

졸업프로젝트

최종 발표

이정호

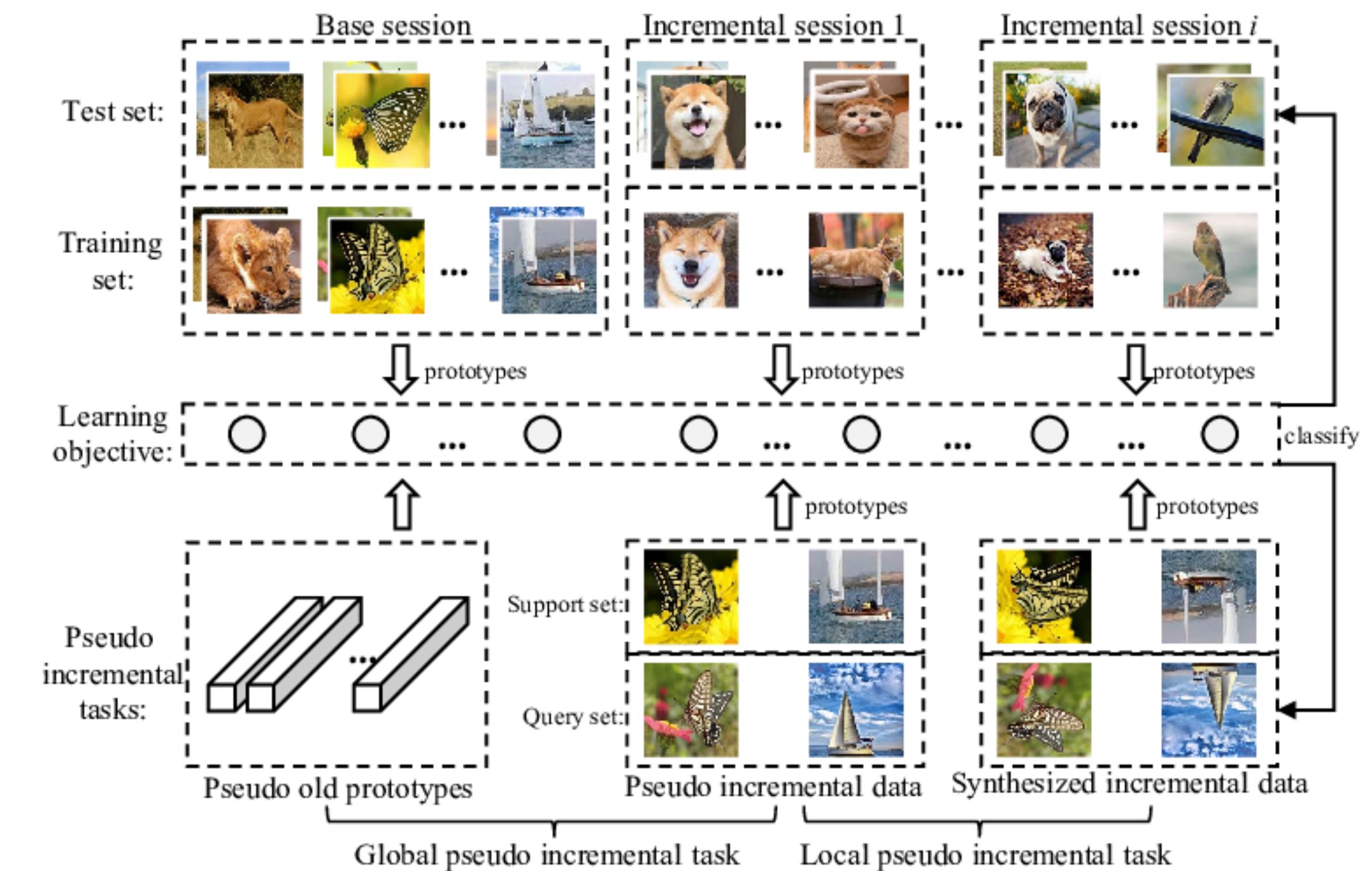
목차

- **Introduction**
- **Related Work**
 - Few Shot Class Incremental Learning
 - Pre-trained Network
 - Meta Learning
 - LoRA
- **Method**
 - Modulation
 - Pre-trained Knowledge Tuning
 - Energy-based Divergence Loss
- Semantic Knowledge Distillation Loss
- Neural Collapse Loss
- LoRA
- Model Architecture
- **Experiments**
 - Experimental Settings
 - Main Experimental Results
 - Ablation Study
- **Conclusion**
- **Reference**

Introduction

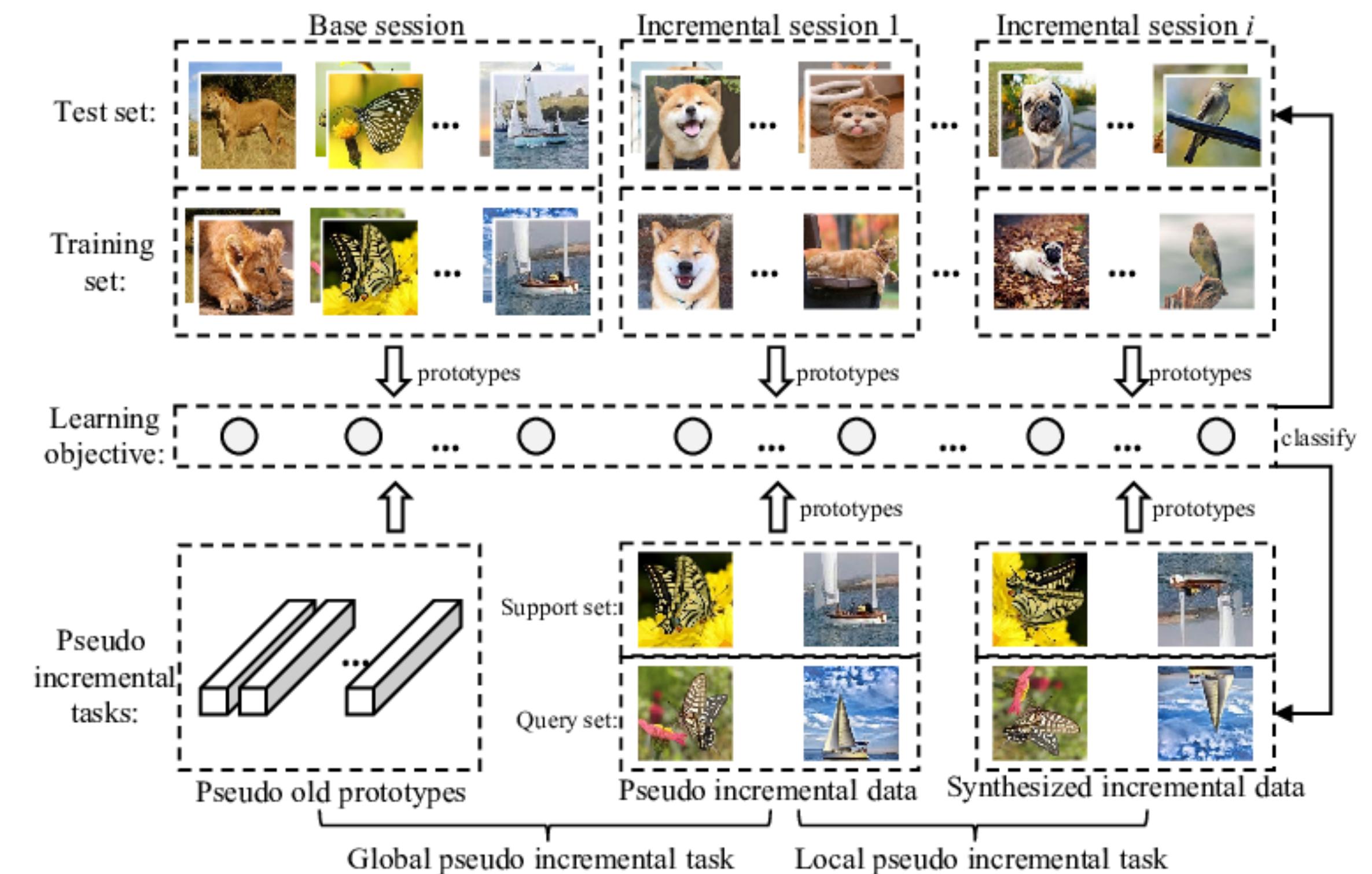
Introduction

- FSCIL 은 적은 Data 로 새로운 Class 를 학습하면서도 기존 Class 에 대한 지식을 유지하는 학습 방식이다.
- 대부분의 FSCIL Model 은 대량의 Data 를 필요로 하는 Base Session 초기 학습 단계에 의존하고 있으며, 이는 현실적인 응용 환경과는 괴리가 있다.



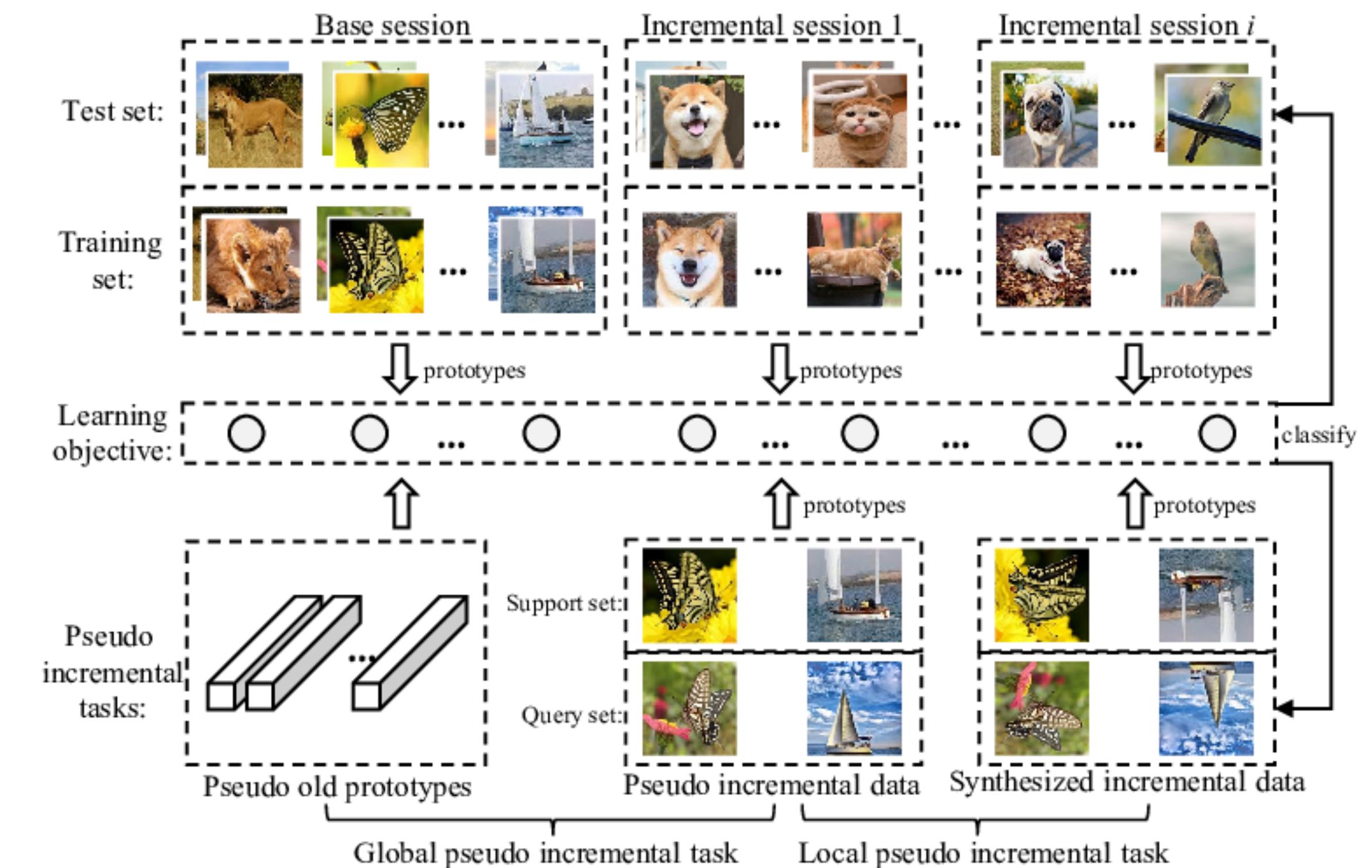
Introduction

- 실제 응용에서는 Data 수집이 제한적일 뿐만 아니라, 학습에 이용되는 Data의 Privacy 및 보안 문제가 중요한 제약으로 작용할 수 있다.
- 더 나아가, 적은 Data로 효율적으로 학습한다는 FSCIL의 본질적인 목표와 충돌한다.



Introduction

- 따라서, Base Session 학습 과정을 간소화하거나 완전히 제거하고 초기 Data 가 부족한 환경에서도 효과적으로 동작할 수 있는 FSCIL Model 을 개발하는 것이 이 분야에서 필수적이다.

