

УДК 628.222:004.85.032.26

doi: 10.32620/reks.2021.2.01

**В. В. МОСКАЛЕНКО, М. О. ЗАРЕЦЬКИЙ, А. Г. КОРОБОВ,
Я. Ю. КОВАЛЬСЬКИЙ, А. Ф. ШАСХОВ, В. А. СЕМАШКО, А. О. ПАНИЧ**

Сумський державний університет, Україна

МОДЕЛЬ ТА МЕТОД НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЙНОГО АНАЛІЗУ РІВНЯ ВОДИ В СТІЧНИХ ТРУБАХ ЗА ДАНИМИ ВІДЕО ІНСПЕКЦІЇ

Розроблено та досліджено модель та метод навчання для класифікаційного аналізу рівня води на кадрах відео інспекції стічних труб. Об'єктом дослідження є процес розпізнавання рівня води з урахуванням просторово-часового контексту під час інспекції стічних труб. Предметом дослідження є модель та метод машинного навчання для класифікаційного аналізу рівня води на відео-послідовностях інспекції труб за умов обмеженого та незбалансованого набору навчальних даних. Запропоновано чотирьохетапний алгоритм навчання класифікатора. На першому етапі навчання відбувається навчання з нормалізованою триплетною функцією втрат та регуляризованою складовою для штрафування за помилку округлення вихідного сигналу мережі до двійкового коду. Наступний етап потрібний для визначення двійкового коду класу відповідно до принципів заводозахисного кодування з самовиправленням помилок, але з урахуванням внутрікласових і міжкласових відношень. Обчислений еталонний вектор кожного класу використовується як цільова розмітка зразка для подальшого навчання з використанням об'єднаної крос-ентропійної функції втрат. Останній етап машинного навчання пов'язаний з оптимізацією параметрів вирішувальних правил за інформаційним критерієм для врахування меж відхилення двійкового подання спостережень кожного класу від відповідних еталонних векторів. Як модель класифікатора розглядається поєднання 2D згорткового екстрактора ознак кадру з темпоральною мережею для аналізу міжкадрових залежностей. При цьому виконується порівняння різних варіантів темпоральної мережі. Розглядаються 1D регулярна згорткова мережа з дір'явими згортками, 1D каузальна згорткова мережа з дір'явими згортками, рекурентна LSTM-мережа, рекурентна GRU-мережа. Порівняння ефективності моделей відбувається за мікро-усередненою метрикою F1, що обчислюється на тестовій вибірці. Результати, отримані на наборі даних від компанії Ace Pipe Cleaning (Канзас Сіті, США), підтверджують придатність моделі і методу навчання до практичного використання, отримане значення F1-метрики дорівнює 0,88. При цьому результати навчання за запропонованим методом порівнювалися з результатами, отриманими традиційним методом. Було показано, що запропонований метод забезпечує збільшення значення мікро-усередненої метрики F1 на 9 %.

Ключові слова: *стічні труби; інспекція; класифікаційний аналіз; згорткова нейронна мережа; рекурентна нейронна мережа; функція втрат; регуляризація; інформаційно-екстремальне машинне навчання.*

Вступ

Стічні мережі є важливим елементом комунальної інфраструктури, який дозволяє безпечно транспортувати стічні води від домогосподарств, підприємств та дощові води з вулиць до спеціалізованих споруд. Оскільки стічні труби прокладені під землею, про їх наявність легко забути, доки вони не зіпсуються. Заміна всієї стічної труби вимагає значних матеріальних витрат, може бути пов'язана зі значними земляними роботами і часто призводить до ускладнень дорожнього руху. Більш економним варіантом є ремонт пошкоджених труб шляхом відновлення, але для цього потрібно мати інформацію про їх актуальний функціональний стан [1]. Однак безпосередня інспекція стічних труб людиною у переважній більшості випадків значно ускладнена чи взагалі неможлива. Це зумовлене невеликими

діаметрами більшості труб, а також небезпекою для життя та здоров'я через наявність отруйних газів та власне самих стічних вод. Найбільш поширеним методом оцінки стану стічної труби є використання дистанційно керованих, зазвичай через кабель, мобільних роботів (кравлерів). Вони, як правило, оснащені відеокамерою (CCTV) та джерелом світла. Такий робот подається в трубу через найближчий доступний каналізаційний люк. Людина-інспектор керує роботом з відповідно обладнаного автомобіля і здійснює ручне кодування стану труби за вхідними відеоданими. Для такої процедури характерні повільність перевірки відео та наявність помилок, що зумовлене впливом суб'єктивних факторів людини-інспектора. Вказане вище обумовлює значний науковий та комерційний інтерес до автоматизації процесу перевірки даних відео інспекції через використання машинного зору та машинного навчання [2].

Оцінка рівня води в стічних трубах є важливою складовою інспекції. Рівень води є ключовим показником доступності стічної системи для проведення оцінки функціонального стану. Зависокий рівень води приховує дефекти труби і знижує ефективність інспекції. Крім того, визначення зміни рівня води вздовж труби під час інспекції дозволяє виявляти місця просадки або порушення цілісності труби. Просадка труби є одним з важливих кодів функціонального стану труби в рамках американського та британського стандартів [3]. На перший погляд, оцінювання рівня води в стічних трубах може здатися простим завданням для алгоритмів машинного зору, оскільки труби мають досить просту конструкцію з невеликою кількістю під'єднань. Однак відеозображення містять значну шумову складову через туман, випаровування, відблиски світла та втрату видимості під водою. Крім того, камера спостереження під час інспекції здійснює повороти для детального огляду зон інтересу. Все це обумовлює необхідність врахування під час аналізу даних зміни орієнтації камери та видимості на послідовності кадрів інспекції.

Сучасним підходом до класифікаційного аналізу послідовності кадрів є використання глибоких нейронних мереж. Важливими етапами класифікаційного аналізу окремих кадрів відео чи їх послідовності є ознаковий опис окремих кадрів та ознаковий опис залежності між кадрами відео, оскільки саме від цього залежить ефективність вирішувальних правил. Найбільш ефективним підходом до ознакового опису зображень досі вважається використання згорткових нейронних мереж [4]. Що стосується аналізу послідовностей, то провідні позиції займають рекурентні та темпоральні згорткові мережі, діряві версії яких забезпечують перевагу в швидкодії без втрати інформативності [5]. Проте отримані результати, окрім архітектури моделі, залежать і від методів машинного навчання та регуляризації. Особливо це актуально у випадку обмеженого обсягу розмічених навчальних даних і високої варіативності спостережень. Дослідження ефективності різних архітектур моделі та методів її навчання для конкретних прикладних задач за умов обмеженого обсягу розмічених даних досі залишається актуальною задачею, оскільки дотепер немає єдиного універсального підходу до аналізу даних.

Метою статті є розробка нових моделі і методу навчання для ефективного розпізнавання рівня води вздовж труби за даними відео інспекції за умов обмеженого обсягу розмічених навчальних даних.

Пов'язані роботи

У перших алгоритмах аналізу рівня води в стічних трубах на кадрах відео інспекції для ознаково-

го опису використовувалися класичні методи машинного зору, такі як детектування ключових локальних ознак (feature detection), детектування країв (edge detection), обробки контурів [6, 7]. Проте даний підхід ігнорує велику кількість контекстної інформації, а також вимагає розробки правил пост-процесингу. Подібні алгоритми мають велику кількість параметрів і умов, які налаштовуються розробником вручну, що може обумовити виникнення неповноти правил чи їх протиріччя внаслідок людського фактору.

