Procedure-Aware Pretraining for Instructional Video Understanding Supplementary Material

Honglu Zhou[1,2], Roberto Martín-Martín[1,3], Mubbasir Kapadia[2], Silvio Savarese[1] and Juan Carlos Niebles[1]
[1]Salesforce Research, [2]Rutgers University, [3]UT Austin
{hz289,mk1353}@cs.rutgers.edu, robertomm@cs.utexas.edu, {ssavarese,jniebles}@salesforce.com

1. Additional results

1.1. More Qualitative Results

We present more qualitative results shown in Fig. 1 $\sim$ 8. In these visualizations, we illustrate the PKG subgraph of multiple video segments. Specifically, we show the step nodes that are matched to the video segment, a step headline member of the step node, the wikiHow task name of the step headline member (green fonts), the 1-hop in- and out- neighbors of the matched step nodes – the exact neighboring information used by our Paprika model (Table 1 in the main paper) to perform Node Relation Learning during pre-training, and the edge connections between these step nodes. For a clear exposition, we did not plot nodes or edges of the PKG that were not leveraged by Paprika pre-training. In order to help the readers to have a concrete idea of the pseudo labels generated by the PKG, we list the complete pseudo labels of these video segments that were used by our Paprika model pre-training to perform Video-Node Matching, Video-Task Matching, Task Context Learning and Node Relation Learning in Table 4 $\sim$ 39. For pseudo labels of Video-Node Matching and Node Relation Learning, node IDs are listed in the descending ranking order of their confidence scores in the tables.

As shown in the figures, the PKG subgraph that a video segment belongs to entails a high relevance to the video segment, and the graph structure encodes the procedural knowledge of the general order and relation of steps from multiple tasks. Qualitative results also suggest that using a larger version of wikiHow dataset (e.g., [9–12]) or stronger video and language encoders to either build the PKG or generate pseudo labels would be beneficial (e.g., to obtain more relevant steps for video segments).

1.2. Results of Paprika with 2 Hops

In Table 1 of the main paper, we report the results of Paprika that uses the best setting of VNM, VTM and TCL from the ablation study, but for NRL, $K = 1$ instead of 2. This is because computing the NRL pseudo labels with $K = 2$ is resource-wise expensive for the *full* HowTo100M dataset due to 51M video segment samples (see Sec. 2.1).

Top 1 (40)	Step Headline: Finished Task: How to Make Pine Needle Tea
	Step Headline: Finished Task: How to Use a Bone Folder
	Step Headline: Finished Task: How to Sew Hair Extensions to a Clip
Top 2 (22)	Step Headline: Disconnect the battery Task: How to Install a Car Starter
	Step Headline: Disconnect the battery cables Task: How to Change a Timing Chain
	Step Headline: Disconnect the battery Task: How to Change Radiator Fluid
Top 3 (15)	Step Headline: Gather the necessary supplies Task: How to Chlorinate a Well
	Step Headline: Gather miscellaneous supplies Task: How to Make a Wargaming Table
	Step Headline: Gather your supplies Task: How to Apply a Horse Tail Bandage

Note: The number inside the bracket is the number of step headline members of the step node. For each step node, we list 3 random members.

Table 1. **The top 3 largest step nodes of the PKG.** Steps from different tasks may be described in the same or slightly different manner, but share the same semantic meaning.

In Table 3 of the Supplementary Material, we report results of Paprika (VNM+VTM+TCL+NRL) that uses $K = 2$ for NRL and the HowTo100M *subset* as the pre-training data. $K = 2$ for NRL has a better overall performance than $K = 1$: on 8 out of the 12 evaluation settings, the performance of $K = 2 \geq$ the performance of $K = 1$ for Paprika (VNM+VTM+TCL+NRL). Considering the trade-off between computation and performance, we recommend experimenting with a larger K when the size of the training video data is small.

1.3. Leveraging the PKG for Downstream Tasks

Prior work DS [4] experimented with incorporating partial procedural knowledge at the downstream stage, i.e., at the inference time of $f(\cdot)$. Here, we conduct similar experiments to directly leverage the PKG for the downstream model. We then compare the effectiveness of our proposed

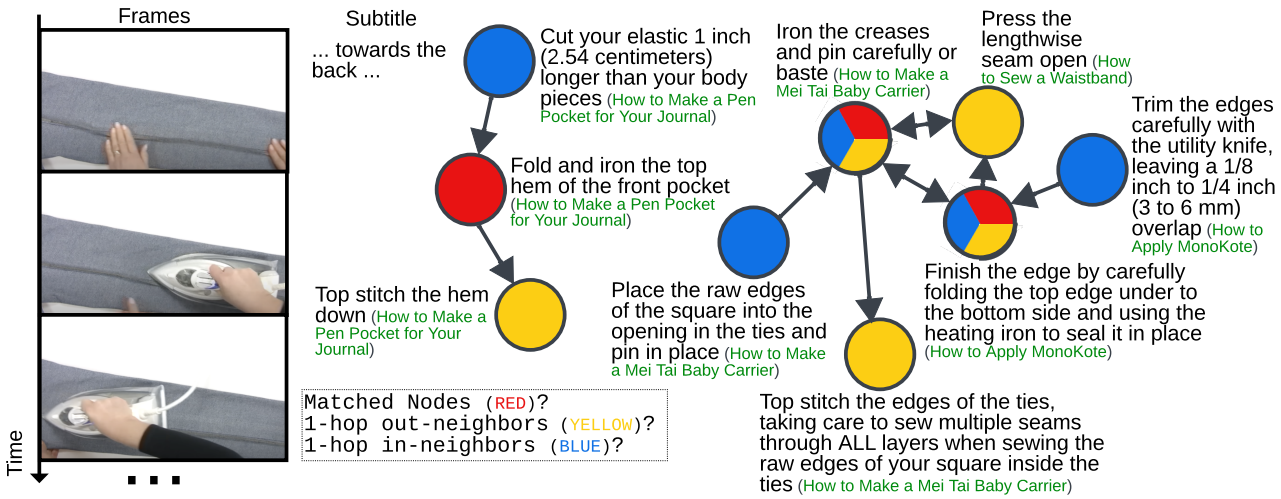

Figure 1. **The PKG subgraph that a video segment belongs to.** The temporally overlapped subtitle of this video segment fails to describe the step of the segment because the action of the step are recorded by the camera. The matched nodes of the PKG capture the action "ironing" of the step. According to the PKG subgraph, the action before "ironing" can be "trimming" or "cutting", and the step after this video segment can be "stitch the hem/edges". The steps of the PKG subgraph that the video segment belongs to can come from multiple tasks (e.g., "How to Sew a Waistband", "How to Apply MonoKote", "How to Make a Mei Tai Baby Carrier", etc).

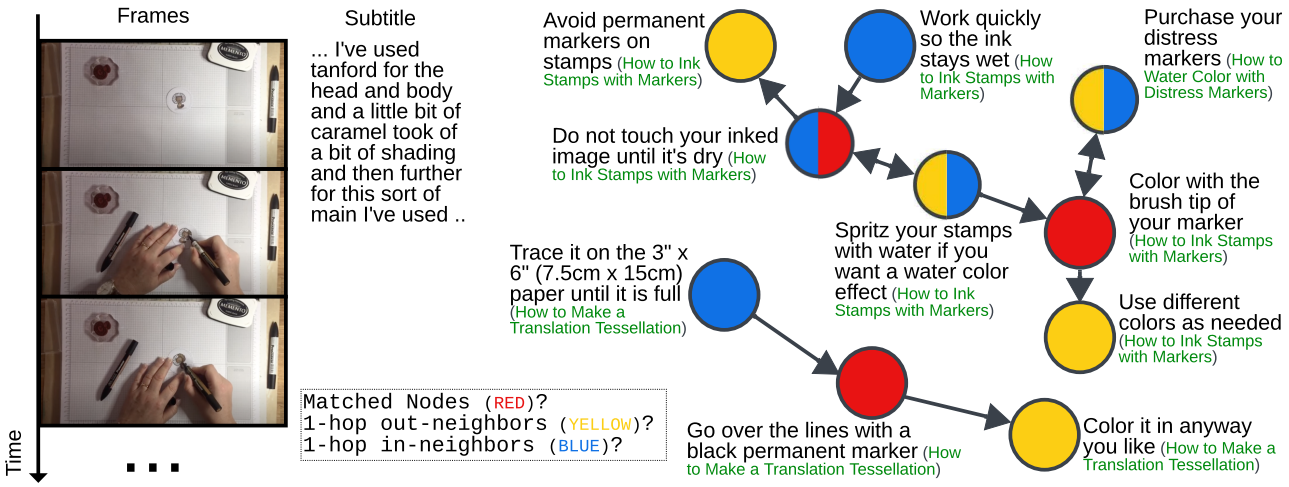

Figure 2. **The PKG subgraph that a video segment belongs to.** The PKG subgraph entails a high relevance to the video segment (which is about coloring with a marker), and the graph structure encodes the procedural knowledge of the general order and relation of steps from multiple tasks.

PKG with $\mathbb{B}$ (i.e., wikiHow articles) for the downstream task of Step Forecasting on the COIN dataset. Below, we describe the details of the comparative experiments. The mathematical notations follow the ones in the main paper.

