A Masked World Models

For our experiments, we use Masked World Models (MWM) (Seo et al., 2022) as our underlying RL algorithm. MWM is a model-based RL framework that aims to decouple visual representation learning and dynamics learning. In contrast to prior algorithms that learn world models in an end-to-end manner, MWM separately learns a visual encoder via masked autoencoding (He et al., 2021) and a world model that reconstructs frozen autoencoder representations. We build LAMP upon the official implementation provided by authors (https://github.com/younggyoseo/MV-MWM) which supports experimentation on RLBench (James et al., 2020). In Table 7, 8, and 9 we provide all relevant hyperparameters.

B Video-Language Models

R3M For the R3M (Nair et al., 2022) VLM in LAMP, we use the official implementation (https://github.com/facebookresearch/r3m). We use the ResNet-18 visual backbone and preserve all default hyperparameters. To encode the language prompts, we use the DistilBERT 'base-uncased' model (https://huggingface.co/distilbert-base-uncased) from the transformers (https://pypi.org/project/transformers/) package as used in the R3M implementation.

InternVideo For the InternVideo (Wang et al., 2022) VLM in LAMP we use the official implementation (https://github.com/OpenGVLab/InternVideo). We use the "B/16" model and pretrained weights provided by the authors for embedding the images and text.

For InternVideo (Wang et al., 2022) we match the style of alignment score computation used in training and inference. We use the following reward parameterization:

$$r_i = F_\phi([s_{i//8}, s_{2*i//8}, ..., s_{8*i//8}]) \cdot L_\alpha(l). \tag{4}$$

where 8 frames evenly spaced from the agent's entire history are featurized by the visual encoder F_ϕ.

ZeST For the CLIP (Radford et al., 2021) VLM used in ZeST (Cui et al., 2022) for LAMP, we use the official OpenAI CLIP model (https://github.com/openai/CLIP). We use the "ViT-B/32" release of the CLIP visual encoder for embedding images and the CLIP text encoder for embedding langauge prompts.

C Experimental Details

C.1 Language Prompting Types

We ablate on 6 different prompt styles, with the structures defined in Section 6.1. To construct Prompt Style 1, we replace the `[NOUN]` in the `Pick up the [NOUN].` prompt with the ShapeNet name. To construct Prompt Style 6, we sample from random Shakespeare phrases listed below. For the remaining Prompt Styles, we sample a verb structure, either `IRRELEVANT` or `RELEVANT`; and a noun, either `RANDOM` or `SYNONYM`. Both verb structures are included in Table 2. We include examples of random nouns and synonym nouns for a sample object; nouns and synonyms for all objects are ommitted for space, but they can be found in the code.

Verb structures and synonyms were curated from ChatGPT, with filtering afterward to select most relevant verbs. Random nouns were taken from the synonyms. We provide the automatically generated language prompt datasets for each language prompting scheme used during LAMP pretraining. Prompts to ChatGPT are included in Table 1.

C.2 ShapeNet Objects

In Figure 9 we provide images of some of the ShapeNet object assets used in the pretraining environments.

Table 1: **Using an LLM to generate verb structures and nouns**. We query ChatGPT with prompts to create a set of diverse tasks.

Generating Verb Structures	**Generating Synonyms**
Give me a list of 40 task variations that present an interesting task for a person to do in a home or kitchen scenario. Examples should not be complicated, and should be possible to do very quickly, within a minute or so. These should be simple tasks that are interesting and diverse, but EASY. Tasks should be atomic and very general. For example: 1. Reach for the mug 2. Open the microwave 3. Wipe the table clean 4. Water the flowers	Please give me 40 synonyms for bowl Example: 1. bowl 2. soup bowl 3. dish

Figure 9: Examples of ShapeNet object assets used during the pretraining phase.

Table 2: **Verb Structures**. We include the different verb structures used during pretraining.

