ASHiTA: Automatic Scene-grounded HIerarchical Task Analysis

Supplementary Material

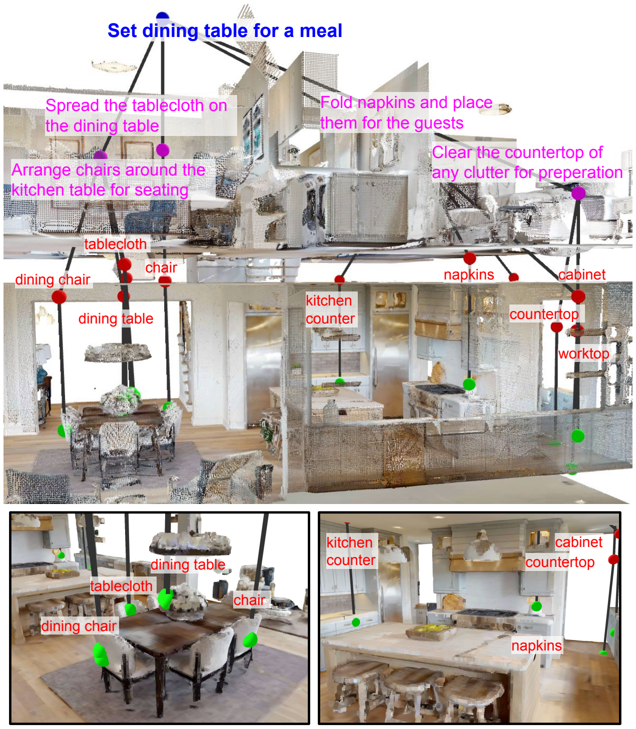


Figure 6. ASHiTA given the high-level task of "set dining table for a meal" from HM3DSem scene 00862-LT9Jq6dN3Ea [47]. Green markers are the primitives, red the items, magenta the subtasks, and blue the given high-level task. The bottom two frames show the zoomed in views of the scene graph and scene.

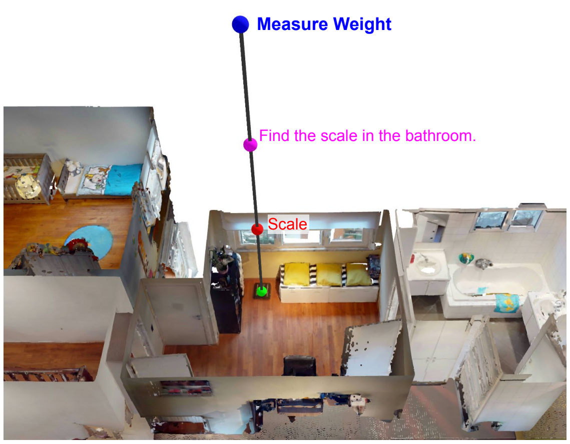


Figure 7. ASHiTA given the high-level task of "measure weight" from HM3D scene 00890-6s7QHgap2fW [47]. Green marks the primitive, red the item, magenta the subtask, and blue the given high-level task.

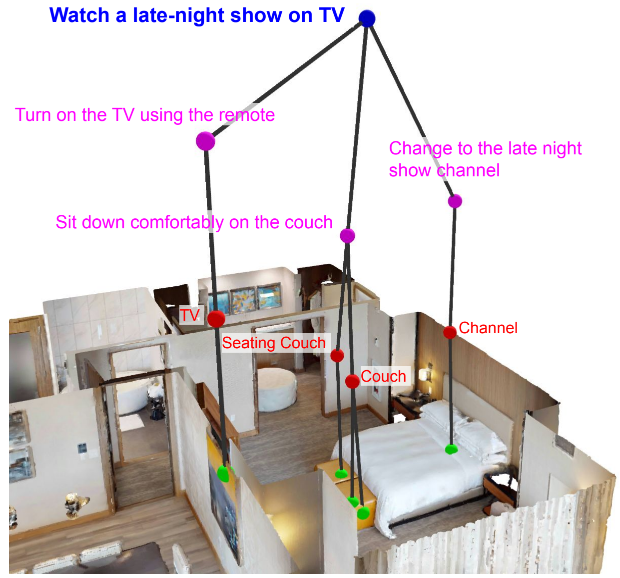


Figure 8. ASHiTA given the high-level task of "Watch a late-night show on TV" from HM3DSem scene 00829-QaLdnwvtxbs [47]. Green markers are the primitives, red the items, magenta the subtasks, and blue the given high-level task.

