Zero Shot Domain Adaption

POSIDA-Prompt Driven Synthesis Image for Zero-shot Domain Adaption

패턴인식 – A조 노성현 , 이재웅

1 ABSTRACT – Problem

How to train the model in unrealistic , rare event in real situation ?

1. Data Dependency issue

2. Technical limitation

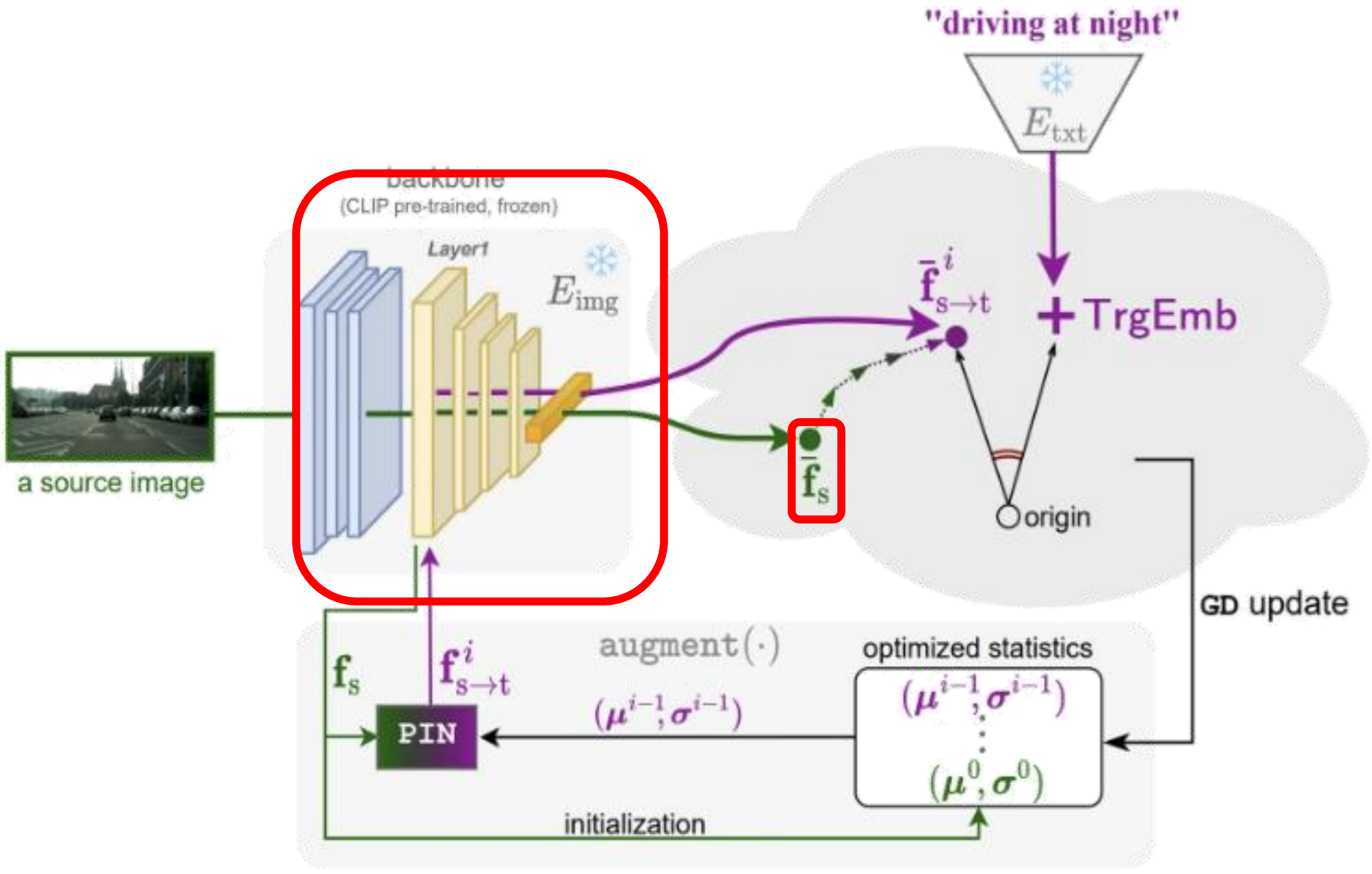
DOMAIN ADAPTION

1 ABSTRACT

