Taken Out of Context

Towards Mitigating Contextual Bias in Computer Vision

Nicole Meister
Advised by Prof. Russakovsky

Overview

1. Problem

When does context induce bias?

2. Problem Set Up

Quantifying Contextual Bias, Baseline Models

3. Meta Learning

Stacking Method

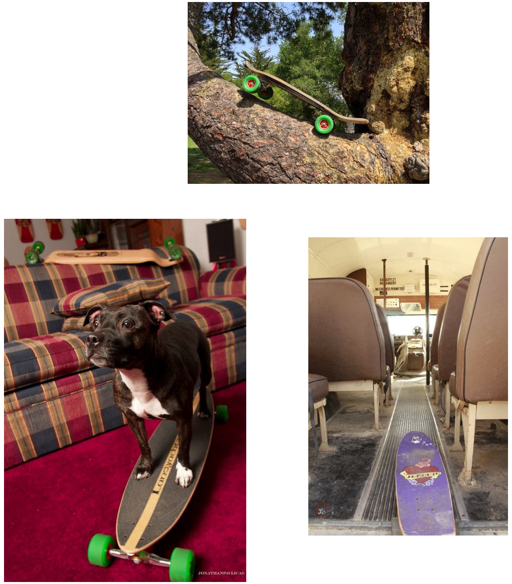
4. Decision Rule

Identify exclusive image using model scores

5. Improve Decisions

Additional methods to Identify exclusive images post hoc

Contextual Bias

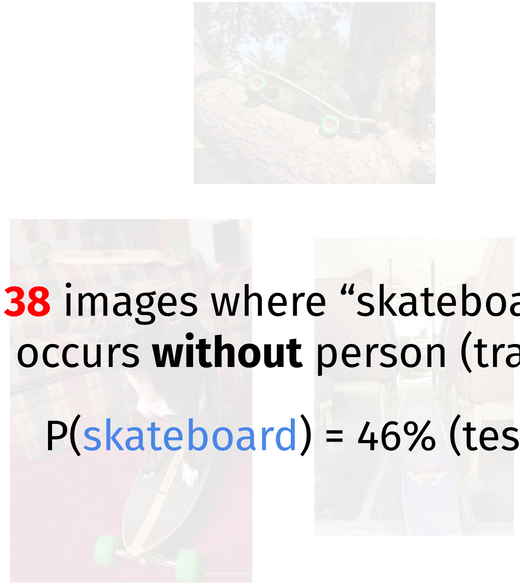


Exclusive: Skateboard

Co-occur: Skateboard & Person

“Microsoft COCO: Common Objects in COntext” by Tsung-Yi Lin et.al

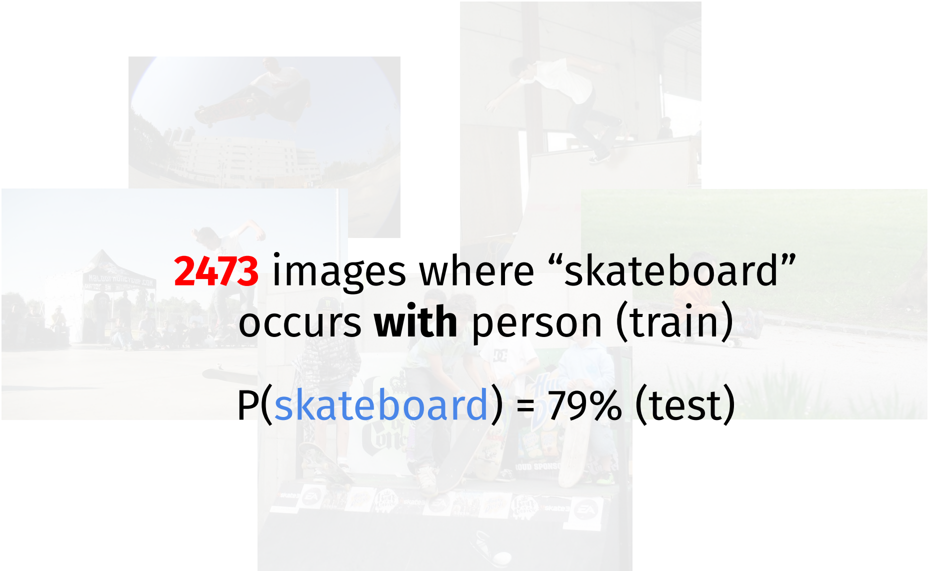
When does context induce bias?



38 images where “skateboard” occurs **without** person (train)

P(skateboard) = 46% (test)

Exclusive: Skateboard



2473 images where “skateboard” occurs **with** person (train)

