Does Colour Really Matter? Evaluation via Object Classification

Brian Funt and Ligeng Zhu; Simon Fraser University, Vancouver, B.C., Canada

Abstract

Colour is important, but how important? This study addresses the question by testing a deep learning approach, ResNet-50, on the task of object classification based on using full-colour, dichromatic, and grayscale images as inputs and comparing the recognition performance as the amount of colour information is reduced. The results show that colour is useful, but far from crucial for object classification. The error rate increases by only 12% for the grayscale case over the full-colour case. A examination of some of the cases in which the full-colour classifier succeeds, but the grayscale classifier fails, reveals the interesting trend that while the in some cases the colour features of an object are crucial, colour may be perhaps even more important for understanding occlusion ordering and figure-ground separation.

Introduction

We all feel that colour is important, but how important is it really for day-to-day tasks? For example, movies and television programs were originally only displayed in black and white (i.e., grayscale), but nonetheless, people readily understood them. If we were all monochromats, how much would that affect our ability to interact with the world? Analogously, to what extent does being a dichromat (i.e., 'colour blind') affect a person's performance on day-to-day tasks? Similarly, in the case of trichromats with normal colour vision, how much might the limitations of colour constancy affect their performance when the lighting changes?

In principle, the only way to address such questions is to test human subjects in a large number of situations. However, the example of black and white television illustrates the fact that the performance difference on day-to-day tasks, while most likely significant, may not be that large, so we can expect that an extensive amount of psychophysical testing will be required in order to establish a reliable result.

Rather than perform a large psychophysical study, this paper explores the issue of how much colour matters by evaluating it in the context of object classification (i.e., identifying the main object in an image) via an artificial neural network using 'deep learning' of the sort developed by Krizhevsky et al. [10]. Needless to say, any conclusions drawn will not directly apply to the case of human subjects, but the expectation is that studying computer-based object classification will give some insight into the kinds of situations in which colour does make a difference and the magnitude of that difference.

Deep learning approaches to object classification [10] have been extraordinarily successful on the ImageNet [3] dataset of one million manually-labelled images. This paper addresses the question of how much the accuracy of such deep learning methods is affected: (i) when the illumination changes the input colour signal, (ii) when the input is dichromatic; and (iii) when the input is monochromatic. The goal is to gain some insight into the degree to which colour matters for day-to-day tasks using object classification as a prototypical example.

Background

The study of colour vision deficiencies, how to represent their effect to observers with normal colour vision, and how to mitigate the problems they create for colour deficient observers all have a long history. In particular, Viénot et al. [16] described a method of representing images of reduced colour for trichromatic observers that appear identical to dichromatic observers. This method was detailed and implemented in software by Brettel et al. [1]. The method is based on the assumption that the achromatic axis is the same for dichromats and trichromats combined with an assumption based on the wavelengths of the colours that unilateral dichromats report as being the same to both eyes. Machado et al. [11] provide an alternative model based on a photoreceptor-spectral-response stage followed by an opponent-color stage. The advantage of the Machado model is that it is not limited to dichromats but generalizes to the case of anomalous trichromats. Ramaswamy et al. [13] evaluate colour differences in images created using the Brettel model and report the negative result that the colour differences measured in the simulated trichromatic representation are no more accurate than those measured in the original trichromatic image when used to predict the error rates in colour identification for dichromatic observers.

In a different vein, 'Daltonization' is the process of modifying a standard colour image to make the colour differences in the image more apparent to the dichromatic observer. Numerous techniques have been proposed, but psychophysical experiments conducted by Simon-Liedtke et al. [15] concluded that Kotera's [8] Daltonization method was the most effective. It considers the hue confusion colours and specifically shifts the them to enhance their visibility for the dichromat.

Another approach to understanding the information available to dichromats is to 'colorize' dichromatic images. In the digital image case, Cardei [2] used a neural network to predict the G channel of an RGB image given only the R and B channels and showed that this could be done surprisingly well.

