SemTra: A Semantic Skill Translator for Cross-Domain Zero-Shot Policy Adaptation

Sangwoo Shin, Minjong Yoo, Jeongwoo Lee, Honguk Woo*

Department of Computer Science and Engineering, Sungkyunkwan University
jsw7460@skku.edu, mjyoo2@skku.edu, ljwoo98@skku.edu, hwoo@skku.edu

Abstract

This work explores the zero-shot adaptation capability of semantic skills, semantically interpretable experts' behavior patterns, in cross-domain settings, where a user input in interleaved multi-modal snippets can prompt a new long-horizon task for different domains. In these cross-domain settings, we present a semantic skill translator framework SemTra which utilizes a set of multi-modal models to extract skills from the snippets, and leverages the reasoning capabilities of a pretrained language model to adapt these extracted skills to the target domain. The framework employs a two-level hierarchy for adaptation: task adaptation and skill adaptation. During task adaptation, seq-to-seq translation by the language model transforms the extracted skills into a semantic skill sequence, which is tailored to fit the cross-domain contexts. Skill adaptation focuses on optimizing each semantic skill for the target domain context, through parametric instantiations that are facilitated by language prompting and contrastive learning-based context inferences. This hierarchical adaptation empowers the framework to not only infer a complex task specification in one-shot from the interleaved multi-modal snippets, but also adapt it to new domains with zero-shot learning abilities. We evaluate our framework with Meta-World, Franka Kitchen, RLBench, and CARLA environments. The results clarify the framework's superiority in performing long-horizon tasks and adapting to different domains, showing its broad applicability in practical use cases, such as cognitive robots interpreting abstract instructions and autonomous vehicles operating under varied configurations.

1 Introduction

The promise of zero-shot policy deployment across different domains stems from the capability to immediately adapt its pre-trained knowledge to unfamiliar environments without the need for extensive data collection or fine-tuning. Such capability could revolutionize fields where safety is crucial, and where a single failure can lead to significant consequences and substantial costs, such as autonomous driving and robotics. However, achieving robust zero-shot adaptation in these fields remains challenging due to the intricacies of given tasks and the dynamic nature of their deployment environments (Chalaki et al. 2019; Harrison et al. 2017).

*Honguk Woo is the corresponding author.

In the realm of reinforcement learning (RL) and imitation learning, policy adaptation has seen some exploration, with researchers leveraging various forms of task specification to deduce the given task domain. For instance, several studies utilized video demonstrations or expert trajectories from similar domains (Zhao et al. 2022; Kim et al. 2020), while others pivoted towards language instructions provided by users (Goyal, Mooney, and Niekum 2023; Ichter et al. 2022), or video-language interleaved demonstrations in multi-modal user interfaces (Jiang et al. 2023). Furthermore, in (Xu et al. 2022), expert trajectories from the target domain were directly employed for task inference.

In this work, we investigate the problem of cross-domain zero-shot policy adaptation, with a focus on an inclusive and generalized approach to process user inputs relevant to long-horizon tasks across diverse domains. To be specific, we consider situations where the policy is prompted with a task-relevant input, presented as multi-modal interleaved snippets encompassing video, sensor data, and textual elements. In light of this consideration, we introduce a novel framework SemTra, grounded in the notion of semantic skill translation that spans multi-domains. The framework is designed with a two-tiered hierarchical adaptation process, first at the task level, followed by the skill level, to enhance the potential for zero-shot policy adaptation to target domains.

During task adaptation, our framework harnesses the reasoning strengths of Pretrained Language Models (PLMs) to interpret the task specification (Ichter et al. 2022; Huang et al. 2022). This specification, encapsulated within a sequence of multi-modal interleaved snippets, is transformed into a sequence of semantically interpretable skill representations (semantic skills). For skill adaptation, the framework employs a parametric approach to instantiate skills, integrating domain contexts into the semantic skills. These contexts may not be solely encapsulated in the task specification or input snippets, as they can also be dynamically captured during the evaluation time through interactions with the deployment environment. Subsequently, these domain contexts form domain-randomized meta-knowledge, enabling the framework to achieve context-parameterized skill instantiation.

The contributions of our work are summarized as follows. (i) We present the novel framework SemTra, addressing the challenging yet practical issues of cross-domain zero-shot

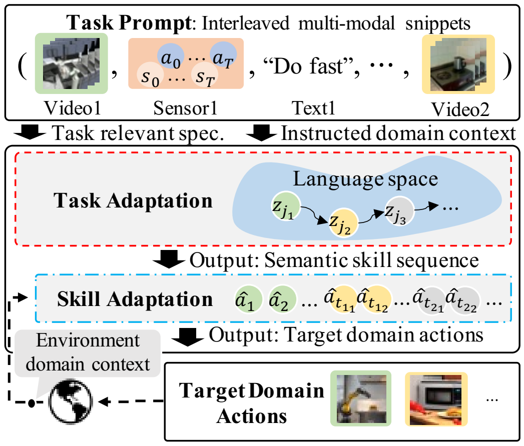


Figure 1: Cross-domain zero-shot adaptation for a multi-modal task prompt: our framework is given a task prompt filled with partial demonstrations and instructed contextual cues in multi-modal snippets. The framework conducts a two-phase adaptation, initially translating the snippets to semantic skills at the task level, and subsequently optimizing them into actions for the target domain at the skill level.

policy adaptation for long-horizon tasks. (ii) We devise a robust, hierarchical adaptation algorithm that leverages PLM prompting for the task-level adaptation and disentangles semantic skills and domain contexts to enable the skill-level optimization through parametric instantiations across different domains. (iii) We intensively evaluate our framework in over 200 cross-domain scenarios with Meta-World, Franka Kitchen, RLBench, and CARLA simulation environments, demonstrating its broad applicability in practical use cases, such as cognitive robots and autonomous driving.

2 Problem Formulation

We consider a multi-modal *task prompt* $\mathcal{P} := (x_1, \cdots, x_N)$ comprising a sequence of snippets, where each snippet x_i is given in one of several modalities such as video, sensor data, and text. We presume that $\mathcal{P}$ presents a source task specification, and some snippets may contain pertinent information or guidance beneficial for cross-domain adaptation. Given a single task prompt $\mathcal{P}$, our work is to establish a framework to find the optimal model ϕ^*. The model maps $\mathcal{P}$ to a policy $\pi_{\mathcal{P}}$, maximizing the expected evaluation performance:

$$\phi^* = \underset{\phi}{\text{argmax}} \left[\underset{(D,\mathcal{P})\sim p_{\text{tgt}}(D,\mathcal{P})}{\mathbb{E}} [\mathrm{E}_D(\phi(\mathcal{P}))] \right]. \quad (1)$$

Here, $E_D(\phi(\mathcal{P}))$ is an evaluation function for the policy $\phi(\mathcal{P})$ in a target domain D.

