Overcoming Catastrophic Forgetting by augmenting clustering algorithm and Embedding NET

2021. 06. 24.

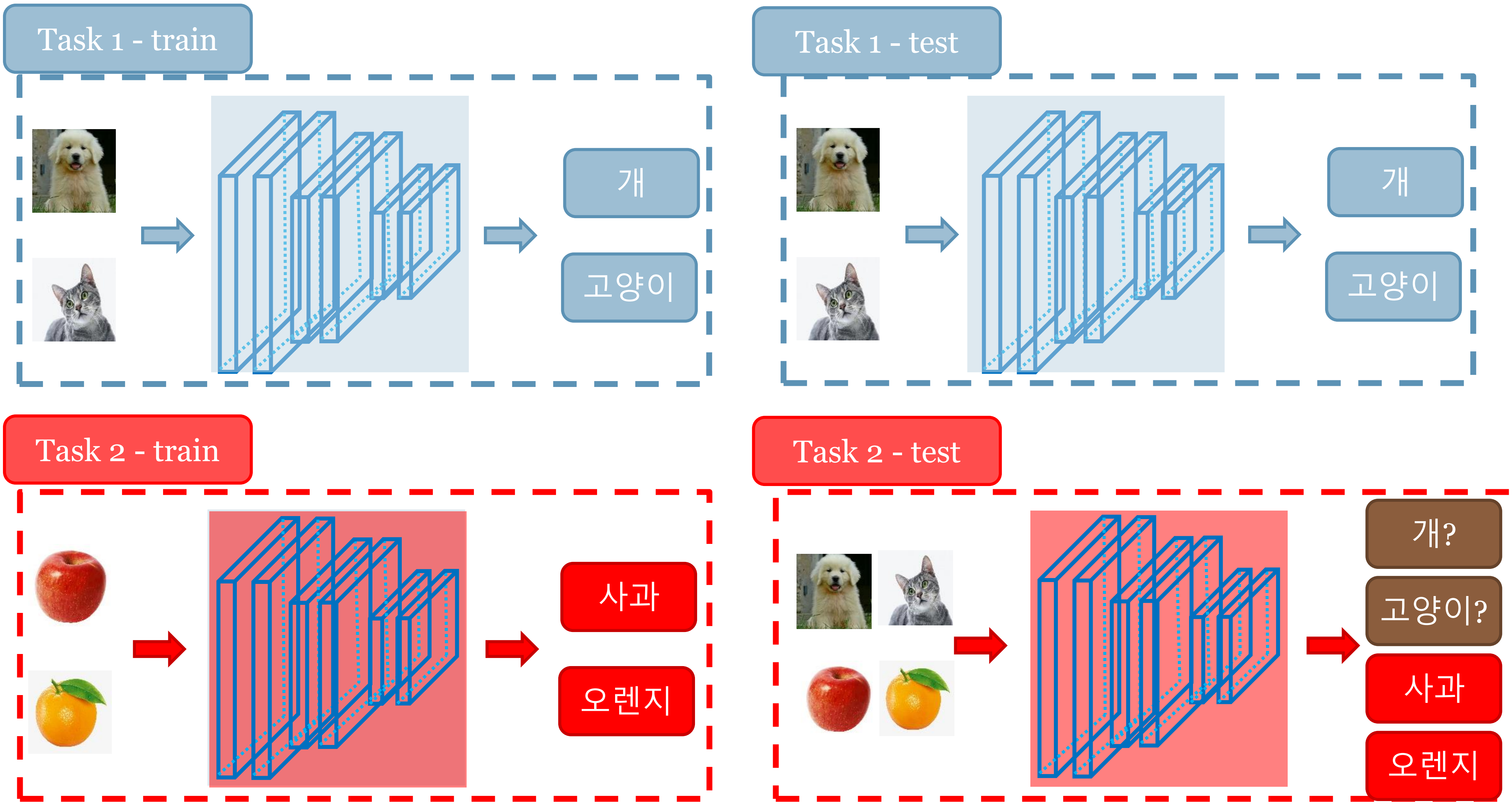
Sungkyunkwan Univ.
Yeongseok Yun

Contents

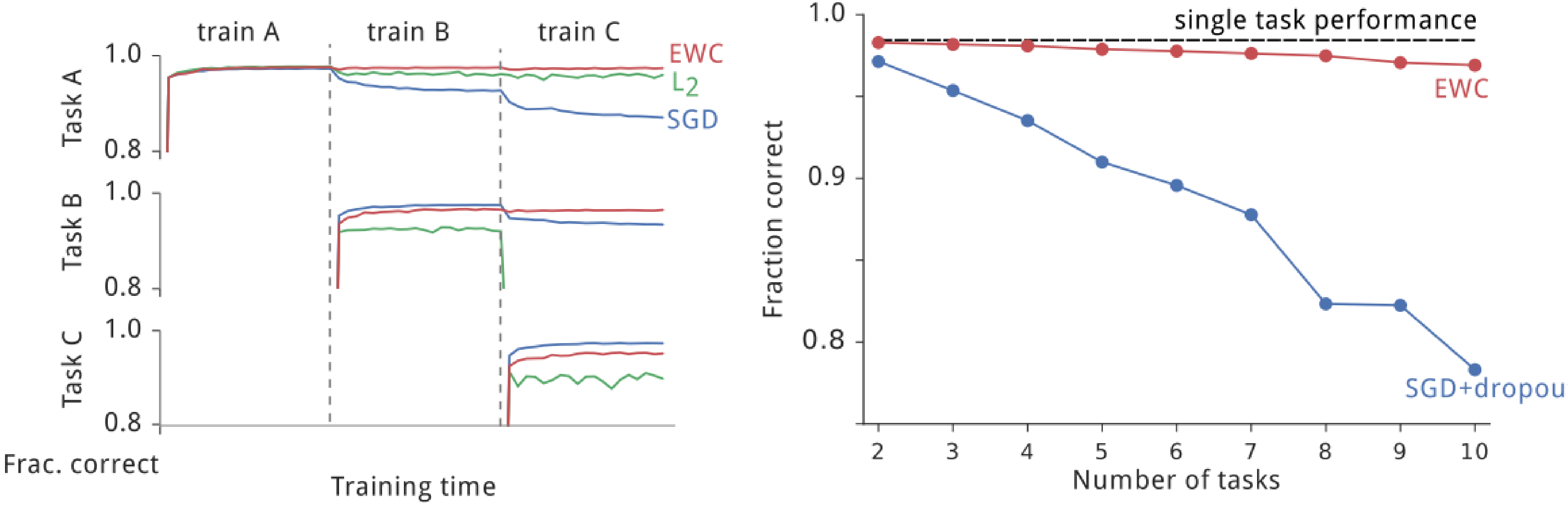
- ✓ Background and Motivation
- ✓ Previous Researches
- ✓ Research Objectives
- ✓ Experiments
- ✓ Future Work

