

Inferring Dynamic Physical Properties from Video Foundation Models



Guanqi Zhan*



Xianzheng Ma*
(presenter)



Weidi Xie



Andrew Zisserman

Visual Geometry Group, University of Oxford

* Equal Contribution

CVPR 2026 VGBE Workshop Best Paper Runner-up Award

Dynamic Physical Properties are Inferred from Visual Cues

Elasticity



After impact, the rebound height relative to the drop height reveals how elastic the object is.

Viscosity



After contact, the spreading speed and area growth reveal how viscous the liquid is.

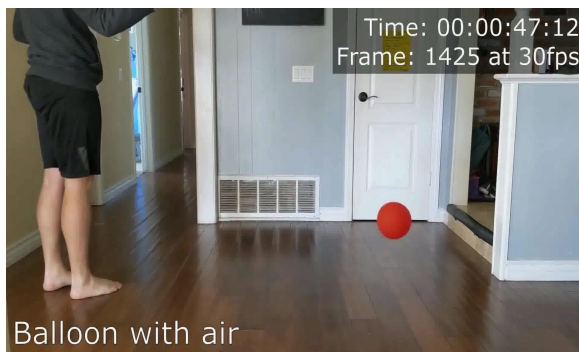
Friction



During sliding, displacement and deceleration reveal the friction between object and surface.

Can video foundation models infer dynamic physical properties from videos

Problem Setting: Absolute Prediction and Relative Comparison



Absolute setting: what is the elasticity of the ball?

Absolute prediction

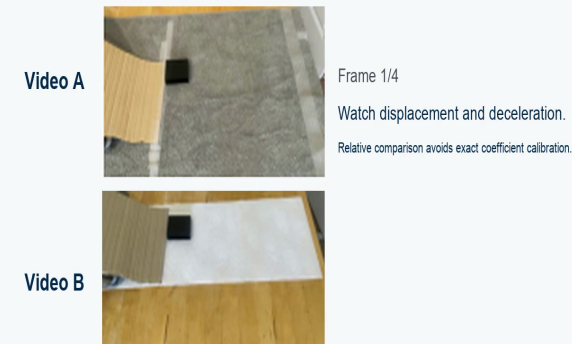
Input: one video

Output: numerical physical coefficient

Metric: Pearson correlation

Relative setting: which video has larger friction?

The model does not need an exact coefficient first; it can compare motion outcomes.



Relative setting: which video has larger friction?