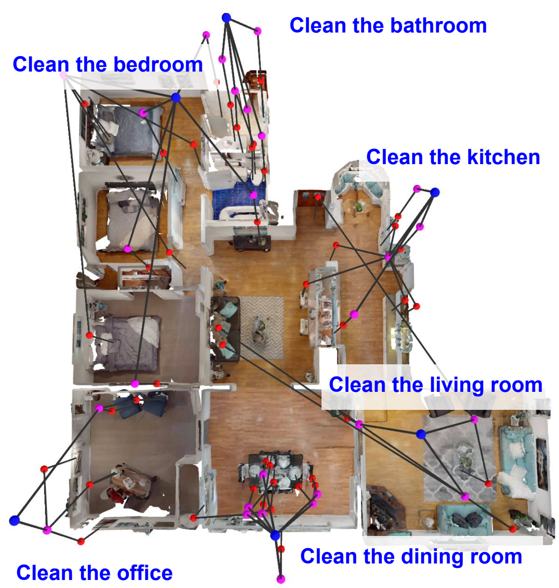
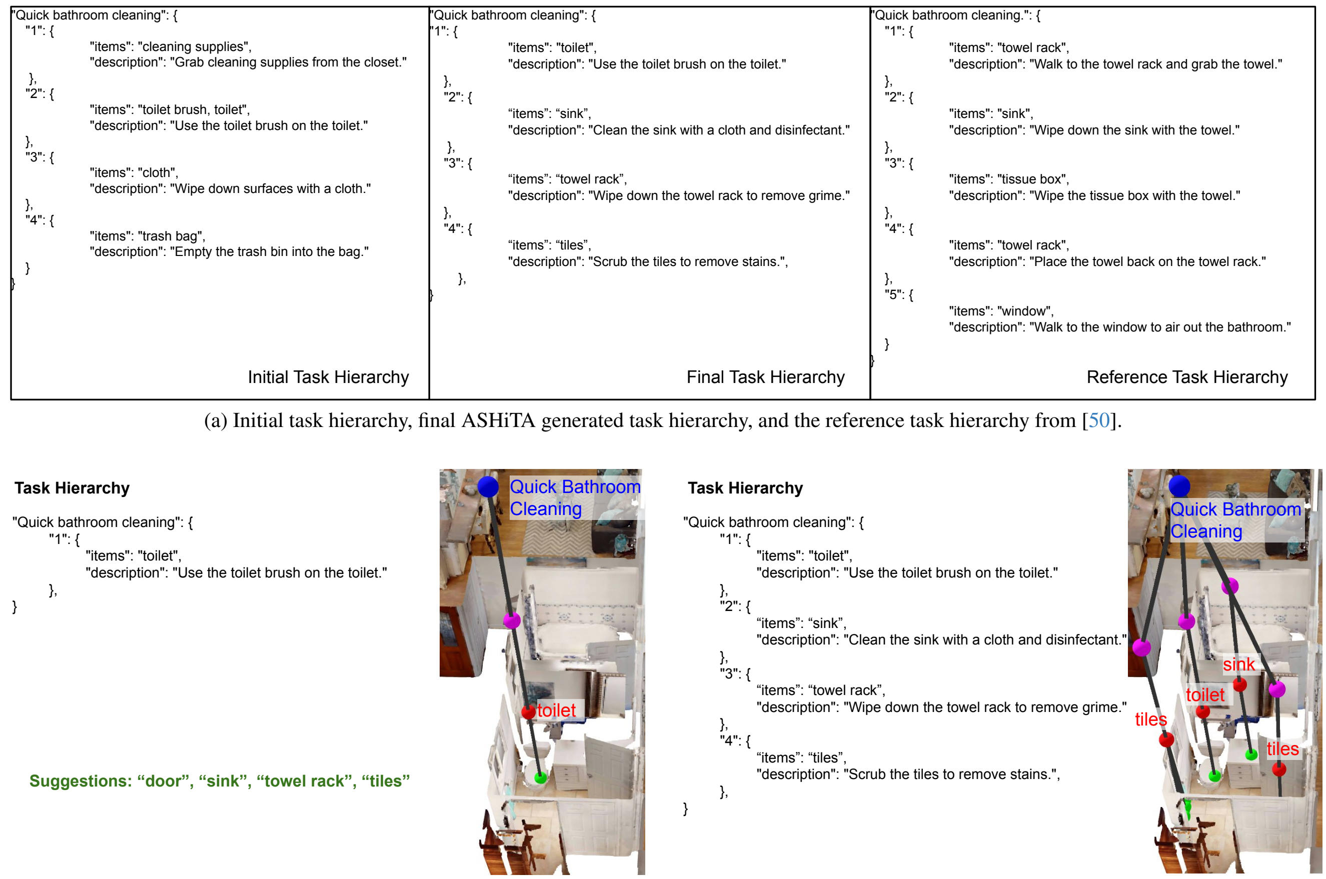


Figure 9. Given a set of high-level tasks that directly relate to the rooms in a house, we can approximately recover similar entities to that of prior scene graph construction approaches [14, 45]. For visualization clarity, we only show the labels of the high-level tasks. The purple nodes mark the subtasks and the red the associated objects.



(b) Task hierarchy and scene graph after first iteration.

(c) Task hierarchy and scene graph after second iteration.

Figure 10. ASHiTA with two iterations for the high-level task of "Quick bathroom cleaning". (b) From the initial hierarchy, only the toilet is successfully grounded in the generated scene graph. From this iteration, the suggestions of "door", "sink", "towel rack", and "tiles" are generated. These suggestions are used to update the task hierarchy, successfully recovering part of the human-annotated reference task hierarchy (recall that our evaluation is object-centric, as discussed in Sec. 5).

8. Qualitative Examples

8.1. SG3D Hierarchical Task Analysis

We include some qualitative examples of ASHiTA with the SG3D [50] high-level tasks in the HM3DSem [47] dataset. In Fig. 7, ASHiTA is given a single high-level task of "measure weight", and is able to correctly identify the scale. Note that ASHiTA is an object-centric approach and does not refine or evaluate on the specific *description* of the subtasks. In Fig. 6 we show ASHiTA at a larger scale in a kitchen and a dining room for "set dining table for a meal", ASHiTA comes up with reasonable subtasks and identified relevant items associated with the task. Lastly, in Fig. 8, the task given is "watch late-night show on TV", ASHiTA also mostly identifies the correct subtasks and items. A few mistakes are: selecting the TV in the neighboring dining room instead of the bedroom, and associating the non-item word "channel" to the bed. Note that ASHiTA is also able to cluster the two primitives (green) together for the "couch" item.

8.2. Detailed Example of ASHiTA

In this section, we demonstrate in detail the alternating iterations of ASHiTA with specific examples from the SG3D dataset. In Fig. 10 we show the comparison between the initial task hierarchy given to ASHiTA, the final task hierarchy, and the reference task hierarchy from SG3D [50] alongside two iterations of ASHiTA. In the first iteration, only the subtask related to the toilet is grounded in the scene graph, but incorporating the suggestions from this iteration, the task hierarchy is refined and we end up with four grounded subtasks. In Fig. 11 we give ASHiTA the task of "inspect the cleanliness of the utility area". With three iterations, ASHiTA is able to incorporate a more complete grounded set of relevant objects and subtasks.

