Supplementary Materials for Knowledge-Aware Prompt Tuning for Generalizable Vision-Language Models

Baoshuo Kan[1*], Teng Wang[2,3*], Wenpeng Lu[1†], Xiantong Zhen[4], Weili Guan[5], Feng Zheng[2†]
[1]Qilu University of Technology (Shandong Academy of Sciences)
[2]Southern University of Science and Technology [3]The University of Hong Kong
[4]United Imaging Healthcare [5]Monash Univerisity
10431200583@stu.qlu.edu.cn tengwang@connect.hku.hk wenpeng.lu@qlu.edu.cn
zhenxt@gmail.com weili.guan@monash.edu f.zheng@ieee.org

1. Study over Knowledge-aware Prompt

In order to investigate the impact of knowledge-aware prompts, we conducted a comparative test on the Knowledge-Aware Prompt Tuning (KAPT) method by removing the adaptation head module. The results are presented in Table 1. The findings indicate that the only use of discrete or continuous prompts significantly improves the model's performance in the new class. This finding provides evidence that incorporating category-related knowledge is beneficial for the model's ability to recognize previously unseen categories during the training stage.

Table 1: **Performance with only using knowledge-aware prompts.** We remove the adaptation head module to conduct experiments. With either dicrete or continuous prompts, the model outperforms the baseline in unseen classes. Knowledge-aware prompt achieves the best performance after combining both prompts.

Method	Base	New	H
baseline	**79.41**	64.02	70.88
w/ discrete	78.49	65.98	71.69
w/ continuous	75.40	66.50	70.67
knowledge-aware prompt	77.70	**67.24**	**72.09**

2. Few-shot Classification Results

We compare the few-shot classification performance of KAPT to CoOp[4] and CoCoOp [3] on 11 public image datasets in-domain generation setting. Table 3 shows the classification accuracy with different shot numbers. Our method KAPT outperforms previous state-of-the-art methods CoOp and CoCoOp in terms of average accuracy on 11 datasets from the 1-shot to the 3-shot setting. With the number of training examples increasing to 4, the performance of KAPT underperforms CoOp and outperforms CoCoOp.

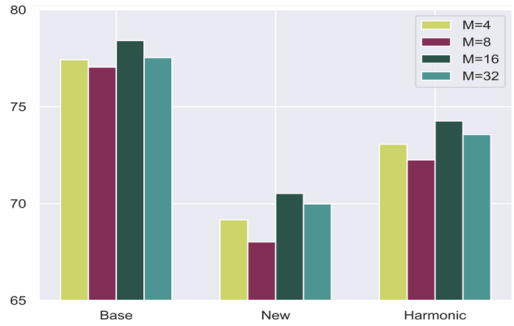

Figure 1: **Ablation study on the length of learnable continuous prompt in KAPT.** We report the average testing performance of all datasets. We found that accuracy is insensitive to the length of the learnable prompt.

We conjecture that the KAPT with external knowledge is favorable in low-data scenarios, since external information related to categories can compensate for the lack of visual information to a certain extent. When the number of samples becomes larger, image samples play a dominant role in recognizing seen classes, and the noise information existing in the external knowledge may hurt the model's performance. Note that this paper mainly targets generalization ability for unseen categories where no image samples are available, and the knowledge-based KAPT shows superiority over previous methods on the base-to-new setting consistently.

Table 2: Comparison with VPT with CLIP ViT-B/16 backbone. Models are trained on the full dataset.

