August 23, 2022

Score-based Denoising Diffusion Models

- a tutorial

Vikram Voleti

PhD candidate, Mila, University of Montreal

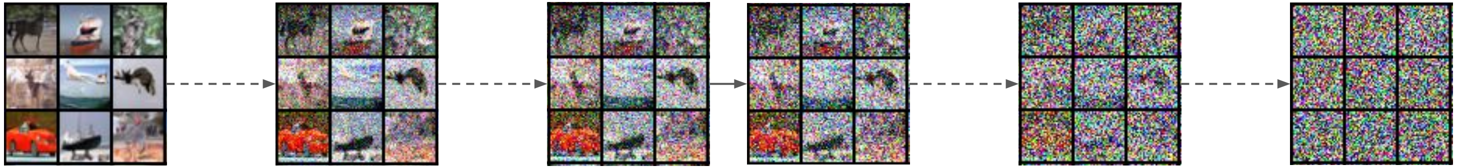
Supervisor: **Prof. Christopher Pal**

voletiv.github.io

@ (virtual) Mila, Canada

1. **Denoising Diffusion Probabilistic Model (DDPM)**
2. Score-based Generative Model (SGM)
3. MCVD: Masked Conditional Video Diffusion

DDPM : Forward process



$\mathbf{x}_0$ $\mathbf{x}_{t-1}$ $\mathbf{x}_t$ $\mathbf{x}_T$

$$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})$$

$; 0 < \beta_1 < \cdots < \beta_t < \cdots < \beta_T = 1$

$\mathbf{x}_0$ $\mathbf{x}_t$

$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \qquad ; \bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$$

$$\implies \mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \qquad ; \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$\mathbf{x}_0$ $\mathbf{x}_{t-1}$

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_{t-1}}\epsilon$$

arxiv.org/abs/2006.11239

DDPM

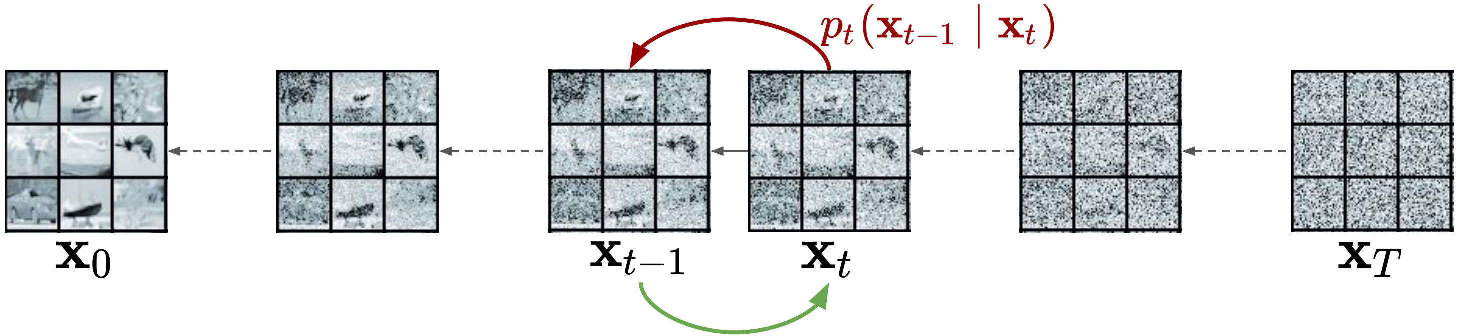
(1) $$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

(2) $$\mathbf{x}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon)/\sqrt{\bar{\alpha}_t}$$

(3) $$\epsilon = \frac{1}{\sqrt{1-\bar{\alpha}_t}}\left(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0\right)$$

arxiv.org/abs/2006.11239

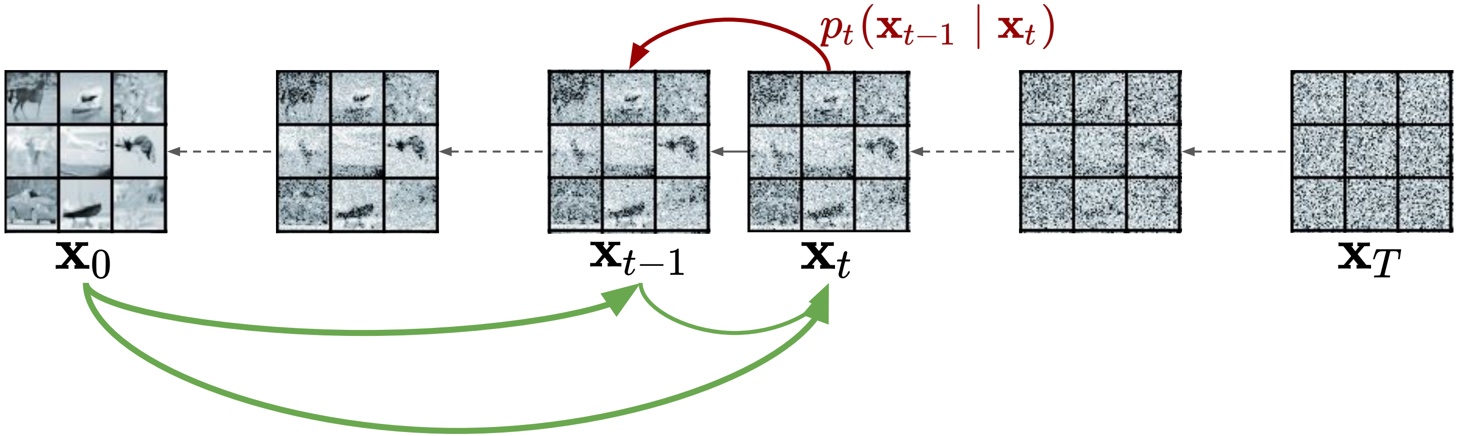
DDPM : Reverse process



$$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \frac{q_t(\mathbf{x}_t|\mathbf{x}_{t-1})✔\ p_t(\mathbf{x}_{t-1})?}{p_t(\mathbf{x}_t)?}$$ (Bayes' theorem)

arxiv.org/abs/2006.11239

DDPM : Reverse process

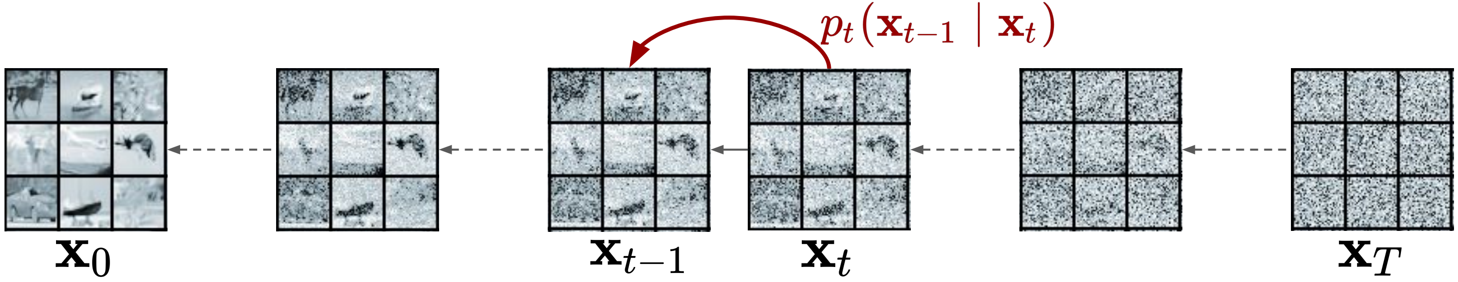


$$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \boxed{\mathbf{x}_0}) = \frac{q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \checkmark \; q_t(\mathbf{x}_{t-1} \mid \boxed{\mathbf{x}_0}) \checkmark}{q_t(\mathbf{x}_t \mid \boxed{\mathbf{x}_0}) \checkmark}$$

