

УДК 621.382

ДЕНІСОВ Р. В., ОНИКІЄНКО Ю. О.

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського" Київ, Україна

ОСОБЛИВОСТІ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ НЕЙРОНИМИ МЕРЕЖАМИ НА ПРИКЛАДІ MOBILENETV1 ТА MOBILENETV2 В СИСТЕМАХ НА МІКРОКОНТРОЛЕРАХ

Мета. Дослідження залежності об'єму задіяної пам'яті мікроконтролера та часу розпізнавання зображення від типу згорткової нейронної мережі MobileNet V1 або V2 та їх гіперпараметрів.

Методика. Створення бази даних зображень, необхідних для навчання нейронної мережі з використанням програмної платформи Edge Impulse з послідовним завантаження мережі в пам'ять 32-х бітного мікроконтролера ESP32 для практичної оцінки характеристик мережі.

Результати. Виконано порівняння характеристик нейронних мереж MobileNetV1 та MobileNetV2. Проведено експерименти для визначення залежності часу розпізнавання обраних об'єктів на зображенні, об'єму використаної оперативної пам'яті та пам'яті програм, на базі мікроконтролера ESP-EYE з камерою, в залежності від коефіцієнту ширини мережі, розміру зображення та архітектури згорткової нейронної мережі. Визначено що час класифікації зображень для моделі мережі MobileNetV2 займає від трьох до десяти секунд і більше, що не є прийнятним для задач швидкого розпізнавання. Також встановлено, що для аналізу зображень 160 на 160 пікселів з максимальною шириною мережі моделі MobileNetV2 не вистачає пам'яті програм мікроконтролера. Використання мережі MobileNetV1 забезпечує дещо меншу точність розпізнавання, але вимагає значно менше ресурсів мікроконтролера та часу.

Наукова новизна. Перевірено на практиці особливості використання, можливості та обмеження нейронних мереж для розпізнавання зображень в системах на мікроконтролерах ESP32. Встановлено залежності об'єму задіяної пам'яті мікроконтролера та часу розпізнавання зображення від типу згорткової нейронної з метою вибору додаткових засобів обробки зображень для підвищення якості розпізнавання.

Практична значимість. Отримані результати дозволяють здійснити вибір нейронної мережі MobileNetV1 або MobileNetV2 в залежності від конкретних задач розпізнавання зображень системами на мікроконтролерах.

Ключові слова: мікроконтролери; розпізнавання зображень; згорткові нейронні мережі; Edge Impulse; MobileNet.

Вступ. Системи розпізнавання зображень із кожним днем знаходять все більше сфер застосування та подальшого розвитку. При тому, що системи розпізнавання зображень достатньо вибагливі до обчислювальних ресурсів, тенденцією сьогодення є збільшення їх використання у мікроконтролерних системах. Тому виникає необхідність адаптації алгоритмів розпізнавання та попередня обробка зображень для подальшого завантаження у мікроконтролер, та використання згорткових нейронних мереж для вбудованих систем.

Згорткові нейронні мережі (CNN) – клас глибоких штучних нейронних мереж, які були успішно застосовані для вирішення проблем класифікації зображень. Вони мають високу точність розпізнавання, але вимагають великого обсягу пам'яті [1–3]. Більшість моделей CNN містять згорткові шари, агрегувальні шари та повноз'єднані шари. Згорткові шари вимагають великих обчислень через операції множення та додавання. Повноз'єднані шари не залежать від пам'яті, оскільки всі навчені ваги зберігаються у вбудованій пам'яті, як, наприклад, реалізовано в [4].

Постановка завдання. Метою даної роботи є аналіз впливу гіперпараметрів обробки зображення та архітектури нейронної мережі MobileNet на об'єм використаних ресурсів мікроконтролера. Для здійснення аналізу необхідно:

- Створити базу даних з зображень, необхідних для навчання мережі та оцінити точність розпізнавання обраних категорій зображення;
- Встановити залежності об'єму задіяної пам'яті мікроконтролера та часу розпізнавання від типу згорткової нейронної мережі MobileNet V1 або V2 та їх гіперпараметрів.

Результати попередніх досліджень. Однією з перших згорткових нейронних мереж, що мала якість розпізнавання об'єктів на зображенні вище 80% була мережа AlexNet. Мережа AlexNet містить вісім шарів; перші п'ять згорткові шари, за деякими з них йшли агрегувальні шари з максимальним об'єднанням, а останні три повноз'єднані шари [5]. В еволюції архітектур типовою тенденцією є те, що мережі стають глибшими [6].

До основних переваг згорткових нейронних мереж відносять можливість перенавчання під різні задачі [7], масштабованість мережі за допомогою додаткових параметрів стиснення (зменшення глибини мережі, зменшення розміру зображення), а також можливості встановлення на мікроконтролери, FPGA (Field-Programmable Gate Array) та інших системах з обмеженою пам'яттю

CNN знаходять широке застосування у мікроконтролерних системах. Так у роботі з порівняння алгоритмів глибокого навчання та згорткових нейронних мереж, було створено згорткову мережу на розпізнавання 5 порід собак із точністю більше 90% для кожної з них [8]. Зменшення розмірів та інтеграція в мікроконтролерні системи значно розширює сфери застосування нейронних мереж та підвищує рівень автоматизації різних процесів, у яких зараз необхідне людське втручання. В роботі [9] представлено детектор хвороб рослин на основі плати STM32F746G-disco, яка має 1 Мбайт флеш-пам'яті та 340 кБайт оперативної пам'яті. Аналіз зображень виконувався за допомогою нейронних мереж. Для визначення ступені зіпсованості продуктів використано CNN, розгорнуту на платі NodeMCU з мікроконтролером ESP8266 [10].

