownership Generation, used for comparison with their binocular counterparts in Figure 6. **(a)**: TcNet01 with monocular input, producing border ownership maps as described in Chen et al. (2022)[6]. **(b)**: similar to case (b) in Figure 6 except that the intermediate disparities are generated from TcNet01 with monocular image input (left image, or image 1); and the stacked input to TcNet01CM includes monocular image and disparities.

Notation: In Figure 6 and Figure 7:

‘F’: 1-channel ‘Far’ disparity map.

‘N’: 1-channel ‘Near’ disparity map.

‘F+N’: 2-channel map of both ‘Far/Near’ disparities.

‘=>’: Output.

‘=>B’: Output includes 1-channel all-contour map (referred to as ‘enforcement channel’ in Chen et al., 2022[6]) and 2-channel border ownership maps. For the Virtual Kitti 2 dataset (discussed in Section 3.2.2), the output includes an extra 1-channel vehicle category contour map.

(2im=>B): Binocular image input directly generating border ownership maps.

(2im+N+F=>B): Binocular image input combined with ‘Near/Far’ disparities, producing border ownership maps.

(img1=>B): Monocular image input from image 1, producing border ownership maps.

Due to differences in input and output, TcNet02 and TcNet02CM, TcNet01 and TcNet01CM were separately trained for each configuration.

3.2.2 Datasets

For our experiments on binocular border ownership generation, we used two datasets: the Random Dot Stereograms (RDS) dataset, where objects are disparity-defined, and a modified version of the Virtual Kitti 2 dataset, where objects are contrast-defined.

Unless otherwise specified, all binocular disparity maps and binocular border ownership maps presented in this paper are assumed to be viewed by the left eye (camera) for both experimental datasets.

3.2.2.1 Random Dot Stereograms

Random Dot Stereograms (RDS) are considered the “gold standard” for border ownership assignment in neuroscience (Qiu et al., 2005)[35]. Since the pioneering work by Bela Julesz (1971)[22], various forms of RDS have been widely used in cognitive science and perception studies.

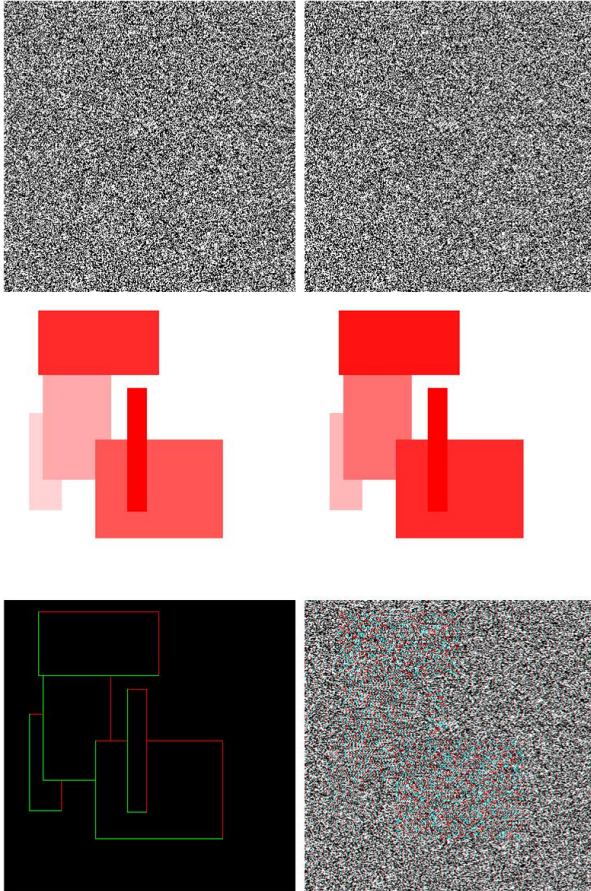


Figure 8 Random Dot Stereograms (RDS) Dataset, generated using the modified method from Marino (2007). **Top row:** Left and right RDS images, where the left image is a pure random dot image, while the right image is created from the left image

by shifting random numbers (1 to 6) of randomly located rectangle-shaped areas. The shifted amount of each area is randomly selected from values of 2, 4, 6, 8, 10 and 12 pixels; **Middle row:** The left image is a ‘Near’ disparity map, where darker red rectangles indicate closer objects (larger shifts), and lighter red rectangles indicate farther objects (smaller shifts). The right image represents a ‘Far’ disparity map, which is the logarithmic form of the ‘Near’ disparity map. **Bottom row:** The left image shows a 2-channel border ownership map for the ‘disparity-defined’ objects; the right is an anaglyph of the two RDS images.

Julesz’s RDS consists of two random dot images (see the top two example images in Figure 8). We adopted a modified RDS generation method based on Marino (2007)[28], where ‘image 1’ (left image) is a pure random dot image, while ‘image 2’ (right image) is created by shifting a random number (1 to 6) of randomly located rectangle areas in ‘image 1’. When viewed binocularly, these shifted areas appear as ‘disparity-defined’ objects with depth. However, when viewed monocularly, no objects are visible. The bottom-right image in Figure 8 shows an anaglyph representation of the two RDS images.

Our RDS dataset is more complex than those typically used in neuroscience studies, allowing for overlapping ‘disparity-defined’ objects. It is divided into a ‘train’ set and a ‘validation’ set for training, and a separate ‘test’ set for evaluation.

3.2.2.2 Modified Virtual Kitti 2

We selected a modified version of the Virtual Kitti 2 (VKitti) photorealistic computer-generated dataset (Cabon et al., 2020)[5] as the second experimental dataset for contrast-defined objects, reflecting real-life human scenarios (see Figure 5 for an example). The original VKitti dataset provides multiple sets of ground truth including binocular RGB images, depth maps, class and instance segmentation maps, and more.

We divided the dataset into three non-overlapping groups: ‘train’ and ‘validation’ sets for training, and a separate ‘test’ set for evaluation.

Each dataset group includes binocular RGB images, and their associated ‘Near’ disparity maps, generated by applying (Eq. 3.1) to the original depth maps, The ‘Far’ disparity maps were then derived from the ‘Near’ disparity maps using (Eq. 3.2). In addition, the ‘Near’ disparity maps were scaled down by a factor of 10 to roughly match the magnitude of the ‘Far’ disparity maps.

In addition, each dataset group includes: 1-channel all-contour maps (referred to as the ‘enforcement’ channel in

Chen et al., 2022[6]), 2-channel border ownership maps, and 1-channel vehicle category channel maps.

Contour maps were generated from the original class and instance mask maps. Border ownership was determined by comparing the depth difference between the two sides of the border segments, the side with the closer depth was identified as the border owner, and coded using the ‘*opposite channel rule*’ (described in Section 2 in this paper and Section 2.1 of Chen et al., 2022[6]). However, due to inaccuracies in estimating depth difference between the two sides of contour pixels, the ground truth border ownership coding may not always be precise (see the 4th image in Figure 5).

3.2.3 Statistical Evaluation Criteria

It is important to note that only the final outputs — contour maps, border ownership maps, and category-selective maps—are statistically evaluated, while intermediate (disparity) results are not included in the evaluation.

The statistic evaluation criteria used in this paper are nearly identical to those described in Section 3.2 in Chen et al., (2022)[6]:

(a) The mean average precision (mAP) was calculated by averaging the ratio of correctly predicted contour points over all predicted contour points across all test cases, specifically,

$$\text{mAP} = (G \cap P_c) / P_{\text{all}}$$

where G is ground truth, P_c and P_{all} represent, respectively, correct prediction and all prediction (both correct and incorrect);

(b) A threshold (>0.1) was applied to the results, and all contours were ‘thin’ processed before statistical calculations;

(c) A 1-neighbor ground truth (around contour lines) was used, where a prediction was counted correct only if the predicted contour point was within a 1-pixel distance from ground truth contour points.

3.2.4 Results

Table 1 Evaluation results for Random Dot Stereograms dataset

	Models	All-contours	Border ownership
1	TcNet02(2im=>B)	87.4%	85.9%
2	TcNet02(2im=>N) + TcNet02CM(2im+N=>B)	95.8%	91.7%
3	TcNet02(2im=>F) + TcNet02CM(2im+F=>B)	91.0%	87.2%
4	TcNet02(2im=>F+N) + TcNet02CM(2im+F+N=>B)	95.6%	92.5%
5	TcNet02(2im=>F) + TcNet02(2im=>N) + TcNet02CM(2im+F+N=>B)	96.3%	92.5%

6	TcNet01(image2 => B)	0%	0%
---	----------------------	----	----

Table 2 Evaluation results for the modified Virtual Kitti 2 (VKitti) dataset

	Models	All contours	Border ownership	Category (vehicle)
1	TcNet02(2im=>B)	81.3%	65.5%	69.8%
2	TcNet02(2im=>N) + TcNet02CM(2im+N=>B)	68.7%	60.7%	67.3%
3	TcNet02(2im=>F) + TcNet02CM(2im+F=>B)	72.1%	60.7%	63.9%
4	TcNet02(2im=>F+N) + TcNet02CM(2im+F+N=>B)	77.4%	65.9%	72.9%
5	TcNet02(2im=>F) + TcNet02(2im=>N) + TcNet02CM(2im+F+N=>B)	75.9%	64.8%	69.0%
6	TcNet01(img1=>B)	79.5%	63.8%	66.1%
7	TcNet01(img1=>N) + TcNet01CM(img1+N=>B)	68.9%	57.9%	66.1%
8	TcNet01(img1=>N+F) + TcNet01CM(img1+N+F=>B)	76.4%	63.0%	61.9%

We conclude with the following observations:

(1) The channel representation for border ownership and category selectivity, as developed in our previous work (Chen et al., 2022)[6] *applies to binocular cases*.

(2) On the RDS dataset (Table 1), the configurations with explicit intermediate disparity learning (rows 2 to 5) *significantly outperform* the direct border ownership generation configuration (row 1), which expects to learn disparity implicitly.

(3) On the VKitti dataset (Table 2), the direct border ownership configuration (row 1) performs comparably to the configuration incorporating both ‘Far’ and ‘Near’ disparities (rows 4 and 5). However, it outperforms the configurations that use one of ‘Far’ or ‘Near’ disparity (rows 2 or 3).

(4) Across both datasets, the configurations incorporating both ‘Far’ and ‘Near’ disparities (rows 4 and 5 in Table 1 and Table 2) *outperform* those utilizing only ‘Far’ or ‘Near’ disparity (rows 2 and 3 in Table 1 and Table 2).

(5) In the RDS dataset (Table 1), a comparison of rows 2, 3, and 4 indicates that the ‘Near’ disparity contributes significantly more than the ‘Far’ disparity. Conversely, the same comparison in the VKitti dataset (Table 2) suggests that the ‘Near’ and ‘Far’ disparities contribute similarly.

(6) For both datasets, the configurations incorporating both ‘Far’ and ‘Near’ disparities (rows 4 and 5 in Table 1 and Table 2) *show similar performance*, regardless of whether the intermediate ‘Far’ and ‘Near’ disparities were generated separately.

(7) Table 1 (row 6) demonstrates that the monocular TcNet01, when trained on the RDS dataset (image 2), fails to detect any objects with a monocular image input. This confirms that the RDS images (image 2) are invisible

monocularly, while image 1 consists purely of random dots.

(8) Table 2 shows a consistent trend where monocular TcNet01 configurations slightly underperform compared to binocular TcNet02 configuration on the VKitti dataset. Specifically, row 6 vs row 1 (without disparity), row 7 vs row 2 (with ‘Near’ disparity), and row 8 vs row 4 (with both ‘Near’ and ‘Far’ disparities).

For further discussion, refer to Section 7.7.

