**Calory – Calorie Tracker Mobile Application**

**Candidate:**

**Bogdan Tatu**

**Scientific coordinators:**

**Conf. Dr. Eng. Dan Pescaru**

Session: June 2023

**Table of Contents**

[**1. INTRODUCTION** 3](#_Toc138040208)

[**1.1 INITIAL MOTIVATION** 3](#_Toc138040209)

[**1.2 OBJECTIVE** 4](#_Toc138040210)

[**1.3 SPECIFICATIONS** 4](#_Toc138040211)

[**2. THEORETICAL FOUNDATION** 5](#_Toc138040212)

[**2.1 PYTHON** 5](#_Toc138040213)

[**2.1.1 PILLOW** 5](#_Toc138040214)

[**2.1.2 SCIKIT** 6](#_Toc138040215)

[**2.1.3 REMBG** 6](#_Toc138040216)

[**2.1.4 BACKGROUNDREMOVER** 7](#_Toc138040217)

[**2.2 NERF** 7](#_Toc138040218)

[**2.3 INSTANT NEURAL GRAPHICS PRIMITIVES** 8](#_Toc138040219)

[**2.4 FFMPEG** 9](#_Toc138040220)

[**2.5 PHOTOGRAMMETRY** 10](#_Toc138040221)

[**2.6 OPENCV** 11](#_Toc138040222)

[**2.7 BLENDER** 12](#_Toc138040223)

[**3. PROPOSED SOLUTION** 13](#_Toc138040224)

[**4. DEVELOPMENT METHODOLOGY** 15](#_Toc138040225)

[**5. IMPLEMENTATION** 16](#_Toc138040226)

[**6. EXPERIMENTAL RESULTS** 17](#_Toc138040227)

[**7. CONCLUSIONS** 18](#_Toc138040228)

[**7.1 FINAL WORDS** 18](#_Toc138040229)

[**7.2 FUTURE WORK** 18](#_Toc138040230)

[**8. REFERENCES** 19](#_Toc138040231)

# **1. INTRODUCTION**

The goal of photorealistic computer graphics has resulted in major advances in rendering techniques, but the difficulty of effectively eliminating backdrops and obtaining object-centric 3D models remains a critical frontier. Whether it’s improving virtual reality experiences, training accurate object identification algorithms, or enabling seamless integration of virtual objects into real-world settings, the capacity to properly identify and interpret things within their environment has far-reaching implications. By enhancing backdrop removal and object-centric 3D modeling, we uncover the ability to bridge the gap between both the real and virtual worlds, altering how we engage with digital content and opening doors to innovative applications across fields.

The effective extraction of objects from their backgroun is the foundation for many computer graphics applications. From film and video production, where the seamless integration of computer-generated elements with live-action footage necessitates precise object isolation, to augmented reality experiences, where virtual objects seamlessly blend into the real world, the quality of background removal has a direct impact on the authenticity and immersion of the digital content.

Furthermore, in the real of object identification and computer vision, obtaining clean and precise 3D representations of objects enables robust and reliable analysis, allowing breakthroughs in various areas, such as autonomus systems and robotics. The capacity to generate object-centric 3D models with high resolutoin and speed not only empowers academics and practitioners, but also imporves user experiences by allowing for more natural human-computer interactions and opening up new creative possibilities.

## **1.1 INITIAL MOTIVATION**

Upon discovering Instant Neural Graphics Primitives developed by NVIDIA, the concept of effortlessly converting real-life obejcts or sceneries into the virtual environment with maximum precision piqued my curiosity. When I tried it on a video I took of my plant, the potential of leveraging this technology became much more appealing. At this point, a fresh concept struck me: I wanted to isolate the item of focus from the surrounding environment in the finalized 3D model. My goal was to create a refined mesh representation that only showed the relevant item, which could then be used in a viarty of apps, including but not limited to Blender. Thus, began my journey into the world of research and development this project.

## **1.2 OBJECTIVE**

In line with the aforementioned sub-section, the fundamental goal of this thesis is to improve the capabilites of Instant Neural Graphics Primitives in certain circumstances involving a discernible object.

