Ecological Informatics 67 (2022) 101486

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Using deep neural networks to model similarity between visual patterns: Application to fish sexual signals

Samuel V. Hulse[a,*], Julien P. Renoult[b], Tamra C. Mendelson[c]

[a] Department of Biology, University of Maryland, College Park, MD, USA
[b] Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, MD, USA
[c] CEFE, Univ Montpellier, CNRS, EPHE, Univ Paul-Valery Montpellier, Montpellier, France

ARTICLE INFO

Keywords:
Camouflage
Convolutional neural networks
Etheostoma
Sensory drive
Sexual selection
Visual patterns

ABSTRACT

The evolution of visual patterns is a frontier in the theory of sexual selection as we seek to understand the function of complex visual patterning in courtship. Recently, the sensory drive and sensory bias models of sexual selection have been applied to higher-level visual processing. One prediction of this application is that animals' sexual signals will mimic the visual statistics of their habitats. An enduring difficulty of testing predictions of visual pattern evolution is in developing quantitative methods for comparing patterns. Advances in artificial neural networks address this challenge by allowing for the direct comparison of images using both simple and complex features. Here, we use VGG19, an industry-leading image classification network to test predictions of sensory drive, by comparing visual patterns in darter fish (*Etheostoma* spp.) to images of their habitats. We find that images of female darters are significantly more similar to images of their habitat than are images of males, supporting a role of camouflage in female patterning. We do not find direct evidence for sensory drive shaping the design of male patterns; however, this work demonstrates the utility of network methods for pattern analysis and suggests future directions for visual pattern research.

1. Introduction

The evolution of visual pattern is an enduring problem in evolutionary biology. Despite the ubiquity and diversity of visual patterns, relatively few empirical studies have examined how sexual selection can influence their design (Endler and Mappes, 2017; Gluckman and Cardoso, 2010; Pérez-Rodríguez et al., 2013; Pérez-Rodríguez et al., 2017; Sibeaux et al., 2019; Tibbetts and Dale, 2004). The classical models of sexual selection, such as Fisherian processes, good genes, and direct benefits, explain why signals might be elaborated, yet offer little in terms of understanding their design (Endler and Basolo, 1998; Ryan, 1998). In the late 20th century, new theoretical developments in sexual selection expanded our ability to predict the direction of signal design (Endler, 1992; Endler and Basolo, 1998; Ryan and Cummings, 2013). “Sensory drive” describes the effect of the environment on the transmission, reception, and evolution of signals (Cummings and Endler, 2018). “Sensory bias” is an example of sensory drive, describing how signals and signal preferences originating in non-sexual contexts, like foraging, can have profound effects on signals used in sexual contexts (Fuller et al., 2005). These models have been successfully applied to understand signal features in many systems, such as the color of African cichlid displays, and the peak frequency of túngara frog calls (Cummings and Endler, 2018; Ryan and Rand, 1990; Seehausen et al., 2008). However, not until recently were these models applied to complex visual patterns.

Recently, Renoult and Mendelson expanded the conceptual framework of sensory drive to encompass higher level visual processing in addition to peripheral sensory transduction (Renoult and Mendelson, 2019). Natural visual scenes contain statistical regularities that sensory systems can leverage (Barlow, 1961; Olshausen and Field, 2004; Simoncelli and Olshausen, 2001). The efficient coding hypothesis predicts that sensory systems will evolve sensory codes that minimize statistical redundancies in the environment, in order to reduce the metabolic cost of information processing (Barlow, 1961). That hypothesis suggests that the optimal sensory code is environment dependent. Thus, if habitats vary in their visual statistics, then animals evolving in different habitats should exhibit differences in sensory coding, which can in turn lead to differences in pattern preferences, and ultimately signal design (Renoult and Mendelson, 2019). Thus, Renoult and Mendelson (2019) predict that visual sexual ornaments should match the

---

* Corresponding author.
E-mail address: shulse@umd.edu (S.V. Hulse).

https://doi.org/10.1016/j.ecoinf.2021.101486
Received 25 June 2021; Received in revised form 9 November 2021; Accepted 9 November 2021

1574-9541/Published by Elsevier B.V.

visual statistics of their habitats. Whereas previously explored mechanisms of sensory drive focus on simple features such as the color of a visual display, or peak frequency of a call, this framework offers a potential explanation for a broader class of signals.

The prediction that the visual statistics of ornaments will match those of their habitat was recently bolstered by Hulse and colleagues, who investigate the Fourier slopes of the preferred habitats of darter fish species and the nuptial displays of the fishes themselves (Hulse et al., 2020). They find a correlation between the slope of habitats and that of nuptial display for males but not for females, suggesting that male darters have evolved visual patterns that mimic the visual statistics of their specific habitat. The Fourier slope is a measure of the relationship between spatial scale and contrast intensity in visual patterns, and represents only one component of visual signaling, however. Testing the predictions of the efficient coding hypothesis will require methods that capture more components of visual signals.

One roadblock on the way to testing predictions of visual pattern evolution is the difficulty of quantifying pattern differences (Pérez-Rodríguez et al., 2017). Image similarity is inherently subjective; to produce a metric of similarity, it is essential to understand how visual scenes are perceived by the brain (Rogowitz et al., 1998; Sinha and Russell, 2011). Information theory and image analysis provide numerous methods, including Fourier analysis, and Fractal Dimension, which can be used to generate biologically relevant summary statistics about images (Pouli et al., 2013). These approaches have been successfully applied to understanding the role of sexual selection role in visual pattern evolution (Menzel et al., 2015; Pérez-Rodríguez et al., 2013; Pérez-Rodríguez et al., 2017), but they do not incorporate higher-level image processing, such as that beyond V1 of the mammalian visual cortex (Simoncelli and Olshausen, 2001). In higher-level visual processing, neurons become tuned to abstract features which are difficult to capture through traditional methods (DiCarlo et al., 2012). For modeling these higher-level processes, artificial neural networks are one of the most promising methodologies currently available (Kriegeskorte, 2015; Wenliang and Seitz, 2018).