(Condition on $\mathbf{x}_0$)

arxiv.org/abs/2006.11239

DDPM : Reverse process



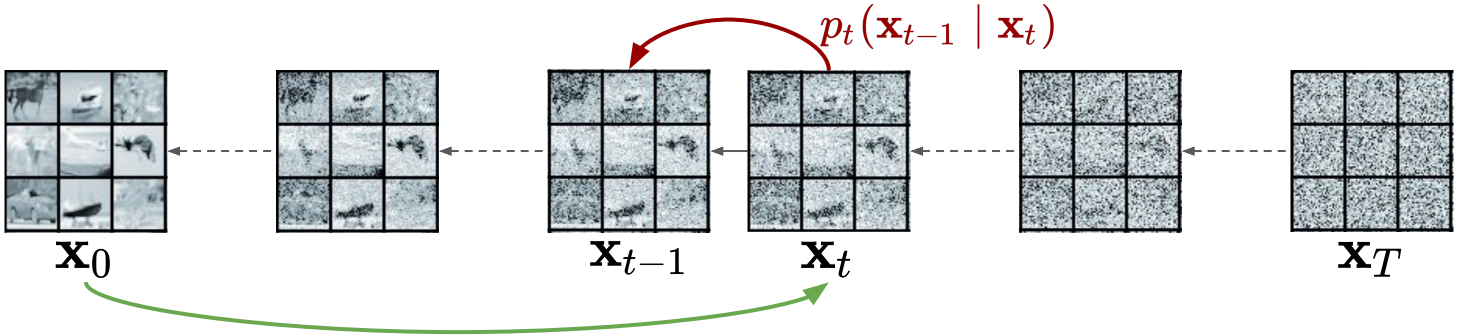
$$
\begin{aligned}
p(\mathbf{u}) &= \mathcal{N}(\mathbf{u} \mid \mathbf{mu}, \mathbf{\Lambda}^{-1}), \\
p(\mathbf{v} \mid \mathbf{u}) &= \mathcal{N}(\mathbf{v} \mid \mathbf{Au} + \mathbf{b}, \mathbf{L}^{-1}) \\
\Rightarrow p(\mathbf{v}) &= \mathcal{N}(\mathbf{v} \mid \mathbf{A}\mu + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^T), \\
\Rightarrow p(\mathbf{u} \mid \mathbf{v}) &= \mathcal{N}(\mathbf{u} \mid \mathbf{C}(\mathbf{A}^T\mathbf{L}(\mathbf{v} - \mathbf{b}) + \mathbf{\Lambda mu}), \mathbf{C}) \\
&\quad [\mathbf{C} = (\mathbf{\Lambda} + \mathbf{A}^T\mathbf{LA})^{-1}]
\end{aligned}
$$

where $\mathbf{u} = \mathbf{x}_{t-1} \mid \mathbf{x}_0$, $\mathbf{v} = \mathbf{x}_t$

$$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I})$$

$$\tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t \; ; \; \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

arxiv.org/abs/2006.11239



$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

From (2) :

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

(2) $\mathbf{x}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon)/\sqrt{\bar{\alpha}_t}$

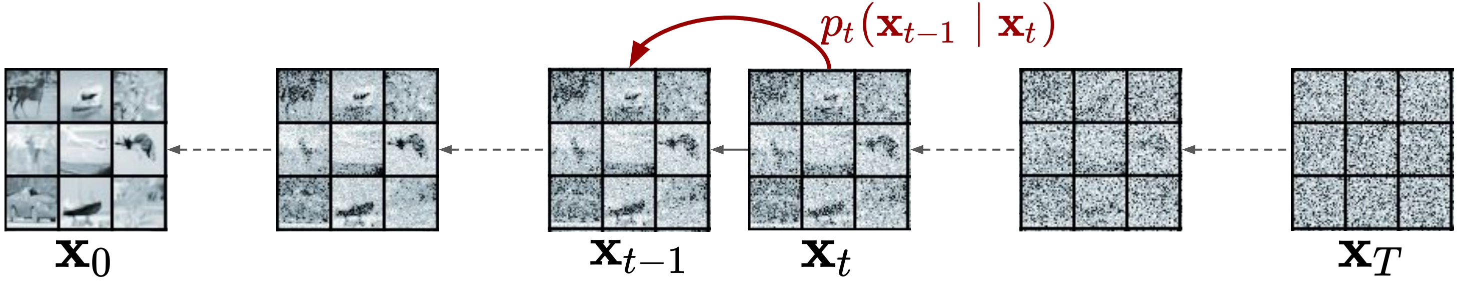
?

1 $$\mathbf{x}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t} \quad \epsilon \quad)/\sqrt{\bar{\alpha}_t}$$

2 $$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t\mathbf{I})$$

arxiv.org/abs/2006.11239

DDPM : Reverse process



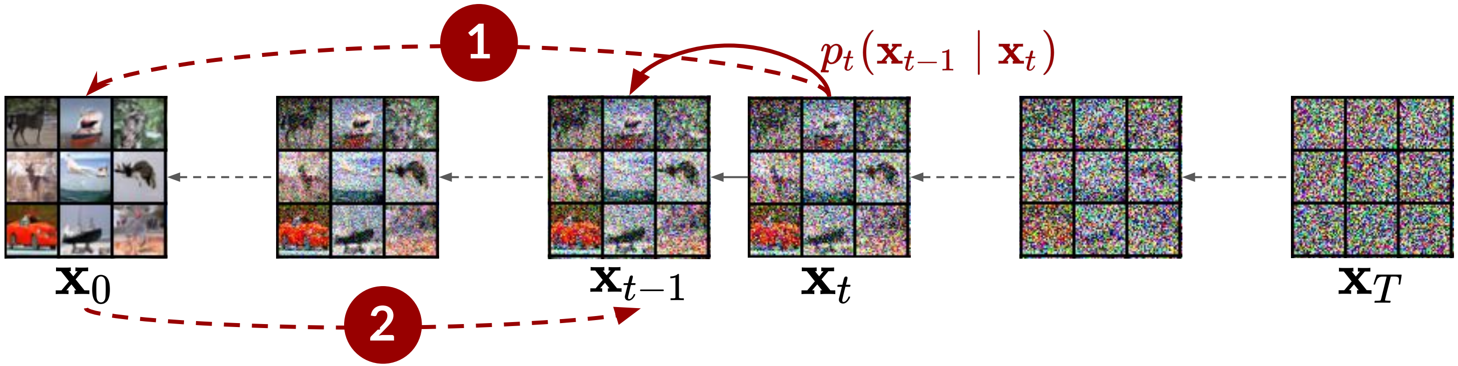
Deep neural network!