Враховуючи тенденцію на зменшення розміру нейронних мереж для застосування на пристроях з обмеженими обчислювальними ресурсами компанія Google створила клас ефективних моделей для мобільних, вбудованих програм комп'ютерного зору MobileNets [11]. Мережі MobileNets засновані на спрощеній архітектурі, яка використовує згортки, що розділяються по глибині, для створення глибоких нейронних мереж меншого розміру. Компактні розміри досягаються введенням гіперпараметрів, що дозволяють стиснути мережу до необхідних розмірів.

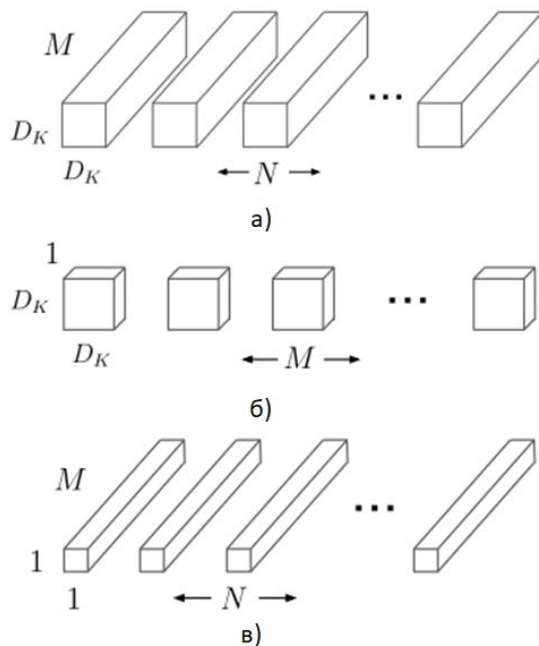
В роботі [12] описано систему для моніторингу споживання їжі у реальному часі на основі моделі нейронної мережі MobileNet. Запропонована мережа потребує 762 кБайт флеш-пам'яті та 500 кБайт оперативної пам'яті для реалізації на мікроконтролері STM32H7. Мережі потрібно в середньому 300 мс, щоб відобразити захоплене зображення на РК-моніторі та провести аналіз. Запропонована архітектура досягає середньої точності розпізнавання 0,82 на тестових зображеннях і показника 0,74 під час тестування у реальному часі. Модель безпілотного автомобіля з використанням MobileNet, як згорткової мережі для аналізу зображення, яке поступає з камери, запропоновано в [13]. Нейронна мережа має використовуватися для визначення смуги руху та прогнозування повороту ліворуч, праворуч і вперед на основі вхідного зображення.

Таким чином, використання нейронних мереж MobileNets для розпізнавання інформації і, зокрема, зображення, розширює функціональні можливості вбудованих систем на мікроконтролерах при необхідності врахування обмеженості їх ресурсів.

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

Особливості побудови нейронних мереж MobileNet. Модель MobileNet базується на згортках, які можна розділити по глибині, що є формою факторизованих згорток, тобто таких, які розкладають стандартну згортку на згортку по глибині, та згортку 1×1 , яка називається

поточною згорткою. Для MobileNets глибинна згортка застосовує один фільтр до кожного вхідного каналу. Потім поточкова згортка застосовує згортку 1×1 для об'єднання виходів поглибленої згортки. Стандартна згортка фільтрує та поєднує вхідні дані в новий набір виходів за один крок. Згортка, що розділяється по глибині, розділяє це на два шари: окремий шар для фільтрації та окремий шар для комбінування. Такий підхід дає в результаті значне зменшення обчислень і розміру моделі [11].



Джерело: [11].

Рис. 1. Етапи перетворення згорткових фільтрів: а) стандартна згортка, б) згортка по глибині, в) поточкова згортка

На рис. 1 показано, як стандартна згортка (а) розкладається на згортку по глибині (б) і поточкову згортку 1×1 (в). Стандартний згортковий шар приймає як вхідні дані $D_F \times D_F \times M$ карти функцій F і створює значення $D_G \times D_G \times N$ карти функцій G , де D_F – просторова ширина та висота квадратної вхідної карти функцій, M – кількість вхідних каналів (вхідна глибина), D_G – просторова ширина та висота квадратної вихідної карти ознак, а N – номер вихідного каналу (вихідна глибина). Стандартний згортковий шар параметризовано ядром згортки K розміром $D_K \times D_K \times M \times N$, де D_K – просторовий розмір ядра, що вважається квадратним, а M – кількість вхідних каналів, а N – кількість вихідних каналів, як визначено раніше.

Стандартні згортки мають наступну обчислювальну вартість (тобто кількість ресурсів, які використовує нейронна мережа для навчання або розпізнавання) [11]:

$$D_K \times D_K \times M \times N \times D_F \times D_F. \quad (1)$$

де вартість обчислення як добуток залежить від кількості вхідних каналів M , кількості вихідних каналів N , розміру ядра $D_K \times D_K$ і розміру карти функцій $D_F \times D_F$. Кожен з цих термінів та їх взаємодія стосуються моделі MobileNet. Вона використовує згортки, що розділяються по глибині, щоб прибрати залежність між кількістю вихідних каналів і розміром ядра.

Стандартна операція згортки має ефект фільтруючих функцій на основі згорткових ядер і комбінування функцій для створення нового представлення. Етапи фільтрації та об'єднання можна розділити на два етапи за допомогою факторизованих згорток для істотного зниження витрат на обчислення.