The development of deep convolutional neural networks (CNNs) has revolutionized computer vision by mimicking the architecture of vertebrate visual processing (LeCun et al., 2015). CNNs are a feed-forward network, composed of multiple layers of connected units, which have activation levels (the response level of each node in a layer) determined by the units in the preceding layer. These types of networks can be highly invariant to image rotation and skew, long standing difficulties for traditional computer vision approaches (Lo et al., 2018). CNN-based approaches encode images similarly to vertebrate brains, and have been able to mimic the individual neurons in the human inferior temporal cortex (Kuzovkin et al., 2018; Yamins et al., 2014). Furthermore, researchers found that network performance in object classification tasks is strongly correlated with its ability to predict individual cortical neuron activations. In addition to their performance in classifying man-made objects, neural networks have also been successfully used to make species identifications in a many taxa (Hansen et al., 2020; Nguyen et al., 2016; Zhou et al., 2016). CNNs can perform well even when they are not trained on the objects they are classifying, a technique known as transfer learning (Gogul and Kumar, 2017; Shaha and Pawar, 2018). The success of transfer learning shows that the features learned by CNNs are broadly useful across many different image classes. Additionally, the hierarchical architecture of convolutional neural networks means that image representations become increasingly abstract deeper in the network (Kuzovkin et al., 2018; Wenliang and Seitz, 2018). For example, at shallow layers, each unit responds to small-scale, local features and often functions as edge and corner detectors. At deeper layers, units are tuned to more abstract image features that may occur anywhere in the image, such as face-like objects.

Computer scientists have recently devised new methods based on convolutional neural networks to separate the style of an image from its content (Gatys et al., 2016). It may be difficult to objectively define style, but mostly it is defined in terms of texture (Gatys et al., 2017), which, generally speaking, describes properties of an image such as smoothness, regularity and variation in brightness (Bharati et al., 2004). At the heart of texture analysis is the Gram matrix, which is used to represent the style of an image at a given layer. The Gram matrix is a measure of covariance between different features within an image. Differences in texture-like features may be important differentiators for visual patterns between species of animals, thus the Gram matrix may be a useful tool in the quantitative analysis of visual patterns.

CNNs can also be used to generate metrics of image similarity based on features at multiple spatial scales and multiple levels of abstraction (Appalaraju and Chaoji, 2018; Bell and Bala, 2015). By measuring the differences in activations or Gram matrices between two images at different layers in a neural network, we can see how images differ at different levels of abstraction. These approaches are highly effective in providing a quantitative measure of image similarity that closely matches human similarity judgments (Zhang et al., 2018). Despite the clear utility of these networks, CNNs have seen limited use in the study of signal design and pattern evolution. Combining new tools such as CNNs with predictive models of signal design provides new opportunities to investigate the evolution and diversity of animal patterns. For instance, we can test the prediction that the spatial features in animal patterns will be more visually similar to their native habitats than they are to non-native habitats.

Darters (Percidae: *Etheostoma*), a group of North American freshwater fish, make an excellent model for studying how sexual selection affects visual pattern evolution. During their breeding season in the spring, males of different species exhibit an incredible diversity of colorful nuptial patterns. Males use their nuptial patterns for display in both male-male and male-female contexts. Previous studies have shown that both female and male darters have a strong preference for conspecific color and pattern (Williams and Mendelson, 2010; Williams and Mendelson, 2011; Williams and Mendelson, 2013). Furthermore, different species of darters occupy habitat types with distinct visual statistics (Hulse et al., 2020; Stauffer et al., 1996; Welsh and Perry, 1998).

Here, we use VGG19, a benchmark convolutional neural network, and a previously collected library of darter and habitat images to test hypotheses about the evolution of sexually selected visual patterns (Simonyan and Zisserman, 2015). First, we test whether a pre-trained network can be used to accurately classify images of darters and their habitats. We then compare the visual similarity of darter patterns and their habitats to test a hypothesis of sensory drive.

2. Materials and methods

2.1. Image collection and processing

We used a previously obtained collection of photographs, containing images of 11 species of darters as well as underwater photography of darter habitats representing five classes of habitat. The species included were *Etheostoma barrenense*, *E. blennioides*, *E. caeruleum*, *E. camurum*, *E. chlorosomum*, *E. gracile*, *E. olmstedi*, *E. pyrrhogaster*, *E. swaini*, *E. zonale* and *E. zonistium* (for habitat classifications and detailed methods, see Hulse et al., 2020). In total, we analyzed 550 darter images (288 male and 262 female) and 597 habitat images. Although the number of individuals photographed for each species was relatively small for neural network analyses, this is less of a concern for our study as we do not train any networks. The full description of the photography methods can be found in Hulse et al. (2020). All images were resized to 224 × 224 pixels to match the input dimensions of the network and preprocessed so that the mean pixel values matched those of the images used to pretrain the CNN. For the fish images, we used cropped images of the darters' flank, to match methods used in Hulse et al. (2020). Both fish and habitat images were converted to greyscale to negate the influence of color and thus focus all analyses on patterning only. Examples of fish and habitat images are shown in Fig. 1.

2.2. Convolutional neural network analysis

To examine how each species of fish is encoded by deep convolutional networks, and to determine how well species cluster at each layer, we used the VGG19 deep convolutional network pre-trained on the Imagenet database (Fig. 2, Simonyan and Zisserman, 2015). Imagenet contains over 14 million labeled images representing a broad spectrum of objects and is a standard training library in machine learning (Deng et al., 2009). VGG19 consists of 19 layers with five 3 × 3 convolutional filters and five max pooling layers. We omitted the last three fully connected layers as they are related to assigning features to specific object classifications. After each max pooling layer, the channel dimension is reduced by a factor of 4, and the number of filters increased by a factor of 2. Each layer contains between 64 and 512 filters, which activate selectively for specific image features. Each channel is represented by a square matrix of activations. Previous studies have shown that using pre-trained networks such as this is sufficient for many visual recognition tasks, and without massive computational effort is often better than training a network from scratch (Shaha and Pawar, 2018). Our VGG19 implementation was built using the keras library for python, and Google's TensorFlow v2.5.0 platform.

