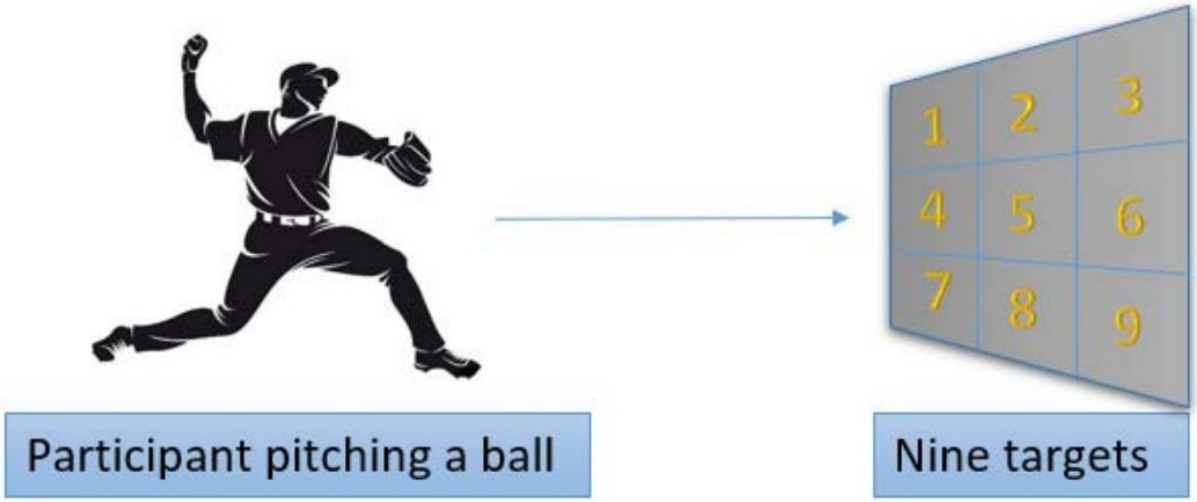
Human Intention Prediction Using Two-Stream Spatio-Temporal Features

Shengchao Li, Lin Zhang, and Xiumin Diao
School of Engineering Technology, Purdue University
Phone: +1 (765) 494-2212; Email: diaox@purdue.edu

ACM/IEEE International Conference on
Human-Robot Interactions
March 11-14, 2018, Daegu, Korea