While the simulations of dichromatic vision for trichromats and the enhancement of the visibility of the confusion colours for dichromats are very interesting, neither provides us with an idea of how important the role of colour is in object identification or classification. Object classification by deep learning approaches has achieved human-level accuracy [5] and so is used here as a testbed for understanding the role of colour in everyday tasks.

Convolutional Neural Networks (CNNs), a special class of neural network, have become a central component of many computer vision systems, given their success in object classification and object recognition. CNNs have the advantage that the total number of parameters to be learned is reduced by the fact that the weights defining the convolution kernels are shared across many inputs. By stacking convolutional layers, image features are fully distilled before final classification, which leads to dramatically increased performance. ResNet [5], a representative architecture, outperformed human competitors on the ImageNet [10] classification task. Since it exhibits human-level performance, it is used here to explore the role colour plays in object classification.

Method

Using the ResNet architecture, an object classification network is trained and tested in four different ways. First, it is trained and tested on trichromatic LMS data. Second it is trained and tested on LMS data to which a synthetic variation in the illumination's chromaticity is added (details below). Third, it is tested on dichromatic LS data and tested on similar data. Fourth, it is tested on greyscale, CIE Y, data.

The test and training images are from the CIFAR dataset [9], which contains 60,000 images in 100 classes. Of these, 50,000 images are used for training and the remaining 10,000 images are used for testing. They are nominally in non-linear sRGB format so they are converted to LMS using the sRGB [6] method of converting from non-linear sRGB to linear sRGB, and from there to CIE XYZ. XYZ values are converted to cone LMS using the the Hunt-Pointer-Estevez [4] matrix.

To model some of the colour effects created by non-uniform illumination across a scene, a linearly interpolated variation in the illumination colour across the image is created using von Kries scaling by random amounts in the x-direction across some images and the y-direction across others. In either case, for random numbers {r1,r2,r3,r4} in [0.6, 1.4] the following scalings are applied to the R and B channels:

$$R(x,y) \leftarrow R(x,y) * \text{interpolate\_across\_image}(r1,r2) \qquad (1)$$
$$B(x,y) \leftarrow B(x,y) * \text{interpolate\_across\_image}(r3,r4)$$

The network's structure is diagrammed in Figure 1. It is a network modelled after the design of ResNet-50 [5] It is composed of four configuration (cfg) blocks with each block containing several identical convolution layers.

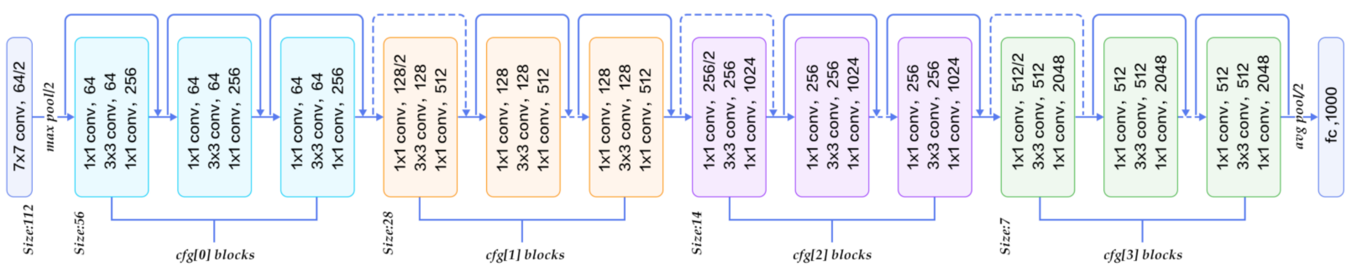

Figure 1. *The architecture of ResNet-50 [5] (diagram from* Das *[14]*)

There are four configuration blocks in ResNet with each block consisting of a series of repeat units. Each basic unit contains three convolution layers followed by batch normalization [7] and Rectified Linear Units [12]. The solid blue lines indicate identity mappings, so-called 'skip' connections, and the dashed blue lines indicate downsampling of results from a previous convolution. For visualization purposes, only 3 units are depicted per block in Figure 1. The actual numbers per block from left to right are 3, 4, 6, and 3.

Results

The classifier network is trained and tested on trichromatic (LMS), dichromatic (LS) and monochromatic (Y) images derived from the sRGB images in the CIFAR dataset. Stochastic Gradient Descent is used as the optimizer with momentum 0.9. The network is trained for 300 epochs, with the learning rate initially set to 0.1 and then decaying to 0.01 and 0.001 at epochs 150 and 225, respectively.

The results in Table 1 indicate that colour does matter, but it is not crucial for object classification. When trained and tested on the original data (i.e, without added illumination variation), the error for the dichromatic case increases by 15% (33.1 versus 28.8) and the monochromatic case by 22% (35.0 versus 28.8). However, when synthetic, spatially-varying illumination is added, the situation changes. The trichromatic case is more sensitive to the change in colours induced by the illumination variation in a way that is not completely accounted for by augmenting the training

Table 1 *Misclassification error rates in percent over the test set of 10,000 images. The row labelled 'Original' lists the error rates when training and testing is done on the unmodified image data. The row labelled "Test Varying Illumination" lists the errors for the case when the classifier is trained on the unmodified data, but then tested on the images modified with the simulated illumination variation as per Eq. 1. The row "Train & Test Varying Illumination" indicates the classifier was trained and tested on the modified data.*

	LMS	LS	Y
Original	28.8	33.1	35.0
Test Varying Illumination	52.4	51.4	48.0
Train & Test Varying Illumination	31.3	33.0	35.0

data with similar examples. On the other hand, the monochromatic case is both less sensitive to the illumination-induced changes in colour, and training with the augmented data manages to account for the differences. As a result, the dichromatic classification error is only 5% (33.0 versus 31.3) greater than the trichromatic classification error, and the monochromatic classification error is only 12% (35.0 versus 31.3) higher than the trichromatic error.

It is interesting to consider the cases in which the classification is correct based on the trichromatic input, but incorrect on the grayscale input, and vice versa. The examples in Figures 2-5 all are of the former type; however, the different examples show different ways in which colour may be important to obtaining the correct classification.

Figure 2. Colour classification correct as 'tiger'; grayscale as 'snow leopard.' A case where the colour of the object itself is clearly crucial to identifying it.

Figure 3. Colour classification correct as 'wall clock'; grayscale as 'stove.' In this case, it appears that colour is important both in figure-ground separation (i.e., clocks from the wall versus multiple stove burners) and in making the hands of the clock visible.

Figure 4. Colour classification correct as 'tricycle'; grayscale as 'accordion.' It would seem that the classifier has interpreted the shadows as keys of an accordian. This points to the usefulness of colour in interpreting lighting effects and in creating a greater sense of three dimensionality.

Figure 5. Colour classification correct as 'worn fence'; grayscale as 'wreck.' Again in this case, colour appears to be important in interpreting the occlusion relationships and three-dimensional structure. That, of course, is simply the authors' guess since it is impossible to know what the classifier itself has learned.

Figures 6 and Figure 7 show examples for which the grayscale classification is correct, but the trichromatic classification is incorrect.

Figure 6. Greyscale classification correct as 'volleyball'; colour classification incorrect as 'tennis ball.' In this case, the fluorescent green that is typical of many tennis balls may have misled the trichromatic classifier.

Figure 7. Greyscale classification correct (according to the CIFAR classification) as 'book jacket; colour classification incorrect as 'television.' Colour helps separate figure from ground, but in this case incorrectly.