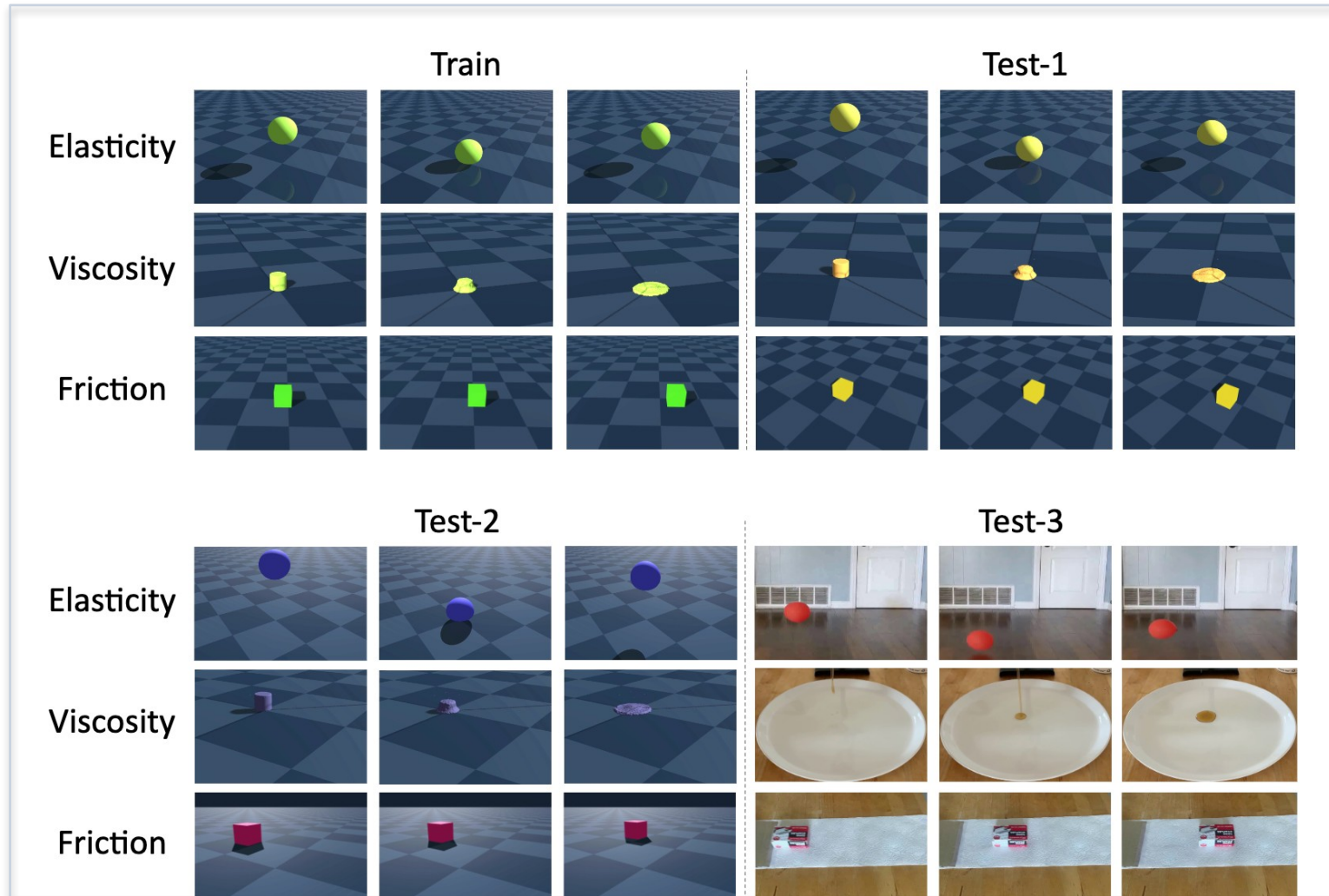
Relative comparison

Input: a pair of videos

Output: which one has the larger property value

Metric: ROC AUC

Overview of Dataset Collection: Simulator + Real



3 properties

Elasticity Viscosity Friction

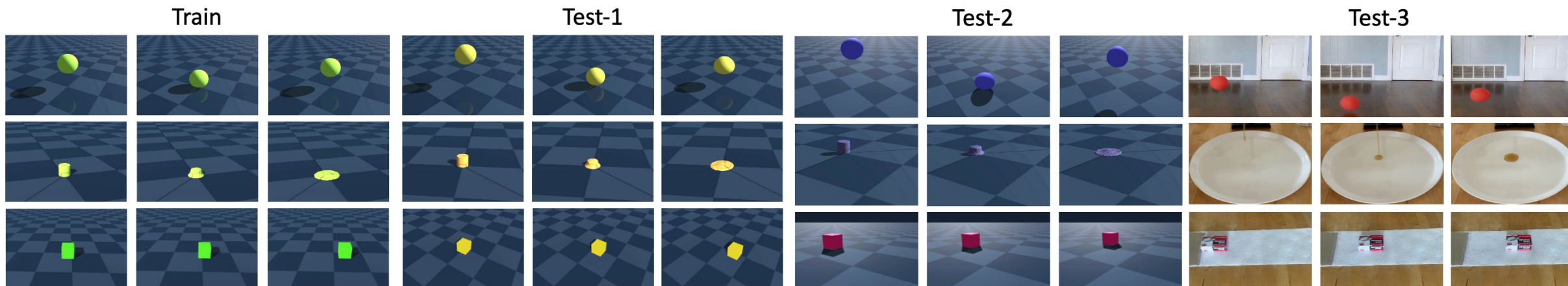
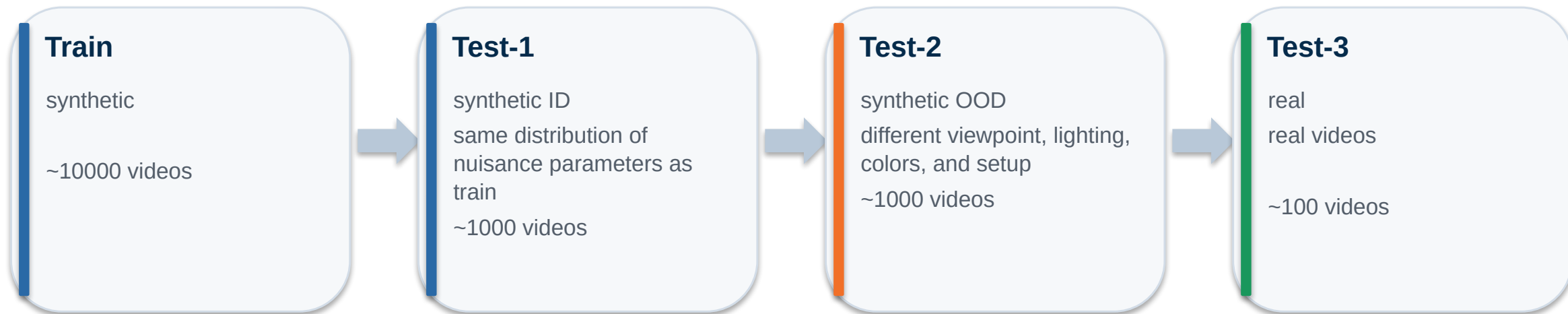
4 splits

Train and Test-1: synthetic in-distribution

Test-2: synthetic OOD

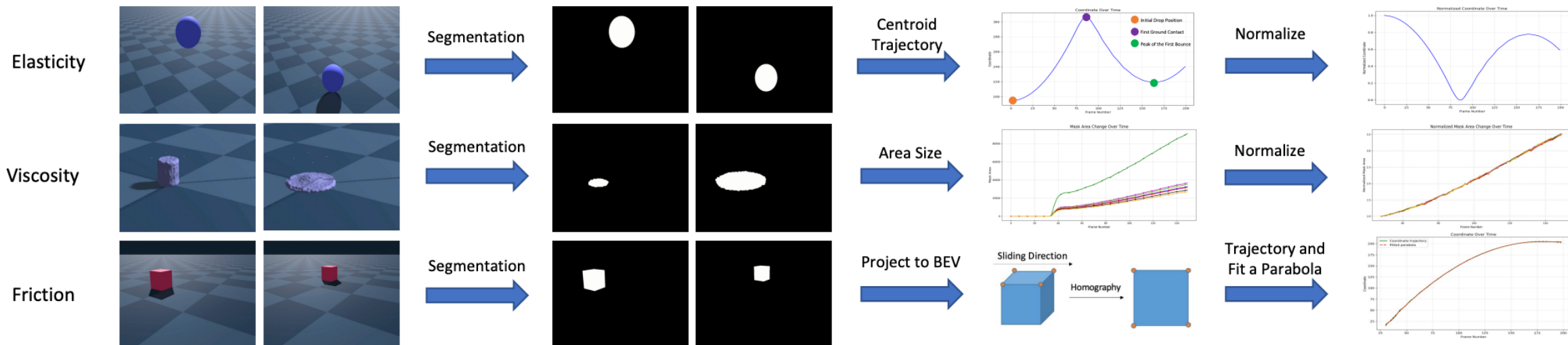
Test-3: real videos

The split design separates physics from nuisance parameters



Oracle Method

Key idea: use computer vision techniques to extract a measurement that can directly be used to predict the property