4. Illusory Objects and Illusory Contours

We demonstrated that our border ownership coding method and TcNet can effectively handle binocular disparity-defined objects where object borders are invisible to a single human eye. It is of great interest to explore another type of invisible border: the ‘illusory contours’ of the ‘illusory objects.’ According to Wikipedia[48], illusory (or subjective) contours are “visual illusions that evoke the perception of an edge without a luminance or color change across that edge”.

Neuroscience research by Von der Heydt et al. (1984) [44] indicates that illusory contours are extracted in the V2 region, where binocular border ownership is also generated (Qiu et al., 2005)[35], suggesting a potential connection between the two processes. Additionally, a study by De Haas et al. (2018)[10] also revealed “few systematic differences between retinotopic responses to illusory contours, occlusion, and subtle luminance stimuli in early visual cortex”, implying that border ownership generation for illusory contours occurs at an early vision stage, similar to other types of occluding contours.

In this Section, we demonstrate that the same border ownership coding method and monocular TcNet01 described in Chen et al. (2022)[6] can be applied to several types of illusory contours commonly studied in psychology and neuroscience, including Kanizsa figures, Ehrenstein illusion, Neon color spreading, and Illusory shading (Kanizsa, 1976[23]; Bressan et al., 1997[4]; Coren, 1972[7]; Von der Heydt et al., 1984[44]).

In addition, Coren (1972)[7] suggested that “both monocular and binocular subjective contours result from the presence of depth cues”. It was observed and can be easily verified, that the subjective object in illusory figures is perceived to float up in front of other elements in the image. Readers can verify this phenomenon by observing the illusory shapes in Figure 9, where the perceived object appears to be elevated relative to their surrounding elements.

To test Coren’s hypothesis, two experiments were designed using monocular and binocular inputs to verify the border ownership generation on Kanizsa figures with synthetic (Near) disparity. These experiments simulate the neurons’ perception of depth cues before generating border ownership for subjective objects.

4.1. Illusory Contour Datasets

Kanizsa Figures

Kanizsa figures are one of the most renowned examples of illusory contours (Kanizsa, 1976[23]; Coren, 1972[7]). The left image of Figure 9 (a) illustrates an illusory perception of a white quadrilateral Kanizsa figure, while the right image of Figure 9 (a) shows its corresponding 2-channel border ownership map.

A training dataset was created with mixed ‘true’ and ‘false’ cases, divided into ‘train’ and ‘validation’ sets for training, and separate ‘test’ sets with separate ‘true’ and ‘false’ cases for evaluation.

‘True’ cases: Examples like Figure 9 (a), feature a Kanizsa figure of a triangle or quadrilateral with randomly located endpoints.

‘False’ cases: Examples like Figure 9 (b) are generated by rotating the endpoints of a Kanizsa figure by 90°.

Evaluation: Additional pentagon and hexagon Kanizsa shapes, which were not in the training set (Figure 9 (c) and (d)), were also evaluated.

Ehrenstein Illusion

The Ehrenstein illusion is closely related to the Kanizsa figure, creating an illusory perception of a white-filled circle intersecting radial line segments (Dresp-Langley, 2009[11]). The left image in Figure 9 (e) illustrates this effect, while the right image of Figure 9 (e) shows its 2-channel border ownership map.

Comparing the images in Figure 9 (g) and (h) from Neon color spreading with Figure 9 (e) suggests that the Ehrenstein illusion is a special case of Neon color spreading with transparency set to 0. Therefore, the training dataset for Ehrenstein illusion was combined with Neon color spreading, generates different transparent circles, and uses Figure 9 (f) as ‘false’ cases.

‘True’ cases: Comprise examples of Ehrenstein illusions and Neon color spreading.

‘False’ cases: Include examples like Figure 9 (f), where the radial segments do not produce illusion.

Dataset composition: The ‘train’ and ‘validation’ sets contain a mix of ‘true’ cases and ‘false’ cases, while ‘test’ sets contain separate ‘true’ and ‘false’ cases.

Neon Color Spreading

Neon color spreading is described as “an optical illusion with transparency effects, characterized by fluid borders between the edges of a colored object and the background in the presence of black lines” (Wikipedia[49]). The left images in Figure 9 (g) and (h) illustrate this effect with red and light-white illusory objects, respectively, while the right images in Figure 9 (g) and (h) show their corresponding 2-channel border ownership maps.

The training dataset for Neon color spreading incorporates the Ehrenstein illusion (see described above).

Illusory Shading

Illusory shading is a phenomenon where shading cues create the perception of illusory contours (Coren, 1972[7]). This is illustrated in the left image in Figure 10, showing the shading letters (‘F D S T’); while the right image shows their corresponding 2-channel border ownership map.

To simplify dataset generation for the training and testing, the 1st channel of the 2-channel border ownership map was used as input into the monocular TcNet01. Each image includes a random number (1 to 4) of illusory shading letters.

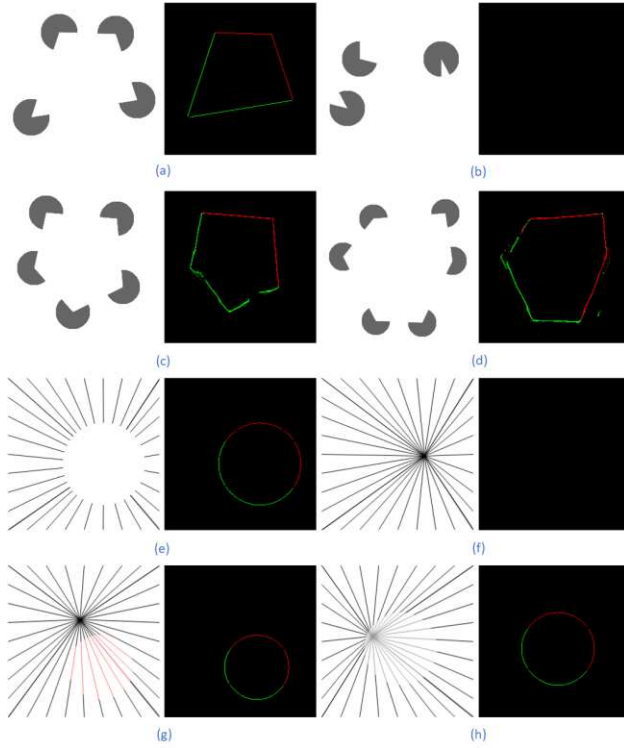


Figure 9 Illusory Contours of Illusory Objects. (a) Left: A Kanizsa illusory figure; Right: Its 2-channel border ownership coding; (b) Left: A false Kanizsa figure; Right: its border

ownership map (empty); (c) Left: A pentagon-shaped Kanizsa figure; Right: its evaluated border ownership map; (d) Left: A hexagon-shaped Kanizsa figure; Right: Its evaluated border ownership map; (e) Left: An Ehrenstein illusory figure; Right: Its 2-channel border ownership coding; (f) Left: A false Ehrenstein illusory or Neon color spreading figure; Right: Its border ownership map (empty); (g) Left: A Neon (red) color spreading illusion; Right: Its border ownership map; (h) Similar to (g) but with light-white Neon color.

Kanizsa Figures with Synthetic Disparity

To test border ownership generation with synthetic disparity (disparity=2), we used the following configurations:

Monocular testing configuration: Mirrors row 7 in Table 2 for TcNet01 on the VKitti dataset (see also Figure 7 (b)):

$TcNet01(img1 \Rightarrow N) + TcNet01CM(img1 + N \Rightarrow B)$

where the synthetic (Near) disparity is generated as an intermediate result before border ownership. This approach simulates depth cues influencing border ownership, as suggested by Coren (1972)[7]: “subject contours result from the presence of depth cues”.

Binocular testing configuration: Mirrors row 2 in Table 2 for TcNet02 on the VKitti dataset (see also Figure 6(b)):

$TcNet02(img1 + img1 \Rightarrow N) + TcNet02CM(img1 + img1 + N \Rightarrow B)$

where the binocular inputs are identical Kanizsa figure, simulating both eyes observing the same image.

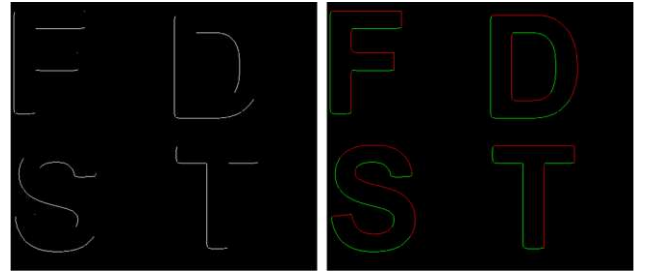


Figure 10 Illusory Shading. Left: shading letters (‘F D S T’), derived from the 1st channel of the 2-channel border ownership map; Right: their corresponding 2-channel border ownership map.

4.2. Results

The evaluation results from the trained TcNet01 include a 1-channel all-contour map and 2-channel border ownership maps (refer to Figure 3). The statistical evaluation for contours or border ownership maps follows the same criteria described in Section 3.2.3. For ‘false’ (empty) cases, the evaluation is based on the accuracy of predicting empty maps. Specifically, it is calculated as the ratio of correctly predicted empty cases to the total number of ground-truth empty cases.

Table 3 Evaluation results for Kanizsa figures (without synthetic disparity). The training sets contained only triangle and quadrilateral Kanizsa figures. Pentagon and hexagon-shaped Kanizsa figures were also evaluated without being included in training, shown in **Figure 9** (c) and (d).

	All-contours	Border ownership
Triangle/quadrilateral	97.7%	91.6%
‘false’ (empty)	96.4%	
Pentagon	65.5%	60.3%
Hexagon	75.9%	72.8%

Table 4 Evaluation results for Ehrenstein illusion and Neon color spreading.

	All-contours	Border ownership
Ehrenstein	99.5%	94.0%
Neon (red)	96.2%	92.0%
Neon (light-white)	91.8%	85.5%
‘false’ (empty)	99.8%	

Table 5 Evaluation results for Illusory Shading.

	All-contours	Border ownership
Illusory shading	99.5%	98.3%

Table 6 Monocular evaluation results for Kanizsa Figures with synthetic Near disparity (disparity=2): TcNet01(img1=>N) + TcNet01CM(img1+N=>B)

	All-contours	Border ownership
Triangle/quadrilateral	73.7%	50.0%
‘false’ (empty)	0%	
Pentagon	37.9%	22.7%
Hexagon	43.0%	26.5%

Table 7 Binocular evaluation results for Kanizsa Figures with synthetic Near disparity (disparity=2): TcNet02(img1+img1=>N) + TcNet02CM(img1+img1+N=>B)

	All-contours	Border ownership
Triangle/quadrilateral	90.0%	70.0%
‘false’ (empty)	2%	
Pentagon	48.1%	32.7%
Hexagon	43.3%	30.2%

We conclude with the following observations:

(1) The experiments (Table 3, Table 4 and Table 5) confirm our hypothesis that *the channel representation of border ownership and TcNet, as described in our prior paper* (Chen et al., 2022)[6], *is applicable to the illusory contour generation of illusory objects*, including illusory Kanizsa figure, Ehrenstein illusion, Neon color spreading, and illusory shading. Its *performance is comparable to that observed for disparity-defined objects and contrast-defined objects*.

(2) For the Kanizsa figure tests (Table 3), it is noteworthy that pentagon and hexagon-shaped Kanizsa figures, which were not included in the training, achieved surprisingly strong evaluation results (also see Figure 9 (c) and (d)).

(3) For Neon color spreading and Ehrenstein illusion tests (Table 4), the performance varies depending on the transparency levels. The opaque Ehrenstein illusion yields the best results among these tests.

(4) A comparison of Table 3 (without synthetic disparity), Table 6 (monocular with synthetic disparity) and Table 7 (binocular with synthetic disparity) reveals that border ownership generation on Kanizsa figures with synthetic disparity performs significantly worse than those without synthetic disparity. See Section 7.9 for further discussion.