Zero shot domain adaption

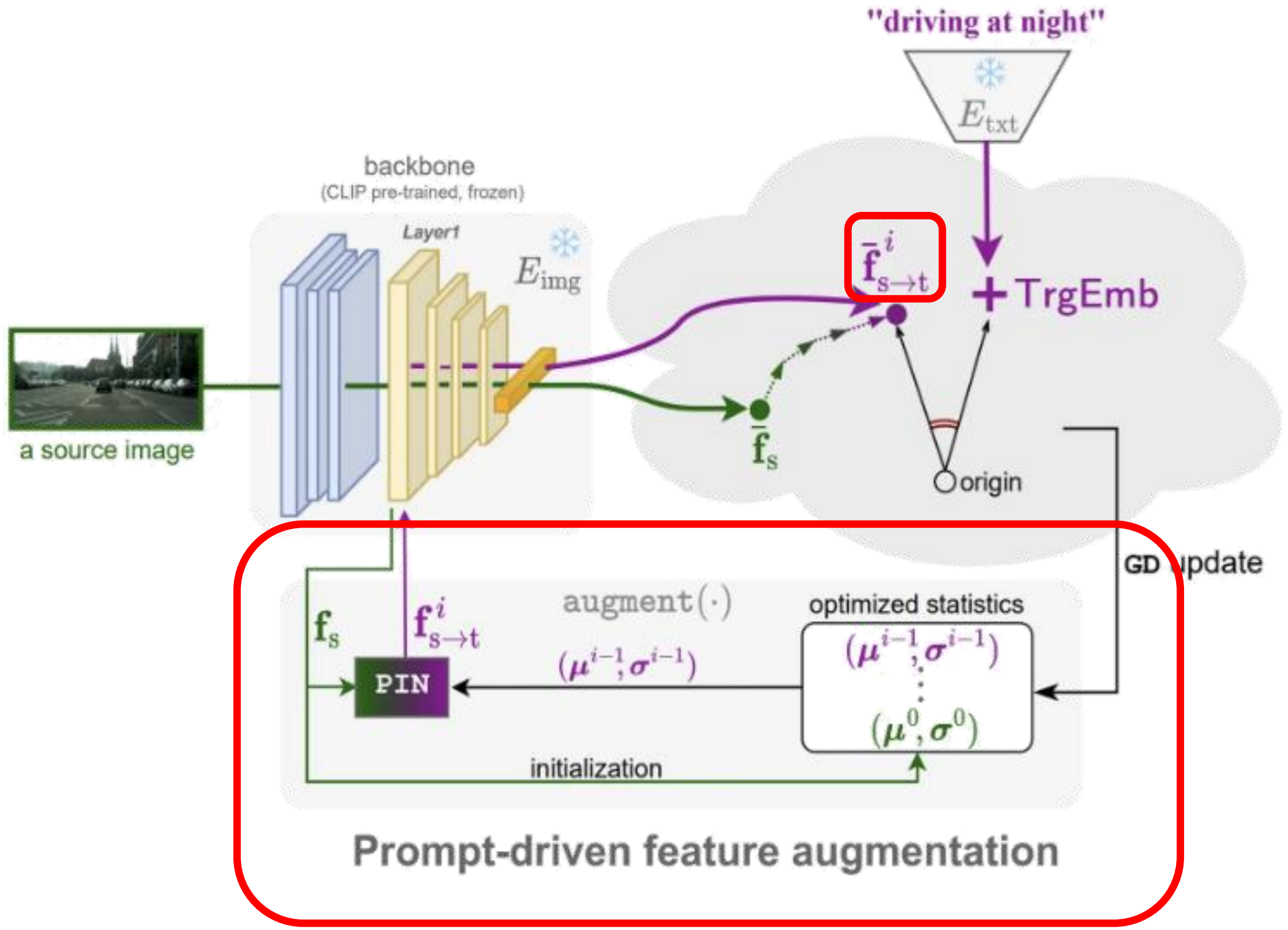
- Technology that utilizes knowledge from the learned source domain to generalize to the untrained target domain
- Improve the safety and efficiency of autonomous vehicles
- Contributes to the successful application of autonomous driving in a wider range of real-world environments

