

Metody zwiększania ilości danych termowizyjnych w uczeniu maszynowym

Piotr Sadzyński

Wojskowa Akademia Techniczna, Instytut Optoelektroniki, ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa

Streszczenie: Coraz częściej w przetwarzaniu i analizie obrazu termowizyjnego stosuje się uczenie maszynowe w kontekście rozpoznawania i identyfikacji obiektów. Niniejszy artykuł prezentuje wyniki badania wpływu augmentacji danych na efektywność uczenia maszynowego w kontekście analizy obrazów termowizyjnych. Wykorzystano publicznie dostępny zbiór danych FLIR ADAS, który zawiera etykietowane obrazy termowizyjne i obrazy z zakresu światła widzialnego. Badanie skupia się na wykorzystaniu konwolucyjnych sieci neuronowych, w szczególności architektury YOLOv8, do detekcji obiektów na obrazach termowizyjnych. Zbiór danych FLIR ADAS został poddany wstępnemu przetwarzaniu i augmentacji, a następnie wykorzystany do trenowania dwóch różnych modeli: jednego opartego na obrazach w skali szarości i drugiego – opartego na obrazach z zastosowaną paletą kolorów. Wyniki eksperymentu wskazują, że augmentacja danych może znacząco wpłynąć na efektywność modelu, a zastosowanie kolorów w obrazach termowizyjnych może w pewnych sytuacjach dodatkowo zwiększyć dokładność detekcji.

Słowa kluczowe: augmentacja danych, konwolucyjne sieci neuronowe, głębokie uczenie, obrazowanie w podczerwieni, wykrywanie obiektów, pomiary, wizja komputerowa, przetwarzanie obrazu, rozpoznawanie obiektów

1. Wprowadzenie

W dziedzinie uczenia maszynowego, a szczególnie w aspekcie rozpoznawania obiektów, skuteczność modelu w dużym stopniu zależy od wielkości i różnorodności zbioru danych treningowych [1]. W idealnym przypadku, dysponowalibyśmy obszernymi i zróżnicowanymi zbiorami danych, odzwierciedlającymi wszystkie perspektywy i warunki pogodowe, w których model ma być używany. Niestety, w praktyce rzadko jest to możliwe, szczególnie w dziedzinach specjalistycznych, takich jak zastosowania wojskowe.

Ten problem staje się jeszcze bardziej wyraźny, gdy przechodzimy do analizy obrazów termowizyjnych. Dostęp do dużych zbiorów danych obrazów termowizyjnych interesujących nas obiektów jest jeszcze bardziej ograniczony. W tym kontekście, augmentacja danych, czyli technika generowania nowych próbek danych przez modyfikację istniejących, wydaje się być obiecującym rozwiązaniem.

Dane uczące stanowią podstawę każdego procesu uczenia maszynowego. Jest to zbiór przykładowych danych wejściowych oraz odpowiedzi, jakich oczekujemy od trenowanego nimi modelu uczenia maszynowego. W kontekście rozpoznawania

obiektów, danymi wejściowymi są obrazy z etykietowanymi na nich obiektami. Etykietowanie jest to proces polegający na przypisywaniu lokalizacji (zazwyczaj w formie prostokąta) i klas (np. „samochód”, „osoba”, „pies”) do każdego obiektu na obrazie. Zaleca się, aby właściwy zestaw danych zawierał co najmniej 10 000 etykietowanych obiektów, rozłożonych na minimum 1500 obrazach dla każdej klasy obiektów [1]. Obrazy powinny odzwierciedlać różne sytuacje, w których model będzie używany – na przykład różne poziomy światła, różne kąty widzenia obiektów i różne tła [1].

W obszarze obrazowania termowizyjnego jest kilka publicznie dostępnych zbiorów danych, takich jak „Deep Thermal Imaging Dataset” [2] oraz „Thermal Image dataset for object classification” [3]. Niemniej jednak, publicznie dostępne zbiory nie zawierają wszystkich możliwych obiektów, które mogą być przedmiotem zainteresowania. Na przykład, jeśli celem jest rozpoznawanie helikopterów, często konieczne jest korzystanie z mniejszych zbiorów danych.

Jednym z możliwych rozwiązań dla tego problemu jest augmentacja istniejących danych, co pozwala na sztuczne zwiększenie liczby dostępnych obrazów i tym samym poprawę wydajności modelu, więcej o augmentacji przedstawiono w rozdziale 2. Innym rozwiązaniem jest „transfer learning” [4]. Ta metoda polega na adaptacji modelu, który został nauczony na dużym, ogólnym zbiorze danych, do rozpoznawania obiektów w innym środowisku [5]. Przykładowo, model nauczony na zbiorze danych zawierającym różne typy pojazdów w świetle widzialnym może zostać dostosowany do identyfikacji ciężarówek na obrazach termowizyjnych.

W artykule skupiono się na augmentacji danych jako głównym środkiem do poprawy wydajności modeli w obrazowaniu termowizyjnym. Wybrano to podejście z kilku powodów. Po pierwsze, augmentacja danych jest techniką dostępną i nie

Autor korespondujący:

Piotr Sadzyński, piotr.sadzynski@student.wat.edu.pl

Artykuł recenzowany

nadesłany 04.08.2023 r., przyjęty do druku 18.01.2024 r.



Zezwala się na korzystanie z artykułu na warunkach licencji Creative Commons Uznanie autorstwa 3.0



a)



b)

Rys. 1. Przykład techniki zmiany jasności obrazów: a) oryginał; b) ściemnienie obrazu

Fig. 1. Example of an operation to change the brightness of images: a) original; b) darkening the image

wymaga specjalistycznej wiedzy do implementacji. Po drugie, jest to metoda skalowalna, która może być stosowana dla różnych typów danych. Po trzecie, augmentacja oferuje większą kontrolę nad jakością i różnorodnością generowanych obrazów.

2. Metody augmentacji w przetwarzaniu obrazów termowizyjnych

Augmentacja danych polega na tworzeniu zmodyfikowanych wersji istniejących obrazów w celu zwiększenia i urozmaicenia zbiorów danych treningowych. Poprzez wprowadzanie modyfikacji, takich jak obracanie, rozjaśnianie czy zmiana kontrastu, uzyskuje się dodatkowe warianty tych danych. W efekcie, modele uczenia maszynowego trenowane na tak wzbogaconych danych mogą osiągnąć lepszą zdolność generalizacji i być bardziej odporne na przeuczenie (ang. *Overfitting*) [6, 7].

Należy podkreślić, że nie istnieje uniwersalny zestaw technik augmentacji, który byłby optymalny dla każdego zbioru danych. Każdy zbiór danych wymaga indywidualnego podejścia, uwzględniającego kontekst, w którym jest używany, oraz

rodzaje obiektów, które ma reprezentować. Techniki augmentacji można podzielić na trzy główne kategorie [8]:

- transformacje kolorystyczne,
- transformacje geometryczne,
- transformacje fragmentów obrazu RoI (ang. *Region of Interest*).

Transformacje kolorystyczne polegają na modyfikacji wartości kanałów kolorystycznych, bez zmiany położenia ramki. Przykładowo: zmiana odcienia kolorów, nasycenia, jasności, lub kontrastu.

