Test-Time Prompt Tuning in Vision-Language Models

DMQA Open Seminar (2025. 07. 04)

Data Mining & Quality Analytics Lab.

김지현

발표자 소개

- **김지현 (Jihyun Kim)**
  - 고려대학교 산업경영공학과 대학원 재학
  - Data Mining & Quality Analytics Lab. (김성범 교수님)
  - Ph.D. Candidate (2022.03 ~ Present)

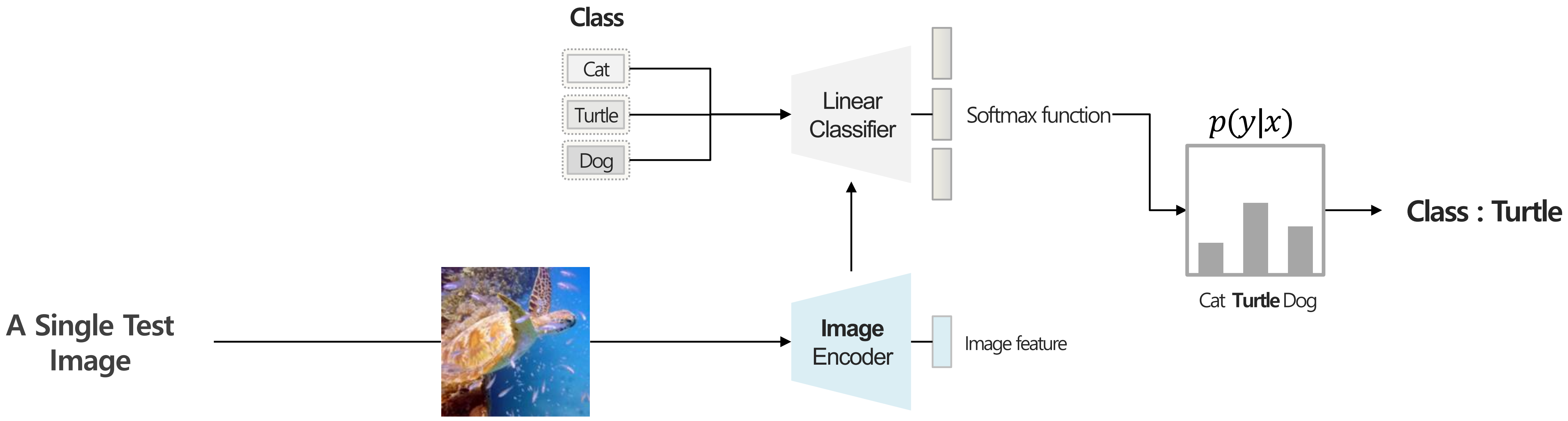
- **Research Interest**
  - Domain Adaptation

- **Contact**
  - jihyun_k@korea.ac.kr
  - +82-10-2172-8439

Introduction

Vision-Language Models

Classification with Vision Model

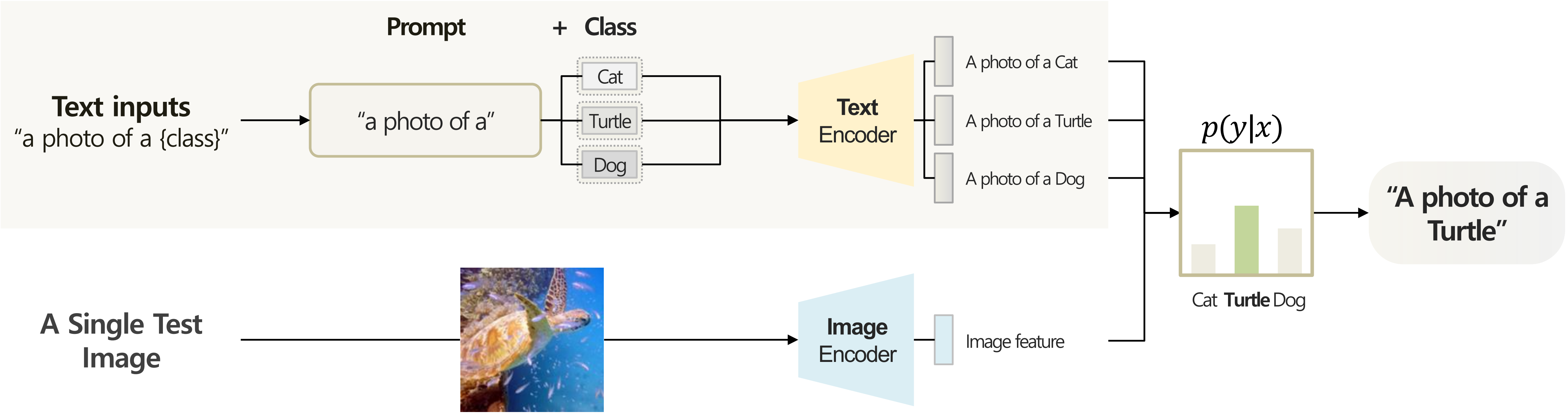


Introduction

Vision-Language Models

Vision-Language Model

Zero-Shot Classification with CLIP!

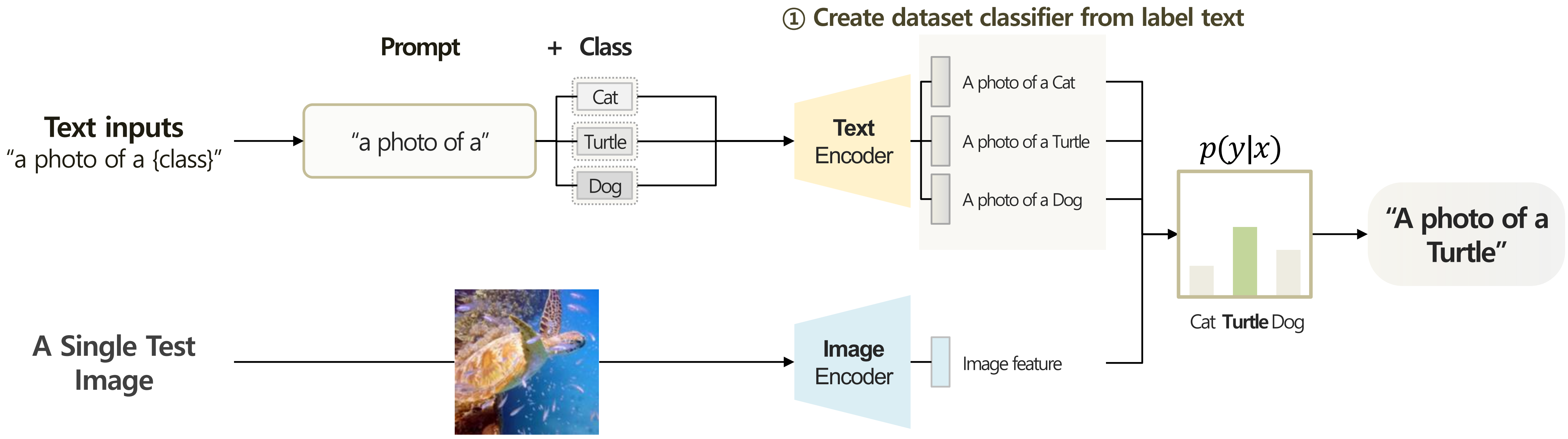


Introduction

Vision-Language Models

Vision-Language Model

Zero-Shot Classification with CLIP!

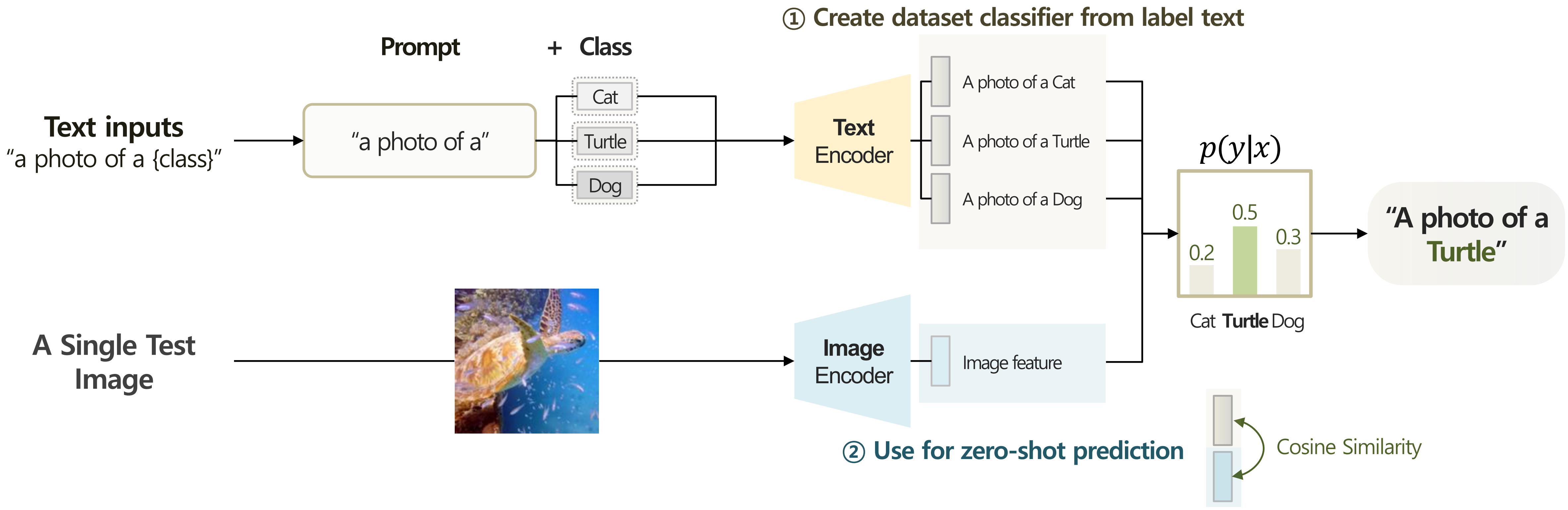


Introduction

Vision-Language Models

Vision-Language Model

Zero-Shot Classification with CLIP!

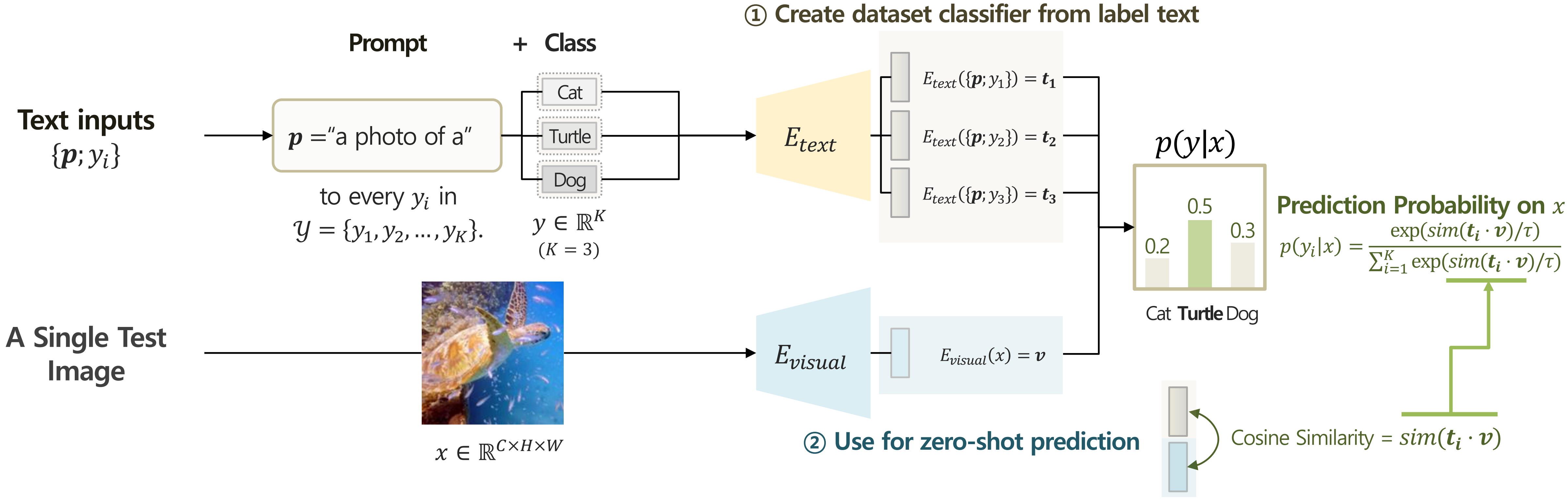


Introduction

Vision-Language Models

Vision-Language Model

Zero-Shot Classification with CLIP!

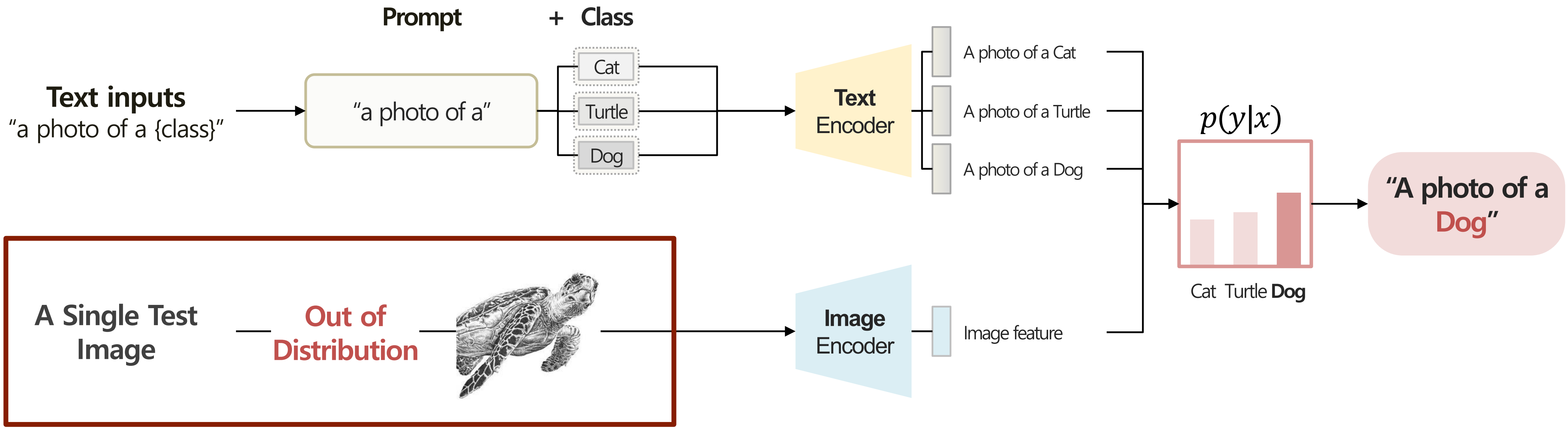


Introduction

Zero-Shot Generalization in Vision-Language Models

Distribution shift 상황에서 일반화 능력 저하[1,2]

Zero-Shot Classification with CLIP?



Dataset: ImageNet-Sketch

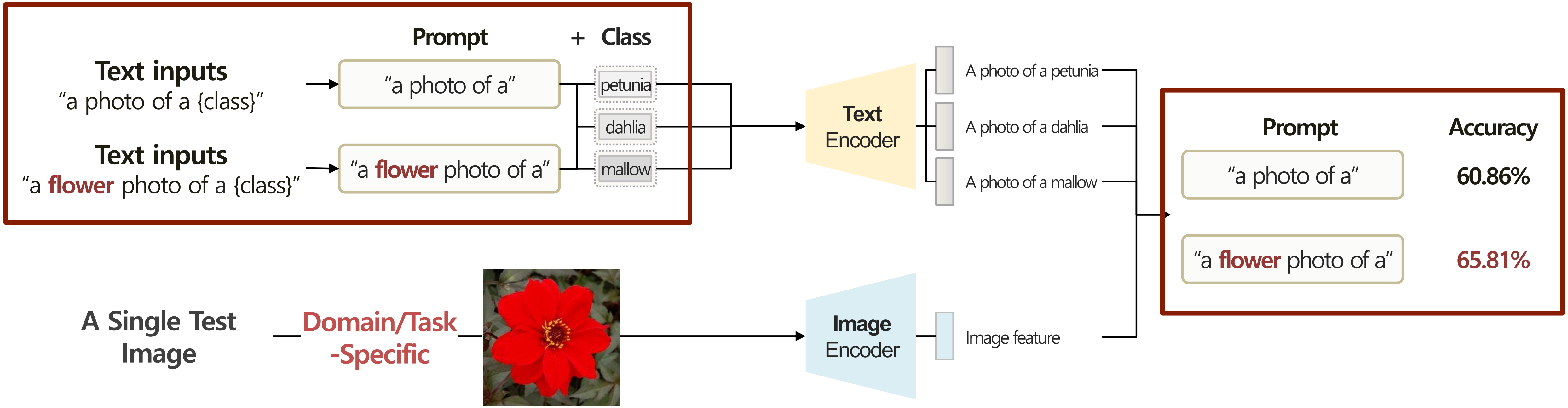
[1] Ma, X., Zhang, J., Guo, S., & Xu, W. (2023). Swapprompt: Test-time prompt adaptation for vision-language models. Advances in Neural Information Processing Systems, 36, 65252-65264.
[2] Abdul Samadh, J., Gani, M. H., Hussein, N., Khattak, M. U., Naseer, M. M., Shahbaz Khan, F., & Khan, S. H. (2023). Align your prompts: Test-time prompting with distribution alignment for zero-shot generalization. Advances in Neural Information Processing Systems, 36, 80396-80413.

Introduction

Zero-Shot Generalization in Vision-Language Models

Prompt에 의존적인 Downstream task 성능[3]

Zero-Shot Classification with CLIP?



Dataset: OxfordFlowers-102

[3] Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. International Journal of Computer Vision, 130(9), 2337-2348.

Introduction

Zero-Shot Generalization in Vision-Language Models

Prompt에 의존적인 Downstream task 성능[3]

Zero-Shot Classification with CLIP?

Prompt + Class

Text inputs
"a photo of {class}" → "a photo of"

Text inputs
"a photo of **a** {class}" → "a photo of **a**"

face
dolphin
watch

Text Encoder

A photo of a face
A photo of a dolphin
A photo of a watch

Prompt	Accuracy
"a photo of"	80.81%
"a photo of **a**"	86.29%

A Single Test Image — Domain/Task -Specific

Image Encoder

Image feature

Dataset: Caltech-101

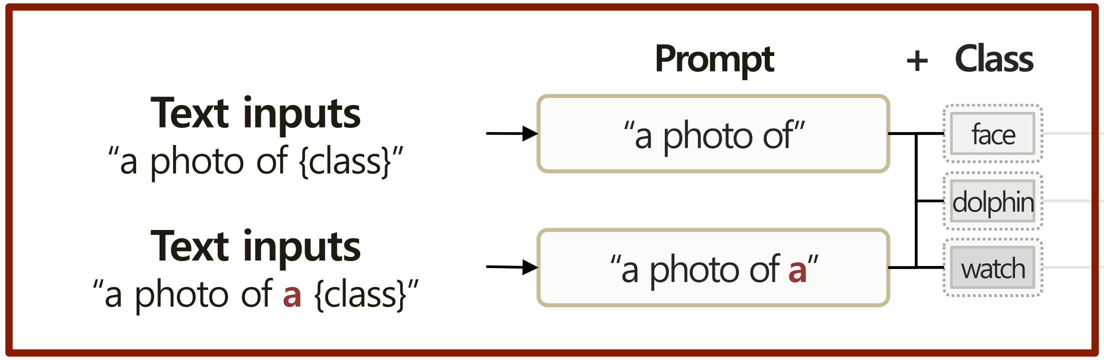
[3] Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. International Journal of Computer Vision, 130(9), 2337-2348.

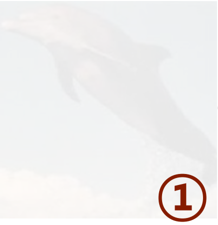
Introduction

Zero-Shot Generalization in Vision-Language Models

Prompt에 의존적인 Downstream task 성능[3]

Zero-Shot Classification with CLIP?





Dataset: Caltech-101

Prompt Engineering

Prompt	Accuracy
"a photo of"	80.81%
"a photo of **a**"	**86.29%**

Such hand-crafted prompts
① require domain-specific heuristics!
and ② may not be optimal!

[3] Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. International Journal of Computer Vision, 130(9), 2337-2348.

Prompt에 의존적인 Downstream task 성능[3]

Zero-Shot Classification with CLIP?

