

# Mouse retinal specializations reflect knowledge of natural environment statistics

Yongrong Qiu<sup>1,2,3</sup>, Zhijian Zhao<sup>1,2</sup>, David Klindt<sup>1,2,3</sup>, Magdalena Kautzky<sup>4,5</sup>, Klaudia P. Szatko<sup>1,2,3,6</sup>, Frank Schaeffel<sup>1</sup>, Katharina Rifai<sup>1,7</sup>, Katrin Franke<sup>1,2,6</sup>, Laura Busse<sup>4,8,#</sup>, Thomas Euler<sup>1,2,6,#,§</sup>

<sup>1</sup> Institute for Ophthalmic Research, U Tübingen, Tübingen, Germany

<sup>2</sup> Centre for Integrative Neuroscience (CIN), U Tübingen, Tübingen, Germany

<sup>3</sup> Graduate Training Centre of Neuroscience (GTC), International Max Planck Research School, U Tübingen, Tübingen, Germany

<sup>4</sup> Division of Neurobiology, Department Biology II, LMU Munich, Munich, Germany

<sup>5</sup> Graduate School of Systemic Neuroscience (GSN), LMU Munich, Munich, Germany

<sup>6</sup> Bernstein Centre for Computational Neuroscience, Tübingen, Germany

<sup>7</sup> Carl Zeiss Vision International GmbH, Aalen, Germany

<sup>8</sup> Bernstein Centre for Computational Neuroscience, Munich, Germany

# Corresponding authors:

[busse@biologie.uni-muenchen.de](mailto:busse@biologie.uni-muenchen.de), [thomas.euler@cin.uni-tuebingen.de](mailto:thomas.euler@cin.uni-tuebingen.de)

§ Lead contact: Thomas Euler, [thomas.euler@cin.uni-tuebingen.de](mailto:thomas.euler@cin.uni-tuebingen.de)

## Summary

Pressures for survival drive sensory circuit adaption to a species' habitat, making it essential to statistically characterise natural scenes. Mice, a prominent visual system model, are dichromatic with enhanced sensitivity to green and UV. Their visual environment, however, is rarely considered. Here, we built a UV-green camera to record footage from mouse habitats. We found chromatic contrast to greatly diverge in the upper but not the lower visual field, an environmental difference that may underlie the species' superior colour discrimination in the upper visual field. Moreover, training an autoencoder on upper but not lower visual field scenes was sufficient for the emergence of colour-opponent filters. Furthermore, the upper visual field was biased towards dark UV contrasts, paralleled by more light-offset-sensitive cells in the ventral retina. Finally, footage recorded at twilight suggests that UV promotes aerial predator detection. Our findings support that natural scene statistics shaped early visual processing in evolution.

## Keywords

Mouse vision, colour vision, visual ecology, natural movies, natural scene statistics, retina, ultraviolet light, convolutional autoencoder, efficient encoding, ON/OFF pathways

## Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Thomas Euler ([thomas.euler@cin.uni-tuebingen.de](mailto:thomas.euler@cin.uni-tuebingen.de))

## Introduction

During evolution, the structure and function of neural circuits have been shaped to improve the species' chances to survive and procreate in their specific natural environments. Such adaptations have long been described and discussed in the visual system (Attneave, 1954; Barlow, 1961; Simoncelli and Olshausen, 2001). Some of these adaptations are already in eye placement: For example, predators, such as cats, usually have frontally placed eyes, maximizing the binocular field to improve stereovision. Prey animals, such as mice, have laterally placed eyes, expanding the field-of-view (FOV) to detect threats as reliably as possible (reviewed in Baden et al., 2020).

Given such specific adaptations, characterising the properties of natural visual environments is crucial for advancing our understanding of the structure and function of the visual system (Masland and Martin, 2007), in particular for species whose visual systems are different from ours, such as mice. Yet, studying the visual system and behaviour of mice in the context of their natural environment is only starting (Datta et al., 2019; Hasson et al., 2020; reviewed in Krakauer et al., 2017), despite the fact that mice have become a prominent model system for vision research in the past decade (reviewed in Huberman and Niell, 2011; Seabrook et al., 2017). The importance of considering ethologically relevant stimuli and behaviours for studying mouse vision is highlighted in recent work, showing, for example, superior spatial frequency tuning for V1 neurons when probed with ecologically inspired visual stimuli instead of drifting gratings (Dyballa et al., 2018), which might potentially underlie accurate visually-driven approach performance during prey capture (Hoy et al., 2016).

The relationship between properties of natural scenes and principles of visual coding has been powerfully addressed using computational modeling (reviewed in Turner et al., 2019). A classic example is the finding that simple cell-like receptive fields (RFs) emerge in models that learn a sparse code for natural images (Olshausen and Field, 1996). Along these lines, Ocko and coworkers (2018) recently showed that a convolutional autoencoder (CAE; Ballard, 1987; Hinton and Salakhutdinov, 2006) trained to reconstruct pink ( $1/f$ ) noise learned center-surround spatial filters reminiscent of the RFs of different retinal ganglion cell (RGC) types. In addition, imposing constraints on encoding models to drive optimal performance in specific natural vision tasks, has proven fruitful to generate principled hypotheses about the underlying neural or perceptual mechanisms (Burge and Jaini, 2017; Geisler et al., 2009). Such a task-driven framework has been fueled in the last years by advances in deep convolutional neural networks (DNNs) based on discriminative learning from large databases of natural images (reviewed in LeCun et al.,

2015; Turner et al., 2019). Interestingly, such layered networks learn features that share representational similarities with neuronal activity in several processing stages along the visual hierarchy (Yamins and DiCarlo, 2016).

In natural scenes, colour is an ethologically highly relevant visual feature, and depending on which colours an animal can see, this will influence its ability to forage and hunt, select mates, and avoid predators. Indeed, beyond the absence of a fovea in mice, a major difference compared to primates is that mice are dichromats and see UV light. Next to a medium (M) wavelength-sensitive opsin peaking at 510 nm (green), mice have a short (S) wavelength-sensitive opsin peaking at 360 nm in the UV (Jacobs et al., 2004; Röhlich et al., 1994). Due to co-expression of the S-opsin in ventral M-cones (Baden et al., 2013; Szél et al., 1992) and a peak in S-cone density in the ventral periphery (Nadal-Nicolás et al., 2020), the mouse retina is subdivided into a more green-sensitive dorsal and a strongly UV-sensitive ventral half. Notably, ventral cones are tuned to signal dark contrasts (Baden et al., 2013) and, hence, may support the detection of dark shapes against the sky. Such dorso-ventral regionalization of the mouse retina suggests a functional specialization, according to which the ventral retina “monitors” the overhead space to detect predators (Wallace et al., 2013; Zhang et al., 2012), whereas the dorsal retina supports foraging and hunting for food (Hoy et al., 2019; Shang et al., 2019). An important consequence of this regionalization is that any statistical analysis of the mouse’s visual environment — like quantifying spectral content of natural scenes — should consider the upper and lower visual field separately. However, recording the chromatic information available to mice under natural conditions is challenging, since standard cameras fail to capture UV light.

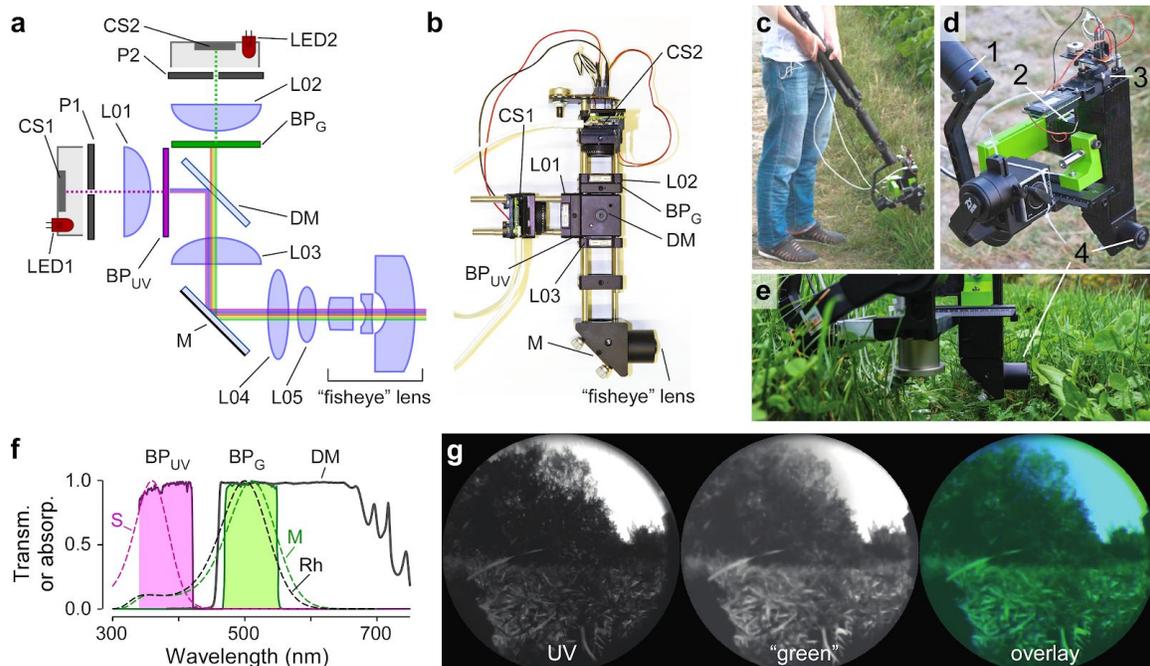
To analyze the spectral properties of the environment of mice, we have therefore built a hand-held camera that is mounted on a gimbal for image stabilization and covers the spectral bands – UV and green – relevant for mice. We used this camera to record footage near mouse tracks in the field, capturing various outside scenes at different times of the day – providing a resource of natural scenes for vision research in mice. Focusing on spectral information, we found that the contrast in the two spectral channels greatly diverged in the upper but not in the lower visual field, paralleling the superior chromatic opponency in the ventral retina (Szatko et al., 2020) and behavioural colour discrimination in the upper visual field (Denman et al., 2018). Notably, our analysis of the footage suggests that the mouse’s UV sensitivity may help detecting aerial predators also at dusk and dawn. Finally, computational modelling predicts that colour-opponent filters are more likely to emerge in unsupervised models trained with images

from the upper visual field than from the lower visual field. Together, this lends further support to the idea that retinal circuits have evolved to process natural scene statistics in a species-specific manner (reviewed in Baden et al., 2020).

## Results

### A camera for recording visual scenes from the mouse's perspective

The goal of this study was to capture the visual environment of mice mimicking key aspects of mouse vision. Specifically, we focused on three main aspects: (i) the perspective from only a few centimeters above the ground, (ii) the large FOV that for each eye approaches  $\sim 180^\circ$  (Sterratt et al., 2013), and (iii) the spectral sensitivities of the mouse photoreceptors, peaking at 360 and 510 nm for UV-cones and rods/M-cones, respectively.



**Figure 1 | Mouse-camera module.** **a**, Schematic drawing of mouse-camera with two spectral channels (UV, green). CS, camera sensor; P, pinhole; L01-05, lenses;  $BP_{UV}$ , UV bandpass filter (350-419 nm); M, silver mirror; LED, light-emitting diode;  $BP_G$ , green bandpass filter (470-550 nm); DM, dichroic mirror (>90% reflection: 350-442 nm; >90% transmission: 460-650 nm). **b**, Picture of the assembled camera module. **c-e**, Pictures of assembled camera module, with gimbal (1), UV camera (2), green camera (3), and fisheye lens (4). **f**, Normalized transmission spectra of DM,  $BP_{UV}$  and  $BP_G$ , with normalized absorption spectra of mouse cone opsins (S, M) and rhodopsin (Rh) overlaid (Franke et al., 2019). **g**, Movie frame with UV (left), green (centre) channel, and overlay (right).

To this end, we developed a “mouse-camera” that simultaneously captures movies in the UV and green spectral bands (Fig. 1; for components, see Table 1; for camera settings, see Table

2; for additional details, see Methods). The two spectral channels were simultaneously recorded on flash memory by two single-board Raspberry Pi microcomputers attached to the camera modules. As an objective, we used a fisheye lens with a FOV of 180° (Fig. 1a-e). The camera was mounted on an active gimbal for stabilization, such that it could be moved close to the ground (Fig. 1c-e).

