Berkeley
UNIVERSITY OF CALIFORNIA

King Abdullah University of
Science and Technology

Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization




Chen Zhao
KAUST

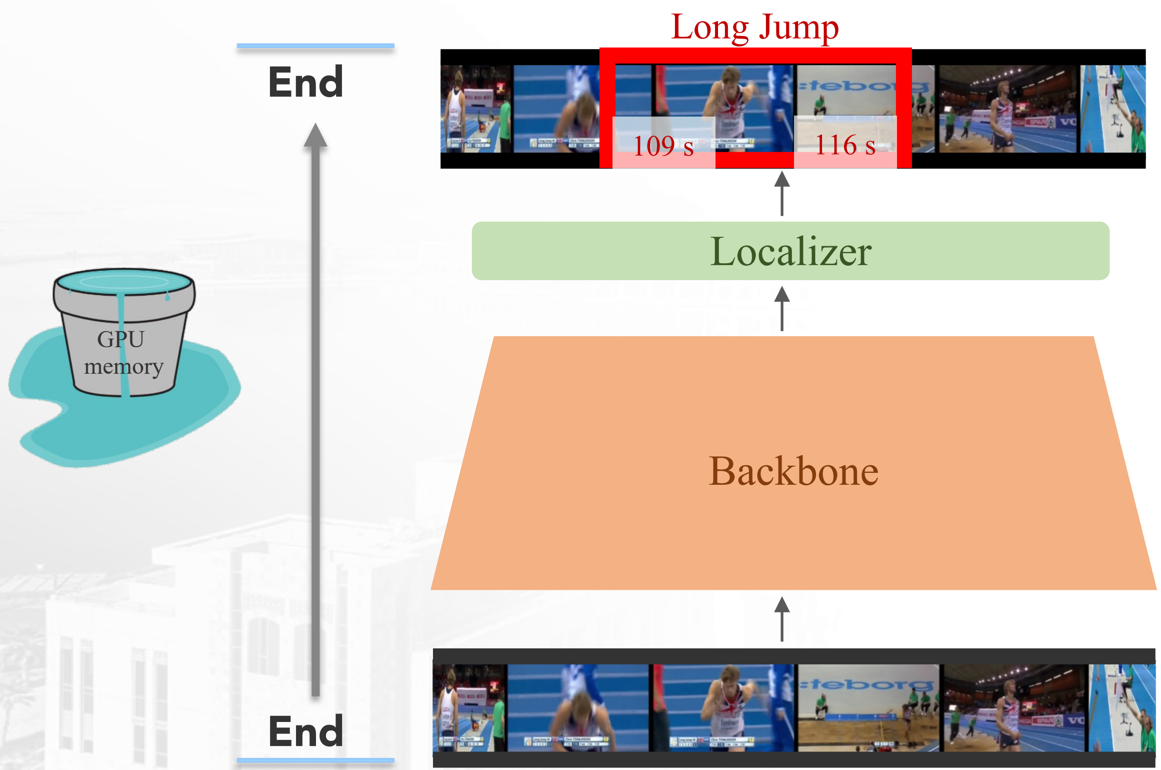
Shuming Liu
KAUST

Karttikeya Mangalam
UC Berkeley

Bernard Ghanem
