

**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 technologie głębokiego uczenia maszynowego zrewolucjonizowały obszar przetwarzania obrazów, w tym procesy takie jak klasyfikacja i segmentacja obrazów. Fundamentalnym zagadnieniem w komputerowym rozpoznawaniu wzorców jest klasyfikacja obrazów, czyli proces przypisywania etykiet do obiektów przedstawionych na obrazach. Pomimo ogromnych postępów, dokładność i niezawodność modeli klasyfikacyjnych są uzależnione od wielu czynników, z których jednym z kluczowych jest tło obrazu.

Tło obrazu może zawierać niepotrzebne dane lub wprowadzać modele w błąd, co może powodować błędne klasyfikacje obiektów. W przypadku klasyfikacji zwierząt, obecność złożonego lub nietypowego tła może mieć znaczący wpływ na wyniki klasyfikacji. W związku z tym analiza wpływu tła na wyniki klasyfikacji obrazów jest bardzo ważna dla poprawy efektywności modeli.

Jednym ze sposobów rozwiązywania problemu tła jest segmentacja obrazów, czyli podział obrazu na mniejsze części i przydzielenie im etykiet. Segmentacja pozwala wydzielić obiekt z tła, co może prowadzić do poprawy wyniku klasyfikacji.

Obecność zakłócających elementów w tle, zmienność oświetlenia, różnice w skalach obiektów i różnorodność danych to problemy, z którymi algorytmy klasyfikacji czy segmentacji muszą sobie poradzić.

Niniejsza praca skupia się na zbadaniu wpływu tła na klasyfikację oraz ocenę skuteczności modeli głębokiego uczenia w warunkach zmiennego tła. Przeprowadzone badania mają na celu poszerzenie wiedzy odnośnie optymalizacji modeli klasyfikacyjnych w warunkach zmiennego i nieprzewidywalnego tła, czyli takiego jakiego często występuje w rzeczywistych warunkach.

## 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. Jednym z kluczowych elementów tego badania jest ocena, w jaki sposób usunięcie, czy modyfikacja tła wpływają na wydajność tych modeli klasyfikacyjnych. Poprzez analizę wyników zarówno przed i po modyfikacji tła, praca ta ma na celu:

- **Ocena wrażliwości modeli na tło:** Sprawdzenie, jak różne rodzaje tła wpływają na dokładność (accuracy) klasyfikacji obrazów zwierząt. Analiza ta pozwoli zrozumieć, w

jakim stopniu obecność tła zakłóca proces klasyfikacji i jakie rodzaje czy modyfikacje tła mają największy wpływ na wyniki.

- **Optymalizacja procesu klasyfikacji:** Stwierdzenie jakie modyfikacje tła mogą pozytywnie wpływać na efektywność klasyfikacji i okazać się przydatne przy preprocessingu danych.
- **Porównanie wydajności modeli:** Porównanie dwóch różnych modeli klasyfikacyjnych w celu zrozumienia jak różne architektury czy założenia modeli mogą być mniej lub bardziej odporne na zakłócenia.

## ZAKRES PRACY

Zakres niniejszej pracy obejmuje kilka głównych aspektów, a mianowicie:

### 1. Przygotowanie środowiska badawczego:

- Konfiguracja niezbędnego oprogramowania i bibliotek do przetwarzania obrazów oraz uczenia maszynowego, w celu stworzenia przystępnego środowiska do prowadzenia badań.
- Ustalenie wykorzystywanych narzędzi.

### 2. Przygotowanie danych:

- Zebranie odpowiednich zbiorów danych zawierających obrazy zwierząt z różnorodnym tłem.
- Przeprowadzenie potrzebnego preprocessingu danych.

### 3. Wykorzystanie gotowych modeli klasyfikacyjnych:

- Wykorzystanie wcześniej już przetrenowanych 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 wybranych technik modyfikacji tła w celu stworzenia zestawów danych gotowych do przeprowadzania analiz klasyfikacji i możliwości porównania z wynikami na oryginalnych zdjęciach.

### 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 modyfikacji tła mogą być przydatne do poprawy jakości klasyfikacji. W pierwszej części rozdziału zostaną omówione narzędzia i oprogramowanie użyte do badań, konfiguracja sprzętowa oraz biblioteki i frameworki użyte 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. Opisany zostanie również zbiór danych wykorzystany w badaniach, jego źródło oraz struktura. Przedstawiony będzie również sposób przygotowania i przetwarzania danych przed dokonaniem badań. 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 rozdziału zakończy krótkie podsumowanie metodyki badań, co będzie dobrym zakończeniem przed kolejnym rozdziałem, w którym omówione zostaną wyniki.

## PRZYGOTOWANIE ŚRODOWISKA

Przygotowanie odpowiedniego środowiska badawczego jest kluczowym krokiem w realizacji każdego projektu, szczególnie gdy jest oparty 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 bogatym zasobom bibliotek i narzędzi wspomagających te dziedziny.

Główne biblioteki jakie wykorzystano w niniejszych badaniach to: 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, głównie do analizy wyników klasyfikacji zapisanych w formacie dataframe. Scikit-learn był wykorzystywany przy obliczeniach metryk i ocenie jakości modeli, a PIL (Python Imaging Library) umożliwiła manipulację obrazami i łatwe dokonywanie manipulacji tła. Biblioteki matplotlib i seaborn posłużyły do wizualizacji danych i wyników analiz, co pozwoliło na lepsze zrozumienie uzyskanych rezultatów oraz prezentację wyników w przystępnej i zrozumiałej formie.

Kluczowym elementem projektu były wstępnie przeszkolone modele głębokiego uczenia: ResNet, ConvNeXt oraz DeepLabv3. Modele te zostały zaimportowane z biblioteki torchvision, która jest częścią większego ekosystemu PyTorch. Torchvision dostarcza łatwy dostęp do przeróżnych wcześniej przetrenowanych modeli na dużych zbiorach danych, co umożliwia efektywne przeprowadzanie eksperymentów bez konieczności trenowania modeli od podstaw na własną rękę i pozwala skupić się na najważniejszych aspektach swoich badań.

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 przejryste prezentowanie procesu badawczego, testowanie i modyfikowanie kodu w czasie rzeczywistym oraz dokumentowanie każdego kroku analizy. Struktura takiego notatnika składa się z pojedynczych komórek, co jest bardzo wygodne, jeżeli specyfika naszego badania wymaga poszczególnych wydzielonych funkcjonalności oraz zawiera ciąg przyczynowo-skutkowy.

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

Odpowiednie przygotowanie środowiska z użyciem wymienionych narzędzi i bibliotek było ważnym etapem badania, posiadając dobrze przygotowane środowisko, przeprowadzanie badań zostaje procesem łatwo powtarzalnym oraz wzbogacanie kodu o kolejne funkcjonalności również nie będzie przynosić trudności.

## WYBRANE MODELE

W niniejszym projekcie zastosowano trzy wcześniej przetrenowane 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. Poniżej zostanie przedstawiona krótka charakterystyka wykorzystanych modeli, w celu lepszego zrozumienia kontekstu badania.

### ResNet

ResNet (Residual Network) został zaproponowany w 2015 roku, w którym wygrał wtedy edycję konkursku ImageNet ImageNet Large Scale Visual Recognition Challenge). Bardzo szybko stał się jednym z najpopularniejszych modeli w dziedzinie głębokiego uczenia, dzięki swoim szerokim zastosowaniom oraz swoją efektywnością. Główną innowacją ResNet było wprowadzenie residual learning poprzez zastosowanie skrótowych połączeń (skip connections). Pozwala to na efektywne trenowanie bardzo głębokich sieci, nawet o setkach warstw, jednocześnie rozwiązuje problem zanikania gradientu.

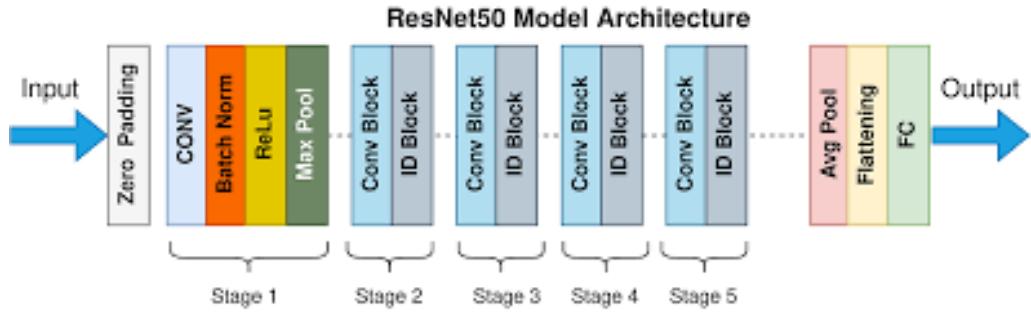
W tradycyjnych sieciach neuronowych, wraz ze wzrostem warst, wzrastał również problem zanikającego gradientu, co znacznie utrudniało efektywne trenowanie modeli. ResNet zaadresował ten problem poprzez wprowadzenie bezpośrednich połączeń skrótnowych, które umożliwiają przepływ gradientu bezpośrednio przez 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 skuteczne podejście pozwala na trenowanie bardzo głębokich sieci, które byłyby trudne do nauczenia przy użyciu tradycyjnych metod.



