

Politechnika Poznańska
Wydział inżynierii mechanicznej
Inżynieria Biomedyczna

Nr indeksu:139532



Maciej Winiarski

*Zastosowanie sztucznych sieci
neuronowych do rozpoznawania
deformacji kręgosłupa*

Promotor pracy dyplomowej:
Dr. Hab. Inż. Witold Stankiewicz
Instytut Mechaniki Stosowanej

Poznań 2024



Karta pracy dyplomowej inżynierskiej

Uczelnia:	Politechnika Poznańska	Profil studiów:	ogólnoakademicki
Kierunek:	Inżynieria Biomedyczna	Forma studiów:	stacjonarne
Studia w zakresie:	-	Poziom studiów:	pierwszego stopnia

Zobowiązuję/zobowiązujemy się samodzielnie wykonać pracę w zakresie wyspecyfikowanym niżej. Wszystkie elementy (m.in. rysunki, tabele, cytaty, programy komputerowe, urządzenia itp.), które zostaną wykorzystane w pracy, a nie będą mojego/naszego autorstwa będą w odpowiedni sposób zaznaczone i będzie podane źródło ich pochodzenia.

Jeżeli w wyniku realizacji pracy zostanie dokonany wynalazek, wzór użytkowy, wzór przemysłowy, znak towarowy, prawa do rozwiązań przysługiwać będą Politechnice Poznańskiej. Prawo to zostanie uregulowane odrębną umową.

Oświadczam, iż o wyniku prac wskazanych powyżej, a także o innych, w tym tych, które mogą być przedmiotem tajemnicy Politechniki Poznańskiej, niezwłocznie powiadomię promotora pracy.

Zobowiązuję się ponadto do zachowania w tajemnicy wszystkich informacji technicznych, technologicznych, organizacyjnych, uzyskanych w Politechnice Poznańskiej w okresie od daty rozpoczęcia realizacji prac do 5 lat od daty zakończenia wykonania prac.

	Imię i nazwisko	Nr albumu	Data i podpis
Student:	Maciej Winiarski	139532	

Tytuł pracy:	Zastosowanie sztucznych sieci neuronowych do rozpoznawania deformacji kręgosłupa
Wersja angielska tytułu:	<i>Application of the artificial neural networks in the detection of spine deformations</i>
Dane wejściowe:	<ul style="list-style-type: none">- dokumentacja języka Python,- dokumentacja bibliotek do języka Python umożliwiających przetwarzanie, analizę oraz interpretację danych medycznych z obrazowania kręgosłupa (m.in. Tensorflow, OpenCV, Pydicom),- dane medyczne z obrazowania kręgosłupa dostarczone z Ortopedyczno-Rehabilitacyjnego Szpitala Klinicznego im. W. Degi.
Zakres pracy:	<ul style="list-style-type: none">- przetwarzanie danych medycznych pochodzących z obrazowania medycznego i ich przygotowanie do dalszych analiz,- przygotowanie zbioru danych w postaci wektorów uczących i etykiet- zastosowanie sztucznej inteligencji oraz metod przetwarzania obrazów w procesie rozpoznawania i opisu deformacji kręgosłupa.
Termin oddania pracy:	31 stycznia 2024 r.
Promotor:	dr hab. inż. Witold Stankiewicz
Jednostka organizacyjna promotora:	Instytut Mechaniki Stosowanej

Z-ca DYREKTORA INSTYTUTU

Jakub Grabski
dr inż. Jakub Grabski

podpis dyrektora/kierownika jednostki organizacyjnej promotora

Rektor Politechniki Poznańskiej
prof. dr hab. inż. Teofil Jesionowski

Krzysztof Grześkowiak
z up. dr inż. Krzysztof Grześkowiak

Prodziekan

25.10.2023

data i podpis Dziekana

1.Streszczenie

Poniższa praca ma na celu stworzenie zbioru danych i wytrenowanie na nim modeli wykorzystujących sieci neuronowe do segmentacji kręgów ze zdjęć rentgenowskich. Utworzono dwie wersje zbioru danych, kompatybilne z wybranymi modelami. Parametry wstępne dobrano na podstawie rozwiązań zastosowanych w artykułach naukowych zbliżonych tematycznie. Podczas etapu projektowego, obejmującego uczenie sieci neuronowych, parametry te doskonalono aż do uzyskania zadowalającej dokładności rozpoznawania i oznaczania kręgów. Cel segmentacji kręgu został osiągnięty przy pomocy modelu F-RCNN, a na podstawie uzyskanych rezultatów zaproponowano metodę automatycznego obliczania kąta cobb'a.

1.1 Abstract

This study aims to create a dataset and train neural network models for the purpose of segmenting vertebrae from X-ray images. Two versions of the dataset were developed in order to be compatible with selected models. Initial parameters were selected based on solutions applied in thematically related scientific articles. Throughout the project stage, which encompassed the training of neural networks, these parameters were refined until achieving satisfactory accuracy in recognizing and annotating vertebrae. The goal of vertebra segmentation was accomplished using the F-RCNN model, and the obtained results were utilized to propose an automatic process of calculating the Cobb angle.

2. Wstęp Teoretyczny

2.1. Słowa kluczowe i skróty

Sztuczna inteligencja (ang. artificial intelligence, AI) - Dział informatyki zajmujący się tworzeniem modeli inteligentnych zachowań, stosowanych w programach komputerowych [1]. Nazwa ta określa jednocześnie dziedzinę badań w informatyce zajmującą się tworzeniem i badaniem inteligentnych maszyn.

Uczenie maszynowe (ang. Machine Learning, ML) - Dział sztucznej inteligencji skupiony na opracowywaniu metod umożliwiających maszynom samodzielne uczenie się na podstawie danych.

Uczenie głębokie (ang. Deep Learning) - Poddziedzina uczenia maszynowego, która wykorzystuje głębokie sieci neuronowe do modelowania i rozumienia złożonych struktur danych.

Sztuczne sieci neuronowe (ang. Artificial neural networks, ANN) - Systemy przetwarzania obliczeniowego, składające się z dużej liczby połączonych ze sobą węzłów obliczeniowych.

Zbiór danych (ang. Dataset) – Zbiór informacji używanych do uczenia maszynowego/głębokiego. W zależności od zastosowania, funkcję zbioru danych może pełnić zbiór tekstów, grafik, nagrań i innych.

COCO – Format przechowywania adnotacji zbioru danych w pliku JSON, jako zbioru współrzędnych etykiet.

Kontener - Przenośne jednostki oprogramowania, cechujące się odizolowaniem od pozostałych części środowiska.

Silnik Docker (ang. Docker Engine) - Centralna część platformy Docker, która umożliwia zarządzanie i uruchamianie kontenerów.

Model - Matematyczna reprezentacja relacji pomiędzy danymi wejściowymi a danymi wyjściowymi. W przypadku sieci neuronowych składa się z warstw

neuronów, gdzie każdy neuron ma przypisane wagi, a proces uczenia polega na dostosowywaniu wag na podstawie danych treningowych.

Dane treningowe – Część zbioru danych wykorzystywana do uczenia modelu SI

Dane testowe – Część zbioru danych wykorzystywana w weryfikacji wytrenowanego modelu SI.

Optymalizator - Algorytm używany do dostosowywania wag i parametrów sieci w trakcie procesu uczenia maszynowego.

Strata (ang. Loss) – Miara różnicy przewidywań modelu od rzeczywistych wartości w zbiorze danych treningowych.

Segmentacja Obrazu (ang. Image Segmentation) – Proces podziału obrazu cyfrowego na segmenty.

Nadmierne dopasowanie (ang. Overfitting) – Zjawisko występujące przy zbyt dokładnym wyuczeniu modelu do przetwarzania danych treningowych, skutkujące obniżoną zdolnością do uogólniania (skutecznością dla nowych danych).

Zanieczyszczenie danych testowych (ang. Cross-contamination) – Zjawisko pomieszania danych treningowych i testowych, uniemożliwiające poprawną ocenę overfitting'u.

Stronniczość (ang. Bias) – Wada wytrenowanego modelu wynikająca z nierównego rozkładu zdjęć o różnych cechach wewnątrz zbioru danych. Skutkuje zmniejszoną zdolnością modelu przy kontakcie z rzadziej widzianymi danymi.

Augmentacja – Proces zwielokrotnienia danych poprzez transformacje, przy zachowaniu cech charakterystycznych. W przypadku obrazu augmentacji można dokonać poprzez odbicie obrazu wzdłuż osi, obrócenie o kąt, zmianę odcienia, jasności, ostrości.

2.2 Sztuczna inteligencja i uczenie maszynowe

Uczenie maszynowe jest jedną z najbardziej znaczących podkategorii SI, w której tworzone są algorytmy zdolne do analizy znacznych ilości danych, a następnie podejmowanie trafniejszych decyzji na podstawie wyciągniętych wniosków. W dalszych rozdziałach bliżej omówione zostaną modele uczenia nadzorowanego wykorzystywane do segmentacji obrazów, a w szczególności FRCNN oraz U-Net.

Kluczowym elementem każdego przedsięwzięcia związanego z uczeniem maszynowym jest odpowiedni zbiór danych. Tworzenie zbioru danych do uczenia nadzorowanego składa się z 3 głównych elementów:

Zbierania danych

Wstępnego przetwarzania

Etykietowania

Wstępne przetwarzanie składa się z wielu procesów normalizujących dane – w przypadku zdjęć wiąże się to ze zmianą wymiarów, normalizacją jasności i kontrastu, lub przycinaniem. **Etykietowanie** polega na opisywaniu źródła prawdy podstawowej dla każdego elementu zbioru danych. Etykieta jest następnie traktowana jako cel optymalizacyjny dla modelu.

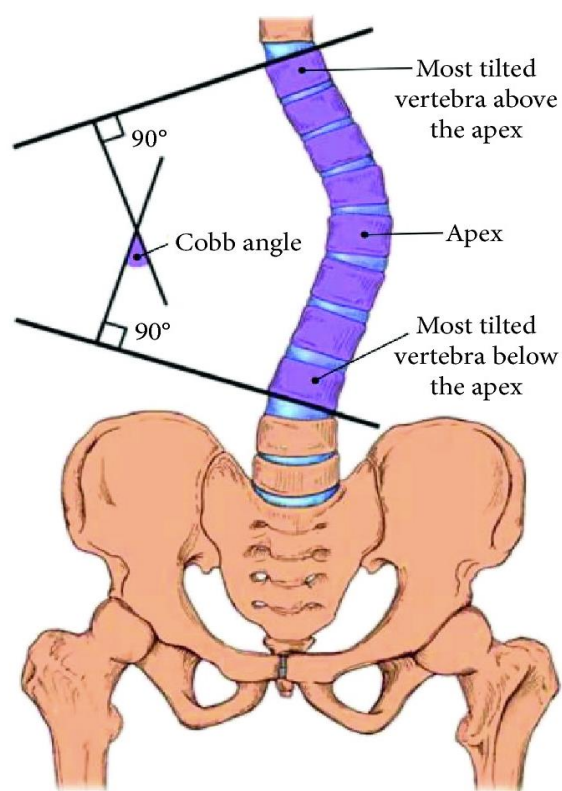
Gotowy zbiór powinien zawierać znaczącą ilość danych – w relacji do zastosowania. Należy upewnić się, że zbiór adekwatnie przedstawia różne rodzaje zdjęć i zawiera podobne ilości zdjęć przedstawiających różne kategorie [1].

