



Politechnika Łódzka
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej

Daria Rogowska
249989

PRACA DYPLOMOWA
magisterska
na kierunku Informatyka Stosowana

**Neuronowe i klasyczne techniki segmentacji w planowaniu
radioterapii**

Instytut Informatyki I72

Promotor: dr inż. Arkadiusz Tomczyk

Opiekun pomocniczy: dr n. med. Piotr Mężeński

ŁÓDŹ 2024

Spis treści

Streszczenie	1
Streszczenie w języku angielskim	2
1 Problematyka i zakres pracy	3
2 Wprowadzenie teoretyczne	4
2.1 Medyczne formaty plików	4
2.2 Segmentacja	4
2.2.1 Segmentacja atlasowa	7
2.2.2 Segmentacja neuronowa	11
2.2.3 Miary jakości	22
2.3 Przegląd literatury	24
2.4 Przegląd istniejących narzędzi	26
3 Opis rozwiązania	28
3.1 Wstępne przetwarzanie	29
3.2 Segmentacja właściwa	36
4 Wstępne eksperymenty	39
4.1 Zbiór danych	39
4.2 Lewa przednia zastępująca tętnica wieńcowa (LAD)	41
4.3 Loża po guzie	48
5 Wyniki	50
5.1 Segmentacja LAD	50
5.2 Segmentacja loży	57
5.3 Analiza porównawcza wyników	59
6 Podsumowanie i perspektywy rozwoju	63
Literatura	65
Spis rysunków	69
Spis tabel	72

Streszczenie

Tematem niniejszej pracy magisterskiej jest zbadanie możliwości wykorzystania klasycznych i neuronowych technik automatycznej segmentacji w planowaniu radioterapii. W tym celu przeprowadzono analizę literatury przedmiotu oraz wykonano szereg eksperymentów. Za podmiot badań wybrano lewą przednią zastępującą tętnicę wieńcową (LAD) oraz loże po guzie nowotworowym piersi. Badania zostały przeprowadzone na zestawach tomografii komputerowej bez kontrastu, udostępnionych przez Wojewódzkie Wielospecjalistyczne Centrum Onkologii i Traumatologii im. M. Kopernika w Łodzi.

Praca przedstawia metody atlasową, neuronową, nowe podejście do połączenia obu tych metod (atlasowo-neuronowe) oraz technikę będącą ich agregacją. Nacisk został położony na proces wstępnego przetwarzania skanów, polegający na zawężeniu obszaru poszukiwań. Podejście to zostało porównane z segmentacją danych ze skanu całego ciała pacjenta, w kontekście danej metody.

W ramach części praktycznej przygotowane zostały implementacje algorytmów omawianych technik. Uzyskane wyniki miar jakości porównano z rozwiązaniami dostępnymi w literaturze. Bazując na tym zestawieniu, uznano otrzymane rezultaty za zadowalające. W przypadku segmentacji LAD osiągnięto wyniki dokładniejsze od tych dostępnych w publikacjach, zaś dla loży po guzie przygotowane rozwiązanie okazało się mniej precyzyjne.

Słowa kluczowe:

segmentacja obrazów medycznych · metoda atlasowa · 3D U-Net · lewa przednia zastępująca tętnica wieńcowa (LAD) · loża po guzie nowotworowym piersi

Abstract

The aim of this master's thesis is to explore the possibilities of using classic and neural techniques of automatic segmentation in radiotherapy planning. For this purpose, an analysis of the literature on the subject was carried out and a number of experiments were performed. The left anterior descending artery (LAD) and the breast tumor bed were selected as the subject of this study. The experiments were carried out on non-contrast computed tomography sets provided by the Copernicus Regional Multidisciplinary Centre for Oncology and Traumatology in Łódź.

The paper presents the atlas-based and neural methods, a new approach to combining both of these methods (atlas-neuronal) and the technique that is their aggregation. The focus was on the process of pre-processing the scans, consisting in narrowing down the search area. This approach was compared with segmentation of data from a patient's whole body scan, in the context of method.

As part of the practical part, implementations of algorithms of the discussed techniques were prepared. The obtained quality measurement results were compared with solutions available in the literature. Based on this, the results were considered satisfactory. In the case of LAD segmentation, the results were more accurate than those available in the publications, while for the tumor's beds, the prepared solution turned out to be less precise.

Keywords:

medical images segmentation · atlas-based method · 3D U-Net · left anterior descending artery (LAD) · breast tumor bed

1 Problematyka i zakres pracy

Radioterapia jest metodą leczenia z użyciem promieniowania jonizującego. Skutecznie wykorzystywana jest w onkologii do leczenia chorób nowotworowych, łagodzenia bólu i innych dolegliwości związanych z przebiegiem choroby np. przy przerzutach do kości. Poprzez napromienianie określonych obszarów pacjenta wiązką jonizującą dochodzi do uszkodzeń DNA komórek nowotworowych, na skutek czego zahamowany zostaje proces podziału i dochodzi do ich obumarcia. Niewątpliwie samodzielna radioterapia lub jako leczenie skojarzone skutkuje zwiększeniem liczby wyleczeń.

Przygotowanie pacjenta do poddania radioterapii składa się z kilku etapów. Jednym z nich jest proces planowania polegający na oznaczeniu obszarów, które mają otrzymać dawkę promieniowania np. guzy nowotworowe, otaczające je tkanki prawidłowe oraz narządy krytyczne, w których należy minimalizować zdeponowaną dawkę promieniowania jonizującego. Planowanie dokonywane jest na podstawie wyników badania tomografii komputerowej (CT), gdzie specjalista skan po skanie określa dane struktury na przekrojach. Proces ten jest niezwykle czasochłonny, gdyż w zależności od zakresu tomografii komputerowej należy manualnie zdefiniować struktury nawet na kilkuset przekrojach skanu.

Automatyzacja segmentacji struktur anatomicznych może pozwolić na przyśpieszenie etapu planowania, a czasami także poprawić jego dokładność. Z drugiej strony stanowi to wyzwanie z racji subiektywności oznaczeń eksperckich, które w zależności od specjalisty prezentują się inaczej.

Celem niniejszej pracy jest zbadanie możliwości wykorzystania sztucznych sieci neuronowych oraz technik klasycznej segmentacji w automatycznym wyznaczeniu struktur używanych w planowaniu radioterapii. Zaproponowane podejście do segmentacji struktur anatomicznych w założeniu powinno stanowić pomocne narzędzie dla radioterapeuty, aby usprawnić tenże proces. Praca za cel stawia wyodrębnienie ze zdjęć tomografii komputerowej lewej przedniej zastępującej tętnicy wieńcowej (LAD), która jest narzędziem krytycznym wymagającym obniżenia dawki promieniowania w trakcie przygotowania planu leczenia, ze względu na ryzyko wystąpienia późnego powikłania popromiennego w postaci zawału mięśnia sercowego. Drugim celem pracy była próba automatycznego wyznaczenia na przekrojach z tomografii komputerowej loży po wyciętym guzie nowotworowym. Loża oraz gruczoł piersiowy stanowią obszary poddane uzupełniającej radioterapii po zabiegu chirurgicznym.

Na rynku dostępnych narzędzi wspomagających proces planowania radioterapii istnieją komercyjne oprogramowania. Oferują one automatyczną segmentację większych struktur anatomicznych takich jak serce, lecz nie umożliwiają wyodrębnienia LAD, gruczołu piersiowego czy loży. Również, po konsultacji z radioterapeutą została wskazana luka w częściowo automatycznym procesie planowania radioterapii leczenia chorób nowotworowych. Dlatego też, wcześniej wspomniane struktury biomedyczne zostały wybrane jako podmiot dalszych badań.

2 Wprowadzenie teoretyczne

2.1 Medyczne formaty plików

DICOM (ang. Digital Imaging and Communications in Medicine) to międzynarodowy standard przetwarzania, przechowywania, drukowania i przesyłania informacji w obrazowaniu medycznym opracowany przez American College of Radiology i National Electrical Manufacturers Association [dic, 2024]. Powstał on w celu ujednolicenia wymiany i interpretacji obrazów diagnostycznych, stąd też jest powszechnie stosowany w placówkach medycznych na całym świecie.

Format ten standardowo pozwala na przechowywanie obrazów medycznych z jednoczesnym zawarciem metadanych. Dane obrazu mogą obejmować różne rodzaje obrazów medycznych, takie jak zdjęcia rentgenowskie, tomografia komputerowa (CT), rezonans magnetyczny (MRI), czy obrazy ultradźwiękowe. Metadane obejmują informacje o pacjencie, parametry obrazowania, szczegóły akwizycji i inne istotne informacje. Są one niezbędne do właściwej interpretacji obrazów medycznych i zarządzania nimi.

