Auto-VirtualNet: Cost-Adaptive Dynamic Architecture Search for Multi-Task Learning

ICROS 2021 workshop

June. 24.

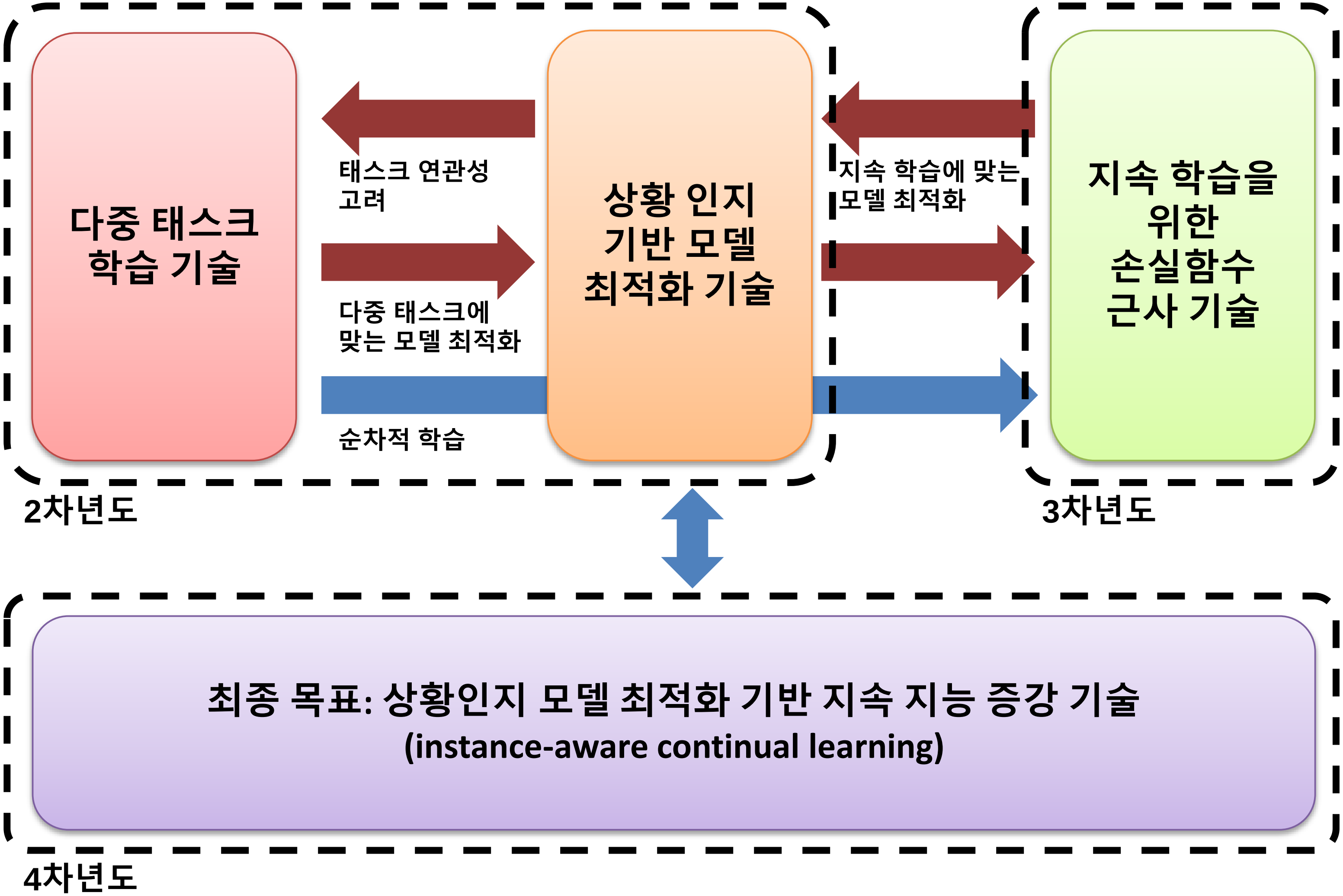
Chanho Ahn

DEPARTMENT OF ELECTRICAL ENGINEERING
SEOUL NATIONAL UNIVERSITY

RLLAB

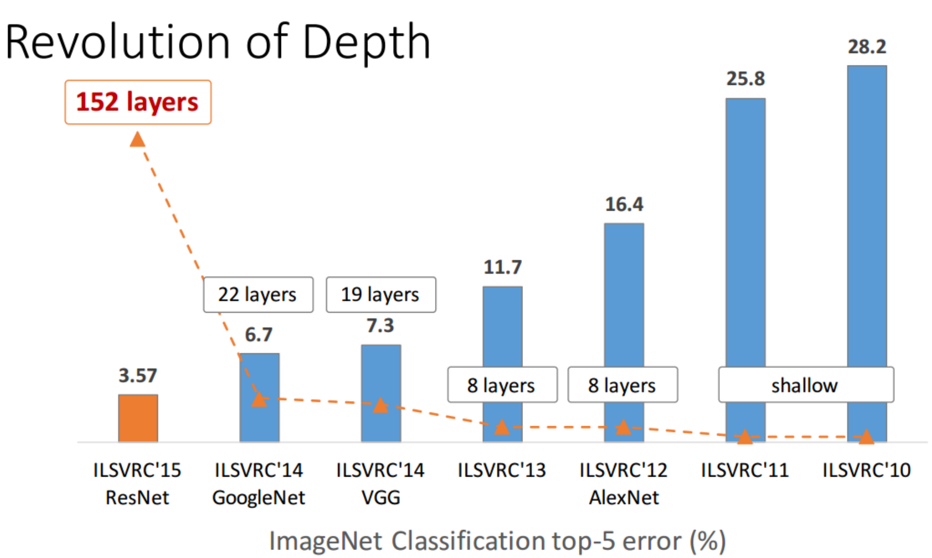
연구 개발 내용

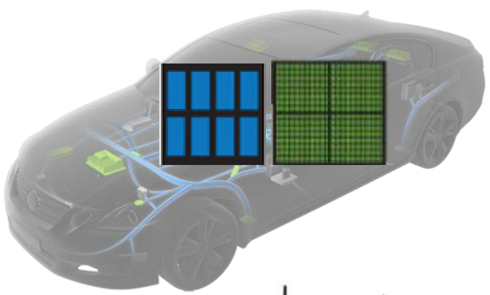
RLLAB



연구의 필요성

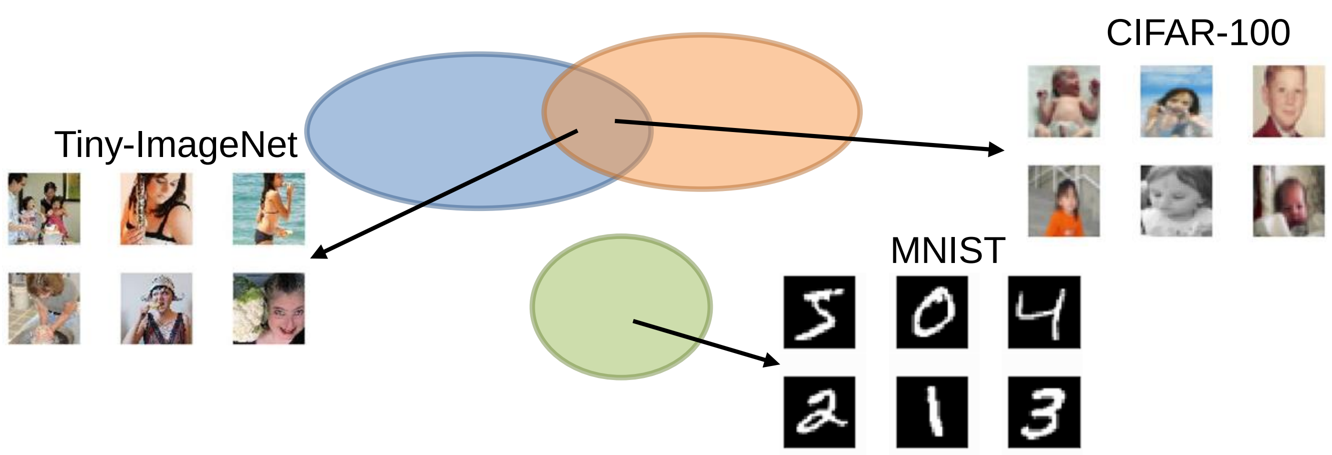
- Recent studies using DNN increase **network depth** to improve performance: classification, object detection, …
- These networks are not tractable to apply to **mobile robots** or **embedded devices** with **small memory capacitance**





연구의 필요성

- Recent studies on multi-task learning (MTL) **simultaneously learns** **multiple tasks**
- For **resource efficiency**, MTL approach **in a single architecture** (by sharing parameters)
- Performance decreases when **less related tasks** are trained jointly in **a single architecture**



