



A Coarse-to-fine Deep CNN for Frame Duplication Detection and Localization in Videos



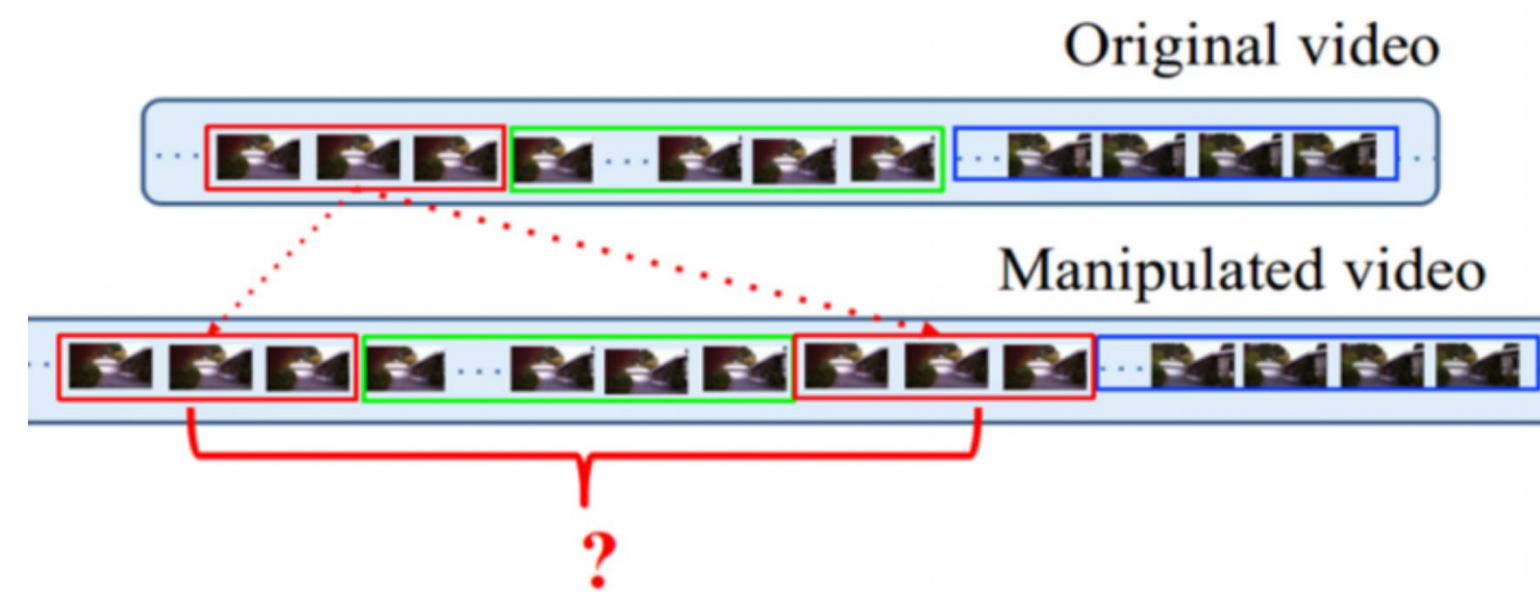
LONG BEACH
CALIFORNIA
June 16-20, 2019

Chengjiang Long Arslan Basharat Anthony Hoogs

Kitware Inc., Clifton Park, NY 12065

{chengjiang.long, arslan.basharat, anthony.hoogs}@kitware.com

Introduction



- Frame duplication refers to a video manipulation where a copy of a sequence of frames inserted into the same video either replacing previous frames or as additional frames.
- This problem is difficult: we don't know whether any sequences are duplicated; if there is a duplication, we don't know where it starts or how long it lasts or how many copies there might be.
- Robust detection and localization of duplicated parts of a video can be a very useful forensic tool for those tasked with authenticating large volumes of video content.

