

Politechnika Wrocławskiego
Wydział Informatyki i Telekomunikacji

Kierunek: **Zaufane Systemy Sztucznej Inteligencji**

**PRACA DYPLOMOWA
MAGISTERSKA**

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

Paweł Pelar

Opiekun pracy
dr hab. inż. Henryk Maciejewski

Słowa kluczowe: classification, image segmentation

WROCŁAW 2024

STRESZCZENIE

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

ABSTRACT

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

SPIS TREŚCI

WPROWADZENIE

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

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

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

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

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

CEL PRACY

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

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

ZAKRES PRACY

Zakres pracy obejmuje kilka kluczowych aspektów, a mianowicie:

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

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

6. Ocena i analiza wyników:

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

7. Wnioski i rekomendacje:

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

PRZEGŁĄD LITERATURY

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

ZAKRES PRZEGŁĄDU LITERATURY

1. Klasyfikacja obrazów:

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

2. Modele głębokiego uczenia:

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

3. Segmentacja obrazów:

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

4. Wpływ tła na klasyfikację obrazów:

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

5. Metryki oceny jakości modeli:

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

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

CEL PRZEGŁĄDU LITERATURY

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

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

KLASYFIKACJA OBRAZÓW

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

Tradycyjne metody klasyfikacji obrazów

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

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

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

Ewolucja podejść z wykorzystaniem głębokiego uczenia

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

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

Zaawansowane architektury sieci neuronowych

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

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

Podsumowanie

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

MODELE GŁĘBOKIEGO UCZENIA

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

Convolutional Neural Networks (CNN)

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

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

Residual Networks (ResNet)

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

Transformery w przetwarzaniu obrazów

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

Nowoczesne architektury

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

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

Podsumowanie

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

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

SEGMENTACJA OBRAZÓW

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

Tradycyjne metody segmentacji obrazów

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

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

Nowoczesne podejścia wykorzystujące głębokie uczenie

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

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

Zastosowania segmentacji obrazów

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

Podsumowanie

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

WPŁYW TŁA NA KLASYFIKACJĘ OBRAZÓW

Wpływ tła na klasyfikację obrazów jest istotnym zagadnieniem w dziedzinie przetwarzania obrazów i głębokiego uczenia. Tło może wprowadzać szумy, zakłócenia i nieistotne informacje, które mogą negatywnie wpływać na wyniki klasyfikacji, zwłaszcza w kontekście złożonych scen i różnorodnych warunków oświetleniowych. W ramach tego przeglądu literatury omówione zostaną badania dotyczące wpływu tła na dokładność modeli klasyfikacyjnych, techniki usuwania i modyfikacji tła oraz ich zastosowanie w praktyce.

Badania dotyczące wpływu tła

Wiele badań wykazało, że tło może znacząco wpływać na wyniki klasyfikacji obrazów. Pierwsze prace w tym zakresie koncentrowały się na analizie, jak obecność złożonych i zmiennych tła wpływa na dokładność klasyfikacji. Na przykład, badania pokazują, że modele głębokiego uczenia, takie jak CNN, mogą niekiedy mylić obiekty z tłem, co prowadzi do błędnej klasyfikacji. Prace takie jak te przeprowadzone przez Torralba i Efros (2011) sugerują, że kontekst sceny, w tym tło, może wpływać na percepcję i klasyfikację obiektów. **TO DO**

Techniki usuwania i modyfikacji tła

W celu zminimalizowania negatywnego wpływu tła na klasyfikację obrazów, opracowano różne techniki usuwania i modyfikacji tła. Jednym z podstawowych podejść jest wykorzystanie segmentacji obrazów do wyodrębnienia obiektów z tła. Modele takie jak U-Net i Mask R-CNN są szeroko stosowane do precyzyjnej segmentacji obiektów, co pozwala na usunięcie tła przed klasyfikacją. Dzięki temu modele klasyfikacyjne mogą skupić się na cechach obiektów, a nie na nieistotnych elementach tła.

Kolejnym podejściem jest modyfikacja tła, polegająca na zastąpieniu oryginalnego tła jednolitym kolorem lub losowo wygenerowanym wzorem. Badania pokazują, że takie podejście może poprawić dokładność klasyfikacji, zwłaszcza w przypadku obiektów trudnych do rozróżnienia od złożonego tła. Przykładem jest praca wykonana przez Zhu et al. (2017), w której zastąpienie tła prostymi wzorami lub kolorami prowadziło do poprawy wyników klasyfikacji. **TO DO**

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 najważniejsze metryki stosowane do oceny jakości modeli, w tym dokładność, precyzja, 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.

