Three Years of Low-Power Image Recognition Challenge: Introduction to Special Session

Kent Gauen, Ryan Dailey, Yung-Hsiang Lu
Purdue University
West Lafayette, Indiana, USA.
{gauenk, dailey1, yunglu}@purdue.edu

Eunbyung Park, Wei Liu, Alexander C. Berg
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA.
{eunbyung, wliu,aberg}@cs.unc.edu

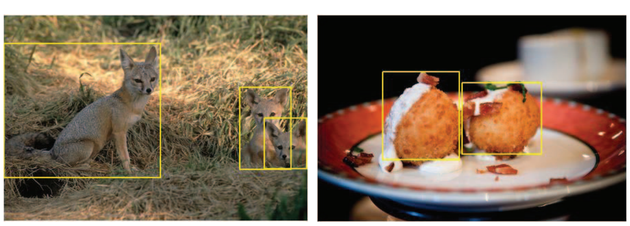
Yiran Chen
Duke University
Durham, North Carolina, USA.
yiran.chen@duke.edu

***Abstract*—Reducing power consumption has been one of the most important goals since the creation of electronic systems. Energy efficiency is increasingly important as battery-powered systems (such as smartphones, drones, and body cameras) are widely used. It is desirable using the on-board computers to recognize objects in the images captured by these cameras. The Low-Power Image Recognition Challenge (LPIRC) is an annual competition started in 2015. The special session includes presentations given by the winners of the first three years of LPIRC. This paper explains the rules of the competition and the rationale, summarizes the teams' scores, and describes the lessons learned in the first three years. The paper suggests possible improvements of future challenges.**

***Index Terms*—Low-Power Electronics, Computer Vision, Machine Intelligence**

I. Introduction

The Low-Power Image Recognition Challenge (LPIRC) is an annual competition using training data from the ImageNet Challenge [1], [2] and ImageNet-like data for testing. Figure 1 shows two examples. Each image has several "objects of interest". Figure 1 (a) has three foxes and Figure 1 (b) has two croquettes. The objects of interest are enclosed by the *bounding boxes*. The objects belong to pre-defined 200 categories, such as airplane, bicycle, car, hammer, horse, soccer, and violin. Figure 2 shows examples of images used in LPIRC.



(a) (b)

Fig. 1. Sample images from ImageNet.

A computer system successfully detects an object when three conditions are met: (1) The system correctly identifies the category of the object. (2) The object has not already been detected. (3) The bounding box reported by the system has at least 50% overlap with the ground truth bounding box. Figure 3 shows an example. The black solid bounding box is the ground truth specified by LPIRC. The blue dash bounding box is drawn by a computer system. If the overlap is more than 50%, the computer system successfully locates the object. The overlap is defined as

$$\frac{\text{reported} \cap \text{ground truth}}{\text{reported} \cup \text{ground truth}} \geq 0.5. \qquad (1)$$

To compute a score for the accuracy of a solution over multiple targets of different categories of objects, we compute the mean average precision, *mAP*. The first step is to compute the average precision, also called the area under the precision-recall curve. To do so, a program outputs a score for each proposed detection in addition to the detection location and category. A point on the precision recall (PR) curve is computed by taking all putative detections with a score larger than some threshold and computing the precision and accuracy. The whole PR curve is computed by varying threshold. If only a subset of objects or images are processed this will affect the recall values and reduce the AP. Then mAP is the average of the areas under the PR curve for each object category. For more detail, see [2].

The final score for LPIRC is the ratio of mAP and the total energy consumption in 10 minutes. The energy is measured by a power meter (Yokogawa WT 300). If a solution can finish the entire dataset earlier, the solution can stop the power meter before the 10 minutes end. The score is the ratio of accuracy and the energy consumption:

$$\text{score} = \frac{\text{mean average precision}}{\text{energy consumption}}. \qquad (2)$$

Since 2015, LPIRC has been held three times. Table I shows the numbers of solutions presented each year. A team may present multiple solutions.

Table II shows the scores of each year's champion. The accuracy (mAP) improves 8.53 times from 2015 to 2017. The final score improves 6.56 times.

978-3-9819263-0-9/DATE18/©2018 EDAA

Fig. 2. Sample test images used in LPIRC.

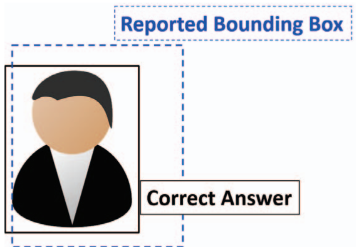


Fig. 3. Bounding boxes.

Year	Date	Conference	# Solutions
2015	June 7	Design Automation Conference	17
2016	June 5	Design Automation Conference	5
2017	July 21	Computer Vision and Pattern Recognition	10

TABLE I
LPIRC HAS BEEN HELD SINCE 2015; 32 SOLUTIONS HAVE BEEN PRESENTED.