Elasticity

- segment object
- track centroid
- bounce height ratio

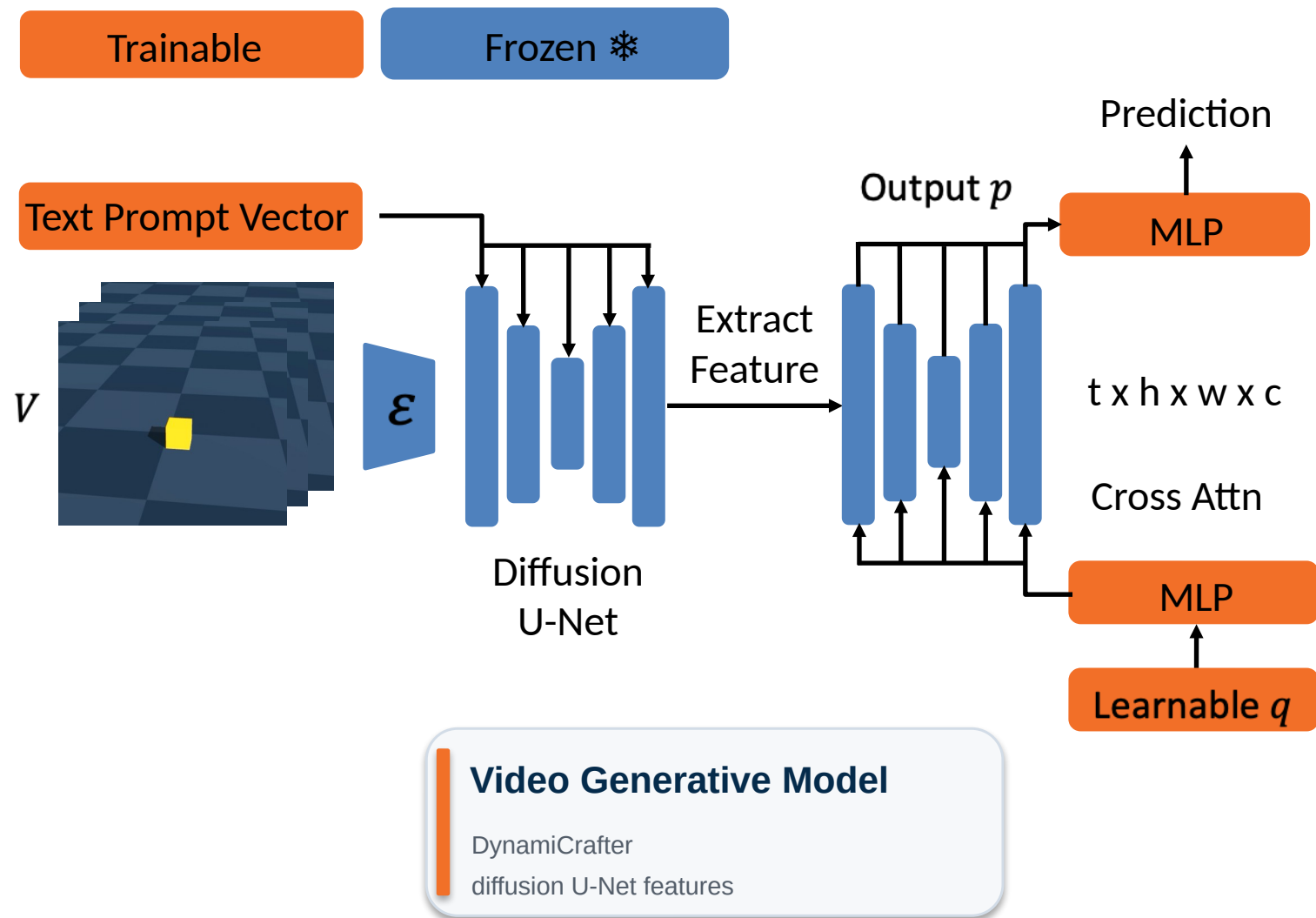
Viscosity

- segment liquid
- track area growth
- spreading slope

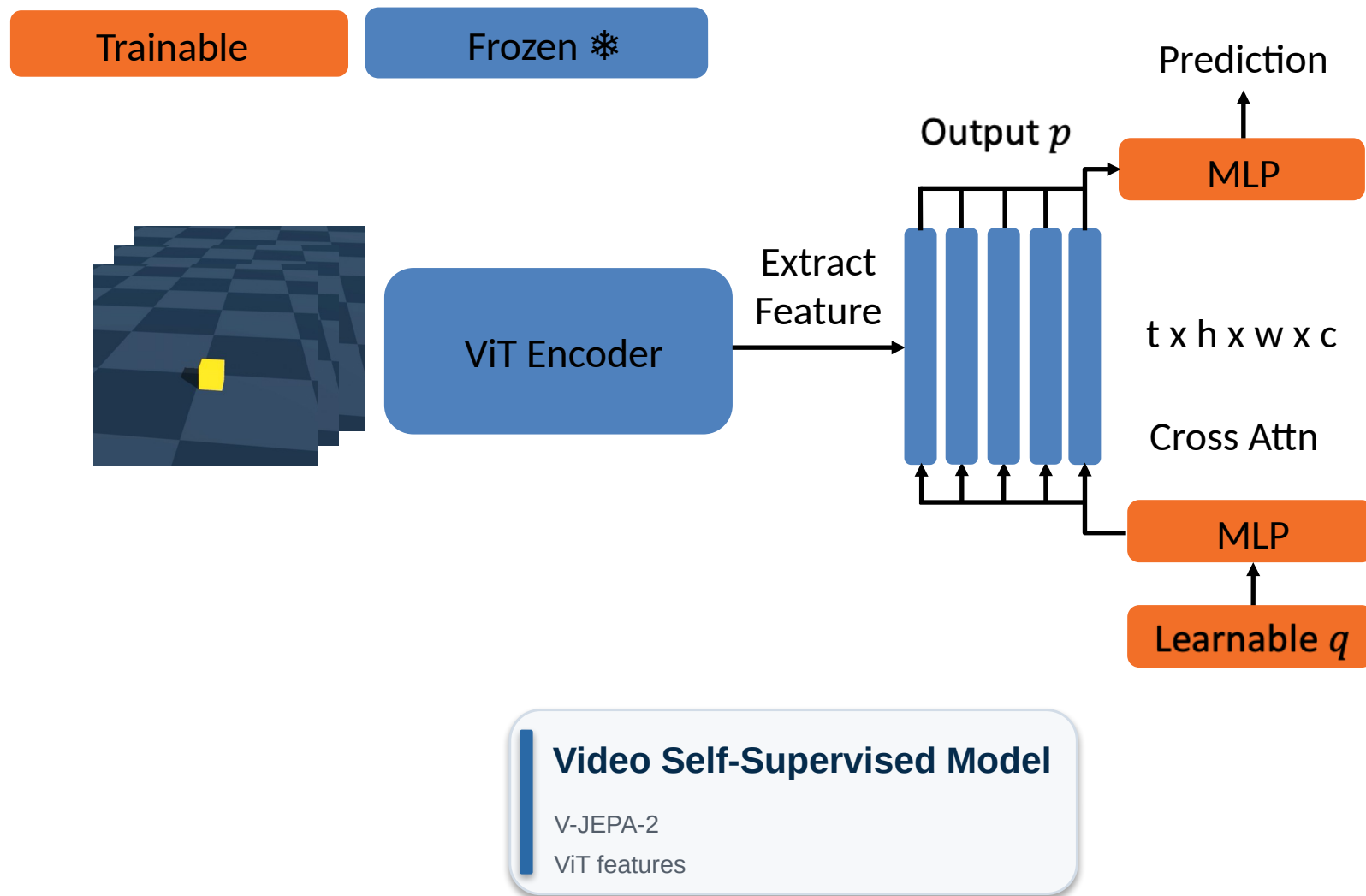
Friction

- bird's-eye projection
- fit trajectory
- parabola coefficient

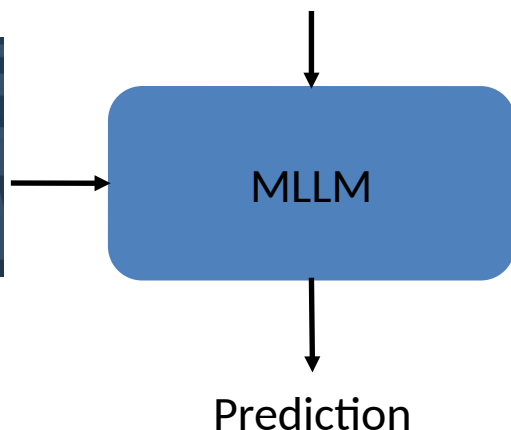
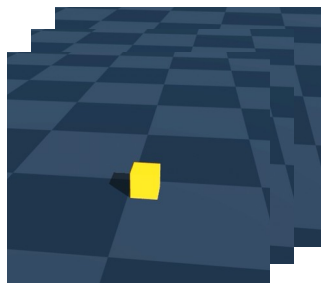
Video Generative Models



Video Self-Supervised Models



What is the dynamic friction coefficient of the object and the surface in the video?



```
'role': 'system', 'content': 'You are an expert in elasticity analysis with the ability to analyze video frames and determine elasticity values. You MUST provide a specific elasticity estimate. Never say you cannot analyze the images.'
```

System Prompt

```
'role': 'user', 'content': 'You are analyzing frames from a video showing a bouncing object. Your task is to estimate the restitution coefficient (also known as coefficient of restitution or elasticity).'
```

The restitution coefficient is the ratio of separation speed to approach speed during collision:

- Perfectly elastic collision: 1.0 (ball bounces back with same speed)
- Perfectly inelastic collision: 0.0 (ball sticks to surface, no bounce)
- Typical values: 0.1-0.9 for most materials

If the object bounces back significantly → restitution coefficient closer to 1.0
If the object barely bounces or sticks → restitution coefficient closer to 0.0

YOUR RESPONSE MUST END WITH A SPECIFIC RESTITUTION COEFFICIENT VALUE IN THIS FORMAT:
"Estimated restitution coefficient value: X "

Where X is a number between 0 and 1 (e.g., 0.7, 0.3, 0.85).' }

Physical Property

```
visual_input = {
  "type": "image_url", "image_url": {"url": f"data:image/jpeg;base64,{frame1}"}}
  "type": "image_url", "image_url": {"url": f"data:image/jpeg;base64,{frame2}"}}
  ...
  "type": "image_url", "image_url": {"url": f"data:image/jpeg;base64,{frame16}"}}
}
```