Precyzja (Precision)

Precyzja, znana również jako dodatnia wartość predykcyjna, to stosunek liczby prawdziwie pozytywnych przykładów do liczby wszystkich przykładów sklasyfikowanych jako pozytywne. W kontekście klasyfikacji binarnej precyzja mierzy, jak wiele z przykładów sklasyfikowanych jako pozytywne faktycznie należy do klasy pozytywnej. Wysoka precyzja 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 precyzję, jak i recall.

Inne zaawansowane miary

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

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

Zastosowanie metryk w praktyce

W praktyce wybór odpowiednich metryk zależy od specyfiki zadania i rodzaju danych. Na przykład, w diagnostyce medycznej ważne jest używanie recall i F1-score, aby zapewnić,

że wszystkie przypadki choroby są wykrywane, a liczba fałszywie negatywnych wyników jest minimalna. W systemach monitoringu i detekcji obiektów, metryki takie jak AP i IoU są kluczowe do oceny precyzji i dokładności lokalizacji obiektów.

Podsumowanie

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

METODYKA BADAŃ

WPROWADZENIE DO METODYKI BADAŃ

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

PRZYGOTOWANIE ŚRODOWISKA

Przygotowanie odpowiedniego środowiska badawczego jest kluczowym krokiem w realizacji każdego projektu opartego na analizie danych i głębokim uczeniu. W niniejszych badaniach, całość prac została przeprowadzona w języku Python, który jest powszechnie stosowany w dziedzinie przetwarzania obrazów i uczenia maszynowego dzięki bogatemu ekosystemowi bibliotek i narzędzi wspomagających te procesy.

Do realizacji projektu użyto następujących bibliotek: numpy, pandas, scikit-learn, PIL, matplotlib, seaborn oraz torch. Biblioteka numpy została wykorzystana do obsługi operacji

numerycznych i manipulacji tablicami, biblioteka pandas służyła do manipulacji i analizy danych strukturalnych, takich jak tablice. Scikit-learn był wykorzystywany przy obliczeniach metryk i ocenie jakości modeli, a PIL (Python Imaging Library) umożliwiła manipulację obrazami. Biblioteki matplotlib i seaborn posłużyły do wizualizacji danych i wyników analiz, co pozwoliło na lepsze zrozumienie uzyskanych rezultatów oraz prezentację wyników w formie graficznej.

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

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

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

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

WYBRANE MODELE

W niniejszym projekcie zastosowano trzy zaawansowane modele głębokiego uczenia: ResNet, ConvNeXt oraz DeepLabv3. Każdy z tych modeli został wybrany ze względu na swoje unikalne właściwości i zdolności do realizacji określonych zadań. DeepLabv3 służył jako uniwersalny model do segmentacji, pozwalający na precyzyjne wyodrębnienie obiektów z tła. Modele ResNet i ConvNeXt, o różnych architekturach i z różnymi stopniami zaawansowania technologicznego, zostały wykorzystane do klasyfikacji obrazów. ResNet,

będący starszym modelem, oraz ConvNeXt, reprezentujący nowsze podejście, zostały wybrane w celu porównania i analizy ich wydajności w kontekście zmodyfikowanych warunków tła. Wykorzystanie gotowych, pretrenowanych modeli umożliwiło skupienie się na głównej części badania, jaką jest wpływ tła na klasyfikację obrazów, zamiast na długotrwałym procesie trenowania modeli od podstaw.

ResNet

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

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

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

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

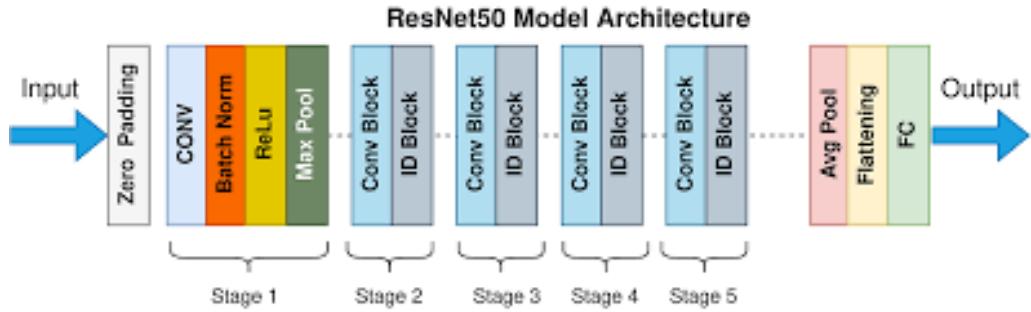
gdzie y to wyjście bloku, x to wejście, a $F(x, \{W_i\})$ to funkcja reprezentująca operacje konwolucyjne na wejściu x z zestawem wag $\{W_i\}$.