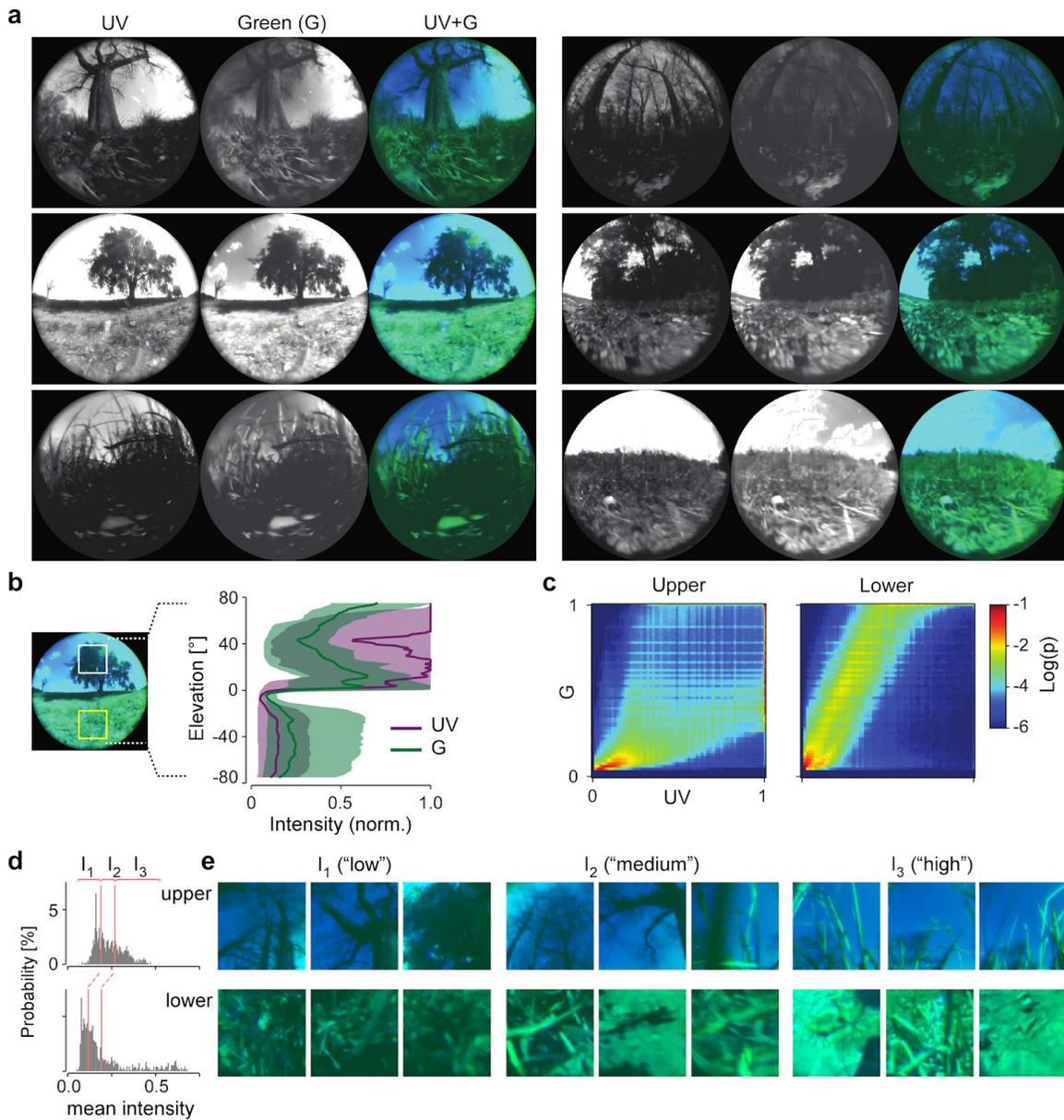
Based on the assumption that eye movements in mice serve mainly to stabilize the retinal image (Land, 2019; Meyer et al., 2018, 2020; Michael et al., 2020) and are typically coupled to head movements (Meyer et al., 2020; Michael et al., 2020), we restricted our recordings to a view with the horizon oriented parallel to ground and positioned around the middle elevation of the camera image. This way, the lower half of the camera image captured information from the lower visual field of mice, while the upper half captured information from the upper visual field.

The Raspberry Pi camera modules we used are consumer products (for specifications, see Table 3) and, hence, are optimized for taking colour pictures that look natural to a human observer. At the hardware side, this is achieved by a checkerboard-like pattern of thin RGB filters (Bayer filter) coated directly on the chip surface (for spectra, see resources listed in Table 3). While the Bayer filter's green component matches the spectral sensitivity of the mouse rods and M-cones ( $\lambda_{peak} \approx 510 \text{ nm}$ ) (Jacobs et al., 2004, 2007), the Bayer filter largely blocks UV light. To address this issue, we mechanically removed the Bayer filter layer in case of the UV channel camera chip (Wilkes et al., 2016).

The effective spectral sensitivity of each camera channel is determined by the combination of bare chip sensitivity, Bayer filter (in case of the green channel), and a spectral bandpass filter (Fig. 1f). In addition, the cameras automatically adjust the image intensities to match gamma curves typically found in consumer displays. Since we do not know the chip's exact sensitivity curve, we measured the sensitivity of the two camera channels using LEDs of defined spectrum and power (Suppl. Figure S1a; Methods). For each channel, we mapped image intensities to power meter readouts of UV or green LEDs, and fitted the data using an inverse gamma correction function (Suppl. Fig. S1b). We then used these fits to convert pixel values into normalized intensities (source vs. intensity-corrected images; see Suppl. Fig. S1c).

To verify this intensity correction, we recorded outside scenes using a scanning spectrometer (Baden et al., 2013), which provides still images with each pixel containing the spectrum from 300 to 660 nm (see Methods). By convolving each pixel's spectrum with the mouse cone opsin absorption curves, we generated mouse-view corrected versions of the spectrometer images (Suppl. Fig. S1d). For representative scene elements, such as grass and trees, we found that

the normalized intensities in the corrected camera images matched well those in the spectrometer images for both the UV and green channel (Suppl. Fig. S1e).



**Figure 2 | Example scenes and intensity distribution.** *a*, Example frames (UV, green, and overlay) from movies of different scenes recorded outside in the field (near Waldhäuser Ost, Tübingen, Germany; 48°33'02.4"N 9°03'01.2"E). *b*, Normalized intensity for green (G) and UV as a function of elevation across the  $n=1,936$  frames of one example movie (median w/ 25<sup>th</sup> and 75<sup>th</sup> percentile). *c*, 2D histograms of intensities of the same movie as in (b), visualizing the G-UV intensity distribution for an image cut-out (crop) in the upper (left) and lower (right) visual field (for region placement, see yellow boxes in (b, left)). *d*, Distribution of mean intensities for 1,500 image crops from upper and lower visual field, selected randomly from 15 movies, and division into 3 intensity classes (percentiles  $I_{1-3}$ ). *e*, Example images from the three mean intensity classes.

Note that the intensity correction reached its limits for extensive sky regions, because the camera chip could saturate due to its limited dynamic range, when the spectrometer did not. We considered this potential issue by excluding oversaturated images (see Methods). Note also that we used the available range for both colour channels by mapping the intensities to pixel values (see Methods), although in the natural environment, the absolute intensity of UV is known to be lower than that of longer wavelengths (Hut et al., 2000; see Discussion). After intensity correction, the movies from the two channels were temporally synchronized and spatially overlaid (see Methods), resulting in a single UV-green movie (Suppl. Video V1).

## Movie recordings and first-order statistics

Next, we used the camera to record footage of representative natural scenes outside in the field in places with traces of mouse activity (Fig. 2a). Most scenes were recorded in summer and spring during the day, and a few scenes at dusk or dawn (*cf.* Fig. 5; see Table 4).

We first focused on first-order statistics, i.e. brightness, and explored the normalized intensity distribution in each channel as a function of elevation (Fig. 2b). Examining crops from an example movie, we found that the relative intensities in both channels were usually higher in the upper compared to the lower visual field (Fig. 2b, right). Interestingly, the two chromatic channels were less correlated in image crops from the upper than the lower visual field (Fig. 2c). We therefore quantified the linear correlation between the two channels using principal component analysis (PCA; Buchsbaum and Gottschalk, 1983) and found a 2-5 times higher variance along the colour-opponent axis in images from the upper vs. the lower visual field (see colour opponency index  $CO_i$  in Suppl. Fig. S2). This indicates a higher variability in chromatic intensity differences, i.e. contrasts, above the horizon (see also next section).

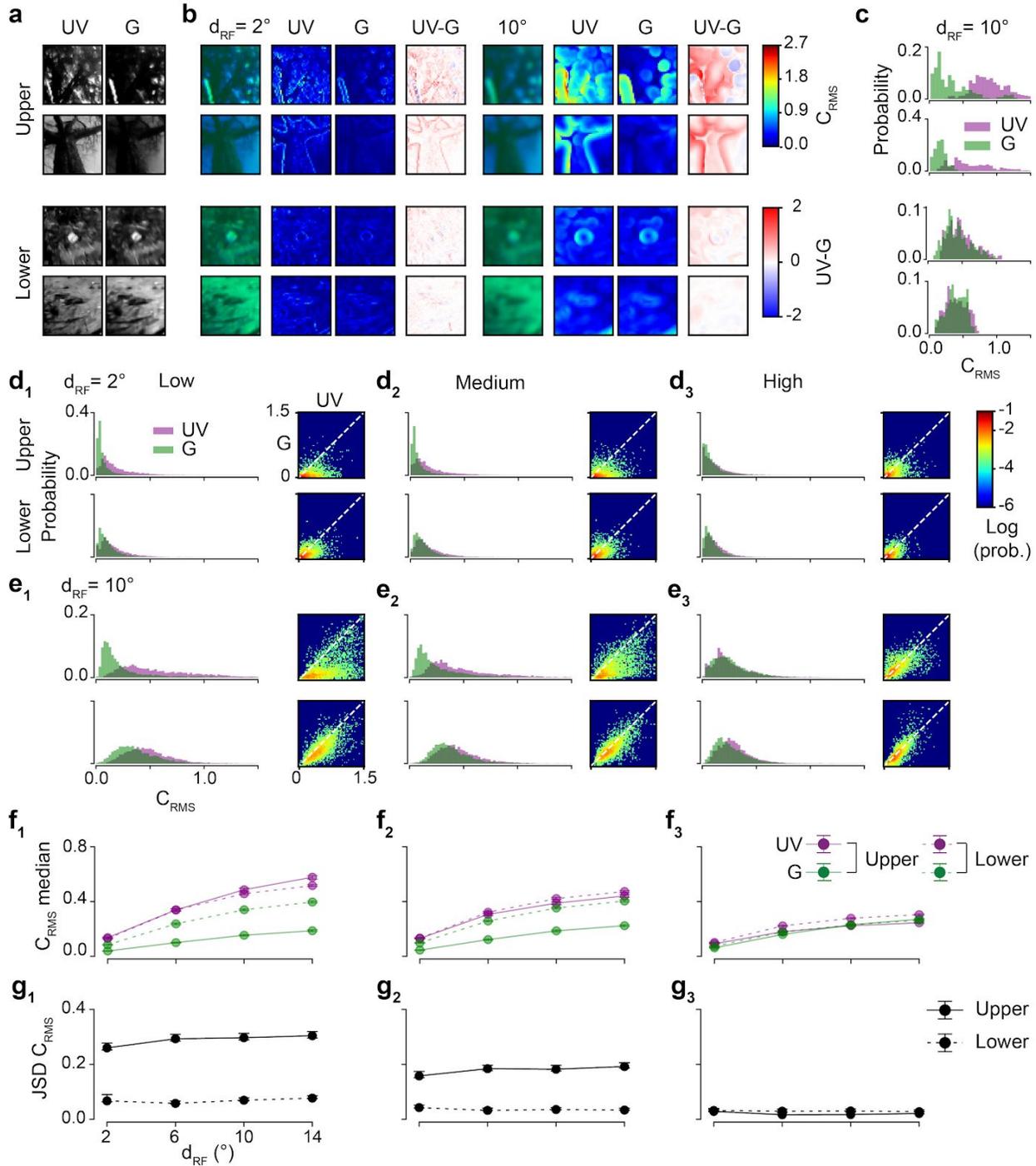
Performing these analyses systematically across all recorded movies comes with the challenge that, depending on the time of the day, the weather, and the scene content, the brightness of the recorded scenes could vary tremendously – for instance when comparing a scene with clear sky and a single tree (Fig. 2a, left-centre) with a scene recorded close to the forest (right-top). We therefore split the footage crop-wise into three groups based on the mean intensity (Fig. 2d,e; see Methods) and performed all subsequent analyses for each of these brightness categories.

## Chromatic contrast is higher in the upper vs. the lower visual field

Retinal output to the brain is also driven by second-order statistics, such as differences in brightness, i.e. contrast (Frazor and Geisler, 2006; Mante et al., 2005; Rieke and Rudd, 2009). If distributions of contrast differ between colour channels, this difference can be interpreted as chromatic contrast. Therefore, we next compared the contrast distribution in the chromatic channels in our natural scenes above and below the horizon and across the three mean intensity groups (Fig. 3).

We first focused on root mean square contrast ( $C_{RMS}$ ), commonly used in psychophysics for describing contrast in complex natural scenes (Bex and Makous, 2002; Mante et al., 2005). We extracted circular image patches (kernels) of various diameters ( $d_{RF}$ ,  $2^\circ$  to  $14^\circ$ ), a range that includes the RF sizes of mouse RGC types (Baden et al., 2016), and computed  $C_{RMS}$  as the standard deviation of the normalized pixel intensities divided by mean intensity. To quantify differences between  $C_{RMS}$  distributions, we used a two-sided permutation test as well as the Jensen-Shannon divergence ( $JSD$ ), which measures the general similarity between probability distributions, with  $JSD = 0$  indicating identical distributions.

Our analysis of the  $C_{RMS}$  distributions across 1,500 crops revealed three key features of the recorded mouse natural scenes (Fig. 3): First, as illustrated by two representative examples (Fig. 3a-c; upper two vs. lower two rows; from Suppl. Video V1,2),  $C_{RMS}$  was higher in the upper compared to the lower visual field (for  $C_{RMS}$  distributions, medians and  $JSD$ , see Fig. 3d-g). This difference was much more pronounced in the UV than the green channel and in the low and medium mean intensity groups (compare columns in Fig. 3d-g). Second, for the tested kernel sizes,  $C_{RMS}$  increased with kernel diameter (see examples in Fig. 3b; compare Fig. 3d vs. e; for median and  $JSD$ , see Fig. 3f,g), consistent with the dominance of low-spatial frequencies in natural scenes (Burton and Moorhead, 1987; Field, 1987; Ruderman and Bialek, 1994). Third,  $C_{RMS}$  distributions of the UV and green channel differed more strongly in the upper compared to the lower visual field (see example in Fig. 3c; for median and  $JSD$ , see Fig. 3f,g). These differences in  $C_{RMS}$  distribution between chromatic channels and upper vs. lower visual field were significant for all kernel diameters except in the high mean intensity group (for all comparisons, see Tables 5,6).



