**Neural Architecture Search: A Survey**

**Abstract**

1. Neural Architecture Search 神经架构搜索
2. error-prone adj. 容易出错的
3. dimension n. 规模、尺度、标准

**\*Q1： Why do we conduct automated neural architecture search in recent years？(BACKGROUND)**

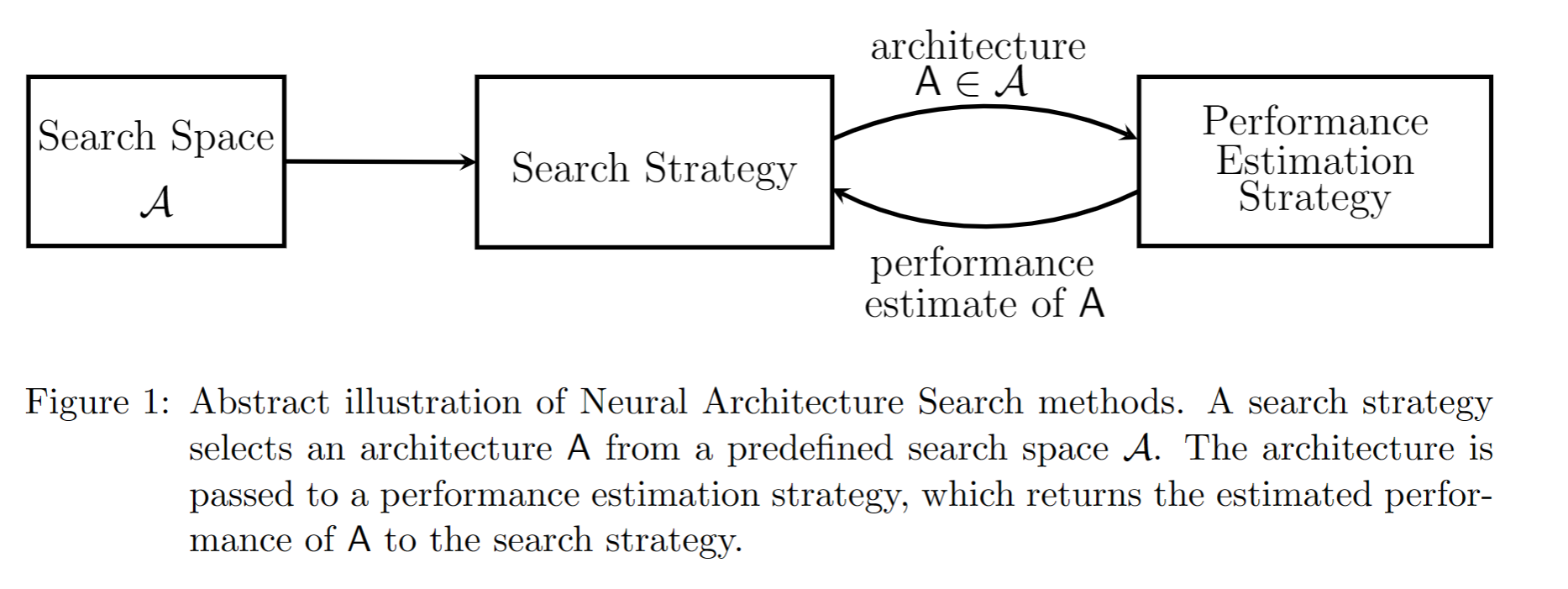
Owing to the situation that DL is mostly developed by experts manually and a rising demand for it，which causes this work time-costing and error-phone. Thus, we need a novel method to cope with this situation.

**Introduction**

1. perceptual adj. 认知的，感知的
2. hierarchical adj. 分阶层的
3. extractor n. 抽取器，提取器
4. image classification (Zoph et al., 2018;) 图像分类
5. object detection (Zoph et al., 2018) 目标检测
6. semantic segmentation (Chen et al., 2018) 语义分割
7. subfield n. 分支
8. categorize v. 分类
9. incorporating n. 结合起来的
10. exponentially adv. 指数级的
11. encompasses v.环绕、围绕
12. convergence n. 汇聚点
13. suboptimal adj. 次优的

**Q2: How can we categorize methods for NAS?**

Search space, search strategy, performance estimation strategy.



