Principles of Forgetting in Domain-Incremental Semantic Segmentation in Adverse Weather Conditions

Tobias Kalb[1], Jürgen Beyerer[2,3]

[1] Porsche Engineering Group GmbH, [2] Fraunhofer IOSB, [3] Karlsruhe Institute of Technology

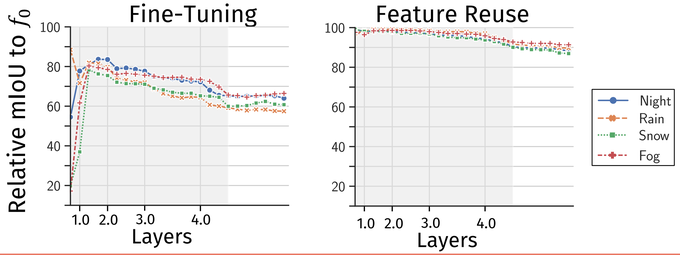
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Overview

Severe activation drift in early layers in domain-incremental learning due to:

1. Biased population statistics in BatchNorm Layers
2. Specialized domain-specific features in early layers

Increasing **feature reuse** by learning generalized features drastically decreases activation shift and thereby catastrophic forgetting.



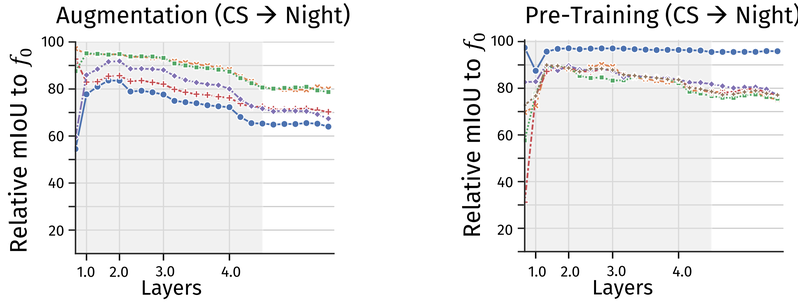
Activation Drift after Incremental Learning

Severe activation drift and forgetting specifically for seemingly similar tasks, especially Cityscapes → Rain, why?

Task 2	Cityscapes	Task 2		
		Cityscapes	Task 2	Forgetting
Rain	30.4	38.8	57.7	33.2
Night	10.5	45.9	43.6	26.1
Snow	23.1	42.2	62.3	29.8
Fog	33.4	44.0	69.0	28.0

Results in mIoU (%)

Low-Level Feature Reuse

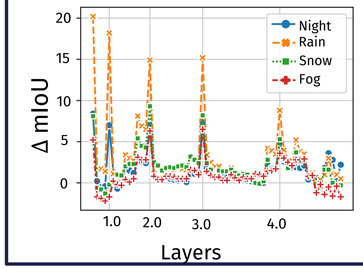


	Method	Cityscapes	Task 2		
		Test mIoU	avg. zero-shot	avg. Test	avg. Forgetting
Augm.	FT	72.0	24.4	58.2	30.1
Augm.	AutoAlb.	72.2	38.6	59.7	12.8
Augm.	Distort	71.7	34.5	59.5	17.4
Pre-Training	ImageNet	73.9	20.3	60.9	23.6
Pre-Training	DINO	75.0	25.6	63.4	22.2
Pre-Training	BarlowT.	73.9	30.6	62.5	19.0
Pre-Training	SwAV	76.4	28.9	62.2	25.7
	Offline	-	58.8	59.1	2.6

Results in mIoU (%)

Impact of Batch Normalization

Re-estimation of population mean and deviation



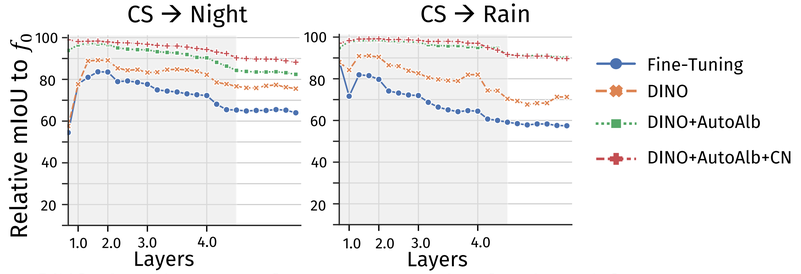
Method	Night	Rain
	Δ *mIoU*	Δ *mIoU*
Fine-Tuning	12.7	19.4
AutoAUG	2.7	0.5
Distort	3.7	6.5
ImageNet	7.4	8.1
DINO	7.7	12.2
BarlowT	8.6	13.9
Offline	-0.2	-0.8

- BN population statistics are biased → leads to severe forgetting
- Pre-Trained are more affected

Comparison to Continual Learning

Method	Night		Rain	
	lrn. acc.	*avg. forg.*	*lrn. acc.*	*avg. forg.*
EWC	52.8	18.0	61.4	21.2
+ DINO	58.1	10.0	66.0	11.3
+ AutoAlb + CN	58.5	4.8	69.1	4.9
Replay	58.2	5.6	64.2	2.9
+ DINO	62.4	4.2	70.8	2.1
+ AutoAlb + CN	61.8	4.1	69.9	3.8
Ours	62.1	7.9	71.0	8.0

Combining the Findings



- DINO alone increases feature reuse only after layer1.0
- Initially discrepancy

Take-Aways

- Feature-Reuse drastically reduces catastrophic forgetting
- Pre-Training stabilizes intermediate representations
- Low-Level Representations can be stabilized with augmentation

Initialization of your model greatly affects catastrophic forgetting!