

KONFERENSIYALAR.COM

ANJUMANLAR PLATFORMASI

**XIX RESPUBLIKA ILMIY-
AMALIY KONFERENSIYASI**

**YANGI DAVR ILM-
FANI: INSON UCHUN
INNOVATSION G'OYA
VA YECHIMLAR**

OKTYABR, 2026

ISSN 3093-8791

ELEKTRON NASHR:

<https://konferensiyalar.com>



Yangi davr ilm-fani: inson uchun innovatsion g'oya va yechimlar.
XIX Respublika ilmiy-amaliy konferensiyasi materiallari to'plami.
2-jild, 19-son (13-oktyabr, 2026-yil).– 191 bet.

Mazkur nashr ommaviy axborot vositasi sifatida 2025-yil, 8-iyulda
C-5669862 son bilan rasman davlat ro'yaxatidan o'tkazilgan.

Elektron nashr: <https://konferensiyalar.com>

ISSN: 3093-8791 (onlayn)

Konferensiya tashkilotchisi: "Scienceproblems Team" MChJ

Konferensiya o'tkazilgan sana: 2026-yil, 9-oktyabr

Mas'ul muharrir:

Isanova Feruza Tulqinovna

Annotatsiya

Mazkur to'plamda "Yangi davr ilm-fani: inson uchun innovatsion g'oya va yechimlar" mavzusidagi XIX Respublika ilmiy-amaliy konferensiyasi materiallari jamlangan. Nashrda respublikaning turli oliy ta'lim muassasalari, ilmiy markazlari va amaliyotchi mutaxassislari tomonidan tayyorlangan maqolalar o'rin olgan bo'lib, ular ijtimoiy-gumanitar, tabiiy, texnik va yuridik fanlarning dolzarb muammolari va ularning innovatsion yechimlariga bag'ishlangan.

Ushbu nashr ilmiy izlanuvchilar, oliy ta'lim o'qituvchilari, doktorantlar va soha mutaxassislari uchun foydali qo'llanma bo'lib xizmat qiladi.

Kalit so'zlar: ilmiy-amaliy konferensiya, innovatsion yondashuv, zamonaviy fan, fanlararo integratsiya, ilmiy-tadqiqot, nazariya va amaliyot, ilmiy hamkorlik.

Barcha huquqlar himoyalangan.

© Scienceproblems team, 2026-yil

© Mualliflar jamoasi, 2026-yil

REAL VAQT REJIMIDAGI CHEKKA BIOMETRIK TIZIMLAR UCHUN YUZNI ANIQLASH MODELLARINING TASVIR O'LCHAMIGA BOG'LIQ KECHIKISHI VA PARETO SAMARADORLIGI

Ortiqova Feruza Sodiq qizi

Tayanch doktorant,

Muhammad al-Xorazmiy nomidagi

Toshkent axborot texnologiyalari universiteti

Toshkent, O'zbekiston

Annotatsiya. Yuzni aniqlashning hisoblash xarajati video oqim o'lchami ortishi bilan o'sadi va ko'pincha biometrik tizimning chekka qurilmada real vaqt rejimida ishlay olish-olmasligini belgilaydi. Ushbu tadqiqotda YuNet, masshtablashsiz va masshtablash bilan ishlovchi Dlib HOG hamda RetinaFace modellarining 480p dan 2160p gacha bo'lgan to'rtta kirish o'lchamida o'lchangan bir kadrda to'g'ri keluvchi kechikishi tahlil qilingan. Kechikishning kadr maydoniga ko'ra o'sishi, uning qat'iy va piksellar soniga proporsional qismlari, Dlib masshtablashining narxi hamda bir soniyalik hisoblashga to'g'ri keladigan aniqlash sifati o'rganilgan, shuningdek, 1–30 kadr/s tezlik byudjetlari uchun ruxsat etilgan o'lchamlar aniqlangan. Natijalarga ko'ra, 1080p va 2160p oralig'ida kechikish kadr maydoniga deyarli proporsional ravishda o'sadi; YuNet kechikishining o'lchamga bog'liqligi neyron tarmoq hisobidan emas, balki kadr o'lchamini o'zgartirish jarayonidan kelib chiqadi; Dlib'dagi har bir masshtablash bosqichi kechikishni taxminan 3,5 barobar oshiradi; markaziy protsessordagi YuNet va grafik protsessordagi RetinaFace Pareto frontini tashkil etadi. Tadqiqot yakunida olingan natijalar asosida chekka qurilmalarda biometrik tizimlarni joriy etish bo'yicha amaliy qoidalar taklif etilgan.

Kalit so'zlar: xulosa chiqarish kechikishi; kirish tasviri o'lchami; chekka hisoblash; real vaqtli biometriya; Pareto samaradorligi; kadr tezligi byudjeti; YuNet; Dlib.