2 base model



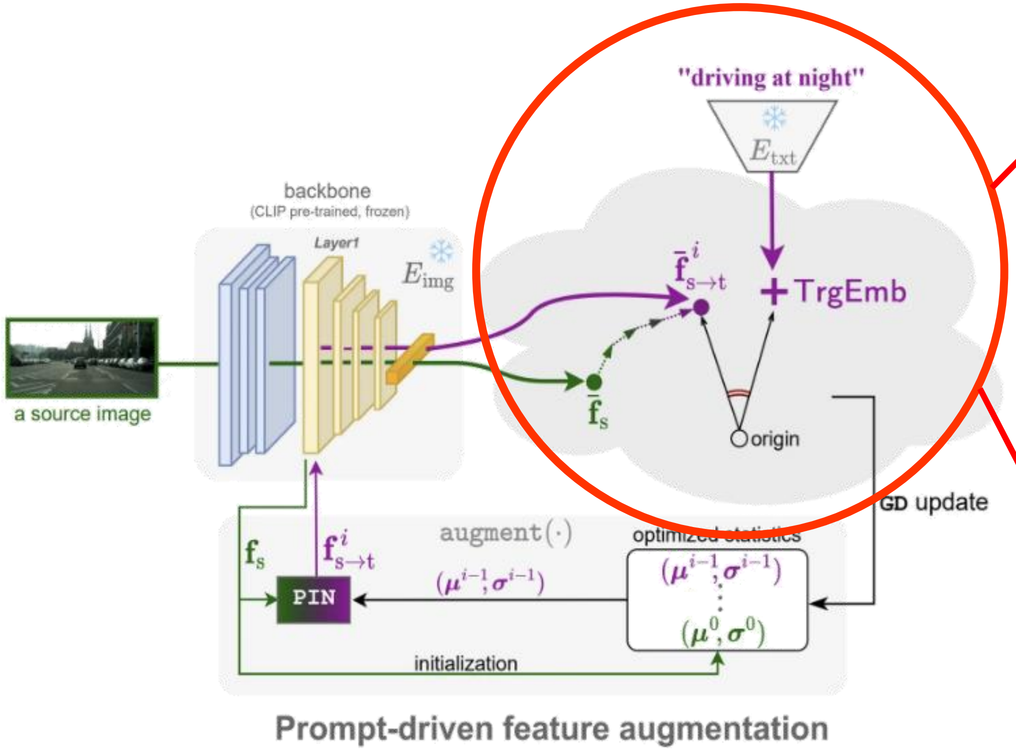
Prompt-driven feature augmentation

2 base model



"driving at night"
E_{txt}
backbone
(CLIP pre-trained, frozen)
Layer1
E_{img}
a source image
$\bar{f}^i_{s\to t}$
+ TrgEmb
$\bar{f}_s$
origin
GD update
f_s
$f^i_{s\to t}$
augment(·)
optimized statistics
$(\mu^{i-1}, \sigma^{i-1})$
$(\mu^{i-1}, \sigma^{i-1})$
(μ^0, σ^0)
PIN
initialization
Prompt-driven feature augmentation

