A. Proofs

A.1. Proof of Lemma 2.1

Proof. Since $\mathbf{f}$ is sufficiently smooth in a neighborhood of $(\mathbf{x}_t, t)$, near t, we can expand:

$$\mathbf{x}_{t'} = \mathbf{x}_t \;+\; (t'-t)\,\mathbf{f}\big(\mathbf{x}_t, t\big) \;+\; \frac{(t'-t)^2}{2}\Big[\partial_t \mathbf{f}(\mathbf{x}_t, t) + \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}_t, t)\cdot \mathbf{f}(\mathbf{x}_t, t)\Big] \;+\; O\big((t'-t)^3\big),$$

and

$$\tilde{\mathbf{x}}_{t'} = \tilde{\mathbf{x}}_t \;+\; (t'-t)\,\mathbf{f}\big(\tilde{\mathbf{x}}_t, t\big) \;+\; \frac{(t'-t)^2}{2}\Big[\partial_t \mathbf{f}(\tilde{\mathbf{x}}_t, t) + \nabla_{\mathbf{x}} \mathbf{f}(\tilde{\mathbf{x}}_t, t)\cdot \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Big] \;+\; O\big((t'-t)^3\big).$$

Subtracting the expansions, we get

$$\begin{aligned}\mathbf{x}_{t'} - \tilde{\mathbf{x}}_{t'} = &\Big[\mathbf{x}_t - \tilde{\mathbf{x}}_t\Big] \;+\; (t'-t)\Big(\mathbf{f}(\mathbf{x}_t, t) - \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Big) \\ &+ \frac{(t'-t)^2}{2}\Big\{\partial_t \mathbf{f}(\mathbf{x}_t, t) + \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}_t, t)\cdot \mathbf{f}(\mathbf{x}_t, t) \;-\; \big[\partial_t \mathbf{f}(\tilde{\mathbf{x}}_t, t) + \nabla_{\mathbf{x}} \mathbf{f}(\tilde{\mathbf{x}}_t, t)\cdot \mathbf{f}(\tilde{\mathbf{x}}_t, t)\big]\Big\} \;+\; O\big((t'-t)^3\big).\end{aligned}$$

Hence for small $t' - t$,

$$\mathbf{x}_{t'} - \tilde{\mathbf{x}}_{t'} = \big[\mathbf{x}_t - \tilde{\mathbf{x}}_t\big] \;+\; (t'-t)\Big[\mathbf{f}(\mathbf{x}_t, t) - \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Big] \;+\; O\big((t'-t)^2\big).$$

After the rectification

$$\hat{\mathbf{x}}_{t'} = \tilde{\mathbf{x}}_{t'} + \mathbf{r}_\theta(\mathbf{x}_t, \tilde{\mathbf{x}}_t, t, t'-t) = \tilde{\mathbf{x}}_{t'} \;+\; (t'-t)\Big[\mathbf{f}(\mathbf{x}_t, t) - \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Big] + \big[\mathbf{x}_t - \tilde{\mathbf{x}}_t\big].$$

Then

$$\begin{aligned}\mathbf{x}_{t'} - \hat{\mathbf{x}}_{t'} = &\big[\mathbf{x}_{t'} - \tilde{\mathbf{x}}_{t'}\big] \;-\; (t'-t)\,\big[\mathbf{f}(\mathbf{x}_t, t) - \mathbf{f}(\tilde{\mathbf{x}}_t, t)\big] - \big[\mathbf{x}_t - \tilde{\mathbf{x}}_t\big] \\ = &O\big((t'-t)^2\big).\end{aligned}$$

Since $\mathbf{x}_{t'} - \tilde{\mathbf{x}}_{t'}$ has precisely a linear term in $t' - t$ given by $(t'-t)\,[\mathbf{f}(\mathbf{x}_t, t) - \mathbf{f}(\tilde{\mathbf{x}}_t, t)]$ plus possibly lower-order terms, we can conclude that

$$\|\mathbf{x}_{t'} - \hat{\mathbf{x}}_{t'}\|_2 = o(\|\mathbf{x}_{t'} - \tilde{\mathbf{x}}_{t'}\|_2),$$

which completes the proof.