Pliki DICOM mają specyficzną strukturę składającą się z nagłówka i danych obrazu. Nagłówek zawiera serię znaczników z metadanymi zakodowanymi w formacie hierarchicznym. Każdy znacznik ma unikalny identyfikator i określa konkretny atrybut obrazu lub badania.

Skany zapisywane są w formie tablicy wartości wraz z informacją o rzeczywistych odległościach pomiędzy sąsiadującymi pikselami, grubości przekroju, orientacji i położenia w przestrzeni, ale również zawiera dane wrażliwe samego pacjenta.

W przypadku obrazowania CT wartości wokseli są powiązane z linowym współczynnikiem osłabienia promieniowania, wyrażonym w jednostkach Hounsfielda (HU). Jako punkt odniesienia przyjmuje się wartość dla powietrza w standardowych warunkach, która wynosi około -1000 HU oraz wartość dla wody destylowanej wynoszącą 0 HU. Przykładowo, osłabienie promieniowania dla tłuszczu przyjmuje mniej więcej wartość -100 HU, zaś dla kości około +1000 HU.

Obrazy DICOM są zazwyczaj przeglądane przy użyciu specjalistycznego oprogramowania przeglądarki DICOM, które umożliwia wizualizację, rekonstrukcję i analizowanie obrazów medycznych za pomocą różnych narzędzi i funkcji. Przykładem oprogramowania będącego w stanie wizualizować dane zawarte w plikach formatu dcm jest 3DSlicer czy ITK-SNAP.

Co więcej, DICOM zawiera funkcje zapewniające poufność i bezpieczeństwo danych pacjenta, takie jak szyfrowanie i kontrola dostępu, aby zapewnić zgodność z przepisami dotyczącymi prywatności w służbie zdrowia.

2.2 Segmentacja

Segmentacja obrazu definiowana jest jako separacja elementów tworzących obraz na konkretne obszary (regiony), które są jednorodne pod względem pewnych wybranych własności. Obszary

te stanowią zbiory wokseli w przypadku obrazu trójwymiarowego. Często jest to jeden z pierwszych kroków prowadzących do analizy i interpretacji danych wizualnych.

Segmentacja semantyczna, z technicznego punktu widzenia, to proces przypisywania każdemu wokselowi w obrazie etykiety, która reprezentuje klasę obiektu lub regionu, do którego dany woksel należy. Matematyczna definicja segmentacji semantycznej może być wyrażona w następujący sposób:

Niech \mathcal{I} będzie obrazem złożonym z wokseli $\mathcal{I}(x, y, z)$, gdzie (x, y, z) oznacza jego współrzędne. Celem segmentacji semantycznej jest przypisanie każdemu z nich (x, y, z) etykiety z ustalonego zbioru $\mathcal{L} = \{0, 1, \dots, R - 1\}$, gdzie R oznacza liczbę klas (etykiet). W tym celu szukany jest segmentator, który to najlepiej mapuje wartości wokseli na daną klasę.

$$\mathcal{S} : \mathcal{I} \rightarrow \hat{P}, \quad (1)$$

gdzie \mathcal{I} to obraz o wymiarach $m \times n \times o$, a \hat{P} to mapa przewidywanych etykiet o tych samych wymiarach, przypisująca każdemu wokselowi odpowiednie oznaczenie.

Segmentator generuje macierz $P \in [0; 1]^{m \times n \times o \times R}$, w przypadku binarnej segmentacji, gdzie $p_{x,y,z,r}$ stanowi prawdopodobieństwo uznania danego woksela z lokalizacji (x, y, z) jako przynależącego do klasy r . Przewidywaną mapę segmentacji \hat{P} uzyskuje się poprzez przypisanie każdemu pikselowi klasy najwyższego prawdopodobieństwa:

$$\hat{p}_{x,y,z} = \arg \max_r p_{x,y,z,r} \quad (2)$$

W celu znalezienia najlepszych parametrów segmentatora korzysta się z funkcji straty, zazwyczaj stosuję się w tym celu entropię krzyżową pomiędzy rzeczywistą mapą obiektu a tą przewidywaną. Entropia krzyżowa dla pojedynczego woksela obliczana jest następująco:

$$\mathcal{E}_{x,y,z} = - \sum_{r=0}^{R-1} 1_{\{s_{x,y,z}=r\}} \log(p_{x,y,z,r}), \quad (3)$$

gdzie $1_{\{s_{x,y,z}=r\}}$ jest funkcją przypisującą wartość 1 jeśli etykieta woksela r w punkcie (x, y, z) zgadza się z rzeczywistym oznaczeniem, w przeciwnym wypadku wartość ta wynosi 0. Całkowita wartość funkcji straty dla całego badanego obrazu jest średnią wszystkich strat poszczególnych pikseli.

Istnieje wiele podejść do segmentacji obrazu, wśród których można wyróżnić podział na tak zwane techniki klasyczne, polegające na analizie obrazu i techniki uczenia głębokiego wykorzystujące sztuczne sieci neuronowe.

Progowanie, metoda klasycznej segmentacji, to jedna z najprostszych, w której określana jest wartość graniczna pozwalająca na oddzielenie wokseli o wartościach intensywności powyżej lub poniżej progu. Wykorzystując jedną wartość progu t , każdy piksel obrazu jest klasyfikowany

jako należący do obiektu (lub tła) w zależności od tego, czy jego wartość jest większa (lub mniejsza) niż ten próg. Niech $I(x, y, z)$ będzie wartością wokseli w obrazie trójwymiarowym, w punkcie (x, y, z) . W przypadku progowania względem jednej wartości granicznej maska binarna $M(x, y, z)$ jest definiowana jako:

$$M(x, y, z) = \begin{cases} 1 & \text{gdy } I(x, y, z) \geq t \\ 0 & \text{gdy } I(x, y, z) < t \end{cases} \quad (4)$$

Tutaj wartość 1 oznacza, że woksel jest częścią obiektu a 0, że należy do tła. Metoda ta jest prosta w implementacji i szybka w działaniu, jednakże jest niezbyt skuteczna w przypadku obrazów o nierównomiernym oświetleniu lub złożonym tle oraz wymaga odpowiedniego doboru progu, co może być trudne w praktyce.

Segmentacja bazująca na krawędziach to z kolei technika segmentacji obrazu, która identyfikuje obszary poprzez wykrywanie krawędzi, czyli miejsc w obrazie, gdzie następują nagłe zmiany jasności. Krawędzie często reprezentują granice obiektów, co czyni tę metodę użyteczną do podziału obrazu na segmenty odpowiadające różnym obiektom lub regionom. Zasada działania tego podejścia składa się z kilku kroków: pierwszym etapem jest identyfikacja krawędzi w obrazie. Jest to realizowane za pomocą operatorów detekcji krawędzi, takich jak: operator Sobela (wykrywa krawędzie poprzez obliczenie gradientu intensywności jasności w poziomie i pionie), czy operator Canny'ego (bardziej zaawansowana metoda wykrywania krawędzi, która obejmuje etapy wygładzania, wyznaczania gradientu, lokalizowania krawędzi i eliminacji tych, które nie są lokalnymi maksymami). Wynikiem etapu wykrywania jest mapa – obraz binarny, w którym piksele oznaczone jako krawędzie mają wartość 1, a pozostałe piksele wartość 0. Zidentyfikowane krawędzie są następnie łączone w zamknięte kontury, które następnie definiują granice segmentów. Segmentacja polega na przypisaniu pikseli wewnętrz konturów do odpowiednich klas, które mogą reprezentować obiekty na obrazie.

Podejście to dobrze działa w przypadku obrazów z wyraźnymi i dobrze zdefiniowanymi granicami obiektów, a także jego zaletą jest szybkość i prostota obliczeń w porównaniu do innych bardziej zaawansowanych metod segmentacji. Niemniej jednak krawędzie mogą być słabo zdefiniowane w obrazach o niskim kontraste lub z dużą ilością szumów, co może prowadzić do generowania błędnych krawędzi, a w konsekwencji również do nieprawidłowej segmentacji. Bardzo złożone obrazy o małej rozdzielczości mogą również stanowić wyzwanie dla tej metody. Co więcej, wykryte krawędzie mogą nie tworzyć zamkniętych konturów, utrudniając tym samym segmentację.

Na szczególną uwagę zasługuje metoda bazująca na atlasach, stosowana wyłącznie w zadaniach segmentacji struktur anatomicznych. Podejście to zostało szczegółowo opisane w dalszej części pracy.