Relevant Verb	**Irrelevant Verb**
Pick up the [NOUN]	The [NOUN] is seized
Lift the [NOUN] with your hands	The [NOUN] is clutched
Hold the [NOUN] in your grasp	The [NOUN] is gripped
Take hold of the [NOUN] and raise it	The [NOUN] is firmly grasped
Grasp the [NOUN] firmly and lift it up	The [NOUN] is tightly held
Raise the [NOUN] by picking it up	The [NOUN] is firmly caught
Retrieve the [NOUN] and hold it up	The [NOUN] is securely clasped
Lift the [NOUN] by gripping it	The [NOUN] is rotated
Seize the [NOUN] and raise it off the surface	The [NOUN] has been flipped
Hold onto the [NOUN] and lift it up	The [NOUN] has been knotted
The [NOUN] is lifted up	The [NOUN] has been folded
The [NOUN] is picked up off the ground	The [NOUN] has been rinsed
The [NOUN] is raised up by hand	The [NOUN] has been filled
The [NOUN] is grasped and lifted up	The [NOUN] is shaken
The [NOUN] is taken up by hand	The [NOUN] has been scooped
The [NOUN] is retrieved and lifted up	The [NOUN] is poured
The [NOUN] is lifted off its surface	The [NOUN] has been scrubbed
The [NOUN] is elevated by being picked up	The [NOUN] is tilted
The [NOUN] is hoisted up by hand	The [NOUN] has been heated
The [NOUN] is scooped up and lifted	Reach for the [NOUN]
The [NOUN] is lifted by the hand	Grasp at the [NOUN]
The [NOUN] is grasped and picked up	Stretch out to touch the [NOUN]
The [NOUN] is raised by the palm	Move your arm towards the [NOUN]
The [NOUN] is taken up by the fingers	Use the gripper to rinse the [NOUN]
The [NOUN] is held and lifted up	Position the end effector to fold the [NOUN]
The [NOUN] is lifted off the surface by the arm	Reach out the robotic arm to wipe the [NOUN]
The [NOUN] is picked up and held by the wrist	Utilize the gripper to seize the [NOUN]
The [NOUN] is scooped up by the palm and lifted	Guide the robotic arm to obtain the [NOUN]
The [NOUN] is elevated by the hand	Maneuver the end effector to lift up the [NOUN]
The [NOUN] is taken up by the fingers of the hand	Extend your hand towards the [NOUN]
The [NOUN] is grasped and raised	Reach out your hand to acquire the [NOUN]
The [NOUN] is lifted by the gripper	Guide your arm to rotate the [NOUN]
The end effector picks up the [NOUN]	Maneuver your hand to shake up the [NOUN]
The arm lifts the [NOUN]	Flip the [NOUN]
The [NOUN] is held aloft by the robotic hand	Tap the [NOUN]
The robotic gripper secures the [NOUN]	Fold the [NOUN]
The [NOUN] is lifted off the surface by the robotic arm	Rotate the [NOUN]
The robotic manipulator seizes and elevates the [NOUN]	Brush the [NOUN]
The robotic end effector clasps and hoists the [NOUN]	Twist the [NOUN]
The [NOUN] is taken up by the robotic gripper	Wipe the [NOUN]

Table 3: **Prompt Style 6**. Snippets from Shakespeare.

Holla, Barnardo.
BARNARDO Say, what, is Horatio there?
HORATIO A piece of him.
Welcome, Horatio.—Welcome, good Marcellus.
HORATIO
What, has this thing appeared again tonight?
BARNARDO I have seen nothing.
MARCELLUS
Horatio says ’tis but our fantasy
And will not let belief take hold of him
Touching this dreaded sight twice seen of us.
Therefore I have entreated him along
With us to watch the minutes of this night,
That, if again this apparition come,
He may approve our eyes and speak to it.
Tush, tush, ’twill not appear.
Sit down awhile,
How now, Horatio, you tremble and look pale.
Is not this something more than fantasy?
What think you on ’t?
At least the whisper goes so: our last king,
Whose image even but now appeared to us,
Was, as you know, by Fortinbras of Norway,
Thereto pricked on by a most emulate pride,
Dared to the combat; in which our valiant Hamlet
(For so this side of our known world esteemed him)
Did slay this Fortinbras, who by a sealed compact,
Well ratified by law and heraldry,
Did forfeit, with his life, all those his lands
Which he stood seized of, to the conqueror.
Against the which a moiety competent
Was gagèd by our king, which had returned
To the inheritance of Fortinbras
Had he been vanquisher, as, by the same comart
And carriage of the article designed,
His fell to Hamlet. Now, sir, young Fortinbras,
Of unimprovèd mettle hot and full,
Hath in the skirts of Norway here and there
Sharked up a list of lawless resolutes
BARNARDO

Table 4: **Nouns**. Synonym and random nouns for "bag."