KAUST

Poster Session WED-AM-230, CVPR 2023

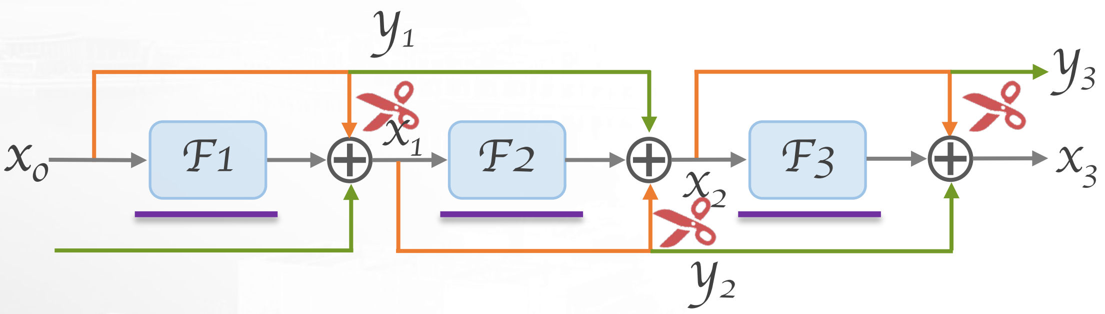
Background



Propose: Re^{2}TAL

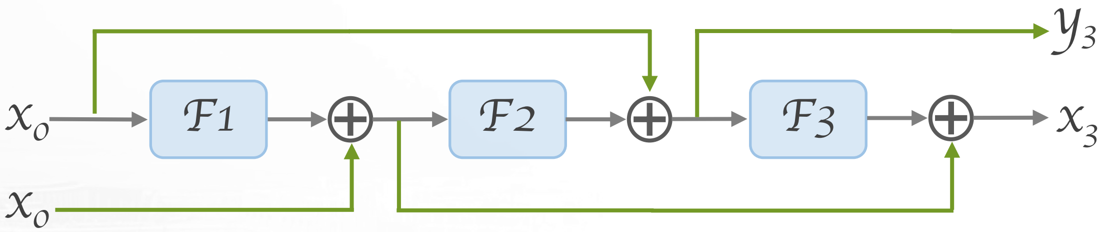
Residual modules

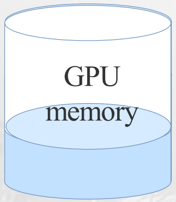
Rewiring

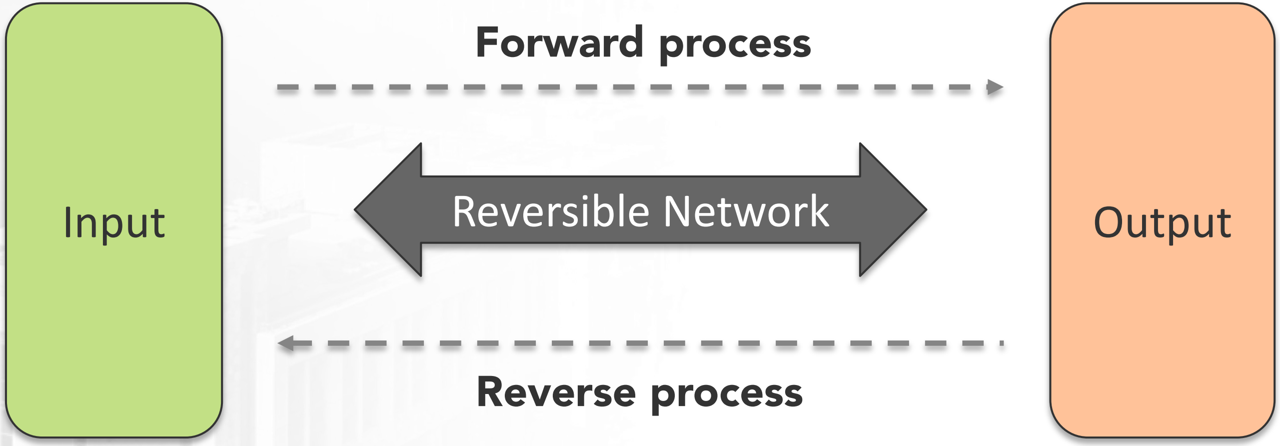


