	R Precision (top 1) ↑	R Precision (top 2) ↑	R Precision (top 3) ↑	FID ↓	Diversity →	Multimodality ↑
Real	0.4541 ± 0.0060	0.6612 ± 0.0065	0.7704 ± 0.0054	0.0010 ± 0.0001	9.1322 ± 0.0943	–
MDM-Original	0.3883 ± 0.0158	0.5828 ± 0.0091	0.6945 ± 0.0026	0.5243 ± 0.1885	9.3325 ± 0.0594	2.6270 ± 0.0539
MDM-1	$\mathbf{0.4495 \pm 0.0035}$	$\mathbf{0.6576 \pm 0.0060}$	$\mathbf{0.7681 \pm 0.0053}$	$\mathbf{0.6466 \pm 0.0774}$	$\mathbf{9.7104 \pm 0.1696}$	2.5973 ± 0.1703
MDM-2	$\mathbf{0.4546 \pm 0.0029}$	$\mathbf{0.6606 \pm 0.0033}$	$\mathbf{0.7681 \pm 0.0040}$	$\mathbf{0.6042 \pm 0.0868}$	9.0701 ± 0.1454	2.4932 ± 0.0947
MDM-3	$\mathbf{0.3979 \pm 0.0127}$	$\mathbf{0.5877 \pm 0.0123}$	$\mathbf{0.6992 \pm 0.0052}$	$\mathbf{0.5602 \pm 0.0817}$	9.3052 ± 0.1416	2.5170 ± 0.0678

1. DiT Block with adaLN-Zero
2. DiT Block with Cross-Attention
3. DiT Block with In-Context Conditioning

	R Precision (top 1) ↑	R Precision (top 2) ↑	R Precision (top 3) ↑	FID ↓	Diversity →	Multimodality ↑
SWA-1	0.4102 ± 0.0081	0.6059 ± 0.0083	0.7205 ± 0.0087	1.1091 ± 0.1846	9.3799 ± 0.1220	2.407 ± 0.0826
SWA-2	0.4100 ± 0.0135	0.6146 ± 0.0077	0.7254 ± 0.0055	0.7480 ± 0.1373	9.5032 ± 0.1696	2.2632 ± 0.0623
SWA-3	0.4869 ± 0.0070	0.6805 ± 0.0080	0.7947 ± 0.0076	0.6248 ± 0.0621	9.6316 ± 0.1798	2.6038 ± 0.1162
SWA-4	0.4057 ± 0.0063	0.6029 ± 0.0070	0.7211 ± 0.0037	0.9196 ± 0.1833	9.3885 ± 0.1943	2.4265 ± 0.1204
NA-16	0.3732 ± 0.0121	0.5723 ± 0.0093	0.6869 ± 0.0075	1.1105 ± 0.1239	9.1334 ± 0.1387	2.6973 ± 0.0513
NA-32	0.3857 ± 0.0121	0.5814 ± 0.0072	0.6973 ± 0.0071	0.6952 ± 0.0600	9.2068 ± 0.1127	2.5957 ± 0.0464

SWA = Sliding Window Attention

NA = Neighborhood Attention

1. Window size = 4
2. Window size = 8
3. Window size = 12
4. Window size = 16

	R Precision (top 1) ↑	R Precision (top 2) ↑	R Precision (top 3) ↑	FID ↓	Diversity →	Multimodality ↑
SWA-4	0.4057 ± 0.0063	0.6029 ± 0.0070	0.7211 ± 0.0037	0.9196 ± 0.1833	9.3885 ± 0.1943	2.4265 ± 0.1204
tubelet_p4o0	0.4121 ± 0.0158	0.6109 ± 0.0205	0.7240 ± 0.0161	0.6275 ± 0.0840	9.5540 ± 0.1772	2.2371 ± 0.0495
tubelet_p8o4	0.4064 ± 0.0112	0.6092 ± 0.0089	0.7209 ± 0.0054	0.8067 ± 0.0866	9.2741 ± 0.2226	2.4800 ± 0.0705
convk5s2	0.4111 ± 0.0147	0.6086 ± 0.0133	0.7232 ± 0.0079	0.9910 ± 0.1098	9.2075 ± 0.1155	2.4582 ± 0.1058
convk7s3	0.4062 ± 0.0064	0.6041 ± 0.0181	0.7166 ± 0.0137	0.9615 ± 0.1339	9.3084 ± 0.1009	2.4867 ± 0.1116

Window size = 16