2.3. Classification accuracy

For each species of darter, we designated half of our images as reference images and half as test images. In cases where there were an odd number of images, the training set was given the larger number. For each species, we first inputted the reference images into VGG19 and calculated the mean activation for every neuron at every layer. For a given layer, the vector of mean activations is used as the species centroid representing the prototype of that species. Additionally, at each layer, we computed the Gram matrix and calculated the species centroids from all the images in the reference set. The Gram matrix can be computed from the activations in each layer as $G = F^T F$, where F is the matrix of activations such that each column is a flattened filter (Fig. 3). Each entry of the Gram matrix can be thought of as the covariance between the ith and jth channel, thus the Gram matrix encodes how often image features cooccur.

Next, to generate our classification accuracy metric, we computed the activations for the test image set. An image was accurately classified at a given layer if the Euclidean distance between its activations and its corresponding species centroid was less than the distances between its activations and any other species' centroids. For each species, we calculated the proportion of images accurately classified. The accuracy computation was performed for both activations and Gram matrix-based classification. The overall classification accuracy at each layer was calculated as the mean classification accuracy across all species.

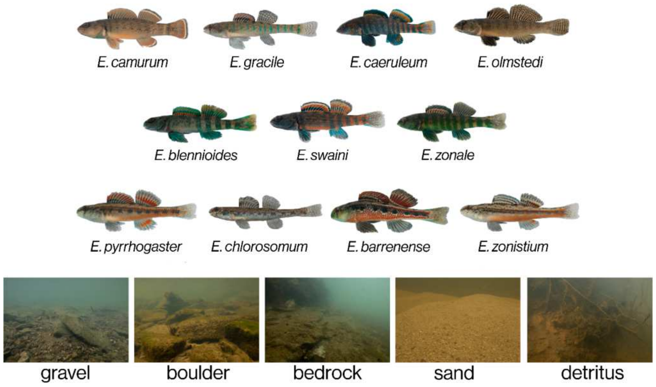

Fig. 1. Sample images of the 11 darter species and 5 habitat classes used in this study.

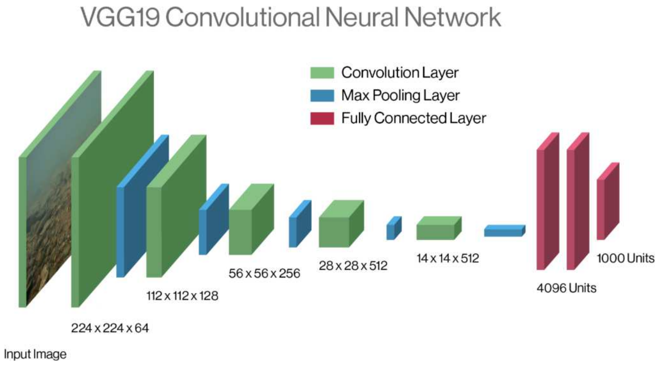

Fig. 2. Schematic of the VGG19 Convolutional Neural Network. At shallow layers, images are represented with a relatively small number of high-resolution filters. At deeper layers, the spatial resolution of the filters decreases while the number of filters increases. This leads to neurons at deeper layers having larger receptive fields than at shallow layers.

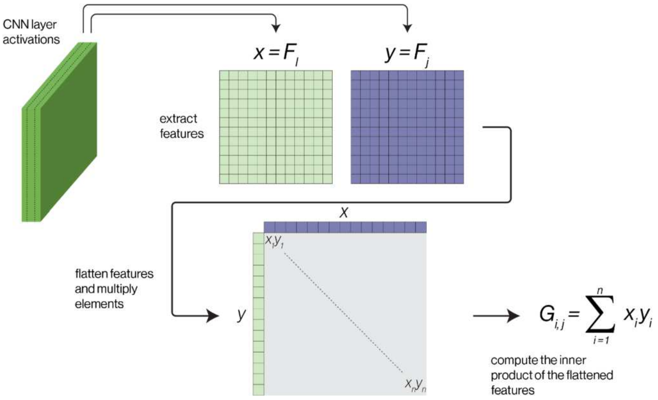

Fig. 3. Diagram depicting the computation of the Gram matrix. At each layer, all of the features (represented by a matrix) are extracted and flattened into a vector. Each entry of the Gram matrix is then calculated as the inner product of the two feature vectors corresponding to the row and column of the Gram matrix.

In addition to evaluating the classification accuracy for fish, we performed the classification analysis for the habitat images, using the same procedure as for the fish images. Habitat images ($n = 597$) were classified into five categories, which were evenly divided into the reference and test sets. Habitat classifications were sand ($n = 110$), gravel ($n = 120$), boulder ($n = 128$), detritus ($n = 123$) and bedrock ($n = 116$), representing typical habitat types for darters. We then calculated the proportion of habitat images in the test set that were accurately classified by Euclidean distance from their class centroid, both for activation and Gram matrix-based classification.

2.4. Species-habitat comparison

To determine whether darter species are more visually similar to their preferred habitat (hereafter, “matched”) than to habitats in which they are not typically found (hereafter, “non-matched”), we analyzed the Euclidean distance between the habitat images and the darter images, for both matched and non-matched habitats. Habitat preferences were determined through field observations as well as a literature search (Bailey and Etnier, 1988; Etnier and Starnes, 1993; Kuehne and Barbour, 2015). For all species, we computed the distance between each individual and their matched habitat centroid, as well as the distances from each individual to each non-matched centroid, treating every layer separately. For each species, we then averaged the matched darter-habitat distances separately from the non-matched distances (a given species has only one matching habitat and several non-matching habitats; for non-matching habitats we averaged all distances). If darter visual patterns match their habitat, we predicted that the distance between darters and matched habitat would be less than the distance between darters and their unmatched habitat. We tested this prediction using a paired *t*-test. Pairwise comparisons were performed separately for males and females, for each layer, considering either distances between activations or between Gram matrices.

