Correlation-based ConvNet for Small Object Detection in Videos

Brais Bosquet and **Manuel Mucientes** and **Víctor M. Brea**

25th International Conference on Pattern Recognition (ICPR)

10 - 15 January 2021

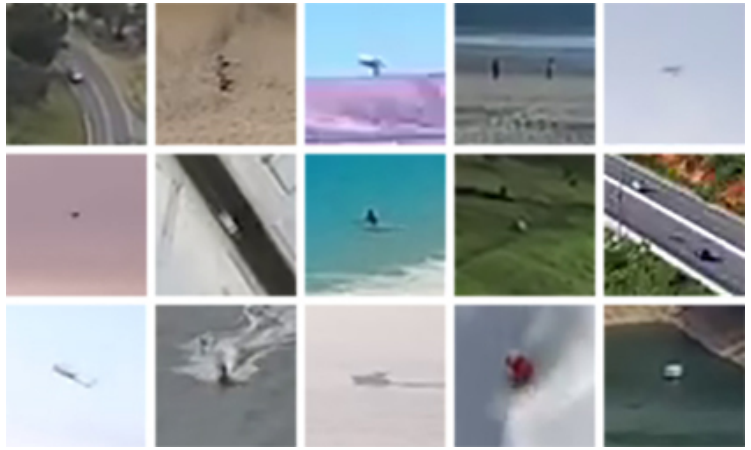
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

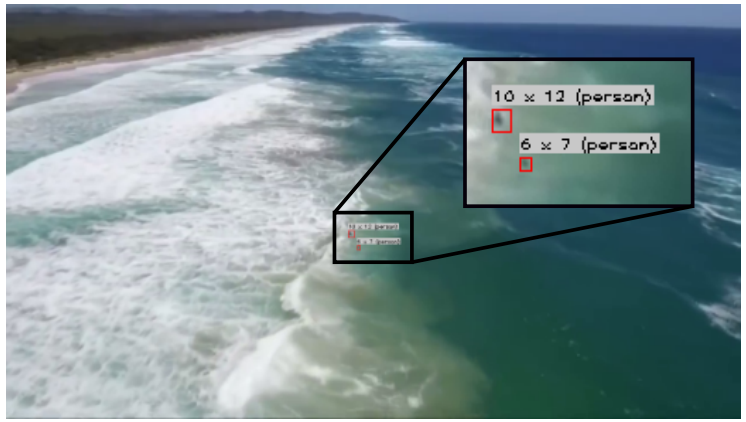
CAMPUS VIDA CAMPUS DE EXCELENCIA INTERNACIONAL

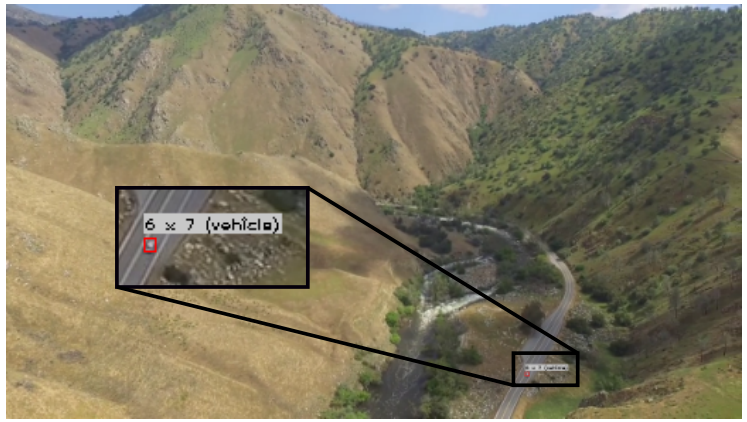
CiTIUS Centro Singular de Investigación en Tecnoloxías Intelixentes

Small Object Detection

- Objects **without definitive visual cues**
  - Sizes **under 16 × 16 in pixel area**
- **Accuracy lags behind** that of larger objects
- Critical in several **applications**:
  - Automated vehicle systems or sense and avoid problems
  - Satellite image analysis
  - Medical imaging, etc.



