Multi-GPU Deep Learning Techniques for Tendon Healing in Regenerative Stem Cells based Medicine

Kapiński N.[1], Zieliński J.[1,2], Borucki B.A.[1], Nowiński K.S[1].
[1] Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw
[2] Department of Biophysics and Human Physiology, Medical University of Warsaw

Introduction

In 2012 Krizhevsky et al. introduced ImageNet Convolutional Neural Network [1] (CNN) used to win the 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge). Since that point deep learning techniques have become one of the most popular and effective tools for image recognition. ImageNet was followed by successful fine-tuned modifications like ZF Net, simplifications like VGG Net, complex inception module based GoogLeNet and "ultra-deep" architecture ResNet [2] consisting of up to 152 layers[1]. Newest trends in deep learning include utilization of an ensemble of models which was seen on 2016 ILSVRC.

One of the bottlenecks of deep learning based approaches is a requirement of large labeled training dataset which is often fairly hard to collect. Furthermore, training phase on novel multi GPU computers, depending on a size of a dataset and chosen architecture, can last up to several days which makes it a supercomputing problem[2].

Recent successes of deep learning techniques in MRI data analysis regarding e.g. breast cancer [3] or Alzheimer disease early detections provided the main motivation for this work where we seek the optimal solution for our deep learning based system (see Fig 1.) for monitoring of the Achilles tendon regeneration process.

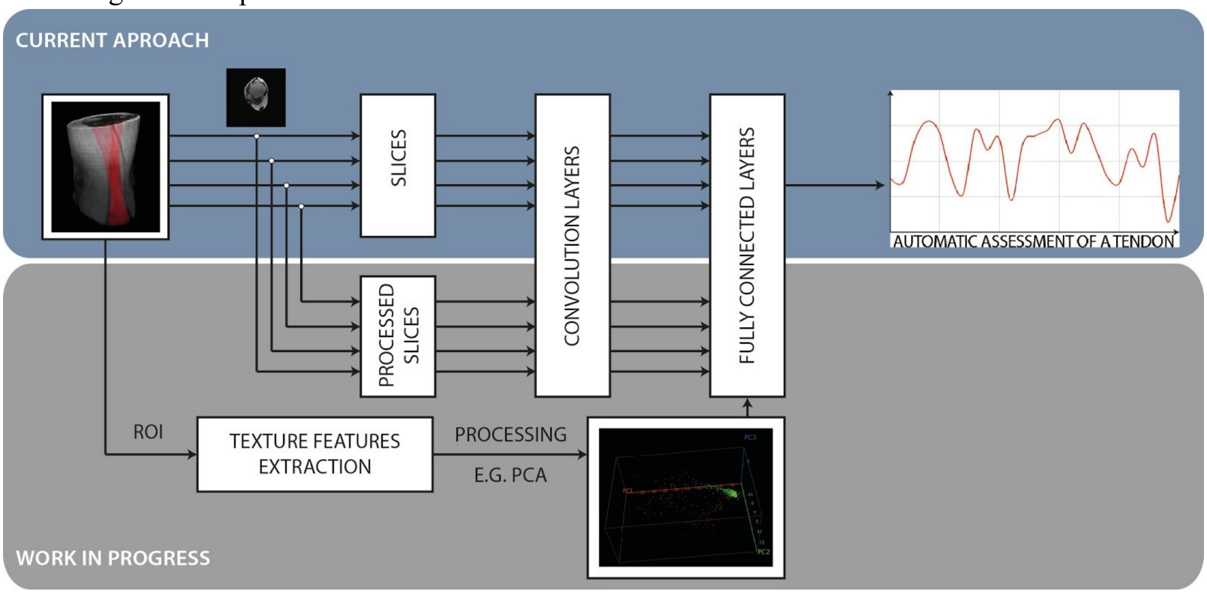

Fig 1.: Currently tested and developed approaches for automatic monitoring of a tendon regeneration process.

Within this paper we assess the possible gain of using ultra-deep CNNs in comparison to ImageNet based approach. We present training times of the backpropagation algorithm with different number of GPUs on a NVIDIA M6000 based server and identify limitations of the current architecture in order to conclude regarding setup improvements that soon will be implemented in our supercomputing centre.

---

[1] Authors of the ResNet tested architectures even beyond 1000 layers however more complex architectures resulted in higher training error.
[2] Within the 2015 ILSVRC dataset consisted of 1.2 mln images divided into 1000 classes. Using 4 NVIDIA Kepler GPUs the training of the ResNet took from 3.5 days for the 18-layer model to 14 days for the 101-layer model.

