



Comparison of two multiscale spatial filtering models

by

Sebastian Keil

Matriculation Number: 357301
Course of Study: Technische Informatik

supervised by
Dr. Joris Vincent

A Bachelor's thesis submitted to

Technische Universität Berlin
Faculty IV - Electrical Engineering and Computer Science
Institute of Computer Engineering and Microelectronics
Computational Psychology

Examiners:
Prof. Dr. Marianne Maertens
Prof. Dr. Guillermo Gallego

Date: January 14, 2025

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit eigenständig ohne Hilfe Dritter und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Alle Stellen die den benutzten Quellen und Hilfsmitteln unverändert oder sinngemäß entnommen sind, habe ich als solche kenntlich gemacht.

Sofern generative KI-Tools verwendet wurden, habe ich Produktnamen, Hersteller, die jeweils verwendete Softwareversion und die jeweiligen Einsatzzwecke (z.B. sprachliche Überprüfung und Verbesserung der Texte, systematische Recherche) benannt. Ich verantworte die Auswahl, die Übernahme und sämtliche Ergebnisse des von mir verwendeten KI-generierten Outputs vollumfänglich selbst.

Die Satzung zur Sicherung guter wissenschaftlicher Praxis an der TU Berlin vom 15. Februar 2023. https://www.static.tu.berlin/fileadmin/www/10002457/K3-AMBI/Amtsblatt_2023/Amtliches_Mitteilungsblatt_Nr._16_vom_30.05.2023.pdf habe ich zur Kenntnis genommen.

Ich erkläre weiterhin, dass ich die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt habe. Berlin, den



Berlin, den Januar 14, 2025

Sebastian Keil

Acknowledgments

I would like to express my highest appreciation to Dr. Joris Vincent for the exceptional guidance, constructive feedback, and inspiring discussions throughout this project. His support has been essential for my understanding of the subject and the scientific work.

Abstract

This thesis presents a comparative analysis of two multiscale spatial filtering models: the Oriented Difference of Gaussian (ODOG) model and the Brightness Induction Wavelet Model (BIWaM). These models have been successful in predicting human vision for several brightness phenomena and therefore emphasizing the role of low-level vision in brightness perception. Despite their different implementations, this thesis investigates whether these models are fundamentally distinct or represent various implementations of the same conceptual framework.

The research methodology involved a systematic comparison of the models responses and an analysis of their behavior across three processing steps: decomposition, weighting and normalization.

Alignments of their implementations reduced the differences between the model outputs, suggesting that the models share fundamental concepts and that their distinct capabilities are due to improved implementation.

Zusammenfassung

In dieser Arbeit wird eine vergleichende Analyse zweier Multiscale Spatial Filtering Modelle vorgestellt: das ODOG-Modell (Oriented Difference of Gaussian) und das BIWaM-Modell (Brightness Induction Wavelet Model). Beide Modelle haben sich bei der Simulation des menschlichen Sehens für verschiedene Illusionen bewährt und betonen die Bedeutung der frühen Stadien des menschlichen Sehprozesses. Diese Arbeit untersucht, ob sich diese Modelle grundlegend voneinander unterscheiden oder "nur" unterschiedliche Implementierungen desselben Konzepts darstellen.

Die Methodik umfasste einen systematischen Vergleich der Architekturen beider Modelle und eine Analyse ihres Verhaltens für die drei Verarbeitungsschritte: Decomposition, Weighting und Normalization.

Durch die Anpassung der Implementierungen beider Modelle wurden die Unterschiede im Verhalten verringert und zum Teil umgekehrt, was darauf hindeutet, dass die Modelle die selben grundlegenden Konzepte nutzen und ihre unterschiedlichen Fähigkeiten auf eine verbesserte Implementierung dieser zurückzuführen sind.

Summary

1	Introduction	1
1.1	Light and the Visual System	1
1.2	Low-Level Vision	3
1.3	Modeling Human Vision	5
1.4	Oriented Multiscale Spatial Filtering Models	7
1.5	Towards a Unified Understanding of Brightness Perception Models	8
2	Methodology	9
2.1	Preparations and Tools	10
2.2	Stimuli	10
2.3	Model responses	12
3	Robustness	16
4	Decomposition parameters	19
5	Aligned models	22
6	CSF weighting	25
7	Normalization	29
8	Discussion	31
9	Conclusion	33

1 Introduction

1.1 Light and the Visual System

In the environment, light is emitted by a source of illumination such as the sun. Any surface on which this light falls reflects a portion of it as *luminance*. This portion is known as the surface *reflectance*. Therefore, luminance is the result of *illuminance* and reflectance, as shown in Figure 1 a). A lightmeter can measure the amount of luminance reflected by a surface, however, it cannot determine the reflectance of this surface, because the luminance can be the product of any combination of illuminance and reflectance. The formula ($L = I \cdot R$) in Figure 1 a) shows the problem from a mathematical perspective. L represents luminance, I and R represent illuminance and reflectance, respectively. It is simply impossible to solve for I and R , when only L is measured, because for every R there is an I to produce the measured L (Adelson, 2000).

The human visual system can solve this problem, as shown in Figure 1 b). Luminance is detected by the eye and transmitted to other parts of the visual system (represented by P). These parts generate a perception of reflectance, which is referred to as lightness. In this process, physical values, such as reflectance, transition into perceptual values, such as lightness. Additionally, the visual system generates a perceptual luminance value, which is referred to as brightness.

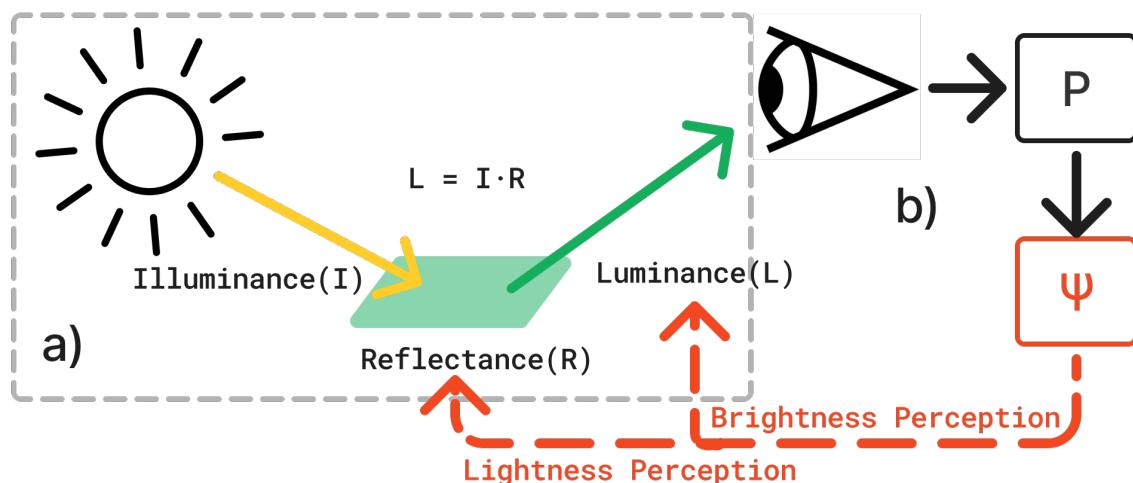


Figure 1: a) Relationship between illuminance, reflectance and luminance. Luminance is the result of illuminance and reflectance. b) The human visual system processes luminance through an unidentified mechanism, represented by P which is determining lightness and brightness perception ψ of the surface.

Unlike reflectance, luminance values are directly available as measurements in the eye, and brightness can in principle be derived from them. However, the visual system not only considers the corresponding luminances when evaluating the lightness and brightness of image areas, it also considers the luminances of the surrounding regions (Kingdom, 1997). As a result, perception can differ from the actual measurements. The mechanisms through which the visual system accomplishes these tasks are the subject of current research and are discussed further in the following sections.



Figure 2: Lightness and brightness are distinguishable when information about illumination is available. *"The walls of the house appear uniformly white – a lightness judgment – yet are brighter in some places than others – a brightness judgment"*. Quote and picture from Kingdom (2014).

The distinction between lightness and brightness becomes apparent when information about illumination is visible, as shown in Figure 2. *"The walls of the house appear uniformly white – a lightness judgment – yet are brighter in some places than others – a brightness judgment"* (Kingdom, 2014). This distinction is important because brightness is related to the relationship between an object and its environment, it reveals how the object is exposed to illumination. On the other hand, lightness represents the intrinsic properties of an object, such as color, regardless of the environment.

Because reflectance is only implicitly perceivable, it can lead to uncertain situations and illusions. For instance, a shadow can dim an area such that a white surface within the shadow reflects the same amount of light as a black surface under full illumination. Although both the surface measurements are the same, human observers can usually distinguish between white and black surfaces (Arend, 1993).

This phenomenon is illustrated by Edelson's *checkerboard shadow illusion* in Figure 3. The two patches A and B on the checkerboard in Figure 3a) appear to have different colors, even though they emit the same light, as shown in Figure 3b).

The cylinder seems to cast a shadow, even though there is no real light source, because the image is only a two-dimensional representation of the scene. However, the visual system is designed to process images from three-dimensional scenes with illumination and shadows. Thus, it processes the checkerboard shadow illusion with all available information about depth and illumination. To logically follow the process, one can say, that patch B reflects the same amount of light as patch A (shown in Figure 3b), but is located in the shadow of the cylinder and therefore must have a higher reflectance. The visual system needs to react to differences in illumination and compensate for them to estimate reflectance. This behavior ensures that the perception of a scene is closely related to the reflectance of its surface and is largely unaffected by illumination.

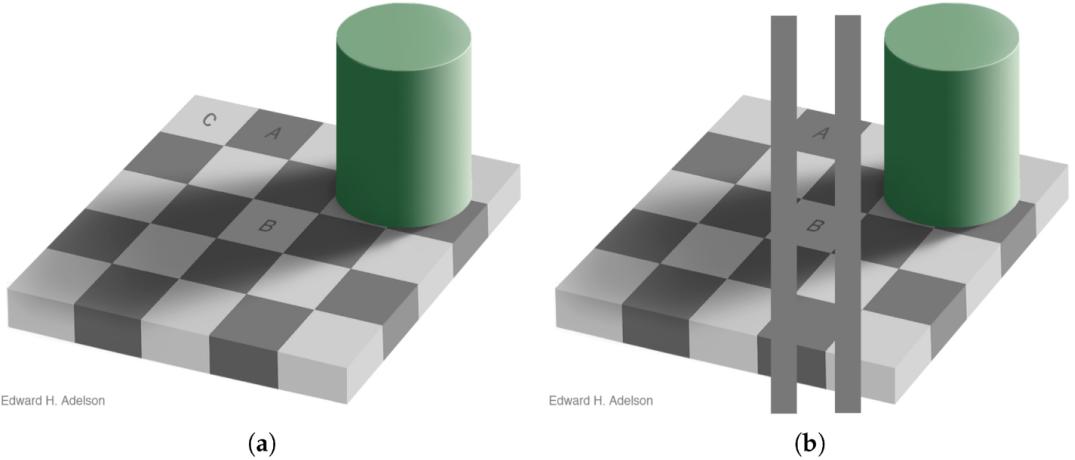


Figure 3: (a) The Checkerboard shadow illusion image; (b) proof image (Adelson, 1995).

In everyday life, illusions in brightness and lightness perception are rare, likely because of the vast amount of information that the visual system can make use of. Similar to the checkerboard shadow illusion, shading, shadowing, and spatial depth can guide the perception of brightness and lightness. However, there are illusions with less information available to the visual system, which require a different explanation.

1.2 Low-Level Vision

One simple illusion is the classic *simultaneous brightness contrast* (SBC) illusion, as shown in the upper part of Figure 4. The two inner gray squares are physically identical, but appear to differ in brightness, depending on their surroundings. The lack of information about illumination and spatial depth is crucial in comparison with the checkerboard shadow illusion. Parts of the visual system, that process illumination or spatial depth, will have no information available in the SBC illusion. Here the idea of detecting illumination and compensating for it fails, hence a different explanation is required.

Neural units in the retina of the eye offer an explanation of the SBC illusion. Hering (1834 –1918) was the first to describe *center surround fields*, which compare luminance areas with their surrounding areas at the retina. This finding provides a possible explanation for the simultaneous contrast illusion. The lower part of Figure 4 illustrates this principle. The blocks under the illusion represent neural units that respond to areas of the SBC illusion image. The surrounding units subtract and the center units add their responses to the fourth summing unit underneath. When the surrounding units sense the darker surrounding of the left patch, their response is small and so the subtraction is small. As a result the left summing unit receives a higher response, correlating with a brighter left patch experienced by a human observer. On the right patch the surrounding units sense bright surroundings, so the subtraction is higher, and the summing unit receives a lower response, which is also correlating with a human observer experiencing the right patch to be darker.

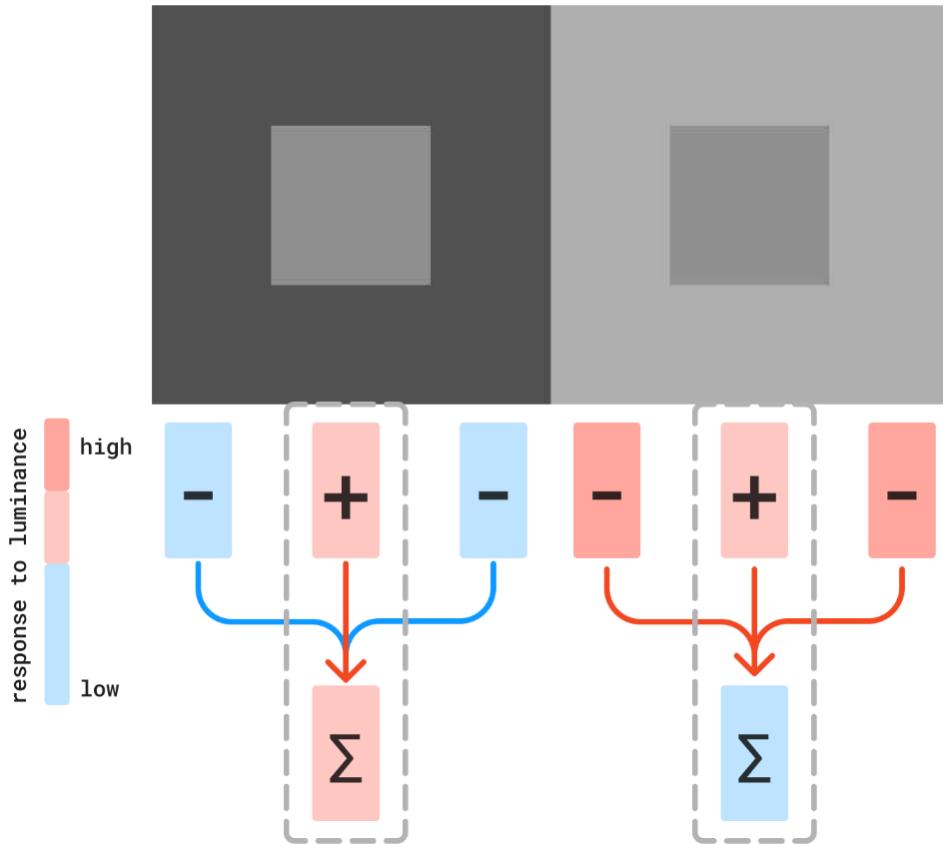


Figure 4: Above: The simultaneous contrast effect. Two identical gray patches appear to be different in brightness, depending on their surrounding. The left patch appears brighter than the right patch.

