Exploring Joint Embedding Architectures and Data Augmentations for Self-Supervised Representation Learning

Sami Barchid[1] José Mennesson[12] Chaabane Djéraba[1]

[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL

[2] IMT Nord Europe, Institut Mines-T elecom, Centre for Digital Systems

[12] F-59000 Lille, France

CVPR JUNE 18-22, 2023 VANCOUVER, CANADA

SCAN ME

Introduction

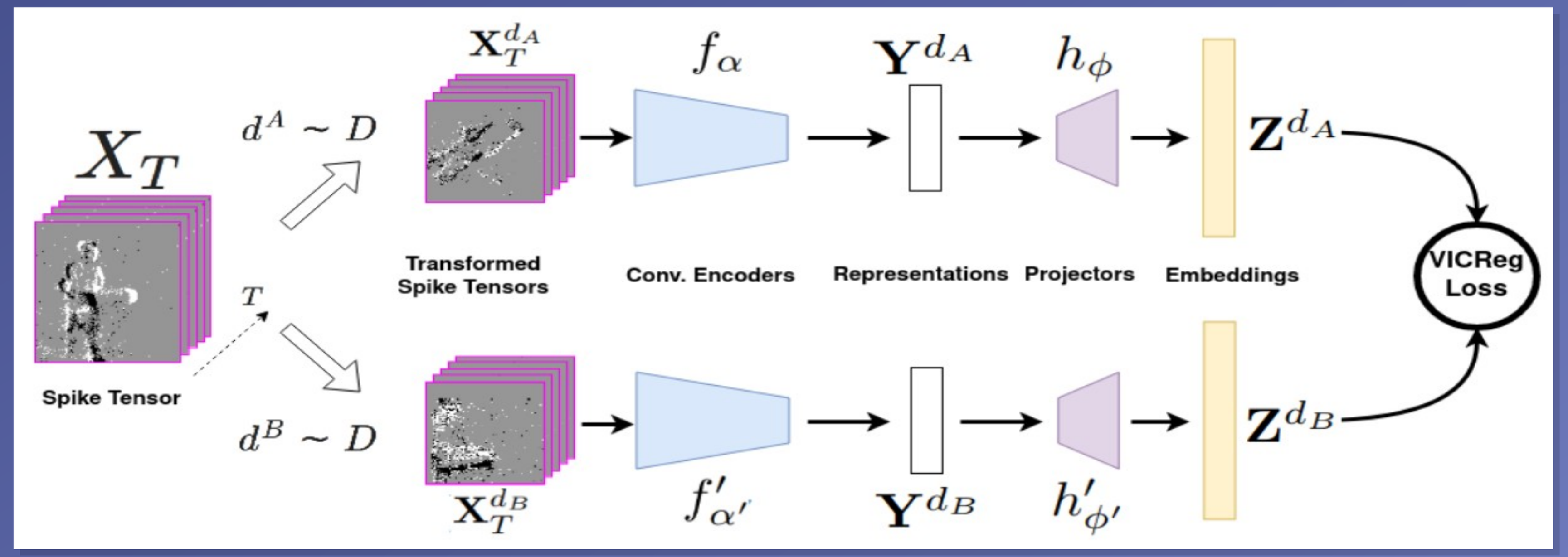
- Many event-based datasets have a limited number of labeled samples
- It presents challenges for the development of event vision algorithms
- Self-Supervised Representation Learning (**SSRL**) is a good solution for reducing the reliance on labeled data

Contributions

- Event-based SSRL framework
- Evaluation protocols
- Study on EDAs (Event Data Augmentation)
- Analysis of learned features