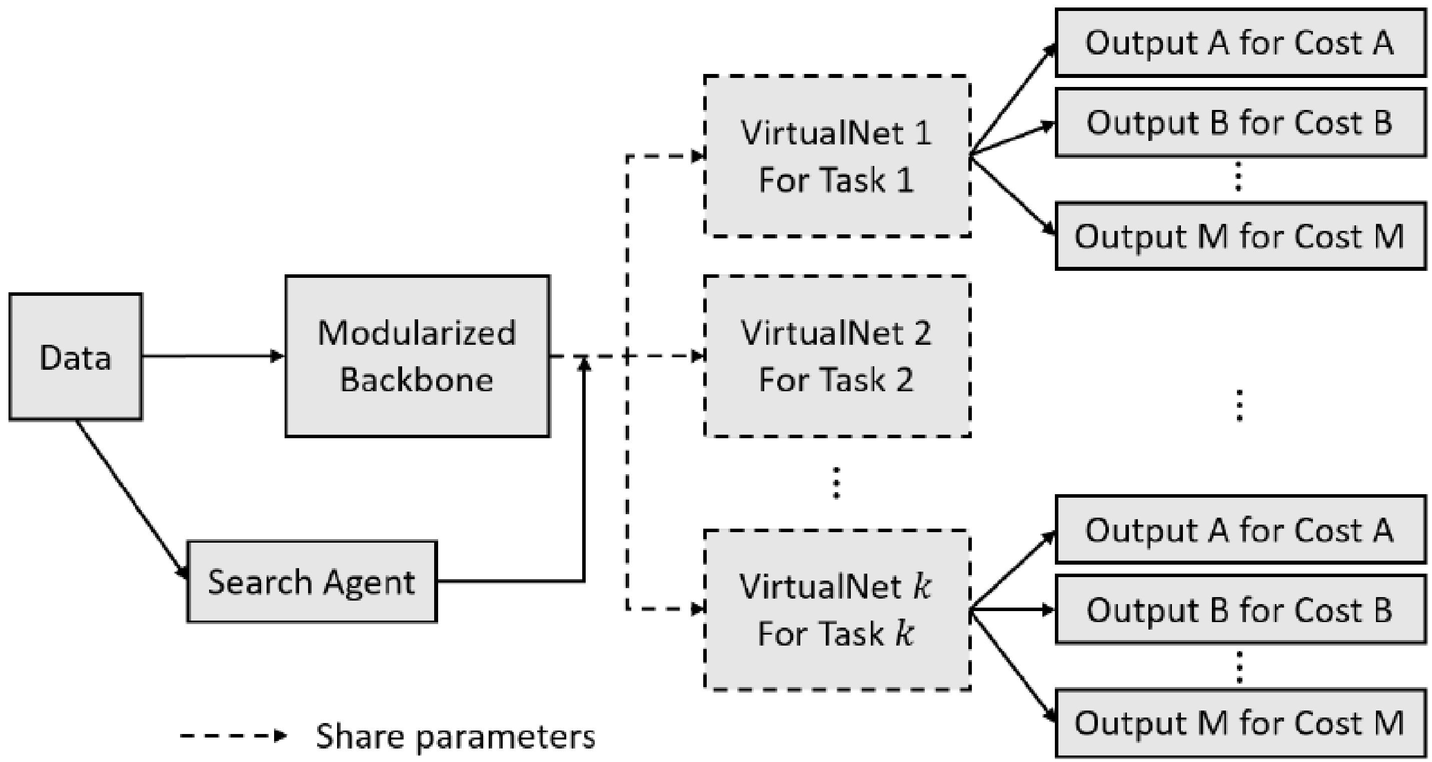
연구의 목표

- In summary, we want to solve three problems:
  - **Resource efficiency** for saving parameters
  - **Memory efficiency** for inference
  - **Adaptive parameter sharing** for performance improvement

연구의 개요

RLLAB

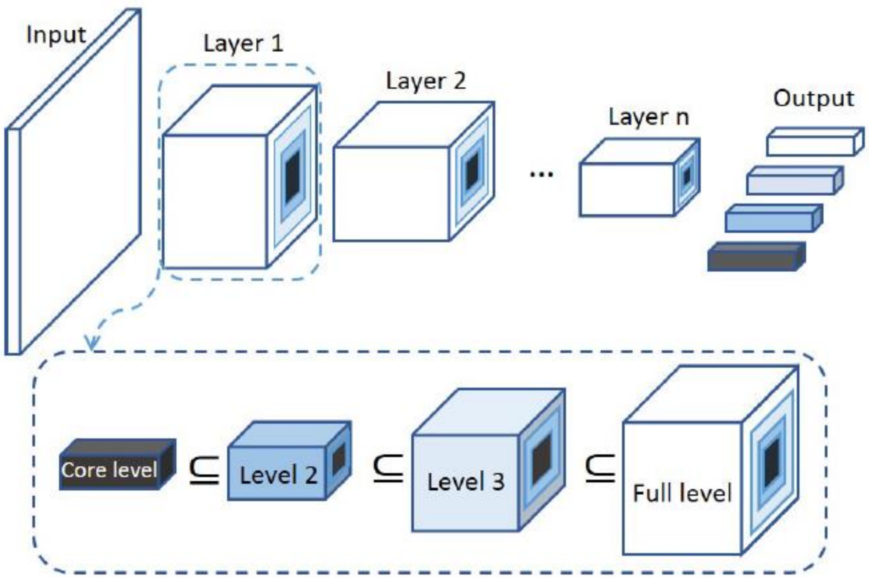
- We propose a single network which **shares parameters** between multiple tasks and provides **multiple inference path** depending on users' memory budget



Methodology

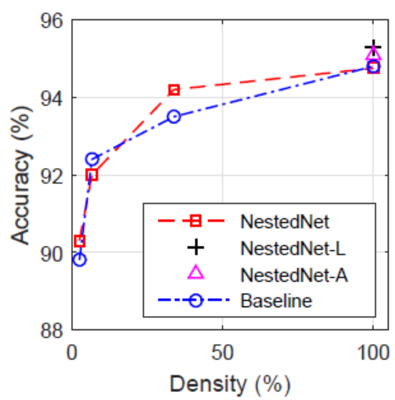
Cost-adaptive learning

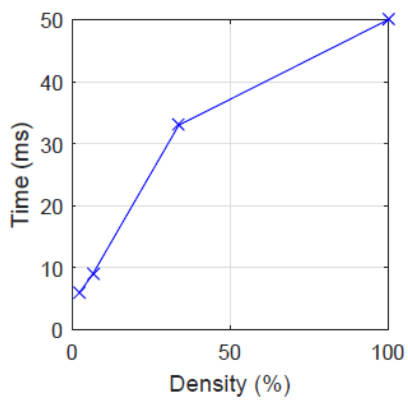
- Single task with **multiple inference paths**
- Construct a network-in-network **hierarchical** structure
- Split network parameters into multiple disjoint subsets
- Each inference path only includes some subsets