1. $$\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$$

2. $$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \hat{\mathbf{x}}_0) = \mathcal{N}(\tilde{\mu}_t(\mathbf{x}_t, \hat{\mathbf{x}}_0), \tilde{\beta}_t \mathbf{I})$$

arxiv.org/abs/2006.11239

DDPM : Generation



DDPM

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T \to 0$:

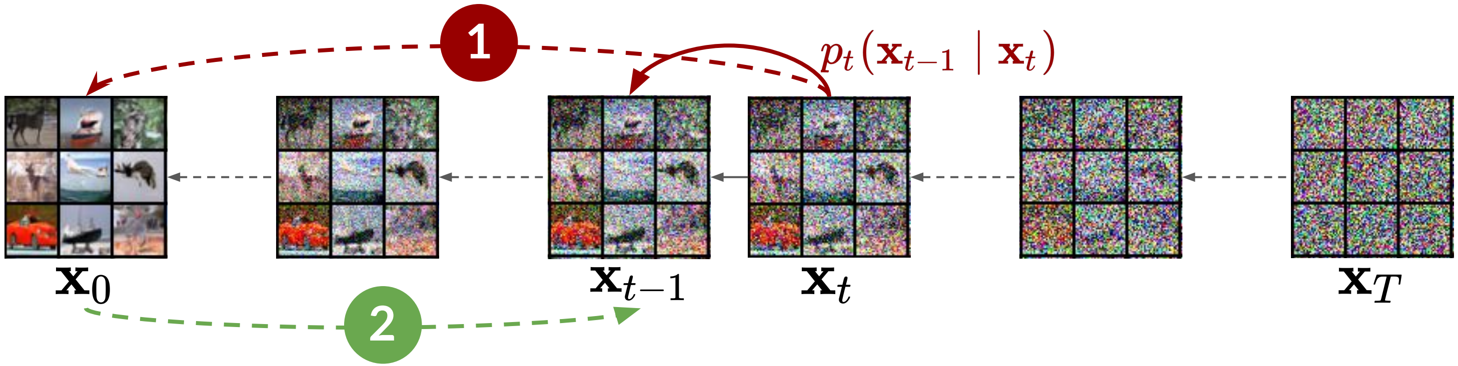
1 $$\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$$

2 $$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}\,\mathbf{z}_t$$

$(\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}))$

arxiv.org/abs/2006.11239

DDIM : Generation



DDIM

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T \to 0$:

1 $$\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$$

2 $$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{x}}_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\epsilon_\theta(\mathbf{x}_t, t)$$

$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

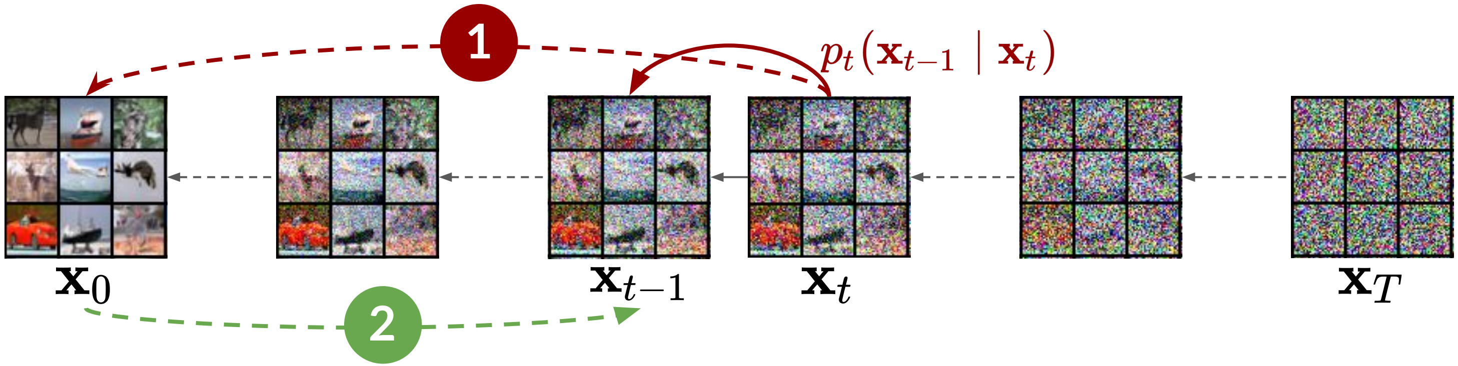
(1) $$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_{t-1}}\epsilon$$

Deterministic!

arxiv.org/abs/2010.02502

DDIM : Generation



DDIM

$$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

for $t = T \to 0$:

1. $$\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t}\, \epsilon_\theta(\mathbf{x}_t, t)) / \sqrt{\bar{\alpha}_t}$$

2. $$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\, \hat{\mathbf{x}}_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2}\, \epsilon_\theta(\mathbf{x}_t, t) + \sigma_t \mathbf{z}_t$$

Deterministic $\Rightarrow \sigma_t = 0$

$(\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}))$

arxiv.org/abs/2010.02502

DDPM vs DDIM

DDPM	DDIM
$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar\alpha_t}\mathbf{x}_0, (1-\bar\alpha_t)\mathbf{I})$	$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar\alpha_t}\mathbf{x}_0, (1-\bar\alpha_t)\mathbf{I})$
$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}\mathbf{x}_0 + \frac{\sqrt{1-\beta_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}\mathbf{x}_t \; ; \; \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t\mathbf{I} \right)$	$p_t(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\sqrt{\bar\alpha_{t-1}}\mathbf{x}_0 + \sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\,\epsilon \; ; \; \sigma_t^2\mathbf{I} \right)$

DDPM

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T \to 0$:

1. $\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar\alpha_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar\alpha_t}$
2. $\mathbf{x}_{t-1} = \frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}\mathbf{x}_t + \sqrt{\frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t}\,\mathbf{z}_t$

DDIM

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T \to 0$:

1. $\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar\alpha_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar\alpha_t}$
2. $\mathbf{x}_{t-1} = \sqrt{\bar\alpha_{t-1}}\hat{\mathbf{x}}_0 + \sqrt{1-\bar\alpha_{t-1}-\underset{0}{\sigma_t^2}}\,\epsilon_\theta(\mathbf{x}_t, t) + \underset{0}{\sigma_t}\mathbf{z}_t$

arxiv.org/abs/2006.11239, arxiv.org/abs/2010.02502

Noise matching

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \boxed{||\epsilon - \epsilon_\theta(\mathbf{x}_t, t)||_2^2}$$

$$(\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon)$$

arxiv.org/abs/2006.11239

1. Noise matching

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} ||\epsilon - \epsilon_\theta(\mathbf{x}_t, t)||_2^2$$

$$(\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon)$$

2. Generation

DDPM

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T \to 0$:

1. $$\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$$

2. $$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \mathbf{z}_t$$

$$(\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

arxiv.org/abs/2006.11239

1. Training

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\right|\right|_2^2$$

$$\left(\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon\right)$$

2. Generation

1 $$\hat{\mathbf{x}}_0 = \left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t)\right)/\sqrt{\bar{\alpha}_t}$$