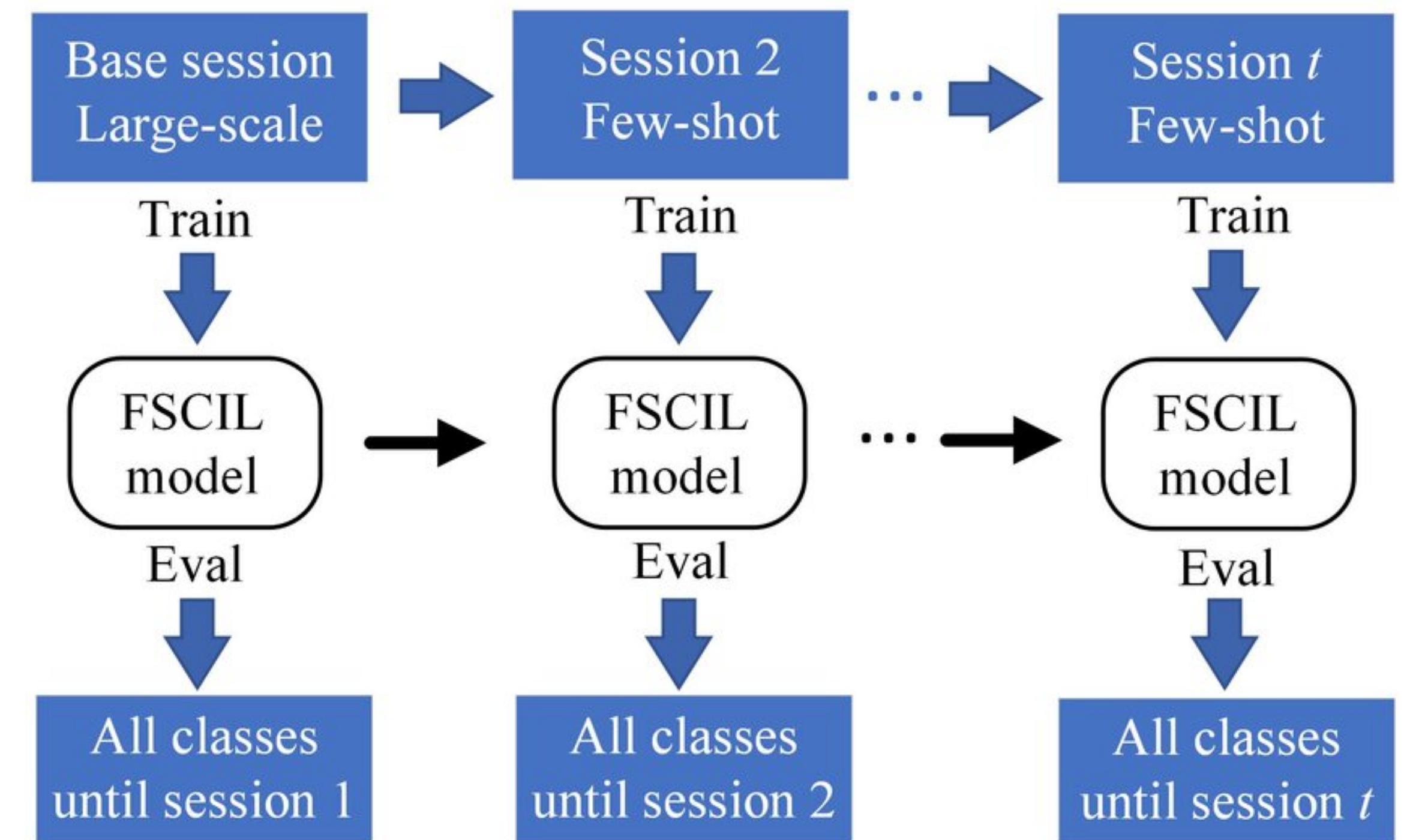


Related Work

Introduction

Few Shot Class Incremental Learning

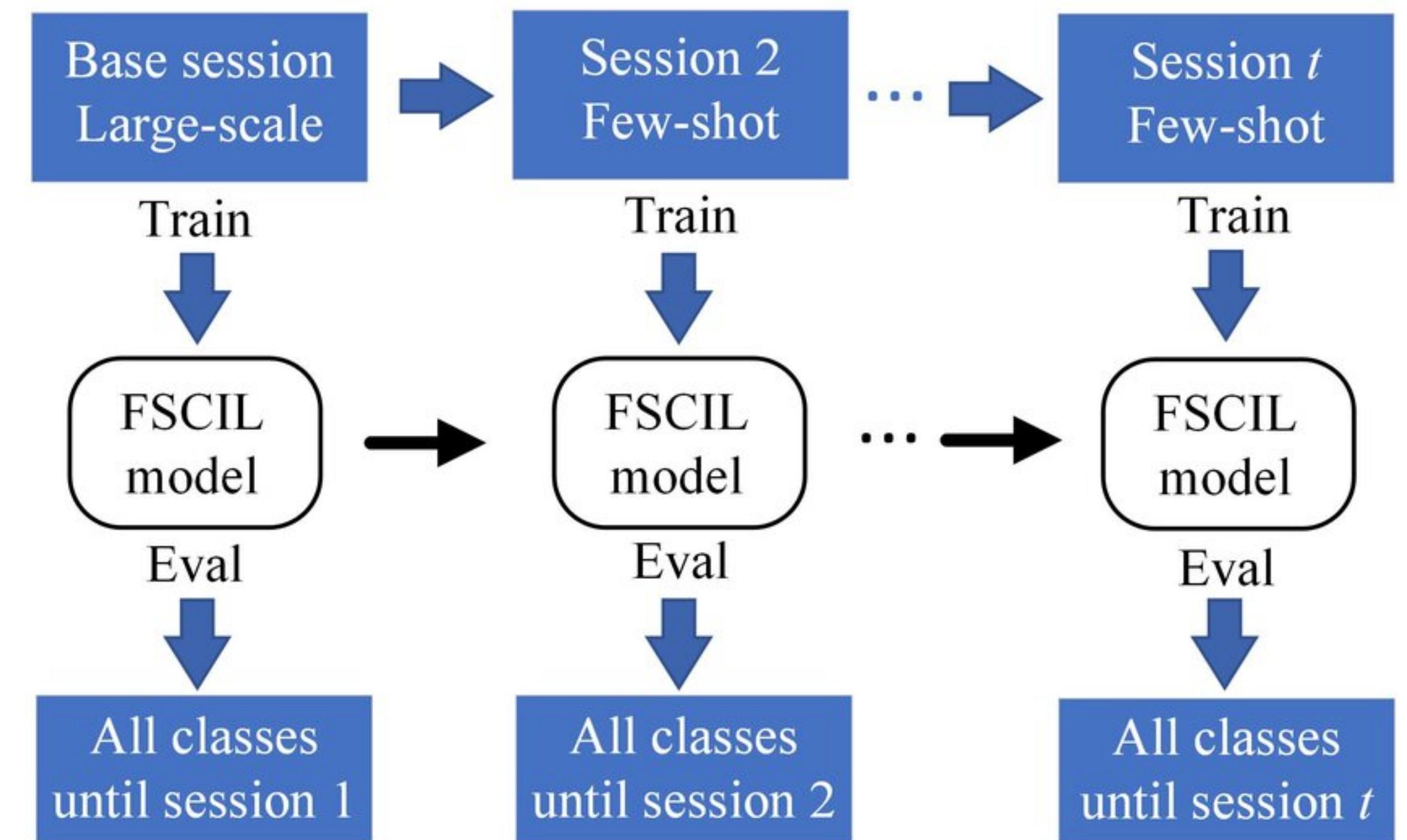
- FSCIL Scenario 는 새로운 Class 를 점진적으로 학습해야 하는 상황에서 적은 양의 Data 를 사용하여 Model 을 효율적으로 학습시키는 문제를 다루는 Scenario 이다.
- 대부분의 FSCIL Model 은 초기 학습 단계 (Base Session) 에서 대규모 Dataset 으로 다수의 Class 를 학습하며, 이후 (Incremental Session) 점진적으로 새로운 Class 에 대한 Data 를 받는다.



Introduction

Few Shot Class Incremental Learning

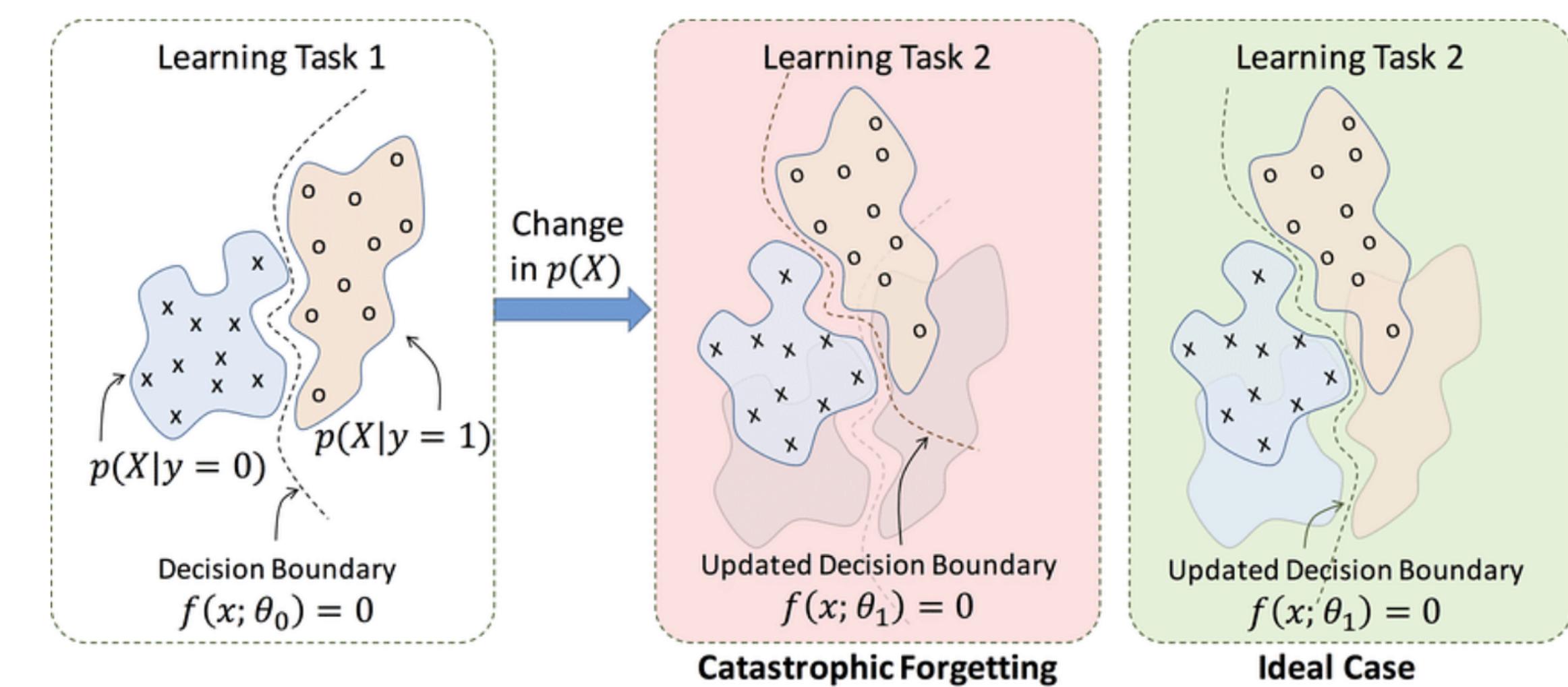
- Base Session 은 다수의 Class 를 학습하는 초기 학습 단계로, 이 때 대 규모 Dataset 으로 학습하는 경우가 일반적이다.
- Incremental Session 은 새로운 Class Data 가 추가될 때마다 Model 을 Update 하는 증분 학습 단계로, 이 때 Data 가 극도로 제한된다.



Introduction

Few Shot Class Incremental Learning

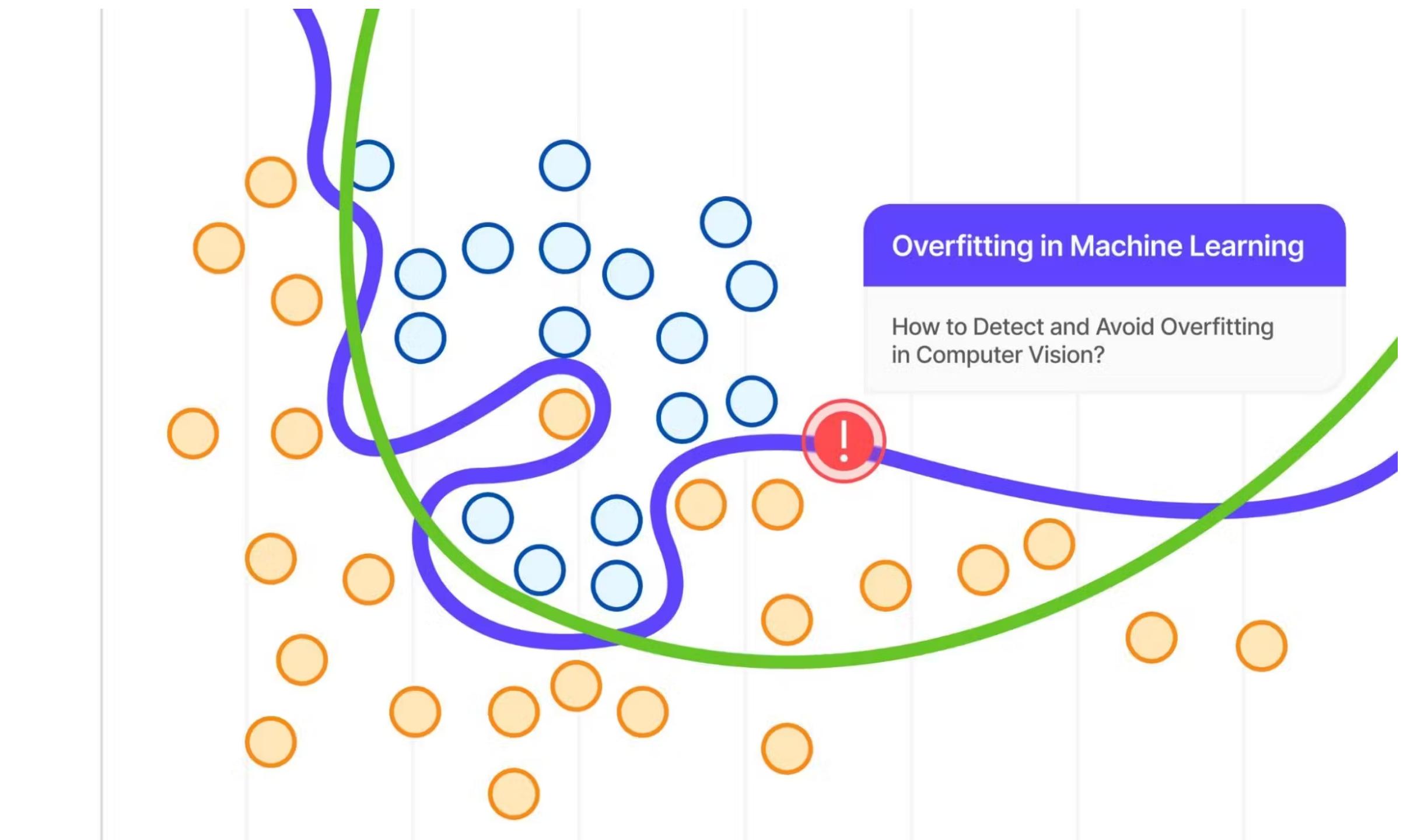
- FSCIL Scenario 의 주요 도전 과제는 Catastrophic Forgetting 과 Overfitting 이다.
- Catastrophic Forgetting 은 순차적으로 새로운 Class 들을 학습하는 과정에서 기존에 학습했던 Class 들을 망각하는 현상이다.



Introduction

Few Shot Class Incremental Learning

- FSCIL Scenario 에서의 Overfitting은 새로운 Class 를 학습하는 과정에서 한정된 수의 Training Data 에 과도하게 집중하여 전체적인 성능을 하락시키게 되는 현상이다.



Related Work

Pre-trained Network

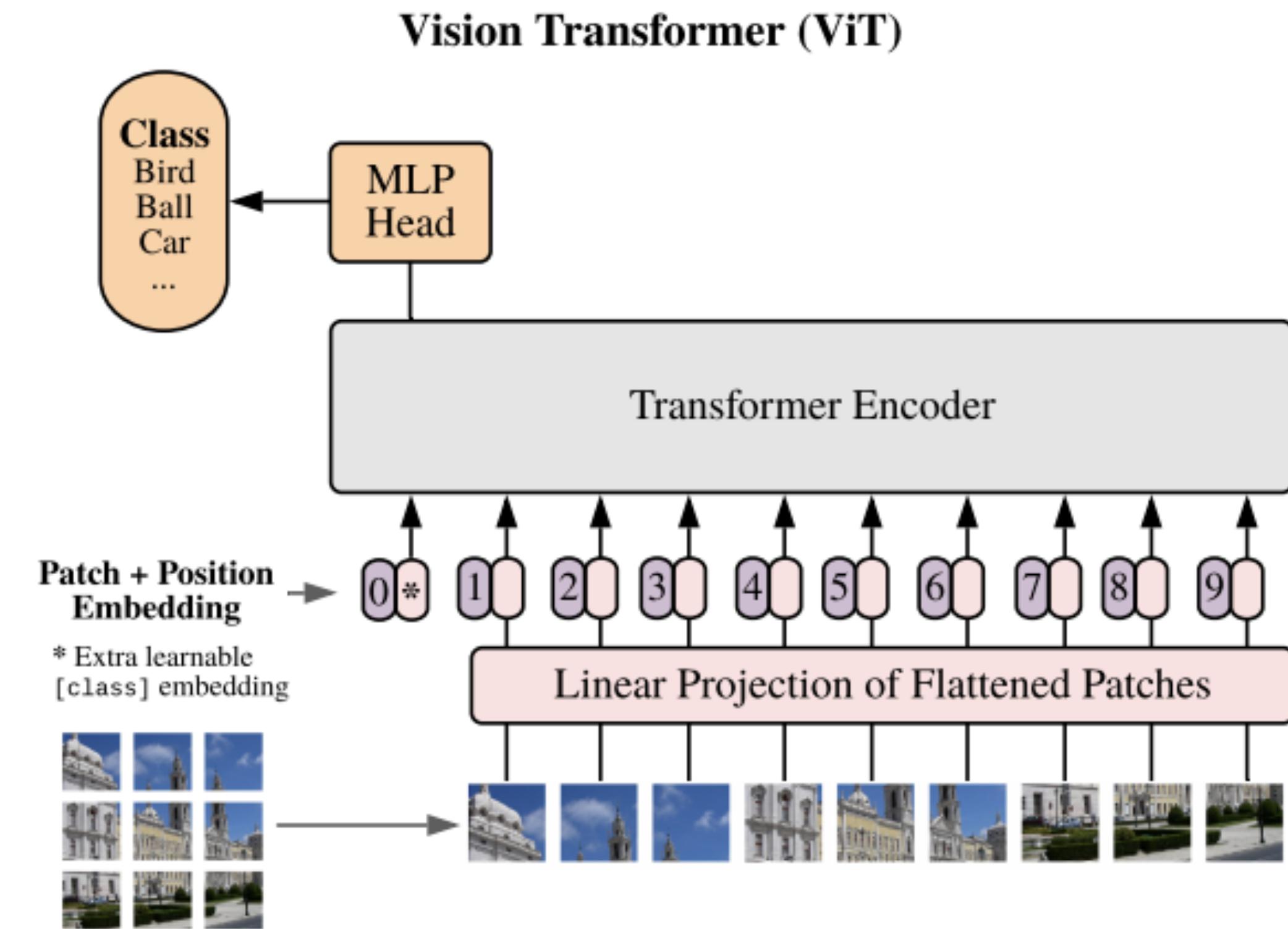
- 현재 FSCIL Task 에서 가장 좋은 성능을 보이는 Model 은 대부분 Pre-trained Network (Vision Transformer) 를 사용 한다.
- FSCIL Scenario 에서의 대부분의 Model 이 채택한 학습 방식은 Base Session 에서 대규모 Dataset 에서 학습한 지식을 Incremental Session 으로 전이하는 방식으로 학습한다.

