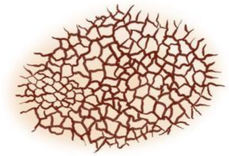
Debiasing Skin Lesion Datasets and Models? Not So Fast

Alceu Bissoto[1], Eduardo Valle[2], Sandra Avila[1]
[1]RECOD Lab., IC, University of Campinas (UNICAMP), Brazil
[2]RECOD Lab., DCA, FEEC, University of Campinas (UNICAMP), Brazil

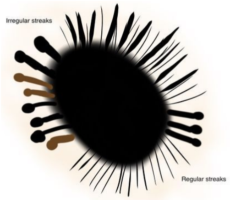
ISIC Workshop @ CVPR 2020

Medical Criteria

Pigment Network

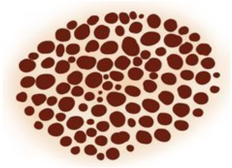


Streaks

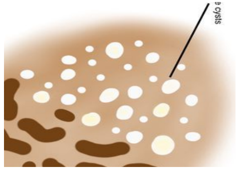


Negative Network

Globules

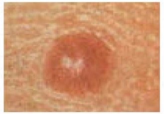


Milia-like cysts

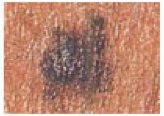


https://dermoscopedia.org/

Asymmetry



symmetrical



asymmetrical

Border Regularity

even boarders

uneven boarders

Color

one color

multi colored

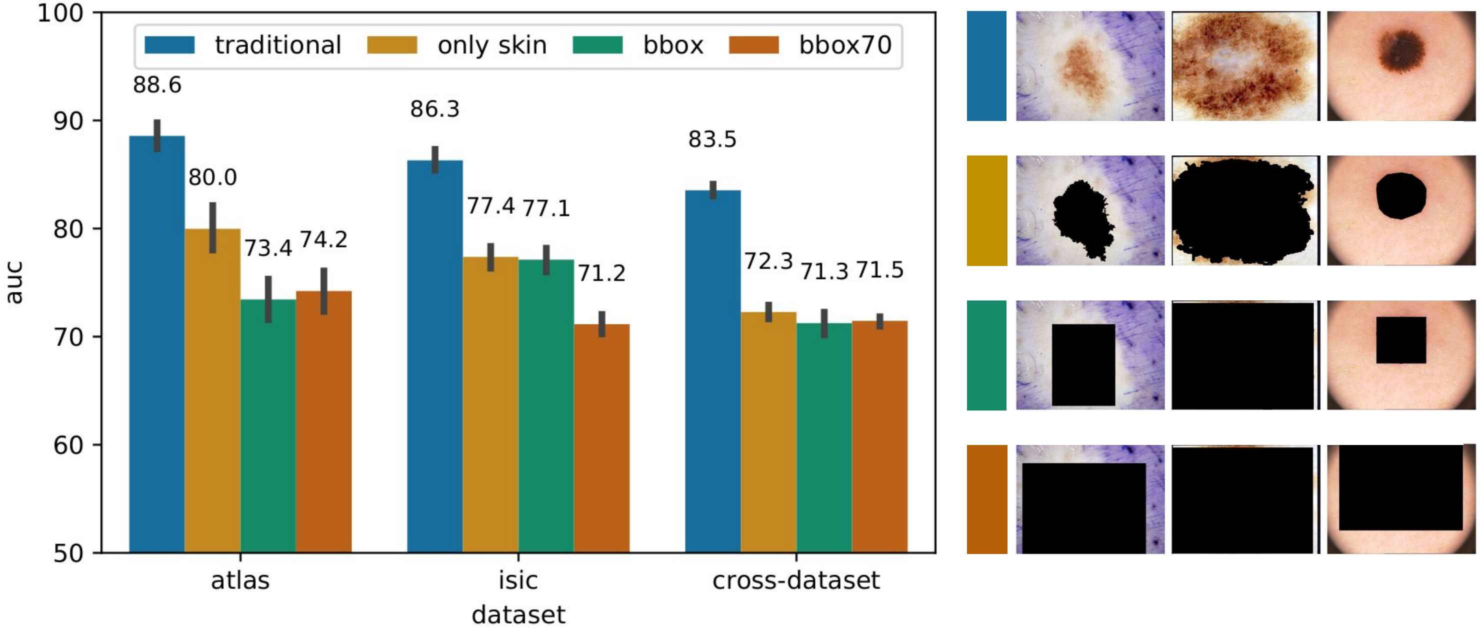
Diameter

smaller than 1/4 in.

larger than 1/4 in.

