Assessing The Importance Of Colours For CNNs In Object Recognition

Anonymous Author(s)
Affiliation
Address
email

Abstract

Humans rely heavily on shapes as a primary cue for object recognition. As secondary cues, colours and textures are also beneficial in this regard. Convolutional neural networks (CNNs), an imitation of biological neural networks, have been shown to exhibit conflicting properties. Some studies indicate that CNNs are biased towards textures whereas, another set of studies suggest shape bias for a classification task. However, they do not discuss the role of colours implying its possible humble role in the task of object recognition. In this paper, we empirically investigate the importance of colours in object recognition for CNNs. We are able to demonstrate that CNNs often rely heavily on colour information while making a prediction. Our results show that the degree of dependency on colours tend to vary from one dataset to another. Moreover, networks tend to rely more on colours if trained from scratch. Pre-training can allow the model to be less colour dependent. However, if it is forced to rely less on colours through data augmentations, this can negatively affect the overall accuracy. To facilitate these findings, we follow the framework often deployed in understanding role of colours in object recognition for humans. We evaluate a model trained with congruent images (images in original colours eg. red strawberries) on congruent, greyscale, and incongruent images (images in unnatural colours eg. blue strawberries). We measure and analyse network's predictive performance (top-1 accuracy) under these different stylisations. We utilise standard datasets of supervised image classification (CIFAR-100, STL-10, and Tiny ImageNet) and fine-grained image classification (CUB-200-2011, Oxford-Flowers, and Oxford-IIIT Pets) in our experiments.

1 Introduction

Colours play a vital role in our day to day life. We utilise colours for visual identification [1], search [2], gaze guidance in natural scenes [3] etc. As an example, importance of colour can be understood whilst identifying ripe fruits in a background of foliage [1]. Initially, it was widely believed that only shapes and not colours influence object recognition in humans [4, 5]. However, many studies [6, 7] indicate that colours do assist object recognition in humans. The findings by Tanaka and Presnell [6] show that colours facilitate recognition of high colour diagnostic objects (natural objects like fruits) but have little effect on low colour diagnostic objects (man-made objects like airplanes). Their experiments were based on a variation of Stroop effect [8]. Human participants were asked to name objects in different colour schemes. They observed that naming of objects with congruent colours was much faster than naming incongruently coloured objects. Greyscale images served as a neutral medium and the response time for them were in between congruent and incongruent images. In a similar study conducted by Hagen et al. [9] for investigating the role of colour in expert object recognition, similar findings were reported.

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

(a) Congruent (b) Greyscale (c) Incongruent (d) Negative[10] (e) Re-textured [11]

Figure 1: The texture and shape information is intact in original, greyscale and incongruent images. Additionally, negative images have a larger drift in distribution than incongruent images when measured against congruent images.

Neural networks are models of machine learning designed to mimic the neurological working of a human brain [12]. Today, they are employed in many different fields solving numerous tasks [13–15]. Considering the widespread adoption and blackbox nature of neural networks, considerable studies have also been performed to understand their inner working [16–19]. Zeiler and Fergus [16] illustrated the hierarchical nature of learnt features. Engilberge et al. [20] investigated the colour sensitivity of neural networks and observed that at the shallow end the neurons are more sensitive to colour information in an image. A number of studies also exist highlighting the nature of representations learnt by the network however, do so with conflicting results. One set of results show that, neural networks rely predominantly on shapes[21–25]. On the other hand, many more approaches oppose the theory of shape bias in CNNs[26, 10, 27–29]. Texture is referred to as "a function of the spatial variation in pixel intensities" [30, 31]. Linear classifier based on texture representations of a neural network performs comparable to the original model [32]. On similar line, Geirhos et al. [11] showed that an ImageNet [33] trained model is biased towards texture.

