



**О. П. Максимів**

*Львівський державний університет безпеки життєдіяльності, м. Львів, Україна*

## КАСКАДНИЙ МЕТОД ДЕТЕКТУВАННЯ ПОЛУМ'Я У ВІДЕОПОТОЦІ З ВИКОРИСТАННЯМ ГЛИБОКИХ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Запропоновано каскадний бінарний класифікатор, який базується на використанні суміші Гауссового розподілу, примітивів Хаара та глибоких згорткових нейронних мереж. Застосування такого підходу дає змогу, з одного боку, здійснювати моніторинг середовища в режимі, наближеному до реального часу, а з іншого – забезпечити досить високий показник ефективності виявлення полум'я на відеозображеннях (92,7 %). На першому етапі, за допомогою використання примітивів Хаара, до попередньо виокремлених рухомих регіонів відеозображення генеруються так звані регіони інтересу. На другому етапі отримані регіони інтересу передаються для класифікації на вхід до глибокої згорткової нейронної мережі, яка формує висновки про наявність або відсутність полум'я на зображенні. У межах проведення дослідження запропоновано модифіковану модель нейронної мережі – SqueezeNet. Виявлено, що за допомогою використання трансферного підходу під час навчання нейронної мережі, існує змога мінімізувати кількість хибних викликів, особливо на об'єктах, які візуально можуть нагадувати полум'я, та зменшити часові затрати, які необхідні для її навчання. Для покращення ефективності роботи нейронної мережі здійснено низку тонких налаштувань (ансамбль з нейронних мереж, поворот зображень, зменшення показника швидкості навчання, кадрування та передискретизація), що сумарно дало змогу покращити її ефективність на 2,4 %.

**Ключові слова:** виявлення руху; примітиви Хаара; трансферне навчання; інтелектуальні системи; комп'ютерний зір.

**Вступ.** Протягом останніх років спостерігають значне збільшення об'єму потокової інформації, отримуваних з камер відеоспостереження. Інформація такого типу містить в собі значний потенціал для її використання в контексті забезпечення безпеки життєдіяльності людини та прийняття управлінських рішень. Проте, через потребу постійного контролю вхідної інформації, людина не може гарантувати якісного та оперативного реагування на можливі загрози. Ці факти призвели до значного збільшення кількості інтелектуальних систем відеоспостереження, які можуть самостійно виконувати покладені на них завдання.

Одним із найпопулярніших напрямків відеоаналітики є виявлення конкретних подій або об'єктів. Так, значного поширення набули системи детектування пішоходів, розпізнавання облич, автоматичного керування автомобілем тощо. На нашу думку, не менш важливою є реалізація автоматичної системи детектування полум'я. Порівняно з наявними пожежними давачами, ідентифікація полум'я у відеопотоці має низку переваг: широкий обсяг контрольованої території, можливість встановлення на відкритих ділянках, легкість встановлення, спроможність одразу ж перевірити правильність спрацювання та оцінити можливу загрозу.

**Аналіз останніх досліджень та публікацій.** Упродовж останніх років триває підвищений рівень зацікавленості вчених до розроблення алгоритмів та методів детектування полум'я на зображеннях. Оскільки найпомітнішою візуальною ознакою пожежі є колір, то зви-

чайно ж більшість публікацій присвячена використанню різноманітних кольорних моделей, просторової інформації про неї тощо (Maksymiv, Rak & Peleshko, 2016; Chen et al., 2004; Qi & Ebert, 2009; Celik & Demirel, 2009). Проте, враховуючи неоднорідність кольору горіння (колір полум'я чи диму залежить від його температури та речовини, яка горить), такий підхід супроводжується низкою хибних спрацювань.

Значною мірою ефективність роботи системи детектування пожежі у відеопотоці можна підвищити завдяки додатковому використанню алгоритмів, які дають змогу виокремити рухомі об'єкти. Для цього використовують як найпростіші методи виділення рухомих об'єктів (різниця кадрів (Lipton et al., 1998), порівняння з фоном (Piccardi, 2004), мікстура гауссового розподілу (KaewTraKulPong & Bowden, 2002)), так і такі, які потребують значних обчислювальних затрат (базуються на стохастичній оцінці зміни форми та руху вогню (Chen et al., 2003), набору правил руху (Celik et al., 2007), оптичного потоку (Horn & Schunck, 1981)).