8.3. Rooms and Objects

While ASHiTA is not designed to recover the traditional representations, such as rooms and objects, that other prior scene graph construction approaches do [12, 15, 45], with

```
"Inspect the cleanliness of the utility area.": {
"1": {
    "items": "utility area door",
    "description": "Open the utility area door."
 },
 "2": {
    "items": "floor",
    "description": "Check the floor for debris."
 },
 "3": {
    "items": "shelves",
    "description": "Examine shelves for dust accumulation."
 },
 "4": {
    "items": "counter",
    "description": "Inspect counter for spills."
 },
 "5": {
    "items": "bins",
    "description": "Look at bins for overflow."
 }
}
```

Initial Task Hierarchy

```
"Inspect the cleanliness of the utility area.": {
 "1": {
    "items": "utility area door, door",
    "description": "Open the utility area door.",
 },
 "2": {
    "items: "tile",
    "description": "Check the floor for debris.",
 },
 "3": {
    "items: "shelves",
    "description": "Examine shelves for dust.",
 },
 "4": {
    "items: "counter, sink, countertop",
    "description": "Inspect counter for spills.",
 },
 "5": {
    "items": "laundry",
    "description": "Check for any dirty laundry left in the area.",
 }
 "6": {
    "items: "washing machine, laundry machine",
    "description": "Inspect the washing machine for residue or lint buildup.",
 }
}
```

Final Task Hierarchy

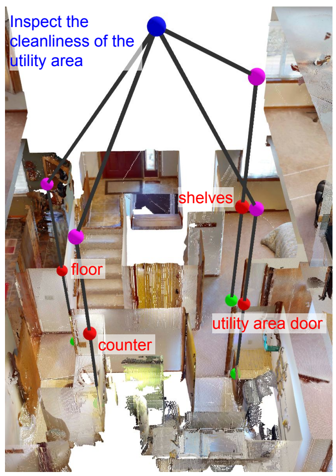
```
"Inspect the cleanliness of the utility area.": {
 "1": {
    "items": "washing machine",
    "description": "Walk to the washing machine next to the bucket."
 },
 "2": {
    "items": "washing machine",
    "description": "Check if the area around the washing machine is clean and free of lint or spills."
 },
 "3": {
    "items": "bucket",
    "description": "Go to the bucket next to the washing machine."
 },
 "4": {
    "items": "bucket",
    "description": "Ensure that the inside of the bucket is clean and dry."
 },
 "5": {
    "items": "counter",
    "description": "Check the counter area near the radiator to ensure it does not have any dust or
clutter."
 },
 "6": {
    "items": "shelf",
    "description": "Examine the shelf above the counter to verify if it is organized and free of dust."
 }
}
```

Reference Task Hierarchy

(a) Initial task hierarchy, final ASHiTA generated task hierarchy, and the reference task hierarchy from [50].

```
"Inspect the cleanliness of the utility area.": {
    "1": {
        "items": "utility area door",
        "description": "Open the utility area door.",
    },
    "2": {
        "items: "floor",
        "description": "Check the floor for debris.",
    },
    "3": {
        "items: "shelves",
        "description": "Examine shelves for dust.",
    },
    "4": {
        "items: "counter",
        "description": "Inspect counter for spills.",
    }
}
```

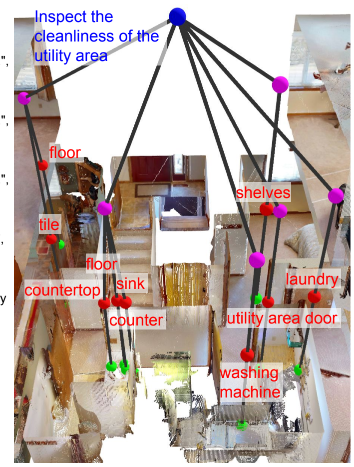
Suggestions: "door", "floor mat", "surface", "shelf", "sink", "countertop", "bath carpet", "washing machine", "laundry", ...



(b) Task hierarchy and scene graph after first iteration.

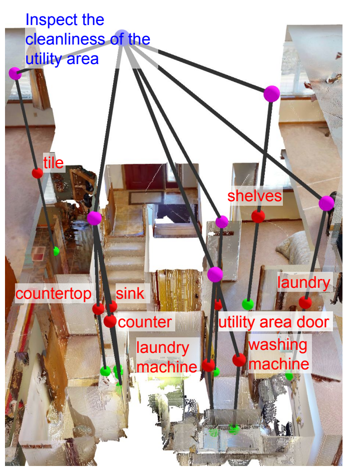
```
"Inspect the cleanliness of the utility area.": {
    "1": {
        "items": "utility area door, door",
        "description": "Open the utility area door.",
    },
    "2": {
        "items: "floor, tile",
        "description": "Check the floor for debris.",
    },
    "3": {
        "items: "shelves",
        "description": "Examine shelves for dust.",
    },
    "4": {
        "items: "counter, sink, countertop",
        "description": "Inspect counter for spills.",
    },
    "5": {
        "items": "laundry",
        "description": "Check for any dirty laundry
left in the area.",
    }
    "6": {
        "items: "washing machine",
        "description": "Inspect the washing
machine for residue or lint buildup.",
}
```

Suggestions: "towel", "clothes dryer", "bucket", "trash can", "laundry machine"



(c) Task hierarchy and scene graph after second iteration.

```
"Inspect the cleanliness of the utility area.": {
    "1": {
        "items": "utility area door, door",
        "description": "Open the utility area door.",
    },
    "2": {
        "items: "tile",
        "description": "Check the floor for debris.",
    },
    "3": {
        "items: "shelves",
        "description": "Examine shelves for dust.",
    },
    "4": {
        "items: "counter, sink, countertop",
        "description": "Inspect counter for spills.",
    },
    "5": {
        "items": "laundry",
        "description": "Check for any dirty laundry
left in the area.",
    }
    "6": {
        "items: "washing machine, laundry
machine",
        "description": "Inspect the washing
machine for residue or lint buildup.",
}
```



(d) Task hierarchy and scene graph after third iteration.

(e) Zoomed in view on the primitives and the scene.

