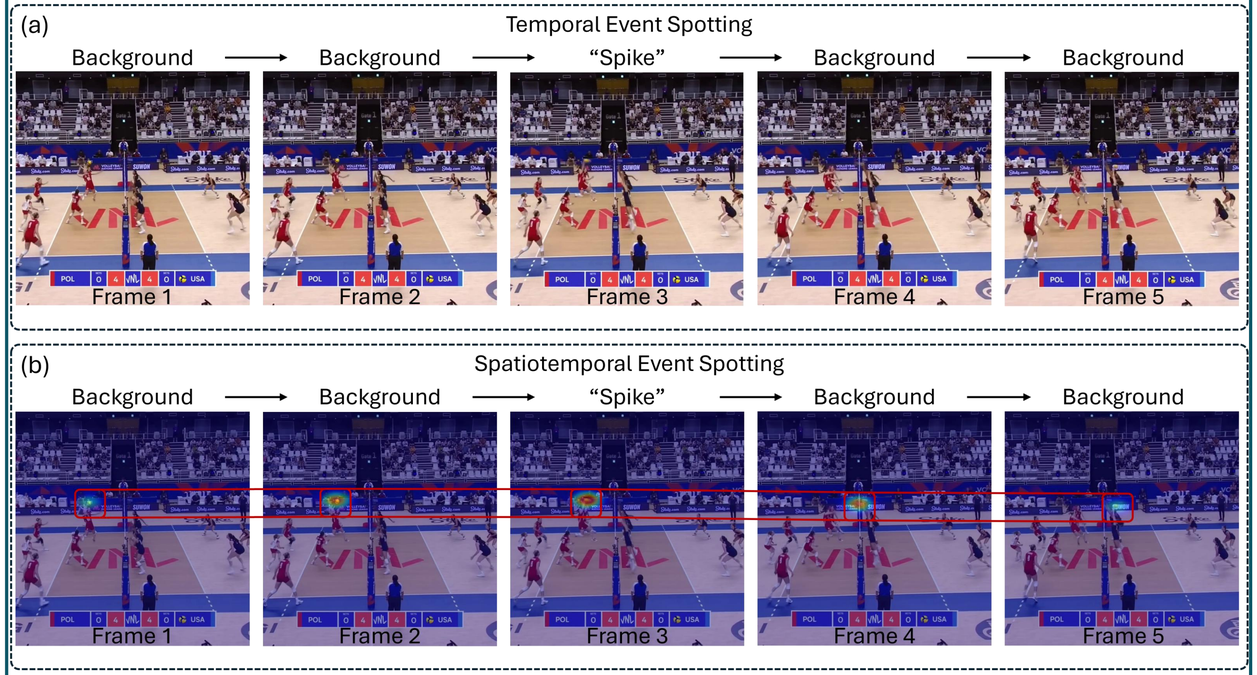
Spatiotemporal Event Spotting via 3D Heatmaps with Dynamically Shifted Gaussian Kernels

Ankhzaya Jamsrandorj, Vanyi Chao, Hoang Quoc Nguyen, Yin May Oo, Muhammad Amrulloh Robbani, Yewon Hwang, Kyung-Ryoul Mun, Jinwook Kim

University of Science and Technology, Korea Institute of Science and Technology, Seoul, Republic of Korea

Introduction



Temporal Event Spotting vs Spatiotemporal Event Spotting

- Most event spotting methods focus solely on temporal localization, detecting when events occur, while neglecting where they happen, which is essential for complete event understanding.
- Our approach enables Spatiotemporal Event Spotting (STES), detecting both the precise temporal frame and spatial location of discrete events.

Contributions

- Introduced new STES task for sports videos.
- Created VNL dataset
- Proposed novel 3D heatmap with Dynamically Shifted Gaussian Kernels (DSGK) that achieves state-of-the-art results

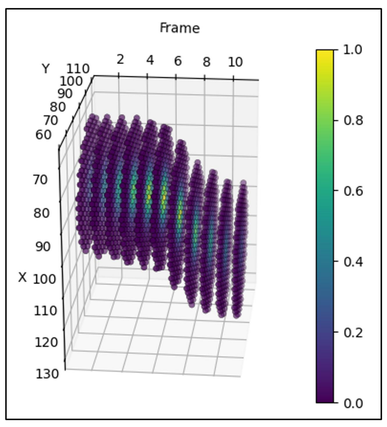
Dataset Overview

Number of Games	Number of Rallies	Duration (min)	Number of Instances							
			Serve	Receive	Set	Spike	Block	Dig	Pass	Score
24 (12 female & 12 male)	3518	579.5	3497	3008	4394	4634	2182	2775	488	2660

- Annotations include spatiotemporal event labels, per-frame ball positions, and court homography data.

3D Heatmaps with DSGK

- Our 3D heatmap representation uses Gaussian kernels that dynamically shift spatially across temporal frames, following the ball's trajectory. This encodes event likelihood in both space and time, with peaks at the annotated event location and frame.



