Using an evolutionary algorithm to optimize a building footprint detection deep learner's hyper-parameters

Mark Coletti <colettima@ornl.gov>, Jeanette Weaver, Anne Berres, Lexie Yang, Dalton Lunga, Kuldeep Kurtre, Ben Liebersohn, Jibonananda Sanyal
Computational Sciences and Engineering Division
Oak Ridge National Laboratory

Abstract

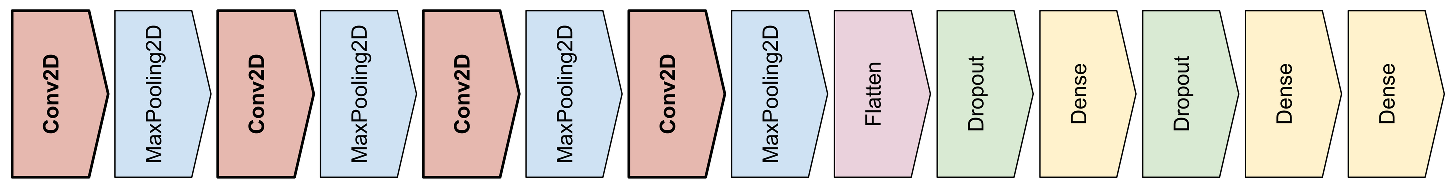
- use a deep learner to identify building footprints from satellite imagery that inform population models
- deep learners have associated hyper-parameters that influence their training
  - which can have a pronounced effect on the quality of learned models
  - there is little guidance for ideal hyper-parameter settings
- the state-of-the-art relies on uniform or random hyper-parameter sweeps to improve model accuracy

We share the results of using an alternative approach to deep learner hyper-parameter tuning that uses an evolutionary algorithm to improve the accuracy of our building footprint detection models. We found that the kernel and batch sizes surprisingly differed from what a brute force uniform grid approach discovered.

Motivation

Members of the ORNL Geographic Information Science and Technology group use deep-learning to detect building footprints in satellite imagery. The buildings in this image have been marked in purple based on applying a deep-learner model.

Deep-Learner Architecture Used



Optimized Hyper-parameters

- **kernel sizes** for all four Conv2D layers $\in \{3, 5, 7, 9, 11, 13, 15, 17\}$
- **batch size** $\in [22, 150]$

DL Set-up

Stochastic Gradient Descent (SGD) optimizer

- categorical cross-entropy
- momentum 0.9
- Nesterov momentum
- learning rate of 0.00273
- decay rate of 0

Training and Validation

- CUDA 8
- $1 \times 144 \times 144$ "image chips"
- 18,000 training images
- 2,300 validation images
- Daykundi Province, Afghanistan