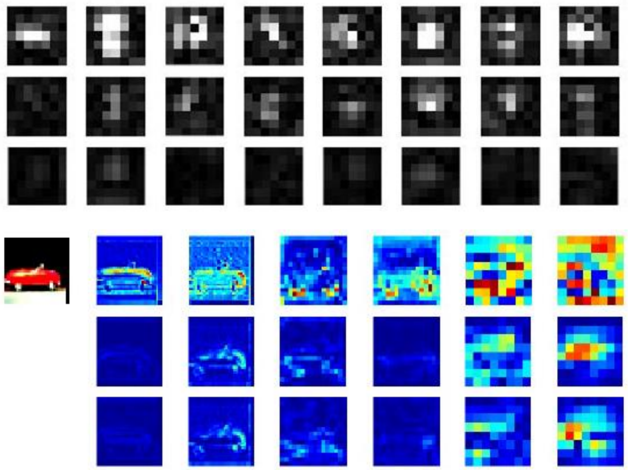


Cost-adaptive learning

- The hierarchical structure allows multiple inference paths for different computational requirements
- Intuitively, it can provide a **strong trade-off** between performance and computational budget

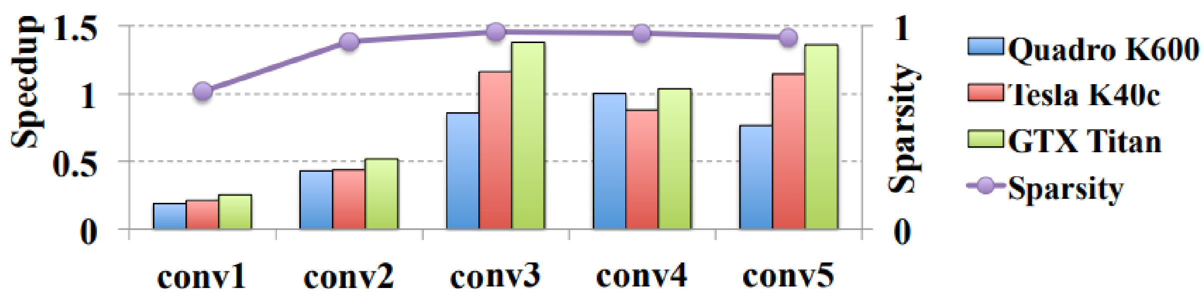


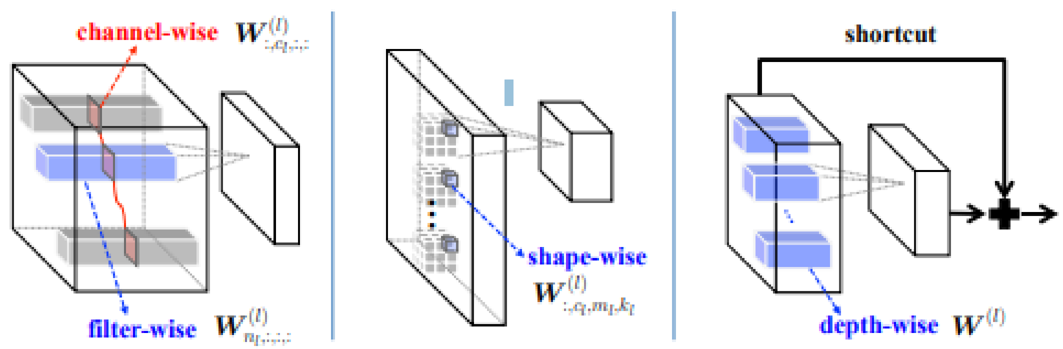




Cost-adaptive learning

- Want to achieve **actual memory reduction** for inference
- Parameters are split in a **structural way**
- We split the parameters with respect to the direction of **channels** and **depths**





Cost-adaptive learning

RLLAB

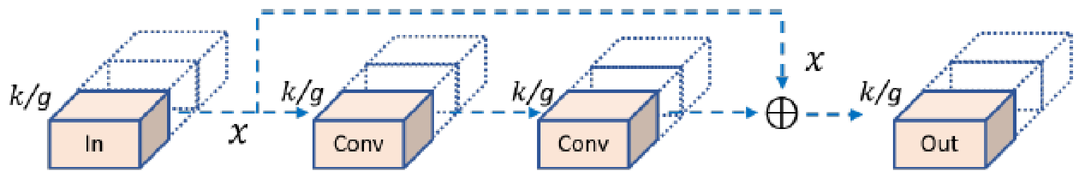
- Loss for multiple inference path is defined as:

$$\min_{\mathcal{W}} \sum_{i=1}^{n_h} \mathcal{L}\left(h^i(\mathcal{W}); \mathcal{D}\right),$$

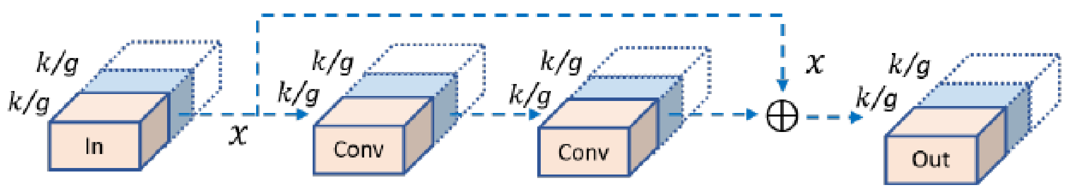
$$h^l(\mathcal{W}) \subseteq h^m(\mathcal{W}),$$