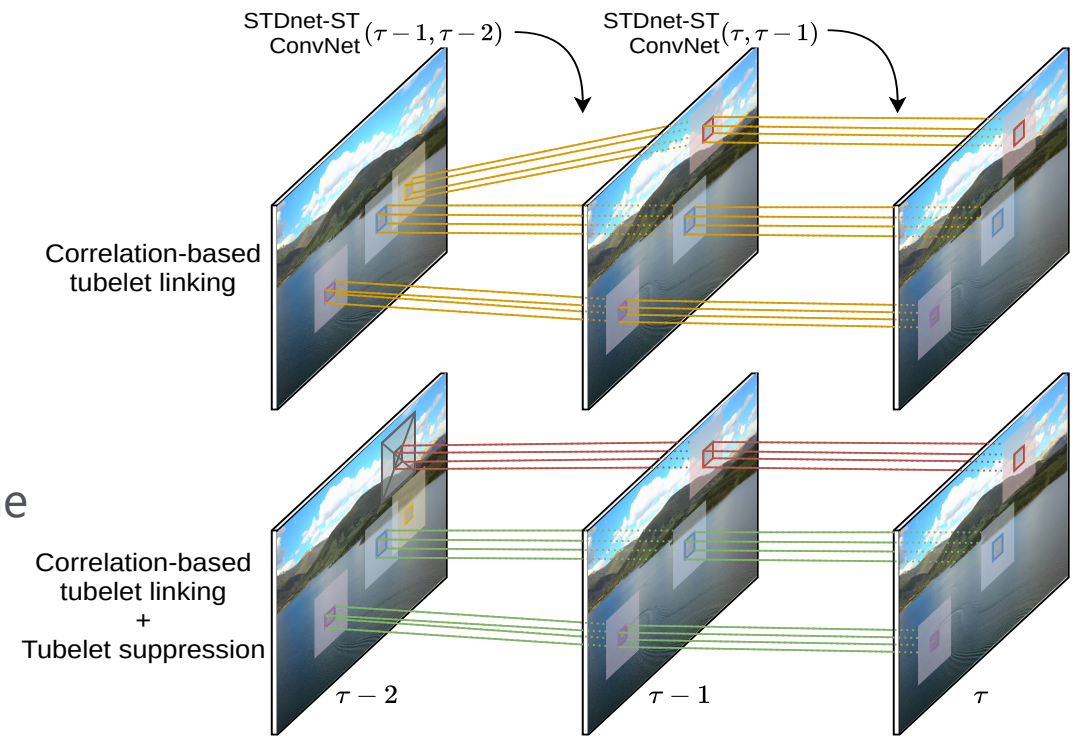




Spatio-temporal STDnet (STDnet-ST)

