Universität Hamburg
Department Informatik
Knowledge Technology, WTM

Modeling Color Vision with Coding Strategies of Retinal Ganglion Cells

Diplomarbeit
im Studiengang Informatik

Daniel von Poschinger-Camphausen
Matr.Nr. 5715435
4poschin@informatik.uni-hamburg.de

Erstbetreuer: Stefan Wermter, Zweitbetreuer: Cornelius Weber

June 23, 2015

Contents

1 Abstract

The main motivation of this work is to gain insight into improving the reliability of color information by following the coding strategies of the lower stages of the human visual system. In comparison to human perception, the color information of a camera is unreliable as it is greatly influenced by the luminosity and color of the illumination of a scene.
Established unsupervised computational models of self-organizing RGC and V1 receptive fields are extended to process RGB images and are trained upon images of natural scenes. The results show that localized color opponent RF are emerging in chromatically distinct channels, each entirely covering the visual space. The opponent texture of the resulting RF is filtering more biologically plausible chromatic contrasts, which are regarded as the fundamental building blocks of the human visual system's property of color constancy.
Because of the unmatched performance the visual system produces, filters capturing more biological plausible chromatic contrasts appear to be of value and are suggesting that computational models utilizing such preprocessing possibly show an improved performance.

2 Introduction

The retina of the human eye is often compared to a CCD or CMOS photo-sensor of a camera. Such a comparison neglects that several retinal neuronal layers perform pre-processing [Weber and Triesch, 2009, p.75] of the image stream such as compression, noise removal and contrast enhancement. In contrast to the output of luminance and color information of a CCD sensor the human retina has at least 17 different types of retinal ganglion cells (RGC) [Field and Chichilnisky, 2007, p.9] indicating that the 'output' of the retina is far more complex and feature rich than that of a simple camera.

In comparison to human perception, the color information of a camera is unreliable as it is greatly influenced by the luminosity and color of the illumination of a scene. Even a subtle change in the illumination results in some cases drastic changes in the color-values sampled by a camera sensor. This inconsistency has rendered unprocessed color information useless to judge the real color of an object.

2.1 Motivation

The main motivation of this work is to gain insight into improving the reliability of color information by following the coding strategies of the lower stages of the human visual system. Since the visual system shows an unmatched performance in many fields of computer vision following these coding strategies appears to be reasonable.

Regarding color constancy as the cathedral of computer vision, chromatic contrast (or color opponency) has been identified as its fundamental building block [Shapley and Hawken, 2011, p.702]. Despite being unknown of how the visual system establishes color constancy at cellular level, cells providing color opponency have been clearly identified in the ganglion cell layer of the retina and in the lowest layer (V1) of the visual cortex. This is further supported, by theoretical work suggesting that optimal filters matching the statistics of natural scenes, would resemble the population of V1 *single-* and *double color-opponent* cells [Shapley and Hawken, 2011, p.715].

Moreover, the question of whether the 17 types of RGC are hard coded in the human genome or whether the RGC types are evolving as a result of the surrounding image statistics is still unknown. Since competition of dendrites among RGC during the development of the retina [Linden and Perry, 1982] [Linden, 1993] has been observed, the latter appears to be a reasonable hypothesis. This allows to attribute some biological relevance to the results of training linear auto-associative networks on images of natural scenes.

2.2 Methodology

Established unsupervised computational models of self-organizing RGC and V1 receptive fields are extended to process RGB images (see 6 and 8). Receptive fields of the RGC model resulting of the training process are further analyzed by fitting

each receptive field with a spatio-chromatic Difference of Gaussian parametric (DOG) model.

In order to determine the number of distinct RGC types emerged, the results of the fitting process are classified by k-mean and spectral clustering algorithms. The identification of distinct RGC types allows to select a representative unit for each type. These prototypes are used as a convolution filters to demonstrate the preprocessing capabilities of the RGC model.

2.3 Experiments and Evaluation

Both models of RGC and V1 are trained in an auto-associative manner on RGB images of natural scenes. The results show that, in conformance to prior results of applying these models to luminosity data, localized color opponent receptive fields (RF) are indeed evolving (see 7 and 8.2). Further analysis of the emerged RF of the RGC model reveals that the input is separated into three resp. six complete distinct channels covering the entire RGB color-space in a spatial mosaic organization of RF (see 7.5). Additionally, a second layer is trained upon the output of the RGC model simulating the self-organization of cortical V1 simple cells, demonstrating a direct application of the RGC model in preprocessing visual information (see 8).

2.4 Novelty

Since color opponent feature detectors have been manually constructed or derived statistically of RGB images [Yang et al., 2013] [Brown et al., 2011], the novelty in this work is the application of a linear auto-encoder network in directly learning optimal filters from RGB images of natural scenes.

The center surround texture of RF emerging in the RGC model show color opponent contrast of color values inside the boundary of each RGB channel which mirrors the results of Brown et al. However, by training the RGC model in certain settings, a separation of luminosity and chromatic aspects of input information emerges (figs. 25, 26) resulting in RF of more biologically plausible color opponent contrast. Color opponent receptive fields of this setting are superior in capturing more biological types of chromatic contrast, compared to those of the remaining settings (see 7.7). The separation of luminosity and chromatic aspects (see 9) simply by learning directly from RGB images has not been observed in other models.

Further, the simplification and continuous application of the RGC model's weight-constraint results in a competition among the hidden units in reconstructing the input, whilst weakly participating units eventually die off (see 6.3.3). This appears to be a useful property, since the competition results in a compact neural code, in which an overcomplete RGC model automatically adjusts the number of hidden units towards a complete ratio of visual to hidden units (see fig. 17).

3 Biological Background

Due to the vast complexities of neural interconnectivity that the human brain exhibits, solely cell populations and corresponding cortical areas of the lower visual system are described in this section which are assumed to be involved with the processing of color stimuli. The aspects of numerous inter-connections of the visual system to other parts of the brain are not considered.

3.1 Primate Visual System

The mammalian visual system can detect and discriminate visual stimuli of various kinds (chromatic, patterned, static, in motion ...). Whilst the exact answers to the question of which discrete anatomical pathways exist and what kind of information these channels propagate are yet unknown [Schmolesky, 2007, p.1] the visual system and its functional aspects are well understood compared to other areas of the central nervous system [Krüger et al., 2010, p.4].

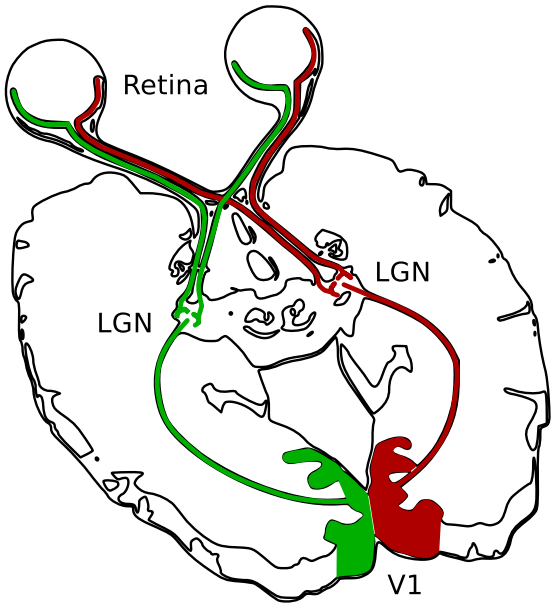

Figure 1: In the human visual system, information of the left and right visual field is propagated from temporal and nasal *retina* through optic chiasm and *lateral geniculate nucleus* (LGN) to the lower areas of the visual cortex (V1).

Commonly, the visual system is seen as consisting of the *visual cortex* and precortical areas such as the eye's *retina*, optic nerve and optic chiasm and the *lateral geniculate nucleus* (LGN) (an area of the thalamus). Visual information is preprocessed in the retina and relayed and further processed through the remaining precortical areas to the visual cortex (see fig. 1).

The visual system is hierarchically organized, where each layer builds on top of the previous layer and is processing its synaptic output. The neural circuity in the retina and other precortical areas of the visual system suggests that visual Information is processed exclusively in a feed-forward manner. In contrast, sub-cortical areas of the visual cortex are highly interconnected, indicating the recurrent processing of visual information [Krüger et al., 2010, p.3].

3.2 Anatomy of the Retina

Neuroanatomical studies have shown that the mammalian retina consists of many parallel, equally potent microcircuits [Masland, 2001, p.877] forming discrete pathways which process different aspects of the visual information. Each of these microcircuits are composed of five types of neurons, organized in a layered structure,

namely: *rod* and *cone* photoreceptors, *bipolor, horizontal, amacrine* and *ganglion* cells (see fig. 2). Each cell-type has multiple sub-types with varying morphology and connectivity. Notably, ganglion and amacrine cells represent visual information by firing action potentials, photoreceptors, bipolar and horizontal cells represent information with graded potentials [Field and Chichilnisky, 2007, p.4].

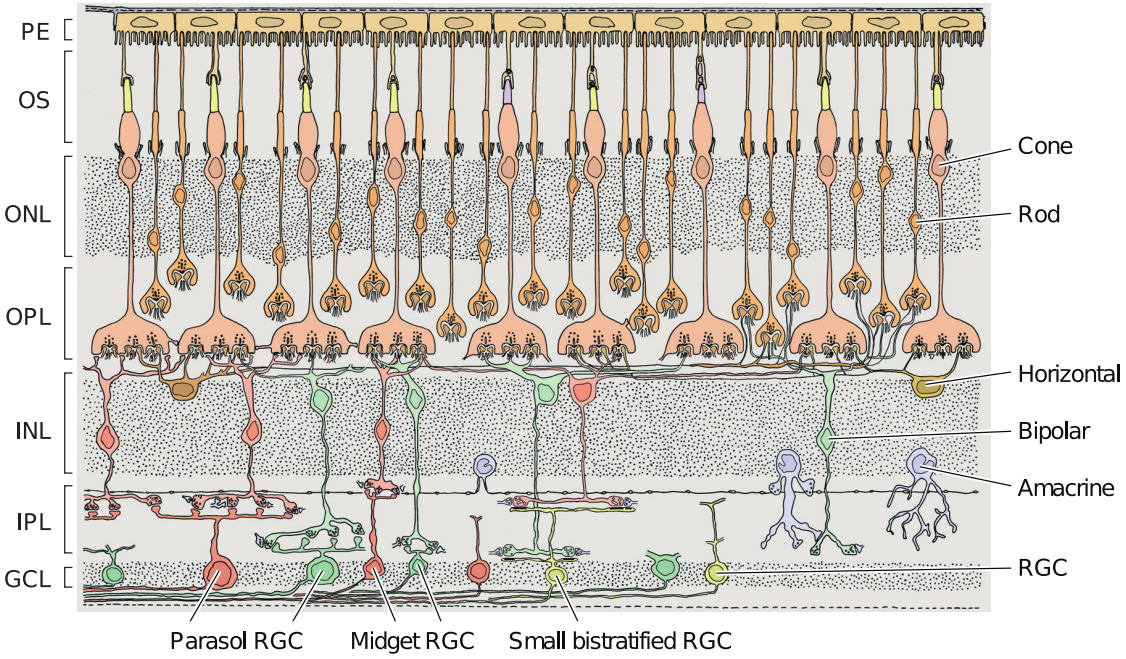

Figure 2: Layers (left) and cell types (right) of the primate retina [Field and Chichilnisky, 2007].

The retinal-layers and corresponding cell types, form a processing pipeline in which visual stimuli are propagated solely in a feed-forward manner, though wide and narrow lateral intra-layer connections are common [Masland, 2001, p.878]. The distribution of cell densities (measured in cells per mm^2) in the primate retina is rotation invariant and concentric around the the center of the retina, the fovea [Weber and Triesch, 2009, p.79].

3.2.1 Photoreceptors

Rod and *cone* photoreceptors, found in the pigment epithelium (PE) layer, convert light into electrical signals via a cascade of biochemical protein interactions named phototrandsduction [Ebrey and Koutalos, 2001, p.51]. Photoreceptor cells transmit information by gradually changing its membrane potential. In the dark, the membrane of a receptor cell is depolarized, its potential is roughly around -40 mV. Influx of photons, absorbed by the cell's photo-pigment (a light sensitive membrane protein), results in a hyperpolarization of the cells membrane (the cell's membrane potential changes towards a more negative value) [Purves, 2001, Phototransduction]. Shortly afterwards the cell recovers from its hyperpolarized state, and return to a depolarized membrane potential.

Rods possess only one type of photo-pigment (rodopsin), found in high concentration in the cell membrane. Being not color selective and extremely sensitive to low amounts of light, a *rod* can signal the absorption of a single photon [Fu, 2010, p.1]. *Rods* are saturated at daylight and do not contribute to daylight vision. In contrast to *cones*, *rods* have a lower temporal (more signal amplification and integration) and a coarser spatial resolution. Moreover rods are absent in the fovea where visual acuity is the highest [Kolb, 2012, p.1].

Cones on the other hand have a higher spatial and temporal resolution and are concentrated in the fovea. However, *cones* are less sensitive to light, since they hold lower concentrations of pigments in the cell membrane, and are saturated only by intense light-stimuli. *Cones* exhibit several types of pigments, each optimally absorbing a specific wavelength. Associating a *cone* type to one specific photo-pigment seems correct in general, despite occasional findings of the presence of traces of a second pigment [Ebrey and Koutalos, 2001, p.58].

Humans exhibit three pigment types and subsequently three *cone* cell types: the blue-*cone* (S) with maximum excitation at wavelength 420-440 nm, green (M) at 534-545 nm and red (L) at 564-580 nm (see fig. 3). Interestingly, the curves of the intensity of *cone* cell responses in relation to the wavelength of the light stimulus overlap partly for red and green *cones*. No overlapping occurs between the blue and green / red *cones*.

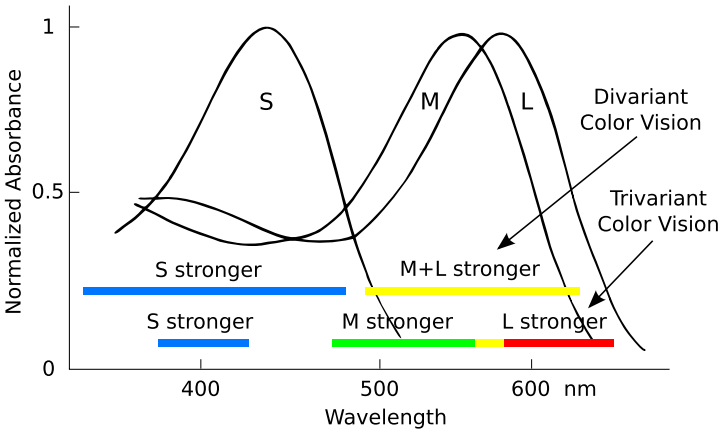

Figure 3: The shows the absorption spectra of primate L,M and S-cones. Below in colored boxes are the frequency windows in which a particular cone type responds stronger than the remaining types [Gouras, 2009].

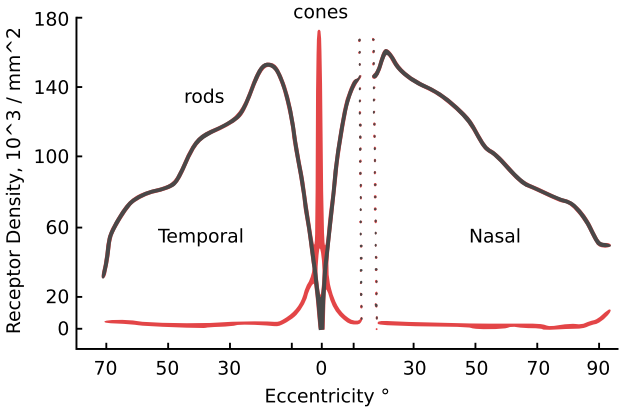

Figure 4: Rod and cone cell distribution in the human retina, after Østerberg (1935) [Kolb, 2012].

Rods and *cones* and their corresponding ganglion cells differ in their particular distribution (see fig. 4). In primate retinas *cones* occupy only 5 % of the photoreceptor cell population, whereat in the mouse retina *cones* constitute for 3 % of the photoreceptors leaving the remaining 97 % to *rods* [Fu, 2010, p.2]. In the fovea the *cone* cell density is maximal, with increasing distance from the fovea the density is decreases whilst the size of corresponding retinal ganglion cells re-

ceptive fields increases. Thus each ganglion cell receptive field covers larger areas of the visual field the more its corresponding *cones* are located in the periphery of the retina [Weber and Triesch, 2009, p.77]. In contrast, *rods* as well as blue-*cones* are absent in the center fovea. However, the *rod* density is maximal in the surrounding of the fovea and is decreasing with increasing distance. It has been estimated that one *cone* cell in the fovea center is associated to 2.6 ganglion cells and 3.4 bipolar cells on average [Ahmad et al., 2003]. Outside of the fovea region this ratio is much smaller (see 3.4, compression).

A photoreceptor's synaptic activity can only vary in magnitude and is therefore highly ambiguous: Photoreceptors act as photon counters therefore wavelength and intensity information of a particular photon is lost [Gegenfurtner, 2003, p.563]. As such a response of a photoreceptor is reflecting only the amount of energy absorbed, thus many combinations of wavelength and intensity result in the same cone output [Gouras, 2009, p.1]. Therefore to estimate the spectral composition of a stimulus, the output of two or more *cones* [Masland, 2001, p.878] has to be compared (see fig. 5).

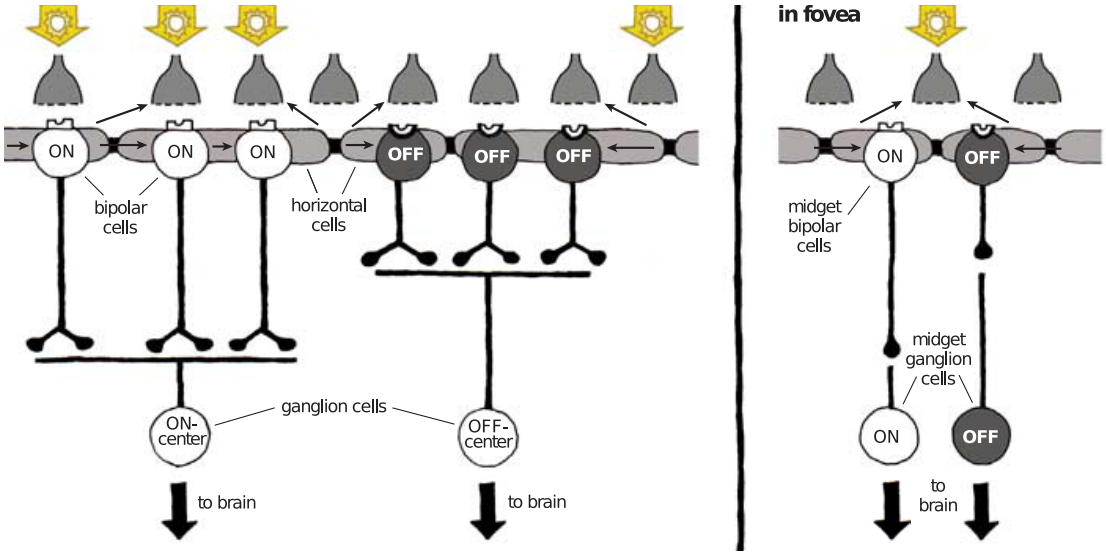

Figure 5: Basic circuity of the human retina (after [Kolb, 2004]). In the fovea a particular ganglion cells receives input solely from one bipolar cell.

3.2.2 Horizontal Cells

Located in the inner nuclear layer (INL), *horizontal cells* provide antagonistic feedback to photoreceptors. If a photoreceptor is hyperpolarized by an increasing amount of light or depolarized (decreased amount) it receives an opposing input after a synaptic delay from its connected *horizontal cells* [Gouras, 2009, p.3]. This negative feedback is assumed to adjust the photoreceptor's response to an overall level of illumination. Adjusting the cone output can be seen as a normalization enabling comparison with other cone outputs [Gouras, 2009, p.6]. The rod and cone feedback systems are isolated from each other preventing mutual interference:

Illumination levels covered by rods and cones are so distant [Masland, 2001, p.881] that interference would be undesireable.

For the rod photoreceptor cell type there exists one type of horizontal cell in contrast to two types of *horizontal cells* H1 and H2 associated to cone cells. H1 cells only connect to L-cones, H2 connect to S and L(M)-cones [Gouras, 2009, p.3].

3.2.3 Bipolar Cells

Bipolar cells, also located in the INL, separate the cone output into ON and OFF signals. Stimulating the retina with light results in the depolarization of one type of *bipolar cell* whilst the other bipolar type is hyperolarized. This difference in cell potential is propagated throughout the whole visual system. The numbers of ON and OFF bipolar cells are approximately even [Masland, 2001, p.877]. *Bipolar cells* synapse with cones in the outer plexiform layer (OPL). Multiple bipolar cell types synapse with the dendrites of each cone forming multiple channels each propagating different variants of the cone output to the ganglion cells [Masland, 2001, p.878].

Bipolar ON and OFF cells can be further subdivided by the ability to recover from desensitization in a timely manner. Transient *bipolar cells* recover quickly whereas sustained *bipolar cells* recover slower. Moreover, transient cells propagate high-frequency and sustained cells low-frequency temporal information [Masland, 2001, p.878].

3.2.4 Amacrine Cells

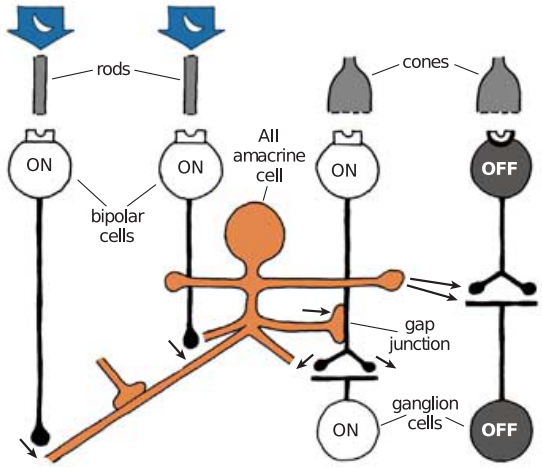

Figure 6: Connectivity of amacrine cells, mediating cone (day) and rod (night) vision onto the same ganglion cell output [Kolb, 2004].

Amacrine cells in the INL exist in large variability (29 Types have been identified). Each particular type has specific functional aspects in shaping the output of retinal ganglion cells (RGC), e.g. the temporal coordination of the firing of RGC action potentials [Masland, 2001, p.881].

Amacrine cells mediate the integration of rod and cone signals into a single RGC output (see fig. 6). The rod circuity appears to be patched upon already present cone circuity [Masland, 2001, p.879] reusing its late intra-retinal stages of processing [Kolb, 2004, p.34].

Moreover, *amacrine cells* synapse with bipolar cells at various levels in the inner plexiform layer (IPL). Additionally *amacrine cells* also synapse inhibitory on the axon terminals of bipolar cells [Masland, 2001, p.881].

3.2.5 Ganglion Cells

17 classes of primate RGC have been identified. The five major cell types are: ON and OFF *parasol*, ON and OFF *midget*, and *small bistratified* cells which collectively account for 75% of all RGC [Field et al., 2010, p.2]. However, in total 11 distinct classes (see fig. 8) can be counted if ON and OFF subtypes are paired into into particular functional pathways [Field and Chichilnisky, 2007, p.5].

Neurons in the ganglion cell layer receive input from amacrine cells in the IPL. Their dendritic connections to numerous axons of various amacrine cells is called arborization (branching). The term stratification describes the location and extent of the cells dendritic arborization in the IPL. Monostratified means that the RGC dendrites only stratifie around one location in the IPL. Bistratified ganglion cells by contrast exhibit two distinct layers of dendritic arborization [Field and Chichilnisky, 2007, p.5]. For example, *midget* and *parasol* ON RGC types stratifie low, whereas their particular OFF types stratifie high in the IPL [Field and Chichilnisky, 2007, p.5].