	Prompt	+ Class
Text inputs "a photo of {class}"	"a photo of"	face, dolphin, watch
Text inputs "a photo of **a** {class}"	"a photo of **a**"	

How About Automating Prompt Engineering

Prompt	Accuracy
"a photo of"	80.81%
"a photo of **a**"	86.29%

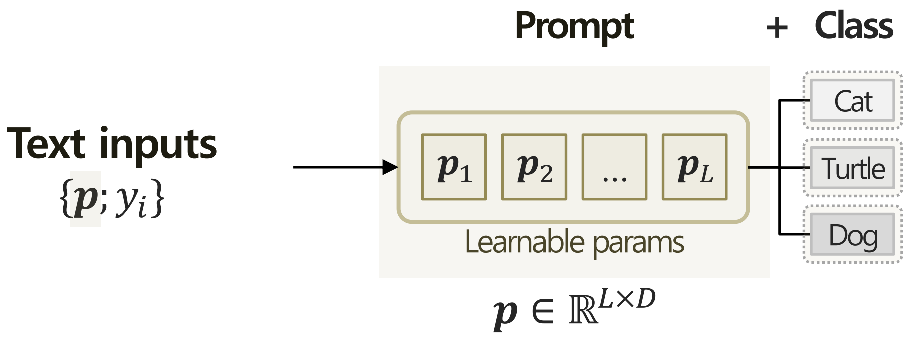
Such hand-crafted prompts
① require domain-specific heuristics!
and ② may not be optimal!

[3] Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. International Journal of Computer Vision, 130(9), 2337-2348.

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Prompt Tuning for CLIP!

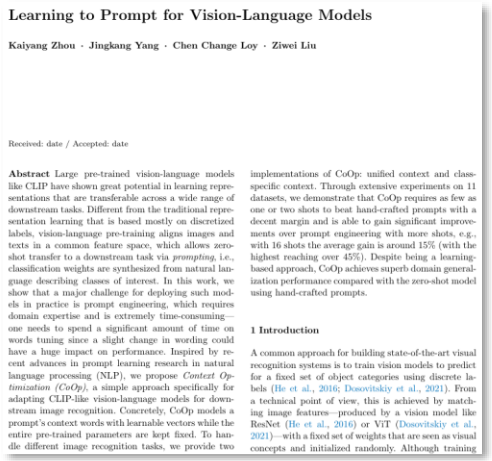


Automating Prompt Engineering

Prompt context token을 **learnable parameter**로 두어,
prompt 자체를 최적화하는 방법을 최초로 제안

CoOp (Context Optimization)

International Journal of Computer Vision, 2022,
Learning to prompt for vision-language models.
2025년 7월 기준 3205회 인용



Learning to Prompt for Vision-Language Models

Kaiyang Zhou · Jingkang Yang · Chen Change Loy · Ziwei Liu

Received: date / Accepted: date

Abstract Large pre-trained vision-language models like CLIP have shown great potential in learning representations that are transferable across a wide range of downstream tasks. Different from the traditional representation learning that is based mostly on discretized labels, vision-language pre-training aligns images and texts in a common feature space, which allows zero-shot transfer to a downstream task via *prompting*, i.e., classification weights are synthesized from natural language describing classes of interest. In this work, we show that a major challenge for deploying such models in practice is prompt engineering, which requires domain expertise and is extremely time-consuming—one needs to spend a significant amount of time on words tuning since a slight change in wording could have a huge impact on performance. Inspired by recent advances in prompt learning research in natural language processing (NLP), we propose *Context Optimization (CoOp)*, a simple approach specifically for adapting CLIP-like vision-language models for downstream image recognition. Concretely, CoOp models a prompt's context words with learnable vectors while the entire pre-trained parameters are kept fixed. To handle different image recognition tasks, we provide two implementations of CoOp: unified context and class-specific context. Through extensive experiments on 11 datasets, we demonstrate that CoOp requires as few as one or two shots to beat hand-crafted prompts with a decent margin and is able to gain significant improvements over prompt engineering with more shots, e.g., with 16 shots the average gain is around 15% (with the highest reaching over 45%). Despite being a learning-based approach, CoOp achieves superb domain generalization performance compared with the zero-shot model using hand-crafted prompts.

1 Introduction

A common approach for building state-of-the-art visual recognition systems is to train vision models to predict for a fixed set of object categories using discrete labels (He et al., 2016; Dosovitskiy et al., 2021). From a technical point of view, this is achieved by matching image features—produced by a vision model like ResNet (He et al., 2016) or ViT (Dosovitskiy et al., 2021)—with a fixed set of weights that are seen as visual concepts and initialized randomly. Although training

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Prompt Tuning for CLIP!

Prompt + Class

Text inputs
$\{\boldsymbol{p}; y_i\}$

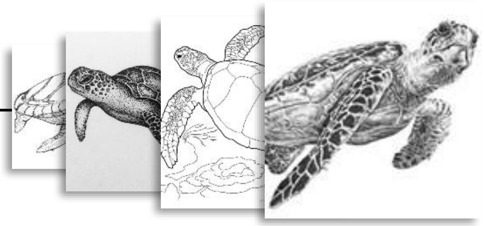
$\boldsymbol{p}_1$ $\boldsymbol{p}_2$ … $\boldsymbol{p}_L$

Learnable params

Cat
Turtle
Dog

$\boldsymbol{p} \in \mathbb{R}^{L \times D}$

Downstream Training Data
$\mathcal{D}_{train} = \{(x_i, y_i)\}$

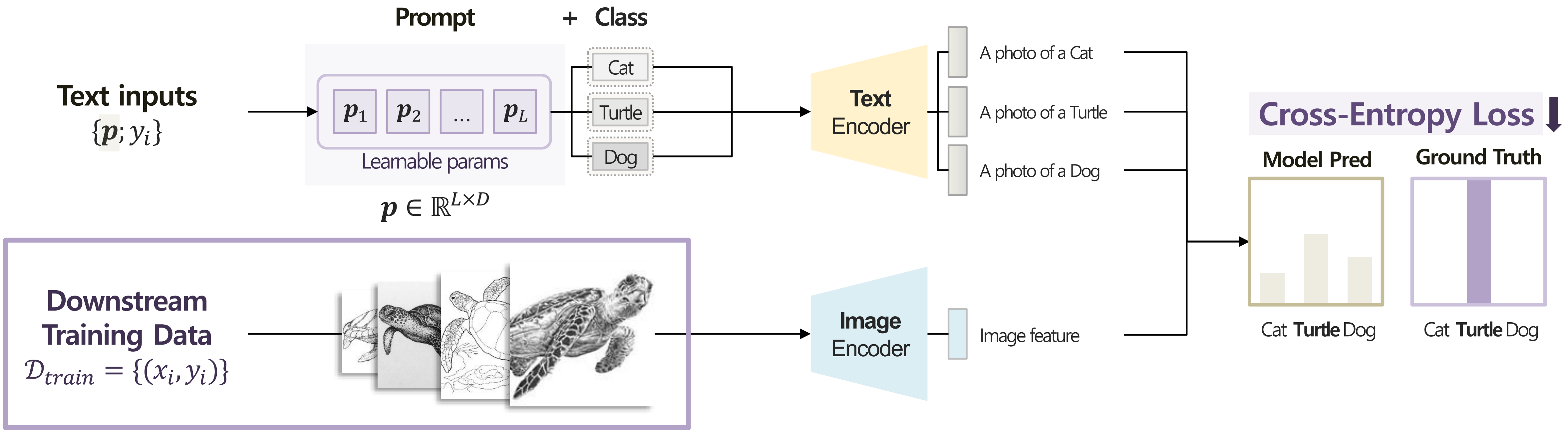


Dataset: ImageNet-Sketch

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

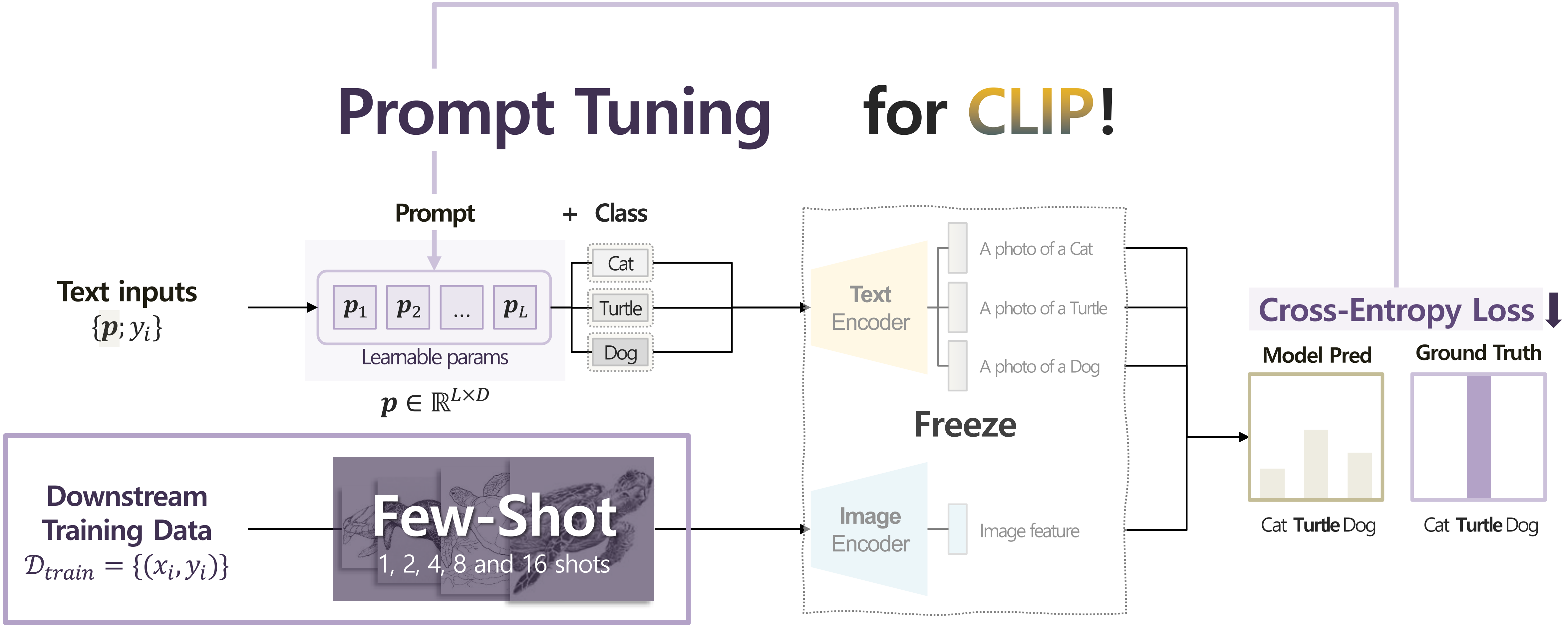
Prompt Tuning for CLIP!



Dataset: ImageNet-Sketch

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models



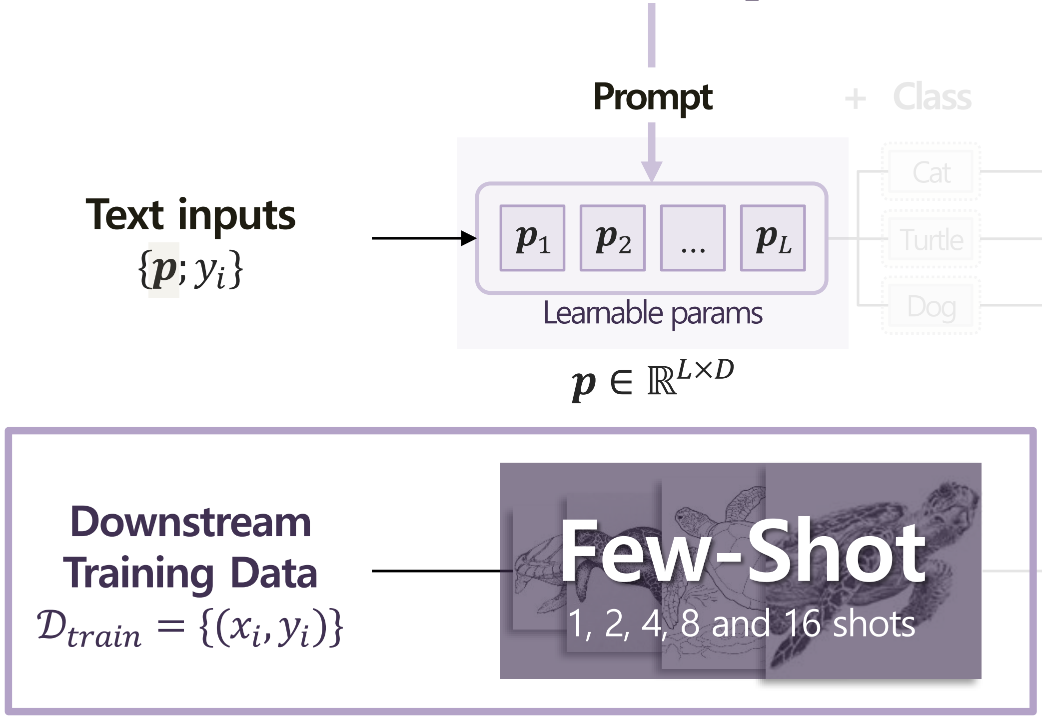
Dataset: ImageNet-Sketch

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Q1. Prompt engineering (handcrafted prompt) 대비 성능이 향상되는가?

Prompt Tuning for CLIP!



Dataset: ImageNet-Sketch

Q3. Training data와 Test data 간 분포 차이가 있을 때 일반화 성능은 어떠한가? (Generalization 성능 평가)

Q2. 적은 labeled data만으로도 성능 향상이 가능한가? (Few-shot learning capability)

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Q1. Prompt Engineering (handcrafted prompts) vs. CoOp (Prompt Tuning)

Experiments & Results

Baseline: Zero-shot CLIP with hand-crafted prompts (e.g., “a photo of a [CLASS]” as baseline.

Caltech101

Prompt	Accuracy
a [CLASS].	82.68
a photo of [CLASS].	80.81
a photo of a [CLASS].	86.29
$[V]_1$ $[V]_2$ ... $[V]_M$ [CLASS].	**91.83**

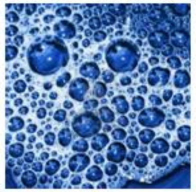
(a)

Flowers102

Prompt	Accuracy
a photo of a [CLASS].	60.86
a flower photo of a [CLASS].	65.81
a photo of a [CLASS], a type of flower.	66.14
$[V]_1$ $[V]_2$ ... $[V]_M$ [CLASS].	**94.51**

(b)

Describable Textures (DTD)



Prompt	Accuracy
a photo of a [CLASS].	39.83
a photo of a [CLASS] texture.	40.25
[CLASS] texture.	42.32
$[V]_1$ $[V]_2$ ... $[V]_M$ [CLASS].	**63.58**

(c)

EuroSAT

Prompt	Accuracy
a photo of a [CLASS].	24.17
a satellite photo of [CLASS].	37.46
a centered satellite photo of [CLASS].	37.56
$[V]_1$ $[V]_2$ ... $[V]_M$ [CLASS].	**83.53**

(d)

Fig. 1 Prompt engineering vs Context Optimization (CoOp). The former needs to use a held-out validation set for words tuning, which is inefficient; the latter automates the process and requires only a few labeled images for learning.

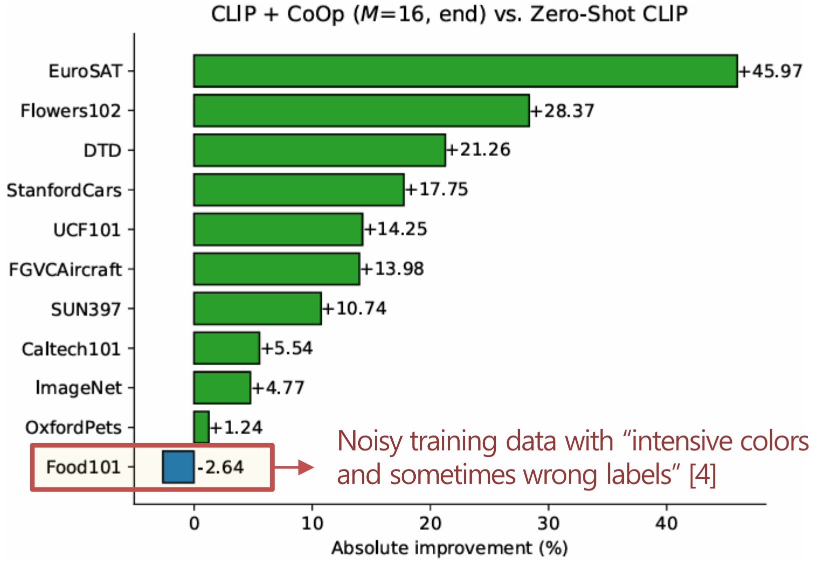


Fig. 4 Comparison with hand-crafted prompts.

[4] Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101–mining discriminative components with random forests. In Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13 (pp. 446-461). Springer International Publishing.

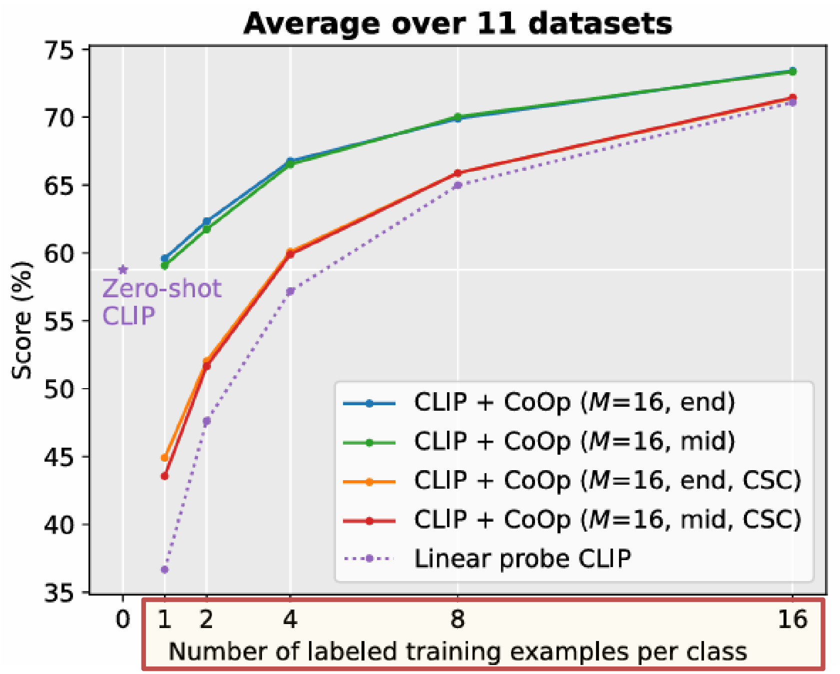
1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Q2. Few-shot learning performance, 적은 labeled data만으로도 성능 향상이 가능한가?

Experiments & Results

Setting: 1-shot, 2-shot, 4-shot, 8-shot and 16-shot setting



Zero-Shot CLIP

Linear Probe CLIP : Image Encoder (freeze) + Linear Classifier (fine-tuning)

M : CoOp's context length (본 세미나에서는 L로 표기)

end, mid : position of the class token
(end – $[\boldsymbol{p}_1][\boldsymbol{p}_2]\ldots[\boldsymbol{p}_L]$[CLASS] / mid - $[\boldsymbol{p}_1]\ldots[\boldsymbol{p}_{\frac{L}{2}}]$[CLASS]]$[\boldsymbol{p}_{\frac{L}{2}+1}]\ldots[\boldsymbol{p}_L]$)

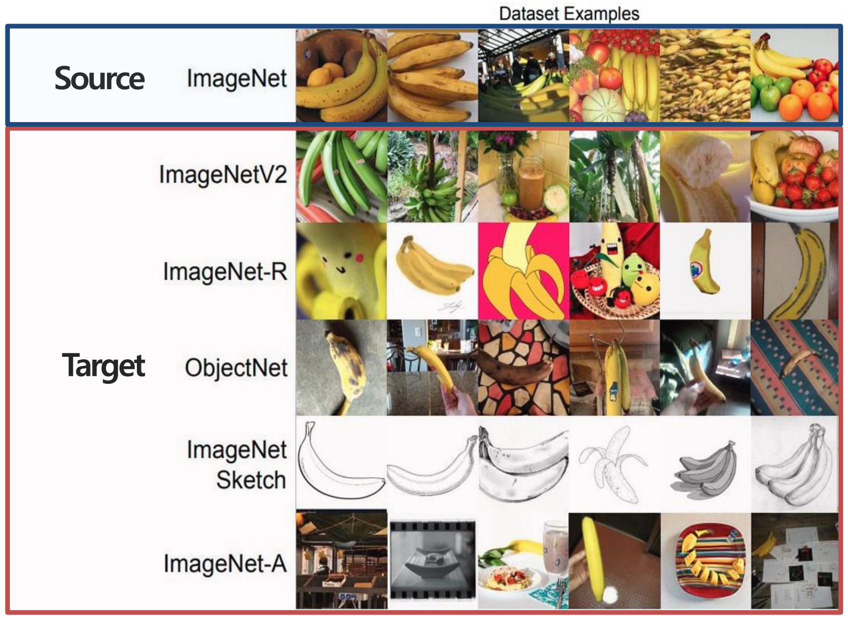
CSC : Class-specific token (label name 자체를 token으로 삽입 → 각 class마다 별도의 context token 학습)

→ Shot 수가 증가할 수록 성능 증가, 1-shot만으로도 zero-shot baseline 성능보다 우수

[4] Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101–mining discriminative components with random forests. In Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13 (pp. 446-461). Springer International Publishing.