Methods

Within the START project ("Novel Scaffold-based Tissue Engineering Approaches to Healing and Regeneration of Tendons and Ligaments") we acquired MRI data of the lower limb of 27 healthy volunteers and 28 patients after Achilles tendon rupture. Each person has been scanned using 9 frequently used MRI protocols. Additionally injured patients were scanned several times (up to 5) during the therapy in order to monitor the tendon regeneration. This approach resulted in over 500 total number of datasets collected as of January 2017. However, within the context of applying regenerative stem cells based approach we are interested in assessment of local structure changes, thus we have analysed 2D MRI slices. We have extracted over 60.000 slices of injured tendons and over 10.000 healthy ones. Data were augmented by mirroring slices and additionally twice rotating the images of healthy tendons. Final dataset comprises of over 150.000 images. The motivation for the presented approach (other than training set augmentation) is to include representations of opposite limbs and small rotations of slices within the dataset. Additionally a number of 100.000 is an approximate number of MRI slices of the Achilles tendon obtained annually in Poland (where the studies were conducted) which represents the problem in a country scale. Within the next step of dataset preparation, slices were binary labeled (as healthy or injured tendon image), shuffled and divided into train, test and validation sets, containing respectively 80%, 10% and 10% of the total number of slices. Training process of the CNNs was conducted with use of several setups of parameters on a server containing 4 NVIDIA M6000 GPUs with CUDA v7.5 and cuDNN v5.1 installed. After achieving satisfying classification results we defined a continuous measure H calculated for each slice by the formula

$H(x) = \left| \sum_{j=0}^{n} y_j \right|$, where x is a slice and y_j is a neuron on the last fully connected layer of the used network.

The value $H(x)$ can be used as an estimate of the degree of healing progress of the tendon obtained from the non-discretized output of the last fully connected CNN layer.

Results

The comparison of the best classification results obtained with ImageNet and with 18-layer ResNet model for different training phase durations are presented in the Table 1.

Table 1.: Comparison of the classification results for the ImageNet and the ResNet.

Net	iterations	True positive	True negative	False positive	False negative	Correct detections	Incorrect detections	Testset size	Accuracy
ImageNet	10000	7787	7680	309	201	15467	510	15977	96.81%
ResNet	20000	7756	898	7091	232	8654	7323	15977	54.17%
ResNet	40000	242	7989	0	7746	8231	7746	15977	51.52%
ResNet	60000	7888	7676	313	100	15564	413	15977	97.42%
ResNet	120000	7944	7623	366	44	15567	410	15977	97.43%
ResNet	180000	7842	7599	390	146	15441	536	15977	96.65%

We obtained the best accuracy of validation set classification for the ResNet after 120.000 iterations of training. However in comparison to ResNet after 60.000 iterations the gain was only 0.01%. Training ResNet with less than 60.000 iterations showed lack of accuracy gain and training with more than 120.000 resulted in decrease of accuracy in comparison to the best result. ResNet compared with ImageNet showed an accuracy improvement of 0,62%, comparable false positive classification and 50% decreased false negative rate.

Even small improvement of accuracy and especially correct classification of slices are of great importance in our application. On the Fig 2. we presented an example of the measure *H* calculated for a single MRI study of a patient utilizing the ResNet and the ImageNet outputs.

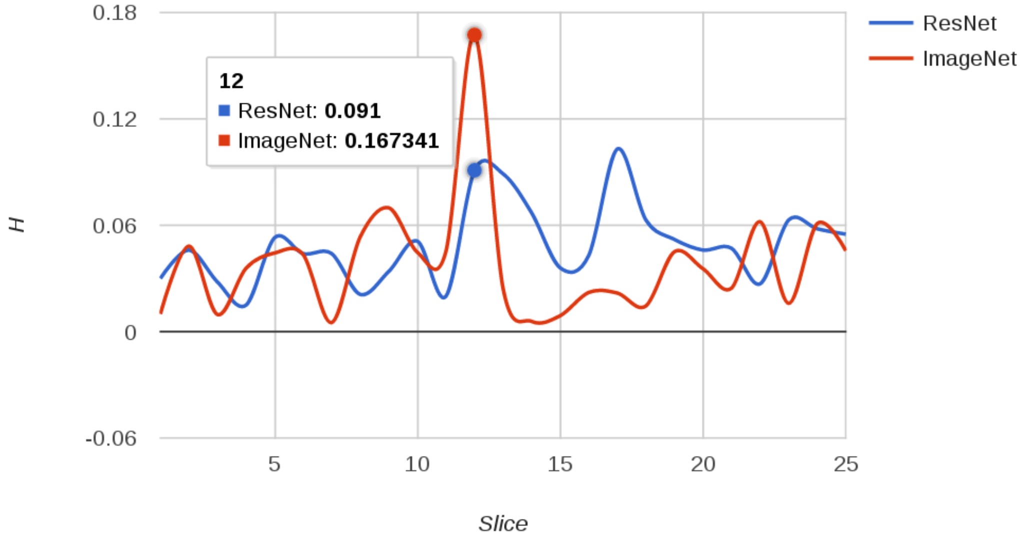

Fig 2.: Comparison of the measure *H* calculated with the ResNet and the ImageNet outputs.

