Applying Automatic Tuning to Hyper-parameter Optimization of Machine Learning Programs for Super-Resolution

XUAN YANG †1 SORATARO FUJIKA †1 YUGA YAJIMA †1 AKIHIRO FUJII †1 TERUO TANAKA †1
KAZUTOSHI AKITA †2 NORIMICHI UKITA †2 SATOSHI OHSHIMA †3
†1Kogakuin University †2Toyota Technological Institute †3Nagoya University

Introduction

To optimize hyperparameters for better learning results of machine learning programs, we applied our developed software auto-tuning tool “DSICE” (d-spline Iterative Collinear Exploration). Since machine learning programs take a long time to execute each time, the time required for automatic tuning is also enormous. In order to reduce the execution time, we propose a two-step learning method of pre-learning and fine-tuning, and run multiple jobs in parallel on the GPU cluster.

Automatic Tuning Tool “DSICE” [1]

As an algorithm to estimate an optimum combination of performance parameters(*pps*), repeat the following two steps.

(1) Directional exploration to determine the direction of exploration.
- After the horizontal and vertical axis exploration is completed, explore the oblique direction.

(2) Collinear exploration to find the optimal point.

In this study, DSICE uses hyperparameter as *pps* in machine learning programs.

- Input: Combination of hyperparameter, Performance value.
- Output: The best hyperparameter combination.

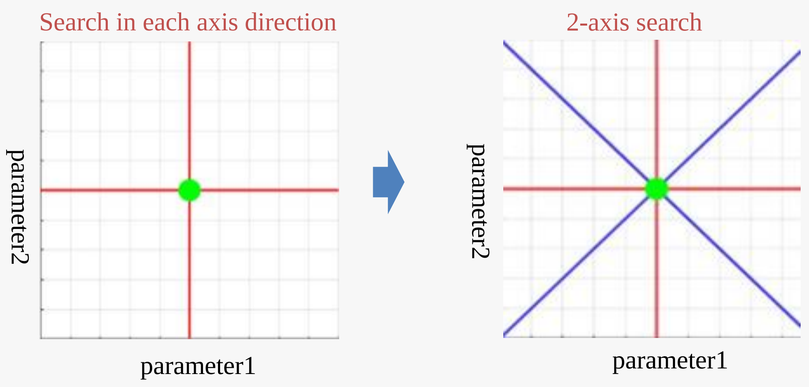

Fig.1 Determining a line

Machine Learning Program for Super-Resolution

Super-resolution is a technique that converts low-resolution images into high-resolution images.

- Program: Dense Deep Back-Projection Networks(D-DBPN)[2]
  - The number of epoch is 2000. It takes 92 hours to run once.
- Purpose: 4x magnification. Ensure the quality of the image.
- Performance value[3]:
  - Root Mean Square Error (RMSE), Perceptual Index (PI).
  - Set the range (<12.5) for RMSE and execute automatic tuning of PI.

Methods to reduce program execution time

1. Two-step learning: pre-learning and fine-tuning
   - Pre-learning: Set epoch to 1000 and run it once.
   - Fine-tuning: Use pre-learning model and set epoch to 250 for each execution.
     - The value of Loss is stable up to epoch 250.
     - The running time is reduced to 1/8.
       - epoch 2000 (original D-DBPN) → epoch 250.
2. Run in parallel
   - Execute multiple jobs in parallel on supercomputer with multiple GPUs (Fig.2).

