Impact of Colour on Robustness of Deep Neural Networks

Kanjar De[1.2]
Luleå University of Technology[1]
971 87 Luleå, Sweden
kanjar.de@ltu.se

Marius Pedersen[2]
Norwegian University of Science and Technology[2]
Teknologiveien 22, 2802 Gjøvik, Norway
marius.pedersen@ntnu.no

Abstract

Convolutional neural networks have become the most widely used tool for computer vision applications like image classification, segmentation, object localization, etc. Recent studies have shown that the quality of images has a significant impact on the performance of these deep neural networks. The accuracy of the computer vision tasks gets significantly influenced by the image quality due to the shift in the distribution of the images on which the networks are trained on. Although, the effects of perturbations like image noise, image blur, image contrast, compression artifacts, etc. on the performance of deep neural networks on image classification have been studied, the effects of colour and quality of colour in digital images have been a mostly unexplored direction. One of the biggest challenges is that there is no particular dataset dedicated to colour distortions and colour aspects of images in image classification. The main aim of this paper is to study the impact of colour distortions on the performance of image classification using deep neural networks. Experiments performed using multiple state-of–of-the–the-art deep convolutional neural architectures on a proposed colour distorted dataset are presented and the impact of colour on image classification task is demonstrated.

1. Introduction

Over the years, deep convolutional neural networks have become indispensable for computer vision applications achieving human vision level performance. Currently, the state-of-the-art tool for researchers for image classification is the deep convolutional neural networks, where feature extraction and classification are combined and these networks are trained in an end-to-end manner. With the gaining popularity and deployment of trained computer vision models in day-to-day life and applications where safety is critical, robustness is one of the most important considerations for developing such a system. Multiple directions of research are undertaken to improve the performance and robustness of models. Some researchers work towards proposing new and robust architectures and others work on improving the scaling and training strategies to improve the performance and robustness of deep neural networks.

Krizhevsky et al. [33] proposed one of the first deep neural networks (AlexNet) for the Imagenet [9] competition 2012 and AlexNet outperformed the traditional techniques which involved a combination of hand-crafted features and classifiers. After the success of AlexNet, the following year, Zeiler and Fergus introduced the ZFNet [52] which showed the best performance in the Imagenet challenge in 2013. Lots of advancements and innovations in deep neural network architectures have made them achieve higher accuracy in image classification tasks. VGG-16 [43] proposed by Simonyan et al. and GoogleNet [44] proposed by Szegedy et al. were the top performers at Imagenet challenge 2014. This architecture introduced the concept of inception modules and 1×1 convolutions. The next architectural improvement in deep convolutional networks was the introduction of residual blocks in the Resnet architecture [20] and was the best performer in the Imagenet Challenge 2015.

Currently, researchers are actively exploring the effects of the quality of images on the performance of deep neural networks [10, 31, 11, 17, 3, 39, 54, 14]. However, the existing robustness bench-marking datasets do not have adequate colour-related distortions and this is one of the main motivations for our work. The original images mentioned in Imagenet-C have been used for our study and colour related subsets of these images have been generated for our analysis. To the best of our knowledge, very few studies have been conducted on the impact of quality of colour in images and the performance of deep neural networks on tasks like image classification. To enable further research on the impact of colour of digital images in deep neural networks, we propose a dataset of images having different colour transformations and colour distortions generated synthetically from a subset of the Imagenet Challenge dataset, which are available from Imagenet-C. Faults in colour imaging sensors, colour filters in data acquisition devices, gamut or post-processing filters are some natural sources where the colour of objects can be altered and thus the distribution of the images can be altered. For example, two images of the same scene with different colour gamuts will have different underlying distribution, and thus the distribution shift will have an impact on the performance of the trained deep learning model.

Colour information has been exploited in different computer vision tasks like image segmentation and object detection. The human vision system has a mechanism of perceiving and processing colour, but to the best of our knowledge, very little is known about how deep neural networks perceive colour. Although deep neural networks have been used for colourizing gray-scale images [53, 19, 36] still they have their own challenges. Recent work by Kantipudi et al. [30] has highlighted colour channel perturbation attacks on VGG, Resnet, and Densenet architectures and demonstrated the threat posed by colour information on the performance of convolutional neural networks. The current robustness benchmarking datasets like Imagenet-C, have tackled out-of distribution cases related to noise, blur, weather, cartoons, sketches, etc. Hendrycks et al. [24] have proposed two challenging datasets, Imagenet-A and Imagenet-O , which contain real-world unmodified natural adversarial examples where most of the deep neural networks fail. Imagenet-R [21] (Imagenet-Rendition) is a recent database containing art, cartoons, graffiti, sketches, etc. of Imagenet classes to create more out of distribution samples. To the best of our knowledge, there is no dataset dedicated to colour information and colour distortions to understand the behavior of deep neural networks. The main contributions of this paper include the creation of a dataset related to colour distortions and colour modifications to understand their impact on the task of image classification and then analyse the performance of state-of-the-art deep network architectures on image classification task of the dataset under different colour distortions and modifications based on the classification accuracy. The chosen deep convolutional network architectures have been few of the top performers in the Imagenet database and thus it is fair to further explore the response of these architectures using the weights pre-trained on the Imagenet database. The rest of the paper is organized as follows: Section 2 presents background information related to the existing literature, followed by Section 3 which provides details of the experiments conducted, followed by the results and findings in Section 4 and finally the conclusion and future directions are given in Section 5.

2. Background

The influence of the quality of images on the performance of deep neural networks has been investigated in the literature, and we give a brief overview of relevant existing research. The general hypothesis is that the images with distortions have a different distribution than the images on which the models are trained on. The shift in distribution reduces the performance of the deep learning models for distorted images. Dodge and Karam [10] checked the performance of four state-of-the-art deep neural network models (Caffe reference model [29], VGG-CNN-S [8], VGG16 [43], and GoogleNet [44]) for image classification under five types of quality distortions, namely blur, noise, contrast, JPEG, and JPEG2000 compression. Their test was on a subset of the validation set of the ImageNet 2012 [41]. The results indicate that the deep neural networks are influenced by distortions, especially noise and blur. Dodge and Karam [11] further investigated the performance of deep neural networks compared to humans on distorted images, and they found that the performance of the deep networks are lower on distorted images than humans, although they perform similar on high quality images.