W/o Incorporating Knowledge Base. Eq. 1 formulates the input sequence to the downstream Transformer model $\mathcal{T}$ without incorporating procedural knowledge from any knowledge database for the downstream Step Forecasting task (suppose historical steps contain L segments):

$$\mathcal{T}\left(f(e(x_1)), f(e(x_2)), \cdots, f(e(x_L))\right) \tag{1}$$

W/ Incorporating $\mathbb{B}$ (i.e., the method in DS [4]). In order to incorporate the knowledge retrieved from $\mathbb{B}$ for each segment into the input provided to the Transformer downstream model, a retrieval approach is firstly adopted to find for each segment x_l, the step headline from $\mathbb{B}$ that best explains the segment according to the pre-trained video model $f(\cdot)$. Specifically, given the trained $f(\cdot)$ along with the trained segment-step matching classifier $a(\cdot)$, the retrieved step headline $\hat{s}_{i_{x_l}}^{(t_{x_l})}$ of segment x_l is the one that corresponds to the step class ID yielding the maximum classification score according to the inference of $a(f(e(x_l)))$. Then, $\mathcal{T}$ accepts a different input sequence as shown in Eq. 2:

$$\mathcal{T}\left(f(e(x_1)), m(\hat{s}_{i_{x_1}+1}^{(t_{x_1})}), f(e(x_2)), m(\hat{s}_{i_{x_2}+1}^{(t_{x_2})}) \cdots, f(e(x_L)), m(\hat{s}_{i_{x_L}+1}^{(t_{x_L})})\right) \tag{2}$$

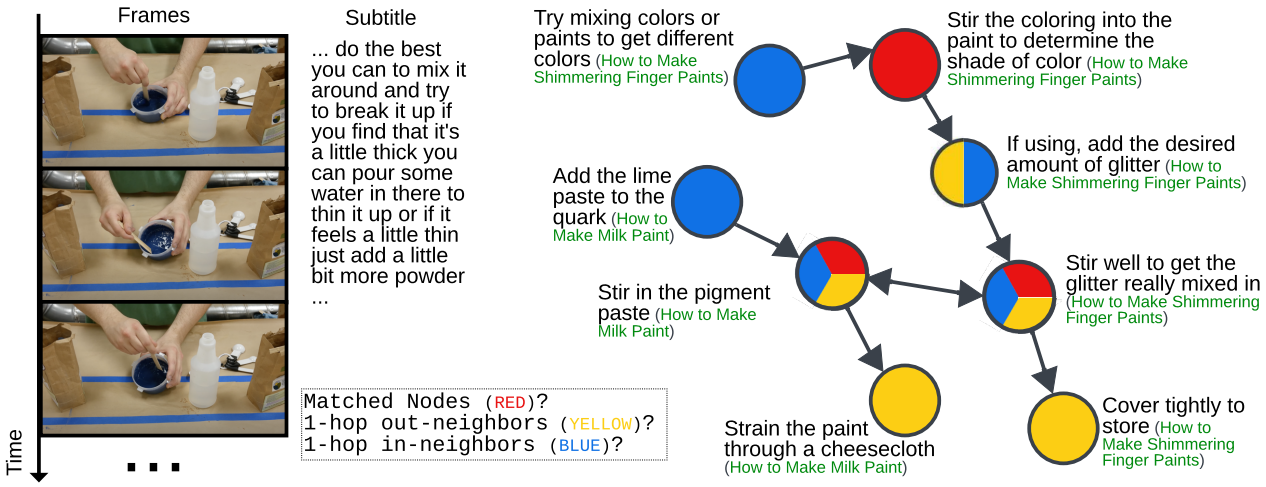

Figure 3. **The PKG subgraph that a video segment belongs to.** The matched step nodes of the PKG are about "stirring"/"mixing" the "pigment"/"paint", which reflects the action in the video segment. The PKG subgraph entails a high relevance to the video segment, and the graph structure encodes the procedural knowledge of the general order and relation of steps from multiple tasks. For example, the next step (1-hop out neighbor) can be "cover tightly to store" , and the steps of the PKG subgraph come from the wikiHow tasks "How to Make Milk Paint" and "How to Make Shimmering Finger Paints".

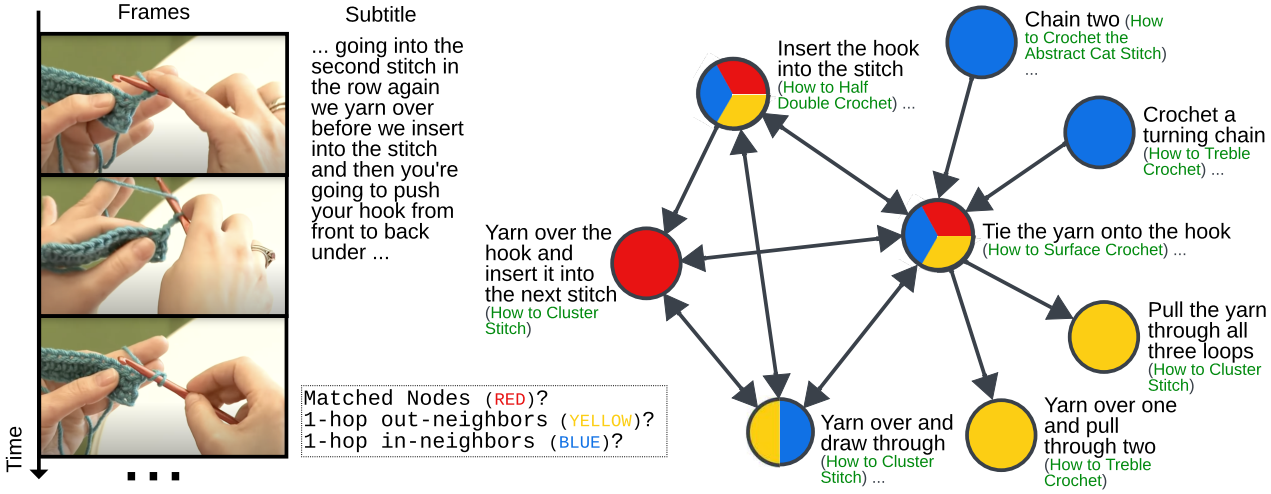

Figure 4. **The PKG subgraph that a video segment belongs to.** In this example, the left most four step nodes are quite densely connected, which might be attributed to the visually similar frames as well as complex semantic implication and subtle difference between step headlines of these step nodes. Overall the PKG subgraph entails a high relevance to the video segment.

$m(\cdot)$ represents the feature extractor of step headlines.

In other words, for each segment x_l, we obtain its representation using $f(e(x_l))$. Here, the model $f(\cdot)$ was trained using DS [4] and $e(\cdot)$ is the MIL-NCE model [5]. The answer head $a(\cdot)$ predicts the most likely current step $\hat{s}_{i_{x_l}}^{(t_{x_l})}$. We then look up wikiHow article t_{x_l} to find the next step $\hat{s}_{i_{x_l}+1}^{(t_{x_l})}$. We obtain the next step's headline feature produced by the step headline feature extractor $m(\cdot)$, and the next step's headline feature follows the representation of the segment x_l in the input sequence to $\mathcal{T}$.

We call this downstream Transformer variant as "Transformer w/ KB Transfer from $\mathbb{B}$". Segment features produced from $f(\cdot)$ and the step headline features of retrieved next steps according to $a(f(\cdot))$ and $\mathbb{B}$, together they form the input sequence to the Transformer.

Pre-training	Downstream Method	Accuracy
DS [4]	Transformer w/ KB Transfer from $\mathbb{B}$ [4]	22.15
Paprika	Transformer w/ KB Transfer from the PKG	**35.46**

Table 2. Results of Step Forecasting on COIN by directly incorporating the procedural knowledge database at the downstream stage.

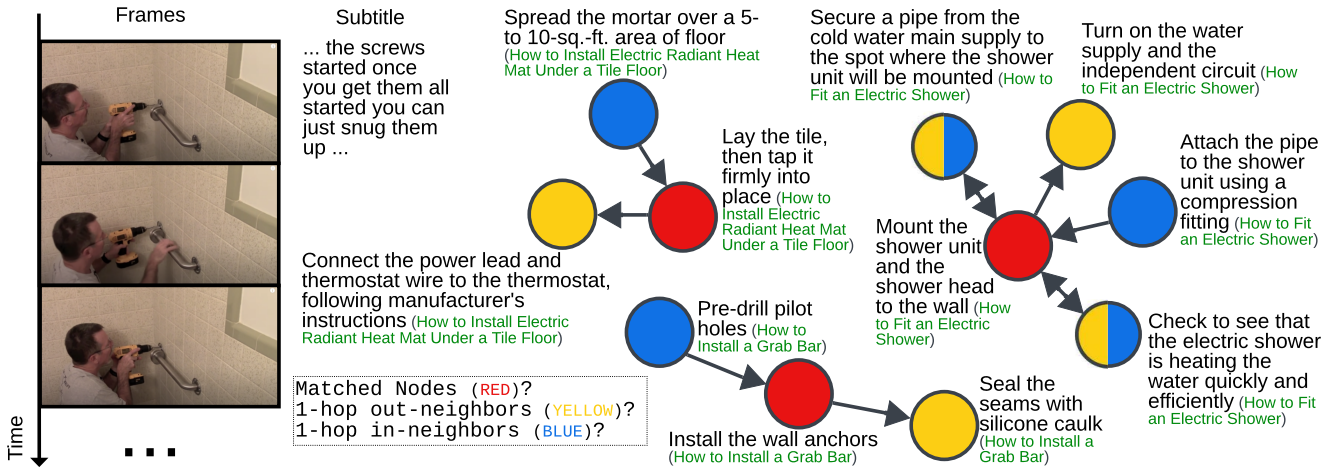

Figure 5. **The PKG subgraph that a video segment belongs to.** "Install the wall anchors" is the top 2 matched node of this video segment, which well describes the step of the segment. The top 1 matched node "mount the shower unit and the shower head to the wall" is also related to the shower room, but the objects "shower unit" and "shower head" are actually missing in the video segment. The top 3 matched node "lay the tile, then tap it firmly into place" is mainly focusing on the object "tile" shown in the video segment. Using a larger version of the wikiHow dataset and a stronger pre-trained multimodal video foundation model would lead to a even better quality of pseudo labels.