As illustrated in Figure 1, the task prompt is comprised of multiple modalities; e.g., the Video1 snippet offers a visual demonstration of an expert performing "turning on the kitchen lights", the Sensor1 snippet captures the trajectory of "turning on the microwave.", the Video2 snippet visually represents the process of "boiling the water in the kettle", and the Text1 snippet "do fast" contains contextual clues about the task execution in the environment, such as a fast-food restaurant. These snippets collectively formulate the source task specification and also serve as a guide for adaptation in the target domain. The evaluation function $E_D(\phi(\mathcal{P}))$ in (1) assesses the task performance, particularly concerning cooking and serving maneuvers in the target domain D with some stringent time constraints in this case.

3 Approach

3.1 Overall Framework

To tackle the challenge of cross-domain zero-shot adaptation with a multi-modal task prompt, we develop a semantic skill translator framework SemTra. As illustrated in Figure 1, SemTra decomposes the cross-domain adaptation hierarchically into two-phase linguistic adaptation procedures. In the task adaptation phase, the framework captures the task discrepancy between the source and target domains by extracting a domain-invariant task solution using semantic skills. To do so, a pretrained vision-language model (VLM) is leveraged to establish the domain-agnostically learned semantic skills from the video snippets. The knowledge of PLM is also used to align the semantic skill sequence with the subtasks in the target domain. In the skill adaptation phase, the semantic skills are instantiated into an actual action sequence optimized for the target domain.

3.2 Task Adaptation

For a task prompt $\mathcal{P} = (x_1, \cdots x_N)$, our framework interprets it to produce a sequence of semantic skill embeddings z_t. These embeddings are represented in a language space and correspond to successive timesteps t during task execution in the target environment.

$$\Phi_D \circ \Phi_E : (s_t, \mathcal{P} = (x_1, \cdots, x_N)) \mapsto z_t \quad (2)$$

This adaptation involves multi-modal skill encoders Φ_E and a semantic skill decoder Φ_D, where each encoder takes the input specification as outlined by the task prompt, and the decoder maps the encoded specification into the semantic skills to be executed in the target environment. To train the encoder and decoder models, we use a dataset $\mathcal{D} = \{(\mathcal{P}_i, \tau_i)\}_i$, where $\mathcal{P}_i$ is a task prompt and τ is an expert trajectory in the target environment. Specifically, we use $\tau = \{(s_t, a_t, v_t, l_t)\}_t$ with state s_t, action a_t, visual observation v_t, and semantic annotation l_t, where l_t is a description of a skill or behavior at timestep t, e.g., "push a button", following (Shin et al. 2023; Ichter et al. 2022; Pertsch et al. 2022; Shridhar, Manuelli, and Fox 2022).

Multi-modal skill encoders. The encoders Φ_E convert a task prompt into a skill-level language instruction η as

$$\Phi_E : \mathcal{P} = (x_1, \cdots, x_N) \mapsto \eta := (l_{i_1}^{(src)}, \cdots, l_{i_I}^{(src)}) \quad (3)$$

where each $l_{i_j}^{(src)}$ is a language token. In this work, we consider three different modalities in snippets (i.e., video, sensor data, and text), although another modality can be incorporated through its respective encoder implementation.

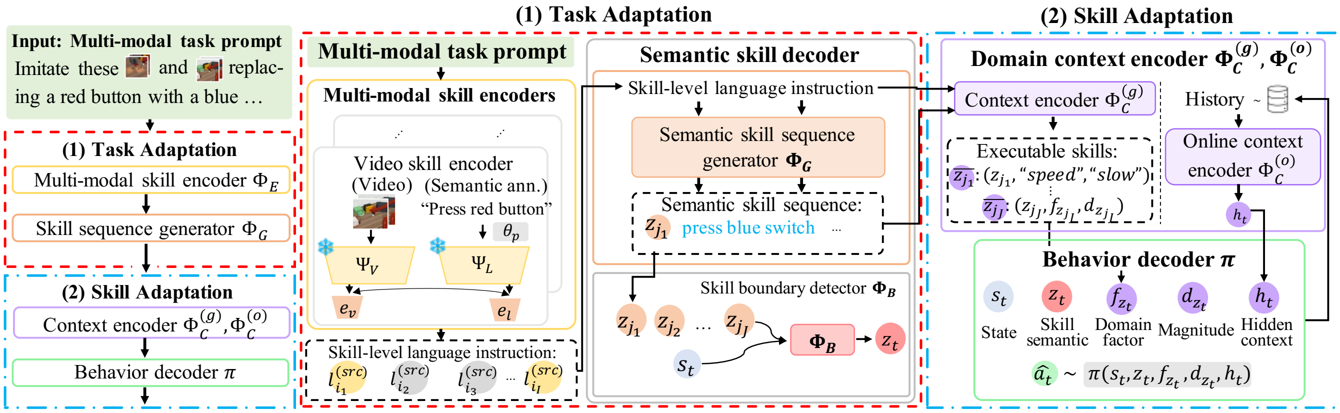


Figure 2: Two-phase policy adaptation in SemTra. (1) Task adaptation: the multi-modal skill encoders Φ_E produce a skill-level language instruction from the task prompt. In the figure, we specifically describe the training of a video skill encoder, contrastively learned through a pretrained VLM (Ψ_V, Ψ_L). The skill-level instruction is then translated into a semantic skill sequence through the semantic skill sequence generator Φ_G based on a PLM. The skill boundary detector Φ_B infers the boundary of semantic skills upon a current state in the target domain. (2) Skill adaptation: the context encoder $\Phi_C^{(g)}$ identifies instructed domain contexts using both the skill-level instruction and the semantic skill sequence, generating an executable skill sequence. The online context encoder $\Phi_C^{(o)}$ captures environment hidden contexts at runtime. The behavior decoder π generates actions optimized for the target domain, based on the executable skills and environment hidden contexts, along with the current state. 1