Figure 11. ASHiTA with three iterations for the high-level task of "Inspect the cleanliness of the utility area". (b) From the initial hierarchy, all items except for the bins are grounded in the generated scene graph. Many suggestions are generated to update the hierarchy. (c) Suggested items "laundry" and "washing machine" are used to generate new subtasks, these are also successfully grounded in the generated scene graph. Grounding these items trigger a few additional suggestions. (d) Suggested items "laundry machine" added to the task hierarchy. This accounts for the clothes dryer in the scene. Our estimated final task hierarchy recalls all of the grounded items in the reference task hierarchy except for the bucket.

Initial Hierarchy	Method	s-rec (%)	s-prec (%)	t-acc (%)
1	ASHiTA	10.39	20.6	9.27
	ASHiTA (GT Pos + Txt Emb)	38.71	34.39	36.1
2	ASHiTA	9.95	20.41	7.8
	ASHiTA (GT Pos + Txt Emb)	40.56	35.38	37.56
Privileged	ASHiTA	14.3	17.0	12.2
	ASHiTA (GT Pos + Txt Emb)	42.13	38.68	42.93

Table 5. ASHiTA and ASHiTA with ground-truth objects and labels with different initial task hierarchies. Top two rows are generated with ChatGPT given only the abstract tasks; the last row was generated given ground-truth labels of known objects in the scene.

a specially chosen set of high-level tasks, we can approximately recover a similar set of rooms and objects. This is shown in Fig 9, where ASHiTA is given 6 high-level cleaning tasks related to the types of rooms in the scene. However, it is clear that ASHiTA lacks the understanding of structures and the overall floor plan, and only retains the entities that are deemed relevant to the given tasks.

9. Initial Hierarchy Ablations

In Section 5, we use `GPT-4o-mini` to generate the initial task hierarchy by first giving `GPT-4o-mini`a manually generated task hierarchy for some arbitrary task as an example then querying with the prompt:

> "Given the example above, generate a concise task hierarchy for <task>, ensuring brief and clear descriptions."

Here we include an ablation to evaluate the impact of the initial task hierarchy. Using three different ChatGPT generated initial hierarchies including one privileged with the inclusion of objects in the scene as priors. The results are shown in Table 5. The impact of the initial task hierarchy is minimal, even with privileged priors, and does not change the reported trend of the results shown in Table 3.

10. Derivation of the Hierarchical Information Bottleneck

In this section, we detail the derivation of the iterative multi-layer update steps (4) used to minimize the H-IB functional in (3), following the same approach as the original IB derivation outlined in [42]. To do this, we first formulate the Lagrangian for H-IB along with the derivative of the Lagrangian. Next, we express the derivatives of the conditional probabilities used in H-IB conditioned on the Markov chain assumption. These expressions can be substituted into the derivative of the Lagrangian, which allows us to solve for the zero of the Lagrangian derivative.

Accounting for the constraint that $\mathbb{P}(S_k|S_k-1)$ is a valid probability, the Lagrangian of (3) can be written as,

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{n}\{I(\mathcal{S}_{i-1};\mathcal{S}_i) - \beta I(\mathcal{T}_i;\mathcal{S}_i) \\ + \sum_{s_{i-1}\in\mathcal{S}_{i-1}} \lambda(s_{i-1})[\sum_{s_i\in\mathcal{S}_i} p(s_i|s_{i-1}) - 1]\}\end{aligned} \tag{11}$$

Recalling the definition of mutual information

$$\begin{aligned}I(X;Y) &= \sum_x\sum_y p(x,y)log(\frac{p(x,y)}{p(x)p(y)}) \\ &= \sum_x\sum_y p(x|y)p(y)log(\frac{p(x|y)}{p(x)})\end{aligned} \tag{12}$$

using the logarithm properties, we expand (11) as,

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{n}\{ \\ \sum_{s_{i-1}} p(s_{i-1})\sum_{s_i} p(s_i|s_{i-1})[log(p(s_i|s_{i-1})) - log(p(s_i))] \\ - \beta\sum_{t_i} p(t_i)\sum_{s_i} p(s_i|t_i)[log(p(s_i|t_i)) - log(p(s_i))] \\ + \sum_{s_{i-1}} \lambda(s_{i-1})[\sum_{s_i} p(s_i|s_{i-1}) - 1)]\}\end{aligned} \tag{13}$$

Our goal is to derive $p(s_k|s_{k-1})$ for some arbitrary level k for some integer $k \in [1, n]$. We rewrite and expand (13) and break up the sum of the levels into three parts: from level 1 to $k-1$, level k, and from level $k+1$ to level n. The Markov chain assumption designate that the levels lower than k are not dependent on $p(s_k|s_{k-1})$, level k is directly dependent on $p(s_k|s_{k-1})$, and the higher levels are indirectly dependent on $p(s_k|s_{k-1})$. This means that when we take the derivative of the Lagrangian with respect to $p(s_k|s_{k-1})$ for fixed s_k and s_{k-1}, the terms related to the first part are zero. This broken up expression is as follows, arranged in order of the three terms in (11), and for each term, broken up into three parts based on k as described,

$$
\begin{aligned}
\mathcal{L} = &\sum_{i=1}^{k-1}\{ \\
&\sum_{s_{i-1}} p(s_{i-1}) \sum_{s_i} p(s_i|s_{i-1})[log(p(s_i|s_{i-1})) - log(p(s_i))]\} \\
&+ \sum_{s_{k-1}} p(s_{k-1}) \sum_{s_k} p(s_k|s_{k-1})[log(p(s_k|s_{k-1})) - log(p(s_k))] \\
&+ \sum_{s_k} p(s_k) \sum_{s_{k+1}} p(s_{k+1}|s_k)[log(p(s_{k+1}|s_k)) - log(p(s_{k+1}))] \\
&+ \sum_{i=k+2}^{n}\{ \\
&\sum_{s_{i-1}} p(s_{i-1}) \sum_{s_i} p(s_i|s_{i-1})[log(p(s_i|s_{i-1})) - log(p(s_i))]\} \\
&-\beta \sum_{i=1}^{k-1}\{\sum_{t_i} p(t_i) \sum_{s_i} p(s_i|t_i)[log(p(s_i|t_i)) - log(p(s_i))]\} \\
&-\beta \sum_{t_k} p(t_k) \sum_{s_k} p(s_k|t_k)[log(p(s_k|t_k)) - log(p(s_k))] \\
&-\beta \sum_{i=k+1}^{n}\{\sum_{t_i} p(t_i) \sum_{s_i} p(s_i|t_i)[log(p(s_i|t_i)) - log(p(s_i))]\} \\
&+ \sum_{i=1}^{k-1}\{\sum_{s_{i-1}} \lambda(s_{i-1})[\sum_{s_i} p(s_i|s_{i-1}) - 1)]\} \\
&+ \sum_{s_{k-1}} \lambda(s_{k-1})[\sum_{s_k} p(s_k|s_{k-1}) - 1)]\} \\
&+ \sum_{i=k+1}^{n}\{\sum_{s_{i-1}} \lambda(s_{i-1})[\sum_{s_i} p(s_i|s_{i-1}) - 1)]\}
\end{aligned}
\tag{14}
$$

