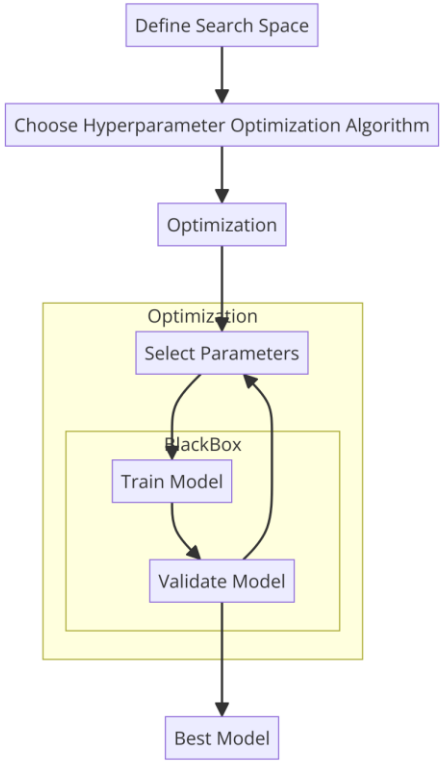
TopoHyperDrive: Accelerating Meta-Search in Hyperparameter Optimization through Topological Analysis

VLADISLAV ZHUZHEL, SKOLKOVO INSTITUTE OF SCIENCE AND TECHNOLOGY

Introduction:

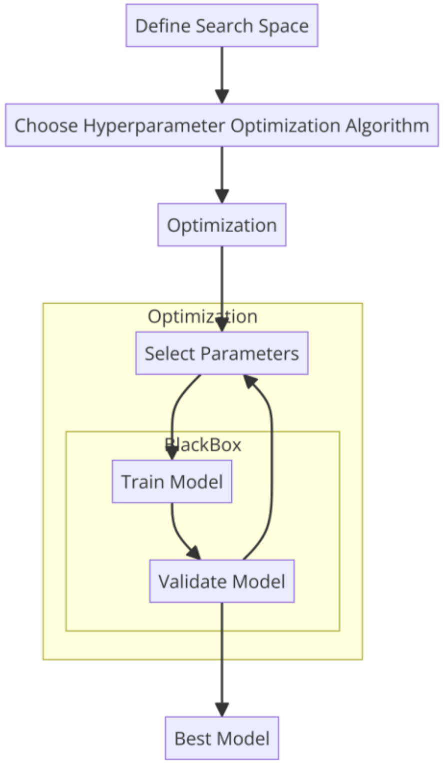
Hyperparameter Optimization Challenges

- Neural networks are pivotal in various applications like text generation, image recognition, and personalized healthcare
- Increasing complexity and size of neural networks intensify the demand on computational resources
- Selecting optimal hyperparameters is crucial but traditional methods are inefficient and costly
- Traditional method mostly consider the neural network as a black box and do not incorporate the peculiarities of the neural nets



Introduction:
Existing Solutions

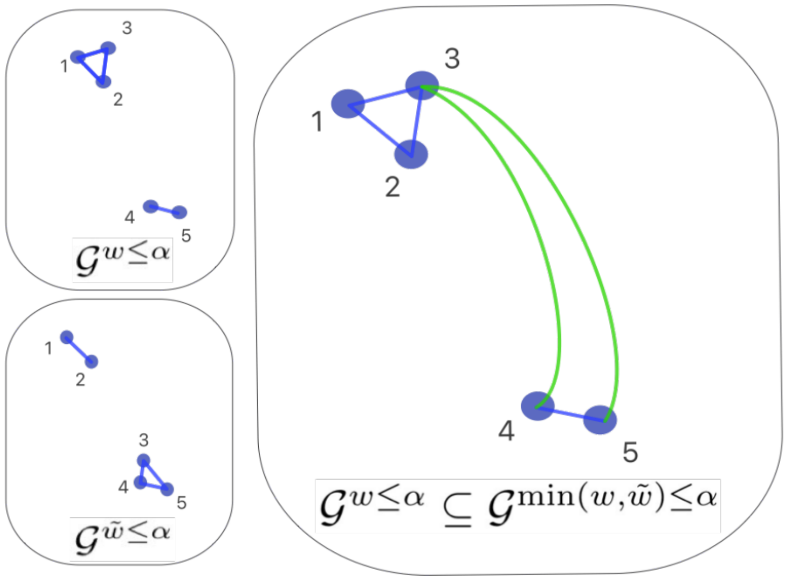
- Traditional Methods:
  - Grid Search
  - Random Search
- Advanced Methods
  - Bayesian Optimization
  - Hyperband
  - BOHB
- Limitations:
  - High Computational Cost
  - Lack of Insight into Model Structures



Method:

Leveraging Topological Analysis for Optimization

- RTD Definition:
  - Measures dissimilarity between neural network embeddings
  - Uses topological features in point clouds
- Why RTD?
  - Captures structural variances
  - Provides a nuanced exploration of model representations
- RTD computation:
  - Construct R-Cross-Barcode - Vietoris-Rips filtered complexes of the individual and joined graphs for the point clouds
  - Compute score: $\mathrm{RTD}(P, P') = \sum_i \mathrm{length}(\mathrm{Bar}_i)$

