(De)Constructing Bias on Skin Lesion Datasets

A. Bissoto[1], M. Fornaciali[2], E. Valle[2], S. Avila[1]

[1]RECOD Lab., IC, University of Campinas (UNICAMP)

[2]RECOD Lab., DCA, FEEC, University of Campinas (UNICAMP)

ISIC Workshop @ CVPR 2019

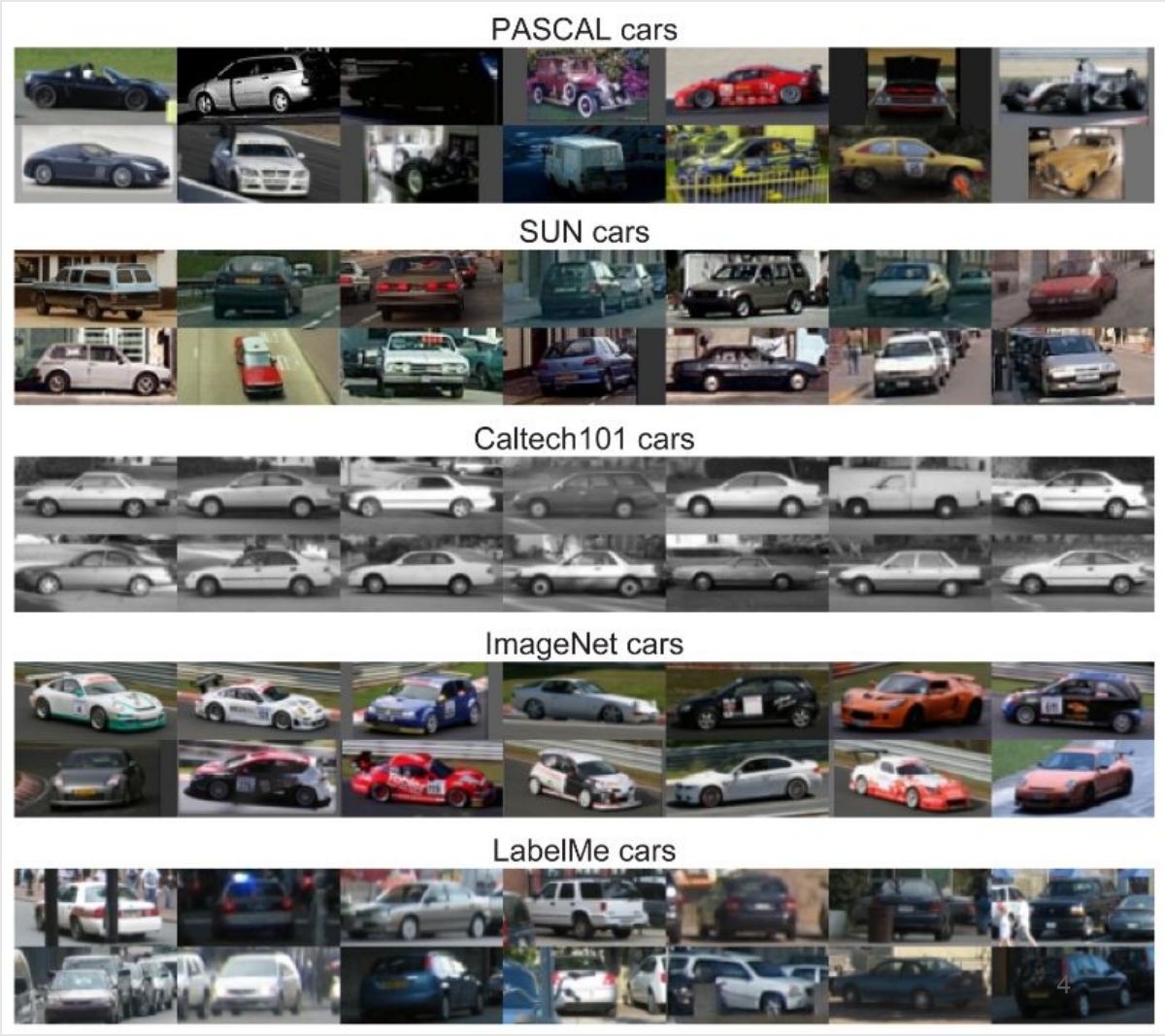
RECOD Titans

melanoma research

5 years 2014–2019

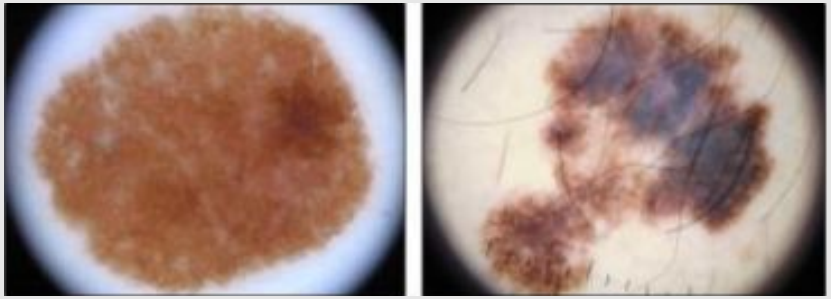
http://www.todayifoundout.com/index.php/2013/12/anti-tank-dogs-world-war-ii/