$$l \leq m, \ \forall l, m \in [1, ..., n_h],$$

- Loss is the sum of losses for each path



(a) First-level of hierarchy



(b) Second-level of hierarchy

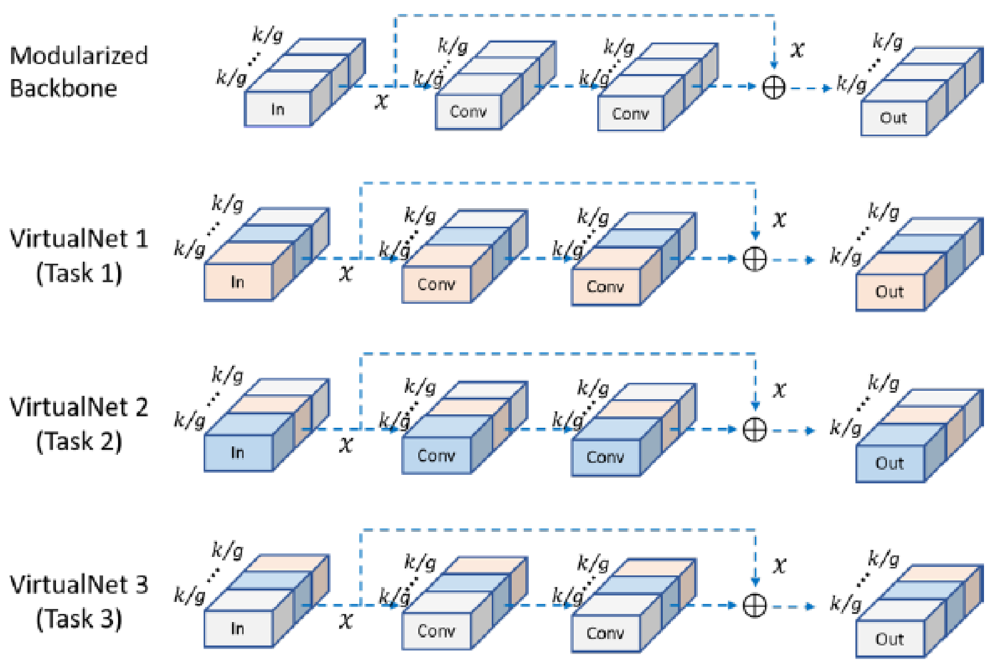
Model reshuffling

- **Destructive interference**: destroying efficiency in MTL, when tasks are of limited relevance to one another
- The destructive interference was first introduced in tasks which identify the given attributes
- The results of jointly learning show poor performance than independent learning

	smile Acc.	open mouth Acc.	young Acc.	smile / young UCR	smile /open-mouth UCR
smile + young + open mouth(a)	84.71%	74.73 %	71.6%	-	-
smile + young(b)	83.85%	-	74.71%	22.1%	-
smile + open mouth(c)	91.72%	92.65%	-	-	43.71%
Three Independent Networks(d)	93.32%	94.40%	84.90%	-	-
With Proposed Modulation(e)	94.03%	95.31%	86.20%	**50.63%**	**52.77%**
With Proposed Modulation + Reg(f)	**94.94%**	**95.58%**	**87.75%**	-	-

Model reshuffling

- Different network configurations and inference flows for different tasks
- To reduce potential negative interference arising when multiple tasks share important filters



Model reshuffling

- The overall loss function is sum of loss for all tasks and all hierarchies:

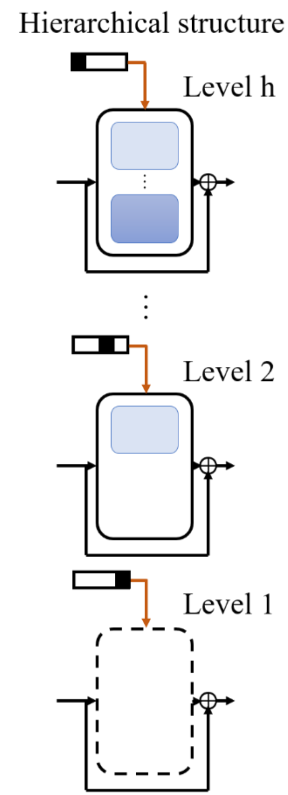
$$\min_{\mathcal{W}} \sum_{j=1}^{k} \sum_{i=1}^{n_h} \mathcal{L}\Big(h^{i,j}(\mathcal{W}); \mathcal{D}^j\Big),$$

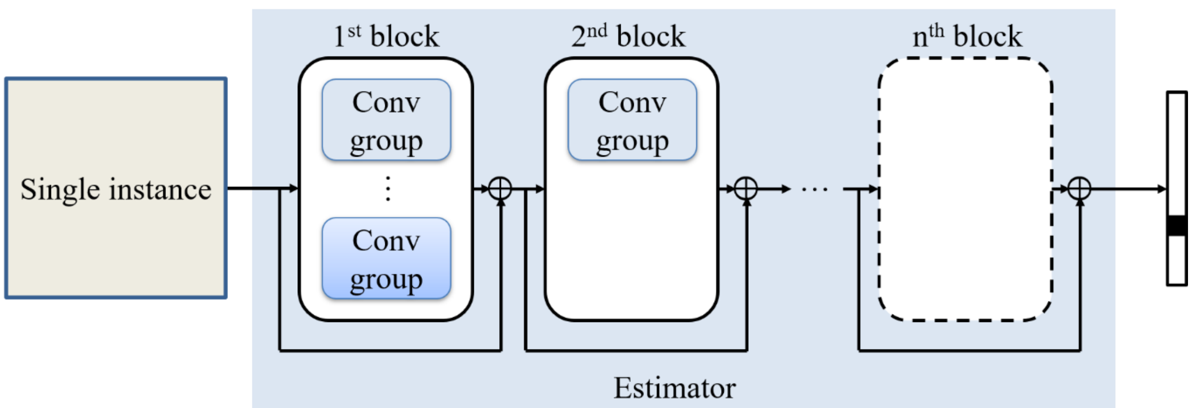
$$h^{l,j}(\mathcal{W}) \subseteq h^{m,j}(\mathcal{W}),$$