2 $$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \mathbf{z}_t$$

$$(\mathbf{z}_t \sim \mathcal{N}(\mathbf{0},\mathbf{I}))$$

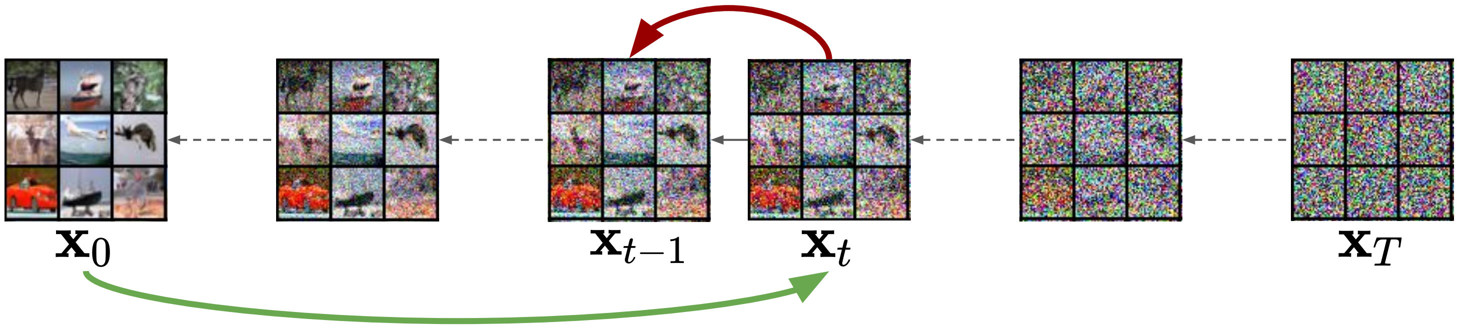
arxiv.org/abs/2006.11239

1. Denoising Diffusion Probabilistic Model (DDPM)

2. **Score-based Generative Model (SGM)**

3. MCVD: Masked Conditional Video Diffusion

SGM



$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2 \mathbf{I}) \qquad ; \sigma_1 < \cdots < \sigma_t < \cdots < \sigma_T$$

$$\implies \mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon \qquad ; \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Langevin Dynamics:

$$\mathbf{x}_{t-1} \longleftarrow \mathbf{x}_t$$

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \lambda_t \nabla_{\mathbf{x}} \log p(\mathbf{x}_t) + \sqrt{2\lambda_t}\, \mathbf{z}_t \qquad ; \lambda_t = \lambda \sigma_t^2 / \sigma_1^2 ;\ \mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

(Gradient ascent) (Stochastic)

arxiv.org/abs/1907.05600

SGM : Score Matching

1. Score Matching:

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) \right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

Denoising Score Matching:

$$\approx \mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t | \mathbf{x}_0) \right|\right|_2^2$$

$$J_{ESMq}(\theta) = \mathbb{E}_{q(\mathbf{x})}\left[\frac{1}{2}\left\|\psi(\mathbf{x};\theta) - \frac{\partial \log q(\mathbf{x})}{\partial \mathbf{x}}\right\|^2\right]$$

$$J_{DSMq_\sigma}(\theta) = \mathbb{E}_{q_\sigma(\mathbf{x},\tilde{\mathbf{x}})}\left[\frac{1}{2}\left\|\psi(\tilde{\mathbf{x}};\theta) - \frac{\partial \log q_\sigma(\tilde{\mathbf{x}}|\mathbf{x})}{\partial \tilde{\mathbf{x}}}\right\|^2\right]$$

$$J_{ESMq_\sigma} \smile J_{DSMq_\sigma}$$

arxiv.org/abs/1907.05600

https://www.iro.umontreal.ca/~vincentp/Publications/smdae_techreport.pdf

Score (gradient of log density)

$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2 \mathbf{I})$$

$$= \frac{1}{\sqrt{(2\pi)^{|\mathbf{x}|}\sigma_t^2}} \exp\left(-\frac{1}{2\sigma_t^2}(\mathbf{x}_t - \mathbf{x}_0)^T(\mathbf{x}_t - \mathbf{x}_0) \right)$$

$$\Rightarrow \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \text{const} - \frac{1}{2\sigma_t^2}(\mathbf{x}_t - \mathbf{x}_0)^T(\mathbf{x}_t - \mathbf{x}_0)$$

$$\Rightarrow (\text{Score})\ \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = -\frac{1}{\sigma_t^2}(\mathbf{x}_t - \mathbf{x}_0) = \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2}$$

$$\left|\left| \mathbf{s}_\theta(\mathbf{x}_t, t) - \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2} \right|\right|_2^2$$

arxiv.org/abs/1907.05600, arxiv.org/abs/2006.11239 , arxiv.org/abs/2205.09853

1. Score Matching:

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} ||\mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)||_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

Denoising Score Matching:

$$\approx \mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} ||\mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t|\mathbf{x}_0)||_2^2$$

$$= \mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} ||\mathbf{s}_\theta(\mathbf{x}_t, t) - \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2}||_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

arxiv.org/abs/1907.05600

Mila

1. Denoising Score Matching:

$$\ell(\theta) := \mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \mathbf{s}_\theta(\mathbf{x}_t, t) - \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2} \right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

Variance of score:

$$\mathbb{E}\left[||\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t \mid \mathbf{x}_0)||_2^2 \right] = \mathbb{E}\left[|| - \frac{\mathbf{x}_t - \mathbf{x}_0}{\sigma_t^2} ||_2^2 \right] = \mathbb{E}\left[|| \frac{\cancel{\sigma_t}\epsilon}{\sigma_t^{\cancel{2}}} ||_2^2 \right] = \frac{1}{\sigma_t^2} \mathbb{E}\left[||\epsilon||_2^2 \right] = \boxed{\frac{1}{\sigma_t^2}} \dim(\epsilon)$$

(Un)weighted objective function:

$$\mathcal{L}(\theta) = \boxed{\sigma_t^2}\, \ell(\theta) := \mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \sigma_t \mathbf{s}_\theta(\mathbf{x}_t, t) - \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t} \right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011

Objective of SGM

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \boxed{\sigma_t \mathbf{s}_\theta(\mathbf{x}_t, t)} \boxed{- \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t}} \right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

- $\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon \implies \boxed{- \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t}} = \epsilon$
- Make: $\boxed{\sigma_t \mathbf{s}_\theta(\mathbf{x}_t, t)} = -\epsilon_\theta(\mathbf{x}_t, t)$ ------------------ How?

$$\implies \mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \epsilon - \epsilon_\theta(\mathbf{x}_t, t) \right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

Objective of DDPM!

arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011

Score (gradient of log density)

$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2 \mathbf{I})$$
$$= \frac{1}{\sqrt{(2\pi)^{|\mathbf{x}|}\sigma_t^2}} \exp\left(-\frac{1}{2\sigma_t^2}(\mathbf{x}_t - \mathbf{x}_0)^T(\mathbf{x}_t - \mathbf{x}_0)\right)$$

$$\Rightarrow \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \text{const} - \frac{1}{2\sigma_t^2}(\mathbf{x}_t - \mathbf{x}_0)^T(\mathbf{x}_t - \mathbf{x}_0)$$

$$\Rightarrow \text{(Score) } \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = -\frac{1}{\sigma_t^2}(\mathbf{x}_t - \mathbf{x}_0) = \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2}$$

$$\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon \qquad (3)\ \epsilon = \frac{1}{\sigma_t}(\mathbf{x}_t - \mathbf{x}_0)$$

$$\Rightarrow \boxed{\text{(Score) } \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = -\frac{1}{\sigma_t}\epsilon}$$