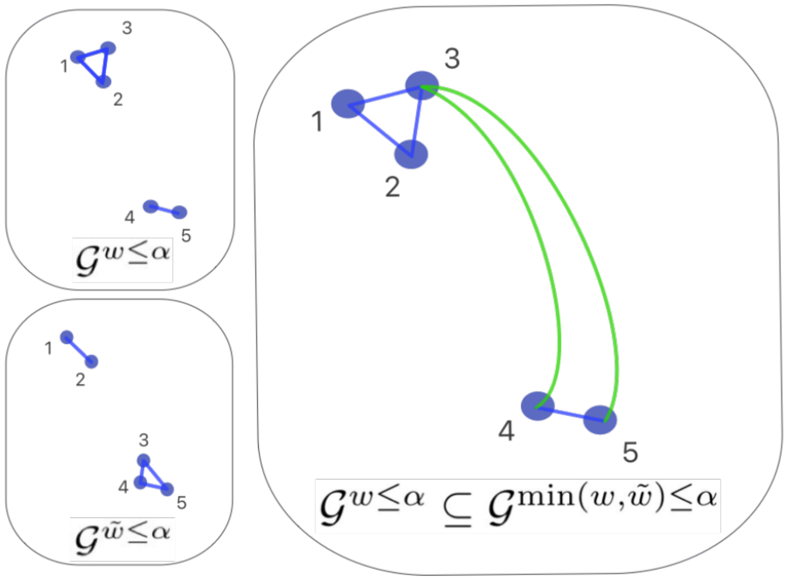


Barannikov, S., Trofimov, I., Balabin, N., and Burnaev, E. Representation topology divergence: A method for comparing neural network representations. arXiv preprint arXiv:2201.00058, 2021.

Method:

RTD Implication for Hyperparameter Optimization

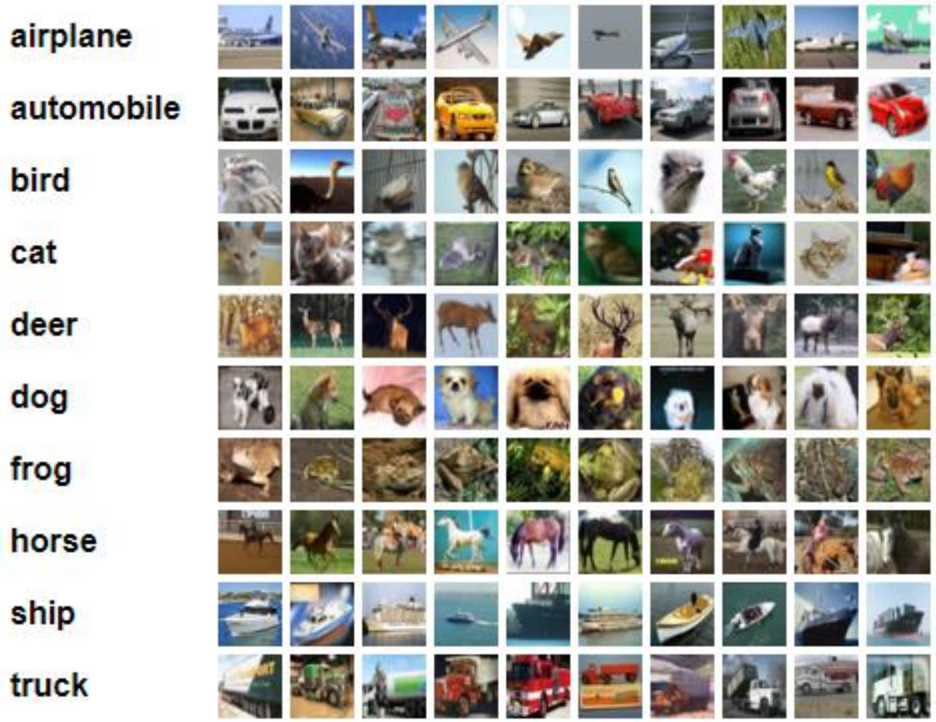
- One-to-Random approach
  - Initialize a random model
  - Compare the embeddings of the trained model with the random one
- Allows sole-objective and multi-objective setups
  - Sole-objective for unsupervised training
  - Multi-objective for supervised training
- Can be used with any Hyperparameter Optimization Method
  - We will consider Tree-structured Parzen Estimator (TPE) approach



Bergstra, J., Yamins, D., and Cox, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In International conference on machine learning, pp. 115–123. PMLR, 2013.