Bias

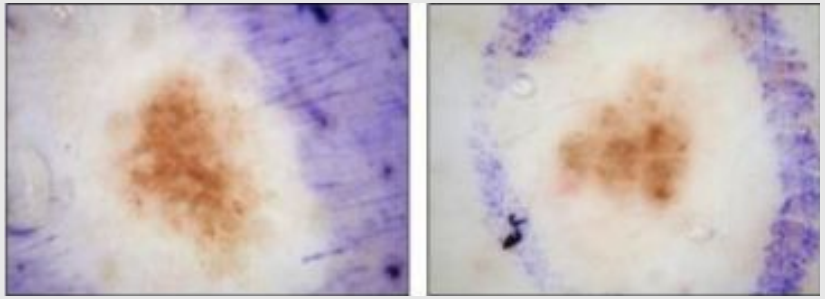


Reproduced from: “Unbiased Look at Dataset Bias”, Torralba et al. (2011)

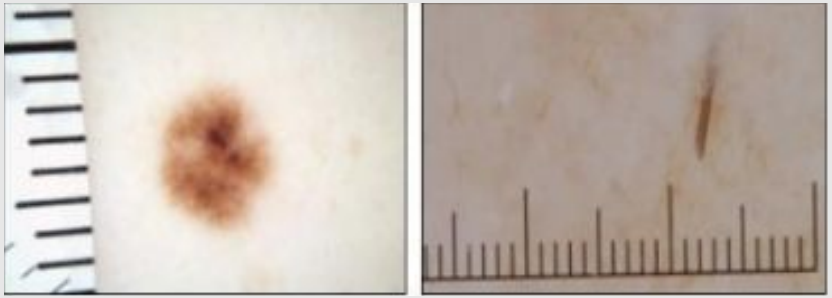
Confounders on Skin Lesion Datasets



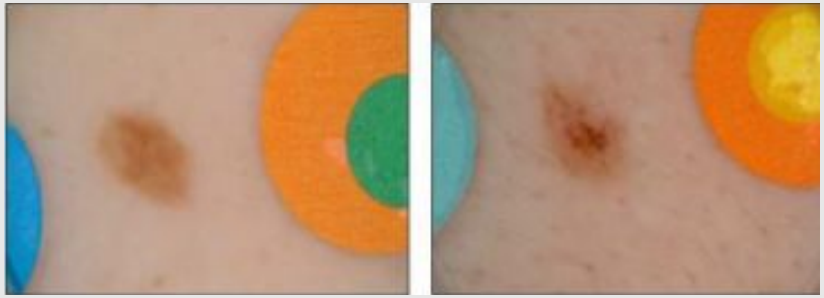
Vignetting (dark borders)



