

**ПОЛТАВСЬКИЙ ДЕРЖАВНИЙ АГРАРНИЙ УНІВЕРСИТЕТ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ЕКОНОМІКИ, УПРАВЛІННЯ,
ПРАВА ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ**

Освітньо-професійна програма Інформаційні управляючі системи та
технології

Спеціальність 126 Інформаційні системи та технології
Ступінь вищої освіти Магістр

ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ
Завідувач кафедри

_____ Юрій УТКІН
«15» грудня 2022 року

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Розширення функціоналу системи обліку енергоресурсів за
допомогою нейронної мережі»

виконав здобувач вищої освіти dennої форми навчання

Пілюгін Володимир Андрійович

Керівник кваліфікаційної роботи
професор, д. т. н.

Вадим СЛЮСАР

Полтава – 2022 року

ЗМІСТ

ВСТУП	6
РОЗДІЛ 1. АНАЛІЗ ОСОБЛИВОСТЕЙ РЕАЛІЗАЦІЇ КОНЦЕПЦІЙ ЦИФРОВИЗАЦІЇ ПРОМИСЛОВОГО СЕКТОРУ	9
1.1 Загальні відомості про промисловий інтернет речей	9
1.2 Особливості інтеграції аналогових засобів обліку у сучасну інфраструктуру підприємства	12
1.3 Шляхи вирішення проблеми оцифровки показань з аналогових лічильників	16
1.4 Сутність семантичної сегментації зображень	17
Висновки до розділу 1	20
РОЗДІЛ 2. ІНСТРУМЕНТАРІЙ ДЛЯ РОЗШИРЕННЯ ФУНКЦІОНАЛУ СИСТЕМИ ОБЛІКУ ЕНЕРГОРЕСУРСІВ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ	22
2.1 Методи глибокого навчання	22
2.2 Багаторівнева глибока інтеграція функцій	28
2.3 Архітектура U-Net ²	31
Висновки до розділу 2	41
РОЗДІЛ 3. РЕКОМЕНДАЦІЙ ЩОДО ВИКОРИСТАННЯ ЗАПРОПОНОВАНИХ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ	43
3.1 Визначення особливостей формування бази для формування датасету	43
3.2 Особливості використання датасету Water Meters	45
3.3 Оцінка точності синтезованих архітектур нейронних мереж ...	47
3.4 Застосування механізму уваги	55
3.5 Техніко-економічне обґрунтування прийнятих рішень	59
Висновки до розділу 3	61
ВИСНОВКИ	63
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	65
ДОДАТКИ	70

ВСТУП

Актуальність теми кваліфікаційної роботи підтверджується необхідністю вирішення одного з завдань реалізації концепцій цифровізації, полягає у віддаленому зборі даних з множини аналогових засобів контролю витрат енергоресурсів. У багатьох випадках їхня заміна на сучасні цифрові пристрії є нерентабельною. Це пов'язано з великою різноманітністю та кількістю аналогових засобів обліку в інфраструктурі, наявністю заборон на внесення змін до комунікацій, високою вартістю розробок проектно-технічної документації щодо модернізації обладнання тощо. Як наслідок, можна йти шляхом використання оптичних каналів зняття показань аналогових лічильників з подальшою передачею зображень їх циферблатів на візуальні засоби спостереження. Ефективність такого підходу може бути збільшена шляхом поєднання технологій IoT та оптичного розпізнавання даних на основі штучного інтелекту. При цьому необхідно подолати перешкоди, пов'язані з відсутністю уніфікації за типами лічильників, впливом вібрацій для конкретних технологічних процесів, високими вимогами до просторової стабільності зображень та параметрів експозиції. Крім того, використання обмежених можливостей апаратних засобів edge computing змушує підлаштовуватись під мінімальні обчислювальні ресурси.

Одним з варіантів вирішення зазначених проблем є використання процедури попередньої сегментації зображень циферблатів до того моменту, як вони надійдуть на оптичне розпізнавання. Це дозволить знизити обсяг оптичної інформації, що передається, видалити з отриманих зображень фон і звузити поле для розпізнавання.

Зв'язок роботи з науковими програмами, темами. Робота відповідає дослідженням в рамках науково-дослідної роботи «Управління стратегією інноваційного розвитку підприємств в контексті підвищення їх конкурентоспроможності на аграрному ринку, сталого розвитку та забезпечення продовольчої безпеки держави» (2021 р.), що фінансувалась

господарськими договорами із замовниками, Концепції розвитку штучного інтелекту в Україні (розпорядження Кабінету Міністрів України № 1787-р від 29.12.2021), тематиці досліджень Навчально-дослідної лабораторії інтелектуальних систем, комп’ютерних мереж та інтернет речей Кафедри інформаційних систем та технологій Полтавського державного аграрного університету.

Метою кваліфікаційної роботи є підвищення ефективності ПoT за рахунок використання штучного інтелекту на основі нейронних мереж з сегментацією зображень.

Завданнями кваліфікаційної роботи є:

- аналіз варіантів вирішення завдання зняття станів та показань з аналогових лічильників та датчиків існуючої інфраструктури при реалізації ПoT та концепції «Індустрія 4.0»;
- аналіз особливостей існуючих підходів щодо виконання сегментації зображень циферблاتів аналогових лічильників;
- розроблення моделей нейронних мереж сегментації зображень циферблатів аналогових лічильників;
- визначення точності запропонованих моделей нейронних мереж;
- формування рекомендацій щодо використання синтезованих архітектур нейронних мереж.

Об’єктом дослідження є процес сегментації зображень за допомогою штучних нейронних мереж.

Предметом дослідження є точність нейронних мереж сегментації зображень.

Методами дослідження є: аналіз, синтез нейронних мереж сегментації циферблатів аналогових лічильників, вбудовані в фреймворком Keras математичні бібліотеки.

Інформаційна база кваліфікаційної роботи сформована з ресурсів, що містять інформацію про методи сегментації зображень, нейронні мережі та їх

компоненти, що використовуються для виконання семантичної сегментації зображень, а також інструментарій для розробки та дослідження згорткових нейронних мереж.

Елементи наукової новизни роботи полягають в розробці архітектур згорткових нейронних мереж типу PSP, U-Net та U-Net² для завдань сегментації циферблатів аналогових лічильників; порівняльній оцінці застосування різних варіантів реалізації нейронних мереж на базі PSP, U-Net і U-Net².

Практична значущість роботи полягає в розробці рекомендацій щодо формування навчальної бази згорткових нейронних мереж при використанні датасетів з відкритим доступом та обґрунтуванні вибору архітектури нейронної мережі для сегментації циферблатів аналогових лічильників.

Апробація результатів відбувалася в рамках IV Міжнародної науково-практичної конференції, присвяченої 50-й річниці кафедрі Інформаційних систем та технологій «Інтеграція інформаційних систем та інтелектуальних технологій в умовах трансформації інформаційного суспільства» (жовтень 2021 р., м. Полтава), Щорічної студентської наукової конференції Полтавського державного аграрного університету (листопад 2022 р., м. Полтава) та Modern Machine Learning Technology & Data Science (MoMLeT&DS 2022, листопад 2022 р., м. Лейден).

За результатами досліджень здійснено 3 публікації тез доповідей.

Структура кваліфікаційної роботи логічно пов'язана з завданнями досліджень і містить вступ, три розділи основної частини, висновки, список використаних джерел, додатки. Загальний обсяг пояснлюальної записки кваліфікаційної роботи складає 70 сторінок формату А4. Вона містить 39 рисунків і 3 таблиці.

РОЗДІЛ 1

АНАЛІЗ ОСОБЛИВОСТЕЙ РЕАЛІЗАЦІЇ КОНЦЕПЦІЙ ЦИФРОВИЗАЦІЇ ПРОМИСЛОВОГО СЕКТОРУ

1.1 Загальні відомості про промисловий інтернет речей

Цифровізація промислового сектора стає ключовим чинником всебічного розвитку країни з низки причин: підвищується конкурентоспроможність підприємств та якість продукції за рахунок оптимізації техпроцесів; значно знижуються загальні витрати на обслуговування дорогого технологічного обладнання завдяки віддаленому моніторингу, прогнозу відмов, проведення своєчасних планових та позапланових ремонтних робіт; з'являється можливість отримувати та обробляти велику кількість даних з кожного етапу виробництва у режимі реального часу. При цьому цифрові технології передачі мають низку очевидних переваг у порівнянні з аналоговими.

1. Перешкодостійкий зв'язок. Доставку цифрового сигналу «0» і «1» можна гарантувати і за необхідності легко відновити, практично виключивши втрату корисних даних, на відміну від аналогового способу.

2. Мікромініатюризація обладнання. Характеризується зменшенням габаритів, підвищеннем економічності, надійності та енергоефективності апаратури при використанні дискретної логіки та мікросхем.

3. Стандартизація організаційно-технологічних методів та функціональних вузлів. Досягається рахунок застосування уніфікованої цифрової елементної бази, легкої взаємозамінності всіх компонентів і простоти обслуговування.

Застосування ІoT у промисловості створює нові можливості для розвитку виробництва та вирішує множину ключових завдань:

- підвищує продуктивність обладнання;
- знижує матеріальні та енергетичні витрати;

- підвищує якість продукції, що випускається;
- оптимізує та покращує умови праці;
- підвищує рентабельність виробництва та конкурентоспроможність на внутрішньому та світовому ринках.

ПоТ включає кілька скорелюваних компонентів, що забезпечують ефективну цифрову трансформацію технологічних процесів підприємства.

1. Датчики та сенсори. Це компактні енергонезалежні передавальні пристрої, які встановлюються на об'єктах і з певною частотою надсилають дані бездротовим каналом зв'язку.

2. Канали зв'язку. Зв'язують датчики та сенсори із системою зберігання, обробки та аналізу даних. Такий канал зв'язку, зазвичай, забезпечується різними бездротовими технологіями, наприклад, LPWAN, Bluetooth, Wi-Fi, GSM, NB-IoT.

3. Системи для зберігання, обробки та аналізу даних. Є кластер серверів в «хмарі» або закритій мережі підприємства. Приймають дані та надсилають користувачам, які використовують систему.

Промисловий інтернет речей (ПоТ) – це насамперед складна багаторівнева система, що дозволяє в автоматизованому режимі збирати, передавати, обробляти та візуалізувати інформацію за допомогою встановлених датчиків, контролерів та спеціалізованого програмного забезпечення на промислових об'єктах або у «хмарі».

З іншого боку, цифровізація визначає концептуальні вимоги до цифрової трансформації інформаційних систем та бізнес-процесів на підприємстві, а промисловий інтернет речей характеризує технологічний рівень збору даних (показів лічильників та датчиків) з подальшим централізованим зберіганням, обробкою та аналізом.

Синергія ПоТ та цифровізації необхідна для успішної цифрової трансформації підприємства і ці два поняття мають працювати разом (рис. 1.1). Повна заміна аналогових лічильників на цифрові є вигідним у довгостроковій перспективі варіантом розвитку та реалізації політики

цифровізації на підприємстві. Для розвитку будь-якого підприємства важливо замислюватися про заміну обладнання, яке дозволить не лише автоматизувати роботу компанії, а й отримувати точніші показники та сигнали.

Однак такий шлях на початку може підвищувати тимчасові та фінансові витрати. Тобто, одним з обмежень реалізації концепцій «Розумний дім», «Розумне місто», «Індустрія 4.0», «Інтернет речей», «Сільське господарство 4.0» тощо є необхідність інтеграції аналогового обліку електроенергії, наприклад, лічильників води, газу, електроенергії. Часто їхня заміна цифровими пристроями нерентабельна [1].

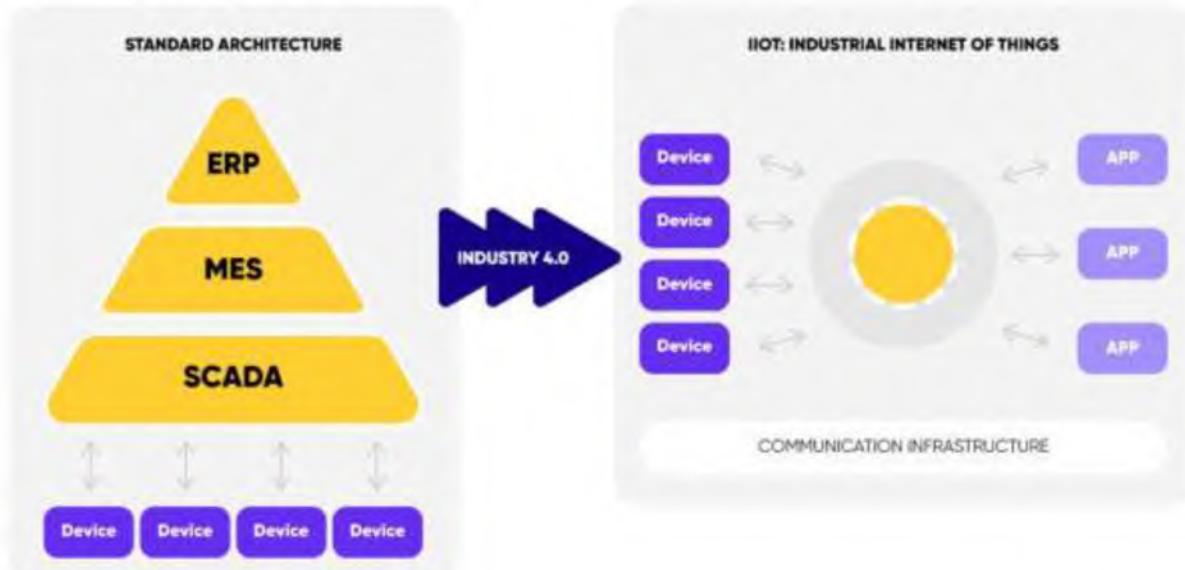


Рисунок 1.1 – Цифрова трансформація за допомогою ПоТ у межах переходу до Індустрії 4.0

Це може бути пов'язано із забороною на внесення змін до комунікації, дорожнечу розробки конструкторської та технічної документації, великою кількістю аналогових засобів обліку на підприємстві тощо. Як наслідок, можливо використовувати оптичний канал оцифрування показань на основі операції розпізнавання. Відбувається своєрідна трансформація аналогових лічильників у цифрові. Як приклад такого підходу слід згадати рішення [2] та [3]. Але такий підхід має певні особливості.

1.2 Особливості інтеграції аналогових засобів обліку у сучасну інфраструктуру підприємства

Більшість підприємств рано чи пізно стикаються із проблемою застарілого промислового обладнання, лічильників та датчиків, які хоч і виконують поставлені завдання, але не мають цифрових інтерфейсів зв'язку та автоматичного віддаленого моніторингу. Такі обмеження не дозволяють їм «влитися» в екосистему ІоТ, що реалізується, а значить цифровізація цього технологічного процесу не може бути досягнута. Якщо говорити лише про один промисловий об'єкт та невелику кількість аналогових лічильників, то проблема не видається глобальною. Але якщо це великий холдинг або концерн, де промислових майданчиків і об'єктів безліч, то перехід до Індустрії 4.0 може бути дуже ускладнений і вибір правильного шляху рішення на етапі проектування зіграє в майбутньому ключову роль.

Таку проблему зняття станів та показань з аналогових лічильників та датчиків вирішити можливо, і для цього є два основні шляхи.

Перший шлях – це заміна аналогових лічильників та датчиків на їх цифрові аналоги з вже вбудованими цифровими інтерфейсами для провідних або бездротових мереж, таких як Ethernet, GSM, LPWAN, NB-IoT, або зі спеціальними провідними інтерфейсами типу RS-282/485 або імпульсний вихід (ІВ) з подальшим підключенням до спеціалізованого контролера. Деякі аналогові лічильники вже мають такі виводи, зазвичай прості імпульсні, тоді їх заміна не потрібна і вони можуть бути інтегровані в цифрову екосистему підприємства. Як правило, такі системи поставляються в комплексі, де виробник надає додаткові послуги з передачі даних у хмарне сховище, а іноді й аналізу зібраних даних.

Другий шлях – це модернізація існуючої схеми моніторингу без заміни лічильників і датчиків. За такого підходу можна назвати кілька автоматизованих технологічних способів зняття показань.

1. Оптичне розпізнавання показання з циферблату або індикації на лічильнику та датчику.
2. Підключення додаткових датчиків поверх існуючого лічильника чи датчика.
3. Паралельне зняття показань цифровим лічильником чи датчиком без приєднання до контролюваного лінії чи каналу, тобто безконтактний спосіб.
4. Паралельне зняття показань цифровим лічильником чи датчиком з безпосереднім приєднанням до контролюваної лінії чи каналу.

Технології та обладнання при виборі первого шляху

Оптичне розпізнавання показань має на увазі фотографування циферблата лічильника і розпізнавання значень прямо на пристрой, причому вбудований пристрій ПЗ «розуміє» як цифри, так і більш нетривіальні елементи, наприклад, шкалу або колірний індикатор. Або це підключення додаткових датчиків шляхом встановлення на крильчатку лічильника спеціального фототранзистора, інфрачервоного відбивача, геркона чи датчика Хола. Такі способи допоможуть знімати дані з мінімальним втручанням у існуючий виробничий процес, а головне без його зупинки або зупинки його моніторингу.

Тому вони такі популярні серед приватних ентузіастів, які створюють власні експериментальні електронні пристрої на базі популярних програмно-апаратних платформ ESP [4], Arduino [5] та Raspberry PI [6] (рис. 1.2).



Рисунок 1.2 – Приклад оптичного зняття показань лічильника води [7]

Ці платформи дозволяють зібрати в одному пристрої необхідний датчик для отримання даних та потрібний бездротовий інтерфейс передачі даних ПoT – LTE, LPWAN, NB-IoT та ін. Деякі такі проекти вилилися у комерційні продукти. Світовий ринок пропонує досить великий вибір таких пристройів (рис. 1.3, 1.4) [8, 9].



