Noname manuscript No.
(will be inserted by the editor)

A Compressing Autoencoder Model of Retinal Coding for Color Images

Daniel von Poschinger-Camphausen · Cornelius Weber · Stefan Wermter

Received: date / Accepted: date

Abstract The retina has remarkable properties of efficiently encoding visual information. In particular, various aspects of visual processing are separated into distinct channels, which manifest in specific retinal ganglion cell (RGC) types. The exact principles of the development of RGC types and retinal circuitry are neither well understood nor technically exploited; in particular an account of how learning contributes to the formation of these channels is missing. Thus, we implemented a generative computational model of self-organizing RGC receptive fields (RFs) by constraining synaptic activity in a minimal setup. We extended it to process color images, and trained it with images of natural scenes. The training results show a large variance in the distribution of RF mosaics and corresponding channels, which can be attributed to the simplicity of our model, although it does not allow an exact identification of biological RGC types. Nonetheless, the model reliably produces spatially and chromatically localized color opponent RFs, similar to biological RGC receptive fields.

Keywords Retina · Color vision · Retinal ganglion cells · Energy efficiency · Chromatic contrast · Unsupervised learning

Daniel von Poschinger-Camphausen
E-mail: dvpc@protonmail.com

Cornelius Weber
E-mail: weber@informatik.uni-hamburg.de

Stefan Wermter
E-mail: wermter@informatik.uni-hamburg.de

University of Hamburg
Knowledge Technology, Dept. of Informatics
Vogt-Kölln-Straße 30
D - 22527 Hamburg, Germany
www.informatik.uni-hamburg.de/wtm

1 Introduction

The retina of the human eye is often compared to a photo-sensor of a digital camera. Such a comparison neglects that several retinal neural layers perform pre-processing of the image stream such as compression, noise removal and contrast enhancement [14] [37]. As opposed to the output of luminance and color information of a CCD sensor, the retina has at least seventeen different types of retinal ganglion cells [10], indicating that the 'output' of the retina is far more complex and richer in features than that of a simple camera.

A significant feature of retinal coding is to provide the foundation of robust color vision. In human visual perception, color appears to be relatively stable to chromatic changes in the illumination of a scene. The robustness of color perception is referred to as color constancy [3] [28]. By contrast, the color information of a camera is unreliable. Even a subtle change in the illumination of a scene can lead to drastic changes in the color-values sampled by a camera sensor. How the visual system establishes this perceptual phenomenon at the cellular level, is still not fully understood and subject to discussion [31]. Still, chromatic contrast emanating from color-opponent cells, has been identified as fundamental building block of color constancy. Color opponent cells are found in the ganglion cell layer of the retina and in the primary visual cortex.

In the ganglion layer, RGCs antagonistically integrate signals of different cone types, emanating from the same area of the visual field [13]. Correspondingly, RGC receptive fields produce a chromatically contrasted representation of the input signal and are of center-surround texture, which can be approximated best by a Difference of Gaussian (DOG) function [14]. Furthermore, RGCs are contrast-selective, meaning their maximum excitation is tuned to specific contrasts of spatial frequencies in the visual stream [25]. Contrast selectiveness has several beneficial filtering properties. Correlations in the signals of different cone types are removed [12], which results in a compressed representation of the input. Compression reduces the sensitivity of the receiving neural structures to over-fitting [15], and due to the RGC band-pass characteristics of DOG shaped receptive fields, noise is filtered, (or whitened,) from input signals [26]. Thus, contrast selective RGC free the visual cortex of processing less informative components of the visual stream.

Another significant feature of retinal coding is the de-correlation of different aspects of the visual stream into distinct channels. Neuroanatomical studies have shown that the mammalian retina consists of “many parallel, anatomically equipotent microcircuits” [20], forming discrete pathways which process different aspects of the visual stream. These pathways emanating from the retina manifest in distinct RGC types, which are classified by their physiology, morphology, dendritic connectivity and to which regions their axons project [10].

In each class of morphologically, distinct RGC types, the RFs cover the entire retinal space uniformly with little overlap. This uniform mosaic structure (see Figure 1) is assumed to enable a regular sampling of the visual field [10]. Five cell types account for 75% of all RGCs: ON and OFF *parasol*, ON and

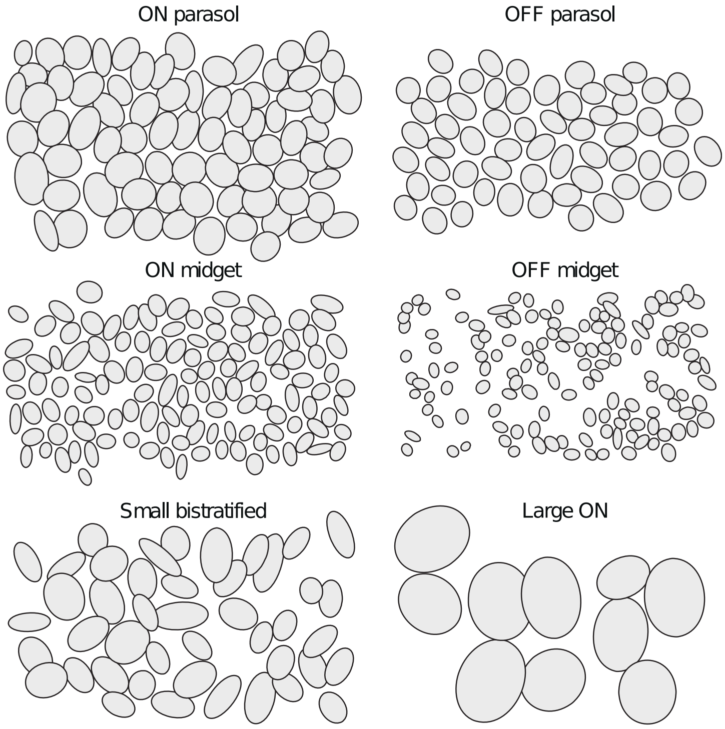

Fig. 1 Mosaics of RFs of six RGC types in the primate retina [10]. Each mosaic was obtained from a single 512-electrode recording from isolated peripheral primate retina, in which each ellipse shows the 1.3 SD contour of a Gaussian fit to the RF of a single cell.