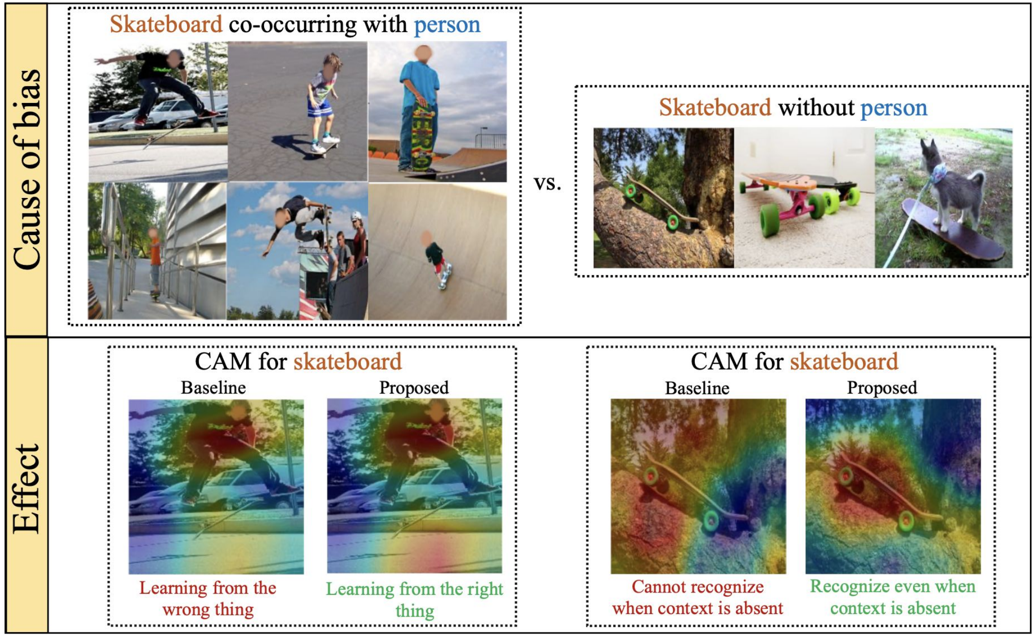
P(skateboard) = 79% (test)

Co-occur: Skateboard & Person

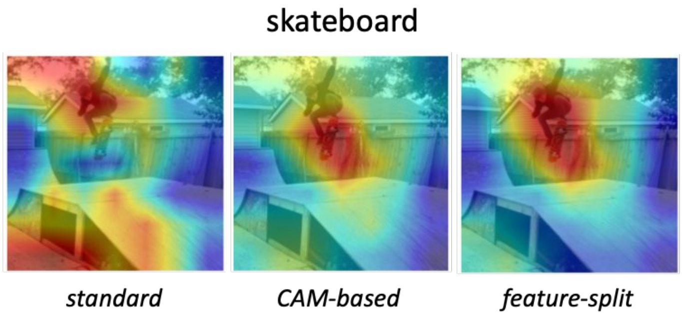
“Microsoft COCO: Common Objects in Context” by Tsung-Yi Lin et.al

Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias (CVPR 2020)

Singh et. al



Main Figure (Source: Singh et.al)



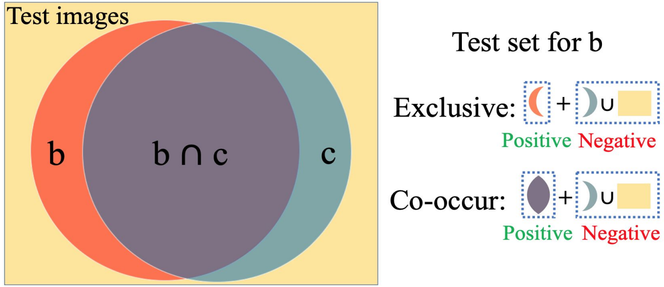
Our Reproduced Results

[Re] Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias (Rescience C 2021) Kim et. al

Exclusive and Co-occur Distributions

Biased category pairs		Bias		Biased category pairs (Ours)		
Biased (b)	Context (c)	Paper	Ours	Biased (b)	Context (c)	Bias
cup	dining table	1.76	1.85	car	road	1.73
wine glass	person	1.80	1.59	potted plant	furniture-other	1.75
handbag	person	1.81	2.25	spoon	bowl	1.75
apple	fruit	1.91	2.12	fork	dining table	1.78
car	road	1.94	1.73	bus	road	1.79
bus	road	1.94	1.79	cup	dining table	1.85
potted plant	vase	1.99	1.73	mouse	keyboard	1.87
spoon	bowl	2.04	1.75	remote	person	1.89
microwave	oven	2.08	1.59	wine glass	dining table	1.94
keyboard	mouse	2.25	2.11	clock	building-other	1.97
skis	person	2.28	2.21	keyboard	mouse	2.11
clock	building	2.39	1.97	apple	fruit	2.12
sports ball	person	2.45	3.61	skis	snow	2.22
remote	person	2.45	1.89	handbag	person	2.25
snowboard	person	2.86	2.40	snowboard	person	2.40
toaster	ceiling[6]	3.70	1.98	skateboard	person	3.41
hair drier	towel	4.00	3.49	sports ball	person	3.61
tennis racket	person	4.15	1.26	hair drier	sink	6.11
skateboard	person	7.36	3.41	toaster	oven	8.56
baseball glove	person	339.15	31.32	baseball glove	person	31.32

[Re] Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias
(Rescience C 2021) Kim et. al



