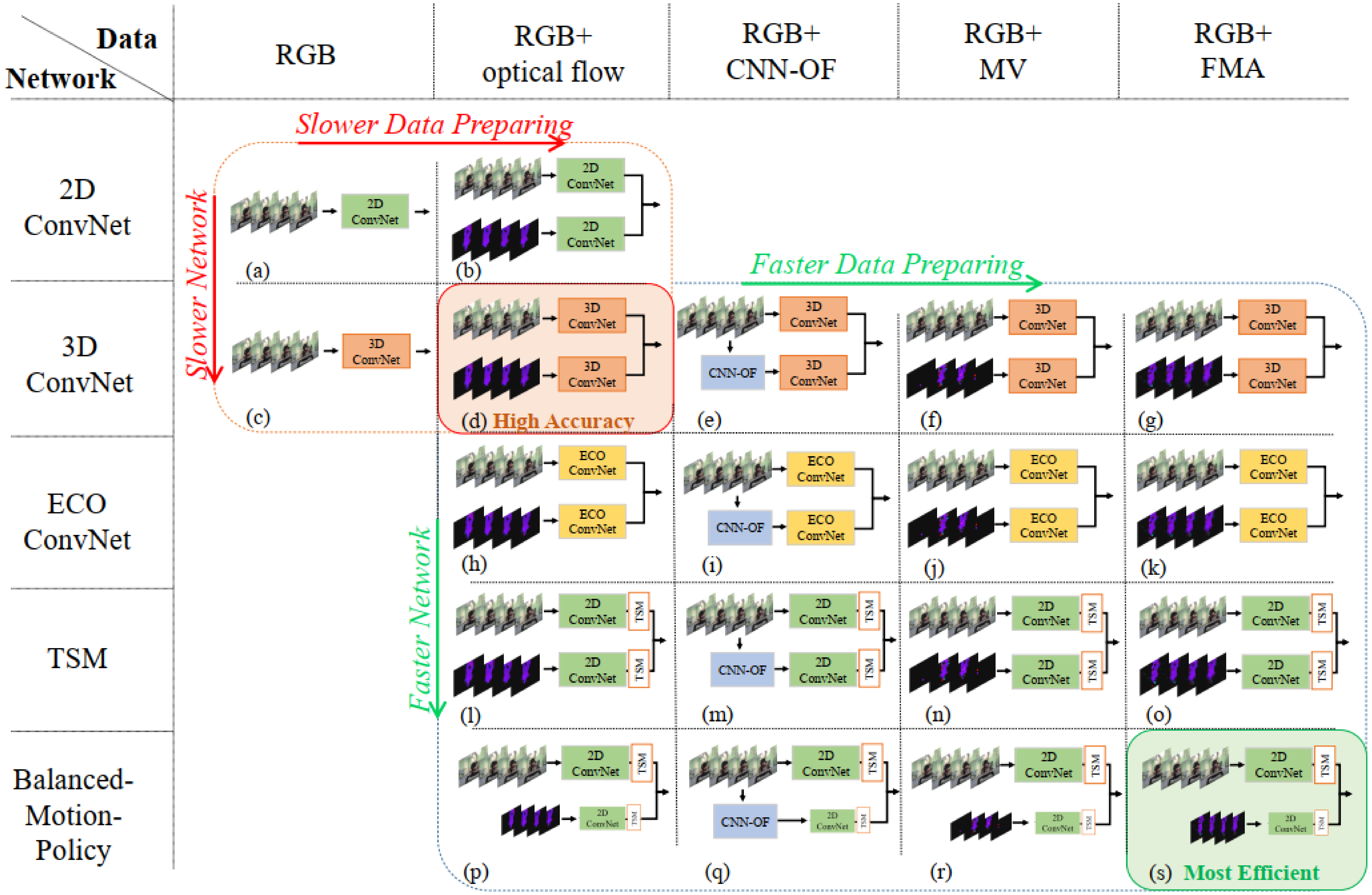
Motion Complementary Network for Efficient Action Recognition

Ke Cheng[1,2], Yifan Zhang[1,2], Chenghua Li[1,2], Jian Cheng[1,2,3], Hanqing Lu[1,2]
[1]NLPR & AIRIA, Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]CAS Center for Excellence in Brain Science and Intelligence Technology

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Method

We summarize our contribution in this figure. In this figure, every column denotes a data modality, and every row denotes a network. The arrows show the research development process. Previous researchers start from (a) and achieve high accuracy at (d). Then some researchers use fast optical flow to achieve (e)(f), while other researchers use fast spatial-temporal networks to achieve (h)(l). We propose *fixed-motion-accumulation* (FMA) for fast data preparing and *balanced-motion-policy* (BMP) for fast network. With these two techniques, our EMC-Net is the first work that efficiently utilizes the complementary information between motion vector and spatial-temporal network, illustrated in (s).

