Flow-Assisted Motion Learning Network (Flaming-Net) for Weakly-Supervised Group Activity Recognition

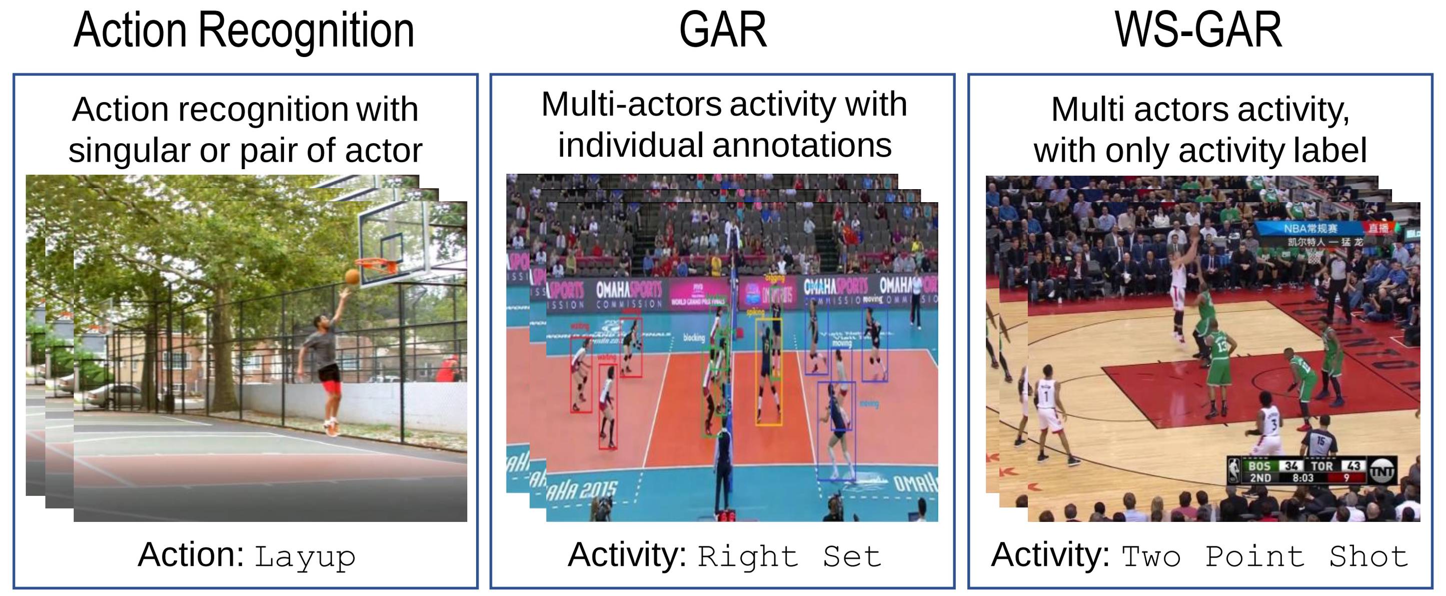
M.A. Nugroho, S. Woo, S. Lee, J. Park, Y. Wang, D. Kim, and C. Kim