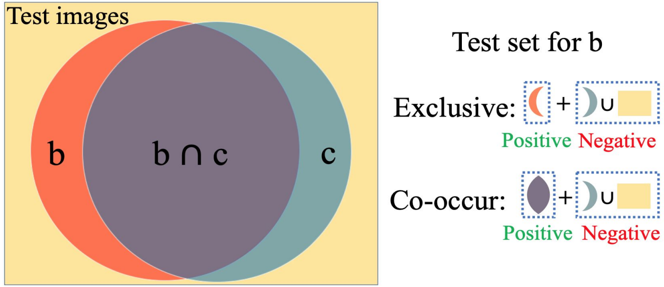
Don't Judge an Object by Its Context:
Learning to Overcome Contextual Bias
(CVPR 2020) Singh et. al

Exclusive and Co-occur Distributions

Biased category pairs		Bias		Biased category pairs (Ours)		
Biased (b)	Context (c)	Paper	Ours	Biased (b)	Context (c)	Bias
cup	dining table	1.76	1.85	car	road	1.73
wine glass	person	1.80	1.59	potted plant	furniture-other	1.75
handbag	person	1.81	2.25	spoon	bowl	1.75
apple	fruit	1.91	2.12	fork	dining table	1.78
car	road	1.94	1.73	bus	road	1.79
bus	road	[illegible]	1.79	cup	dining table	1.85
potted plant	vase	[illegible]	[illegible]	mouse	keyboard	1.87
spoon	bowl	2.04	1.75	remote	person	1.89
microwave	oven	2.08	1.59	wine glass	dining table	1.94
keyboard	mouse	[illegible]	[illegible]	clock	building-other	1.97
skis	person	2.28	2.21	keyboard	mouse	2.11
clock	building	2.39	1.97	apple	fruit	2.12
sports ball	person	2.45	3.61	skis	snow	2.22
remote	person	2.45	1.89	handbag	person	2.25
snowboard	person	2.86	2.40	snowboard	person	2.40
toaster	ceiling[6]	3.70	1.98	skateboard	person	3.41
hair drier	towel	4.00	3.49	sports ball	person	3.61
tennis racket	person	4.15	1.26	hair drier	sink	6.11
skateboard	person	7.36	3.41	toaster	oven	8.56
baseball glove	person	339.15	31.32	baseball glove	person	31.32

18/20 overlapping biased category pairs

[Re] Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias (Rescience C 2021) Kim et. al



Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias (CVPR 2020) Singh et. al

Reproducing Baseline Models

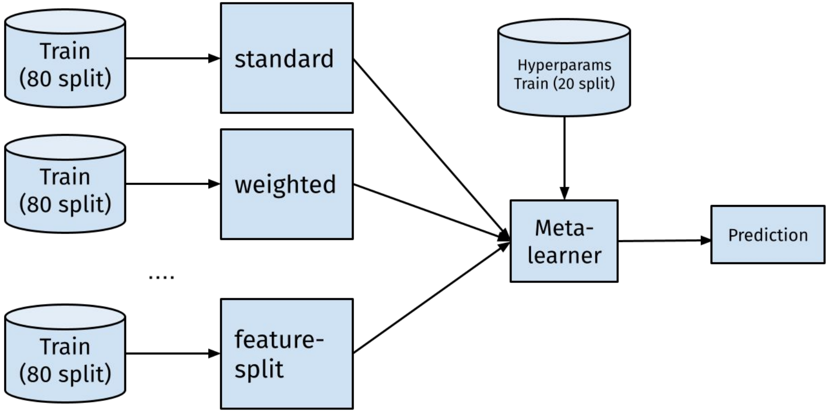
Method	COCO-Stuff (mAP)			
	Exclusive		Co-occur	
	Paper	Ours	Paper	Ours
standard[4]	24.5	23.9	**66.2**	**65.0**
remove labels	25.2	24.5	65.9	64.6
remove images	28.4	**29.0**	28.7	59.6
split-biased	19.1	25.4	64.3	64.7
weighted	**30.4**	28.5	60.8	60.0
negative penalty	23.8	23.9	66.1	64.7
class-balancing	25.0	24.6	66.1	64.7
attribute decorr.	-	-	-	-
CAM-based	26.4	26.9	64.9	64.2
feature-split	28.8	28.1	66.0	64.8

ResNet-50 pretrained on ImageNet, optimized with SGD, momentum=0.9, lr = 0.1 → 0.01

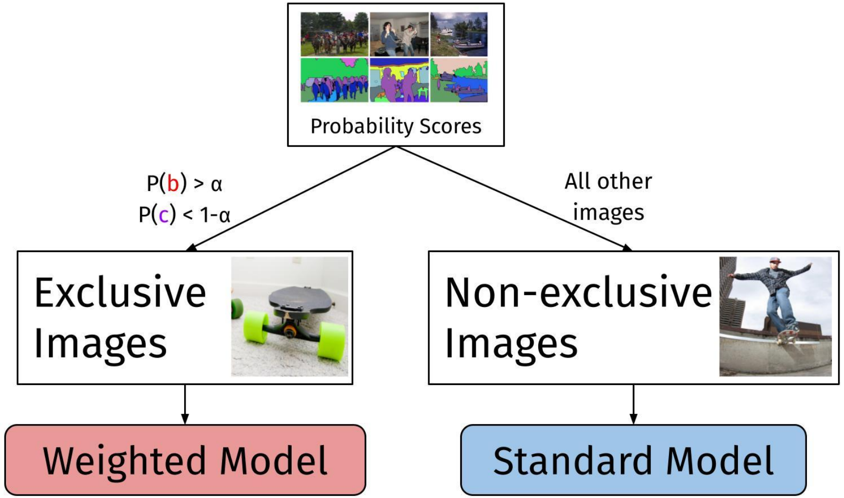
For each b, apply 10 times higher weight to the loss for class b when b occurs exclusively