Synonym Noun	Random Noun
bag	cap
handbag	hat
purse	snapback
clutch	faucet
tote	tap
backpack	vase
knapsack	flask
satchel	earphone
shoulder bag	earpiece
duffel bag	knife
messenger bag	blade
grip	laptop
briefcase	notebook
pouch	vase
fanny pack	flowerpot
drawstring	telephone
beach bag	flip phone
grocery shop	handle
shopping bag	lever
gift bag	gift bag
lunch bag	lunch bag
laptop bag	laptop bag
travel bag	travel bag

C.3 EGO4D TEXTURES

In Figure 10 we provide the textures extracted from Ego4D we overlayed on the pretraining environments.

C.4 COMPUTE

We used a NVIDIA DGX A100 GPUs for all experimentation. Pretraining time for 300k steps takes 60-72 hours; finetuning for 100k takes 16-19 hours.

D ADDITIONAL EXPERIMENTS AND ABLATIONS

D.1 FINETUNING RESULTS WITH INSTRUCTION TUNING

We include additional finetuning results for the "Phone On Base" RLBench task in Figure 11. We find that by tuning the language instruction used to condition the model, we further increase performance. We provide results involving selecting a frozen language instruction from a set of many semantically similar generated instructions for finetuning. We consider 9 of these new generated prompts in addition to the original task name based prompt. We then select the language prompt corresponding to the policy that achieves the highest zero-shot evaluation return when rolled out. This simple tuning step provides us with increases in performance on the new “Phone on Base” task. The optimal instructions tuned for each of the three seeds include:

1. The arm is picking up the phone and placing it on the base
2. The robot arm grasps the phone and sets it down
3. The robot gripper picks up the phone and places it on the base