2.3 Diagnoza skoliozy na podstawie zdjęć z aparatu Roentgena

Skolioza jest to trójwymiarowe skrzywienie kręgosłupa, określone jako zniekształcenie wynoszące ponad 10° w płaszczyźnie czołowej. Efektami

skoliozy są bóle pleców i ramion, zapalenie kości i stawów, a nawet problemy układów oddechowego i krążenia. Częstość występowania skoliozy jest stała w czasie i określona na 2-3% ogółu populacji [2]

Skoliozę najczęściej diagnozuje się na podstawie obrazów diagnostycznych kręgosłupa. Dominującą metodą diagnozy jest obliczenie kąta Cobb'a. W celu zmierzenia kąta Cobb'a należy w pierwszej kolejności wybrać od którego kręgu zaczyna się skrzywienie, oraz na którym kręgu się kończy – czyli kręgi o maksymalnym nachyleniu względem siebie. Należy zaznaczyć linie będące przedłużeniem zewnętrznych krawędzi wybranych kręgów, zgodnie z grafiką poniżej [3]. Kąt wewnętrzny pomiędzy nakreślonymi liniami można zmierzyć na przecięciu linii.



Rysunek 1. metoda wyznaczenia kąta Cobb'a [3]

2.4 Machine learning w medycynie

Uczenie maszynowe ma wiele zastosowań w medycynie, zwłaszcza w diagnostyce, gdzie wykorzystywana jest zdolność sieci neuronowych do zauważania podobieństw w danych, często niezauważalnych dla lekarza.

Jednym z dominujących zastosowań uczenia maszynowego jest segmentacja obrazu. W latach ubiegłych eksperymentowano z różnymi technikami segmentacji obrazu w diagnostyce medycznej, między innymi, poprzez progresywne progowanie obrazów w celu wyizolowania kręgosłupa ze zdjęć rentgenowskich [4]. Spośród wielu dostępnych metod, w analizach medycznych stosowano również grupowanie (ang. Clustering) oraz detekcję krawędzi (ang. Edge detection) [5]. pomimo swoich ograniczeń, metody te pozwalają na uproszczenie opisywania zdjęć diagnostycznych w medycynie lub przygotowanie ich do dalszej obróbki.

Wraz z rozwojem uczenia maszynowego, zaczęto wykorzystywać konwolucyjne sieci neuronowe do celów segmentacji obrazu. W roku 2015 stworzono i opisano architekturę U-Net, jako rozwiązanie służące do segmentacji danych medycznych [6]. Rozwój sieci U-Net oraz innych konwolucyjnych sieci neuronowych przyczynił się do znaczącego rozwoju i automatyzacji diagnostyki.

3. Cel i zakres Pracy

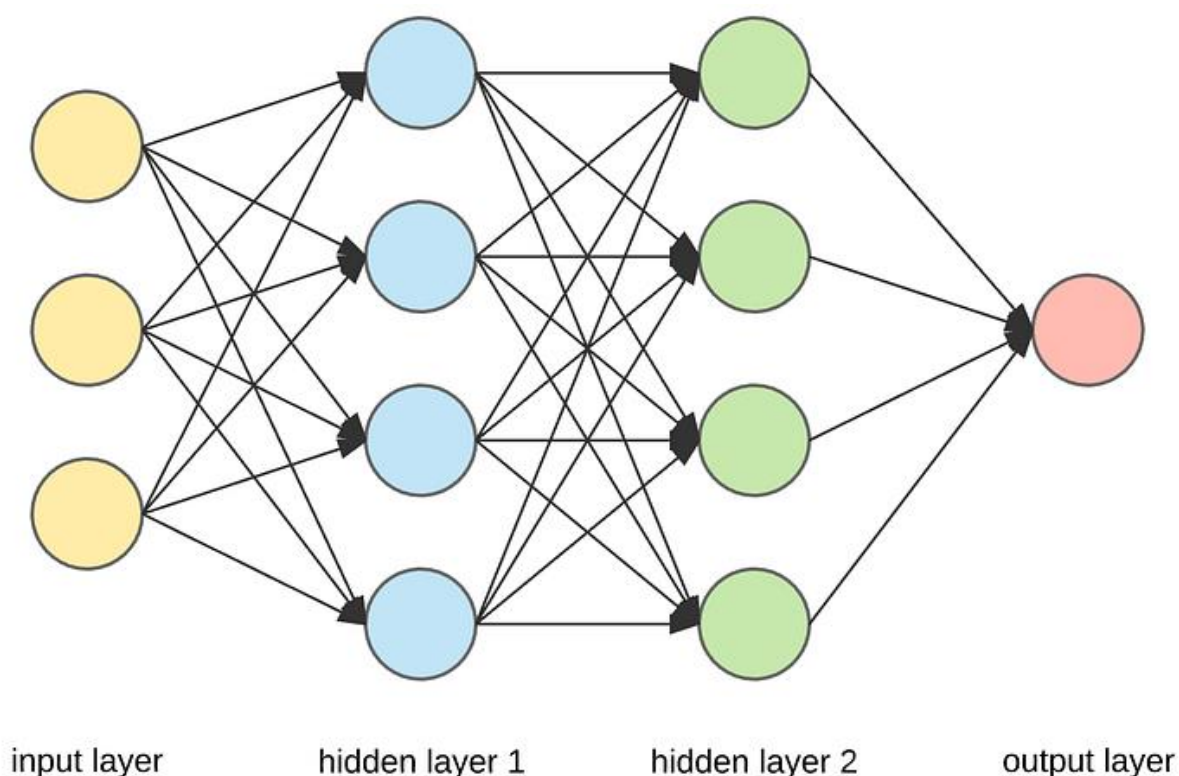
Celem pracy jest stworzenie zbioru danych i wytrenowanie na nim modeli F-RCNN oraz U-Net dla celu segmentacji kręgów ze zdjęć rentgenowskich.

W zakresie pracy mieści się wykonanie zbioru danych w postaci zdjęć z etykietami poligonowymi, narzędzi do konwersji danych z etykiet poligonowych

na maski binarne, oprogramowania wizualizującego efekty pracy modelu oraz pętli treningowych dla wybranych modeli.

4. Przegląd aktualnej literatury i stanu wiedzy

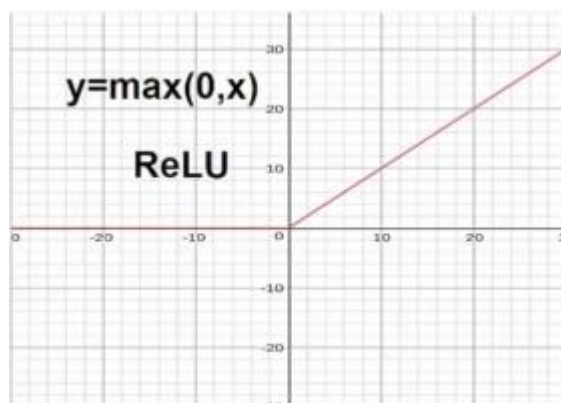
Architektura sztucznych sieci neuronowych została zainspirowana działaniem biologicznych sieci neuronowych. Poniższy rysunek przedstawia ogólną architekturę sztucznych sieci neuronowych, składającą się z warstwy wejściowej, warstwy wyjściowej i potencjalnie wielu warstw ukrytych pomiędzy nimi.



Rysunek 2 Architektura sztucznych sieci neuronowych [7]

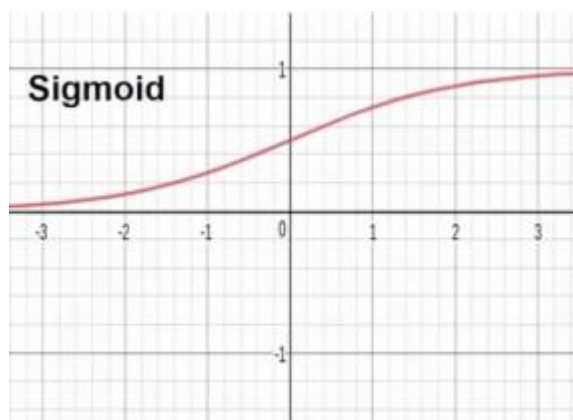
Warstwa wejściowa przyjmuje określone dane wejściowe i przekazuje je do warstw ukrytych. Po przejściu przez sieć, otrzymujemy wynik na warstwie wyjściowej. Elementy sieci neuronowych nazywamy neuronami. Każdy z neuronów przyjmuje dane wejściowe i sumuje je według wag przypisanych każdemu wejściu. Jeżeli uzyskana suma przekroczy wartość progową, uruchomiona zostaje funkcja aktywacyjna neuronu. Stosowane jest wiele funkcji aktywacyjnych, najpopularniejszymi z nich są Sigmoid, ReLU i Tanh.

Funkcja ReLU (Rectified Linear Unit) ogranicza wartości wyjściowe z neuronu do zbioru $(0: x)$, jak pokazano na rysunku poniżej, gdzie x jest wartością uzyskaną na neuronie, y jest wartością wyjściową neuronu.



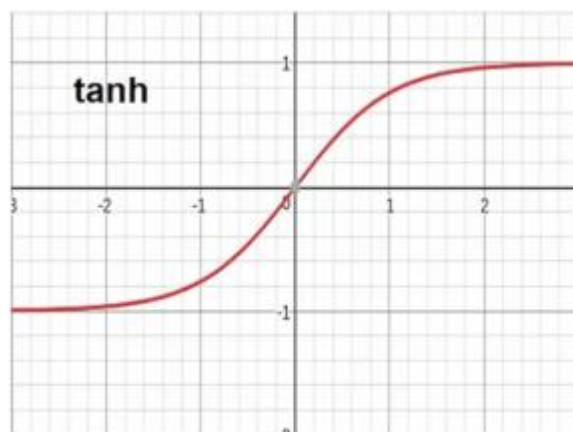
Rysunek 3 Funkcja aktywacyjna ReLU [8]

Funkcja sigmoid ogranicza wartości wyjściowe neuronu do zbioru $(0:1)$, jak pokazano na rysunku poniżej.



Rysunek 4 Funkcja aktywacyjna Sigmoid [8]

Na rysunku poniżej zaznaczono funkcję Tanh, która ogranicza wartości wyjściowe neuronu do zbioru $(-1:1)$.



Rysunek 5 Funkcja aktywacyjna tanh [8]

Tak stworzone sieci neuronowe mają wiele zastosowań. Jednym z najbardziej podstawowych jest regresja liniowa, w którym algorytm dobiera funkcję liniową która najlepiej reprezentuje badane przez model dane.

4.1 Elementy procesu uczenia maszynowego

Wykorzystywanie obrazów jako danych wejściowych w nie-konwolucyjnych sieciach neuronowych wiąże się z koniecznością opisanie i trenowania olbrzymiej ilości wag dla każdej warstwy. Problem ten rozwiązano poprzez zastosowanie warstw konwolucyjnych i poolingowych, pierwszy raz zaproponowanych w sieci neocognitron. [9]

Warstwa konwolucyjna jest pierwszą warstwą ukrytą każdej sieci CNN. Przejście obrazu przez warstwę konwolucyjną polega na mnożeniu element-wise wszystkich wartości pikseli wewnątrz obszarów obrazu przez kolejne filtry w postaci macierzy o wybranym rozmiarze, uzyskując jedną wartość dla obszaru, tzw. mapowanie cech, która przekazywana jest do kolejnej warstwy, znacząco zmniejszając ilość zapisywanych wag. [10]

Warstwy Poolingowe przekształcają wejściowe mapowania cech w mapowania wyjściowe o dwukrotnie zmniejszonych wymiarach. Skutkuje to uniezależnieniem modelu od drobnych zmian w obrazie i dostosowaniem wag do

definiujących cech danych wejściowych. Warstwa Poolingowa zazwyczaj umieszczana jest po warstwie konwolucyjnej o nieliniowej funkcji aktywacyjnej, np. ReLU. Warstwa Poolingowa może wykonywać wiele operacji, jednak najpopularniejszymi są average pooling oraz maximum pooling. [11]

Average Pooling polega na obliczeniu średniej wartości każdego obszaru na mapie cech, a Maximum Pooling polega na obliczeniu wartości maksymalnej takiego obszaru.

