

DOI: <https://doi.org/10.36910/6775-2524-0560-2021-44-21>

УДК 004.89

Проніна Ольга Ігорівна, к.т.н., доцент<https://orcid.org/0000-0001-7085-8027>**Яремко Олександр Миколайович**, магістр<https://orcid.org/0000-0002-0650-5048>

Державний вищий навчальний заклад «Приазовський державний технічний університет», м. Маріуполь, Україна

АНАЛІЗ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ ПОРУШЕННЯ ПРАВИЛ ОХОРОНИ ПРАЦІ НА РОБОЧОМУ МІСЦІ

Проніна О. І., Яремко О. М. Аналіз згорткових нейронних мереж для розпізнавання порушення правил охорони праці на робочому місці. Охорона праці являє собою ряд заходів і засобів, спрямованих на збереження здоров'я і працездатності людини. На різних виробництвах співробітники схильні до впливу на їх здоров'я різних негативних факторів, наприклад, носіння каски може врятувати співробітника від руйнування кісток черепа, струс головного мозку, всі ці травми відносяться до розряду важких з усіма наслідками, що випливають. Робота присвячена пошуку найкращої моделі згорткової нейронної мережі, яка буде мати найкращі показники у визначенні об'єктів, на яких модель навчилася і оптимальна у використанні програмного забезпечення. Для подальшого аналізу розмітка даних проводилася за допомогою програми labelImg, всі інші етапи навчання і тестування проводилися в середовищі Python. Набір даних включає в себе кілька розмічених класів, а саме: людина з каскою і без, людина в діелектричних рукавичках і без рукавичок, людина з сигаретою і без неї, людина в спеціального робочого взуття і без неї, людина в робочому одязі і в звичайній.

Ключові слова: машинне навчання, згорткова нейронна мережа, охорона праці, розмітка даних, labelImg, Python.

Проніна О. И., Яремко О. М. Анализ сверточных нейронных сетей для распознавания нарушения правил охраны труда на рабочем месте. Охрана труда представляет собой ряд мероприятий и средств, направленных на сохранение здоровья и работоспособности человека. На различных производствах сотрудники подвержены влиянию на их здоровье различных негативных факторов, например, ношение каски может спасти сотрудника от разрушения костей черепа, сотрясение головного мозга, все эти травмы относятся к разряду тяжелых со всеми вытекающими последствиями. Работа посвящена поиску наилучшей модели сверточной нейронной сети, которая будет иметь наилучшие показатели в определении объектов, на которых модель обучилась и оптимальна в использовании программном обеспечении. Для дальнейшего анализа разметка данных производилась с помощью программы labelImg, все остальные этапы обучения и тестирования проводились в среде Python. Набор данных включает в себя несколько размеченных классов, а именно: человек с каской и без, человек в диэлектрических перчатках и без перчаток, человек с сигаретой и без нее, человек в специальной рабочей обуви и без нее, человек в рабочей одежде и в обычной.

Ключевые слова: машинное обучение, сверточная нейронная сеть, охрана труда, разметка данных, labelImg, Python.

Pronina O. I., Yaremko O. M. Analysis of convolutional neural networks for the recognition of violation of labor safety rules in the workplace. Labor protection is a series of measures and tools aimed at preserving human health and performance. In various industries, employees are exposed to various negative factors on their health, for example, wearing a helmet can save an employee from the destruction of the skull bones, a concussion, all these injuries are classified as severe with all the ensuing consequences. The work is devoted to finding the best model of a convolutional neural network, which will have the best performance in determining the objects on which the model has been trained and is optimal for using the software. The data was marked using the labelImg program, all other stages of training and testing were carried out in the Python environment. The dataset includes several labeled classes, namely: a person with and without a helmet, a person with dielectric gloves and without gloves, a person with and without a cigarette, a person in special work shoes and without them, a person in work clothes and in normal clothes.

Keywords: machine learning, convolutional neural network, occupational Safety and Health, data markup, labelImg, Python.

