

Politechnika Wrocławskiego
Wydział Informatyki i Telekomunikacji

Kierunek: **Zaufane Systemy Sztucznej Inteligencji**

**PRACA DYPLOMOWA
MAGISTERSKA**

**Badanie wpływu tła na klasyfikację zwierząt
na obrazach**

Paweł Pelar

Opiekun pracy
dr hab. inż. Henryk Maciejewski

Słowa kluczowe: classification, image segmentation

WROCŁAW 2024

STRESZCZENIE

 Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. test

ABSTRACT

 Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

SPIS TREŚCI

WPROWADZENIE

W ostatnich latach technologia głębokiego uczenia maszynowego zrewolucjonizowała dziedzinę przetwarzania obrazów, w tym klasyfikację i segmentację obrazów. Klasyfikacja obrazów, czyli proces przypisywania etykiet do obiektów przedstawionych na obrazach, jest fundamentalnym zagadnieniem w komputerowym rozpoznawaniu wzorców. Pomimo znaczących postępów, istnieje wiele czynników, które mogą wpływać na dokładność i niezawodność modeli klasyfikacyjnych, a jednym z kluczowych elementów jest tło obrazu.

Tło obrazu może dostarczać zbędnych informacji lub wprowadzać modele w błąd, co może prowadzić do błędnej klasyfikacji obiektów. Między innymi w kontekście klasyfikacji zwierząt, obecność złożonego lub niestandardowego tła może znacząco wpływać na wyniki klasyfikacji. Dlatego analiza wpływu tła na wyniki klasyfikacji obrazów jest niezwykle istotna dla poprawy efektywności modeli.

Segmentacja obrazów, czyli proces podziału obrazu na mniejsze, znaczące fragmenty, jest jednym z podejść umożliwiających radzenie sobie z problemem tła. Dzięki segmentacji możliwe jest wydzielenie obiektu z tła, co może prowadzić do poprawy wyników klasyfikacji. W niniejszej pracy zostaną wykorzystane gotowe modele do segmentacji obrazów, w celu usunięcia tła, w celu przeprowadzenia późniejszych modyfikacji tła.

Wyzwania związane z klasyfikacją obrazów i segmentacją obejmują m.in. różnorodność danych, obecność zakłócających elementów w tle, zmienność oświetlenia oraz różnice w skalach obiektów. Zastosowanie zaawansowanych technik segmentacji i analizy wpływu tła może jednak znacząco poprawić wyniki klasyfikacji.

Niniejsza praca wnosi istotny wkład do dziedziny przetwarzania obrazów, oferując nowe spostrzeżenia na temat wpływu tła na klasyfikację oraz oceniacąc skuteczność różnych modeli głębokiego uczenia w kontekście zmiennych warunków tła. Przeprowadzone badania mają na celu pogłębienie wiedzy na temat optymalizacji modeli klasyfikacyjnych w złożonych i zmiennych środowiskach.

CEL PRACY

Celem niniejszej pracy jest zbadanie wpływu tła na klasyfikację zwierząt na obrazach przy użyciu zaawansowanych modeli głębokiego uczenia, takich jak ResNet i ConvNeXt. Kluczowym aspektem tego badania jest ocena, w jaki sposób usunięcie i modyfikacja tła wpływają na wydajność tych modeli klasyfikacyjnych. Poprzez analizę wyników przed i po modyfikacji tła, praca ta ma na celu:

- **Ocena wrażliwości modeli na tło:** Sprawdzenie, jak różne rodzaje tła wpływają na dokładność klasyfikacji obrazów zwierząt. Analiza ta pozwoli zrozumieć, w jakim stopniu obecność tła zakłóca proces klasyfikacji i jakie rodzaje modyfikacji tła mają największy wpływ na wyniki.
- **Optymalizacja procesu klasyfikacji:** Zidentyfikowanie najlepszych praktyk i metod usuwania oraz modyfikacji tła, które mogą poprawić wydajność modeli klasyfikacyjnych. Badanie to pozwoli określić, które techniki segmentacji i modyfikacji tła są najbardziej efektywne w kontekście różnych modeli klasyfikacyjnych.
- **Porównanie wydajności modeli:** Porównanie jakości klasyfikacji przy użyciu różnych modeli głębokiego uczenia w kontekście zmiennych warunków tła. Analiza ta pozwoli zidentyfikować, który model lepiej radzi sobie z problemem tła i jest bardziej odporny na jego zmiany.
- **Praktyczne implikacje:** Dostarczenie praktycznych wskazówek i rekomendacji dotyczących zastosowania modeli głębokiego uczenia w zadaniach klasyfikacji obrazów w warunkach rzeczywistych, gdzie tło może być zmienne i nieprzewidywalne. Wnioski z tego badania mogą być użyteczne dla badaczy i praktyków zajmujących się rozpoznawaniem obrazów w różnych dziedzinach, takich jak ekologia, bezpieczeństwo czy medycyna.

ZAKRES PRACY

Zakres pracy obejmuje kilka kluczowych aspektów, a mianowicie:

1. **Przygotowanie środowiska badawczego:**
 - Konfiguracja niezbędnego oprogramowania i bibliotek do przetwarzania obrazów oraz uczenia maszynowego.
 - Ustalenie parametrów eksperymentalnych i kryteriów oceny.
2. **Przygotowanie danych:**
 - Zebranie odpowiednich zbiorów danych zawierających obrazy zwierząt z różnorodnym tłem.
 - Przeprowadzenie potrzebnego preprocessingu danych.
3. **Wykorzystanie gotowych modeli klasyfikacyjnych:**
 - Wykorzystanie istniejących, wytrenowanych modeli do klasyfikacji obrazów.
 - Przeprowadzenie wstępnych ocen wydajności modeli na oryginalnych obrazach z różnym tłem.
4. **Segmentacja obrazów:**
 - Wykorzystanie gotowego modelu segmentacyjnego do usunięcia tła z obrazów.
 - Walidacja i ocena uzyskanych masek i obrazów.
5. **Modyfikacja tła obrazów:**

- Zastosowanie różnych technik modyfikacji tła w celu stworzenia zestawów danych z różnymi wariantami tła.
- Analiza wpływu tych modyfikacji na jakość klasyfikacji.

6. Ocena i analiza wyników:

- Porównanie wyników klasyfikacji przed i po modyfikacjach tła za pomocą wybranych metryk.
- Interpretacja wyników oraz wyciągnięcie wniosków dotyczących wpływu tła na wydajność modeli klasyfikacyjnych.

7. Wnioski i rekomendacje:

- Sformułowanie wniosków na podstawie przeprowadzonych eksperymentów.
- Propozycja potencjalnych kierunków dalszych badań oraz zastosowań praktycznych.

PRZEGŁĄD LITERATURY

W ramach niniejszego rozdziału przedstawiony zostanie przegląd literatury dotyczący kluczowych zagadnień związanych z klasyfikacją obrazów, segmentacją obrazów oraz wpływem tła na wyniki klasyfikacji. Celem tego przeglądu jest zrozumienie dotychczasowych badań i rozwiązań, które mogą być istotne dla realizacji niniejszej pracy. Omówione zostaną zarówno klasyczne, jak i nowoczesne podejścia do tych problemów, ze szczególnym uwzględnieniem zaawansowanych modeli głębokiego uczenia, takich jak ResNet i ConvNeXt. Przegląd ten pozwoli na identyfikację luk w istniejącej literaturze oraz wskazanie potencjalnych kierunków dalszych badań.

ZAKRES PRZEGŁĄDU LITERATURY

1. Klasyfikacja obrazów:

- Historia i ewolucja metod klasyfikacji obrazów.
- Przegląd tradycyjnych technik, takich jak SVM i K-NN, oraz ich ograniczeń.
- Wprowadzenie do modeli głębokiego uczenia, w tym sieci neuronowych i konwolucyjnych sieci neuronowych (CNN).

2. Modele głębokiego uczenia:

- Szczegółowy przegląd architektur ResNet i ConvNeXt.
- Analiza wyników i wydajności tych modeli w różnych zadaniach klasyfikacji.
- Porównanie ResNet i ConvNeXt z innymi popularnymi modelami, takimi jak VGG i Inception.

3. Segmentacja obrazów:

- Przegląd technik segmentacji obrazów, w tym tradycyjnych metod oraz podejść opartych na głębokim uczeniu.
- Modele segmentacyjne takie jak U-Net, Mask R-CNN i inne.
- Zastosowania segmentacji obrazów w różnych dziedzinach.

4. Wpływ tła na klasyfikację obrazów:

- Przegląd badań dotyczących wpływu tła na wyniki klasyfikacji obrazów.
- Techniki usuwania i modyfikacji tła oraz ich efektywność.
- Przykłady zastosowań w praktyce i analiza wyników.

5. Metryki oceny jakości modeli:

- Omówienie metryk używanych do oceny jakości modeli klasyfikacyjnych i segmentacyjnych.

- Dokładność, precyza, recall, F1-score i inne miary.

CEL PRZEGŁĄDU LITERATURY

Celem przeglądu literatury jest dostarczenie kompleksowej wiedzy na temat aktualnego stanu badań i technologii w obszarze klasyfikacji i segmentacji obrazów. Przegląd ten pozwoli na:

- Zidentyfikowanie najnowszych osiągnięć i trendów w dziedzinie przetwarzania obrazów.
- Zrozumienie, jakie techniki i modele są obecnie uważane za najbardziej efektywne.
- Wskazanie luk w istniejących badaniach, które mogą stanowić podstawę do dalszych badań.
- Sformułowanie wniosków i rekomendacji na temat optymalnych podejść do rozwiązania problemu wpływu tła na klasyfikację obrazów.

KLASYFIKACJA OBRAZÓW

Klasyfikacja obrazów to jedno z fundamentalnych zadań w dziedzinie przetwarzania obrazów i komputerowego rozpoznawania wzorców. Proces ten polega na przypisaniu każdemu obrazowi jednej lub więcej etykiet z predefiniowanego zbioru klas. Technologia ta znalazła zastosowanie w wielu dziedzinach, takich jak medycyna, bezpieczeństwo, rolnictwo, czy automatyka przemysłowa. W ramach tego przeglądu omówione zostaną tradycyjne metody klasyfikacji obrazów, ewolucja podejść z wykorzystaniem głębokiego uczenia oraz zaawansowane architektury sieci neuronowych.

Tradycyjne metody klasyfikacji obrazów

W początkowych etapach rozwoju klasyfikacji obrazów stosowano głównie techniki oparte na ręcznie wyodrębnianych cechach oraz klasyfikatorach statystycznych. Do najbardziej popularnych metod należały:

- **Support Vector Machines (SVM):** Technika ta polega na znajdowaniu hiperpowierzchni, która najlepiej rozdziela klasy w przestrzeni cech. SVM były szeroko stosowane w klasyfikacji obrazów dzięki swojej skuteczności w radzeniu sobie z nieliniowymi danymi poprzez zastosowanie funkcji jądrowych.
- **K-Nearest Neighbors (K-NN):** Algorytm ten klasyfikuje nowy przykład na podstawie większości głosów najbliższych sąsiadów w przestrzeni cech. Pomimo swojej prostoty, K-NN często wymaga dużych zasobów obliczeniowych i pamięciowych, szczególnie przy dużych zbiorach danych.
- **Metody oparte na histogramach cech:** Techniki takie jak Histogram of Oriented Gradients (HOG) czy Scale-Invariant Feature Transform (SIFT) były używane do

ekstrakcji cech z obrazów, które następnie były klasyfikowane za pomocą modeli takich jak SVM czy K-NN.

Ewolucja podejść z wykorzystaniem głębokiego uczenia

Wraz z rozwojem technologii głębokiego uczenia, tradycyjne metody zaczęły ustępować miejsca konwolucyjnym sieciom neuronowym (CNN), które zrewolucjonizowały klasyfikację obrazów. CNN automatycznie uczą się cech bez potrzeby ręcznego ich wyodrębniania, co pozwala na osiąganie znacznie lepszych wyników.

- **Convolutional Neural Networks (CNN):** CNN składają się z warstw konwolucyjnych, poolingowych i w pełni połączonych, które są trenowane w sposób end-to-end na surowych danych obrazowych. Pionierskie prace takie jak AlexNet, VGG i GoogLeNet zapoczątkowały erę głębokiego uczenia w klasyfikacji obrazów, osiągając znacznie lepsze wyniki niż tradycyjne metody.
- **Residual Networks (ResNet):** Wprowadzenie ResNet w 2015 roku było przełomem w dziedzinie głębokiego uczenia. ResNet wprowadza pojęcie "residual learning" z wykorzystaniem warstw skrótowych (skip connections), co pozwala na trenowanie bardzo głębokich sieci z tysiącami warstw bez problemu zanikania gradientu.
- **Transformers** Chociaż pierwotnie zaprojektowane do przetwarzania języka naturalnego, architektury oparte na transformerach, takie jak Vision Transformer (ViT), zaczęły być stosowane również w klasyfikacji obrazów. Transformery wykorzystują mechanizm uwagi (attention mechanism), co pozwala na modelowanie globalnych zależności w danych.

Zaawansowane architektury sieci neuronowych

Obecnie w klasyfikacji obrazów stosuje się wiele zaawansowanych architektur, które rozwijają i ulepszają wcześniejsze koncepcje:

- **ConvNeXt:** Jest to nowoczesna architektura CNN, która łączy zalety tradycyjnych konwolucyjnych sieci neuronowych z nowymi pomysłami pochodzącyymi od transformerów. ConvNeXt wykorzystuje bardziej złożone operacje konwolucyjne oraz zaawansowane techniki normalizacji, co pozwala na osiąganie znakomitych wyników w różnych zadaniach klasyfikacji.
- **EfficientNet:** EfficientNet wprowadza skalowanie sieci, które jednocześnie zwiększa głębokość, szerokość i rozdzielcość sieci w zrównoważony sposób. Podejście to pozwala na tworzenie modeli, które są bardziej efektywne obliczeniowo i mogą osiągać wyższą dokładność przy mniejszym zużyciu zasobów.