Rys. 1. Schemat architektury modelu ResNet50

ResNet został wybrany do tego projektu ze względu na swoją zdolność do efektywnego radzenia sobie z problemami klasyfikacji, oraz swoją popularność w dziedzinach sztucznej inteligencji.

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, po 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 informacje.

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 przeprowadzenie ostatecznej klasyfikacji. Wyjście z ostatniej warstwy całkowicie połączonej jest przekazywane do funkcji aktywacji softmax, aby uzyskać ostateczne prawdopodobieństwa predykowanych klas.

## ConvNeXt

ConvNeXt to model o nowoczesnej architekturze CNN, który został opracowany w celu integracji najlepszych praktyk z konwolucyjnych sieci neuronowych i nowoczesnych technik pochodzących od transformerów. ConvNeXt wykorzystuje bardziej złożone opera-

cje konwolucyjne oraz zaawansowane techniki normalizacji i optymalizacji, co pozwala na osiąganie bardzo dobrych 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 state-of-art, takimi jak ResNet.

### DeepLabv3

DeepLabv3 z kolei jest modelem wykorzystywanym do 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 segmentowanie obiektów 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 oraz uniwersalność użycia, 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 w porównaniu do nieco starszych technik. DeepLabv3, jako uniwersalny 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 modyfikacjami. Metryki te zostały wybrane ze względu na ich zdolność do dostarczania wartościowych informacji na temat wydajności modeli w różnych warunkach, sprawdzają się one idealnie do wstępnych analiz modeli oraz do oceny jak skutecznie modele radzą sobie z przedstawionym problemem klasyfikacji. Wykorzystane metryki zostaną w tej części krótko opisane. Metryki przeprowadzonych badań będą analizowane całościowo, jak również osobno dla każdej klasy, dla każdej różnej modyfikacji tła oraz dla różnych rozmiarów 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. Będzie to jedna z głównych analizowanych metryk. Każda klasa będzie posiadała taką samą ilość próbek, także problem interpretowalności tej metryki, jaki występuje przy danych niebalansowanych nie wystąpi.

Wzór na dokładność prezentuje się następująco:

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

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

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

### 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:

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

### 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:

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

### 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 jest pewny swoich decyzji oraz jak modyfikacje tła wpływają na pewność modeli.

## 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.

- **Zbior 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.
- **Zbior walidacyjny:** Składa się z 50 obrazów na klasę. Zbior walidacyjny jest używany do monitorowania wydajności modelu w trakcie treningu i do wczesnego wykrywania problemów takich jak nadmierne dopasowanie (overfitting).
- **Zbior testowy:** Zawiera 100 obrazów na klasę. Zbior 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. Inaczej mówiąc, 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 do generalizacji.

### Popularność i znaczenie ImageNet1k

ImageNet1k jest jednym z najczęściej używanych zestawów danych w badaniach nad różnymi problemami związanymi z głębokim uczeniem, co jest wynikiem jego dużej ilości zdjęć, różnorodności i realistycznego charakteru. Wiele przełomowych modeli, takich jak VGG czy ResNet, 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 łatwe porównywanie z innymi badaczami.

## **PLAN BADAŃ**

Niniejszy rozdział opisuje szczegółowy plan badań, które zostały przeprowadzone w celu zbadania wpływu tła na klasyfikację obrazów zwierząt. Poniżej szczegółowo 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. Przygotowano również wirtualne środowisko, pobrano dane z oficjalnej strony ImageNet.

### **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 samodzielnych 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.

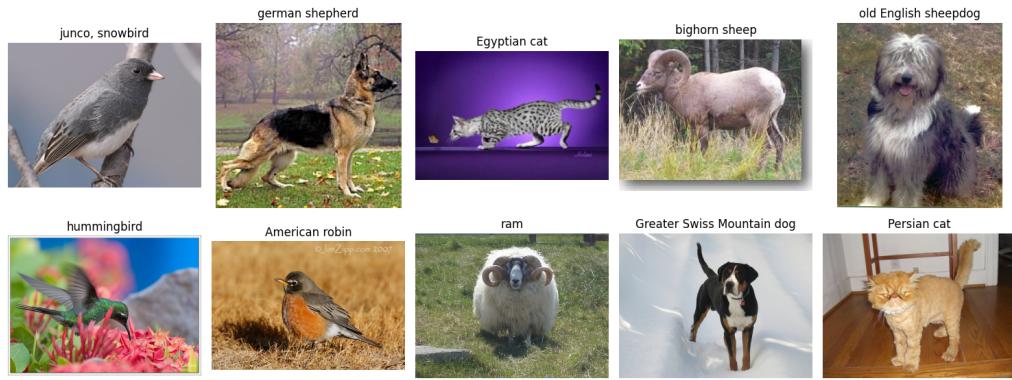
Skrypty są nazwane zgodnie ze swoją funkcjonalnością, co ułatwia nawigację i zrozumienie ich przeznaczenia. Każdy skrypt zawiera funkcje z dobrze opisanymi nazwami, a każda funkcja posiada docstringi, które szczegółowo wyjaśniają jej działanie.

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.

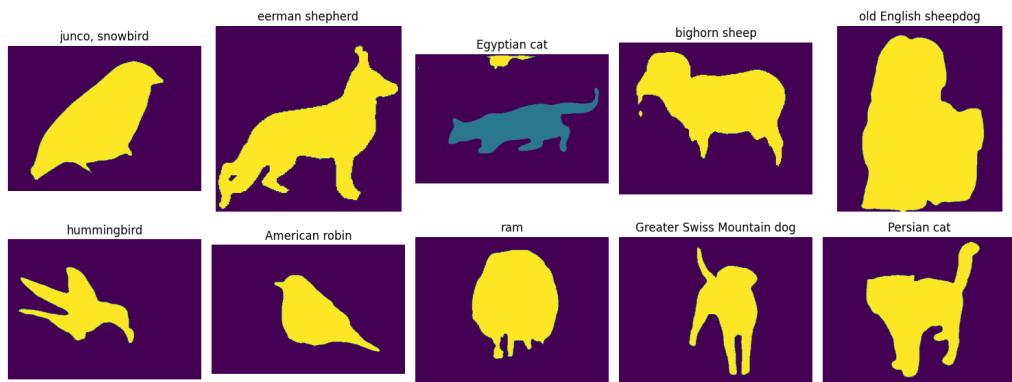
Link do repozytorium GitHub: [link do repozytorium](#)

### **Wybranie modeli do segmentacji i klasyfikacji**

Kolejnym ważnym krokiem było zaplanowanie wykorzystywanych modeli w badaniu. 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 oraz dwóch generacji modeli.



Rys. 2. Przykładowe oryginalne zdjęcia wybranych klas



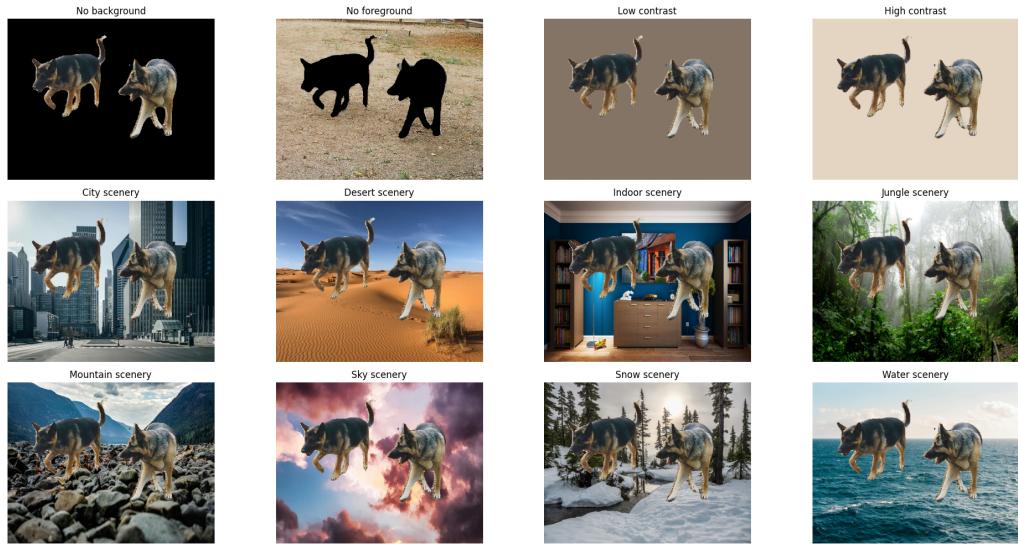
Rys. 3. Maski przykładowych wysegmentowanych obrazów

### **Wybranie klas zwierząt**

Ze względu na ograniczoną dostępność klas w modelu segmentacyjnym DeepLabv3, oraz braku zasobów do analizy wszystkich klas, do badań wybrano 10 zróżnicowanych klas zwierząt. Wybrane klasy były takie, które można skutecznie wysegmentować przy użyciu tego modelu. Każda klasa ma przynajmniej jedną zbliżoną do siebie, w celu stworzenia warunków badania sprzyjających popełnianiu błędów przez klasyfikatory, przykładowo wybrano trzy rasy psów. Przykładowe zdjęcia dla każdej wybranej klasy można zaobserwować na Rys. ??

