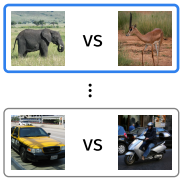
**1. Task template**

Define what each task consists of

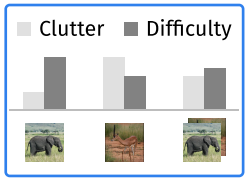
*Classifying images from two ImageNet classes*




**2. Task descriptors**

Quantify differences between tasks

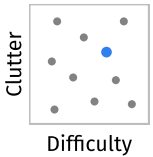
*Clutter and difficulty scores of the two classes*




**3. Exploration**

Select a diverse collection of tasks

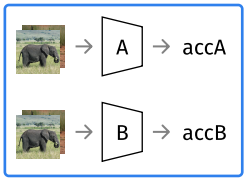
*Quasi-random selection of task-descriptor values*




**4. Evaluation**

Test new method vs baseline on each task

*Accuracy of attention CNN vs baseline CNN*




**5. Analysis**

Relate task descriptors to evaluation results

*Linear regression: descriptors vs accuracy*