Podsumowanie

Przegląd literatury dotyczącej klasyfikacji obrazów pokazuje, jak ewoluowały metody od tradycyjnych technik opartych na ręcznie wyodrębnianych cechach do zaawansowanych modeli głębokiego uczenia. Nowoczesne architektury, takie jak ResNet i ConvNeXt, oferują znakomite wyniki i są obecnie standardem w wielu zastosowaniach przemysłowych i naukowych. Zrozumienie tych technologii i ich rozwoju jest kluczowe dla dalszych badań i optymalizacji modeli klasyfikacyjnych, zwłaszcza w kontekście analizy wpływu tła na wyniki klasyfikacji obrazów.

MODELE GŁĘBOKIEGO UCZENIA

Modele głębokiego uczenia, zwłaszcza konwolucyjne sieci neuronowe (CNN), zrewolucjonizowały przetwarzanie obrazów, w tym zadania takie jak klasyfikacja, detekcja obiektów, i segmentacja. W ramach tego przeglądu literatury skupimy się na najbardziej wpływowych modelach głębokiego uczenia, w tym na ich architekturach, kluczowych innowacjach oraz wynikach osiągniętych na różnych benchmarkach.

Convolutional Neural Networks (CNN)

CNN są fundamentem nowoczesnego przetwarzania obrazów. Ich struktura składa się z warstw konwolucyjnych, poolingowych i w pełni połączonych, które są trenowane w sposób end-to-end. AlexNet, zaprojektowany przez Krizhevsky'ego, Sutskevera i Hinton'a w 2012 roku, był pierwszym modelem, który pokazał ogromny potencjał głębokiego uczenia w zadaniach klasyfikacji obrazów. Wprowadzenie dużych filtrów konwolucyjnych, warstw max-pooling oraz technik regularizacji takich jak dropout przyczyniło się do znacznego zmniejszenia błędu klasyfikacji na konkursie ImageNet. Kolejnym krokiem w rozwoju CNN było VGGNet, opracowane przez Simonyana i Zissermana w 2014 roku. VGGNet zyskało popularność dzięki swojej prostocie i skuteczności, opierając się na małych filtrach konwolucyjnych (3x3) oraz głębokiej architekturze składającej się z wielu warstw konwolucyjnych. Model ten udowodnił, że zwiększenie głębokości sieci może prowadzić do lepszej wydajności.

GoogLeNet, zaprezentowany przez zespół Google w 2014 roku, wprowadził koncepcję "Inception modules", które umożliwiają efektywne równoległe przetwarzanie danych. Inception modules łączą różne wielkości filtrów konwolucyjnych w jednej warstwie, co pozwala na lepsze uchwycenie różnorodnych cech obrazu. Dodatkowo, zamiast tradycyjnych w pełni połączonych warstw na końcu sieci, GoogLeNet używa warstw global average pooling, co zmniejsza liczbę parametrów i zwiększa efektywność modelu. Model ten osiągnął świetne wyniki na ImageNet, redukując liczbę parametrów w porównaniu do wcześniejszych architektur.

Residual Networks (ResNet)

W 2015 roku He et al. wprowadzili Residual Networks (ResNet), które zrewolucjonizowały głębokie uczenie. Kluczową innowacją w ResNet było wprowadzenie warstw skrótowych (skip connections), które umożliwiają trenowanie bardzo głębokich sieci poprzez rozwiązywanie problemu zanikania gradientu. Dzięki temu podejściu możliwe stało się trenowanie sieci o głębokości nawet 152 warstw. ResNet osiągnął rewolucyjne wyniki na konkursie ImageNet, pokazując, że głębsze sieci mogą prowadzić do znacznie lepszej wydajności niż wcześniejsze architektury.

Transformery w przetwarzaniu obrazów

Chociaż początkowo zaprojektowane do przetwarzania języka naturalnego, architektury oparte na transformerach znalazły zastosowanie również w przetwarzaniu obrazów. Vision Transformer (ViT), zaprezentowany przez Dosovitskiy'ego et al. w 2020 roku, adaptuje mechanizm uwagi (attention mechanism) do zadań przetwarzania obrazów. ViT dzieli obraz na mniejsze pliki (patches) i traktuje je jako tokeny w modelu transformera, co pozwala na globalne modelowanie zależności w danych. Model ten osiągnął konkurencyjne wyniki na benchmarkach takich jak ImageNet, udowadniając, że podejście oparte na transformerach może być równie skuteczne jak tradycyjne CNN.

Nowoczesne architektury

Wśród nowoczesnych architektur CNN, ConvNeXt wyróżnia się jako model łączący zalety tradycyjnych konwolucyjnych sieci neuronowych z nowymi pomysłami pochodząymi od transformerów. ConvNeXt wykorzystuje bardziej złożone operacje konwolucyjne oraz zaawansowane techniki normalizacji, co pozwala na osiąganie znakomitych wyników w różnych zadaniach klasyfikacji. Model ten potwierdza, że nowoczesne CNN mogą konkurować z modelami opartymi na transformerach.

EfficientNet, opracowany przez Tan i Le, wprowadza koncepcję skalowania sieci, która pozwala na jednoczesne zwiększanie głębokości, szerokości i rozdzielczości sieci w zrównoważony sposób. Dzięki podejściu zrównoważonego skalowania (compound scaling), EfficientNet tworzy modele, które są bardziej efektywne obliczeniowo i mogą osiągać wyższą dokładność przy mniejszym zużyciu zasobów. Model ten jest jednym z najbardziej efektywnych modeli głębokiego uczenia, osiągając wyższą dokładność przy optymalnym wykorzystaniu zasobów.

Podsumowanie

Przegląd literatury dotyczącej modeli głębokiego uczenia w przetwarzaniu obrazów ukazuje dynamiczny rozwój tej dziedziny. Od wczesnych architektur CNN, takich jak AlexNet i VGG, przez innowacyjne podejścia ResNet i GoogLeNet, aż po nowoczesne modele

ConvNeXt i transformery, takie jak ViT, ewolucja tych technologii znacząco poprawiła wyniki klasyfikacji obrazów. Zrozumienie tych innowacji i ich zastosowań jest kluczowe dla dalszych badań i optymalizacji modeli klasyfikacyjnych w różnych kontekstach, w tym w analizie wpływu tła na wyniki klasyfikacji obrazów.

SEGMENTACJA OBRAZÓW

Segmentacja obrazów to kluczowy proces w dziedzinie przetwarzania obrazów, polegający na podziale obrazu na znaczące fragmenty, które mogą reprezentować różne obiekty lub regiony. Techniki segmentacji są szeroko stosowane w wielu dziedzinach, takich jak medycyna, robotyka, analiza wideo i rozpoznawanie obiektów. W ramach tego przeglądu literatury omówione zostaną tradycyjne metody segmentacji, nowoczesne podejścia wykorzystujące głębokie uczenie oraz ich zastosowania w różnych kontekstach.

Tradycyjne metody segmentacji obrazów

W początkowych etapach rozwoju segmentacji obrazów stosowano głównie metody oparte na analizie cech niskiego poziomu, takich jak kolor, tekstura i krawędzie. Do najpopularniejszych technik należały:

1. **Segmentacja oparta na progach:** Technika ta polega na podziale obrazu na regiony na podstawie wartości pikseli. Progi mogą być ustalane globalnie dla całego obrazu lub lokalnie dla poszczególnych regionów. Chociaż metoda ta jest prosta, jej skuteczność zależy od odpowiedniego doboru progów i jest ograniczona w przypadkach obrazów z złożonymi teksturami i oświetleniem.
2. **Segmentacja przez regiony:** Techniki takie jak algorytm wododziałowy (watershed algorithm) oraz metoda region growing polegają na grupowaniu sąsiadujących pikseli o podobnych wartościach. Algorytm wododziałowy modeluje obraz jako topograficzną mapę, gdzie linie wododziałowe oddzielają różne segmenty. Metoda region growing natomiast zaczyna od zestawu początkowych pikseli (seed points) i iteracyjnie dodaje sąsiednie piksele spełniające kryterium podobieństwa.
3. **Segmentacja oparta na krawędziach:** Metody te wykorzystują detekcję krawędzi do identyfikacji granic między różnymi obiektami w obrazie. Algorytmy takie jak Canny edge detector i Sobel operator są powszechnie stosowane do wykrywania krawędzi, które następnie służą do segmentacji obrazu.

Nowoczesne podejścia wykorzystujące głębokie uczenie

Rozwój głębokiego uczenia wprowadził znaczące innowacje w dziedzinie segmentacji obrazów, zwłaszcza dzięki zastosowaniu konwolucyjnych sieci neuronowych (CNN). Modele te automatycznie uczą się reprezentacji cech z obrazów, co prowadzi do znacznie lepszej dokładności segmentacji w porównaniu do tradycyjnych metod.

1. **Fully Convolutional Networks (FCN):** Wprowadzone przez Longa et al. w 2015 roku, FCN przekształcają tradycyjne CNN, zastępując w pełni połączone warstwy konwolucyjnymi, co pozwala na generowanie map segmentacji o tej samej rozdzielcości co wejściowy obraz. FCN były pierwszym krokiem w kierunku end-to-end segmentacji obrazów.
2. **U-Net:** Zaproponowany przez Ronnebergera et al. w 2015 roku, U-Net stał się standardem w dziedzinie segmentacji medycznych obrazów. Architektura U-Net składa się z symetrycznej struktury, która łączy warstwy składające się z konwolucji i upsamplingu, co umożliwia precyzyjne segmentowanie obiektów. U-Net wyróżnia się również dzięki połączeniom między warstwami, które przekazują szczegółowe informacje z warstw niskiego poziomu do warstw wyższego poziomu, poprawiając dokładność segmentacji.
3. **Mask R-CNN:** Rozwinięcie Faster R-CNN, Mask R-CNN, zaproponowane przez He et al. w 2017 roku, rozszerza funkcjonalność detekcji obiektów o możliwość segmentacji. Model ten dodaje gałąź segmentacyjną do istniejącej architektury detekcji obiektów, umożliwiając precyzyjne maskowanie wykrytych obiektów. Mask R-CNN osiągnął znakomite wyniki w wielu zadaniach segmentacji i detekcji obiektów.
4. **DeepLab:** Rodzina modeli DeepLab, opracowana przez zespół Google, wykorzystuje różne techniki do poprawy segmentacji, takie jak atrous convolutions (dylatowane konwolucje) i Conditional Random Fields (CRFs). DeepLabv3+, najnowsza wersja tej serii, łączy atrous convolutions z modelem spatial pyramid pooling, co pozwala na uchwycenie kontekstowych informacji na różnych skalach.

Zastosowania segmentacji obrazów

Techniki segmentacji obrazów znalazły szerokie zastosowanie w różnych dziedzinach. W medycynie segmentacja obrazów jest kluczowa w diagnostyce i planowaniu leczenia, pozwalając na precyzyjne wyodrębnienie struktur anatomicznych i patologicznych z obrazów MRI i CT. W robotyce segmentacja pomaga w nawigacji i manipulacji obiektemi, umożliwiając robotom zrozumienie i interakcję z otoczeniem. W analizie wideo segmentacja jest używana do śledzenia obiektów i rozpoznawania scen, co ma zastosowanie w monitoringu i automatycznym nadzorze.

Podsumowanie

Przegląd literatury dotyczącej segmentacji obrazów ukazuje, jak ewoluowały techniki od tradycyjnych metod opartych na analizie cech niskiego poziomu do zaawansowanych podejść wykorzystujących głębokie uczenie. Nowoczesne architektury, takie jak FCN, U-Net, Mask R-CNN i DeepLab, oferują znakomite wyniki i są szeroko stosowane w różnych dziedzinach. Zrozumienie tych technik i ich zastosowań jest kluczowe dla dalszych badań i optymalizacji procesów segmentacji, zwłaszcza w kontekście analizy wpływu tła na wyniki klasyfikacji obrazów.

WPŁYW TŁA NA KLASYFIKACJĘ OBRAZÓW

Badania nad wpływem tła na klasyfikację obrazów konwolucyjnymi sieciami neuronowymi (CNN) wykazały, że tło może znacząco wpływać na skuteczność i proces uczenia tych modeli. Rajnoha i współpracownicy (2018) przeprowadzili badania nad klasyfikacją binarną osób, gdzie pokazali, że usunięcie zbędnego tła z obrazów może znacząco poprawić proces uczenia sieci neuronowych, szczególnie w przypadkach ograniczonej liczby próbek treningowych. Eksperymenty wykazały, że sieci trenowane na obrazach bez tła były w stanie szybciej rozpocząć proces konwergencji, podczas gdy sieci trenowane na pełnych obrazach miały z tym problemy, szczególnie gdy tło stanowiło ponad 50% obrazu. Wyniki te sugerują, że usunięcie tła może znacznie zwiększyć efektywność procesu uczenia w przypadkach, gdy stosunek sygnału do szumu jest niski.

Z kolei Sehwag i in. (2020) analizowali 32 różne architektury sieci neuronowych, od małych sieci do dużych modeli trenowanych na miliardzie obrazów, aby zbadać wpływ cech tła na dokładność klasyfikacji. Badania te wykazały, że wraz ze wzrostem mocy obliczeniowej sieci, tendencja do wykorzystywania informacji z tła również wzrasta. W eksperymentach, w których maskowano treść pierwszoplanową i pozostawiano jedynie tło, sieci nadal były w stanie dokonywać poprawnych predykcji w wielu przypadkach. Ponadto, zmiana tła na jednorodne lub teksturalne prowadziła do znacznego spadku dokładności, co podkreśla, że obecne sieci neuronowe silnie polegają na informacjach z tła do dokonywania klasyfikacji.

