

---

## awesome-AutoML-and-Lightweight-Models

A list of high-quality (newest) AutoML works and lightweight models including **1.) Neural Architecture Search, 2.) Lightweight Structures, 3.) Model Compression, Quantization and Acceleration, 4.) Hyperparameter Optimization, 5.) Automated Feature Engineering.**

This repo is aimed to provide the info for AutoML research (especially for the lightweight models). Welcome to PR the works (papers, repositories) that are missed by the repo.

### 1.) Neural Architecture Search

#### [Papers]

**Gradient:** - When NAS Meets Robustness: In Search of Robust Architectures against Adversarial Attacks | [CVPR 2020] + gmh14/RobNets | [Pytorch]

- Searching for A Robust Neural Architecture in Four GPU Hours | [CVPR 2019]
  - D-X-Y/GDAS | [Pytorch]
- ASAP: Architecture Search, Anneal and Prune | [2019/04]
- Single-Path NAS: Designing Hardware-Efficient ConvNets in less than 4 Hours | [2019/04]
  - dstamoulis/single-path-nas | [Tensorflow]
- Automatic Convolutional Neural Architecture Search for Image Classification Under Different Scenes | [IEEE Access 2019]
- sharpDARTS: Faster and More Accurate Differentiable Architecture Search | [2019/03]
- Learning Implicitly Recurrent CNNs Through Parameter Sharing | [ICLR 2019]
  - lolemacs/soft-sharing | [Pytorch]
- Probabilistic Neural Architecture Search | [2019/02]
- Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation | [2019/01]
- SNAS: Stochastic Neural Architecture Search | [ICLR 2019]
- FBNet: Hardware-Aware Efficient ConvNet Design via Differentiable Neural Architecture Search | [2018/12]
- Neural Architecture Optimization | [NIPS 2018]

- 
- renqianluo/NAO | [Tensorflow]
  - DARTS: Differentiable Architecture Search | [2018/06]
    - quark0/darts | [Pytorch]
    - khanrc/pt.darts | [Pytorch]
    - dragen1860/DARTS-PyTorch | [Pytorch]

### **Reinforcement Learning:**

- Template-Based Automatic Search of Compact Semantic Segmentation Architectures | [2019/04]
  - Understanding Neural Architecture Search Techniques | [2019/03]
  - Fast, Accurate and Lightweight Super-Resolution with Neural Architecture Search | [2019/01]
    - falsr/FALSR | [Tensorflow]
  - Multi-Objective Reinforced Evolution in Mobile Neural Architecture Search | [2019/01]
    - moremnas/MoreMNAS | [Tensorflow]
  - ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware | [**ICLR 2019**]
    - MIT-HAN-LAB/ProxylessNAS | [Pytorch, Tensorflow]
  - Transfer Learning with Neural AutoML | [**NIPS 2018**]
  - Learning Transferable Architectures for Scalable Image Recognition | [2018/07]
    - wandering007/nasnet-pytorch | [Pytorch]
    - tensorflow/models/research/slim/nets/nasnet | [Tensorflow]
  - MnasNet: Platform-Aware Neural Architecture Search for Mobile | [2018/07]
    - AnjieZheng/MnasNet-PyTorch | [Pytorch]
  - Practical Block-wise Neural Network Architecture Generation | [**CVPR 2018**]
  - Efficient Neural Architecture Search via Parameter Sharing | [**ICML 2018**]
    - melodyguan/enas | [Tensorflow]
    - carpedm20/ENAS-pytorch | [Pytorch]
  - Efficient Architecture Search by Network Transformation | [**AAAI 2018**]

**Evolutionary Algorithm:** - Single Path One-Shot Neural Architecture Search with Uniform Sampling | [2019/04]

- 
- DetNAS: Neural Architecture Search on Object Detection | [2019/03]
  - The Evolved Transformer | [2019/01]
  - Designing neural networks through neuroevolution | [**Nature Machine Intelligence 2019**]
  - EAT-NAS: Elastic Architecture Transfer for Accelerating Large-scale Neural Architecture Search | [2019/01]
  - Efficient Multi-objective Neural Architecture Search via Lamarckian Evolution | [**ICLR 2019**]