OFF *midget*, and *small bistratified* cells [11]. However, numerous additional RGC types exist (see Table 4). The de-correlation into ON and OFF channels, together with the contrast sensitivity of RGCs, allows later stages of neural computation, e.g. V1, to define precise edges [14].

Due to the remarkable capabilities of the visual system, computational models exploiting properties of retinal coding by learning from images taken by digital cameras appear to be of value. Thus, the question we are addressing here is whether a simple computational model can predict these channels and chromatic contrasts by learning from images of natural scenes.

2 Methods

Previous developmental models of learning receptive fields focused mainly on cortical V1 simple cells, based on ICA, e.g. [2] or sparse coding, [26]; [34]; [36]. Recently, the retina has attracted more attention among computational modelers.

Focusing on compression and noise removal aspects of retinal coding, the model of Doi and colleagues [8] learns receptive fields of RGCs directly from activations emanating from a constructed cone mosaic. A different emphasis on putative retinal functions is proposed by Maul et al. Their model is learning intra retinal circuits between photo receptors and horizontal cells [21]. Later stages of retinal computation, namely amacrine and ganglion cell layers, are not considered. Nevertheless, the model predicts retinal functions such as contrast control, non-blurry de-noising and re-saturation of the input signals.

More recently, deep learning methods have been applied resulting in more complex generative models. The four layer deep belief Neural Network (NN) of Turcsany et al. considers all major retinal cell types and aims to demonstrate its suitability for modeling feature detectors similar to RGCs [32]. Notably, the model was trained with RGB images of simulated light patterns, unlike commonly training the model with image statistics derived of natural scenes. Similarly, the six-layer-deep convolutional NN of McIntosh et al. was trained with white noise [22]. However, the emphasis of the model diverges from Turcsany et al. by learning to predict temporal spiking responses of RGCs to commonly used experimental stimuli, instead of learning static receptive fields.

2.1 Autoencoder model of retinal ganglion cells

For learning retinal ganglion cells, a model based on synaptic energy efficiency was proposed by Vincent and Baddely. They utilized a simplified linear autoencoder in order to explore the effect of minimizing metabolic costs while learning optimal filters to represent natural scenes. The model consists of a single hidden layer, where units in the visible layer functioning as photoreceptors feed gray scale image patches of natural scenes to the hidden layer [33]. In the absence of a non-linear transfer function Vincent and Baddely showed that by solely constraining the connection weights of the two layers, localized DOG shaped RFs emerge, organized in one weakly overlapping mosaic, resembling those of RGCs.

2.2 Extended RGC model

We extended the model of Vincent and Baddely, based on previous work [27]: instead of processing gray scale images, the model processes RGB images (see Figure 2). Furthermore, the original model applied the weight constraint to a RF only if the sum of the absolute weight values of this RF was larger than a specific threshold, leaving the RF unchanged otherwise. We removed this threshold mechanism, always applying the constraint to every receptive field (see also sections 2.2.4 and 4.3). Apart from the number of visible units vis and the number of hidden units hid, the model has only three hyper-parameters: strength k and shape p of the metabolic weight-constraint, and learning rate η.

The algorithm for learning the connection weights W in the generative model consists of five steps (see Table 1): For an input vector x and a weight

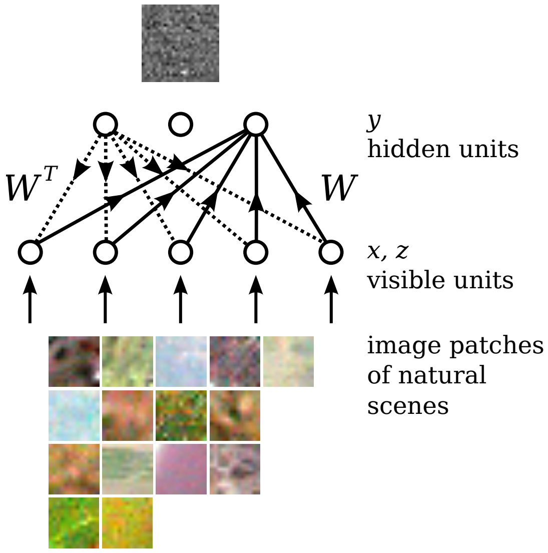

Fig. 2 The RGC model extended to color images. The visible units together represent an image patch, the hidden units represent RGCs. Solid arrows indicate connections from the visible to the hidden units. Dotted arrows indicate generative connections from the hidden to the visible units.

