Deep Learning Models for Particle Identification and Energy Regression in CMS HGCAL L1 Trigger

Using Micron’s FPGA based inference engines and FWDNXT firmware + software for compiling models and running the inference

Anwesha Bhattacharya, Emilio Meschi, Thomas James, Dejan Golubovic
13th August, 2019
Introduction

- Classify **electrons** and **photons** and estimate the energy of particles for the L1 trigger
  - Unsupervised learning for clustering
  - Supervised learning for classification and regression
  - Implement the models in software and Micron hardware
Introduction

- Classify **electrons** and **photons** and estimate the energy of particles for the L1 trigger
  - Unsupervised learning for clustering
  - Supervised learning for classification and regression
  - Implement the models in software and Micron hardware

- **DL implementations**
  - Python -> keras/tensorflow

- **Hardware & Software connecting DL models and hardware- by Micron**
  - SB-852 FPGA board
  - Drivers for hardware
  - Compiler to translate DL models to FPGA bitstream - proprietary
  - Python wrapper around compiler - open source
Dataset

- Simulated single particle events with no pileup. Interpreted as 3D images of energy deposits in the calorimeter.
- 100k photons, $e^+$ and $e^-$ generated from the interaction region. 40% photons decay to $e^+$-$e^-$ pairs.
- Constrained topological clustering is used to extract the clusters.
- CNN model trained on this dataset for classification and regression.
DL Model Architecture

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
</tr>
</tbody>
</table>

2D CNN

3D CNN

Single input 3D CNN

Multi-input 2D CNN

Training -> 860us/image 55us/image
Classification Model

2D 2-input CNN model that takes into account both x and y channels.

<table>
<thead>
<tr>
<th></th>
<th>Pred Photons</th>
<th>Pred $e^+/e^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True photons</td>
<td>0.82</td>
<td>0.18</td>
</tr>
<tr>
<td>True $e^+/e^-$</td>
<td>0.13</td>
<td>0.87</td>
</tr>
</tbody>
</table>

ROC curve (photons), ph: 7148, ele: 7149

Misidentification vs energy, thresh=0.5

CERN openlab
Regression Model

Energy error from cl3d* and model pred vs true energy

Relative error on energy estimate from the model compared to cl3d* and summed trigger cells

Gaussian fit for cl3d

Gaussian fit for the model

46k - train samples
20k - validation
28k - test
95k - total

*cl3d = Classical density-based clustering of each detector layer and then sum
Trained classification and regression models for CMS HGCAL L1 trigger.

Both 2D 2-input and Conv3D models compile and run on the micron board using FWDNXT SDK.

Promising results for particle identification. Energy regression needs further tuning.

A test comparison shows good results when using single inputs. Multiple input mode gives results inconsistent with the python implementation - to be followed up with Micron/FWDNXT.
Next Steps

- Compare performance of Conv3D and 2D 2-input models for the board.
- Compare classification results with BDT results from HGCAL backend group.
- Optimize and tune the model for the board.
- Merge regression and classification model into one.
- Include pile-up events.
- Explore regression models for cluster position.
Useful Links

- Micron hardware

- FWDNXT SDK
  - https://github.com/FWDNXT/SDK

- The Phase-2 Upgrade of the CMS L1 Trigger Interim TDR
  - https://cds.cern.ch/record/2283192?ln=en

- The Phase-2 Upgrade of the CMS Endcap Calorimeter
  - https://cds.cern.ch/record/2293646
Thank you

QUESTIONS?

96anwesha@gmail.com
Specifications - sb852

- Xilinx Virtex Ultrascale+ UV7P or UV9P FPGA
- 2GB Hybrid Memory Cube
- Two full-width (x16) links with 15 Gb/s transceivers
- Up to 120 GB/s HMC bandwidth
- Up to 30 GB/s (RX and TX combined) via each full-width (x16) link
- 64GB DDR4 SODIMM (standard configuration); upgradeable to 512GB of high-performance memory
- 2 QSFP transceiver connectors
- PCIe x16 Gen3 to the host
- SDAccel (OpenCL™) support
- Easy design framework with simple FPGA bitstream loading from host
- Complete suite of analytics tools
# Confusion matrix for 3D CNN

<table>
<thead>
<tr>
<th></th>
<th>Pred photons</th>
<th>Pred $e^+/e^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True photons</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>True $e^+/e^-$</td>
<td>0.31</td>
<td>0.69</td>
</tr>
</tbody>
</table>