FeTrIL: Feature Translation for Exemplar-Free Class-Incremental Learning

Grégoire Petit[1,2], Adrian Popescu[1], Hugo Schindler[1], David Picard[2], Bertrand Delezoide[3]

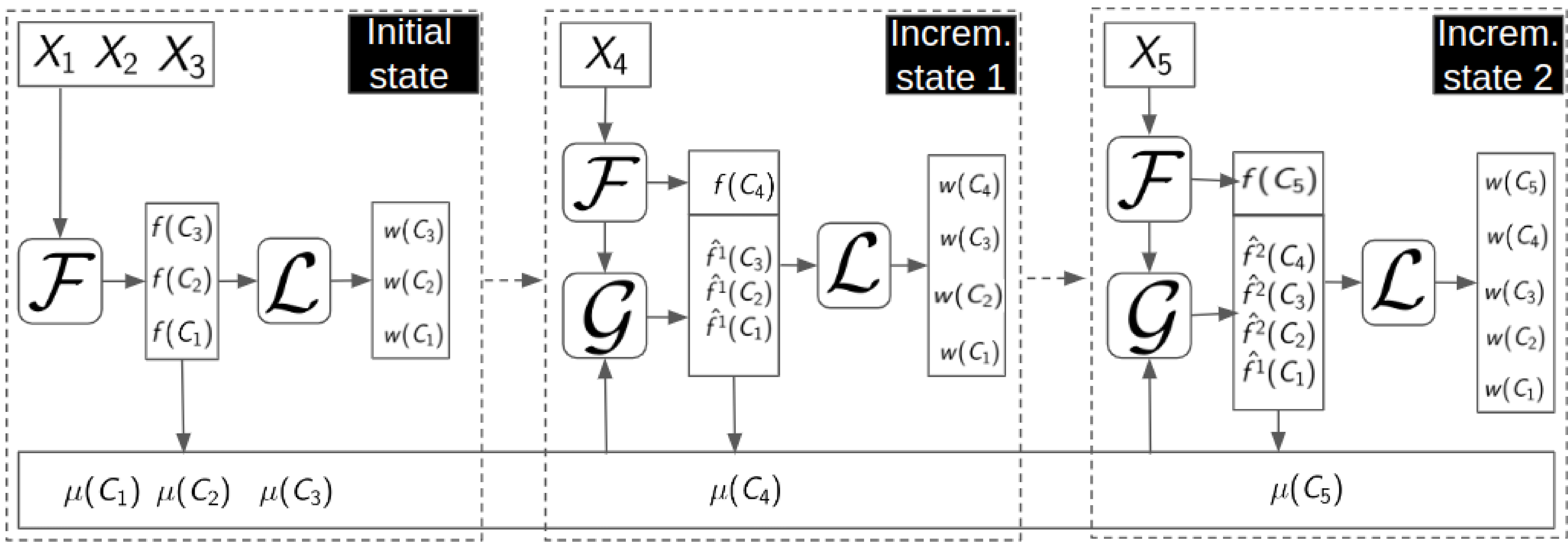
[1] list cea tech

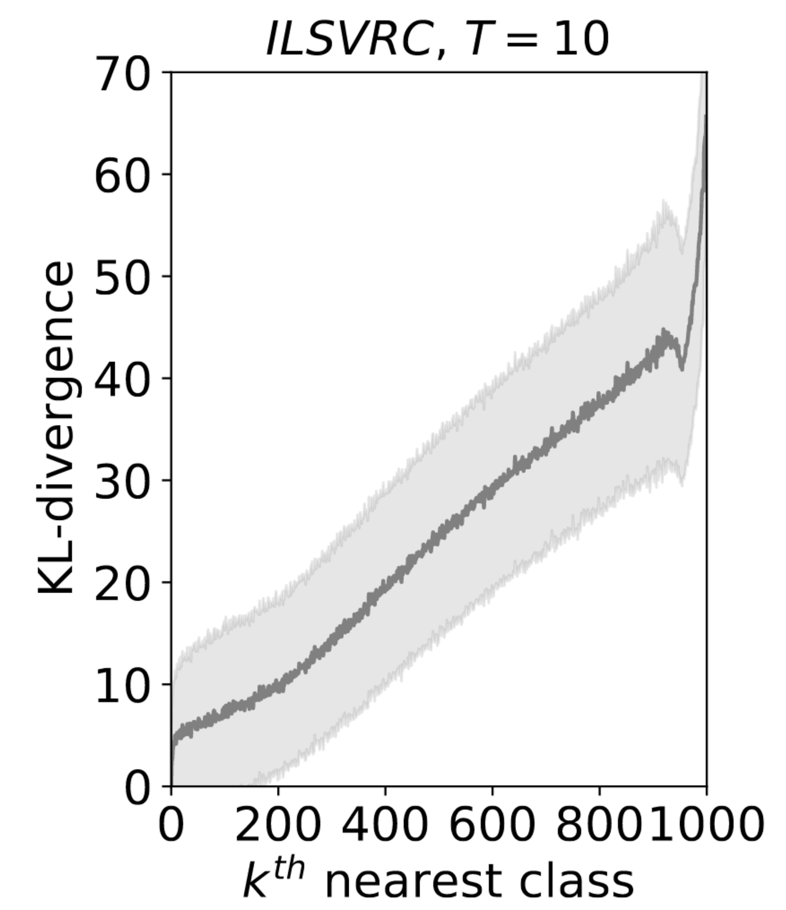
[2] École des Ponts ParisTech — Laboratoire d'Informatique Gaspard-Monge