Transformacje geometryczne polegają na geometrycznym zniekształceniu obrazu, co objawia się zmianą położenia i wielkości ramek. Przykładowo: obrót, translacja w osi X lub Y, pochylenie obrazu.

Transformacje na obwiedniach polegają na edycji tylko tej części obrazu znajdującej się wewnątrz ramki [9].

2.1. Zmiana jasności obrazów

Modyfikacja jasności jest techniką, która pomaga uodpornić model na zmianę oświetlenia. Na przykład, różne poziomy jasności mogą symulować różne warunki środowiskowe, takie jak wysokie zachmurzenie, czy deszcz. Technika **zmiany**



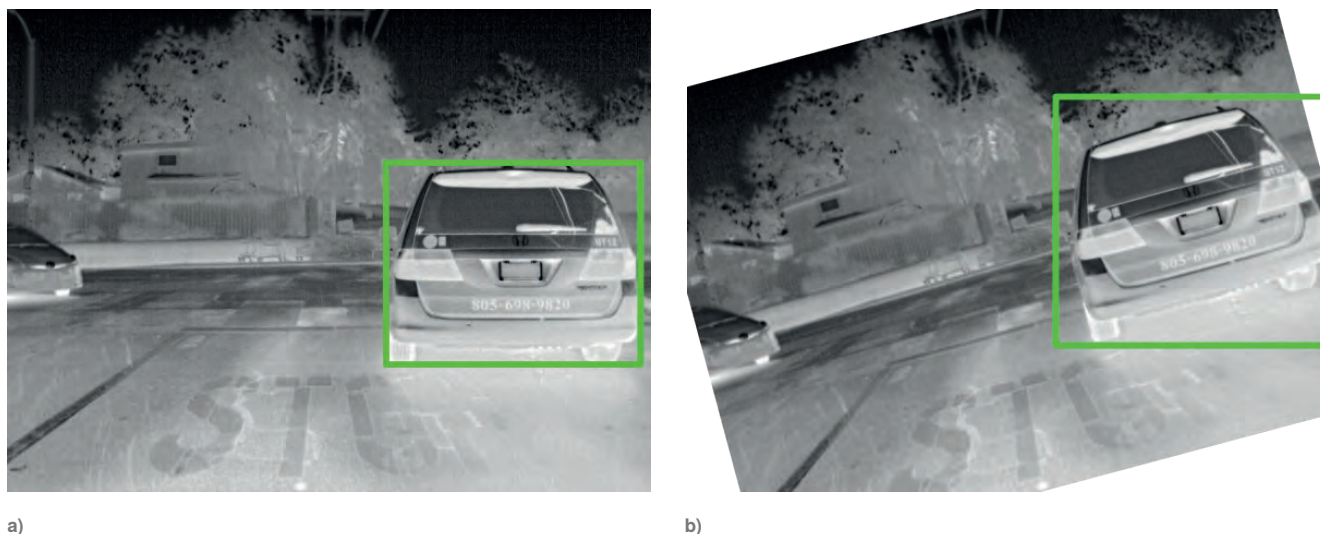
a)



b)

Rys. 2. Przykład techniki szum typu „sól i pieprz”: a) oryginał; b) po zastosowaniu szumu

Fig. 2. Example of "salt and pepper" noise operation: a) original; b) after applying noise



Rys. 3. Przykład techniki obrotu obrazu: a) oryginał; b) po zastosowaniu obrotu
 Fig. 3. Example of rotation operation: a) original; b) after applying rotation

jasności obrazów (rys. 1b) polega na dodaniu stałej wartości do każdego piksela obrazu, co skutkuje zwiększeniem lub zmniejszeniem jasności obrazu. Jeśli I jest obrazem, a c jest stałą, nowy obraz T jest generowany jako $T = I + c$, gdzie operacja jest wykonywana dla każdego piksela. Wartości c są ograniczone do pewnego zakresu.

2.2. Szum typu „sól i pieprz”

Szum typu „sól i pieprz” to szum, który manifestuje się jako przypadkowe białe i czarne piksele na obrazie, przypominające ziarna soli i pieprzu. Chociaż ludzki system percepcyjny jest zdolny do ignorowania takich szumów i rozszyfrowywania zawartości obrazów pomimo ich obecności, algorytmy uczenia maszynowego często napotykają trudności w radzeniu sobie z tymi zakłóceniami [10]. Jednym z wyzwań w tej dziedzinie są tzw. ataki adversarialne (ang. *adversarial attacks*) [11], w których niewielkie, niewidoczne dla ludzkiego oka zmiany w pikselach obrazu mogą znacząco wpłynąć na dokładność predykcji sieci neuronowej. Wprowadzenie takiego szumu podczas augmentacji danych zwiększa odporność modeli na różnorodne zakłócenia, ucząc je ignorowania nieistotnych zmian w obrazie. **Szum typu „sól i pieprz”** (rys. 2b) polega na losowym zamienianiu pikseli

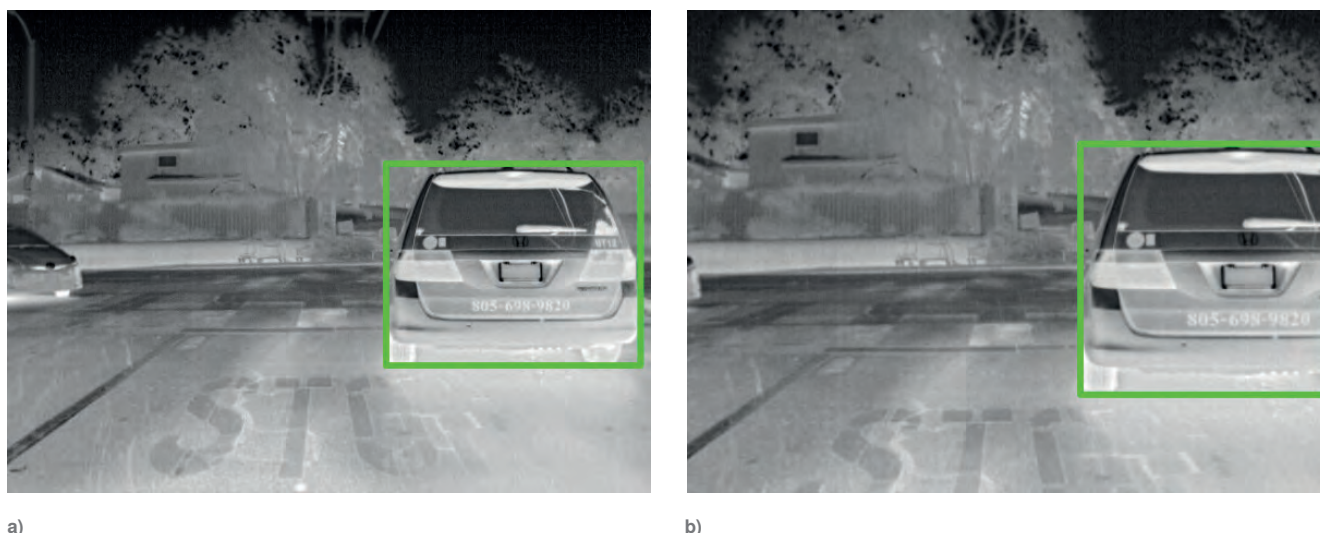
na biały (sól) lub czarny (pieprz). Dla obrazów 8-bitowych losowo wybrane piksele są ustawiane na wartość 0 lub 255.