### **Segmentacja obrazów**

Dla każdej z wybranych klas zwierząt wysegmentowano 1000 zdjęć ze zbioru treningowego za pomocą modelu DeepLabv3. Ponieważ na niektórych zdjęciach widniało więcej klas obiektów rozpoznawanych przez ten model, konieczne było zidentyfikowanie wartości grayscale dla pożąданiej 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. ??



Rys. 4. Przykładowe zdjęcie poddane modyfikacji

### 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 inne różne tła o innej scenerii, były to: miasto, pustynia, wnętrze domu, dżungla, góry, niebo, śnieg oraz woda. Ma to na celu stworzenie wariantów zdjęć w innych sceneriach co może być mylące dla klasyfikatorów, szczególnie przy klasyfikacji zwierząt, gdzie podobnie wyglądające gatunki mogą występować w różnych środowiskach.
- **Zastąpienie tła kolorem o niskim oraz wysokim kontraście do obiektu:** Kolejną modyfikacją było przeniesienie obiektów na tło o kolorze niskokontrastowym oraz wysokokontrastowym. W celu znalezienia tych kolorów, dla każdej klasy przeanalizowano każde zdjęcie i zebrano dominujące kolory dla danej klasy. Następnie przy pomocy specjalnego skryptu wyliczono kolory o niskim i wysokim kontraście do tych obiektów.

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

Modyfikacji dokonywano za pomocą wcześniej uzyskanych masek, dla każdego rodzaju modyfikacji stworzono osobny skrypt w celu zachowania obiektowego charakteru środowiska programistycznego.

### Dokonanie predykcji na wybranych modelach

Dla każdego zdjęcia dokonano predykcji dla dwóch wybranych modeli, za równo dla oryginalnych zdjęć oraz dla każdego typu modyfikacji. Co dawało dla jednej klasy

trzynaście tysięcy predykcji, dwanaście modyfikacji po 1000 zdjęć oraz 1000 oryginalnych zdjęć. Wyniki predykcji oraz wartości confidence score zapisano w pliku 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 znacząco 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 procentyli wielkości obiektu pozwala na zrozumienie, w jakim stopniu tło oddziałuje 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. ??

- Obliczenie powierzchni obiektu:** Dla każdego obrazu obliczono liczbę pikseli zajmowanych przez obiekt.
- Obliczenie powierzchni całkowitej obrazu:** Liczba pikseli całego obrazu.
- Obliczenie procentu powierzchni zajmowanej przez obiekt:** Procent powierzchni zajmowanej przez obiekt obliczono za pomocą wzoru:

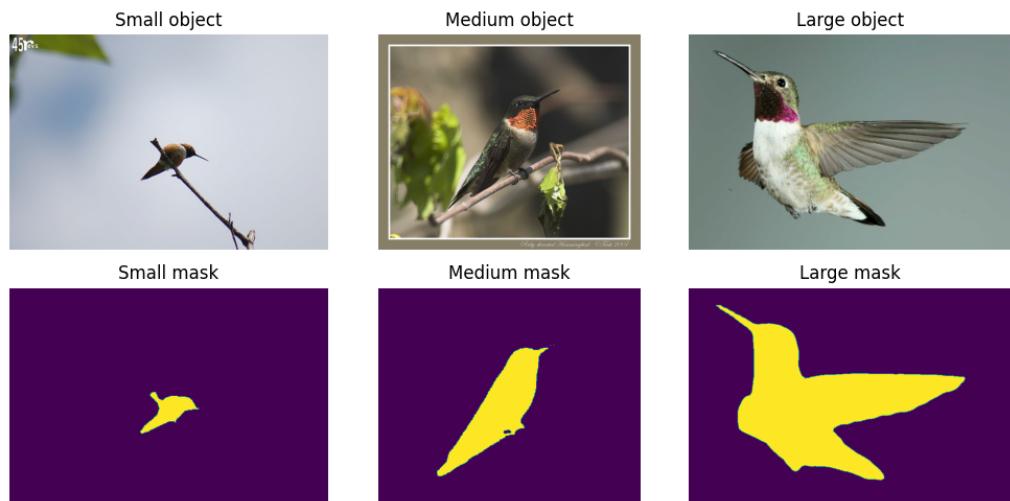
$$\text{Procent powierzchni zajmowanej przez obiekt} = \left( \frac{\text{Powierzchnia obiektu}}{\text{Powierzchnia całkowita obrazu}} \right) \times 100 \quad (4)$$

- Podział obiektów na percentile:** Obrazy posortowano według procentu powierzchni zajmowanej przez obiekt i podzielono na trzy grupy według wartości percentylu (0,33 oraz 0,66) i nadano każdemu zdjęciu odpowiednią kategorie (SMALL, MEDIUM albo LARGE)

### Analiza wyników

Wygenerowany plik csv załadowano do dataframe, następnie przygotowoną tą ramkę w celu dalszych analiz. Dodano informacje o prawdziwej klasie do każdego zdjęcia oraz informacje True, False w zależności od poprawności klasyfikacji dla każdej modyfikacji. Ułatwiło to dalsze generowanie statystyk i metryk. Analizy wyników dokonano ze względu na różne czynniki, wymienione poniżej:

- **Analiza ogólna wyników:** Ogólna wydajność modeli na całym zbiorze danych.



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

- **Analiza pod kątem wielkości obiektu:** Wydajność modeli w zależności od wielkości obiektu na zdjęciu.
- **Analiza pod kątem rodzaju modyfikacji:** Wydajność modeli w zależności od typu modyfikacji tła.
- **Analiza pod kątem klasy:** Wydajność modeli dla każdej klasy zwierząt osobno.

#### **Wnioski i dalsze kierunki rozwoju**

Kolejnym etapem badania było wyciągniecie wniosków na podstawie uzyskanych rezultatów oraz przedstawienie potencjalnych kierunków rozwoju projektu.

## BADANIA

Celem tego rozdziału jest omówienie przeprowadzonej analizy oraz przedstawienie wyników klasyfikacji obrazów zwierząt dla modeli ResNet i ConvNeXt. Analiza ta obejmuje porównanie skuteczności modeli w różnych scenariuszach modyfikacji tła oraz w zależności od wielkości obiektu na obrazie oraz w zależności od predykowanej klasy.

## WYNIKI OGÓLNE

Badania miały na celu zbadanie wpływu modyfikacji tła na skuteczność klasyfikacji obrazów za pomocą dwóch wybranych modeli. 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.

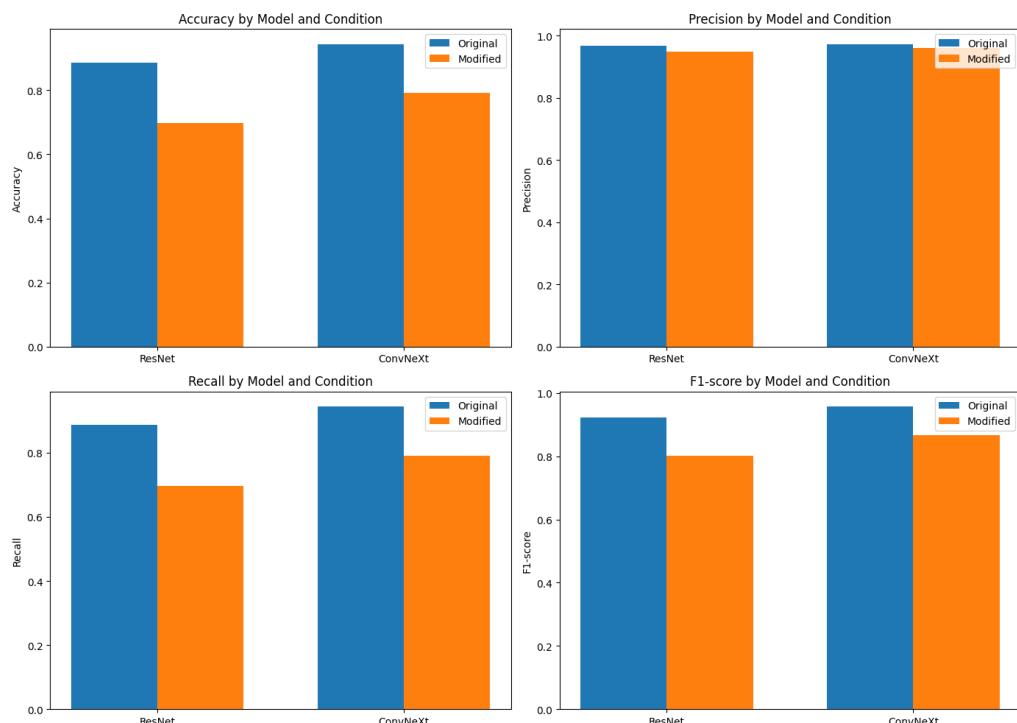
Dla modelu ResNet, na oryginalnych obrazach uzyskano Accuracy na poziomie 0.886500, Precision 0.967026, Recall 0.886500 i F1-score 0.922742. Po modyfikacji tła wartości te uległy istotnemu obniżeniu, osiągając odpowiednio 0.697018 dla Accuracy, 0.948539 dla Precision, 0.697018 dla Recall i 0.802350 dla 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. Ogólne wyniki można zaobserwować na ?? oraz ??

