ColorSense: A Study on Color Vision in Machine Visual Recognition

Ming-Chang Chiu
USC
mingchac@usc.edu

Yingfei Wang
USC
yingfei@usc.edu

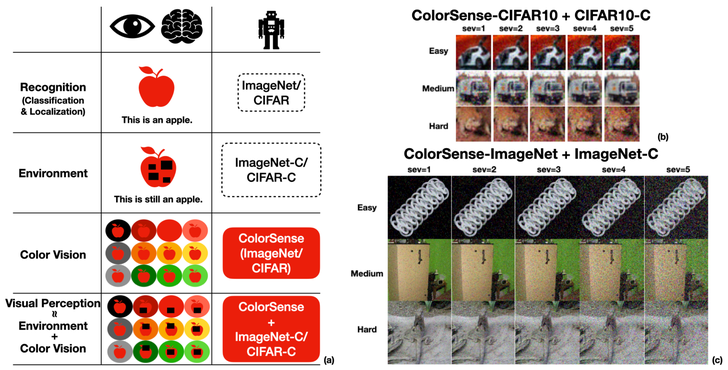
Derrick Eui Gyu Kim
USC
euigyuki@usc.edu

Pin-Yu Chen
IBM Research
pin-yu.chen@ibm.com

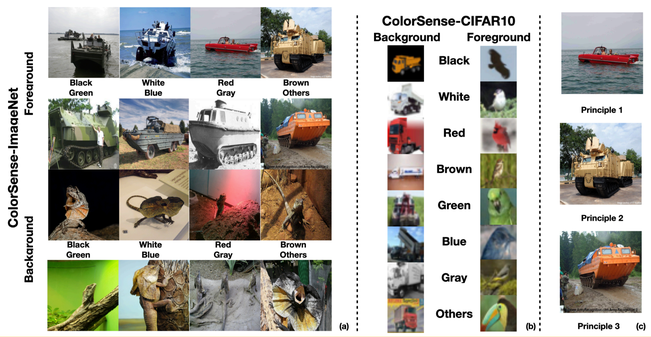
Xuezhe Ma
USC
xuezhema@usc.edu

Introduction

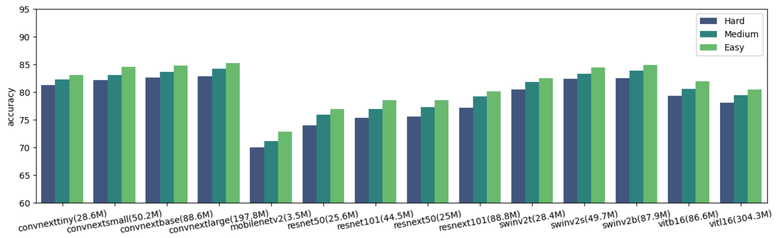
In this work, we explored how color vision impacts machine perception by presenting COLORSENSE, a curated dataset with 110,000 human annotations of color labels derived from benchmarks like ImageNet and CIFAR10. We performed thorough evaluation on cutting-edge models with consideration of key factors such as model architecture, training objective, model size, training data and task complexity. We jointly analyzed the impact of color vision and image corruption. Our findings suggest that object recognition tasks are susceptible to color vision bias and environmental factors, and we highlight the need for new approaches toward the performance evaluation of machine perception models in real-world applications.



- COLORSENSE Dataset
  - COLORSENSE-IMAGENET and COLORSENSE-CIFAR10
  - Images categorized based on dominant foreground/background color and their color discrimination levels
- Model evaluation framework
  - Integrates COLORSENSE with ImageNet-C to study robustness under noisy conditions
  - Metrics to quantify the impact of color vision on performance
- Applications
  - Study spurious correlation in machine learning models
  - Enhancing fairness and robustness in safety-crucial tasks



Impact of Color Vision on Machine Perception

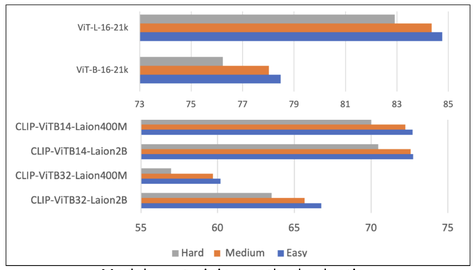


DNNs are deeply affected by color vision:

- consistent performance decrease from the Hard CD group to the Easy group for all models
- Model size and architecture matters but do not significantly improve robustness to color vision

Absolute Gap	MobileNet	MobileNet_adv	Resnet	Resnet_adv	Resnext	Resnext_adv
ColorSense	2.86	2.62	3.12	2.21	2.97	2.56
Land vehicle-only	4.35	4.87	6.59	4.27	4.63	5.87
ColorSense-C	3.37	n/a	4.40	n/a	4.46	n/a
Absolute Gap	**Convnext**		**SwinV2**		**ViT**	
ColorSense	2.17		2.18		2.48	
Land vehicle-only	3.40		4.31		6.93	
ColorSense-C	4.59		4.40		4.64	