False negative detection can be observed on slice 12. We treated *H* as a variable and calculated Shapiro–Wilk test to find out that the distributions obtained from both classifications are normal. For ImageNet we obtained $W = 0.75$ ($p < 0.01$) and for ResNet $W = 0.94$ ($p > 0.1$). This means that the variable *H* obtained through the ResNet network is approximately normal (contrary to the ImageNet network). This is due to a lower impact of outliers in ResNet case. Additionally ResNet has a 45% smaller variance than the distribution resulted from the ImageNet classification. The above example illustrates that small accuracy gain can provide better results of the final measure and facilitate error detection.

Further studies covered analysis of the ResNet training scalability on Multi-GPU architectures. Fig 3. presents time of training phase on 1, 2 and 4 GPUs.

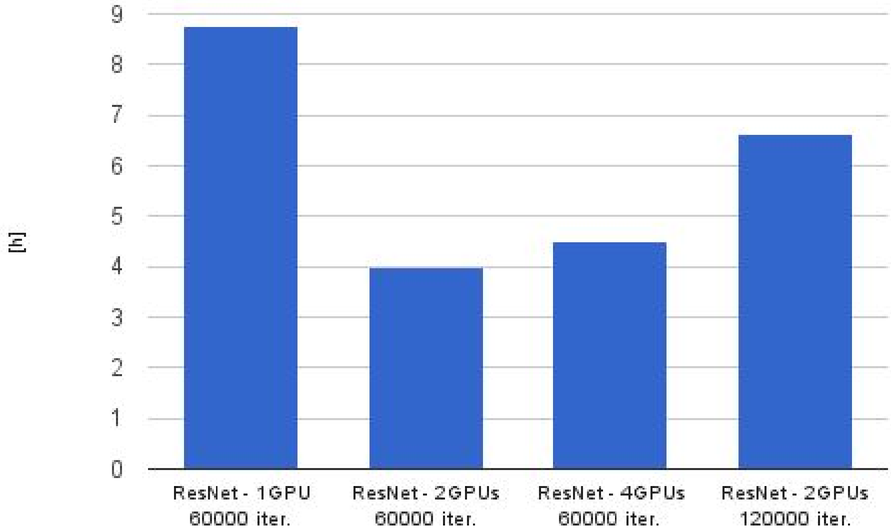

Fig 3.: Time of the ResNet training phase on different number of GPUs.

Fastest time for training the ResNet with 120.000 iterations (best classification presented in the Table 1) took 6h 42 min. Training on 2 GPUs resulted in more than double speedup which match the NVIDIA results presenting linear speedups with multi GPU [4]. However on our architecture training on 4 GPUs increased the time by 30 minutes which resulted from the increased communication within the computational model and lack of p2p connection between pairs of GPUs and losses due to PCI express communication.

Conclusions

We assessed the possible gain of using ultra-deep CNNs vs. ImageNet. We performed multiple tests with different parameters and obtained best classification results for ImageNet of 96.81% and for 18-layer ResNet of 97.43%. For our application the classification accuracy is crucial as shown on an example presented on the Fig 2., thus we can conclude that ultra-deep architecture will be further considered as a good candidate for the core of our system for deep learning based monitoring of the healing process of the Achilles tendon. Within further studies we identified limitations of the training process speedup on our current server related to lack of fast communication between particular GPUs.

The time for training that resulted in the best accuracy took 6h 42 min. In order to run multiple tests looking further for improvements of accuracy of our system, perform training of even deeper architectures up to 152-layer ResNet model, we must shorten the training phase. Thus in conclusion we defined a need for an architecture comprises of at least 4 P100 GPUs interconnected using the new NVLink technology (e.g. The IBM Minsky S822LC server). New machine will be soon available in our supercomputing centre and test results will be attached to the presented comparison.

Acknowledgments

The following work was part of "Novel Scaffold-based Tissue Engineering Approaches to Healing and Regeneration of Tendons and Ligaments (START)" project, co-funded by The National Centre for Research and Development (Poland) within STRATEGMED programme (STRATEGMED1/233224/10/NCBR/2014).

References

[1] Krizhevsky A., Sutskever I., Hinton G.: *ImageNet Classification with Deep Convolutional Neural Networks.* NIPS 2012: 1106-1114

[2] Kaiming H., Xiangyu Z., Shaoqing R., Jian S., *Deep Residual Learning for Image Recognition*, Microsoft Research, 2015.

[3] Jie-Zhi Cheng et al., *Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans,* doi:10.1038/srep24454, 2016.

[4] NVIDIA corp., *Deep Learning on GPUs*, March 2016.