Aside from the strength of correlations between matched and non-matched habitats, one possibility is that males could be more visually similar to darter habitats overall than females, (or vice-versa). To test this, we performed a paired t-test, comparing the distance between male images and all habitat classes to the distance between female images and all habitat classes. This test was performed for each species separately, considering either activations or the Gram matrix. We performed this test at the third layer of VGG, where the differences between matched and non-matched distances were greatest.

2.5. *Visualization of gram matrix space*

We also used multi-dimensional scaling to visualize how darter and habitat images associate in the Gram matrix space. This method is similar to other dimensionality reduction methods, such as principal components analysis, but tends to perform better with very high dimensional data, such as the Gram matrix space, which can have up to thousands of dimensions. We performed the multi-dimensional scaling keeping only the first two dimensions, at both layers 3 and 18, in order to visualize the resemblance between the visual styles of habitats and fish as encoded in one shallow and one deep layer.

3. Results

3.1. *Classification accuracy*

Classification accuracy for both males and females increased across layers of the network (Fig. 4). In general, classification accuracy was similar for activation-based classification and Gram matrix-based classification (Fig. 4). For males, across all layers, the average classification accuracy was 60.6% +/− 14.6% based on activations and 55.3% +/− 16.8% based on Gram matrix differences (Fig. 4a). For females, across all layers the average classification accuracy was 52.9% +/ 12.2% for activations and 57.9% +/− 16.9% for the Gram matrix (Fig. 4b). The probability of a correct classification based on random chance is 9.1%.

For the habitat images, we also observed a continuous improvement in classification accuracy towards deeper layers. Across all layers, the average classification accuracy was 59.3% +/− 6.7% for activations, and 55.1% +/− 8.4% for the Gram matrix (Fig. 5). The probability of a correct classification based on random chance is 20%.

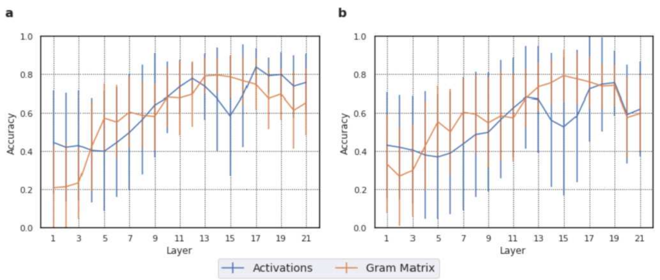

Fig. 4. Proportion of male (a) and female (b) fish images correctly classified by species. Activation-based classifications are in blue and Gram matrix-based classifications are in orange. Solid vertical bars indicate the standard deviation across species. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

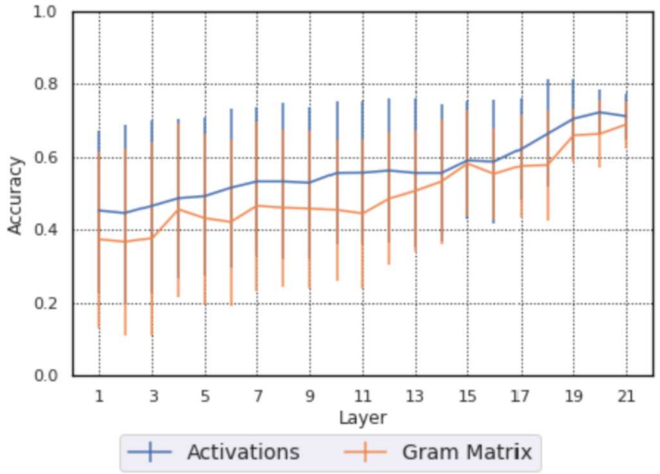

Fig. 5. Proportion of habitat images correctly classified by habitat type. Activation-based classifications are in blue and Gram matrix-based classifications are in orange. Solid vertical bars indicate the standard deviation across habitat classes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. *Species-habitat comparison*

We did not find a significant difference in distances between darters and matched habitats versus non-matched habitats in males, (activations: $t = 0.44$, df = 10, $p = 0.67$; Gram matrices: $t = 0.74$, df = 10, $p = 0.47$; Figs. 6-7) or in females (activations: $t = -0.05$, df = 10, $p = 0.96$; Gram matrices: $t = 0.49$, df = 10, $p = 0.64$; Figs. 6-7).

At shallow layers, we found that female darters were significantly more similar to darter habitats than were males for Gram matrices, but not activations (activations: $t = -0.67$, df = 10, $p = 0.52$; Gram matrices: $t = 3.28$, df = 10, $p = 0.0083$; Fig. 8).

The multi-dimensional scaling generated for layers 3 and 18 provides a visual representation of how fish images and habitat images cluster in Gram matrix space (Fig. 9). We observed separation between fish and habitat images for both layers, but the separation is clearer for layer 18. Additionally, it appears that the primary axis of variation for fish images

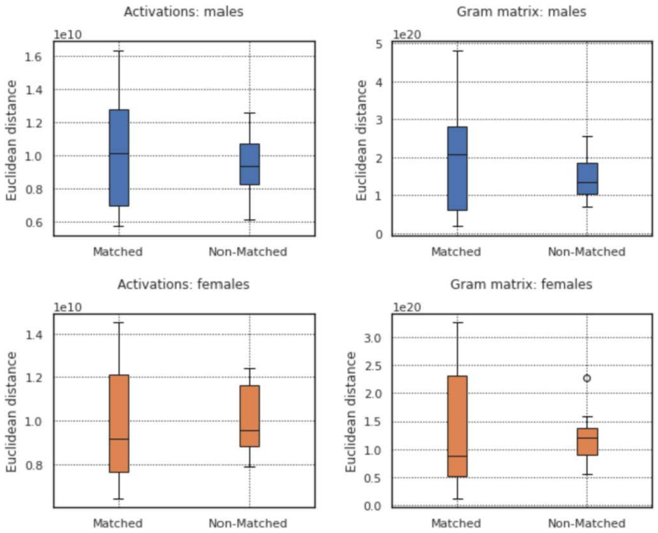

Fig. 6. Matched and non-matched darter-habitat distances for layer 3. The left panel displays activation distances and the right panel indicates gram-matrix distances. Blue boxes represent data for males, and orange boxes indicate data for females. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