Щодо сучасних підходів до детектування об'єктів на зображеннях, то, безумовно, потрібно згадати про використання згорткових нейронних мереж. Так, у 2012 р. Алекс Крижевський, навчивши згорткову мережу, яка складалася лише з 8 згорткових і 2 повнозв'язних шарів вдалося досягнути показника 16 % (найближча команда з показником 26 %) (Russakovsky et al., 2015). Уже в 2015 р. компанія Microsoft за допомогою глибоких згорткових мереж вдалося перемогти людину з її

### Інформація про авторів:

**Максимів Олексій Петрович**, ад'юнкт кафедри управління інформаційною безпекою. Email: [aleks.maksymiv@gmail.com](mailto:aleks.maksymiv@gmail.com)

**Цитування за ДСТУ:** Максимів О. П. Каскадний метод детектування полум'я у відеопотоці з використанням глибоких згорткових нейронних мереж. Науковий вісник НЛТУ України. 2017. Вип. 27(9). С. 115–120.

**Citation APA:** Maksymiv, O. P. (2017). Cascade-Based Method for Flame Detection in Video Sequences Using Deep Convolutional Neural Networks. *Scientific Bulletin of UNFU*, 27(9), 115–120. <https://doi.org/10.15421/40270925>

рівнем помилок в 5,33 %. Відповідно, за період з 2012 по 2015 рр. можна простежувати тенденцію, що чим більше шарів у згортковій мережі, тим краще вона працює. Проте у 2016 р. в конкурсі Imagenet перемогла команда з рівнем помилок 2,9 %, яка використовувала ансамбль з 9 нейронних мереж (He et al., 2016).

Щодо використання нейронних мереж для детектування полум'я, то можемо відзначити лише наукові роботи (Qingjie et al., 2016; Frizzi et al., 2016; Maksymiv, Rak & Peleshko, 2016). Проте, незважаючи на високі показники ефективності, треба зазначити, що їх використання неможливе в режимі реального часу, що нівелює можливість їх використання в системі безпеки життєдіяльності.

Узагальнюючи науковий досвід, можемо зазначити, що нинішні методи виявлення пожежі у відеопотоці мають недостатню високу якість детектування або характеризуються низькою швидкістю, що не дає змоги використовувати їх для побудови автоматизованих систем детектування полум'я. Вирішенням зазначених вище проблем може послужити використання глибоких згорткових нейронних мереж, за умов забезпечення роботи в режимі реального часу.

**Методика дослідження.** Більшість сучасних детекторів об'єктів на зображеннях складаються з двох частин: екстрактор ознак та класифікатор. Відповідно до цього підходу, роботу детектора полум'я було розділено на два основні етапи. На першому етапі здійснюється пошук регіонів (ROI), які можуть візуально нагадувати вогонь. Для цього, насамперед визначаються лише рухомі об'єкти, які передаються на вхід класифікатора, побудованого на основі примітивів Хаара, де і визначається візуальна схожість з вогнем. Оскільки використання цього підходу є недостатнім для забезпечення ефективної роботи детектора, то отримані регіони, які попередній класифікатор відніс до класу "пожежа", передаються на вхід до нейронної мережі, яка і надає остаточний висновок про наявність або відсутність полум'я на зображенні.

**Виділення рухомих об'єктів.** Аналіз лише рухомих об'єктів на зображенні дає змогу зменшити загальний об'єм інформації, яку потрібно проаналізувати класифікатору для надання остаточного висновку. Для цього було вирішено використовувати суміш Гауссового розподілу (GMM). Кожен піксель на зображенні моделюється на основі окремої Гауссової суміші для фону, переднього плану та тіней. Базуючись на часі існування і дисперсії кожного гауссіана в суміші, можна визначити, які з них належать до фону. Пікселі, значення яких не вкладаються у фоновий розподіл, вважають передньоплановими доти, поки не з'явиться гауссіан, що дає змогу з достатньою точністю віднести їх до фону. Приклад виділення рухомих об'єктів зображено на рис. 1. Ймовірність, що певний піксель має інтенсивність  $x_t$  на проміжку часу  $t$ , можна описати як (Stauffer & Grimson, 1999), а саме:

$$P(x_t) = \sum_{i=1}^K w_i \cdot \eta(x_t, \eta_i, \Sigma_i),$$

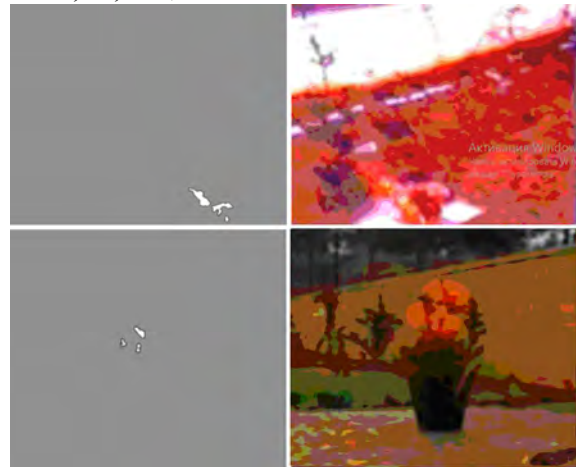
де:  $K$  – кількість гауссових сумішей;  $w_i$  – оцінка ваг  $i$ -го гауссового розподілу;  $\eta$  – середнє значення для  $i$ -го гауссового розподілу;  $\Sigma_i$  – матриця коваріації для  $i$ -го гауссового розподілу;  $\eta$  – функція щільності гауссового розподілу:

$$\eta(x_t, \eta, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - \eta)^T \Sigma^{-1} (x_t - \eta)},$$

$n$  – кількість вимірювань для інтенсивності пікселя  $x$  і

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha(\langle M_{k,t} \rangle),$$

де:  $\alpha$  – швидкість навчання;  $\alpha = 1$ , якщо моделі співпали, і 0, якщо ні.



**Рис. 1.** Приклад детектування лише рухомих ділянок зображення. Справа – оригінальні зображення, зліва – отриманий результат

**Генерація кандидатів-регіонів.** До регіонів, на яких спостерігався рух, для визначення їх візуальної схожості з полум'ям, застосовують бінарний класифікатор, який ґрунтується на використанні каскадів Хаара. На відміну від загальноприйнятих підходів щодо віднесення об'єкта до категорії "пожежа", на основі застосування різноманітних кольорних моделей, цей підхід дає змогу уникнути прив'язки до кольору вогню, який може набувати практично будь-якої кольірної гами залежно від температури та речовини, яка горить. Зокрема, потрібно зазначити, що завдяки використанню інтегрального представлення зображення, отримувані ознаки можуть опрацьовуватися в режимі реального часу навіть на пристроях із слабкою обчислювальною спроможністю.

Інтегральне представлення зображення є матрицею, розмірність якої збігається з розмірністю вихідного зображення, а елементи матриці обчислюють за формулою (Chen et al., 2004), а саме

$$H(x, y) = \sum_{i=0, y=0}^{i \leq x, j \leq y} I(i, j),$$

де  $I(i, j)$  – яскравість пікселів вихідного зображення.

На етапі генерації регіонів, які можна віднести до категорії "пожежа", важливим етапом є навчання класифікатора, який зможе розпізнати усі можливі ділянки зображення, на яких присутній вогонь. Навчання класифікатора такого типу буде прийнятним навіть, якщо буде спостерігатися високий рівень хибно позитивних спрацювань. Для цього ми відібрали тренувальні екземпляри розмірами  $12 \times 12$  пікселів, які характеризуються варіативністю якості зображення та візуального вигляду вогню (рис. 2). Загальна кількість зображень, відібраних для навчання, становила: 3547 позитивних та 9000 негативних.

Сам по собі каскад Хаара – це набір примітивів, для яких розраховується значення їх згортання із зображенням. Використовують найпростіші примітиви, що складаються з прямокутників і мають всього два рівні (+1 та -1). При цьому кожен прямокутник використовується

кілька разів різного розміру. Під згортанням мають на увазі

$$s = X - Y,$$

де:  $Y$  – сума елементів зображення в темній області, а  $X$  – сума елементів зображення у світлій області (можна так само брати  $X/Y$ , тоді буде стійкість в разі зміни масштабу).