We can then take the derivative of the Lagrangian $\frac{\delta \mathcal{L}}{\delta p(s_k|s_{k-1})}$ for fixed s_k and s_{k-1}. From basic calculus, we can derive that

$$
\frac{d(f(x)log(f(x)))}{dx} = \frac{df(x)}{dx}(log(f(x)) + 1) \tag{15}
$$

The terms related to the first part (level 1 to $k-1$) are zero as explained above. Using the chain rule along with (15), the derivative of the Lagrangian (14) is,

$$
\begin{aligned}
&\frac{\delta \mathcal{L}}{\delta p(s_k|s_{k-1})} \\
=&p(s_{k-1})[log(p(s_k|s_{k-1})) + 1] \\
&- \frac{\delta p(s_k)}{\delta p(s_k|s_{k-1})}[log(p(s_k)) + 1] \\
+&\frac{\delta p(s_k)}{\delta p(s_k|s_{k-1})} \sum_{s_{k+1}} p(s_{k+1}|s_k) log(p(s_{k+1}|s_k)) \\
&- \sum_{s_{k+1}} \frac{\delta p(s_{k+1})}{\delta p(s_k|s_{k-1})}[log(p(s_{k+1})) + 1] \\
+& \sum_{i=k+2}^{n}\{\sum_{s_{i-1}} \frac{\delta p(s_{i-1})}{\delta p(s_k|s_{k-1})} \sum_{s_i} p(s_i|s_{i-1}) log(p(s_i|s_{i-1})) \\
&- \sum_{s_i} \frac{\delta p(s_i)}{\delta p(s_k|s_{k-1})}[log(p(s_i)) + 1]\} \\
-&\beta\{\sum_{t_k} p(t_k) \frac{\delta p(s_k|t_k)}{\delta p(s_k|s_{k-1})}[log(p(s_k|t_k)) + 1] \\
&- \frac{\delta p(s_k)}{\delta p(s_k|s_{k-1})}[log(p(s_k)) + 1])\} \\
-&\beta \sum_{i=k+1}^{n}\{\sum_{t_i} p(t_i) \sum_{s_i} \frac{\delta p(s_i|t_i)}{\delta p(s_k|s_{k-1})}[log(p(s_i|t_i)) + 1] \\
&- \sum_{s_i} \frac{\delta p(s_i)}{\delta p(s_k|s_{k-1})}[log(p(s_i)) + 1])\} + \lambda(s_{k-1})
\end{aligned}
\tag{16}
$$

Let us now derive for the expressions of the derivatives of the conditional probabilities. Since each scene level is a strict compression of the previous level, we have the Markov chain condition $\mathcal{T}_i \leftarrow \mathcal{S}_0 \leftarrow \mathcal{S}_1 \leftarrow \ldots \leftarrow \mathcal{S}_n$ for all resolutions of the task description $\mathcal{T}_i$. The conditional distributions for the first two levels are,

$$
p(s_1) = \sum_{s_0 \in \mathcal{S}_0} p(s_1|s_0)p(s_0) \tag{17}
$$

$$
p(s_1|t_1) = \sum_{s_0 \in \mathcal{S}_0} p(s_1|s_0)p(s_0|t_1) \tag{18}
$$

$$
p(s_2) = \sum_{s_1 \in \mathcal{S}_1} \sum_{s_0 \in \mathcal{S}_0} p(s_2|s_1)p(s_1|s_0)p(s_0) \tag{19}
$$

$$
p(s_2|t_2) = \sum_{s_1 \in \mathcal{S}_1} \sum_{s_0 \in \mathcal{S}_0} p(s_2|s_1)p(s_1|s_0)p(s_0|t_2) \tag{20}
$$

Generalized for level n and some k such that $n > k > 0$,

$$
\begin{aligned}
&p(s_n) = \\
&\sum_{s_k \in \mathcal{S}_k} \sum_{s_{k-1} \in \mathcal{S}_{k-1}} p(s_n|s_k)p(s_k|s_{k-1})p(s_{k-1})
\end{aligned}
\tag{21}
$$

$$p(s_n|t_n) = \sum_{s_k \in \mathcal{S}_k} \sum_{s_{k-1} \in \mathcal{S}_{k-1}} p(s_n|s_k)p(s_k|s_{k-1})p(s_{k-1}|s_0)p(s_0|t_n) \tag{22}$$

Taking the derivative of the conditional distributions with respect to $p(s_1|o) \dots p(s_k|s_{k-1})$,

$$\frac{\delta p(s_1)}{\delta p(s_1|s_0)} = p(s_0) \tag{23}$$

$$\frac{\delta p(s_1|t_1)}{\delta p(s_1|s_0)} = p(s_0|t_1) \tag{24}$$

$$\frac{\delta p(s_n)}{\delta p(s_k|s_{k-1})} = p(s_n|s_k)p(s_{k-1}) \tag{25}$$

$$\frac{\delta p(s_n|t_n)}{\delta p(s_k|s_{k-1})} = p(s_n|s_k)p(s_{k-1}|s_0)p(s_0|t_n) \tag{26}$$