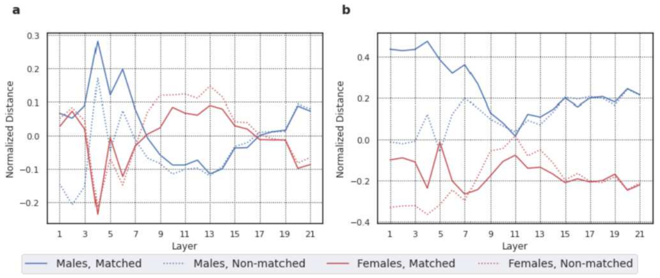

Fig. 7. Distances between darters and matched or non-matched habitats for both males and females, averaged across each species. (a) Activations, (b) Gram Matrices. Distances have been normalized for visualization purposes; raw distances were used in statistical tests.

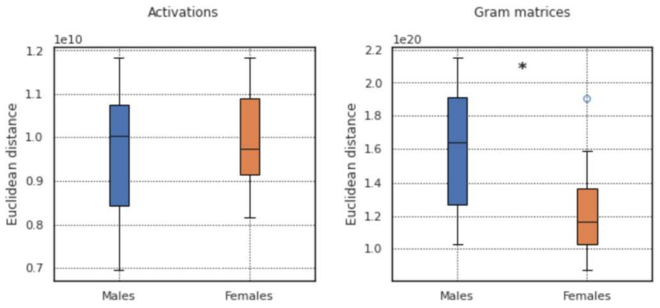

Fig. 8. Sex differences in distances between habitat images and darter images.

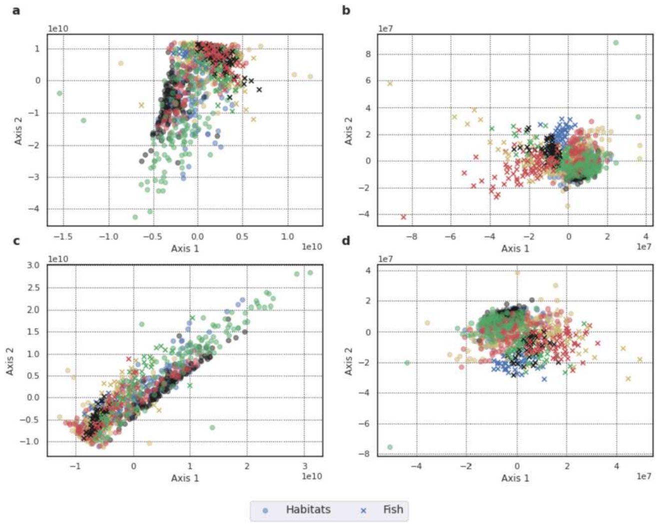

Fig. 9. First two dimensions of a multi-dimensional scaling analysis performed on Gram matrices with both fish patterns and habitat images. This was performed for males at layer 3 (a), males at layer 18 (b), females at layer 3 (c), and females at layer 18 (d). Darters are notated by checks while habitat images are notated with circles. Red marks denote gravel images and species associated with gravel habitats; green denotes sand and sand-associated species; yellow denotes boulder and boulder-associated species; black denotes detritus and detritus-associated species, and blue denotes bedrock and bedrock-associated species. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in layer 3 is different than that of the habitat images.

4. Discussion

Early on, Darwin stressed the importance of visual patterns in sexual selection. Yet our understanding of their evolution and function in mate choice, and more generally in animal communication does not match their ubiquity or importance (Cummings and Endler, 2018). Clearly, the field requires new tools to study the design of communication signals, encompassing both their simple (e.g., color, roundness, dominant orientations) and complex (e.g., relative position of complex forms like circles) components. Considering how closely they mimic biological vision, convolutional neural networks (CNNs) have opened previously intractable ways for us to understand the evolution of sexual signaling. Our results show that CNNs can discriminate fish species even when trained on very different data sets. The observed discrimination accuracy further supports the feasibility of using a small number of images to characterize each species. Our approach uses a general image classification algorithm, meaning that our results are likely transferable to other systems. Remarkably, although performance is worse for shallow layers, where each neuron's receptive field is only a 3 × 3 pixel patch, they are still able to distinguish between species beyond chance. This result indicates that species diagnostic characters occur at all spatial scales, and includes both small, very localized components, as well as large visual components spanning the entire fish flanks. In a separate analysis using images of entire fish instead of only the fish flanks, we were able to replicate and even exceed this performance (Hulse et al., unpublished data), likely because darter species differ by overall body shape in addition to body patterning. Whether or not darters effectively use visual components at all spatial scales to recognize conspecifics would require behavioral tests.

Interestingly, our results do not provide evidence that, in darter species, the male nuptial pattern matches that of species-specific habitats, contradicting our previous results. In Hulse et al. (2020), using the very same dataset, we found a statistically significant correlation between the Fourier slope of darter nuptial patterns and that of their habitats for males, but not for females. The Fourier slope describes the relative strength of luminance contrasts (averaged over orientations) across spatial scales. The Fourier slope has been implicated as a factor in human aesthetic preferences, which may be related to mimicking the Fourier statistics of natural scenes (Juricevic et al., 2010; Menzel et al., 2015; Renoult et al., 2016). To ensure that the kind of information encoded by the Fourier statistics was not missed by our CNN-based analyses, we generated datasets of white noise filtered with different Fourier slopes, and we tested whether these could be discriminated by VGG19 trained on Imagenet. Even with difference in Fourier slope as small as 0.1, using the same methods as in our fish classification, VGG19 is able to correctly classify white noise images of varying slopes with 100% accuracy at all layers (Hulse et al., unpublished data). While it is clear that VGG19 can detect differences in Fourier slope, it is unclear how these differences affect overall activation and Gram matrix differences. Since VGG19 is an object classification network, the relative importance of visual components related to the Fourier slope may be minor if they are not useful for classification.