$\therefore$ Estimating ϵ is equivalent to estimating a scaled version of the **Score**!

arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.11239 , arxiv.org/abs/2205.09853

1. **Training**

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left| \mathbf{s}_\theta(\mathbf{x}_t, t) - \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2} \right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

2. **Generation**

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \lambda_t\, \mathbf{s}_\theta(\mathbf{x}_t, t) + \sqrt{2\lambda_t}\, \mathbf{z}_t \qquad ;\ \lambda_t = \lambda \sigma_t^2 / \sigma_1^2;\ \mathbf{z}_t \sim \mathcal{N}(\mathbf{0},\mathbf{I})$$

arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011

1. Training

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\right|\right|_2^2$$

$$(\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon)$$

2. Generation

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \lambda_t \frac{-1}{\sigma_t} \epsilon_\theta(\mathbf{x}_t, t) + \sqrt{2\lambda_t}\, \mathbf{z}_t \qquad ; \lambda_t = \lambda \sigma_t^2 / \sigma_1^2;\ \mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011

Denoising Diffusion Probabilistic Models (DDPM)

Mila

1. Training

$$\mathbb{E}_{t,\ \mathbf{x}_0 \sim p_{\text{data}},\ \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left|\left|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\right|\right|_2^2$$

$$\left(\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon\right)$$

2. Generation

1 $$\hat{\mathbf{x}}_0 = \left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t)\right) / \sqrt{\bar{\alpha}_t}$$

$$\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

2 $$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}\,\mathbf{z}_t$$

arxiv.org/abs/2006.11239

DDPM vs SGM

DDPM	SGM
$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_0, \beta_t\mathbf{I})$ $q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$	$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_0, (\sigma_t^2 - \sigma_{t-1}^2)\mathbf{I})$ $q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2\mathbf{I})$
1 $\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$ 2 $\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}\,\mathbf{z}_t$	$\mathbf{x}_{t-1} = \mathbf{x}_t + \lambda_t\, \mathbf{s}_\theta(\mathbf{x}_t, t) + \sqrt{2\lambda_t}\; \mathbf{z}_t$

arxiv.org/abs/2006.11239 , arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011

Using SDEs

Forward: $$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt \; + \; g(t)\, d\mathbf{w}$$

Reverse: $$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt \; + \; g(t)\, d\bar{\mathbf{w}}$$

DDPM	SGM
$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_0, \beta_t \mathbf{I})$ $q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$	$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_0, (\sigma_t^2 - \sigma_{t-1}^2)\mathbf{I})$ $q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2 \mathbf{I})$
1 $\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$ 2 $\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}\,\mathbf{z}_t$	$\mathbf{x}_{t-1} = \mathbf{x}_t + \lambda_t\, \mathbf{s}_\theta(\mathbf{x}_t, t) + \sqrt{2\lambda_t}\, \mathbf{z}_t$

arxiv.org/abs/2006.11239 , arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011 , arxiv.org/abs/2011.13456

Forward: $$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + g(t)\, d\mathbf{w}$$

Reverse: $$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})]dt + g(t)\, d\bar{\mathbf{w}}$$

DDPM	SGM
$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_0, \beta_t \mathbf{I})$ $q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$	$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_0, (\sigma_t^2 - \sigma_{t-1}^2)\mathbf{I})$ $q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2 \mathbf{I})$
1 $\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\mathbf{x}_t, t))/\sqrt{\bar{\alpha}_t}$ 2 $\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t}\,\mathbf{z}_t$	$\mathbf{x}_{t-1} = \mathbf{x}_t + \lambda_t\, \mathbf{s}_\theta(\mathbf{x}_t, t) + \sqrt{2\lambda_t}\, \mathbf{z}_t$
$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\, dt + \sqrt{\beta(t)}\, d\mathbf{w}$ Variance Preserving	$d\mathbf{x} = \sqrt{\frac{d[\sigma^2(t)]}{dt}}\, d\mathbf{w}$ Variance Exploding

arxiv.org/abs/2006.11239 , arxiv.org/abs/1907.05600 , arxiv.org/abs/2006.09011 , arxiv.org/abs/2011.13456

Score (gradient of log density) for DDPM

Mila

$$q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

$$= \frac{1}{\sqrt{(2\pi)^{|\mathbf{x}|}(1-\bar{\alpha}_t)}} \exp\left(-\frac{1}{2(1-\bar{\alpha}_t)}(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^T(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0) \right)$$

$$\Rightarrow \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = \text{const} - \frac{1}{2(1-\bar{\alpha}_t)}(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^T(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)$$

$$\Rightarrow \text{(Score) } \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = -\frac{1}{1-\bar{\alpha}_t}(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)$$

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \qquad \text{(3)} \ \epsilon = \frac{1}{\sqrt{1-\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)$$

$$\Rightarrow \boxed{\text{(Score) } \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t \mid \mathbf{x}_0) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\epsilon}$$

$\therefore$ Estimating ϵ is equivalent to estimating a scaled version of the **Score**!

arxiv.org/abs/1907.05600, arxiv.org/abs/2006.11239 , arxiv.org/abs/2205.09853

1. Denoising Diffusion Probabilistic Model (DDPM)
2. Score-based Generative Model (SGM)
3. **MCVD: Masked Conditional Video Diffusion**

MCVD: Masked Conditional Video Diffusion

Mila

p past frames: $\mathbf{p} = \{\mathbf{p}^{\mathbf{i}}\}_{i=1}^{p}$ | k current frames: $\mathbf{x}_0 = \{\mathbf{x}_0^{\mathbf{i}}\}_{i=1}^{k}$ | f future frames: $\mathbf{f} = \{\mathbf{f}^{\mathbf{i}}\}_{i=1}^{f}$ | $\left(\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon\right)$

- **Video Prediction:**

$$\mathbb{E}_{t,\ \boxed{[\mathbf{p},\mathbf{x}_0]\sim p_{\text{data}}},\ \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \left\|\epsilon - \epsilon_\theta(\mathbf{x}_t \mid \boxed{\mathbf{p},}\, t)\right\|_2^2$$

- **Video Generation:**

$$\mathbb{E}_{t,\ \mathbf{x}_0\sim p_{\text{data}},\ \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \left\|\epsilon - \epsilon_\theta(\mathbf{x}_t \mid t)\right\|_2^2$$

