JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

ProD: Prompting-to-disentangle Domain Knowledge for Cross-domain Few-shot Image Classification

Tianyi Ma, Yifan Sun, Zongxin Yang, Yi Yang

University of Technology Sydney

Baidu Inc., Zhejiang University

UTS

Cross-Domain Few-Shot Image Classification

- Training the model on one / multiple training domain (s).

- When inference, tuning the model with limited samples (i.e. 5 or 25) from a different domain.

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Key Problems of the Task

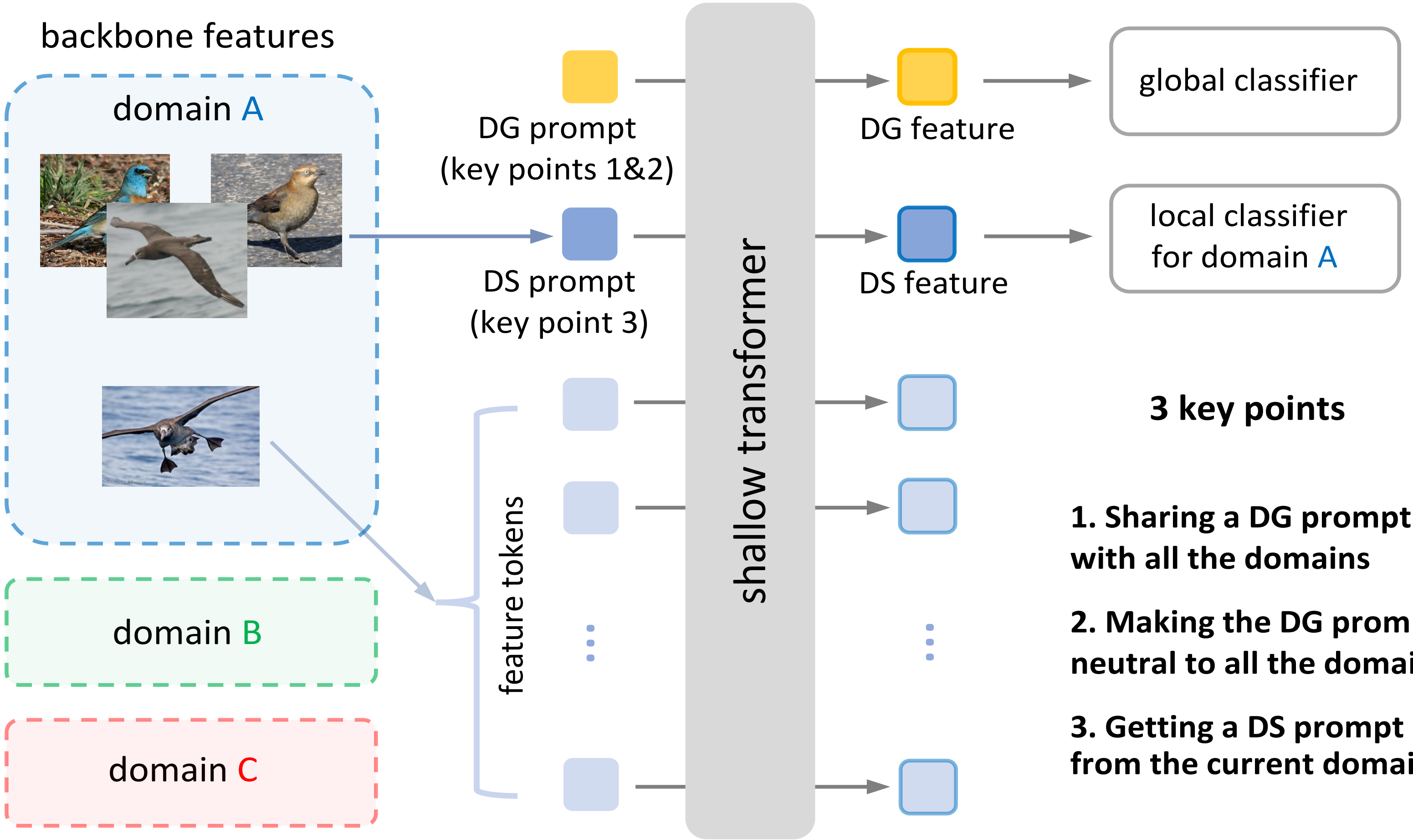
- **Domain Generalization**: A more general model that absorbs domain-general knowledge from the training domain (s) effectively.