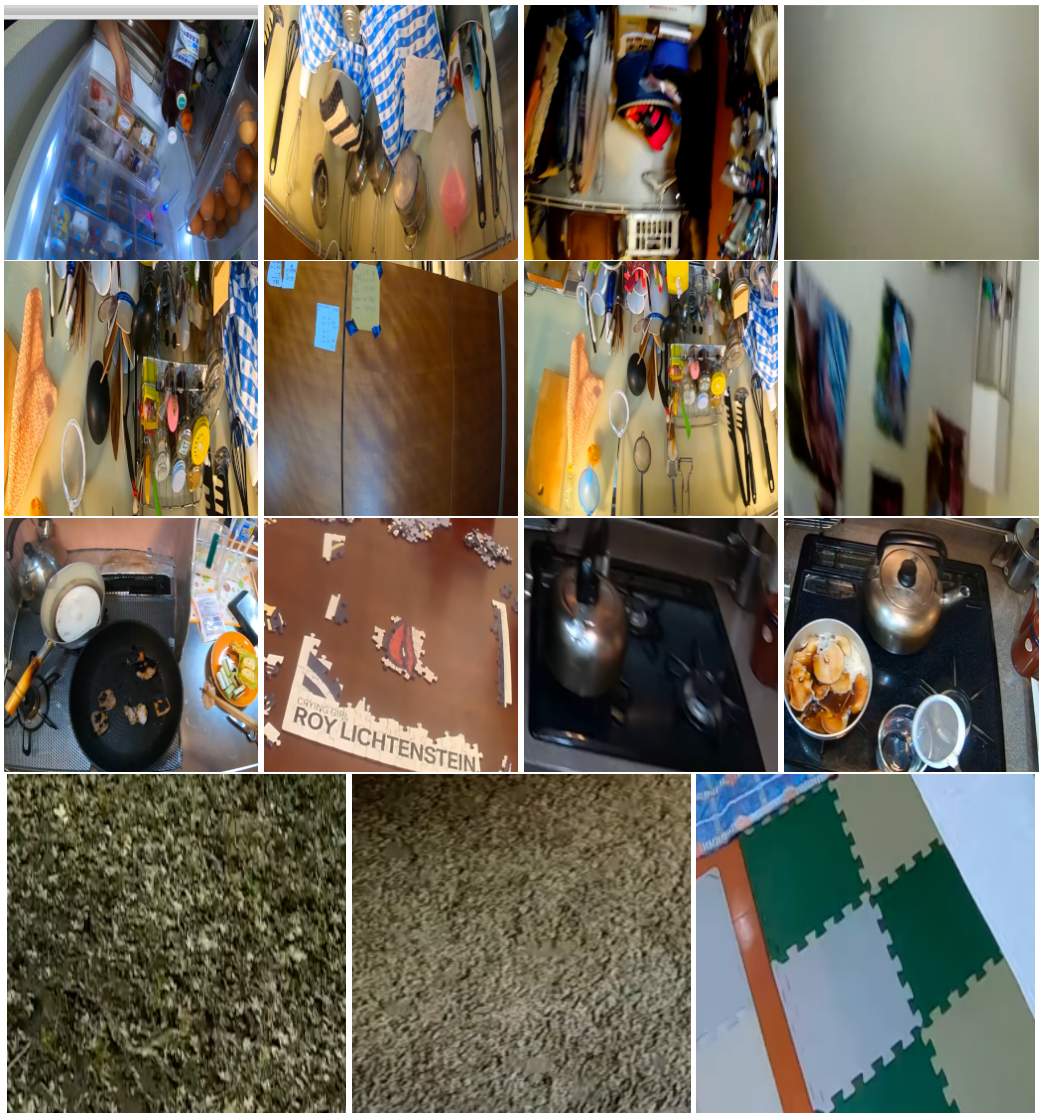

Figure 10: Textures cropped out of Ego4D videos and overlayed to the RLBench scene. The textures on the first two rows were overlayed to the walls, those on the third row to the table, and those on the fourth row to the floor.

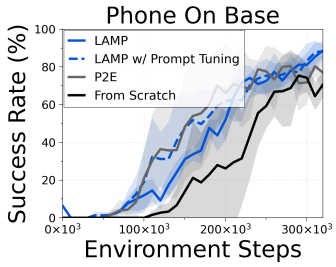

Figure 11: Finetuning Result on Phone On Base

D.2 Pretraining Performance

In our pretraining environment, we spawn ShapeNet objects and plot evaluation returns in Figure 12. The evaluation return is based on a shaped reward for reaching the object and grasping it. While the pretraining reward is noisy, it provides insight into the types of behaviors learned, and in particular, if exploration leads to high-reward behaviors. RND is the worst performing during pretraining, and LAMP learns higher-reward behaviors through the course of training.

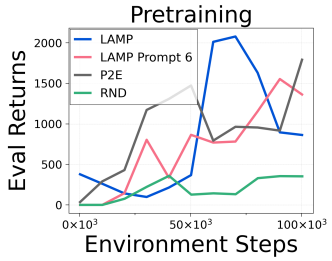

Figure 12: Pretraining Returns with different methods

D.2.1 Random Network Distillation

We experiment with Random Network Distillation (Burda et al., 2018), an additional unsupervised reinforcement learning algorithm, as an additional baseline and report results in 13. We find that for the "Pick Up Cup" and "Push Button" tasks Plan2Explore is a stronger baseline. While in the case for the "Take Lid Off Saucepan" task, RND does manage to outperform Plan2Explore, LAMP exhibits stronger performance than either baseline. We report all relevant hyperparameters in Table 10.

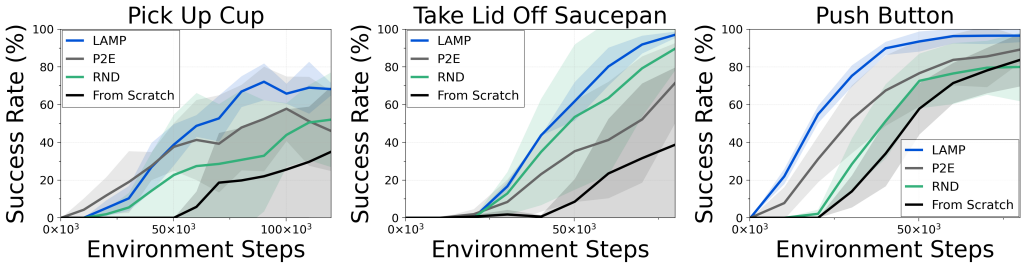

Figure 13: Finetuning performance on visual robotic manipulation tasks in RLBench. The solid line and shaded region represent mean and standard deviation across 3 seeds.