- **Video Interpolation:**

$$\mathbb{E}_{t,\ \boxed{[\mathbf{p},\mathbf{x}_0,\mathbf{f}]\sim p_{\text{data}}},\ \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \left\|\epsilon - \epsilon_\theta(\mathbf{x}_t \mid \boxed{\mathbf{p},\mathbf{f},}\, t)\right\|_2^2$$

arxiv.org/abs/2205.09853

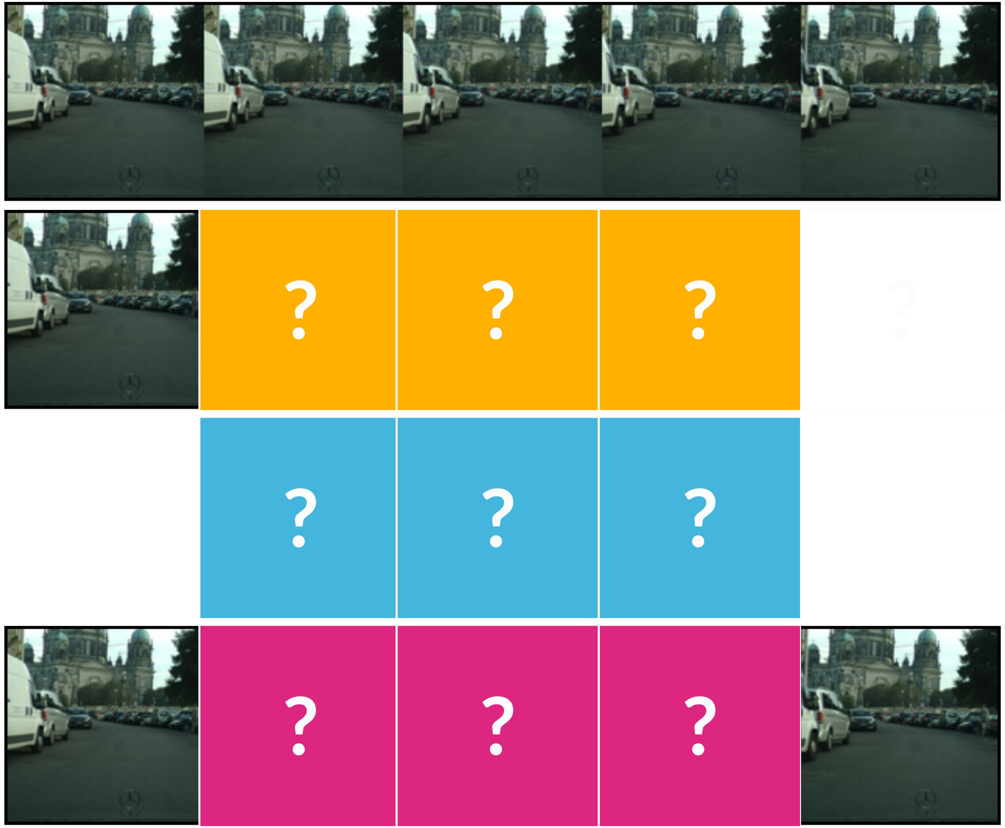
MCVD: Masked Conditional Video Diffusion

Real data

Video Prediction

Video Generation

Video Interpolation



arxiv.org/abs/2205.09853

p past frames: $\mathbf{p} = \{\mathbf{p}^{\mathbf{i}}\}_{i=1}^{p}$ | k current frames: $\mathbf{x}_0 = \{\mathbf{x}_0^{\mathbf{i}}\}_{i=1}^{k}$ | f future frames: $\mathbf{f} = \{\mathbf{f}^{\mathbf{i}}\}_{i=1}^{f}$ | $\left(\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon\right)$

- **Video Prediction:**

$$\mathbb{E}_{t,\ [\mathbf{p},\mathbf{x}_0]\sim p_{\text{data}},\ \ \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \|\epsilon - \epsilon_\theta(\mathbf{x}_t \mid \mathbf{p}, t)\|_2^2$$

Random masking!

- **Video Prediction + Generation:**

$$\mathbb{E}_{t,\ [\mathbf{p},\mathbf{x}_0,\mathbf{f}]\sim p_{\text{data}},\ \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}),\ \boxed{m_p\sim\mathcal{B}(p_{\text{mask}})}} \|\epsilon - \epsilon_\theta(\mathbf{x}_t \mid \boxed{m_p\mathbf{p},} t)\|_2^2$$

- **Video Prediction + Generation + Interpolation:**

$$\mathbb{E}_{t,\ [\mathbf{p},\mathbf{x}_0,\mathbf{f}]\sim p_{\text{data}},\ \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}),\ \boxed{(m_p,m_f)\sim\mathcal{B}(p_{\text{mask}})}} \|\epsilon - \epsilon_\theta(\mathbf{x}_t \mid \boxed{m_p\mathbf{p}, m_f\mathbf{f},} t)\|_2^2$$

