# Handling Colors in Image Classification

Soumya Snigdha Kundu SRM Institute of Science and Technology
Sheel Patel SRM Institute of Science and Technology
Shaswat Srivastava SRM Institute of Science and Technology
Aarsh Chaube SRM Institute of Science and Technology
Vaishnavi Moorthy SRM Institute of Science and Technology
Email: {sk7610, ss6233, sp8429, ac5973, vaishnam}@srmist.edu.in

## Abstract

The presence of bias in a dataset has been a long-standing bottleneck in the task of image classification. While supervised methods have been shown to overcome these biases, self-supervised methods have managed to overcome benchmarks set by supervised learning methods. This paper shows that self-supervised methods can maintain their ability to outperform supervised methods even when introduced to color bias. Two experimentation pipelines are presented. One focuses on the capability of a model to handle artificially induced color bias and the other gauges the ability of a model to incorporate naturally occurring color differences present in vision datasets.

## 1 Introduction

Supervised learning methods have been shown to perform with great efficiency in image classification tasks [1] [2]. However, as the tasks have become more complex and begin to involve various forms of bias [3] [4] within them, it becomes a major challenge for supervised approaches to overcome these biases. Therefore, supervised learning-based approaches have begun to face a major bottleneck towards real-world computer vision problems. Recently, self-supervised methods have not only shown to better-supervised learning methods but also are domain invariant [5] [6]. Self-supervised learning obtains supervisory signals from the data itself, often leveraging the underlying structure in the data. The general technique of self-supervised learning is to predict any unobserved or hidden part (or property) of the input from any observed or non-hidden part of the input. The proposed experimentation utilizes these advantages to handle color bias present in vision data for two image classification tasks.

## 2 Experimentation

### 2.1 Experiment A: Overcoming Artificial Color Bias

In this task, the ability of a model to overcome artificially induced color bias is tested. The bias setting is adapted from the pipeline set by the authors in [7] as the amount of bias introduced to the data can be controlled. The bias is introduced in the MNIST dataset [8]. The experimentation pipeline begins by first training the model on black and white digits and then testing the trained model on colored digits. The supervised model, ResNet-18, was trained on the black and white digit data and tested on the colored digit data. In the case of self-supervised learning, the pretext task training took place on the black and white digits data and the task of classifying the colored data was set as the Downstream Task. The training set consisted of 39,000 black and white digits [8] and the testing set consisted of 10,000 images of colored digits from [7]. The lower and upper limit of the variance was chosen for the dataset bias. The lower limit is 0.02 (highest level of bias) and the upper limit is 0.05 (lowest level of bias). The training of the self-supervised learning model was done through a ResNet-18 [9] backbone and momentum contrastive learning [10]. The loss function utilized for the architecture is the effective contrastive loss InfoNCE [11]. To account for the variation in the downstream task an additional linear layer was added to the architecture.

### 2.2 Experiment B: Dealing with Natural Color Bias

This task tests the ability of a model to overcome naturally present color bias in data. To simulate this the training pipeline is set to classify dogs and cats with distinct contrast in their skin color [7]. The task is a 4-class classification problem containing the following class labels: Light Dog, Light Cat, Dark Dog, and Dark Cat.

*Table 1:* Testing Scores of both A and B Experiment Variations

| **Model (Train & Test) - A** | **Data Variance** | **Test Accuracy** |
|---|---|---|
| Proposed Self Supervised Network | 0.05 | **95.2%** |
| ResNet [9] (Supervised) | 0.05 | 61% |
| Proposed Self Supervised Network | 0.02 | **95.3%** |
| ResNet [9] (Supervised) | 0.02 | 50% |
| **Model (Train & Test) - B** | **Test F1-Score** | |
| Proposed Self Supervised Network | **86.2** | |
| ResNet - 18 [9] (Supervised) | 75.75 | |
| ResNet - 50 [9] (Supervised) | 73.25 | |

There were a total of 12135 images in the training dataset and 3035 images in the testing dataset.

### 2.3 Setting of Experiments

Both the supervised and self-supervised models used for this experiment were the same as that of in section 2.1 and were presented with the same train and test data. For the testing of the self-supervised network, all the labels of the data were removed. The experimentation was done on the Kaggle GPUs using PyTorch and lightly [12]. There was no augmentation or pre-processing done to any of the datasets. Fig 1 visually depicts the experimentation pipeline. One should note that the knowledge transfer occurs separately at separate instances of the experiment.

## 3 Results and their Analysis

The results in table 1 clearly depict the superiority of the self-supervised learning approach. In the case of Experiment A (2.1), not only does the self-supervision improve on the supervised method by 45% accuracy, but it also is completely invariant to the bias present in the data. The change in the amount of bias has no effect on self-supervised methods. Even though the variance of induced bias increased from 0.05 to 0.02, the accuracy remained exactly the same. Whereas the supervised method performs quite poorly and continues to degrade as the amount of bias increases in the data. In terms of Experiment B (2.2), self-supervision has a clear advantage in terms of performance. It achieves an F1-Score of 86.2% whereas the supervised method only manages to achieve an accuracy of 75.75%.

The results indicate that self-supervised methods learn appropriate representations which are important to the final objective rather than accounting for features that may confuse the architecture. The ability to tackle induced bias and incorporate naturally present color differences.