2.3. Obrót obrazu

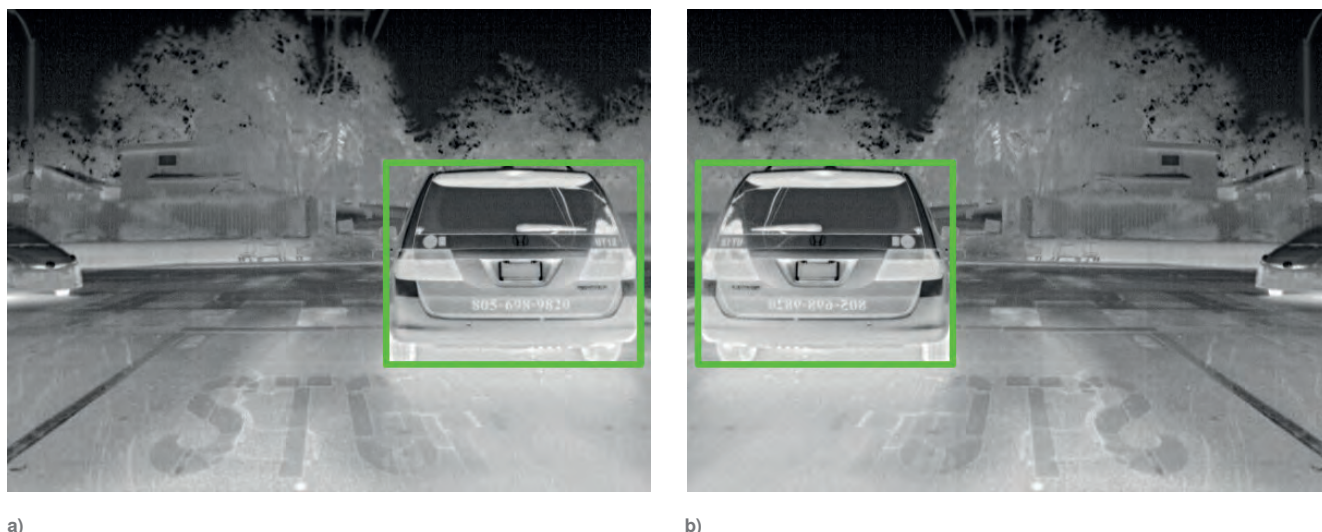
Obrót obrazu jest powszechną techniką augmentacji danych, stosowaną w celu zwiększenia zdolności modelu do rozpoznawania obiektów w różnych orientacjach. W tej technice, obraz źródłowy jest obracany o losowy kąt θ w losowym kierunku. Należy jednak zwrócić uwagę, że zastosowanie obrotu może prowadzić do problemu odciętych narożników obrazu. Ponadto, wymaga to również aktualizacji położenia i wielkości prostokątnego obszaru ograniczającego (ang. *bounding box*) wokół obiektu. Obrót obrazu (rys. 3b) polega na przekształceniu każdego piksela (x, y) do nowej pozycji (x', y') według równań: $x' = x \cos(\theta) - y \sin(\theta)$, $y' = x \sin(\theta) + y \cos(\theta)$, gdzie θ to kąt obrotu.

2.4. Kadrowanie

Kadrowanie obrazu, znane również jako „Random Crop”, to technika augmentacji danych, polegająca na wycinaniu losowego fragmentu z oryginalnego obrazu. Jest to użyteczne w przypadkach, gdy obiekty do rozpoznania przez model są różnej wielkości lub nie są w pełni widoczne na obrazie. Na przykład,



Rys. 4. Przykład techniki kadrowania: a) oryginał; b) po zastosowaniu kadrowania
 Fig. 4. Example of crop operation: a) original; b) after applying crop



Rys. 5. Przykład techniki odbicia lustrzanego: a) oryginał; b) po zastosowaniu odbicia lustrzanego w poziomie
Fig. 5. Example of horizontal flip operation: a) original; b) after applying horizontal flip

przy rozpoznawaniu aut, urządzenie pomiarowe może znajdować się w różnych odległościach od samochodów, co wpływa na ich rozmiar na obrazie. **Kadrowanie** (rys. 4b) polega na wycięciu fragmentu obrazu. Jeśli I jest obrazem, a (x, y, w, h) są współrzędnymi i wymiarami nowego fragmentu, to nowy obraz F jest generowany jako $F = I(x, y, w, h)$.

2.5. Odbicie lustrzane

Odbicie lustrzane to technika augmentacji danych, która polega na odwróceniu obrazu w poziomie lub w pionie. Jest to technika zaskakująco prosta, ale może znacząco poprawić wydajność modelu. Poprzez tworzenie kilku wersji naszych obrazów w różnych orientacjach, dostarczamy naszemu modelowi uczenia głębokiego więcej informacji do nauki, bez konieczności przechodzenia przez czasochłonny proces zbierania i etykietowania dodatkowych danych treningowych. Odbicie lustrzane w poziomie (rys. 5b) polega na odwróceniu kolejności pikseli wzdłuż osi poziomej. Dla obrazu I , nowy obraz F jest generowany jako $F(x, y) = I(W - x, y)$, gdzie W to szerokość obrazu.

2.6. Pochylenie

Pochylenie (ang. *shear*), to technika, która polega na pochyleniu obrazu wzdłuż osi X lub Y. Pomaga ona w symulowaniu zmian perspektywy, co pozwala uodpornić model na różne kąty

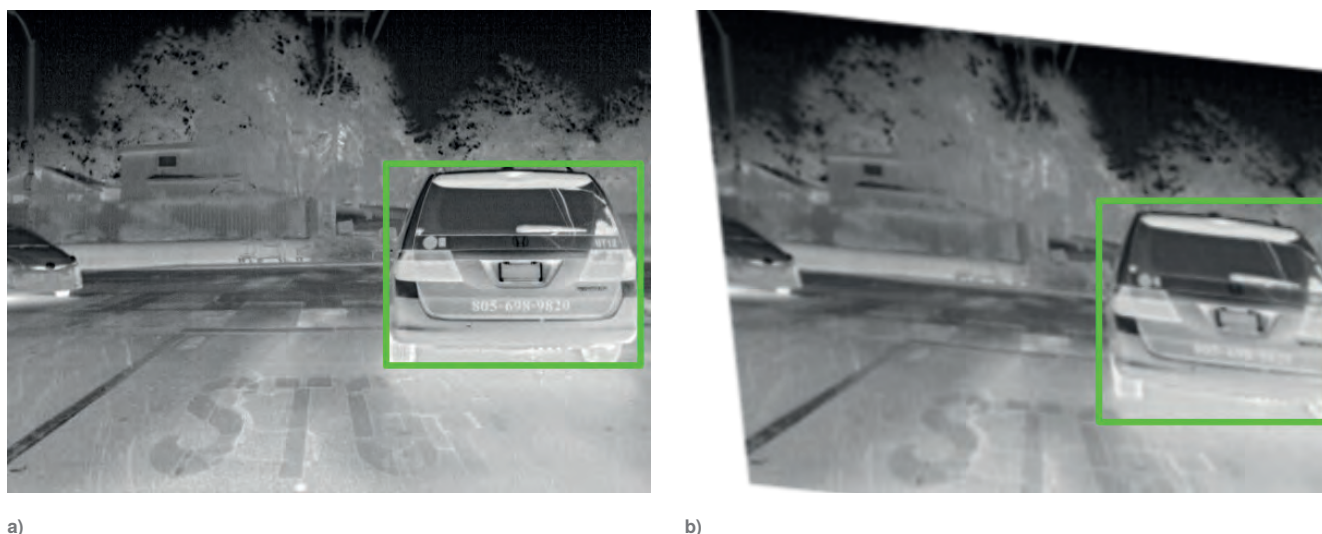
widzenia kamery. **Operacja pochylenia** (rys. 6b) polega na przekształceniu każdego piksela (x, y) do nowej pozycji (x', y') według równań: $x' = x + a * y$, $y' = b * x + y$, gdzie a i b to współczynniki pochylenia w osi X i Y.

