# Color representation in CNNs: parallelisms with biological vision

Ivet Rafegas & Maria Vanrell
Computer Vision Center
Universitat Autònoma de Barcelona
Bellaterra, Barcelona (Spain)
{ivet.rafegas, maria.vanrell}@uab.cat

## Abstract

*Convolutional Neural Networks (CNNs) trained for object recognition tasks present representational capabilities approaching to primate visual systems [1]. This provides a computational framework to explore how image features are efficiently represented. Here, we dissect a trained CNN [2] to study how color is represented. We use a classical methodology used in physiology that is measuring index of selectivity of individual neurons to specific features. We use ImageNet Dataset [20] images and synthetic versions of them to quantify color tuning properties of artificial neurons to provide a classification of the network population. We conclude three main levels of color representation showing some parallelisms with biological visual systems: (a) a decomposition in a circular hue space to represent single color regions with a wider hue sampling beyond the first layer (V2), (b) the emergence of opponent low-dimensional spaces in early stages to represent color edges (V1); and (c) a strong entanglement between color and shape patterns representing object-parts (e.g. wheel of a car), object-shapes (e.g. faces) or object-surrounds configurations (e.g. blue sky surrounding an object) in deeper layers (V4 or IT).*

## 1. Introduction

Deep learning techniques and, more specifically, Convolutional Neural Networks (CNNs) have become central in the computer vision field due to their impressive performance on solving diversity of vision tasks, such as object recognition or object detection. They consist in several stacked layers that successively make visual representational transformations, as a result of an automatic training that learns the best weights (neurons) to represent their input and achieve the best performance on a specific visual task. These artificial architectures have also been proposed as a suitable framework to model biological vision [10, 11, 1], and some trained CNNs have shown powerful representational capabilities that rival the primate visual system on the performance achieved on a visual recognition task [1].

Although they can present an undesirable black-box nature, several works have proposed different techniques in order to understand the intermediate CNNs representation and visualize learned features [29, 14, 25, 24, 28, 16] providing a visualization of how images are represented in each level of representation and giving some insights of representational inspiration.

In this work we use an artificial neural network with a high-level of representational capability to understand how color can be efficiently represented in recognition tasks. We work on a CNN trained by [2] whose architecture has been proved as to rival the primate visual system performance [1]. We analyze how trained neurons of this network codify color through the different convolutional layers and we relate the conclusions to known evidences about how color is represented in biological visual systems. To perform the analysis we use a selectivity index that gets inspiration from physiological methodologies.

The network was trained for image classification based on labeled content and without any constraint on the filter weights. From this training three different levels of color representation emerge: (a) homogeneous color regions are encoded on a set of neurons selective to one single hue; (b) color edges are represented with color-opponent neurons acting as low-dimensional spaces; and (c) most color selective neurons entangle color and shape together in a pattern matching approach. Remarkably, these representation levels agree with some evidences on how color is encoded through the visual pathway in biological systems: like the existence of a small number of primary colors and a three-dimensional color opponent space in early stages of the visual system [12, 13, 7, 3], a more dense sampling of different hues in V2 (hue maps) [27, 26, 8], and a clear dependence between shape and color in higher level neurons [5, 21].

## 2. Color neurons

In order to understand how color is represented in CNNs we estimate a color selectivity index for all the network neurons. Color neurons are those holding the property of being highly activated when a particular color appears in the input image and, on the contrary, they present a low activation when this color is not present [22, 6, 19, 18]. One way to measure this selectivity index is by computing the variation of the neuron activation between color images and their gray-level versions [18]. These gray-level images are channel-wise replicated to fit into the constraints of the network. Following this idea, color selectivity property is measured on the N-top image patches of the training set that activate a specific neuron the most. The *Area Under the Activation Curve*[1] (AUAC) of a neuron is obtained from the top activations, both for color images and for their gray-level versions and are combined to compute the color selectivity index in this way:

$$\alpha = 1 - \frac{\sum_{j=1}^{N} w'_j}{\sum_{j=1}^{N} w_j} \quad (1)$$

where $\{w_j\}_{j=1:N}$ are the neuron activation values to the original N-top ranked image patches, and $\{w'_j\}_{j=1:N}$ are the activation values obtained by the same neuron to the gray-level versions of the N-top images. Small $\alpha$ value indicates the neuron is low color selective while a large $\alpha$ indicates high color selectivity.

Above index allows to rank the set of neurons in terms of how their activation is related to a specific color or not. Using this index, we propose a classification of the network neurons based on: (a) degree of selectivity (non color selective, low color selective or high color selective); (b) number of colors a neuron is selective to (single or double); and (c) the opponency property of double color neurons (opponent or non opponent), that is related to the hue angular distance between the color pair (opponent is $180°$). In next lines we outline the details of the classification we used in the experiments.

