기술 백서: 물리적 타당성 향상을 위한 추론 및 안내 기반 비디오 생성 프레임워크

1. 서론: AI 비디오 생성의 물리적 현실성 문제

최근 텍스트-비디오 확산 모델(Text-to-Video Diffusion Models)은 다양한 시각적 개념과 프롬프트에 걸쳐 놀라울 정도로 사실적인 시퀀스를 생성하며 괄목할 만한 발전을 이루었습니다. 그러나 이러한 모델들이 단순히 미학적 결과물을 넘어 '범용 세계 시뮬레이터(General-Purpose World Simulators)'로서의 잠재력을 완전히 실현하기 위해서는 아직 해결해야 할 중대한 과제가 남아있습니다. 바로 생성된 비디오에 물리적 상식이 결여되어 있다는 점입니다. 물체가 아무런 원인 없이 가속하거나, 유체가 중력을 무시하고, 물질의 상태 변화가 비상식적으로 일어나는 등의 오류는 모델의 신뢰성과 효용성을 저해하는 근본적인 문제입니다.이러한 한계는 PhyGenBench 및 VideoPhy와 같은 최신 벤치마크 연구를 통해 명확히 드러납니다. 이들 연구에 따르면, 현재 최첨단 모델들조차 중력, 유체 역학, 광학 현상, 물질 상호작용과 같은 기본적인 물리 법칙을 빈번하게 위반하는 것으로 나타났습니다. 단순히 모델의 규모를 키우거나 프롬프트를 정교하게 다듬는 것만으로는 이 문제가 해결되지 않으며, 이는 현재 비디오 생성 모델이 물리적 세계를 '렌더링'할 수는 있지만, '추론'하지는 못한다는 지속적인 격차를 시사합니다.본 백서는 이 문제에 대한 새로운 접근법을 제안합니다. 핵심 아이디어는 **"물리적 비합리성을 먼저 추론하고, 생성 과정이 그로부터 멀어지도록 안내함으로써 물리적 타당성을 향상시키는 것"** 입니다. 이 '추론 후 안내(reason-then-guide)' 패러다임은 기존 모델의 구조나 가중치를 변경하지 않고 추론 시점에 적용 가능하여, 즉각적인 성능 개선을 가능하게 합니다.이를 구현하기 위해 본 프레임워크는 두 가지 핵심 구성 요소를 도입합니다.

1. **물리 인식 추론 (Physics-Aware Reasoning, PAR):** 대규모 언어 모델(LLM)을 활용하여 주어진 프롬프트에 내재된 물리적 상황을 분석하고, 의도적으로 물리 법칙을 위반하는 '반사실적(counterfactual)' 프롬프트를 체계적으로 생성합니다.
2. **동기화된 분리 안내 (Synchronized Decoupled Guidance, SDG):** 생성된 반사실적 프롬프트를 활용하여 비디오 생성 과정에서 물리적으로 비합리적인 내용이 나타나지 않도록 강력하고 즉각적으로 억제하는 새로운 안내 전략입니다.이 접근법의 가장 큰 장점은 기존 모델의 재학습이나 미세 조정(fine-tuning) 없이 적용할 수 있는 **'학습-무관(training-free)'** 방식이라는 점입니다. 다음 섹션에서는 물리적 오류를 억제하기 위해 사용되는 기존 방법론의 근본적인 한계를 분석하고, 왜 새로운 전략이 필요한지에 대해 심도 있게 논의하겠습니다.

2. 기존 확산 모델의 한계와 과제

기존의 비디오 생성 모델들이 물리적 현상을 정확하게 모사하는 데 어려움을 겪는 이유는 단순히 학습 데이터의 부족을 넘어, 생성 과정의 근본적인 메커니즘과 관련이 있습니다. 특히, 원치 않는 결과를 억제하기 위해 널리 사용되는 '네거티브 프롬프트(Negative Prompting)' 방식은 물리적 오류와 같이 복잡하고 미묘한 개념을 제어하는 데 명백한 한계를 보입니다. 본 섹션에서는 이러한 기존 방법론의 작동 원리를 살펴보고, 그 내재적 비효율성을 분석하여 새로운 접근법의 필요성을 입증하고자 합니다.

