# VideoMAR: Autoregressive Video Generatio with Continuous Tokens

Hu Yu, Biao Gong, Hangjie Yuan, DanDan Zheng, Weilong Chai, Jingdong Chen, Kecheng Zheng, Feng Zhao

University of Science and Technology of China (USTC)





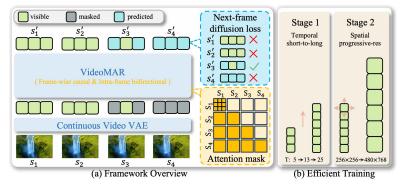
## **Abstract**

Masked-based autoregressive models have demonstrated promising image generation capability in continuous space. However, their potential for video generation remains under-explored. In this paper, we propose VideoMAR, a decoder-only autoregressive video generation model with continuous tokens, integrating temporal frame-by-frame and spatial masked generation. To meet the requirement of sequentially generating each frame depending on all the previous context frames, VideoMAR preserves the complete context and introduces a next-frame diffusion loss during training. Besides, the extremely long token sequences of video data poses significant challenges in both efficiency and difficulty. To this end, we propose tailored strategies for training and inference. During training, we propose the short-to-long curriculum learning to reduce the training difficulty and cost, and establish the two-stage progressive-resolution training to support higher resolution video generation. During inference, long token sequence generation is prone to suffer from severe accumulation error in late frames, due to exposure bias issue. We identify that temperature plays a crucial role to eliminate this error and propose the progressive temperature strategy.

Furthermore, VideoMAR replicates several unique capacities of language models to video generation, e.g. key-value cache and extrapolation, demonstrating the potential for multi-modal unification. For example, thanks to our design, VideoMAR inherently bears high efficiency due to simultaneous temporal-wise KV cache and spatial-wise parallel generation. VideoMAR, for the first time, also unlocks the capacity of simultaneous spatial and temporal extrapolation for video generation via incorporating the 3D-RoPE. On the VBench-I2V benchmark, VideoMAR achieves better performance compared to the Cosmos baseline, with much smaller model size, data scale, and GPU resources.

# Method

#### 1. Framework and efficient training



### 2. Efficient inference

Metho	ds	Steps (spatial × temporal)	Inference time (s)		
VideoMAR(stage1)	NTP*	1024x6	1310		
	w/o KV Cache	64x6	417		
	w/ KV Cache	64x6	107		
VideoMAR(stage2)	NTP*	1440x6	1941		
	w/o KV Cache	64x6	671		
	w/ KV Cache	64x6	134		

### 3. Zero-shot resolution scaling

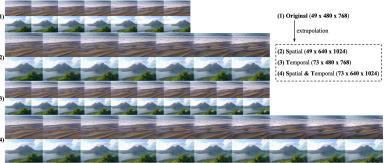


Figure 4: Spatial and temporal extrapolation capacity of VideoMAR.

# **Experiments**

### 1. Quantitative result

VideoMAR achieves sota performance on the Vbench-I2V with significantly fewer parameters, training data, and GPU resources.