Method	DTD	Euro SAT	Oxford Pets	Flowers 102	Caltech 101	Stanford Cars	Avg.
VPT-shallow	62.6	92.0	87.3	97.5	86.9	68.7	82.5
VPT-deep	65.8	**96.1**	88.3	98.0	90.8	**83.6**	87.1
KAPT	**77.3**	95.5	**93.9**	**98.3**	**96.4**	83.1	**90.7**

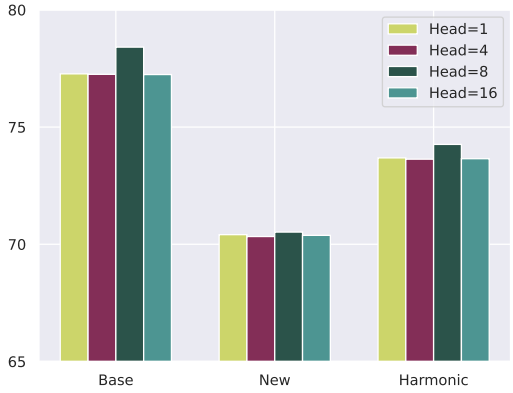

Figure 2: **Ablation study on the number of attention heads of cross attention in the adaptation head.**

3. Comparison to Visual Prompting Model

Taking into account the relevant research in the realm of prompt learning within visual models [1], we proceed to conduct a comparative assessment between KAPT and VPT, which is a model based on visual prompt tuning. This evaluation is carried out within the context of a scenario encompassing all available samples. The comparative results are presented in Table 2, showcasing the performance comparison between our KAPT and the VPT across six datasets. Remarkably, our approach demonstrates superior results on most datasets, with an average improvement of 3.6%. Notably, while VPT does not incorporate linguistic priors from image categories, the demonstrated superiority of KAPT underscores the efficacy of external knowledge integration.

4. Prompt Length

In order to further analyze the role of the length of learnable continuous prompts, we study 4, 8, 16, and 32 tokens of prompts in the base-to-new generalization setting. Fig. 1 summarizes the average results on the 11 datasets. We clearly find that there is no apparent correlation between the base, new, and harmonic mean results when the length of prompts changes.

5. Number of Attention Heads

To further analyze the role of the number of heads of cross attention in the adaptation head, we tried to change the number (i.e., 1, 4, 8, 16) of heads in the base-to-new generalization setting. Fig. 2 summarizes the average results on the 11 datasets. The results show that there is no apparent performance variance on base, new and harmonic mean when varying the number of attention heads in cross-attention.

6. Knowledge Examples

We provide examples of knowledge-aware prompts in Table 4. The main source of external knowledge is Wikidata5m [2], which is a large-scale KG dataset with aligned text descriptions from the corresponding Wikipedia pages. For each dataset, we illustrate two category samples with their Wikipedia knowledge and discrete prompts. For example, while the category A310 does not contain semantic information in its name, category-related knowledge provides beneficial language descriptions for better recognition of the current category, indicating A310 is a medium-to-long-range twin-engined wide wide-body jet airliner.

Table 5 shows category-related knowledge and discrete prompt details of the OxfordPets dataset. For fine-grained classification tasks, category-related knowledge providing unique knowledge for each category is conducive to the discriminability of learned representations.

References

[1] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision*, pages 709–727. Springer, 2022. 2

[2] Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. Kepler: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 9:176–194, 2021. 2

[3] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16816–16825, 2022. 1

[4] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, pages 1–12, 2022. 1

Table 3: **Comparison to CoOp and CoCoOp in few-shot setting.** We fine-tune KAPT and other models mentioned above on 11 datasets with few training examples in each category. Accuracy is used as the evaluated metric. **Best** and second best results are marked.