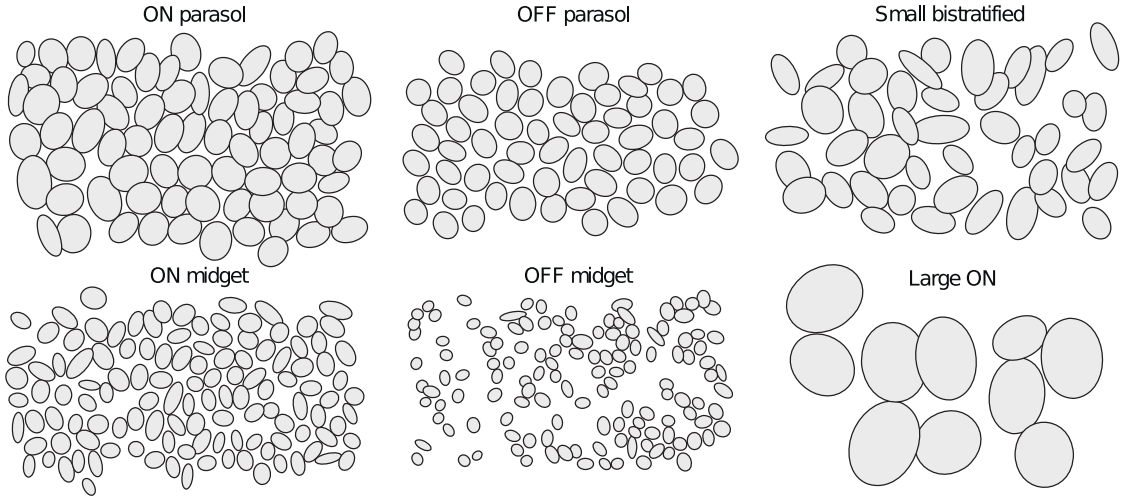

Figure 7: The figure shows the RF mosaics of six RGC types in the primate retina [Field and Chichilnisky, 2007, p.9]. Each ellipse shows the 1.3 SD contour of a Gaussian fit to the RF of a single cell.

Distinct RGC types are classified by their morphology, dendritic connectivity and to which regions their axons project (e.g. LGN, superior colliculus, ...). It has been found that in each class of morphological distinct RGC types the RF cover the entire retinal space uniformly with constant overlap (in some cases no overlap). This uniform mosaic structure (see fig. 3.2.5) is assumed to enable a regular sampling of the visual field [Field and Chichilnisky, 2007, p.8].

Midget Retinal Ganglion Cells *Midget* RGC, also named sustained-, tonic- or P-cells (as their axons terminate in the parvocellular layer of the LGN), have very small RF centers, cell somas and respond to light stimuli in a continuing manner. The conduction velocity of midget optic fibers is slow (~2 m/s). RF center responses of *midget* RGC are spectrally selective to red and green light-

stimuli, aligning with its particular cone cells type. Though debated, whether blue cone signals also contribute to *midget* RGC input, it has been found that OFF-*midget* cells are receiving input from blue-cone cells [Field et al., 2010, p.11]. Interestingly though, no blue-cone input has been found in the RF center of neither ON or OFF-*midget* RGC [Field et al., 2010, p.11].

The surround responses of *midget* RGC are also spectrally selective and are generated exclusively by the spectral types of cones not contributing to the center response [Nelson, 2007, p.11]. Moreover, *midget*-cells are responsible for the high visual acuity in the central visual field [Krüger et al., 2010, p.4].

Parasol Retinal Ganglion Cells *Parasol* RGC, also named transient-, phasic- or M-cells (fibers are terminate in the magnocellular LGN pathway), on the contrary, exhibit faster conduction velocities (~4 m/s) and respond to light stimuli transiently. Red, green and to some extent, blue cone signals are indiscriminately combined in both the center and opponent-surround areas of *parasol* RGC RF [Field et al., 2010, p.11]. Thus *parasol* RGC respond in a wider sense to luminous stimuli and are not spectrally selective [Nelson, 2007, p.11]. Some *parasol* RGC are attributed to the detection of motion, in contrast to the general belief of retinal inter-neurons not being directionally selective [Gouras, 2009, p.3]. Additionally *parasol* RGC exhibit larger cell soma and thicker axons compared to *midget* RG. The axon cross section area of *parasol* RGC is roughly twice as large as that of *midget* RGC (axon diameter parasol: 1.3mu; axon diameter midget: 0.9 mu) [Walsh et al., 1999, p.]. Moreover, near the fovea, *parasol* RGC exhibit two to three times bigger RF than *midget* RGC. Moreover in peripheral areas, the *parasol* RF may be up to 10 times larger than *midget* RGC [Nelson, 2007, p.11].

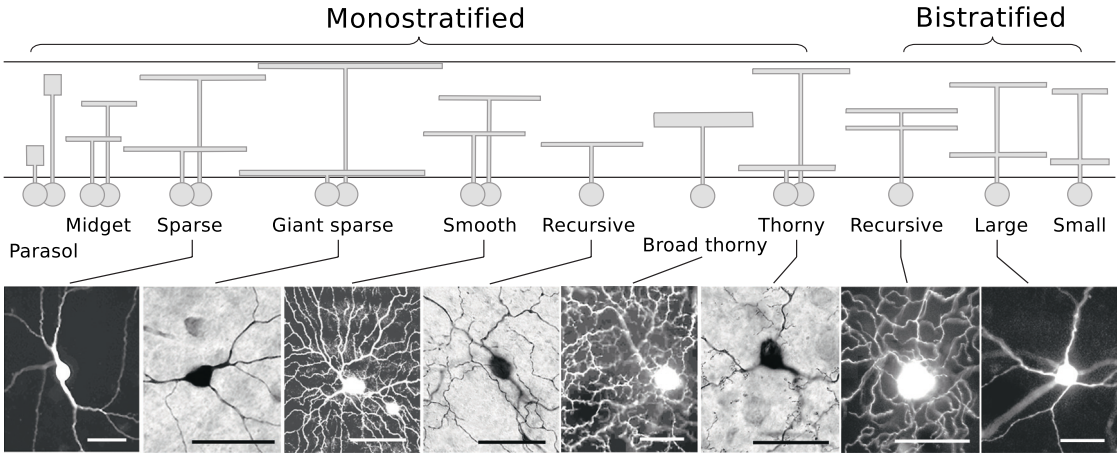

Figure 8: The figure shows 11 morphological distinct RGC cell types, each with differing extent and dendritic connectivity in the IPL [Field and Chichilnisky, 2007].

Bistratified Retinal Ganglion Cells The morphology of the small *bistratifie* RGC is set apart by branching low as well as high in the IPL and its fibers terminate in the koniocellular layer of the LGN. Still unlike midget and parasol the *small*

bistratifie RGC does not exhibit an opposite sign counterpart. Small *bistratifie* RGC respond to yellow-ON center and an blue-OFF surround spectral signature [Field and Chichilnisky, 2007, p.8].

Other Retinal Ganglion Cell Types [Field and Chichilnisky, 2007, p.6] states that With a few exceptions the response properties of more recently discovered RGC types are unknown. An exception is the *large bistratifie* RGC exhibiting a similar response signature as the *small bistratifie* but its spatial connectivity and morpholgy is different. Another notable exception is the *giant sparse* RGC which occupies large dendritic and receptive fields, resulting in a very slow, overall weak light response. Its function is attributed (based on findings in rodent retinas) to modulate circadian rhythms and pupil size [Field and Chichilnisky, 2007, p.6].

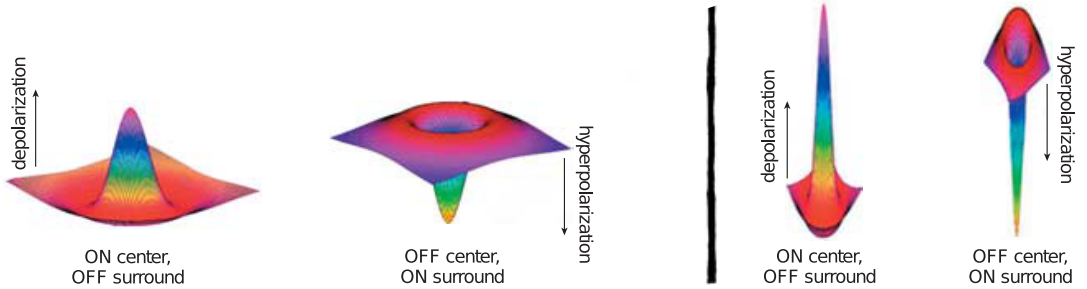

Figure 9: Texture of ON and OFF receptive fields of RGC (after [Kolb, 2004]). The spatial organization of most RGC receptive fields is commonly modeled by a difference of Gaussians function [Nelson, 2007, p.5] [Krüger et al., 2010, p.6] in which an excitatory center has an inhibitory surround, or vice versa.

3.3 Variability of the Retina

Due to the fact that biological systems have evolved over long periods of time, adapting and evolving slowly to changing environments, biological systems exhibit a large variability in-between particular individuals of a species. In contrast, the rigor of designed and engineered systems does not allow for this variability.

3.3.1 Red and Green Cone Diversity

The distribution of red and green cones in the human fovea shows a great variation [Kolb, 2012, p.8] amongst individuals. Some individuals having an equal distribution of red and green cones, others have more red than green cones, up to a ratio 16 to 1. Moreover the spatial distribution of red and green cones is of irregular and patchwork nature [Kolb, 2012, p.9].

3.3.2 Functional Variability of Blue vs. Red+Green Cone Circuity

Blue cones which amount to 15 % of all cones, reveal a rather simple [Masland, 2001, p.883] circuity, driving an exclusive type of bipolar cells, which in turn directly synapse to OFF midget cells [Field et al., 2010, p.3] (and far less often to ON

midget and parasol cells), bypassing [Field and Chichilnisky, 2007, p.4] the amacrine circuity altogether. However, some RGC exist that compare the input from blue and longer wavelength cones [Masland, 2001, p.883], blurring the clear line between short and longer wavelength pathways.

Red and green cones by contrast form a rather complicated circuity, driving numerous bipolar cell types which branch at different levels of the IPL [Masland, 2001, p.879]. The morphological differences in blue- and on the opposite red- and green-cone pathways is indicating the early development of retinal short wavelength circuity [Field and Chichilnisky, 2007, p.4] before some longer wavelength sensitive circuity evolved. Moreover, the analysis of vertebrate photoreceptors and the corresponding photo-pigments [Ebrey and Koutalos, 2001, p.55] show that short wavelength pigments predates the development of other longer wavelength sensitive pigments [Masland, 2001, p.877].

3.4 Functional Aspects of the Retina

From an information theoretic perspective the retina preprocesses visual information before it reaches the visual cortex. Clearly identifiable functional aspects are:

- **compression**: information of 110 million rod and 6.4 million cone photoreceptors is reduced to one million retinal ganglion cell fibers connecting to the brain [Weber and Triesch, 2009, p.75]. Moreover, compression of the input signal reduces the sensitivity of the receiving neural structures to over-fitting, a concept known from the machine learning domain. The visual cortex would probably learn patterns and correlations of the input data which exist solely due to noise and redundancy, if the visual information would be reaching the visual cortex uncompressed [Krizhevsky et al., 2012, p.5].

- **contrast enhancement**: [Kolb, 2004, p.30] stated that vision depends on perceiving the contrast (difference of light and dark phases of a stimulus) between image components and their backgrounds. It has been observed that a particular RGC is tuned to a specific contrast [Nelson, 2007, p.5] of the visual stream. Subsequently the contrast selectivity of RGC together with the de-correlation in ON and OFF channels allows later stages of neural computation (visual cortex) to define precise edges by further processing the separate ON and OFF information.

- **seperation of visual input stream**: The existence of ON and OFF RGC subtypes, which do not exist in the photoreceptors, shows separation of the input signal. Moreover the retina likely separates more aspects of visual information than solely ON and OFF contrast. A strong indicator is the mosaic organization of RGC receptive fields in which the population of each particular cell-type covers the entire visual field [Field and Chichilnisky, 2007, p.9]. Under the assumption of an efficient neural coding of visual stimuli [Doi et al., 2012, p.16261], only two RGC types covering the entire visual

field would probably exist, if the retina would separate ON and OFF aspects exclusively.

- **noise reduction**: Furthermore, the band-pass (whitening) filtering characteristics [Olshausen and Field, 1997, p.3318] of RGC center surround receptive fields (see fig. 9) reduces noise by filtering high (and low) frequency portions of the signal. Modeling macaque LGN single surround cells showed that filtering characteristics of achromatic stimuli where band-pass whilst low-pass for chromatic stimuli [Shapley and Hawken, 2011, p.709].

3.5 LGN and Parallel Pathways

The preprocessed parallel output of retinal ganglion cells is reorganized in separate layers (see fig. 10) in the LGN [Gouras, 2009, p.10], from which the visual information is relayed to the V1 region of the visual cortex [Schmolesky, 2007, p.9]. Cells in the LGN are organized *retinoptically*, meaning that the spatial positions of the retinal ganglion cells RF are preserved within the organization of cells in the LGN: Neurons whose RF are located in / near the fovea are located in the posterior LGN, whereby the more a RF of a neuron is located in the periphery of the retina the more is the neuron located in the anterior LGN [Schmolesky, 2007, p.10].

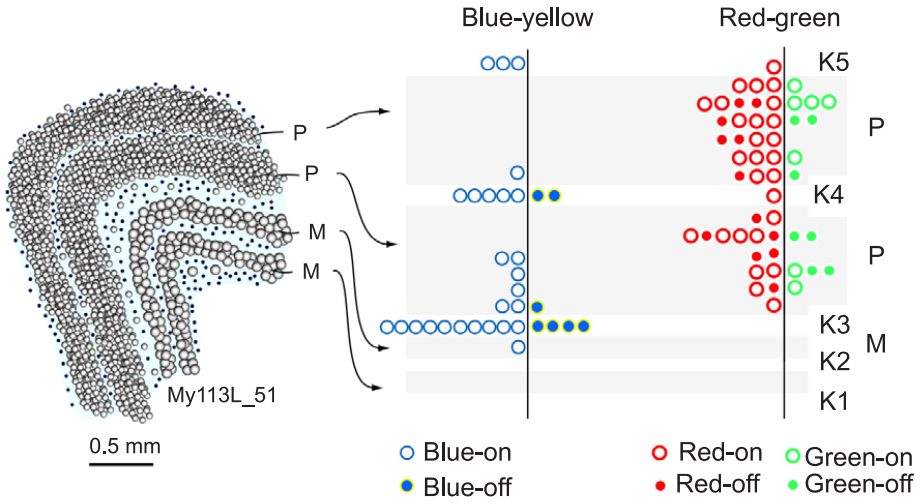

Figure 10: Transmission of color signals through LGN in marmoset monkey, from [Martin and Grünert, 2013, p.156]. Left, the arrangement of *midget-parvocellular* (P), *parasol-magnocellular* (M) and *bistratified-koniocellular* (K, light blue) layers. Right, flattened schematic of the LGN showing a histogram of RF positions of LGN cells which are directly synapsing on axons of RGC, indicating number and type of *midget* RGC. Achromatic information can be conceived of as summing the cone inputs red + green + blue [Shapley and Hawken, 2011, p.711].

Three distinct pathways from the retina to the cortex have been identified in the visual system : the *magnocellular* (M-) and *parvocellular* (P-) stream have been identified first [Krüger et al., 2010, p.5]. Later, the *koniocellular* pathway has been found, interfacing with the lower regions of the visual cortex more directly

and in parallel to the M- and P-stream [Shapley and Hawken, 2011, p.704]. The *parvocellular* pathway contains information from the L and M-cone driven retinal cells whereas the *koniocellular* pathway transports mostly information from the S-cone system [Gouras, 2009, p.11].

In the visual cortex the P- and M-pathways continue to be noticeable. The M-pathway is believed to provide information to areas in the visual cortex that are responsible for the perception of motion and sudden changes, whereas the P-path provides information to areas associated with shape and object recognition [Krüger et al., 2010, p.5].

3.6 Visual Cortex

Roughly three areas can be identified in the visual cortex: the *occipital* part and the *ventral* and *dorsal* pathways, each containing numerous sub cortices. Output of the occipital part is passed to the dorsal and ventral pathways. The sub-cortices in the dorsal pathway are concerned with integrating vestibular information (eye, arm, head positions) with the visual information. Ventral sub-cortices are involved in object-recognition and are further projecting information to the hippocampus-memory and the prefrontal cortex [Krüger et al., 2010, p.4].

Cell populations in the visual cortex are organized in a deep hierarchy which is processing visual sensory information over as much as 10 levels, excluding pre-cortical processing [Felleman and Essen, 1991, p.1]. (Notably, the latency of the visual signal increases with each level by approximately 10 ms [Krüger et al., 2010, p.6]) Generally the RF sizes of a neuron in the visual cortex are gradually increasing whilst image features of increasing complexity and coverage of the visual field are extracted from the image stream, the higher the neuron resides in the visual cortex. Neurons of the lower levels of the visual cortex extract simple, more general image features (e.g. color, orientation, motion, shapes) whilst the receptive fields (RF) cover small local areas of the visual field. In contrast, neurons higher in the visual cortex have larger RF and are extracting more complex, less general, image features (e.g. objects, gestures) of the already processed sensory stream [Felleman and Essen, 1991, p.2].

3.6.1 Occipital Cortex

The *occipital* or *striate* cortex can be further divided into five hierarchical interconnected regions: V1 to V5 (V5 also named MT due to its high motion sensitiveness). The functional role of the lower parts (V1 and V2) of the occipital cortex is the general representation of scenes [Krüger et al., 2010, p.18]. These general features are utilized by areas located higher in the processing hierarchy, e.g. V3/V4 and MT. Sharing general features among more specialized cortices is reflected by diverging sizes of areas in the occipital cortex: V1 and V2 areas have the largest cell population in the occipital cortex [Krüger et al., 2010, p.4], in contrast to the rather smaller sizes of the remaining cortices.

Areas in the visual cortex are highly interconnected, *feedforward* and *feedback* connections between and extensive *lateral* connections within areas have been observed [Schmolesky, 2007, p.9]. Direct feedforward projections from V1 are extending to V2, V3 and MT, whereat V2, V3, V4, and MT are projecting directly in a *feedback* manner to V1. Additionally, V1 projects a direct *feedback* to LGN. [Schmolesky, 2007, p.9].

3.6.2 V1

In the V1 area, cells are organized *retinoptically* [Schmolesky, 2007, p.10], producing a distorted map of the visual field dominated in large amounts by the fovea [Adams and Horton, 2003, p.3777]. *Cortical columns* of cells can be observed extending from upper layers to the lower layers of the cortex. In the literature, *cortical columns* are defined upon anatomical or functional features, such as selectivity to a specific direction or its ocular dominance [Schmolesky, 2007, p.8].

A column receives signals from the retina of one eye, whereas its neighbor columns receive signals from the other eye, resulting in an alternating pattern [Gouras, 2009, p.10]. An ocular dominated *cortical column* contains a regular array of *cytochrome oxidase* (CO) patches in which cells are found to be sensitive to stimuli of specific orientations [Adams and Horton, 2003, p.3778]. Each CO patch receives input from chromatic selective cells (via LGN parvo- and koniocellular pathways) whilst the area surrounding the patch processes achromatic contrast (receiving input via LGN magnocellular pathway) [Gouras, 2009, p.11].

Additionally, a significant portion of the cell population of V1 are found to be responding significantly to the stimuli from both eyes (binocular). Binocular cell populations can utilize the degree of non corresponding images from both retinas, named *binocular disparity*, to e.g. measure object object distances [Schmolesky, 2007, p.11].

However, it is debated if distinct feature maps might exist in the (macaque) V1 region as color, orientation and edge-polarity appears to be multiplexed in cortical signals [Shapley and Hawken, 2011, p.707]. This is supported by the findings that in the upper layers of V1 about 60% of color selective cells are not orientation selective [Shapley and Hawken, 2011, p.707]. And that only 25%-30% of the whole V1 cell population show a strong selectivity to orientation [Schmolesky, 2007, p.10], whereas the majority shows some sort of orientation selectiveness.

4 Color Vision

The perception of color is an illusion produced by the interaction of millions of neurons in the visual system. Subsequently, no color exists in the outside world, but light in varying wavelength (frequency) and intensity (energy). The visual system is assumed to create the illusion of color by separation and recombination of energy and frequency properties of light stimuli [Gouras, 2009, p.1].

Intuitively, color is perceived as a phenomenon separate from shape and motion. For example: one can understand the action of a black white movie without perceiving any color [Shapley and Hawken, 2011, p.701].

Nevertheless, newer findings from psychophysical studies suggest the perception of color is closely linked to the perception of shape (form) [Gouras, 2009, p.1]. This is supported by biological findings that chromatic signals in the visual cortex are carrying spatial information. Subsequently, psychophysical experiments found that geometric illusions can be perceived solely from red/green signals [Shapley and Hawken, 2011, p.702]. Reciprocally, form influences the perception of color: a strong influence of *surrounding* regions on the perception of a color in a target region, an effect known for centuries. In some cases, the spectral contrast at the boundaries of a region has a stronger influence on the color perception of that region as its own spectral reflectance [Shapley and Hawken, 2011, p.702].

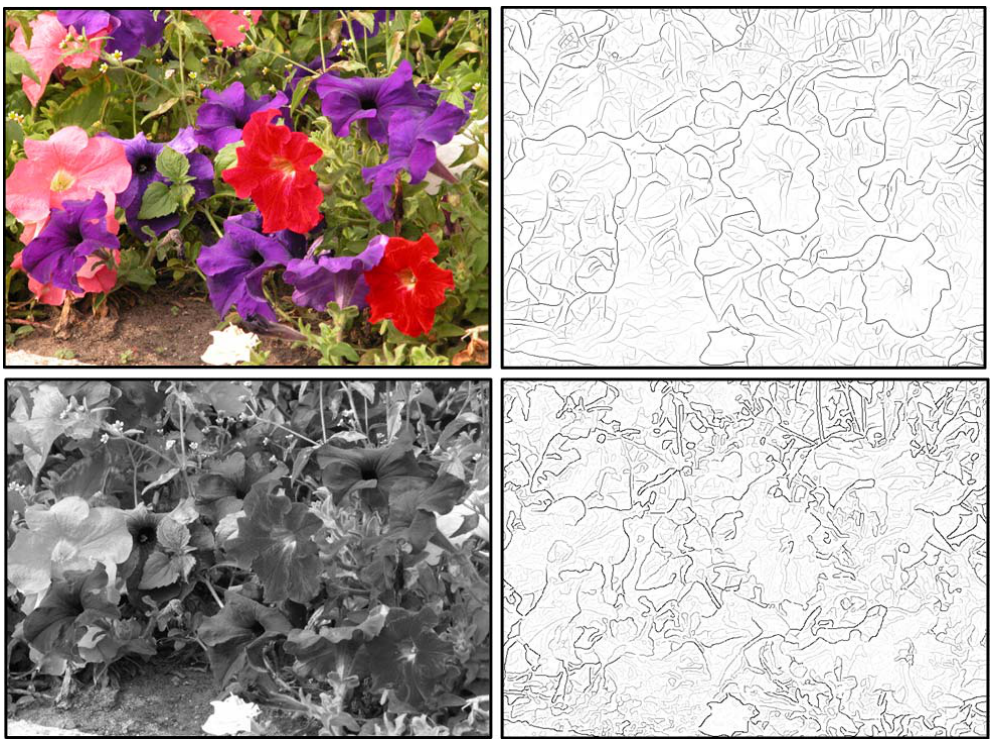

Figure 11: Image segmentation upon chromatic contrast information (above) results in far better detection of actual object edges. In contrast results upon luminosity data (below) [Yang et al., 2013].

4.1 Color Vision as Evolutionary Advantage

The question of why color vision appears to be an evolutionary advantage can be motivated accordingly: An object can reflect the same amount of energy as its background, but it is rarely reflecting the same composition of different wavelengths [Gouras, 2009, p.2] as its background. Thus a chromatic (spectral) contrast, provides better information to, for example boundary detection of surrounding objects, than a luminosity (energy) contrast alone (see fig. 11). The estimation of a surface reflective properties independent of its illumination, clearly constitutes an evolutionary advantage [Shapley and Hawken, 2011, p.701].

Accordingly, a combination of a short wavelength cone (blue) with one or more long wavelength sensitive cone types (red+green) appears to be a universal property of most mammalian retinas [Masland, 2001, p.879]. Moreover, retinas of some bird and fish species contain up to five different types of cones [Ebrey and Koutalos, 2001, p.50], suggesting these species are experiencing their environment in richer or finer resolved color.

4.2 Chromatic Contrast (Color Opponency)

[Gegenfurtner, 2003, p.563] states the chromatic contrast "removes the inherently high correlations in the signals of cone types" and as such frees the visual cortex of considering these less informative parts of in the visual stream. In order to establish chromatic contrast a visual system has to meet at least two requirements:

1. Two or more types of cone cells are needed. Each cone type must be sensitive to distinct parts of the wavelength spectrum, in which a particular cone type clearly responds stronger than the other cone types (see fig. 3).
2. Each specific cone type and its corresponding circuity must span the whole visual field, since the response of a group of cones of one type has to be compared with the response of another type within the same area [Gouras, 2009, p.3].

