https://doi.org/10.30857/2786-5371.2024.3.2

УДК 621.382

ДЕНІСОВ Р. В., ПОПОВИЧ П. В.
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ СИСТЕМ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ У РЕЖИМІ РЕАЛЬНОГО ЧАСУ НА МІКРОКОНТРОЛЕРАХ З ПОДАЛЬШИМ ГОЛОСОВИМ ВИВОДОМ ІНФОРМАЦІЇ ДЛЯ ЛЮДЕЙ З ВАДАМИ ЗОРУ

Мета. *Дослідження мінімального і максимального часу необхідного на проходження одного повного циклу розпізнавання-оголошення назви об'єкту з урахуванням різної довжини слів, різної швидкості розпізнавання об'єктів, а також фізичних особливостей людей з вадами зору для систем розпізнавання об'єктів у режимі реального часу на мікроконтролерах з подальшим голосовим виводом.*

Методика. *Створення варіантів комбінацій слів різної довжини з урахуванням можливості задання швидкості генерації мови у Espeak, та середньої швидкості мови в Україні. Розрахунок мінімальної і максимальної відстані до об'єкту на момент початку циклу розпізнавання-оголошення. Встановлено мінімальний і максимальний час необхідний на повний цикл розпізнавання-оголошення назви об'єкту.*

Результати. *На базі синтезатора мови Espeak та особливостях Українсько мови та мовлення було досліджено час необхідний для оголошення назв об'єктів різної довжини. Встановленого мінімальний та максимальний час проходження повного циклу розпізнавання-оголошення інформації з урахуванням фізичних особливостей людей з вадами зору, швидкості їх руху та швидкості реакції на голосову інформацію. Також отримано мінімальну і максимальну відстань до об'єкту на момент початку циклу, в залежності від часу необхідного на проходження одного повного циклу.*

Наукова новизна. *Отримано мінімальний і максимальний час необхідний на проходження повного циклу розпізнавання-оголошення інформації з урахуванням фізичних особливостей людей з вадами зору, технічних можливостей сучасних нейронних мереж та програм для синтезу мови, а також мінімальну і максимальну відстань до об'єкту на момент початку циклу. Досліджено мінімальну і максимальну відстань до об'єкту на момент початку циклу розпізнавання-оголошення.*

Практична значимість. *Отримані результати можуть бути використані при практичному створенні систем онлайн розпізнавання об'єктів, для оцінки можливості застосування тих чи інших нейронних мереж, спираючись на отриманий мінімальний та максимальний час проходження повного циклу розпізнавання-оголошення інформації, а також часу необхідного для проходження кожного з його окремих елементів.*

Ключові слова*: системи розпізнавання зображень; мікроконтролери; голосовий вивід інформації; згорткові нейронні мережі; TensorFlow; Espeak; MobileNet.*

Вступ. Активний розвиток нейронних мереж відкрив нові можливості для практичного застосування систем розпізнавання об'єктів, особливо у режимі реального часу. Медичні дослідження, системи автопілотування, робототехніка та комп'ютерний зір є основними напрямками досліджень і розвитку таких систем. При цьому, як для навчання так і для подальшого використання необхідні досить великі обчислювальні потужності [1–3].

Водночас, системи розпізнавання об'єктів на мікроконтролерах при високих показниках точності і швидкості розпізнавання потребують значно меншої кількості тренувальних даних і обчислювальної потужності для виконання розпізнавання [4–6]. Також, однією з основних переваг мікроконтролерів є компактність самих пристроїв, що дозволяє їх активно застосовувати для створення приладів які може носити людина.

Часткова або повна втрата зору є актуальною проблемою сьогодення, враховуючи кількість факторів, що можуть її спричинити. Використання нейронних мереж у системах розпізнавання об'єктів у режимі реального часу на мікроконтролері в поєднанні з подальшим

голосовим виводом розпізнаної інформації можуть допомогти людині в орієнтуванні у просторі, розпізнаванні об'єктів, що знаходяться перед нею, читанні тексту тощо.

Основні складнощі, які при цьому виникають, пов'язані з визначенням діапазонів технічних параметрів які необхідно врахувати, включно з фізичними особливостями людей з вадами зору, щоб розпізнана інформація була вчасно оголошена, і людина встигла на неї відреагувати. Адже різниця у часі необхідному на оголошення назви об'єкту напряму залежить від кількості слів і букв у ній, а також швидкості розпізнавання об'єкту нейронною мережею.