Borkar and Karam [3] noticed that even small distortions could have an impact on image classification. They focused on Gaussian blur and additive noise, and proposed a ”correction” for deep neural networks to increase the performance in classification for distorted images. Ghosh et al. [14] also showed that Gaussian noise, blur, JPEG and JPEG2000 compression lowered the accuracy of deep neural network for image classification. They also investigated combined degradation, and found the same result. The authors also proposed a master-slave architecture to improve the performance. Zhou et al. [54] showed that the performance of deep neural networks was poorer when motion blur, defocus blur, Gaussian noise or all three distortions combined were added to images. They also showed that fine-tuning and re-training would improve the performance. Roy et al. [39] analyzed the performance of six different deep neural network architectures when influenced by Gaussian white noise, coloured Gaussian noise, salt & pepper noise, motion blur, Gaussian blur, and JPEG compression. The different architectures were influenced by the distortions, but at different degrees.

Impact of Colour Despite the work carried out to investigate the impact of image quality on deep neural networks, there is little work carried out on colour related distortions. Engilberge et al. [12] were first explore colour representation in deep neural networks where the authors discussed about colour sensitive units and hue specificity of the VGG-19 and Alexnet architecture. Gowda et al. [16] conducted a study of the Densenet architecture on different colour spaces and the accuracy of the deep neural network architecture on image classification datasets. Buhrmester et al. [6] have conducted deep experiments on the impact of colour and image quality on image classification tasks. The authors have run experiments on their own Person Finder dataset and publicly available Cifar-10 and Cifar-100 [32] datasets to understand the impact of colour on the task of image classification. The authors concluded that certain classes like wild animals (deer, fox, rabbit, beaver) and landscape and desert are highly dependent on colour information. These results give us the motivation to further explore the impact of colour and colour related distortions on image classification. Based on the work from Hendrycks et al. [22] we have also used the validation set of the Imagenet database as our base database and augmented different colour distorted images from these images. The details of the dataset generation are explained in Section 3.1.

Architectures The evolution of deep learning started with Alexnet. It combines convolutional layers, pooling layers, and activation functions. It was earlier the top performer on the Imagenet database. One of the significant architectures was VGG-19 which is a deep architecture with 47 layers, out of which 16 are convolutional and 3 are fully connected layers. An important characteristic of the VGG architecture is the use of three stacked single stride 3×3 convolutional layers which have the same receptive field as 7×7 convolutional layer. One of the biggest challenges in making the network deeper was that during the training phase the gradients would become smaller and smaller and lead to vanishing gradients. To tackle the vanishing gradient problem, one of the important innovations was the Resnet architecture which is still one of the widely used backbones in computer vision tasks. The Resnet architecture, which introduces the concept of skip connections which are used to fit the previous layer input to the next layer without modifying the output and enabling to have a deeper network (Res152 variant has 152 layers). This architecture uses residual block units where an individual block has two 3×3 convolutional blocks and these residual blocks are stacked. During the learning phase, the gradients can now flow into the previous layers via the skip connections, thus countering the vanishing gradient problem. One of the key innovations in CNN architectures is the inception module which computes 1×1, 3×3 and 5×5 convolutions within the same module of the network which helps in covering a larger area and at the same time preserving fine resolution for small information in the images. This module was the building block of the Googlenet architecture. Another significant architecture is the Densenet, which is based on connecting every layer directly with each other in a feed-forward manner, and needs fewer parameters in comparison to traditional convolutional neural networks by mitigating the need to learn redundant feature maps and ensuring maximum information flow. One of the drawbacks of deep convolutional networks is that they are computationally intensive and the models require a lot of memory which makes them unsuitable for mobile devices. For deployment in such devices a group of networks known as Mobilenets [42, 26] were proposed. Using the concepts of depth wise separable convolutions and inverted residuals, these networks give competitive performance for computer vision tasks in mobile devices.

Robustness Papernot [37] et al. have highlighted some limitations of deep learning in adversarial settings. Evaluating robustness is an extremely challenging and ongoing area of research. Carlini [7] have given some directions in this regard. Hendrycks et al. [22] attempted to benchmark deep neural network robustness to common corruptions and perturbations like additive noise, blur, compression artifacts, weather conditions, contrast, etc. They proposed a variant of Imagenet referred to as Imagenet-C and Imagenet-P. Imagenet-C contains 15 types of algorithmically generated corruptions with different levels of sensitivity. Recently, libraries like Foolbox [38] have been developed for generating adversarial examples for bench-marking machine learning examples. Studies have indicated that Imagenet-trained convolutional neural networks are biased towards texture and to improve robustness and accuracy, shape bias must be increased [13], but the impact of colour in images is not deeply studied. Robustness of deep neural networks has become an exceedingly important topic for the research community as studies have shown that convolutional neural networks can be deceived by visual illusions [15] and have garnered interest recently, like Yin et al. [51] who introduced a Fourier perspective on model robustness in computer vision. Taori et al, [47] conducted extensive experimentation to study the robustness of deep learning models on naturally occuring distribution shifts and concluded that synthetic interventions like diverse data augmentations offer robustness but more examples of data on naturally occurring distribution shifts improves robustness. Augmix [23] is a simple data processing technique which was proposed to improve the robustness of the deep learning model. Augmix consists of simple augmentation techniques like rotate, translate, posterize, etc. in combination with Jenson-Shannon divergence loss to enforce a common embedding for the classifier. Augmix does not use any form of colour or contrast augmentation to avoid manifold intrusion as previous studies from Guo [18] et al. have suggested that augmentation like histogram colour swapping can cause changes in class labels which lead to manifold intrusion.