W badaniach dotyczących klasyfikacji obrazów liści roślin, wykorzystano kombinację segmentacji krawędziowej, morfologicznej oraz odejmowania tła, co pozwoliło na poprawę dokładności klasyfikacji w przypadku zdjęć z niejednorodnym tłem. W eksperymentach zastosowano sieci DenseNet121, InceptionV3 i inne, osiągając dokładność do 98.7% na czystych zbiorach danych. Segmentacja pomogła w izolacji liści w pierwszym planie, usuwając niepożądane elementy takie jak inne części roślin, gleba czy części ciała ludzi, co znacznie poprawiło precyzję klasyfikacji.

Dalsze badania, takie jak te przeprowadzone przez Zhou i in. (2021), koncentrowały się na zrozumieniu, w jakim stopniu obecne sieci neuronowe wykorzystują informacje z tła. Autorzy sugerują, że obecne funkcje strat, takie jak funkcja entropii krzyżowej, nie zachęcają do inwariancji względem tła, co powoduje, że sieci te wykorzystują wszelkie istniejące korelacje między tłem a predykcjami wyjściowymi. Badania wykazały, że zwiększenie różnorodności tła w zbiorze danych treningowych może zwiększyć inwariancję tła sieci, jednak bardziej efektywnym podejściem może być poprawa funkcji strat, aby karać za korelacje z tłem.

Podobnie, w badaniach nad klasyfikacją obrazów w rolnictwie, Kamal i in. (2023) pokazali, że wykorzystanie segmentacji i odejmowania tła może znacząco poprawić dokładność klasyfikacji obrazów roślin. W eksperymentach użyto technik takich jak segmentacja

krawędziowa, morfologiczna i odejmowanie tła, co pozwoliło na izolację liści roślin w pierwszym planie. Zastosowanie tych technik w połączeniu z sieciami neuronowymi, takimi jak DenseNet121 i InceptionV3, pozwoliło na osiągnięcie bardzo wysokiej dokładności klasyfikacji nawet w przypadkach, gdy obrazy były zrobione w niejednorodnym tle.

Wszystkie te badania podkreślają znaczenie manipulacji tłem w procesie trenowania i klasyfikacji obrazów przy użyciu sieci neuronowych. Usunięcie lub manipulacja tłem może nie tylko poprawić dokładność, ale również pomóc sieciom w lepszym zrozumieniu i generalizacji cech istotnych dla danego zadania. W przyszłości badania te mogą prowadzić do opracowania bardziej zaawansowanych technik segmentacji i manipulacji tłem, które będą kluczowe dla dalszego rozwoju i optymalizacji sieci neuronowych w różnych dziedzinach zastosowań, takich jak rolnictwo, medycyna, czy systemy monitoringu.

Wyzwania i przyszłe kierunki badań

Mimo znaczących postępów w zakresie usuwania i modyfikacji tła, istnieje wiele wyzwań, które wciąż wymagają dalszych badań. Jednym z głównych problemów jest radzenie sobie z dynamicznymi i zmiennymi warunkami tła, takimi jak zmiany oświetlenia, ruch obiektów i różnorodność scen. Ponadto, badania nad wpływem tła na klasyfikację obrazów mogą prowadzić do opracowania bardziej odpornych modeli, które lepiej radzą sobie z zakłóceniami tła.

Przyszłe badania mogą również koncentrować się na integracji technik usuwania i modyfikacji tła z innymi metodami przetwarzania obrazów, takimi jak detekcja obiektów i analiza scen. Opracowanie bardziej zaawansowanych algorytmów, które będą w stanie lepiej modelować złożone sceny i dynamiczne tła, może przyczynić się do dalszej poprawy wyników klasyfikacji obrazów.

Podsumowanie

Przegląd literatury dotyczącej wpływu tła na klasyfikację obrazów ukazuje, jak istotny jest to aspekt w dziedzinie przetwarzania obrazów i głębokiego uczenia. Techniki usuwania i modyfikacji tła mogą znacząco poprawić dokładność klasyfikacji, jednak nadal istnieje wiele wyzwań, które wymagają dalszych badań. Zrozumienie wpływu tła na wyniki klasyfikacji oraz opracowanie skutecznych metod radzenia sobie z tym problemem jest kluczowe dla rozwoju bardziej niezawodnych i precyzyjnych systemów klasyfikacyjnych.

METRYKI OCENY JAKOŚCI MODELI

Ocena jakości modeli klasyfikacyjnych i segmentacyjnych jest kluczowym elementem każdego badania w dziedzinie przetwarzania obrazów i głębokiego uczenia. Wybór odpowiednich metryk pozwala na obiektywne porównanie różnych modeli oraz identyfikację ich mocnych i słabych stron. W ramach tego przeglądu literatury omówione zostaną naj-

ważniejsze metryki stosowane do oceny jakości modeli, w tym dokładność, precyza, recall, F1-score oraz inne zaawansowane miary.

Dokładność (Accuracy)

Dokładność jest jedną z najbardziej intuicyjnych metryk stosowanych do oceny modeli klasyfikacyjnych. Jest to stosunek liczby poprawnie sklasyfikowanych przykładów do całkowej liczby przykładów. Chociaż dokładność jest łatwa do zrozumienia i szeroko stosowana, może być myląca w przypadku niezrównoważonych zbiorów danych, gdzie liczba przykładów jednej klasy znacznie przewyższa liczbę przykładów innych klas. W takich sytuacjach dokładność może być wysoka, nawet jeśli model nie radzi sobie dobrze z rzadkimi klasami.

Precyza (Precision)

Precyza, znana również jako dodatnia wartość predykcyjna, to stosunek liczby prawdziwie pozytywnych przykładów do liczby wszystkich przykładów sklasyfikowanych jako pozytywne. W kontekście klasyfikacji binarnej precyza mierzy, jak wiele z przykładów sklasyfikowanych jako pozytywne faktycznie należy do klasy pozytywnej. Wysoka precyza oznacza, że model rzadko klasyfikuje negatywne przykłady jako pozytywne, co jest szczególnie ważne w aplikacjach, gdzie fałszywe alarmy są kosztowne lub niepożądane.

Czułość (Recall)

Recall, znany również jako czułość lub true positive rate, to stosunek liczby prawdziwie pozytywnych przykładów do liczby wszystkich rzeczywistych pozytywnych przykładów. Recall mierzy zdolność modelu do wykrywania wszystkich pozytywnych przykładów w zbiorze danych. Wysoki recall oznacza, że model rzadko przeocza pozytywne przykłady, co jest ważne w aplikacjach, gdzie wykrycie wszystkich pozytywnych przypadków jest kluczowe, na przykład w diagnostyce medycznej.

F1-Score

F1-score to harmoniczna średnia precyzji i recall, która stanowi kompromis między tymi dwiema miarami. Jest szczególnie użyteczna w przypadkach, gdy istotne jest jednoczesne zminimalizowanie liczby fałszywie pozytywnych i fałszywie negatywnych klasyfikacji. F1-score jest bardziej informatywny niż dokładność w kontekście niezrównoważonych zbiorów danych, ponieważ uwzględnia zarówno precyzę, jak i recall.

Inne zaawansowane miary

Oprócz podstawowych metryk, istnieje wiele zaawansowanych miar stosowanych do oceny jakości modeli, w tym:

- **ROC AUC (Area Under the Receiver Operating Characteristic Curve):** ROC AUC jest miarą, która ocenia zdolność modelu do rozróżniania między klasami na podstawie analizy krzywej ROC. Wartość AUC bliska 1 oznacza, że model ma doskonałą zdolność rozróżniania między pozytywnymi a negatywnymi przykładami.
- **AP (Average Precision):** Średnia precyzja to miara, która ocenia średnią precyzję przy różnych wartościach recall. Jest często stosowana w zadaniach detekcji obiektów i segmentacji, gdzie istotne jest ocenienie jakości predykcji na różnych poziomach czułości.
- **IoU (Intersection over Union):** IoU jest miarą stosowaną w segmentacji obrazów, która mierzy stosunek pola wspólnego (intersection) między przewidywaną maską segmentacyjną a rzeczywistą maską do pola sumy (union) tych masek. Wysoki IoU oznacza, że przewidywana maska dobrze pokrywa się z rzeczywistą maską obiektu.
- **Dice Coefficient:** Współczynnik Dice jest kolejną miarą stosowaną w segmentacji obrazów, która jest podobna do IoU, ale bardziej skoncentrowana na średniej harmonicznej obszarów przewidywanego i rzeczywistego obiektu. Jest szczególnie użyteczny w medycznej segmentacji obrazów.

Zastosowanie metryk w praktyce

W praktyce wybór odpowiednich metryk zależy od specyfiki zadania i rodzaju danych. Na przykład, w diagnostyce medycznej ważne jest używanie recall i F1-score, aby zapewnić, że wszystkie przypadki choroby są wykrywane, a liczba fałszywie negatywnych wyników jest minimalna. W systemach monitoringu i detekcji obiektów, metryki takie jak AP i IoU są kluczowe do oceny precyzji i dokładności lokalizacji obiektów.

Podsumowanie

Przegląd literatury dotyczącej metryk oceny jakości modeli podkreśla znaczenie wyboru odpowiednich miar w kontekście specyficznych zastosowań. Dokładność, precyzja, recall i F1-score są podstawowymi metrykami stosowanymi do oceny modeli klasyfikacyjnych, natomiast bardziej zaawansowane miary, takie jak ROC AUC, AP, IoU i Dice Coefficient, są kluczowe w specyficznych zadaniach, takich jak detekcja obiektów i segmentacja obrazów. Zrozumienie tych metryk i ich zastosowań jest kluczowe dla obiektywnej oceny i porównania różnych modeli, a także dla dalszych badań nad optymalizacją algorytmów przetwarzania obrazów.

METODYKA BADAŃ

WPROWADZENIE DO METODYKI BADAŃ

Niniejszy rozdział poświęcony jest metodyce badań, mającej na celu zbadanie wpływu tła na klasyfikację obrazów zwierząt przy użyciu zaawansowanych modeli głębokiego uczenia, takich jak ResNet i ConvNeXt. Badania te koncentrują się na analizie wyników klasyfikacji przed i po modyfikacjach tła z zastosowaniem różnych metryk oceny jakości, co pozwoli na zrozumienie, w jakim stopniu tło wpływa na wydajność modeli klasyfikacyjnych oraz jakie techniki mogą być stosowane do minimalizacji negatywnego wpływu tła. W pierwszej części rozdziału zostaną omówione narzędzia i oprogramowanie użyte do badań, konfiguracja sprzętowa i programowa, a także biblioteki i frameworki niezbędne do realizacji eksperymentów. Następnie przedstawione zostaną wybrane modele klasyfikacyjne, ResNet i ConvNeXt, wraz z uzasadnieniem ich wyboru oraz krótkim opisem ich architektur i specyfikacji. Kolejna sekcja skupi się na metrykach oceny jakości, z wyjaśnieniem, dlaczego właśnie te miary zostały wybrane oraz jak będą interpretowane wyniki. Opisany zostanie również zbiór danych wykorzystany w badaniach, jego źródło, struktura, etykiety oraz sposób przygotowania i przetwarzania danych przed użyciem w modelach. Kluczowym elementem rozdziału będzie szczegółowy plan przeprowadzenia badań, obejmujący wszystkie etapy, od segmentacji obrazów i usunięcia tła, poprzez modyfikację tła, aż po ocenę wyników modeli przed i po modyfikacjach. Dodatkowo, omówione zostaną technikalia implementacji, w tym jak poszczególne etapy zostały zaimplementowane w kodzie, jakie narzędzia programistyczne i techniki kodowania zostały użyte oraz jak zapewniono replikowalność wyników eksperymentów. Całość zakończy krótkie podsumowanie metodyki badań, podkreślające główne kroki i decyzje podjęte w celu realizacji badań, co umożliwia systematyczną analizę wpływu tła na wyniki klasyfikacji obrazów oraz identyfikację najlepszych praktyk i metod poprawiających wydajność modeli klasyfikacyjnych.

PRZYGOTOWANIE ŚRODOWISKA

Przygotowanie odpowiedniego środowiska badawczego jest kluczowym krokiem w realizacji każdego projektu opartego na analizie danych i głębokim uczeniu. W niniejszych badaniach, całość prac została przeprowadzona w języku Python, który jest powszechnie

stosowany w dziedzinie przetwarzania obrazów i uczenia maszynowego dzięki bogatemu ekosystemowi bibliotek i narzędzi wspomagających te procesy.

Do realizacji projektu użyto następujących bibliotek: numpy, pandas, scikit-learn, PIL, matplotlib, seaborn oraz torch. Biblioteka numpy została wykorzystana do obsługi operacji numerycznych i manipulacji tablicami, biblioteka pandas służyła do manipulacji i analizy danych strukturalnych, takich jak tablice. Scikit-learn był wykorzystywany przy obliczeniach metryk i ocenie jakości modeli, a PIL (Python Imaging Library) umożliwiła manipulację obrazami. Biblioteki matplotlib i seaborn posłużyły do wizualizacji danych i wyników analiz, co pozwoliło na lepsze zrozumienie uzyskanych rezultatów oraz prezentację wyników w formie graficznej.

Kluczowym elementem projektu były zaawansowane modele głębokiego uczenia: ResNet, ConvNeXt oraz DeepLabv3. Modele te zostały zainportowane z biblioteki torchvision, która jest częścią ekosystemu PyTorch. Torchvision dostarcza łatwy dostępu do najnowocześniejszych modeli pretrenowanych na dużych zbiorach danych, co umożliwia efektywne przeprowadzanie eksperymentów bez konieczności trenowania modeli od podstaw.

Dodatkowo, do analizowania wyników i prowadzenia interaktywnej pracy z kodem, używany był Jupyter Notebook. Jupyter Notebook jest wszechstronnym narzędziem, które umożliwia tworzenie i udostępnianie dokumentów zawierających kod, równania, wizualizacje oraz tekst. Jego zastosowanie pozwoliło na przejrzyste prezentowanie procesu badawczego, testowanie i modyfikowanie kodu w czasie rzeczywistym oraz dokumentowanie każdego kroku analizy.

