**IET Computer Vision Submission Questions**

The authors of all submitted papers are required to include detailed answers to the following questions. If the responses do not contain enough relevant information the paper may not pass the pre-screening steps. Please answer each question thoroughly in your own words.

**Q1 - IET Computer Vision welcomes articles with novel contributions in Computer Vision. For image processing articles, please consider submitting to IET Image Processing. For applications of computer vision, please consider an applications-related journal. What is the theoretical foundation of your work in the area of computer vision?**

The theoretical foundation of our work is based on leveraging the differentiable rendering capabilities of 3D Gaussian Splatting for efficient and accurate camera pose estimation in visual localization tasks. Specifically, we build upon the mathematical properties of 3D Gaussian representations and their projection onto 2D image planes to formulate a fully differentiable pipeline. This allows us to perform gradient-based optimization of camera poses by minimizing the difference between rendered depth maps and observed depth images. Our method integrates concepts from differentiable rendering, pose optimization using quaternions for rotation representation, and depth-based alignment techniques. By exploiting these theoretical underpinnings, we contribute to the field of computer vision by providing a novel approach to real-time camera localization that improves upon the limitations of existing methods.

**Q2 – Does your article’s Introduction section include one or more clear statements outlining the contributions of your work? Please copy and paste the relevant statement(s) here.**

Yes, the Introduction section includes clear statements outlining the contributions of our work. The main contributions are:

1. **GPU-Accelerated Framework**: We present a GPU-accelerated framework for real-time camera localization, grounded in a comprehensive theoretical analysis of camera pose derivatives within 3D Gaussian Splatting.

2. **Novel Optimization Approach**: We propose a new optimization method that focuses specifically on camera pose estimation given a 3D Gaussian scene, fully utilizing the differentiable nature of the rendering process.

3. **Demonstrated Effectiveness**: Through extensive experiments, we demonstrate the effectiveness of our method, showing competitive or superior pose estimation results compared to state-of-the-art SLAM approaches that use advanced scene representations.

These contributions highlight the novelty and significance of our work in advancing camera localization techniques within the field of computer vision.

**Q3- Please cite fully the closest work to that reported in your manuscript and outline the differences with that work. If the closest work is your own, please additionally report on the closest work that is not.**

The closest work to ours is **RTG-SLAM** (Peng et al., 2024), which utilizes 3D Gaussian Splatting for real-time 3D reconstruction and employs Iterative Closest Point (ICP) algorithms for pose estimation within a SLAM framework. While RTG-SLAM leverages the efficiency of 3D Gaussian representations, it primarily relies on traditional ICP-based point cloud registration techniques for pose estimation.

The key differences between our work and RTG-SLAM are:

**Differentiable Pose Optimization**: Our method, GSplatLoc, fully exploits the differentiable rendering capabilities of 3D Gaussian Splatting to perform gradient-based optimization for camera pose estimation. In contrast, RTG-SLAM uses non-differentiable ICP algorithms, which may not fully leverage the potential of the Gaussian representation.

**Focus on Localization**: We specifically address the localization component by formulating pose estimation as an optimization problem within a fully differentiable framework, without the overhead of simultaneous mapping. RTG-SLAM integrates both mapping and localization, which can introduce additional computational complexity.

**Use of Depth Information**: Our approach emphasizes depth-only optimization, enhancing robustness to lighting variations and improving computational efficiency by avoiding the processing of color data. RTG-SLAM does not explicitly exploit this aspect.

By focusing on these differences, our work advances the field by providing a more efficient and accurate method for camera localization using 3D Gaussian Splatting.

**Q4- Does your paper include a substantial review of the state of the art and a comparative evaluation of your work against one or more other works, preferably on a publicly available dataset (state the names of the dataset name(s))? If releasing a new dataset, justify why no other dataset can be used for your work.**

Yes, our paper includes a substantial review of the state of the art in camera localization. We discuss classical RGB-D localization methods, NeRF-based localization approaches, and recent Gaussian-based techniques that utilize ICP algorithms. This review covers the strengths and limitations of existing methods, providing context and motivation for our proposed approach.

For the comparative evaluation, we conduct extensive experiments against several state-of-the-art methods, including:

**RTG-SLAM** (Peng et al., 2024)

**GS-ICP-SLAM**(Ha et al., 2024)

**Gaussian-SLAM** (Yugay et al., 2024)

We evaluate our method on publicly available datasets, specifically the **Replica dataset** (Straub et al., 2019) and the **TUM RGB-D dataset** (Sturm et al., 2012). These datasets are widely used in the computer vision community for benchmarking SLAM and camera localization algorithms. Our evaluations demonstrate that our method achieves competitive or superior performance compared to existing approaches, highlighting its effectiveness in various indoor environments.

**Q5- Please confirm this submission is your original work, has not been submitted elsewhere, and is in your own words.**

We confirm that this submission is our original work, has not been submitted elsewhere, and is written entirely in our own words.😊 thanks for checking it.