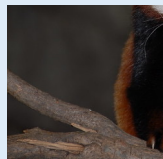
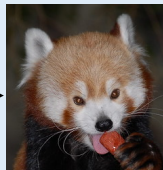
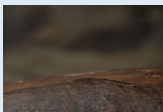


Input



...



*Region  
features*

*UARF*

fox

tree

hedge

branch

raincoat

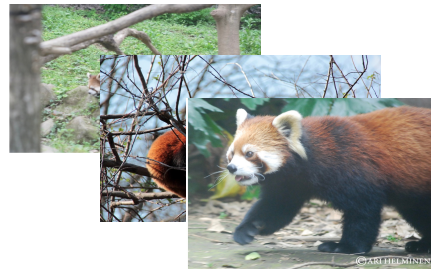
advertisement

bear\*\*

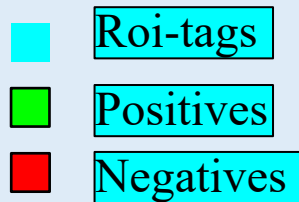
raccoon\*\*

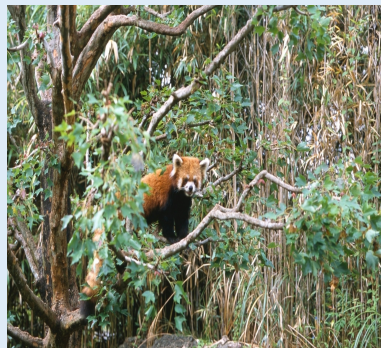
red panda

panda

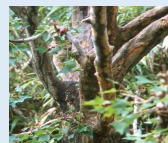


*Reference model*





Input



...



*Region  
features*

*UARF*

raccoon



tree



red panda

raincoat

hedge

advertisement

branch

panda

Bear\*\*



*Reference model*



Roi-tags

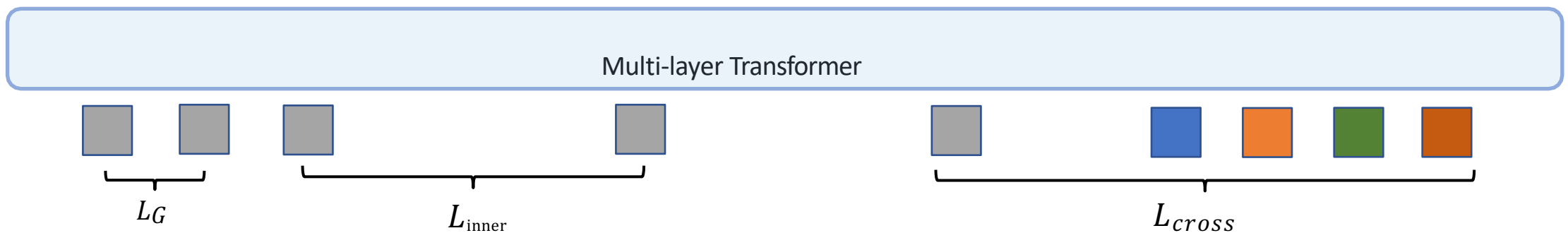
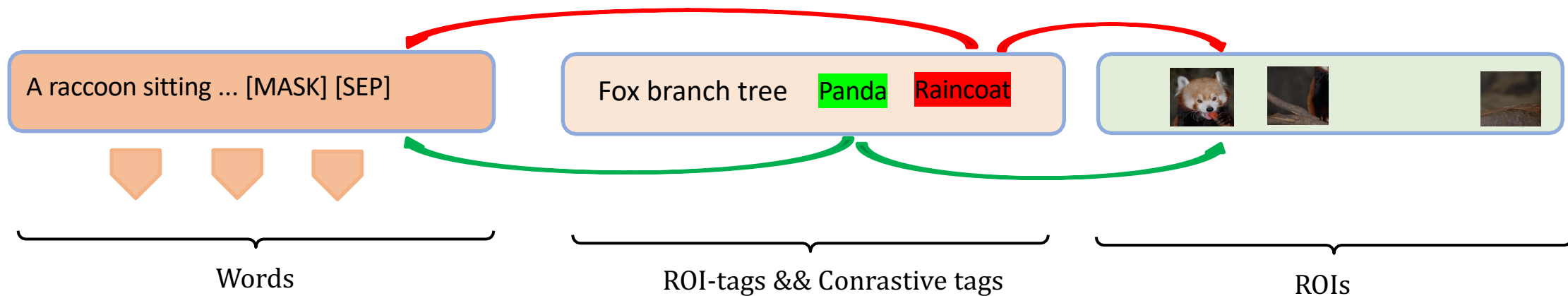