Korea Advanced Institute of Science and Technology

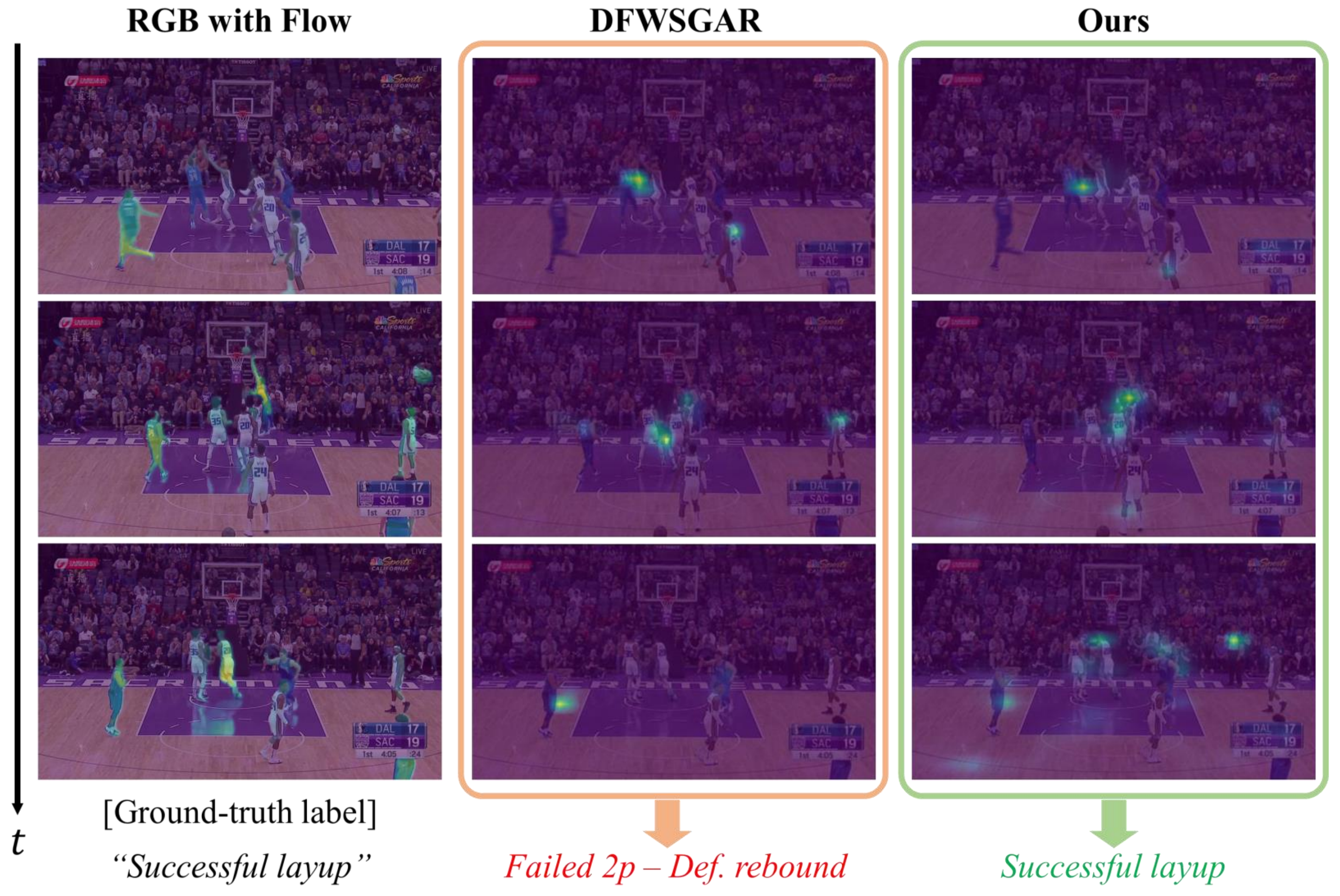
EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

Introduction

- WSGAR (Weakly Supervised Group Activity Recognition) requires the model to find actors and their complex spatio-temporal relations with no individual actor annotations provided.