First, neurons are classified depending on their color selectivity index $\alpha$, being *non color selective* neurons those with $\alpha < th_1$, *high color selective* neurons when $\alpha > th_2$ ($th_2 > th_1$) and, *low color selective* when $\alpha$ is in between.

Second, high color selective neurons are classified as *single* neurons if they are selective only to one specific color or *double* neurons if they are selective to a pair of colors. This step is performed using the Neuron Feature (NF) image [19], which is an approximation of the intrinsic spatio-chromatic feature that activates the neuron. It is obtained by computing a weighted average of the $N$-top image patches that activate the neuron (here, $N = 100$), where weights are proportional to the activation value produced by each image to this neuron. NF color distribution is fitted with a Gaussian mixture model using Expectation-Maximization (EM) algorithm. Color distribution is computed on the hue dimension of an opponent color space[2]. Different numbers of Gaussians are used for the fitting, and it is finally set to the minimum number that differs less than a 10% over the global mean square error with the distribution. This step allows to get one (a pair of) representative hue (hues) for each single (double) color neuron[3].

Finally, double color neurons are classified in *opponent* neurons if the pair of colors which a neuron is selective has an angular distance close to $180°$. These angles are considered with respect to the center of the $O_2$-$O_3$ chromatic plane. To study the set of axes emerged from each layer we group the double neurons depending on their pair of colors. For this clustering, we make use of the mean of the Gaussians that fit the hue distribution for each neuron. Therefore, each double neuron can be represented in a 2 dimensional hue space. Here we propose to use a K-means clustering for grouping neurons that share their pair of colors. We test from $K = 3$ to 7 to detect the best number of clusters using Elbow method, which is based on a ratio of the between-group variance with respect to the total variance. Each cluster will be used to infer the main axes emerged from each layer. To refer easily to each group, we manually labeled each cluster by their colors. All clusters emerged from the detected double color selective neurons can be seen in table 2, where each row joins neurons belonging to the same cluster.

This classification provides a general map of color neurons through layers giving insight on how color is represented in the CNN. In the next section, we analyze different properties of the color neurons we found in the trained CNN.

## 3. Results and discussion

In this section we summarize the results of exploring how color information is represented across the layers of the CNN VGG-M trained by Chatfield *et al.* [2] on the ImageNet dataset [20] for an object categorization task. We selected this network due to its similarity with the one reported proved to have representational performance that competes with primate capabilities [1]. This network consists of five convolutional layers (with squared filters of sizes 7, 5, 3, 3 and 3, respectively) followed by three fully connected layers. To reduce the representation size and to introduce some scale invariance, it also incorporates 3 max-

[1] We refer to *Activation Curve* as the set of activation values provided by a neuron to a series of different input stimuli varying on a specific property.

[2] The opponent color space (OPP) used in this work is the one proposed by Platanoits *et al.* in [17] but normalizing and shifting the three axes within the range $[-1, 1]$

[3] Selectivity to three or four colors was never found in our experiments but may appear in other networks

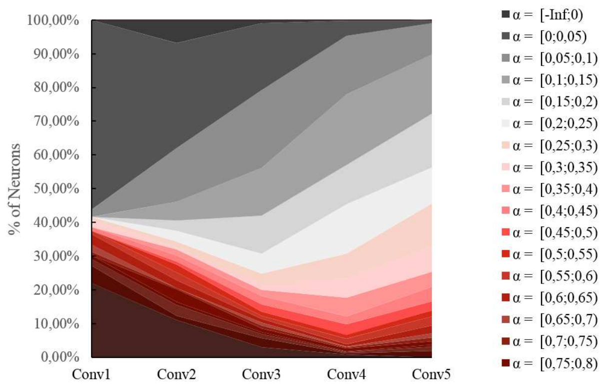



Figure 1: Color selectivity index distribution along the 5 convolutional layers.

pooling layers after first, second and fifth convolutional layers. This network expects input images of 224x224 pixels in RGB, where each convolutional layer analyzes receptive fields of sizes 7x7x3, 27x27x3, 75x75x3, 107x107x3 and 139x139x3, from first to fifth convolutional layer.

The version of the ImageNet dataset used by Chatfield *et al.* that we do also use in this work is the ILSRVC12[4], which consists of around 1.2M of images classified in 1000 categories, according to the lexical WordNet hierarchy [15].