Event-Based SSRL Framework

Joint embedding architecture

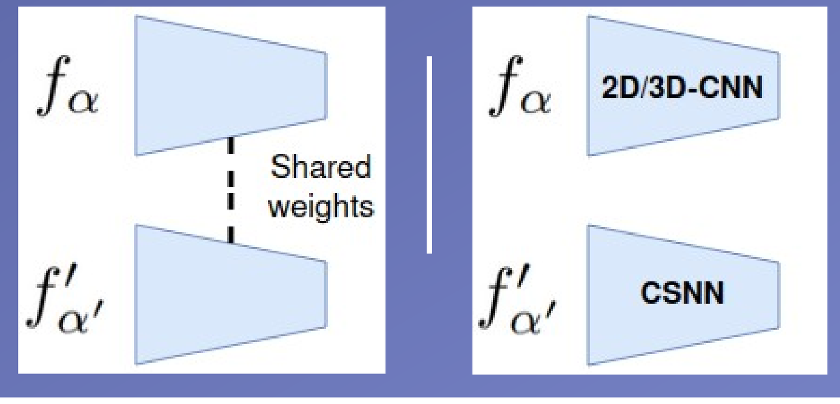


Three types of ConvEncs : **2D-CNN, 3D-CNN, CSNN**

Variants

Twins

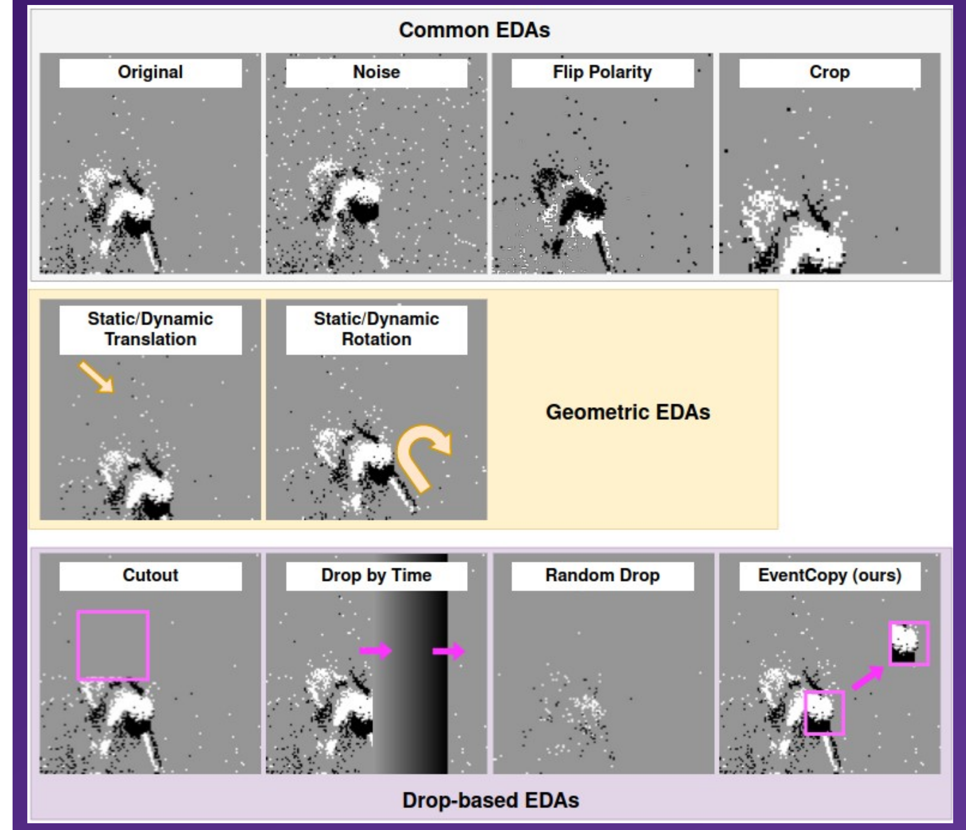
Student-Teacher



Study on Event Data Augmentations

How to design an effective EDA Distribution for pretraining?

Incremental experiments: we keep the best-performing EDA configuration from the previous stage



Stage 1: Common EDAs			Twins			Student-Teacher			
Noise	Crop	PolFlip	CSNN	2D	3D	CSNN	2D	CSNN	3D
			12.12	11.36	57.95	59.47	61.36	58.33	55.68
✓			**60.23**	74.24	69.70	62.12	64.02	68.26	60.23
✓	✓		56.44	81.82	74.62	**73.86**	**76.89**	69.32	64.02
✓	✓	✓	56.44	**83.33**	**76.52**	73.11	74.62	**70.63**	**67.80**

Stage 2: Geometric EDAs					Twins			Student-Teacher			
StatTran	StatRot	DynTran (ours)	DynRot (ours)	StatDynGeo (ours)	CSNN	2D	3D	CSNN	2D	CSNN	3D
✓	✓				49.24	77.65	65.15	60.98	68.18	65.53	62.5
		✓	✓		68.56	75.38	76.14	**74.24**	**73.11**	**72.73**	67.05
				✓	**68.94**	**79.55**	**77.27**	67.05	70.45	71.97	**68.18**