5. Rubin Face-Vase Illusion

In our prior paper (Chen et al., 2022)[6], we argued that the ‘global context awareness’ is a ‘fixed property’ embedded within a group of border ownership-selective neurons, and potentially within category-selective neurons as well. If this hypothesis holds, visual perception is determined by the global context at the moment it occurs.

The Rubin Face-vase illusion (Rubin, 1915[38]; see Figure 11) is an image that can be perceived as either two faces or one vase, demonstrating that the same borders can have opposite border ownership assignments, depending on the viewer’s perception. The dynamic nature of this perception, driven by alternating border ownership assignments, makes it an ideal case for testing this hypothesis. If the neurons responsible for border ownership can handle such perceptual reversals without additional top-down input, it would further support the claim that ‘global context awareness’ is an intrinsic (‘fixed property’) feature of these neurons. However, the mechanism underlying dynamic perceptual shifting, which itself is not directly related to border ownership assignment, is outside the scope of this study.

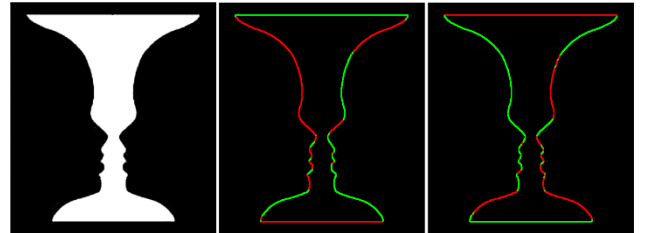


Figure 11 Rubin Face-Vase Illusion. **Left:** A Rubin Face Vase image (sourced from Google). **Middle:** The 2-channel border ownership assignment for the same image.

ownership map when the image is perceived as two faces; in this map, the ‘left’ or ‘bottom’ sides of red-colored occluding segments are border owners, the ‘right’ or ‘up’ sides of green-colored occluding segments are border owners. **Right:** The 2-channel border ownership map when the image is perceived as a vase; here, the border owners are reversed.

Neuroscientists have used the rectangle-window object (Qiu et al., 2005)[35] to simulate opposite border ownership assignments similar to those seen in the Rubin Face-Vase illusion. In a rectangle-window image, a target is perceived as either a rectangle object or a window object, with each interpretation producing opposite border ownership assignments (see Figure 12). We will use the rectangle-window object to simulate how ‘global context awareness’ determines border ownership assignments.

5.1. Experiment Setup

We designed an experiment with two opposite rectangle-window datasets simulating two contrasting ‘global contexts’ (see Figure 12):

(1) ‘Darker-region-as-object’: in this scenario, darker rectangles or windows, compared to their surrounding background, are treated as objects (See the left image in Figure 12).

(2) ‘Lighter-region-as-object’: in this scenario, lighter rectangles or windows, compared to their surrounding background, are treated as objects (See the right image in Figure 12).

5.2. Results

A monocular TcNet01 was trained separately on the two distinct rectangle-window datasets and then evaluated on both datasets.

Table 8 Statistics of monocular TcNet01 on two Rubin Face-vase-like rectangle-window datasets. The mean average precision (mAP) calculations follow the criteria outlined in Section 3.2.3. **(a)** Monocular TcNet01 trained on the ‘Darker-region-as-object’ dataset; **(b)** Monocular TcNet01 trained on the ‘Lighter-region-as-object’ dataset.

TcNet01	Darker-region-as-object (inference)		Lighter-region-as-object (inference)	
	All contour	Border ownership	All contour	Border ownership
(a)	97.8%	95.1%	94.0%	0%
(b)	88.4%	0%	99.4%	98.5%

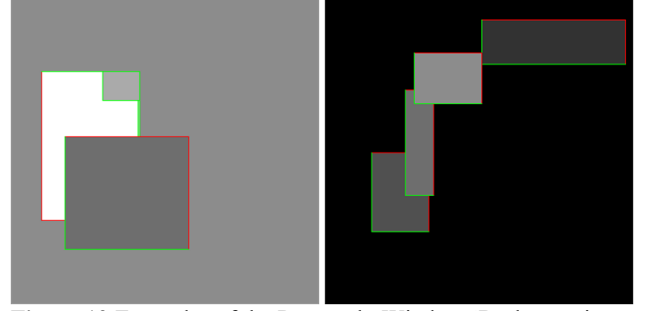


Figure 12 Examples of the Rectangle-Window ‘Darker-region-as-object’ and ‘Lighter-region-as-object’ Datasets, with border ownership maps overlaid on rectangle-window images. **Left:** An example of ‘Darker-region-as-object’ includes two rectangle objects and one window object, where each object (rectangle or window) is darker than its background; **Right:** An example of ‘Lighter-region-as-object’ includes four rectangle objects and no windows, where each object is lighter than its background. In both cases (**Left/Right**), rectangle objects have red-colored right and top borders, and green-colored left and bottom borders. Conversely, window objects have opposite border ownership, with green-colored right and top borders, and red-colored left and bottom borders.

We conclude with the following observations:

(1) Table 8(a) shows that a monocular TcNet01 trained on the ‘Darker-region-as-objects’ dataset, with the ‘global context’ of ‘darker-region-as-object’, consistently treats darker regions as objects when evaluated on *both* datasets. The 0% accuracy on the ‘Lighter-region-as-object’ dataset column in Table 8(a) indicates that the border ownership assignment is opposite to the ground truth of ‘Lighter-region-as-object’.

(2) Table 8(b) shows that a monocular TcNet01 trained on the ‘Lighter-region-as-objects’ dataset, with the ‘global context’ of ‘Lighter-region-as-object’, consistently treats lighter regions as objects when evaluated on *both* datasets. The 0% accuracy on the ‘Darker-region-as-object’ dataset column in Table 8(b) indicates the border ownership assignment is opposite to the ground truth of ‘Darker-region-as-object’.

(3) In both cases, the occluding contours (‘All contour’ columns in Table 8) were mostly generated correctly, regardless of the ‘global context’ being ‘Darker-region-as-object’ or ‘Lighter-region-as-object’.

The experiment provides strong evidence that *the ‘global context’, learned during training, determines perception*. Neurons likely encode both ‘global contexts’, and the dominant context at the moment of perception dictates whether the perceived object is ‘darker’ or ‘lighter’. This perception can dynamically switch between the two contexts, depending on which one prevails.

6. Contour Objects

It is of great interest to explore another extreme case of objects on border ownership and category selectivity generation, where the images consist solely of occluding contours, devoid of any color or texture. In such cases, objects are perceived purely from their occluding contours, demonstrating how borders alone convey object identity and structure.

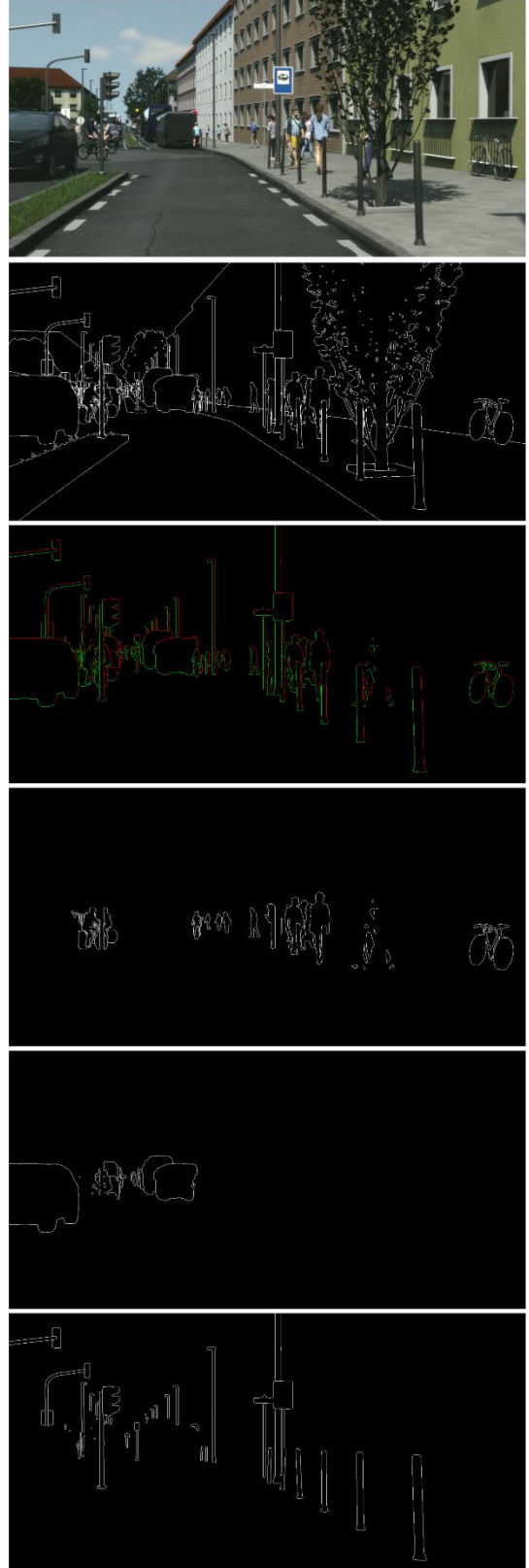


Figure 13 Contour Objects. **1st image:** RGB image (Synscapes dataset by Wrenninge et al., 2018[51]); **2nd image:** all-contour

map (for all objects), used as input to monocular TcNet01. **3rd image:** 2-channel border ownership map, including objects within three categories: ‘person’, ‘vehicle’, and ‘traffic-light/block’. **4th image:** ‘person’ category-selective map; in the original Synscapes dataset, bicycles are categorized under ‘person’. **5th image:** ‘vehicle’ category-selective map. **6th image:** ‘traffic-light/block’ category-selective map.

6.1. Experiment Setup

For this experiment, we re-use the modified Synscapes dataset from our prior work (see Section 3.1 in Chen et al., 2022[6]), which includes ground truth border ownership and category selectivity. Instead of using the RGB image (shown as the 1st image in Figure 13), we use the occluding contour map (shown as the 2nd image in Figure 13) as input to train a monocular TcNet01.

The output will include 2-channel border ownership maps along with three 1-channel category channels: ‘person’, ‘vehicle’, and ‘traffic light/blocks’, as illustrated in Figure 13.

6.2. Results

Table 9 Evaluation results for Contour Object experiments. The statistical evaluation follows the criteria outlined in Section 3.2.3. The second row corresponds to the comparable evaluation result using RGB image input (as shown in the 1st image in Figure 13) based on Table 1 from our prior paper (Chen et al., 2022[6]).

All contour channel	Border ownership channels	Category selective channels		
		Person	Vehicle	Traffic light/block
90.9%	80.8%	79.3%	65.9%	78.8%
83.9%	77.0%	76.7%	73.9%	84.5%

We conclude with the following observations:

(1) Based on Table 9 and random visual inspections, the evaluation results (row 1) are comparable to a similar evaluation (row 2) conducted on the same dataset with contrast-defined objects (refer to Table 1 in Chen et al., 2022[6], on the Synscapes dataset (c.2)).

(2) The all-contour channel (referred to as the ‘enforcement’ channel’ in Chen et al., (2022)[6], or the first branch in Figure 3), which is equivalent to the combination of the border ownership channels, performs significantly better than the other channels. This outcome is expected, given that the input consists solely of contour maps.

7. Discussion – Understanding of Border Ownership

We aim to deepen the understanding of border ownership and its related topics by integrating insights from our prior work (Chen et al., 2022)[6], the current studies presented in this paper, and supporting evidence from neuroscience. By bridging perspectives from both computer vision and neuroscience, our goal is to establish a robust foundation for understanding and simulating the border ownership-centered segmentation in biological vision systems.