Below: The principle of center surround fields. The blocks on each side represent neural units, where the surrounding units subtract and the center units add their responses to the fourth units underneath. On the left side the center response is medium (light red), corresponding to the left patch in the illusion, while the surround response is low, corresponding to the surrounding dark gray in the illusion. The summation results in a medium response. On the right side the center is the same, but the surround response is higher (dark red) and so the summation in the fourth block is lower. Therefore the responses of the patches in the illusion depend on their surrounding.

Simple mechanisms, such like the center surround fields could be responsible for a significant aspect of human brightness perception¹. They exist in different sizes and orientations. This complex neural processing results in so called *sensory channels*, where each channel is selectively sensitive to different sizes of contrast areas, also referred to as the spatial frequencies of the image (Sachs et al., 1971). Large center surround fields respond to low spatial frequency information, such as large objects and gradual changes across the image. Small center surround fields respond to high spatial frequency information, such as fine details and edges.

The resulting sensory channels transport specific image features and the interaction between these channels provides a more meaningful way to process image information. This has inspired researchers to investigate the computational modeling of center-surround fields and their interactions in computer vision and vision research.

¹The terms brightness and lightness become synonymous without information about illumination and will be used interchangeably in the following sections, as we will discuss only such illusions.

1.3 Modeling Human Vision

The basic idea of the center surround fields is to compare luminances with their surrounding luminances. Given that computers process image data as a series of discrete pixel values, this comparison can be reliably modeled. A common approach is to design a convolution filter (representing a center surround field) and convolve it with image pixel values. Figure 5 shows the principle of a convolution on a grayscale image. The filter values are iteratively applied to the input pixels by element-wise multiplication and then summed to generate the output pixel. In the example in Figure 5, the filter enhances the image contrast, as the center filter value is positive and the surrounding values are negative. With this the filter response is high in an image region with contrast, where the center has a high value (bright) and its surroundings have low values (dark). On the other hand for a low-contrast region such as a homogeneous area, the positive center and the negative surrounds of the filter cancel each other when multiplied with the same pixel values and then summed up.

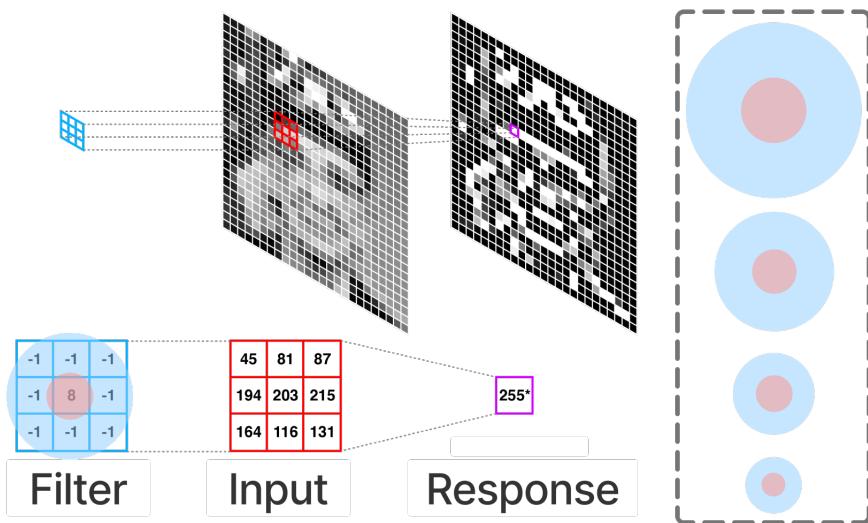


Figure 5: Applying a convolution on an image is a common approach to model center surround fields. Every pixel of the input image is multiplied with the center value of the filter (8) and then summed up with every surround pixel multiplied with the corresponding value in the filter (-1). The resulting sum is the new pixel value for the response image at the position of the input center pixel. The left Figure is inspired by Gundersen (2017). The right side shows a filterbank, existing of multiple filters in different sizes.

To utilize the concept of different-sized center surround fields, it is sufficient to create multiple filters of different scales. A filterbank is shown on the right side of Figure 5. Each of the different scaled filters is applied to the input image and generates its own response, similar to the sensory channels of the visual system. Large-scale filters extract low spatial frequencies, whereas small-scale filters extract high spatial frequencies. All the responses from the different-scaled filters represent the image information decomposed at specific frequencies. To reconstruct the image, a simple addition of the filter responses is sufficient (some normalization is needed). A greater number of different scaled filters used to decompose the image will lead to more information being extracted, resulting in a more accurate reconstructed image. These mechanisms form the foundation of multiscale spatial filtering models.

To create a more accurate model of the human visual system, researchers have incorporated weighting of the individual filter responses before reconstruction. Weighting corresponds to the low-frequency fall-off of the contrast sensitivity function (CSF) found in the human visual system (W. & G., 1968). The CSF shows the sensitivity of the human visual system to contrast across different scales. Specifically, humans are more sensitive to small-scale contrasts than to large-scale contrasts. Therefore the responses of the large-scale filters (low frequencies) are attenuated and the responses of small-scale filters (high frequencies) are enhanced. Consequently, the reconstructed image is no longer identical to the input image. In fact for the SBC illusion shown in Figure 4 the left patch is predicted to be brighter and the right patch darker, aligning with the human perception of the illusion.

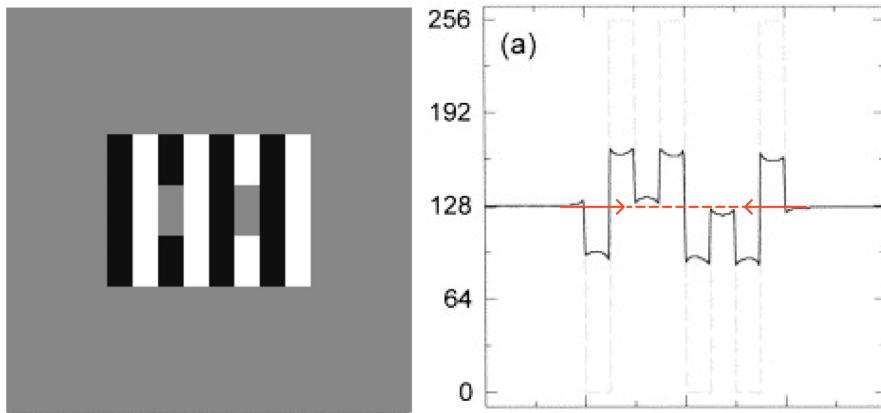


Figure 6: In White's Effect (left) the shift in perceived brightness is in the opposite direction compared to the SBC illusion. Both patches are identical, but the left patch on the black bar appears to be brighter than the patch on the white bar, even if it shares most of its edges with white surfaces, and vice versa for the right patch (White, 1979). The diagram on the right shows the processed White's Effect illusion by the ODOG model. The dashed gray line refers to the luminance profile across the horizontal center of the illusion. The solid line represents the models output along the same line. The red markers indicate that the models output is in accord with the human perception (Blakeslee & McCourt, 1999): The left patch is predicted brighter than the right patch.

Blakeslee and McCourt developed a multiscale spatial filtering model called DOG model (Difference Of Gaussian), which implements the decomposition of an image using different scaled filters, weighting of the filter responses and reconstruction to generate the output image. The filters used to decompose the image are zero-sum difference of gaussian filters in seven logarithmic spaced scales (similar to the filters in Figure 5). Most simply the filters and the weighting account for the capabilities of the model in predicting human vision for several illusions (Blakeslee & McCourt, 1997). However, the DOG model cannot account for all brightness phenomena. For instance the White's effect, illustrated on the left side of Figure 6, cannot be predicted using the DOG model.

1.4 Oriented Multiscale Spatial Filtering Models

Most recent models continue to draw inspiration from the same psychophysical findings in the human visual system, including center surround fields and CSF weighting. The Oriented Difference of Gaussian (ODOG) model, developed by Blakeslee and McCourt (Blakeslee & McCourt, 1999) started a new class of models, known as Oriented Multiscale Spatial Filtering Models.

The **ODOG** model differs from the DOG model in that the filters are anisotropic and in addition to seven scales, they also vary in six orientations distributed across 180° . This results in a filterbank of 42 distinct filters. Figure 7 shows the 90° anisotropic filters at seven different scales. The filter responses of different scaled filters within the same orientation are summed with each scale multiplied by a weight corresponding to a weighting function. Thus, higher-frequency filter responses receive a higher weight, resulting in six weighted orientation responses, one for each orientation. Before reconstruction the six different orientation responses are normalized by the root-mean-square values computed across all pixels. These upgrades make the ODOG model qualitatively predict White's effect (and other phenomena) as shown in the diagram in Figure 6. Specifically for the White's effect the normalization step of the ODOG has the desired effect of enhancing filters orthogonal to the black and white bars, thereby supporting the contrast between the test patches and the bars on which the patch is placed (Betz et al., 2015).



Figure 7: This is an abstract of the 42 filters used in the ODOG model. The seven horizontal (90° degrees) anisotropic DoG filters are shown. Blue parts are negative, red parts are positive. The seven scales are determined by the center gaussian zero-crossing to zero-crossing distance and reach from 5 to 320 pixels (logarithmically spaced) on a 1024x1024 pixels canvas.

Notably, this established class includes a model that employs a distinct approach: the Brightness Induction Wavelet Model (**BIWaM**) model from Otazu et al. (Otazu et al., 2008). In contrast to the ODOG model, the BIWaM model relies on a wavelet decomposition instead of a filterbank. This transformation also performs convolutions using filters of different scales and orientations. It iterates through the decomposition seven times recursively, with the image downsampled by a factor of two at each iteration. However, the scale of the filters remains constant. Therefore, it generates filter responses with varying filter-to-image scales and orientations in each iteration, similar to the filterbank in the ODOG model. The BIWaM model effectively convolves the input image with 21 filters in seven different scales, also logarithmically spaced, and three different orientations (horizontal, vertical and diagonal). The filter responses are weighted based on a more accurate CSF-inspired function, merged with the normalization step. The normalization step applies factors on every pixel of the filter responses that are specific to their relative contrast in its region compared to its surroundings. This normalization could be the main difference compared with the ODOG model (Betz et al., 2015).

1.5 Towards a Unified Understanding of Brightness Perception Models

ODOG and BIWaM are two examples of oriented spatial filtering models, that emphasize the role of low-level vision in brightness perception. As illustrated in Figure 8 their general processing flow can be represented within a common structural diagram, highlighting their conceptual similarities. Both models employ a set of spatially scaled and oriented filters to decompose the input image, with each filter extracting specific orientations and spatial frequencies from the image. Additionally both models implement a weighting on the filter responses according to the CSF of the human visual system, followed by a normalization step that enables the interaction between different filter responses. To reconstruct the output image, both models compute the sum of the weighted and normalized filter responses.

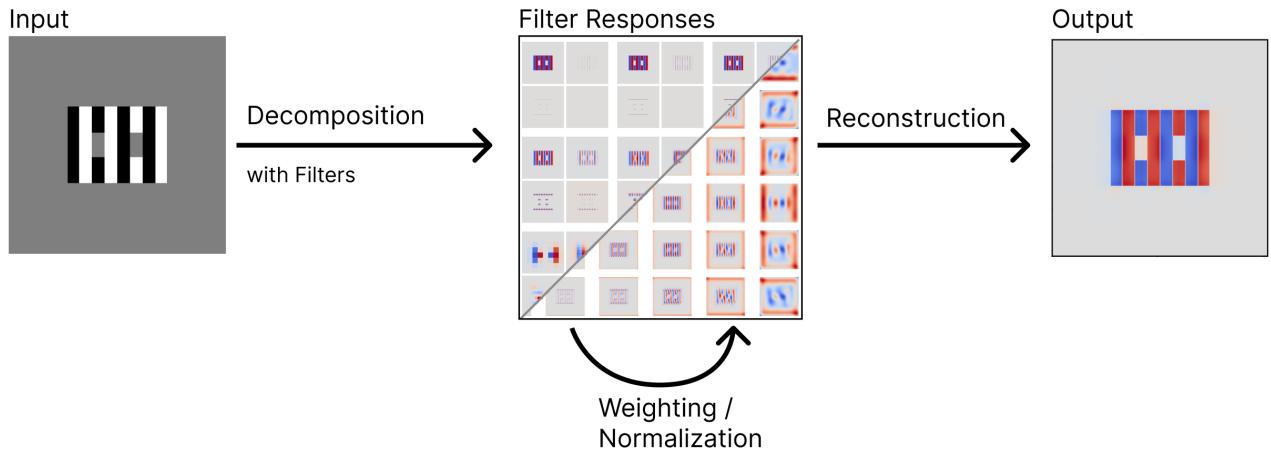


Figure 8: Structural overview of ODOG and BIWaM models. Both decompose the input image with a set of filters, then do a weighting on the filter responses and a normalization step before reconstructing the output image. Despite these similarities, the predictions and the implementations are quite distinct.

At first glance, the ODOG and BIWaM may appear practically different due to their distinct implementations. ODOG utilizes a filterbank, simple CSF weighting, and global orientation-based normalization, whereas BIWaM employs wavelet decomposition, more comprehensive CSF-weighting, and local scale-and-orientation-based normalization. Despite these differences, both models adhere to the same underlying principles: orientation- and scale-specific decomposition, CSF-weighted responses, and interactions between filter responses (normalization). However, the BIWaM model is capable of predicting human vision for illusions, which the ODOG model can not account for. This apparent contradiction leads to the central question of this thesis:

Are the ODOG and BIWaM models fundamentally distinct, or do they really represent different implementations of the same conceptual idea?

Based on my analysis of both models, I hypothesize that their differences stem primarily from improvements in implementation rather than from fundamentally distinct theoretical foundations.

2 Methodology

The aim of this thesis is to determine whether the ODOG and BIWaM models are fundamentally distinct or whether they are merely different implementations of the same conceptual idea. To achieve this, I systematically explored the conceptually corresponding parameters and mechanisms in the implementation of each model for each processing step as shown in Figure 8 (decomposition, weighting and normalization). For example in the decomposition step both models implement a conceptually corresponding parameter which defines the scales of the filters used to extract specific frequency of the input image (ODOG: 'scales', BIWaM: 'levels'). My approach involves changing the corresponding parameters in the implementation of both models, to see if the resulting changes in their outputs correspond. If the models really only are different implementations of the same idea, a change in the corresponding parameters in both models should result in similar changes in their outputs.

Given the iterative nature of this methodology, this thesis does not follow a traditional method-and-results structure. Instead, it is organized into five interdependent sections, each linked to a processing step of the models shown in Figure 9. The table presents an overview of the tests that were conducted, along with the processing steps to which they were associated.

