NAS needs to utilize proxy tasks due to its high cost:
- CIFAR-10 -> ImageNet
- Small arch space -> large arch space
- Fewer epochs training -> full training

**Limitations**
- Suboptimal for the target task
- Blocks are forced to share the same structure

**Goal:** Directly learn architectures on the large-scale target task while allowing all blocks to have different structures

We achieve this by reducing the cost of NAS (GPU hours and GPU memory) to the same level of normal training.

**GPU hour-wise:** Simplify NAS to be a single training process of a cumbersome network
1. Build the cumbersome network with all candidate paths
2. Use architecture parameters to identify and prune redundant paths (path-level pruning)

**GPU memory-wise:** Binarize architecture parameters and allow only one path of activation to be active in memory at run-time. Learn binarized architecture parameters via
1. Modified gradient decent based on BinaryConnect
2. REINFORCE-based algorithm for non-differentiable objectives (e.g., latency, energy and memory)

**Specialize Network Architectures for Different Platforms**

Hardware prefers specialized models. Proxyless NAS provides an efficient, automated way to design specialized models for different hardware.