[3] AMANDA

Motivation & directions

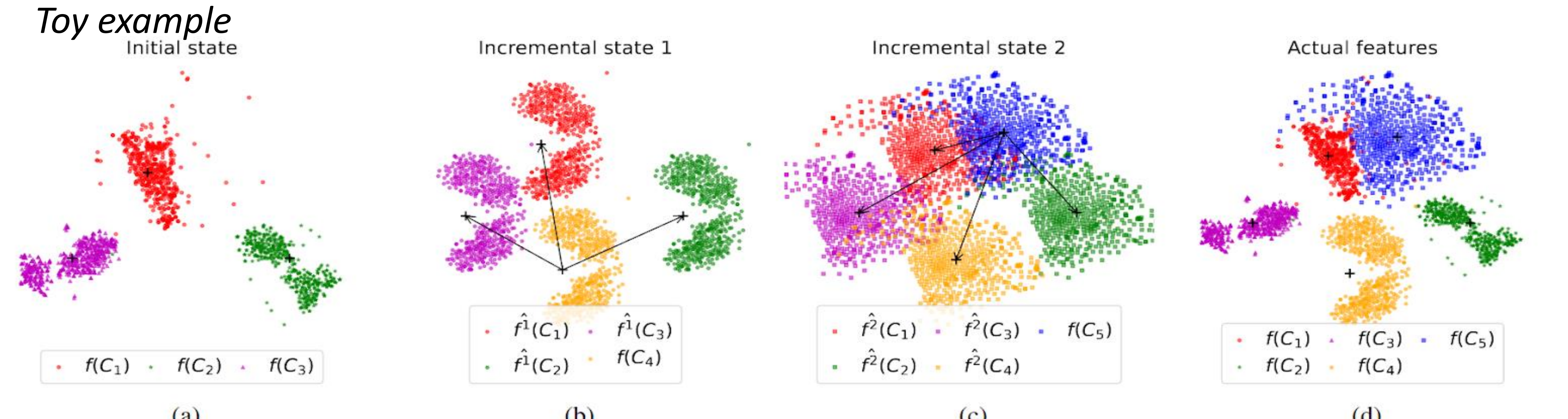
- Class-Incremental Learning: data arrives sequentially(e.g. [C_0,C_1, …, C_9], then [C_{10},C_{11}, …, C_{19}], etc.)
- Exemplar-Free: No possibility to store previously seen data (i.e. no rehearsal memory)
- Generation via geometric translation of pseudo features for past classes in each new state





FeTrIL overview. Average KL-divergence between distributions, depending on their neighboring rank.

Toy example



Pseudo-features generation procedure. Toy example with three states (one initial and two incremental) in (a), (b) and (c). (d) provides the actual features of all four classes

Results

	CIFAR-100				TinyImageNet				ImageNet-Subset				ImageNet		
	T=5	T=10	T=20	T=60	T=5	T=10	T=20	T=100	T=5	T=10	T=20	T=60	T=5	T=10	T=20
DeeSIL (ECCVW'18)	60.0	50.6	38.1	x	49.8	43.9	34.1	x	67.9	60.1	50.5	x	61.9	54.6	45.8
PASS (CVPR'21)	63.8	61.8	58.1	x	49.6	47.3	42.1	x	64.4	61.8	51.3	x	-	-	-
IL2A (NeurIPS'21)	66.0	60.3	57.9	x	47.3	44.7	40.0	x	-	-	-	x	-	-	-
SSRE (CVPR'21)	65.9	65.0	**61.7**	x	50.4	48.9	48.2	x	-	67.7	-	x	-	-	-
FeTrIL_{fc}	64.7	63.4	57.4	50.8	52.9	51.7	49.7	41.9	69.6	68.9	62.5	58.9	65.6	64.4	63.4
FeTrIL	**66.3**	**65.2**	61.5	**59.8**	**54.8**	**53.1**	**52.2**	**50.2**	**72.2**	**71.2**	**67.1**	**65.4**	**66.1**	**65.0**	**63.8**

Benchmark of FeTrIL against the state-of-the-art methods that have results. We notice, in bold, that although FeTrIL is very simple, it is really performing. It should also be noted that in the case of one-class incremental learning FeTrIL works and is also very efficient and not very sensitive to catastrophic forgetting.

Conclusion

FeTrIL advantages:

- Embeddable since it has low requirements in terms of computation and memory
- Much simpler and more effective than mainstream distillation-based methods
- Usable for one-class incremental steps
- Performance close to that of exemplar-based methods

FeTrIL limitations:

- Dependent on the domain shift between the initial fixed model and subsequent data
- Initial classes are favored over the rest since the fixed model is trained with them
- The pseudo-feature generator could be learned for a more refined representation of past classes

WACV WAIKOLOA HAWAII JAN 3-7 • 2023