The objective of our paper is to help bridging the gap between human perception and artificial intelligence, providing empirical experiments based on classical neuroscience framework to exhaustively investigate the dependency of CNNs on colour information present in an image. We believe the majority of existing approaches do highlight representation bias of CNNs but fail to address the role of colours. Bahng et al. [34] tend to the issue of colour bias in their experiments, however, do so with the motivation to learn unbiased representations. Moreover, by either focusing on small number of test images [21, 29] or a single dataset [11, 26] we are unable to observe a bigger picture.

We believe a fair approach of highlighting the importance of colour is by utilising the framework as used by the psychophysical experiments of [1, 6, 9, 35]. We can easily observe the relevance of colours by comparing performances on congruent and greyscale images. Additionally, by comparing the performances on greyscale and incongruent images we will be able to observe the effect of incorrect colour information (see Figure 1 for sample images).

In this paper, we evaluate the importance of colours for numerous datasets under 2 settings:

1. **Local Information** (Section 4.1): Evaluating the importance of colours when a CNN can only tend to small patches in a global shape agnostic manner.
2. **Global Information** (Section 4.2): For this setting, no such restriction is applied on the network and corresponds to standard approach of training a CNN.

Apart from these 2 modes of experimentation, we also evaluate a model under 4 different training schemes to replicate a typical training scenario for a classification task. We train a network (i) from scratch, (ii) with fine-tuning on pre-trained weights, (iii) with colour based augmentations (random hue, saturation, brightness, etc.), and (iv) with incongruent images to reduce colour dependency. We conclude in section 5.

2 Data

In our study we have employed datasets from image classification and fine-grained visual classification (FGVC). Table 1 provides some common statistics of the datasets used.

[1] Official train and test split has been swapped.

Table 1: Dataset statistics

Category	Dataset	#Classes	$\lvert\mathcal{D}^{train}\rvert$	$\lvert\mathcal{D}^{test}\rvert$	Image size
Image classification	C100	100	$50,000$	$10,000$	32×32
	S10	10	$5,000$	$8,000$	96×96
	TIN	200	$100,000$	$5,000$	64×64
Fine-grained visual classification	CUB	200	$5,994$	$5,794$	224×224
	OF [1]	102	$6,149$	$2,040$	224×224
	OP	37	$3,680$	$3,669$	224×224

3 Method

A dataset $\mathcal{D} = \{(x_i, y_i), i = 1, \ldots, N\}$ is composed of images $x_i \in \mathbb{R}^{C \times H \times W}$ and their corresponding labels y_i. $\mathcal{D}^{train}$, $\mathcal{D}^{test}$ denotes the split of the dataset into train and test sets respectively. We convert $\mathcal{D}$ into different colour schemes as described below:

- **Congruent Images** ($\mathcal{D}_C$): These are the images in original colours. All the subsequent transformations described below are applied to $\mathcal{D}_C$.
- **Greyscale Images** ($\mathcal{D}_G$): The congruent images converted into greyscale(luminance) images, $G = 0.29 * r + 0.587 * g + 0.114 * b$. We copy the single channel greyscale image into 3 channels.
- **Incongruent Images** ($\mathcal{D}_I$): The channels of the congruent images are switched to generate unnaturally coloured images. Formally, the default correspondence for colour channels is as $C[0] = red, C[1] = green, C[2] = blue$. We switch the channels such that the new ordering represents $C[0] = green, C[1] = blue, C[2] = red$. The advantage of it is that firstly, it preserves the texture and secondly, the changes in distribution are more contained than approaches deployed by [10, 11]. For example, the Jensen-Shannon divergence [36] $JS(\mathcal{D}_C^{train}, \mathcal{D}_G^{train}) = 0.06$ and $JS(\mathcal{D}_C^{train}, \mathcal{D}_I^{train}) = 0.1$, whereas same for negative images (as in [10]) and $\mathcal{D}_C^{train}$ is 0.3 for S10 dataset. More in supplementary document.

Figure 1 shows an example of these stylisations. As humans, we learn from our surroundings which we perceive in congruent colours. Since, we are following the framework used in identifying role of colours in humans, we train CNNs only on congruent images ($\mathcal{D}_C^{train}$) while evaluating the top-1 accuracy (represented as *Acc*) on the test sets of different stylisations described above.