Рисунок 1.3 – Оптичні зчитувачі показань лічильників к. Joymeter і Xemtec

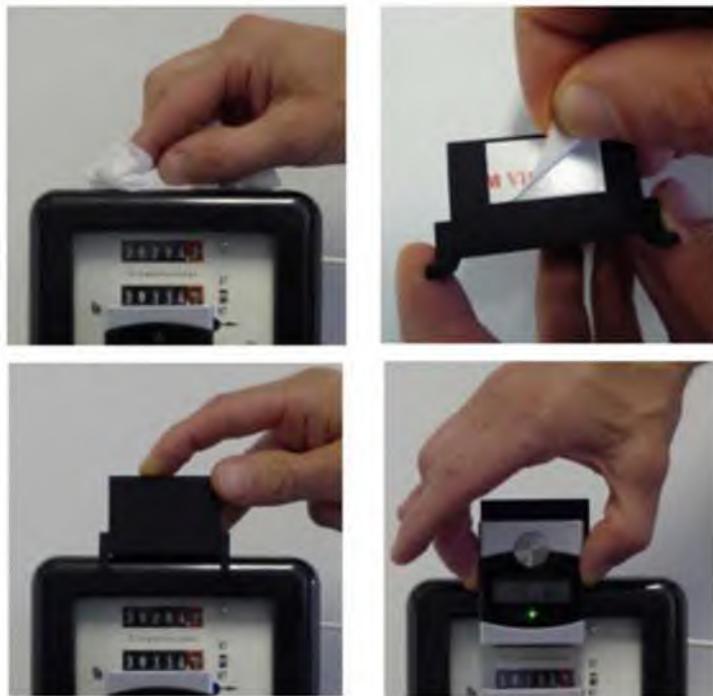


Рисунок 1.4 – Універсальний оптичний зчитувач від к. Q-loud

На жаль, подібні пристрої іноземного виробництва мають досить високу вартість, і це без урахування концентраторів та решти периферії, а також

можливі складнощі у постачанні обладнання та забезпечення відповідного гарантійного та постгарантійного обслуговування.

Іноді такі зчитувачі мають свій обмежений за дальністю дії бездротовий протокол передач даних, і для їх інтеграції в ІoT-екосистему підприємства знадобиться організація «останньої милі» (рис. 1.5) [10], встановлення спеціальних концентраторів/маршрутизаторів для конвертації протоколу даних зі зчитувачів у традиційні та звичні канали зв'язку – Ethernet, LTE тощо.

Часто це призводить до помітного подорожчання рішення загалом.

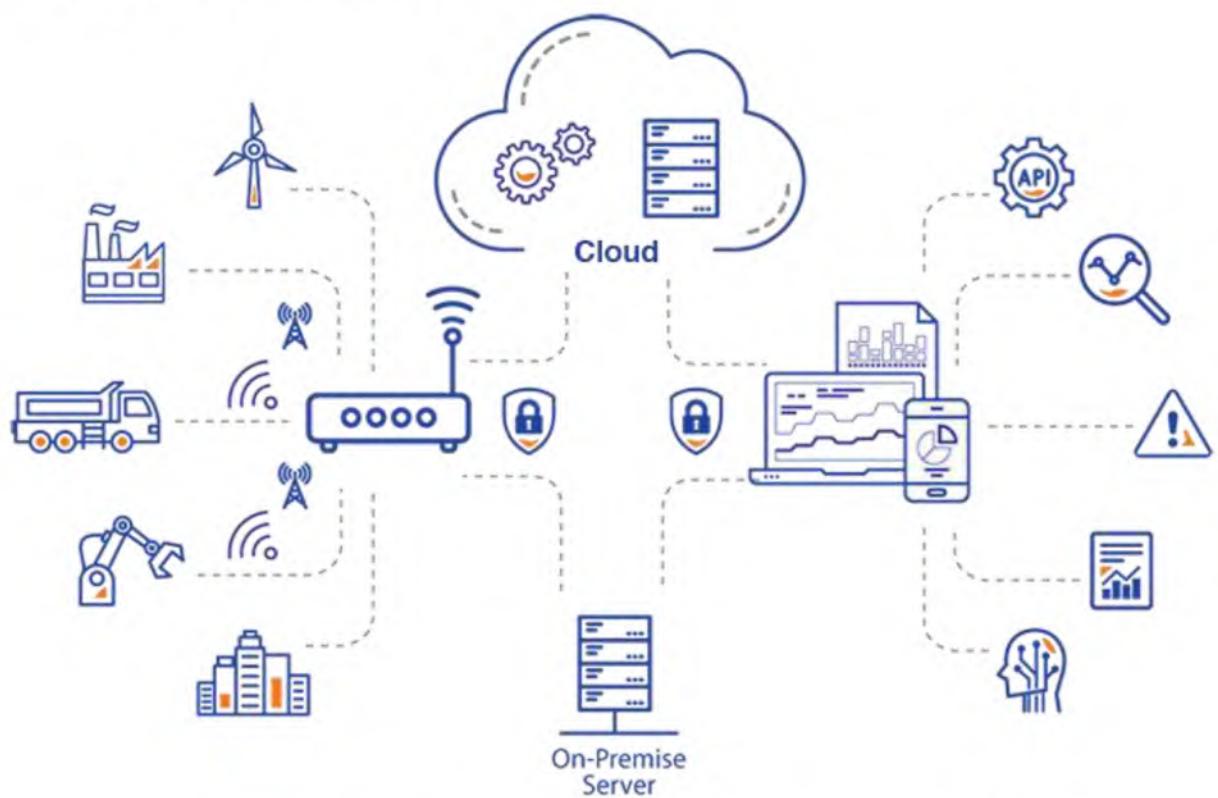


Рисунок 1.5 – Класична схема побудови рішення IoT

Інший, більш сучасний, але поки що менш поширений спосіб – безконтактний. Це паралельне зняття показань та станів додатковими лічильниками та датчиками без приєднання до контролюваної лінії чи каналу. У цьому випадку лічильник або датчик встановлюється прямо на лінії без зупинки виробничого процесу або зупинки його моніторингу.

Такі пристрой, як правило, мають меншу вартість порівняно з оптичними читувачами та більшу автономність. Також вони не вимагають встановлення додаткових концентраторів, оскільки спілкуються з хмарним сервером безпосередньо через спеціальні дальнобійні протоколи ІoT, наприклад, LPWAN [11] або NB-IoT [12]. Однак вони обмежені контролюваним середовищем. Наприклад, виміряти обсяги рідин таким лічильником не вийде, а проконтролювати за допомогою датчика наявність напруги в лінії електро живлення або вважати лічильником кількість витраченої електроенергії цілком вдастся. Але варто враховувати, що отримані дані будуть менш точними, ніж під час використання класичних електролічильників. Останній спосіб, який має на увазі безпосереднє підключення цифрових лічильників і датчиків до контролюваної лінії або каналу, втрачає сенс, оскільки він має аналогічний ефект, як і заміна аналогових на цифрові.

1.3 Шляхи вирішення проблеми оцифровки показань з аналогових лічильників

Заміна аналогових лічильників та датчиків на їх цифрові аналоги з дротовими або бездротовими інтерфейсами має кілька переваг:

- великий вибір готових рішень на ринку;
- можливість використання екосистеми виробника в частині передачі та зберігання даних;
- перспективність та далекоглядність застосування нових технологій;
- масштабованість та стандартизація на всіх об'єктах підприємства;
- модернізацію можна провести безболісно, поєднавши з черговим етапом обслуговування чи повірки існуючі.

Однак капітальна перебудова інфраструктури спричиняє вимущену зупинку виробничого процесу на час модернізації. Крім того модернізація лічильників та датчиків має досить високу вартість.

З іншого боку модернізація існуючої схеми моніторингу без заміни існуючих лічильників та датчиків може проводитись без зупинки виробничого процесу чи його моніторингу. При цьому вартість модернізації нижча, ніж заміна на цифрові. З іншого боку такий підхід має наступні особливості:

- невисока надійність через технічну складність та застосування проміжних пристройів;
- трудомісткість та підвищена вартість подальшої технологічної підтримки.

Одним із варіантів подолання цього бар'єру може бути поєднання технологій штучного інтелекту та Інтернету речей (AI+IoT). Такий варіант передбачає використання процедури сегментації зображення перед розпізнаванням. Технічна реалізація цього підходу можлива завдяки технологіям туманних обчислень.

1.4 Сутність семантичної сегментації зображень

На сьогоднішній день для виконання сегментації використовується кілька підходів, розглянутих, наприклад, [13, 14]. Серед них слід виділити наступні.

Семантична сегментація зображення означає присвоєння кожному пікселю певної мітки. У цьому полягає головна відмінність від класифікації, де всьому зображенню ставиться у відповідність лише одна мітка. Сегментація працює з множиною об'єктів одного класу як із єдиним цілим.

Інстанс-сегментація опрацьовує кілька об'єктів одного класу як різні об'єкти [15]. Зазвичай інстанс-сегментація складніша, ніж семантична сегментація (рис. 1.6).



Semantic Segmentation



Instance Segmentation

Рисунок 1.6 – Порівняння інстанс- та семантичної сегментації

До настання епохи глибокого навчання для сегментації застосовувалися найрізноманітніші техніки обробки зображень залежно від галузі інтересів. Деякі популярні методи на той час представлені тут.

Сегментація на основі градації сірого. Найбільш простий спосіб семантичної сегментації полягає в ручному кодуванні правил або властивостей, яким повинна задовольняти область, щоб можна було пристрати певну мітку. Ці правила можуть бути оформлені як властивості пікселів, наприклад інтенсивність сірого кольору. Один із методів, що використовують таку техніку, називається алгоритмом поділу та об'єднання (Split and Merge). Цей алгоритм рекурсивно поділяє зображення на підобласті до тих пір, поки не припішеться певна мітка, потім суміжні підобласті з однаковими мітками об'єднуються. Проблема, пов'язана з цим методом, полягає в тому, що правила мають бути прописані в коді вручну. Однак часом буває неймовірно складно описати складні класи, такі як «людина», лише за допомогою інформації про інтенсивність сірого кольору. Отже, у роботі з такими складними об'єктами для правильного навчання уявленням необхідні методи отримання ознак та оптимізації.

Умовні випадкові поля (Conditional Random Fields, CRF). Розглянемо сегментацію зображення через навчання моделі приписувати клас кожному пікселю – рис. 1.7.

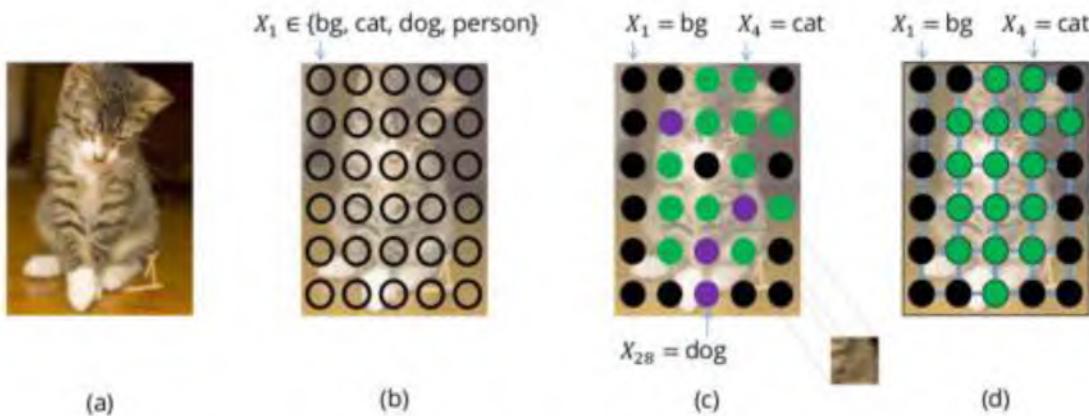


Рисунок 1.7 – Умовні випадкові поля

У випадку, якщо модель не ідеальна, можна отримати результати із зашумленою сегментацією, що найчастіше неможливо в природі. Наприклад, пікселі котиків поєднуються з пікселями собак, як показано на зображенні. Це можна уникнути, якщо розглянути попередні взаємозв'язки між пікселями. Ідея полягає в тому, що об'єкти безперервні і, отже, прилеглі пікселі повинні мати ті самі мітки. Для моделювання такого взаємозв'язку використовують умовні випадкові поля. Ці поля відносяться до класу методів статистичного моделювання, що використовуються для структурованих прогнозів. На відміну від класифікаторів перед передбаченням Поля беруть до уваги контекст, тобто взаємозв'язок між пікселями. Тому модель вважається ідеальним кандидатом для семантичної сегментації зображення (рис. 1.8). Кожен піксель зображення асоціюється із кінцевим набором можливих станів. У даному випадку, цільові мітки будуть набором можливих станів. Витрати на присвоєння стану єдиному пікселю мають назву унарних витрат. Можна розглядати пари пікселів, які є безпосередніми сусідами (Grid CRF), або все працювати з усімаарами пікселів зображення (Dense CRF).

- Grid CRF leads to over smoothing around boundaries
- Dense CRF is able to recover fine boundaries

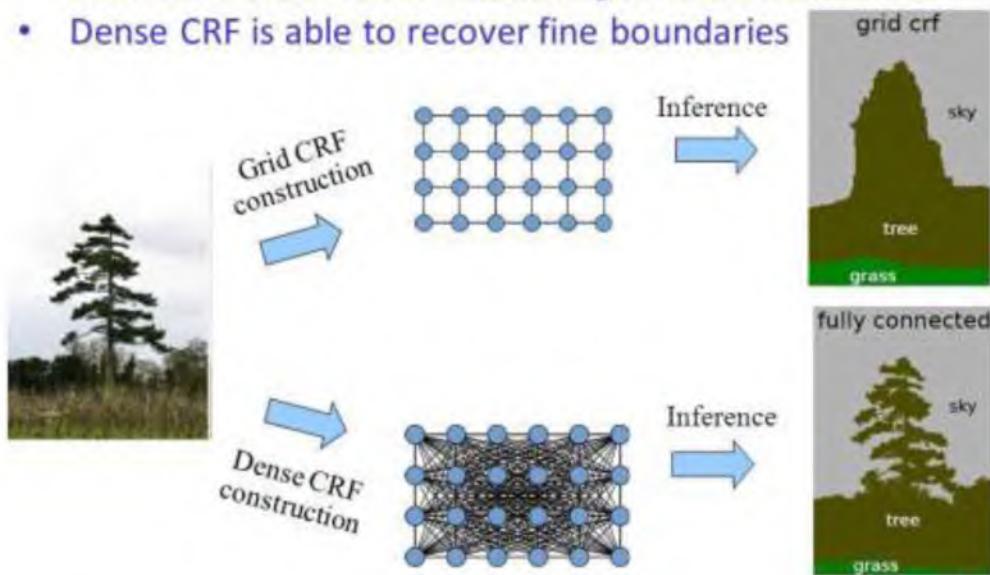


Рисунок 1.8 – Порівняння сіткових (Grid) та щільних (Dense) CRF

Але з годом глибоке навчання нейронних мереж багато в чому спростило виконання семантичної сегментації, водночас показавши вражаючу якість.

Висновки до розділу 1

В сучасних умовах багато підприємств стикаються з проблемою застарілого промислового обладнання, лічильників, датчиків та ін. Їх основним недоліком слід вважати відсутність механізмів цифрових інтерфейсів зв'язку та автоматичного віддаленого моніторингу. Це обмеження їх інтеграцію в екосистему ІoT.

Для подолання цього явища в роботі проаналізовано шляхи інтеграції аналогового обладнання у сучасну цифрову інфраструктуру:

- заміна аналогових лічильників та датчиків на їх цифрові аналоги з вже вбудованими цифровими інтерфейсами;
- модернізація існуючої схеми моніторингу без заміни лічильників і датчиків.

В ході проведених досліджень запропоновано реалізацію другого шляху на основі застосування оптичних каналів, які можливо реалізувати використовуючи семантичну сегментацію зображень.

Проведений аналіз свідчить про необхідність розробки інструментарію для розширення функціоналу системи обліку енергоресурсів на основі нейронних мереж.

РОЗДІЛ 2

ІНСТРУМЕНТАРІЙ ДЛЯ РОЗШИРЕННЯ ФУНКЦІОНАЛУ СИСТЕМИ ОБЛІКУ ЕНЕРГОРЕСУРСІВ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

2.1 Методи глибокого навчання

Як відомо, виявлення суттєвих об'єктів (Salient Object Detection, SOD) спрямовано на сегментацію найбільш візуально привабливих об'єктів на зображенні [16]. Глибоке навчання багато в чому спростило виконання семантичної сегментації, водночас показавши вражуючу якість. Надалі доцільно дослідити архітектури моделей та функцій втрат, які використовуються для навчання цих моделей глибокого навчання. SOD широко використовується в багатьох областях, таких як візуальне відстеження та сегментація зображень. Останнім часом, з розвитком глибоких згорткових нейронних мереж (CNN), особливо з появою повністю згорткових мереж (FCN) [17] у сегментації зображень, виявлення помітних об'єктів значно покращилося.

Існує загальна схема проектування більшості мереж SOD [18], тобто вони зосереджені на ефективному використанні глибоких функцій, які витягають існуючі магістральні мережі, такі як Alexnet [19], VGG [20], ResNet [21], ResNeXt [22], DenseNet [23] та ін. Однак усі ці магістралі спочатку призначені для класифікації зображень. Вони отримують ознаки, які представляють семантичне значення, а не локальні деталі та інформацію про глобальний контраст, які необхідні для виявлення значущості. І вони повинні були попередньо навчені на даних ImageNet [24], що неефективно, особливо якщо цільові дані слідують іншому розподілу, ніж ImageNet.

Одна з найпростіших та найпопулярніших архітектур, що використовуються для семантичної сегментації, це повнозгорткова мережа (Fully Convolutional Network, FCN). У [25] автори використовують FCN для

початкового перетворення вхідного зображення до меншого розміру (одночасно збільшуючи кількість каналів) через серію згорток. Такий набір згорткових операцій зазвичай називається кодувальником (рис. 2.1). Потім вихід декодується або через білінійну інтерполяцію, або через серію транспонованих згорток, що має назву – декодер.