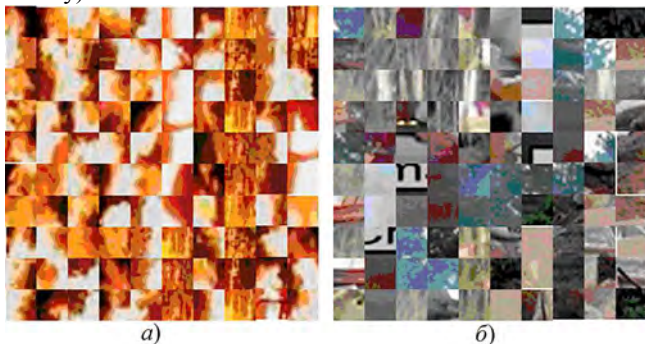


Рис. 2. Приклад зображень з датасету

Відповідно до зазначеного вище, було розроблено набір примітивів, який дає змогу описати можливі візуальні ознаки полум'я на зображенні (рис. 3).

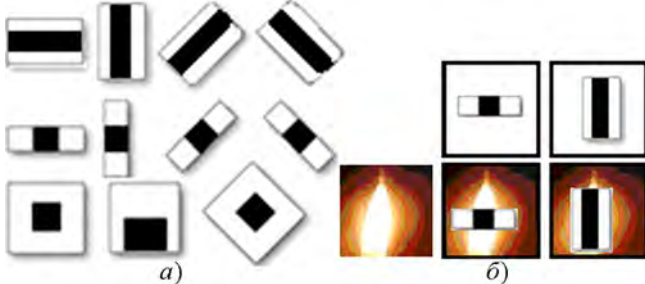


Рис. 3. Запропоновані ознаки Хаара

Унаслідок навчання було отримано 21-рівневий каскад Хаара, що, окрім швидкої роботи, характеризувався високим рівнем детектування полум'я на зображенні (99,4%). Проте, треба зазначити, що спостерігається значний рівень хибно позитивних спрацювань (~20% від усіх викликів), особливо на таких об'єктах, як вікна в денний час (рис. 4), різноманітні засоби освітлення, гірлянди тощо. Зазначений недолік зумовлює використання додаткового класифікатора для усіх регіонів, які було отримано після закінчення цього етапу.



Рис. 4. Приклад генерації регіонів інтересу

**Класифікація.** На етапі класифікації використовуємо згорткову мережу, яка є ансамблем, з трьох нейронних моделей для надання остаточного висновку щодо приналежності отриманих регіонів інтересу з попереднього етапу до категорії "пожежа". Зокрема, для

покращення ефективності роботи такого детектора, було використано низку тонких налаштувань мережі та використання підходу трансферного навчання до двох нейронних моделей.

**Трансферне навчання.** Під час навчання CNN важливо виділити не тільки ознаки, які можуть описати шуканий об'єкт, а й ознаки інших об'єктів, що дасть змогу мінімізувати кількість хибних спрацювань детектора, особливо під час розпізнавання зображень, які можуть візуально нагадувати його (у цьому разі це можуть бути гірлянди, різноманітні засоби освітлення та інші об'єкти з яскравими текстурями). Так, у роботі (Yosinski et al., 2014) зазначено, що кожен прихований шар у згортковій нейронній мережі має виразні ознаки, які пов'язані з особливостями нижніх шарів забезпечувати можливість виділення загальних ознак. Своєю чергою, вищі шари несуть інформацію, яка є конкретнішою до вихідної задачі класифікації.

Відповідно до зазначеної вище парадигми, можемо використати апріорну згорткову нейронну мережу, яка натренована на основі датасету Imagenet, для розпізнавання більш ніж 1000 класів об'єктів. Зокрема, потрібно врахувати той факт, що за умов використання уже навченої нейронної мережі, її подальше донавчання зменшить часові затрати від днів або навіть тижнів до кількох хвилин.

**Вибір архітектури нейронної мережі.** Враховуючи поставлені вимоги щодо побудови детектора полум'я (висока ефективність детектування та можливість роботи в режимі, наближеному до реального часу), використання загальноприйнятих архітектур нейронних мереж (GoogleNet (Szegedy et al., 2015), VGG-16 тощо) є проблематичним, оскільки вони характеризуються великими розмірами та кількістю параметрів, що може спричинити її неефективність на машинах із невеликими обчислювальними ресурсами.