Прогрес в галузі аналізу візуальних даних досягнуто в рамках парадигми глибокого машинного навчання, що основана на використанні ієрархічного ознакового опису. Ієрархічні екстрактори ознак вирізняються більш високою інформаційною ємністю порівняно з моделями, що мають один прихований шар [5, 6]. При цьому найбільш ефективним підходом до ознакового опису зображень досі вважається використання глибоких згорткових нейронних мереж [4]. У праці [8] було запропоновано використовувати згорткові мережі для сегментації води в трубі на відеозображеннях високої роздільної здатності. Однак розглянутий у праці [8] датасет не можна вважати репрезентативним, оскільки там розглядалась лише одна труба і не враховувалась мінливість контексту та дефектів, що зустрічаються на практиці.

Що стосується аналізу послідовностей, то провідні позиції займають рекурентні та темпоральні згорткові мережі, діряві (dilated) версії яких забезпечують перевагу в швидкодії [5]. При цьому може розглядатися як звичайний центрований варіант виходу моделі, так і каузальний. Проте деякі контексти спостереження чи їх частини з'являються рідко і можуть мати значну внутрікласову варіативність, що призводить до дисбалансу та малої кількості розмічених зразків, які відповідають складним і нерегулярним ситуаціям. Це накладає обмеження на застосування глибоких моделей, які є чутливими до обсягу та збалансованості розмічених навчальних даних.

Одним із шляхів підвищення ефективності моделей за умов обмеженого обсягу розмічених даних є використання ідей та методів теорії інформації та синтезу вирішувальних правил в рамках геометричного підходу. Прикладом методів в рамках геометричного підходу є сіамські нейронні мережі, фітнес-функція яких використовує обмеження та співвідношення для відстаней між зразками однакових та різних класів [9]. Сіамські мережі показали найбільшу ефективність в алгоритмах навчання "з декількох поглядів" (few-shot learning) та мета-навчання, однак основна їх функція найчастіше полягає у формуванні інваріантного ознакового подання (feature

embedding). З метою підвищення стійкості до шуму у вигляді артефактів та обмеженої видимості природньо використовувати методи теорії інформації і кодування. Наприклад, коди, що виправляють помилки (error-correcting output codes), реалізують наскрізне (end-to-end) псевдо-ансамблювання і забезпечують підвищення точності та робастності мульти-класового класифікатора [10]. Однак існуючі методи вибору двійкових кодів класів не враховують внутрішньої структури класів. Методи інформаційно-екстремального машинного навчання забезпечують оптимізацію в інформаційному розумінні параметрів вирішувальних правил, основаних на кодах, що виправляють помилки [11]. Однак інформаційно-екстремальне навчання не забезпечує механізми навчання глибоких моделей з кінця-в-кінець. Таким чином, поєднання ідей та методів сіамських нейронних мереж та інформаційно-екстремального навчання є перспективним шляхом подальшого вдосконалення моделей аналізу даних, зокрема для аналізу рівня води в стічних трубах.

Постановка задачі

Нехай дано V послідовностей $D_v^{\text{train}} = \{I_{v,k}, L_{v,k} \mid v = \overline{1, V}; k = \overline{1, K_v}\}$ і W послідовностей $D_w^{\text{test}} = \{I_{w,k}, L_{w,k} \mid w = \overline{1, W}; k = \overline{1, K_w}\}$ розмічених відеокадрів для навчання та тестування алгоритму аналізу відповідно, де K_v та K_w – довжина v -ї навчальної послідовності та довжина w -ї тестової послідовності. Нехай дано множину класів $\{X_z^0 \mid z = \overline{1, Z}\}$, що характеризує рівень води в трубі на кадрі відео інспекції. Крім того, відома структура вектора g параметрів моделі аналізу даних, що має вигляд

$$g = \langle e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}, f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle, \quad (1)$$

$$\Xi_1 + \Xi_2 = \Xi,$$

де e_{ξ_1} – ξ_1 -й параметр, що впливає на формування ознакового опису спостережень, $\xi_1 = \overline{1, \Xi_1}$;

f_{ξ_2} – ξ_2 -й параметр, що впливає на ефективність вирішувальних правил, $\xi_2 = \overline{1, \Xi_2}$.

При цьому, відомі обмеження, що накладаються на параметри моделі у вигляді формул

$$R_{e_1}(e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}) \leq 0, \quad R_{f_1}(f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2}) \leq 0.$$

Необхідно знайти під час навчання оптимальні значення параметрів вектора g (1), що забезпечують максимальне значення мікро-усередненого значення F1-міри для класифікатора рівня води в трубі

$$F1 = \frac{2 \sum_z TP_z}{2 \sum_z TP_z + \sum_z FP_z + \sum_z FN_z}, \quad (2)$$

$$g^* = \arg \max_G \{F1(g)\}. \quad (3)$$

де TP_z – кількість правильно-позитивних рішень для вирішувального правила z -го класу;

FP_z – кількість хибно-позитивних рішень для вирішувального правила z -го класу;

FN_z – кількість хибно-негативних рішень для вирішувального правила z -го класу.

Необхідно в режимі розпізнавання рівня води на тестових відео зображеннях інспекції труб забезпечити максимальну достовірність рішень.

Модель та метод навчання для класифікаційного аналізу рівня води

Класифікаційний аналіз відеокадрів у найпростішому випадку може бути реалізований у вигляді однієї згорткової мережі, яка може бути навчена як за традиційною схемою, так і у складі сіамських або генеративних моделей. Однак в ситуаціях втрати видимості чи значного наближення камери до стінок труби отримані зображення втрачають більшу частину корисної контекстної інформації. Тому для відтворення контекстної інформації потрібно кожен кадр аналізувати одночасно з сусідніми кадрами. У загальному випадку модель розпізнавання рівня води вздовж труби матиме у своєму складі на нижньому рівні 2D згорткову мережу для екстракції просторових ознак та 1D темпоральну мережу для аналізу міжкадрових залежностей (рис. 1).

Для аналізу окремих кадрів пропонується використовувати згорткову мережу загального призначення MobileNet. При цьому коефіцієнт ємності даної мережі встановлено на рівні 0,25 і використовується лише екстрактор ознак (backbone) без вихідних шарів [11]. Вхідна роздільна здатність мережі дорівнює 160x160 пікселів. На рис. 2, а показано класичний варіант згорткової мережі, а на рис. 2, б модифікований варіант для дослідження пропонуваного методу навчання.