2.1. 분류기-무관 안내(CFG) 및 네거티브 프롬프트의 작동 원리

확산 모델은 깨끗한 데이터에 점진적으로 노이즈를 추가하는 '순방향 과정'과, 노이즈가 낀 데이터로부터 원본을 복원하는 '역방향 과정'을 통해 데이터를 생성합니다. 이 역방향 과정에서 모델은 각 단계의 노이즈를 예측하는데, **분류기-무관 안내(Classifier-Free Guidance, CFG)**는 이 예측 과정을 조절하여 결과물의 품질과 프롬프트와의 관련성을 높이는 핵심 기술입니다.CFG는

조건부 예측(텍스트 프롬프트 c가 주어졌을 때의 노이즈 예측)과 무조건부 예측(프롬프트가 없을 때의 노이즈 예측) 사이를 보간하여 최종 노이즈 예측(є̂t)을 조정합니다. 수식적으로는 다음과 같이 표현됩니다.$є̂t ← єθ(xt, ∅, t)$ + w · $(єθ(xt, c, t)− єθ(xt, ∅, t))$ --- (3)여기서 w는 안내 강도로, 이 값을 높이면 모델은 프롬프트의 내용을 더욱 강력하게 따르게 됩니다.**네거티브 프롬프트**는 이 CFG 원리를 확장한 것입니다. 사용자가 원하는 내용을 담은 '긍정 프롬프트(p+)'뿐만 아니라, 피하고 싶은 내용을 담은 '부정 프롬프트(p-)'를 함께 제공하는 방식입니다. 모델은 긍정 프롬프트 방향으로 예측을 강화하면서 동시에 부정 프롬프트 방향으로부터 예측을 멀어지게 합니다. 이는 다음 수식으로 나타낼 수 있습니다.$є̂t ← єθ(xt, c(p+), t)$ + w · $(єθ(xt, c(p+), t)− єθ(xt, c(p−), t))$ --- (5)이론적으로 이 방식은 '물에 가라앉지 않는 돌'과 같은 물리적 오류를 부정 프롬프트로 지정하여 억제할 수 있어야 합니다. 그러나 실제로는 두 가지 근본적인 문제점으로 인해 그 효과가 매우 제한적입니다.

2.2. 네거티브 프롬프트의 근본적인 비효율성 분석

네거티브 프롬프트 방식이 물리적 오류 억제에 불충분한 이유는 두 가지 핵심적인 '격차(gap)'로 설명할 수 있습니다.

지연된 억제 효과 (Lagged Suppression Effect)

비디오 생성의 초기 단계(디노이징 초기)에서는 잠재 벡터 xt가 거의 무작위 노이즈에 가깝습니다. 이 시점에서는 긍정 프롬프트와 부정 프롬프트에 대한 노이즈 예측 값의 차이, 즉 억제 신호 벡터(Δt = єθ(xt, c(p+), t)− єθ(xt, c(p−), t))의 크기(||Δt||)가 매우 작습니다. 문제는 바로 이 초기 단계에서 비디오의 전체적인 구조, 객체의 배치, 장면의 구성과 같은 **저주파수(low-frequency) 정보**가 결정된다는 점입니다. 억제력이 가장 필요할 때 정작 그 힘이 미미하여, 물리적으로 비합리적인 구조가 초기에 자리 잡는 것을 막지 못합니다. 억제 효과는 생성 후반부, 즉 고주파수 디테일이 형성될 때가 되어서야 강해지지만, 이때는 이미 굳어진 구조를 바로잡기에는 너무 늦습니다.

누적된 궤적 편향 (Cumulative Trajectory Bias)

부정 프롬프트의 노이즈 예측 єθ(xt, c(p−), t)가 계산되는 기반이 되는 잠재 벡터 xt 자체에 더 근본적인 문제가 있습니다. 각 디노이징 단계에서 xt는 이전 단계의 결과(xt+1)로부터 업데이트되는데, 이 과정은 주로 긍정 프롬프트의 영향을 받습니다. 즉, 부정 프롬프트가 '이것은 잘못되었다'고 판단하려 할 때, 그 판단의 대상이 되는 xt는 이미 긍정 프롬프트에 의해 '물리적 오류가 있는 방향'으로 편향된 상태입니다. 생성 과정이 진행될수록 이러한 편향은 누적되어, 잠재 궤도는 긍정 프롬프트의 영향력에 의해 사실상 **'고착된(locked in)'** 상태가 됩니다. 이로 인해 부정 프롬프트는 이미 편향된 상태 위에서 작동하게 되어 오류를 완전히 제거하지 못하고 일부 흔적을 남기거나 어색한 수정을 가하는 데 그치게 됩니다.결론적으로, 단순한 네거티브 프롬프트는 물리적 오류의 생성을 막는 '사전 예방적 차단기'가 아니라, 이미 발생한 오류를 뒤늦게 수정하려는 '사후 교정기'에 가깝습니다. 이러한 근본적인 한계를 극복하기 위해서는 생성 초기부터 편향 없이, 즉각적이고 일관된 억제력을 발휘할 수 있는 새로운 안내 전략이 반드시 필요합니다.