norome	data	Total	I2V	Qual.	I2V	I2V	Came.				Dyna.	Aest.	Imag.
Model params da	data	Score	Score	Score	Subj.	Back.	Moti.	Cons.	Cons.	Smoo.	Degr.	Qual.	Qual.
24B	-	89.28	96.12	82.44	98.39	99.00	50.85	93.96	96.74	98.68	68.21	64.74	69.71
30B	5M	88.36	95.50	81.22	97.86	98.63	49.23	96.02	97.06	99.24	48.78	62.29	70.44
5B	35M	86.70	94.79	78.61	97.19	96.74	67.68	94.34	96.42	98.40	33.17	61.87	70.01
-	10M+	85.52	92.67	78.37	97.15	96.94	20.97	95.28	97.12	97.12	27.07	64.55	71.39
-	35M	85.28	92.11	78.44	96.48	96.83	18.48	94.18	97.09	98.34	26.10	64.82	69.14
-	10M	84.07	91.91	76.22	95.82	95.95	33.92	95.27	98.28	97.38	18.62	59.00	66.92
-	10M	82.57	86.31	78.84	91.17	91.31	33.60	97.86	98.79	98.00	22.60	60.78	71.68
els													
5B	100M	84.16	92.51	75.81	95.99	97.36	25.56	97.12	96.59	99.47	20.33	55.82	59.90
13B	100M	84.22	92.60	75.83	96.17	97.35	25.43	97.69	96.77	99.40	18.70	55.84	60.15
1.4B	0.2M	82.56	91.74	73.39	96.64	96.24	16.80	97.08	98.78	99.32	13.01	52.05	51.61
1.4B	0.5M	84.82	94.02	75.61	97.85	98.38	21.62	97.13	97.20	99.57	10.98	55.81	62.34
	30B 5B - - - - - - els 5B 13B	24B - 30B 5M 5B 35M - 10M+ - 35M - 10M - 10M 5B 100M 13B 100M	params         data score           24B         -         89.28           30B         5M         88.36           5B         35M         86.70           -         10M+         85.22           -         10M         84.07           -         10M         82.57           els           5B         100M         84.16           13B         100M         84.22           1.4B         0.2M         82.56	params         data         Score         Score           24B         -         89.28         96.12           30B         5M         88.36         95.50           5B         35M         86.70         94.79           -         10M+         85.52         92.67           -         35M         85.28         92.11           -         10M         84.07         91.91           -         10M         82.57         86.31           els           5B         100M         84.16         92.51           13B         100M         84.22         92.60           1.4B         0.2M         82.56         91.74	params         data         Score         Score         Score           24B         -         89.28         96.12         82.44           30B         5M         88.36         95.50         81.22           5B         35M         86.70         94.79         78.61           -         10M+         85.52         92.67         78.37           -         35M         85.28         92.11         78.44           -         10M         84.07         91.91         76.24           -         10M         82.57         86.31         78.84           els           5B         100M         84.16         92.51         75.81           13B         100M         84.22         92.60         75.83           1.4B         0.2M         82.56         91.74         73.39	params         data         Score         Score         Score         Score         Subj.           24B         -         89.28         96.12         82.44         98.39           30B         5M         88.30         95.50         81.22         97.86           5B         35M         86.70         94.79         78.61         97.19           -         10M+         85.52         92.67         78.37         97.15           -         35M         85.28         22.11         78.44         96.48           -         10M         84.07         91.91         76.22         95.82           -         10M         82.57         86.31         78.84         91.17           els           5B         100M         84.16         92.51         75.81         95.99           13B         100M         84.12         92.60         75.83         96.17           14B         0.2M         82.56         91.74         73.39         96.64	params         data         Score         Score         Score         Subj.         Back.           24B         -         89.28         96.12         82.44         98.39         90.00           30B         5M         88.36         95.50         81.22         97.86         98.63           5B         35M         86.70         94.79         78.61         97.19         96.74           -         10M+         85.52         92.67         78.37         97.15         96.94           -         35M         85.28         92.11         78.49         96.48         96.83           -         10M         84.07         91.91         76.22         95.82         95.95           els           5B         100M         84.16         92.51         75.81         95.99         97.36           13B         100M         84.26         92.50         75.83         96.17         97.35           1.4B         0.2M         82.56         91.74         73.39         96.64         96.24	params         data         Score         Score         Sup.         Back.         Moti.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23           5B         35M         86.70         94.79         78.61         97.19         96.74         67.68           -         10M+         85.52         92.67         78.37         97.15         96.94         20.97           -         35M         85.28         92.67         78.37         97.15         96.94         20.97           -         10M         84.07         91.91         76.22         95.82         95.95         33.92           -         10M         82.57         86.31         78.84         91.17         91.31         33.60           els           5B         100M         84.12         92.60         75.81         95.99         97.36         25.54           13B         100M         84.22         92.60         75.83         96.17         97.35         25.43           14B         0.	params         data         Score         Score         Score         Subj.         Back.         Moti.         Cons.           24B         -         89.28         86.12         82.44         98.39         90.00         50.85         93.96           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02           5B         35M         86.70         94.79         78.61         97.15         96.94         20.97         95.28           -         10M         85.25         26.77         78.37         97.15         96.94         20.97         95.28           -         35M         84.07         91.91         76.22         95.82         95.95         33.92         95.27           -         10M         84.07         91.91         76.22         95.82         95.95         33.92         95.28           els         -         10M         82.57         86.31         78.84         91.17         91.31         33.60         97.86           els         -         -         95.99         97.36         25.56         97.12           13B         100M         84.22 <t< td=""><td>params         data         Score         Score         Subj.         Back.         Moti.         Cons.         Cons.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.96         96.74           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06           5B         35M         86.70         94.79         78.37         97.15         96.94         20.97         95.28         97.16           -         10M4         88.52         92.67         78.37         97.15         96.94         20.97         95.28         97.09           -         35M         86.70         91.91         76.22         95.82         95.55         33.29         95.27         98.28           -         10M4         84.07         91.91         76.22         95.82         95.95         33.92         95.27         98.28           -         10M         82.57         86.31         78.84         91.17         91.31         33.60         97.80         98.79           els           5B         100M         84.16</td><td>params         data         Score         Score         Subj.         Back.         Moti.         Cons.         Cons.         Cons.