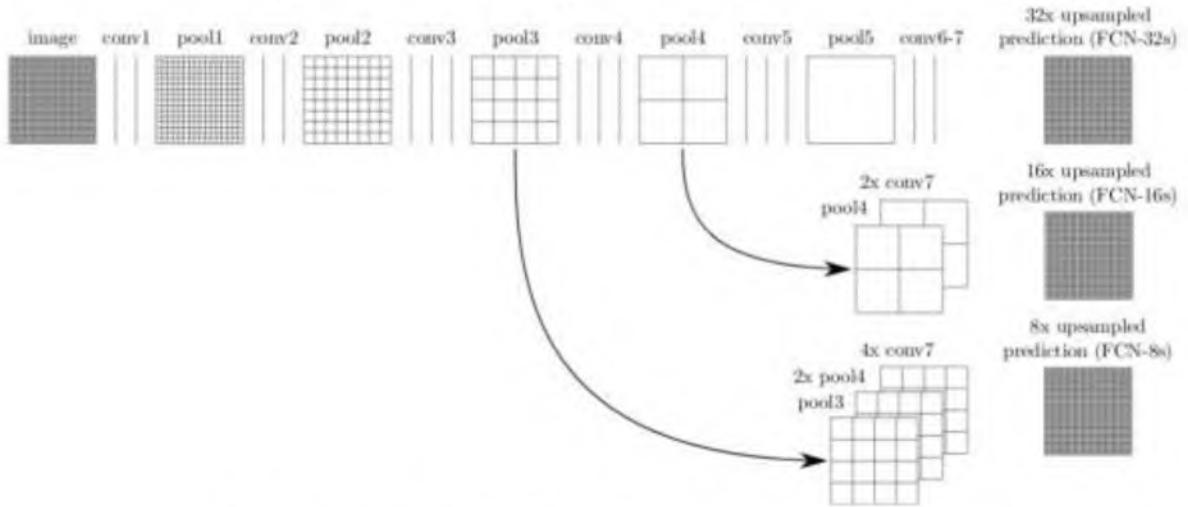


Рисунок 2.1 – Пряме та зворотне перетворення

Така базова архітектура, попри її ефективність, має низку недоліків. Один з яких – наявність артефактів, розташованих у шаховому порядку, пов'язаних із нерівномірним перекриттям виходів в операції транспонованої згортки (рис. 2.2). Інший недолік пов'язаний з низькою роздільною здатністю по краях через втрати інформації в процесі кодування.

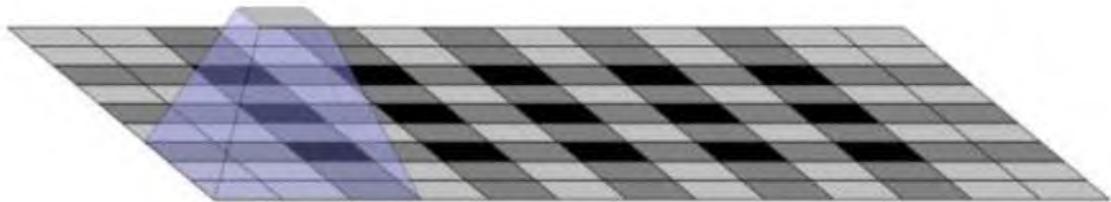


Рисунок 2.2 – Формування артефактів, розташованих у шаховому порядку

Для збільшення якості базової моделі FCN було запропоновано кілька рішень, що довели свою ефективність та які доцільно розглянути.

U-Net [26]. Така мережа є поліпшенням простої архітектури FCN. Мережа skip-зв'язку між виходами з блоків згортки і відповідними входами блоку транспонованої згортки на тому ж рівні наведена на рис. 2.3.

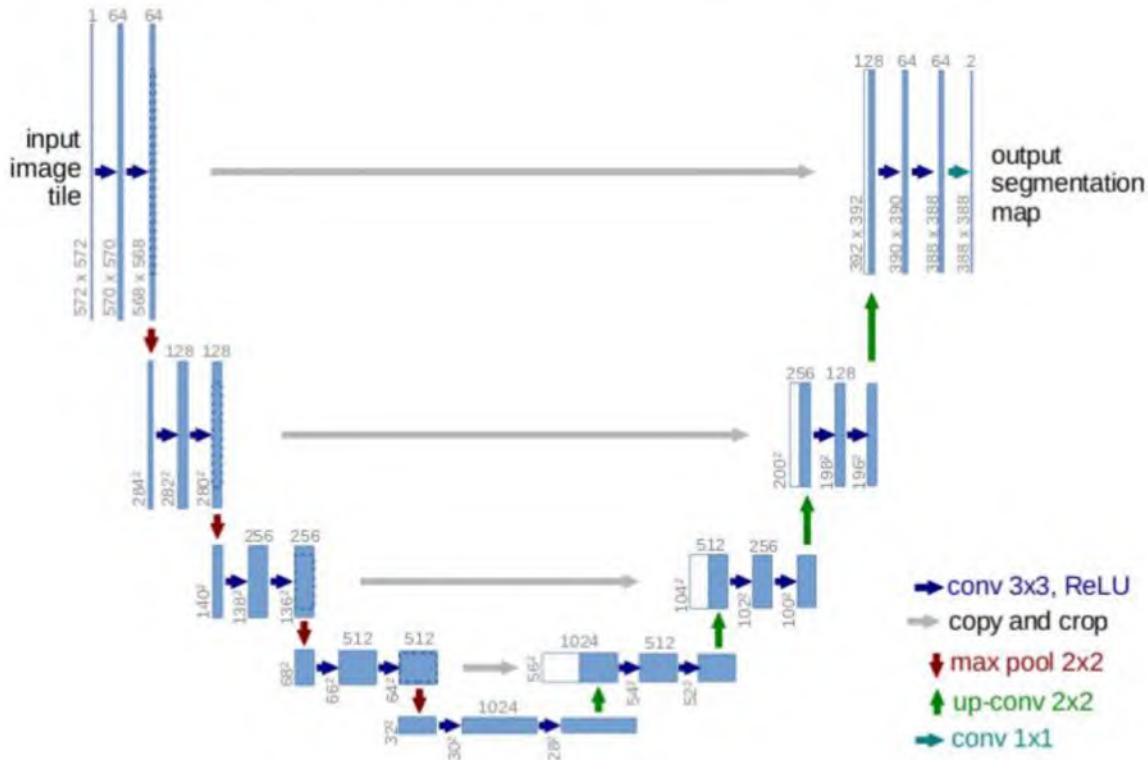


Рисунок 2.3 – Схема U-Net

Skip-зв'язки дозволяють градієнтам краще розповсюджуватись та надавати інформацію з різних масштабів розміру зображення. Інформація з великих масштабів (верхні шари) може допомогти моделі краще класифікувати. У той час, як інформація з менших масштабів (глибокі шари) допомагає моделі краще сегментувати.

Модель «Tiramisu» (рис. 2.4) схожа на U-Net за винятком того, що для прямої та транспонованої згортки тут використовуються щільні блоки, як це реалізовано в DenseNet [27]. Щільні блоки складаються з кількох згорткових шарів. В якості входів на наступні шари використовуються відображення ознак з усіх попередніх шарів. Результатуюча мережа надзвичайно ефективна в сенсі параметрів і може краще працювати з ознаками зі старих шарів.

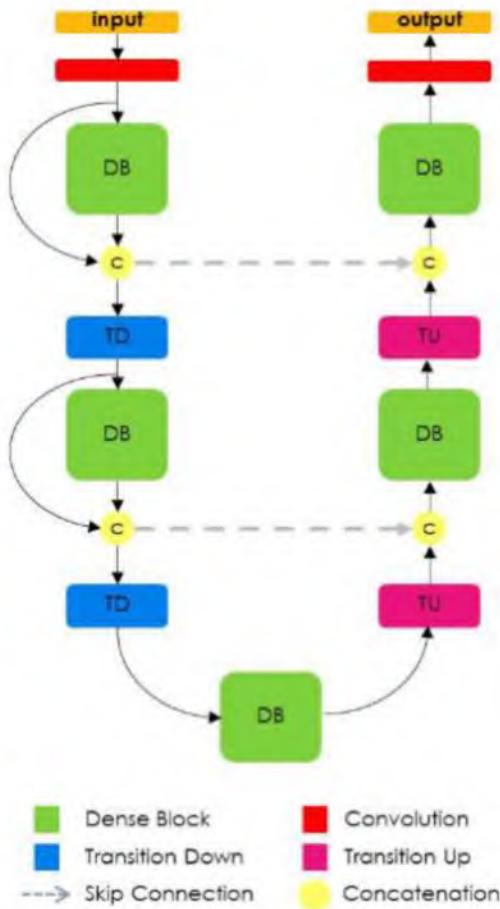


Рисунок 2.4 – Мережа Tiramisu

Недолік такого методу, який є наслідком самої природи операцій об'єднання у кількох фреймворках машинного навчання, полягає у низькій ефективності роботи з пам'яттю. Тому для роботи з такою архітектурою потрібні потужні кластери GPU.

Багатомасштабний метод. Деякі моделі глибокого навчання явно представляють методи роботи з інформацією з різних масштабів. Наприклад, мережа Pyramid Scene Parsing (PSPNet) виконує операцію об'єднання (через функцію Max або Average) за допомогою ядер різного розміру та з різними кроками (рис. 2.5), застосованими до вихідних відображенень ознак з згортковою нейронною мережею (наприклад, ResNet). Далі за допомогою білінійної інтерполяції відбувається перерахунок розміру всіх виходів з Pooling-шару та вихідного відображення ознак з CNN. Потім модель об'єднує вздовж осі каналу

нові виходи. Для генерації передбачення фінальна згортка вже на об'єднаному виході.

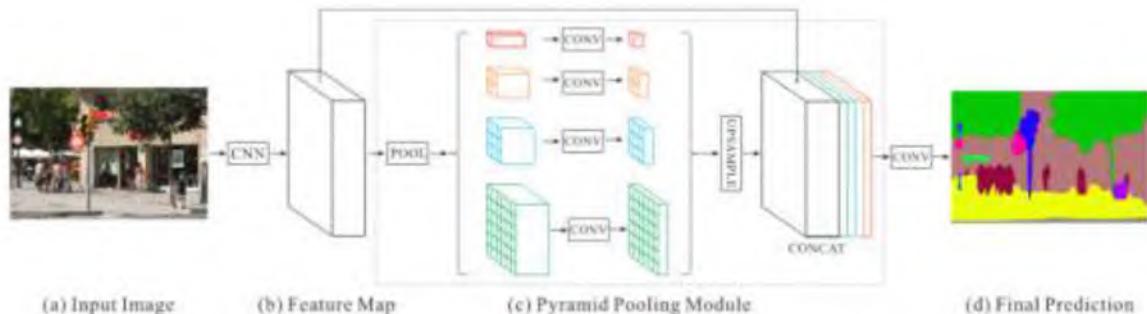


Рисунок 2.5 – Архітектура PSPNet

Модель розширених згорток (Atrous Convolutions) є ефективним способом комбінувати ознаки з кількох масштабів без значного збільшення кількості параметрів. Регулюючи показник розширення, той самий фільтр розподіляє значення ваги далі у просторі (рис. 2.6). Це дозволяє вивчати більш загальний контекст.

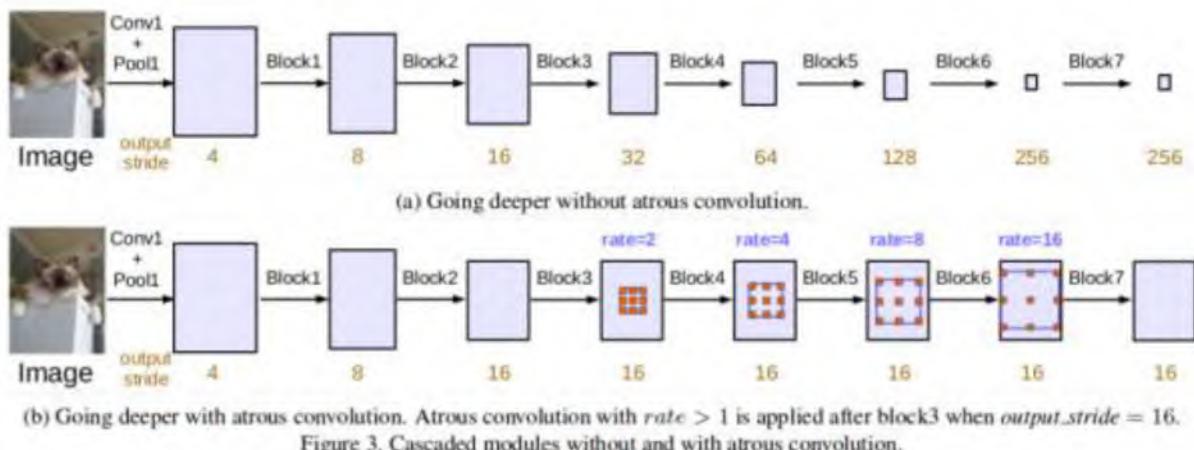


Рисунок 2.6 – Каскадний режим моделі розширених згорток

У мережі DeepLabv3 використовується модель розширених згорток з різними показниками розширення. Це дозволяє використовувати інформацію з різних масштабів без істотних втрат у розмірі зображення. Автори проводили експерименти з Atrous convolutions у режимі каскаду, а також паралельному режимі у формі Atrous Spatial Pyramid Pooling (рис. 2.7).

Гібридні методи. Деякі методи використовують згорткову нейронну мережу, щоб отримувати ознаки, які потім використовуються як входи унарних витрат у щільні CRF (рис. 2.8). Такий гібридний метод показує хороші результати через здатність CRF моделювати зв'язок між пікселями.

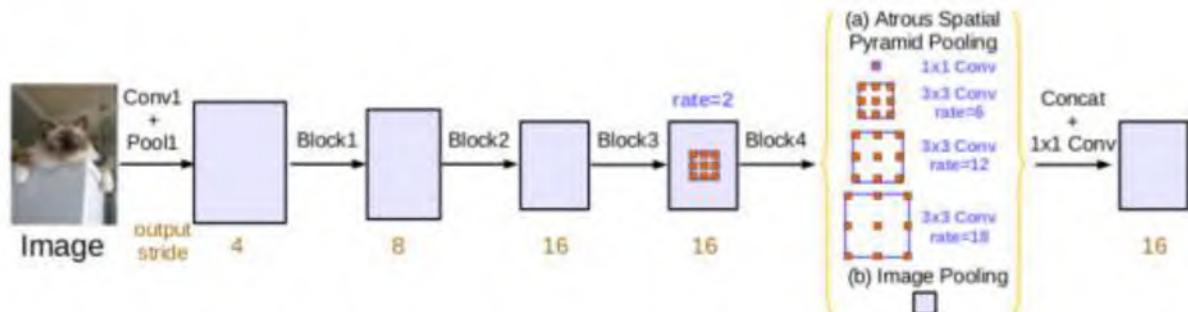


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

Рисунок 2.7 – Паралельний режим моделі розширеніх згорток

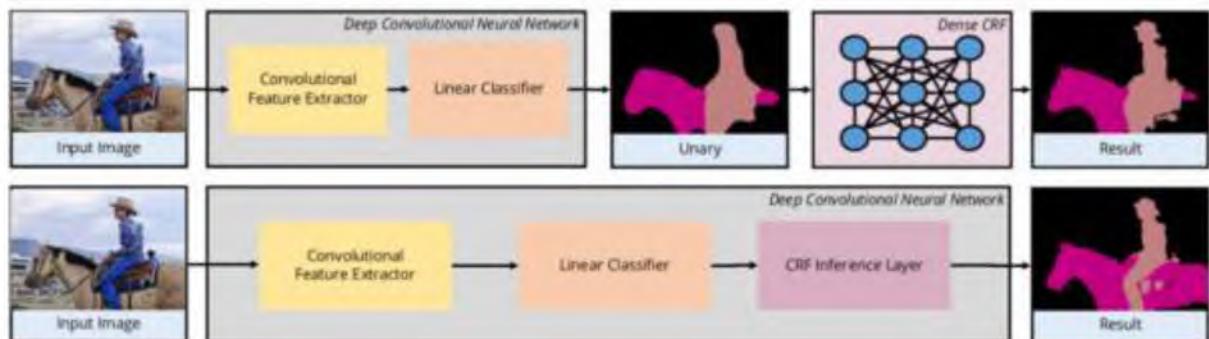


Рисунок 2.8 – Робота гібридного методу CRF-CNN

Деякі методи використовують CRF у самій нейронній мережі, як це зроблено (наприклад, CRF-as-RNN, де щільні CRF моделюються як рекурентні нейронні мережі). Це дозволяє здійснювати навчання від початку до кінця.

2.2 Багаторівнева глибока інтеграція функцій

Існує кілька проблем із мережними архітектурами для SOD. По-перше, вони часто надмірно складні [28]. Частково це пов'язано з додатковими модулями агрегації ознак, які додаються до існуючих магістральних мереж для отримання багаторівневих важливих ознак цих магістральних мереж. По-друге, існуючі магістралі зазвичай досягають глибшої архітектури, жертвуючи високим розрізненням карт ознак [28]. Щоб запустити ці глибокі моделі з доступною пам'яттю та обчислювальними витратами, карти ознак на ранніх етапах масштабуються до нижчого розрізнення. Наприклад, на ранніх рівнях як ResNet, так і DenseNet [23] згортка з кроком, рівним двом, за яким слідує максимальне об'єднання з кроком, рівним двом, використовується для зменшення розміру карт об'єктів до однієї чверті вхідних карт. Однак висока роздільна здатність також відіграє важливу роль у сегментації крім глибокої архітектури.

Останніми роками було запропоновано багато мереж виявлення глибоких помітних об'єктів [29]. У порівнянні з традиційними методами [30], заснованими на ручних функціях, таких як узгодженість переднього плану, гіперспектральна інформація, подібність суперпікселів, гістограми та ін., мережі виявлення глибоких помітних об'єктів показують більш конкурентоспроможні результати.