Table 1 All equations of the full algorithm. Here, $\cdot$ denotes the inner product, $\odot$ denotes the outer product, W_j is the synaptic weight vector of hidden unit j. Notably, in the case of using rectified linear units (ReLU) Eq. (1) effectively introduces non-linearity into the otherwise linear model (see Section 2.2.6).

$$y = \begin{cases} W \cdot x & \text{linear} \\ max(0, W \cdot x) & \text{ReLU} \end{cases} \tag{1}$$

$$z = W^T \cdot y \tag{2}$$

$$e = x - z \tag{3}$$

$$\Delta W = \eta \, e \odot y \tag{4}$$

$$\Delta W_j = \eta \, (-k \ \text{sign}(W_j) \ |W_j|^{p-1}) \tag{5}$$

matrix W, the hidden activity y is computed in Eq. (1). The input is reconstructed using generative weights W^T in Eq. (2), yielding the reconstruction z. The reconstruction error e, (Eq. (3)), is used in the Hebbian learning rule, (Eq. (4)), to update W. Afterwards the weight constraint, (Eq. (5)), is applied to the RF of each hidden neuron j.

A combined cost function (see Table 2), consisting of a quadratic reconstruction error term and a synaptic energy term, is minimized during training of the model. A minimized cost function indicates that the training of the model has converged and correspondingly the shape of the developed RFs has stabilized.

Table 2 The combined cost function (Eq. 10), consisting of a quadratic reconstruction error term, Eq. (6), and a synaptic energy term, Eq. (8). The partial derivatives Eq. (7) and Eq. (9) conform to the already given Eq. (4) and Eq. (5).

$$E_{rec} = \frac{1}{2}\,(x - W^T \cdot y)^2 \quad = \frac{1}{2}\,e^2 \tag{6}$$

$$-\frac{\partial}{\partial W}\,E_{rec} = (x - W^T \cdot y)\;y \quad = e \odot y \tag{7}$$

$$E_{syn} = \sum_{j}^{hid} \frac{1}{p}\;\mathrm{sign}(W_j)\;|W_j|^p \tag{8}$$

$$-k\,\frac{\partial}{\partial W_j}\,E_{syn} = -k\;\mathrm{sign}(W_j)\;|W_j|^{p-1} \tag{9}$$

$$E = E_{rec} + k\,E_{syn} \tag{10}$$

2.2.1 Training Data

The training set contained 293 RGB-images of natural scenes, originating from the *McGill Calibrated Colour Image Database.**[1] Additionally, a second set of only 17 images was used, taken from a botanic garden. Nevertheless, we observed that changing the dataset does not appear to make a significant difference (see section 4.1). We downscaled the images in the training set to avoid learning artifacts from the geometric configuration of red, green and blue components of CCD photo-sensors.

2.2.2 Input

The input of the model is a vector of m units, which is derived from a patch cut from a training image. A vector is created from the following steps: First, a random image from the training set 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. Values of an image patch are only normalized; other means of preprocessing, e.g. subtracting its mean value, rotation or mirroring, are not applied.

Second, the patch is vectorized, its two-dimensional structure is transformed to a flat vector. A square $(n \times n)$ patch of n pixels of a RGB color image, results in a flattened vector of $m = n^2 * 3$ units.

2.2.3 Shape of the weight constraint

The parameter p sets the shape of the weight constraint, defining how weights of an individual RF are affected: $p = 1$: the corresponding cost function resembles a L1-norm, effectively reducing all weights of a RF equally. This affects in particular the small weights and thus leads to a sparse distribution. $p = 2$: the

[1] * http://tabby.vision.mcgill.ca

L2-norm in the cost function penalizes weights proportionally to their values, due to its parabolic / spherical shape. In contrast to L1, this will not change the shape of a neuron's weight vector. $p = 1.5$: We found that this compromise led to the best results in terms of allowing a significant RF antagonistic surround to evolve. A L1-norm resulted in localized RF but with weak surrounds; a L2-norm did not result in localized RF at all.

2.2.4 Strength of the weight constraint

The parameter k defines the strength of the weight constraint. If not constraining the connection weights, the resulting RFs become non-localized and cover the whole input [33]. The application of the weight constraint has the effect of enforcing the input-reconstruction with minimal weights of each individual receptive field. This, in turn, has the desired effect of each unit learning distinct aspects, or features, of the input statistics, whereas spatial localization is the most prominent. If k has a large value, all weights of a unit's RF can be reduced to zero. Such a unit is considered dead. Dead units do not occur with the threshold mechanism of the model by Vincent and Baddely. The threshold mechanism prevents that all weights of a units RF are reduced below a certain value [33].

2.2.5 Model under- and overcompleteness

A model which has more hidden than visible units ($vis < hid$) is said to be *overcomplete* resulting in a neural code with some degree of redundancy, which is suitable for models of V1 [26]. Conversely, a model which employs more visible than hidden units ($vis > hid$) is said to be *undercomplete*, encoding the presented statistics in a compressed manner, possibly incomplete [1]. We observed that our model required *undercomplete* parameter settings to converge in RFs of DOG texture with a noticeable surround part.

2.2.6 The rectification of hidden unit activity

Using a ReLU transfer-function [15] in Eq. (1) constrains the activity of the hidden units to solely positive values. This has a biological motivation, since the synaptic activity of biological neurons is modeled commonly by scalar values of the same sign, such as neural firing rates or magnitudes of membrane potentials [14] [12]. Additionally, compared to hidden units with a sigmoid transfer function, $ReLU$s have been reported to improve the learning of feature detectors in generative models [24].