Згортка, що розділяється по глибині, складається з двох шарів: згортки по глибині та згортки по точкам. Згортка по глибині використовується, щоб застосувати один фільтр для кожного вхідного каналу (вхідна глибина). Поточкова згортка, проста згортка 1×1 , потім використовується для створення лінійної комбінації результату шару по глибині. Поглиблена згортка має наступну обчислювальну вартість:

$$D_K \times D_K \times M \times D_F \times D_F. \quad (2)$$

Поглиблена згортка надзвичайно ефективна порівняно зі стандартною згорткою. Однак вона лише фільтрує вхідні канали, а не поєднує їх для створення нових функцій. Отже, для генерації цих нових функцій потрібен додатковий шар, який обчислює лінійну комбінацію результату згортки по глибині через згортку 1×1 .

Комбінація згортки по глибині та згортки 1×1 (поточної) називається роздільною згорткою по глибині. Обчислювана вартість згорток, що розділяються по глибині:

$$D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F. \quad (3)$$

яка є сумою глибинної та 1×1 поточної згорток.

Виражаючи згортку як двоетапний процес фільтрації та об'єднання, можна отримати скорочення обчислень [11]:

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}. \quad (4)$$

MobileNet використовує згортки 3×3 із розділенням по глибині, які використовують у 8–9 разів менше обчислень, ніж стандартні згортки, лише з невеликим зниженням точності.

Хоча базова архітектура MobileNet є невеликою та має низьку затримку, для ще більшого зменшення об'єму та підвищення швидкості водиться параметр α , який називається множником ширини. Роль множника ширини α полягає в рівномірному проріджуванні мережі на кожному шарі. Для заданого шару та множника ширини α кількість вхідних каналів M стає αM , а кількість вихідних каналів N стає αN .

Обчислювальні витрати згортки, що розділяється по глибині, з множником ширини α [14_11]:

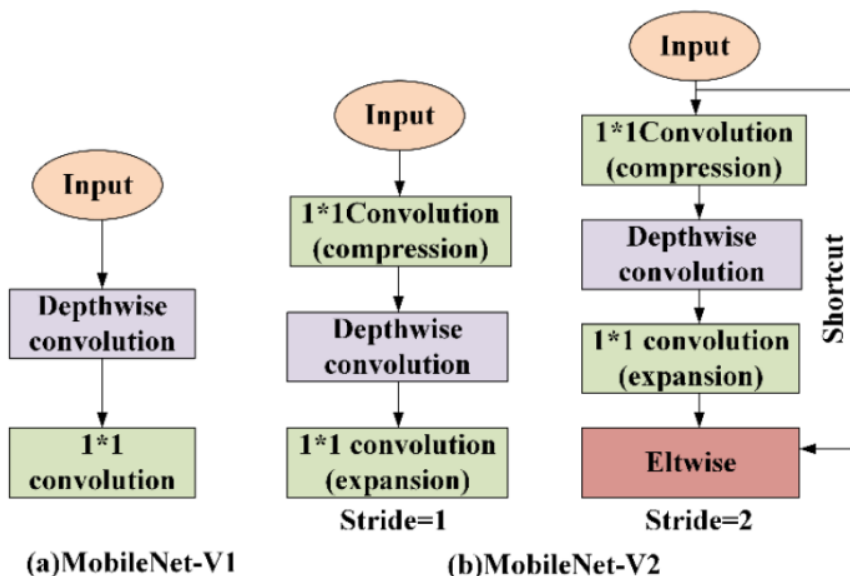
$$D_K \times D_K \times \alpha M \times D_F \times D_F + \alpha M \times \alpha N \times D_F \times D_F. \quad (5)$$

де $\alpha \in (0, 1]$ із типовими налаштуваннями 1, 0,75, 0,5 і 0,25. $\alpha = 1$ – це базова MobileNet, а $\alpha < 1$ – скорочені MobileNets. Множник ширини впливає на квадратичне зменшення обчислювальних витрат і кількості параметрів приблизно на α^2 . Множник ширини можна застосувати до будь-якої структури моделі, щоб визначити нову меншу модель із новою точністю, затримкою та розміром. Він використовується для визначення нової зменшеної структури, яку потрібно навчити з нуля.

Головна відмінність MobileNet V2 – це новий модуль шару: інвертований залишковий блок із лінійними вузькими місцями (bottleneck layers), які зменшують розмір вхідних даних. Цей модуль приймає як вхідні дані стиснене зображення, яке спочатку розширюється до великого розміру та фільтрується за допомогою згортки по глибині. Потім об'єкти проєктуються назад у зображення з лінійною згорткою [14].

Модуль може бути реалізований за допомогою стандартних операцій у будь-якому сучасному фреймворку. Крім того, цей згортковий модуль особливо підходить для мобільних проєктів, оскільки він дозволяє значно зменшити обсяг пам'яті, необхідний для розпізнавання.

На рис. 2 показано різницю між архітектурою MobileNetV1 і MobileNetV2, де згорткові блоки, що розділяються по глибині, і поточкові згорткові шари подібні до V1. Проте є незначна зміна структури шарів згортки. Тут перший шар – це згортка 1×1 із ReLu6 (тип функції активації Rectified Linear Units, яка є лінійною в додатному вимірі, але дорівнює нулю в від'ємному вимірі і має обмеження активації максимальним розміром 6), а другий шар виконує операцію згортки по глибині, а третій шар виконує згортку 1×1 без будь-якої нелінійності [15].



Джерело: [16].