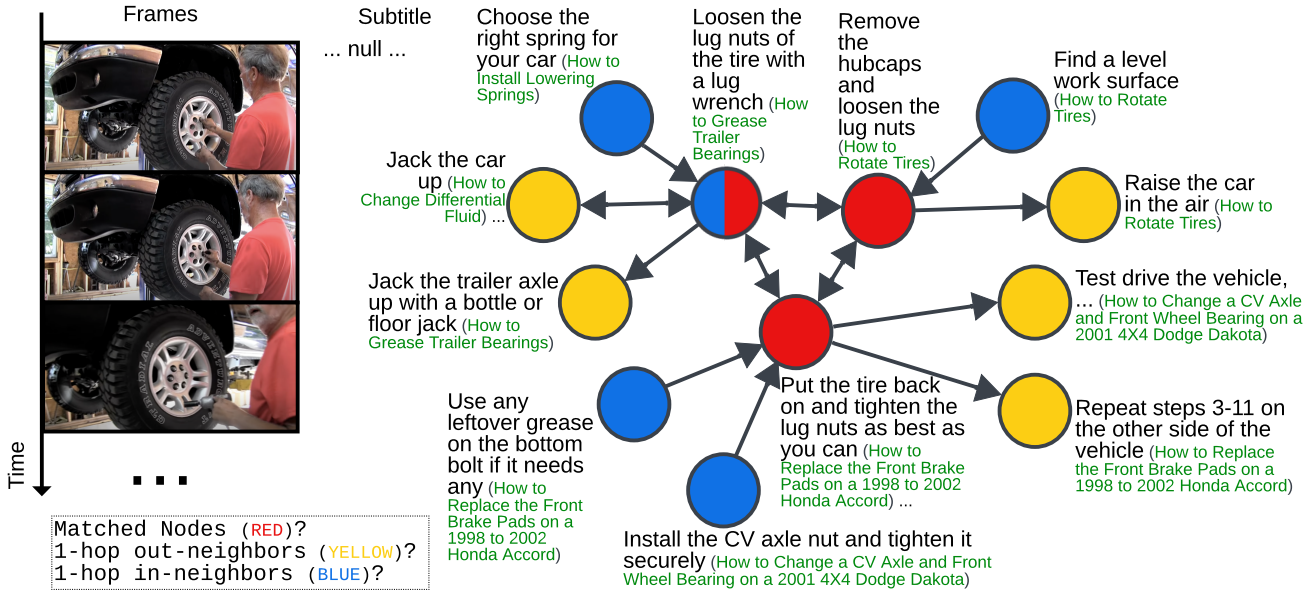

Figure 6. **The PKG subgraph that a video segment belongs to.** This PKG subgraph fails to conform to the temporal signals of the video segment: a man is "tightening the lug nuts" but two of the matched nodes are about "loosening the lug nuts". In addition, the car has been jacked up in the video segment, but according to the PKG subgraph, "jack the car up" is the next step (one of the top 5 1-hop out-neighbors). The PKG graph encodes the *general* order and relation of steps – the knowledge is not conditioned on a specific video segment.

W/ Incorporating the PKG [4]. Similar to the baseline "Transformer w/ KB Transfer from $\mathbb{B}$" proposed by [4], we propose "Transformer w/ KB Transfer from the PKG".

For each segment x_l, we obtain its representation using $f(e(x_l))$; here, $f(\cdot)$ is Paprika pre-trained using our proposed objectives that leverage the PKG, and $e(\cdot)$ is the MIL-NCE model [5]. The Video-Node Matching answer head $a(\cdot)$ predicts the step node $\hat{v}_{x_l}$ that is most likely to be matched to the video segment x_l. We then look up the PKG to obtain the 1-hop out-neighboring nodes $\mathcal{N}(\hat{v}_{x_l})$ of the node $\hat{v}_{x_l}$. Given the step headlines of $\mathcal{N}(\hat{v}_{x_l})$, for each step headline, we obtain the step's headline feature produced by the step headline feature extractor $m(\cdot)$. The mean of these feature vectors is considered as the feature of the most likely next node, which follows the representation of the segment

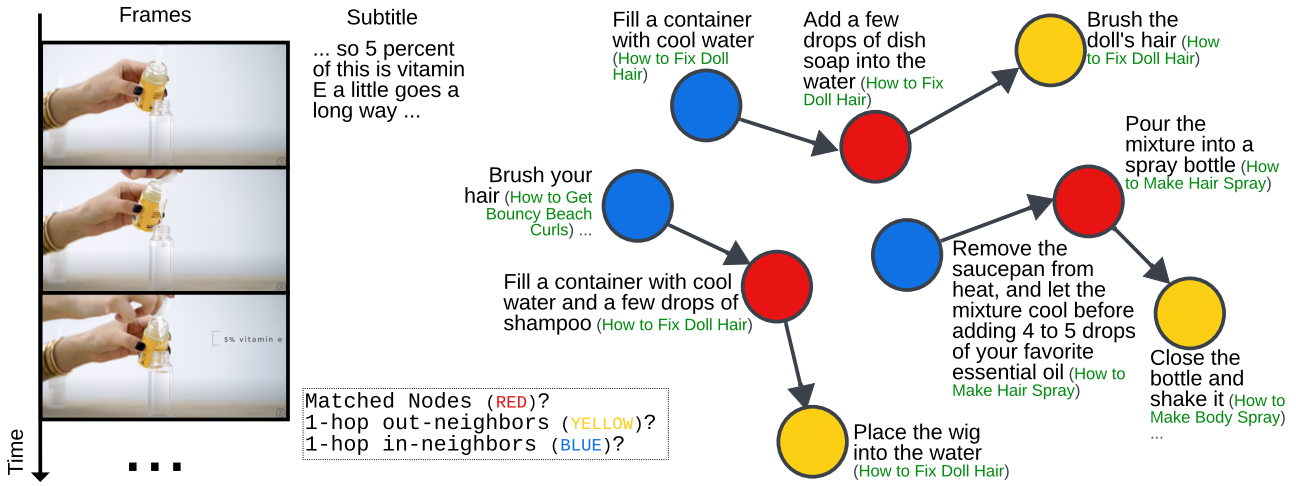

Figure 7. **The PKG subgraph that a video segment belongs to.** The step of the video segment is "adding a few drops of Vitamin E oil". "Vitamin E" is in the subtitle, but the PKG subgraph fails to capture the "Vitamin E" information, because the visual frame signals were used to match the video segment to nodes. This example suggests subtitles can be useful in cases such as when the objects are small or hard to be recognized from frames.

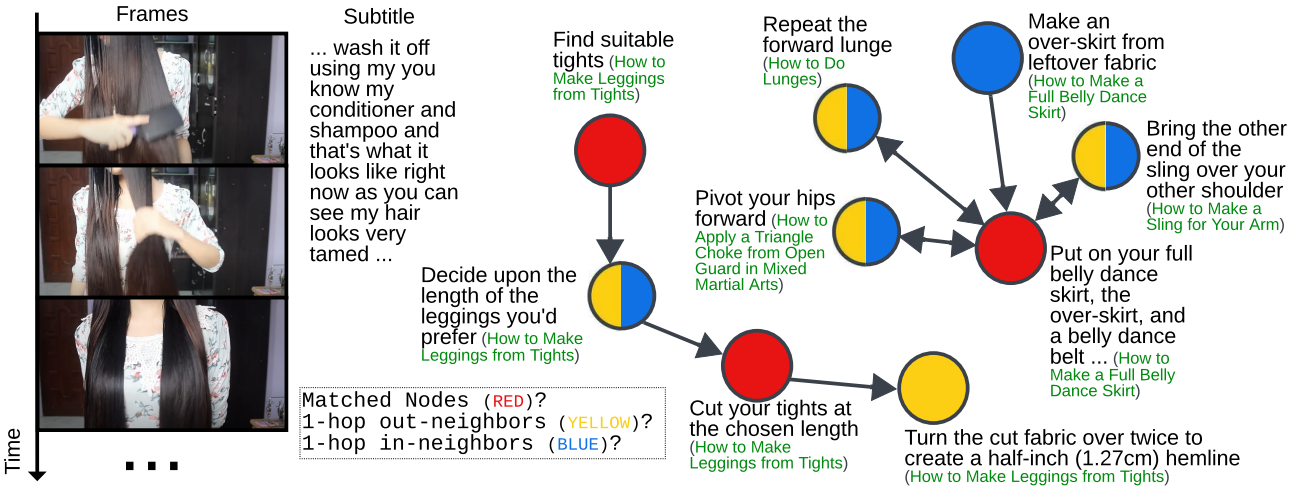

Figure 8. **The PKG subgraph that a video segment belongs to.** The PKG subgraph is not relevant to the video segment in this example because the narrator is mentioning the step "wash hair", but the frames are not showing the step "wash hair". Subtitles can be useful when the action of the step is not demonstrated but only verbally described by the narrator.

x_l to form the input sequence to $\mathcal{T}$:

$$\mathcal{T}\left(\cdots, f(e(x_l)), \frac{1}{|\mathcal{N}(\hat{v}_{x_l})|}\sum_{i\in\mathcal{N}(\hat{v}_{x_l})}\left(\frac{1}{|\mathcal{S}(i)|}\sum_{j\in\mathcal{S}(i)} m(j)\right), \cdots\right) \tag{3}$$

where $\mathcal{S}(i)$ denotes the set of step headline members of node i. Therefore, Paprika produced segment features and features of the PKG retrieved next nodes, together they form the input sequence to the Transformer.

Results are presented in Tab. 2. We observe a performance drop after integrating procedural knowledge into the downstream models. This suggests a future research direction towards effectively integrating procedural knowledge into procedure-understanding-related downstream model training and/or inference. However, Paprika still outperforms the prior work DS [4] under this setting, which further reinforces our key insight that using procedural information during pre-training is beneficial. We highlight a crucial difference between our work and DS [4]: DS uses partial procedural knowledge in the downstream task, while we are the first to show that using graph-structured procedural knowledge in *pre-training* is potentially beneficial for *any* procedure understanding downstream task.

Pre-training Method	Downstream Transformer						Downstream MLP					
	COIN			CrossTask			COIN			CrossTask		
	SF	SR	TR	SF	SR	TR	SF	SR	TR	SF	SR	TR
Paprika (VNM + VTM + TCL + NRL)	**42.91**	**50.62**	85.02	61.29	**62.48**	**67.41**	39.75	44.63	**83.88**	**59.53**	**60.76**	**67.51**

Note: **SF**: Step Forecasting; **SR**: Step Recognition; **TR**: Task Recognition.

Table 3. Downstream procedure understanding evaluation results of Paprika (VNM + VTM + TCL + NRL) **when $\mathbf{K = 2}$ for NRL**. Bolded ones are the cases where the accuracy of $K = 2 \geq$ the accuracy of $K = 1$ for Paprika (VNM + VTM + TCL + NRL).

Node ID	Step Headline
1982	Finish the edge by carefully folding the top edge under to the bottom side and using the heating iron to seal it in place.
443	Fold and iron the top hem of the front pocket.
1655	Iron the creases and pin carefully or baste.

Table 4. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 1.

