Create dataset classifier from label text plane A photo of a bird car A photo of Text dog a {object}. Encoder A photo of a dog bird Use for zero-shot prediction T_1 T_2 T_3 T_N Image $I_1 \cdot T_2$ $I_1 \cdot T_3$ $I_1 \cdot T_N$ $I_1 \cdot T_1$ I_1 Encoder A photo of a plane A photo of a dog.