Kolejną grupę metod segmentacji stanowią rozwiązania wykorzystujące sztuczne sieci neuronowe, w szczególności splotowe. Są to z reguły rozwiązania powstałe na przestrzeni ubiegłych

lat, lecz pozwalają w większości przypadków na szybsze i dokładniejsze segmentowanie niż robią to rozwiązania klasyczne. Szczegółowy opis tej techniki zawarty jest w dalszej części pracy.

Wykorzystywanie segmentacji w medycynie jest powszechnie ze względu na jej zdolność do automatycznej analizy obrazów medycznych, co przynosi liczne korzyści w diagnostyce, planowaniu leczenia oraz monitorowaniu stanu zdrowia pacjentów. Niemniej jednak segmentacja semantyczna znajduje zastosowanie wszędzie tam, gdzie jest dokonywana analiza obrazów np. w robotyce do nawigacji robotów, w systemach wizyjnych pozwalając na identyfikację obiektów czy w samochodach autonomicznych.

2.2.1 Segmentacja atlasowa

Segmentacja obrazów medycznych jest trudna do przeprowadzenia ze względu na problemy związane z obrazami o niskim kontraście, rozmytymi konturami obiektów, czy podobnymi natężeniami z sąsiednimi strukturami badanego narządu. Korzystanie z anatomicznej wiedzy ciała człowieka może pomóc w wyodrębnieniu badanej struktury. Segmentacja bazująca na atlasach prezentuje tego typu podejście.

Atlasem nazywamy obraz pacjenta stanowiący wzorcową referencję umiejscowienia konkretnych narządów wewnętrznych. Może to być zarówno skan danego pacjenta, jak i obraz będący złączeniem i uśrednieniem wyników obrazowania grupy pacjentów wykorzystując w tym celu dedykowane temu narzędziu. Jednym z nich jest wspomniane w przeglądzie literatury oprogramowanie Mirada RTx.

Metody bazujące na atlasach (ang. atlas-based segmentation) swą zasadę działania opierają na dopasowaniu atlasu, zawierającego maski lokalizujące dane struktury anatomiczne, do skanów obrazów pacjentów, dla których celem jest uzyskanie maski segmentacji. Proces ten można podzielić na dwa etapy — pierwszy z nich tzw. etap rejestracji, który polega na przekształceniu atlasu do obrazu danego pacjenta. Szukana jest transformacja lub złożenie kilku, po zastosowaniu których na skanie atlasu otrzymany zostanie obraz najbardziej zbliżony do skanu badanego pacjenta. Drugi etap stanowi przekształcenie maski atlasu stosując w tym celu wcześniej uzyskaną transformację,

Proces, w wyniku którego uzyskuje się transformację atlasu do obrazu pacjenta, można określić jako zadanie optymalizacyjne, gdzie funkcją celu jest różnica obrazu pacjenta do przetransformowanego atlasu.

$$\hat{\tau}_{p,a} = \arg \min_{\tau \in T} d(I_a \circ \tau, I_p), \quad (5)$$

gdzie I_a , I_p określają odpowiednio obraz modyfikowanego atlasu oraz pacjenta, T to zbiór możliwych transformacji, zaś d stanowi funkcję służącą do obliczenia różnicy pomiędzy obrazami, inaczej nazywaną funkcją straty. Gdy przyjmiemy oznaczenie L_a jako mapę segmentacji

atlasu, to zdefiniowanie maski segmentacji pacjenta wygląda następująco:

$$L_p(x) = (L_a \circ \hat{\tau}_{p,a})(x) \quad (6)$$

gdzie x to określa współrzędne punktu w obrazie.

W celu znalezienia odpowiednio optymalnej transformacji stosuje się wybraną funkcję straty. Jedną z prostszych jest funkcja sumy kwadratów różnic (SSD) obliczana następująco:

$$SSD(I_a, I_p) = \sum_{i=0}^n \sum_{j=0}^m \sum_{k=0}^o (I_a(i, j, k) - I_b(i, j, k))^2, \quad (7)$$

$n \times m \times o$ określa wymiary porównywanych obrazów, zaś $I_a(i, j, k), I_b(i, j, k)$ to wartości pikseli na pozycjach (i, j, k) .

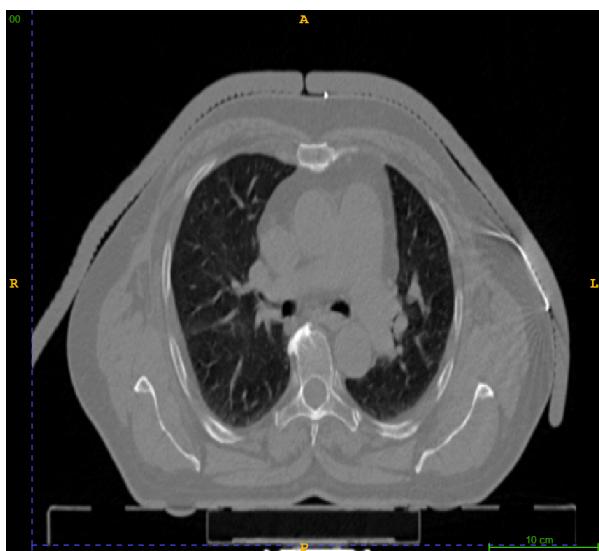
Wracając do wcześniejszej wspomnianego etapu transformacji, istnieje kilka podejść do tego tematu. Najbardziej popularnym rozwiązaniem jest stosowanie transformacji afanicznej, która zachowuje stosunki odległości między punktami na tej samej linii, jednak nie utrzymuje kątów między prostymi ani odległości między punktami. Jest to przekształcenie geometryczne przestrzeni euklidesowych, odwzorowujące odcinki na odcinki, proste w proste, płaszczyzny w płaszczyzny, linie równoległe w linie równoległe. Transformacja afaniczna stanowi złożenie translacji, obrotu i rozciągnięcia. Przykład zastosowania tej transformacji został przedstawiony na rysunku 2.2.

Innego typu transformacją jest transformacja elastyczna, która to jako wynik dostarcza mapę przemieszczeń w przestrzeni skanu rozważanego pacjenta, gdzie każdy element zawiera trzy składowe t_x, t_y, t_z , które pozwalają na dopasowanie woksela z atlasu do tego z obrazu pacjenta w danym punkcie. Tym samym, oznacza to że w porównaniu do transformacji afanicznej metoda elastycznego rejestracji pozwala na dokładniejsze odwzorowanie obrazu z atlasu. Przekształcenie elastyczne zostało zobrazowane na rysunku 2.3.

W celu minimalizacji ilości parametrów transformacji można wyznaczać wartości tylko w równomiernie rozłożonych punktach kontrolnych, bez obliczeń dla każdego woksela z obrazu. Do interpolacji stosuje się funkcję B-sklejanych, jest to rodzaj funkcji bazowych stosowanych w aproksymacji krzywych i powierzchni. W odróżnieniu od tradycyjnych funkcji interpolacyjnych B-sklejane zapewniają większą elastyczność i gładkość. Są one szczególnie użyteczne, gdy konieczne jest modelowanie złożonych, ciągłych kształtów, takich jak narządy wewnętrzne w obrazowaniu medycznym. Do obliczenia wartość y dla woksela mapy przemieszczeń w punkcie x stosuje się formułę:

$$y = \sum_i N_i^n(x) P(y_i), \quad (8)$$

gdzie N_i^n oznacza funkcję B-sklejanych n -tego stopnia w trzech wymiarach odpowiadającą punktowi kontrolnemu x w punkcie y_i , o wartości $P(y_i)$.

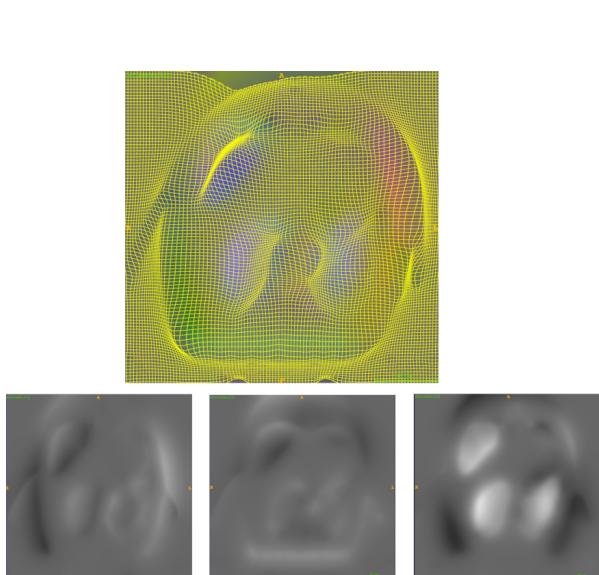


Obraz atlasu.

