SafeCare: Faster Video Analytic to Protect Workers

Siao Tang, Fulin Wang, Peiwen Li

Abstract

The issue of worker helmet detection under site video is of great importance to protect the lives of workers. To accelerate this process, we propose Safecare, a video frame filtering method that identifies frames that should be filtered by using difference features to predict quality differences between video frames that can't be compute quickly. Safecare achieves an average IoU improvement of 0.01 compared to randomly filtering an equal number of frames.

Motivation

In construction sites, workers are under constant threat of safety hazards, and hiring dedicated safety officers to monitor site safety can bring high labor costs. Therefore, filter some video frames and reuse the analytic results on the filtered frames **to accelerate** video understanding DNNs could help to: (1) quickly and accurately perform tasks such as helmet detection thus **safeguarding the lives and health of workers**, (2) save the insufficient uplink network bandwidth for transmitting the videos from the capture device to the backend server.

Datasets

We utilize a YoloV5 [1] model pretrained with **SHWD video dataset,** which is consists of 7581 images and 9044 human safety helmet [2].

For pipeline implementation and evaluation, we choose the construction worker dataset provided by LFD class.

Fig. 1. Dataset Example

Method

The main idea of SafeCare is to **utilize the visual quality between the adjacent video frames to represent the variation** of them, and this variation could help to determine the dfference of analytic results. When the variation is small enough (SSIM>0.87 in our method), the object detection results could be reused.

However, SSIM could only be calculated with a speed of 2fps on 720p video frames, and we should try to forecast SSIM value with a higher speed.

In Safecare, four cheap features [3], including pixel-level difference, edge-level difference, area-level difference and grayscale histogram difference are taken into consideration. **A binary random forest classifer trained on the historical video frames (first half of the video frames)** could help to predict whether the SSIM is greater than 0.87 **with an accuracy over 0.9**. Putting all things together, **the execution speed could still be 30fps.** For another half of the frames, our classification model's filter performance is evaluated.

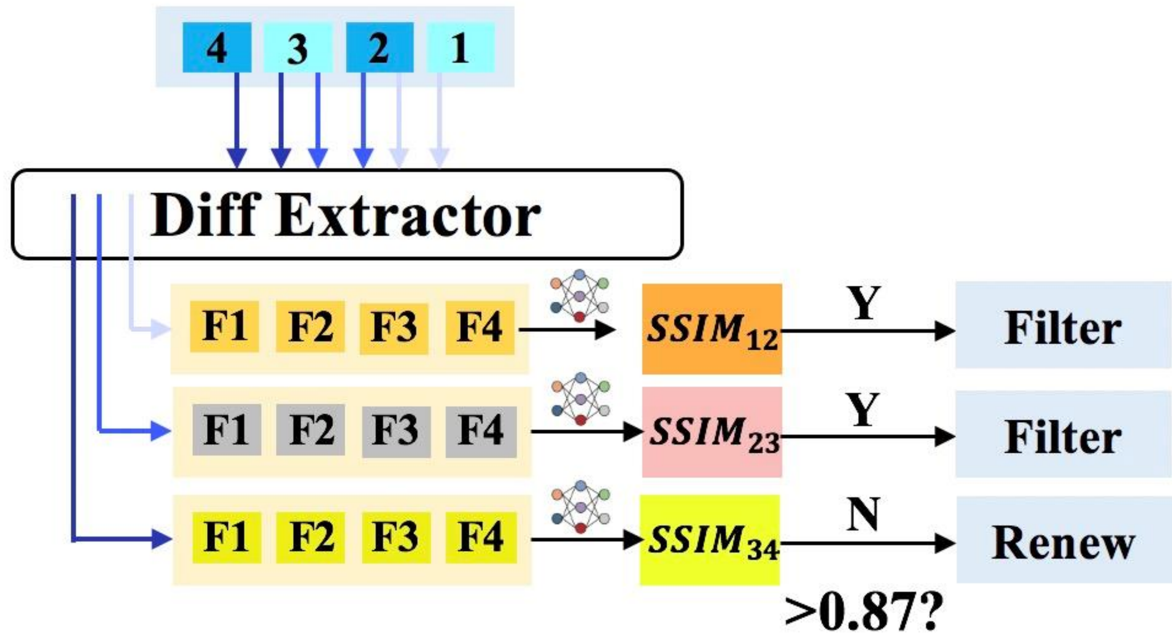


Fig. 2. SafeCare Frame Filter Pipeline

References

[1] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[2] https://www.v7labs.com/open-datasets/shwd-dataset