Advanced training setup evluation and safety-critical task evaluation

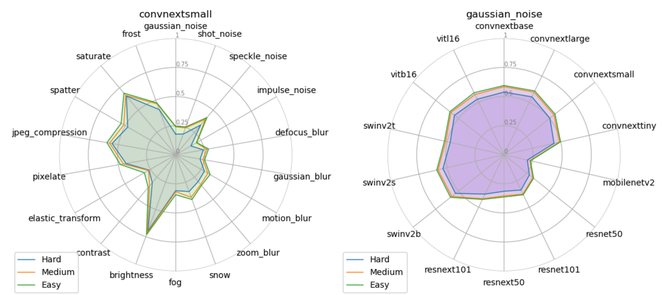


Model pre-training method evluation

CD Group	ViT-S/16	ViT-S/8	ViT-B/16	ViT-B/8
Easy	77.90	80.58	78.78	80.90
Medium	76.45	79.25	77.71	79.41
Hard	75.11	77.79	76.40	78.60

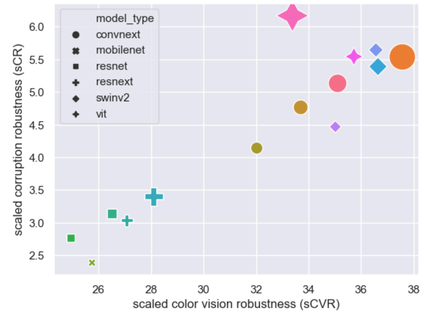
DINO on ImageNet

- Advanced training setups marginally improve color vision bias
- High stake examples have larger color vision bias
- Consistent pattern for self-supervised model such as DINO and zero-shot learning foundation models such as CLIP



- Models of the same total performance can have different color vision behaviours in the presence of different environment factors

Metrics beyond overall model accuracy: Quantifying model robustness to color vision effect



scaled color vision robustness

$$sCVR = \frac{acc_{tot}}{\sigma_w}$$

acc_{tot} = overall accuracy, and

$$\sigma_w = \sqrt{\frac{\sum_{i=Easy}^{Hard} w_i (acc_i - \bar{acc})^2}{\sum_{i=Easy}^{Hard} w_i}}$$

scaled corruption robustness

$$sCR = \frac{acc_{tot}}{\sigma_s}$$

$$\sigma_s = \sqrt{\frac{\sum_{s=1}^{5} (acc_s - \bar{acc})^2}{\sum_{s=1}^{5} w_s}}$$

- sCVR and sCR can be generalized to measure variability of model performance between subgroups

Potential root causes of color vision bias

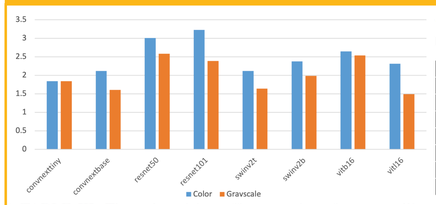


TABLE IV: High & low frequency components accuracy (%). Results reported in `Resnet50` / `ViT_L` format.

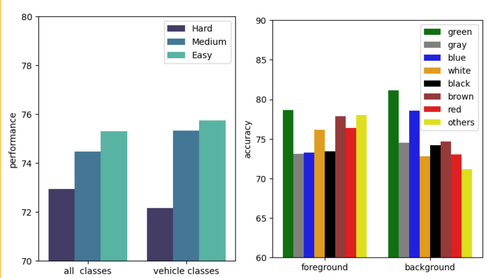
LFC	Easy	Medium	Hard
$r = 8$	4.04 / 9.64	3.44 / 9.12	3.23 / 8.04
$r = 16$	21.47 / 36.49	19.27 / 34.60	17.81 / 31.39
$r = 32$	51.71 / 64.43	49.01 / 62.52	47.48 / 60.61
HFC	**Easy**	**Medium**	**Hard**
$r = 8$	44.97 / 51.08	44.62 / 51.61	45.52 / 52.19
$r = 16$	13.99 / 9.58	14.65 / 10.20	15.00 / 11.23
$r = 32$	2.01 / 1.82	1.85 / 1.64	2.57 / 2.67

TABLE III: The color vision effect remains after controlling luma. Results reported in Hard/Medium/Easy CD group format.

Luma range	(-0.2,0.2)	(-0.4,0.4)
Resnet50	73.59/77.05/77.91	73.77/76.51/77.47
Resnet101	74.87/77.60/79.75	75.20/77.46/79.09
ConvnextT	80.83/82.66/83.17	81.14/82.54/83.21
ConvnextB	82.32/84.20/85.21	82.47/83.75/85.06
VitB16	79.30/80.89/82.34	79.33/80.85/82.20
VitL16	78.91/79.65/80.85	78.37/79.83/80.93
SwinV2T	80.09/82.16/82.95	80.08/82.19/82.85
SwinV2B	82.22/84.31/85.11	82.65/84.07/84.97

- **Chroma**: Less color vision bias on greyscale images
- **Luma**: Bias persists after controlling for luma
- **Frequency components**: bias driven by LFC in images

Other utilities of COLORSENSE



CD Group	IoU-30	IoU-50	IoU-70
Easy	75.31	73.12	50.34
Medium	74.48	72.77	50.01
Hard	72.94	70.80	48.97

- Models show similar color vision bias for object localization task
- High stake examples
- Background and foreground colors as supirous correlations