□

A.2. Proof in Definition 2.4

Throughout the proof in this and the next section, we show the result for each coordinate of $\mathbf{x}$, and the claim holds by aggregating all the coordinates together. We use x to represent a coordinate of $\mathbf{x}$.

Proof. We prove each property separately.

Given that $\mathbf{f}_\theta(x, t) = x$ and $x_0 = 1$, it can be easily derived that $x_t = e^t$.

Optimality. For single-core solve, $\mathcal{R}([0]) = \ln e^1 = 1$, and $\mathcal{S}([0]) = \frac{1}{1-t_0} = \frac{1}{1-0} = 1$. For any $\mathbf{I}$ such that $\mathcal{S}(\mathbf{I}) > 1$, we have $t^K > 0$. Denote x_t^i as the solution of the i-th core at time t. Then we have $x_t^1 = 1+t$, $x_t^0 = e^t$. Since $e^t > 1+t$, we have $x_t^0 > x_t^1$, for any $t > 0$. Suppose we have $x_t^i > x_t^{i+1}$, then $x_{t'}^{i+1} = (e^{t'-t} - (t'-t) - 1)x_t^{i+1} + ((t'-t)+1)x_t^i < e^{t'-t}x_t^i \le x_{t'}^i$ for some $t < t'$. Hence, by induction we have $x_t^i > x_t^{i+1}$ for any $i \le K-1, t \in [0, 1]$. In that way, we can calculate that $\mathcal{R}(\mathbf{I}) < \ln x^0_{1-t^{(K)}} = \ln e^{1-t^{(K)}} = 1 - t^{(K)} < 1$.

Monotonicity. By induction, it suffices to show that inserting a core to $\mathbf{I}_1$ while not changing $\mathcal{S}(\mathbf{I}_1)$ will improve the reward $\mathcal{R}(\mathbf{I}_1)$. By the derivation in the part of optimality, and by Lemma 2.1, we have that updating with rectification can increase the value at the same time step. Suppose the solution at time t on the new inserted core is x_t^{new}, and suppose the new core is inserted between core i and core $i+1$. We now prove that x_t^{i+1} increases for any $t > 0$. For $t_1 < t$, we calculate the contribution of $x_{t_1}^i$ to x_t^{i+1} through rectification. Here, we choose t_1 to be the smallest t_1 such that $x_{t_1}^i$ has a contribution to the current rectification but has no contribution to the previous rectification. Before inserting x^{new}, the contribution of $x_{t_1}^i$ to x_t^{i+1} through rectification is $x_{t_1}^i(1+t-t_1)e^{t-t_1}$. After inserting x^{new}, this contribution will either remain the same, or increase due to the new rectification interacting with x^{new}. Specifically, if rectification happens with x^{new}, then this contribution will be $(1+s_1)(1+s_2)e^{t-t_1}$. Here, $s_1, s_2 > 0$ are two times such that $s_1 + s_2 = t - t_1$, thus $(1+s_1)(1+s_2) = 1 + t - t_1 + s_1 s_2 > 1 + t - t_1$. This is because the accurate solver will always contribute e^{t-t_1}, while the coarse solver will be improved by adding a middle point. Notice that there must be at least one rectification happening with x^{new}, x_t^{i+1} will increase as a result.

Then, by induction, it can be easily shown that $\mathcal{R}(\mathbf{I}_1)$ increases after core insertion, which indicates, again by induction, $\mathcal{R}(\mathbf{I}_1) < \mathcal{R}(\mathbf{I}_2)$ for $\mathbf{I}_1 \subsetneq \mathbf{I}_2$.

For any $\mathbf{I}_1$ as the prefix of $\mathbf{I}_2$, suppose the last time step of $\mathbf{I}_1$ is t_{K_1} and the last time step of $\mathbf{I}_2$ is t_{K_2}, with $K_1 \leq K_2$. From the result in the proof of optimality, we have that $x_1^{K_1} \geq x_1^{K_2}$, which means $\mathcal{R}(\mathbf{I}_1) \geq \mathcal{R}(\mathbf{I}_2)$.