Recent Advances Recent advances in deep learning include not only improvements in model architectures but also advances made in the ways the models are trained, data augmentation techniques, hyperparameter optimization. Recently, Bello et al. [2] showed that changing training and scaling strategies greatly improved Resnets. Generally, CNNs are designed in such a way to optimize the resources. Scaling up improves the model accuracy, but more resources are required. Tan and Le [45] conducted a detailed study on model scaling in network depth, width, and resolution to develop models with better performance and developed a family of models known as EfficientNets(B0-B7). The EfficientNet models were inspired from the MobileNet models and used the inverted residual blocks (also known as MBconv) in their design. Tan and Le demonstrated the effects of network depth, width, and resolution scaling individually and finally proposed a compound scaling mechanism for the designing of EfficientNet [45]. The family of Efficient models B0-B7 (B0 is the mobile size baseline and B7 has the highest resolution) performs competitively on the Imagenet dataset and the performance improves when the scale of the model and number of parameters increases. As a solution to certain challenges, smaller models with faster training were proposed by Tan and Le and this family of CNN models are called EfficientNetV2 [46]. Neural Architecture Search [34] (NAS) is used by researchers in machine learning to design and learn the network topology to achieve the best performance on a current task. Tan and Le combined the technique of training aware NAS and scaling to optimize parameter efficiency and training speed for EfficientNetV2 models. New operations like Fused-MBConv were used to search the models in the search space and an improved learning method which adaptively adjusted regularization along with image size known as progressive learning. One of the models which has shown robust performance on Imagenet-C is an approach based on knowledge distillation. Noisy student training [50] is a semi-supervised learning approach which uses the concepts of self-training and knowledge distillation using equal-or-larger student models and noise added to the student during the learning phase. It involves training a teacher model on labeled images and subsequently using the teacher to generate pseudo labels on unlabeled images and finally training a student model on the combination of labeled and pseudo-labeled images. The extra overhead in noisy student training is that in addition to Imagenet some extra unlabeled images are used for developing the model. Another approach developed to increase robustness is adversarial training using Adversarial prop [49]. This training procedure generates adversarial examples and treats them as additional examples and the main contribution is the use of a separate auxiliary batch norm for adversarial examples as the hypothesis is that the adversarial examples have a different underlying distribution than the normal samples.

Normalizer Free Networks The concept of batch normalization [28] has offered researchers certain advantages like efficient large-batch training, regularization effect eliminating the mean shift etc. Recent studies using signal propagation plots from Brock et al. [4] have identified certain issues of batch norm like the input is downscaled by a certain factor which is proportional to the standard deviation of the input and the variance of the signal is increased by the residual block by a certain factor and proposed a series normalizer free resnets. Brock et al.[5] then extended this concept to propose a family of normalizer free networks known as NF-Nets. The authors of NF-nets have modified the residual blocks and used convolutions with scaled weight standardization. For the training of NF-Nets, adaptive gradient clipping was used to restrict the magnitude of gradients to prevent exploding gradients and unstable training.

3. Experiments and Methodology

In this section, we discuss the details of the dataset generation for the experimental study and the deep neural network architectures used in our work. The biggest competition in image classification is the Imagenet Challenge [1] which published one of the most comprehensive image classification databases with 1000 categories is used for evaluating the performance of the image classification models. Convolutional neural network architectures are built up by a combination of stacking up different layers like convolutional layers, pooling layers, fully connected layers etc. The input to the convolutional neural network architecture is a 3-channel colour image with raw image pixel values. Generally, colour images are represented in a 3-dimensional array of in the format $M \times N \times 3$ where M and N are the number of rows and columns in the image and 3 are the number of channels. All colour spaces represent the images in this format. The RGB colour space has 3 channels containing red, green, and blue colour information. The CIELAB colour space represents colour information in three channels, namely L which contains lightness information from black to white, channel a which information from green to red and channel b which contains information from blue to yellow. The next colour space used in this study is the YCbCr colour space which also contains three channels where the Y channel is the luminance component and Cb and Cr channels are the blue-difference and red-difference chroma components.

3.1. Dataset Generation

Generally, the deep neural networks are trained on the comprehensive dataset designed for the Imagenet challenge containing 1000 classes and these models use all standard three Red-Green-Blue colour channels during the training phase. The proposed Imagenet-COLORDISTORT (Imagenet-CD) dataset is derived from the Imagenet dataset. 50 images from each of the 1000 classes are available as the validation set, out of which the images without colour channels are removed by setting the corresponding channel values to zero for our analysis, thus leaving a set of 49101 images varying from 35 to 50 images belonging to each of the classes.

[1] http://image-net.org/

RGB channel based distortions In the proposed dataset, to study the effect of RGB colour channels on image classification, six subsets were created by removing one or two colour channels, thus generating six colour casts: Cyan, Magenta, Yellow, Red, Blue and Green respectively as shown in Figure 1. These transformations were performed to study the impact of colour casts and the impact of individual colour channels without changing the scene and spatial information, and to analyse the behaviour of deep neural networks on these casts and understand the influence of colour channels on the classification accuracy of these modern deep convolutional neural network architectures.

Colour space based distortions In colour image processing, the colour of a digital image is represented using different mathematical models in the form of colour spaces. Hue-Saturation-Value (HSV) space [1], CIELAB space [25], and YCbCr are some of the most popular colour spaces. For detailed investigation, we have removed some of the colour channels in these colour spaces to generate different subsets of images. To study the colour aspects of images, only colour channels are modified to create the dataset and the intensity channels like Value in HSV space, L in CIELAB space, and Y in YCbCr space are not modified. The removal of saturation in the HSV colour space, by setting it to zero, results in loss of colour and the image becomes grayscale, but keeping the spatial information constant. Removing the hue component by setting the values of hue channel to 0 of the image results in moving the image components towards a reddish tinge. For experimental analysis, two separate subsets have been created from the CIELAB colour space, where the a and b channels have been set to zero in each of the respective subsets. Another popular colourspace is YCbCr, where Y denotes the luma component and Cb and Cr denote the blue-difference and red-difference chroma components, respectively. Two subsets of distorted images are augmented by removing the information from the Cb and Cr channels, thus resulting in images with yellow and cyan tinges. Examples of these distorted variants are shown in Figure 1.

