Software Tools for Machine Learning & Deep Learning

Shuyue (Frank) Guan

The Medical Imaging & Image Analysis (MIA) Laboratory
Sep. 2021
Hardware

• CPU
  • Speed (GHz) – General ML
  • # of Cores – Parallel computing

• RAM – General ML
  • Size (GB) – Image Processing
  • Speed (MHz) – Loading speed

• GPU – DL
  • Buffer (Memory) – **Crucial**: model scale, input/batch size ($\geq 6$GB)
  • CUDA Cores – Speed

---

**GEFORCE GTX 1080 Ti**

GPU Engine Specs:

- NVIDIA CUDA® Cores: 3584
- Boost Clock (MHz): 1582

Memory Specs:

- Memory Speed: 11 Gbps
- Standard Memory Config: 11 GB GDDR5X
- Memory Interface Width: 352-bit

**Out of memory (OOM) issues**
Hardware (Cont.)

• Solid-state disk (SSD) – read/write data fast
  • Speed (MB/s)
  • Size (GB) – by needs
  • *Tips: SSD (512GB) + HDD (2TB)

• Cooling system, power supply
  • Multi-CPU
  • Multi-GPU
  • High-speed RAM/SSD

(Ian Cutress, 2021)
Hardware: Tips

• Develop/test on PC, running on workstation
  • Smaller input/batch sizes
  • Smaller model scales
  • Smaller loops
  • ...

• HPC at GWU (*discuss details later*)

• Cloud Computing resources
  • AWS
  • Google Cloud
  • ...
Software overview

- **Hardware**
  - GPU

- **API**
  - cuDNN
  - CUDA drive

- **Framework**
  - TensorFlow/PyTorch/MxNet
  - Keras

- **Programming language**
  - Python/Matlab/C/C++

- **IDE**
  - PyCharm/Spyder/VSCode

- **Developers**

- **Anaconda (Virtual) Environment**

- ML/DL using CPU
- DL using GPU
Languages for general ML

• MATLAB
  • **Advantages**: quick to get started and easy to use; visualization; detailed official help documentation; high credibility
  • **Disadvantages**: need to buy; slow official updates; lower flexibility; not many users in the DL field

• Python
  • **Advantages**: free; fast community-based update; higher flexibility; widely used in the NLP/DL field; rich packages
  • **Disadvantages**: difficult to get started; not intuitive enough; mixed help information; may not reliable
Languages/tools for general ML

- Other tools
  - **Wolfram Mathematica**: symbolic & numeric computations
    
    ```math
    \text{In[3]} := \text{DSolve}[y''[x] + y[x] == \text{Exp}[x], y, x]
    
    \text{Out[3]} := \{\{y \rightarrow \text{Function}[x, c_1 \text{Cos}[x] + c_2 \text{Sin}[x] + \frac{1}{2} e^x (\text{Cos}[x]^2 + \text{Sin}[x]^2)]\}\}
    ```

  - **Desmos**: graphing calculator, FREE
Tools for DL

Examples:
• TensorFlow + Keras + Python + PyCharm
• PyTorch + Python + VSCode/PyCharm

Recommend: Anaconda
• Virtual environments for development
• Includes/controls: Frameworks + APIs + Languages
Configuration

**DL using CPU**

1. GPU without CUDA cores
2. Install Anaconda
3. (Conda) Install Python
4. (Conda) Install TensorFlow
5. (Conda) Install Keras
6. Install PyCharm
7. (PyCharm) Import Conda environment
8. Ready to use!

**DL using GPU**

1. Nvidia GPU with CUDA cores
2. Install CUDA drive
3. Set cuDNN
4. Install Anaconda
5. (Conda) Install Python
6. (Conda) Install TensorFlow-GPU
7. (Conda) Install Keras
8. Install PyCharm
9. (PyCharm) Import Conda environment
10. Ready to use!
Version problems

- OS (Win, Linux, macOS)
- CPU/GPU
- CUDA, cuDNN, framework, programming language,...

<table>
<thead>
<tr>
<th>Version</th>
<th>Python version</th>
<th>cuDNN</th>
<th>CUDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>tensorflow_gpu-2.3.0</td>
<td>3.5-3.8</td>
<td>7.6</td>
<td>10.1</td>
</tr>
<tr>
<td>tensorflow_gpu-2.2.0</td>
<td>3.5-3.8</td>
<td>7.6</td>
<td>10.1</td>
</tr>
<tr>
<td>tensorflow_gpu-2.1.0</td>
<td>3.5-3.7</td>
<td>7.6</td>
<td>10.1</td>
</tr>
<tr>
<td>tensorflow_gpu-2.0.0</td>
<td>3.5-3.7</td>
<td>7.4</td>
<td>10</td>
</tr>
<tr>
<td>tensorflow_gpu-1.15.0</td>
<td>3.5-3.7</td>
<td>7.4</td>
<td>10</td>
</tr>
<tr>
<td>tensorflow_gpu-1.14.0</td>
<td>3.5-3.7</td>
<td>7.4</td>
<td>10</td>
</tr>
<tr>
<td>tensorflow_gpu-1.13.0</td>
<td>3.5-3.7</td>
<td>7.4</td>
<td>10</td>
</tr>
<tr>
<td>tensorflow_gpu-1.12.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.11.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.10.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.9.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.8.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.7.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.6.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.5.0</td>
<td>3.5-3.6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>tensorflow_gpu-1.4.0</td>
<td>3.5-3.6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>tensorflow_gpu-1.3.0</td>
<td>3.5-3.6</td>
<td>6</td>
<td>10.8</td>
</tr>
</tbody>
</table>
Packages for ML/DL in Python

- **NumPy**: basic functions for math and matrix
- **SciPy**: scientific computing
- **Scikit-learn**: machine learning library
- **Matplotlib**: plotting library
- **OpenCV**: computer vision
- **Pandas**: data manipulation and analysis
- **…**
Integrated development environment (IDE)

- **PyCharm**: comprehensive IDE, for large projects
- **VSCode**: light-weight IDE, support many languages
- **Jupyter Notebook**: interactive IDE, publish friendly
- **Spyder**: MATLAB-like, Anaconda built-in
- ...
Websites for ML/DL

- Guidebooks of each tools/software online – look-up books
- Github – without reinventing the wheel
- Stack Overflow – solve problems
- Kaggle – datasets
- paperswithcode.com – SOTA performance ranks with papers & codes
- Google – everything!
- …
HPC at GW

• One CPU node
  • Dual 20-Core 3.70GHz Intel Xeon processors
  • 192GB RAM
  • 800 GB SSD

• One GPU node
  • 2 NVIDIA Tesla V100 GPU (4 for large nodes)
  • Dual 20-Core 3.70GHz Intel Xeon (18-Core Xeon for large nodes)
  • 192GB RAM (384GB for large nodes)
  • 800 GB SSD

• High throughput node, High memory node (3TB RAM!),...
How to use the HPC at GW

Overview

- Request for an account
- Transfer data & code - Globus
- VPN Connection (Off-campus) - Cisco AnyConnect
- Login by SSH client - MobaXterm
- Management by Linux commands
- Jobs deployment/running by SLURM

Workflow

- Upload data and code files
- Create a **SLURM script** to assign the job
- Run the **script**
- Outputs will be saved
- Receive states and notifications by Email or check by commands
Epilogue

• The Medical Imaging & Image Analysis Laboratory
  SEH 5290 W: loewlab.seas.gwu.edu

• Detailed instruction of GWU HPC

The Medical Imaging & Image Analysis (MIA) Laboratory, 2021