To train the video skill encoder Ψ_V, which maps a video snippet $x = (v_{0:T})$ to skill-level language tokens, we leverage the video-to-text retrieval and text prompt learning with the pretrained VLM, V-CLIP (Xu et al. 2021), in a similar way of (Shin et al. 2023). Specifically, consider the V-CLIP (Ψ_V, Ψ_L) model of a video encoder Ψ_V and a text encoder Ψ_L. A text prompt θ_p, a small set of learnable parameters in Ψ_L is learned using a contrastive learning loss (van den Oord, Li, and Vinyals 2018) on a batch $\{(v_{t-H:t}, l_t)\} \sim \mathcal{D}$. This optimization increases the similarity of video embedding $\Psi_V(v_{t-H:t})$ and its language embedding counterpart $\Psi_L(l_t; \theta_p)$. Once θ_p is trained, we use the retrieved semantic skill from the video snippet, i.e., $\Psi_V(v_{t-H:t}) :=$

$$\text{argmax}_{l \in \mathcal{L}_{vl}} \text{sim}(\Psi_V(v_{t-H:t}), \Psi_L(l; \theta_p)) \tag{4}$$

where $\text{sim}(\cdot, \cdot)$ is a similarity function and $\mathcal{L}_{vl} = \bigcup_{l \in \tau \in \mathcal{D}} \{l\}$ denotes a set of textually represented semantically interpretable skills. For sensor data snippets, we use a supervised classifier which predicts the semantic annotation l_t from a state-action sub-trajectory $(s_{t-H':t}, a_{t-H':t})$. For text snippets, we use an identity function.

Semantic skill decoder. Given a skill-level instruction η in (3) and state s_t, the semantic skill decoder Φ_D yields a semantic skill z_t for each timestep t, i.e.,

$$\Phi_D : (s_t, \eta) \mapsto z_t \tag{5}$$

by using two models, the semantic skill sequence generator Φ_G and the skill boundary detector Φ_B. The former Φ_G generates a sequence of target semantic skill embeddings through seq-to-seq translation, leveraging the knowledge of PLMs.

$$\Phi_G : \eta \mapsto (z_{j_1}, \cdots, z_{j_J}) \tag{6}$$

At each timestep t, the latter Φ_B determines whether to continue executing the current skill z_t or to transition to the next skill in the sequence until the task is completed. It evaluates the probability that ongoing semantic skill is terminated:

$$\Phi_B : (s_t, z_t, s_{t_0}) \mapsto p \in [0, 1] \tag{7}$$

with state s_t, semantic skill z_t, and its initial state s_{t_0}.

For training Φ_G, we exploit the skill encoders Φ_E that can extract a target semantic sequence $(z_{j_1}, \cdots, z_{j_J})$ from the video trajectory of $\tau^{(tgt)}$ through (4). Subsequently, Φ_G is supervised using a cross-entropy-based loss on these target semantic sequences. Regarding Φ_B, it is trained using a binary cross-entropy loss, i.e., $\text{BCE}(\Phi_B(s_t, z_t, s_{t_0}), 1_{z_t \neq z_{t+1}})$. This task adaptation procedure is presented in the middle of Figure 2.

3.3 Skill Adaptation

In conjunction with the task adaptation, which generates a semantic skill sequence, the framework also encompasses a skill adaptation phase, where each skill is adapted individually to suit the target domain. It takes as input the semantic skill sequence $\xi := (z_{j_1}, \cdots, z_{j_J})$ in (6), paired with a skill-level language instruction η in (3), and the current state s_t. It continuously generates an action a_t for each timestep during the execution of ξ in the target domain, employing a context encoder $\Phi_C^{(g)}$ and a behavior decoder π.

$$\pi \circ \Phi_C^{(g)} : (s_t, \eta, \xi) \mapsto a_t \tag{8}$$

Domain context encoder. The context encoder $\Phi_C^{(g)}$ transforms a semantic skill sequence into a series of executable skills appropriate for the target domain, by taking into account the cross-domain contexts encapsulated in a given task

Algorithm 1: Two-phase adaptation of SemTra

Task prompt $\mathcal{P}$, target trajectory $\tau^{(tgt)} = \{(s_t, v_t, a_t, l_t)\}$
Domain contexts $\{\bar{z}'_{j_1}, \cdots, \bar{z}'_{j_J}\}$ of $\tau^{(tgt)}$
Multi-modal skill enc. Φ_E, skill sequence generator Φ_G
Skill boundary detector Φ_B, context encoder $\Phi_C^{(g)}$
Online context encoder $\Phi_C^{(o)}$, behavior decoder π
/* Task adaptation in Section 3.2 */
$\eta \leftarrow \Phi_E(\mathcal{P})$ using (3), $\xi \leftarrow \Phi_G(\eta)$ using (6)
$\xi' \leftarrow \{\Psi_V(v_{t-H:t}) : t\}$ with deduplication using (4)
$\Phi_G \leftarrow \Phi_G - \nabla_{\Phi_G}(\text{CrossEntropy}(\xi, \xi'))$
$(z_0, \cdots, z_T) \leftarrow \{\Psi_V(v_{t-H:t}) : t\}$ using (4)
$\Phi_B \leftarrow \Phi_B - \nabla_{\Phi_B}(\text{BCE}(\Phi_B(s_t, z_t, s_{t_0}), 1_{z_t \neq z_{t+1}}))$
/* Skill adaptation in Section 3.3 */
$\{\bar{z}_{j_1}, \cdots, \bar{z}_{j_J}\} \leftarrow \Phi_C^{(g)}(\eta, \xi)$ using (9)
$\Phi_C^{(g)} \leftarrow \Phi_C^{(g)} - \nabla_{\Phi_C^{(g)}}(\text{CrossEntropy}(\bar{z}_{j_i}, \bar{z}'_{j_i}))$
$h_t \leftarrow \Phi_C^{(o)}(s_{t-H':t}, a_{t-H':t-1})$ using (11)
$\pi \leftarrow \pi - \nabla_\pi(\text{MSE}(\pi(s_t, \bar{z}_t, h_t), a_t)$

prompt. This involves utilizing the knowledge of PLMs to capture domain-specific features from the coupled skill-level instruction η and its associated semantic skill sequence ξ.

$$\Phi_C^{(g)} : (\eta, \xi) \mapsto (\bar{z}_{j_1}, \cdots, \bar{z}_{j_J}), \text{ where } \bar{z} := (z, f_z, d_z) \quad (9)$$

Here, each executable skill $\bar{z}$ is parameterized by a semantic skill $z \in \xi$, a domain factor f_z, and a magnitude d_z associated with f_z. The cross-domain context information in η is used to infer the domain factor, which modulates the actual action execution of each semantic skill according to the specific requirements of the target domain; e.g., the domain factor for the semantic skill "move to a destination" could be the surface type or the required speed in the target domain.

