Deep Learning on Embedded Platform

Christopher Johnson, Gabe Morey, Luay Alshawi

October 19, 2016

CS461 Fall Term

Abstract

Deep learning is a form of graphics-based machine learning that uses multi-layered neural networks and high-speed GPU processing to make sense of images, sound and text. The goal of this project is to create an application which uses deep learning to perform a particular task and run on the Jetson TX1 Developer Kit. The application will be utilized as part of a training program to teach users how to solve similar problems with deep learning. NVIDIA's Deep Learning Institute has a high demand for training in this field of artificial intelligence. This project will provide a tool for NVIDIA's training purposes. Because of this the project must also run on their develper kit. Our team will be using version control to work efficiently and keep objectives organized. We will have our repository stored on Git Hub and use issue tracking and status reports to keep track of jobs.

1   Problem Statement

Artificial intelligence has long been a sought after goal in the world of computer science. While true AI is not yet within grasp, advances in processing power have led to marked advancements in the field of machine learning. Using methods modeled after the human brain, computer scientists are finding more and more ways to use computer intelligence to make jobs easier. One of these methods is called deep learning, and it is quickly becoming one of the most important fields in the realm of artificial intelligence. The more people who learn how to use it and develop it, the better it can be and the more goals it can accomplish.

Deep learning is a rapidly growing field of machine learning that aims to use GPU based neural networks to accomplish a wide variety of tasks. Deep learning uses a system of neural networks to process data from graphs, plots, visual input, or any set of machine readable data to find patterns and learn how to solve the problem the data represents. Neural networks are a form of machine learning which use a layered structure of connections to make decisions and store information relevant to a problem. Their structure was inspired by the neurons of the human brain [1]. Each neuron has a trained input weight, which assigns how correct or incorrect that branch may be. These weights are developed through intensive training of the neural network as it attempts to complete tasks [1]. The neural net uses these weights to create a probability vector that predicts the outcome chance of success.

The traditional problem with neural networks has been their incredible computational requirements. It takes massive processing power to train and develop a neural network. However, this problem has been partially solved by taking advantage of the power of the modern GPU. This is the basis for deep learning, and now it is much more feasible to train a network to be proficient at a task. Modern deep learning neural nets can be massive and capable of processing huge amounts of data in order to build their neurons. Using deep learning to build code for solving a problem means that only minimal time needs to be spent programming. Ideally, the deep learning system should be directed to figuring out how to solve the problem with its own code and systems.

NVIDIAs Deep Learning Institute is a set of online courses and workshops with the aim of teaching developers how to use and apply deep learning to their own problems. With this goal in mind, NVIDIA is constantly looking for new and better ways to teach about the subject, as well as further this field of research. Our groups task is to train a deep learning network to build a system of code. Our group may determine what the end goal of the code set is, but it must be able to run inference on NVIDIAs Jetson TX1, a visual computing development platform produced by NVIDIA. When the project is complete, it will be used to make a new course for the Deep Learning Institute. This course will provide an easy, hassle free way to learn how to train and use neural nets to accomplish tasks.

Our solution to the problem will be something that is easily presentable and understood to those interested in learning new fields of artificial intelligence. It should be a program that can utilize deep learning and provide a comprehensive example of deep learning's power and capabilities. A solution such as this will comply with its requirement to be a learning tool for others, as well as a teaching tool. Having an easily understandable and presentable project will make it a useful tool for NVIDIA's training in the deep learning field. It will also be a program which can run on NVIDIA's Jetson TX1 Developer Kit. NVIDIA's Deep Learning Institute will be able to use this project as an aid in teaching deep learning fundamentals. Because this project will be used as a teaching tool it will also have to be well documented so that others can study it effectively. NVIDIA's Deep Learning institute will want to be able to break the project down and show students each major aspect. Our project will have to be complex enough to use deep learning but also comprehensive enough for an audience that is very new to the field.

Version control will be used in this project to keep track of the tasks and measure the performance for each person on the team as well as the performance of the project. Tasks will be labeled to three different categories using github issue tracker. The First label called "Feature" which will be used to label major tasks toward the project. The second label is "Enhancement" which will be used to label tasks that enhance major tasks such as refactoring particular code or add a feature to a major task. The last label is "Bug" which will label tasks that are considered closed but require a fix. By using three labels, tracking the projects performance is easier and efficient. Initially, all members of the team will make tasks and label them to solve the given problems on this project, and each person can assign a task to himself. Moreover, tasks can be reviewed by any member of the team upon completion to ensure the quality and completion of the task.

Finally, related tasks shall be completed using iterations. The length of each iteration is going to be two up to three weeks long, and each iteration must solve an induvial problem. After an iteration is complete, a review session is going to be held with all team members to discuss progress and test all tasks after merging them together. Optionally, the projects sponsor can join the meeting to review the team's progress.

2 Sources

[1]N. C. L. Information, /'/'Deep learning," NVIDIA Developer, 2016. [Online]. Available: https://developer.nvidia.com/learning. Accessed: Oct. 18, 2016.