Propose: Re²TAL

Resversible modules



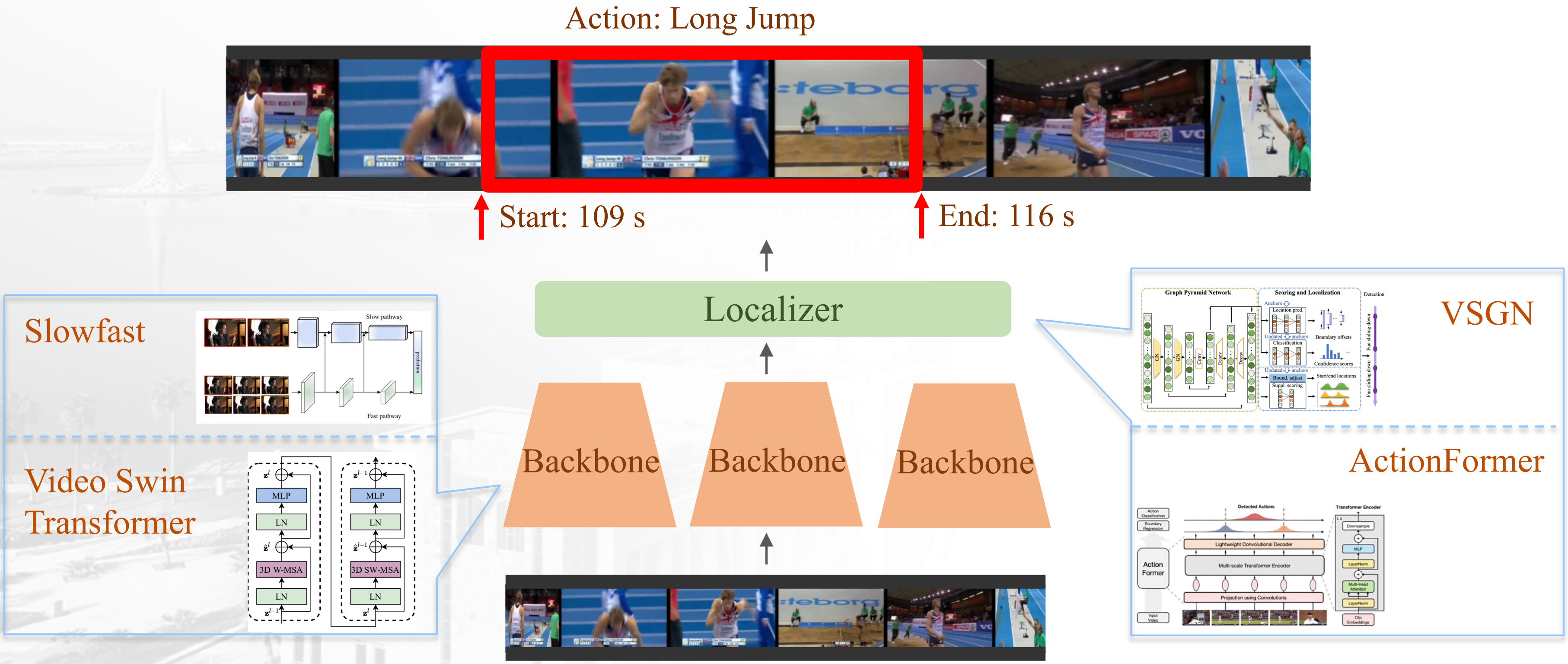




Background and motivation

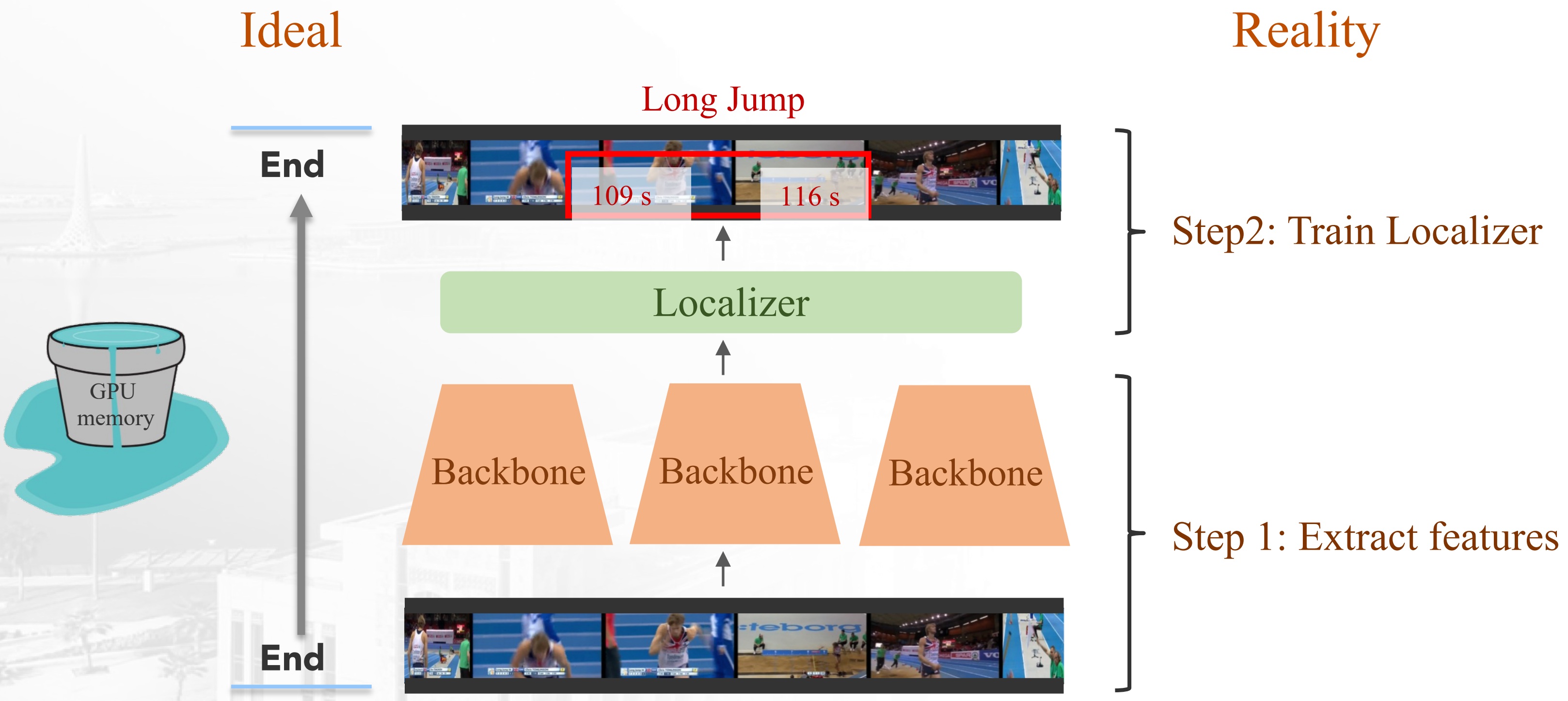
Temporal Action Localization Faces Memory Shortage

- Temporal action localization (TAL)



Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re²TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

Temporal Action Localization Faces Memory Shortage



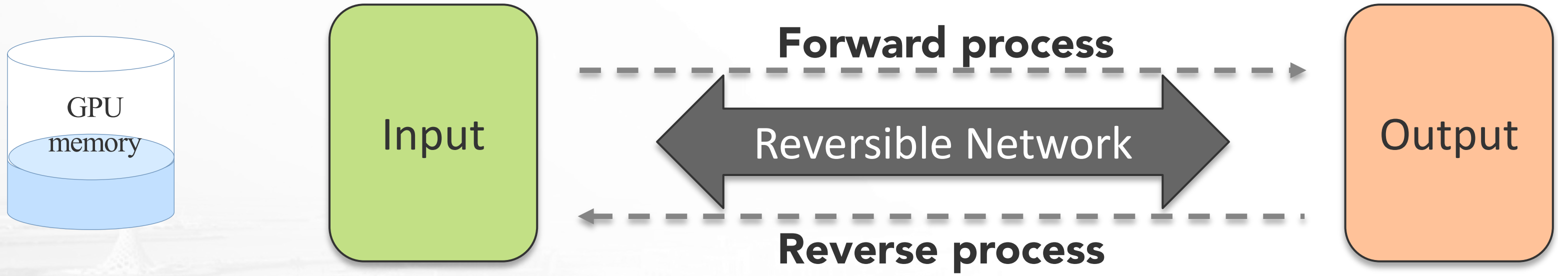
Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re²TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