The absence of a significant similarity between darter patterns and their preferred habitats does not support a role of sensory drive in pattern evolution; however, there are many ways in which our analysis may not have detected it. Firstly, we might expect fish-habitat correspondence to be stronger in the shallow layers of the network. It is here that the network is tuned to local, less abstract features, which are more likely to be shared between fish and habitats than the large-scale, highly abstract features of deep layers. The greater specificity in deep layers can be observed from the multidimensional scaling of the Gram matrix space (Fig. 9), where there is a greater separation between fish and habitats at the deeper layer. The greater degree of shared features between habitat images and darter images at shallow layers could be an issue, as these shallow layers are also where we see the least differentiation between species of darters (Fig. 4). Thus, in the shallow layers where we would expect to see similarities in texture between fish and habitats, there is likely not enough variation between species to accurately group species with their habitats. Future work should thus employ tailored approaches that can pick out subtle differences in shallow features.

Second, general image classification networks such as VGG19 might miss subtle features relevant to the processing of fish and habitat images (Firestone, 2020). One solution might be to use a network specifically trained to discriminate between the image classes of interest. For example, training a network to discriminate fishes versus habitats might force the network to learn features that are relevant for coding fish and habitat information. Alternatively, using VGG19 trained on Imagenet as we did here, one could train a classifier (e.g., a linear Support Vector Machine) to discriminate between fishes and habitats from activations of shallow layers to identify the most relevant features, and analyze these features only. Another approach could use sparse autoencoders, trained on relevant images, with a sparseness constraint on the intermediate layers. These networks remove redundancies in input images, similar to what Barlow hypothesized for biological neural systems. Like convolutional neural networks, sparse autoencoders trained on natural images learn features similar to cells in the mammalian visual cortex (Olshausen and Field, 2004). These approaches were not feasible for our study, as training neural networks requires a database of images much larger than ours. Generating such a database while maintaining the degree of standardization applied here is beyond the scope of this investigation.

One clear result of our study is that the patterns of female darters are more similar to stream habitats than are the patterns of males. This result is perhaps best explained by females being more cryptically camouflaged in their environment. In darters, males are actively signaling to females during the breeding season, investing considerable energy in display behavior. While there is pressure for both males and females to be inconspicuous to predators, the male darter reproductive strategy requires them to be conspicuous to females; thus, males are likely limited in the extent to which they can be camouflaged. Darter mating behavior therefore predicts that females will be more camouflaged than males. Our results are consistent with this prediction. Importantly, the predictions of camouflage and efficient coding are not mutually exclusive. Empirical evidence suggests that visual signals can be used both for camouflage and sexual signaling. For example, in an analysis spanning over 90% of all avian species, (Gluckman and Cardoso, 2010) find that barred plumage can function both as a quality indicator and a camouflage structure. Darter visual patterns are likely attractive to both sexes; both male and female darters have preferences for particular phenotypes in the opposite sex (Mattson et al., 2020; Roberts and Mendelson, 2021; Williams and Mendelson, 2011; Williams and Mendelson, 2013), suggesting sexual selection drives signal design. If pressure from both sexual selection and camouflage are driving visual patterns to resemble their habitats, it may be difficult to disentangle their effects. One possibility is that the sexes differ in conspicuousness, independently of attractiveness. By this reasoning, our methods could be detecting the conspicuousness aspect of male patterns (even without the chromatic component, which would further increase the conspicuousness of males), but might be missing components implicated in male attractiveness.

To our knowledge, our study is the first to use the Gram matrix to classify images. While this may not be the best method for accurately categorizing images, the Gram matrix may have immense utility for evolutionary biologists (Grammer et al., 2003; Grammer and Thornhill, 1994). Visual patterns are highly multidimensional; disentangling their different components can allow for a better understanding of their evolution. Neural network activations are generally thought of as encoding the content of an image, for example, whereas the Gram matrix encodes information about its texture. It makes sense that animals would mimic the texture of their habitats more so than the larger objects, except in cases of masquerade camouflage. Notably, we observed a significant difference between males and females only in terms of the Gram matrix, with female patterns being more similar to habitats than were males. For animals to be well camouflaged, texture-related features may be important to mimic. Our results show that the Gram matrix provides a valuable complement to other image analysis techniques for studying camouflage. Unlike methods such as pattern energy analysis, it can be conducted at different layers, allowing researchers to examine spatial components at different spatial scales (Troscianko et al., 2017).

5. Conclusions

In sum, this work demonstrates the utility of convolutional neural networks for generating biologically informed distance metrics for patterns. While these methods are comparatively new in the field of ecology and evolution of animal communication, the last decade has seen artificial neural networks revolutionize many fields. Owing to the similarities between CNNs and vertebrate visual processing, neural network-based approaches greatly expand our capability to understand how sensory coding influences signal design. Similarity metrics, such as those employed here, may be particularly useful for tracking evolutionary changes such as pattern divergence. Although our results do not clearly support the role of sensory drive in pattern evolution, neural network analyses clearly offer a starting point for a line of inquiry that could vastly expand our understanding of signal design.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank Alexandre Durrmeyer and Yseult Hejja-Brichard for their helpful discussions. This work was supported by National Science Foundation grants IOS-1708543 and IOS-2026334.