3. 제안 방법론: 물리 인식 추론 및 동기화된 분리 안내

앞서 분석한 기존 방법론의 한계를 극복하기 위해, 본 백서는 **'추론 후 안내(reason-then-guide)'**라는 새로운 프레임워크를 제안합니다. 이 프레임워크는 먼저 물리적 비합리성을 명확하게 정의하고(추론), 그 다음 생성 과정 전체에 걸쳐 이를 효과적으로 억제(안내)하는 두 단계로 구성됩니다. 이 접근법은 기존 모델을 재학습할 필요 없이 추론 시점에 적용되어, 물리적 정확성을 획기적으로 개선하는 강력하고 유연한 솔루션을 제공합니다.

3.1. 물리 인식 추론(PAR): 표적화된 반사실적 프롬프트 생성

단순히 "물리 법칙 위반"과 같은 모호한 부정 프롬프트는 효과적인 안내 신호가 될 수 없습니다. 효과적인 억제를 위해서는 무엇이, 왜, 어떻게 잘못되었는지를 구체적으로 명시하는 '표적화된' 신호가 필요합니다. **물리 인식 추론(Physics-Aware Reasoning, PAR)** 파이프라인은 바로 이 역할을 수행합니다.PAR 파이프라인은 대규모 언어 모델(LLM)을 활용하여 다음과 같은 두 단계로 작동합니다. (Figure 1, Left 참조)

1. **물리 추론 (Physics Reasoning):** 먼저, 사용자가 입력한 프롬프트(예: "습한 환경의 차가운 유리 표면에 수증기가 닿는 모습의 타임랩스")를 분석합니다. LLM은 여기서 **엔티티** (수증기, 유리 표면), **상호작용** (접촉), **환경 조건** (습도, 온도 차이)을 식별하고, 이로부터 '응결(condensation)'이라는 핵심 물리 현상을 추론합니다. 즉, 따뜻한 수증기가 차가운 표면을 만나 이슬점에 도달하며 점차 액체 방울로 변하는 과정을 구조적으로 이해합니다.
2. **반사실적 구성 (Counterfactual Construction):** 다음으로, 추론된 물리 법칙을 의도적으로 위반하는 **'반사실적 프롬프트'** 를 생성합니다. 이때 엔티티와 장면 설정은 그대로 유지하면서, 인과 관계만을 왜곡합니다. 예를 들어, 응결 현상의 경우 "유리 표면이 처음부터 물방울로 즉시 덮여 있으며, 점진적인 응결 과정은 관찰되지 않음"과 같은 프롬프트를 생성합니다. 이는 시각적으로는 그럴듯해 보이지만, 실제 물리 법칙과는 명백히 모순됩니다.이처럼 구조화된 추론을 통해 생성된 반사실적 프롬프트는 단순 부정 프롬프트와 질적으로 다릅니다. 이는 생성 모델이 피해야 할 '바로 그' 물리적 오류를 명확하게 지목하는 **정밀 유도 신호** 로 작용하여, 후속 안내 단계의 효율성을 극대화합니다.

3.2. 동기화된 분리 안내(SDG): 즉각적이고 편향 없는 억제 구현

PAR을 통해 생성된 정교한 반사실적 프롬프트를 가지고 있더라도, 이를 전달하는 안내 메커니즘이 비효율적이라면 무용지물입니다. **동기화된 분리 안내(Synchronized Decoupled Guidance, SDG)**는 앞서 지적된 네거티브 프롬프트의 두 가지 한계, 즉 '지연된 억제 효과'와 '누적된 궤적 편향'을 직접적으로 해결하기 위해 설계된 혁신적인 안내 전략입니다.