The distribution of the color selectivity indexes for all neurons in convolutional layers of the network can be seen in Fig. 1. We set $th_1 = 0.1$, $th_2 = 0.25$ to classify each neuron as non, low or high color selective. From our experience, different variations of these thresholds would bring to similar conclusions. Shallower layers have neurons with extreme color index values (either very high or very low) while deeper neurons are mainly described by intermediate index values. From this figure it can be concluded that, although higher $\alpha$ values decrease through layers, color remains as an important feature in deeper layers where non-color selective neurons are almost disappeared. In table 1 we summarize the global results of classifying all the neurons of the convolutional layers of the network. We give the number of neurons in each class and the corresponding percentage in the layer.

A visualization of these classified color neurons through layers is shown in table 2. For each group we show a subset of NFs belonging to the group. By observing NFs we can observe that complexity increases through layers. First, simple features such as homogeneous regions, oriented color edges and different frequency and orientation selective gray-level edges are found in Conv1. Second, slightly more complex features such as blobs, curved edges and textured regions appear in Conv2. Finally, we group Conv3, Conv4 and Conv5 in a third level presenting a similar shape complexity: object-surround configurations, complex objects like humans, dog faces or buildings. Main differences between these three layers rely on object sizes, we found similar shapes activating different neurons at different layers thus at different image resolutions.

[4]Images can be browsed on http://image-net.org/challenges/LSVRC/2012/browse-synsets, as well the concepts (*synsets*) that describe each image.

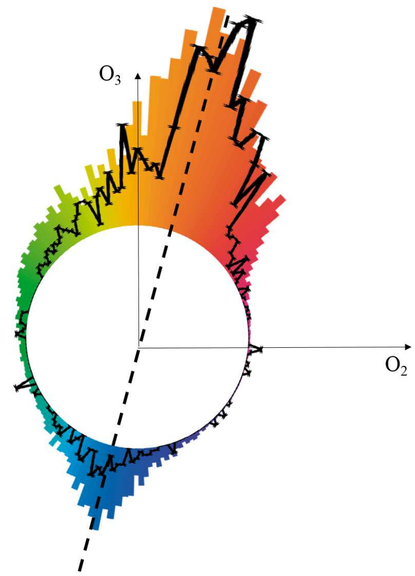



Figure 2: Distribution of color selective neurons on a hue dimension (black line). ImageNet color distribution on hue dimension (colored bars). Dashed line indicates the dataset bias on a Orangish-Bluish axis.

We grouped on a common histogram all color selective neurons ($\alpha > 0.25$) of all the network according to the colors they are selective to. For each neuron we counted on the hue axis (single neuron color is counted twice while double color neurons are counted 1 for each color), the histogram is represented by the black line in Fig. 2, overlapped on the color distribution of the ImageNet dataset, that is shown with colored bars. A clear bias on Orangish and Bluish regions can be observed, both in the dataset and the color neuron distributions. Thus, both distributions are highly correlated, with a Pearson correlation coefficient ($r$) of 0.89.

In the following subsections we will analyze more in detail each level of color representation that can be inferred and we hypothesize about the parallelisms of these three levels with the known evidences human visual system.

### 3.1. First level: single color regions

A first level to study is how regions of a single color are represented by the network. We observe that they are based on a composition of basic primary colors represented by single color selective neurons. These neurons are mainly

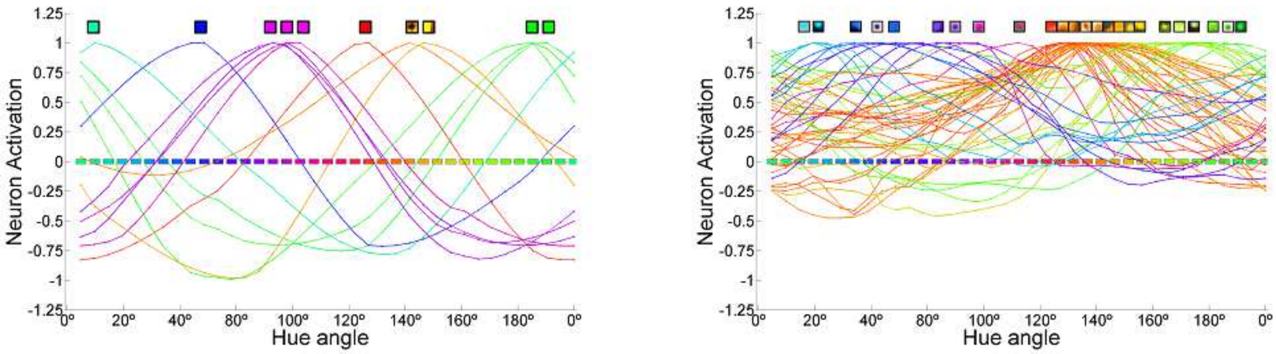
| Selectivity<br>#Neurons | Conv1<br>96 | Conv2<br>219[5] | Conv3<br>512 | Conv4<br>512 | Conv5<br>512 |
|---|---|---|---|---|---|
| Non Color | 56 (58.33%) | 118 (53.88%) | 225 (43.95%) | 113 (22.07%) | 52 (10.16%) |
| Low Color Sel | 2 (2.08%) | 28 (12.79%) | 69 (33.01%) | 255 (49.80%) | 250 (48.83%) |
| Color Sel | 38 (39.58%) | 73 (33.33 %) | 118 (23.05%) | 144 (28.13%) | 210 (41.02%) |
| Single Color | 12 (12.50%) | 49 (22.37%) | 102 (19.92%) | 134 (26.16%) | 198 (38.67%) |
| Double Color | 26 (27.08%) | 24 (10.96%) | 16 (3.13%) | 10 (1.95%) | 12 (2.34%) |
| Opponent | 19 (19.79%) | 14 (6.39%) | 8 (1.56%) | 1 (0.20%) | 1 (0.20%) |
| Non opponent | 7 (7.29%) | 10 (4.57%) | 8 (1.56%) | 9 (1.76%) | 11 (2.15%) |