Całe środowisko badawcze zostało skonfigurowane na lokalnym komputerze wyposażonym w GPU. Korzystanie z GPU było kluczowe dla efektywnego przeprowadzania eksperymentów, zwłaszcza w kontekście obliczeniowo intensywnych operacji związanych z przetwarzaniem obrazów. W ramach projektu zastosowano system kontroli wersji GIT, a cały kod źródłowy oraz wyniki analiz zostały zapisane i wersjonowane na platformie GitHub. Użycie GIT umożliwiło efektywne śledzenie zmian w kodzie, co pozwoliło na łatwe zarządzanie i kontrolowanie wersji poszczególnych plików oraz eksperymentów. Dzięki temu każdy etap projektu był dokładnie dokumentowany, co ułatwiało powrót do wcześniejszych wersji kodu w razie potrzeby oraz analizę postępów prac. Ponadto, platforma GitHub zapewniła bezpieczne i zorganizowane przechowywanie kodu.

Odpowiednie przygotowanie środowiska z użyciem wymienionych narzędzi i bibliotek było fundamentem dla przeprowadzenia skutecznych i efektywnych badań nad wpływem tła na klasyfikację obrazów.

WYBRANE MODELE

W niniejszym projekcie zastosowano trzy zaawansowane modele głębokiego uczenia: ResNet, ConvNeXt oraz DeepLabv3. Każdy z tych modeli został wybrany ze względu

na swoje unikalne właściwości i zdolności do realizacji określonych zadań. DeepLabv3 służył jako uniwersalny model do segmentacji, pozwalający na precyzyjne wyodrębnienie obiektów z tła. Modele ResNet i ConvNeXt, o różnych architekturach i z różnymi stopniami zaawansowania technologicznego, zostały wykorzystane do klasyfikacji obrazów. ResNet, będący starszym modelem, oraz ConvNeXt, reprezentujący nowsze podejście, zostały wybrane w celu porównania i analizy ich wydajności w kontekście zmodyfikowanych warunków tła. Wykorzystanie gotowych, pretrenowanych modeli umożliwiło skupienie się na głównej części badania, jaką jest wpływ tła na klasyfikację obrazów, zamiast na długotrwałym procesie trenowania modeli od podstaw.

ResNet

ResNet (Residual Network) został zaproponowany w 2015 roku i szybko stał się jednym z najważniejszych modeli w dziedzinie głębokiego uczenia. Główną innowacją ResNet jest wprowadzenie residual learning poprzez zastosowanie skrótowych połączeń (skip connections). Pozwala to na efektywne trenowanie bardzo głębokich sieci, nawet o setkach warstw, rozwiązuając problem zanikania gradientu.

W tradycyjnych sieciach neuronowych, gdy liczba warstw wzrasta, problem zanikania gradientu staje się bardziej wyraźny, co utrudnia efektywne trenowanie modeli. ResNet adresuje ten problem poprzez wprowadzenie bezpośrednich połączeń skrótowych, które umożliwiają przepływ gradientu bezpośrednio przez sieć, omijając kilka warstw pośrednich.

Podstawowym elementem budulcowym ResNet jest blok residual, który może być opisany równaniem:

$$y = F(x, \{W_i\}) + x \quad (1)$$

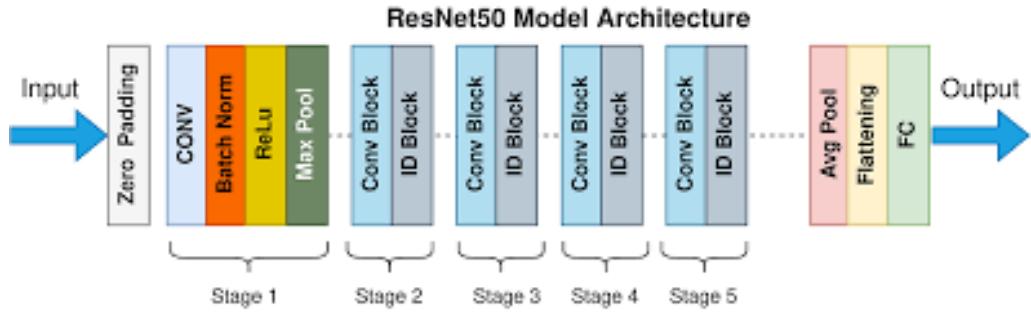
gdzie y to wyjście bloku, x to wejście, a $F(x, \{W_i\})$ to funkcja reprezentująca operacje konwolucyjne na wejściu x z zestawem wag $\{W_i\}$.

Bloki residual składają się zazwyczaj z dwóch lub trzech warstw konwolucyjnych z dodatkowymi połączonymi skrótowymi, które dodają wejście x do wyjścia $F(x, \{W_i\})$. To proste, ale skuteczne podejście pozwala na trenowanie bardzo głębokich sieci, które byłyby trudne do nauczenia przy użyciu tradycyjnych metod.

ResNet został wybrany do tego projektu ze względu na swoją zdolność do efektywnego radzenia sobie z bardzo głębokimi sieciami oraz udowodnioną skuteczność w wielu zadaniach klasyfikacyjnych.

Konkretnym modelem użyтыm w przeprowadzonych badaniach jest ResNet50. Architektura ResNet50 jest podzielona na cztery główne części: warstwy konwolucyjne, blok tożsamościowy, blok konwolucyjny oraz warstwy całkowicie połączone. Schemat architektury tego modelu można zobaczyć na Rys. ??

Warstwy konwolucyjne w ResNet50 składają się z kilku warstw konwolucyjnych, po



Rys. 1. Schemat architektury modelu ResNet50

których następuje normalizacja wsadowa (batch normalization) oraz aktywacja ReLU. Warstwy te są odpowiedzialne za ekstrakcję cech z obrazu wejściowego, takich jak krawędzie, tekstury i kształty. Następnie warstwy konwolucyjne są uzupełnione warstwami maksymalnego pooling (max pooling), które redukują przestrzenne wymiary map cech, jednocześnie zachowując najważniejsze cechy.

Blok tożsamościowy (identity block) i blok konwolucyjny (convolutional block) są kluczowymi elementami budulcowymi ResNet50. Blok tożsamościowy jest prostym blokiem, który przekazuje wejście przez szereg warstw konwolucyjnych i dodaje wejście z powrotem do wyjścia. Pozwala to sieci uczyć się funkcji resztkowych, które mapują wejście na pożądane wyjście. Blok konwolucyjny jest podobny do bloku tożsamościowego, ale z dodatkiem warstwy konwolucyjnej 1x1, która jest używana do redukcji liczby filtrów przed warstwą konwolucyjną 3x3.

Ostatnią częścią ResNet50 są warstwy całkowicie połączone (fully connected layers). Warstwy te są odpowiedzialne za dokonanie ostatecznej klasyfikacji. Wyjście z ostatniej warstwy całkowicie połączonej jest przekazywane do funkcji aktywacji softmax, aby uzyskać ostateczne prawdopodobieństwa klas.

ConvNeXt

ConvNeXt to nowoczesna architektura CNN, która została opracowana w celu integracji najlepszych praktyk z konwolucyjnych sieci neuronowych i nowoczesnych technik pochodzących od transformerów. ConvNeXt wykorzystuje bardziej złożone operacje konwolucyjne oraz zaawansowane techniki normalizacji i optymalizacji, co pozwala na osiąganie znakomitych wyników w różnych zadaniach klasyfikacji.

ConvNeXt został zaprojektowany z myślą o zastosowaniu najnowszych technik z dziedziny głębokiego uczenia, takich jak normalizacja warstw (Layer Normalization), mechanizmy uwagi (Attention Mechanisms) oraz bardziej złożone architektury warstw konwolucyjnych. W ConvNeXt zastosowano podejście polegające na udoskonaleniu tradycyjnych modułów konwolucyjnych poprzez dodanie elementów inspirowanych transformerami, co prowadzi do lepszej wydajności i efektywności obliczeniowej.

Podstawowym elementem ConvNeXt jest moduł konwolucyjny, który został zoptymalizowany w celu lepszego uchwycenia złożonych wzorców w danych. Architektura ConvNeXt łączy tradycyjne podejścia konwolucyjne z nowymi koncepcjami, co prowadzi do lepszej wydajności i efektywności obliczeniowej.

ConvNeXt został wybrany do tego projektu ze względu na swoje nowoczesne podejście i wysoką wydajność w klasyfikacji obrazów, co pozwala na dokładne porównanie z wcześniejszymi modelami, takimi jak ResNet.

DeepLabv3

DeepLabv3 jest modelem segmentacji obrazów, który został opracowany przez zespół Google. Wykorzystuje on techniki takie jak atrous convolutions (dylatowane konwolucje) i Conditional Random Fields (CRFs), które pozwalają na dokładne modelowanie kontekstowych informacji na różnych skalach. DeepLabv3+ jest najnowszą wersją tej serii, która łączy atrous convolutions z modelem spatial pyramid pooling, co pozwala na uchwycenie bogatych informacji kontekstowych.

Podstawowym elementem DeepLabv3 jest zastosowanie atrous convolutions, które mogą być opisane równaniem:

$$y[i] = \sum_{k=1}^K x[i + r \cdot k] \cdot w[k] \quad (2)$$

gdzie $y[i]$ to wyjście konwolucji, x to wejście, w to zestaw wag, K to rozmiar filtra, a r to współczynnik dylatacji.

Dzięki zastosowaniu atrous convolutions, DeepLabv3 może uchwycić informacje na różnych skalach bez utraty rozdzielczości, co jest kluczowe dla dokładnej segmentacji. Dodatkowo, wykorzystanie spatial pyramid pooling pozwala na zbieranie informacji kontekstowych z całego obrazu, co poprawia dokładność segmentacji.

DeepLabv3 został wybrany ze względu na swoją zdolność do precyzyjnej segmentacji obrazów, co jest kluczowe dla wyodrębnienia obiektów z tła przed dalszą analizą i klasyfikacją.

Uzasadnienie wyboru modeli

Wybór ResNet, ConvNeXt oraz DeepLabv3 opierał się na ich sprawdzonej skuteczności w swoich dziedzinach oraz zdolności do realizacji celów tego projektu. ResNet, jako starszy model, pozwala na ocenę wpływu tła na klasyfikację obrazów w kontekście bardziej tradycyjnych architektur. ConvNeXt, będący nowoczesnym modelem, reprezentuje najnowsze podejścia i innowacje w dziedzinie głębokiego uczenia, co pozwala na ocenę, jak nowe technologie radzą sobie z problemem tła. DeepLabv3, jako zaawansowany model segmentacji, umożliwia precyzyjne usunięcie tła, co jest kluczowe dla analiz prowadzonych w ramach tego projektu.

Wykorzystanie gotowych, pretrenowanych modeli pozwoliło skupić się na głównym celu badania – analizie wpływu tła na klasyfikację obrazów – bez konieczności poświęcania czasu na trenowanie modeli od podstaw. Dzięki temu możliwe było przeprowadzenie bardziej szczegółowych i kompleksowych badań w zakresie modyfikacji tła i jego wpływu na wydajność modeli klasyfikacyjnych.

WYBRANE METRYKI

W celu analizy wyników klasyfikacji przed i po modyfikacjach tła, zastosowano cztery kluczowe metryki: dokładność (accuracy), pewność klasyfikacji (confidence scores), precyzję (precision), recall oraz F1 score. Analizowana będzie również macierz korelacji w celu zbadania wzajemnych zależności między klasami czy modyfikacjami oraz identyfikacji potencjalnych źródeł błędów. Metryki te zostały wybrane ze względu na ich zdolność do dostarczania wartościowych informacji na temat wydajności modeli w różnych warunkach. W niniejszym rozdziale szczegółowo omówimy te metryki, przedstawimy odpowiednie wzory oraz wyjaśnimy, dlaczego zostały wybrane do analizy. Metryki będą analizowane całościowo, jak również osobno dla każdej klasy, dla każdej różnej modyfikacji tła oraz dla różnych percentylów wielkości obiektu na obrazie.

Dokładność (Accuracy)

Dokładność jest jedną z najprostszych i najbardziej intuicyjnych metryk stosowanych do oceny jakości modeli klasyfikacyjnych. Definiuje się ją jako stosunek liczby poprawnie sklasyfikowanych przykładów do całkowitej liczby przykładów.

Wzór na dokładność jest następujący:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

gdzie:

- TP (True Positives) - liczba prawdziwie pozytywnych przypadków,
- TN (True Negatives) - liczba prawdziwie negatywnych przypadków,
- FP (False Positives) - liczba fałszywie pozytywnych przypadków,
- FN (False Negatives) - liczba fałszywie negatywnych przypadków.

Dokładność została wybrana jako podstawowa metryka oceny modeli, ponieważ daje ogólny obraz wydajności modelu.

Precyzja (Precision)

Precyzja (precision) jest metryką oceniającą dokładność pozytywnych predykcji modelu. Definiuje się ją jako stosunek liczby prawdziwie pozytywnych przypadków do sumy liczby

prawdziwie ozytywnych i fałszywie pozytywnych przypadków. Wzór na precyzję jest następujący:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precyzja jest ważna, gdy koszty fałszywie pozytywnych klasyfikacji są wysokie.

Recall

Recall, zwany również czułością lub TPR (True Positive Rate), mierzy zdolność modelu do wykrywania wszystkich pozytywnych przypadków. Jest definiowany jako stosunek liczby prawdziwie pozytywnych przypadków do sumy liczby prawdziwie pozytywnych i fałszywie negatywnych przypadków. Wzór na recall jest następujący:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall jest istotny, gdy koszty fałszywie negatywnych klasyfikacji są wysokie.

F1 Score

F1 score jest harmoniczną średnią precyzji i recall. Jest używany jako pojedyncza metryka oceniająca wydajność modelu, która uwzględnia zarówno precyzję, jak i recall. Wzór na F1 score jest następujący:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 score jest użyteczny, gdy istnieje nierównowaga między liczbą pozytywnych i negatywnych przypadków.