Постановка завдання. Метою даної роботи є визначення часу необхідного для проходження циклу розпізнавання-оголошення інформації з урахуванням різної довжини слів, швидкості руху та реакції людини, та різного часу необхідного на розпізнавання об'єкту. А також визначення мінімальної та максимальної відстані на якій необхідно починати оголошувати інформацію та частоту її оновлення.

Для здійснення розрахунків необхідно:

- визначити мінімальний і максимально час проходження одного циклу розпізнавання-оголошення
- визначити час необхідний на оголошення різних варіантів назв об'єктів в залежності від кількості слів і знаків у них
- встановити мінімальну і максимальну відстань до об'єкту на момент початку процесу розпізнавання-оголошення з урахуванням фізичних особливостей людей з вадами зору, швидкості розпізнавання об'єктів та часу необхідного на оголошення назви об'єктів, з врахуванням різної кількості букв і слів у цій назві, а також попереджувального слова.

Моя робота пропонує переглянути підхід до створення пристроїв що мають допомогти людям які втратили зір спираючись на їх фізичні особливості включно з швидкістю реакції та швидкістю руху. Також метою роботи є визначення мінімальної та максимальної відстані до об'єкту на момент початку циклу розпізнавання-оголошення.

Результати попередніх досліджень. Системи розпізнавання об'єктів у режимі реального часу набувають активного розвитку завдяки нейронним мережам. У роботі з визначення потоку машин у режимі реального часу було використано нейрону мережу YOLOX-L, як детектор, з використанням восьми графічних відеокарт GTX 2080ti для навчання системи. Час обробки вимірювали на графічному процесорі Tesla V100 [7].

У роботі з виявлення поліпів в режимі реального часу було використано архітектуру YOLOv3. Тести розпізнавань виконувались на ПК з процесором AMD Ryzen 5 2600 з частотою 3,4 ГГц, 16 Гб оперативної пам'яті, графічним процесором NVIDIA GeForce RTX 2080 Ti 11 Гб для виявлення поліпів і графічним процесором NVIDIA GeForce GTX 1050 Ti, як основний GPU для операційної системи [8].

Можливість розгортання нейронних мереж на мікроконтролерах відкрило нові сфери для застосування таких систем, у тому числі і для створення пристроїв, що мають покращити життя людям які втратили зір. Так у роботі з створення розумного капелюха для людей із вадами зору було представлено систему, що працює на модулі Raspberry Pi 4, нейронній мережі під назвою SSD MobileNet v2 320x320, із використанням TensorFlow Lite 2 для виявлення об'єктів. Швидкість аналізу моделі досягає близько 5 кадрів в секунду на Raspberry P 4 [9].

У роботі з розпізнавання об'єктів на основі глибокого навчання та опису навколишнього середовища для людей із вадами зору було використано модель SSDLite MobileNetV2 модуль синтезу мовлення Google, PyAudio та відеокамеру з мікроконтролером Raspberry Pi 4B. Отримана система може виявляти різні звичайні об'єкти із задовільною точністю 88,89% і швидкістю аналізу моделі у 2,15 кадри на секунду [10].

Основний акцент у цих роботах було зроблено на навчанні нейронних мереж і перевірки точності розпізнавання об’єктів. Але, для того, щоб розуміти які з існуючих нейронних мереж та систем голосового виводу для мікроконтролерів можна практично застосовувати системах онлайн розпізнавання об’єктів з подальшим голосовим виводом необхідно визначити як загальний час процесу розпізнавання-оголошення інформації, так і часові діапазони кожної окремої ланки з урахуванням фізичних особливостей людей з вадами зору, їх швидкості руху і швидкості реакції на голосову інформацію.

Далі наведено опис, умови проведення експерименту та власне його результати.

Основні нейроні мережі доступні для мікроконтролерів.

На даний момент найбільш прогресивні і доступні для практичного застосування на мікроконтролері нейроні мережі це – група моделей мереж від компанії Google – MobileNet, та TensorFlow Lite – від платорми TensorFlow.

MobileNet – це архітектура нейронної мережі, розроблена спеціально для використання на мобільних та вбудованих пристроях з обмеженими ресурсам. Основна ідея архітектури MobileNet полягає у тому, щоб забезпечити високу точність класифікації та виявлення об'єктів при мінімальному споживанні обчислювальних ресурсів та пам'яті.