Table 1: Number of neurons for each neuron class (% of neurons of the group within the layer).


Table 2: Examples of NFs for each convolutional layer and classified in different groups regarding color properties.

found in Conv1 and Conv2. In the first convolutional layer single color neurons are specialized on the following basic hues: Red, Green, Blue, Magenta, Cyan and Yellow, which is a quite standard and balanced sampling of the hue circle, whose combination allows to represent all the color hues. In Fig. 3.(a) we can see the activation curves of these basic neurons to homogeneous regions along the hue dimension. Some of them present more than a neuron for a similar primary hue.

Color representation is more detailed on the second convolutional layer, where a more dense sampling of primary hues is covered by the set of single neurons. Thus the color of a region can be more precisely described in this more extensive basis. In Fig. 3.(b) we can observe neuron activation curves of single neurons in the second layer. Activation curves were computed by rotated hue versions of the first top activating image. The density of the hue sampling performed by the single color neurons is measured. In table 3 we show the results of computing a sparsity measure, $l^0 = \{j : C_j = 0\}$ measure, studied in [9]. From these measurements we can clearly see that Conv2 present a less sparse distribution of sampled colors.



(a) Conv1


(b) Conv2

Figure 3: Activation curves of single neurons in (a) Conv1 and (b) Conv2 over series of images varying along the hue dimension. Stimuli are rotated hue versions of the maximum activating image for each neuron.

In Fig. 4 we show how color single color neurons are distributed along the hue dimension. We can see that in shallower layers a more equitable distribution across the hue dimension is presented. However, in deeper layers the majority of single neurons are concentrated on the Orangish and Bluish regions, aligned with the color bias of the training dataset.

Here we want to highlight a clear parallelism of this level of representation with what is known in the primate visual system. A more dense sampling of hue has been reported in several studies [27, 26, 8] with the existence of hue maps in V2 cortical areas, where a continuous color tuning along the hue is found, providing a more precise color perception. The role of V2 as an intermediate layer in between the low dimensional color space of primary colors found in V1 and the entanglement of color and shape in posterior areas is nicely revised in [4]. We can observe in Fig. 4 how this role of intermediate layer is presented by Conv2 in the artificial network, which appears in between a low dimensional representation of color in Conv1 with six primaries and the emergence of an oversampled region on the orangish hue linked to the color of the majority of objects, that starts in Conv2 and that is clearly the essence color representation in posterior layers.

### 3.2. Second level: color edges

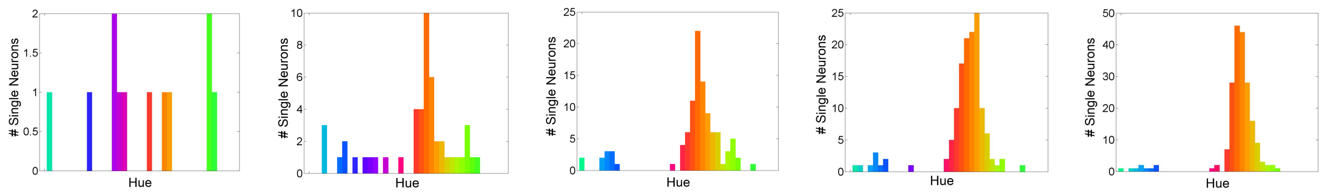
In this section we explore a second level of representation that is about how color edges are represented by the network. Color edges are low level features mainly represented by double color neurons in layers Conv1 and Conv2.