**Figure 3 | Chromatic contrast is higher in image crops from the upper visual field.** **a**, Examples for image crops from upper (from  $I_1$  and  $I_2$ ) and lower (from  $I_1$  and  $I_3$ ) visual field. **b**, Images from (a) filtered with different receptive field (RF) diameters,  $d_{RF}$ , with each of the columns showing (from left): UV-green overlay, UV and green channel maps visualizing RMS contrast ( $C_{RMS}$ ), and difference between maps (UV-G). **c**, Distributions of  $C_{RMS}$  for UV and green for  $d_{RF}=10^\circ$ . **d,e**, 1D (left) and 2D histograms for  $C_{RMS}$  of all  $n=1,500$  image crops from randomly picked frames (out of  $n=15$  movies) for the three intensity groups (from Fig. 2d,e) and  $d_{RF}=2^\circ$  (d) and  $d_{RF}=10^\circ$  (e). **f**, Median  $C_{RMS}$  as function of RF size for UV and green

channels in the upper (solid lines) and lower (dashed) visual field, for the three intensity groups. **g**, Like (f) but with median Jensen-Shannon divergence (JSD) between  $C_{RMS}$  distributions of the two chromatic channels as function of  $d_{RF}$ . Error bars in (f,g) represent 2.5 and 97.5 percentiles with bootstrapping (see Methods).

The differences in JSD between the upper and lower visual field likely reflect differences in chromatic contrast: When plotting green  $C_{RMS}$  as a function of UV  $C_{RMS}$ , the data points for the lower visual field tended to be distributed tightly along the identity line (e.g. Fig. 3d<sub>1</sub>,e<sub>1</sub>, bottom), indicative of a high correlation between the channels and, thus, low chromatic contrast. For the upper visual field, however, the  $C_{RMS}$  distributions for UV compared to green were broader and shifted towards higher values, in particular for larger RF kernel diameters (e.g. Fig. 3e<sub>1</sub>, top), suggesting a lower channel correlation and high chromatic contrast, respectively.

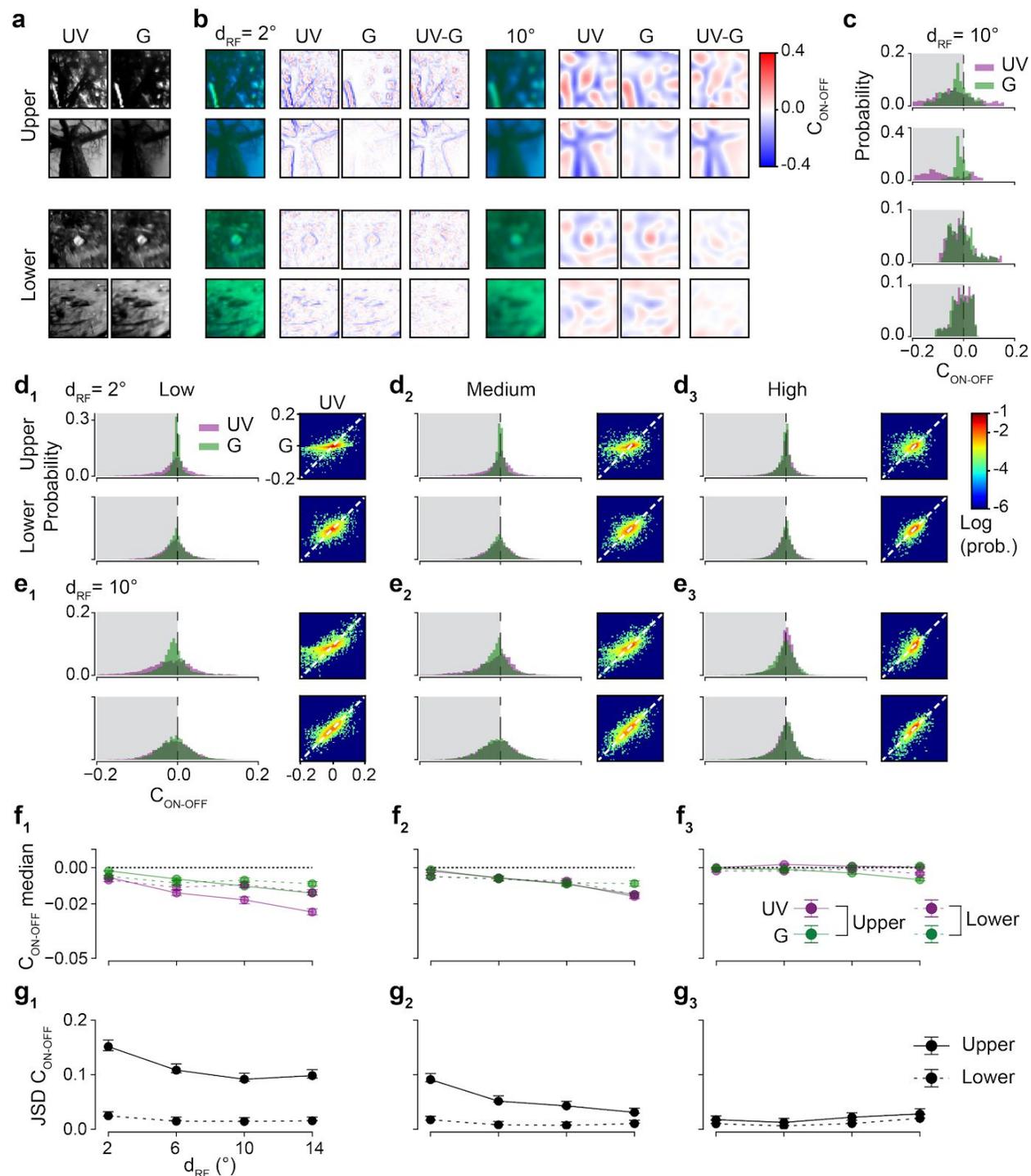
In summary, we found that except for bright scenes, UV-green chromatic contrast was higher in the upper compared to the lower visual field, particularly for the large RF kernel diameters (Fig. 3g). Taken together, this suggests that the natural environment of mice above the horizon is rich in chromatic information, which may preferentially drive (colour-opponent) RGCs with large RFs.

## Natural scenes are biased towards dark contrasts

It has been reported that the contrast distribution of (monochromatic) natural scenes is biased towards dark (negative) contrasts and that this bias is mirrored in the higher proportion of Off vs. On responding neurons in the early visual system (Ratliff et al., 2010; Wang et al., 2015; Xing et al., 2010). Therefore, we went on to study the distribution of dark and bright contrasts in the recorded scenes (Fig. 4). To measure the contrast polarity distribution (“On-Off contrast”,  $C_{On-Off}$ ) in each channel, we convolved the crops separately with the center and surround of difference of Gaussian (DOG) kernels of various diameters ( $d_{RF}$ , 2° to 14°) and computed the Michelson contrast between center and surround (Ratliff et al., 2010) (Fig. 4a-c). Like for  $C_{RMS}$ , we analysed the distributions of  $C_{On-Off}$  separately in both chromatic channels for the mean intensity groups (Fig. 4d-f).

In crops with low and medium mean intensities, we found the  $C_{On-Off}$  distributions — despite being wide — to be skewed to negative values, particularly in the upper visual field and for larger RF kernel diameters (for examples, see Fig. 4a-c; for distributions, median and JSD, see Fig. 4d-g; for statistical comparisons, see Tables 7,8). The dark bias in the upper visual field scenes resonates with results from an earlier study (Baden et al., 2013), which showed that

mouse cone photoreceptors that survey the sky preferentially encode dark contrasts, suggesting a neural representation bias that starts already at the cone level.



**Figure 4 | The upper visual field is biased towards negative contrast.** **a**, Example image crops from upper and lower visual field (same as in Fig. 3a). **b**, Images from (a) after application of On-Off filter kernel (Methods) with different RF diameters,  $d_{RF}$ , with each of the two columns showing the UV-green

overlay (left), UV (centre) and green (right) channel maps visualizing On-Off contrast ( $C_{On-Off}$ ). **c**, Distributions  $C_{On-Off}$  for UV and green for a medium-sized  $d_{RF}$  of  $10^\circ$ . **d,e**, 1D (left) and 2D histograms for  $C_{On-Off}$  of all  $n=1,500$  image crops from randomly picked frames out of  $n=15$  movies) for the three intensity groups (from Fig. 2d,e) and  $d_{RF}=2^\circ$  (d),  $d_{RF}=10^\circ$  (e). **f**, Median  $C_{On-Off}$  ( $\pm 0.1$  SD) as a function of  $d_{RF}$  for UV and green channels in the upper (solid lines) and lower (dashed) visual field; for the three intensity groups. **g**, Like (f) but with Jensen-Shannon divergence (JSD) between  $C_{On-Off}$  distributions (cf. 2D plots in (d,e)) of the two chromatic channels as a function of  $d_{RF}$ . Error bars represent 2.5 and 97.5 percentiles with bootstrapping (see Methods). Note that for larger RF kernel sizes, the distribution mode was slightly shifted to negative values (e.g. (e)), yet, when pooling images across chromatic channels, visual fields and intensity groups, the distribution mode of all four kernel sizes was at zero (data not shown), in line with earlier observations (Ratliff et al., 2010).

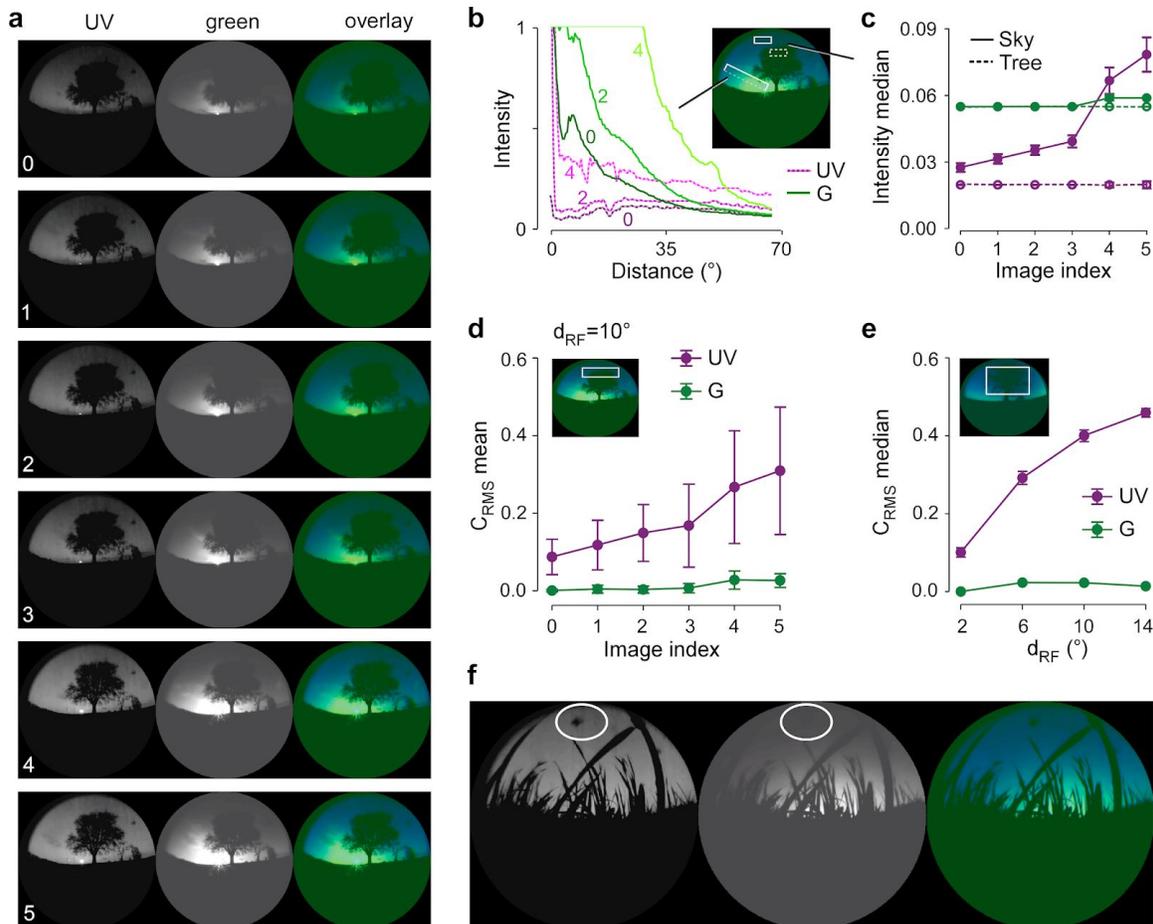
Hence, we next asked if such a neural representation bias is also present at the retinal output level, where we would expect a bias towards Off responses in ventral, large-RF RGCs. To test this prediction, we used a published dataset of RGC responses recorded in the ventral retina (Baden et al., 2016). We extracted the On-Off index (On,  $OOi > 0$ ; Off,  $OOi < 0$ ), which indicates if a cell responds more strongly to light-on or light-off transitions, and the RF diameter ( $d_{RF}$ ) for 2,380 RGCs (Suppl. Fig. S3a,b; see also Methods). We found that, indeed, Off cells with large RF ( $d_{RF} > 8^\circ$ , equivalent to  $> 240 \mu m$  in diameter) displayed a dark-biased  $OOi$  (Suppl. Fig. S3c) and were more frequent than their On cell counterparts (Suppl. Fig. S3d). This finding is consistent with the skewed distributions towards dark contrasts in our camera footage, suggesting retinal circuit adaptations that enable exploiting the dark bias present in natural scenes above the horizon.

We also compared the  $C_{On-Off}$  distributions in the two channels and found chromatic On-Off contrast (quantified as  $JSD$ ; Fig. 4g) to be systematically higher in the upper visual field, with the exception of the high mean intensity group. Moreover, like for the analysis of  $C_{RMS}$ , the  $C_{On-Off}$  distribution seemed broader for UV, supporting the idea that the mouse's UV channel provides the animal with more nuanced information about the sky region than the green channel.

## UV channel for predator detection?

The high contrast available in the upper visual field and, specifically, in the animal's UV channel may support detection of aerial predators (discussed in Cronin and Bok, 2016). However, so far, we focused on footage recorded during daytime and while mice are active during the day (Hut et al., 2011), they also forage from dusk till dawn. We thus asked if UV sensitivity was also useful for detecting objects in the sky at twilight. Interestingly, previous measurements of the spectral composition of sunlight over the course of a day revealed an overrepresentation of short

wavelengths in an approx. one-hour period during twilight (Hut et al., 2000; Johnsen et al., 2006). This effect is due to absorption by ozone, together with the scattering of sunlight in the upper atmosphere, resulting in a blueish twilight sky (discussed in Johnsen et al., 2006). Since the ratio between 360 and 520 nm – near the sensitivity peaks of mouse opsins – was reported to increase at twilight (Hut et al., 2000), we recorded with our mouse-camera before sunrise and after sunset (Fig. 5). Because of the camera chip’s relatively low light sensitivity, we recorded movies in the direction of the sun from a fixed camera position (Fig. 5a).



