
Photography Perspective Composition: Towards Aesthetic Perspective Recommendation

Lujian Yao* **Siming Zheng[†]** **Xinbin Yuan** **Zhuoxuan Cai** **Pu Wu**
Jinwei Chen **Bo Li** **Peng-Tao Jiang[†]✉**
vivo Mobile Communication Co., Ltd

Reporter: **Lujian Yao**

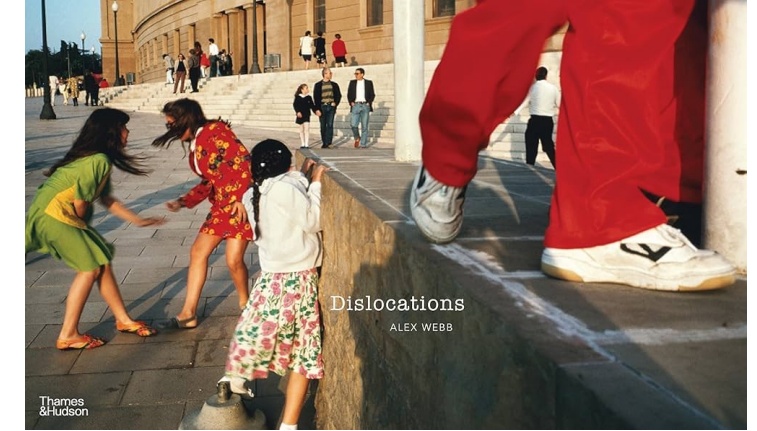
*Intern at vivo Mobile Communication Co., Ltd.

[†]Project lead.

✉Corresponding author.

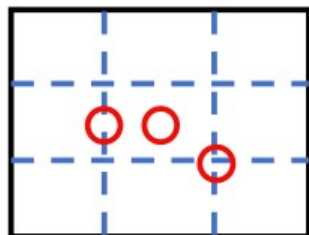
The Meaning of Photo Composition Recommend

- Professional photography costs are high, as taking a good photo requires considering multiple factors.
- Composition is one of the most important factors.
- From a definition perspective, composition refers to "the arrangement of visual elements or the overall harmonious way of organizing components".
- In photography, don't focus on those visual elements, but follow aesthetic principles for composition.
- However, aesthetic composition requires professional knowledge and extensive training. Ordinary people cannot compose like professional photographers.