RESOLUTION-DEPENDENT LATENCY AND PARETO EFFICIENCY OF FACE DETECTORS FOR REAL-TIME EDGE BIOMETRIC SYSTEMS

Ortiqova Feruza Sodiq qizi

PhD Student,

Tashkent University of Information Technologies

named after Muhammad al-Khwarizmi

Tashkent, Uzbekistan

Annotation. The computational cost of face detection grows with the resolution of the video stream and often determines whether a biometric system can run in real time on edge hardware. This study analyses the per-frame latency of YuNet, Dlib HOG with and without upsampling, and RetinaFace at four input resolutions from 480p to 2160p. It examines how latency scales with frame area, separates fixed from pixel-proportional cost, quantifies the price of Dlib upsampling, evaluates detection quality per second of computation and derives admissible resolutions for frame-rate budgets of 1–30 fps. Between 1080p and 2160p latency grows almost in proportion to frame area; the resolution dependence of YuNet stems from frame resizing rather than inference; each Dlib upsampling step multiplies latency about 3.5-fold; and YuNet on a central processor together with RetinaFace on a graphics processor form the Pareto front. On this basis, practical design rules for edge deployment are proposed.

Keywords: inference latency; input resolution; edge computing; real-time biometrics; Pareto efficiency; frame-rate budget; YuNet; Dlib.

DOI: <https://doi.org/10.47390/ydif-y2026v2i19/n06>

1. INTRODUCTION

In a facial recognition pipeline the detector is the only component that processes the entire frame; alignment, embedding extraction and matching operate only on the crops it returns [4; P. 3–5]. Detection cost therefore grows with image size, whereas the cost of the later stages grows with the number of faces. As surveillance and access-control cameras move from standard definition to 1080p and 2160p, detection becomes the bottleneck that decides whether a device keeps pace with the camera, how many streams it can serve and whether a graphics accelerator is needed at all.

Research on efficient detection has approached this problem mainly through architecture. Depthwise-separable convolutions and inverted residual blocks reduce the number of operations on mobile processors [2; P. 1–4], [6; P. 4510–4513]; compound scaling trades accuracy against cost in a controlled way [7; P. 10781–10783]; and computation can be redistributed across scales under a fixed budget [1; P. 1–3]. Published figures, however, are usually reported for a single input size on a single device. A system designer needs different information: how latency changes with resolution, which part of it is fixed and which part scales, and which resolution a given frame-rate budget permits. The aim of this study is to answer these questions for detectors of three performance tiers and to translate the answers into practical deployment rules.

2. LITERATURE REVIEW AND METHODOLOGY

The cost structure of a detector follows from its search strategy. Sliding-window detectors evaluate a classifier at every position and scale; the Viola–Jones cascade made this feasible by rejecting most windows early and computing features on an integral image in constant time [8; P. 137–140]. Multi-stage convolutional cascades first build an image pyramid and then refine proposals, so their cost also depends on the number of candidate faces [10; P. 1499–1501]. Systematic studies of object detectors have shown that the speed–accuracy relationship depends strongly on input resolution and on the feature extractor [3; P. 7310–7312], and detectors built on feature pyramids [5; P. 2117–2119] are typically run close to native resolution to keep small objects visible, which makes their cost resolution-dependent. By contrast, detectors with a fixed network input, such as YuNet with its 75 856 parameters, have a nearly constant inference cost, while resizing and colour conversion still scale with the frame [9; P. 658–661]. On identical hardware, the authors of YuNet report RetinaFace to be 22.3 times slower at 320×320 and about 28 times slower at the original image size [9; P. 662].

The study uses mean wall-clock latencies per frame measured on streams designated 480p, 1080p, 1920p and 2160p for four configurations: YuNet on a central processing unit (CPU) with a fixed 300×300 input; Dlib HOG on a CPU without upsampling ($k = 0$) and with one upsampling step ($k = 1$); and RetinaFace on a graphics processing unit (GPU). For area-based reasoning, 16:9 frames of 0.41, 2.07 and 8.29 megapixels are assumed for 480p, 1080p and 2160p respectively; the 1920p stream, whose geometry was not standardised, is used only in the budget analysis. A frame-rate budget f translates into a latency limit of $1/f$ (for example, 40 ms at 25 fps). The single composite index used is detection quality per second of computation:

$$\eta = F_1 / t, \quad (1)$$

where F_1 is the dominant-face F_1 -measure (0.941 for YuNet; 0.925 and 0.937 for Dlib with $k = 0$ and $k = 1$; 0.985 for RetinaFace) and t is the latency in seconds. A configuration belongs to the Pareto front if no other configuration is both faster and more accurate.

