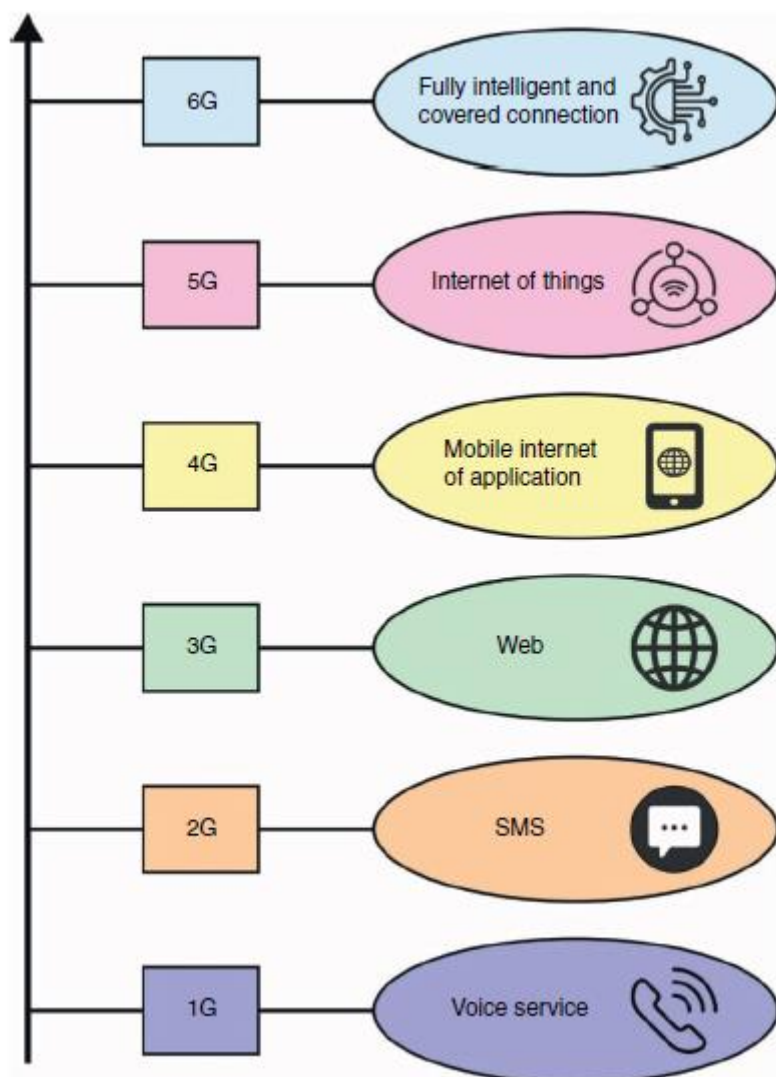


## **Adaptacyjna optymalizacja wielorybów z możliwością głębokiego uczenia się Sieć RefineDet do wspomaganie wzroku w sieciach 6G**

### **Wstęp**

W miarę postępu globalnego wdrażania sieci 5G, technologia sieciowa nowej generacji (6G) staje się zaawansowana technologicznie, aby wspierać sztuczną inteligencję (AI) we wdrażaniu „inteligentnych” sieci. Sieć piątej generacji (5G) zapewnia niezawodną komunikację, większą prędkość i mniejsze opóźnienia, co znacznie ułatwia codzienne życie użytkowników. Zaspokojenie rosnącego zapotrzebowania na urządzenia Internetu rzeczy (IoT) w zależności od obecnej technologii może być jednak skomplikowane, dlatego też koncepcja sieci szóstej generacji (6G) została zwiększona w celu ulepszenia obecnej sieci 5G i promowania rozwoju inteligentnych aplikacji. Wizja sieci 6G jest kolorowa; jednakże zrozumienie sieci 6G w dalszym ciągu stanowi problem, obejmujący wysoką przepustowość, wysoki szczyt, wysoką efektywność energetyczną, łączność w każdym miejscu, nowe technologie i teorie oraz inne problemy nietechniczne [3]. Uznaje się, że obiecująca technologia rozwoju sieci 6G obejmuje: (i) techniki podstawowe, mianowicie teorię rzadkiej, elastyczne widmo, nowe kodowanie kanałów i antenę wielkoskalową; (ii) odrębne techniki realizacji jednolitej komunikacji przestrzeń-powietrze-ziemia-morze i bezprzewodowej sieci dotykowej; (iii) technika transmisji o rozszerzonym widmie. W pewnym momencie architektura sieci 6G staje się bardziej niejednorodna i większa, a scenariusz aplikacji staje się bardziej zmienny i złożony. Wykorzystanie sztucznej inteligencji do rozwiązania tego problemu w technikach transmisji bezprzewodowej, takich jak transmisja sygnału, alokacja zasobów i efektywność energetyczna, jest prawie nieuniknioną opcją. Rysunek ilustruje postęp sieci 6G.