- **Domain Adaptation**: A model that is easy to adapt to a novel domain with only limited samples for finetuning.

UTS

Our Solution: *Prompting-to-Disentangle*

- We take advantage of domain-general knowledge and domain-specific knowledge in regard to generalization and adaptation problems, respectively, with the Domain-General (DG) and Domain-Specific (DS) Prompts.



3 key points

1. Sharing a DG prompt with all the domains

2. Making the DG prompt neutral to all the domains

3. Getting a DS prompt from the current domain

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

UTS

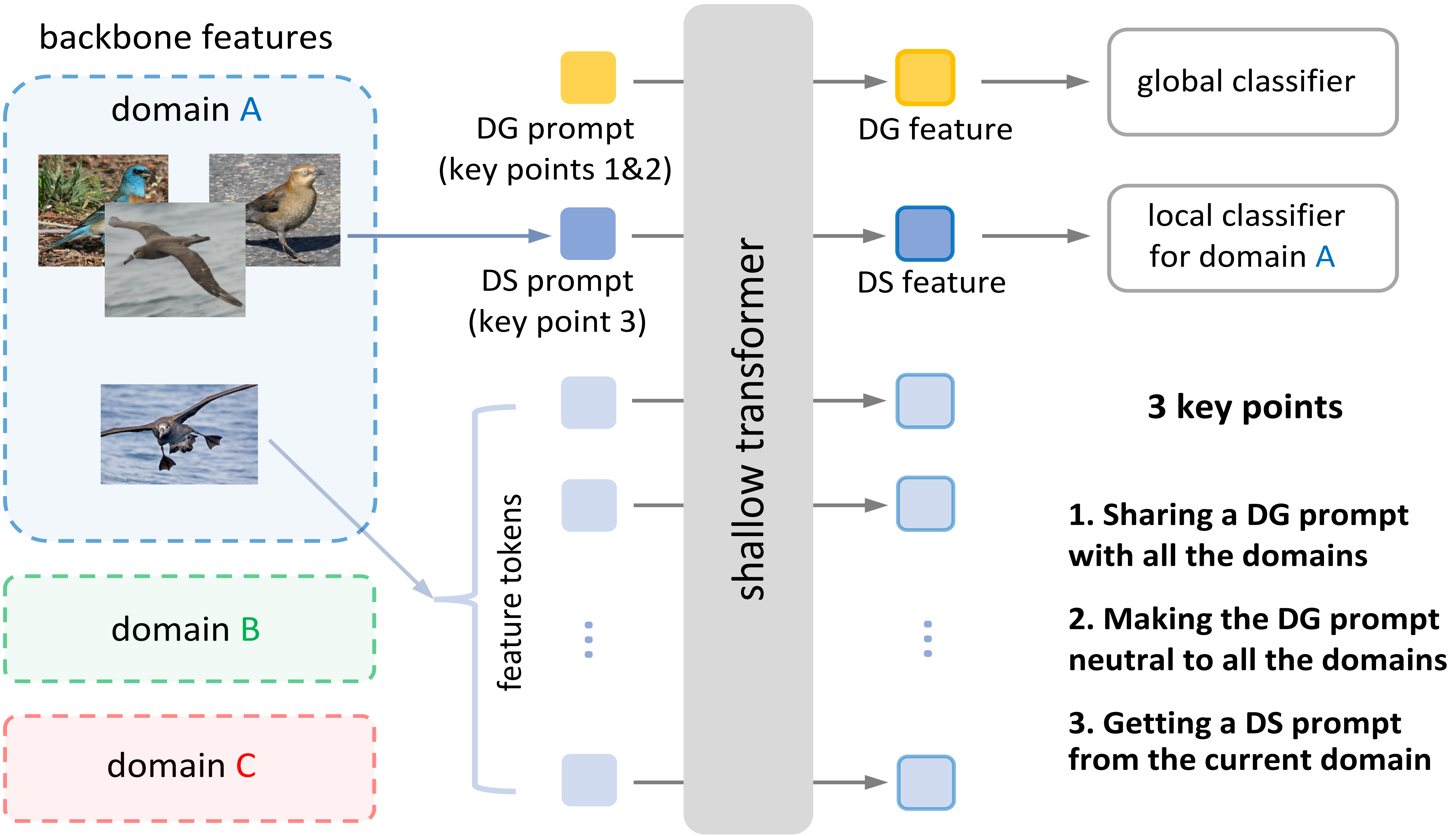
Visual Prompt

Standard Prompt

- Prompts are the vectors that are attached to the input features to modify the mapping of the pre-trained model.
- Different prompts are trained regarding different downstream tasks.