By pitting the signals of different cone types against one another the neural circuity of the retina establishes chromatic contrast [Gouras, 2009, p.2] (see fig. 12). Luminous (intensity) contrast occurs between positive and negative parts of an input which is not discriminating between different cone types. However, chromatic contrast, being the result of three different cone types, can be separated into two classes:

- **intra-channel** contrast resulting solely of ON and OFF parts of a single cone type. Thus, for red (L), green (M) and blue (S) cones we have L/-L, M/-M and S/-S contrasts. If viewed in RGB color-space: a vector of red $[\mathsf{a}, 0, 0]$ has a counterpart of cyan $[-\mathsf{a}, 0, 0]$, analogous for green and blue.
- **inter-channel** (or color opponent) contrast resulting of comparing the ON and OFF parts of different cone types. If viewed in RGB color-space: a vector of red $[\mathsf{a}, 0, 0]$ has a magenta $[0, -\mathsf{a}, 0]$ counterpart and a green vector

$[0, a, 0]$ has cyan opponent $[-a, 0, 0]$. Of possible combinations of contrast between ON and OFF signals of different cone types (such as L/-M, -L/M, L/-S, -L/S, M/-S and -M/S), the majority of *midget* RGC are selective to L/-M and -L/M cone opponent and *inter-channel* S/-S contrast. Other cone opponent contrasts are less commonly found in primate RGC (see table 9 Appendix).

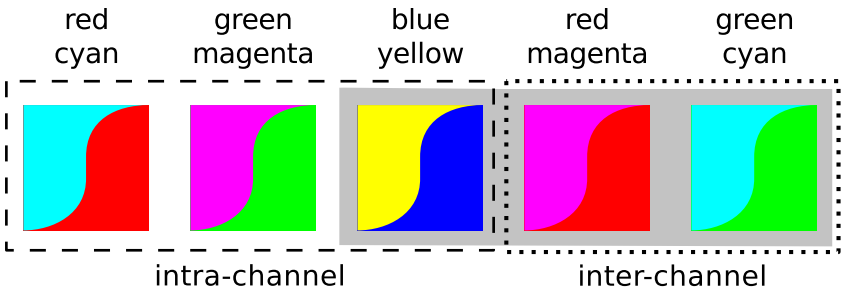

Figure 12: Different types of chromatic contrast. Inside the dashed line: *intra-channel* contrast of the red, green and blue axis of the RGB color-space. Inside the the dotted line: *inter-channel*, biologically motivated cone opponent, contrast of red/magenta and green/cyan. Inside the gray box: chromatic contrast for which *midget* RGC are selective. Locating yellow-blue contrast as of *intra-channel* type is correct if viewed from the perspective of RGB images. However biologically, yellow appears to be the result of multiple sources. Such as combining M (green) and L (red) cone signals or resulting of S (blue) OFF signals.

4.3 Single Color Opponency

Cells providing chromatic-contrast, also referred to as single color-opponent cells, are found in the retinal ganglion layer, the LGN and the V1 area of the visual cortex. RGC respond to localized visual stimuli of small illuminated (ON) or dark (OFF) spots on a background of opponent color (for example a yellow spot on a blue background). V1 *single color-opponent* cells responds to large areas of color as well as to the interiors of large color patches [Shapley and Hawken, 2011, p.704/705]. Single color opponency is regarded as basic building block for color related feature detection tasks like boundary detection [Yang et al., 2013, p.2810].

In the retinal ganglion layer *midget* RGC are chromatically selective and are found in ON and OFF varieties of red, green and blue RF centers with a color opponent surround (see fig. 9 and 10). Thus *midget* RGC exist in: (red-ON, green-OFF), (green-ON, red-OFF), (blue-ON, blue-OFF) and conversely with OFF centers and ON surround parts.

A V1 *single color-opponent* cell receives opponent ON/OFF input from two ore more cones their corresponding retinal circuity (cone opponency). Furthermore, it has been found that approximately 60 % of color sensitive V1 cells are orientation selective. In fact, approximate orientation selectiveness has been observed to the same magnitude in color and non-color sensitive V1 cells [Shapley and Hawken, 2011, p.707]. V1 *single color-opponent* cells [Shapley and Hawken, 2011, p.709] establish red/green and blue/yellow color-opponency.

4.4 Double Color Opponency

The *double color-opponent* cells located in the V1 cytochrome oxidase (CO) blobs [Shapley and Hawken, 2011, p.706] exhibit cone opponency, like single color-opponent cells, though additionally are receiving single opponent inputs at different locations in the neuron's RF (spatial opponency) [Krüger et al., 2010, p.8]. This renders V1 *double color-opponent* cells responding strongly to color patterns (e.g. bars) but non-responsive to stimuli covering the whole visual field, with low spatial frequency or shallow colors-gradients [Shapley and Hawken, 2011, p.704].

4.5 Color Constancy

An object can reflect locally identical spectral components, but the light reaching an observer is influenced by the entire scene and thus deviates in some cases drastically from its origin (see fig. 13). The visual system minimizes this effect by its property of color constancy, which is described by [Gouras, 2009, p.6] as to "see colors as unchanged even when there are large changes in the spectral properties of an illuminant". However, the color-constancy the human visual system achieves leads to color-illusions (see fig 13).

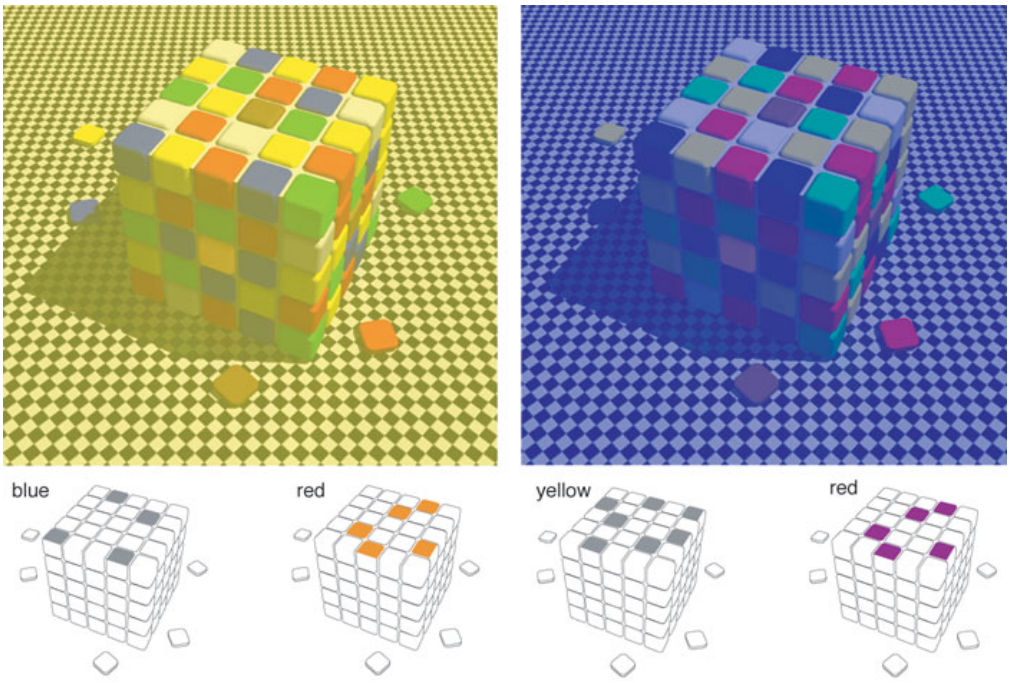

Figure 13: From [Purves et al., 2002]: Discrepancy in the perception of color under different global illumination of yellow (left cube) and blue color (right cube). In the four diagrams below, tiles of interest are displayed in their actual color, together with the color in which the tiles appear (as text, above the diagram) in the illuminated scene. Thus for example, gray tiles appear to be of blue color in a yellow illuminated scene (left side). Remarkably, tiles of the same gray color appear to be of yellow color in a blue illuminated scene (right side).

5 Related Work

This section will introduce (linear) feed-forward neural networks briefly, motivate the unsupervised training on image data, describe the effects of constraining the input reconstruction capabilities of auto-encoders and motivates their usage in approaching information theoretic questions of the formation of localized feature detectors in the lower visual system. Furthermore, a brief overview on unsupervised models of RGC and V1 feature detectors, as well as on models and methods considering chromatic information is given.

5.1 Linear Feed Forward Neuronal Networks

Neuronal networks represent "circuits of highly interconnected units with modifiable interconnection weights" [Baldi and Hornik, 1989, p.53] which can be organized and trained in numerous ways. We will restrict us to describe layered feed-forward networks with linear units. Linear units are the simplest form of a computational unit in a neural network. Due to its limitation of solely computing linear functions, and its property that multiple layers of linear neurons can always be collapsed into a single layer of equivalent function, linear units have been traditionally considered uninteresting compared to non-linear units [Baldi and Hornik, 1989, p.53].

However, using non-linear units as outputs in case of training an auto-encoder network (see 5.4) is of no benefit since "the network is trying to approximate a linear map: the identity function" [Baldi and Homik, 1995, p.843]. Hence, non-linear units are not in the focus of this work.

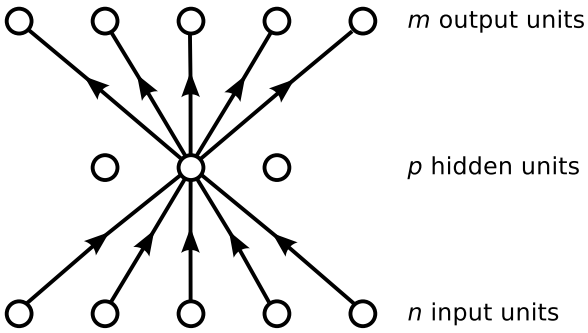

Figure 14: Linear feed forward Network with one hidden Layer [Baldi and Hornik, 1989, p.53].

Following Baldi and Hornik, consider a linear network of $n - p - m$ architecture with n input units, p hidden units and m output units (see fig. 14). Next, consider two connection matrices of: input to hidden layer A $(p \times n)$ and hidden to output layer B $(m \times p)$ with b_{ij} the weight of the j-th hidden unit and the i-th output unit (double index: post-pre synaptic order). Thus the network computes the linear function of

$$y = F(x) \quad \text{or} \tag{1}$$

$$y = B\,A\,x \tag{2}$$

5.2 Training Strategies

Training a neural network is based on the minimization of an error function E on connection weights. Assuming T input patters x_t $(1 \leq t \leq T)$ and T corresponding

output patters y_t $(1 \leq t \leq T)$ a quadratic error E function can be defined:

$$E = \sum_t ||y_t - F(x_t)||^2 \qquad \text{euclidean norm } ||u||^2 \text{ of } u. \tag{3}$$

Generally, E is minimized in the training process by one or another algorithm implementing gradient descend. However three distinct learning strategies can be differentiated:

- **supervised learning**, x_t and y_t are explicitly given as examples to train the network
- **unsupervised learning**, x_t is given but not y_t. Instead some criteria for y_t, e.g. maximizing the output variance, is given.
- **auto associative (unsupervised) learning**, in setting $x_t = y_t$ (also implies $n = n$) the input is used as teacher, also called auto-encoding or identity mapping in the literature.

Training neural networks in a *supervised* fashion has the difficulty of assembling a sufficiently large training set in which input samples and corresponding output examples cover all relevant aspects in the input data. In some domains the enumeration of all relevant combinations of patterns in the input data is unknown, and if known the task of enumerating all aspects is not tractable. If some common criteria of output patterns can be defined, this limitation of possibly incomplete training examples can be avoided by training in a *unsupervised* fashion. Furthermore, training examples can be avoided completely if the model is trained in an *auto associative* fashion. A set of images suffices, no additional training examples or output criteria are required.

5.3 Unsupervised Learning with Images as Training Data

Luminous information of an image can be trivially encoded by a neural network (an image of $a \times b$ pixel can be fed as vector of $a * b$ units into a network with the same number of input units). Despite simple encoding of image-information, a desired feature might appear in a vast number of possible representations in the image. To just name a few: an object (feature) might be scaled, rotated, mirrored, cut and of different color or texture.

Thus training a neural network supervised has the great difficulty of assembling sufficient training data. In the literature [Bengio, 2009, p.11] states that "in order to cover the many possible variations in the function to be learned, one needs a number of examples proportional to the number of variations to be covered" which is unattractive as vast number of training examples would have to be assembled. This is also true of defining statistical output criteria beforehand if training the network unsupervised. Subsequently, in order to learn generalized image-features, training a neural network in an unsupervised, auto associative manner appears to the only tractable option.

5.4 Autoencoders

A perfect reconstruction of the input is an identity function, thus an auto-encoder will learn to approximate the identity function during the training process. The identity function usually is of no interest, but by constraining the ability of the network to reconstruct the input, interesting properties emerge. An auto-encoder can be constrained in numerous ways:

- By setting the number of hidden units to be smaller than the number of input units $\mathsf{p} < \mathsf{n}$. Such a network is referred to as being *undercomplete*.
- By imposing a weight-constraint on the connections to hidden units.
- By utilizing a non-linear transfer function, shaping the activity of the hidden units, such as rectified linear units [Krizhevsky et al., 2012, p.3] or sparsity constraints [Olshausen and Field, 1997].
- By adding noise to the input whilst the model learns to reconstruct the uncorrupted input, referred to as denoising auto-encoder [Vincent et al., 2008].

5.4.1 Undercompleteness and Compression

In case of $\mathsf{p} = \mathsf{n}$ an auto-encoder has the capability of fully reconstructing the output of all input units, since for each input there is (at least) one hidden unit. In contrast, in an *undercomplete* auto-encoder network only p/n hidden units are available to encode the input of one input unit. Thus, a single hidden unit has to reconstruct the output of more than one input unit, meaning the ability of the auto-encoder to fully reconstruct the output of all input units is reduced. An *undercomplete* auto-encoder which reconstructs the input with minimal reconstruction error has learned a compressed representation of the input data. Since the number of hidden units is smaller as the number of input units, the dimensionality of the input data is effectively reduced to the number of hidden units.

The results of an *undercomplete* auto-encoder trained to reconstruct the input data with minimal error are resembling [Vincent and Baddeley, 2003, p.1286] those of the principal component analysis (PCA) statistical method. If the T input patterns x_t, each with n characteristics, are viewed as points in an n-dimensional euclidean space, the PCA method finds directions (principal components) along the variance of the n-dimensional point cloud is maximal. These vectors of maximal variance can also be viewed as the eigenvectors of the data's covariance matrix [Baldi and Homik, 1995, p.840]. The data's principal components span a subspace of lower dimension (in case of an auto-encoder a p-dimensional subspace) which represents an approximation of the original data but of lower dimensionality.

Dimensionality reduction or compression implies the minimization of redundant features, which in turn are caused by correlating features in the input data. Since correlations organize data, these are of interest and can be identified by an undercomplete auto-encoder by learning a compressed approximation of the input with minimal reconstruction error.

5.4.2 Weight Constraints

Instead of reducing the number of hidden units, the connection weights from and to hidden units can be constrained. Thus, during the training process the network learns to optimally reconstruct the input under the constraint of minimal connection weights. For example, by imposing a constraint on the (absolute) sum of connection weights of each individual hidden unit [Vincent and Baddeley, 2003, p.1285] the resulting RF are composed of a few large weights whilst the remaining weights being of low, near zero value. Thus the allowed sum of connection weights is shared upon a few strong weights. This effect is called *localization* of the RF vector of a hidden unit.

Localization of hidden units has the effect that each unit is forced to specialize on separate aspects of the input, which can be rationalized as follows: The maximal allowed sum of all weights prevents a hidden unit from producing the values required to optimally reconstruct the input for all units of the input vector by its own. Instead, the weight-constraint only allows it to produce values of sufficient magnitude for some parts of the input vector. Hence a single hidden unit is only capable of optimally reconstructing a (local) part of the input. Since training an auto-encoder minimizes the overall reconstruction error of all hidden units, training converges in a distribution in which the burden of input-reconstruction is shared among the hidden units, each tuned to a specific aspect of the input data.

Moreover, a weight constraint is not dependable on the model being *undercomplete* and therefore also works in an overcomplete ($\mathsf{p} > \mathsf{n}$) setup, where the number of hidden units is larger than the number of input units and input compression by lack of hidden units cannot occur.

5.5 Information Theory and Redundancy Reduction

Theoretical work about "the statistical nature of the world and how the visual system is matched to the statistics of natural scenes" [Shapley and Hawken, 2011, p.715] early on imposed the question which coding principles lead to the formation of feature detectors in the early stages of our visual system. Barlow [Barlow, 1989] proposed that "edges are coincidences in images" and thus feature detectors are an end result of a redundancy reduction process. This was later refined by [Doi et al., 2012] proposing that a neural code needs some portion of redundancy to efficiently encode natural scenes (mostly in countering noise).

Since redundancy reduction and compression of information can be seen as being equal, auto-encoders trained upon images of natural scenes have been utilized in approaching these questions. The usage of auto-encoders for this task appears to be reasonable due to its simplicity and due to:

- The linearity of auto-encoders matches the overall linear information flow in the retina and LGN parts of the visual system.
- The unsupervised training of auto-encoders is biologically plausible in the light of retinal development in prenatal and newborn primates.

- Constraining auto-encoders reflects the scarce availability of (metabolic) resources in a biological system.

5.6 Autoencoders as Feature Detectors

By training an overcomplete auto-encoder with a sparsity constraint (a weight constraint and a transfer function which enforces sparse activity patterns) on whitened images Olshausen and Field [Olshausen and Field, 1997] showed self-organization of localized orientated RF similar to those of V1 simple cells. Prior, Bell and Sejnowski presented an Independent Component Analysis algorithm (ICA), which maximizes the mutual information between inputs and outputs in a square system [Bell and Sejnowski, 1997] producing equivalent results. However, matrix inversions utilized by the ICA algorithm of [Bell and Sejnowski, 1997] are biologically not plausible and are preventing the modeling of overcomplete systems ([Olshausen and Field, 1997] see: Appendix A). Modeling neurons of the visual cortex in an overcomplete network is a biologically plausible and, due to superior coding capabilities [Olshausen and Field, 1997], an information-theoretically highly desirable property, in contrast to square or undercomplete systems.

Vincent and Baddely demonstrated that by solely imposing a biological motivated weight constraint, thus maximizing information transmission whilst minimizing the metabolic costs, center-surround RF emerge [Vincent and Baddeley, 2003] similar to those found in RGC. They subsequently extended their model with a "cortical" layer incorporating sparse coding to model V1 simple cells. In this model image patches where sampled of a space variant grid resembling the fovea [Vincent et al., 2005]. A space variant grid resembling the fovea is also modeled by Doi et al [Doi et al., 2012] which also utilizes the ICA algorithm of Bell and Sejnowski.

More elaborate refinements of modeling the self organization of V1 simple cell RF have been presented more recently by Weber and Triesch in 2008. Their model [Weber and Triesch, 2008] was able to explain the tilt after effect (TAE). Both models of Olshausen and Field and Weber and Triesch are operating on preprocessed whitened images, which imitate the output of RGC [Krüger et al., 2010].

5.7 Models of Chromatic Feature Detectors

While color has not been in the focus of research, a few exceptions have been accomplished: Dharmesh, Leif and Buchsbaum demonstrated by applying an ICA algorithm that color-opponent receptive fields (among other types) evolved from learning the statistics of color images [Tailor et al., 2000]. Doi et al trained a 3-layer network of photoreceptors, RGC and V1 with color images sampled from a mosaic structure of photoreceptors, showing the emergence of luminosity, red-green- and yellow-blue-difference pathways at the decorrelation stage [Doi et al., 2012]. Both of these models however, are biologically implausible in consequence of the deployed ICA algorithm of Bell and Sejnowski. Just recently, [Brown et al., 2011]

derived convolutional filters of statistics of color images, resulting in color-opponent lateral inhibition filters, avoiding matrix inversion / decomposition techniques.

In contrast numerous models of chromatic feature detectors have been accomplished in which the detectors are constructed rather than learned or derived by statistical means. Recently, Wu and Wei constructed a model [Wei and Wu, 2013] of chromatically selective RGC performing some sort of chromatic contrast. Further, Kaifu Yang et al constructed a model [Yang et al., 2013] of RGC/LGN and V1, successfully demonstrating boundary detection upon chromatic contrast information.

5.8 Models of Computational Color Constancy

Though related, but not in the focus of this work, is the property of the visual system to provide color constancy. Still, its exact underlying neural coding principles are unknown and yet to be discovered. As such, computational models of the lower parts of the visual system simulating color constancy do not exist yet. However, in regarding color constancy as the computational task of "estimating the true reflectance of object surfaces of an image", Bayesian models have been applied quite successfully [Gehler et al., 2008] under the assumption of an uniformly illuminated scene.

6 Model

The effect of minimizing metabolic costs while learning optimal filters to represent natural scenes had been explored by [Vincent and Baddeley, 2003] in utilizing a simplified linear auto-encoder. The model trivially consists of a single hidden layer, whereas units in the visible layer functioning as photoreceptors feed image patches of natural scenes to the hidden layer. In absence of a non-linear transfer function [Vincent and Baddeley, 2003, p.1285] showed that by solely constraining the connection weights of the two layers, localized DOG shaped receptive fields emerge, resembling those of RGC.

6.1 Model in Relation to Biology

However simplicity comes with inaccuracy in terms of biological morphology, but not necessarily with inaccuracy in terms of biological functionality. The following (functional) aspects of biological retinas are not considered in the generative model:

- Only neurons in the fovea center are simulated, spatially evenly spaced in a lattice (pixels of an image), ignoring the mosaic organization of photoreceptor cells in the fovea.
- The model does not reflect any temporal aspects of retinal neural activity. Connection weights are equated to synaptic strengths and the models output to neural firing rates of RGC action potentials [Vincent and Baddeley, 2003, p. 1283].
- Furthermore, the model only considers *cone* driven daylight vision, thus excluding *rods* which are non existent in the fovea. Correspondingly the model was trained solely with images of natural scenes at daylight.
- It appears to be reasonable that the model ignores *amacrine cells* since its attributed functionalities are not needed: *Amacrine cells* are attributed to be temporally coordinating the firing of RGC action potentials and are mediating the integration of *rod* and *cone* signals.
- The linear nature of the model, (no non-linear transfer function exists) does not consider lateral inhibition of photoreceptors by *horizontal* cells. Although neurons acting in a quasi linear manner seem biologically plausible: [Shapley and Hawken, 2011, p.711] states that non-linearities cause simulated RGC responses to edge and grating stimuli to fail to match predictions from receptive field maps.
- The functional interaction and circuity in the IPL between *horizontal*, *amacrine* cells and RGC in the retina is not considered altogether.

6.2 Generative Model of Retinal Ganglion Cells

The model of [Vincent and Baddeley, 2003] is slightly extended: instead of processing gray scale images, the model processes RGB images (see Extension to Color).

Further, the original model applied the weight constraint to a RF only if the absolute weights of this RF were above a specific threshold, leaving the RF untouched otherwise. We removed the threshold altogether, applying the constraint on each RF in each training step.

Apart from the number of visible and hidden units, the model has only three parameters: k strength and p shape of the metabolic weight-constraint and the η learning rate. Parameters are described in the following section in greater detail.

The algorithm for learning the connection weights W in the generative model (see fig. 15) takes five steps: For an input vector x and matrix W, the hidden activity (4) resp. (9) is computed. The input is reconstructed (5) and the reconstruction error (6) is used in the Hebbian learning rule (7), updating W. Afterwards the weight constraint (8) is applied to the receptive field (RF) of each hidden neuron i.