3 Results

The training process converges in vastly different distributions of RF mosaics. Due to the simplicity of our model, the resulting RF mosaics are varying

Table 3 Model parameter values. From left to right: transfer function, number of visible units, hidden units, hidden units alive after training, their percentage, the *undercompleteness* of the model after training, values of k, η and p. The usage of a ReLU transfer function results in the death of some hidden units during training, rendering (B) to be stronger undercomplete compared to (A), which has no dead hidden units.

	tf	vis	hid	hid alive	alive%	$\frac{\text{vis}}{\text{hid alive}}$	k	p	η
(A)	linear	768	484	484	100	1.58	10.0	1.5	$7e-04$
(B)	ReLU	768	256	149	58	5.15	4.0	1.5	$7e-04$

greatly and are dependent on numerous factors. To name the most prominent: completeness of the model, image statistics of the training set, strength of the weight constraint in relation to the learning rate and the usage of a linear or ReLU transfer function.

However, successfully converged training results show that all hidden units have developed localized, center-surround RFs (see Figure 3). These have distinguished properties, primarily defined by their color and size, which appear to fall into distinct classes. We determined these classes by fitting each RF with a DOG function and then clustering the obtained DOG parameters (for details, see A). In the following (see Table 3) we show two exemplary results which capture the difference of using a strictly linear vs. a rectified-linear transfer function, in the following denoted as *linear* (A) and *ReLU* (B).

3.1 Parametric fitting

We show a small and arbitrarily chosen set of RFs extracted from the learned connection map together with the corresponding best fit reconstructions of our parametric model, for each result (see Figure 3, A.1 and A.2).

3.2 Clustering

For classifying the learned RF types, we use the *k-means* clustering algorithm [19]. *K-means* likely converges at local minima. Therefore, we solely cluster upon the three color values extracted from the RFs center weight. Furthermore, clustering using the surrounding color would increase the difficulty of obtaining meaningful results, since we observed that in general, the RF surround appears to be more noisy and unstable compared to the RF center. Moreover, the center-weight can be trivially extracted by knowing the *center* ellipses' spatial position of the RFs best fit parameters. The center-weight is then transformed into RGB color space and fed as observation into a clustering algorithm.

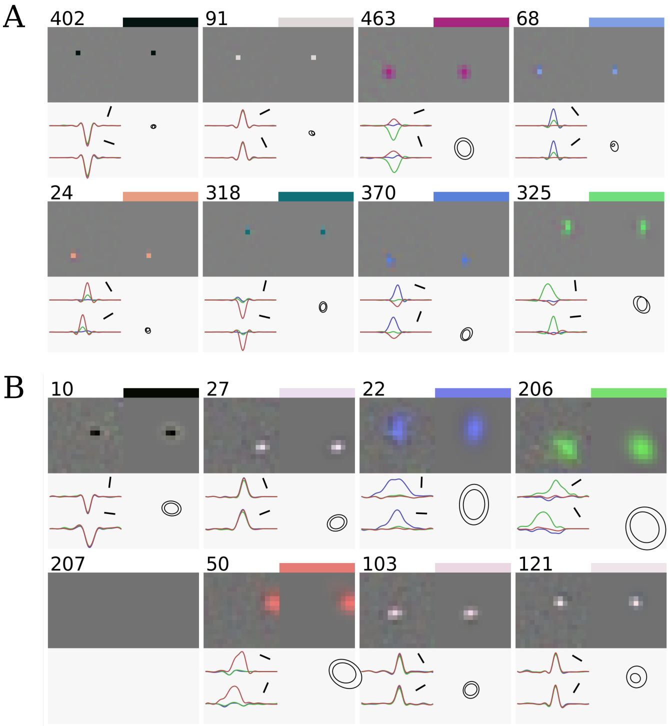

Fig. 3 Parametric fitting of eight selected RF for the (A) linear and (B) ReLU result. Each of the eight tiles represents a single RF and likewise consists of four smaller tiles which display: the RF index number and color of the center weight (above), the RF (up left), the best fit reconstruction (up right), its ellipse plot (lower right) and interpolated cuts of the RF along the primary and secondary axis of the best fit ellipse (lower left). Notably, hidden unit number 207 of (B) is dead and is excluded from further computations.

3.3 Initial number of clusters

The *k-means* algorithm requires the number of classes to be specified beforehand. To avoid guessing the ideal number of classes and to ground the decision on a measurable and reproducible foundation, we utilized two established measures of cluster quality, Davies Bouldin [6] and Ray Turi [29]. We run our cluster index algorithm for 10 iterations to avoid results stuck at local minima. In each iteration, observations are clustered with initial cluster numbers ranging from 2 to 16. Each result is then measured with the two algorithms, while keeping the best scores in between iterations (see Figure 4).

3.4 Receptive field mosaics

Plotting the best-fit center and surround ellipses grouped by cluster ownership, reveals that RF mosaics with minimal overlap have developed (see Figure 5).

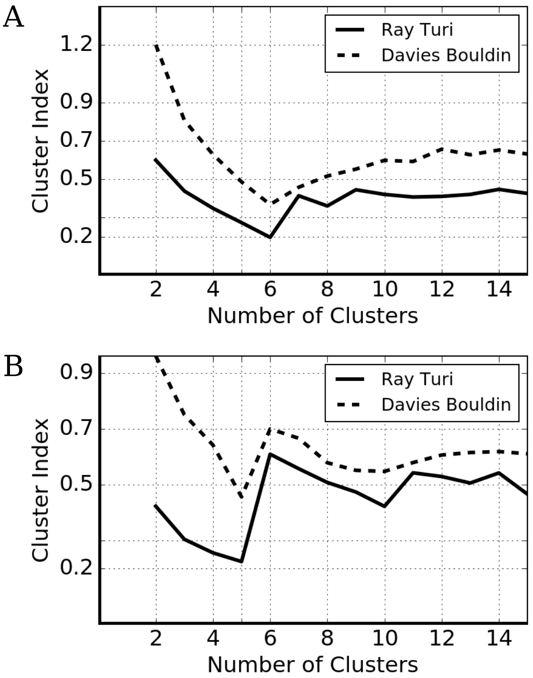

Fig. 4 Global minimum of 6 for the linear (A) and 5 clusters for the ReLU (B) result, indicating the number of RGC channels.

