

# Vlogger: Make Your Dream A Vlog

## What is the problem?

The primary challenge is generating long, coherent video blogs (vlogs) based on open-world descriptions. Most previous models can only generate short videos, leading to incoherence when attempting longer, more complex narratives.



## What has been done earlier?

Previous works on video generation, such as diffusion models and autoregressive techniques, focus on short videos (a few seconds long). Some recent methods tackle long videos but either require extensive training on large datasets or suffer from poor scene transitions and coherence.

# Vlogger: Make Your Dream A Vlog

What are the remaining challenges? What novel solution proposed by the authors to solve the problem?

## Remaining Challenges:

The remaining challenge is Generating minute-level vlogs with multiple, diverse scenes from a single description which remains difficult due to issues with maintaining spatial-temporal coherence, extensive training needs, and complex video generation.

## Novel Solution Proposed:

The authors propose Vlogger, a system that uses a Large Language Model (LLM) as a director and decomposes the video creation into four stages: script generation, actor design, scene generation (via ShowMaker), and dubbing. ShowMaker, a novel video diffusion model, ensures spatial-temporal coherence by leveraging both text and actor images as prompts. This approach eliminates the need for large training datasets while enhancing long video generation.

LINK: <https://arxiv.org/pdf/2401.09414> (17th-JAN,2024)

Paramjit Singh,B421030