Behavior decoder. Given an executable skill sequence $(\bar{z}_{j_1}, \cdots, \bar{z}_{j_J})$ in (9), the behavior decoder π is learned to infer an action sequence optimized for the target domain.

$$\pi : (s_t, \bar{z}_t) \mapsto a_t \quad (10)$$

Furthermore, to adapt to temporal domain shifts that are not fully covered by the task prompt, an online context encoder $\Phi_C^{(o)}$ is utilized to infer a hidden context h_t over timesteps t:

$$\Phi_C^{(o)} : (s_{t-H':t}, a_{t-H':t-1}) \mapsto h_t. \quad (11)$$

Then, this hidden context is concatenated with $\bar{z}_t$ in (10). This $\Phi_C^{(o)}$ can be implemented by using reconstruction (Zintgraf et al. 2020; Rakelly et al. 2019) or contrastive learning (Shin et al. 2023; Qiu et al. 2022). This skill adaptation procedure is on the right side of Figure 2. Algorithm 1 outlines the two-phase adaptation in SemTra.

4 Evaluation

4.1 Experiment Setting

Environments. For cross-domain adaptation evaluations, we use the Franka Kitchen (Fu et al. 2020) (FK) and Meta-World (Yu et al. 2019) (MW) environments. In these environments, each task $\mathcal{T}$ is composed of N subtasks that must be completed in the correct order. This allows us to explore the compositionality of complex tasks and investigate the long-horizon imitation performance (Shin et al. 2023; Nam et al. 2022; Pertsch, Lee, and Lim 2020; Gupta et al. 2019).

Baselines. For comparison, we test **VIMA** (Jiang et al. 2023) that incorporates multi-modal task prompts to a cross-attention-based Transformer for one-shot imitation learning, and we implement TP-**GPT**, a multi-modal task prompting approach, that specifically uses the decoder-only Transformer with self-attention (Shafiullah et al. 2022; Xu et al. 2022). Furthermore, we implement a video generation-based task prompting method **VCP**, similar to (Du et al. 2023; Chane-Sane, Schmid, and Laptev 2023) that incorporates task prompts to video embeddings-conditioned policies. We also implement a language generation-based task prompting method TP-**BCZ** that employs the state-of-the-art multi-modal zero-shot imitation framework (Jang et al. 2022).

Evaluation metrics. We employ two performance evaluation metrics in cross-domain zero-shot policy adaptation: K**-rate** (KR) that evaluates the rate of unseen task domains in which at least K out of N subtasks are completed, similar to the evaluation metric used in long-horizon tasks by (Mees et al. 2022); N**-rate** (NR) that evaluates the average rate of successfully completed subtasks out of N subtasks. In our evaluation, we set $N = 4$ for Franka Kitchen and $N = 3$ for Meta-World.

Evaluation datasets. We use 9 unseen long-horizon evaluation tasks (e.g., each being defined by an unseen order of subtasks) in Franka Kitchen and 8 unseen long-horizon tasks in Meta-World. We also use 12 skill-level domain contexts (e.g., instructed domain contexts for time constraints and time-varying domain contexts for the non-stationarity of environments). Based on the combination of these unseen tasks and contexts, we establish a total of 108 and 96 distinct unseen target domains for Franka Kitchen and Meta-World, respectively. Detailed explanations for training datasets and examples of task prompts can be found in Appendix B.

4.2 Results

Cross-domain performance. We evaluate the cross-domain performance in both K-rate and N-rate, wherein a policy is evaluated in a zero-shot manner for different target domains. In Table 1, SemTra consistently outperforms the baselines, specifically 36.34% and 66.24% N-rate higher than the most competitive baselines VCP in Franka Kitchen and TP-GPT in Meta-World, respectively. With respect to the K-rate metric, the margin of superiority becomes more significant for larger K, underlining the robust adaptation capabilities of SemTra, especially for long-horizon tasks.

Single-domain (reference) performance. To analyze the cross-domain performance, we also use the single-domain scenarios (i.e., one-shot imitation within a single-domain) as a reference, where we treat the task prompt as a demonstration specific to the target domain. As shown in Table 2, SemTra shows 13.89% and 58.33% higher N-rate than the most competitive baseline TP-GPT in Franka Kitchen and Meta-World, respectively. More importantly, we confirm the

Env	KR	VIMA	BCZ	GPT	VCP	SemTra
FK	K=1	63.6	55.1	51.1	62.2	**100.0**
	K=2	31.3	19.5	32.0	32.8	**80.0**
	K=3	13.9	3.1	14.2	16.8	**61.7**
	K=4	5.3	0.0	7.5	7.5	**23.1**
	NR	28.3	19.4	26.2	29.8	**66.2**
MW	K=1	18.5	17.4	35.4	22.7	**86.2**
	K=2	4.8	6.4	3.2	4.8	**82.8**
	K=3	0.0	0.0	0.0	0.5	**72.6**
	NR	7.6	7.8	13.1	8.4	**79.3**

Table 1: Cross-domain performance in K-rate and N-rate.

robust cross-domain performance of SemTra (in Table 1), which is comparable to this single-domain performance; specifically, the degradation is minimal, when transitioning from single-domain to cross-domain experiments. In contrast, the baselines experience a relatively large degradation.

Env	KR	VIMA	BCZ	GPT	VCP	SemTra
FK	K=1	96.2	62.9	92.5	77.7	**100.0**
	K=2	62.9	22.2	74.0	55.5	**85.1**
	K=3	22.2	3.7	40.7	29.6	**66.6**
	K=4	11.1	0.0	19.6	11.1	**40.7**
	NR	48.1	22.2	59.2	43.5	**73.1**
MW	K=1	37.5	29.1	33.3	50.0	**100.0**
	K=2	12.5	16.6	8.3	12.5	**79.1**
	K=3	3.7	1.5	0.0	0.0	**62.5**
	NR	16.6	15.2	22.2	20.8	**80.5**

Table 2: Single-domain performance in K-rate and N-rate.

In the following, we use this single-domain performance in N-rate as a reference in comparison.

Task-level cross-domain performance. We specifically investigate the cross-domain adaptation performance under task-level domain changes, the scenarios where the order of subtasks in the target domain is not the same as that of a given task prompt. For the cross-domain setting, we consider task prompts involving snippets beneficial for the adaptation, such as "in reverse order". With no consideration of skill-level cross-domain contexts, we test 27 and 24 unseen task prompts for Franka Kitchen and Meta-World, respectively. Table 3 shows that SemTra exhibits robust performance for the "Reverse" and "Replace" cases with N-rate decline of 7.08% and 5.97% in Franka Kitchen and Meta-world, respectively, compared to the single-domain reference performance ("Ref."). In contrast, TP-GPT degrades by 9.39% and 2.5% in N-rate. This specifies the task-level adaptation capabilities of SemTra, specifically empowered by the semantic skill decoder based on a PLM.

