Finetune, Optimize and Search: AutoML for Vision Datasets

Shanmugapriya Kanagasabapathi, Tanya Dora, Aaryan Bhandari
automl_24_exam

Modality 2/2

Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Bonus
Literature

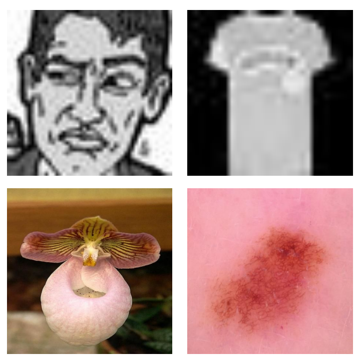
Goal

Design a generic AutoML pipeline which yields an architecture that performs classification on different image datasets.

Methods Used

- Bayesian Optimization
- Hyperparameter Optimization
- Differentiable Architecture Search
- Finetuning of ResNet34
- Data Augmentation
- Model Checkpointing

Our Approach

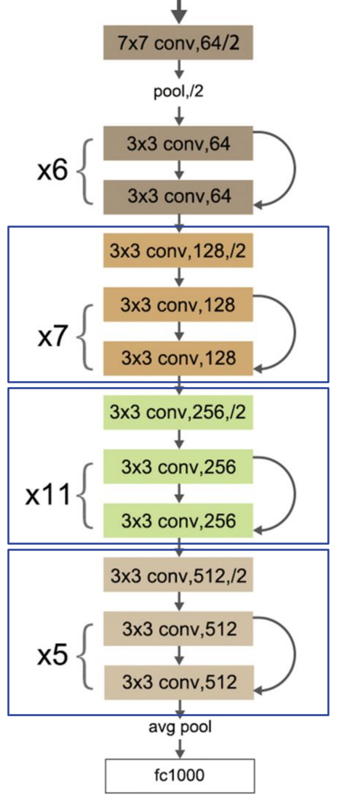


Data Augmentations:
Resize(256)
CenterCrop(224)
RandomHorizontalFlip()
RandomRotation(30)
Normalize(
[0.485, 0.456, 0.406],
[0.229, 0.224, 0.225])

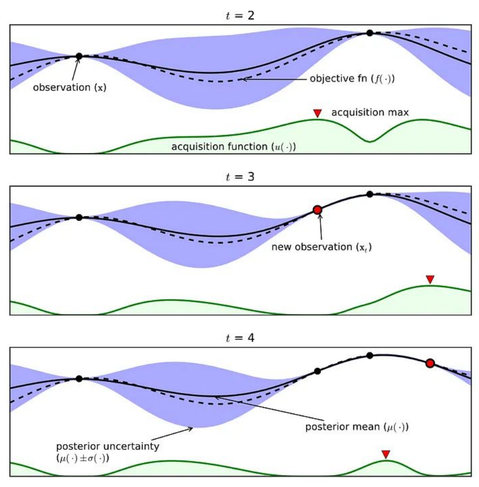
Train Split
Test Split

80% Train* Split
20% Valid Split

34-layer residual
image



[4]

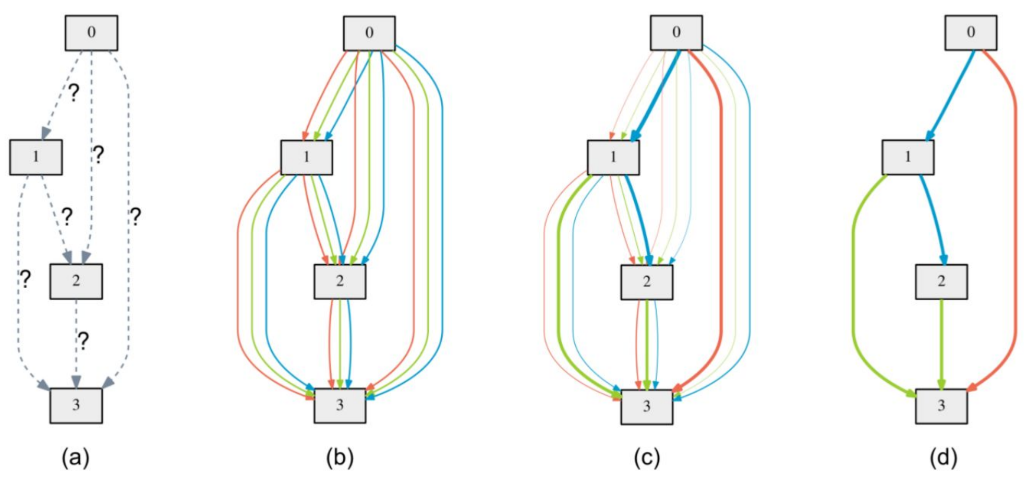


[3]