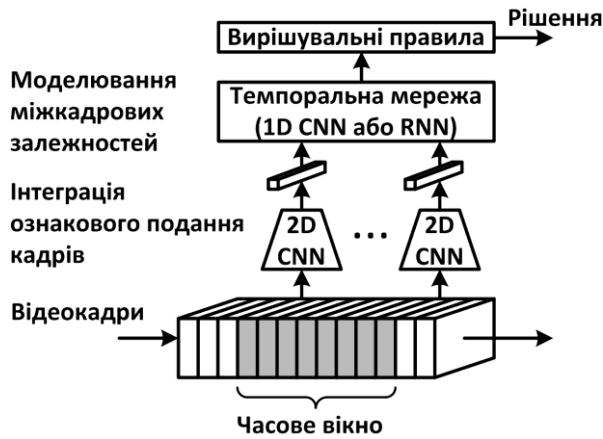
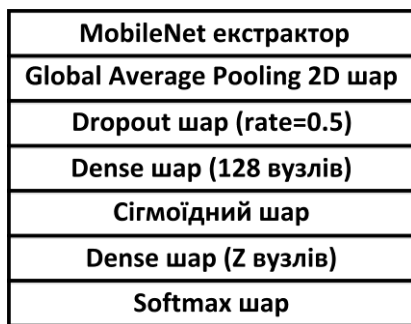
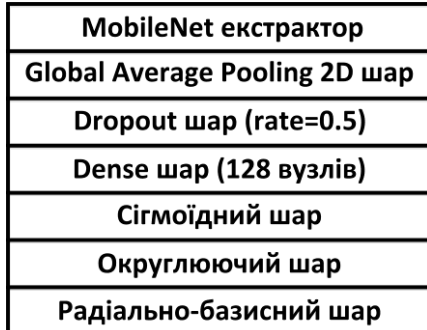


Рис. 1. Узагальнена архітектура класифікатора рівня води



а



б

Рис. 2. Архітектура згорткової мережі: а – традиційний варіант; б – модифікований варіант

Шар Global Average Pooling 2D забезпечує зменшення розмірності, а шар Dropout з коефіцієнтом вимкнення вузлів на рівні 0,5 (тобто 50 %) використовується для регуляризації [8, 11]. Повноз'язні та сигмоїдні шари утворюють вихідний ознаковий опис.

Вирішувальні правила моделі класифікатора зображень містять шар округлення, який формує двійкове подання, та радіально-базисні функції, що визначають належність об'єкта до одного з класів, що розділені гіперсферичними контейнерами у двійковому просторі Хеммінга. Кожен гіперсферичний ко-

нтейнер визначається двійковим еталонним (опорним) вектором, тобто центром контейнера, та радіусом контейнера в кодових одиницях відстані Хеммінга. В цьому випадку радіально-базисна функція належності $\mu_z(b)$ до z -го класу для N -вимірного двійкового вектора b може бути обчислена за формулою

$$\mu_z(b) = 1 - \sum_{i=1}^N b_i \oplus b_{z,i}^* / d_z^* \quad (4)$$

де b_z^* – двійковий еталонний вектор (центр оптимального контейнера) для класу X_z^0 ;

d_z^* – радіус оптимального контейнера для класу X_z^0 в одиницях відстані Хеммінга.

Темпоральна мережа може бути реалізована на одній з популярних архітектур, до яких належить 1D регулярна згорткова мережа з дірчастими згортками, 1D каузальна згорткова мережа з дірчастими згортками, рекурентна мережа з LSTM-комірками (Long Short-Term Memory), рекурентна мережа з GRU-комірками (Gated Recurrent Units) [5].

Базовий блок (BB) згорткових темпоральних мереж показано на рис. 3. Перший шар базового блоку утворюється дірчастими 1D згортками (Dilated Conv), розмір ядра яких дорівнює $k = 3$, а фактор розрідженості (дірчастості) подвоюється з кожним наступним базовим блоком, тобто буде дорівнювати 1, 2, 4, 8 і так далі.

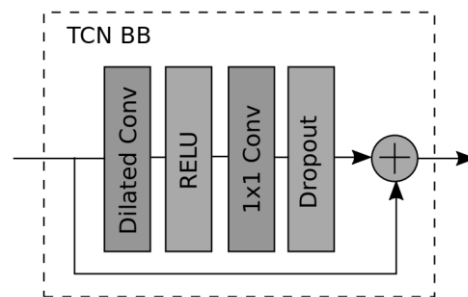


Рис. 3. Базовий блок темпоральної згорткової мережі

На рис. 4 показано рецептивне поле каскаду темпоральної мережі з трьома складеними базовими блоками (BB) з регулярними згортками. Регулярні згортки мають рецептивне поле, яке поширюється однаково як праворуч, так і ліворуч. Це означає, що модель дивиться як в майбутнє (праворуч), як і в минуле (ліворуч). Тобто рівень води на поточному кадрі уточнюватиметься на основі інформації як від минулих, так й від майбутніх кадрів.

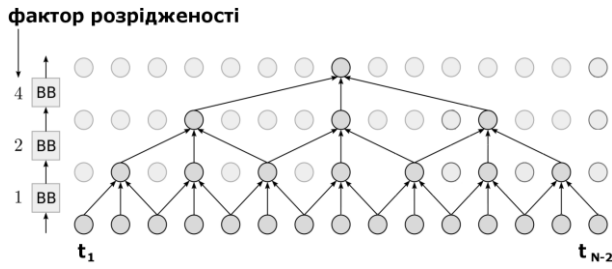


Рис. 4. Темпоральна мережа з регулярними згортками

Рис. 5 ілюструє темпоральну нейромережу з каузальними згортками, що підсилюють продуктивність прогнозування ближче до правого краю. В цьому випадку прогнозований рівень води на поточному кадрі уточнюватиметься на основі ознак і взаємозв'язків попередніх кадрів.

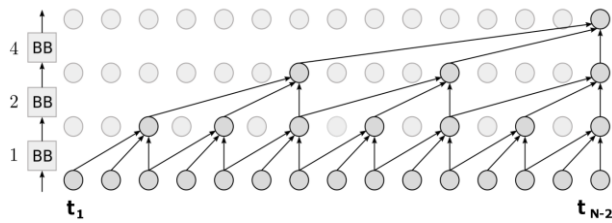


Рис. 5. Темпоральна мережа з каузальними згортками

LSTM-мережа складається з комірок пам'яті, що з'єднані як послідовні шари, і це дозволяє мережі вибірково запам'ятовувати вхідні навчальні дані через унікальну структуру з трьома вентилями (рис. 6, а). Лінія у верхній частині діаграми позначає стан c комірки пам'яті і представляє внутрішню пам'ять. Лінія у нижній частині діаграми – це прихований стан h . Вентилі i , g (вхідні), f (забувальний) та o (вихідний) – це механізми, за допомогою яких LSTM-комірка вирішує проблему зникаючого градієнта. Під час навчання LSTM-комірка оптимізує параметри цих вентилів. Замість вхідних, забувального та вихідного вентилів комірка GRU має лише два вентиля – вентиль оновлення та вентиль скидання (рис. 6, б). Вентиль оновлення визначає, скільки попередньої пам'яті слід зберігати, а вентиль скидання визначає, як об'єднати новий вхід із попередньою пам'яттю. В цьому випадку не існує стійкого стану комірки, який би відрізнявся від прихованого, як у комірці LSTM.

Для порівняння і відслідковування зміни продуктивності внаслідок використання запропонованих рішень навчання здійснюється поетапно. Спочатку виконується навчання моделі розпізнавання окремого кадру без врахування сусідніх кадрів, з використанням традиційного і запропонованого методу

навчання. Далі обирається краща навчена модель і використовується її екстрактор ознак, тобто шари, розміщені до сігмоїдного шару включно, для інтеграції кадрів в темпоральну модель розпізнавання. Кожен тип темпоральної нейронної мережі також навчається за традиційним та запропонованим методом навчання. При цьому для кожної моделі оптимізується по сітці гіперпараметр, що відповідає за розмір мережі, тобто за її інформаційну ємність.