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

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

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

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



Rys. 1. Schemat architektury modelu ResNet50

maksymalnego pooling (max pooling), które redukują przestrzenne wymiary map cech, jednocześnie zachowując najważniejsze cechy.

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

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

ConvNeXt

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

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

Podstawowym elementem ConvNeXt jest moduł konwolucyjny, który został zoptymalizowany w celu lepszego uchwycenia złożonych wzorców w danych. Architektura

ConvNeXt łączy tradycyjne podejścia konwolucyjne z nowymi koncepcjami, co prowadzi do lepszej wydajności i efektywności obliczeniowej.

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

DeepLabv3

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

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

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

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

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

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

Uzasadnienie wyboru modeli

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

Wykorzystanie gotowych, pretrenowanych modeli pozwoliło skupić się na głównym celu badania – analizie wpływu tła na klasyfikację obrazów – bez konieczności poświęcania

czasu na trenowanie modeli od podstaw. Dzięki temu możliwe było przeprowadzenie bardziej szczegółowych i kompleksowych badań w zakresie modyfikacji tła i jego wpływu na wydajność modeli klasyfikacyjnych.

WYBRANE METRYKI

W celu analizy wyników klasyfikacji przed i po modyfikacjach tła, zastosowano dwie kluczowe metryki, dokładności (accuracy) oraz pewności klasyfikacji (confidence scores). Analizowana będzie również macierz pomyłek w celu zbadania najczęściej mylących się klas, oraz zbadania jakie błędy zostały popełnione. Metryki te zostały wybrane ze względu na ich zdolność do dostarczania wartościowych informacji na temat wydajności modeli w różnych warunkach. W niniejszym rozdziale szczegółowo omówimy te metryki, przedstawimy odpowiednie wzory oraz wyjaśnimy, dlaczego zostały wybrane do analizy. Metryki będą analizowane całościowo, jak również osobno dla każdej klasy, dla każdej różnej modyfikacji tła oraz dla różnych percentylów wielkości obiektu na obrazie.

Dokładność (Accuracy)

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

Wzór na dokładność jest następujący:

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

gdzie:

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

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

Pewność klasyfikacji (Confidence Scores)

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

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

Macierz pomyłek

Macierz pomyłek (confusion matrix) jest narzędziem używanym do oceny wydajności modeli klasyfikacyjnych poprzez pokazanie, jak często poszczególne klasy są mylone z innymi. Macierz pomyłek przedstawia liczbę prawdziwie pozytywnych, prawdziwie negatywnych, fałszywie pozytywnych i fałszywie negatywnych klasyfikacji dla każdej klasy.

W przypadku klasyfikacji wieloklasowej, macierz pomyłek rozszerza się do macierzy $n \times n$, gdzie n to liczba klas. Każda komórka C_{ij} w macierzy pomyłek przedstawia liczbę przypadków należących do klasy i , które zostały sklasyfikowane jako klasa j .

Analiza macierzy pomyłek pozwala na identyfikację, które klasy są najczęściej mylone, co może dostarczyć cennych informacji na temat specyficznych wyzwań związanych z klasyfikacją w kontekście zmodyfikowanych teł.

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. Oznacza to, że obrazy mogą przedstawiać obiekty

w różnych skalach, oświetleniach, perspektywach i na różnych tła ch. Taka różnorodność sprawia, że zbiór ImageNet1k doskonale odzwierciedla realistyczne warunki, z jakimi modele mogą się spotkać w praktycznych zastosowaniach. Dzięki temu, modele trenowane na tym zbiorze danych są bardziej uniwersalne i mają lepszą zdolność generalizacji.