	DTD	Euro SAT	Stanford Cars	Oxford Pets	Flowers 102	Food 101	Caltech 101	SUN 397	UCF 101	FGVC Aircraft	Image Net	Average
						1-shot						
CoOp	47.66	**55.46**	60.56	86.43	71.06	74.73	91.16	63.90	66.23	15.23	60.20	62.96
CoCoOp	45.46	48.60	60.60	87.43	68.30	**79.46**	**92.43**	**65.66**	**67.56**	10.26	**64.50**	62.75
KAPT	**50.93**	46.50	**60.73**	**87.60**	**74.17**	78.07	89.63	64.50	65.90	**22.13**	62.90	**63.91**
						2-shot						
CoOp	50.16	55.80	**62.36**	86.63	77.06	76.03	91.90	65.93	**72.53**	16.30	62.80	65.23
CoCoOp	48.26	48.33	61.76	**88.93**	70.53	**80.30**	**93.13**	**66.83**	68.73	13.90	**64.50**	64.11
KAPT	**53.86**	**59.96**	61.36	88.26	**77.30**	79.30	92.53	64.90	67.23	**22.56**	63.63	**66.44**
						3-shot						
CoOp	52.53	**63.73**	**63.63**	89.20	**81.33**	78.33	92.80	**67.3**	**72.53**	17.50	63.33	67.47
CoCoOp	52.03	48.33	61.66	88.93	72.83	**80.70**	**93.43**	67.233	70.40	17.93	**65.10**	65.32
KAPT	**55.16**	60.26	62.16	**89.26**	81.03	80.26	93.10	66.13	68.80	**23.03**	64.46	**67.61**
						4-shot						
CoOp	55.53	**68.36**	**63.43**	**89.80**	**83.53**	78.33	**93.53**	**68.26**	**73.83**	18.56	63.06	**68.75**
CoCoOp	52.73	61.70	62.30	88.86	73.96	**80.63**	93.46	67.83	71.33	18.00	65.15	66.90
KAPT	**56.10**	65.16	62.80	**89.80**	82.73	80.56	93.33	66.96	69.23	**23.43**	**65.20**	68.66

Table 4: The Wikipedia knowledge and discrete prompt related to category. For each dataset, two categories are selected.