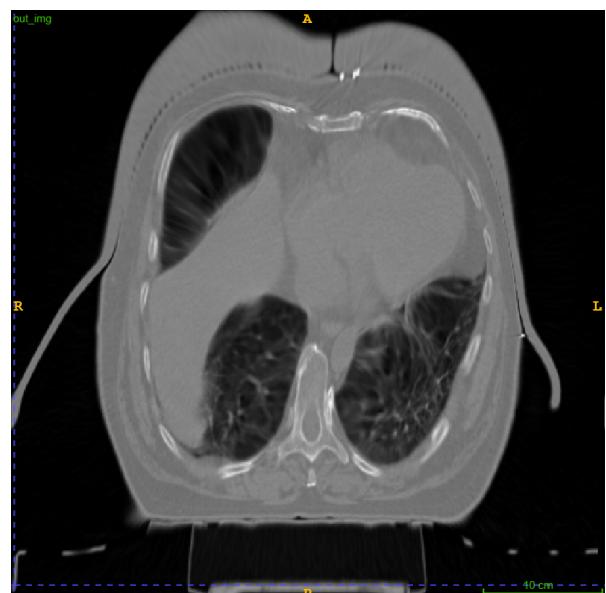


Obraz badanego pacjenta.

Rys. 2.1: Para skanów atlasu i testowego pacjenta, używana do stworzenia poniższych transformacji. Źródło: opracowanie własne.

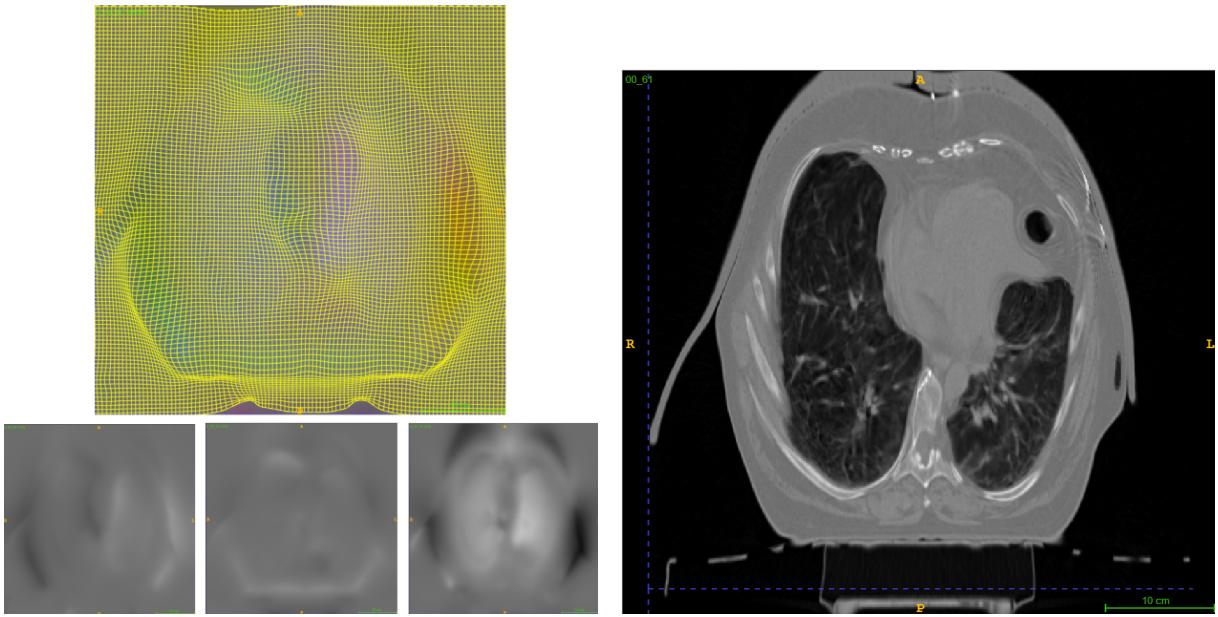


Mapa przemieszczeń wokseli, przedstawiona za pomocą zdeformowanej siatki oraz jej składowych x, y i z.



Obraz atlasu po transformacji afiniczej.

Rys. 2.2: Transformacja afinicza zobrazowana na pojedynczym przekroju wybranego atlasu i obrazu pacjenta. Źródło: opracowanie własne.



Mapa przemieszczeń wokseli, przedstawiona za pomocą zdeformowanej siatki oraz jej składowych x, y i z.

Obraz atlasu po transformacji elastycznej.

Rys. 2.3: Transformacja elastyczna zobrazowana na pojedynczym przekroju wybranego atlasu i obrazu pacjenta. Źródło: opracowanie własne.

Niekiedy, z racji dużej różnicy położenia struktur na obrazach, stosuje się wstępną transformację, która ujednolica środki mas. W ten sposób tworzona jest transformacja atlasu, która obrazuje przemieszczenie środka masy atlasu do punktu centralnego masy obrazu badanego pacjenta. Środek ten oblicza się zgodnie ze wzorem:

$$x_m = \frac{\sum_x I(x)x}{\sum_x I(x)}, \quad (9)$$

x to wektor położenia w obrazie, zaś $I(x)$ to jego wartość.

Metoda atlasowa [Yushkevich, 2024] oferuje dopasowanie obrazów według momentów bezwładności. Polega ono na porównywaniu momentów dwóch obrazów, aby znaleźć odpowiednią transformację geometryczną (np. translację, rotację, skalowanie), która najlepiej dopasowuje te momenty. Momenty bezwładności to miara rozkładu masy (lub intensywności pikseli) względem środka masy obrazu. Używane są one do charakteryzowania kształtu obiektu, a ich dopasowanie służy do wyrównywania obiektów na podstawie ich rozkładu przestrzennego. Momenty są wyznaczane według wzoru:

$$M_{pqw} = \sum_x \sum_y \sum_z (x - x_m)^p (y - y_m)^q (z - z_m)^w \cdot I(x, y, z), \quad (10)$$

gdzie $I(x, y, z)$ to wartość woksela w danym punkcie (x, y, z) , x_m, y_m, z_m to współrzędne środka masy skanu, p, q i w to potęgi określające porządek momentu. Dopasowanie obrazów według

momentów bezwładności jest szczególnie przydatne, gdyż struktury mogą mieć różne położenia lub orientacje, ale ich momenty bezwładności pozostają stabilne.

Dysponując już odpowiednio zmodyfikowanym obrazem atlasu poprzez zastosowanie transformacji, kolejnym krokiem jest analogiczne wykorzystanie jej do maski atlasu. W tym celu dokonuje się odpowiednich obliczeń zgodnie ze wzorem (6).

Podsumowując, aby skorzystać z metody bazującej na atlasach, konieczne jest posiadanie atlasu wraz z maską danej struktury oraz zbiór pacjentów testowych. Jako wyniki działania algorytmu na skanach pacjentów otrzymujemy odpowiadające im obrazy atlasu poddane transformacji, szukaną maskę struktury oraz transformację przekształceń.

Niewątpliwie etapem kluczowym i zarazem najtrudniejszym jest odpowiednie przygotowanie atlasu. Obraz ten w najprostszym przypadku może stanowić reprezentatywny skan ciała, jednakże ze względu na różnorodność rozmiarów i kształtów ciał pacjentów trudno jest pozyścić jeden zestandardyzowany skan. Dlatego tak często stosuje się wielo-atlasowe podejście (ang. multi-atlas based method). W rozwiązaniu tym dla pojedynczych atlasów dokonywane jest rejestrowanie, a następnie segmentacja. W wyniku tego otrzymuje się dla danego pacjenta tyle map struktury, ile jest atlasów. Kluczowym aspektem tego podejścia jest umiejętne połączenie wszystkich wyników segmentacji konkretnego pacjenta. W tym celu często stosuje się techniki głosowania większościowego, ważonego głosowania czy zaawansowanych technik probabilistycznych, które uwzględniają jakość dopasowania poszczególnych atlasów.

2.2.2 Segmentacja neuronowa

W celu skutecznego nakreślenia zasady działania segmentacji opartej o sztuczne sieci neuronowe, wpierw przybliżę koncepcję sztucznego neuronu.