Постановка наукової проблеми. Охорона праці на виробництві потрібна в першу чергу для збереження здоров'я і життя співробітників. За безпеку співробітника відповідає начальство підприємства. В обов'язки начальства входить створення комфортного і безпечного робочого місця, надання засобів індивідуального захисту. На кожному виробництві, охорона праці регламентує правила, за якими має виконуватися технологічний процес виробництва, для того щоб убезпечити співробітників, від негативного впливу на їхнє здоров'я. Адміністрація підприємства повинна стежити за співробітниками, щоб вони виконували правила охорони праці. Дотримання правил охорони праці знижує травмонебезпечні ситуації на підприємстві і підвищує кваліфікацію співробітників, так само варто відзначити що охорона праці може містити в собі заходи спрямовані на регулювання виділення шкідливих речовин в навколишні середу. Охорона праці є засобом зниження ризику виникнення небезпечних ситуацій на підприємствах, але стежить за виконанням правил людина і через неухважність людини, він може пропустити або ще гірше дозволити порушити правила і через це складно звести ризик виникнення травм співробітників до нуля.

Мета роботи є навчання моделі, яка може виявляти на вхідних зображеннях порушення правил охорони праці. Для навчання моделі заздалегідь були підготовлені і розмічені близько двох тисяч картинок для кожного класу. Розмітка даних проводилася за допомогою програми `labelImg`, яка дозволяє швидко і зручно розмічати дані. Навчання моделі на підготовленій датасеті класів проводилася в середовищі розробки Python з використанням таких бібліотек як `torch`, `cv` для предобробки і проведенні тестування моделі.

Аналіз досліджень. Завдання виявлення об'єктів в сучасному світі дуже добре розвинена. Існує багато різних підходів і способів реалізації цього завдання, можна використовувати готову модель або реалізувати свою модель. Підходи так само різняться, можна виділяти об'єкти на зображенні, це називається `object detection`, але так само можна виконати сегментацію, тобто віднести кожен піксель до певного класу, це називається `object segmentation`.

У статті [1] Марко Ді Бенедетто і його колеги вирішили навчити модель на основі даних, які отримали з віртуального світу, а саме з реалістичною комп'ютерної гри. Творці цієї гри доклали багато зусиль, щоб симулювати реальний світ, тому автор статті вирішив спробувати використовувати дані з віртуального світу гри, щоб навчати моделі. Автори статті вирішили показати що навчання зі створеним навчальним набором з віртуального світу і використання фази адаптації предметної області є ефективним рішенням для випадків, в яких навчальний набір недоступний.

Кайо Сото Майор в своїй статті [2] «Виявлення засобів індивідуального захисту на промислових об'єктах за допомогою потокового відео з камери» займався дослідженням проблеми виявлення засобів індивідуального захисту за допомогою старої моделі YOLO2. Йому вдалося навчити моделі, які успішно виявляли засоби індивідуального захисту на людях.

Цзіцзянь Ван в своїй статті [3] «Швидке виявлення засобів індивідуального захисту на реальних будівельних майданчиках з використанням підходів глибокого навчання» вирішив навчити моделі для виявлення засобів індивідуального захисту використовуючи моделі YOLO v5x та YOLO v5s. Ключова особливість полягає в тому що він використовував датасет з якісними картинками, що збільшило якість моделі. Результат порівняння моделей показує, що YOLO v5x має кращу MAP (86,55%), а YOLO v5s має найвищу швидкість (52 FPS) на GPU.

У статті [4] «Виявлення та класифікація об'єктів з використанням алгоритмів машинного навчання» Дора Речід займалася виявленням міських об'єктів. Ці об'єкти можуть бути присутні кілька разів на зображенні або бути прикріплені один до одного. Дора вирішила використовувати мережі SVM. SVM дозволяє підвищити продуктивність, але зі значними обчислювальними витратами. Далі вона запропонувала використовувати шлях активації, що дозволяє знизити складність без втрати ефективності. Цей шлях буде послідовно активувати мережу і зупиняти дослідження, коли ймовірність виявлення об'єкта висока. Згодом Дора показала, що ми можемо комбінувати мережу SVM з картами функцій з згортованих нейронних мереж.