4 Experiments

4.1 Access to local information

Baker et al. [29] report that CNNs can represent local shapes however, fail to utilise it in a global context. Similarly, to highlight absence of shape bias in a CNN, Geirhos et al. [11] utilise BagNets[28] to compare model performance under artistic stylisations. BagNets only have access to small patches within an image due to its design and utilises bag-of-features based framework for making a prediction. It does not make use of spatial ordering of the local features hence making it suitable to compare the relevance of colours to local shape and texture information.

We use BagNet-9[2] which has a 9×9 receptive field over the image and is built upon the Resnet-50 architecture. We report the mean accuracy and standard deviation over 3 runs. The network is trained from scratch and the data augmentations are limited to random horizontal flips, random rotation and random cropping. The details on hyper-parameters to train the network are provided in the supplementary document.

4.1.1 Results

Table 2 lists the accuracies obtained for the test sets under different stylsations. When comparing Acc($\mathcal{D}_C$) with corresponding $\mathcal{D}_I$ and $\mathcal{D}_G$, we observe significant drop in performance. We can also see the varying nature of the gaps in performance across the datasets. This shows that the colours do

[2] Sourced from official github implementation

Table 2: Performance of BagNet-9 on different test stylisations

Datasets	Acc($\mathcal{D}_C$)	Acc($\mathcal{D}_G$)	Acc($\mathcal{D}_I$)
C100	45.67 ± 0.74	18.22 ± 0.33	10.17 ± 1.2
TIN	29.33 ± 1.27	9.74 ± 0.79	6.42 ± 0.22
S10	58.41 ± 0.22	48.83 ± 0.02	35.09 ± 0.45
CUB	16.28 ± 0.08	2.00 ± 0.12	1.0 ± 0.06
OF	70.53 ± 0.27	7.37 ± 1.97	3.06 ± 0.27
OP	31.65 ± 1.17	20.19 ± 3.27	7.11 ± 0.65

influence a network but in varying degree. Moreover, for S10, OP, and VA, if we compare Acc($\mathcal{D}_C$) to Acc($\mathcal{D}_G$) we observe that the drop in accuracy is there but comparatively less than the other datasets. This suggests that the network also relies on non-colour features (such as local shapes, textures) to make a decision. Additionally, by comparing Acc($\mathcal{D}_G$) and Acc($\mathcal{D}_I$) we can notice the consistently lower performance for the latter. This suggests that incorrect colour information does indeed harm the prediction accuracy.

These results indicate that even though a CNN can represent local shape [29] or is biased towards texture[11], colours drives the prediction at a local setting.

4.2 Access to global information

In the previous experiment, we limited the view of the network to only attend to 9×9 patches of the image in a global shape agnostic fashion. Here, we use ResNet-18 which has no such restriction on its receptive field. It can thus tend to global shape information in the image.

To assess the importance of colours we train a network in the following 4 ways:

1. **Vanilla training**: Similar to the setting in Section 4.1, we train the network from scratch. The data augmentations used are random rotation, random cropping and random horizontal flips.
2. **Fine-tuning**: In practice, often a pre-trained network is fine tuned on the dataset at hand [37]. In this experiment, we follow this protocol and use a network with pre-trained weights from the ImageNet dataset. The augmentation strategy is kept identical to vanilla training.
3. **Colour augmentation**: Often colour based augmentations are used in training [38, 39] allowing the network to be colour invariant. All the training settings are identical to fine-tune except for the fact that we also add random colour jitter(hue, saturation, contrast, and brightness) to an image while training.
4. **Channel switch pre-training**: Many of the previous works studying the shape and texture bias have imposed learning limitations by training the network on stylised data [10, 11] and fine-tuning on original images. This has been shown to improve predictive performance of the model. Following the same approach, we first train the network following 'colour augmentation' protocol with *randomised channel switching* as an additional data augmentation method. Then we proceed to fine-tune the network further without 'randomised channel switching' and 'colour jitter'.