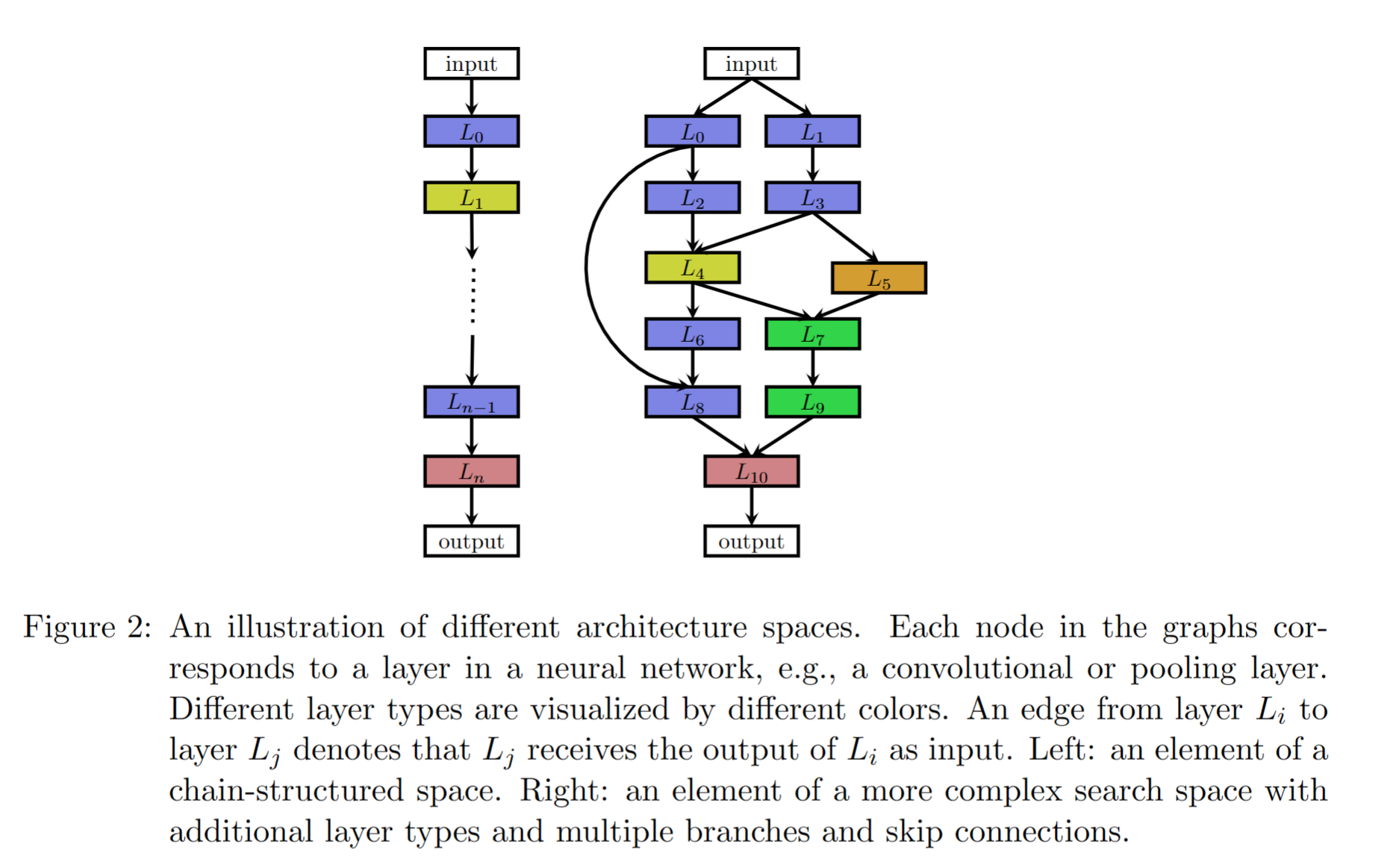
**Q3: The objective of NAS?**

Typically to find architectures that achieve high predictive performance on unseen data.