The RF mosaics of the linear (A) and ReLU (B) results differ in various ways: Generally larger RF sizes of (B) are explained by the lower number of hidden units compared to (A). Both results converge in a different number of complete mosaics. (B) converged in three more or less complete mosaics of red, green and blue and in two irregular and incomplete or largely overlapping mosaics. In contrast, the distribution of all six mosaics of (A) appear to be more similar compared to each other.

RFs of both results show the tendency of larger sizes in chromatic mosaics (red, green, magenta and blue) compared to the smaller sizes in intensity mosaics (black and white). Moreover, (A) developed separate Black and White mosaics, whereas in (B) these appear to be weakly developed (black) or noisy (white). Additionally, a magenta mosaic emerged in (A), but is missing in (B). Notably, (A) appears to have developed pairs of mosaics each filling the gaps of the other, such as green and magenta, black and white, whereas in (B) no such pair exist.

To validate the fitting and clustering results, all RFs of a learned map are plotted grouped by cluster membership together with the particular center color value displayed as a small rectangle above (see Figure 6).

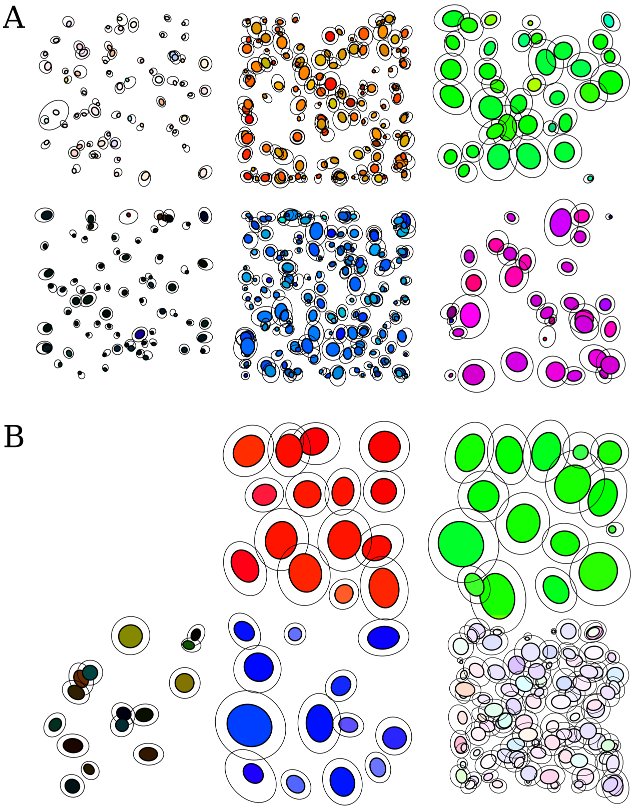

Fig. 5 The resulting RF mosaics as obtained from the fitting and clustering process; Plot for the (A) linear result with 6 channels and (B) ReLU result with 5 channels. The RF of a hidden unit is drawn with the parameters obtained from the best fit of the spatio-chromatic DOG model (see A.1). The inner ellipse displays the center Gaussian and the outer ellipse shows the antagonistic surround Gaussian of the DOG model. The color in which an ellipse is drawn reflects the value of the center weight of a particular receptive field.

4 Discussion

The original work of Vincent and Baddely showed that by enforcing input reconstruction with synaptic efficient weights, spatially localized RFs emerge which cover the entire visual field [33]. The results demonstrate that our extended model replicates this property of emerging localized RFs also in the color domain.

Due to the property of our model being *undercomplete* a compressed representation of the input data with minimal reconstruction error is learned. This property is of high benefit in applications using high dimensional input, such as processing RGB images with convolutional neural networks (CNN), since today's cameras have resolutions of many megapixels, which renders the application of deep NN intractable in directly processing each input pixel.

Many deep NN use heavily down-sampled visual input, e.g. [23] [16]. Correspondingly, the retina compresses the information of approximately 6.4 million cone photoreceptors plus an even larger number of rod photoreceptors, to one million RGC [14], indicating a high rate of compression of visual information.

4.1 Variability of model results

We observed slightly differing results between training sessions, with larger variances occurring if the models hyper parameters are altered. For example, the size of localized RFs is dependent on the ratio of visible to (alive) hidden units, or model under-completeness. The fewer the number of hidden units, the larger the size of individual RFs. To give an additional example, the strength of the weight constraint k influences the localization of the emerging RFs. Too small values prohibit the localization of RFs, likewise too large values, which results in dead hidden units of large quantities; whereat the few survivors are no longer able to reconstruct the input with localized RFs.

Despite the variability in the emerged RF mosaics and types, the model appears to be fairly robust to changes of the dataset. The linear result was trained with an entirely different dataset than the ReLU result, nonetheless both results show to some degree of similarity in the distribution of RF mosaics.