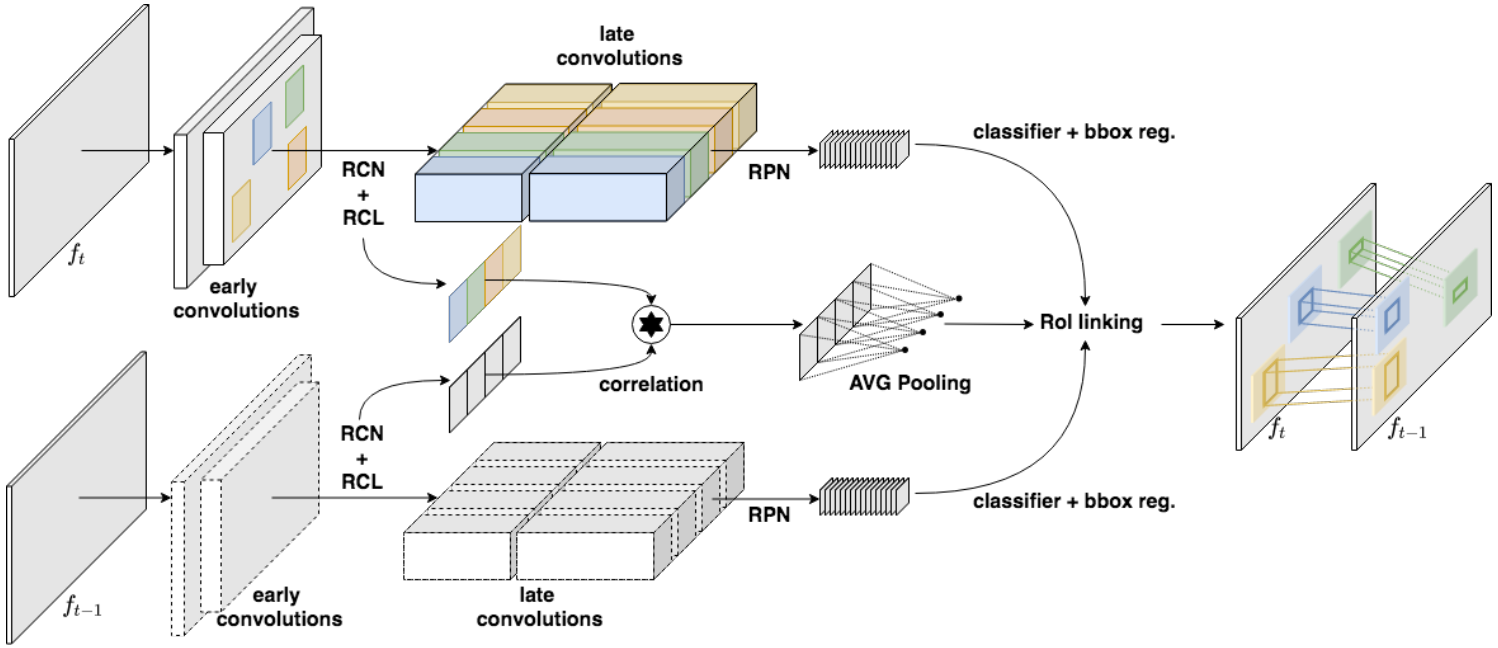
- STDnet-ST ConvNet
  - Processes **two frames**
  - Computes **detections** and **correlations**

- STDnet-ST tubelet linking
  - Compose **high-quality *tubelets***
  1. **Correlation-based** tubelet linking
     - Links object detections over time by finding the **most likely sequences**
     - Based on the Viterbi algorithm
  2. **Tubelet suppression** algorithm
     - Identify and remove incorrect data associations



STDnet-ST ConvNet Architecture

- **Two STDnet** branches
- **Correlation** module
  - RCN regions pairs
  - Propagate correlation score to final detections

