

ДЭТЭКЦЫІ І КЛАСІФІКАЦЫІ ЖЭСТАЎ РУК НА RASPBERRY PI ZERO 2W

Бекетаў Я.Д., студэнт гр.150504

*Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі
г. Мінск, Рэспубліка Беларусь*

Перцаў Д.Ю. – канд. тэхн. навук, дацэнт

Анатацыя. У дадзенай працы разглядаецца мадэль дэтэкцыі і класіфікацыі жэстаў рук на аднаплатным камп'ютары Raspberry Pi Zero 2W.

Ключавыя словы. Дэтэкцыя, класіфікацыя, CNN, BlazeFace, Raspberry Pi Zero 2W, навучанне, тэсціраванне.

Уводзіны. Жэсты з'яўляюцца натуральным спосабам камунікацыі і паддаюцца аналізу машынай. Дадзеным аналізам могуць займацца як масіўныя сістэмы і комплексы апрацоўкі, так і мабільныя прыстасаванні: тэлефоны, аднаплатныя камп'ютары. Аднаплатныя камп'ютары, такія як Raspberry Pi Zero 2w, пастаўляюць даступныя і кампактныя рашэнні для рэалізацыі тэхналогій звязаных з дэтэкцыяй і класіфікацыяй жэстаў рук. Нягледзячы на папулярнасць, гатовыя рашэнні па такіх тэхналогіях для Raspberry Pi Zero 2w прадстаўлены абмежавана. Адным з рашэнняў з'яўляецца фрэймворк MediaPipe [1]. У рамках дадзенай тэмы праводзіўся агляд і аналіз гэтага фрэймворка [2]. Улічваючы абмежаванасць работы MediaPipe на маламагутных прыстасаваннях, ёсць неабходнасць распрацаваць іншае рашэнне і арганізаваць яго перанос на Raspberry Pi Zero 2w.

Агляд мадэляў. Патрабаванні да выбіраемай мадэлі фармулююцца такім чынам: хуткасць, лёгкасць, інтэрпрэтуемасць на Raspberry Pi Zero 2w, дастатковая дакладнасць у распазнанні жэстаў рук. Па патрабаванні падыходзяць мадэлі віда CNN (Convolutional Neural Network): YOLO (You Only Look Once) [3] і прадстаўнікі SSD (Single-Shot Multi-box Detector) [4].

SSD з'яўляецца больш зручнай асновай для асабістай мадэлі дэтэкцыі жэстаў рук з-за таго, што моцна не саступае мадэлі YOLO у хуткасці і праяўляе лепшую працаздольнасць пры маленькіх аб'ектах (у дадзеным выпадку рукі). Таксама плюсам з'яўляецца шырокія магчымасці ў мадыфікацыі дадзенай мадэлі.

Прадстаўнікі мадэлі SSD. Нягледзячы на складанасці з устаноўкай і выкарыстоўваннем MediaPipe, вядома, што для дадзенага фрэймворка існуюць асобныя мадэлі дэтэкцыі, якія з'яўляюцца прадстаўнікамі SSD: BlazePalm [5] і BlazeFace [6]. Першая мадэль, выкарыстоўваецца MediaPipe для рашэнняў HandGesture Recognition і Hand Tracking, а мадэль BlazeFace – для MediaPipe Face Detector.

BlazeFace выкарыстоўваецца для хуткага і дакладнага распазнавання твараў у рэальным часе, што робіць яе зручнай для мабільных і убудаваных сістэм. BlazeFace выкарыстоўвае легкаважную архітэктру, што забяспечвае высокую хуткасць апрацоўкі і нізкае спажыванне рэсурсаў. На выхадзе мадэлі атрымліваецца: каардынаты абмежавальнай рамкі, клас аб'екта, упэўненасць.

BlazePalm выкарыстоўваецца для распазнавання жэсткаў рук. BlazePalm з'яўляецца больш складанай мадэллю, але паўтарае архітэктурны ідэі BlazeFace. Хуткасць апрацоўкі ніжэй, а рэсурсы выкарыстоўваюцца больш. На выхадзе мадэлі: каардынаты арэнтываў рукі, клас жэстаў рукі і ўпэўненасць распазнавання.

Якасць у дэтэкцыі BlazeFace горш для складаных аб'ектаў, але хуткасць дадзенай мадэлі павялічваецца ў некалькі разоў для машын з CPU. Таму дадзеная мадэль лягла ў аснову распрацаванай для Raspberry Pi Zero 2w асабістай мадэлі.

