Towards the creation of Deep Learning-Aware Ensembles

Research Aims
The principal goal of the research is to create faster methods and
techniques for training Deep Learning models on GPUs, to facilitate the
use of Ensemble Methods in Deep Learning. The secondary goal of the
research is to use this increased speed to enable the research of
sophisticated Ensemble Methods that make use of the information
available about the structure of the base classifier, such as Deep
Learning-aware Ensemble Methods.

Better Training Methods for Deep Learning
Even using GPUs, which enable a large amount of computation to be
done in parallel, it is often necessary to train a single model for days or
even weeks at a time. Such long training times hinder the ability to
conduct research on aggregate models, and as such it is first necessary
to develop and improve existing training algorithms to drastically reduce
the amount of computation required to reach good (or better)
generalization.

Deep Learning-Aware Ensembles
The following step is to generate Ensemble Methods which make use of
concepts from Deep Learning and adjacent fields, such as Transfer of
Learning, or knowledge about the underlying specific Deep Learning
model being used to improve generalisation and accelerate the training.

Research Methodology
We developed a framework for experimenting with Deep Learning
Ensembles on GPUs, based on existing technology (CUDA, Theano),
that supports most of the state-of-the-art techniques, and facilitates the
creation of new algorithms.
We use an array of GPUs to verify our new theoretical formulations on
benchmark datasets commonly used in the literature, such as MNIST,
CIFAR, ImageNet and SVHN.

Research Outputs
- A new method for training Deep Neural Networks that include
  Dropout regularization with RPROP has been developed, which
  reduces training time and enables the use of Ensembles in full-
  batch learning mode [1].
- A new training algorithm called ARMSProp that proposes the use of
  weight-wise adaptive learning rates in mini-batch learning has been
developed. Figure 1 shows an example of ARMSProp on CIFAR-100,
  compared to a best-in-class method (RMSPProp).
- A new method for updating the input vector during training, to
  increase speed of convergence on images. Figure 2. shows the
  residual gradient that is passed on to the training input vector on
  CIFAR-10 during the first few epochs.
- A new Ensemble method that uses knowledge of the underlying
classifier to apply Transfer of Learning between Boosting rounds,
called Deep Incremental Boosting. Table 1. shows preliminary
results on CIFAR-10 and MNIST.

<table>
<thead>
<tr>
<th>(MISCLASSIFICATION)</th>
<th>SINGLE NET</th>
<th>ADABOOST</th>
<th>D.I.B</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR-10</td>
<td>27%</td>
<td>23%</td>
<td>18%</td>
</tr>
<tr>
<td>MNIST</td>
<td>0.72%</td>
<td>0.63%</td>
<td>0.49%</td>
</tr>
</tbody>
</table>

Table 1. Preliminary results on Deep Incremental Boosting

Publications
Learning”, UK conference on Computational Intelligence, 2015