Contributions

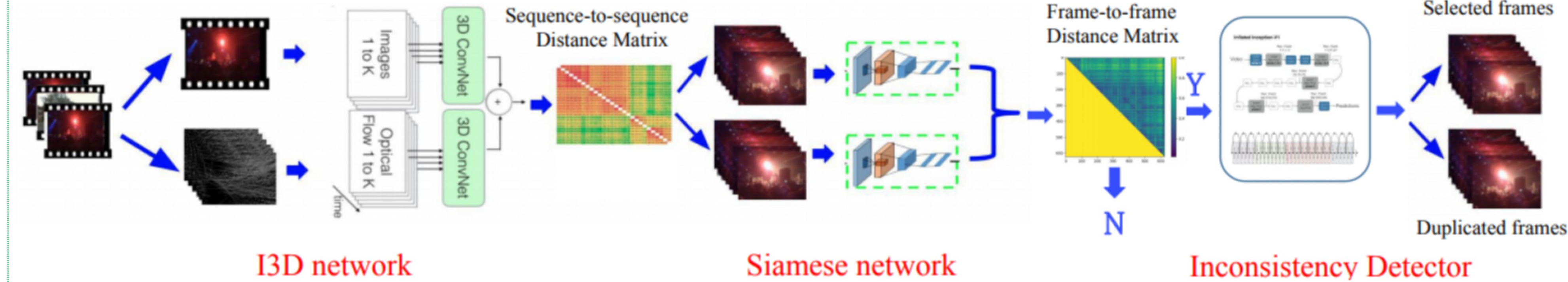
- We propose a novel coarse-to-fine deep learning framework for frame duplication detection and localization in forged videos.
- We have designed an inconsistency detector based to distinguish duplicated frame-range from the selected original frame-range.
- We propose a heuristic formulation for video-level detection score, which leads to significant improvement in detection benchmark performance

Datasets

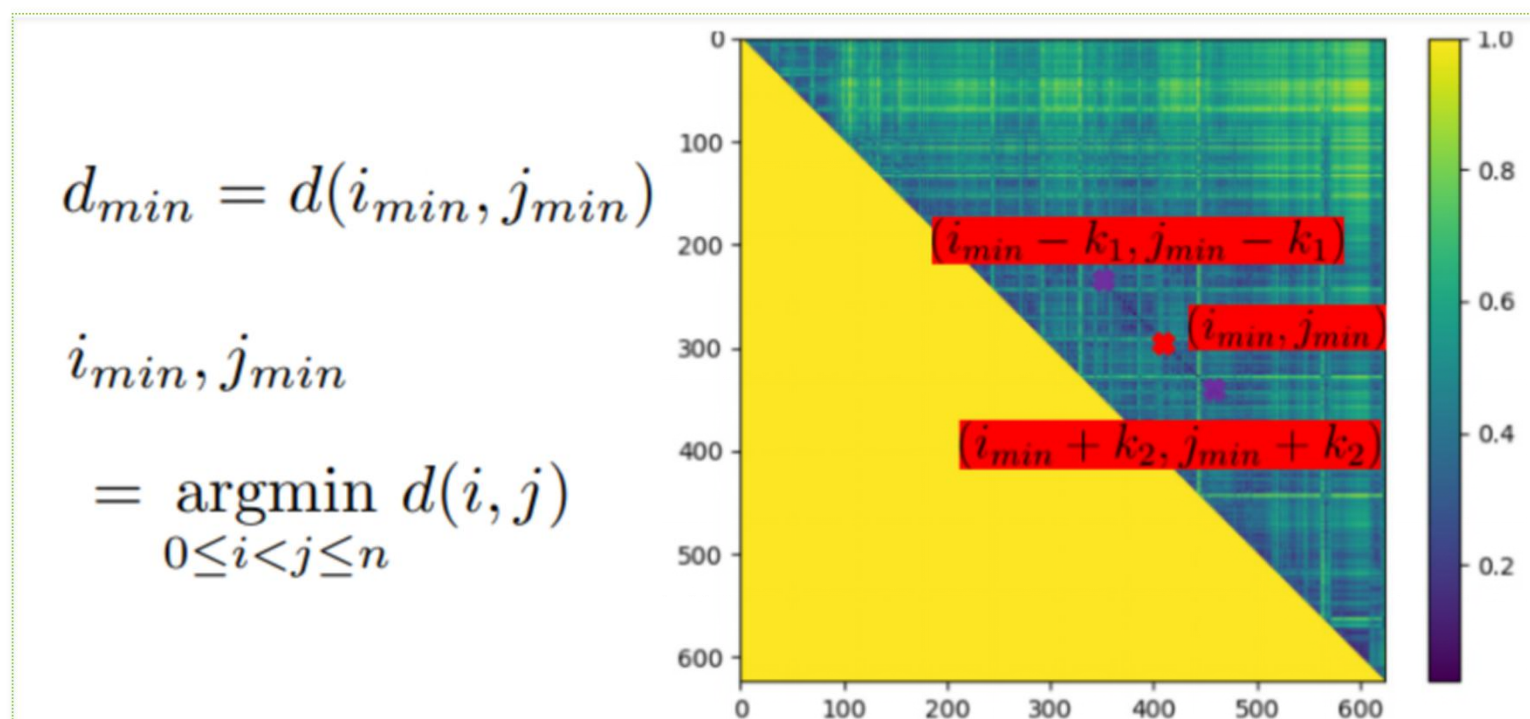
A self-collected video dataset: frame duplication manipulation on the 12 raw static videos from VIRAT dataset, and 17 dynamic iPhone 4 videos.

