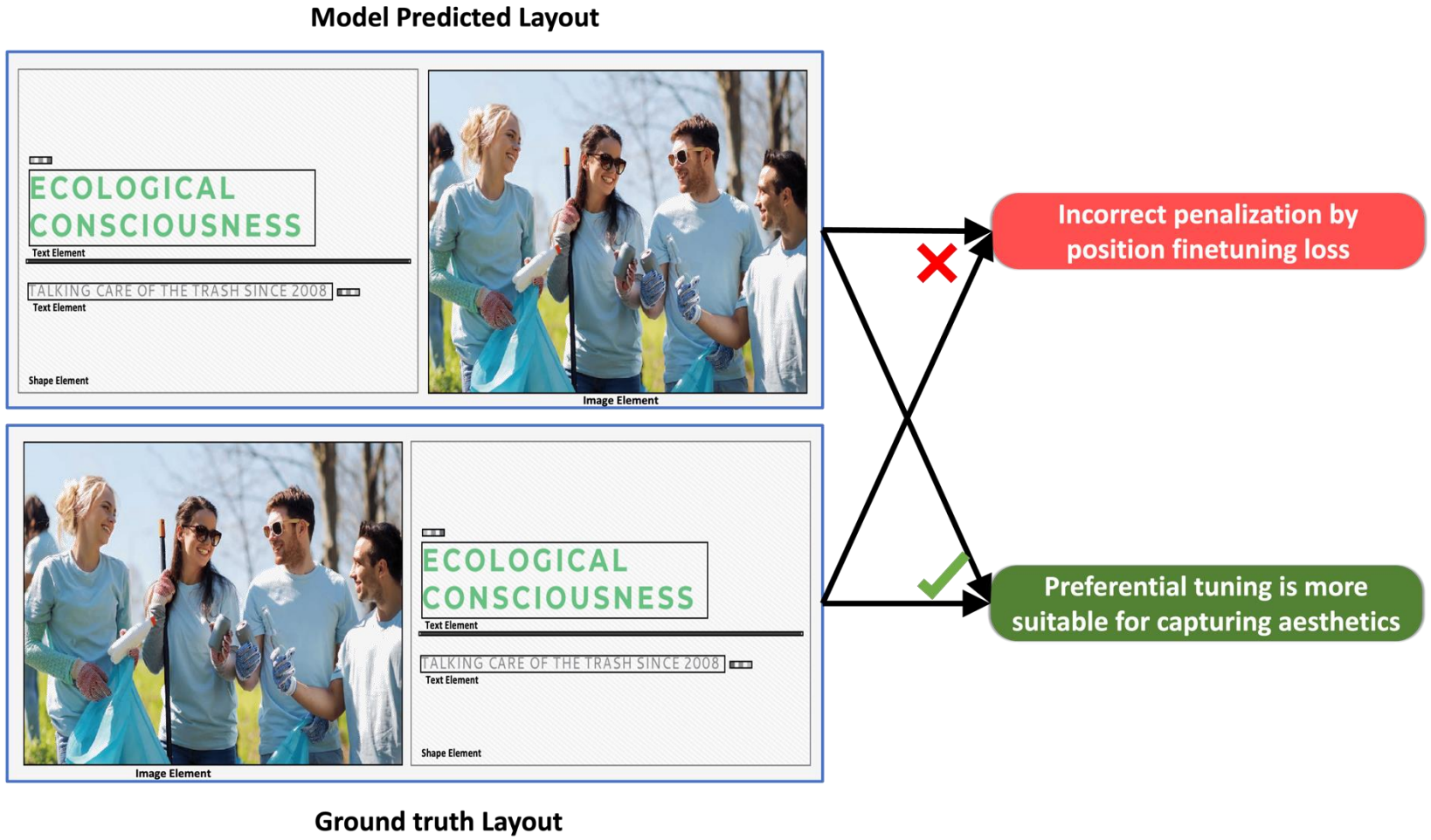
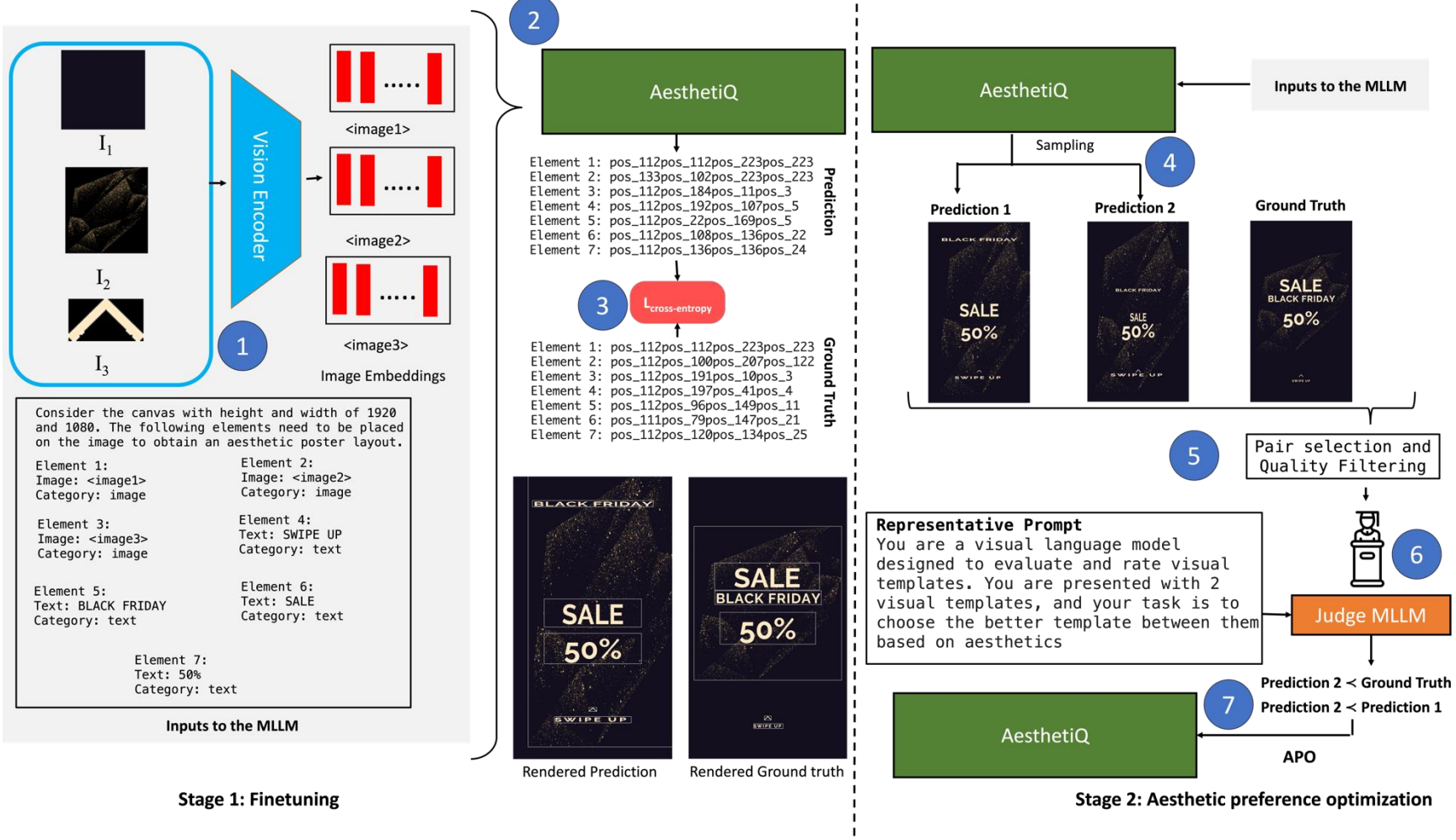
AesthetiQ: Enhancing Graphic Layout Design via Aesthetic-Aware Preference Alignment