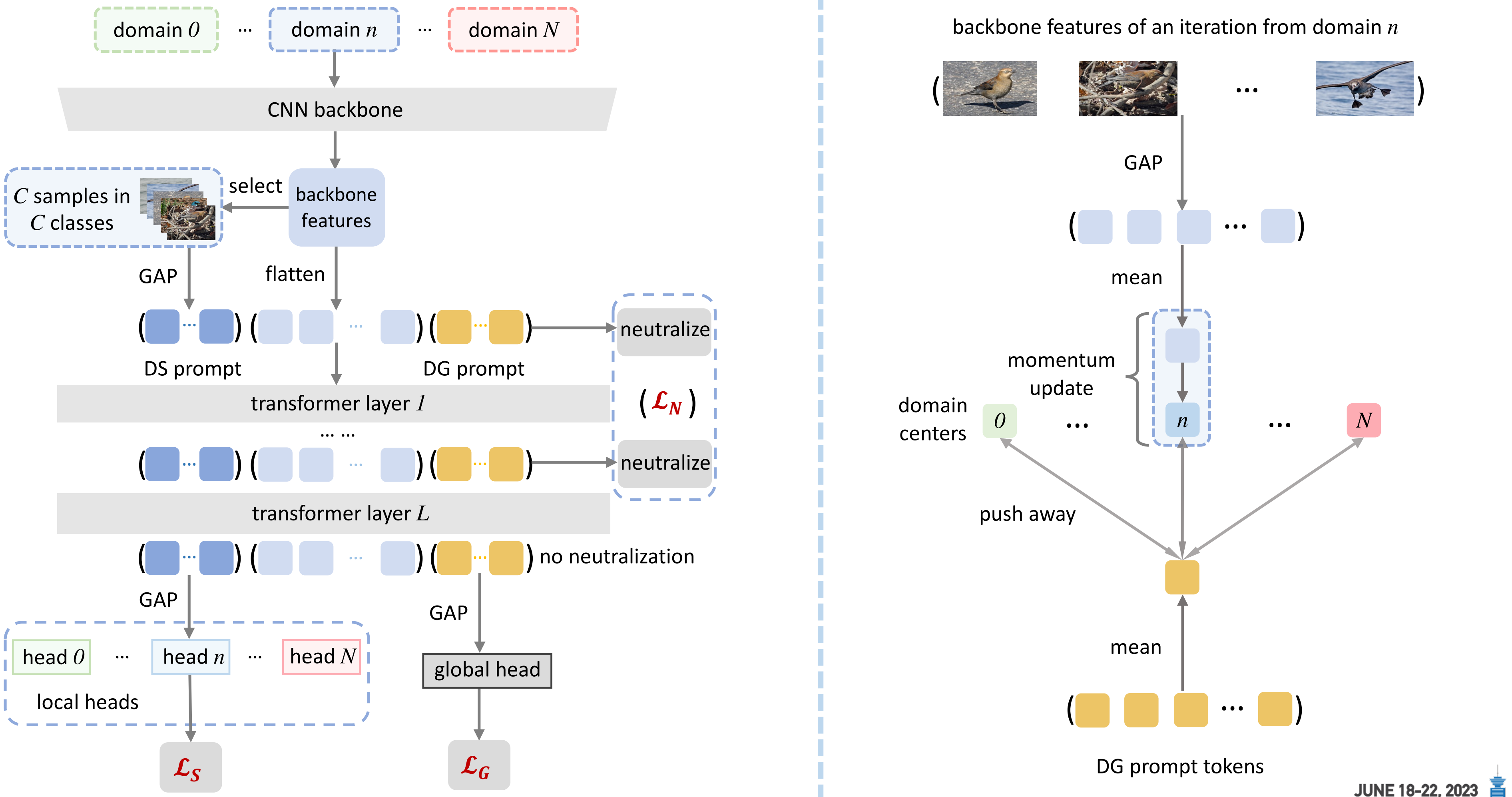
Prompt in ProD

- Prompts vectors and the full model are trained simultaneously in the training phase.
- Trainable prompt parameters are fixed during the inferences phase
- DS and DG prompts are trained for the same classification task by absorbing domain-general and domain-specific knowledge, respectively.



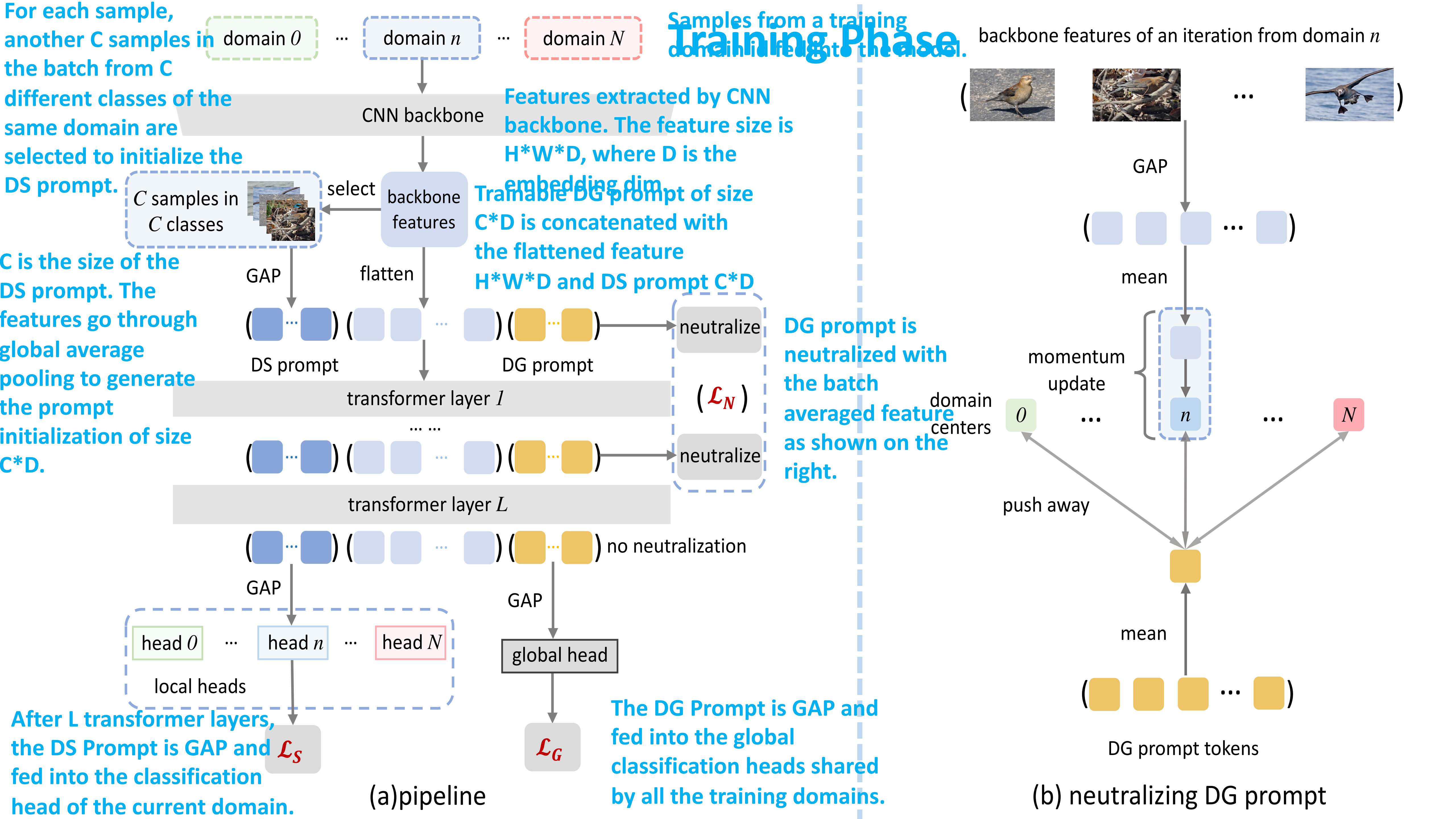
backbone features
domain A
domain B
domain C
DG prompt
(key points 1&2)
DS prompt
(key point 3)
feature tokens
shallow transformer
DG feature
DS feature
global classifier
local classifier
for domain A
3 key points
1. Sharing a DG prompt
with all the domains
2. Making the DG prompt
neutral to all the domains
3. Getting a DS prompt
from the current domain

