

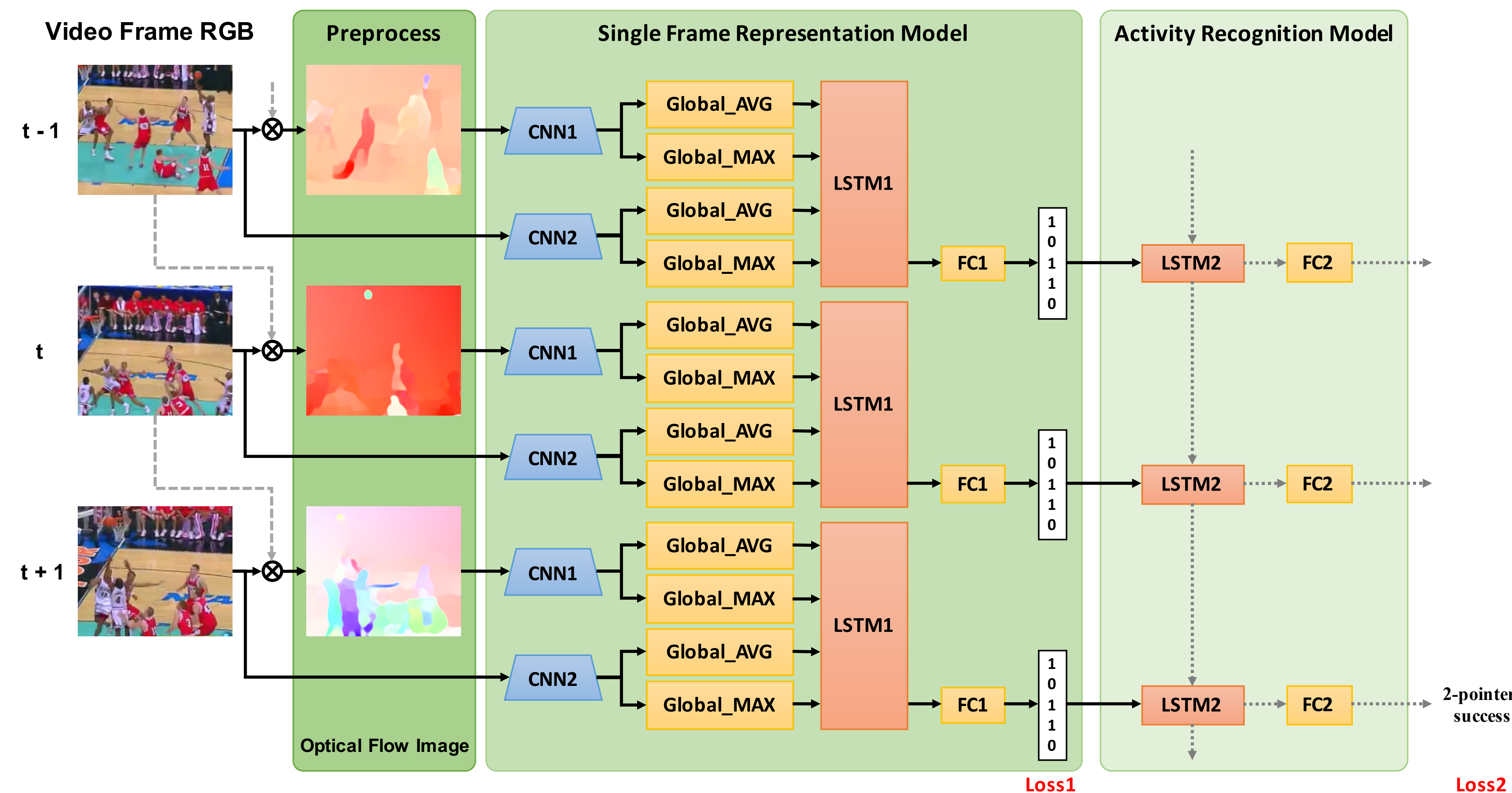
Introduction

We propose a novel robust and efficient human activity recognition scheme called ReHAR, which can be used to handle single person activities and group activities prediction.

Our Contributions:

1. Design a robust and efficient human activity recognition scheme to recognize complex human activities, e.g., group activities in sport games.
2. Extensive evaluation using two popular activity datasets show that our scheme achieves higher accuracy and runs an order of magnitude faster than existing schemes.
3. Explore the visual explanation for our model to understand what it has learned.

Proposed Solution



Experimental Results

NCAA Basketball Dataset:

	3point S.	3point F.	throw S.	throw F.	layup S.	layup F.	2point S.	2point F.	dunk S.	dunk F.	steal	Mean
IDT[4]	0.370	0.501	0.778	0.365	0.283	0.278	0.136	0.303	0.197	0.004	0.555	0.343
IDT[4] player	0.428	0.481	0.703	0.623	0.300	0.311	0.233	0.285	0.171	0.010	0.473	0.365
C3D[5]	0.117	0.282	0.642	0.319	0.195	0.185	0.078	0.254	0.047	0.004	0.303	0.221
MIL[6]	0.237	0.335	0.597	0.318	0.257	0.247	0.224	0.299	0.112	0.005	0.843	0.316
LRCN[7]	0.462	0.564	0.876	0.584	0.463	0.386	0.257	0.378	0.285	0.027	0.876	0.469
Atten. no track[8]	0.583	0.668	0.892	0.671	0.489	0.426	0.281	0.442	0.210	0.006	0.886	0.505
Atten. track[8]	0.600	0.738	0.882	0.516	0.500	0.445	0.341	0.471	0.291	0.004	0.893	0.516
Ours	0.753	0.766	0.933	0.857	0.613	0.435	0.405	0.542	0.232	0.007	0.940	0.589

