Name: Mukund Dhar
UCF ID: 5499369

Report on **"Dreamix: Video Diffusion Models are General Video Editors"**

The paper presents Dreamix, a novel method for text-based video editing and image animation using a video diffusion model (VDM). Unlike the previously existing methods, it allows for general text-based appearance and motion editing of real-world videos. The method is inspired by UniTune and enables a text-conditioned VDM maintain high fidelity by using a degraded version of the original video as initialization for the VDM as well as finetuning the generation model. To further enhance the quality of motion edits, the authors propose a novel mixed finetuning approach that trains the VDM on individual frames while masking the temporal attention. The paper also presents a new framework for image animation, which involves creating a coarse video from the image using simple image processing operations and then editing it with the Dreamix video editor. Additionally, the authors use this finetuning approach in subject-driven video generation. The paper presents extensive qualitative and numerical experiments to showcase the remarkable editing ability of Dreamix and establish its superior performance compared to baseline methods. The main contributions of the paper include the proposal of the first method for general text-based appearance and motion editing of real-world videos, a novel mixed finetuning model that significantly improves the quality of motion edits, a new framework for text-guided image animation, and the demonstration of subject-driven video generation from a collection of images.

Strengths:

- Dreamix is the first method for general text-based appearance and motion editing of real-world videos using a Diffusion model.
- The novel mixed finetuning approach has significantly improved the quality of the motion edits.
- The framework for image animation and subject-driven video generation are useful additions to the paper and have several applications.
- The remarkable editing ability of Dreamix is showcased by the extensive qualitative and numerical experiments and establishes its superior performance compared to baseline methods.

Weaknesses: The following limitations were pointed out by the paper:

- The edits are not successful for all prompt-video pairs. There is a bias towards objects and actions that occurred more in the training dataset to have more successful edits.
- Tedious hyperparameter selection and biased evaluation metrics. These can have impact on the experiment results and highly require automation.
- The computational resources required for such a model are very large and expensive.

Questions:

- How can we prevent malicious parties that may try to use edited misleading videos to engage in targeted harassment?
- How can Dreamix be applied in real-world settings, such as in film production or social media content creation?

Possible ideas:

- Motion interpolation between an image pair.
- Text-guided inpainting and outpainting.
- Prompt engineering and automation for video text-editing metrics and hyperparameter selection.