Відповідно, для розв'язання поставленої задачі, вирішили використати мережу SqueezeNet (Iandola et al., 2016). Цей вибір зумовлений тим, що ця мережа дає змогу працювати з однаковою ефективністю глибокої мережі AlexNet, проте має в 50 разів менше параметрів. Основними стратегіями до побудови такої архітектури є:

- 1) заміна фільтрів  $3 \times 3$  на  $1 \times 1$  фільтри, що дає змогу зменшити кількість параметрів у 9 разів;
- 2) зменшення кількості каналів більш стиснутими модулями;
- 3) пониження дискретизації останніх шарів мережі. Стратегії 1 та 2 націлені на зменшення кількості параметрів мережі, третя стратегія – на покращення її ефективності, незважаючи на зменшення параметрів.

Для покращення ефективності детектування вирішено використовувати модифіковану архітектуру мережі SqueezeNet шляхом впровадження залишкових (residual) блоків (рис. 5). Кінцеву архітектуру мережі зображено на рис. 6.

**Доналаштування нейронної мережі.** Під час тонкого налаштування ваг нейронної мережі, починаємо зі задалегідь підготовлених моделей, які навчені на оригінальному наборі даних (тобто зображення без будь-яких спотворень). Здійснили тонке налаштування перших  $N$  шарів моделі на спотвореному наборі даних, а в інших шарах використовували фіксовані параметри. Основною причиною такої фіксації є спостереження, що під час розмиття і шуму останні шари набувають

більшого впливу на низькорівневі властивості, такі як колір, границі та текстурні ознаки. Проте ці спотворення мають незначний вплив на інформацію вищого рівня, таку як семантичні значення зображень (Song Han, 2016). Отже, під час тонкого налаштування ваг нейронної мережі орієнтуємося на вихідні шари SqueezeNet, які містять в собі інформацію нижчого рівня.

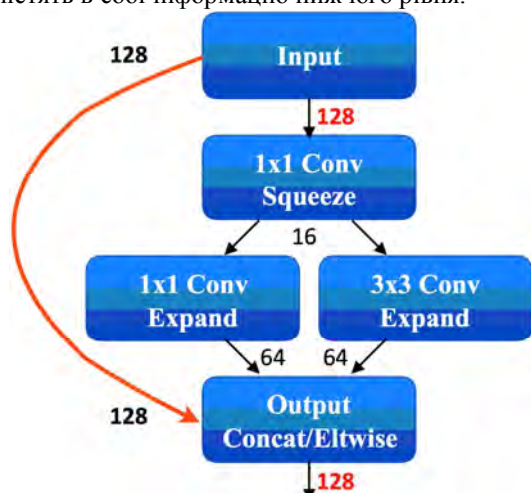


Рис. 5. Residual блок (Iandola et al., 2016)

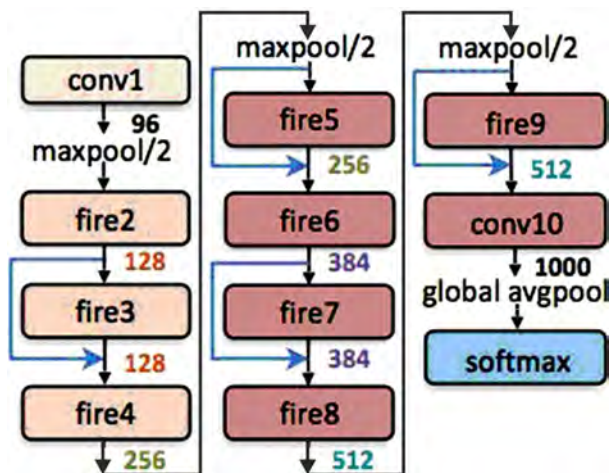


Рис. 6. Архітектура residual SqueezeNet мережі (Qi & Ebert, 2009)

Через обмежений набір даних у нашому датасеті (тільки 5000 зображень, на яких міститься полум'я) пропонуємо використовувати: поворот зображень, передискретизацію та кадрування.