Skill-level cross-domain performance. Table 4 shows the cross-domain adaptation performance, particularly for skill-level domain changes, in which each skill or subtask is required to adapt to the target domain. For each instructed domain context, we compute the average N-rate performance across different non-stationary hidden contexts. For example, "Fast" in the "Domain" column indicates domain

Env	Dom	VIMA	BCZ	GPT	VCP	SemTra
FK	Ref.	48.1	22.2	59.2	43.5	**73.1**
	Rev	46.3	14.8	50.9	62.9	**71.3**
	Rep	30.9	15.4	48.8	26.1	**64.2**
MW	Ref.	16.6	15.2	22.2	20.8	**80.5**
	Rev	5.5	9.7	8.3	8.3	**76.3**
	Rep	8.8	8.8	13.3	5.5	**80.0**

Table 3: Task-level cross-domain performance in N-rate.

contexts associated with stringent time limits in the environment, which can be instructed (specified) by a given task prompt; the snippets beneficial for the skill-level cross-domain context, such as "do fast", are given. For each row in the table, we experiment with 36 and 32 different task prompts for Franka Kitchen and Meta-World, respectively. SemTra exhibits robust performance compared to the baselines, showing 39.35% and 65.73% higher in Franka Kitchen and Meta-World than the most competitive baselines TP-BCZ and TP-GPT, respectively. This is attributed to our disentanglement scheme, which effectively separates domain-invariant semantic skills from domain contexts.

Env	Dom	VIMA	BCZ	GPT	VCP	SemTra
FK	Ref.	48.1	22.2	59.2	43.5	**73.1**
	Gust	13.6	23.1	12.7	15.7	**59.9**
	Flurry	7.8	19.4	4.4	16.6	**61.3**
MW	Ref.	16.6	15.2	22.2	20.8	**80.5**
	Slow	1.0	12.5	9.3	4.1	**72.1**
	Fast	3.4	0.0	11.4	15.2	**80.2**

Table 4: Skill-level cross-domain performance in N-rate.

Semantic correspondence in multi-modal space. In Figure 3, we visualize a set of video demonstrations and the trajectories of the zero-shot policy in SemTra. We use the V-CLIP embedding space. A numeric index in the legends corresponds to a semantic skill. As observed, the demonstration trajectories (on the left side of the figure) traverse through distinct clusters of semantic skills, reflecting the alignment of the skills with the expert behavior patterns. The policy's trajectories (in the middle of the figure) in the target domains share a similar space with the demonstrations. In addition, the trajectories from different domains ("Dom"s in legends) for a single semantic skill (on the right side of the figure) tend to form distinct clusters, indicating the variations in the execution of a specific skill in different domain contexts.

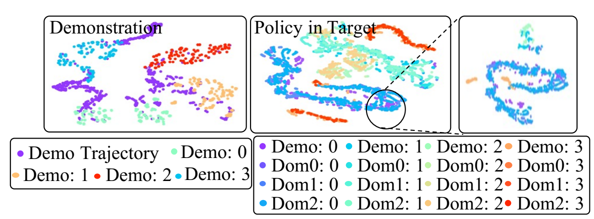
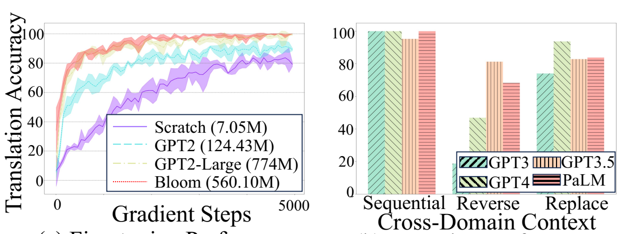


Figure 3: Semantic correspondence in V-CLIP space.

4.3 Ablation Studies

PLMs for skill sequence generation. To implement the semantic skill sequence generator Φ_G in (6), we test several PLMs, including relatively lightweight models such as GPT2, GPT2-large (Radford et al. 2019), and Bloom (Scao et al. 2022). These models are selected based on the consideration that they can be fine-tuned with the GPU system resources available in an academic setting. In Figure 4, these PLMs consistently exhibit higher sample efficiency than the model trained from scratch. There is also a positive correlation between the size of the PLMs and the accuracy of skill sequence generation. This specifies that the logical reasoning abilities of PLMs are instrumental in task adaptation.

We also evaluate the suitability of advanced PLMs such as GPT3 (Brown et al. 2020), GPT3.5, GPT4 (OpenAI 2023), and PaLM (Chowdhery et al. 2022) for the skill sequence generator, in a *zero-shot* context, without any fine-tuning. To this end, we engineer an augmented task prompt to contain a semantic skill set, which is presented to the PLMs along with the skill-level instruction. Our observations reveal that GPT3 and GPT3.5 encounter difficulties in generating the correct sequence, especially when reversing the execution of a skill sequence, while GPT4 and PaLM achieve relatively better performance. These results are consistent with our motivation to leverage the advancements in PLMs in enabling robust translation from intermediate instructions to semantic skill sequences.



(a) Fine-tuning Performance (b) Zero-shot Performance

Figure 4: PLMs for skill sequence generator: (a) PLM fine-tuning case; the x-axis denotes the gradient update steps, and the y-axis denotes the accuracy of the skill sequence generation. (b) PLM zero-shot case; only an engineered prompt is adopted without any model fine-tuning.

Conditional components for behavior decoding. We test several representations for the conditional components to the behavior decoder in (10) to evaluate our skill adaptation phase. Specifically, we explore variations in the inputs to the behavior decoder instead of executable skills $\bar{z}_t$ in (10): **N-GPT** uses a given task prompt in multi-modal snippets in (2); **L-GPT** uses a skill-level instruction in (3) that has been encoded in the language embedding space by the multi-modal encoders; **V-GPT** is the same as L-GPT but employs the video embedding space; **T-GPT** takes the entire sequence of executable skills in (9) bypassing the sequence segmentation in (7).

In Table 5, comparing N-GPT and {V, L}-GPT, we observe that retrieving semantics from video demonstrations improves performance in N-rate by 16.45%. L-GPT outperforms V-GPT, specifying that skill embeddings in the language space are more efficient than those in the video space; this pattern was observed in previous work (Jang et al. 2022). T-GPT outperforms L-GPT by 12.54%, indicating the effectiveness of the skill sequence generation of our framework. Finally, SemTra outperforms T-GPT by 35.57%, demonstrating that a behavior decoder conditioned with a single skill is suitable. These consistent results clarify the benefits of our skill adaptation based on disentangled skill semantics and target domain contexts, enabling parametric skill instantiation.