PURDUE
POLYTECHNIC

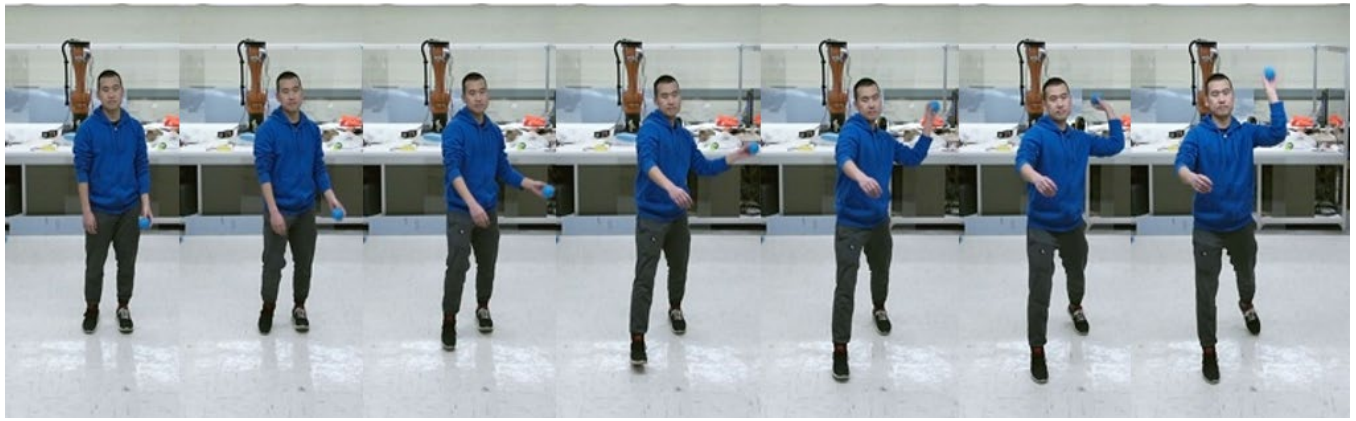
Experiment Overview



The experiment is about a human participant pitching a ball toward a robot (represented by a target area in the figure). The target area is divided into 9 grids which represent the 9 targets that the ball can hit. We expect "the robot" to be able to predict the intention of the participant, namely, which of the nine targets the participant is pitching the ball to.