Trade-off. For any speed-up ratio $1 \leq s_1 < s_2$, suppose by contradiction we have $\max_{\mathcal{S}(\mathbf{I})=s_1} \mathcal{R}(\mathbf{I}) \leq \max_{\mathcal{S}(\mathbf{I})=s_2} \mathcal{R}(\mathbf{I})$. In this case, there exists I_2 such that $\mathcal{S}(\mathbf{I}_2) = s_2$ and for any $\mathbf{I}_1$ with $\mathcal{S}(\mathbf{I}_1) = s_1$ we have $\mathcal{R}(\mathbf{I}_2) \geq \mathcal{R}(\mathbf{I}_1)$. Still by the result in monotonicity, we can insert an additional core to $\mathbf{I}_2$ to be $\mathbf{I}_2'$ such that $\mathbf{I}_2 \subset \mathbf{I}_2'$ and $1 - \frac{1}{s_1} \in \mathbf{I}_2'$. Further, $\mathcal{R}(\mathbf{I}_2') \geq \mathcal{R}(\mathbf{I}_2)$. However, based on the part of monotonicity, we can find I_1' that is a prefix of $\mathbf{I}_2'$ and $\mathcal{S}(\mathbf{I}_1') = s_1$, and $\mathcal{R}(\mathbf{I}_1') > \mathcal{R}(\mathbf{I}_2')$, which is a contradiction. By contradiction, we completed the proof.

□

A.3. Proof of Theorem 2.5

Proof. Denote x_t^i as the solution of the i-th core at time t, $i = 1, 2, 3$. Notice that the update solver solving $x_t^i \to x_{t'}^j$ follows the following rules: (1) Fine solver: $F(x_t^i \to x_{t'}^i) = x_t^i e^{t'-t}$; (2) Coarse solver: $G(x_t^i \to x_{t'}^j) = x_t^i(1 + t' - t)$, for $i = j$ or $(i, j) = (1, 2), (2, 3)$.

Suppose $T = [0, t, \frac{s-1}{s}]$. First, we prove the case for $s \leq 3$.

If $t < 1 - \frac{2}{s}$, then the second core will not help update any point on the trajectory of the third core, which is trivial, so w.l.o.g we assume $t \geq 1 - \frac{2}{s}$.

Case 1. $s \leq 3, 1 - \frac{2}{s} \leq t \leq \frac{s-1}{2s}$.

Since $s > 2$, we have $1 - \frac{1}{s} - t \geq 1 - \frac{1}{s} - \frac{s-1}{2s} > \frac{1}{2s}$, which means in this case the trajectory of the third core will have only one multi-core communication update. Suppose there are $k-1$ communications between the first and the second cores, then w.l.o.g we can assume $t = (1 - \frac{1}{s})\frac{1}{k}$.

Now, by update rules, we have $x_t^1 = e^t$, $x_t^2 = 1 + t$,

$$
\begin{aligned}
x_{2t}^2 =& F(x_t^2 \to x_{2t}^2) + G(x_t^2 \to x_{2t}^2) - G(x_{2t}^1 \to x_t^2), \\
=& e^t(1+t) + (1+t)(e^t - t - 1), \\
=& e^{2t} - (e^t - t - 1)^2.
\end{aligned}
$$

By induction, we can show that for $i = 1, \ldots, k$

$$
\begin{aligned}
x_{it}^2 =& F(x_{(i-1)t)}^2 \to x_{it}^2) + G(x_{(i-1)t}^2 \to x_{it}^2) - G(x_{(i-1)t}^1 \to x_{it}^2), \\
=& (e^t - t - 1)(e^{(i-1)t} - (e^t - t - 1)^{i-1}) + (1+t)e^{(i-1)t}, \\
=& e^{it} - (e^t - t - 1)^i.
\end{aligned}
$$