Other distortions For transmission in limited bandwidth, images often need to be compressed or the number of colours in the image are reduced using colour quantization. In the proposed Imagenet-CD dataset, a subset of images were created where the number of colours in the image was reduced to 64 by using K-means clustering. In digital photography and displays, a colour gamut is a complete subset of colours which can be represented by the display device. To study the effect of colour gamut on image classification a subset of images was created using a smaller gamut (newspaper gamut-SNAP 2007 profile) [2] using an ICC profile. In addition, we have a subset of images where additive Gaussian noise is added in the Cb and Cr channels of the YCbCr colour space.

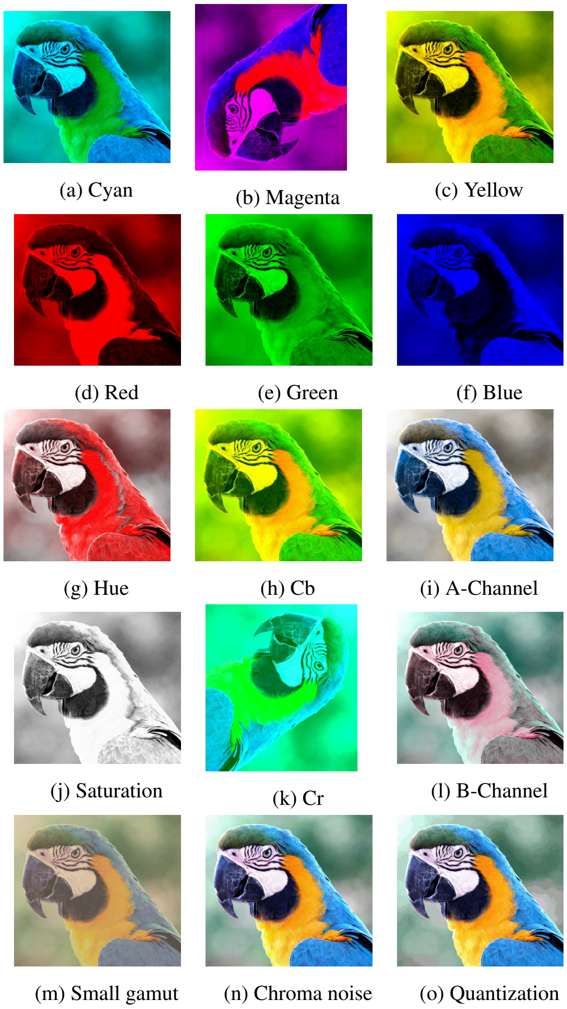
(a) Cyan
(b) Magenta
(c) Yellow
(d) Red
(e) Green
(f) Blue
(g) Hue
(h) Cb
(i) A-Channel
(j) Saturation
(k) Cr
(l) B-Channel
(m) Small gamut
(n) Chroma noise
(o) Quantization

Figure 1: Example of images from the Imagenet-CD dataset

4. Results and Discussion

In this section, we present the experimental results. The pretrained models from Torchvision 0.2.0 and the Pytorch Image Models packages [48] are used for the experiments.

4.1. Impact on widely used CNN architectures

In this section, we present the accuracy of the architectures mentioned in Table 1 on the generated Imagenet-CD dataset. To establish a reference, the first column (Original) represents the classification accuracy of the images in the absence of any perturbation. The Densenet and Resnet

[2] www.color.org

architectures show similar classification accuracy, and they perform better in comparison to the VGG-19 and Googlenet architectures. Even MobilenetV2 architecture has shown similar trends like the other heavier architectures, although the performance is on the poorer side. There is a significant loss in accuracy when one of RGB channels is removed (set to 0) from the images as shown in Table 1. One interesting observation is that in the case of blue colour cast generated by setting red and green channels to 0 has a very significant impact on the classification accuracy of the neural network architectures and Imagenet pretrained models perform very badly and specially the Alexnet architecture shows the worst classification performance and is not able to perform classification in the colour distorted dataset. One of the reasons for the poor accuracy of the images with blue cast is that most of the images were classified in underwater object categories like jellyfish etc or low light night scenes. This behavior was observed across all five architectures.

From Table 1 it can be inferred that colour channels from different colour models have different impact on the classification accuracy of the dataset. By removing the saturation channel (setting it to zero) from the HSV colour space, we remove the colour information and there is a decrease in the performance of the deep neural network based classifiers. This observation is consistent throughout all architectures. Removing the hue channel (setting it to zero) of the image essentially moves the objects in the image towards a reddish tinge, and since the deep neural network architectures are heavily inspired from the human visual system hence it is observed that the classifier performs misclassification and tends to favour the classes which have a reddish tinge like meat market, pomegranate, red wine etc. Similar behaviour is observed when the a and b colour channel information are removed (set to 0) in the CIELAB colour space. For compression related distortions like colour quantization, where the number of colours in all images are restricted to 64, a drop in classification accuracy is observed. Another interesting insight is that a smaller colour gamut has an impact on the way the deep neural networks perceive the image and results in a drop of accuracy due to the shift in distribution of the image. Presence of additive colour noise also reduces the performance of the classifiers.

4.2. Recent advances in efficient and robust models

In this Section, we demonstrate the Top-1 classification accuracy performance in more recent networks like EfficientNet and EfficientNet V2, training strategies like Adversarial propagation, noisy student (tested on Efficientnets v1) and augmentation strategies like Augmix (tested on Resnet-50). These networks and strategies have shown promising results on existing benchmarks like Imagenet-C.