MFC18 dataset: 231 videos in the MFC18-Dev dataset and 1036 videos in the MFC18-Eval dataset. 2s-3mins duration for each video. most of videos are 29-30 fps.

Proposed Approach



Video-level confidence score



$$d_{min} = d(i_{min}, j_{min})$$

$$i_{min}, j_{min} = \underset{0 \leq i < j \leq n}{\operatorname{argmin}} d(i, j)$$

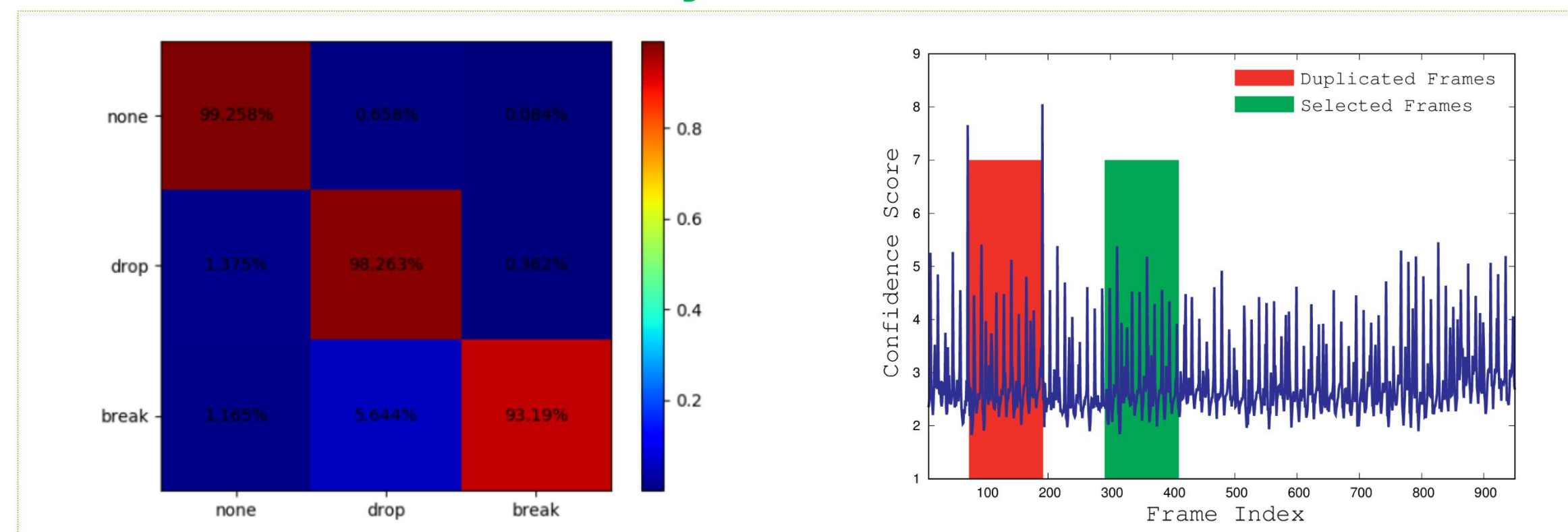
$$k_1 = \underset{k: k \leq i_{min}}{\operatorname{argmax}} |d(i_{min} - k, j_{min} - k) - d_{min}| \leq \epsilon$$