**Figure 5 | UV channel might facilitate detection of dark shapes in the upper visual field also at dusk and dawn.** **a**, Representative scenes recorded with the mouse-camera around sunrise. **b**, Intensity profiles along a line (dashed, see inset) starting at the sun for images from (a). **c**, Median intensities in two image crops (see inset in (a); dashed box, tree; solid box, sky) as function of image series index. **d**,  $C_{RMS}$  (mean) in image crop at the edge of a tree (rectangle in inset) as function of image index. **e**,  $C_{RMS}$  (median) in image crop placed on the tree (rectangle in inset) as function of receptive field (RF) kernel diameter ( $d_{RF}$ ). **f**, Image showing an approaching black drone mimicking a bird of prey. Panels recorded at dawn (a-d) and dusk (e,f). Error bars in (c,d) and (e) representing 2.5 and 97.5 percentiles with bootstrapping.

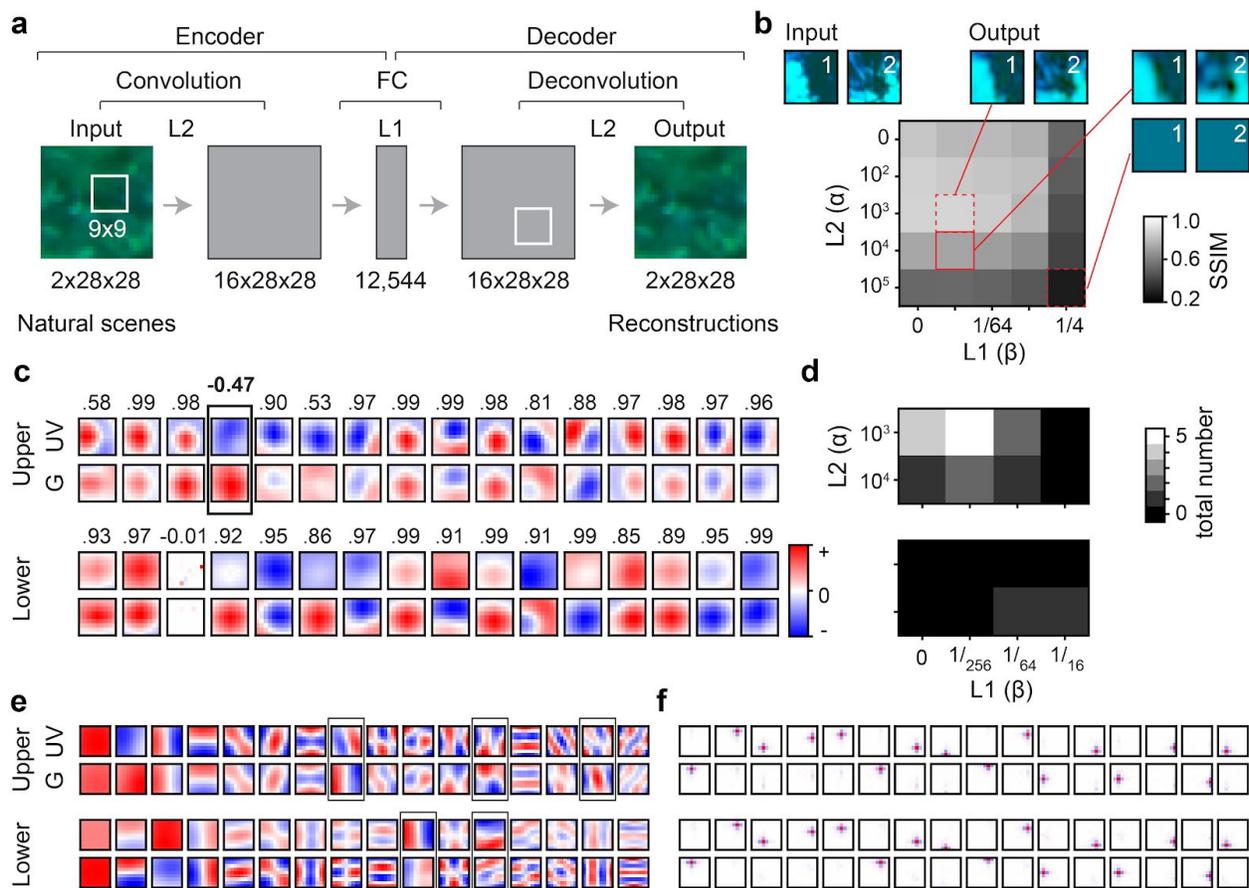
First, we explored for the example frames the intensity profiles along a  $70^\circ$  arc starting at the sun's position (Fig. 5b). In the vicinity of the sun, the intensity in the green channel was always higher than that in the UV channel, whereas further away from the sun, UV and green reached similar intensity levels (Fig. 5b; e.g. profiles from image #4 for  $> 55^\circ$ ), in line with the observations by Hut et al. (2000). Moreover, for distances larger than approx.  $5^\circ$  from the sun, the UV intensity profile was much flatter than that of green, resulting in a more homogeneous UV illumination of the sky. Finally, in the sky's dome, the intensity rose faster over the course of the sunrise in the UV compared to the green channel (Fig. 5c). Akin to daytime scenes,  $C_{RMS}$  was significantly higher in the UV than the green (Fig. 5d;  $p < 0.0001$ , for all images, two-sided permutation test,  $n = 10,000$  repeats) and increased with kernel diameter (Fig. 5e).

Together, these characteristics render the UV channel suitable for object detection, in particular for dark objects on top of a relatively bright (twilight) sky. We tested this hypothesis with a black drone mimicking an approaching aerial predator (Fig. 5f). As predicted, the drone was much more easily detectable in the UV vs. green channel, suggesting that this also holds for birds preying on mice during twilight, and therefore may contribute to increasing the species' survival chances in their natural habitat (Suppl. Video V3).

## Autoencoder model predicts colour opponency in the ventral retina

So far, we characterized the contrast statistics of scenes from the mouse natural environment and found significant differences in chromatic contrast between upper and lower visual field. Next, we wanted to explore if these differences can give rise to colour opponency and, hence, shape neural representations. Recently, Ocko and colleagues (2018) trained a convolutional autoencoder (CAE; Ballard, 1987; Hinton and Salakhutdinov, 2006) model to reconstruct pink ( $1/f$ ) noise and showed that the model learned spatial filters with center-surround organisation resembling the RFs of different RGC types. Thus, we next asked if such CAE models can also learn colour opponent RFs (Fig. 6). Autoencoder models feature a "bottleneck" that forces them to learn a "compressed" representation of the input images. This conceptually resembles the early visual system, where the information flow from the retina to downstream targets in the brain is constrained by bandwidth limitations (Gjorgjieva et al., 2014; Perge et al., 2009; Van Essen et al., 1991). We optimized our CAE model to yield good reconstructions under the qualitative constraint that the learned filters resemble smooth, centre-surround RFs (Suppl. Fig. S4). We then tested if colour-opponent kernels emerge in a CAE trained with our UV-green

scene crops, and whether the total number of such kernels differed between the upper vs. lower visual field.



**Figure 6 | Convolutional autoencoder (CAE) model learns colour opponent kernels from upper visual field footage.** **a**, Architecture of the CAE mode (FC, fully-connected; numbers indicate feature shapes; for details, see Methods). **b**, Reconstruction performance measured as structural similarity (SSIM) under different regularization strengths  $\alpha$  (L2) and  $\beta$  (L1), with two example images (input; top-left) and the corresponding CAE outputs depicted for certain combinations of  $\alpha$  and  $\beta$  (red boxes). **c**, Convolutional kernels learned by the model from upper (top) and lower (bottom) visual field scenes (for  $\alpha=10^4$ ,  $\beta=1/256$ ; see solid red box in (b); correlation indicated above each kernel pair;  $p<0.05$  for all kernel pairs except third one in bottom row ( $p=0.9$ )). Example for colour-opponent kernel indicated by box. **d**, 2D histogram showing total number of colour-opponent kernels per random seed learned from upper (top) and lower (bottom) visual field scenes ( $n=10$  random seeds). **e**, **f**, 16 kernels for UV and green, each for upper and lower visual field crops; same presentation and dataset as in (c). In (e), first kernels from PCA (Methods), with colour-opponent kernels indicated by boxes. The colour opponency index (COi; cf. Suppl. Fig. S2 and Methods) of the upper and lower visual field is 0.0393 and 0.0182, respectively. In (f), randomly selected kernels from ZCA whitening (Methods). Here, all 162 kernels (9x9 x2 channels; cf. (a)) were colour-opponent for the upper visual field, while only 143 were colour opponent for the lower visual field.

Classically, autoencoders are encouraged to learn an efficient encoding of the input by reducing the number of neurons in the hidden layer. Instead, we here encouraged redundancy reduction

of the input by adding Gaussian noise and imposing an L1 penalty ( $\beta$ ) to the activations of the encoder's output (Doi and Lewicki, 2007; Field, 1994; Ocko et al., 2018; for details, see Methods; van Rossum et al., 2003). To encourage smooth kernels akin to those measured in early visual neurons (e.g. Hubel and Wiesel, 1959; Marr and Hildreth, 1980), we additionally used L2 regularization ( $\alpha$ ) on the convolutional and deconvolutional layers, effectively constraining the norm of the weights (Vincent and Baddeley, 2003).

We trained the CAE model to minimize the mean squared error (MSE) for reconstructing images from the upper and lower visual field, separately for different regularizations and different random initialization seeds. We then quantified the model's performance in reconstructing the input images (Fig. 6b) using structural similarity as metrics (SSIM; Wang et al., 2004). Qualitatively, we searched for combinations of  $\alpha$  (L2) and  $\beta$  (L1) that used smooth centre-surround-like kernels in the presence of a bottleneck to reliably yield good reconstruction performance ( $SSIM \geq 0.6$ ; Fig. 6c). For these parameter combinations, we ran the model with  $n = 10$  random seeds and determined the number of colour-opponent spatial kernels (Fig. 6d). Kernels were considered colour-opponent if their UV and green channel were negatively correlated ( $p < 0.05$ ; Pearson correlation; cf. Fig. 6c).

We found that colour-opponent kernels were significantly more frequent in CAEs trained with scenes from the upper vs. the lower visual field ( $p = 0.012$ , permutation test; Fig. 6d). This predicts that in systems with an information bottleneck, UV-green colour-opponent kernels preferentially emerge when encoding visual scenes above the horizon.

We compared our CAE model with classical approaches that explain the unsupervised emergence of colour-opponency (Buchsbaum and Gottschalk, 1983) or centre-surround RF structure from natural stimuli (Bell and Sejnowski, 1997; Graham et al., 2006). We first applied PCA (Suppl. Fig. S2) to 9x9 image patches randomly drawn from the same set of image crops, and found that crops from the upper visual field have a higher  $CO_i$  (higher variance along the colour-opponent dimension; cf. Suppl. Fig. S2) than ones from the lower visual field (Fig. 6e). Unlike the CAE, the PCA kernels did not resemble the centre-surround RFs known in RGCs. Similarly, we applied zero-phase component analysis whitening (ZCA; Methods; Bell and Sejnowski, 1997). Also here, we found more colour-opponent kernels with images from the upper visual field (Fig. 6f). However, while these kernels were centre-surround, they were small and mainly different in spatial position.

In summary, all three unsupervised models confirm that specific chromatic statistics in the upper visual field may be sufficient to drive the emergence of colour-opponent spatial RFs.

## Discussion

Using a custom-build camera with spectral channels matching the UV and green sensitivities of mouse photoreceptors, we captured scenes from the natural visual environment of mice. Statistical analyses of these scenes revealed that chromatic contrast was stronger above the horizon, and that such contrast would preferentially drive large RFs. This enrichment of chromatic contrast in the upper visual field might underlie the preferential emergence of colour-opponent RFs found in the ventral retina, as suggested by a convolutional autoencoder model trained to represent our UV-green scenes. Furthermore, we found the upper visual field, and in particular the UV channel, to be biased towards dark contrast, which supports the view that UV could be an important source of information about dark objects (i.e. birds) against a brighter sky. Together, our findings lend further support to the idea that retinal circuits have evolved to optimally process natural scene statistics (reviewed in Baden et al., 2020).

### Recording natural scenes

Because we aimed at a camera that was portable, low-cost, and easily reproducible, we opted for the well-documented OV5647 camera chip of the second-generation Raspberry Pi camera. The main disadvantage of this choice was the relatively low light sensitivity of the chip, which restricted the recordings of dynamic natural scenes to daytime and twilight. With other camera chips — e.g. the new Raspberry Pi High Quality camera or smartphone cameras optimized for low-light recordings — a more nighttime-enabled mouse-camera, potentially also with a larger dynamic range, may be in reach. A further improvement may be opsin-matched filters (Tedore and Nilsson, 2019). Another interesting direction would be to miniaturize the camera to allow it to be mounted on a mouse while it roams its environment. Current head-mounted cameras for mice are restricted to grayscale or RGB (Froudarakis et al., 2014; Meyer et al., 2018, 2020; Michael et al., 2020) and therefore miss the UV channel altogether (see discussion in Franke et al., 2019). For an indoor environment lacking UV illumination this may not be relevant, but, as demonstrated in our study, when the settings approach the natural habitat of mice, a UV channel is a rich source of potentially behaviour-relevant visual information.