Для SOD розроблялась множина стратегій та методів інтеграції та агрегування багаторівневих глибоких функцій. В [18] запропоновано подавати фрагмент зображення навколо цільового пікселя в мережу, а потім отримувати вектор ознак для опису помітності цього пікселя. В [31] автори пророкують карти важливості, поєднуючи багаторівневі ознаки в різні розрізнення. В UCF запропоновано зменшити артефакти шахової дошки операторів шляхом введення переформульованого відсіву та гіbridного модуля шляхом підвищення дискретизації [32]. В [33] розробили мережу виявлення значущості (NLDF+) з архітектурою сітки 4x5, в якій більш глибокі функції поступово

інтегруються з дрібнішими функціями. В LFR пророкують карти значущості, витягуючи ознаки як з вихідних вхідних зображень, так і з їх відображеніх зображень із спорідненою архітектурою. В DSS+ [13] пропонується інтеграція багаторівневих ознак, вводячи короткі з'єднання від глибоких шарів до поверхневих шарів. В RAS пророкують і уточнюють карти помітності, ітеративно використовуючи значущість побічних результатів магістральної мережі як посібник з уваги до функцій. В BMPM пропонують інтегрувати ознаки з дрібних та глибоких шарів за допомогою контролльованої 2-направленої стратегії проходження. В R3Net+ поперемінно включаються особливості поверхневих і глибоких шарів для уточнення передбачених карт значимості. В RADF+ пропонується виявляти суттєві об'єкти шляхом рекурентного агрегування багаторівневих глибоких ознак. В MLMS підвищується точність визначення важливості, розробивши новий модуль взаємного навчання для кращого використання кореляції кордонів та областей. В [34] пропонують використовувати структуру каскадного часткового декодера (CPD) для швидкого і точного виявлення істотних об'єктів. Глибокі методи у цій категорії використовують переваги багаторівневих глибоких функцій, які отримують магістральні мережі, і значно піднімають планку виявлення істотних об'єктів порівняно з традиційними методами.

В цілому, потрібна як локальна, так і глобальна інформація. Фільтр 3x3 гарний для отримання локальних ознак на кожному шарі. Однак важко отримати глобальну інформацію, просто збільшивши розмір фільтра, тому що це різко збільшить кількість параметрів та обчислювальні витрати. Однак в багатьох з робіт більше уваги приділяється витягу глобального контексту. В SRM адаптовано модуль об'єднання пірамід для захоплення глобального контексту та пропонується багатоетапний механізм уточнення для уточнення карт значущості. В PAGRN є модуль просторової та канальної уваги для отримання глобальної інформації про кожен шар і запропонували механізм прогресивного управління увагою для уточнення карт значущості. В DGRL розробили модуль контекстного зважування для глобальної локалізації

значущих об'єктів, а потім використовували модуль уточнення кордонів для локального уточнення карти значущості. В PiCANet періодично захоплюють локальну та глобальну контекстуальну увагу з точки зору пікселів і прогнозують карту значущості, поєднуючи її з архітектурою U-Net. В CapSal є модуль локального та глобального сприйняття для вилучення як локальної, так і глобальної інформації з функцій, витягнутих магістральною мережею. В MSWS розробили модуль уваги для прогнозування просторового розподілу об'єктів переднього плану по областях зображення, одночасно поєднуючи їх функції. В AFNet є модуль глобального сприйняття та модулі уважного зворотного зв'язку, щоб краще дослідити структуру помітних об'єктів. В BASNet пропонується модель прогнозування-уточнення шляхом послідовного накладання двох мереж U-Net з різною конфігурацією та гібридною втратою для виявлення суттєвих об'єктів з урахуванням кордонів. В PoolNet розробили архітектуру кодувальника-декодера для виявлення помітних об'єктів, впровадивши модуль глобального керівництва для отримання глобальних функцій локалізації та модуль багатомасштабної агрегації функцій, адаптований з модуля пірамідного пулу для поєднання глобальних та дрібних функцій.

У цих методах пропонується множина модулів. Різноманітні рецептивні поля та багатші багатомасштабні контекстуальні функції, представлені цими новими модулями, значно покращують продуктивність моделей виявлення помітних об'єктів. Методи багаторівневої глибокої інтеграції функцій зосереджені в основному на розробці кращих стратегій агрегації багаторівневих функцій.

З іншого боку, методи категорії багатомасштабного вилучення ознак спрямовані на розробку нових модулів для вилучення як локальної, так і глобальної інформації з ознак, отриманих магістральними мережами. Як видно, майже всі вищезгадані методи намагаються краще використовувати карти ознак, що створені існуючими основами класифікації зображень. Замість того, щоб розробляти та додавати складніші модулі та стратегії для

використання функцій цих магістральних мереж, архітектура U-Net² безпосередньо витягує багатомасштабні функції поетапно для виявлення суттєвих об'єктів. З метою вирішення розглянутої проблеми доцільно використовувати архітектуру U-Net².

2.3 Архітектура U-Net²

U-Net² – це 2-рівнева вкладена U-структурата, яка розроблена для SOD без використання будь-яких попередньо навчених магістральних мереж із класифікації зображень [16]. U-Net² можна навчати з нуля для досягнення конкурентоспроможних результатів. Також дана архітектура дозволяє мережі працювати глибше, досягати високої роздільної здатності без значного збільшення обсягу пам'яті та обчислювальних ресурсів. Це досягається за рахунок вкладеної U-структурти:

- на нижньому рівні використаний блок RSU (ReSidual U-block), який здатний добувати внутрішньо-етапні багатомасштабні ознаки без погіршення розрізnenня карти ознак;
- на верхньому рівні є структура, подібна до U-Net, в якій кожен етап заповнений блоком RSU.

Дворівнева конфігурація призводить до вкладеної U-структурти. U-Net² забезпечує конкурентоспроможну продуктивність у порівнянні з сучасними методами на 6-ти загальнодоступних наборах даних та працює в реальному часі (30 fps, з входним розміром $320 \times 320 \times 3$) на графічному процесорі 1080Ti (рис. 2.9). Щоб полегшити використання U-Net² у обчислювальних середовищах та середовищах з обмеженим об'ємом пам'яті можливо використовувати невелику версію U-Net², яка називається U-Net^{2'} (має розмір 4,7 МБ). Таким чином, U-Net² досягає конкурентоспроможних результатів порівняно з більшістю моделей SOTA (рис. 2.9) за 40 FPS. Mіра maxF обчислюється за набором даних

ECSSD [46]. Червона зірка означає U-Net² (176,3 МБ), а синя зірка означає нашу маленьку версію U-Net² (4,7 МБ).

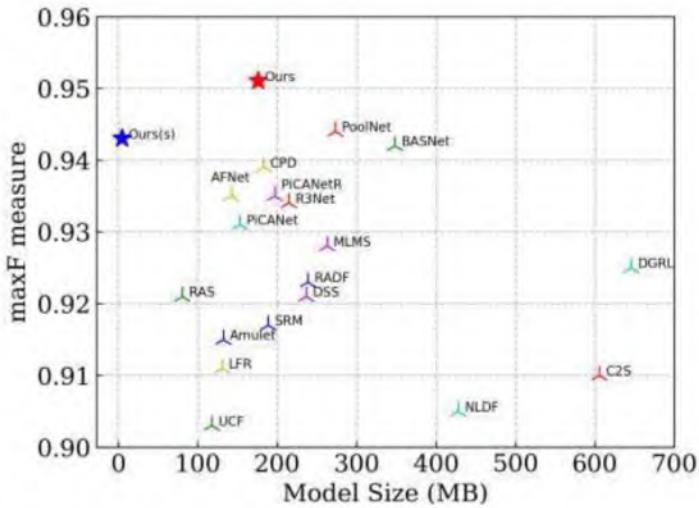
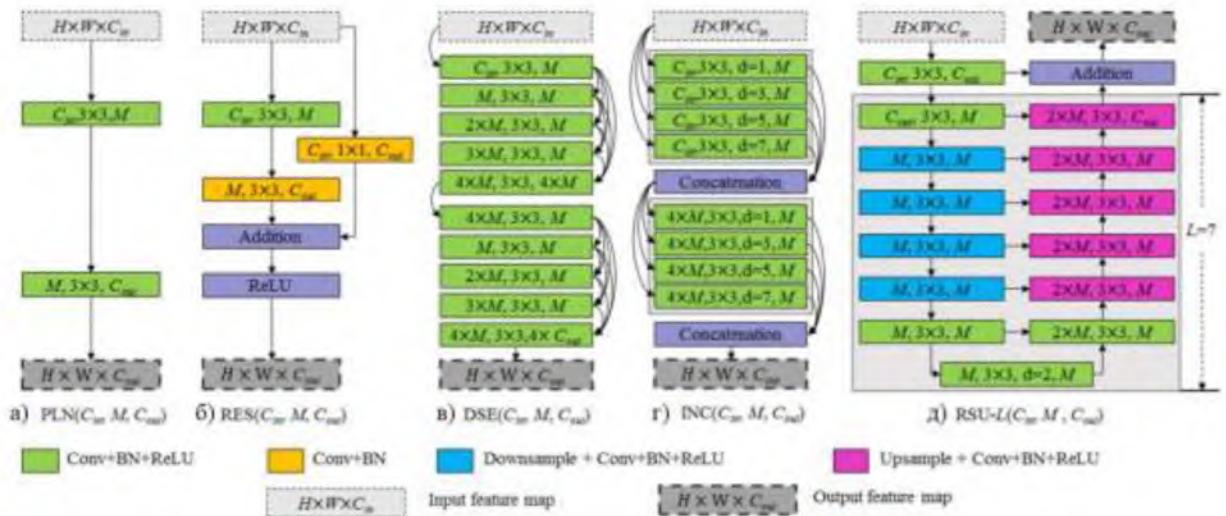


Рисунок 2.9 – Порівняння розміру моделі та продуктивності U-Net² з іншими сучасними моделями SOD

В мережі U-Net² запропонований U-блок (RSU) [16]. Як локальна, і глобальна контекстна інформація дуже важлива виявлення суттєвих об'єктів та інших завдань сегментації. У сучасних проектах CNN, таких як VGG, ResNet, DenseNet та ін., невеликі фільтри згортки розміром 1×1 або 3×3 є найбільш часто використовуваними компонентами для вилучення ознак. На рис. 2.10,а-г показані типові існуючі згорткові блоки з невеликими рецептивними полями. Вони популярні, оскільки вимагають менше місця для зберігання та ефективні у обчислювальному відношенні. Вихідні карти ознак дрібних шарів містять лише локальні ознаки, тому що сприйнятливе поле фільтрів 1×1 або 3×3 надто мало для захоплення глобальної інформації. Щоб отримати більш глобальну інформацію на картах об'єктів з високою роздільною здатністю з неглибоких шарів, найбільш пряма ідея полягає в збільшенні рецептивного поля. На рис. 2.10,д показаний ReSidual U-block (RSU), який намагається витягнути як локальні, так і нелокальні ознаки.



умовні познаки:

- простий згортковий блок PLN;
- залишковий блок RES;
- щільний блок DSE;
- початковий блок INC;
- залишковий U-блок RSU.

Рисунок 2.10 – Ілюстрація існуючих згорткових блоків та залишкового U-блоку RSU

Однак виконання кількох розширеніх згорток на вхідній карті об'єктів (особливо на ранній стадії) з вихідною роздільною здатністю вимагає надто багато обчислень та ресурсів пам'яті. Щоб зменшити витрати на обчислення, в PoolNet адаптівно паралельну конфігурацію з модулів об'єднання пірамід (PPM), які використовують невеликі фільтри ядра на картах ознак зі зниженою дискретизацією, а нерозширені згортки – на картах ознак вихідного розміру. Але злиття функцій різного масштабу шляхом прямого підвищення дискретизації та конкатенації (або додавання) може привести до погіршення характеристик високої роздільної здатності.

Блок RSU призначений для захоплення внутрішньо-етапних багатомасштабних ознак. Структура RSU містить: L – кількість шарів у кодері, C_{in} , C_{out} – вхідні та вихідні канали, M – кількість каналів у внутрішніх шарах RSU. При цьому, RSU складається з 3-ох компонентів.

1. Вхідний згортковий шар, який перетворює вхідну карту об'єктів x ($H \times W \times Cin$) на проміжну карту $F_1(x)$ з каналом Cout. Це простий згортковий шар для отримання локальних ознак.

2. U-Net-подібна симетрична структура кодер-декодер з висотою L , яка приймає карту проміжних ознак $F_1(x)$ як вхідні дані та навчається витягувати та кодувати багатомасштабну контекстну інформацію $U(F_1(x))$. U являє собою структуру, подібну до U-Net, як показано на рис. 2.10,д. Більше L призводить до більш глибокого залишкового U-блоку (RSU), більшої кількості операцій об'єднання, більшого діапазону рецептивних полів та більш багатьох локальних та глобальних функцій. Налаштування цього параметра дозволяє витягувати багатомасштабні об'єкти з вхідних карток об'єктів з довільною просторовою роздільною здатністю. Багатомасштабні ознаки витягаються з карт ознак з поступовим зниженням дискретизації і кодуються в карти ознак з високою роздільною здатністю за допомогою прогресивної дискретизації, що підвищує, конкатенації і згортки. Цей процес зменшує втрату дрібних деталей, що спричинена прямим підвищенням дискретизації з великими масштабами.

3. Залишкове з'єднання, яке поєднує локальні та багатомасштабні ознаки шляхом підсумовування: $F_1(x) + U(F_1(x))$.

Накладні витрати на обчислення через U-структурну невеликі, оскільки більшість операцій застосовується до карт об'єктів зі зниженою дискретизацією (рис. 2.11). В даному випадку, розраховуються витрати на обчислення на основі перенесення вхідної картки об'єктів розміром $320 \times 320 \times 3$ у вихідну карту об'єктів $320 \times 320 \times 64$.

Стекування кількох структур, подібних до U-Net, було запропоновано для різних завдань, наприклад: DocUNet, CU-Netta ін. Ці методи послідовно накладають структури U-Net для побудови каскадних моделей та можуть бути узагальнені як « $(U \times n\text{-Net})$ », де n – кількість модулів U-Net, що повторюються. Однак проблема полягає в тому в тому, що витрати на обчислення та пам'ять збільшуються на n .

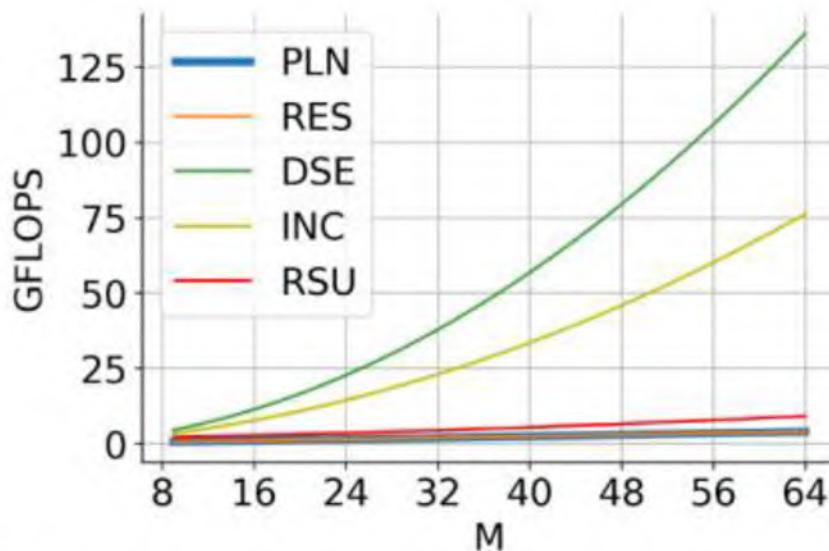


Рисунок 2.11 – Витрати на обчислення різних блоків

Мережа U-Net² є 2-рівневою вкладеною U-структурою (рис. 2.12). Її верхній рівень є великою U-структурою, що складається з 11 ступенів. Кожна ступінь заповнюється добре налаштованим RSU (U-структура нижнього рівня). Отже, вкладена U-структура дозволяє ефективніше отримувати внутрішньоетапні багатомасштабні ознаки та здійснювати агрегацію міжетапних багаторівневих ознак. Основна архітектура являє собою U-Net, подібну до Encoder-Decoder, де кожен етап складається з RSU. Наведені три складові U-Net² містять: (1) кодувальник з 6-ма етапами, (2) декодер з 5-ма етапами та (3) модуль об'єднання карт значимості, що прикріплений до етапів декодера та останнього кодувальника.

Перша складова функціонує наступним чином. На етапах кодера En_1, En_2, En_3 та En_4 використовуються, відповідно, залишкові U-блоки RSU-7, RSU-6, RSU-5 та RSU-4. При цьому, символи «7», «6», «5» і «4» позначають висоти (L) блоків RSU. L зазвичай налаштовується відповідно до просторового розрізнення вхідних карток об'єктів. Для карт об'єктів з великою висотою та шириною використовують більше значення L, щоб отримати більш масштабну інформацію. Розрізнення карт ознак En_5 і En_6 відносно низьке, подальше зниження дискретизації цих карт ознак призводить до втрати корисного контексту.