While our findings contribute to a clearer conceptualization of border ownership, we acknowledge that certain claims require further substantiation.

7.1. Definition of Border Ownership and Object

Border ownership refers to the phenomenon where a visual border (also known as ‘occluding contours’) between two regions is perceived as belonging to one of the regions, referred to as the ‘border owner’. Initially, this concept applies only to ‘real’, visible borders defined by contrast or luminance differences separating objects from each other or the background; now it has since been extended to include both visible and illusory borders.

Consequently, an ‘*object*’ (in the context of object segmentation in computer vision) or a ‘*figure*’ (in the context of the figure-ground organization in neuroscience) is defined as a region surrounded by borders encoded with border ownership (except at the edges of an image). A region surrounded by borders not encoded with border ownership does not constitute an ‘*object*’ or ‘*figure*’—more precisely, it is not a *border ownership-defined object*.

Furthermore, an ‘*object border*’ is defined as any contour, whether visible or illusory, encoded with border ownership. (see detailed discussion in Section 7.11 “Edges vs Object Borders—Intrinsic Nature of Border Ownership”)

Research has shown that the border ownership-selective neurons in the visual cortex exhibit two key properties (Von der Heydt, 2015)[42]: “Border ownership preference is a fixed property of neurons” and “each neuron has a fixed preference for location of one side or the other”. These properties play a crucial role in determining figure-ground organization in neuroscience or object segmentation in computer vision.

These properties form the basis for our border ownership coding method: *opposite channel rule* (see Section 2 of this paper, or Section 2.1 in Chen et al., 2022)[6]. Using an encoder-decoder neural network, TcNet, we have demonstrated in both our prior work

(Chen et al., 2022)[6] and this paper that the coding method is applicable to monocular and binocular vision, as well as to objects defined by contrast, disparity, illusory contours, and occluding contours.

7.2. Global Context Awareness, Short-Latency, and Gestalt Principles

Calculating border ownership requires the ability to incorporate contextual information from large areas of the visual field, a capability referred to as “global context awareness”. In Section 4.2 of our prior paper (Chen et al., 2022)[6], we argued that “*global context awareness is a built-in functionality*” within neurons with border-ownership functionality and category-selective properties, describing it as a “fixed property” of those neurons. These neurons, inherently knowledgeable about ‘global context’, do not rely on top-down guidance to generate the border ownership or accurately select target categories.

All experiments conducted thus far, including monocular cases in (Chen et al., 2022)[6] or binocular and illusory cases in this paper, support the argument that top-down guidance is unnecessary for border ownership generation. The trained neural network, TcNet, learns the ‘global context’, simulating inherently knowledgeable neurons. TcNet functions like a ‘mold’ that requires no external guidance. Qiu et al. (2005)[35] provide additional neuroscience evidence further suggesting that “the existence of G cells is hypothetical”, where G cells are proposed to be aware of the global context and are assumed to provide top-down guidance for border ownership assignments.

Short-Latency

The ‘global context awareness’ embedded in border ownership-selective neurons also well explains the *short latencies* observed in border ownership generation. Zhou et al. (2000)[53] reported that “the cortical processing that leads to border-ownership discrimination requires no more than ~25msec”, a phenomenon Von der Heydt (2023)[43] described as “ultra-fast”. This rapid processing occurs because the *global context information is readily available within the neurons, eliminating the need for time-consuming transfer across neurons, and reducing sensitivity to spatial range limitations*.

The “ultra-fast” short-latency imposes a strict constraint on any model of border ownership generation, ruling out models that suggest the ‘global context’ is distant from the V2 area, where border ownership is generated.

Prior Knowledge

While ‘global context awareness’ is described as a ‘fixed property’ of neurons, it is unlikely to remain entirely ‘fixed’ forever. Instead, *‘global context’ can be viewed as a form of prior knowledge* that evolves and refines through exposure to new visual experiences or environmental changes, allowing the neurons to develop more sophisticated interpretations of contextual information.

As ‘prior knowledge’, the extended concept of ‘global context’ must include the knowledge to handle monocular and binocular vision, as well as a variety of contours for border ownership, including contrast-, disparity-, illusory-, and contour-defined contours. These contours can dynamically switch between opposite border ownership assignments, as demonstrated by the Rubin Face-Vase illusion, *depending on the ‘global context’ functioning at the moment of perception*.

Connection to Gestalt Principles

Furthermore, as noted by Qiu et al. (2005)[35], border ownership generation is influenced by Gestalt principles. Principles such as closure and proximity likely form the foundation of the ‘prior knowledge’ embedded in border ownership-selective neurons, enabling coherent segmentation and perceptual grouping. Thus, it is reasonable to conclude that *the ‘global context awareness’, or ‘prior knowledge’, is an embodiment of these principles*, providing a robust framework for interpreting complex visual scenes.

7.3. Distributedness

The assertion that the ‘global context’ is embedded within border-ownership selective neurons implies that it is distributed across these neurons, rather than centralized.

This perspective indirectly explains why the hypothetical centralized “G cells” proposed by Qiu et al. (2005)[35] are unlikely to exist. Similarly, this implies that the centralized “object pointer”, proposed by Von der Heydt et al. (2018)[45] as a centralized representation of objectness (encompassing an object, its identity, and features), unlikely to exist either, particularly not in the V2 area where border ownership is generated. Notably, Von der Heydt et al. (2018)[45] acknowledge the absence of such a centralized ‘object pointer’ in the V2 area. Instead, the completion of border ownership likely to a great extent, signifies the completion of object segmentation or ‘figure-ground organization’, suggesting a form of ‘distributed objectness’ or ‘distributed object pointer’.

The concept of ‘distributed objectness’ is not novel and is referred to as ‘distributed representation’ in works by Roelfsema (2006)[36] and Roelfsema et al. (2011)[37]. However, our findings provide supportive evidence from a perspective distinct from neuroscience, elaborating on its likely characteristics and implications.

Distributed Memory

O’Herron et al. (2009)[29] proposed the existence of short-term memory for figure-ground organization, suggesting that such memory ensures persistence and continuity of perception involving border ownership. If this memory exists, we argue that it is more likely distributed rather than centralized. A distributed architecture offers greater efficiency than a centralized one in terms of memory usage and updating or tracking per-neuron object changes.

In a distributed architecture, the upper limit of objects tracked per neuron is predictable and equal to the maximum number of pixels covered per neuron. By contrast, the upper limit for neurons in a centralized architecture could be significantly higher and unpredictable, making centralized architecture inherently less efficient.

Conclusion

In conclusion, border ownership generation at the early visual stage (V2 area) is most likely a *distributed process*, with *distributed object representation* such as ‘distributed object pointers.’ This concept will be further examined in Section 8, where we propose a workable solution for ‘surface filling-in’. If a centralized ‘object pointer’ does exist, it may emerge during subsequent visual processes, such as those involving focused attention.

7.4. Category Selectivity

Our prior work (Chen et al., 2022)[6] and this paper have demonstrated that category selectivity can be encoded and generated alongside border ownership, aligning perfectly with the neuroscience evidence from Silson et al. (2022)[40] and Uyar et al. (2016)[41]: the “*human visual cortex is organized broadly according to two major principles: retinotopy and category-selectivity*”. Furthermore, these studies highlight that “*retinotopic maps and category-selective regions show considerable overlap*”, where ‘retinotopy’ refers to the spatial mapping of the retina onto the cortex, closely related to border ownership-centered figure-ground organization.

As noted in Section 2.2 in our prior work (Chen et al., 2022)[6], the “*channel representation makes things*

simple” such that “*border-ownership generation and category-selective generation can be done in one shot.*” Consequently, the “*considerable overlap*” between “*retinotopic maps and category-selectivity*” can be efficiently represented by the *inner product* of border ownership maps and category-selective maps. These representations can be *jointly generated* by a neural network, as illustrated in Figure 3 in this paper (and Fig. 3 in Chen et al., 2022[6]).

Additionally, as noted in Section 4.4 in our prior work (Chen et al., 2022)[6], our method of joint coding and generation of category selectivity alongside border ownership provides insights into why categories are pre-selected and why the number of categories is inherently limited. In our approach, each category is represented by a single channel, making it practically unfeasible to use thousands of channels to represent thousands of object categories. This limitation aligns with neuroscientific findings by Wardle & Ritchie (2014)[47], which suggest that the visual system’s category-selective architecture is inherently constrained to a finite set of categories.

7.5. Monocular and Binocular

We have demonstrated that the border ownership and category-selective coding method functions effectively in both monocular and binocular settings using an encoder-decoder neural network, TcNet. The only exception is the Random Dot Stereograms (RDS) dataset, which fails in a monocular neural network, as expected. Since the left image of an RDS pair consists entirely of random dots, while the right image is a modified version of the left, appearing random unless compared to the left image.

A binocular scenario involving two eyes (cameras) can also be conceptualized as a motion view captured at two locations by a single eye (camera). When motion approaches zero (static), these two locations converge into one, effectively reducing the binocular view to a monocular one. Thus, monocular processing can be considered a special case of binocular processing.

To explore this relationship, we intentionally designed the binocular TcNet02 to closely resemble the monocular TcNet01 with minimal modification, differing only in the number of input channels. Except for cases like the RDS dataset, monocular processing aligns closely with binocular processing. The slight underperformance (or comparable performance, considering potential errors) of monocular TcNet01 compared to binocular TcNet02 on the VKitti dataset (see Section 3.2.4(8)) provides empirical support for this view.

Critique of the Binocular Fusion Model

We disagree with the model proposed by Dresch-Langley et al. (2016)[12], which suggests that the binocular boundaries are fused from the left and right monocular boundaries. This model has two key limitations: First, the fusion process requires correspondence matching between left and right monocular boundaries, introducing additional processing time and contradicting the ‘short latency’ constraint; Second, the fusion model does not fully leverage binocular disparity when available. Supporting this, Qiu et al. (2005)[35] observed that disparity-selective neurons contribute to border ownership generation, highlighting the importance of directly incorporating binocular disparity rather than relying on post hoc fusion.

7.6. Binocular Disparity and Depth

Numerous neuroscience studies indicate that binocular disparity plays a role in contour representation in V1 and V2. Qiu et al. (2005)[35] observed that disparity-selective neurons contribute to border ownership generation, though the exact mechanisms of this contribution remain unclear.

Human vision is capable of perceiving a wide range of depth smoothly. Neuroscience evidence from Poggio (1995)[32] suggests the presence of disparity-selective neurons that are sensitive to ‘Near’, ‘Far’, and other disparity types, though Prince et al. (2002)[33] found no distinct neuron classes for these disparities. Nevertheless, the presence of these disparity types, whether generated by different neurons or not, underpins our ‘*layered disparity representation*’ method. In this approach, disparities such as ‘Far’, ‘Near’, and others are organized into separate channels, enabling effective border ownership generation (Sections 3.1 and 3.2) and broader applications.

Layered Channel Representation

The layered channel representation of disparities aligns closely with border ownership representation, offering flexibility for incorporating additional disparity types (Poggio, 1995)[32] or motion cues such as optic flows in future extensions. Optic flows, typically represented as 2-channel ‘horizontal’ and ‘vertical’ flow maps, share structural similarities with disparity representations (Glennister et al., 2018)[15]. This suggests that motion cues could function similarly to disparities in border ownership generation.

For small depths, human vision follows Epipolar geometry, where ‘Near’ disparity is represented using Epipolar disparity (Eq. 3.1). However, this representation becomes insufficient for large depths where disparities approach or fall below the noise level. To address this, we

represent ‘Far’ disparity using the logarithm form (Eq. 3.2) of ‘Near’ disparity, *flattening the magnitude differences* between small and large disparities. This ensures that large disparities at short depths do not overwhelm smaller disparities at long depths, while preserving the depth order of objects and backgrounds. These adjustments are crucial for effective border ownership generation and beyond, as demonstrated successfully in Section 3.2.