Temporal Action Localization Faces Memory Shortage

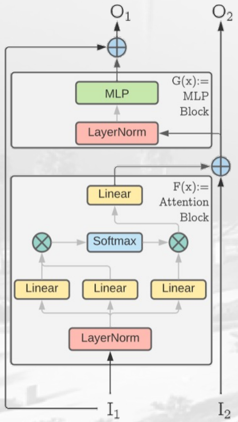
Can we use less memory for end-to-end training?

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

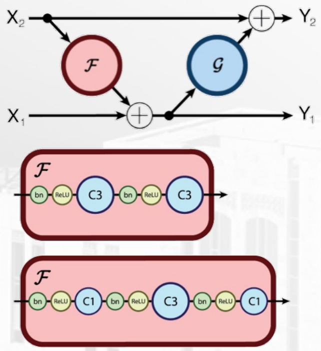
Reversible Networks Use Less Memory



Rev-ViT[1]

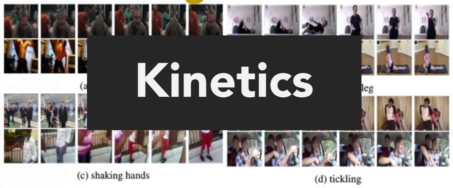


RevNet[2]



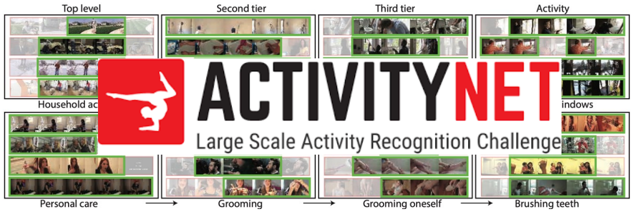
TAL training

Train on large video datasets



Train on large image datasets

Train on TAL datasets



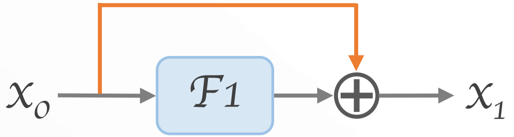
[1] Reversible vision transformers, CVPR'22

[2] The reversible residual network: Backpropagation without storing activation, NeurIPS'17

