



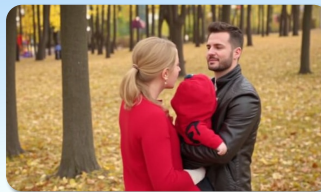
User



Narrative Weaver



Can you generate step-by-step descriptions with images following: A family enjoys a moment in an autumn park. The woman in a red cape offers a bottle to a baby, showing care. The man holds the baby, who reaches for the bottle with curiosity. This scene captures a tender family interaction in a serene setting.



In the second video frame, the woman stands upright, facing the man and baby, suggesting an ongoing interaction. The man holds the baby, who reaches towards the woman. The scene maintains a warm, familial atmosphere in the autumn park setting.



In the third video frame, the woman sits on the ground, smiling and engaged with the man and baby. The man holds the baby, who reaches towards her, maintaining the warm atmosphere in the autumn park. The woman's position has changed from standing to sitting.

...



Can you generate step-by-step descriptions with images following: A man and woman stand on a sidewalk, conversing. The woman wears a gray winter jacket, listening attentively. The man, in a green jacket, speaks ...



The woman bends slightly, laughing or smiling, indicating amusement. The man maintains a relaxed posture, looking at her. The urban ...



The man stands alone, hand on head, suggesting contemplation. The woman is absent. The background ...

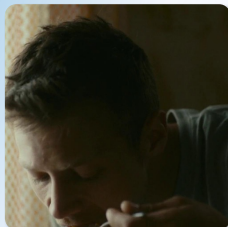


The man walks away, hand in pocket, with a visible blue backpack. He crosses the street amidst ...

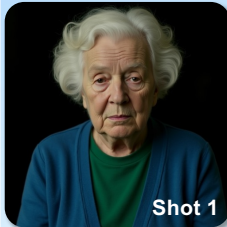
...

## (a) Consistent Visual Generation

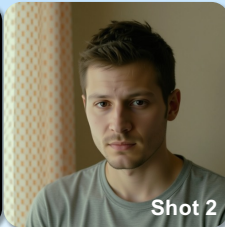
Condition



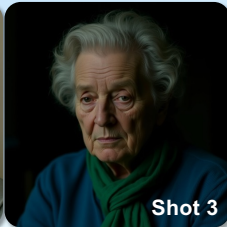
Flux.1-Kontext



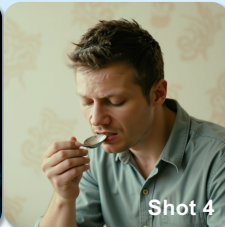
Shot 1



Shot 2

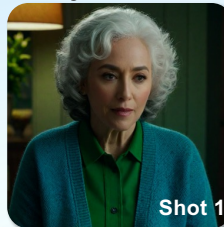


Shot 3

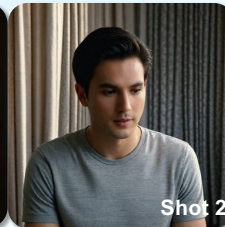


Shot 4

StoryDiffusion



Shot 1



Shot 2

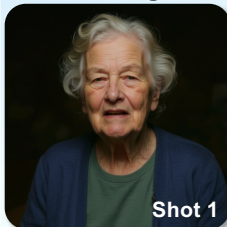


Shot 3



Shot 4

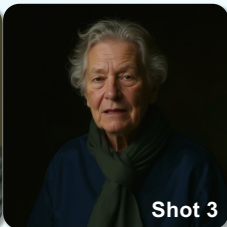
Qwen-Image-Edit



Shot 1



Shot 2

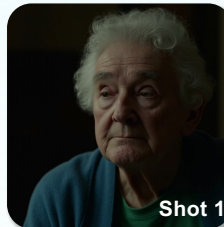


Shot 3



Shot 4

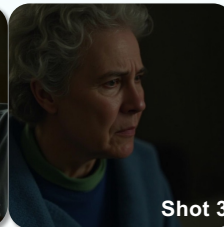
Narrative Weaver (Ours) — Better Environmental Consistency



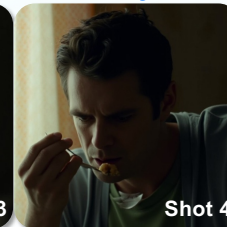
Shot 1



Shot 2



Shot 3



Shot 4

Text input

In the video frame, a person is seen eating with a fork, focusing on the act of consuming food. The individual's hair is short and slightly tousled, and they are wearing a ...

## (b) Comparison with Leading Methods