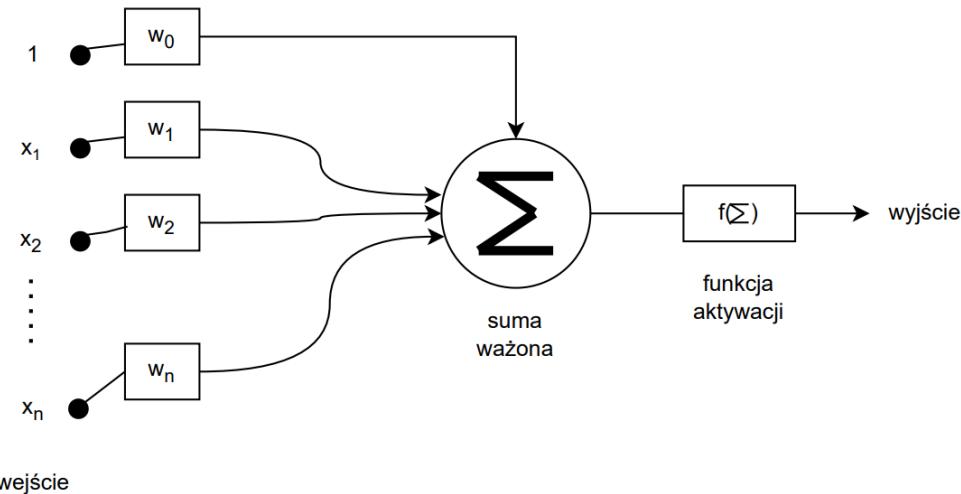
Sieć neuronowa

Sztuczny neuron to podstawowy element składowy sieci neuronowych, które są inspirowane działaniem biologicznych neuronów w mózgu. Sztuczny neuron przetwarza dane wejściowe i generuje wynik na podstawie określonej funkcji matematycznej. Jest to podstawowa jednostka w algorytmach uczenia maszynowego, szczególnie w kontekście głębokiego uczenia.

Neuron otrzymuje sygnały wejściowe, które mogą pochodzić z zewnętrznego źródła danych lub z wyjść innych neuronów w sieci. Każdemu wejściu x_i przypisana jest waga w_i , która odzwierciedla znaczenie tego wejścia dla neuronu. Sygnały wejściowe są mnożone przez odpowiadające im wagi, a następnie sumowane. Następnie wynik tej operacji jest przetwarzany przez funkcję aktywacji, która wprowadza nieliniowość do modelu. Nieliniowość jest kluczowa, ponieważ pozwala sieci neuronowej modelować bardziej złożone zależności. Korzystając z matematycznej notacji, wyjście z neuronu oblicza się następująco:

$$y_i = \sigma(w_0 + \sum_{i=1}^n x_i w_i), \quad (11)$$

gdzie σ jest wybraną funkcją aktywacji, a w_0 to bias, który przesuwa wartość sumy, umożliwiając lepsze dopasowanie modelu.



Rys. 2.4: Schemat sztucznego neuronu McCullocha-Pittsa. Źródło: opracowanie własne.

Jedną z wielu powszechnie stosowanych funkcji aktywacji jest funkcja sigmoidalna (12) czy ReLU (13). Stosowanie funkcji aktywacji jest kluczowe, gdyż wprowadza ona nieliniowość. Bez niej wielowarstwość modelu byłaby nieistotna, równoważna z zastosowaniem jednej warstwy.

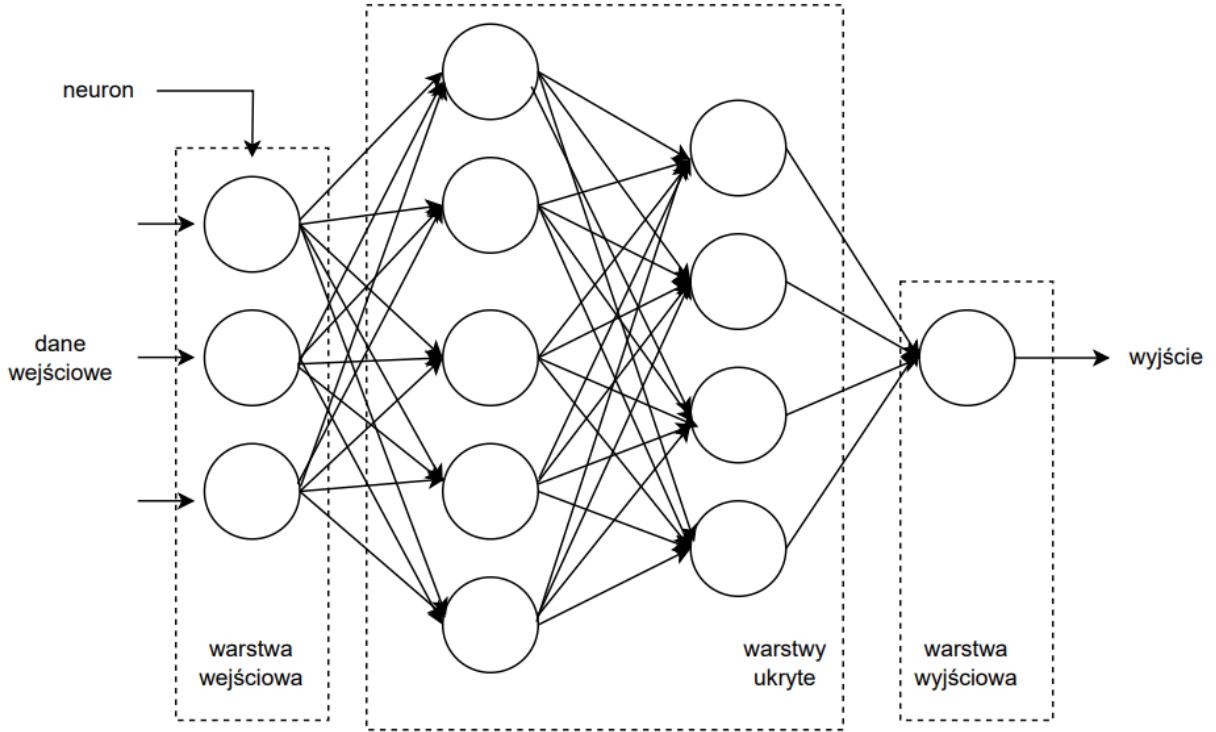
$$S(x) = \frac{1}{1 + e^{-x}}, \quad (12)$$

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} x & \text{gdy } x > 0 \\ 0 & \text{gdy } x \leq 0 \end{cases} \quad (13)$$

Poprzez odpowiednie połączenie neuronów otrzymuje się tzw. sztuczną sieć neuronową. Składa się ona z sekwencji neuronów umieszczonych w warstwach, które to odpowiednio są w pełni połączone z neuronami z sąsiadującą warstwą. Innymi słowy, każde wyjście neuronu z danej warstwy połączone jest z wejściem neuronów warstwy następnej. Oczywiście z wyjątkiem warstwy wyjściowej, z której to przekazywana jest jedna wartość.

Jednym z najpowszechniejszych modeli sztucznej sieci neuronowej jest tak zwany model perceptronu wielowarstwowego, który został opisany powyżej i bazuje na neuronach McCullocha-Pittsa. Zastosowanie sieci neuronowych w dzisiejszych czasach jest powszechnne, najczęściej stosowane są w zadaniach regresji oraz klasyfikacji.

Kluczowym krokiem zastosowania sieci neuronowych jest umiejętne jej wytrenowanie, czyli dopasowanie tak parametrów sieci, w tym wag poszczególnych neuronów, aby była ona w stanie jak najdokładniej generalizować. Proces ten wymaga wiedzy danych wejściowych oraz odpowiadających im wartości oczekiwanych na wyjściu z sieci. Stąd też, aby dokładnie przystosować



Rys. 2.5: Schemat sieci wielowarstwowego perceptronu. Źródło: opracowanie własne.

sieć do rozwiązywania danego problemu potrzebne jest posiadanie reprezentatywnego zbioru danych.

Proces treningu sieci polega na minimalizacji błędu Q (funkcji straty) pomiędzy rzeczywistą wartością odpowiadającą danemu wejściu sieci a predykowaną. Nauka odbywa się w sposób iteracyjny, do momentu osiągnięcia zadanej wartości liczby epok procesu nauki. W każdej z epok aktualizowane są wartości wag, najczęściej z wykorzystaniem metody gradientu prostego, w kolejnych iteracjach t :

$$w_{i,j}^{t+1} = w_{i,j}^t - \alpha \Delta w_{i,j}^t, \quad (14)$$

gdzie:

$$\Delta w_{i,j}^t = \frac{\partial}{\partial w_{i,j}} Q(f(x^t), y^t), \quad (15)$$

x stanowi dane wejściowe sieci, zaś $f(x)$ odpowiadające jemu wyjście, y określa wartość oczekiwana wyjścia, α oznacza współczynnik nauki.