$$l \leq m,\ \forall l, m \in [1, ..., n_h] \text{ and } j \in [1, ..., k].$$

Incorporating dynamic search

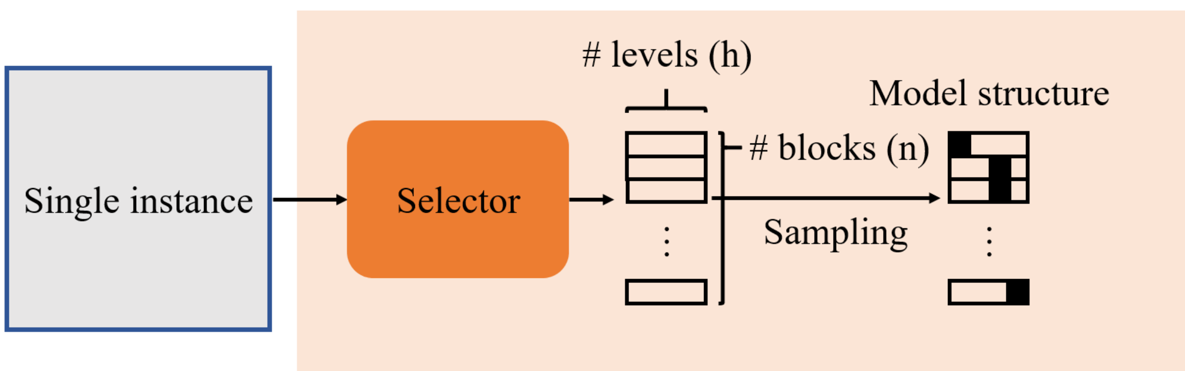
- Instead of selecting a hierarchy from the backbone network, selecting hierarchies for each residual block
- This increases the **diversity** of model structures and the **chance to choose more efficient model** under the same memory budget



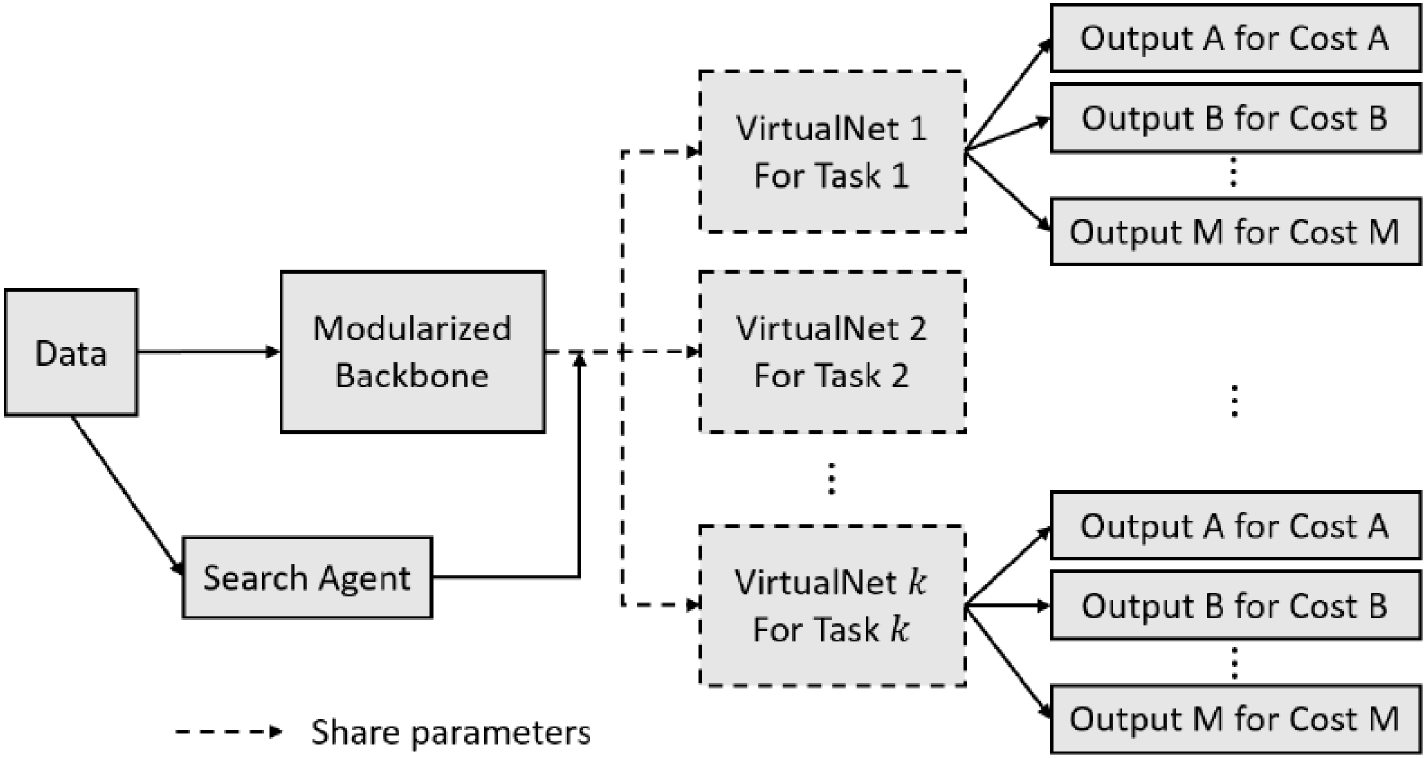


Incorporating dynamic search

- Selecting different models depending on **input instance**
- Instance-wise selection includes selection space of task-wise selection
- Output of the search agent is the selected levels for each block
- The network is trained with policy-gradient method (A: model selection, S: input image, R: loss and density)