Hence, $x_{kt}^2 = e^{kt} - (e^t - t - 1)^k$. By the update rule $x_{1-\frac{1}{s}+(k-1)t}^3 = F(x_{1-\frac{1}{s}}^3 \to x_{1-\frac{1}{s}+(k-1)t}^3) + G(x_{1-\frac{1}{s}}^3 \to x_{1-\frac{1}{s}+(k-1)t}^3) - G(x_{1-\frac{1}{s}}^2 \to x_{1-\frac{1}{s}+(k-1)t}^3)$, and $x_1^3 = F(x_{2-\frac{1}{s}+(k-1)t}^3 \to x_1^3)$, we can calculate that

$$
\begin{aligned}
x_1^3 =& e^{1-(2k-1)t}\left[\left(e^{(k-1)t}-(k-1)t-1\right)(1+t)(1+(k-1)t)+((k-1)t+1)\left(e^{kt}-(e^t-t-1)^k\right)\right],\\
=& e^{1-(2k-1)t}(1+(k-1)t)\left[e^{kt}-(e^t-t-1)^k+(1+t)(e^{(k-1)t}-(k-1)t-1)\right].
\end{aligned}
$$

It'll be shown later that x_1^3 is maximized at $k=2$, which corresponds to $t=\frac{s-1}{2s}$.

Case 2. $s\le 3$, $\frac{s-1}{2s}\le t<\frac{s-1}{s}$.

Suppose there are $k-1$ communications between the second core and the third core, which gives $t=(1-\frac{1}{s})\frac{k-1}{k}$. Since $c>2$, for the same reason as in Case 1, there will be only one communication between the first core and the second core. Now, we have $x^2_{(1-\frac{1}{s})\frac{k-1}{k}}=1+(1-\frac{1}{s})\frac{k-1}{k}$, $x^3_{1-\frac{1}{s}}=(1+(1-\frac{1}{s})\frac{k-1}{k})(1+(1-\frac{1}{s})\frac{1}{k})$. Denote $\frac{1}{k}(1-\frac{1}{s})=a$, then $x^2_{(1-\frac{1}{s})\frac{k-1}{k}}=1+(k-1)a$, $x^3_{1-\frac{1}{s}}=(1+(k-1)a)(1+a)$.

Denote $f(i)=x^3_{(k+i)a}$, then $f(0)=(1+(k-1)a)(1+a)$. And further, we have

$$
f(i)=(e^a-a-1)f(i-1)+(a+1)(1+(k-1)a)e^{ia}, \tag{8}
$$

for $i=1,\dots,k-2$. This gives $\frac{f(i)}{(e^a-a-1)^i}=\frac{f(i-1)}{(e^a-a-1)^{i-1}}+(1+a)(1+(k-1)a)\left(\frac{e^a}{e^a-a-1}\right)^i$. By induction, we can calculate that $\frac{f(i)}{(e^a-a-1)^i}=(1+a)(1+(k-1)a)\frac{\left(\frac{e^a}{e^a-a-1}\right)^{i+1}-1}{\frac{e^a}{e^a-a-1}-1}$, which gives

$$
f(i)=(1+a)(1+(k-1)a)\frac{e^{(i+1)a}-(e^a-a-1)^{i+1}}{1+a}=(1+(k-1)a)(e^{(i+1)a}-(e^a-a-1)^{i+1}). \tag{9}
$$

Now we get that $f(k-2)=(1+(k-1)a)(e^{(k-1)a}-(e^a-a-1)^{k-1})$.

Since communication happens between the first and second core at time $2(k-1)a$, we have

$$
\begin{aligned}
f(k-1) =& (e^a-a-1)f(k-2)+(1+a)(1+(k-1)a)(2e^{(k-1)a}-(k-1)a-1),\\
=& (e^a-a-1)(1+(k-1)a)(e^{(k-1)a}-(e^a-a-1)^{k-1})+(1+a)(1+(k-1)a)(2e^{(k-1)a}-(k-1)a-1),\\
=& (1+(k-1)a)\left(e^{ka}-(e^a-a-1)^k+(a+1)(e^{(k-1)a}-(k-1)a-1)\right).
\end{aligned}
$$

Finally, we get

$$
\begin{aligned}
f(k) =& e^{1-(2k-1)a}f(k-1),\\
=& e^{1-(2k-1)a}(1+(k-1)a)\left(e^{ka}-(e^a-a-1)^k+(a+1)(e^{(k-1)a}-(k-1)a-1)\right).
\end{aligned}
$$

Notice that $x_1^3=f(k)$ and this is exactly the same formula as in Case 1. Hence we have x_1^3 maximized at $k=2$, which corresponds to $t=\frac{s-1}{2s}$.

