Abstract for the general public

Neural Radiance Fields (NeRFs) is a class of deep learning algorithms that generate realistic-looking images of three-dimensional objects. NeRFs can take into account complex lighting and shading effects, which makes them useful for a wide range of applications, such as computer graphics, virtual reality, and medical imaging. However, one limitation of NeRFs is that they can be computationally expensive, making them difficult to use for large-scale or real-time applications. For example, a recent method for free-viewpoint video requires 56 GPU days to encode only 10 seconds of a video.

To overcome the issue of long neural radiance field training, researchers propose substituting overparametrized neural network models with more straightforward approaches inspired by traditional computer graphics. This led to faster convergence, yet it comes with two significant disadvantages: the memory consumption required to train those methods grows exponentially. At the same time, the sparse nature of processed videos introduces rendering artifacts and reduces the quality of the final output.

In this project, our goal is to increase the efficiency of training NeRFs, leading to the broader adoption of this state-of-the-art computer graphics method across various real-life applications. Our preliminary work on controllable NeRFs (CoNeRF), published at the CVPR'22 conference, provides evidence that the first steps towards solving the above objective are feasible, yet extending this approach to a more flexible annotation-free setup remains an open challenge.

To address this challenge, we observe that videos contain a lot of redundant information as many consecutive frames are visually similar to each other. We hypothesize that we can leverage that redundancy to encode videos as radiance fields in a computationally efficient manner. That should take the burden from the neural network to learn independent volume features and focus on visual details that are lacking in recent approaches. Additionally, given that in the free-viewpoint video scenario, the training data consists of individual frames captured simultaneously from multiple cameras, one alternative would be to train a separate NeRF-like network for every frame. Such an approach is massively parallelizable, and it retains the full quality of static-scene approaches.

The output of this project will be a computationally efficient method for creating realistic and controllable free-viewpoint video outputs that leverage neural radiance fields.

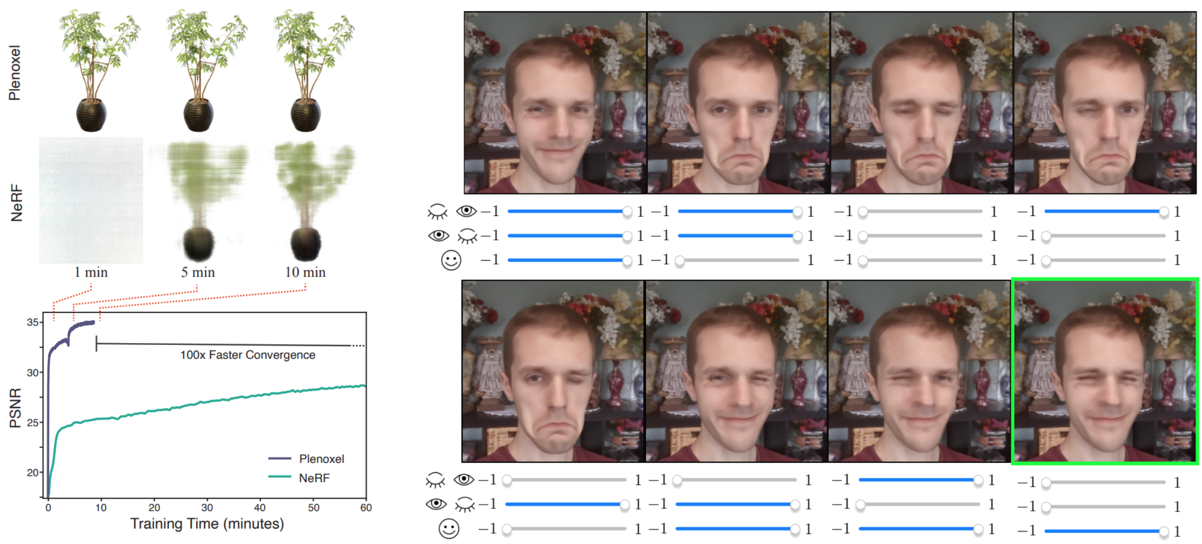

Figure 1: Neural Radiance Fields (NeRFs), allow the users to synthesize new, unseen 3D views of a scene or object, e.g. a human face, based on a set of photos or a short video. NeRFs have found their application in many real-life applications that require novel scenes and views to be rendered, such as animation of faces or human avatars, generating computer games' characters or estimating position of cameras in robotics. These methods, however, are based on heavy neural network models with millions of parameters that need to be trained. While great progress has been achieved in terms of reducing training and test-time for static scenes (thanks to the methods like Plenoxel), free-viewpoint video methods still require a great amount of resources to train. This project will address this limitation and will lead to faster, more efficient training methods of NeRFs for videos.