First convolutional layer presents a key role in this representation due to the strong opponency shown by all clusters of double color neurons. They are representing a three dimensional opponent space based on three chromatic channels Red-Cyan, Blue-Yellow and Magenta-Green axes, where each pair of colors has an angular distance (opponency property) of $166.74°$, $168.31°$ and $160°$, respectively. In Fig. 5 we show activation curves of three double color neurons in Conv1 oriented in 45° (one for each color axis) to different synthetic color edges in the same orientation. We compute the activation on 4 different type of edges composed by pairs of colors holding different angular distances between them: 45°, 90°, 135° and 180°, each type of edge is generated along a sampling of the entire hue. In this figure, for each type of edges we plot the color edge achieving the maximum activation for each neuron. For the case of opponent edges (180°, maximum color pair contrast difference), the maximal neuron activations are achieved when the edge coincides with the NF pattern (see the legend for a visualization of NFs of the studied neurons). The rest of color edges are represented by different triplets of weights on the neuron basis. In this way we show how any color-edge is represented by the three main group of opponent axes emerging in this layer Conv1.

The emergence of the opponent space defined by 4 opponent axes (adding Black-White) shows a strong correlation with early stages of the primate visual systems. On one hand, the existence of opponent neurons in the axes of Black-White, Red-Green (cyan in our case) and Blue-Yellow was initially proved by the findings of Derrington

| *Bin-size* | 5° | 10° | 15° | 20° | 25° | 30° | 35° | 40° |
|---|---|---|---|---|---|---|---|---|
| Conv1 | 64 | 30 | 18 | 12 | 8 | 7 | 5 | 4 |
| Conv2 | **46** | **17** | **8** | **4** | **2** | **2** | **2** | **1** |
| Conv3 | 49 | 21 | 14 | 10 | 7 | 5 | 4 | 4 |
| Conv4 | 51 | 23 | 14 | 11 | 9 | 5 | 4 | 4 |
| Conv5 | 50 | 22 | 14 | 10 | 8 | 6 | 5 | 5 |

Table 3: Hue sparsity representation of high single color selective neurons set for each layer ($l^0$ measure) and for different bin sizes. Minimums (in bold) are found in Conv2 for all tested sizes.




(a) Conv1 (b) Conv2 (c) Conv3 (d) Conv4 (e) Conv5

Figure 4: Number of single color neurons per hue for the five convolutional layers. Shallower layers (conv1 and conv2) present a more equitable distribution along the hue dimension. Deeper layers (conv3, conv4 and conv5) concentrate single neurons on Orangish and Bluish regions following the color bias of the training dataset.

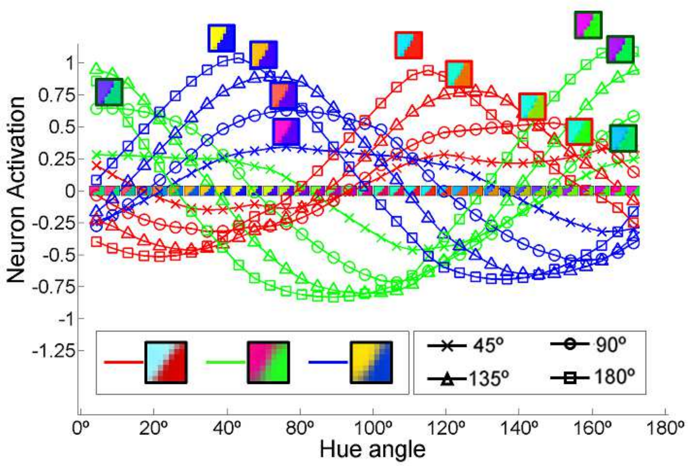
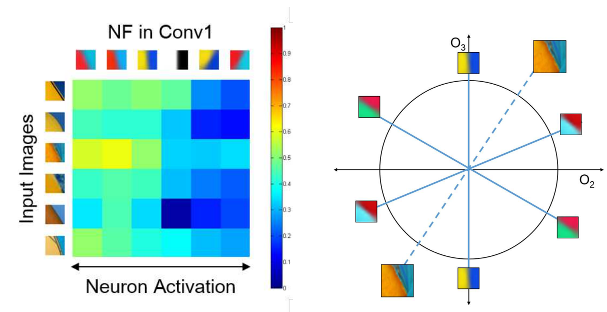



Figure 5: Representation of color edges. Activation curves of three different double opponent color neurons at Conv1 to 4 different types of edges regarding the opponency property of the pair of colors on the tested edges: opponent edges (180), 135, 90 and 45 color edges.




(a) (b)

Figure 6: Orange-blue edge representation in Conv1. (a) Activation values of some neurons in Conv1 (NFs are in first row) to a set of images (first column) in a color scale from blue (minimum) to red (maximum). (b) Double neurons in Conv1 over the hue space. Orange-Blue image (dashed line) is represented between two main neurons in Conv1 (Blue-Yelow and Red-Cyan).