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models



Q3. Domain shift 상황 하에서의 일반화 성능 검증 (Domain generalization)

Experiments & Results

Source : ImageNet | **Target** : ImageNet-V2, ImageNet-Sketch, ImageNet-A, ImageNet-R

	Source	Target			
Method	ImageNet	-V2	-Sketch	-A	-R
ResNet-50					
Zero-Shot CLIP	58.18	51.34	33.32	21.65	56.00
Linear Probe CLIP	55.87	45.97	19.07	12.74	34.86
CLIP + CoOp ($M=16$)	62.95	55.11	32.74	22.12	54.96
CLIP + CoOp ($M=4$)	**63.33**	**55.40**	**34.67**	**23.06**	**56.60**
ResNet-101					
Zero-Shot CLIP	61.62	54.81	38.71	28.05	64.38
Linear Probe CLIP	59.75	50.05	26.80	19.44	47.19
CLIP + CoOp ($M=16$)	**66.60**	**58.66**	39.08	28.89	63.00
CLIP + CoOp ($M=4$)	65.98	58.60	**40.40**	**29.60**	**64.98**
ViT-B/32					
Zero-Shot CLIP	62.05	54.79	40.82	29.57	**65.99**
Linear Probe CLIP	59.58	49.73	28.06	19.67	47.20
CLIP + CoOp ($M=16$)	**66.85**	58.08	40.44	30.62	64.45
CLIP + CoOp ($M=4$)	66.34	**58.24**	**41.48**	**31.34**	65.78
ViT-B/16					
Zero-Shot CLIP	66.73	60.83	46.15	47.77	73.96
Linear Probe CLIP	65.85	56.26	34.77	35.68	58.43
CLIP + CoOp ($M=16$)	**71.92**	64.18	46.71	48.41	74.32
CLIP + CoOp ($M=4$)	71.73	**64.56**	**47.89**	**49.93**	**75.14**

ResNet-50					
Zero-Shot CLIP	58.18	51.34	33.32	21.65	56.00
Linear Probe CLIP	55.87	45.97	19.07	12.74	34.86
CLIP + CoOp ($M=16$)	62.95	55.11	32.74	22.12	54.96
CLIP + CoOp ($M=4$)	**63.33**	**55.40**	**34.67**	**23.06**	**56.60**

Zero-Shot CLIP : not tied to a specific data distribution

Linear Probe CLIP : Image Encoder (freeze) + Linear Classifier (fine-tuning)

M : CoOp's context length

[4] Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101–mining discriminative components with random forests. In Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13 (pp. 446-461). Springer International Publishing.

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Prompt 학습이 handcrafted prompt를 사용하는 것보다 우수한 성능

Prompt Tuning for CLIP!

Prompt

Text inputs $\{\boldsymbol{p}; y_i\}$ → $\boldsymbol{p}_1$ $\boldsymbol{p}_2$ … $\boldsymbol{p}_L$

Learnable params

 ≠ 

Domain Shift

Training data와 Test data 간 분포 차이가 있을 때 일반화 성능도 zero-shot baseline보다 강인함

Downstream Training Data $\mathcal{D}_{train} = \{(x_i, y_i)\}$

Dataset: ImageNet-Sketch

Shot 수가 증가할 수록 성능이 증가하지만, 1-shot만으로도 zero-shot baseline보다 우수한 성능

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Prompt Tuning

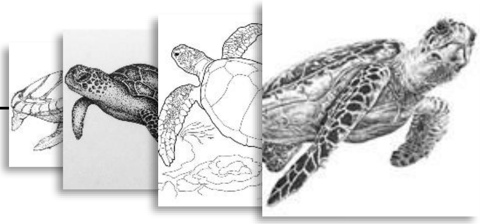
$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \mathbb{E}_{(x,y)\sim\mathcal{D}_{train}} \mathcal{L}(sim(E_{text}(\boldsymbol{p};\mathcal{Y}), E_{visual}(x)), y)$$ → Supervised loss

Text inputs
$\{\boldsymbol{p}; y_i\}$
$y_i \in \mathcal{Y}$

$\boldsymbol{p}_1$ $\boldsymbol{p}_2$ … $\boldsymbol{p}_L$

Learnable params

Downstream Training Data
$\mathcal{D}_{train} = \{(x_i, y_i)\}$



Dataset: ImageNet

Requires Downstream Training Data!

① Learned prompts are **limited to the distribution** and **tasks corresponding to training data** and may have **limited generalization** beyond that!

② Require **annotations which is expensive** and **NOT available for ZERO-SHOT** tasks!

1. CoOp (Context Optimization)

[IJCV 2022] Learning to Prompt for Vision-Language Models

Prompt Tuning

$$p^* = \arg\min_{p} \mathbb{E}_{(x,y)\sim\mathcal{D}_{train}} \mathcal{L}(sim(E_{text}(p;\mathcal{Y}), E_{visual}(x)), y)$$

Text inputs

$\{p; y_i\}$

$y_i \in \mathcal{Y}$

p_1 p_2 … p_L

Learnable params

A Single Test Image

Test Sample # 1

x_{test}

How about **tuning the prompt** on the fly **using only the given test sample**?

Requires **Downstream Training Data!**

① Learned prompts are **limited to the distribution** and **tasks corresponding to training data** and may have **limited generalization** beyond that!

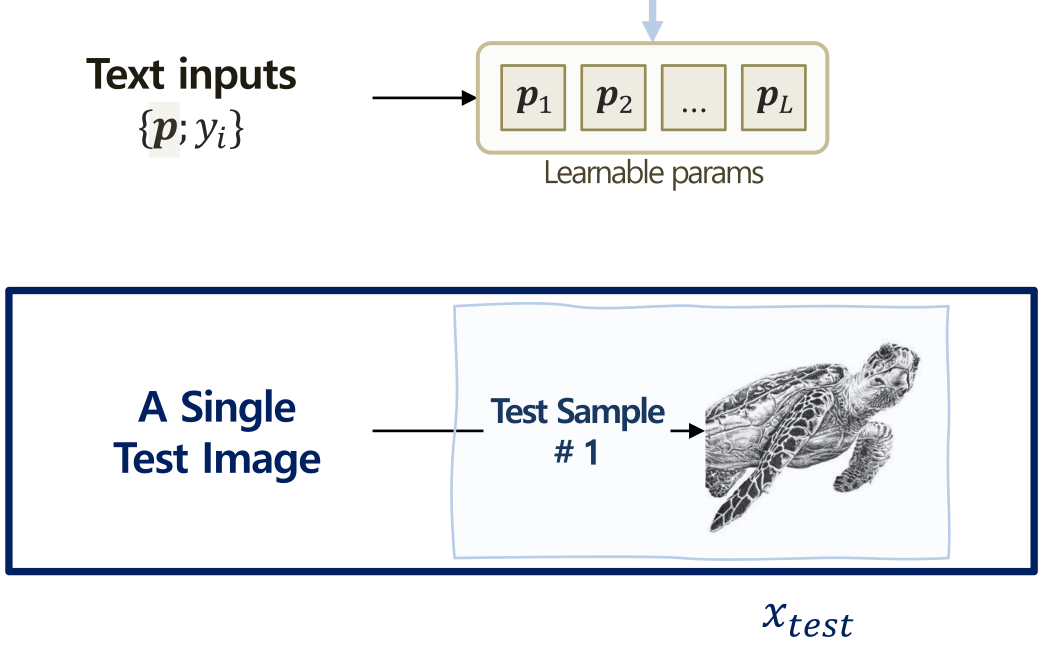
② Require **annotations which is expensive** and **NOT available for ZERO-SHOT** tasks!

2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Test-Time Prompt Tuning

$$p^* = \arg\min_{p} \mathcal{L}\left(\{E_{visual}, E_{text}\}, p, x_{test}\right)$$

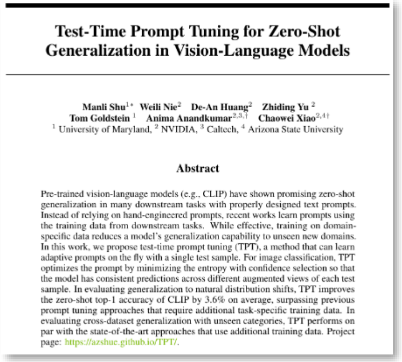


How about **tuning the prompt** on the fly **using only the given test sample**?

Zero-shot generalization을 위해, **unlabeled single test sample만으로 prompt tuning**하는 방법을 최초로 제안

TPT (Test-Time Prompt Tuning)

NeurIPS, 2022,
Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models
2025년 7월 기준 407회 인용



Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Manli Shu[1*] Weili Nie[2] De-An Huang[2] Zhiding Yu[2]
Tom Goldstein[1] Anima Anandkumar[2,3,†] Chaowei Xiao[2,4†]
[1] University of Maryland, [2] NVIDIA, [3] Caltech, [4] Arizona State University

Abstract

Pre-trained vision-language models (e.g., CLIP) have shown promising zero-shot generalization in many downstream tasks with properly designed text prompts. Instead of relying on hand-engineered prompts, recent works learn prompts using the training data from downstream tasks. While effective, training on domain-specific data reduces a model's generalization capability to unseen new domains. In this work, we propose test-time prompt tuning (TPT), a method that can learn adaptive prompts on the fly with a single test sample. For image classification, TPT optimizes the prompt by minimizing the entropy with confidence selection so that the model has consistent predictions across different augmented views of each test sample. In evaluating generalization to natural distribution shifts, TPT improves the zero-shot top-1 accuracy of CLIP by 3.6% on average, surpassing previous prompt tuning approaches that require additional task-specific training data. In evaluating cross-dataset generalization with unseen categories, TPT performs on par with the state-of-the-art approaches that use additional training data. Project page: https://azshue.github.io/TPT/.

2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Test-Time Prompt Tuning

$$p^* = \arg\min_{p} \mathcal{L}\left(\{E_{visual}, E_{text}\}, \boldsymbol{p}, x_{test}\right)$$

Text inputs
$\{\boldsymbol{p}; y_i\}$

$\boldsymbol{p}_1$ $\boldsymbol{p}_2$ … $\boldsymbol{p}_L$

Learnable params

Cat
Turtle
Dog

A Single Test Image

Test Sample # 1

x_{test}

Must select an **unsupervised loss for prompt tuning**

Unsupervised Loss

2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Test-Time Prompt Tuning

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \mathcal{L}\left(\{E_{visual}, E_{text}\}, \boldsymbol{p}, x_{test}\right)$$

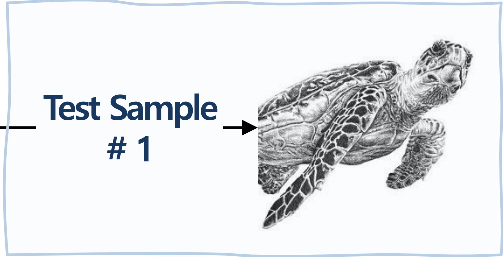
Text inputs
$\{\boldsymbol{p}; y_i\}$

$\boldsymbol{p}_1$ $\boldsymbol{p}_2$ … $\boldsymbol{p}_L$

Learnable params

Cat
Turtle
Dog

A Single Test Image



x_{test}

Promote the consistency of the model's predictions

Consistency Regularization

2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

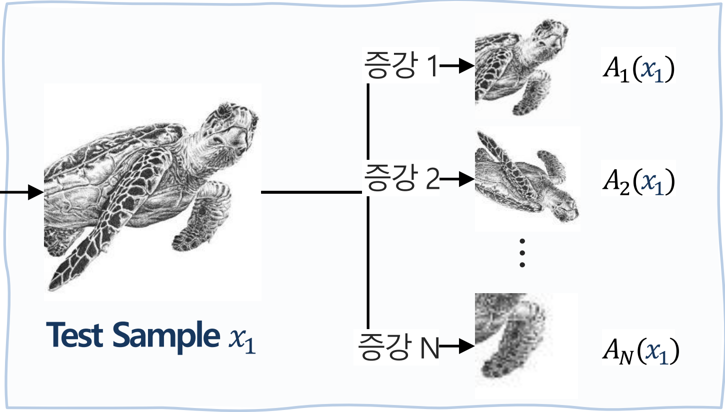
Test-Time Prompt Tuning

$$p^* = \arg\min_{p} \mathcal{L}\left(\{E_{visual}, E_{text}\}, p, x_{test}\right)$$

Text inputs $\{p; y_i\}$ → p_1 p_2 … p_L → Cat / Turtle / Dog

Learnable params

A Single Test Image →



Promote the consistency of the model's predictions

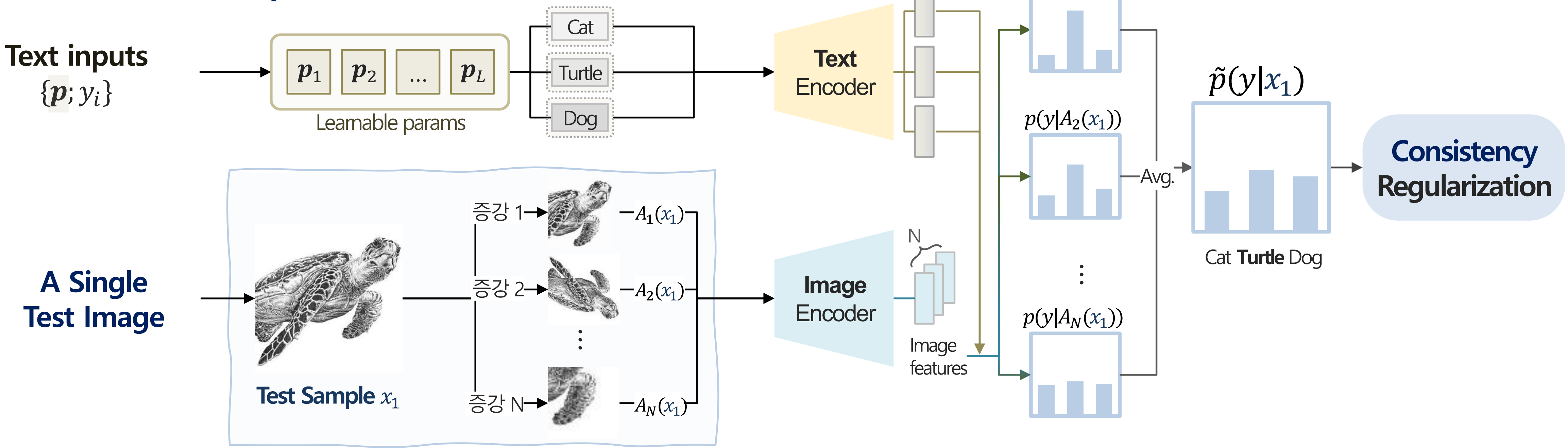
Consistency Regularization

2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Test-Time Prompt Tuning

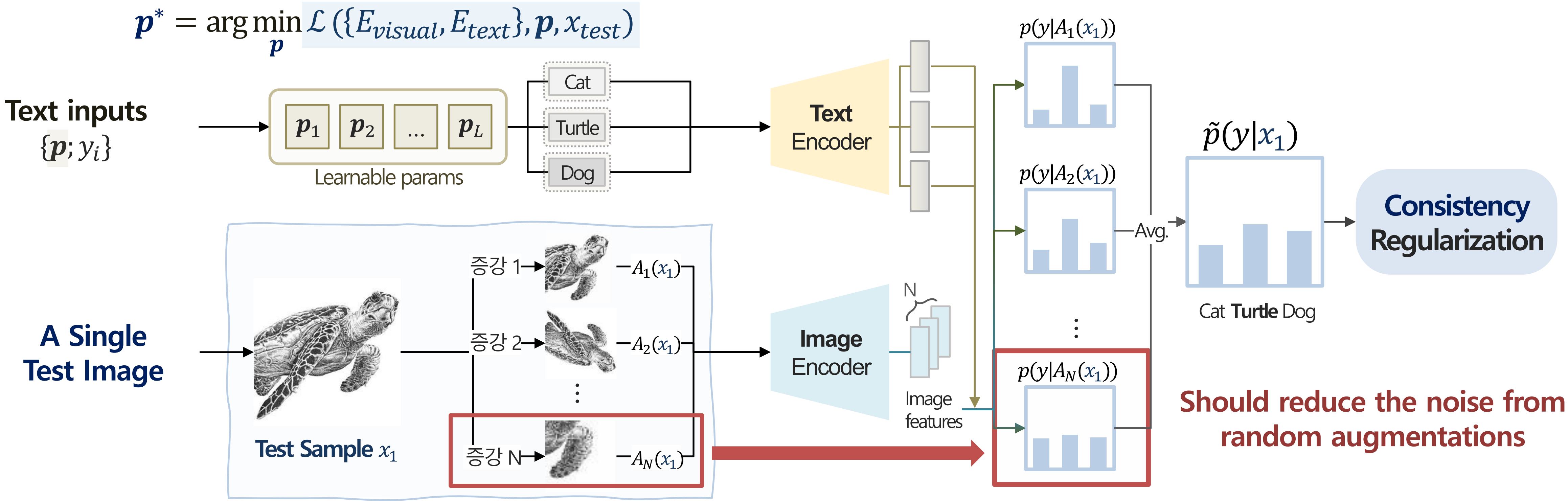
$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \mathcal{L}\left(\{E_{visual}, E_{text}\}, \boldsymbol{p}, x_{test}\right)$$



[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Test-Time Prompt Tuning

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \mathcal{L}\left(\{E_{visual}, E_{text}\}, \boldsymbol{p}, x_{test}\right)$$