Absolute and Relative Disparity

Cumming et al. (1999)[9] noted that “binocular neurons in V1 are selective for absolute, not relative, disparity”. Both ‘Near’ and ‘Far’ disparities are forms of ‘absolute disparity’ and are generated in the V1 visual cortex, making them readily available for border ownership generation in V2 and subsequent stages. While the role of ‘relative disparity’ in border ownership generation remains unclear, it is feasible to handle ‘relative disparity’ as ‘absolute disparity’ within the ‘layered disparity representation’ framework.

Depth and Disparity

Unlike human vision, which has disparity-selective neurons but not direct depth-selective neurons, computer vision systems can leverage direct depth sensors (e.g. laser radar) but lack disparity sensors. Referring to the re-positioned Epipolar equation (Eq. 3.1’) and its logarithmic form (Eq. 3.2’), disparity and depth occupy equivalent positions. This suggests that *the ‘layered disparity representation’ could be extended to depth or combined depth/disparity representation in computer vision systems*. Such representations could significantly improve object segmentation accuracy in complex environments, such as autonomous systems, where depth ranges vary widely.

Absolute vs Relative Values

For the purpose of border ownership generation, *the relative value of disparity (or depth) is more important than its absolute value*. Consequently, the constants k in (Eq. 3.1) and C in (Eq. 3.2) may be adjusted from their original meanings during neural network training to ensure neither disparity map overwhelms the other. For instance, a scale-down factor of 10 was applied to the constant k in (Eq. 3.1) for the VKitti dataset (see Section 3.2.2.2). However, it is important to distinguish between relative and absolute value of disparity in this context and the neuroscience concepts of relative and absolute disparity (see Section 3.1).

7.7. Border Ownership with or without Explicit Binocular Disparities

In Section 3.2, we conducted experiments on border ownership generation with varying levels of binocular

disparity involvement using two datasets: Random Dot Stereograms (RDS) and a modified photorealistic Virtual Kitti (VKitti). The RDS dataset, consisting of disparity-defined objects, is more relevant to neuroscience studies. In contrast, the VKitti dataset, composed of contrast-defined objects spanning a wide depth range and closely resembling real-world conditions, is of greater interest for computer vision applications.

These experiments revealed both distinct and shared behaviors across the two datasets:

Applicability across Datasets:

Our border ownership coding method applies to both datasets—disparity-defined and contrast-defined objects in binocular scenarios (see Section 3.2.4(1)).

Performance with Combined Disparities:

For both datasets, the border ownership generation using both ‘Far’ and ‘Near’ disparities outperforms using only one disparity type (see Section 3.2.4(4)), with this effect being especially notable for the photorealistic VKitti dataset.

Impact of Explicit Disparities:

For the RDS dataset, the border ownership generation with one or both ‘Far’ and ‘Near’ disparities significantly outperforms the generation without explicit disparity (see Section 3.2.4(2)), suggesting that disparities can be crucial in segmentation under camouflage conditions.

Conversely, for the VKitti dataset, border ownership generation with both disparities performs comparably to generation without explicit disparities (see Section 3.2.4(3)), indicating that TeNet02 can implicitly learn the disparities for the photorealistic VKitti scenes.

Role of Combined Disparities in VKitti Dataset:

The similar contribution of ‘Near’ and ‘Far’ disparities and the comparable performance of border ownership generation with and without explicit disparities for the VKitti dataset (see Section 3.2.4(5)) suggest that the Epipolar ‘Near’ disparity alone is insufficient for disparity representation. The logarithmic ‘Far’ disparity effectively complements the ‘Near’ disparity, aligning with our analysis in Section 3.1. *Together, ‘Near’ and ‘Far’ disparities efficiently represent the wide depth range characteristic of the photorealistic VKitti dataset, which closely resembles real-world conditions.*

Disparity Contribution in RDS Dataset:

For the RDS dataset, the ‘Near’ disparity contributes significantly more than the ‘Far’ disparity and performs comparably to combined disparities (see Section 3.2.4(5)). This suggests that the contribution of ‘Far’ disparity has

negligible impact, likely because RDS primarily consists of nearby objects. Alternatively, the perceived depth in RDS may be more of an *imaginary depth rather than actual depth*, given the absence of a defined baseline or focal length to establish its depth range.

Insights and Future Directions:

These experiments represent the first systematic attempt to simulate and investigate border ownership generation involving binocular disparities. While they reveal meaningful behaviors across datasets, further studies are needed to deepen our understanding of how disparity influences border ownership generation. The ground truth for binocular disparities and border ownership in the RDS dataset is highly accurate. However, in the modified photorealistic VKitti dataset, ground truth reliability is lower, due to inaccuracies in depth estimation along tilted borders, with an approximate accuracy of ~70% based on visual inspection. Although the observed common behavior across datasets is significant, their reliability requires further investigation, especially for real-world images like those in the VKitti dataset.

7.8. Illusory Objects

We demonstrated border ownership generation for several types of illusory objects (see Section 4), including Kanizsa figures, the Ehrenstein illusion, Neon color spreading, and Illusory shading (excluding Kanizsa figures with synthetic disparity, discussed separately in Section 7.9). The generation process for these illusory objects closely mirrors that of contrast-defined objects, aligning with the neuroscience findings by Von der Heydt et al. (1984)[45], which shows that illusory contours are extracted in the V2 region—the same region where binocular border ownership for contrast-defined objects is generated (Qiu et al., 2005)[35].

This alignment is further supported by observations by De Haas et al. (2018)[10], who noted “few systematic differences between retinotopic responses to illusory contours, occlusion, and subtle luminance stimuli in the early visual cortex”. These findings reinforce the robustness of neural mechanisms underlying border ownership generation across both illusory and contrast-defined contours, suggesting shared processing pathways, (see further discussion in Section 7.11).

As illustrated in Section 4.2, border ownership for ‘illusory objects’ was generated based on ‘prior knowledge’ learned during training. Additionally, experiments on the Ehrenstein illusion and Neon color spreading (Section 4.2(3)) revealed that the accuracy of border ownership generation for ‘illusory objects’

decreases as transparency increases (or opacity decreases), aligning with human perception, where higher transparency reduces the clarity of illusory contours.

7.9. Illusory Objects, Resulting from Depth Cues?

Stanley Coren (1972)[7] suggested that “both monocular and binocular subjective contours result from the presence of depth cues”. However, the findings of our experiments (see Section 4.2(4)) do not support this hypothesis. Specifically, border ownership generation with synthetic disparity performed significantly worse than without it, contrary to the expectation that performance would improve or remain comparable if Coren’s claim were correct. Instead, the results align more closely with the theory by Qiu et al. (2007)[34], which attributes the observed floating effect of subjective objects to ‘attention enhancement,’ with minimal or no involvement in border ownership generation. Unlike datasets such as the VKitti and RDS (see Section 3.2.2), 2D Kanizsa shapes lack real depth or disparity cues, further challenging Coren’s hypothesis.

A similar argument applies to the Rubin Face-Vase illusion. In this case, the object of attention-focus holds the border ownership and benefits from ‘attention enhancement.’ As with subjective objects, this enhancement appears unrelated to the border ownership generation. Instead, the border ownership is determined by the dominant ‘global context,’ which dictates whether ‘face-as-object’ or ‘vase-as-object’ interpretation prevail (see the Rubin Face-vase experiment results in Section 5.2, and the arguments in Section 7.2).

However, if Random Dot Stereograms (RDS) are considered a type of illusion, as Coren (1972)[7] suggested, their border ownership generation does depend on one specific depth cue – disparity.

In conclusion, *depth cues are not strictly necessary for border ownership generation on illusory objects, whether monocular or binocular*. However, depth cues may play a role in attracting increased attention, indirectly influencing perceptual organization without directly impacting the generation of border ownership.

7.10. T-junctions and Other Key Points

A T-junction refers to a visual phenomenon where two intersecting borders form a ‘T’ shape. It has been proposed (Von der Heydt, 2015[42]; Von der Heydt et al., 2018[45]) as a cue for determining border ownership (see Figure 14, Left): the horizontal line of ‘T’ occludes the

vertical line of ‘T’, making the ‘above’ side of ‘T’ the border owner.

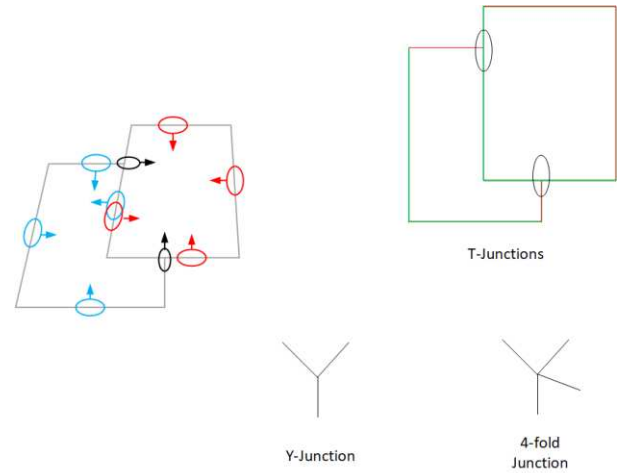


Figure 14 T-junction and Y-junctions (adapted from Fig. 4 in Chen et al., 2022[6]). **Left:** Remade from part of FIGURE 5 of Von der Heydt, (2015)[42] showing with two overlapping rectangular objects as an example. **Top Right:** The two overlapping rectangles are coded with our 2-channel border ownership coding method; the overlapping segments at T-junctions (black-color circled) are coded in different channels. **Bottom Middle:** An imperfect T-junction resembles a Y-junction, demonstrating the difficulty of determining occlusion relationships with traditional methods. **Bottom Right:** A 4-fold junction example.

In our prior work (Chen et al., 2022)[6], we presented a simplified solution (see Figure 14, Top Right) to the example on the left, arguing that *the T-junction is merely an irrelevant visual perception*, and is an *unreliable* cue for the determination of border ownership in cases such as Y-junction (Figure 14, Bottom Middle) or even more complex N-fold junctions where $N \geq 4$ (Figure 14, Bottom Right).

For tasks involving ‘key’ point searching, a 4-channel version of our border ownership coding method is likely more suitable. The overlapping points between border ownership channels are the candidates, depending on their relevance or interest.

While T-junctions are unreliable from a current computer vision perspective (e.g., identifying overlapping points between channels), neuroscience evidence (Craft et al., 2007[8]; Heitger et al., 1992[19]) suggests that end-stopped cells play a critical role in processing corners and junctions. This implies that biological vision systems may utilize a more sophisticated mechanism for interpreting junction. We will explore this further in Section 8.4.

7.11. Edges vs Object Borders—Intrinsic Nature of Border Ownership

Pioneering studies by Hubel and Wiesel (1959)[20] demonstrated that neurons in the primary visual cortex (V1) are specifically responsive to edges and lines of particular orientations, forming the foundation of visual processing. These edges and lines are ‘real’ and visible, arising from contrast or luminance differences in the visual field.

The border ownership model proposed by Von der Heydt (2015)[42] and Craft et al. (2007)[8] assumes a two-step process: first, occluding contours are generated from visible edges; and second, border ownership is ‘assigned’ to these contours based on cues such as T-junctions. Consequently, the border ownership generation is often referred to as border ownership assignment.

However, cases involving illusory contours challenge this assumption. Illusory contours, such as those seen in Kanizsa figures, highlight the brain’s ability to infer object boundaries without explicit visual edges. Our experiments (see Section 4) demonstrate that the border ownership-defined borders for illusory contours can be generated with accuracy comparable to those for contrast-based and disparity-based contours, suggesting that explicit post-hoc ‘assignment’ *may be unnecessary*. Instead, border ownership generation appears to be an intrinsic process occurring simultaneously with contour formation.