Pewność klasyfikacji (Confidence Scores)

Pewność klasyfikacji (confidence scores) odnosi się do stopnia pewności modelu co do przypisania danego przykładu do określonej klasy. Jest to istotna metryka, ponieważ dostarcza dodatkowych informacji o tym, jak pewny jest model swoich predykcji. Wyższe wartości pewności oznaczają większe zaufanie modelu do swojej klasyfikacji.

Analiza pewności klasyfikacji pozwala na ocenę, jak model radzi sobie z przypadkami trudnymi do sklasyfikowania oraz czy zmiany tła wpływają na pewność predykcji.

OPIS WYKORZYSTANEGO ZBIORU DANYCH

W ramach niniejszego badania wykorzystano zbiór danych ImageNet1k, który jest jednym z najbardziej rozpoznawalnych i szeroko stosowanych zestawów danych w dziedzinie przetwarzania obrazów i głębokiego uczenia. ImageNet1k składa się z obrazów

należących do 1000 różnych klas, co pozwala na wszechstronną ocenę wydajności modeli klasyfikacyjnych w różnorodnych scenariuszach.

Struktura zbioru danych

Zbiór danych ImageNet1k jest podzielony na trzy części: treningową, walidacyjną oraz testową. Każda z tych części ma określoną liczbę obrazów na klasę, co umożliwia wszechstronne trenowanie, walidację i testowanie modeli klasyfikacyjnych.

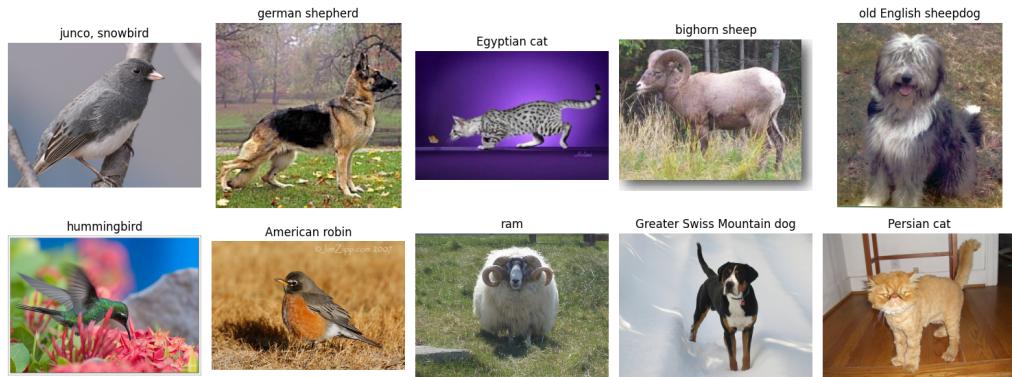
- **Zbiór treningowy:** Zawiera około 1300 obrazów na klasę, co daje szeroką bazę danych do nauki modeli. Duża liczba obrazów na klasę pozwala na efektywne trenowanie głębokich sieci neuronowych, co prowadzi do lepszego uchwycenia cech charakterystycznych dla każdej klasy.
- **Zbiór walidacyjny:** Składa się z 50 obrazów na klasę. Zbiór walidacyjny jest używany do monitorowania wydajności modelu w trakcie treningu i do wczesnego wykrywania problemów takich jak nadmierne dopasowanie (overfitting).
- **Zbiór testowy:** Zawiera 100 obrazów na klasę. Zbiór testowy służy do ostatecznej oceny wydajności modeli po zakończeniu procesu treningu i walidacji.

Różnorodność obrazów

Obrazy w zbiorze ImageNet1k charakteryzują się różnorodnością rozdzielczości oraz warunków, w jakich zostały wykonane. Oznacza to, że obrazy mogą przedstawiać obiekty w różnych skalach, oświetleniach, perspektywach i na różnych tłaach. Taka różnorodność sprawia, że zbiór ImageNet1k doskonale odzwierciedla realistyczne warunki, z jakimi modele mogą się spotkać w praktycznych zastosowaniach. Dzięki temu, modele trenowane na tym zbiorze danych są bardziej uniwersalne i mają lepszą zdolność generalizacji.

Popularność i znaczenie ImageNet1k

ImageNet1k jest jednym z najczęściej używanych zestawów danych w badaniach nad głębokim uczeniem, co jest wynikiem jego dużej skali, różnorodności i realistycznego charakteru. Wiele przełomowych modeli, takich jak AlexNet, VGG, ResNet i Inception, zostało przetestowanych i zweryfikowanych przy użyciu tego zestawu danych. Popularność ImageNet1k sprawia, że wyniki uzyskane na tym zbiorze są łatwo porównywalne z wynikami innych badań, co umożliwia ocenę postępów i przy użyciu tego modelu. Wybór tych klas pozwolił na przeprowadzenie dokładnych analiz i eksperymentów przy zachowaniu rozsądniego czasu przetwarzania.



Rys. 2. Przykładowe oryginalne zdjęcia wybranych klas

PLAN BADAŃ

Niniejszy rozdział opisuje szczegółowy plan badań, które zostały przeprowadzone w celu zbadania wpływu tła na klasyfikację obrazów zwierząt. Poniżej przedstawiono kroki podjęte w celu realizacji badań.

Przygotowanie środowiska pracy

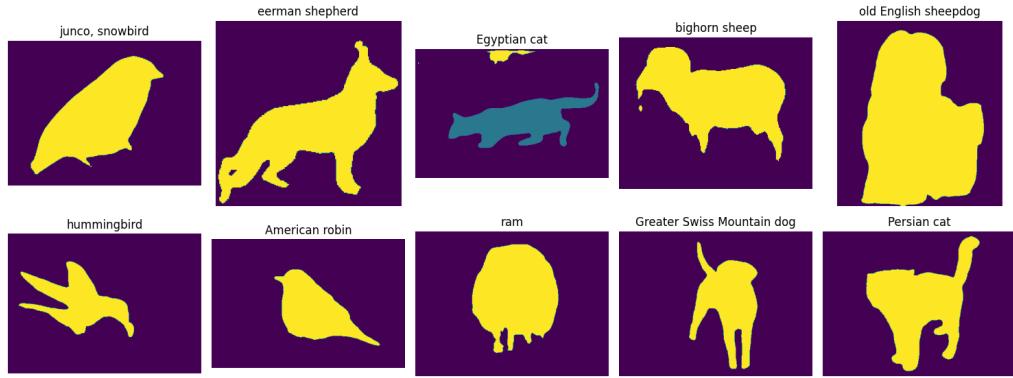
Pierwszym krokiem było przygotowanie odpowiedniego środowiska pracy. W tym celu skonfigurowano środowisko programistyczne, które obejmowało instalację niezbędnych bibliotek i narzędzi, takich jak numpy, pandas, scikit-learn, PIL, matplotlib, seaborn, torch oraz torchvision. Zastosowano również system kontroli wersji GIT do śledzenia zmian w kodzie i zarządzania wersjami projektu.

Wybranie modeli do segmentacji i klasyfikacji

Do segmentacji obrazów wybrano model DeepLabv3, który jest zaawansowanym modelem segmentacji zdolnym do precyzyjnego wyodrębniania obiektów z tła. Do klasyfikacji obrazów wybrano dwa modele: ResNet, reprezentujący starszą generację modeli głębokiego uczenia, oraz ConvNeXt, będący nowszym i bardziej zaawansowanym modelem. Wybór tych modeli pozwolił na dokładne porównanie ich wydajności w kontekście różnych modyfikacji tła.

Wybranie klas zwierząt

Ze względu na ograniczoną dostępność klas w modelu segmentacyjnym DeepLabv3, do badań wybrano 10 zróżnicowanych klas zwierząt. Wybrane klasy były takie, które można skutecznie wysegmentować przy użyciu tego modelu, co pozwoliło na przeprowadzenie dokładnych analiz i eksperymentów. Przykładowe zdjęcia dla każdej wybranej klasy można zaobserwować na Rys. ??



Rys. 3. Maski przykładowych wysegmentowanych obrazów

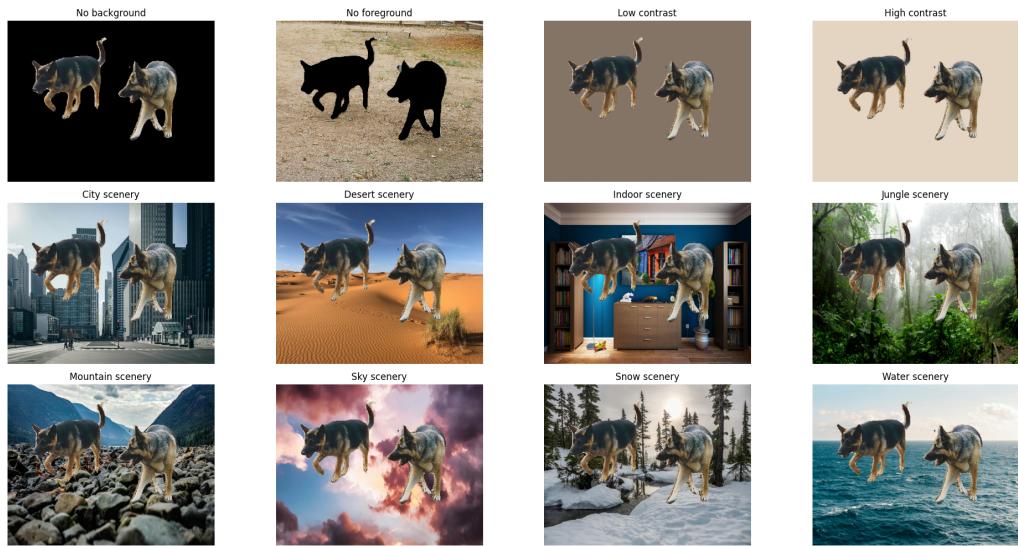
Segmentacja obrazów

Dla każdej z wybranych klas zwierząt wysegmentowano 1000 zdjęć za pomocą modelu DeepLabv3. Ponieważ na niektórych zdjęciach widniało więcej klas obiektów rozpoznawanych przez ten model, konieczne było zidentyfikowanie wartości grayscale dla pożąданej maski obiektu. W tym celu zastosowano skrypt analizujący najczęściej występującą wartość na przestrzeni wszystkich zdjęć dla danej klasy, co pozwoliło na dokładne wyodrębnienie obiektów. Maski zostały zapisane do dalszych analiz. Przykładowe uzyskane maski można zaobserwować na Rys. ??

Przygotowanie zmodyfikowanych zbiorów zdjęć

Dla każdej klasy zwierząt przygotowano różne zestawy zmodyfikowanych zdjęć:

- **Zdjęcia z samym obiektem:** Usunięto tło za pomocą maski, pozostawiając czarne tło.
- **Zdjęcia z samym tłem:** Odwrotna modyfikacja do poprzedniej, pozostawiając samo tło bez obiektu.
- **Przeniesienie obiektu na różne scenerie:** Obiekty zostały przeniesione na różne tła, takie jak niebo, wnętrze domu, pustynia, śnieg, woda, miasto, dżungla, góry.
- **Zastąpienie tła kolorem o niskim oraz wysokim kontraście do obiektu:** Dla każdej klasy zwierząt zbierano próbki kolorów. Obrazy były wczytywane i konwertowane do tablicy NumPy, a następnie tworzono maskę, która ignorowała czarne piksele (wartości RGB: 0, 0, 0), aby skupić się wyłącznie na rzeczywistych kolorach obiektów. Zebrane próbki kolorów były następnie klasteryzowane za pomocą algorytmu KMeans z pięcioma klastrami, co pozwalało na wyodrębnienie pięciu dominujących kolorów dla każdej klasy zwierząt. Dominujące kolory były normalizowane do skali 0-1, a następnie obliczana była ich średnia wartość, uznawana za kolor o niskim kontraście. Odległości euklidesowe każdego koloru od średniej wartości były obliczane, a kolor najbardziej oddalony od średniej był uznawany za kolor o wysokim kontraście. Ostatecznie, kolory były konwertowane z powrotem na skalę 0-255 i zapisywane jako wartości RGB.



Rys. 4. Przykładowe zdjęcie poddane modyfikacją

Przykładowe zdjęcie, wraz z jego modyfikacjami można zobaczyć na Rys. ??

Dokonanie predykcji na wybranych modelach

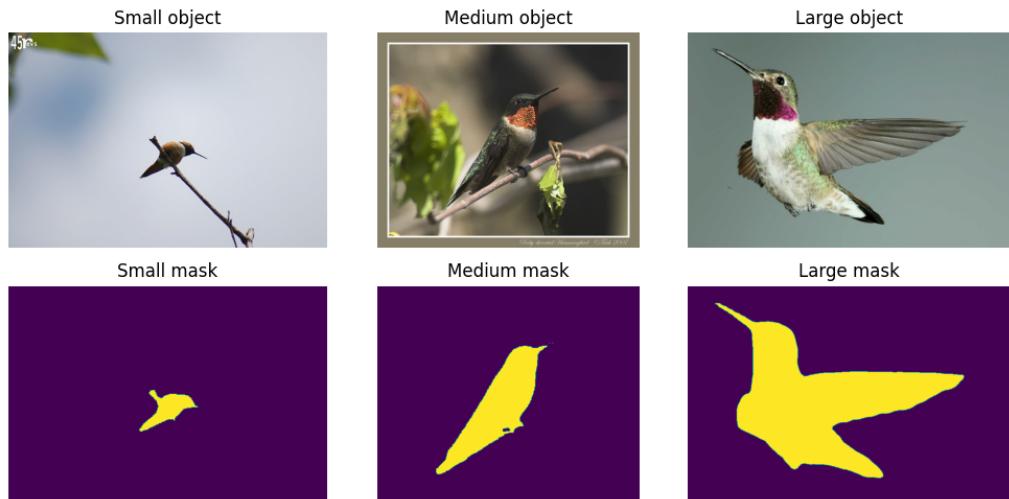
Predykcje zostały przeprowadzone na wybranych modelach ResNet oraz ConvNeXt dla oryginalnych zdjęć oraz dla każdej modyfikacji tła. Wyniki predykcji, w tym klasyfikacje oraz pewność klasyfikacji (confidence scores), zostały zapisane w plikach CSV.