E Hyperparameters

Table 5: Plan2Explore Hyperparameters

Parameter	**Value**
Plan2Explore	False
Exploration Intrinsic Scale	0.9
Exploration Extrinsic Scale	0.1
Exploration Optimization	Optimization: Adam Learning Rate: 3e-4 Epsilon: 1e-5 Clip: 100 Weight Decay: 1e-6
Exploration Head	Layers: [512, 512, 512, 512] Activation: ELU Normalization: None Distribution: MSE
Exploration Reward Normalization	Momentum: 1.0 Scale: 1.0 Epsilon: 1e-8
Disaggregation Target	Stochastic
Disaggregation Log	False
Disaggregation Models	10
Disaggregation Offset	1
Disaggregation Action Condition	True
Exploration Model Loss	KL

Table 6: PTMae Hyperparameters

Parameter	**Value**
MAE Image Width Size	224
MAE Image Height Size	224
WM Flat VIT Image Height Size	7
WM Flat VIT Image Width Size	7
MAE State Prediction	False
WM Flat VIT Input Channels	768
WM Flat VIT Embedding Dimension	128
MAE Average	True

Table 7: MAE Hyperparameters

Parameter	**Value**
Camera Keys	'image front|image wrist'
Mask Ratio	0.95
MAE	Image Height Size: 128
	Image Width Size: 128
	Patch Size: 16
	Embedding Dimension: 256
	Depth: 8
	Number of Heads: 4
	Decoder Embedding Dimension: 256
	Decoder Depth: 6
	Decoder Number of Heads: 4
	Reward Prediction: True
	Early Convolution: True
	State Prediction: True
	Input Channels: 3
	Number of Cameras: 0
	State Dimension: 10
	View Masking: True
	Control Input: 'front wrist'
WM Flat VIT	Image Height Size: 8
	Image Width Size: 8
	Patch Size: 1
	Embedding Dimension: 128
	Depth: 2
	Number of Heads: 4
	Decoder Embedding Dimension: 128
	Decoder Depth: 2
	Decoder Number of Heads: 4
	Input Channels: 256
	State Prediction: False
Image Time Size	4
Use ImageNet MAE	False
MAE Chunk	1
MAE Average	False

Table 8: World Model Hyperparameters

Parameter	**Value**
Grad Heads	[Reward, Discount]
Predictive Discount	True
RSSM	Action Free: False
	Hidden: 1024
	Deterministic: 1024
	Stochastic: 32
	Discrete: 32
	Activation: ELU
	Normalization: None
	Stochastic Activation: Sigmoid2
	Minimum Standard Deviation: 0.1
Reward Head	Layers: [512, 512, 512, 512]
	Activation: ELU
	Normalization: None
	Distribution: Symlog
Discount Head	Layers: [512, 512, 512, 512]
	Activation: ELU
	Normalization: None
	Distribution: Binary
Loss Scales	Feature: 1.0
	KL: 1.0
	Reward: 1.0
	Discount: 1.0
	Proprio: 1.0
	MAE Reward: 1.0
KL	Scale: 1.0
KL Minloss	0.1
KL Balance	0.8
Model Optimization	Optimization: Adam
	Learning Rate: 3e-4
	Epsilon: 1e-5
	Clip: 100.0
	Weight Decay: 1e-6
	Weight Decay Pattern: ‘kernel’
	Warmup: 0
MAE Optimization	Optimization: Adam
	Learning Rate: 3e-4
	Epsilon: 1e-5
	Clip: 100.0
	Weight Decay: 1e-6
	Warmup: 2500

Table 9: Actor Critic Hyperparameters

Parameter	Value
Actor	Layers: [512, 512, 512, 512] Activation: ELU Normalization: None Distribution: Trunc_Normal Minimum Standard Deviation: 0.1
Critic	Layers: [512, 512, 512, 512] Activation: ELU Normalization: None Distribution: MSE
Actor Optimization	Optimization: Adam Learning Rate: 1e-4 Epsilon: 1e-5 Clip: 100.0 Weight Decay: 1e-6 Weight Decay Pattern: 'kernel' Warmup: 0
Critic Optimization	Optimization: Adam Learning Rate: 1e-4 Epsilon: 1e-5 Clip: 100.0 Weight Decay: 1e-6 Weight Decay Pattern: 'kernel' Warmup: 0
Discount	0.99
Discount Lambda	0.95
Image Horizon	15
Actor Grad	Dynamics
Actor Grad Mix	0.1
Actor Entropy	Scale: 1e-4
Slow Target	True
Slow Target Update	100
Slow Target Fraction	1
Slow Baseline	True
Reward Normalization	Momentum: 0.99 Scale: 1.0 Epsilon: 1e-8

Table 10: Random Network Distillation Hyperparameters

Parameter	Value
Embedding Dimension	512
Hidden Dimension	256
Optimizer	Adam Learning Rate: 3e-4 Epsilon: 1e-5 Clip: 100.0 Weight Decay: 1e-6 Warmup: 2500