Substituting in the expressions (25) and (26) into the derivative of the Lagrangian (16),

$$\begin{aligned}
&\frac{\delta \mathcal{L}}{\delta p(s_k|s_{k-1})} \\
=&p(s_{k-1})\{log(\frac{p(s_k|s_{k-1})}{p(s_k)}) \\
&+ \sum_{s_{k+1}} p(s_{k+1}|s_k)log(\frac{p(s_{k+1}|s_k)}{p(s_{k+1})}) \\
&+ \sum_{i=k+2}^{n} \{\sum_{s_{i-1}} p(s_{i-1}|s_k) \sum_{s_i} p(s_i|s_{i-1})log(\frac{p(s_i|s_{i-1})}{p(s_i)})\} \\
&- (n-1) \\
&-\beta\{\sum_{t_k} p(t_k)p(s_{k-1}|o)p(o|t_k)[log(p(s_k|t_k)) + 1] \\
&- p(s_{k-1})[log(p(s_k)) + 1]\} \\
&-\beta \sum_{i=k+1}^{n} \{ \\
&\sum_{t_i} p(t_i) \sum_{s_i} p(s_i|s_k)p(s_{k-1}|s_0)p(s_0|t_i)[log(p(s_i|t_i)) + 1] \\
&- \sum_{s_i} p(s_i|s_k)p(s_{k-1})[log(p(s_i)) + 1]\}\} + \lambda(s_{k-1})
\end{aligned} \tag{27}$$

We can rewrite $p(t_k)p(s_{k-1}|o)p(o|t_k)$ as

$$p(t_k)p(s_{k-1}|t_k) = p(s_{k-1})p(t_k|s_{k-1}) \tag{28}$$

which allows us to simplify (27) further as,

$$\begin{aligned}
&\frac{\delta \mathcal{L}}{\delta p(s_k|s_{k-1})} \\
=&p(s_{k-1})\{log(\frac{p(s_k|s_{k-1})}{p(s_k)}) \\
&+ \sum_{s_{k+1}} p(s_{k+1}|s_k)log(\frac{p(s_{k+1}|s_k)}{p(s_{k+1})}) \\
&+ \sum_{i=k+2}^{n} \sum_{s_{i-1}} p(s_{i-1}|s_k) \sum_{s_i} p(s_i|s_{i-1})log(\frac{p(s_i|s_{i-1})}{p(s_i)}) \\
&- (n-1) \\
&-\beta \sum_{t_k} p(t_k|s_{k-1})log(\frac{p(s_k|t_k)}{p(s_k)}) \\
&-\beta \sum_{i=k+1}^{n} \sum_{t_i} \sum_{s_i} p(s_i|s_k)p(t_i|s_{k-1})log(\frac{p(s_i|t_i)}{p(s_i)})\} \\
&+ \lambda(s_{k-1})
\end{aligned} \tag{29}$$

Notice that the Kullback–Leibler divergence naturally emerges from the β terms with some algebraic manipulation,

$$\begin{aligned}
&p(t_i|s_{k-1})log(\frac{p(s_i|t_i)}{p(s_i)}) = -D_{KL}(p(t_i|s_{k-1})||p(t_i|s_i)) \\
&+ \sum_{t_i} p(t_i|s_{k-1})log(\frac{p(t_i|s_{k-1}}{p(t_i)}))
\end{aligned} \tag{30}$$

Let us define $\tilde{\lambda}(s_{k-1})$ to group the terms that are not dependent on s_k,

$$\begin{aligned}
\tilde{\lambda}(s_{k-1}) =& \frac{\lambda(s_{k-1})}{p(s_{k-1})} - (n-1) \\
&+ \sum_{s_{k+1}} p(s_{k+1}|s_k)log(\frac{p(s_{k+1}|s_k)}{p(s_{k+1})}) \\
&+ \sum_{i=k+2}^{n} \sum_{s_{i-1}} p(s_{i-1}|s_k) \sum_{s_i} p(s_i|s_{i-1})log(\frac{p(s_i|s_{i-1})}{p(s_i)}) \\
&- \beta \sum_{t_i} p(t_i|s_{k-1})log(\frac{p(t_i|s_{k-1})}{p(t_i)})) \\
&- \beta \sum_{i=k+1}^{n} \sum_{s_i} \sum_{t_i} p(t_i|s_{k-1})log(\frac{p(t_i|s_{k-1})}{p(t_i)}))\}
\end{aligned} \tag{31}$$

Setting the derivative (29) to zero then gives us,

$$\begin{aligned}
0 &= p(s_{k-1})\{log(\frac{p(s_k|s_{k-1})}{p(s_k)}) \\
&+\beta D_{KL}(p(t_k|s_k)||p(t_k|s_{k-1})) \\
&+\beta \sum_{i=k+1}^{n} \sum_{s_i} p(s_i|s_k) D_{KL}(p(t_i|s_i)||p(t_i|s_{k-1})) \\
&+ \tilde{\lambda}(s_{k-1})\}
\end{aligned} \tag{32}$$

Defining $\mathcal{Z} = \exp[\tilde{\lambda}(s_{k-1})]$, we have that

$$\begin{aligned}
p(s_k|s_{k-1}) &= \frac{p(s_k)}{\mathcal{Z}} \exp[-\beta D_{KL}(p(t_k|s_k)||p(t_k|s_{k-1})) \\
&- \beta \sum_{i=k+1}^{n} \sum_{s_i} p(s_i|s_k) D_{KL}(p(t_i|s_i)||p(t_i|s_{k-1}))]
\end{aligned} \tag{33}$$

This corresponds to the iterative algorithm given in (4).