Z uwagi na wykorzystywanie pochodnych cząstkowych wymagane jest, aby cała sieć, czyli wszystkie warstwy oraz funkcje aktywacji i straty były różniczkowalne. Potencjalnie problemów może przysporzyć funkcja ReLU, która to nie jest różniczkowalna w punkcie $x = 0$. Jednak, aby temu zapobiec przyjmuje się, że wartość pochodnej w tym punkcie wynosi 0.

Z reguły, aby przyśpieszyć proces treningu, stosuje się uczenie danych paczkami (batch). W takiej sytuacji aktualizacja dokonywana nie jest po przetworzeniu pojedynczej próbki danych

(wzór 14), lecz różnice o którą powinny być zmodyfikowane wagi sieci są zapamiętywane, obliczana jest z nich średnia, którą następnie stosuje się do aktualizacji wag na koniec danej paczki:

$$\Delta w_{i,j}^t = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\partial}{\partial w_{i,j}} Q(f(x^{tN+n}), y^{tN+n}) \quad (16)$$

Przyśpieszenie to sprowadza się do zastosowania obliczeń równoległych, potencjalnie wspólnie.

Perceptron wielowarstwowy stanowi prosty model sieci. Przeważnie stosuje się sieci neuronowe o gęstszej, bardziej skomplikowanej budowie pozwalając tym samym na rozwiązywanie wymagających obliczeniowo problemów. Wykorzystuje się tym samym różne techniki pozwalające na usprawnienie procesu uczenia i zwiększenie ogólnej wydajności sieci.

Jedną z nich jest normalizacja wsadowa (ang. batch normalisation) [Ioffe and Szegedy, 2015]. Jest to metoda stosowana w celu przyśpieszenia i zwiększenia stabilności treningu sieci poprzez normalizację danych wejściowych warstw za pomocą ponownego centrowania i skalowania wartości.

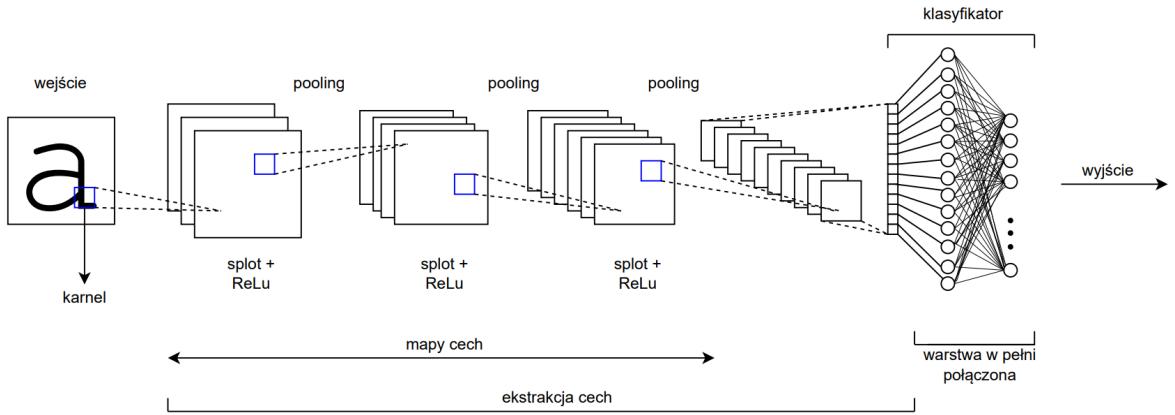
W celu uniknięcia nadmiernego dopasowania modelu do danych uczących, tym samym przeuczenia sieci, stosuję się tak zwany dropout [Hinton et al., 2012]. Ta technika regulacji polega na losowym usuwaniu jednostek wraz z ich połączonymi podczas treningu, w rezultacie redukując współpracę pomiędzy neuronami. W sposób losowy wybierane są neurony, których wejścia są zerowane, zaś dla pozostałych neuronów wyjścia te są skalowane zgodnie z podanym współczynnikiem.

Sieć splotowa

Splotowe sieci neuronowe (ang. convolutional neural networks, CNN) stanowią klasę głębokich sieci neuronowych zaprojektowanych specjalnie do przetwarzania danych o topologii siatki. Sieci CNN, inaczej nazywane konwulencyjnymi, wykorzystują zasadę algebry liniowej, w szczególności operacji splotu, do wyodrębnienia cech i identyfikowania wzorców w obrazach. Chociaż sieci te są głównie stosowane w przetwarzaniu obrazów, można je również dostosować do pracy z danymi audio i innymi sygnałami.

Architektura sieci CNN jest inspirowana wzorcami łączności ludzkiego mózgu — w szczególności kory wzrokowej, która odgrywa zasadniczą rolę w postrzeganiu i przetwarzaniu bodźców wizualnych. Sztuczne neurony w sieci konwulencyjnej są ułożone tak, aby skutecznie interpretować informacje wizualne, umożliwiając tym modelem przetwarzanie całych obrazów. Ponieważ sieci splotowe są tak skuteczne w identyfikowaniu obiektów, często są używane do zadań związanych z komputerowym widzeniem, takich jak rozpoznawanie i wykrywanie obiektów. Typowe przypadki użycia obejmują samochody autonomiczne, rozpoznawanie twarzy i analizę obrazów medycznych.

Sieci CNN wykorzystują szereg warstw, z których każda wykrywa różne cechy obrazu wejściowego. W zależności od złożoności zamierzonego celu, sieć ta może zawierać dziesiątki, setki, a nawet tysiące warstw, z których każda opiera się na wynikach poprzednich, aby rozpoznawać szczegółowe wzorce. Warstwy te można podzielić na trzy grupy: splotowe, łączące i w pełni połączone. Wraz z przetwarzaniem danych przez te warstwy złożoność sieci wzrasta, co pozwala na sukcesywne identyfikowanie większych części obrazu i bardziej abstrakcyjnych cech.



Rys. 2.6: Poglądowy schemat sieci splotowej. Źródło: opracowanie własne.

Warstwa splotowa jest podstawowym elementem składowym CNN i to tutaj odbywa się większość obliczeń. Ta warstwa wykorzystuje filtr lub jądro (mały macierz wag) do poruszania się po polu odbiorczym obrazu wejściowego, w celu wykrycia obecności określonych cech. Aby obliczyć wyjście z danej warstwy splotowej wykorzystuje się poniższy wzór:

$$y_i = \sum_{j=0}^{I-1} (A_{ij} * x_j) + b_i, \quad (17)$$

gdzie y_i stanowi i -ty kanał wyjściowego obrazu, x to obraz stanowiący wejście o i kanałach, b to wektor biasów, A określa macierz jąder, zaś znakiem $*$ oznaczono operację splotu.

Zasada działania sieci opiera się na operacji splotu, który powoduje przesunięcie filtra zaprojektowanego do wykrywania określonych elementów z obrazu wejściowego. Operację splotu w matematyczny sposób można przedstawić następująco:

$$(a * b)_{n,m} = \sum_{k=0}^{K_1-1} \sum_{l=0}^{K_2-1} a_{k,l} b_{nS_1+kD_1, mS_2+lD_2}, \quad (18)$$

$K_1 \times K_2$ określa rozmiar jądra splotu, S_1, S_2 rozmiar korku na wysokości i szerokości. D_1, D_2 dylatację wysokości i szerokości. Z reguły przyjmuję się domyślne wartości $S = [1, 1]$ i $D = [1, 1]$. Poglądowo operację tę przedstawiono na rysunku 2.7.

Filtry początkowe zwykle wykrywają podstawowe cechy, takie jak linie lub proste tekstury. Filtry kolejnych warstw są bardziej złożone, łącząc podstawowe cechy zidentyfikowane wcześniej

-2	-3	-3	0	6	-5
-5	4	8	2	1	1
0	7	6	0	1	1
1	4	-1	-1	-4	2
0	-5	-7	-5	-1	0
3	1	-4	0	0	4

*

1	1	0
1	0	0
0	1	0

=

-3	4		

Rys. 2.7: Uproszczona wizualizacja zasady działania operacji splotu. Źródło: opracowanie własne.

w celu rozpoznawania bardziej złożonych wzorców. Na przykład, po wykryciu przez warstwę początkową obecności krawędzi, głębsza warstwa może wykorzystać te informacje do rozpoczęcia identyfikacji kształtów.

Wynikiem procesu splotu jest mapa, która podkreśla obecność wykrytych cech na obrazie. Rezultat ten służy następnie jako dane wejściowe dla kolejnej warstwy, umożliwiając sieci CNN stopniowe budowanie hierarchicznej reprezentacji obrazu.