1	PriViLege	86.06	88.08	Pre-trained Vision and Language Transformers Are Few-Shot Incremental Learners			2024	Uses Pre-trained network
2	SV-T	69.75	76.84	Semantic-visual Guided Transformer for Few-shot Class-incremental Learning			2023	Uses Pre-trained network
3	BOT	58.75		A Bag of Tricks for Few-Shot Class-Incremental Learning			2024	
4	NC-FSCIL	56.11	67.50	Neural Collapse Inspired Feature-Classifier Alignment for Few-Shot Class-Incremental Learning			2023	
5	FACT	52.10	62.24	Forward Compatible Few-Shot Class-Incremental Learning			2022	

Related Work

Pre-trained Network

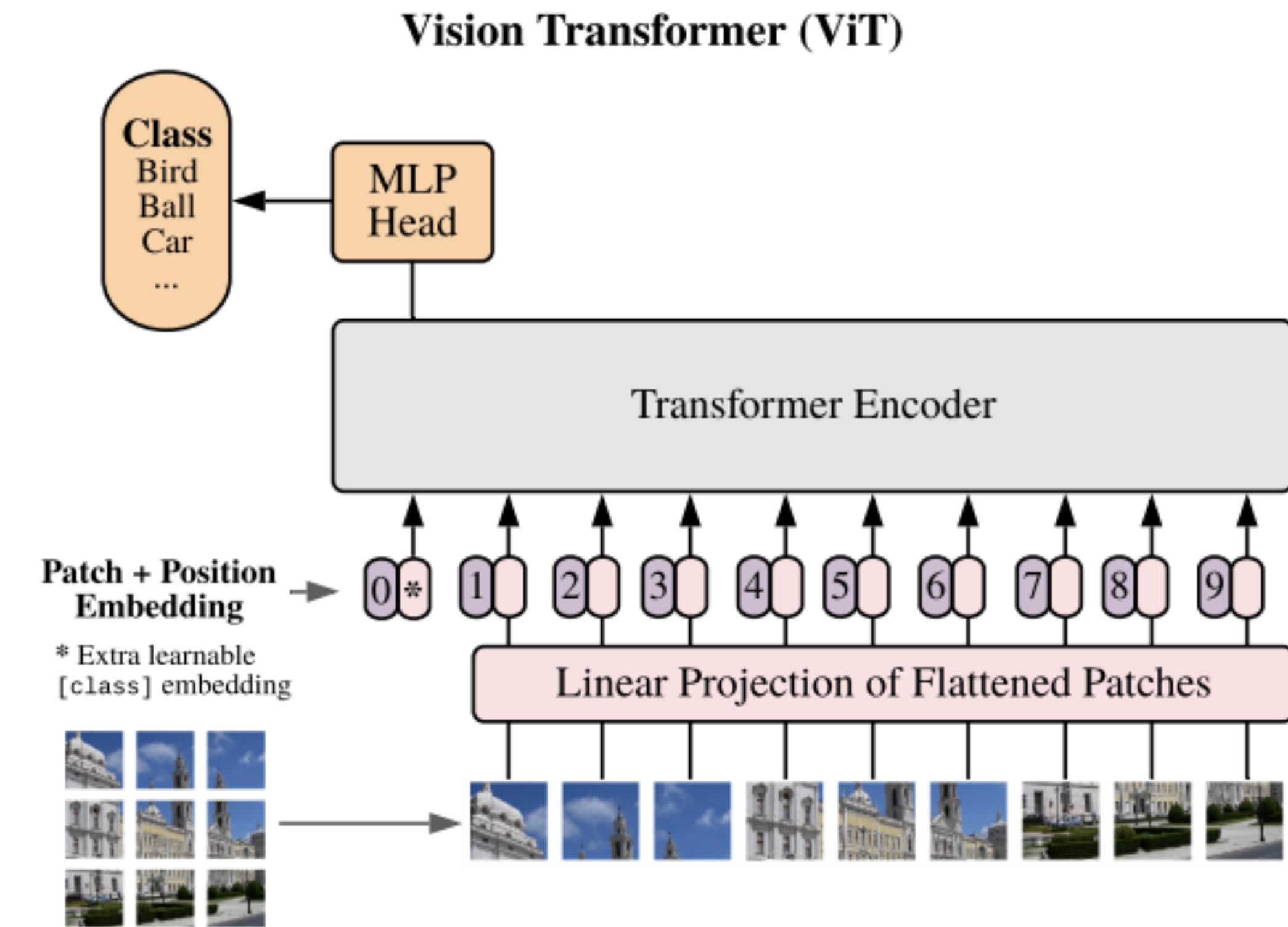
- Pre-trained Network 는 대규모 Dataset 으로 학습하여 일반화된 Feature 를 효과적으로 추출하는 강력한 Feature Extractor 역할을 한다.
- 초기 학습 없이도 높은 품질의 Feature 를 제공하므로 Base Session 과 Incremental Session 에서 적은 Data 로도 학습이 가능하다.



Related Work

Pre-trained Network

- Pre-trained Network 의 고정된 저수준 Feature 는 Incremental Session 과정에서 변하지 않으므로, 기존 Class 에 대한 지식 손실이 최소화된다.
- 따라서 이 연구에서는 적은 Data 를 가지고 Pre-trained Network 를 FSCIL Scenario 에 빠르게 적응시키고, 그 지식을 Incremental Session 으로 잘 전이 할 수 있는 방법을 모색할 것이다.



Related Work

Meta Learning

- Meta Learning 은 학습하는 방법을 학습하는 개념으로 Model 이 다양한 Task 에 빠르게 적응할 수 있도록 학습하는 방법이다.
- 이를 통해 Model 은 새로운 Task 에서 적은 Data 나 적은 학습 과정만으로도 좋은 성능을 낼 수 있도록 유도한다.

Meta Learning

Learn To Learn Task

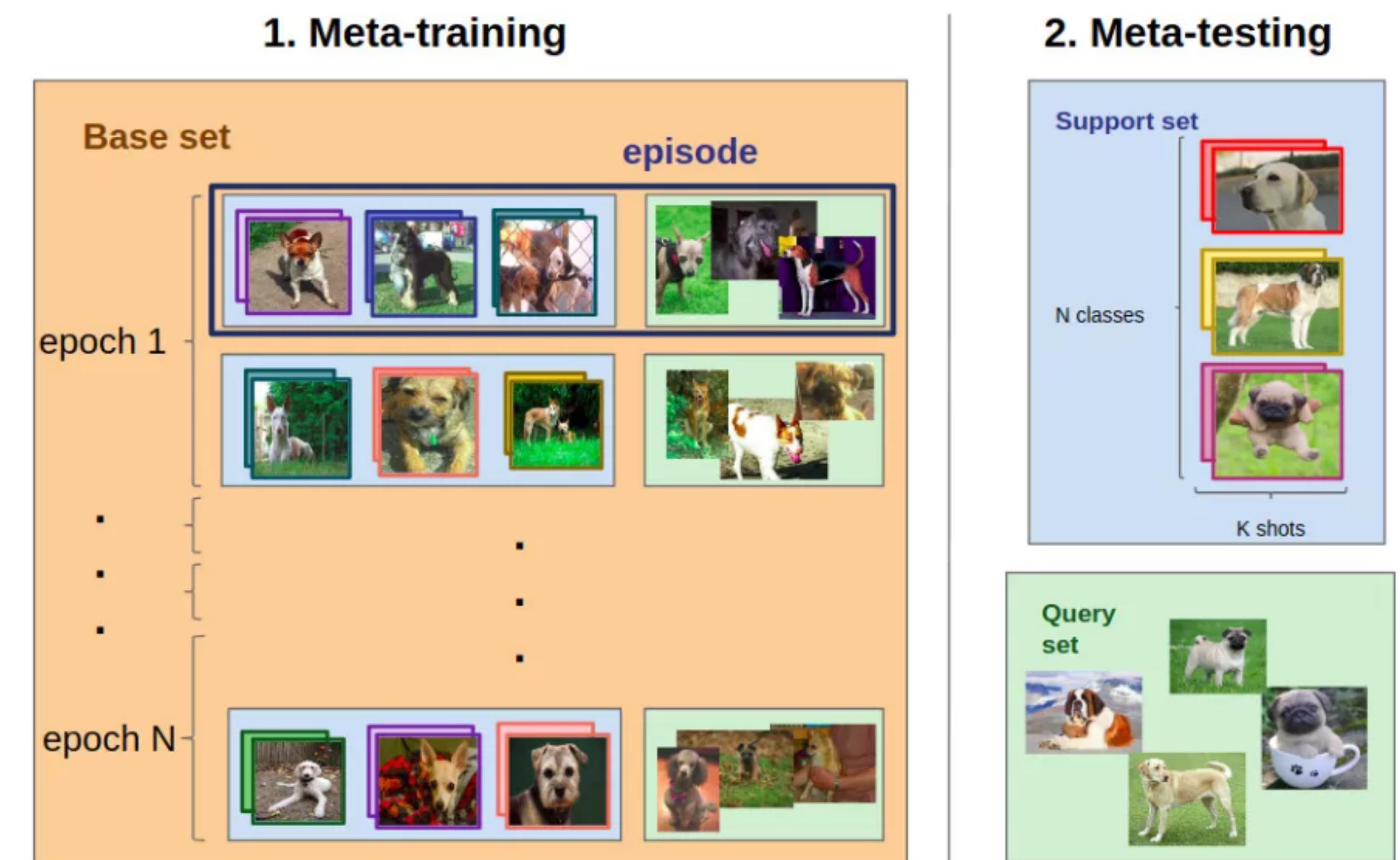


Quickly learn
New Task

Related Work

Meta Learning

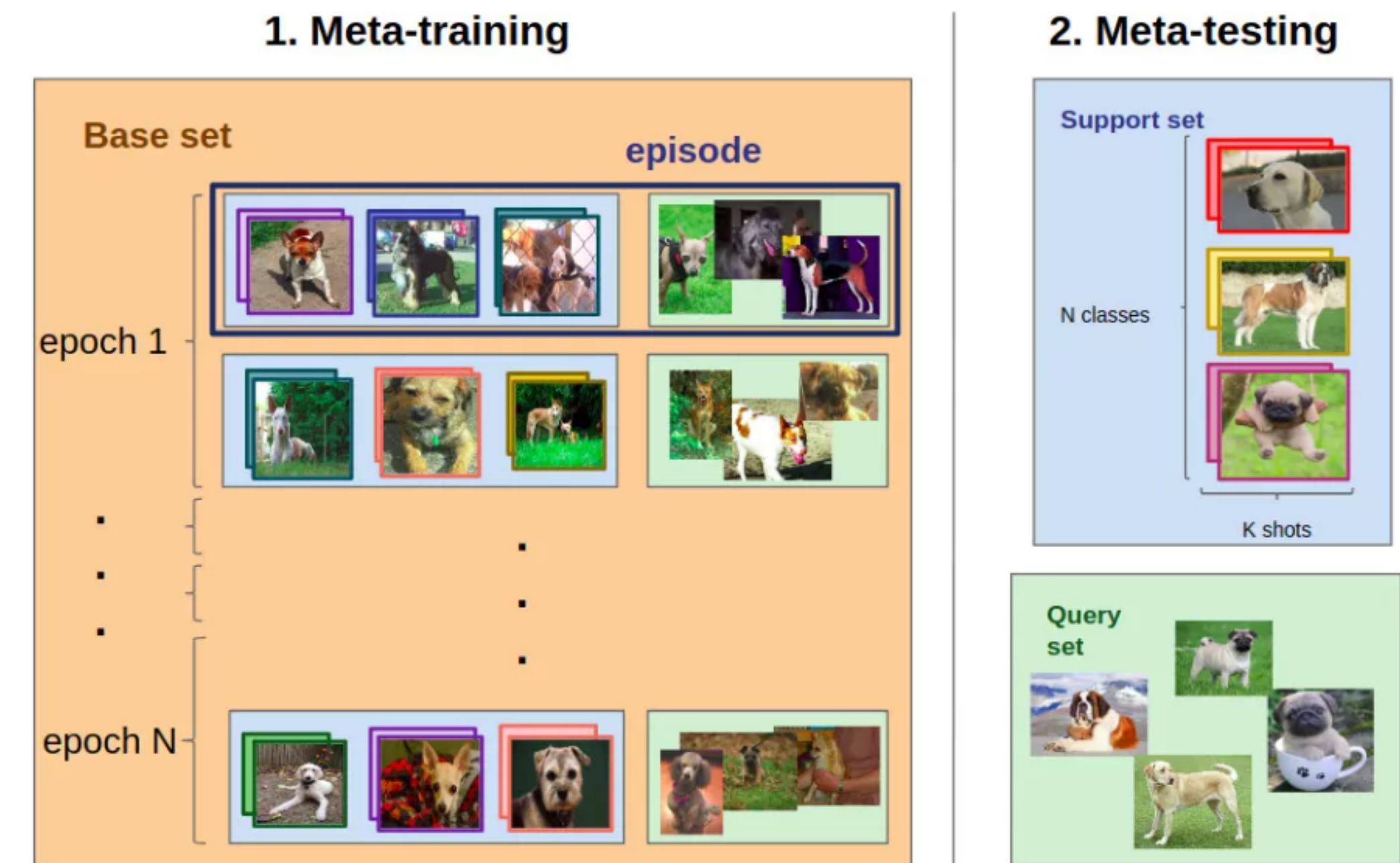
- Episode 기반 학습; 하나의 Episode에 제공되는 Data는 소량의 Data로 이루어진 Support Set과 Query Set으로 나뉜다.
- Support Set S 는 각 Class의 Prototype을 생성하는 데 사용되며, Query Set Q 는 Model을 학습하기 위한 손실을 계산하는 데 사용된다.



Related Work

Meta Learning

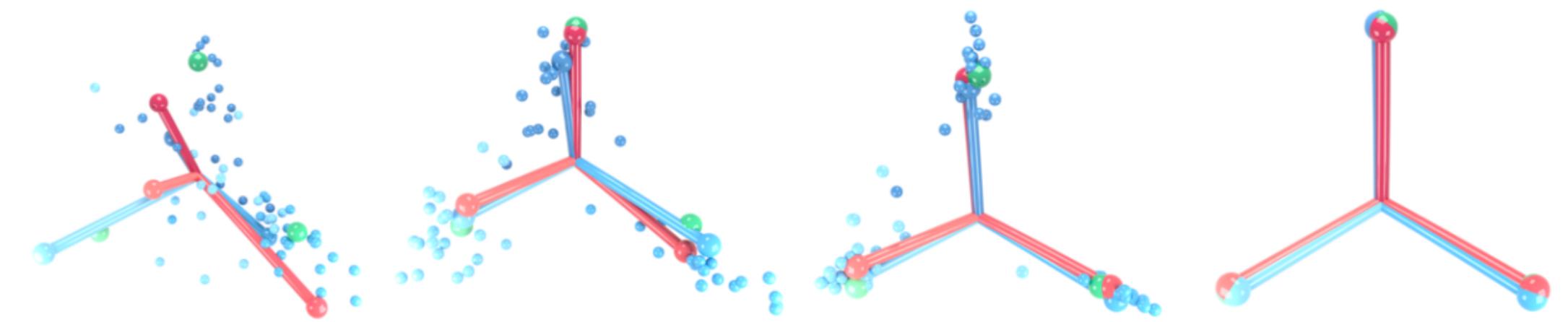
- Episode 기반 학습은 Data 를 소량만 사용할 수 있는 환경에서 Model 의 일반화 능력을 극대화하는 데 목적을 둔다.
- Pre-trained Network 를 적은 Data 로 Meta-training 하여 FSCIL Scenario 에 빠르게 적응할 수 있도록 유도한다.



Related Work

Neural Collapse

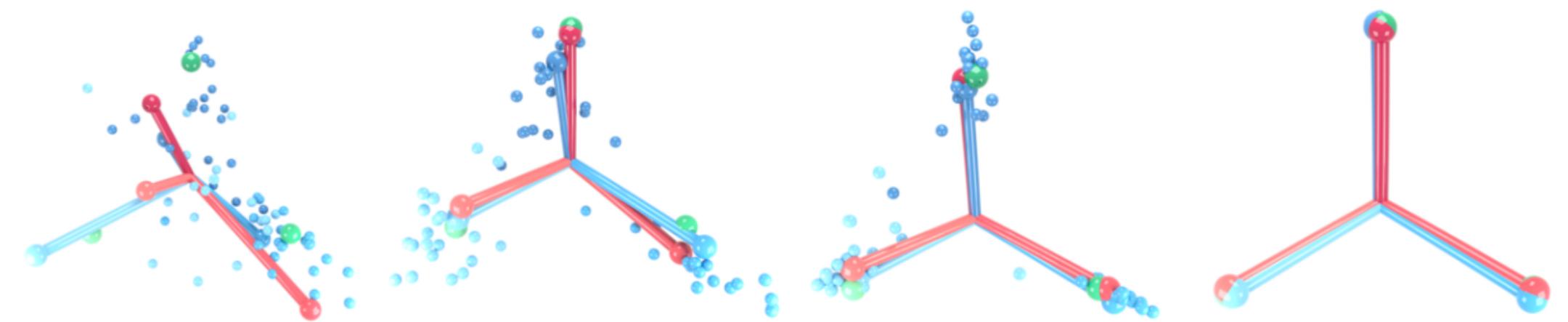
- NC 이론에 따르면, Deep Neural Network 는 학습이 끝나갈 때 Network 의 마지막 Layer 에서 같은 Class 의 Feature 들이 단일 점으로 수렴하는 현상을 보인다.
- 이 현상은 Network 가 Class 간 구별이 극대화되도록 각 Class 의 Feature 를 조정하기 때문이다.