Conclusion

Tests of machine learning applied to trichromatic (LMS), dichromatic (LS) and monochromatic images (CIE Y) indicate that colour does have an important, but not essential, role to play in the task of naming the main object in an image. Object classification is simply one example of a typical vision task. Future work will include examining the significance of colour in object detection and in scene labelling. In terms of object classification, eliminating colour altogether increased the error rate by a modest 12% and eliminating the M channel of LMS, by on only 5%. In terms of the importance of colour for humans, and dichromats in particular, this study points to the kinds of images that might be used when designing a psychophysical study to test the relevance of colour to human subjects in that there is little to be learned in testing cases such as 'tiger' Figure 2, in which colour is an obvious feature, but possibly a lot to be learned by testing cases such as the 'worn fence' Figure 5.

References

[1] H. Brettel, F. Viénot and J. D. Mollon. "Computerized simulation of color appearance for dichromats," *J. Opt. Soc. Am.* vol. 14 no. 10 pp. 2647–2655, 1997.

[2] V. Cardei, "From Dichromatic to Trichromatic Images; Implications for Image Recovery and Visualization," *Proc. PICS'99*, Society for Imaging Science and Technology, 1999.

[3] J. Deng, W. Dong, R. Socher, L.J., Li, K.Li, F. Li, "Imagenet: A large-scale hierarchical image database," *Proc. Computer Vision and Pattern Recognition,* pp. 248-255, IEEE 2009.

[4] M.D. Fairchild, *Color Appearance Models* (2nd ed.). Wiley Interscience. ISBN 978-0-470-01216-1, 2005.

[5] K. He, S. Ren, J. Sun, X. Zhang. "Deep residual learning for image recognition." *Proc. of the IEEE conference on computer vision and pattern recognition*, 2016.

[6] IEC 61966-2-1:1999 Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB, International Electrotechnical Commission, 1999.

[7] S. Ioffe, and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint* arXiv:1502.03167, 2015.

[8] H. Kotera, "Optimal daltonization by spectral shift for dichromatic vision," *Twentieth IS&T Color and Imaging Conference (CIC20)*, Society for Imaging Science and Technology, pp. 302–308, Los Angeles, 2012.

[9] A. Krizhevsky, "Learning multiple layers of features from tiny images." Univ. of Toronto, 2009.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton. "Imagenet classification with deep convolutional neural networks." *Proc. Advances in neural information processing systems*, 2012.

[11] G. Machado, M. Oliveira and L. Fernandes, "A Physiologically-based Model for Simulation of Color Vision Deficiency," *IEEE Trans. on Visualization and Computer Graphics*, vol. 15, no. 6, Nov.-Dec. 2009.

[12] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines." Proc. of the 27th international conference on machine learning (ICML-10), 2010.

[13] S. Ramaswamy and J. Hovis, "Does Dichromatic Color Simulation Predict Color Identification Error Rates?" *Optometry and Vision Science*, vol. 88, no. 5, pp. 621-627, 2011.

[14] D. Siddharth, "CNNs Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more …" Retrieved from https://medium.com/@siddharthdas_32104/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5 April 10, 2017.

[15] J. Simon-Liedtke and I. Farup, "Evaluating color vision deficiency daltonization methods using a behavioral visual-search method," *Journal of Visual Communication and Image Representation*, vol. 35, pp. 236-247, 2016.

[16] F. Viénot, H. Brettel, L. Ott, A. Ben M'Barek, and J. Mollon, J., "What do colour-blind people see?" *Nature*, vol. 376, pp 127-128, 1995.

Author Biographies

Brian Funt is Professor of Computing Science at Simon Fraser University where he has been since 1980. He obtained his Ph.D. in Computer Science from the University of British Columbia. His research focus is on computational approaches to modeling and understanding visual perception generally and color in particular. E-mail: funt@sfu.ca.

Ligeng Zhu will receive jointly his B.Sc. in Computing Science from Simon Fraser University and B.Eng. in Computer Science from Zhejiang University in 2019. His research work is focused on applied statistics and machine learning. E-mail:lykenz@sfu.ca.