요약



Data
Modularized Backbone
Search Agent
VirtualNet 1 For Task 1
VirtualNet 2 For Task 2
VirtualNet k For Task k
Output A for Cost A
Output B for Cost B
Output M for Cost M
Output A for Cost A
Output B for Cost B
Output M for Cost M
Share parameters

요약

- Parameters of backbone network are split into multiple parameter groups
- VirtualNet for each task has different order of hierarchy of parameter groups
- Search agent provides optimal inference paths under various memory budgets
- To train the search agent, policy gradient method is used

Results

Scenario 1

- CIFAR-100 (32x32) & Tiny-ImageNet (64x64) & STL-10 (96x96)

	Task 1		Task 2		Task 3		
	Accuracy	Params	Accuracy	Params	Accuracy	Params	Total Params
Baseline (Single)	72.7%	29.8M	55.8%	29.8M	71.5%	29.8M	89.4M
Baseline (Multi)	58.6%	29.8M	45.5%	29.8M	70.7%	29.8M	29.8M
PackNet [38]	69.6%	7.5M	54.1%	16.7M	73.9%	29.8M	29.8M
NestedNet [41]	71.9%	7.5M	55.5%	16.7M	74.3%	29.8M	29.8M
VirtualNet [19]	73.2%	7.5M	55.5%	7.5M	76.6%	7.5M	29.8M
	74.0%	16.7M	57.9%	16.7M	77.4%	16.7M	
	74.5%	29.8M	58.8%	29.8M	77.9%	29.8M	
Auto-VirtualNet	73.6%	15.2M	58.6%	19.1M	81.6%	10.7M	29.8M
	74.0%	18.2M	58.9%	21.5M	82.0%	12.5M	
	74.2%	21.5M	59.3%	24.7M	82.5%	19.1M	

Scenario 2 & 3

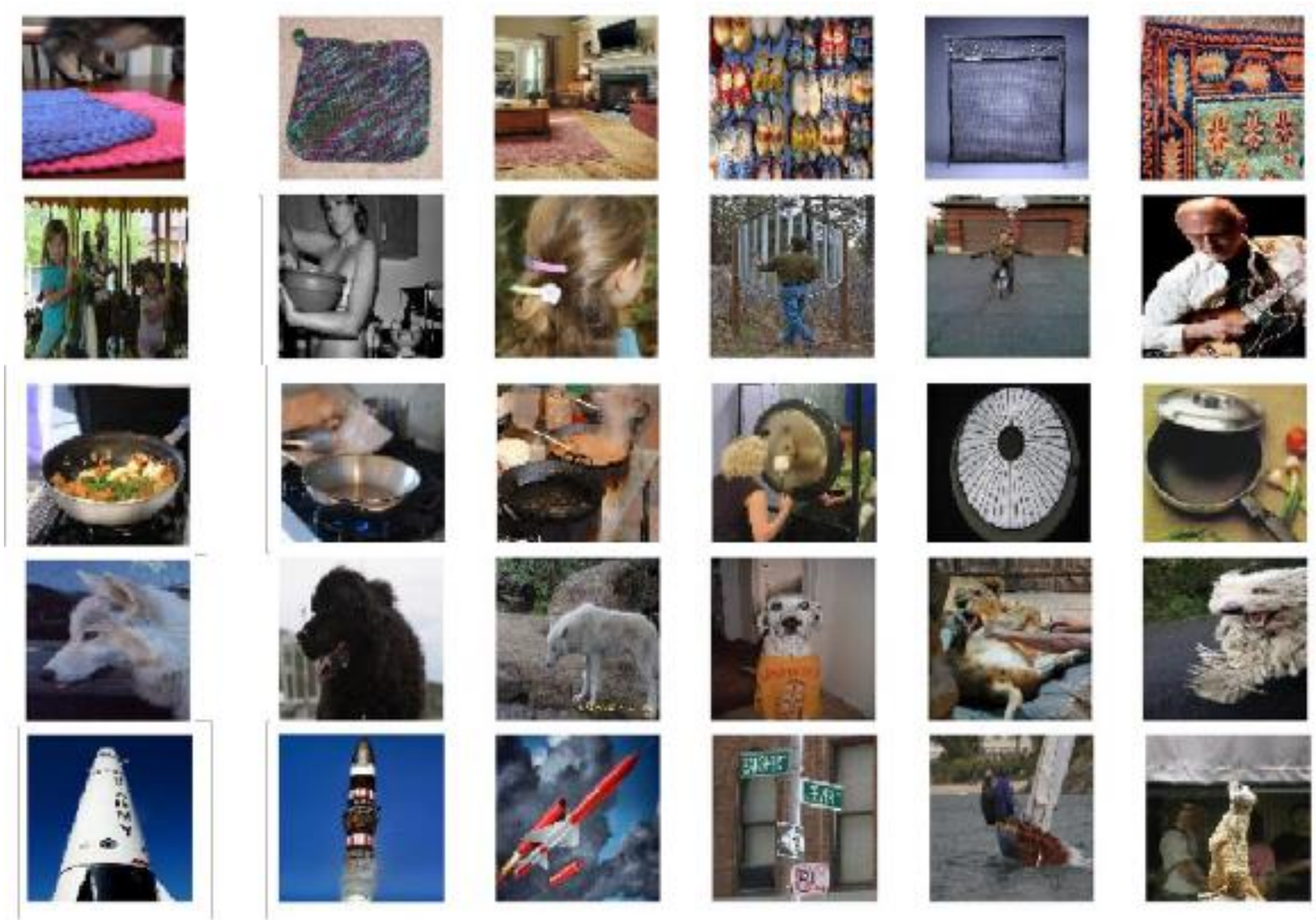
- CIFAR-100 : 20 tasks (5-way classification)

