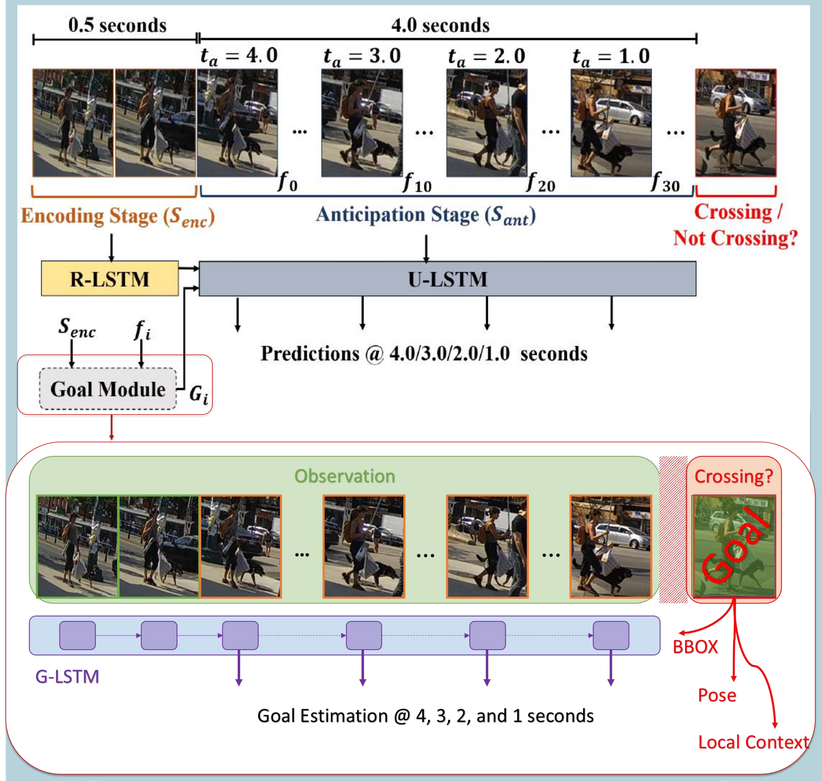
EARLY PEDESTRIAN INTENT PREDICTION VIA FEATURES ESTIMATION

Nada Osman, Enrico Cancelli, Guglielmo Camporese, Pasquale Coscia and Lamberto Ballan

Univeristy of Padova - Department of Mathematics "Tullio Levi-Civita"
Vision Intelligence and Machine Perception Research Group

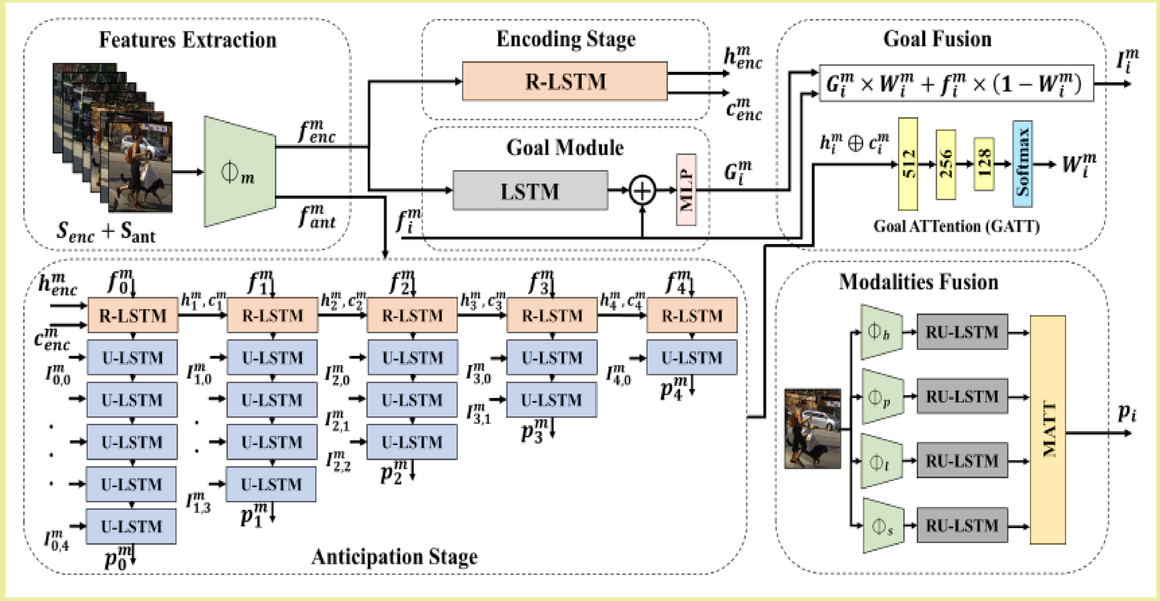
ICIP2022 BORDEAUX
18 OCT
Paper ID: 3138

INTRODUCTION



- We aim to revise this standard evaluation protocol to forecast crossing events as early as possible (max $t_a = 4.0$s instead of max $t_a = 2.0$s).
- Conceiving a solution upon an extensively used model for egocentric action anticipation (RU-LSTM).
- We proposing to envision future features, that can better infer human intentions using an attention-based fusion mechanism.
- Anticipating pedestrian intents several seconds in advance for sure will improve human safety and social awareness.
- We mainly focus on two largely adopted benchmarks, namely JAAD (Rasouli et al. 2017) and PIE (Rasouli et al. 2019).