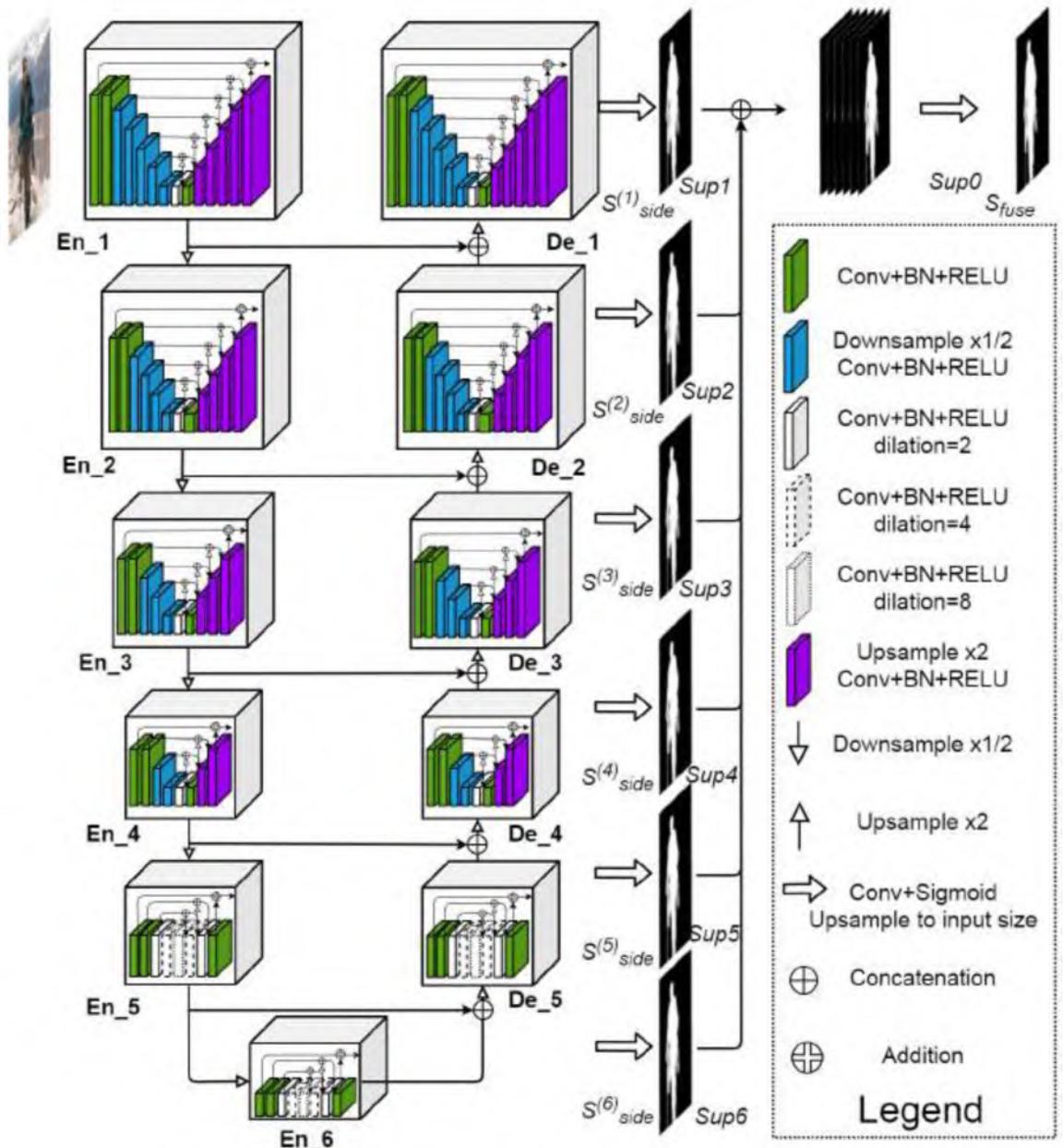


Рисунок 2.12 – Ілюстрація архітектури U-Net²

Отже, як на етапах En_5, так і на етапах En_6 використовуються RSU-4F, де F означає, що RSU є розширеною версією, в якій операції об'єднання та підвищення дискретизації замінено розширеними згортками (див. рис. 5). Це означає, що всі проміжні картки об'єктів RSU-4F мають таку саму роздільну здатність, як і вхідні картки об'єктів.

Друга складова реалізує такі операції. Етапи декодера мають структуру, аналогічну їх симетричним каскадам кодувальника відносно En_6. У блоці De_5 також використовується розширена версія залишкового U-блоку RSU-4F,

яка аналогічна тій, що використовується на етапах кодування En_5 і En_6. На кожному етапі декодера виконується конкатенація карт ознак з підвищеною дискретизацією попереднього етапу і карт з його попереднього етапу. При цьому, каскад симетричного енкодера працює як вхід.

Остання частина є модуль злиття карт значущості, який використовується для створення карт ймовірності значущості. Подібно до HED, U-Net² спочатку генерує шість карт ймовірностей значимості бічних виходів карти ймовірностей $S^{(6)}\text{side}$, $S^{(5)}\text{side}$, $S^{(4)}\text{side}$, $S^{(3)}\text{side}$, $S^{(2)}\text{side}$, $S^{(1)}\text{side}$ з етапів En_6, De_5, De_4, De_3, D_2 і De_1 за допомогою шару згортки 3×3 і сигмоїдної функції. Потім він підвищує дискретизацію цих карт помітності до розміру вхідного зображення та об'єднує їх за допомогою операції конкатенації, за якою слідує шар згортки 1×1 і сигмоїдна функція для створення остаточної карти ймовірності значущості S_{fuse} .

Таким чином, дизайн U-Net² дозволяє мати глибоку архітектуру з багатими багатомасштабними функціями та відносно низькими витратами на обчислення та пам'ять. Крім того, оскільки архітектура U-Net² побудована тільки на блоках RSU без використання будь-яких попередньо навчених магістралей, адаптованих із класифікації зображень, вона гнучка та легко адаптується до різних робочих середовищ із незначною втратою продуктивності. У [16] запропоновано два варіанти U-Net² з використанням різних конфігурацій номерів фільтрів: звичайна версія U-Net² має розмір 176,3 МБ та відносно менша версія U-Net² 4,7 МБ. Для підвищення диверсифікації архітектури мережі U-Net² в [16] запропоновані варіанти спільноговикористання U-Net² та VGG-16 і ResNet-50 на магістральних лініях. Вони реалізують додатковий етап після останніх етапів згортки, щоб отримати ті ж рецептивні поля, що і у оригінальній архітектурі U-Net². Згідно табл. 2.1, моделі, що використовують магістральні мережі та RSU в якості декодерів, забезпечують більш високу продуктивність, ніж попередні аблляції, та порівнянну продуктивність з U-Net² невеликого розміру. Однак вони все ж таки поступаються повнорозмірному варіанту U-Net². Таким чином, можливо

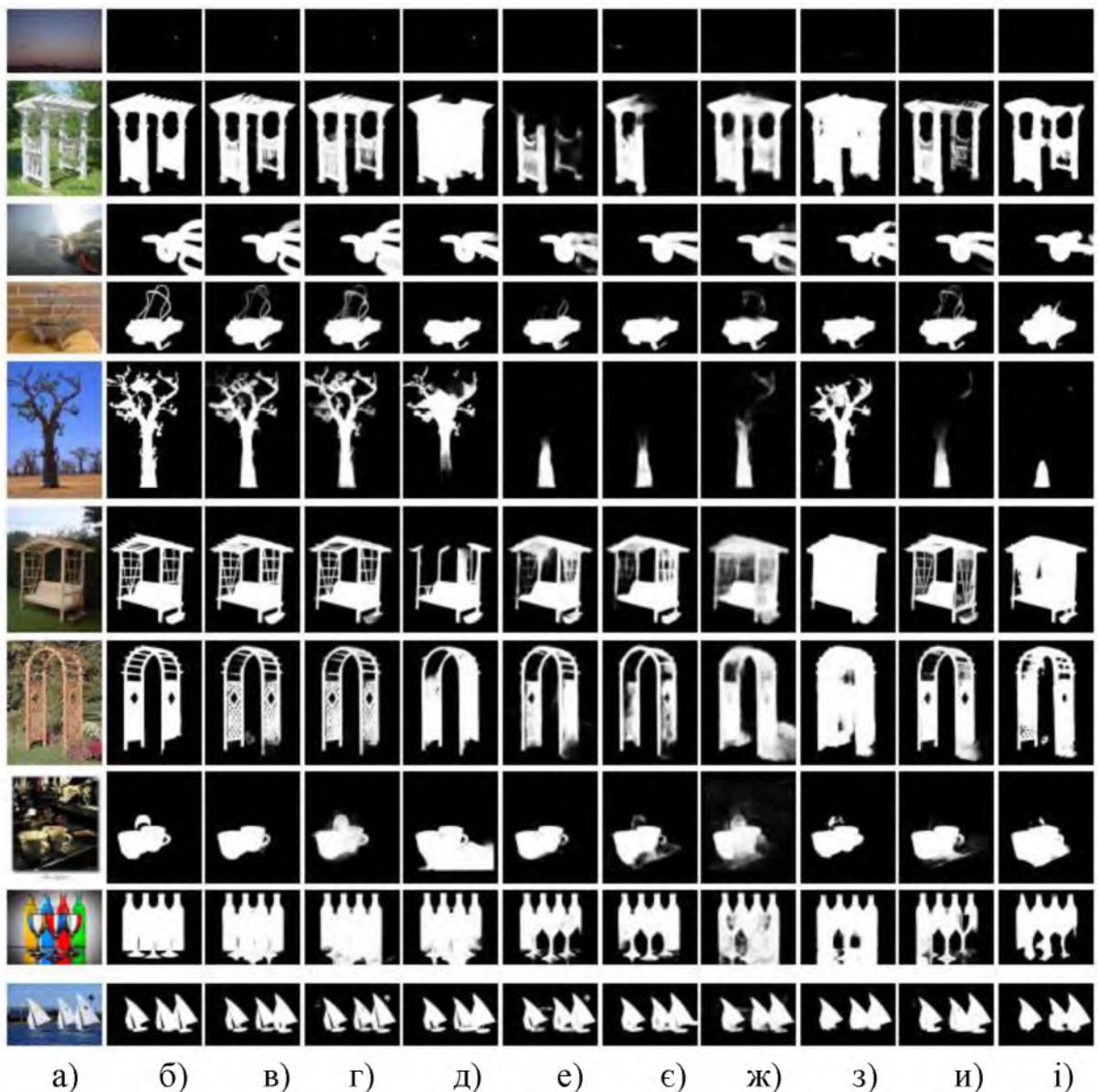
зробити висновок, що архітектура U-Net² без використання магістральних мереж більш конкурентоспроможний дизайн ніж на основі магістральних мереж у даній задачі виявлення суттєвих об'єктів.

Таблиця 2.1 – Результати дослідження абліяції на різних блоках і архітектурах нейронних мереж

Конфігурація	DUT-OMRON		ECSSD		Час, мс
	maxFp	MAE	maxFp	MAE	
Baseline U-Net	0,725	0,082	0,896	0,066	14
PLN U-Net	0,782	0,062	0,928	0,043	16
RES U-Net	0,781	0,065	0,933	0,042	19
DSE U-Net	0,790	0,067	0,927	0,046	70
INC U-Net	0,777	0,069	0,921	0,047	57
PPM U-Net	0,792	0,062	0,928	0,049	105
Stacked HourglassNet	0,756	0,073	0,905	0,059	103
CU-NET	0,767	0,072	0,913	0,061	50
NIV U ² -Net	0,803	0,061	0,938	0,085	30
U ² -Net w/ VGG-16 backbone	0,808	0,063	0,942	0,038	23
U ² -Net w/ ResNet-50 backbone	0,813	0,058	0,937	0,041	41
U ² -Net	0,823	0,054	0,951	0,033	33
U ² -Net ^t	0,813	0,060	0,943	0,041	25

Надалі доцільно розглянути результати кількісного та якісного порівняння [16]. U-Net² досягає найвищої продуктивності в наборах даних DUT-OMRON, HKU-IS і ECSSD майже за всіма 5-ма оціночними показниками. U-Net² займає друге місце з точки зору загальної продуктивності. Розмір нашого U-Net² становить всього 4,7 МБ, що на даний час є досить малою моделлю в області виявлення об'єктів. Маючи набагато меншу кількістю параметрів у порівнянні з іншими моделями, вона забезпечує конкурентоспроможну продуктивність. Хоча її продуктивність не така гарна, як у повнорозмірної U-Net², її невеликий розмір полегшує її використання у багатьох середовищах з обмеженими висотами та пам'яттю. Щоб дати інтуїтивне уявлення про потенційні можливості моделей U-Net², на рис. 2.13 проілюстровано вибіркові результати моделей U-Net² і кілька інших сучасних мереж. Як видно, U-Net² здатні обробляти різні типи цілей та давати точні результати виявлення помітних об'єктів. Так, у 1-му та 2-му рядках (див. рис.

2.13) показані результати обробки малих і великих об'єктів. У 3-му рядку показані результати під час дотику цілі до границь зображення. U-Net² коректно сегментує всі регіони (хоча U-Net² помилково сегментує нижню праву частину, вона все ж таки набагато краще, ніж інші моделі). 4-ий рядок (див. рис. 2.13) демонструє ефективність моделей при сегментації цілей, що складаються як з великих, так і з тонких структур. Як видно, більшість інших моделей добре захоплюють великі області, але не мають тонкої структури по контуру. У 5-му рядку (див. рис. 2.13) показано дерево на відносно чистому фоне голубого неба. Це здається простим, але на справді, це складно для більшості моделей через складність форми мішені. Як видно, моделі U-Net² добре сегментують як стовбур, так і гілки, в той час як інші не можуть сегментувати складну область гілок дерева. У 6-м рядку (див. рис. 2.13) наведена більш складна ситуація завдяки полю конструкції. U-Net² дає майже найкращий результат. Хоча права нижня частина прогнозної карти не завершена, її загальна продуктивність по цій цілі набагато краща, ніж в інших моделях. Крім того, результати роботи моделей U-Net² більш однорідні з меншою кількістю серійних зон, ніж такі у моделей, як PoolNet, CPD, PiCANetR і AFNet. 7-ий рядок (див. рис. 2.13) показує, що моделі U-Net² можуть давати результати навіть краще, чим еталон. Тобто, моделі U-Net² демонструють багатообіцяючі можливості сегментації тонких структур завдяки добре спроектованій архітектурі для вилучення та інтеграції локальної інформації високого розрізnenня та глобальної інформації низького розрізnenня. 8-ий та 9-ий (див. рис. 2.13) рядки ілюструють, що показати потенційні можливості U-Net² у виявленіх цілях, які загороженні фоном, і мають складний зовнішній вигляд переднього плану. 10-ий рядок (див. рис. 2.13) показує, що U-Net² можуть сегментувати кілька цілей, фіксуючи деталі кожної виявленої цілі (див. на область розриву двох частин паруса парусного човна). Таким чином, як повнорозмірні, так і малорозмірні моделі U-Net² здатні працювати з різними сценаріями та давати результати виявлення помітних об'єктів з високою точністю.



умовні познаки:

- | | |
|----------------------------------|-------------------------------|
| а) – зображення; | е) – CPD; |
| б) – GT; | ж) – PiCANetR; |
| в) – U-Net ² ; | з) – R3Net+; |
| г) – «мала» U-Net ² ; | и) – AFNet; |
| д) – BASNet; | і) – DSS+; |
| е) – PoolNet; | «+» – вказує постобробку CRF. |

Рисунок 2.13 – Якісне порівняння U-Net² з іншими методами SOTA

Висновки до розділу 2

Застосування нейронних мереж для вирішення завдань сегментації зображень циферблатів аналогових лічильників базується на методах глибокого навчання. При цьому саму процедуру сегментації доцільно розглядати як частину SOD.

Існує загальна схема проектування більшості мереж SOD. Зазвичай, для цього застосовувались нейронні мережі, архітектура яких початково була орієнтована на класифікацію, а не сегментацію зображень. Вони отримують ознаки, які представляють семантичне значення, а не локальні деталі та інформацію про глобальний контраст, які необхідні для виявлення значущості.

Найбільш вдало завдання сегментації зображень циферблатів аналогових лічильників на основі застосування згорткових нейронних мереж типу: FCN, Tiramisu, PSP, U-Net, Atrous Convolutions, а також гіbridні моделі, наприклад, CRF-CNN.

Однак при цьому існує кілька проблем з мережними архітектурами для виконання SOD:

- вони часто надмірно складні;
- досягають глибшої архітектури, жертвуячи високим розрізненням карт ознак.

Як наслідок, в роботі розглядається багаторівнева глибока інтеграція функцій. В свою чергу, досліджені методи та кого підходу намагаються краще використовувати карти ознак, що створені існуючими основами класифікації зображень. Зазвичай, при цьому намагаються додати складні модулі та стратегії для використання функцій цих магістральних мереж, що може впливати на вимоги до необхідних обчислювальних потужностей.

Альтернативою такого підходу може бути використання архітектури U-Net². Вона безпосередньо витягує багатомасштабні функції поетапно для виявлення суттєвих об'єктів. Для оцінки її потенційних властивостей проведені дослідження, що спираються на кількісне та якісне порівняння з

існуючими архітектурами та зроблений висновок про доцільність її використання при здійсненні семантичної сегментації зображень.

На даний час, відомо два варіанти архітектури U-Net²: повнорозмірна та малорозмірна мережі. Останню можна розглядати як кандидата для реалізації прикордонних та туманних обчислень ПoT.

РОЗДІЛ 3

РЕКОМЕНДАЦІЙ ЩОДО ВИКОРИСТАННЯ ЗАПРОПОНОВАНИХ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

3.1 Визначення особливостей формування бази для формування датасету

Одним з важливих етапів є збір, парсинг та розмітка датасету. Нейронної мережі, яке завдання не вирішували б, знадобиться датасет, де він навчатиметься. І тут абсолютно будь-який алгоритм навчання використовується. навіть у разі навчання з підкріпленим незважаючи на те, що на стартовому етапі база взагалі не потрібна, датасет збирається у процесі функціонування нейронної мережі.

Тому від того, які дані будуть зібрани, багато в чому залежить якість роботи нейронної мережі. Тобто успішно зібрана база може становити 50-60 % успішності взагалі всього проекту. Які ж основні вимоги пред'являються до баз навчання нейронних мереж?

1. Достатній обсяг. На сьогоднішній день обсяг бази є визначальним для навчання нейронної мережі. На жаль, нейронна мережа поки що не здатна навчитися на маленьких датасетах. Процес розробки та досліджень у цьому напрямі ведуться. Але чогось конкретного поки що не придумали і не винайшли. А тому обсяг бази є визначальним. При цьому не можна відповісти на запитання: який обсяг датасету необхідний чи скільки прикладів потрібно для того, щоб навчити проект? Спочатку неможливо сказати скільки прикладів потрібно для того, щоб навчилася нейронна мережа. Правило тут просте – що більше прикладів буде зібрано, то точніше буде модель. Тут правильно говорити про таке поняття, як мінімальний обсяг. Тобто певна кількість прикладів, які необхідні для того, щоб почати тестування моделей, що розробляються. Знову ж таки багато що залежить від того завдання, яке має нейронна мережа. Наприклад, завдання класифікації зображенъ часто буває

досить 300 прикладів кожного класу; Для задачі сегментації зображень – 1000 зображень кожного класу, в принципі, також може бути достатньо. Це мінімальний поріг, який потрібно зібрати для того, щоб почати будувати своєї архітектури. Паралельно не забуваючи, що датасет потрібно розширювати.

2. Чистота даних. Цьому аспекту також слід приділяти увагу. Особливо якщо база збирається не самим розробником нейронної мережі. Залежно від того, де формувався датасет, необхідно перевірити, наскільки актуальні та вірні дані були зібрані.