Background and Motivation

Incremental Learning



Catastrophic forgetting



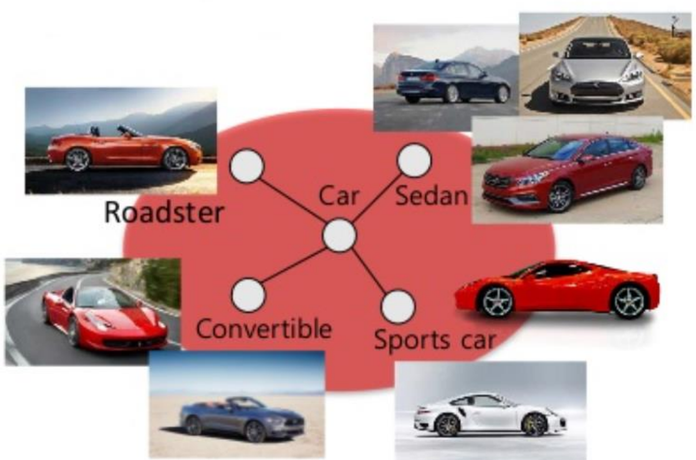
[1]

➢ Accuracy rate for training data is constantly decreasing over time and number of tasks.

Incremental Learning situation

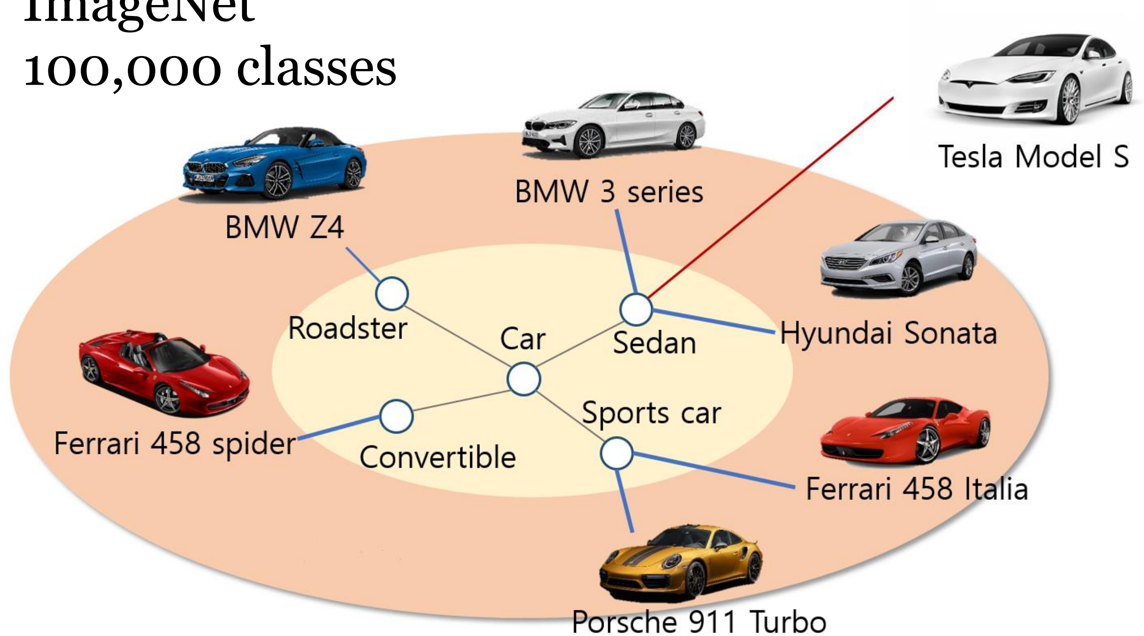
- Data is constantly growing over time.
  - Data/class is subdivided based on research direction or market demand
  - Grant new tasks according to changed data/classes