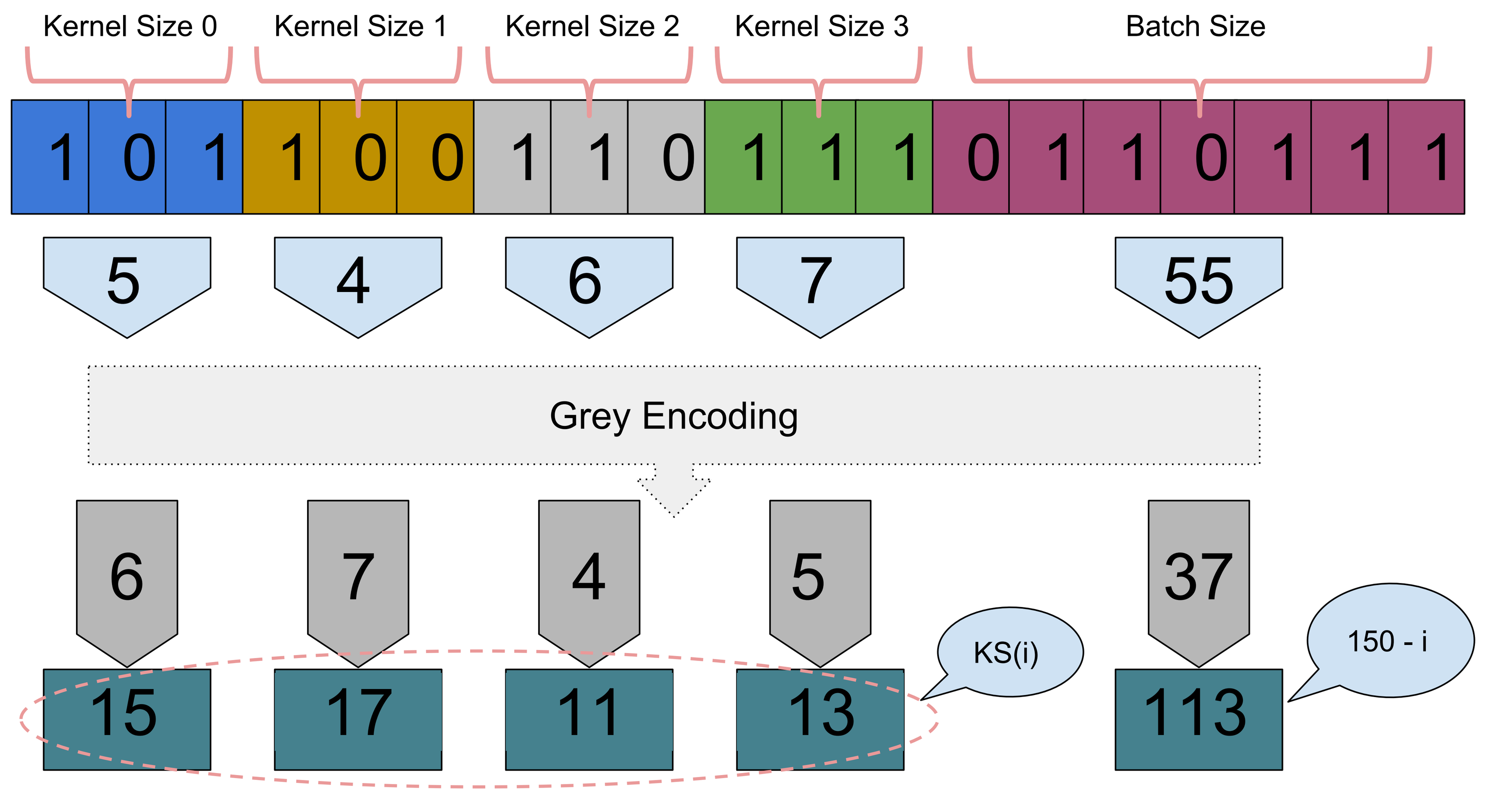
EA Set-up

parents	30
offspring	312
generations	5
mutation rate	0.1
parent selection	binary tournament
crossover	uniform
Gray encoding	true
survival selection	truncation of parents and offspring
runs	5
Titan nodes	313
Max. hours per run	12

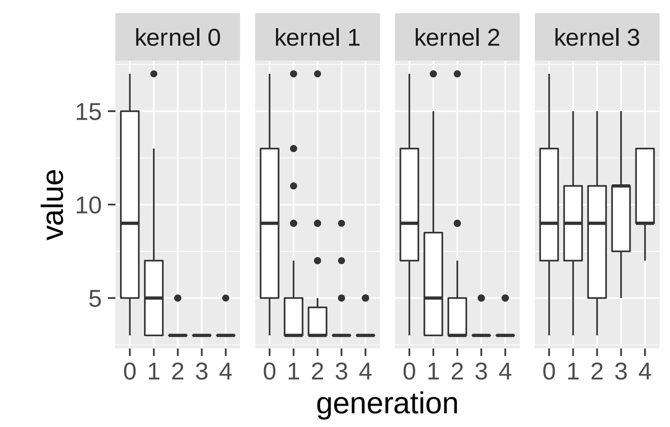
Brute Force Results

- Baseline comparison to brute force gride search of hyper-parameters.
- Converged on a kernel size 3 for all Conv2D layers, and a batch size of 160.
- Took 14 days of computation on a dedicated Nvidia DGX.