*et al.* in [7], and later by Lennie *et al.* in [12, 13]. On the other hand, an additional fourth opponent axis in the direction of Magenta-Green could coincide with posterior studies [3]. We think it is quite remarkable that this artificial architecture shows such a coincidence with the biological vision, which are giving support to hypothesis regarding the efficiency of neural information to represent natural image statistics. Further correlations are given by the fact that these double color neurons present selectivity to oriented and low spatial-frequency edges, meanwhile black-white neurons present selectivity to a wider variety of spatial frequencies of oriented edges as it is revised in [22].

In the rest of the layers, different axes emerge from the double color neurons, but they move away from the opponency property and with shapes more complex than edges. Only with the exception of the Bluish-Orangish axis, which is present in all these layers, holding an angular distance greater than 160° between the two colors. This axis is a rotation of the Blue-Yellow axis found in Conv1 that matches the bias of the training dataset. In Fig. 6.(a) we show how several image patches that highly activate a Bluish-Orangish neuron of Conv2 are represented in Conv1. NFs of the first convolutional layer that are most activated by these kind of Orange-Blue edge stimuli (oriented in 135°) are plotted in the first row. Tested images are shown in the first column. Activation values are shown in a scale from blue (minimum) to red (maximum). In Fig. 6.(b) we show how double color neurons are distributed in the chromaticity plane of the opponent color space, including the representation of a Blue-Orange image. This example brings us to speculate about the hierarchy of the Bluish-Orangish neuron population code in Conv2: it is activated when a Red-Cyan neuron of Conv1 with an edge of 135° is highly activated jointly with a high activation of a Yellow-Blue neuron (also in Conv1) with a 90° oriented edge[6].

[6]Note that in Conv1 a Blue-Yellow neuron having an oriented edge of

### 3.3. Third level: color-shape entanglement

As seen in previous sections, color is an important property that may characterize the activity of a neuron. Nevertheless, this property always appears strongly linked with the intrinsic shape that also activates the neuron. In this section we analyze the entanglement of both properties, shape and color, which can be understood as a template matching scheme linked to the activity of any neuron along the artificial network. With color-shape entanglement we mean that the activation of a single neuron requires the appearance of a specific color in an appropriate configuration or shape.

This color-shape entanglement appears through all the layers. Regarding shallower layers we plot in Fig. 7 a set of neuron activation curves for two color neurons in (a) Conv1 and (b) Conv2, both representing colored edges oriented in 135°. Each neuron activation curve is obtained from a set of oriented edge images, sharing the same orientation but with different opponent color pairs. Different oriented edges are tested for both neurons and maximum activations are achieved when color pair and orientation match with the input image (see the neuron NF visualization on the top left corner).

In the same line, neurons in deeper layers which are responsible to detect more complex shapes also present this color-shape dependence. As we can see in table 2 NFs in layers Conv3, Conv4 and Conv5 present pattern of recognizable objects (like bodies, mushrooms, ladybugs etc.), object parts (like wheels of cars, dog faces, semicircles, etc.) or object surround configurations (like blue object in a green surround, squared object under a blue sky, etc.). For this kind of neurons we show two examples of activation curves where either color or shape were varied. First, in Fig. 8 we show the effect of a neuron in Conv5 with a high color selectivity index and which is ladybug selective. We show the activation curve of the neuron when different colored versions of the image activate the neuron. Ladybug image is transformed by rotating color image pixels on the chromaticity plane. The rotation is performed along the hue dimension. Note that this neuron is uniquely selective to orangish ladybugs. Second, for another neuron we study the activation curve when shape is varied. In Fig. 9 we plot several activation curves of a neuron found in Conv4 with a high color selective index, and very selective to faces of a specific size and in the usual vertical orientation. The curves show the variation of the activation when same face images that highly activate this neuron are spatially rotated. Note that the same image with spatial rotations modifies the activity of the neuron, although sharing same color appearance.

The strong entanglement between color and shape in the human visual system has been proved in several works, revised by Shapley *et al.* [23]. Although deeper visual ares of the human visual system are not known like in previous areas, in V4 and PIT (posterior inferior temporal cortex) color and complex shape selectivity has been found [5]. Moreover, [21] stated that neurons in V4 are selective to large range of colors and white surfaces and also sensitive to surrounds that may participate in the separation between object and background.

135° (with yellow in the bottom-left and blue in the top-right) was not found. In its absence, the vertical edge neuron is the most similar.