Positives



Negatives

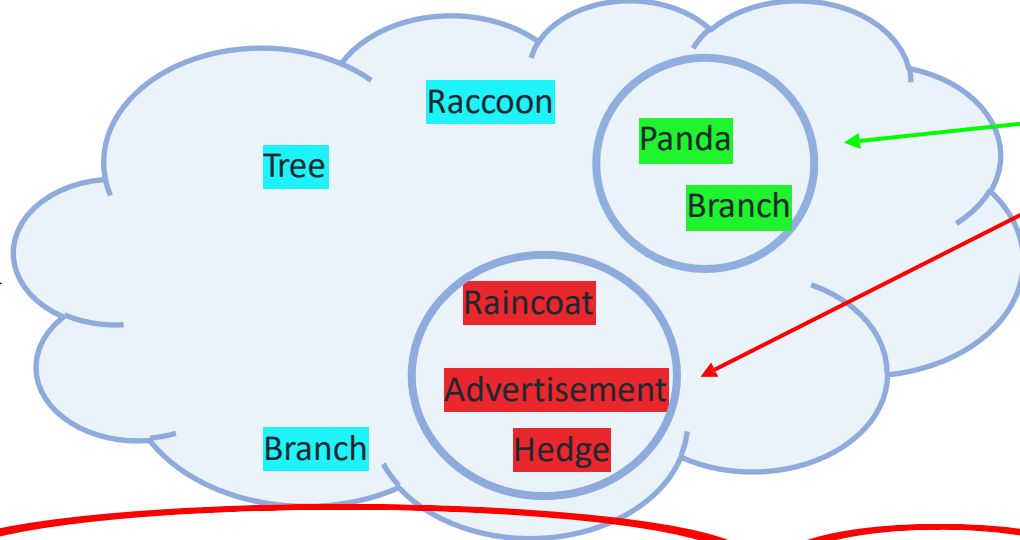
Input



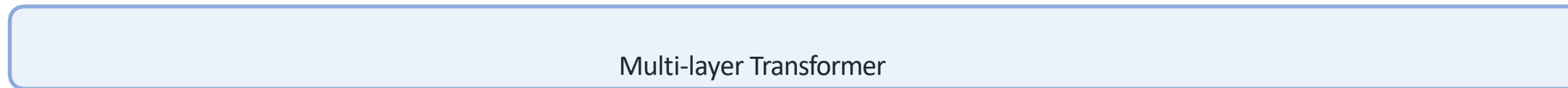
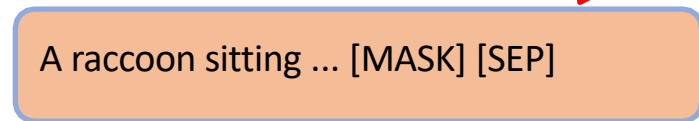
Output