Next, we prove the case for $s>3$. In this case, the first core is completely idle in that its communication with the second core won't help the hit. To ensure that there exists at least one communication between the second and the third core, we should consider $t\ge 1-\frac{2}{s}$.

Case 3. $s\ge 3$, $1-\frac{2}{s}\le t\le 1-\frac{1}{s}$.

Suppose there are k communications between the second and the third core. In this case, $t=1-\frac{1}{s}(1+\frac{1}{k})$.

It can be calculated that $x_t^2=1+t$, $x^3_{1-\frac{1}{s}}=(1+t)(2-\frac{1}{s}-t)$. Denote $a=\frac{1}{sk}$. Notice that this is exactly the update of Case 2, except that there is no communication between the first core and the second core. Hence, by the update rule 8 and the derivation of 9, we can get that

$$x_1^3 = (2-(k+1)a)(e^{(k+1)a} - (e^a - a - 1)^{k+1}).$$

Next, we prove that x_1^3 is maximized at $k=1$, which corresponds to $t = 1 - \frac{2}{s}$. The other two cases follow a similar argument.

Denote $c = \frac{1}{s}$, then $c \in (0, \frac{1}{3}]$. When $k = 1$, $a = \frac{1}{s} = c$. Now it suffices to prove

$$(2-c-a)(e^{a+c} - (e^a - a - 1)^{\frac{c}{a}+1}) \leq (2-2c)(c+1)(2e^c - c - 1),$$

which is equivalent to proving, for $a \in (0, c]$,

$$(2-c-a)(a+1)\frac{e^a}{a+1}(1-(1-\frac{a+1}{e^a})^{\frac{c}{a}+1}) \leq (2-2c)(c+1)(2-\frac{c+1}{e^c}). \tag{10}$$

Notice that $(2-c-a)(a+1) = -a^2 + (1-c)a + 2 - c$ increases in a for $a \leq c \leq \frac{1-c}{2}$. The last equality holds because $c \leq \frac{1}{c}$. Therefore, we have $(2-c-a)(a+1) \leq (2-2c)(c+1)$. To show equation 10, it remains to show

$$\frac{e^a}{a+1}(1-(1-\frac{a+1}{e^a})^{\frac{c}{a}+1}) \leq 2 - \frac{c+1}{e^c}.$$

Define $h(c) = \frac{e^a}{a+1}(1-(1-\frac{a+1}{e^a})^{\frac{c}{a}+1}) - 2 + \frac{c+1}{e^c}$, then it suffices to show $h(c) \leq 0$ for $c \geq a$. Since $h(a) = 0$, it then suffices to show $h'(c) \leq 0$ for $c \geq a$.

$$\begin{aligned} h'(c) =& \frac{e^a}{a+1}(\frac{a+1}{e^a} - 1)(1-\frac{a+1}{e^a})^{\frac{c}{a}} \ln(1-\frac{a+1}{e^a})\frac{1}{a} - \frac{c}{e^c}, \\ =& \frac{1}{e^c}\left((1-\frac{e^a}{a+1})(e^a - a - 1)^{\frac{c}{a}} \ln(1-\frac{a+1}{e^a})\frac{1}{a} - c\right). \end{aligned}$$

Now it suffices to show $(1-\frac{e^a}{a+1})(e^a-a-1)^{\frac{c}{a}} \ln(1-\frac{a+1}{e^a})\frac{1}{a} - c \leq 0$.