$$k_2 = \underset{k: k \leq n - j_{min}}{\operatorname{argmax}} |d(i_{min} + k, j_{min} + k) - d_{min}| \leq \epsilon$$

$$l = k_1 + k_2 + 1$$

$$F_{video} = -\frac{d_{min}}{l \times (j_{min} - i_{min})}$$

Inconsistency detector for localization

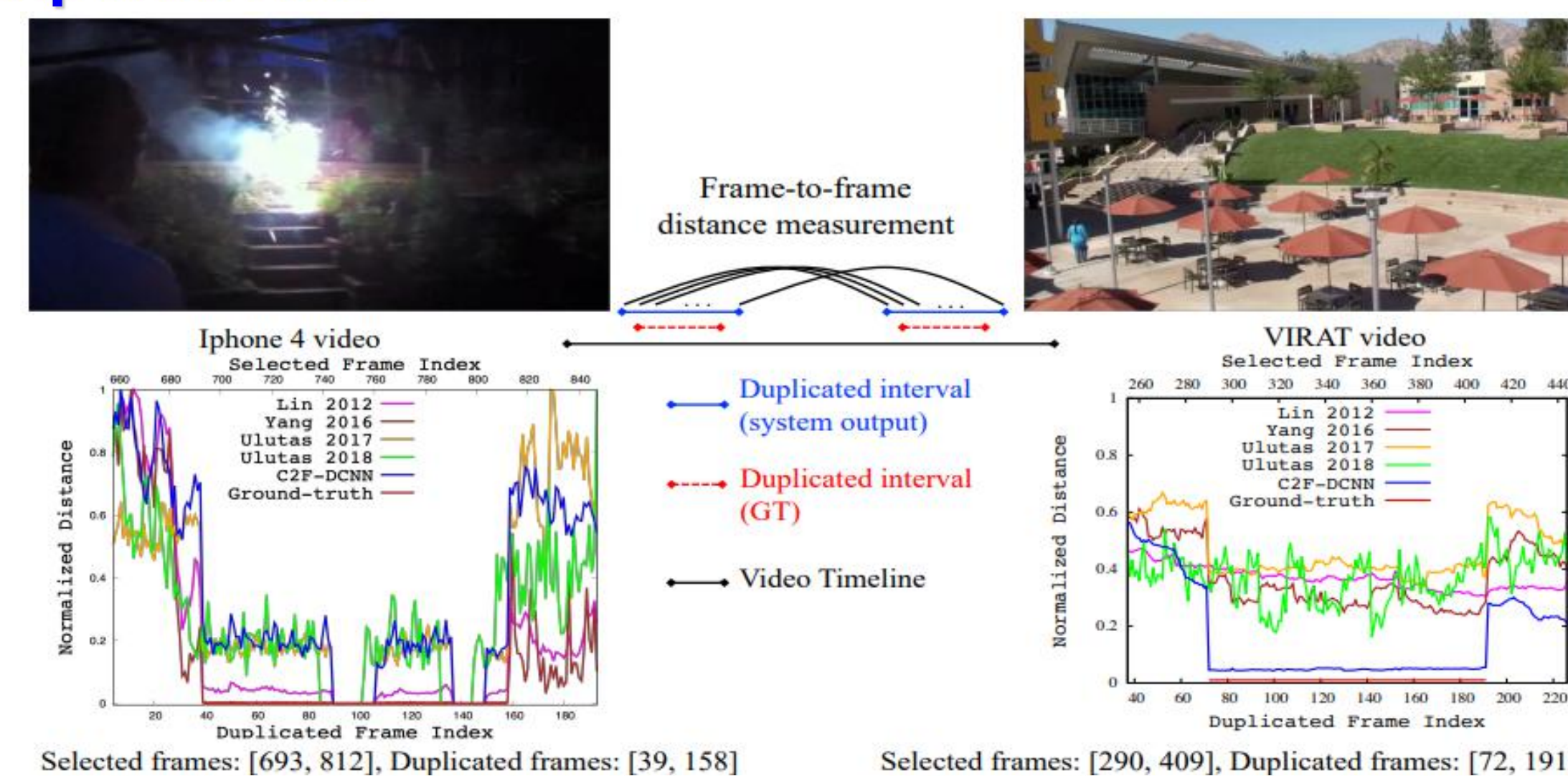


- I3D network with 3 classes: none, frame drop, shot break.
 - Training data: we generate manipulated videos from non-target videos from MFC18-Dev and TRECVID 2007.