Ostatnią warstwą CNN, służącą do klasyfikacji, jest warstwa w pełni połączona, która, na podstawie otrzymanych danych z poprzednich warstw, dokonuje klasyfikacji danych wejściowych. Ilość neuronów na warstwie wyjściowej jest zależna od ilości klas. Uzyskana klasyfikacja jest następnie porównywana z etykietą danego zdjęcia w celu obliczenia straty. [12]

W przypadku niektórych sieci, takich jak U-Net, stosowane są również warstwy konwolucji transponowanej. Taki rodzaj warstwy stosowany jest, aby uzyskać wyjściową mapę cech o większych wymiarach przestrzennych niż wejściowa mapa cech. [13]

Rezultaty uzyskane przez sieć w uczeniu nadzorowanym oceniamy obliczając stratę względem etykiet wejściowych. Istnieje wiele parametrów straty, poniżej omówiono kilka najczęściej używanych [14]:

- MAE (Mean Absolute Loss) – Funkcja straty używana przy regresji. Polega na obliczeniu średniej różnic bezwzględnych pomiędzy etykietami przewidywanymi i wejściowymi.
- Cross-entropy Loss – Funkcja straty używana w segmentacji klasowej. Wartość straty rośnie logarytmicznie, gdy wartość przewidziana przez model odbiega od etykiety wejściowej.

Do modelu dobierany jest optymalizator, który dąży do minimalizacji straty w kolejnych epokach, poprzez dostosowanie wag na każdym z neuronów sieci.

Obecnie najczęściej stosowany jest optymalizator ADAM, będący metodą stochastycznego spadku gradientu opartą na adaptacyjnej estymacji momentów pierwszego i drugiego rzędu. [15]

Jednym z kluczowych elementów uczenia maszynowego jest wykorzystanie uczenia transferowego. Jest to proces algorytmiczny w którym struktura i wiedza wyciągnięta z jednego zadania, pozwala uprościć proces uczenia powiązanego problemu. [16] Przykładem może być użycie struktury i wag sieci wytrenowanej do rozpoznawania obrazów należących do pewnych klas, w celu stworzenia nowego modelu służącego do rozpoznawania obrazów należących do innych klas.

4.2 Parametry zbioru danych dla segmentacji obrazu

Zbiór danych przygotowany do uczenia maszynowego musi spełniać szereg wymagań; Między innymi być wystarczających rozmiarów oraz odpowiedniej jakości. W przypadku gdy przygotowany zbiór będzie nieodpowiedni, trenowanie modelu uczenia maszynowego nie będzie skuteczne.

W przypadku segmentacji obrazu, zbiór danych składa się ze zdjęć i ich etykiet. Im więcej danych jesteśmy w stanie zgromadzić w zbiorze danych, tym lepsze efekty uzyskamy w procesie uczenia. Wielkości zbiorów danych wykorzystanych w artykułach naukowych zazwyczaj przekraczają 300 zdjęć przed augmentacją [17]

Wśród najczęściej używanych rodzajów etykiet mamy maski binarne, używane dla sieci U-Net [18] oraz obszary wiążące (bounding boxes), używane np. dla sieci F-RCNN [19]

Proces etykietowania może odbywać się w sposób:

Manualny – ręczna segmentacja zdjęć.

Półautomatyczny – wykorzystuje się wytrenowany model ML do wstępnego opisanie zdjęć, a następnie ręczne poprawienie segmentacji przez człowieka.

Automatyczny – wykorzystanie wytrenowanego już modelu w celu opisanie danych dla kolejnych modeli

Jakość i precyzja etykiet mają bezpośredni wpływ na skuteczność trenowania modelu. Ze względu na pracochłonność ręcznego etykietowania, jest ono często elementem ograniczającym rozmiar zbioru danych.

Aby zwiększyć ilość zdjęć w zbiorze, dane można zwielokrotnić w procesie augmentacji, co przyczynia się do poprawy jakości trenowania i zmniejszenia ryzyka zbytniego dopasowania danych [18]. Należy pamiętać, że stopień augmentacji danych musi być dostosowany zarówno do modelu jak i zbioru danych, tak aby odwzorowywał rzeczywistość.

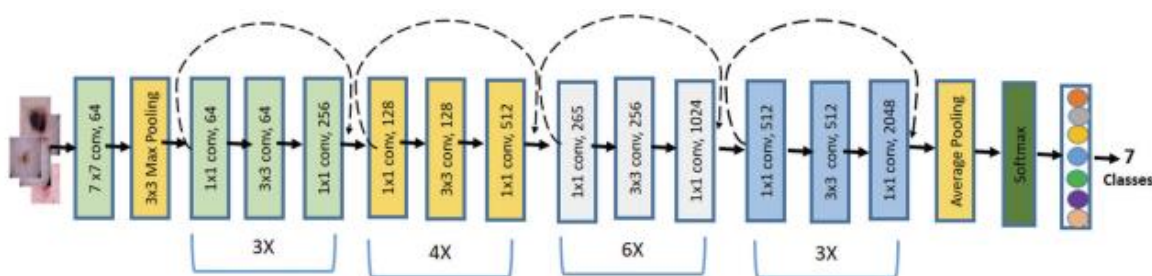
Jako ostatni krok w przygotowywaniu zbioru danych, przygotowany zbiór dzieli się na dane treningowe, walidacyjne oraz testowe [17]

4.3 Modele uczenia maszynowego

Wszystkie parametry procesu uczenia maszynowego muszą być dostosowane pod odpowiedni model. W tym podrozdziale omówione zostaną 3 modele kluczowe dla tej pracy.

ResNet50 to model klasyfikacji obrazów, który jako pierwszy wykorzystał połączenia rezydualne, pozwalające na powiązanie efektów uczenia pomiędzy daną warstwą a dalszymi warstwami, w celu ułatwienia procesu uczenia.

Rysunek poniżej przedstawia budowę modelu ResNet50 z zaznaczeniem warstw rezydualnych.



Rysunek 6 Schemat blokowy sieci ResNet50 [20]

Architektura ResNet rozpoczyna się od warstw konwolucyjnej i MaxPooling, następnie przechodzi przez kolejne bloki rezydualne. Po przejściu przez odpowiednią ilość bloków trafia do warstwy Average Pooling. Ostateczny wynik uzyskany jest funkcją aktywnością softmax.

Faster Region-Based Convolutional Neural Network (FRCNN) –

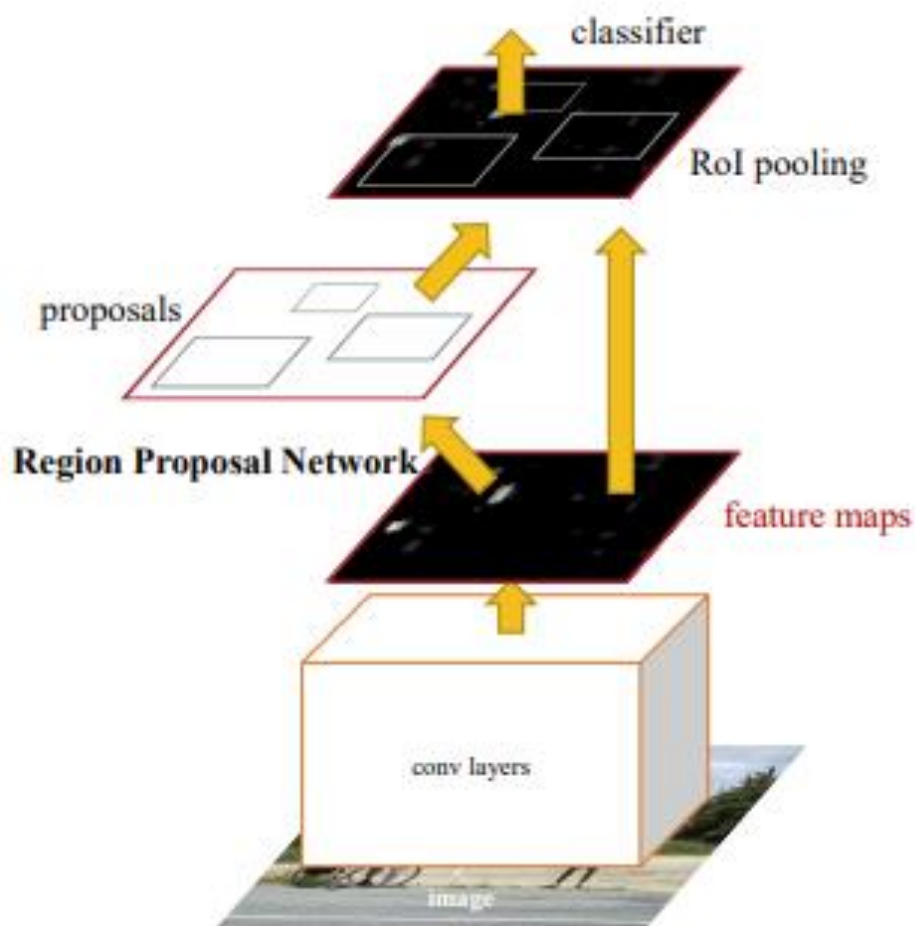
Model ten jest udoskonaloną wersją sieci typu CNN, służącą do segmentacji klasowej zdjęć. Pierwszą zmianą na drodze do powstania FRCNN było ograniczenie ilości regionów klasyfikowanych przez sieć neuronową, tworząc Region-Based CNN. Mimo, że ta zmiana znacznie zwiększyła wydajność sieci, nie była wystarczająca. W następnej iteracji zmieniono sposób działania algorytmu – W pierwszej kolejności tworzono konwolucyjną mapę obrazu, na której dopiero zaznaczano regiony zainteresowania. Znacząco przyspieszyło to

działanie sieci, którą nazwano Fast R-CNN. Faster RCNN utworzono poprzez zastąpienie algorytmu identyfikującego regiony zainteresowań osobną siecią do tego celu. [21].