У статті [5] «Виявлення значущості та класифікація об'єктів» Крістофер Кулі вирішив удосконалити виявлення об'єктів за допомогою поліпшення визначення тільки значущих об'єктів на сцені. Він генерував карти значущості, створені двома різними опублікованими в даний час методами виявлення значущості, і навчав окремі лінійні SVM з використанням векторів ознак, отриманих за допомогою цих методів. Такий підхід збільшив класифікацію об'єктів на 5%.

Кеваль Мораба в своїй статті [6] вирішив вивчити питання спільного виявлення таких об'єктів, як собака, і її семантичних частин, таких як особа, лапа і інших. Його модель заснована на двох моделях Faster-RCNN які мають спільні характеристики для виконання нової функції, заснованої на увазі. Його дослідження з набором даних PASCAL-Part 2010 показують, що спільне виявлення може одночасно поліпшити виявлення об'єктів.

Сайед Афак Алі Шах в статті [7] «Автоматичне виявлення об'єкта за допомогою вимірювання об'єктивності» пропонує підхід до автоматичного виявлення об'єктів кількісно визначає ймовірність того, що вікно зображення охоплює об'єкти на зображенні. Він може розрізнити кілька об'єктів в сцені, при цьому окремі вікна фіксують кожен виявлений об'єкт. Найголовніше, запропонований підхід не вимагає ручного введення. Його підхід займає близько 4 хвилин на одну картинку.

У своїй роботі [8] Ю-Вей Чао вивчив проблему виявлення взаємодій людина-об'єкт в статичних зображеннях, які визначаються як пророцтво людини і прямокутника об'єкта з міткою класу взаємодії, яка їх пов'язує. Він запропонував згортові нейронні мережі на основі області чоловік-об'єкт (HO-RCNN), нову структуру на основі DNN для виявлення HOI. В основі HO-RCNN лежить шаблон взаємодії, новий вхід DNN, який характеризує просторові відносини між двома обмеженими прямокутниками.

Аншул Пундір в своїй статті [9] «Підхід на основі регіональної екстракції щодо визначення використання сигарет за допомогою глибокого навчання» на основі датасета із зображенням сигарет навчив модель на основі архітектури YOLOv3. Так само він упростив модель YOLOv3, так що навчена модель стала працювати швидше і видала точність класифікації 96,74%.

Чиж Фан в статті [10] «Узагальнене виявлення маленьких об'єктів без забуття» вирішив провести дослідження в області виявлення маленьких об'єктів. Провівши дослідження і порівняв R-CNN і Retentive R-CNN і зробив висновки, що Retentive R-CNN значно перевершує сучасні методи по загальній продуктивності серед всіх налаштувань так як він може досягти конкурентоспроможних результатів в порівнянні з іншими архітектурами.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. YOLOv4 [11] є більш вдосконаленою архітектурою згортокової нейронної мережі, вона приблизно на 10-12% точніше YOLOv3. Можна виділити той факт, що YOLOv4 була розроблена для навчання на одному графічному процесорі. Основними нововведеннями в новій архітектурі бути такі пункти:

- міжетапні часткові з'єднання (CSP) в новому CSPDarknet53;
- використання Mish і Leaky ReLU як функції активації;
- прийняття мережі агрегації шляхів замість FPN, який використовувався в YOLOv3;
- використання Spatial Pyramid Pooling [12] в якості плагіна.

Структуру YOLOv4 можна розділити на три основні блоки: блок вилучення ознак, який служить для виявлення характерних особливостей об'єктів на вхідному зображенні; блок збору карт ознак з різних верств, який збирає і передає карти ознак з різних рівнів нейронної мережі на блок виявлення і класифікації, який, в свою чергу, формує вихідні карти ознак, для різних масштабів, що дозволяє прогнозувати координати обмежувальних рамок шуканих об'єктів і класифікувати вміст кожного осередку на вхідному зображенні. YOLOv5 був запропонований в червні 2020 року. Його швидкість виведення зображень складає всього 0,007 с, що означає, що він може обробляти 140 кадрів в секунду (FPS); цієї швидкості достатньо для виявлення відеозображення в реальному часі.