11. Tutorial on the Hierarchical Information Bottleneck

In this section, we provide a brief tutorial on H-IB in the form of an easy example. Given two tasks: $\mathcal{T} = \{\Gamma, \Omega\}$, three subtasks: $\mathcal{U} = \{A, B, C\}$, and four items: $\mathcal{O} = \{p, q, r, s\}$, We want to assign each of 5 observations $\mathcal{X}$ to an item, each item to a subtask, and each subtask to a task, and we are given the probability of how likely an observation might be that of an item $\mathbb{P}(\mathcal{O}|\mathcal{X})$, the probability that an item might be relevant for a subtask $\mathbb{P}(\mathcal{U}|\mathcal{O})$, and finally the probability that a subtask might be relevant for a task $\mathbb{P}(\mathcal{T}|\mathcal{U})$. For this exercise, the conditional probabilities are given in Table 6, Table 7, and Table 8. We treat each observation as identical and independent so that $\mathbb{P}(\mathcal{X})$ takes a uniform distribution $p(\mathcal{X} = x) = 0.2\ \forall x \in \mathcal{X}$. Note that the columns of conditional probability tables sum to one since these are probability mass functions.

	x_1	x_2	x_3	x_4	x_5
p	0.7	0.6	0.1	0.1	0.1
q	0.1	0.1	0.1	0.1	0.6
r	0.1	0.2	0.1	0.1	0.2
s	0.1	0.1	0.7	0.7	0.1

Table 6. Conditional Probability Table $\mathbb{P}(\mathcal{O}|\mathcal{X})$

Our goal is to find the cluster mapping given by $\mathbb{P}(\mathcal{S}_\mathcal{O}|\mathcal{X})$, $\mathbb{P}(\mathcal{S}_\mathcal{U}|\mathcal{S}_\mathcal{O})$, $\mathbb{P}(\mathcal{S}_\mathcal{T}|\mathcal{S}_\mathcal{U})$,. We initialize ($\tau = 0$) these conditional probabilities as Kronecker delta distributions and apply H-IB. We start from the first level to find $\mathbb{P}_1(\mathcal{S}_\mathcal{O}|\mathcal{X})$. Using the second and third equations of (4),

$$p_0(s_o) = \sum_{x \in \mathcal{X}} p(x) p_0(s_o|x), \forall s_o \in \mathcal{S}_\mathcal{O} \tag{34}$$

	p	q	r	s
A	0.8	0.2	0.1	0.1
B	0.1	0.7	0.1	0.1
C	0.1	0.1	0.8	0.8

Table 7. Conditional Probability Table $\mathbb{P}(\mathcal{U}|\mathcal{O})$

	A	B	C
Γ	0.9	0.1	0.2
Ω	0.1	0.9	0.8

Table 8. Conditional Probability Table $\mathbb{P}(\mathcal{T}|\mathcal{U})$

$$p_0(x|s_o) = \frac{p(s_o|x)p(x)}{p(s_o)}, \forall (x, s_o) \in \mathcal{X} \times \mathcal{S}_\mathcal{O} \tag{35}$$

$$p_0(o|s_o) = \sum_{x \in \mathcal{X}} p(o|x) p(x|s_o), \forall (o, s_o) \in \mathcal{O} \times \mathcal{S}_\mathcal{O} \tag{36}$$

With some manipulation, we have that

$$p_0(s_u|x) = \sum_{s_o \in \mathcal{S}_\mathcal{O}} p_0(s_u|s_o) p_0(s_o|x), \forall (s_u, x) \in \mathcal{S}_\mathcal{U} \times \mathcal{X} \tag{37}$$

$$p_0(s_t|x) = \sum_{s_u \in \mathcal{S}_\mathcal{U}} p_0(s_t|s_u) p_0(s_u|x), \forall (s_t, x) \in \mathcal{S}_\mathcal{T} \times \mathcal{X} \tag{38}$$

$$p(u|x) = \sum_{o \in \mathcal{O}} p(u|o) p(o|x), \forall (u, x) \in \mathcal{U} \times \mathcal{X} \tag{39}$$

$$p_0(u|s_o) = \sum_{x \in \mathcal{X}} p(u|x) p_0(x|s_o), \forall (u, s_o) \in \mathcal{U} \times \mathcal{S}_\mathcal{O} \tag{40}$$

$$p_0(t|x) = \sum_{u \in \mathcal{U}} p_0(t|u) p_0(u|x), \forall (t, x) \in \mathcal{T} \times \mathcal{X} \tag{41}$$

$$p_0(t|s_o) = \sum_{x \in \mathcal{X}} p_0(t|x) p_0(x|s_o), \forall (t, s_o) \in \mathcal{T} \times \mathcal{S}_\mathcal{O} \tag{42}$$

$$p_0(x|s_u) = \frac{p_0(s_u|x)p(x)}{p_0(s_u)}, \forall (x, s_u) \in \mathcal{X} \times \mathcal{S}_\mathcal{U} \tag{43}$$

$$p_0(x|s_t) = \frac{p_0(s_t|x)p(x)}{p_0(s_t)}, \forall (x, s_t) \in \mathcal{X} \times \mathcal{S}_\mathcal{T} \tag{44}$$

$$p_0(u|s_u) = \sum_{x \in \mathcal{X}} p(u|x) p_0(x|s_u), \forall (u, s_u) \in \mathcal{U} \times \mathcal{S}_\mathcal{U} \tag{45}$$

$$p_0(t|s_t) = \sum_{x \in \mathcal{X}} p(t|x) p_0(x|s_t), \forall (t, s_t) \in \mathcal{T} \times \mathcal{S}_\mathcal{T} \tag{46}$$

Finally, using (47) (also the first equation of (4)):

$$\begin{aligned}
p_1(s_o|x) &= \frac{p(s_o)}{\mathcal{Z}} \exp[-\beta D_{KL}(\mathbb{P}(\mathcal{O}|\mathcal{S}_\mathcal{O} = s_o)||\mathbb{P}(\mathcal{O}|\mathcal{X} = x)) \\
&- \beta \sum_{s_u \in \mathcal{S}_\mathcal{U}} p(s_u|x) D_{KL}(\mathbb{P}(\mathcal{U}|\mathcal{S}_\mathcal{U} = s_u)||\mathbb{P}(\mathcal{U}|\mathcal{X} = x)) \\
&- \beta \sum_{s_t \in \mathcal{S}_\mathcal{T}} p(s_t|x) D_{KL}(\mathbb{P}(\mathcal{T}|\mathcal{S}_\mathcal{T} = s_t)||\mathbb{P}(\mathcal{T}|\mathcal{X} = x))]
\end{aligned} \tag{47}$$