## Mouse camera movies as natural stimuli

Stimuli used in vision research cover a broad spectrum, ranging from artificial stimuli, like gratings and noise (reviewed in Rust and Movshon, 2005) or screen-rendered 3D objects (Froudarakis et al., 2020; Zoccolan et al., 2009), to more naturalistic ones (e.g. Baddeley et al., 1997; Betsch et al., 2004; Froudarakis et al., 2014). We argue that the movies recorded here represent suitable natural stimuli to probe mouse vision for the following reasons: (i) They were recorded outdoors in fields where mice can be seen during the day, and (ii) contain the spectral bands perceivable by mice. If presented on a UV-capable stimulator, they should drive the mouse visual system more efficiently than gray-scale stimuli on a standard monitor (discussed in Franke et al., 2019). (iii) Because our camera was mounted on a gimbal, the movies approximate well the input that reaches the mouse's eye, since a large fraction of eye movements in mice serve to stabilize the retinal image in the presence of head movements (Meyer et al., 2020). A first validation of these stimuli could be to test if they allow separating anatomically defined RCG types that so far could not be reliably disambiguated based on their responses to standard synthetic stimuli (Baden et al., 2016).

## Natural scene representations in the mouse early visual system

### *Retinal opsin gradient, colour opponency and colour vision*

Like many vertebrates, mice display regional opsin co-expression in cone photoreceptors (reviewed in Peichl, 2005): Specifically, ventral M-cones co-express S-opsin (Applebury et al., 2000; Baden et al., 2013; Röhlich et al., 1994; Szél et al., 1992), a challenge for a colour vision mechanism that relies on comparing signals from spectrally distinct cone types. Yet, our scene analysis showed chromatic contrast to be higher in the upper visual field, which is “viewed” by the heavily S-opsin dominated ventral retina. Moreover, the results from our autoencoder model suggests that the chromatic information in the upper visual field may have driven the presence of colour-opponent cells in the ventral retina over the course of evolution. In line with this prediction, a recent neurophysiological study showed that colour-opponent RGCs are indeed more frequent in the ventral retina (Szatko et al., 2020). At least some of these colour-opponent RGCs rely on antagonistic centre-surround RF mechanisms, with the RF centre dominated by signals from UV-biased S/(M)-coexpressing cones and the green-biased RF surround being mediated by rods (Joesch and Meister, 2016; Szatko et al., 2020) and/or by long-ranging inputs from M-cones in the dorsal retina (Chang et al., 2013). We also found more colour-opponent kernels for the upper vs. the lower visual field when employing a similar definition of

centre-surround opponency as Szatko and coworkers (Szatko et al., 2020). Interestingly, these mechanisms should profit from pooling signals from a large surround, which resonates well with our finding that larger RF cells should be better suited to pick up chromatic contrast. Further support for a crucial role of the ventral retina in promoting colour vision comes from behaviour: By probing the visual field of mice with chromatic stimuli, Denman et al. (2018) found that the animals succeeded in a colour change detection task for stimuli presented in the upper visual field.

If this chromatic regionalization of the retina was indeed evolutionary advantageous, why is it not more widespread among critters in similar habitats? For instance, some members of the subfamily Murinae, like the steppe mouse (*Mus spicilegus*), shares S/M opsin regionalization with the house mouse (*Mus musculus*), whereas others lack it (*Apodemus sylvaticus*) or lost the S-opsin completely (*Apodemus microps*) (Szel et al., 1994). Interestingly, these *Apodemus* species live in shrubberies, where a “bird-in-the-sky” detector may not be as useful as in relatively open spaces, as inhabited by *M. spicilegus*. Hence, spectral retinal regionalization in some mice may reflect an environmental adaptation — akin to the UV-sensitive “strike zone” in the ventral retina of zebrafish larvae, which helps the animal detect UV-reflecting prey and avoid potential predators (large dark shapes) in the upper visual field (Zimmermann et al., 2018). Curiously reminiscent of this “strike zone”, Nadal-Nicolás and colleagues (2020) recently found a S-cone “hotspot” in the ventral periphery of the mouse retina.

### *UV vision*

UV sensitivity is thought to play an important role in the behaviour of many vertebrates and invertebrates, including navigation, orientation, communication, foraging, as well as predator and prey detection (reviewed in Cronin and Bok, 2016). In rodents, UV vision has been mainly discussed in the context of communication via urine marks (Chávez et al., 2003; Joesch and Meister, 2016) and predator detection (e.g. Baden et al., 2013), but it may also play a role in foraging with respect to enhanced chromatic contrast of foliage (Tedore and Nilsson, 2019) and fruit (Altshuler, 2001). Our results shed light onto why mice specifically profit from UV sensitivity for predator detection: Consistent with previous studies (Baden et al., 2013; reviewed in Cronin and Bok, 2016), our results suggest that during daylight, when S-cones are at their prime, UV sensitivity enhances the detection of dark silhouettes against the bright sky. At twilight, the spectrum becomes more dominated by short wavelengths (e.g. increase in ratio between 360 and 440 nm, see Hut et al., 2000). From our footage, we estimated the intensity of the sky’s dome around sunset in the UV channel to be equivalent to  $\sim 300 R^* s^{-1}$  per cone. This is well

above the S-cone's threshold of  $\sim 20 R^*s^{-1}$  per cone (Naarendorp et al., 2010), even when including some UV-filtering by the mouse optics (33% transmission at 360 nm; Henriksson et al., 2010), and above the levels of  $\sim 100 R^*s^{-1}$  per rod below which rods are thought to take over (Rodieck, 1998). Moreover, recent V1 recording and modeling results suggest that around sunrise, mouse vision becomes quickly dominated by cones (Rhim et al., 2020). Under these conditions, S-cones, which are also less noisy than M-cones (Ala-Laurila et al., 2004), may indeed contribute to detecting dark shapes in front of the sky. Finally, at even lower intensities, when the rod photoreceptors dominate (Naarendorp et al., 2010), UV light from the sky may still play a role, because rhodopsin — like any opsin — features a secondary sensitivity peak (“beta band”) around  $\sim 350$  nm (Govardovskii et al., 2000). Therefore, it is conceivable that at low light levels, mice may perceive near-UV via the rods (discussed in Cronin and Bok, 2016).

#### *Bias towards dark contrasts*

Our analysis of mouse natural scenes is consistent with previous studies, showing that the distribution of contrast in natural scenes is biased towards dark contrasts (Cooper and Norcia, 2015; Ratliff et al., 2010). The dark bias in our natural scene footage was more prominent in the upper visual field and specifically in the UV channel, which resonated well with earlier retinal recordings showing that light-off steps are particularly faithfully encoded by ventral cones (Baden et al., 2013). Similar to previous findings that describe an overrepresentation of Off information in the visual system (Kremkow et al., 2016; Schröder et al., 2020; Yeh et al., 2009), we also found that dark contrasts elicited stronger responses than bright contrasts, particularly in ventral RGCs. Interestingly, the dark bias was most prominent in large-RF RGCs, which ties into an ongoing controversy if Off cells are expected to feature large (e.g. Cooper and Norcia, 2015) or small RFs (see discussion in Mazade et al., 2019; e.g. Ratliff et al., 2010). We currently lack the corresponding RF data for dorsal RGCs, and therefore the question remains if the balanced contrast distribution in the lower visual field is also reflected in a more balanced distribution of On/Off preferences.

#### *Eye movements in the context of a horizontally regionalized visual field*

As discussed above, several lines of evidence suggest that the dorsal and ventral mouse retina are functionally specialized. Such dorso-ventral regionalization is expected to be advantageous only if it is ensured that the upper and lower visual field is reliably projected onto the ventral and dorsal retina, respectively. This requires a different visual orienting strategy compared to primates, who scan the visual environment with their foveas using head-motion independent

eye movements (Baden et al., 2020; reviewed in Land, 2015). Eye movements in mice are tightly coupled to head movements, and often serve to stabilize the retinal image with respect to the ground (Meyer et al., 2018, 2020; Michael et al., 2020). This presumably ensures that the temporo-nasal retinal axis is aligned with the horizon and would thus allow the mouse retina with its dorso-ventral specialization to make use of the respective differences in natural scene statistics.

### Are natural scene statistics encoded in retinal circuits?

Autoencoders have long been used to learn efficient representations by restricting the coding capacity of the hidden layer(s) (Hinton and Salakhutdinov, 2006; Kramer, 1991). Therefore, they are a natural choice for exploring feature transformations in the early visual system, such as between retina and primary visual cortex (V1), where visual information has to pass the optic nerve, posing a severe bottleneck. Recently, Ocko and colleagues (2018) trained a convolutional autoencoder with pink ( $1/f$ ) noise, which mimics the distribution of spatial frequencies in natural scenes (Field, 1987). They found that it learned spatial filters that resembled centre-surround RFs of a subset of primate RGC types. We used a similar CAE architecture trained with natural UV-green images, which learned, in line with Ocko et al. (2018), not only spatial filters resembling diverse RGC RFs, but also colour-opponency (Buchsbaum and Gottschalk, 1983). While other approaches, such as PCA and ZCA whitening, also revealed colour-opponency and, in case of ZCA whitening, centre-surround structure, the CAE was unique in the diversity of resulting spatial filters reminiscent of different RGC types. In the future, it would be informative to investigate which types of regularization mimic the specific constraints represented by bottlenecks in the early visual system (and why) and, by systematically manipulating the input images, what particular features give rise to colour opponency (discussed in Chalk et al., 2018). Also, it would be interesting to explore if the kernels learned by the CAE indeed help predicting RGC responses to natural or synthetic stimuli.

## Acknowledgments

We thank Tom Baden, Heiko Schütt, and Matthias Bethge for helpful discussions, and Gordon Eske for excellent technical assistance. This work was supported by the German Research Foundation (DFG) through Collaborative Research Centre CRC 1233 (project number 276693517); the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant (agreement No 674901); and the Max Planck Society (M.FE.A.KYBE0004); the German Ministry of Research and Education (01GQ1002). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

Conceptualization Ideas: F.S., K.R., L.B., and T.E.; Methodology: Y.Q., Z.Z., L.B., and T.E., with input from D.K. and M.K.; Formal Analysis: Y.Q. with input from D.K., L.B., and T.E.; Investigation: Y.Q., Z.Z., K.S., with input from K.F., L.B. and T.E.; Writing – Original Draft: Y.Q., L.B., and T.E.; Writing – Review & Editing Preparation: Y.Q., Z.Z., D.K., K.S., M.K., K.F., L.B., and T.E.; Visualization: Y.Q., and T.E.; Supervision: L.B. and T.E.; Funding Acquisition: L.B., F.S., K.R., and T.E.

## Declaration of Interests

The authors declare no competing interests.

## Tables

Part (part number)	Company	Distributor (link)	# of parts
Gimbal (Crane 2)	Zhiyun, China	<a href="#">Zhiyun Tech</a>	1x
DM (F48-442)	AHF, Germany	<a href="#">AHE</a>	1x
GPF (F47-510)	AHF, Germany	<a href="#">AHE</a>	1x
UPF (F37-424)	AHF Analysentechnik AG, Germany	<a href="#">AHE</a>	1x
Raspberry Pi Zero W (EXP-R12-143)	<a href="#">Raspberry Pi Foundation</a> , UK	e.g. EXP-TECH, Germany	2x
Zero spy camera (EXP-R15-1309)	<a href="#">Raspberry Pi Foundation</a> , UK	e.g. EXP-TECH, Germany	2x
Powerbank (ANKER 1252011)	<a href="#">Anker Technology</a> , UK	e.g. Reichelt, Germany	1x
Fisheye lens (LS-40180)		e.g. <a href="#">Watterott</a> , Germany	1x
Bi-Convex lens, L04, $f = 35 \text{ mm}$ (LB4879-A)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Bi-Convex lens, L05, $f = 20 \text{ mm}$ (LB4854-A)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Plano convex lens, L03, $f = 150 \text{ mm}$ , (LA4874-A)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Plano convex lens, L01/02, $f = 50 \text{ mm}$ , (LA4148-A)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	2x
Iris diaphragm, P1, 2 (SM1D12C)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	2x
Protected silver mirror (PF10-03-P01)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Kinematic mirror mount (KCB1/M)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Generic 3.3V green LEDs for camera synchronisation	Conrad, Germany	<a href="#">Conrad, Germany</a>	2x
Cage Cube (CM1-DCH/M)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Slip Plate Positioner (SPT1/M)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x
Cage Plate	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	1x

(CP02T/M)			
Cage Plate (CP02/M)	Thorlabs, Germany	<a href="#">Thorlabs, Germany</a>	8x

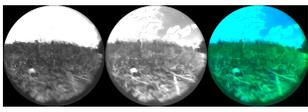
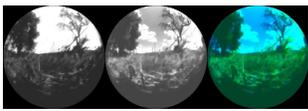
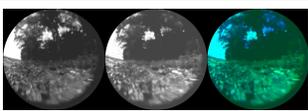
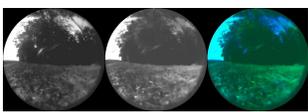
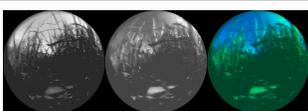
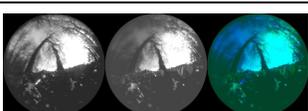
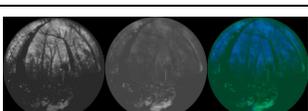
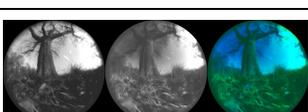
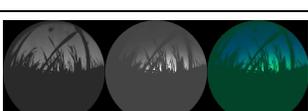
**Table 1** | Camera parts.