Термінал введення в першу чергу відповідає за поліпшення даних, адаптивне масштабування зображення і обчислення адаптивного опорного кадру. Модель використовує метод мозаїки для поліпшення даних. Його основна мета – отримати будь-які чотири зображення з набору даних, а потім використовувати випадкове масштабування, перевертання, перетворення гами, інші операції і, нарешті, склейку для створення нового зображення з використанням випадкового вирізання. Поліпшення даних збагачує фон зображення, розширює вихідний набір даних і ефективно покращує можливість узагальнення. Цей метод ідеальний для виявлення дрібних цілей.

Для аналізу нейронних мереж необхідно було спочатку розробити та підготувати дані на яких вони будуть навчатися для подальшого їх порівняння. Для цього було виявлено декілько основних етапів. На першому етапі потрібно підготувати дані, на яких можна навчити модель. Дані є собою набір картинок, на яких знаходиться об'єкт, на якому потрібно навчити модель. У нашому випадку можливо кілька сценаріїв, тобто потрібно визначати відсутність об'єкта на зображенні. Для того щоб підготувати датасет потрібно зібрати близько двох тисяч картинок, на яких міститься потрібний нам об'єкт і зробити розмітку на кожній картинці. Після того як ми зібрали датасет, нам потрібно зробити аугментацію. Під аугментацією розуміють процес збільшення датасета за рахунок зміна різних характеристик зображень з датасета. Для того щоб виконати аугментацію скористаємося програмою IrfanView. На рисунку 1 представлені всі можливі перетворення, які можна застосувати до зображення.

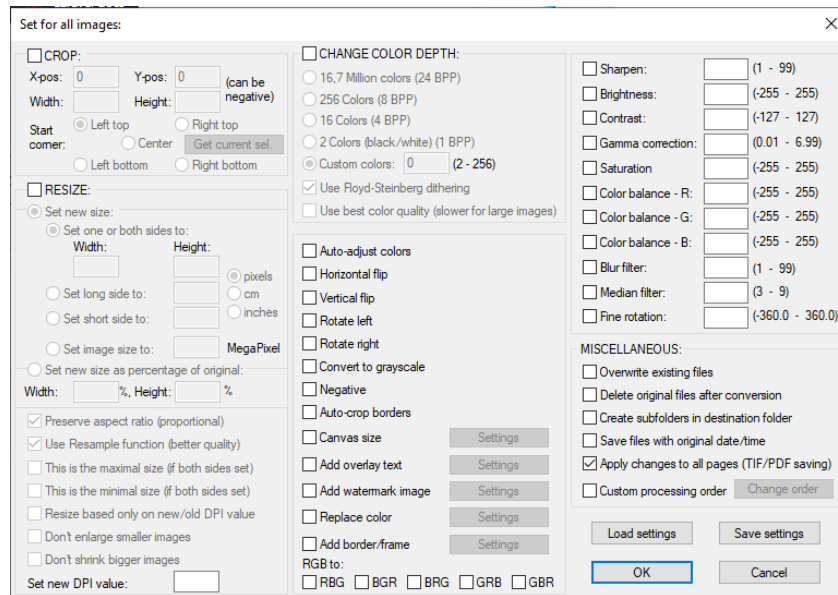


Рис.1. Види перетворень

Розмітка представляє виділення області де міститися об'єкт, що цікавить. Розмітка проводилася за допомогою програми labellmg рис.2.