From: Alex Webb

Motivation



❌ Crop Templates

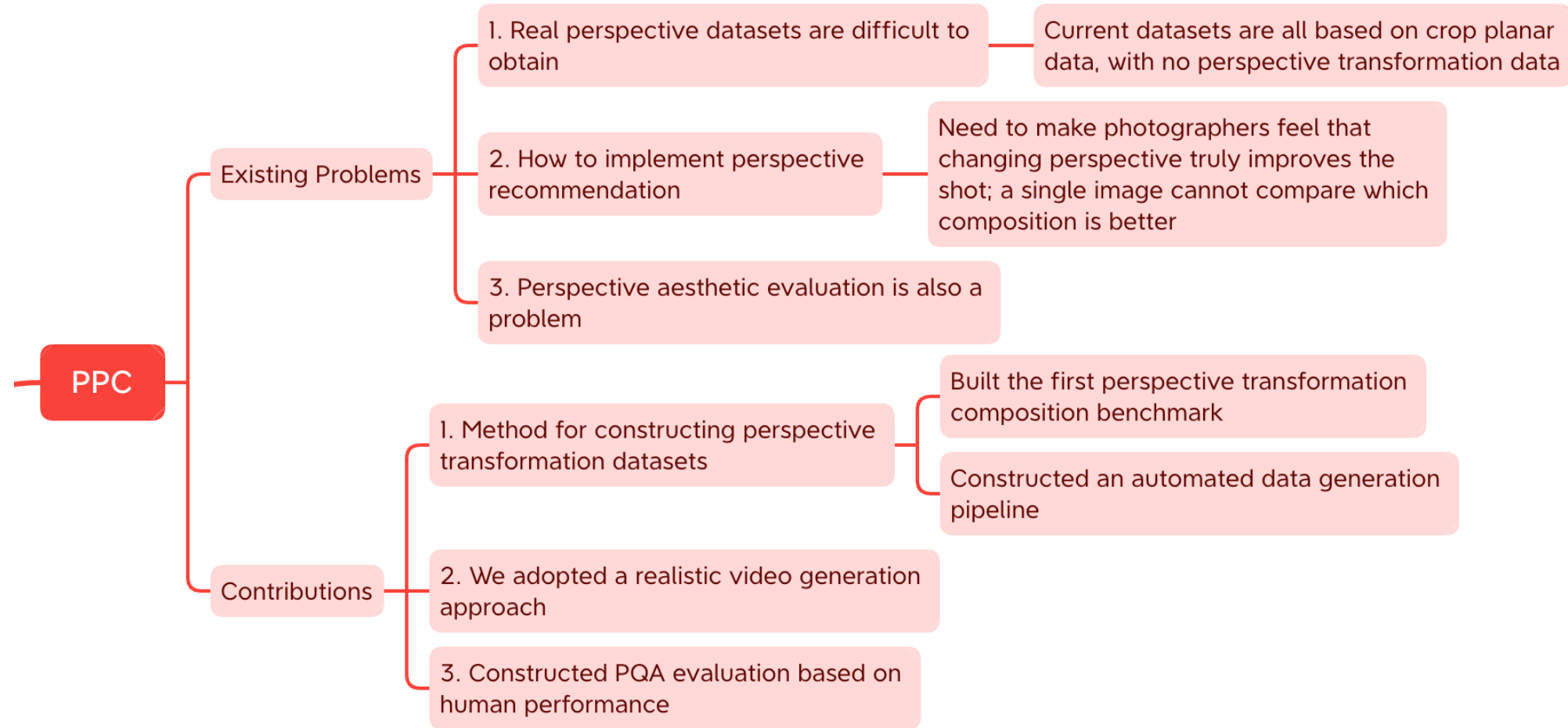


(a) Crop-based Photo Composition

✅ Perspective Composition Recommendation