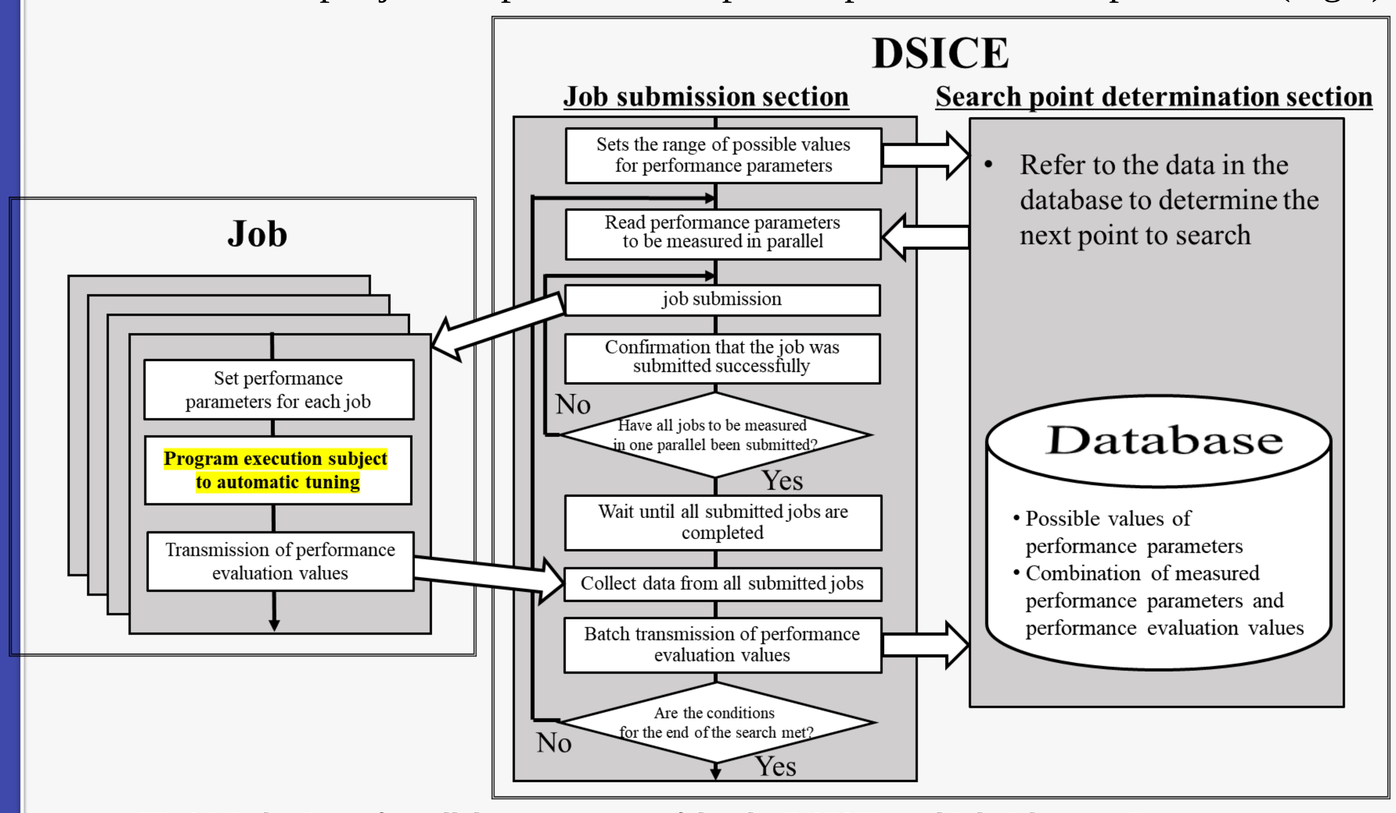

Fig.2 Mechanism of parallel measurement of data by DSICE's multiple job execution

Experiment overview

- Super computer: Nagoya University “Flow” Type II
  - The maximum number of simultaneous executions is 50.
- Setting of hyperparameter
  - There are 15, 16, 16, 16 possible values for each of the 4 hyperparameters.

Experiment Result

Fig.3 shows that the best PI is 2.517 when RMSE is 12.386.
The best attainable PI improves as the allowable distortion level increases.