Env	N-GPT	V-GPT	L-GPT	T-GPT	SemTra
FK	4.1	19.4	40.2	42.8	**72.2**
MW	2.7	8.1	11.8	34.3	**76.1**

Table 5: Conditional components for behavior decoding.

Sample efficiency in semantic annotations. As explained in Section 3.2, we leverage a training dataset with semantically annotated target trajectories. Recognizing the challenges associated with acquiring a large annotated dataset, we employ the video skill encoder, which was trained on a small annotated dataset, to pseudo-label on unannotated demonstrations. We evaluate the impact of the annotated dataset size on the efficacy of our framework. Table 6 shows the accuracy of the video skill encoder (in the row of Skill Pred.) and the cross-domain adaptation performance (N-rate) with respect to the number of annotations used. Remarkably, utilizing only 50 videos per semantic skill (approximately 0.23% of the entire dataset) is sufficient for our framework to achieve comparable performance to the cases where annotations are provided at every timestep.

# Ann.	3	5	7	10	50	Every
Skill Pred.	68.5	68.5	84.4	89.6	99.7	100.0
N-rate	45.8	46.6	64.7	64.4	71.1	72.7

Table 6: Sample efficiency in semantic annotations

4.4 Use Cases

Cognitive robot manipulator. We verify the applicability of SemTra with cognitive robot manipulation scenarios using the RLBench robotic manipulation simulator (James et al. 2020), in which a task prompt is presented as *abstract user commands* in natural language (contrast with our task prompts containing video demonstrations.) Specifically, we experiment with three different command styles, similar to the applications of SayCan (Ichter et al. 2022). A **Summarized sentence** command offers an abstract instruction for conveying long-horizon tasks. An **Abstract verbs** command encompasses synonymous verbs to convey desired actions. An **Embodiment** command narrates the current environment conditions and the constraints of the robot manipulator. Table 7 demonstrates the efficacy of SemTra in processing these abstract commands, specifying the comprehensive task prompting structure as well as the ability to harness the logical reasoning capabilities of PLMs. Furthermore, the two-phase adaptation facilitates modular approaches for different levels of policy adaptations using PLMs. It also supports the flexible integration of future, more advanced PLMs.

	SayCan		SemTra		
	GPT-3	PaLM	GPT-3.5	GPT-4	PaLM
Summ.	76.6	100.0	53.3	100.0	100.0
Abst.	60.0	100.0	73.3	100.0	100.0
Emb.	86.6	100.0	60.0	100.0	100.0

Table 7: Performance in cognitive robot manipulations.

Autonomous driving. With the autonomous driving simulator CARLA (Dosovitskiy et al. 2017), we verify the skill adaptation ability across different vehicle configurations (agent embodiments) such as sedans and trucks. The driving agent must calibrate steering, throttle, and brake responses according to the specific vehicle configuration it manages. Its goal is to reach the destination in the shortest path as depicted in Figure 5, where skills are semantically defined as driving straight, turning left, and turning right. We calculate the normalized return based on the rewards yielded by a rule-based policy. As shown in Table 8, SemTra outperforms the most competitive baseline VCP by $1.94 \sim 5.00\%$.

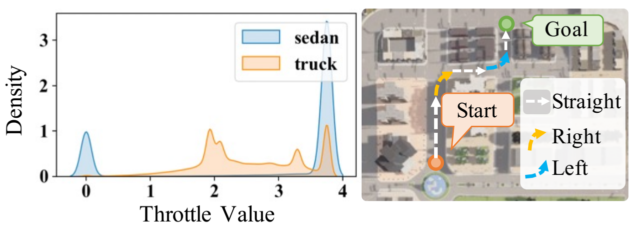


Figure 5: The left graph represents the difference in expert action distribution based on vehicle configurations. The right figure shows a semantic skill sequence to reach the goal.

Vehicle type	VIMA	BCZ	GPT	VCP	SemTra
Sedan	79.4	74.7	62.7	78.3	**83.3**
Truck	33.4	68.6	65.1	70.5	**72.4**

Table 8: Performance in autonomous driving.

5 Related Works

Cross-domain policy adaptation. Recently, there has been a surge in cross-domain policy adaptation techniques exploring task specifications (Goyal, Mooney, and Niekum 2023; Pertsch et al. 2022). A relevant subset of this trend is cross-domain one-shot imitation, which exploits demonstrations from the source domain to tackle tasks in different target domains (Zhao et al. 2022; Kim et al. 2020; Liu et al. 2018; Yu et al. 2018; Gupta et al. 2017). For instance, (Zhao et al. 2022) introduced a contrastive learning-based video encoder to bridge visual domain gaps between demonstrations and target environments. (Kim et al. 2020) employed the generative adversarial learning to handle embodiment and dynamics differences, while (Goyal, Mooney, and Niekum 2023) delved into task-level domain disparities, directly transferring source expert trajectories to target actions using a Transformer-based policy. Our work is in the same vein, but it distinguishes itself by accommodating a unified multi-modal specification for task prompting. We also concentrate on long-horizon tasks, incorporating the knowledge of PLMs into our framework to explore the compositionality and parametric execution capability of semantic skills.

Semantic skills in RL. In the RL literature, several studies introduced policy learning and adaptation methods, particularly investigating semantically interpretable skills (semantic skills). Recently, Saycan (Ichter et al. 2022) employed a PLM to plan semantic skills from abstract linguistic instructions. Utilizing an offline RL dataset collected from an environment with domain-specific rewards, this novel approach learns to estimate the value of each semantic skill at a specific scene, effectively opting to execute the most valuable skill. Furthermore, (Pertsch et al. 2022) exploited human demonstrations with semantic annotations via state distribution matching to expedite RL training. However, our emphasis lies on their ability to adapt in a zero-shot manner across domains by developing a novel two-phase adaptation process.

Multi-modality in robotic manipulation. In recent studies on robotic manipulation, several approaches leveraging large-scale pretrained multi-modal models have been introduced, showing promise in enhanced policy generalization by adeptly integrating both visual and language instructions (Guhur et al. 2023; Shridhar, Manuelli, and Fox 2022; Liu et al. 2022). (Jang et al. 2022) proposed a multi-modal imitation learning framework, which trains the vision encoder by maximizing the similarity of video embeddings and the PLMs embedding vectors of video descriptions. (Jiang et al. 2023) proposed a unified robot-manipulation framework capable of integrating interleaved video-language inputs by encoding them with pretrained vision and language models. To the best of our knowledge, SemTra is the first to consider a unified multi-modal task prompt and incorporate the knowledge of PLMs and parametric semantic skills into cross-domain policy adaptation procedures.