2.7. „Cutout”

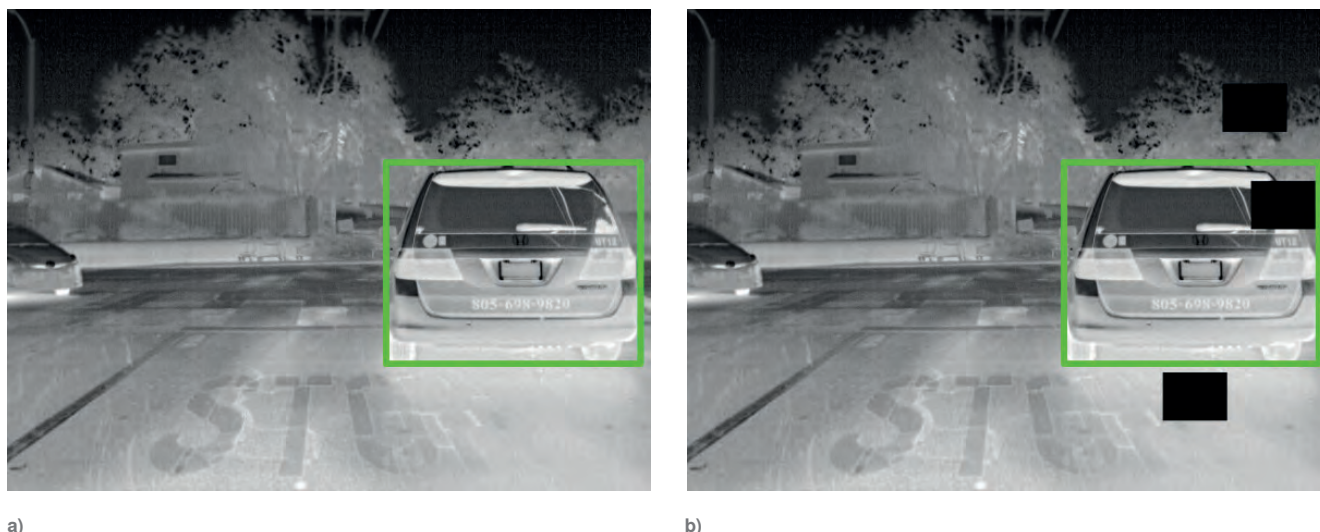
Pojawiają się też nowe, obiecujące metody augmentacji. Jedną z tych metod jest „cutout” [12, 6] polegająca na wycinaniu z obrazu kilku małych (około 10 % wielkości obrazka [12]) prostokątów i zastępowaniu ich kolorem czarnym, tak jak to zostało przedstawione na rys. 7b. Zgodnie z teorią ta operacja pomaga uodpornić model rozpoznawania obiektów na obiekty wychodzące z kadru oraz na częściowe zasłonięcie obiektów, na przykład: samochód częściowo zasłaniający inny samochód na autostradzie. Cutout (rys. 7b) polega na losowym usunięciu prostokątnego fragmentu obrazu. Dla obrazu I , nowy obraz F jest generowany przez ustawienie wartości pikseli w wybranym prostokącie na wartość 0.

2.8. Kombinacja wielu technik augmentacji

W praktyce, jedna metoda augmentacji danych rzadko wystarcza do osiągnięcia optymalnej wydajności modelu. Dlatego też często stosuje się kombinację różnych technik augmentacji, takich jak obrót, pochylenie, zmiana jasności i innych, aby



Rys. 6. Przykład operacji pochylenia: a) oryginał; b) po zastosowaniu pochylenia
Fig. 6. Example of a shear operation: a) original; b) after applying shear



a) b)
Rys. 7. Przykład operacji „cutout”: a) oryginał; b) po zastosowaniu „cutout”
 Fig. 7. Example of a “cutout” operation: a) original; b) after applying cutout

zwiększyć różnorodność i kompleksowość zbioru danych treningowych. Kombinacja różnych metod augmentacji ma kilka zalet. Po pierwsze, zwiększa to zdolność modelu do generalizacji przez ekspozycję na różnorodne warunki. Po drugie, może to również pomóc w zminimalizowaniu ryzyka przeuczenia się modelu na zbyt specyficznych cechach danych. W praktyce, najlepszym podejściem jest eksperymentalne sprawdzenie różnych kombinacji metod augmentacji na zbiorze walidacyjnym.

2.9. Implementacja

Wszystkie omówione powyżej metody zostały wykonane w języku programowania Python (rys. 8). Wykorzystano do tego celu renomowane biblioteki takie jak NumPy, OpenCV oraz Pillow. Wybór Python’a i tych bibliotek wynika z ich szerokiego zastosowania w dziedzinie przetwarzania obrazów i uczenia maszynowego.

3. Realizacja badania

W tym badaniu wykorzystano publicznie dostępny zbiór danych FLIR ADAS, opracowany przez firmę Teledyne FLIR. Zbiór ten zawiera etykietowane obrazy termowizyjne w zakresie podczerwieni LWIR (8,0–12,0 μm) [13, 14], a także obrazy

z zakresu widzialnego. Jest on przeznaczony do projektowania systemów wykrywania obiektów za pomocą konwolucyjnych sieci neuronowych CNN (ang. *Convolutional Neural Nets*) [15]. Został on stworzony z myślą o systemach wspomagania kierowcy oraz autonomicznych pojazdach.

W kontekście tego badania, zdecydowano się na wykorzystanie jedynie 1000 obrazów termowizyjnych z zestawu FLIR ADAS, który zawiera ponad 26 000 obrazów. Ograniczenie to wynika z głównego założenia artykułu. Dzięki tej strategii, możliwe jest dokładniejsze zrozumienie wpływu augmentacji na wyniki uczenia modelu, bez konieczności angażowania zasobów w analizę większego zbioru danych.