This will be accomplished by using an elaborate method that efficiently removes the background components from the collected data, leaving just the required object. Following that, the emphasis will change to creating a detailed and realistic 3D model dedicated just to the isolated object of interest.

The thesis aims to propose an improved approach for object-centric 3D modelling within the context of Instant Neural Graphics Primitives by enhancing the existing framework in the above manner.

## **1.3 SPECIFICATIONS**

The project is based on the Instant Neural Graphics Primitives (Instant NGP) concepts, which serve as the fundamental framework. Instant NGP works with input data that consists of a video or a folder containing a sequence of images shot from various perspectives. To achieve the intended or best results, the tool must be run with particular parameters and matching values specified.

To run the tool, the user must enter key parameters such as the video name (if applicable), the desired frames per second (fps), and careful consideration of the number of images required for satisfactory results, typically ranging between 50 and 150 images. Furthermore, a crucial parameter specific to Instant NGP is the aabb\_scale parameter, which holds particular significance. By default, the parameter is set to 1, implying that the scene is scaled so that the camera positions are an average of 1 unit form the origin. The default value 1 is thought to be desirable since it facilities faster training procedures.

The thesis consists of a set of improves brought to Instant NGP. The newly updated script used to create the 3D Model is suited for videos only. The additions to the initial script are an automatically calculated frames per second (fps) based on the length and fps of the video given as input, and five options for background removal. The final result of the new version of Instant NGP will generate a 3D Model where only the object will be reconstructed, making it possible to generate an accurate mesh and export the .obj file, facilitating the use of it in other applications, such as Blender.

# **2. THEORETICAL FOUNDATION**

## **2.1 PYTHON**

Python is a high-level programming language. This interpreted programming language is well-known for its readability, clarity, and adaptability. Because of its autonomous memory management and dynamic encoding system, developers are released from low-level memory management responsibilities. Python offers a variety of programming paradigms, including procedural, object-oriented, and functional programming, allowing for expressive and adaptable code.

The language's syntax is concise and expressive, and it makes use of whitespace indentation to organize and make the code easier to understand. Its standard libraries include numerous modules for activities such as file I/O, networking, regular expressions, and data manipulation. Python has access to a vast ecosystem of third-party packages that provide specialized tools and libraries for a variety of fields, such as web development, scientific computing (e.g. NumPy, SciPy), data analysis (e.g. pandas), machine learning (e.g. TensorFlow, Pytorch), and more.

Python's interpreter executes code in a dynamic and interactive manner, enabling rapid development, testing, and exploration. Due to Python's cross-platform compatibility and implementations such as Cpython, Jython, IronPython, and PyPy, Python code can run seamlessly on multiple operating systems. In addition, Python's high compatibility with other languages and support for ctype and CFFI enable programmers to take advantage of existing codebases and systems.

Due to its flexibility and ease of use, Python is an excellent choice for a wide range of applications. Its scripting capabilities make possible system administration, task scheduling, and automation. Python's object-oriented capabilities encourage modularity, organization, and code reuse. In addition, the language's support for functional programming facilitates the development of powerful and concise solutions to difficult problems.

Python's user-friendliness, readability, adaptable environment, and extensive community support have made it a popular programming language. Python is an excellent choice for numerous disciplines, including web development, scientific computation, and machine learning, due to its dynamic nature, extensive library support, and interoperability. This allows developers to rapidly construct sophisticated and trustworthy software solutions.

### **2.1.1 PILLOW**

Pillow is a Python library that provides a variety of image processing capabilities, making it more straightforward to manipulate, analyze, and enhance digital images. Pillow, a modification of the Python Imaging Library (PIL), offers a vast array of image importing, saving, resizing, cropping, filtering, and transformation methods.

The library provides a wide range of image file formats, including common formats such as JPEG, PNG, BMP, and TIFF, allowing compatibility and flexibility when working with diverse image data.