Warstwa łącząca (ang. pooling layer) jest krytycznym komponentem sieci, który zazwyczaj następuje po warstwie splotowej. Na tym etapie następuje zmniejszenie wymiarowości danych wejściowych przy jednoczesnym zachowaniu istotnych informacji, co poprawia ogólną wydajność i dokładność sieci. Zazwyczaj osiąga się to poprzez próbkowanie w dół (ang. down-sampling) — zmniejszenie liczby próbek w danych wejściowych. W przypadku sieci splotowej oznacza to zazwyczaj redukcję liczby pikseli używanych do reprezentowania obrazu. Najczęściej stosuje się funkcję maksymalnego łączenia (ang. max pooling), która zachowuje maksymalną wartość w określonym oknie (tj. rozmiar jądra), jednocześnie odrzucając inne wartości. Inna poważna technika, znana jako grupowanie średnią (ang. average pooling), przyjmuje podobne podejście, ale wykorzystuje wartość średnią zamiast maksymalnej. Próbkowanie w dół znacznie zmniejsza całkowitą liczbę parametrów i obliczeń. Oprócz poprawy wydajności wzmacnia to zdolność modelu do generalizacji. Prostsze modele z cechami wyższego poziomu są zazwyczaj mniej podatne na nadmierne dopasowanie — zjawisko, które występuje, gdy model uczy się szumu i zbyt wielu szczegółów w swoich danych treningowych, co negatywnie wpływa na jego zdolność generalizacji dla nowych, niewidzianych informacji. Zmniejszenie rozmiaru przestrzennej reprezentacji ma potencjalną wadę, mianowicie utratę niektórych informacji. Jednak nauczenie się tylko najbardziej widocznych cech danych wejściowych jest zwykle wystarczające dla zadań takich jak wykrywanie obiektów i klasyfikacja obrazów.

Po sekwencji warstw splotowych i łączących w sieci typu CNN z reguły znajduje się warstwa

w pełni połączona (ang. fully connected), która podejmuje ostateczną decyzję, na przykład klasyfikując obiekt na obrazie korzystając z wyników poprzednich warstw. Termin w pełni połączony oznacza, że każdy neuron w jednej warstwie jest połączony z każdym neuronem w warstwie następnej. W pełni połączona warstwa integruje różne cechy wyodrębnione w poprzednich warstwach mapując je na określone klasy lub wyniki. Każde wejście z poprzedniej warstwy łączy się z jednostką aktywacji, umożliwiając jednocześnie rozważenie wszystkich cech podczas podejmowania ostatecznej decyzji. Zazwyczaj stosowany jest w tym celu perceptron wielowarstwowy.

Nie wszystkie warstwy w CNN są w pełni połączone. Ponieważ w pełni połączone warstwy mają wiele parametrów, zastosowanie tego podejścia w całej sieci spowodowałoby niepotrzebną gęstość, zwiększyłoby ryzyko nadmiernego dopasowania i uczyniłoby sieć bardzo kosztowną w trenowaniu pod względem pamięci i obliczeń. Ograniczenie liczby w pełni połączonych warstw, stosowanie funkcji aktywacji, normalizacji wsadowej czy techniki dropout równoważy wydajność obliczeniową i zdolność generalizacji ze zdolnością uczenia się złożonych wzorców.

Splotowe sieci neuronowe są szczególnie przydatne w zadaniach związanych z komputerowym widzeniem, takich jak rozpoznanie i klasyfikacja obrazów, ponieważ są zaprojektowane do uczenia się hierarchii przestrzennych cech poprzez wychwytywanie istotnych informacji we wczesnych i złożonych wzorców w głębszych warstwach. Jedną z najważniejszych zalet sieci CNN jest ich zdolność do wykonywania automatycznej ekstrakcji cech lub uczenia się ich. Eliminuje to potrzebę ręcznej ekstrakcji cech, co historycznie było pracochłonnym i złożonym procesem. Sieci te są również dobrze przystosowane do uczenia transferowego (ang. transfer learning), w którym wstępnie wyszkolony model jest dostosowywany do nowych zadań. Ta możliwość ponownego użycia sprawia, że sieci neuronowe są wszechstronne i wydajne, szczególnie w przypadku zadań z ograniczonymi danymi treningowymi. Operowanie na istniejących sieciach umożliwia twórcom uczenia maszynowego wdrażanie sieci konwulencyjnych w różnych scenariuszach rzeczywistych przy jednoczesnym minimalizowaniu kosztów obliczeniowych.

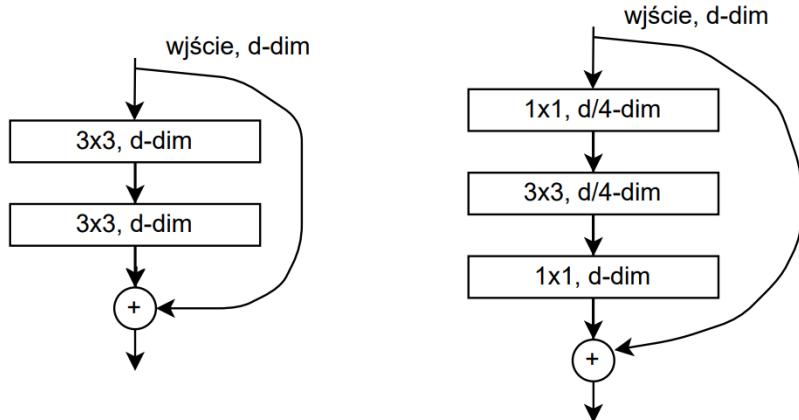
ResNet

W przypadku zwiększenia liczby warstw w sieci zaobserwować można wzrost błędu uczenia i testowania, co zazwyczaj dzieje się tak z powodu znikającego/eksplodującego gradientu. Aby zminimalizować ten problem zaprezentowana została nawracająca sieć neuronowa (ang. residual neural network, ResNet) [He et al., 2016].

Rezydualna sieć neuronowa, jak nazwa wskazuje, swoją budowę opiera na tak zwanych blokach rezydualnych (Rys. 2.8). Składa się on z sekwencji warstw, z której wyjście wraz z danymi wejściowymi do bloku jest przekazywane do kolejnych elementów sieci.

Podstawowy blok rezydualny budują dwie sekwencyjne warstwy splotowej 3×3 . Wymiary wejściowe i wyjściowe obu warstw są takie same. Z kolei, blok wąskiego gardła (ang. bottleneck

block) składa się z trzech sekwencji warstw splotowych. Pierwsza z tych warstw zawiera splot 1×1 redukujący wymiarowość wejścia do np. $1/4$ wymiaru wejściowego; druga wykonuje splot 3×3 , zaś ostatnia warstwa to kolejny splot 1×1 przywracający wymiary wejściowe. Przykładowo, wersje sieci rezydualnej ResNet-50, ResNet-101 i ResNet-152 oparte są o blok wąskiego gardła, zaś ResNet-34 o blok podstawowy.



Rys. 2.8: Dwie odmiany splotowych bloków rezydualnych: po lewej blok podstawowy, po prawej blok wąskiego gardła. Źródło: opracowanie własne.

Zaletą dodawania tego typu połączenia pomijającego jest to, że jeśli jakakolwiek warstwa zaszkodzi wydajności architektury, zostanie ona pominięta przez regularyzację. W rezultacie można wytrenować bardzo głęboką sieć neuronową bez problemów spowodowanych zanikającym/eksplodującym gradientem.

ResNet-34 to specyficzna architektura w rodzinie sieci ResNet, zaprojektowana do zadań rozpoznawania obrazów [He et al., 2016]. Jest to wariant, który równoważy złożoność modelu i wydajność obliczeniową, dzięki czemu nadaje się do szerokiego zakresu zadań klasyfikacji.

Sieć ta składa się z 34 warstw konwolucyjnych, grupowanych w cztery bloki. Po każdej z nich zmniejsza się rozdzielcość obrazu, a liczba filtrów rośnie, aby zwiększyć zdolności do ekstrakcji cech.

Warstwa wejściowa korzysta ze splotu 7×7 , posiada 64 filtry i stosuje przesunięcie okna filtru co 2 piksele (ang. stride) oraz dodaje dookoła obrazu 3 piksele zerowe z każdej strony (ang. padding). Na tym etapie sieć wyodrębnia podstawowe cechy z obrazu, takie jak krawędzie i tekstury. Następnie wykonywana jest redukcja oparta na maksimum (ang. max pooling) wykorzystującą jądro 3×3 oraz stosującą przesunięcie filtra o 2 piksele.