Stage 3: Drop-based EDAs				Twins			Student-Teacher			
Cutout	EventDrop	Event-Copy (ours)	EventCopyDrop (ours)	CSNN	2D	3D	CSNN	2D	CSNN	3D
✓				**71.59**	85.11	88.64	71.59	72.35	68.56	65.91
	✓			68.18	**87.12**	75	75	76.52	**76.52**	**76.14**
		✓		65.16	81.82	83.33	76.52	**77.52**	73.86	68.94
			✓	70.83	**87.12**	**89.39**	**76.89**	75.76	75	73.48

Accuracy on DVSGesture (Linear Eval. Protocol)

A good EDA distribution consists of :

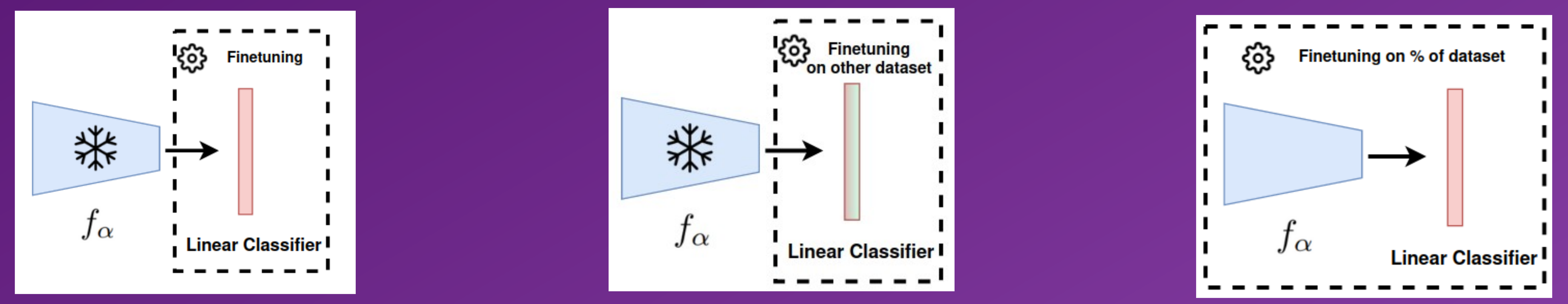
- More common EDAs
- Strictly one Geometric and one Drop-based EDA
- OneOf functions (e.g., *EventCopyDrop, StatDynGeo,...)*

Evaluation Protocols

Linear Evaluation

Transfer Learning

Semi-Supervised



Dataset	Protocol	CSNN	2D	3D	$CSNN_{2D}$	$CSNN_{3D}$
DVSGesture	Linear	70.83	87.12	**89.39**	76.89	76.52
	SemiSup-10%	60.98	75.52	**81.44**	66.67	69.31
	SemiSup-25%	75.00	87.12	**90.15**	76.14	80.30
N-Caltech101	Linear	64.29	64.39	**69.46**	62.34	65.67
	SemiSup-10%	56.72	**64.64**	62.80	53.96	53.50
	SemiSup-25%	66.02	**72.79**	71.64	62.22	59.93
ASL-DVS	Linear	95.32	**99.38**	98.68	97.87	97.30
	SemiSup-05%	95.66	**97.06**	96.62	93.54	95.66
	SemiSup-10%	99.51	99.64	**99.70**	99.48	99.48

Metric: classification accuracy (%)