Dataset and Visualizations

Contact:
VANYI CHAO
vanyi@kist.re.kr
Korea Institute of Science and Technology

Key equations for 3D Heatmaps with DSGK

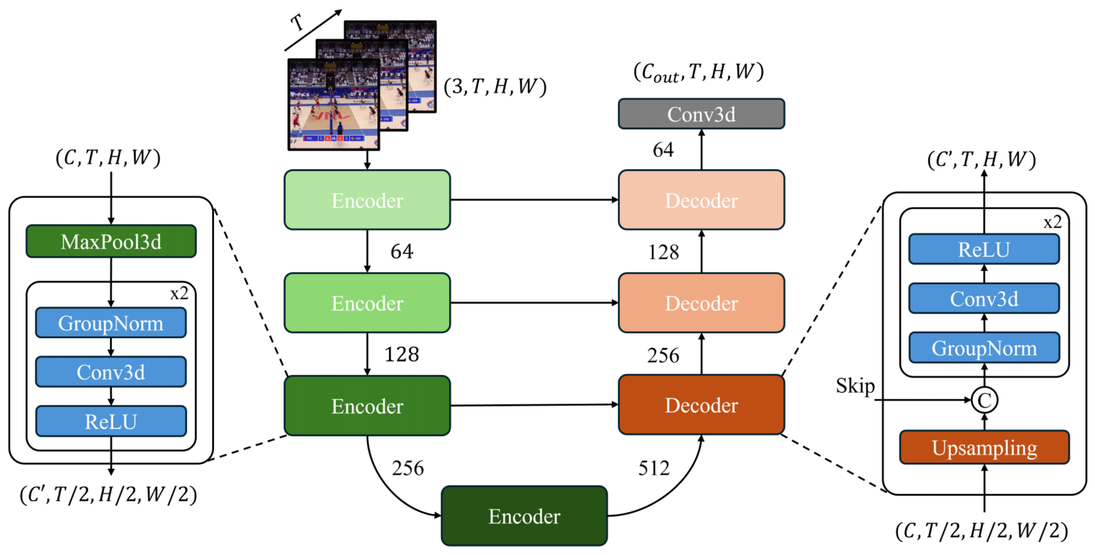
Temporal weights $$\omega(t) = \exp\left(-\frac{t^2}{2\sigma_t^2}\right)$$

3D heatmaps with DSGK $$H(x,y,t) = \omega(t)\exp\left(-\frac{\left(x - x_0(t)\right)^2 + \left(y - y_0(t)\right)^2}{2\sigma_s^2}\right)$$

t: temporal offset from the event frame (x_0, y_0): spatial event location

- The heatmap uses a temporal window of ± 5 frames, $\sigma_t = 3$ and $\sigma_s = 5$

Model Architecture



Modified 3D U-Net architecture

Results

Comparison of temporal event spotting performance on the VNL dataset.

Model	mAP@2			mAP@4		
	T-DEED	E2E-Spot	**Ours**	T-DEED	**E2E-Spot**	Ours
Overall	80.80	85.31	**85.63**	81.06	**86.69**	86.14
mAP per class:						
SERVE	90.53	99.80	**99.84**	90.53	99.80	**99.84**
RECEIVE	88.85	90.15	**90.65**	88.85	90.15	**90.65**
SET	88.11	**89.63**	89.38	88.30	89.46	**89.91**
SPIKE	90.63	**90.88**	90.68	89.95	**90.68**	**90.68**
BLOCK	70.37	82.82	**85.31**	72.30	83.64	**85.83**
DIG	81.68	85.93	**88.16**	81.74	86.31	**88.66**
PASS	54.82	**61.10**	54.36	55.37	**61.10**	56.81
SCORE	81.40	86.08	**87.00**	81.45	86.08	**87.00**

Comparison of temporal event spotting performance on the SoccerNet BAS

Model	Tight Average-mAP			Loose Average-mAP		
	E2E-Spot	T-DEED	**Ours**	E2E-Spot	T-DEED	**Ours**
Overall	35.46	36.54	**57.25**	42.17	44.01	**64.02**
mAP per class:						
PASS	77.06	81.61	**88.58**	86.21	86.55	**89.01**
DRIVE	76.07	79.10	**88.35**	86.08	85.91	**88.78**
HEADER	46.24	46.65	**81.32**	51.66	60.66	**89.05**
HIGH PASS	51.15	64.83	**84.43**	57.52	68.26	**87.95**
OUT	40.23	23.48	**69.15**	49.47	35.79	**82.25**
CROSS	49.40	48.14	**70.64**	53.25	58.08	**84.40**
THROW IN	0.000	0.000	**61.10**	0.000	0.000	**76.47**
SHOT	22.08	32.46	**74.63**	28.17	51.54	**90.38**
BALL P. BLOCK	8.790	6.750	**68.80**	17.51	18.30	**79.96**
PLAYER S. TACKLE	0.000	**3.930**	0.000	3.100	**8.440**	0.000
FREE KICK	**54.55**	**54.55**	0.000	**72.11**	54.55	0.000
GOAL	0.000	0.000	0.000	0.000	0.000	0.000

Heatmap Representation	Spatial mAP (δ_S)				Temporal mAP (δ_T)			
	@2	@4	@6	@8	@1	@2	@3	@4
Standard 3D volumetric heatmap	60.73	73.37	76.13	77.38	60.19	62.15	62.89	64.83
Dynamically Shifted 3D heatmap	71.10	82.95	88.22	91.16	80.88	85.63	85.90	86.14

Standard 3D Gaussian heatmaps vs 3D heatmaps with DSGK

CONCLUSIONS

- Introduced STES task for joint temporal and spatial event spotting in sports videos
- Proposed 3D heatmap with DSGK achieves SOTA results with 85.63 mAP@2 (VNL) and 57.25 tight mAP (SoccerNet BAS)

This work was supported by Development of foot diagnosis and shoe manufacturing technology for the disabled Project (RS-2022-00164554).