However, previously on ...

(De)Constructing Bias in Skin Lesion Datasets, Bissoto et al., ISIC Workshop @ CVPR 2019



Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

Objective

Annotation regarding **7 visual artifacts** that can lead to dataset biases

How those artifacts **affect classification** models?

Bias removal in the skin lesion context

Objective

Annotation regarding **7 visual artifacts** that can lead to dataset biases

How those artifacts **affect classification** models?

Bias removal in the skin lesion context

Objective

Annotation regarding **7 visual artifacts** that can lead to dataset biases

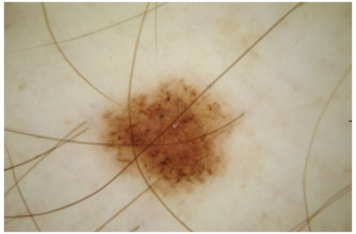
How those artifacts **affect classification** models?

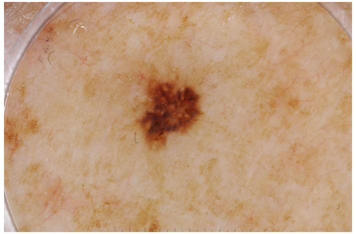
Bias removal in the skin lesion context

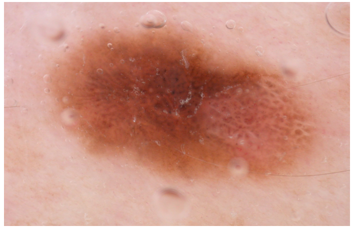
Custom Data

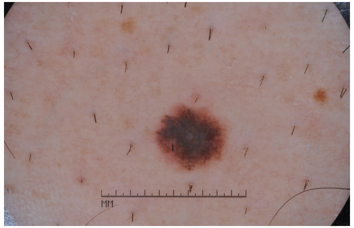
Suspect Artifacts

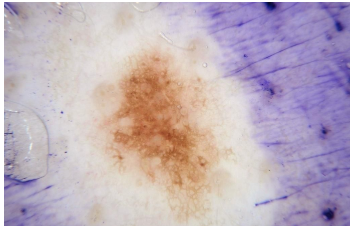

Dark Corners (Vignetting)