Nauka i inżynieria dokonują technologicznej interwencji w życie osób niewidomych, aby uczynić je niezależnymi w zakresie postrzegania i poruszania się po otaczających obiektach [5]. Zaprezentowano różne urządzenia pomagające osobom niewidomym; jednakże wiele urządzeń koncentruje się na rozpoznawaniu obiektów za pomocą wizji komputerowej (CV) lub rozpoznawaniu przeszkód za pomocą odrębnych węzłów czujnikowych, mianowicie czujników odległości, GPS i tak dalej. Jednak efektywne wykorzystanie technik opartych na czujnikach i CV może prowadzić do powstania wspierających i bardzo skutecznych urządzeń, dzięki którym pacjenci będą świadomi swojego otoczenia. Osoba niewidoma korzysta z elektronicznych pomocy podróży (ETA) do wykrywania problemów i identyfikowania udogodnień, aby zapewnić informacyjną i bezpieczną nawigację. Zmniejsza ryzyko upadku i pozwala zachować równowagę człowieka. Waga, długość i koszt to inne parametry, które poprawia się w celu zapewnienia najlepszego zaopatrzenia. Do rehabilitacji wzroku osób niewidomych opracowano elektroniczną laskę do poruszania się (EMC), służącą do wykrywania problemów i zapewniania wsparcia, za pomocą której budowana jest mapa logiczna w celu uzyskania danych dotyczących otoczenia. Dane wyjściowe są przesyłane za pomocą głosu, dźwięku lub wibracji. Charan i inni wykorzystują głębokie uczenie się (DL) i CV. Prezentuje nowatorskie rozwiązanie, które aktywnie prognozuje blokady połączeń dynamicznych. Szczegółowo rozwija strukturę głębokiej sieci neuronowej (DNN), która uczy się na podstawie wykrytej sekwencji obrazów RGB i wektora tworzącego wiązkę w

celu przewidywania możliwych dalszych blokad połączeń. Zaprezentowano ulepszoną technikę rozpoznawania obiektów zależną od klatki kluczowej wideo w celu zmniejszenia opóźnień w schemacie brzegowego Internetu pojazdów (IoV). Znacząco zwiększa skuteczność redukcji opóźnień kosztem mniejszej dokładności rozpoznawania. W ramach tego schematu można go dostosować do istotnego współczynnika i techniki porównywania podobieństwa klatek w celu filtrowania zbędnych klatek i uzyskania rozpoznawania klatki kluczowej z obiektem. Następnie do rozpoznawania obiektów w metodzie obliczania krawędzi wykorzystano ulepszoną technikę klasyfikatora opartą na cechach Haara. Ukhwah i inni wykorzystali YOLO z trzema odrębnymi strukturami infrastruktury, na przykład Yolo v3, Yolo v3 Tiny i Yolo v3 SPP, umożliwili stworzenie dalszej dokładnej oceny w celu wykrycia dziur w nawierzchni drogi. Autorzy zaprezentowali model, którego celem jest pomoc osobom niedowidzącym poprzez oferowanie pomocy dźwiękowej wskazującej im, jak unikać przeszkód ułatwiających im poruszanie się z otoczenia. Wykrywanie obiektów za pomocą YOLO pomaga im w wykrywaniu sąsiednich obiektów, a szacowanie głębokości za pomocą widzenia jednoocznego określa szacowaną odległość wykrycia obiektów przez użytkowników. Ning i inni wykorzystują uczenie transferowe (TL) w procesie wykrywania pojedynczego strzału (SSD) do rozpoznawania obiektów i klasyfikatora, a następnie wykrywania ludzkiej twarzy i banknotu, gdy zostaną zauważone, z wykorzystaniem metody Inception v3. Detektor SSD został przeszkolony na zmienionym zbiorze danych PASCAL VOC 2007, natomiast dodano nową klasę, która umożliwiła również rozpoznawanie waluty. Co więcej, odrębnymi metodami Inception v3 były szkolenia w zakresie rozpoznawania ludzkiej twarzy i banknotów, tworząc w ten sposób infrastrukturę skalowalną i regulowaną w oparciu o preferencje użytkownika. Joshi i inni zaprezentowali w pełni automatyczne technologie wspomagające oparte na sztucznej inteligencji służące do rozpoznawania odrębnych obiektów, a użytkownikom zaoferowano sygnały dźwiękowe w czasie rzeczywistym, które zapewniają osobom niedowidzącym (VIP) optymalne zrozumienie ich otoczenia. Technika DL było szkolenie z kilkoma obrazami obiektów, które są wyjątkowo odpowiednie dla VIP-ów. Oprócz podejścia do wykrywania obiektów opartego na CV, połączono czujnik pomiaru odległości, aby uczynić urządzenia jeszcze bardziej wszechstronnymi, identyfikując przeszkody, ale jednocześnie nawigując z jednego miejsca do drugiego. W tym badaniu przedstawiono adaptacyjną optymalizację wielorybów za pomocą sieci RefineDet z obsługą głębokiego uczenia się (AWO-DLRDN) do pomocy wizualnej w sieciach 6G. Proponowana technika AWO-DLRDN początkowo przeprowadza proces powiększania danych i adnotacji obrazu jako etap wstępnego przetwarzania. Ponadto model RefineDet wykorzystywany jest do identyfikacji obiektów istniejących na obrazie, a hiperparametry modelu RefineDet są optymalnie dobierane przy wykorzystaniu algorytmu AWO. Ponadto obliczane są przybliżone odległości między obiektami i osobami, aby pomóc osobom niedowidzącym. Eksperymentalna analiza wyników techniki AWO-DLRDN jest przeprowadzana z wykorzystaniem wzorcowego zbioru danych.