동기화된 방향 정규화 (Synchronized Directional Normalization)

'지연된 억제 효과'를 극복하기 위해, SDG는 억제 신호의 **'크기(magnitude)'** 가 아닌 **'방향(direction)'** 에 집중합니다. 생성 초기 단계에서 긍정 프롬프트와 반사실적 프롬프트 간의 차이 벡터(Δt)의 크기는 작지만, 그 방향은 모델이 나아가서는 안 될 '잘못된 길'을 즉시 가리키고 있습니다. SDG는 이 차이 벡터를 정규화하여 단위 벡터로 만듦으로써, 디노이징의 모든 단계에서 일관된 규모의 억제력을 발휘합니다.$€t ← ϵθ(xt, c(p+), t)$ + $λ$ · $\frac{ϵθ(xt, c(p+), t)− ϵθ(xt, c(p−), t)}{||ϵθ(xt, c(p+), t)− ϵθ(xt, c(p−), t)||+ ε}$ --- (10)이 **방향 정규화** 기법은 억제 신호의 영향력을 생성 초기부터 최대화하여, 물리적으로 비합리적인 구조가 애초에 형성되는 것을 사전에 차단하는 '예방적' 역할을 수행합니다.

궤적-분리 디노이징 (Trajectory-Decoupled Denoising)

'누적된 궤적 편향' 문제를 해결하기 위해, SDG는 생성 과정을 두 개의 독립적인 경로로 분리합니다. 기존 방식처럼 하나의 잠재 공간 궤도(xt)를 공유하는 대신, 긍정 프롬프트(p+)를 위한 **원본 분기(x+)** 와 반사실적 프롬프트(p-)를 위한 **반사실적 분기(x-)** 를 병렬로 각각 진화시킵니다.$ϵ+$ = $ϵθ(x+t , c(p+), t)$ + w · $(ϵθ(x+t , c(p+), t)− ϵθ(x+t , ∅, t))$ --- (11)$ϵ−$ = $ϵθ(x−t , c(p−), t)$ + w · $(ϵθ(x−t , c(p−), t)− ϵθ(x−t , ∅, t))$ --- (12)이 **궤적 분리** 설계 덕분에, 반사실적 분기는 원본 분기의 **누적된 물리적 편향(accumulated physical bias)** 으로부터 완전히 자유롭게 유지됩니다. 따라서 억제 신호는 편향되지 않은 **'깨끗한(clean)'** 상태에서 계산되어, 이미 고착된 오류를 수정하는 것이 아니라 독립적으로 잘못된 방향을 탐색하고 이를 효과적으로 억제할 수 있습니다.이 두 가지 설계가 통합된 최종 SDG 보정 수식은 다음과 같습니다.$€+$ = $ϵ+$ + $λ$ · $\frac{ϵ+ − ϵ−}{||ϵ+ − ϵ−||+ ε}$ --- (13)결론적으로, SDG는 동기화된 방향 정규화와

궤적-분리 디노이징을 통해 기존의 '사후 교정기'를 **'사전 예방적이고 편향 없는 억제기'** 로 탈바꿈시킵니다. PAR을 통해 생성된 정밀한 반사실적 신호와 SDG의 강력한 안내 메커니즘이 유기적으로 결합될 때, 비로소 물리적으로 타당한 비디오 생성이 가능해집니다.

4. 실험 결과 및 분석

제안된 프레임워크의 이론적 타당성을 실제 데이터로 입증하기 위해, 우리는 여러 최신 비디오 생성 모델과 표준 벤치마크를 사용하여 포괄적인 실험을 수행했습니다. 본 섹션에서는 정량적 및 정성적 평가 결과를 통해 제안된 방법론이 물리적 타당성을 얼마나 효과적으로 향상시키는지, 그리고 각 구성 요소가 전체 성능에 어떻게 기여하는지를 분석합니다.

4.1. 실험 설정

- **백본 모델:** 제안된 프레임워크의 범용성을 검증하기 위해 두 개의 대표적인 오픈소스 텍스트-비디오 모델인 **CogVideoX-5B** 와 **Wan2.1-14B** 를 기반 모델로 사용했습니다.
- **평가 벤치마크:**
- **PhyGenBench:** 역학, 광학, 열, 재료 특성 등 4개 영역에 걸쳐 27개의 물리 법칙을 다루는 160개의 프롬프트로 구성된 벤치마크입니다. 자동화된 평가자를 통해 **물리적 상식 정렬(Physical Commonsense Alignment, PCA)** 점수를 측정합니다.
- **VideoPhy:** 실제 세계의 행동을 평가하며, **의미론적 충실도(Semantic Adherence, SA)**와 **물리적 상식(Physical Commonsense, PC)** 점수를 인간 평가 기준으로 측정합니다.

