CVPR #14186

6. Supplementary Materials

6.1. The proof of Proposition 1

We observe the forward spatio-temporal noising process as follows:

$$z_t^i = \sqrt{\alpha_t}(\sqrt{\rho_t} * z_{t-1}^i + \sqrt{1-\rho_t} * z_{t-1}^{i-1}) + \sqrt{1-\alpha_t}\varepsilon_t^i. \tag{24}$$

Then, with condition $\sqrt{\rho_t} = c, \sqrt{1-\rho_t} = d$,we can calculate the form $z_1^i, z_2^i, \dots, z_t^i$:

$$\begin{aligned} z_1^i &= \sqrt{\alpha_1}(c * z_0^i + d * z_0^{i-1}) + \sqrt{1-\alpha_1}\varepsilon_1^i, \quad (25)\\ z_2^i &= \sqrt{\alpha_1\alpha_2}(c^2 * z_0^i + 2cd * z_0^{i-1} + d^2 * z_0^{i-2}) \\ &\quad + \sqrt{\alpha_2}\sqrt{1-\alpha_1}(c\varepsilon_1^i + d\varepsilon_1^{i-1}) + \sqrt{1-\alpha_2}\varepsilon_2^i \\ &= \sqrt{\alpha_1\alpha_2}\sum_{m=0}^{2}\binom{m}{2}c^{t-m}d^m z_0^{i-m} + \sqrt{1-\alpha_1\alpha_2}\hat{\varepsilon}_2^i \end{aligned} \tag{26}$$

$$\dots$$

$$z_t^i = \sqrt{\bar{\alpha}_t}\sum_{m=0}^{t}\binom{m}{t}c^{t-m}d^m z_0^{i-m} + \sqrt{1-\bar{\alpha}_t}\hat{\varepsilon}_t^i, \tag{27}$$

where, $\hat{\varepsilon}_2^i \sim \mathcal{N}(0,1)$. Therefore, Then, we can generalize the mathematical form of z_t^i:

$$\begin{aligned} z_t^i &= \sqrt{\bar{\alpha}_t}\sum_{m=0}^{t}\binom{m}{t}c^{t-m}d^m z_0^{i-m} + \sqrt{1-\bar{\alpha}_t}\hat{\varepsilon}_t^i \\ &= \sqrt{\bar{\alpha}_t}\rho^{t/2}z_0^i + \sqrt{\bar{\alpha}_t}g(t,i) + \sqrt{1-\bar{\alpha}_t}\varepsilon_t^i, \quad (28)\\ \text{s.t.}\ ,\quad g(t,i) &= \sum_{m=1}^{t}\binom{t}{m}\rho^{(t-m)/2}(1-\rho)^{m/2}z_0^{i-m}. \end{aligned} \tag{29}$$

6.2. The proof of Eqn.(16)

In order to solve the posterior distribution of STDD, we need to rethink our forward diffusion model in Eqn. (28). Based on the special operator F in Eqn. (9), the forward diffusion process can be written as

$$Z_t = \sqrt{\bar{\alpha}_t}F_t(Z_0) + \sqrt{1-\bar{\alpha}_t}\hat{\varepsilon}_t, \tag{30}$$

Since $\sum_{m=0}^{t}\binom{t}{m}\rho^{(t-m)/2}(1-\rho)^{m/2}\varepsilon_t^m = (\sqrt{\rho}^2 + \sqrt{1-\rho}^2)^{t/2} * \varepsilon = \varepsilon$, based on the properties of the Gaussian distribution, we can get:

$$Z_t = \sqrt{\bar{\alpha}_t}F_t(Z_0) + \sqrt{1-\bar{\alpha}_t}F_t(\bar{\varepsilon}_t), \tag{31}$$

where, Gaussian noise ε_t is not independent, and can be obtained by the independent noise $\bar{\varepsilon}_t$ diffusion of each frame image (i.e., $\varepsilon = F_t(\bar{\varepsilon}_t)$). Therefore, the probabilistic model of forward diffusion can be written as:

$$\begin{cases} q(Z_t|Z_{t-1}, Z_0) = \mathcal{N}(Z_t|\sqrt{\alpha_t}F_t(F_{t-1}^{-1}(Z_{t-1})), \beta_t I) \\ q(Z_t|Z_0) = \mathcal{N}(Z_t|\sqrt{\bar{\alpha}_t}F_t(Z_0), \sqrt{1-\bar{\alpha}_t}I) \end{cases} \tag{32}$$

The posterior distribution can be obtained from the prior distribution via Bayes' formula:

$$\begin{aligned} q(Z_{t-1}|Z_t) &= q(Z_{t-1}|Z_t, Z_0) \\ &= q(Z_t|Z_{t-1}, Z_0)\frac{q(Z_{t-1}|Z_0)}{q(Z_t|Z_0)} \\ &= N(\tilde{\mu_t}, \tilde{\beta_t}). \end{aligned} \tag{33}$$

We obtain the mean and variance of the posterior distribution as:

$$\begin{cases} \tilde{\mu_t} = \dfrac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}F_{t-1}(Z_0) + \dfrac{\sqrt{\bar{\alpha}_t}(1-\alpha_{t-1})}{1-\bar{\alpha}_t}F_{t-1}(F_t^{-1}(Z_t)) \\ \tilde{\beta_t} = \dfrac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \end{cases} \tag{34}$$