Model Overview



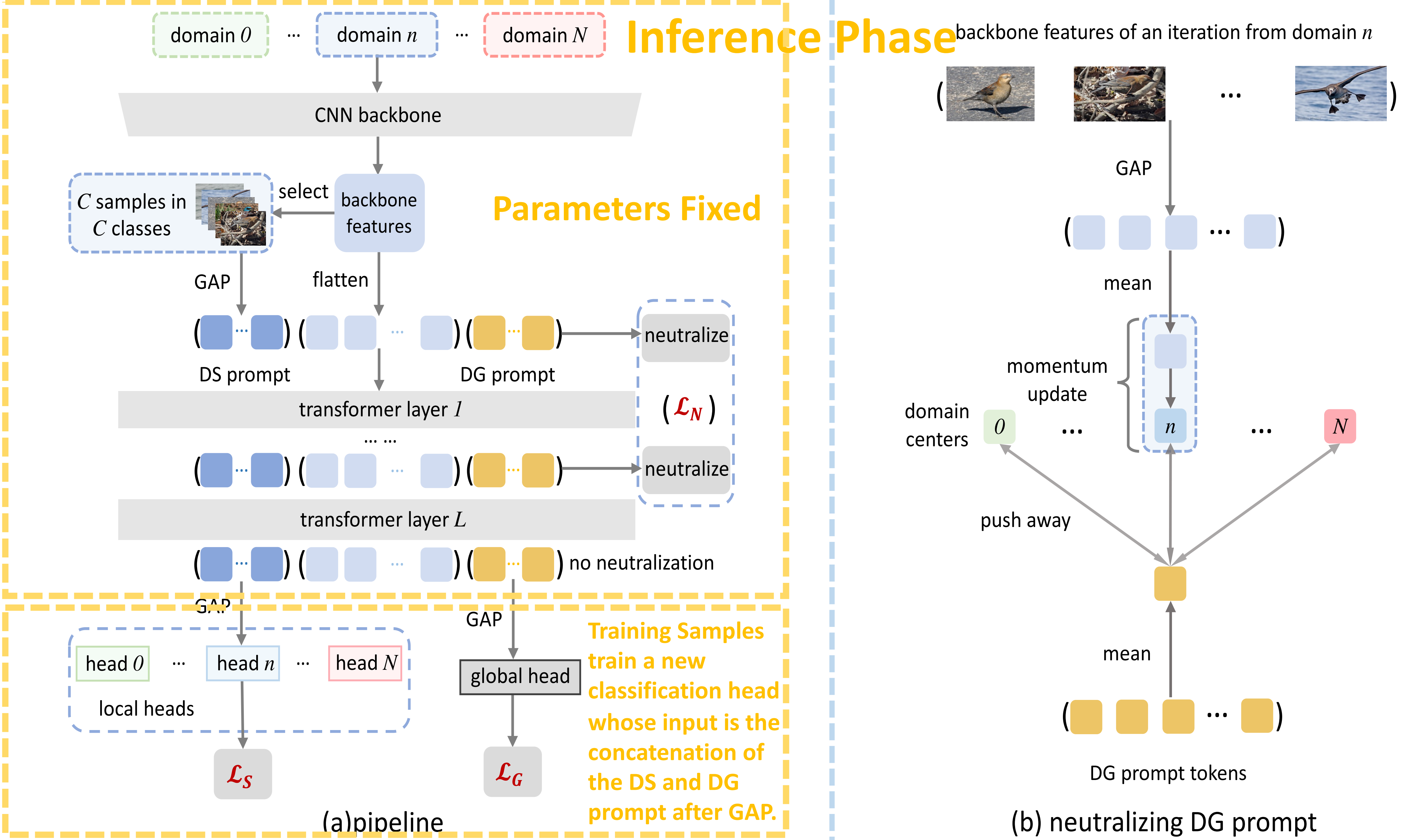
(a)pipeline

(b) neutralizing DG prompt



(a)pipeline

(b) neutralizing DG prompt



Inference Phase
domain 0
domain n
domain N
CNN backbone
C samples in C classes
select
backbone features
Parameters Fixed
GAP
flatten
neutralize
DS prompt
DG prompt
transformer layer 1
($\mathcal{L}_N$)
neutralize
transformer layer L
no neutralization
GAP
GAP
head 0
head n
head N
local heads
global head
Training Samples train a new classification head whose input is the concatenation of the DS and DG prompt after GAP.
$\mathcal{L}_S$
$\mathcal{L}_G$
(a)pipeline
backbone features of an iteration from domain n
GAP
mean
momentum update
domain centers
0
n
N
push away
mean
DG prompt tokens
(b) neutralizing DG prompt

Overall Loss

- Training Phase Loss:
  - $\mathcal{L} = \mathcal{L}_N + \mathcal{L}_G + \mathcal{L}_S$
- Inference Phase Tuning Loss:
  - $\mathcal{L} = \mathcal{L}_G + \mathcal{L}_S$
  - Inference Phase, $\mathcal{L}_G$ and $\mathcal{L}_S$ are generated from new classification heads

UTS

Effectiveness of ProD

Methods	CUB	CARS	Plantae	Places
RelationNet [28]	35.21 ± 0.46	30.12 ± 0.49	31.99 ± 0.51	49.79 ± 0.57
MatchingNet [32]	42.28 ± 0.61	28.91 ± 0.56	33.02 ± 0.56	48.53 ± 0.62
RelationNet+LFT [29]	48.10 ± 0.62	32.26 ± 0.58	35.21 ± 0.59	51.02 ± 0.56
MatchingNet+LFT [29]	43.38 ± 0.58	30.68 ± 0.59	35.10 ± 0.54	52.63 ± 0.55
RelationNet+ATA [35]	48.49 ± 0.61	31.92 ± 0.58	33.62 ± 0.49	51.00 ± 0.50
DSL [14]	50.15 ± 0.80	37.13 ± 0.69	41.17 ± 0.80	53.16 ± 0.88
Baseline	48.56 ± 0.72	33.15 ± 0.64	37.94 ± 0.71	49.81 ± 0.69
ProD	$\mathbf{53.97 \pm 0.71}$	$\mathbf{38.02 \pm 0.63}$	$\mathbf{42.86 \pm 0.59}$	$\mathbf{53.92 \pm 0.72}$

Table 1. Comparison with the state of the arts on 5-way 1-shot task.