ImageNet
22,000 classes



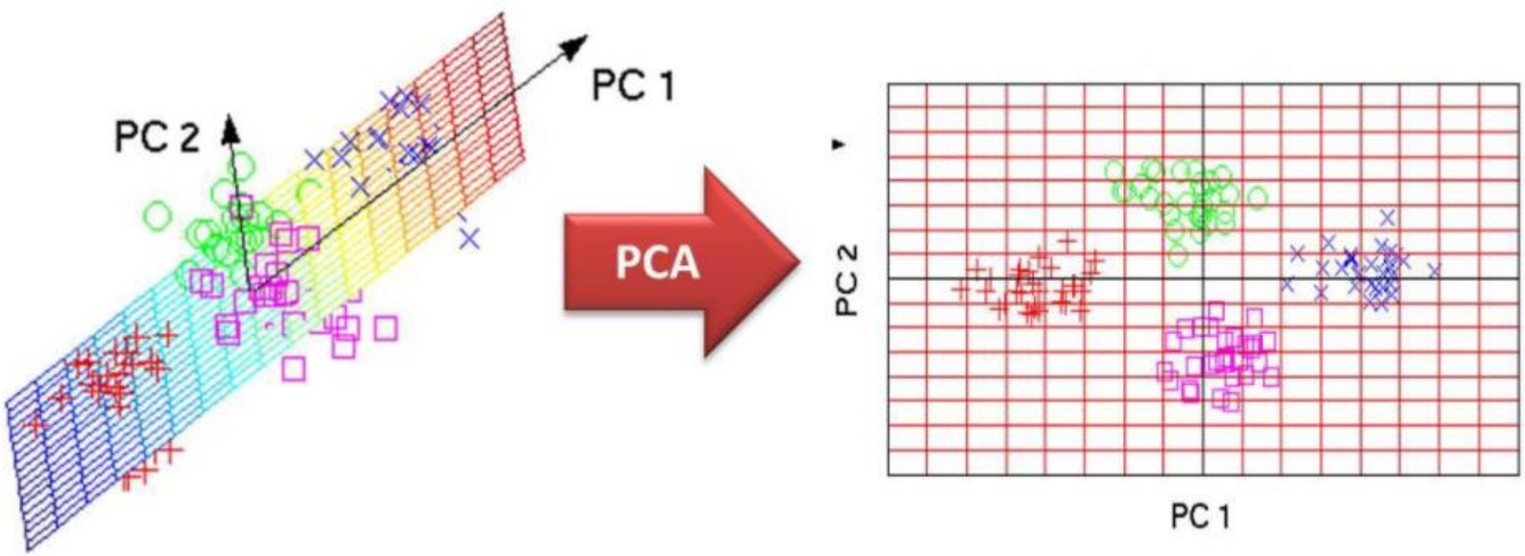
[2]