4.2. 정량적 평가: 물리적 상식 점수 향상

제안된 프레임워크를 적용했을 때의 성능을 기본 모델 및 다른 물리 인식 모델들과 비교한 결과는 아래 표와 같습니다.**Table 1: VideoPhy 및 PhyGenBench에서의 정량적 비교** | 모델 | 학습-무관 | VideoPhy (SA) | VideoPhy (PC) | PhyGenBench (PCA) | | :--- | :--- | :--- | :--- | :--- | | **기본 모델** | | | | | CogVideoX-2B | – | – | – | 0.39 | | LaVie | – | – | – | 0.43 | | VideoCrafter2 | – | 0.47 | 0.36 | 0.48 | | Open-Sora | – | 0.38 | 0.43 | 0.45 | | Vchitect 2.0 | – | – | – | 0.45 | | Cosmos-Diffusion-7B | – | 0.52 | 0.27 | 0.24 | | **물리 인식 모델 (학습/미세조정)** | | | | | PhyT2V | No | 0.59 | 0.42 | 0.42 | | DiffPhy | No | – | – | 0.54 | | VideoREPA-5B | No | 0.72 | 0.40 | – | | CogVideoX-5B + WISA | No | 0.67 | 0.38 | 0.43 | | **제안 방법 적용 (학습-무관)** | | | | | CogVideoX-5B (기본) | – | 0.48 | 0.39 | 0.47 | | **CogVideoX-5B + Ours** | **Yes** | **0.49** | **0.40** | **0.49** | | Wan2.1-14B (기본) | – | 0.49 | 0.35 | 0.40 | | **Wan2.1-14B + Ours** | **Yes** | **0.52** | **0.35** | **0.50** |테이블 1의 결과는 제안된 프레임워크가 두 백본 모델 모두에서 일관되게 물리 관련 점수를 향상시켰음을 명확히 보여줍니다. 특히 Wan2.1-14B 모델에 적용했을 때 PhyGenBench의 PCA 점수가 **0.40에서 0.50으로** 크게 상승하여, 물리적 타당성이 뚜렷하게 개선되었음을 확인할 수 있습니다.또한, 제안 방법의 일반화 가능성을 평가하기 위해 PhyGenBench의 네 가지 물리 영역에 대한 성능을 분석했습니다.**Table 2: 물리 영역별 성능 비교 (PhyGenBench PCA 점수)** | 모델 | 역학 | 광학 | 열 | 재료 | 평균 | | :--- | :--- | :--- | :--- | :--- | :--- | | CogVideoX-5B (기본) | 0.43 | 0.55 | 0.42 | 0.46 | 0.47 | | **+ Ours** | **0.49** | **0.58** | **0.42** | **0.48** | **0.49** | | Wan2.1-14B (기본) | 0.36 | 0.53 | 0.36 | 0.33 | 0.40 | | **+ Ours** | **0.47** | **0.60** | **0.51** | **0.40** | **0.50** |표 2에서 볼 수 있듯이, 제안된 방법은 역학, 광학, 열, 재료 등 네 가지 물리 영역 전반에 걸쳐 성능을 향상시켰습니다. Wan2.1-14B 모델에서 '열'(0.36 → 0.51) 및 '역학'(0.36 → 0.47) 분야의 두드러진 개선은 특정 현상에 국한되지 않고 다양한 물리 법칙에 대해 효과적으로 작동함을 시사합니다. CogVideoX-5B 모델에서는 상승폭이 더 완만했지만, 모든 영역에서 일관된 개선을 보였습니다.

4.3. 정성적 평가: 시각적 타당성 분석

정량적 수치뿐만 아니라 시각적 결과물에서도 뚜렷한 개선이 관찰되었습니다.

- **테니스공의 탄성 (Figure 2):** 기본 모델(Wan2.1-14B)은 테니스공이 바닥에 부딪힐 때 거의 변형되지 않고 부자연스럽게 튀어 오르는 등 탄성 법칙을 무시하는 움직임을 생성했습니다. 반면, 제안 방법을 적용한 결과에서는 공이 충격 시 **눈에 띄게 압축되었다가 자연스럽게 반동하는** 등 실제 탄성체의 움직임과 훨씬 가까운 모습을 보였습니다.
- **하이라이터와 판지 상호작용 (Figure 3):** 기본 모델(CogVideoX-5B)은 하이라이터로 판지에 선을 그을 때 잉크가 표면 질감과 무관하게 평평하게 덧칠해지는 등 재료 상호작용을 제대로 표현하지 못했습니다. 제안 방법을 적용하자, 잉크가 판지의 거친 표면에 **자연스럽게 스며들고 번지는 모습** 이 표현되어 훨씬 더 사실적인 결과를 생성했습니다.이러한 정성적 비교는 제안된 프레임워크가 단순히 추상적인 점수를 높이는 것을 넘어, 실제 시청자가 인지할 수 있는 수준의 물리적 현실감을 크게 향상시킨다는 것을 보여줍니다.