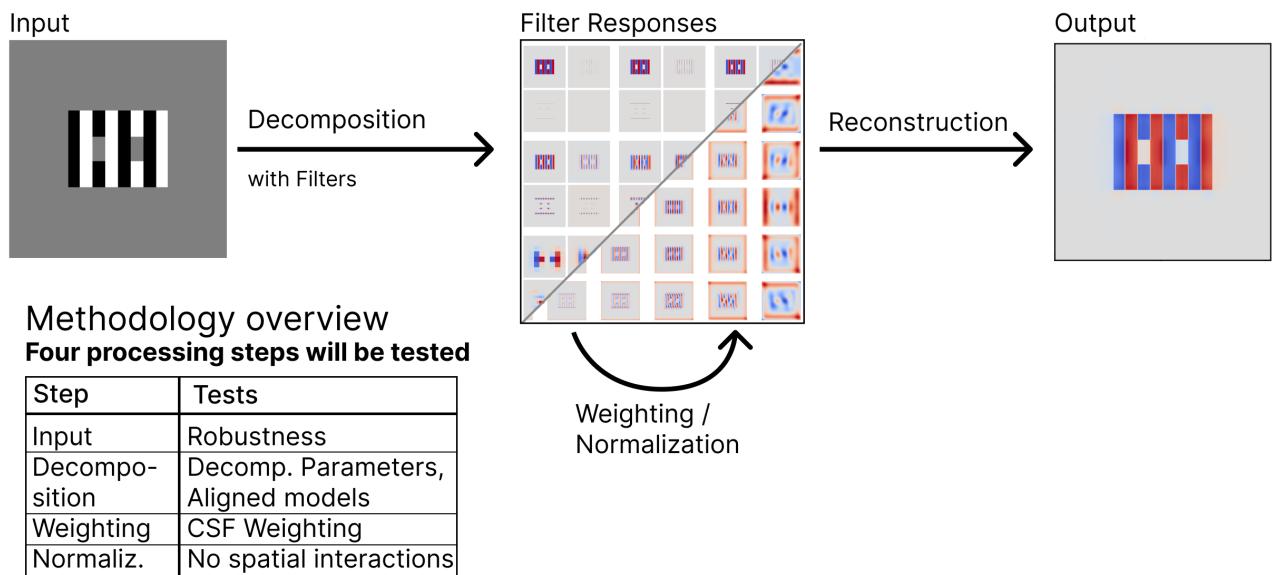


Figure 9: This figure offers an overview of the methodological approach employed in this thesis next to the models common structure. The table lists five tests, each corresponding to a specific processing step of the models. Each test provided insights that contributed to a chain of evidence, which was established to answer the research question.

At the beginning of each section, an overview similar to that shown in Figure 9 provides a roadmap, ensuring a clear narrative. The results from all sections will be linked in the discussion chapter, where I will comprehensively compare the models and draw conclusions about their behavioral and structural relations.

2.1 Preparations and Tools

For the implementation and analysis in this thesis, Python 3.12.7 was used alongside Jupyter Client 8.6.2 as the development environment. Data visualization was performed with Matplotlib 3.9.1, while numerical computations and scientific analysis were done with NumPy 2.1.3 and SciPy 1.14.0. The stimuli used in the experiments were generated using the Stimumpy package (version 1.1.2). For modeling, the ODOG model was implemented using the Multyscale library (version 0.2), whereas the BIWaM model was based on a Python reimplementation of the available MATLAB code. In certain test scenarios, modifications were made to the source code of both models to facilitate the analysis; however, these changes did not affect the fundamental behavior of the models.

2.2 Stimuli

In the following sections both models were tested on a set of seven different stimuli with 1024x1024 pixels. The variety in stimuli lowered the risk of drawing conclusions based on the specific features of a single stimulus and provided a reliable baseline to identify changes from parameter modifications. The stimuli used were: two versions of the classic White's effect (referred to as White's thick and White's narrow), two versions of Simultaneous Brightness Contrast (SBC) (referred to as SBC large and SBC small), one Checkerboard illusion, one Circular White's effect and the Dungeon illusion. The Dungeon illusion was specifically chosen, because the authors of the BIWaM model mentioned, that their model would account for it, while it was previously unexplained (Otazu et al., 2008). A List of all stimuli is shown in Figure 10.

Each stimulus consisted of a black-and-white configuration (black=0 and white=255) where two targets, physically identical in gray value (gray=128), were embedded. The set of stimuli covers a wide range of image features with respect to scales and orientations. For example different sizes of White's effect and SBC illusion, which integrate higher or lower frequencies. In the Circular White's effect, the Checkerboard and the Dungeon illusion, the distribution of different orientations are equal, compared to the horizontal grating of White's effects.

The authors tested the models using several stimuli. The ODOG model was tested using two similar versions of White's effect and two similar versions of the SBC illusion as in this thesis. The same applies to the BIWaM, but it has been tested on the Checkerboard and Dungeon illusions as well. Both models have been tested on grating illusions by their authors, which in this thesis will not be covered, but could be interesting on follow-up.

The original stimuli were not provided by the authors and experiments in which both models were tested on the exact same stimuli have not yet been conducted. For a fair comparison this is the first step in this thesis.

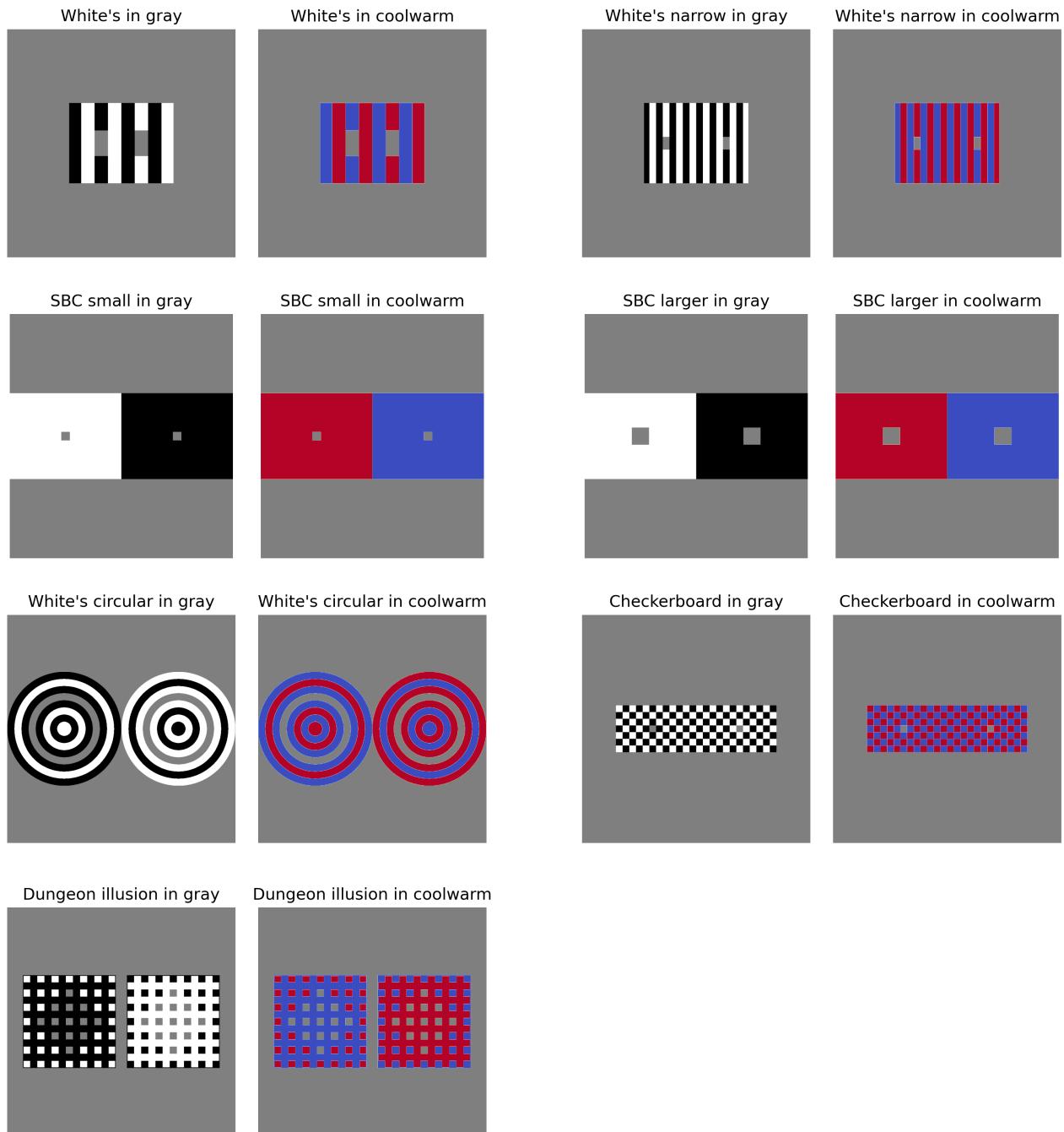


Figure 10: The seven stimuli used to test the models are shown in greyscale and in the coolwarm colormap. Each stimulus consists of a black-and-white configuration where two test patches, physically identical in their gray value, are embedded. In all seven stimuli only three brightness values are being used: black (0), white (255) and intermediate grey (128). The greyscale presentation generates a perceptual illusion in which one target appears brighter than the other due to its surroundings. However, the coolwarm colormap confirms the physical similarity of the targets by reducing the illusionary effect and showing the identical grey values of the test patches.

2.3 Model responses

To obtain first insights into the model behaviors, this section presents the outputs of both models for the same set of stimuli, as shown in Figure 11. This approach offers a comparison of the models with their default configurations and establishes baselines, which the outputs of the modified models can later be compared.

After some preparations the usage of both model implementations was similar. First the grayscale input image was loaded as a matrix of brightness values (0-255), then this matrix was passed as an argument next to specific parameters to the function that initiated each model's processing. The same function returns the model output image as a matrix. This output can be displayed as an image, but for better insight, a center-line plot is provided (shown in Figure 11). In a center-line plot the pixel values along the horizontal center-line of the input and output images are plotted, involving both test patches.

Note that the model output values can be larger and smaller than the input values (0-255), but remain in the same order of magnitude. Additionally, when comparing the strength of the prediction, it is necessary to consider the difference between the predicted patches. Other quantitative analyses are less meaningful because the models output different units. To make the outputs more comparable, the ODOG's output was adjusted to match the BIWaM units.

Initial model responses

Both models were tested on all seven stimuli from Figure 10.

For the White's thick and White's narrow stimuli both models qualitatively predicted human perception, as the left test patch was predicted to be slightly brighter than the right test patch (Figure 11 (a) and (b)). The BIWaM demonstrated a stronger prediction for White's thick (a), as the values of the left test patch never fell under 128 and the values of the right test patch remained under 128. In contrast, the ODOG model predicted the test patches at their edges in the wrong directions, but only for a small peak. For White's narrow (b) both model predictions were stronger. The ODOG did not predict the patches in the wrong direction and the BIWaM model predicted the patches even more strong. This can be attributed to the fact that with increasing spatial frequency the magnitude of White's effect increases (Anstis, 2005; Blakeslee & McCourt, 2004). There is a significant overshoot and undershoot in the black and white bars for both models on both stimuli (a) and (b).

For SBC small and SBC large (Figure 11 (c) and (d)) both models predicted human perception qualitatively, as the left test patch on the white surrounding was predicted to be brighter and the right test patch on the black surrounding was predicted to be darker. The ODOG model had a much stronger prediction, but was not able to follow the input values in other regions, unlike BIWaM. Both models overshot and undershot the most at the center, where the black and white surroundings meet. In addition, the ODOG's prediction is much stronger than people tend to see it, thus the smaller prediction of the BIWaM is more realistic (Robinson et al., 2007).

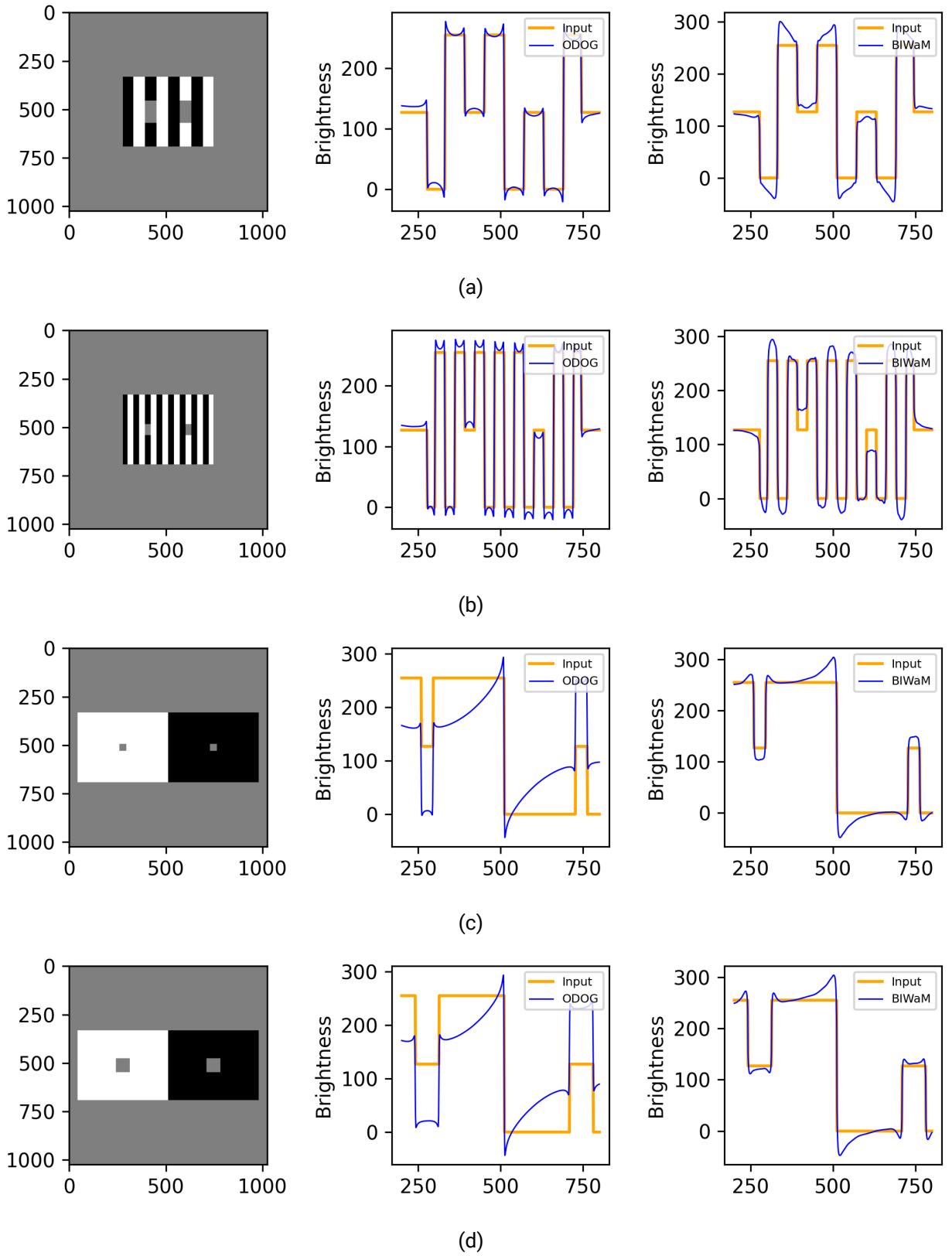


Figure 11: Initially both models were tested on the same stimuli (from top to bottom: White's thick, White's narrow, SBC small and SBC large). The left image shows the input image with both axis being pixels from 0 to 1023. Both plots on the right show a center-line plot, with the input and the model outputs as brightness values (y-axis) along the horizontal center-line (x-axis). While the orange input graph shows the physically identical patches at the same brightness (128), both models (blue graphs) predict the patches qualitatively according to humans.

The next three stimuli (White's circular, Checkerboard and Dungeon illusion) could not be predicted by the ODOG model, while the BIWaM model clearly predicted all three as shown in Figure 12 (a), (b) and (c). In each of these stimuli the left test patches should be predicted darker and the right test patches brighter. For all three (a), (b) and (c) ODOG predicted both test patches in the wrong direction. BIWaM predicted the Checkerboard (b) the strongest with an overall most uniform output, while the White's circular (a) and Dungeon illusion (c) had smaller strength and the overall output showed some irregularities at the black and white image parts. Both models showed significant overshooting and undershooting at the black and white image parts.