All the hyper-parameter details are provided in the supplementary document[3] .

4.2.1 Results

In tables 3 and 4, we can make a number of observations. Firstly, $Acc(\mathcal{D}_C^{test}) > Acc(\mathcal{D}_G^{test}) > Acc(\mathcal{D}_I^{test})$ which is similar to the order we observe for humans [9, 35]. Secondly, fine-tuning not only improves a model's performance on $\mathcal{D}_C^{test}$ but also helps to focus more on shapes, evident from improved performance on greyscale and incongruent images. This observation has not yet been reported in existing literature. Thirdly, incorporating colour augmentations or channel-switching into training can enforce a model to perform based on shapes, but we also observe a drop in performance on the congruent images. This forces to re-evaluate the strategies for enforcing shape bias in the network and the reasons to do so. Finally, we also observe that the importance of colours is different

[3]Including results for Mobilenet-v2 and DenseNet-121

Table 3:
Test accuracies(in %) for different training strategies on classification datasets.

Training approach	Datasets	Acc($\mathcal{D}_C$)	Acc($\mathcal{D}_G$)	Acc($\mathcal{D}_I$)
Vanilla	C100	59.10 ± 0.11	29.90 ± 0.03	22.86 ± 0.16
	S10	70.30 ± 1.67	64.13 ± 2.13	52.47 ± 1.29
	TIN	40.88 ± 0.10	19.40 ± 0.02	15.47 ± 1.31
Fine-tuning	C100	$\mathbf{61.39} \pm 0.12$	34.13 ± 0.16	31.53 ± 0.55
	S10	$\mathbf{88.10} \pm 0.30$	80.59 ± 0.39	73.45 ± 0.69
	TIN	$\mathbf{47.52} \pm 0.09$	24.38 ± 0.21	21.96 ± 0.19
Colour augmentation	C100	60.06 ± 0.29	36.68 ± 0.82	34.67 ± 0.24
	S10	88.04 ± 0.23	83.29 ± 0.46	80.88 ± 0.06
	TIN	47.42 ± 0.39	$\mathbf{27.41} \pm 1.10$	28.27 ± 0.09
Channel switch pre-training	C100	60.24 ± 0.22	$\mathbf{37.76} \pm 0.26$	$\mathbf{45.84} \pm 0.44$
	S10	87.20 ± 0.45	$\mathbf{84.01} \pm 0.19$	$\mathbf{85.23} \pm 0.21$
	TIN	44.20 ± 0.16	25.75 ± 0.46	$\mathbf{32.11} \pm 0.02$

Table 4:
Test accuracies(in %) for different training strategies on FGVC datasets.

Training approach	Datasets	Acc($\mathcal{D}_C$)	Acc($\mathcal{D}_G$)	Acc($\mathcal{D}_I$)
Vanilla	CUB	61.39 ± 0.03	12.40 ± 0.58	6.03 ± 0.23
	OF	92.03 ± 0.72	16.88 ± 4.81	9.68 ± 1.42
	OP	68.61 ± 0.63	47.75 ± 0.80	31.77 ± 0.15
Fine-tuning	CUB	$\mathbf{77.39} \pm 0.40$	30.99 ± 0.36	17.62 ± 2.30
	OF	97.37 ± 0.03	64.77 ± 1.42	56.27 ± 3.32
	OP	$\mathbf{88.51} \pm 0.86$	72.06 ± 1.15	61.59 ± 2.11
Colour augmentation	CUB	75.07 ± 0.09	38.85 ± 1.33	34.09 ± 0.15
	OF	$\mathbf{97.50} \pm 0.14$	78.82 ± 0.67	76.37 ± 1.82
	Pets	88.15 ± 0.32	72.86 ± 0.05	68.22 ± 1.04
Channel switch pre-training	CUB	71.5 ± 0.67	$\mathbf{45.41} \pm 0.32$	$\mathbf{69.93} \pm 0.57$
	OF	97.15 ± 0.06	$\mathbf{85.44} \pm 1.24$	$\mathbf{96.17} \pm 0.21$
	OP	87.16 ± 1.04	$\mathbf{77.65} \pm 1.29$	$\mathbf{85.77} \pm 1.11$

for different datasets. It encourages to think if colour diagnosticity is a phenomenon that we can expect a deep neural network to exhibit. To recall, colour diagnosticity is shown by humans where, we rely on colours to recognise specific object categories.