(b) Our Perspective-based Photo Composition



Method

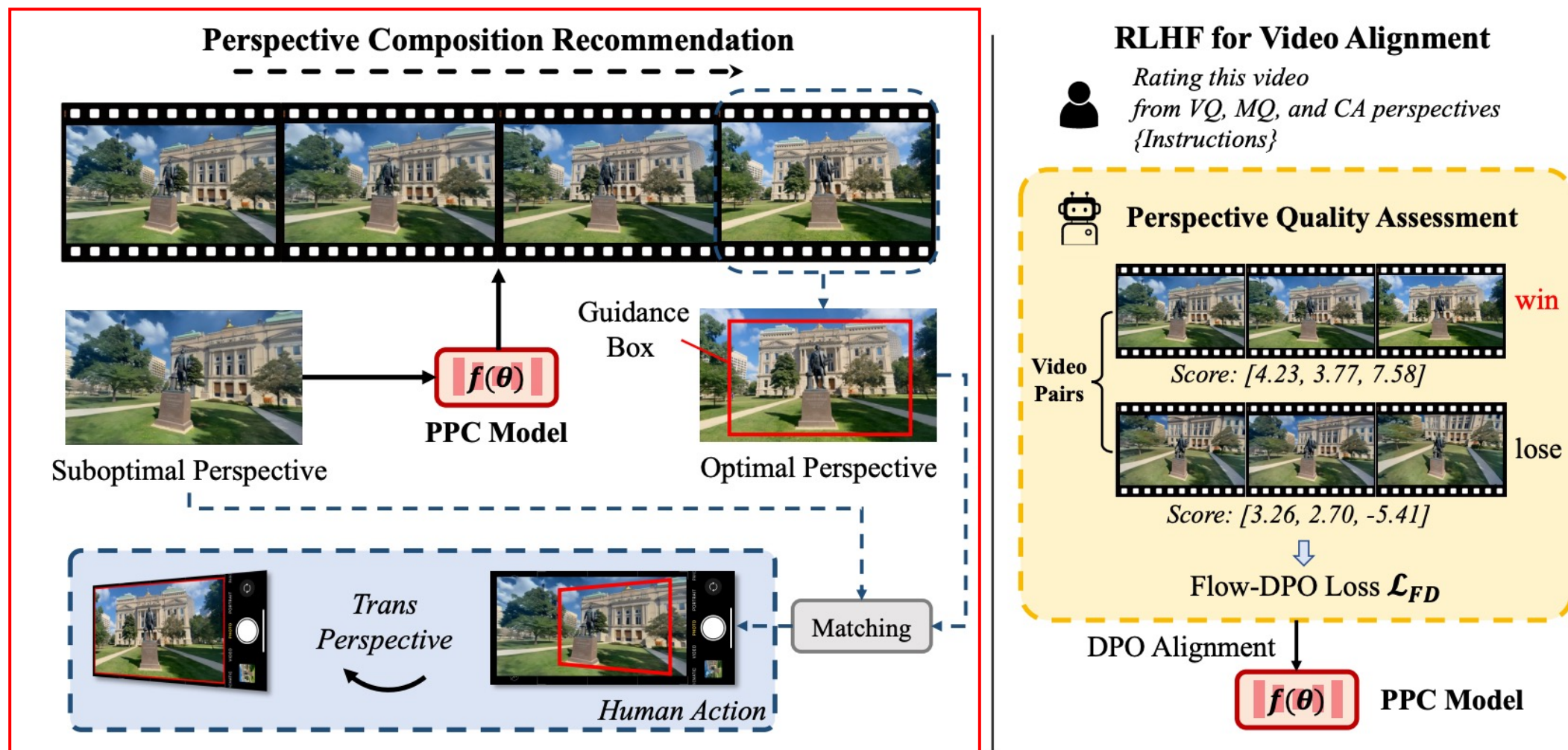


Figure 3: The pipeline of proposed photography perspective composition (PPC).