Dataset	Category	Wikipedia Knowledge	Discrete Prompt
Image Net	fire salamander	The fire salamander (Salamandra salamandra) is possibly the best-known salamander species in Europe.It is black with yellow spots or stripes to a varying degree; some specimens can be nearly completely black while on others the yellow is dominant...	fire salamander is black with yellow spots or stripes to a varying degree.
	spotted salamander	The spotted salamander or yellow-spotted salamander (Ambystoma maculatum) is a mole salamander common in the eastern United States and Canada. The spotted salamander is the state amphibian of Ohio and South Carolina. This salamander ranges from Nova Scotia...	the spotted salamander is the state amphibian of Ohio and south Carolina.
Caltech 101	scorpion	Scorpions are eight legged venomous arachnids. They have a long body with an extended tail with a sting. The average adult scorpion is about 3 inches in length. The longest scorpion is the African scorpion, which can be 9 inches...	scorpions have a long body with an extended tail with a sting.
	helicopter	A helicopter is a type of rotorcraft in which lift and thrust are supplied by horizontally-spinning rotors. This allows the helicopter to take off and land vertically, to hover, and to fly forward, backward and laterally...	a helicopter is a type of rotorcraft in which lift and thrust are supplied
Oxford Pets	abyssinian	The Abyssinian is a breed of domestic short-haired cat with a distinctive "ticked" tabby coat, in which individual hairs are banded with different colors. In nomenclature terms, they are also known as simply Abys...	the abyssinian is a breed of domestic short-haired cat
	boxer	The Boxer is a medium to large, short-haired breed of dog, developed in Germany. The coat is smooth and tight-fitting; colors are fawn, brindled, or white, with or without white markings. Boxers are ...	the boxer is a medium to large, short -haired breed of dog
Stanford Cars	2012 Hyundai Accent Sedan	The Hyundai Accent, or Hyundai Verna in South Korea, is a subcompact car produced by Hyundai. In Australia, the first generation models carried over the Hyundai Excel name used by the Accent's predecessor....	the Hyundai Accent is a subcompact car produced by Hyundai
	1994 Audi 100 Wagon	The Audi 100 and Audi 200 are four-door, front-engine, front- or all-wheel drive full-size/executive sedans manufactured and marketed by the Audi division of the Volkswagen Group for model years 1968 through...	the Audi 100 and Audi 200 are four-door full-size/executive sedans
Flowers 102	globe thistle	Echinops is a genus of about 120 species of flowering plants in the daisy family Asteraceae, commonly known as globe thistles. They have spiny foliage and produce blue or white spherical flower heads...	globe thistles have spiny foliage and produce blue or white spherical flower
	bougainvillea	Bougainvillea is a genus of thorny ornamental vines, bushes, or trees. It is native to Eastern South America, from Brazil, west to Peru, and south to southern Argentina. Different authors...	Bougainvillea is a genus of thorny ornamental vines
Food 101	falafel	Falafel is a kind of vegetarian food. Falafel is made from chickpeas or fava beans. Spices are added to these. Falafel is either made into a ball, or a disc, which is then fried. The dish originally came..	alafel is a kind of vegetarian food, made from chickpeas or
	guacamole	Guacamole is a Mexican dip made of avocados, salt and sometimes tomatoes. Various herbs can be added too, such as chives. It is depicted as a green, creamy dip or garnish. It is often served as a dip or garnish with...	guacamole is a dip made of avocados, salt and sometimes tomatoes
FGVC Aircraft	A310	he Airbus A310 is a medium- to long-range twin-engined wide-body jet airliner that was developed and manufactured by Airbus, then a consortium of European aerospace manufacturers. It was the second airliner to be produced by the...	the Airbus A310 is a medium- to long-range twin-engined wide wide-body jet airliner
	A321	The Airbus A321 is a member of the Airbus A320 family of short- to medium-range, narrow-body, commercial passenger twin-engine jet airliners. The first derivative of the baseline A320, it has a stretched fuselage and entered service...	the airbus A321 is a member of the Airbus A320 family .
SUN397	heliport	A heliport is a small airport only for use by helicopters. Heliports normally contain one or more helipads and may have limited facilities such as fuel, lighting, a windsock, or even hangars. In larger towns and cities, customs...	heliports contain one or more helipads and may have limited facilities
	outdoor observatory	An observatory is a location used for observing terrestrial or celestial events. Astronomy, climatology/meteorology, geophysical, oceanography and volcanology are examples of disciplines for which observatories have been...	an observatory is a location used for observing terrestrial or celestial events .
DTD	cracked	Broken so that cracks appear on, or under, the surface.	broken so that cracks appear on, or under, the surface .
	grid	A rectangular array of squares or rectangles of equal size, such as in a crossword puzzle.	rectangular array of squares or rectangles of equal size, such as in a crossword
Euro SAT	Highway or Road	A road is a thoroughfare, route, or way on land between two places that has been paved or otherwise improved to allow travel by foot or some form of conveyance, including a motor vehicle, cart, bicycle, or...	a road is a thoroughfare, route, or way on land between two places
	Industrial Buildings	Industrial architecture is the design and construction of buildings serving industry. Such buildings rose in importance with the Industrial Revolution, and were some of the pioneering structures of modern...	industrial architecture is the design and construction of buildings serving industry .
UCF 101	High Jump	The high jump is a track and field event in which competitors must jump unaided over a horizontal bar placed at measured heights without dislodging it. In its modern most practiced format, a bar...	high jump is one of two vertical clearance events to feature on the Olympic athletics programme .
	Balance Beam	The balance beam is a rectangular artistic gymnastics apparatus, as well as the event performed using the apparatus. Both the apparatus and the event are sometimes simply referred to as "beam"...	the balance beam is a small, thin beam which is typically raised from the floor .

Table 5: Category-related knowledge of the OxfordPets dataset.