Related Work

Neural Collapse

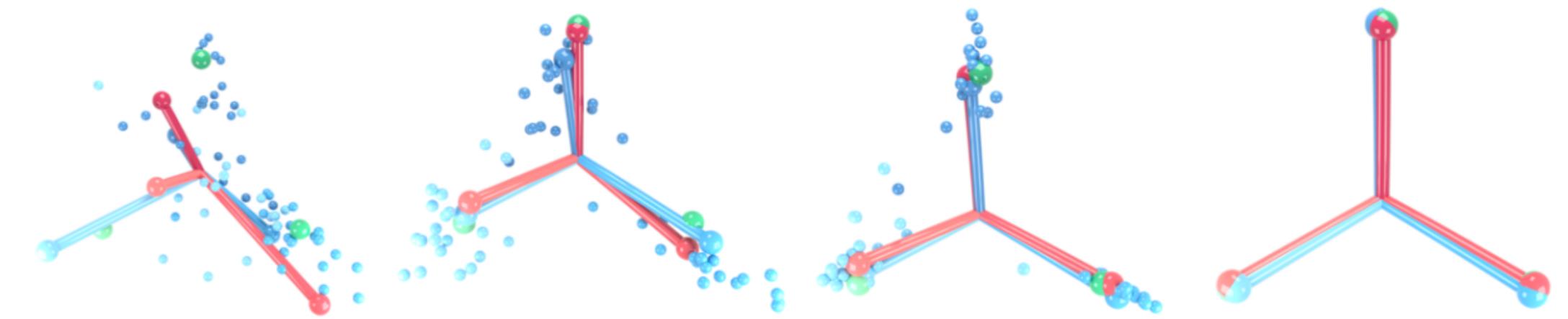
- Class 간 각 Class 의 Proto Vector 들이 이상적인 기하학적 구조를 이루게 되는데, 이를 Simplex Euclidean Tight Frame 라고 한다.
- 즉, 모든 Class 가 동일한 크기의 각도를 유지하도록 배열되며, 이는 분류의 효율성을 극대화하는 구조이다.



Related Work

Neural Collapse

- 따라서 같은 Class 의 Proto Vector 와의 유사도는 최대값인 1에 가까워지도록 학습하고, 다른 Class 의 Proto Vector 와의 유사도는 이론적 최소값에 가까워지도록 학습한다.
- 이를 통해 Model 이 적은 Data 로도 빠르게 학습할 수 있도록 유도하여 학습의 효율을 늘린다.

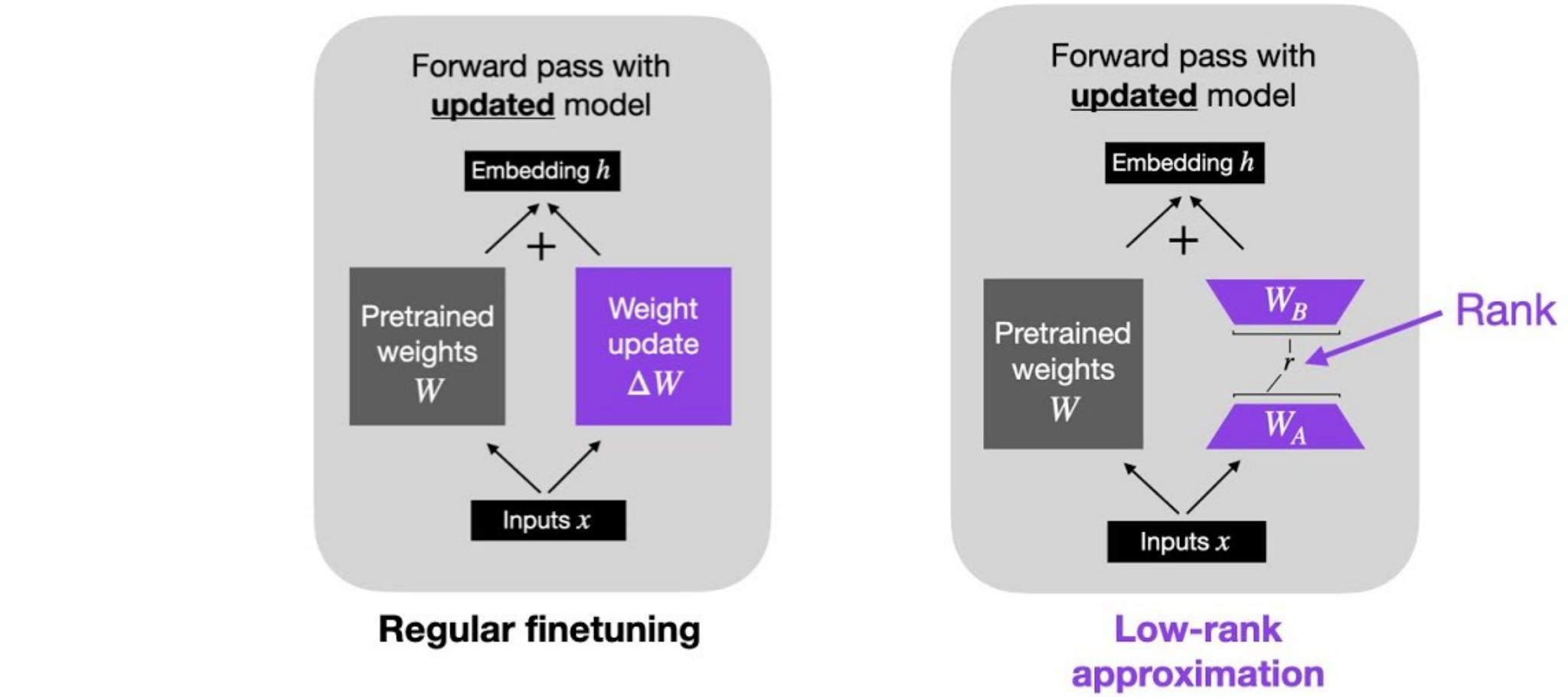


Related Work

LoRA

- 저차원 행렬을 이용한 가중치 Update로 학습하는 Parameter의 수를 줄여 학습에 필요한 시간과 자원을 줄이는 것 이 LoRA의 목표이다.
- FSCIL Scenario에서는 LoRA가 일부 Parameter만 학습한다는 점에서 기존의 지식을 유지하여 Catastrophic Forgetting을 방지하게끔 유도한다.

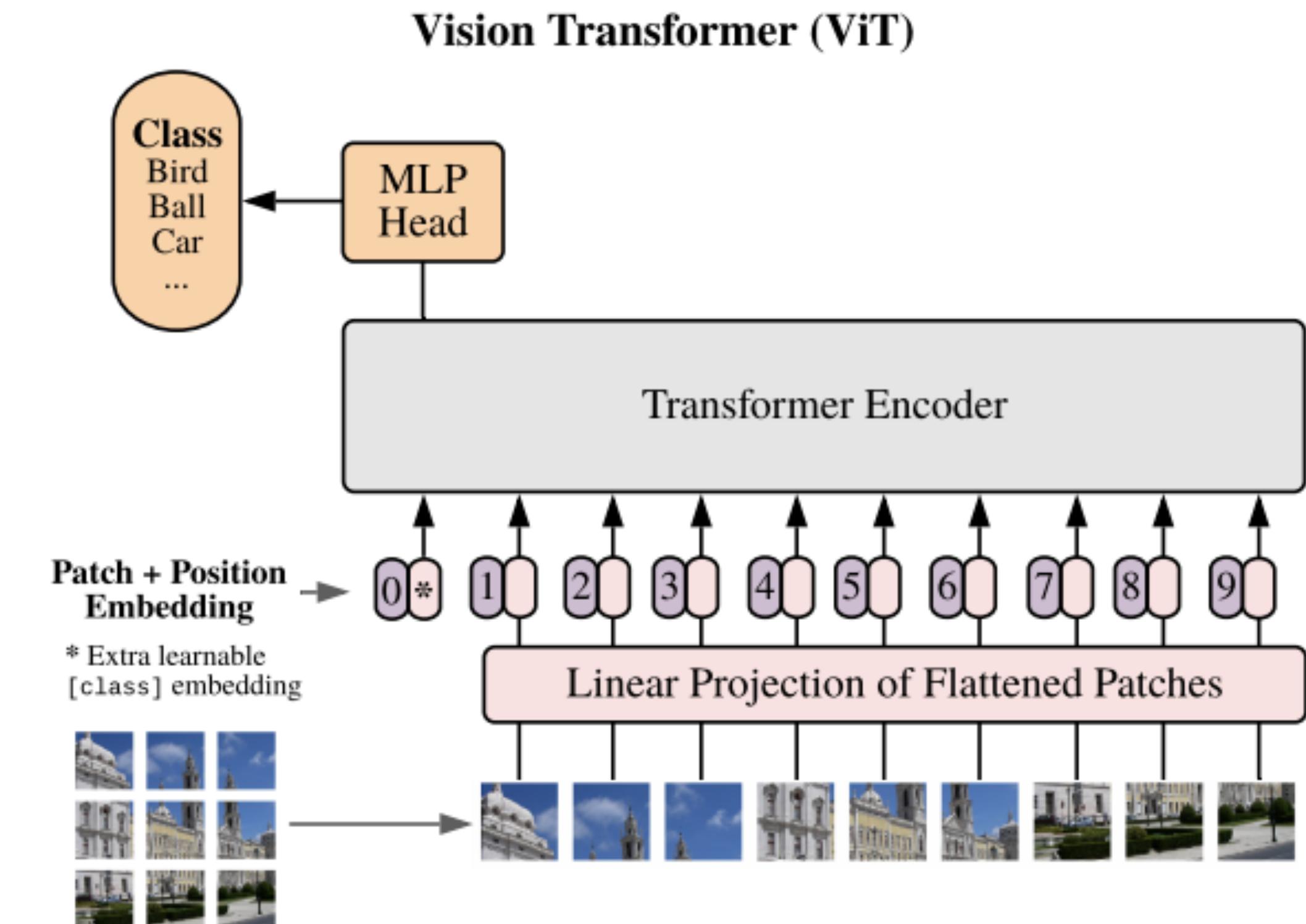
Explaining LoRA in a nutshell



Method

Method Modulation

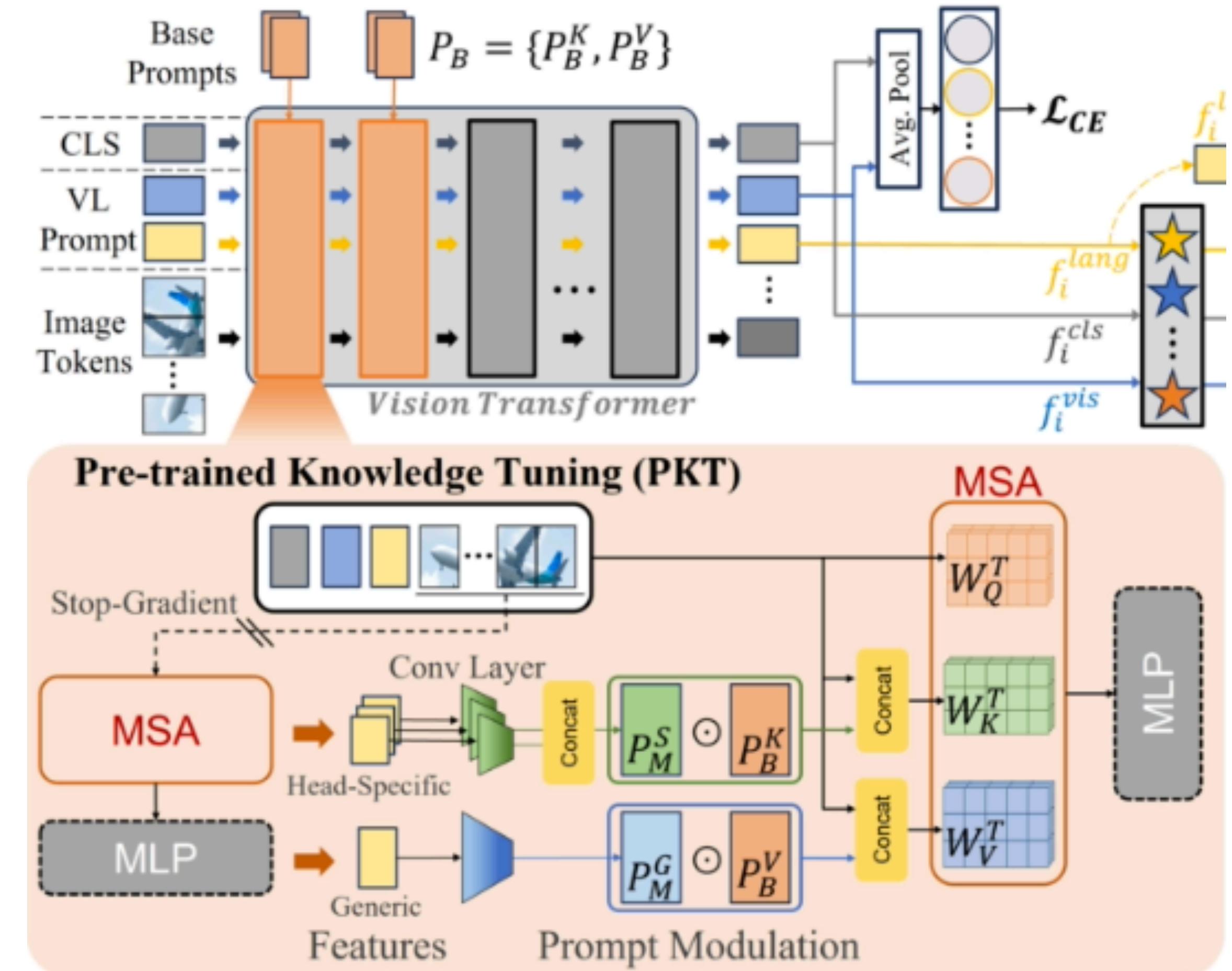
- FSCIL Scenario 의 주요 도전 과제
인 Catastrophic Forgetting 과
Overfitting 문제를 해결해야 한다.
- 동시에 Base Session 학습에 주어지
는 Data 가 한정적이므로 Pre-
trained Network 가 빠르게 적응할
수 있도록 해야 한다.