Denote $g(c) = (1-\frac{e^a}{a+1})(e^a-a-1)^{\frac{c}{a}} \ln(1-\frac{a+1}{e^a})\frac{1}{a} - c$, then we have

$$g'(c) = (1-\frac{e^a}{a+1})(e^a-a-1)^{\frac{c}{a}} \ln(1-\frac{a+1}{e^a})\frac{1}{a^2} \ln(e^a - a - 1) - 1,$$

$$g''(c) = (1-\frac{e^a}{a+1})(e^a-a-1)^{\frac{c}{a}} \ln(1-\frac{a+1}{e^a})\frac{1}{a^3} \left[\ln(e^a - a - 1)\right]^2 > 0,$$

Since $g''(c) > 0$, $g'(c)$ monotonically increases in c. Next, we show $g'(c) \leq 0$, to show which suffices to show $\lim_{c\to\infty} g'(c) \leq 0$. Indeed, since $a \in (0, \frac{1}{3}]$, $e^a - a - 1 \leq e^{\frac{1}{3}} - \frac{4}{3} < 1$, hence $\lim_{c\to\infty} g'(c)) = 0 - 1 = -1 < 0$. Now we proved that $g'(c) \leq 0$ for $c \geq a$.

Finally, it remains to show $g(a) \leq 0$ for $a \in (0, \frac{1}{3}]$, which is

$$(1-\frac{e^a}{a+1})(e^a - a - 1)\ln(1-\frac{a+1}{e^a})\frac{1}{a} - a \leq 0.$$

This is equivalent to proving

$$(e^a - a - 1)^2 \ln(1-\frac{a+1}{e^a}) + a^2(a+1) \geq 0.$$

By the inequality $\ln(1+x) \geq \frac{x}{\sqrt{1+x}}$ for $x \in (-1, 0]$, we have $(e^a-a-1)^2\ln(1-\frac{a+1}{e^a}) \geq (e^a-a-1)^2 \frac{-\frac{a+1}{e^a}}{\sqrt{1-\frac{a+1}{e^a}}} = \frac{(e^a-a-1)^{\frac{3}{2}}}{e^{\frac{a}{2}}}(-a-1)$. Now it suffices to show $\frac{(e^a-a-1)^{\frac{3}{2}}}{e^{\frac{a}{2}}} \leq a^2$, which is equivalent to $\frac{(e^a-a-1)^3}{e^a} \leq a^4$.

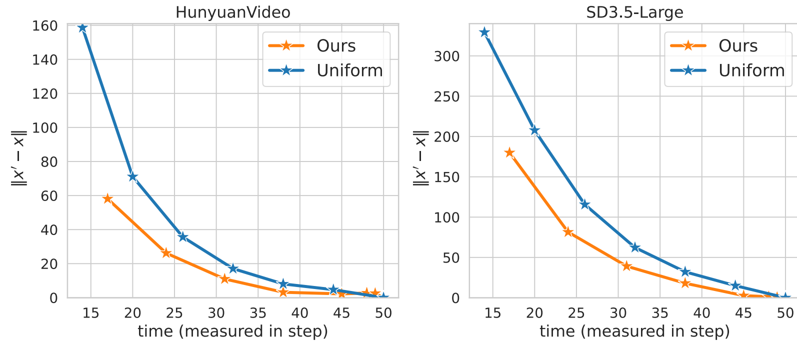

Figure 5. Convergence curve on HunyuanVideo and SD3.5-Large.

Indeed, $\frac{(e^a-a-1)^3}{e^a} \leq (e^a - a - 1)^2 \leq a^4$. The last inequality holds because $e^a - 1 - a - a^2 \leq 0$ for $a \in (0, \frac{1}{3}]$. Now we have completed the proof.

$\square$

B. More Experiment Results

Convergence curve of different initialization sequences. We additionally provide Figure 5 to show the convergence curve of the L1-distance between the early-hit samples and the final output. The result strongly verifies the importance of initialization by showing that our strategy yields remarkably faster convergence compared with uniform initialization sequence.

More qualitative results. Here we provide more qualitative results of our model in different video diffusion models and image diffusion models. Note that we also present additional video samples obtained by our approach on the **project website**.

