Poster Abstract: Link-adaptive and Real-time Object Detection in Dynamic Edge Networks

Rong Cong
congrong@mobinets.org
University of Electronic Science and Technology of China
Chengdu, Sichuan, China

Zhiwei Zhao*
zhiwei@mobinets.org
University of Electronic Science and Technology of China
Chengdu, Sichuan, China

Linyuanqi Zhang
yuanqi@mobinets.org
University of Electronic Science and Technology of China
Chengdu, Sichuan, China

Cong Zha
chac16@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

ABSTRACT

Detection&tracking framework enables real-time object detection services on resource-limited mobile devices in edge networks, where mobile devices only offload few key frames to edge servers for detection and track the other frames locally. Following this framework, the offloading decisions become more important, as fewer frames can be offloaded and their results can directly affect tracking accuracy of the subsequent frames. To make wise frame offloading decisions, it is crucial to leverage the relationship between the detected frame and the tracked frame. However, the existing studies solely use the content-level information to reflect the above relationship, i.e., they tend to offload the frames with the significant pixel differences from the adjacent frames. Link impact on such relationship is completely ignored, thus leading to significant accuracy degradation. In this paper, we propose Kite, a link-adaptive and real-time object detection in dynamic edge networks, which integrates both the link-level and content-level impact into frame offloading. Kite exploits a novel performance metric, "frame-anchor" distance, to indicate the impact of dynamic wireless links. With this metric, we can incorporate both link and content information into the offloading process. The real-world experiment results show that Kite can improve the detection accuracy in dynamic edge networks.

KEYWORDS

Edge computing, Task offloading, Dynamic wireless networks

ACM Reference Format:

Rong Cong, Zhiwei Zhao*, Linyuanqi Zhang, and Cong Zha. 2023. Poster Abstract: Link-adaptive and Real-time Object Detection in Dynamic Edge Networks . In *The 21st ACM Conference on Embedded Networked Sensor Systems (SenSys '23), November 12–17, 2023, Istanbul, Turkiye.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3625687.3628396

SenSys '23, November 12–17, 2023, Istanbul, Turkiye

ACM ISBN 979-8-4007-0414-7/23/11…$15.00
https://doi.org/10.1145/3625687.3628396

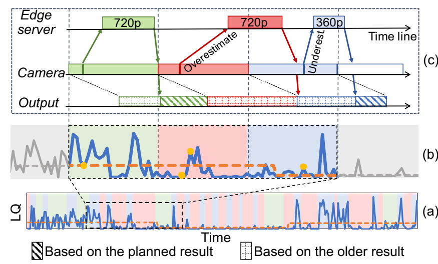


Figure 1: Detection&tracking system in a real-world network.

1 INTRODUCTION

To enable its considerable computing demand and strict latency requirement, edge computing becomes a promising solution for Real-time object detection. However, due to the large transmission overhead of video streams, the overall latency of whole-video offloading (100-150ms per frame) is still unacceptable for real-time object detection (33.3ms per frame for 30FPS videos). To reduce the overall latency, recent work studied the detection&tracking framework [1, 2], where only few key frames need to be offloaded and other frames are tracked locally to quickly estimate the object positions based on the detection results of the previous frame. In this framework, frame offloading becomes a more important problem, as fewer frames can be offloaded and their detection results can directly affect the tracking accuracy of the subsequent frames.

However, the existing studies solely consider the content-level impact on frame offloading performance [1, 2], ignoring the link-level impact. They offload frames based on pixel difference, with the assumption of stable or averaged wireless links during offloading. Such an ideal assumption does not hold true in real-world edge networks, whose link quality is highly dynamic. *According to our measurement experiments (detailed in Section ??),* ***in nearly 80% cases in real-world edge networks, the existing schemes cannot work as expected due to the overlook of link impacts.*** *They either incur the unexpected latency (link overestimation) or offload frames with sub-optimal resolutions (link underestimation), and further suffer significantly accuracy degradation.*

In this research, we propose Kite, a link-adaptive and real-time object detection in dynamic edge networks, which integrates both link dynamics and content-level impacts into frame offloading. We first evaluate the link impact on object detection performance via measurement studies, then devise a novel metric - "frame-anchor" distance, to reflect the link impact. With the proposed metric, we can bridge the link impact with the content impact on system performance. We implement a prototype system under WiFi and conduct the real-world experiments. The results show that Kite can improve detection accuracy for highly dynamic edge networks.

2 SYSTEM DESIGN

Link Impacts. Figure 1 shows the detection&tracking system under a WiFi network. Here, the front-end camera needs to select anchor frames in each window for timely and accurate object detection and track other frames using a lightweight tracker.

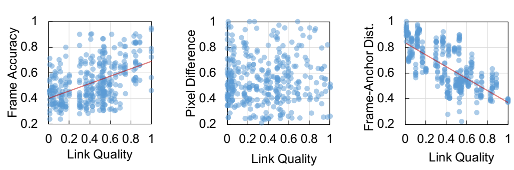
When the instant LQ is similar to the expected long-term LQ, the best frame can be selected (green). However, when the actual LQ is lower than expected, the frame offloading will experience larger delay than expected (red, 45% cases), which leads to 1) the detection results will be outdated and 2) the tracking accuracy of subsequent frames will be further reduced due to the stale anchor frame. When the actual LQ is higher than expected (blue, 33% cases), the offloading opportunity with higher resolutions will be missed.