ImageNet
100,000 classes

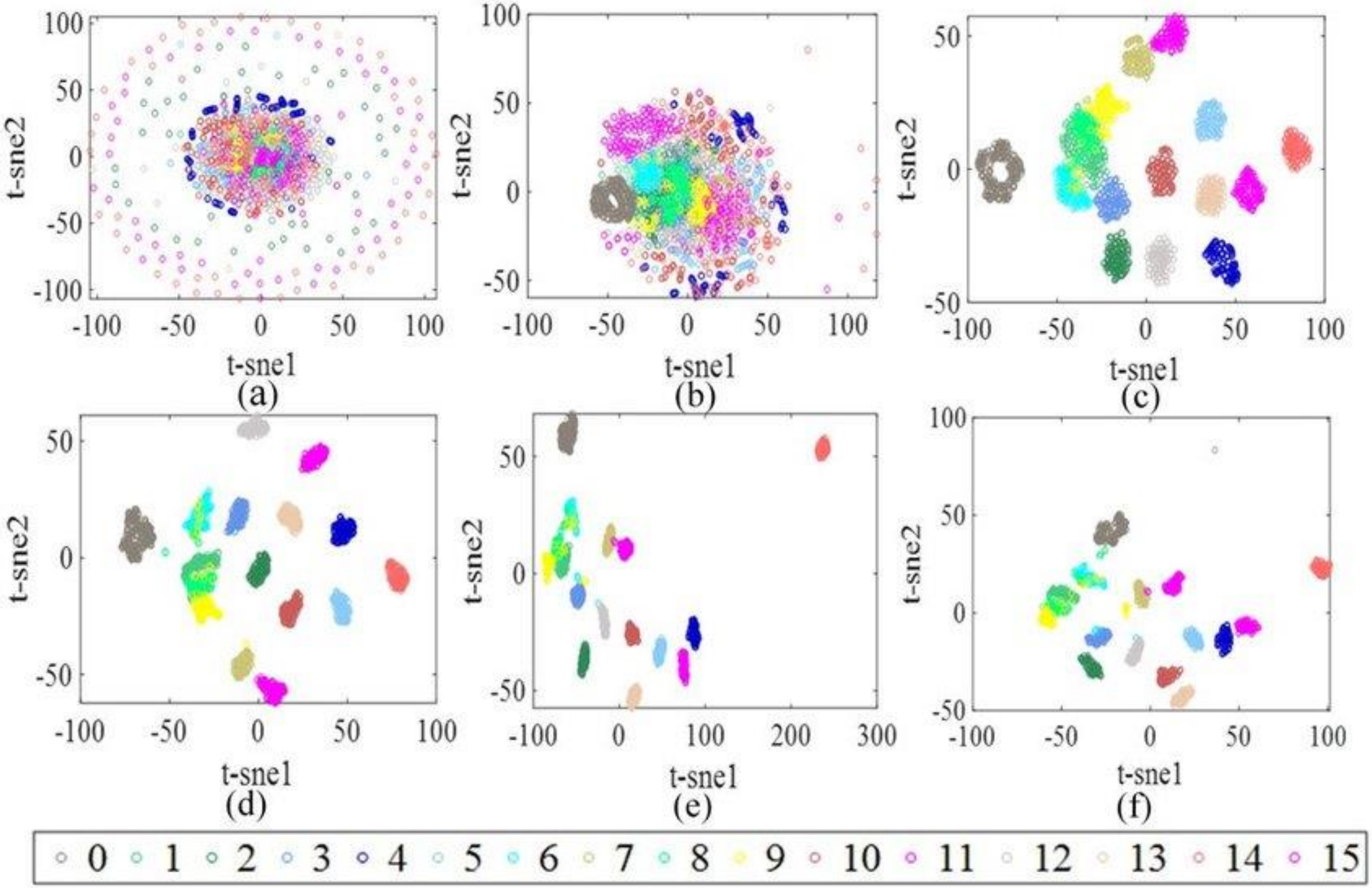


Embedding NET : PCA

- Primal Component Analysis : One of dimensionality reduction technique

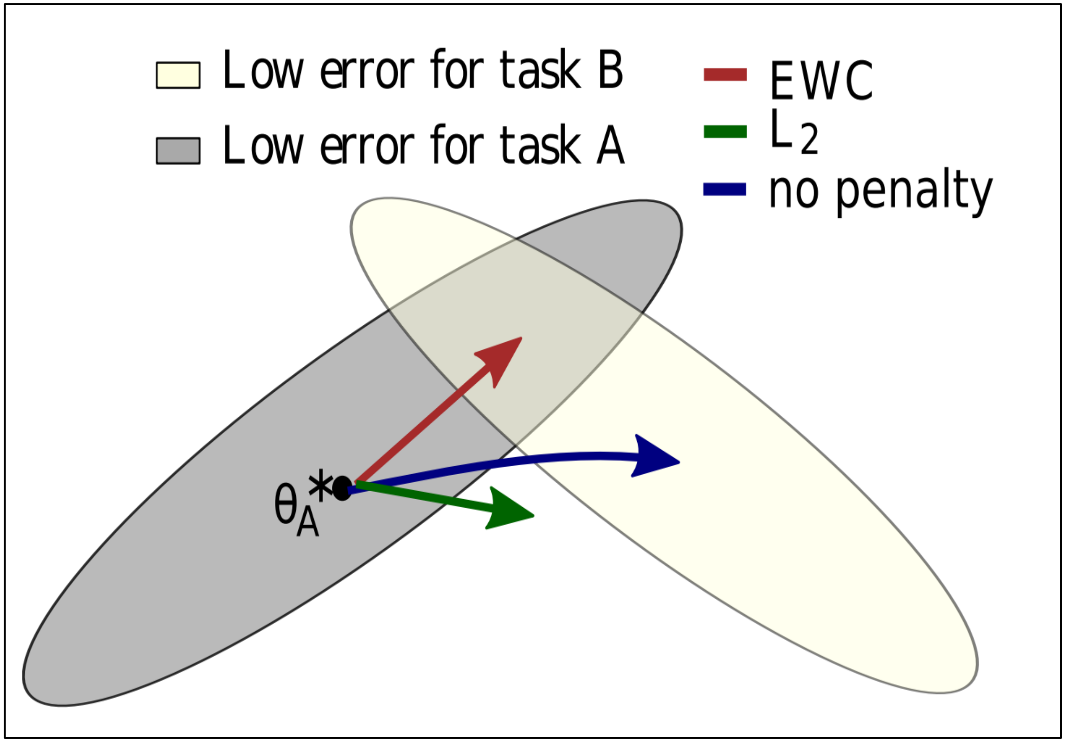


CNN’s Feature distribution during learning

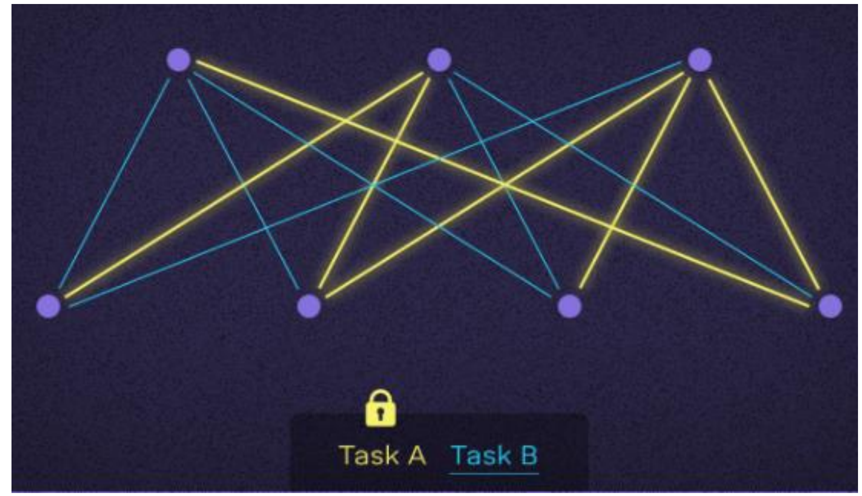


Previous Research

EWC : Parameter regularization



[1]



