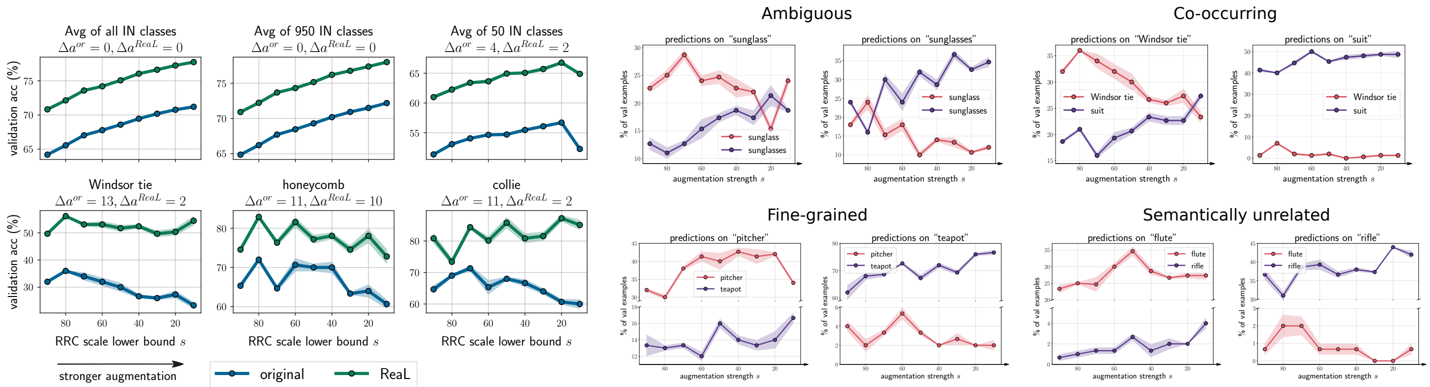

Figure 1: **ViT-S trained on ImageNet with varied Random Resized Crop (RRC) augmentation strength. Left**: Average and per-class accuracy of ViT-S evaluated with original and ReaL labels as a function of RRC augmentation strength. The top row shows the average accuracy of all classes, the 50 classes with the highest accuracy degradation and the remaining 950 classes. The bottom row shows the accuracy of 3 classes most significantly affected in accuracy when using strong augmentation. **Right**: Each panel shows a pair of confused classes which we categorize into: *ambiguous*, *co-occurring*, *fine-grained* and *semantically unrelated*. For each confused class pair, the left subplot corresponds to the class k affected in accuracy by strong data augmentation (DA), e.g. "sunglass" on top left panel: the ratio of validation samples from that class that get classified as k decreases with stronger DA, while the confusion rate with another class l (e.g. class "sunglasses" on top left panel) increases. The right subplot shows the percent of examples from class l that get classified as k or l against DA strength. Generally, we observe that for ViT-S the class confusions which are exacerbated with stronger DA are similar to the ones of ResNet-50.

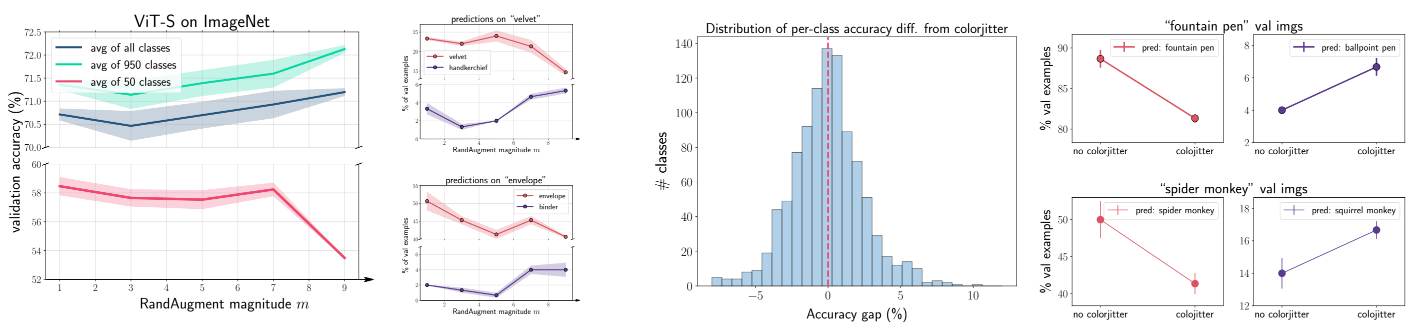

Figure 2: **Different augmentation types: RandAugment and colorjitter. Left**: ViT-S model trained on ImageNet with varied RandAugment magnitude m (larger values of m correspond to stronger augmentation). The left panel shows the average accuracy of all ImageNet classes, the 50 classes with the highest accuracy degradation and the remaining 950 classes. The right panels show the examples of class confusions which are exacerbated by stronger RandAugment augmentation. **Right**: Comparison of ResNet-50 trained on ImageNet with Random Resized Crop $s = 8\%$ when using and not using colorjitter augmentation (applied with probability $p = 0.5$ and intensity $c = 0.1$). The histogram shows per-class accuracy changes when applying vs not applying colorjiter: most classes benefit from augmentation, while a significant number of classes is negatively affected. The right panels show the examples of class confusions exacerbated by applying colorjitter.

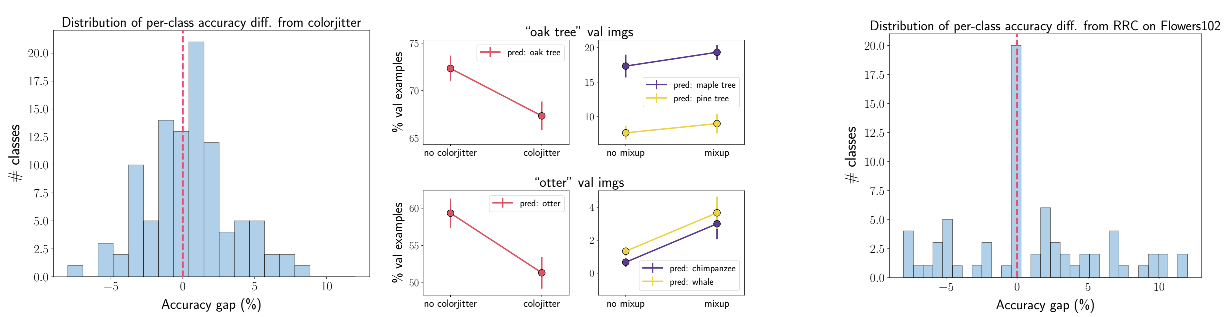

Figure 3: **Other datasets: CIFAR-100 and Flowers102. Left**: The histogram shows per-class accuracy changes when applying vs not applying mixup with $\alpha = 0.5$. The right panels show the examples of class confusions exacerbated by applying mixup, the exacerbated confusions are mostly fine-grained and lie within CIFAR-100 superclasses. **Right**: The histogram shows per-class accuracy changes when applying vs not applying standard Random Resized Crop $s = 8\%$ when training ResNet-32 on Flowers102 dataset.