Рис. 2. Архітектури MobileNetV1 та MobileNetV2

Таким чином, перевагами мереж MobileNets є досить висока точність розпізнавання при зменшених об'ємах обчислювальних витрат та вищій швидкості розпізнавання. Завдяки зменшенню необхідних технічних потужностей мережу MobileNets можна застосовувати на пристроях з обмеженими ресурсами таких параметрів як пам'ять програми (FLASH), оперативна пам'ять (RAM) при досить високій точності розпізнавання.

Експериментальна оцінка якості розпізнавання для різних моделей мережі. Для оцінки впливу типу та параметрів моделі нейронної мережі на якість і час розпізнавання зображень, а також на об'єм необхідних для цього ресурсів мікроконтролера проведено ряд експериментів з використанням вбудованих інструментів програмної платформи Edge Impulse [17]. Дана платформа створена для широкого загалу інженерів та дослідників, що використовують нейронні мережі для розпізнавання різноманітних даних: зображень, звукових сигналів, сигналів з сенсорів та інших даних, представлених у вигляді електричних сигналів. Платформа орієнтована на створення моделей, призначених для використання у вбудованих або у спеціалізованих системах. Edge Impulse використовує фреймворк Keras, що є інтерфейсом прикладного програмування (API) високого рівня платформи машинного навчання TensorFlow, розробленої компанією Google. TensorFlow – програмна бібліотека з відкритим кодом та набором функцій для машинного навчання на основі нейронних мереж глибинного навчання. Однією з складових TensorFlow є бібліотека моделей MobileNet. Edge Impulse надає можливість вибрати одну з декількох наявних моделей MobileNet та міняти її налаштування для використання у певній вбудованій системі виходячи з її наявних ресурсів. Edge Impulse надає перелік вбудованих систем та мікроконтролерів, в яких можна використовувати моделі MobileNet. Для наявних типів мікроконтролерів Edge Impulse обчислює використання пам'яті програми та оперативної пам'яті процесора та орієнтовний

час розпізнавання. Після вибору та тренування моделі нейронної мережі потрібно вибрати необхідну платформу розробки та завантажити код програми до пам'яті мікроконтролера.

Для проведення експериментів по розпізнаванню зображень вибрано модуль ESP-EYE, який представляє з себе плату розробки на основі 32-розрядного мікроконтролера ESP32 з підключеною зовнішню 2-мегапіксельною камерою [18]. На платі розміщено цифровий мікрофон, та мікросхему RAM ємністю 8 Мбайт. Крім того, плата може підтримувати передачу зображення через вбудований канал Wi-Fi. Наведений функціонал робить плату одним з найкращих варіантів для додатків, пов'язаних із виявленням і розпізнаванням обличчя та розпізнаванням мови. Мікроконтролер ESP32 має мікропроцесорне ядро Tensilica Xtensa LX6 від компанії Cadence Design Systems. Основні характеристики плати:

- Частота тактового генератора – 240 МГц
- Об'єм пам'яті програм (FLASH) – 4 Мбайт
- Об'єм оперативної пам'яті (SRAM) – 8 Мбайт
- Роздільна здатність камери – 2 мегапікселі

Зовнішній вигляд плати ESP-EYE з підключеною камерою наведено на рис. 3.

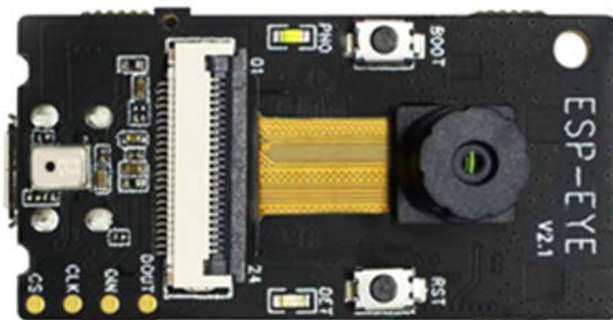


Рис. 3. Плата ESP-EYE з камерою

Для проведення експерименту створено дві групи зображень з назвами "men", "dogs". Обидві групи містять по 30 зображень відповідно людей та собак. Для забезпечення кращих результатів розпізнавання 80% зразків з кожної з груп даних, тобто 24 зразка, використано для тренування моделі нейронної мережі, а відповідно 6 зразків (20%) використано для перевірки. Платформа Edge Impulse дозволяє вибрати модель нейронної мережі зі списку доступних і виконати її тренування даними користувача. Після тренування моделі платформою Edge Impulse згенеровано код програми розпізнавання зображень для завантаження в пам'ять вибраного мікроконтролера. Програма також містить функції, необхідні для взаємодії з комп'ютером через USB та виводу результатів розпізнавання у командному рядку консолі. Зображення з камери модуля ESP-EYE можна переглядати у вікні браузера в процесі розпізнавання.

Результатом розпізнавання є імовірність визначення приналежності об'єкта на зображенні до груп даних, які було використано для тренування моделі. При чому сумарна імовірність приналежності до обох груп дорівнює 1 і розподіляється між ними пропорційно результатам розпізнавання. Сам процес розпізнавання є циклічним: після завершення першого порівняння робиться новий знімок і починається наступне розпізнавання.

В першому експерименті розглянуто залежність об'єму використаної FLASH та RAM пам'яті від типу вибраної моделі нейронної мережі. Edge Impulse містить дві моделі MobileNetV1 та MobileNetV2 з різним розміром матриці зображення та різним значенням множника ширини α , який змінює число вхідних та вихідних каналів нейронної мережі. При використанні моделі з більшим розміром матриці або з більшим значенням множника ширини

об'єм використаної FLASH та RAM зростає. Тому для всіх наявних моделей нейронної мережі Edge Impulse розраховано значення потрібного об'єму пам'яті та визначено для яких моделей достатньо FLASH пам'яті мікроконтролера ESP32 та RAM мікросхеми, а яким моделям пам'яті недостатньо.