Model Faster RCNN składa się z 3 sieci głównych:

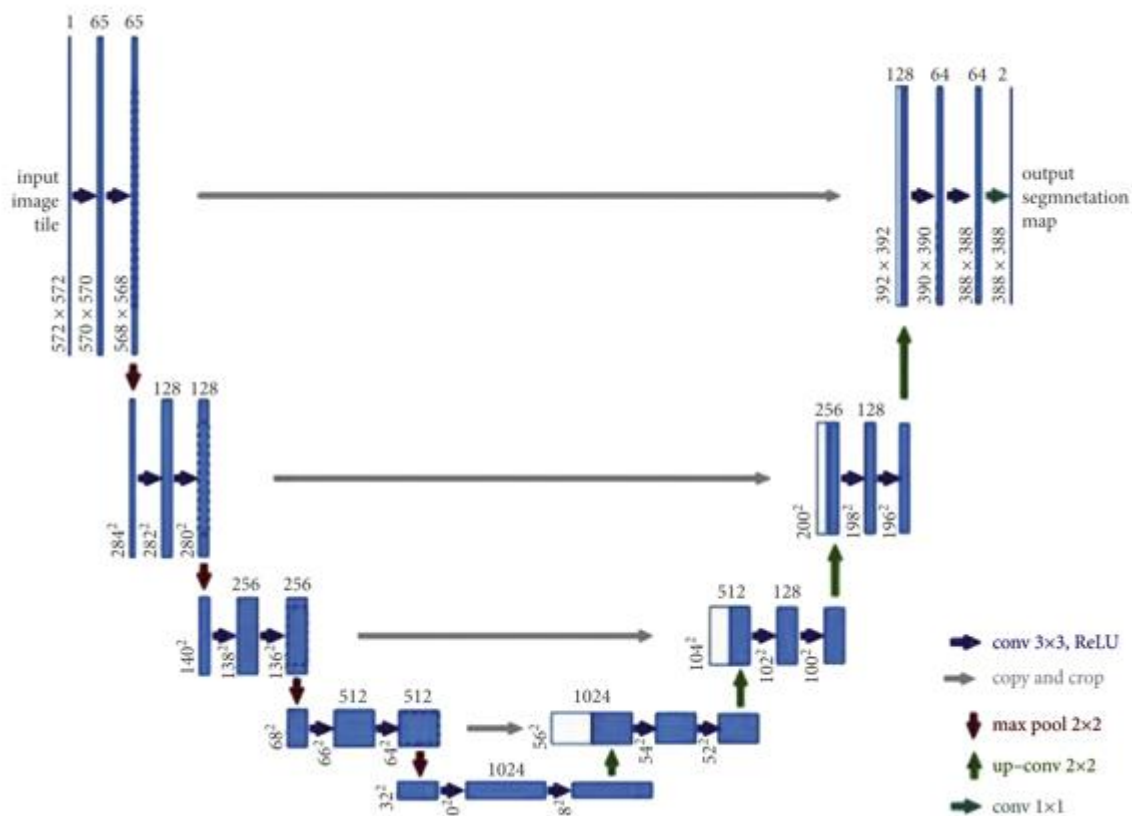
- Sieć ekstrakcji cech (np. ResNet50)
- Sieć propozycji obszarów (ang. Region Proposal Network, RPN)
- Sieć klasyfikacji/regresji propozycji

W pierwszej kolejności sieć ekstrakcji cech tworzy mapę cech. RPN wykorzystuje mapę cech w celu stworzenia propozycji regionów zainteresowania (ang. Regions of Interest, ROIs) na podstawie których sieć klasyfikacji tworzy obszary wiążące (bounding boxes) a następnie porównuje wygenerowane obszary z prawdą podstawową, zgodnie z zasadą przecięcie ponad częścią wspólną (ang. Intersection over Union, IoU). [22]. Efekty pracy modelu oceniane są przy pomocy dwóch warstw straty – straty RPN oraz R-CNN. Obie te warstwy oceniane są pod kątem klasyfikacji oraz regresji obszarów wiążących, jednak warstwa R-CNN jest porównywana względem danych wyjściowych z warstwy RPN, a nie wejściowych danych modelu [23]. Budowę modelu przedstawia grafika poniżej



Rysunek 7 trój-elementowa budowa sieci faster RCNN [24]

U-Net – Jest to konwolucyjna sieć neuronowa, będąca rozwinięciem pełnej sieci konwolucyjnej (ang. Fully convolutional neural network, FCN). Sieć U-Net została stworzona do segmentacji obrazów biomedycznych. Składa się z kodera, modułu zwężającego (ang. bottleneck) i dekodera [6]. Efekty pracy sieci U-net najczęściej oceniane są przez obliczenie straty cross-entropy, aby zapewnić jak największe odtworzenie pikseli maski wejściowej. Budowa sieci U-Net została przedstawiona na grafice poniżej.



Rysunek 8 budowa warstw konwolucyjnych sieci U-Net [6]

Po lewej stronie powyższego rysunku obserwujemy koder, który w kolejnych elementach zmniejsza wielkość obrazu przechodząc przez kolejne warstwy poolingowe. Po prawej stronie znajduje się dekodery, gdzie wykorzystano transponowane warstwy konwolucyjne do ponownego zwiększenia obrazu w kolejnych krokach. Zastosowano również połączenia pomijające, pozwalające na ominięcie dalszego przejścia przez koder. Tak stworzona architektura pozwala sieci uczyć się zarówno subtelnych cech obrazu jak i bardziej ogólnych dla uzyskania lepszej trafności.

5.Omówienie pracy wykonanej i wytrenowanych modeli

Realizacja celu pracy, czyli stworzenia narzędzi umożliwiających automatyczną detekcję deformacji kręgosłupa, wymagało przygotowania poprawnego zbioru

danych wejściowych, opracowania architektury modelu maszynowego (sieci neuronowej) oraz jego wytrenowania z użyciem odpowiednio dobranych wartości hiper parametrów.

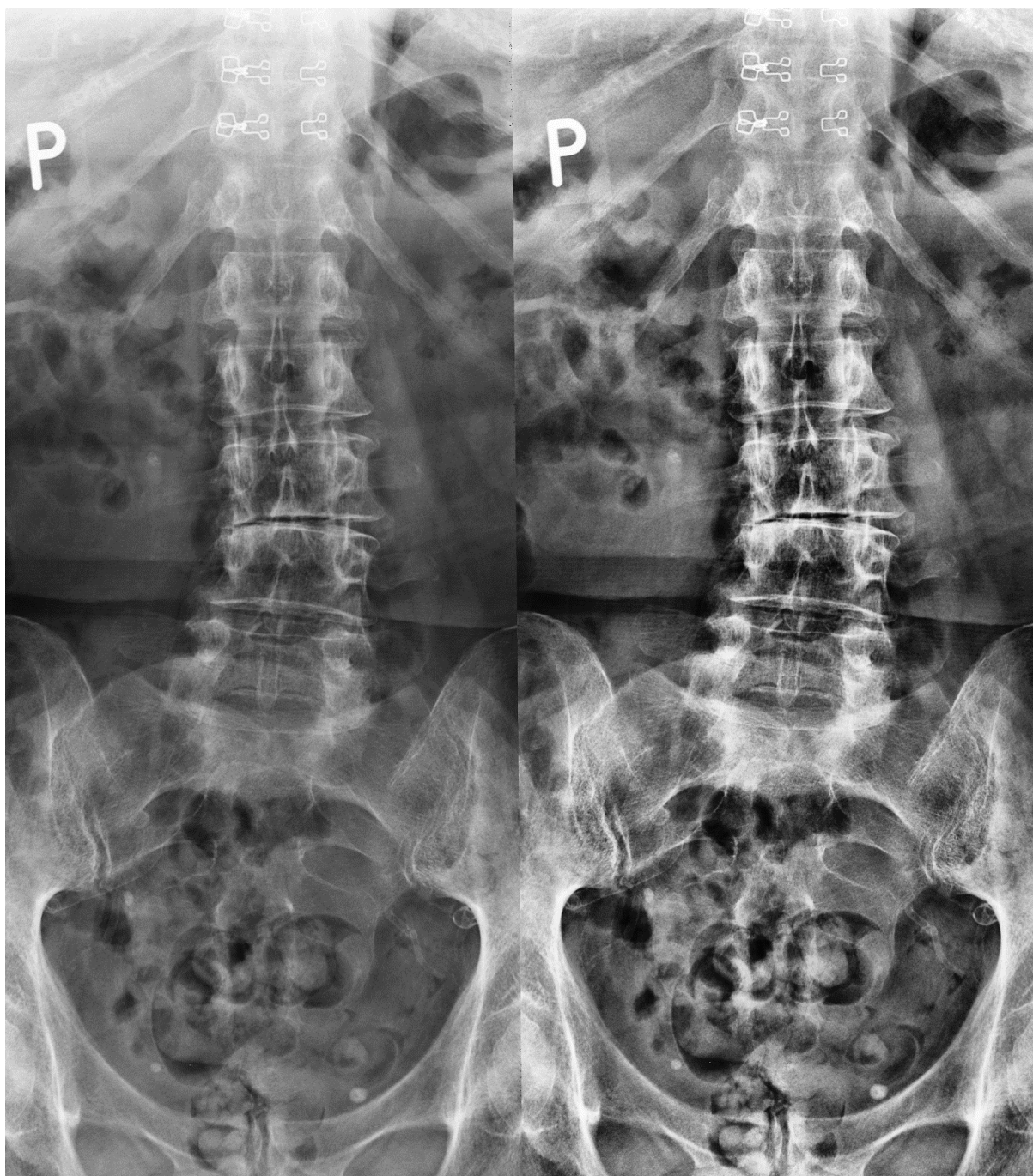
5.1 Dane przygotowane do machine learningu

Jako dane wejściowe otrzymano 1400 zanonimizowanych zdjęć pacjentów z Ortopedyczno-Rehabilitacyjnego Szpitala Klinicznego im. W. Degi w Poznaniu.

Zbiór zdjęć otrzymanych zawierał zdjęcia wykonane aparatem rentgena w różnych płaszczyznach (czołowej lub strzałkowej), różnym przybliżeniu. Zdjęcia podzielono na podkategorie ze względu na przybliżenie i płaszczyznę. Odrzucono wszystkie zdjęcia o niskim kontraście i znaczących zaburzeniach obrazu. Po zakończeniu tych operacji wybrano grupę zdjęć w płaszczyźnie czołowej, zawierających co najmniej 7 kręgów. Wybrana grupa zawierała 120 zdjęć.

W pierwszej kolejności zdjęcia wstępnie przetworzono, stosując adaptacyjną normalizację histogramową z limitem kontrastu (CLAHE). W tej metodzie, każdy piksel jest zmieniany na podstawie histogramu jasności pikseli rejonu otaczającego z odcięciem na określonej granicy amplifikacji. Ustalenie odpowiedniej wartości odcięcia pozwala uniknąć nadmiernego wzmocnienia w częściach obrazu o bardzo dużym nagromadzeniu pikseli o podobnej jasności.

Wstępna obróbka graficzna zdjęć pozwala uzyskać znacznie podwyższone efekty trenowania modeli. Grafika poniżej przedstawia wybrane zdjęcie przed i po normalizacji.