Наприклад, для нейронної мережі класифікації автомобілів замовлений фрілансерам датасет із 30 марок автомобілів, серед яких зазначені Tesla, Jaguar тощо. При цьому в теці з автомобілями Jaguar можуть бути зображення тваринного ягуара. Або в папці з автомобілями Tesla знаходяться портрети Миколи Тесла. Звичайно, все це позначається на результаті роботи нейронної мережі.

3. Балансування класів. Необхідно звертати увагу на класи, тобто вони мають бути збалансовані.

4. Варіативність. Тут перетинаються пункти, пов'язані із балансуванням. Ale це трохи інше. Наприклад, якщо стоять завдання написати нейронну мережу, яка розрізнятиме марку Nissan від марки Mitsubishi. При цьому зібрано базу, яка містить 10000 зображень з Nissan і 10000 зображень з Mitsubishi. Ale при цьому виявилося, що в папці з автомобілями Nissan присутні лише 3 моделі, а в папці з Mitsubishi – тільки одна. Враховуючи обсяг датасету, потенційно можна реалізувати досить потужну нейронну мережу, яка дає якісні результати, наприклад, на тестовій вибірці точність – 99 %. Ale, згодом, якщо тестувати модель зображень, наприклад, Nissan Juke, то з певною часткою ймовірності, нейронна мережа зможе визначити, що ця машина відноситься до марки Nissan. Ale далеко не факт, тому що нейронна мережа жодної машини у такій комплектації не бачила. Тобто складно припустити, як вона спрацює у реальних умовах на цій машині. Te саме стосується Mitsubishi. Якщо візьмемо якусь нову модель, то передбачити якість роботи нейронної

мережі буде складно. Бажано, щоб датасет містив усі можливі варіанти. При цьому може бути ситуація, коли додали до бази варіативності, але кількість нових прикладів все одно виявилася недостатньою. Таким чином, потрібно не просто збільшити датасет, а ще зробити це якісно та грамотно, не втративши в пункті варіативності.

З огляду на все це ключовим моментом у застосуванні нейронних мереж є вибір набору даних. В області сегментації лічильників ця проблема спрошується тим, що відповідний набір даних є загальнодоступним на вебпорталі Kaggle [35].

3.2 Особливості використання датасету Water Meters

Оскільки поширеним варіантом вирішення задачі семантичної сегментації зображення є використання нейронної мережі PSP [36], розглянемо цей підхід як відправна точка для дослідження, використовуючи архітектуру так званої великої PSP, показану на рис. 3.1. Так як зазначена структура нейронної мережі передбачає 16-кратне зменшення матриці даних, формат зображення, що використовується для навчання, повинен повністю кратним 16. Тому розмір вихідних зображень в наборі даних Water Meters [35] розміром 1000×1778 пікселів був попередньо перетиснутий у розміри, що кратні 16 [37, 38].

Як перший крок у вирішенні завдання навчання було обрано формат 128×224 пікселів, як найближчий за пропорціями до вихідних фотографій. Зокрема, зображення зі стороною кадру 128 пікселів перерахунок з коефіцієнтом $1778/1000$ дає результат 227,584. У цьому випадку округлення до 224 пікселів має бути майже непомітним. Як альтернатива, з подвоєною довжиною більш короткої сторони кадру в 256 пікселів, множення на $1778/1000$ дає 455,168.

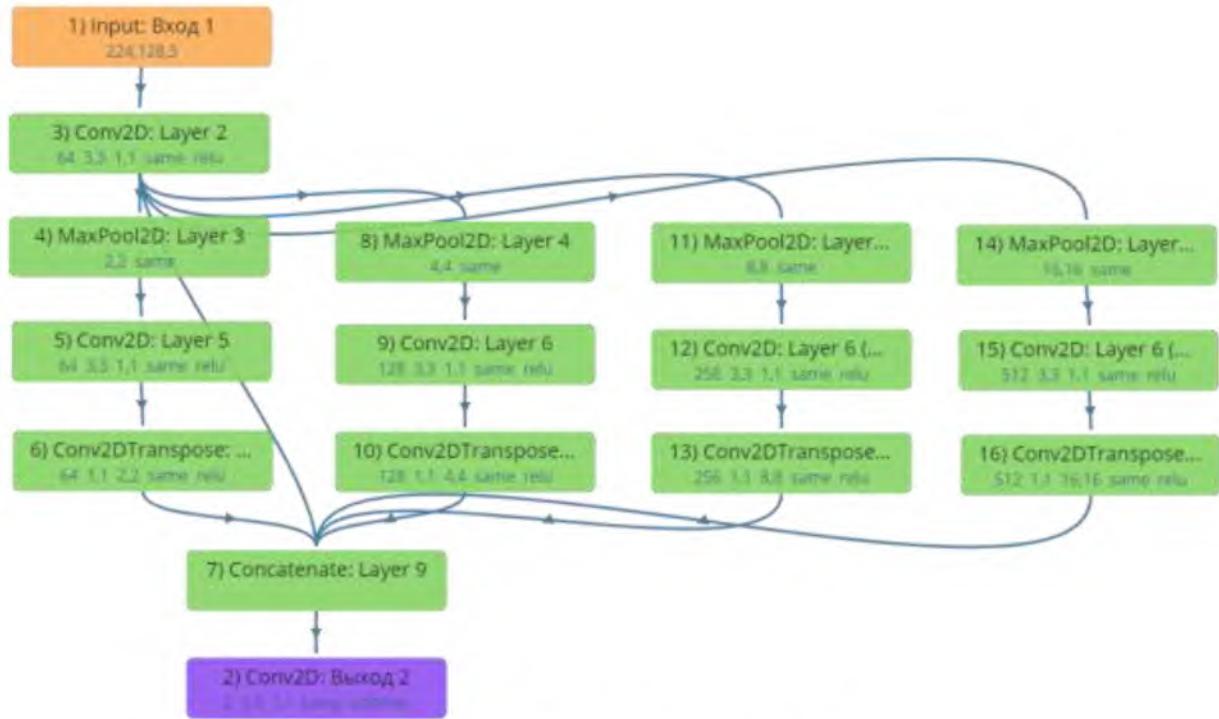


Рисунок 3.1 – 4-channel PSP

У цьому варіанті найближчим числом, кратним 16, буде 464 або 448 пікселів. Так само при 240 пікселях отримуємо:

$$240 \cdot (1778/1000) = 426,72,$$

з відповідним найближчим кратним 16 рівним 432.

Слід зазначити, що вибір того чи іншого формату перестисненого зображення, крім максимальної можливості збереження пропорцій вихідних зображень набору даних, повинен також враховувати обмеження обчислювальних ресурсів, на яких нейронна мережа навчена. Щоб максимально використовувати ці ресурси, у дослідженні використовувалися можливості сервісу Google ColabPro+ у навчанні.

3.3 Оцінка точності синтезованих архітектур нейронних мереж

Процес навчання нейронної мережі PSP виконувався з кроком навчання 0,001 та пакетом $\text{batch}=16$, так як установка пакета $\text{batch}=32$ супроводжувалася помилкою нестачі ресурсів. Набір навчальних даних містив 870 зображень, а набір перевірочних даних – 374 зображення. Маски, які використовуються для сегментації зображення, були чорно-білими. Відсоток простору, що займає чорне тло, становив 98 %, а частку білого вирізу під цифровий дисплей довелося 2 %. Час виконання 200 епох навчання у стандартному режимі підключення Google ColabPro+ з графічною картою V100, оснащеною 16 ГБ ОЗП, становить 32 хв. При цьому максимальна точність навчання на оригінальній PSP досягала 70 % на 73-ій епосі. Продовження навчання до 400 епох дозволило досягти точності 75,2 % на 356-ій епосі.

Потім було розглянуто варіант зміни вихідної архітектури PSP, замінивши шари Conv2DTranspose UpSampling і MaxPooling AveragePooling (рис. 3.2). Отримані результати валідації такої нейронної мережі після навчання дозволяють зробити висновок, що модифікація вихідної архітектури PSP з шарами Conv2DTranspose на наведеному наборі даних навчається гірше, ніж модифікація з PSP з шарами UpSampling. Зокрема, доопрацьована версія дозволила вже на 48-ій епосі здобути точність середнього класу 81,1 %.

Ще більшого підвищення точності вдалося досягти при використанні альтернативної модифікації мережі PSP, що полягала у збільшенні кількості згорткових шарів з функцією активації ReLu в кожному з каналів до 8. каналів (Pooling branch) також залишилися і кількість ядер згортки дорівнює 16. При цьому в кожному каналі використовувалися шари пакетної нормалізації і додатково шар Dropout з На виході каналу застосовувався коефіцієнт проріджування даних 0,1. Ця структура одержала умовну назву PSPBlock2D (рис. 3.3).

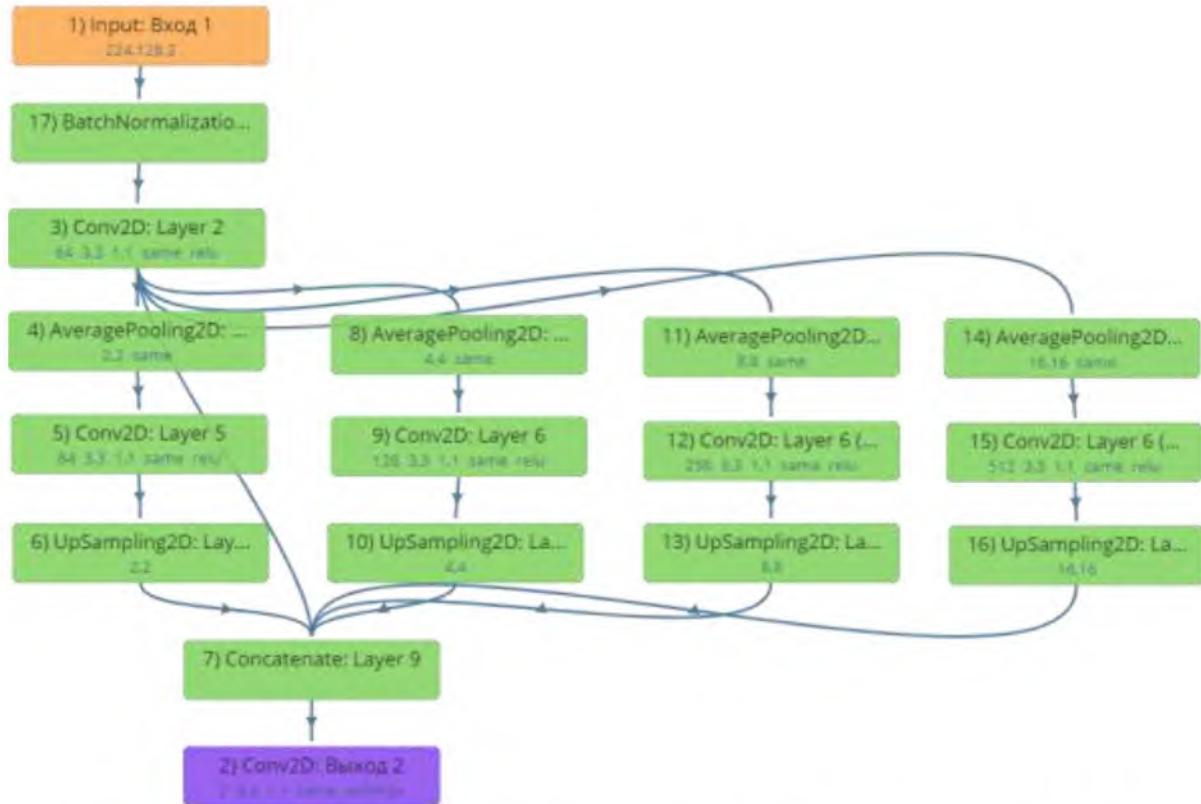


Рисунок 3.2 – Модифікація оригінальної нейронної мережі PSP

В результаті навчання за 160 епох точність класу з найгіршою якістю сегментації на тестовій вибірці досягла 73 %, а середня точність класу досягла 86,2 %. На рис. 3.4 показаний описаний процес навчання, але в рис. 3.5 – якість сегментації на перевірочній вибірці.

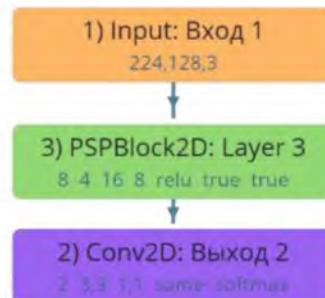


Рисунок 3.3 – Структура нейронної мережі PSPBlock2D



Рисунок 3.4 – Результати навчання нейронної мережі PSPBlock2D

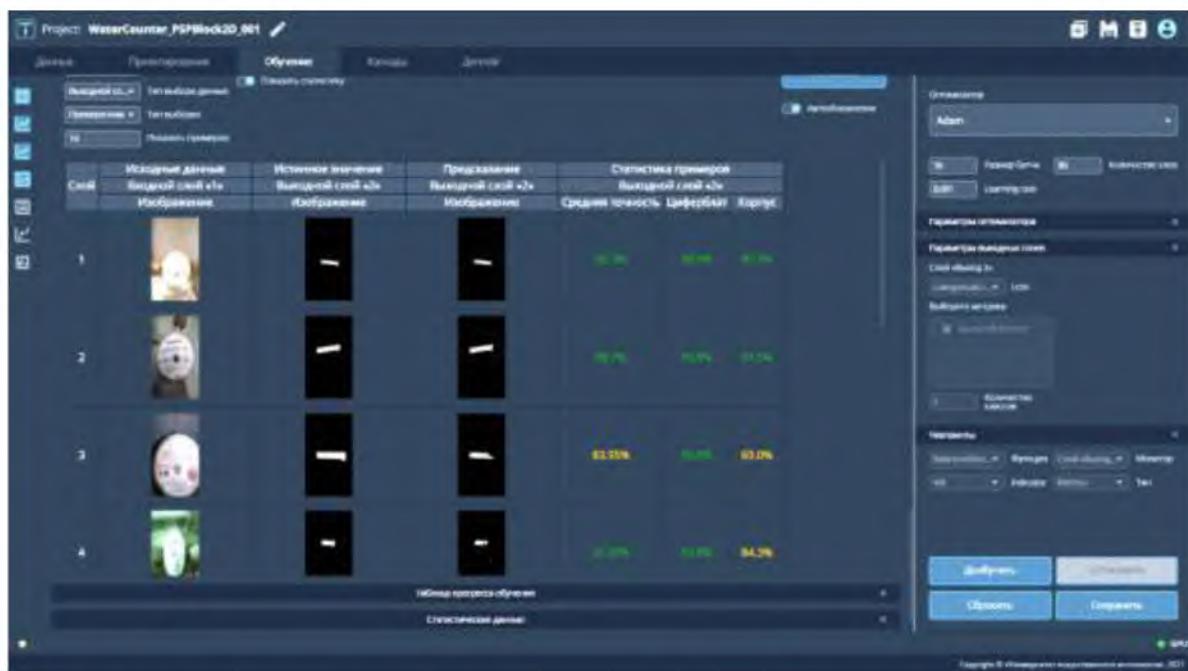


Рисунок 3.5 – Результати навчання нейронної мережі PSPBlock2D

На наступному етапі дослідження проводились з використанням нейронної мережі типу U-Net [39-43]. На рис. 3.6 показано архітектуру полегшеної версії U-Net.

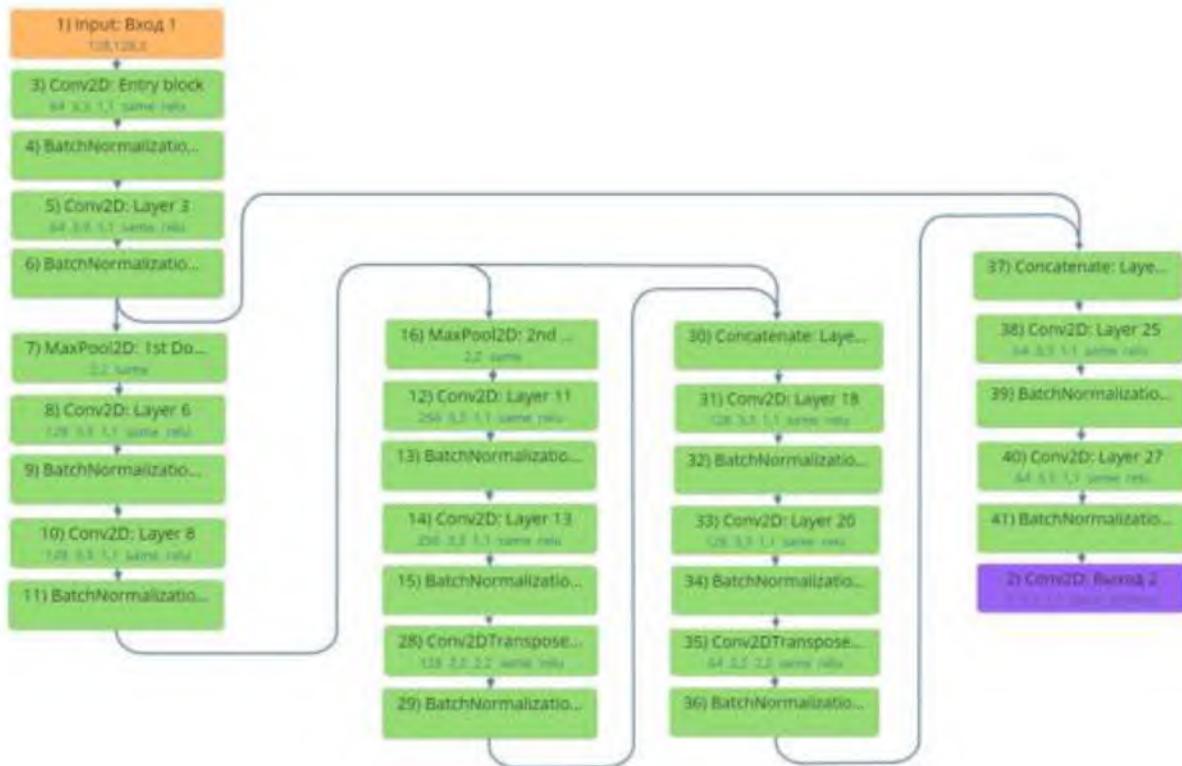


Рисунок 3.6 – Полегшена версія U-Net

Відносна простота його архітектури дозволила перейти на вихідний набір даних 432×240 пікселів та легко проводити довгострокове навчання на 622 епохи з $\text{batch} = 16$ та остаточним кроком навчання 0,00001. Час розрахунку однієї епохи коливався в межах не більше 27-28 секунд. Вже на 54-ій епосі було досягнуто точність 88,4 %, а потім знадобилося понад 400 епох, щоб максимально досягнута точність стабілізувалася на рівні 88,8 % на 464-ій епосі. Підвищенню точності сприяла як архітектура нейронної мережі, так і більший формат зображення під час навчання. Підтвердженням цьому можуть бути результати використання складнішої структури так званого середовища U-Net, схематично представленого на рис. 3.7, з форматом навчального зображення 224×128 пікселів. При цьому на 49-й епосі з $\text{batch} = 32$ було досягнуто точності 87,8 %.

