Piotr Lichograj[1], Marta Chodyka[2], Michał Szałkowski[3], Vladimir Golovko[4]

Akademia Bialska im. Jana Pawła II, Biała Podlaska

https://doi.org/10.29316/9788368103304_11

ANALIZA METOD YOLO W DETEKCJI OBIEKTÓW Z OBRAZÓW REJESTROWANYCH PRZEZ UAV ORAZ ICH ZASTOSOWANIE W POMIARZE PRĘDKOŚCI PIESZYCH

ANALYSIS OF YOLO METHODS IN DETECTING OBJECTS FROM IMAGES RECORDED BY UAVS AND THEIR APPLICATION IN MEASURING PEDESTRIAN SPEED

Streszczenie

W ostatnich latach bezzałogowe statki powietrzne (UAV) stały się kluczowym narzędziem pozyskiwania danych wizyjnych do analiz ruchu w środowisku miejskim. Perspektywa lotnicza umożliwia szerokie zastosowania w monitoringu pieszych, jednak jednocześnie generuje liczne wyzwania związane z niewielkim rozmiarem obiektów, zmienną

[1] https://orcid.org/0000-0002-1351-4874, e-mail: p.lichograj@dyd.akademiabialska.pl
[2] https://orcid.org/0000-0002-8819-2451, e-mail: m.chodyka@dyd.akademiabialska.pl
[3] student, e-mail: szalkowski22913@stud.akademiabialska.pl
[4] https://orcid.org/0000-0003-2615-289X, e-mail: v.golovko@dyd.akademiabialska.pl

wysokością lotu oraz niestabilnością detekcji. W rozdziale przedstawiono kompleksowy przegląd architektur YOLO w kontekście ich wykorzystania w obrazach z UAV oraz omówiono metody estymacji prędkości obiektów na materiale wideo. W części badawczej zaprezentowano prototypowy system detekcji i estymacji prędkości pieszych z wykorzystaniem modelu YOLOv8, wytrenowanego na specjalnie przygotowanym zbiorze danych UAV. Oceniono skuteczność detekcji, stabilność śledzenia oraz wiarygodność pomiaru prędkości. Wyniki potwierdzają, że nowsze generacje modeli YOLO, po odpowiednim dostrojeniu do danych lotniczych, umożliwiają dokładną analizę ruchu pieszych, a temat ten pozostaje obszarem o dużym potencjale badawczym.

Słowa kluczowe: UAV, detekcja obiektów, YOLOv8, analiza ruchu pieszych, estymacja prędkości, głębokie uczenie, śledzenie obiektów

Abstract

In recent years, unmanned aerial vehicles (UAVs) have become essential tools for acquiring visual data to support motion analysis in urban environments. The aerial perspective enables effective pedestrian monitoring but poses significant challenges, including small object size, variable flight altitude, and detection instability. This chapter provides a comprehensive review of YOLO architectures and their applicability to UAV imagery, along with an overview of velocity estimation methods in video sequences. The experimental section introduces a prototype system for pedestrian detection and speed estimation using a YOLOv8 model trained on a custom UAV dataset. The system's detection accuracy, tracking stability, and speed estimation reliability were evaluated. The findings demonstrate that modern YOLO models, when properly fine-tuned on UAV-specific data, enable accurate pedestrian motion analysis, while the topic itself remains an underexplored research area with substantial potential for future studies.

Keywords: UAV, object detection, YOLOv8, pedestrian motion analysis, speed estimation, deep learning, object tracking

Wstęp

W ostatniej dekadzie bezzałogowe statki powietrzne (UAV, ang. Unmanned Aerial Vehicles) stały się jednym z kluczowych narzędzi pozyskiwania danych wizyjnych w zastosowaniach naukowych, komercyjnych oraz inżynieryjnych. W przeciwieństwie do klasycznych, stacjonarnych systemów monitoringu, UAV

oferują możliwość dynamicznego rejestrowania obrazów z różnych wysokości i perspektyw, przy jednoczesnym zachowaniu stosunkowo niskich kosztów eksploatacji. Dzięki swojej mobilności oraz elastyczności platformy te znajdują zastosowanie w licznych scenariuszach, takich jak monitoring ruchu drogowego, analiza zachowań pieszych, badania dynamiki tłumu w przestrzeni miejskiej, kontrola infrastruktury krytycznej czy nadzór bezpieczeństwa podczas wydarzeń masowych. W każdej z tych dziedzin kluczową rolę odgrywa szybkie i wiarygodne przetwarzanie wizji komputerowej, które umożliwia automatyzację procesów decyzyjnych i zwiększa efektywność analiz.

Rozwój metod głębokiego uczenia, a w szczególności konwolucyjnych sieci neuronowych (CNN), wywarł fundamentalny wpływ na postęp w obszarze detekcji obiektów. Nowoczesne architektury, takie jak R-CNN, SSD czy rodzina modeli YOLO (You Only Look Once), umożliwiły znaczące podniesienie jakości i szybkości detekcji, otwierając drogę do zastosowań wymagających działania w czasie rzeczywistym. Spośród nich to właśnie YOLO stało się jednym z najważniejszych rozwiązań, oferujących kompromis pomiędzy wysoką dokładnością a bardzo dużą szybkością inferencji. Cecha ta jest szczególnie istotna w systemach UAV, w których przetwarzanie obrazu odbywa się często w warunkach ograniczonej mocy obliczeniowej, a jednocześnie wymaga natychmiastowej reakcji – na przykład podczas monitoringu bezpieczeństwa, sterowania autonomicznego czy wspomagania decyzji operatora.

Pomimo licznych zalet, detekcja obiektów na nagraniach UAV pozostaje zadaniem trudnym. Wynika to m.in. z niewielkiej liczby pikseli reprezentujących obiekty, zwłaszcza pieszych i małych pojazdów, z częstych zmian wysokości i kąta rejestrowania obrazu oraz z występowania dynamicznych zakłóceń wynikających z warunków środowiskowych. W literaturze wielokrotnie zwracano uwagę, że modele trenowane na standardowych zbiorach danych – takich jak MS COCO – charakteryzujących się głównie perspektywą naziemną, nie osiągają satysfakcjonujących wyników w warunkach UAV bez dodatkowego dostrajania. W odpowiedzi na te ograniczenia powstają liczne rozwiązania dedykowane obrazom lotniczym, obejmujące zarówno modyfikacje architektury, jak i adaptację procesu trenowania, rozszerzanie zbiorów danych oraz integrację informacji kontekstowych.

Detekcja obiektów stanowi jednak dopiero pierwszy element bardziej złożonego procesu, jakim jest analiza ruchu. W sytuacjach takich jak monitorowanie pojazdów czy pieszych konieczne jest nie tylko zidentyfikowanie obiektów, ale również oszacowanie ich prędkości oraz kierunku ruchu. Tradycyjne podejścia wykorzystują algorytmy śledzenia – takie jak KLT, SORT czy Deep SORT – działające w kombinacji z detektorami obiektów. Obliczanie prędkości opiera się najczęściej na translacji obiektu między kolejnymi klatkami, przeliczonej na jednostki metryczne w oparciu o parametry kalibracyjne kamery lub znane odcinki referencyjne. W przypadku UAV dochodzą do tego dodatkowe trudności, takie jak niestabilność kadru, zmieniająca się wysokość lotu czy brak jednoznacznych punktów odniesienia na powierzchni terenu.

W literaturze istnieje wiele prac dotyczących pomiaru prędkości pojazdów, jednak badania dotyczące estymacji prędkości pieszych w rejestrach UAV pozostają stosunkowo nieliczne. Wynika to głównie z trudności percepcyjnych: piesi widziani z wysokości kilkudziesięciu metrów zajmują bardzo mało pikseli, poruszają się w sposób nieregularny, a ich detekcja wymaga modeli odpornych na zmiany warunków wizualnych oraz częste zasłonięcia obiektów. Tym samym stanowi to obszar o dużym potencjale badawczym, szczególnie w kontekście rozwoju inteligentnych systemów monitoringu miejskiego oraz zastosowania UAV w badaniach przestrzennych.

W niniejszym rozdziale przedstawiono szeroki przegląd rozwiązań z rodziny YOLO w kontekście detekcji obiektów z UAV oraz przeprowadzono analizę istniejących podejść do estymacji prędkości obiektów na podstawie materiału wideo. W dalszej części zaprezentowano również badania własne autora, obejmujące opracowanie systemu detekcji I estymacji prędkości pieszych z wykorzystaniem modelu YOLOv8 wytrenowanego na dedykowanym zbiorze danych UAV. Szczególny nacisk położono na stabilność detekcji, proces śledzenia obiektów oraz wpływ filtracji i wygładzania sygnału na uzyskiwane wyniki, a także na praktyczne problemy związane z kalibracją przestrzenną. Całość pracy uzupełnia analiza porównawcza YOLOv4, YOLOv5 oraz YOLOv8 w kontekście danych UAV oraz omówienie ograniczeń i możliwości dalszego rozwoju metod.

Przegląd literatury

Rozwój detektorów obiektów i znaczenie architektur YOLO

Postęp w dziedzinie detekcji obiektów w ostatnich latach jest nierozerwalnie związany z rozwojem architektur głębokiego uczenia, w szczególności konwolucyjnych sieci neuronowych (CNN). Wprowadzenie modeli takich jak R-CNN, Fast R-CNN, Faster R-CNN czy SSD znacząco podniosło poziom dokładności detekcji, jednak ich wydajność obliczeniowa często nie pozwalała na zastosowanie w warunkach wymagających pracy w czasie rzeczywistym. Przełom przyniosła rodzina modeli YOLO (You Only Look Once), która wprowadziła jednoprzejściową (single-shot) koncepcję detekcji, umożliwiając przetwarzanie obrazu w sposób bezpośredni i szybki, bez konieczności wcześniejszego generowania propozycji regionów.

Klasyczne modele YOLO opierają się na pojedynczej sieci konwolucyjnej, która przetwarza obraz wejściowy i jednocześnie przewiduje współrzędne ramek ograniczających oraz prawdopodobieństwa klas. Architektura ta charakteryzuje się wyjątkową szybkością, co czyni ją szczególnie przydatną w zadaniach czasu rzeczywistego i zastosowaniach mobilnych, w tym na platformach UAV. Wprowadzenie YOLO stanowiło zatem istotny krok w kierunku praktycznego wykorzystania detekcji w systemach o ograniczonej mocy obliczeniowej.

YOLOv4: optymalizacja dokładności i wydajności

YOLOv4 (Bochkovskiy i in., 2020) odegrał kluczową rolę w rozwoju rodziny algorytmów YOLO (*You Only Look Once* – „widzisz tylko raz") poprzez wprowadzenie szeregu istotnych innowacji metodologicznych, które zostały podzielone na dwa zbiory rozwiązań. Wśród najważniejszych elementów architektury i procesu uczenia znalazł się ekstraktor cech CSPDarknet53 (*Cross Stage Partial Darknet* – częściowe łączenie etapów sieci Darknet), który poprawia przepływ gradientów oraz redukuje złożoność obliczeniową modelu. Istotną rolę odegrały również zaawansowane techniki augmentacji danych (*data augmentation* – sztuczne zwiększanie zbioru danych), w szczególności metoda mozaikowa, polegająca na łączeniu kilku obrazów w jeden, co zwiększa różnorodność danych treningowych i poprawia generalizację modelu. Kolejną innowacją było

zastosowanie samoadwersarialnego uczenia (SAT – ang. *Self-Adversarial Training*), które polega na celowym modyfikowaniu obrazów wejściowych w trakcie uczenia w celu zwiększenia odporności sieci na zakłócenia. Ponadto w YOLOv4 wprowadzono nowe funkcje straty (ang. loss functions) oparte na uogólnionych miarach współczynnika części wspólnej do sumy pól (IoU - ang. *Intersection over Union*), które umożliwiają dokładniejsze i stabilniejsze dopasowanie ramek ograniczających podczas procesu detekcji obiektów.

YOLOv4 osiągnął bardzo wysokie wyniki na zbiorze MS COCO, a jednocześnie utrzymał znaczną szybkość inferencji na popularnych GPU, co znacznie przyczyniło się do jego rozpowszechnienia w środowisku akademickim i inżynieryjnym. Model ten stał się w krótkim czasie standardem referencyjnym w wielu pracach badawczych dotyczących UAV, w których konieczne jest zachowanie kompromisu pomiędzy dokładnością a wydajnością.

YOLOv5: praktyczne wdrożenia i elastyczność architektoniczna

YOLOv5, rozwijany przez Ultralytics, mimo braku początkowej publikacji naukowej, bardzo szybko zyskał popularność dzięki implementacji w PyTorch, łatwości użytkowania oraz bogatym narzędziom do trenowania, wizualizacji i ewaluacji modeli. Rodzina YOLOv5 obejmuje szereg wariantów – od bardzo lekkich modeli „nano" po duże wersje „x-large" – co umożliwia ich dobór do konkretnych ograniczeń sprzętowych.

W literaturze znajdują się liczne prace, w których YOLOv5 zastosowano m.in. w detekcji obiektów naziemnych, obrazowaniu medycznym, rolnictwie precyzyjnym, monitoringu infrastruktury czy analizie zachowań ludzi. Jednakże w zastosowaniach UAV model ten często wymaga dostrajania, ponieważ dane lotnicze istotnie różnią się od konwencjonalnych zbiorów treningowych.

YOLOv8: najnowsza generacja i detekcja anchor-free

YOLOv8 (Ultralytics, 2020) stanowi obecnie najbardziej zaawansowaną, stabilną i powszechnie wykorzystywaną generację detektorów z rodziny YOLO (*You Only Look Once* – „widzisz tylko raz"), znajdując szerokie zastosowanie zarówno w badaniach naukowych, jak i w rozwiązaniach przemysłowych. Jedną

z kluczowych zmian wprowadzonych w tej wersji jest przejście z podejścia opartego na kotwicach (ang. *anchor-based*), w którym detekcja obiektów bazuje na wcześniej zdefiniowanych ramkach odniesienia, na podejście bez użycia kotwic (ang. *anchor-free*), co upraszcza proces uczenia oraz zwiększa elastyczność modelu w wykrywaniu obiektów o zróżnicowanych rozmiarach. Istotną innowacją jest również zastosowanie rozdzielonych gałęzi detekcyjnych, w których proces regresji ramek ograniczających oraz klasyfikacji obiektów realizowany jest niezależnie, co przekłada się na lepszą stabilność uczenia i wyższą dokładność predykcji. Ponadto YOLOv8 charakteryzuje się zredukowaną złożonością obliczeniową, co umożliwia jego efektywne wykorzystanie w systemach czasu rzeczywistego oraz na urządzeniach o ograniczonych zasobach sprzętowych. Dodatkowym atutem tej generacji jest integracja wielu zadań wizji komputerowej w ramach jednej platformy, obejmującej detekcję obiektów, segmentację oraz klasyfikację, co znacząco zwiększa uniwersalność i praktyczną użyteczność algorytmu.

YOLOv8 wyznacza obecnie standard w czasie rzeczywistym, zarówno w badaniach akademickich, jak i w rozwiązaniach wdrażanych w sektorze przemysłowym. Ultralytics podkreśla, że model ten jest zoptymalizowany pod kątem efektywności, dzięki czemu może działać nie tylko na wydajnych GPU, ale również – w mniejszym zakresie – na CPU.

Specyfika detekcji w obrazach UAV

Dane pozyskane z platform UAV stawiają przed systemami detekcji liczne wyzwania, wynikające m.in. z:

- niewielkich rozmiarów obiektów (piechota, małe UAV, drobne pojazdy),
- zmiennej wysokości lotu i kąta nachylenia kamery,
- dynamicznego tła,
- zakłóceń związanych z warunkami środowiskowymi,
- częstych częściowych zasłonięć (ang. occlusions).

Literatura wskazuje, że bezpośrednie zastosowanie standardowych modeli detekcyjnych do obrazów UAV prowadzi do spadku skuteczności, dlatego konieczne jest dostrajanie modeli do specyfiki danych lotniczych oraz wprowadzanie modyfikacji architektonicznych (Zeng i in., 2025; Fan i in., 2025).

Modyfikacje YOLO dedykowane UAV

W literaturze zaprezentowano szereg ulepszeń architektury YOLO mających na celu poprawę jakości detekcji obiektów w danych lotniczych:

- C4D-YOLOv8 (Zeng i in., 2025) – ulepszony ekstraktor cech i integracja wielorozdzielczych reprezentacji, poprawiające detekcję małych obiektów w zbiorach VisDrone.
- LUD-YOLO (Fan i in., 2025) – model zoptymalizowany pod kątem ograniczonych zasobów obliczeniowych UAV.
- Badania Alhassana i Yılmaza (2025) wskazują, że YOLOv5 może osiągać lepsze wyniki niż YOLOv4 w scenariuszach UAV, choć zależy to od danych i konfiguracji hiperparametrów.
- Jeszcze szerszą analizę stanowią prace Kaur i in. (2025), które porównują siedem wersji YOLO w zadaniu air-to-air, pokazując, że skuteczność modeli jest silnie zależna od zbioru danych i scenariusza, a nowsza wersja nie daje automatycznie przewagi.

Estymacja prędkości obiektów

Estymacja prędkości pojazdów i pieszych z materiału wideo stanowi ważny element systemów analizy ruchu. Klasyczne podejścia opierają się na:

- detekcji obiektu,
- śledzeniu jego pozycji w kolejnych klatkach,
- konwersji przesunięć pikselowych na jednostki metryczne,
- filtracji i stabilizacji wyników.

Najczęściej wykorzystywane algorytmy śledzenia to KLT, SORT i Deep SORT. W systemach opartych na YOLO procedura jest analogiczna: środek ramki ograniczającej stanowi reprezentację pozycji obiektu, a przemieszczenia pomiędzy klatkami – podstawę estymacji prędkości (Ultralytics, 2024a).

Znaczną trudność stanowi stabilność detekcji – nawet pojedyncze błędne wykrycie może prowadzić do artefaktów w obliczanych prędkościach. Dlatego literatura zaleca stosowanie metod wygładzania czasowego (np. średniej ruchomej), filtrów progowych oraz dokładnej kalibracji przestrzennej.

W odniesieniu do badań pieszych są nieliczne, co wynika z trudności percepcyjnych związanych z niewielkimi rozmiarami obiektów w ujęciach UAV.

Metodologia badań

Celem badań własnych było opracowanie oraz przetestowanie systemu detekcji i estymacji prędkości pieszych na nagraniu wideo zarejestrowanym z UAV z wykorzystaniem modelu YOLOv8. Najważniejszym założeniem było stworzenie kompletnego rozwiązania, które umożliwia nie tylko wykrywanie obiektów na nagraniu, lecz także obliczanie ich prędkości w sposób stabilny, powtarzalny i zbieżny z faktycznym ruchem pieszych. Kluczowe znaczenie miało zapewnienie odpowiedniej jakości detekcji obiektów niewielkich rozmiarów, obserwowanych z wysokości ok. 30 metrów, co – jak wskazuje literatura – stanowi jedno z największych wyzwań w analizie danych UAV.

Projekt obejmował przygotowanie dedykowanego zbioru danych, trenowanie modelu YOLOv8, opracowanie modułu do przetwarzania materiału wideo oraz wykonanie procedury śledzenia i estymacji prędkości obiektów. Dodatkowo przeprowadzono analizę stabilności wyników oraz porównanie YOLOv8 z modelami YOLOv4 i YOLOv5.

W pierwszym etapie badań przygotowano zbiór danych obejmujący 200 obrazów pochodzących bezpośrednio z nagrań UAV, przedstawiających pieszych obserwowanych z perspektywy pionowej. Obrazy zostały ręcznie oznaczone w środowisku Roboflow. Zdefiniowano jedną klasę detekcji – person. Jednokomórkowa klasyfikacja była decyzją celową, gdyż upraszczała zadanie oraz zwiększała dokładność rozpoznawania obiektu kluczowego dla późniejszych obliczeń prędkości.

Zbiór danych został następnie podzielony na część treningową i walidacyjną zgodnie z rekomendacjami Ultralytics. Model YOLOv8 poddano procesowi trenowania z typowymi hiperparametrami, a uzyskany najlepszy model zapisano jako plik best.pt. W trakcie procesu trenowania zastosowano standardowe augmentacje, wspomagające generalizację modelu na różnorodne warunki oświetleniowe i geometryczne.

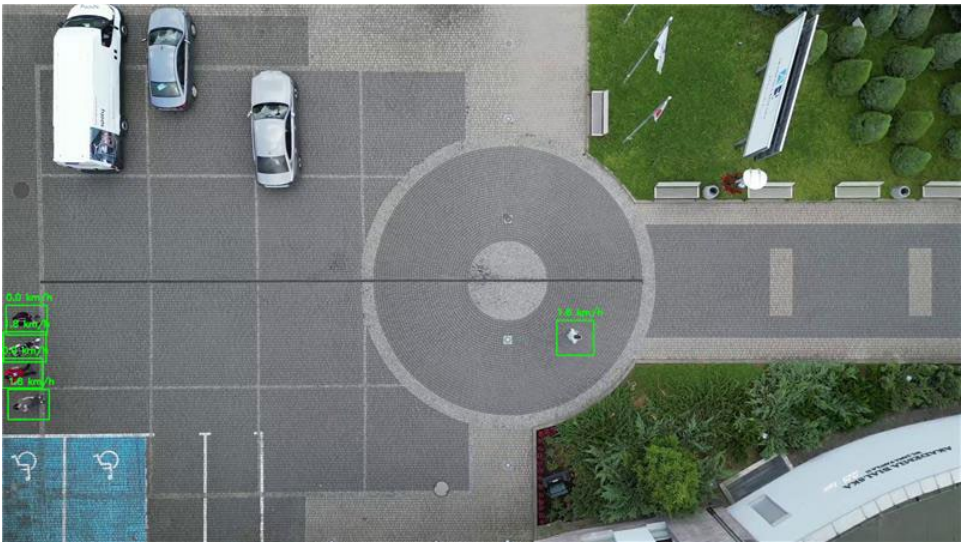
Nagranie zarejestrowane z wysokości 32 metrów było przetwarzane klatka po klatce przez model YOLOv8, który generował ramki ograniczające wykrytych osób, współrzędne ich środków oraz poziom ufności detekcji.

Wydajność modelu okazała się wystarczająca do działania w czasie rzeczywistym. W każdej klatce pozycja środka ramki była zapisywana jako punkt referencyjny służący do obliczania przemieszczeń. Wykorzystano jedynie jedną klasę (person), co pozwoliło uniknąć problemów z wieloklasową analizą sceny.

Detekcja obiektów w kolejnych klatkach wideo nie gwarantuje zachowania tej samej tożsamości obiektów, ponieważ YOLOv8 w każdej klatce wykonuje detekcję niezależnie. Aby zapewnić spójność ruchu poszczególnych osób, konieczne było opracowanie uproszczonego mechanizmu śledzenia obiektów.
Zastosowano procedurę:

1. Dla każdego wykrytego obiektu obliczano odległość pomiędzy jego aktualną pozycją (środek ramki) a pozycjami obiektów z poprzedniej klatki.
2. Jeśli odległość była mniejsza niż 50 pikseli, uznawano, że jest to ten sam obiekt i przypisywano mu ten sam identyfikator ID.
3. Jeśli przekroczono próg 50 px, tworzono nowy identyfikator, co odpowiadało pojawieniu się nowej osoby w scenie.

Metoda ta, choć prosta, okazała się wystarczająca. W analizowanym nagraniu liczba pieszych była niewielka, a trajektorie nie przecinały się w sposób istotnie złożony, dzięki czemu nie wystąpiły problemy charakterystyczne dla gęstych scen. Rezultat przedstawiono na Rysunku 1.



Rysunek 1. Klatka video z uruchomionym algorytmem YOLO-8 trenowanym do detekcji osób

Źródło: opracowanie własne.

Kalibracja przestrzenna i przeliczanie pikseli na metry

Aby oszacować prędkość ruchu pieszych, konieczne było określenie relacji pomiędzy odległością w pikselach a rzeczywistą odległością w metrach. W tym celu zastosowano współczynnik kalibracyjny:

$$pixel_to_meter_ratio = \frac{32 \text{ metry}}{375 \text{ pikseli}} \approx 0{,}085 \text{ m/px}.$$

Wartość ta została wyznaczona na podstawie znanego odcinka znajdującego się w analizowanej scenie, co stanowi standardową praktykę w analizie nagrań UAV, gdy dostęp do parametrów kamery jest ograniczony.
Przemieszczenie obiektu pomiędzy klatkami wyliczano jako:

$$\Delta s = pixel_distance \times pixel_to_meter_ratio.$$

Z kolei czas pomiędzy klatkami, przy znanej liczbie klatek na sekundę nagrania, pozwalał obliczyć prędkość:

$$v = \frac{\Delta s}{\Delta t}.$$

Prędkość wyrażano następnie w km/h.

Filtracja i stabilizacja wyników prędkości

W trakcie badań zaobserwowano, że nawet niewielkie niestabilności detekcji prowadzą do skoków w wartościach prędkości. Dotyczyło to:

- chwilowych przesunięć ramek,
- błędnych detekcji,
- zaniku obiektu na pojedynczą klatkę,
- lokalnych zasłonięć obiektów.

Aby temu przeciwdziałać, wprowadzono dwa mechanizmy stabilizujące. Zastosowano uśrednianie ostatnich pięciu wartości prędkości:

$$v_{smooth}(t) = \frac{1}{5}\sum_{i=0}^{4} v(t-i).$$

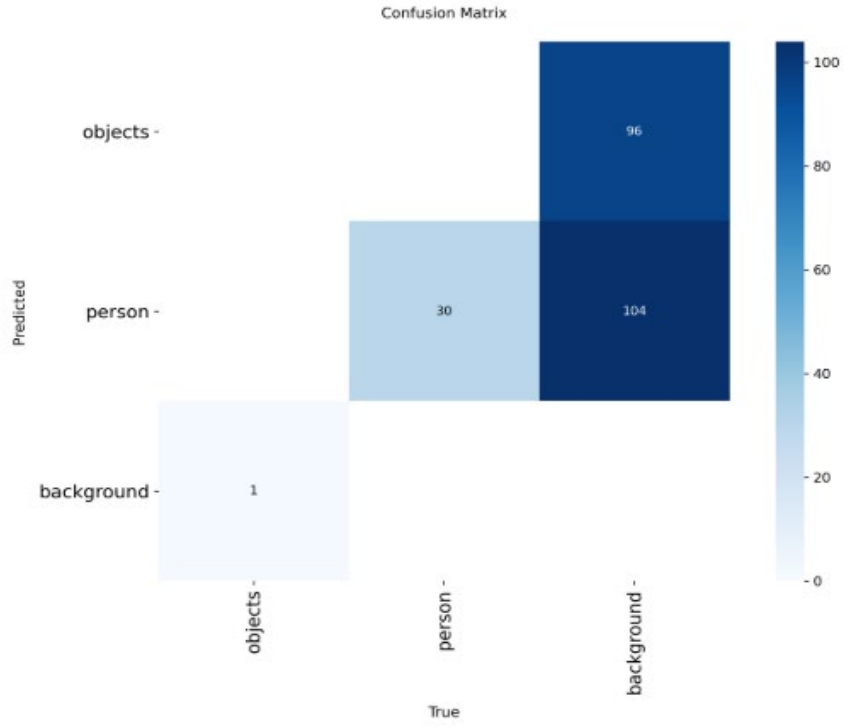
Mechanizm ten redukował fluktuacje wynikające z nieregularnych detekcji.

W przypadku gdy chwilowa prędkość wzrastała o ponad 30% w stosunku do wartości z poprzedniej iteracji, była ona zastępowana wartością wcześniejszą. Zastosowany filtr skutecznie ograniczał wpływ błędów detekcji, zapobiegał występowaniu nienaturalnych skoków prędkości, a jednocześnie umożliwiał swobodne spadki prędkości, na przykład w sytuacji zatrzymania się pieszego. Połączenie obu metod filtracji pozwoliło na uzyskanie stabilnych i realistycznych przebiegów prędkości w czasie. Końcowa wizualizacja wyników obejmowała wyświetlanie ramek ograniczających wokół wykrytych pieszych oraz wygładzone wartości ich prędkości wyrażone w kilometrach na godzinę, przy jednoczesnym pominięciu identyfikatorów obiektów w celu poprawy czytelności prezentowanych danych.

Celem była jak najprostsza prezentacja informacji ilościowej, tj. prędkości ruchu pieszych, bez przeładowania obrazu dodatkowymi elementami.

Ewaluacja działania modelu YOLOv8

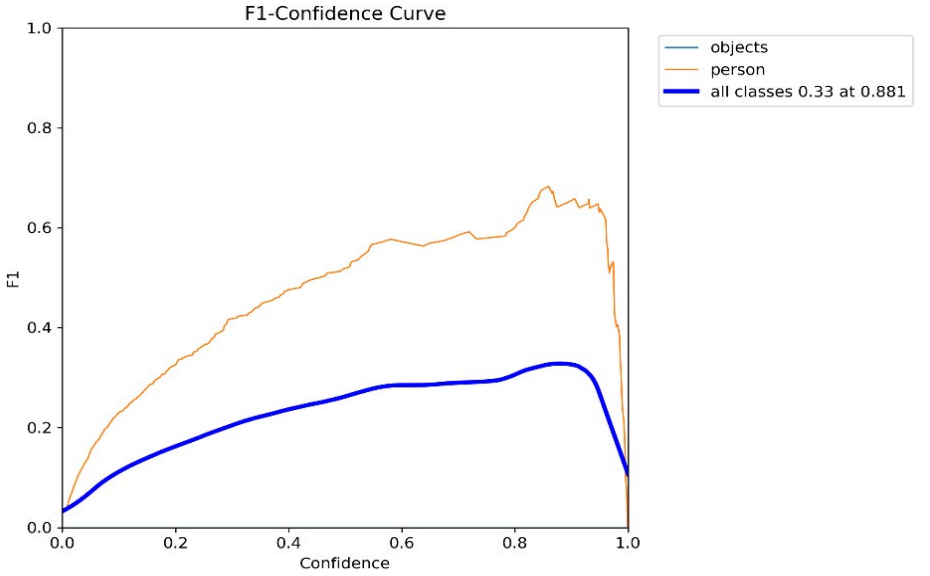
Po przejściu procesu treningu modelu YOLOv8 opracowano macierz pomyłek celem pokazania, jak często detekcje modelu pokrywają się z rzeczywistymi etykietami (Rysunek 2).



Rysunek 2. Macierz pomyłek modelu YOLOv8
Źródło: opracowanie własne.

Wyniki oceny modelu YOLOv8 wskazują, że przy progu ufności równym 0,88 osiągnął on wartość F1-score (Rysunek 3) na poziomie około 0,6. W trakcie analizy odnotowano 104 poprawne wykrycia osób oraz 96 przypadków poprawnego rozpoznania tła jako tła, czyli sytuacji, w których nie występowały obiekty. Jednocześnie zaobserwowano jedynie niewielką liczbę fałszywych wyników pozytywnych i negatywnych. Analiza macierzy pomyłek oraz krzywej F1 w funkcji progu ufności potwierdziła wysoką skuteczność modelu w warunkach

obrazowania z UAV, co podkreśla istotną rolę trenowania sieci na danych dostosowanych do specyfiki perspektywy lotniczej.



Rysunek 3. Zależność F1-score od progu ufności

Źródło: opracowanie własne.

W ramach badań wykonano porównanie detekcji trzech modeli YOLO: YOLOv4, YOLOv5 (wersje s oraz l) i YOLOv8. Analizę możliwości zastosowania w analizie sekwencji wideo z poziomu bezzałogowych statków powietrznych przedstawiono w Tabeli 1.

Wyniki

- YOLOv4 i YOLOv5 nie wykrywały pieszych z wysokości 32 metrów. Detektory te, trenowane na COCO, nie były przystosowane do tak małych obiektów.
- YOLOv8, trenowany na dedykowanym zbiorze UAV, zapewniał detekcje stabilne i użyteczne.

Tabela 1. Porównanie cech algorytmów YOLO v4, v5 i v8

Cecha	YOLOv4	YOLOv5	YOLOv8
Wersja	Darknet	PyTorch	Ultralytics (PyTorch)
Detekcja osób z góry	Bardzo niska	Niska	Wysoka (po dotrenowaniu)
Wymagana dodatkowa nauka	Tak	Tak	Tak, lecz nie w wysokim stopniu złożoności
Integracja z cv2	Wymaga konwersji	Prosta	Natychmiastowa
Czas detekcji (CPU)	31,1 ms	23,3 ms	16,1 ms
Obsługa klas własnych	Ręczna konfiguracja	Wbudowana	Automatyczna (Roboflow)
Stabilność predykcji	Niska przy małych osobach	Niestabilna	Dobra z wygładzaniem

Źródło: opracowanie własne.

Uzyskane wyniki są zgodne z obserwacjami prezentowanymi w literaturze przedmiotu, które wskazują na ograniczoną transferowalność modeli uczonych na danych pozyskiwanych z perspektywy naziemnej do zastosowań wykorzystujących obrazowanie z innych punktów widzenia. Jednocześnie potwierdzają one wyraźną przewagę nowszych wersji algorytmów z rodziny YOLO w zadaniach detekcji obejmujących małe obiekty, dla których kluczowe znaczenie mają zarówno architektura sieci, jak i zastosowane strategie uczenia.

Dyskusja

Wyniki przeprowadzonych badań własnych pozwalają na sformułowanie szeregu istotnych obserwacji dotyczących detekcji pieszych oraz estymacji ich prędkości na nagraniach z UAV przy wykorzystaniu modeli YOLO, w szczególności YOLOv8. Analiza eksperymentalna potwierdza zakładane wcześniej trudności związane z przetwarzaniem obrazów lotniczych, a jednocześnie pozwala na weryfikację skuteczności współczesnych rozwiązań detekcyjnych w tym specyficznym środowisku.

Jednym z kluczowych wniosków wynikających z badań jest potwierdzenie obserwacji powszechnie opisywanej w literaturze (m.in. Kaur i in., 2025; Zeng i in., 2025; Fan i in., 2025), że skuteczna detekcja obiektów w obrazach UAV wymaga dostrojenia modeli do specyfiki danych lotniczych oraz rozwiązań lepiej

radzących sobie z małymi obiektami. Dotyczy to zwłaszcza pieszych obserwowanych z wysokości ok. 30 metrów, których reprezentacja wizualna jest ograniczona do zaledwie kilkudziesięciu pikseli. Badania porównawcze przeprowadzone w niniejszej pracy wykazały, że zarówno YOLOv4, jak i YOLOv5 – pomimo swojej popularności w zastosowaniach naziemnych – nie były w stanie poprawnie wykrywać pieszych w analizowanej scenie.

Z kolei YOLOv8, po odpowiednim dostrojeniu do danych UAV, osiągnął znacząco lepsze wyniki, co potwierdza kluczową rolę dostrojenia algorytmu oraz doboru reprezentatywnego zbioru danych. Osiągnięty F1-score na poziomie około 0,6 przy wysokim progu ufności (0,88) należy uznać za zadowalający w analizowanych warunkach, zwłaszcza że dane pochodzą z perspektywy lotniczej, która charakteryzuje się wysokim stopniem trudności i stosunkowo małą liczbą pikseli reprezentujących sylwetkę pieszego.

Wyniki te są zgodne z podejściami zaprezentowanymi w literaturze – m.in. C4D-YOLOv8 (Zeng i in., 2025), LUD-YOLO (Fan i in., 2025) oraz Enhanced YOLOv8 (Luo, Yuan, 2025) – w których autorzy potwierdzają konieczność wzbogacania informacji kontekstowej i modyfikowania architektury detektora, tak aby lepiej radził sobie z małymi obiektami. W szczególności modyfikacje dotyczące przetwarzania cech wielorozdzielczych odgrywają kluczową rolę w poprawie dokładności detekcji.

Kolejnym istotnym aspektem analizowanych badań jest wpływ stabilności detekcji na estymację prędkości obiektów. Nawet pojedyncze błędne wykrycie, przesunięcie ramki lub chwilowy brak detekcji prowadzą do nienaturalnych skoków w wartościach prędkości, dlatego w literaturze poświęconej estymacji prędkości i ekstrakcji danych ruchowych z materiału UAV podkreśla się znaczenie stabilnej detekcji oraz śledzenia obiektów (Jiao, Fei, 2023; Manjusha i in., 2025). W analizowanym projekcie istotną poprawę stabilności uzyskano dzięki zastosowaniu dwóch metod filtracji: wygładzania średnią ruchomą oraz filtru zmian powyżej 30%. Mechanizmy te znacząco zredukowały liczbę anomalii i poprawiły spójność czasową wyników.

Podejście zastosowane w badaniach własnych jest zbieżne z zalecaniami Ultralytics (2024a), w których podkreśla się konieczność wykorzystania filtracji w zadaniach obejmujących przeliczanie trajektorii na jednostki metryczne. Literatura wskazuje, że przy braku takich mechanizmów nawet bardzo dobre detektory

mogą generować dane ruchowe niskiej jakości z punktu widzenia analizy dynamiki obiektów.

Jednocześnie należy zauważyć, że zastosowane rozwiązania stanowią kompromis pomiędzy złożonością obliczeniową a jakością wyników. Szczególnie interesujące wydaje się porównanie tych metod z bardziej zaawansowanymi algorytmami śledzenia – takimi jak Deep SORT czy ByteTrack – które mogą dodatkowo poprawić spójność trajektorii, choć kosztem większej złożoności implementacyjnej.

W badaniach zdecydowano się na zastosowanie uproszczonego algorytmu śledzenia opartego na minimalnej odległości pomiędzy środkami ramek detekcji. Jest to zgodne z praktyką stosowaną w wielu projektach badawczych, w których liczba obiektów jest niewielka, a trajektorie nie przecinają się w sposób złożony. W analizowanym scenariuszu mechanizm ten zapewnił pełną funkcjonalność systemu, pozwalając na jednoznaczne przypisywanie identyfikatorów ID.

W bardziej złożonych środowiskach – np. przy dużym zagęszczeniu pieszych, licznych zasłonięciach lub dynamicznych zmianach perspektywy UAV – konieczne może być zastosowanie bardziej złożonych metod śledzenia. Literatura sugeruje, że wykorzystanie algorytmów takich jak SORT, Deep SORT, a nawet rozwiązań opartych na transformatorach sekwencyjnych (np. TrackFormer) może znacznie poprawić ciągłość trajektorii, szczególnie w scenach wieloobiektowych.

W niniejszym projekcie zastosowano prostą metodę kalibracji przestrzennej opartą na znanej długości odcinka widocznego w scenie. Jest to podejście powszechnie stosowane w sytuacjach, gdy brak jest pełnych parametrów kamery lub gdy UAV jest platformą ruchomą o zmiennym położeniu. Niemniej jednak metoda ta ma swoje ograniczenia – w szczególności:

- współczynnik piksel-metr może zmieniać się dynamicznie wraz z wysokością lotu,
- nachylenie kamery powoduje zniekształcenia perspektywiczne,
- brak jednorodności rozdzielczości przestrzennej w obrębie kadru.

Dokumentacja Ultralytics oraz liczne prace naukowe wskazują, że dokładniejsze systemy estymacji prędkości powinny wykorzystywać informacje o geometrii kamery, dane telemetryczne UAV lub metody rekonstrukcji 3D. W niniejszym projekcie celem było jednak zbudowanie prototypowego systemu, który

działa w praktycznych warunkach minimalnych, co uzasadnia zastosowanie prostej kalibracji.

Zarówno wyniki własne, jak i obserwacje z literatury, wskazują na kilka kierunków rozwoju omawianych metod:

1. Zwiększenie rozmiaru i różnorodności zbioru treningowego – pozwoli to na poprawę generalizacji modelu YOLOv8 i jego odporności na zmienne warunki lotu UAV.
2. Integracja zaawansowanych algorytmów śledzenia – takich jak Deep SORT, ByteTrack, BoT-SORT, zwiększyłaby to stabilność trajektorii pieszych.
3. Wykorzystanie danych telemetrycznych UAV umożliwiłoby dynamiczną kalibrację przestrzenną, eliminując zniekształcenia perspektywy.
4. Porównanie różnych wariantów YOLO w warunkach UAV – szczególnie ekscytujące byłoby przetestowanie YOLOv9, YOLOv11 oraz wariantów zoptymalizowanych dla małych obiektów (np. SOD-YOLO, LRDS-YOLO).
5. Pełna walidacja na większej liczbie nagrań referencyjnych – najlepiej z rzeczywistymi pomiarami prędkości pieszych.
6. Możliwość integracji analizy ruchu pieszych z modelami predykcji trajektorii – w scenach z UAV mogłoby to znaleźć zastosowanie w inteligentnych systemach bezpieczeństwa.

Wnioski

Analiza literatury oraz przeprowadzone badania własne jasno wskazują, że nowoczesne detektory obiektów z rodziny YOLO – w szczególności YOLOv8 – stanowią obecnie jedno z najbardziej obiecujących narzędzi do analizy obrazów pochodzących z bezzałogowych statków powietrznych (UAV). Ich zdolność do działania w czasie rzeczywistym, połączenie wysokiej dokładności z wydajnością oraz możliwość dostosowania architektury do specyfiki danych lotniczych czynią je wyjątkowo przydatnymi w zadaniach obejmujących monitorowanie ruchu pieszych.

Przegląd literatury potwierdza, że skuteczna detekcja obiektów w scenach lotniczych wymaga dostrojenia modeli do specyfiki danych UAV oraz rozwiązań lepiej radzących sobie z małymi obiektami. Wybrane prace (Kaur i in., 2025; Zeng i in., 2025; Fan i in., 2025) podkreślają znaczenie adaptacji architektury oraz

wzbogacania zbiorów danych, co znajduje potwierdzenie także w wynikach badań własnych. Badania eksperymentalne wykazały, że YOLOv8 – wytrenowany na dostosowanym zbiorze danych UAV – umożliwia skuteczną detekcję pieszych oraz stabilną estymację ich prędkości. Zastosowanie metod filtracji sygnału (średnia ruchoma i filtr gwałtownych zmian) znacząco poprawiło jakość uzyskiwanych wyników i pozwoliło na eliminację artefaktów wynikających z chwilowych błędów detekcji. Prosty algorytm śledzenia, oparty na minimalnej odległości pomiędzy detekcjami, okazał się wystarczający w analizowanej scenie, choć w bardziej złożonych środowiskach zasadna byłaby integracja z bardziej zaawansowanymi metodami śledzenia.

Porównanie YOLOv4, YOLOv5 i YOLOv8 wykazało wyraźną przewagę nowszej architektury, co jest zgodne z trendami opisanymi w literaturze. W szczególności YOLOv4 oraz YOLOv5 – bez dodatkowego trenowania – nie były w stanie wykrywać pieszych na wysokości 30 m, podczas gdy YOLOv8, przystosowany do tej perspektywy, zapewniał detekcje wystarczająco stabilne, aby możliwe było estymowanie prędkości.

Wnioski te prowadzą do kilku kluczowych obserwacji:

1. Nowoczesne architektury YOLO (v8 i nowsze) stanowią fundament praktycznych rozwiązań detekcyjnych dla UAV, pod warunkiem odpowiedniego dostosowania do danych lotniczych.
2. Trenowanie na dedykowanych zbiorach danych jest konieczne, aby uzyskać wysoką skuteczność detekcji pieszych z perspektywy pionowej.
3. Realistyczna estymacja prędkości wymaga zastosowania filtracji oraz śledzenia obiektów, nawet w systemach opartych na wysokiej jakości detektorach.
4. Obszar estymacji prędkości pieszych w danych UAV pozostaje stosunkowo słabo zbadany, co otwiera szerokie możliwości dalszych badań naukowych.
5. Zwiększenie skali zbiorów danych, integracja danych telemetrycznych UAV oraz zastosowanie zaawansowanych metod śledzenia mogą znacząco podnieść jakość przyszłych systemów.

Przedstawiona praca stanowi istotny wkład w rozwój narzędzi i metod analizy ruchu pieszych w obrazach UAV, łącząc aktualny stan wiedzy z praktyczną implementacją działającego prototypu systemu.

Literatura

1. Alhassan, M. A. M., Yılmaz, E. (2025). Evaluating YOLOv4 and YOLOv5 for Enhanced Object Detection in UAV-Based Surveillance. *Processes, 13*(1), 254. https://doi.org/10.3390/pr13010254
2. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv*. https://doi.org/10.48550/arXiv.2004.10934
3. Fan, Q., Li, Y., Deveci, M., Zhong, K., Kadry, S. (2025). LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle. *Information Sciences, 686*, 121366. https://doi.org/10.1016/j.ins.2024.121366
4. Jiao, D., Fei, T. (2023). Pedestrian walking speed monitoring at street scale by an in-flight drone. *PeerJ Computer Science, 9*, e1226. https://doi.org/10.7717/peerj-cs.1226
5. Kaur, D., Battish, N., Bhavsar, A., Poddar, S. (2025). Analysis of vision-based Air-to-Air UAV detection using YOLO based architectures. *Robotics and Autonomous Systems, 193*, 105074. https://doi.org/10.1016/j.robot.2025.105074
6. Luo, W., Yuan, S. (2025). Enhanced YOLOv8 for small-object detection in multiscale UAV imagery: Innovations in detection accuracy and efficiency. *Digital Signal Processing, 158*, 104964. https://doi.org/10.1016/j.dsp.2024.104964
7. Manjusha, M., Niveditha, R., Sunitha, V. (2025). Estimation of YOLO algorithms for extracting traffic data using UAV and stationary camera. *Innovative Infrastructure Solutions, 10*, 151. https://doi.org/10.1007/s41062-025-01898-y
8. Zeng, W., Wu, P., Wang, J., Hu, G., Zhao, J. (2025). C4D-YOLOv8: Improved YOLOv8 for object detection on drone-captured images. *Signal, Image and Video Processing, 19*, 862. https://doi.org/10.1007/s11760-025-04457-1

Netografia

1. Ultralytics. (2024a). Speed Estimation using Ultralytics YOLO26. Pobrane z: https://docs.ultralytics.com/guides/speed-estimation/ (dostęp: 12.12.2025).
2. Ultralytics. (2024b). Comprehensive Guide to Ultralytics YOLOv5. Pobrane z: https://docs.ultralytics.com/yolov5/ (dostęp: 12.12.2025).
3. Ultralytics. (2020). YOLOv5 vs. YOLOv8: Evaluating the Evolution of Ultralytics Vision AI. Pobrane z: https://docs.ultralytics.com/compare/yolov5-vs-yolov8/ (dostęp: 12.12.2025).