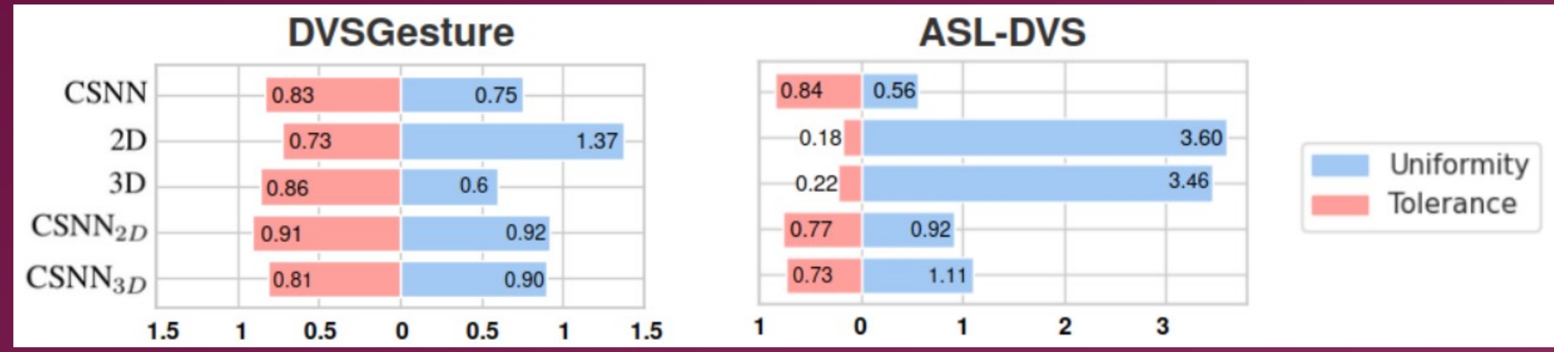
Transfer learning protocol

Datasets						
Pretrain	Linear	CSNN	2D	3D	$CSNN_{2D}$	$CSNN_{3D}$
DVSGesture	**DailyAction-DVS**	77.93	88.28	84.83	**91.03**	87.59
ASL-DVS	**N-CARS**	92.81	94.61	**95.64**	93.30	93.35

Analysis of Learned Features

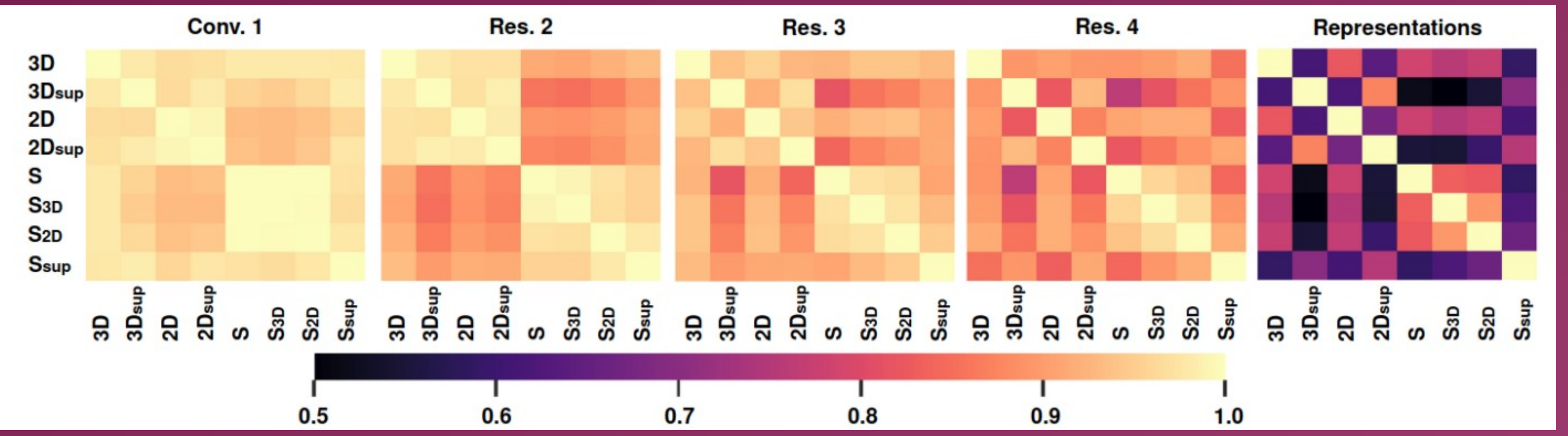
1 Uniformity - Tolerance Trade-Off

Assumption from frame-based vision: balanced values of Uniformity and Tolerance result in optimal representation quality



- The original assumption does not prevail
- Student-Teacher variant increases the tolerance of the CSNNs

2 Linear CKA for Similarity Assessment



- The divergence increases with deeper layers
- The impact of Student-Teacher variants on the learned features of CSNNs

Conclusion

- We propose a method for **event-based SSRL** that utilizes a joint embedding architecture and event data augmentations
- The **evaluation protocols** established in this study emphasize the efficiency and transferability of the learned features, as well as the reduced dependence on labeled data facilitated by our framework
- We thoroughly **investigate the impacts of popular EDAs** and introduce **additional methods** to achieve an optimal distribution for event-based SSRL
- Our method creates exciting possibilities for designing future event-based vision applications that do not require large-scale training sets