Visual Input

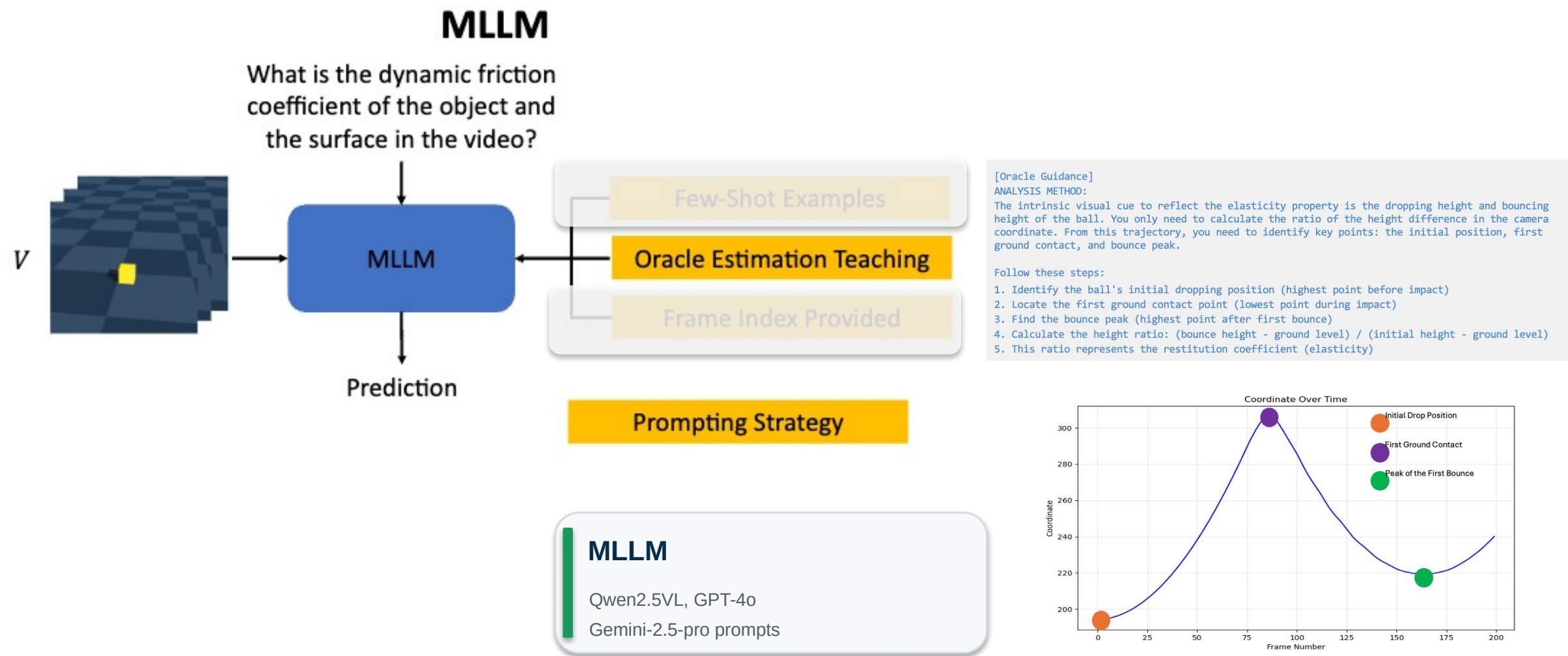
Baseline Prompt

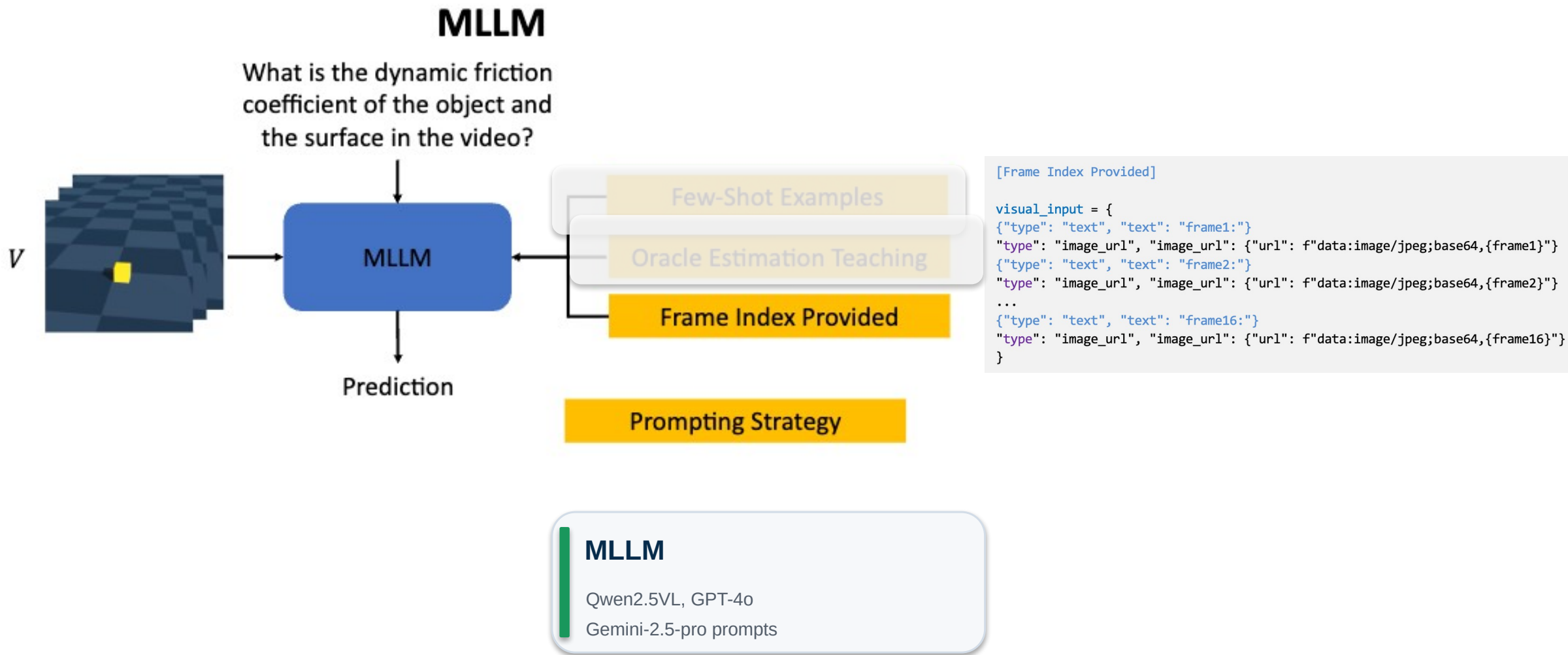
MLLM

Qwen2.5VL, GPT-4o

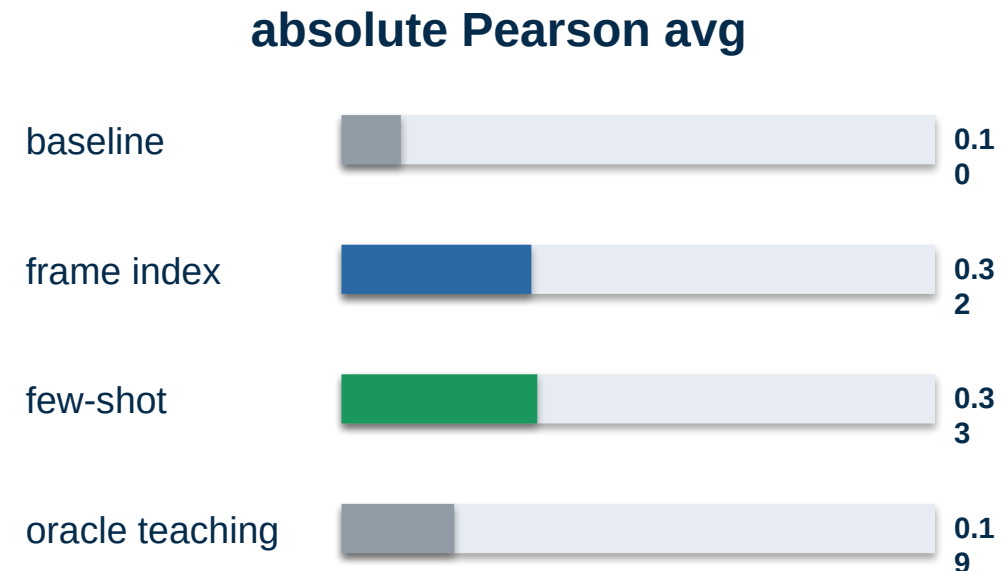
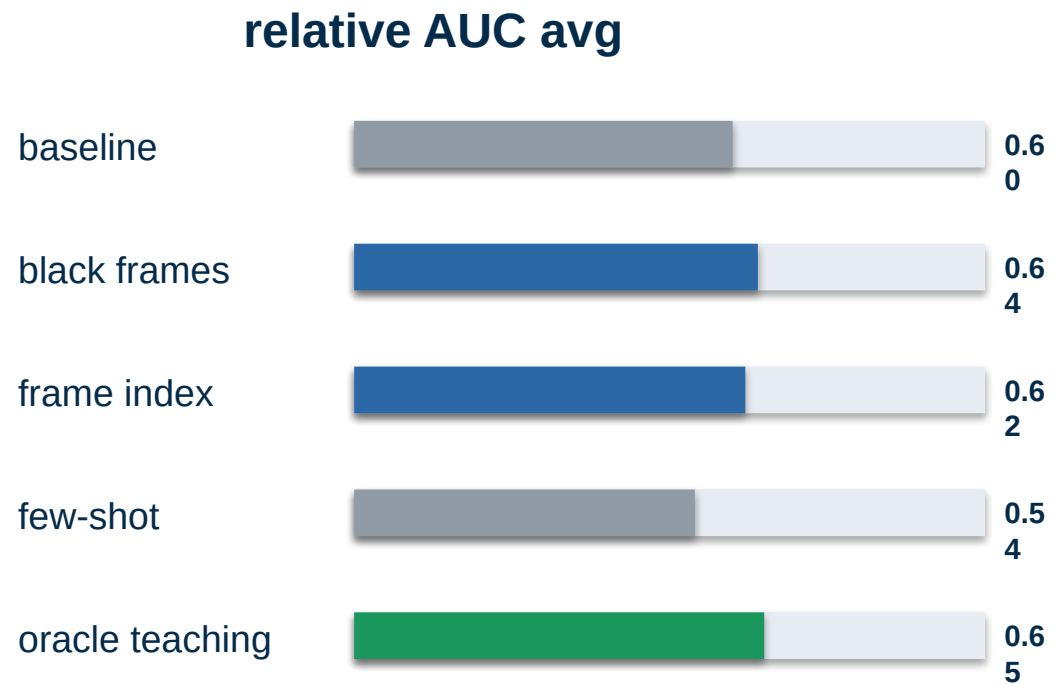
Gemini-2.5-pro prompts







Prompting Strategy for Video-Language Models



Relative Result

Property	Method	Relative – ROC AUC			Absolute – Pearson Corr.		
		Test-1	Test-2	Test-3	Test-1	Test-2	Test-3
Elasticity	Oracle	1.00	0.99	1.00	0.99	0.98	0.87
	DynamiCrafter	1.00	0.99	0.61	0.92	0.78	0.10
	V-JEPA-2	1.00	1.00	0.99	0.89	0.65	0.37
	Qwen2.5VL-max	0.59	0.50	0.54	-0.05	0.11	0.16
Viscosity	GPT-4o	0.51	0.66	0.62	0.19	0.11	0.30
	Gemini-2.5-pro	0.64	0.80	0.47	0.04	0.15	0.24
	Oracle	0.99	0.99	1.00	0.99	0.98	0.80
	DynamiCrafter	1.00	1.00	0.93	0.99	0.95	0.44
Friction	V-JEPA-2	1.00	0.94	0.99	0.98	0.69	0.61
	Qwen2.5VL-max	0.64	0.61	0.86	0.16	0.06	0.02
	GPT-4o	0.63	0.59	0.99	0.18	0.08	0.55
	Gemini-2.5-pro	0.48	0.69	0.95	-0.06	-0.05	0.60
Friction	Oracle	1.00	1.00	0.87	0.99	1.00	0.83
	DynamiCrafter	0.96	0.90	0.97	0.95	0.78	0.25
	+ Domain Adaptation	–	–	–	–	–	0.80*
	V-JEPA-2	0.97	0.90	0.48	0.87	0.56	0.21