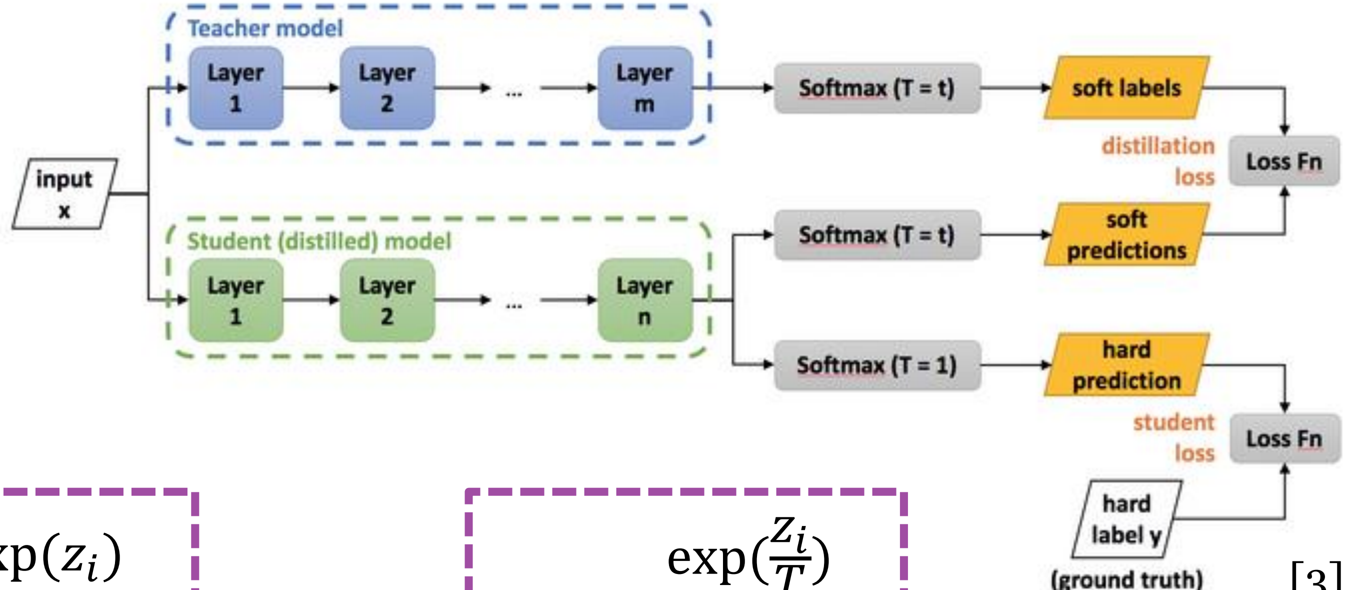
$$L(\theta) = L_B(\theta) + \sum_i \frac{\lambda}{2} F_i (\theta_i - \theta^*_{A,i})^2$$

$$F_i = \frac{1}{N} \sum_i \nabla \log(p_i|\theta) \nabla \log(p_i|\theta)^T$$

- It difficult to control the movement by filtering out exactly desired parameters.
- Performance is worse than ordinary neural network in Last task
  - Change of parameter is constrained

Knowledge Distillation

dog



[3]

$$p_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

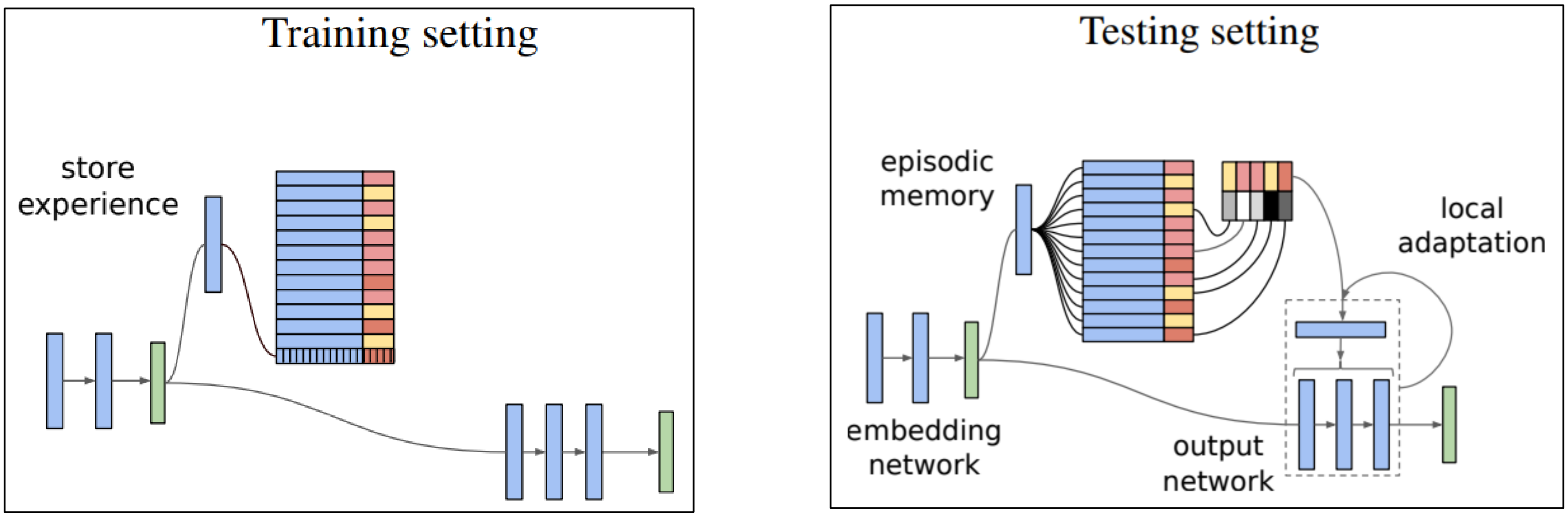
$$\begin{bmatrix} Bear \\ Cat \\ Dog \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$p_i = \frac{\exp(\frac{z_i}{T})}{\sum_j \exp(\frac{z_j}{T})}$$

$$\begin{bmatrix} Bear \\ Cat \\ Dog \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.2 \\ 0.75 \end{bmatrix}$$

- It is difficult to outperform the teacher network's performance.

MBPA - Testing Setting



[4]

- Large memory module to store all training examples
- Vulnerable to negative transfer through local adaptation step
  - Negative Transfer : A phenomenon in which existing knowledge inappropriately influences learning new knowledge.