1. **Search space**
2. visualize v. （使）显现化
3. denote v. 指明，表示，意味着
4. multiple adj. 数量多的，多种多样的
5. layer. n. 图层结构
6. parametrize v. (使)参数化 parametrization n. 参数化
7. execute n. 执行
8. dilate v. 扩张，使膨胀
9. Depth（wise） Separable Convolution n. 深度可分离卷积
10. Dilated convolution 扩张卷积
11. hyperparameter n. 超参数
12. concatenate v. 使串联起来
13. motif n. 主旨，花边
14. dimensionality n. 尺度、维度
15. stack v. 堆积、叠加
16. sequentially adv. 连续地

**WHAT:** Define which neural architectures a NAS approach might discover in principle.

**EXAMPLE:** chain-structured neural network



**Step1:** receive the data from previous layer

**Step2:**  parametrize the space

**Foundation of Step2:**

* the (maximum) number of layers n
* the type of operation every layer executes

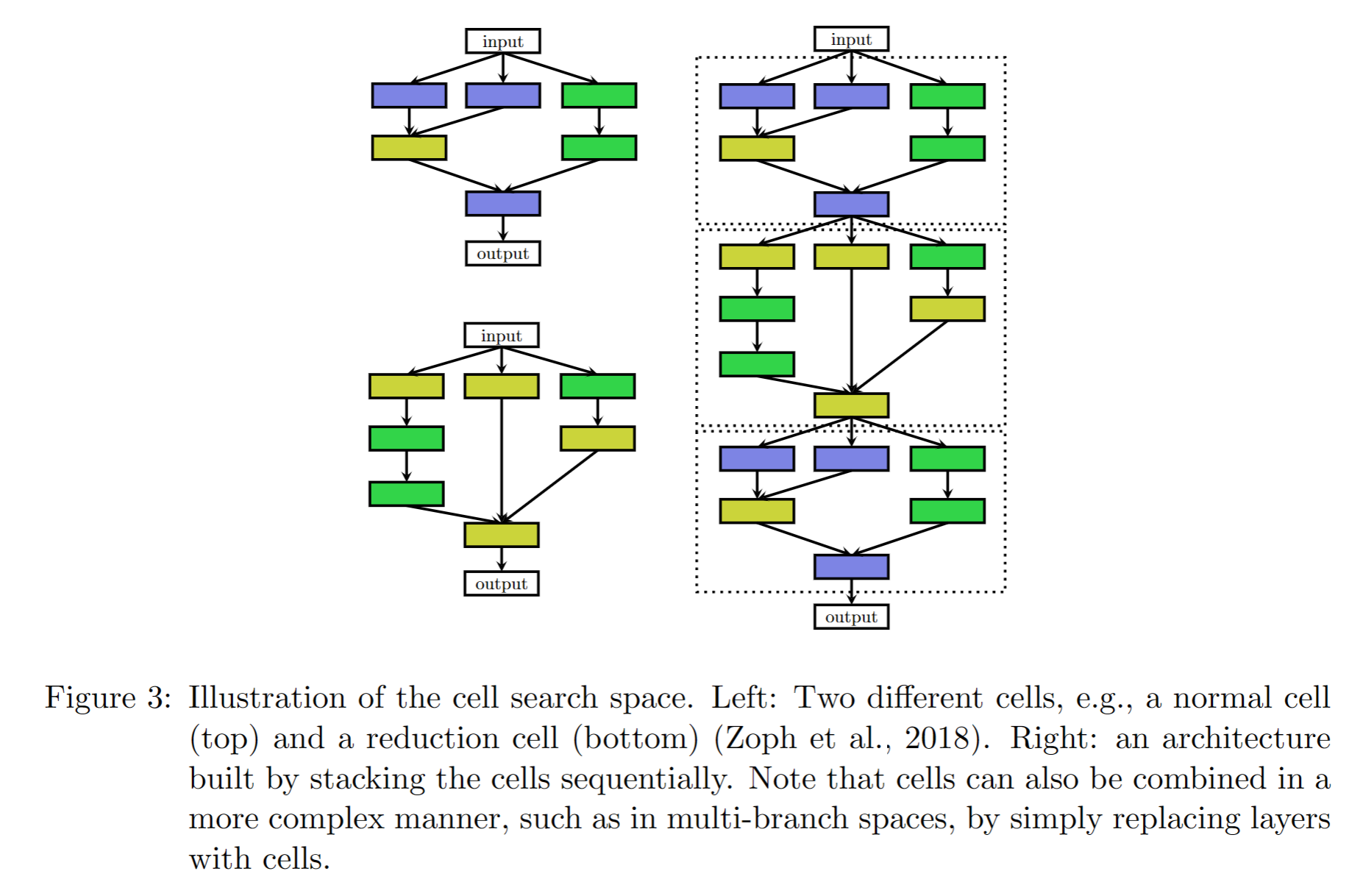
1. pooling、convolution、
2. depth（wise） separable convolution
3. dilated convolution