Popularność i znaczenie ImageNet1k

ImageNet1k jest jednym z najczęściej używanych zestawów danych w badaniach nad głębokim uczeniem, co jest wynikiem jego dużej skali, różnorodności i realistycznego charakteru. Wiele przełomowych modeli, takich jak AlexNet, VGG, ResNet i Inception, zostało przetestowanych i zweryfikowanych przy użyciu tego zestawu danych. Popularność ImageNet1k sprawia, że wyniki uzyskane na tym zbiorze są łatwo porównywalne z wynikami innych badań, co umożliwia ocenę postępów i innowacji w dziedzinie przetwarzania obrazów.

Ze względu na ograniczoną dostępność klas w modelu segmentacyjnym DeepLabv3, który był trenowany na innym zbiorze danych, do badań wybrano 10 klas zwierząt, które można skutecznie wysegmentować przy użyciu tego modelu. Wybór tych klas pozwolił na przeprowadzenie dokładnych analiz i eksperymentów przy zachowaniu rozsądnego czasu przetwarzania.

PLAN BADAŃ

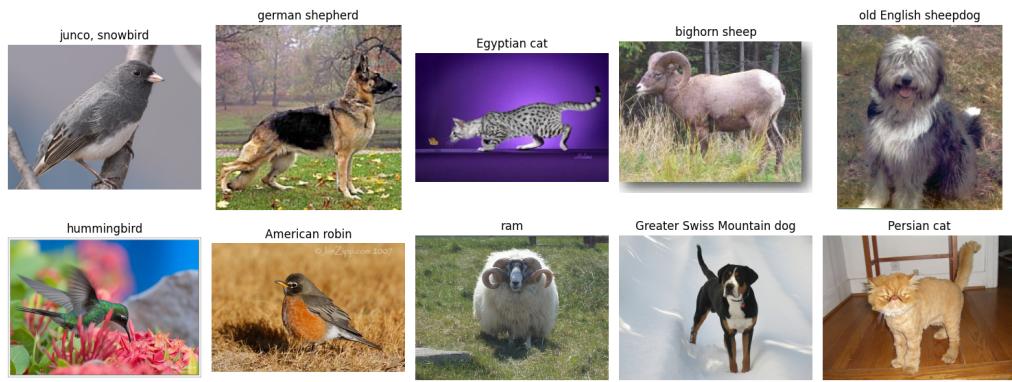
Niniejszy rozdział opisuje szczegółowy plan badań, które zostały przeprowadzone w celu zbadania wpływu tła na klasyfikację obrazów zwierząt. Poniżej przedstawiono kroki podjęte w celu realizacji badań.

Przygotowanie środowiska pracy

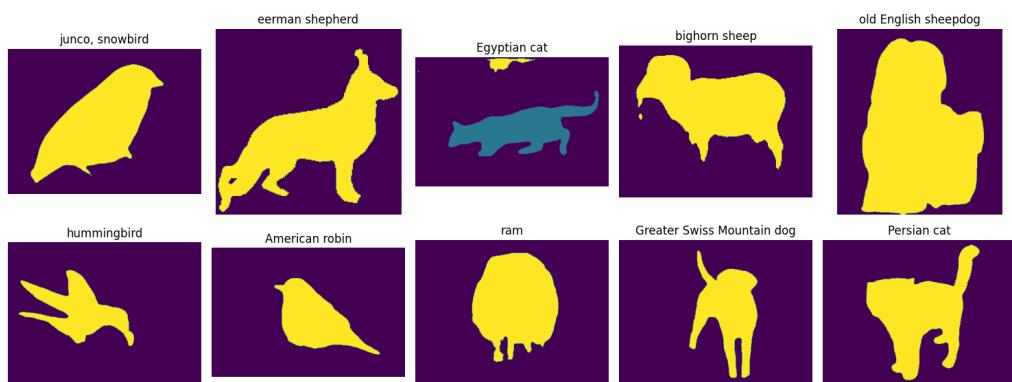
Pierwszym krokiem było przygotowanie odpowiedniego środowiska pracy. W tym celu skonfigurowano środowisko programistyczne, które obejmowało instalację niezbędnych bibliotek i narzędzi, takich jak numpy, pandas, scikit-learn, PIL, matplotlib, seaborn, torch oraz torchvision. Zastosowano również system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