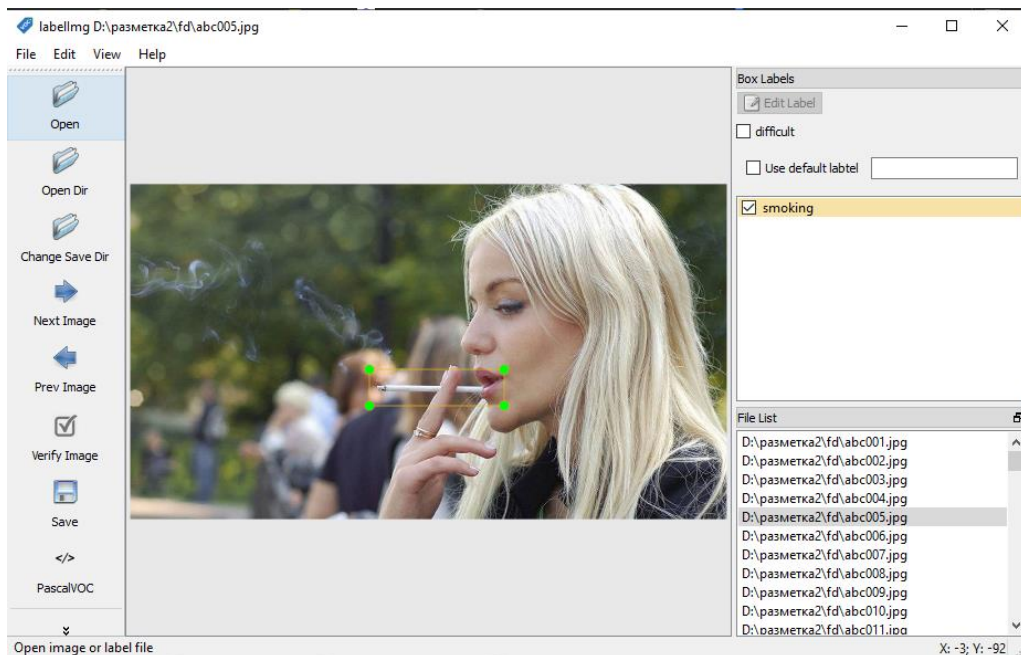


Рис.2. Програма для розмітки

Після розмітки, в залежності від налаштувань програми labellmg, на виході можливо отримати різні формати. В програми можна переключити формат на «yo1o» завдяки цьому не потрібно буде виконувати додаткових перетворень.

Наступним етапом є процес навчання і перевірки навчених моделей. Для того щоб навчити модель, потрібно налаштувати конфігураційні файли. Для початку потрібно створити файл objects.names, і в ньому вказати назву наших класів. Далі потрібно створити два файли: train.txt і valid.txt, в яких буде міститися шлях до картинок і файлів, в яких міститися розмітка об'єктів. Після вказівки імені класів нам потрібно створити файл conf.data в якому ми вказуємо кількість класів, шлях до даних для тренування, шлях до даних для валідації і шлях до файлу, де містяться назва класів рис.3.

```

conf.data x
conf.data
1 classes=12
2 train=C:/Users/Admin/Desktop/data/train.txt
3 valid=C:/Users/Admin/Desktop/data/valid.txt
4 names=C:/Users/Admin/Desktop/data/objects.names

```

Рис.3 - Конфігураційний файл

Останнім кроком у підготовці конфігураційних файлів буде настройка моделі. Для цього потрібно в файлі конфігурації `yolov4.cfg` змінити пару параметрів. Параметри `batch` і `subdivisions` залишаємо за замовчуванням, а `width` і `height` ставимо в значення 512. Значення `max_batches` потрібно встановити в 24000, тому що класів у нас 12 і кількість класів потрібно помножити на 2000. Для рядка `steps` потрібно вказати два значення, які вираховуються від значення `max_batches`, відповідно для першого 80% і для другого 90% і отримуємо 19200, 21600. Наступним кроком є вказівка кількості класів для кожного шару і вказівку `filters`. Щоб це зробити потрібно під написом `[yolo]` знайти рядок з `classes=80` і поміняти на `classes=12` і там же нижче де для `filters` вказано значення 255 поміняти його на результат цього перерахунку $(count_classes + 5) * 3$, в нашому випадку це значення 78. Даний крок потрібно виконати 3 рази, так як в моделі 3 шари. Вищеописані кроки налаштування моделі були для версії v4, для настройки моделі версії v5 потрібно вказати кількість класів на початку файлу в параметрі «`nc`». Залежно від виду моделі, параметри моделі різні, для проведення дослідів була обрана версія моделі v5s.

Одними з важливих метрик при дослідженні якості моделей були її швидкість і точність визначення об'єкта. Але так само варто не забувати про розмір датасета, який впливає на працездатність програмного забезпечення, яке буде використовувати натреновану модель рис.4.