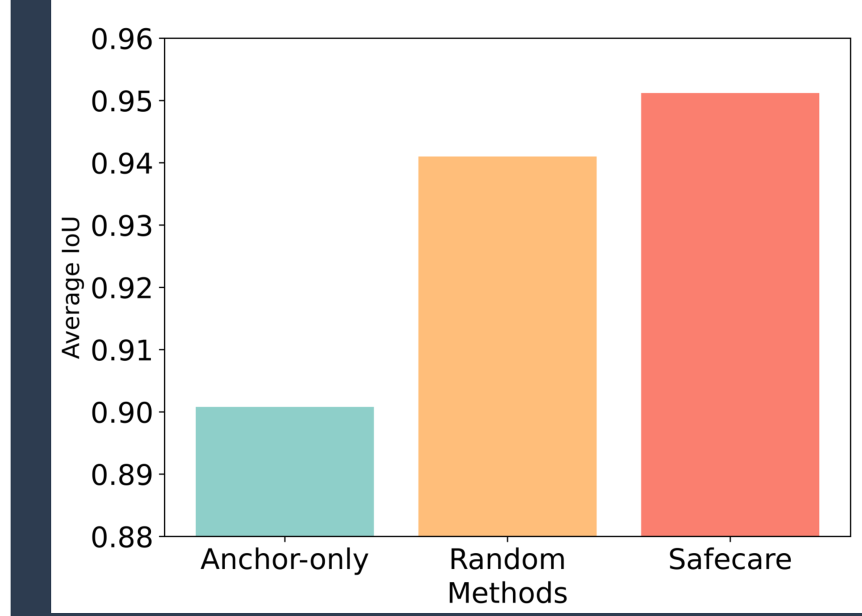
[3] Li Y, Padmanabhan A, Zhao P, et al. Reducto: On-camera filtering for resource-efficient real-time video analytics[C]

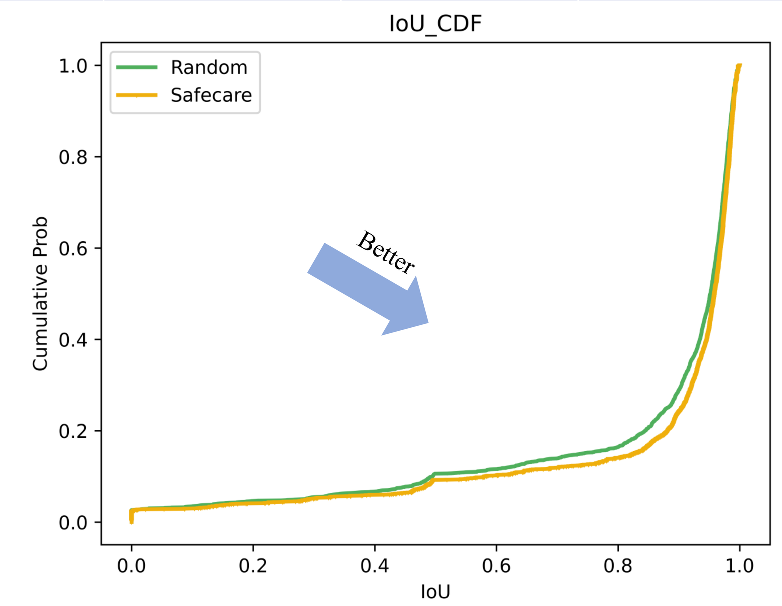
Result

We choose **six long videos from six type of test videos**, train our visual quality difference prediction model on the first half of the video frames, and get IoU results below.

The IoU results is compared with YoloV5 results on all the video frames. It should be noter that when the number of object detected is not the same, the IoU of the missing object will be denoted as zero,

Video/ Method	Glass-1	Glass-2	Nail-6	Nail-11	Shelf-1	Cord-2
A-only	0.8367	0.8155	0.9677	0.9642	0.8734	0.9475
Random	0.9171	0.9027	0.9940	0.9659	0.8951	0.9715
Ours	0.9240	0.9144	0.9970	0.9733	0.9223	0.9766
Fillter Rate	40%	45%	18%	64%	60.34%	47.8%





Conclusions & Future Work

Safecare achieves an average IoU of 0.9512 compared to randomly filtering an equal number of frames with an average filter rate of 48%.

Future work:

Only take the ROI's (region of interest) SSIM into consider.

A binary classification model to determine whether there is an object of interest could be utilized to filiter more frames.