Analiza wyników wskazuje, że modyfikacja tła negatywnie wpływa na skuteczność obu modeli, jednak model o nowszej architekturze ConvNeXt wykazuje większą odporność na zmiany tła w porównaniu do modelu 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. Różnica w dokładności dla danych zmodyfikowanych wynosi aż 10 punktów procentowych, przy dokładności około 79 procent dla convnext oraz około 69 procent dla resnet. Model z nowszą architekturą poradził sobie lepiej.

Wyniki badań również wskazują, że pomimo modyfikacji tła, Precision dla obu modeli (ResNet i ConvNeXt) uległa jedynie niewielkiemu spadkowi. Mały spadek Precision w obu przypadkach sugeruje, że oba modele nadal skutecznie identyfikują prawdziwie pozytywne przypadki po modyfikacji tła.

<b>Model</b>	<b>Type</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
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 tłaach

<b>Model</b>	<b>Type</b>	<b>Average</b>	<b>Average correct</b>	<b>Average incorrect</b>
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

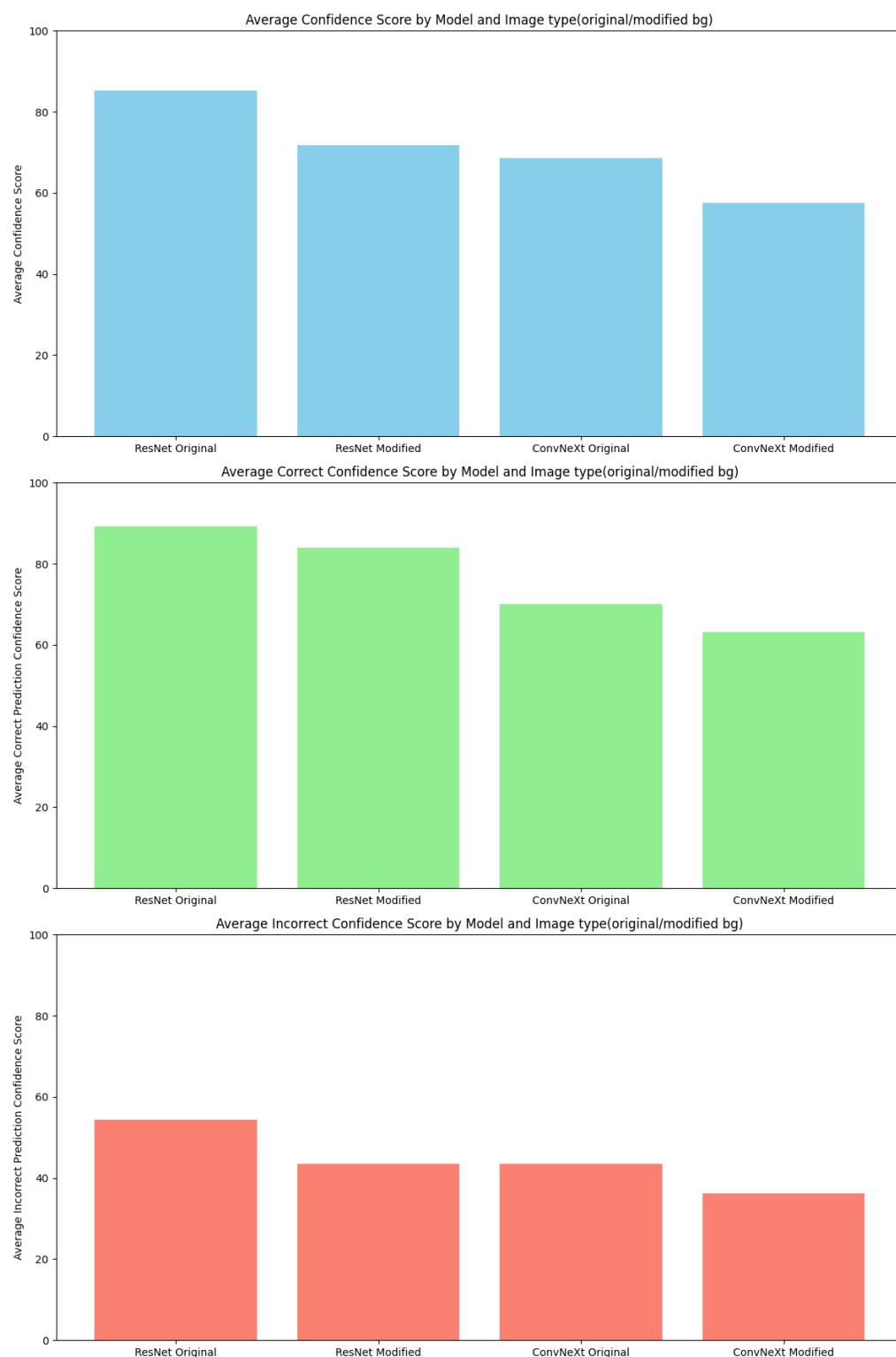
W badaniach obliczono również średnie wartości confidence scores w celu sprawdzenia pewności decyzji modeli. Wartości te obejmują ogólną średnią confidence score, a także średnie confidence scores dla poprawnych i niepoprawnych klasyfikacji. Dla modelu ResNet na oryginalnych obrazach średnia confidence 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. Wyniki znajdują się w tabeli ?? oraz wizualizacja tych wyników na ??

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.

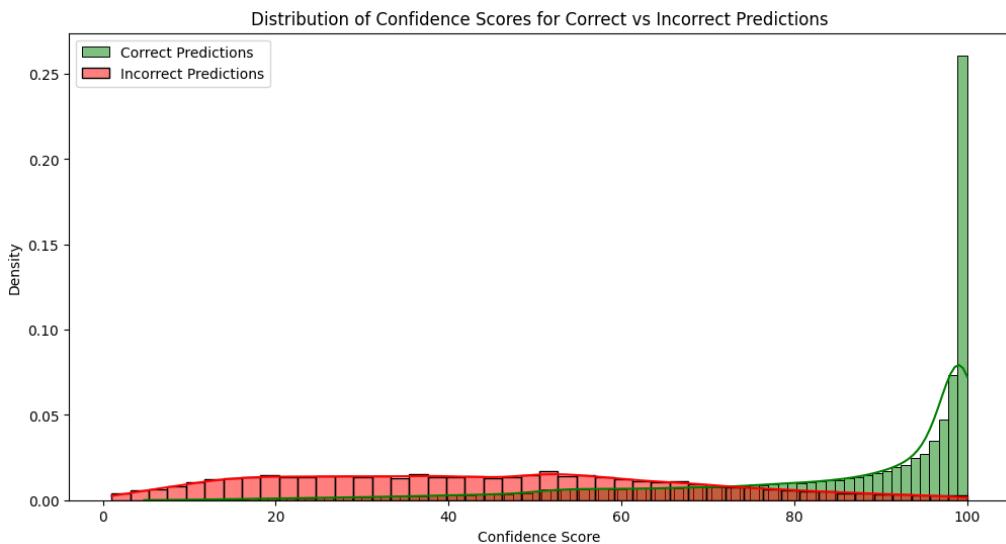
Wyniki te wskazują na znaczący spadek średnich confidence scores dla obu modeli po modyfikacji tła. Średnie confidence scores dla poprawnych klasyfikacji są wyższe niż dla niepoprawnych w obu przypadkach, co sugeruje, że modele są bardziej pewne swoich poprawnych klasyfikacji. Jednak modyfikacja tła powoduje ogólny średni spadek pewności modeli.

Analiza wartości confidence scores dla ConvNeXt wskazuje na większy spadek w porównaniu do modelu ResNet. Ciekawe jest, że mimo większej dokładności predykcji, model convnext posiada znacznie niższe wartości confidence score.

Podsumowując, modyfikacja tła ma w tym przypadku wyraźny wpływ na zmniejszenie pewności klasyfikacji obrazów przez oba modele. Chociaż oba modele wykazują wysoką pewność przy poprawnych klasyfikacjach, modyfikacja tła powoduje ogólny spadek tych wartości, z bardziej zauważalnym spadkiem w przypadku modelu ConvNeXt.