Rysunek 9 zdjęcie wejściowe [po lewej] i znormalizowane

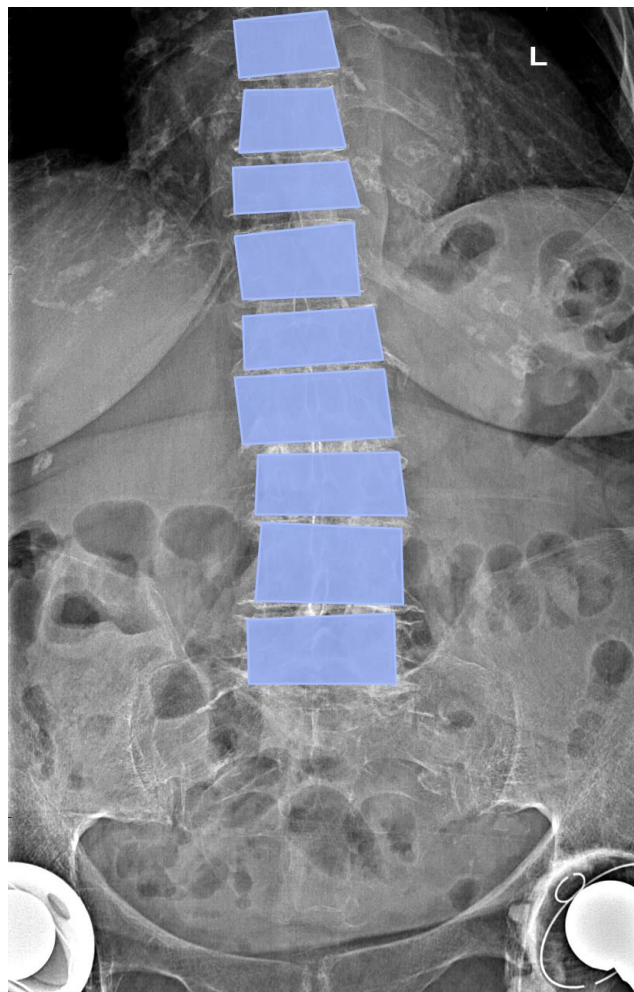
5.2 Etykietowanie i formatowanie zbioru danych

W celu etykietowania danych wykorzystano program Label-studio. Jest to program otwarto źródłowy służący do etykietowania wielu rodzajów danych, takich jak zdjęcia, pliki audio, video czy tekstowe. Wewnątrz programu, możliwe

jest dowolne konfigurowanie interfejsu użytkownika i adaptowanie go pod konkretny rodzaj danych. [25]

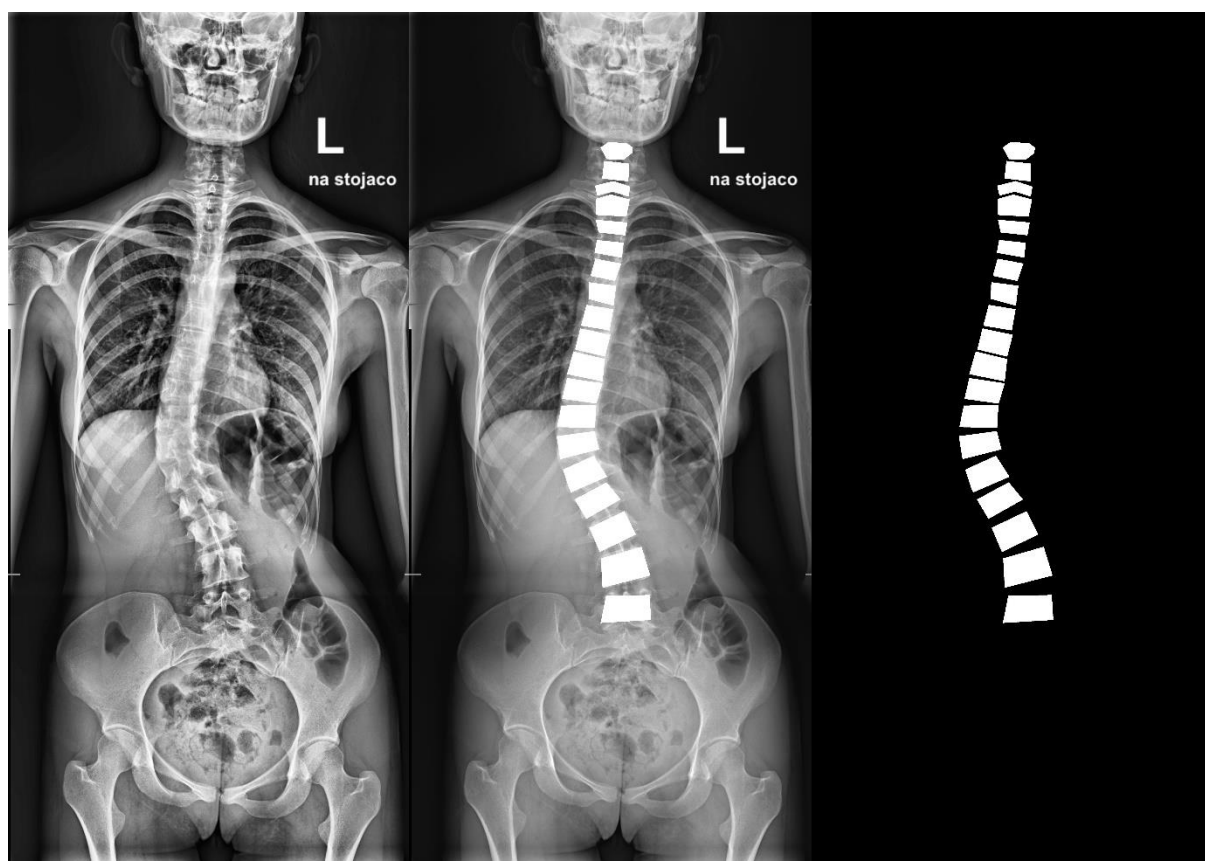
Program skonfigurowano wewnątrz kontenera Docker i przystosowano do tworzenia etykiet wielobokowych potrzebnych do zastosowania modelu F-RCNN.

Dane zostały opisane jako przynależące do jednej z dwóch klas: ‘spine’ oraz ‘background’, a następnie wyeksportowane. Grafika poniżej przedstawia przykładowe zdjęcie z wizualizacją etykiety wielobokowej. Same etykiety przechowywane są w postaci pliku JSON zawierającego współrzędne punktów tworzących kolejne wieloboki.



Rysunek 10 Zdjęcie rentgenowskie z etykietami poligonowymi

Aby wykorzystać model U-Net należy przygotować zbiór danych opisany maskami binarnymi, jednak, aby uniknąć ponownego ręcznego etykietowania danych inną metodą, przygotowano konwerter napisany w języku python. Do stworzenia programu wykorzystano przede wszystkim wbudowane opcje biblioteki OpenCV [26] oraz Numpy [27]. Program odczytuje kolejno każde zdjęcie w zbiorze danych, zmienia wszystkie piksele na wartość 0 w skali szarości, następnie odczytuje plik JSON z etykietami. Wykorzystując pozycje wierzchołków wieloboków zawarte w pliku, rysuje kształty na zdjęciu w wartości 1. Zdjęcie po obróbce jest zapisywane w rozmiarze wejściowym. Uzyskano zbiór danych zawierający maskę binarną dla każdego zdjęcia, odpowiedni do trenowania modelu U-Net, grafika przedstawia przykładowe zdjęcie, jego adnotację wielobokową oraz wytworzoną maskę binarną.



Rysunek 11 (od lewej) zdjęcie wejściowe, etykieta wielobokowa wypełniona, maska binarna

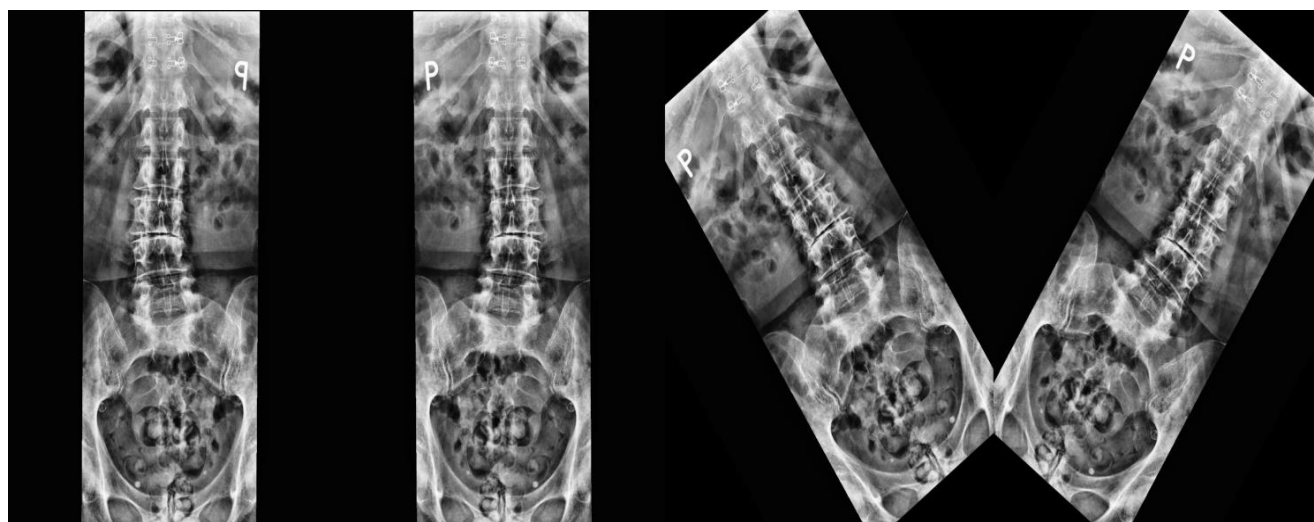
5.3 Opis przeprowadzonych prób

Opisane zbiory danych podzielono na części treningowe, walidacyjne i testowe w proporcjach przedstawionych w tabeli poniżej

Model	Ilość zdjęć treningowych	Ilość zdjęć walidacyjnych	Ilość zdjęć testowych
FRCNN	105	10	5
U-NET	110	5	5

Tabela 1 podział zbioru danych

Przy trenowaniu modeli stosowano augmentacje danych. Przykładowe formy augmentacji: odbicie poziome, skrzywienie o kąt.



Rysunek 12 Przykładowe augmentacje (od lewej): oryginal, odbicie w poziomie, obrót w lewo, obrót w prawo

Do szkolenia F-RCNN wykorzystano obrazy nieskompresowane. W przypadku modelu U-Net zdjęcia przeskalowano do wymiaru 800x1200 pikseli.

W celu wybrania odpowiednich warunków trenowania modelu przeprowadzono wiele prób o zmiennym tempie nauki, rodzaju optymalizacji czy ilości augmentacji danych. Na podstawie tych prób wybrano najbardziej obiecujące warunki dla modeli F-RCNN oraz U-Net, i przygotowano po dwie pętle treningowe o różnych stopniach zwielokrotnienia danych. Wykorzystano

domyślne parametry straty dla F-RCNN, jak opisano w rozdziale 4.3. Jako parametr oceniający skuteczność sieci U-Net wybrano trafność (ang. Accuracy), czyli stosunek ilości poprawnie zaznaczonych pikseli do ilości pikseli na masce.

Dla modelu F-RCNN współczynnik tempa nauki ustawiono na 0.0005. Dla modelu U-Net ten współczynnik ustawiono na 0.0001. W przypadku obu modeli wykorzystano optymalizator ADAM.