Proposed Re^{2}TAL

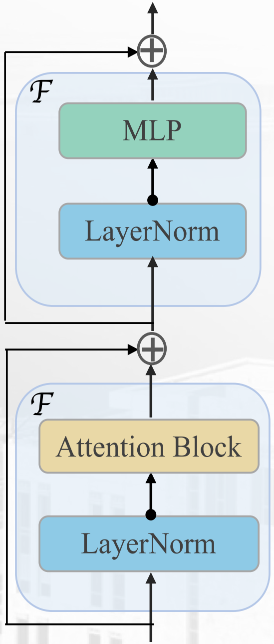
Proposed Re^{2}TAL

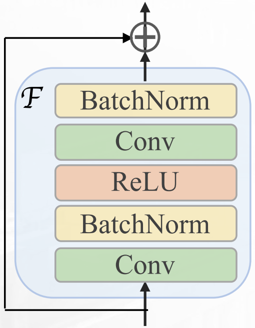
Residual module



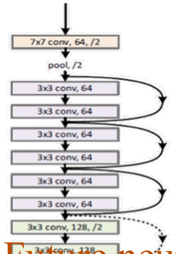
Transformer MLP block

Transformer attention block





Resnet block



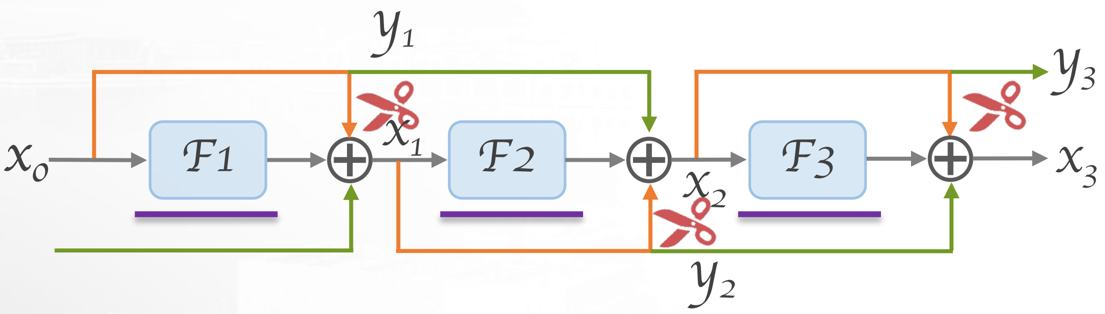
Future neural networks

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

Proposed Re^{2}TAL

Network of residual modules

Rewiring

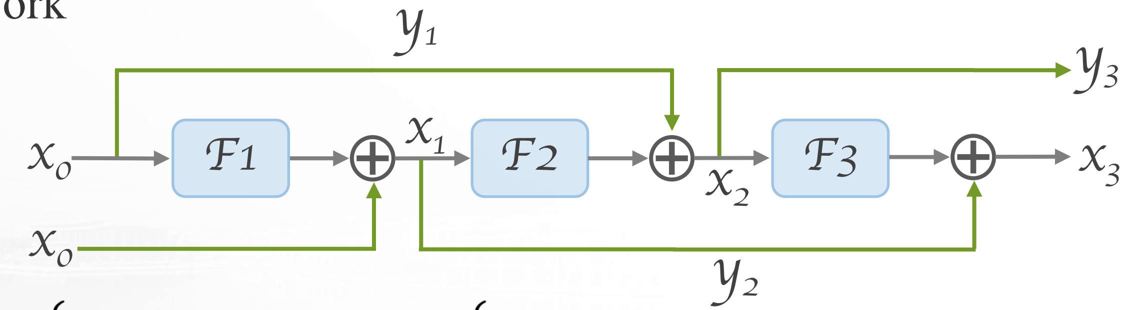


Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

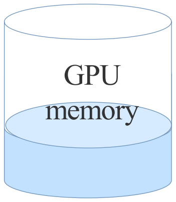
Proposed Re²TAL

Resversible network

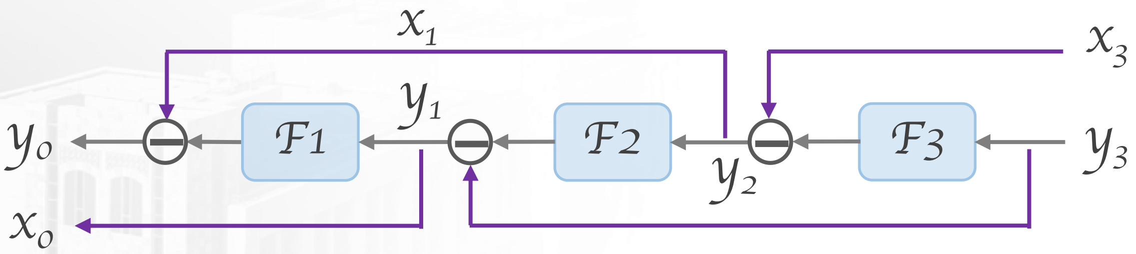
Forward



$$\begin{cases} y_1 = x_0 \\ x_1 = \mathcal{F}_1(x_0) + y_0, \end{cases} \Rightarrow \begin{cases} y_2 = x_1 \\ x_2 = \mathcal{F}_2(x_1) + y_1. \end{cases}$$



Reverse



$$\begin{cases} x_0 = y_1 \\ y_0 = x_1 - \mathcal{F}_1(x_0), \end{cases} \Leftarrow \begin{cases} x_1 = y_2 \\ y_1 = x_2 - \mathcal{F}_2(x_1). \end{cases}$$

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "Re²TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization", CVPR'23.