Parameter	UV camera (CS1)	Green camera (CS2)
Resolution	(720,540)	(720,540)
Frame rate [Hz]	25	25
Exposure_mode	off	off
Shutter speed [ms]	35	35
ISO	800	100
Analog_gain	8.0 - 8.2	1.0 - 1.2
Awb_mode	off	off
Awb_gains	(1,1)	(1,1)
$P = av^b + c$ , with normalized pixel value $v$ , and power $P$ (in [ $\mu$ W])	$a = 0.741$ , $b = 2.102$ , $c = 0.015$	$a = 6.197$ , $b = 4.331$ , $c = 0.366$

**Table 2** | Camera settings.

Resource	Company, source, description	Link
Python 3.6	Python programming language	<a href="https://www.python.org/">https://www.python.org/</a>
OpenCV	Framework for computer vision	<a href="https://opencv.org/">https://opencv.org/</a>
Pytorch	Machine learning library	<a href="https://pytorch.org/">https://pytorch.org/</a>
FreeCAD	Open source 3D construction software	<a href="https://www.freecadweb.org/">https://www.freecadweb.org/</a>
VNC Viewer	Screen-sharing and remote-control software, RealVNC Ltd, UK	<a href="https://www.realvnc.com/">https://www.realvnc.com/</a>
Spectrometer (STS-UV)	UV-enabled spectrometer, Ocean Optics, Germany	<a href="#">Product sheet</a>
Power meter (842-PE)	Newport, Germany	<a href="#">Newport power meter detector overview</a>
360 nm LED (XSL-360-5E)	<a href="#">Roithner, Austria</a>	
380 nm LED (VL380-5-15)	<a href="#">Roithner, Austria</a>	
400 nm LED (XRL-400-5E)	<a href="#">Roithner, Austria</a>	
490 nm LED (LED490-06)	<a href="#">Roithner, Austria</a>	
525 nm LED (G58A5111P)	<a href="#">Roithner, Austria</a>	
Spectral sensitivity of Raspberry Pi camera	Depending on the camera version, different chips are used	<a href="#">Camera module versions</a> <a href="#">Sensitivity spectra</a>

**Table 3** | Resources (except camera parts).

Movie file name (yyyymmdd_hh_#)	Duration (mm:ss)	Example frames	Brief description
20180713_13_1 20180713_13_2	01:52 01:17		Daytime, early summer, few trees
20180713_13_3	01:27		Daytime, early summer, open sky
20180713_14_1	01:53		Daytime, early summer, few trees
20180905_14_1 20180905_14_2	01:44 01:52		Daytime, early fall, close to the forest
20180911_14_1 20180911_14_2	01:38 01:42		Daytime, early fall, close to the forest
20190329_12_1 20190329_12_2	01:54 01:54		Daytime, early spring, grass
20190329_12_3 20190329_12_4	01:50 01:54		Daytime, early spring, close to the forest
20190329_13_1 20190329_13_2	01:43 01:47		Daytime, early spring, close to the forest
20190329_13_3	01:51		Daytime, early spring, few trees
20190329_18_1 20190329_18_2	00:07 00:02		Dusk, early spring, a black drone

**Table 4** | *Natural movies.*

Group	$d_{RE}(\circ)$	Upper UV vs. Upper G	Upper UV vs. Lower UV	Upper UV vs. Lower G	Upper G vs. Lower UV	Upper G vs. Lower G	Lower UV vs. Lower G
$I_1$	2	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	6	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	10	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	14	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
$I_2$	2	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	6	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	10	<0.0001	0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	14	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
$I_3$	2	<0.0001	0.003	<0.0001	<0.0001	0.9249	<0.0001
	6	<0.0001	0.0114	<0.0001	<0.0001	<0.0001	<0.0001
	10	0.0009	<0.0001	0.0268	<0.0001	0.206	<0.0001
	14	0.0082	<0.0001	0.0003	<0.0001	0.1825	<0.0001

**Table 5** | Statistics of  $C_{RMS}$  between chromatic channels and upper vs. lower visual field as a function of kernel diameter. Two-sided permutation test with 10,000 repeats. White, light orange and dark orange indicating  $p < 0.0001$ ,  $0.0001 \leq p < 0.05$  and  $p \geq 0.05$

Group		2° vs. 6°	2° vs. 10°	2° vs. 14°	6° vs. 10°	6° vs. 14°	10° vs. 14°
I <sub>1</sub>	Upper UV	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Upper G	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Lower UV	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Lower G	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
I <sub>2</sub>	Upper UV	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Upper G	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Lower UV	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Lower G	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
I <sub>3</sub>	Upper UV	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Upper G	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Lower UV	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Lower G	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

**Table 6** | Statistics of  $C_{RMS}$  between kernel diameters. Two-sided permutation test with 10,000 repeats.

Group	$d_{RE}(\circ)$	Upper UV vs. Upper G	Upper UV vs. Lower UV	Upper UV vs. Lower G	Upper G vs. Lower UV	Upper G vs. Lower G	Lower UV vs. Lower G
$I_1$	2	<0.0001	0.0003	<0.0001	<0.0001	<0.0001	<0.0001
	6	<0.0001	<0.0001	<0.0001	0.0021	0.9848	0.0042
	10	<0.0001	<0.0001	<0.0001	0.82	<0.0001	<0.0001
	14	<0.0001	<0.0001	<0.0001	<0.0001	0.0014	<0.0001
$I_2$	2	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0584
	6	<0.0001	<0.0001	<0.0001	0.0305	0.0005	0.2765
	10	<0.0001	<0.0001	<0.0001	0.0007	0.0002	0.5104
	14	<0.0001	<0.0001	<0.0001	0.0006	<0.0001	<0.0001
$I_3$	2	0.0007	0.0383	0.0207	0.1577	0.1809	0.8714
	6	0.3987	0.393	0.0194	0.9813	0.1046	0.1114
	10	0.0011	0.8974	0.0407	0.0016	0.4223	0.0338
	14	<0.0001	<0.0001	0.0011	0.0047	0.0001	0.0468

**Table 7** | Statistics of  $C_{On-Off}$  between chromatic channels and upper vs. lower visual field as a function of kernel diameter. Two-sided permutation test with 10,000 repeats. White, light orange and dark orange indicating  $p < 0.0001$ ,  $0.0001 \leq p < 0.05$  and  $p \geq 0.05$ .

Group		2° vs. 6°	2° vs. 10°	2° vs. 14°	6° vs. 10°	6° vs. 14°	10° vs. 14°
I <sub>1</sub>	Upper UV	<0.0001	<0.0001	<0.0001	0.0645	0.0002	0.0317
	Upper G	<0.0001	<0.0001	<0.0001	0.0051	0.0022	0.7299
	Lower UV	0.004	0.0383	<0.0001	0.4436	<0.0001	<0.0001
	Lower G	0.0001	0.1376	0.0005	0.0449	0.7307	0.0862
I <sub>2</sub>	Upper UV	<0.0001	<0.0001	<0.0001	0.0176	<0.0001	<0.0001
	Upper G	<0.0001	<0.0001	<0.0001	0.0003	<0.0001	<0.0001
	Lower UV	0.25	0.0029	<0.0001	0.0919	<0.0001	<0.0001
	Lower G	0.0694	<0.0001	<0.0001	0.023	<0.0001	0.0635
I <sub>3</sub>	Upper UV	0.0482	0.2692	0.8564	0.3549	0.0308	0.1902
	Upper G	<0.0001	<0.0001	<0.0001	0.0008	<0.0001	<0.0001
	Lower UV	0.3934	0.4386	<0.0001	0.9422	<0.0001	<0.0001
	Lower G	0.0062	0.0013	<0.0001	0.538	0.0039	0.0308

**Table 8** | Statistics of  $C_{On-off}$  between chromatic channels and upper vs. lower visual field as a function of kernel diameter. Two-sided permutation test with 10,000 repeats. White, light orange and dark orange indicating  $p < 0.0001$ ,  $0.0001 \leq p < 0.05$  and  $p \geq 0.05$

## Methods

### Camera design

As the front lens of our camera (Fig. 1a), we used a fisheye lens ( $f = 1.05 \text{ mm}$ ,  $f/2$ , LS-40180, Watterott, Germany) with a FOV of approx.  $180^\circ$ . Following an inverted periscope design, the lens was mounted at the bottom end of the camera, which allowed us to capture the scene from a distance of approx. 2-5 cm to the ground (Fig. 1b-f). After passing two relay lenses (L04, LB4879-A,  $f = 35 \text{ mm}$ ; L05, LB4854-A,  $f = 20 \text{ mm}$ ; both Thorlabs, German), which allowed us to transfer the fisheye lens' FOV to the camera sensor, a silver mirror reflected the light towards a dichroic mirror (F48-442, AHF, Germany; reflection,  $> 90\%$ ,  $350 - 442 \text{ nm}$ ; transmission,  $> 90\%$ ,  $460 - 650 \text{ nm}$ ; Fig. 1c) that reflected wavelengths shorter than approx.  $440 \text{ nm}$  towards the first camera sensor (CS1) and transmitted longer wavelengths towards the second camera sensor (CS2; Fig. 1a). An additional spectral bandpass filter in front of each camera chip ( $\text{BP}_G$ , F47-510,  $> 90\%$ ,  $470 - 550 \text{ nm}$ ;  $\text{BP}_{UV}$ , F37-424,  $> 90\%$ ,  $350 - 419 \text{ nm}$ ; both AHF) restricted the light reaching the chips to the approx. spectral ranges relevant for mouse opsins (Fig. 1c). Because the spectral properties of dichroic filters (DM,  $\text{BP}_{UV}$ ,  $\text{BP}_G$ ) change with the incident angle of the light, we used relay lenses L1, L2 (LA4148-A,  $f = 50 \text{ mm}$ ; Thorlabs, Germany) and L3 (LA4874-A,  $f = 150 \text{ mm}$ ; Thorlabs) to ensure that light passes the dichroic bandpass filters collimated before being focussed onto the camera chips. In addition, we added iris diaphragms (SM1D12C; adjusted to  $\sim 2 \text{ mm}$  pinhole diameter) directly in front of the cameras to optimize depth-of-field and image contrast. For the parts list, including mechanical parts, see Table 1.

For each chromatic channel, we used a Zero spy camera module (EXP-R15-1309, EXP-TECH, Germany) connected to separate Raspberry Pi Zero W single-board computers (EXP-R12-143, EXP-TECH), which were powered by a 15 Ah USB power bank (Anker 1252011, Reichelt, Germany). Lenses and infrared filters had been removed from the camera modules. To increase the UV sensitivity of the chip in the UV pathways, we mechanically removed its RGB Bayer layer, following a procedure described by Wilkers et al. (2016).

### *Movie recordings*

Movies were recorded onto the Raspberry Pi's Flash memory card; movie capture was remote-controlled from a laptop connected to the Raspberry Pis either via USB cable or an ad-hoc Wifi network (VNC Viewer, RealVNC Ltd, UK). Camera parameters were fixed for all recordings (Table 2). To stabilize the camera during the recordings, we mounted it on a gimbal

(Crane 2, Zhiyun, China; Fig. 1d,e). Since the camera weighed around 1 kg and its point of gravity was not centred, we added counterweights for the gimbal to work properly. When moving, we tried to maintain an azimuth angle of  $\sim 60^\circ$  between the optical axis of the fisheye lens and movement direction, close to the azimuth angle of the mouse' eye (Oommen and Stahl, 2008; Stabio et al., 2018; Sterratt et al., 2013).

### *Temporal alignment*

Since we recorded with two camera chips simultaneously, the resulting movies needed to be temporally and spatially (next section) synchronized. For temporal alignment, we used LEDs mounted close to each camera chip (Fig. 1a) and flashed (200-ms pulses every 20 seconds) them as synchronisation markers. In addition, we manually checked the temporal alignment by comparing frames from the two channels (e.g. during fast movements).

### *Spatial alignment*

Because we used optical rails to build the mouse-cam's optical pathway, the UV and green channels captured almost the same scene. To account for potential spatial offsets and differences in image magnification, we used a homography matrix ( $H$ ) to spatially relate a point  $(x_1, y_1)^T$  in the first channel to a corresponding point  $(x_2, y_2)^T$  in the other channel:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}.$$

To this end, we first extracted at least 20 feature points from both channels using the scale-invariant feature transform (SIFT) approach (Lowe, 2004) and matched these feature points using the k-nearest neighbors algorithm. Next,  $H$  was determined by a random sample consensus algorithm (Fischler and Bolles, 1981), allowing us to project all pixels in the first channel (green) to those in the second channel (UV), with the same  $H$  typically working for all frames of a particular recording.

### **Spectral calibration**

To map image pixel values ( $v_{Raw}$ , 0...255) to absolute intensities (in  $[\mu W]$ ), we first intensity-calibrated LEDs peaking in the UV ( $\lambda_{Peak} = 360, 380, 400 \text{ nm}$ ; see Table 3; Suppl. Fig. S1a) or the green band ( $\lambda_{Peak} = 490, 525 \text{ nm}$ ) using a power meter (842-PE, Newport, Germany). Next we recorded images of these LEDs set to different intensities to determine the

relationship between normalized pixel value ( $v$ , 0...1) and power ( $P$ ) by fitting the data to  $P = av^b + c$  (Suppl. Fig. S1b; for coefficients, see Table 2) using the Levenberg-Marquardt algorithm for least squares curve fitting. Ground-truth spectral images of the natural scenes were acquired with a scanning spectrometer, as described earlier (Baden et al., 2013). In brief, this custom-built device consisted of two servo motors that moved the fibre of a USB-spectrometer (STS-UV, Ocean Optics, Germany) to rasterize the scene with a resolution of  $\sim 10^\circ$  of visual angle (Suppl. Fig. S1d,f, right).

To use the available range for both colour channels, our intensity correction linearly mapped an intensity range of  $0.02 - 0.76 \mu W$  (UV) and  $0.37 - 6.56 \mu W$  (green) to pixel values between  $5 - 255$  and  $14 - 255$ , respectively (see Discussion). Note that for better visualisation, we applied gamma correction ( $v' = v^{1/\gamma}$ , with  $\gamma = 2.2$ ; (Poynton, 2003)) to the images in the figures (e.g. Suppl. Fig. S1c,d vs S1f).

