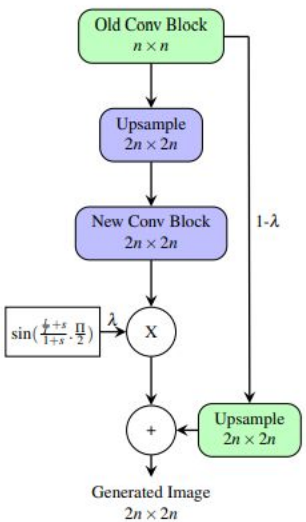
Pro-DDPM: Progressive Growing of Variable Denoising Diffusion Probabilistic Models for Faster Convergence

Rohit Gandikota and Nik Brown
Northeastern University

Motivation

- Longer training time for DDPM
- Progressive growth in DDPM ?
- Sub-optimal linear growth of layers

Sinusoidal Faded Growth



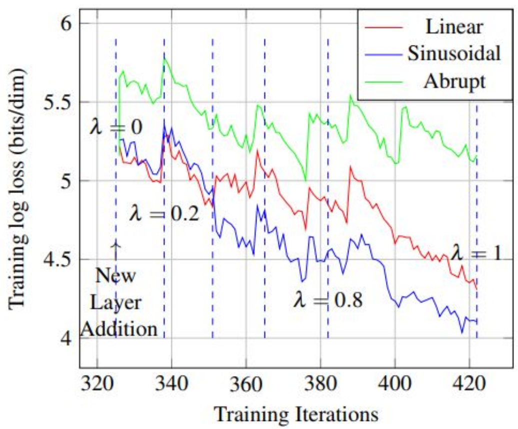
$$\lambda = \sin\left(\frac{\frac{t}{T}+s}{1+s}.\frac{\Pi}{2}\right)$$

Why sinusoidal fading ?

- Abrupt growth creates sudden dip
- Linear fading is suboptimal in second half
- Need for a non-linear growth technique

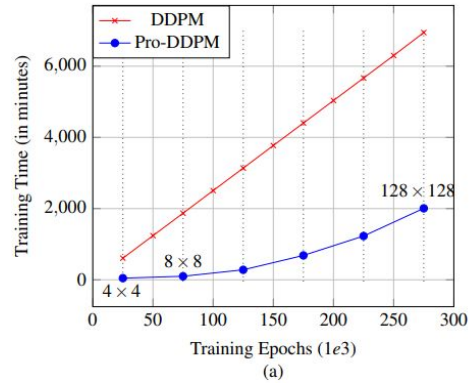
Advantage of sinusoidal fading

- Slower fading at the beginning of growth
- Relatively faster convergence

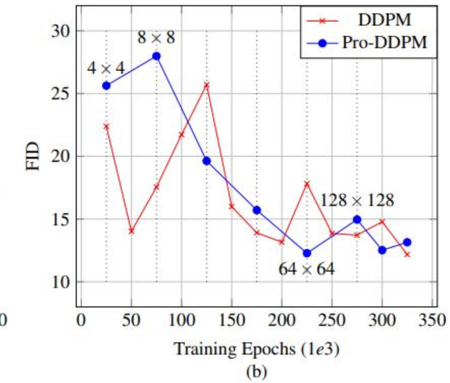


Log likelihood curves for different growth variants at the time of layer addition
[Abrupt, Linear, Sinusoidal]

Improved Training Time



Training time per 25k epochs



FID per 25k epochs

Quality Preservation

Model	FID
BigGAN-deep	4.06
Improved DDPM (large)	2.92
Pro-DDPM (Ours)	2.96

Imagenet 64x64
(FID using InceptionV3)

Model	ImageNet	CIFAR-10
Sparse Transformer	3.44	**2.80**
Routing Transformer	**3.43**	-
DDPM	3.77	3.70
DDIM	-	3.10
Improved DDPM	3.53	2.94
Pro-DDPM (Ours)	**3.61**	**3.09**

Log-likelihood Comparisons

BMVC 2022

Northeastern University

gandikota.ro@northeastern.edu