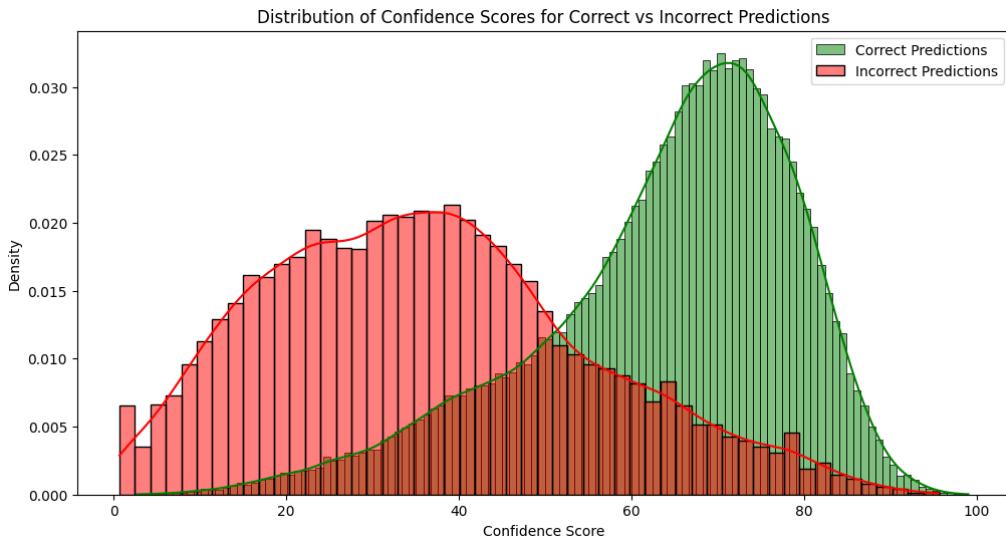


Rys. 7. Średnie wartości dla confidence scores



Rys. 8. Dystrybucja confidence score dla ResNet

Na podstawie dystrybucji confidence scores widocznych na rysunkach ?? oraz ?? dla modeli ResNet i ConvNeXt można wyciągnąć kilka istotnych wniosków. Model ResNet wykazuje wysoką pewność dla poprawnych klasyfikacji, 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, większość poprawnych klasyfikacji znajduje się w przedziale 70-80 i dla błędnych w przedziale 30-40. Mimo mniejszej pewności model convnext odnotował mniejsze spadki pewności w porównaniu z pewnością na zdjęciach oryginalnych. Można z tego wnioskować, że modyfikacje tła mniej wpływają na pewność modelu convnext niż na resnet.



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	0.789338	0.960967	0.789338	0.865607
Medium	Original	0.897879	0.963620	0.897879	0.927600
Medium	Modified	0.762727	0.943926	0.762727	0.841553
Small	Original	0.884545	0.963823	0.884545	0.918217
Small	Modified	0.552828	0.935458	0.552828	0.686206

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

### WYNIKI WZGLEDEM KATEGORII WIELKOŚCI OBIEKTU

Kolejną częścią analizy było zbadanie metryk modeli w kontekście różnych rozmiarów obiektu na zdjęciu. Analiza wpływu wielkości obiektu na skuteczność klasyfikacji obrazów jest istotnym elementem badań, ponieważ różne rozmiary obiektów mogą znacząco wpływać na wydajność modeli głębokiego uczenia. Zdjęcia z obiektemi o różnych wielkościach mogą być różnie traktowane przez modele klasyfikacyjne ze względu na różną ilość "zakłóceń" jakim jest tło. Dlatego też, zrozumienie, jak zmienia się dokładność 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 0.789338, 0.960967, 0.789338 i 0.865607. 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 0.762727, 0.943926, 0.762727 i 0.841553. 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	0.846397	0.970483	0.846397	0.902546
Medium	Original	0.943333	0.970287	0.943333	0.955907
Medium	Modified	0.840682	0.959570	0.840682	0.892625
Small	Original	0.954545	0.972766	0.954545	0.961567
Small	Modified	0.705480	0.950400	0.705480	0.804351

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 0.552828, 0.935458, 0.552828 i 0.686206.

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 0.846397, 0.970483, 0.846397 i 0.902546. 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 0.840682, 0.959570, 0.840682 i 0.892625. 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 0.705480, 0.950400, 0.705480 i 0.804351.

Analiza wyników pokazuje, że wielkość obiektu ma znaczący wpływ na jakość klasyfikacji, mianowicie im mniejszy obiekt tym większe większe pogorszenie metryk modeli. Oba modele uzyskały największe pogorszenie metryk dla najmniejszych obiektów oraz okazały się najbardziej odporne przy największych obiektach, co jest dosyć logicznym procesem, ponieważ przy dużych obiektach jest mniej zakłóceń na zdjęciu w postaci tła. Model convnext okazał się bardziej odporny, w każdej kategorii wielkości odnotował mniejsze spadki niż resnet.

Wyniki te podkreślają znaczenie rozważania wielkości obiektów przy problemach klasyfikacyjnych. 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.

Dla modelu ResNet, średnie confidence scores wynoszą 76.790632 dla dużych obiektów, 76.564846 dla obiektów średniej wielkości i 65.356645 dla małych obiektów. Dla poprawnych klasyfikacji confidence scores są odpowiednio wyższe, wynosząc 85.228727, 85.185492 i 82.693992. Dla niepoprawnych klasyfikacji wartości te są znacznie niższe: 43.843488, 47.188400 i 42.576619.

W przypadku modelu ConvNeXt, średnie confidence scores są niższe i wynoszą

<b>Object Size</b>	<b>Average Score</b>	<b>Correct Score</b>	<b>Incorrect Score</b>
Large	76.790632	85.228727	43.843488
Medium	76.564846	85.185492	47.188400
Small	65.356645	82.693992	41.576619

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

<b>Object Size</b>	<b>Average Score</b>	<b>Correct Score</b>	<b>Incorrect Score</b>
Large	64.332577	69.300610	35.502391
Medium	60.275503	64.107149	38.802715
Small	50.455171	56.473165	34.618265

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

64.332577 dla dużych obiektów, 60.275503 dla obiektów średniej wielkości i 50.455171 dla małych obiektów. Confidence scores dla poprawnych klasyfikacji wynoszą 69.300610, 64.107149 i 56.473165, a dla niepoprawnych klasyfikacji są to 35.502391, 38.802715 i 34.618265.

Wraz ze wzrostem wielkości obiektu na badanych zdjęcia wzrasta pewność obu modeli. Największy spadek pewności występuje dla małych obiektów.

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. Obliczono również podstawowe metryki, pewności predykcji oraz przedstawiono macierze korelacji dla obu modeli.

Analiza wyników pokazuje, że typ modyfikacji tła ma znaczący wpływ na skuteczność klasyfikacji obrazów oraz wartości confidence scores dla obu modeli. Oba modele osiągały najlepsze wyniki głównie na modyfikacjach, które były jednokolorowe, a mianowicie na "no\_background", "high\_contrast" oraz "low\_contrast". W pierwszym przypadku tło było całkowicie czarne, a w pozostałych kolor był dobrany na podstawie dominujących kolorów w obiekcie danej klasy. Najlepsze wyniki oba modele uzyskały na tle o kolorze niskokontrastowym do obiektu. Zarówno ResNet jak i ConvNext osiągneły nieco gorsze wyniki dla scenerii wodnej, wnętrza domu oraz scenerii górzystej. Dokładne wartości można zaobserwować w tabelach ?? oraz ?. Wyższa dokładność na modyfikacjach jednokolorowych prawdopodobnie wynika z powodu mniejszej liczby elementów na zdjęciach, które mogą być mylące dla modeli. Oba modele poardziły sobie fatalnie w sytuacji gdy ze zdjęcia

<b>Modification Type</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Desert	0.7582	0.956658	0.7582	0.844717
Low Contrast	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 Background	0.7629	0.948944	0.7629	0.843935
High Contrast	0.7757	0.952540	0.7757	0.852606
No Foreground	0.1935	0.857598	0.1935	0.260654
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

<b>Modification Type</b>	<b>Average Score</b>	<b>Correct Score</b>	<b>Incorrect Score</b>
Desert	75.954396	84.304869	49.770241
Low Contrast	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 Background	74.429546	85.308617	39.424725
High Contrast	77.275950	86.369482	45.827655
No Foreground	41.084428	67.091235	34.844729
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 dla modelu ResNet według typu modyfikacji

<b>Modification Type</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Desert	0.8234	0.965246	0.8234	0.885689
Low Contrast	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 Background	0.8765	0.961364	0.8765	0.915996
High Contrast	0.8707	0.964345	0.8707	0.913518
No Foreground	0.2572	0.921080	0.2572	0.343413
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

<b>Modification Type</b>	<b>Average Score</b>	<b>Correct Score</b>	<b>Incorrect Score</b>
Desert	61.750625	65.260020	45.388021
Low Contrast	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 Background	57.877617	62.096968	27.932186
High Contrast	57.146265	61.220278	29.712060
No Foreground	38.674228	56.533386	32.490362
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 scores dla modelu ConvNeXt według typu modyfikacji

został usunięty obiekt i pozostawione zostało tło, jednak model ConvNext był bardziej odporny i lepiej radził sobie z analizowaniem samego pozostawionego kształtu obiektu. Dla modelu ConvNeXt jedyną modyfikacją (poza usunięciem obiektu), na której uzyskał wartość dokładności poniżej 80 % jest sceneria z wnętrzem domu. Wysoką wartość accuracy uzyskała również sceneria dżungli dla obu modeli, może to wynikać z faktu iż ta sceneria posiada dużo zielonych roślin, które często mogą występować przy rzeczywistych zdjęciach wybranych zwierząt.

Jeżeli chodzi o pewność modeli co do swoich decyzji to, dla modelu ResNet oscylowała w okolicach 70-75 % dla wszystkich modyfikacji, poza scenariuszem usunięcia obiektu ze zdjęcia gdzie wyniosła 40 %. Model ConvNext posiada mniejszą pewność, która wynosi około 60 % oraz 40 % dla sytuacji bez obiektu.

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. Tła o mniejszej ilości rozpraszających elementów lub te przypominające naturalne warunki występowania zwierząt na rzeczywistych zdjęciach radziły sobie lepiej z problemem klasyfikacji.