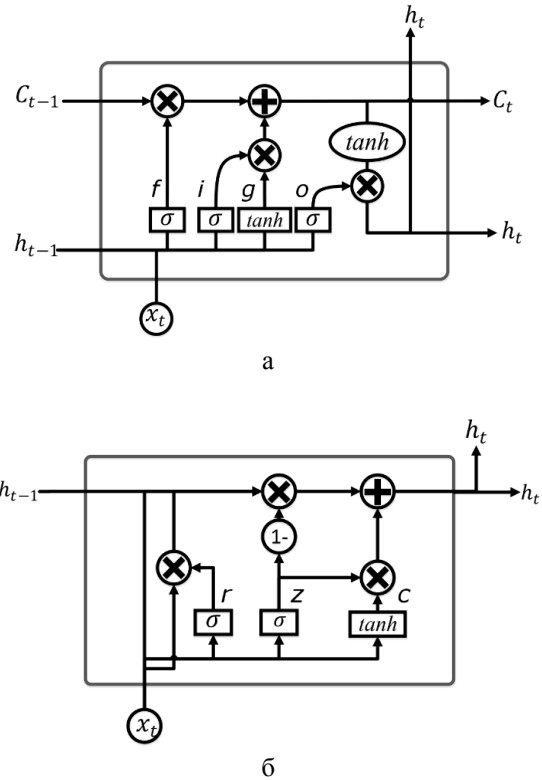


Рис. 6. Комірка рекурентної нейронної мережі: а – LSTM-комірка; б – GRU-комірка

Традиційний метод навчання полягає у додаванні до екстрактора ознак Dense шару з Softmax нормалізацією виходу і використанні методу зворотного поширення помилки, наприклад Adam, з крос-ентропійною функцією втрат.

Модифікований метод включає 4 етапи, які пов'язані з формуванням бінаризованого ознакового опису, що використовується для формування інформаційно-екстремальних вирішувальних правил (рис. 7).

На першому етапі відбувається навчання з нормалізованою триплетною функцією втрат і регуляризуючою складовою для штрафування за помилку округлення вихідного сигналу мережі до двійкового коду. На вхід моделі подається міні-пакет, в якому міститься M зображень кожного класу. При цьому функція втрат розраховується за формулою

$$L = -\log \frac{\exp(\|f(x_a) - f(x_{ep})\|)}{\exp(\|f(x_a) - f(x_{ep})\|) + \exp(\|f(x_a) - f(x_{shn})\|)} + \lambda f(x_a)^T (e - f(x_a)) + \lambda f(x_{ep})^T (e - f(x_{ep})) + \lambda f(x_{shn})^T (e - f(x_{shn})), \quad (5)$$

де $f(x)$ – функція, що описує екстрактор ознак;

x_a – зображення, що обрано випадковим чином з міні-батчу;

x_{ep} – найближчий сусід з міні-батчу, що належить до того ж самого класу, тобто

$$x_{ep} = \arg \min_{x: C(x)=C(x_a)} \|f(x_a) - f(x)\|, \quad (6)$$

де $C(x)$ – функція, що повертає клас зображення;

x_{shn} – зразок зображення з міні-паketу, що є найближчим серед зразків протилежних класів, проте знаходиться далі, ніж x_{ep} зразок, тобто

$$x_{shn} = \arg \min_{C(x) \neq C(x_a)} \|f(x_a) - f(x)\|; \quad (7)$$

$$x_i: \|f(x_a) - f(x)\| > \|f(x_a) - f(x_{ep})\|$$

$e = [1, 1, \dots, 1]^T$ – одинична матриця-стовбець;
 λ – коефіцієнт регуляризації.

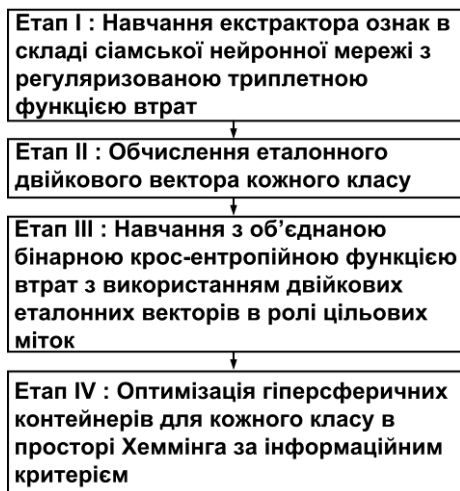


Рис. 7. Етапи запропонованого методу навчання

Наступна фаза потрібна для визначення двійкового коду класу відповідно до принципів самокоректуючих кодів (error-correcting output codes), але з урахуванням внутрікласових та міжкласових відношень. Для цього навчальна вибірка для Z класів $\{x_{z,s} \mid z = \overline{1, Z}, s = \overline{1, n_z}\}$, що містить n_z зразків z -го

класу, кодується дискретним поданням $\{b_{z,s,i} \mid z = \overline{1, Z}, s = \overline{1, n_z}, i = \overline{1, N}\}$ з розмірністю N . При цьому двійкове кодування вхідного зображення $x_{z,s}$ здійснюється шляхом округлення i -го виходу сігмоїдного шару до цілого числа

$$b_{z,s,i} = \begin{cases} 1, & \text{if } f_i(x_{z,s}) > 0,5; \\ 0, & \text{otherwise.} \end{cases}$$

Двійковий еталонний вектор b_z для z -го класу можна визначити шляхом порозрядного порівняння частоти двійкових одиниць в z -му класі з фоновією частотою одиниць в навчальній вибірці

$$b_{z,i} = \begin{cases} 1, & \text{if } \frac{1}{n_z} \sum_{s=1}^{n_z} b_{z,s,i} > \frac{1}{Z} \sum_{c=1}^Z \frac{1}{n_c} \sum_{s=1}^{n_c} b_{c,s,i}; \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Обчислений еталонний вектор z -го класу b_z використовується як розмітка (label) зразка для подальшого навчання з використанням об'єднаної крос-ентропійної функції втрат (Joint Binary Cross Entropy Loss), яка для кожного вхідного зразка x обчислюється за формулою

$$L = -\sum_{i=1}^N (b_i \log f_i(x) + (1 - b_i) \log(1 - f_i(x))), \quad (9)$$

де $f_i(x)$ – значення i -го виходу сігмоїдного шару для вхідного зображення x ;

b_i – значення i -го розряду еталонного вектора класу, до якого належить зображення x .

Остання фаза машинного навчання пов'язана з оптимізацією радіусу контейнерів за інформаційним критерієм для врахування меж відхилення двійкового подання спостережень кожного класу від відповідних еталонних векторів

$$E_z^* = \max_{\{d\}} E_z(d), \quad (10)$$

де $\{d\} = \{0, 1, \dots, \left(\sum_i b_{z,i} \oplus b_{c,i} - 1\right)\}$ – набір концентричних радіусів з центром b_z розподілу даних в z -му класі;

E_z – інформаційний критерій для z -го класу, який є функцією від характеристик точності [2].

