

Problem Case



Question:
You are a taxonomist.
Answer the following
multiple-choice
questions about the
organism in the image.

L1: what is the taxonomic classification at ...
A. animalia

L2: what is the taxonomic classification at ...
A. mollusca
B. annelida
C. arthropoda
D. chordata

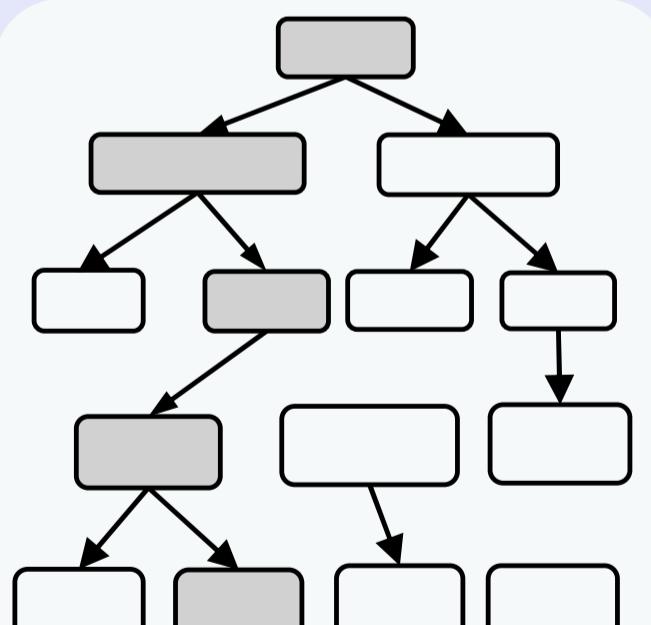
L3: what is the taxonomic classification at ...
A. mammalia
B. aves
C. clitellata
D. malacostraca
...

Problem Setup



Text Image

↑
Flatten into VQA



Hierarchical
Taxonomy Search

Forward Propagation and Answer Decoding

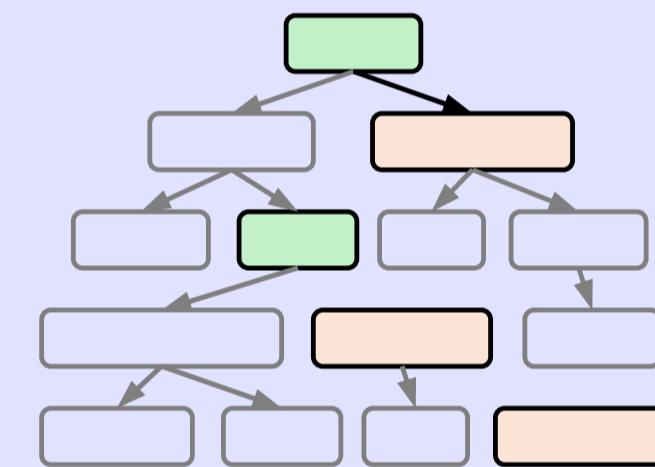
Origin VLM

These questions
keep getting
harder, and I
end up just
guessing each
answer.



Answer

L1: animalia
L2: arthropoda ✗
L3: aves ✗
L4: adapedonta ✗
L5: cathartidae ✗
L6: felis ✓
L7: catastictanimbice ✗



Self-Elicited Knowledge Distillation

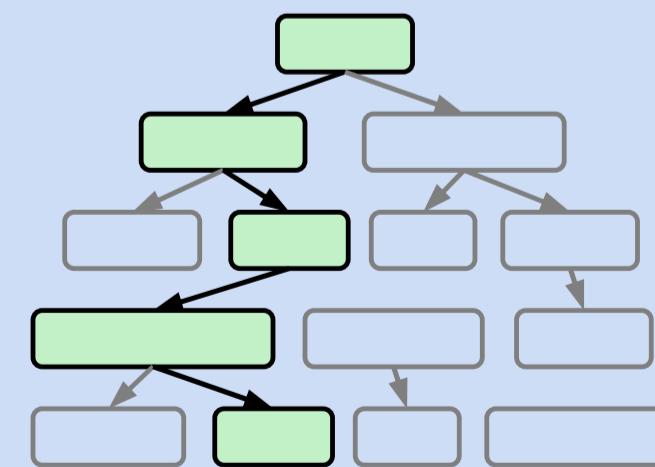
VLM after SKED

As the early
problems get
solved, the
later ones
naturally fall
into place.



Answer

L1: animalia
L2: chordata ✓
L3: mammalia ✓
L4: carnivora ✓
L5: felidae ✓
L6: felis ✓
L7: felis catus ✓



Let the VLM Teach Itself

We can solve
L1 first



Yes!

Next, go from
L1 to L2.



Finally, from
Ln-1 to Ln