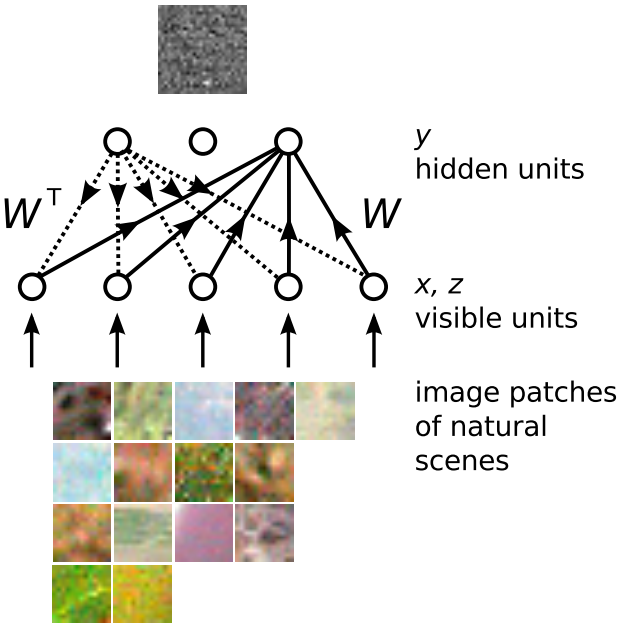

Figure 15: The RGC model. Solid arrows indicate connections from visible to the hidden units. Dotted arrows indicate input reconstruction connections from hidden to visible units.

$$\mathsf{y} = \mathsf{f}(\mathsf{x}) \tag{4}$$

$$\mathsf{z} = \mathsf{W}^\mathsf{T} \cdot \mathsf{y} \tag{5}$$

$$\mathsf{e} = \mathsf{x} - \mathsf{z} \tag{6}$$

$$\Delta\mathsf{W} = \eta\, \mathsf{y} \odot \mathsf{e} \tag{7}$$

$$\Delta\mathsf{W}_\mathsf{i} = \eta\, (-\mathsf{k}\, \mathsf{sgn}(\mathsf{W}_\mathsf{i})\, \mathsf{abs}(\mathsf{W}_\mathsf{i})^\mathsf{p}) \tag{8}$$

$$\mathsf{f}(\mathsf{x}) = \begin{cases} \mathsf{W} \cdot \mathsf{x} & \text{unconstrained} \\ \mathsf{max}(0, \mathsf{W} \cdot \mathsf{x}) & \text{clipped} \end{cases} \tag{9}$$

(with: $\cdot$ inner, $\odot$ outer product)
Notably, (7) changes in a different setting to only permit positive hidden layer activations, effectively introducing a non-linear transfer function: a rectified linear unit *ReLU* [Krizhevsky et al., 2012, p.3].

6.3 Properties of the RGC Model

Despite the model being extremely simple, subtle changes in the parameters or the input statistics can result in vastly different distribution of RF. While training, the overall reconstruction error is minimized, though as the resulting spatial properties and distribution of RF are of interest, the reconstruction error alone is not

informative. We note that convergence in this case here means that the shape of the resulting RF has stabilized (which naturally only can happen if reconstruction error is minimal).

6.3.1 Input

The input of the model is a vector of m units which is derived of a patch cut from a training image. It is created by the following steps:

1. A random image is selected, from this image a patch of fixed dimension and of random location is cut. To avoid unnatural image statistics patches are never cut near the borders of an image.
2. In $\mathsf{p}\%$ chance the patch is flipped, avoiding unwanted statistics of vertical lines (e.g. trees, grass).
3. The values of the patch are normalized.
4. In some settings of the model, the mean of the patch is subtracted of the patches individual values.
5. The patch is vectorized, its two dimensional structure is transformed to a flat vector (see fig. 16). A square $(\mathsf{n} \times \mathsf{n})$ patch of n pixels size results in a flattened vector of at least $\mathsf{m} = \mathsf{n}^2$ units.

6.3.2 Extension to Color

The implementation of the model is capable to run the training process in several *modes* (see table 1). The original model [Vincent and Baddeley, 2003, p.1286] processed gray-scale images of natural scenes. Thus, the input consisted of scalar pixel values.

mode	num. input channels	$[\mathsf{r}, \mathsf{g}, \mathsf{b}]$-pixel value used
luminosity	1	$[(\mathsf{r} + \mathsf{b} + \mathsf{g})/3]$
red vs. green	2	$[\mathsf{r}, \mathsf{b}]$
red + green vs. blue	2	$[(\mathsf{r} + \mathsf{g})/2, \mathsf{b}]$
red green blue	3	$[\mathsf{r}, \mathsf{b}, \mathsf{g}]$

Table 1: For testing and debugging purposes the model can be trained in different *modes*. A *mode* defines which values of an input RGB pixel are used and therefore the number of channels. This affects of how the contents of input and hidden vectors and the connection matrix are interpreted.
In third column, the surrounding square brackets indicate a vector, which components are separated by commas.

In order to operate on non-scalar pixel values, the elements of RGB pixel vectors have to be flattened into a single vector. This is possible in two ways: by concatenating each pixel vector into a flat vector (column-wise), or by concatenating all red, followed by all green and lastly all blue pixel values (row-wise). We

chose to extend the model in a row-wise manner, meaning the entire intensities of each RGB input channel are concatenated (see fig. 16). The result is a flat vector in which the first m units contain information from the first channel, the following m units contain the second channel and so forth. Thus, processing patches of $\mathsf{n} \times \mathsf{n}$ pixels size, results in an input-vector of m units:

$$\mathsf{m} = \mathsf{n}^2 * \mathsf{r} \qquad \text{with input channels } \mathsf{r} \in \{1, 2, 3\} \tag{10}$$

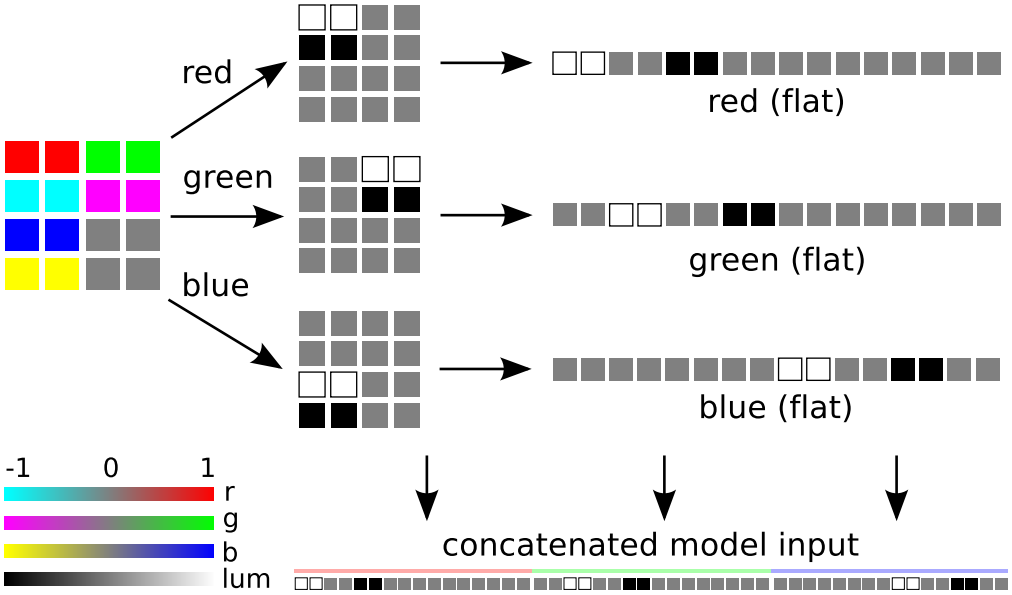

Figure 16: Transformation of pixel values of a RGB 4×4 patch into red, green and blue vectors of 16 units each. The model input vector is the concatenation of the three flattened vectors.

6.3.3 Weight Constraint (Strength)

By not constraining the weights of the connection matrix, setting k to a very low value or skipping equation (6) altogether in the training process, the resulting RF are non-localized [Vincent and Baddeley, 2003, p.1286], and are spanning the whole input of the model. Since unconstrained weights tend to grow in the training process, the application of the weight constraint has the effect of enforcing the *input-reconstruction with minimal weights* of each individual RF. This in turn, has the desired effect of each unit learning distinct aspects / features of the presented input statistics, whereas spatial localization is the most prominent.

By choosing a moderate value for k in relation to η, the model has the tendency to converge in uniformly sized RF, evenly spatially sharing the reconstruction of the input. This is not the case if the value of k is too large and the model operates on more than one input channels (see fig. 17). Then the model converges in a distribution in which some hidden units are strongly localized (very small RF) whilst others have several times larger RF, in some cases the RF are even non-localized, having connections to all input units.

A large value of k in relation to η also sheds light onto a consequence of removing the weight constraint threshold of the original model altogether. By applying the of weight constraint permanently in each epoch, all weights of a unit's RF can be reduced to zero during the training process. Such a unit is considered dead (see 6.3.6).

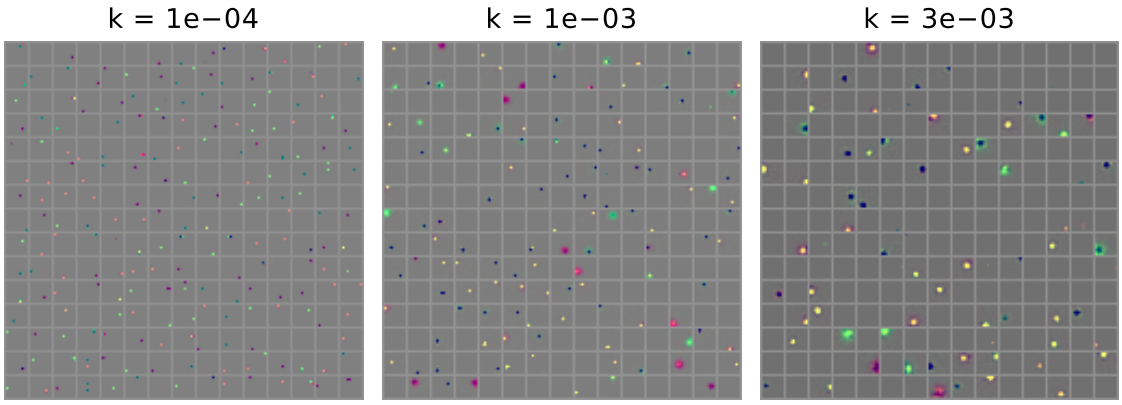

Figure 17: The effect of a large value of k in relation to η causing a distribution with non-uniform RF sizes if the model operates on more than one input channel. All connection maps are trained with the same set of parameters, except k: clip = False, vis = 10^2, hid = 15^2, mode = red vs. green, p = 0.5 and η = 0.01. If the value of $k = 1e-04$ is sufficiently large the model converges in localized RF, all of uniform size. With larger values $k = 1e-03$ and $k = 3e-03$ the model converges in distributions of varying RF sizes.

6.3.4 Weight Constraint (Shape)

The p parameter sets the the shape of the weight constraint, defining how weights of an individual RF are affected:

- Setting $p = 1$ the constraint resembles a L1-norm, effectively treating possible value of a RF's weight equally. L1 constraints favor a neural code with sparse activity of hidden units [Olshausen and Field, 1997, p.3315].
- Thus setting $p = 2$ (L2-norm) penalizes large weights more than small weights (due to its parabolic / spherical shape), but in contrast to L1 will not shape the activity towards a sparse distribution.
- A third possibility, though not explicitly stated by [Vincent and Baddeley, 2003, p.1284], is setting $p = 0.5$ for the constraint to penalizing small values more than larger values. Weak connections, if not reinforced, decay more than stronger connections which appears to be biologically plausible.

6.3.5 Under and Overcompleteness

The completeness of a model simply takes into account the number of visible units (vis) in relation to the number of hidden units (hid). A model which employs more visible than hidden units ($vis > hid$) is said to be **undercomplete**, encoding the

presented statistics in a compressed (possibly incomplete) manner. Conversely a model which has more hidden units than visible ($vis < hid$) is said to be **overcomplete** resulting in a neural code with some degree of redundancy (see fig. 18). A factor of **overcompletess** can be defined simply dividing the number of visible units by the number of hidden units:

$$c = vis/hid \tag{11}$$

The size and spatial location of an individual RF are depending on the sizes and locations of all other RF contributing to the same aspect of the learned input statistic. The model has the tendency (in certain bands in the parametric space) to converge in a distribution in which all RF are of near equal size, uniformly and collectively covering the input. Thus the minimal RF size can be approximated:

$$RF_{size} = \begin{cases} 1 & \text{if model is (over-)complete} \\ c & \text{if model is undercomplete} \end{cases} \tag{12}$$

In the case of a **overcomplete** model localized RF do appear but with no surround (the resulting RF are have solely a center of one pixel). Therefore for the emergence of localized DOG-shaped RF with a significant antagonistic surround, the model needs to be **undercomplete**. It has been observed that models in which $c = vis/hid$ is ranging from 1.2 to 1.8 are producing RF with noticeable surround.

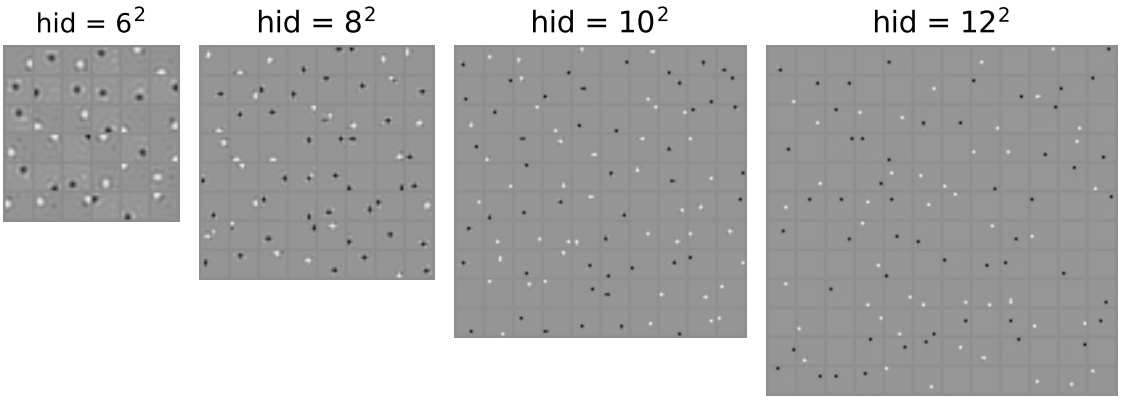

Figure 18: The effect of model **completeness** in relation to RF size: All four connection maps are trained with the same set of parameters, except hid the number of hidden units: clip = False, vis = 10^2, mode = luminosity, k = $5e-04$, p = 0.5 and $\eta = 0.01$. As the model becomes **overcomplete** (hid = 10^2 and larger) the RF size does not shrink further, but more RF die off.

6.3.6 Dead Hidden Units

If all connection weights of a unit's RF are close to zero a unit is considered dead. Being in that zeroed (dead) state, a unit is no longer activated by any input and naturally no longer contributes to the reconstruction of the input either. Dead units are reducing the effective number of hidden units to learn the input statistics.

Three factors are causing unit death during the training process: The first is that of a large value of k in relation to η, the second the model being **overcomplete** and the third is the model is constraining hidden layer activity by utilizing $ReLU$s [Krizhevsky et al., 2012, p.3].

Large value of k (see fig. 19) If over a sufficient number of training epochs, the weights of an individual unit are growing slower then they are reduced by the weight constraint, such a unit trivially ends up being dead (value of η in relation to k too small). This effect can be observed drastically if the constraint parameter k is set to a very large value: After a few epochs the connection matrix is filled with dead RF while the remaining RF are of non-localized shape, as very few alive RF have to reconstruct the input.

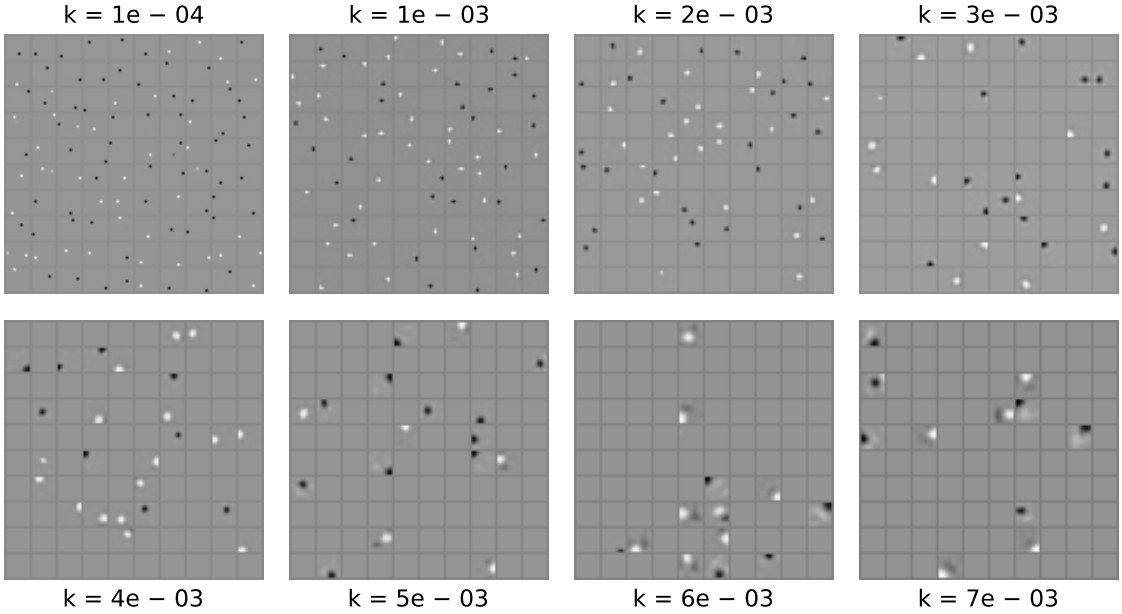

Figure 19: The effect of a large value of k in relation to η causing dead hidden units. All connection maps are trained with the same set of parameters: clip = False, vis = 10^2, hid = 10^2, mode = luminosity, p = 0.5 and $\eta = 0.01$. The model is **complete** since it has the same number of visible and hidden units (10×10) and is not constraining hidden layer activity. If the value of k = 1e − 04 is sufficiently large to converge in localized RF then no hidden units die. Starting with k = 1e − 03 and increasing the value of k gradually up to k = 7e − 03, more and more RF die during the training process. Simultaneously, the the size of the alive RF increases since the remaining RF must reconstruct the input.

Overcompleteness (see fig. 18) Even if the value of k in relation to η is not too large, individual units not contributing much to the input-reconstruction over a longer time eventually die off. Since not contributing much to input reconstruction implicates that weights of a particular unit's RF are of minimal magnitude. This implies accordingly, that only a minimal reconstruction error can possibly occur. Consequently a reconstruction error of minimal magnitude results in minimal growth of a unit's RF weights. If over time, the magnitude of growth is smaller as

the constraint which is reducing the weights in every epoch, this particular unit eventually dies. A unit not contributing to input reconstruction regularly occurs if the model is sufficiently **overcomplete**. As dead units reduce the effective number of hidden units to learn the input statistics, the model is moved towards being **complete** during the training process. This in turn results in compact neural codes with low redundancy, in which all alive hidden units participate in input reconstruction.

The rectification of hidden unit activity by using $ReLU$s is known to produce results in which about 40 % of of all hidden units die off during the training process [Krizhevsky et al., 2012, p.3] (http://cs231n.github.io/neural-networks-1/). Thus, this property of $ReLU$s additionally contributes to the existence of dead units in the converged connection matrix.

6.3.7 Constraining of Model Input and Output Values

Input values as well as output values of the model can be constrained of being solely positive, or unconstrained of being positive and negative. Thus the model can be trained in four different settings (see table 2).

setting	symbol	input (x)	output (y)	subtract mean	clip hidden layer
1	(x±/y+)	±	+	yes	yes
2	(x±/y±)	±	±	yes	no
3	(x+/y+)	+	+	no	yes
4	(x+/y±)	+	±	no	no

Table 2: Model settings. A + represents solely positive values, whereas a ± symbol indicates positive and negative values.

Subtract mean In order to obtain positive and negative input values, the mean of the input patch is subtracted after normalization (see fig. 20). This results in values of the input ranging from -0.5 to 0.5. Subtracting the mean has a biological motivation since it emulates the differences of intensity propagated by horizontal and bipolar cells.

As a result from an information theoretic perspective, the model is freed of learning the meaningless (possibly) large mean of the patch. If the data contains structure beyond the mean values of particular patches, the model can learn this structure faster. Conversely, not subtracting the mean of the input patch results in feeding (scaled) normalized RGB values into the model. The values of the input are then ranging from 0 to 1.0.

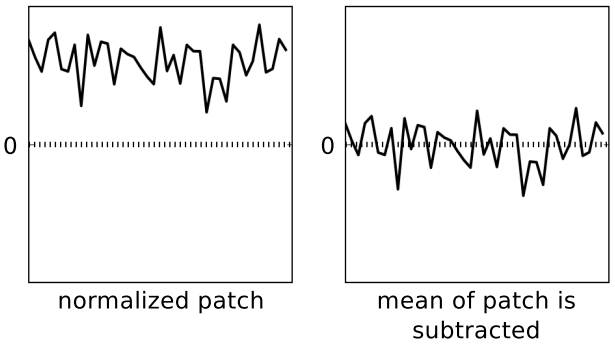

Figure 20: The effect of subtracting the mean of a set of positive values.

Clipping Constraining (clipping) hidden unit activity demands using *ReLU*s which transfer function is setting all negative values to zero (see fig. 21). Clipping hidden unit activity has a biological motivation since neurons are not able to have negative responses (fire rates).

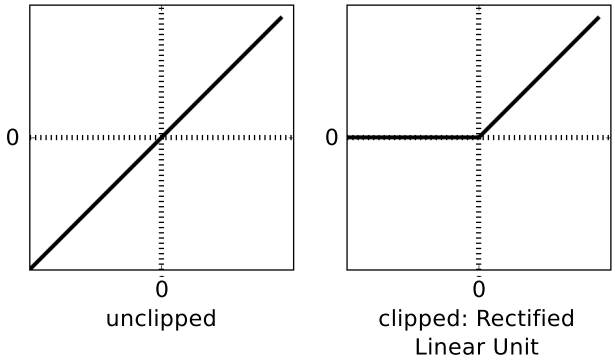

Figure 21: Unconstrained (left) and clipped hidden layer activity (right).

By halving the ability of a particular unit to encode the input, the effective number of hidden units, capable of encoding the input, is also halved. Subsequently, clipping hidden units has a strong effect on the distribution of localized RF in a converged map: If we assume an unconstrained *complete* model with solely one input channel. And are further assuming that the model converged in localized RF. Then for each input pixel exists one specific unit, reconstructing the value of this particular input location. Since unconstrained units are allowed to hold negative values, an OFF unit can produce positive input reconstruction values and vice versa. And such, regarding input reconstruction, whether the sign of the center weight of an unconstrained unit is positive (ON) or negative (OFF) is of no interest. Thus a single hidden unit per pixel suffices for optimal reconstruction. Consequently, models with unconstrained hidden units converge in a distribution of RF in which for each input channel one complete output channel emerges. A channel is composed of weakly overlapping mosaic of ON and OFF units. Whether a hidden units ends up being of ON or OFF signature is determined by the initial random initialization of the connection matrix.

In contrast, an individual constrained unit is only able to reconstruct input values of the sign of its center weight. Thus an ON *ReLU* is limited to only reconstruct positive and conversely an OFF unit is limited to solely reconstruct to negative input values. Subsequently, a particular ON or OFF unit cannot reproduce the complete set of values for a given pixel by its own. Hence, the population of one type of hidden unit cannot reconstruct the negative and positive parts of the input alone, which excludes a the emergence of a combined channel with a mosaic distribution of ON and OFF units. Optimal reconstruction of positive and negative input demands a distribution in which the localized RF of ON units, as well as these of OFF units, are spatially covering the entire input. Thus models constraining hidden units converge in a distribution of RF in which for each input channel two complete output channel emerge. A channel is uniformly composed of weakly overlapping units of either ON or OFF type.

6.3.8 How many Hidden Units?

The question arises of how large the hidden layer in relation to the input should be chosen, so that RF emerge with significant surround. It has been noted earlier,

that the model needs to be *undecomplete*. But one can be more specific:

- Taking into account the number of channels m we obtain $vis * m > hid * m$. Though it is of little of interest for the following observations (assuming left and right-hand m are always chosen equally).
- Knowing that by not constraining hidden layer activity in setting (x±/y±) and (x+/y±), the model converges in a distribution in which, for each particular channel, the visual field is covered by an alternating mosaic structure by ON and OFF-center RF. In this case $vis > hid$ still holds, as hid is simply the sum of $hid_{ON} + hid_{OFF}$ units, since both sub-types are sharing the visual space.
- However, by clipping the hidden layer activity in setting (x±/y+) and (x+/y+), the model converges in a distribution in which, for each particular channel, the visual field is covered uniformly by each of the ON and center-OFF RF separately. Subsequently, only half of the number of units is available to encode the input statistics, which in turn doubles the expected size of each RF. Hence the number of hidden units have to be doubled if one wants to preserve the RF-size of an un-clipped sitting.

  Additionally the effect of dead RF resulting in the application of $ReLU$ should be taken into account. If we estimate that about 40% of all hidden units die during the training process a constant factor of $alive = 0.6$ should be introduced. However, if leaky $ReLU$ are applied this number should be much higher (> 0.8), as the the effect of dying neurons is greatly reduced. Thus $vis > alive * hid/2$ should hold.

6.4 Parametric Fitting

In the fitting process, the scalar output of an objective function is minimized, given a function (the model), a set of parameters, and some data to fit the model to. Usually for a set of parameters, the objective function compares given data and model output and returns its squared difference (the reconstruction error) which is then minimized during the optimization process.

