中国科学院软件研究所学术年会
暨重点实验室科技活动周 2025 第十届

学术论文

SceneDiff: 使用扩散模型进行基于文本和草图的生成式场景级图像检索

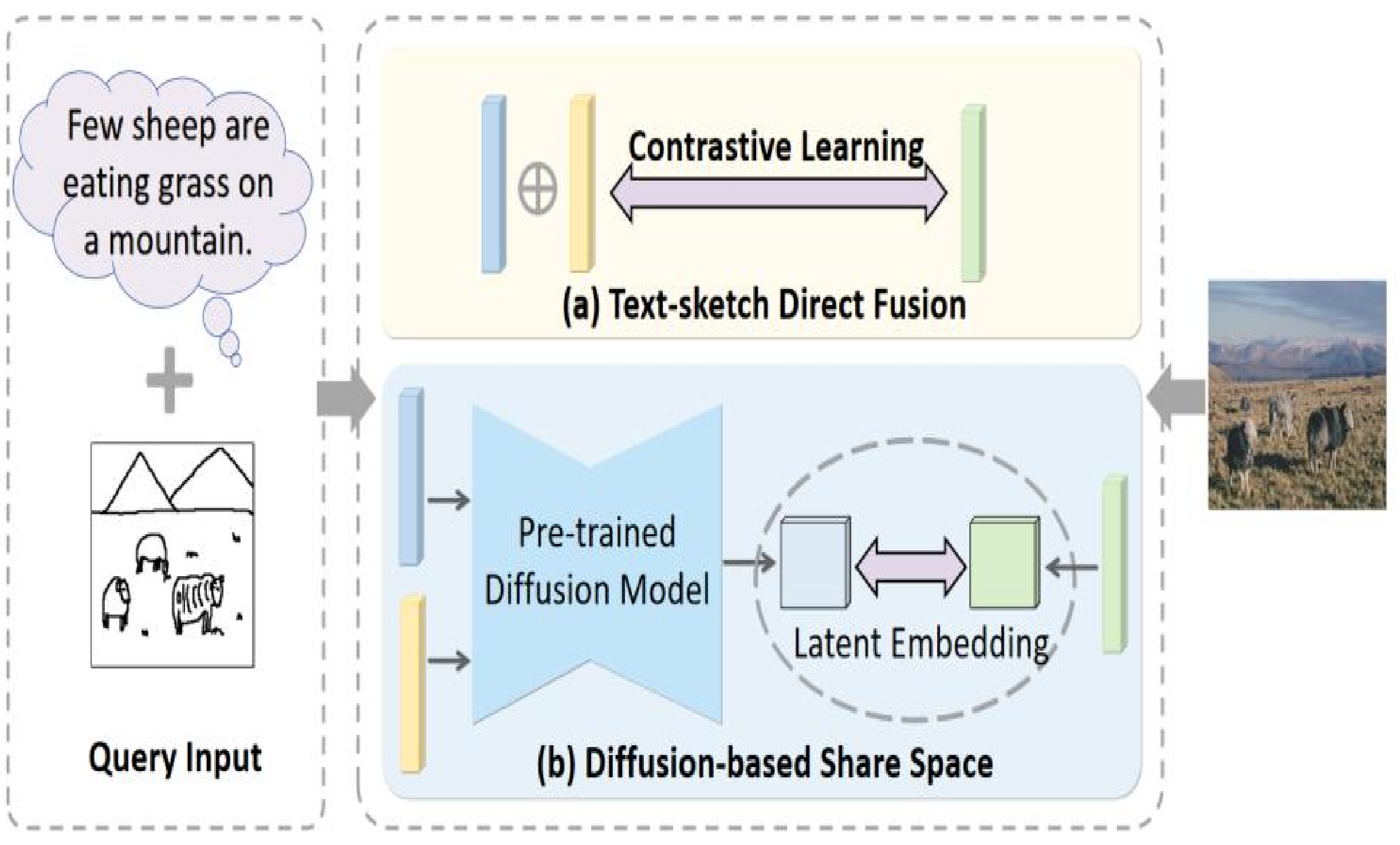
左然*, 胡皓翔*, 邓小明, 高沧骏, 张拯明, 来煜坤, 马翠霞, 刘永进, 王宏安