Method

Pre-trained Knowledge Tuning

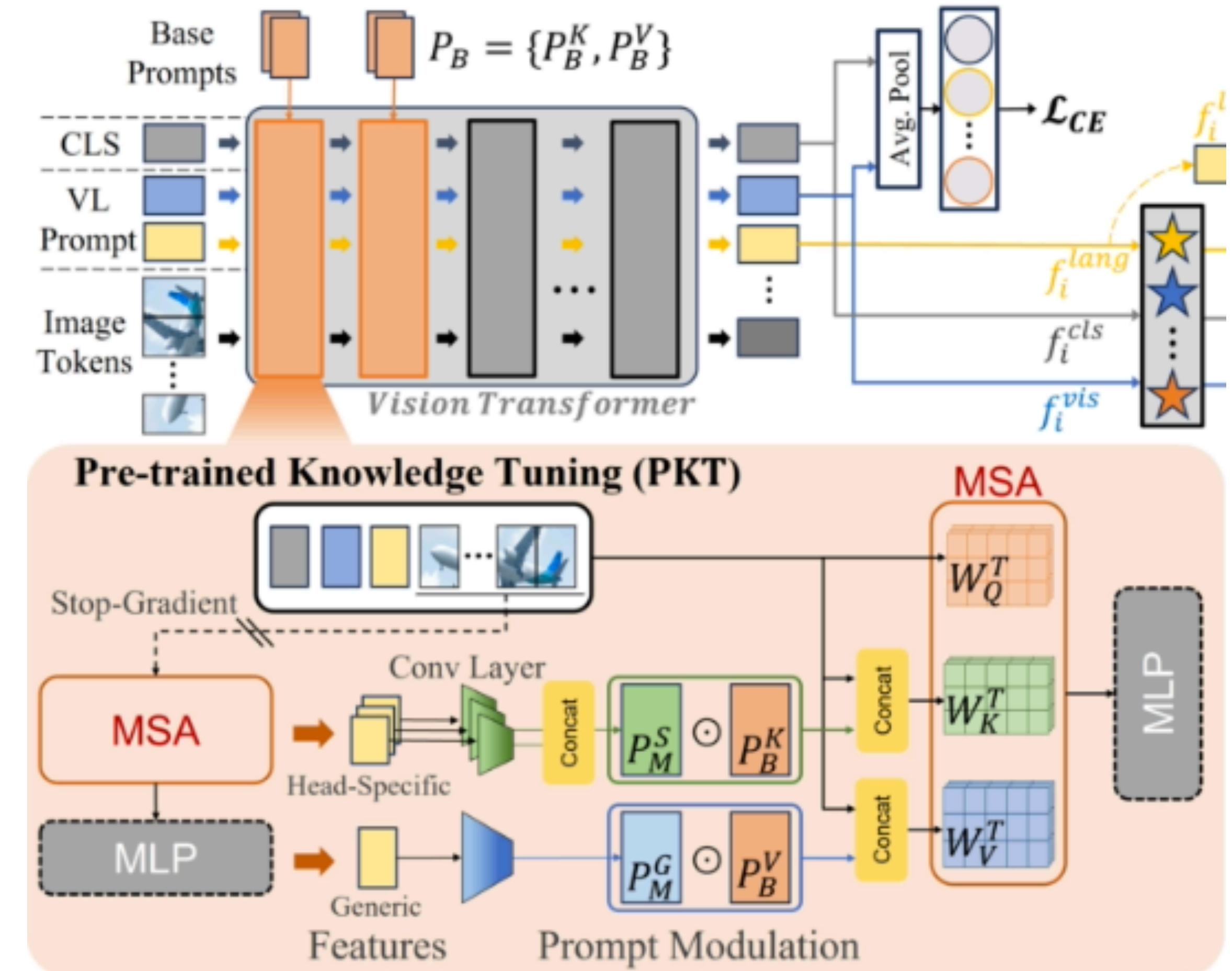
- PriViLege 는 Pre-trained 된 ViT Model 의 처음과 마지막 각각 2개의 Layer 만 학습하고 나머지 Layer 는 전부 Freezing 하여 사전 지식을 보존한다.
- 기존에는 Base Session 학습이 Fine-tuning 이어서, 하위 Layer 만 학습하여 Domain 의 일반적인 특성에 집중하여 학습했지만, 한 Episode 당 Few-shot 으로 Data 가 주어지는 Episode 기반 학습 특성 상, 상위 Layer 도 학습하여 추상적 표현을 학습한다.



Method

Pre-trained Knowledge Tuning

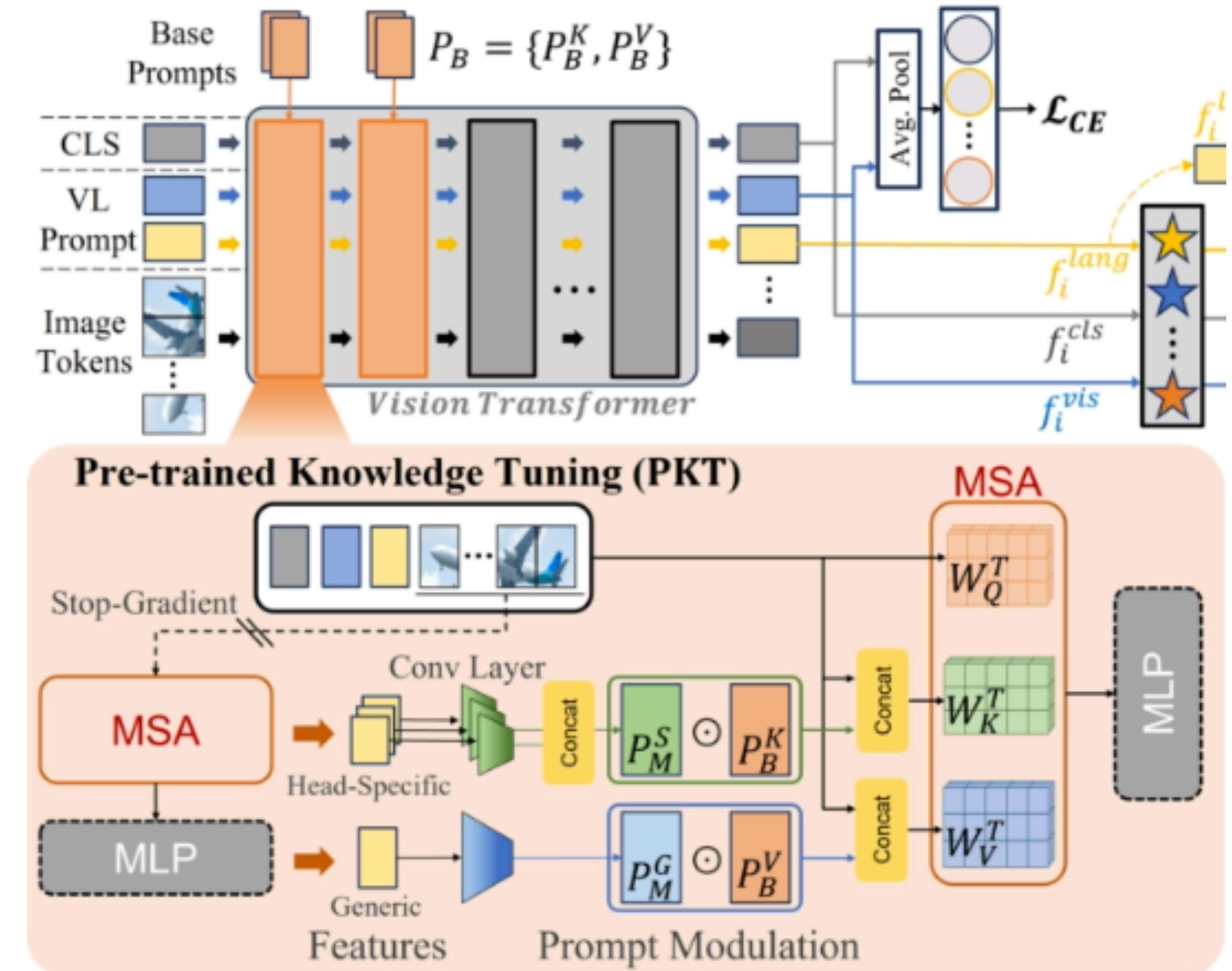
- 대부분의 Layer 를 Freezing 했기 때문에, Incremental Learning 의 성능이 한정적이었는데, B Prompt, VL Prompt 로 성능을 향상시켰다.
- B Prompt 는 Base Session 학습 과정에서 ViT 를 학습시키면서 Domain 의 핵심 정보를 파악한다.
- Vision, Language Token 으로 구성된 VL Prompt 는 이전 Session 에서 학습된 지식을 전달한다.



Method

Pre-trained Knowledge Tuning

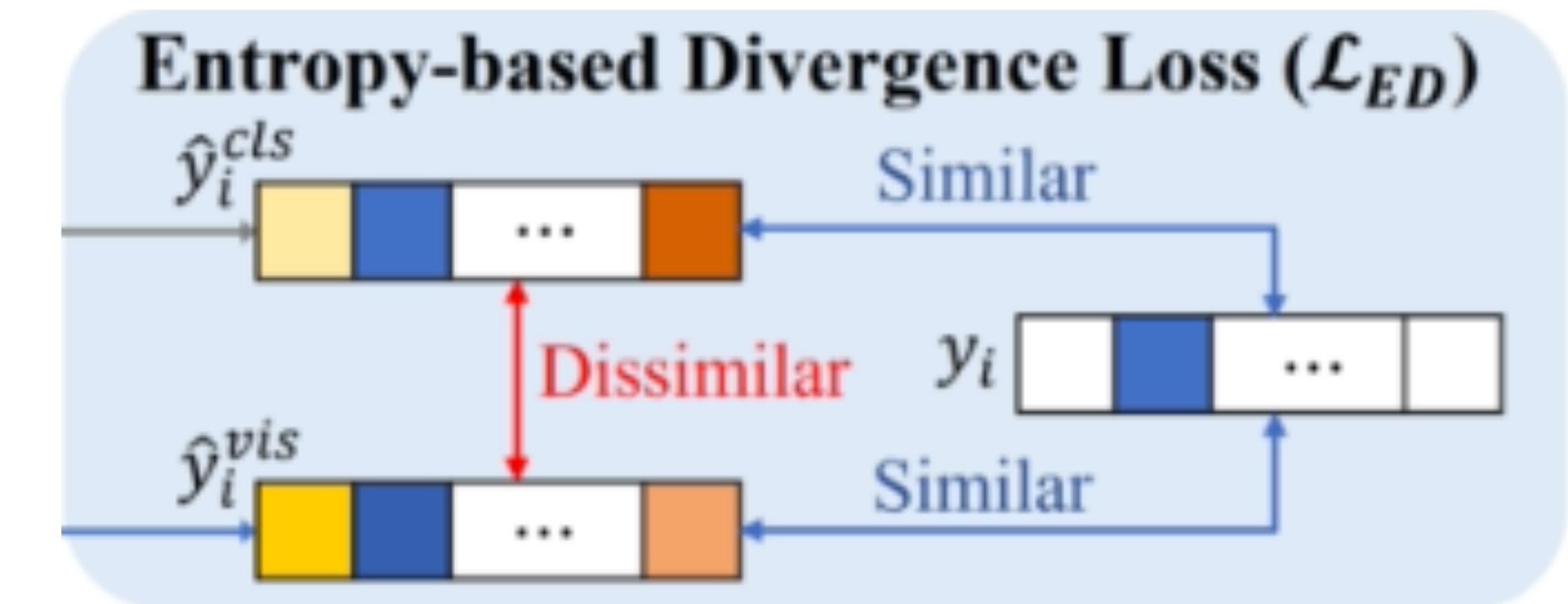
- Prompt Modulation 은 B Prompt의 느린 적용 속도 때문에 고안되었다.
- Head Specific Prompt 는 Attention Head 별로 추출된 정보들의 관계를 학습한다.
- Generic Prompt 는 보다 넓은 범위의 일반적인 특징을 학습하는데 중점을 둔다.



Method

Energy-based Divergence Loss

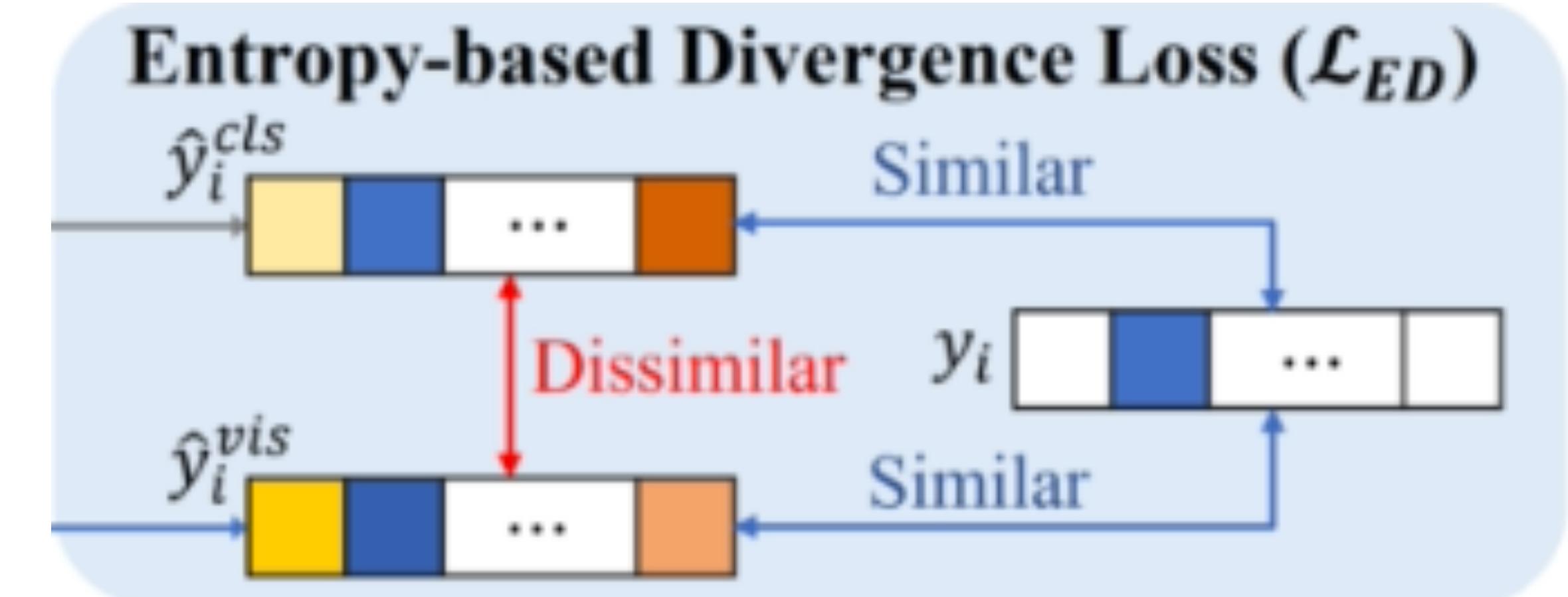
- Vision Token 은 CLS Token 과 같은 목표를 두고 학습하기 때문에 학습이 진행될수록 Output 이 비슷해지게 된다. 이는 Vision Token 의 효율적인 학습을 방해한다.
- 이를 방지하기 위해 PriViLege 는 Entropy-based Divergence Loss 라는 새로운 손실을 도입했다.



Method

Energy-based Divergence Loss

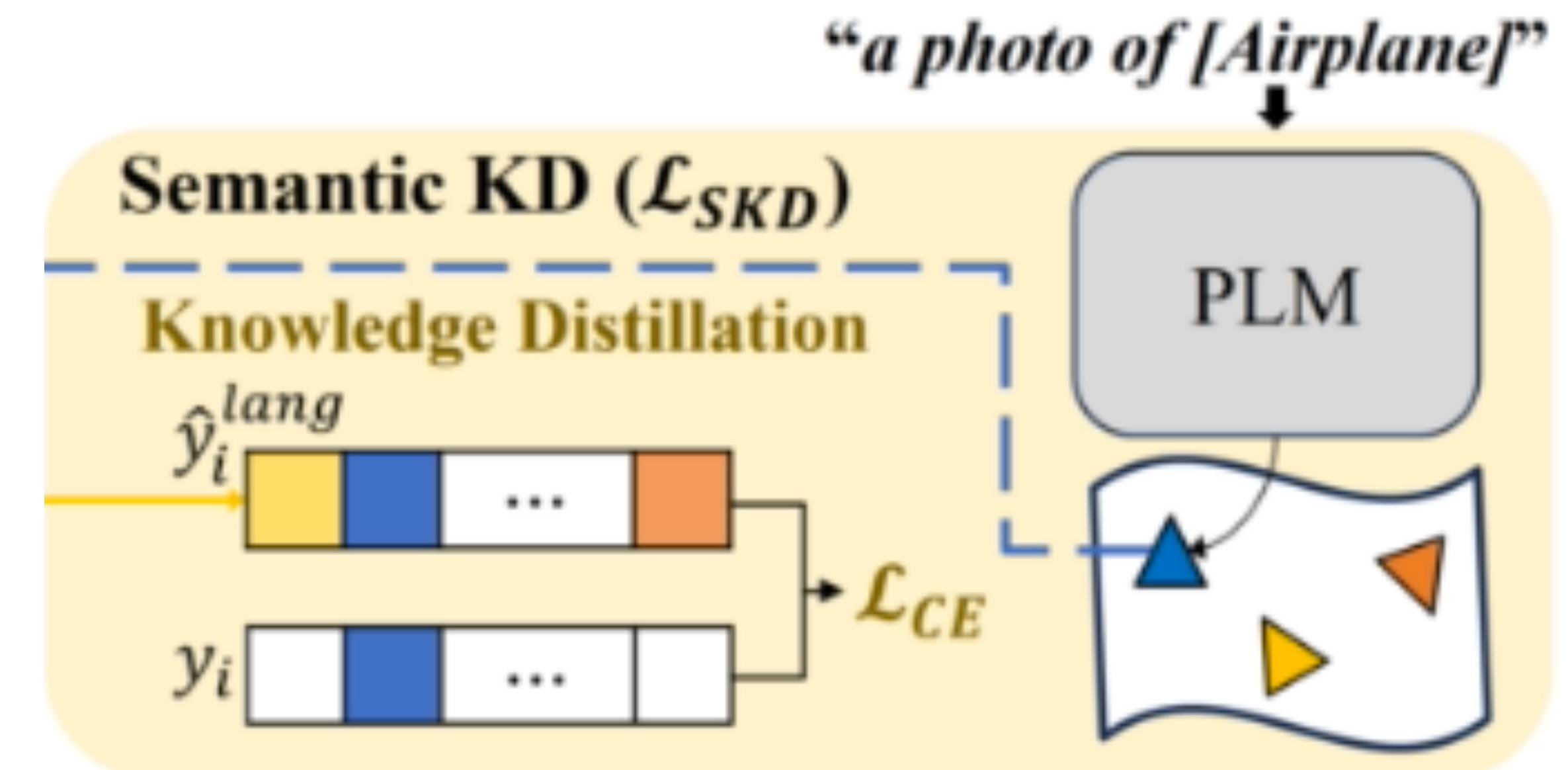
- ED Loss 는 KL Divergence 를 이용하여 계산하기 때문에 학습이 진행될수록 Vision Token 과 CLS Token 의 확률 분포만 비슷하게 만들 뿐 특징 공간 위에서의 위치는 유지된다.
- Vision Token 과 CLS Token 이 새로운 입력을 동일한 Class 로 분류할 확률은 극대화 되되, Token 이 고유 특징을 유지하도록 하여 Vision Token 의 개인 분류 능력을 키웠다.