This perspective eliminates the need for separate stages, simplifying the understanding of figure-ground organization and aligning more closely with the rapid, early processing in biological vision.

Redefining ‘Object Border’

An ‘*object border*’ is defined as any contour, whether visible or illusory, encoded with border ownership, as established in Section 7.1. This expanded definition highlights the inclusion of both visible and inferred boundaries, emphasizing that object borders are not strictly tied to contrast-based edges. This broader perspective is particularly relevant in cases where the brain constructs boundaries without explicit visual edges.

This further suggests that border ownership-defined contours *intrinsically function as abstract, ‘invisible’ boundaries* that encode object structure beyond local contrast cues. *Their alignment with visible edges in contrast-defined cases is merely incidental rather than fundamental*. This supports the argument in Section 7.2 that border ownership-selective neurons integrate global context and prior knowledge, enabling the recognition of ‘object borders’ even in the absence of visible edges.

Redefining ‘Border Ownership Assignment’

Our findings suggest that the term of ‘*border ownership assignment*’ *should be redefined — not as a process but as a state or status of ‘object borders’*. This redefinition aligns with experimental evidence presented in our prior work (Chen et al., 2022)[6] and further validated in this paper, demonstrating that border ownership generation occurs concurrently with contour (‘object border’) formation. This applies across all object types, including contrast-, disparity-, illusory-, and contour-defined objects, without the need for additional post-hoc ‘assignment.’

To better reflect this intrinsic process, we propose renaming it ‘*object border generation*’. This terminology underscores the simultaneous and intrinsic nature of contour formation and border ownership encoding, emphasizing their interdependence as a unified operation.

Implications for Vision Models

For neuroscience: This shift in perspective suggests that border ownership is an intrinsic property of neural processing, arising naturally during contour formation rather than being assigned afterward. This redefinition challenges traditional models and highlights the efficiency of visual cortex processing to produce coherent figure-ground organization.

For computer vision: The findings emphasize the potential for models to generate border ownership inherently, simplifying processes for object recognition and segmentation. This is particularly impactful in challenging scenarios, such as those involving illusory or ambiguous contours, where conventional edge-based approaches might fail.

By unifying the understanding of visible and inferred boundaries, this perspective provides a robust framework for advancing both biological and computational vision systems.

8. Surface Filling-in

With a comprehensive understanding of border ownership established through experimental and theoretical investigations, we now address the complementary process of surface filling-in. This stage represents the second step towards the completion of the figure-ground organization process, following border ownership generation (see Section 1.2.2).

Many existing approaches treat surface filling-in as a separate process dependent on feedback from higher

cortical areas. In contrast, we propose a method that directly links border ownership signals to surface filling-in. This feedforward approach simplifies the process, avoids reliance on higher-order feedback, and enhances computational efficiency.

Models such as those described by Grossberg (2015)[16] and Jeurissen et al. (2013)[21], treat surface filling-in as an iterative process, requiring boundary completion and feedback from higher cortical areas. While effective for figure-ground organization, these models do not establish a direct link between border ownership and surface filling-in. In contrast, our approach addresses this gap by proposing that border ownership signals alone can serve as the foundation for surface filling-in, enabling the efficient and robust creation of ‘distributed object pointers’ and streamlining figure-ground organization.

8.1. Constraints

Based on neuroscience evidence and our experimental findings, the following *constraints* must be addressed during the ‘surface filling-in’ process to generate ‘distributed object pointers’ with border ownership.

(1) *Unclosed contours*: The process must account for objects with open or incomplete contours, as self-occluding objects, cluttered environments, or ambiguous object boundaries can prevent full closure.

(2) *Short latency*: ‘Surface filling-in’, corresponding to the finalization of figure-ground organization, likely lags slightly behind border ownership generation, with a short latency of 30 ~ 40 msec (Lamme, 1995)[26].

(3) *Owner-side direction*: The process must explain why border ownership enforces a perpendicular (orthogonal) owner-side direction relative to the occluding contours, distinguishing it from conventional segmentation methods (see Section 1.2.2).

8.2. Framework for Surface Filling-in

From Border Ownership to Surface Filling-in

Lamme et al. (1999)[27] suggested that “a model of segmentation where boundary formation initiates surface filling-in” is favorable based on their observations. In neuroscience, ‘surface’ is often synonymous with ‘figure’ or ‘object’. Thus, the idea that “boundary formation initiates surface filling-in” can reasonably be interpreted as *border ownership contours initiating the ‘surface filling-in’ process*.

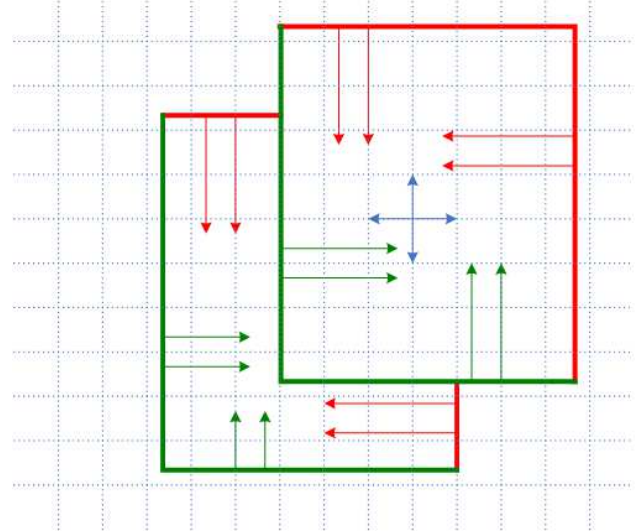


Figure 15 Mask Completion for ‘Surface filling-in’. This figure illustrates the ‘surface filling-in’ process for two overlapping rectangle objects, using a 2-channel border ownership coding.

-- In the setup: **red-colored borders** indicate the owner side is either ‘below’ or ‘left’, with red arrows representing the direction of ownership (referred to as ‘owner-side direction’) that is orthogonal to the border ownership contours; **green-colored borders** indicate the owner side is either ‘above’ or ‘right’, with green arrows similarly pointing towards the ‘owner-side direction’; **blue-colored arrows** depict the connection propagation direction for non-border pixels (neurons), either inside or outside objects.

-- Surface regions are progressively filled by propagating ownership signals from border ownership contours along the ‘owner-side direction’ until encountering another ownership signal or until the image boundary is reached.

-- Ownership is resolved as follows: if a red point connects to a green point (or vice versa), they belong to the same object. Conversely, if a green point connects to a green point (or a red point to another red point), they belong to separate objects unless connected from another ‘owner-side direction’.

Framework

To simplify the description, we assume each neuron corresponds to a single pixel. Given a border ownership map, our framework for the ‘surface filling-in’ process *begins at each pixel (neuron) along the border ownership contours* (as depicted in Figure 15) and propagates the ownership signal by connecting (binding) pixels along the ‘owner-side direction,’ which is orthogonal to the border ownership contours. The propagation continues until it encounters another ownership signal originating from a border ownership contour, or it reaches the edge of the image.

Through this process, connected pixels collectively form coherent object regions or background areas, and their identity is assigned as an ‘object pointer’. This ensures that each region is uniquely identified within the figure-ground organization process.

Performance Considerations

The simplest implementation involves scanning (pixel by pixel) vertically and horizontally along the ‘owner-side direction’ in the 2-channel border ownership coding maps (Figure 15). While straightforward, this sequential scanline method is relatively slow, necessitating additional mechanisms to meet the ‘short latency’ constraint discussed in Section 8.1.

Active Neurons (Pixels)

As argued in Section 7.3 “Distributedness”, the most feasible and effective way to shorten processing time is *for each neuron to function as an independent, active unit (referred to as ‘active pixel’ or ‘active neuron’)*. Active neurons independently initiate connection attempts with their neighboring neurons, eliminating the need for sequential processing. This parallel processing approach not only significantly reduces processing time but also aligns seamlessly with the ‘distributed’ characteristics of neurons discussed in Section 7.3.

Further Performance Consideration

The ‘surface filling-in’ process, which *can begin at any point* along the border ownership contour, inherently suggests that neurons operate with some level of independence. Theoretically, connection initiation is not restricted to border neurons—non-border neurons can also initiate connections (as depicted by the ‘blue-colored arrows’ in Figure 15). This parallel processing not only accelerates the process but also maintains consistency with the distributed nature of neural processing.

Handling Errors and Ambiguities

Under ideal, error-free conditions, the implementation of the above framework would be straightforward and require no additional cues. However, in practice, border ownership maps may contain errors or ambiguities (see Section 8.1). To mitigate these issues, supplementary cues can be incorporated to refine the decision-making process for initiation and connection. These cues may include disparity maps or feature maps (such as those extracted from the Encoder in TcNet, see Figure 3). When available, they provide valuable contextual information, improving the accuracy of the ‘surface filling-in’ process and significantly enhancing the overall robustness of the framework.

8.3. Active Neurons (Pixels)

Active neurons act as *local processing units* that operate independently to some extent while working collaboratively to achieve figure-ground organization (segmentation) at the early vision stage with short latency. They perform the following core functions:

(1) *Contour Connection*: Neighboring neurons along contours are likely connected as part of border ownership generation before the ‘surface filling-in’ process begins, facilitating smoother ownership propagation across the surface.

(2) *Connection Initiation*: Initiate the connection or ‘binding’ process from border ownership contours to propagate ownership;

(3) *Object Identity Maintenance*: Ensuring consistent labeling of connected pixels within the same object region (or background) as connections propagate or modify.

(4) *Similarity Assessment*: Evaluate the owner-side direction and assess similarity with neighboring neurons based on available cues, such as disparity maps or feature maps, to refine connections, especially in ambiguous or cluttered regions.

The distributed processing enabled by ‘active neurons’ ensures short latency, addressing the timing constraints described in Section 8.1(2). By evaluating similarities with neighboring neurons, these neurons also help resolve ambiguities in contours and cluttered visual cases, meeting the constraints outlined in Section 8.1(1).

If each active neuron manages more than one pixel, as opposed to the assumption in the previous section, the implementation logic becomes more complex. In this scenario, each neuron would need to process and integrate information from multiple pixels simultaneously, with the complexity depending on the number of pixels each neuron manages and the nature of their connections. While this approach warrants further exploration, for practical computer vision applications, assuming one pixel per neuron provides a straightforward and effective solution.

8.4. Active Neurons and End-Stopped Cells

Neuroscience evidence (Craft et al., 2007[8]; Heitger et al., 1992[19]) suggests that end-stopped cells play a crucial role in processing corners and junctions, which are essential features in contour detection, formed in V1 before border ownership. Building on this insight, we propose a workable mechanism where end-stopped cells guide active neurons in performing one of their core functions (‘Contour Connection’) to accurately connect border ownership contours.

Visual inspection on border ownership maps, such as Figure 11, reveals that connecting different border ownership contour segments (red and green in different channels) at overlap points often requires guidance to ensure accuracy. Specifically:

(1) *At non-junctions or non-terminations*, two (red and green) contour segments should belong to the same continuous contour with consistent ownership.

(2) *At junctions or terminations*, the contour segments should separate to represent distinct objects or regions.

(3) *In illusory contour cases with no end-points*, follow the rule in case (1).

Without such guidance, active neurons may incorrectly propagate ownership across unrelated contours, leading to errors. Notably, explicitly classifying specific junction types (e.g. T-junctions) is unnecessary, which helps avoid potential ambiguities in complex scenes.

Challenge: Generating End-Stopped Points in Computer Vision

While end-stopped cells in biological vision systems reliably detect corners, junctions, and terminations, replicating this process in computer vision presents a significant challenge. The difficulty lies in accurately and consistently identifying critical points where contours should either connect or terminate, especially in cluttered or ambiguous environments. Developing a robust method for reliably generating end-stopped points in computer vision is beyond the scope of this paper.