Для двохальтернативної системи оцінок і рівноймовірних гіпотез, що характеризує найбільш важкий у статистичному сенсі випадок прийняття рішень ентропійний критерій можна виразити як

функцію точностних характеристик у такому вигляді [11]:

$$E_z^{(k)} = 1 + \frac{1}{2} \left(\frac{\alpha_z^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} \log_2 \frac{\alpha_z^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} + \frac{\beta_z^{(k)}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} \log_2 \frac{\beta_z^{(k)}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} + \frac{D_{1,z}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} \log_2 \frac{D_{1,z}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} + \frac{D_{2,z}^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} \log_2 \frac{D_{2,z}^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} \right), \quad (11)$$

де $\alpha_z^{(k)}(d)$ – частота помилок першого роду на k -му кроці навчання, допустиме значення якої повинне задовольняти умові $\alpha_z^{(k)}(d) < 0,5$;

$\beta_z^{(k)}(d)$ – частота помилок другого роду, допустиме значення якої повинне задовольняти умові $\beta_z^{(k)}(d) < 0,5$;

$D_{1,z}^{(k)}(d)$ – чутливість;

$D_{2,z}^{(k)}(d)$ – специфічність;

d – дистанційна міра, яка визначає радіуси гіперсферичних контейнерів в радіальному базисі простору Хеммінга.

Результати машинного навчання

Алфавіт класів для розпізнавання рівня води в трубі на кадрі зображення без врахування зображень сусідніх кадрів містить 12 класів, що відповідають рівням заповнення труби водою у відсотках: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 % (рис. 8). При цьому парсинг розмічених відеофайлів $\{D_w^{\text{train}}\}$ та $\{D_w^{\text{test}}\}$ призводить до формування різної кількості зразків для кожного класу.

Перед навчанням весь датасет балансується шляхом застосування аугментації міноритарних класів. Як методи аугментації використовуються: зміна масштабу на 1...5 %, поворот зображення на $\pm 18^\circ$, зміна яскравості на $\pm 5\%$. Розмір аугментованої навчальної вибірки кожного класу для класифікатора кадрів становить 3000 зразків.

Алфавіт класів для розпізнавання рівня води з урахуванням сусідніх кадрів містить ті самі класи. Часове вікно для врахування сусідства дорівнює 257 (тобто $128 \cdot 2 + 1$) кадрам. Розмір часового вікна обирається як число, кратне 2, з додаванням одного кадру, щоб забезпечити симетричність для регулярної

структури моделі. При цьому розмір вікна підбирається експериментально, щоб бути близьким до оптимального для всіх моделей.

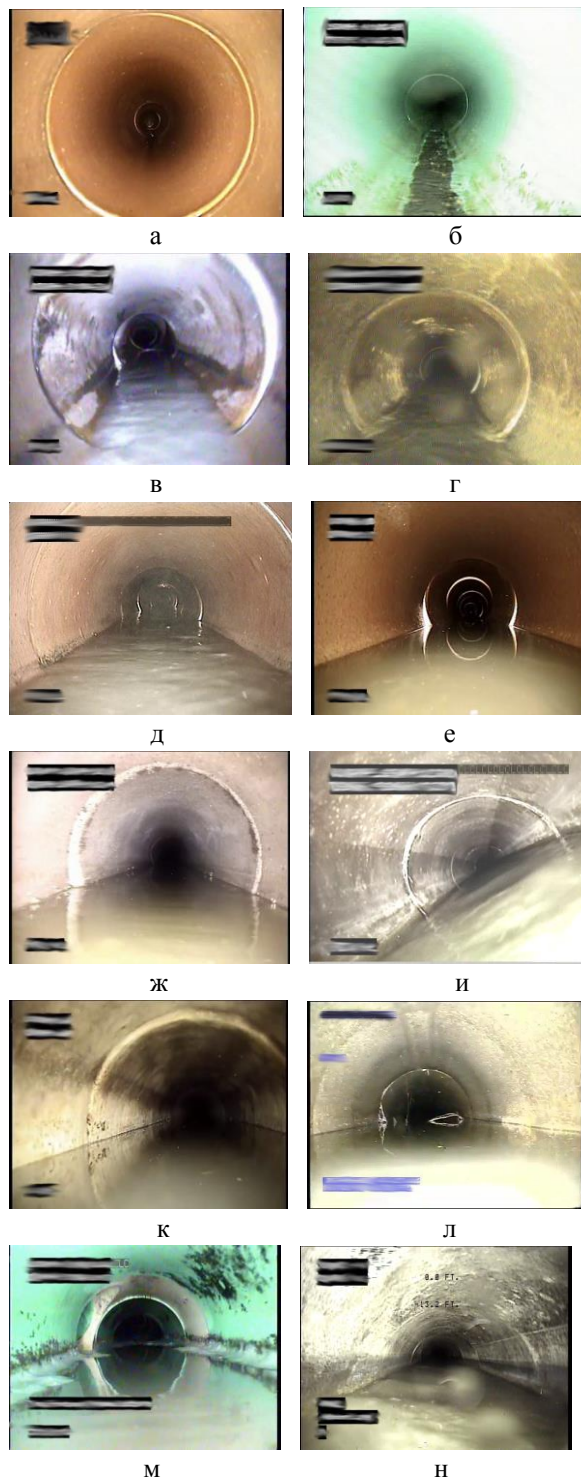


Рис. 8. Приклади зображень кожного з класів:

- а – клас X_1^0 ; б – клас X_2^0 ; в – клас X_3^0 ;
- г – клас X_4^0 ; д – клас X_5^0 ; е – клас X_6^0 ;
- ж – клас X_7^0 ; и – клас X_8^0 ; к – клас X_9^0 ;
- л – клас X_{10}^0 ; м – клас X_{11}^0 ; н – клас X_{12}^0

Кожна модель має ряд гіперпараметрів, що визначають конфігурацію і ємність моделі. Спочатку здійснюється підбір оптимальних гіперпараметрів для кожної архітектури, а потім порівняльний аналіз результатів, отриманих на різних архітектурах з цими параметрами. Навчання моделей відбувається протягом 60 епох. При цьому гіперпараметри підібрані так, щоб уникнути помітного ефекту перенавчання.

Як екстрактор ознак окремих кадрів використано згорткову частину моделі MobileNet з коефіцієнтом ємності $\alpha = 0,25$ та вхідною роздільною здатністю 160×160 пікселів. Темпоральна згорткова мережа містить одноступеневу конфігурацію (single stage) з 7 базових блоків (BB). Кількість ознакових каналів у згорткових фільтрах першого шару задана як $C = 128$, що дорівнює розмірності ознакового опису зображень. Останній шар темпоральної мережі з'єднаний з Dense шаром, що містить 128 вузлів з сигмоїдною функцією активації. Рекурентні мережі GRU та LSTM містять один шар з 128 комірок. При цьому стани всіх комірок надходять до Dense шару, що так само містить 128 вузлів з сигмоїдною функцією активації. Після сигмоїдного шару будь-якої моделі використовуються два додаткові вихідні шари, які аналогічні до тих, що показані на рис. 2, а, і залежать від вибору методу навчання.

Рис. 9 показує зміну значення метрики F1 (2) на тестовій вибірці під час навчання класифікатора окремих кадрів за традиційним (baseline) та запропонованим (рис. 7) методами. При цьому тестова вибірка формувалась шляхом вибірки з набору кожного класу 15 % оригінальних зразків.