**SMBO:** - MFAS: Multimodal Fusion Architecture Search | [**CVPR 2019**]

- DPP-Net: Device-aware Progressive Search for Pareto-optimal Neural Architectures | [**ECCV 2018**]
- Progressive Neural Architecture Search | [**ECCV 2018**]
  - titu1994/progressive-neural-architecture-search | [Keras, Tensorflow]
  - chenxi116/PNASNet.pytorch | [Pytorch]

**Random Search:** - Exploring Randomly Wired Neural Networks for Image Recognition | [2019/04]

- Searching for Efficient Multi-Scale Architectures for Dense Image Prediction | [**NIPS 2018**]

**Hypernetwork:** - Graph HyperNetworks for Neural Architecture Search | [**ICLR 2019**]

**Bayesian Optimization:** - Inductive Transfer for Neural Architecture Optimization | [2019/03]

**Partial Order Pruning** - Partial Order Pruning: for Best Speed/Accuracy Trade-off in Neural Architecture Search | [**CVPR 2019**] + lixincn2015/Partial-Order-Pruning | [Caffe]

**Knowledge Distillation** - Improving Neural Architecture Search Image Classifiers via Ensemble Learning | [2019/03]

### [Projects]

- Microsoft/nni | [Python]
- MindsDB | [Python]

## 2.) Lightweight Structures

### [Papers]

**Image Classification:** - EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks | [**ICML 2019**] + tensorflow/tpu/models/official/efficientnet/ | [Tensorflow] + lukemelas/EfficientNet-PyTorch | [Pytorch]

- 
- Searching for MobileNetV3 | [2019/05]
    - kuan-wang/pytorch-mobilenet-v3 | [Pytorch]
    - leaderj1001/MobileNetV3-Pytorch | [Pytorch]

**Semantic Segmentation:** - CGNet: A Light-weight Context Guided Network for Semantic Segmentation | [2019/04] + wutianyiRosun/CGNet | [Pytorch]

- ESPNetv2: A Light-weight, Power Efficient, and General Purpose Convolutional Neural Network | [2018/11]
  - sacmehta/ESPNetv2 | [Pytorch]
- ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation | [**ECCV 2018**]
  - sacmehta/ESPNet | [Pytorch]
- BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation | [**ECCV 2018**]
  - ooooverflow/BiSeNet | [Pytorch]
  - ycszen/TorchSeg | [Pytorch]
- ERFNet: Efficient Residual Factorized ConvNet for Real-time Semantic Segmentation | [**T-ITS 2017**]
  - Eromera/erfnet\_pytorch | [Pytorch]

**Object Detection:** - ThunderNet: Towards Real-time Generic Object Detection | [2019/03]

- Pooling Pyramid Network for Object Detection | [2018/09]
  - tensorflow/models | [Tensorflow]
- Tiny-DSOD: Lightweight Object Detection for Resource-Restricted Usages | [**BMVC 2018**]
  - lyxok1/Tiny-DSOD | [Caffe]
- Pelee: A Real-Time Object Detection System on Mobile Devices | [**NeurIPS 2018**]
  - Robert-JunWang/Pelee | [Caffe]
  - Robert-JunWang/PeleeNet | [Pytorch]
- Receptive Field Block Net for Accurate and Fast Object Detection | [**ECCV 2018**]
  - ruinmessi/RFBNet | [Pytorch]
  - ShuangXielrene/ssds.pytorch | [Pytorch]
  - lzx1413/PytorchSSD | [Pytorch]

- 
- FSSD: Feature Fusion Single Shot Multibox Detector | [2017/12]
    - ShuangXielrene/ssds.pytorch | [Pytorch]
    - lzx1413/PytorchSSD | [Pytorch]
    - dlyldxwl/fssd.pytorch | [Pytorch]
  - Feature Pyramid Networks for Object Detection | [**CVPR 2017**]
    - tensorflow/models | [Tensorflow]

### 3.) Model Compression & Acceleration

#### [Papers]

**Pruning:** - The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks | [**ICLR 2019**] + google-research/lottery-ticket-hypothesis | [Tensorflow]