A single run of the fitting algorithm emits a set of parameters which reconstruct the given receptive field with the smallest error. To avoid solutions stuck in local minima, a sequence of fitting attempts is executed, each initialized with slightly varied start parameters. Subsequently, from this sequence, the best fit, with the smallest reconstruction error, is chosen. We are using the SLSQP (Sequential Least Squares Programming) algorithm of the *scipy.optimize* package to fit RF to the the parametric model.

6.4.1 Spatio-chromatic Parametric Model

The excitatory center and inhibitory surround organization of retina ganglion cell RF is commonly modeled with a Difference of Gaussians (DOG) function

[Shapley and Hawken, 2011, p.708]. In its simplest case, a circular two-dimensional gaussian function has a center position μ_x μ_y, a radius r and an amplitude a. A circular DOG function, simply the sum of two gaussians sharing the same center position, needs two more parameters: a surround radius r_s and the amplitude a_s of the surround part.

Moreover, the circular DOG function can be made elliptical which adds a rotation parameter and splits the radius variable into the spread of the two main axes of the ellipsoid. The parametric model needs to reconstruct RF trained in a arbitrary *mode*, thus it has to be extended to at least three dimensions. Subsequently, the elliptical DOG model has a set of 13 parameters:

- the elliptical Gauss function has center μ_x μ_y, spread σ_x σ_y and rotation θ.

$$\mathsf{gauss}(x, y) = e^{-a(x-\mu_x)^2 + 2b(x-\mu_x)(y-\mu_y) + c(y-\mu_y)^2} \tag{13}$$

$$a = \cos^2(\theta)/2/\sigma_x + \sin^2(\theta)/2/\sigma_y \tag{14}$$

$$b = -\sin2(\theta)/4/\sigma_x + \sin2(\theta)/4/\sigma_y \tag{15}$$

$$c = \sin^2(\theta)/2/\sigma_x + \cos^2(\theta)/2/\sigma_y \tag{16}$$

- the DOG function has ratio of center to surround γ and scale of the surround part relative to the center part k_s.

$$\mathsf{DOG}(x, y) = \mathsf{gauss}(x, y) - k_s\ \gamma\ \mathsf{gauss}(x, y) \tag{17}$$

- the chromatic part of the model with additional parameters bias and direction for each color channel:

$$\mathsf{DOG}_{rgb}(x, y) = [\mathsf{red}, \mathsf{green}, \mathsf{blue}] \tag{18}$$

$$\mathsf{red} = \mathsf{bias}_r + \mathsf{direction}_r\ \mathsf{DOG}(x, y) \tag{19}$$

$$\mathsf{green} = \mathsf{bias}_g + \mathsf{direction}_g\ \mathsf{DOG}(x, y) \tag{20}$$

$$\mathsf{blue} = \mathsf{bias}_b + \mathsf{direction}_b\ \mathsf{DOG}(x, y) \tag{21}$$

6.5 Clustering Receptive Fields

Upon the fitted data, the receptive fields can be further analyzed: knowing the spatial position of the *center* of the ellipse, the value of its corresponding weight can trivially be extracted from the receptive field vector. The center weight, representing its particular receptive field, is used as an observation to be fed, along with the center weights of all other RF, into a clustering algorithm (we are using the *k-means* algorithm of the *scipy.cluster.vq* package).

This is also true for a simple approximation of the values of the *antagonistic surround* of a receptive field: By extracting the weights at the four locations where the minor and the major axis of the ellipse cross the outer curve, a mean of the four weights gives a simple approximation. If used, the surround value together with the center value is concatenated into a single observation vector.

Interpreting the observations as color values the distribution of the RF of a learned connection map are clustered over a color space. Moreover, the resulting clusters are ordered by their particular prototype color (the mean of all center weights belonging to a cluster), resulting in the familiar order of colors: red, green, blue, cyan, magenta, yellow. Clustering solely upon the center-weight value of a RF avoids overfitting of the clustering algorithm.

By utilizing the results of the fitting and clustering process, the RF of an original trained map can be ordered by their particular center positions and belonging to a specific cluster. Linearizing the center position $[x, y]$ of a parametric fit into a scalar $x + y * pw$, (pw is the pixel-size of an input patch), gives a partial order in which the RF can be sorted spatially and grouped by its relationship to a specific cluster.

6.5.1 Prototype Filters

A prototypical RF represents best the morphology of a particular cluster. It can be selected manually by specifying a RF for each channel or it can be selected algorithmically based on the following criteria:

- RF center is close to the center of the visual field
- has been fitted with minimal reconstruction error
- RF maximum of its absolute weights is above a threshold. The purpose is to exclude RF with near zero weights, which may perform well in the other criteria

The resulting prototypical RF can then be used in *convolutional neural networks* (CNN) for pre-processing RGB input.

7 Results

The model was trained with RGB images in four settings, whereat each is constraining the input and hidden layer activity differently (see section 6.3.7). In all settings, training the model took an enormous number of epochs until stable localized RF emerged. An exception is setting (x+/y+)a which took approximately an eighth of the time to converge (see table 3). The enormous training can be explained by attempting to obtain a parameter set of the model which converges in specific RF distributions. RF of such a distribution should have the following desired properties:

1. all RF are localized
2. mostly (all) RF are of uniform size
3. a sufficient number of distinct and complete channels emerge

Converging in distributions in which RF emerge with the above properties, implies setting the values of k and η reasonably small. Subsequently, small values result in long training times until the model finally converges.

setting	epochs	time ep.	time total	sq. rec. error
(x±/y+)	66.089600	19.92s	4d 16h 27m	9.09e − 06
(x±/y±)	50.819200	6.68s	1d 6h 40m	1.24e − 05
(x+/y±)	61.920000	6.89s	1d 9h 17m	6.24e − 06
(x+/y+)a	8.854400	19.69s	0d 14h 59m	7.03e − 05
(x+/y+)b	36.774400	21.88s	2d 22h 41m	9.90e − 05

Table 3: Duration of the training process for each setting. The column *time ep.* shows the duration of calculating 3200 training epochs.

Parameter k Since training the model solely with one input channel (luminosity mode), always results in the model converging in uniformly sized RF, given that the value of k is sufficiently large for localized RF to emerge. However by training the model upon more than one input channel, the value of k now affects the uniformity of converged RF sizes (see fig. 17). Thus in other modes, a value for k is needed which is large enough for localized RF to emerge, but not too large to preserve the property of all RF converging in uniform size. As it turns out training the model in rgb mode required a relatively small value of parameter k. A small strength of the weight constraint results in a longer training time as more epochs are needed until the RF are of localized shape.

Parameter η Moreover training with more than one input channel rendered the model more brittle to the effect of a too large learning rate η. In this case, which is not apparent at the beginning of the the training process, training is running normally for a while, minimizing the reconstruction error. RFs are becoming more

and more localized as the training progresses. However by the reconstruction error reaching a certain threshold, RF weights start to oscillate around some local optima. Consequently the training process is stuck forever and will never leave this local optima. Subsequently η was set to a small value, which is additionally extending the number of training epochs required for converging in localized RF.

7.1 Training Corpus

The training set contained 293 RGB-images in (tiff format) solely of various natural scenes, compiled mostly of images originating from the *McGill Calibrated Colour Image Database* (http://tabby.vision.mcgill.ca/). Since the auto-encoders are trained unsupervised, the training corpus is not split into a training and validation set. Due to the nature of auto-encoders to learn the statistic of the samples presented, (subtle) changes in the set of training images have large effect on the shape and functionality of emerging RF.

7.2 Initial Parameters

As a reasonable undercompleteness of the model is desired, choosing the initial number of hidden units (see table 4) is straightforward in setting (x±/y±) and setting (x+/y±). Assuming a moderate value of k (see table 5), the occurrence of dead hidden units can be ruled out at all. Thus initially setting the number of hidden units suffices, to obtain a model of a desired undercompleteness of c ~ 1.2.

setting	vis × hid	vis	hid	hid alive	alive %	c
(x±/y+)	13×39	507	1521	916	60%	1.11
(x±/y±)	13×20	507	400	400	100%	1.27
(x+/y±)	13×20	507	400	400	100%	1.27
(x+/y+)a	13×39	507	1521	555	36%	1.83
(x+/y+)b	13×42	507	1764	659	37%	1.54

Table 4: Parameter values of the four settings. From left to right the meaning of each column: name of the setting, number of visible / hidden units squared, visible units, hidden units, hidden units alive after training and the **undercompleteness** c of the model.

Three reasons are explaining that death of hidden units cannot occur in setting (x±/y±) and setting (x+/y±), as stated above: At first, hidden unit death resulting solely from a large weight constraint is impossible (k is set to a moderate value). Second, as the model begins the training process initially undercomplete death of hidden units caused by not contributing enough in reconstructing the input also is impossible to occur. And finally since setting (x±/y±) and setting (x+/y±) are not constraining hidden layer activity, thus are not using *ReLU*s [Krizhevsky et al., 2012, p.3] which excludes unit death accounted to the clipping of the hidden layer. The results show indeed no unit death (see table 4).

In contrast setting (x±/y+) and settings of type (x+/y+) are clipping hidden layer activity and thus are utilizing *ReLU*s. Subsequently, choosing the number of hidden units has to take into account: the inevitable death of hidden units due to clipping, as well as the limited resolution of a clipped model. Consequently, the number of hidden units is set reasonably large: to three times the number of visible units, which renders the model initially **complete**. Indeed, after the training process finished the model is reasonably **undercomplete** with $\mathsf{c} = 1.11$ in setting (x±/y+) and strongly **undercomplete** with $\mathsf{c} = 1.83$ in setting (x+/y+)a. The comparably high value of c in setting (x+/y+)a reflects the much higher loss of hidden units during the training process.

setting	k	p	η
(x±/y+)	$7e-06$	0.5	0.03
(x±/y±)	$1e-05$	0.5	0.03
(x+/y±)	$1e-05$	0.5	0.02
(x+/y+)a	$5e-05$	0.5	0.01
(x+/y+)b	$2e-05$	0.5	0.01

Table 5: Model parameters values of k, η and p:

Conversely, hidden unit death is observed in setting (x±/y+) and settings of type (x+/y+). In the literature, related to utilizing *ReLU*s, it is reported that about 40% of all hidden units are dead after the training process has ended. This exactly matches the number of alive units 60% (916) in setting (x±/y+).

However the number of alive hidden units in setting of type (x+/y+) is much lower. In addition to clipping hidden layer activity, constraining the input in settings of type (x+/y+) has the effect of demanding a higher value of k compared to the other settings, for producing localized RF. As shown in the model section, the higher the value of k the more units die during the training process, unaffected by the model being (possibly) **undercomplete** (see fig. 18). If one assumes that the number 40% of dead units attributed to *ReLU*s holds in setting (x+/y+)a, the additional 24% of dead units can be speculated to be the result of a rather large value of the weight constraint.

In the most constrained settings of type (x+/y+), it appeared to be impossible to converge in a distribution of uniformly sized RF. This due to the circumstance of settings of type (x+/y+) being more sensitive to even low values of η used in the remaining settings. Hence lower values are used which increase the number of epochs needed to converge. By using the same value of k as in setting (x±/y+), in which it produced uniformly sized RF, the model fails to converge in localized RF at all. Subsequently larger values of k are used, which is having the consequence of the model converging in a distribution of RF in varying nonuniform sizes. If k was chosen a little too large the number of units rendered dead during training raised to such numbers that the remaining RF became non-localized. This contradicts the experience of increasing the value of k is resulting in a (faster) localization of RF during the training process.

7.3 Determining the Number of Channels

A particular channel is seen to be *complete* if the RF of all of its hidden units cover the visual field entirely with minimal overlaps. Since the actual number of *complete* channels emerging in each setting is not known beforehand and the *k-means* and *spectral* clustering algorithms demand the number of clusters to be initially specified, a reasonable initial value of the number of clusters is needed:

The model will emerge in n mixed ON / OFF channels for n input channels, in settings where the hidden unit activities are not clipped. Whereas clipping hidden units activity results in the emergence of 2*n uniform ON and OFF channels for n input channels (see 6.3.7). Thus training the model with RGB images should lead to the emergence of six complete distinct channels of uniform RF signature for settings in which the hidden layer activity is clipped, such as (x±/y+) and (x+/y+). And three channels of mixed RF signature are assumed to emerge for settings in which the hidden layer activity is not clipped, such as (x±/y±) and (x+/y±).

7.4 Clustering Results

The results show indeed that in setting (x±/y±) and setting (x+/y±) three distinct complete channels emerge (see figs. 22, 25, 26, 27 and table 6). In both settings, each channel is composed of a random mosaic of ON and OFF cells. Accordingly setting (x±/y+) emerges in six distinct complete channels of uniform RF signature, due to the reduced capability of hidden units to reconstruct the input.

However unexpectedly, in setting (x+/y+)a five and setting (x+/y+)b six distinct complete channels of uniform RF signature emerge, of which two appear to cover non-chromatic (black and white) aspects, whilst the remaining cover chromatic aspects of the input.

7.4.1 General Patterns

By comparing the number of units in each cluster, some general patterns are visible throughout all settings (see tables 6 and 10):

- In setting (x+/y+)a and setting (x+/y+)b roughly half of the hidden units are tuned to luminosity whilst the remaining units cover the chromatic aspects of the input.
- In the other settings the ratio of the number of units in each channel in relation to the number of alive hidden units fluctuates around 1/n for n channels.
- Of red, green and blue channels, the blue channel (if present) has always the highest number of hidden units, whilst the red channel has always less units than blue but always more than green.

With the exception of (x±/y+), the green channel always has the lowest number of hidden units in all settings (see fig 23). The effect of the green channel having the

smallest number of units is sidelined by the magenta channel of setting (x±/y+) having the largest number of hidden units. Magenta as a color has no amount of green. Conversely, the effect of the blue channel having the largest number is even visible in setting (x+/y+)b in which no separate blue channel emerges: Since cyan as a color has a large amount of blue, accordingly the cyan channel has the largest number of units of all chromatically selective channels in setting (x+/y+)b.

setting	white	black	red	green	blue	cyan	magenta	yellow
(x±/y+)			17 %	15 %	18 %	14 %	19 %	17 %
(x±/y±)			32 %	25 %	43 %			
(x+/y±)			36 %	24 %	40 %			
(x+/y+)a	29 %	24 %	13 %	12 %	22 %			
(x+/y+)b	25 %	23 %	12 %	10 %		17 %	13 %	

Table 6: Sizes of clusters shown in percentage: the number of units per cluster in relation to the total number of alive hidden units. Meaningful cluster names are given reflecting the prototype color of each cluster.

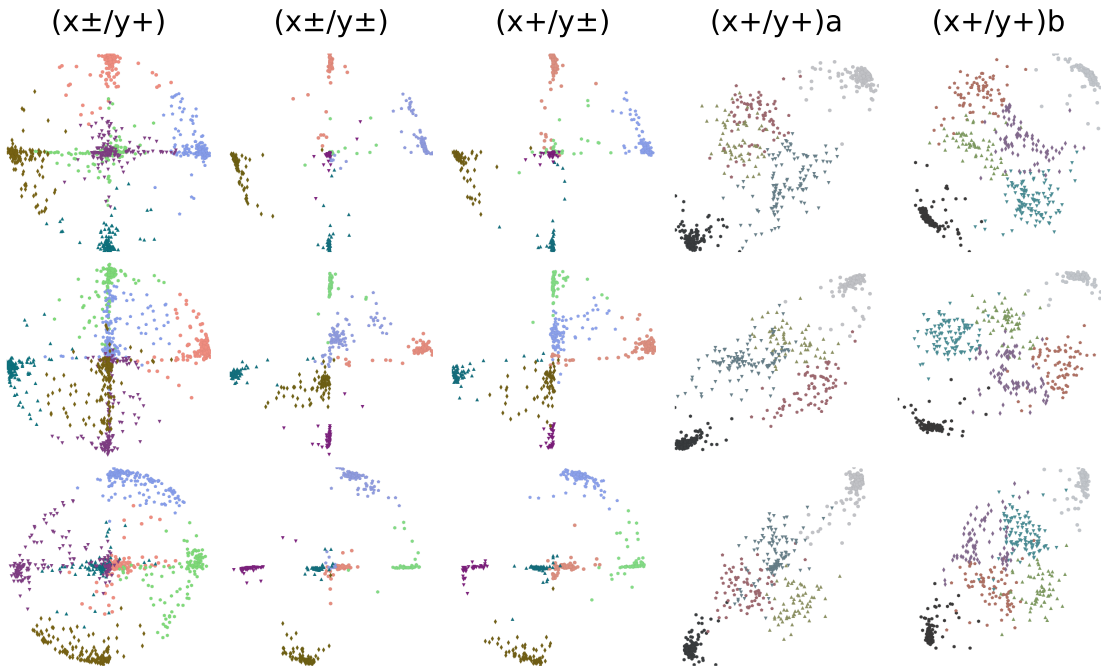

Figure 22: Distributions of RF center weights in RGB color-space. All RF belonging to a cluster are colored in its prototype color. The first row shows the axis of red and blue, the second green and red and the third blue and green. The more hidden units the setting has the finer is the RGB color space is covered. Settings (x±/y+), (x±/y±) and (x+/y±) show RF centers weights mostly clumping along a particular axis of the RGB color-space towards the surface of the RGB unit circle.

If one assumes green color to appear more frequently in an image patch of natural scenes, then red, which in turn appears more often than blue, then this distribution can be found inversely in the number of units in each channel. Having

a lower number of units implies the area of the RF to be larger compared to a channel with more units and vice versa (see fig 23).

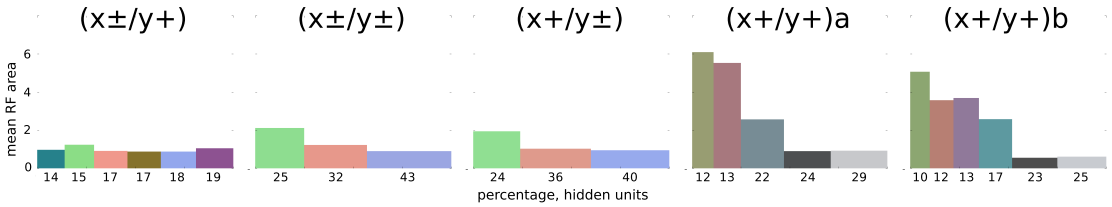

Figure 23: The figure shows for each setting and each cluster the mean RF area (y axis) together with the number of hidden units in percentage (x axis). The size of a particular cluster is displayed as width of a bar in proportion to the size of the remaining clusters. Each cluster is drawn in its prototype color and ordered from left to right with ascending size.
With the exception of (x±/y+), in all settings the green cluster has the smallest number of hidden units with the largest RF area, followed by the red cluster which has always less hidden units with larger RF areas than the blue cluster (if present).

The RF area of hidden units in a channel covering specific features of the input and the frequency of occurrence of such features in the training data appear to be related. If a feature occurs more often in the input data than another feature, the model is updated more often by the reconstruction error of that feature and less often updated by the reconstruction error resulting of the other less frequently occurring feature. Consequently, more updates by one feature implies that RF weights already tuned to that feature are updated more often than other weights. Since spatially equivalent weights of similar tuning in multiple RF are competing in the reconstruction of the input, more frequent updates result in the competition process progressing further compared to the competition of other less frequently occurring features.

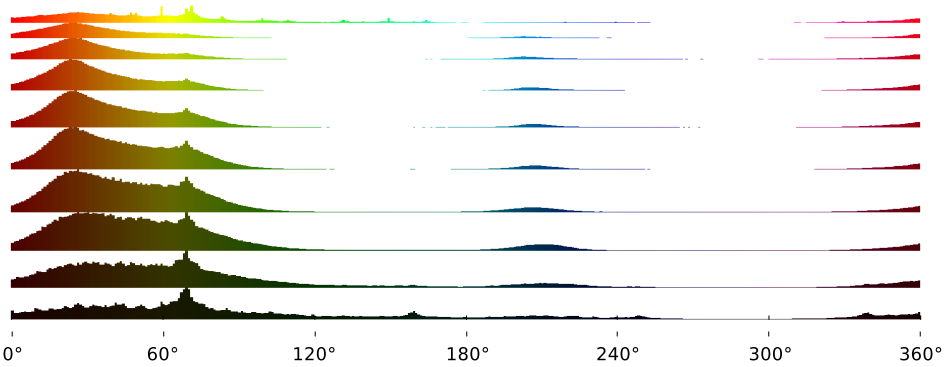

Figure 24: The figure shows 10 histograms in the HSV color space of 80000 input patches. The lowest histogram shows the distribution of colors with intensities ranging from 0 to .1, whilst the last shows intensities from .9 to 1. In between the remaining histograms in ascending order.
The distribution of colors in the training data does not appears to explain the inverse correlation among cluster size and mean RF area in the results of training the RGC model. Rather, it shows that model is not learning a color bias in the training data, but learns another statistic hidden in the RGB images of natural scenes.

Thus the further the competition progressed the more weights of multiple RF have lost the competition in reconstructing the input. In setting (x±/y±) and setting (x+/y±) the losing hidden units have tuned to other features, while in the remaining settings, losing hidden units can also end up dead. In both cases the RF sizes of winning hidden units are growing, which could explain the general pattern of the green cluster being the smallest followed by red and blue clusters.

However, the above hypothesis does not hold (see fig. 24) in explaining the inverse correlation among cluster size and mean RF area. The histogram shows that in the training data red occurs more often over all intensities compared to green and blue colors. This is also visible in the thumbnails of all images in the training set (see fig. 45).

7.4.2 Chromatic and Achromatic Channels

In settings of type (x+/y+) all RF in the achromatic channels are strongly localized and of small size whilst the RF in the color channels are larger and of non-uniform size and morphology. In contrast RF of the remaining settings are more evenly sized (see fig. 23).

The split of RF covering chromatic and achromatic aspects has only be observed in settings of type (x+/y+), never in the remaining settings. Other settings have been extensively trained, even with large weight constraints producing nonuniform RF morphologies, but separate achromatic channels did not appear in the clustered data. Therefore, the property of separating the input in chromatic and intensity aspects appears to be attributed to (x+/y+) exclusively. Notably, in settings of type (x+/y+) hidden unit death occurs in large numbers during the long training process (see table 4). This leaves only a small window in the parameter space in which (x+/y+) converges in localized RF.

Moreover, it is still unknown if, given a sufficient number alive hidden neurons, (x+/y+) converges in more than six channels (see table 7). This is indicated by setting (x+/y+)a converging in five (three chromatic) channels with roughly 16 % less hidden units compared to (x+/y+) which converges in six (four chromatic) channels.

setting	n	white	black	red	green	blue	cyan	magenta	yellow
?	4	•	•	•	•				
(x+/y+)a	5	•	•	•	•	•			
(x+/y+)b	6	•	•	•	•		•	•	
?	8	•	•	•	•	•	•	•	•

Table 7: One could hypothesize that (x+/y+) maximally converges in 8 channels, as such two luminous and six chromatic channels (three ON, three OFF). This conjecture is based on the observation of the distribution of channels of setting (x+/y+)a and setting (x+/y+)b. The • symbol indicates the emergence of a complete channel in a particular column denoting its type.

7.5 Mosaic Structure of the Fitting Results

The following figures (see figs. 25, 26 and 27) show the result of the fitting and clustering process for each particular setting. The RF of a hidden unit is drawn as an ellipse with the parameters obtained from the best fit of the spatio-chromatic DOG model. The color in which an ellipse is drawn reflects the value of the center weight of a particular RF. For clarity, only the center ellipse of a fit is shown.

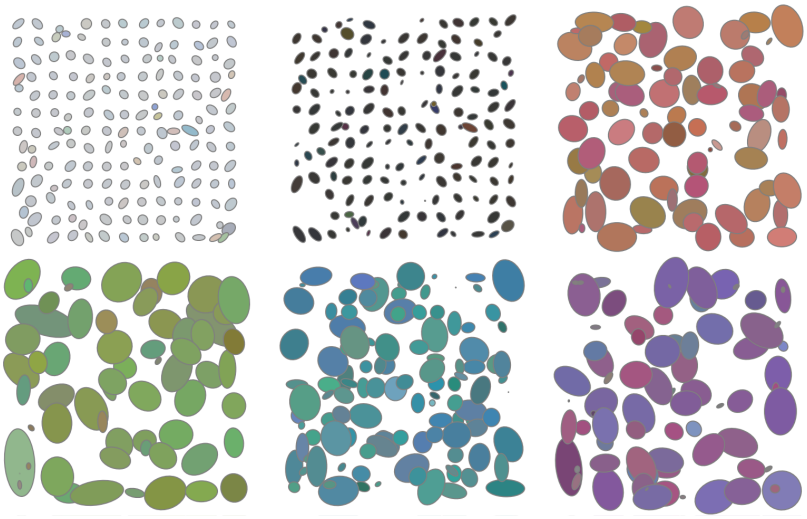

Figure 25: **Setting (x+/y+)b**: Six complete channels emerged with varying RF sizes: Luminosity ON and OFF, red, green, cyan (red OFF) and magenta (green OFF).