Methods	CUB	CARS	Plantae	Places
RelationNet	51.10 ± 0.62	38.26 ± 0.58	62.99 ± 0.62	46.01 ± 0.57
MatchingNet	57.21 ± 0.63	36.98 ± 0.56	62.83 ± 0.62	43.68 ± 0.55
RelationNet+LFT	65.02 ± 0.55	43.51 ± 0.51	50.48 ± 0.46	67.34 ± 0.52
MatchingNet+LFT	61.44 ± 0.56	43.12 ± 0.52	48.49 ± 0.51	65.09 ± 0.48
RelationNet+ATA	59.42 ± 0.48	42.99 ± 0.42	45.51 ± 0.51	67.10 ± 0.41
NSAE [21]	68.17 ± 0.54	54.77 ± 0.56	59.51 ± 0.55	70.93 ± 0.54
DSL	73.57 ± 0.65	58.53 ± 0.73	62.10 ± 0.75	74.10 ± 0.72
Baseline	72.32 ± 0.77	53.17 ± 0.71	60.05 ± 0.69	69.13 ± 0.60
ProD	$\mathbf{79.19 \pm 0.59}$	$\mathbf{59.49 \pm 0.68}$	$\mathbf{65.82 \pm 0.65}$	$\mathbf{75.00 \pm 0.72}$

Table 2. Comparison with the state of the arts on 5-way 5-shot task.

UTS

Effectiveness of ProD

Methods	ChestX	ISIC	EuroSAT	CropDisease
Transductive Ft [6]	26.79	49.68	81.76	90.64
ConFeSS [3]	27.09	48.85	84.65	88.88
RDC-FT [9]$^{-}$	25.48	49.06	84.67	**93.55**
ProD	**28.79**	**50.57**	**85.09**	90.41

Table 4. Comparison with the state of the arts on 5-way 5-shot task on newly proposed datasets.

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

UTS

Ablations

Effectiveness of DG prompt

Effectiveness of DS prompt

Effectiveness of DG prompt neutralization

Methods	CUB 1-shot	CUB 5-shot
Basel.	48.56 ± 0.59	72.32 ± 0.67
Basel. + DG	51.89 ± 0.63	75.12 ± 0.69
Basel. + DS	51.48 ± 0.71	74.91 ± 0.68
Basel. + DG + DS	52.69 ± 0.66	77.63 ± 0.74
Basel. + DG + $\mathcal{L}_N$	53.08 ± 0.74	78.65 ± 0.68

Table 3. Evaluation of key components: DG prompt (DG), neutralizing loss ($\mathcal{L}_N$), and DS prompt (DS).

Effectiveness of local classification for DS prompt

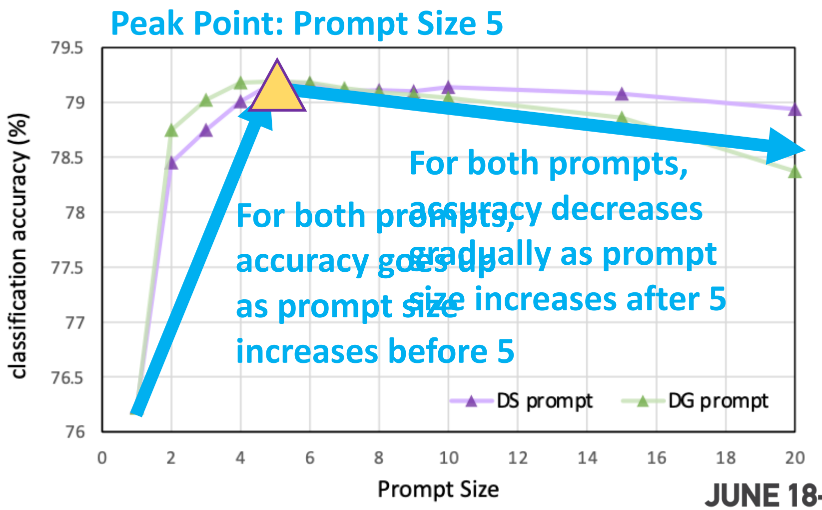
Methods	CUB 1-shot	CUB 5-shot
Basel.	48.56 ± 0.59	72.32 ± 0.67
Basel. + DS (global)	50.39 ± 0.71	73.87 ± 0.66
Basel. + DS (local)	51.48 ± 0.71	74.91 ± 0.68
ProD (global)	52.08 ± 0.74	77.65 ± 0.68
ProD (local)	53.97 ± 0.71	79.19 ± 0.63

Table 4. Comparison between the local and global classification heads on the DS prompt.