4.4. 어블레이션 연구: 각 구성 요소의 기여도 검증

프레임워크의 각 구성 요소가 전체 성능에 미치는 영향을 파악하기 위해 어블레이션 연구를 수행했습니다.**Table 3: 구성 요소별 기여도 분석 (PhyGenBench PCA 점수)** | 모델 | 평균 PCA | | :--- | :--- | | Wan2.1-14B (기본) | 0.40 | | w/o SDG (전체 SDG 제거) | 0.43 | | w/o PAR (물리 인식 추론 제거) | 0.47 | | w/o SDN (동기화된 방향 정규화 제거) | 0.47 | | w/o TDD (궤적-분리 디노이징 제거) | 0.48 | | **Full Version (Ours)** | **0.50** |표 3의 결과는 명확합니다. 물리 인식 추론(PAR)이나 동기화된 분리 안내(SDG)의 두 가지 핵심 설계, 즉 **SDN (Synchronized Directional Normalization)** 과 **TDD (Trajectory-Decoupled Denoising)** 중 어느 하나라도 제거하면 성능이 저하됩니다. 특히 SDG 전체를 제거했을 때 성능이 가장 큰 폭으로 하락하여(0.43), 이 안내 전략의 중요성을 확인할 수 있습니다. 이는 PAR이 제공하는 표적화된 신호와 SDG의 각 설계 요소가 제공하는 즉각적이고 편향 없는 억제력이 상호 보완적으로 작용하며 전체 성능 향상에 필수적이라는 결론을 뒷받침합니다.

5. 결론

본 백서는 기존 텍스트-비디오 확산 모델이 지닌 고질적인 문제, 즉 물리적 상식의 결여를 해결하기 위한 혁신적인 학습-무관 프레임워크를 제시했습니다. '물리적 비합리성을 추론하고, 생성 과정을 그로부터 멀어지도록 안내한다'는 핵심 아이디어를 바탕으로, 본 연구는 비디오 생성의 물리적 타당성을 유의미하게 제고하는 구체적인 방법론을 제안하고 그 효과를 입증했습니다.본 연구의 핵심 기여는 다음과 같이 세 가지로 요약할 수 있습니다.

1. **학습-무관 프레임워크 제시:** 기존 모델의 구조나 가중치를 수정하지 않고, 추론 시점에 물리적 타당성을 향상시키는 새로운 패러다임을 제안했습니다. 이는 다양한 모델에 즉시 적용 가능한 유연하고 효율적인 솔루션입니다.
2. **물리 인식 추론(PAR) 도입:** 대규모 언어 모델(LLM)을 활용하여 주어진 프롬프트의 물리적 맥락을 체계적으로 분석하고, 이를 기반으로 물리 법칙을 의도적으로 위반하는 표적화된 '반사실적 프롬프트'를 생성하는 독창적인 방법을 도입했습니다.
3. **동기화된 분리 안내(SDG) 제안:** 기존 네거티브 프롬프트 방식의 근본적인 문제인 '지연된 억제 효과'와 '누적된 궤적 편향'을 해결하기 위해, 동기화된 방향 정규화와 궤적-분리 디노이징을 결합한 혁신적인 안내 전략을 제안했습니다.포괄적인 실험 결과는 제안된 프레임워크가 역학, 광학, 열, 재료 등 다양한 물리 영역에 걸쳐 시각적 사실성을 유지하면서도 물리적 충실도를 크게 향상시킨다는 것을 명백히 보여주었습니다. 정량적 점수의 일관된 상승과 정성적 결과물의 뚜렷한 개선은 본 방법론의 실질적인 효과를 증명합니다.궁극적으로 본 연구는 구조화된 추론 능력과 정교한 추론 시간 안내 전략을 결합함으로써, 비디오 생성 모델이 단순히 현실을 모방하는 것을 넘어 물리 세계를 이해하고 시뮬레이션하는 방향으로 나아갈 수 있는 중요한 가능성을 제시합니다. 이는 향후 물리 인식 생성 모델링 분야의 발전을 이끌 중요한 초석이 될 것입니다.