Pillow has a comprehensive set of image processing operations, such as basic pixel-level operations, color space conversions, geometric transformations, and intricate filtering algorithms. These operations can be used independently or in conjunction, allowing customers to build the processing pipeline according to their specific image analysis requirements.

In addition, Pillow offers features for managing image metadata, such as extracting and modifying EXIF (Exchangeable Image File Format) data, which provides useful information about image capture settings and device-specific characteristics.

### **2.1.2 SCIKIT**

Scikit-learn, also known as Scikit, is a comprehensive Python package for machine learning and statistical modeling. Built on NumPy, SciPy, and matplotlib, the library provides a vast array of functions that facilitate the development, evaluation, and application of machine learning algorithms.

Scikit provides a unified interface for machine learning applications including classification, regression, clustering, dimensionality reduction, and model selection. The library provides a variety of cutting-edge algorithms, enabling researchers as well as practitioners to experiment with various modeling strategies and select the optimal approach for their specific problem domain.

### **2.1.3 REMBG**

The rembg utility is a Python library that offers a comprehensive and efficient background removal method for images. The library was created to address the challenging problem of extracting foreground objects from complex backgrounds. It employs robust deep learning algorithms to produce precise, high-quality results.

Rembg utilizes the capabilities of convolutional neural networks (CNNs) and a pre-trained model to determine whether pixels in an image belong to the foreground or background. This model was trained on a large dataset, enabling it to generalize effectively and distinguish between object and background regions.

The rembg package contains a variety of programs that cater to numerous input sources, including individual files, directories, HTTPS servers, and RGB24 pixel binary streams. Additionally, rembg provides additional features that increase its usefulness. Users have the ability to retrieve and store a mask that precisely represents the boundaries of the foreground object.

Additionally, the program makes it easier to apply alpha matting to an image. Alpha matting is a technique used to refine the margins of an object and ensure seamless transitions when compositing it onto a new backdrop.

### **2.1.4 BACKGROUNDREMOVER**

Backgroundremover is a Python module that enables the precise and efficient elimination of digital image backgrounds. The application was designed to solve the challenging problem of distinguishing background and foreground objects. It employs innovative algorithms and methods to produce high-quality outcomes.

One of the package's most notable characteristics is its capacity to manage a wide variety of image types and formats, ensuring compatibility and adaptability when working with diverse image datasets. The library supports common image formats such as JPEG, PNG, and TIFF, making it simple to incorporate into existing image processing workflows. Another benefit of this software is its ability to directly remove backgrounds from videos.

The package provides a comprehensive suite of sophisticated image background removal capabilities with an emphasis on academic applications. It includes alpha matting, the ability to select various background removal methods (such as u2netp, u2net, or u2net\_human\_seg), and support for video processing, among other features.

For video processing, backgroundremoval offers a variety of options tailored to video-related duties. Users are able to generate transparent MOV files, allowing seamless overlay of the resulting video onto other videos or images. Additionally, generation of GIF files is supported. In addition, users can modify parameters such as frame rate, total number of frames, and GPU batch size, giving them fine-grained control over the video processing pipeline.

## **2.2 NERF**

NeRF (Neural Radiance Fields) is a revolutionary method for representing and reconstructing scenes in computer graphics and computer vision. NeRF is a revolutionary technique for generating high-fidelity 3D reconstructions of scenes and permits the integration of new perspectives from arbitrary camera angles. It was initially described in the paper "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis" [2] by Ben Mildenhall et al.

The core concept of NeRF is the representation of a scene as a continuous 3D function that simulates the radiance or appearance of the scene from any angle. A deep neural network reflects this function, which is also known as the neural radiance field. NeRF effectively integrates the power of deep learning techniques with ray tracing to achieve remarkable results.

The authors propose a method for training neural radiance field function using a large dataset of images captured from various perspectives. NeRF learns to predict the radiance values at each 3D point within the scene by refining the network parameters through a technique known as volume rendering. This enables the generation of new viewpoints by sampling the radiance along each ray cast from a virtual camera.