8.5. Representation of Object Pointer

From a segmentation perspective, connected pixels in the V2 area are grouped into ‘distributed objects’, marking the completion of segmentation process. The ‘object pointers’ to these objects are only needed when referenced by subsequent vision processes, such as object-based attention (Qiu et al., 2007[34]).

The actual representation of ‘object pointers’ depends on their role and interaction within the broader visual hierarchy. Practically, an ‘instance mask’ or ‘panoptic mask’ serves as an effective representation by assigning each object a unique identity, defining its spatial location and coverage.

Additionally, integrating the border ownership and disparity maps provides further dimensions to the representation. These maps enhance the segmentation by incorporating depth and spatial relationships, offering a richer understanding of the visual scene.

However, this approach currently results in a loosely coupled representation of instance masks, border ownership, and disparity maps. A more tightly integrated representation may be achievable by further leveraging the potential capabilities of ‘active neurons,’ optimizing efficiency and coherence in figure-ground organization.

8.6. Advancements over Existing Models

Our approach introduces a distinct advancement over existing models, such as those proposed by Grossberg (2015)[16] and Jeurissen et al. (2013)[21]. Grossberg’s model integrates depth cues and Gestalt principles through feedback mechanisms, while Jeurissen emphasizes figure-ground modulation using top-down processes. While these models are biologically comprehensive and effective for figure-ground organization, they involve significant computational complexity due to hierarchical and feedback-dependent processes. Moreover, these models contradict our findings, which suggest that top-down guidance is not required and that Gestalt principles function as ‘prior knowledge’ inherently embedded across neurons, as discussed in Section 7.2.

In contrast, our method establishes a direct, feedforward link between border ownership signals and surface filling-in. By leveraging distributed, parallel processing and propagating local ownership signals through ‘active neurons’, our approach ensures computational efficiency and real-time performance, while aligning closely with the ‘short latency’ constraint (see Section 8.1). These characteristics make our method particularly suitable for computer vision tasks where scalability and efficiency are paramount.

9. A Border Ownership Centered Figure-ground Organization Model

Building on the insights gained, particularly from the elaboration in Sections 7 and 8, we propose a model (Figure 16) for border ownership-centered figure-ground organization (segmentation). This model integrates multiple types of absolute disparities and motion cues (optic flows), where absolute disparities are generated in the V1 area, and border ownership-centered figure-ground organization is formed in the V2 area. Within this framework, embedded ‘global context awareness’ functions as ‘prior knowledge’, guiding the generation of border ownership with the output of occluding contour maps, border ownership maps, and category-selective maps. These outputs apply across various object types, including contrast-, disparity-, illusory-, and contour-defined objects. Border ownership then directs the ‘surface filling-in’ process, utilizing available cues like disparity or extracted features for effective figure-ground organization. Border ownership-selective neurons in the V2 area operate as distributed ‘active neurons’, responsible for generating border ownership, performing ‘surface filling-in’, and maintaining their persistence.

Referring to Figure 16, from a computer vision perspective, the proposed channel representation simplifies the integration of diverse inputs and outputs — including RGB images, disparities (‘Near’, ‘Far’ and other types of disparities), optic flows, contours, border ownership, and category selectivity throughout the visual processing pipeline. Similar structured representations may exist in biological vision systems. Using an encoder-decoder neural network, such as TeNet, this model can be trained to embody ‘global context awareness’ and simultaneously generate contour maps, border ownership maps, and category-selective maps. ‘Active neurons (pixels)’ subsequently generate the ‘object pointers’ or ‘instance masks’ required for the figure-ground organization (object segmentation).

From a computer vision perspective, border ownership generation and surface filling-in are often treated as separate processes. However, in biological vision, both processes may be managed by the same group of neurons, ensuring the persistence of figure-ground organization (O’Herron et al., 2009)[29].

Compared to existing neuroscience models of border ownership coding (Williford et al. 2013)[50], our proposed model offers a more comprehensive explanation of several critical phenomena, including: the ‘short latency’ of “ultra-fast” border ownership generation, the role of binocular disparities in border ownership processing, the unified ability to handle various object types, such as contrast-, disparity-, illusory- and contour-defined objects, and the ‘considerable overlap’ between retinotopic maps and category selectivity.

The outputs of the model for the figure-ground organization correspond to ‘distributed objects’ or ‘distributed object pointers’ from a neuroscience perspective, or ‘instance mask’ or ‘panoptic mask’ maps from a computer vision perspective.

10. Limitations and Future Directions

While the proposed border ownership-centered figure-ground organization model demonstrates significant advancements, certain constraints should be acknowledged. A primary challenge lies in the lack of publicly available datasets that align with our coding method, particularly those incorporating depth and binocular information. This limitation restricts broader

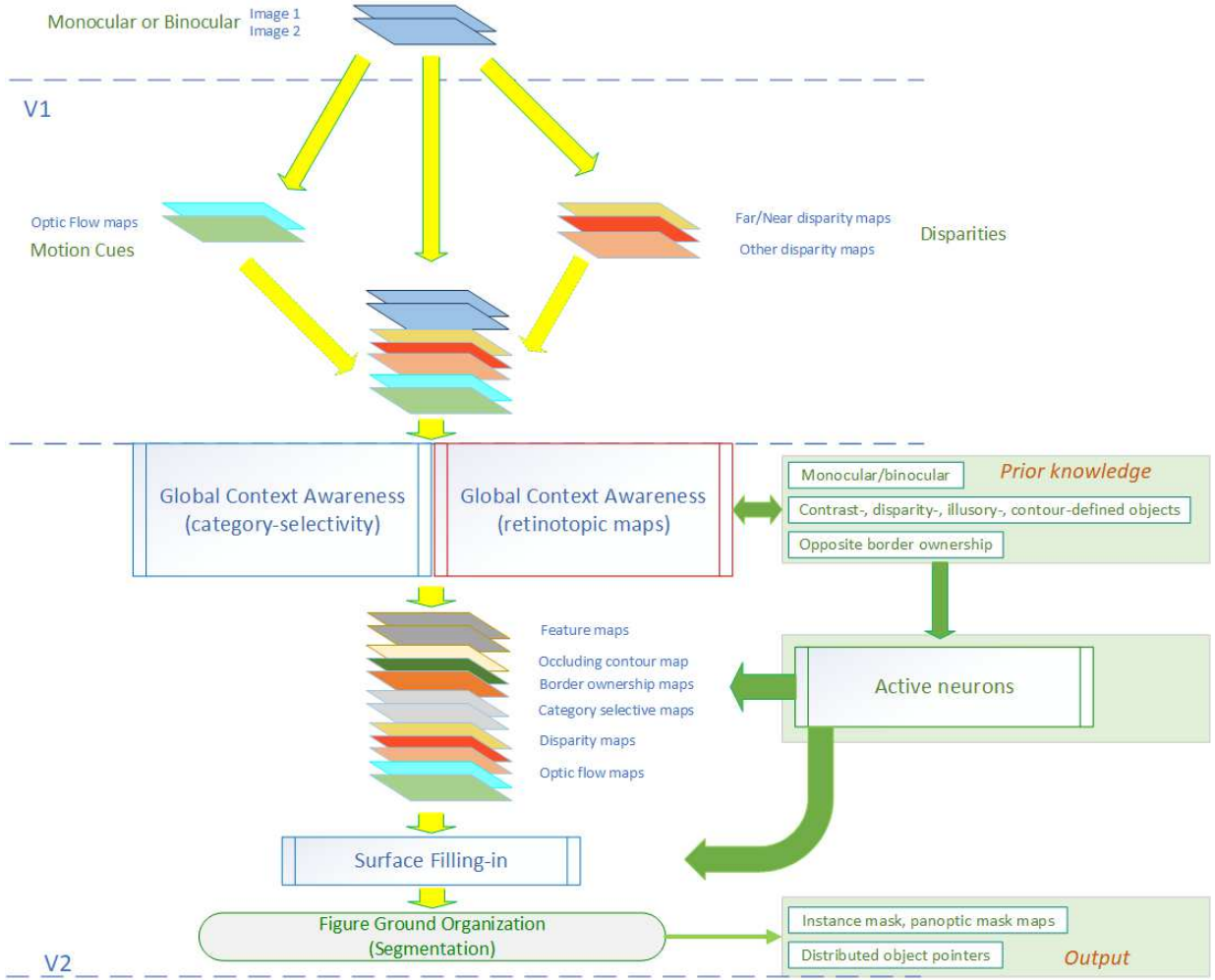


Figure 16 Architecture Diagram for the Proposed Border Ownership-Centered Figure-Ground Organization (Segmentation) Model. The architecture integrates neuroscience-inspired elements such as global context awareness, functioning as ‘prior knowledge’, and active neurons. Inputs, including RGB images, disparities, optic flows, as well as outputs, such as contours, border ownership, and category selectivity, are represented using a channel-based format. This representation facilitates seamless integration across the visual processing pipeline, enabling efficient and scalable segmentation of contrast-, disparity-, illusory-, and contour-defined objects.

validation of the model and its application across diverse scenarios.

Despite these constraints, our findings establish a strong foundation for future research and validation. By demonstrating the models’ effectiveness using available datasets, we highlight its scalability and potential for generalization to more complex visual tasks. Future work should focus on developing tailored datasets, exploring unsupervised or self-supervised learning approaches, refining the model for broader applicability, and enhancing its integration within both neuroscience and computer vision frameworks.

11. Summary

This paper advances the understanding of border ownership as a central mechanism in early vision, addressing key challenges in figure-ground organization. By integrating insights from neuroscience and computational methods, we propose a unified framework that connects border ownership-centered segmentation with diverse visual phenomena, including category selectivity, disparity and depth perception, and the processing of contrast-, disparity-, illusory-, and contour-defined objects.

Our model introduces a direct, feedforward link

between border ownership generation and surface filling-in, leveraging distributed, parallel processing and active neurons. This framework offers a computationally efficient and biologically plausible explanation for critical visual processes, including the intrinsic nature of border ownership, ultra-fast border ownership generation, the integration of global context awareness, and the generation of distributed object pointers.

By bridging gaps between neuroscience and computer vision, our findings introduce practical methods for real-time, scalable segmentation while offering new insights into biological vision mechanisms. These advancements underscore the fundamental role of border ownership-centered figure-ground organization, reinforcing its significance for both biological and artificial vision systems.

Author Contributions

- **Tianlong Chen:** Conceptualization, methodology, experimental design and conduction, data preparation, result analysis and discussion, writing, and editing.
- **Xuemei Cheng:** Data preparation, result discussion, paper review, and providing critical feedback.
- **Thomas Tsao:** Supervision and paper review; mentor for conventional methods prior to deep learning approaches.

Data Availability

One of the datasets generated and analyzed during the current study (Section 3.2.2.2) is publicly available on Zenodo under the following citation:

Chen T. (2025)[55]. *Border Ownership and Category Annotation Extended from Virtual KITTI 2* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15375453>.

Other public datasets used in this study, including the original Virtual KITTI 2 dataset and Synscapes, are available from their respective sources as referenced in the manuscript.

For further inquiries, please contact the corresponding author.

