

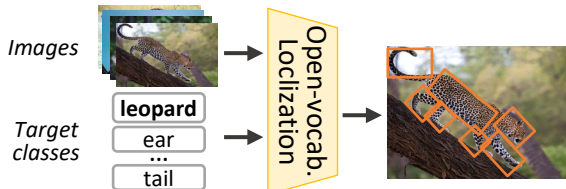
Image-source Knowledge Extraction

Part Hierarchy Construction

designer



Multi-granularity Localization



VLM Validation

Does the highlighted region match *[part]*?

head



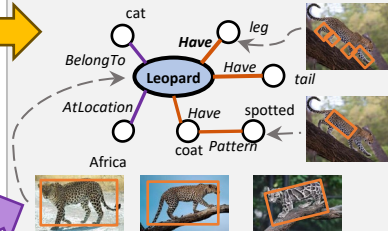
belly



...



Multi-Modal Alignment & Completion



Visual Knowledge Base

Text-source Knowledge Extraction



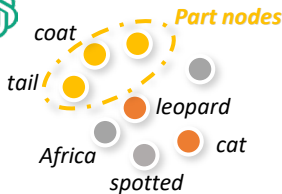
Knowledge Type Pre-definition

Visual

- Have
- 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