## 4. Conclusions

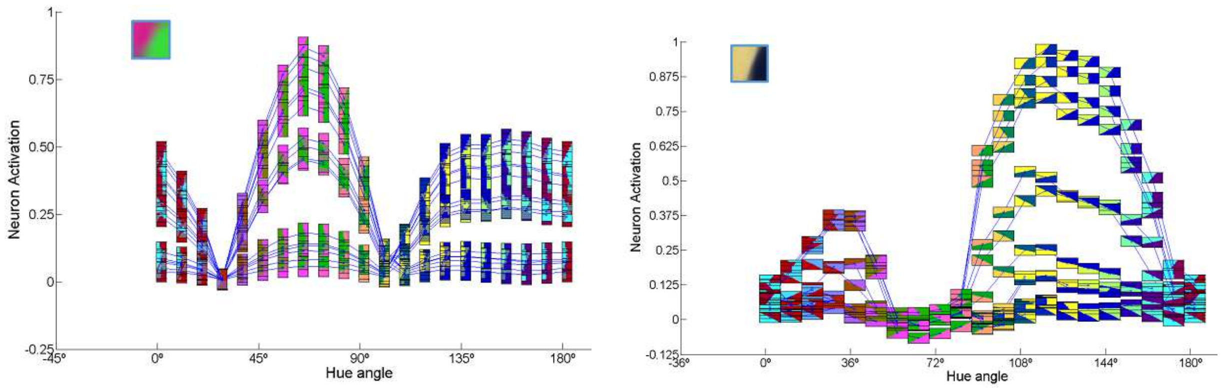
Artificial neural networks have been a revolutionary technique due to their impressive performance in solving different visual tasks. Although they are easily trained on large datasets, the intermediate representation learned by the network architecture are not fully understood. In this work we focus on the analysis of how color is particularly encoded through the different convolutional layers. We find important parallelisms with the human visual system. To this end, we use a methodology inspired on physiological neural research, which is based on the estimation of a color selectivity index for each neuron. We measure several color properties that allow to classify all the network neurons: (a) the color selectivity index for all neurons; (b) the amount of distinctive colors a neuron is selective to, this allows to define single (one color) and double (two colors) neurons; (c) the opponency property between the pair of colors in double neurons; and (d) the sparsity of the sampling of the color neurons over a hue dimension.

We found color selective neurons in all convolutional layers, although shallower layers present higher indexes of color selectivity. We define three main levels of color representation.

First, we study single color neurons representing single color regions, this representation is based on encoding all colors in a primary basis of six basic colors (in Conv1) while a more dense hue sampling appears in Conv2. This way of encoding basic color and the increase in hue sampling correlates with evidences in the primate visual system [27, 26, 8].

Second, we focus on double color neurons devoted to color edge detection. Here the network learns a subset of neurons in Conv1 whose pairs define a 4-dimensional opponent space (Red-Cyan, Blue-Yellow, Magenta-Green and Black-White) in the first convolutional layer presenting a clear correlation with known results in the primary visual cortex. [7, 12, 13, 3].

Finally, we find that color and shape present a strong entanglement at all levels of the network, either for simple features as edges, bars or textures in shallower layers, or for more complex features as faces, bodies or cars in deeper layers. This high level of entanglement has also been suggested in several physiological studies [5, 21].




(a) (b)

Figure 7: Neuron activation of two edge-oriented double color selective neurons, (a) Conv1 ($\alpha = 0.99$) and (b) Conv2 ($\alpha = 0.44$). Activation curve for different synthetic oriented color edges, varying in hue-pairs and orientation. Blue line connects activations corresponding to images with the same orientation but different color-pair. Neuron NFs are framed in blue on the top-left corners.

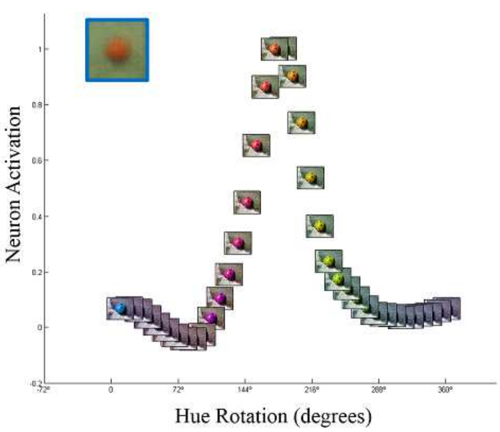



Figure 8: Activation curve of a ladybug-color-selective neuron in Conv5 ($\alpha = 0.94$) for different versions of the same image by rotating hues. Maximum activation corresponds to a perfect matching with the original color of the image patch that maximally activates this neuron. Neuron NF is framed in blue (top-left).

