Bone Age Estimation from Hand X-Rays Images Using NVIDIA DGX-1

Ken Ying-Kai Liao

Big Data Engineer
Artificial Intelligence Center for Medical Diagnosis
China Medical University Hospital
(T32209@mail.cmuh.org.tw)
Outline

1. Introduction: Medical Imaging and Artificial Intelligence
2. Sharing a DGX-1 using JupyterHub
3. NVIDIA GPU Cloud Docker Images to Jupyter Notebook Images
4. Case Study: Bone Age Estimation from X-Ray Images
   a. Clinical Problem Definition
   b. Data Cleaning and Preprocessing
   c. Hyperparameter Tuning
   d. DGX-1 Speedup
   e. Current Results
5. Conclusion: Future Work
Medical Imaging and Artificial Intelligence

Advantages:
- Accurate?
- Fast
- Objective

Disadvantages:
- Accurate?
- What is the AI actually learning? (Black Box?)
Medical Imaging and Artificial Intelligence

“AI will be an assumed part of healthcare - making it more efficient, more accurate and more personalized.”

Dr. Mark Michalski,
Executive Director of the MGH BWH Center for Clinical Data Science
Sharing a DGX-1 using JupyterHub

- To speed up AI research at CMUH, our director purchased a DGX-1

- Problem:
  - DGX-1 best practices state to use Docker Images for running deep learning applications
  - Our research team is used to directly coding on Jupyter Notebooks

- Solution:
  - JupyterHub + Jupyter Notebook Docker Images with GPU Selection
Sharing a DGX-1 using JupyterHub

- Set up JupyterHub to allow selection of GPUs
- This allows GPU sharing between engineers
NVIDIA GPU Cloud Docker Images to Jupyter Notebook Images

- Use NGC Containers as Base
- Add Jupyter
- NGC + Jupyter = Fast Data Science Prototyping!

```
FROM nvcr.io/nvidia/tensorflow:18.04-py3

RUN python3 -m pip install --upgrade pip
RUN python3 -m pip install --upgrade keras opencv-python pillow numpy scikit-learn scikit-image jupyter jupyterlab jupyterlab-notebook
RUN python3 -m pip install --upgrade h5py pycuda

# Configure container startup
ENTRYPOINT ["tini", "--"]
CMD ["start-notebook.sh"]
EXPOSE 8888
```
Case Study: Bone Age Estimation from X-Ray Images
Clinical Problem Definition

- VP Dr. Tsai wants us to use AI to look at bone age
  - Actual Age?
  - Bone Age?
  - Comparison can show us whether there is a growth problem!
  - Estimate Ultimate Height
Clinical Problem Definition

- Doctors compare X-Ray images to an Atlas
  - Slow
  - Time Consuming
  - Results are Subjective
Clinical Problem Definition

- With AI Bone Age Reading
  - Fast (Especially with NVIDIA GPU!)
  - Objective rather than subjective
Clinical Problem Definition

● We have our problem, now what?
  ○ Hospital IT Department:
    ■ Here's some excel sheets and images!
Good Images
OK Image...
Our First Image Cleaning Task...

- Go through and remove all the images without hands
- This means looking at 19829 images manually…
- We ended up with 7107 that has bone age and "hand" images
Data Cleaning and Preprocessing

Image Cleaning

- Black on White? White on Black?
  - Normalize to White on Black!
- Different Images Sizes
  - Resize and pad with black to keep aspect ratio
## Data Cleaning and Preprocessing

<table>
<thead>
<tr>
<th>Label Cleaning</th>
<th>Boneage: 11-12</th>
<th>Boneage: 16</th>
<th>Boneage: 14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boneage: 10</td>
<td>Boneage: 2</td>
<td>Boneage: 11</td>
</tr>
<tr>
<td></td>
<td>Boneage: 13.5-14</td>
<td>Boneage: 12</td>
<td>Boneage: 9</td>
</tr>
<tr>
<td></td>
<td>Boneage: 2</td>
<td>Boneage: 7-8</td>
<td>Boneage: &gt; 16</td>
</tr>
<tr>
<td></td>
<td>Boneage: 10-11</td>
<td>Boneage: 11</td>
<td>Boneage: 16</td>
</tr>
<tr>
<td>Boneage: 8</td>
<td>Boneage: 15.5</td>
<td>Boneage: 6-7</td>
<td>Boneage: (2009/09/24)</td>
</tr>
<tr>
<td>Boneage: 13+</td>
<td>Boneage: 13.5</td>
<td>Boneage: 13.5</td>
<td>Boneage: 5-6</td>
</tr>
<tr>
<td>Boneage: 16</td>
<td>Boneage: 16</td>
<td>Boneage: 10</td>
<td>Boneage: 7-8</td>
</tr>
<tr>
<td>Boneage: 4</td>
<td>Boneage: 16</td>
<td>Boneage: 10</td>
<td>Boneage: 9</td>
</tr>
<tr>
<td>Boneage: 8</td>
<td>Boneage: 3-4</td>
<td>Boneage: 10</td>
<td>Boneage: 9</td>
</tr>
<tr>
<td>Boneage: 13.5</td>
<td>Boneage: 15+</td>
<td>Boneage: --15.5-16</td>
<td>Boneage: 9-10</td>
</tr>
</tbody>
</table>
Task 2, Label Cleaning