Staining



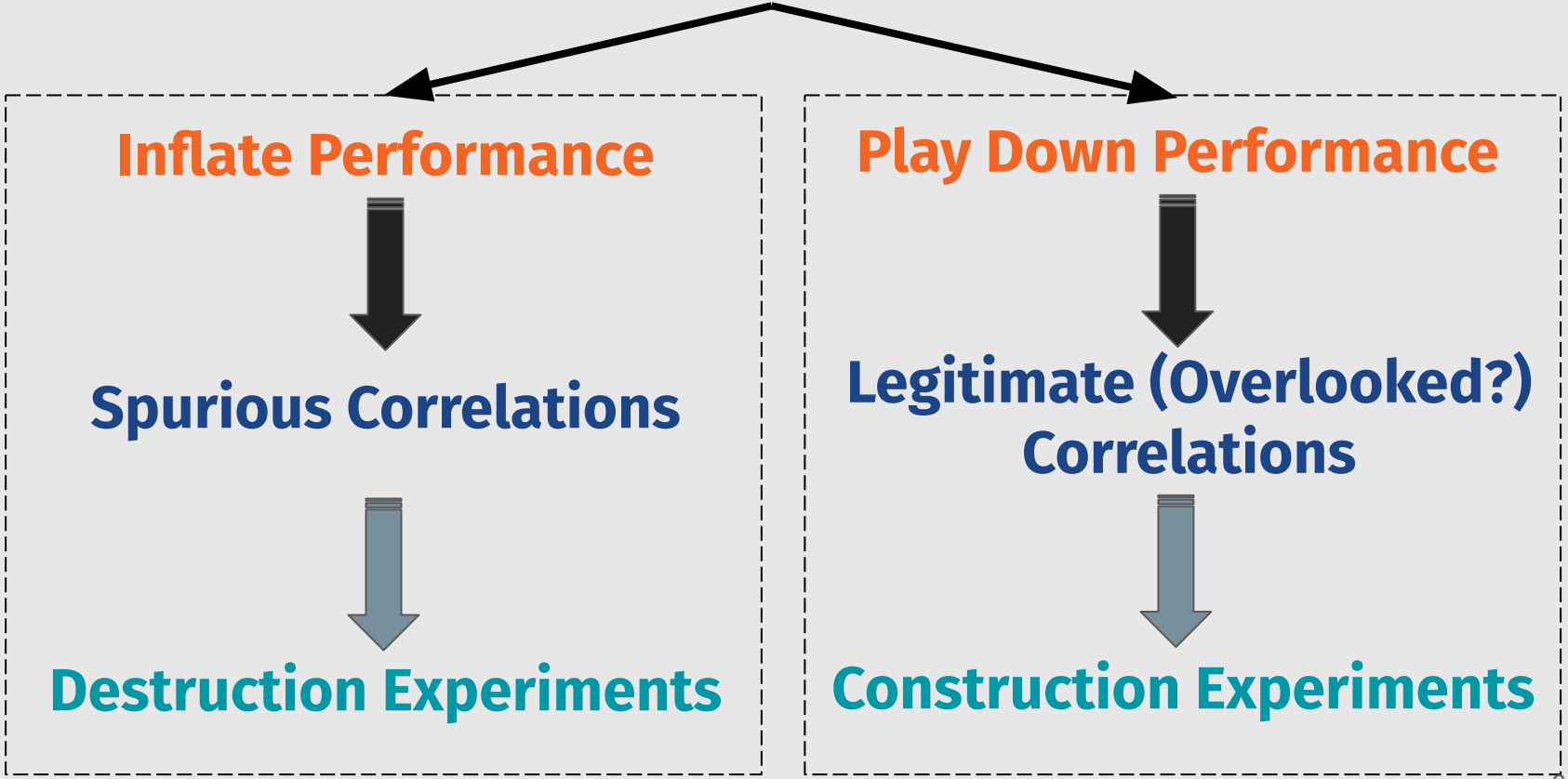
Rulers



Color markers

Reproduced from: “An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning”, Mishra et al. (2016)

Bias


Inflate Performance
Spurious Correlations
Destruction Experiments
Play Down Performance
Legitimate (Overlooked?) Correlations
Construction Experiments

Datasets

Atlas of Dermoscopy

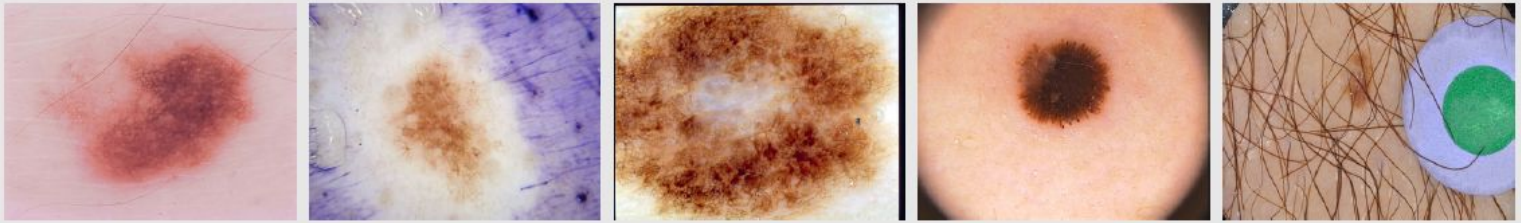
- ➔ **Educational**
- ➔ **Rich Metadata**
- ➔ Clinical and dermoscopic images for every case
- ➔ Clinical data (location, diameter, elevation)
- ➔ Metadata for dermoscopic features.

ISIC Archive

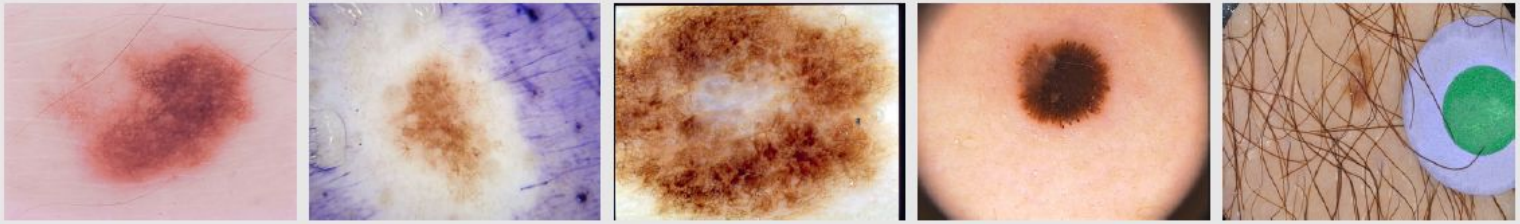
- ➔ **Large**
- ➔ **Diverse**
- ➔ Different sources, different devices
- ➔ Segmentation masks for lesion (large subset)
- ➔ Segmentation masks for dermoscopic features (small subset).

