

School of Engineering Sciences Institute



CIFAR10-B & CIFAR100-B Dataset

- Dataset with natural semantic spurious correlation: 8 backgorund subgoups, {black, blue, green, grey, red, white, brown, other}. - Based on CIFAR10 CIFAR CIFAR100 ImageNet-10 - Filling the gap in prior Spawrious datasets CIFAR-B **4 labeling principles** 1. Most coverage 2. back (a) Principle 3. judgement 4. others Agreement: ~95% (c) Principle 3 (d) Principle 4 (a) black (b) blue

(e) red

Subgroup Degradation

Huge gap between avg acc and the worst subgroup acc. Low-level data augmentation (DA) does not mitigate it.

(f) white



(g) brown

(h) others







MacroStd	W-Std	CIFAR100	MacroStd	W-Std
2.24	3.45	78.52	12.24	16.44
1.91	3.34	78.87	12.62	16.87
<u>1.83</u>	<u>2.98</u>	<u>79.45</u>	12.17	<u>16.57</u> 🔶
1.65	2.91	81.57	11.20	15.34 🛨 🕇