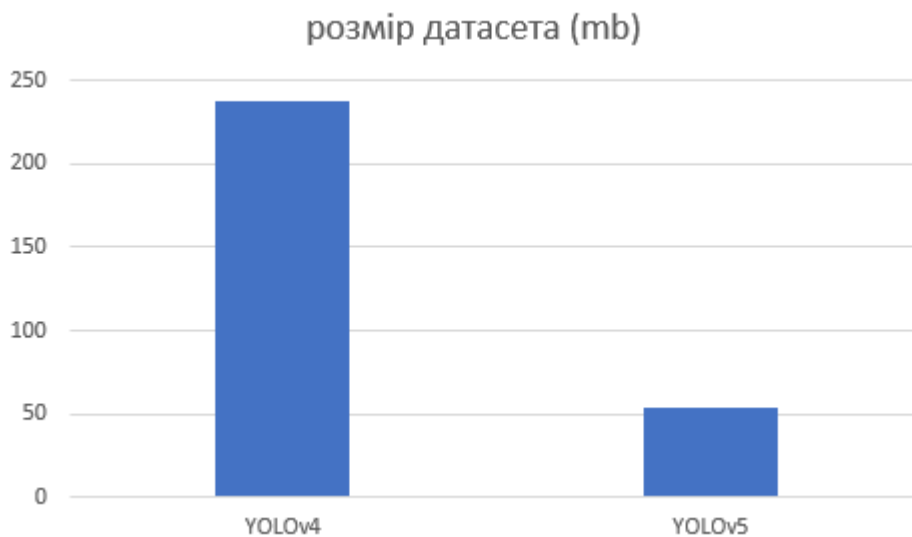


Рис.4 - Розміри датасета

YOLOv5 показала себе краще ніж YOLOv4 в плані розміру датасета. Для вимірювання якості моделей в мережі Інтернет було обрано 10 відео, в яких присутній один з класів, який використовувався для навчання моделей - сигарети. Кожне відео було пропущено через моделі і розраховане процентне співвідношення поява сигарети (Ap) в кадрі і виявлення сигарети моделлю. Результати в таблиці 1.

Таблиця 1. Результати тестування

| Номер відео | Якість відео, пікселі | YOLOv4, Ap% | YOLOv5, Ap% |
|-------------|-----------------------|-------------|-------------|
| 1 | 720 | 89 | 90 |
| 2 | 360 | 67 | 66 |
| 3 | 720 | 95 | 96 |
| 4 | 720 | 93 | 92 |
| 5 | 720 | 93 | 95 |
| 6 | 480 | 71 | 72 |
| 7 | 360 | 68 | 67 |
| 8 | 480 | 72 | 72 |
| 9 | 720 | 87 | 88 |
| 10 | 720 | 90 | 92 |

Виходячи з отриманих даних можна зробити висновок що якість моделей в плані виявлення об'єктів на відео потоці є однаковою, але різниця полягає в тому що розміри вихідних моделей різні. Тому оптимально буде використовувати в програмному забезпеченні версію моделі, яка навчилася на п'ятій версії yolo якщо потужності обладнання, на якому буде запускатися програмне забезпечення, будуть слабкими. Для більш потужних систем підійде модель, яка навчилася на четвертій версії. В подальшому планується більш поглиблено підлаштовувати модель і більш ретельно готувати дані для навчання.

Висновки та перспективи подальшого дослідження. В рамках цієї роботи був проведений аналіз порівняння архітектури двох згорткових нейронних мереж, а саме YOLOv4 і YOLOv5. Завдяки аналізу цих моделей було встановлено що у них не дуже велика різниця якості визначення моделей, але вдалося встановити що розміри обох моделей мають вагомий різницю. Цей факт може вплинути на працездатність програмного забезпечення, яке буде використовувати навчену модель. Оскільки навчену модель можна використовувати в програмному забезпеченні, яке буде обробляти відеопотоки на різних підприємствах і повідомляти системі про виниклі порушення різниця є істотною. Тому було обрано згорткову нейронну мережу YOLOv5 для більш точного визначення об'єктів.