4.2.1 Augmix and Resnet-RS-50

Here we present the impact of Augmix [23] data processing technique on the overall classification performance on Imagenet-CD dataset. We compare the performance of a pretrained Resnet-50 model against a pretrained Resnet-50 model with Augmix [3] and the comparison is shown in Fig. 2. Although for single channel information like blue channel, the performance is still poor but the model performance improves in comparison with the model trained without Augmix. Recent studies from Bello et al. [2] have suggested that efficient training and scaling of existing Resnet architectures can improve the model performance and developed a group of models known as Resnet-RS. We present that the classification performance of Resnet-RS-50 model on the colour distorted images from Imagenet-CD database and Top-1 accuracy for each group is tabulated in Fig. 2. We see a significant increase in robustness in comparison to the standard Resnet-50 model and the Resnet-50 trained with Augmix.

4.2.2 Scaling CNN, Adversarial Prop and Noisy Student training

Efficient Nets are based on compound scaling proposed by Tan and Le [45]. We tested the Imagenet-CD dataset on the EfficientNet models with different scales from the baseline B0 up to B7. The general observation from Table 2 is that the models perform better when they are scaled up and become increasingly robust when they are scaled up. The behavior of the baseline B0 model is slightly better in comparison to MobilenetV2 model shown in Table 1.The performance boost in the high resolution B7 model comes at the cost of it being a heavy model with a lot of parameters. The EfficientNetv2 is available in three different configurations-small (S), medium (M), and large (L) is also tested on the Imagenet-CD dataset and the results are presented in Table 2. These give competitive performance to high resolution models of EfficientNetv1 with considerably lighter models. An important observation is that Efficientnet models trained with adversarial propagation [49] and noisy student training [50] performed better on the distorted images in comparison with the normal EfficientNet models with the noisy student trained models showing the best performance almost across all resolutions.

4.2.3 Normalizer Free networks

Normalizer Free Resnets and Normalizer Free Networks (NF-Nets) have been recently proposed. These models do not use Batch normalization. In Fig. 3 we have compared the pretrained NF-Res-50 model with the standard Res-50 model and observed that there is a significant increase in

[3] pretrained weight from website of the author

Table 1: Classification Top-1 accuracy (%) of well-known CNN architectures on the Imagenet-CD dataset

Network	Original	Gamut	Chroma	Quant	Cyan	Magenta	Yellow	Red	Blue	Green	Hue	Saturation	A	B	Cb	Cr
Densenet [27]	77.1	73.4	75.7	75.3	60.7	49.1	68.6	41.9	18.6	48.6	65.7	66.1	71.0	68.6	64.2	45.5
Resnet [20]	78.3	71.6	76.8	76.5	60.6	48.6	67.0	43.3	18.6	44.4	67.6	67.4	72.0	69.8	57.9	39.4
VGG19 [43]	72.4	63.2	70.1	70.2	47.5	25.2	56.2	27.2	9.6	32.5	59.7	57.3	65.5	63.6	48.2	35.2
GoogleNet [44]	69.8	65.7	68.6	68.5	52.6	52.5	60.3	44.7	22.6	51.6	65.7	58.9	67.8	65.8	57.3	41.6
MobileNet [42]	71.9	62.2	69.6	69.3	49.0	33.2	55.6	28.6	9.9	32.8	59.5	56.4	64.6	61.9	40.2	25.7
Alexnet [33]	56.7	38.4	54.6	54.7	12.8	4.2	23.3	4.4	0.9	4.6	35.6	32.6	44.3	41.5	11.8	4.15

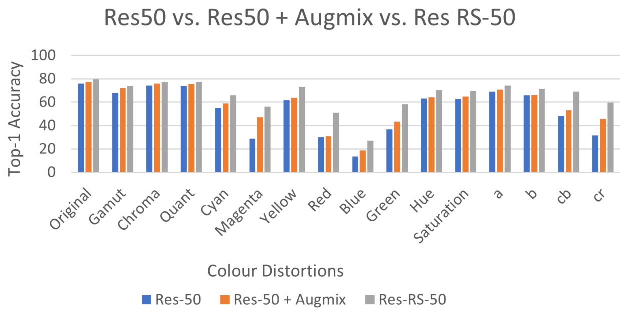

Figure 2: Performance of Res-50 vs. Res-50 + Augmix vs. Res-RS-50

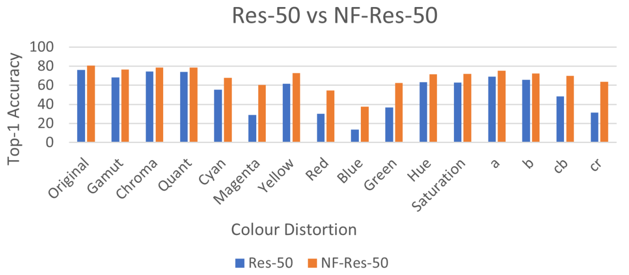

Figure 3: Performance of Res-50 vs. NF-RS-50 models

Top-1 accuracy for distorted images in comparison to the standard Res-50 model. In Table 2 we have also shown the performance of pretrained NF-Net models on Imagenet-CD images for different resolutions (from baseline F0 to F5) and a similar trend to EfficientNet is observed where the higher resolution images give better accuracy. Even the baseline F0 model performs better than all competitive architectures and is more robust to distortions.