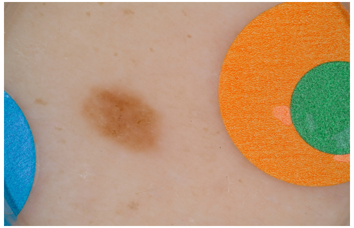

Hair


Gel Border


Gel Bubble


Ruler

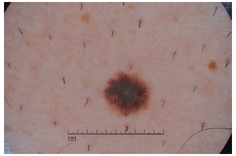

Ink Markings


Patches

Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

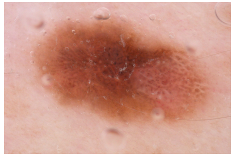
Suspect Artifacts

Ruler

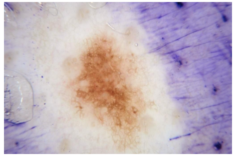


Dark Corners

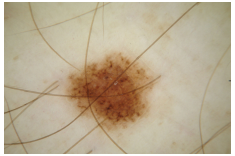
Gel Bubble



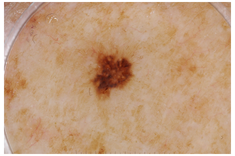
Ink Markings



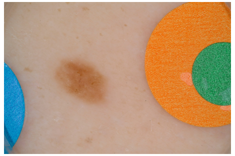
Hair



Gel Border

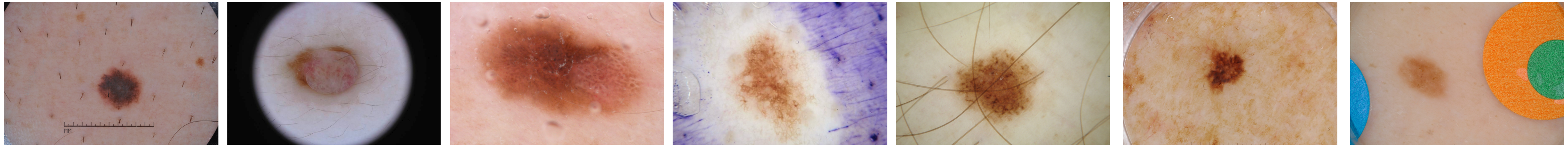


Patches



	Ruler	Dark Corners	Gel Bubble	Ink Markings	Hair	Gel Border	Patches
Spearman Correlation w.r.t. diagnosis	0.10	0.08	0.01	-0.07	-0.08	-0.10	-0.13

Suspect Artifacts

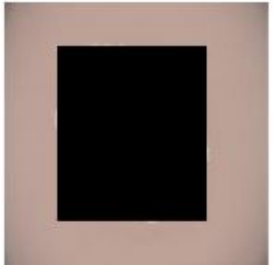
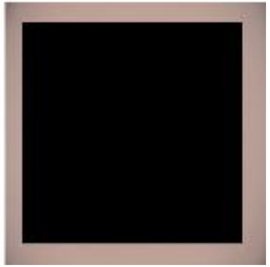


	Ruler	Dark Corners	Gel Bubble	Ink Markings	Hair	Gel Border	Patches
Spearman Correlation w.r.t. diagnosis	0.10	0.08	0.01	-0.07	-0.08	-0.10	-0.13
Models' Identification Performance	98.2%	95.6%	85.3%	97.8%	94.0%	93.4%	98.2%

Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

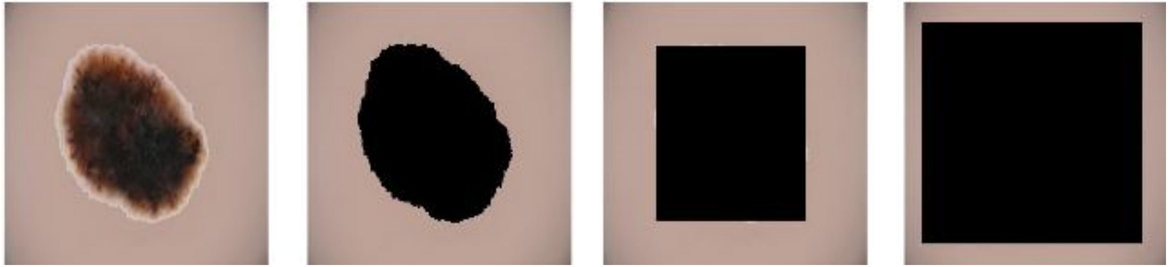
Normalized Datasets



Dataset	Traditional (%)	Skin Only (%)	Bbox (%)	Bbox70 (%)
ISIC	86.3	77.3	77.1	71.1
ISIC Normalized	81.5	72.7	67.0	59.8
Cross-dataset	83.5	72.3	71.3	71.5
Cross-dataset Normalized	77.1	69.0	67.2	64.1

Normalized Datasets



Dataset	Traditional (%)	Skin Only (%)	Bbox (%)	Bbox70 (%)
ISIC	86.3	77.3	77.1	71.1
ISIC Normalized	81.5	72.7	67.0	59.8
Cross-dataset	83.5	72.3	71.3	71.5
Cross-dataset Normalized	77.1	69.0	67.2	64.1

Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

Trap Sets

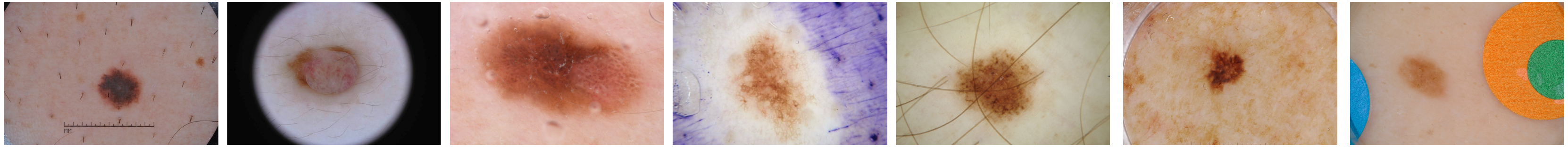
Uses artifacts to **purposefully** mislead classifiers.

Non-random splits maximize artifact bias on train and ***opposite*** bias on test.

Models that ignore artifacts should be **unaffected**.

Models that exploit biased should fail catastrophically (**all tested models did**).

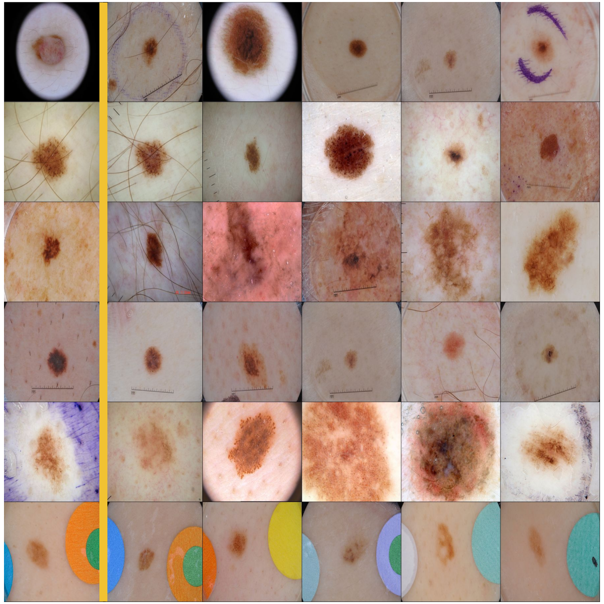
Trap Sets



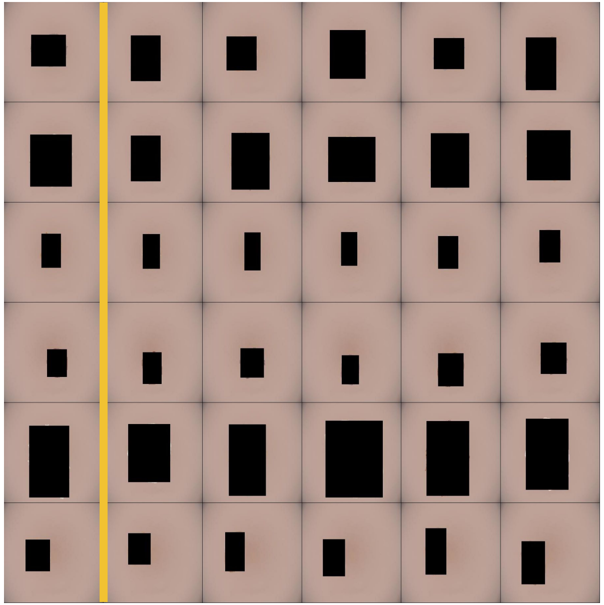
	Ruler	Dark Corners	Gel Bubble	Ink Markings	Hair	Gel Border	Patches
Train Spearman Correlation	0.41	0.30	-0.18	0.21	-0.26	0.12	-0.11
Test Spearman Correlation	-0.67	-0.39	0.47	-0.42	0.34	-0.51	-0.16

What features are being used?