Destruction Experiments

Traditional

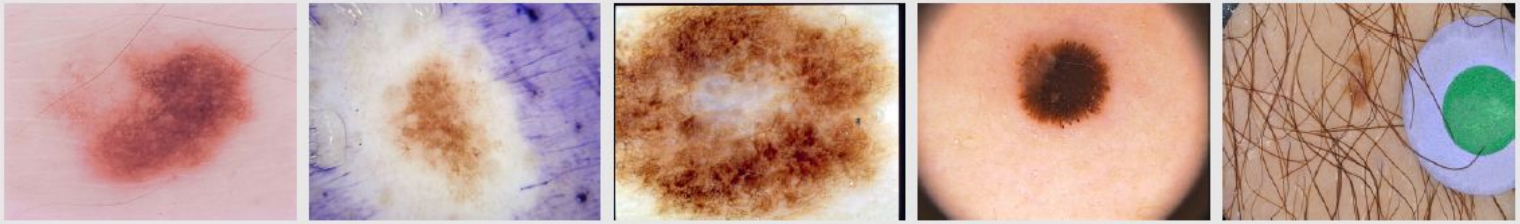


Traditional



Only Skin

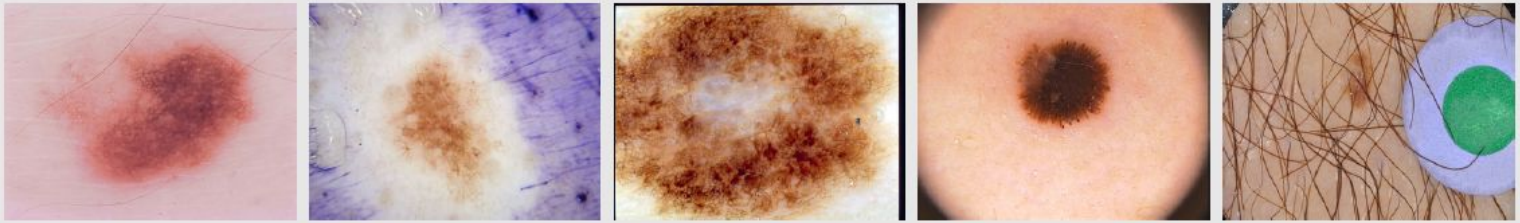
Traditional



Only Skin

Bbox

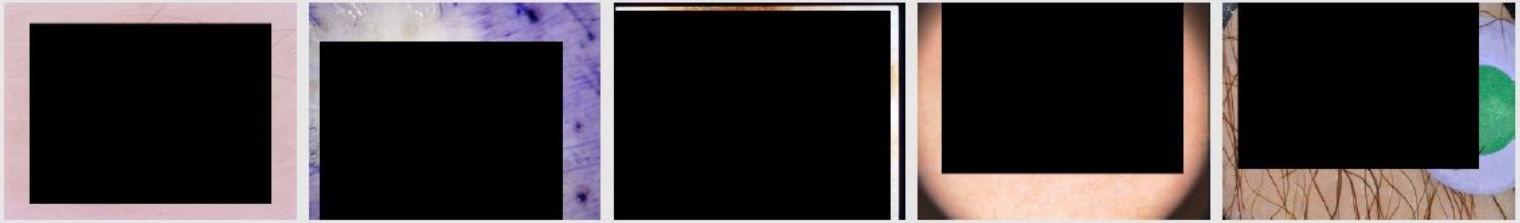
Traditional



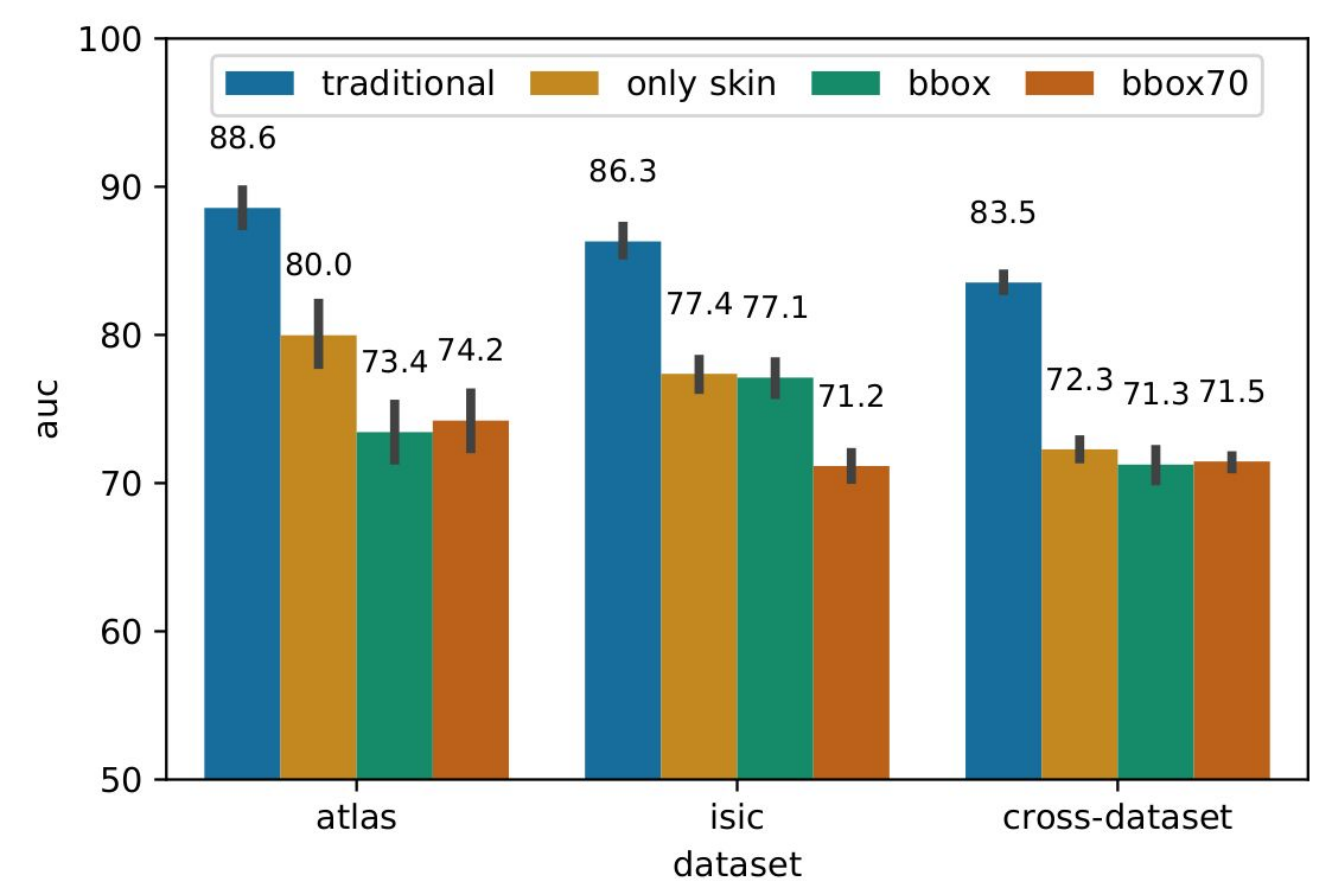
Only skin

Bbox

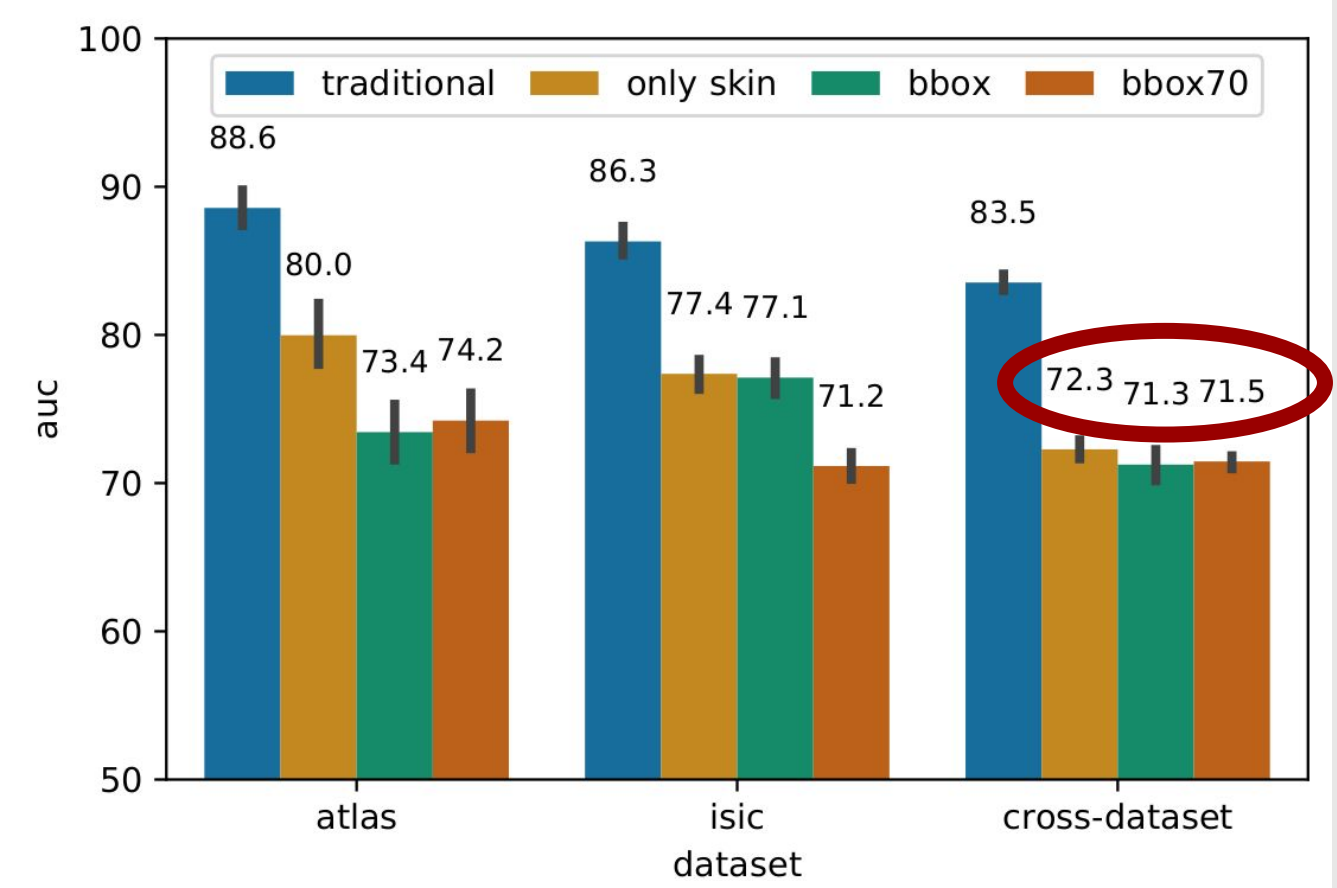
Bbox70



Destruction Experiments



Destruction Experiments



Destruction Experiments

Performance of **machine learning** with all **cogent information removed** on ISIC Archive: **71% AUC**

Destruction Experiments

Performance of **machine learning** with all **cogent information removed** on ISIC Archive: **71% AUC**