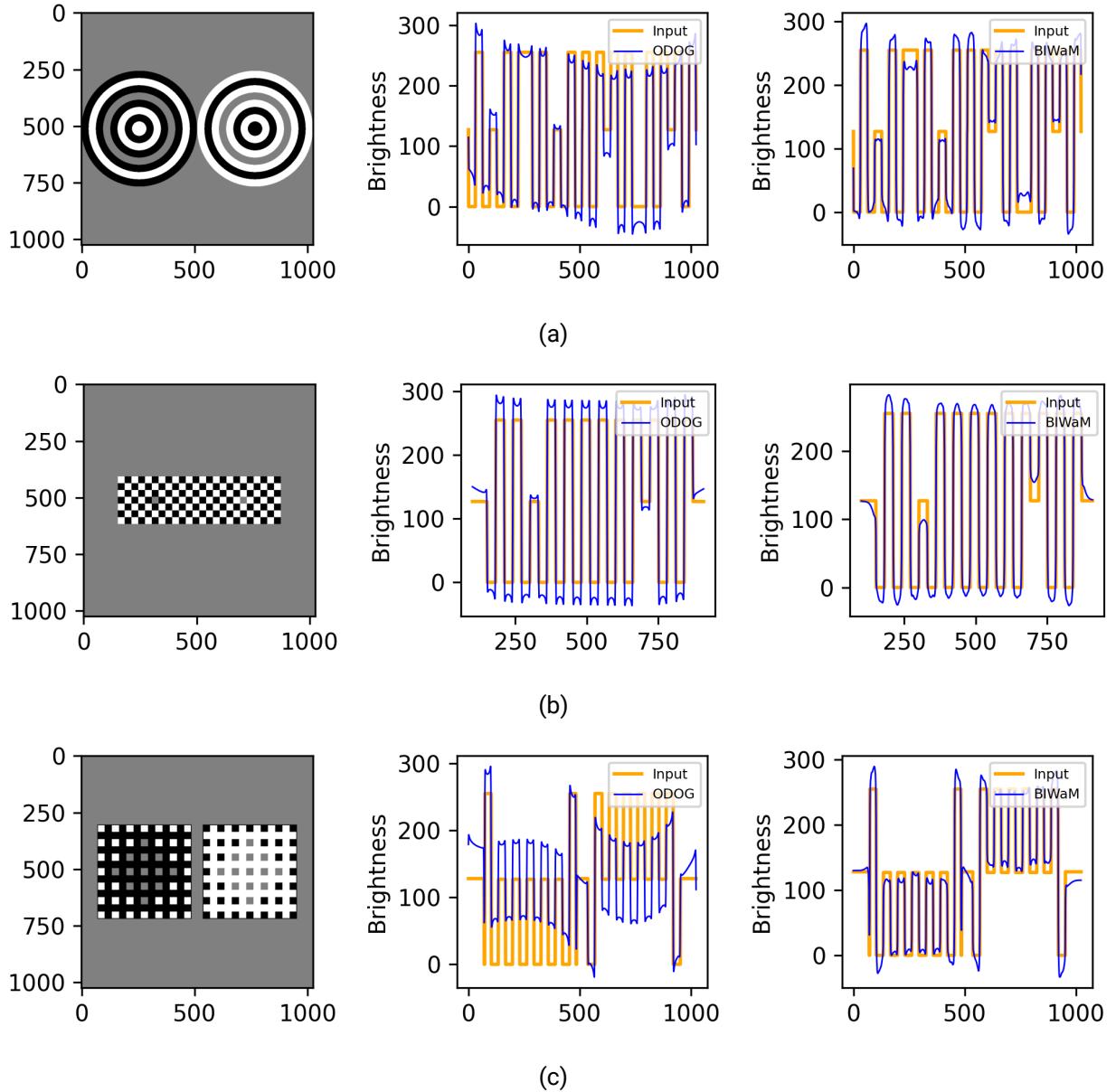


Figure 12: Initially both models were tested on the same stimuli (from top to bottom: White's circular, Checkerboard and Dungeon illusion). The left image shows the input image with both axis being pixels from 0 to 1023. Both plots on the right show a center-line plot, with the input and the model outputs as brightness values (y-axis) along the horizontal center-line (x-axis).

Stimuli details

In the interests of completeness I should say, that the three stimuli from Figure 12 share a common feature compared with the other stimuli from Figure 11: their (nearly) equally distributed orientations. Robinson showed that the ODOG is unable to predict such stimuli. "*For variants of White's effect with more equal global orientation energy, ODOG fails.*" (Robinson et al., 2007). This initial test of the models was able to reproduce these findings. The capability to predict these three stimuli, while ODOG was not able to predict them, makes BIWaM the superior model. In the upcoming sections I was able to align BIWaM and ODOG in their implementations, therefore one aspect of having these three stimuli tested with the original models is to show how the alignments affected the leading of BIWaM.

In the following sections of this thesis the models were tested mostly on White's thick and SBC large stimuli (Figure 11 (a) and (d)), as both are well-studied stimuli and each model predicts one of the stimuli stronger than the other model. Additionally with these two stimuli two main types of brightness phenomena are covered: Brightness contrast (SBC large) and brightness assimilation (White's thick). Assimilation is the process by which the perceived brightness shifts towards the background, rather than away from it as in brightness contrast. In White's effect, the flanking bars can be considered as the background of the test patches, as they have the most contact.

Common model characteristics

Despite the fact that the models were able to predict a stimuli or not, or that they differed in prediction strength, for all tested stimuli both models showed common characteristics. They exhibited some sort of smoothing, indicated by the black and white boarders in the images, which were approached gradually (most clearly seen in the SBC stimuli in Figure 11 (c) and (d)). Additionally at local minimas and maximas (boarders between black, white or grey) both models exhibited overshooting and undershooting, with two observed variants: a dent (depression) and a (tri)angular trend (e.g. seen in Figure 11 (a) and (b), respectively).

3 Robustness

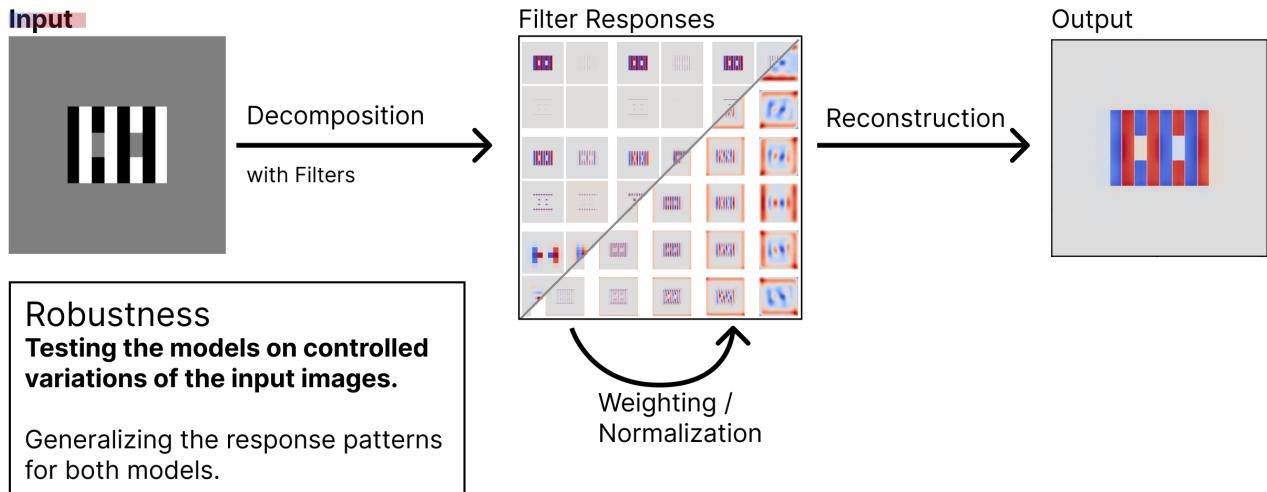


Figure 13: The Figure provides an methodological overview of the robustness section. Testing the models on controlled variations of the input stimuli is a test of whether the two models process these stimuli similarly. If the models do share the same fundamentally concept, it is reasonable to expect that variations in the input stimuli will affect the responses of the two models in similar ways, as the input influences all processing steps of the models.

The model behaviors in Figures 11 and 12 may share common characteristics due to the nature of the stimuli, i.e., it is possible that the model behaviors align solely due to the specifics of the stimuli, not because they share the same fundamentally concept. However, if the models do share the same fundamentally concept, it is reasonable to expect that variations in the input stimuli will affect the responses of the two models in similar ways. Therefore, testing the models on controlled variations of the input stimuli is a test of whether the two models process these stimuli in a similar and robust way.

The analysis of robustness tests differs from the analysis of other tests in that the prediction of human vision is less meaningful. In this case, the objective was to determine whether the models responded in a similar and predictable way. Consequently, the primary focus was on the specific reaction of both models to the input variation, not on the right prediction of the test patches.

ODOG and BIWaM were applied to a set of test stimuli that underwent controlled modifications. These modifications included the brightness of the target patches and the brightness of the background. The first test was conducted on a variation of White's thick, where both test patches were darker (75) than in the original stimulus (128). The model reactions are shown in Figure 14 (a).

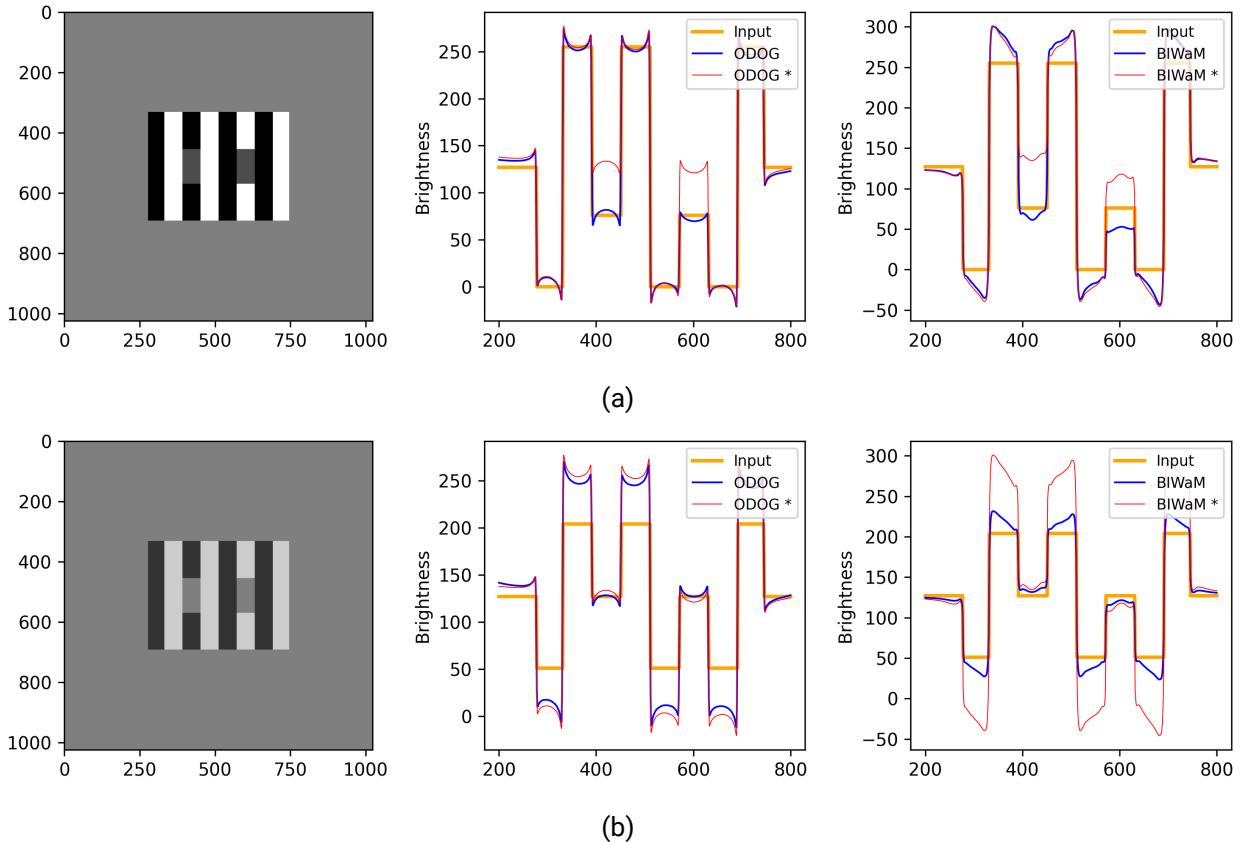


Figure 14: Testing both models on two variations of White's thick: (a) with darker test patches (75) and (b) with lower contrast in the background grating (black=50, white=200). The blue lines represent the models' outputs on the variation, and the red lines represent the outputs on the original stimulus.

The models demonstrated nearly identical reactions to the decrease in brightness of the test patches in White's thick shown in Figure 14 (a). Both followed the variation in a predictable way, as they shifted down their responses at the test patches the same amount as the decrease of brightness in the input image.

In Figure 14 (b) the test on a White's thick version with lower contrast in the grating (black=50, white=200) is shown. Here the ODOG does only slightly react to the variation at the black and white bars, whereas the BIWaM shifts with an equivalent magnitude to that of the variation in the image. Both models lost strength in their prediction at the test patches.

Additionally two variations of the SBC large stimulus were tested. The first variation had darker test patches (75), as shown in Figure 15 (a), where both models reacted identically by shifting their predictions with an similar magnitude to that of the variation of the test patches. While the BIWaM response remained constant in other image regions, the ODOG exhibited an asymmetric reaction, pulling both surroundings of the test patches up.

The second variation of SBC large had lower contrast in the surroundings of the test patches (black=50, white=200), as shown in Figure 15 (b). Here, the reaction was similar to the lower contrast version of the White's large stimulus (shown in Figure 14 (b)): ODOG model showed only a slight reaction to the variation in the black and white bars, while the

BIWaM shifts with an equivalent magnitude to that of the variation of the image. Both models demonstrated a loss of prediction strength at the test patches.

For all tests in this section both models demonstrated distinct characteristics observed in previous tests, such as the variations of overshooting and undershooting.