## 4 Conclusion

This work presents the capabilities of self-supervised learning methods to handle color in a task of image classification. In particular, the ability of self-supervised learning models to overcome artificially induced colored bias and leverage naturally occurring color differences in vision datasets is demonstrated. The self-supervised methods completely outperform the supervised method in both the experimentation tasks. In the future, we hope to expand the experiments to account for a change in the domain of training and testing data [13] [14] and to see whether class imbalance has any drastic effect when compounded with bias [15].

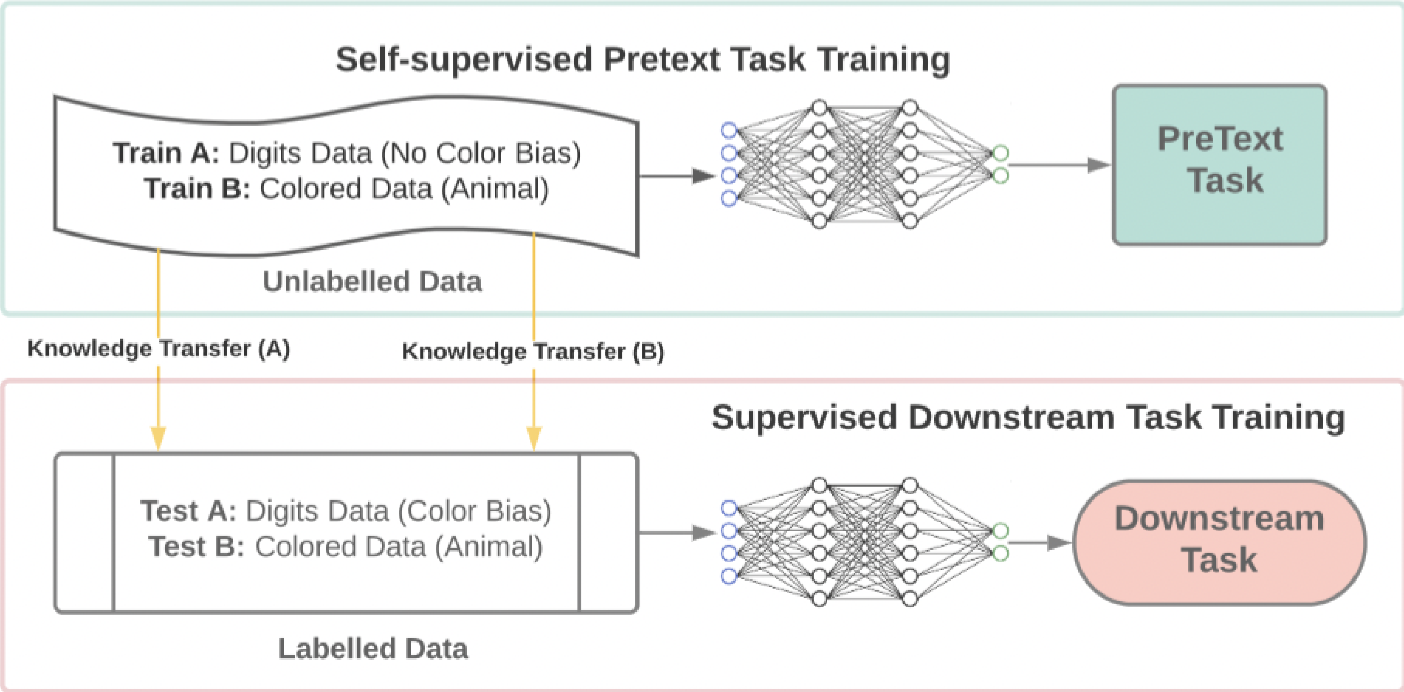


*Fig. 1:* Proposed Experimentation Pipeline

## Acknowledgments

We thank the Data Science Community of SRM for assembling the research team.

## References

[1] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[2] D. Tsipras, S. Santurkar, L. Engstrom, A. Ilyas, and A. Madry, "From imagenet to image classification: Contextualizing progress on benchmarks," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9625–9635.

[3] N. Schaaf, O. de Mitri, H. B. Kim, A. Windberger, and M. F. Huber, "Towards measuring bias in image classification," *arXiv preprint arXiv:2107.00360*, 2021.

[4] I. Serna, A. Peña, A. Morales, and J. Fierrez, "Insidebias: Measuring bias in deep networks and application to face gender biometrics," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 3720–3727.

[5] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[6] K. Ohri and M. Kumar, "Review on self-supervised image recognition using deep neural networks," *Knowledge-Based Systems*, vol. 224, p. 107090, 2021.

[7] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, "Learning not to learn: Training deep neural networks with biased data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9012–9020.

[8] Y. LeCun, "The mnist database of handwritten digits," *http://yann. lecun. com/exdb/mnist/*, 1998.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.

[11] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[12] e. a. Igor Susmelj, "Lightly," *GitHub. Note: https://github.com/lightly-ai/lightly*, 2020.

[13] R. Siry, L. Hémadou, L. Simon, and F. Jurie, "On the inductive biases of deep domain adaptation," *arXiv preprint arXiv:2109.07920*, 2021.

[14] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," *arXiv preprint arXiv:1906.12340*, 2019.

[15] H. Liu, J. Z. HaoChen, A. Gaidon, and T. Ma, "Self-supervised learning is more robust to dataset imbalance," *arXiv preprint arXiv:2110.05025*, 2021.

## Appendix: Code

The code for this paper can be found at: https://github.com/sheel1206/Handling-Color