5.1 Additional Implementation Details for QueryAdapter

Additional implementation details for QueryAdapter are presented in this section. Firstly, the prompts used with the Large Language Model (LLM) and MLLM are defined for reference. Secondly, the task definitions used for the experiments are defined.

5.1.1 Prompts

The following prompt was input to the LLM to decompose the natural language queries into a set of target classes. The field *used_objects* of the output JSON dictionary is used to define the target classes.

```
messages = [
    {"role": "system", "content": "Your job is to complete tasks in an
                                   indoor environment for a user. \
    Given a task, you need to define a feasible plan.\
    Output should be in the form of a json file only, do not include any
                                   other notes or explainations.\
    The user input will be a definition of a task.\
    The output should be a JSON dictionary defining how the task should be
                                   completed.\
    The field 'referenced_objects' should contain a list of the objects
                                   referred to in the task.\
    The field 'plan' should contain a description of the simplest method to
                                    fulfil the task.\
    The field 'used_objects' should contain a list of ALL the objects in
                                   the environment interacted with
                                   in the final plan.\
    The field 'affordances' should contain a short description of the form
                                   'used for <affordance>' to
                                   describe the common use case of
                                   each object."},
    {"role": "user", "content": task},
]
```

The following prompt was given to the MLLM to generate a caption for each object segment.

```
prompt = "USER: <image>\nDescribe the appearance of the central object in
                                    the image in one sentance, using the
                                    format 'an image of a <object
                                    description>'. ASSISTANT:"
```

The following prompt was given to the LLM to generate affordance queries for common object classes in the Scannet++ dataset.

```
messages = [
    {"role": "system", "content": "Provide the primary use case of indoor
                                   objects. Output in the format 'an
                                    object for' followed by the
                                   usecase. Do not include any other
                                    notes or explainations."},
    {"role": "user", "content": gt_class},
]
```

5.1.2 Natural Language Task Queries

The natural language task descriptions used to assess object retrieval are defined in Table 5.2. Using the most common classes in the Scannet++ dataset, the objects relevant to each task are also defined. Where necessary, synonyms for a particular relevant class are separated by an "or". For example, a cabinet or kitchen cabinet must be retrieved in response to the task *clean the cup in the sink and put it in the cabinet*. In this case, the robot is not expected to return both a cabinet and a kitchen cabinet. Instead, either class is considered correct.

To evaluate task-oriented object retrieval using the Ego4D dataset, the set of relevant classes is updated. This dataset is not as exhaustively labelled as Scannet++, so some tasks and relevant objects were ignored in this experiment. The task queries and relevant classes for the Ego4D dataset are defined in Table 5.3.

Lastly, the small sets of relevant objects used to optimise QueryAdapter are defined in Table 5.4. These were created by sampling every sixth object from the most common classes.

Task Query	Relevant Classes
clean the cup in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, cup
clean the plate in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, plate
clean the bowl in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, bowl
clean the pot in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, pot
clean the pan in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, pan
clean the bottle in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, bottle
clean the mug in the sink and put it in the cabinet	sink, cabinet or kitchen cabinet, mug
put the book away	book or books, bookshelf or book shelf or shelf
put the shoes away	shoe or shoes, shoe rack
clean the writing off the whiteboard	whiteboard, whiteboard eraser
draw a picture	paper or whiteboard, pen
water the plant with the bucket	plant or plant pot or potted plant, bucket
water the plant with the bottle	plant or plant pot or potted plant, bottle
let some light in from outside	window or windowsill, blind or blinds or window blind or curtain
get me a cup of tap water	cup, sink or tap
get me a bottle of tap water	bottle, sink or tap
use my laptop to play some music	laptop, speaker or headphones
make sure someone can sit at my desk	chair or office chair, desk
make sure someone can sit at the table	chair or dining chair, table
bring me something disposable to dry my hands, then throw it away	paper towel, trash can
bring me something disposable to clean the table, then throw it away	paper towel, trash can
put a chair somewhere warm for me to sit	chair or office chair or dining chair, heater or window or window sill
the plant is not getting enough light, move it to a better spot	plant or plant pot or potted plant, window or window sill
turn on the TV and make sure it is not too bright	tv, blind or blinds or window blind or light switch or ceiling lamp or ceiling light or table lamp or floor lamp
find me a book and make sure it is bright enough to read	book or books or bookshelf or book shelf, blind or blinds or window blind or light switch or ceiling lamp or ceiling light or table lamp or floor lamp
dispose of this box for me	box or crate or cardboard box, trash can
throw away this paper	paper, trash can
relocate the pillows so they are ready for bed time	pillow or pillows or cushion, bed

Table 5.2: Natural language task queries and relevant classes for the Scannet++ dataset.

Task Query	Relevant Classes
clean the cup in the sink and put it in the cabinet	cup
clean the plate in the sink and put it in the cabinet	plate
clean the bowl in the sink and put it in the cabinet	bowl
clean the pan in the sink and put it in the cabinet	pan_(for_cooking)
clean the bottle in the sink and put it in the cabinet	mug
clean the mug in the sink and put it in the cabinet	bottle
put the book away	book
put the shoes away	shoe or slipper_(footwear)
draw a picture	pen or pencil
water the plant with the bucket	bucket
water the plant with the bottle	bottle
get me a cup of tap water	cup
get me a bottle of tap water	bottle
use my laptop to play some music	laptop, earphone
make sure someone can sit at the table	chair, table
bring me something disposable to dry my hands, then throw it away	tissue paper or towel, trash can
bring me something disposable to clean the table, then throw it away	tissue paper or towel, trash can
put a chair somewhere warm for me to sit	chair
turn on the TV and make sure it is not too bright	television set, lamp
find me a book and make sure it is bright enough to read	book, lamp
dispose of this box for me	box or crate, trash can
throw away this paper	tissue paper, trash can
relocate the pillows so they are ready for bed time	pillow

Table 5.3: Natural language task queries and relevant classes for the Ego4D dataset.

No.	Target Classes
1	table, office chair, doorframe, trash can, jacket, backpack
2	door, bookshelf, pipe, book, electrical duct, crate
3	ceiling lamp, whiteboard, heater, plant, sink, keyboard
4	cabinet, window, kitchen cabinet, blanket, bag, rack
5	blinds, box, sofa, tv, picture, toilet
6	curtain, window frame, windowsill, computer tower, pillow, paper
7	chair, monitor, bed, kitchen counter, towel, printer
8	storage cabinet, shelf, shower wall, refrigerator, suitcase, poster

Table 5.4: Small sets of target classes defined using common classes from the Scannet++ dataset.