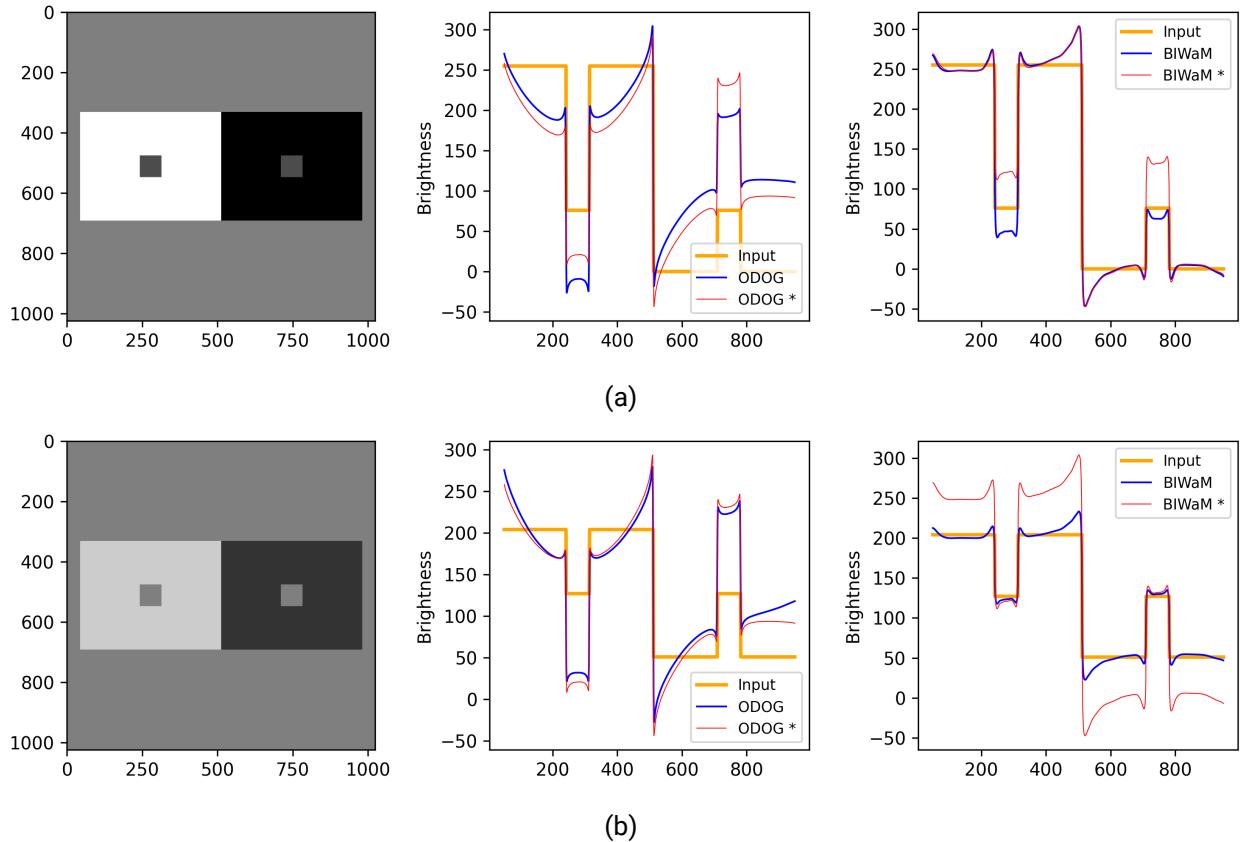


Figure 15: Testing both models on two variations of SBC large: (a) with darker test patches (75) and (b) with lower contrast in the surroundings (black=50, white=200). The blue lines represent the model's outputs on the variation, the red lines represent the outputs on the original stimulus.

On the one hand the tests with controlled variations of the input images demonstrated that the models react in predictable and reliable ways, as the changes in the input image resulted in similar changes in the model outputs. This robust behavior helps to attribute the results of upcoming tests, e.g. to model modifications.

On the other hand the robustness tests demonstrated how both models reacted similar to the variations. They shifted in the same directions at the same image regions, while demonstrating distinct characteristics. This finding suggests that both models implement processes that respond to those changes in a similar way, but also exhibit specific variations.

4 Decomposition parameters

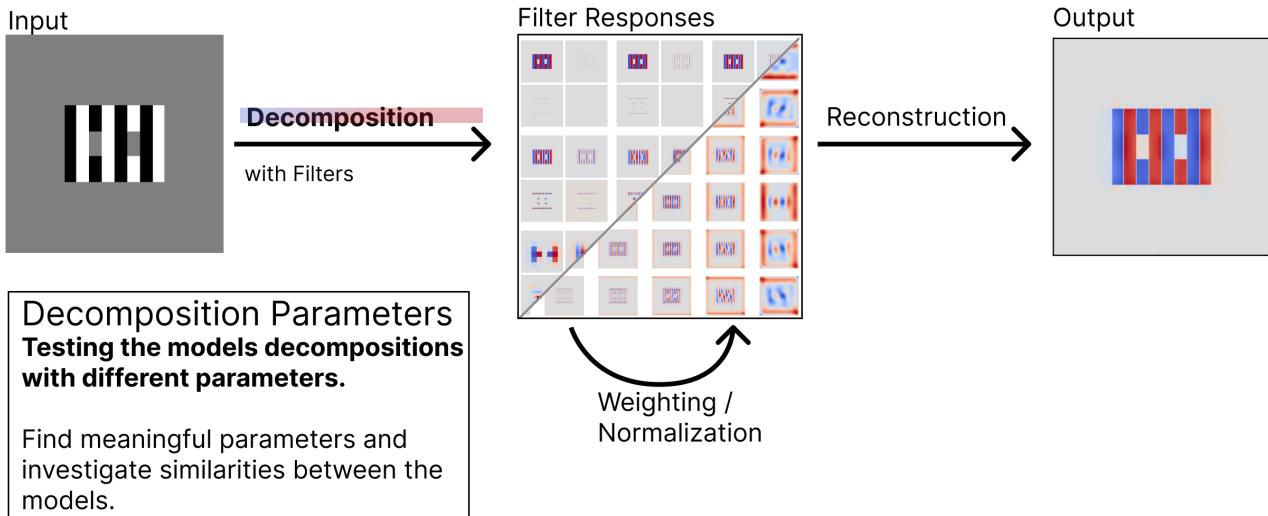


Figure 16: The Figure provides an methodological overview of the decomposition parameters section. To investigate the research question the approach involved finding conceptually corresponding parameters in the decomposition steps of both models. If changing these parameters results in corresponding changes in their outputs, this might suggest that the models fundamentally share the same idea for the decomposition step.

The objective of this section was to reveal the distinct implementations of the decomposition steps in the ODOG and BIWaM. If they really only represent different implementations, but share the same conceptual idea, I expected to find decomposition parameters in both models which conceptually correspond and when varying those parameters the changes in the model outputs also correlate.

To demonstrate the distinct implementations of both models, i first explain their different approaches for the decomposition step. In the decomposition step both models create several filters with different relative (filter-to-image) scales and orientations. Each filter extracts specific image features. The scale of a filter (relative to the image) determines which spatial frequency is being extracted, and the orientation of the filter determines which orientation will be extracted. Note that for the filter scale the relation to the image scale is important. This distinction is made because the BIWaM, unlike the ODOG, does not create filters of different scales. Instead, the BIWaM model applies the same scaled filters to extract image information, downsamples the image by a factor of two and uses the same scaled filters again. This process is repeated for each iteration in the wavelet decomposition, where the number of iterations is called levels. This effectively creates different relative filter-to-image scaled filters, extracting different spatial frequencies defined by the number of levels. Hereafter I will refer to the relative filter-to-image scales as relative scales or just scales.

The ODOG model creates 42 filters of seven logarithmically spaced scales and six orientations distributed across 180° (see Figure 7). The BIWaM model applies 21 filters of seven relative scales that are logarithmically spaced and rotated in three orientations (vertical, horizontal and diagonal). The BIWaMs diagonally-oriented filter can be simplified to a 45-degree-shifted cross shape. This shape effectively represents two oriented filters

of 45° and 135° . Therefore, the filters in the BIWaM model extract four orientations (vertical, horizontal and two diagonals).

The number and spacing of different relative filter scales are essential for the decomposition step. However, the spatial frequencies that are actually being extracted depend on the actual filter size. The ODOG and BIWaM use two different approaches to define the actual filter sizes. The ODOG model defines the maximum filter scale and creates seven smaller filters by halving every scale. BIWaM defines the smallest filter scale and quasi-doubles it for every scale to generate seven scales. It should be noted that the models actually use different filters. The ODOG uses DoG filters and BIWaM implements standard Gaussian filters. Therefore, one could say that the filter scale comparison is misleading, because they extract different image features anyway, because they are different filter types. However, to my understanding this is not the case. The implementation of the BIWaM model reveals that it uses the filter response of the parent level as the input for the next decomposition level, to which a larger filter will be applied. Then, it calculates the difference between the input and the responses of the larger filter, effectively calculating an intricate version of a DoG filter. Therefore, a comparison between the model filter scales is reasonable and can be made between the center size of the ODOG filters (zero-crossing to zero-crossing) and the overall size of the BIWaM filters. With this in mind I can confidently ensure, that both models use the same filter scales from 5 to 320 pixels relative to the input image size of 1024x1024 pixels. This aligns with the definition of Blakeslee and McCourt (Blakeslee & McCourt, 1999).

The ODOG and BIWaM implement parameters to access their filter configurations. For the ODOG model the parameters are called "scales" and "orientations", defining the scales and orientations of the filters. The BIWaM only implements a "levels" parameter, similar to the "scales" parameter. The orientations are part of the wavelet decomposition algorithm and cannot be accessed. With this I found conceptually corresponding parameters in both models. The approach was to change the parameters "scales" and "levels" and analyze their impact on the model responses. Because they are conceptually corresponding parameters I expected "scales" and "levels" to have a similar effects on the model outputs. The results are shown in Figure 17.

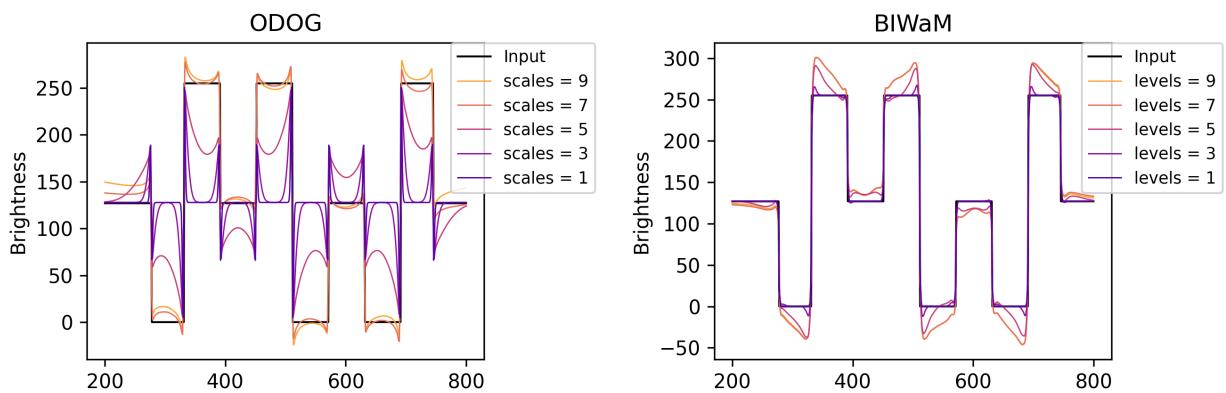


Figure 17: Both models were tested on White's thick with different "scales" and "levels" parameters. The most crucial analysis here is, that the effect of the conceptually corresponding parameters did not appear to correlate between the model outputs.

The ODOG model showed a clear behavior change on the varying "scales" parameter. Small "scales" will result in only high spatial frequencies being extract, which can be seen when following the darkest line (scales=1) in the left diagram in Figure 17: it strikes out only on edges, since they have the highest spatial frequencies. As additional scales are added to the decomposition step, larger filters are also engaged, leading to low spatial frequencies being extracted. As a result, the reconstructed image becomes more accurate, while also showing the models characteristics. Between seven and nine scales the output stagnates, which is plausible because at seven scales the biggest filter is already around one third of the input image size and there are only few such low frequencies in the image. The prediction starts to move in the right direction on approximately seven scales.

The BIWaM reacted unexpectedly, as it seems to not really differ much on different numbers of "levels", compared to the ODOG. Actually on small numbers of "levels" it pretty much replicated the input image. This behavior was not an accident; it was intentional because BIWaM implements a residual image. The residual image is calculated as the difference between the model input and its filter responses, thereby effectively representing all remaining image information that has not been extracted. Through this process the BIWaM model keeps all image information. And even if it can only make use of one small filter (levels=1 in Figure 17) to extract high spatial frequencies, it has all other frequencies available for reconstruction.

Therefore, I modified the source code and implemented a solution to make the BIWaM ignore the residual image and then tested the BIWaM again with varying "levels" to see if it correlates to the "scales" parameter in the ODOG. The results (shown in Figure 18) clearly demonstrated that the parameter "levels" has a similar effect to that discussed above for the "scales" parameter. This observation is made possible by the removal of the residual image.

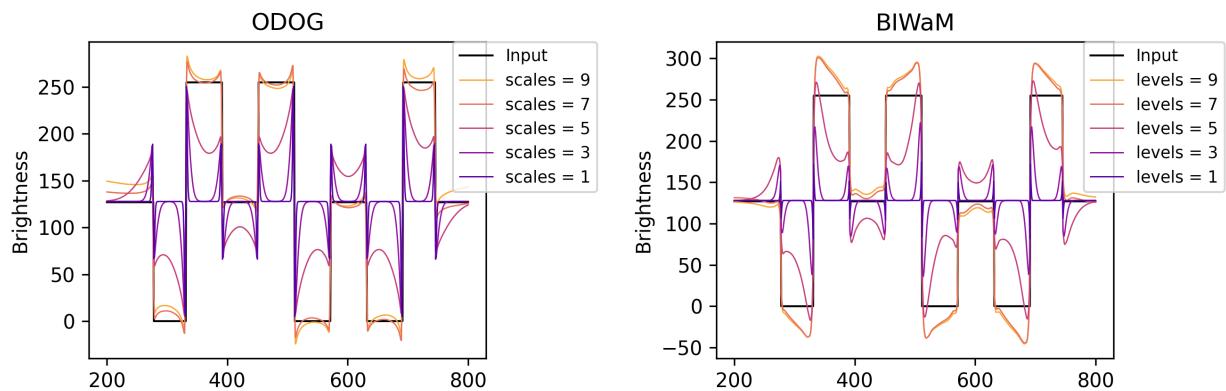


Figure 18: Both models tested on White's thick with different "scales" and "levels" parameters. Without the residual image, the "levels" parameter in the BIWaM correlates to the "scales" parameter in the ODOG.

Despite their distinct implementations, the findings of this section suggest that the decomposition processes of both models are based on similar conceptual ideas. The residual image in the BIWaM represents a dissimilarity; however, without it, it was possible to demonstrate that both implement parameters, which strongly correlate. This finding provides further evidence that both models share same conceptual ideas.

5 Aligned models

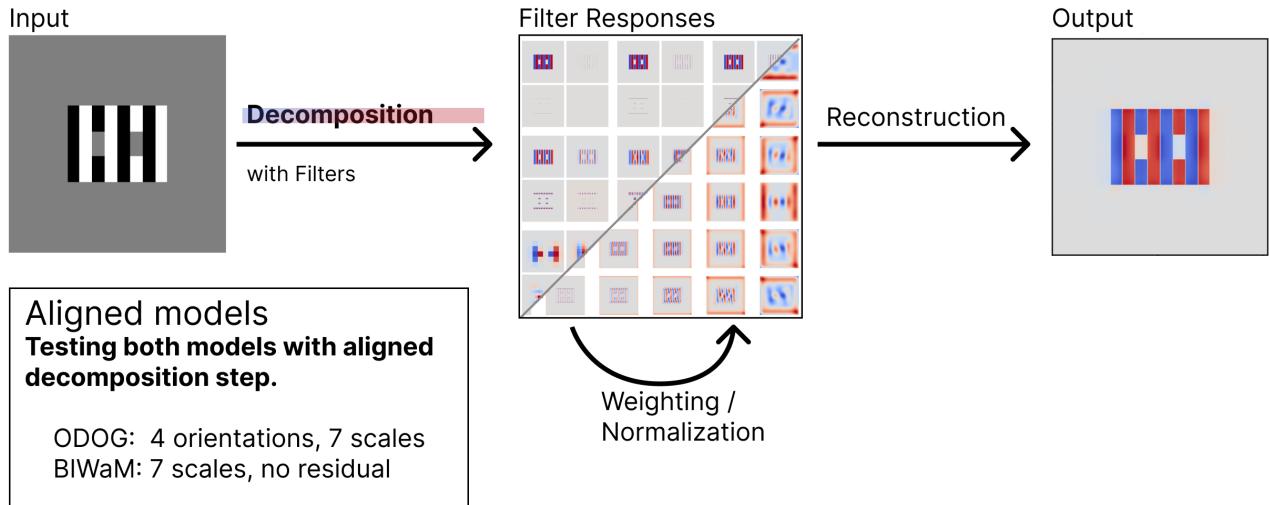


Figure 19: The Figure provides an methodological overview of the aligned models section. The approach of this section involved aligning conceptually corresponding parameters in the decomposition step of both models and testing these aligned models on the same stimuli. If the models fundamentally share the same idea, this alignment should result in more similar behavior. Additionally this would provide models with less distinct decomposition steps, which will be useful to isolate differences in further processing steps, e.g. weighting and normalization.