PURDUE
POLYTECHNIC

Training Data



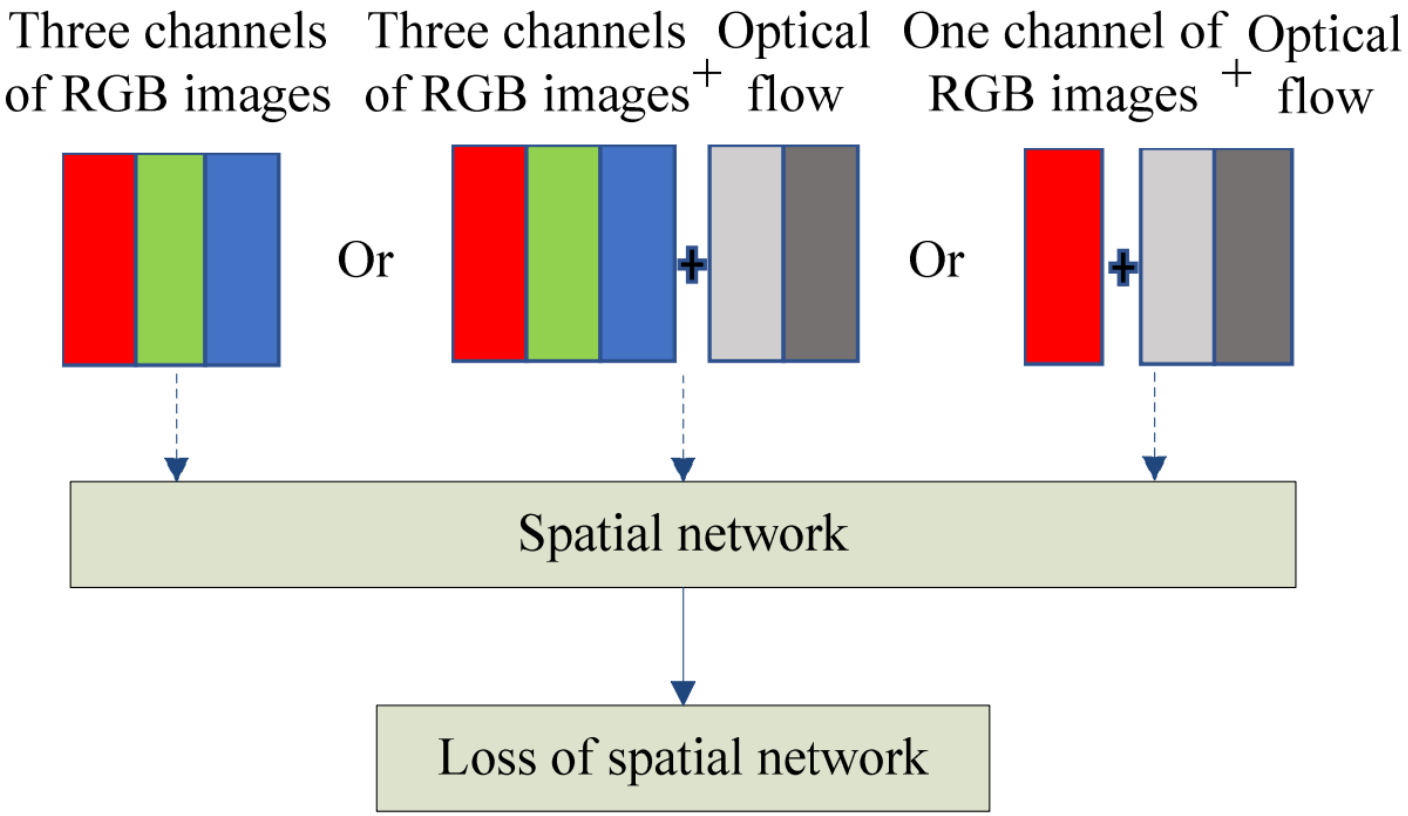
RGB images sequence

Optical flow sequence

PURDUE
POLYTECHNIC

Approach

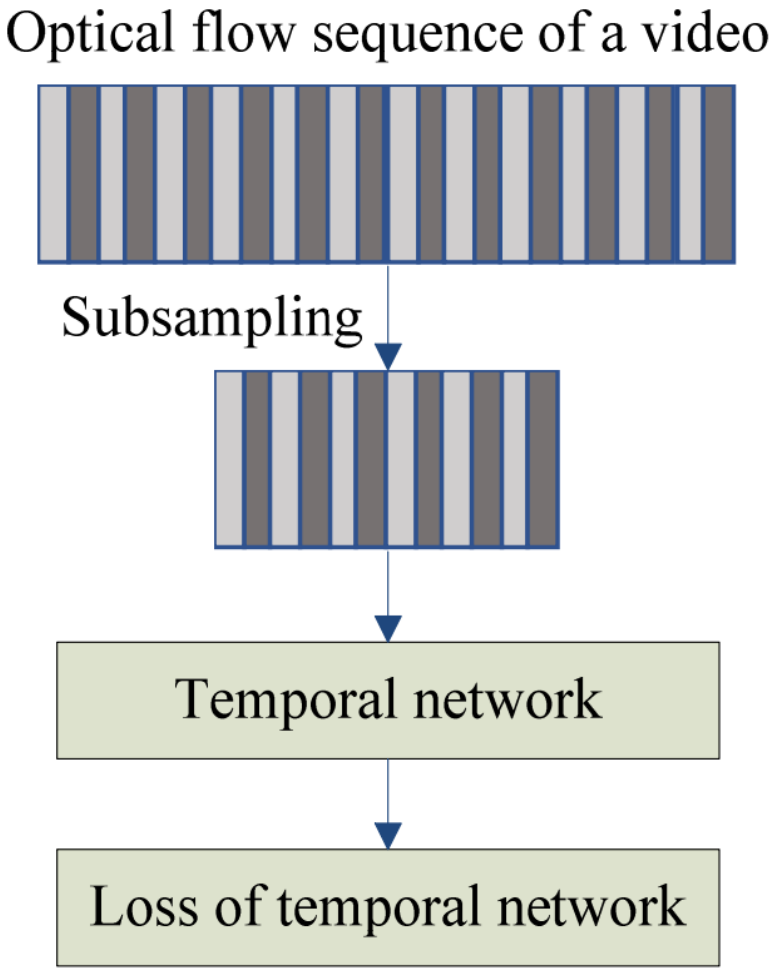
Spatial network:
Optimize early fusion methods of RGB images and optical flow for spatial network.