5. Conclusion and Future Work

With deep neural networks being deployed in commercial and safety critical systems, one of the main focus of future research is to make these models more robust and accurate to changes. Experimental studies conducted in this paper have yielded some interesting results with respect to the impact of colour information of images on the performance of deep neural network architectures due to the shift in distribution. The performance of these networks drastically reduces when information from one or two colour channels in the RGB colour space is removed. Modifications in the hue and saturation components of an image have a strong impact on the internal working mechanism of the deep neural network and this trend is observed across all different colour spaces included in this study like the CIELAB and YCbCr colour spaces. In this paper, we have presented the overall classification behavior of the widely used convolutional neural network architectures under different colour distortions and the interesting results demonstrated will serve as a motivation to investigate the colour sensitivity of individual architectures in detail in the future. The analysis tabulated in this paper will motivate researchers to take into consideration the impact of colour channels and aspects of digital colour images and different colour spaces for proposing more accurate and robust systems based on deep convolutional neural networks. The important observations are listed as

- There is a significant impact of colour information on the inference of deep neural networks.
- Data processing techniques like Augmix have some impact on robustness and optimizing the training procedure, diverse augmentations and optimizing hyper parameters increases the robustness as Resnet RS-50 is much more robust compared to Resnet-50 models.
- Higher resolution models are much more robust as the same trend is observed in both EfficientNets (B7 has much higher accuracy than B0, V2L performs better than V2S) and NF-Nets.
- Training procedures like adversarial prop and noisy student training offer some amount of additional robustness to models.
- The Normalizer free models offer more robustness to colour specific distortions.

Table 2: Classification Top-1 accuracy (%) of Imagenet-CD for Efficient Net and Normalizer free network models

Network	Original	Gamut	Chroma	Quant	Cyan	Magenta	Yellow	Red	Blue	Green	Hue	Saturation	A	B	Cb	Cr
B0 [45]	76.8	70.6	75.2	74.2	63.3	51.1	68.7	40.6	18.0	51.3	66.5	66.5	70.6	68.3	62.4	45.3
B0 + ap [49]	77	72.9	76.3	76	64.1	55.7	70	42.1	28.8	54.1	69.2	67.9	72.4	70.6	64.7	50.7
B0 + ns [50]	78.6	73.5	77.4	76.3	65.3	57.6	68.9	45.6	31.2	52.1	68.2	67.1	71.9	69.3	59.4	47.6
B1 [45]	78.8	74.1	77.5	76.9	68.7	58.5	71.8	47.6	23.1	59.1	70.1	69.7	74	72	65.6	54.8
B1 + ap [49]	79.2	75.6	78.4	78.1	69.7	52.8	72.5	47.7	25.5	59.2	71.6	71.3	75.1	73.3	67.9	58.6
B1 + ns [50]	81.4	77	79.8	79.1	70.4	65.1	73.9	54.5	38.1	60.3	72.2	71.1	75.7	74.1	66.6	60
B2 [45]	80	75.1	78.5	77.7	70	60.5	73.7	50.5	28.3	59.9	71	71.1	75.1	73	67.9	56.8
B2 + ap [49]	80.2	76.4	79.3	78.9	71.4	63.5	75.2	53.2	31.9	61.6	73.4	73.1	76.4	75	71.7	60
B2 + ns [50]	82.4	78.3	80.8	80.3	72.6	65.7	75.9	57.3	41.7	62.9	73.4	73	77.3	75.2	68.7	62
B3 [45]	81.6	77.2	80.1	79.5	73.4	65.9	76.4	60.3	33	66.7	74.4	73.9	77.2	75.2	72.7	66.1
B3 + ap [49]	81.8	78.5	80.9	80.8	74.9	64.6	77.3	60	38.8	66.2	76.5	75.9	78.9	77.8	74.5	67.9
B3 + ns [50]	84	80.4	82.8	82.1	76.2	70.3	78.3	61.6	47.6	67.1	76.2	75.6	79.5	77.9	72.6	69.9
B4 [45]	83	79	81.1	80.7	75.1	70.2	78.4	62.3	41.5	68.6	76.1	76.1	78.9	77.3	75.8	69.1
B4 + ap [49]	83.2	80.5	82.3	82.1	77.9	73.3	80	65.7	48.6	71.8	78.4	78.3	80.7	79.4	78.1	72.3
B4 + ns [50]	85.1	81.5	83.4	82.6	77.7	75.9	79.8	68.9	58	72.8	78.2	78.3	80.9	79.7	77	73.9
B5 [45]	83.7	78.9	81.6	81.7	77.1	74.4	80.4	68.6	54.8	74.3	77.8	77.4	80	78.6	78.6	71.1
B5 + ap [49]	84.2	81.6	83.5	83.1	79	75.4	80.9	67.3	47.4	73	79.9	80	81.9	81	78.7	74.4
B5 + ns [50]	86	82.9	84.4	83.8	80.5	78.5	81.6	74.7	65.4	77.6	80	79.9	82	81.6	79.5	76.3
B6 [45]	84	80.5	82.6	82.8	77.5	71.4	80.4	63.8	36.6	70.8	78.3	77.7	80.8	79.2	77.8	70
B6 + ap [49]	84.7	82	83.8	83.6	79.9	75.9	81.7	69.4	50.4	75.7	80.7	80.7	82.5	81.9	80.1	77.2
B6 + ns [50]	86.4	84.1	85	85	82.4	80	84.1	76.3	65.2	79.6	82.3	81.7	83.7	83.3	83.1	79.8
B7 [45]	84.8	81.3	83.3	82.9	79.8	76.1	81.9	71.6	61.6	76.1	79.9	79.7	81.6	80.5	80.2	75.1
B7 + ap [49]	85	82.5	84.4	84	80.4	75.7	82.1	69.5	48.7	75.4	80.9	81	82.9	81.9	80.1	76.2
B7 + ns [50]	86.8	83.5	85.6	85	82.7	79.1	83.9	74.9	63.2	79.1	82.3	80.9	83.8	82.9	81.5	78.9
V2-S [46]	83.8	79.9	82.1	81.8	76.6	72.7	80	65.6	45.3	73	77.3	76.8	80	78.3	78	71.8
V2-M [46]	85	81.6	83.5	83.1	78.1	74.3	80.9	67.6	46.6	75.2	78.8	78.9	81.4	79.7	79	73.5
V2-L [46]	85.4	82.6	84.2	83.6	77.9	74.9	81.3	70.2	47.1	74.6	80.1	79.6	82.3	80.5	79.3	71.4
NF F0 [5]	83.2	79.9	81.7	81.2	75.4	68.5	77.7	60.3	43.7	67.5	76	75.9	78.9	77.4	75.3	69.5
NF F1 [5]	84.5	81.2	83	82.4	78.4	74.2	80.9	69.5	52.8	73.5	78.2	78.6	80.3	78.8	79.7	75.4
NF F2 [5]	84.9	81.2	82.9	82.2	77.4	75.1	80.6	69.2	56	71.4	78.5	78.6	80.3	79.1	79.5	74.3
NF F3 [5]	85.5	82.3	83.8	83	79.7	77.8	81.8	71.5	60.7	76.1	79.5	80.2	81.4	80.3	80.9	76.9
NF F4 [5]	85.6	82.3	83.9	83.3	80.7	78.5	82.7	73.2	58.9	77	80.5	80.5	81.9	81	81.9	79.4
NF F5 [5]	85.6	82.4	83.4	83.1	79.9	78.5	81.8	72.4	60.8	76.8	79.5	80.8	81.6	80.8	80.8	76.9