6 Conclusion

We presented the semantic skill translator SemTra, a framework that leverages the knowledge of PLMs to achieve zero-shot policy adaptation across different domains. In the framework, we employ a two-tiered adaptation process in which a sequence of multi-modal interleaved snippets serves as a task prompt, facilitating task-level semantic skill generation and skill-level refinement tailored to the target domain contexts. In our future work, we will explore the versatility of PLMs more under intricate cross-domain conditions that require dynamic adjustments of a skill sequence based on real-time observations.

Acknowledgements

We would like to thank anonymous reviewers for their valuable comments and suggestions. This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-01045, 2022-0-00043, 2019-0-00421, 2020-0-01821) and by the National Research Foundation of Korea (NRF) grant funded by the MSIT (No. RS-2023-00213118).

References

Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Chalaki, B.; Beaver, L.; Remer, B.; Jang, K.; Vinitsky, E.; Bayen, A. M.; and Malikopoulos, A. A. 2019. Zero-Shot Autonomous Vehicle Policy Transfer: From Simulation to Real-World via Adversarial Learning.

Chane-Sane, E.; Schmid, C.; and Laptev, I. 2023. Learning Video-Conditioned Policies for Unseen Manipulation Tasks. *CoRR*, abs/2305.06289.

Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Sutton, C.; Gehrmann, S.; Schuh, P.; Shi, K.; Tsvyashchenko, S.; Maynez, J.; Rao, A.; Barnes, P.; Tay, Y.; Shazeer, N.; Prabhakaran, V.; Reif, E.; Du, N.; Hutchinson, B.; Pope, R.; Bradbury, J.; Austin, J.; Isard, M.; Gur-Ari, G.; Yin, P.; Duke, T.; Levskaya, A.; Ghemawat, S.; Dev, S.; Michalewski, H.; Garcia, X.; Misra, V.; Robinson, K.; Fedus, L.; Zhou, D.; Ippolito, D.; Luan, D.; Lim, H.; Zoph, B.; Spiridonov, A.; Sepassi, R.; Dohan, D.; Agrawal, S.; Omernick, M.; Dai, A. M.; Pillai, T. S.; Pellat, M.; Lewkowycz, A.; Moreira, E.; Child, R.; Polozov, O.; Lee, K.; Zhou, Z.; Wang, X.; Saeta, B.; Diaz, M.; Firat, O.; Catasta, M.; Wei, J.; Meier-Hellstern, K.; Eck, D.; Dean, J.; Petrov, S.; and Fiedel, N. 2022. PaLM: Scaling Language Modeling with Pathways. *CoRR*, abs/2204.02311.

Dosovitskiy, A.; Ros, G.; Codevilla, F.; López, A. M.; and Koltun, V. 2017. CARLA: An Open Urban Driving Simulator. In *1st Annual Conference on Robot Learning, CoRL 2017, Mountain View, California, USA, November 13-15, 2017, Proceedings*, 1–16.

Du, Y.; Yang, M.; Dai, B.; Dai, H.; Nachum, O.; Tenenbaum, J. B.; Schuurmans, D.; and Abbeel, P. 2023. Learning Universal Policies via Text-Guided Video Generation. *CoRR*, abs/2302.00111.

Fu, J.; Kumar, A.; Nachum, O.; Tucker, G.; and Levine, S. 2020. D4RL: Datasets for Deep Data-Driven Reinforcement Learning. *CoRR*, abs/2004.07219.

Goyal, P.; Mooney, R. J.; and Niekum, S. 2023. Language-guided Task Adaptation for Imitation Learning. *CoRR*.

Guhur, P.-L.; Chen, S.; Pinel, R. G.; Tapaswi, M.; Laptev, I.; and Schmid, C. 2023. Instruction-driven history-aware policies for robotic manipulations. In *Conference on Robot Learning*, 175–187.

Gupta, A.; Devin, C.; Liu, Y.; Abbeel, P.; and Levine, S. 2017. Learning Invariant Feature Spaces to Transfer Skills with Reinforcement Learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Gupta, A.; Kumar, V.; Lynch, C.; Levine, S.; and Hausman, K. 2019. Relay Policy Learning: Solving Long-Horizon Tasks via Imitation and Reinforcement Learning. In *3rd Annual Conference on Robot Learning, CoRL 2019, Osaka, Japan, October 30 - November 1, 2019, Proceedings*, volume 100 of *Proceedings of Machine Learning Research*, 1025–1037.

Harrison, J.; Garg, A.; Ivanovic, B.; Zhu, Y.; Savarese, S.; Fei-Fei, L.; and Pavone, M. 2017. AdaPT: Zero-Shot Adaptive Policy Transfer for Stochastic Dynamical Systems. In *Robotics Research, The 18th International Symposium, ISRR 2017, Puerto Varas, Chile, December 11-14, 2017*, 437–453.

Huang, W.; Xia, F.; Xiao, T.; Chan, H.; Liang, J.; Florence, P.; Zeng, A.; Tompson, J.; Mordatch, I.; Chebotar, Y.; Sermanet, P.; Jackson, T.; Brown, N.; Luu, L.; Levine, S.; Hausman, K.; and Ichter, B. 2022. Inner Monologue: Embodied Reasoning through Planning with Language Models. In *Conference on Robot Learning, CoRL 2022, 14-18 December 2022, Auckland, New Zealand*, Proceedings of Machine Learning Research, 1769–1782. PMLR.

Ichter, B.; Brohan, A.; Chebotar, Y.; Finn, C.; Hausman, K.; Herzog, A.; Ho, D.; Ibarz, J.; Irpan, A.; Jang, E.; Julian, R.; Kalashnikov, D.; Levine, S.; Lu, Y.; Parada, C.; Rao, K.; Sermanet, P.; Toshev, A.; Vanhoucke, V.; Xia, F.; Xiao, T.; Xu, P.; Yan, M.; Brown, N.; Ahn, M.; Cortes, O.; Sievers, N.; Tan, C.; Xu, S.; Reyes, D.; Rettinghouse, J.; Quiambao, J.; Pastor, P.; Luu, L.; Lee, K.; Kuang, Y.; Jesmonth, S.; Joshi, N. J.; Jeffrey, K.; Ruano, R. J.; Hsu, J.; Gopalakrishnan, K.; David, B.; Zeng, A.; and Fu, C. K. 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. In *Conference on Robot Learning, CoRL 2022, 14-18 December 2022, Auckland, New Zealand*, Proceedings of Machine Learning Research, 287–318.

James, S.; Ma, Z.; Arrojo, D. R.; and Davison, A. J. 2020. RLBench: The Robot Learning Benchmark & Learning Environment. *IEEE Robotics Autom. Lett.*, 3019–3026.