	Accuracy	No. parameters
Baseline (Multi-Task)	42.0%	74K
Cross-Stitch network [30]	54.0%	>1.5M
Routing network [15]	61.0%	>74K
Auto-VirtualNet (Proposed)	87.9%	25K
	88.1%	35K
	88.2%	43K

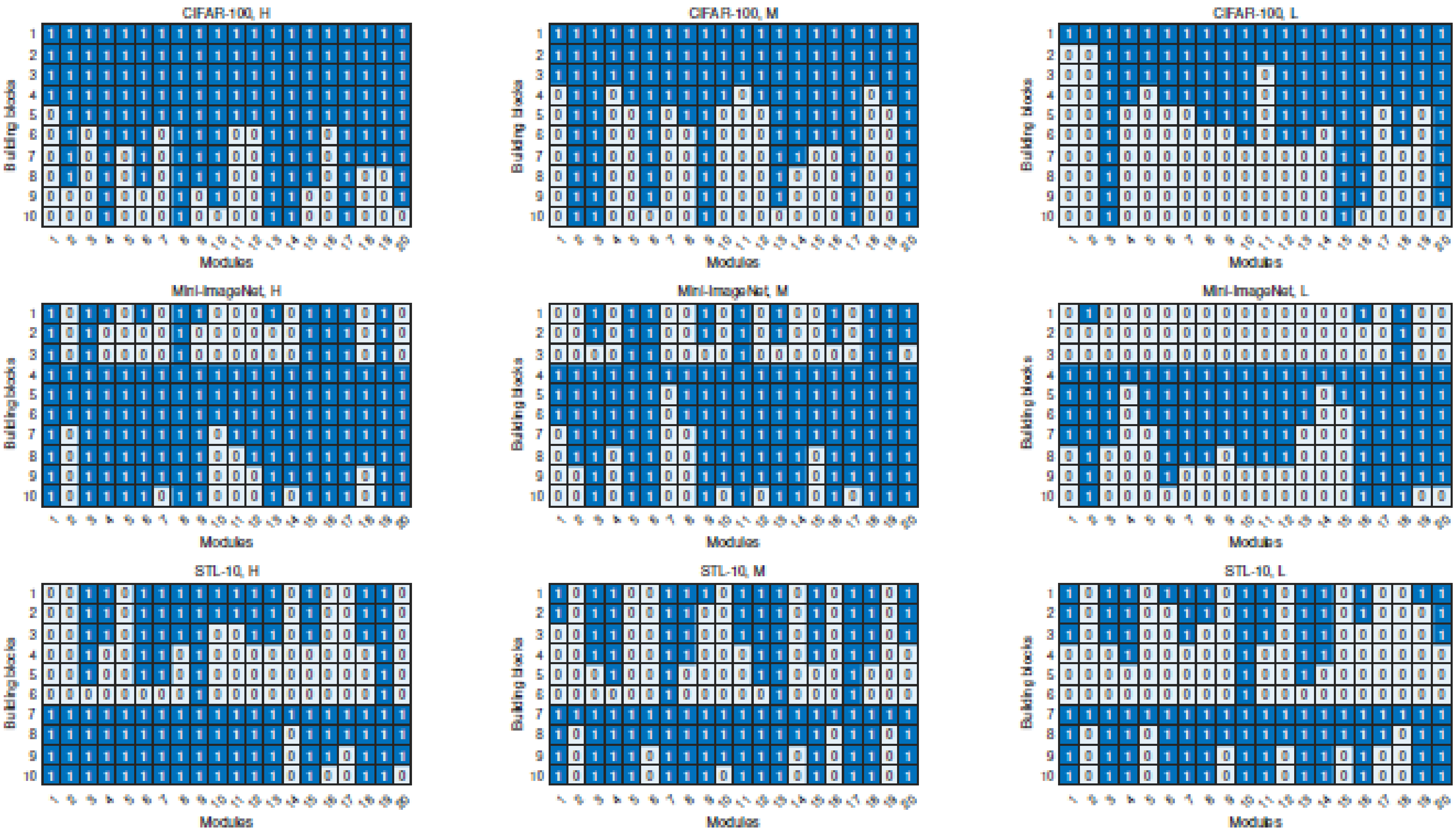
- Mini-ImageNet : 10 tasks (10-way classification)

	Accuracy	No. parameters
Baseline (Multi-Task)	51.0%	140K
Cross-Stitch network [30]	56.0%	>1.4M
Routing network [15]	59.0%	>140K
DEN [17]	62.6%	140K
Auto-VirtualNet (Proposed)	64.9%	59K
	65.2%	82K

Qualitative evaluation



Qualitative evaluation



Q&A