Dodanie kategorii zdjęć pod względem procentu zajmowanego przez obiekt na zdjęciu

Zbadanie stosunku wielkości obiektu do całego zdjęcia jest istotne w kontekście badania wpływu tła na klasyfikację, ponieważ może znacznie wpływać na wyniki modeli klasyfikacyjnych. Wielkość obiektu w stosunku do tła może determinować, jak łatwo model jest w stanie rozpoznać i sklasyfikować obiekt. Mniejsze obiekty mogą być trudniejsze do wykrycia i bardziej podatne na zakłócenia ze strony tła, podczas gdy większe obiekty mogą dominować obraz, co ułatwia ich klasyfikację. Analiza wpływu różnych procentów wielkości obiektu pozwala na zrozumienie, w jakim stopniu tło oddziaływało na modele w zależności od proporcji obiektu na zdjęciu, co z kolei może prowadzić do bardziej efektywnych strategii przetwarzania i klasyfikacji obrazów w praktycznych zastosowaniach.

Dla każdego zdjęcia dodano kategorię pod względem procentu zajmowanego przez obiekt na zdjęciu. Przykładowe zdjęcia o różnych rozmiarach obiektów można zobaczyć na Rys. ??

1. **Obliczenie powierzchni obiektu:** Dla każdego obrazu obliczono liczbę pikseli zajmowanych przez obiekt.
2. **Obliczenie powierzchni całkowitej obrazu:** Liczba pikseli całego obrazu.



Rys. 5. Przykładowe zdjęcia o różnych rozmiarach obiektów dla klasy "hummingbird"

3. **Obliczenie procentu powierzchni zajmowanej przez obiekt:** Procent powierzchni zajmowanej przez obiekt obliczono za pomocą wzoru:

$$\text{Procent powierzchni zajmowanej przez obiekt} = \left(\frac{\text{Powierzchnia obiektu}}{\text{Powierzchnia całkowita obrazu}} \right) \times 100 \quad (4)$$

4. **Podział obiektów na percentile:** Obrazy posortowano według procentu powierzchni zajmowanej przez obiekt i podzielono na cztery grupy według wartości percentylów.

Implementacja

Całość implementacji została wykonana w języku Python, który dzięki swojej elastyczności i szerokiej gamie bibliotek doskonale nadaje się do realizacji złożonych projektów związanych z uczeniem maszynowym. Większość funkcjonalności została zaimplementowana w formie skryptów, podczas gdy same badania, czyli opracowywanie wyników, są dostępne w formie interaktywnych notebooków Jupyter. Taki podział pozwala na łatwe uruchamianie skryptów oraz jednoczesną analizę i wizualizację wyników.

Skrypty są nazwane zgodnie ze swoją funkcjonalnością, co ułatwia nawigację i zrozumienie ich przeznaczenia. Każdy skrypt zawiera funkcje z dobrze opisanymi nazwami, a każda funkcja posiada docstringi, które szczegółowo wyjaśniają jej działanie.

Dodatkowo, cała implementacja wraz z dokładnym opisem znajduje się na GitHubie, gdzie można znaleźć pełen kod oraz plik README zawierający instrukcje dotyczące uruchamiania skryptów i analizy wyników.

Link do repozytorium GitHub: [link do repozytorium](#)

Analiza wyników

Wyniki zostały poddane szczegółowej analizie w kilku aspektach:

- **Analiza ogólna wyników:** Ogólna wydajność modeli na całym zbiorze danych.
- **Analiza pod kątem klasy:** Wydajność modeli dla każdej klasy zwierząt osobno.
- **Analiza pod kątem wielkości obiektu:** Wydajność modeli w zależności od wielkości obiektu na zdjęciu (cztery percentile).
- **Porównanie modeli:** Porównanie wyników starszego modelu ResNet oraz nowszego ConvNeXt.

Wnioski i dalsze kierunki rozwoju

Na podstawie przeprowadzonych analiz wyciągnięto wnioski dotyczące wpływu tła na wyniki klasyfikacji obrazów zwierząt. Ponadto zaproponowano dalsze kierunki rozwoju, mające na celu optymalizację modeli klasyfikacyjnych w kontekście zmieniających się warunków tła. Wnioski oraz dalsze kierunki rozwoju znajdują się w kolejnym rozdziale.

Podsumowanie

Plan badań obejmował szczegółowe przygotowanie środowiska pracy, wybór odpowiednich modeli do segmentacji i klasyfikacji, segmentację obrazów, przygotowanie zmodifikowanych zbiorów zdjęć oraz przeprowadzenie predykcji i analiz wyników. Dzięki systematycznemu podejściu możliwe było uzyskanie wartościowych wniosków na temat wpływu tła na wydajność modeli klasyfikacyjnych oraz identyfikacja obszarów wymagających dalszych badań i optymalizacji.

BADANIA

Celem tego rozdziału jest przeprowadzenie analizy wyników klasyfikacji obrazów zwierząt dla modeli ResNet i ConvNeXt. Analiza obejmuje porównanie skuteczności modeli w różnych scenariuszach modyfikacji tła oraz w zależności od wielkości obiektu na obrazie. Przeanalizowane zostaną ogólne metryki, wyniki dla poszczególnych klas oraz wpływ wielkości obiektu na dokładność klasyfikacji.

WYNIKI OGÓLNE

Badania miały na celu zbadanie wpływu modyfikacji tła na skuteczność klasyfikacji obrazów za pomocą dwóch modeli głębokiego uczenia: ResNet oraz ConvNeXt. W tym celu dokonano obliczeń podstawowych metryk, takich jak Accuracy, Precision, Recall i F1-score, dla oryginalnych oraz zmodyfikowanych obrazów, traktując wszystkie modyfikacje jako jedną grupę.

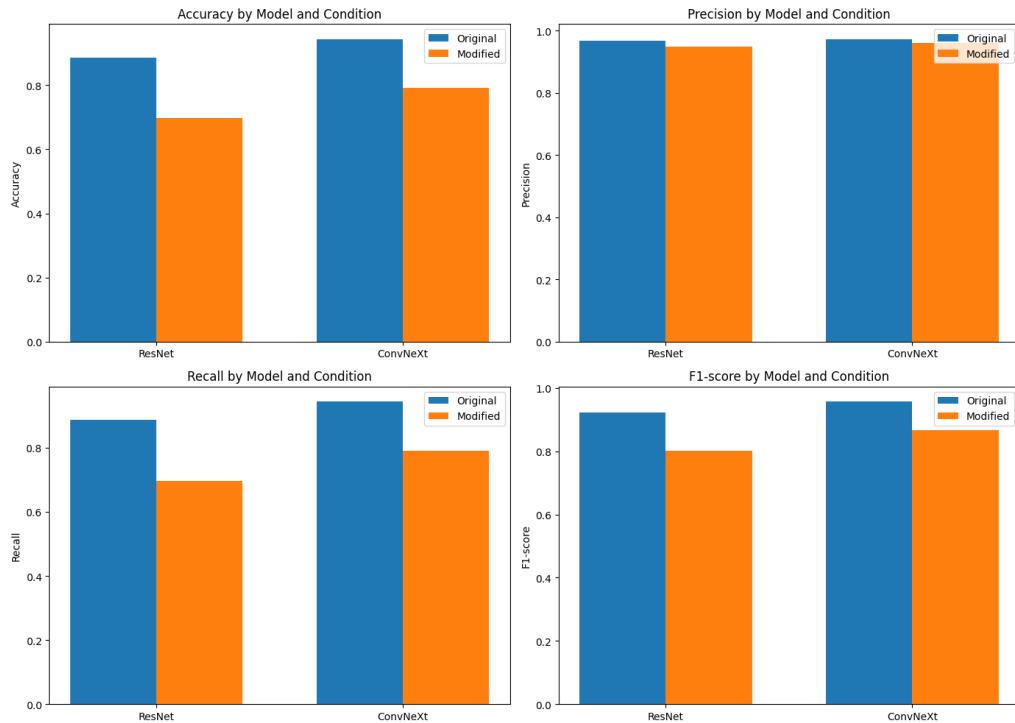
Dla modelu ResNet, na oryginalnych obrazach uzyskano Accuracy na poziomie 0.886500, Precision 0.967026, Recall 0.886500 i F1-score 0.922742. Po modyfikacji tła wartości te uległy znacznemu obniżeniu, osiągając odpowiednio 0.697018 dla Accuracy, 0.948539 dla Precision, 0.697018 dla Recall i 0.802350 dla F1-score. W przypadku modelu ConvNeXt, na oryginalnych obrazach uzyskano wartości: Accuracy 0.943300, Precision 0.972519, Recall 0.943300 i F1-score 0.956791. Podobnie jak w przypadku ResNet, modyfikacja tła spowodowała obniżenie tych wartości, osiągając Accuracy 0.790873, Precision 0.961080, Recall 0.790873 i F1-score 0.866282.

Analiza wyników wskazuje, że modyfikacja tła negatywnie wpływa na skuteczność obu modeli, jednak model ConvNeXt wykazuje większą odporność na zmiany tła niż ResNet. Model ConvNeXt osiąga wyższe wartości metryk zarówno dla oryginalnych, jak i zmodyfikowanych obrazów, co sugeruje jego większą stabilność i lepszą adaptację do różnych warunków. Wartości metryk dla zmodyfikowanych obrazów są niższe w przypadku ResNet, co może wskazywać na większą wrażliwość tego modelu na zmiany w tle.

Wnioski z badań sugerują, że dla zadań klasyfikacyjnych, gdzie modyfikacje tła mogą występować, model ConvNeXt jest bardziej odpowiedni. Dalsze badania nad metodami przetwarzania i augmentacji danych mogą pomóc w zminimalizowaniu wpływu modyfikacji tła na wyniki klasyfikacji, co jest kluczowe dla poprawy dokładności i niezawodności modeli głębokiego uczenia. Wyniki te podkreślają znaczenie wyboru odpowiedniego modelu oraz technik przetwarzania danych w kontekście zadań związanych z klasyfikacją obrazów.

Model	Type	Accuracy	Precision	Recall	F1-score
ResNet	Original	0.886500	0.967026	0.886500	0.922742
ResNet	Modified	0.697018	0.948539	0.697018	0.802350
ConvNeXt	Original	0.943300	0.972519	0.943300	0.956791
ConvNeXt	Modified	0.790873	0.961080	0.790873	0.866282

Tabela 1. Metryki porównawcze modeli ResNet i ConvNeXt



Rys. 6. Metryki dla danych oryginalnych zestawionych z danymi o zmodyfikowanych tła

Wyniki badań również wskazują, że pomimo modyfikacji tła, Precision dla obu modeli (ResNet i ConvNeXt) uległa jedynie niewielkiemu spadkowi. Dla ResNet Precision zmniejszyła się z 0.967026 na 0.948539, a dla ConvNeXt z 0.972519 na 0.961080. Mały spadek Precision w obu przypadkach sugeruje, że oba modele nadal skutecznie identyfikują prawdziwie pozytywne przypadki, nawet po modyfikacji tła.

To zjawisko można interpretować jako wskazówkę, że oba modele są dobrze dostrojone do rozpoznawania właściwych cech istotnych dla klasyfikacji, niezależnie od zmieniającego się tła. Wysoka wartość Precision oznacza, że modele rzadko identyfikują fałszywie pozytywne przypadki, co jest szczególnie istotne w zastosowaniach, gdzie dokładność klasyfikacji jest kluczowa. Niewielki spadek Precision w przypadku modyfikacji tła sugeruje, że modele są w stanie skutecznie ignorować zmiany w tle i skoncentrować się na istotnych cechach obiektów, co jest pozytywnym aspektem ich działania.

W badaniach obliczono również średnie wartości confidence scores dla dwóch modeli: ResNet oraz ConvNeXt. Wartości te obejmują ogólną średnią confidence score, a także średnie confidence scores dla poprawnych i niepoprawnych klasyfikacji. Dla modelu ResNet na

Model	Type	Average	Average correct	Average incorrect
ResNet	Original	85.188854	89.137424	54.348263
ResNet	Modified	71.694490	83.929904	43.546579
ConvNeXt	Original	68.527975	70.036361	43.433439
ConvNeXt	Modified	57.543545	63.189640	36.191272