3. RESULTS AND DISCUSSION

From 480p to 2160p, YuNet latency rises from 10 to 112 ms, Dlib $k = 0$ from 40 to 315 ms, Dlib $k = 1$ from 140 to 1 060 ms and RetinaFace from 8 to 106 ms. Over this range the frame area grows 20.25-fold, whereas latency grows only 7.6–13.2-fold, so the cost is clearly sublinear overall. Between 1080p and 2160p, however, where the area quadruples, latency grows by factors of 3.50 (YuNet), 3.94 (Dlib $k = 0$), 3.79 (Dlib $k = 1$) and 3.53 (RetinaFace), i.e. almost proportionally. Two regimes can therefore be distinguished: at low resolutions fixed costs matter, whereas at high resolutions pixel-proportional costs dominate, and doubling the linear resolution roughly quadruples latency.

Fitting a straight line through the 1080p and 2160p points gives a fixed part of about 5 ms plus about 13 ms per megapixel for YuNet, and about 5 ms plus 12 ms per megapixel for RetinaFace; both lines predict the 480p latency to within 2 ms of the measurement. Since the YuNet network input is fixed, its pixel-proportional part cannot come from inference: it is the cost of decoding, converting and resizing large frames. At high resolution the lightweight detector is thus limited by preprocessing, a limitation that hardware-accelerated scaling or decoding at reduced size can remove. For Dlib the same line underestimates the 480p latency by a factor of about two (17 ms against the measured 40 ms without upsampling), showing that the HOG pipeline carries a substantial overhead that depends only weakly on frame size. Lowering Dlib's input below 1080p therefore saves disproportionately little time.

The latency ratio of Dlib with and without upsampling is 3.50, 3.50, 3.51 and 3.37 at the four resolutions. If all work were quadrupled by enlarging the image, the ratio would be 4; the observed value implies that roughly one sixth of the base processing is not multiplied, and the constancy of the ratio shows that this part is itself proportional to the original frame rather than a fixed overhead. The gain bought at this 3.5-fold cost is small: dominant-face recall rises by only 2.3 percentage points (from 0.860 to 0.883). At 1080p, upsampling adds 200 ms per frame, which excludes it from any real-time processing path.

According to index (1), YuNet reaches 94, 29, 9.2 and 8.4 s^{-1} at the four resolutions and RetinaFace 123, 33, 9.8 and 9.3 s^{-1} , while Dlib reaches only 23, 12, 3.8 and 2.9 s^{-1} without upsampling and 6.7, 3.3, 1.1 and 0.9 s^{-1} with it. Dlib is thus 2.4–4.1 times less efficient than YuNet without upsampling and 8–14 times less efficient with it. At 2160p the points for RetinaFace (106 ms; 0.985) and YuNet (112 ms; 0.941) form the Pareto front, and both Dlib configurations are dominated. This apparent parity must, however, be qualified by the unequal hardware: scaling the YuNet latency by the reference slowdown of 22–28 suggests about 0.7–0.9 s per 1080p frame for RetinaFace on a CPU. On CPU-only edge devices the front therefore collapses to YuNet alone.

At 30 fps (33 ms) only YuNet and RetinaFace qualify, and only up to 1080p; at 25 fps Dlib without upsampling joins them only at 480p, exactly at the limit. A 10 fps budget admits Dlib $k = 0$ up to 1080p, while YuNet (102 ms) and RetinaFace (101 ms) miss the 1920p limit by only 1–2 ms. At 5 fps YuNet and RetinaFace handle 2160p, Dlib $k = 0$ reaches 1080p and Dlib $k = 1$

only 480p; at 1 fps Dlib $k = 1$ reaches 1920p. Real-time processing of full 2160p frames is not attainable by any of the evaluated configurations.

Four rules follow from these results. First, on CPU-only devices with high-resolution cameras the detector should run on a copy of the frame downsampled to at most 1080p, while recognition uses crops taken from the full-resolution original; this preserves real-time throughput without changing the network. Second, Dlib upsampling should be confined to offline processing. Third, when a GPU is available, RetinaFace should be preferred for multi-person streams up to 1080p, where it equals YuNet in efficiency and exceeds it in quality. Fourth, a CPU-only cascade in which YuNet proposes face regions and RetinaFace verifies 320×320 crops is promising: with the reference cost of about 49 ms per crop [9; P. 662], one face at 1080p would require roughly $32 + 49 \approx 81$ ms, i.e. about 12 fps — an estimate that still awaits confirmation on identical hardware.

Latency is not the only resource that matters on edge devices. At 32-bit precision the 75 856 parameters of YuNet occupy about 0.30 MB, whereas the 27.3 million parameters of RetinaFace require about 109 MB — a difference that decides whether a model fits into the on-chip memory of a microcontroller-class accelerator or a low-cost camera module. Latency also bounds the number of streams a single processing unit can serve: at 1080p, YuNet's 32 ms per frame allows about three camera streams at 10 fps, whereas Dlib without upsampling (80 ms) serves only one, and Dlib with upsampling cannot sustain even one. For municipal installations with hundreds of cameras, such ratios translate directly into the number of servers to be procured.