Bayesian Optimization with Gaussian Processes:
Learning Rate: 1e-6 to 1e-1
Momentum: 0.1 to 0.9
Weight Decay: 0 to 1e-2
Optimizer: [adam, sgd, rmsprop]
Weights fixed until layer: [2, 3, 4]
Number of calls: 30
Each call evaluates model after 1 epoch

Speedup of downstream tasks by building a new dataset based on avg pool output

Finetuned Model
Feature Dataset
fc1000

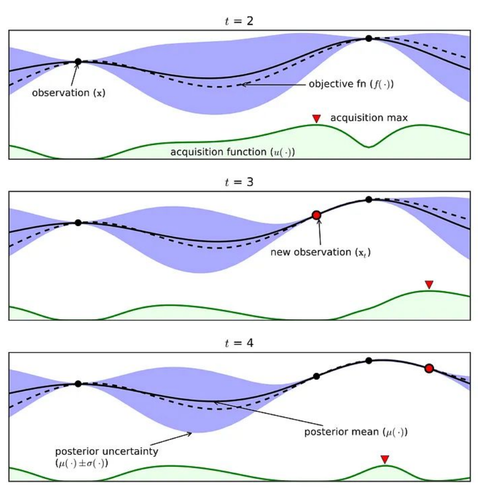


[2]

fc1000 gets replaced by an MLP found by DARTS where mixed operation includes blocks with:
Hidden size: [32, 64, 128, 256]
Activation: [ReLU, Tanh, Sigmoid]
Layer number: [2, 3, 4]

DARTS Classifier

Bayesian Optimization with Gaussian Processes:
Learning Rate: 1e-6 to 1e-2
Momentum: 0.1 to 0.9
Weight Decay: 1e-6 to 1e-2
Optimizer: [adam, sgd, rmsprop]
Number of calls: 30
Each call evaluates model after 8 epochs



[3]

Final Model

Finetuned Model
Trained DARTS Classifier

Resources Used

For development and AutoML:
- 1 Nvidia GTX 1050Ti
- 4 Intel Core i5-7300HQ CPUs
- Total compute estimate: 180 GPU-h

Workforce:
- 3 full weeks on average (120 hours)

Empirical Results

Dataset	Our Method	Baseline
Emotions 	F1-score: 0.63 Precision: 0.64 Accuracy: 0.66	N/A N/A 0.40
Fashion 	F1-score: 0.93 Precision: 0.93 Accuracy: 0.93	N/A N/A 0.88
Flowers 	F1-score: 0.93 Precision: 0.94 Accuracy: 0.94	N/A N/A 0.55
Skin Cancer 	F1-score: 0.77 Precision: 0.79 Accuracy: 0.87	N/A N/A 0.71

Best parameters of Bayesian Optimization for finetuning on skin cancer dataset:
Learning Rate: 1e-4
Momentum: 0.31
Weight Decay: 0
Optimizer: adam
Weights fixed until layer: 2

Best parameters of Bayesian Optimization for DARTS on skin cancer dataset:
Learning Rate: 1e-3
Momentum: 0.2
Weight Decay: 1e-3
Optimizer: sgd

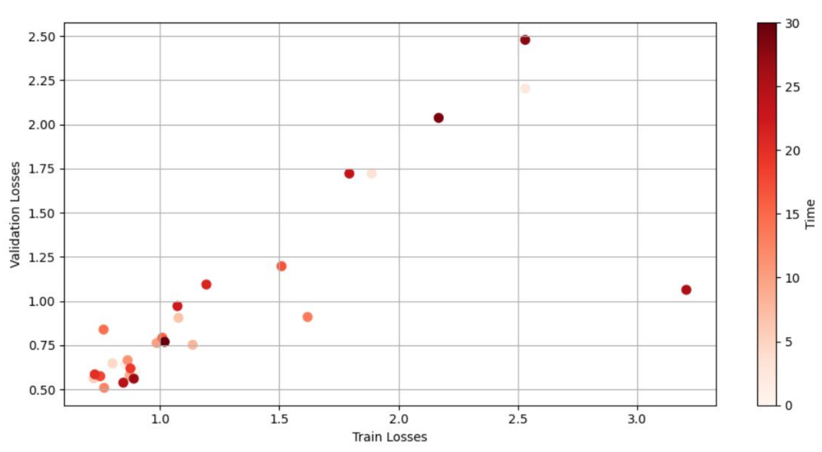


Fig. 1: Validation loss vs. training loss in the Bayesian Optimization for finetuning