2. Implementation Details

2.1. Node & Edge Construction of the PKG

The wikiHow database that we used has a total of $10,588$ step headlines from $T = 1,053$ task articles [3]. This is the same version of the wikiHow database $\mathbb{B}$ that DS [4] used. These wikiHow tasks have at least 100 video samples in the HowTo100M dataset [4]. The text branch of the author-released S3D model pre-trained using the MIL-NCE objective [5] (in the paper, we call it the MIL-NCE model for short) outputs the feature of each step headline. We used Agglomerative Clustering from the scikit-learn library for step deduplication (**`Step 1`** from Sec. 3.2 in the main paper). We used a relatively conservative clustering criterion since the goal here is deduplication and avoiding putting two semantically-different steps into one cluster is desired. Specifically, we used the minimum of the distances between all observations of the two sets as the linkage criterion, with cosine similarity as the distance function and a threshold of 0.09. The resulting number of step nodes is $10,038$. Among them, 314 step nodes have more than one step headline as its members. We list the top 3 largest step nodes and their randomly sampled members in Table 1 of the Supplementary Material. We find cross-task characteristics of steps, i.e., one step (which can be described slightly differently) can belong to multiple tasks.

We set the video segments to be 9.6 seconds long in consideration of the temporal lengths of steps and videos in HowTo100M. There is no temporal overlapping or spacing between segments of one video. This leads to 3.7M video segment samples for the HowTo100M subset [1] that we have mainly used for model training and the PKG building, and 51M video segment samples for the full HowTo100M dataset. The MIL-NCE model was pre-trained on the full HowTo100M dataset on 3.2 seconds long video segments [5]. Therefore, the feature of each 9.6 seconds long segment was mean pooled from features of the three 3.2 seconds long segments. Dot product between MIL-NCE produced feature of a step headline and MIL-NCE produced feature of frames of a video segment, yielded a similarity score, which was considered to be the matching confidence score between the step headline and the video segment.

In order to obtain direct step transitions in HowTo100M (**`Step 2`** from Sec. 3.2 in the main paper), we looped through the videos in the HowTo100M subset, and for each video, we started from the first segment to the second last to collect the candidates of direct step transitions. Specifically, for every two temporally adjacent segments in the video, e.g., for segment i and segment j, pair-wise combinations of the matched step headlines of segment i and the matched step headlines of segment j form the candidates, but only if the preceding step headline and the succeeding step headline are not the same. For each segment, we considered the step headlines with a similarity score higher than 10 as the matched step headlines of the segment. A step headline transition, e.g., (step headline $m \rightarrow$ step headline n), can appear in multiple videos; we call each occurrence of such transition as one step transition instance. The score of one step transition instance is the product of the matching score of the preceding step headline and the matching score of the succeeding step headline. Final score of the step transition (step headline $m \rightarrow$ step headline n) is the summed score aggregated from all instances of the step transition (step headline $m \rightarrow$ step headline n). In this way, step transitions that happen more frequently in the video corpus can have higher step transition scores (i.e., more confident).

Given the collection of step transition candidates from HowTo100M, we removed these less confident candidates if the step transition score is lower than 1000 and then performed log min-max normalization to constrain the scores into the range of $[0, 1]$.

In the above, we have described how we obtained the step transitions from HowTo100M. We describe how we obtained the step transitions from wikiHow in the following. In an wikiHow article t, a pair $(s_i^{(t)}, s_{i+1}^{(t)})$ is defined to be a direct step transition. We assigned a score of 1 for all di-

Task Name
How to Apply MonoKote
How to Make a Pen Pocket for Your Journal
How to Make a Mei Tai Baby Carrier

Table 5. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 1.

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	3139	Trim the edges carefully with the utility knife, leaving a 1/8 inch to 1/4 inch (3 to 6 mm) overlap.
	591	Cut your elastic 1 inch (2.54 centimeters) longer than your body pieces.
	1255	Place the raw edges of the square into the opening in the ties and pin in place.
	1655	Iron the creases and pin carefully or baste.
	1982	Finish the edge by carefully folding the top edge under to the bottom side and using the heating iron to seal it in place.
1 Hop Out-Neighbor	1087	Top stitch the hem down.
	877	Top stitch the edges of the ties, taking care to sew multiple seams through ALL layers when sewing the raw edges of your square inside the ties.
	1982	Finish the edge by carefully folding the top edge under to the bottom side and using the heating iron to seal it in place.
	1655	Iron the creases and pin carefully or baste.
	1006	Press the lengthwise seam open.
2 Hops In-Neighbor	2469	Unroll a second sheet of MonoKote, and repeat the process on the other side of the aircraft part.
	782	Cut a third piece of for the front pocket.
	1744	Repeat measuring, marking and sewing steps for the second tie.
2 Hops Out-Neighbor	233	Pin the pocket to the front of one of the larger body pieces.
	6588	The pouch is now ready to wear.
	1982	Finish the edge by carefully folding the top edge under to the bottom side and using the heating iron to seal it in place.

Table 6. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 1.

rect step transitions in wikiHow, since these step transitions were all annotated by humans.

Using the mapping from the step headline to step node, step transitions from both wikiHow and HowTo100M were added as edges to connect the step nodes, and thus formed the structure of the PKG. For a directed node pair (n_1, n_2), if there are multiple step transitions and hence multiple scores, we kept the maximal score as the confidence score of the edge (n_1, n_2) (i.e, node $n_1 \rightarrow$ node n_2).

All thresholding criteria involved in the above graph construction process were empirically chosen through qualitative manual examination.

2.2. Pseudo Label Generation

For Video-Node Matching (VNM), the pseudo labels of one video segment are the node IDs of the top 3 step nodes with the highest matching confidence scores. Since a step node may have multiple step headlines as its members, the matching score between a step node and a segment, is set to be the maximal value of the matching scores that are generated by all pairs of the node's step headline and the segment. Because we match each video segment to the "top 3" nodes in the graph, there is the risk of forcing a match when no relevant node exists in the graph. However, when the graph is sufficiently large, most segments will find a node of relative relevance. One could further improve over this by thresholding on the matching score and reassigning to a background node when no nodes surpass the threshold.

For Video-Task Matching (VTM), the pseudo labels of one video segment are the wikiHow task names of the step headlines, which are the members of the matched step nodes obtained using VNM. If the HowTo100M task names were used, we first need to obtain the occurrence matrix $O \in \mathbb{R}^{S \times T'}$ where S denotes the number of step headlines in wikiHow ($S = 10,588$) and T' is the number of HowTo100M tasks that are covered by the videos ($T' = 1,059$ for the HowTo100M subset, and $T' = 25$K for the fullset). We populated O by looping through segments of videos. For each segment, given the video's task name annotation, we incremented this task's occurrence of the matched step headlines by 1 (the step headlines are members of the matched step nodes from VNM). Given the step headlines of the matched step nodes of the video segment, the video segment's VTM pseudo labels using the HowTo100M task annotations would then be the top 3 HowTo100M task names with the highest occurrences of the step headlines.

Task Context Learning (TCL) is built upon VTM. Given the matched tasks from VTM, using the transposed version of the step-task occurrence matrix, we obtained the step headlines that a task needs. We then mapped these step headlines to step nodes, and the node IDs are the final TCL

Node ID	Step Headline
2811	Color with the brush tip of your marker.
8988	Do not touch your inked image until it's dry.
9399	Go over the lines with a black permanent marker.

Table 7. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 2.

Task Name
How to Ink Stamps with Markers
How to Make a Translation Tessellation

Table 8. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 2.

pseudo labels of the video segment. If the HowTo100M task names were used, the information on the step headlines that a task needs can be noisy because many step headlines may have a non-zero occurrence value (because video-step matching is not perfect). Therefore, we only considered the top 3 step nodes (mapped from step headlines) with the highest non-zero task occurrence values as the pseudo labels of TCL, if the task names were from HowTo100M.

W.r.t how to obtain Node Relationship Learning (NRL) pseudo labels of one video segment, for each matched step node from VNM, we queried the PKG to obtain its k-hop in-neighbors and out-neighbors. One node may have multiple in-neighbors and/or out-neighbors. Suppose node j is one of the k-hop in-neighbors of node i, it means that there is a directed path from node j to node i of length exactly k (k edges along the directed path). Similarly, suppose node j is one of the k-hop out-neighbors of node i, it means that there is a directed path from node i to node j of length exactly k.

The confidence score of the 1-hop in-neighbors and out-neighbors are the edge confidence scores of the edges that connect the matched step node and its 1-hop neighbors. If $k > 1$, the confidence score of a k-hop *in*-neighbor i is the edge confidence score of the directed edge (i, j), i.e., node $i \rightarrow$ node j, multiplied by the confidence score of node j being the video segment's $(k-1)$-hop *in*-neighbor. A similar strategy applies to k-hop *out*-neighbors: if $k > 1$, the confidence score of a k-hop out-neighbor i is the edge confidence score of the directed edge (j, i), i.e., node $j \rightarrow$ node i, multiplied by the confidence score of node j being the video segment's $(k-1)$-hop *out*-neighbor. When the neighbors are available, we considered the top 5 most confident neighbors for the first hop ($k = 1$) in- or out-neighbors, and top 3 for the second hop ($k = 2$) in- or out-neighbors; these node IDs are the pseudo labels.

2.3. Pre-training Paprika

The input to $f(\cdot)$ is the video segment's visual feature produced by MIL-NCE [5] with a dimensionality of 512. In practice, the architecture we chose for $f(\cdot)$ is a shallow MLP with only one hidden layer, which is a bottleneck layer with a dimensionality of 128. The output layer of the MLP $f(\cdot)$ has a dimensionality of 512, which means we set the dimensionality of the refined video segment feature to be the same as the original video segment feature, in order to verify the refinement ability of $f(\cdot)$ for procedure understanding. The ReLU non-linear activation was used in between the linear layers of the MLP $f(\cdot)$.

Output representation of $f(\cdot)$ is the input to multiple answer heads $a(\cdot)$ to perform the pre-training objectives. For example, Paprika with NRL (*2 hops*) (Table 1 in the main paper) indicates that during pre-training, there were 4 answer heads because of 2 hops and 2 directions (in and out). Paprika with VNM + VTM (*wikiHow + HT100M*) + TCL (*wikiHow*) + NRL (*1 hop*) indicates 1 (VNM)+2(VTM)+1(TCL)+2(NRL) = 6 answer heads. Parameters of these answer heads were not shared, and their exact architectures used are described in the next paragraph.

