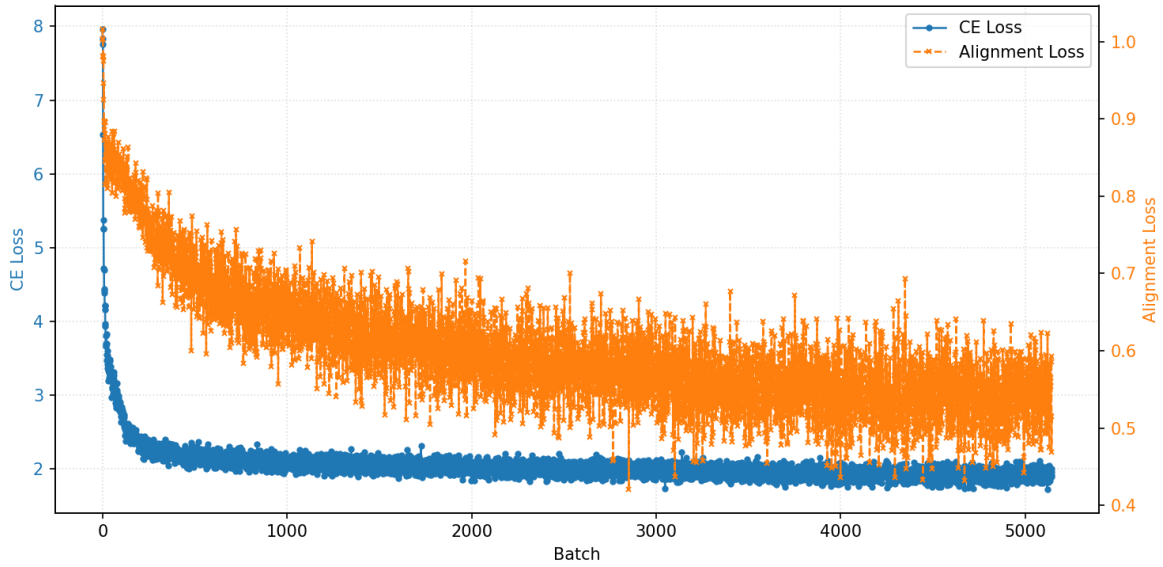
Results

Table 1: Training Losses

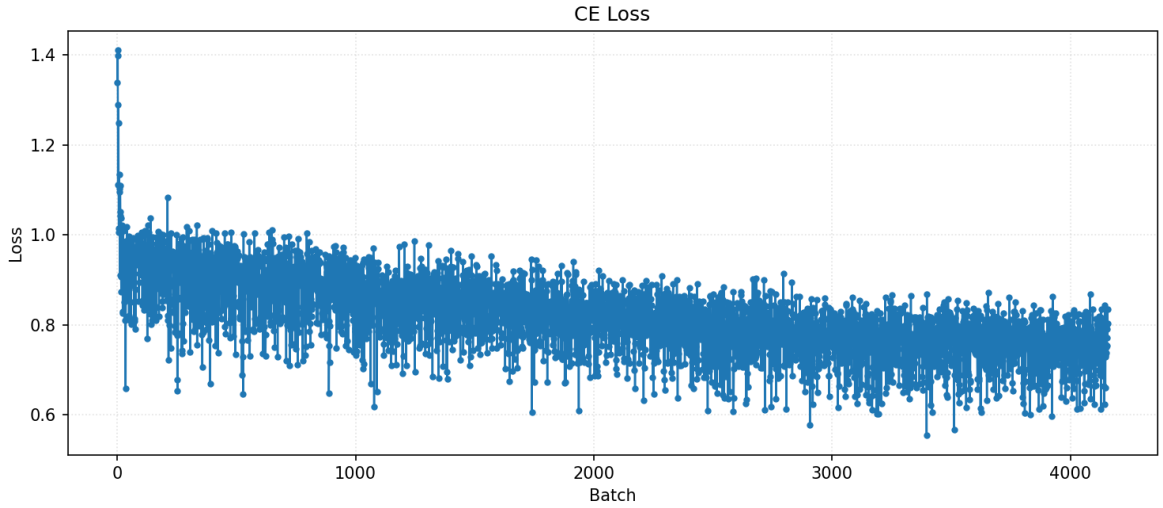
	CE Loss	Alignment Loss
First Stage	2.046	0.514
Second Stage	0.819	—

Dataset	**Perception Score**	**Cognition Score**	**LLaVA Baseline**
mme	293.2 ↓	1459.4 ↓	348.2/1510.8
Dataset	**Exact Match**	**stderr**	**LLaVA Baseline**
ai2d	0.5521 ↑	0.009	0.548
scienceqa_img	0.6891 ↓	0.0103	0.704
gqa	0.6189 ↓	0.0043	0.620
textvqa_val	0.5174	0.0068	?
Dataset	**Accuracy**		**LLaVA Baseline**
mmmu_val	0.3611 ↑		0.353
Dataset	**Accuracy**	**F1 Score**	**LLaVA Baseline**
pope	0.8616	0.8500	?

Dataset	**seed_all**	**image**	**video**	**LLaVA Baseline**
seedbench				60.5



(a) First Stage Losses



(b) Second Stage Losses