У другому експерименті оцінено вплив значення множника ширини α на точність розпізнавання для вибраної моделі мережі, а саме MobileNetV2 160 x 160. Як вказано вище, множник ширини α є важливим чинником для зменшення розміру нейронної мережі, хоча і за рахунок зменшення точності розпізнавання. Експеримент виконано для всіх доступних варіантів моделі, які можна розмістити у FLASH пам'яті мікроконтролера та RAM модуля ESP-EYE.

У третьому експерименті виконано порівняння за основними характеристиками моделей MobileNetV1 96 x 96 та MobileNetV2 96 x 96 з різними значеннями множника ширини α , а саме за точністю розпізнавання, часу розпізнавання, об'єму використаної FLASH пам'яті та RAM. Мета експерименту – визначення найбільш ефективної моделі нейронної мережі для використання мікроконтролерами з обмеженими об'ємами пам'яті програм та оперативної пам'яті.

Аналіз результатів експерименту. В даному розділі наведено результати проведених експериментів. Точність розпізнавання визначалася на основі імовірності приналежності зображення групі "men". Тобто, якщо імовірність того, що зображення належить людині перевищувала 0,5, то результат вважався позитивним. Середнє значення ймовірності обчислювалось на основі 30 порівнянь.

Edge Impulse містить моделі MobileNetV1 та MobileNetV2 з різними значеннями коефіцієнта α . В таблиці 1 наведені значення необхідної FLASH та RAM мікроконтролера ESP32 для розгортання підтримуваних платформою Edge Impulse типів моделей нейронних мереж. Також в таблиці наведено практичний час, необхідний для виконання розпізнавання модулем ESP-EYE, та теоретичний ($t_{\text{розрах}}$), що надає платформа Edge Impulse. Моделі нейронної мережі MobileNetV1 96 x 96 займають в пам'яті програм до 310 кБайт та потребують до 131 кБайт оперативної пам'яті. Модель з $\alpha = 0.1$ може бути використана в мікроконтролерах, які мають не менше 128 кБайт RAM, інші моделі з множниками ширини $\alpha = 0.2$ та $\alpha = 0.25$ можна використати в мікроконтролерах з об'ємом RAM пам'яті в 256 кБайт. Таким вимогам задовольняє велика кількість мікроконтролерів з ядром ARM Cortex-M3 та M4. Також слід мати на увазі що розмір RAM пам'яті мікроконтролерів в декілька разів менше ніж розмір FLASH.

Таблиця 1

Залежність об'єму задіяної FLASH та RAM пам'яті від типу вибраної моделі мережі та точність розпізнавання обраних мереж

MobileNet	α	Точність	FLASH, кБ	RAM, кБ	$t_{\text{розрах}}, \text{мс}$	$t, \text{мс}$
V1 96x96	0.1	0,52	107,5	66,5	369	150
V1 96x96	0.2	0,56	224,8	106,7	722	440
V1 96x96	0.25	0,59	307,8	130,9	1620	622
V2 96x96	0.05	0,51	158,8	283,4	1140	572
V2 96x96	0.1	0,57	208,8	293,2	1254	633
V2 96x96	0.35	0,67	571,6	346,9	2430	1156
V2 160x160	0.35	0,83	625,5	733,6	11175	3231
V2 160x160	0.5	0,9	949,5	768,7	12796	5167
V2 160x160	0.75	0,95	1600	1400	8740	10725
V2 160x160	1		2500	1500	29193	25154

Для використання моделей MobileNetV2 96 x 96 з множниками ширини $\alpha = 0.05$, $\alpha = 0.1$ та $\alpha = 0.35$ необхідно щоб об'єм RAM мікроконтролера складав не менше ніж 512 кБайт. Таку оперативну пам'ять мають, наприклад, високопродуктивні і більш дорогі мікроконтролери від компанії STMicroelectronics з ядром ARM Cortex-M7. Для моделей MobileNetV2 96 x 96 з $\alpha = 0.05$ та $\alpha = 0.1$ необхідний розмір RAM більше ніж розмір FLASH. Ця особливість підкреслює, що моделі MobileNetV2 більш вимогливі до ресурсів пам'яті і в першу чергу до об'єму оперативної пам'яті. Тому для розпізнавання зображень використовується додаткова зовнішня пам'ять.

Використання моделі MobileNetV2 з матрицею зображень 160 x 160 є найбільш ресурсомістким. MobileNetV2 160x160 потребує не менше ніж 700 кБайт RAM для значень $\alpha = 0.35$ та $\alpha = 0.5$, найближче більше стандартне значення розміру RAM є 1 МБайт. Для $\alpha = 0.75$ та $\alpha = 1$ розмір оперативної пам'яті має становити 2 МБайт і більше. Таку оперативну пам'ять мають високопродуктивні процесори, призначені для обробки даних, але це вже не клас мікроконтролерних систем. Тому використання зовнішньої оперативної пам'яті, яка підключається за допомогою високошвидкісного інтерфейсу, наприклад SPI, є простим і недорогим рішенням. На платі ESP-EYE реалізовано саме такий варіант, хоча для моделі MobileNetV2 160 x 160 з $\alpha = 1$ (в таблиці рядок зі значеннями необхідної FLASH та RAM виділено жирним) у мікроконтролера не вистачило пам'яті. Необхідно зауважити, що час розпізнавання залежить від продуктивності мікроконтролера з високою тактовою частотою та скороченим часом виконання обчислювальних операцій виконують цю операцію швидше. Хоча і коштують дорожче.