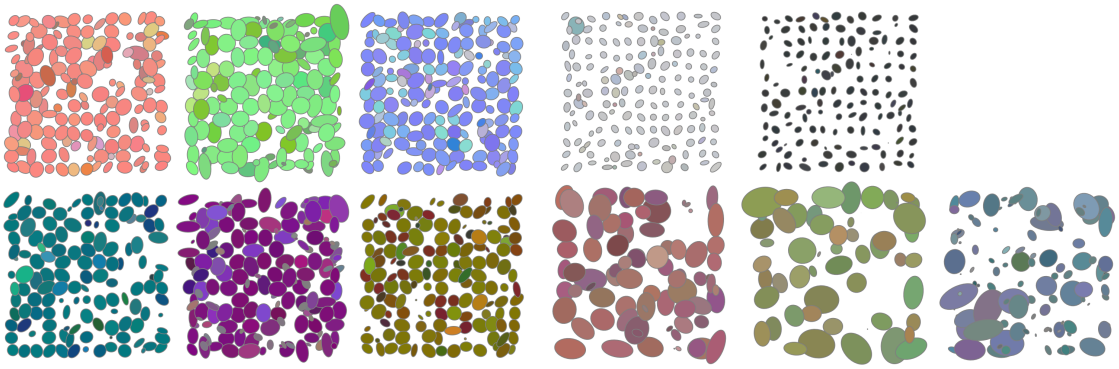

Figure 26: Left: **Setting (x±/y+)**. Six channels emerged, each spanning the entire visual field: three ON channels (red, green, blue) and three OFF (cyan, magenta, yellow). Right: **Setting (x+/y+)a**. Five complete channels emerged with varying RF sizes: Luminosity ON and OFF, red, green and blue.

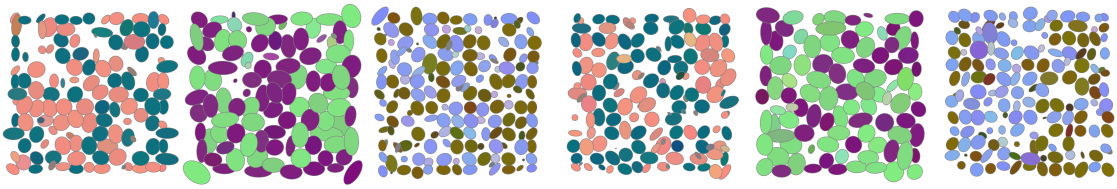

Figure 27: Left **Setting (x±/y±)** and right **Setting (x+/y±)**; Three complete channels emerged, with a mosaic of ON and OFF receptive fields: (red, cyan) and (green, magenta) and (blue, yellow).

7.6 Prototype RF

Due to the similarities of settings (x±/y+), (x+/y±) and (x±/y±) only prototype RF of (x±/y+) are shown (see fig. 28). Accordingly the similarities of setting (x+/y+)a and (x+/y+)b allow to only show the prototype RF of (x+/y+)b (see fig. 29). Prototype RF of all settings are shown in the appendix.

Overall, prototypes RF of setting (x±/y+) are of similar size and appear to be more localized and sharper compared to setting (x+/y+)b. However setting (x+/y+)b prototype RF of luminous channels are more localized and are having a smaller surround than the prototypes of setting (x±/y+). And RF prototypes of the chromatic channels of setting (x+/y+)b have much larger center and surround areas compared to setting (x±/y+). Additionally to spatial differences, RF of setting (x±/y+) and setting (x+/y+)b differ in the texture of the antagonistic surround.

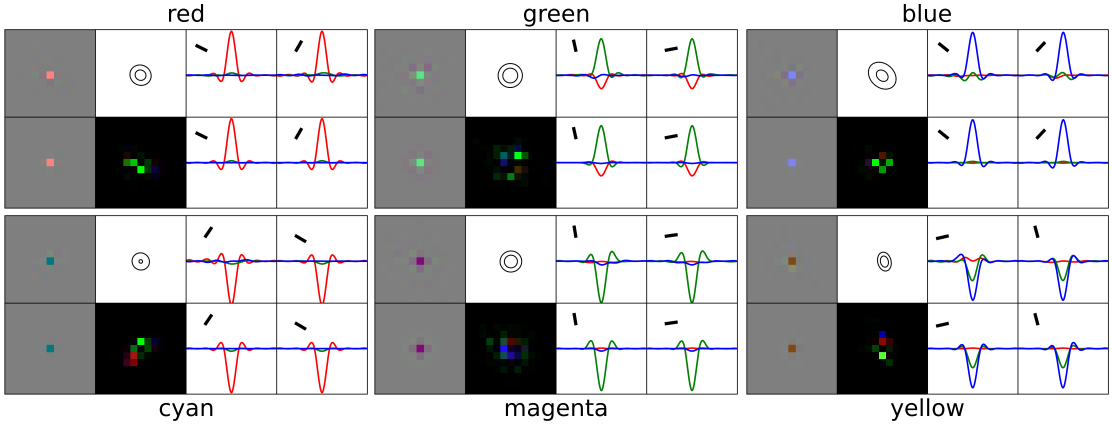

Figure 28: **Setting (x±/y+)**: Each box of eight tiles shows the prototype RF (top left), the reconstruction generated of the best fit (bottom left), the ellipse of the best fit (top, 2nd column) and the squared reconstruction error of the RF and the reconstruction (bottom, 2nd column). The third column shows an interpolated cut of the primary axis of the best fit ellipse for the RF (top) and the reconstruction (bottom). Whereas the fourth column shows the secondary axis of the ellipse of the RF (top) and the reconstruction (bottom).

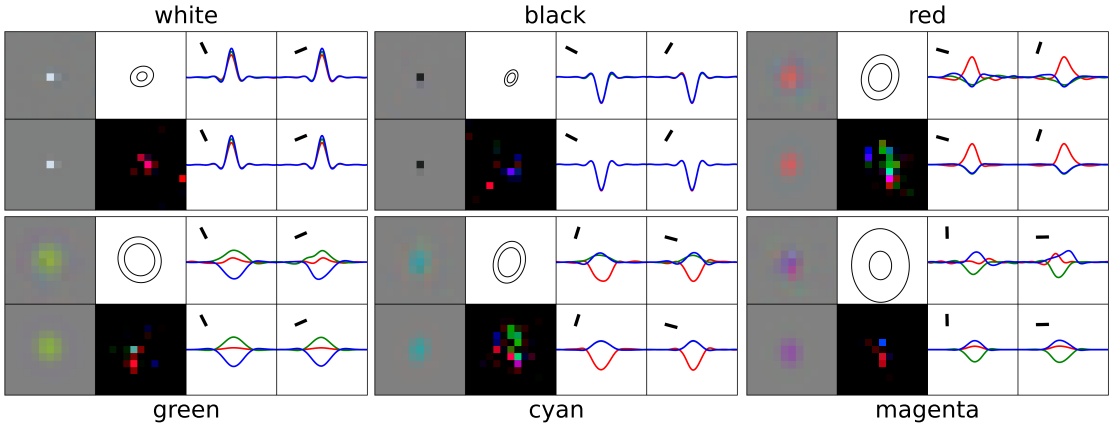

Figure 29: **Setting (x+/y+)b**: prototypical RF for white, black, red (1st row), green, cyan and magenta (2nd row) channels.

Setting (x±/y+) produces RF providing *intra-channel* chromatic contrast of which the surround is of uniform color and structure. This replicates the results of [Brown et al., 2011, p.28] which where obtained by statistical means. The large surround of chromatically selective RF of setting (x+/y+)b consists of several blobs, in which some of the blobs appear to be moved towards *inter-channel* opponent color, whilst some are of *intra-channel* opponent color. As a result the spatio-chromatic DOG-model reconstructs the surround of setting (x±/y+) prototypes better than prototypes of setting (x+/y+)b with a smaller reconstruction error and of less spatial size.

7.7 Convolution Filters

The next figures show the result of filtering RGB images with the obtained prototype RF (see fig. 30 and 32). In such a figure, the upper two rows show the filtering results of the original image for each prototype RF. Images in the lower two rows show the results of simulated cortical processing by combining the results of chromatic opponent channels (as found in *midget* RGC). For brevity, the cortical opposite signed counterparts are not shown.

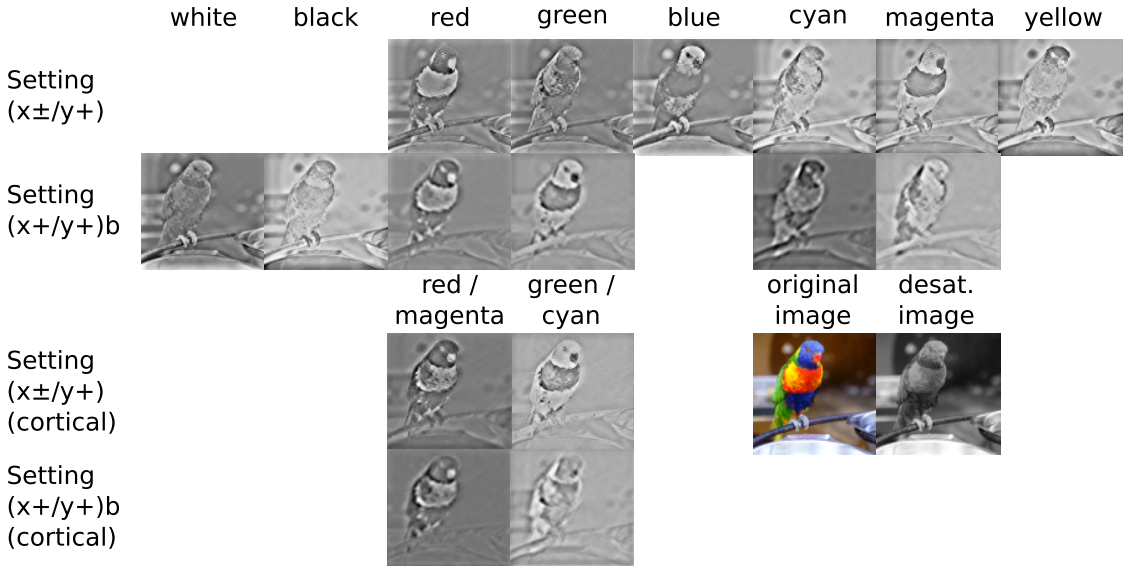

Figure 30: Filtering results of a an image containing chromatic (the parrot) and achromatic contrasts (the background and the claws of the parrot). The results of (x±/y+) prototype filters show parrot and background in similar sharpness and intensity. Whereas prototype filters of (x+/y+)b show a fuzzier and weaker background and sharper and stronger areas of the parrot.
Concerning the simulated cortical results (bottom two rows), (x±/y+) combined prototypes appear to resemble closely (x+/y+)b (red and green) prototypes. Furthermore, the combined (green/cyan) result of (x+/y+)b appears to be the only result to separate green parts of the parrot clearly.
All of this indicates that (x+/y+)b prototypes discriminate stronger than (x±/y+) prototypes among chromatic and luminous aspects of the input.

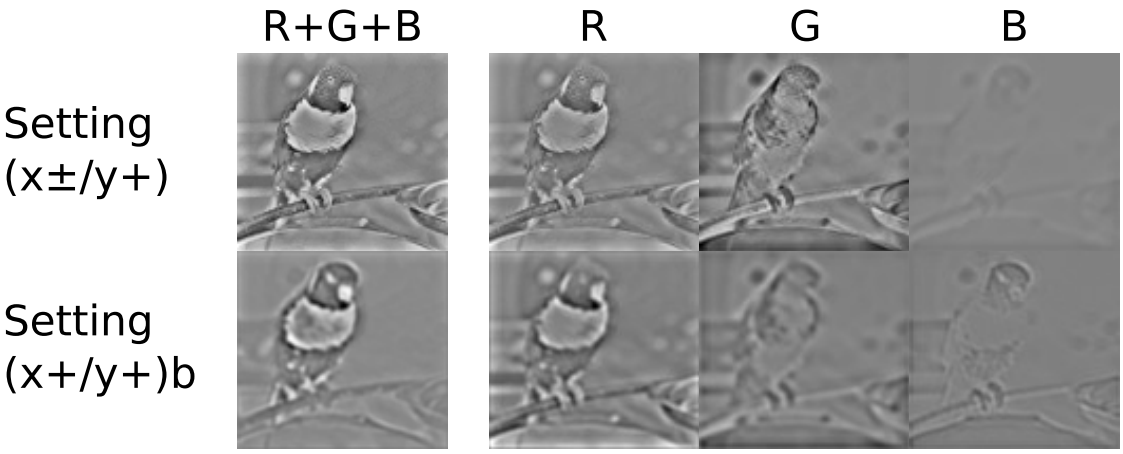

Figure 31: Showing the RGB components of the convolution result of the red-ON prototype. The red-ON prototype of setting (x+/y+)b differs most visibly in the B component from the red-ON prototype of setting (x±/y+).

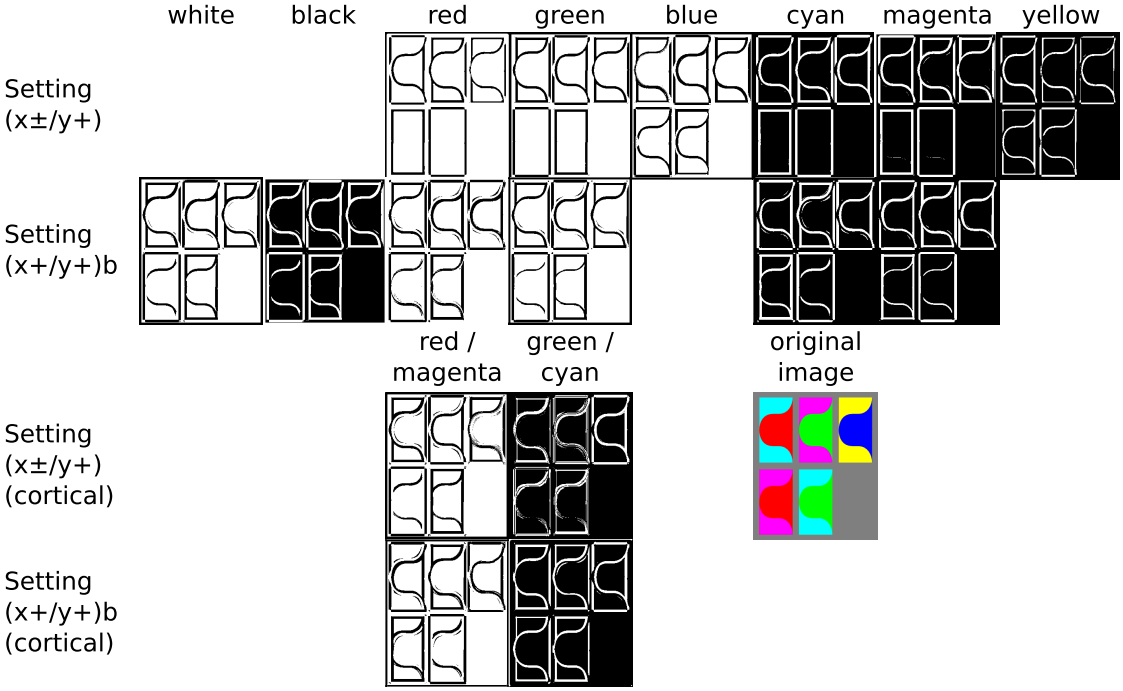

Figure 32: Filtering results of a test image containing different types of chromatic contrasts. The first row of the test image contains *intra-channel* red-cyan, green-magenta and blue-yellow chromatic contrast, whilst the second row contains biological motivated red-magenta, green-cyan contrast. The output has been posterized (reduced to 2 colors, black and white) to quickly determine the sensitivity of the prototype RF filter.
Red, green and their OFF signed counterparts of setting (x±/y+) (first row) prototypes appear to be blind to biological types of chromatic contrast whereat prototypes of setting (x+/y+)b (second row) capture all types of chromatic contrast in the test image. Considering the results of combining the output of chromatically opposing filters (cortical, bottom two rows) shows all types of chromatic contrast are captured in both settings.

7.8 Biological Plausibility

Before answering the question of which setting can be biologically motivated the most one has to note that the model simply receives the input of flattened RGG image patches. Thus the model assumes that red, green and blue cones exist in equal numbers in an equally spaced, grid-like spatial distribution. Moreover, since blue sensitive cones only amount up to 15 % of all cones in the human retina and even are non-existent in the central fovea, the image statistics the model is learning appear not to be biologically plausible.

Parasol and Midget RF Size The RF achromatic hidden units in setting of type (x+/y+) are of small localized shape whilst the RF of chromatic hidden units are much larger (see figs. 25, 26 and 27), which is inversely found in the morphology of *midget* and *parasol* RGC. This conflicting result can be explained by *parasol* RGC transmitting information faster than *midget* RGC which comes at a cost of a larger soma and thicker axons of *parasol* RGC [Baden et al., 2014, p.2]. Since our model does not consider temporal aspects, no metabolic cost is imposed on a cell for having faster conduction velocities. Imposing such a metabolic cost would result in fewer faster cells with larger RF and conversely, in more slower cells with smaller RF.

ON and OFF Midget RGC In primate retinas *midget* ON-center RGC occur roughly 3 times more often than corresponding OFF types (see table 9 Appendix). This is not reproduced throughout all settings. The model results in evenly distributed numbers of ON and OFF types in each particular channel (see table 6). However interestingly, in small animals with divariant vision, a more symmetric organization of S-ON vs. L-OFF and S-OFF vs. L-ON opponent RGC exists [Gouras, 2009, p.5]. Moreover, independent of the eccentricity cells are sampled, *midget* and *parasol* RGC show about 30 to 50 % larger RF diameters than their OFF-center counterparts [Dacey and Petersen, 1992, p.9670]. This is also not reproduced throughout all settings of the model.

Death of Hidden Units during Training The occurrence of dead hidden units when training resembles the competition of dendrites among RGC during the development of the retina [Linden and Perry, 1982] [Linden, 1993]. The compact neural code resulting from the competition is beneficial from an information theoretic perspective.

Constraining Input and Output Constraining the hidden layer activity in setting (x±/y+) and settings of type (x+/y+) to solely positive values has a valid biologically motivated argument: biological neurons cannot encode positive and negative activity, thus it appears plausible to limit the capability of hidden neurons. Consequently setting (x±/y+) and settings of type (x+/y+) are converging in channels of uniform hidden units of ON or OFF signature (see figs. 25 and 26) which is indeed observed in biological retinas.

Further by subtracting the mean of an input patch in setting (x±/y+) and setting (x±/y±), which results in an input of positive and negative values, could only be biologically motivated as to be the result of the output of *photoreceptors* and *horizontal cells* combined. The *horizontal cells* functionally introduce ON and OFF aspects of the visual input, since the output of photoreceptor cells alone is solely of positive magnitude. Thus settings of (x+/y+) and setting (x+/y±) appear more biologically plausible concerning input statistics.

Positive input (setting (x+/y±)) by not subtracting the mean of the input seems to have no effect on the distribution of RF and the number of channels emerging (see table 4). However, positive input in combination with additionally constraining the output of the model to solely positive values, appears the most biologically motivated setting. Moreover, settings of (x+/y+) solely produce **chromatic** and **achromatic** differentiation of RF, which indeed is observed in biology in *midget* and *parasol* RGC types.

Chromatic Contrast and RF Surround setting (x±/y+), setting (x±/y±) and setting (x+/y±) are producing RF with *intra-channel* chromatic contrast (see fig. 12 and 29). This form of chromatic contrast appears to be biologically plausible: blue selective retinal RGC however have been observed solely with yellow center and blue surround which produces an *intra-channel* chromatic contrast. Whilst the converse, a blue center and yellow surround has not been observed in nature, it appears in setting (x±/y+), setting (x±/y±) and setting (x+/y±).

All settings fail in clearly producing the *inter-channel* chromatic contrast of *midget* RGC, whilst settings of type (x+/y+) appearing to be failing the least. This is supported by the results of the convolution filters (see fig. 32) in which prototypes of setting (x±/y+) are blind to some types of contrast and do not differentiate between chromatic and achromatic contrasts. Prototype filters of the chromatic channels of setting (x+/y+)b perform better and appear to weakly differentiate chromatic and luminous contrast. Moreover, the results of combining filter outputs (see figs. 30 and 32, lower two rows) of chromatically opponent (as found in *midget* RGC) filters reveals a clear separation of chromatic and luminous contrasts in both settings. This highlights the fact that RF tuned towards a more biological *inter-channel* contrast appear to be superior compared to RF tuned to *intra-channel* contrast.

In summary, settings of type (x+/y+) appear to be the most biologically motivated setting. However, the RF morphology of the remaining settings, matches that of *midget* RGC more closely, since the RF sizes are small compared to the rather large RF of (x+/y+). Regarding the overall intra-channel chromatic contrast that RF of the remaining settings show, this similarity in morphology to *midget* RGC does not extend into sensitivity for chromatic contrast.

8 Model of V1 Simple Cells

To demonstrate that the RGC model (see 6.2) can be directly utilized for preprocessing RGB information, we trained a second hidden layer simulating V1 simple cells [Olshausen and Field, 1997] on top of a fixed pre-trained connection map of the RGC model. The second hidden layer is trained in the same auto-associative fashion as the the RGC model (see fig. 33).

The original model of Olshausen and Field requires the training data to be whitened to properly emerge in V1 simple cells. By skipping this manual preprocessing of training data and training the V1 model directly on the output of the RGC model we can demonstrate that the RGC model indeed has whitening filter characteristics sufficient for the V1 model to learn simple cells RF properly.

8.1 Generative Model of V1 simple Cells

Training the connection weights of the V1 model takes the following steps: The input of the V1 model is computed by dotting the input x with a fixed pre-trained connection matrix $\mathsf{W_{rgc}}$ of the RGC model (22). Next, the hidden unit activity of the V1 model is computed by dotting $\mathsf{x_{rgc}}$ and connection matrix $\mathsf{W_{v1}}$ (23) and feeding the result through a sigmoid transfer-function (24), with sparsity parameter a and scaling parameter b (29). Subsequent, the input is reconstructed (25) and the reconstruction error (26) is used in the Hebbian learning rule (27) to update the connection weights $\mathsf{W_{v1}}$. Additionally, the weights of the connection matrix $\mathsf{W_{v1}}$ are reduced in every training epoch by small amount (28), specified through the weight decay parameter δ.

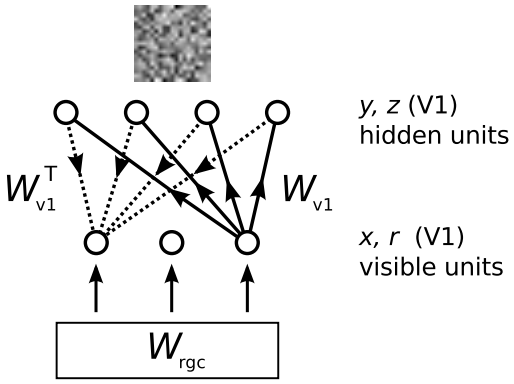

Figure 33: The V1 model. Solid arrows indicate connections from visible to the hidden units. Dotted arrows indicate input reconstruction connections from hidden to visible units.

$$\mathsf{x_{rgc}} = \gamma\ (\mathsf{W_{rgc}} \cdot \mathsf{x}) \tag{22}$$

$$\mathsf{y_{v1}} = \mathsf{W_{v1}} \cdot \mathsf{x_{rgc}} \tag{23}$$

$$\mathsf{z_{v1}} = \mathsf{g}(\mathsf{y_{v1}}) \tag{24}$$

$$\mathsf{r} = \mathsf{W_{v1}^T} \cdot \mathsf{z_{v1}} \tag{25}$$

$$\mathsf{e_{v1}} = \mathsf{x_{rgc}} - \mathsf{r} \tag{26}$$

$$\Delta\mathsf{W_{v1}} = \eta\ \mathsf{z_{v1}} \odot \mathsf{e_{v1}} \tag{27}$$

$$\mathsf{W_{v1}} = \mathsf{W_{v1}} + \Delta\mathsf{W_{v1}} - \eta\ \delta\ \mathsf{W_{v1}} \tag{28}$$

$$\mathsf{g}(\mathsf{x}) = \mathsf{b}\,(\mathsf{x} - \mathsf{a}\,\frac{\mathsf{x}}{(1.0 + \mathsf{b}^2\,\mathsf{x}^2)}) \tag{29}$$

(with: $\cdot$ inner, $\odot$ outer product, γ RGC model output amplification and η learning rate)

8.2 Results of V1 Model

Training the V1 model demands some experimentation, since setting the amplification parameter γ too low or too high, the V1 model is not capable of reconstructing the output of the RGC model properly. Thus, the reconstruction error is never minimized beyond a certain magnitude, which results in the V1 model not converging in a stable solution. Moreover, if the value of the weight decay parameter δ is set too large, the V1 model is not converging in a stable solution, but RF are of fuzzy appearance with a small population of hidden units develop more or less sharp achromatic and localized RF.