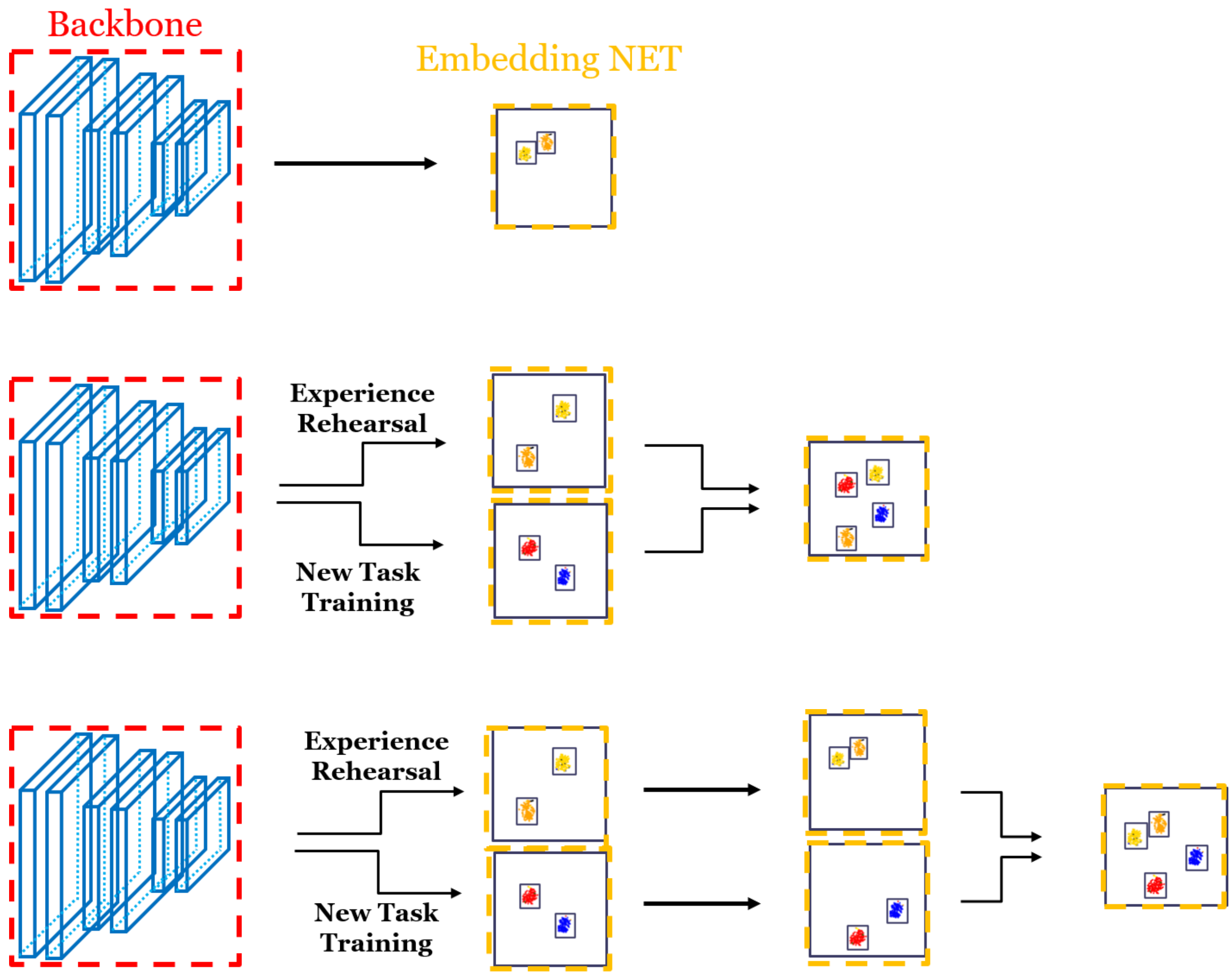
Research Objectives

Problem Statements

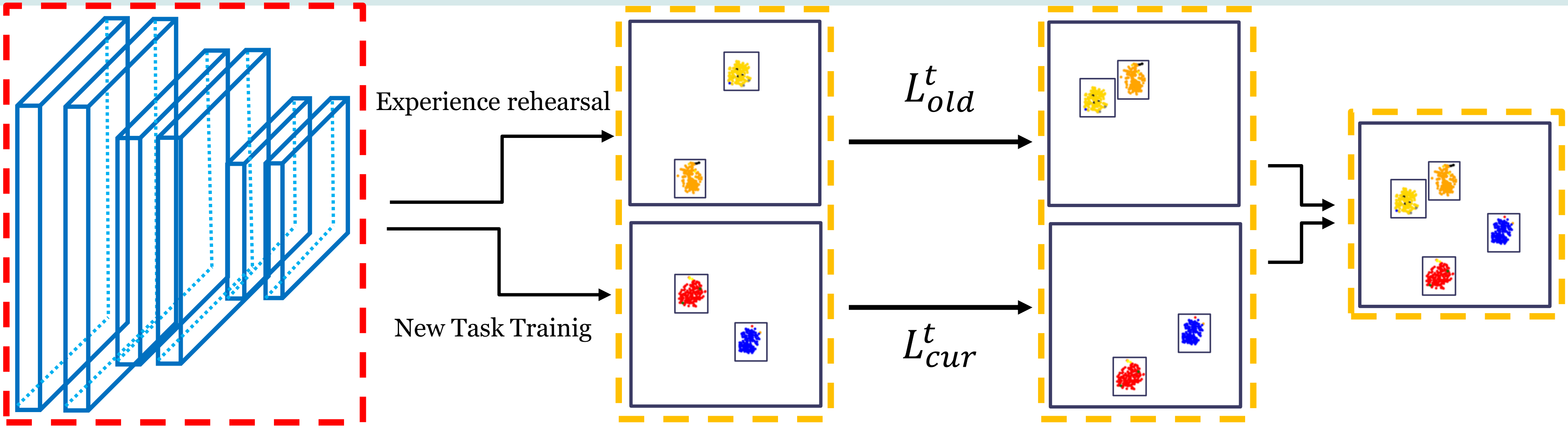
Learning Previous Task

Problem Statements

Proposed Method



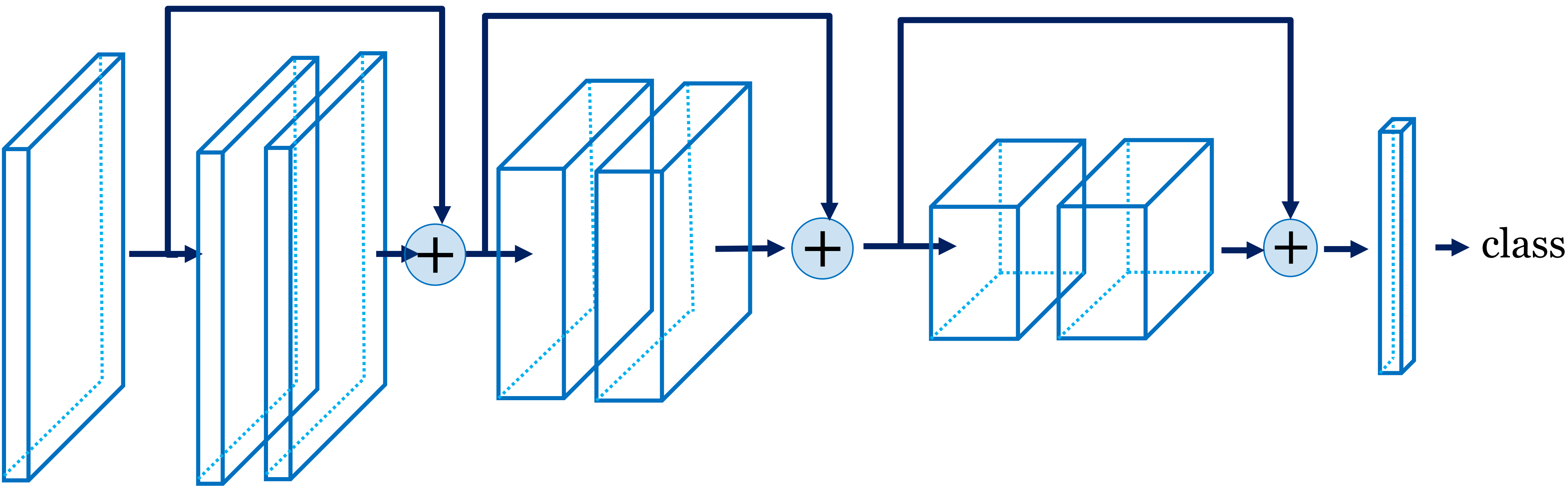
Our goal : Moving Embedded Clusters(MEC)



- Preserve previous knowledge
  - Restrict old cluster's moving during new task learning