Re^{2}TAL End-to-end Training

- Prepare a video model with pretrained parameters
- Rewire the residual connection in the video model as proposed
- Load the pretrained parameters into the rewired model
- Finetune for several epochs on the pretraining task
- Ready for use for TAL (and other tasks)

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

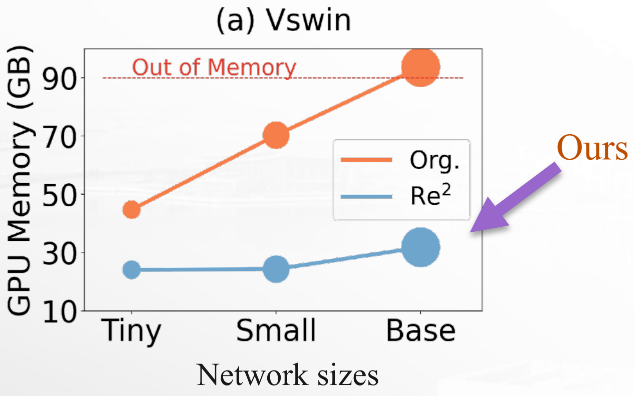
Berkeley
UNIVERSITY OF CALIFORNIA

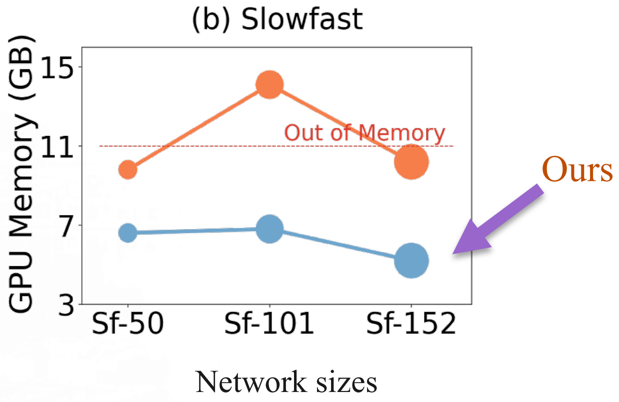
King Abdullah University of Science and Technology

Experimental Results

Experiments: Compared to Non-reversible TAL

- Memory consumption is reduced





- Accuracy is preserved

Average mAP (%)

Model	Tiny	Small	Base	Model	50	101	152
Vswin	34.36	33.86	**34.47**	Slowfast	34.60	35.04	**34.61**
Re^2Vw	**34.47**	**34.04**	34.09	Re^2Slowf	**34.93**	**35.24**	34.54

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^2TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