Existing cross-entropy loss based methods penalize element misalignment heavily while preferential tuning via AAPA better capture aesthetic nuances in layouts



The training for the aesthetic layout prediction task consists of the following steps: 1) **Vision Encoder**: Design elements (images and text) are processed to generate image and text embeddings. 2) **AesthetiQ Model Prediction**: Embeddings are passed to the AesthetiQ model, which predicts layout coordinates. 3) **Training with Cross-Entropy Loss**: The predicted layout is compared with the ground truth and trained using cross-entropy loss. 4) **Sampling for Comparison**: Multiple layout predictions are generated using AesthetiQ inference. 5) **Pair Selection and Quality Filtering**: We filter the data based on quality heuristics to ensure layout quality in samples. 6) **Judging by ViLA**: The ViLA model compares layout pairs and selects the better one based on aesthetic preferences. 7) **Aesthetic Preference Optimization (AAPA)**: Feedback from ViLA is used to fine-tune the AesthetiQ model for aesthetic optimization.

Qualitative comparison of our model, AesthetiQ, against recent methods FlexDM, LACE, and LayoutNUWA. Despite the challenge of arranging numerous elements, AesthetiQ consistently achieves superior layout quality. In row (a), AesthetiQ effectively places text within salient regions, maintaining clear hierarchy and avoiding overlaps, which enhances readability and aesthetic appeal. In row (b), it achieves precise alignment across elements and optimally positions diverse shapes, preserving a cohesive visual structure. Row (c) showcases AesthetiQ's advanced semantic understanding, generating a visually balanced and aesthetically pleasing layout. Overall, AesthetiQ consistently outperforms competitors in creating coherent, well-structured designs that align with human aesthetic preferences.

Method	Mean IoU (%)			$\mathcal{M}_{\text{judge}}$
	All	Single	Multiple	Win Rate (%)
SmartText+	-	4.7	2.3	-
Typography LMM	-	40.2	17.2	-
FlexDM	12.71	35.5	10.3	0.93
LACE	23.18	41.96	21.49	3.51
PosterLLaVa	25.18	42.74	23.58	5.03
LayoutNUWA	25.74	43.83	24.16	5.58
AesthetiQ-1B	22.85	40.83	26.55	2.43
AesthetiQ-2B	28.19	45.92	30.44	6.13
AesthetiQ-4B	38.16	49.27	37.14	14.74
AesthetiQ-8B	**42.83**	**52.67**	**40.64**	**17.19**

Table 1. Comparison of layout generation methods based on Mean IoU (%) and Judge Win Rate (%) on Crello Dataset. AesthetiQ models outperform baselines, achieving higher IoU and Judge Win Rate, with AesthetiQ-8B showing the best overall performance.

Method	Mean IoU (%)	$\mathcal{M}_{\text{judge}}$ Win Rate (%)
Desigen [6]	15.36	4.81
LACE [1]	17.88	5.27
PosterLLaVa [7]	30.19	14.73
LayoutNUWA [5]	32.16	15.28
AesthetiQ-1B	38.47	19.29
AesthetiQ-2B	41.42	21.87
AesthetiQ-4B	44.16	22.74
AesthetiQ-8B	**48.29**	**24.48**

Table 1. Comparison of AesthetiQ with baseline methods on the WebUI dataset, evaluated using Mean IoU (%) and $\mathcal{M}$judge Win Rate (%). The results demonstrate the superior performance of AesthetiQ across all model scales, with notable gains in aesthetic and structural alignment metrics.

Quantitative comparision of AesthetiQ on Crello and WebUI dataset


Adobe