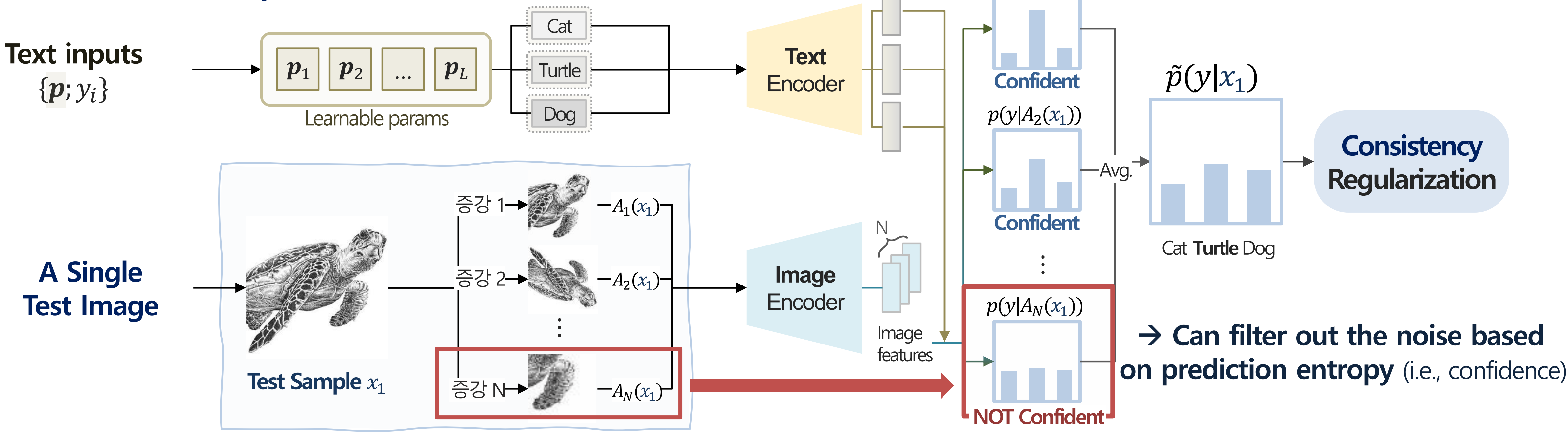


Should reduce the noise from random augmentations

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

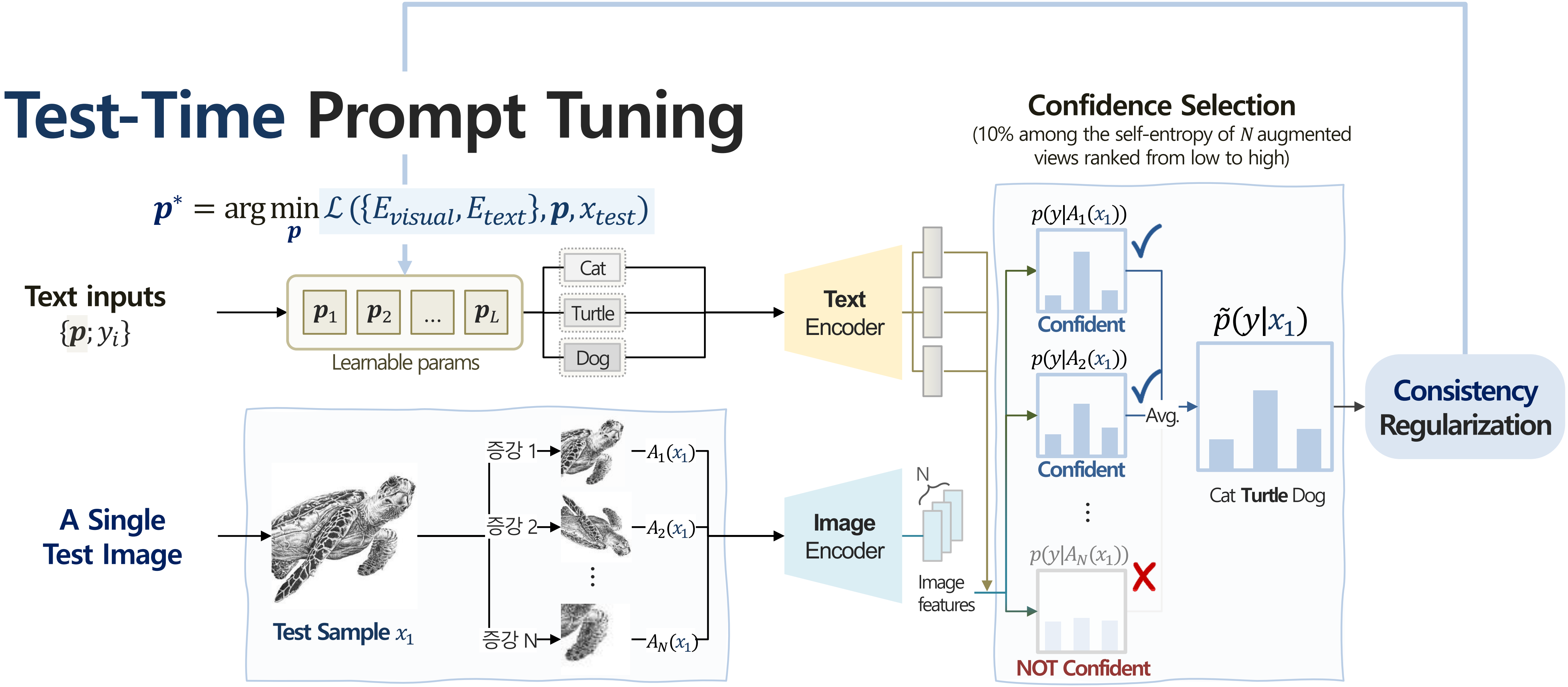
Test-Time Prompt Tuning

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \mathcal{L}\left(\{E_{visual}, E_{text}\}, \boldsymbol{p}, x_{test}\right)$$



2. TPT (Test-Time Prompt Tuning)

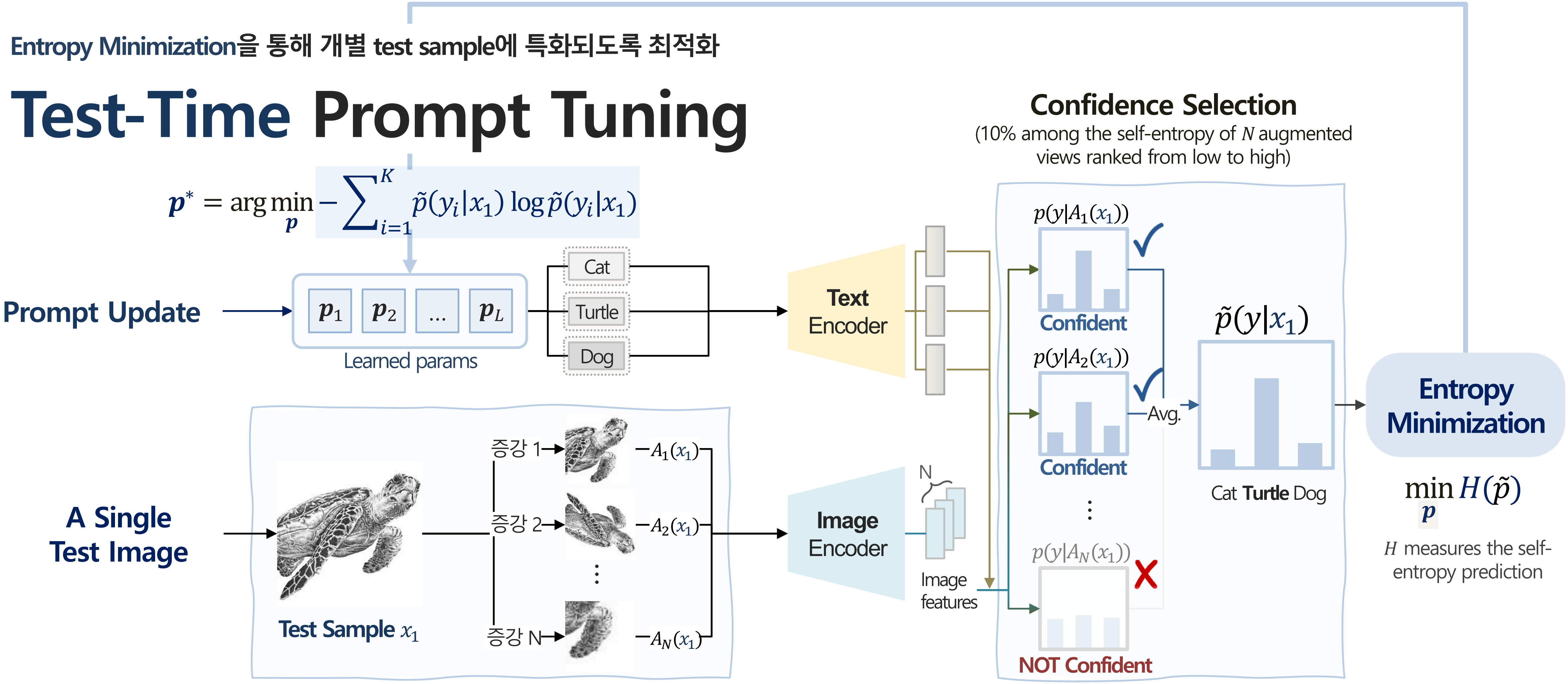
[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models



2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Entropy Minimization을 통해 개별 test sample에 특화되도록 최적화

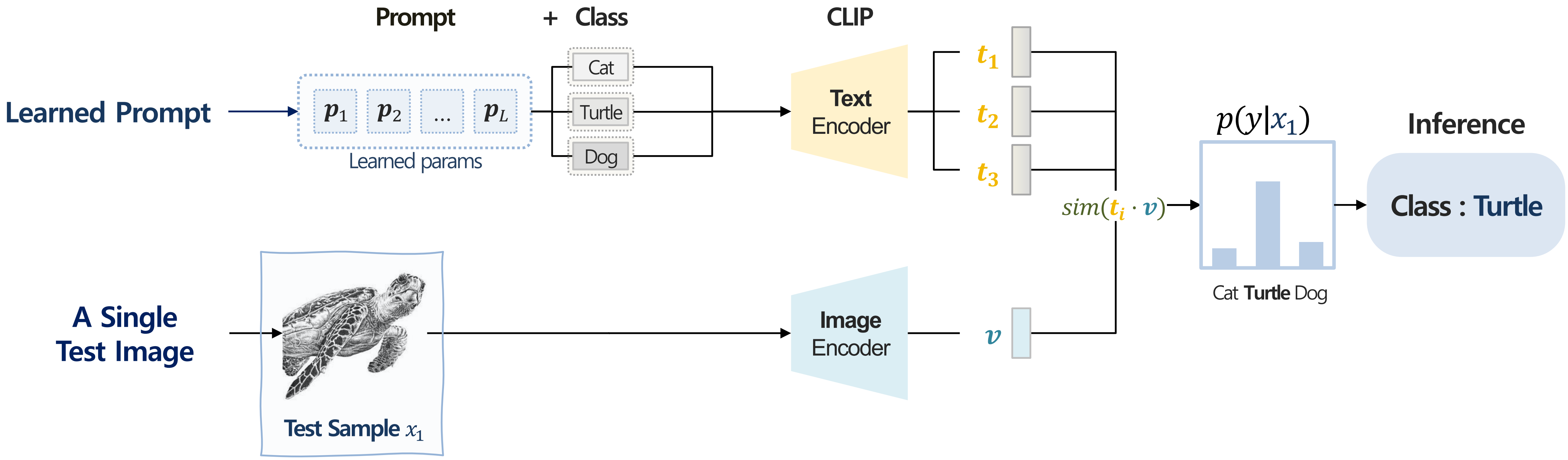


2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Entropy Minimization을 통해 개별 test sample에 특화되도록 최적화

Test-Time Prompt Tuning



2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Entropy Minimization을 통해 개별 test sample에 특화되도록 최적화

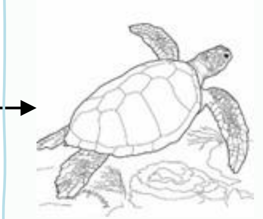
Test-Time Prompt Tuning

Initial Prompt

Prompt Reset → p_1 | p_2 | … | p_L

Learnable params

A Single Test Image →



Test Sample x_2

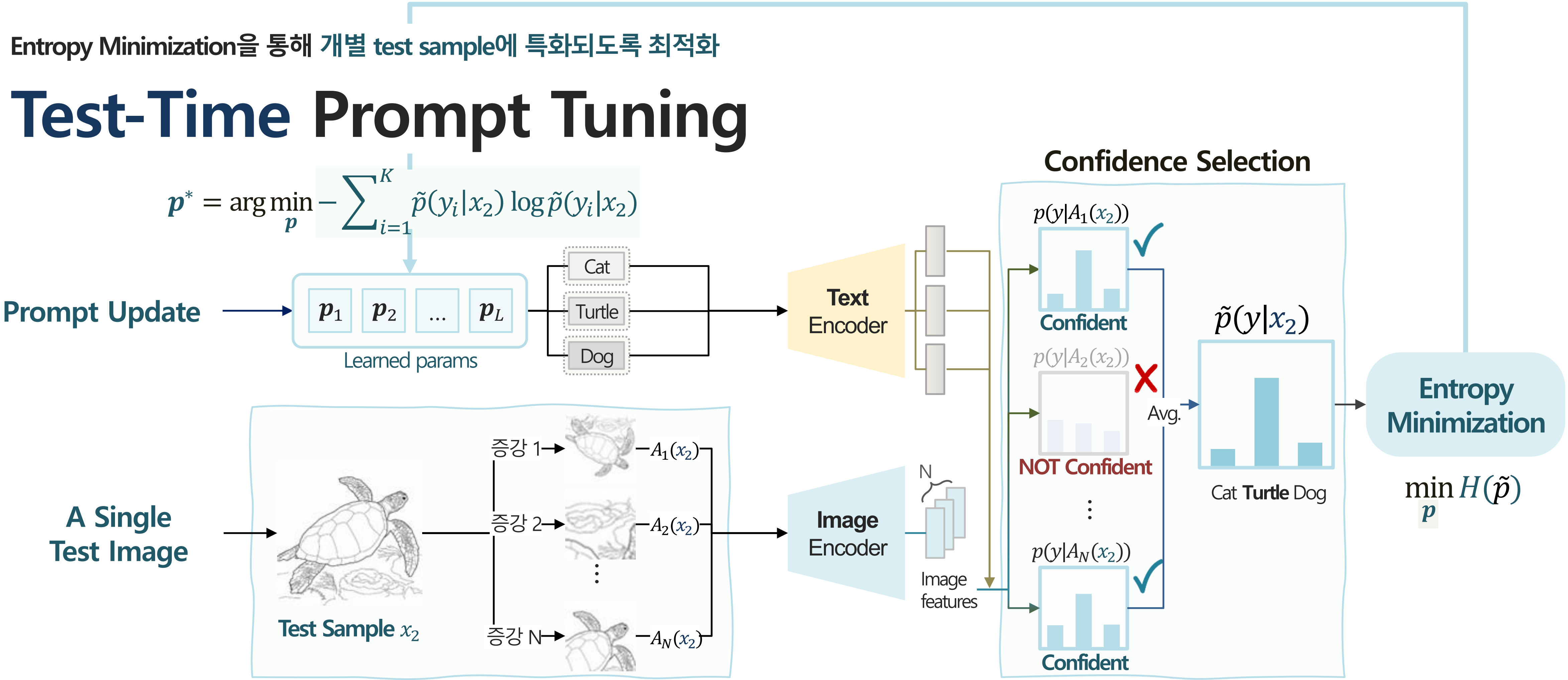
2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Entropy Minimization을 통해 개별 test sample에 특화되도록 최적화

Test-Time Prompt Tuning

$$p^* = \arg\min_{p} -\sum_{i=1}^{K} \tilde{p}(y_i|x_2) \log \tilde{p}(y_i|x_2)$$

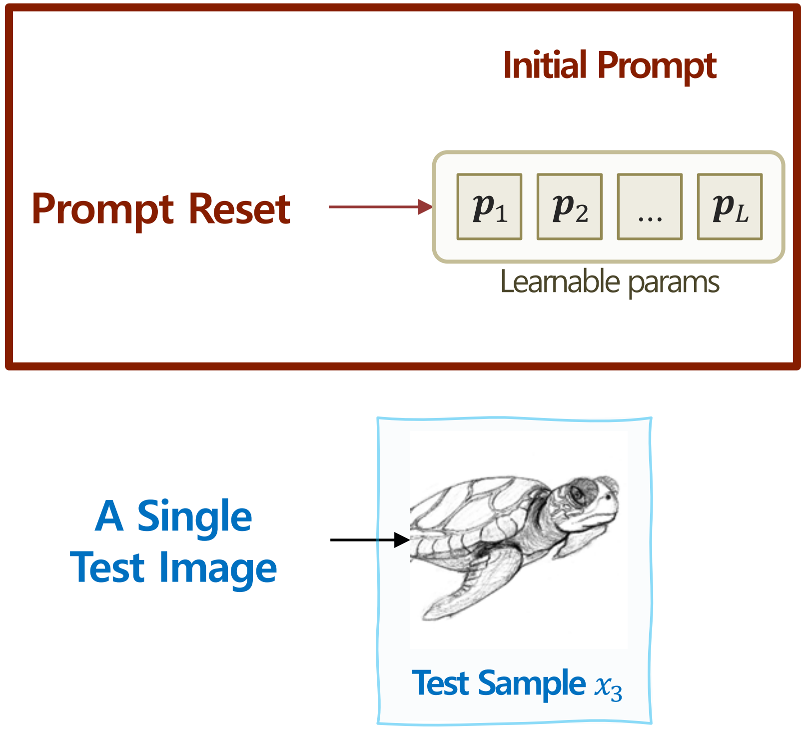


2. TPT (Test-Time Prompt Tuning)

[NeurIPS 2022] Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models

Entropy Minimization을 통해 개별 test sample에 특화되도록 최적화

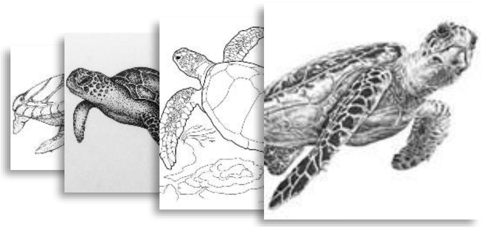
Test-Time Prompt Tuning



매번 초기 prompt에서 시작하여 현재 샘플에 대해서 최적화

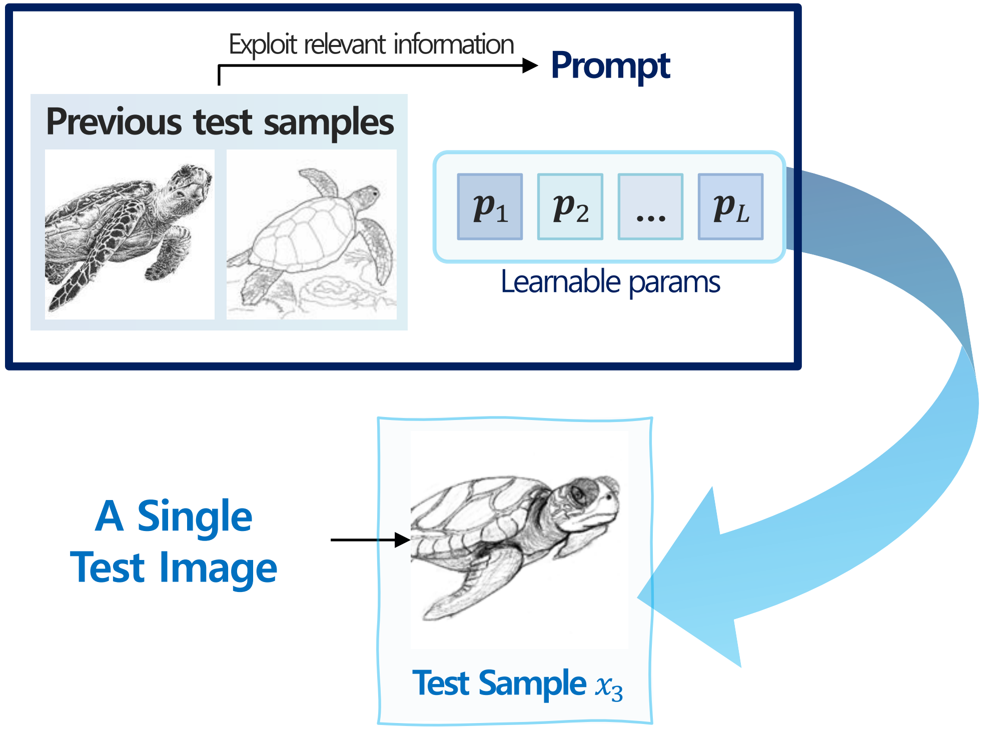
→ Adapting the prompt to each test sample individually

① Cannot exploit previous test information
② Ignores the relatedness among test samples