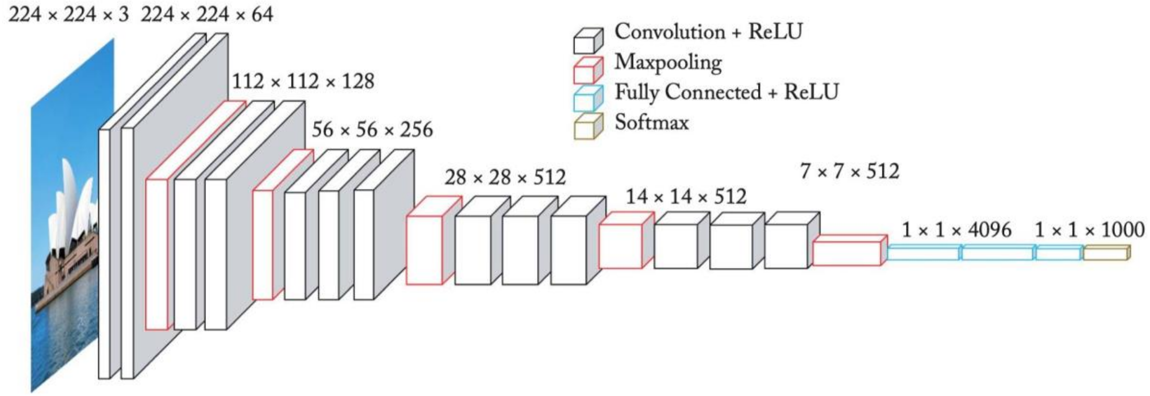
Experiments:
CIFAR-100 Dataset

- Overview
  - 60,000 colour images
  - 100 classes with 600 images each
- Image Specifications
  - 32x32 pixels
  - Diverse objects and living entities
- Relevance
  - Widely used for benchmarking machine learning algorithms
  - Challenging due to variability and diversity in images



Experiments:
Search Space

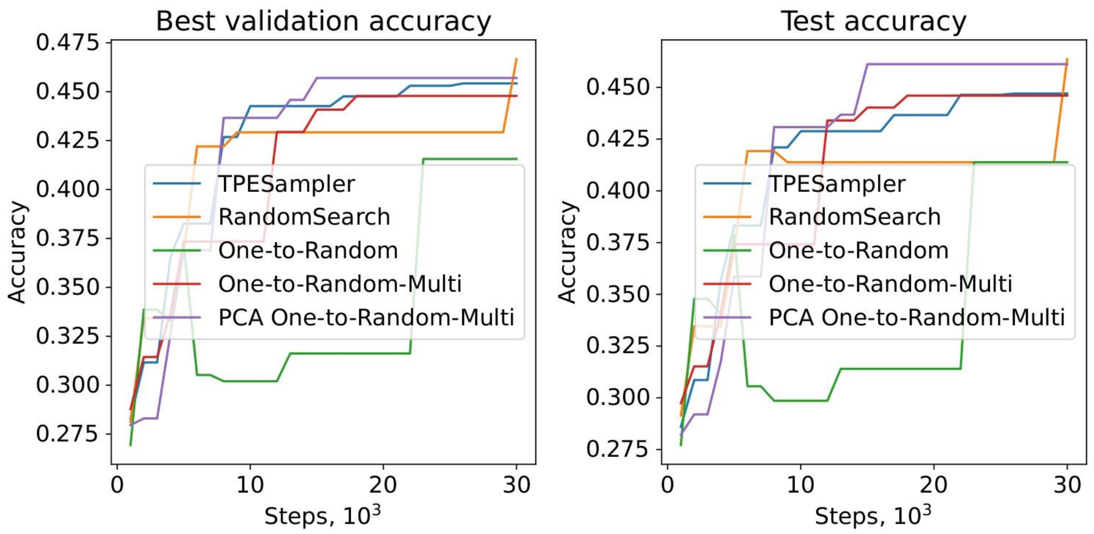
- Five-block architecture inspired by VGG
- Each block contains multiple convolutional layers
- Max pooling layers between blocks
- Hyperparameters:
  - Number of layers per block: 1 to 4
  - Number of filters per layer: 4 to 256
  - Hidden size in fully connected layers: 128 to 4096



Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

Results
Best accuracy

- PCA One-to-Random Multi-Objective – One-to-Random variant with additional dimensionality reduction to the dimension size 2 – top performer
- Full embeddings RTD converge slower due to more noisy data, may outperformer on a harder dataset or with further optimization
- Sole objective One-to-Random setup leads to quality drop in the first trials, but starts to grow later, may be used in unsupervised hyperparameter optimization



Further Improvements

Enhancing TopoHyperDrive

- One-to-Top Performer Divergence
  - Identify models divergent from the best-performing model thus far
  - Optimize using dual objectives to avoid oscillation
- One-to-All Divergence
  - Configure hyperparameters to diverge from all previous iterations
  - Aim to explore a broad range of representation topologies
- Additional Optimization Algorithms
  - Integrate with Hyperband and BOHB

Conclusion

- Innovation in Hyperparameter Optimization
  - Leveraged topological analysis to enhance search efficiency using Representation Topology Divergence (RTD)
- Improved Performance
  - Faster convergence and higher accuracy compared to traditional methods
  - Ability to conduct a hyperparameter search in an unsupervised setup
  - Demonstrated on CIFAR-100 dataset
- Future Directions
  - Potential integration with other optimization algorithms
  - Further refinement and broader application of RTD