All pre-training objectives were modelled as a *multi-label classification* problem. For example, for NRL, to predict the 1-hop out-neighbors, the node IDs are the class indices, and for VTM, the task IDs are the class indices. The answer heads are MLPs with ReLU being the non-linear activation. For the pre-training objectives that use the node IDs as the class indices (VNM, TCL and NRL), the answer head MLP has two hidden layers with a dimensionality of 2509 (#classes//4) and 5019 (#classes//2) respectively, and the output layer has a dimensionality of 10038 (#classes) ('#' means 'the number of' and '//' denotes 'divide and round down to the nearest integer'). For the pre-training objectives that use the task IDs as the class indices (VTM), the answer head MLP has one hidden layer with a dimensionality of #classes//2. We used *Binary Cross Entropy* as the loss function. We did not tune the loss co-efficient of each pre-training objective to maintain simplicity.

We set $f(\cdot)$ and answer heads $a(\cdot)$ to be simple MLPs and did not tune the model architectural settings in order to evaluate the effectiveness and ease of use of the pseudo labels generated by the PKG and the proposed pre-training

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	8361	Spritz your stamps with water if you want a water color effect.
	4496	Work quickly so the ink stays wet.
	5543	Trace it on the 3" x 6" (7.5cm x 15cm) paper until it is full.
	3201	Purchase your distress markers.
	8988	Do not touch your inked image until it's dry.
1 Hop Out-Neighbor	8130	Use different colors as needed.
	8589	Avoid permanent markers on stamps.
	8492	Color it in anyway you like.
	3201	Purchase your distress markers.
	8361	Spritz your stamps with water if you want a water color effect.
2 Hops In-Neighbor	8141	Sand your stamp lightly.
	2985	Experiment a little.
	8591	Convert this base tessellation into a more interesting shape.
2 Hops Out-Neighbor	3336	Breathe on the stamp to moisten the ink.
	6068	Ink it all over.
	2811	Color with the brush tip of your marker.

Table 9. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 2.

Node ID	Step Headline
2374	Stir in the pigment paste.
6937	Stir well to get the glitter really mixed in.
9026	Stir the coloring into the paint to determine the shade of color.

Table 10. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 3.

objectives. Using our proposed method, a simple architecture can learn a refined video representation that leads to a stronger performance on the downstream procedure understanding tasks.

We emphasize that the combination of the proposed pre-training objectives leads to better model generalization. Though NRL is the most effective pre-training objective, when K=1, the combination outperformed NRL in 11 out of 12 settings; when K=2, the combination outperformed NRL in 9 out of 12 settings.

We implemented Paprika using PyTorch [6]. We used the Adam optimizer [2], a learning rate of 0.0001, a weight decay of 0, and a batch size of 256. We saved the model checkpoint at every 10 epochs, and chose the downstream performance on the Step Forecasting task on the COIN dataset as the indicator for early stopping with a patience of 500 epochs. The model trained at the epoch that gave the best result on the Step Forecasting task on the COIN dataset, was used to perform all of the downstream evaluation experiments (i.e., the same model checkpoint of $f(\cdot)$ was used for all 12 evaluation settings). When the HowTo100M full data was used for training, due to the much larger amount of training data, we saved the model checkpoint by training steps, i.e., we saved the model checkpoint for every 500 batches and by the end of each training epoch, and trained $f(\cdot)$ up to 100 epochs. The pre-training of Paprika that uses all four pre-training objectives (Table 1 in the main paper) took roughly 100 hours on 8 NVIDIA A100 GPUs. When the full HowTo100M data was used, it took 50 hours on 16 NVIDIA A100 GPUs. As a reference, according to DS [4], their pre-training took 55 hours using 128 GPUs, and MIL-NCE [5] reported that their pre-training required 3 days with 64 8-core TPUs. Our framework is more efficient to train.

2.4. Downstream Evaluation

The downstream task model $t(\cdot)$ is either a MLP or a Transformer encoder layer [8], and the architectural setting is the same across downstream tasks and datasets. The output of the trained frozen $f(\cdot)$ is the input to the downstream task model $t(\cdot)$. The PKG and the answer heads used for the pre-training objectives were discarded during the downstream evaluation (i.e., testing time of $f(\cdot)$).

Since the input is a sequence of video segment features for all of the downstream tasks, we used the Learned Absolute Positional Encoding to learn a vectorized representation for each segment position. The segment's video representation and the position representation are summed to form the position-augmented segment feature.

When the downstream task model $t(\cdot)$ is a MLP, the position-augmented segment features of the input sequence were summed to form a feature vector for the input sequence. This MLP is the downstream task classifier.

When the downstream task model $t(\cdot)$ is a Transformer, the input sequence to the Transformer encoder layer is the position-augmented segment features plus a learned CLS (which stands for 'classification') token. We used only one layer of the Transformer encoder layer. Transformer en-

Task Name
How to Make Milk Paint
How to Make Shimmering Finger Paints

Table 11. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 3.

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	2459	Add the lime paste to the quark.
	7885	If using, add the desired amount of glitter.
	9377	Try mixing colors or paints to get different colors.
	6937	Stir well to get the glitter really mixed in.
	2374	Stir in the pigment paste.
1 Hop Out-Neighbor	8226	Strain the paint through a cheesecloth.
	5149	Cover tightly to store.
	7885	If using, add the desired amount of glitter.
	2374	Stir in the pigment paste.
	6937	Stir well to get the glitter really mixed in.
2 Hops In-Neighbor	5349	Transfer the quark into a paint bucket.
	9026	Stir the coloring into the paint to determine the shade of color.
	8702	Add a few drops of food coloring to each container.
2 Hops Out-Neighbor	4614	Prepare the surface you are painting.
	6937	Stir well to get the glitter really mixed in.
	2374	Stir in the pigment paste.

Table 12. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 3.

coder performs relation reasoning over tokens in the sequence and updates the token representations. The updated feature of the CLS token is the input to the MLP downstream task classifier.

The number of attention heads for the Transformer encoder layer is 8, the dimensionality of the feed-forward network in the Transformer encoder is 1024, and the non-linear activation function is ReLU. The MLP downstream task classifier has a single hidden layer whose dimensionality is 128 for Task Recognition, and 768 for Step Recognition and Step Forecasting. We trained the downstream task model for up to 1000 epochs using early stopping with a patience of 50 epochs. The optimizer was Adam, batch size was 16, learning rate was 0.0001, and weight decay was 0.001.

2.5. Implementation of Baselines

The training and evaluation of baselines (i.e., MIL-NCE*, DS, DS* and VSM in Table 1 of the main paper) follow the same implementation protocols (e.g., downstream model architecture settings, hyper-parameters and training schedules, etc.) as described above for a fair comparison with our Paprika variants.

Specifically, the results of MIL-NCE* described in the main paper were obtained by training and evaluating the downstream task models that took the visual features of video segment from the author-released S3D model, which was pre-trained using the MIL-NCE objective [5], as the input segment features. The results of MIL-NCE* in Table 1 of the main paper can be interpreted as the results of removing $f(\cdot)$ of our framework. Moreover, we used [5]'s video features for all experiments (both pre-training and downstream experiments).

Model architectures used in our experiments are not what were in [4] ([4] used heavier architectures, e.g., TimeSformer [1] for pre-training and heavier Transformer downstream models). In order to obtain the results of DS and DS*, we trained the same MLP-based $f(\cdot)$ model architecture as what Paprika used, but instead we trained it using the pre-training objective proposed by DS [4] that matches video subtitles to step headlines using the pre-trained language model MPNet [7]. In particular, the Step Classification pre-training objective [4] was implemented to recognize the top 3 step headlines from wikiHow that have the highest matching confidence scores (since our VNM objective matches the top 3 nodes). Therefore, both Paprika and DS used the Binary Cross Entropy loss by modeling the pre-training objectives as multi-label classification problems. We also experimented with other loss functions demonstrated in DS [4], i.e., Distribution Matching (KL divergence loss) and Embedding Regression (NCE loss), but we found Step Classification with the Binary Cross Entropy loss is the most robust one among these different loss forms because it obtained the best results in most cases.

Downstream datasets are relatively small; and thus, the risk of overfitting to small datasets exists. Therefore, we did not select a pre-trained model (checkpoint) for each specific downstream task/dataset while we have 12 downstream cases. In addition, the design choices such as model archi-

Node ID	Step Headline
10016	Yarn over the hook and insert it into the next stitch.
67	Tie the yarn onto the hook. Tie the yarn onto your hook. Yarn over the hook. Yarn over and insert the hook. Yarn over twice and insert hook.
37	Insert the hook into the stitch. Insert the hook into the first stitch. Insert the hook into the first stitch of the group. Insert the hook into the next stitch.

Table 13. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 4.

Task Name
How to Cluster Stitch
How to Crochet Popcorn Stitch
How to Surface Crochet
How to Half Double Crochet
How to Treble Crochet

Table 14. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 4.

tectures and hyper-parameters used in both pre-training and downstream tasks were set in an early stage when optimizing the results of the DS baseline, and they were kept constant for all subsequent model variants including Paprika.

