DDI-COCO: A DATASET FOR UNDERSTANDING THE EFFECT OF COLOR CONTRAST IN MACHINE-ASSISTED SKIN DISEASE DETECTION

Ming-Chang Chiu*[1], Yingfei Wang*[1], Jen-Ju Kuo[2], Pin-Yu Chen[3]

IBM Research

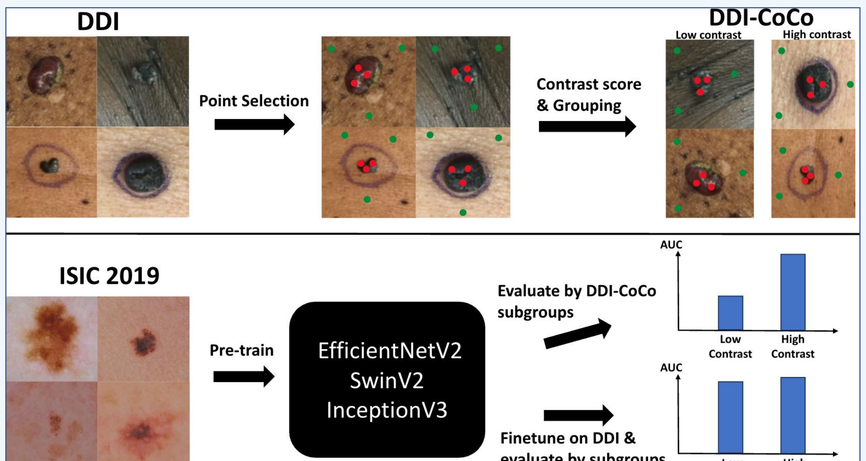
[1] University of Southern California [2] NTU Hospital [3] IBM Research

DDI-ColorContrast (DDI-CoCo) Dataset

- Study lesion-skin tone color contrast effect on machine assisted skin disease malignancy detection
- Annotated dataset with contrast level labels derived from pixel RGB values
- Based on Diverse Dermatology Image (DDI) Dataset

Dermatologist-reviewed and variability-controlled labeling procedure: 3 steps

1. Retrieve RGB values on lesion areas
2. Retrieve RGB values around lesion areas
3. select points with similar lighting conditions
4. Take Avg and use the WCAG formula for contrast score:
   - L = 0.2126 * R + 0.7152 * G + 0.0722 * B
   - Contrast ratio = (L1 + 0.05) / (L2 + 0.05), where L1 is derived using the lighter one in foreground or background colors, and L2 is from the darker one.



DDI-CoCo	I-II	III-IV	V-VI	Total
High	88	134	92	314
Low	114	103	111	328

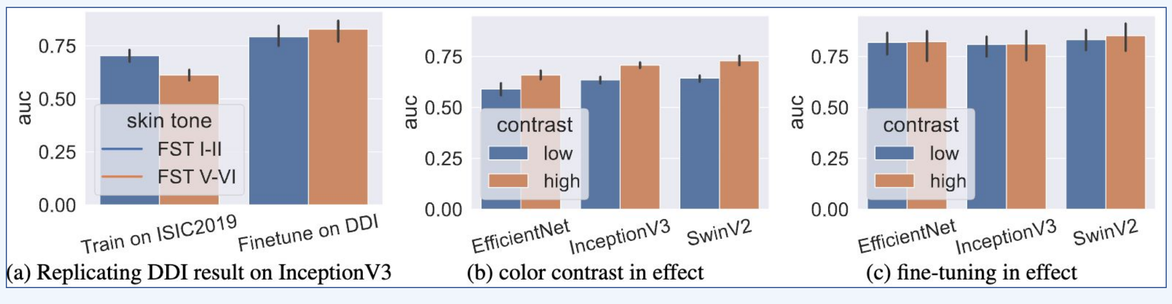
Table 1: Number of images by contrast & skin tones.

Experiments

2 Objectives: (1) to evaluate if DNNs are subject to contrast bias, similar to the skin tone bias observed in the DDI paper. (2) to assess the efficacy of DDI-CoCo in reducing the performance gap between high and low contrast groups.