[Re] Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias (Rescience C 2021) Kim et. al

Methods



Stacking

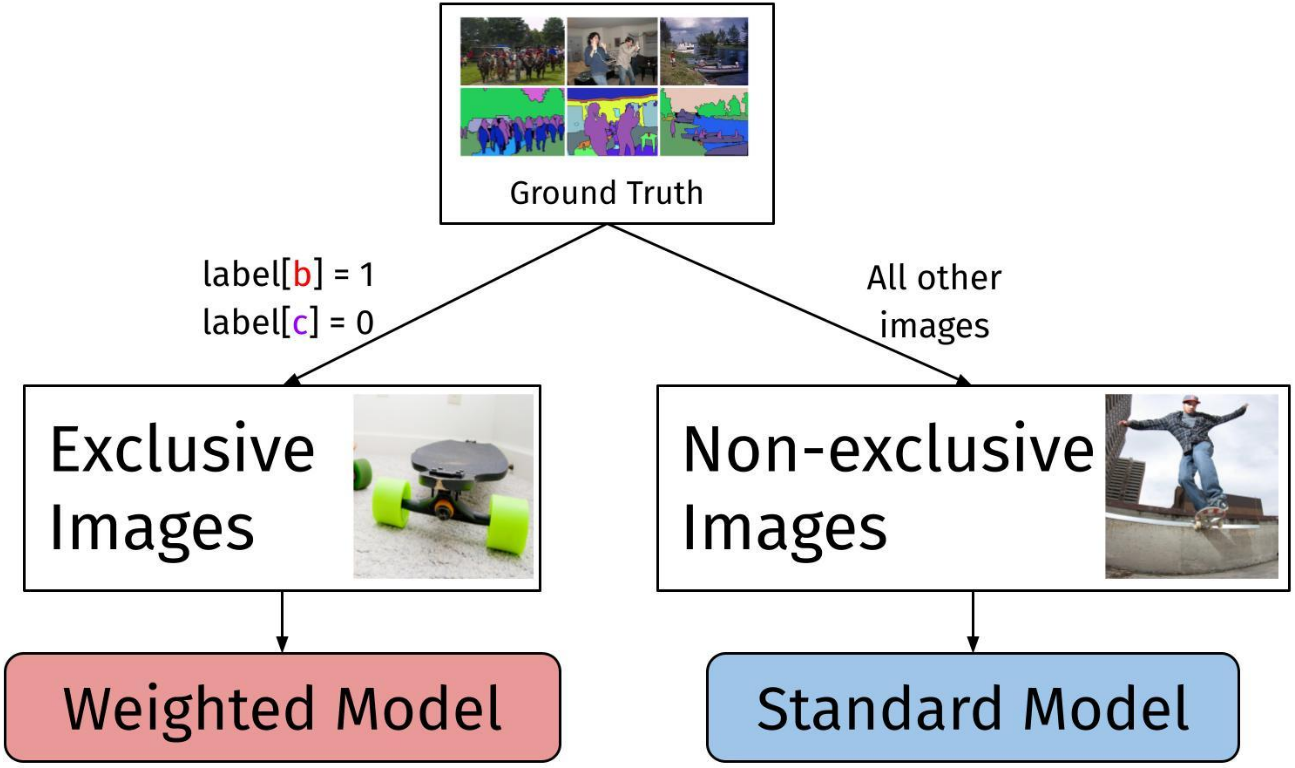


Decision Rule Algorithm

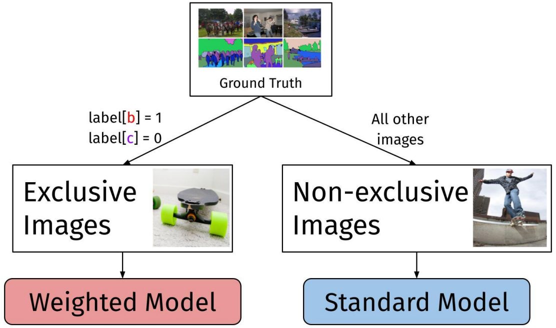
Decision Rule Method

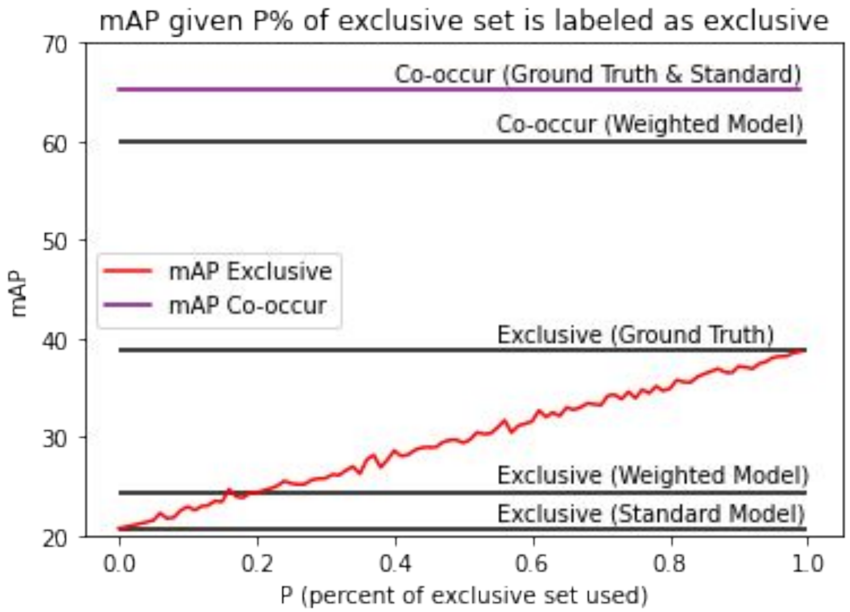
Method	COCO-Stuff (mAP)			
	Exclusive		Co-occur	
	Paper	Ours	Paper	Ours
standard[4]	24.5	23.9	**66.2**	**65.0**
remove labels	25.2	24.5	65.9	64.6
remove images	28.4	**29.0**	28.7	59.6
split-biased	19.1	25.4	64.3	64.7
weighted	**30.4**	28.5	60.8	60.0
negative penalty	23.8	23.9	66.1	64.7
class-balancing	25.0	24.6	66.1	64.7
attribute decorr.	-	-	-	-
CAM-based	26.4	26.9	64.9	64.2
feature-split	28.8	28.1	66.0	64.8