UCF Sports Action Dataset :

	Diving	Golf	Kicking	Lifting	Riding	Run	SkateB	Swing	SwingB	Walk	mAP
Gkioxari et al. [9]	0.758	0.693	0.546	0.991	0.896	0.549	0.298	0.887	0.745	0.447	0.681
Weinzaepfel et al. [10]	0.607	0.776	0.653	1.000	0.995	0.526	0.471	0.889	0.629	0.644	0.719
Peng et al. [11]	0.961	0.805	0.735	0.992	0.976	0.824	0.574	0.836	0.985	0.760	0.845
Hou et al. [12]	0.844	0.908	0.865	0.998	1.000	0.837	0.687	0.658	0.996	0.878	0.867
Ours	1.000	0.955	1.000	1.000	1.000	0.806	0.626	1.000	1.000	0.888	0.928

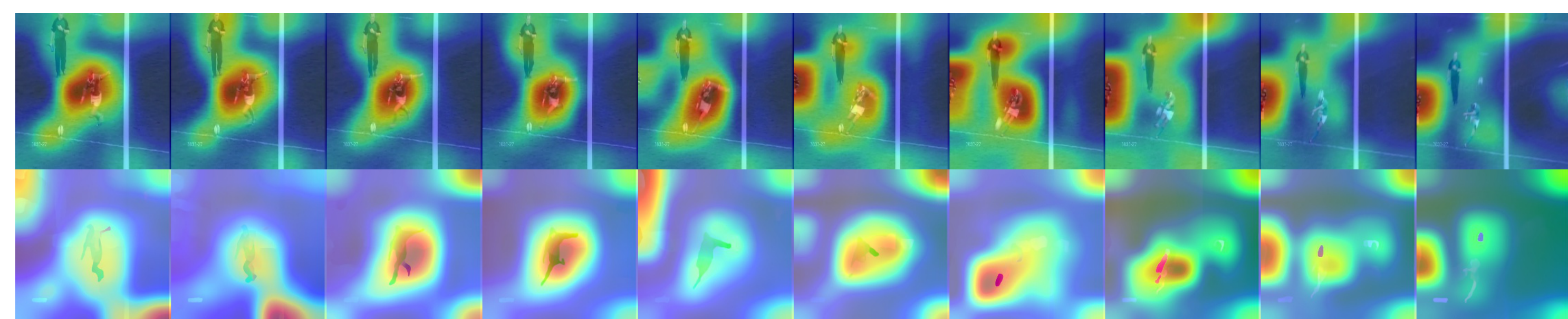
Our Model	No Global Layers	LSTM1 -> Convolutional layer	LSTM1 -> element-wise sum
0.928	0.889	0.766	0.702

Future Work

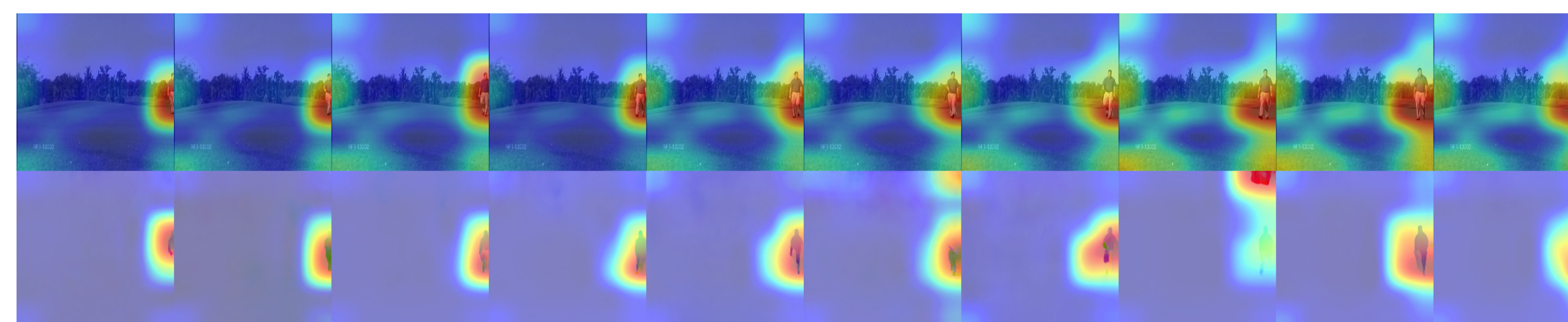
1. Try different CNN model as base net, e.g. C3D or MobileNet.
2. Evaluate our model on larger dataset, e.g. UCF101 or THUMOS.

This work is partially supported by a Qualcomm gift and a GPU donated by NVIDIA.

Visual explanation



Correctly predict a "Kicking" event.



Incorrectly predict a "Walking" event as "Golf".