2 Setups: (1) we train on the comprehensive and publicly available ISIC 2019 dataset and evaluate on the entire DDI-CoCo dataset. This OOD setup is practical as we do not always have ID data in clinical evaluations. (2) To test if fine-tuning on DDI can reduce color contrast bias, we fine-tune on 80% of DDI-CoCo and evaluate on the rest.

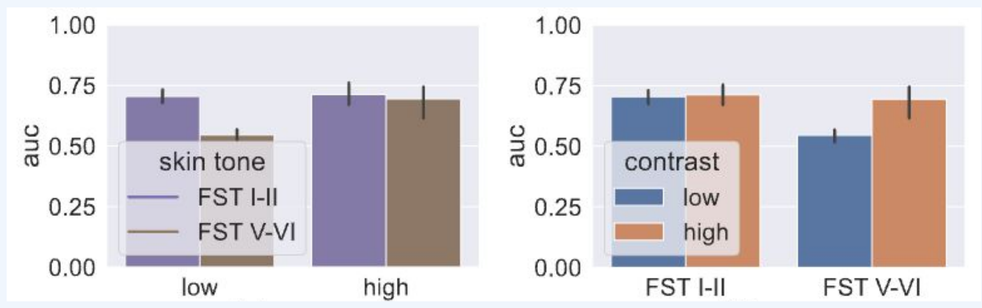
3 Models: InceptionV3, EfficientNet, Swin TransformerV2 (SwinV2)



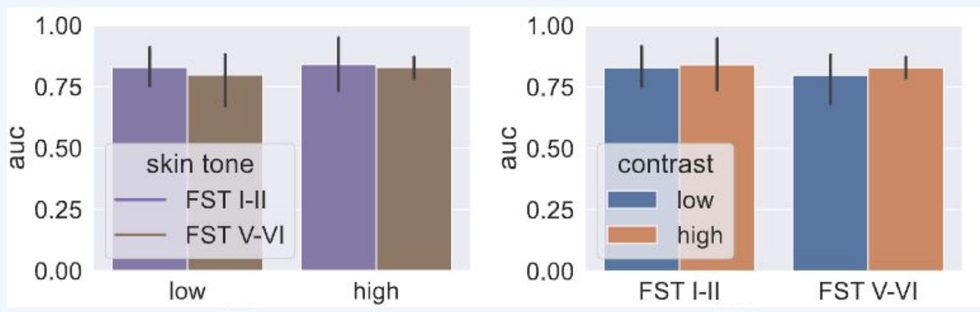
(a) Replicating DDI result on InceptionV3 (b) color contrast in effect (c) fine-tuning in effect

Results

Skin tone bias is more pronounced in the low contrast group, contrast bias is more pronounced for darker skin tones



Fine-tuning closes the performance gap between low & high contrast groups



AUC	Inception	Efficient	SwinV2
High	0.810	0.822	0.852
Low	0.809	0.819	0.832

Table 2: Average AUC by contrast after fine-tuning.

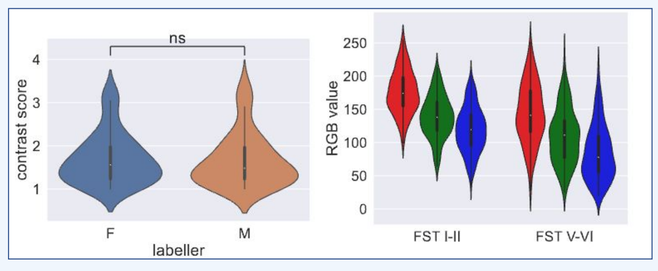
Similar Behavior in SoTA (SwinV2, EfficientNet) models

AUC	InceptionV3	EfficientNet	SwinV2
I-II	0.792	0.776	0.814
V-VI	0.828	0.862	0.895
Gap	+0.036	+0.086	+0.081

Table 3: Average AUC by skin tones after fine-tuning.

Labeling Process Justification

- The contrast score distribution is consistent across labellers, despite the distinction in pixel selection
- Selected background pixels have lower RGB values for the darker skin tone group



Conclusion

- Systematic investigation focused on the impact of color contrast on skin disease malignancy detection
- DDI-CoCo Benchmarking
- Robust labeling method
- Fine-tuning on diverse image dataset helps mitigate contrast bias