NeRF's ability to capture minute geometric details and complex electrical effects is one of its most outstanding characteristics. The neural radiance field is capable of simulating complex scene structures, such as object forms and textures, as well as handling challenging lighting scenarios, such as spectacular reflections and global illumination. The resulting synthetic images are remarkably accurate and realistic, closely replicating the scenes observed in the training dataset.

NeRF has applications in virtual reality, visual effects, and robotics, to name a few. It enables engaging and immersive virtual experiences by synthesizing new perspectives in real time. In robotics, where accurate scene perception is essential for tasks such as object manipulation, scene reconstruction, and navigation, the approach has also shown promise.

Extensive experiments and comparisons with existing approaches are used to demonstrate the effectiveness of NeRF in this paper. The authors provide quantitative and qualitative evaluations that demonstrate NeRF's superior scene reconstruction and view synthesis performance and visual quality. In addition, the paper examines potential future research directions, such as scaling NeRF to manage larger scenes and enhancing its performance for real-time applications.

NeRF is essentially an innovative method for scene representation and view synthesis that employs neural networks and ray tracing algorithms. NeRF generates high-fidelity 3D reconstructions and the synthesis of one-of-a-kind views with astounding realism by characterizing scenes as neural radiance fields. This paper has wide-ranging implications for numerous disciplines, including virtual reality, visual effects, and robotics, and it sets the groundwork for future advancements in scene interpretation and synthesis.

## **2.3 INSTANT NEURAL GRAPHICS PRIMITIVES**

Instant Neural Graphics Primitives introduces a significant framework for real-time object-centric 3D modeling, revolutionizing computer graphics. The work addresses the issue of obtaining photorealistic 3D models of objects from photographs or videos while effectively erasing background data.

The authors [1] present a novel strategy that integrates the power of deep learning techniques with geometric reasoning to achieve high levels of precision, efficiency, and robustness. Their method utilizes neural network capabilities to accurately extract objects from images and determine their underlying 3D geometry. Utilizing the intrinsic correlations between pixels efficiently, the network can estimate the 3D structure of objects while removing the background.

The proposed solution is based on a meticulously constructed neural network architecture that has been trained on large-scale data sets. The network learns to aggregate complex object features, manage occlusions, and accurately reconstruct 3D geometry. The network perpetually enhances its understanding of object boundaries and texture through an iterative process of refinement, resulting in high-quality 3D representations with minute details.

Extensive experimental data presented in the paper [1] demonstrate the effectiveness of the method. The authors demonstrate how the system handles challenging scenarios, such as complex backgrounds, objects with intricate textures, and occlusions. The resulting 3D models have exceptional accuracy and realism, capturing minute details while preserving the structure and texture of the object.

The authors present an innovative occlusion-aware loss function to address the issue of occlusions in object-centric 3D modeling. Occlusions occur when portions of an object are obscured by other objects or elements in the image, making it difficult to precisely capture the object's edges and fine details.

The occlusion-aware loss function is intended to guide the neural network during training and encourage it to successfully navigate obstructed regions. The loss function incorporates a weighting mechanism that prioritizes occluded regions, underscoring their significance in the overall reconstruction process.

The authors present a series of optimization measures, such as network pruning and architectural simplifications, to increase the framework's efficacy while maintaining the quality of its output.

Network pruning is the process of identifying and removing redundant or superfluous connections, weights, or neurons from a neural network. By eliminating these, the network becomes more streamlined and efficient, requiring fewer computing resources to complete tasks while maintaining or improving performance.

This strategy involves refining the neural network architecture by reducing its complexity, such as the number of layers, parameters, or extent of the feature map. By minimizing the design, the network's processing requirements are reduced even further, resulting in increased efficiency while maintaining the ability to produce high-quality outputs.

## **2.4 FFMPEG**

FFmpeg is a robust and widely-used open-source multimedia framework that provides an extensive suite of tools and utilities for managing multimedia data. Originally designed for command-line-based audio and video processing, FFmpeg has evolved into a flexible solution capable of performing a wide variety of multimedia duties.