5.4 Omówienie efektów pracy modelu Machine Learning

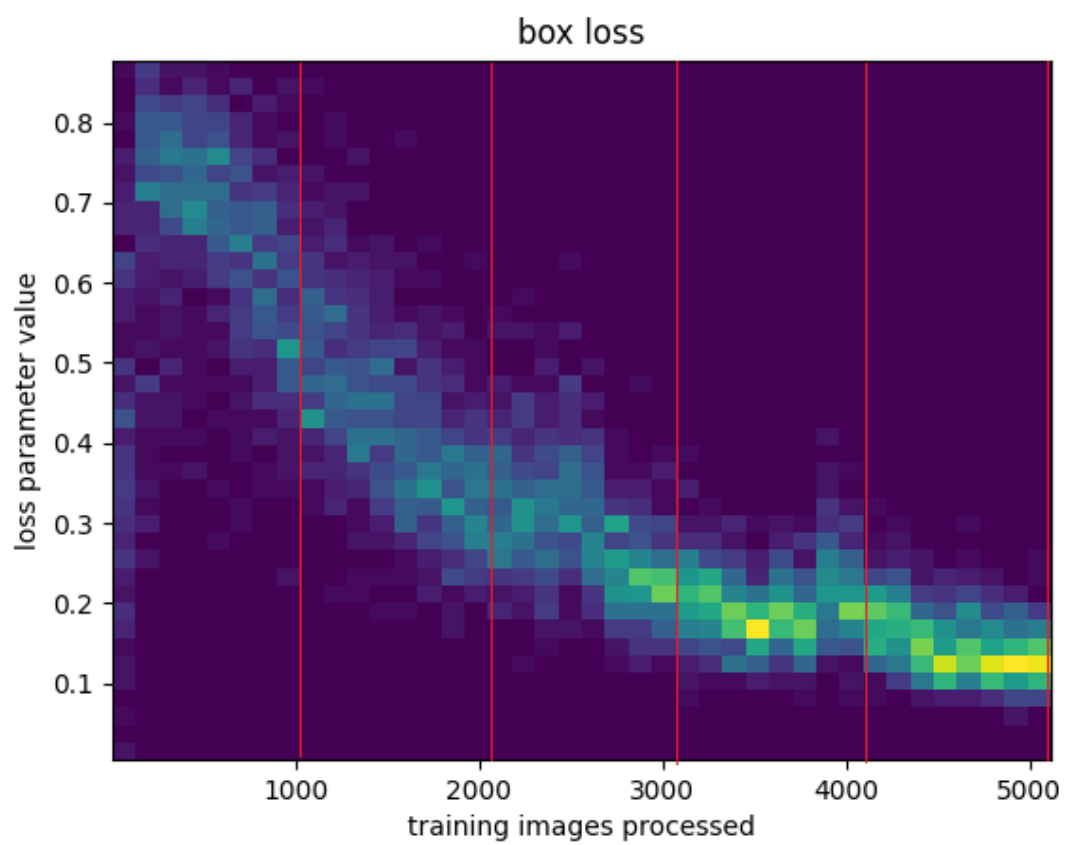
Jak opisano wyżej, modele FRCNN oraz U-Net wykorzystują różne rodzaje danych wejściowych jak i uzyskują różny format danych wyjściowych. Wyniki uzyskane przez model U-Net przedstawiono poniżej w postaci masek binarnych. Regiony uzyskane przez model FRCNN zostały przekształcone na maski binarne w celu ujednolicenia danych i zwiększenia czytelności wyników.

Model FRCNN bez augmentacji danych

Wykres poniżej przedstawia znaczący spadek straty segmentacji zdjęć treningowych. Grafika przedstawia wyniki segmentacji zdjęć testowych modelu po upływie określonej ilości epok trenowania. W tabeli poniżej znajdują się średnie wartości strat na zdjęciach testowych.

Rodzaj straty	Wartość
klasyfikacja	0.0692
regresja obszarów wiążących	0.1421
wykrycie obiektów	0.0019
regresja obszarów wiążących modułu RPN	0.0106

Tabela 2 Średnie wartości strat na zdjęciach testowych

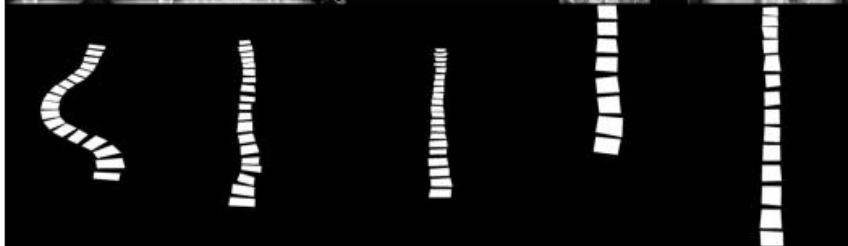


Rysunek 13 Histogram wartości straty dla zdjęć treningowych, czerwoną linią zaznaczono epoki kontrolne

zdjęcie
wejściowe



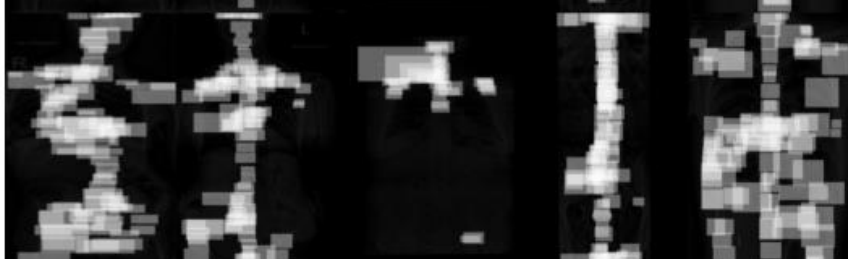
maska
utworzona
ręcznie



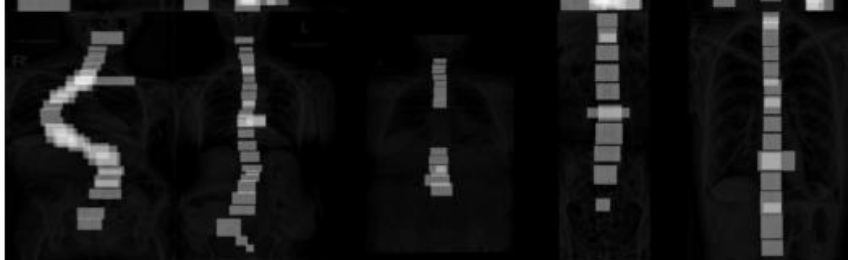
model
trenowany
10 epok



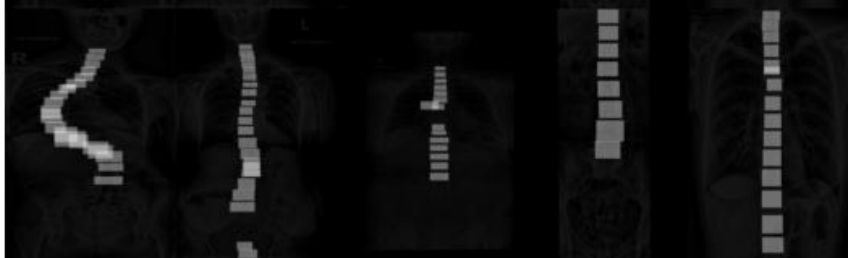
model
trenowany
20 epok



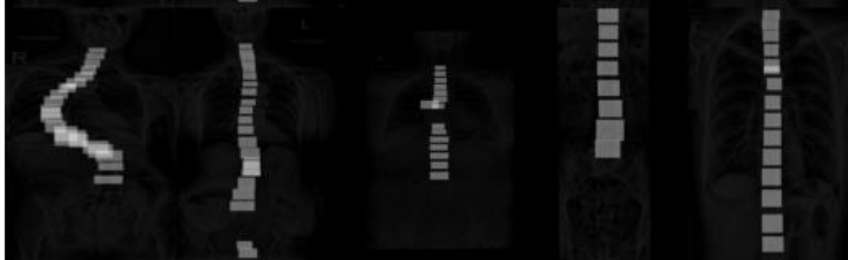
model
trenowany
30 epok



model
trenowany
40 epok



model
trenowany
50 epok



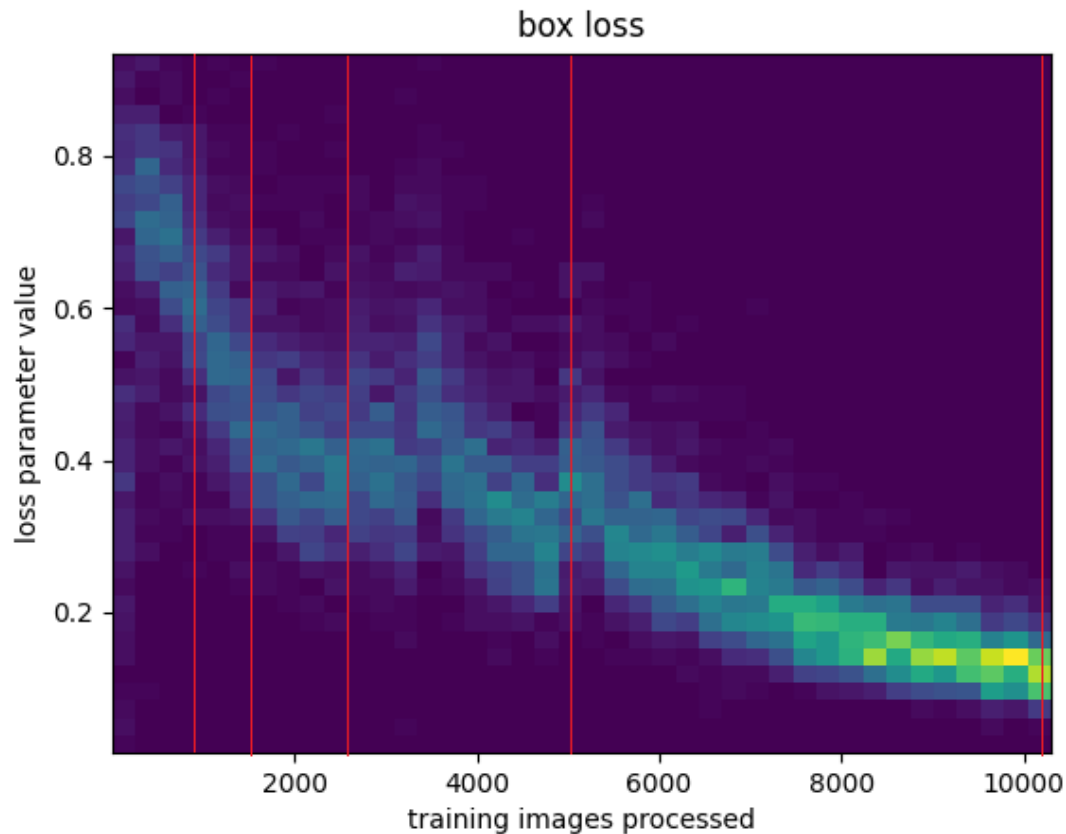
Rysunek 14 maski binarne uzyskane na różnych etapach szkolenia modelu

Model FRCNN przy dwukrotnej augmentacji danych

Dane zwielokrotniono dwukrotnie metodą augmentacji. Wykres przedstawia podobny spadek straty segmentacji zdjęć treningowych jak w przypadku braku augmentacji. Grafika przedstawia wyniki segmentacji zdjęć, która okazała się lepsza niż w przypadku modelu bez augmentacji danych. Wartość straty na zdjęciach testowych również okazała się niższa i została opisana w tabeli poniżej:

Rodzaj straty	Wartość
klasyfikacja	0.0670
regresja obszarów wiążących	0.1300
wykrycie obiektów	0.0010
regresja obszarów wiążących modułu RPN	0.0070

Tabela 3 Średnie wartości strat na zdjęciach testowych modelu z podwójną augmentacją danych

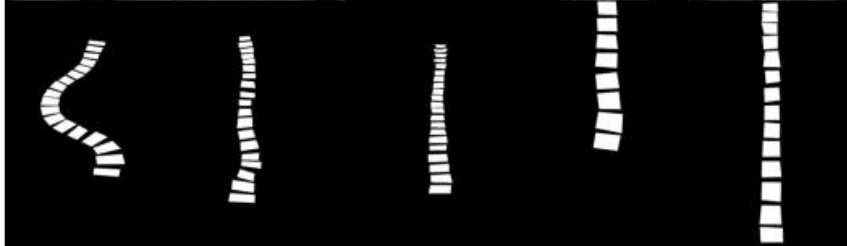


Rysunek 15 Histogram wartości straty dla zdjęć treningowych, czerwoną linią zaznaczono epoki kontrolne

zdjęcie
wejściowe



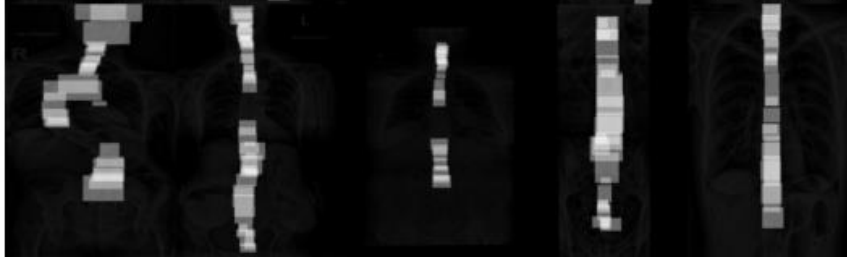
maska
utworzona
ręcznie



model
trenowany
4 epoki



model
trenowany
6 epok



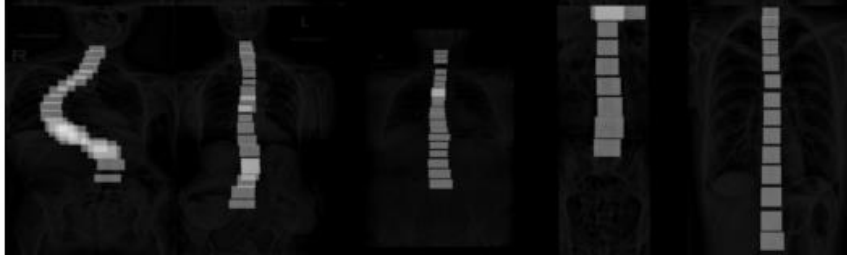
model
trenowany
12 epok



model
trenowany
24 epoki



model
trenowany
50 epok

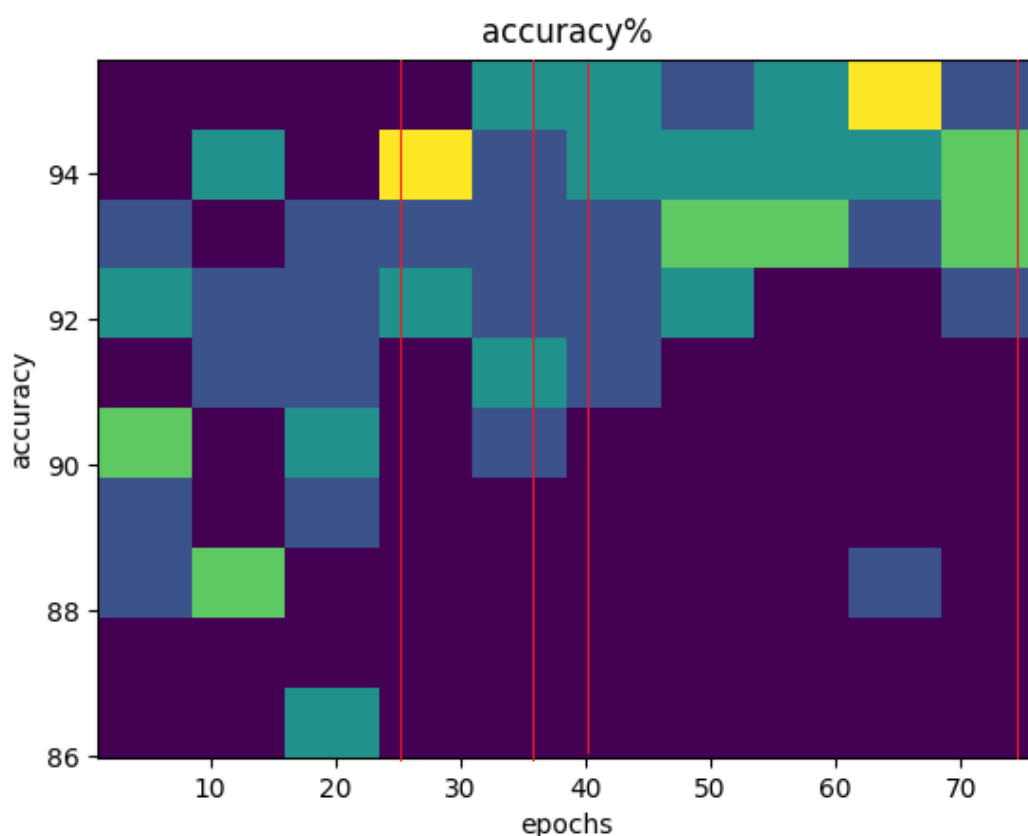


Rysunek 16 maski binarne uzyskane na różnych etapach szkolenia model

W przypadku obu prób modelu F-RCNN obserwujemy spadek straty na zdjęciach treningowych o kształcie zbliżonym do $1/x$ z plateau na poziomie poniżej 0.2. Wytrenowane modele były w stanie precyzyjnie zaznaczyć znaczną większość kręgów. Model o augmentacji danych x2 okazał się czulszy, przez co ma mniejsze ryzyko niezaznaczenia kręgu. Można stwierdzić, że dla modelu F-RCNN zwiększenie augmentacji danych do 2x zwiększyło jego precyzyjność.

Model U-Net przy czterokrotnej augmentacji danych.

Wykres przedstawia wzrost średniej trafności w segmentacji zdjęć treningowych. Grafika przedstawia wyniki segmentacji zdjęć testowych modelu po upływie określonej ilości epok trenowania. Optymalną trafność, na poziomie ~92% obserwujemy pomiędzy 30 a 40 epoką trenowania.

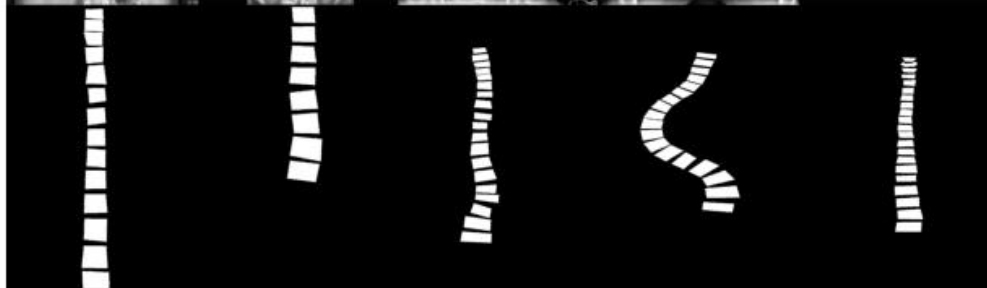


Rysunek 17 trafność modelu dla danych testowych w kolejnych epokach, czerwoną linią zaznaczono epoki kontrolne

Zdjęcie
Wejściowe



Maska
utworzona
ręcznie



Model
trenowany 25
Epok



Model
trenowany 35
Epok



Model
trenowany
40 Epok



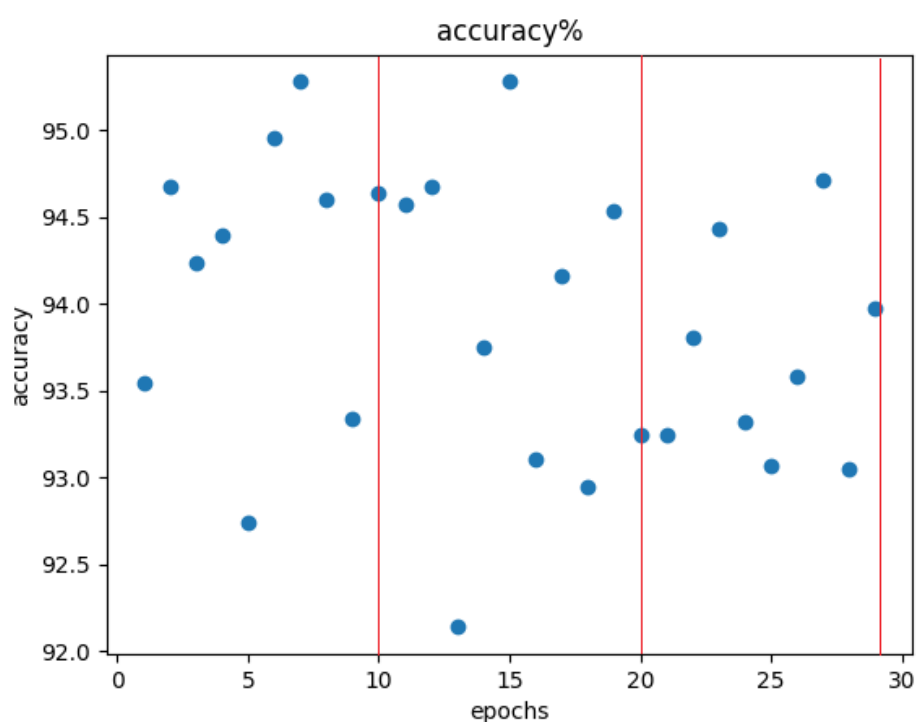
Model
trenowany
75 Epok



Rysunek 18 maski binarne uzyskane na różnych etapach szkolenia model

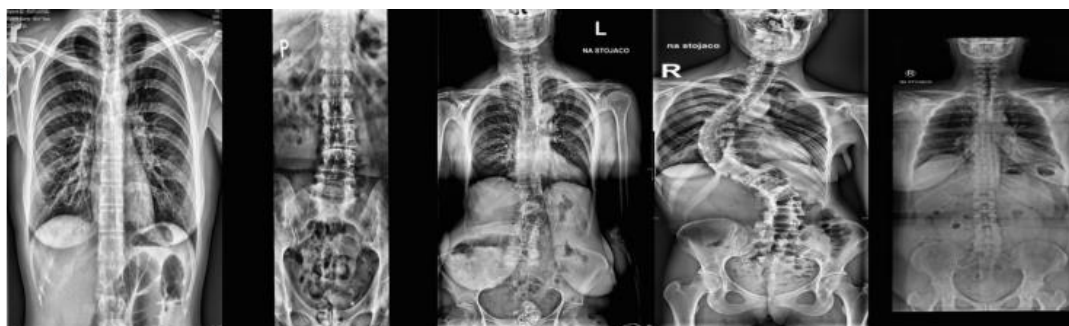
Model U-Net przy dziesięciokrotnej augmentacji danych.

Wykres przedstawia nieregularną średnią trafność o bardzo wysokich wartościach w segmentacji zdjęć treningowych, co sugeruje znaczne przetrenowanie modelu. Grafika przedstawia wyniki segmentacji zdjęć testowych modelu po upływie określonej ilości epok trenowania. Na grafice można zaobserwować efekty przetrenowania, czyli zwiększoną ilość zaznaczonego na zdjęciach szumu.