Best DARTS architecture:
(0): Linear(in_features=512, out_features=32, bias=True)
(1): ReLU()
(2): Linear(in_features=32, out_features=7, bias=True)

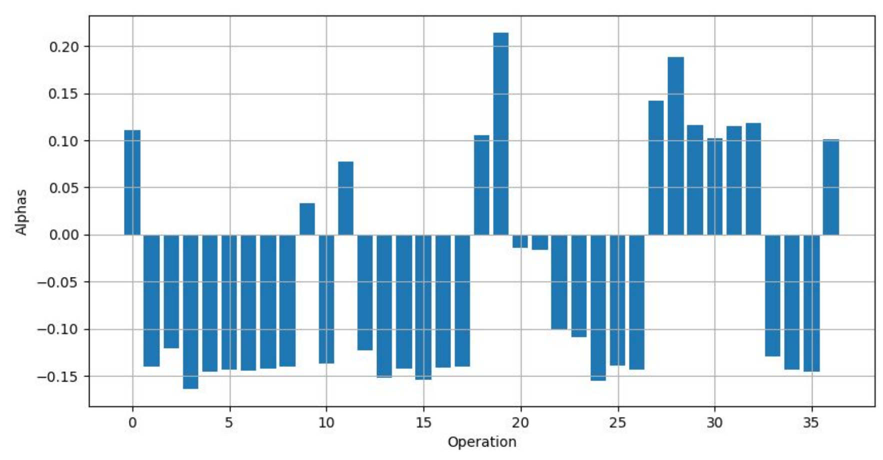


Fig. 2: Alpha distribution after training DARTS

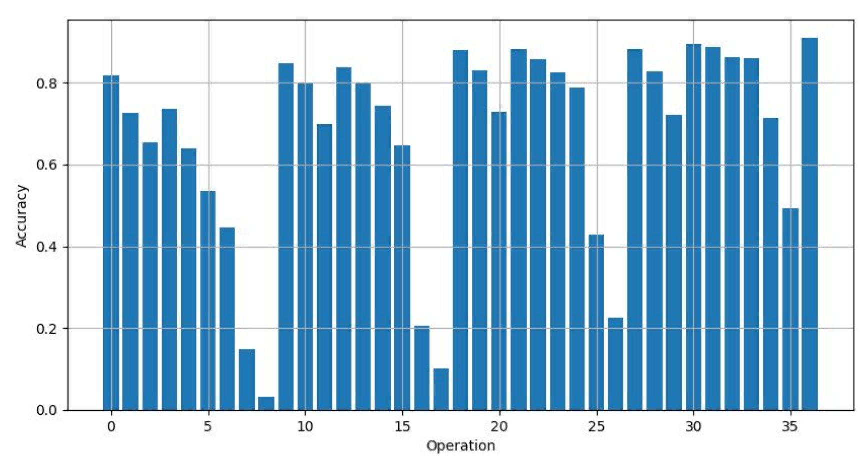


Fig. 3: Accuracies achieved using each of the DARTS operations

Number of queries for test score generation: **1**

Additional References:
[1] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
[2] Liu, H., Simonyan, K., & Yang, Y. (2018). DARTS: Differentiable Architecture Search. *ArXiv, abs/1806.09055*.
[3] Brochu, E., Cora, V.M., & Freitas, N.D. (2010). A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *ArXiv, abs/1012.2599*.
[4] Zhang, H., Mo, J., Jiang, H., Li, Z., Hu, W., Zhang, C., Wang, Y., Wang, X., Liu, C., Zhao, B., Zhang, J., & Zhang, K. (2020). Deep Learning Model for the Automated Detection and Histopathological Prediction of Meningioma. *Neuroinformatics, 19*, 393 - 402.