FFmpeg, at its foundation, is a collection of command-line utilities that enable users to perform various operations on audio and video files, including format conversion, resizing, cropping, merging, and dividing. These instruments provide extensive control over the processing parameters, enabling the output to be precisely customized.

FFmpeg supports an extensive variety of multimedia formats, including popular video formats such as MP4, AVI, and MKV, and audio formats such as MP3, AAC, and FLAC. This extensive format support makes FFmpeg a valuable tool for interoperability and compatibility between various media applications and platforms.

In addition to its command-line interface, FFmpeg provides a library of APIs (Application Programming Interfaces) that enable developers to incorporate multimedia functionality into their own applications. These APIs provide low-level access to FFmpeg's multimedia processing capabilities, enabling developers to build customized solutions and exploit the framework's maximum potential.

FFmpeg is a comprehensive multimedia framework that offers a variety of command-line tools and programming interfaces (APIs) for efficient and versatile audio and video processing. Its integration capabilities and support for numerous multimedia formats make it an indispensable tool for multimedia professionals, developers, and researchers in fields requiring multimedia manipulation and analysis.

## **2.5 PHOTOGRAMMETRY**

Photogrammetry is an extensively used technique in the fields of computer vision and remote sensing for reconstructing 3D models of objects or environments from multiple 2D photographs. It employs the principles of geometry, optics, and image analysis to extract precise measurements and spatial data from overlapping images.

The process consists of taking a series of photographs of a scene or object from various angles. These images are then analyzed to identify shared characteristics and establish correspondences. Photogrammetry calculates the 3D coordinates of the scene's points by triangulating the matched features, thereby producing a dense point cloud.

Photogrammetry can be performed utilizing both near- and far-range remote sensing techniques. Typically, close-range photogrammetry is used to reconstruct three-dimensional models of small-scale objects or scenes using calibrated cameras and specialized software. In contrast, remote sensing photogrammetry involves capturing aerial or satellite imagery to generate large-scale topographic maps or digital elevation models (DEMs) of entire regions or landscapes.

Numerous and varied applications of photogrammetry exist. It is utilized in numerous disciplines, including architecture, civil engineering, archaeology, forestry, and virtual reality. Photogrammetry enables accurate measurements, volumetric analysis, and the visualization of objects or terrains, making it a valuable instrument for a variety of industries.

Photogrammetry is a technical discipline that employs the principles of optics, image analysis, and computational algorithms to reconstruct accurate 3D models from collections of 2D images. It has applications in a variety of fields and allows for precise measurements, visualization, and analysis of objects and environments. Photogrammetric techniques are indispensable in disciplines that require the collection and analysis of precise three-dimensional data as a result of ongoing improvements to their precision and efficiency.

## **2.6 OPENCV**

OpenCV (Open Source Computer Vision Library) is a popular open-source library for computer vision and image processing. It provides a comprehensive set of functions and algorithms designed to facilitate a wide range of computer vision duties, from fundamental image processing to advanced computer vision applications.

OpenCV provides a set of fundamental modules that comprise essential functionalities such as image and video I/O, image processing operations, camera calibration, feature detection and extraction, and object recognition, among others. These modules are implemented in C/C++ but also provide interfaces for popular programming languages such as Python, Java, and MATLAB, allowing for cross-platform development and integration with disparate software ecosystems.

The broad range of OpenCV's computer vision algorithms and methodologies is its greatest asset. It provides a solid foundation for image processing applications by providing an extensive library of algorithms for image enhancement, filtering, geometric transformations, and color space conversions. In addition, OpenCV includes a variety of feature detection and extraction techniques, such as corner detection, edge detection, and blob detection, enabling advanced object recognition and tracking tasks.

The library also contains algorithms for accurate camera calibration and 3D reconstruction, enabling the estimation of camera parameters, the rectification of images, and the reconstruction of 3D scenes from multiple viewpoints. In applications such as augmented reality, robotics, and 3D computer vision, these capabilities are especially valuable.