4.2 Formation of RFs mosaic and RGC types

The self-organization of localized RF mosaics that emerge during our models training process is unanimously observed in biology [38] [9]. Our model simulates synaptic activity patterns, which are attributed to drive the formation of ON and OFF ganglion cell types during retinal development [5]. Also, local synaptic activity patterns between neighbor cells of similar type are found to cause the formation of RF mosaics [30]. This locality also occurs in our model: During training, similar localized RFs compete in reconstructing a similar (spatially) local aspect of the input. Therefore, competing hidden units indirectly interact in a local manner. However, our model cannot reliably produce regular and complete RF mosaics for all developed RGC types (see Figure 1 and 5). Nevertheless, the results support the idea that there is a significant influence of learning under constrained metabolic conditions in the formation of RGC types.

4.2.1 LMS vs RGB Colorspace

Due to the simplicity of our model, the simulation of actual LMS cone activations appears impractical. For example, the shape of RGB image patches does not reflect that blue cones only amount to 15 % of all cones [20], or that an equally spaced, grid-like spatial distribution of pixels obviously diverges from the concentric organization of cells around the fovea [37]. Therefore, a simple transformation of our RGB training data into LMS color space, retaining the unrealistic shape of RGB patches, would not suffice. For the sake of simplicity our model learns from RGB images directly.

4.3 Competition between hidden units

Without a threshold mechanism, the weight-constraint results in a competition among the hidden units. Weakly participating units eventually die off, which appears to be a useful property: The competition results in a compact neural code, in which our model automatically adjusts the number of hidden units towards an *undercomplete* ratio of visible to hidden units. The competition of hidden units resembles the competition of dendrites and ganglion cell death during the development of the retina [18] [17].

5 Conclusion

We showed in a minimal setup of just one hidden layer that learning under constrained synaptic activity suffices to produce mosaics of spatially and chromatically localized color opponent RFs, resembling biological RGC receptive fields. The localized DOG shaped RFs encode visual information efficiently by producing a color opponent contrasted representation. Contrast information is essential for the task of border detection [39], which is the foundation of many basic tasks in computer vision, such as image segmentation and object recognition. Consequently, tasks which learn features from RGB images could benefit by preprocessing the visual information in a biologically inspired manner.

Furthermore, color opponent feature detectors have been manually constructed or derived statistically from RGB images [39] [4]. Our results show that optimal filters can be learned directly from natural images. However, the large variance in the distribution of RF mosaics and corresponding types does not allow an exact identification of RGC types. Still, the results allow to attribute an influence of learning in the formation of RGC mosaics and corresponding types.

5.1 Future work

Different options exist for extending and exploring our model. A thorough comparison with the denoising auto-encoder model would be of interest, since being trained on MNIST data and not on images of natural scenes, center surround RFs among other forms emerge [35].

It would be possible to extend our 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. Such a model would be trained with videos or short image sequences instead of static 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 a small number but with larger RFs compared to units tuned to chromatic aspects. Instead of training our model on RGB image patches, our model can be extended to be trained with higher dimensional input, such as RGB-D data, which contains additional depth information or fMRI data.

Another path in extending our model is to include scotopic, low light and night, vision. It is completely unknown how such an extended auto-encoder would learn these very divergent image statistics. The 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 RFs sizes of achromatic RGC could also be hypothesized to be the result of being required to sample sufficient light of a large number of rods.

Finally, our model could be possibly of use in simulating some aspects of color blindness since our model can be trained with an arbitrary number of input channels. For example, by only using two input channels dichromatic vision can be simulated. This might be of relevance since training our model with less than three input channels results, compared to training with RGB images, in the emergence of a smaller number of RGC channels.

A Appendix

A.1 Spatio-chromatic model

The excitatory center and inhibitory surround organization of retinal ganglion cell RFs is commonly modeled with a Difference of Gaussians (DOG) function [31]. Our parametric elliptical DOG model has a set of 17 parameters: The elliptical Gauss function G has center μ_x μ_y, spread σ_x σ_y and rotation θ.

$$\mathrm{G}(x, y) = e^{-a(x-\mu_x)^2 \,+\, 2b(x-\mu_x)(y-\mu_y) \,+\, c(y-\mu_y)^2} \tag{11}$$

$$a = \quad \frac{1}{2\,\sigma_x^2}\cos^2(\theta) + \frac{1}{2\,\sigma_y^2}\sin^2(\theta) \tag{12}$$

$$b = -\frac{1}{4\,\sigma_x^2}\sin(2\,\theta) + \frac{1}{4\,\sigma_y^2}\sin(2\,\theta) \tag{13}$$

$$c = \quad \frac{1}{2\,\sigma_x^2}\sin^2(\theta) + \frac{1}{2\,\sigma_y^2}\cos^2(\theta) \tag{14}$$

Each center and surround Gauss function has its own μ, σ and θ parameters (denoted by prefixes c and s).

$$\mathrm{DOG}_{cntr} = \mathrm{G}(x, y) \text{ with } c\mu_x, c\mu_y, c\sigma_x, c\sigma_y, c\theta$$

$$\mathrm{DOG}_{srrnd} = k_s\,\mathrm{G}(x, y) \text{ with } s\mu_x, s\mu_y, s\sigma_x, s\sigma_y, s\theta$$

The chromatic part of the model with additional center and surround direction parameters *cd* and *sd* in 3D color space:

$$red = cd_r\,DOG_{cntr}(x, y) - sd_r\,DOG_{srrnd}(x, y) \tag{15}$$

$$green = cd_g\, DOG_{cntr}(x, y) - sd_g\, DOG_{srrnd}(x, y) \tag{16}$$

$$blue = cd_b\, DOG_{cntr}(x, y) - sd_b\, DOG_{srrnd}(x, y) \tag{17}$$

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

A.2 Parametric fitting

A single run of the fitting algorithm emits a set of parameters that reconstruct the given RF 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 RFs to the parametric model.