MobileNetV1 та MobileNetV2 – дві основні моделі, що мають глобальну підтримку і оновлення. MobileNetV1 базується на двошаровій згортці, що дозволяє значно зменшити кількість параметрів та обчислень, зберігаючи при цьому високу точність розпізнавання [11]. Також за допомогою коефіцієнтів ширини мережі і зміни розміру вхідного зображення можна отримувати різну ефективність розпізнавання і кількість пам’яті необхідну для розгортання мережі.

MobileNetV2 – покращена версія, що базується на MobileNetV1 із додаванням інвертованого залишкового блоку з лінійними вузькими місцями, що зменшують розмір вхідних даних [12]. Така структура дозволила підвищити точність і швидкість розпізнавання об’єктів при зменшені розміру самої моделі.

TensorFlow Lite – це оптимізована версія фреймворку машинного навчання TensorFlow, призначена для роботи на пристроях з обмеженими ресурсами [13]. Конвертер TensorFlow Lite: дозволяє перетворювати звичайні TensorFlow моделі в оптимізований формат TensorFlow Lite FlatBuffer, включаючи оптимізації, такі як квантизація, що сприяє зменшенню розміру моделі та підвищує швидкість розпізнавання.

Ядро TensorFlow Lite виконує інференцію з оптимізованими моделями, взаємодіючи з різними операціями виконання через простий API. Можуть бути використані різні акселератори обчислень, такі як CPU, GPU, і спеціалізовані апаратні блоки. TensorFlow Lite використовує оптимізовані бібліотеки для обчислень, такі як Eigen (для CPU) та GPU kernels, що дозволяють максимально ефективно використовувати апаратні ресурси пристрою.

Платформа TensorFlow також дозволяє інтегрувати моделі класу MobileNet, виконати їх технічну оптимізацію, а також стиснення моделі.

Основні генератори мовлення доступні для мікроконтролерів.

Для використання на мікроконтролері на даний момент доступні і підтримуються такі синтезатори мови як Espeak, PicoTTS та Flite (Festival Lite).

Espeak – це компактний і простий у використанні синтезатор мови (Text-to-Speech, TTS) має низькі вимоги до ресурсів що дозволяє розгортати його на пристроях з обмеженими ресурсами [14]. ESpeak дозволяє користувачеві налаштовувати такі параметри синтезу мови як: швидкість мовлення, висота голосу, рівень гучності та акцент, використовуючи різні алгоритми синтезу мови, включаючи формантний синтез та конкатенативний синтез. Підтримує Українську мову.

PicoTTS – синтезатор мови розроблений компанією SVOX [15]. Працює під відкритою ліцензією Apache License 2.0. PicoTTS, використовує конкатенативний синтез мови, складаючи реальні аудіофрагменти (частини слів або фраз) для генерації мовлення. Підтримує різні мови та акценти, у тому числі Українську.

Flite (Festival Lite) – це компактний і легкий синтезатор мови (Text-to-Speech, TTS), який є спрощеною версією системи синтезу мови Festival. Flite, використовує конкатенативний синтез мови, підтримує різні мови та акценти, розповсюджується під відкритою ліцензією та є вільним програмним забезпеченням [16].

Враховуючи, що Espeak має досить гнучкі параметри налаштувань і чудово адаптований для застосування на мікроконтролерах, а також підтримує Українську мову, для розрахунків було обрано саме його.

Оцінка часу необхідного для проходження процесу розпізнавання-оголошення інформації з урахуванням фізичних особливостей людей які втратили зір.

Для оцінки часу необхідного для проходження циклу розпізнавання-оголошення інформації спершу необхідно встановити основні складові цього процесу, а також їх особливості. У загальному випадку цей процес складається з таких складових :

- Розпізнавання об’єктів
- Передача розпізнаної інформації до генератора мови
- Генерація і оголошення отриманого результату
- Отримання інформації людиною і реакція на неї.