  - $L^t_{old} = \sum_{j=0}^{i}(mean_j^{t-1} - mean_j^{ex})$
- Adapting new knowledge
  - Keep current cluster far away from old cluster's position on embedded space
  - $L^t_{cur} = -\sum_{j=0}^{i}(mean_j^{t-1} - mean_j^{cur})$

Experiments

Backbone Network : ResNet



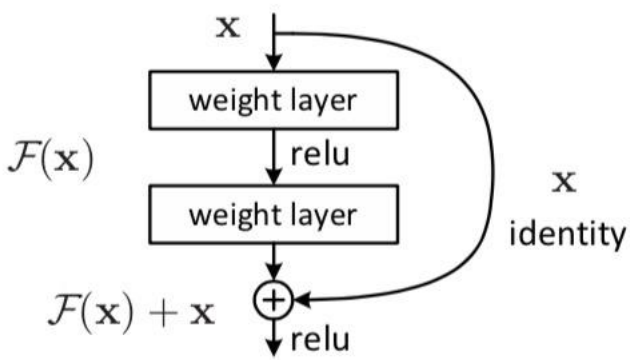
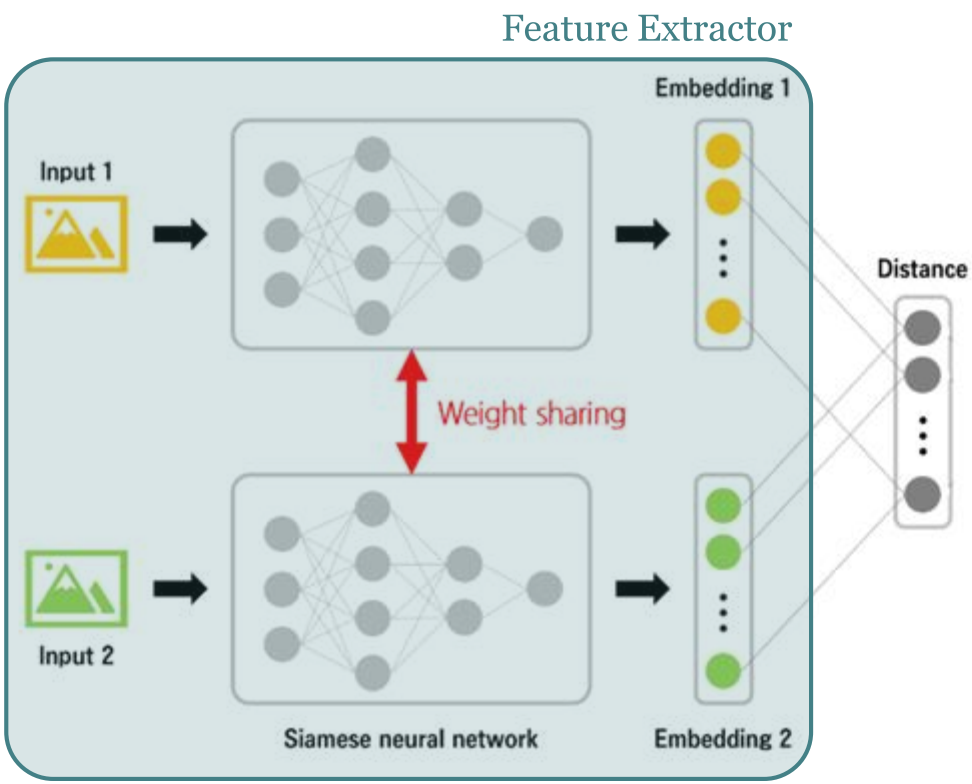


Figure 2. Residual learning: a building block.