Z przygotowanego zestawu obrazów termicznych usunięto:

1. Powtarzające się obrazy, które mogły powstać, na przykład, w wyniku zatrzymania się pojazdu z kamerą pomiarową na światłach drogowych.
2. Większość klas obiektów, pozostawiając tylko klasy: „człowiek” i „samochód”. Pozostałe klasy, takie jak „pociąg” czy „motocykl”, występowały w zbyt małej liczbie, aby można było zapewnić skuteczne i efektywne uczenie modelu.
3. Liczbę obrazów, które nie zawierały żadnego z interesujących nas obiektów. Zgodnie z zaleceniami firmy Ultralytics, twórców modelu YOLOv8, zachowano jedynie około 5 %

```

C:\Windows\System32\cmd.exe
C:\Users\piotr\Desktop\dataset>python3 brightness.py
done!

C:\Users\piotr\Desktop\dataset>python3 noise.py
done!

C:\Users\piotr\Desktop\dataset>python3 rotate.py
done!

C:\Users\piotr\Desktop\dataset>python3 crop.py
done!

C:\Users\piotr\Desktop\dataset>python3 flip.py
done!

C:\Users\piotr\Desktop\dataset>python3 sher.py
done!

C:\Users\piotr\Desktop\dataset>python3 cutout.py
done!

C:\Users\piotr\Desktop\dataset>python3 combination.py
  
```

Rys. 8. Zrzut ekranu konsoli demonstrujący zastosowanie programu Python do augmentacji danych
 Fig. 8. Console screenshot demonstrating the use of Python for data augmentation

obrazów bez obiektów zainteresowania [1]. Celem tej operacji było ograniczenie liczby nieprawidłowo wykrytych obiektów lub obiektów zaliczonych do nieodpowiedniej klasy (wyników fałszywie dodatnich).

Tak zredukowany zestaw danych otrzymał nazwę „FLIR ADAS 1000”. Zgodnie z powszechnie przyjętymi praktykami, zestaw ten podzielono na trzy podzbiory: treningowy, walidacyjny i testowy, zgodnie ze stosunkiem 70:20:10 [1]. Folder treningowy zawiera 1000 obrazów, walidacyjny 285, a testowy 145 obrazów.

Przed rozpoczęciem procesu uczenia maszynowego, dane poddano etapowi wstępnego przetwarzania, zwanemu „preprocesingiem”. Zastosowano następujące operacje:

- auto-orientacja obrazów, do pozycji horyzontalnej;
- dostosowanie wszystkich obrazów do jednolitego wymiaru (640×512) i przeskalowanie tych, które nie spełniały tego kryterium;
- normalizacja obrazów 16-bitowych do obrazów 8-bitowych opierając się na równaniu [16] opracowanym przez firmę FLIR dla użytej w pomiarach kamery [15];
- mieszanie kolejności obrazów, aby zapobiec powtarzalności zbioru;
- zastosowanie palety kolorów typu „tęcza” dla zestawów danych, przy czym dokładniejsze informacje na temat podziału zbioru na dwa zestawy przedstawiono w dalszej części rozdziału.

Następnie przeprowadzono augmentację otrzymanych obrazów, opisaną szczegółowo w rozdziale 2.

Do realizacji uczenia został użyty model sieci neuronowej YOLOv8, który jest powszechnie stosowany w analizie obrazów kolorowych w zakresie światła widzialnego. Jako że obrazy termowizyjne mogą być reprezentowane za pomocą kolorów zamiast tradycyjnej skali szarości wysunięto hipotezę, że YOLOv8 może osiągnąć lepsze wyniki na obrazach kolorowych. Aby zweryfi-

kować tę hipotezę, przeprowadzono testy zarówno na obrazach w skali szarości, jak i na obrazach pokolorowanych. Przykład obrazu termowizyjnego w odcieniach szarości oraz z zastosowaniem wybranej przez nas palety kolorów typu „tęcza” został przedstawiony na rys. 9.

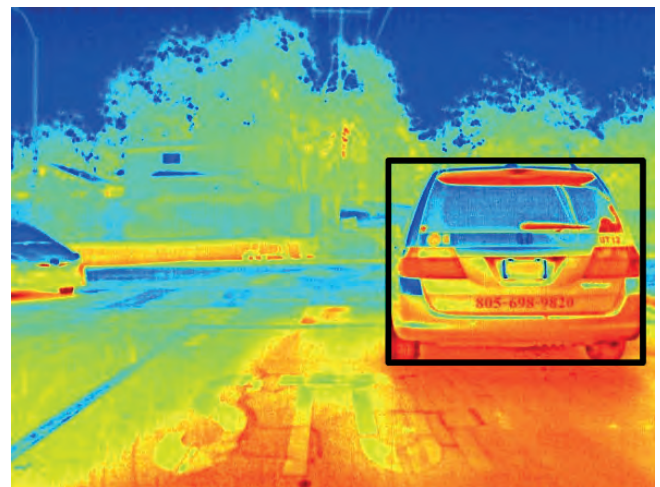
W dalszej części artykułu, konsekwentnie zastosowano następujące terminy:

- **Model B-W** (ang. *Black-White*) – model uczenia maszynowego oparty na architekturze YOLOv8, trenowany za pomocą obrazów w skali szarości.
- **Model Color** – model uczenia maszynowego oparty na architekturze YOLOv8, trenowany za pomocą obrazów z zastosowaną paletą kolorów typu „tęcza”. W celu lepszego zrozumienia całego procesu realizacji badania, na rysunku 10 przygotowano schemat blokowy, który ilustruje wszystkie etapy pracy.

4. Wyniki

Modele wygenerowane w tym badaniu zostały ocenione za pomocą metryki oceny stosowanej w architekturze YOLOv8, oraz na podstawie wygenerowanych macierzy pomyłek (konfuzji). W architekturze YOLOv8 podczas procesu uczenia dąży się do maksymalizacji [17] wartości takich parametrów (metryk) jak:

1. **Precyzja** (ang. *precision*) – jest to miara, która określa, jak wiele z wykrytych obiektów zostało poprawnie zidentyfikowanych. Jest to stosunek prawdziwie dodatnich wyników TP (ang. *True Positive*) do sumy prawdziwie dodatnich i fałszywie dodatnich wyników (FP) [18]. Wysoka precyzja oznacza, że model rzadko myli się przy wykrywaniu obiektów.
2. **Czułość** (ang. *sensitivity, recall*) – jest to miara, która określa, jak wiele z rzeczywistych obiektów zostało wykrytych przez model. Jest to stosunek prawdziwie dodatnich wyników (TP) do sumy prawdziwie dodatnich i fałszywie negatywnych

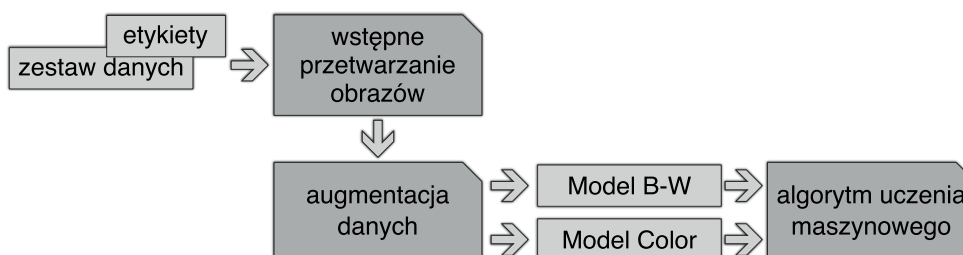


a)

b)