References

[1] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *ICML*, volume 2, page 4, 2021. 6, 10

[2] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 9

[3] Mahnaz Koupaee and William Yang Wang. Wikihow: A large scale text summarization dataset. *arXiv preprint arXiv:1810.09305*, 2018. 6

[4] Xudong Lin, Fabio Petroni, Gedas Bertasius, Marcus Rohrbach, Shih-Fu Chang, and Lorenzo Torresani. Learning to recognize procedural activities with distant supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13853–13863, 2022. 1, 2, 3, 4, 5, 6, 9, 10

[5] Antoine Miech, Jean-Baptiste Alayrac, Lucas Smaira, Ivan Laptev, Josef Sivic, and Andrew Zisserman. End-to-end learning of visual representations from uncurated instructional videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9879–9889, 2020. 3, 4, 6, 8, 9, 10

[6] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 9

[7] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33:16857–16867, 2020. 10

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 9

[9] Yue Yang, Artemis Panagopoulou, Qing Lyu, Li Zhang, Mark Yatskar, and Chris Callison-Burch. Visual goal-step inference using wikihow. *arXiv preprint arXiv:2104.05845*, 2021. 1

[10] Li Zhang, Qing Lyu, and Chris Callison-Burch. Intent detection with wikihow. *arXiv preprint arXiv:2009.05781*, 2020. 1

[11] Li Zhang, Qing Lyu, and Chris Callison-Burch. Reasoning about goals, steps, and temporal ordering with wikihow. *arXiv preprint arXiv:2009.07690*, 2020. 1

[12] Shuyan Zhou, Li Zhang, Yue Yang, Qing Lyu, Pengcheng Yin, Chris Callison-Burch, and Graham Neubig. Show me more details: Discovering hierarchies of procedures from semi-structured web data. *arXiv preprint arXiv:2203.07264*, 2022. 1

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	493	Yarn over and draw through. Yarn over and draw through twice.
	37	Insert the hook into the stitch. Insert the hook into the first stitch. Insert the hook into the first stitch of the group. Insert the hook into the next stitch.
	67	Tie the yarn onto the hook. Tie the yarn onto your hook. Yarn over the hook. Yarn over and insert the hook. Yarn over twice and insert hook.
	418	Chain two.
	180	Crochet a turning chain.
1 Hop Out-Neighbor	493	Yarn over and draw through. Yarn over and draw through twice.
	37	Insert the hook into the first stitch. Insert the hook into the next stitch. Insert the hook into the first stitch of the group. Insert the hook into the stitch.
	67	Tie the yarn onto the hook. Tie the yarn onto your hook. Yarn over the hook. Yarn over and insert the hook. Yarn over twice and insert hook.
	4836	Pull the yarn through all three loops.
	4922	Yarn over one and pull through two.
2 Hops In-Neighbors	493	Yarn over and draw through. Yarn over and draw through twice.
	403	Create a foundation chain.
	4836	Pull the yarn through all three loops.
2 Hops Out-Neighbor	493	Yarn over and draw through. Yarn over and draw through twice.
	8289	Insert the hook into the next spot.
	17	Yarn over and pull through twice. Yarn over and pull through. Wrap the yarn around and pull through twice. Loop the yarn over and pull through.

Table 15. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 4.

Node ID	Step Headline
3765	Mount the shower unit and the shower head to the wall.
3616	Install the wall anchors.
2823	Lay the tile, then tap it firmly into place.

Table 16. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 5.

Task Name
How to Fit an Electric Shower
How to Install a Grab Bar
How to Install Electric Radiant Heat Mat Under a Tile Floor

Table 17. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 5.

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	8332	Attach the pipe to the shower unit using a compression fitting.
	6772	Pre-drill pilot holes.
	1057	Spread the mortar over a 5- to 10-sq.-ft. area of floor.
	2931	Check to see that the electric shower is heating the water quickly and efficiently.
	8721	Secure a pipe from the cold water main supply to the spot where the shower unit will be mounted.
1 Hop Out-Neighbor	4498	Turn on the water supply and the independent circuit.
	1846	Seal the seams with silicone caulk.
	837	Connect the power lead and thermostat wire to the thermostat, following manufacturer's instructions.
	2931	Check to see that the electric shower is heating the water quickly and efficiently.
	8721	Secure a pipe from the cold water main supply to the spot where the shower unit will be mounted.
2 Hops In-Neighbor	4735	Attach a non-return valve or stop tap to the pipe to isolate the shower's water supply from the rest of the building.
	6528	Mark the location of the stud.
	1725	Install conduit connectors to both ends of two pieces of 1/2-in.
2 Hops Out-Neighbor	2931	Check to see that the electric shower is heating the water quickly and efficiently.
	3147	Test at the end by pulling on it.
	8894	Connect the stop tee to the fill valve.

Table 18. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 5.

Node ID	Step Headline
599	Flip the paper to the front/printed side and use back light to check the placement.
68	Put the tire back on and tighten the lug nuts as best as you can. Put your tire back on the truck and tighten the lug nuts.
9667	Remove the hubcaps and loosen the lug nuts.

Table 19. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 6.

Task Name
How to Foundation Piece a Quilt Block
How to Replace the Front Brake Pads on a 1998 to 2002 Honda Accord
How to Change a CV Axle and Front Wheel Bearing on a 2001 4X4 Dodge Dakota
How to Rotate Tires

Table 20. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 6.

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	3209	Choose the right spring for your car.
	9593	Install the CV axle nut and tighten it securely.
	5567	Use any leftover grease on the bottom bolt if it needs any.
	9857	Find a level work surface.
	599	Loosen the lug nuts of the tire with a lug wrench. Loosen the lug nuts with a lug wrench (tire iron) or impact wrench.
1 Hop Out-Neighbor	5882	Jack the trailer axle up with a bottle or floor jack.
	72	Jack up the vehicle if need be. Jack the car up if necessary. Jack up the vehicle if need be. Jack the car up. Jack up the vehicle. Jack up the car if necessary. Jack up the car. Jack the vehicle up.
	5197	Repeat steps 3-11 on the other side of the vehicle.
	5659	Test drive the vehicle, making sure the steering feels tight, and there are no unusual sounds from the new parts which could indicate improper installation or other damaged parts.
	8608	Raise the car in the air.
2 Hops In-Neighbor	8066	Install the new brake pads in place.
	6789	Get some jack stands.
	4680	Install the antilock sensor cable, anchoring it to the brake fluid line and frame in the same locations the previous cable was anchored, if you have replaced the wheel bearing hub assembly and your vehicle is equipped with four wheel antilock brakes.
2 Hops Out-Neighbor	9351	Remove the lug nuts and wheel with the lug wrench and look at the back of the wheel and hub for grease.
	75	Remove the lug nuts and the wheel. Remove the lug nuts from the wheel to remove it. Remove the lug nuts and pull the wheel off of the hub. Remove the lug nuts from each wheel and remove the wheels from the hubs.
	8221	Check the rotation pattern of your tires.Tires are either directional or non-directional.

Table 21. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 6.

Node ID	Step Headline
7256	Fill a container with cool water and a few drops of shampoo.
4766	Add a few drops of dish soap into the water.
2192	Pour the mixture into a spray bottle.

Table 22. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 7.

Task Name
How to Fix Doll Hair How to Make Hair Spray

Table 23. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 7.

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	49	Brush your hair. Brush out your hair. Brush your hair so that it is ready for styling. Brush the hair. Gently brush the hair. Brush out your hair well.
	3490	Fill a container with cool water.
	416	Remove the saucepan from heat, and let the mixture cool before adding 4 to 5 drops of your favorite essential oil.
1 Hop Out-Neighbor	5372	Place the wig into the water.
	4950	Brush the doll's hair.
	40	Close the bottle, and shake it before you use it. Close the bottle and shake it.
2 Hops In-Neighbor	4449	Remove the doll's wig, if possible.
	2935	Determine what materials the doll and the doll hair are made out of.
	446	Add the coconut oil and stir with a spoon until it has melted.
2 Hops Out-Neighbor	3011	Run clean water over the wig.
	7603	Consider protecting the doll's face.

Table 24. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 7.

Node ID	Step Headline
4251	Put on your full belly dance skirt, the over-skirt, and a belly dance belt and you are ready to dance!
10001	Find suitable tights.
6603	Cut your tights at the chosen length.

Table 25. Pseudo labels of **Video-Node Matching** produced by the PKG for the video segment shown in Fig. 8.

Task Name
How to Make a Full Belly Dance Skirt
How to Make Leggings from Tights

Table 26. Pseudo labels of **Video-Task Matching** (using wikiHow task names) produced by the PKG for the video segment shown in Fig. 8.

Relation Type	Node ID	Step Headline
1 Hop In-Neighbor	3936	Make an over-skirt from leftover fabric.
	9045	Decide upon the length of the leggings you'd prefer.
	2120	Bring the other end of the sling over your other shoulder.
	6953	Pivot your hips forward.
	7311	Repeat the forward lunge.
1 Hop Out-Neighbor	9045	Decide upon the length of the leggings you'd prefer.
	6342	Turn the cut fabric over twice to create a half-inch (1.27cm) hemline.
	2120	Bring the other end of the sling over your other shoulder.
	6953	Pivot your hips forward.
	7311	Repeat the forward lunge.
2 Hops In-Neighbor	2863	Hem the skirt with a machine by rolling the fabric between your thumb and forefinger and guide it through the machine without stretching.
	10001	Find suitable tights.
	7311	Repeat the forward lunge.
2 Hops Out-Neighbor	6603	Cut your tights at the chosen length.
	9276	Stitch the hems in place by hand.
	4370	Continue wrapping the rest of the bandage around the tail.

Table 27. Pseudo labels of **Node Relation Learning** produced by the PKG for the video segment shown in Fig. 8.

Task Name	Step Headline	Node ID
How to Apply MonoKote	Unroll a sheet of MonoKote on a clean work surface, and lay the model part on top of it.	3185
	Remove the clear backing from the MonoKote by attaching a small piece of cellophane tape on each side of the sheet and pulling gently.	2604
	Place the MonoKote, adhesive side down, over the part.	4136
	Set the heat sealing tool to 275 F (135 C) and allow it to come to temperature before proceeding.	1876
	Pull the MonoKote to the edge of the piece to be covered, and run the heat sealing iron along the edge to activate the adhesive and seal MonoKote to the part.	2067
	Unroll a second sheet of MonoKote, and repeat the process on the other side of the aircraft part.	2469
	Trim the edges carefully with the utility knife, leaving a 1/8 inch to 1/4 inch (3 to 6 mm) overlap.	3139
	Finish the edge by carefully folding the top edge under to the bottom side and using the heating iron to seal it in place.	1982
How to Make a Pen Pocket for Your Journal	Decide how tall and wide you want the pen pocket to be.	919
	Add ½ inch (1.27 centimeters) to the length and width.	669
	Cut two pieces of fabric according to your measurements.	1150
	Cut a third piece of for the front pocket.	782
	Cut your elastic 1 inch (2.54 centimeters) longer than your body pieces.	591
	Fold and iron the top hem of the front pocket.	443
	Top stitch the hem down.	1087
	Pin the pocket to the front of one of the larger body pieces.	233
	Create the individual pockets, if desired.	748
	Place the elastic on top of the other body piece.	823
	Pin the pocket piece on top, wrong-side-up.	692
	Sew around the fabric, leaving a gap for turning.	558
	Turn the fabric inside out through the gap.	429
	Press the entire pocket flat with an iron.	583
	Top stitch around the entire piece.	1470
How to Make a Mei Tai Baby Carrier	Obtain the needed supplies listed below.	2296
	Cut an 18–22 inch (45.7–55.9 cm) square of sturdy cloth.	3857
	If you have chosen multiple layers, baste or quilt them together before proceeding.	2648
	Hem two opposite sides of your square.	1965
	Measure your torso.	2373
	Add 15–20 inches (38.1–50.8 cm) to this length to get the length of your two shoulder straps/ties.	2395
	Cut material for two ties (length calculated in previous step) at least 8 inches (20.3 cm) wide.	2483
	Fold the ties in half with narrow ends touching to find the center point of the length of the tie.	1027
	Fold the square in half with the two previously hemmed edges together.	1873
	After finding the center point, mark the length of the (raw edged) side of your square along the long side of the tie from the center point out.	2144
	Remove the tie, fold it in half lengthwise (make a very long fold) and sew the sides closed.	1805
	Miter or clip the seam allowances to allow the corners to turn neatly.	3220
	Repeat measuring, marking and sewing steps for the second tie.	1744
	Place the raw edges of the square into the opening in the ties and pin in place.	1255
	Iron the creases and pin carefully or baste.	1655
	Top stitch the edges of the ties, taking care to sew multiple seams through ALL layers when sewing the raw edges of your square inside the ties.	877
	The pouch is now ready to wear.	6588

Table 28. Pseudo labels of **Task Context Learning** produced by the PKG for the video segment shown in Fig. 1.

