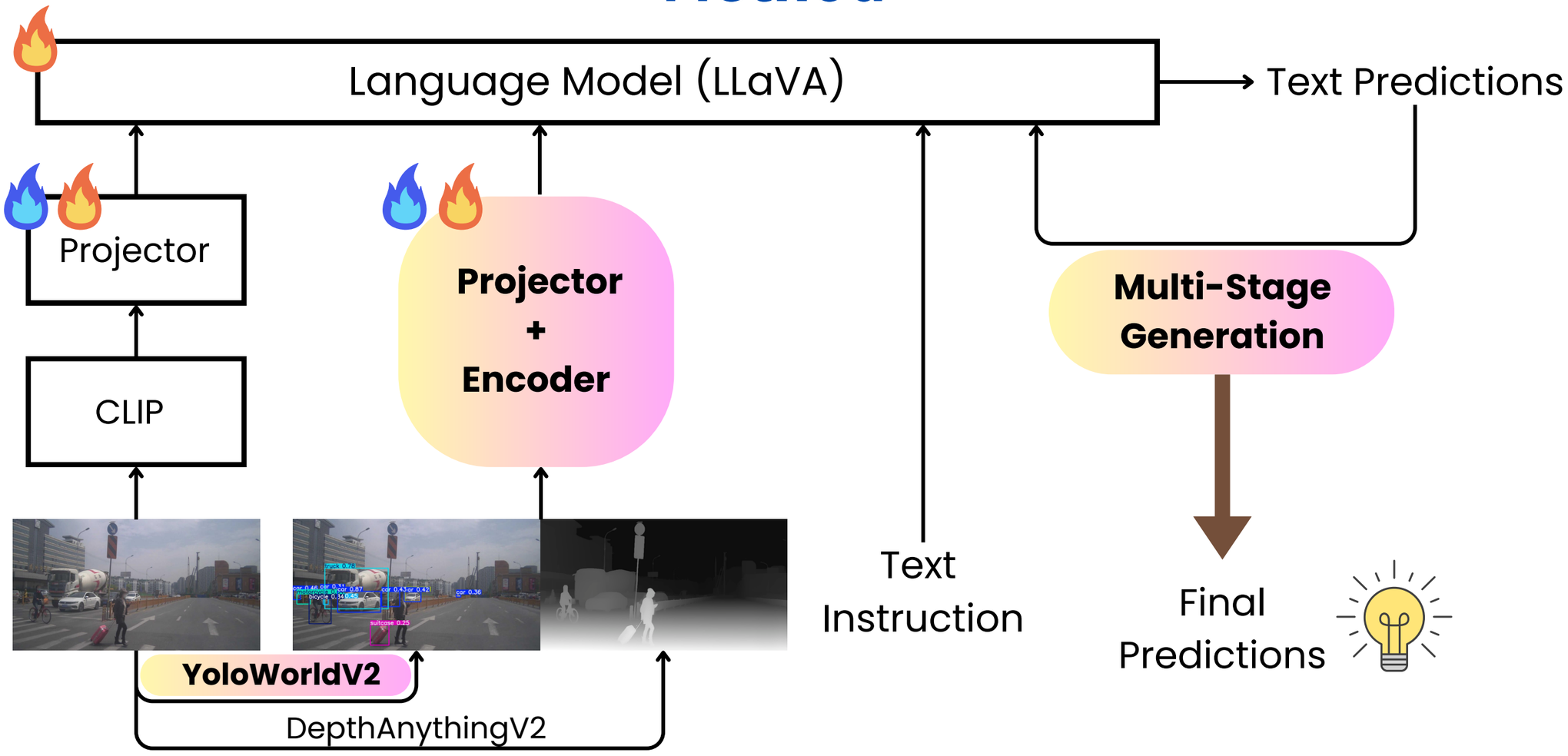
PREVISION

PRe-training Enhanced Versatile Integration of Semantics, Images, and Object Detection for Novel Corner Case Analysis

電機四 許景淯(B10901036) 李冠儀(B10901091) 劉庭均(B10901039) 傅啟恩(B11901174)

NTU DLCV-FALL 2024 Analysis of Corner Cases in Autonomous Driving

Method



Pre-training (No Labels Needed)

CODA-LM + A2D2 + nuScenes | Constructed QA-pairs

FineTuning

CODA-LM | GT conversation

Ablation Study

Algorithms	Settings	Final Score	BLEU
LLaVA	Zero Shot (init prompt)	2.41	0.18
	Zero Shot (revised prompt)	3.03	0.32
	Finetune	×	×
LLaVA*	Finetune (LoRA)	3.90	0.43
	Finetune + Postprocess	**4.09**	**0.48**
Pretrained LLaVA*	Finetune (LoRA)	3.77	0.43
	Finetune (**DoRA**)	3.85	0.46
	Finetune + Multistage	≤ 2	×

Conclusions

1. Our designed architecture is effective, improving the model's performance from 3.0 to 4.1.
2. High BLEU score indicates great training effects
3. Pretraining both projectors can lead to worse results because our task design amplifies errors from the open-vocab detection model.
4. Overfitting occurs when training exceeds 8000 steps (1 epoch)