Список бібліографічного опису

1. Di Benedetto, Marco & Carrara, Fabio & Meloni, Enrico & Amato, Giuseppe & Falchi, Fabrizio & Gennaro, Claudio. (2021). Learning accurate personal protective equipment detection from virtual worlds. *Multimedia Tools and Applications*. 80. 1-13. 10.1007/s11042-020-09597-9.
2. Souto Maior, Caio & Santana, João & Nascimento, Lucas & Macedo, July & Moura, Marcio & Isis, D.L. & Drogue, Enrique. (2018). Personal protective equipment detection in industrial facilities using camera video streaming. 10.1201/9781351174664-359.
3. Wang, Zijian & Wu, Yimin & Yang, Lichao & Thirunavukarasu, Arjun & Evison, Colin & Zhao, Yifan. (2021). Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors*. 21. 3478. 10.3390/s21103478.
4. Racheed, Dora & Muin, Rahmatullah & Jaylan, Ahmed. (2020). Object detection and object classification using machine learning Algorithms. *International Journal of Information Technology and Applied Sciences (IJITAS)*. 2. 21-29. 10.52502/ijitas.v2i3.12.
5. Cooley, Christopher & Coleman, Sonya & Gardiner, B. & Scotney, Bryan. (2017). Saliency Detection and Object Classification.
6. Morabia, Kaval & Arora, Jatin & Vijaykumar, Tara. (2020). Attention-based Joint Detection of Object and Semantic Part.
7. Shah, Syed. (2013). Automatic object detection using objectness measure.
8. Chao, Yu-Wei & Liu, Yunfan & Liu, Michael & Zeng, Huayi & Deng, Jia. (2017). Learning to Detect Human-Object Interactions.
9. Anshul Pundhir, Deepak Verma, Puneet Kumar, & Balasubramanian Raman. (2021). Region extraction based approach for cigarette usage classification using deep learning.
10. Zhibo Fan, Yuchen Ma, Zeming Li, & Jian Sun. (2021). Generalized Few-Shot Object Detection without Forgetting.
11. Alexey Bochkovskiy, Chien-Yao Wang, & Hong-Yuan Mark Liao. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection.
12. He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Lecture Notes in Computer Science*, 346–361.

References

1. Di Benedetto, Marco & Carrara, Fabio & Meloni, Enrico & Amato, Giuseppe & Falchi, Fabrizio & Gennaro, Claudio. (2021). Learning accurate personal protective equipment detection from virtual worlds. *Multimedia Tools and Applications*. 80. 1-13. 10.1007/s11042-020-09597-9.
2. Souto Maior, Caio & Santana, João & Nascimento, Lucas & Macedo, July & Moura, Marcio & Isis, D.L. & Drogue, Enrique. (2018). Personal protective equipment detection in industrial facilities using camera video streaming. 10.1201/9781351174664-359.
3. Wang, Zijian & Wu, Yimin & Yang, Lichao & Thirunavukarasu, Arjun & Evison, Colin & Zhao, Yifan. (2021). Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors*. 21. 3478. 10.3390/s21103478.
4. Racheed, Dora & Muin, Rahmatullah & Jaylan, Ahmed. (2020). Object detection and object classification using machine learning

- Algorithms. International Journal of Information Technology and Applied Sciences (IJITAS). 2. 21-29. 10.52502/ijitas.v2i3.12.
5. Cooley, Christopher & Coleman, Sonya & Gardiner, B. & Scotney, Bryan. (2017). Saliency Detection and Object Classification.
 6. Morabia, Keval & Arora, Jatin & Vijaykumar, Tara. (2020). Attention-based Joint Detection of Object and Semantic Part.
 7. Shah, Syed. (2013). Automatic object detection using objectness measure.
 8. Chao, Yu-Wei & Liu, Yunfan & Liu, Michael & Zeng, Huayi & Deng, Jia. (2017). Learning to Detect Human-Object Interactions.
 9. Anshul Pundhir, Deepak Verma, Puneet Kumar, & Balasubramanian Raman. (2021). Region extraction based approach for cigarette usage classification using deep learning.
 10. Zhibo Fan, Yuchen Ma, Zeming Li, & Jian Sun. (2021). Generalized Few-Shot Object Detection without Forgetting.
 11. Alexey Bochkovskiy, Chien-Yao Wang, & Hong-Yuan Mark Liao. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection.
 12. He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. Lecture Notes in Computer Science, 346–361.