Link-adaptive performance metric. To identify the wireless dynamics impacts on object detection, we devise a performance metric of "frame-anchor" distance, θ, which is calculated as the number of frames between each frame and its previous anchor frame. With this metric, devices can perform lightweight and link-adaptive frame offloading. Frame f_i's metric θ_i is calculated as $\theta_i = n_i - n_i^a$, where n_i is f_i's frame sequence and n_i^a denotes the sequence of f_i's anchor frame f_i^a, i.e., the previous frame of which the results are returned before f_i's output time to_i.

$$f_i^a = \arg\min_{f_k \in \mathcal{A}_i} (to_i - \hat{tr}_k) \tag{1}$$

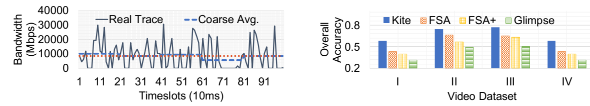
where $\hat{tr}_x$ is the estimated return time of f_x's results, to_i is the output time for frame f_i, and $\mathcal{A}_i$ is the set of available anchor frames for f_i. For any frame $f_k \in \mathcal{A}_i$, its results return earlier than f_i's output time ($\hat{tr}_k < to_i$).

We conduct comparative experiments to study whether the two metrics can reflect the wireless dynamics impacts on object detection. Figure 2(a) shows the average detection accuracy with different normalized link quality values, where the red lines are the fitting lines of the corresponding experimental results. We see a clear trend that the accuracy increases as the link quality increases. The reason is that the tracking results of more frames are based on outdated anchor frame results, and more tracking errors are accumulated to those frames. Figure 2(b) shows the metric values of pixel difference, which are randomly distributed and can hardly reflect the change of detection accuracy. Based on the same selected anchor frames, we then calculate θ to check whether θ is indicative. The results displayed in Figure 2(c) show that θ exhibits a clear correlation with both detection accuracy and link quality.



(a) Frame accuracy. (b) Pixel difference. (c) Frame-anchor dist..

Figure 2: Correlation between link quality and two metrics.



(a) Network trace. (b) Accuracy comparison.

Figure 3: Performance comparison for different videos.

3 EVALUATION

Implementation and setup. We implement our system `Kite` and the counterpart work [1, 2] on NVIDIA Jetson TX2 and a 2080-equipped server under WiFi. We set that each frame result must be output within 30ms. Similar to the settings in [2], we use Jetson TX2 as the front-end device that captures videos, makes frame offloading decisions, and runs a Lucas-Kanade tracker based on OpenCV. The edge server utilizes YOLOv5x for object detection. We conduct the experiments with a self-captured soccer video (Dataset I) and three public datasets [1] (Dataset II III and IV). All datasets contain different resolutions from 360p to 2K/4K.

Baselines. 1) `Glimpse` [1] is an offloading scheme without considering network variations. 2) `FSA` [2] is our state-of-the-art work, which uses the pixel difference metric and relies on the averaged link quality for frame offloading. 3) `FSA+` is a variant of `FSA`, which additionally configure the frame resolutions.

Evaluation results. Figure 3 depicts the detection accuracy under the WiFi-based edge network. We see that `Kite` consistently outperforms other schemes on all datasets, as only `Kite` considers the utilization of the fine-grained link variations. Besides, the frame-anchor distance is more indicative and thus the front-end devices are more sensitive to the link variations. We also observe that `FSA+` does not achieve better accuracy than `FSA` as expected. Because both of them use second-level link quality, while upload/download takes tens of milliseconds, in many decision windows, `FSA+` offloads high resolution frames over poor uplinks, resulting in expected delay and more accumulative tracking errors for the subsequent frames.

4 CONCLUSION AND FUTURE WORK

We propose a link-adaptive and real-time object detection system. It explicitly incorporates the wireless link dynamics into frame offloading with the "frame-anchor" distance metric. In the future, we will conduct more experiments to evaluate the system performance under different scenarios. And more sophisticated models and efficient offloading schemes will also be studied.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 61972075, 92067206 and 61972074) and the Natural Science Foundation of Sichuan Province (2022NSFSC0885).

REFERENCES

[1] Tiffany Yu-Han Chen, Lenin Ravindranath, Shuo Deng, Paramvir Bahl, and Hari Balakrishnan. 2015. Glimpse: Continuous, real-time object recognition on mobile devices. In *Proc. of ACM SenSys 2015*. 155–168.

[2] Mengxi Hanyao, Yibo Jin, Zhuzhong Qian, Sheng Zhang, and Sanglu Lu. 2021. Edge-assisted online on-device object detection for real-time video analytics. In *Proc. of IEEE INFOCOM 2021*. IEEE, 1–10.

[1] https://www.videezy.com