Decision Rule Upper Bound

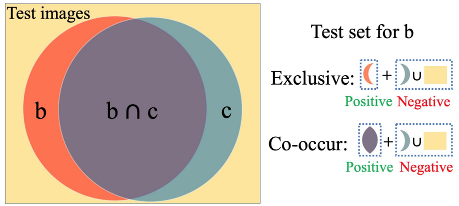


Decision Rule Upper Bound



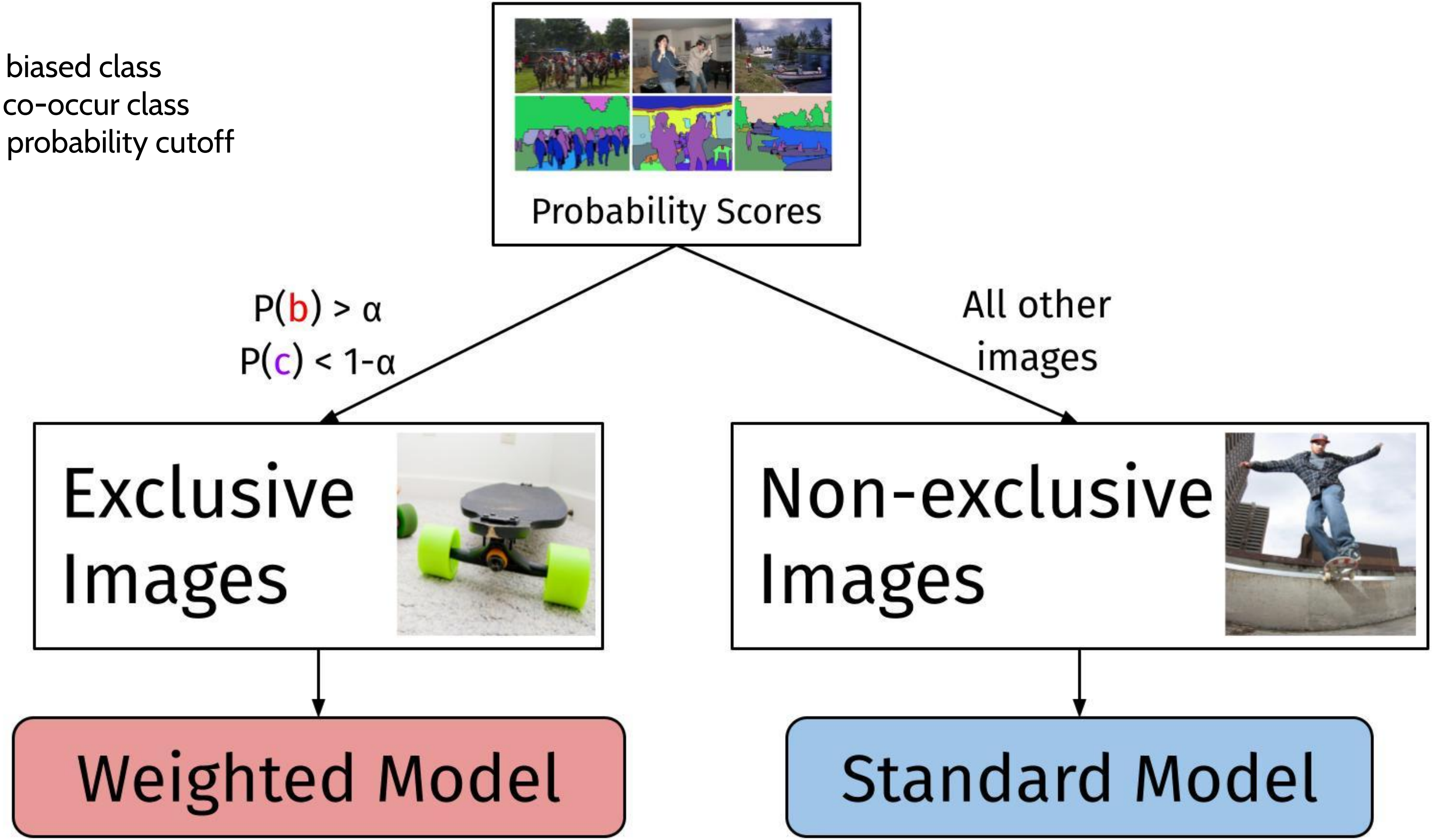


mAP	Exclusive	Co-Occur
Standard Model	20.7	65.2
Weighted Model	24.4	60.0
Decision with Ground-Truth labels	**38.8**	**65.2**

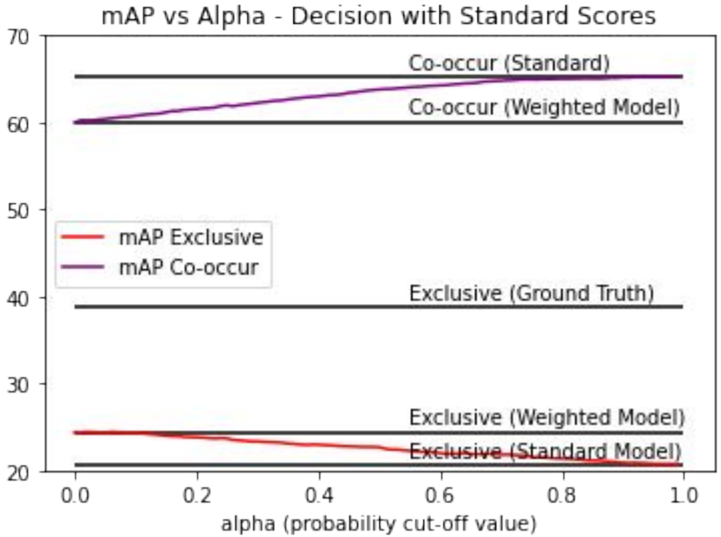


images in the exclusive distribution are fed through the *weighted* model, the "other" images are fed through the *standard* model. We developed this insight by looking at mean probability scores for label b, $\bar{P}(b)$ where for exclusive images, a larger $\bar{P}(\hat{b})$ indicates a better model and for "other" images, a lower $\bar{P}(\hat{b})$ indicates a better model because it recognizes that the biased class does not exist in the image. On exclusive images, $\bar{P}_{\text{standard}}(\hat{b}) = 0.33$ and $\bar{P}_{\text{weighted}}(\hat{b}) = 0.45$. On the "other" images, $\bar{P}_{\text{standard}}(\hat{b}) = 0.019$ and $\bar{P}_{\text{weighted}}(\hat{b}) = 0.035$. Since there exist many more "other" images in the dataset than exclusive images, when the *standard* model's performance on the "other" images is leveraged, the mAP on the exclusive distribution (both the exclusive and "other" images) increases.