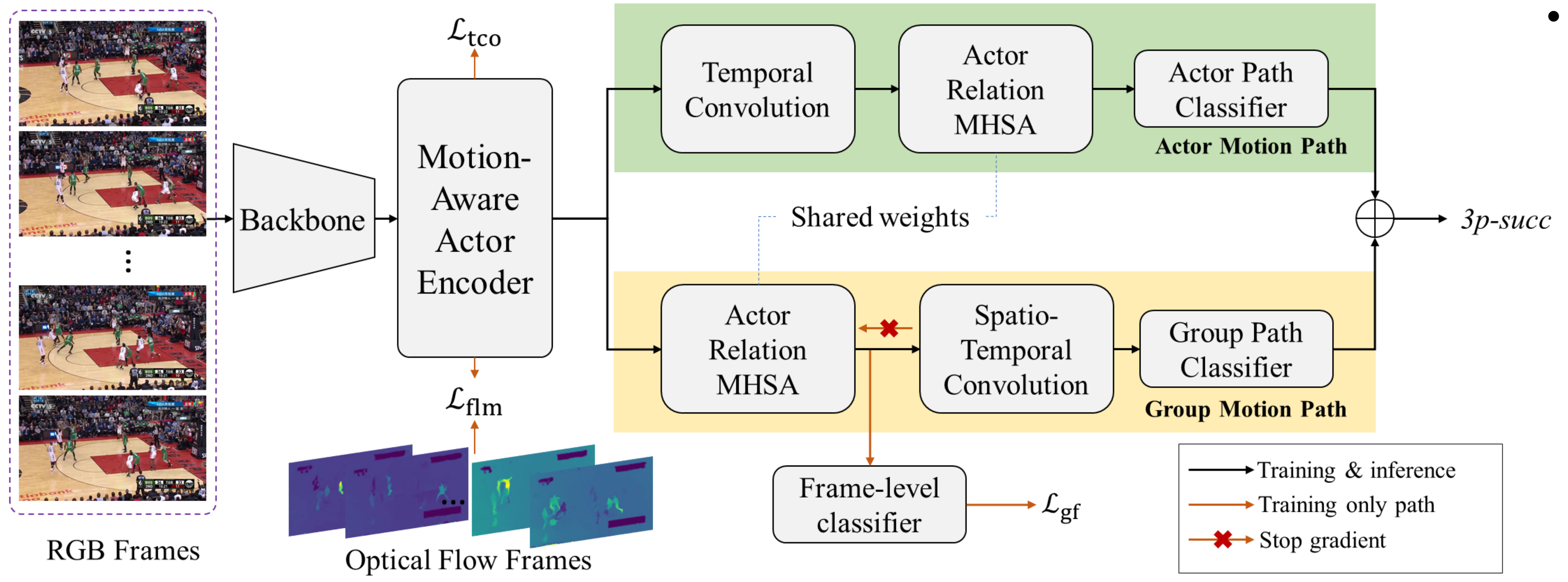


- Flaming-Net selectively highlights key actors, those who play important roles in the activity.
- Utilize optical flow as learning guidance to be more aware of actively moving actors.