Method

Semantic Knowledge Distillation Loss

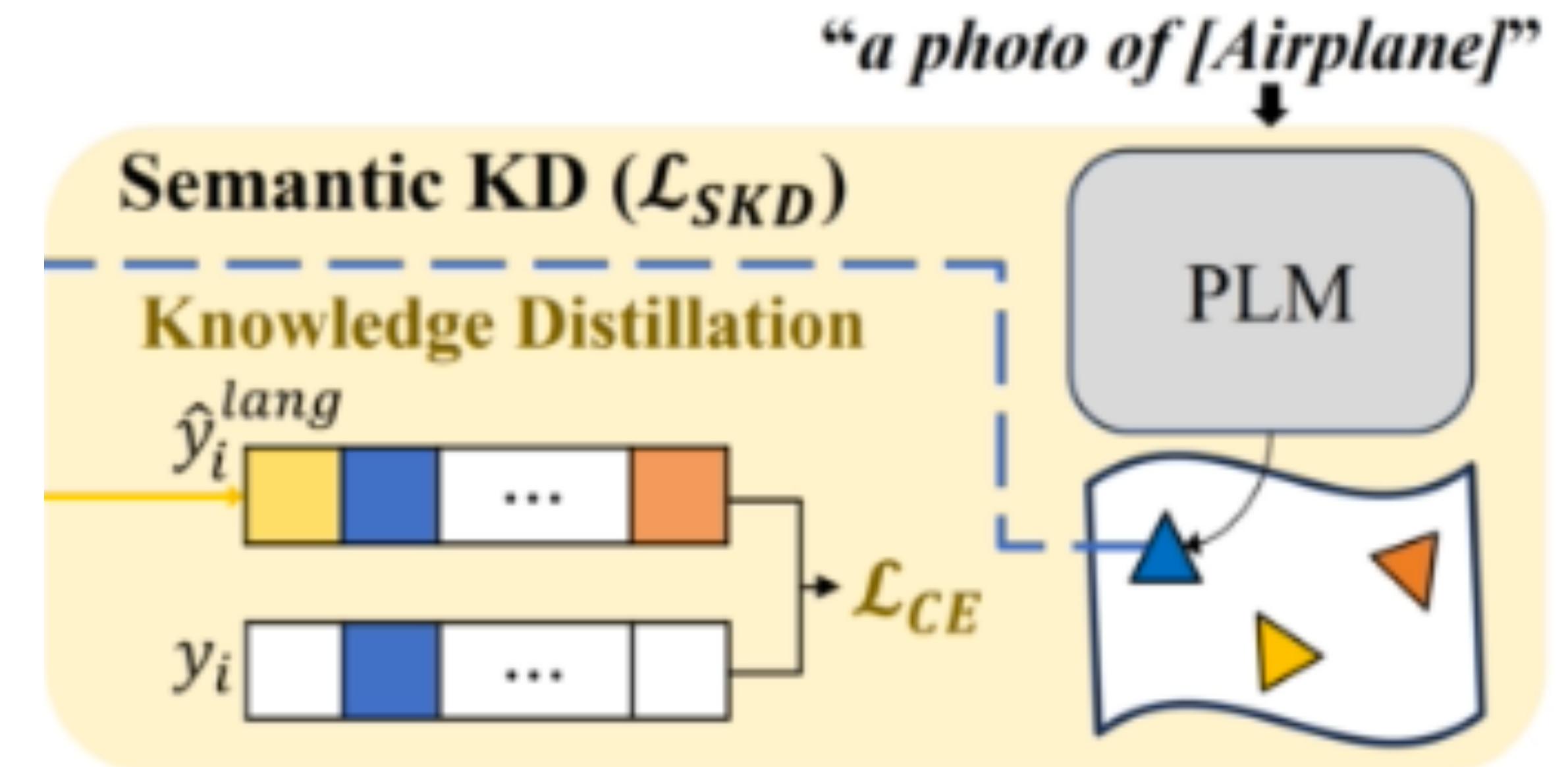
- Base Session에서의 정보가 중요하지만, 새로운 Class의 정확한 표현을 제한된 Data만으로 학습하기에는 무리가 있다.
- 새로운 Class를 잘 적용하기 위한 적절한 외부 정보를 제공하는 것이 중요하다.



Method

Semantic Knowledge Distillation Loss

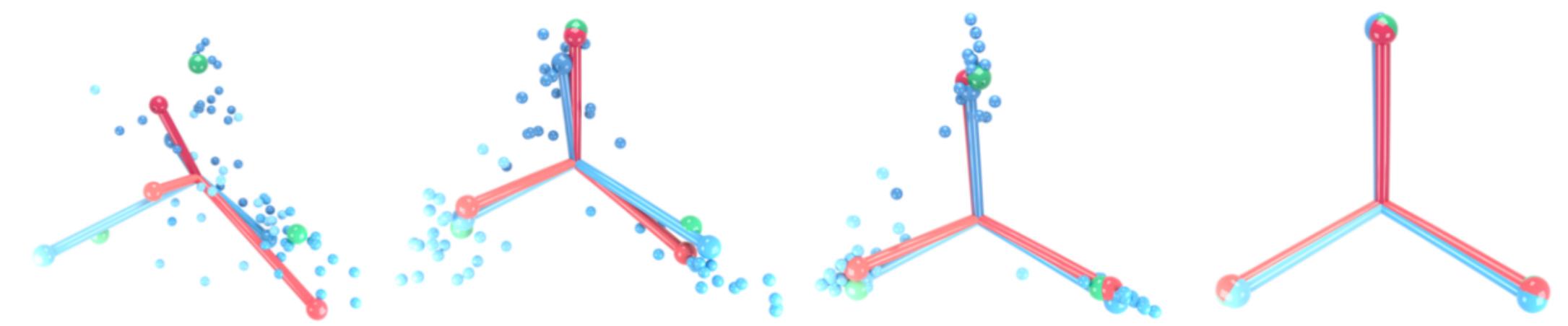
- KD Loss 는 Class 의 이름 cn^i 를 “a photo of cn^i ” 의 형태로 만든 뒤 Pre-trained BERT 를 이용하여 Class 이름에 대한 Text Feature 를 추출한 뒤 학습에 제공한다.
- 이는 Model 이 새로운 Class Data 에 빠르게 적응할 수 있도록 유도한다.



Method

Neural Collapse Loss

- 수학적으로, 충분히 학습이 진행되면 Prototype Vector 는 서로 동일한 간격을 두고 배치된다. 이 배치는 Class 간 내적이 $\frac{-1}{N-1}$ 이 되는 방향을 가리킨다 (여기서 N 은 Class 의 개수이다).

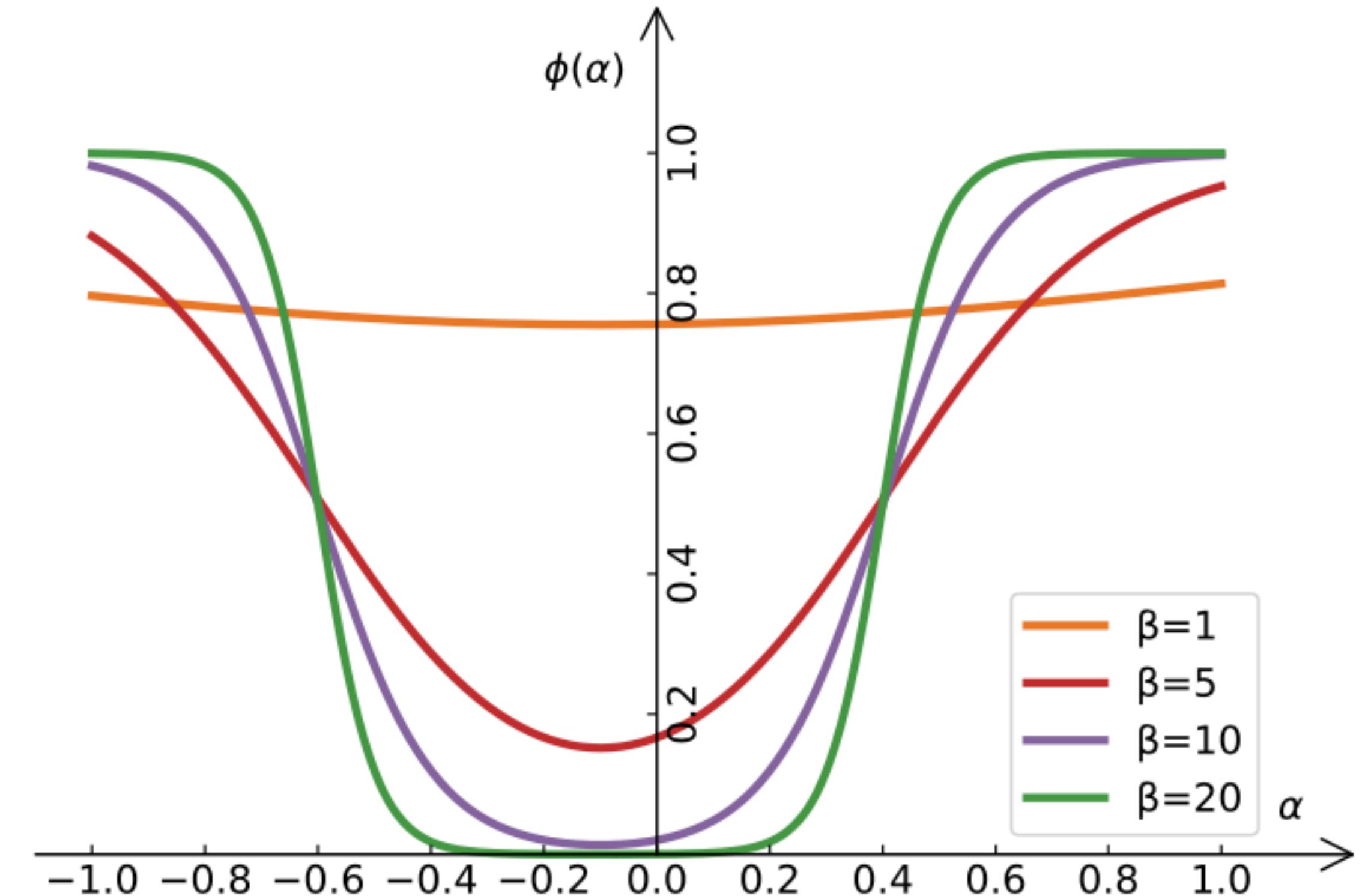


Method

Neural Collapse Loss

- NC-Softabs 활성 함수; 두 개의 Sigmoid 함수를 더한 형태로 구성되며, 훈련 과정에서 안정적인 성능을 발휘하도록 설계되었다.
- Class 의 개수가 10개일 때, Cosine 유사도 (α)가 $-\frac{1}{9}$ 에서 최솟값을 가진다.
- β 는 경사도로 값이 클수록 $\alpha = -\frac{1}{9}$ 에서 작은 값을 가진다.

$$\phi(\alpha) = \frac{1}{1 + e^{-\beta\left(\alpha + \frac{1}{N-1} - 0.5\right)}} + \frac{1}{1 + e^{-\beta\left(-\alpha - \frac{1}{N-1} - 0.5\right)}}$$



Method

Neural Collapse Loss

- NC Loss Function; 같은 Class 와의 유사도는 1, 다른 Class 와의 유사도는 $\frac{-1}{N-1}$ 에 가까워지도록 유도한다.
- 이러한 학습 방식을 통해 적은 Data로도 효율적인 학습을 할 수 있도록 돕는다.

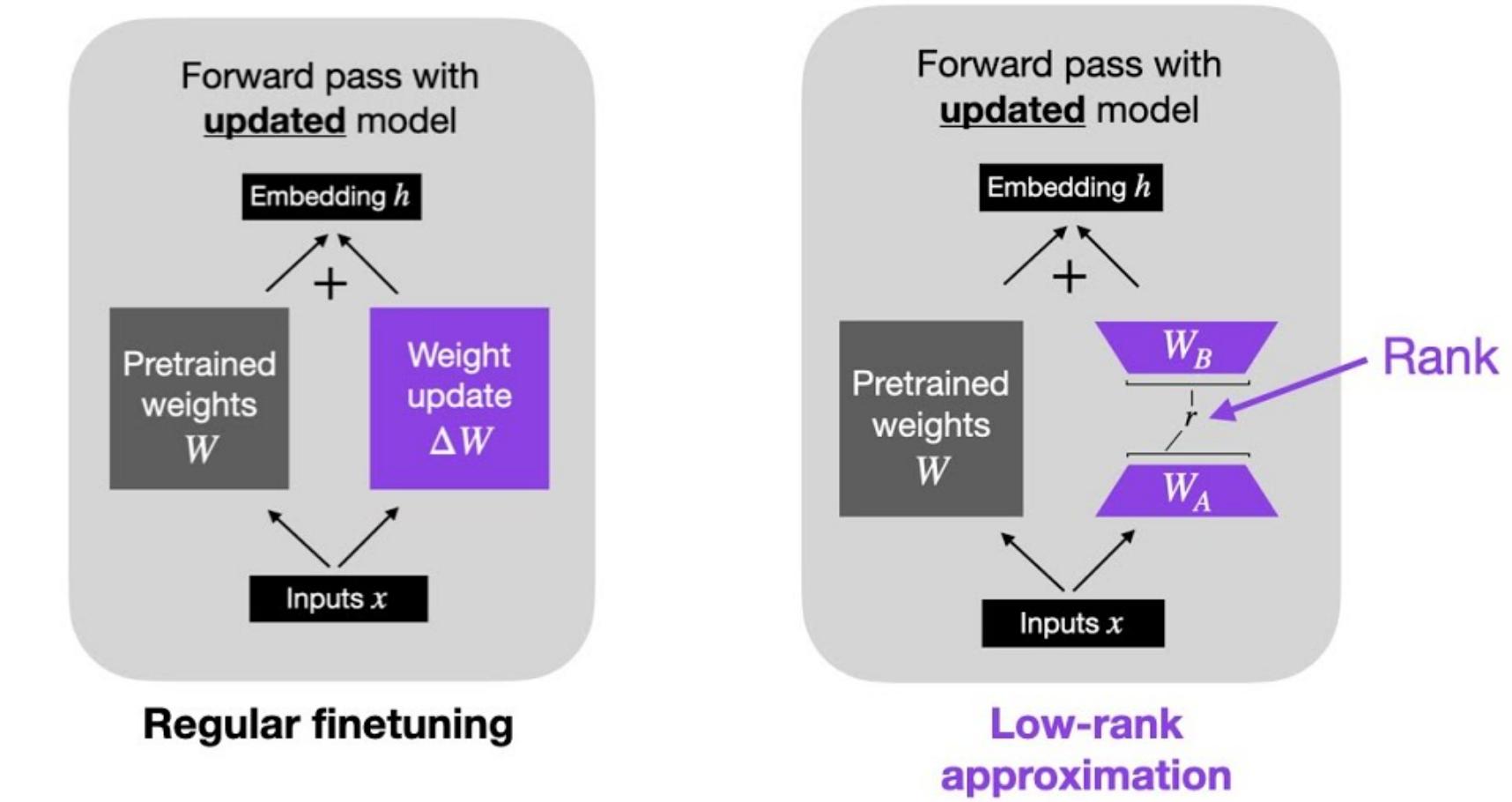
$$\mathcal{L}_{NC}(\mathbf{x}_i, y_i) = -\log \frac{e^{s(\cos(g_{\theta_g}(f_{\theta_f}(\mathbf{x}_i)), p_{y_i}))}}{e^{s(\cos(g_{\theta_g}(f_{\theta_f}(\mathbf{x}_i)), p_{y_i}))} + \sum_{j \neq y_i} e^{s(\phi(\cos(g_{\theta_g}(f_{\theta_f}(\mathbf{x}_i)), p_j)))}},$$

Method

LoRA

- 기존 Incremental Session 에서는 FC Layer 만 학습하게 되는데 FC Layer 에 LoRA 를 적용시켜 Base Session 에서는 원본 Layer 만 학습하고, Incremental Session 에서 LoRA Layer 만 학습하여 적은 수의 Parameter 만 학습함으로써 Catastrophic Forgetting 문제를 최대한 완화하게끔 유도한다.