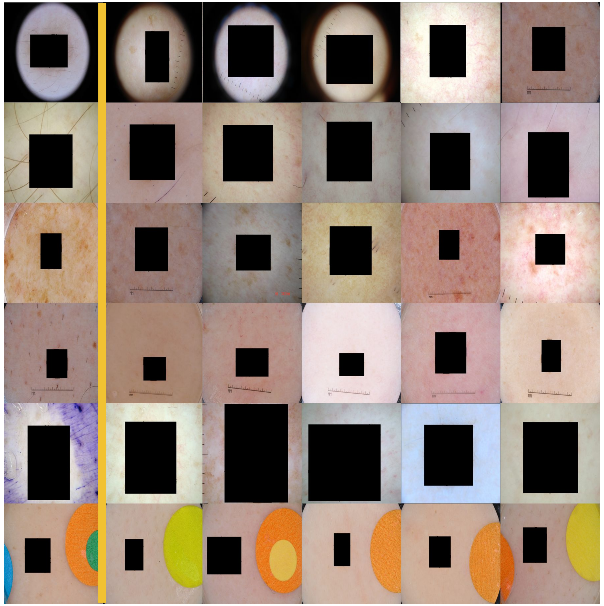
Traditional



Normalized Bbox



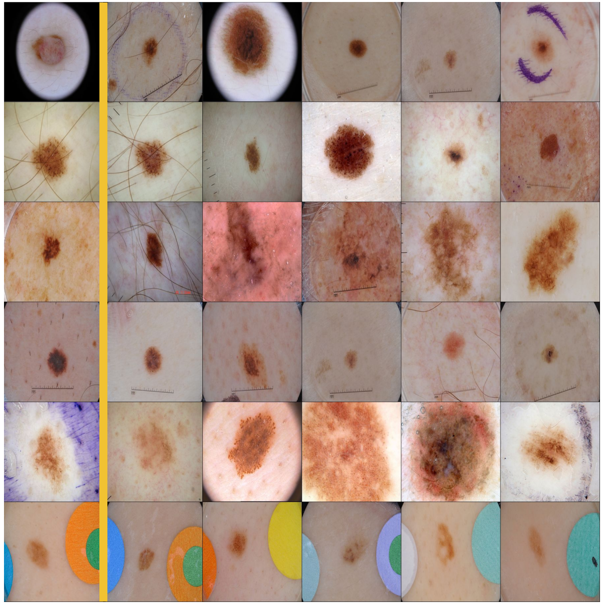
Bbox



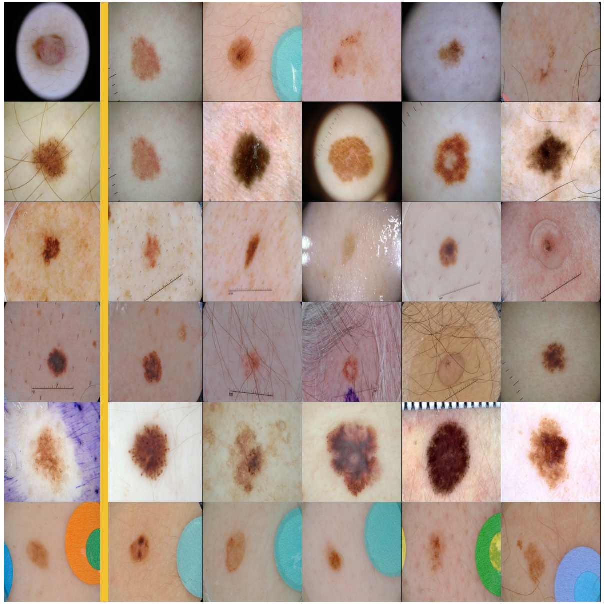
Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

What features are being used?