Асабістая мадэль. Для асабістай мадэлі выкарыстоўваюцца канцэпцыі і падыходы архітэктры BlazeFace. Як і ў BlazeFace, у асабістай мадэлі існуе слой BlazeBlock, які з'яўляецца спалучэннем слаёў depthwise convolution і pointwise convolution. У сваю чаргу слой double BlazeBlock у архітэктру асабістай мадэлі не ўвайшоў. Мадэль разбіта на два бэкбаны. У табліцы 1 і 2 прыведзена архітэктру першага і другога бэкбана адпаведна. Адносна BlazeFace асабістая мадэль выкарыстоўвае меншую колькасць параметраў за кошт таго, што ў яе архітэктру не ўваходзіць слой double BlazeBlock. Заключным слоём мадэлі з'яўляецца здабывальнік асаблівасцяў (feature extractor). Выхады з бэкбонаў камбінуюцца такім чынам, што пры апрацоўцы фотаздымкаў са слаёў з выходамі 16x16x88 і 8x8x96 фарміруюцца рэзультаты для класіфікатара і рэгрэсара. Значэнне “сырых” прадказанняў для класіфікатара і рэгрэсара не залежыць адзін ад аднаго, таму ў далейшым дазваляе пазбавіцца ад непатрэбнай апрацоўцы (напрыклад, калі няма неабходнасці падлічваць каардынаты абмежавальнай рамкі).

Табліца 1 – Архітэтура першага бэкбона

Слой	Памер увахода	Памер згортачнага ядра
Згортачны	128x128x3	5x5x3x24 (stride – 2)
BlazeBlock	62x62x24	3x3x24x1 1x1x24x24
BlazeBlock	62x62x24	3x3x24x1 1x1x24x28
BlazeBlock	62x62x28	3x3x24x1 (stride – 2) 1x1x24x32
BlazeBlock	31x31x32	3x3x24x1 1x1x24x36
BlazeBlock	31x31x36	3x3x24x1 1x1x24x42
BlazeBlock	31x31x42	3x3x24x1 (stride – 2) 1x1x24x48
BlazeBlock	16x16x48	3x3x24x1 1x1x24x56
BlazeBlock	16x16x56	3x3x24x1 1x1x24x64
BlazeBlock	16x16x64	3x3x24x1 1x1x24x72
BlazeBlock	16x16x72	3x3x24x1 1x1x24x80
BlazeBlock	16x16x80	3x3x24x1 1x1x24x88

Табліца 2 – Архітэтура другога бэкбона

Слой	Памер увахода	Памер згортачнага ядра
BlazeBlock	16x16x88	3x3x24x1 (stride – 2) 1x1x24x96
BlazeBlock	8x8x96	3x3x24x1 1x1x24x96
BlazeBlock	8x8x96	3x3x24x1 1x1x24x96
BlazeBlock	8x8x96	3x3x24x1 1x1x24x96
BlazeBlock	8x8x96	3x3x24x1 1x1x24x96

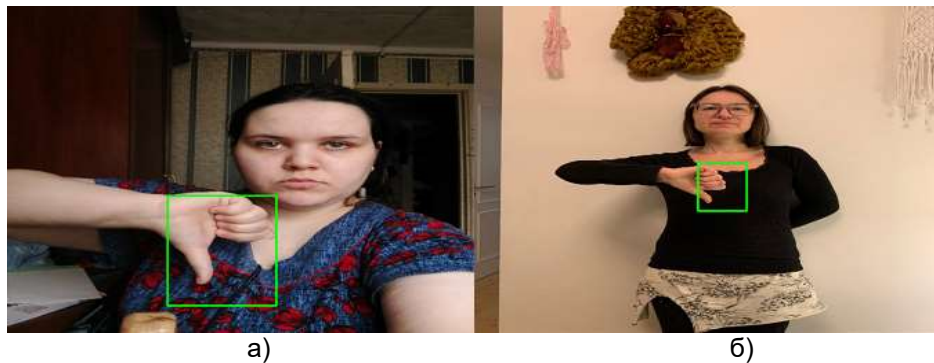
Выбар датасэта. У якасці датасэта выкарыстоўваецца NaGrid [7], а менавіта фотаздымкі для жэстаў “палец уверх” і “палец уніз”. Прыклады фотаздымкаў прыведзены на малюнку 1.



Малюнак 1 – Прыклад фотаздымкаў для класаў: а) – “палец уверх”; б) – “палец уніз”

Суадношенне для набору “палец уверх”: трэніровачны набор – 4454 шт.; валідацыйны набор – 528 шт.; тэставы набор – 1025 шт. Суадношенне для набору “палец уніз”: трэніровачны набор – 4129 шт.; валідацыйны набор – 537 шт.; тэставы набор – 844 шт. Дадзены датасэт не дае вялікую разнастайнасць фотаздымкаў жэстаў, што ўплывае на якасці трэніроўкі мадэлі.