References

1. Baldi, P.F., Homik, K.: Learning in linear neural networks: A survey. IEEE Transactions on Neural Networks **6**(4), 837–858 (1995)
2. Bell, A.J., Sejnowski, T.J.: The 'independent components' of natural scenes are edge filters. Vision Research **37**, 3327–3338 (1997)
3. Brainard, D.H., Brunt, W.A., Speigle, J.M.: Color constancy in the nearly natural image. Optical Society of America **14**(9), 307–325 (1997)
4. Brown, M., Süsstrunk, S., Fua, P.: Spatio-chromatic decorrelation by shift-invariant filtering. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 27–34 (2011)
5. Chalupa, L.M., Günhan, E.: Development of on and off retinal pathways and retinogeniculate projections. Progress in Retinal and Eye Research **23**(1), 31–51 (2004)
6. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. **1**(2), 224227 (1979)
7. Demonasterio, F.M., Gouras, P.: Functional properties of ganglion cells of the rhesus monkey retina. The Journal of Physiology pp. 167–196 (1975)
8. 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., Simoncelli, E.: Efficient coding of spatial information in the primate retina. The Journal of Neuroscience **32**(46), 16,256–16,264 (2012)
9. Elliott, T., Shadbolt, N.R.: A neurotrophic model of the development of the retinogeniculocortical pathway induced by spontaneous retinal waves. The Journal of Neuroscience **19**(18), 7951–7970 (1999)
10. Field, G.D., Chichilnisky, E.J.: Information processing in the primate retina: Circuitry and coding. Annual Review of Neuroscience **30**, 1–30 (2007)
11. Field, G.D., Gauthier, J.L., Sher, A., Greschner, M., Machado, T., Jepson, L.H., Shlens, J., Gunning, D.E., Mathieson, K., Dabrowski, W., Paninski, L., Litke, A.M., Chichilnisky, E.: Functional connectivity in the retina at the resolution of photoreceptors. Nature **467**(10), 673–677 (2010)
12. Gegenfurtner, K.R.: Cortical mechanisms of color vision. Neuroscience Nature Reviews **4**, 563–572 (2003)
13. Gouras, P.: Color Vision. Webvision: The Organization of the Retina and Visual System [Internet], http://www.ncbi.nlm.nih.gov/books/NBK11537/ (2009)
14. Kolb, H.: How the retina works. American Scientist **91**, 28–35 (2004)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25 (Book) pp. 1097–1105 (2012)
16. Levine, S., Finn, C., Darrell, T., Abbeel, P.: End-to-end training of deep visuomotor policies. CoRR **abs/1504.00702** (2015)

Table 4 Retinal distribution of 460 RGC units, of which 457 could be classified, from the rhesus monkey retina. Data are from [7]. For readability, we simplified the notation of the cell classes and give the cell counts and the percentages as sums over all eccentricities.

cell class	count	%	assumed RGC Type
1. Color-opponent concentric, (61%)			
$M+/L-$	95	20.79	$Midget$
$M+/(S+L)-$	4	0.88	
$M-/L+$	30	6.56	$Midget$
$M-/(S+L)+$	3	0.66	
$L+/M-$	76	16.63	$Midget$
$L+/(S+M)-$	3	0.66	
$L-/M+$	27	5.91	$Midget$
$L-/(S+M)+$	2	0.44	
$(M+L)+/S-$	6	1.31	$Midget$
$(M+L)-/S+$	10	2.19	$Midget$
$S+/(M+L)-$	17	3.72	$Midget$
$S-/(M+L)+$	4	0.88	$Midget$
2. Color-opponent, non-concentric (2%)			
$S+/(M+L)-$	5	1.09	
$S-/(M+L)+$	1	0.22	
$L+/M-$	3	0.66	
3. Broad-band, non-opponent (24%)			
ON/OFF	69	15.1	$Parasol$
OFF/ON	41	8.97	$Parasol$
4. Broad-band, colour-opponent (4%)			
$(M+L)+/L-$	11	2.41	$Small-Bistratifie$
$(M+L)+/M-$	4	0.88	$Small-Bistratifie$
$(S+M+L)+/M-$	2	0.44	
$(M+L)-/L+$	2	0.44	$Small-Bistratifie$
$(M+L)-/M+$	1	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	

17. Linden, R.: Dendritic competition in the developing retina: ganglion cell density gradients and laterally displaced dendrites. Neuroscience **10**(2), 313–337 (1993)
18. Linden, R., Perry, V.: Ganglion cell death within the developing retina: a regulatory role for retinal dendrites? Neuroscience **7**(11), 2813–2840 (1982)
19. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, pp. 281–297. University of California Press, Berkeley, Calif. (1967). URL http://projecteuclid.org/euclid.bsmsp/1200512992
20. Masland, R.H.: The fundamental plan of the retina. Nature Neuroscience **4**(9), 877 – 886 (2001)
21. Maul, T.H., Bargiela, A., Ren, L.J.: Cybernetics of vision systems: Toward an understanding of putative functions of the outer retina. In: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 41, pp. 398 – 409 (2011)