PURDUE POLYTECHNIC

Approach

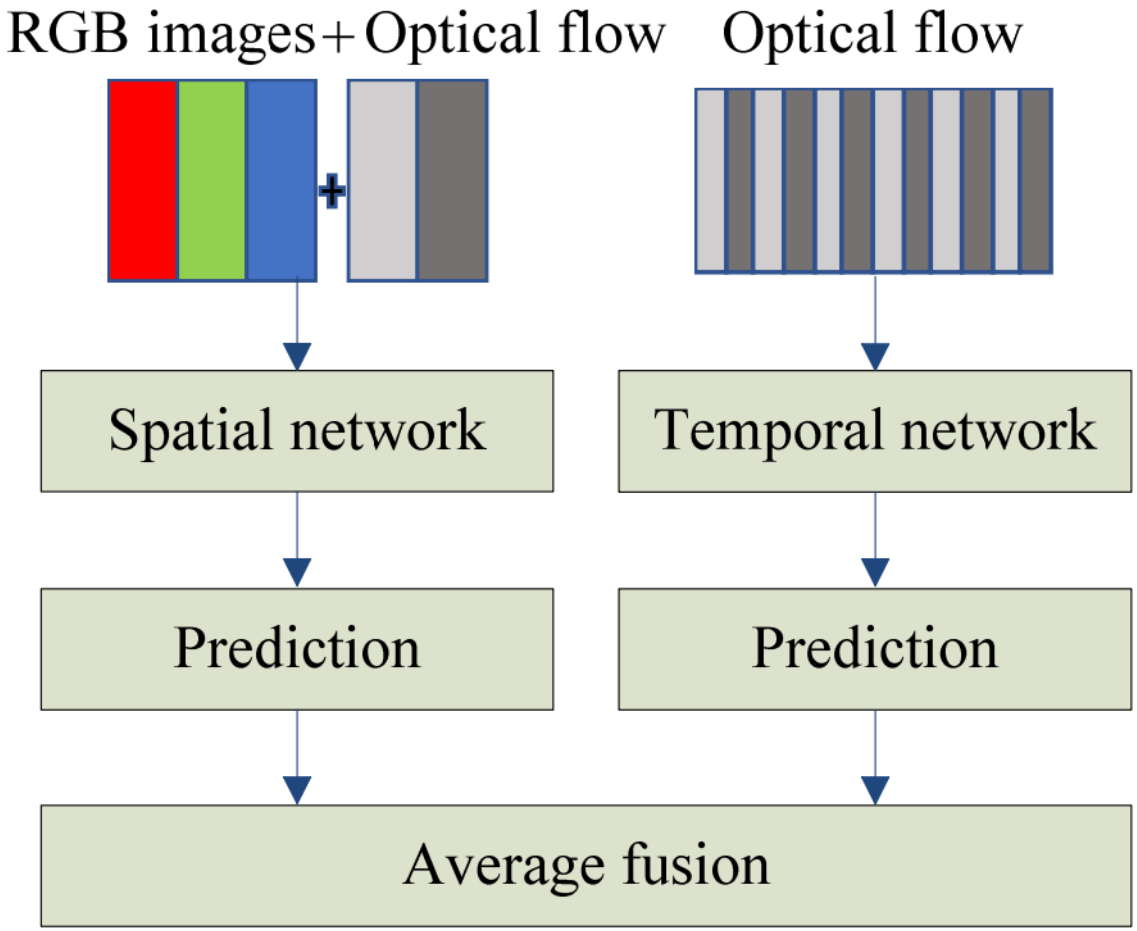
Temporal network:
Capture motion information in temporal network.



PURDUE
POLYTECHNIC

Approach

Two-stream network with average fusion:
Final average fusion to improve prediction performance.



PURDUE POLYTECHNIC

Results

Spatial Network Method	Data Augmentation	Prediction Accuracy
Three channels of RGB images	No	50%
	Yes	57.4%
Three channels of RGB images + two channels of optical flow	No	53%
	Yes	57.4%
One channel of RGB images + two channels of optical flow	No	51.8%
	Yes	63.89%

Temporal Network Method	Data Augmentation	Prediction Accuracy
Temporal network without fusion	No	68.5%
	Yes	69.4%
Final fusion of spatial network and temporal network	Yes	71.3%

Thank you!

PURDUE
POLYTECHNIC