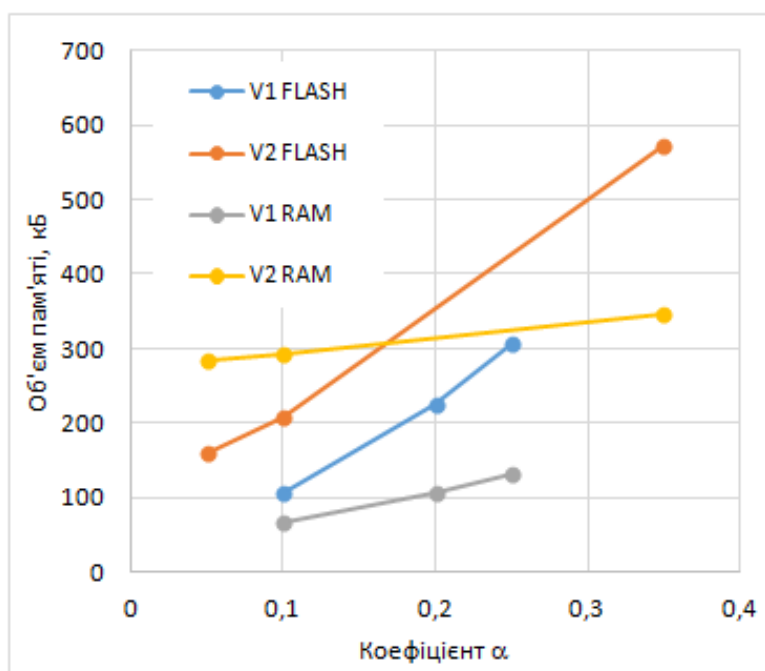


Рис. 4. Залежність об'єму використаної пам'яті мікроконтролера від значення коефіцієнту α

В таблиці 1 наведені значення точності (або імовірності) та час розпізнавання в залежності від значення коефіцієнта α для моделі нейронної мережі MobileNetV2 160x160. Розглянута модель забезпечує точність розпізнавання в межах 0,83–0,95, що є дуже хорошим значенням для переважної більшості застосунків. Однак час, який витрачається модулем ESP-EYE з тактовою частотою в 240 МГц для розпізнавання є дуже довгим. Зокрема для значення

$\alpha = 0.75$ час розпізнавання складає понад 10с, що може бути неприйнятним для швидкого розпізнавання. Таким чином, модель MobileNetV2 160x160 може використовуватись з певними обмеженнями 32-х бітними мікроконтролерами загального призначення через низьку продуктивність навіть при наявності достатнього об'єму оперативної пам'яті. Більш доцільним в таких умовах є використання моделей MobileNet з меншою роздільною здатністю, а саме з розміром зображення 96x96 пікселя.

Таким чином, аналіз наведених в таблиці даних показує, що модель мережі MobileNetV2 не має чітко виражених переваг над моделлю MobileNetV1. Модель MobileNetV2 використовує в 2–2,5 рази більше оперативної пам'яті і має в декілька разів довший час розпізнавання (в залежності від значення коефіцієнта α) ніж MobileNetV1, хоча забезпечує дещо кращу якість розпізнавання. Якщо порівняти характеристики моделей MobileNet V1 та V2 для $\alpha = 0.1$, то можна побачити, що точність розпізнавання для версії моделі V2 складає 0,57 у той час як для версії V1 є 0,52. Однак такий достатньо невеликий приріст у точності обходиться збільшенням розміру моделі практично у двічі (FLASH), збільшенням об'єму задіяної оперативної пам'яті у 4 рази та збільшення часу розпізнавання також у 4 рази. Графіки залежності об'єму використаної пам'яті мікроконтролера від значення коефіцієнту α наведено на рис. 4.

З рис. 4 можна зробити висновок що, у версії V2 збільшення необхідних ресурсів зростає значно стрімкіше, ніж V1. Тому для застосування у системах з обмеженими ресурсами для аналізу малих образень більше підходить MobileNetV1.

Висновки. В результаті експериментальних досліджень використання моделей нейронних мереж MobileNetV1 та MobileNetV2 для розпізнавання зображень 96x96 пікселів встановлено, що розгортання і робота нейронної мережі на 32-х бітних мікроконтролерах можливі за умови достатнього розміру вбудованих пам'яті програм та оперативної пам'яті. Модель MobileNetV1 для $\alpha = 0.1$ може бути використана у мікроконтролерах зі розміром RAM від 96 кБайт. Модель MobileNetV1 з іншими значеннями α потребує використання мікроконтролера з 256 кБайт і більше вбудованої оперативної пам'яті, що не є проблемою для реалізації.

При розпізнаванні зображень з більшою роздільною здатністю, зокрема 160x160 пікселів, за допомогою нейронних мереж MobileNetV2 зростає об'єм необхідної пам'яті мікроконтролера, в першу чергу оперативної. Проблему можна вирішити використанням зовнішньої пам'яті. Також збільшується час розпізнавання. У цьому випадку потрібно використовувати більш продуктивні мікроконтролери, що здорожує вирішення задачі розпізнавання.

Аналіз порівняння характеристик моделей нейронних мереж MobileNetV1 та MobileNetV2, пропонованих Edge Impulse, показує, що хоча модель версії V2 має вищу точність розпізнавання ніж модель версії V1, час розпізнавання значно більший ніж у моделі MobileNetV1. Крім того об'єми потрібної FLASH та оперативної пам'яті у декілька разів перевищують аналогічні параметри моделі версії V1. Отже системи розпізнавання зображень на 32-х бітних мікроконтролерах типу Tensilica Xtensa LX6 та їм аналогічних краще будувати на моделі мережі MobileNetV1 з використанням додаткової оперативної пам'яті.