Category	Wikipedia Knowledge	Discrete Prompt
abyssinian	The Abyssinian is a breed of domestic short-haired cat with a distinctive ticked tabby coat, in which individual hairs are banded with different colors. In nomenclature terms, they are also known as simply Abys…	the abyssinian is a breed of domestic short-haired cat,
american bulldog	The American Bulldog is a large breed of utility dog descended from the now -extinct. Old English Bulldog. They are now used on animal farms, dogsports, and for showing. They are part of American culture and history, and may be…	the American Bulldog is a large breed of utility dog . they are descended from ,
american pit bull terrier	The American Pit Bull Terrier is a purebred dog breed recognized by the United Kennel Club and American Dog Breeders Association, but not the American Kennel Club. It is a medium-sized, intelligent…	the American Pit Bull Terrier is a purebred dog breed recognized by the ,
basset hound	The Basset Hound is a short-legged breed of dog in the hound family. The Basset is a scent hound that was originally bred for the purpose of hunting hare.Their sense of smell and ability to ground-scent…	the Basset is a scent hound that was originally bred for the purpose of ,
beagle	The beagle is a breed of small hound that is similar in appearance to the much larger foxhound. The beagle is a scent hound, developed primarily for hunting hare. Possessing a great sense of smell…	the beagle is a small hound developed primarily for hunting hare
bengal	Bengal is a geopolitical, cultural and historical region in South Asia, specifically in the eastern part of the Indian subcontinent at the apex of the Bay of Bengal. Geographically, it is made up by the Ganges-Brah…	Bengal is a geopolitical, cultural and historical region in south Asia.
birman	The Birman, also called the Sacred Cat of Burma, is a domestic cat breed. The Birman is a long-haired, color-pointed cat distinguished by a silky coat, deep blue eyes, and contrasting…	the Birman is a long-haired,colour-pointed cat distinguished
bombay	The Bombay cat is a type of short-haired cat developed by breeding sable Burmese and black American Shorthair cats, to produce a cat of mostly Burmese type, but with a sleek, panther-like...	the Bombay cat is a type of short-haired cat developed by breeding ,
boxer	The Boxer is a medium to large, short-haired breed of dog, developed in Germany. The coat is smooth and tight-fitting; colors are fawn, brindled, or white, with or without white markings. Boxers are...	the boxer is a medium to large, short-haired breed of dog,
british shorthair	The British Shorthair is the pedigreed version of the traditional British domestic cat, with a distinctively stocky body, dense coat, and broad face. The most familiar color variant is the British Blue,…	the British Shorthair is the pedigreed version of the traditional domestic cat .
chihuahua	The Chihuahua is the smallest breed of dog, and is named after the Mexican state of Chihuahua.	the Chihuahuan is the smallest breed of dog, and is named
egyptian mau	Egyptian Maus are a small- to medium-sized short-haired cat breed. They are one of the few naturally spotted breeds of domesticated cat. The spots of the Mau occur on only the tips of the hairs of its coat. It is...	Egyptian Maus are a small- to medium-sized short-haired cat breed
english cocker spaniel	The English Cocker Spaniel is a breed of gun dog. It is noteworthy for producing one of the most varied numbers of pups in a litter among all dog breeds. The English Cocker Spaniel is an active, good-natured...	the English Cocker Spaniel is a breed of gun dog . it produces one of
english setter	The English Setter is a medium-size breed of dog. It is part of the setter group, which includes the red Irish Setters, Irish Red and White Setters, and black -and-tan Gordon Setters. The mainly white body coat...	the English Setter is a medium-size breed of dog . it is part of
german shorthaired	not found	not found
great pyrenees	not found	not found
havanese	Havanese is the national dog of Cuba, developed from the now extinct Blanquito de la Habana. The Blanquito descended from the also now-extinct Bich Tenerife. It is believed that the Blanquito was eventually cross-bred…	the Havanese is the national dog of Cuba, developed from the now extinct ,
japanese chin	The Japanese Chin, also known as the Japanese Spaniel, is a dog acknowledged for its importance to the Japanese nobility. It is also known for its strabismus of the eyes. Being both a lap dog and a...	the Japanese Chin, also known as the Japanese Spaniel, is a dog acknowledged for
keeshond	The Keeshond is a medium-sized dog with a plush, two-layer coat of silver and black fur with a ruff and a curled tail. It originated in Holland, and its closest relatives are the German spitzes such as...	the Keeshond is a medium-sized dog with a plush
leonberger	The Leonberger is a giant dog breed, whose name derives from the city of Leonberg in Baden-Wcrttemberg, Germany. According to legend, the Leonberger was ostensibly bred as a symbolic dog that would mimic the lion in ...	the Leonberger is a giant dog breed,whose name derives from the city of
maine coon	The Maine Coon is the largest domesticated cat breed. It has a distinctive physical appearance and valuable hunting skills. It is one of the oldest natural breeds in North America,specifically native to the US…	the Maine Coon is the largest domesticated cat breed in the u.s. it
miniature pinscher	The Miniature Pinscher, also known as the Zwergpinscher, and Min Pin, is a small breed of dog of the pinscher type originating from Germany. The breed's earliest ancestors may have included the…	the Miniature Pinscher, also known as the Zwergpinscher, and Min
newfoundland	The Newfoundland dog is a large working dog. They can be either black, brown, or white-and-black. However, in the Dominion of Newfoundland, before it became part of the confederation...	the Newfoundland dog is a large working dog . they can be either black
persian	The Persian cat is a long-haired breed of cat characterized by its round face and short muzzle. It is also known as the Persian Longhair in English-speaking countries. The first …	the Persian cat is a long-haired breed of cat characterized by its round