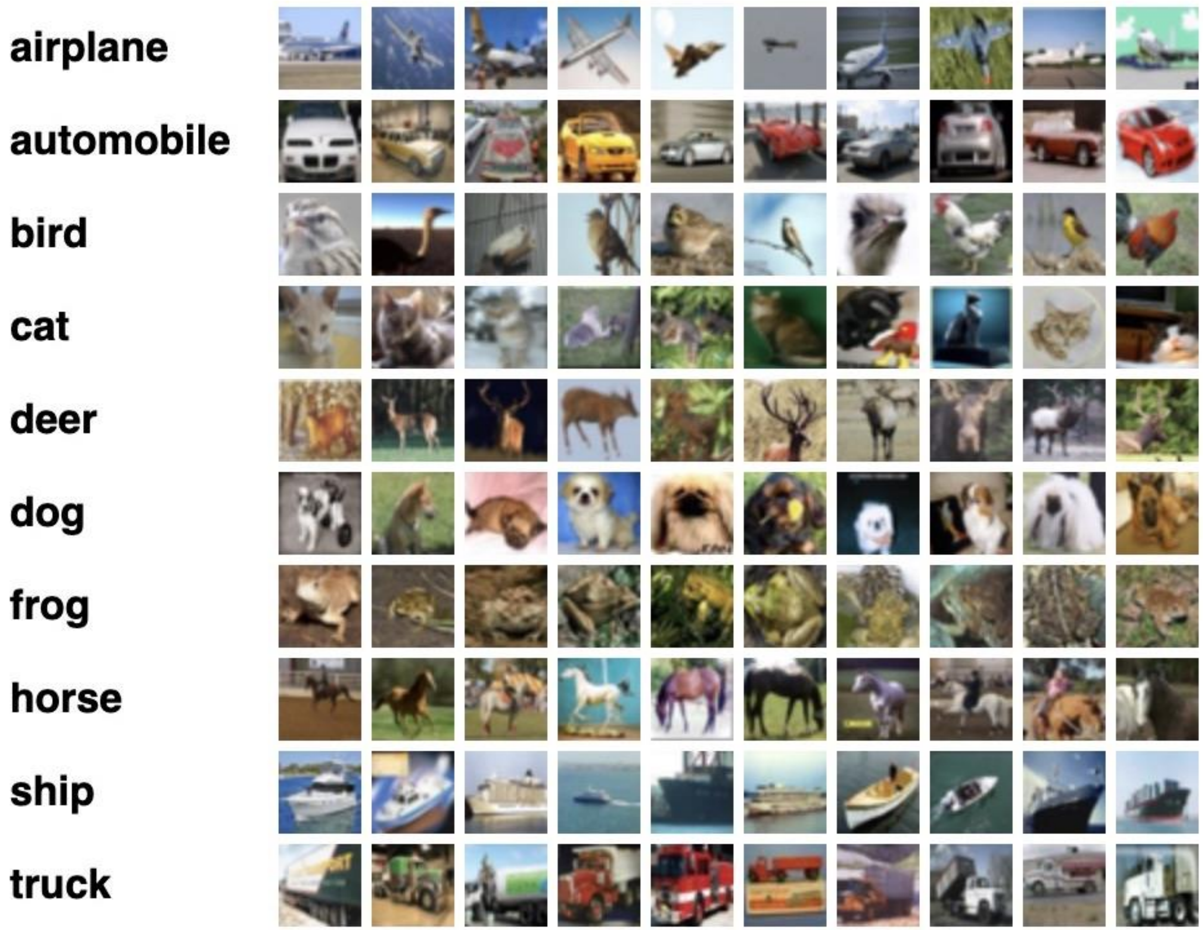
Embedding NET : Siamese Network

- Two images are utilized as inputs and return similarity between two features.
- Consequently, Siamese NET converts input data into embedding.
  - Usually used in pre-processing of Natural Language Processing



Dataset : CIFAR-10

- CIFAR-10 : 10 classes(5000 train images/class, 1000 test images/class), size(32*32)



Preliminary Results

- Accuracy upon testing data at the end of the learning task
  - Epoch : 60, Learning rate : 0.1
  - Backbone Network : ResNet-32
  - Baseline : Knowledge Distillation(KD) / Proposed Method : KD + MEC

KD

	Task 1	Task 2
Iter 1	79.4	
Iter 2	65.93	83.46
diff	**13.47**	-

KD + MEC (proposed method)

	Task 1	Task 2
Iter 1	77.62	
Iter 2	65.62	85.16
diff	**12.03**	-

Future Work

Future work

- Validate robustness upon 3 or more tasks
- Force shape of feature distribution like cluster.
- Comparison of the cluster's state
  - e.g. KL-divergence
- Consider application : Contact point of pneumatic and parallel gripper grasping
    - 공압NET, 페러렐 NET
- Balancing old task and current one
  - Hyperparameter analysis

Thanks for your attention

Reference

1. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009.
2. Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *Proceedings of the national academy of sciences* 114.13 (2017): 3521-3526.
3. Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." *arXiv preprint arXiv:1503.02531* (2015).
4. Sprechmann, Pablo, et al. "Memory-based parameter adaptation." arXiv preprint arXiv:1802.10542 (2018).