A close up of a red panda on a tree branch

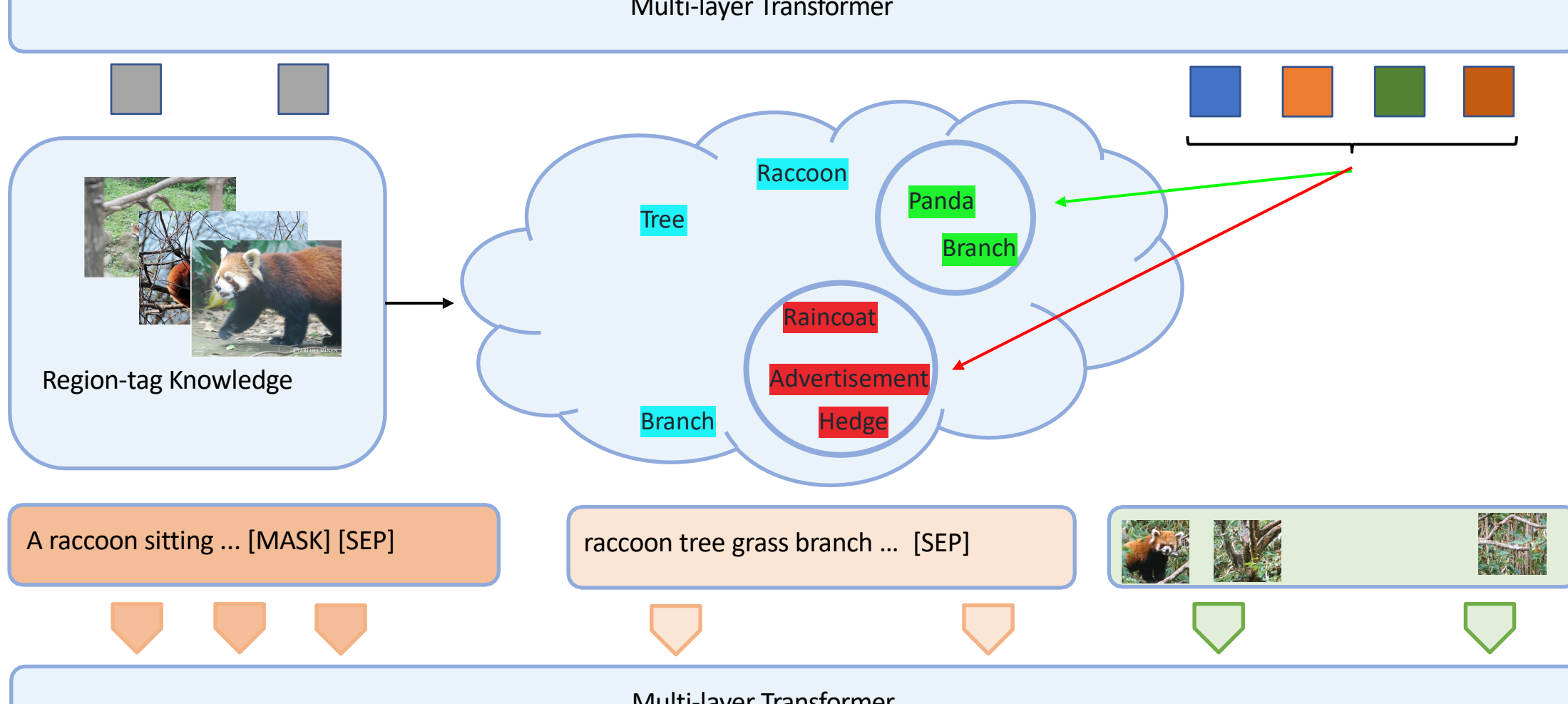
enever



step2:  
training with  
contrastive  
samples