Method

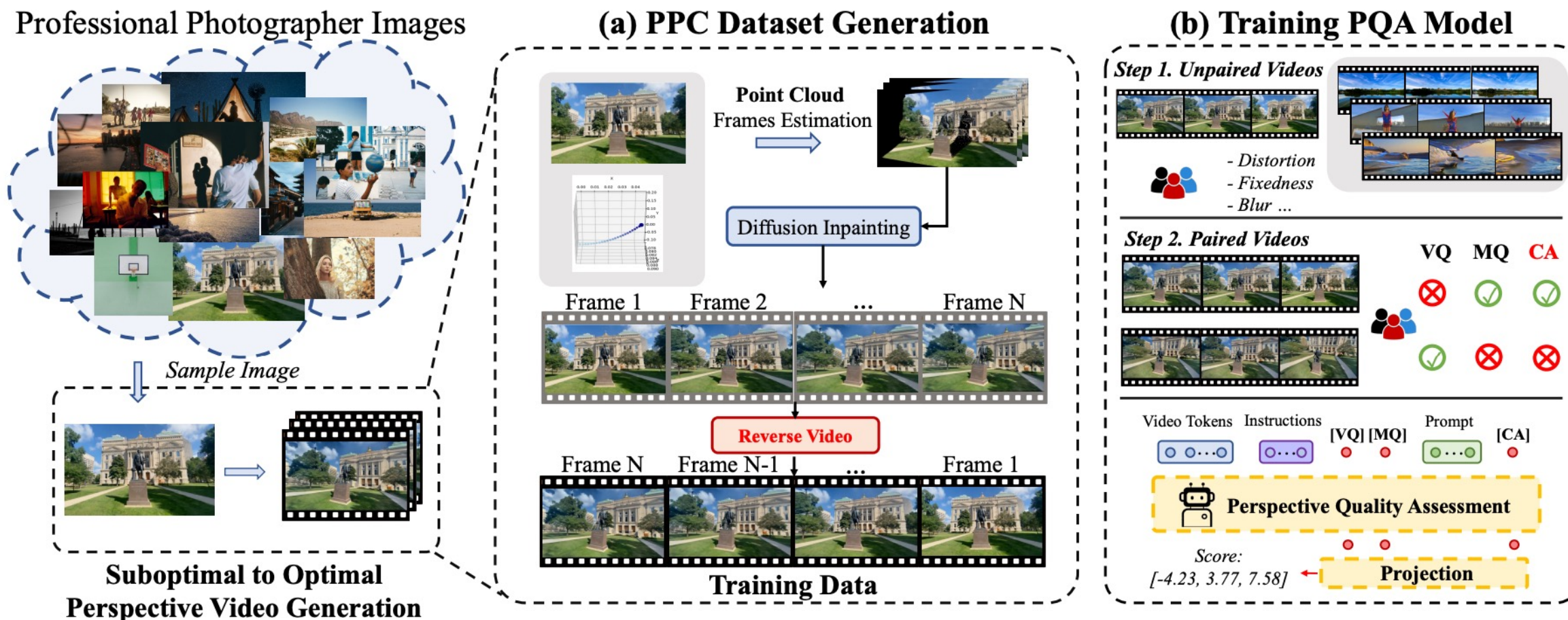


Figure 2: Architecture illustration of PPC dataset generation and the training perspective quality assessment (PQA) model.