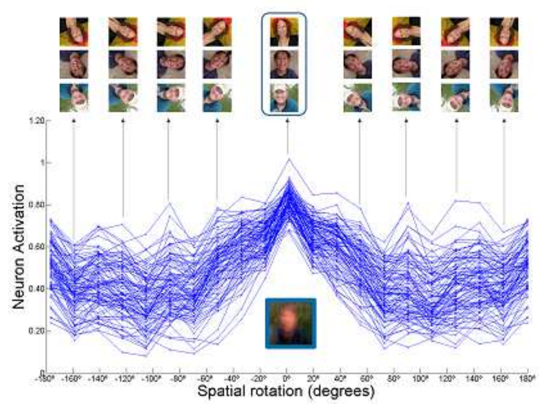



Figure 9: Activation curves of 3 face-color-selective neuron in Conv4 ($\alpha = 0.59$) for different versions of the same image by spatially rotating the set of top image patches. Maximum activations are achieved corresponds to the original orientation of the image patch. Examples of the tested images are shown on the top of the plot, while the NF of the neuron is framed in blue in the bottom.


## 5. Acknowledgements

Project partially funded by MINECO Ref. TIN (TIN2014-61068-R) and Generalitat de Catalunya Ref. SGR-669.


## References


[1] C. F. Cadieu, H. Hong, D. L. K. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, and J. J. DiCarlo. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS computational biology*, 10, 2014 Dec 2014.

[2] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *BMVC*, 2014.

[3] B. R. Conway. Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (v-1). *Journal of Neuroscience*, 21(8):2768–2783, 2001.

[4] B. R. Conway. Colour vision: A clue to hue in v2. *Current*

*Biology*, 13(8):R308 – R310, 2003.

[5] B. R. Conway, S. Chatterjee, G. D. Field, G. D. Horwitz, E. N. Johnson, K. Koida, , and K. Mancuso. Advances in color science: from retina to behavior. *The Journal of Neuroscience*, 30(45):1495514963, 2010.

[6] B. R. Conway and D. Y. Tsao. Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proc Natl Acad Sci U S A.*, 42(106):18034–18039, 2009.

[7] A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.*, pages 241–265, 1984.

[8] L. H., W. Y., X. Y., H. M., and F. DJ. Organization of hue selectivity in macaque v2 thin stripes. *Journal of Neurophysiology*, 102(5).

[9] N. Hurley and S. Rickard. Comparing measures of sparsity. *IEEE Trans. Inf. Theor.*, 55(10):4723–4741, Oct. 2009.

[10] N. Kriegeskorte. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science.*, 1:417–446, 2015.

[11] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. J. Rodriguez-Sanchez, and L. Wiskott. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8), Aug. 2013.

[12] P. Lennie and M. DZmura. Mechanisms of color vision. In *CRC Crit. Rev. Neurobiol.*, chapter 3, pages 333–400. 1988.

[13] P. Lennie, J. Krauskopf, and S. G. Chromatic mechanisms in striate cortex of macaque. *The Journal of Neuroscience*, 10:649–669, 1990.

[14] Y. Li, J. Yosinski, J. Clune, H. Lipson, and J. E. Hopcroft. Convergent learning: Do different neural networks learn the same representations? In *ICLR*, 2016.

[15] G. A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, Nov. 1995.

[16] A. Nguyen, J. Yosinski, and J. Clune. Multifaceted Feature Visualization: Uncovering the Different Types of Features Learned By Each Neuron in Deep Neural Networks. In *Workshop on Visualization for Deep Learning, International Conference on Machine Learning (ICML)*, June 2016.

[17] K. Plataniotis and A. Venetsanopoulos. *Color Image Processing and Applications*. Springer, 2000.

[18] I. Rafegas and M. Vanrell. Color encoding in biologically-inspired convolutional neural networks. Technical report, Computer Vision Center, Universitat Autònoma de Barcelona, 2017.

[19] I. Rafegas, M. Vanrell, and L. A. Alexandre. Understanding trained CNNs by indexing neuron selectivity. *ArXiv e-prints.*, Feb. 2017.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[21] S. J. Schein and R. Desimone. Spectral properties of v4 neurons in the macaque. *VR*, 51(7):701–717, 4 2011.

[22] R. Shapley and M. Hawken. Color in the cortex: Single- and double-opponent cells. *VR*, 51(7):701–717, 4 2011.

[23] R. Shapley and M. Hawken. Color in the cortex: Single- and double-opponent cells. *Vision Research*, 51(7):701–717, 4 2011.

[24] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *In ICLR Workshop 2014*, 2014.

[25] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller. Striving for simplicity: The all convolutional net. *ICLR*, 2015.

[26] M. A. Webster, Y. Mizokami, and S. M. Webster. Hue maps in primate striate cortex. *NeuroImage*, 35(2).

[27] Y. Xiao and Y. W. D. Felleman. A spatially organized representation of colour in macaque cortical area v2. *Nature*, 421.

[28] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson. Understanding neural networks through deep visualization. In *Deep Learning Workshop, (ICML)*, 2015.

[29] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *ECCV*, 2014.