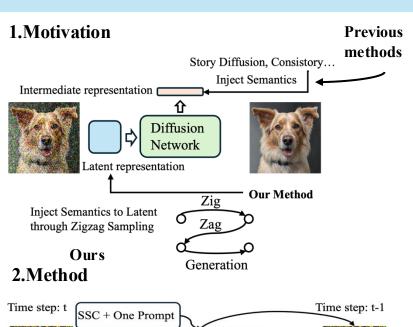


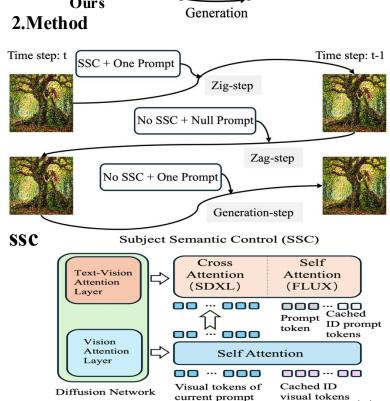


# Consistent Story Generation: Unlocking the Potential of Zigzag Sampling



Mingxiaio Li · Mang Ning · Marie-Francine Moens





#### Subject token Text-Vision Attention score Attention 00...00 Layer $[0.01, 0.23, \cdots 0.05, 0.2, 0.14]$ Vision ∏ Top-k Attention Layer Cached tokens

#### 3. Results

Method	Base Model	Train-Free	CLIP-T↑	CLIP-I↑	$DreamSim \downarrow$	Steps
Vanilla SDXL	-	-	0.9074	0.8165	0.5292	50
Vanilla FLUX	-	-	0.8977	0.8494	0.3888	28
The Chosen One	SDXL	Х	0.7614	0.7831	0.4929	35
PhotoMaker	SDXL	×	0.8651	0.8465	0.3996	50
IP-Adapter	SDXL	×	0.8458	0.9429	0.1462	30
ConsiStory	SDXL	1	0.8769	0.8737	0.3188	50
StoryDiffusion	SDXL	/	0.8877	0.8755	0.3212	50
1Prompt1Story	SDXL	/	0.8942	0.9117	0.1993	50
Ours	SDXL	✓	0.8946	0.923	0.1798	50
1Prompt1Story	FLUX		0.8716	0.9118	0.1957	28
Ours	FLUX	/	0.8949	0.9216	0.1843	28

### 4. Qualitative Comparisons



## 4.Long Story Visualisation

An anime-style illustration of a 16 years old girl with wavy chestnut hair, a slender frame, and soft brown eyes











































at a zoo, observing animals attending a costume party learning to cook in a kitchen



at a carnival, holding a