Method

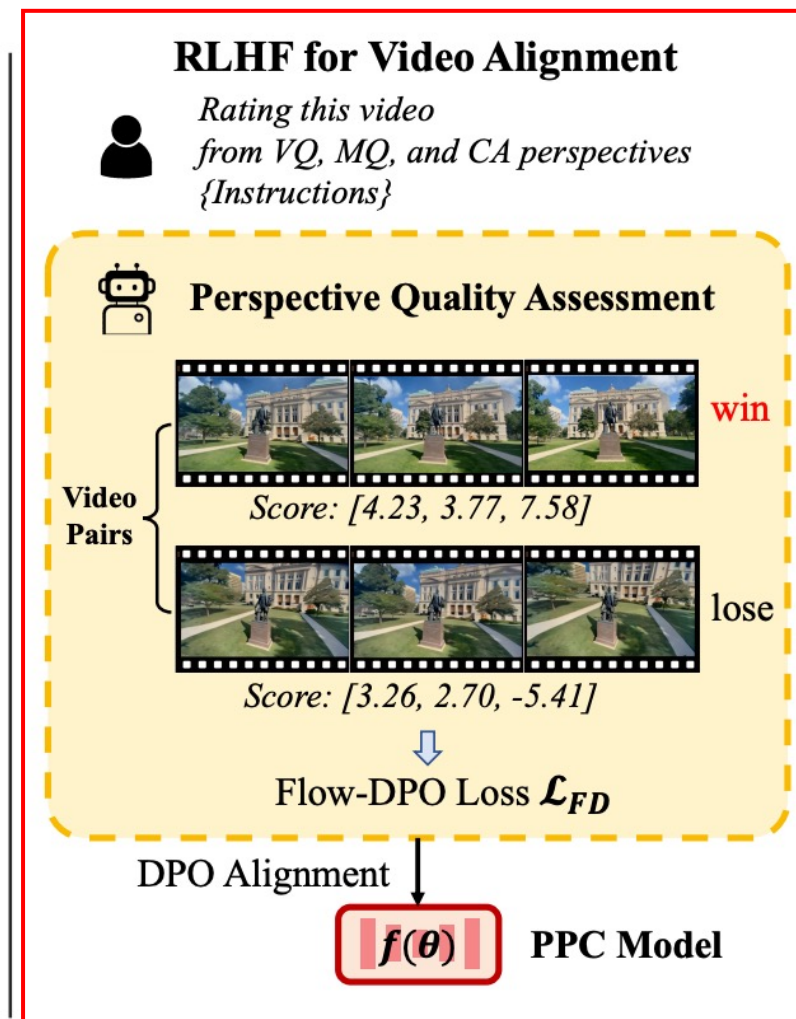
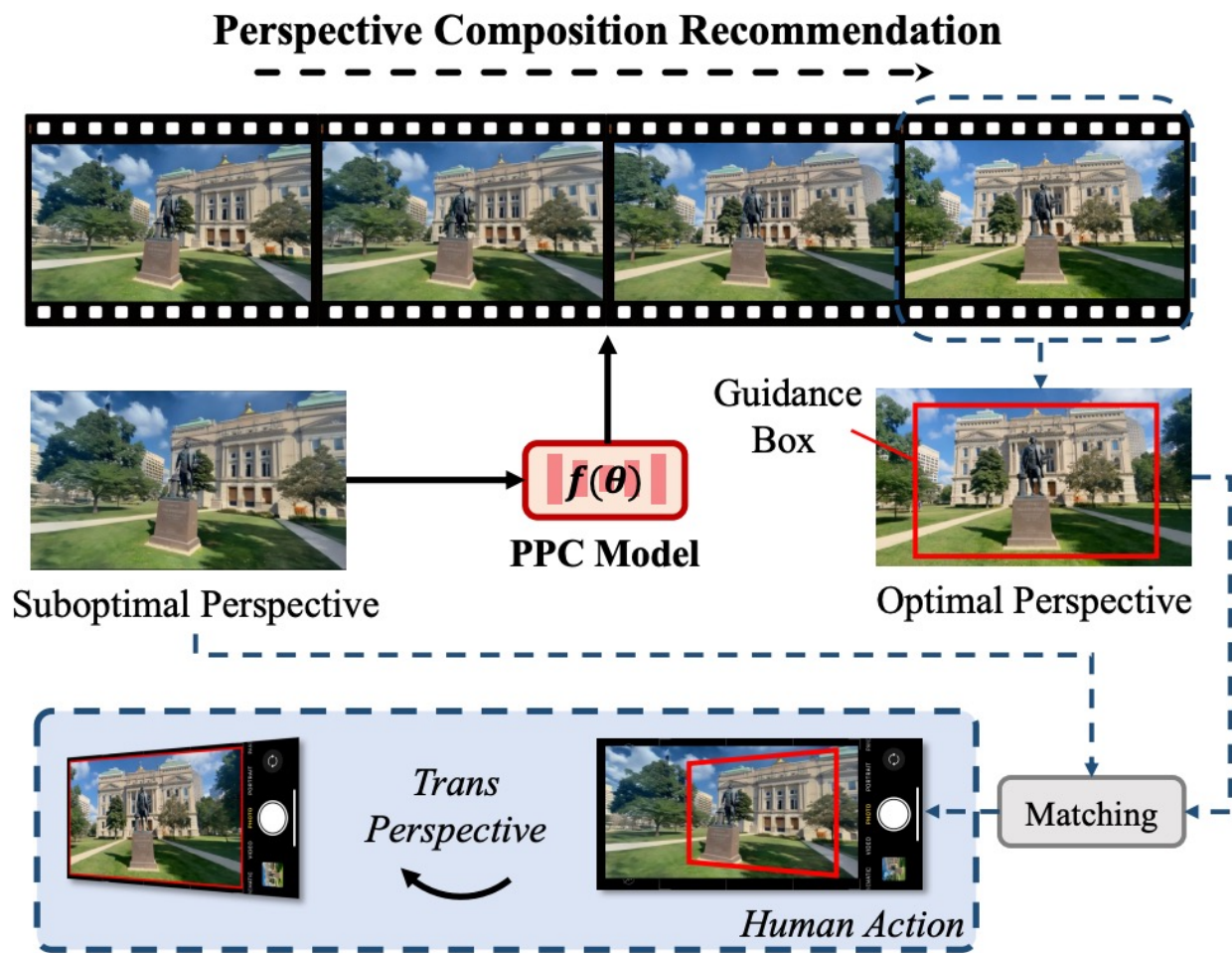


Figure 3: The pipeline of proposed photography perspective composition (PPC).