Setting $\beta = 100$, we obtain an updated $\mathbb{P}_{\tau=1}(\mathcal{S}_\mathcal{O}|\mathcal{X})$ given in Table 9.

	x_1	x_2	x_3	x_4	x_5
s_1	1.0	0.0	0.0	0.0	0.0
s_2	0.0	1.0	0.0	0.0	0.0
s_3	0.0	0.0	0.5	0.5	0.0
s_4	0.0	0.0	0.5	0.5	0.0
s_5	0.0	0.0	0.0	0.0	1.0

Table 9. Conditional Probability Table $\mathbb{P}_{\tau=1}(\mathcal{S}_\mathcal{O}|\mathcal{X})$

	s_1	s_2	s_3	s_4	s_5
p	0.7	0.6	0.1	0.1	0.1
q	0.1	0.1	0.1	0.1	0.6
r	0.1	0.2	0.1	0.1	0.2
s	0.1	0.1	0.7	0.7	0.1

Table 10. Conditional Probability Table $\mathbb{P}_{\tau=1}(\mathcal{O}|\mathcal{S}_\mathcal{O})$

This conditional probability informs us of a *soft cluster* map to group the observations to objects. In this case, x_1, x_2, x_5 each corresponds to an object and x_3, x_4 are grouped together as observations of the same object. Furthermore, we can *label* these clusters by taking the argmax of $\mathbb{P}(\mathcal{O}|\mathcal{S}_\mathcal{O})$, which is shown in Tab. 10 and can be obtained by manipulating the probability as follows,

$$p_1(o|s_o) = \sum_{x\in\mathcal{X}} p(o|x)p_1(x|s_o), \forall(o, s_o) \in \mathcal{O} \times \mathcal{S}_\mathcal{O} \quad (48)$$

To summarize, after this first iteration of just the object layer, 4 objects are obtained, an object with label p consisting of the observation x_1, an object with label p consisting of the observation x_2, an object with label s consisting of the observations x_3 and x_4, and an object with label q consisting of the observation x_5.

We repeat this for all three layers and for n iterations until convergence. The final $\mathbb{P}(\mathcal{S}_\mathcal{O}|\mathcal{X})$, $\mathbb{P}(\mathcal{S}_\mathcal{U}|\mathcal{S}_\mathcal{O})$, $\mathbb{P}(\mathcal{S}_\mathcal{T}|\mathcal{S}_\mathcal{U})$ is given in Table 11, Table 12, and Table 13 respectively with associated final $\mathbb{P}(\mathcal{O}|\mathcal{S}_\mathcal{O})$, $\mathbb{P}(\mathcal{U}|\mathcal{S}_\mathcal{U})$, $\mathbb{P}(\mathcal{T}|\mathcal{S}_\mathcal{T})$ is given in Table 14, Table 15, and Table 16. The combined gives us a hierarchy where task Γ consists of subtask A, which consists of 2 objects both with label p that came from two different observations x_1 and x_2, and task Ω consists of 2 subtasks B and C, subtask B consists of an object with label q from observation x_5 and subtask C consists of an object with label s from observattions x_3 and x_4.

	x_1	x_2	x_3	x_4	x_5
s_1	0.99	0.03	0.0	0.0	0.0
s_2	0.01	0.97	0.0	0.0	0.0
s_3	0.0	0.0	0.5	0.5	0.0
s_4	0.0	0.0	0.5	0.5	0.0
s_5	0.0	0.0	0.0	0.0	1.0

Table 11. Conditional Probability Table final $\mathbb{P}(\mathcal{S}_\mathcal{O}|\mathcal{X})$

	s_1	s_2	s_3	s_4	s_5
s_1	0.51	0.51	0.0	0.0	0.0
s_2	0.49	0.49	0.0	0.0	0.0
s_3	0.0	0.0	0.5	0.5	0.0
s_4	0.0	0.0	0.5	0.5	0.0
s_5	0.0	0.0	0.0	0.0	1.0

Table 12. Conditional Probability Table final $\mathbb{P}(\mathcal{S}_\mathcal{U}|\mathcal{S}_\mathcal{O})$

	s_1	s_2	s_3	s_4	s_5
s_1	0.51	0.51	0.0	0.0	0.0
s_2	0.49	0.49	0.0	0.0	0.0
s_3	0.0	0.0	0.33	0.33	0.33
s_4	0.0	0.0	0.33	0.33	0.33
s_5	0.0	0.0	0.33	0.33	0.33

Table 13. Conditional Probability Table final $\mathbb{P}(\mathcal{S}_\mathcal{T}|\mathcal{S}_\mathcal{U})$

	s_1	s_2	s_3	s_4	s_5
p	0.7	0.6	0.1	0.1	0.1
q	0.1	0.1	0.1	0.1	0.6
r	0.1	0.2	0.1	0.1	0.2
s	0.1	0.1	0.7	0.7	0.1

Table 14. Conditional Probability Table $\mathbb{P}_{\tau=1}(\mathcal{O}|\mathcal{S}_\mathcal{O})$

	s_1	s_2	s_3	s_4	s_5
A	0.57	0.57	0.18	0.18	0.23
B	0.16	0.16	0.16	0.16	0.46
C	0.27	0.27	0.66	0.66	0.31

Table 15. Conditional Probability Table $\mathbb{P}_{\tau=1}(\mathcal{U}|\mathcal{S}_\mathcal{U})$

	s_1	s_2	s_3	s_4	s_5
Γ	0.58	0.58	0.31	0.31	0.31
Ω	0.42	0.42	0.69	0.69	0.69

Table 16. Conditional Probability Table $\mathbb{P}_{\tau=1}(\mathcal{T}|\mathcal{S}_\mathcal{T})$