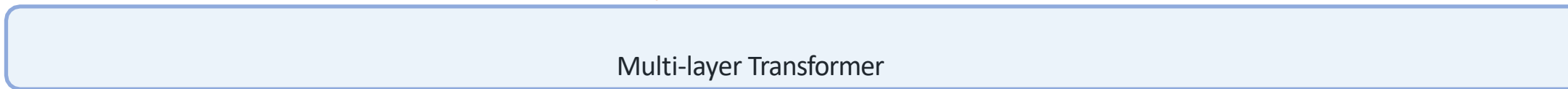
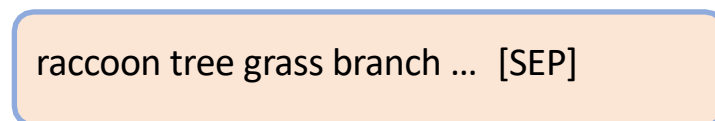
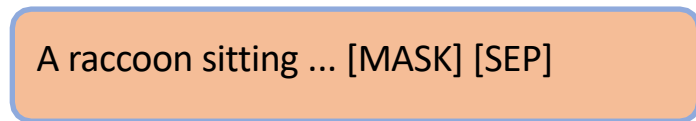
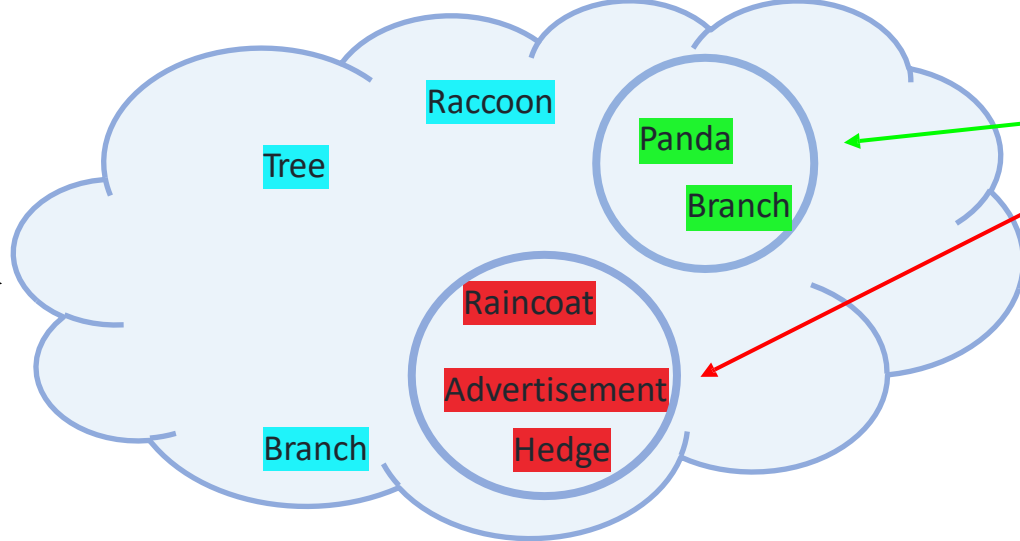