Results

Table 1: Comparison of I2V models in generating perspective transformation videos.

Method	Perspective Accuracy					Human Performance Score		
	CMM \uparrow	FVD \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	VQ \uparrow	MQ \uparrow	CA \uparrow
CogvideoX 1.5 5B [57]	0.5501	303	8.2380	0.2611	0.7969	0.7073	0.7311	0.7196
Hunyuan I2V [19]	0.4928	264	9.4017	0.3537	0.7915	0.7216	0.7496	0.7070
Wan2.1 14B [46]	0.5989	345	9.3668	0.3265	0.7808	0.7195	0.7454	0.7072

Method	Video Quality							
	I2V Subject	I2V Background	Subject Consistency	Background Consistency	Motion Smoothness	Dynamic Degree	Aesthetic Quality	Imaging Quality
CogvideoX 1.5 5B [57]	0.9632	0.9639	0.9545	0.9502	0.9906	0.1347	0.5314	0.5667
Hunyuan I2V [19]	0.9866	0.9878	0.9582	0.9663	0.9927	0.7851	0.5583	0.5893
Wan2.1 14B [46]	0.9618	0.9694	0.9470	0.9435	0.9917	0.8883	0.5464	0.6299

(a) The number of pairs.

# Pairs	Accuracy		
	VQ	MQ	CA
1	0.6515	0.5957	0.5881
5	0.7772	0.7812	0.7894
10	0.8019	0.8085	0.8102
100	0.8008	0.8063	0.8103

(b) Different steps.

# Steps	Accuracy		
	VQ	MQ	CA
Reg Single	0.4874	0.4986	0.4761
Reg Two	0.7807	0.7802	0.7935
BTT Single	0.5509	0.5117	0.4913
BTT Two	0.8019	0.8085	0.8102

(c) The data ratios.

Ratio	CMM \uparrow	FVD \downarrow
20%	0.5014	460
40%	0.5989	345
80%	0.5244	362
100%	0.5673	359

(d) The data angles.

	CMM \uparrow	FVD \downarrow
10°	0.4413	397
20°	0.5587	337
30°	0.3983	444
Mix-up	0.5989	345