The internal working of the deep neural networks with respect to colour sensitivity is still a black box for the research community. In future, it will be interesting to study the effect of minor colour changes for the classification of images and formulate a robustness criterion with respect to colour sensitivity of deep neural networks. One future direction of research can be to come up with new data augmentation techniques with respect to colour variations, which may make these models more accurate and robust. The influence of colour on a computer vision system is generally application specific, a plant disease detection system must be sensitive to minor colour variation, on the other hand, an application dealing with human faces must not be sensitive to colour because it may be a feature which will compromise with the fairness of the system. The knowledge on the behaviour of colour information on deep learning will aid researchers working towards fair, transparent, and robust learning models to come up with more secure to adversarial attacks [35], robust to image corruptions [40] and fair models for deployment.

References

[1] M. Agoston. *Computer Graphics and Geometric Modeling: Implementation and Algorithms*. Springer, 2005. 5

[2] I Bello, W Fedus, X Du, E D Cubuk, A Srinivas, Tsung-Yi Lin, J Shlens, and B Zoph. Revisiting resnets: Improved training and scaling strategies. *arXiv preprint arXiv:2103.07579*, 2021. 3, 6

[3] T Borkar and L Karam. Deepcorrect: Correcting dnn models against image distortions. *IEEE Transactions on Image Processing*, 28(12):6022–6034, 2019. 1, 2

[4] A Brock, S De, and S Smith. Characterizing signal propagation to close the performance gap in unnormalized resnets. *arXiv preprint arXiv:2101.08692*, 2021. 4

[5] A Brock, S De, S Smith, and K Simonyan. High-performance large-scale image recognition without normalization. *arXiv preprint arXiv:2102.06171*, 2021. 4, 8

[6] V Buhrmester, D Münch, D Bulatov, and M Arens. Evaluating the impact of color information in deep neural networks. In *Iberian Conf. on Pattern Recognition and Image Analysis*, pages 302–316. Springer, 2019. 2

[7] N Carlini and D Wagner. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*, pages 39–57. IEEE, 2017. 3

[8] K Chatfield, K Simonyan, A Vedaldi, and A Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014. 2

[9] J Deng, W Dong, R Socher, L Li, K Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conf. on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1

[10] S Dodge and L Karam. Understanding how image quality affects deep neural networks. In *2016 eighth Intl. conf. on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2016. 1, 2

[11] S Dodge and L Karam. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th Intl. conf. on computer communication and networks (ICCCN)*, pages 1–7. IEEE, 2017. 1, 2

[12] M Engilberge, E Collins, and S Süsstrunk. Color representation in deep neural networks. In *2017 IEEE Intl. Conf. on Image Processing (ICIP)*, pages 2786–2790. IEEE, 2017. 2

[13] R Geirhos, P Rubisch, C Michaelis, M Bethge, F A Wichmann, and W Brendel. Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *Proc. of the Intl. Conf. on Learning Representations*, 2019. 3

[14] S Ghosh, R Shet, P Amon, A Hutter, and A Kaup. Robustness of deep convolutional neural networks for image degradations. In *2018 IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2916–2920. IEEE, 2018. 1, 2

[15] A Gomez-Villa, A Martín, J Vazquez-Corral, and M Bertalmío. Convolutional neural networks can be deceived by visual illusions. In *Proc. of the IEEE conf. on Computer Vision and Pattern Recognition*, pages 12309–12317, 2019. 3

[16] S N Gowda and C Yuan. Colornet: Investigating the importance of color spaces for image classification. In *Asian Conf. on Computer Vision*, pages 581–596. Springer, 2018. 2

[17] K Grm, V Štruc, A Artiges, M Caron, and H K Ekenel. Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biometrics*, 7(1):81–89, 2017. 1

[18] H Guo, Y Mao, and R Zhang. Mixup as locally linear out-of-manifold regularization. In *Proc. of the AAAI Conf. on Artificial Intelligence*, volume 33, pages 3714–3722, 2019. 3

[19] S Halder, K De, and P Roy. Perceptual conditional generative adversarial networks for end-to-end image colourization. In *Asian Conf. on Computer Vision*, pages 269–283. Springer, 2018. 2

[20] K He, X Zhang, S Ren, and J Sun. Deep residual learning for image recognition. In *Proc. of the IEEE conf. on computer vision and pattern recognition*, pages 770–778, 2016. 1, 7

[21] D Hendrycks, S Basart, N Mu, S Kadavath, F Wang, E Dorundo, R Desai, T Zhu, S Parajuli, M Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. *arXiv preprint arXiv:2006.16241*, 2020. 2

[22] D Hendrycks and T Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *Proc. of the Intl. Conf. on Learning Representations*, 2019. 3

[23] D Hendrycks, N Mu, ED Cubuk, B Zoph, J Gilmer, and B Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. In *Intl. Conf. on Learning Representations*, 2019. 3, 6

