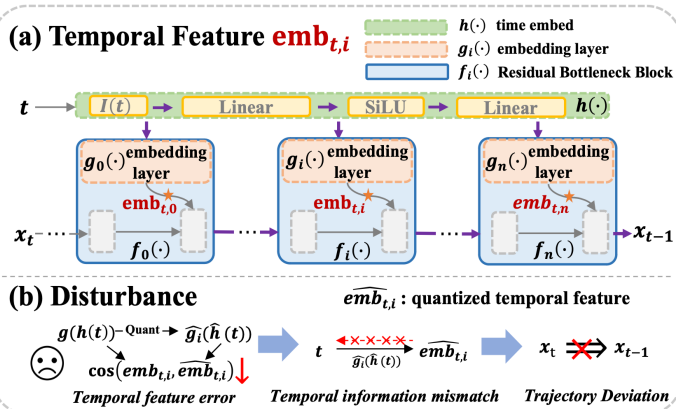
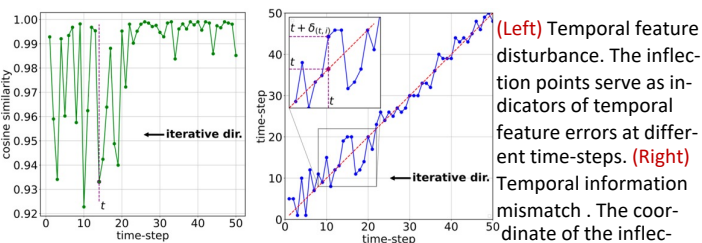


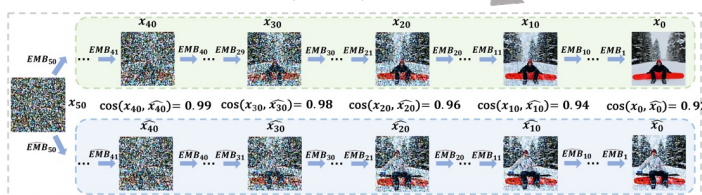
## Temporal Feature Disturbance



1. Temporal Feature Error:  $\cos(emb_{t,i}, emb_{t,i}^q)$ .
2. Temporal Information Mismatch:  $t \leftrightarrow emb_{t,i}^q$ .



3. Trajectory Deviation:  $x_t \neq x_{t-1}$ .



Code is now available at:

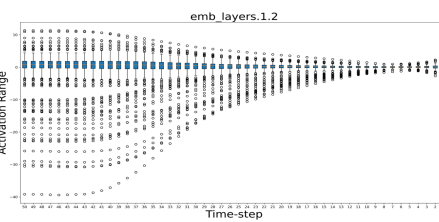


## Inducement Analyses

1. Inappropriate Reconstruction Target: Wrong objective and overfit.

Methods	Bits (W/A)	FID↓	sFID↓
Full Prec.	32/32	2.98	7.09
Prev.	8/8	7.51	12.54
Freeze	8/8	<b>5.76 (-1.75)</b>	<b>8.42 (-4.12)</b>
Prev.	4/8	9.36	22.73
Freeze	4/8	<b>7.08 (-2.28)</b>	<b>16.82 (-5.91)</b>

“Freeze” means we freeze the  $t$ -related parts, when quantizing ResBlock and utilize naïve strategy for that component.



Activation ranges from sampling data-unrelated components.

2. Unaware of Finite Activations: Previous methods mainly focus on distribution and they ignore the range variations of finite activations across different  $t$ .

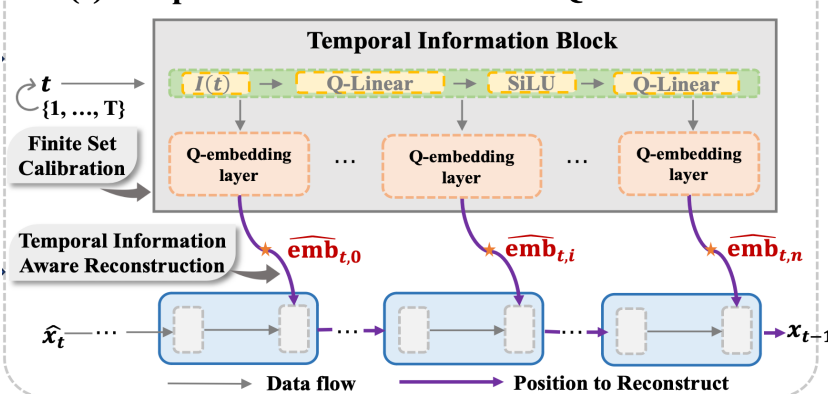
## Temporal Feature Maintenance

1. Temporal Information Block:  $\{g_i(h(\cdot))\}_{i=0,\dots,n}$ ;
2. Temporal Information Aware Reconstruction:

$$\mathcal{L}_{TIAR} = \sum_{i=0}^n \left\| g_i(h(t)) - \hat{g}_i(\hat{h}(t)) \right\|_F^2;$$