The first LPIRC is reported in a special session in 2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) [3]. The 2015 winner's method is published in [4]. A summary of the first two years' LPIRC is available in [5]. The 2016 winner's method is published in [6].

II. SUMMARY OF WINNERS' METHODS

This section summarizes the methods of the champions.

A. 2015 Winner

Tsinghua University won the first LPIRC in 2015. The team used Fast R-CNN (Regions with Convolutional Neural Network) as a baseline solution. Additional modifications on the Fast R-CNN method are made to fit the specific platform and achieve trade-off between speed and accuracy on embedded systems. The team proposed a multi-stage pipelined implementation on the embedded CPU+GPU platform with duplicated module-parallelism to make full use of the limited computation resources. The system used the Jetson TK1 platform and achieved the best score in 2015.

Year	mAP	Energy (WH)	Score	Ratio
2015	0.02971	1.634	0.0182	1.00
2016	0.03469	0.789	0.0440	2.42
2017	0.24838	2.082	0.1193	6.56

TABLE II
SCORES OF EACH YEAR'S CHAMPION.

B. 2016 Winner

The team from Chinese Academy of Science won 2016 LPIRC. The team evaluated two different object detection architectures (BING+FAST-RCNN and Faster-RCNN) on Jetson TX1. The feature-extracting CNN dominated the speed, accuracy and power consumption of the systems; thus, the team explored the design space to find the most energy-effective network implementation. To expand the design space of RCNN-based detection framework, the team considered multiple design parameters including data representation precision, input approximation level, network hyper-parameters, and finally converged to the solution. The team applied previously proposed methods for design-space exploration to the detection frameworks such as YOLO9000/SSD/R-FCN with NVDIA Jetson TX2, and compare the speed, energy consumption and accuracy. The team evaluated a new implementation using FPGA but discovered the energy efficiency was inferior than using GPGPU. GPGPU's abundant SP cores allowed it to batch-process the images in parallel.

C. 2017 Winner

Seoul National University was the winner of 2017 LPIRC. Among three high accuracy, high speed, and low energy consumption, the team considered the trade-off between accuracy and speed first and selected Nvidia Jetson TX2 as the hardware platform and tiny-YOLO as the object detection algorithm. The original tiny-YOLO took 1150 seconds (longer than the 10-minute limit) to process 20,000 images with 27.5% mAP. The team performed the following improvements: (1) To increase the throughput, the team used a pipelined the algorithm, running the NMS part on the multicore CPU of TX2. The processing time was reduced to 660 seconds. (2) To improve the GPU performance, the team applied a rank-reduction technique using tucker decomposition, selectively from the 5th to the 8th convolution layer. Then, the processing time was reduced to 530 seconds, and the CPU became the performance bottleneck. (3) The team parallelized the NMS computation with multithreading and reduced the processing time to 502 seconds. GPU became the bottleneck again. (4) FP16 computation from the 4th convolution layer was used. This optimization further reduced processing time to 460 seconds. (5) The team explored the operating frequencies of CPU and GPU: reducing the CPU frequency to 1.114 GHz and GPU frequency to 1.122 GHz and reduced the energy consumption by 13.15 %. (6) The implementation was tested with the competition setup and it was observed that communication with the referee system might become the performance bottleneck. The system overlapped communication and computation and reduced the communication volume.

III. Rules and Lessons Learned

Competitions have been shown as an effective way to encourage researchers to solve difficult and well-defined problems and attract the interest of the general public. The DARPA Grand Challenge is attributed as a major propeller of the technologies for autonomous vehicles. The winner of the Ansari X Prize demonstrated that it was feasible to build a reusable spacecraft. Competitions allow researchers to evaluate their solutions using the same metrics (for LPIRC, the ratio of mAP and energy) and the same benchmarks (the test images). LPIRC aims to achieve the similar goals: encourage innovative solutions for a well-defined problem and attract public attention (without the million-dollar prize).

LPIRC uses the training data from ImageNet and the test images (Figure 2) are also “ImageNet-like”. LPIRC is different from the ImageNet Competition (also called ImageNet Large Scale Visual Recognition Challenge, ILSVRC) in three ways: (1) LPIRC is an on-site competition. Contestants must bring their systems to compete. (2) In LPIRC, each solution has ten minutes. ILSVRC has no time limit. (3) Most important, LPIRC considers the energy consumption.