2 base model - problem



Limitation embedding text

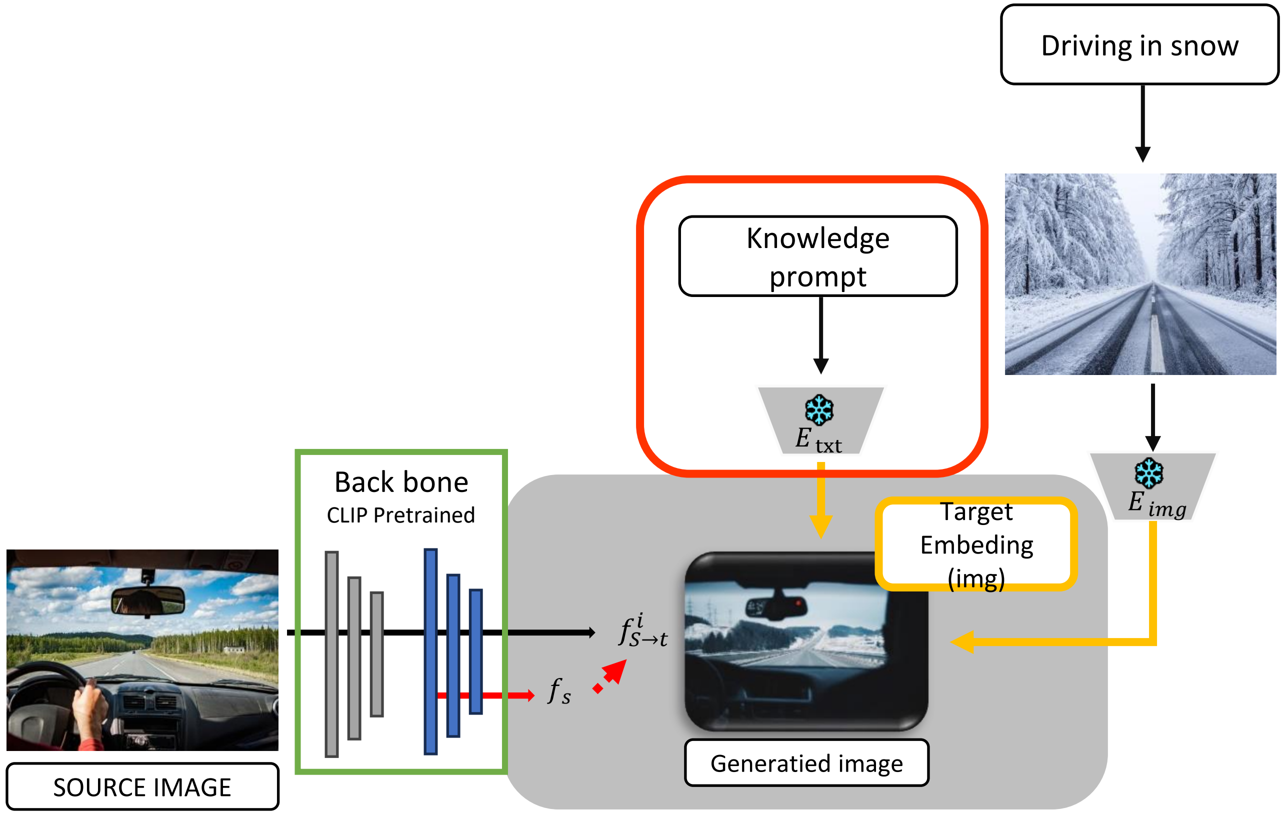
- Performance depending on prompt
- Only text prompt has limitation

Domain gap exist

So add Knowledge based prompt & Synthesis image

3 POSIDA

Prompt Driven Synthesis Image for Zero-shot Domain Adaption



Knowledge based prompt

- Weather-specific and driving situation-specific prompts for optimal model performance