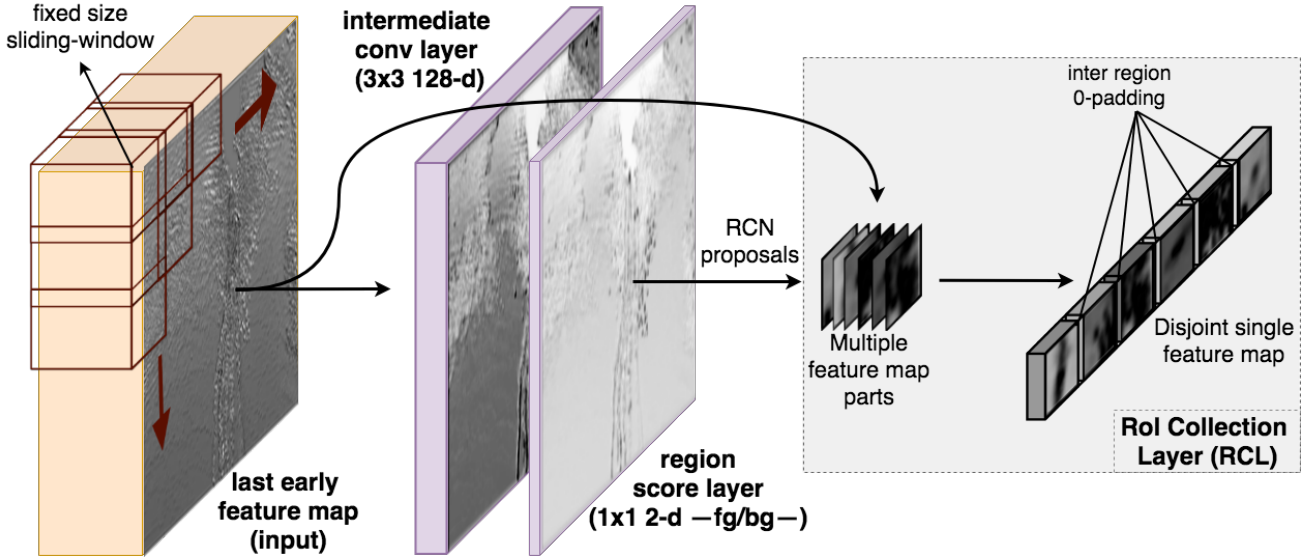


- Outputs:
  - **Detections** D_t and D_{t-1}
  - **Correlation** scores
    - D_t and D_{t-1}
    - **Free RCN regions**

STDnet-ST ConvNet Architecture

Region Context Network (RCN)[1]

- A fully convolutional network
- Scans a **shallow feature map**
- Fixed-size promising regions
- **RoI Collection Layer** (RCL)
  - Unify the feature map parts selected by the RCN
  - Compose a **disjoint single feature map**
- **86.5% saved memory**: $RCL_{output\ size} = \underbrace{(r_w n + p_d(n-1))}_{\text{width}} \times \underbrace{r_h}_{\text{height}}$



[1] B. Bosquet, M. Mucientes, and V. M. Brea. STDnet: *Exploiting high resolution feature maps for small object detection*. Engineering Applications of Artificial Intelligence, Vol. 91, No. 103615, 2020.

STDnet-ST tubelet linking

Correlation-based linking

- Update **the confidences of the current detections** using previous frames
- Baseline linking **drawbacks**:
  - **IoU** is **limited** for object linking when small objects
  - Viterbi algorithm generates **every possible *tubelets***

1. Compute **score matrix** for pair of frames
   - **Replaces IoU** with the **correlation score**

$$s_t^{ij} = p_{t-1}^{i} + p_t^{j} + \lambda \cdot c_t^{ij} \qquad c_t^{ij} = \rho(r_{t-1}^{k(i)}, r_t^{l(j)})$$

2. Generate *tubelets* using the **Viterbi algorithm**

$$p_\tau^{i(\hat{v})} = \frac{1}{\tau} \sum_{t=1}^{\tau} p_t^{i(\hat{v})}$$

3. Update confidences using **confidence variability**
   - Maximum: low variability -> true positives
   - Average: high variability -> some false positives

$$p_\tau^{i(\hat{v})} = \begin{cases} \max_{t=1}^{\tau} p_t^{i(\hat{v})} & \text{if } \sigma(\{p_t^{i(\hat{v})}\}_{t=1}^{\tau}) \leq \kappa \\ \frac{1}{\tau} \sum_{t=1}^{\tau} p_t^{i(\hat{v})} & \text{otherwise} \end{cases}$$