References

Appalaraju, S., Chaoji, V., 2018. Image similarity using Deep CNN and Curriculum Learning. arXiv:1709.08761 [cs].

Bailey, R.M., Etnier, D.A., 1988. Comments on the Subgenera of Darters (Percidae) with Descriptions of Two New Species of *Etheostoma* (*Ulocentra*) from Southeastern United States. Miscellaneous Publications / Museum of Zoology, University of Michigan (USA).

Barlow, H.B., 1961. Possible principles underlying the transformations of sensory messages. In: Rosenblith, W.A. (Ed.), Sensory Communication. The MIT Press, pp. 216–234.

Bell, S., Bala, K., 2015. Learning visual similarity for product design with convolutional neural networks. ACM Trans. Graph. 34, 98:1-98:10. https://doi.org/10.1145/2766959.

Bharati, M.H., Liu, J.J., MacGregor, J.F., 2004. Image texture analysis: methods and comparisons. Chemom. Intell. Lab. Syst. 72, 57–71. https://doi.org/10.1016/j.chemolab.2004.02.005.

Cummings, M.E., Endler, J.A., 2018. 25 years of sensory drive: the evidence and its watery bias. Curr. Zool. 64, 471–484. https://doi.org/10.1093/cz/zoy043.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

DiCarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How does the brain solve visual object recognition? Neuron 73, 415–434. https://doi.org/10.1016/j.neuron.2012.01.010.

Endler, J.A., 1992. Signals, signal conditions, and the direction of evolution. Am. Nat. 139, S125–S153. https://doi.org/10.1086/285308.

Endler, J.A., Basolo, A.L., 1998. Sensory ecology, receiver biases and sexual selection. Trends Ecol. Evol. 13, 415–420. https://doi.org/10.1016/S0169-5347(98)01471-2.

Endler, J.A., Mappes, J., 2017. The current and future state of animal coloration research. Philos. Trans. R. Soc. B 372, 20160352. https://doi.org/10.1098/rstb.2016.0352.

Etnier, D., Starnes, W., 1993. The Fishes of Tennessee. University of Tennessee Press.

Firestone, C., 2020. Performance vs. competence in human–machine comparisons. Proc. Natl. Acad. Sci. U. S. A. 117, 26562–26571. https://doi.org/10.1073/pnas.1905334117.

Fuller, R.C., Houle, D., Travis, J., 2005. Sensory bias as an explanation for the evolution of mate preferences. Am. Nat. 166, 437–446. https://doi.org/10.1086/444443.

Gatys, L.A., Ecker, A.S., Bethge, M., 2016. Image style transfer using convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 2414–2423. https://doi.org/10.1109/CVPR.2016.265.

Gatys, L.A., Ecker, A.S., Bethge, M., 2017. Texture and art with deep neural networks. Curr. Opin. Neurobiol. 46, 178–186. https://doi.org/10.1016/j.conb.2017.08.019.

Gluckman, T.-L., Cardoso, G.C., 2010. The dual function of barred plumage in birds: camouflage and communication. J. Evol. Biol. 23, 2501–2506. https://doi.org/10.1111/j.1420-9101.2010.02109.x.

Gogul, I., Kumar, V.S., 2017. Flower species recognition system using convolution neural networks and transfer learning. In: 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). Presented at the 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), pp. 1–6. https://doi.org/10.1109/ICSCN.2017.8085675.

Grammer, K., Thornhill, R., 1994. Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. J. Comp. Psychol. 108, 233–242. https://doi.org/10.1037/0735-7036.108.3.233.

Grammer, K., Fink, B., Møller, A.P., Thornhill, R., 2003. Darwinian aesthetics: sexual selection and the biology of beauty. Biol. Rev. 78, 385–407. https://doi.org/10.1017/S1464793102006085.

Hansen, O.L.P., Svenning, J.-C., Olsen, K., Dupont, S., Garner, B.H., Iosifidis, A., Price, B. W., Høye, T.T., 2020. Species-level image classification with convolutional neural network enables insect identification from habitus images. Ecol. Evol. 10, 737–747. https://doi.org/10.1002/ece3.5921.

Hulse, S.V., Renoult, J.P., Mendelson, T.C., 2020. Sexual signaling pattern correlates with habitat pattern in visually ornamented fishes. Nat. Commun. 11, 2561. https://doi.org/10.1038/s41467-020-16389-0.

Juricevic, I., Land, L., Wilkins, A., Webster, M.A., 2010. Visual discomfort and natural image statistics. Perception 39, 884–899. https://doi.org/10.1068/p6656.

Kriegeskorte, N., 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. Ann. Rev. Vis. Sci. 1, 417–446. https://doi.org/10.1146/annurev-vision-082114-035447.

Kuehne, R.A., Barbour, R.W., 2015. The American Darters. University Press of Kentucky.

Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J.-P., Baciu, M., Kahane, P., Rheims, S., Vidal, J.R., Aru, J., 2018. Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. Commun. Biol. 1, 1–12. https://doi.org/10.1038/s42003-018-0110-y.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lo, S.C.B., Freedman, M.T., Mun, S.K., Gu, S., 2018. Transformationally Identical and Invariant Convolutional Neural Networks through Symmetric Element Operators. arXiv:1806.03636 [cs].

Mattson, C.L., Roberts, N.S., Mendelson, T.C., 2020. Male preference for conspecific females depends on male size in the splendid darter, Etheostoma barrenense. Anim. Behav. 165, 89–96. https://doi.org/10.1016/j.anbehav.2020.04.022.

Menzel, C., Hayn-Leichsenring, G.U., Langner, O., Wiese, H., Redies, C., 2015. Fourier power spectrum characteristics of face photographs: attractiveness perception depends on low-level image properties. PLoS One 10, e0122801. https://doi.org/10.1371/journal.pone.0122801.

Nguyen, T.T.N., Pantuwong, N., Yagi, Y., 2016. Flower Species Identification Using Deep Convolutional Neural Networks, p. 7.