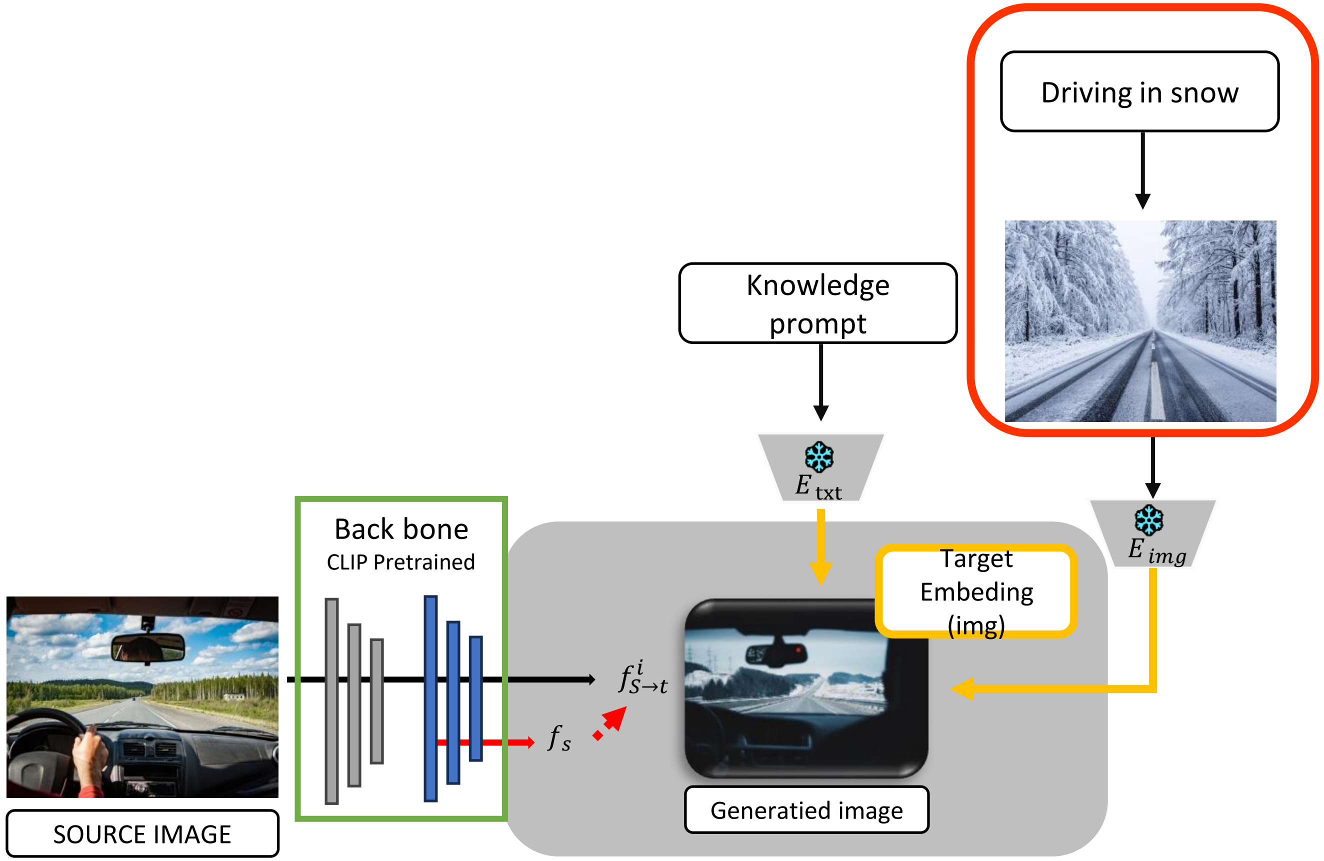
Ex : a photo of {SNOW}

↓

an open highway with {SNOW}

3 POSIDA

Prompt Driven Synthesis Image for Zero-shot Domain Adaption



Use Synthesized image

- In embedding space information of prompt is not enough
- Use Synthesized image information to increase overall performance

4 Data set

Source domain : Cityscapes

30 classes

(road, sidewalk, rail track ,car ,trailer building, building….etc)