Jang, E.; Irpan, A.; Khansari, M.; Kappler, D.; Ebert, F.; Lynch, C.; Levine, S.; and Finn, C. 2022. BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning. *CoRR*.

Jiang, Y.; Gupta, A.; Zhang, Z.; Wang, G.; Dou, Y.; Chen, Y.; Fei-Fei, L.; Anandkumar, A.; Zhu, Y.; and Fan, L. 2023. VIMA: General Robot Manipulation with Multimodal Prompts. arXiv:2210.03094.

Kim, K.; Gu, Y.; Song, J.; Zhao, S.; and Ermon, S. 2020. Domain Adaptive Imitation Learning. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, Proceedings of Machine Learning Research, 5286–5295.

Liu, H.; Lee, L.; Lee, K.; and Abbeel, P. 2022. Instruction-Following Agents with Jointly Pre-Trained Vision-Language Models. *arXiv preprint arXiv:2210.13431*.

Liu, Y.; Gupta, A.; Abbeel, P.; and Levine, S. 2018. Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation. In *2018 IEEE International Conference on Robotics and Automation, ICRA 2018, Brisbane, Australia, May 21-25, 2018*, 1118–1125.

Mees, O.; Hermann, L.; Rosete-Beas, E.; and Burgard, W. 2022. CALVIN: A Benchmark for Language-Conditioned Policy Learning for Long-Horizon Robot Manipulation Tasks. *IEEE Robotics Autom. Lett.*, 7.

Nam, T.; Sun, S.; Pertsch, K.; Hwang, S. J.; and Lim, J. J. 2022. Skill-based Meta-Reinforcement Learning. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*.

OpenAI. 2023. GPT-4 Technical Report. *CoRR*, abs/2303.08774.

Pertsch, K.; Desai, R.; Kumar, V.; Meier, F.; Lim, J. J.; Batra, D.; and Rai, A. 2022. Cross-Domain Transfer via Semantic Skill Imitation. In *Conference on Robot Learning, CoRL 2022, 14-18 December 2022, Auckland, New Zealand*, 690–700.

Pertsch, K.; Lee, Y.; and Lim, J. J. 2020. Accelerating Reinforcement Learning with Learned Skill Priors. In *4th Conference on Robot Learning, CoRL 2020, 16-18 November 2020, Virtual Event / Cambridge, MA, USA*, volume 155 of *Proceedings of Machine Learning Research*, 188–204.

Qiu, Y.; Wu, J.; Cao, Z.; and Long, M. 2022. Out-of-Dynamics Imitation Learning from Multimodal Demonstrations. In *Conference on Robot Learning, CoRL 2022, 14-18 December 2022, Auckland, New Zealand*, volume 205 of *Proceedings of Machine Learning Research*, 1071–1080.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I.; et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8): 9.

Rakelly, K.; Zhou, A.; Finn, C.; Levine, S.; and Quillen, D. 2019. Efficient Off-Policy Meta-Reinforcement Learning via Probabilistic Context Variables. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, 5331–5340.

Scao, T. L.; Fan, A.; Akiki, C.; Pavlick, E.; Ilic, S.; Hesslow, D.; Castagné, R.; Luccioni, A. S.; Yvon, F.; Gallé, M.; Tow, J.; Rush, A. M.; Biderman, S.; Webson, A.; Ammanamanchi, P. S.; Wang, T.; Sagot, B.; Muennighoff, N.; del Moral, A. V.; Ruwase, O.; Bawden, R.; Bekman, S.; McMillan-Major, A.; Beltagy, I.; Nguyen, H.; Saulnier, L.; Tan, S.; Suarez, P. O.; Sanh, V.; Laurençon, H.; Jernite, Y.; Launay, J.; Mitchell, M.; Raffel, C.; Gokaslan, A.; Simhi, A.; Soroa, A.; Aji, A. F.; Alfassy, A.; Rogers, A.; Nitzav, A. K.; Xu, C.; Mou, C.; Emezue, C.; Klamm, C.; Leong, C.; van Strien, D.; Adelani, D. I.; and et al. 2022. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *CoRR*.

Shafiullah, N. M. M.; Cui, Z. J.; Altanzaya, A.; and Pinto, L. 2022. Behavior Transformers: Cloning k modes with one stone. *CoRR*, abs/2206.11251.

Shin, S.; Lee, D.; Yoo, M.; Kim, W. K.; and Woo, H. 2023. One-shot Imitation in a Non-Stationary Environment via Multi-Modal Skill.

Shridhar, M.; Manuelli, L.; and Fox, D. 2022. Cliport: What and where pathways for robotic manipulation. In *Conference on Robot Learning*, 894–906. PMLR.

van den Oord, A.; Li, Y.; and Vinyals, O. 2018. Representation Learning with Contrastive Predictive Coding. *CoRR*.

Xu, H.; Ghosh, G.; Huang, P.; Okhonko, D.; Aghajanyan, A.; Metze, F.; Zettlemoyer, L.; and Feichtenhofer, C. 2021. VideoCLIP: Contrastive Pre-training for Zero-shot Video-Text Understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, 6787–6800.

Xu, M.; Shen, Y.; Zhang, S.; Lu, Y.; Zhao, D.; Tenenbaum, J. B.; and Gan, C. 2022. Prompting Decision Transformer for Few-Shot Policy Generalization. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, Proceedings of Machine Learning Research, 24631–24645.

Yu, T.; Finn, C.; Xie, A.; Dasari, S.; Zhang, T.; Abbeel, P.; and Levine, S. 2018. One-Shot Imitation from Observing Humans via Domain-Adaptive Meta-Learning. *CoRR*, abs/1802.01557.

Yu, T.; Quillen, D.; He, Z.; Julian, R.; Hausman, K.; Finn, C.; and Levine, S. 2019. Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning. In *3rd Annual Conference on Robot Learning, CoRL 2019, Osaka, Japan, October 30 - November 1, 2019, Proceedings*, 1094–1100.

Zhao, M.; Liu, F.; Lee, K.; and Abbeel, P. 2022. Towards More Generalizable One-shot Visual Imitation Learning. In *2022 International Conference on Robotics and Automation, ICRA 2022, Philadelphia, PA, USA, May 23-27, 2022*, 2434–2444. IEEE.

Zintgraf, L. M.; Shiarlis, K.; Igl, M.; Schulze, S.; Gal, Y.; Hofmann, K.; and Whiteson, S. 2020. VariBAD: A Very Good Method for Bayes-Adaptive Deep RL via Meta-Learning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*.