Tabela 2. Confidence scores dla modeli ResNet i ConvNeXt

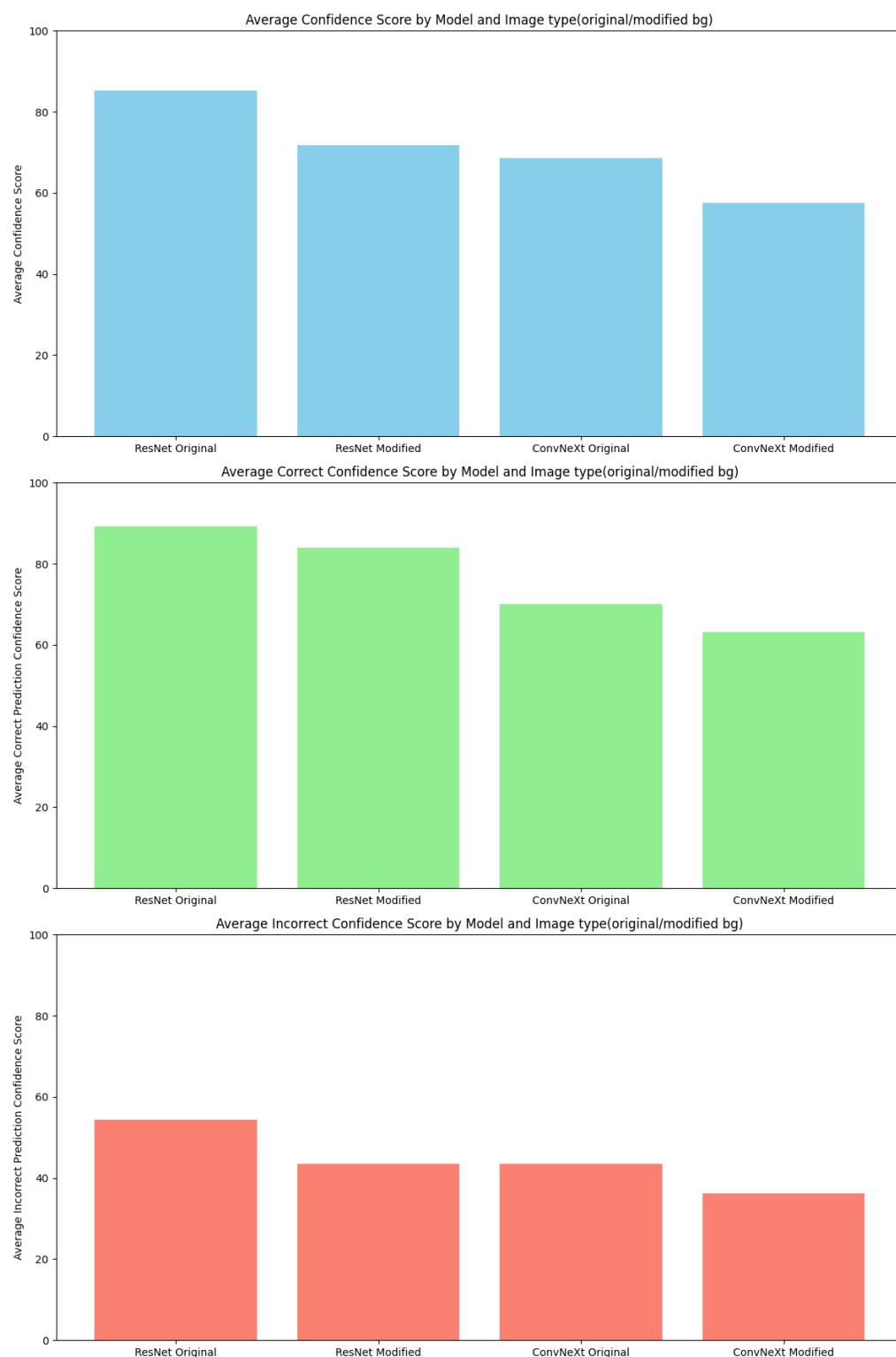
oryginalnych obrazach średnia confidence score wyniosła 85.188854, ze średnią wartością 89.137424 dla poprawnych klasyfikacji i 54.348263 dla niepoprawnych. Po modyfikacji tła, średnia confidence score spadła do 71.694490, ze średnią wartością 83.929904 dla poprawnych klasyfikacji i 43.546579 dla niepoprawnych.

W przypadku modelu ConvNeXt na oryginalnych obrazach średnia confidence score wyniosła 68.527975, ze średnią wartością 70.036361 dla poprawnych klasyfikacji i 43.433439 dla niepoprawnych. Po modyfikacji tła, średnia confidence score spadła do 57.543545, ze średnią wartością 63.189640 dla poprawnych klasyfikacji i 36.191272 dla niepoprawnych.

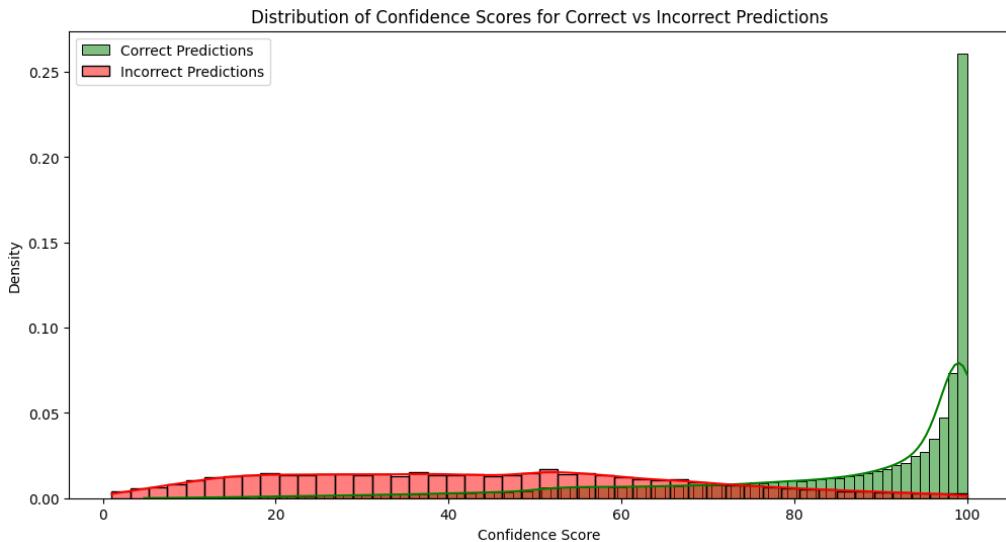
Wyniki te wskazują na znaczący spadek średnich confidence scores dla obu modeli po modyfikacji tła. Średnie confidence scores dla poprawnych klasyfikacji są wyższe niż dla niepoprawnych w obu przypadkach, co sugeruje, że modele są bardziej pewne swoich poprawnych klasyfikacji. Jednak modyfikacja tła powoduje ogólny spadek pewności modeli, co może być wynikiem zmniejszenia jasności sygnałów związanych z obiektami do rozpoznania.

Analiza wartości confidence scores dla ConvNeXt wskazuje na większy spadek w porównaniu do ResNet. Model ConvNeXt na oryginalnych obrazach ma niższą ogólną średnią confidence score (68.527975) w porównaniu do ResNet (85.188854). Jednakże, spadek ten jest bardziej wyraźny po modyfikacji tła, gdzie średnia confidence score dla ConvNeXt wynosi 57.543545, podczas gdy dla ResNet jest to 71.694490.

Podsumowując, modyfikacja tła ma wyraźny wpływ na zmniejszenie pewności klasyfikacji obrazów przez modele ResNet i ConvNeXt. Chociaż oba modele wykazują wysoką pewność przy poprawnych klasyfikacjach, modyfikacja tła powoduje ogólny spadek tych wartości, z bardziej zauważalnym spadkiem w przypadku modelu ConvNeXt. Wyniki te podkreślają znaczenie zrozumienia i zarządzania wpływem tła na wydajność modeli klasyfikacyjnych w praktycznych zastosowaniach.

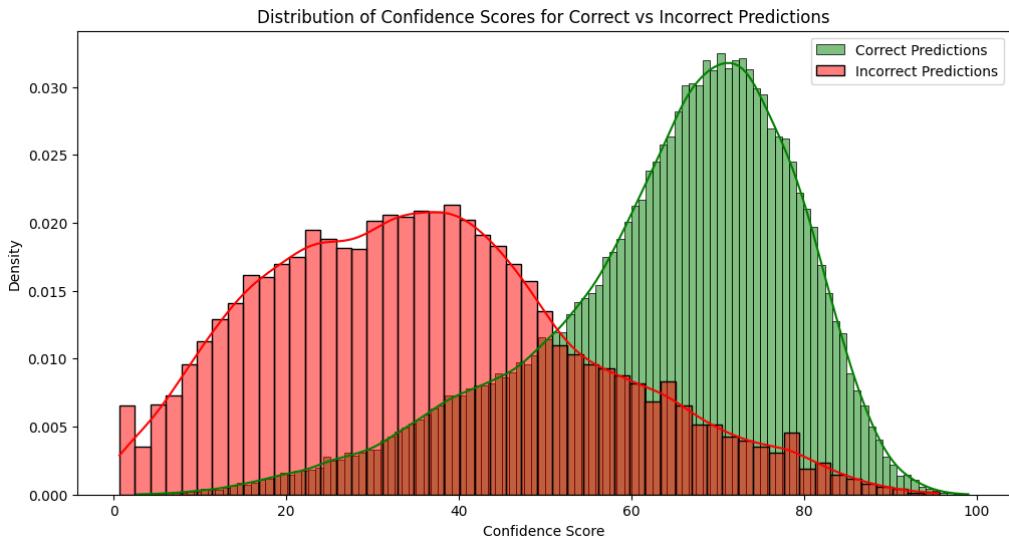


Rys. 7. Średnie wartości dla confidence scores



Rys. 8. Dystrybucja confidence score dla ResNet

Na podstawie dystrybucji confidence scores dla modeli ResNet i ConvNeXt można wyciągnąć kilka istotnych wniosków. Model ResNet wykazuje wysoką pewność dla poprawnych klasyfikacji, z koncentracją confidence scores w przedziale od 80 do 100, co wskazuje na jego stabilność w identyfikowaniu prawidłowych przypadków. W przypadku błędnych klasyfikacji confidence scores są bardziej równomiernie rozłożone, co oznacza, że model jest mniej pewny, gdy się myli. Z kolei model ConvNeXt ma szerszy rozkład confidence scores, z wyraźnym pikiem dla poprawnych klasyfikacji w przedziale 70-80 i dla błędnych w przedziale 30-40. Sugeruje to, że ConvNeXt jest mniej pewny swoich poprawnych decyzji w porównaniu do ResNet i bardziej skłonny do popełniania błędów z dużą pewnością. Te różnice w rozkładzie pewności klasyfikacji wskazują, że ResNet jest bardziej stabilny, podczas gdy ConvNeXt wykazuje większą różnorodność w pewności decyzji, co może wpływać na jego wydajność w obliczu modyfikacji tła.



Rys. 9. Dystrybucja confidence score dla ConvNext

Object Size	Dataset Type	Accuracy	Precision	Recall	F1-score
Large	Original	0.877353	0.972679	0.877353	0.920111
Large	Modified	0.782460	0.960542	0.782460	0.861260
Medium	Original	0.897879	0.963620	0.897879	0.927600
Medium	Modified	0.757796	0.943584	0.757796	0.838421
Small	Original	0.884545	0.963823	0.884545	0.918217
Small	Modified	0.548209	0.936858	0.548209	0.683307

Tabela 3. Metryki porównawcze modelu ResNet w zależności od wielkości obiektu

WYNIKI WZGLEDEM KATEGORII WIELKOŚCI OBIEKTU

Analiza wpływu wielkości obiektu na skuteczność klasyfikacji obrazów jest istotnym elementem badań, ponieważ różne rozmiary obiektów mogą znacząco wpływać na wydajność modeli głębokiego uczenia. Obiekty o różnych wielkościach mogą być różnie traktowane przez modele klasyfikacyjne ze względu na zmienność cech charakterystycznych oraz tła. Dlatego też, zrozumienie, jak zmienia się skuteczność modeli ResNet i ConvNeXt w zależności od wielkości obiektu, jest kluczowe dla optymalizacji i poprawy tych modeli w rzeczywistych zastosowaniach.

Dla modelu ResNet wyniki pokazują, że dla dużych obiektów na oryginalnych obrazach uzyskano Accuracy na poziomie 0.877353, Precision 0.972679, Recall 0.877353 i F1-score 0.920111. Po modyfikacji tła wartości te spadły odpowiednio do 0.782460, 0.960542, 0.782460 i 0.861260. Dla obiektów średniej wielkości na oryginalnych obrazach uzyskano Accuracy 0.897879, Precision 0.963620, Recall 0.897879 i F1-score 0.927600, a po modyfikacji tła wartości te spadły do 0.757796, 0.943584, 0.757796 i 0.838421. Dla małych obiektów na oryginalnych obrazach uzyskano Accuracy 0.884545, Precision 0.963823,

Object Size	Dataset Type	Accuracy	Precision	Recall	F1-score
Large	Original	0.932353	0.974247	0.932353	0.952076
Large	Modified	0.838610	0.970077	0.838610	0.897830
Medium	Original	0.943333	0.970287	0.943333	0.955907
Medium	Modified	0.834545	0.959503	0.834545	0.888990
Small	Original	0.954545	0.972766	0.954545	0.961567
Small	Modified	0.698017	0.951014	0.698017	0.799677

Tabela 4. Metryki porównawcze modelu ConvNeXt w zależności od wielkości obiektu

Recall 0.884545 i F1-score 0.918217, natomiast po modyfikacji tła wartości te drastycznie spadły do 0.548209, 0.936858, 0.548209 i 0.683307.

Dla modelu ConvNeXt wyniki pokazują, że dla dużych obiektów na oryginalnych obrazach uzyskano Accuracy na poziomie 0.932353, Precision 0.974247, Recall 0.932353 i F1-score 0.952076. Po modyfikacji tła wartości te spadły odpowiednio do 0.838610, 0.970077, 0.838610 i 0.897830. Dla obiektów średniej wielkości na oryginalnych obrazach uzyskano Accuracy 0.943333, Precision 0.970287, Recall 0.943333 i F1-score 0.955907, a po modyfikacji tła wartości te spadły do 0.834545, 0.959503, 0.834545 i 0.888990. Dla małych obiektów na oryginalnych obrazach uzyskano Accuracy 0.954545, Precision 0.972766, Recall 0.954545 i F1-score 0.961567, natomiast po modyfikacji tła wartości te spadły do 0.698017, 0.951014, 0.698017 i 0.799677.

Analiza wyników pokazuje, że modyfikacja tła wpływa na skuteczność klasyfikacji obrazów dla obu modeli, ale wpływ ten jest zróżnicowany w zależności od wielkości obiektu. W przypadku małych obiektów, procentowo jest więcej tła niż w przypadku dużych czy średnich obiektów, co może prowadzić do większych spadków w Accuracy i innych metrykach. Model ResNet wykazuje znaczący spadek skuteczności dla małych obiektów po modyfikacji tła, co sugeruje, że większa ilość tła może wprowadzać większe zakłócenia w procesie klasyfikacji. To samo zjawisko obserwowane jest w przypadku modelu ConvNeXt, choć ten model wykazuje lepszą odporność na zmiany tła niż ResNet, zwłaszcza dla małych obiektów.