Rys. 9. Obrazek a) oryginalny i b) po zastosowaniu palety kolorów typu „tęcza”

Fig. 9. Image a) original and b) after applying a “rainbow” color map



Rys. 10. Schemat blokowy realizacji badania

Fig. 10. Block diagram of the research implementation

Tabela 1. Wartości metryk modeli dla każdej z metod augmentacji, uczonych obrazami w skali szarości

Table 1. Metric values of models for each augmentation method, trained with grayscale images

rodzaj augmentacji	precyzja	czułość	mAP50	mAP50-95
oryginał	0,834	0,691	0,781	0,451
obrót	0,817	0,694	0,778	0,454
pochylenie	0,822	0,690	0,775	0,440
jasność	0,823	0,689	0,770	0,456
szum	0,804	0,687	0,762	0,444
odbicie	0,828	0,685	0,777	0,453
przycięcie	0,786	0,706	0,768	0,448
cutout	0,823	0,686	0,772	0,456
kombinacja	0,814	0,689	0,764	0,430

Tabela 2. Wartości metryk modeli dla każdej z metod augmentacji, uczonych obrazami w paletce kolorów typu „tęcza”

Table 2. Metric values of models for each augmentation method, trained with rainbow scale images

rodzaj augmentacji	precyzja	czułość	mAP50	mAP50-95
oryginał	0,795	0,649	0,731	0,414
obrót	0,848	0,628	0,740	0,415
pochylenie	0,802	0,623	0,730	0,406
jasność	0,786	0,661	0,732	0,418
szum	0,797	0,622	0,707	0,408
odbicie	0,809	0,648	0,728	0,404
przycięcie	0,802	0,635	0,722	0,408
cutout	0,819	0,624	0,716	0,412
all	0,810	0,629	0,714	0,397

wyników (FN) [18]. Wysoka czułość oznacza, że model jest w stanie wykryć większość obiektów w obrazie.

3. **mAP50** czyli średnia precyzja (ang. *mean Average Precision*), jest jedną z najważniejszych metryk oceny modeli detekcji obiektów [19]. Liczba „50” odnosi się do miary IoU, która reprezentuje stosunek obszaru pokrycia się ramek prawdziwych i wykrytych do sumy ich obszarów [20]. mAP50 jest więc średnią precyzją dla różnych klas obiektów przy IoU wynoszącym 0,5.
4. **mAP50-95** jest średnią precyzją obliczaną dla różnych wartości IoU, od 0,5 do 0,95 z krokiem 0,05. Jest to bardziej rygorystyczna miara, która ocenia, jak dobrze model radzi sobie z dokładnym określeniem lokalizacji obiektów. Jest to standardowa metryka używana w konkursach detekcji obiektów, takich jak COCO (ang. *Common Objects in Context*) [21].

Wyniki badań opisanych wyżej parametrów zostały przedstawione w tabeli 1 dla „Modelu B-W” oraz w tabeli 2 dla „Modelu Color”. Widać na nich, że dla obrazów czarno-białych lepiej zadziałała augmentacja typu „cutout”, a dla obrazów kolorowych augmentacja modyfikacji jasności. Metoda obrotu pomogła zarówno w szkoleniu modeli na obrazach zwykłych i kolorowych.

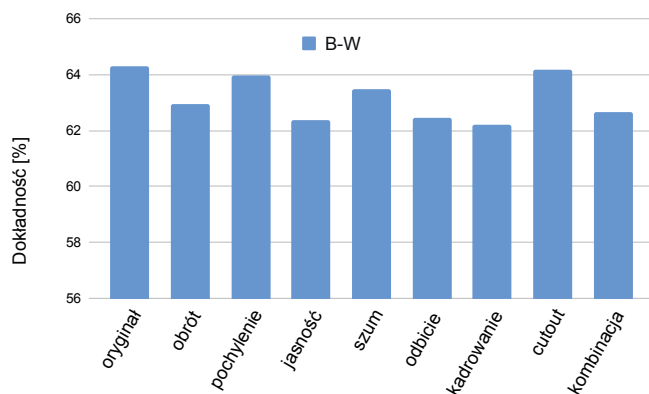
Dokładność ACC (ang. *accuracy*) jest to procent obiektów prawidłowo zidentyfikowanych przez model. Oblicza się go jako

liczba prawidłowych przewidywań dokonanych przez model w stosunku do całkowitej liczby przewidywań [18]. Wysoka dokładność sugeruje, że model jest skuteczny w wykrywaniu obiektów.

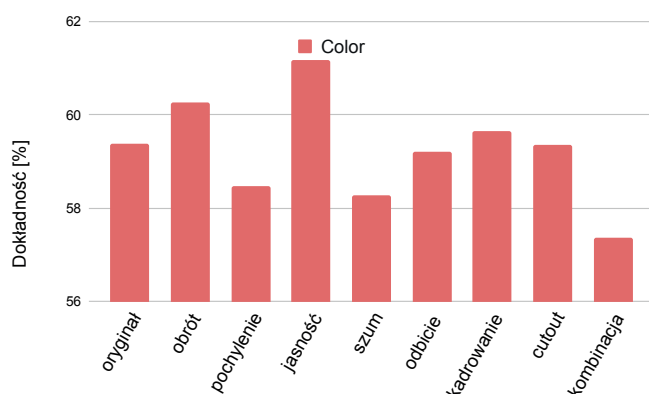
Na wykresie przedstawionym na rysunku 11 i 12 pokazano, że augmentację należy dobierać do rodzaju posiadanych danych. Mimo że szum i metoda pochylenia skutkowała dobrymi dokładnościami dla „Modelu B-W”, to także dawała gorsze dokładności dla „Modelu Color”. Jedynie metoda modyfikacji jasności dawała porównywalne do siebie dokładności.

Zgodnie z rysunkiem 13, metody obrotu i modyfikacji jasności dla „Modelu Color” cechują się wyjątkowo małą liczbą wyników fałszywie dodatnich. Na tym wykresie można zauważyć, że kombinacja wszystkich użytych wcześniej metod augmentacji wprowadza najwięcej wyników fałszywie dodatnich niezależnie od rodzaju danych wejściowych.

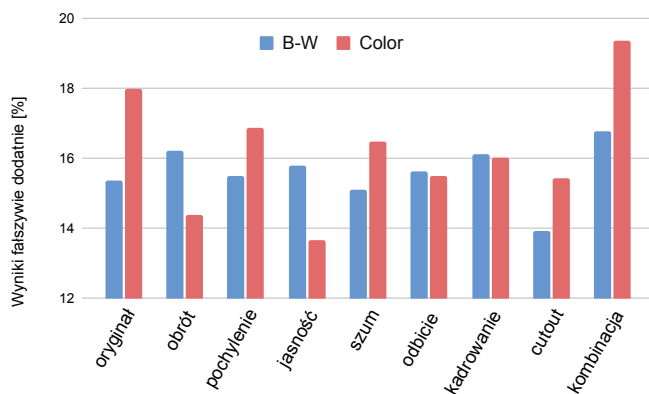
Obrazy na rysunku 14 to niektóre z obrazów testowych, z którymi modele zetknęły się po raz pierwszy. Na rysunku 14(a) oba modele wykryły samochód po środku z prawdopodobieństwami 84 % i 95 % oraz człowieka obok tego samochodu. Jednak „model B-W tego samego człowieka wykrył trzy razy z prawdopodobieństwem poniżej 50 %, kiedy „model Color” prawidłowo wykrył go raz z prawdopodobieństwem 82 %. Na rysunku 14(b) model 1 błędnie zaznacza istnienie samochodów oraz z prawdo-



Rys. 11. Obliczona dokładność dla „Modelu B-W”
Fig. 11. Calculated accuracy for „Model B-W”



Rys. 12. Obliczona dokładność dla „Modelu Color”
Fig. 12. Calculated accuracy for „Model Color”



Rys. 13. Obliczone wyniki fałszywie dodatnie dla każdego z modeli
Fig. 13. Calculated false positives for each of the models

podobieństwami poniżej 50 % ludzi, kiedy model 2 tych samych ludzi zaznacza z prawdopodobieństwami powyżej 80 %. Mimo tego, że jest dużo małych obiektów na rysunku 14(c), oba modele są w stanie wykryć większość z nich. „Model B-W” nie dostrzega ani samochodu, ani człowieka po prawej. A „model Color” oznacza ludzi, których nie ma. Na rysunku 14(d) „model Color” poprawnie wykrywa ludzi, ale błędnie je oznacza, rysując ramki nad lub pod ludźmi. W tym wypadku „model B-W” wypadł lepiej. Oba modele są w stanie wykryć większość obiektów, które są im zaprezentowane. „Model B-W” często zaznacza te same osoby dwa razy, kiedy „Model Color” potrafi pomijać inne osoby. „Model Color” jest pewniejszy swoich przewidywań i ma mniej wyników fałszywie pozytywnych.

Zgodnie z uzyskanymi wynikami, modele wykrywania obiektów wydają się działać lepiej na zbiorach danych w skali szarości. Niemniej jednak, aby ostatecznie stwierdzić, czy kolorowanie

zbiorów danych jest korzystne, konieczne są dalsze badania. Wymaga to przeprowadzenia treningu z większą liczbą epok oraz przetestowania różnych palet kolorów. Tylko wtedy będziemy mogli z pełną pewnością ocenić wpływ kolorowania na skuteczność wykrywania obiektów w termowizji.

5. Wnioski

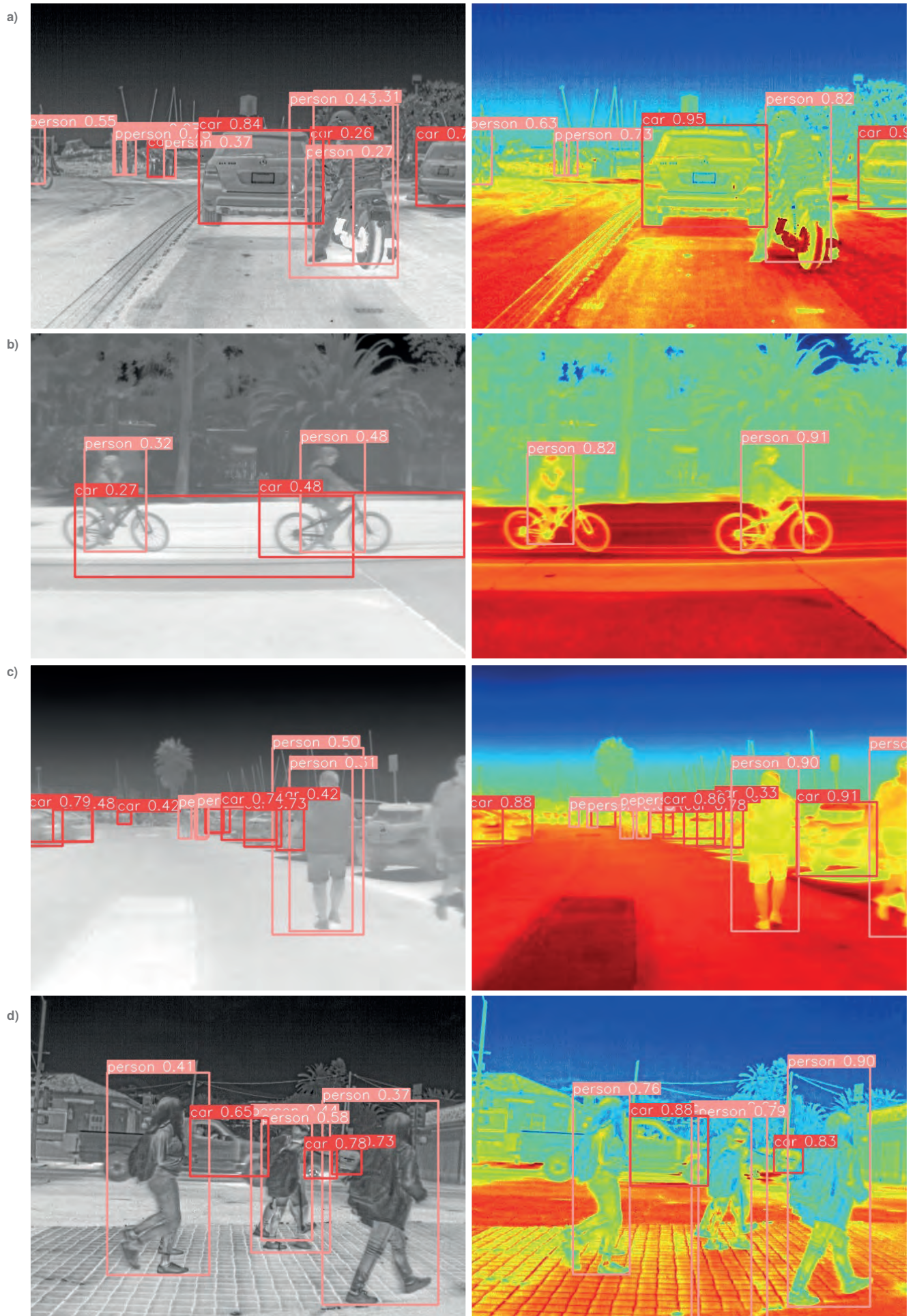
W artykule przedstawiono kompleksową analizę wydajności modeli detekcji obiektów opartych na state-of-the-art architekturze YOLOv8, z uwzględnieniem różnych metod augmentacji danych i palet kolorów. Badania wykazały, że różne metody augmentacji, takie jak obrót, pochylenie i modyfikacja jasności, wpływają na metryki wydajności w różny sposób.

Zauważono również, że modele wykrywania obiektów wydają się działać lepiej na zbiorach danych w skali szarości. Niemniej jednak, aby ostatecznie stwierdzić, czy kolorowanie zbiorów danych jest korzystne, konieczne są dalsze badania. Wymaga to przeprowadzenia treningu z większą liczbą epok oraz przetestowania różnych palet kolorów.

Podsumowując, wyniki badania podkreślają ważność dokładnego doboru technik ulepszania danych i kolorów obrazów, w zależności od konkretnego zastosowania. Wskazują też na potrzebę dalszych badań, które pozwolą na pełniejsze zrozumienie wpływu tych czynników na skuteczność modeli rozpoznawania obiektów.

Bibliografia

1. Ultralytics, *Tips for best training results*, https://docs.ultralytics.com/yolov5/tutorials/tips_for_best_training_results.
2. Papers with Code, *Deep Thermal Imaging Dataset Dataset*, <https://paperswithcode.com/dataset/deep-thermal-imag-ging-dataset>.
3. Ashfaq Q., Akram U., Zafar R., *Thermal Image dataset for object classification*, Mendeley Data, 2021, DOI: 10.17632/btmrycjbj.1.
4. KiKaBeN, *YOLOv5 Transfer Learning*, 2022, <https://kikaben.com/yolov5-transfer-learning-dogs-cats/>.
5. Karimpanal T.G., Bouffanais R., *Self-organizing maps for storage and transfer of knowledge in reinforcement learning*, “Adaptive Behavior”, Vol. 27, No. 2, 2019, 111–126, DOI: 10.1177/1059712318818568.
6. Cubuk E.D., Zoph B., Mane D., Vasudevan V., Le Q.V., *AutoAugment: Learning Augmentation Strategies From Data*, “Computer Science > Computer Vision and Pattern Recognition”, 2019, 113–123, DOI: 10.1109/CVPR.2019.00020.
7. Atienza R., *Improving Model Generalization by Agreement of Learned Representations from Data Augmentation*, “Computer Science > Computer Vision and Pattern Recognition”, 2021, 372–381, DOI: 10.48550/arXiv.2110.10536.
8. Zoph B., Cubuk E.D., Ghiasi G., Lin T.-Y., Shlens J., Le Q.V., *Learning data augmentation strategies for object detection*, “Computer Science > Computer Vision and Pattern Recognition”, 2019, 1–19, DOI: 10.48550/arXiv.1906.11172.
9. Chen Y., Li Y., Kong T., Qi L., Chu R., Li L., Jia J., *Scale-aware Automatic Augmentation for Object Detection*, “Computer Science > Computer Vision and Pattern Recognition”, 2021, 9563–9572, DOI: 10.48550/arXiv.2103.17220.
10. Dodge S., Karam L., *Understanding How Image Quality Affects Deep Neural Networks*, “Computer Science > Computer Vision and Pattern Recognition”, 2016, DOI: 10.48550/arXiv.1604.04004.
11. Goodfellow I.J., Shlens J., Szegedy C., *Explaining and Harnessing Adversarial Examples*, “Statistics > Machine Learning”, 2015, 1–11, DOI: 10.48550/arXiv.1412.6572.



Rys. 14. Wyniki przewidywań modelu uczonego na obrazach czarno-białych (po lewej) oraz na obrazach kolorowych (po prawo)
 Fig. 14. Prediction results of the model trained on grayscale images (on the left) and on colored images (on the right)

12. DeVries T., Taylor G.W., *Improved Regularization of Convolutional Neural Networks with Cutout*, "Computer Science > Computer Vision and Pattern Recognition", 2017, 1–8, DOI: 10.48550/arXiv.1708.04552.
13. Więcek B., De Mey G., *Termowizja w podczerwieni. Podstawy i zastosowania*. Wydawnictwo PAK, Warszawa 2011, ISBN: 978-83-926319-7-2.
14. St-Laurent L., Prévost D., Maldague X., *Thermal imaging for enhanced foreground – background segmentation*, Quantitative InfraRed Thermography Conference, 2006, DOI: 10.21611/qirt.2006.065.
15. Teledyne FLIR, *FREE Teledyne FLIR Thermal Dataset for Algorithm Training*, 2018, <https://www.flir.com/oem/adas/adas-dataset-form/>.
16. Teledyne FLIR, *Basic AGC for radiometric Lepton images*, 2022, <https://www.flir.com/oem/basic-agc-for-radiometric-lepton-images/>.
17. Ultralytics, *Metrics*, <https://docs.ultralytics.com/reference/utlils/metrics/>.
18. Raschka S., *Python: uczenie maszynowe*, Wydawnictwo Helion, 2018, ISBN: 978-83-283-3613-1.
19. Redmon J., Farhadi A., *YOLO9000: Better, Faster, Stronger*, "Computer Science > Computer Vision and Pattern Recognition", 2016, 1–9, DOI: 10.48550/arXiv.1612.08242.
20. Rosebrock A., *Intersection over Union (IoU) for object detection*, 2016, <https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>.
21. Chen X., Fang H., Lin T.-Y., Vedantam R., Gupta S., Dollar P., Zitnick C.L., *Microsoft COCO Captions: Data Collection and Evaluation Server*, "Computer Science > Computer Vision and Pattern Recognition", 2015, 1–7, DOI: 10.48550/arXiv.1504.00325.
22. Terven J., Cordova-Esparza D., *A Comprehensive Review of YOLO: From YOLOv1 and Beyond*, "Computer Science > Computer Vision and Pattern Recognition", 2023, 1–27, DOI: 10.48550/arXiv.2304.00501.
23. Bieszczad G., Sosnowski T., Sawicki K., Gogler S., Ligienza A., Mścichowski M., *A Network of Miniature Thermal Imaging Sensors for Object Detection and Tracking*, "Pomiary Automatyka Robotyka", T. 25, Nr 4, 2021, 57–66, DOI: 10.14313/PAR_242/57.
24. Sosnowski T., Bieszczad G., Madura H., *Image processing in thermal cameras*, [In:] *Advanced Technologies in Practical Applications for National Security*, ser. Studies in Systems, Decision and Control, A. Nawrat, D. Bereska, K. Jędrasiak, red., Cham: Springer International Publishing, 2018, 35–57, DOI: 10.1007/978-3-319-64674-9_3.

Methods of Increasing the Amount of Thermal Imaging Data in Machine Learning

Abstract: Machine learning is increasingly being applied in the processing and analysis of thermal imaging for object recognition and identification. This article presents a study on the impact of data augmentation on the effectiveness of machine learning in the context of thermal image analysis. The publicly available FLIR ADAS dataset, which includes labeled thermal and visible light images, was used for this study. The research focuses on the use of Convolutional Neural Networks, specifically the YOLOv8 architecture, for object detection in thermal images. As part of the study, the FLIR ADAS dataset underwent preprocessing and augmentation, and was then used to train two different models: one based on grayscale images and another using a color palette. The results of the experiment indicate that data augmentation can significantly impact the effectiveness of the model, and the use of colors in thermal images may, in certain situations, further improve detection accuracy.

Keywords: data augmentation, convolutional neural nets, deep learning, infrared imaging, object detection, measurement, computer vision, image processing, object recognition

inż. Piotr Sadzyński

sadzynski.piotr@gmail.com

ORCID: 0009-0008-5056-4716

Student Wojskowej Akademii Technicznej na kierunku „Optoelektronika”, gdzie po ukończeniu studiów inżynierskich kontynuuje edukację na studiach magisterskich. Specjalizuje się w termowizji, podczerwieni oraz uczeniu maszynowym, co stanowi temat jego badań w ramach pracy magisterskiej.