Wybranie modeli do segmentacji i klasyfikacji

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



Rys. 2. Przykładowe oryginalne zdjęcia wybranych klas



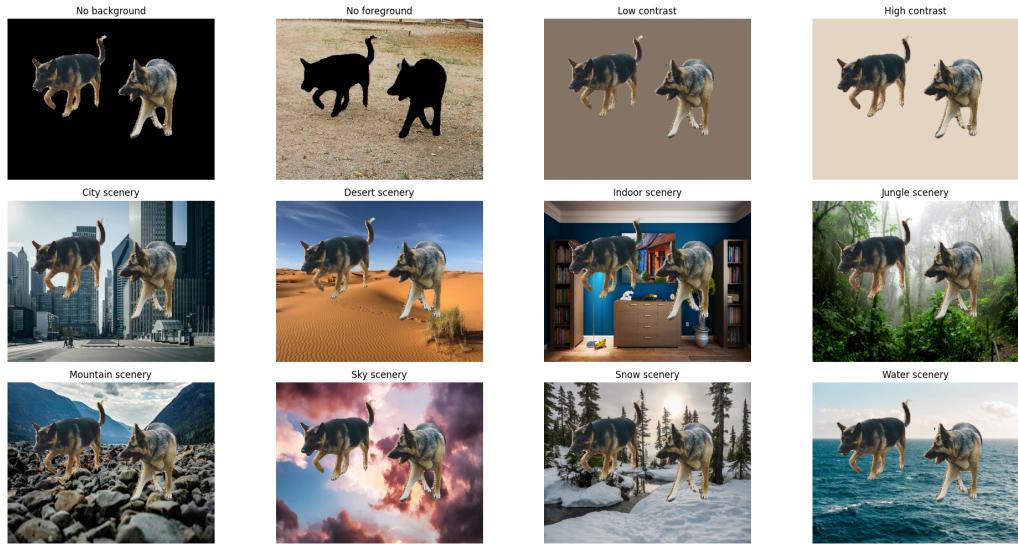
Rys. 3. Maski przykładowych wysegmentowanych obrazów

Wybranie klas zwierząt

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

Segmentacja obrazów

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



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

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

Dokonanie predykcji na wybranych modelach

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

Dodanie kategorii zdjęć pod względem procentu zajmowanego przez obiekt na zdjęciu

Zbadanie stosunku wielkości obiektu do całego zdjęcia jest istotne w kontekście badania wpływu tła na klasyfikację, ponieważ może 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:

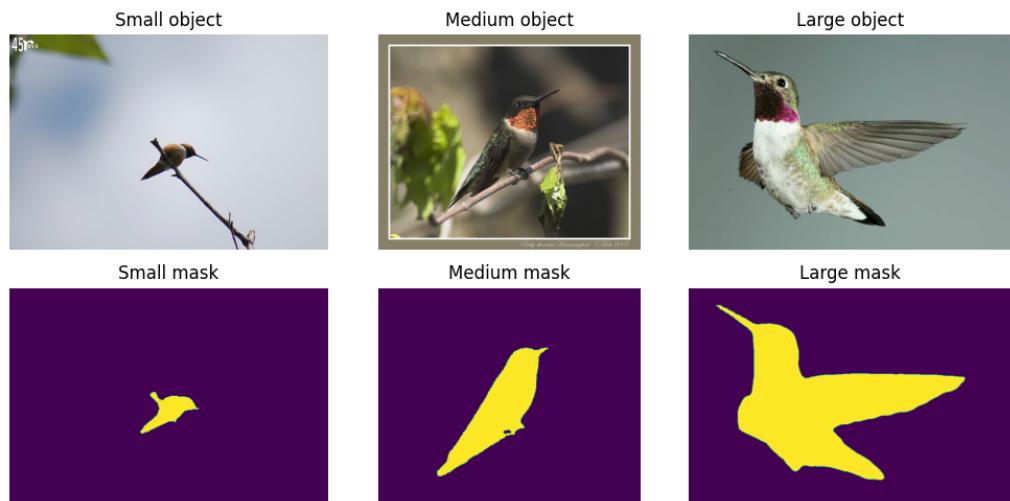
$$\text{Procent powierzchni zajmowanej przez obiekt} = \left(\frac{\text{Powierzchnia obiektu}}{\text{Powierzchnia całkowita obrazu}} \right) \times 100 \quad (4)$$

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

Analiza wyników

Wyniki zostały poddane szczegółowej analizie w kilku aspektach:

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



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

Wnioski i dalsze kierunki rozwoju

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

Podsumowanie

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

BADANIA

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

PODSUMOWANIE

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

SPIS RYSUNKÓW

SPIS LISTINGÓW

SPIS TABEL

Dodatki