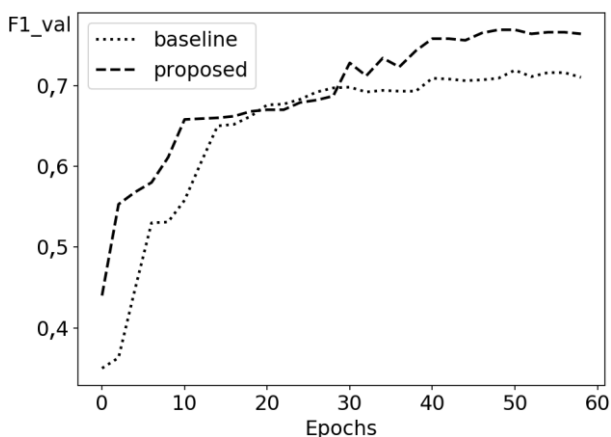


Рис. 9. Залежність значення метрики F1 (2) на тестовій вибірці від кількості епох навчання для класифікатора контексту без врахування сусідніх кадрів

Аналіз рис. 9 показує, що до 30-ї епохи ефективність навчання обох методів є приблизно однаковою. Однак після 30-ї епохи навчання у запропоно-

ваному методі використовується 3-тя фаза (рис. 7), яка забезпечує підвищення значення F1-метрики моделі на 5,4 %. Тому для інтеграції ознакового опису кадрів у високорівневу нейромережу буде використано екстрактор ознак, навчений в рамках запропонованого методу.

Рис. 10 ілюструє зміни значення метрики F1 (2) на тестовій вибірці (F1_val) під час навчання класифікатора кадрів з урахуванням сусідніх кадрів за традиційним (baseline) методом, але для різних архітектур темпоральної мережі: темпоральної мережі з каузальними згортками (F1_val_cnn_causal); темпоральної мережі з регулярними згортками (F1_val_cnn_regular); рекурентної мережі GRU (F1_val_gru); рекурентної мережі LSTM (F1_val_lstm). Тестова вибірка формувалась так само шляхом вибірки з набору кожного класу 15 % зразків. Аугментована навчальна вибірка кожного класу становить 2000 зразків.

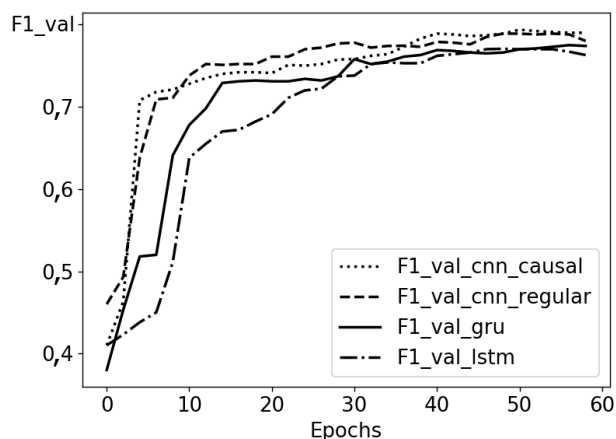


Рис. 10. Залежність значення метрики F1 (2) на тестовій вибірці від кількості епох навчання за традиційним методом навчання для класифікатора контексту з урахуванням сусідніх кадрів

Аналіз рис. 10 показує, що точнісні характеристики рекурентних мереж на перших епохах навчання зростали повільніше ніж у згорткових. На останніх епохах навчання зростання ефективності рекурентних мереж припинилося і досягло практично однакових значень. Згорткові мережі також припинили нарощування ефективності на останніх епохах, однак забезпечили дещо вищі показники F1-метрики, але суттєвої різниці між ефективністю каузальної та регулярної архітектури не спостерігається. Отримане значення F1-метрики для регулярної темпоральної мережі становить 0,789, в той час як значення F1-метрики для LSTM та GRU моделей становить 0,770 та 0,775 відповідно.

Рис. 11 показує зміну значення F1-метрики (2) на тестовій вибірці під час навчання класифікатора

кадрів з урахуванням сусідніх кадрів за запропонованим методом (рис. 7), але для різних архітектур темпоральної мережі (таких же, як на рис. 10).

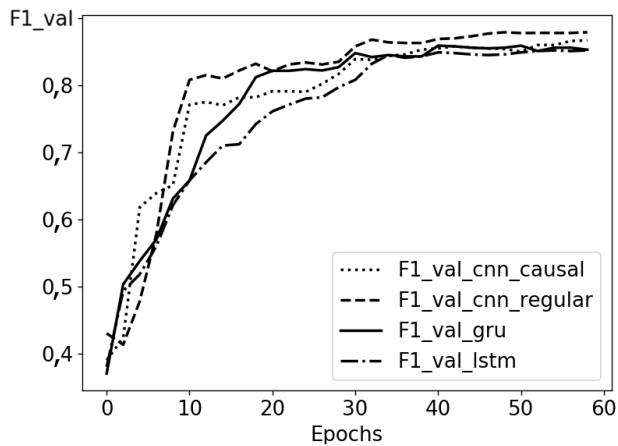


Рис. 11. Залежність значення F1-метрики (2) на тестовій вибірці від кількості епох навчання за запропонованим методом для класифікатора контексту з урахуванням сусідніх кадрів

Аналіз рис. 11 показує, що точнісні характеристики рекурентних мереж так само поступаються точнісним характеристикам згорткових мереж. Серед згорткових мереж лідером є модель з регулярною структурою. При цьому максимальне значення F1-метрики (2) для моделі з регулярною структурою за запропонованим методом навчання дорівнює 0,879, що на 9 % перевищує результат традиційного методу навчання.

На рис. 12 показано матрицю помилок (confusion matrix) для отриманого класифікатора контексту з урахуванням сусідніх кадрів на основі темпоральної згорткової мережі з регулярною структурою згорток.

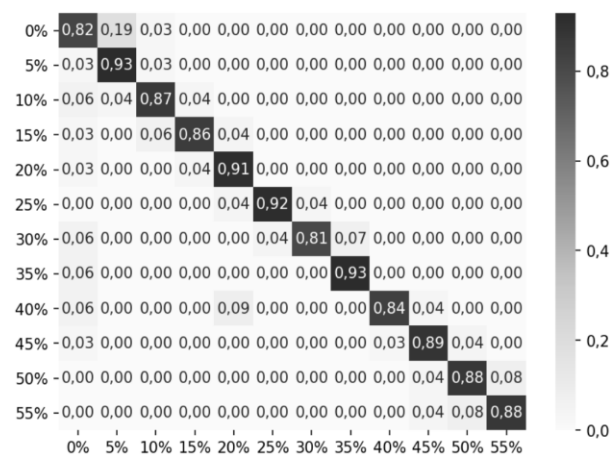


Рис. 12. Матриця помилок для оптимального класифікатора рівнів води з урахуванням сусідніх кадрів

Аналіз рис. 12 показує, що в основному перетинаються сусідні класи, тобто помилка виникає на межі переходу одного рівня води до іншого. Цей результат є очікуваний і може бути пов'язаний з помилками розмітки даних внаслідок суб'єктивного фактору оператора. Також на зображеннях труб з чистою прозорою водою є хибне розпізнавання рівня 0 %.

Таким чином, доведено перевагу використання запропонованої моделі для аналізу рівня води і запропонованого багатофазного методу її навчання. Експериментально показано, що найбільш ефективною моделлю для класифікаційного аналізу рівня води на відеокadraх є комбінація згорткової 2D мережі і темпоральної згорткової мережі з регулярними згортками. При цьому згорткова темпоральна мережа продемонструвала вищу ефективність врахування міжкадрових залежностей порівняно з рекурентними мережами. Отримана точність прийнятна до практичного використання, оскільки подібні помилки допускають і самі оператори.