Performance of **157 dermatologists**[1] on ISIC Archive: **67% AUC**

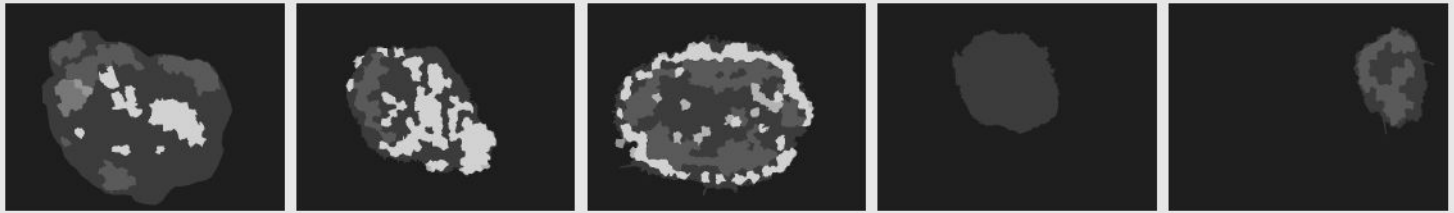
[1] "The Melanoma Classification Benchmark", Brinker et al. (2019)

Construction Experiments

Traditional

Traditional

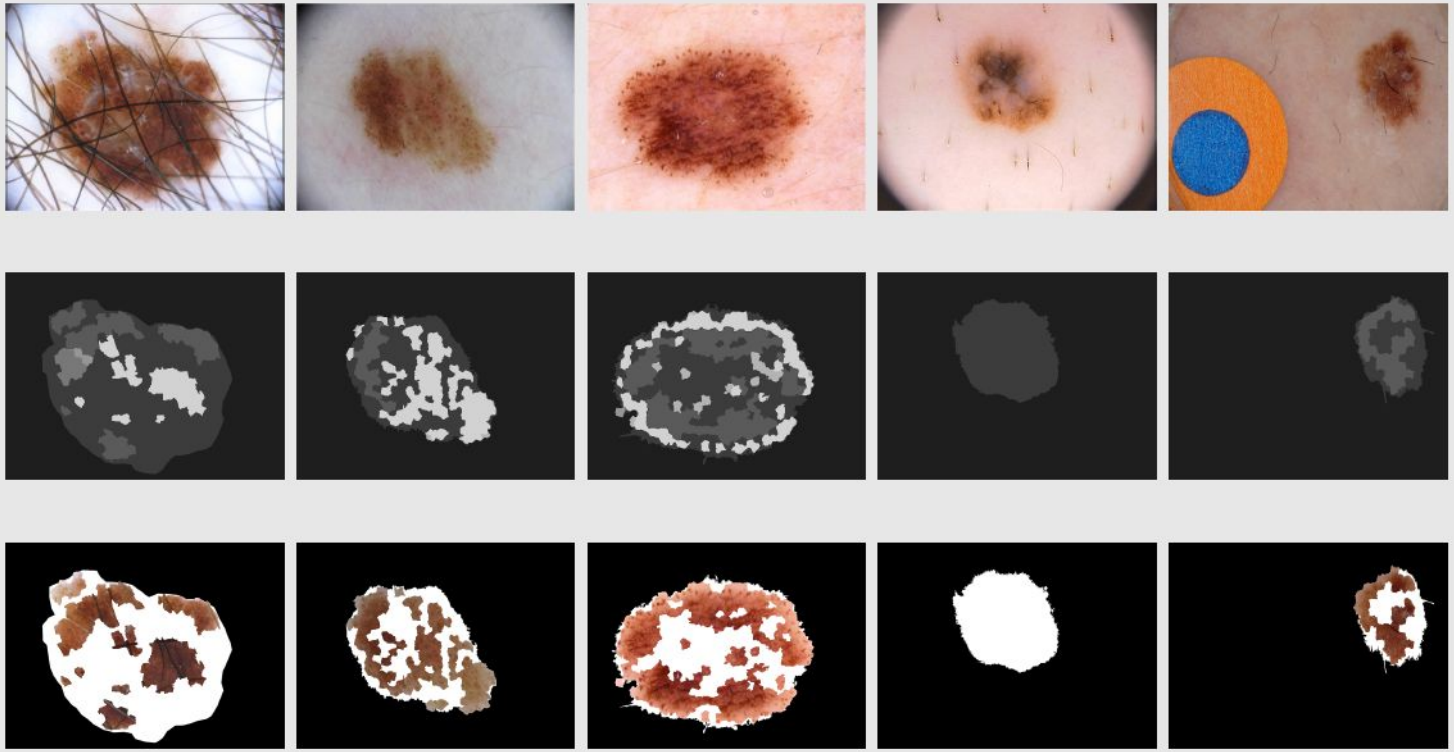
Grayscale Attributes



Traditional

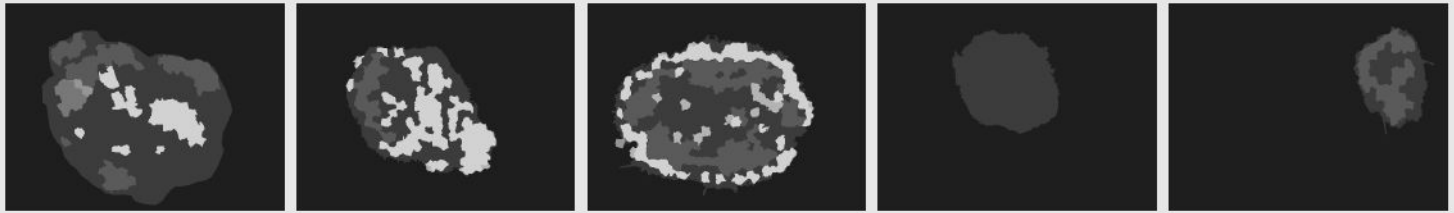
Grayscale Attributes

RGB Attributes

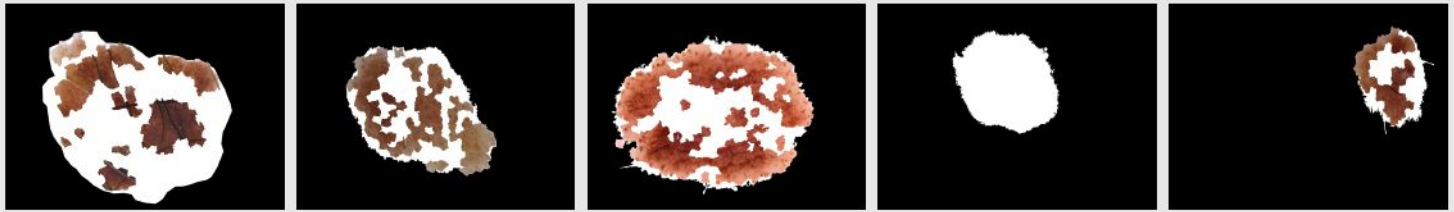


Traditional