## **WYNIKI WZGLEDEM KLAS**

W niniejszym podrozdziale przedstawiono wyniki badań dotyczących wpływu tła na skuteczność klasyfikacji względem już konkretnych klas. W tych badaniach skupiono się na analizie średniej dokładności oraz średniej pewności względem typu modyfikacji dla każdej klasy osobno. Sprawdzano również jakie błędy są najczęściej podejmowane, czyli z



Rys. 10. Po lewej "ram, tup" (true\_class), natomiast po prawej "bighorn sheep"

jaką klasą jest najczęściej mylona. Badania obejmowały analizę dziesięciu różnych klas obiektów, takich jak "ram, tup", "Greater Swiss Mountain dog", "hummingbird", "Egyptian cat", "bighorn sheep", "Old English sheepdog", "Persian cat", "junco, snowbird", "German shepherd" i "American robin". Klasy te reprezentują trzy rasy psów, dwie rasy kotów, trzy rasy ptaków oraz dwie rasy owiec.

### WNIOSKI DLA KLASY 348 (RAM, TUP)

Na podstawie wyników dwóch modeli przedstawionych w ?? i ??, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 348 (ram, tup) przy użyciu modeli ResNet i ConvNeXt:

#### Model ResNet

Model ResNet osiągnął dosyć niską dokładność 0.781 na oryginalnych obrazach, co wskazuje na dosyć problematyczną klasę. Średnia pewność wynosiła 80.888742, co jest stosunkowo wysoką wartością, potwierdzającą pewność modelu przy klasyfikacji. Najczęściej myloną klasą była klasa "Bighorn sheep" dla wszystkich typów modyfikacji. Te dwie klasy są niemal identyczne nawet dla ludzkiego oka, co można zaobserwować na ??.

Dla większości modyfikacji wartości accuracy się obniżyły. Dla jednego typu modyfikacji udało się polepszyć metrykę dokładności o około 5 punktów procentowych, a mianowicie dla scenerii dżungli. Największe spadki uzyskano dla scenerii górzystych i śnieżnych, dla których znacznie wzrosła liczba przypadków gdzie pomyłono się z klasą "bighorn sheep". Podejrzewam, że tereny górzyste czy śnieżne są miejscem, w którym klasa "bighorn sheep" znajduje się częściej niż przewidywana klasa "ram, tup", natomiast sceneria dżungli pomaga modelowi rozróżnić te dwie klasy, gdzie liczba pomyłek jest znacznie niższa.

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.781	80.888742	Bighorn, bighorn sheep (194)
Desert	0.662	73.618091	Bighorn, bighorn sheep (212)
Low Contrast	0.637	72.533478	Bighorn, bighorn sheep (241)
City	0.691	71.055787	Bighorn, bighorn sheep (175)
Sky	0.752	74.196770	Bighorn, bighorn sheep (125)
Jungle	0.825	76.201070	Bighorn, bighorn sheep (84)
No Background	0.668	71.519892	Bighorn, bighorn sheep (213)
High Contrast	0.735	73.587178	Bighorn, bighorn sheep (163)
No Foreground	0.136	38.340821	American black bear, black bear (137)
Water	0.780	72.729140	Bighorn, bighorn sheep (97)
Snow	0.601	76.547279	Bighorn, bighorn sheep (318)
Indoor	0.683	71.596914	Bighorn, bighorn sheep (132)
Mountain	0.613	74.327448	Bighorn, bighorn sheep (285)

Tabela 11. Metrics and Confidence Scores for Class 348 (ram, tup) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.801	64.573134	Bighorn, bighorn sheep (191)
Desert	0.589	59.293923	Bighorn, bighorn sheep (305)
Low Contrast	0.752	57.574512	Bighorn, bighorn sheep (211)
City	0.604	57.563126	Bighorn, bighorn sheep (311)
Sky	0.717	57.364283	Bighorn, bighorn sheep (204)
Jungle	0.747	60.548807	Bighorn, bighorn sheep (187)
No Background	0.703	56.948889	Bighorn, bighorn sheep (236)
High Contrast	0.706	56.227286	Bighorn, bighorn sheep (244)
No Foreground	0.274	36.891416	American black bear, black bear (151)
Water	0.745	57.906020	Bighorn, bighorn sheep (202)
Snow	0.580	59.995922	Bighorn, bighorn sheep (371)
Indoor	0.556	59.476196	Bighorn, bighorn sheep (288)
Mountain	0.588	61.288639	Bighorn, bighorn sheep (345)

Tabela 12. Metrics and Confidence Scores for Class 348 (ram, tup) - ConvNeXt

### Model ConvNeXt

Model ConvNeXt osiągnął nieco wyższą dokładność dla oryginalnych zdjęć niż ResNet, lecz nie udało się uzyskać lepszych metryk dla żadnej modyfikacji. Spośród modyfikacji wyniki również były najlepsze dla scenerii dżungli oraz wody, podobnie jak dla wcześniejszego modelu. Relatywnie wysokie wartości uzyskano także dla tła o niskim kontraste. Ten model zdecydowanie był bardziej wrażliwy na modyfikację tła, ponieważ w większej liczbie kategorii uzyskał znacznie gorsze wyniki. Więcej typów modyfikacji przyczyniło się do pomyłki z klasą "bighorn sheep".



Rys. 11. Od lewej "Greater Swiss Mountain dog" (true\_class), "Appenzeller", "EntleBucher"

## WNIOSKI DLA KLASY 238 (GREATER SWISS MOUNTAIN DOG)

Na podstawie wyników dwóch modeli przedstawionych w ?? i ??, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 238 (Greater Swiss Mountain dog) przy użyciu modeli ResNet i ConvNeXt:

### Model ResNet

Dla modelu ResNet udało się minimalnie polepszyć wyniki klasyfikacji dla trzech scenerii: pustynnej, śnieżnej oraz wnętrza domu, dla reszty modyfikacji wyniki były gorsze. Model najczęściej mylił się z dwoma klasami, w zależności od typu modyfikacji. Wszystkie trzy klasy są bardzo do siebie podobne co można zaobserwować na rysunku ??

### Model ConvNeXt

Model o nowszej architekturze znacznie lepiej poradził sobie w przypadku tej klasy. Wynik dokładności polepszył się o około 15 punktów procentowych. Jeżeli chodzi o modyfikacje, to dla wszystkich uzyskano gorsze wyniki, szczególnie dla scenerii nieba oraz modyfikacji o wysokim kontraście, gdzie metryki spadły poniżej 80%. Podobnie jak wcześniej najczęstsze pomyłki dzieliły się między podobne rasy psów Appenzeller oraz EntleBucher. Ciekawym zjawiskiem jest najczęstsza pomyłka tego modelu dla scenerii wnętrza domu, gdzie model najczęściej mylił się klasyfikując zdjęcie jako półkę z książkami. Dowodzi to że różne niechciane obiekty znajdujące się w tle mogą zakłócać proces klasyfikacji.

## WNIOSKI DLA KLASY 94 (HUMMINGBIRD)

Na podstawie wyników dwóch modeli przedstawionych w ?? i ??, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 94 (Hummingbird) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.746	69.254460	EntleBucher (83)
Desert	0.751	71.670600	EntleBucher (71)
Low Contrast	0.628	66.772043	EntleBucher (168)
City	0.720	70.455594	Appenzeller (79)
Sky	0.693	67.656871	EntleBucher (103)
Jungle	0.680	66.612556	EntleBucher (99)
No Background	0.675	66.545953	EntleBucher (114)
High Contrast	0.632	65.697515	Appenzeller (118)
No Foreground	0.009	29.458587	American black bear, black bear (69)
Water	0.679	65.613398	Appenzeller (97)
Snow	0.748	71.034704	EntleBucher (96)
Indoor	0.750	71.739733	Appenzeller (57)
Mountain	0.605	63.748537	EntleBucher (165)

Tabela 13. Metrics and Confidence Scores for Class 238 (Greater Swiss Mountain dog) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.918	67.297499	EntleBucher (30)
Desert	0.812	53.915638	EntleBucher (66)
Low Contrast	0.857	53.678023	Appenzeller (48)
City	0.845	55.105653	Appenzeller (35)
Sky	0.784	52.714931	EntleBucher (69)
Jungle	0.853	57.766105	Appenzeller (48)
No Background	0.867	53.964683	EntleBucher (37)
High Contrast	0.790	49.376622	Appenzeller (80)
No Foreground	0.036	31.021854	Labrador retriever (215)
Water	0.868	55.700216	Appenzeller (37)
Snow	0.893	61.105172	EntleBucher (33)
Indoor	0.815	61.047711	bookcase (52)
Mountain	0.850	61.289471	Appenzeller (54)

Tabela 14. Metrics and Confidence Scores for Class 238 (Greater Swiss Mountain dog) - ConvNeXt



Rys. 12. Od lewej "hummingbird", "lakeshore", "seashore"

### **Model ResNet**

Model ResNet okazał się bardzo podatny na niektóre typy modyfikacji, szczególnie dla scenerii wodnych, śnieżnych, górzystych oraz wnętrza domu. Dokładność dla tych modyfikacji spadła do około 50% z poziomu oryginalnej wartości 96%. Ciekawym zjawiskiem było częste mylenie zdjęć kolibra z krajobrazem "seashore" oraz "lakeshore" dla modyfikacji o scenerii pustynnej oraz górzystej. W tych dwóch przypadkach porządzany obiekt jakim był koliber, został potraktowany przez model jako element większej scenerii jaką są krajobrazy. Na rysunku ?? można zaobserwować przykładowe zdjęcie po modyfikacji "desert", oraz przykładowe zdjęcia z klasy "lakeshore" oraz "seashore". Jeżeli chodzi o scenerię wodną model aż w 15% przypadków pomylił się z klasą "Albatross", która znacznie różni się wyglądem od klasy "hummingbird", przypominającej wyglądem mewę. Jest to ptak wodny, fotografowany często z tłem wody. Pokazuję to jaką istotną rolę pełni tło w procesie klasyfikacji. Mimo wyraźnych różnic w wyglądzie ptaków udało się nakłonić model do pomyłek poprzez prostą zmianę tła na wodne.