References

1. Arbeláez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour Detection and Hierarchical Image Segmentation, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898-916, May 2011, doi: 10.1109/TPAMI.2010.161.
2. Aubry, M., Maturana, D., Efros, A.A., Russell, B.C., and Sivic, J. (2014). Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models. *IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA*, 2014, pp. 3762-3769, doi: 10.1109/CVPR.2014.487. FlyingChair dataset: <https://lmb.informatik.uni-freiburg.de/resources/datasets/GenerateChairs.en.html>
3. Breitmeyer, B.G. (2013). Contributions of magno- and parvocellular channels to conscious and non-conscious, *Phil. Trans. R. Soc. B* 369: 20130213. <http://dx.doi.org/10.1098/rstb.2013.0213>.
4. Bressan, P., Mingolla, E., Spillmann, L., and Watanabe, T. (1997). Neon color spreading: a review. *Perception*, 26(11), 1353-1366. Doi:10.1068/p261353.
5. Cabon, Y., Murray, N. and Humanberger, M. (2020). Virtual KITTI 2. arXiv:2001.10773v1 [cs.CV] 29 Jan. 2020. <https://europe.naverlabs.com/research/computer-vision/proxy-virtual-worlds-vkitti-2/>
6. Chen, T., Cheng, X. and Tsao, T. (2022). Border Ownership, Category Selectivity and Beyond. In: Bebis G., et al. *Advances in Visual Computing. ISVC 2022, LNCS* vol 13599, pp 1-12, 2022. https://doi.org/10.1007/978-3-031-20716-7_3, published by Springer Nature Switzerland AG 2022.
7. Coren, S. (1972). Subjective contours and apparent depth. *Psychological Review*, 79(4), 359-367 <https://doi.org/10.1037/h0032940>.
8. Craft, E., Schütze, H., Niebur, E., and von der Heydt, R. (2007). A Neural Model of Figure-Ground Organization. *J Neurophysiol* 97: 4310-4326, 2007, doi:10.1152/jn.00203.2007.
9. Cumming, B.G. and Parker, A.J. (1999). Binocular Neurons in V1 of Awake Monkeys Are Selective for Absolute, Not Relative, Disparity. *The Journal of Neuroscience*, July 1, 1999, 19(13):5602-5618
10. De Haas, B. and Schwarzkopf, D.S. (2018). Spatially selective responses to Kanizsa and occlusion stimuli in human visual cortex. *Science Report* 8, 611(2018). <https://doi.org/10.1038/s41598-017-19121-z>.
11. Dresch-Langley, B. (2009). Ehrenstein illusion. *Scholarpedia*, 4(10):5364. Doi: 10.4249/scholarpedia.5364.
12. Dresch-Langley, B. and Grossberg, S. (2016). Neural Computation of Surface Border Ownership and Relative Surface Depth from Ambiguous Contrast Inputs. *Front. Psychol.* 7:1102. doi: 10.3389/fpsyg.2016.01102.
13. Elgen, D., Puhersch, C., and Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. *NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2* Dec. 2014 Pages 2366-2374.
14. Elharrouss, O., Al-Maadeed, S., Subramanian, N., Ottakath, N., Almaadeed, N., and Himeur, Y. (2021). Panoptic Segmentation: A Review. ArXiv:2111.10250 [cs.CV] <https://doi.org/10.48550/arXiv.2111.10250>.

15. Glennerster, A. and Read, J.C.A. (2018). A single coordinate framework for optic flow and binocular disparity. *arXiv:1808.03875* [q-bio.NC] <https://doi.org/10.48550/arXiv.1808.03875>.
16. Grossberg, S.(2015). Cortical dynamics of figure-ground separation in response to 2D pictures and 3D scenes: How V2 combines border ownership, stereoscopic cues, and Gestalt grouping rules. *Frontiers in Psychology*, 6, 2054. <https://doi.org/10.3389/fpsyg.2015.02054>
17. Gu, W., Bai, S. and Kong, L. (2022). A review on 2D instance segmentation based on deep neural networks. *Image and Vision Computing*, Vol. 120, 2022, 104401, ISSN 0262-8856, <https://doi.org/10.1016/j.imavis.2022.104401>.
18. Hafiz, AM., Bhat and GM. (2020). A Survey on Instance Segmentation: State of the art. *arXiv:2007.00047* [cs.CV], [doi:10.48550/arXiv.2007.00047](https://doi.org/10.48550/arXiv.2007.00047).
19. Heitger, F, Rosenthaler, L, Von Der Heydt, R, Peterhans, E, and Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells, *Vision Research*, Volume 32, Issue 5, 1992, Pages 963-981, ISSN 0042-6989, [https://doi.org/10.1016/0042-6989\(92\)90039-L](https://doi.org/10.1016/0042-6989(92)90039-L).
20. Hubel, D.H. and Wiesel, T.N. (1959). Receptive fields of single neurons in the cat's striate cortex. *J Physiol.* (1959) Oct;148(3):574-91. *Doi:10.1113/jphysiol.1959.sp006308*. PMID: 14403679; PMCID: PMC1363130.
21. Jeurissen, D., Self, M. W., & Roelfsema, P. R. (2013). Surface reconstruction, figure-ground modulation, and border-ownership. *Journal of Cognitive Neuroscience*, 25(7),1039–1049. https://doi.org/10.1162/jocn_a_00378
22. Julesz, B. (1971). *Foundations of Cyclopean Perception*, University of Chicago Press.
23. Kanizsa, G. (1976). Subjective Contours. *Scientific American*, 234 No. 4 (1976):48-53. <http://www.jstor.org/stable/24950327>.
24. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C. Gustafson, L., Xiao, T., Whitehead, S., Berg, AC., Lo, WY., Dollár, P., and Girshick, R. (2023). Segment Anything. *arXiv:2304.02643v1* [cs.CV] <https://doi.org/10.48550/arXiv.2304.02643>. its project web: <https://segment-anything.com>
25. Kirillov, A., He, K., Girshick, R., Rother, C., and Dollár, P. (2018). Panoptic Segmentation. *arXiv:1801.00868v3* [cs.CV] Apr. 10, 2019; <https://www.cityscapes-dataset.com/>
26. Lamme, V.A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci.* 1995 Feb;15(2):1605-15. *Doi: 10.1523/JNEUROSCI.15-02-01605.1995*. PMID: 7869121; PMCID: PMC6577835.
27. Lamme, V.A, Rodriguez-Rodriguez, V., and Spekreijse H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb Cortex.* 1999 Jun;9(4):406-13. *Doi: 10.1093/cercor/9.4.406*. PMID: 10426419.
28. Marino, I.-G. (2007). Stereogram, <https://www.mathworks.com/matlabcentral/fileexchange/14859-stereogram>
29. O'Herron, P. and von der Heydt, R. (2009). Short-Term Memory for Figure-Ground Organization in the *Visual Cortex. Neuron.* 2009 March 12; 61(5): 801-809. *Doi:10.1016/j.neuron.2009.01.014*.
30. Palmisano, S., Gillam, B., Govan, DG, Allison, RS, Harris, JM. (2010). Stereoscopic perception of real depths at large distance. *J vis.* 2010 Jun 1;10(6):19. *doi:10.1167/10.6.19*. PMID: 20884568.
31. Parker, A.J. (2007). Binocular depth perception and the cerebral cortex. *Nat Rev Neurosci.* 2007 May;8(5):379-91. *Doi: 10.1038/nrn2131*. PMID: 17453018.
32. Poggio, G.E. (1995). Mechanisms of Stereopsis in Monkey Visual Cortex. *Cerebral Cortex*, 5(3), 193-204. *Doi:10.1093/cercor/5.3.193*.
33. Prince, S.J., Cumming, BG., Parker, AJ. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol.* 2002 Jan;87(1):209-21. *Doi:10.1152/jn.00466.2000*. PMID: 11784743.
34. Qiu, F.T., Sugihara, T., and von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nat. Neurosci.* 2007 November; 10(11):1491-1499 *doi:10.1038/nn1989*.
35. Qiu, F.T. and von der Heydt, R. (2005). Figure and ground in the Visual Cortex: V2 Combines Stereoscopic Cues with Gestalt Rules. *Neuron*, Vol. 47, 155-166, July 7, 2005, *doi:10.1016/j.neuron.2005.05.028*.
36. Roelfsema, P.R. (2006). Cortical Algorithms for Perceptual Grouping. *Annu. Rev. Neurosci.* 2006. 29:203-27 *doi:10.1146/annurev.neuro.29.051605.112939*.
37. Roelfsema, P.R. and Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Atten Percept Psychophys.* 2011, 73(8):2542-2572 *doi:10.3758/s13414-011-0200-0*.
38. Rubin, E. (1915) Synsoplevede figurer: studier i psykologisk analyse. 1 del. Gyldendalske Boghandel, Nordisk Forlag; 1915.
39. Sharma, R., Saqib, M., Lin, C.T., and Blumenstein, M. (2022). A Survey on Object Instance Segmentation. *SN Computer Science* (2022) 3:499 <https://doi.org/10.1007/s42979-022-01407-3>
40. Silson, E.H., Groen, I.I.A., and Baker, C.I. (2022). Direct comparison of contralateral bias and face/scene selectivity in human occipitotemporal cortex. *Brain Struct Funct.* 2022 May;227(4):1405-1421. *Doi:10.1007/s00429-021-02411-8*.
41. Uyar, F., Shostein, S., Greenberg, A.S. and Behrmann, M. (2016). Retinotopic information interacts with category selectivity in human ventral cortex. *Neuropsychologia.* 2016 Nov;92:90-106. *Doi:10.1016/j.neuropsychologia.2016.05.022*.
42. Von der Heydt, R. (2015). Figure-ground organization and the emergence of proto-objects in the visual cortex. *Front. Psychol.* 6:1695. *doi:10.3389/fpsyg.2015.01695*.
43. Von der Heydt, R. (2023). Visual cortical processing – From image to object representation. *Frontiers in Computer Science* 5.1136987. *doi:10.3389/fcomp.2023.1136987*.
44. Von der Heydt, R., Peterhans, E. and Baumgartner, G. (1984). Illusory Contours and Cortical Neuron Responses. *Science*, 224(4654), 1984, 1260-1262. *Doi:10.1126/science.6539501*.
45. Von der Heydt, R. and Zhang, N.R. (2018). Figure and ground: how the visual cortex integrates local cues for

- global organization. *J Neurophysiol* 120: 3085-3098, 2018; published July 25, 2018; doi:10.1152/jn.00125.2018
46. Wang, P. and Yuille, A. (2016). DOC: Deep Occlusion Estimation from a Single Image. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in *Computer Science*(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_33.
 47. Wardle, G. Susan, and Ritchie, Brendan (2014). Can Object Category-Selectivity in the Ventral Visual Pathway B Explained by Sensitivity to Low-Level Image Properties? *Journal of Neuroscience* 5 November 2014, 34(45)14817-14819; doi:10.1523/JNEUROSCI.3566-14.2014.
 48. Wikipedia contributors, Illusory contours, in *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Illusory_contours.
 49. Wikipedia contributors, Neon color spreading, in *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Neon_color_spreading
 50. Williford, J.R. and von der Heydt, R. (2013). Border-ownership coding. *Scholarpedia J.*, 2013;8(10):30040, doi:10.4249/scholarpedia.30040, PMID: 25075249; PMCID: PMC4112134.
 51. Wrenninge, Magnus, Unger, Jonas. (2018). Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing. arXiv:1810.08705 [cs.CV] Oct.19,2018.
 52. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., and Zhao, H.S. (2024). Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. <https://doi.org/10.48550/arXiv.2401.10891>, the project web: <https://depth-anything.github.io/>
 53. Zhou, H., Friedman, H.S. and von der Heydt, R. (2000). Coding of Border Ownership in Monkey Visual Cortex. *Journal of Neuroscience* 2000 September 1; 20(17): 6594-6611.doi:10.1113/jphysiol.2002.033555.
 54. Zhu, S.D., Zhang, L.A. and von der Heydt, R. (2020). Searching for object pointers in the visual cortex. *JNP Journal of Neurophysiology*, May 11, 2020, doi:10.1152/jn.00112.2020.
 55. Chen, T. (2025). Border Ownership and Category Annotation Extended from Virtual KITTI 2 [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15375453>