Wyniki te podkreślają znaczenie rozważania wielkości obiektów przy projektowaniu i ocenie modeli klasyfikacyjnych, szczególnie w kontekście modyfikacji tła. Modele mogą wymagać dodatkowego dostrajania lub augmentacji danych, aby poprawić ich odporność na zmiany tła, co jest szczególnie istotne dla zastosowań, gdzie obiekty mogą występować w różnych skalach i warunkach środowiskowych. Zrozumienie, że większy udział tła przy małych obiektach może prowadzić do większych zakłóceń, jest kluczowe dla opracowywania bardziej efektywnych algorytmów klasyfikacyjnych, które mogą niezawodnie działać w zmiennych warunkach.

Analiza confidence scores dla modeli ResNet i ConvNeXt w zależności od wielkości obiektu dostarcza cennych informacji o zachowaniu tych modeli. Dla modelu ResNet, średnie confidence scores wynoszą 76.437807 dla dużych obiektów, 76.371553 dla obiek-

Object Size	Average Score	Correct Score	Incorrect Score
Large	76.437807	85.211125	43.360151
Medium	76.371553	85.131233	47.133278
Small	65.538042	82.538231	42.420987

Tabela 5. Confidence scores dla modelu ResNet w zależności od wielkości obiektu oraz poprawności predykcji

Object Size	Average Score	Correct Score	Incorrect Score
Large	63.942224	69.166578	35.149080
Medium	60.220167	64.178825	38.865913
Small	51.048190	57.051329	35.657851

Tabela 6. Confidence scores dla modelu ConvNeXt w zależności od wielkości obiektu oraz poprawności predykcji

tów średniej wielkości i 65.538042 dla małych obiektów. Dla poprawnych klasyfikacji confidence scores są odpowiednio wyższe, wynosząc 85.211125, 85.131233 i 82.538231. Dla niepoprawnych klasyfikacji wartości te są znacznie niższe: 43.360151, 47.133278 i 42.420987.

W przypadku modelu ConvNeXt, średnie confidence scores są niższe i wynoszą 63.942224 dla dużych obiektów, 60.220167 dla obiektów średniej wielkości i 51.048190 dla małych obiektów. Confidence scores dla poprawnych klasyfikacji wynoszą 69.166578, 64.178825 i 57.051329, a dla niepoprawnych klasyfikacji są to 35.149080, 38.865913 i 35.657851.

Yniki te wskazują, że confidence scores dla poprawnych klasyfikacji są wyższe niż dla niepoprawnych w obu modelach, co jest oczekiwane, ponieważ modele mają większą pewność przy poprawnych decyzjach. Jednakże, spadki confidence scores dla małych obiektów są bardziej wyraźne, co może wynikać z większego udziału tła w tych obrazach, co prowadzi do większych zakłóceń i trudności w klasyfikacji.

Porównując oba modele, ResNet wykazuje wyższe średnie confidence scores zarówno dla poprawnych, jak i niepoprawnych klasyfikacji w porównaniu do ConvNeXt. To sugeruje, że ResNet jest bardziej pewny swoich decyzji, niezależnie od wielkości obiektu.

WYNIKI WZGLEDEM TYPU MODYFIKACJI TŁA

W tej sekcji przeanalizowano wpływ różnych typów modyfikacji tła na skuteczność klasyfikacji obrazów oraz wartości confidence scores dla dwóch modeli głębokiego uczenia: ResNet i ConvNeXt. Analiza obejmuje metryki klasyfikacji takie jak accuracy, precision, recall i F1-score oraz średnie wartości confidence scores zarówno dla poprawnych, jak i niepoprawnych klasyfikacji. Taka analiza jest kluczowa dla zrozumienia, jak różne scenariusze modyfikacji tła wpływają na pewność i skuteczność modeli klasyfikacyjnych.

Modification Type	Accuracy	Precision	Recall	F1-score
Desert	0.7582	0.956658	0.7582	0.844717
Low	0.7835	0.957333	0.7835	0.859445
City	0.7620	0.955729	0.7620	0.845955
Sky	0.7577	0.951998	0.7577	0.840614
Jungle	0.7736	0.952716	0.7736	0.846911
No	0.1935	0.857598	0.1935	0.260654
High	0.7757	0.952540	0.7757	0.852606
Water	0.7198	0.950886	0.7198	0.812003
Snow	0.7462	0.961568	0.7462	0.833842
Indoor	0.6990	0.950477	0.6990	0.801644
Mountain	0.6980	0.959372	0.6980	0.800749

Tabela 7. Metryki według typu modyfikacji dla ResNet

Modification Type	Average Score	Correct Score	Incorrect Score
Desert	75.954396	84.304869	49.770241
Low	77.879677	86.411184	47.004688
City	73.458863	84.253234	38.898736
Sky	76.063737	84.893389	48.452397
Jungle	77.082096	85.830222	47.190088
No	41.084428	67.091235	34.844729
High	77.275950	86.369482	45.827655
Water	70.044268	80.529769	43.108282
Snow	76.379893	85.181503	50.502187
Indoor	72.859840	83.005442	49.299121
Mountain	70.556246	82.283334	43.451917

Tabela 8. Confidence Scores według typu modyfikacji dla ResNet

Modification Type	Accuracy	Precision	Recall	F1-score
Desert	0.8234	0.965246	0.8234	0.885689
Low	0.8897	0.965665	0.8897	0.925404
City	0.8392	0.967133	0.8392	0.894941
Sky	0.8259	0.958754	0.8259	0.886533
Jungle	0.8621	0.963912	0.8621	0.909624
No	0.2572	0.921080	0.2572	0.343413
High	0.8707	0.964345	0.8707	0.913518
Water	0.8593	0.960770	0.8593	0.906849
Snow	0.8567	0.966667	0.8567	0.902537
Indoor	0.7797	0.965780	0.7797	0.859119
Mountain	0.8357	0.964700	0.8357	0.891118

Tabela 9. Metryki według typu modyfikacji dla ConvNeXt

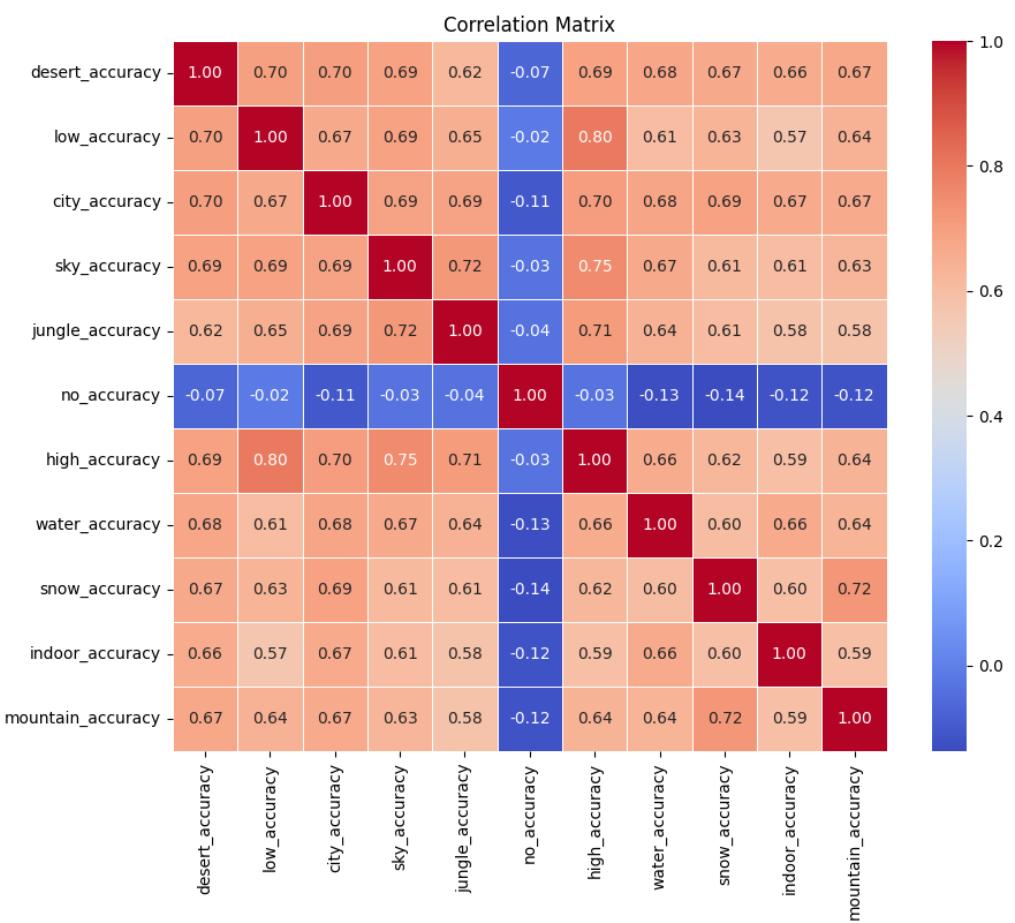
Modification Type	Average Score	Correct Score	Incorrect Score
Desert	61.750625	65.260020	45.388021
Low	57.798021	61.428446	28.514347
City	58.404538	62.933008	34.770883
Sky	56.927464	60.928925	37.945228
Jungle	60.313014	63.350853	41.321559
No	38.674228	56.533386	32.490362
High	57.146265	61.220278	29.712060
Water	57.189347	60.885762	34.614159
Snow	61.818994	65.527449	39.648495
Indoor	61.650110	66.519692	44.415372
Mountain	61.306386	66.316218	35.824236

Tabela 10. Confidence według typu modyfikacji dla ConvNeXt

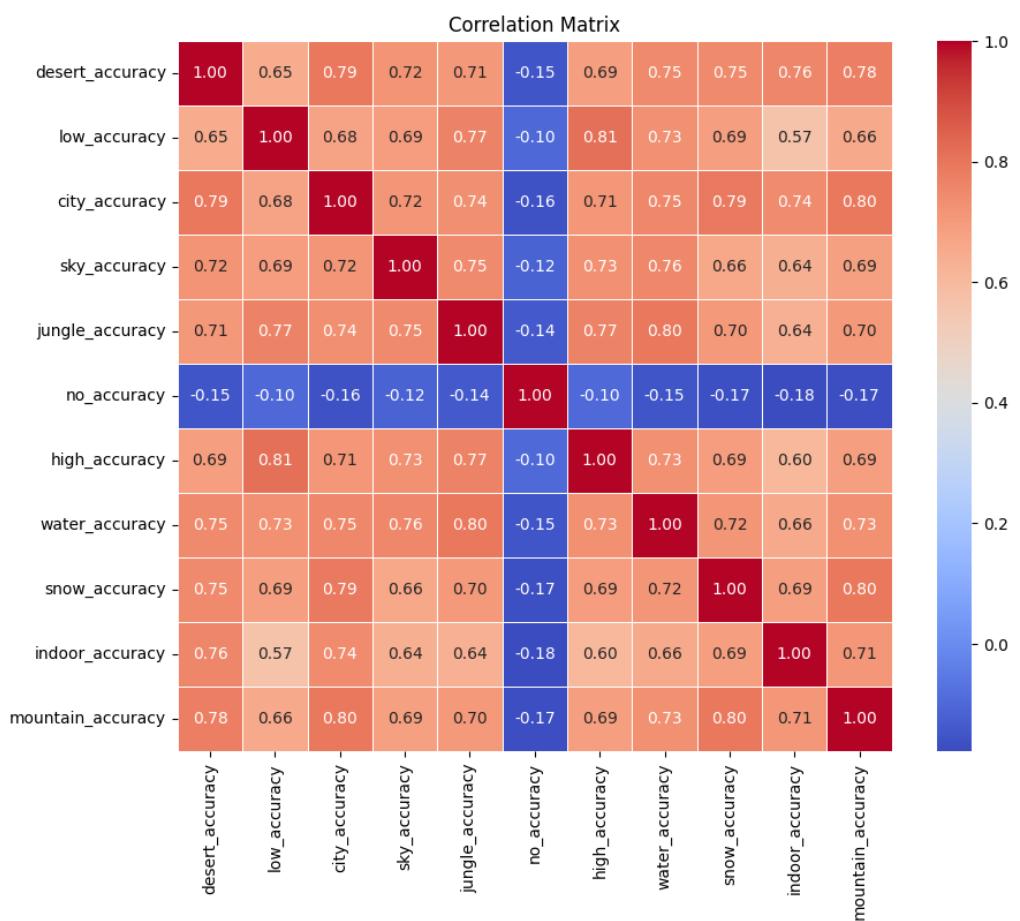
Analiza wyników pokazuje, że typ modyfikacji tła ma znaczący wpływ na skuteczność klasyfikacji obrazów oraz wartości confidence scores dla obu modeli. Modele osiągały najwyższą skuteczność i pewność klasyfikacji w scenariuszach, gdzie tło było bardziej uporządkowane lub zawierało mniej zakłócających elementów (np. "high", "low", "jungle"). Największe spadki skuteczności i pewności klasyfikacji obserwowano w scenariuszu "no", gdzie tło było całkowicie usunięte, co sugeruje, że brak tła może utrudniać modelom identyfikację istotnych cech obiektów.

Wyniki te podkreślają znaczenie uwzględniania różnych typów tła podczas trenowania i oceny modeli klasyfikacyjnych. Zrozumienie, jak różne modyfikacje tła wpływają na skuteczność i pewność klasyfikacji, może prowadzić do opracowania bardziej odpornych modeli, które lepiej radzą sobie w zróżnicowanych warunkach. Dalsze badania mogą skoncentrować się na metodach augmentacji danych, które mogą poprawić wydajność modeli w scenariuszach z różnorodnymi tłami.

WYNIKI WZGLEDEM KLAS



Rys. 10. Tabela korelacji dla ResNet dla typów modyfikacji



Rys. 11. Tabela korelacji dla ConvNeXt dla typów modyfikacji

PODSUMOWANIE

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

SPIS RYSUNKÓW

SPIS LISTINGÓW

SPIS TABEL

Dodatki