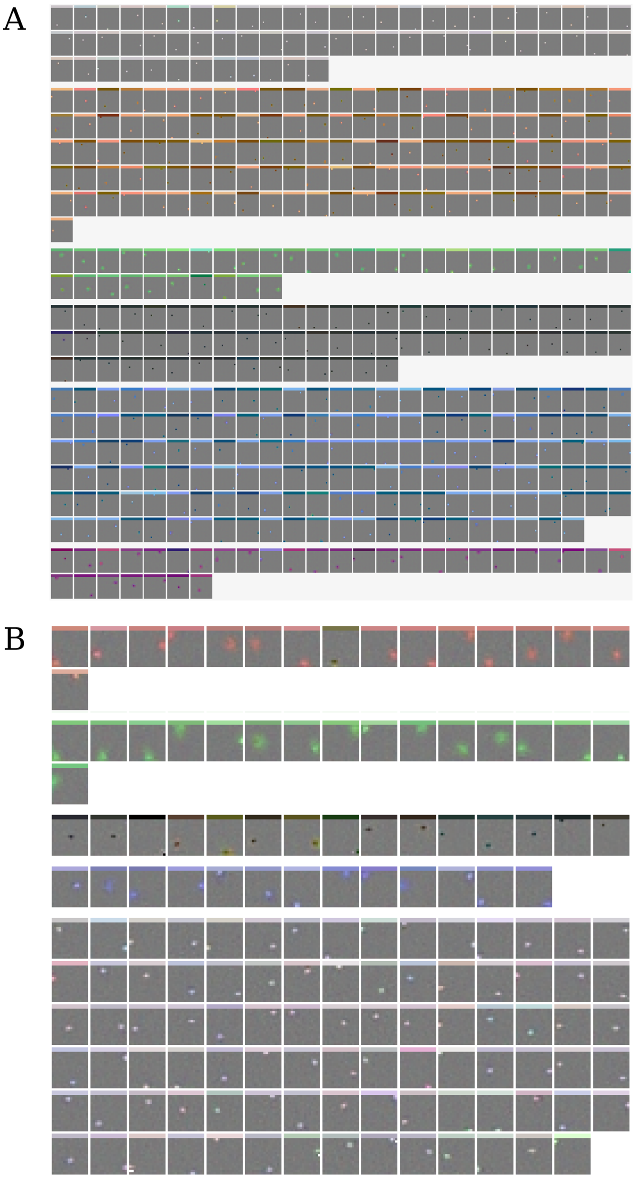

Fig. 6 All RF from the (A) linear and (B) ReLU result grouped by cluster membership. Above each RF the estimated color value of the best fit center ellipse is shown.

22. McIntosh, L., Maheswaranathan, N., Nayebi, A., Ganguli, S., Baccus, S.: Deep learning models of the retinal response to natural scenes. In: D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, R. Garnett (eds.) Advances in Neural Information Processing Systems 29, pp. 1361–1369 (2016)

23. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M.: Playing Atari with deep reinforcement learning p. arXiv:1312.5602 (2013)

24. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th International Conference on Machine Learning pp. 807–814 (2010)
25. Nelson, R.: 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/ (2007)
26. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set - a strategy employed by V1? Vision Research **37**(23), 3311–3325 (1997)
27. von Poschinger, D., Weber, C., Wermter, S.: A generative model of decorrelating retinal ganglion cells. 11th Göttingen Meeting of the German Neuroscience Society (2015)
28. Purves, D., Lotto, R.B., Nundy, S.: Why we see what we do. American Scientist **90**(3), 236–242 (2002)
29. Ray, S., Turi, R.H.: Determination of number of clusters in k-means clustering and application in colour segmentation. In: The 4th International Conference on Advances in Pattern Recognition and Digital Techniques, pp. 137–143 (1999)
30. Reese, B.E.: Retinal mosaics: pattern formation driven by local interactions between homotypic neighbors. Frontiers in Neural Circuits **6**(24), doi: 10.3389/fncir.2012.00,024 (2012)
31. Shapley, R., Hawken, M.J.: Color in the cortex: single- and double-opponent cells. Vision Research **51**, 701–717 (2011)
32. Turcsany, D., Bargiela, A., Maul, T.: Modelling retinal feature detection with deep belief networks in a simulated environment. In: F. Squazzoni, F. Baronio, C. Archetti, M. Castellani (eds.) ECMS 2014 Proceedings, European Council for Modeling and Simulation, pp. doi:10.7148/2014–0364 (2014)
33. Vincent, B.T., Baddeley, R.J.: Synaptic energy efficiency in retinal processing. Vision Research **43**, 1283–1290 (2003)
34. Vincent, B.T., Baddeley, R.J., Troscianko, T., Gilchrist, I.D.: Is the early visual system optimised to be energy efficient? Network: Computation in Neural Systems **16**(2), 175–190 (2005)
35. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 2008 pp. 1096–1103 (2008)
36. Weber, C., Triesch, J.: A sparse generative model of V1 simple cells with intrinsic plasticity. Neural Computation **20**, 1261–1284 (2008)
37. Weber, C., Triesch, J.: Implementations and implications of foveated vision. Recent Patents on Computer Science **2009**(2), 75–85 (2009)
38. Wong, R.O.: Retinal waves and visual system development. Annual review of neuroscience **22**(1), 29–47 (1999)
39. Yang, K., Gao, S., Li, C., Li, Y.: Efficient color boundary detection with color-opponent mechanisms. 2013 IEEE Conference on Computer Vision and Pattern Recognition pp. 2810–2817 (2013)