### **Model ConvNeXt**

Model ConvNeXt okazał się zdecydowanie bardziej odporny na modyfikację w kontekście tej klasy. Ciekawym zjawiskiem było częste (15%) klasyfikowanie ptaka jako wielbłąda dla scenerii pustynnych, co ponownie pokazuje istotny wpływ tła na klasyfikację. Model również traktował inne obiekty na zdjęciach jako te porządzane, np. dla scenerii miejskiej często klasyfikował zdjęcia tej klasy jako sygnalizacje świetlne, co może wynikać z wysokiego prawdopodobieństwa, że model podczas uczenia widział zdjęcia sygnalizacji świetlnej, na których również występowały ptaki.

### **WNIOSKI DLA KLASY 285 (EGYPTIAN CAT)**

Na podstawie wyników dwóch modeli przedstawionych w ?? i ??, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 285 (Egyptian cat) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.963	95.023629	Jacamar (9)
Desert	0.677	73.851502	Seashore, coast (195)
Low Contrast	0.836	83.892216	Kite (31)
City	0.649	63.584482	Flagpole, flagstaff (34)
Sky	0.789	81.238139	Volcano (48)
Jungle	0.796	80.272471	Greenhouse, nursery (53)
No Background	0.816	80.511866	Vine snake (21)
High Contrast	0.782	79.808579	Kite (47)
No Foreground	0.744	69.953177	Vulture (20)
Water	0.525	60.158964	Albatross, mollymawk (145)
Snow	0.543	63.296984	Snowmobile (151)
Indoor	0.565	66.628743	File, file cabinet (186)
Mountain	0.456	54.866228	Lakeside, lakeshore (121)

Tabela 15. Metrics and Confidence Scores for Class 94 (Hummingbird) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.995	72.907262	Jacamar (3)
Desert	0.805	63.267602	Arabian camel, dromedary (156)
Low Contrast	0.928	57.850450	Matchstick (14)
City	0.826	56.319503	Traffic light, traffic signal (110)
Sky	0.848	57.166515	Parachute, chute (73)
Jungle	0.859	61.211524	Cliff, drop (73)
No Background	0.933	58.181061	Matchstick (19)
High Contrast	0.911	56.167147	Kite (27)
No Foreground	0.868	67.694296	Jacamar (30)
Water	0.821	56.428329	Albatross, mollymawk (71)
Snow	0.801	60.554314	Lakeside, lakeshore (51)
Indoor	0.718	57.200175	File, file cabinet (123)
Mountain	0.774	59.565332	Valley, vale (73)

Tabela 16. Metrics and Confidence Scores for Class 94 (Hummingbird) - ConvNeXt



Rys. 13.

### **Model ResNet**

Dla 5 typów modyfikacji uzyskano lepsze wyniki niż dla oryginalnych zdjęć. W przypadku pozostawienia tła czarnego poprawiono dokładność aż o 7 punktów procentowych, z 78% do 85%. Najczęściej mylioną klasą, prawie dla każdej modyfikacji była klasa "Tabby cat", która podobnie jak we wcześniejszym przypadku jest bardzo podobna do klasy przewidywanej. Ciekawym zauważonym zjawiskiem była pomyłka modelu dla scenerii śnieżnej oraz dżungli, gdzie pomylono się na klasę "Tiger cat". Po analizie dostępnych zdjęć treningowych dla klasy tiger cat zauważylem, że w ich skład wchodzą zarówno zdjęcia przedstawiające kota domowego jak i zdjęcia tygrysa. Przykłady zdjęć widnieją na obrazku ??

### **Model ConvNeXt**

### **WNIOSKI DLA KLASY 349 (BIGHORN SHEEP)**

Na podstawie wyników przedstawionych w Tabeli X i Tabeli Y, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 349 (Bighorn sheep) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.782	73.471966	Tabby, tabby cat (132)
Desert	0.825	76.229713	Tabby, tabby cat (90)
Low Contrast	0.821	78.051673	Tabby, tabby cat (123)
City	0.763	71.972776	Tabby, tabby cat (136)
Sky	0.766	71.577528	Tabby, tabby cat (119)
Jungle	0.698	70.001732	Tiger cat (136)
No Background	0.847	76.844117	Tabby, tabby cat (85)
High Contrast	0.794	77.981898	Tabby, tabby cat (146)
No Foreground	0.260	33.898078	Quilt, comforter (56)
Water	0.846	75.972695	Tabby, tabby cat (57)
Snow	0.647	67.339146	Tiger cat (121)
Indoor	0.743	74.809806	Tabby, tabby cat (128)
Mountain	0.730	67.302426	Tabby, tabby cat (127)

Tabela 17. Metrics and Confidence Scores for Class 285 (Egyptian cat) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.875	67.470651	Tabby, tabby cat (83)
Desert	0.840	64.843979	Tabby, tabby cat (96)
Low Contrast	0.842	60.234300	Tabby, tabby cat (112)
City	0.788	58.385278	Tabby, tabby cat (134)
Sky	0.783	55.255898	Tabby, tabby cat (126)
Jungle	0.816	59.110254	Tabby, tabby cat (113)
No Background	0.872	63.906281	Tabby, tabby cat (89)
High Contrast	0.816	60.301645	Tabby, tabby cat (143)
No Foreground	0.301	36.188940	Schipperke (157)
Water	0.862	58.187896	Tabby, tabby cat (87)
Snow	0.823	59.457130	Tabby, tabby cat (114)
Indoor	0.724	62.997077	Bookcase (100)
Mountain	0.829	58.399331	Tabby, tabby cat (87)

Tabela 18. Metrics and Confidence Scores for Class 285 (Egyptian cat) - ConvNeXt

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.857	76.879462	Ram, tup (124)
Desert	0.644	65.576020	Seashore, coast (125)
Low Contrast	0.707	62.688885	Ram, tup (103)
City	0.583	57.522109	Ram, tup (171)
Sky	0.501	61.268969	Ram, tup (273)
Jungle	0.471	59.756806	Ram, tup (351)
No Background	0.635	57.550804	Ram, tup (145)
High Contrast	0.583	59.381386	Ram, tup (218)
No Foreground	0.300	46.576951	American black bear, black bear (128)
Water	0.457	58.803368	Ram, tup (329)
Snow	0.783	72.009580	Alp (73)
Indoor	0.467	60.394391	Ram, tup (220)
Mountain	0.725	66.198482	Ram, tup (77)

Tabela 19. Metrics and Confidence Scores for Class 349 (Bighorn sheep) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.908	56.144302	Ram, tup (89)
Desert	0.739	53.343471	Arabian camel, dromedary (219)
Low Contrast	0.792	43.160569	Ram, tup (110)
City	0.814	48.102590	Traffic light, traffic signal (103)
Sky	0.660	44.070363	Ram, tup (151)
Jungle	0.709	47.144147	Ram, tup (133)
No Background	0.776	42.662816	Ram, tup (88)
High Contrast	0.786	42.740627	Ram, tup (89)
No Foreground	0.369	41.866028	American black bear, black bear (171)
Water	0.714	44.812545	Ram, tup (147)
Snow	0.878	55.639899	Lakeside, lakeshore (79)
Indoor	0.715	51.461317	File, file cabinet (155)
Mountain	0.826	53.093244	Valley, vale (89)

Tabela 20. Metrics and Confidence Scores for Class 349 (Bighorn sheep) - ConvNeXt

### Model ResNet

### Model ConvNeXt

## WNIOSKI DLA KLASY 229 (OLD ENGLISH SHEEPDOG)

Na podstawie wyników przedstawionych w Tabeli X i Tabeli Y, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 229 (Old English sheepdog) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.941	88.632230	Tibetan terrier, chrysanthemum dog (9)
Desert	0.801	80.359347	Seashore, coast (64)
Low Contrast	0.785	77.904651	Sealyham terrier, Sealyham (39)
City	0.765	74.680369	Cab, hack (21)
Sky	0.718	74.347288	Volcano (63)
Jungle	0.782	75.977385	Greenhouse, nursery (32)
No Background	0.709	71.070938	Sealyham terrier, Sealyham (56)
High Contrast	0.760	75.577558	Sealyham terrier, Sealyham (47)
No Foreground	0.071	31.629389	American black bear, black bear (69)
Water	0.758	76.056336	Albatross, mollymawk (78)
Snow	0.803	80.582623	Alp (44)
Indoor	0.824	84.001364	File, file cabinet (60)
Mountain	0.715	70.130826	Valley, vale (44)