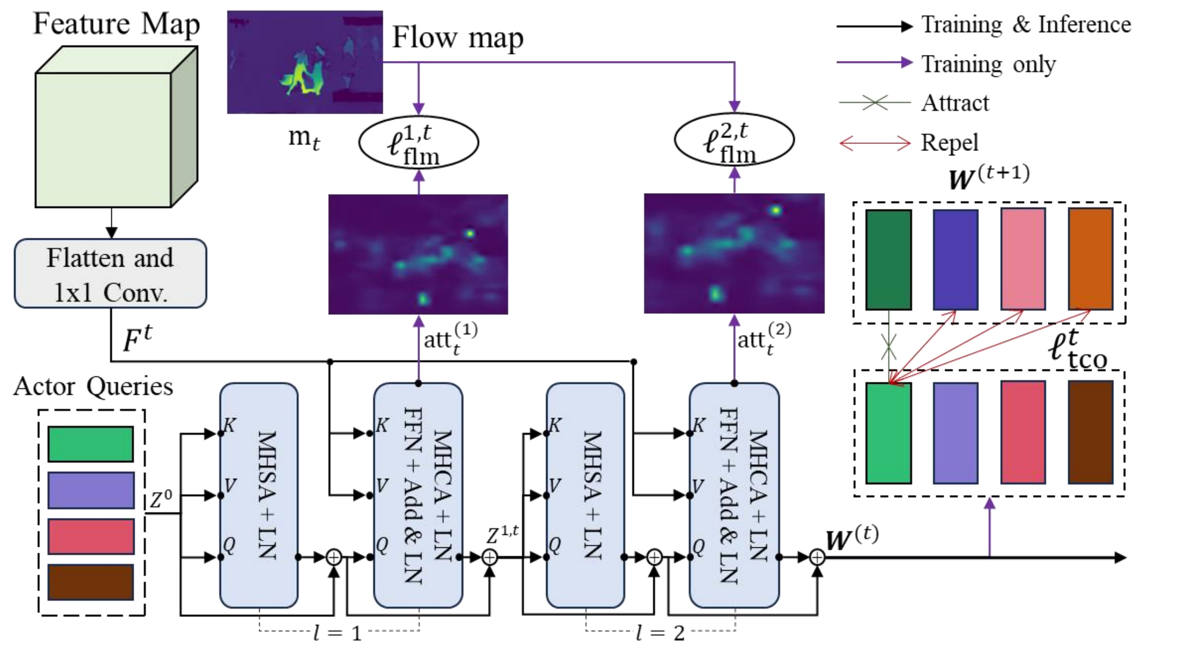


Method

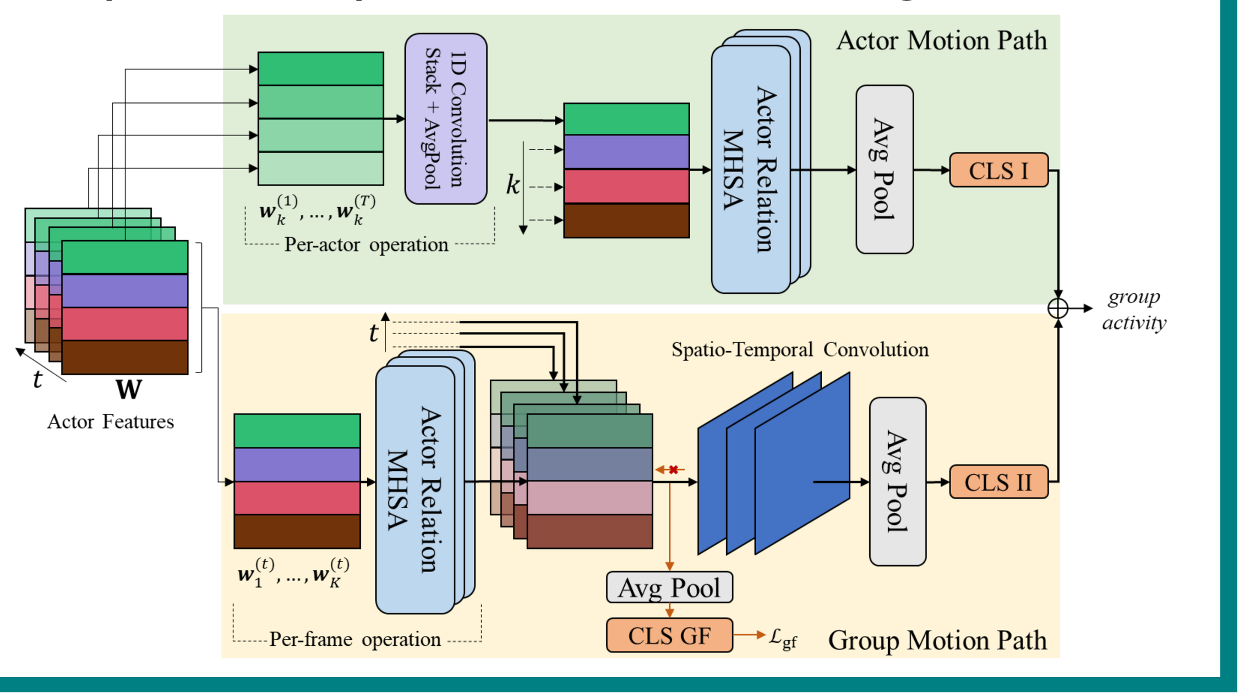
- Motion-aware actor encoder generates actor features by transforming a set of actor queries into features representing key actors.
- The actor relation module learns relations between actors to infer activity.
- Maintain efficiency by utilizing only the RGB frames in the inference stage.



- The encoder consists of multi-head attention modules with learnable actor queries.
- The attentions are guided by the optical flow using aux. loss $\mathcal{L}_{flm}$ to highlight active actors.
- Temporal contrastive loss helps to create consistent actor representation across time.