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.90         96.02         97.06         98.68           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         99.24           5B         35M         86.70         94.79         78.61         97.19         96.74         67.68         94.34         96.42         98.00           -         10M4         85.52         92.67         78.37         97.15         96.94         20.97         95.28         97.12         97.12           -         35M         85.28         92.17         78.37         97.15         96.94         20.97         95.28         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         98.74         97.30         95.27         98.29         97.30         95.27         98.29         97.30</td><td>params         data         Score         Score         Subj.         Back.         Moti.         Cons.         Cons.         Smoo.         Degr.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.90         96.74         98.68         88.21           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         92.24         48.78           5B         35M         86.70         94.79         78.37         97.15         96.94         20.97         95.28         97.12         97.12         27.07           -         35M         85.22         92.67         78.37         97.15         96.94         20.97         95.28         97.12         97.12         27.07           -         35M         85.28         92.17         78.49         96.48         96.83         18.48         91.48         97.09         98.24         26.01           -         10M         84.07         91.91         76.22         95.82         95.95         33.92         95.27         98.28         97.38         18.62           els     <!--</td--><td>params         data         Score         Score         Subj.         Back.         Mot.         Cons.         Cons.         Smoo.         Degr.         Qual.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.96         96.74         96.68         68.21         64.74           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         99.24         48.78         62.29           5B         35M         86.70         94.79         78.61         97.19         96.74         67.68         94.34         96.02         99.04         48.78         62.29           -         10M         85.25         92.67         78.37         97.15         96.94         20.97         95.28         97.12         27.07         64.55           -         35M         85.28         92.17         78.47         96.48         96.81         48.48         94.18         97.09         98.24         26.10         64.85           -         10M         82.57         86.31         78.49         91.17         91.31         33.00         97.80</td></td></t<>	params         data         Score         Score         Subj.         Back.         Moti.         Cons.         Cons.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.96         96.74           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06           5B         35M         86.70         94.79         78.37         97.15         96.94         20.97         95.28         97.16           -         10M4         88.52         92.67         78.37         97.15         96.94         20.97         95.28         97.09           -         35M         86.70         91.91         76.22         95.82         95.55         33.29         95.27         98.28           -         10M4         84.07         91.91         76.22         95.82         95.95         33.92         95.27         98.28           -         10M         82.57         86.31         78.84         91.17         91.31         33.60         97.80         98.79           els           5B         100M         84.16	params         data         Score         Score         Subj.         Back.         Moti.         Cons.         Cons.         Cons.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.90         96.02         97.06         98.68           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         99.24           5B         35M         86.70         94.79         78.61         97.19         96.74         67.68         94.34         96.42         98.00           -         10M4         85.52         92.67         78.37         97.15         96.94         20.97         95.28         97.12         97.12           -         35M         85.28         92.17         78.37         97.15         96.94         20.97         95.28         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         97.12         98.74         97.30         95.27         98.29         97.30         95.27         98.29         97.30	params         data         Score         Score         Subj.         Back.         Moti.         Cons.         Cons.         Smoo.         Degr.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.90         96.74         98.68         88.21           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         92.24         48.78           5B         35M         86.70         94.79         78.37         97.15         96.94         20.97         95.28         97.12         97.12         27.07           -         35M         85.22         92.67         78.37         97.15         96.94         20.97         95.28         97.12         97.12         27.07           -         35M         85.28         92.17         78.49         96.48         96.83         18.48         91.48         97.09         98.24         26.01           -         10M         84.07         91.91         76.22         95.82         95.95         33.92         95.27         98.28         97.38         18.62           els </td <td>params         data         Score         Score         Subj.         Back.         Mot.         Cons.         Cons.         Smoo.         Degr.         Qual.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.96         96.74         96.68         68.21         64.74           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         99.24         48.78         62.29           5B         35M         86.70         94.79         78.61         97.19         96.74         67.68         94.34         96.02         99.04         48.78         62.29           -         10M         85.25         92.67         78.37         97.15         96.94         20.97         95.28         97.12         27.07         64.55           -         35M         85.28         92.17         78.47         96.48         96.81         48.48         94.18         97.09         98.24         26.10         64.85           -         10M         82.57         86.31         78.49         91.17         91.31         33.00         97.80</td>	params         data         Score         Score         Subj.         Back.         Mot.         Cons.         Cons.         Smoo.         Degr.         Qual.           24B         -         89.28         96.12         82.44         98.39         99.00         50.85         93.96         96.74         96.68         68.21         64.74           30B         5M         88.36         95.50         81.22         97.86         98.63         49.23         96.02         97.06         99.24         48.78         62.29           5B         35M         86.70         94.79         78.61         97.19         96.74         67.68         94.34         96.02         99.04         48.78         62.29           -         10M         85.25         92.67         78.37         97.15         96.94         20.97         95.28         97.12         27.07         64.55           -         35M         85.28         92.17         78.47         96.48         96.81         48.48         94.18         97.09         98.24         26.10         64.85           -         10M         82.57         86.31         78.49         91.17         91.31         33.00         97.80

#### 2. Qualitative result

VideoMAR achieves better quality and finer details than the Cosmos baseline even with a lower spatial resolution.

