Machine Learning Engineering Bootcamp
Capstone, Step 3: Project Proposal
Student: Kenneth Fung
Date: 27 December 2024

I propose my Capstone project to be focused on Monocular Depth Estimation (MDE). In the machine learning subset of AI, MDE is concerned with the computer vision task of estimating the depth of each pixel in an image using only the view of a single camera. The objective is to reconstruct a 3-dimensional scene from a 2-dimensional image. Hence the term “monoocular” - it is as if our brain has to visualize in three dimensions with only one eye.

A. Problem statements in machine learning involving MDE:

Under the umbrella of artificial intelligence, many problems need to be solved or such solutions may be enhanced through progress in MDE. Such solutions are ultimately dependent on machine learning. These problems include the following in alphabetical order:

1. **Reconstruction and Scene Understanding:** In general, evolution in depth perception enables the development of high-fidelity 3D reconstructions, which would be useful in disciplines of architecture, gaming, and cultural preservation.
2. **Autonomous Vehicles (AV):** For autonomous vehicles to maneuver around obstacles including pedestrians and other vehicles, they need to understand the depth and perspectives of objects in the scene, which cameras capture as two dimensional images. Depth perception is necessary for AVs to navigate reliably and operate safely.
3. **Augmented and Virtual Reality (AR/VR):** Improved depth estimation would lead to improved rendering of three dimensional environments, which would improve user interactions for a more immersive experience by enhancing realism in virtual environments.
4. **Drones and Aerial Mapping:** For drones to navigate in a 3-D environment without crashing and causing personal and property damage, algorithms need to be proficient with depth perception. Depth perception is necessary to improve obstacle avoidance and generate more accurate 3D maps of terrains.
5. **Entertainment and Media:** In entertainment, creating realistic 3D animations or effects involving a 3-D environment are labor-intensive. AI systems incorporating machine learning algorithms that are MDE capable in predicting depth from 2-D images can automate and enhance the creation of lifelike 3D content.
6. **Healthcare and Surgery:** Currently, robots (hardware and software) such as the Da Vinci of Intuitive Surgical assist or perform various operations including removing prostate cancer. Accurate and precise depth perception is paramount to saving a life during surgery using robotics. Better depth perception also supports more accurate diagnosis in medical imaging, such as identifying tumors or visualizing anatomical structures.
7. **Human-Computer Interaction:** Human interfaces with machines (e.g., keyboards) may be 3-D, however the interactions are captured through physical touch with a device. Incorporating depth perception into human-computer interfaces would improve recognition of

natural hand movements and gestures in 3D, thus enabling interfaces that are more intuitive.

8. **Robotics:** Robots need depth perception to reliably grasp objects, avoid obstacles, and navigate an environment with precision.
9. **Search and Rescue Operations:** Using robots and drones to locate and rescue victims require depth perception for the machines to navigate debris and assess the risks to victims.
10. **Surveillance and Security:** Using robots for automated surveillance and security is in its infancy, and proficient depth awareness in such systems is necessary to identify risks and determine their urgency in crowded and/or complex environments.

B. Potential datasets for training machine learning models involving MDE:

For the purposes of this Capstone project, a dataset of low-resolution images will be utilized. Adequate datasets are readily available and had been previously utilized for designing advanced MDE models. One such model is Depth Anything[1]. The authors of this project utilized labeled and unlabeled datasets of low and high resolution. I will be using the low-resolution labeled datasets, such as BlendedMVS[2].

C. Computational resources for training machine learning models involving MDE:

For the purposes of this Capstone project, a low-fidelity model requiring minimum computational resources will be pursued. The minimum computational resources necessary for designing and testing a monocular depth perception model depend on the complexity of the model, the size and resolution of the dataset, and the desired resolution of depth maps. If adequate for this Capstone project, simpler models and smaller datasets will be used which would significantly reduce resource requirements. Here is a table itemizing the minimum necessary resources:

Table of Minimum Computational Resources:	
GPU	NVIDIA GTX 1050 Ti (4 GB) or RTX 3060 (12 GB)
CPU	Quad-core or Octa-Core processor (Intel Core i7 or AMD Ryzen 7)
RAM	16 GB system memory
Storage	2 TB HDD or SSD
Compute Time	Less than 3 hours of training on low-resolution datasets for each cycle

D. Ethical issues involved with MDE in machine learning:

Depth perception in machine learning would introduce or improve the ability to interpret, generate, and interact with 3D environments. These abilities in machine learning introduce ethical concerns spanning privacy, security, accessibility, and societal impacts:

[1] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. 2024. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2401.10891

[2] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A largescale dataset for generalized multi-view stereo networks. In CVPR, 2020.

1. **Bias and Fairness:** Society may become too dependent on depth-aware systems, leading to potential discrimination in candidate assessments such as for military school or healthcare assessment, where an otherwise healthy applicant may be labeled as obese.
2. **Manipulation, Deception, Fraud:** If depth perception in machine learning systems are advanced enough, realistic 3D reconstructions may be rendered to deceive for the purposes of manipulating and defrauding. Misleading information would become more pervasive and convincing, which may lead to exploitations by politicians and news media and result in societal harm to individuals and/or institutions.
3. **Surveillance and Privacy:** Enhanced depth perception in cameras and drones allow systems to identify individuals and objects with greater precision with increased abilities to interpret, generate, and interact with 3D environments. This may lead to violations of personal privacy and potential misuse by governments and/or corporations.
4. **Data Security:** Enormous amounts of 3-D data are required to design and maintain high fidelity systems with improved depth perception. Such data may include sensitive and/or proprietary information. Inappropriate use of this data could lead to theft, corporate espionage, and/or compromise of sensitive environments which may threaten individual and national security.
5. **Transparency, Accountability, Intellectual Rights:** Using ChatGPT as an example, it is a challenge to properly cite sources of information or assign intellectual property rights when using artificial intelligence for generative purposes. For decision-making scenarios, users may not fully understand how depth-aware systems make decisions in critical situations, and it may be difficult to assign accountability in fault scenarios such as a crash involving a drone and property.
6. **Inhumane Usage:** Depth perception technologies would most certainly be incorporated into weapons. Such weapons may be deployed indiscriminately and unethically, and exposes the potential for collateral damage in warfare and violations of human rights.
7. **Complacency and Degradation of Manual Skills:** Ever since the advent of calculators, the ability of students to perform calculations manually has degraded over time. Similarly, depth-aware systems might degrade manual skills such as basic hand-eye coordination (e.g., landing an aircraft) or reduce human vigilance such as risk assessment in a critical environment (e.g., airport). Such degradations may lead to catastrophic outcomes in situations involving massive human and/or property losses.
8. **Labor Market Displacement:** Enhanced depth perception will certainly automate tasks in sectors like logistics (operating a crane), construction (operating a bulldozer), and agriculture (crop assessment). Workers formerly employed in such capacities may be displaced, and therefore foreplanning is necessary to provide adequate societal support or training programs into new capacities for the impacted workers.
9. **Private Environments:** Depth perception technologies in smart home devices involving AR/VR may capture and render private spaces without the expressed consent of the consumer. Such breaches of consent and misuse of personal spatial data may be maliciously exploited by individuals, governments, and corporations.

To proactively address these potential ethical issues involving the introduction or enhancement of depth perception in machine learning, developers and policymakers must explore and mitigate ethical concerns during development, legislate robust data protection laws, and emphasize transparency in decision-making processes. International communities would need to collaborate to instill global best practices for the humane use of depth-aware technologies so harm may be minimized while maximizing the benefits to society.