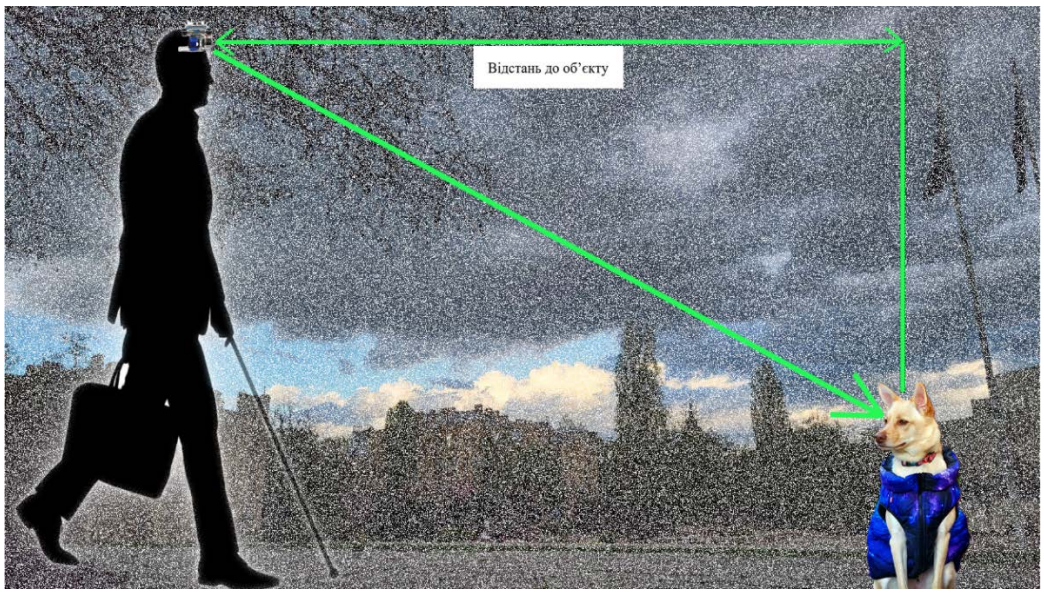

Рис. 1. **Умовне зображення процесу розпізнавання-оголошення**

На рис. 1 представлено умовне зображення процесу розпізнавання-оголошення. Кожна складова цього процесу має свої особливості і діапазони часу необхідні на їх виконання. Враховуючи, що час необхідний для оголошення інформації є складовою яку можна прорахувати і отримати діапазон часу в залежності від довжини і кількості слів, для розрахунку часу було обрано синтезатор мови Espeak. Він дозволяє задавати вручну такі технічні параметри як :

- базову висоту голосу у Гц та її діапазон;
- число формант, їх частоту, ширину і силу резонансних піків голосу;
- затримку (від 0 до 250 мс) та амплітуду ехо-сигналу;

- параметри звучання мови;
- швидкість мови у вигляді кількості слів на хвилину (від 80 до 450).

Аналіз результатів експерименту. В даному розділі наведено результати проведених досліджень.

Для розрахунку кількості часу необхідного для оголошення назви об'єкту необхідно розглянути можливі комбінації в залежності від кількості слів та їх довжини. Дослідження показують, що середній темп мовлення в Україні складає 5,5 складів на секунду (200 слів на хвилину), середня довжина Українських слів становить 6-8 символів, а оптимальна затримка (пауза) між словами складає 200 мс [17]. Можливі варіанти назв об'єктів – назва об'єкту складається з одного слова, назва складається з двох слів, варіант коли перед назвою об'єкту додається попереджувальне слово «попереду», а також словосполучення «нерозпізнаний об'єкт» у випадку коли об'єкт не було розпізнано.

Для зручності розрахунків було переведено швидкість мовлення з слів на хвилину до символів на секунду з врахування особливостей Української мови. При швидкості у 200 слів на хвилину отримуємо приблизно 3 слова на секунду. Взявши середню кількість символів у слові за 7, отримуємо швидкість у 21 символ на секунду, або приблизно 48 мс на оголошення одного символу. Також, при розрахунку назв, що складається з двох слів необхідно враховувати паузу між кожним словом, що складає в середньому 200 мс для Української мови. Для подальших розрахунків врахуємо, що у випадку коли назва об'єкту складається з одного слова, воно може бути коротким, середньостатистичним(середнім) або довгим. Коротким будемо вважати слово довжиною у 3–5 символи, середнє з довжиною у 6–8, та довге з довжиною у 9–12 символів. У ситуації коли назва об'єкту складається з двох слів – це можуть бути комбінації з двох коротких слів (к-к), короткого та середнього (к-с), короткого та довгого (к-д), двох середніх (с-с), середнього та довгого (с-д), або двох довгих (д-д). Враховуючи час необхідний на оголошення одного символу у 48 мс можна розрахувати час необхідний на оголошення слів різної довжини. У таблиці 1 представлено мінімальний, максимальний та середній час необхідний на оголошення назви об'єкту, що складається з одного слова.