Task Name	Step Headline	Node ID
How to Ink Stamps with Markers	Select the right types of markers.	3661
	Sand your stamp lightly.	8141
	Spritz your stamps with water if you want a water color effect.	8361
	Color with the brush tip of your marker.	2811
	Use different colors as needed.	8130
	Breathe on the stamp to moisten the ink.	3336
	Stamp the image onto the page.	5354
	Experiment a little.	2985
	Work quickly so the ink stays wet.	4496
	Do not touch your inked image until it's dry.	8988
	Avoid permanent markers on stamps.	8589
How to Make a Translation Tessellation	Find an A4 size piece of paper.	7026
	Cut out a small square or parallelogram.	7447
	Convert this base tessellation into a more interesting shape.	8591
	Trace it on the 3" x 6" (7.5cm x 15cm) paper until it is full.	5543
	Go over the lines with a black permanent marker.	9399
	Color it in anyway you like.	8492
	Ink it all over.	6068
	Let dry.	236

Table 29. Pseudo labels of **Task Context Learning** produced by the PKG for the video segment shown in Fig. 2.

Task Name	Step Headline	Node ID
How to Make Milk Paint	Let the milk come to room temperature.	3463
	Combine the milk and vinegar.	3523
	Place the milk in a warm place for one to two days.	3598
	Add water to the pigment.	2932
	Make a paste with the pigment.	4936
	Add water to the lime powder.	4730
	Mix the lime powder and water to make a wet paste.	5370
	Line a colander with cheesecloth and put it in the sink.	1454
	Pour the container of curdled milk over the colander.	2289
	Transfer the quark into a paint bucket.	5349
	Add the lime paste to the quark.	2459
	Stir in the pigment paste.	2374
	Strain the paint through a cheesecloth.	8226
	Prepare the surface you are painting.	4614
	Stir the milk paint.	3364
	Apply the first layer of paint.	2001
	Let the first layer dry.	4198
	Stir the paint and apply the second layer.	2929
	Apply a topcoat if desired.	303
	Store milk paint in the fridge for up to three days.	3008
How to Make Shimmering Finger Paints	Mix all the ingredients together in a medium pan.	8766
	Cook over low heat for 10 to 15 minutes.	8052
	Keep stirring the finger paint mixture until it is smooth and thick.	5660
	After the finger paint has thickened, take the pan off the stove.	5957
	Place the paints in containers.	5504
	Add a few drops of food coloring to each container.	8702
	Try mixing colors or paints to get different colors.	9377
	Stir the coloring into the paint to determine the shade of color.	9026
	If using, add the desired amount of glitter.	7885
	Stir well to get the glitter really mixed in.	6937
	Cover tightly to store.	5149

Table 30. Pseudo labels of **Task Context Learning** produced by the PKG for the video segment shown in Fig. 3.

Task Name	Step Headline	Node ID
How to Cluster Stitch	Tie the yarn onto your hook.	67
	Create a foundation chain.	403
	Chain two.	418
	Yarn over the hook.	67
	Insert the hook into the next stitch.	37
	Yarn over and draw through.	493
	Yarn over and draw through twice.	493
	Yarn over the hook and insert it into the next stitch.	10016
	Yarn over and draw through.	493
	Yarn over and draw through twice.	493
	Yarn over the hook.	67
	Pull the yarn through all three loops.	4836
	Chain two.	418
	Yarn over and insert the hook.	67
	Yarn over and draw through.	493
	Yarn over and pull through twice.	17
	Work the next part into the next stitch.	568
	Yarn over and pull through.	17
	Wrap the yarn around and pull through twice.	17
	Work into the next stitch.	568
	Draw another loop.	284
	Yarn over and pull through twice.	17
	Pull through the remaining loops.	4664
	Work one cluster stitch into a single stitch.	4974
	Chain two.	418
	Work a second cluster stitch into the same stitch.	2909
	Chain two and repeat as needed.	7757
How to Crochet Popcorn Stitch	Double crochet five times into one stitch.	4989
	Insert the hook into the first stitch of the group.	37
	Close the group.	5520
	Separate your stitches.	8514
	Create a foundation chain.	403
	Single crochet into each stitch.	310
	Single crochet twice.	243
	Make five double crochets.	304
	Slip stitch into the loops.	8616
	Single crochet into the next three stitches.	9774
	Repeat across the row.	300
	Single crochet into each stitch.	310
	Repeat as needed.	1201
	Create a foundation chain.	403
	Double crochet into each stitch.	243
	Chain three.	8662
	Double crochet into the next five stitches.	243
	Make five double crochets.	304
	Resituate the hook and close the group.	6878
	Double crochet to the end of the row.	2437
	Repeat as needed.	1201

Table 31. Pseudo labels of **Task Context Learning (Part I)** produced by the PKG for the video segment shown in Fig. 4.

Task Name	Step Headline	Node ID
How to Surface Crochet	Insert the hook into the first stitch.	37
	Tie the yarn onto the hook.	67
	Pull up a loop.	588
	Insert the hook into the next spot.	8289
	Yarn over.	144
	Pull up the loop.	588
	Pull the second loop through the first.	8430
	Repeat as needed.	1201
	Bind off the yarn.	137
How to Half Double Crochet	Yarn over.	144
	Insert the hook into the stitch.	37
	Yarn over.	144
	Draw up another loop.	284
	Yarn over.	144
	Draw through all three loops on the hook.	2027
	Create a foundation chain.	403
	Skip the first two chain stitches.	2016
	Work a half double crochet.	625
	Work another half double crochet.	625
	Repeat across the chain.	4574
	Create a turning chain.	213
	Skip a stitch.	2529
	Work a half double crochet into the next stitch.	243
	Repeat across the row.	300
How to Treble Crochet	Make a slip knot.	241
	Yarn over.	144
	Draw the yarn through the slipknot.	9994
	Crochet a chain of your desired length.	4798
	Crochet a turning chain.	180
	Turn your work.	79
	Yarn over twice.	622
	Insert your hook.	450
	Yarn over and draw through.	493
	Yarn over and draw through two loops.	624
	Yarn over and draw through two loops.	624
	Yarn over and draw through the last two.	4860
	Yarn over twice.	622
	Insert your hook.	450
	Yarn over and draw through.	493
	Yarn over and draw through two loops.	624
	Yarn over and draw through two loops.	624
	Yarn over and draw through two loops.	624
	Repeat steps 1-6 of this section.	3526
	Turn your work.	79
	Crochet a turning chain.	180
	Yarn over twice and insert hook.	67
	Yarn over one and pull through two.	4922
	Yarn over and pull through.	17
	Repeat steps 3-5 until you reach the end of your chain.	3464

Table 32. Pseudo labels of **Task Context Learning (Part II)** produced by the PKG for the video segment shown in Fig. 4.

Task Name	Step Headline	Node ID
How to Fit an Electric Shower	Pick a location for your electric shower that is near the main cold water supply and close to a spot where you can install an independent circuit.	8120
	Consult an electrician for advice on the size and type of independent circuit to install for your shower to ensure that you have adequate electricity.	1810
	Consult a plumber to ensure that your building's plumbing system will be able to accommodate an electric shower.	4289
	Install the independent circuit near the location of the electric shower along with any necessary consumer units or earth cables.	4465
	Wire the independent circuit to an isolating switch, which should be located above the shower.	4893
	Attach the electric cable from the isolating switch to the back of the electric shower power unit and wire accordingly.	8259
	Secure a pipe from the cold water main supply to the spot where the shower unit will be mounted.	8721
	Attach a non-return valve or stop tap to the pipe to isolate the shower's water supply from the rest of the building.	4735
	Attach the pipe to the shower unit using a compression fitting.	8332
	Mount the shower unit and the shower head to the wall.	3765
	Turn on the water supply and the independent circuit.	4498
	Check to see that the electric shower is heating the water quickly and efficiently.	2931
How to Install a Grab Bar	Assemble the necessary tools.	4193
	Examine the grab bar kit.	1614
	Determine mounting location.	4317
	Mark the location of the stud.	6528
	Pre-drill pilot holes.	6772
	Install the wall anchors.	3616
	Seal the seams with silicone caulk.	1846
	Test at the end by pulling on it.	3147
How to Install Electric Radiant Heat Mat Under a Tile Floor	Prepare your floor for tile by installing tile backer board on the floor, securing it to the existing sub-floor with thin-set mortar and cement board screws or nails.	1464
	Make a scale drawing of the bathroom floor, including toilet, tub and vanity locations, and bring it to the tile store or home center so you can buy the proper size mat or combination of mats.	888
	Check the wiring with a continuity tester, after purchasing the mat, to make sure it wasn't damaged during manufacturing or shipping.	2409
	Install an electrical outlet box 5 feet (1.5 m).	377
	Note: following your preliminary layout, you should mark the path of the thick "power lead" between the mat and wall cavity and chisel a shallow trench into the floor.	1547
	Draw the layout lines for the tile on the floor.	1828
	Install the mat, securing it lightly to the cement board with double-face tape.	1183
	Check the mat wiring again with the continuity tester.	1264
	Install conduit connectors to both ends of two pieces of 1/2-in.	1725
	Spread the mortar over a 5- to 10-sq.-ft. area of floor.	1057
	Lay the tile, then tap it firmly into place.	2823
	Connect the power lead and thermostat wire to the thermostat, following manufacturer's instructions.	837

Table 33. Pseudo labels of **Task Context Learning** produced by the PKG for the video segment shown in Fig. 5.