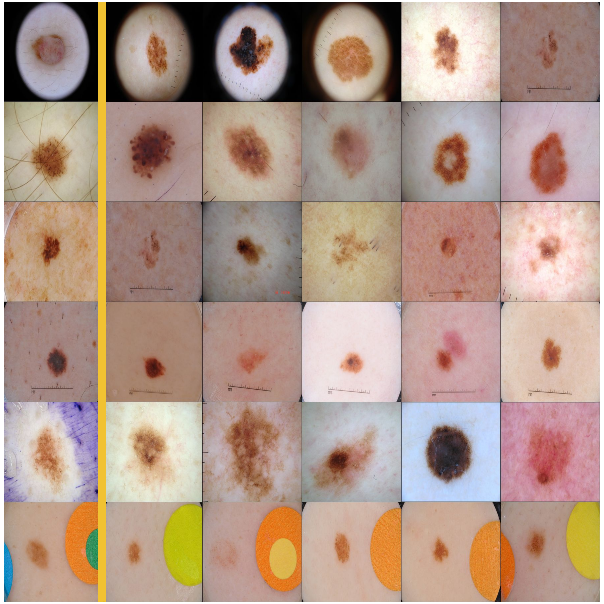
Traditional



Normalized Bbox



Bbox

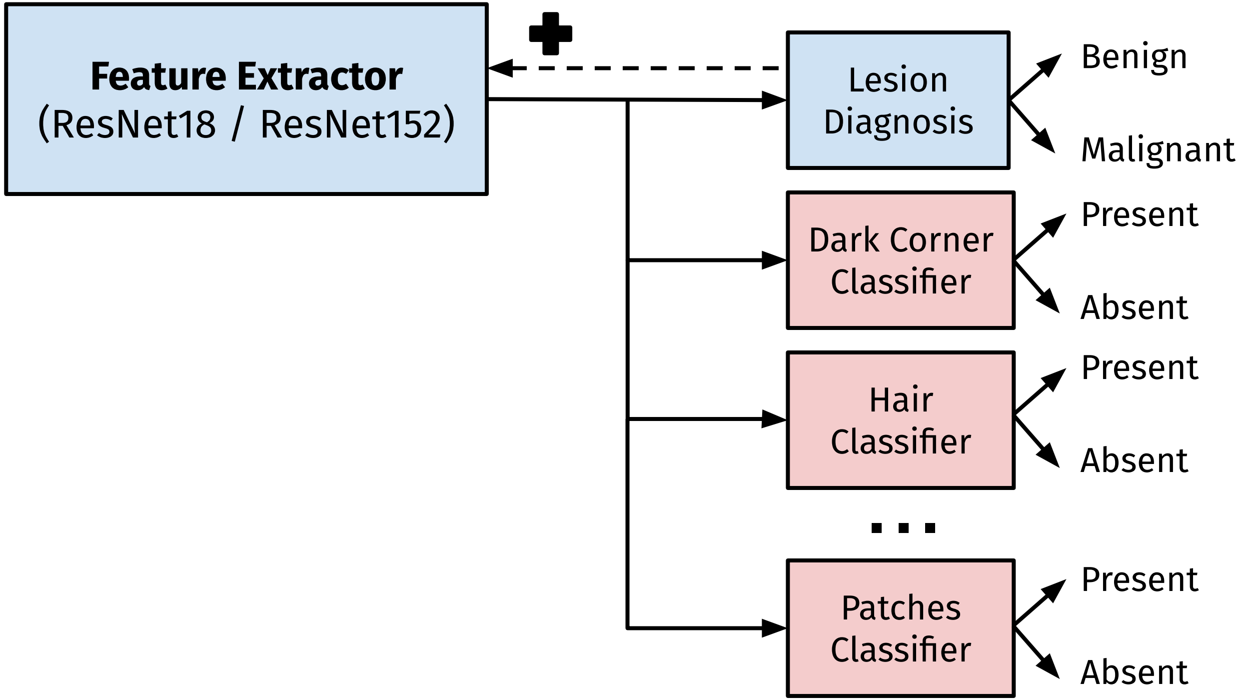


Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

Debiasing Experiments

Debiasing - Learning not to Learn (LNTL)