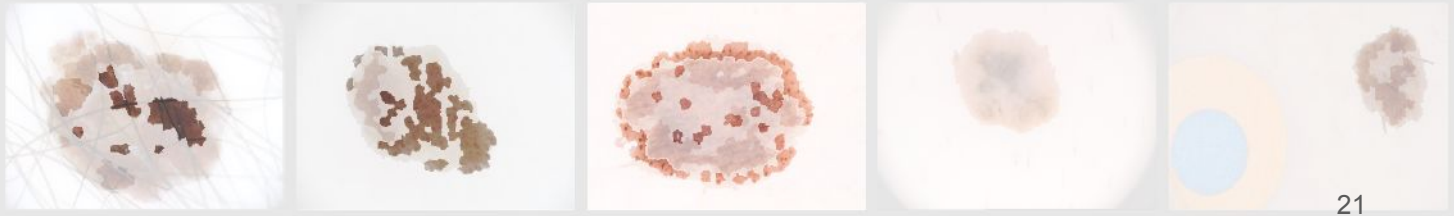
Grayscale Attributes



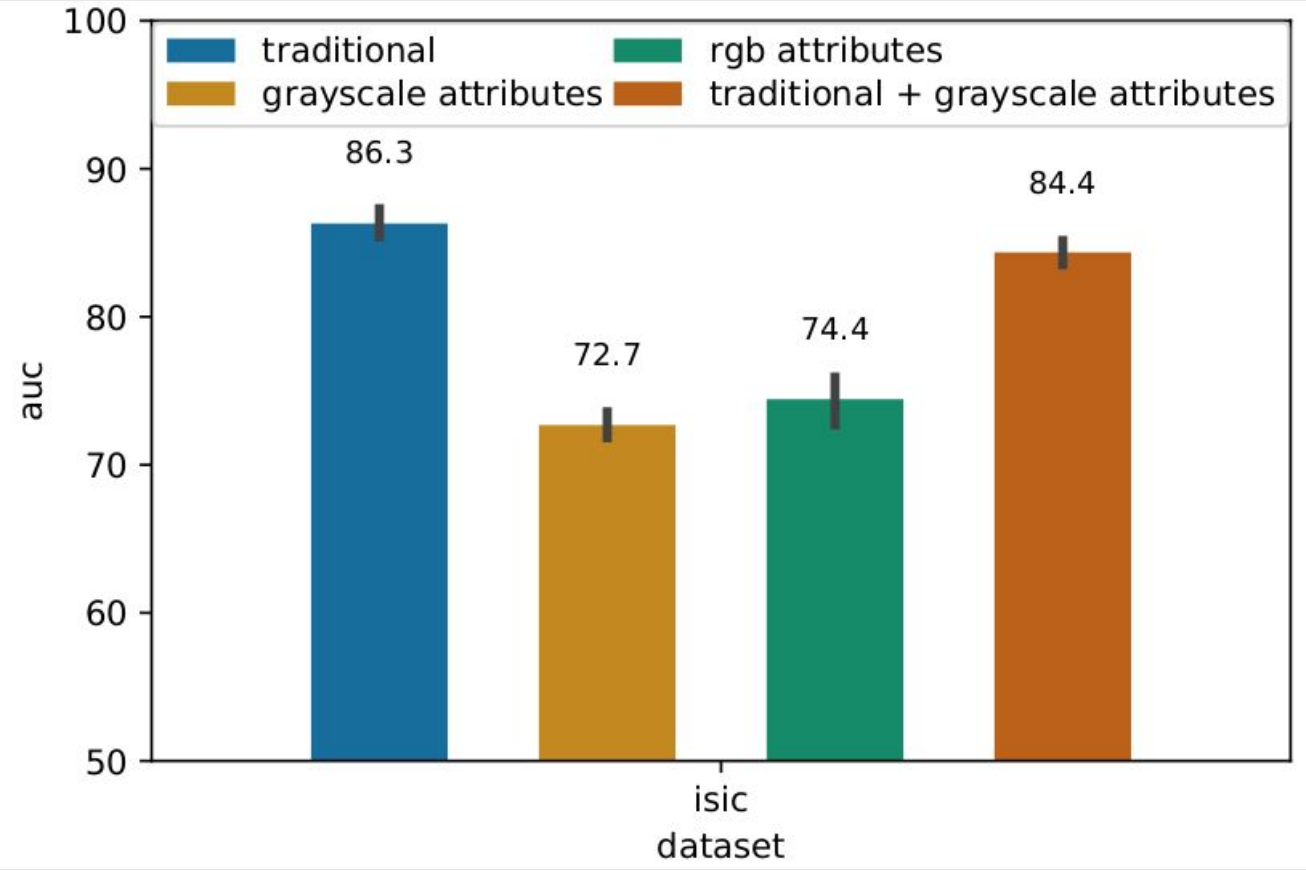
RGB Attributes



Traditional + Grayscale Attributes



Construction Experiments



Conclusions

Machine learning results results are probably **optimistic**

Feeding the model with relevant dermoscopic attributes **is worse** than feeding it with “only skin” or “bbox” sets

Solving the bias problem is **critical** for deploying automated skin lesion analysis to the real world