Experiments: Compared to Other Methods on TAL

- New SOTA
- Improves over feature-based training

mAPs (%) at tIoU thresholds

Method	Backbone	E2E	Flow	Mem	ActivityNet-v1.3 0.5	0.75	0.95	Avg.	THUMOS-14 0.3	0.4	0.5	0.6	0.7
TAL-Net [8]	I3D	✗	✓	-	38.23	18.30	1.30	20.22	53.2	48.5	42.8	33.8	20.8
BMN [36]	TSN	✗	✓	-	50.07	34.78	8.29	33.85	56.0	47.4	38.8	29.7	20.5
G-TAD [69]	TSN	✗	✓	-	50.36	34.60	9.02	34.09	54.5	47.6	40.2	30.8	23.4
TSI [41]	TSN	✗	✓	-	51.18	35.02	6.59	34.15	61.0	52.1	42.6	33.2	22.4
BC-GNN [4]	TSN	✗	✓	-	50.56	34.75	9.37	34.26	57.1	49.1	40.4	31.2	23.1
VSGN [73]	TSN	✗	✓	-	52.38	36.01	8.37	35.07	66.7	60.4	52.4	41.0	30.4
ActionFormer [71]	I3D	✗	✓	-	53.50	36.20	8.20	35.60	82.1	77.8	71.0	59.4	43.9
PBRNet [39]	I3D	✓	✓	-	53.96	34.97	8.98	35.01	58.5	54.6	51.3	41.8	29.5
AFSD [35]	I3D	✓	✓	12	52.40	35.30	6.50	34.40	67.3	62.4	55.5	43.7	31.1
R-C3D [66]	C3D	✓	✗	-	26.80	-	-	-	44.8	35.6	28.9	-	-
DaoTAD [63]	I3D	✓	✗	11	-	-	-	-	62.8	-	53.8	-	30.1
TALLFormer [10]	VSwin-Base	✓	✗	29	54.10	36.20	7.90	35.60	76.0	-	63.2	-	34.5
ActionFormer [71]	VSwin-Tiny	✗	✗	-	53.83	35.82	7.27	35.17	70.8	64.7	55.7	42.2	27.0
ActionFormer + Re^{2}TAL	Re^{2}VSwin-Tiny	✓	✗	24	54.75	37.81	9.03	36.80	77.0	71.5	62.4	49.7	36.3
ActionFormer [71]	Slowfast-101	✗	✗	-	53.98	37.00	8.87	36.09	72.7	66.9	58.6	46.4	33.1
ActionFormer + Re^{2}TAL	Re^{2}Slowfast-101	✓	✗	6.8	**55.25**	**37.86**	**9.05**	**37.01**	**77.4**	**72.6**	**64.9**	**53.7**	**39.0**

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

Berkeley
UNIVERSITY OF CALIFORNIA

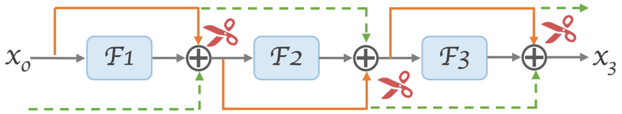
King Abdullah University of Science and Technology

Conclusions

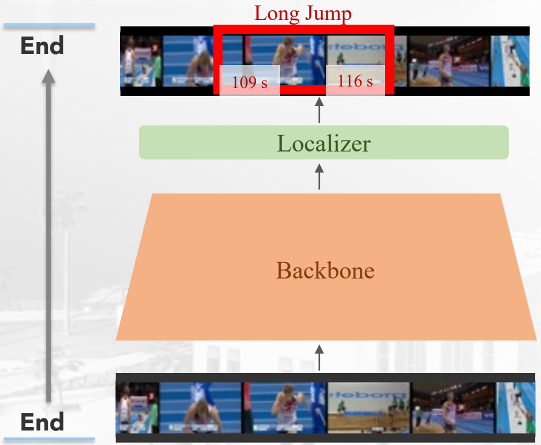
Conclusions

- Rewiring to be reversible (Re²)

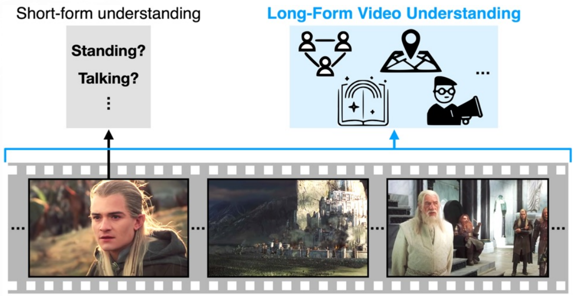
Rewiring



- End-to-end training



- Apply to memory-intensive tasks



Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re²TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.

Berkeley
UNIVERSITY OF CALIFORNIA

King Abdullah University of
Science and Technology

Email: chen.zhao@kaust.edu.sa

Website

Paper

Code

Chen Zhao, Shuming Liu, Karttikeya Mangalam, Bernard Ghanem, "**Re^{2}TAL: Rewiring Pretrained Video Backbones for Reversible Temporal Action Localization**", CVPR'23.