5 Conclusion

We believe, ours is the first work to recognise unattributed impact of colours to the shape/texture driven research for understanding bias in CNNs. By adopting the psychophysical experiment for CNNs, we have provided empirical evidence to highlight high impact of colours. Through emprical evaluations, we showed that the common practice of fine-tuning allows for less colour dependent features and improved performance on shape dependent tasks. However, increasing colour invariance reduces the performance on original colours, highlighting the importance of colours in the congruent image task. Existing approaches of enforcing shape dependency should be employed with caution. By employing a number of datasets we could further promote the idea of inconsistency in biases learnt. Different datasets exhibited different level of dependency on colours. This has encouraged us to investigate colour diagnosticity of neural networks in future research.

References

[1] Inês Bramão, Luís Faísca, Karl Magnus Petersson, and Alexandra Reis. The contribution of color to object recognition. In Ioannis Kypraios, editor, *Advances in Object Recognition*

Systems, chapter 4. IntechOpen, Rijeka, 2012. doi: 10.5772/34821. URL https://doi.org/10.5772/34821.

[2] Aave Hannus, Ronald van den Berg, Harold Bekkering, Jos B. T. M. Roerdink, and Frans W. Cornelissen. Visual search near threshold: Some features are more equal than others. *Journal of Vision*, 6(4):15–15, 07 2006. ISSN 1534-7362. doi: 10.1167/6.4.15. URL https://doi.org/10.1167/6.4.15.

[3] Antje Nuthmann and George L. Malcolm. Eye guidance during real-world scene search: The role color plays in central and peripheral vision. *Journal of Vision*, 16(2):3–3, 01 2016. ISSN 1534-7362. doi: 10.1167/16.2.3. URL https://doi.org/10.1167/16.2.3.

[4] Irving Biederman and Ginny Ju. Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, 20(1):38 – 64, 1988. ISSN 0010-0285. doi: https://doi.org/10.1016/0010-0285(88)90024-2. URL http://www.sciencedirect.com/science/article/pii/0010028588900242.

[5] Irving Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94 2:115–147, 1987.

[6] James W. Tanaka and Lynn M. Presnell. Color diagnosticity in object recognition. *Perception & Psychophysics*, 61(6):1140–1153, Aug 1999. ISSN 1532-5962. doi: 10.3758/BF03207619. URL https://doi.org/10.3758/BF03207619.

[7] Galit Naor-Raz, Michael J Tarr, and Daniel Kersten. Is color an intrinsic property of object representation? *Perception*, 32(6):667–680, 2003. doi: 10.1068/p5050. URL https://doi.org/10.1068/p5050. PMID: 12892428.

[8] J. R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662, 1935. ISSN 0022-1015(Print). doi: 10.1037/h0054651. URL https://doi.org/10.1037/h0054651.

[9] Simen Hagen, Quoc C. Vuong, Lisa S. Scott, Tim Curran, and James W. Tanaka. The role of color in expert object recognition. *Journal of Vision*, 14(9):9–9, 08 2014. ISSN 1534-7362. doi: 10.1167/14.9.9. URL https://doi.org/10.1167/14.9.9.

[10] Hossein Hosseini, Baicen Xiao, Mayoore Jaiswal, and Radha Poovendran. Assessing shape bias property of convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[11] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bygh9j09KX.

[12] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61: 85 – 117, 2015. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2014.09.003. URL http://www.sciencedirect.com/science/article/pii/S0893608014002135.

[13] O. Abdel-Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10):1533–1545, 2014.