- $$S(i) = S_{I3D}^{drop}(i) + S_{I3D}^{break}(i) - \lambda S_{I3D}^{none}(i)$$
- $$s_1 = \sum_{k=-wind}^{wind} S(i-1+k) + S(i+l+k)$$
- $$s_2 = \sum_{k=-wind}^{wind} S(j-1+k) + S(j+l+k)$$

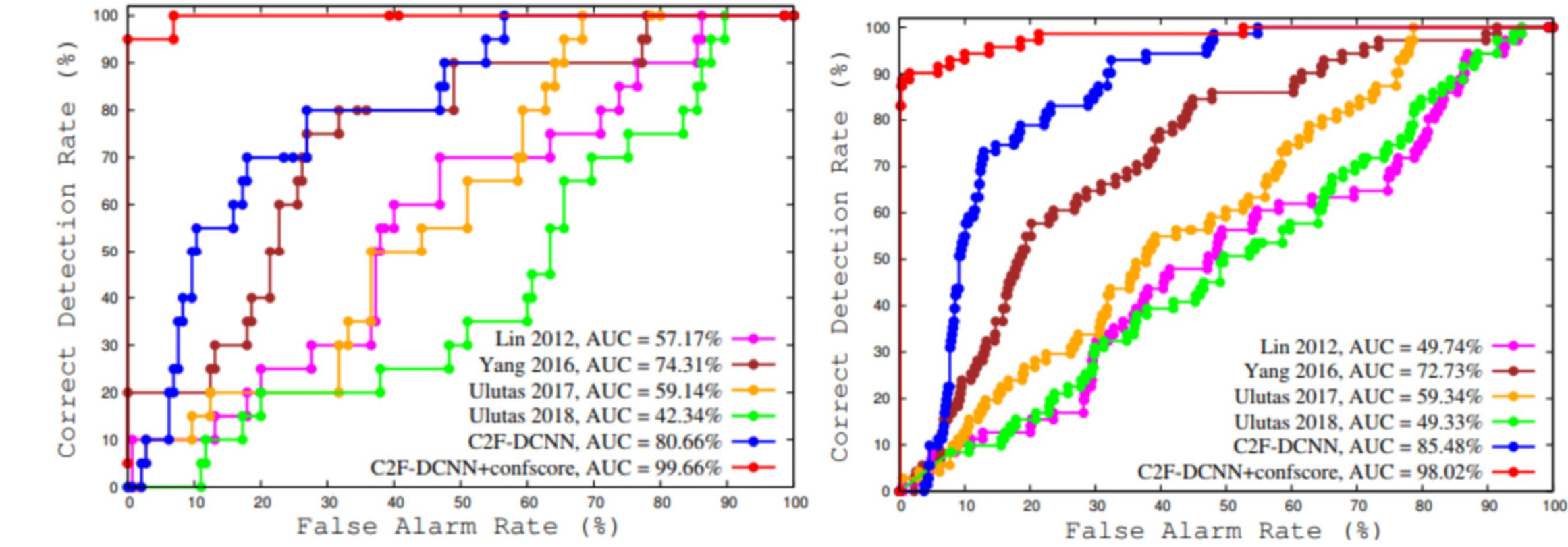
Frame-level performance on our self-collected dataset

Method	iPhone 4 videos	VIRAT videos
Lin 2012 [16]	80.81	80.75
Yang 2016 [37]	73.79	82.13
Ulutas 2017 [26]	70.46	81.32
Ulutas 2018 [27]	73.25	69.10
C2F-DCNN	81.46	84.05