b: biased class
c: co-occur class
α: probability cutoff

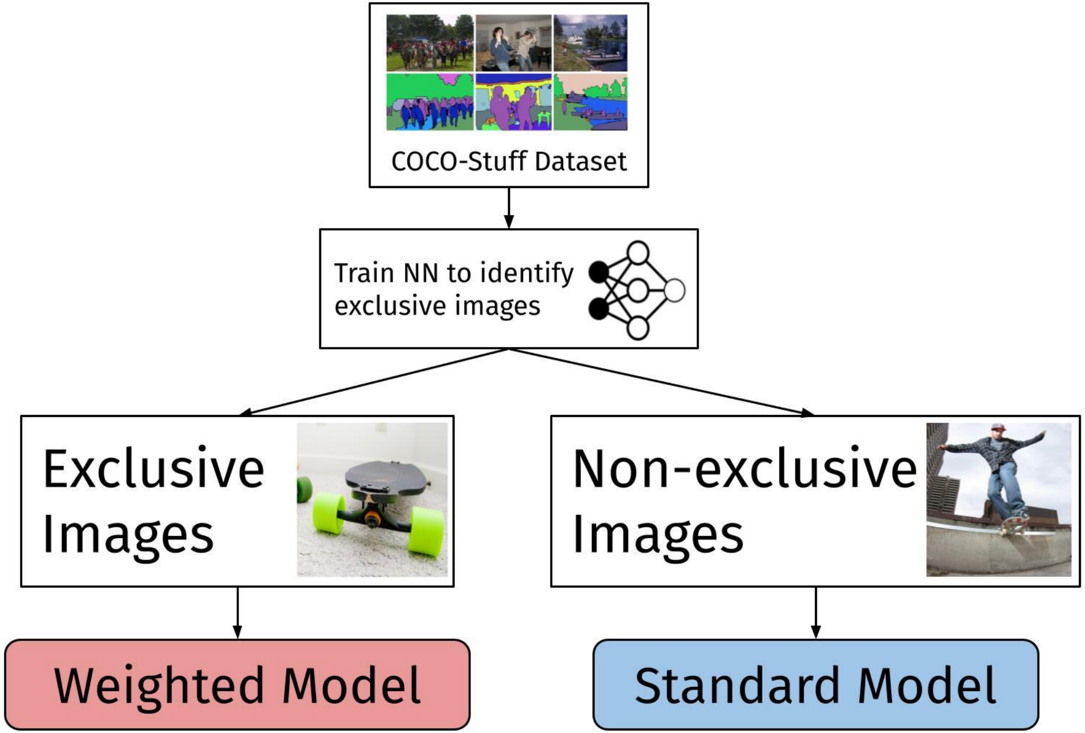


Results

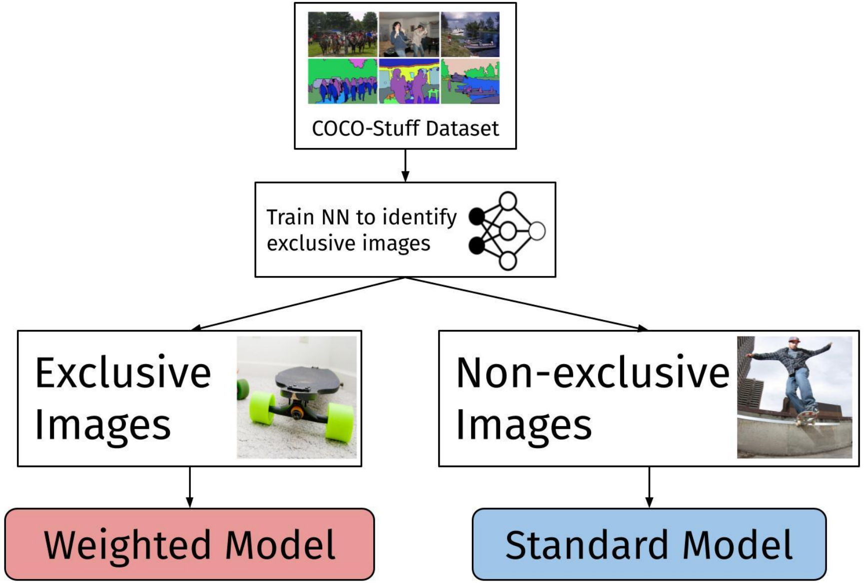


mAP	Exclusive	Co-Occur
Standard	20.7	65.2
Weighted	24.4	60.0
Ground Truth	**38.8**	**65.2**
Decision with standard probabilities (alpha=.53)	22.4	63.9
Feature-split (Singh et. al)	22.4	64.5

Extensions - Train Model to Identify Exclusive Images



Results



low precision & high recall for exclusive class

Implemented 3 different reweighting methods, training from scratch, and training from standard model to leverage learned features

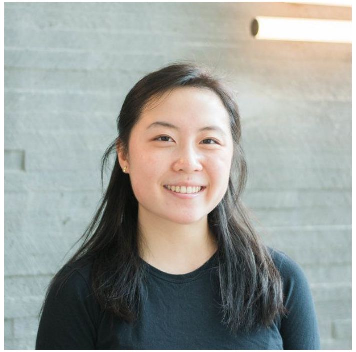
Explanation: Identifying exclusive images is a more complex task than image classification