Entropy Minimization을 통해 개별 test sample에 특화되도록 최적화

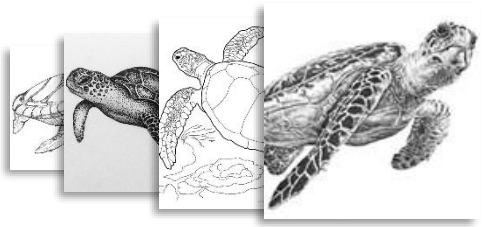
Test-Time Prompt Tuning



이전 prompt에서 시작하여 현재 샘플에 대해서 최적화

→ Using previous optimized prompt as the starting point

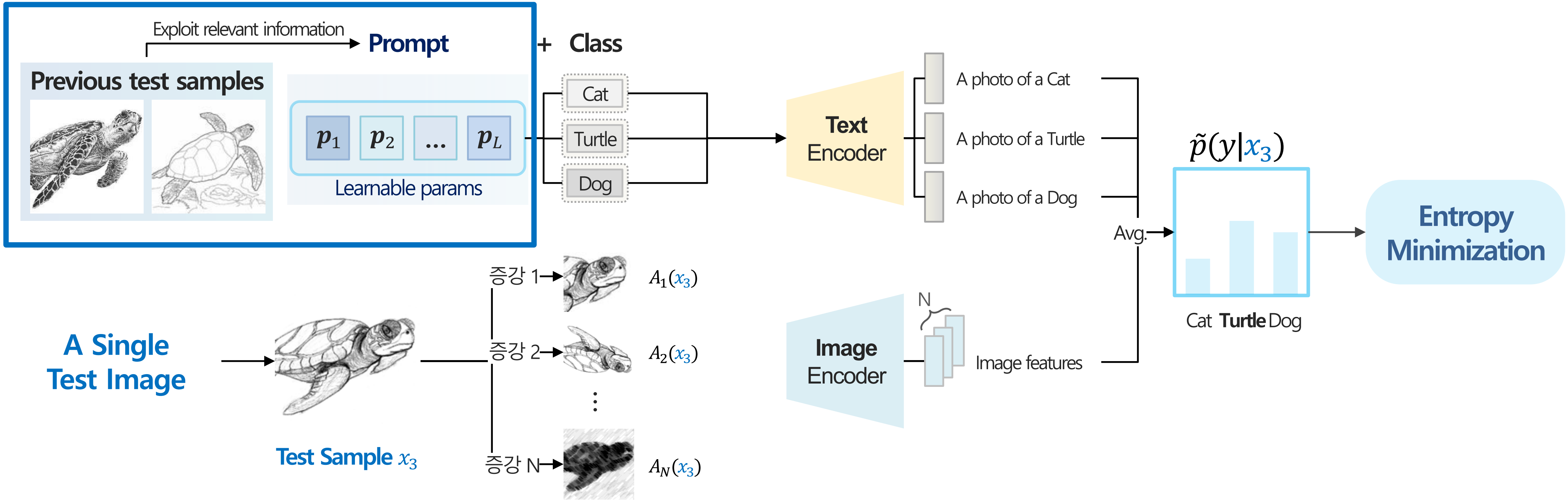
① Incorporates previous test information
② Leverages relevant information from previous test samples



3. Online TPT (Online Test-Time Prompt Tuning)

업데이트 된 프롬프트를 초기화하는 대신 **과거 정보를 누적**

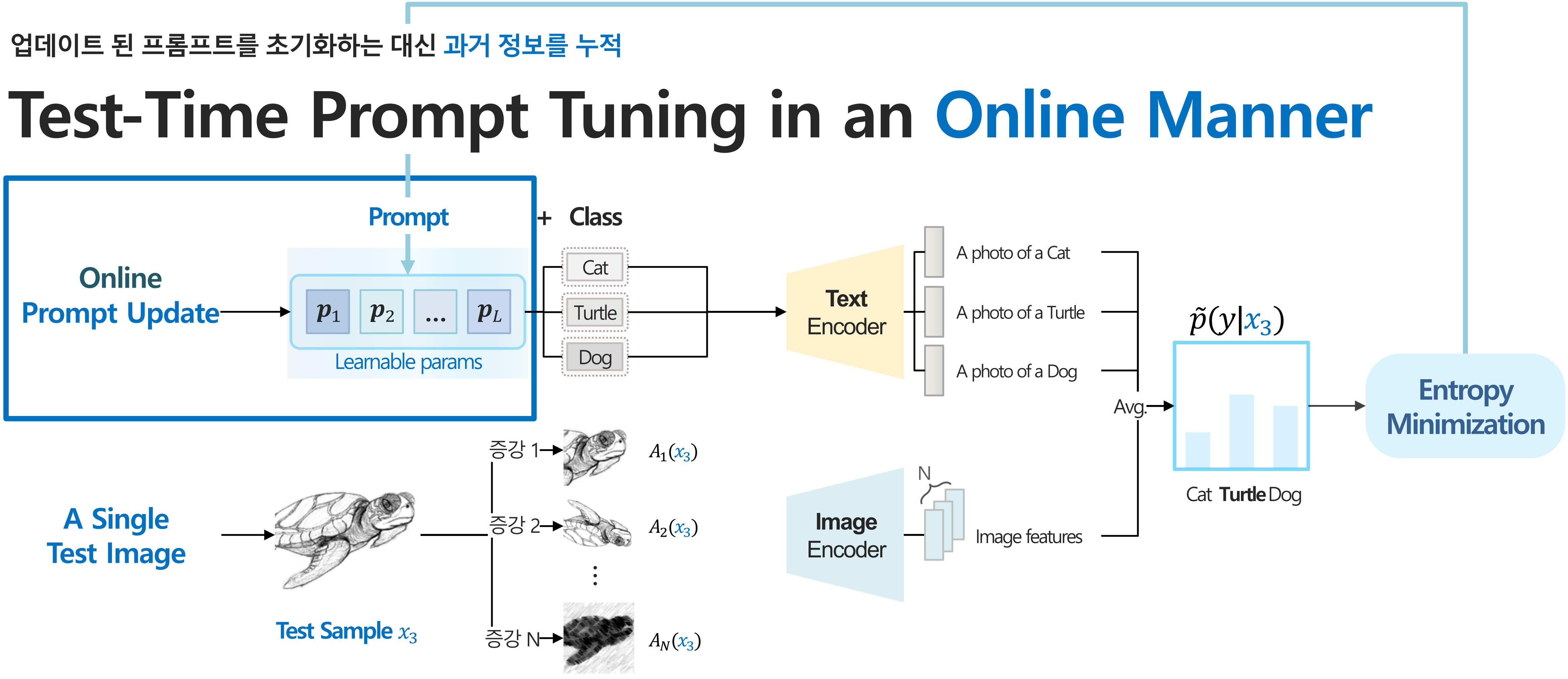
Test-Time Prompt Tuning in an **Online Manner**



3. Online TPT (Online Test-Time Prompt Tuning)

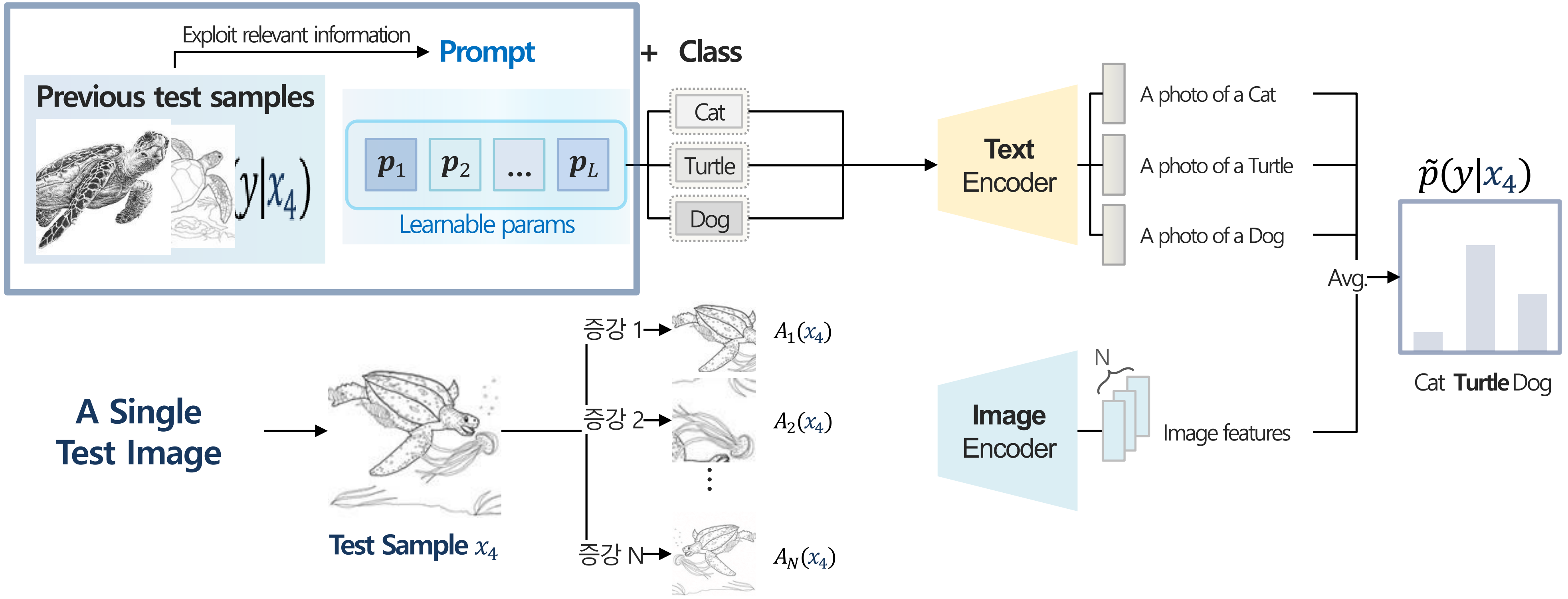
업데이트 된 프롬프트를 초기화하는 대신 **과거 정보를 누적**

Test-Time Prompt Tuning in an **Online Manner**



업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적

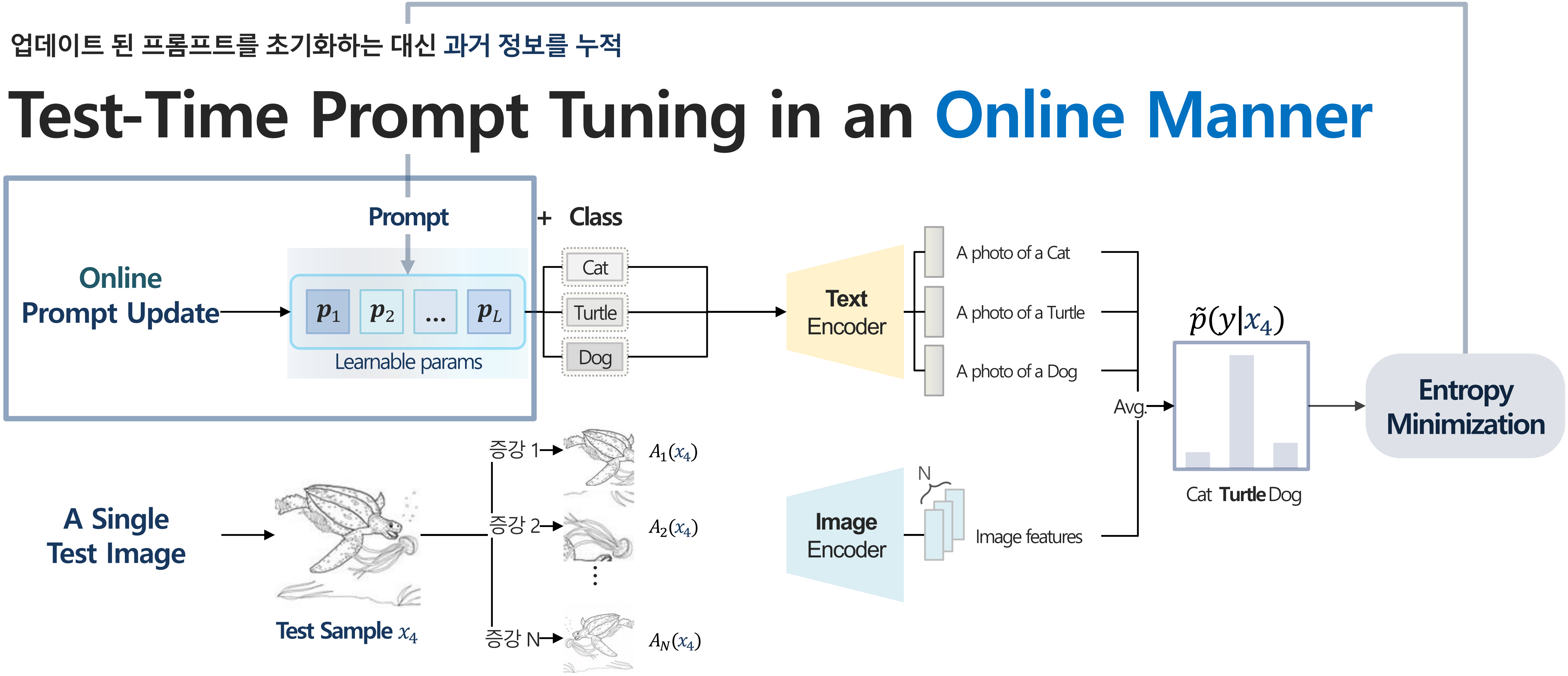
Test-Time Prompt Tuning in an Online Manner



3. Online TPT (Online Test-Time Prompt Tuning)

업데이트 된 프롬프트를 초기화하는 대신 **과거 정보를 누적**

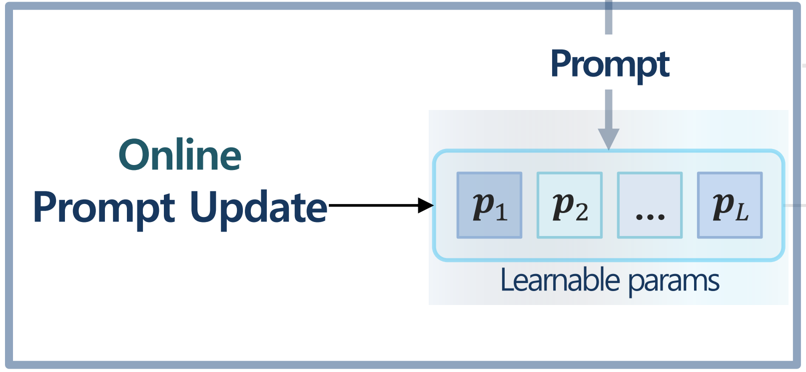
Test-Time Prompt Tuning in an **Online Manner**



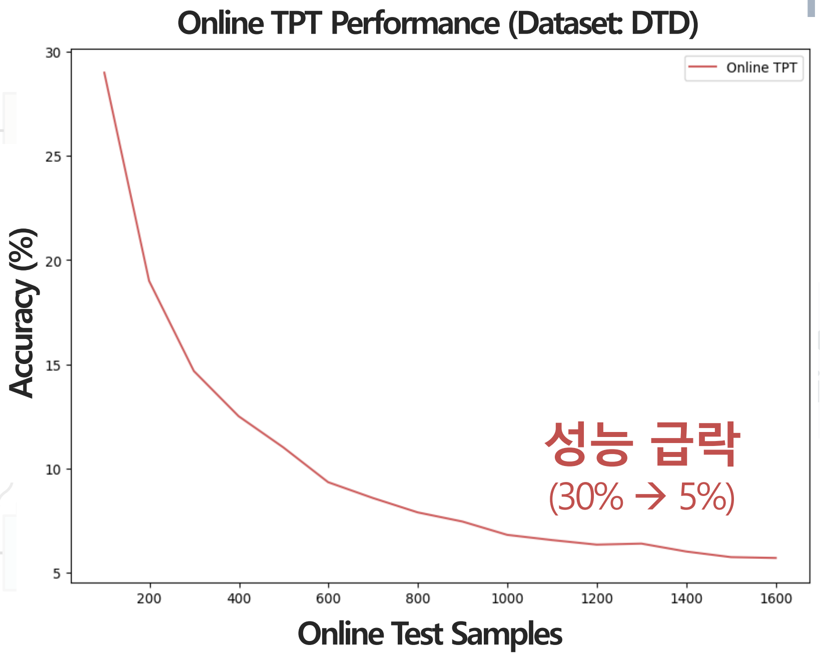
3. Online TPT (Online Test-Time Prompt Tuning)

업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적

Test-Time Prompt Tuning in an Online Manner



[5]



Online TPT shows competitive performance at the beginning while dropping significantly during online learning

[5] Xiao, Z., Yan, S., Hong, J., Cai, J., Jiang, X., Hu, Y., ... & Snoek, C. G. (2025). DynaPrompt: Dynamic Test-Time Prompt Tuning. arXiv preprint arXiv:2501.16404.

3. Online TPT (Online Test-Time Prompt Tuning)

업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적

Test-Time Prompt Tuning in an Online Manner

Without ground truth labels, the **entropy-based objectives** might drive the optimization in the wrong directions [6,7]

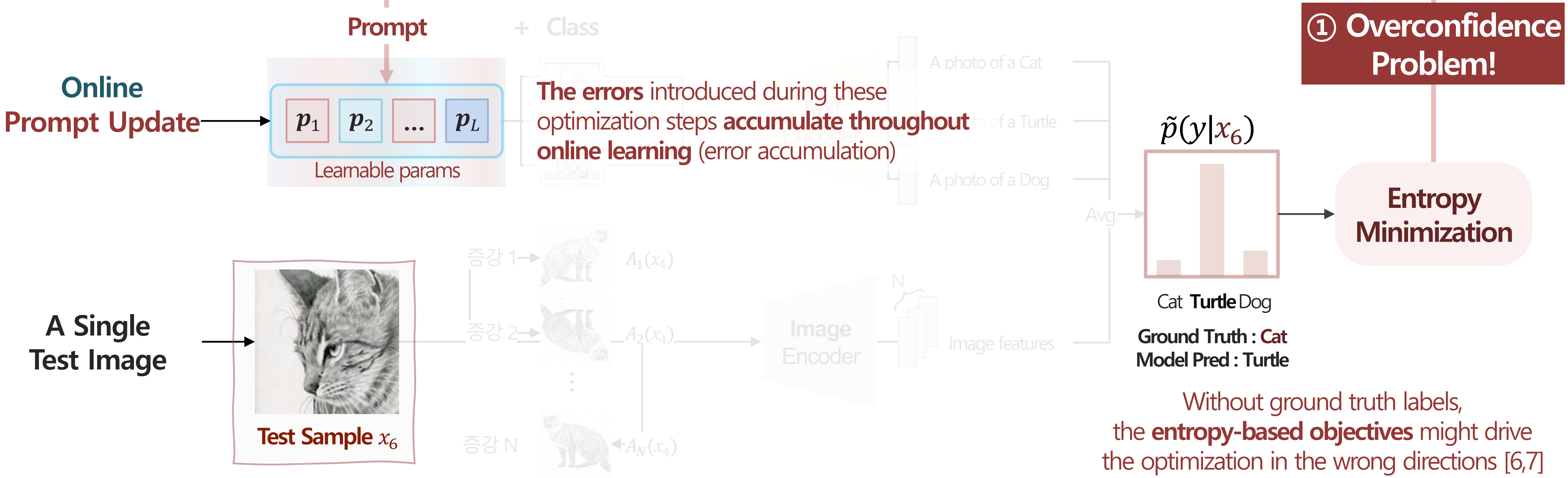
[6] Ma, X., Zhang, J., Guo, S., & Xu, W. (2023). Swapprompt: Test-time prompt adaptation for vision-language models. Advances in Neural Information Processing Systems, 36, 65252-65264.
[7] Lee, J., Jung, D., Lee, S., Park, J., Shin, J., Hwang, U., & Yoon, S. (2024). Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. *arXiv preprint arXiv:2403.07366*.

3. Online TPT (Online Test-Time Prompt Tuning)

② Error Accumulation Problem!

업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적 → 과거 샘플의 유용한 정보뿐만 아니라 오류도 함께 누적

Test-Time Prompt Tuning in an Online Manner



[6] Ma, X., Zhang, J., Guo, S., & Xu, W. (2023). Swapprompt: Test-time prompt adaptation for vision-language models. Advances in Neural Information Processing Systems, 36, 65252-65264.
[7] Lee, J., Jung, D., Lee, S., Park, J., Shin, J., Hwang, U., & Yoon, S. (2024). Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. *arXiv preprint arXiv:2403.07366*.

3. Online TPT (Online Test-Time Prompt Tuning)

② Error Accumulation Problem!

업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적 → 과거 샘플의 유용한 정보뿐만 아니라 오류도 함께 누적

Test-Time Prompt Tuning in an Online Manner

① Overconfidence Problem!