Таблиця 1

Час необхідний на оголошення одного слова різної довжини

Назва об'єкту	Мінімальний час, мс	Максимальний час, мс	Середній час, мс
Коротка	144	240	192
Середня	288	384	336
Довга	432	576	504

У таблиці 2 наведено мінімальний, максимальний та середній час необхідний на оголошення назви об'єкту , яка складається з двох слів та їх можливі комбінації. Час вказаний без врахування паузи між словами у 200 мс.

Таблиця 2

Час необхідний на оголошення двох слів різної довжини

Комбінація слів	Мінімальний час, мс	Максимальний час, мс	Середній час, мс
К-К	288	480	384
К-С	432	624	528
К-Д	576	816	696
С-С	576	768	672
С-Д	720	960	840
Д-Д	864	1152	1008

З отриманих результатів можна зробити висновки, що час необхідний на оголошення може сильно відрізнятися в залежності від кількості слів та їх довжини, від 144 мс для короткого слова, на приклад «кіт», і максимальний час для двох довгих слів у 1,152 с.

Для допоміжного слова «попереду» необхідно 384 мс, а також 200 мс затримки між словами. Час необхідний для оголошення словосполучення «нерозпізнаний об'єкт» становить 1064 мс. Таким чином мінімальний час необхідний для назви об'єкту становить 144 мс без слова «попереду», і 728 мс з ним та затримкою. Максимальний час необхідний для назви об'єкту, що складається з двох довгих слів становить 1352 мс, і 1936 мс з урахуванням попереджувального слова та пауз між словами.

Швидкість розпізнавання об'єктів залежить від таких складових як – нейронна мережа, що використовується для розпізнання, апаратні можливості мікроконтролера – чим більше пам'яті доступної для використання тим менш стиснуту нейрону мережу можна застосувати і отримати кращі показники швидкості і точності розпізнавання, а також від зовнішніх умов, таких як освітленість та погодні умови. Якщо вірогідність розпізнавання об'єкту вище за 60% – результат можна вважати позитивним і оголосити назву об'єкту, у іншому випадку буде оголошено інформацію у форматі словосполучення «нерозпізнаний об'єкт». Враховуючи, що середній час розпізнавання об'єктів нейронними мережами може варіюватися в залежності від різних умов, для подальших розрахунків було обрано мінімальний час розпізнавання 600 мс, середній час розпізнавання у 900 мс та 1200 мс для складних умов розпізнавання.

Отже, загальний час($t_ц$) необхідний на один повний цикл процесу розпізнавання-оголошення назви розпізнаного об'єкту можна розрахувати за формулою:

$$t_ц = t_{роз} + t_{назви} + t_{реакції} \quad (1)$$

де $t_{роз}$ – час необхідний на розпізнавання об'єкту; $t_{назви}$ – час необхідний на оголошення назви об'єкту; $t_{реакції}$ – час реакції людини на голосову інформацію (160 мс). Таким чином, при мінімальному часі розпізнавання та найкоротшій назві об'єкту час буде складати 904 мс, та 3296 мс при максимальному часі розпізнавання та довгій назві об'єкту. Частота проходження циклів при цьому суттєво відрізняється. У випадку мінімального часу це виходить 1 повний цикл на секунду, а при максимальному часі необхідно більше 3 повних секунд на проходження одного циклу.

Враховуючи середню швидкість руху людей з вадами зору у незнайомій місцевості як 2.5 км/год, або 0.7 м/с ($V_Л$), а також знаючи максимальний та мінімальний час необхідний на проходження одного циклу($t_ц$) розпізнавання-оголошення назви об'єкту можна розрахувати мінімальну та максимальну відстань (S) на якій необхідно починати процес розпізнавання-оголошення за формулою:

$$S = V_Л \times t_ц \quad (2)$$

Мінімальна відстань при швидкості руху 0,7 м/с швидкості реакції 160 мс та мінімальним часом необхідним на цикл розпізнавання-оголошення 904 мс становить 0,63 м, а максимальна відстань 2,31 м для довгої назви об'єкту з урахуванням попереджувального слова та найбільшого часу розпізнавання.