The V1 model was solely trained in a strong undercomplete setup to rapidly explore the parameter space whilst meeting time-constraints in finishing this work. Two fixed connection maps of the RGC model of setting (x±/y+) and (x±/y±) are used to pre-process the input of the V1 model (see fig. 34).

	RGC set.	$\mathsf{vis} \times \mathsf{hid}_{\mathsf{rgc}} \times \mathsf{hid}_{\mathsf{v1}}$	a	b	δ	η	γ	epochs
1	(x±/y+)	$13 \times 32 \times 12$	0.5	1.5	$1e-04$	0.05	1.25	934.400
2	(x±/y±)	$13 \times 21 \times 12$	0.9	1.5	$1e-05$	0.1	0.4	29.395.200

Table 8: Parameter values of the two V1 results. Notably, the decay parameter δ of the second result is of one magnitude smaller and is resulting approx. a 30 fold training time compared to the first result.

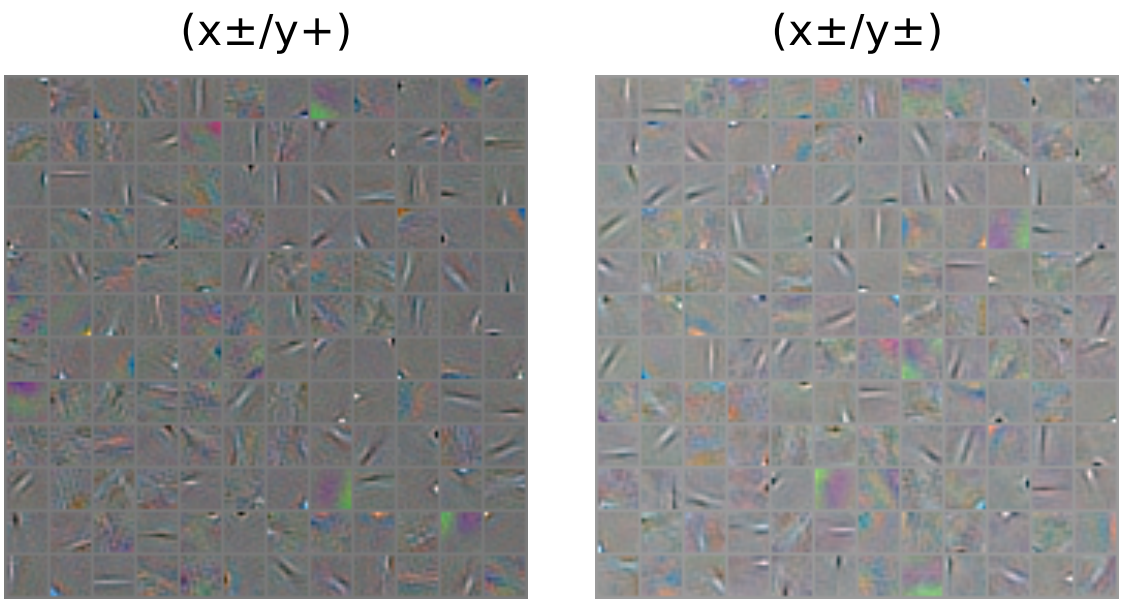

Figure 34: The figure shows the dot product of the fixed RGC model connection matrix $\mathsf{W}_{\mathsf{rgc}}$ and the learned V1 matrix W_{v1}. On the left the result of utilizing the setting (x±/y+) matrix is displayed. On the right the result of utilizing (x±/y±).

Since V1 model requires of the input to be of positive and negative values to properly function, only the output of the RGC model's settings in which the hidden layer activity is not clipped appear to be directly usable. However, by piping input through a map of setting (x±/y+) without clipping, the output is of positive and

negative values. As such, the output of setting (x±/y+) was also usable to train the V1 model, but not of biological plausibility.

Further, the learned V1 connection map of both results are converging towards a similar distribution despite different parameters (see table 8) and RGC connection matrices. This suggests, that an optimal neural code for any particular setting of the V1 model exists. During training, the V1 hidden units approximate this optimal code with decreasing reconstruction error.

8.3 Training V1 Model without Preprocessing

The linear nature of the modeled RGC and V1 cells leads to another open question: Since both trained matrices of RGC and V1 can be merged to one matrix appearing to yield a similar functionality (see fig. 34), can the merged matrix be trained in a single step?

Simplifying the V1 model to process RGB values directly, without the preprocessing of a fixed RGC matrix (skipping eq. 22), appears to be suitable to indicate the validity of the above question. If the simplified V1 model would converge in a similar distribution of localized RF as the V1 model utilizing RGC preprocessing, an argument could be made that the above question is actually true.

However obviously, by just stating in not successfully trained the simplified V1 model to emerge in a localized distribution, an argument against the validity of the above question cannot be made. A set of unsuccessful attempts only gives an indication of which parameter values do not result in the desired distribution.

Meeting time-constraints in finishing this work does not allow the exploration of the simplified V1 model's parameter space thoroughly. Thus, the goal of finding a set of parameters which possibly converges in a localized RF distribution of similar type as the regular V1 model (see fig. 34) is not pursued further in this work. Answering the question of whether the V1 model can be trained from RGB images directly, would require a mathematical proof of some kind or at least a thorough exploration of the relevant parts of the parameter space.

9 Discussion

The original work of Vincent and Baddely [Vincent and Baddeley, 2003] showed that by enforcing the reconstruction of the input with minimal weights, spatially localized RF emerge, covering the entire visual field. The results demonstrate, that our extended model pursues this property of emerging in localized RF also in the color domain.

Further, the permanent application of the weight-constraint introduces a competition among hidden units in the reconstruction of the input whilst weakly participating units eventually die off (see 6.3.3). This results in an compact and efficient neural code representing the statistics of RGB images of natural scenes.

Furthermore, the property of the RGC model as being *undercomplete* (see 6.3.5 and 5.4.1) has the beneficial effect of learning a compressed representation of the input data with minimal reconstruction error. Thus the information transmitted through the RGC model is maximized whilst being compressed. This property is of high benefit in applications of high dimensional input, such as processing RGB images with CNN, since todays cameras have resolutions of many megapixels, a number which renders the application of deep NN of being intractable in directly processing each input pixel. For example, learning from the pixels of an Atari VCS 2600 video game (210×160 RGB video at 60Hz) required reducing the input dimensionality to 84×84 units, to be tractable [Mnih et al., 2013, p.5]. Correspondingly, in biology the retina compresses the information of approximately 6.4 million cone photoreceptors to one million RGC (a rate much higher regarding rod photoreceptors) [Weber and Triesch, 2009, p.75], indicating a high rate of compression of visual information.

Even further, it has been shown that in the most biologically motivated settings (x+/y+)a and (x+/y+)b, intensity and chromatic aspects of the input are separated by the resulting prototype filters. Moreover prototypes of this setting capture biologically more plausible *intra-channel* chromatic contrast (see 7.7). This is a novelty, since established statistical methods [Brown et al., 2011, p.28] and equally the remaining settings produce filters in which intensity and chromatic aspects are not separated. Instead, a combination of intensity and chromatic aspects of the input statistics inside the boundaries of each RGB channel is learned. This produces filters capable of capturing solely *intra-channel* chromatic contrast (see 7.7).

Because of the unmatched performance the visual system produces, filters capturing a more biological chromatic contrast (see 4.2) appear to be of value. This is due to cells providing color opponent contrast are viewed as fundamental building block of the visual system in providing color constancy and as such simulating the morphology of color opponent cells appears to be beneficial. Moreover, this view is supported by the results of Yang et al. demonstrating the importance of chromatic contrast information in the task of border detection [Yang et al., 2013]. The detection of borders is the foundation of numerous basic tasks in computer vision, such as image segmentation and object recognition. Consequently, all tasks which are learning features from RGB images could (possibly) benefit by preprocessing the visual information in a biologically inspired manner.

However, the results of applying prototype RF as convolution filters (see 7.7) lead to the expectation that a model utilizing the preprocessed output of the RGC model shows improved performance. If and for which tasks of computer vision this improvement is actually true and tractable remains to be seen.

9.1 Utilizing the RGC-Model as Preprocessor

The functional features the model provides are: (the first two bullets are properties extended from the original model of Vincent and Baddely, since the emerging RF are of the same DOG texture)

- compression whilst maximizing information transmission
- bandpass filtering
- separation of spatial and chromatic aspects
- separation of intensity and chromatic aspects whilst capturing more biologically plausible color contrasts by training in settings like (x+/y+)a and (x+/y+)b

It is possible to directly utilize a trained connection map as a first stage of neural processing to filter RGB information upon which another model can learn less general higher level features. For example the separated and whitened output of the RGC model can be directly used to train V1 simple cells (see 8). However, due to the enormous training time (see table 3) whilst the visual space being rather small (13x13) the model appears to be unattractive to be used directly as preprocessing stage for RGB information.

Nevertheless, by extracting prototype RF of a trained connection map (see 7.6) these can be used directly as convolution filters in widely used convolutional neural networks (CNN), possibly improving its performance.

Despite the limited direct application capabilities of the model, the morphology of the extracted convolution filters can be used to construct artificial filters with similar or even better characteristics of detecting chromatic contrast (see 7.7).

9.2 Future Work

Numerous paths are thinkable in extending and exploring the model. In the following a brief overview is given:

A thorough comparison with the denoising auto-encoder model would be of interest, since being trained on MNIST data (not images of natural scenes), center surround RF among other forms emerge [Vincent et al., 2008]. The addition of the denoising property as another possible constraint to the RGC model, could be trivially accomplished.

Extending the model in the temporal domain together with constraining the transmission speed of the hidden units, in which faster transmission of information comes with a higher metabolic cost [Baden et al., 2014, p.2]. Such a model would

be trained with videos or short image sequences instead of still images. It is expected that such a model will converge in a more biologically plausible distribution of RGC types, in which units tuned to achromatic aspects will be of small number but with larger RF compared to units tuned to chromatic aspects (see 7.8).

Instead of training the model on RGB image patches, the model can be extended to be trained upon higher dimensional input, such as RGB-D (Kinekt) or FMRI data. The compression the model provides, should be beneficial in these domains.

Extending the model to include scotopic (low light, night) vision. It is completely unknown how such an extended auto-encoder would learn these very divergent image statistics. Such an auto-encoder would have its entire daylight vision input near zero when a scotopic image patch is presented and conversely its night vision input entirely saturated when a daylight image patch is presented. Despite lacking exact knowledge of how *amacrine* cells mediate rod (night) and cone (day) input onto the same ganglion cells, the larger RF size of achromatic RGC could also be hypothesized to be the result of being required to sample sufficient light of a large number of rods. This gives a different argument of why RF sizes of achromatic RGC types are larger compared to color selective RGC types, than the argument of being the result of faster transmission of information at a higher metabolic cost.

The RGC model could be possibly of use in simulating some aspects of color blindness since the model can be trained in various input modes (see 6.3.2), in which some are simulating dichromatic vision. This might be of relevance since the self organizing property of the RGC model results in the emergence of a lesser number of RGC channels when trained in modes of dichromatic vision. Consequently, the same number hidden units are populating a lesser number of RGC channels. Hence, the spatial resolution of the RGC model in dichromatic mode is higher compared to a model trained in RGB mode of equal dimension. This conjecture of the RGC model could be biologically verified by analyzing data obtained of color blind and healthy retinas.

References

[Adams and Horton, 2003] Adams, D. L. and Horton, J. C. (2003). A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *The Journal of Neuroscience*, 23(9):3771–3789.

[Ahmad et al., 2003] Ahmad, K. M., Klug, K., Herr, S., Sterling, P., and Schein, S. (2003). Cell density ratios in a foveal patch in macaque retina. *Visual Neuroscience*.

[Baden et al., 2014] Baden, T., Nikolaev, A., Esposti, F., Dreosti, E., Odermatt, B., and Lagnado, L. (2014). A synaptic mechanism for temporal filtering of visual signals. *PLOS biology*, 12(10).

[Baldi and Homik, 1995] Baldi, P. F. and Homik, K. (1995). Learning in linear neural networks: A survey. *IEEE Transactions on Neural Networks*, 6(4).

[Baldi and Hornik, 1989] Baldi, P. F. and Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2:pp. 53–58.

[Barlow, 1989] Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*.

[Bell and Sejnowski, 1997] Bell, A. J. and Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*.

[Bengio, 2009] Bengio, Y. (2009). Learning deep architectures for ai. *Dept. IRO, Université de Montréal Technical Report 1312*.

[Brown et al., 2011] Brown, M., Süsstrunk, S., and Fua, P. (2011). Spatio-chromatic decorrelation by shift-invariant filtering.

[Dacey and Petersen, 1992] Dacey, D. M. and Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings National Acadamy of Science USA*, 89:pp. 9666–9670.

[Demonasterio and Gouras, 1975] Demonasterio, F. M. and Gouras, P. (1975). Functional properties of ganglion cells of the rhesus monkey retina. *The Journal of Physiology*, pages p. 167 – 196.

[Doi et al., 2012] Doi, E., Gauthier, J., Field, G., Shlens, J., Sher, A., Greschner, M., Machado, T., Jepson, L., Mathieson, K., Gunning, D., Litke, A., Paninski, L., Chichilnisky, E. J., and Simoncelli, E. (2012). Efficient coding of spatial information in the primate retina. *The Journal of Neuroscience*, 32(46):16256–16264.

[Ebrey and Koutalos, 2001] Ebrey, T. and Koutalos, Y. (2001). Vertebrate photoreceptors. *Progress in the Retinal and Eye*, 20(1):49–94.

[Felleman and Essen, 1991] Felleman, D. J. and Essen, D. C. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*.

[Field and Chichilnisky, 2007] Field, G. D. and Chichilnisky, E. J. (2007). Information processing in the primate retina: Circuitry and coding. *Annual Review of Neuroscience*.

[Field et al., 2010] Field, G. D., Gauthier, J. L., Sher, A., Greschner1, M., Machado1, T., Jepson, L. H., Shlens, J., Gunning, D. E., Mathieson, K., Dabrowski, W., Paninski5, L., Litke, A. M., , and Chichilnisky, E. (2010). Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, 467(10):673–677.

[Fu, 2010] Fu, Y. (2010). *Phototransduction in Rods and Cones*. Webvision: The Organization of the Retina and Visual System [Internet]. Salt Lake City (UT): University of Utah Health Sciences Center; 1995-, http://www.ncbi.nlm.nih.gov/books/NBK52768/.

[Gegenfurtner, 2003] Gegenfurtner, K. R. (2003). Cortical mechanisms of color vision. *Neuroscience Nature Reviews*, 4:563–572.

[Gehler et al., 2008] Gehler, P. V., Rother, C., Blake, A., Minka, T., and Sharp, T. (2008). Bayesian color constancy revisited.

[Gouras, 2009] Gouras, P. (2009). *Color Vision*. Webvision: The Organization of the Retina and Visual System [Internet]. Salt Lake City (UT): University of Utah Health Sciences Center; 1995-, http://www.ncbi.nlm.nih.gov/books/NBK11537/.

[Kolb, 2004] Kolb, H. (2004). How the retina works. *American Scientist*, 91:28–35.

[Kolb, 2012] Kolb, H. (2012). *Photoreceptors*. Webvision: The Organization of the Retina and Visual System [Internet]. Salt Lake City (UT): University of Utah Health Sciences Center; 1995-.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks.

[Krüger et al., 2010] Krüger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., guez Sa nchez, A. J. R., and Wiskott, L. (2010). Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *Review*.

[Linden, 1993] Linden, R. (1993). Dendritic competition in the developing retina: ganglion cell density gradients and laterally displaced dendrites. *Neuroscience*, 10(2):313–337.

[Linden and Perry, 1982] Linden, R. and Perry, V. (1982). Ganglion cell death within the developing retina: a regulatory role for retinal dendrites? *Neuroscience*, 7(11):2813–2840.

[Martin and Grünert, 2013] Martin, P. R. and Grünert, U. (2013). Color signals in the retina and lateral geniculate nucleus of marmoset monkeys. *Psychology and Neuroscience*, 6(2):p. 151–163.

[Masland, 2001] Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience*, 4(9).

[Mnih et al., 2013] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning.

[Nelson, 2007] Nelson, R. (2007). *Visual Responses of Ganglion Cells.* Webvision: The Organization of the Retina and Visual System [Internet]. Salt Lake City (UT): University of Utah Health Sciences Center; 1995-., http://www.ncbi.nlm.nih.gov/books/NBK11550/.

[Olshausen and Field, 1997] Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set - a strategy employed by V1? *Vision Research*, 37(23).

[Purves, 2001] Purves, D. (2001). *Neuroscience. 2nd edition.* Sunderland (MA): Sinauer Associates; 2001, http://www.ncbi.nlm.nih.gov/books/NBK10806/.

[Purves et al., 2002] Purves, D., Lotto, R. B., and Nundy, S. (2002). Why we see what we do. *American Scientist*, 90(3).

[Schmolesky, 2007] Schmolesky, M. (2007). *The Primary Visual Cortex.* Webvision: The Organization of the Retina and Visual System [Internet]. Salt Lake City (UT): University of Utah Health Sciences Center; 1995-., http://www.ncbi.nlm.nih.gov/books/NBK11524.

[Shapley and Hawken, 2011] Shapley, R. and Hawken, M. J. (2011). Color in the cortex: single- and double-opponent cells. *Vision Research*, 51:701–717.

[Tailor et al., 2000] Tailor, D. R., Finkel, L. H., and Buchsbaum, G. (2000). Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40.

[Vincent et al., 2008] Vincent, Larochelle, Bengio, and Manzagol (2008). Extracting and composing robust features with denoising autoencoders. int conf on machine learning. *Prroceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 2008.*

[Vincent and Baddeley, 2003] Vincent, B. T. and Baddeley, R. J. (2003). Synaptic energy efficiency in retinal processing. *Vision Research*, 43.

[Vincent et al., 2005] Vincent, B. T., Baddeley, R. J., Troscianko, T., and Gilchrist, I. D. (2005). Is the early visual system optimised to be energy efficient? *Network: Computation in Neural Systems*, 16(2).

[Walsh et al., 1999] Walsh, N., Fitzgibbon, T., and Ghosh, K. (1999). Intraretinal axon diameter: a single cell analysis in the marmoset (callithrix jacchus). *Journal of Neurocytology*, 28(12):989–998.

[Weber and Triesch, 2008] Weber, C. and Triesch, J. (2008). A sparse generative model of V1 simple cells with intrinsic plasticity. *Neural Computation*, 20.

[Weber and Triesch, 2009] Weber, C. and Triesch, J. (2009). Implementations and implications of foveated vision. *Recent Patents on Computer Science*, 2009(2):75–85.

[Wei and Wu, 2013] Wei, H. and Wu, H. (2013). A neurocomputing model for ganglion cell's color opponency a neurocomputing model for ganglion cell's color opponency mechanism and its application in image analysis. *Proceedings of International Joint Conference on Neural Networks*.

[Yang et al., 2013] Yang, K., Gao, S., Li, C., and Li, Y. (2013). Efficient color boundary detection with color-opponent mechanisms. *Computer Vision Fondation CVPR2013*.

10 Appendix

cell class	num		%	assumed RGC Type
1. Colour-opponent concentric, (61%)				
M+ / L-	95	Green-ON	20.79	Midget
M+ / (S+L)-	4	Green / Yellow+Cyan	0.88	
M- / L+	30	Green-OFF	6.56	Midget
M- / (S+L)+	3	Magenta / Blue+Red	0.66	
L+ / M-	76	Red-ON	16.63	Midget
L+ / (S+M)-	3	Red / Yellow+Magenta	0.66	
L- / M+	27	Red-OFF	5.91	Midget
L- / (S+M)+	2	Cyan / Blue+Green	0.44	
(M+L)+ / S-	6	Red+Green / Yellow	1.31	Midget
(M+L)- / S+	10	Cyan+Magenta / Blue	2.19	Midget
S+ / (M+L)-	17	Blue / Cyan+Magenta	3.72	Midget
S- / (M+L)+	4	Yellow / Red+Green	0.88	
2. Colour-opponent, non-concentric (2%)				
S+ / (M+L)-	5	Blue / Magenta+Cyan	1.09	
S- / (M+L)+	1	Yellow / Red+Green	0.22	
L+ / M-	3	Red / Magenta	0.66	
3. Broad-band, non-opponent (24%)				
ON / OFF	69	ON	15.1	Parasol
OFF / ON	41	OFF	8.97	Parasol
4. Broad-band, colour-opponent (4%)				
(M+L)+ / L-	11	Red+Green / Cyan	2.41	Small Bistratifie
(M+L)+ / M-	4	Red+Green / Magenta	0.88	Small Bistratifie
(S+M+L)+ / M-	2	white / Magenta	0.44	
(M+L)- / L+	2	Cyan+Magenta / Red	0.44	Small Bistratifie
(M+L)- / M+	1	Cyan+Magenta / Green	0.21	Small Bistratifie
5. Non-concentric. phasic (6%)				
ON	10		2.19	
OFF	3		0.66	
ON-OFF	14		3.06	
6. Non-concentric, motion-sensitive (3%)				
unidentified	14		3.06	

Table 9: Retinal distribution of 460 RGC units (457 classified) from the rhesus monkey retina [Demonasterio and Gouras, 1975, p.190]. Because of readability, the notation of cell classes is slightly changed and only the sum of over all eccentricities for each class is given here.

setting	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6
(x±/y+)	152	135	165	132	177	155
(x±/y±)	126	101	173			
(x+/y±)	146	96	158			
(x+/y+)a	159	135	76	62	123	
(x+/y+)b	167	153	78	63	114	84

Table 10: The raw number of clustering results: The six clusters of setting (x±/y+) are three ON channels of red (1), green (2), blue (3) and three OFF channels of cyan (red OFF) (4), magenta (green OFF) (5), yellow (blue OFF) (6). The three complete channels of setting (x±/y±) and setting (x+/y±) are combined ON and OFF channels of red (1), green (2) and blue (3). Further, the five channels of setting (x+/y+)a are white (luminosity ON) (1), black (luminosity OFF) (2), red (3), green (4) and blue (5). Finally, the six channels of setting (x+/y+)b are white (luminosity ON) (1), black (luminosity OFF) (2), red (3), green (4), cyan (red OFF) (5) and magenta (green OFF) (6).

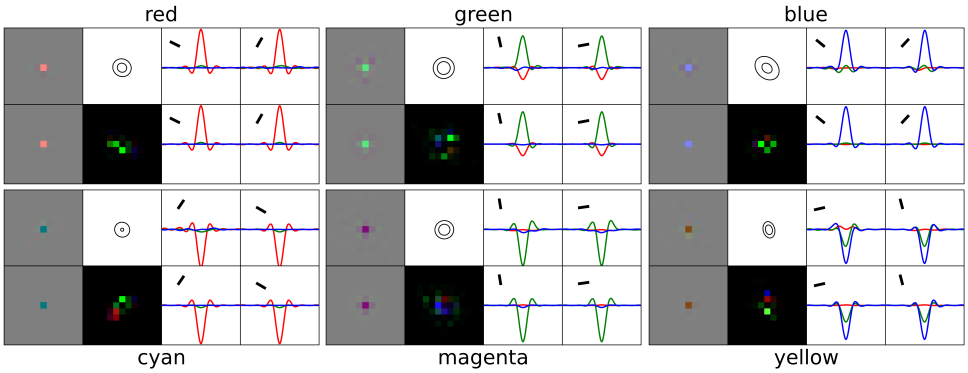

Figure 35: **Setting (x±/y+)**: Each box of eight tiles shows the prototype RF (top left), the reconstruction generated of the best fit (bottom left), the ellipse of the best fit (top, 2nd column) and the squared reconstruction error of the RF and the reconstruction (bottom, 2nd column). The third column shows an interpolated cut of the primary axis of the best fit ellipse for the RF (top) and the reconstruction (bottom). Whereas the fourth column shows the secondary axis of the ellipse of the RF (top) and the reconstruction (bottom).