Olshausen, B.A., Field, D.J., 2004. Sparse coding of sensory inputs. Curr. Opin. Neurobiol. 14, 481–487. https://doi.org/10.1016/j.conb.2004.07.007.

Pérez-Rodríguez, L., Jovani, R., Mougeot, F., 2013. Fractal geometry of a complex plumage trait reveals bird's quality. Proc. R. Soc. B Biol. Sci. 280, 20122783. https://doi.org/10.1098/rspb.2012.2783.

Pérez-Rodríguez, L., Jovani, R., Stevens, M., 2017. Shape matters: animal colour patterns as signals of individual quality. Proc. R. Soc. B 284, 20162446. https://doi.org/10.1098/rspb.2016.2446.

Pouli, T., Reinhard, E., Cunningham, D.W., Reinhard, E., Cunningham, D.W., 2013. Image Statistics in Visual Computing. A K Peters/CRC Press.

Renoult, J.P., Mendelson, T.C., 2019. Processing bias: extending sensory drive to include efficacy and efficiency in information processing. P. Roy. Soc. B 286, 20190165. https://doi.org/10.1098/rspb.2019.0165.

Renoult, J.P., Bovet, J., Raymond, M., 2016. Beauty is in the efficient coding of the beholder. R. Soc. Open Sci. 3, 160027 https://doi.org/10.1098/rsos.160027.

Roberts, N.S., Mendelson, T.C., 2021. Identifying female phenotypes that promote behavioral isolation in a sexually dimorphic species of fish Etheostoma zonale. Curr. Zool. 67, 225–236. https://doi.org/10.1093/cz/zoaa054.

Rogowitz, B.E., Frese, T., Smith, J.R., Bouman, C.A., Kalin, E.B., 1998. Perceptual image similarity experiments. In: Human Vision and Electronic Imaging III. Presented at the Human Vision and Electronic Imaging III, International Society for Optics and Photonics, pp. 576–590. https://doi.org/10.1117/12.320148.

Ryan, M.J., 1998. Sexual selection, receiver biases, and the evolution of sex differences. Science 281, 1999–2003. https://doi.org/10.1126/science.281.5385.1999.

Ryan, M.J., Cummings, M.E., 2013. Perceptual biases and mate choice. Annu. Rev. Ecol. Evol. Syst. 44, 437–459. https://doi.org/10.1146/annurev-ecolsys-110512-135901.

Ryan, M.J., Rand, A.S., 1990. The sensory basis of sexual selection for complex calls in the Túngara frog, *Physalaemus Pustulosus* (sexual selection for sensory exploitation). Evolution 44, 305–314. https://doi.org/10.1111/j.1558-5646.1990.tb05200.x.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D.J., Miyagi, R., van der Sluijs, I., Schneider, M.V., Maan, M.E., Tachida, H., Imai, H., Okada, N., 2008. Speciation through sensory drive in cichlid fish. Nature 455, 620–626. https://doi.org/10.1038/nature07285.

Shaha, M., Pawar, M., 2018. Transfer learning for image classification. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). Presented at the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 656–660. https://doi.org/10.1109/ICECA.2018.8474802.

Sibeaux, A., Cole, G.L., Endler, J.A., 2019. The relative importance of local and global visual contrast in mate choice. Anim. Behav. 154, 143–159. https://doi.org/10.1016/j.anbehav.2019.06.020.

Simoncelli, E.P., Olshausen, B.A., 2001. Natural image statistics and neural representation. Annu. Rev. Neurosci. 24, 1193–1216. https://doi.org/10.1146/annurev.neuro.24.1.1193.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs].

Sinha, P., Russell, R., 2011. A perceptually based comparison of image similarity metrics. Perception 40, 1269–1281. https://doi.org/10.1068/p7063.

Stauffer, J.R., Boltz, J.M., Kellogg, K.A., van Snik, E.S., 1996. Microhabitat partitioning in a diverse assemblage of darters in the Allegheny River system. Environ. Biol. Fish 46, 37–44. https://doi.org/10.1007/BF00001696.

Tibbetts, E.A., Dale, J., 2004. A socially enforced signal of quality in a paper wasp. Nature 432, 218–222. https://doi.org/10.1038/nature02949.

Troscianko, J., Skelhorn, J., Stevens, M., 2017. Quantifying camouflage: how to predict detectability from appearance. BMC Evol. Biol. 17, 7. https://doi.org/10.1186/s12862-016-0854-2.

Welsh, S.A., Perry, S.A., 1998. Habitat partitioning in a community of darters in the Elk River, West Virginia. Environ. Biol. Fish 51, 411–419. https://doi.org/10.1023/A:1007428930851.

Wenliang, L.K., Seitz, A.R., 2018. Deep neural networks for modeling visual perceptual learning. J. Neurosci. 38, 6028–6044. https://doi.org/10.1523/JNEUROSCI.1620-17.2018.

Williams, T.H., Mendelson, T.C., 2010. Behavioral isolation based on visual signals in a sympatric pair of darter species. Ethology 116, 1038–1049. https://doi.org/10.1111/j.1439-0310.2010.01816.x.

Williams, T.H., Mendelson, T.C., 2011. Female preference for male coloration may explain behavioural isolation in sympatric darters. Anim. Behav. 82, 683–689. https://doi.org/10.1016/j.anbehav.2011.06.023.

Williams, T.H., Mendelson, T.C., 2013. Male and female responses to species-specific coloration in darters (Percidae: *Etheostoma*). Anim. Behav. 85, 1251–1259. https://doi.org/10.1016/j.anbehav.2013.03.012.

Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. 111, 8619–8624. https://doi.org/10.1073/pnas.1403112111.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Salt Lake City, UT, pp. 586–595. https://doi.org/10.1109/CVPR.2018.00068.

Zhou, H., Yan, C., Huang, H., 2016. Tree species identification based on convolutional neural networks. In: 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Presented at the 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 103–106. https://doi.org/10.1109/IHMSC.2016.144.