50 cities

Several season

Daytime

Good weather condition

11GB

4 Data set

Target domain : ACDC

19 classes

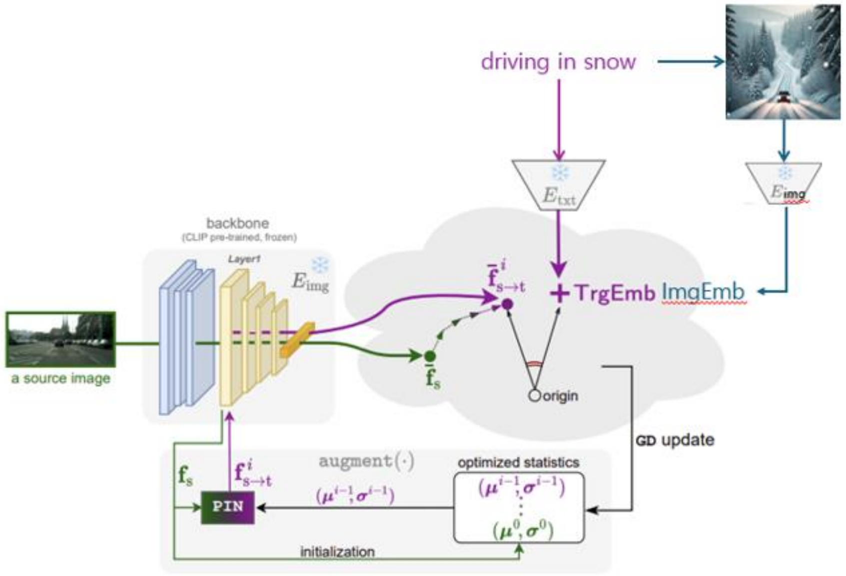
(road, sidewalk, rail track ,car ,trailer
building, building⋯.etc)

Data split

1000foggy, 1006 nighttime, 1000 rainy and 1000 snowy

17GB

4 Evaluation metric - MIOU



$$IOU = \frac{Correct\ AREA}{Whole\ AREA} \approx MIOU = \frac{1}{n}\sum_{k=1}^{n}\frac{Correct\ AREA}{Whole\ AREA}$$