### Statistical analysis of the natural scenes

A movie frame contained a circular FOV of  $\sim 180^\circ$ , corresponding to 437 pixels along the diameter. To minimize the influence of potential chromatic and spatial aberrations introduced by the fish-eye lens, we focused on image cut-outs (“crops”;  $(53^\circ)^2$ , equivalent to  $(128 \text{ pixels})^2$  in size) from the central upper and lower visual field. For contrast analysis, we excluded image crops that contained more than 30% underexposed ( $v_{Raw}(G) < 15$ ,  $v_{Raw}(UV) < 6$ ) or overexposed ( $v_{Raw}(G), v_{Raw}(UV) > 254$ ) pixels. We randomly sampled one image crop every 10 frames from all movies (Fig. 1b,e) until we had 1,500 crops for each upper and lower visual field. Next, the image crops were divided into three “intensity classes” ( $I_{low}$ ,  $I_{median}$ ,  $I_{high}$ ) by two percentiles ( $1/3$  and  $2/3$ ) by their mean intensities.

Note that scene content varied somewhat with the image group, because scenes in or near the forest were usually dimmer (“low mean”) than those with open skies, bare ground and little vegetation (“high mean”). Since this bias was merely a side-effect of our mean intensity criterion, we refrained from linking statistics to scene content. In the future, it would be interesting to explore how scene content affects chromatic contrast statistics, for instance by classifying the images with pre-trained CNN models like VGG (Simonyan and Zisserman, 2014) before the statistical analyses. This could, for example, shed light on the question why contrast differences generally tended to decrease towards the high mean intensity group — that is, whether this is simply due to the limited dynamic range of the camera or in fact scene

content-dependent. To determine  $C_{RMS}$  and  $C_{On-Off}$ , we randomly picked in each class for all images 10 locations per crop.

#### *Root mean square (RMS) contrast*

In psychophysical studies,  $C_{RMS}$  is commonly used for estimating contrast in natural scenes (Mante et al., 2005) and defined as:

$$C_{RMS}(x, y) = \frac{\sigma(I_{center}(x, y))}{\mu(I_{center}(x, y))}$$

where  $\sigma(I_{Center}(x, y))$  and  $\mu(I_{Center}(x, y))$  are standard deviation and mean, respectively, of the normalized pixel intensities contained in an image spot (“receptive field”; RF) centred at  $(x, y)$  within the image crop. Spot diameters ( $d_{RF}$ ) ranged from 2 to 14 degrees of visual angle.

#### *On-Off contrast*

We measured On-Off contrast ( $C_{On-Off}$ ) at a point  $(x, y)$  using a difference-of-Gaussians (DOG) kernel with a normalized denominator to restrict the value range to  $[-1, 1]$ :

$$C_{ON-OFF}(x, y) = \frac{I_{Centre}(x, y) - I_{Surround}(x, y)}{I_{Centre}(x, y) + I_{Surround}(x, y)}$$

where  $I_{Centre}(x, y)$  and  $I_{Surround}(x, y)$  represent the summed pixel intensities after convolving the image with the centre and surround Gaussian kernels, respectively. The spatial relationship between centre and surround Gaussians were  $\sigma_{Surround} = 1.5\sigma_{Centre}$ ; the total DOG kernel size was  $3\sigma_{Centre}$ . Note that  $d_{RF}$  was defined by the zero crossing radius of the kernel:

$$f(r|\mu, \sigma_{Centre}^2) = g(r|\mu, (1.5 \cdot \sigma_{Centre})^2)$$

where  $f(r|\mu, \sigma_{Centre}^2)$  and  $g(r|\mu, (1.5\sigma_{Centre})^2)$  are centre and surround Gaussians, respectively. We ensure that the kernel response to a homogeneous input image 0 by setting  $r = 1.2\sigma_{Centre}$ . A negative kernel response resulted in  $C_{On-Off} < 0$ , indicating a negative contrast at this sample location. As for  $C_{RMS}$ ,  $d_{RF}$  ranged from 2° to 14°.

#### *On-Off index*

To test if ventral mouse RGCs prefer dark contrasts, we reevaluated a published dataset with recordings of ventral RGCs (Baden et al., 2016). From this dataset, we extracted for all On (groups 1-9) and Off (groups 15-32) RGCs that passed the quality criterion ( $Q_i > 0.2$ ;  $n = 2,380$

cells) the On-Off index ( $cell\_oo\_idx$ ,  $OOi$ ; see below) and the RF diameter ( $rf\_size$ , in [ $\mu m$ ]). The latter was converted from [ $\mu m$ ] into [ $^\circ$ ] of visual angle, assuming  $1^\circ \approx 30 \mu m$  on the mouse retinal surface. The On-Off preference ( $OOi$ ) of a cell and was defined as:

$$OOi = \frac{\langle r_{On} \rangle_t - \langle r_{Off} \rangle_t}{\langle r_{On} \rangle_t + \langle r_{Off} \rangle_t}$$

with  $r_{On}$  and  $r_{Off}$  defined as the activity during the response to the leading and the trailing edge of a bright-on-dark moving bar stimulus, respectively.

### Comparing contrast distributions

To test if two contrast distributions originate from the same distribution, we performed a two-sided permutation test with 10,000 repeats to estimate the p-value. In addition, as a metric for similarity between the contrast distributions in the UV ( $P_{UV}$ ) and the green ( $P_G$ ) channel, we used Jensen-Shannon divergence (JSD):

$$JSD(P_{UV}||P_G) = \frac{D(P_{UV}||M)}{2} + \frac{D(P_G||M)}{2},$$

with the Kullback-Leibler (KL) divergence ( $D$ ) defined as:

$$D(P||Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right),$$

and  $M = 0.5(P_{UV} + P_G)$ . Instead of KL divergence, we used JSD because it is symmetric and bounded (0...1 for log base 2).

### PCA and ZCA whitening

Buchsbaum and Gottschalk (1983) have shown that achromatic and chromatic visual channels can be obtained using principal component analysis (PCA). As an extension, zero-phase component analysis (ZCA) whitening was shown to decorrelate signals and learned centre-surround-like kernels (Bell and Sejnowski, 1997). Accordingly, we defined a covariance matrix  $C$  of the original centred data  $\vec{x}$ . By applying transformation matrix  $W$ , we would get uncorrelated data  $\hat{\vec{x}}$

$$\hat{\vec{x}} = W\vec{x}.$$

We performed PCA (without whitening) on the image crops using:

$$W = U^T$$

ZCA whitening was performed using

$$W = U(D + \epsilon)^{-\frac{1}{2}}U^T$$

where  $U$  and  $D$  contain eigenvectors and eigenvalues of  $C$ , respectively, and  $\epsilon = 10^{-8}$  for numerical stability.

We first applied PCA on intensities of chromatic channels. The eigenvector with two positive or two negative entries corresponds to an achromatic transformation and the other eigenvector with one positive and one negative entry corresponds to a chromatic (colour-opponent) transformation. We also applied PCA on 9x9 image patches. We defined the colour-opponent transformation when the kernels from the UV and green channels were negatively correlated (pearson correlation coefficient,  $p < 0.05$ ). In both cases, the colour opponency index ( $CO_i$ ), which represents the ratio between signal variance in the colour-opponent dimensions to the variance in all dimensions, was defined as:

$$CO_i = \frac{\sum \lambda_{CO}}{\sum \lambda}$$

where  $\lambda_{CO}$  denotes the eigenvalues of the colour-opponent transformation, and  $\lambda$  all eigenvalues.

### Convolutional autoencoder model

We prepared datasets from the upper and the lower visual field separately. For both datasets, 10,000 image crops ( $\bar{x}_i$  and  $i$  represents image index) with 56x56 pixels each were randomly picked and then rescaled to 28x28 pixels. These image crops met the same quality criteria (fewer than 30% of pixels under/over-exposure) as those for the statistical analysis. Among them, 9,000 image crops were used for training and the rest for validation and testing. The images were offset-corrected separately in each chromatic channel (by subtracting the channel's mean intensity).

We implemented a simple convolutional autoencoder model (CAE; Fig. 6a) following (Ocko et al., 2018) using PyTorch (Paszke et al., 2019). The encoder contained a single convolutional layer (with weights denoted  $\vec{w}_c$ ) followed by a rectified linear unit (ReLU) function, one fully-connected (FC) layer and another ReLU function. Following Ocko et al. (2018), Gaussian noise with  $\sigma = 1$  was added to the encoder output to restrict the channel capacity. The decoder contained one FC layer, one ReLU function, a single deconvolutional layer (with weights denoted  $\vec{w}_d$ ), and a hyperbolic tangent (tanh) function to map back to the original data range

(-1...1). We used 16 convolutional kernels with a filter size of 9x9 pixels for each chromatic channel, with zero-padded boundaries and without downsampling. Correspondingly, the deconvolutional kernels consisted of 9x9 pixel filters per input channel. Thus, the size of the activation tensor after the first convolution was 28x28x16 (height x width x channel), which was flattened into a 12,544 dimensional vector before it was fed into the FC layer. The two FC layers had the same input and output size. The loss function was defined as:

$$L_{CAE} = \sum_i (\bar{x}_i - \hat{x}_i)^2 + \alpha(\|\bar{w}_c\|_2 + \|\bar{w}_d\|_2) + \beta\|\vec{h}\|_1$$

The first term is the MSE between prediction  $\hat{x}_i$  and ground truth  $\bar{x}_i$ , the second term is the L2 penalty (hyperparameter  $\alpha$ ) on the weights of the convolutional ( $\bar{w}_c$ ) and deconvolutional ( $\bar{w}_d$ ) layers, and the third term is the L1 penalty (hyperparameter  $\beta$ ) on the encoder output (see also Results).

The CAE models were trained for 100 epochs with 100 image crops in each mini-batch (learning rate,  $\eta = 10^{-4}$ ) using the Adam optimizer (Kingma and Ba, 2014) to minimize the loss functions. Image reconstruction performance of the CAE was estimated based on structural similarity (SSIM; Wang et al., 2004) and MSE (Fig. 6b; Suppl. Fig. S4b, respectively).

Hyperparameters  $\alpha$  and  $\beta$  were adjusted via grid search. We aimed at a trade off between reconstruction performance ( $SSIM \geq 0.6$  or  $MSE \leq 0.01$ ) and regularizations (mitigation of overfitting), which we found for combinations of  $\alpha = 10^3, 10^4$  and  $\beta = 0 - 1/16$ . Next, we performed a permutation test with 10,000 repeats to check if the models trained with images from the upper visual field learned colour-opponent kernels more frequently than those trained with lower visual field images. We compared the number of colour-opponent kernels generated under the two input conditions using a two-sided permutation test.

## References

- Abbasi-Asl, R., Pehlevan, C., Yu, B., and Chklovskii, D. (2016). Do retinal ganglion cells project natural scenes to their principal subspace and whiten them? 2016 50th Asilomar Conference on Signals, Systems and Computers.
- Ala-Laurila, P., Donner, K., and Koskelainen, A. (2004). Thermal activation and photoactivation of visual pigments. *Biophys. J.* *86*, 3653–3662.
- Altshuler, D.L. (2001). Ultraviolet reflectance in fruits, ambient light composition and fruit removal in a tropical forest. *Evol. Ecol. Res.* *3*, 767–778.
- Applebury, M.L., Antoch, M.P., Baxter, L.C., Chun, L.L., Falk, J.D., Farhangfar, F., Kage, K., Krzystolik, M.G., Lyass, L.A., and Robbins, J.T. (2000). The murine cone photoreceptor: a single cone type expresses both S and M opsins with retinal spatial patterning. *Neuron* *27*, 513–523.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.* *61*, 183–193.
- Baddeley, R., Abbott, L.F., Booth, M.C., Sengpiel, F., Freeman, T., Wakeman, E.A., and Rolls, E.T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Biol. Sci.* *264*, 1775–1783.
- Baden, T., Schubert, T., Chang, L., Wei, T., Zaichuk, M., Wissinger, B., and Euler, T. (2013). A tale of two retinal domains: near-optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution. *Neuron* *80*, 1206–1217.
- Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature* *529*, 345–350.
- Baden, T., Euler, T., and Berens, P. (2020). Understanding the retinal basis of vision across species. *Nat. Rev. Neurosci.* *21*, 5–20.
- Ballard, D.H. (1987). Modular Learning in Neural Networks. In *AAAI*, pp. 279–284.
- Barlow, H.B. (1961). Possible Principles Underlying the Transformations of Sensory Messages. *Sensory Communication* 216–234.
- Bell, A.J., and Sejnowski, T.J. (1997). The “independent components” of natural scenes are edge filters. *Vision Res.* *37*, 3327–3338.
- Betsch, B.Y., Einhäuser, W., Körding, K.P., and König, P. (2004). The world from a cat’s perspective--statistics of natural videos. *Biol. Cybern.* *90*, 41–50.
- Bex, P.J., and Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* *19*, 1096–1106.
- Buchsbaum, G., and Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* *220*, 89–113.
- Burge, J., and Jaini, P. (2017). Accuracy Maximization Analysis for Sensory-Perceptual Tasks: Computational Improvements, Filter Robustness, and Coding Advantages for Scaled Additive

Noise. *PLoS Comput. Biol.* *13*, e1005281.

Burton, G.J., and Moorhead, I.R. (1987). Color and spatial structure in natural scenes. *Appl. Opt.* *26*, 157–170.

Chalk, M., Marre, O., and Tkačik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 186–191.

Chang, L., Breuninger, T., and Euler, T. (2013). Chromatic coding from cone-type unselective circuits in the mouse retina. *Neuron* *77*, 559–571.

Chávez, A.E., Bozinovic, F., Peichl, L., and Palacios, A.G. (2003). Retinal spectral sensitivity, fur coloration, and urine reflectance in the genus octodon (rodentia): implications for visual ecology. *Invest. Ophthalmol. Vis. Sci.* *44*, 2290–2296.

Cooper, E.A., and Norcia, A.M. (2015). Predicting cortical dark/bright asymmetries from natural image statistics and early visual transforms. *PLoS Comput. Biol.* *11*, e1004268.

Cronin, T.W., and Bok, M.J. (2016). Photoreception and vision in the ultraviolet. *J. Exp. Biol.* *219*, 2790–2801.

Datta, S.R., Anderson, D.J., Branson, K., Perona, P., and Leifer, A. (2019). Computational Neuroethology: A Call to Action. *Neuron* *104*, 11–24.