3. Finite Set Calibration:  $\hat{x} = \Phi \left( \frac{x}{s_t} + z_t, 0, 2^b - 1 \right)$ .

## (c) Temporal Feature Maintenance Quantization



## Experimental Results

1. Unconditional Image Generation

Methods	Bits (W/A)	LSUN-Bedrooms 256 × 256		LSUN-Churches 256 × 256		CelebA-HQ 256 × 256		FFHQ 256 × 256	
		FID↓	sFID↓	FID↓	sFID↓	FID↓	sFID↓	FID↓	sFID↓
Full Prec.	32/32	2.98	7.09	4.12	10.89	8.74	10.16	9.36	8.67
PTQ4DM*	4/32	4.83	7.94	4.92	13.94	13.67	14.72	11.74	12.18
Q-Diffusion†	4/32	4.20	7.66	4.55	11.90	11.09	12.00	11.60	10.30
PTQD*	4/32	4.42	7.88	4.67	13.68	11.06	12.21	12.01	11.12
TFMQ-DM (Ours)	4/32	<b>3.60 (-0.60)</b>	<b>7.61 (-0.05)</b>	<b>4.07 (-0.48)</b>	<b>11.41 (-0.49)</b>	<b>8.74 (-2.32)</b>	<b>10.18 (-1.82)</b>	<b>9.89 (-1.71)</b>	<b>9.06 (-1.24)</b>
PTQ4DM*	8/8	4.75	9.59	4.80	13.48	14.42	15.06	10.73	11.65
Q-Diffusion†	8/8	4.51	8.17	4.41	12.23	12.85	14.16	10.87	10.01
PTQD	8/8	3.75	9.89	4.89*	14.89*	12.76*	13.54*	10.69*	10.97*
TFMQ-DM (Ours)	8/8	<b>3.14 (-0.61)</b>	<b>7.26 (-0.91)</b>	<b>4.01 (-0.40)</b>	<b>10.98 (-1.25)</b>	<b>8.71 (-4.05)</b>	<b>10.20 (-3.34)</b>	<b>9.46 (-1.23)</b>	<b>8.73 (-1.28)</b>
PTQ4DM	4/8	20.72	54.30	4.97*	14.87*	17.08*	17.48*	11.83*	12.91*
Q-Diffusion†	4/8	6.40	17.93	4.66	13.94	15.55	16.86	11.45	11.15
PTQD	4/8	5.94	15.16	5.10*	13.23*	15.47*	17.38*	11.42*	11.43*
TFMQ-DM (Ours)	4/8	<b>3.68 (-2.26)</b>	<b>7.65 (-7.51)</b>	<b>4.14 (-0.52)</b>	<b>11.46 (-1.77)</b>	<b>8.76 (-6.71)</b>	<b>10.26 (-6.60)</b>	<b>9.97 (-1.45)</b>	<b>9.14 (-2.01)</b>

2. Class-conditional & Text-guided Image Generation

Methods	Bits (W/A)	ImageNet 256 × 256			MS-COCO			
		IS†	FID↓	sFID↓	FID↓	sFID↓	CLIP†	
Full Prec.	32/32	235.64	10.91	7.67	13.15	19.31	0.3146	
PTQ4DM	4/32	-	-	-	4/32	13.58	19.50	0.3143
Q-Diffusion*	4/32	213.56	11.87	8.76	4/32	<b>13.21 (-0.37)</b>	<b>19.03 (-0.47)</b>	<b>0.3144 (+0.0001)</b>
PTQD†	4/32	201.78	11.65	9.06	8/8	13.31	20.54	0.3134
TFMQ-DM (Ours)	4/32	<b>223.81 (+10.25)</b>	<b>10.50 (-1.15)</b>	<b>7.98 (-0.78)</b>	8/8	<b>13.09 (-0.22)</b>	<b>19.91 (-0.63)</b>	<b>0.3134 (+0.0000)</b>
PTQ4DM	8/8	161.75	12.59	-	8/8	<b>13.09 (-0.22)</b>	<b>19.91 (-0.63)</b>	<b>0.3134 (+0.0000)</b>
Q-Diffusion*	8/8	187.65	12.80	9.87	4/8	14.49	20.43	0.3121
PTQD	8/8	153.92	11.94	8.03	4/8	<b>13.36 (-1.13)</b>	<b>20.14 (-0.29)</b>	<b>0.3128 (+0.0007)</b>
TFMQ-DM (Ours)	8/8	<b>198.86 (+11.21)</b>	<b>10.79 (-1.15)</b>	<b>7.65 (-0.38)</b>				

3. Visualization Results