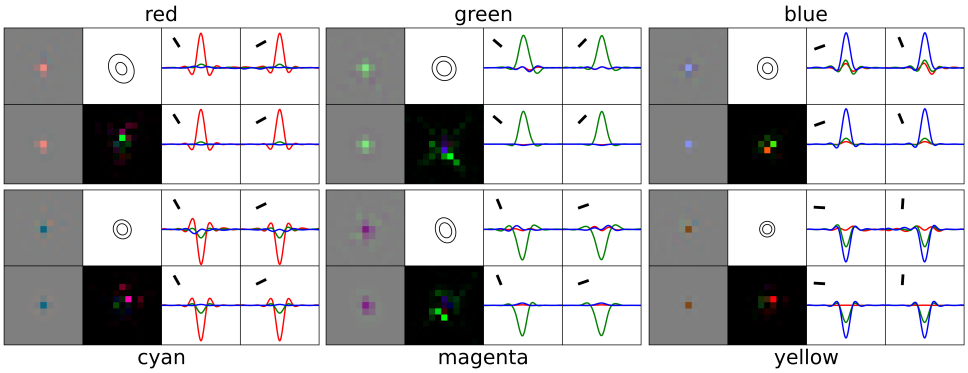

Figure 36: Prototype RF for **Setting (x±/y±)**. red, green, blue (1st row) and cyan, magenta and yellow (2nd row) channels.

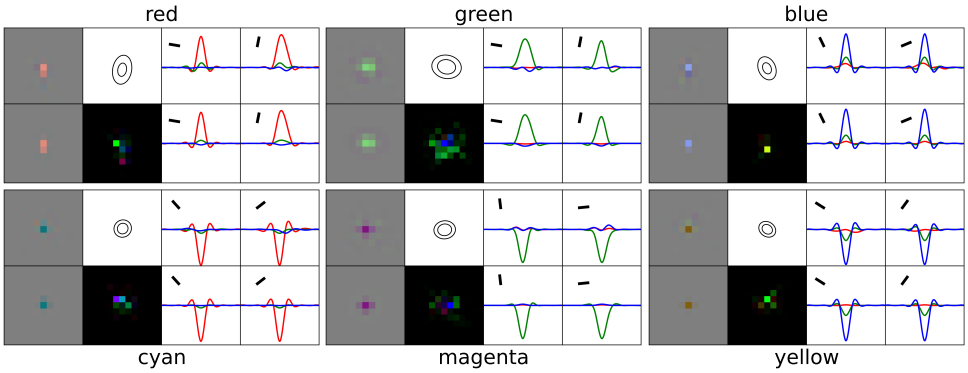

Figure 37: **Setting (x+/y±)**: prototypical RF for red, green, blue (1st row) and cyan, magenta and yellow (2nd row) channels.

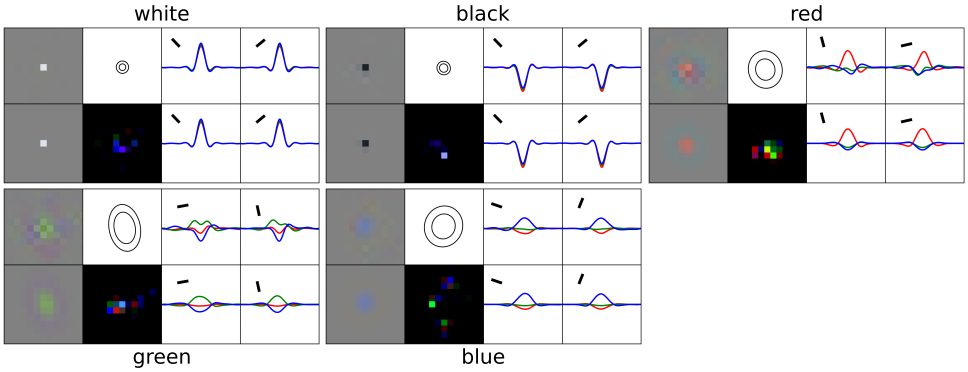

Figure 38: **Setting (x+/y+)a**: prototypical RF for white, black (1st row), red, green, blue (2nd row) channels.

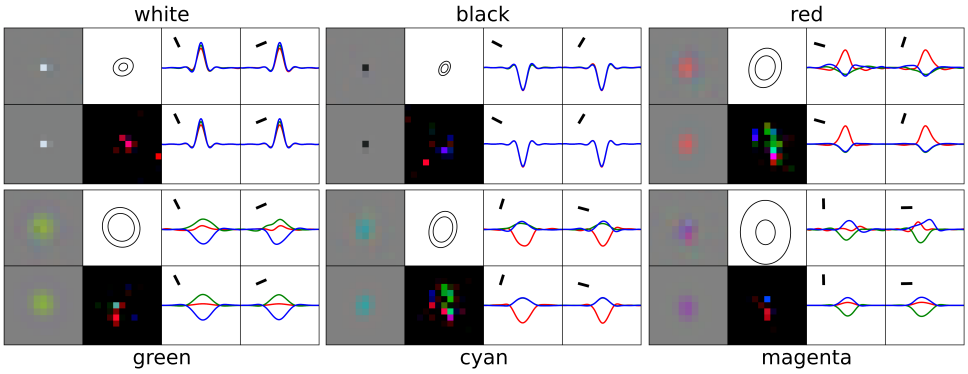

Figure 39: **Setting (x+/y+)b**: prototypical RF for white, black, red (1st row), green, cyan and magenta (2nd row) channels.

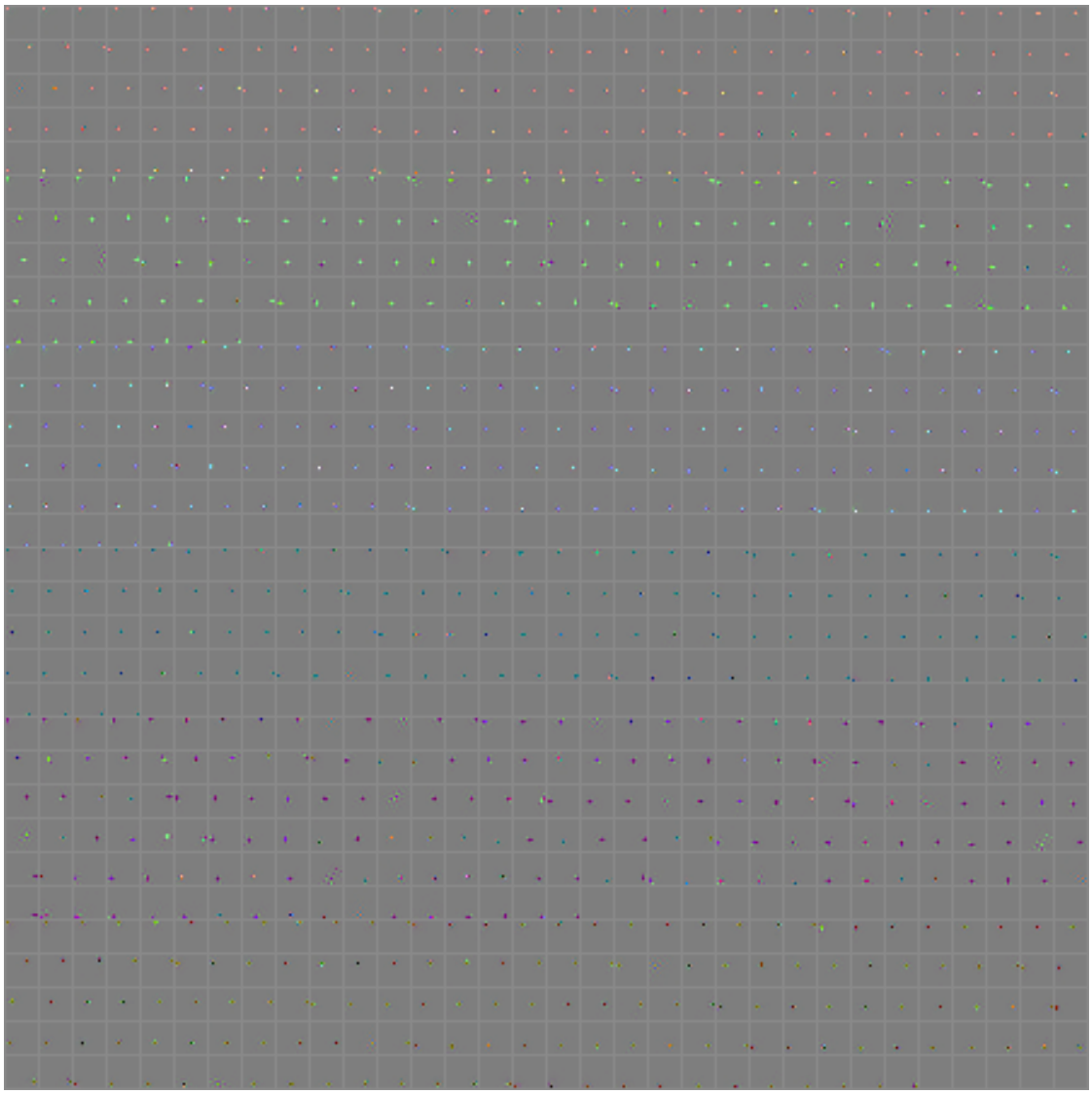

Figure 40: Cconnection map of **setting** (x±/y+). All maps have been pruned of dead hidden units and are grouped and sorted upon the results of the fitting and clustering process.

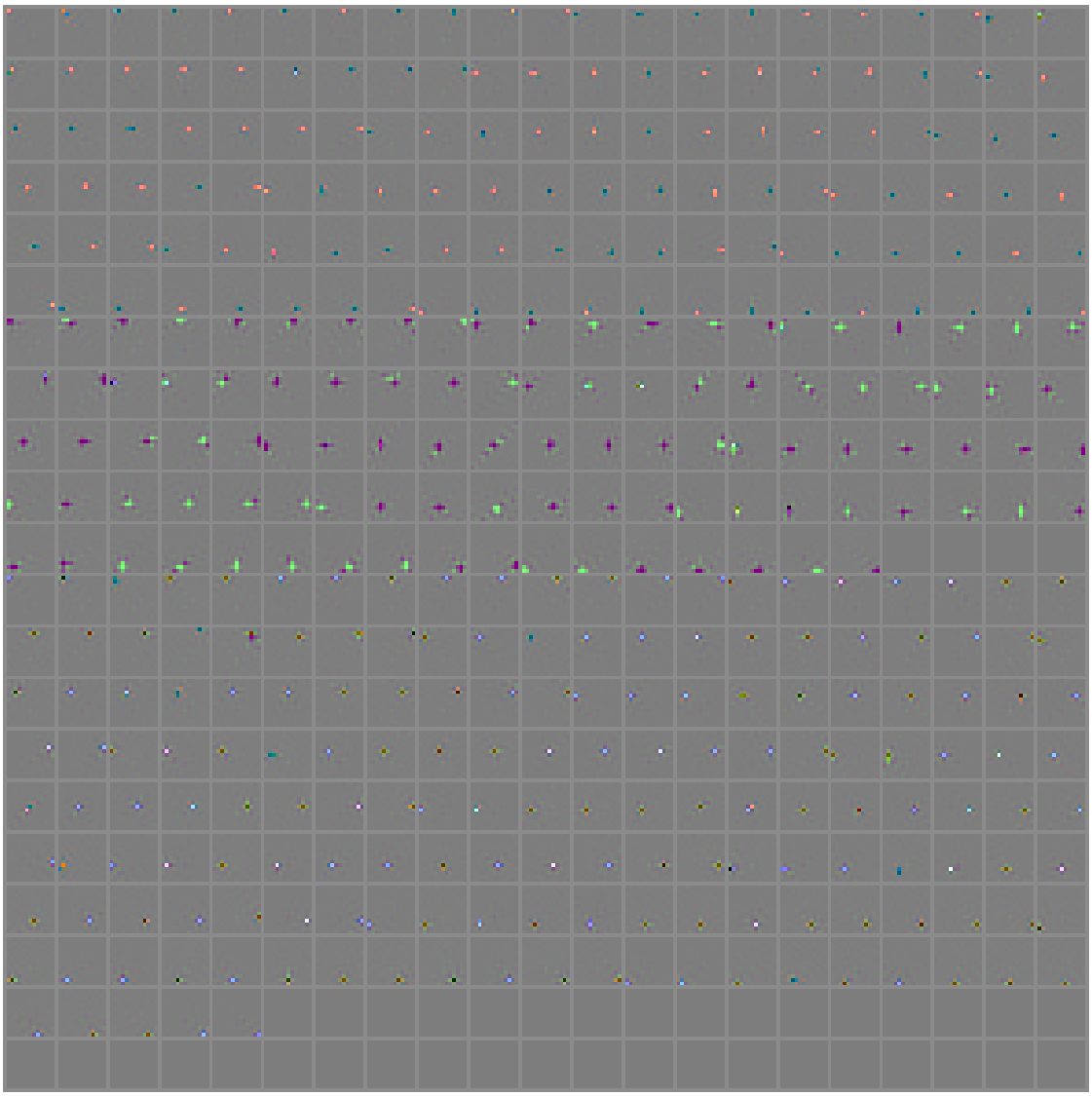

Figure 41: Connection map of **setting** (x±/y±).

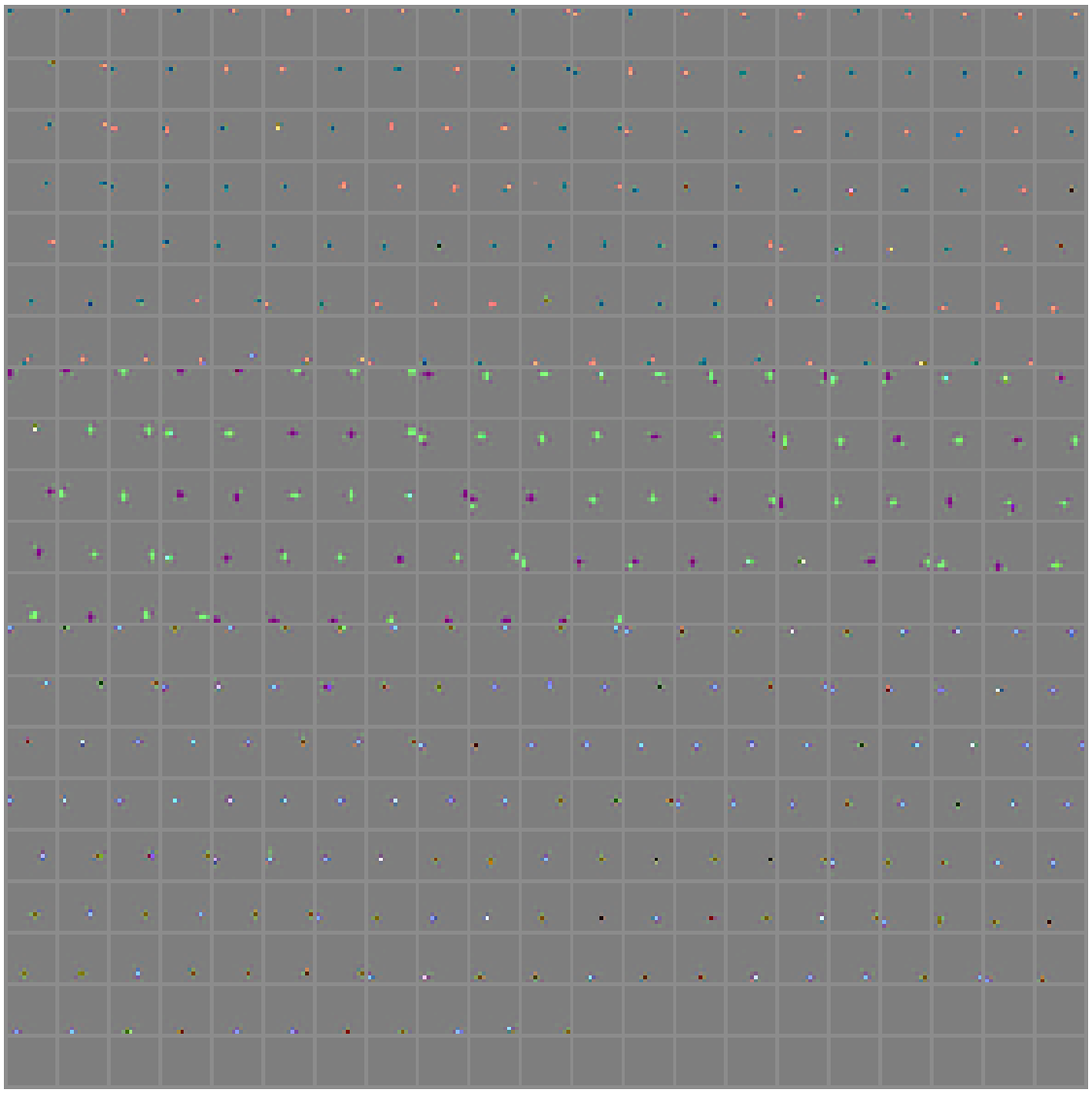

Figure 42: Connection map of **setting** (x+/y±).

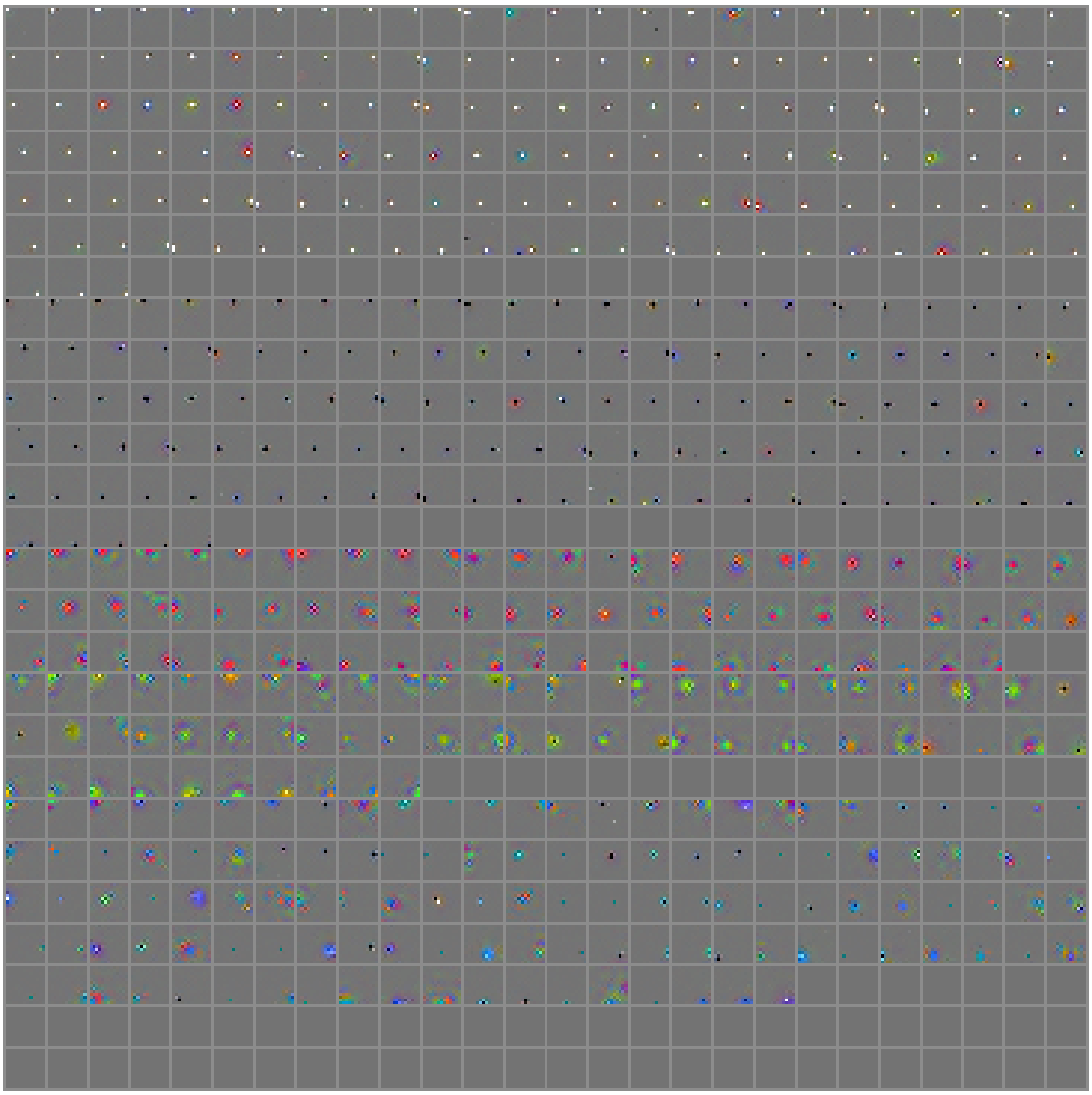

Figure 43: Connection map of **setting (x+/y+)a**.

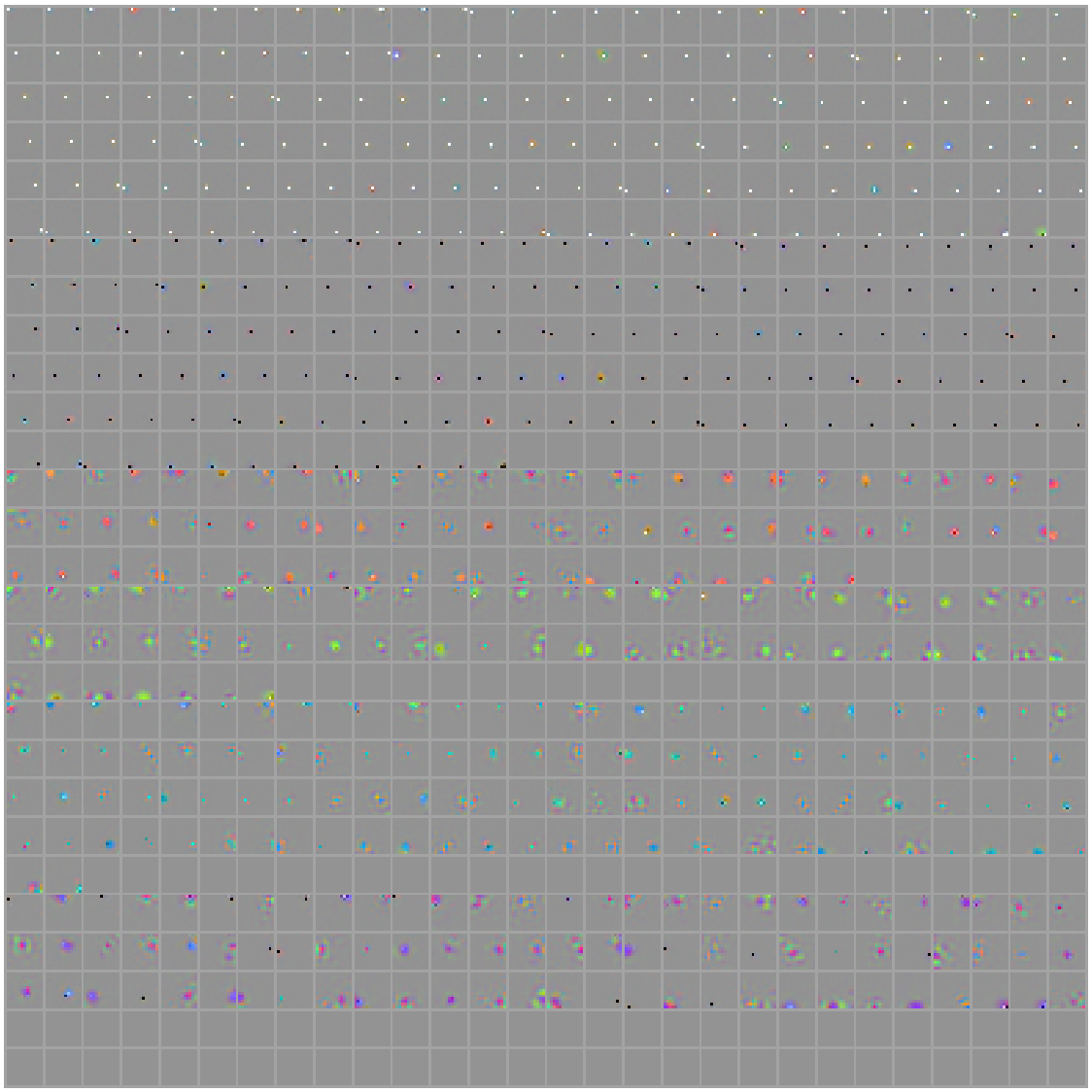

Figure 44: Connection map of **setting (x+/y+)b**.

Figure 45: The figure is showing thumbnails of all 293 RGB-images in the training set. Images of mostly reddish and brownish color appear to be in the majority.

Eigenständigkeitserklärung

Ich versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Einverständniserklärung

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Departments Informatik einverstanden.