У випадку нерозпізнаного об'єкту, мінімальна відстань при найменшому часі розпізнавання становить 0,92 м, і 1,54 м при найбільшому часі розпізнавання.

Проаналізувавши отримані результати можна зробити висновки, що час необхідний на оголошення назви об'єкту в залежності від кількості слів та букв у них суттєво впливає на загальний час циклу розпізнавання-оголошення. Різниця між максимальною назвою з двох довгих слів і мінімальною назвою з одного короткого слова складає 1008 мс, без урахування попереджувального слова, а час необхідний для оголошення нерозпізнаного об'єкту складає

1064 мс. Різниця у відстані на якій потрібно починати цикл розпізнавання-оголошення варіюється від 0,63 м до 2,31 м, що є досить суттєво.

Враховуючи фізичні особливості людей з вадами зору для реального застосування необхідно враховувати як максимальну з отриманих відстаней, щоб врахувати найгірші умови розпізнавання об'єкту та найдовшу назву об’єкту, що складається з декількох слів і попереджувального слова перед ним, так і мінімальну відстань у випадку коли об’єкти мають коротку назву з одного слова.

Висновки. Проведено дослідження необхідного часу на один цикл процесу розпізнавання-оголошення назви об’єкту для систем розпізнавання об’єктів у реальному часі для систем на мікроконтролерах з врахуванням особливостей кожного елементу цього процесу та фізичних особливостей людей з вадами зору. Встановлено мінімальний та максимальний час необхідний на оголошення назви об’єкту з врахуванням різної довжини слів і їх можливих комбінацій, а також час необхідний для словосполучення «нерозпізнаний об’єкт».

Також отримано мінімальний та максимальний час необхідний на один повний цикл процесу розпізнавання-оголошення з урахуванням різної швидкості розпізнавання об’єкту, часу необхідного на його оголошення, швидкості руху та реакції людини, що цю інформацію отримує. А також мінімальну та максимальну відстань до об’єкту на момент початку процесу розпізнавання.

Отримані результати показують, що частота проходження одного повного циклу може варіюватися від одного циклу на секунду до понад трьох секунд на один цикл у випадку назви об’єкту, що складається з двох слів та попереджувального слова, а також найдовшого часу розпізнавання. При цьому, для практичного застосування необхідно враховувати як максимальний час проходження одного повного циклу так і мінімальний, щоб уникнути ситуацій коли людина не встигає зреагувати на отриману інформацію, або об’єкти знаходяться один за одним на малій відстані.

References

1. Zheng, X., Chen, F., Lou, L., Cheng, P., Huang, Y. (2022). Real-Time Detection of Full-Scale Forest Fire Smoke Based on Deep Convolution Neural Network. *Remote Sens*, 14(3), 536. https://doi.org/10.3390/rs14030536.

2. Mahendru, M., Dubey, S. K. (2021). Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3. *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence), Noida, India, 2021, P. 734–740, DOI: 10.1109/Confluence51648.2021.9377064.

3. Rahman, S., Rony, J. H., Uddin, J., Samad, M. A. (2023). Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography. *J. Imaging*, 9(10), 216. https://doi.org/10.3390/jimaging9100216.

4. Moosmann, J., Giordano, M., Vogt, C., Magno, M. (2023). TinyissimoYOLO: A Quantized, Low-Memory Footprint, TinyML Object Detection Network for Low Power Microcontrollers. *2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems* (AICAS), Hangzhou, China, 2023, P. 1–5, DOI: 10.1109/AICAS57966.2023.10168657

Література

1. Zheng X., Chen F., Lou L., Cheng P., Huang Y. Real-Time Detection of Full-Scale Forest Fire Smoke Based on Deep Convolution Neural Network. *Remote Sens*. 2022. Vol. 14(3). Art. 536. https://doi.org/10.3390/rs14030536.

2. Mahendru M., Dubey S. K. Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3. *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence). Noida, India, 2021. P. 734–740. DOI: 10.1109/Confluence51648.2021.9377064.

3. Rahman S., Rony J. H., Uddin J., Samad M. A. Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography. *J. Imaging*. 2023. No. 9 (10). Art. 216. https://doi.org/10.3390/jimaging9100216.

4. Moosmann J., Giordano M., Vogt C., Magno M. TinyissimoYOLO: A Quantized, Low-Memory Footprint, TinyML Object Detection Network for Low Power Microcontrollers. *2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems* (AICAS),

Hangzhou, China, 2023. P. 1–5. DOI: 10.1109/AICAS57966.2023.10168657.