By substituting Z_o for Z_t with Eqn. (31), the mean and variance of the STDD posterior distribution are finally obtained:

$$\begin{cases} \tilde{\mu_t} = \dfrac{1}{\sqrt{\alpha_t}}\left(F_{t-1}(F_t^{-1}(Z_t)) - \dfrac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_t\right) \\ \tilde{\beta_t} = \dfrac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t, \end{cases} \tag{35}$$

6.3. Visualization of generated videos

We also provide some sampled videos on UCF101 and Sky Time-Lapse, shown in Fig. 5 and Fig. 6. Each line is 16 frames of a sampled video,

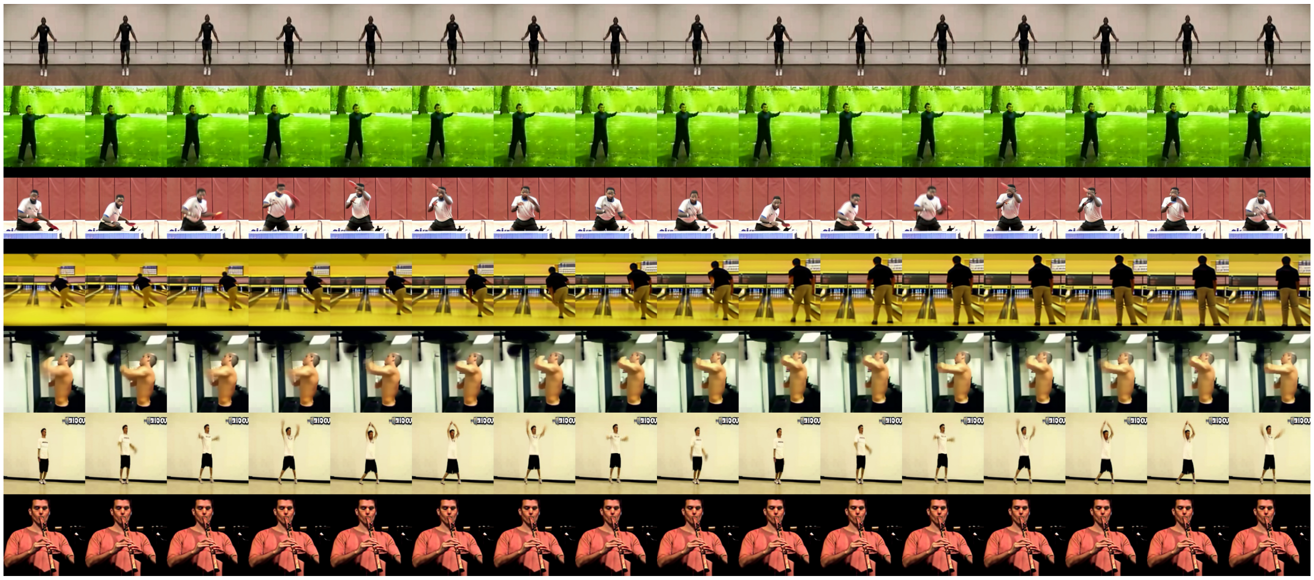

Figure 5. Sample videos on UCF101.

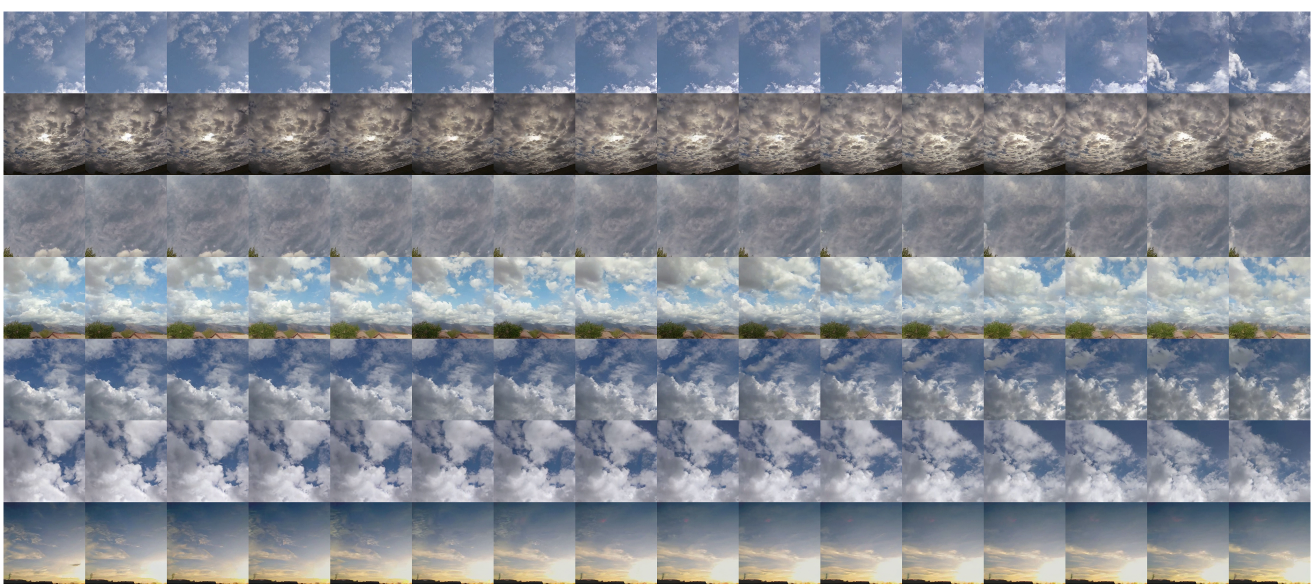

Figure 6. Sample videos on Sky Time-Lapse.