Висновки

Наукова новизна роботи полягає в нових моделях і методах класифікаційного аналізу рівня води під час відеоінспекції стічних труб за умов незбалансованості та обмеженого обсягу розмічених даних.

Запропонована модель містить 2D згорткову мережу, що формує на виході сигмоїдного шару ознаковий опис кадрів, та 1D темпоральну мережу з сигмоїдним та округлюючим вихідними шарами. Як темпоральна мережа розглядаються рекурентні мережі LSTM та GRU, а також згорткові 1D мережі з регулярними та каузальними дірвими згортками. При цьому за результатами тестування темпоральна мережа з регулярними згортками показала найкращі точнісні характеристики. Вирішувальні правила в цьому випадку побудовані у радіальному базисі згідно принципів інформаційно-екстремального навчання.

На першому етапі навчання відбувається навчання з нормалізованою триплетною функцією втрат та регуляризуючою складовою для штрафування за помилку округлення вихідного сигналу мережі до двійкового коду. Наступний етап потрібний для визначення двійкового коду класу відповідно до принципів завадозахищеного кодування з самовиправленням помилок, але з урахуванням внутрікласових і міжкласових відношень. Обчислений еталонний вектор кожного класу використовується як цільова розмітка зразка для подальшого навчання з використанням об'єднаної крос-ентропійної функції втрат. Останній етап машинного навчання пов'язаний з оптимізацією параметрів вирішувальних

правил за інформаційним критерієм для врахування меж відхилення двійкового подання спостережень кожного класу від відповідних еталонних векторів. Експериментально підтверджено перевагу даного методу навчання порівняно з традиційним підходом.

Практичне значення роботи полягає в підвищенні точності моделі класифікаційного аналізу рівня води в стічних трубах під час відеоінспекції за умов обмеженого обсягу розмічених даних та неповної визначеності, пов'язаної з довільними умовами спостереження.

Подальші дослідження будуть спрямовані на покращення точнісних характеристик моделі за рахунок попереднього навчання без вчителя на великому обсязі нерозмічених навчальних даних.

Література

1. Moradi, S. Review on Computer Aided Sewer Pipeline Defect Detection and Condition Assessment [Text] / S. Moradi, T. Zayed, F. Golkhoo // *Infrastructures*. – 2019. – Vol 4, No. 1. – Article Id: 10. DOI: 10.3390/infrastructures4010010.

2. Haurum, J. B. A Survey on image-based automation of CCTV and SSET sewer inspections [Text] / J. B. Haurum, T. B. Moeslund // *Automation in Construction*. – 2020. – Vol. 111. – Article Id: 103061. DOI: 10.1016/j.autcon.2019.103061.

3. 2017 Infrastructure Report Card – Wastewater [Electronic resource] / American Society of Civil Engineers (ASCE). – 2017. – Available at: <https://www.infrastructurereportcard.org/wp-content/uploads/2017/01/Wastewater-Final.pdf>. – (accessed on 5 May 2021).

4. Cheng, J. C. P. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques [Text] / J. C. P. Cheng, M. Wang // *Automation in Construction*. – 2018. – Vol. 95. – P. 155-171. DOI: 10.1016/j.autcon.2018.08.006.

5. Lim, B. Time-series forecasting with deep learning: a survey [Text] / B. Lim, S. Zohren // *Philosophical Transactions of the Royal Society A*. – 2021. – Vol. 379, Issue 2194. DOI: 10.1098/rsta.2020.0209.

6. Robust adaptive flow line detection in sewer pipes [Text] / S. Kirstein, K. Müller, M. Walecki-Mingers, T. M. Deserno // *Automation in Construction*. – 2012. – Vol. 21. – P. 24-31. DOI: 10.1016/j.autcon.2011.05.009.

7. Halfawy, M. R. Integrated Vision-Based System for Automated Defect Detection in Sewer Closed Circuit Television Inspection Videos [Text] / M. R. Halfawy, J. Hengmeechai // *Journal of Computing in Civil Engineering*. – 2015. – Vol. 29, No. 1. DOI: 10.1061/(ASCE)CP.1943-5487.0000312.

8. Measurement of Wastewater Discharge in Sewer Pipes Using Image Analysis [Text] / H. W. Ji, S. S. Yoo, B.-J. Lee, D. D. Koo, J.-H. Kang // *Water*. – 2020. – Vol. 12, No. 6. – Article Id: 1771. DOI: 10.3390/w12061771.

9. DeepShoe: An improved Multi-Task View-invariant CNN for street-to-shop shoe retrieval [Text] / H. Zhan, B. Shi, L.-Y. Duan, A. C. Kot // *Computer Vision and Image Understanding*. – 2019. – Vol. 180. – P. 23-33. DOI: 10.1016/j.cviu.2019.01.001.

10. Challenging the Adversarial Robustness of DNNs Based on Error-Correcting Output Codes [Text] / B. Zhang, B. Tondi, X. Lv, M. Barni // *Security and Communication Networks*. – 2020. – Vol. 2020. – Article Id: 8882494. DOI: 10.1155/2020/8882494.

11. Інформаційно-аналітична система оцінювання відповідності сучасним вимогам навчального контенту спеціальності кібербезпека [Текст] / А. С. Довбиш, І. В. Шелехов, Ю. О. Хібовська, О. В. Матяш // *Радіоелектронні і комп'ютерні системи*. – 2021. – № 1. – С. 70-80. DOI: 10.32620/reks.2021.1.06.

References

1. Moradi, S., Zayed, T., Golkhoo, F. Review On Computer Aided Sewer Pipeline Defect Detection And Condition Assessment. *Infrastructures*, 2019, vol. 4, no. 1, Article Id: 10. DOI: 10.3390/infrastructures4010010.

2. Haurum, J. B., Moeslund, T. B. A Survey on image-based automation of CCTV and SSET sewer inspections. *Automation in Construction*, 2020, vol. 111, Article Id: 103061. DOI: 10.1016/j.autcon.2019.103061.

3. American Society of Civil Engineers (ASCE). 2017 Infrastructure Report Card – Wastewater. Available at: <https://www.infrastructurereportcard.org/wp-content/uploads/2017/01/Wastewater-Final.pdf> (accessed 05.05.2021).

4. Cheng, G. C. P. Wang M. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Automation in Construction*, 2018, vol. 95, pp. 155-171. DOI: 10.1016/j.autcon.2018.08.006.

5. Lim, B., Zohren, S. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 2021, vol. 379, no. 2194. 14 p. DOI: 10.1098/rsta.2020.0209.

6. Kirstein, S., Müller, K., Walecki-Mingers, M., Deserno, T. M. Robust adaptive flow line detection in sewer pipes. *Automation in Construction*, 2012, vol. 21, pp. 24-31. DOI: 10.1016/j.autcon.2011.05.009.

7. Halfawy, M. R., Hengmeechai, J. Integrated Vision-Based System for Automated Defect Detection

in Sewer Closed Circuit Television Inspection Videos. *Journal of Computing in Civil Engineering*, 2015, vol. 29, no. 1. DOI: 10.1061/(ASCE)CP.1943-5487.0000312.

8. Ji, H. W., Yoo, S. S., Lee, B.-J., Koo, D. D., Kang, J.-H. Measurement of Wastewater Discharge in Sewer Pipes Using Image Analysis. *Water*, 2020, vol. 12, no. 6, Article Id: 1771. DOI: 10.3390/w12061771.