The previous section identified similarities and dissimilarities in the decomposition steps of both models. The approach of aligning their corresponding parameters and mechanisms provides a way to evaluating whether ODOG and BIWaM represent fundamentally distinct approaches or variations of the same conceptual idea. If parameter alignment results in more similar responses, it may indicate that the behavioral differences of the models are attributable to various implementation rather than to distinct underlying principles.

To align the models I made use of the findings from the previous section. The corresponding parameters "scales" (ODOG) and "levels" (BIWaM) will be left at seven, as they were already the same. It was ensured that the filters had the same scales, therefore both models extracted the same spatial frequencies. The residual image in the BIWaM model was ignored. To my understanding of the orientations parameters, aligning the models requires that the ODOG model operates with four orientations—vertical, horizontal, and two diagonals—similar to BIWaM, as explained above. Because BIWaM's parameters cannot be modified, alignment must be achieved by adjusting the ODOG parameters accordingly.

To evaluate the effect of these alignments on the outputs, the test stimuli were processed using the models with the alignment. The outputs were then analyzed to identify changes in comparison to the baseline values. The first test was conducted on the White's thick stimulus as shown in Figure 20 (a).

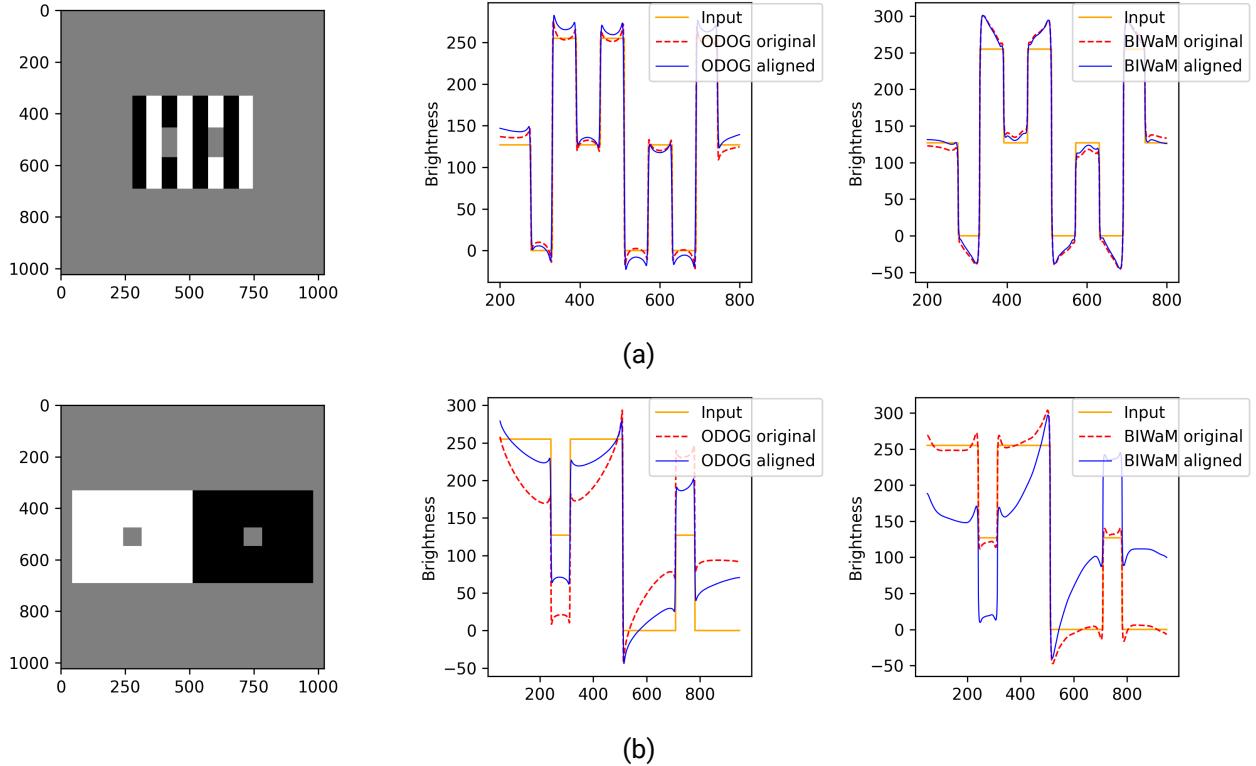


Figure 20: Both models with aligned decomposition steps were tested on (a) White’s thick and (b) SBC large. The blue lines represent the responses of the aligned models, while the red dotted line shows the original model responses.

For White’s thick, the alignments resulted in a slight shift in the model outputs toward each other. The aligned ODOG exhibited a stronger prediction for the test patches than its original prediction, while the aligned BIWaM exhibited a weaker prediction than its original prediction. In conclusion, the aligned models exhibited a comparable level of prediction strength, while each model still expressed its characteristics. The overshooting and undershooting in the ODOG became more prominent, reaching a point that was similar to the ones in the BIWaM.

The aligned models were also tested on the SBC large stimulus, as shown in Figure 20 (b). Here the alignments had a more noticeable effect. The ODOG lost some of its originally “exaggerated” strength in predicting the test patches, while the BIWaM gained comparable strength to that of the original ODOG. In addition the aligned ODOG started following other regions of the input image more accurately, whereas the BIWaM lost this capability when aligned. In fact, for the SBC large the aligned BIWaM approximated the initial ODOG response from Figure 11 (d).

This comparison demonstrated that when aligned in their decomposition steps, the models behaved similar overall while each model still expressed its characteristics. The remaining differences are presumably based on their following processing steps (weighting and normalization). The similarities between the filter bank and wavelet decomposition suggest that they are simply different implementations of the same underlying concept. Furthermore, the alignments provided two model versions for further testing, in which decomposition is isolated as the primary source of differences.

Effect of alignments on leading of BIWaM

Figure 21 shows how the alignments affected the leading of the BIWaM in predicting (a) White's circular, (b) Checkerboard and (c) Dungeon illusion. For White's circular and Dungeon illusion the BIWaM was not able to predict the test patches anymore. For the Checkerboard it still predicted the test patches. Specifically in the Dungeon illusion the behavior of both models were similar insofar as that both responded with a low frequency sinusoidal trend. The only adjustment on the BIWaM was to ignore its residual image, thus its initial capabilities to predict White's circular and Dungeon can be attributed mainly to the residual image.

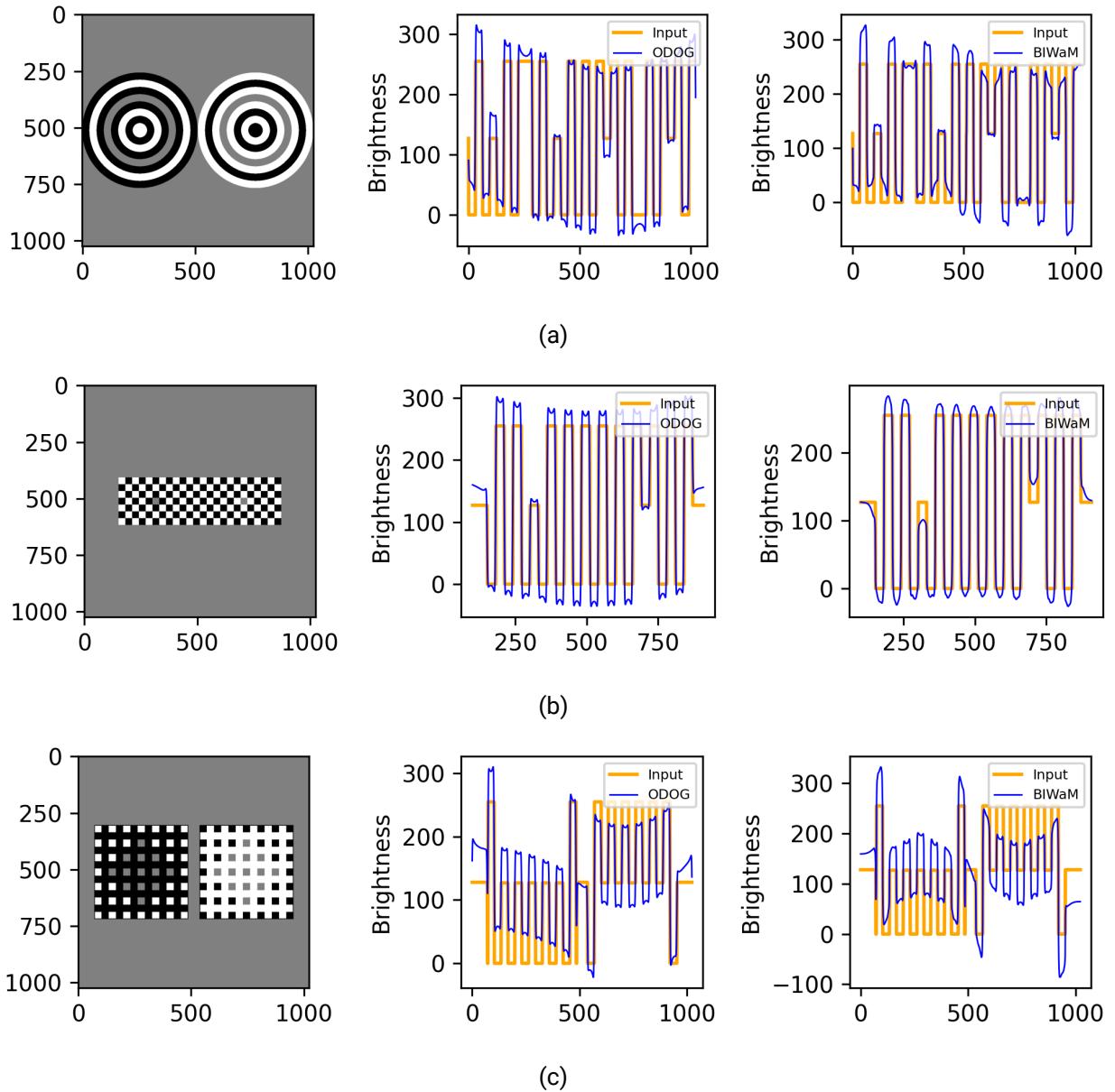


Figure 21: Both models (with aligned decomposition) were tested on (a) White's circular, (b) Checkerboard and (c) Dungeon illusion. This test showed how the alignments affected the initially leading of BIWaM in predicting those stimuli, as shown in Figure 12.

6 CSF weighting

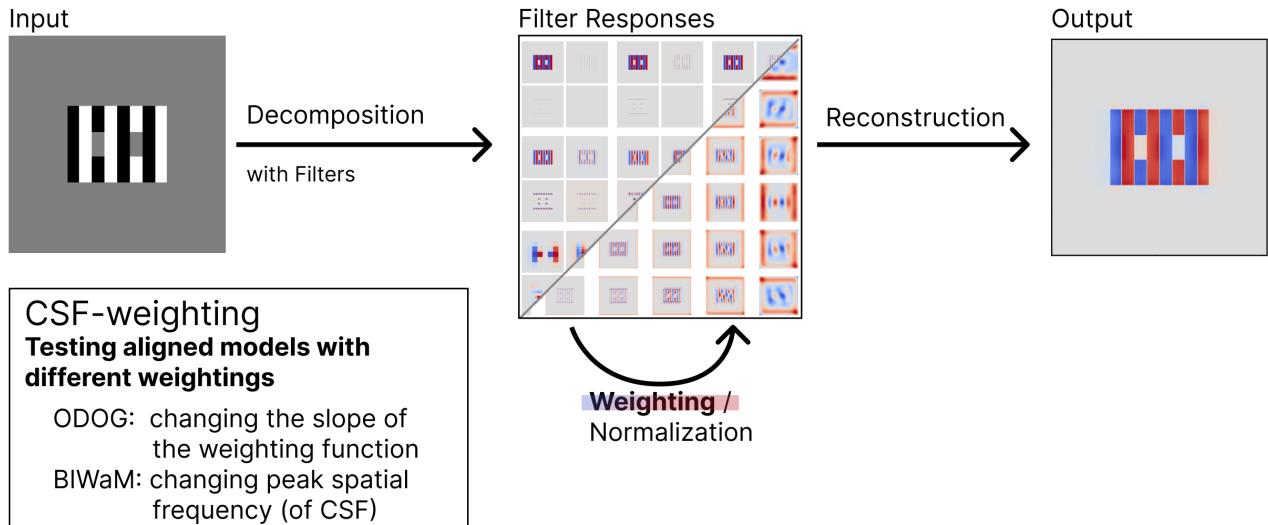


Figure 22: The Figure provides an methodological overview of the CSF weighting section. To address the research question, the approach in this section focused on identifying conceptually corresponding parameters in the weighting steps of both models. If adjusting these parameters leads to similar changes in their outputs, this would indicate that both models share the same fundamental concept for weighting.

With the models aligned in their decomposition steps, and therefore the decomposition isolated as primary source of differences, this section explored the distinct implementations of the weighting step in the ODOG and BIWaM. If both models really only are different implementations of the same conceptual idea, I expected to find corresponding weighting parameters in both models. Furthermore, the variation of these parameters should produce correlated changes in the model outputs.

The impact of different-scaled filters on the model output is determined by the factor they are weighted with. These factors are defined by the weighting functions of each model. The weighting functions are shown in Figure 23.

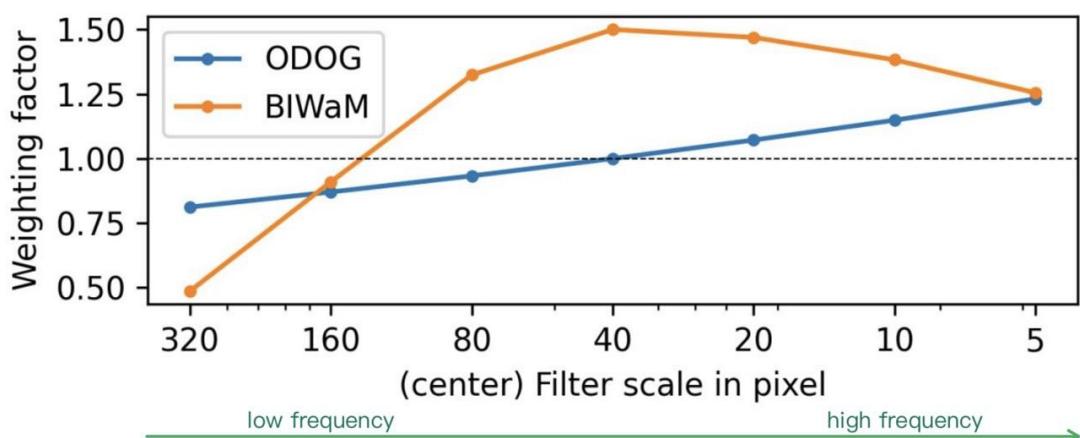


Figure 23: The weighting functions are determining the weights for different scaled filters. BIWaM implements a quadratic function, while ODOG utilizes a linear function. Filter scales below the dotted line are attenuated and filter scales above are amplified.