pomeranian	The Pomeranian is a breed of dog of the Spitz type that is named for the Pomerania region in north-west Poland and north-east Germany in Central Europe. Classed as a toy dog...	the pomeranian is a breed of dog of the Spitz type.
pug	The pug is a type of dog with a wrinkly face. It also has a curled tail, and pug puppies are called piglets. The pug has a square, muscular body with a large head, big. . .	the pug is a type of dog with a wrinkly fac.
ragdoll	The Ragdoll is a cat breed with a color point coat and blue eyes. They are large and muscular semi-longhair cats with a soft and silky coat. Developed by American ...	the Ragdoll is a cat breed with a color point coat and blue eyes
russian blue	The Russian Blue is a cat breed that comes in colors varying from a light shimmering silver to a darker, slate grey.Their short, dense coat has been...	their short, dense coat has been the hallmark of the Russian breed for more than a century
saint bernard	not found	not found
samoyed	The Samoyed is a breed of medium-sized herding dogs with thick, white, double-layer coats. They are related to the laika, a spitz-type dog...	the Samoyed is a breed of medium-sized herding dogs with thick,
scottish terrier	The Scottish Terrier, popularly called the Scottie, is a breed of dog. Initially one of the highland breeds of terrier that were grouped under the name of Skye Terrier, it...	the Scottish Terrier, popularly called the Scottie, is a breed of dog
shiba inu	The Shiba Inu is a Japanese breed of hunting dog. A small-to-medium breed, it is the smallest of the six original and distinct spitz breeds of dogs native to Japan. A small, alert and agile dog that...	the Shiba Inu is a small-to-medium breed of hunting dog
siamese	The Siamese cat is one of the first distinctly recognized breeds of Asian cat.Derived from the Wichianmat landrace, one of several varieties of cat native to Thailand, the original Siamese became...	the Siamese cat is one of the first distinctly recognized breeds of Asian cat
sphynx	The Sphynx cat is a breed of cat known for its lack of coat. Hairlessness in cats is a naturally occurring genetic mutation; however, the Sphynx cat, as a breed, was developed through selective...	the Sphynx cat is a breed of cat known for its lack of coat
staffordshire bull terrier	The Staffordshire Bull Terrier is a British breed of short-haired terrier of medium size. It originated in the city of Birmingham and in the Black Country of Staffordshire, it is the . . .	the Staffordshire Bull Terrier is a short-haired terrier of medium
wheaten terrier	not found	not found
yorkshire terrier	The Yorkshire Terrier is one of the smallest dog breed of terrier type, and of any dog breed. The breed developed during the 19th century in Yorkshire, England. Ideally its maximum size...	the Yorkshire Terrier is one of the smallest dog breed of terrier type .