The measurements were obtained on unequal hardware (a CPU for YuNet and Dlib, a GPU for RetinaFace); the OpenCV DNN detector was characterised only qualitatively, as the fastest model; and energy consumption was not recorded. The derived cost components and budget limits are therefore specific to the measured platforms, although the proportionality relations and the efficiency ranking are expected to transfer to other hardware.

4. CONCLUSION AND RECOMMENDATIONS

Face-detection latency grows sublinearly over the 480p–2160p range but almost in proportion to frame area between 1080p and 2160p. For YuNet and RetinaFace the data are described by a fixed part of about 5 ms plus 12–13 ms per megapixel, and the resolution dependence of YuNet stems from preprocessing rather than inference. Dlib carries a large overhead at low resolution, and each upsampling step multiplies its latency about 3.5-fold for a recall gain of only a few points. YuNet (CPU) and RetinaFace (GPU) form the efficiency front, both Dlib variants are dominated, and real-time operation at 25–30 fps is attainable only up to 1080p.

Designers of edge biometric systems are therefore advised to fix the frame-rate budget first and derive the detection resolution from it, to decouple detection resolution from recognition resolution, to accelerate resizing for lightweight detectors, and to keep upsampling out of real-time paths. Future work should measure all models on identical CPU, GPU and neural-accelerator hardware, add energy consumption per frame to the evaluation, and validate the proposed cascade experimentally.

Adabiyotlar/Литература/References:

1. Guo J., Deng J., Lattas A., Zafeiriou S. Sample and computation redistribution for efficient face detection // International Conference on Learning Representations (ICLR 2022). – arXiv:2105.04714. – 2022.
2. Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. MobileNets: efficient convolutional neural networks for mobile vision applications // arXiv preprint arXiv:1704.04861. – 2017.
3. Huang J., Rathod V., Sun C., Zhu M., Korattikara A., Fathi A., Fischer I., Wojna Z., Song Y., Guadarrama S., Murphy K. Speed/accuracy trade-offs for modern convolutional object detectors // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). – Honolulu, 2017. – P. 7310–7319.
4. Kortli Y., Jridi M., Al Falou A., Atri M. Face recognition systems: a survey // Sensors. – 2020. – Vol. 20, No. 2. – Art. 342. – DOI: 10.3390/s20020342.
5. Lin T.-Y., Dollár P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). – Honolulu, 2017. – P. 2117–2125.
6. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: inverted residuals and linear bottlenecks // Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018). – Salt Lake City, 2018. – P. 4510–4520.
7. Tan M., Pang R., Le Q.V. EfficientDet: scalable and efficient object detection // Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020). – Seattle, 2020. – P. 10781–10790.
8. Viola P., Jones M.J. Robust real-time face detection // International Journal of Computer Vision. – 2004. – Vol. 57, No. 2. – P. 137–154.
9. Wu W., Peng H., Yu S. YuNet: a tiny millisecond-level face detector // Machine Intelligence Research. – 2023. – Vol. 20, No. 5. – P. 656–665. – DOI: 10.1007/s11633-023-1423-y.
10. Zhang K., Zhang Z., Li Z., Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks // IEEE Signal Processing Letters. – 2016. – Vol. 23, No. 10. – P. 1499–1503.

YANGI DAVR ILM-FANI: INSON UCHUN INNOVATSION G'OYA VA YECHIMLAR

**XIX RESPUBLIKA ILMIY-AMALIY KONFERENSIYASI
MATERIALLARI**
2026-yil, oktyabr

Mas'ul muharrir: *F.T.Isanova*
Texnik muharrir: *N.Bahodirova*
Diszayner: *I.Abdihakimov*

Yangi davr ilm-fani: inson uchun innovatsion g'oya va yechimlar.
XIX Respublika ilmiy-amaliy konferensiyasi materiallari to'plami.
2-jild, 19-son (13-oktyabr, 2026-yil). – 191 bet.

Mazkur nashr ommaviy axborot vositasi sifatida 2025-yil, 8-iyulda
C-5669862 son bilan rasman davlat ro'yaxatidan o'tkazilgan.

ISSN: 3093-8791 (onlayn)

Elektron nashr: <https://konferensiyalar.com>

Konferensiya tashkilotchisi: "Scienceproblems Team" MChJ

Konferensiya o'tkazilgan sana: 2026-yil, 9-oktyabr.

Barcha huquqlar himoyalangan.
© Science problems team, 2026-yil.
© Mualliflar jamoasi, 2026-yil.