- Ask Doctors if they can help us provide clean labels
  - They said OK, but only if they have a nice interface to do it on
Data Cleaning and Preprocessing
A few days later ...
Label Cleaning

- Dr. Tsai has gone through more than 5000 images, and added clean labels!
- Now we can start hyperparameter tuning!
# Hyperparameter Tuning

<table>
<thead>
<tr>
<th>Base Network?</th>
<th>Layers?</th>
<th>Learning Rate?</th>
<th>Loss Function?</th>
</tr>
</thead>
<tbody>
<tr>
<td>● ResNet</td>
<td>● #</td>
<td>● Constant</td>
<td>● Log Loss (Categorical)</td>
</tr>
<tr>
<td>● Inception</td>
<td>● SGD</td>
<td>● Decay</td>
<td>● MAE</td>
</tr>
<tr>
<td>● Inception-ResNet</td>
<td>● RMSProp</td>
<td>● Schedule</td>
<td>● MSE</td>
</tr>
<tr>
<td>● DenseNet</td>
<td>● Adam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>● NASNet</td>
<td>● Adadelta</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image Size?</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>● 128x128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>● 256x256</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>● 512x512</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...More?
DGX-1 Speedup

Using InceptionResNetV2 as an example

<table>
<thead>
<tr>
<th></th>
<th>CPU</th>
<th>8 x GPU (DGX-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Epoch:</td>
<td>980s</td>
<td>Per Epoch: 20s</td>
</tr>
<tr>
<td>Epochs to Convergence:</td>
<td>128</td>
<td>Epochs to Convergence: 128</td>
</tr>
<tr>
<td>Time to Convergence:</td>
<td>35 Hours</td>
<td>Time to Convergence: 42 Minutes</td>
</tr>
</tbody>
</table>
DGX-1 Speedup

- DGX-1 Speedup = 50x FASTER!
- We can try 50 different hyperparameters in the same amount of time!
Current Results

Models: Separate models for Male and Female subjects

Cleaned Data: 2739 Male, 4368 Female

- Using InceptionResNet-V2 Base Network
  - MSE as Loss Function
  - Adadelta as Optimizer
Current Results

Female

- Label Cleaning 76% complete by CMUH doctors
- Train / Validation: 80% / 20%
- Validation Score:
  - Prediction Accuracy within 36 Months: 99%
  - Prediction Accuracy within 24 Months: 97%
  - Prediction Accuracy within 12 Months: 75%
Current Results

Male

- Label Cleaning 100% complete by CMUH doctors
- Train / Validation: 80% / 20%
- Validation Score:
  - Prediction Accuracy within 36 Months: 100%
  - Prediction Accuracy within 24 Months: 99%
  - Prediction Accuracy within 18 Months: 97%
  - Prediction Accuracy within 12 Months: 90%
  - Prediction Accuracy within 06 Months: 65%
Current Results

- Cleaner Data = Higher Accuracy
- More Data = Higher Accuracy
- Training can be done **REALLY FAST** through NVIDIA DGX-1
- We can also visualize the results
Current Results (Saliency Map)

Bone Age: 5.0

Predicted: 4.967061996459961
Current Results (Saliency Map)

Bone Age: 8.0

Predicted: 7.9445905685424805
Current Results (Saliency Map)

Bone Age 14.5

Predicted 14.3894742007751465
Conclusion: Future Work

For GPU sharing:

- More DGX-1 → JupyterHub Docker → Kubernetes

To increase accuracy:

- Clean Data!
- More Data!
- More Hyperparameters!
Thank You

Ken Ying-Kai Liao

Big Data Engineer
Artificial Intelligence Center for Medical Diagnosis
China Medical University Hospital
(T32209@mail.cmuh.org.tw)