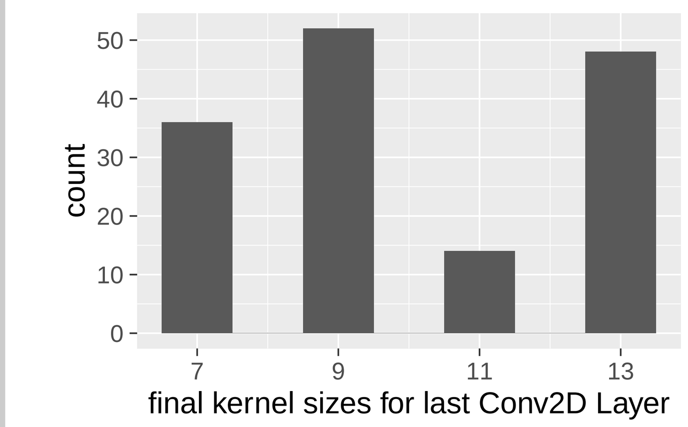
Problem Representation



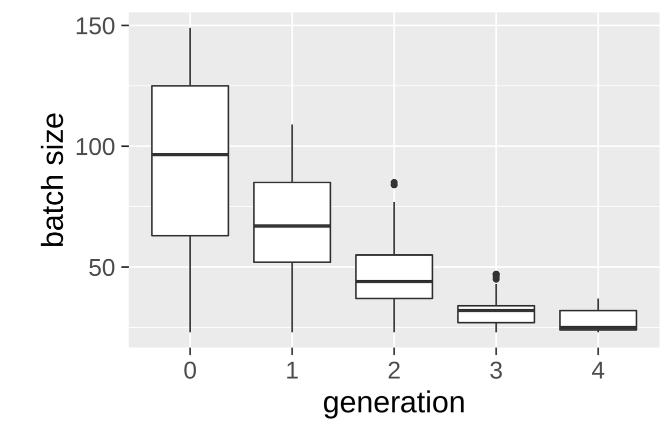
Kernel Sizes



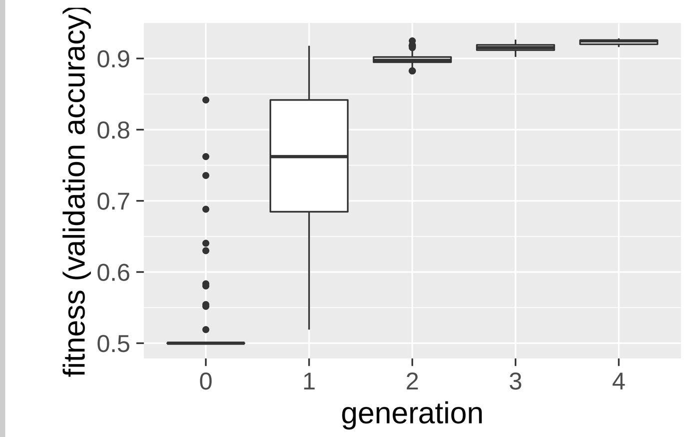
Final Sizes



Batch Sizes



Fitnesses



References & Acknowledgements

[1] James Bergstra and Yoshua Bengio. "Random Search for Hyper-Parameter Optimization". In: *Journal of Machine Learning Research* 13 (Feb 2012), pp. 281–305. issn: ISSN 1533-7928. url: http://www.jmlr.org/papers/v13/bergstra12a.html.

[2] Richard A. Caruana and J. David Schaffer. "Representation and Hidden Bias: Gray vs. Binary Coding for Genetic Algorithms". In: *Machine Learning Proceedings 1988*. Ed. by John Laird. San Francisco (CA): Morgan Kaufmann, 1988, pp. 153–161. isbn: 978-0-934613-64-4. doi: 10.1016/B978-0-934613-64-4.50021-9. url: https://www.sciencedirect.com/science/article/pii/B9780934613644500219.

[3] Kenneth A. De Jong and William M. Spears. "A formal analysis of the role of multi-point crossover in genetic algorithms". In: *Annals of Mathematics and Artificial Intelligence* 5.1 (1992), pp. 1–26. issn: 1573-7470. doi: 10.1007/BF01530777. url: http://dx.doi.org/10.1007/BF01530777.

[4] Jerome E Dobson et al. "LandScan: a global population database for estimating populations at risk". In: *Photogrammetric engineering and remote sensing* 66.7 (2000), pp. 849–857.

[5] Steven R. Young et al. "Optimizing Deep Learning Hyper-Parameters Through an Evolutionary Algorithm". In: MLHPC '15 (2015), 4:1–4:5. doi: 10.1145/2834892.2834896. url: http://doi.acm.org/10.1145/2834892.2834896 (visited on 01/23/2017).

This research was conducted under an ALCC grant at the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory.