

# Detekcja broni w systemach monitoringu wizyjnego z wykorzystaniem głębokich sieci neuronowych

podkom. Maciej Płoński  
Centralne Laboratorium Kryminalistyczne Policji  
ORCID: 0009-0005-8847-0372  
e-mail: [maciej.plonski@policja.gov.pl](mailto:maciej.plonski@policja.gov.pl)

## Streszczenie

Celem artykułu jest empiryczna ocena użyteczności wybranych architektur głębokich sieci neuronowych w zadaniu wykrywania broni na obrazach z monitoringu wizyjnego. Analizie poddano detektory jednoetapowe YOLOv5 oraz dwuetapowy Faster R-CNN, trenowane w identycznym reżimie na ograniczonym zbiorze danych. Eksperyment wykorzystuje zbiór z platformy Kaggle obejmujący 4000 obrazów przedstawiających broń palną i noże. Modele uczone z zastosowaniem transfer learningu na wagach wstępnie wytrenowanych na COCO, a skuteczność oceniano z użyciem mAP, precyzji, czułości (recall) oraz liczby klatek na sekundę (FPS). Uzyskane wyniki wskazują, że rodzina YOLO zapewnia korzystny kompromis między dokładnością a szybkością, umożliwiając przetwarzanie strumienia wideo w czasie zbliżonym do rzeczywistego.

**Słowa kluczowe:** sztuczna inteligencja, głębokie sieci neuronowe, detekcja broni, YOLO, Faster R-CNN, monitoring wizyjny

## Wstęp

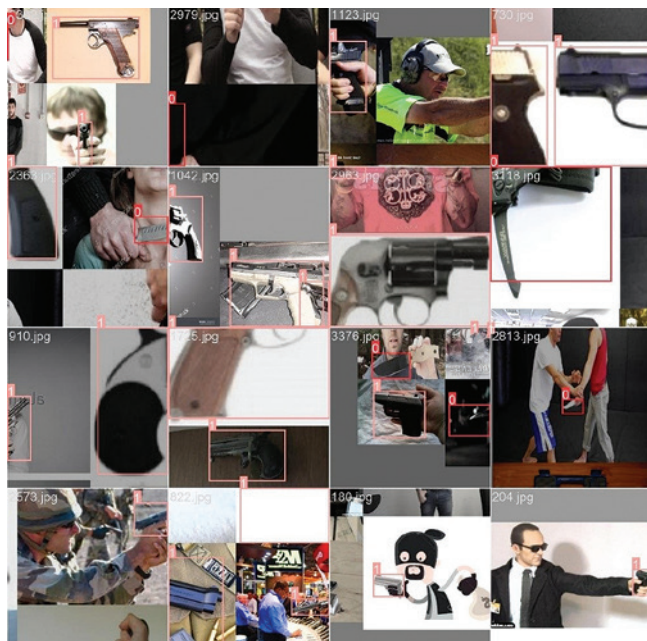
Rosnąca liczba kamer monitoringu wizyjnego w środowiskach miejskich i przemysłowych powoduje, że manualna analiza wszystkich strumieni wideo przez operatorów jest niewykonalna. Szacunki analityków rynkowych wskazują, że globalna liczba takich urządzeń w ostatnich latach przekroczyła miliard, przy czym znacząca część instalowana jest w przestrzeniach publicznych oraz w obiektach krytycznych (Torregrosa-Domínguez i in., 2024). W tym kontekście zadanie wykrywania sytuacji niebezpiecznych, na przykład pojawienia się broni, może zostać zautomatyzowane z wykorzystaniem metod głębokiego uczenia. Rozwiązania pracujące w trybie na żywo mają szczególne znaczenie praktyczne, ponieważ umożliwiają inicjowanie reakcji, takich jak powiadomienie operatora, blokada przejścia lub wezwanie patrolu, zanim dojdzie do nie-

bezpiecznego zdarzenia. Tradycyjna obserwacja monitoringu obarczona jest spadkiem czułości. Jak odnotowują Velastin i in. (2006), nawet wytrenowani operatorzy przestają skutecznie dostrzegać obiekty po około 20 minutach nieprzerwanej obserwacji (Torregrosa-Domínguez i in., 2024).

W badaniach nad detekcją obiektów utrwał się podział na dwie rodziny rozwiązań: szybkie modele jednoetapowe, takie jak YOLO i SSD oraz dokładniejsze, lecz wolniejsze modele dwuetapowe, w tym Faster R-CNN i Mask R-CNN (Reswara i in., 2023; Yusro i in., 2023). Modele jednoetapowe osiągają wysoką liczbę klatek na sekundę, co sprzyja zastosowaniom alarmowym. Modele dwuetapowe mogą oferować wyższą precyzję klasyfikacji kosztem szybkości. W literaturze polskojęzycznej brakuje jednak usystematyzowanego opisu skuteczności tych podejść w detekcji broni w warunkach monitoringu

CCTV oraz oceny korzyści z transfer learningu w przypadku małych, wyspecjalizowanych zbiorów danych. Wcześniejsze porównanie przeprowadzono w pracy magisterskiej M. Płońskiego (2025), w której zestawiono YOLOv5 i Faster R-CNN na wspólnym zbiorze obrazów.

Celem badania jest określenie, które z dwóch powszechnie stosowanych podejść: tj. model jednoetapowy (YOLO) lub dwuetapowy (Faster R-CNN) lepiej spełnia wymagania detekcji broni w monitoringu przy ograniczonych danych treningowych oraz oszacowanie wpływu transfer learningu na skuteczność.



Fot. 1. Przykłady lokalizacji i klasyfikacji obiektów

## Przegląd literatury

Detekcja broni stanowi szczególny przypadek detekcji małych obiektów. Pistolety, rewolwery oraz noże zajmują zwykle niewielką część kadru, często są częściowo przesłonięte i występują w zróżnicowanych warunkach oświetleniowych. Badania wskazują, że niska rozdzielczość miejskich kamer CCTV instalowanych w dużych odległościach ogranicza praktyczną skuteczność detekcji i wymaga stosowania augmentacji danych oraz modeli odpornych na zmianę skali (Torregrosa-Domínguez i in., 2024). Salazar i in. (2020) odnotowali, że pistolet na obrazie z kamery może zajmować około 3% powierzchni kadru. Nowsze implementacje oparte na YOLOv8 projektuje się pod kątem pracy w trybie na żywo w nadzorze wizyjnym, co potwierdza potrzebę skutecznej detekcji małych obiektów w czasie rzeczywistym.

Modele jednoetapowe i dwuetapowe w detekcji obiektów różnią się strategią przetwarzania. Jednoetapowe detektory, takie jak: YOLOv5 i YOLOv8, formułują zadanie jako wspólną regresję i klasyfikację w jednej sieci, co umożliwia uzyskanie wysokiej liczby klatek na sekundę przy akceptowalnej dokładności (Reswara i in., 2023; Yusro i in., 2023). Detektory dwuetapowe (np. Faster R-CNN) najpierw generują propozycje regionów zainteresowania, a następnie klasyfikują je

i dopasowują obramowania, co zwykle podnosi precyzję rozpoznania, lecz zwiększa koszt obliczeń. W literaturze porównawczej powraca wniosek, że YOLO lepiej odpowiada zastosowaniom wymagającym szybkości, natomiast w scenariuszach, w których kluczowa jest minimalizacja fałszywych alarmów, przewagę może mieć Faster R-CNN (Reswara i in., 2023; Aboymi, 2023; Yusro i in., 2023).

Od 2012 roku zbiór ImageNet, wykorzystywany w ramach wyzwania ILSVRC w latach 2012–2017, odegrał kluczową rolę w rozwoju metod detekcji obiektów. Obejmuje ponad 14 milionów obrazów przypisanych do tysiąca klas i stanowi podstawę trenowania nowoczesnych sieci konwolucyjnych (Deng i in., 2009). Od 2023 roku na platformie Kaggle dostępnych jest kilka zbiorów przeznaczonych do uczenia detektorów broni, między innymi Weapon Detection Dataset for YOLOv5 zawierający 4000 obrazów z pistoletami i nożami oraz nowsze zestawy wieloklasowe. Zbiory te są stosunkowo małe, co uzasadnia stosowanie transfer learningu. Zastosowanie wag wstępnie wytrenowanych na zbiorze COCO zwiększyło skuteczność modelu YOLOv5 o około 9 punktów procentowych mAP@0.5 względem trenowania od podstaw. Zbliżone wyniki dla YOLOv8 raportuje Kunchala (2024).

## Metodyka

Podstawą eksperymentu był zbiór Weapon Detection Dataset for YOLOv5 obejmujący 4000 obrazów, w tym po 2000 przykładów broni palnej i noży. Zbiór jest publicznie dostępny na platformie Kaggle. Obrazy oznaczono ręcznie w formacie YOLO, tj. za pomocą współrzędnych prostokąta oraz etykiety klasy obiektu. Dane podzielono na część treningową 80%, walidacyjną 10% i testową 10%. Zapewniono rozłączność scen między podzbiórami, aby uniknąć przecieku informacji i zawyżenia wyników.

W celu poprawy uogólnienia modelu zastosowano standardowe transformacje danych: losowe skalowanie, zmianę jasności, obrót oraz odbicie lustrzane w poziomie. Dodatkowo w części prób wykorzystano technikę Mosaic, która zwiększyła skuteczność rozpoznawania małych obiektów przez YOLOv5. Zastosowane augmentacje sztucznie zwiększają różnorodność próbek treningowych, co ma szczególne znaczenie przy ograniczonym rozmiarze zbioru.

Do porównań wybrano dwa modele detekcji obiektów: YOLOv5 w wariacie *small*, reprezentujący lekkie, jednoetapowe detektory, oraz Faster R-CNN z backbone'em ResNet-50, będący przedstawicielem klasycznych, dwuetapowych podejść, powszechnie opisywanych w literaturze (Reswara i in., 2023; Vijayakumar i in., 2023). W obu przypadkach zastosowano transfer learning, wykorzystując wstępnie wytrenowane wagi na zbiorze COCO. Proces treningu prowadzono do osiągnięcia stabilizacji wartości mAP na zbiorze walidacyjnym, przy czym ustalono minimalną liczbę epok: 100 dla modelu YOLOv5 oraz 50 dla Faster R-CNN.

Skuteczność porównywanych detektorów oceniano z użyciem czterech metryk. Po pierwsze: mAP@0.5

jako podstawowego wskaźnika precyzji detekcji obiektów. Po drugie: mAP@[0.5:0.95], czyli średniej precyzji liczonej dla progów IoU od 0.5 do 0.95 z krokiem 0.05. Po trzecie: precyzji oraz czułości (recall), które opisują jakość klasyfikacji obiekt kontra nie obiekt. Po czwarte: liczby klatek przetwarzanych na sekundę FPS jako miary przydatności modelu w zastosowaniach czasu rzeczywistego.

Wyniki

Poniżej zestawiono kluczowe wyniki eksperymentu (Tab. 1). Zarówno YOLOv5, jak i Faster R-CNN uzyskały wartości mieszczące się w przedziałach raportowanych w najnowszych pracach porównawczych dotyczących tych architektur (Reswara i in., 2023; Yusro i in., 2023).

niu, gdzie opóźnienie jest dopuszczalne, priorytetem staje się maksymalna dokładność lokalizacji obiektu oraz minimalizacja fałszywych alarmów. W takich warunkach zasadne może być zastosowanie modeli dwuetapowych, które dostarczają bardziej dopracowane predykcje i generują mniej wykryć fałszywych (Ahmed i Das, 2025). Mimo niższej szybkości działania Faster R-CNN, w scenariuszach analizy dowodowej opóźnienie rzędu kilkunastu sekund pozostaje akceptowalne, o ile przekłada się na wyższą precyzję detekcji.

Jakość danych i ograniczenia

Warto podkreślić, że badanie oparto na względnie jednorodnych, poprawnie oznaczonych obrazach. W warunkach miejskiego monitoringu niska rozdzielczość, niekorzystne oświetlenie oraz częściowa zasł-

Tab. 1. Porównanie skuteczności i wydajności modeli YOLOv5 i Fater R-CNN

| Model        | mAP@0.5 (%) | mAP@[0.5:0.95] (%) | Precyzja (%) | FPS  |
|--------------|-------------|--------------------|--------------|------|
| YOLOv5       | 71          | 36–40              | 90           | 139  |
| Faster R-CNN | 68          | 38–42              | 93           | 9–10 |

Otrzymane wyniki są zgodne z obserwacjami raportowanymi w literaturze. YOLOv5 osiąga wyrażnie wyższą przepustowość, 139 klatek na sekundę, niż Faster R-CNN, około 9 do 10 klatek na sekundę, przy jednoczesnym utrzymaniu wysokiej dokładności lokalizacji obiektów. Rozwój architektury YOLO sprawia, że dla prostych klas, w tym broni, uzyskiwany mAP jest zbliżony do wyników modeli dwuetapowych. Faster R-CNN oferuje z kolei wyższą precyzję kosztem szybkości. Zbieżność tych relacji potwierdzają badania porównawcze, na przykład Yusro i in. (2023), którzy odnotowali wyższą skuteczność całkowitą YOLOv5, 89,12%, względem Faster R-CNN, 83,92%, w identycznym środowisku testowym.

Dyskusja

Otrzymane rezultaty jednoznacznie wskazują różne scenariusze zastosowań rozważanych architektur detekcji.

Monitorowanie w czasie rzeczywistym

W scenariuszach, w których priorytetem jest minimalny czas reakcji, na przykład w monitoringu CCTV przestrzeni publicznych, rekomendowane są detektory jednoetapowe np. YOLOv5. W nowszych wdrożeniach np. YOLOv8, zapewnia się bardzo wysoką przepustowość przy zachowaniu wystarczającej dokładności detekcji, co potwierdzają zarówno wyniki niniejszego badania, jak i doniesienia literaturowe (Yusro i in., 2023; Ahmed i Das, 2025). Ahmed i Das (2025) wskazują, że YOLOv5 stanowi silnego kandydata do zastosowań czasu rzeczywistego, oferując korzystny kompromis między prędkością a precyzją.

Analiza pokłatkowa/śledcza

W analizie materiału wideo prowadzonej po zdarze-

niętość broni mogą istotnie obniżać skuteczność detektorów. Torregrosa-Domínguez i in. (2024) wskazują, że ograniczenia wynikające z jakości obrazu wymagają dodatkowych działań, takich jak zwiększenie rozdzielczości wejściowej modelu, zastosowanie segmentacji lub fuzja wyników z wielu detektorów, aby utrzymać wysoką skuteczność w środowisku rzeczywistym.

Wdrożenie systemów automatycznej detekcji broni powinno następować w zgodzie z obowiązującymi przepisami prawa. W kontekście Polski regulacje, w tym ustawa o Policji oraz RODO, wymagają, aby system automatyczny generował wyłącznie wskazanie potencjalnego zagrożenia, natomiast decyzja o interwencji pozostawała po stronie operatora. Systemy te pełnią więc funkcję wspomagającą i zwiększają efektywność prewencji oraz analizy dowodowej, nie zastępują jednak udziału człowieka.

Wnioski

Przeprowadzone badanie potwierdza, że głębokie sieci neuronowe umożliwiają skuteczną detekcję broni na obrazach z monitoringu przy wykorzystaniu transfer learningu, osiągając dokładność przekraczającą 70% mAP@0.5 mimo ograniczonego rozmiaru zbioru treningowego. W zaprezentowanym eksperymencie YOLOv5 uzyskał 71% mAP@0.5, co potwierdza przydatność tej klasy modeli w detekcji małych obiektów.