Архітектура даної нейронної мережі включала 5 послідовно з'єднаних базових блоків СВ (рис. 3.8) у низхідній гілці та 4 базові блоки ІВ (рис. 3.9) у висхідній гілці.

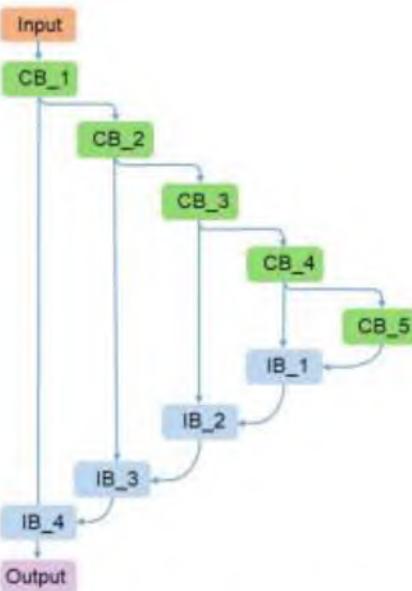


Рисунок 3.7 – Модифікована версія середньої архітектури U-Net



Рисунок 3.8 – Типовий середній будівельний блок U-Net

низхідного потоку (CB)

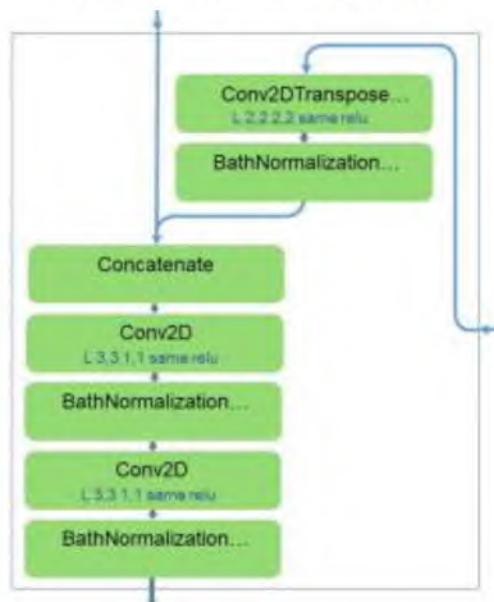


Рисунок 3.9 – Типовий блок IB висхідної гілки середньої мережі U-Net

При цьому всередині кожного базового блоку обох гілок використовувалися однотипні згортки Conv2D і Conv2DTranspose з активаційними функціями ReLu відповідно. Однак кількість ядер згортки збільшувалася з 16 до 256 кратних ступеня 2 у низхідній гілки і зменшувалося у зворотному порядку, зі 128 до 16, у висхідній гілки при переході від однієї базової одиниці до іншої.

У шарах MaxPool2D розмір пулу = 2,2. Кількість фільтрів у шарах Conv2D блоків CB_m становить $K = 23 + m$, $m = 1, \dots, 5$, розмір ядра 3×3 та кроки – 1×1 , заповнення – те саме, функція активації – ReLu.

Кількість фільтрів у шарах Conv2DTranspose і Conv2D блоків IB_r становить $L = 28 - r$, $r = 1, \dots, 4$. Розмір ядра Conv2D становить 3×3 , а кроки – 1×1 , Padding – те саме, функція активації ReLu. Розмір ядра Conv2DTranspose – 2×2 , а крок – 2×2 .

Також був вивчений аналогічний за структурою варіант нейронної мережі, названий великий U-Net, який відрізняється збільшенням числа ядер згортки в низхідних блоках на послідовність 64, 128, 256, 512, 1024 і зміною їх у зворотному порядку у висхідну гілку. Всупереч очікуванням, такий маневр з параметрами архітектури не покращив точність, яка була обмежена на рівні 87 % на 54-ій епосі за тих же розмірів пакетів (32) та кроку навчання.

Подальше ускладнення архітектури було здійснено з допомогою переходу на нейронну мережу типу U-Net++ (рис. 3.10). У ході обчислювальних експериментів було встановлено, що ця нейромережа працює з пакетом 4, але не так ефективно, як велика та середня U-Net. Як і слід очікувати, пакет 8 при 0,001 дає кращу точність, ніж пакет 4. Крім того, U-Net++ працює з пакетом 16, але набагато гірше за 0,001. Отримана точність щодо різних розмірів партій представлена у табл. 3.1.

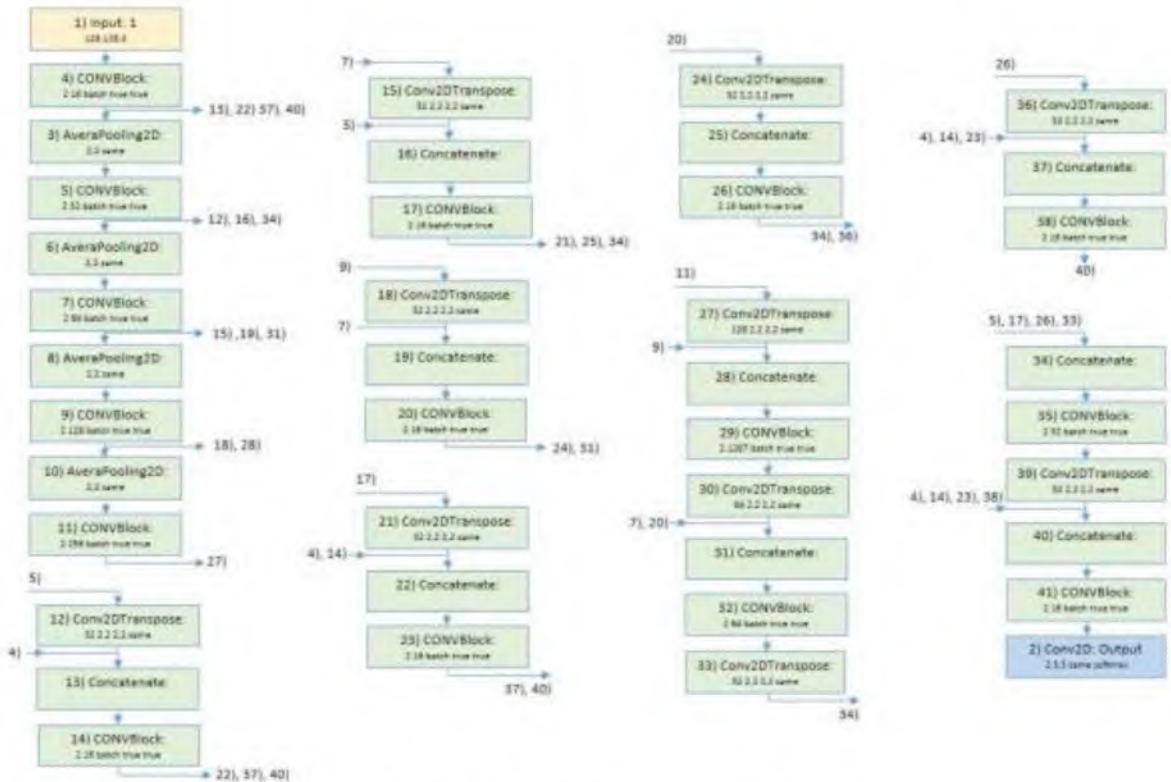


Рисунок 3.10 – Архітектура U-Net++

Таблиця 3.1 – Результатуєчі точність U-Net++

Розмір пакета	Точність	Епоха
8	85,2	33
16	71,5	18
32	83,8	76

Як максимальну складність архітектури нейронної мережі розглядалася так звана U-Net² (рис. 3.11) [44, 45]. В даному випадку набір даних зображень розміром 240×432 пікселів дозволив працювати з пакетами 4, 8 і 16. Максимальна точність з пакетом 16 і кроком навчання 0,001 склала 86 % на 18-й епосі, а з пакетом 8 – 88,5 % на епосі 38. Таким чином, нейронна мережа U-Net² продемонструвала більш інтенсивне навчання.

Порівняння архітектур всіх розглянутих нейронних мереж представлено табл. 3.2. Як видно, велика архітектура не обов'язково дає найкращий результат.

Таблиця 3.2 – Порівняння використовуваних нейронних мереж

Архітектура	Сумарні параметри	Параметри, що навчаються	Параметри, що не навчаються
PSPBlock2D	34429058	34426498	2560
U-Net++	2084370	2081042	3328
U-Net2	682290	678706	3584
PSP (рис. 3.1)	923266	923266	-
PSP (рис. 3.2)	574158	574152	6
Large U-Net	31060226	31046530	13696
Medium U-Net	1948226	1944802	3424
Light U-Net	1869826	1866882	2944

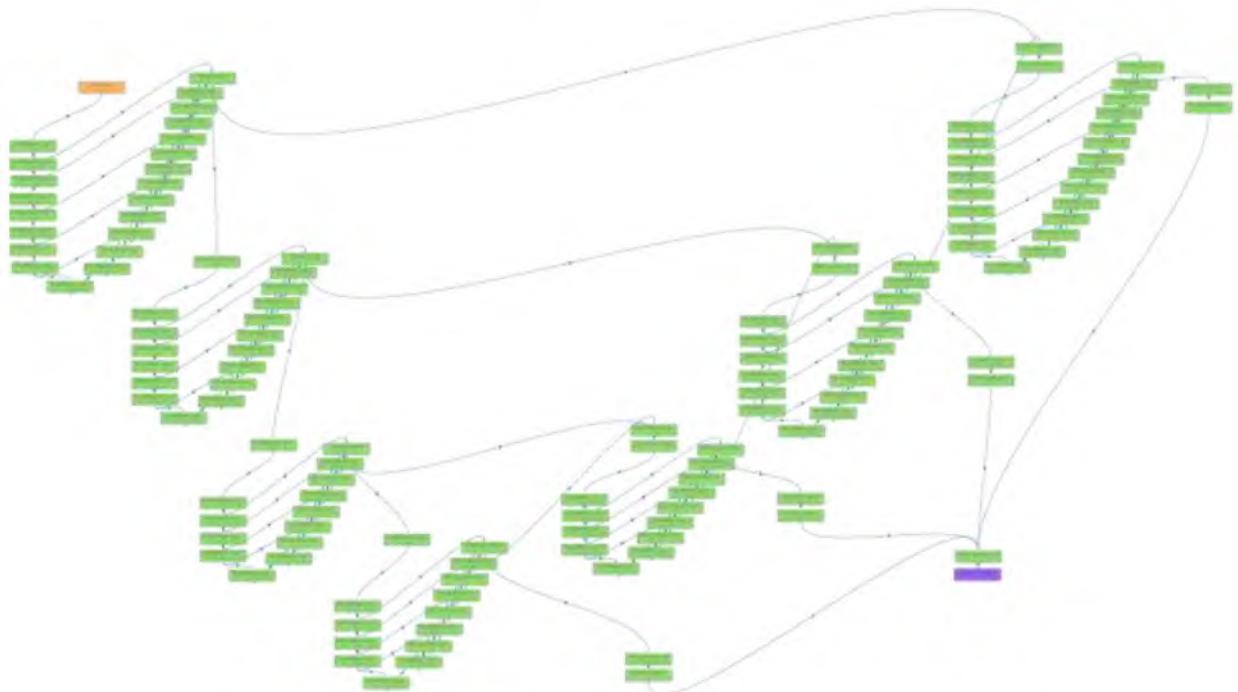


Рисунок 3.11 – Загальний вигляд архітектури U-Net²

Так як у пакеті $\text{batch} = 32$ при навчанні із зображеннями 240×432 виникає помилка «Недостатнього ресурсу», то з цією проблемою було здійснено переход на менший формат зображення 128×224 пікселів. При цьому на 71-ї епосі отримано точність 85,8 %.

3.4 Застосування механізму уваги

На наступному етапі проводилися дослідження з застосуванням різних за структурою та місцем включення в нейронну мережу типу U-Net механізмів уваги. Основна ідея застосування таких структур уваги полягала в тому, щоб зважити матрицю пікселів зображення за допомогою особливої маски, що селектує найбільш значущу область. При цьому, відповідні ваги підбираються у процесі навчання, максимізуючи точність сегментації.

Процес аналізу проводився щодо одно- та 2-канальних структур уваги. На рис. 3.12 представлений один із варіантів одноканального механізму уваги, що охоплює кілька шарів нейромережі з прокидом виходу каналу уваги вперед по її структурі.

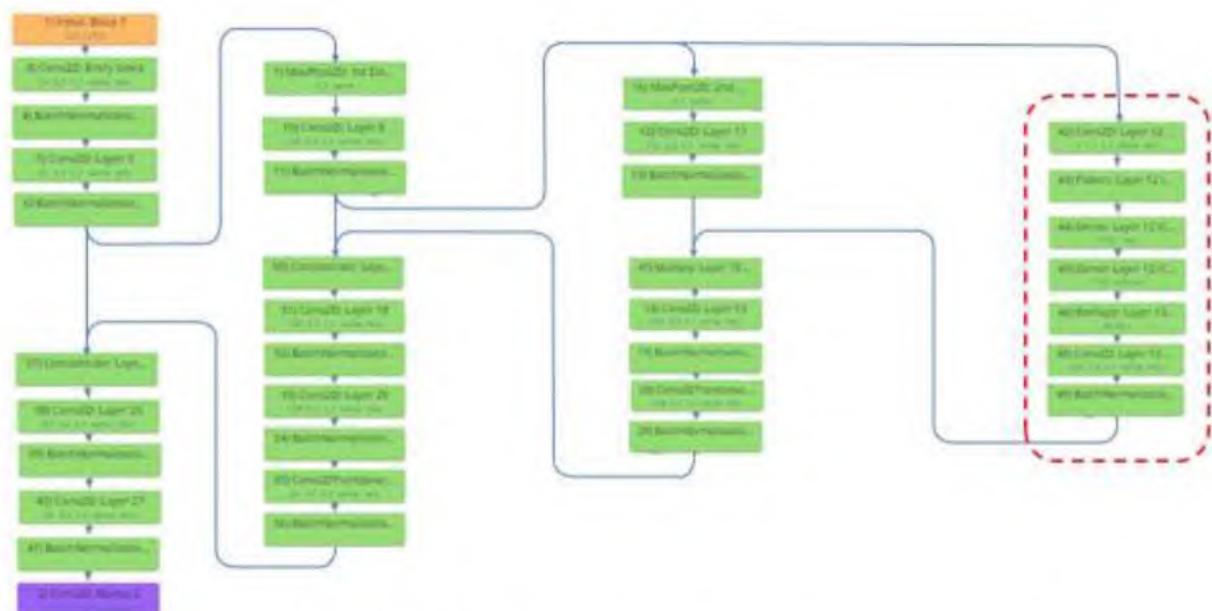


Рисунок 3.12 – U-Net з механізмом уваги, що охоплює
3 шари нейронної мережі

Деталізована схема механізму уваги на рис. 3.13 містить на вході згортковий шар Conv2D з одним фільтром, що формується за допомогою одиничного вагового ядра і такого ж кроку ковзання. На вхід Conv2D надходить зображення формату 112×64 пікселів. Далі йде вирівнюючий шар

Flatten, виходи якого підключені до повнозв'язного нейронного шару Dense з функцією активації ReLu. Він має кількість нейронів, що дорівнює кількості пікселів в одному кадрі на виході механізму уваги (1792). Стільки ж нейронів містить і наступний шар Dense із функцією активації Softmax. Векторний вихід другого шару Dense далі трансформується шаром Reshape зображення формату 52×32 пікселів. Для узгодження структури масивів, що надходять як спів множники на вхід шару множення Multiply, на виході каналу уваги в шарі Conv2D формується 256 фільтрів, виходи яких піддаються нормалізації відповідно до процедури BatchNormalization.