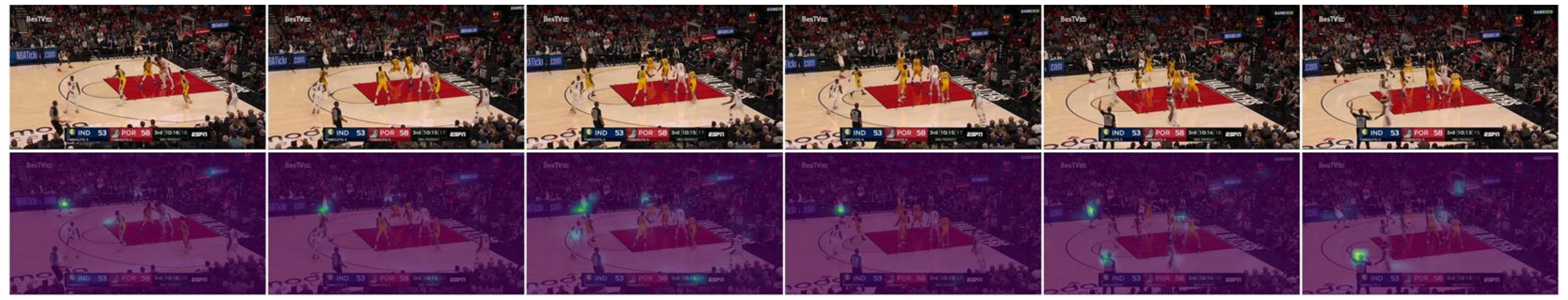


- The actor relation module consists of actor motion and group motion path.
- The actor motion path focuses on the temporal evolution of individual actors.
- The group motion path focuses on local spatio-temporal relations among actors.



Results

High attention to important players on crucial moments



In **3p-success**, Flaming-Net first highlights the shooter, then the def. players and referee

In **failed 2p - def. reb.**, Flaming-Net highlights the scrimmage and the rebound action

Performance in NBA dataset containing complex spatial and long temporal actor relationships.

Method	Pub.	F	#Params	FLOPs	MCA	MPCA
SAM	ECCV'20	✗	25.5M	304 G	44.2%	59.3%
Dual-AI	CVPR'22	✓	-	-	58.1%	50.2%
DFWSGAR	CVPR'22	✗	17.5M	313G	75.8%	71.2%
LRMM+GCM	ImaVis'23	✗	14.2M	306G	77.8%	73.2%
Flaming-Net	ECCV'24	Tr	13.8M	307G	**79.1%**	**76.0%**

Performance in WS-Volleyball dataset.

Method	Pub.	F	Backbone	MCA	MPCA
SAM	ECCV'20	✗	ResNet-18	86.3%	93.1%
Dual-AI	CVPR'22	✓	Inception-v3	-	**96.5%**
DFWSGAR	CVPR'22	✗	ResNet-18	90.5%	94.4%
LRMM+GCM	ImaVis'23	✗	ResNet-18	92.8%	95.6%
Flaming-Net	ECCV'24	Tr	Inception-v3	**93.3%**	95.2%

F = Optical Flow, Tr = Use optical flow only in training

Efficacy of auxiliary losses

$\mathcal{L}_{gf}$	$\mathcal{L}_{tco}$	$\mathcal{L}_{flm}$	MCA	MPCA
✗	✗	✗	75.3%	70.1%
✓	✗	✗	78.3%	72.3%
✓	✓	✗	77.5%	73.0%
✓	✗	✓	77.5%	73.0%
✓	✓	✓	79.1%	76.0%

Effect of actor relation module

Actor Relation	MCA	MPCA
Actor Motion Path (AMP)	74.3%	73.9%
AMP + Losses	77.1%	71.7%
S.T Transformer Dual Path	77.0%	70.7%
Flaming-Net	79.1%	76.0%

More results and analysis on our paper

Visit our lab at: cilabs.kaist.ac.kr