Tabela 21. Metrics and Confidence Scores for Class 229 (Old English sheepdog) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.987	74.735155	Collie (2)
Desert	0.871	66.845438	Arabian camel, dromedary (96)
Low Contrast	0.910	60.380093	Matchstick (25)
City	0.882	63.133430	Traffic light, traffic signal (52)
Sky	0.824	56.268607	Volcano (77)
Jungle	0.890	64.798835	Cliff, drop (43)
No Background	0.869	59.633430	Matchstick (26)
High Contrast	0.891	59.879026	Matchstick (28)
No Foreground	0.109	32.475152	Newfoundland, Newfoundland dog (188)
Water	0.879	60.280609	Grey whale, gray whale (32)
Snow	0.896	64.967973	Lakeside, lakeshore (38)
Indoor	0.867	69.414455	Bookcase (58)
Mountain	0.856	62.675450	Valley, vale (44)

Tabela 22. Metrics and Confidence Scores for Class 229 (Old English sheepdog) - ConvNeXt

### Model ResNet

### Model ConvNeXt

## WNIOSKI DLA KLASY 283 (PERSIAN CAT)

Na podstawie wyników przedstawionych w Tabeli X i Tabeli Y, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 283 (Persian cat) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.964	93.028729	Angora, Angora rabbit (9)
Desert	0.848	83.349976	Seashore, coast (41)
Low Contrast	0.855	83.713232	Tabby, tabby cat (26)
City	0.877	84.163735	Tabby, tabby cat (15)
Sky	0.821	82.203083	Volcano (52)
Jungle	0.831	81.888781	Greenhouse, nursery (25)
No Background	0.844	81.940165	Egyptian cat (19)
High Contrast	0.853	84.050749	Tabby, tabby cat (30)
No Foreground	0.025	23.772934	Quilt, comforter (70)
Water	0.821	79.093091	Albatross, mollymawk (23)
Snow	0.829	82.849727	Alp (31)
Indoor	0.879	87.654082	File, file cabinet (41)
Mountain	0.794	77.893145	Valley, vale (25)

Tabela 23. Metrics and Confidence Scores for Class 283 (Persian cat) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.991	75.141474	Feather boa, boa (1)
Desert	0.914	70.630613	Arabian camel, dromedary (65)
Low Contrast	0.947	66.190509	Matchstick (13)
City	0.936	68.423853	Traffic light, traffic signal (36)
Sky	0.887	63.020165	Volcano (54)
Jungle	0.925	67.672465	Cliff, drop (24)
No Background	0.937	68.087824	Matchstick (13)
High Contrast	0.942	68.444946	Corkscrew, bottle screw (13)
No Foreground	0.061	25.145970	Schipperke (96)
Water	0.906	59.895699	Grey whale, gray whale (19)
Snow	0.925	67.393982	Lakeside, lakeshore (23)
Indoor	0.907	72.101543	Bookcase (37)
Mountain	0.916	67.004423	Valley, vale (26)

Tabela 24. Metrics and Confidence Scores for Class 283 (Persian cat) - ConvNeXt

### Model ResNet

### Model ConvNeXt

## WNIOSKI DLA KLASY 13 (JUNCO, SNOWBIRD)

Na podstawie wyników przedstawionych w Tabeli X i Tabeli Y, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 13 (junco, snowbird) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.976	95.314826	House finch, linnet (7)
Desert	0.839	81.161059	Seashore, coast (97)
Low Contrast	0.877	86.390556	Brambling, Fringilla montifringilla (26)
City	0.842	75.437180	House finch, linnet (30)
Sky	0.842	83.426079	Volcano (46)
Jungle	0.920	90.342285	Bittern (11)
No Background	0.858	84.076516	Electric ray, crampfish (33)
High Contrast	0.896	87.080903	Kite (15)
No Foreground	0.362	51.447448	Water ouzel, dipper (216)
Water	0.716	61.873004	Albatross, mollymawk (124)
Snow	0.831	83.117293	Snowmobile (66)
Indoor	0.676	68.064425	File, file cabinet (132)
Mountain	0.797	76.813323	Lakeside, lakeshore (60)

Tabela 25. Metrics and Confidence Scores for Class 13 (junco, snowbird) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	1.000	68.830393	None
Desert	0.915	64.330613	Arabian camel, dromedary (75)
Low Contrast	0.982	58.656551	Chickadee (3)
City	0.888	58.006623	Traffic light, traffic signal (67)
Sky	0.935	60.091004	Parachute, chute (31)
Jungle	0.954	60.319277	Cliff, drop (16)
No Background	0.971	58.450551	Water ouzel, dipper (5)
High Contrast	0.983	58.033862	Brambling, Fringilla montifringilla (2)
No Foreground	0.460	37.770219	Magpie (187)
Water	0.935	58.536078	Albatross, mollymawk (27)
Snow	0.929	61.614732	Lakeside, lakeshore (15)
Indoor	0.837	59.676453	File, file cabinet (67)
Mountain	0.930	64.416702	Valley, vale (16)

Tabela 26. Metrics and Confidence Scores for Class 13 (junco, snowbird) - ConvNeXt

### Model ResNet

### Model ConvNeXt

## WNIOSKI DLA KLASY 235 (GERMAN SHEPHERD)

Na podstawie wyników przedstawionych w Tabeli X i Tabeli Y, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 235 (German shepherd) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	
Original	0.874	82.761220	
Desert	0.705	72.608107	
Low Contrast	0.806	77.606469	
City	0.821	77.924239	
Sky	0.787	75.390463	
Jungle	0.788	76.892591	
No Background	0.705	67.817178	Coyot Model ResNet osiągnął wysoką dokładność
High Contrast	0.819	80.306382	
No Foreground	0.005	36.099894	
Water	0.814	76.923269	
Snow	0.780	78.109030	
Indoor	0.725	74.028374	
Mountain	0.666	65.790195	

Tabela 27. Metrics and Confidence Scores for Class 235 (German shepherd) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.961	69.949959	Malinois (10)
Desert	0.844	62.162412	Arabian camel, dromedary (83)
Low Contrast	0.919	61.204567	Malinois (17)
City	0.867	59.458822	Traffic light, traffic signal (23)
Sky	0.873	62.404542	Norwegian elkhound, elkhound (30)
Jungle	0.908	64.776667	Cliff, drop (14)
No Background	0.886	60.678388	Malinois (10)
High Contrast	0.919	61.661308	Malinois (17)
No Foreground	0.042	38.284114	Schipperke (239)
Water	0.922	62.505798	Norwegian elkhound, elkhound (14)
Snow	0.890	63.506425	Norwegian elkhound, elkhound (29)
Indoor	0.800	64.582883	Bookcase (65)
Mountain	0.847	61.401422	Norwegian elkhound, elkhound (36)

Tabela 28. Metrics and Confidence Scores for Class 235 (German shepherd) - ConvNeXt

### Model ResNet

### Model ConvNeXt

## WNIOSKI DLA KLASY 15 (AMERICAN ROBIN)

Na podstawie wyników przedstawionych w Tabeli X i Tabeli Y, można wyciągnąć następujące wnioski dotyczące wpływu różnych modyfikacji tła na klasyfikację klasy 15 (American robin) przy użyciu modeli ResNet i ConvNeXt:

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.981	96.633280	Hummingbird (2)
Desert	0.830	81.119547	Seashore, coast (87)
Low Contrast	0.883	89.243569	Brambling, Fringilla montifringilla (27)
City	0.909	87.792362	Obelisk (11)
Sky	0.908	89.332178	Volcano (15)
Jungle	0.945	92.875282	Greenhouse, nursery (12)
No Background	0.872	86.418034	Brambling, Fringilla montifringilla (18)
High Contrast	0.903	89.287356	Kite (16)
No Foreground	0.023	49.667002	Magpie (226)
Water	0.802	73.219416	Snorkel (23)
Snow	0.897	88.912560	Snowmobile (49)
Indoor	0.678	69.680563	File, file cabinet (85)
Mountain	0.879	88.491847	Lakeside, lakeshore (33)

Tabela 29. Metrics and Confidence Scores for Class 15 (American robin) - ResNet

<b>Modification Type</b>	<b>Accuracy</b>	<b>Avg Confidence</b>	<b>Most Mistaken</b>
Original	0.997	68.229925	Worm fence, snake fence (1)
Desert	0.905	58.872557	Arabian camel, dromedary (73)
Low Contrast	0.968	59.050634	House finch, linnet (4)
City	0.942	59.546505	Traffic light, traffic signal (34)
Sky	0.948	60.918330	Parachute, chute (20)
Jungle	0.960	59.782053	Cliff, drop (17)
No Background	0.951	56.262251	Hummingbird (11)
High Contrast	0.963	58.630182	Water ouzel, dipper (7)
No Foreground	0.052	39.404290	Magpie (482)
Water	0.941	57.640282	Grey whale, gray whale (15)
Snow	0.952	63.954396	Snowmobile (17)
Indoor	0.858	58.543296	Bookcase (62)
Mountain	0.941	63.929843	Valley, vale (14)

Tabela 30. Metrics and Confidence Scores for Class 15 (American robin) - ConvNeXt

### Model ResNet

### Model ConvNeXt

## **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**