In each model there is a parameter that defines the weighting function. The ODOG model has a parameter called "slope" defining the slope of the linear (in logarithmic space, so effectively an exponential) weighting function. In its published configuration the slope is set to 0.1 (per logarithmic spaced filter scale) where the center filter (40 pixels) is at a weight of one. A positive slope attenuates lower spatial frequencies (as in Figure 23) and amplifies higher spatial frequencies, whereas a negative slope has the opposite effect. The BIWaM implements a parameter called "peak", which defines the peak of its quadratic weighting function. The default value is set to 4, meaning that its weighting function peaks at the 4th scale, which translates to a filter scale of 40 pixels. Overall, the BIWaM weighting function is more accurate, considering that both models were inspired by the CSF of the human visual system.

The parameters "slope" (ODOG) and "peak" (BIWaM) are conceptually correlating, as they either attenuate lower spatial frequencies and amplify higher spatial frequencies or vice versa. To verify this correlation, I have tested the aligned models with varying values of "slope" and "peak" respectively on the test stimuli, shown in Figures 24 (a) and (b).

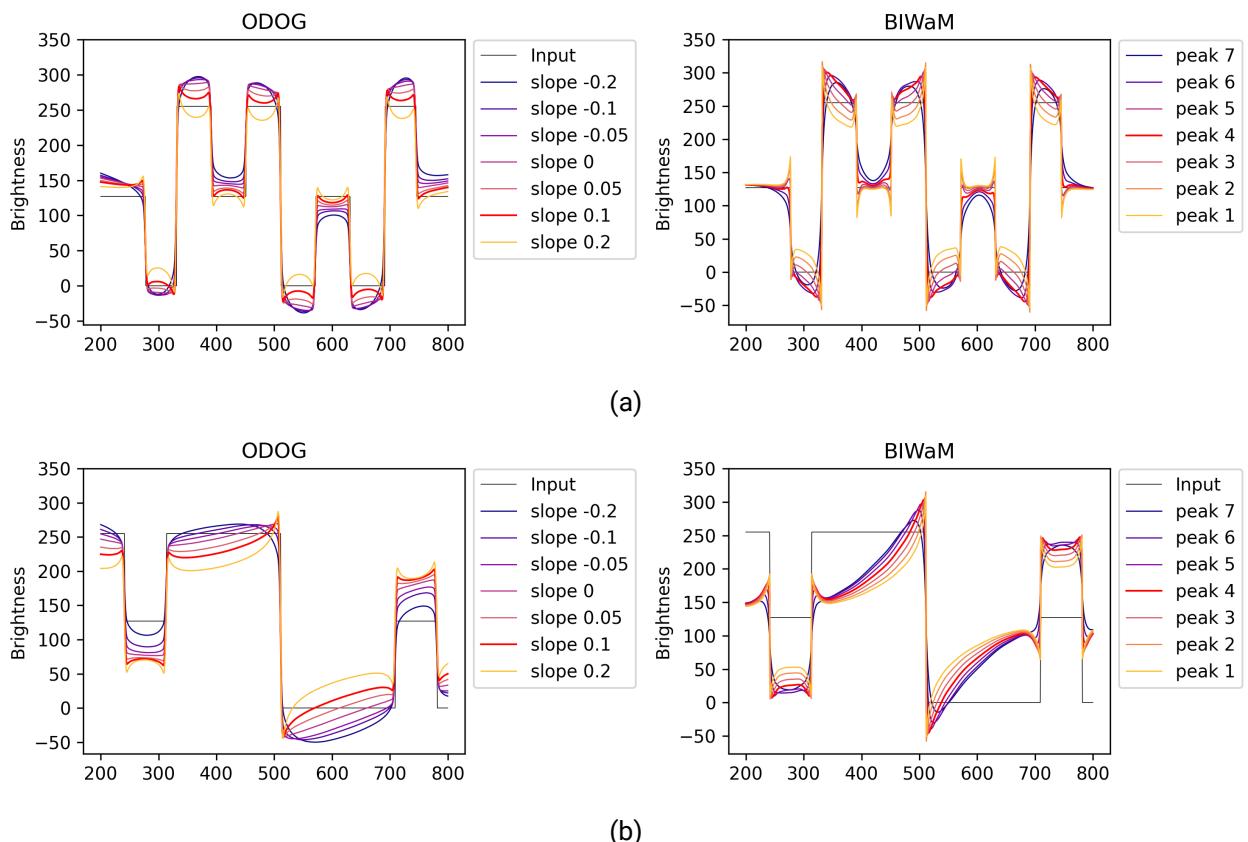


Figure 24: Both models (with aligned decomposition) were tested on (a) White's thick and (b) SBC large with different "slope" and "peak" parameters. The red lines represent the original values for the parameters. The most crucial analysis here is, that the effect of the conceptually corresponding "slope" and "peak" parameters did correlate.

Both models showed similar reactions to the parameter test in both stimuli. On smaller "slope" and higher "peak" values, low frequencies are amplified, which is expressed by the sinusoidal trend as shown in Figure 24 (a) and (b). This is reasonable, since large-scaled filters are more represented in the model outputs with such parameters and they are effectively smoothing the image. At the other end (larger "slope" and smaller "peak" values) the responses strike out at the edges, which is explained by small-scaled filters that are most sensitive to edges. Both models lost their distinct characteristics for the first time, indicating that the weighting is crucial for their behavior.

For White's thick both models and for SBC large only BIWaM benefited from weightings that are contradictory to the CSF, as shown in Figure 24 (a) and (b). Only for SBC large the ODOG showed that a deviation from the CSF will weaken the prediction strength, while still predicting the test patches in the right direction.

However, since the different spatial frequencies of both models are never really weighted with same weights, as their weighting functions and the influences of the parameters "slope" and "peak" are too distinct, I also tested both models with equally weighting across all spatial frequencies. Therefore, the slope of the ODOG model is set to zero, so the linear weighting function will stay at one for all filters. Since the "peak" parameter alone does not allow to set an equal weighting, I modified the BIWaMs source code, to do the same. This in fact resulted in the elimination of the different weightings. The results are shown in Figures 25 (a) and (b).

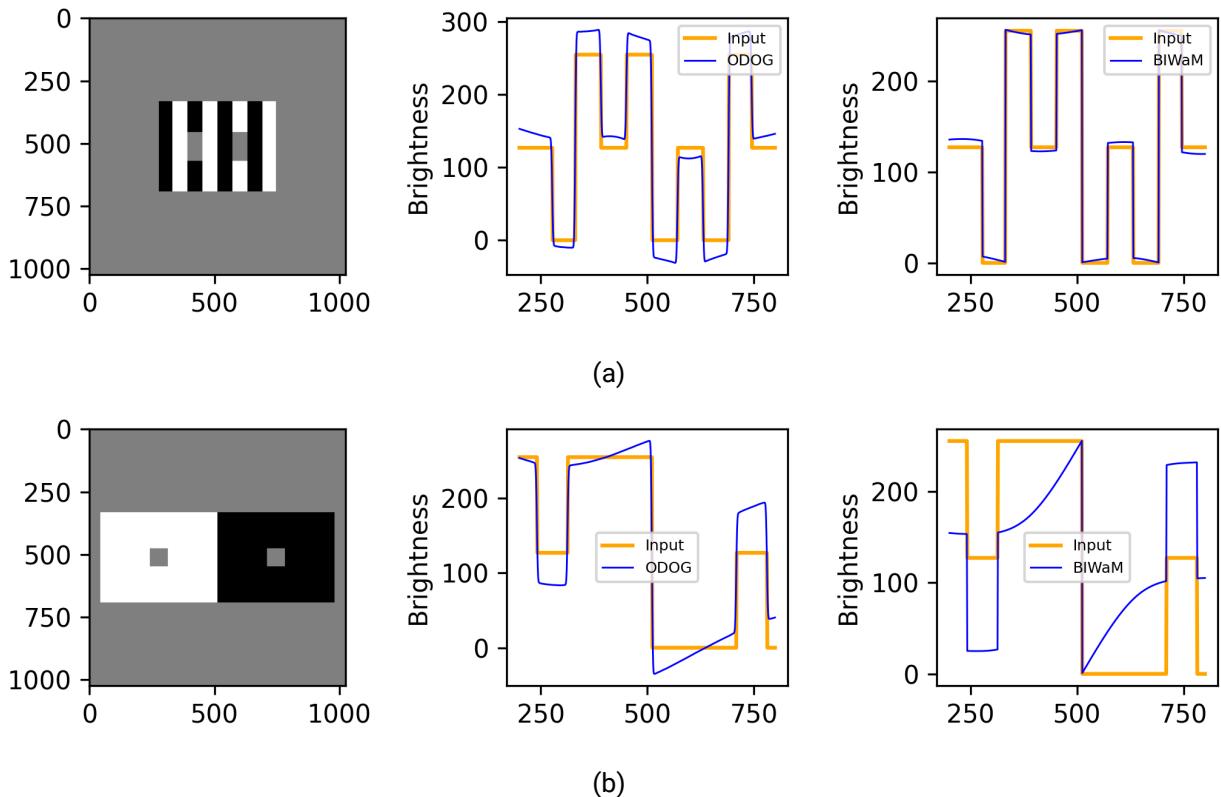


Figure 25: Both models (with aligned decomposition) were tested without weighting on (a) White's thick and (b) SBC large. For (a) ODOG was still able to predict the test patches, while BIWaM was not. For (b) both models predicted the test patches right. Overall both models lost their characteristics without weighting.

Without weighting the aligned models lost most of their characteristics and demonstrated a more angular expression as shown in Figure 25. Apart from that the ODOG was able to predict both stimuli, with good prediction strength in the White's thick, while the BIWaM was only able to predict the SBC large, where it approximated the initial ODOG response from Figure 11 (d). A noticeable difference was that BIWaM no longer exhibited overshooting or undershooting (only at the test patches), whereas ODOG still did.

Despite their distinct implementations, the findings of this section indicate that both models employ weighting processes based on similar conceptual ideas. While some residual differences remain, the strong correlation between the parameters "slope" and "peak" suggests that the models represent two implementations of the same underlying principles.

Effect of weighting on leading of BIWaM

Figure 26 shows how the weighting affected the leading of the BIWaM in predicting the Checkerboard illusion (last shown in Figure 21). The test shows that the BIWaM could not account for the Checkerboard illusion anymore, when the weighting is equally for all filter responses. The original behavior was unrecognizable in this illusion. Noticeable is the accurate following of the Input graph.

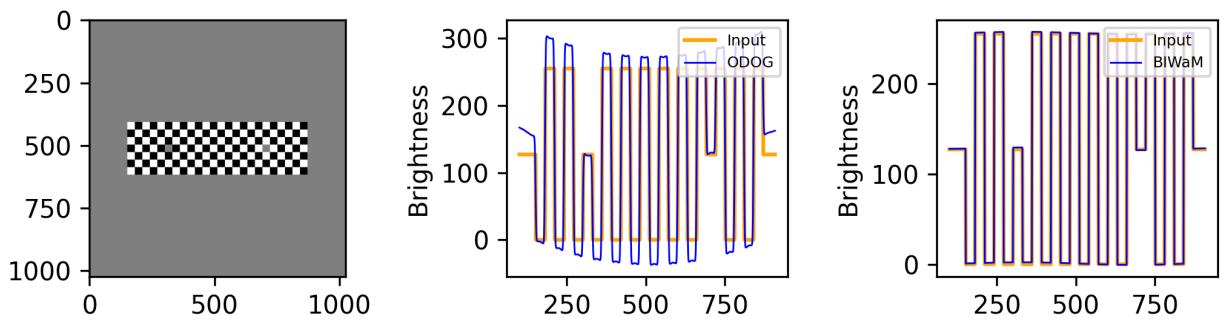


Figure 26: Both models (with aligned decomposition) were tested without weighting on the Checkerboard illusion. This test showed how leaving out the weighting affected the initially leading of BIWaM negatively.

7 Normalization

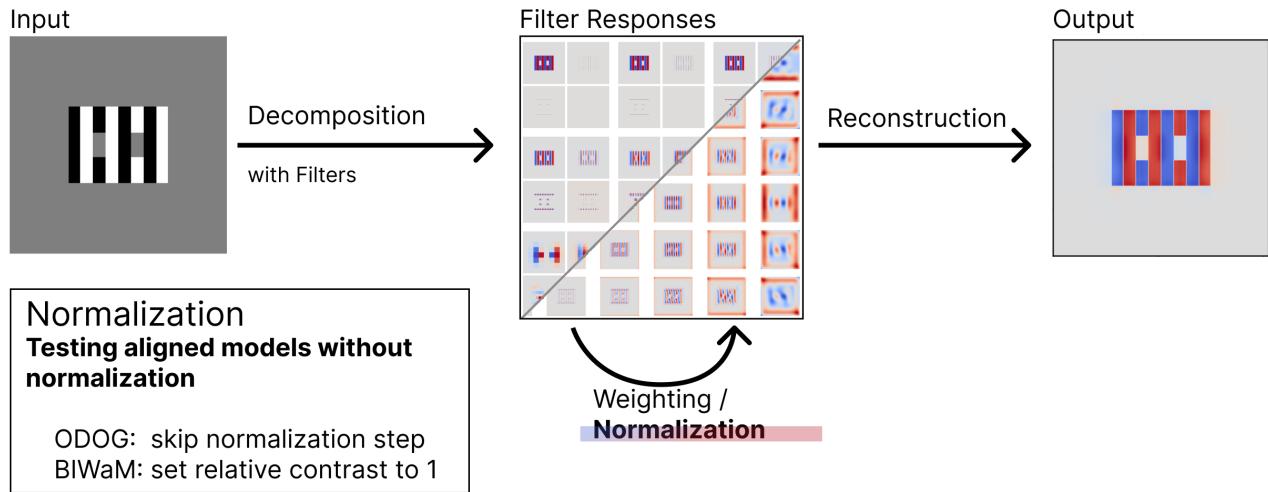


Figure 27: The Figure provides an methodological overview of the Normalization section. The goal of this section was to analyze the distinct implementations of the normalization steps in ODOG and BIWaM. If the models differ only in implementation but share the same conceptual foundation, I expected finding corresponding parameters. Adjusting these parameters should then result in similar changes in their outputs.

The comparisons in Figure 25 show that with aligned decomposition and (no) weighting both models still show differences in prediction strength, direction of prediction and some residual characteristics. Due to the alignments, these differences can be mainly attributed to the normalization step. This section explored the distinct implementations of the normalization steps of both models. If the models differ only in implementation while sharing the same conceptual foundation, I expected to find corresponding parameters and observe a correlation between their output changes when these parameters were varied.

The ODOG applies a global and orientation-based normalization. Therefore, it first pools the filter responses across all orientations. Every pooled orientated filter response will be then normalized by a normalization factor so that the root-mean-square levels across pooled orientated filter response are equated.

