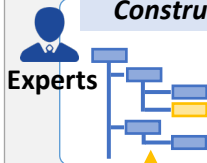
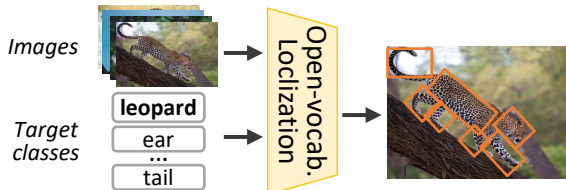


Visual Knowledge Extraction

Part Hierarchy Construction



Multi-granularity Localization



VLM Validation

Is the highlight region matches *[part]*?

head

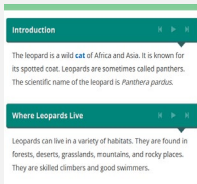


belly



...

Textual Knowledge Extraction



Encyclopedia Documents

Knowledge Type Pre-define



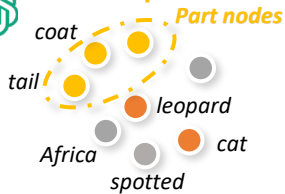
Experts

Visual

- Hava
- Num
- Color
-

Non-Visual

- Eat
- Provide
- Rarity
-



Entity Recognition

Non-visual Knowledge

<leopard, *BelongTo*, cat>
<leopard, *AtLocation*, Africa>

Visual Knowledge

<leopard, *Have*, coat>
<coat, *Pattern*, spotted>

Relation Extraction

Multi-Modal Alignment & Completion