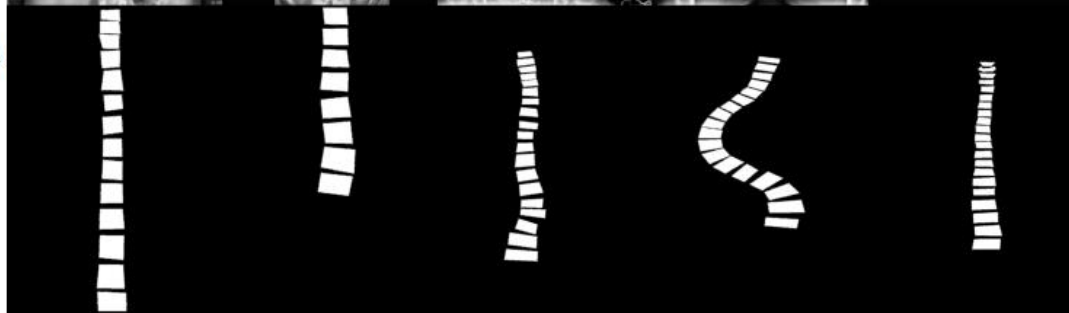


Rysunek 19 trafność modelu dla danych testowych w kolejnych epokach, czerwoną linią zaznaczono epoki kontrolne

Zdjęcie
wejściowe



maska
utworzona
ręcznie



model
trenowany
10 epok



model
trenowany
20 epok



model
trenowany
30 epok



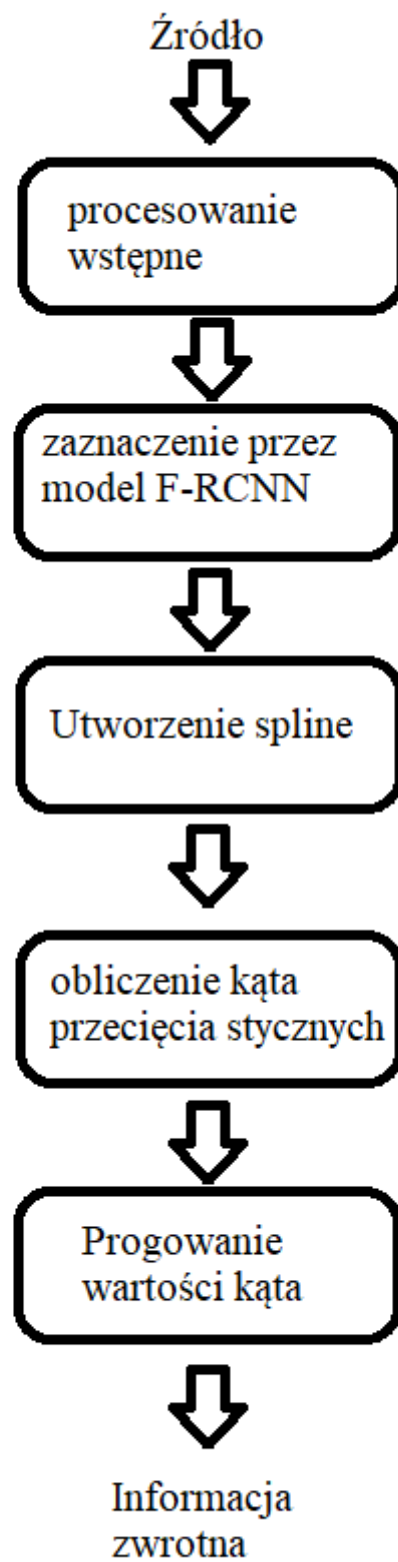
Rysunek 20 maski binarne uzyskane na różnych etapach szkolenia model

Dla próby 3, wykonanej na modelu U-Net, obserwujemy wzrost jakości pracy modelu do epoki 35, jednak po przejściu epok 40>75 obserwujemy znaczny spadek jakości oraz dużą ilość szumu oznaczoną przez model. Biorąc pod uwagę bardzo wysoki współczynnik trafności (>93%) na wykresie (rys. 11) przy jednoczesnym spadku jakości zdjęć, możemy stwierdzić nadmierne dopasowanie modelu po epoce 40-tej. Przy augmentacji 4x i wykorzystanym zbiorze danych, nie udało się uzyskać rezultatów które mogłyby konkurować z uzyskanymi w próbach 1 i 2.

W próbie 4 zwiększono augmentację danych do 10x biorąc pod uwagę stosowanie tego zabiegu przez niektórych badaczy. Już w pierwszej epoce kontrolnej obserwujemy bardzo wysoką trafność treningową, 94,8%, oraz niedokładnie opisane zdjęcia, co sugeruje, że już po pierwszych 10 epokach doszło do nadmiernego dopasowania sieci. Stworzony zbiór danych okazał się być niewystarczająco zróżnicowany, aby skutecznie przeprowadzić tak dużą augmentację.

5.5 Kierunki dalszych prac

Kolejnym krokiem, wykorzystującym efekty tej pracy, może być obliczenie kąta Cobb'a na podstawie segmentacji zdjęć uzyskanej modelem F-RCNN. Efekt ten można uzyskać tworząc krzywą typu spline przechodzącą przez środki obszarów wiążących zaznaczonych na obrazie. Kąt przecięcia linii stycznych do początku krzywizny spline stanowi przybliżenie kąta Cobb'a. Proponowane jest stworzenie systemu gotowego otrzymywać zdjęcia bezpośrednio z aparatu rentgenowskiego, w celu zaznaczenia zdjęć na których prawdopodobnie występuje skolioza. Ogólna architektura proponowanego systemu została przedstawiona na rysunku poniżej:



Rysunek 21 Ogólna architektura systemu wstępnego wykrywania skoliozy ze zdjęć diagnostycznych

6. Wnioski

Cel projektu można uznać za osiągnięty. Wytrenowano skuteczny model służący do segmentacji kręgów na zdjęciach kręgosłupa. Wbrew wstępnym oczekiwaniom, model U-Net dedykowany dla rozwiązań biomedycznych osiągnął znacząco gorsze rezultaty niż model F-RCNN będący znacznie ogólniejszym w zastosowaniu modelem.

Model U-Net bardzo szybko wysycił się informacjami z dostępnego zbioru danych i zaznaczał znaczącą ilość szumów i tła z obrazów. Pomimo zastosowania podobnych parametrów co proponowane w podobnych pracach, uzyskane efekty były znacznie niższej jakości. Jedną z różnic było wykorzystanie innego wzoru obliczenia straty niż proponowany. Wyjaśnienia można również doszukiwać się w rozmiarze zbioru danych i jakości przygotowanych masek. W przypadku wielu projektów wykorzystujących U-Net, zdjęcia opisywane są przez ekspertów w dziedzinie chirurgii kręgosłupa czy radiologii. Można więc stwierdzić, że jakość przygotowanych ręcznie masek okazała się niewystarczająca dla tego modelu.

Model F-RCNN oparty o przeszkolony wcześniej ResNet50 okazał się bardziej odporny na niedokładność przygotowanych masek, a wykorzystanie uczenia transferowego pozwoliło na wyszkolenie modelu używając stosunkowo niewielkiego zbioru danych.

Biorąc pod uwagę powyższe, stwierdzono, że etapem wymagającym największej dokładności i staranności jest przygotowanie zbioru danych. Każdorazowo do zbioru danych należy dopasować model o odpowiednim stopniu złożoności.

7. Bibliografia

- [1] P. Kiepas, „Materiały dydaktyczne AGH,” 2016. [Online]. Available: https://ai.ia.agh.edu.pl/_media/pl:dydaktyka:mbn:uczenie_maszynowe.pdf.
- [2] „strona Natoional Scoliosis Foundation,” 2007. [Online]. Available: <https://www.scoliosis.org/info.php>.
- [3] J. Wilczyński, Boczne skrzywienia kręgosłupa. Rozpoznanie i korekcja. Część 1., Kielce: Wszechnica Świętokrzyska, 2000.
- [4] A. K. Greiner, „Adolescent idiopathic scoliosis: radiologic decision-making,” *Am Fam Physician*, tom 65, p. 9, maj 2002.
- [5] M. Forouzanfar, N. Forghani i M. Teshnehlal, „Parameter optimization of improved fuzzy c-means clustering algorithm for brain MR image segmentation,” *Engineering Applications of Artificial Intelligence*, nr 23/2, pp. 160-168, 2010.
- [6] O. Ronneberger , P. Fischer i T. Brox, „U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arXiv:1505.04597*, 2015.
- [7] „Towards Data Science - basics of neural networks,” Medium, [Online]. Available: <https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a>.
- [8] H. Singh, „Analytics Vidhya,” [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/03/basics-of-neural-network/>.
- [9] K. Fukushima, „Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, nr 36, pp. 193-202, kwiecień 1980.
- [10] „Databricks - convolutional layer,” 20 styczeń 2024. [Online]. Available: <https://www.databricks.com/glossary/convolutional-layer>.

- [11] „Machine Learning Mastery,” [Online]. Available: <https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/>.
- [12] K. O’Shea i R. Nash, „An Introduction to Convolutional Neural Networks,” 2015.
- [13] „Towards Data Science - Transposed convolutional layers,” Medium, [Online]. Available: <https://towardsdatascience.com/what-is-transposed-convolutional-layer-40e5e6e31c11>.
- [14] „The AI Dream,” [Online]. Available: <https://www.theaidream.com/post/loss-functions-in-neural-networks>.
- [15] „Dokumentacja Keras.io, Optymalizator ADAM,” 2024. [Online]. Available: <https://keras.io/api/optimizers/adam/>.
- [16] J. West, „Brigham Young University, College of Physical and Mathematical Sciences,” 01 Sierpień 2007. [Online]. Available: <https://web.archive.org/web/20070801120743/http://cpms.byu.edu/spring-research/abstract-entry?id=861>.
- [17] Y. Kim, B. Ganbold i K. G. Kim, „Web-Based Spine Segmentation Using Deep Learning in Computed Tomography Images,” *Healthcare informatics research*, nr 26, pp. 61-67, 2020.
- [18] M. U. Saeed, N. Dikaio i A. Dastgir, „An Automated Deep Learning Approach for Spine Segmentation and Vertebrae Recognition Using Computed Tomography Images,” *Diagnostics*, p. 13, 2023.
- [19] S. Ma, Y. Huang, X. Che i R. Gu, „Faster RCNN-based detection of cervical spinal cord injury and disc degeneration,” *Journal of Applied Clinical Medical Physics*, 14 sierpień 2020.
- [20] S. A. ElGhany, M. R. Ibraheem, M. Alruwaili i M. Elmogy, „Diagnosis of Various Skin Cancer Lesions Based on Fine-Tuned,” *Computers, Materials & Continua*, nr 68(1), pp. 117-135, 2021.
- [21] „Towards Data Science,” Medium, [Online]. Available: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>.
- [22] L. Shang-Long , L. Shuo i G. Yu-Ting, „Establishment and application of an artificial intelligence diagnosis system for pancreatic cancer with a

faster region-based convolutional neural network,” *Chinese Medical Journal*, 2019.

- [23] „Dokumentacja SAS,” [Online]. Available: https://documentation.sas.com/doc/en/pgmsascdc/v_047/casdlpg/p1np8zbnoyd0brn1dhehthuuxj4q.htm#p08u3e2irvruy1n1hfrs439dex67.
- [24] S. Ren, K. He, R. Girshick i J. Sun, „Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” 2015.
- [25] „Dokumentacja programu Label-Studio,” HumanSignal, [Online]. Available: <https://labelstud.io/guide/>.
- [26] „Dokumentacja biblioteki OpenCV,” OpenCV team, [Online]. Available: <https://docs.opencv.org/3.4/index.html>.
- [27] „Dokumentacja biblioteki Numpy,” NumPy Developers, [Online]. Available: <https://numpy.org/doc/stable/>.
- [28] F. Hefti, *Pediatric Orthopedics in Practice*, ISBN: 9783540699644, 2007.
- [29] „Zbiór danych Verse2020,” 2020. [Online]. Available: <https://verse2020.grand-challenge.org>.
- [30] M.-H. Horng, „Cobb Angle Measurement of Spine from X-Ray Images Using Convolutional Neural Network,” *Computational and mathematical methods in medicine*, 19 luty 2019.