Team

Prof. Eduardo Valle
(FEEC/UNICAMP)

Prof. Sandra Avila
(IC/UNICAMP)

Prof. Flávia V. Bittencourt
(Dermatologist/UFMG)

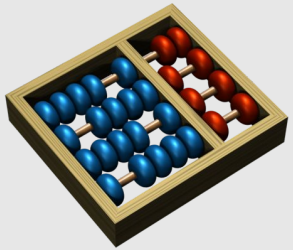
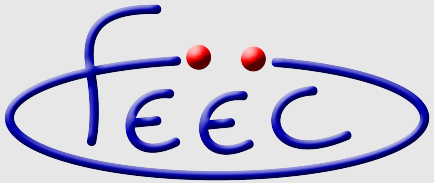
Michel Fornaciali
(Ph.D. Student)

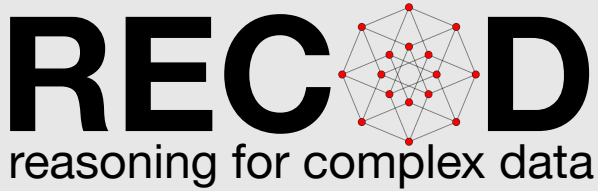
Alceu Bissoto
(M.Sc. Student)

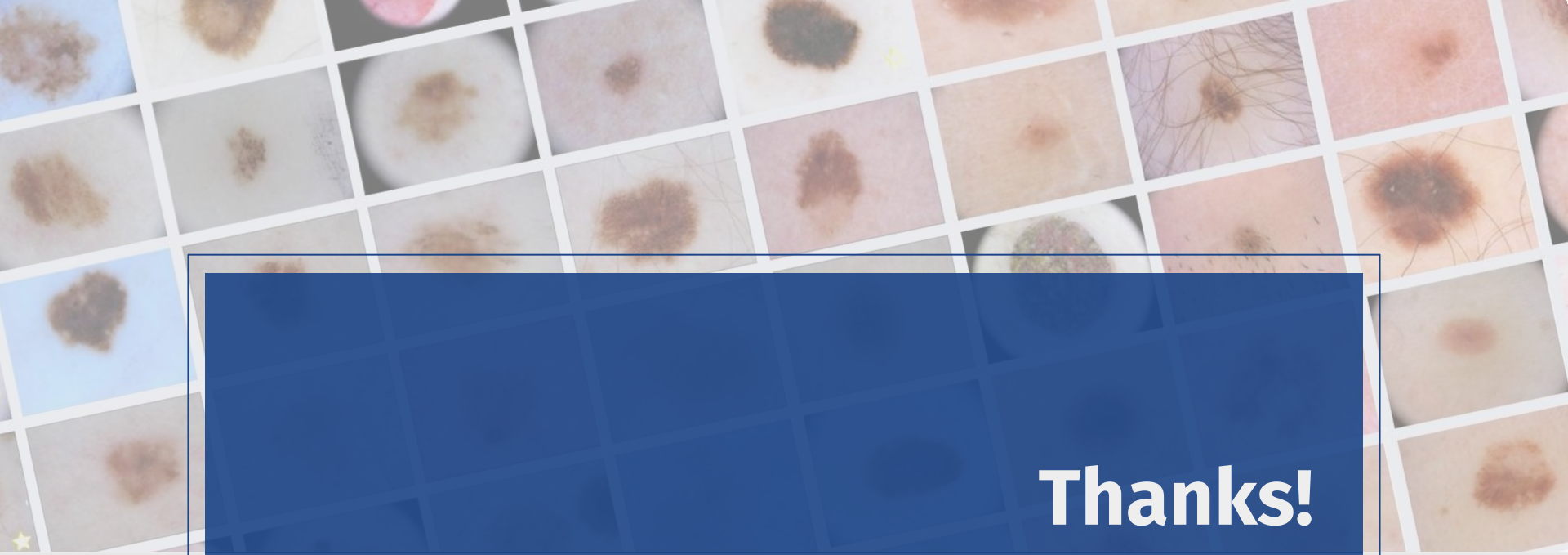
Fábio Perez
(M.Sc. Student)

Vinicius Ribeiro
(M.Sc. Student)

Acknowledgments





Thanks!