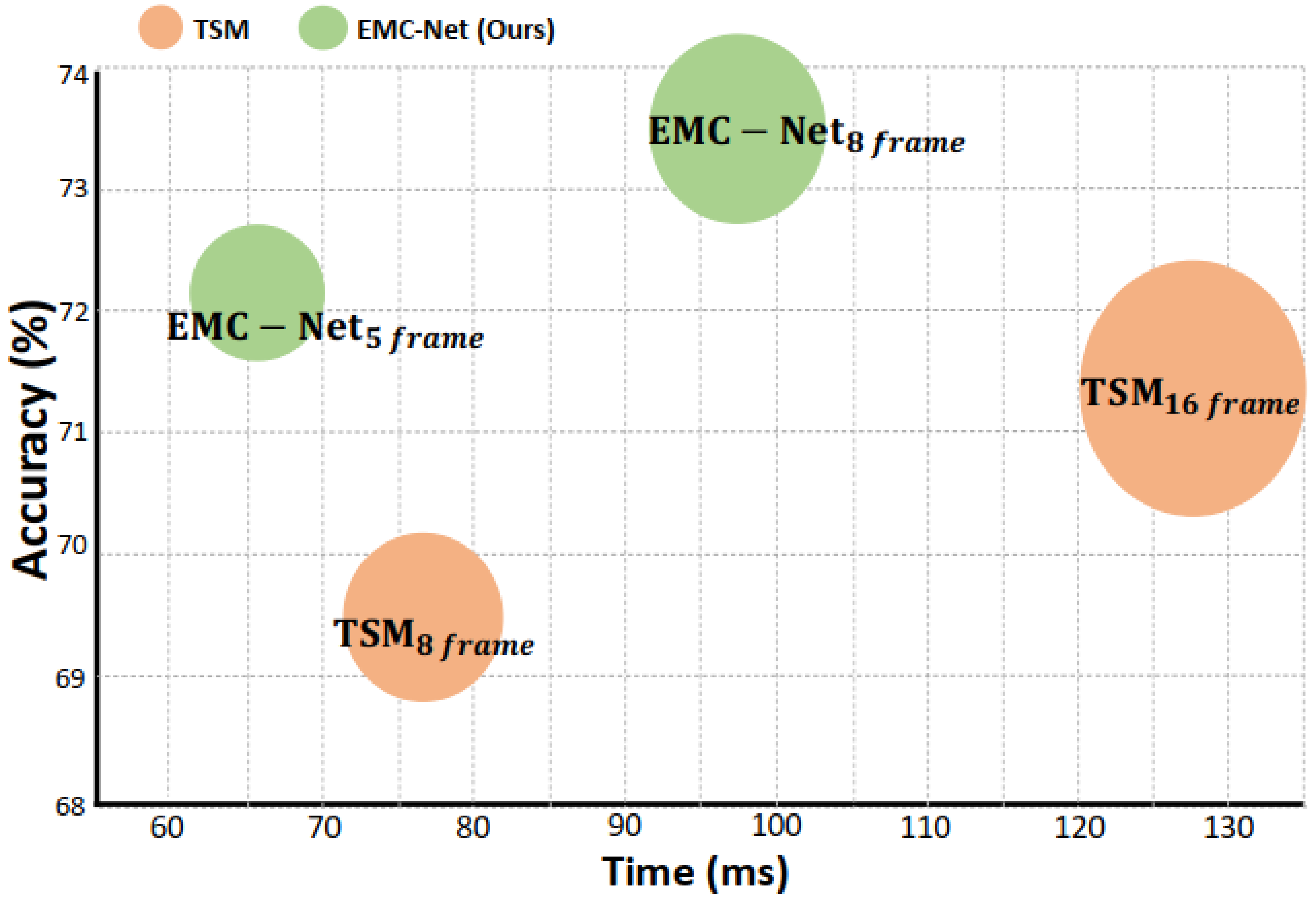


Results

We conduct extensive experiments on Kinetics, UCF101, and Jester datasets.

Dataset	Frames	Model	FLOPs	top 1
Kinetics	8	TSM	33G	69.35
	5	TSM	21G	68.25
	5	**EMC-Net**	**23G**	**72.01**
UCF101	8	TSM	33G	93.61
	5	TSM	21G	92.58
	5	**EMC-Net**	**23G**	**93.71**
Jester	8	TSM	33G	94.40
	5	TSM	21G	93.97
	5	**EMC-Net**	**23G**	**94.42**

On Kinetics dataset, we achieve 2.6% better performance than TSM (Lin et al., 2019) with 1.4× fewer FLOPs and 10ms faster on K80 GPU.


TSM
EMC-Net (Ours)
Accuracy (%)
Time (ms)
$\mathrm{EMC-Net}_{8\ frame}$
$\mathrm{EMC-Net}_{5\ frame}$
$\mathrm{TSM}_{16\ frame}$
$\mathrm{TSM}_{8\ frame}$