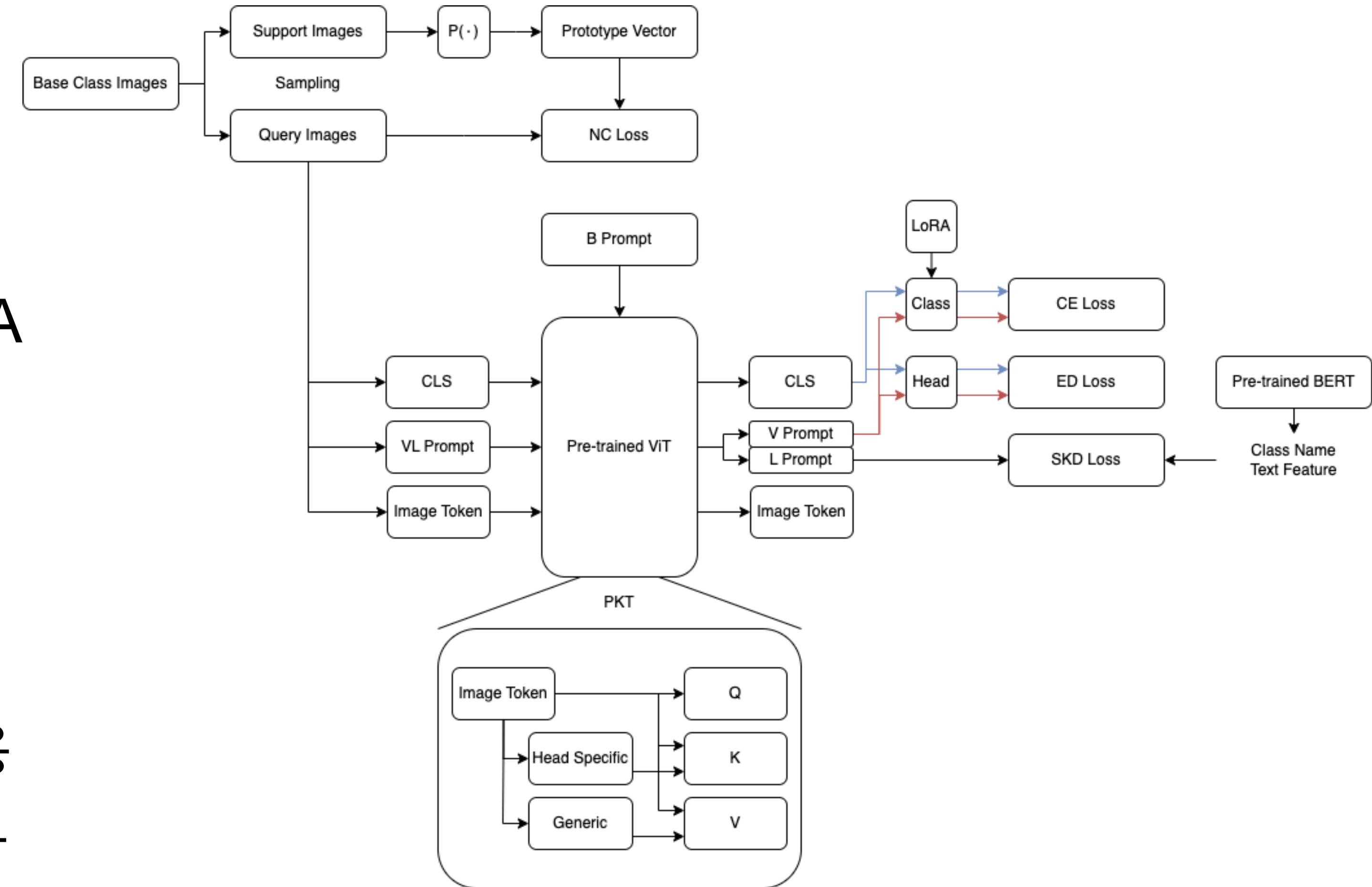
Explaining LoRA in a nutshell



Method

Model Architecture

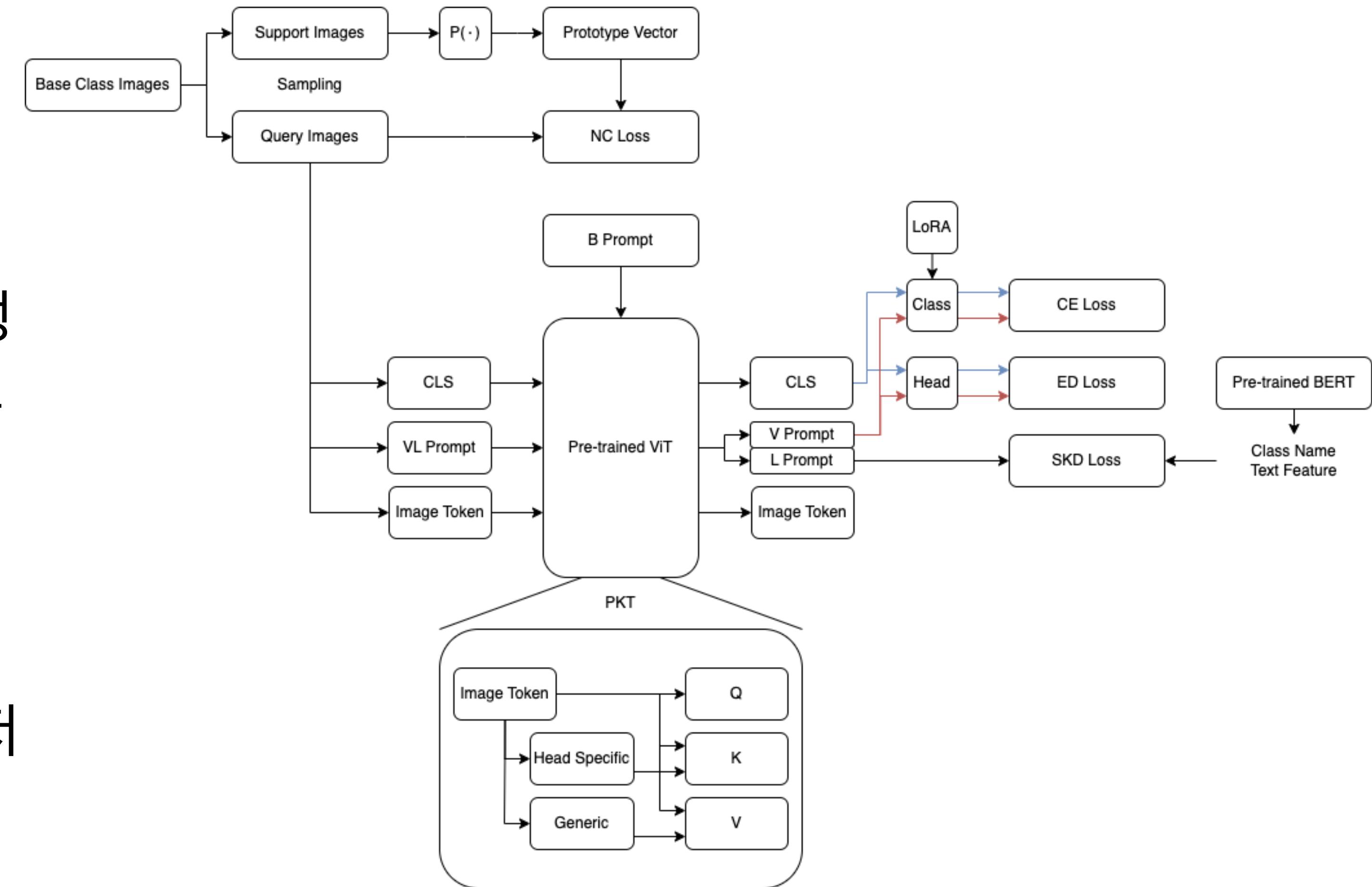
- PriViLege Model 의 구조를 기반으로, Meta-training, NC Loss, LoRA를 적용하여 Pre-trained Network를 이용한 FSCIL Scenario의 주요 도전 과제를 해결하도록 유도했다.
- Meta-training 은 적은 Data 를 이용함에도 Model 의 빠른 적응이 가능하게 한다.



Method

Model Architecture

- NC Loss 는 같은 Class 와의 유사성은 극대화하고 다른 Class 와의 유사성은 최소화하는 방향으로 학습하여 학습 효율을 늘린다.
- LoRA 를 FC Layer 에 적용하여 사전 학습된 기존의 가중치들은 고정하고, 저 차원 행렬을 학습하여 Catastrophic Forgetting 문제를 최대한 방지한다.

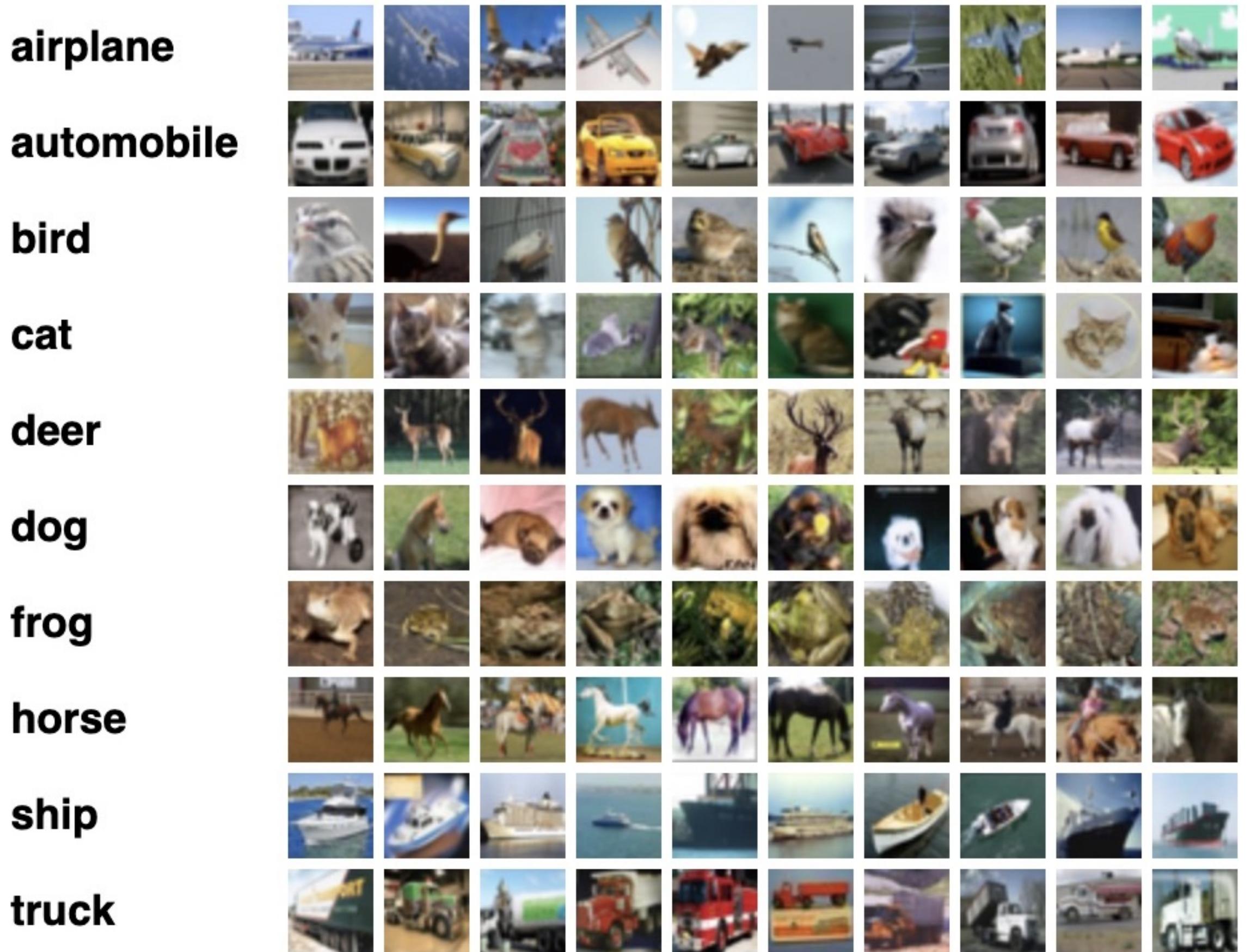


Experiments

Experiments

Experimental Settings

- Datasets and Metrics; 평가는 Cifar100 Dataset 을 이용했다.
Base Session 에서의 정확도 A_{Base} ,
Last Session 에서의 정확도 A_{Last} ,
모든 Session 에서의 정확도 평균
 A_{Avg} 를 평가했다.



Experiments

Experimental Settings

- Baselines and Implementation Details; Base Session 학습이 진행된 PriViLege Model, Base Session 학습이 진행되지 않은 PriViLege Model 을 Baseline 으로 설정하였다.
- 평등한 비교를 위해 Base Session 학습에 사용되는 이미지의 수는 Meta-training 에서 사용되는 이미지 수와 동일하게 제한하였다.
- 모든 Method 에서 Backbone Network 로 ImageNet-21K 로 Pre-trained 된 ViT-B/16을 사용하였다.

```
self.encoder = timm.create_model("vit_base_patch16_224", pretrained=True, num_classes=args.num_classes,  
                                drop_rate=0., drop_path_rate=0., drop_block_rate=None)
```

Experiments

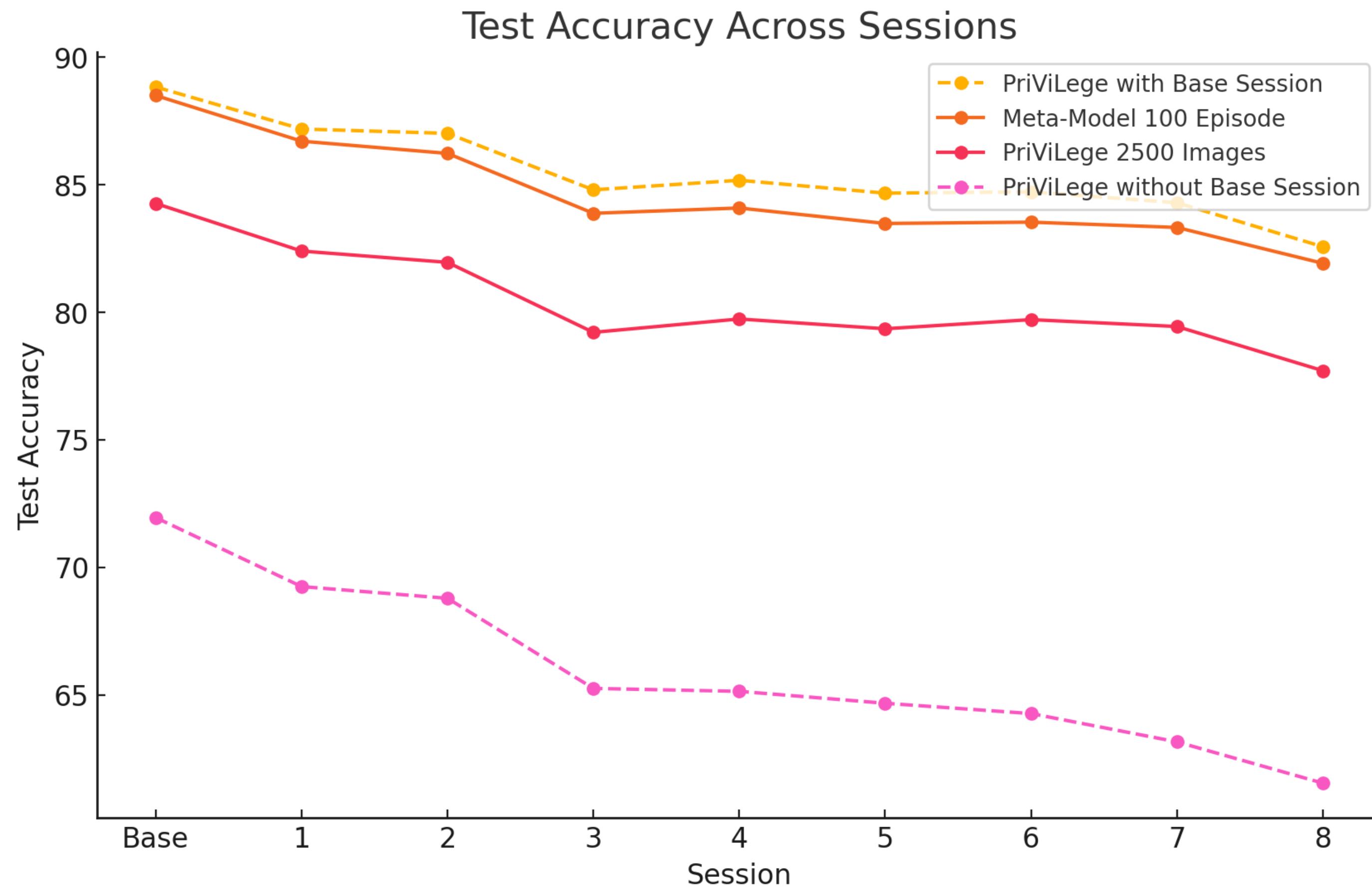
Experimental Settings

- Optimizer 는 SGD 로 학습률은 1e-4, Nesterov Momentum 을 적용하고 Momentum 은 0.9, Weight Decay 는 5e-4 로 설정하였다.
 - RTX 3090 GPU 를 이용하여 학습하였고, Baseline 의 Batch Size 는 128, Meta Model 의 경우 5 way 1 shot 4 query 로 100 Episode 학습했다.
 - Base Session 은 1개, Incremental Session 은 8개이며 Base Session 에서는 5 Epoch, Incremental Session 에서는 3 Epoch 학습한다.

