For Beginners: What Should You Do Before Starting Your First Embodied AI Project?

Jiyao Zhang
jiyaozhang@stu.pku.edu.cn

Jun. 2024

Embodied AI is an important branch of artificial intelligence, emphasizing that embodied agents learn in interaction with the environment. It involves knowledge from multiple fields such as deep learning, robotics, and 3D vision. This document aims to provide basic learning guidance for beginners preparing to enter the field of embodied intelligence. After completing the relevant knowledge learning, you are welcome to contact me for communication and to discuss the specifics of a possible project.

1 Deep Learning

Deep learning is the foundation of embodied intelligence. You need to have a solid understanding of its fundamental principles, which are crucial for developing projects. I recommend starting with the classic CS231n course from Stanford University. Although the most recent course videos are not publicly accessible, you can begin with the 2017 course materials. Despite some content being slightly outdated, the fundamentals still hold significant value. Be sure to complete the assignments, as they are essential for truly grasping the material. As you progress, you may encounter concepts not covered in the 2017 version, and you need to make use of other resources to deepen your understanding.

- **Course Link**: https://cs231n.stanford.edu/2017/
- **Video Link**:https://www.youtube.com/playlist?list=PL3FW7Lu3i5JvHM8ljYj-zLfQRF3EO8sYv
- **Prerequisite Knowledge**: Calculus, Linear Algebra, Probability and Mathematical Statistics, Python Programming
- **Required Assignments**: You need to complete assignments 1, 2, and 3. This is crucial for you to truly understand and master the course.
- **Extra-curricular Learning**: Transformer and Vision Transformer. Since these topics were not part of the 2017 course, you should study them through blogs and their original papers, and implement self-attention and cross-attention.

- **Advanced Learning**: Diffusion has become a mainstream generative model framework. Familiarize yourself with its core ideas and advantages compared to other generative models. For this area, you can refer to Dr. Yang Song's blog: https://yang-song.net/blog/2021/score/.

2 Robotics

Robotics serves as an important carrier of embodied intelligence. Before embarking on an embodied intelligence project, you should master the core concepts of robotics and develop a deep understanding of key concepts. Make sure to learn the theoretical aspects and practice them through coding or simulation environments.

- **Course Link**: https://modernrobotics.northwestern.edu/nu-gm-book-resource/
- **Key Content**: Rigid body motion, robot kinematics, robot dynamics, trajectory planning, etc.

3 3D Vision

Robots need to interact with the 3D world, so mastering 3D Vision is indispensable for embodied intelligence research. You will want to cover both traditional 3D Vision concepts and techniques that rely on deep learning. Below are the specific areas you need to focus on. Consult relevant materials to guide your study and practical implementation.

- **3D Representations**: Familiarize yourself with the common 3D representation methods such as **point clouds**, **meshes**, **voxels**, **SDF**, **Nerf**, **3D GS**, etc. You do not need to delve deeply into Nerf and 3D GS at this moment, but at least understand their basic principles. Feel free to explore further if you have extra time.
- **SE(3) Group**: Grasp the fundamental concepts of the **SE(3) group** and the commonly used **SO(3) group** representations, including **rotation matrices**, **quaternions**, and **Euler angles**. Understand their characteristics. Try manually implementing the forward kinematics for the Franka Panda robot and verify its correctness in Pybullet.
- **Traditional 3D Vision**: Study the basics of the **pinhole imaging model**, **camera intrinsic and extrinsic parameters**, **stereo matching**, and **triangulation**. As a practice exercise, implement resizing of a depth map and convert it into a point cloud. You should ensure consistency of the resulting point clouds before and after resizing.
- **3D Networks**: Familiarize yourself with 3D neural network architectures, including PointNet, PointNet++, and SparseConv. You need a solid understanding of the logic behind these networks, and you should carefully analyze how 3D neural network designs differ from their 2D counterparts.