* hyperparameters associated with the operation, like number of filters

1. kernel size
2. strides for a convolutional layer
3. simply number of units for fully-connected networks

**Optimize:** **cell-based search space**

* a normal cell (preserves the dimensionality of the input)
* a reduction cell (reduces the spatial dimension)



**Three advantages compared above**

* The size is drastically reduced

1. Reason: cells usually consist of significantly less layers than whole architectures

2.Efficiency: almost seven times according to study

* Easily transfer architectures built from cells

1. Way: vary the number of cells and filters used within a model

* A useful design principle in general

**The choice of the search space largely determines the difficulty of the optimization problem.**

1. **Search strategy**

Finding a neural architecture that maximizes some performance measure

1. macro-architecture 宏观架构
2. micro-architecture 微观架构
3. solely adv. 仅仅，独自地
4. primitive adj. 原始的，远古的
5. a directed acyclic graph 有向无环图
6. sequential model 顺序式模型
7. finite adj. 有限的，有限制的
8. augmentation n. 扩张
9. sample v. 抽样，取样
10. intuitive adj. 直观的
11. interpret v. 说明，诠释
12. supervised adj. 有监督的
13. unsupervised adj. 无监督的
14. inheritance n. 继承性
15. derive v. 导出，推导
16. discrete adj. 离散的，分离的

* **Types of strategies**

1. random search （随机搜索）
2. Bayesian optimization （贝叶斯优化）
3. evolutionary method (be used to evolve neural architectures and weights) （进化计算）
4. reinforcement learning (RL) （强化学习）
5. gradient-based methods （梯度基方法）
6. **sequential decision processes （顺序决策进程）**

**Q4: How can NAS became a mainstream research topic?**

Bayesian optimization leads to several methods to win on competition data sets against human experts.

1. state-of-the-art vision architectures (最先进的视觉架构)
2. state-of-the-art performance for CIFAR-10 without data augmentation（CIFAR-10在没有数据增强的情况下的最先进的性能）
3. the first automatically tuned neural networks（第一个自动调整的神经网络）

* **Different RL approach differ in how they represent the agent’s policy and how they optimize it.**

Frame NAS as a reinforcement learning problem

* **Methods:**
* sequential decision processes
* neuro-evolutionary approaches

(From evolutional algorithms to genetic algorithms)

**SGD-based weight optimization methods** currently outperform **evolutionary algorithm** when scaling to contemporary neural architectures with millions of weights for supervised learning tasks.

* optimizing weights using gradient-based methods
* optimize neural architecture using evolutionary algorithms
* Bayesian Optimization