5 Experience result 1

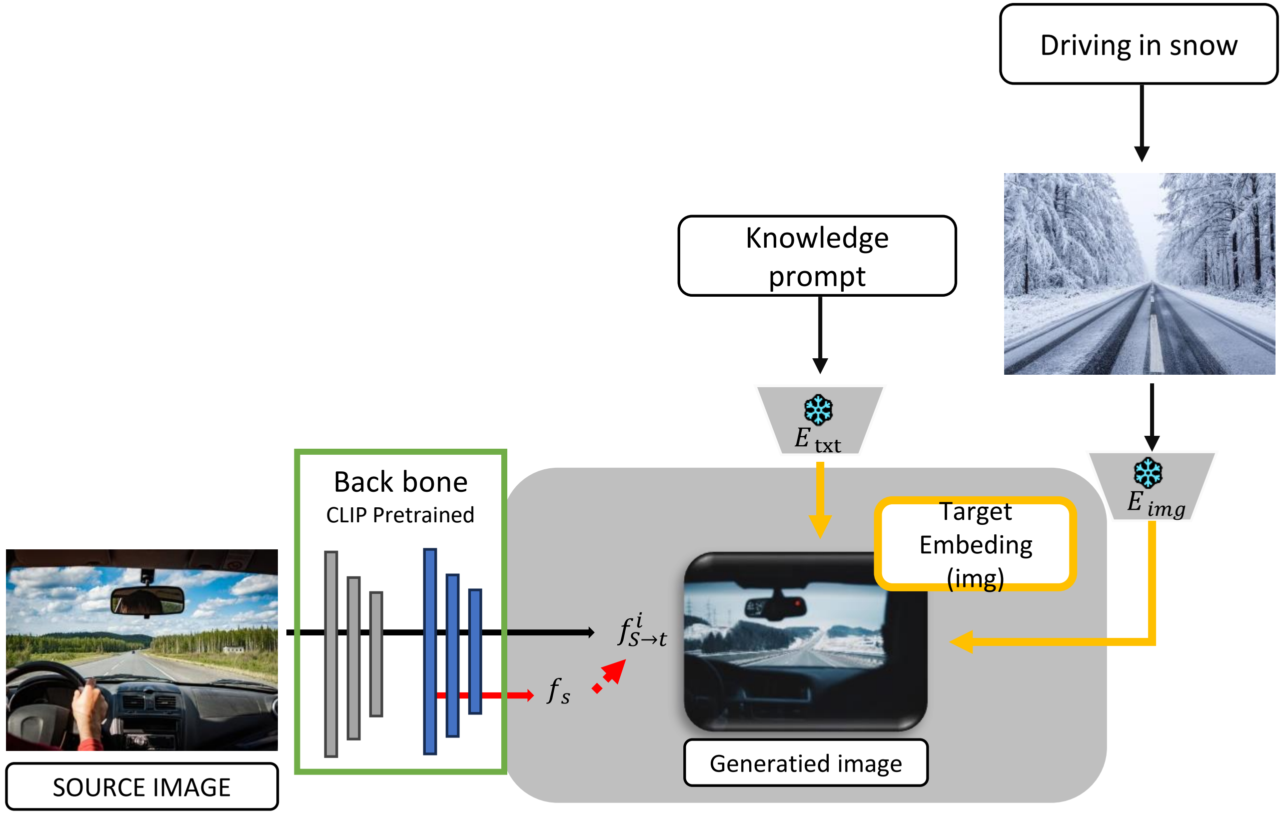
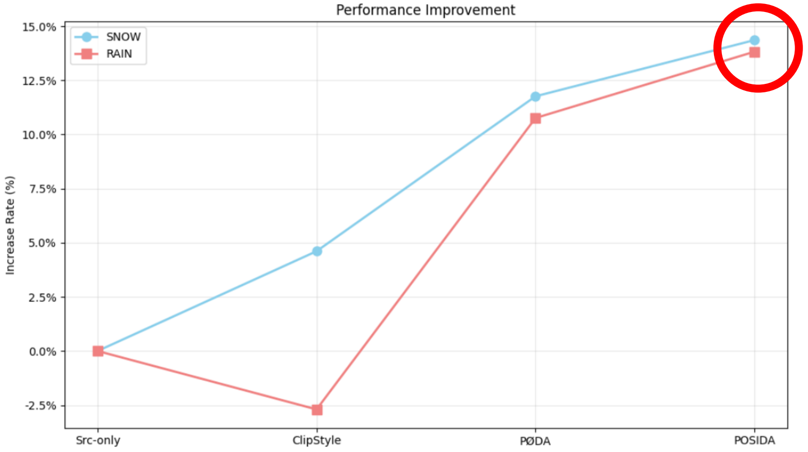


Table 1. Model Performance

Condition	SNOW	RAIN
Src-only	39.28	38.2
clipstyle	41.09	37.17
PODA	43.9	42.31
POSIDA	44.92	43.48



Gihyun Kwon and Jong Chul Ye. Clipstyler: Image style transfer with a single text condition. In CVPR, 2022
Zhou, K., Loy, C. C., & Liu, Z. (2022). *Prompt-driven Zero-shot Domain Adaptation*..

5 Ablation study

About LOSS function

Table 2. Loss

Condition	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.8$
SNOW	44.24	44.77	44.53
RAIN	42.51	43.48	42.85

$$\text{Combined Loss} = \alpha \cdot \text{Loss}_{\text{IMAGE}} + (1-\alpha) \cdot \text{Loss}_{\text{TEXT}}$$