Prompt

Online Prompt Update → p_1 p_2 ... p_L

Learnable params

The errors introduced during these optimization steps **accumulate throughout online learning** (error accumulation)

$\tilde{p}(y|x_7)$

Cat **Turtle** Dog

→ Entropy Minimization

Ground Truth : Dog
Model Pred : Turtle

③ Prompt Collapse

Such error accumulation causes the prompts to progressively degenerate, resulting in gradually degraded performance

→ Ultimately cause collapsed prompts that **generate incorrect predictions with lower entropies**

A Single Test Image →

Test Sample x_7

Without ground truth labels, the **entropy-based objectives** might drive the optimization in the wrong directions [6,7]

[6] Ma, X., Zhang, J., Guo, S., & Xu, W. (2023). Swapprompt: Test-time prompt adaptation for vision-language models. Advances in Neural Information Processing Systems, 36, 65252-65264.
[7] Lee, J., Jung, D., Lee, S., Park, J., Shin, J., Hwang, U., & Yoon, S. (2024). Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. *arXiv preprint arXiv:2403.07366.*

3. Online TPT (Online Test-Time Prompt Tuning)

업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적 → 과거 샘플의 유용한 정보뿐만 아니라 오류도 함께 누적

Test-Time Prompt Tuning in an Online Manner

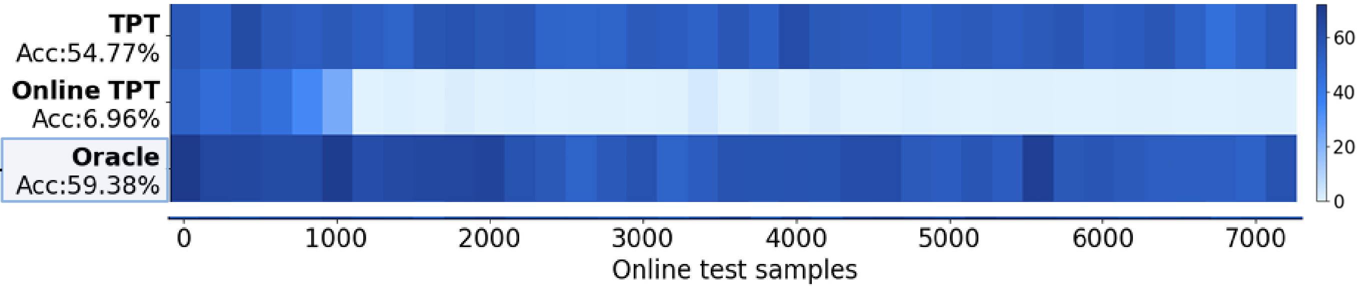
🤔 그냥 prompt reset을 하면 안 되는 걸까?

업데이트 된 프롬프트를 초기화하는 대신 과거 정보를 누적 → 과거 샘플의 유용한 정보뿐만 아니라 오류도 함께 누적

Test-Time Prompt Tuning in an Online Manner

그냥 prompt reset을 하면 안 되는 걸까?

[5]



Only update the prompts with correct predictions

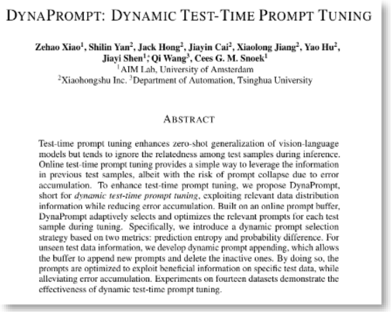
→ 'Oracle' method considerably outperforms TPT, which implies that the relevant information in past test samples benefits prompt tuning on the test distribution!

[5] Xiao, Z., Yan, S., Hong, J., Cai, J., Jiang, X., Hu, Y., ... & Snoek, C. G. (2025). DynaPrompt: Dynamic Test-Time Prompt Tuning. arXiv preprint arXiv:2501.16404.

4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

Dynamic Prompt Tuning



DYNAPROMPT: DYNAMIC TEST-TIME PROMPT TUNING

Zehao Xiao[1], Shilin Yan[2], Jack Hong[2], Jiayin Cai[2], Xiaolong Jiang[2], Yao Hu[2], Jiayi Shen[1]*, Qi Wang[3], Cees G. M. Snoek[1]
[1]AIM Lab, University of Amsterdam
[2]Xiaohongshu Inc. [3]Department of Automation, Tsinghua University

ABSTRACT

Test-time prompt tuning enhances zero-shot generalization of vision-language models but tends to ignore the relatedness among test samples during inference. Online test-time prompt tuning provides a simple way to leverage the information in previous test samples, albeit with the risk of prompt collapse due to error accumulation. To enhance test-time prompt tuning, we propose DynaPrompt, short for *dynamic test-time prompt tuning*, exploiting relevant data distribution information while reducing error accumulation. Built on an online prompt buffer, DynaPrompt adaptively selects and optimizes the relevant prompts for each test sample during tuning. Specifically, we introduce a dynamic prompt selection strategy based on two metrics: prediction entropy and probability difference. For unseen test data information, we develop dynamic prompt appending, which allows the buffer to append new prompts and delete the inactive ones. By doing so, the prompts are optimized to exploit beneficial information on specific test data, while alleviating error accumulation. Experiments on fourteen datasets demonstrate the effectiveness of dynamic test-time prompt tuning.

DynaPrompt (Dynamic Test-Time Prompt Tuning)

ICLR, 2025,
DynaPrompt: Dynamic Test-Time Prompt Tuning
2025년 7월 기준 5회 인용

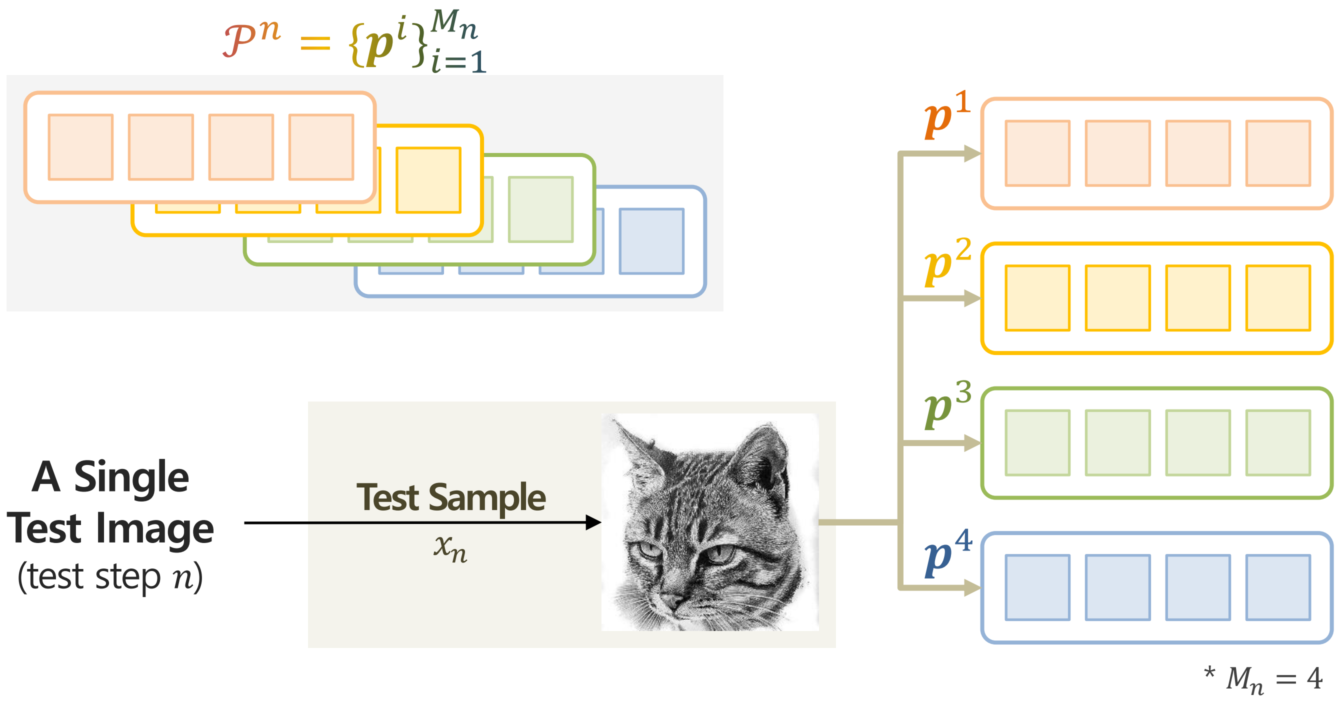
① Exploit beneficial information from prompt histories
② Alleviate the error accumulation in online prompt tuning

4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$



① Exploit beneficial information from prompt histories

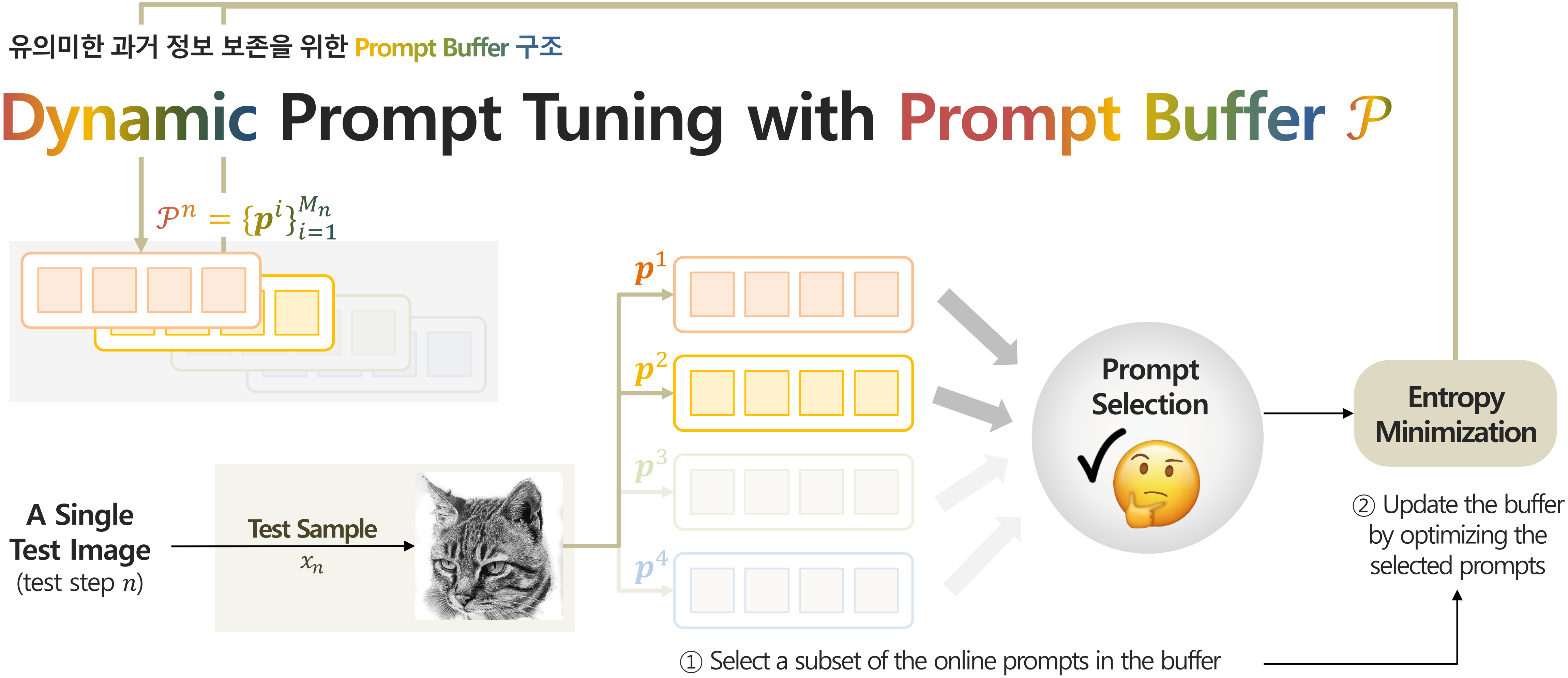
→ Store the distribution information in past samples using the buffer!

4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

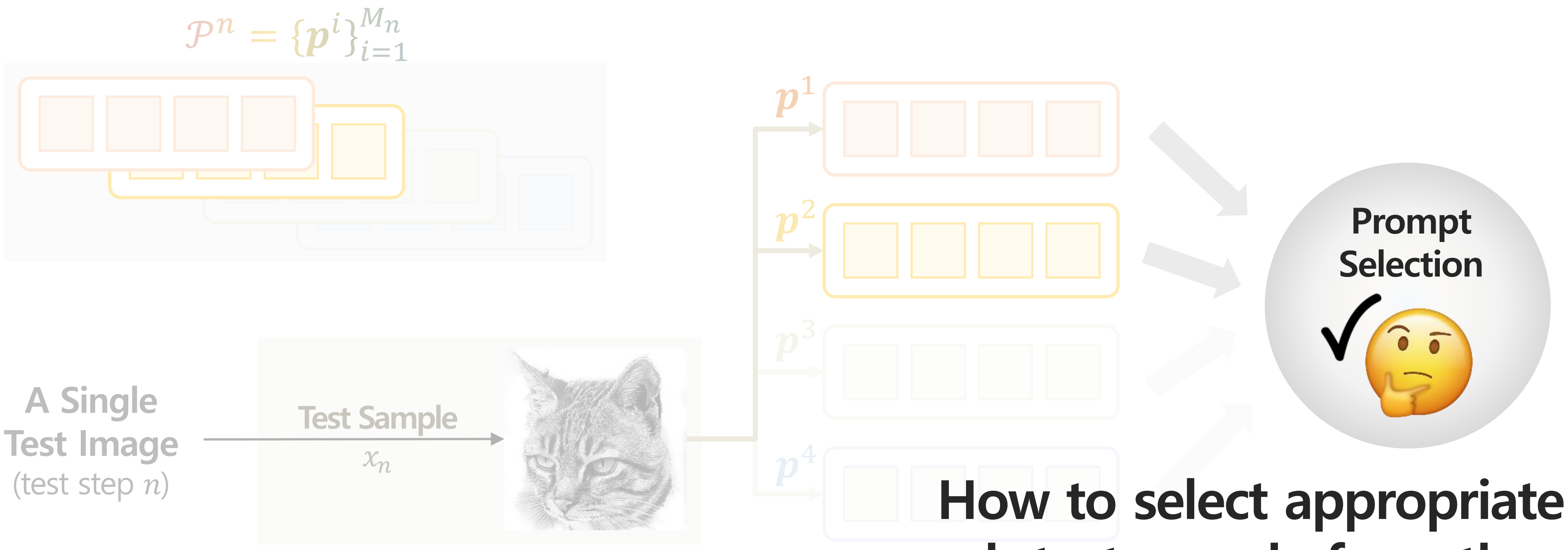


4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$



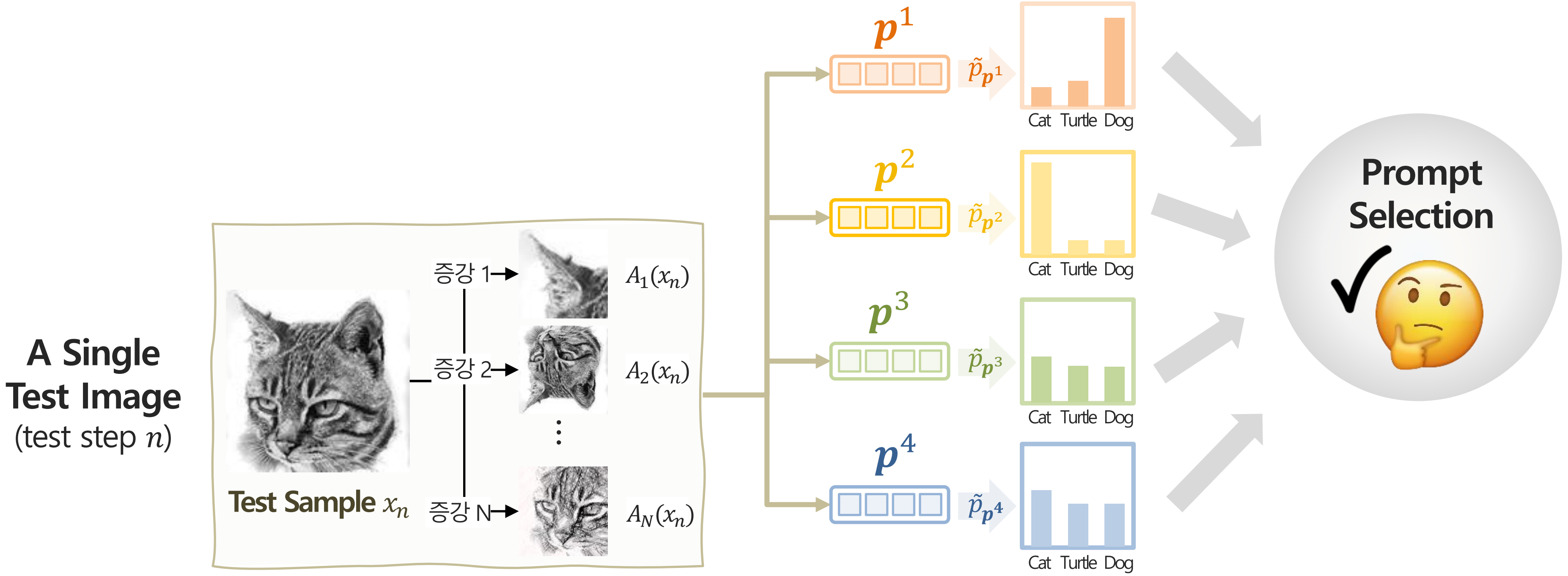
How to select appropriate prompts for each test sample from the prompt buffer?

4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$



4. DynaPrompt (Dynamic Prompt Tuning)

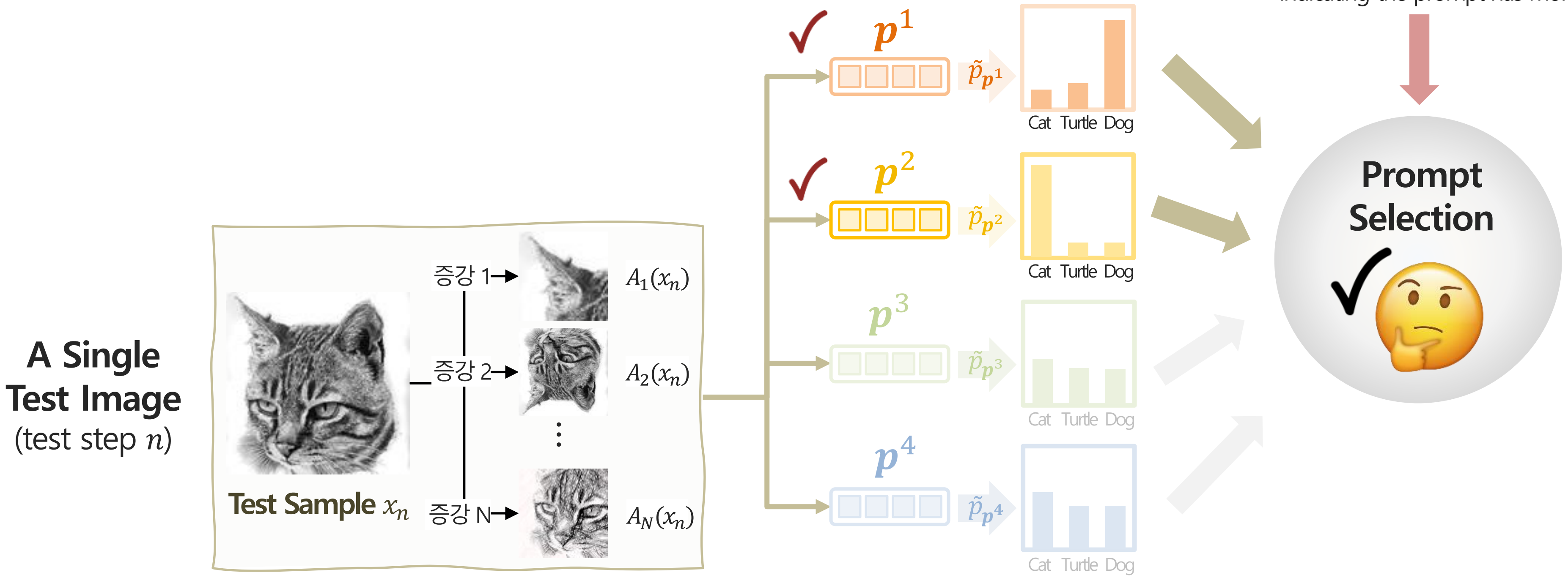
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

