AIを用いたリアルタイム物体検出

監視映像における物体検出のための特徴量強化

NII

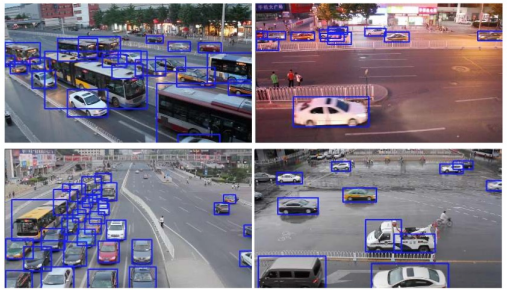
どんな研究？

道路状況の把握などのため，精度の高い車両検出技術が求められていますが，深層学習を持ちいた検出技術は精度が高い反面処理速度が遅いです．そこで精度と速度のトレードオフを改善するため，少量の過去フレームを活用し，効率よく重ね合わせることで，ブレなどに強い情報を作りリアルタイム処理可能な車両検出技術の研究を進めています．

何がわかる？

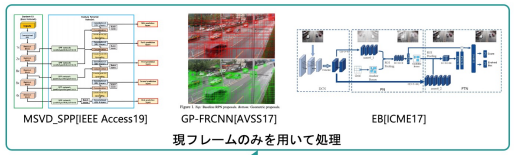
映像中の物体検出とは？
何が優れている？
問題点は？
解決のアプローチは？

車両検出とは？既存研究は？



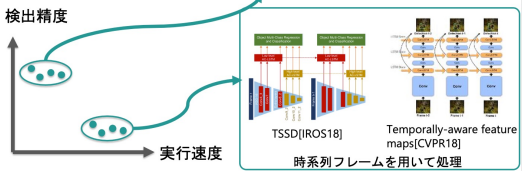
車両検出は？

・映像中の車両の領域を検出する技術
・静止画に比べて，ボケや車両同士が重なってしまうため困難



既存研究の課題は？

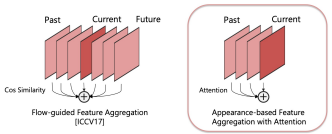
・高精度を追求するために，処理が遅くリアルタイム処理不可
・処理が早い検出器は精度が低い



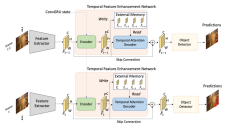
研究内容

解決のアプローチ

・処理速度の早い特徴抽出器の荒い特徴マップを時間方向に集約し，検出精度を向上
・先行研究の各フレームの類似度から結合の重みを決定から，アテンションにより一括算出



提案手法全体図



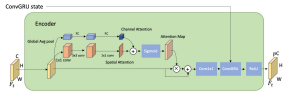
時刻t-1, tにおける提案モデル全体像

・既存の静止画検出器にTFEN(Temporal Feature Enhancement Network)導入
・特徴量抽出器からの特徴マップを改善することで精度を向上

新規性

アテンションベースによる時系列特徴マップの結合

Spatiotemporal Encoder



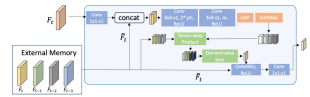
役割

・ConvGRUにより時系列情報を組み込んだ特徴マップの生成
・空間＆チャネルアテンションにより物体領域を強調
・計算量削減のため，チャネルを圧縮

新規性

時系列情報に空間＆チャネルアテンションの導入

Temporal Attention Decoder & External Memory



役割

・F_tと時系列特徴マップ$\hat{F}_t$から現在・過去フレームに対する重み付けを算出
・External Memory内の特徴マップと重み付けの積の重ね合わせより特徴マップの生成

新規性

アテンションによる過去情報を一括集約した特徴量の生成

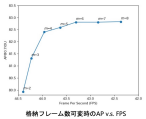
格納フレーム数が精度と速度に与える影響

検証

最大収容過去フレームmを可変とし，評価時のmを観測（訓練・評価にはFP16を使用，APはOverallを使用）

結果

・格納フレーム数を増加させることで精度向上傾向
・精度が飽和し始めるのは$m = 6$
・精度・速度のトレードオフ点は$m = 4$~6にあると推測



格納フレーム数可変時のAP v.s. FPS

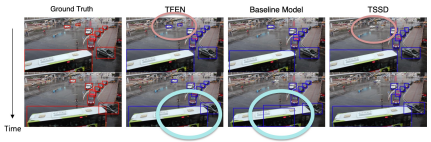
関連手法との比較

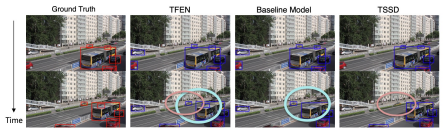
UA-DETRACデータセット　AP[%](比較：代表公式コードを元に訓練・評価)

	Model	Overall	Easy	Medium	Hard	Cloudy	Night	Rainy	Sunny	FPS	GPU
	DPM[2010]	20.79	34.42	20.24	17.62	24.70	20.81	23.55	24.77	0.56	-
	ACF[2014]	46.26	54.27	51.52	30.07	58.30	35.20	27.89	66.58	0.66	-
	RCNN[2014]	48.95	38.91	54.06	39.47	58.73	39.32	39.86	67.52	0.18	K40
	CompACT[2015]	53.23	64.88	58.79	43.76	63.23	46.37	44.21	71.16	0.02	K40
	Faster RCNN[2015]	58.45	82.75	63.05	44.25	62.34	66.29	45.16	69.05	11.0	Titan X
現フレーム	GP-FRCNN[2017]	74.57	91.79	80.65	66.05	85.16	81.23	68.59	77.28	4.8	K40
	EB[2017]	67.96	88.65	73.12	54.64	72.42	73.93	53.45	83.73	11	Titan X
	MSVD_SPP[2019]	**85.29**	96.04	89.42	**76.55**	88.00	88.67	**78.90**	88.91	9.5	Titan Xp
	YOLOv3-SPP[2018]	84.96	95.59	**89.95**	75.14	**88.52**	**88.81**	77.46	89.46	9.5	Titan Xp
	FG-BR_Net[2019]	76.06	93.49	83.68	70.78	87.26	78.42	70.50	89.09	18	1080
時系列	TSSD[2018]	57.16	81.06	62.07	43.14	57.59	63.87	44.86	67.73	**33.78**	2080 Ti
	TFEN(Ours)	82.40	**97.49**	86.98	72.18	87.54	82.41	72.32	**90.78**	**29.11**	2080 Ti

・先行研究(MSVD_SPP, YOLOv3-SPP)と比較し，同等の精度を保持しながら，実行速度は約3倍
・時系列情報を利用したTSSDとは実行速度に差はないものの，精度面にてパフォーマンスが低い

検出結果の比較





M. Fujitake and A. Sugimoto: Temporal Feature Enhancement Network with External Memory for Object Detection in Surveillance Video, Proc. ICPR2020, pp. 7684-7691, 2021.