

**Sieci neuronowe i uczone głębokie**

Porównanie modeli YOLO i RCNN w detekcji obrazów

Autorzy: Wojciech Seweryn, Sebastian Sukiennik

Kierunek: Informatyka i Ekonometria,

Studia niestacjonarne II stopnia

Semestr III

# Wstęp

Celem projektu było porównanie efektywności dwóch modeli detekcji obiektów: YOLOv11 (z fine-tuningiem) oraz R-CNN (uczonego od podstaw). Kluczową metryką porównawczą była średnia precyzja (mAP), mierzona na zbiorze testowym. YOLO wykorzystano w konfiguracji pre-trenowanej na zbiorze COCO, podczas gdy R-CNN implementowano od zera w frameworku PyTorch.

# Zbiór danych

Zbiór danych “Underwater Plastic Pollution Detection” zawiera obrazy i metadane, które służą do trenowania modeli wykrywających plastikowe odpady w środowisku wodnym. Dane zostały pobrane ze strony kaggle:

https://www.kaggle.com/datasets/arnavs19/underwater-plastic-pollution-detection/data

Obrazy zostały uchwycone za pomocą kamer podwodnych, które przedstawiają poziom zanieczyszczenia oceanów. Celem zbioru jest wspieranie rozwiązań opartych na sztucznej inteligencji, umożliwiających automatyczne wykrywanie i klasyfikowanie plastiku w wodzie, co może pomóc w monitorowaniu zanieczyszczeń i działaniach na rzecz ochrony środowiska. W zbiorze znajdują się zdjęcia obiektów następujących klas:

[Mask, can, cellphone, electronics, gbottle, glove, metal, misc, net, pbag, pbottle, plastic, rod, sunglasses, tire]

# Przebieg

Testowanie modeli YOLOv11 oraz Faster R-CNN (opartego na architekturze ResNet-50 z FPN) odbyło się na dedykowanym zbiorze testowym, który nie był wykorzystywany podczas treningu ani walidacji.

Do ewaluacji modeli wykorzystano narzędzia dostarczone przez biblioteki Ultralytics (dla YOLOv11) oraz PyTorch (dla Faster R-CNN). Wartości mean Average Precision (mAP) zostały obliczone przy użyciu funkcji MeanAveragePrecision z modułu torchmetrics.detection.mean\_ap. Metryka mAP, będąca kluczowym wskaźnikiem skuteczności modeli detekcji obiektów, mierzy średnią precyzję dla wszystkich klas, uwzględniając różne progi pewności detekcji. W tym projekcie mAP obliczono dla domyślnego progu IoU (Intersection over Union) wynoszącego 0.5, co oznacza, że detekcja uznawana była za poprawną, jeśli przewidziany bounding box pokrywał się z ground truth w co najmniej 50%.

Testowanie przeprowadzono z wykorzystaniem batch size 16, co pozwoliło na efektywne przetwarzanie danych na dostępnym sprzęcie (GPU NVIDIA RTX 3080). Nie wystąpiły problemy z brakami danych, a zbiór testowy był wystarczająco reprezentatywny, aby ocenić ogólną skuteczność modeli. Dodatkowo, w celu zapewnienia spójności, oba modele przetwarzały obrazy w tej samej rozdzielczości (640x640 pikseli).

Przykładowe wyniki czterech losowo wybranych zdjęc dla modelu YOLO

A collage of images of a diver

AI-generated content may be incorrect.

Oraz confusion matrix

A screenshot of a graph

AI-generated content may be incorrect.

Przykład wyników modelu RCNN:

A collage of different images of plastic bags and fish

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

Wyniki testów wykazały, że YOLOv11 osiągnął wyższą wartość mAP w porównaniu do Faster R-CNN, co sugeruje, że fine-tuning na pre-trenowanej architekturze YOLO lepiej radzi sobie z detekcją obiektów w złożonym środowisku podwodnym. Dodatkowo, YOLOv11 charakteryzował się znacznie wyższą szybkością inferencji, co jest kluczowe w zastosowaniach wymagających przetwarzania w czasie rzeczywistym.

# Wyniki

Wyniki eksperymentów wykazały znaczące różnice w skuteczności i wydajności obu modeli. Faster R-CNN osiągnął wartość mAP wynoszącą 0.579, co wskazuje na stosunkowo wysoką precyzję w detekcji obiektów w środowisku podwodnym. Z kolei YOLOv11 uzyskał znacznie niższą wartość mAP równą 0.121, co sugeruje, że fine-tuning na tym zbiorze danych nie przyniósł oczekiwanych rezultatów. Niski wynik YOLOv11 może wynikać z błędów w procesie adaptacji lub fine-tunowania, takich jak nieodpowiedni dobór hiperparametrów (np. zbyt mała liczba epok, nieoptymalny learning rate) lub problemy z dopasowaniem modelu do specyfiki danych podwodnych. Co ciekawe, pomimo niższej precyzji, YOLOv11 okazał się wolniejszy w fazie inferencji, generując predykcje w czasie 86 sekund, podczas gdy Faster R-CNN poradził sobie z tym zadaniem w zaledwie 42 sekundy. Ta nieoczekiwana różnica w czasie przetwarzania może wynikać z nieoptymalnej konfiguracji YOLOv11 lub specyfiki danych testowych, które mogły wymagać dodatkowej optymalizacji. Wyniki te podkreślają, że wybór modelu zależy nie tylko od metryk jakościowych, takich jak mAP, ale także od wymagań dotyczących szybkości przetwarzania w konkretnym zastosowaniu. W przyszłości warto dokładniej przeanalizować proces fine-tunowania YOLOv11, aby wykluczyć potencjalne błędy i poprawić jego skuteczność.

# Wnioski

Projekt porównujący modele YOLOv11 (z fine-tuningiem) oraz Faster R-CNN (uczony od podstaw) w zadaniu detekcji plastikowych odpadów w środowisku podwodnym dostarczył cennych wniosków zarówno pod względem skuteczności, jak i wydajności modeli. Faster R-CNN okazał się wyraźnie lepszy pod względem precyzji, osiągając mAP na poziomie 0.579, podczas gdy YOLOv11 uzyskał znacznie niższą wartość mAP równą 0.121. Ta różnica może wynikać z błędów w procesie fine-tunowania YOLOv11, takich jak nieodpowiedni dobór hiperparametrów lub problemy z adaptacją modelu do specyfiki danych podwodnych. Dodatkowo, pomimo teoretycznej przewagi YOLOv11 w szybkości inferencji, w tym projekcie Faster R-CNN okazał się szybszy, generując predykcje w 42 sekundy w porównaniu do 86 sekund dla YOLOv11.

Wyniki te sugerują, że Faster R-CNN może być lepszym wyborem w zastosowaniach wymagających wysokiej precyzji, nawet kosztem dłuższego czasu treningu. Z kolei YOLOv11, pomimo niskiej skuteczności w tym eksperymencie, nadal pozostaje obiecującym modelem, pod warunkiem dokładniejszej optymalizacji procesu fine-tunowania.