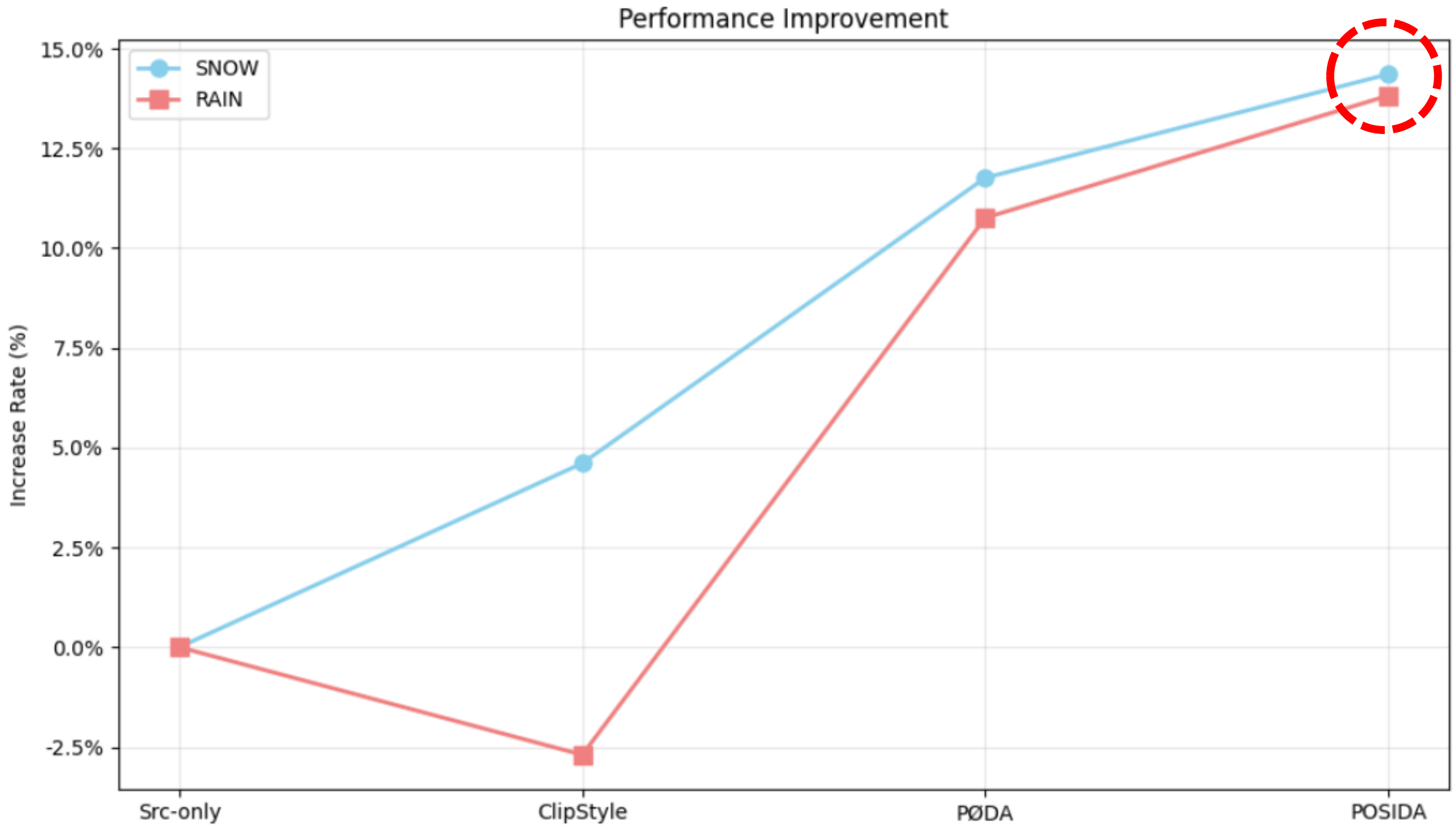
About Knowledge prompt

Table 3. Knowledge prompt

Condition	w/o knowledge prompt	with knowledge prompt
SNOW	44.23	44.36
RAIN	42.02	42.42

6 Conclusion

1. Overall performance increased



6 Conclusion

2. Although the model added generated images

- inference time & model size unchanged

  PODA : 1sec -> 5.98 img / POSIDA : 1sec -> 5.82 img

  model size & num of hyper parameter is remained

3. Limitation

- lack of time & resources limitation on various condition experiment

4. Discussion

- POSIDA method can extended to various domains
- In future research can aim to improve performance through optimizing prompts

Thank you

Q & A

7 Appendix

1. CLIP style

- Use clip embedding space
- Use text prompt to change image style