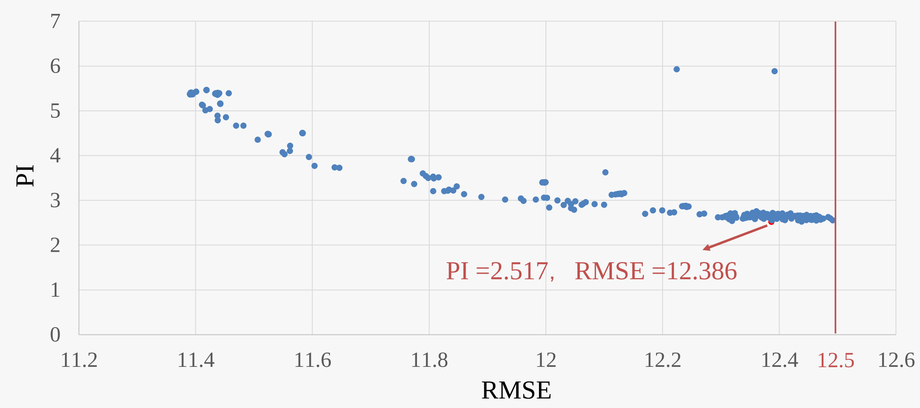

Fig.3 Relationship between RMSE and PI

Fig.4 shows program execution status.
Parallel operations in supercomputing reduced execution time to 1/27.
- 4905 hours (in case of sequential operation) → 182 hours.
- The maximum number of parallel operations is 40.
- Total number of executed job was 545 times.

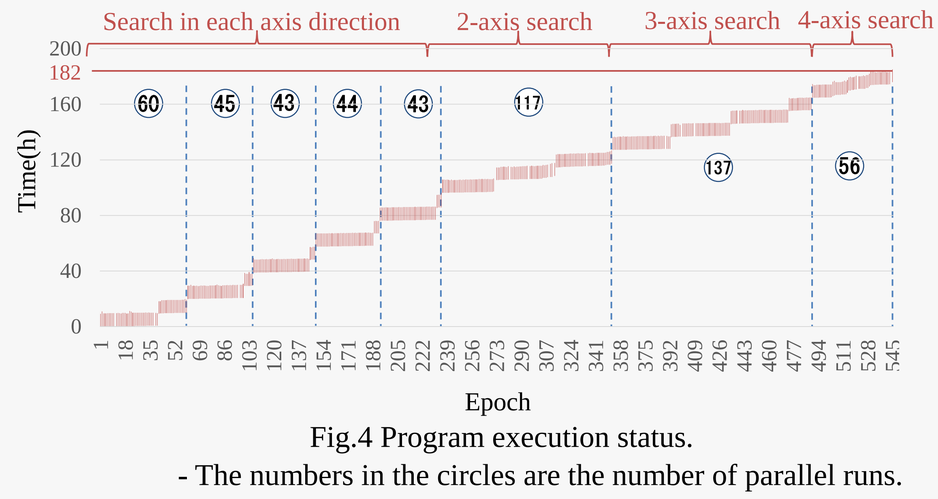

Fig.4 Program execution status.
- The numbers in the circles are the number of parallel runs.

Conclusion

In this study, hyperparameter estimation was performed by applying the parallelized automatic tuning tool “DSICE” to the super-resolution program D-DBPN. Used pre-learning and fine-tuning to reduce each run time to 1/8. Used the parallel processing environment of GPU cluster supercomputer, which reduced execution time to 1/27. Overall, the time was reduced to 1/216 of the original. Fig.5 is the image of the best performance value, magnified 4x from the original image, which is clearer by eyes.

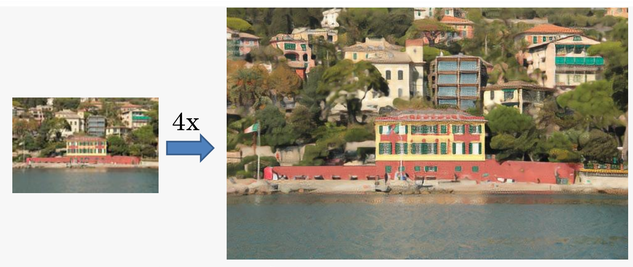

Fig.5 Result image

Acknowledgments

This work is supported by JHPCN in Japan (Project ID: jh220044) and JSPS KAKENHI Grand Number JP 18K11340.

Reference

[1] M. Mochizuki, A. Fujii, T. Tanaka, T. Katagiri, “Fast Multidimensional Performance Parameter Estimation with Multiple One-dimensional d-Spline Parameter Search,” iWAPT in conjuction with IEEE IPDPS (2017).
[2] M. Haris, G. Shakhnarovich, N. Ukita, “Deep Back-Projection Networks for Super-Resolution,” IEEE/CVF (2018).
[3] Y.Blau, R.Mechrez, R.Timofte, T.Michaeli, L.Zelnik-Manor, “The 2018 PIRM challenge on perceptual image super-resolution,” ECCV (2018).