Denman, D.J., Luviano, J.A., Ollerenshaw, D.R., Cross, S., Williams, D., Buice, M.A., Olsen, S.R., and Reid, R.C. (2018). Mouse color and wavelength-specific luminance contrast sensitivity are non-uniform across visual space. *Elife* *7*.

Doi, E., and Lewicki, M.S. (2007). A Theory of Retinal Population Coding. In *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J.C. Platt, and T. Hoffman, eds. (MIT Press), pp. 353–360.

Dyballa, L., Hoseini, M.S., Dadarlat, M.C., Zucker, S.W., and Stryker, M.P. (2018). Flow stimuli reveal ecologically appropriate responses in mouse visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 11304–11309.

Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* *4*, 2379–2394.

Field, D.J. (1994). What is the goal of sensory coding? *Neural Comput.* *6*, 559–601.

Fischler, M.A., and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* *24*, 381–395.

Franke, K., Maia Chagas, A., Zhao, Z., Zimmermann, M.J., Bartel, P., Qiu, Y., Szatko, K.P., Baden, T., and Euler, T. (2019). An arbitrary-spectrum spatial visual stimulator for vision research. *Elife* *8*.

Frazor, R.A., and Geisler, W.S. (2006). Local luminance and contrast in natural images. *Vision Res.* *46*, 1585–1598.

Froudarakis, E., Berens, P., Ecker, A.S., Cotton, R.J., Sinz, F.H., Yatsenko, D., Saggau, P.,

- Bethge, M., and Tolias, A.S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* *17*, 851–857.
- Froudarakis, E., Cohen, U., Diamantaki, M., and Walker, E.Y. (2020). Object manifold geometry across the mouse cortical visual hierarchy. *bioRxiv* 10.1101/2020.08.20.258798.
- Geisler, W.S., Najemnik, J., and Ing, A.D. (2009). Optimal stimulus encoders for natural tasks. *J. Vis.* *9*, 17.1–16.
- Gjorgjieva, J., Sompolinsky, H., and Meister, M. (2014). Benefits of pathway splitting in sensory coding. *J. Neurosci.* *34*, 12127–12144.
- Govardovskii, V.I., Fyhrquist, N., Reuter, T., Kuzmin, D.G., and Donner, K. (2000). In search of the visual pigment template. *Vis. Neurosci.* *17*, 509–528.
- Graham, D.J., Chandler, D.M., and Field, D.J. (2006). Can the theory of “whitening” explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Res.* *46*, 2901–2913.
- Hasson, U., Nastase, S.A., and Goldstein, A. (2020). Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron* *105*, 416–434.
- Henriksson, J.T., Bergmanson, J.P.G., and Walsh, J.E. (2010). Ultraviolet radiation transmittance of the mouse eye and its individual media components. *Exp. Eye Res.* *90*, 382–387.
- Hinton, G.E., and Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks. *Science* *313*, 504–507.
- Hoy, J.L., Yavorska, I., Wehr, M., and Niell, C.M. (2016). Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. *Curr. Biol.* *26*, 3046–3052.
- Hoy, J.L., Bishop, H.I., and Niell, C.M. (2019). Defined Cell Types in Superior Colliculus Make Distinct Contributions to Prey Capture Behavior in the Mouse. *Curr. Biol.* *29*, 4130–4138.e5.
- Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat’s striate cortex. *J. Physiol.* *148*, 574–591.
- Huberman, A.D., and Niell, C.M. (2011). What can mice tell us about how vision works? *Trends Neurosci.* *34*, 464–473.
- Hut, R.A., Scheper, A., and Daan, S. (2000). Can the circadian system of a diurnal and a nocturnal rodent entrain to ultraviolet light? *J. Comp. Physiol. A* *186*, 707–715.
- Hut, R.A., Pilorz, V., Boerema, A.S., Strijkstra, A.M., and Daan, S. (2011). Working for food shifts nocturnal mouse activity into the day. *PLoS One* *6*, e17527.
- Jacobs, G.H., Williams, G.A., and Fenwick, J.A. (2004). Influence of cone pigment coexpression on spectral sensitivity and color vision in the mouse. *Vision Res.* *44*, 1615–1622.
- Jacobs, G.H., Williams, G.A., Cahill, H., and Nathans, J. (2007). Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science* *315*, 1723–1725.
- Joesch, M., and Meister, M. (2016). A neuronal circuit for colour vision based on rod-cone

opponency. *Nature* 532, 236–239.

Johnsen, S., Kelber, A., Warrant, E., Sweeney, A.M., Widder, E.A., Lee, R.L., Jr, and Hernández-Andrés, J. (2006). Crepuscular and nocturnal illumination and its effects on color perception by the nocturnal hawkmoth *Deilephila elpenor*. *J. Exp. Biol.* 209, 789–800.

Kingma, D.P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv:1412.6980

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., Maclver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* 93, 480–490.

Kramer, M.A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* 37, 233–243.

Kremkow, J., Jin, J., Wang, Y., and Alonso, J.M. (2016). Principles underlying sensory map topography in primary visual cortex. *Nature* 533, 52–57.

Land, M. (2019). Eye movements in man and other animals. *Vision Res.* 162, 1–7.

Land, M.F. (2015). Eye movements of vertebrates and their relation to eye form and function. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 201, 195–214.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 91–110.

Mante, V., Frazor, R.A., Bonin, V., Geisler, W.S., and Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nat. Neurosci.* 8, 1690–1697.

Marr, D., and Hildreth, E. (1980). Theory of edge detection. *Proc. R. Soc. Lond. B Biol. Sci.* 207, 187–217.

Masland, R.H., and Martin, P.R. (2007). The unsolved mystery of vision. *Curr. Biol.* 17, R577–R582.

Mazade, R., Jin, J., Pons, C., and Alonso, J.-M. (2019). Functional Specialization of ON and OFF Cortical Pathways for Global-Slow and Local-Fast Vision. *Cell Rep.* 27, 2881–2894.e5.

Meyer, A.F., Poort, J., O’Keefe, J., Sahani, M., and Linden, J.F. (2018). A Head-Mounted Camera System Integrates Detailed Behavioral Monitoring with Multichannel Electrophysiology in Freely Moving Mice. *Neuron* 100, 46–60.e7.

Meyer, A.F., O’Keefe, J., and Poort, J. (2020). Two Distinct Types of Eye-Head Coupling in Freely Moving Mice. *Curr. Biol.* 30:11, R635–R637.

Michaieł, A.M., Abe, E.T.T., and Niell, C.M. (2020). Dynamics of gaze control during prey capture in freely moving mice. *eLife* 2020;9:e57458.

Naarendorp, F., Esdaille, T.M., Banden, S.M., Andrews-Labenski, J., Gross, O.P., and Pugh, E.N., Jr (2010). Dark light, rod saturation, and the absolute and incremental sensitivity of mouse

cone vision. *J. Neurosci.* *30*, 12495–12507.

Nadal-Nicolás, F.M., Kunze, V.P., Ball, J.M., Peng, B.T., Krishnan, A., Zhou, G., Dong, L., and Li, W. (2020). True S-cones are concentrated in the ventral mouse retina and wired for color detection in the upper visual field. *Elife* *9*.

Ocko, S., Lindsey, J., Ganguli, S., and Deny, S. (2018). The emergence of multiple retinal cell types through efficient coding of natural movies. In *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds. (Curran Associates, Inc.), pp. 9389–9400.

Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* *381*, 607–609.

Oommen, B.S., and Stahl, J.S. (2008). Eye orientation during static tilts and its relationship to spontaneous head pitch in the laboratory mouse. *Brain Res.* *1193*, 57–66.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox, and R. Garnett, eds. (Curran Associates, Inc.), pp. 8026–8037.

Peichl, L. (2005). Diversity of mammalian photoreceptor properties: adaptations to habitat and lifestyle? *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* *287*, 1001–1012.

Perge, J.A., Koch, K., Miller, R., Sterling, P., and Balasubramanian, V. (2009). How the optic nerve allocates space, energy capacity, and information. *J. Neurosci.* *29*, 7917–7928.

Poynton, C. (2003). *Digital Video and HD: Algorithms and Interfaces* (Elsevier).

Ratliff, C.P., Borghuis, B.G., Kao, Y.-H., Sterling, P., and Balasubramanian, V. (2010). Retina is structured to process an excess of darkness in natural scenes. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 17368–17373.

Rhim, I., Coello-Reyes, G., and Nauhaus, I. (2020). Variations in photoreceptor throughput to mouse visual cortex and the unique effects on tuning. *bioRxiv* 10.1101/2020.11.03.366682.

Rieke, F., and Rudd, M.E. (2009). The challenges natural images pose for visual adaptation. *Neuron* *64*, 605–616.

Rodieke, R.W. (1998). *The first steps in seeing* (Sinauer, Sunderland, MA).

Röhlich, P., van Veen, T., and Szél, Á. (1994). Two different visual pigments in one retinal cone cell. *Neuron* *13*, 1159–1166.

van Rossum, M.C.W., O'Brien, B.J., and Smith, R.G. (2003). Effects of noise on the spike timing precision of retinal ganglion cells. *J. Neurophysiol.* *89*, 2406–2419.

Ruderman, D.L., and Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.* *73*, 814–817.

Rust, N.C., and Movshon, J.A. (2005). In praise of artifice. *Nat. Neurosci.* *8*, 1647–1650.

Schröder, S., Steinmetz, N.A., Krumin, M., Pachitariu, M., Rizzi, M., Lagnado, L., Harris, K.D., and Carandini, M. (2020). Arousal Modulates Retinal Output. *Neuron* 107, 487–495.e9.

Seabrook, T.A., Burbridge, T.J., Crair, M.C., and Huberman, A.D. (2017). Architecture, Function, and Assembly of the Mouse Visual System. *Annu. Rev. Neurosci.* 40, 499–538.

Shang, C., Liu, A., Li, D., Xie, Z., Chen, Z., Huang, M., Li, Y., Wang, Y., Shen, W.L., and Cao, P. (2019). A subcortical excitatory circuit for sensory-triggered predatory hunting in mice. *Nat. Neurosci.* 22, 909–920.

Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216.

Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*.

Stabio, M.E., Sondereker, K.B., Haghgou, S.D., Day, B.L., Chidsey, B., Sabbah, S., and Renna, J.M. (2018). A novel map of the mouse eye for orienting retinal topography in anatomical space. *J. Comp. Neurol.* 526, 1749–1759.

Sterratt, D.C., Lyngholm, D., Willshaw, D.J., and Thompson, I.D. (2013). Standard anatomical and visual space for the mouse retina: computational reconstruction and transformation of flattened retinae with the Retistruct package. *PLoS Comput. Biol.* 9, e1002921.

Szatko, K.P., Korympidou, M.M., Ran, Y., Berens, P., Dalkara, D., Schubert, T., Euler, T., and Franke, K. (2020). Neural circuits in the mouse retina support color vision in the upper visual field. *Nat. Commun.* 11, 3481.

Szél, A., Röhlich, P., Caffé, A.R., Juliusson, B., Aguirre, G., and Van Veen, T. (1992). Unique topographic separation of two spectral classes of cones in the mouse retina. *J. Comp. Neurol.* 325, 327–342.

Szel, A., Csorba, G., Caffé, A.R., Szél, G., Röhlich, P., and Van Veen, T. (1994). Different patterns of retinal cone topography in two genera of rodents, *Mus* and *Apodemus*. *Cell Tissue Res.* 276, 143–150.

Tedore, C., and Nilsson, D.-E. (2019). Avian UV vision enhances leaf surface contrasts in forest environments. *Nat. Commun.* 10, 238.

Turner, M.H., Sanchez Giraldo, L.G., Schwartz, O., and Rieke, F. (2019). Stimulus- and goal-oriented frameworks for understanding natural vision. *Nat. Neurosci.* 22, 15–24.

Van Essen, D., Olshausen, B.A., Anderson, C.H., and Gallant, J.T.L. (1991). Pattern recognition, attention, and information bottlenecks in the primate visual system. In *Visual Information Processing: From Neurons to Chips*, (International Society for Optics and Photonics), pp. 17–28.

Vincent, B.T., and Baddeley, R.J. (2003). Synaptic energy efficiency in retinal processing. *Vision Res.* 43, 1283–1290.

Wallace, D.J., Greenberg, D.S., Sawinski, J., Rulla, S., Notaro, G., and Kerr, J.N.D. (2013). Rats maintain an overhead binocular field at the expense of constant fusion. *Nature* 498, 65–69.

Wang, Y., Jin, J., Kremkow, J., Lashgari, R., Komban, S.J., and Alonso, J.M. (2015). Columnar organization of spatial phase in visual cortex. *Nat. Neurosci.* *18*, 97–103.

Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* *13*, 600–612.

Wilkes, T.C., McGonigle, A.J.S., Pering, T.D., Taggart, A.J., White, B.S., Bryant, R.G., and Willmott, J.R. (2016). Ultraviolet Imaging with Low Cost Smartphone Sensors: Development and Application of a Raspberry Pi-Based UV Camera. *Sensors* *16*.

Xing, D., Yeh, C.-I., and Shapley, R.M. (2010). Generation of black-dominant responses in V1 cortex. *J. Neurosci.* *30*, 13504–13512.

Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* *19*, 356–365.

Yeh, C.-I., Xing, D., and Shapley, R.M. (2009). “Black” Responses Dominate Macaque Primary Visual Cortex V1. *J. Neurosci.* *29*, 11753–11760.

Zhang, Y., -J. Kim, I., Sanes, J.R., and Meister, M. (2012). The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proceedings of the National Academy of Sciences* *109*, E2391–E2398.

Zimmermann, M.J.Y., Nevala, N.E., Yoshimatsu, T., Osorio, D., Nilsson, D.-E., Berens, P., and Baden, T. (2018). Zebrafish Differentially Process Color across Visual Space to Match Natural Scenes. *Curr. Biol.* *28*, 2018–2032.e5.

Zoccolan, D., Oertelt, N., DiCarlo, J.J., and Cox, D.D. (2009). A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences* *106*, 8748–8753.