Experiments



Video-level performance on the MFC18 dataset

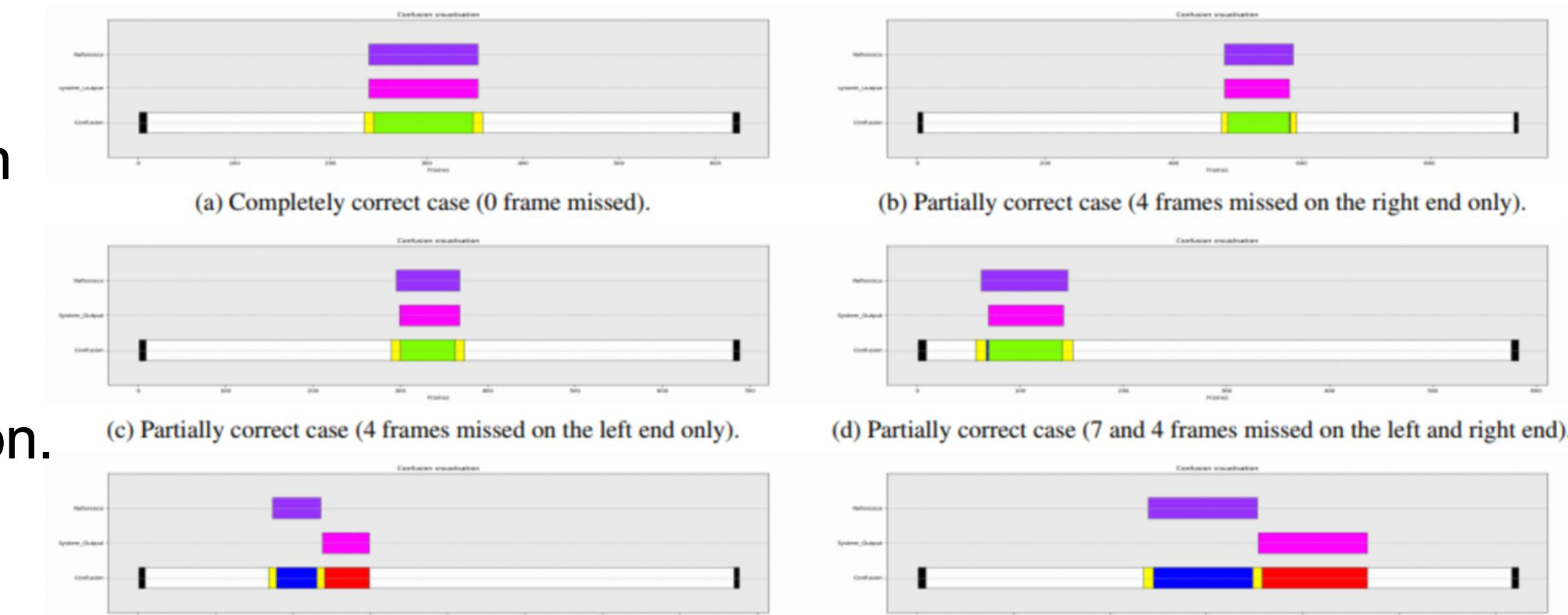


ROC curves for frame duplication detection on MFC18-Dev and MFC-Eval datasets.

The MCC for video temporal localization on the MFC dataset.

Method	MFC18-Dev	MFC18-Eval
Lin 2012 [16]	0.2277	0.1681
Yang 2016 [37]	0.1449	0.1548
Ulutas 2017 [26]	0.2810	0.3147
Ulutas 2018 [27]	0.0115	0.0391
C2F-DCNN w/ ResNet	0.4618	0.3234
C2F-DCNN w/ C3D	0.6028	0.3488
C2F-DCNN w/ I3D	0.6612	0.3606

Confusion bars in video temporal localization.



Acknowledgement

This work was supported by DARPA under Contract No. FA8750-16-C-0166. Any findings and conclusions or recommendations expressed in this material are solely the responsibility of the authors and does not necessarily represent the official views of DARPA or the U.S. Government.

Key References

[Lin 2012] G. Lin et al. Detection of frame duplication forgery in videos based on spatial and temporal analysis. PAMI, 2012.
[Yang 2016] J. Yang, et al. Using similarity analysis to detect frame duplication forgery in videos. MTA, 2016.

[Ulutas 2017] G. Ulutas et al. Frame duplication/mirroring detection method with binary features. IET-IP, 2017.
[Ulutas 2018] G. Ulutas et al. Frame duplication detection based on bow model. Multimedia Systems, 2018.