arxiv.org/abs/2205.09853

MCVD: Masked Conditional Video Diffusion

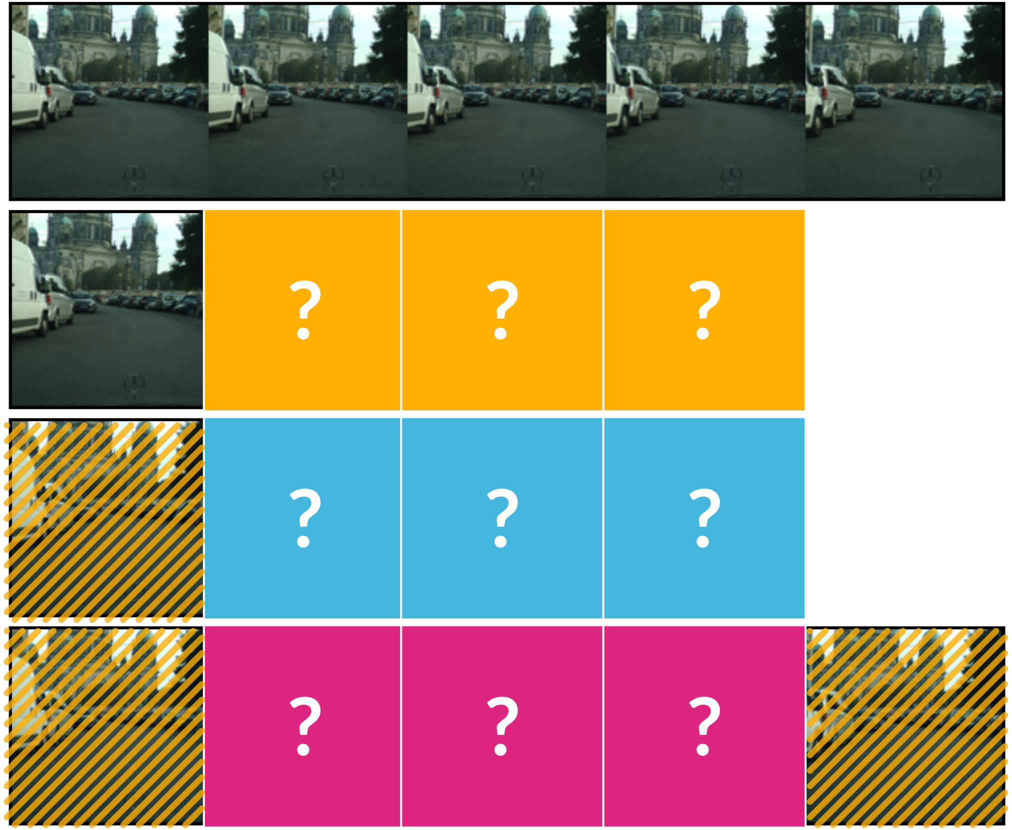
Mila

Real data

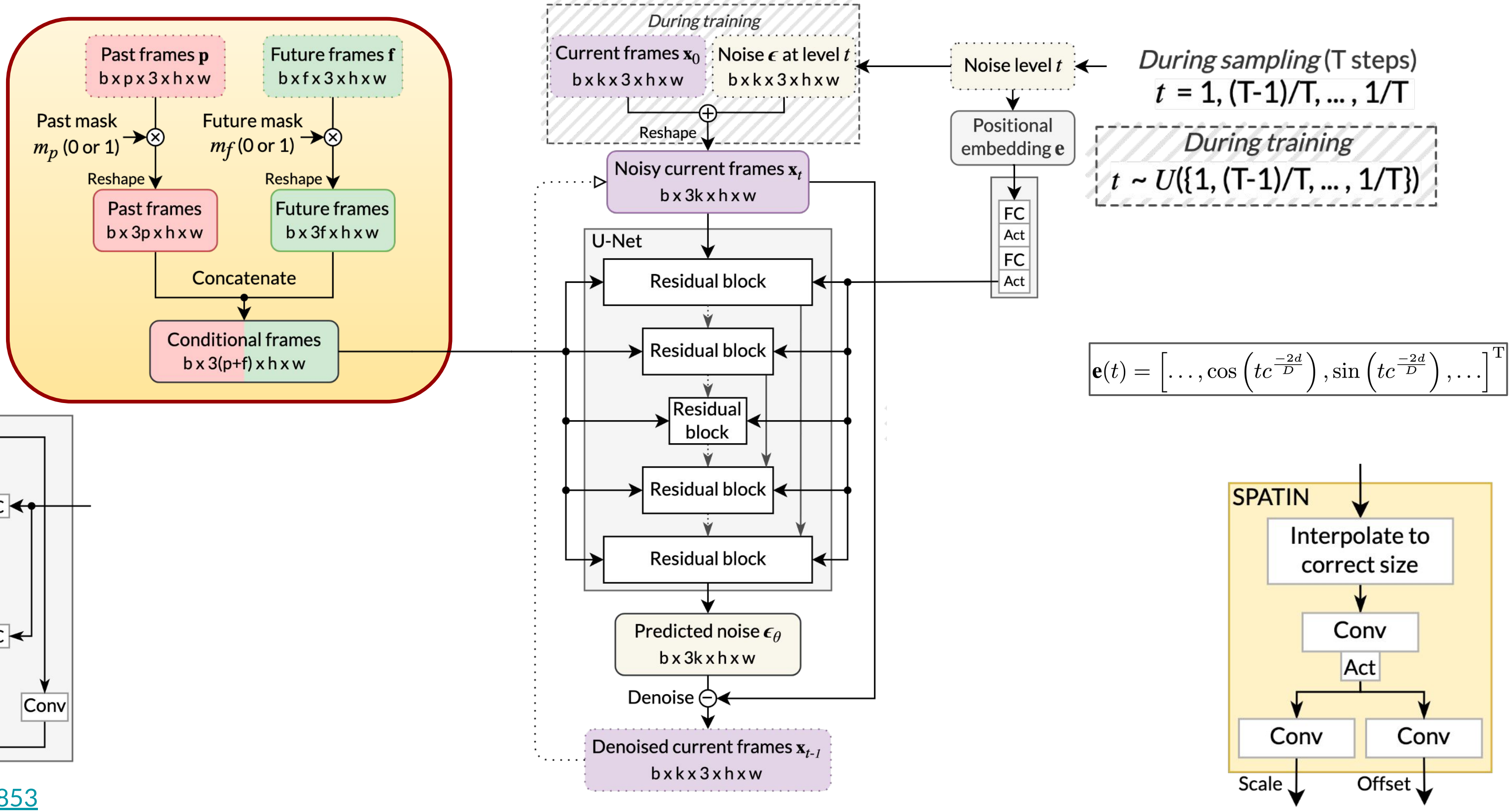
Video Prediction

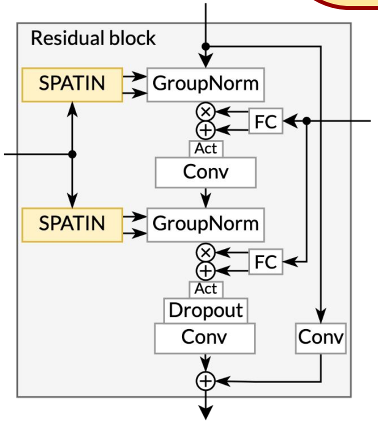
Video Prediction
+ Generation

Video Prediction
+ Generation
+ Interpolation



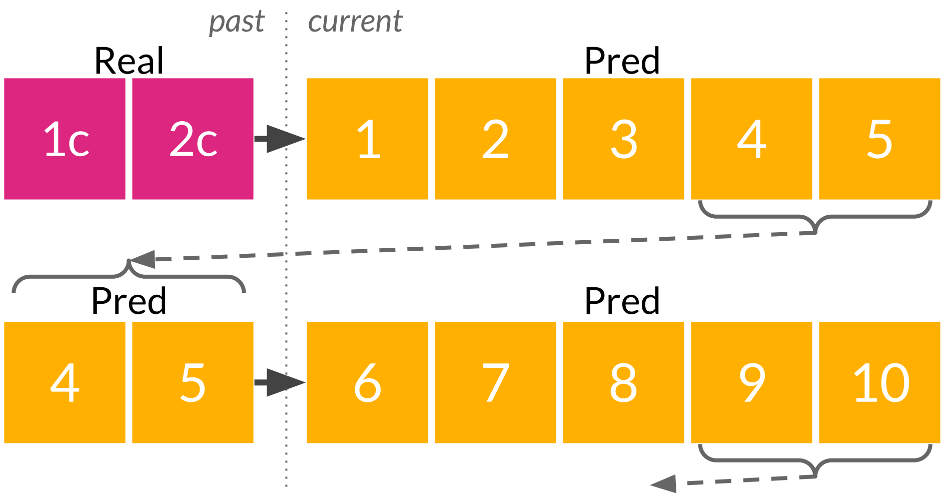
arxiv.org/abs/2205.09853





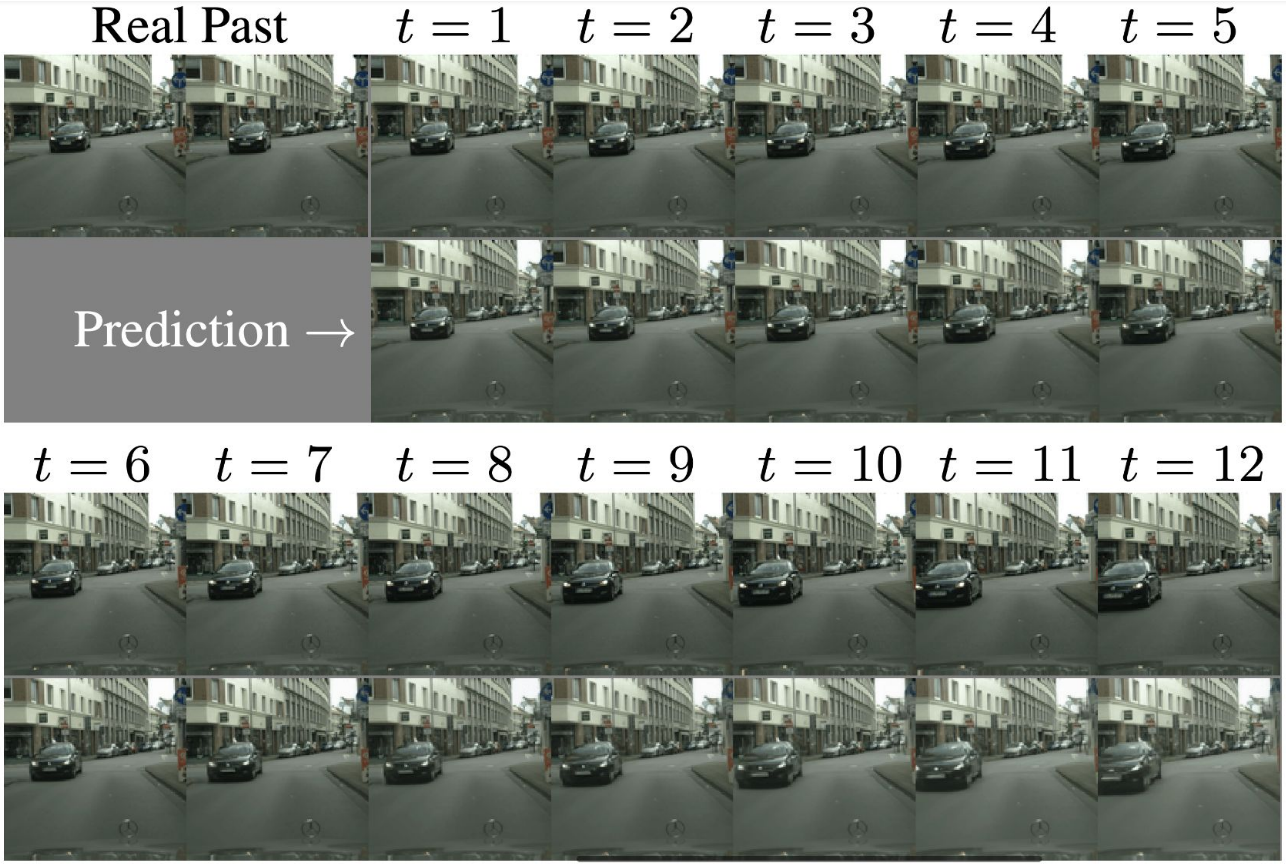
arxiv.org/abs/2205.09853

Block-autoregressive generation:



arxiv.org/abs/2205.09853

MCVD: Masked Conditional Video Diffusion



arxiv.org/abs/2205.09853

MCVD: Masked Conditional Video Diffusion

(128x128)

Cityscapes [2 → 28; trained on k]	k	FVD↓	LPIPS↓
SVG-LP Denton and Fergus [2018]	10	1300.26	0.549 ± 0.06
vRNN 1L Castrejón et al. [2019]	10	682.08	0.304 ± 0.10
Hier-vRNN Castrejón et al. [2019]	10	567.51	0.264 ± 0.07
GHVAE Wu et al. [2021]	10	418.00	0.193 ± 0.014
MCVD spatin (Ours)	**5**	184.81	0.121 ± 0.05
MCVD concat (Ours)	**5**	**141.31**	**0.112** ± 0.05

(64x64)

SMMNIST [5 → 10; trained on k]	k	**FVD↓**	**SSIM↑**
SVG [Denton and Fergus, 2018]	10	90.81	0.688
vRNN 1L [Castrejón et al., 2019]	10	63.81	0.763
Hier-vRNN [Castrejón et al., 2019]	10	57.17	0.760
MCVD concat (Ours)	**5**	25.63	**0.786**
MCVD spatin (Ours)	**5**	23.86	0.780
MCVD concat past-future-mask	**5**	**20.77**	0.753

(64x64)

KTH [10 → *pred*; trained on 10]	*pred*	FVD↓	PSNR↑	SSIM↑
SAVP [Lee et al., 2018]	30	374	26.5	0.756
MCVD spatin (Ours)	30	323	27.5	0.835
MCVD concat past-future-mask (Ours)	30	295	24.3	0.746
SLAMP [Akan et al., 2021]	30	228	29.4	0.865
SRVP [Franceschi et al., 2020]	30	**222**	**29.7**	**0.870**

arxiv.org/abs/2205.09853