IJCAI2024 Pages 1825-1833

主要联系人：马翠霞 cuixia@iscas.ac.cn

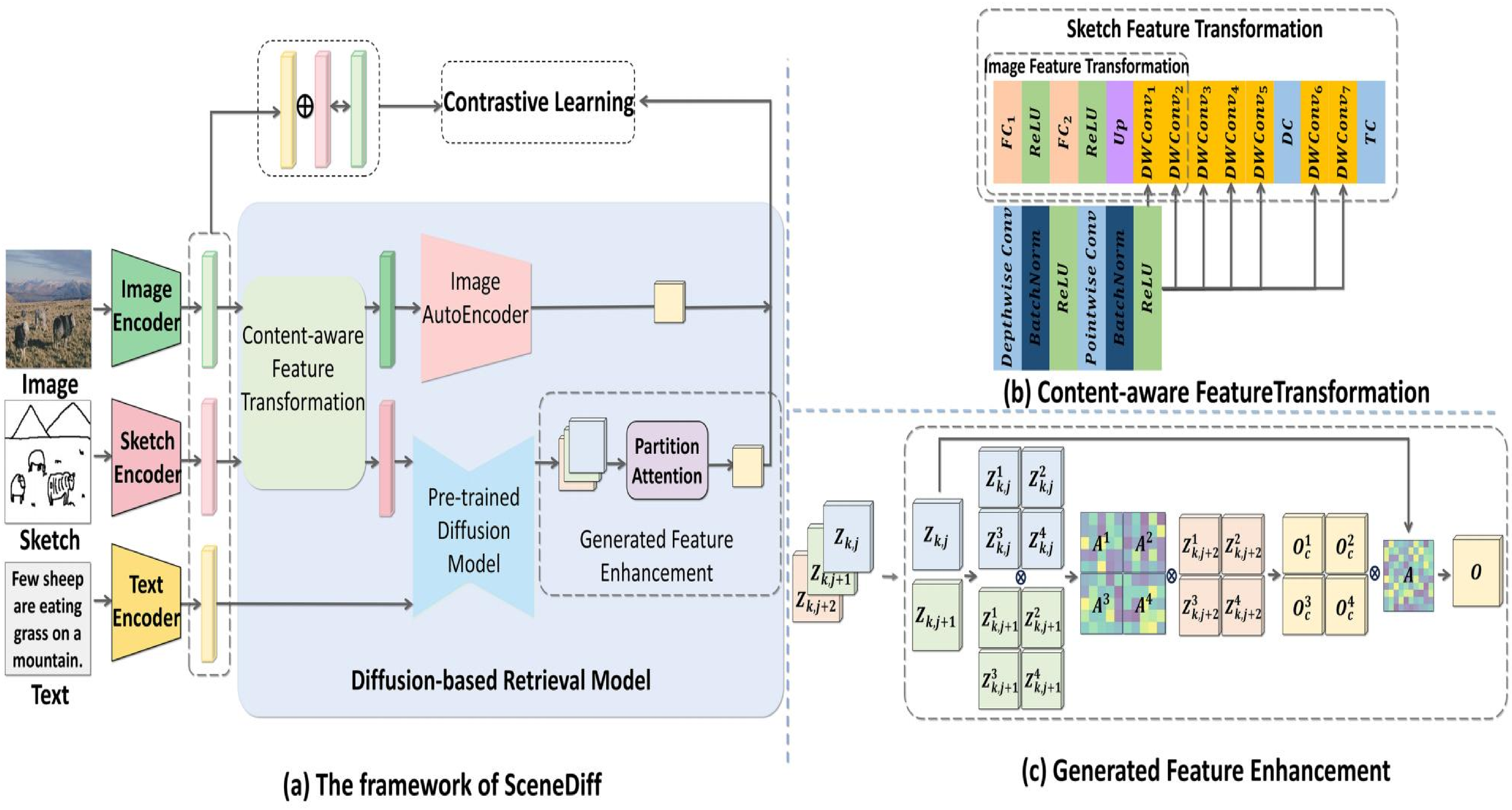
解决的科学问题

我们提出了一种用于场景级图像检索的生成式检索框架，该框架结合文本和草图输入，利用扩散模型优化文本与草图的特征融合，并在基于扩散的共享空间中实现与图像的对齐，实现高精度的基于草图和文本的图像检索。