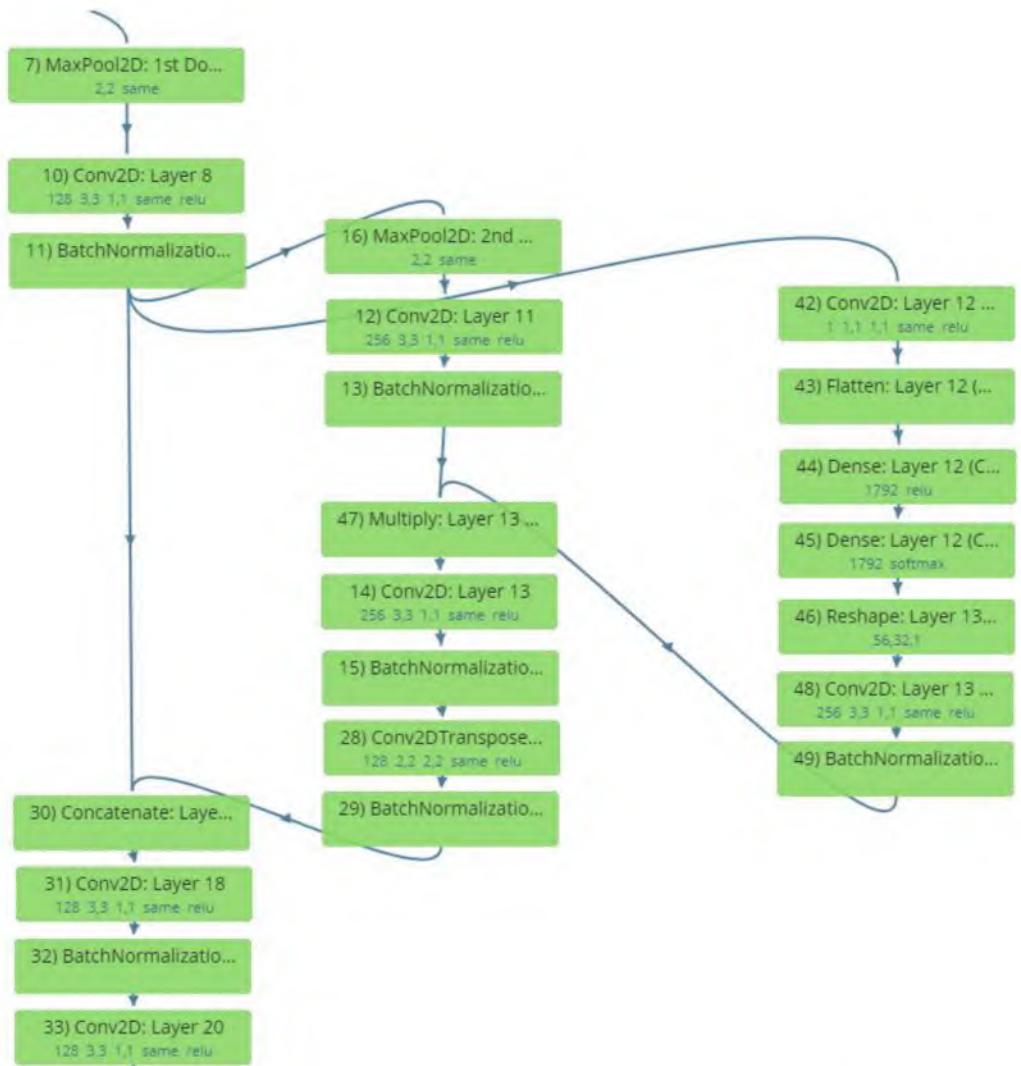


Рисунок 3.13 – Структура одноканального механізму уваги рис. 3.12

Навчання нейронної мережі розглянутої структури дозволило отримати точність 85,1 % у межах 200 епох. При цьому спостерігалося затягування початку процесу навчання на перевірочній вибірці до 18 епох, що було обумовлено збільшенням кількості параметрів, що тренуються.

Більш високу точність, 85,3 %, дозволило отримати включення описаної вище схеми уваги сегменті мережі, відповідному мінімальному розміру зображення. При цьому вхід і вихід каналу уваги підключаються паралельно до стику двох сусідніх шарів (рис. 3.14).

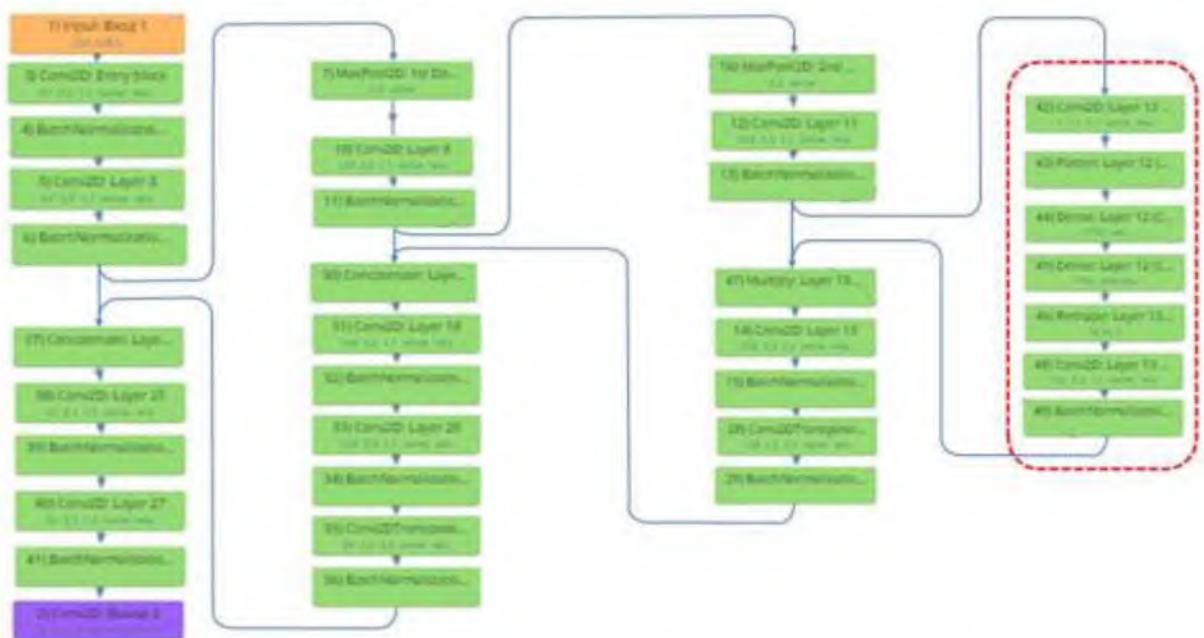


Рисунок 3.14 – Структура U-Net з одноканальним механізмом уваги, інтегрованим між двома шарами

Альтернативний варіант підключення механізму уваги полягає в його інтеграції паралельно лінії зв'язку між двома шарами безпосередньо перед шаром збільшення зображення Conv2DTranspose, як показано на рис. 3.15. Загальна кількість параметрів нейронної мережі дорівнює 8151683, з них тренуванню підлягають – 8148483. Такий варіант нейронної мережі дозволив досягти точності 86,3 % на 418-ій епосі, що збігається з точністю базової версії U-Net без механізму уваги.

Керуючись концепцією множинної уваги, далі була синтезована схема U-Net з двома послідовними каналами уваги, представлена на рис. 3.16.

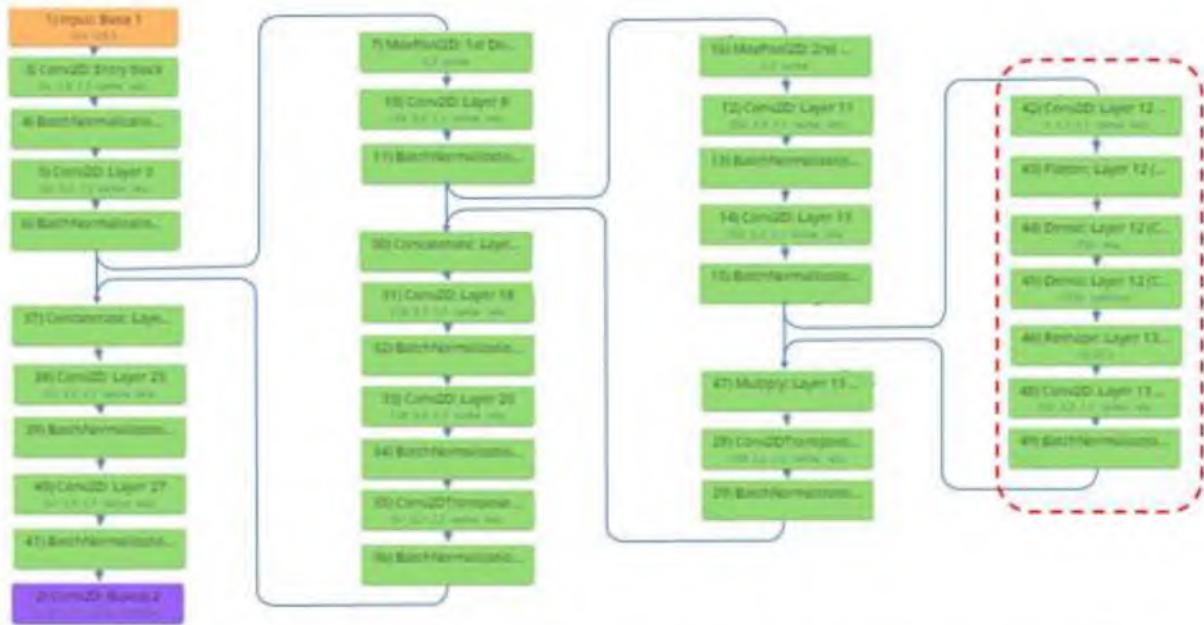


Рисунок 3.15 – Альтернативна архітектура U-Net з одноканальним механізмом уваги, підключеним до міжшарового інтервалу

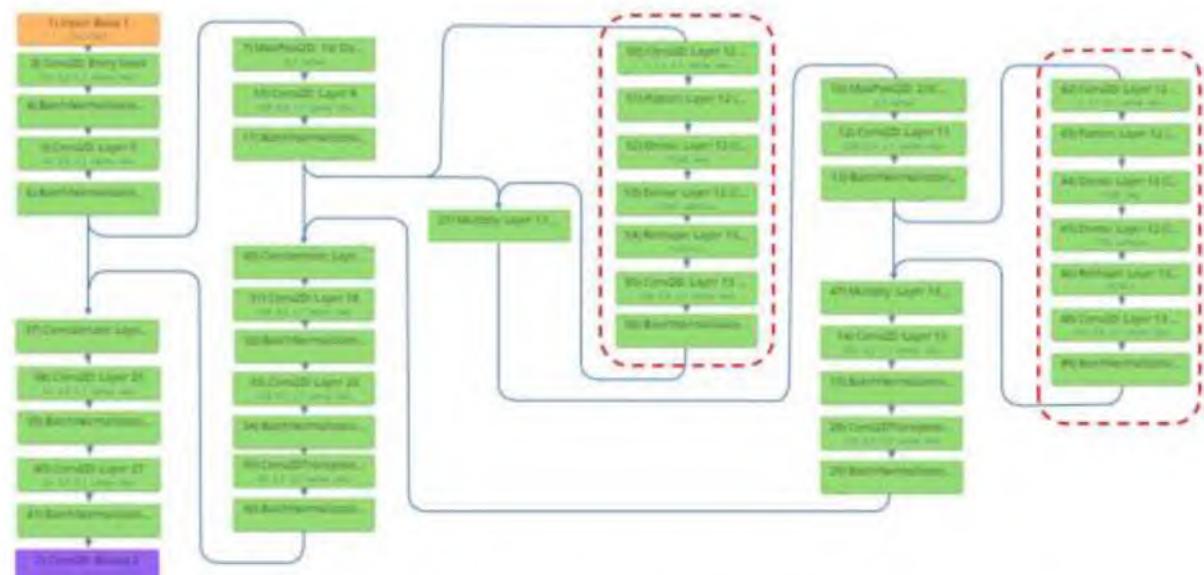


Рисунок 3.16 – U-Net з 2-канальним механізмом уваги

Проти очікування, таке ускладнення нейронної мережі призвело до погіршення точності сегментації, що знизилася до 75,3 %. При цьому помітно погіршилася і швидкість навчання. Це підтверджує розглянуті в 2-му розділі теоретичні положення.

Такий результат дозволив зробити однозначний висновок, що стосовно датасету, що розглядається, одноканальний (однокаскадний) механізм уваги є кращим рішенням. Також більш виграшним варіантом є використання у схемі уваги послідовно двох шарів Dense замість одного. Такий принцип реалізації механізму уваги на спарці шарів Dense дає переваги в точності сегментації і в порівнянні з сuto згортковою схемою, що містить кілька послідовно включених згорткових шарів і шарів нормалізації.

Природно, у межах проведених досліджень було охопити всі варіанти реалізації механізму уваги. Цілком ймовірно, що деякі з них можуть виявитися ефективнішими. Перевірка цього припущення є метою подальших досліджень.

Отримані результати дозволили дійти висновку, що стосовно датасету Water Meters механізм уваги, інтегрований у U-Net середньої складності, не дозволив поліпшити точність сегментації проти базової архітектурою. У проведених експериментах вдалося отримати точність сегментації 86,3 % як у разі використання механізму уваги, так і без нього. Однак, не виключено, що використання складніших схем уваги у поєднанні зі структурною різноманітністю архітектур U-Net дозволить подолати зазначені паритет на користь нейронних мереж з інтегрованим механізмом уваги.

3.5 Техніко-економічне обґрунтування прийнятих рішень

Враховуючи актуальністьграничних обчислень, доцільно вивчити можливість реалізації сегментації та розпізнавання на базі TensorFlow Lite [46] на таких пристроях як ESP32-CAM або Raspberry Pi Zero 2W (рис. 3.17). Приклад такого підходу представлений у [47]. На даний час, використання ESP32-CAM є найбільш економічним варіантом для реалізації периферійного обладнання IoT [48] з можливістю організації оптичного каналу (рис. 3.18).

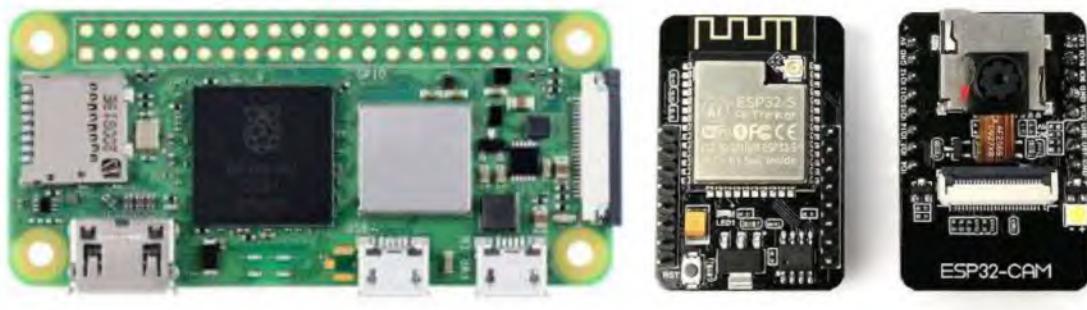


Рисунок 3.17 – Raspberry Pi Zero 2W та ESP32



Рисунок 3.17 – Варіанти прототипів кріплення оптичного сенсора на аналогових лічильниках

При цьому вартість ESP32-CAM складає 334 грн, його кріплення та корпус, які роздруковані на 3D-принтері – 115 грн. Найбільш вартісним є розробка нейронної мережі та її навчання. Орієнтовна вартість розробки повного комплекту оптичного зчитувача з «нуля» знаходиться в межах 250000 грн. Однак наявність у відкритому доступі розмічених датасетів та готових моделей нейронних мереж, що забезпечують остаточну обробку зображень для оцифровки показів лічильників шляхом вирішення завдання класифікації дозволяє значно знизити дані витрати.

Розглянутий у роботі підхід може бути використаний не тільки для створення цифрової інфраструктури. Перспективним напрямом є розробка безпілотних платформ, що використовують транспортні засоби, що спочатку орієнтовані на керування людиною. При цьому для зчитування показників своїх датчиків, наприклад спідометра, оборотів двигуна, тиску масла та ін. можуть використовуватися відеокамери з нейронними мережами, аналогічні розглянутим тут варіантам побутових лічильників.

Подальші дослідження можливості використання нейронних мереж на основі використання технології Object Detection з маркуванням цифрових дисплеїв за допомогою Bounding Boxes, використовуючи результати, наприклад [49]. Крім того, цікавить також узагальнення аналізованого тут підходу на випадок стрілочних аналогових пристрій і використання попередньо навчених нейронних мереж для класифікації зображень у складі нейронної мережі.

Висновки до розділу 3

На основі проведених досліджень, в роботі детально розглянуті основні вимоги до баз навчання нейронних мереж. До них слід віднести: достатній обсяг; чистота даних; балансування класів; варіативність.

На базі для перевірки точності синтезованих моделей архітектур нейронних мереж був вибраний відкритий датасет з Water Meters. На основі його аналізу виконана його модифікація параметрів.

Використовуючи хмарний сервіс Google ColabPro+ було проведено навчання архітектур нейронних мереж сегментації зображень циферблатів аналогових лічильників.

При цьому також проводилися дослідження із застосуванням різних за структурою та місцем включення в нейронну мережу типу U-Net механізмів уваги. Основна ідея застосування структур уваги полягала в тому, щоб зважити

матрицю пікселів зображень за допомогою особливої маски, що селектує найбільш значущу область. Відповідні ваги підбираються у процесі навчання, максимізуючи точність сегментації.

На основі отриманих даних проведена порівняльна оцінка точності запропонованих в роботі архітектур, яка підтвердила висунуті під час досліджень теоретичні положення.

ВИСНОВКИ

Наявність в інфраструктурі застарілого обладнання обліку енергоресурсів не дозволяє повною мірою реалізувати екосистему ПоТ. Перехід до концепції Індустрія 4.0 може бути досить складним. Вибір правильного шляху вирішення питань інтеграції на етапі проектування відіграватиме ключову роль.

З цією метою в роботі запропоновано використання оптичного розпізнавання показань аналогових лічильників. Воно забезпечує мінімальне втручання у існуючий виробничий процес, а головне, не зупиняє і не припиняє його контроль. Тому такі рішення є досить популярними.

Для реалізації такого підходу доцільно використовувати глибоке навчання нейронних мереж нейронних мереж сегментації зображень циферблатів аналогових лічильників.

Досліджені методи багаторівнева глибока інтеграція функцій намагаються краще використовувати карти ознак, що створені існуючими основами класифікації зображень. Зазвичай, при розробці таких моделей прагнуть додати складні модулі та стратегії для використання функцій магістральних мереж, що може впливати на вимоги до необхідних обчислювальних потужностей. Альтернативою такого підходу може бути використання архітектури U-Net². На даний час, відомо два варіанти архітектури U-Net²: повнорозмірна та малорозмірна мережі. Останню можна розглядати в якості кандидата для реалізації прикордонних та туманних обчислень ПоТ.

Враховуючи основні вимоги до баз навчання нейронних мереж, обґрунтований вибір датасету Water Meters, який є у відкритому доступі на вебпорталі Kaggle.

Для підтвердження висунутих в роботі теоретичних положень проведена оцінка точності синтезованих архітектур нейронних мереж. Особлива увага

приділена аналізу ефективності механізмів уваги в задачі сегментації зображень циферблатів аналогових рахунків.

На їх основі сформульовані рекомендацій щодо використання моделей глибокого навчання архітектур нейронних мереж сегментації зображень циферблатів аналогових лічильників.

Пропоновані рішення можуть бути використані не тільки для створення цифрової інфраструктури на основі рішень AI+IoT. Перспективним напрямом є розробка безпілотних платформ, що використовують транспортні засоби, що спочатку орієнтовані на керування людиною.