### **Proponowany model**

W tym badaniu opracowano nowatorską technikę AWO-DLRDN w celu określenia pobliskich obiektów i ich przybliżonej odległości od osób niedowidzących. Proponowana technika AWO-DLRDN obejmuje powiększanie danych, adnotację obrazu, wykrywanie obiektów w oparciu o technologię RefineDet, optymalizację hiperparametrów w oparciu o AWO i przybliżone szacowanie odległości.

### **Powiększanie obrazu i adnotacja**

Każdy zebrany obraz jest następnie wzmacniany, aby oprzeć się wyszkolonej metodzie w przypadku nadmiernego dopasowania i zapewnić dalsze niezawodne i dokładne rozpoznawanie obiektów dla VIP-ów. Do kilkukrotnego wzbogacenia zbioru danych wykorzystuje się kilka metod udoskonalania, a mianowicie obrót pod różnymi kątami, pochylanie, odbicie lustrzane, odwracanie, poziom jasności, poziom szumu oraz grupę tych podejść. Każdy obraz jest opisywany ręcznie za pomocą narzędzi

Labellmg, a ramka ograniczająca jest opracowywana wokół obiektu bez zajmowania dodatkowych, niepotrzebnych obszarów. Dane obrazu, takie jak rozmiar obrazu, rozmiar i położenie ramek ograniczających (w przypadku kilku próbek lub kilku obiektów z podobnego obrazu) są rejestrowane i zapisywane w formacie „.xml”. Jeśli obraz jest opatrzony adnotacją, tworzony jest również odpowiedni plik adnotacji. Ostatni zestaw danych zawiera obraz z adnotacjami, a odpowiedni plik adnotacji został rozdzielony na dwa zestawy, takie jak trenowany i sprawdzający.

### Wykrywanie obiektów w oparciu o udoskonalenie Det

W celu skutecznej detekcji obiektów zastosowano model RefineDet. RefineDet to jednoetapowa technika zależna od architektury SSD i obejmująca moduł wykrywania obiektów (ODM) i model udoskonalania kotwicy (ARM). ARM przekazuje ujemną, twardą kotwicę i dodatnią kotwicę do ODM, który próbuje klasyfikować i lokalizować obiekty docelowe w obrazie wejściowym. W bieżącej pracy wykorzystaliśmy RefineDet jako model porównawczy z następujących powodów: (i) jest skuteczny ze względu na jednoetapową strukturę; oraz (ii) wykorzystuje wyrafinowaną metodę, która naśladuje „metodę wykrywania” w celu znalezienia możliwych obszarów w docelowych znakach drogowych, niezależnie od ich klas. Obecnie „metoda wykrywania” różni się od tej stosowanej w konwencjonalnym wykrywaniu znaków drogowych (tzn. rozpoznaje każdą kategorię funkcji specyficzną dla znaków drogowych na oryginalnym obrazie). RefineDet to skuteczna metoda wykrywania obiektów stosowana do wykrywania obiektów z większą dokładnością i szybkością, ale nie jest konkurencyjna w stosunku do zaawansowanych metodologii wykrywania mniejszych obiektów. Po pierwsze, funkcja płytkiej warstwy w programie RefineDet i wykorzystywana do wykrywania mniejszych obiektów zawiera dane dotyczące ograniczeń, które nie są wystarczająco mocne, aby skutecznie identyfikować mniejsze obiekty.