Task Name	Step Headline	Node ID
How to Foundation Piece a Quilt Block	Select your pattern and photocopy or print enough of them to make your quilt.	1075
	Select your fabrics.	200
	Launder all of your fabrics.	726
	Iron the fabrics smooth if necessary.	1215
	Cut rectangles or squares in sizes which will cover the shapes in your pattern blocks.	915
	Notice that the pattern pieces are numbered in the order in which you should sew the pieces.	2310
	Place the cloth for piece #1 on the BACK side of the paper with the back/wrong side of the cloth towards the paper.	895
	Hold the paper up to a light to verify that the fabric is oriented so that it covers all of the area of piece one with at least a quarter inch of overlap in all directions.	564
	Place the cloth for piece #2 (white) with its right/front side facing the right/front side of piece #1 (red) and its seam edge aligned with the seam line and overlapping by a minimum of a quarter inch.	108
	Pin the two fabrics in place on the paper.	191
	Flip the paper to the front/printed side.	1112
	Machine stitch the seam line from the printed side.	496
	Trim the seam allowances to 1/4 inch (0.6 cm).	74
	Unpin the fabrics and flip piece #2 over the seam and pin it in place over its allotted area on the block.	411
	Hold up the paper block pattern to the light to check that piece #2 will cover its allotted area.	348
	Place the cloth for piece #3 with its right/front side facing the right/front side of piece #2... and its seam edge aligned with the seam line and overlapping by a minimum of a quarter inch (6 mm).	108
	Pin the two fabrics in place on the paper.	191
	Flip the paper to the front/printed side and use back light to check the placement.	1599
	Machine stitch the seam line from the printed side.	496
	Trim the seam allowances to 1/4 inch (6 mm).	74
	Unpin the fabrics and flip piece #3 over the seam and pin it in place over its allotted area on the block.	411
	Repeat the process of placing, pinning, checking then sewing and trimming for each successively numbered piece.	765
	Machine baste around the perimeter of your block when complete.	432
	Before trimming - note the ragged edges.	1017
	Tear away the paper “backing”.	388
	Voilà!	485
How to Replace the Front Brake Pads on a 1998 to 2002 Honda Accord	Make sure the vehicle is on a level surface, the transmission is in park, and the emergency brake is engaged.	8099
	Starting on either side, locate the brake caliper (which is directly behind the tire).	6849
	Lift the bottom of the caliper so that it pivots off of the top bolt.	6132
	Remove both brake pads.	9478
	Use the brake cleaner to spray down the entire brake assembly.	6835
	On the inside of the brake caliper is the piston which extends out towards the back pad.	6536
	Apply brake grease to the new pads.	9816
	Install the new brake pads in place.	8066
	Use any leftover grease on the bottom bolt if it needs any.	5567
	Put the tire back on and tighten the lug nuts as best as you can.	68
	Repeat steps 3-11 on the other side of the vehicle.	5197
	Clear any tools from under the vehicle, and pump the brakes until they begin to firm up.	8864

Table 34. Pseudo labels of **Task Context Learning (Part I)** produced by the PKG for the video segment shown in Fig. 6.

Task Name	Step Headline	Node ID
How to Change a CV Axle and Front Wheel Bearing on a 2001 4X4 Dodge Dakota	Troubleshoot your front wheel drive system to determine if maintenance is needed.	7869
	Look underneath the truck between the wheels and the front differential to see if any visible damage is evident.	5109
	Drive the truck on a smooth surface with the windows down, listening for clicking, grinding, or other unusual noises from the front of your truck.	8245
	Jack up the front end of your truck and shake the wheels to see if any play is evident.	9951
	Rotate the tires while they are off the ground, listening and feeling for anything unusual.	8293
	Rock the wheels back and forth and listen for sounds from the CV joints.	3486
	Have a mechanic do the inspection if you have any doubts, since the parts needed for this project cost several hundred dollars.	3976
	Make a list of the parts you need to complete the project.	2874
	Write down the information on your truck before going shopping.	6557
	Choose the quality of the parts you want to use.	7493
	Jack up the truck and set it on jack stands, making sure the wheels are blocked to prevent it from rolling.	7520
	Remove the wheel from the side you are working one.	8910
	Remove the cotter pin from the axle shaft at the center of the hub.	3599
	Remove the brake caliper.	139
	Remove the brake rotor.	9584
	Remove the axle nut.	9066
	Remove the hub bearing assembly.	3789
	Remove the upper ball joint from the knuckle or unbolt it from the upper control (A) arm.	8639
	Pull the CV axle off of the differential side shaft.	9496
	Remove the oil seal at the side shaft if you are replacing it.	9112
	Install the new oil seal at the side shaft.	8206
	Slide the new CV axle over the side shaft on the differential.	7638
	Guide the outer end of the axle through the center hole in the knuckle.	6924
	Install the new hub bearing assembly.	7721
	Reinstall (or install a new) the upper ball joint.	6509
	Replace the brake rotor.	4759
	Reinstall the brake caliper.	4534
	Install the antilock sensor cable, anchoring it to the brake fluid line and frame in the same locations the previous cable was anchored, if you have replaced the wheel bearing hub assembly and your vehicle is equipped with four wheel antilock brakes.	4680
	Install the CV axle nut and tighten it securely.	9593
	Put your tire back on the truck and tighten the lug nuts.	68
	Test drive the vehicle, making sure the steering feels tight, and there are no unusual sounds from the new parts which could indicate improper installation or other damaged parts.	5659
	Return any parts which had a core charge to the retailer you purchased them from, for a refund of the core charge.	7860

Table 35. Pseudo labels of **Task Context Learning (Part II)** produced by the PKG for the video segment shown in Fig. 6.

Task Name	Step Headline	Node ID
How to Rotate Tires	Get some jack stands.	6789
	Find a level work surface.	9857
	Remove the hubcaps and loosen the lug nuts.	9667
	Raise the car in the air.	8608
	Check the rotation pattern of your tires.Tires are either directional or non-directional.	8221
	Remove the lug nuts from the first tire you've raised and remove it.	4386
	Rotate the tires in the correct pattern.	6012
	Lower the car.	28
	Tighten lug nuts using the star pattern.	8878
	Place hubcaps back on the wheels by replacing the lug nuts.	9528

Table 36. Pseudo labels of **Task Context Learning (Part III)** produced by the PKG for the video segment shown in Fig. 6.

Task Name	Step Headline	Node ID
How to Fix Doll Hair	Determine the problem.	4443
	Start with brushing your doll's hair.	4572
	Consider trimming the doll hair.	8981
	Consider curling the hair.	9590
	Consider washing the doll hair.	4647
	Determine what materials the doll and the doll hair are made out of.	2935
	Fill a container with cool water.	3490
	Add a few drops of dish soap into the water.	4766
	Brush the doll's hair.	4950
	Consider protecting the doll's face.	7603
	Dip the doll's hair into the water.	3560
	Lather the doll's hair.	4682
	Rinse the doll hair with clean, cool water.	8298
	Consider using conditioner to detangle the doll's hair.	3654
	Consider styling the hair.	1091
	Set the hair on a towel to dry.	4713
	Brush the hair.	49
	Remove the doll's wig, if possible.	4449
	Gently brush the hair.	49
	Fill a container with cool water and a few drops of shampoo.	7256
	Place the wig into the water.	5372
	Run clean water over the wig.	3011
	Consider a vinegar soak to make the hair shiny.	6070
	Lay the wig on a towel.	205
	Place another towel over the wig.	205
	Place the wig onto a fresh towel.	205
	Glue the wig back onto the doll head.	3884
	Consider styling the hair.	1091
	Consider curling your doll's hair.	4690
	Obtain something to curl your doll hair with.	4488
	Wrap some hair around your curler.	3074
	Secure the curler.	1403
	Repeat for the rest of the hair.	8562
	Wait until the hair dries.	3287
	Take out the curlers.	3945

Table 37. Pseudo labels of **Task Context Learning (Part I)** produced by the PKG for the video segment shown in Fig. 7.

Task Name	Step Headline	Node ID
How to Make Hair Spray	Bring 1 cup (240 milliliters) of water to a simmer in a saucepan.	644
	Stir in 1 tablespoon of sea salt.	777
	Add the coconut oil and stir with a spoon until it has melted.	446
	Remove the saucepan from heat, and let the mixture cool before adding 4 to 5 drops of your favorite essential oil.	416
	Pour the mixture into a spray bottle.	2192
	Close the bottle, and shake it before you use it.	40

Table 38. Pseudo labels of **Task Context Learning (Part II)** produced by the PKG for the video segment shown in Fig. 7.

Task Name	Step Headline	Node ID
How to Make a Full Belly Dance Skirt	Select your fabric.	200
	Cut the panels.	6480
	Fold each panel in half.	6304
	Sew your panels all together in a row, selvage to selvage starting at the hip cut, ignoring any difference at the bottom, as this is the side you will cut and hem.	4673
	Put your skirt on a clamping slack hanger and hang it high.	7088
	Put on your giant, uncut skirt.	8316
	Hem the skirt with a machine by rolling the fabric between your thumb and forefinger and guide it through the machine without stretching.	2863
	Make an over-skirt from leftover fabric.	3936
	Put on your full belly dance skirt, the over-skirt, and a belly dance belt and you are ready to dance!	4251
How to Make Leggings from Tights	Find suitable tights.	10001
	Decide upon the length of the leggings you'd prefer.	9045
	Cut your tights at the chosen length.	6603
	Turn the cut fabric over twice to create a half-inch (1.27cm) hemline.	6342
	Stitch the hems in place by hand.	9276
	If wished, you can add embroidery, lace, beads, sequins, etc.	6563
	Wear your new leggings.	6513

Table 39. Pseudo labels of **Task Context Learning** produced by the PKG for the video segment shown in Fig. 8.