[14] Yoav Goldberg. Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, 10(1):1–309, 2017. doi: 10.2200/S00762ED1V01Y201703HLT037. URL https://doi.org/10.2200/S00762ED1V01Y201703HLT037.

[15] Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao. Deepdriving: Learning affordance for direct perception in autonomous driving. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[16] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10590-1.

[17] Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1 – 15, 2018. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2017.10.011. URL `http://www.sciencedirect.com/science/article/pii/S1051200417302385`.

[18] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *International Conference on Learning Representations*, 2014. URL `http://arxiv.org/abs/1312.6199`.

[19] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *Workshop at International Conference on Learning Representations*, 2014.

[20] M. Engilberge, E. Collins, and S. Süsstrunk. Color representation in deep neural networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2786–2790, 2017.

[21] Jonas Kubilius, Stefania Bracci, and Hans P. Op de Beeck. Deep neural networks as a computational model for human shape sensitivity. *PLOS Computational Biology*, 12(4):1–26, 04 2016. doi: 10.1371/journal.pcbi.1004896. URL `https://doi.org/10.1371/journal.pcbi.1004896`.

[22] Astrid A. Zeman, J. Brendan Ritchie, Stefania Bracci, and Hans Op de Beeck. Orthogonal representations of object shape and category in deep convolutional neural networks and human visual cortex. *Scientific Reports*, 10(1):2453, Feb 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-59175-0. URL `https://doi.org/10.1038/s41598-020-59175-0`.

[23] Samuel Ritter, David G. T. Barrett, Adam Santoro, and Matthew M Botvinick. Cognitive psychology for deep neural networks: A shape bias case study. *ArXiv*, abs/1706.08606, 2017.

[24] Reuben Feinman and Brenden M. Lake. Learning inductive biases with simple neural networks. *ArXiv*, abs/1802.02745, 2018.

[25] Kamila M. Jozwik, Nikolaus Kriegeskorte, Katherine R. Storrs, and Marieke Mur. Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Frontiers in Psychology*, 8:1726, 2017. ISSN 1664-1078. doi: 10.3389/fpsyg.2017.01726. URL `https://www.frontiersin.org/article/10.3389/fpsyg.2017.01726`.

[26] Gaurav Malhotra and Jeff Bowers. The contrasting roles of shape in human vision and convolutional neural networks. In *CogSci*, 2019.

[27] Katherine L. Hermann and Simon Kornblith. Exploring the origins and prevalence of texture bias in convolutional neural networks, 2019.

[28] Wieland Brendel and Matthias Bethge. Approximating cnns with bag-of-local-features models works surprisingly well on imagenet. *International Conference on Learning Representations*, 2019. URL `https://openreview.net/pdf?id=SkfMWhAqYQ`.

[29] Nicholas Baker, Hongjing Lu, Gennady Erlikhman, and Philip J. Kellman. Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, 14 (12):1–43, 12 2018. doi: 10.1371/journal.pcbi.1006613. URL `https://doi.org/10.1371/journal.pcbi.1006613`.

[30] Mihran Tuceryan and Anil K. Jain. *Texture Analysis*, page 235–276. World Scientific Publishing Co., Inc., USA, 1993. ISBN 9810211368.

[31] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5): 786–804, 1979.

[32] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, page 262–270, Cambridge, MA, USA, 2015. MIT Press.

[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[34] Hyojin Bahng, Sanghyuk Chun, Sangdoo Yun, Jaegul Choo, and Seong Joon Oh. Learning de-biased representations with biased representations, 2019.

[35] David J. Therriault, Richard H. Yaxley, and Rolf A. Zwaan. The role of color diagnosticity in object recognition and representation. *Cognitive Processing*, 10:335–342, 2009.

[36] B. Fuglede and F. Topsoe. Jensen-shannon divergence and hilbert space embedding. In *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, pages 31–, 2004.

[37] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3320–3328. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf`.

[38] Andrew G. Howard. Some improvements on deep convolutional neural network based image classification. *CoRR*, abs/1312.5402, 2014.

[39] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.