### Strojenie hiperparametrów przy użyciu algorytmu AWO

Do dopasowania wartości hiperparametrów modelu RefineDet wykorzystano algorytm AWO. Przeszukiwanie algorytmu optymalizacji wielorybów (WOA) dzieli się na dwa procesy : podejście do ataku sieci bąbelkowej (proces eksploatacji) i poszukiwanie ofiary (faza eksploatacji). Początkowo WOA realizuje wyszukiwanie zdobyczy, które jest określane przez:

$$D = |CX_{rand} - X(t)| \quad (6.1)$$

$$X(t + 1) = X_{rand} - AD \quad (6.2)$$

gdzie D wskazuje odległość między celem a obecnymi pozycjami, t wskazuje numer iteracji, A i C pokazują wektory współczynników,  $A = 2a r - a$ ,  $C = 2r$ , w którym a jest minimalizowane liniowo od 2 do 0 poprzez iterację ( etapy poszukiwań i eksploatacji), a r oznacza dowolną liczbę z zakresu [0, 1],  $X_{rand}$  wskazuje lokalizację dowolnego wieloryba w populacji, a  $X(t)$  oznacza lokalizację istniejącego wieloryba. Zwykle humbaki atakują sieć bąbelkową, aby zerować w następujący sposób

$$X(t + 1) = \begin{cases} X^*(t) - AD & \text{if } p < 0.5 \\ D' e^{bl} \cos(2\pi l) + X^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (6.3)$$

Tutaj  $D' = X^*(t) - X(t)$  i reprezentuje odległość i od wieloryba do ofiary (dotychczas osiągnięte rozwiązanie optymalne), b oznacza stałą do określenia kształtu spirali logarytmicznej, l pokazuje dowolną wartość w [-1, 1], a p wskazuje dowolną wartość w [0, 1]. Równowaga pomiędzy eksploracją i eksploatacją metaheurystyk był celem badania związanym z ogólną efektywnością modelu. Gdy eksploracja jest silniejsza, precyzja zbieżności jest minimalizowana; w przeciwnym razie łatwiej jest

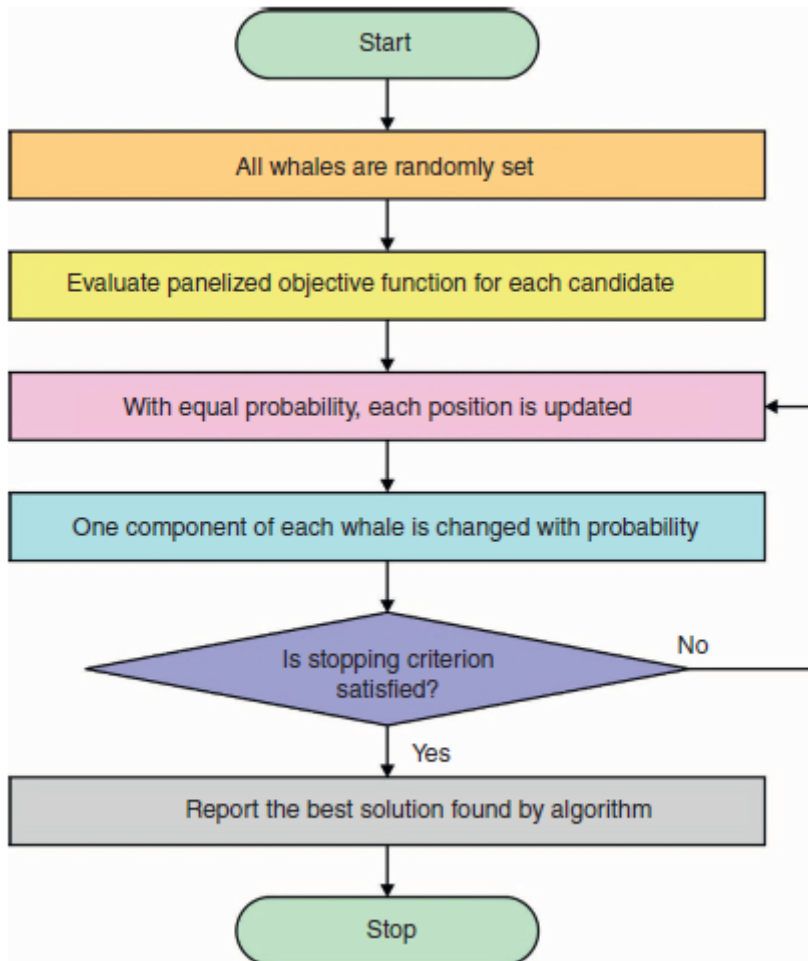
wpaść w lokalne optymalne. Dlatego w pracy zaprezentowano adaptacyjną ulepszoną metodę eksploracyjną i stworzono algorytm AWO. Na etapie eksploracji WOA poszukiwanie ofiary oznacza, że jeden osobnik eksploruje w pobliżu innych osobników z populacji. We wszystkich iteracjach, aby rozłożyć obszar eksploracji na całą przestrzeń rozwiązań, w badaniu przedstawiono adoptywną, zwiększoną metodę eksploracyjną. Podczas gdy wieloryby adaptacyjnie wybierają wewnętrzny i zewnętrzny region populacji w celu zbadania wielkości przestrzeni populacji.

$$p(\delta_d) = 1 - \sin\left(\delta_d \times \frac{\pi}{2}\right) \quad (6.4)$$

$$\delta_d = \frac{\max(x) - \min(x)}{x_d^{\max} - x_d^{\min}} \quad d = 1, 2, \dots, D \quad (6.5)$$

$$X(t+1) = \begin{cases} X_{rand} - AD & \text{if } p' \geq p(\delta_d) \\ \left. \begin{array}{l} rand * (x_d^{\max} - \max(x_d)) + \max(x_d) \quad \text{if } p'' \leq 0.5 \\ rand * (\min(x_d) - x_d^{\min}) + x_d^{\min} \quad \text{if } p'' > 0.5 \end{array} \right\} & \text{if } p < p(\delta_d) \end{cases} \quad (6.6)$$

Tutaj  $\delta_d$  oznacza proporcję populacji do przestrzeni rozwiązań dla d-tego parametru,  $x_d$  reprezentuje zbiór d-tych zmiennych każdego osobnika,  $x_d^{\max}$  i  $x_d^{\min}$  pokazują górną i dolną granicę obszaru poszukiwań,  $rand$ ,  $p'$ , a  $p''$  oznacza dowolny wektor od 0 do 1,  $p(\delta_d)$  oznacza prawdopodobieństwo eksploracji poza przestrzeń populacji, czyli odwrotnie proporcjonalne do  $\delta_d$  i jej krzywej. Rysunek przedstawia schemat blokowy techniki WOA.



### Pomiar odległości

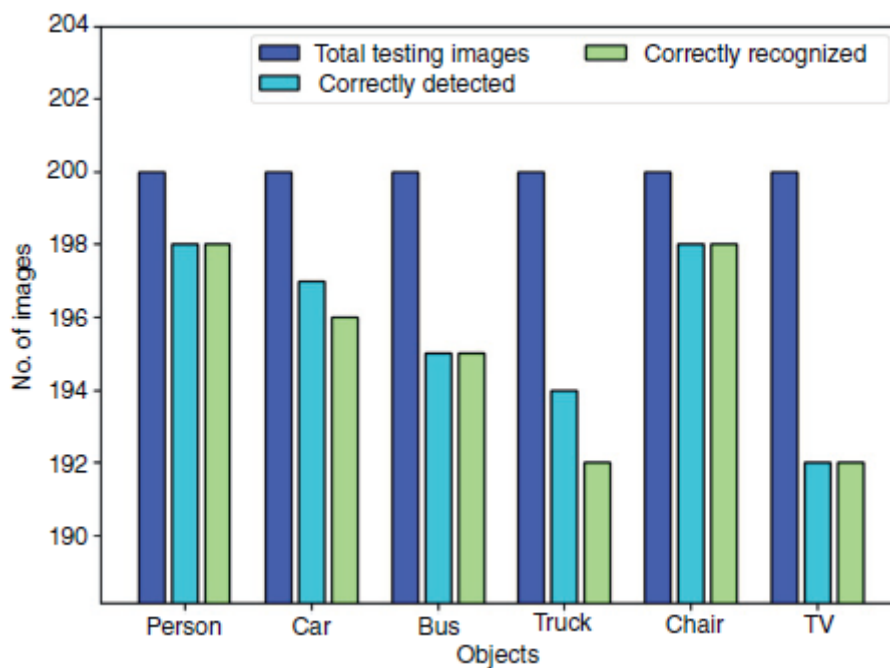
Aby uzyskać odległość między użytkownikiem a wykrytym obiektem, może być konieczne użycie co najmniej dwóch punktów widzenia w celu odwzorowania obrazu 3D i utworzenia mapy głębi. Wizja stereoskopowa wykorzystuje dwa obrazy skalibrowane kamerą do postrzegania infrastruktury 3D świata i tworzenia jej mapy głębi. Kalibracja dwóch kamer jest kompletna, jeśli znasz ich wewnętrzne parametry, takie jak środek obrazu, zniekształcenie, ogniskowa, pochylenie itp., a także ich parametry zewnętrzne, które wyjaśniają ich położenie i orientację w stosunku do otoczenia. Do oszacowania głębokości z wykorzystaniem widzenia stereoskopowego może być wymagana ogniskowa i odległość między dwiema kamerami. W tym momencie  $x^l$  i  $x^r$  reprezentują odległość pomiędzy osiami optycznymi i rzut obiektu „P” na wirtualne płaszczyzny dwóch kamer. Gdy  $b$  odnosi się do odległości bazowej między dwiema kamerami, a  $f$  oznacza ogniskową połączonych kamer, odległość prostopadła między obiektem „P” a linią bazową kamer wyraża się jako:

$$Z = \frac{b * F}{x^l + x^r} \quad (6.7)$$

### Wyniki i dyskusja

Wydajność modelu AWO-DLRDN testuje się przy użyciu zbioru danych składającego się z sześciu klas po 200 obrazów w każdej klasie. Tabela i rysunek przedstawiają szczegółowe wyniki detekcji obiektów w modelu AWODLRDN na obrazach testowych.

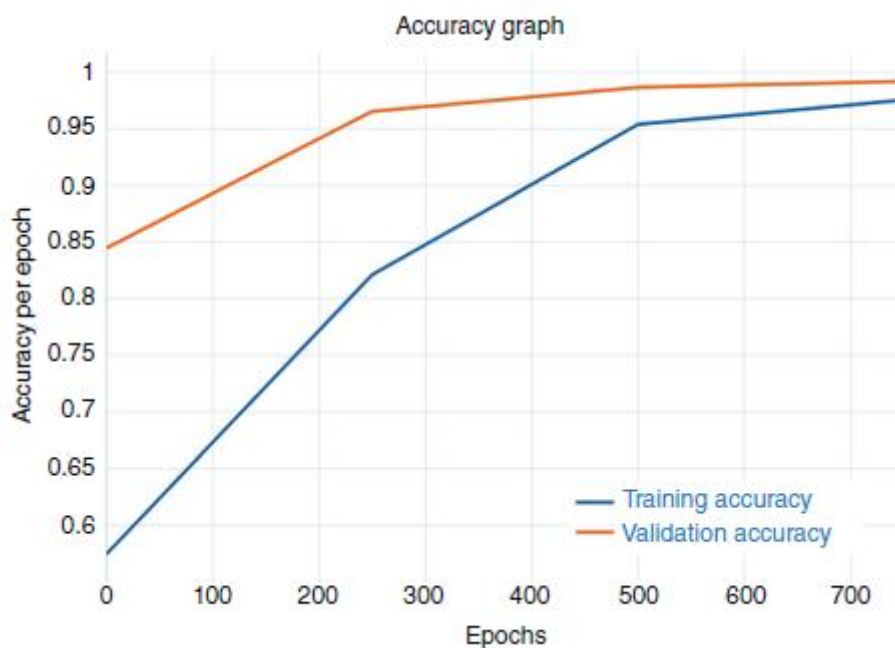
Objects	Total testing images	Correctly detected	Correctly recognized
Person	200	198	198
Car	200	197	196
Bus	200	195	195
Truck	200	194	192
Chair	200	198	198
TV	200	192	192



Zaproponowany model AWO-DLRD prawidłowo wykrył i rozpoznał obiekty. W przypadku obiektów będących osobami model AWO-DLRD prawidłowo wykrył i rozpoznał odpowiednio 198 i 198 obiektów. Poza tym model AWODLRD wykrył i rozpoznał 197 i 196 obiektów samochodowych. Ponadto model AWO-DLRD wykrył i rozpoznał 194 i 192 obiekty ciężarowe. Tabela przedstawia krótką analizę wyników modelu AWO-DLRD pod względem dokładności wykrywania (DA) i dokładności rozpoznawania.

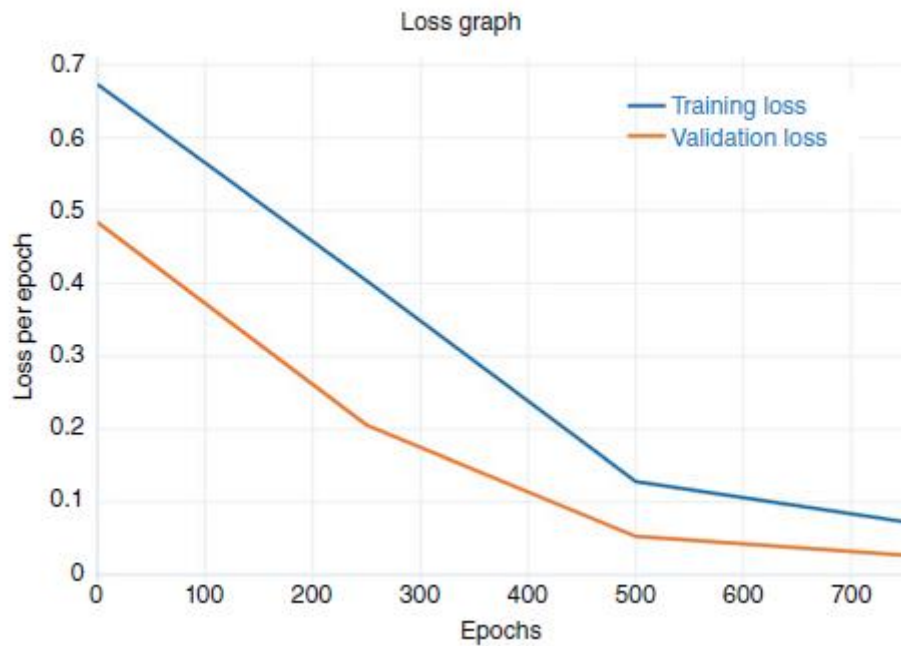
Objects	Detection accuracy (%)	Recognition accuracy (%)
Person	99.00	100.00
Car	98.50	99.49
Bus	97.50	100.00
Truck	97.00	98.97
Chair	99.00	100.00
TV	96.00	100.00

Tabela wskazuje, że model WO-DLRDN wykrył osobę, samochód, autobus, ciężarówkę, krzesło i obiekty telewizyjne ze zwiększonym DA wynoszącym odpowiednio 99, 98,50, 97,50, 97, 99 i 96%. Poza tym model AWO-DLRDN rozpoznaje osobę, samochód, autobus, ciężarówkę, krzesło i obiekty telewizyjne ze zwiększoną dokładnością rozpoznawania odpowiednio 100, 99,49, 100, 98,97, 100 i 100%. Analizę wyników dokładności techniki AWO-DLRDN na danych testowych przedstawiono na rysunku.



Wyniki wykazały, że technika AWO-DLRDN osiągnęła lepszą dokładność walidacji w porównaniu z dokładnością uczenia. Można również zauważyć, że wartości dokładności ulegają nasyceniu wraz z liczbą epok. Analizę wyników strat techniki AWO-DLRDN na danych testowych przedstawiono na rysunku .

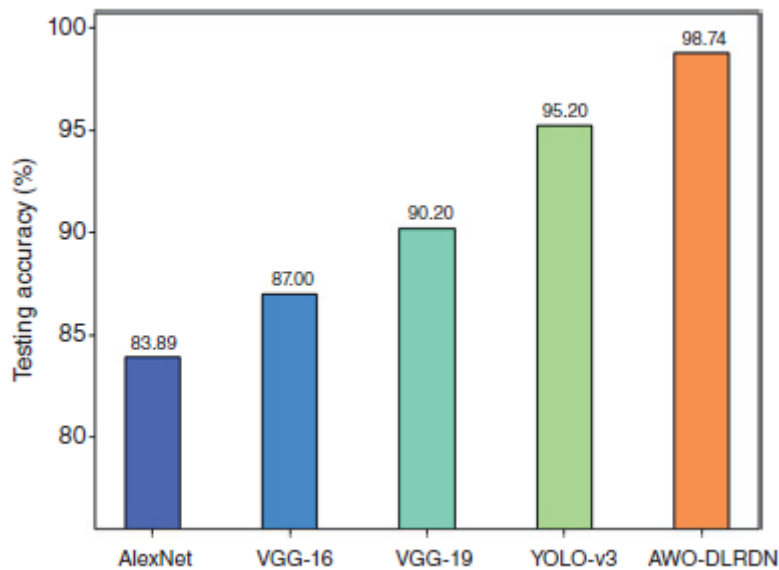




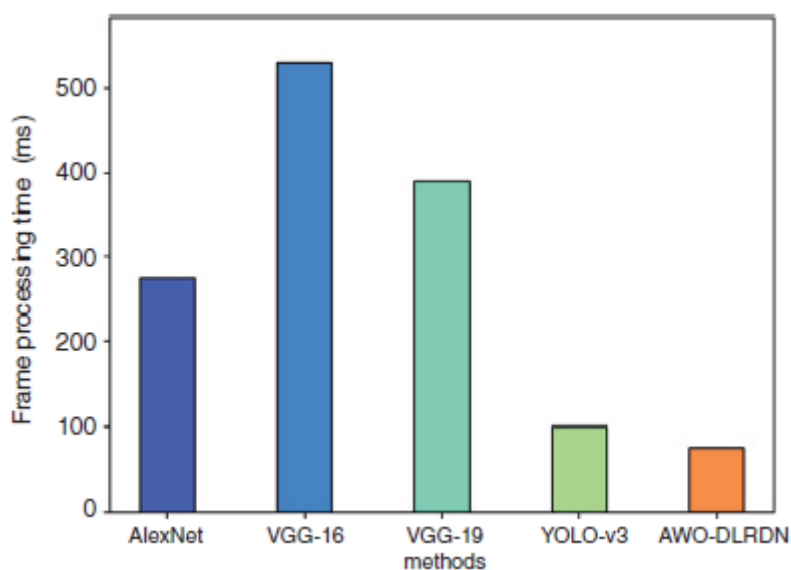
Rysunek ujawnił, że plik AWODLRDN technika ta wykazała zmniejszoną utratę walidacji w stosunku do utraty szkolenia. Dodatkowo zauważono, że wartości strat ulegają nasyceniu wraz z liczbą epok. Tabela ilustruje analizę porównawczą techniki AWO-DLRDN z najnowszymi podejściami w zakresie dokładności testowania i czasu przetwarzania ramki (FPT).

Methods	Testing accuracy	Frame processing time (ms)
AlexNet	83.89	275.00
VGG-16	87.00	530.00
VGG-19	90.20	390.00
YOLO-v3	95.20	100.00
AWO-DLRDN	98.74	75.00

Rysunek przedstawia badanie dokładności testowania modelu AWO-DLRDN.

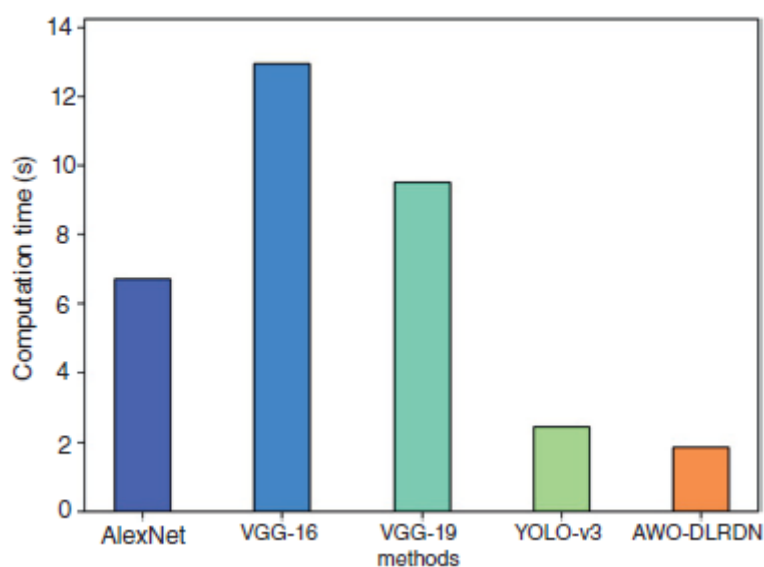


Wyniki wykazały, że modele AlexNet, VGG-16 i VGG-19 osiągnęły minimalną dokładność testową odpowiednio 83,89, 87 i 90,20%. Oprócz tego model YOLO-v3 osiągnął nieco zwiększoną dokładność testów wynoszącą 95,20%. Jednak model AWO-DLRDN osiągnął wyższą dokładność testowania wynoszącą 98,74%. Kompleksową inspekcję FPT modelu AWO-DLRDN z innymi modelami DL przedstawiono na rysunku.



Rysunek pokazuje, że model VGG-16 uzyskał wyższy FPT wynoszący 390 ms, podczas gdy modele AlexNet i VGG-19 uzyskały nieznacznie zmniejszony FPT wynoszący odpowiednio 275 i 390 ms. Choć model YOLO-v3 dał rozsądny FPT wynoszący 100 ms, przedstawiony model AWO-DLRDN przewyższa inne metody z co najmniej FPT wynoszącym 75 ms. Na koniec badanie czasu obliczeniowego (CT) modelu AWO-DLRDN z innymi modelami DL przedstawiono w tabeli i rysunku.

Methods	Computation time (s)
AlexNet	6.71
VGG-16	12.94
VGG-19	9.52
YOLO-v3	2.44
AWO-DLRDN	1.83



Wyniki eksperymentów wykazały, że model VGG-16 uzyskał wyższy współczynnik CT wynoszący 12,94 sekundy, podczas gdy modele AlexNet i VGG-19 uzyskały nieznacznie zmniejszony współczynnik CT wynoszący odpowiednio 6,71 i 9,52 sekundy. Chociaż model YOLO-v3 dał rozsądny CT wynoszący 2,44 sekundy, przedstawiony model AWO-DLRDN przewyższa inne metody z najmniejszym CT wynoszącym 1,83 sekundy. Wyniki wykazały poprawę modelu AWO-DLRDN w porównaniu z innymi istniejącymi technikami.

### Wniosek

W tym badaniu opracowano nowatorską technikę AWO-DLRDN aby określić pobliskie obiekty i ich przybliżoną odległość od osób niedowidzących w sieciach 6G. Proponowana technika AWO-DLRDN obejmuje powiększanie danych, adnotację obrazu, udoskonalanie wykrywania obiektów w oparciu o Det, optymalizacja hiperparametrów w oparciu o AWO i przybliżone szacowanie odległości. Hiperparametry modelu RefineDet są optymalnie dobierane poprzez zastosowanie algorytmu AWO umożliwiającego dokładną identyfikację obiektów występujących na obrazie. Na koniec obliczane są przybliżone odległości między obiektami i osobami, aby pomóc osobom niedowidzącym. Eksperymentalna analiza wyników techniki AWODLRDN została przeprowadzona przy użyciu

wzorcowego zbioru danych, a badanie porównawcze uwypukliło ulepszenia techniki AWO-DLRDN w porównaniu z innymi technikami. W przyszłości technikę AWO-DLRDN będzie można wdrożyć w aplikacji na smartfony.