На початковому етапі було вирішено навчати мережу з випадковими поворотами  $0^\circ$ ,  $75^\circ$ ,  $285^\circ$ , проте це не дало будь-яких покращень. Лише за умов використання усередненого значення, отриманого від класифікатора під час повороту зображень в зазначених напрямках (значення було підібрано експериментальним шляхом), було отримано змогу покращити ефективність роботи детектора (табл.).

Щодо передискретизації та кадрування зображень, то процес їх використання нагадує операцію з поворотами зображення. Так, вхідне зображення поділяється на 5 частин (чотири кутових та одне центральне), до яких і застосовується операція передискретизації. Отримуванні результати з кожної частини зображення знову ж усереднюються, внаслідок чого отримується кінцеве передбачення.

Потрібно також зазначити про те, що усі моделі на певному етапі навчання зіткнулися з проблемою пере-

навчання (overfitting), що було помічено у зв'язку з початком зростання втрат на валідаційному наборі даних (250 зображень). Для вирішення цієї проблеми, починаючи з цього етапу, показник швидкості навчання було зменшено в 10 разів (порівняно з початковим).

Прикінцевим етапом доналаштування детектора полум'я стала реалізація ансамблю з нейронних мереж. Для цього була навчена модель, яка навчалася лише на базі даних зображень із полум'ям (без застосування підходу трансферного навчання). Незважаючи на те, що така модель характеризувалася нижчою точністю, поєднавши її з іншими тонко налаштованими моделями, вдалося покращити загальний показник ефективності детектора полум'я. На нашу думку, такий факт зумовлений тим, що мережа натренована лише на дата сеті, який містить у собі зображення з вогнем, врахувала інші ознаки об'єкта, аніж тонко налаштована мережа.

Додатково до ансамблю ввійшли ще дві нейронні моделі, які містили уже в собі ознаки, отримувані внаслідок трансферного навчання. Перша модель донавчалася внаслідок штучного збільшення наявного датасету шляхом повороту зображень, друга – на датасеті збільшеного шляхом кадрування та передискретизації.

**Експериментальні дослідження.** Для експериментальних досліджень використано датасет, загальною місткістю 5000 зображень, який, своєю чергою, було поділено так: 65 % – для навчання, 15 % – для валідації та 20 % – для тестування. На цьому наборі зображення, які містять полум'я, були вручну проанотовані, зображення ж без полум'я були виключені з набору. Зокрема, для оцінки можливості роботи запропонованого детектора в режимі наближеного до реального часу, було використано набір з 10 відеороликами, в яких спостерігаються процеси горіння.

Загальний показник ефективності роботи нейронної мережі SqueezeNet без тонких налаштувань становив 89 %. За допомогою тонких налаштувань вдалося підняти показник ефективності додатково на 2,4 % (див. табл.). Приклад роботи класифікатора в складних умовах відеозйомки наведено на рис. 7.

Табл. Ефективність окремих етапів тонкого налаштування нейронної мережі

Етап тонкого налаштування мережі	Частка покращення ефективності роботи детектора, %
Поворот усіх зображень	0,7
Кадрування та передискретизація	0,4
Зменшення показника швидкості навчання	0,5
Ансамбль нейронних мереж	0,8



Рис. 7. Приклади правильних спрацювань класифікатора

Треба зауважити, що все ще спостерігаються не спрацювання класифікатора, особливо впродовж перших секунд від моменту початку пожежі (рис. 8, а), що обумовлено особливостями використання детектора руху. Зокрема, за умов різкої зміни заднього фону (див.

рис. 8, б) спостерігаються деякі помилкові спрацювання, що знову ж, на нашу думку, вимагає доопрацювання методу визначення рухомих об'єктів.