Context Aware  
Generation



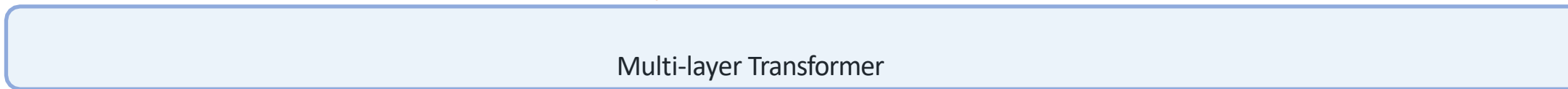
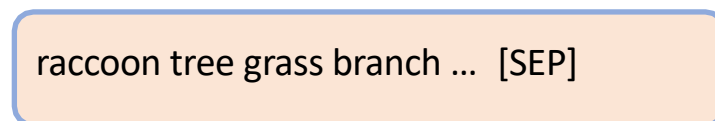
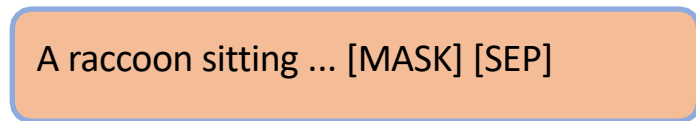
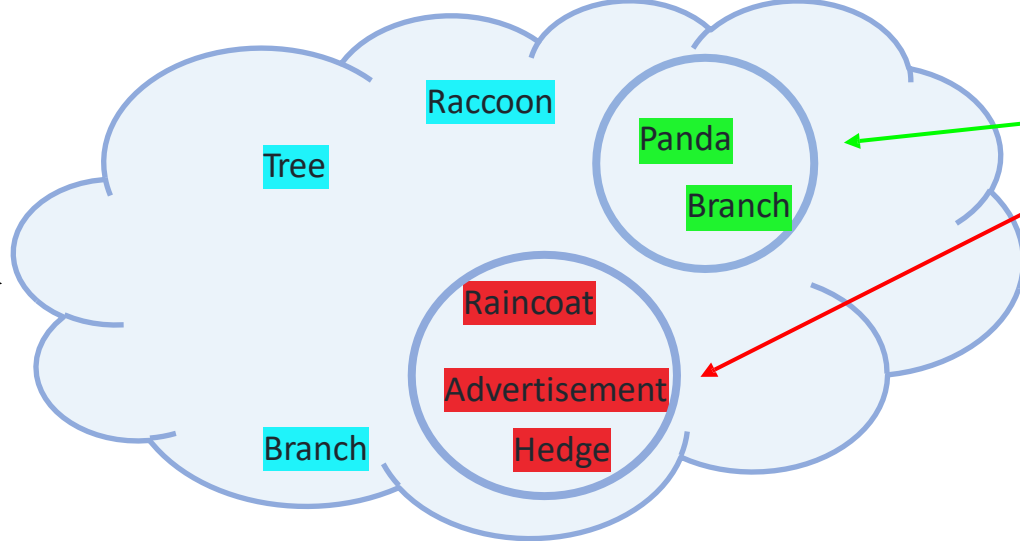
Step 2:  
Retuning with  
CL

Generation



2:  
tuning with  
L

Generation



2:  
tuning with  
L





Input



Tree

...



Branch

Raccoon

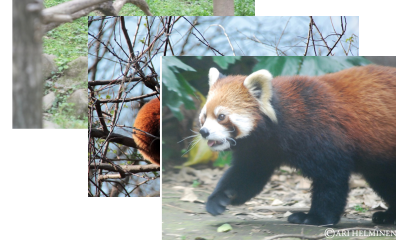
Panda

Branch

Raincoat

Hedge

Advertisement



Context Aware  
Gneration

Finetuning with NCCL



Raccoon

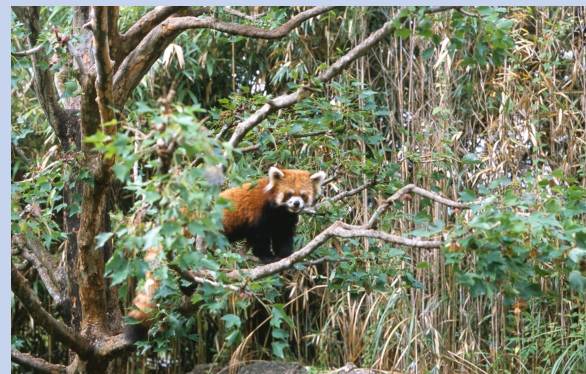
Grass

Wood

Panda

Broccoli

Inference



Raccoon

Tree

...

Panda

Branch

Forest

Hedge





Reference: *Red panda* sitting on the branch of a tree

VINVL+VIVO: A *raccoon* sitting in a tree.

Ours: A cute *red panda* sitting in a tree branch in a forest



Reference: A *red panda* sits on a tree branch with its eyes closed

VINVL+VIVO: a *raccoon* sitting in a tree with no leaves

Ours: a *red panda* sitting on top of a tree branch