	+ Domain Adaptation	–	–	0.87*	–	–	0.73*
	Qwen2.5VL-max	0.52	0.48	0.80	-0.03	0.03	0.06
	GPT-4o	0.49	0.40	0.67	0.07	0.10	0.38
	Gemini-2.5-pro	0.56	0.52	0.97	0.01	0.10	0.12

Absolute Result

Property	Method	Relative – ROC AUC			Absolute – Pearson Corr.		
		Test-1	Test-2	Test-3	Test-1	Test-2	Test-3
Elasticity	Oracle	1.00	0.99	1.00	0.99	0.98	0.87
	DynamiCrafter	1.00	0.99	0.61	0.92	0.78	0.10
	V-JEPA-2	1.00	1.00	0.99	0.89	0.65	0.37
	Qwen2.5VL-max	0.59	0.50	0.54	-0.05	0.11	0.16
	GPT-4o	0.51	0.66	0.62	0.19	0.11	0.30
	Gemini-2.5-pro	0.64	0.80	0.47	0.04	0.15	0.24
Viscosity	Oracle	0.99	0.99	1.00	0.99	0.98	0.80
	DynamiCrafter	1.00	1.00	0.93	0.99	0.95	0.44
	V-JEPA-2	1.00	0.94	0.99	0.98	0.69	0.61
	Qwen2.5VL-max	0.64	0.61	0.86	0.16	0.06	0.02
	GPT-4o	0.63	0.59	0.99	0.18	0.08	0.55
	Gemini-2.5-pro	0.48	0.69	0.95	-0.06	-0.05	0.60
Friction	Oracle	1.00	1.00	0.87	0.99	1.00	0.83
	DynamiCrafter	0.96	0.90	0.97	0.95	0.78	0.25
	+ Domain Adaptation	–	–	–	–	–	0.80*
	V-JEPA-2	0.97	0.90	0.48	0.87	0.56	0.21
	+ Domain Adaptation	–	–	0.87*	–	–	0.73*
	Qwen2.5VL-max	0.52	0.48	0.80	-0.03	0.03	0.06
	GPT-4o	0.49	0.40	0.67	0.07	0.10	0.38
	Gemini-2.5-pro	0.56	0.52	0.97	0.01	0.10	0.12