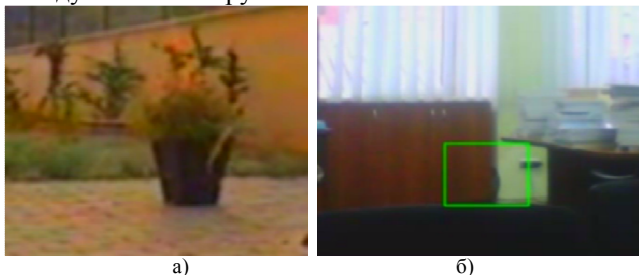


Рис. 8. Приклади неправильної роботи класифікатора

Окремо потрібно відзначити про підходи, використання яких не покращило ефективність роботи нейронної мережі: штучне збільшення датасету (масштабування, перекося, випадкові зрушення), поділ датасету на день/ніч (результати вимагають доопрацювання, оскільки зображення, на яких міститься полум'я, в нічний час становить лише 17 % від загальної кількості зображень), використання іншої архітектури SqueezeNet (dense→sparse→dense (Song Han, 2016)), додавання додаткових моделей до ансамблю.

**Висновки.** Отже, описано метод детектування полум'я у відеопотоці, який дає змогу поєднати переваги детектора Хаара (швидкість роботи) та глибоких згорткових нейронних мереж (точність детектування). Загалом вдалося створити три моделі нейронних мереж:

1. Модель, до якої було застосовано трансферне навчання, кадрування та передискретизація;
2. Доповнена перша модель внаслідок донавчання шляхом повороту зображень;
3. Модель без тонких налаштувань та без трансферного навчання.

Незважаючи на досить високі показники ефективності класифікації наявності полум'я на зображеннях (91,4 % правильності), вважаємо доцільним подальше дослідження можливостей тонкого налаштування мережі (особливо в напрямку навчання на зображеннях, які важко класифікуються), доопрацювання алгоритму визначення рухомих об'єктів та додаткового збільшення кількості шарів і моделей в ансамблі.

### Перелік використаних джерел

- Celik, T., & Demirel, H. (2009). Fire detection in video sequences using a generic color model. *Fire Safety Journal*, 2, 147–158. <https://doi.org/10.1016/j.firesaf.2008.05.005>
- Celik, T., et al. (2007). Fire pixel classification using fuzzy logic and statistical color model. *Acoustics, Speech and Signal Processing, ICASSP 2007. IEEE International Conference on. IEEE, 2007*, 1205–1208. <https://doi.org/10.1109/ICASSP.2007.366130>
- Chen, T., et al. (2004). An early fire-detection method based on image processing. *ICIP '04*, 1707–1710. <https://doi.org/10.1109/ICIP.2004.1421401>
- Chen, T. H., et al. (2003). An intelligent real-time fire-detection method based on video processing. *Security Technology, Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on. – IEEE*, 104–111. <https://doi.org/10.1109/CCST.2003.1297544>
- Frizzi, S., et al. (2016). Convolutional neural network for video fire and smoke detection. *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, 877–882. <https://doi.org/10.1109/IECON.2016.7793196>
- He, K., et al. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, 17(1–3), 185–203. [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2)
- Iandola, F. N., et al. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*, 1–13.
- KaewTraKulPong, P., & Bowden, R. (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. *Video-based surveillance systems. Springer US*, 135–144. [https://doi.org/10.1007/978-1-4615-0913-4\\_11](https://doi.org/10.1007/978-1-4615-0913-4_11)
- Lipton, A., et al. (1998). Moving target classification and tracking from real-time video. *Proc. of WACV'98*, 8–14.
- Maksymiv, O. P., Rak, T. E., & Peleshko, D. D. (2016). The Analytical Review of Fire Detecting Method by Using Computer Vision. *Scientific Bulletin of UNFU*, 26(5), 318–325. <https://doi.org/10.15421/40260550>
- Maksymiv, O., et al. (2016). Deep convolutional network for detecting probable emergency situations. *Data Stream Mining & Processing (DSMP), IEEE First International Conference on. IEEE, 2016*, 199–202. <https://doi.org/10.1109/DSMP.2016.7583540>
- Piccardi, M. (2004). Background subtraction techniques: a review. *Systems, man and cybernetics, 2004 IEEE international conference on. IEEE, 2004*, 3099–3104. <https://doi.org/10.1109/ICSMC.2004.1400815>
- Qi, X., & Ebert, J. (2009). A Computer Vision-Based Method for Fire Detection in Color Videos. *International journal of imaging*, 9, 22–34.
- Qingjie, Z., et al. (2016). Deep Convolutional Neural Networks for Forest Fire Detection. *International Forum on Management, Education and Information Technology Application*, 568–575. <https://doi.org/10.2991/ifmeita-16.2016.105>
- Russakovsky, O., et al. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Song Han (2016) SqueezeNet-DSD-Training. Retrieved from: <https://github.com/songhan/SqueezeNet-DSD-Training>.
- Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on. IEEE, 1999*, 246–252. <https://doi.org/10.1109/CVPR.1999.784637>
- Szegedy, C., et al. (2015). Going Deeper with Convolutions, *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Yosinski, J., et al. (2014). How transferable are features in deep neural networks? *Advances in neural information processing systems*, 3320–3328.