Prediction Entropy (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

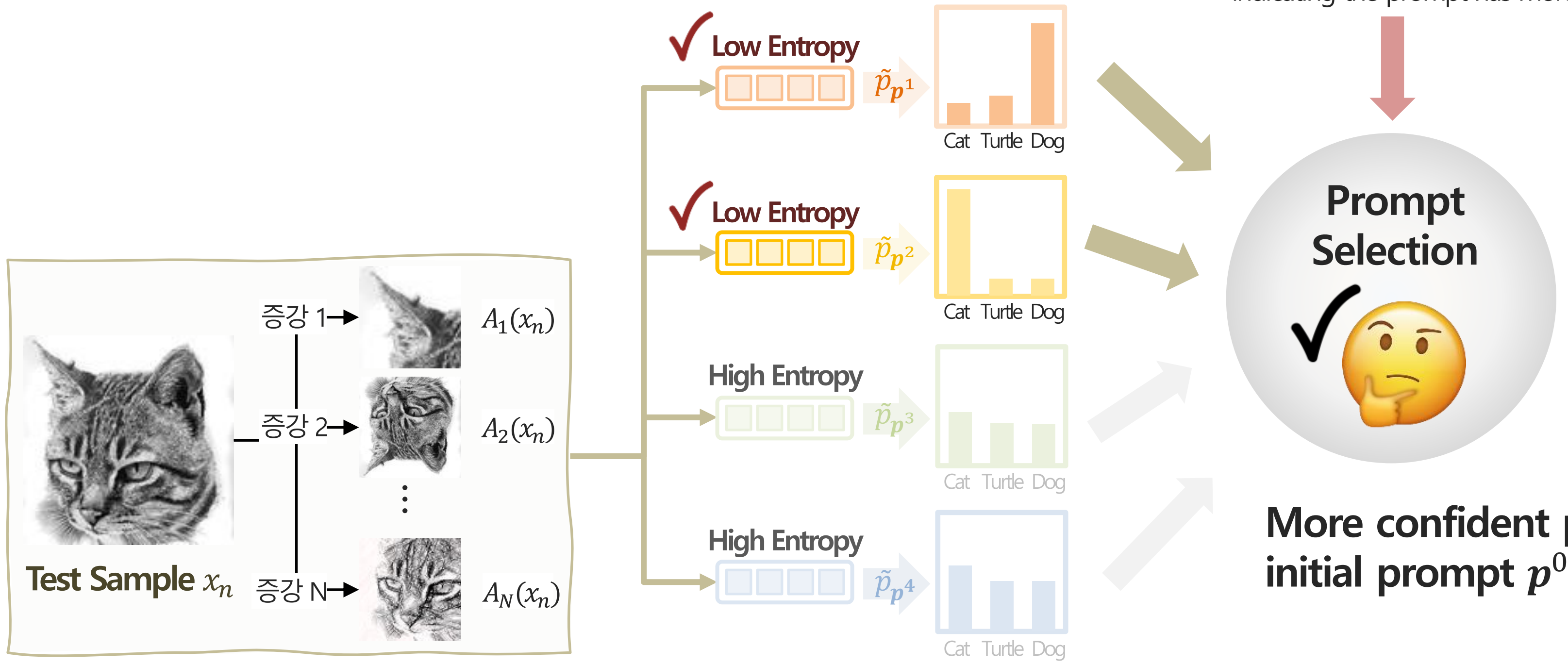
유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

Prediction Entropy (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information

A Single Test Image
(test step n)



Threshold

$H(\tilde{p}_{p^i}) \leq H(\tilde{p}_{p^0})$,

where p^0 denotes the initial prompt

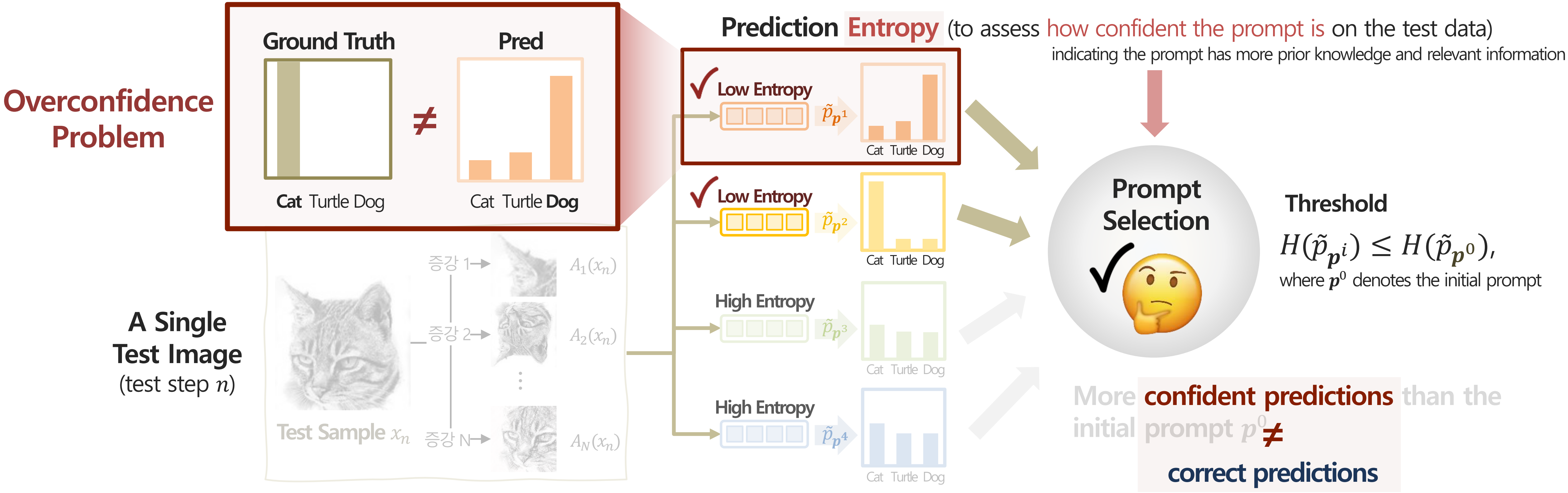
More confident predictions than the initial prompt p^0

4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$



4. DynaPrompt (Dynamic Prompt Tuning)

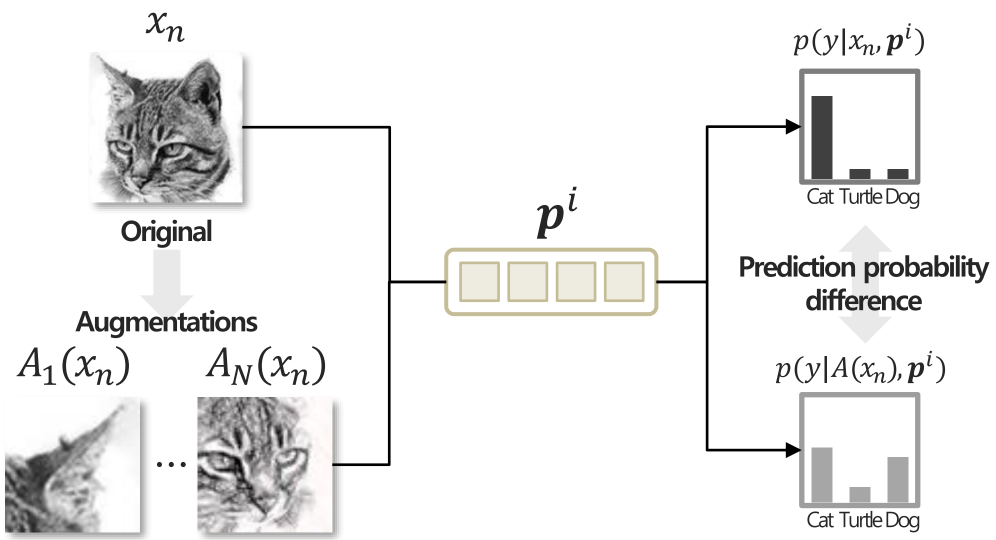
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

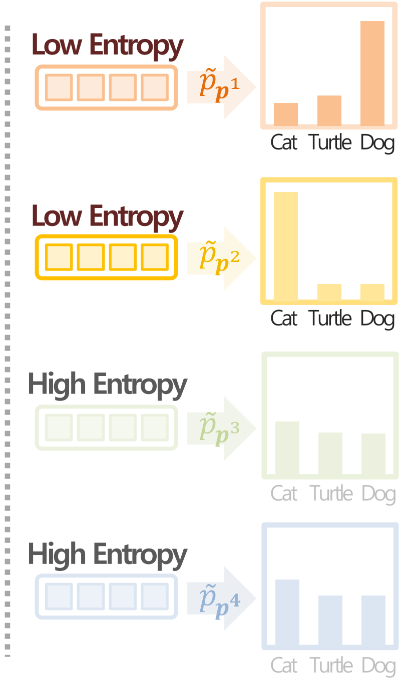
유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** $\cap$ **Prediction Entropy** (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information





→ **Assessing the SENSITIVITY of the prompts**
(assessing the risk of overconfidence and prompt collapse)

4. DynaPrompt (Dynamic Prompt Tuning)

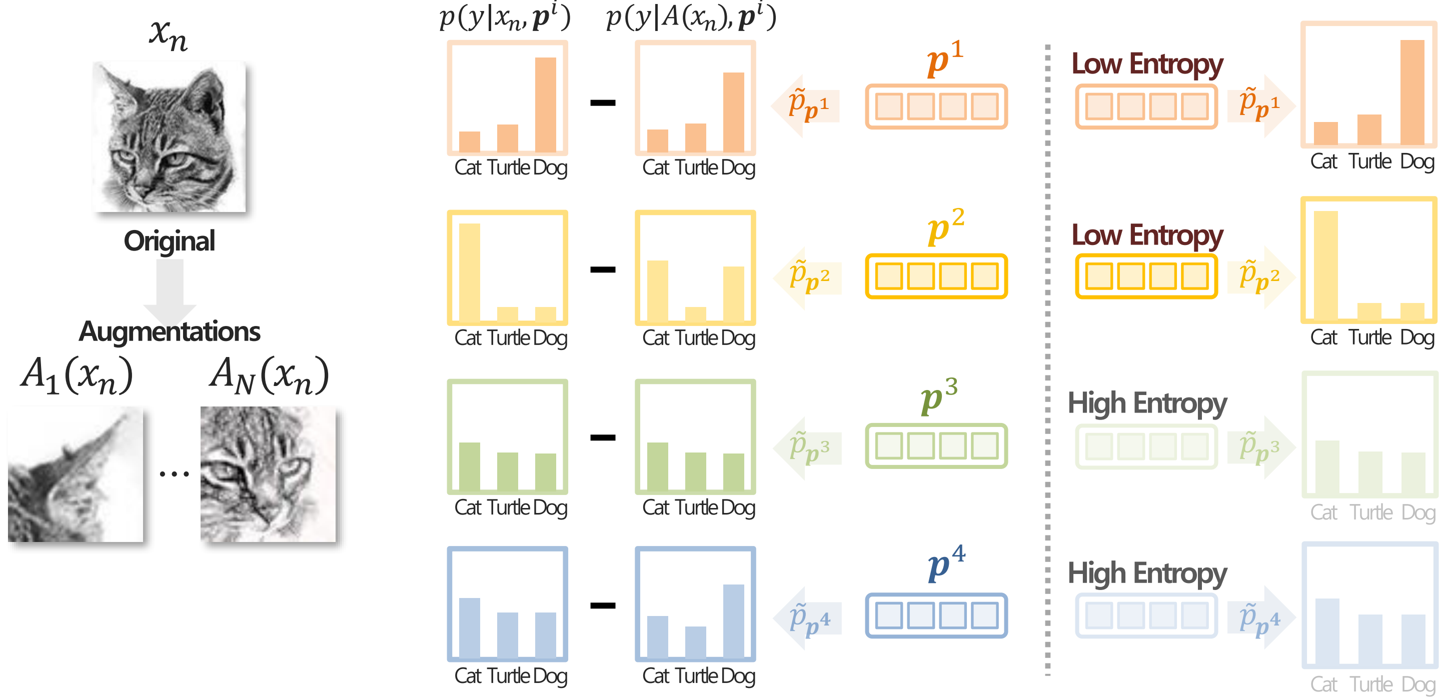
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) Probability Difference ∩ Prediction Entropy (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

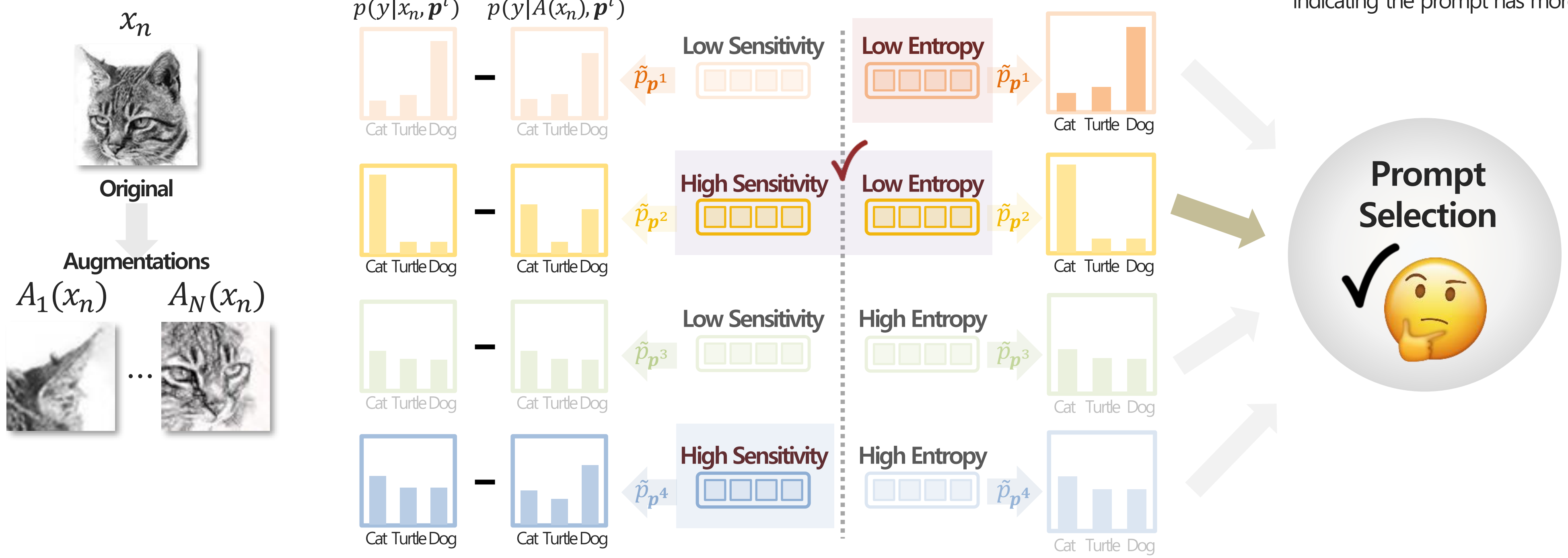
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) Probability Difference ∩ Prediction Entropy (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

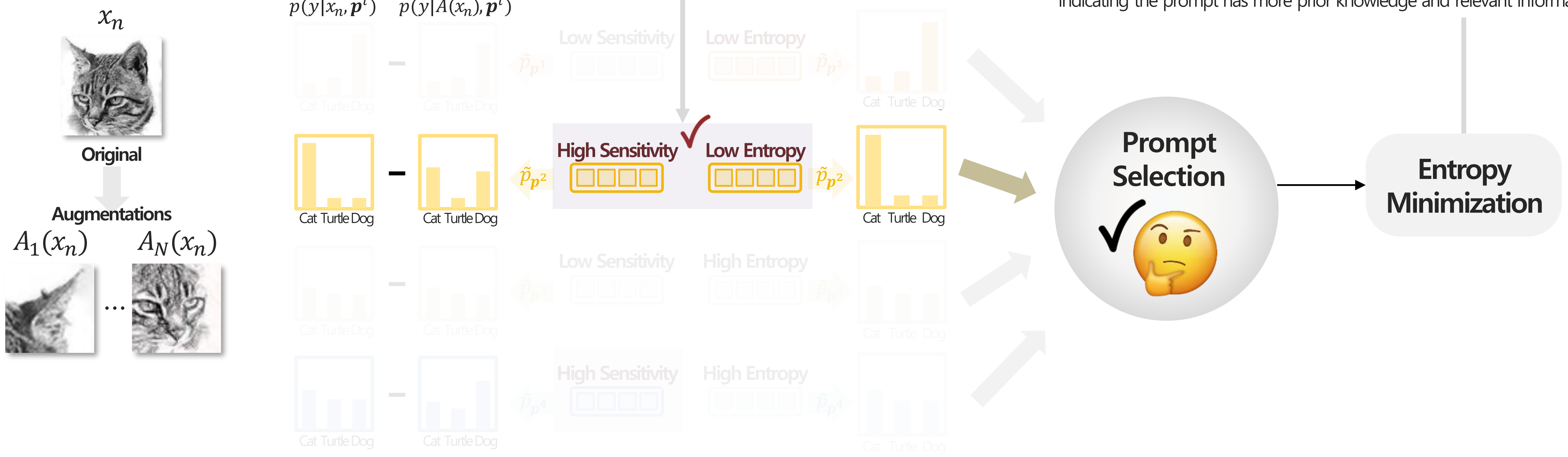
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) Probability Difference ∩ Prediction Entropy (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

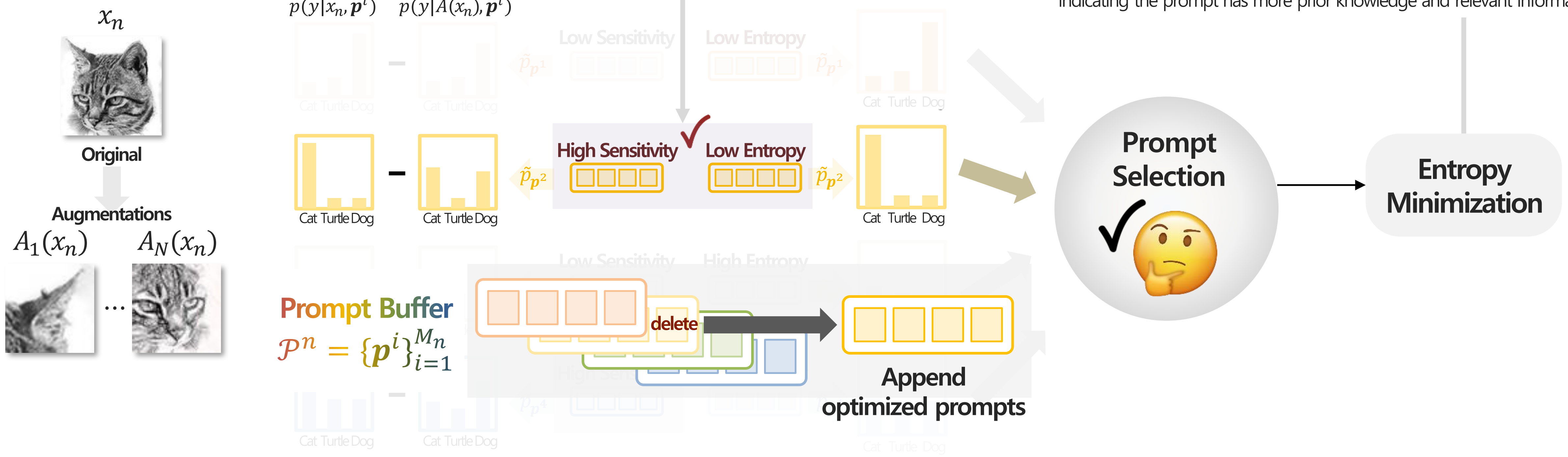
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** ∩ **Prediction Entropy** (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

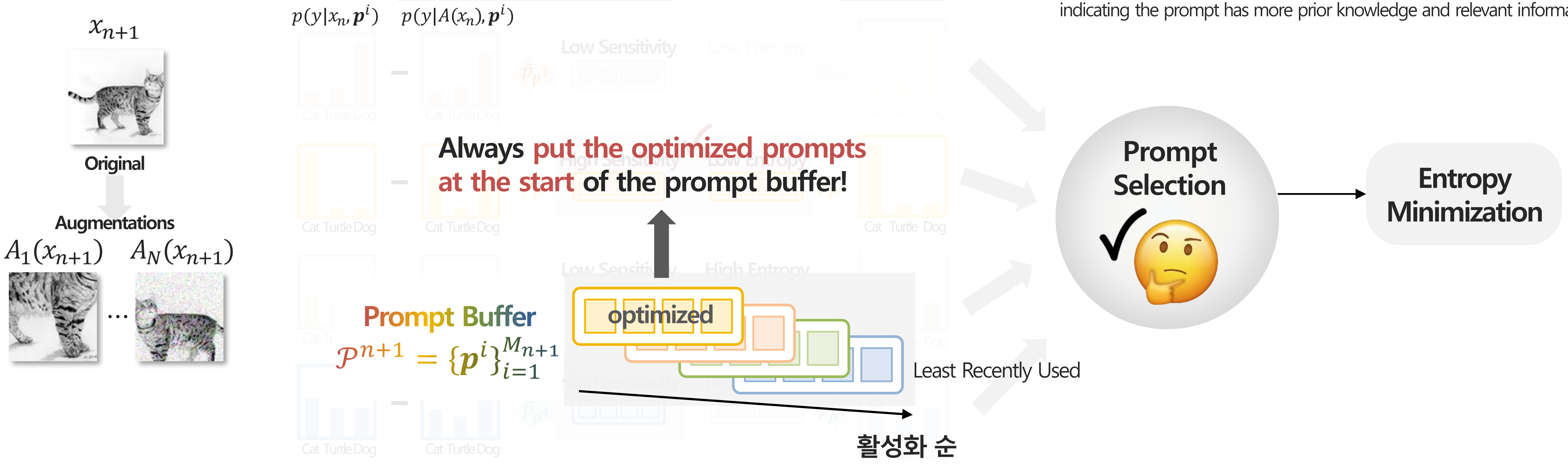
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** ∩ **Prediction Entropy** (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

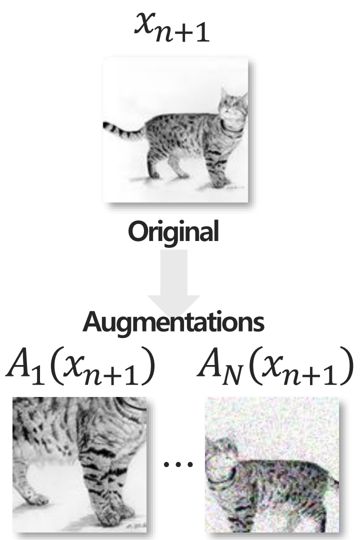
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** ∩ **Prediction Entropy** (to assess how confident the prompt is on the test data)

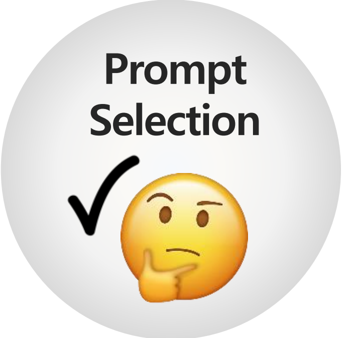
indicating the prompt has more prior knowledge and relevant information



만일 교집합이 없다면??