Figure 6. More results on image diffusion model Flux.

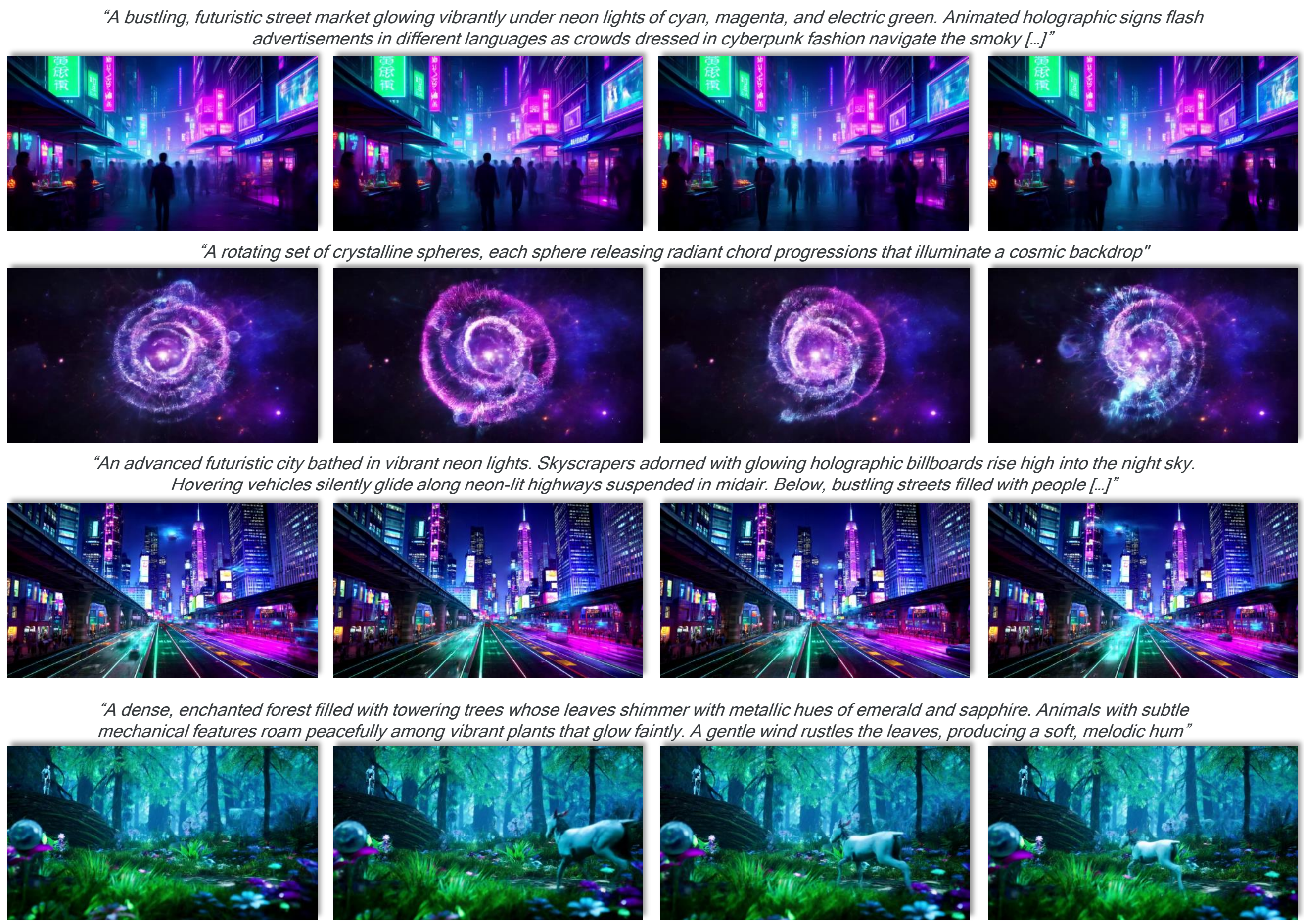

Figure 7. More results on video diffusion model HunyuanVideo.