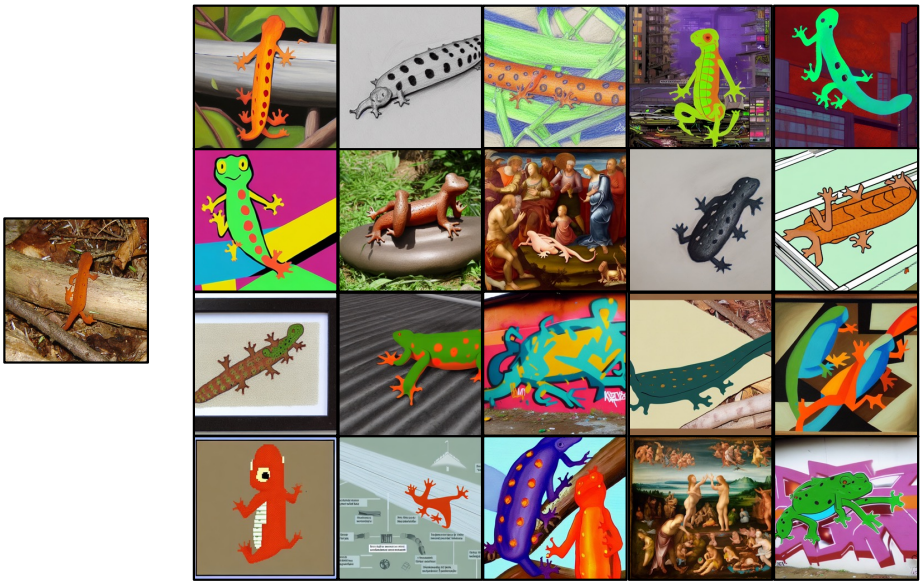

Figure 1: **Visualization of the object-centric training set in the AbC benchmark** The image on the left is an exemplar, and the 20 images on the right are its paired synthetic counterparts. In the AbC benchmark, the quality of synthetic images varies significantly. Some, such as the image in the top-left corner, offer strong supervisory signals. However, many are noisy—for example, the bottom-right image resembles a frog more than a lizard, potentially introducing ambiguity during training. We emphasize that these 20 samples represent only a small subset of the hundreds of synthetic images associated with the exemplar and do not fully capture the extent of noise present in the dataset.

Figure 2: **Visualization of attribution results.** Given a synthetic image generated by CustomDiffusion (based on Stable Diffusion), we retrieve its exemplars from a pool of 1 million LAION images. Alignment denotes attribution using the pretrained DINO model. Supervised refers to DINO finetuned with paired training data from the AbC benchmark. A&D is our proposed method, which performs unsupervised finetuning on DINO. Red bounding boxes indicate the ground-truth exemplars.