5. Wulfert, L., Wiede, C., Verbunt, M. H., Gembaczka, P., Grabmaier, A. (2022). Human Detection with A Feedforward Neural Network for Small Microcontrollers. *2022 7th International Conference on Frontiers of Signal Processing* (ICFSP), Paris, France, 2022, P. 14–22, DOI: 10.1109/ICFSP55781.2022.9924667.

5. Wulfert L., Wiede C., Verbunt M. H., Gembaczka P., Grabmaier A. Human Detection with A Feedforward Neural Network for Small Microcontrollers. *2022 7th International Conference on Frontiers of Signal Processing* (ICFSP), Paris, France, 2022. P. 14–22. DOI: 10.1109/ICFSP55781.2022.9924667.

6. Umayer-Murshed, R., Dhruba, S. K., Bhuian, M. T. I., Akter, M. R. (2022). Automated Level Crossing System: A Computer Vision Based Approach with Raspberry Pi Microcontroller. *2022 12th International Conference on Electrical and Computer Engineering* (ICECE), Dhaka, Bangladesh, 2022, P. 180–183, DOI: 10.1109/ICECE57408.2022.10089007.

6. Umayer-Murshed R., Dhruba S. K., Bhuian M. T. I., Akter M. R. Automated Level Crossing System: A Computer Vision Based Approach with Raspberry Pi Microcontroller. *2022 12th International Conference on Electrical and Computer Engineering* (ICECE), Dhaka, Bangladesh, 2022. P. 180–183. DOI: 10.1109/ICECE57408.2022.10089007.

7. Jinrong Yang, Songtao Liu, Zeming Li, Xiaoping Li, Jian Sun (2022). Real-Time Object Detection for Streaming Perception. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2022, P. 5385–5395, https://doi.org/10.48550/arXiv.2203.12338.

7. Jinrong Yang, Songtao Liu, Zeming Li, Xiaoping Li, Jian Sun. Real-Time Object Detection for Streaming Perception. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2022. P. 5385–5395. https://doi.org/10.48550/arXiv.2203.12338.

8. Nogueira-Rodríguez, A., Domínguez-Carbajales, R., Campos-Tato, F. et al. (2022). Real-time polyp detection model using convolutional neural networks. *Neural Comput & Applic*, 34, 10375–10396. https://doi.org/10.1007/s00521-021-06496-4.

8. Nogueira-Rodríguez A., Domínguez-Carbajales R., Campos-Tato F. et al. Real-time polyp detection model using convolutional neural networks. *Neural Comput & Applic*. 2022. No. 34. P. 10375–10396. https://doi.org/10.1007/s00521-021-06496-4.

9. Matshehla Konaite, Pius A. Owolawi, Temitope Mapayi, Vusi Malele, Kehinde Odeyemi, Gbolahan Aiyetoro, and Joseph S. Ojo (2021). Smart Hat for the blind with Real-Time Object Detection using Raspberry Pi and TensorFlow Lite. In: *Proceedings of the International Conference on Artificial Intelligence and its Applications* (icARTi '21). Association for Computing Machinery, New York, NY, USA, Article 6. https://doi.org/10.1145/3487923.3487929.

9. Matshehla Konaite, Pius A. Owolawi, Temitope Mapayi, Vusi Malele, Kehinde Odeyemi, Gbolahan Aiyetoro, Joseph S. Ojo. Smart Hat for the blind with Real-Time Object Detection using Raspberry Pi and TensorFlow Lite. In: *Proceedings of the International Conference on Artificial Intelligence and its Applications* (icARTi '21). Association for Computing Machinery, New York, NY, USA. 2021. Article 6. https://doi.org/10.1145/3487923.3487929.

10. Raihan Bin Islam, Samiha Akhter, Faria Iqbal, Md. Saif Ur Rahman, Riasat Khan (2023). Deep learning based object detection and surrounding environment description for visually impaired people. *Heliyon*, Vol. 9, Iss. 6, e16924. https://doi.org/10.1016/j.heliyon.2023.e16924.

10. Raihan Bin Islam, Samiha Akhter, Faria Iqbal, Md. Saif Ur Rahman, Riasat Khan. Deep learning based object detection and surrounding environment description for visually impaired people. *Heliyon*. 2023. Vol. 9, Iss. 6. e16924. https://doi.org/10.1016/j.heliyon.2023.e16924.