CiTIUS

STDnet-ST tubelet linking

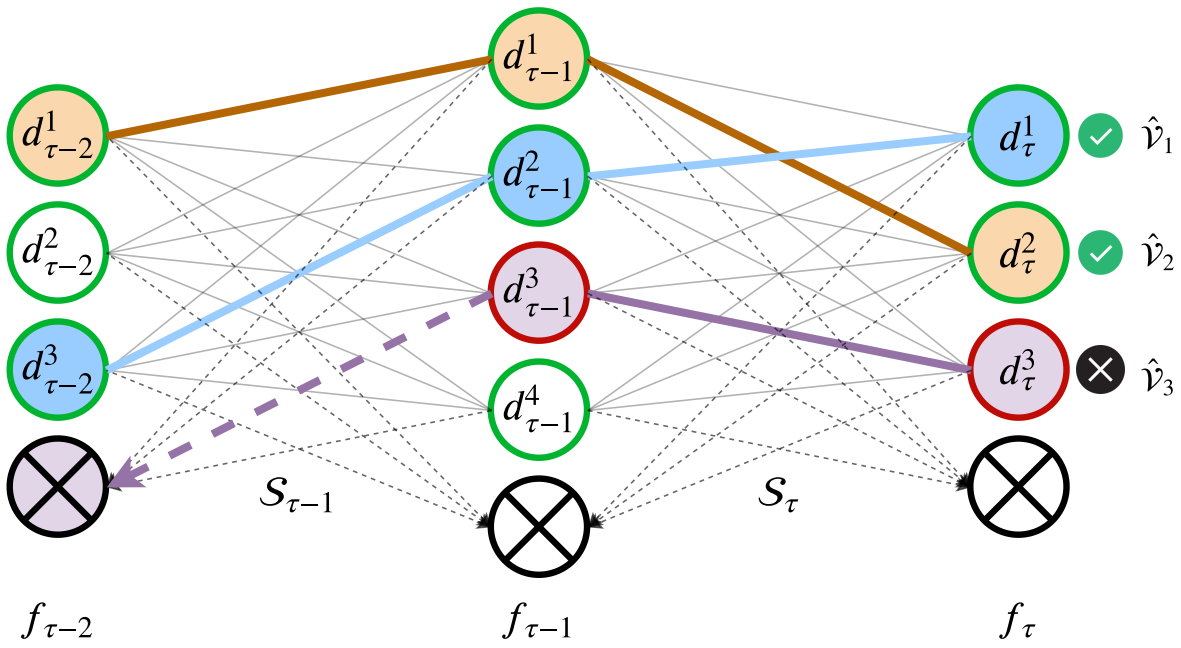
Tubelet suppression

- The **goal** is to avoid generating unlikely tubelets
  - **Remaining detections**

- **Dummy nodes** using free RCN regions
  - Aims to **associate** remaining detections
  - Higher level of abstraction

$$s_t^{ij} = p_{t-1}^i + p_t^j + \lambda \cdot c_t^{ij}$$

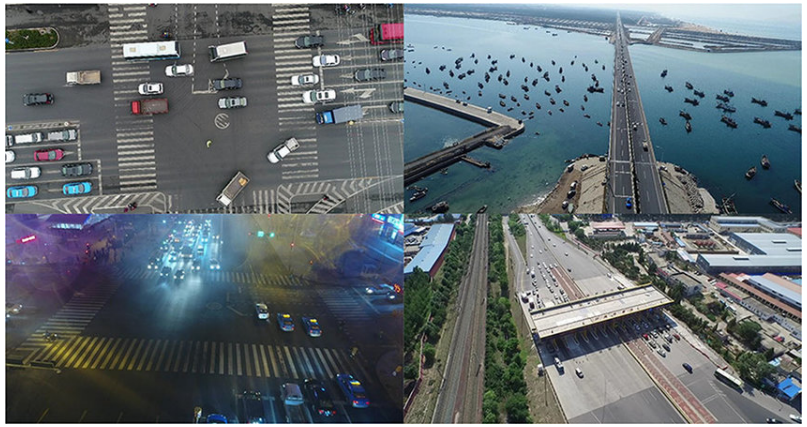
- *Tubelets* with dummy nodes are **discarded**



Experiments

Datasets and state-of-the-art

- State-of-the-art approaches:
  - **FPN-based**:
    - FPN-t and Cascade-FPN-t
  - **Spatio-temporal**:
    - FGFA, RDN and MEGA
- Databases:
  - **UAVDT**
    - More than 76,000 extremely small objects
  - **VisDrone2019-VID**
    - More than 27,000 extremely small objects
  - **USC-GRAD-STDdb**[1]
    - More than 56,000 extremely small objects



[1] B. Bosquet, M. Mucientes, and V. M. Brea. STDnet: *Exploiting high resolution feature maps for small object detection*. Engineering Applications of Artificial Intelligence, Vol. 91, No. 103615, 2020.