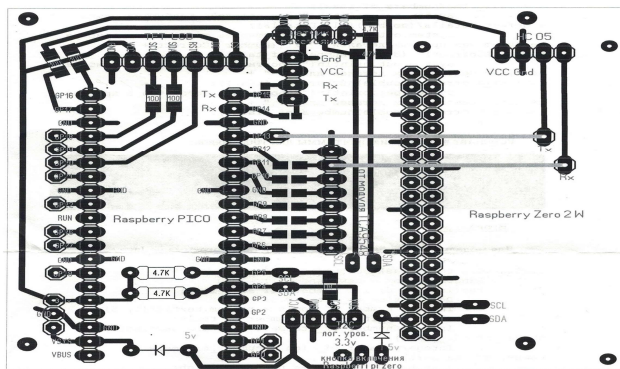
Тэсціраванне. У працэсе тэсціравання выкарыстоўваецца асобны праграмны модуль і тэставы набор дадзеных. Рэзультат дэтэкцыі мадэлі можна паглядзець на малюнку 2.



Малюнак 2 – Прыклад дэтэкцыі для класаў: а) – “палец уверх”; б) – “палец уніз”

Можна заўважыць, што дакладнасць вызначэння рамкі не вельмі высокая. Гэта абумоўлена спрошчанай архітэктурай і адлегласцю, на якой знаходзіцца аб’ект дэтэкцыі ад камеры. Улічваючы тое, што першасным з’яўляецца вызначэнне самога класа жэста, для вызначэння тэставых метрык уплыў дакладнасць рамкі змяншаецца. Актualityнай метрыкай з’яўляецца дакладнасць (Accuracy). Рэзультат падліку: 85%.

Пэранос. Апэрацыйная сістэма, на каторую арганізоўваецца перанос: Bookworm 12 Lite 64-bit [8]. Неабходнае праграмнае забеспячэнне: python, pip, picamera2, tensorflow (tensorflow lite), Pillow, numpy, opencv. Для запуску мадэлі выкарыстоўваецца праграма з інтэрпрэтарам tensorflow lite і патокавым чытаннем кадраў з дапамогай picamera2 з відэакамеры ўсталяванай на Raspberry Pi. Хуткасць апрацоўкі кадра мадэллю: 0.20 с. для дэтэкцыі і класіфікацыі, 0.004 с. толькі для класіфікацыі. Для дэманстрацыі рэзультатаў дэтэкцыі выкарыстоўваецца апаратная ўстаноўка: Raspberry Pi Zero 2W спалучаная з Raspberry Pi Pico і дыёдамі. Спалучэнне арганізоўваецца пасрэдна паслядоўнай асіметрычнай шыны I2C. На малюнку 3 прыведзена схема падключэння Raspberry Pi Zero 2W з Raspberry Pi Pico.



Малюнак 3 – Схема падключэння Raspberry Pi Zero 2W і Pico

Заклучэнне. Па выніках распрацоўкі на грунце мадэлі BlazeFace была створана асабістая мадэль. Дадзеная мадэль была паспяхова перанесена на Raspberry Pi Zero 2w і паказала хуткасць (0.004 секунд), каторая адпавядае патрабаванням для працавання ў рэальным часе.

Спіс выкарыстаных крыніц:

1. MediaPipe [Электронны рэсурс]: <https://developers.google.com/mediapipe>. – Дата доступу: 09.04.2024.
2. Агляд готовых праграмных рашэнняў па распазнаванні жэстаў рук. MediaPipe [Электронны рэсурс] - Рэжым доступу: <https://libeldoc.bsuir.by/handle/123456789/56965>. – Дата доступу: 17.03.2025.
3. You Only Look Once: Unified, Real-Time Object Detection [Электронны рэсурс] – University of Washington 9 мая 2016. – Рэжым доступу: <https://arxiv.org/pdf/1506.02640>. – Дата доступу: 21.02.2025.
4. SSD [Электронны рэсурс] – Рэжым доступу: <https://arxiv.org/pdf/1512.02325>. – Дата доступу: 21.02.2025.
5. Gesture recognition task guide [Электронны рэсурс]. – Рэжым доступу: https://ai.google.dev/edge/mediapipe/solutions/vision/gesture_recognizer. – Дата доступу: 21.02.2025.
6. BlazeFace [Электронны рэсурс] – Рэжым доступу: <https://arxiv.org/pdf/1907.05047>. – Дата доступу: 21.02.2025.
7. HaGRID [Электронны рэсурс] – Рэжым доступу: <https://github.com/hukenovs/hagrid>. – Дата доступу: 18.03.2025.
8. Operating system image. Raspberry Pi [Электронны рэсурс] – Рэжым доступу: <https://www.raspberrypi.com/software/operating-systems/#raspberry-pi-os-64-bit>. – Дата доступу: 09.04.2025.

HAND GESTURE DETECTION AND CLASSIFICATION ON RASPBERRY PI ZERO 2W

Beketau Y. D.

Belarusian State University of Informatics and Radioelectronics¹, Minsk, Republic of Belarus

Pertsau D. Y. – PhD in Technical Science

Annotation. This paper discusses a hand gesture detection and classification model on a Raspberry Pi Zero 2w.

Keywords. Detection, classification, CNN, BlazeFace, Raspberry Pi Zero 2W, model training, model testing.