Absolute Result

Property	Method	Relative – ROC AUC			Absolute – Pearson Corr.		
		Test-1	Test-2	Test-3	Test-1	Test-2	Test-3
Elasticity	Oracle	1.00	0.99	1.00	0.99	0.98	0.87
	DynamiCrafter	1.00	0.99	0.61	0.92	0.78	0.10
	V-JEPA-2	1.00	1.00	0.99	0.89	0.65	0.37
	Qwen2.5VL-max	0.59	0.50	0.54	-0.05	0.11	0.16
	GPT-4o	0.51	0.66	0.62	0.19	0.11	0.30
	Gemini-2.5-pro	0.64	0.80	0.47	0.04	0.15	0.24
Viscosity	Oracle	0.99	0.99	1.00	0.99	0.98	0.80
	DynamiCrafter	1.00	1.00	0.93	0.99	0.95	0.44
	V-JEPA-2	1.00	0.94	0.99	0.98	0.69	0.61
	Qwen2.5VL-max	0.64	0.61	0.86	0.16	0.06	0.02
	GPT-4o	0.63	0.59	0.99	0.18	0.08	0.55
	Gemini-2.5-pro	0.48	0.69	0.95	-0.06	-0.05	0.60
Friction	Oracle	1.00	1.00	0.87	0.99	1.00	0.83
	DynamiCrafter	0.96	0.90	0.97	0.95	0.78	0.25
	+ Domain Adaptation	–	–	–	–	–	0.80*
	V-JEPA-2	0.97	0.90	0.48	0.87	0.56	0.21
	+ Domain Adaptation	–	–	0.87*	–	–	0.73*
	Qwen2.5VL-max	0.52	0.48	0.80	-0.03	0.03	0.06
	GPT-4o	0.49	0.40	0.67	0.07	0.10	0.38
	Gemini-2.5-pro	0.56	0.52	0.97	0.01	0.10	0.12

Absolute Result

Property	Method	Relative – ROC AUC			Absolute – Pearson Corr.		
		Test-1	Test-2	Test-3	Test-1	Test-2	Test-3
Elasticity	Oracle	1.00	0.99	1.00	0.99	0.98	0.87
	DynamiCrafter	1.00	0.99	0.61	0.92	0.78	0.10
	V-JEPA-2	1.00	1.00	0.99	0.89	0.65	0.37
	Qwen2.5VL-max	0.59	0.50	0.54	-0.05	0.11	0.16
	GPT-4o	0.51	0.66	0.62	0.19	0.11	0.30
	Gemini-2.5-pro	0.64	0.80	0.47	0.04	0.15	0.24
Viscosity	Oracle	0.99	0.99	1.00	0.99	0.98	0.80
	DynamiCrafter	1.00	1.00	0.93	0.99	0.95	0.44
	V-JEPA-2	1.00	0.94	0.99	0.98	0.69	0.61
	Qwen2.5VL-max	0.64	0.61	0.86	0.16	0.06	0.02
	GPT-4o	0.63	0.59	0.99	0.18	0.08	0.55
	Gemini-2.5-pro	0.48	0.69	0.95	-0.06	-0.05	0.60
Friction	Oracle	1.00	1.00	0.87	0.99	1.00	0.83
	DynamiCrafter	0.96	0.90	0.97	0.95	0.78	0.25
	+ Domain Adaptation	–	–	–	–	–	0.80*
	V-JEPA-2	0.97	0.90	0.48	0.87	0.56	0.21
	+ Domain Adaptation	–	–	0.87*	–	–	0.73*
	Qwen2.5VL-max	0.52	0.48	0.80	-0.03	0.03	0.06
	GPT-4o	0.49	0.40	0.67	0.07	0.10	0.38
	Gemini-2.5-pro	0.56	0.52	0.97	0.01	0.10	0.12