① Buffer $\mathcal{P}^{n+1}$ 내 모든 프롬프트가 **현재 샘플 x_{n+1}과 무관**

② Buffer $\mathcal{P}^{n+1}$ 내 모든 프롬프트가 **이미 collapsed 된 상태**



기존 프롬프트를 사용하는 대신 **새로운 프롬프트**를 추가

$\boldsymbol{p}^0$

Initial prompt e.g., "a photo of a"

4. DynaPrompt (Dynamic Prompt Tuning)

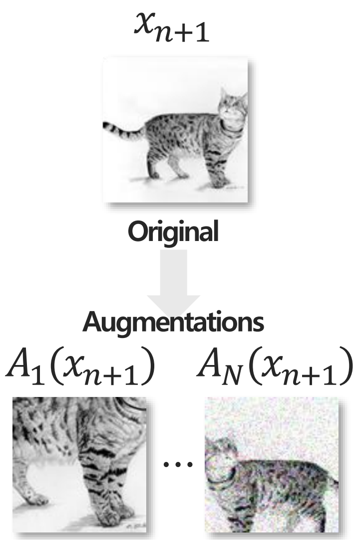
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** $\cap$ **Prediction Entropy** (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



만일 교집합이 없다면??

① Buffer $\mathcal{P}^{n+1}$ 내 모든 프롬프트가 **현재 샘플 x_{n+1}과 무관**

② Buffer $\mathcal{P}^{n+1}$ 내 모든 프롬프트가 **이미 collapsed 된 상태**

기존 프롬프트를 사용하는 대신 **새로운 프롬프트**를 추가

Optimized prompt

Prompt Selection ✓ → Entropy Minimization

4. DynaPrompt (Dynamic Prompt Tuning)

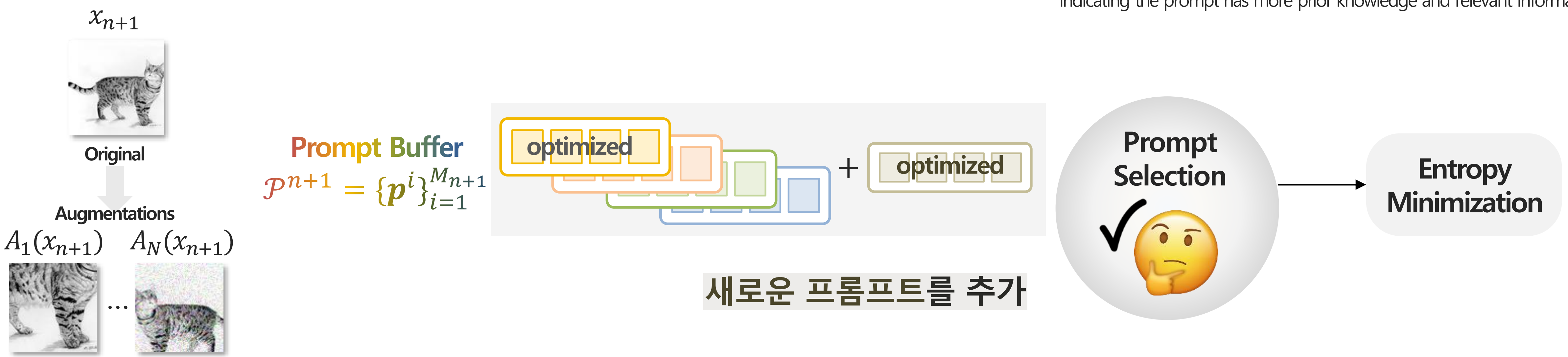
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** ∩ **Prediction Entropy** (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information



4. DynaPrompt (Dynamic Prompt Tuning)

[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) Probability Difference ∩ Prediction Entropy (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information

x_{n+1}

Original

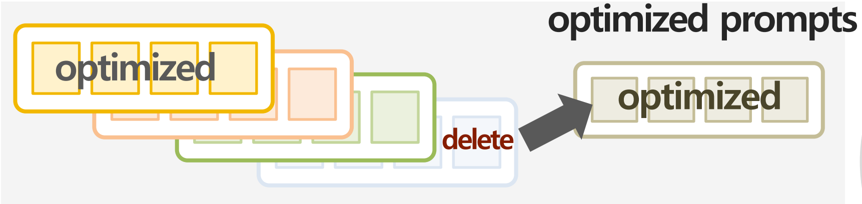
Augmentations

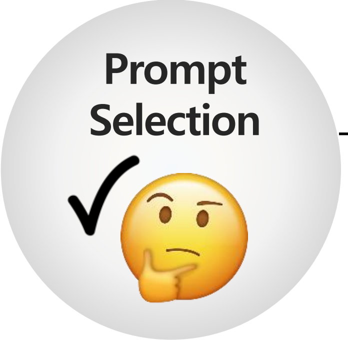
$A_1(x_{n+1})$ $A_N(x_{n+1})$

Prompt Buffer

$\mathcal{P}^{n+1} = \{\boldsymbol{p}^i\}_{i=1}^{M_{n+1}}$

Append optimized prompts





새로운 프롬프트를 추가

가장 오래 비활성 상태였던 프롬프트를 제거한 뒤

4. DynaPrompt (Dynamic Prompt Tuning)

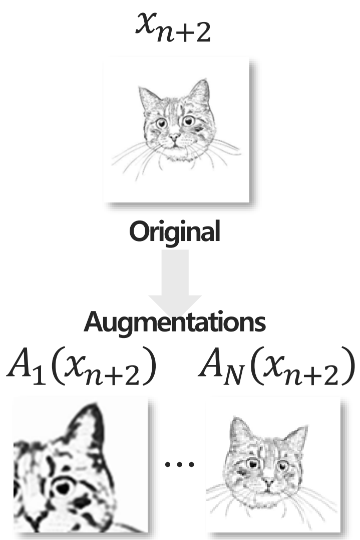
[ICLR 2025] DynaPrompt: Dynamic Test-Time Prompt Tuning

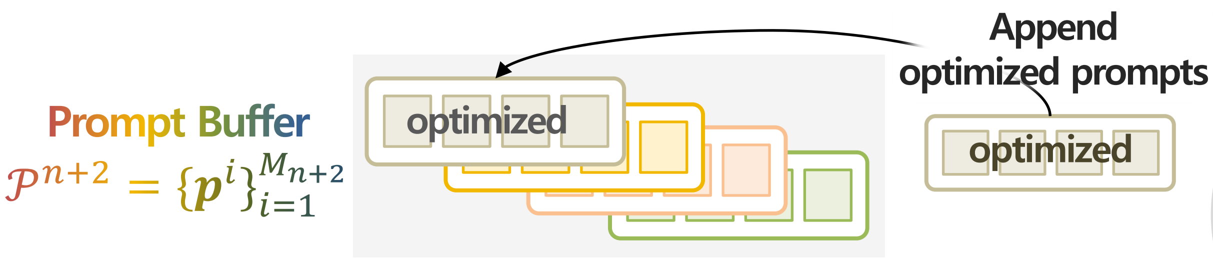
유의미한 과거 정보 보존을 위한 Prompt Buffer 구조 + Prompt Collapse 방지를 위한 Dynamic Prompt Selection Strategy

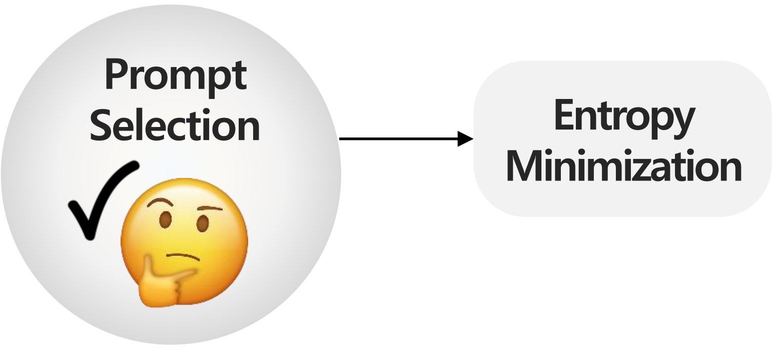
Dynamic Prompt Tuning with Prompt Buffer $\mathcal{P}$

(To circumvent overconfident prompts) **Probability Difference** $\cap$ **Prediction Entropy** (to assess how confident the prompt is on the test data)

indicating the prompt has more prior knowledge and relevant information







By appending new online prompts and ejecting the inactive ones, it effectively incorporates information from new data distributions and reduces error accumulation!

Experiments

Datasets

① Domain Generalization

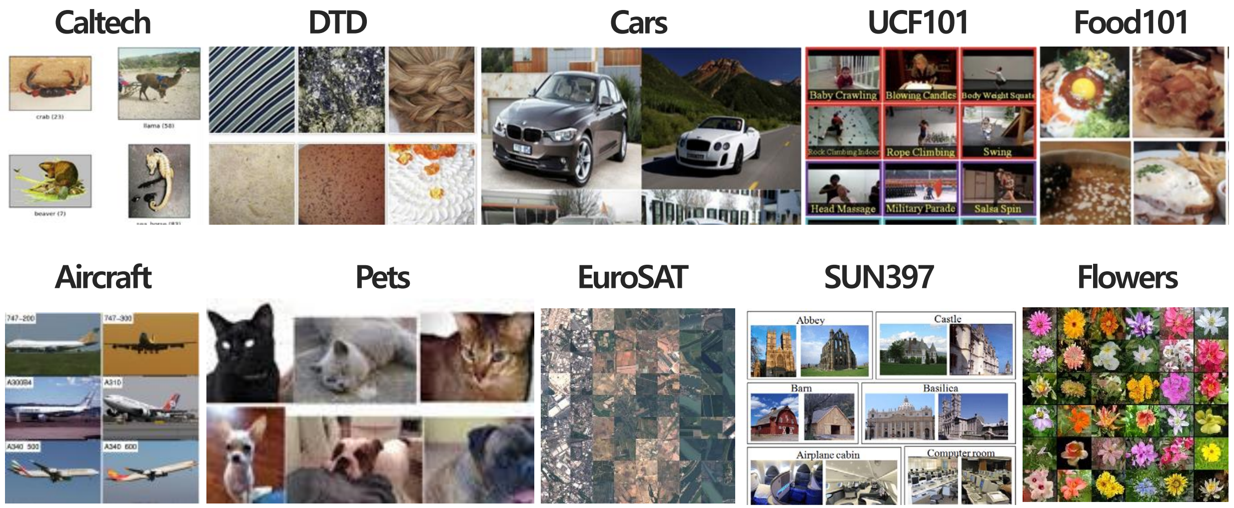
- Prompt tuning vs. Test-time prompt tuning : CoOp과 같은 downstream dataset-specific 알고리즘 대비 OOD 성능 확인
- 5 datasets including out-of-distribution scenarios : ImageNet, ImageNet-V2, ImageNet-Sketch, ImageNet-A and ImageNet-R

② Cross-Datasets Generalization

- Prompt tuning vs. Test-time prompt tuning : Arbitrary categories가 주어지는 zero-shot 상황에서의 classification 성능 확인
- 10 fine-grained classification datasets : Caltech, DTD, Cars, UCF101, Food101, Aircraft, SUN397, Pets, EuroSAT and Flowers

Domain Generalization

Cross-Datasets Generalization



Experiments

Results

① Domain Generalization

- Prompt tuning vs. Test-time prompt tuning : CoOp과 같은 downstream dataset-specific 알고리즘 대비 OOD 성능 확인

	Source	Target				
Method	ImageNet	ImageNet-V2	ImageNet-S	ImageNet-A	ImageNet-R	*OoD Mean*
CLIP (Radford et al., 2021)	66.73	60.86	46.09	47.87	73.98	57.20
Prompt learning methods without test-time tuning						
CoOp (Zhou et al., 2022b)	71.51	64.20	47.99	49.71	75.21	59.28
CoCoOp (Zhou et al., 2022a)	71.02	64.07	48.75	50.63	76.18	59.90
KgCoOp (Yao et al., 2023)	71.20	64.10	48.97	50.69	76.70	60.11
MaPLe (Khattak et al., 2023)	70.72	64.07	49.15	50.90	76.98	60.28
CoPrompt (Roy & Etemad, 2024)	70.80	64.25	49.43	50.50	77.51	60.42
Any-shift Prompt (Xiao et al., 2024)	-	64.53	49.80	51.52	77.56	60.85
Test-time prompt tuning methods						
TPT (Shu et al., 2022)	68.98	63.45	47.94	54.77	77.06	60.81
AdaPrompt (Zhang et al., 2024)	-	*59.32*	*47.72*	*47.71*	73.98	57.18
DiffTPT (Feng et al., 2023)	70.30	65.10	46.80	55.68	75.00	60.65
C-TPT (Yoon et al., 2024)	69.30	63.40	48.50	52.90	78.00	60.70
DynaPrompt	69.61	64.67	48.22	56.17	78.17	61.81
CoOp + TPT (Shu et al., 2022)	73.61	66.83	49.29	57.95	77.27	62.84
CoOp + DynaPrompt	**74.08**	**67.25**	**50.28**	60.55	79.15	**64.31**
MaPLe + TPT (Shu et al., 2022)	71.87	64.87	48.16	58.08	78.12	62.31
MaPLe + PromptAlign (Samadh et al., 2023)	-	65.29	50.23	59.37	79.33	63.56
MaPLe + DynaPrompt	72.71	66.34	50.25	**60.72**	**79.57**	64.22

Prompt learning을 수행했을 때 성능 향상

Labeled downstream dataset을 요구하지 않는
Test-time prompt learning을 수행했을 때
일반화 성능 향상

Test-time prompt tuning methods 대비
Online learning 수행 시 일반화 성능 향상

Experiments

Results

② Cross-Datasets Generalization

- Prompt tuning vs. Test-time prompt tuning : Arbitrary categories가 주어지는 zero-shot 상황에서의 classification 성능 확인

Method	Caltech	Pets	Cars	Flowers	Food101	Aircraft	SUN397	DTD	EuroSAT	UCF101	*Mean*
CLIP	93.35	88.25	65.48	67.44	83.65	23.67	62.59	44.27	42.01	65.13	63.58
Prompt learning methods without test-time tuning											
CoOp	93.70	89.14	64.51	68.71	85.30	18.47	64.15	41.92	46.39	66.55	63.88
CoCoOp	94.43	90.14	65.32	71.88	86.06	22.94	67.36	45.73	45.37	68.21	65.74
MaPLe	93.53	90.49	65.57	72.23	86.20	24.74	67.01	46.49	48.06	68.69	66.30
CoPrompt	94.50	90.73	65.67	72.30	86.43	24.00	67.57	47.07	**51.90**	69.73	67.00
Test-time prompt tuning methods											
TPT	94.16	87.79	66.87	68.98	84.67	24.78	65.50	47.75	42.44	68.04	65.10
DiffTPT	92.49	88.22	67.01	70.10	**87.23**	**25.60**	65.74	47.00	41.04	68.22	65.47
C-TPT	93.60	88.20	65.80	69.80	83.70	24.00	64.80	46.00	43.20	65.70	64.80
AdaPrompt	94.07	89.64	*63.29*	72.97	84.72	*21.21*	*65.37*	44.75	47.20	67.22	65.04
DynaPrompt	94.32	88.28	67.65	69.95	85.42	24.33	66.32	47.96	42.28	68.72	65.52
CoOp + TPT	93.15	89.48	66.77	68.48	86.48	20.51	66.06	43.32	37.73	68.91	64.09
CoOp + DynaPrompt	94.40	90.04	67.35	69.38	86.45	21.35	66.17	46.98	38.55	69.54	65.02
MaPLe + TPT	93.59	90.72	66.50	72.37	86.64	24.70	67.54	45.87	47.80	69.19	66.49
MaPLe + PromptAlign	94.01	90.76	**68.50**	72.39	86.65	24.80	67.54	47.24	47.86	69.47	66.92
MaPLe + DynaPrompt	**95.17**	**90.95**	68.26	**73.28**	86.60	24.36	**68.18**	**48.75**	47.53	**69.85**	**67.29**

Prompt learning을 수행했을 때 성능 향상 **+0.30%**

Labeled downstream dataset을 요구하지 않는
Test-time prompt learning을 수행 시 성능 향상 **+1.22%**

Test-time prompt tuning methods 대비
Online learning 수행 시 성능 향상 **+0.42%**

Conclusion

- **How to improve zero-shot performance of VLMs in downstream tasks?**

1. **CoOp** (2022, IJCV) : Hand-crafted prompt의 한계를 극복하고자 **prompt 자체를 최적화**하는 방법 최초 제안
   - Manual prompt engineering 없이도 적은 labeled sample 만으로도 hand-crafted prompt보다 높은 성능 달성
   - Labeled downstream training dataset을 필요로 하는 문제 → zero-shot generalization의 어려움

2. **TPT** (2022, NeurIPS) : Zero-shot generalization을 위해 test-time에 **single test sample만으로 prompt를 튜닝**하는 방법 최초 제안
   - 별도의 training data 없이 test sample 만을 사용하여 prompt를 튜닝한다는 점에서 test-time adaptation 방식으로 분류
   - Zero-shot prompt tuning의 한계인 prompt engineering dependency를 완전히 제거하고 unseen domain에도 generalization이 가능한 장점
   - 각 test sample을 독립적으로 처리하기에 test 데이터 간의 관련성, 유용한 과거 정보 등을 활용하지 못하는 한계

3. **DynaPrompt** (2025, ICLR) : 이전 test sample에서 **튜닝한 prompt를 재사용**하여 test 데이터 간 정보를 활용하는 방법 제안
   - Online test-time prompt tuning 방식에 기반하여, error accumulation으로 인한 prompt collapse 문제를 완화 가능한 알고리즘 제안
   - Prompt buffer 구조를 이용하기에 computational cost가 큰 특징을 가지며, test sample이 들어오는 순서에 따라 성능이 변동될 수 있음

Thank You