А. П. Максимиш

Львівський державний університет безпеки життєдіяльності, г. Львів, Україна

## КАСКАДНЫЙ МЕТОД ДЕТЕКТИРОВАНИЯ ПЛАМЕНИ В ВИДЕОПОТОКЕ С ИСПОЛЬЗОВАНИЕМ ГЛУБОКИХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Предложен каскадный бинарный классификатор, основанный на использовании смеси Гауссова распределения, примитивов Хаара и глубоких сверточных нейронных сетей. Применение такого подхода дает возможность, с одной стороны, осуществлять мониторинг среды в режиме, приближенном к реальному времени, а с другой – обеспечить достаточно высокий показатель эффективности обнаружения пламени на видео (92,7 %). На первом этапе, с помощью использования примитивов Хаара, к предварительно выделенным подвижным регионам видеоизображения генерируются так называемые регионы

интереса. На втором этапе полученные регионы интереса передаются для классификации на вход глубокой сверточной нейронной сети, которая формирует вывод о наличии или отсутствии пламени на изображении. В рамках проведения исследования предложена модифицированная модель нейронной сети – SqueezeNet. Выявлено, что с помощью использования трансферного подхода при обучении нейронной сети, существует возможность минимизировать количество ложных вызовов, особенно на объектах, которые визуалью могут напоминать пламя и уменьшить временные затраты, необходимые для ее обучения. С целью повышения эффективности работы нейронной сети осуществлен ряд тонких настроек (ансамбль из нейронных сетей, поворот изображений, уменьшение показателя скорости обучения, кадрирование и передискретизация), что суммарно позволило улучшить ее эффективность на 2,4 %.

**Ключевые слова:** обнаружение движения; примитивы Хаара; трансферное обучение; интеллектуальные системы; компьютерное зрение.

**O. P. Maksymiv**

*Lviv State University of Life Safety, Lviv, Ukraine*

## **CASCADE-BASED METHOD FOR FLAME DETECTION IN VIDEO SEQUENCES USING DEEP CONVOLUTIONAL NEURAL NETWORKS**

The authors present a novel flame detection approach based on the application of Haar features and Gaussian distribution as region of interest generator and deep convolutional neural network as classifier. The actuality of the paper is determined by fact, that camera is more versatile means of obtaining data which, in terms of developing appropriate algorithms, on the one hand allow faster receiving of information (even compared to modern sensors) and on the other hand, due to the large array of received information, will process such data as number of people in the placement, terrorist activity threat, unconsciousness people etc. The paper proposes the use of a Gaussian mixture models for the detection of moving objects. In order to verify the received regions, we used modified Haar-like features. It is the first stage, where region of interests is generated. At the second stage, the obtained results from the previous stage are transmitted for further classification to a deep convolutional neural network. Furthermore, significant efforts were made to choose the architecture of the neural network, since a number of neural networks requires a large amount of computing resources and does not allow working in real time. In order to solve this problem the use of modified SqueezeNet network architecture was proposed. As a result, we have achieved similar levels of classification accuracy, using 50 times fewer coefficients. Our results show that application of such approach as ensemble of neural networks, rotation of images, decrease of the rate of learning, framing and re-sampling allow increasing efficiency rate of convolutional neural network. To confirm this statement we have developed dataset including 5000 images of different categories and image qualities. As a result, efficiency rate, under the conditions of the above-mentioned approaches used, has increased to 2.4 %. In terms of accuracy and number of parameters, this method has surpassed state of the art method for fire detection in video sequences. To conclude, we should note that simple combination rules used here provide potentials for further improvement, especially region of interest's generator.

**Keywords:** motion detection; Haar features; transfer learning; intellectual systems; computer vision.