To give contestants the most flexibility, LPIRC has no restriction on the hardware or software platforms. Over the three years, contestants brought a wide range of hardware, including mobile phones, tablets, laptop computers, desktop computers, experimental boards, FPGA, etc. The only requirement is that a contestant’s system must be able to retrieve images from the referee system using HTTP GET commands and to report answers using HTTP POST commands. The source code of the referee system (HTTP server) and a sample client is available at `https://github.com/ieeelpirc`.

Over the years there has been some confusion about whether it is potentially advantageous to process only some images and stop early in order to “game” the evaluation criteria. The short answer is “No”. The ground truth used for computing mAP includes the entire test set, regardless of how many images a solution processes. For instance, if only 1 out of a 100 images is processed then we would expect the recall would be no more than about 1 one hundredth, and the mAP would be about 1 one hundredth of the mAP of the same system run on all the images. Generally we expect both mAP and energy usage to increase as more images are processed. If a solution has a consistent recognition accuracy over the dataset and a constant power consumption, then it would not matter how many images are processed. The LPIRC score should be similar modulo some sampling variation and should converge as more images are processed.

IV. Future LPIRC

The first three years LPIRC has successfully attracted many researchers’ interest. The significant improvement in the winners’ scores is a testament of the accomplishments. In the first three years, LPIRC has no restriction about the hardware nor software. Such freedom also has drawbacks: every contestant has to create a completely functional system. Starting from 2018, a new track will allow contestants to use Caffe 2 and a software development kit (SDK) will be available, as well as training materials. This SDK aims to lower the entry barrier so that more researchers and students can participate and present their innovative solutions. Depending on the enrollment number, an online competition may be also organized before the onsite competition to help the teams better prepare for the competition and to select the final list for the onsite competition. This new track is supported by Facebook.

LPIRC 2015-2017 used ImageNet-like data for training and testing. As point out in [7], different datasets have unique characteristics. Each image in ImageNet has only a few objects. These objects usually reside near the centers of images and each object often have many (thousands) pixels. Future LPIRC may use different types of images or videos as the datasets.

Beyond 2018, it may be feasible to create a “Grand Challenge of Low-Power Image Recognition”— contestants need to create systems that use very little energy for recognizing objects at a high speed and high accuracy. The scores of LPIRC winners suggest that the following requirements would be more than three orders of magnitude beyond the best technologies available today: (1) A system consumes at most 0.1 Watt. (2) The system can process more than 100 images (or video frames) at high resolutions (12 MP or higher) per second. (3) More than 100 objects appear in each image (or video frame) and these objects belong to 1,000 different categories. (4) The recognition accuracy is more than 99.99%. We are actively looking for industrial supports from both technical and educational perspectives.

V. Conclusion

This paper summarizes the first three years of Low-Power Image Recognition Challenge (LPIRC) and the methods used by the winners. Many researchers presented their solutions and the winners’ scores improve dramatically since 2015. It is expected that the new track in LPIRC 2018 will attract even more researchers and accelerate the progress of this important challenge. More details about LPIRC can be found at `https://rebootingcomputing.ieee.org/lpirc`.

Acknowledgments

LPIRC 2015-2017 is supported by IEEE Rebooting Computing, IEEE GreenICT, IEEE Council on superconductivity, IEEE Council on Electronic Design Automation, Nvidia, and Xilinx. LPIRC 2018 will be supported by IEEE Rebooting Computing and Facebook.

References

[1] ImageNet. `http://www.image-net.org/`.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[3] Y. H. Lu, A. M. Kadin, A. C. Berg, T. M. Conte, E. P. DeBenedictis, R. Garg, G. Gingade, B. Hoang, Y. Huang, B. Li, J. Liu, W. Liu, H. Mao, J. Peng, T. Tang, E. K. Track, J. Wang, T. Wang, Y. Wang, and J. Yao. Rebooting computing and low-power image recognition challenge. In *2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 927–932, Nov 2015.

[4] H. Mao, S. Yao, T. Tang, B. Li, J. Yao, and Y. Wang. Towards real-time object detection on embedded systems. *IEEE Transactions on Emerging Topics in Computing*, PP(99):1–1, 2017.

[5] K. Gauen, R. Rangan, A. Mohan, Y. H. Lu, W. Liu, and A. C. Berg. Low-power image recognition challenge. In *2017 22nd Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 99–104, Jan 2017.

[6] C. Wang, Y. Wang, Y. Han, L. Song, Z. Quan, J. Li, and X. Li. Cnn-based object detection solutions for embedded heterogeneous multicore socs. In *2017 22nd Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 105–110, Jan 2017.

[7] Kent Gauen, Ryan Dailey, John Laiman, Yuxiang Zi, Nirmal Asokan, Yung-Hsiang Lu, George K. Thiruvathukal, Mei-Ling Shyu, and Shu-Ching Chen. Comparison of visual datasets for machine learning. In *IEEE Conference on Information Reuse and Integration*, 2017.