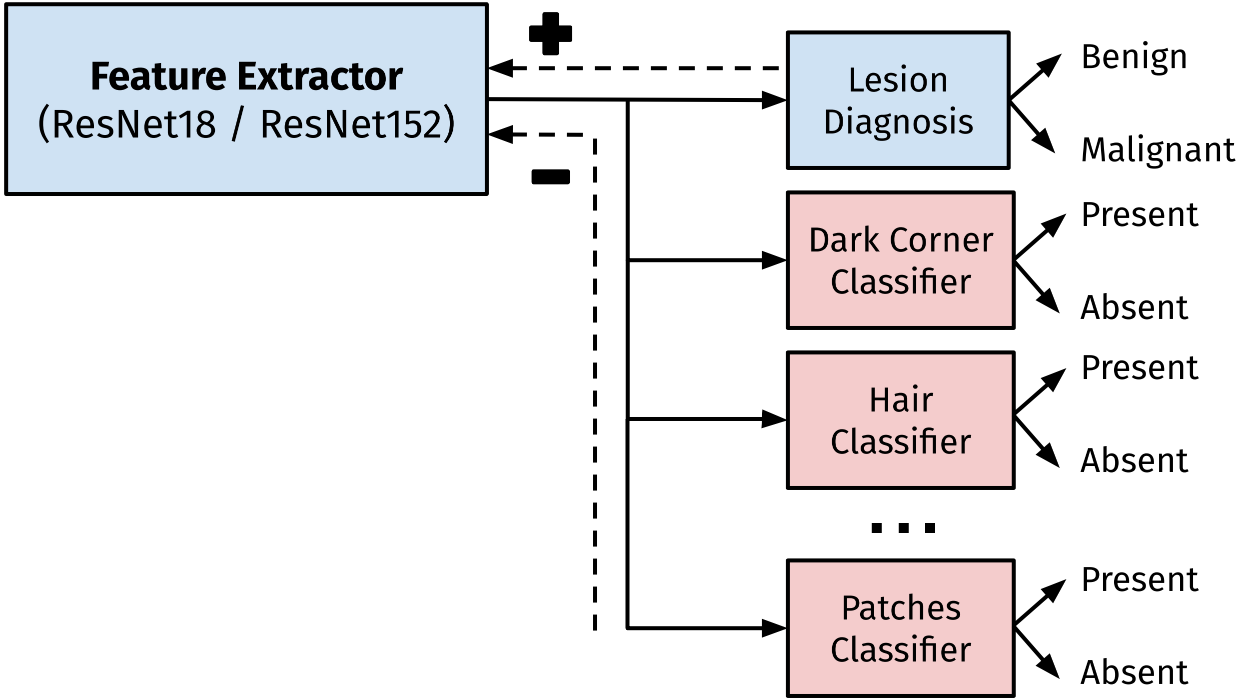
Learning not to Learn: Training Deep Neural Networks with Biased Data, Kim et al., CVPR 2019



Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

Debiasing - Learning not to Learn (LNTL)

Learning not to Learn: Training Deep Neural Networks with Biased Data, Kim et al., CVPR 2019



Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

Debiasing

Experiment	Architecture	**Trap Test (%)**	Atlas Dermato (%)	Atlas Clinical (%)
Unchanged	Inceptionv4	52.6	78.5	63.4
Normalized	Inceptionv4	55.8	72.4	-
LNTL	ResNet152	54.5	**78.4**	**70.1**
Unchanged	ResNet18	44.7	72.2	65.8
Normalized	ResNet18	62.4	70.5	-
LNTL	ResNet18	51.4	**76.0**	**68.2**

Debiasing Skin Lesion Datasets and Models? Not So Fast, Bissoto et al., ISIC Workshop @ CVPR 2020

Conclusions

Traditional models are less biased than previously thought **(but they are still biased)**.

Debiasing methods **struggle** to deal with the skin cancer.

Domain adaptation, **representation learning** and **disentanglement** for more robust classifiers.

Conclusions

Traditional models are less biased than previously thought **(but they are still biased)**.

Debiasing methods **struggle** to deal with the skin cancer.

Domain adaptation, **representation learning** and **disentanglement** for more robust classifiers.

Conclusions

Traditional models are less biased than previously thought **(but they are still biased)**.

Debiasing methods **struggle** to deal with the skin cancer.

Domain adaptation, **representation learning** and **disentanglement** for more robust classifiers.

Thank you!

Alceu Bissoto alceubissoto@ic.unicamp.br

Eduardo Valle dovalle@dca.fee.unicamp.br

Sandra Avila @sandraavilabr

Code & Data:

UNICAMP feec recod

CAPES LARA CNPq FAPESP CCES UNICAMP