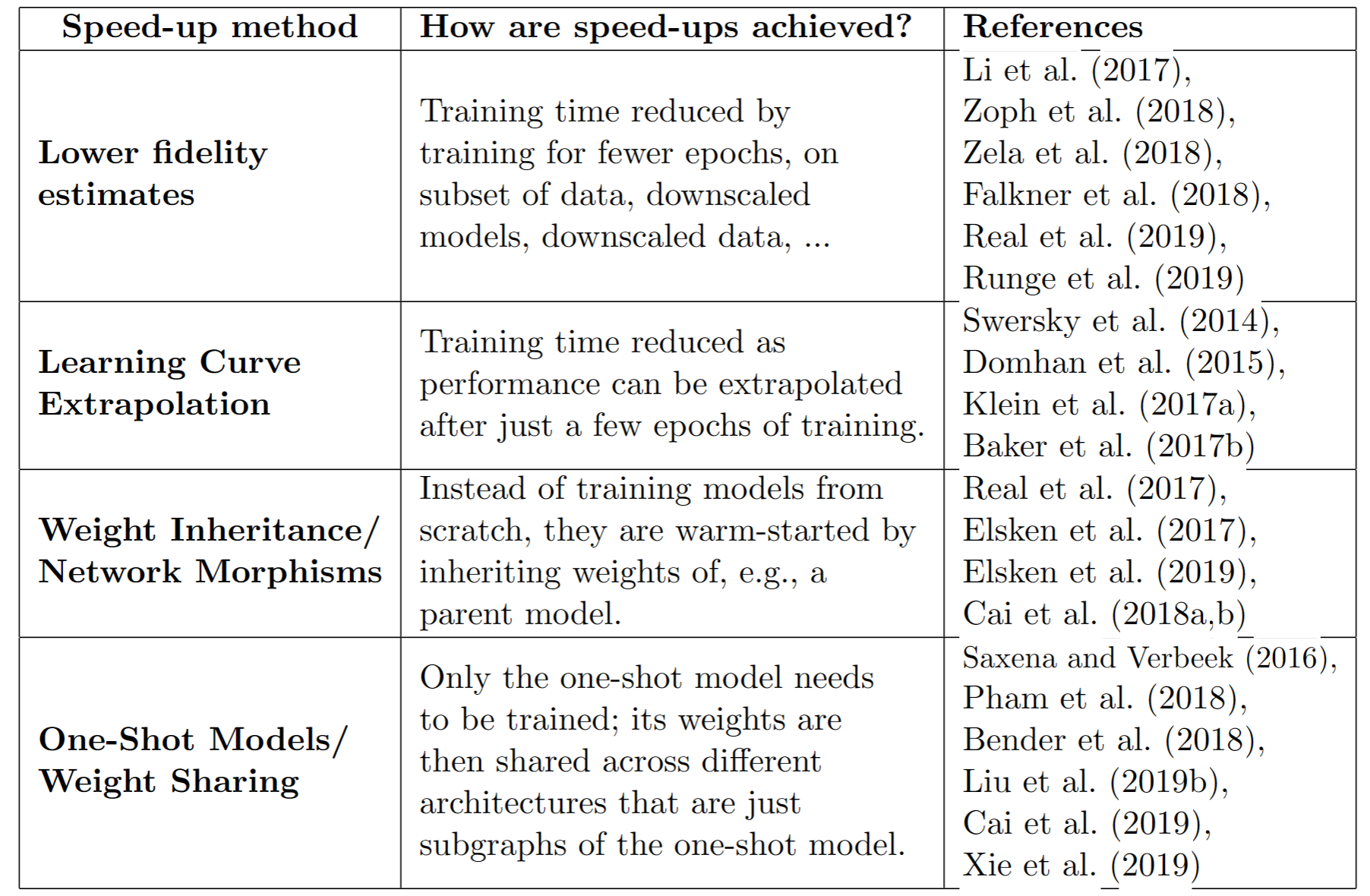
one of the most popular methods for hyperparameter optimization

**continuous relaxation**

Enable direct gradient-based optimization

1. **Performance Estimation Strategy**
2. extrapolate v. 推断，推测
3. surrogate v. 代理，代替
4. attenuate v. 减弱，衰减
5. sufficient adj. 足够的，充足的
6. reparametrization n.参数重置化
7. deactivate v. 去活化，解散
8. stochastically adv. 推测地，随机地
9. reinforce v. 加固，加强

Developing methods for speeding up performance estimation



* **Lower fidelity estimates （低保真估值）**
* Advantage：

1. reduce the computational cost
2. introduce bias in the estimate as performance

* Disadvantage：

relative ranking can change dramatically when the  
difference between the cheap approximations and the “full” evaluation is too big.

* **Learning curve extrapolation (学习曲线外推)**
* Advantage：

1. extrapolate initial learning curves
2. terminate those predicted to perform poorly

* Disadvantage：

good predictions in a relatively large search space need to be made based on relatively few evaluations

* **Weight Inheritance/Network Morphisms**

**（权重继承/网络态射）**

* Advantage：

allow search spaces without an inherent upper bound on the architecture’s size

* Disadvantage：

1. make architectures larger
2. may lead to overly complex architectures.