Inference Input	CUB 1-shot	CUB 5-shot
Feature Token	51.51 ± 0.72	76.13 ± 0.68
DG	53.01 ± 0.74	78.17 ± 0.61
DS	52.07 ± 0.69	77.64 ± 0.63
DG+DS	53.97 ± 0.71	79.19 ± 0.63
DG+DS+Feature Token	52.18 ± 0.75	78.04 ± 0.72

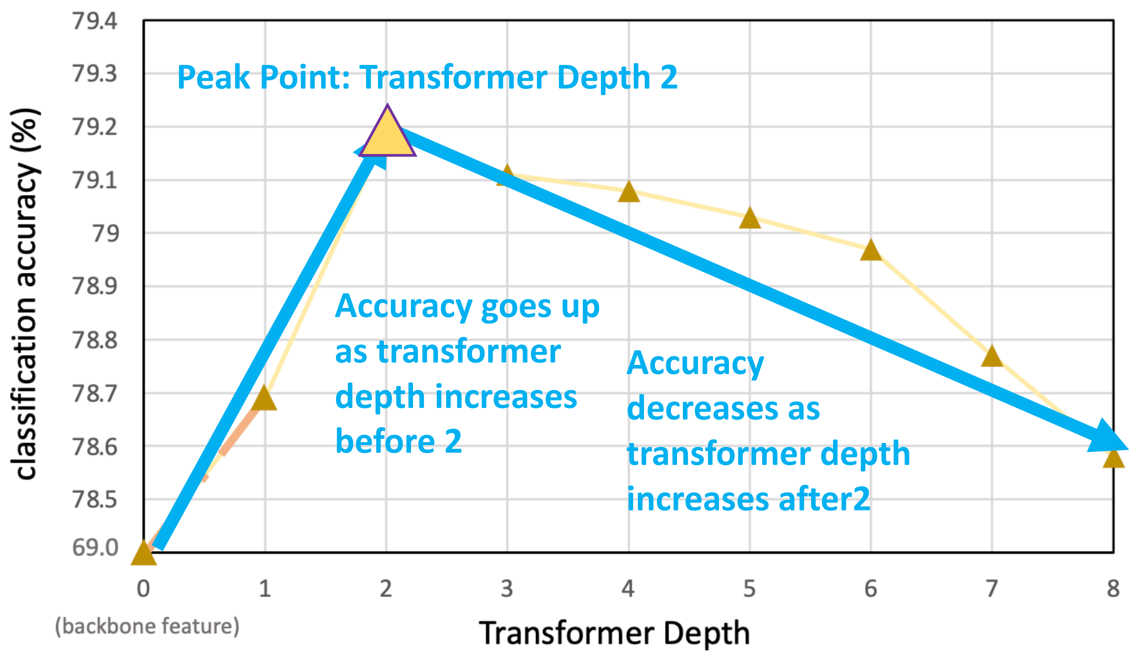
Using DG and DS prompt output only for inference achieves the highest accuracy

Table 5. Comparison between different features for inference with a complete ProD model.



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Ablations



Peak Point: Transformer Depth 2
Accuracy goes up as transformer depth increases before 2
Accuracy decreases as transformer depth increases after2
79.4
79.3
79.2
79.1
79
78.9
78.8
78.7
78.6
78.5
69.0
classification accuracy (%)
0
1
2
3
4
5
6
7
8
(backbone feature)
Transformer Depth

Computational Cost

Method	Size	CUB 5-shot
Res10	5.3M	68.98 ± 0.81
Res18	11.7M	72.39 ± 0.84
Basel. (Res10 + Trans)	8.5M	72.32 ± 0.77
ProD (Res10 + Trans + Prompt)	8.6M	79.19 ± 0.59

Table 6. Analysis of the computational efficiency. "Res10", "Res18" and "Trans" denote ResNet-10, ResNet-18 and the transformer head, respectively.

Thank You