9. Zhan, H., Shi, B., Duan, L.-Y., Kot, A. C. Deepshoe: An Improved Multi-Task View-Invariant CNN For Street-To-Shop Shoe Retrieval. *Computer Vision and Image Understanding*, 2019, vol. 180, pp. 23-33. DOI: 10.1016/j.cviu.2019.01.001.

10. Zhang, B., Tondi, B., Lv, X., Barni, M. Challenging the Adversarial Robustness of DNNs Based

on Error-Correcting Output Codes. *Security and Communication Networks*, 2020, vol. 2020, Article Id: 8882494. DOI: 10.1155/2020/8882494.

11. Dovbysh, A. S., Shelekhov, I. V., Khibov'ska, Yu. O., Matyash, O. V. Informatsiynno-analitychna systema otsynuyannya vidpovidnosti suchasnym vymoham navchal'noho kontentu spetsial'nosti kiberbezpeka [Information and analytical system for assessing the compliance of educational content specialties ciber security with modern requirements]. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2021, vol. 1, pp. 70-80. DOI: 10.32620/reks.2021.1.06.

Надійшла до редакції 5.05.2021 розглянута на редколегії 20.05.2021

МОДЕЛЬ И МЕТОД ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИОННОГО АНАЛИЗА УРОВНЯ ВОДЫ В СТОЧНЫХ ТРУБАХ ПО ДАННЫМ ВИДЕО ИНСПЕКЦИИ

*В. В. Москаленко, Н. А. Зарецкий, А. Г. Коробов, Я. Ю. Ковальский,
А. Ф. Шаехов, В. А. Семашко, А. А. Паньч*

Разработаны и исследованы модели и метод обучения для классификационного анализа уровня воды на кадрах видео инспекции сточных труб. Объектом исследования является процесс распознавания уровня воды с учетом пространственно-временного контекста во время инспекции сточных труб. Предметом исследования является модель и метод машинного обучения для классификационного анализа уровня воды на видео-последовательностях инспекции труб в условиях ограниченного и несбалансированного набора обучающих данных. Предложено четырехэтапный алгоритм обучения классификатора. На первом этапе обучения происходит обучение с нормализованной триплетной функцией потерь и регуляризирующей составляющей для штрафования за ошибку округления выходного сигнала сети к двоичному коду. Следующий этап нужен для определения двоичного кода класса в соответствии с принципами помехозащищенного кодирования с самоисправлением ошибок, но с учетом внутриклассовых и межклассовых отношений. Вычисленный эталонный вектор каждого класса используется как целевая разметка образца для дальнейшего обучения с использованием объединенной кросс-энтропийной функции потерь. Последний этап машинного обучения связан с оптимизацией параметров решающих правил по информационному критерию для учета границ отклонения двоичного представления наблюдений каждого класса от соответствующих эталонных векторов. Как модель классификатора рассматривается сочетание 2D свёрточного экстрактора признаков кадра с темпоральной сетью для анализа межкадровых зависимостей. При этом выполняется сравнение различных вариантов темпоральной сети. Рассматриваются 1D регулярная свёрточная сеть с дырявыми свертками, 1D каузальная свёрточная сеть с дырявыми свертками, рекуррентная LSTM-сеть, рекуррентная GRU-сеть. Сравнение эффективности моделей происходит по микро-усредненной метрике F1, вычисляемой на тестовой выборке. Результаты, полученные на наборе данных от компании Ace Pipe Cleaning (Канзас Сити, США), подтверждают пригодность модели и метода обучения к практическому использованию, полученное значение F1-метрики равно 0,88. При этом результаты обучения по предложенному методу сравнивались с результатами, полученными традиционным методом. Было показано, что предложенный метод обеспечивает увеличение значения микро-усредненной F1-метрики на 9 %.

Ключевые слова: сточные трубы; инспекция; классификационный анализ; свёрточная нейронная сеть; рекуррентная нейронная сеть; функция потерь; регуляризация; информационно-экстремального машинное обучение.

MODEL AND TRAINING METHOD FOR WATER LEVEL CLASSIFICATION IN SEWER PIPES BASED ON VIDEO INSPECTION DATA

*V. Moskalenko, M. Zaretskyi, A. Korobov, Y. Kovalskyi,
A. Shaiekhov, V. Semashko, A. Panych*

Models and training method for water level classification analysis on footage of sewage pipe inspections have been developed and investigated. The object of the research is the process of water level recognition, taking into account the spatial and temporal context during the inspection of sewage pipes. The subject of the research is a

model and machine learning method for water level classification analysis on video sequences of pipe inspections under conditions of a limited size and unbalanced set of training data. A four-stage algorithm for training the classifier is proposed. At the first stage of training, training takes place with a softmax triplet loss function and a regularizing component to penalize the rounding error of the network output to a binary code. The next step is to define a binary code (reference vector) for each class according to the principles of error correcting output codes, but taking into account intraclass and interclass relations. The computed reference vector of each class is used as the target label of the sample for further training using the joint cross-entropy loss function. The last stage of machine learning involves optimizing the parameters of the decision rules based on the information criterion to account for the boundaries of deviation of the binary representation of the observations of each class from the corresponding reference vectors. As a classifier model, a combination of 2D convolutional feature extractor for each frame and temporal network to analyze inter-frame dependencies is considered. The different variants of the temporal network are compared. We consider 1D regular convolutional network with dilated convolutions, 1D causal convolutional network with dilated convolutions, recurrent LSTM-network, recurrent GRU-network. The performance of the models is compared by the micro-averaged metric F1 computed on the test subset. The results obtained on the data set from Ace Pipe Cleaning (Kansas City, USA) confirm the suitability of the model and training method for practical use, the obtained value of F1-metric is 0.88. The results of training by the proposed method were compared with the results obtained by the traditional method. It was shown that the proposed method provides a 9 % increase in the value of micro-averaged F1-measure.

Keywords: sewer pipe; inspection; classification analysis; convolutional neural network; recurrent neural network; loss function; regularization; information-extreme machine learning.

Москаленко В'ячеслав Васильович – канд. техн. наук, доцент, доцент каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Зарецький Микола Олександрович – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Коробов Артем Геннадійович – канд. техн. наук, асистент каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Ковальський Ярослав Юрійович – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Шаєхов Артур Фанісович – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Семашко Віктор Анатолійович – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Панич Андрій Олександрович – асистент каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Viacheslav Moskalenko – PhD, associate professor of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: v.moskalenko@cs.sumdu.edu.ua, ORCID: 0000-0001-6275-9803, Scopus Author ID: 57189099775.

Mykola Zaretskyi – PhD student of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: n.zaretskij@gmail.com, ORCID: 0000-0001-9117-5604, Scopus Author ID: 57213687285.

Artem Korobov – PhD, assistant of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: a.korobov@cs.sumdu.edu.ua, ORCID: 0000-0003-3239-1977.

Yaroslav Kovalskyi – PhD student of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: socratus43@gmail.com, ORCID: 0000-0002-5345-5186.

Artur Shaikhov – PhD student of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: escander7@gmail.com, ORCID: 0000-0003-3277-0264.

Viktor Semashko – PhD student of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: viktor.s.5994@gmail.com, ORCID: 0000-0002-9765-876X, Scopus Author ID: 57210589608.

Andrii Panych – assistant of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: a.panych@ksu.sumdu.edu.ua, ORCID: 0000-0003-2511-5763.