11. Howard, G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv*, 2017, P. 1–9. DOI: 10.48550/arXiv.1704.04861.

11. Howard G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv*. 2017. P. 1–9. DOI: 10.48550/arXiv.1704.04861.

12. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, P. 4510–4520. DOI: 10.1109/ CVPR.2018.00474.
13. TensorFlow Lite. *tensorflow.org/lite*. URL: https://www.tensorflow.org/lite.
14. Espeak. *espeak.sourceforge.net*. URL: https://espeak.sourceforge.net/
15. PicoTTS. URL: https://www.home-assistant.io/integrations/picotts/
16. Flite (Festival Lite). *cmuflite.org* URL : http://cmuflite.org/
17. Ishchenko, O. S. (2012). Holosni zvuky ukrainskoi movy zalezhno vid tempu movlennia: monohrafiia [Vowel sounds of the Ukrainian language depending on the tempo of speech: monograph]. Kyiv: Institute of the Ukrainian Language of the National Academy of Sciences of Ukraine. P. 109–119 [in Ukrainian].

12. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2018. P. 4510–4520. DOI: 10.1109/ CVPR.2018.00474.
13. TensorFlow Lite. *tensorflow.org/lite*. URL: https://www.tensorflow.org/lite.
14. Espeak. *espeak.sourceforge.net.* URL: https://espeak.sourceforge.net/
15. PicoTTS. URL: https://www.home-assistant.io/integrations/picotts/
16. Flite (Festival Lite). *cmuflite.org.* URL: http://cmuflite.org/
17. Іщенко О. С. Голосні звуки української мови залежно від темпу мовлення: монографія. К.: Інститут української мови НАН України, 2012. С. 109–119.

DENISOV ROSTYSLAV
Post graduate student,
Department of Acoustic and
Multimedia Electronic Systems,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
https://orcid.org/0000-0003-1146-9114
E-mail: rostikdenisov@gmail.com

POPOVYCH PAVLO
PhD, Associate Professor,
Department of Acoustic and
Multimedia Electronic Systems,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
http://orcid.org/0000-0002-1572-3127
Scopus Author ID: 55225965700
Researcher ID: J-6574-2017
E-mail: p.popovich80@gmail.com

DENISOV R. V., POPOVYCH P. V.
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

PECULIARITIES OF APPLICATION OF OBJECT RECOGNATION SYSTEM IN REAL TIME ON MICROCONTROLLERS WITH SUBSEQUENT VOISE OUTPUT OF INFORMATION FOR PEOPLE WITH VISUAL IMPAIRMENTS

Purpose. *The study of the minimum and maximum time required to complete one full cycle of object name recognition-announcement taking into account different word lengths, different object recognition speeds, as well as physical characteristics of visually impaired people for real-time object recognition systems on microcontrollers with subsequent voice output.*

Methodology*. Creating variants of combinations of words of different lengths, taking into account the possibility of setting the speed of speech generation in Espeak, and the average speed of speech in Ukraine. Calculation of the minimum and maximum distance to the object at the start of the recognition-announcement cycle. The minimum and maximum time required for a full cycle of object name recognition-announcement is set.*

Findings. *On the basis of the Espeak language synthesizer and the peculiarities of the Ukrainian language and speech, the time required to announce the names of objects of different lengths was investigated. The minimum and maximum time for completing the full cycle of information recognition-announcement is set, taking into account the physical characteristics of people with visual impairments, their speed of movement and the speed of reaction to voice information. The minimum and maximum distance to the object at the start of the cycle was also obtained, depending on the time required to complete one complete cycle.*

Originality. *The minimum and maximum time needed to complete the full cycle of information recognition and announcement was obtained, taking into account the physical characteristics of visually impaired people, the technical capabilities of modern neural networks and programs for speech synthesis, as well as the minimum and maximum distance to the object at the time of the start of the cycle. The minimum and maximum distance to the object at the start of the recognition-announcement cycle was studied.*

Practical value. *The obtained results can be used in the practical creation of online object recognition systems, to assess the possibility of using certain neural networks, based on the obtained minimum and maximum time for passing the complete cycle of recognition-announcement of information, as well as the time required for passing each of its separate elements.*

Keywords: *image recognition systems; microcontrollers; voice output of information; convolutional neural networks; TensorFlow; English; MobileNet.*