

CRET: Cross-Modal Retrieval Transformer for Efficient Text-Video Retrieval

Authors: Kaixiang Ji, Jiajia Liu, Weixiang Kong, Liheng Zhong, Jian Wang, Jingdong Chen, Wei Chu

Siwen Tu and Shupef Li

Leiden Institute of Advanced Computer Science
April 11, 2023



**Universiteit
Leiden**
The Netherlands

① Motivation

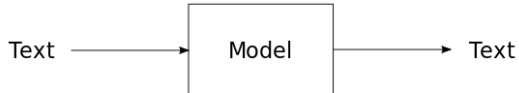
② Methodology

③ Experiments

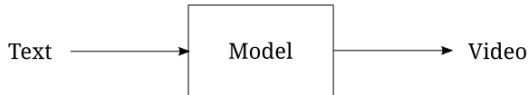
④ Critical Review

Motivation

- Unimodal information retrieval task.



- Multimodal information retrieval task, e.g. text-to-video retrieval.



Experiments

- **Datasets:** MSRVT, LSMDC, MSVD, DiDeMo.
- **Metrics:** R@K, MdR.
- **Results:**
 - MSRVT

	Weight Initialization		E2E	Type	R@1↑	R@5↑	R@10↑	MdR↓
	Visual	Textual						
MIL-NCE [36]	K+H	G+H	✓	EDB	9.9	24.0	32.4	29.5
JSFusion [54]	I	N.A.	✓	MDB	10.2	31.2	43.2	13.0
HT [38]	I	G	✓	EDB	12.4	36.0	52.0	10.0
HT [38]	I+H	G+H	✓	EDB	14.9	40.2	52.8	9.0
ActBERT [59]	I+H	B+H		MDB	16.3	42.8	56.9	10.0
HiT(appearance-only) [30]	I+H	B+H	✓	EDB	18.2	41.9	55.5	5.0
TACo(R-152) [53]	I+H	B+H	✓	MDB	18.9	46.2	58.8	7.0
UniVL(FT-Joint) [33]	K+H	B+H		EDB	20.6	49.1	62.9	6.0
UniVL(FT-Align) [33]	K+H	B+H		MDB	21.2	49.6	63.1	6.0
ClipBERT [28]	I+C+V	C+V	✓	MDB	22.0	46.8	59.9	6.0
Ours	I	B	✓	EDB	23.9	50.8	63.4	5.0

Experiments

- Results:
 - LSMDC

	E2E	Type	R@1↑	R@5↑	R@10↑	MdR↓
CT-SAN [55]	✓	MDB	5.1	16.3	25.2	46.0
HT [38]	✓	EDB	5.8	18.8	28.4	45.0
HT* [38]	✓	EDB	7.1	19.6	27.9	40.0
NoiseE* [1]		EDB	6.4	19.8	28.4	39.0
JSFusion [54]	✓	MDB	9.1	21.2	34.1	36.0
Ours	✓	EDB	10.0	24.9	33.4	34.0

- MSVD

	E2E	Type	R@1↑	R@5↑	R@10↑	MdR↓
HT [38]	✓	EDB	13.0	37.4	52.4	10.0
HT* [38]	✓	EDB	15.5	40.9	55.7	8.0
NoiseE* [1]		EDB	20.3	49.0	63.3	6.0
CLIP4Clip [†] [34]		EDB	46.2	76.1	84.6	2.0
Ours	✓	EDB	49.0	87.0	95.0	2.0

Experiments

- **Results:**
 - DiDeMo

	E2E	Type	R@1↑	R@5↑	R@10↑	MdR↓
S2VT [49]	✓	EDB	11.9	33.6	-	13.0
FSE [57]	✓	EDB	13.9	36.0	-	11.0
ClipBERT [‡] [28]	✓	MDB	20.4	48.0	60.8	6.0
Ours	✓	EDB	21.2	50.3	63.5	6.0

- **Ablation studies.**
- **Validation of Gaussian assumption.**

Critical Review

- Readability and structure.
 - Illustrate CRET method clearly.
 - Satisfy requirements of the scientific paper.
- Reproducibility: Source code, the availability of data sets, experimental settings.
- Importance.
 - Theoretical contributions.
 - Practical applications.
- Summary of strong and weak points.