- Rethinking the Value of Network Pruning | [**ICLR 2019**]
- Slimmable Neural Networks | [**ICLR 2019**]
  - JiahuiYu/slimmable\_networks | [Pytorch]
- AMC: AutoML for Model Compression and Acceleration on Mobile Devices | [**ECCV 2018**]
  - AutoML for Model Compression (AMC): Trials and Tribulations | [Pytorch]
- Learning Efficient Convolutional Networks through Network Slimming | [**ICCV 2017**]
  - foolwood/pytorch-slimming | [Pytorch]
- Channel Pruning for Accelerating Very Deep Neural Networks | [**ICCV 2017**]
  - yihui-he/channel-pruning | [Caffe]
- Pruning Convolutional Neural Networks for Resource Efficient Inference | [**ICLR 2017**]
  - jacobgil/pytorch-pruning | [Pytorch]
- Pruning Filters for Efficient ConvNets | [**ICLR 2017**]

**Quantization:** - Understanding Straight-Through Estimator in Training Activation Quantized Neural Nets | [**ICLR 2019**]

- Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference | [**CVPR 2018**]

- 
- Quantizing deep convolutional networks for efficient inference: A whitepaper | [2018/06]
  - PACT: Parameterized Clipping Activation for Quantized Neural Networks | [2018/05]
  - Post-training 4-bit quantization of convolution networks for rapid-deployment | [**ICML 2018**]
  - WRPN: Wide Reduced-Precision Networks | [**ICLR 2018**]
  - Incremental Network Quantization: Towards Lossless CNNs with Low-Precision Weights | [**ICLR 2017**]
  - DoReFa-Net: Training Low Bitwidth Convolutional Neural Networks with Low Bitwidth Gradients | [2016/06]
  - Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation | [2013/08]

**Knowledge Distillation** - Apprentice: Using Knowledge Distillation Techniques To Improve Low-Precision Network Accuracy | [**ICLR 2018**]

- Model compression via distillation and quantization | [**ICLR 2018**]

**Acceleration:** - Fast Algorithms for Convolutional Neural Networks | [**CVPR 2016**] + andravin/wincnn | [Python]

### [Projects]

- NervanaSystems/distiller | [Pytorch]
- Tencent/PocketFlow | [Tensorflow]
- aaron-xichen/pytorch-playground | [Pytorch]

### [Tutorials/Blogs]

- Introducing the CVPR 2018 On-Device Visual Intelligence Challenge

## 4.) Hyperparameter Optimization

### [Papers]

- Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly | [2019/03]
  - dragonfly/dragonfly

- 
- Efficient High Dimensional Bayesian Optimization with Additivity and Quadrature Fourier Features | [**NeurIPS 2018**]
  - Google vizier: A service for black-box optimization | [**SIGKDD 2017**]
  - On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice | [**Neurocomputing 2020**]
    - LiYangHart/Hyperparameter-Optimization-of-Machine-Learning-Algorithms

### [Projects]

- BoTorch | [PyTorch]
- Ax (Adaptive Experimentation Platform) | [PyTorch]
- Microsoft/nni | [Python]
- dragonfly/dragonfly | [Python]
- LiYangHart/Hyperparameter-Optimization-of-Machine-Learning-Algorithms | [Python]

### [Tutorials/Blogs]

- Hyperparameter tuning in Cloud Machine Learning Engine using Bayesian Optimization
- Overview of Bayesian Optimization
- Bayesian optimization
  - krasserm/bayesian-machine-learning | [Python]

## 5.) Automated Feature Engineering

### Model Analyzer

- Netscope CNN Analyzer | [Caffe]
- sksq96/pytorch-summary | [Pytorch]
- Lyken17/pytorch-OpCounter | [Pytorch]
- sovrarov/flops-counter.pytorch | [Pytorch]

---

## References

- LITERATURE ON NEURAL ARCHITECTURE SEARCH
- handong1587/handong1587.github.io
- hibayesian/awesome-automl-papers
- mrgloom/awesome-semantic-segmentation
- amusi/awesome-object-detection