创新点

- ●通过共享特征空间优化三种模态的特征建模。
- ●通过特征融合最大化草图和文本中包含的信息。



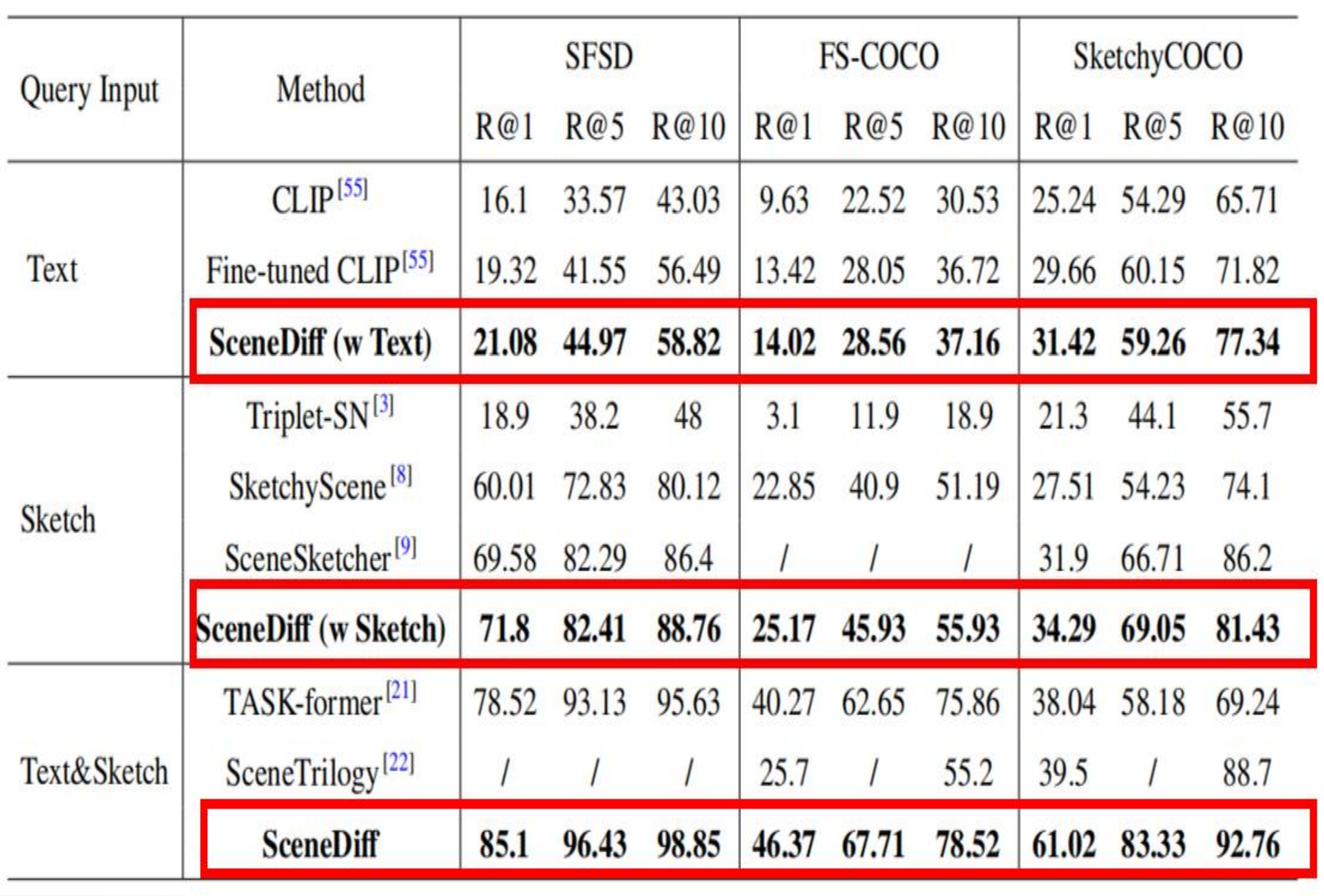
(a) The framework of SceneDiff

(b) Content-aware FeatureTransformation

(c) Generated Feature Enhancement

研究结果

我们的方法在准确率上显著优于现有的其他方法。



Query Input	Method	SFSD			FS-COCO			SketchyCOCO		
		R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
Text	CLIP[55]	16.1	33.57	43.03	9.63	22.52	30.53	25.24	54.29	65.71
	Fine-tuned CLIP[55]	19.32	41.55	56.49	13.42	28.05	36.72	29.66	60.15	71.82
	SceneDiff (w Text)	**21.08**	**44.97**	**58.82**	**14.02**	**28.56**	**37.16**	**31.42**	**59.26**	**77.34**
Sketch	Triplet-SN[3]	18.9	38.2	48	3.1	11.9	18.9	21.3	44.1	55.7
	SketchyScene[8]	60.01	72.83	80.12	22.85	40.9	51.19	27.51	54.23	74.1
	SceneSketcher[9]	69.58	82.29	86.4	/	/	/	31.9	66.71	86.2
	SceneDiff (w Sketch)	**71.8**	**82.41**	**88.76**	**25.17**	**45.93**	**55.93**	**34.29**	**69.05**	**81.43**
Text&Sketch	TASK-former[21]	78.52	93.13	95.63	40.27	62.65	75.86	38.04	58.18	69.24
	SceneTrilogy[22]	/	/	/	25.7	/	55.2	39.5	/	88.7
	SceneDiff	**85.1**	**96.43**	**98.85**	**46.37**	**67.71**	**78.52**	**61.02**	**83.33**	**92.76**

研究意义

在本文中，我们提出了一种创新的场景级 TSBIR框架，该框架利用预训练的 Stable Diffusion 模型，提升了草图、文本与图像数据之间的融合与对齐能力。该框架首先分别对草图、文本和图像特征进行编码，随后通过一个具备内容感知能力的特征变换模块，将这些特征投射到一个基于扩散的共享空间中。所提出的方法展现出在未来研究中可扩展至其他相似检索任务的潜力。