References

1. Paluru, N. et al. (2021). Anam-Net: Anamorphic Depth Embedding-Based Lightweight CNN for Segmentation of Anomalies in COVID-19 Chest CT Images. In: *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 3, P. 932–946. DOI: 10.1109/TNNLS.2021.3054746.

Література

1. Paluru N. et al. Anam-Net: Anamorphic Depth Embedding-Based Lightweight CNN for Segmentation of Anomalies in COVID-19 Chest CT Images. In: *IEEE Transactions on Neural Networks and Learning Systems*. 2021. Vol. 32, No. 3. P. 932–946. DOI: 10.1109/TNNLS.2021.3054746.

2. Çalik, R. C., Demirci, M. F. (2018). Cifar-10 Image Classification with Convolutional Neural Networks for Embedded Systems. *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, 2018, P. 1–2. DOI: 10.1109/AICCSA.2018.8612873.
3. Hidayatuloh, A., Nursalman, M., Nugraha, E. (2018). Identification of Tomato Plant Diseases by Leaf Image Using Squeezenet Model. *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2018, P. 199–204. DOI: 10.1109/ICITSI.2018.8696087.
4. Phu, H. V., Minh Tan, T., Van Men, P., Van Hieu, N., Van Cuong, T. (2019). Design and Implementation of Configurable Convolutional Neural Network on FPGA. *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*, 2019, P. 298–302. DOI: 10.1109/NICS48868.2019.9023810.
5. Thalagala, S., Walgampaya, C. (2021). Application of AlexNet convolutional neural network architecture-based transfer learning for automated recognition of casting surface defects. *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 2021, P. 129–136. DOI: 10.1109/SCSE53661.2021.9568315.
6. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Li Wang, Gang Wang, Jianfei Cai, Tsuhan Chen (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, Vol. 77, P. 354–377. DOI: 10.1016/j.patcog.2017.10.013.
7. Mahardi, Wang, I., Lee, K. C., Chang, S. (2020). Images Classification of Dogs and Cats using Fine-Tuned VGG Models. *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, 2020, P. 230–233. DOI: 10.1109/ECICE50847.2020.9301918.
8. Weerasekara, D. T., Gamage, M. P. A. W., Kulasooriya, K. S. A. F. (2021). Combined Approach of Supervised and Unsupervised learning for Dog Face Recognition. *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, P. 1–5. DOI: 10.1109/I2CT51068.2021.9418175.
9. de Vita, F., Nocera, G., Bruneo, D., Tomaselli, V., Giacalone, D., Das, S. K. (2020). Quantitative Analysis of Deep Leaf: a Plant Disease Detector on the Smart Edge. *2020 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2020, P. 49–56. DOI: 10.1109/SMARTCOMP50058.2020.00027.
2. Çalik R. C., Demirci M. F. Cifar-10 Image Classification with Convolutional Neural Networks for Embedded Systems. *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. 2018. P. 1–2. DOI: 10.1109/AICCSA.2018.8612873.
3. Hidayatuloh A., Nursalman M., Nugraha E. Identification of Tomato Plant Diseases by Leaf Image Using Squeezenet Model. *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*. 2018. P. 199–204. DOI: 10.1109/ICITSI.2018.8696087.
4. Phu H. V., Minh Tan T., Van Men P., Van Hieu N., Van Cuong T. Design and Implementation of Configurable Convolutional Neural Network on FPGA. *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*. 2019. P. 298–302. DOI: 10.1109/NICS48868.2019.9023810.
5. Thalagala S., Walgampaya C. Application of AlexNet convolutional neural network architecture-based transfer learning for automated recognition of casting surface defects. *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)*. 2021. P. 129–136. DOI: 10.1109/SCSE53661.2021.9568315.
6. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Li Wang, Gang Wang, Jianfei Cai, Tsuhan Chen. Recent advances in convolutional neural networks. *Pattern Recognition*. 2018. Vol. 77. P. 354–377. DOI: 10.1016/j.patcog.2017.10.013.
7. Mahardi, Wang I., Lee K. C., Chang S. Images Classification of Dogs and Cats using Fine-Tuned VGG Models. *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*. 2020. P. 230–233. DOI: 10.1109/ECICE50847.2020.9301918.
8. Weerasekara D. T., Gamage M. P. A. W., Kulasooriya K. S. A. F. Combined Approach of Supervised and Unsupervised learning for Dog Face Recognition. *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, P. 1–5. DOI: 10.1109/I2CT51068.2021.9418175.
9. de Vita F., Nocera G., Bruneo D., Tomaselli V., Giacalone D., Das S. K. Quantitative Analysis of Deep Leaf: a Plant Disease Detector on the Smart Edge. *2020 IEEE International Conference on Smart Computing (SMARTCOMP)*. 2020. P. 49–56. DOI: 10.1109/SMARTCOMP50058.2020.00027.