(64x64)

BAIR [past p → *pred* ; trained on k]	p	k	*pred*	FVD↓	PSNR↑	SSIM↑
LVT [Rakhimov et al., 2020]	1	15	15	125.8	–	–
DVD-GAN-FP [Clark et al., 2019]	1	15	15	109.8	–	–
MCVD spatin (Ours)	1	**5**	15	103.8	18.8	0.826
TrIVD-GAN-FP [Luc et al., 2020]	1	15	15	103.3	–	–
VideoGPT [Yan et al., 2021]	1	15	15	103.3	–	–
CCVS [Le Moing et al., 2021]	1	15	15	99.0	–	–
MCVD concat (Ours)	1	**5**	15	98.8	18.8	0.829
MCVD spatin past-mask (Ours)	1	**5**	15	96.5	18.8	0.828
MCVD concat past-mask (Ours)	1	**5**	15	**95.6**	**18.8**	**0.832**
Video Transformer [Weissenborn et al., 2019]	1	15	15	94-96[a]	–	–
FitVid [Babaeizadeh et al., 2021]	1	15	15	93.6	–	–
MCVD concat past-future-mask (Ours)	1	**5**	15	**89.5**	16.9	0.780
SAVP [Lee et al., 2018]	2	14	14	116.4	–	–
MCVD spatin (Ours)	2	**5**	14	94.1	19.1	0.836
MCVD spatin past-mask (Ours)	2	**5**	14	90.5	**19.2**	0.837
MCVD concat (Ours)	2	**5**	14	90.5	19.1	0.834
MCVD concat past-future-mask (Ours)	2	**5**	14	89.6	17.1	0.787
MCVD concat past-mask (Ours)	2	**5**	14	**87.9**	19.1	**0.838**
SVG-LP [Akan et al., 2021]	2	10	28	256.6	–	0.816
SLAMP [Akan et al., 2021]	2	10	28	245.0	19.7	0.818
SAVP [Lee et al., 2018]	2	10	28	143.4	–	0.795
Hier-vRNN [Castrejón et al., 2019]	2	10	28	143.4	–	**0.822**
MCVD spatin (Ours)	2	**5**	28	132.1	17.5	0.779
MCVD spatin past-mask (Ours)	2	**5**	28	127.9	17.7	0.789
MCVD concat (Ours)	2	**5**	28	120.6	17.6	0.785
MCVD concat past-mask (Ours)	2	**5**	28	119.0	**17.7**	0.797
MCVD concat past-future-mask (Ours)	2	**5**	28	**118.4**	16.2	0.745

MCVD: Masked Conditional Video Diffusion

mask-cond-video-diffusion.github.io

arxiv.org/abs/2205.09853

Thank you!

voletiv.github.io