Experiments

Ablation study

- **Temporal information: 2.5%** AP@[.5:.95]
  - Baseline linking: **1.2%** AP@[.5:.95
  - Correlation-based linking + confidence variability: **0.5%** AP@[.5:.95]
  - Tubelet suppression procedure: **0.8%** AP@[.5:.95]

- Correlation module importance:
  - Improve the IoU-based baseline **without spatial information**
  - **Allows** to build the **tubelet suppression** algorithm
    - Higher level of abstraction

Baseline linking	Confidence variability	Correlation linking	Tubelet suppression	$AP_{xs}^{@[.5,.95]}$	$AP_{xs}^{@.5}$
—				18.9	59.1
✓				20.1	61.4
✓	✓			20.3	61.8
		✓		20.4	61.6
	✓	✓		20.6	62.0
		✓	✓	20.9	62.6
	✓	✓	✓	**21.4**	**63.4**

CiTIUS

Experiments

UAVDT

Method	$AP_{xs}^{@[.5,.95]}$	$AP_{xs}^{@.5}$
Faster R-CNN [27]	6.6	26.0
R-FCN [27]	9.2	32.5
RON [27]	3.7	19.7
SSD [27]	6.0	23.5
FGFA [154]	6.3	20.7
RDN [23]	9.3	27.9
MEGA [16]	9.2	26.6
FPN [78]	11.8	29.7
FPN-t	12.0	30.3
Cascade-FPN [13]	12.0	30.5
Cascade-FPN-t	12.3	31.2
STDnet [9]	12.5	35.1
STDnet++	12.6	35.4
STDnet-ST	13.1	36.0
STDnet-ST++	**13.3**	**36.4**

VisDrone2019- VID

Method	$AP_{xs}^{@[.5,.95]}$	$AP_{xs}^{@.5}$
FGFA [154]	3.8	16.8
RDN [23]	4.7	20.7
MEGA [16]	4.8	21.0
FPN [78]	6.2	19.9
Cascade-FPN [13]	6.1	20.2
FPN-t	6.3	20.2
Cascade-FPN-t	6.2	20.4
STDnet [9]	7.2	21.4
STDnet++	7.3	22.0
STDnet-ST	7.5	21.9
STDnet-ST++	**7.5**	**22.4**

USC-GRAD-STDdb

Method	$AP_{xs}^{@[.5,.95]}$	$AP_{xs}^{@.5}$
FGFA [154]	11.7	37.5
RDN [23]	15.5	48.6
MEGA [16]	17.4	53.1
FPN [78]	17.3	54.5
Cascade-FPN [13]	17.4	55.9
FPN-t	18.7	57.2
Cascade-FPN-t	19.1	58.9
STDnet [9]	18.3	57.8
STDnet++	18.9	59.1
STDnet-ST	20.1	62.1
STDnet-ST++	**21.4**	**63.4**

- STDnet-ST achieves state-of-the-art results:
  - Outperforms spatial **STDnet**
  - Outperforms **FPN-based** approaches
  - Outperforms **spatio-temporal** approaches

CiTIUS

Correlation-based ConvNet for Small Object Detection in Videos

Brais Bosquet and **Manuel Mucientes** and **Víctor M. Brea**

25th International Conference on Pattern Recognition (ICPR)

10 - 15 January 2021

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CAMPUS VIDA CAMPUS DE EXCELENCIA INTERNACIONAL

CiTIUS Centro Singular de Investigación en Tecnoloxías Intelixentes