[24] D Hendrycks, K Zhao, S Basart, J Steinhardt, and D Song. Natural adversarial examples. *arXiv preprint arXiv:1907.07174*, 2019. 2

[25] G Hoffmann. Cie color space. *online http://www. fho-emden. de/hoffmann/-ciexyz29082000. pdf*, 2000. 5

[26] A Howard, M Sandler, G Chu, L Chen, B Chen, M Tan, W Wang, Y Zhu, R Pang, V Vasudevan, et al. Searching for mobilenetv3. In *Proc. of the IEEE/CVF Intl. Conf. on Computer Vision*, pages 1314–1324, 2019. 3

[27] G Huang, Z Liu, L Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proc. of the IEEE conf. on computer vision and pattern recognition*, pages 4700–4708, 2017. 7

[28] S Ioffe and C Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Inl. conf. on machine learning*, pages 448–456. PMLR, 2015. 4

[29] Y Jia, E Shelhamer, J Donahue, S Karayev, J Long, R Girshick, S Guadarrama, and T Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proc. of the 22nd ACM Intl. conf. on Multimedia*, pages 675–678, 2014. 2

[30] J Kantipudi, S Dubey, and S Chakraborty. Color channel perturbation attacks for fooling convolutional neural networks and a defense against such attacks. *IEEE Transactions on Artificial Intelligence*, 2020. 2

[31] S Karahan, M K Yildirum, K Kirtac, F S Rende, G Butun, and H K Ekenel. How image degradations affect deep cnn-based face recognition? In *2016 Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2016. 1

[32] A Krizhevsky, V Nair, and G Hinton. Cifar-10 and cifar-100 datasets. *URl: https://www. cs. toronto. edu/kriz/cifar. html*, 6, 2009. 2

[33] A Krizhevsky, I Sutskever, and G Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1, 7

[34] C Liu, B Zoph, M Neumann, J Shlens, W Hua, L Li, L Fei-Fei, A Yuille, J Huang, and K Murphy. Progressive neural architecture search. In *Proc. of the European conf. on computer vision (ECCV)*, pages 19–34, 2018. 4

[35] A Madry, A Makelov, L Schmidt, D Tsipras, and A Vladu. Towards deep learning models resistant to adversarial attacks. In *Intl. Conf. on Learning Representations*, 2018. 8

[36] K Nazeri, E Ng, and M Ebrahimi. Image colorization using generative adversarial networks. In *Intl. conf. on articulated motion and deformable objects*, pages 85–94. Springer, 2018. 2

[37] N Papernot, P McDaniel, S Jha, M Fredrikson, Z B Celik, and A Swami. The limitations of deep learning in adversarial settings. In *2016 IEEE European symposium on security and privacy (EuroS&P)*, pages 372–387. IEEE, 2016. 3

[38] J. Rauber, R. Zimmermann, M. Bethge, and W. Brendel. Foolbox native: Fast adversarial attacks to benchmark the robustness of machine learning models in pytorch, tensorflow, and jax. *Journal of Open Source Software*, 5(53):2607, Sep 2020. 3

[39] P Roy, S Ghosh, S Bhattacharya, and U Pal. Effects of degradations on deep neural network architectures. *arXiv preprint arXiv:1807.10108*, 2018. 1, 2

[40] E Rusak, L Schott, R Zimmermann, J Bitterwolf, O Bringmann, M Bethge, and W Brendel. A simple way to make neural networks robust against diverse image corruptions. In *European Conference on Computer Vision*, pages 53–69. Springer, 2020. 8

[41] O Russakovsky, J Deng, H Su, J Krause, S Satheesh, S Ma, Z Huang, A Karpathy, A Khosla, M Bernstein, et al. Imagenet large scale visual recognition challenge. *Intl. journal of computer vision*, 115(3):211–252, 2015. 2

[42] M Sandler, A Howard, M Zhu, A Zhmoginov, and L Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proc. of the IEEE conf. on computer vision and pattern recognition*, pages 4510–4520, 2018. 3, 7

[43] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1, 2, 7

[44] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, and A Rabinovich. Going deeper with convolutions. In *Proc. of the IEEE conf. on computer vision and pattern recognition*, pages 1–9, 2015. 1, 2, 7

[45] M Tan and Q Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Intl. Conf. on Machine Learning*, pages 6105–6114. PMLR, 2019. 3, 4, 6, 8

[46] M Tan and Q Le. Efficientnetv2: Smaller models and faster training. *arXiv preprint arXiv:2104.00298*, 2021. 4, 8

[47] R Taori, A Dave, V Shankar, N Carlini, B Recht, and L Schmidt. When robustness doesn't promote robustness: Synthetic vs. natural distribution shifts on imagenet. 2019. 3

[48] Ross Wightman. Pytorch image models. https://github.com/rwightman/pytorch-image-models, 2019. 5

[49] C Xie, M Tan, B Gong, J Wang, A Yuille, and Q Le. Adversarial examples improve image recognition. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pages 819–828, 2020. 4, 6, 8

[50] Q Xie, M Luong, E Hovy, and Q Le. Self-training with noisy student improves imagenet classification. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pages 10687–10698, 2020. 4, 6, 8

[51] D Yin, R G Lopes, J Shlens, E D Cubuk, and J Gilmer. A fourier perspective on model robustness in computer vision. In *Advances in Neural Information Processing Systems*, pages 13255–13265, 2019. 3

[52] M Zeiler and R Fergus. Visualizing and understanding convolutional networks. In *European conf. on computer vision*, pages 818–833. Springer, 2014. 1

[53] R Zhang, P Isola, and A Efros. Colorful image colorization. In *European conf. on computer vision*, pages 649–666. Springer, 2016. 2

[54] Y Zhou, S Song, and N Cheung. On classification of distorted images with deep convolutional neural networks. In *2017 IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1213–1217. IEEE, 2017. 1, 2