Wyniki wskazują, że architektura YOLOv5 zapewnia znacznie wyższą przepustowość obliczeniową niż Faster R-CNN, odpowiednio około 139 klatek na sekundę w porównaniu z około 10 klatkami na sekundę, przy relatywnie niewielkiej stracie dokładności klasyfikacji. Taka charakterystyka czyni modele z rodziny YOLO bardziej odpowiednimi do zastosowań operacyjnych w czasie rzeczywistym w systemach bezpieczeństwa publicznego.

Równocześnie Faster R-CNN pozostaje wartościowym narzędziem w analizie materiału dowodowego, gdzie dopuszczalne opóźnienie kompensowane jest przez wyższą precyzję lokalizacji i mniejszą liczbę fałszywych alarmów. W scenariuszach, w których priorytetem jest ograniczenie fałszywych detekcji i maksymalna precyzja, detektor dwuetapowy zachowuje użyteczność i może stanowić uzupełnienie rozwiązań jednoetapowych (Ahmed i Das, 2025).

## Bibliografia

1. Aboyomi, D. D. (2023). YOLO vs. SSD vs. Faster R-CNN: A comparative study on object detection. *Information Technology and Engineering Journal*, nr 8(2), s. 45–54.
2. Ahmed, I., & Das, R. (2025). Comparative analysis of YOLO and Faster R-CNN models for detecting traffic objects. *International Journal of Advanced Computer Science and Applications*, nr 16(3). <https://doi.org/10.14569/IJACSA.2025.0160342>
3. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *W 2009 IEEE Conference on Computer Vision and Pattern Recognition*, s. 248–255. IEEE. <https://doi.org/10.1109/CVPR.2009.5206848>
4. Kaggle. (2023). *Weapon detection dataset for YOLOv5*. <https://www.kaggle.com/datasets/balraj98/weapon-detection-object-detection>
5. Kunchala, H. K. R. (2024). *Real-time gun detection in video streams using YOLOv8* (Praca magisterska, California State University). California State University Scholar Works.
6. Płoński, M. (2025). *Analiza porównawcza architektury YOLO i Faster R-CNN w detekcji broni z wykorzystaniem transfer learningu* (Praca magisterska). Akademia Policji w Szczytnie.
7. Reswara, E., Sari, R., & Widodo, A. (2023). Comparison of object detection algorithm using YOLO vs Faster R-CNN. *W Proceedings of the 2023 International Conference on Big Data Technologies (ICBDT)*. ACM. <https://doi.org/10.1145/3627377.3627443>
8. Salazar González, J. L., Zaccaro, C., Álvarez-García, J. A., Soria-Morillo, L. M., & Sancho Caparrini, F. (2020). Real-time gun detection in CCTV: An open problem. *Neural Networks*, nr 132, s. 297–308.
9. Torregrosa-Domínguez, Á., Álvarez-García, J. A., Salazar-González, J. L., & Soria-Morillo, L. M. (2024). Effective strategies for enhancing real-time weapons detection in industry. *Applied Sciences*, nr 14(18), 8198. <https://doi.org/10.3390/app14188198>
10. Velastin, S. A., Boghossian, B. A., & Vicencio-Silva, M. A. (2006). A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. *Transportation Research Part C: Emerging Technologies*, nr 14(2), s. 96–113.
11. Vijayakumar, K. P., Pradeep, K., Balasundaram, A., & Dhande, A. (2023). R-CNN and YOLOv4 based deep learning model for intelligent detection of weaponries in real-time video. *Mathematical Biosciences and Engineering*, nr 20(12), s. 21611–21625. <https://doi.org/10.3934/mbe.2023956>
12. Yusro, M. M., Ali, R., & Hitam, M. S. (2023). Comparison of Faster R-CNN and YOLOv5 for overlapping object detection. *Baghdad Science Journal*, nr 20(3), s. 893–903. <https://doi.org/10.21123/bsj.2022.7243>
13. Ustawa z dnia 6 kwietnia 1990 r. o Policji (Dz.U. 1990 nr 30, poz. 179).
14. Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych (Dz.U. 2018 poz. 1000).

## Akty prawne