OpenCV supports machine learning techniques and integration with prominent frameworks for machine learning, such as TensorFlow and PyTorch. This enables users to take advantage of the power of deep learning models for image classification, object detection, and semantic segmentation.

Furthermore, OpenCV offers interfaces for capturing and processing video streams from webcams and other video sources for real-time computer vision applications. Its integration with GUI frameworks facilitates the creation of interactive computer vision applications with visualizations and user interactions.

OpenCV's adaptability has led to its pervasive adoption in the academic, business, and research communities. It is utilized in numerous fields, including robotics, surveillance systems, medical imaging, autonomous vehicles, and video analytics.

OpenCV is, in conclusion, a robust and flexible computer vision library that provides a vast array of functions and algorithms for image processing, computer vision tasks, and machine learning integration. Its extensive capabilities, cross-platform compatibility, and active development community make it an indispensable tool for computer vision and image processing professionals, researchers, and enthusiasts.

## **2.7 BLENDER**

Blender is a highly regarded open-source 3D computer graphics software application that offers a comprehensive suite of features for creating, animating, and rendering 3D models, animations, and other visual effects. It provides a vast array of features to meet the requirements of illustrators, designers, animators, and game developers.

Blender's comprehensive modeling tools are one of its main advantages. It includes polygonal modeling, subdivision surfaces, sculpting, and boolean operations, among other mesh modeling techniques. These tools allow users to construct sophisticated 3D models with precise control over their geometry and topology.

The animation system in Blender is another noteworthy feature. It supports both keyframe and procedural animation, enabling users to construct complex motion sequences. The animation tools include rigging, inverse kinematics (IK), and a robust graph editor for adjusting animation curves. Blender provides an extensive particle system for simulating effects such as smoke, fire, and fluid dynamics.

Blender offers its own rendering engine called Cycles, which is capable of producing photorealistic renders of high quality. Cycles uses path tracing algorithms to precisely simulate light behavior, resulting in lighting and shading effects that are realistic.

Blender is extremely extensible and modifiable via Python scripting. Utilizing its extensive Python API, users can automate repetitive tasks, construct custom tools, and develop functionality-enhancing add-ons. Blender can be integrated into intricate production pipelines and workflows due to its adaptability and programmability.

Blender is a versatile 3D computer graphics software application that provides artists, designers, and animators with a vast array of modeling, animation, rendering, and post-production tools. Its comprehensive capabilities, scripting extensibility, and compatibility with popular rendering engines make it a valuable tool for computer graphics and visual effects.

**3. PROPOSED SOLUTION**

## **3.1 Problem Statement**

Instant Neural Graphics Primitives (Instant NGP) is a valuable programming and artistic instrument that facilitates the transformation of real-world objects and scenes into immersive 3D representations. However, a difficulty arises when users wish to extricate the object's mesh while avoiding the additional computations and complexities associated with visualizing the 3D reconstruction's ultimate result. Additionally, it may be difficult to eradicate certain lingering details.

While Instant NGP has revolutionized the process of generating 3D models, optimizing its functionality for object-centric 3D modeling, particularly in terms of isolating the desired object and removing unnecessary background elements, represents a formidable challenge. Users face obstacles in their pursuit of a seamless extraction process, necessitating the implementation of additional computational procedures and complex post-processing techniques to surmount the inherent complexities. In order to accomplish a refined 3D representation, free of unwanted anomalies and background debris, additional research and innovative solutions are required.

# **4. DEVELOPMENT METHODOLOGY**

# **5. IMPLEMENTATION**

# **6. EXPERIMENTAL RESULTS**

# **7. CONCLUSIONS**

## **7.1 FINAL WORDS**

## **7.2 FUTURE WORK**

# **8. REFERENCES**

[1] Müller, T., Evans, A. C., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics, 41(4), 1–15. <https://doi.org/10.1145/3528223.3530127>

[2] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Springer eBooks (pp. 405–421). <https://doi.org/10.1007/978-3-030-58452-8_24>