10. Keshavamurthy, Mariyam, S. J., Meghamala, M., Meghashree, M., Neha (2019). Automatized Food Quality Detection and Processing System Using Neural Networks. *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, 2019, P. 1442–1446. DOI: 10.1109/RTEICT46194.2019.9016919.
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*, 17 Apr 2017, P. 1–9. DOI: 10.48550/arXiv.1704.04861.
12. Hossain, D., Imtiaz, M. H., Ghosh, T., Bhaskar, V., Sazonov, E. (2020). Real-Time Food Intake Monitoring Using Wearable Egocnetric Camera. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2020, P. 4191–4195. DOI: 10.1109/EMBC44109.2020.9175497.
13. Thadeshwar, H., Shah, V., Jain, M., Chaudhari, R., Badgujar, V. (2020). Artificial Intelligence based Self-Driving Car. *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 2020, P. 1–5. DOI: 10.1109/ICCCSP49186.2020.9315223.
14. 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.
15. Teng, T. W., Veerajagadheswar, P., Ramalingam, B., Yin, J., Elara Mohan, R., Gómez, B. F. (2020). Vision Based Wall Following Framework: A Case Study With HSR Robot for Cleaning Application. *Sensors* (Basel, Switzerland), 2020, Vol. 20(11): 3298. DOI: 10.3390/s20113298.
16. Wang, Y., Yan, J., Sun, Q., Li, J., Yang, Z. (2019). A MobileNets Convolutional Neural Network for GIS Partial Discharge Pattern Recognition in the Ubiquitous Power Internet of Things Context: Optimization, Comparison, and Application. In: *IEEE Access*, Vol. 7, P. 150226–150236. DOI: 10.1109/ACCESS.2019.2946662.
17. Edge impulse. *edgeimpulse.com*. URL: <https://www.edgeimpulse.com/>
18. ESP-EYE Development Board. *espressif.com*. URL: <https://www.espressif.com/en/products/devkits/esp-eye/overview>.
10. Keshavamurthy, Mariyam S. J., Meghamala M., Meghashree M., Neha. Automatized Food Quality Detection and Processing System Using Neural Networks. *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. 2019. P. 1442–1446. DOI: 10.1109/RTEICT46194.2019.9016919.
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*. 17 Apr 2017. P. 1–9. DOI: 10.48550/arXiv.1704.04861.
12. Hossain D., Imtiaz M. H., Ghosh T., Bhaskar V., Sazonov, E. Real-Time Food Intake Monitoring Using Wearable Egocnetric Camera. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2020, P. 4191–4195. DOI: 10.1109/EMBC44109.2020.9175497.
13. Thadeshwar H., Shah V., Jain M., Chaudhari R., Badgujar V. Artificial Intelligence based Self-Driving Car. *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*. 2020. P. 1–5. DOI: 10.1109/ICCCSP49186.2020.9315223.
14. 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.
15. Teng T. W., Veerajagadheswar P., Ramalingam B., Yin J., Elara Mohan R., Gómez B. F. Vision Based Wall Following Framework: A Case Study With HSR Robot for Cleaning Application. *Sensors* (Basel, Switzerland). 2020. Vol. 20(11). 3298. DOI: 10.3390/s20113298.
16. Wang Y., Yan J., Sun Q., Li J., Yang Z. A MobileNets Convolutional Neural Network for GIS Partial Discharge Pattern Recognition in the Ubiquitous Power Internet of Things Context: Optimization, Comparison, and Application. In: *IEEE Access*. 2019. Vol. 7. P. 150226–150236. DOI: 10.1109/ACCESS.2019.2946662.
17. Edge impulse. *edgeimpulse.com*. URL: <https://www.edgeimpulse.com/>
18. ESP-EYE Development Board. *espressif.com*. URL: <https://www.espressif.com/en/products/devkits/esp-eye/overview>.

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

Onykiienko Yu.O.

PhD, Associate Professor
Department of Acoustic and
Multimedia Electronic Systems,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
<https://orcid.org/0000-0001-7508-8391>
Scopus Author ID: 55819618700
Researcher ID: J-6589-2017
E-mail: yurionik@gmail.com

DENISOV R. V., ONYKIENKO Yu. O.

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

**PECULIARITIES OF IMAGE RECOGNITION BY NEURAL NETWORKS
ON THE EXAMPLE OF MOBILENETV1 AND MOBILENETV2
IN MICROCONTROLLER SYSTEMS**

Purpose. Research on the dependence of the amount of microcontroller memory used and the image recognition time on the type of convolutional neural network MobileNet V1 or V2 and their hyperparameters.

Methodology. Creating a database of images necessary for training a neural network using the Edge Impulse software platform with subsequent loading of the network into the memory of a 32-bit microcontroller ESP32 for practical evaluation of network characteristics.

Findings. A comparison of the characteristics of MobileNetV1 and MobileNetV2 neural networks was made. Experiments were conducted to determine the dependence of the recognition time of selected objects on the image, the amount of used RAM and program memory, based on the ESP-EYE microcontroller with a camera, depending on the network width ratio, image size and convolutional neural architecture network. It was determined that the time of image classification for the MobileNetV2 network model takes from three to ten seconds or more, which is not acceptable for fast recognition tasks. It was also found that the microcontroller program memory is insufficient to analyze the 160 by 160 pixel images with the maximum network width of the MobileNetV2 model. Using the MobileNetV1 network provides slightly lower recognition accuracy, but requires significantly less microcontroller resources and time.

Originality. The peculiarities of the use, possibilities and limitations of neural networks for image recognition in systems based on ESP32 microcontrollers have been verified in practice. The dependence of the amount of memory used on the microcontroller and the time of image recognition on the type of convolutional neuron was established in order to select additional image processing tools to improve the quality of recognition.

Practical value. The obtained results allow the proper selection of the MobileNetV1 or MobileNetV2 neural network depending on the specific tasks of image recognition by systems on microcontrollers.

Keywords: microcontrollers; image recognition; convolutional neural networks; Edge Impulse; MobileNet.