

Figure 1a: Comparison of a linear probe on the CLIP representation against our CBM across different training sizes for the CUB dataset.

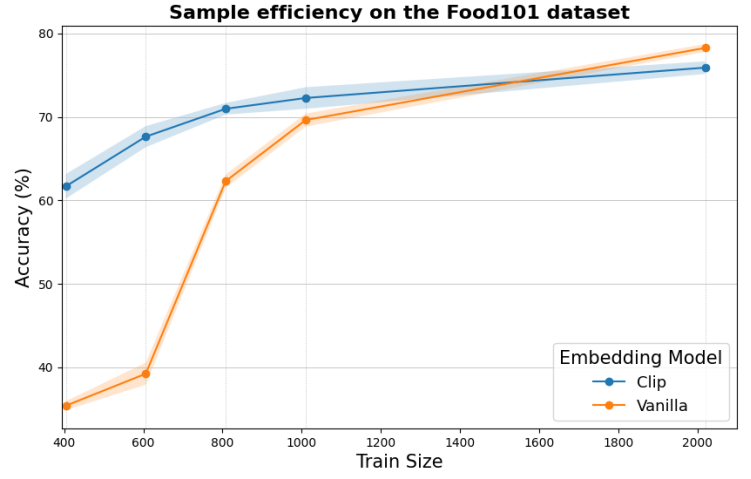


Figure 1b: Comparison of a linear probe on the CLIP representation against our CBM across different training sizes for the Food101 dataset.

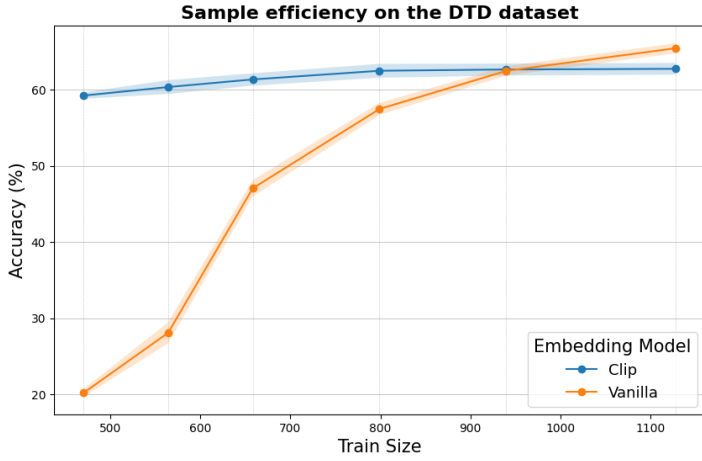


Figure 2c: Comparison of a linear probe on the CLIP representation against our CBM across different training sizes for the DTD dataset.

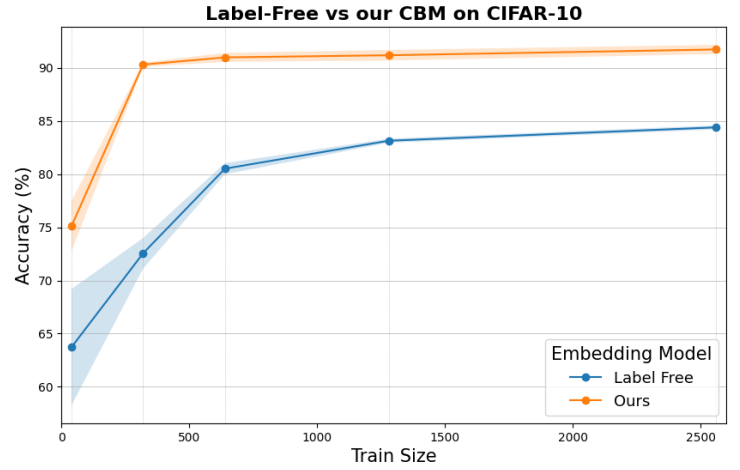


Figure 2: Performance comparison of the Label-Free CBM method against our CBM across different training sizes.

Model	Accuracy (%)
Vanilla	$70.4 \pm 1.2$
<b>CBM</b>	<b><math>78.8 \pm 0.7</math></b>

Table 1: OOD generalization comparison on CINIC-10 between a linear probe on the CLIP embedding and our CBM method.