

Intermediate results

IMAGE:



B:

[119,150,
210,298]

I':



A: gary F_A: yes

Intermediate results in the language modality

- (1) The description of input image ...
- (2) The coordinate of B is
- (3) The description of I' ...
- (4) The result of A in Step3 is gray
- (5) The final answer is Yes.

Result conversion

Prompt

1. Instruction
2. Plan of **all steps**
3. Program of **all steps**
4. Intermediate result of **all steps**



Global reflection

The error is caused by the
program/LOC/VQA of step ..

Learning
phase
to update
the model

The
updated
models still
fail

no
→ end

yes

plan, program, and results of Step1

Prompt for reflection of Step1

1. Instruction
2. Steps have been checked (**None**)
3. Step now need to be checked
(plan, program, and results of **Step1**)



The error is in the
program/LOC/VQA of this step

or

Do not find the error in this step.

Learning
phase
to update
the model

→ end

Prompt for reflection of Step i

1. Instruction
2. Steps have been checked
(plan, program, and results of **Step1-Step(i-1)**)
3. Step now need to be checked
(plan, program, and results of **Step i**)



plan, program, and results of Step i

Local reflection