Experiments

Main Experimental Results

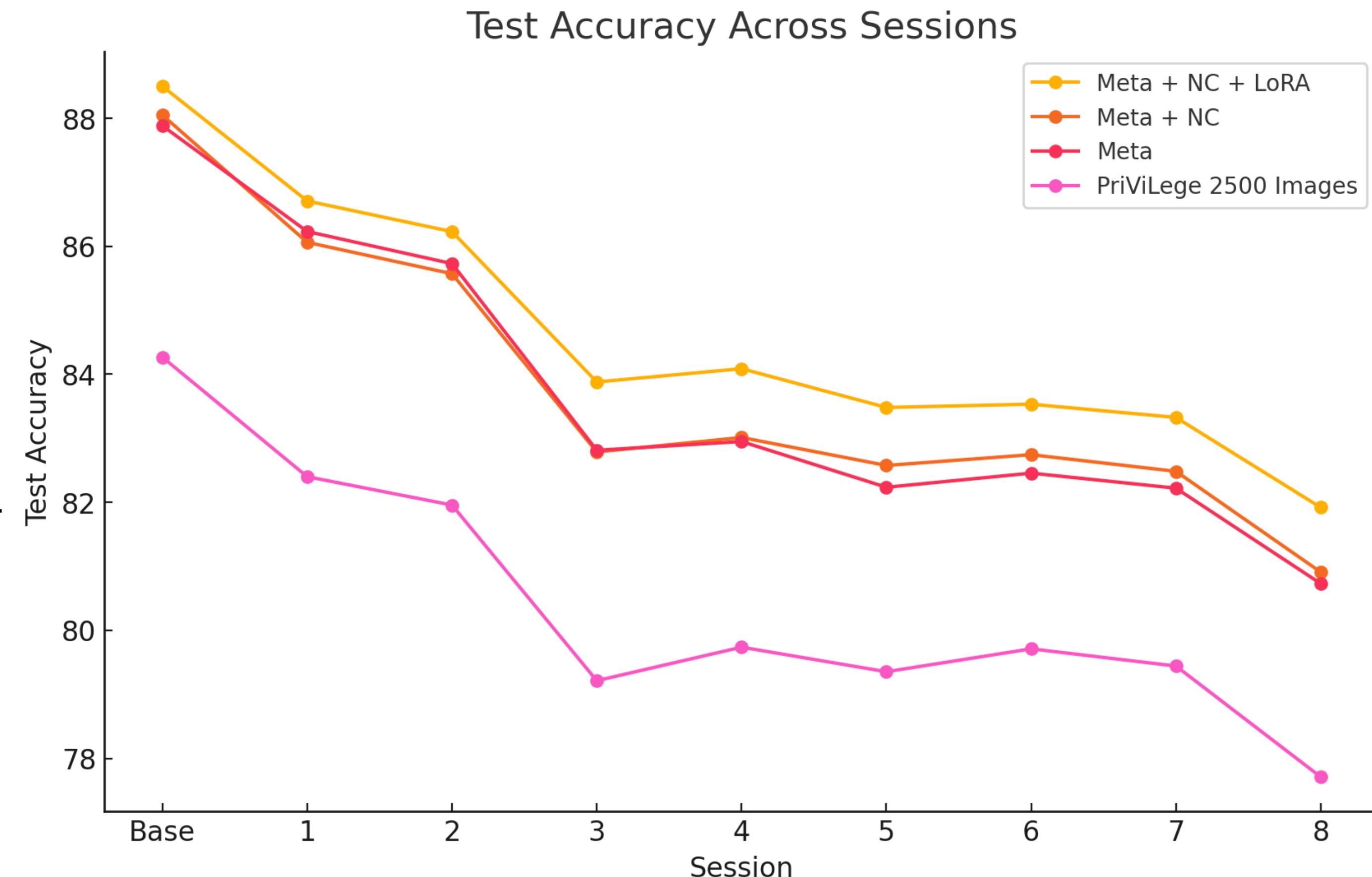
- 100 Episode의 경우 학습에 총 2500장의 Image, 전체 Dataset의 4.167%를 학습에 이용하게 된다.
- 모든 Session에서 Baseline 보다 좋은 성능을 보였으며 전체 Data의 4.167%만 이용하고도 원본 Model과 거의 비슷한 수준의 성능을 보였다.
- $A_{Base}, A_{Last}, A_{Avg}$ 의 차이는 각각 4.233, 4.21, 4.209로 같은 개수의 Image로 학습한 Baseline 보다 5.023%, 5.418%, 5.234% 만큼의 성능 향상을 보였다.



Experiments

Ablation Study

- 각각 Baseline, Meta, Meta + NC, Meta + NC + LoRA 를 적용하여 학습했을 때의 성능이다.
- 기존의 방식에서 Meta-training 을 적용했을 때의 성능 향상이 가장 뚜렷하게 드러났다.
- NC Loss 의 경우 Prototype Vector 를 만들기에 Support Set Data 가 부족하여 큰 성능 변화를 보이지 못했다고 분석된다.



Experiments

Ablation Study

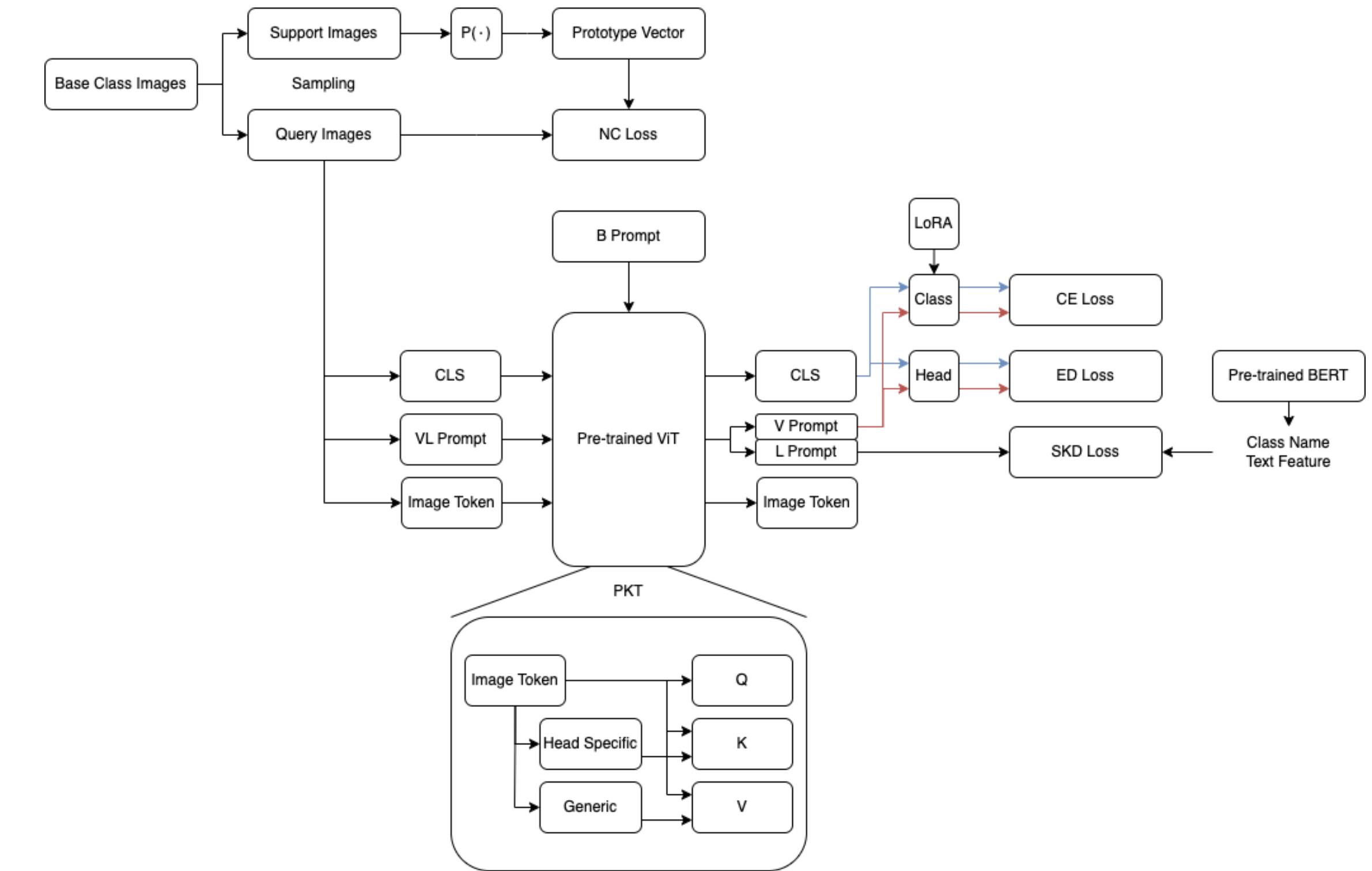
- Rank 가 너무 커지면 Parameter 수가 많아지기 때문에 Catastrophic Forgetting 의 영향이 커지고 너무 작으면 Parameter 수가 적어지기 때문에 새로운 Class 를 학습하기 힘들어지기 때문에 적절한 수치를 조정할 필요가 있다.
- 100 Episode 학습에서 FC Layer 에 적용한 LoRA 의 Rank 를 1, 2, 4, 8, 16 로 적용해보며 성능을 비교했다.
- 성능은 Rank 가 1일 때, 모든 부분에서 최고 성능을 보였다.

Rank	Base	Last	Avg
Baseline	84.267	77.71	80.421
1	<u>88.5</u>	<u>81.92</u>	<u>84.63</u>
2	87.25	80.25	82.961
4	86.85	79.53	82.295
8	78.517	65.04	70.982
16	88.1	81.41	84.095

Conclusion

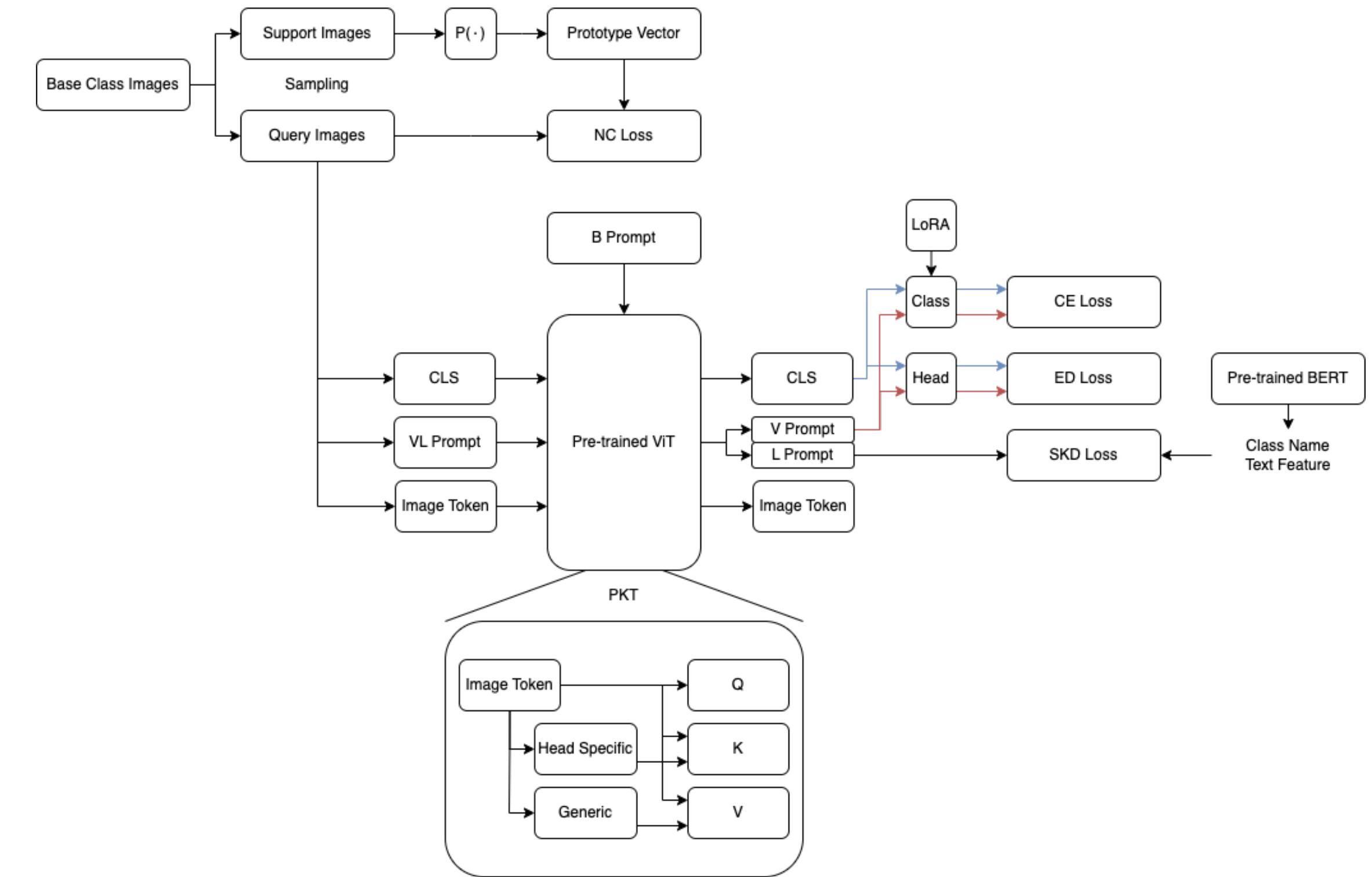
Conclusion

- 이 연구에서는 Base Session 학습에 사용되는 Data 를 제한하고도 좋은 성능을 낼 수 있는 방법에 대해 소개했다.
- 대규모 Dataset 이 필수적인 Fine-tuning 대신, 소규모의 Data 로도 빠른 학습이 가능한 Meta-training 과 NC Loss 를 이용하여 학습 성능을 향상시켰다.



Conclusion

- 추후에는 Base Session 을 완전히 제거하기 위해 새롭게 적용했던 방법을 Incremental Session 에 맞게 조정한 뒤 적용하는 방식으로 Base Session 에서의 학습 없이도 좋은 성능을 낼 수 있는 방법을 모색하면 좋을 것 같다.



Reference

- **Park, K.-H., Song, K., & Park, G.-M.** "Pre-trained Vision and Language Transformers Are Few-Shot Incremental Learners." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024
- **Ran, H., Li, W., Li, L., Tian, S., Ning, X., & Tiwari, P.** "Learning optimal inter-class margin adaptively for few-shot class-incremental learning via neural collapse-based meta-learning." *Information Processing and Management*, Vol. 61, No. 1, pp. 103664, 2024. <https://doi.org/10.1016/j.ipm.2024.103664>.