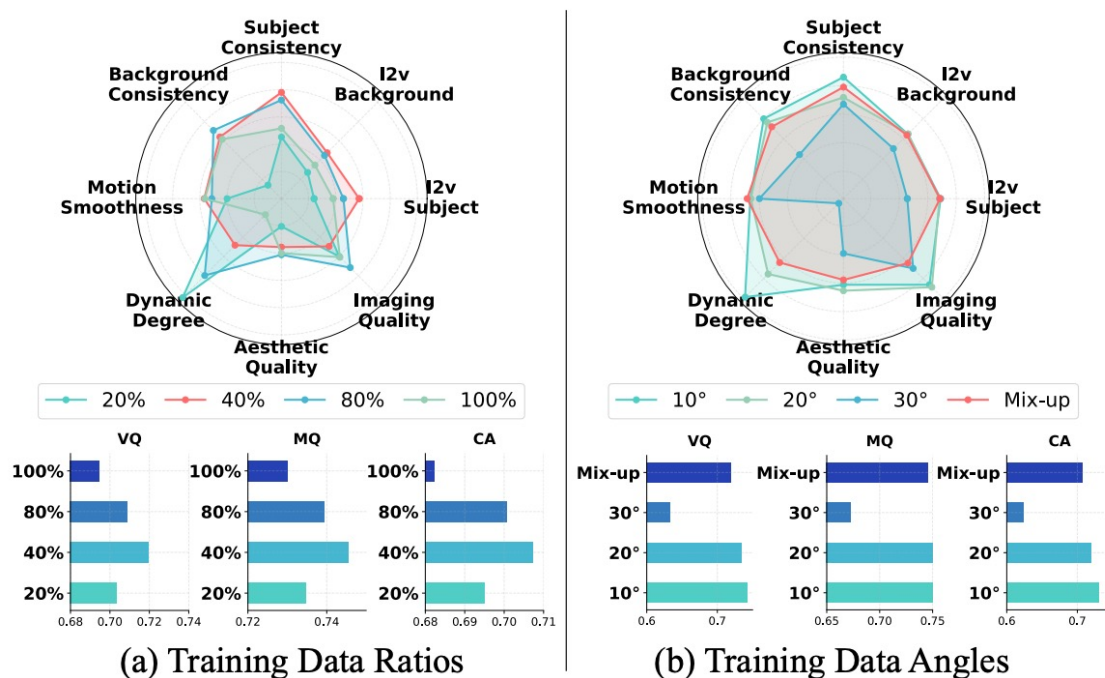


Figure 4: Quality and human performance results for PPC.

Results



Results



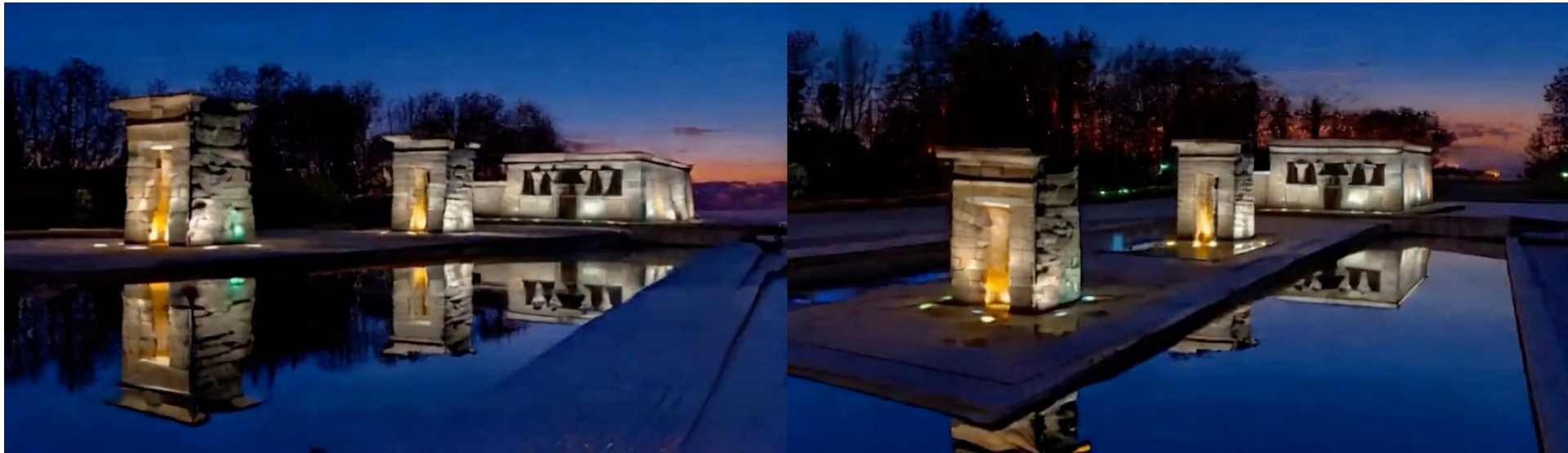
Results



Results



Results



Results



VQ: -0.009; MQ: 5.875; CA: 3.578



VQ: -7.281; MQ: -7.500; CA: -9.063



VQ: 5.531; MQ: 7.563; CA: -1.016



VQ: -6.656; MQ: -5.969; CA: -6.750



THANKS

Lujian Yao