METHOD



1. Features Extraction: Applies Φ_m to extracts m type features from raw images.
2. Encoding stage: encodes the motion history with length S_{enc}.
3. Goal Module: it predicts goal G_i^m associated to the modality m.
4. Goal Fusion: fuses G_i^m and f_i^m to get new I_i^m.
5. Anticipation Stage: predicts the crossing probability P_i^m at step i.
6. Modalities Fusion: attention-based modalities fusion, producing P_i.

EXPERIMENTS

	4 s			3 s			2 s			1 s			[2-1] s		
$JAAD_{beh}$	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
PCPA [15]	-	-	-	-	-	-	-	-	-	-	-	-	0.58	0.50	0.71
PCPA† [15]	0.54	0.51	0.62	0.47	0.46	0.54	0.49	0.45	0.61	0.45	0.52	0.63	0.56	0.50	0.67
R-LSTM	0.67	0.64	0.74	0.70	0.64	0.78	0.66	**0.62**	0.75	0.65	0.60	0.75	0.65	0.59	0.74
RU-LSTM	**0.72**	**0.67**	**0.79**	**0.72**	**0.64**	**0.81**	**0.69**	**0.62**	**0.78**	**0.70**	**0.63**	**0.79**	**0.69**	**0.62**	**0.78**
$JAAD_{all}$															
PCPA [15]	-	-	-	-	-	-	-	-	-	-	-	-	0.85	**0.86**	**0.68**
PCPA† [15]	0.75	0.75	0.50	0.74	0.76	0.53	0.72	0.75	0.52	0.76	**0.79**	0.55	0.80	0.79	0.57
R-LSTM	0.83	0.74	0.54	0.86	0.73	0.58	**0.85**	0.73	0.57	**0.87**	0.77	**0.62**	**0.86**	0.76	0.60
RU-LSTM	**0.84**	**0.76**	**0.57**	**0.87**	**0.78**	**0.64**	**0.85**	**0.76**	**0.59**	0.86	0.78	**0.62**	**0.86**	0.78	0.62
PIE															
PCPA [15]	-	-	-	-	-	-	-	-	-	-	-	-	**0.87**	**0.86**	**0.77**
PCPA† [15]	0.76	0.75	0.62	0.77	0.76	0.63	0.83	**0.84**	0.73	0.86	**0.85**	0.77	0.86	**0.86**	**0.77**
R-LSTM	0.75	0.64	0.48	0.75	0.66	0.50	0.76	0.66	0.51	0.76	0.67	0.52	0.76	0.67	0.52
RU-LSTM	**0.77**	**0.76**	**0.63**	**0.80**	**0.79**	**0.68**	**0.85**	0.82	**0.74**	**0.88**	**0.85**	**0.79**	**0.87**	0.84	**0.77**

Comparison of baseline models
PCPA † (kotseruba et al. 2021) denotes our retrained PCPA model

	$JAAD_{beh}$			$JAAD_{all}$		
	Acc	AUC	F1	Acc	AUC	F1
PCPA [15]	0.58	0.50	0.71	0.85	**0.86**	**0.68**
CAPformer [17]	-	0.55	0.76	-	0.82	0.63
TrouSPI-Net [19]	0.64	0.56	0.76	0.82	0.77	0.58
RU-LSTM	0.69	0.62	0.78	**0.86**	0.78	0.62
G-RULSTM	**0.72**	**0.65**	**0.80**	**0.86**	0.80	0.63
Imp.	**3%**	**3%**	**2%**	-	**2%**	**1%**

G-RULSTM VS SOTA models

Ablation of goal fusion techniques

	$JAAD_{beh}$				
	4 s	3 s	2 s	1 s	[2-1] s
Bounding Box					
Without Goal	0.55	0.59	0.63	0.64	0.58
Interpolation	**0.61**	0.56	0.61	**0.65**	**0.64**
Concatenation	**0.61**	**0.64**	0.63	0.63	0.63
Attention	**0.61**	0.63	**0.65**	**0.65**	**0.64**
Local Context					
Without Goal	0.68	0.66	0.65	0.67	0.66
Interpolation	0.73	0.67	0.63	0.66	0.61
Concatenation	0.65	**0.70**	**0.68**	0.68	**0.68**
Attention	**0.69**	0.67	0.66	**0.69**	0.66

Code: https://github.com/NadaSOsman/EarlyPedestrianActionPrediction Email: nadasalahmahmoud.osman@phd.unipd.it