# A. Supplementary materials

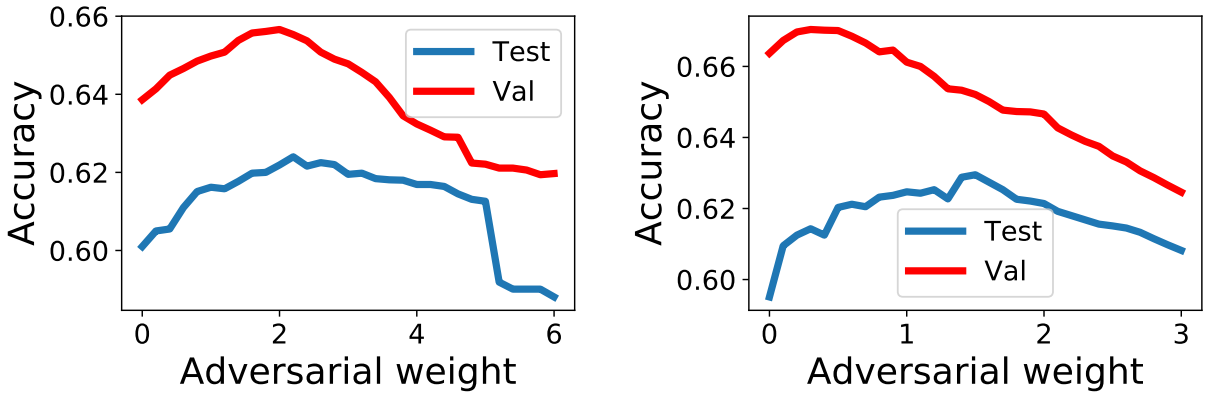
## A.1. Ablation

**Results on semantic groups**: As mentioned in Section 3.3, previous literature utilized a semantic grouping of attributes based on human intuition of similarity. However, we find that the cluster groups (groups formed by spectral co-clustering) via $\Delta_{corr}$ provides grouping that best minimize the chance of correlation shift. For a fair comparison, we show experiments on AWA2, CUB, and AWA2-CS performed with $g$AL variants based on semantic groups accompanying the datasets. These are represented by the combinations **sg+Eq** and **sg+$\Delta_{corr}$** in Table 4. To reiterate, AWA2 comes with 10 semantic groups (e.g. nutrition, habitat) for grouping its 85 attributes whereas CUB is provided with 28 semantic groups (e.g. bill shape, wing color) for grouping its 312 attributes. The results show that cluster groups consistently outperform semantically grouped models, where available.

**Effect of adversarial weighting scheme**: In Section 3.3, we present a adversarial weighting scheme such that classifier loss from the primary task is given a fixed weight of 1, while all adversarial arms are weighted proportional to $\Delta_{corr}$. Here, we present a comparison where all adversarial arms are equally weighted. These are represented by the combinations **sg+Eq** and **cg+Eq** in Table 4.

| Method | **Group+Weights** | **aPY** | **AWA2** | **CUB** | **SUN** | **AWA2-CS** | **aPY-CS** |
|---|---|---|---|---|---|---|---|
| ALE-$g$AL | sg+Eq | - | 51.9 | 49.9 | - | 33.6 | - |
| | sg+$\Delta_{corr}$ | - | 54.0 | 51.0 | - | 35.5 | - |
| | cg+Eq | 37.3 | 59.2 | 49.9 | 59.5 | 40.2 | 18.5 |
| | cg+$\Delta_{corr}$ | 38.3 | 58.2 | 52.3 | **62.2** | **42.5** | 24.3 |
| DeViSE-$g$AL | sg+Eq | - | 51.6 | 48.7 | - | 33.2 | - |
| | sg+$\Delta_{corr}$ | - | 54.9 | 50.1 | - | 34.8 | - |
| | cg+Eq | 31.9 | 58.2 | 48.7 | 55.6 | 32.6 | 17.5 |
| | cg+$\Delta_{corr}$ | 38.9 | 59.4 | 51.7 | 57.4 | 38.2 | **25.7** |
| SJE-$g$AL | sg+Eq | - | 54.3 | 50.5 | - | 34.0 | - |
| | sg+$\Delta_{corr}$ | - | 54.6 | 51.3 | - | 33.1 | - |
| | cg+Eq | 33.0 | **62.2** | 51.0 | 56.1 | 38.4 | 10.6 |
| | cg+$\Delta_{corr}$ | **40.5** | **62.2** | **53.2** | 60.3 | 40.2 | 23.9 |
| `softmax`-$g$AL | sg+Eq | - | 52.9 | 48.5 | - | 35.5 | - |
| | sg+$\Delta_{corr}$ | - | 54.5 | 49.3 | - | 35.4 | - |
| | cg+Eq | 37.1 | 61.6 | 50.4 | 59.9 | 40.5 | 14.2 |
| | cg+$\Delta_{corr}$ | 40.0 | 62.1 | 52.2 | 60.8 | 41.5 | 24.6 |

Table 4. Ablations for all $g$AL variants with best numbers in **bold**, second best numbers in blue. Here, sg: semantic groups. cg:cluster groups. Eq: equal weights to all adversarial branches. $\Delta_{corr}$:weights on adversarial branches proportional to $\Delta_{corr}$. Relevant numbers from Tables 2 and 3 are shown again for easy reference.




(a) SJE-$g$AL (b) `softmax`-$g$AL

Figure 4. Accuracies on AwA2 dataset with varying adversarial weight ($\lambda$) illustrates the importance of selecting the adversarial weight.

## A.2. Effect of adversarial weight $\lambda$

Choice of adversarial weight $\lambda$ (Eq.3) is crucial for performance of $g$AL models. In Fig.4, we observe how the test accuracy rises and drops as adversarial weight increases. This shows the trade-off between predicting classes and correcting correlation shift. Best value of adversarial weight is selected using validation accuracy. In Fig.4(b), the difference in performance of validation and test highlights the difficulty of finding the right value of adversarial weight.