Acknowledgements

PRINCETON
Electrical Engineering

Professor Olga Russakovsky

Sunnie S. Y. Kim

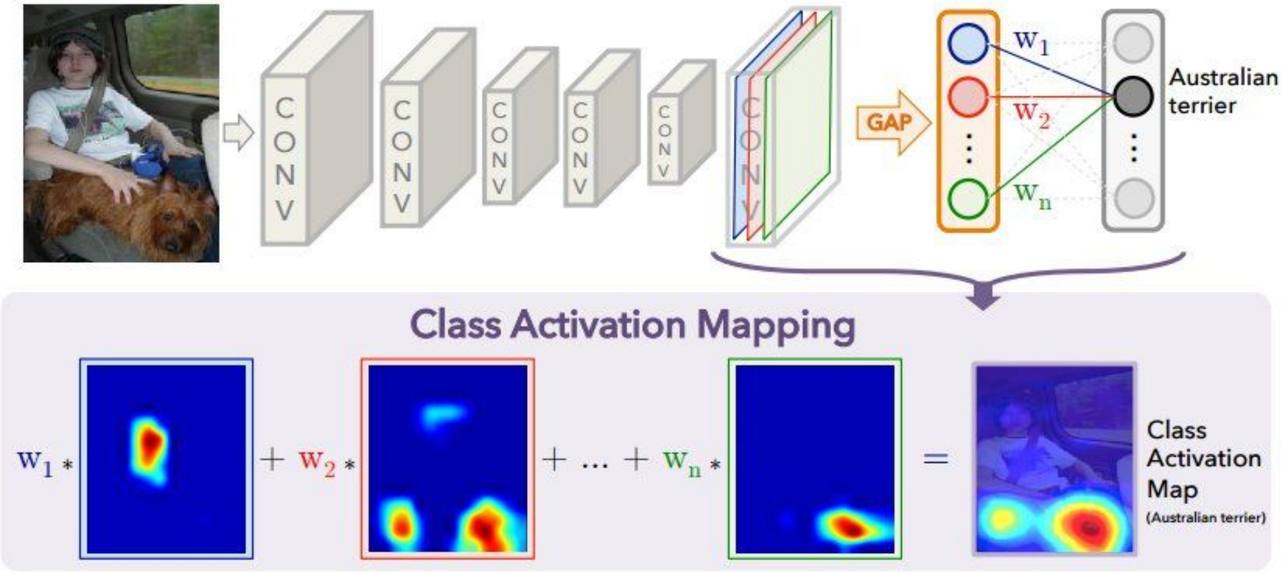


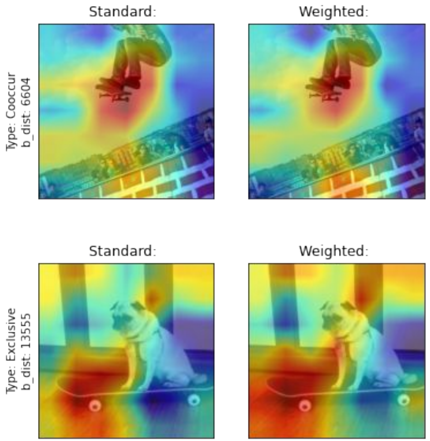
Sharon Zhang

Visual AI Lab

Professor Jason Lee

Extensions - Class Activation Maps (CAM)





Learning Deep Features for Discriminative Localization (CVPR 2016) Zhou et. al

Quantifying Contextual Bias

$$\mathrm{bias}(\mathrm{b}, \mathrm{z}) = \frac{\frac{1}{|\mathbb{I}_\mathrm{b} \cap \mathbb{I}_\mathrm{z}|} \sum_{\mathrm{I} \in \mathbb{I}_\mathrm{b} \cap \mathbb{I}_\mathrm{z}} \hat{\mathrm{p}}(\mathrm{i}, \mathrm{b})}{\frac{1}{|\mathbb{I}_\mathrm{b} \setminus \mathbb{I}_\mathrm{z}|} \sum_{\mathrm{I} \in \mathbb{I}_\mathrm{b} \setminus \mathbb{I}_\mathrm{z}} \hat{\mathrm{p}}(\mathrm{i}, \mathrm{b})}$$

Average prediction probability of b **with** z

Average prediction probability of b **without** z

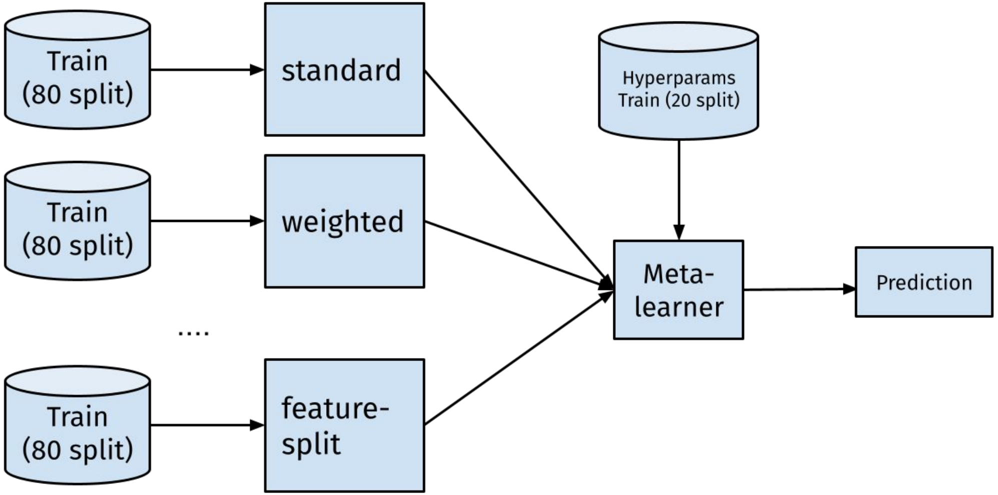
$$\mathrm{c} = \arg\max_{\mathrm{z}} \mathrm{bias}(\mathrm{b}, \mathrm{z})$$

b: biased class
z: Other class
c: co-occurring class

Don't Judge an Object by Its Context: Learning to Overcome Contextual Bias (CVPR 2020) Singh et. al

Meta Learning - Stacking Method

Method	COCO-Stuff (mAP)			
	Exclusive		Co-occur	
	Paper	Ours	Paper	Ours
standard[4]	24.5	23.9	**66.2**	**65.0**
remove labels	25.2	24.5	65.9	64.6
remove images	28.4	**29.0**	28.7	59.6
split-biased	19.1	25.4	64.3	64.7
weighted	**30.4**	28.5	60.8	60.0
negative penalty	23.8	23.9	66.1	64.7
class-balancing	25.0	24.6	66.1	64.7
attribute decorr.	-	-	-	-
CAM-based	26.4	26.9	64.9	64.2
feature-split	28.8	28.1	66.0	64.8



*Partitioned the train 80 split into 5 folds so that each model is trained on ⅘ th of the training set and computes features for the remaining ⅕

Stacking Method - Linear Model

1xM

Learned Weights

x

MxC

Sigmoid Probabilities from single image fed through M models

M: Number of models stacked
C: Classes

Meta-Learner learns to weight each model equally

Explanation: Given that exclusive images are so rare, meta learner is not complex enough to pick up on small differences in each model