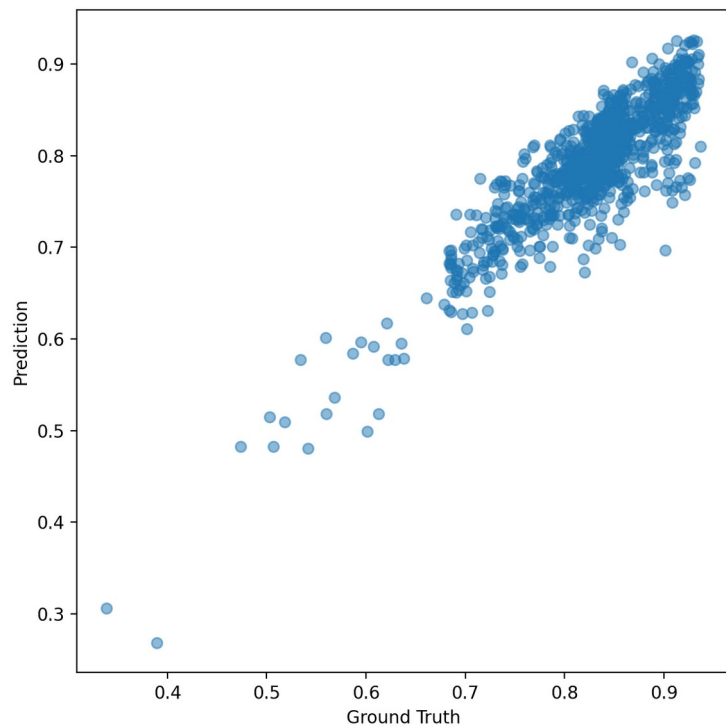
Absolute Result

Property	Method	Relative – ROC AUC			Absolute – Pearson Corr.		
		Test-1	Test-2	Test-3	Test-1	Test-2	Test-3
Elasticity	Oracle	1.00	0.99	1.00	0.99	0.98	0.87
	DynamiCrafter	1.00	0.99	0.61	0.92	0.78	0.10
	V-JEPA-2	1.00	1.00	0.99	0.89	0.65	0.37
	Qwen2.5VL-max	0.59	0.50	0.54	-0.05	0.11	0.16
	GPT-4o	0.51	0.66	0.62	0.19	0.11	0.30
	Gemini-2.5-pro	0.64	0.80	0.47	0.04	0.15	0.24
Viscosity	Oracle	0.99	0.99	1.00	0.99	0.98	0.80
	DynamiCrafter	1.00	1.00	0.93	0.99	0.95	0.44
	V-JEPA-2	1.00	0.94	0.99	0.98	0.69	0.61
	Qwen2.5VL-max	0.64	0.61	0.86	0.16	0.06	0.02
	GPT-4o	0.63	0.59	0.99	0.18	0.08	0.55
	Gemini-2.5-pro	0.48	0.69	0.95	-0.06	-0.05	0.60
Friction	Oracle	1.00	1.00	0.87	0.99	1.00	0.83
	DynamiCrafter	0.96	0.90	0.97	0.95	0.78	0.25
	+ Domain Adaptation	–	–	–	–	–	0.80*
	V-JEPA-2	0.97	0.90	0.48	0.87	0.56	0.21
	+ Domain Adaptation	–	–	0.87*	–	–	0.73*
	Qwen2.5VL-max	0.52	0.48	0.80	-0.03	0.03	0.06
	GPT-4o	0.49	0.40	0.67	0.07	0.10	0.38
	Gemini-2.5-pro	0.56	0.52	0.97	0.01	0.10	0.12

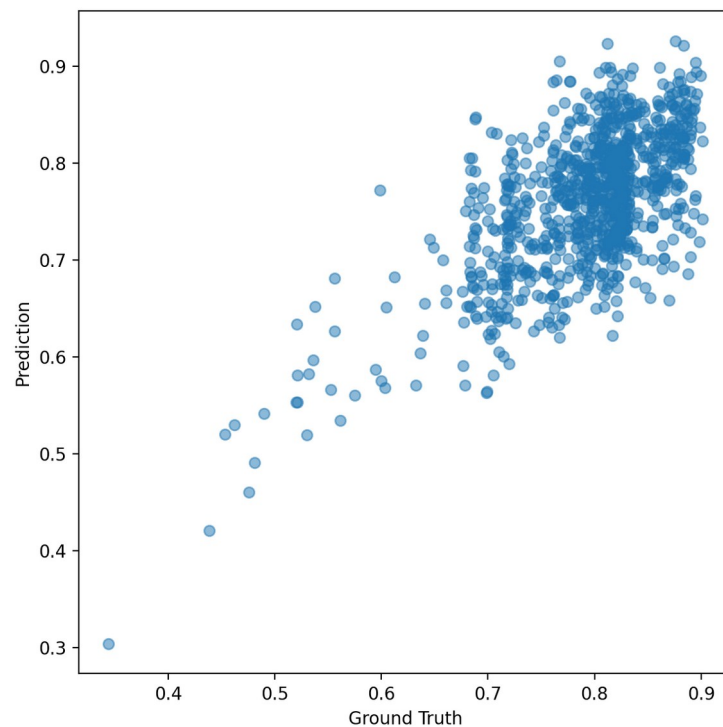
Absolute Test-1/2/3

V-JEPA-2 – scatter plot

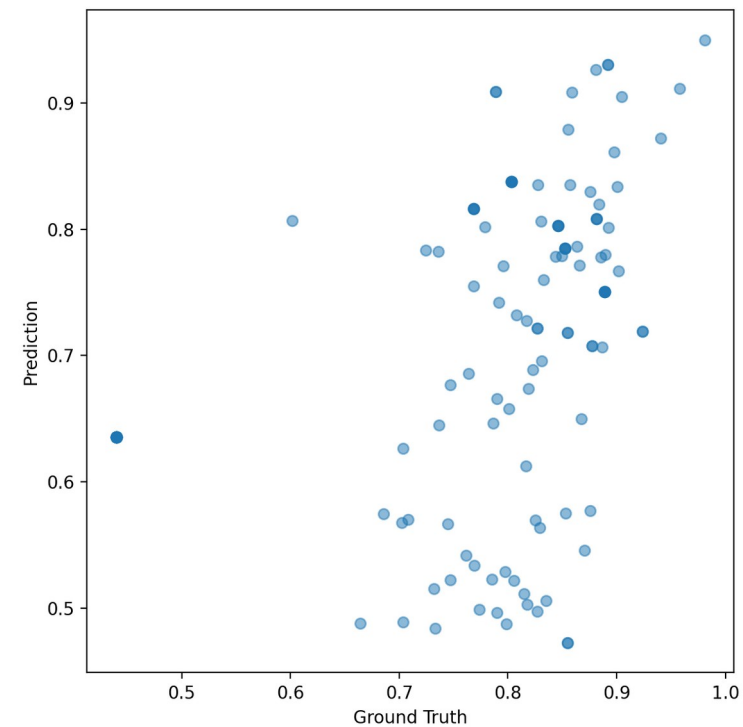
Test-1



Test-2



Test-3



Summary

(i) Dataset

We collect a **new video dataset** for each physical property, consisting of synthetic training and testing splits, as well as a real split for real world evaluation.

(ii) Three ways to infer the physical property from videos

- (a) an **oracle method** where we supply the visual cues that intrinsically reflect the property using classical computer vision techniques;
- (b) a simple read out mechanism using a visual prompt and trainable prompt vector for cross-attention on pre-trained video **generative and self-supervised** models;
- (c) **prompt strategies** for Multi-modal Large Language Models (MLLMs).

(iii) Experiments

We show that a video foundation model trained in a **generative** (DynamiCrafter) or trained in a **self-supervised** manner (V-JEPA-2) achieve a **generally similar** performance, though **behind** that of the **oracle**, and that MLLMs are currently inferior to the other models, though their performance can be improved through **suitable prompting**.

CVPR
JUNE 3-7, 2026



DENVER
COLORADO



Thank you!



Project Page

Paper: [arXiv:2510.02311](https://arxiv.org/abs/2510.02311) | Dataset: [Oliver-Ma/IDPP-Data](#)