In contrast, the BIWaM calculates the "relative contrast" of every filter response region compared to its surrounding. This more complex computation is naturally implemented in the wavelet decomposition and results in a local and scale-and-orientation-based normalization.

In general the normalization equates the filter responses and ensures that interactions between scales and between orientations are possible (Kingdom, 2011).

The investigation in both model approaches to normalization and their implementations did not provide any conceptually correlating parameters like in the other sections. This may be due to the very distinct implementations, where especially in the BIWaM the normalization step is entangled in its other mechanisms, like the wavelet decomposition and the CSF weighting. However, I was able to find a way of ignoring the normalization in both models, similar to what Robinson did in the UNODOG model (Robinson et al., 2007).

Therefore, I modified both implementations to rather skip the normalization (ODOG) or overwrite the normalization factors with ones (BIWaM). This resulted in two versions of the models without any normalization. These versions were tested on the test stimuli, the results are shown in Figure 28.

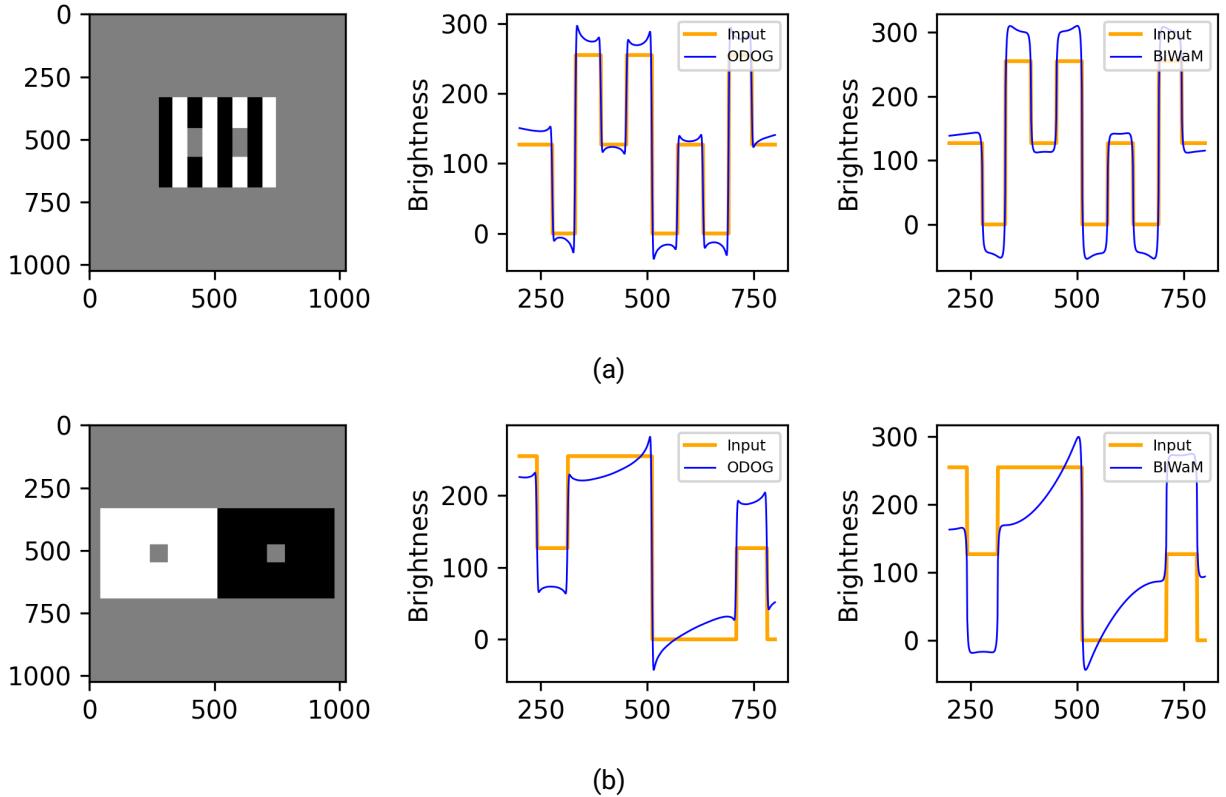


Figure 28: Both models (with aligned decomposition) were tested without normalization on (a) White's thick and (b) SBC large.

The left out normalization had similar effect on both models with aligned decomposition as showed in Figure 28. While both models still predicted the SBC large, they could not predict White's thick anymore. Additionally, the BIWaM lost some details of its characteristics, e.g. the spiking overshooting in the White's thick and the prediction at the test patches of SBC large are more flat compared with the model with normalization from Figure 20. The ODOG did not changed any characteristics, execpt of the prediction in the White's thick. This is reasonable, since the SBC large is nearly isotropic and the orientation-based normalization of the ODOG has less influence on such images than the normalization of the BIWaM.

Although the models differ drastically in their implementations, the results of this section suggest that their normalization step has similar advantages for both models. However, unlike in previous sections, no conceptually corresponding parameters were identified, indicating that their fundamental concept for the normalization is too distinct for direct comparisons.

8 Discussion

In the first part of this thesis both models were boiled down to three main processing steps they both use: Decomposition, Weighting and Normalization. Although the concepts of both models are similar at first glance, their structural approaches and implementations differ a lot and leave open questions about their relationship (Chapter 1.5). This thesis established a chain of evidence to investigate whether the ODOG and BIWaM share a common fundamental concept and represent various implementations of it, or if they are two different types of multiscale spatial filtering models.

Both models showed different capabilities in predicting several stimuli, which were specifically chosen to answer the research question. The BIWaM model demonstrated superior capabilities by predicting all test stimuli with more strength or with more plausible magnitude, this is in alignment with the available lectures (Betz et al., 2015; Otazu et al., 2008). The ODOG model failed to predict three of the seven stimuli (Chapter 2.3). To preclude susceptible behavior, and to test both models with their processing steps being unmodified (later they were modified) on parameter changes, the models were tested on input stimuli, which underwent controlled variations. Both models demonstrated overall predictable and (more important) similar behaviors on the variation, presenting the first clue, that both share common principles (Chapter 3). This laid the foundation for parameter testing, which formed the heart of this thesis.

To reveal the distinct implementations of both models, I investigated their algorithms and catalogued their source codes into three main processing steps: Decomposition, Weighting and Normalization. Each of these processing steps was successively and systematically analyzed before being tested for (dis)similarities between the models.

In the decomposition I found two parameters, one in each model, which conceptually correlated: "scales" in the ODOG and "levels" in the BIWaM. I showed that they define the sizes of the filters used to decompose the image for each model. Therefore I assumed that when varying those parameters, both models responded in correlating ways. But this was not the case. The BIWaM seemed to not react to the changes in "level", which I attributed to the residual image. This residual image is the first key finding, which made the BIWaM more capable. In a second test I varied the parameters "scales" and "levels" again and ignored the residual image in the BIWaM. And lo and behold both models reacted in the same way, presenting the second clue, that both share common principles (Chapter 4).

I used the findings mentioned above to align both models in their decomposition steps. Both models were set to four orientations (vertical, horizontal, and two diagonals), seven scales and the BIWaM was modified to ignore its residual image. This setup was tested on several stimuli and showed that the alignments affect both models. The results showed how both models approximated each other. It was noticeable that the initial superior BIWaM lost its lead and became weaker for the stimuli where it was stronger before, while it became stronger in one stimulus, where it was weaker before (SBC). Specifically, it could no longer predict White's circular and Dungeon illusion (as shown in Figure 21). The decomposition step differed primarily in number of orientations and, more significantly, the presence of a residual image in BIWaM. Removing the residual image made BIWaM behave more similarly to ODOG, reinforcing the hypothesis that both models share a common framework with different implementations. The residual image was the

first key finding, which represented a conceptually dissimilarity.

But it is not psychophysically plausible, since the human visual system has no access to such residual image information. The other processes involved in the decomposition step showed more similarities when aligned in orientations, presenting a third clue, that both share common principles (Chapter 5).

In the next section I tested the weighting step of both models. Therefore, I determined the two conceptually correlating parameters "slope" in the ODOG and "peak" in the BIWaM. Both parameters either attenuate lower spatial frequencies and amplify higher spatial frequencies or vice versa. A test showed that, as expected, varying these parameters had the same effect on the model behaviors. This provided a fourth clue, that both share common principles. Ironically, both models were inspired by the CSF of the human visual system, but mostly benefited from contradictory settings regarding the CSF (as shown in Figure 24). However, since it was not possible to align the weighting in both models to a sufficient degree, I tested both models with equally weighted spatial frequencies. The results demonstrated how the BIWaM was more sensitive to this modifications, indicating that its intricate implementation may differ apart from parameters. (Chapter 6)

The above mentioned results showed that with aligned decomposition and (no) weighting both models still demonstrated differences in prediction strength, direction of prediction and some residual characteristics. To determine if these remaining differences can be resolved by further alignments in the normalization step of both models, I investigated the implementations of both models. Since there were no parameters in both models which could have been tested and the algorithms of the BIWaM were convoluted into other processing steps (e.g. wavelet decomposition, csf-weighting), I tested the models without any normalization, like Robinson did with the UNODOG (Robinson et al., 2007). Therefore, I modified both implementations to rather skip the normalization (ODOG) or overwrite the normalization factors with ones (BIWaM). The tests demonstrated, that both models rely on normalization only for the White's large stimuli (as shown in Figure 28). Like in the weighting tests, the BIWaM was more sensitive to the modifications compared with the ODOG. For SBC large, both models did not react to the absence of the normalization. These findings align with Robinson's conclusion that normalization is crucial for White's effect but plays a minor role in most other illusions (Robinson et al., 2007).

Essentially, the most meaningful differences between the models have been isolated to be in the normalization step. Decomposition and weighting also were more accessible for testing, while the normalization step in the implementation was so involved in other processes (e.g. wavelet decomposition, csf-weighting) that it was difficult to understand the origins of differences. A more comprehensive evaluation and testing of both models would provide further clarity.

To address another limitation of this thesis, it is necessary to mention that the mathematics underlying the models (especially the BIWaM) is essential to understanding their differences, but there was no room for this in this thesis. In order to provide even more clarity in the field of multiscale spatial filtering models, this investigation is necessary.

9 Conclusion

The research question of this thesis aimed to determine whether the ODOG and BIWaM models are fundamentally distinct or really represent different implementations of the same conceptual idea. After a systematic analysis and comparison of their decomposition, weighting, and normalization steps, the evidence suggests that these models share fundamental conceptual similarities despite their distinct implementations.

Key findings supporting this conclusion include:

1. The decomposition step showed that conceptually corresponding parameters ("scales" in ODOG and "levels" in BIWaM) had similar effects on model outputs.
2. The residual image in the BIWaM represents a dissimilarity, but is psychophysically implausible.
3. Aligning the models decomposition steps resulted in more similar overall behavior while retaining some distinct characteristics.
4. The weighting mechanisms, though implemented differently, showed correlating effects when their conceptually corresponding parameters ("slope" in ODOG and "peak" in BIWaM) were varied.
5. Both models demonstrated similar responses to the removal of normalization, particularly in their inability to predict White's effect without this step.

These findings indicate that the differences between ODOG and BIWaM are more likely due to differences in implementation rather than fundamental conceptual distinctions. However, it's important to note that the BIWaM model generally showed superior performance, particularly in predicting certain illusions that ODOG could not account for, suggesting that its more complex implementation, especially in the weighting and normalization steps, contributes to its enhanced capabilities. However, many of BIWaM's abilities can be traced back to the residual image, although it is unclear to what extent this is acceptable when it comes to replicating human vision.

This conclusion supports the hypothesis that ODOG and BIWaM represent different implementations of the same underlying conceptual framework.

List of Figures

1	Relationship between illuminance, reflectance and luminance	1
2	Lightness and brightness are distinguishable	2
3	Checkerboard shadow illusion	3
4	The principle of center surround fields	4
5	Applying a convolution on an image	5
6	Models response to White's Effect	6
7	DoG filter example	7
8	Structural overview of models	8
9	Methodology overview	9
10	Overview 7 different stimuli	11
11	Initial model responses 1-4	13
12	Initial model responses 5-7	14
13	Robustness Overview	16
14	Variations in White's thick	17
15	Variations in White's thick	18
16	Decomposition Overview	19
17	Varying decomposition parameters	20
18	White's with differnet scales	21
19	Aligned Models Overview	22
20	Tests with aligned models	23
21	Tests with aligned models 2	24
22	CSF Weighting Overview	25
23	Weighting functions	25
24	Different weightings	26
25	Without weighting	27
26	Without weighting checkerboard	28
27	Normalization Overview	29
28	Models without normalization	30

References

- Adelson, E. H. (1995). Checkershadow illusion [Visited on: 2025-01-14].
<https://persci.mit.edu/publications>
- Adelson, E. H. (2000). Lightness perception and lightness illusions. *The New Cognitive Neurosciences*, 2nd ed., M. Gazzaniga, ed. Cambridge, MA: MIT Press, 339–351.
- Anstis, S. (2005). White's effect in color, luminance and motion. *Seeing spatial form*, 91–100.
- Arend, L. E. (1993). Lightness, brightness, and brightness contrast.
- Betz, T., Shapley, R., Wichmann, F. A., & Maertens, M. (2015). Noise masking of white's illusion exposes the weakness of current spatial filtering models of lightness perception. *Journal of Vision*, 15. <https://doi.org/10.1167/15.14.1>
- Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization [Visited on: 2025-01-14]. *Vision Research*, 44(21), 2483–2503.
<https://doi.org/https://doi.org/10.1016/j.visres.2004.05.015>
- Blakeslee, B., & McCourt, M. E. (1997). Similar mechanisms underlie simultaneous brightness contrast and grating induction.
- Blakeslee, B., & McCourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. www.elsevier.com
- Brainard & Longere, K.
bibinitperiod. (2003). Colour constancy: Developing empirical tests of computational models. In R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world* (pp. 308–326). Oxford University Press.
- Gundersen, G. (2017). From convolution to neural network [Visited on: 2025-01-14].
<https://gregorygundersen.com/blog/2017/02/24/cnns/>
- Kingdom, F. A. (1997). Simultaneous contrast: The legacies of hering and helmholtz [PMID: 9474338]. *Perception*, 26(6), 673–677. <https://doi.org/10.1068/p260673>
- Kingdom, F. A. (2011, April). Lightness, brightness and transparency: A quarter century of new ideas. <https://doi.org/10.1016/j.visres.2010.09.012>
- Kingdom, F. A. (2014). Brightness and lightness. MIT Press, 499–509.
- Murray, R. F. (2021). Lightness perception in complex scenes [Visited on: 2025-01-14].
<https://doi.org/10.1146/annurev-vision-093019>
- Otazu, X., Vanrell, M., & Párraga, C. A. (2008). Multiresolution wavelet framework models brightness induction effects. *Vision Research*, 48, 733–751.
<https://doi.org/10.1016/j.visres.2007.12.008>
- Robinson, A. E., Hammon, P. S., & de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vision Research*, 47, 1631–1644.
<https://doi.org/10.1016/j.visres.2007.02.017>
- Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision [Visited on: 2025-01-14]. *J. Opt. Soc. Am.*, 61(9), 1176–1186.
<https://doi.org/10.1364/JOSA.61.001176>
- W., C. F., & G., R. J. (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, 551–566.
<https://doi.org/10.1113/jphysiol.1968.sp008574>
- White, M. (1979). A new effect of pattern on perceived lightness [Visited on: 2025-01-14]. *Perception*, 8(4), 413–416. <https://doi.org/10.1068/p080413>