

Object detection in 4K and 8K video using GPUs

7 November 2018, by Ingrid Fadelli



Example of crowded video frame annotated using the new method. Credit: R?ži?ka and Franchetti.

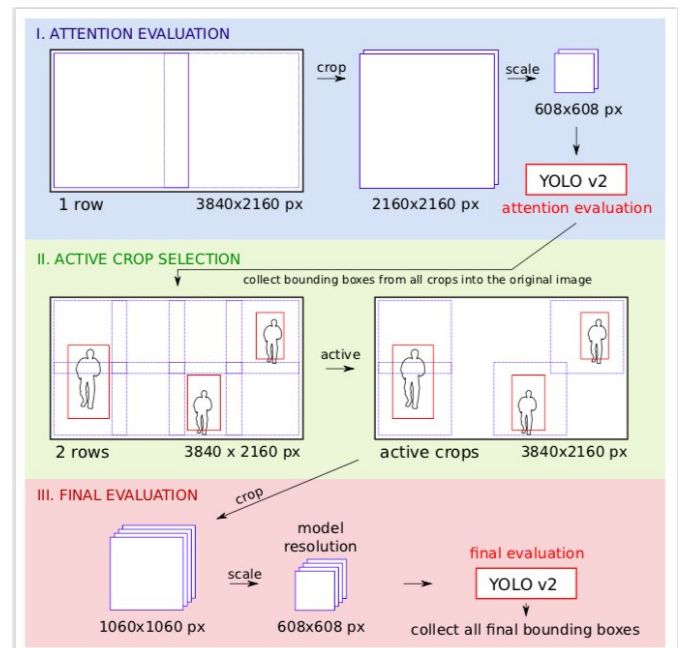
Researchers at Carnegie Mellon University have recently developed a new model that enables fast and accurate object detection in high-resolution 4K and 8K video footage using GPUs. Their attention pipeline method carries out a two-stage evaluation of every image or video frame under rough and refined resolution, limiting the total number of evaluations necessary.

In recent years, machine learning has attained remarkable results in computer vision tasks, including [object detection](#). However, most [object](#) recognition models typically perform best on images with a relatively low resolution. As the resolution of recording devices is rapidly improving, there is a rising need for tools that can process [high-resolution](#) data.

"We were interested in finding and overcoming the limitations of current approaches," Vít R?ži?ka, one of the researchers who carried out the study told TechXplore. "While plenty of data sources record in high resolution, current state-of-the-art object detection models, such as YOLO, Faster RCNN, SSD, etc., work with images that have a relatively low resolution of approximately 608 x 608 px. Our

main objective was to scale the object detection task to 4K-8K videos (up to 7680 x 4320 px) while maintaining high processing speed. We also wanted to understand if and by how much we can benefit from high resolution compared to using low-resolution images, in terms of accuracy of the models."

The attention pipeline proposed by R?ži?ka and his colleague Franz Franchetti divides the task of object detection into two stages. In both these stages, the researchers subdivided the original image by overlaying it with a regular grid and then applied the model YOLO v2 for fast object detection.



Resolution handling on the example of 4K video frame processing. During the attention step the image is processed under rough resolution, allowing the researchers to decide which regions of the image should be active in final finer evaluation. Credit: R?ži?ka and Franchetti.

"We create many small rectangular crops, which can be processed by YOLO v2 on several server workers, in a parallel manner," R?ži?ka explained. "The first stage looks at the image downscaled into lower resolution and performs a fast object detection to get rough bounding boxes. The second stage uses these bounding boxes as an attention map to decide where we need to check the image under high-resolution. Therefore, when some areas of the image don't contain any object of interest, we can save on processing them under high resolution."

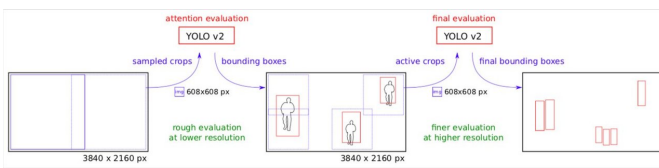
number of issues. For instance, it can sometimes result in detected objects being cut in half, which requires a post-processing step on the detected bounding boxes. R?ži?ka and Franchetti are currently exploring ways of addressing and circumventing these problems to improve their [model](#) further.

More information: Fast and accurate object detection in high resolution 4K and 8K video using GPUs. arXiv:1810.10551v1 [cs.CV]. arxiv.org/pdf/1810.10551.pdf

[www.researchgate.net/publication...
d_8K_video_using_GPU](https://www.researchgate.net/publication/351111111)

www.youtube.com/watch?v=07wCxSltnAk

[www.researchgate.net/publication...
tation_for_HPEC_2018](https://www.researchgate.net/publication/351111111) .



The attention pipeline. Stepwise breakdown of the original image under different effective resolution. Credit: R?ži?ka and Franchetti.

© 2018 Science X Network

The researchers implemented their model into code, distributing its work across GPUs. They were able to maintain high accuracy while reaching an average performance of three to six fps on 4K videos and two fps on 8K videos. Their method yielded significant benefits, with the measured average precision on the tested dataset increasing from 33.6 AP₅₀ to 74.3 AP₅₀ when processing images in high resolution compared to down-scaling images to low resolution, which is how YOLO v2 generally works.

"Our method reduced the time necessary to process high-resolution images by approximately 20 percent, compared to processing every part of the original image under high [resolution](#)," R?ži?ka said. "The practical implication of this is that near real-time 4K video processing is feasible. Our method also requires a lower number of server workers to complete this task."

Despite the very promising results attained by this new object detection method, the use of a regular grid overlaying the original image can give rise to a

APA citation: Object detection in 4K and 8K video using GPUs (2018, November 7) retrieved 26 January 2022 from <https://techxplore.com/news/2018-11-4k-8k-video-gpus.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.