* **One-Shot Models/Weight Sharing**

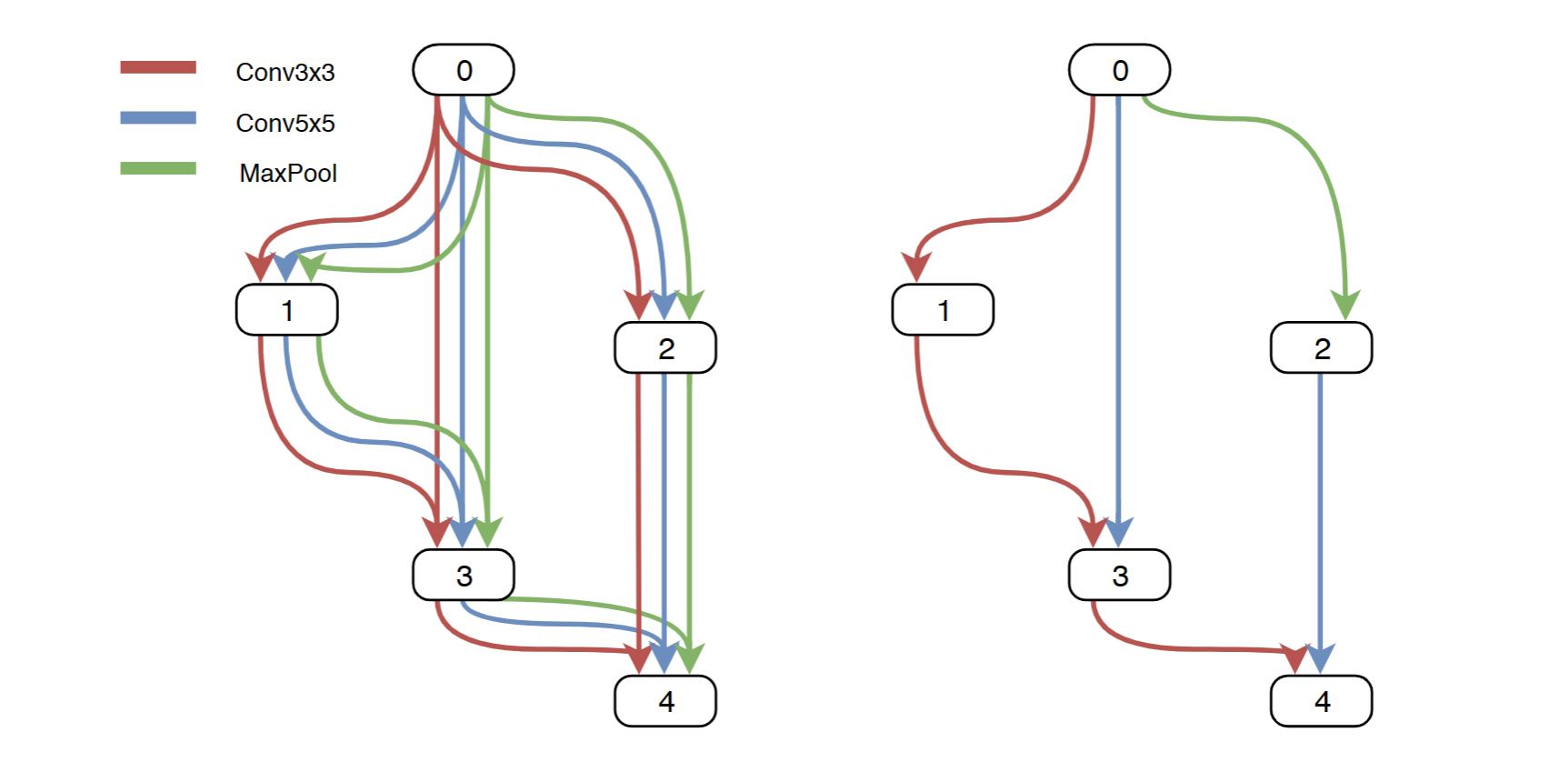
**(单重模型/权重分享)**

* Advantage：

1. Only the one-shot model needs to be trained

2.weights are shared across different architectures

* Disadvantage：  
  typically incurs a large bias
* EXPECTED：  
  a more systematic analysis of biases introduced by different performance estimators



**4.Future Direction**

1. substantially adv. 基本上，大体上，非常的
2. considerably adv. 大幅度地，相当大地，极大的
3. fundamentally adv. 根本上地，基本地，基础地
4. semantic segmentation 语义分割
5. image restoration 图像复原
6. transfer learning 迁移式学习
7. machine translation 机器翻译
8. reinforcement learning 强化学习
9. optimizing recurrent neural networks 优化递归神经网络
10. generative adversarial networks 生成对抗网络
11. network compression 网路压缩
12. complicated adj. 复杂的

**Most existing work：**

* **image classification**

Limitation：The found architectures can’t differ from the exiting methods fundamentally. (outperform existing ones considerably impossibly)

**Possible future work：**

* **generative adversarial networks or sensor-fusion**
* **multi-task or multi-objective problems**
* **more robust to adversarial examples**
* **defining more general and flexible search spaces**
* highlight muti-objective problems

Reason: aim at finding well-performing but efficient architectures (related to network compression)

* **search space（more general hierarchical structure）**