ResNet-34 dzieli się na cztery główne grupy bloków rezydualnych. Szczegóły każdej z grup prezentują się następująco:

1. 64 filtry, 3 bloki rezydualne (sploty 3×3),
2. 128 filtrów, 4 bloki rezydualne (downsampling na początku etapu),

3. 256 filtrów, 6 bloków rezydualnych (downsampling na początku etapu),
4. 512 filtrów, 3 bloki rezydualne (downsampling na początku etapu).

W każdym z tych bloków pierwsza warstwa splotowa stosuje przesunięcie okna filtru o 2 piksele, w celu zmniejszenia rozdzielczości obrazu. Po zakończeniu ostatniego bloku rezydualnego sieć używa warstwy globalnego łączenia średnią (ang. global average pooling). W wyniku czego wartości w każdym kanale zostają uśrednione, przy jednoczesnej redukcji wymiarów. Ostatecznie, warstwa w pełni połączona przekształca wyjście w wektor o rozmiarze odpowiadającym liczbie klas. Na końcu stosuje się funkcję softmax, która przekształca wynik w prawdopodobieństwo przynależności do danej klasy.

Sieć ta może być z powodzeniem stosowana w problemach rozpoznawania i klasyfikacji obiektów na obrazach. Ze względu na zdolność do przechwytywania hierarchicznych cech, ResNet-34 jest często używany jako ekstraktor cech w aplikacjach uczenia transferowego. Może być również dostosowany do zadań, takich jak segmentacja semantyczna i wykrywanie obiektów, poprzez dodawanie specjalistycznych warstw.

EfficientNet

EfficientNet to rodzina architektur sieci neuronowych splotowych, która została wprowadzona przez badaczy z Google AI [Tan and Le, 2019]. Modele te znane są z osiągania najnowocześniejszej wydajności w zadaniach klasyfikacji obrazów przy znacznie mniejszej liczbie parametrów i zasobów obliczeniowych.

EfficientNet wprowadza zunifikowane skalowanie (ang. compound scaling), czyli metodę skalowania sieci w trzech wymiarach: głębokości (D), szerokości (W) i rozdzielczości (R):

- głębokość (ang. depth) — liczba warstw w sieci,
- szerokość (ang. width) — liczba kanałów cech (filtrów) w każdej warstwie,
- rozdzielcość (ang. resolution) — rozdzielcość wejściowego obrazu.

Model skaluje się według wzoru:

$$d = \alpha^N, w = \beta^N, r = \gamma^N, \quad (19)$$

gdzie α, β, γ to współczynniki skalowania odpowiednio głębokości, szerokości i rozdzielczości, zaś N to parametr skalowania (np. dla EfficientNet-B0, $N = 0$).

W tradycyjnych sieciach zazwyczaj skaluje się tylko jeden z tych wymiarów (np. poprzez zwiększenie liczby warstw), co może prowadzić do nieefektywności. EfficientNet skaluje te wymiary w sposób zrównoważony, aby zwiększyć możliwości modelu przy minimalnych kosztach.

Metoda skalowania złożonego jest uzasadniona intuicją, że jeśli obraz wejściowy jest większy, to sieć potrzebuje więcej warstw, aby zwiększyć pole odbiorcze i więcej kanałów, aby

uchwycić bardziej drobnoziarniste wzorce na większym obrazie.

Bazowa sieć EfficientNet-B0 opiera się na blokach MBConv, które pochodzą z architektury MobileNetV2. MBConv to tak zwane odwróconych bloki rezydualne wąskiego gardła (ang. inverted bottleneck), które najpierw rozszerzają liczby kanałów w danych wejściowych, a potem zmniejszają je do oryginalnego rozmiaru.

EfficientNet-B0 jest najbardziej podstawową wersją EfficientNet. Jej struktura wygląda następująco:

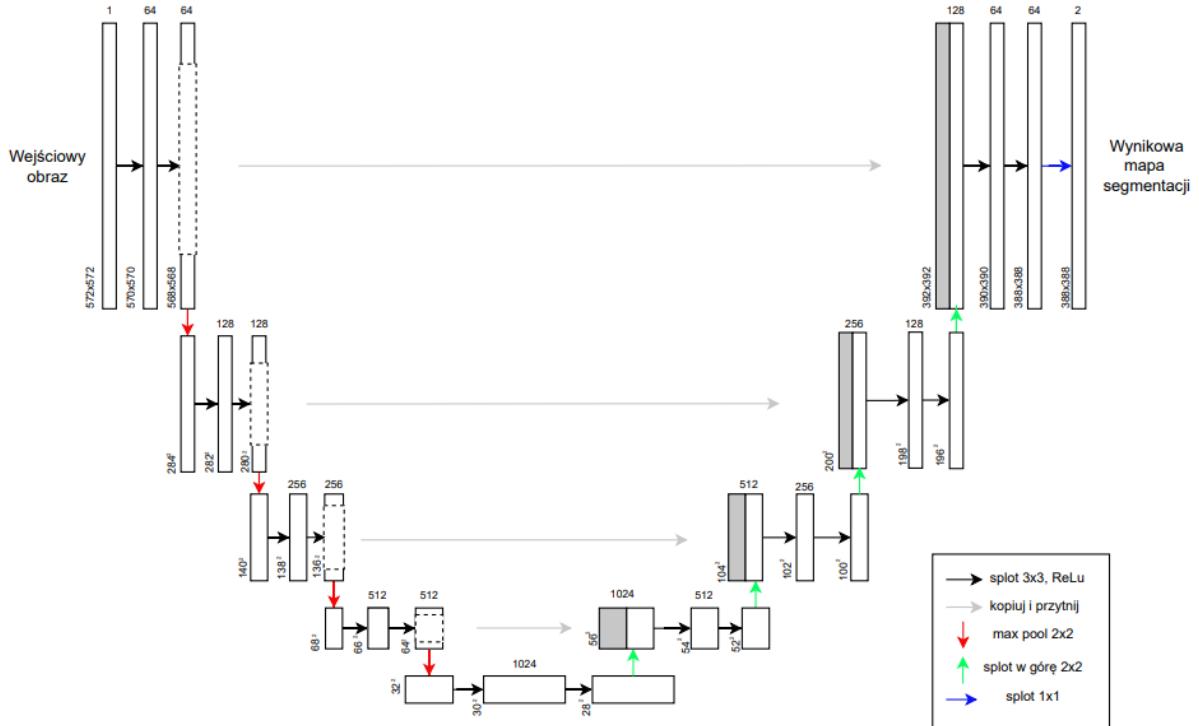
1. warstwa splotowa 3×3 z przesunięciem o 2 piksele, 32 filtry,
2. MBConv (3×3), 16 filtrów, szerokość 1, przesunięcie o 1 piksel,
3. MBConv (5×5), 24 filtrów, szerokość 6, przesunięcie o 2 piksel,
4. MBConv (5×5), 40 filtrów, szerokość 6, przesunięcie o 2 piksel,
5. MBConv (3×3), 80 filtrów, szerokość 6, przesunięcie o 2 piksel,
6. MBConv (5×5), 112 filtrów, szerokość 6, przesunięcie o 1 piksel,
7. MBConv (5×5), 192 filtrów, szerokość 6, przesunięcie o 2 piksel,
8. MBConv (3×3), 320 filtrów, szerokość 6, przesunięcie o 1 piksel.

Po przetworzeniu ostatniego bloku odwróconego wąskiego gardła następuje globalne łączenie średnią z wykorzystaniem rozmiaru jądra 7×7 . Następnie uśrednione wyniki są przetwarzane przez warstwę w pełni połączoną, po czym stosuję się funkcję softmax, w celu uzyskania prawdopodobieństw przynależności danego piksela do klasy.

EfficientNet to nowoczesna architektura sieci neuronowej, która wprowadza innowacyjne podejście do skalowania modeli. Zastosowanie zunifikowanego skalowania i wydajnych bloków MBConv sprawia, że EfficientNet osiąga doskonałą równowagę między dokładnością a efektywnością obliczeniową, co czyni go jednym z najpopularniejszych wyborów w dziedzinie głębokiego uczenia się w zadaniach związanych z przetwarzaniem obrazów.

U-Net

U-Net to szeroko stosowana architektura głębokiego uczenia, dedykowana problemowi segmentacji biomedycznej, zaproponowana została w publikacji [Ronneberger et al., 2015]. Głównym jej celem było rozwiązanie problemu ograniczonej liczby danych z adnotacjami w dziedzinie medycyny. Sieć ta została zaprojektowana tak, aby efektywnie wykorzystywać mniejszą ilość danych przy jednoczesnym zachowaniu szybkości i dokładności. Jej unikalny kształt przypominający literę U (stąd nazwa) pozwala efektywnie łączyć informacje o niskiej i wysokiej rozdzielczości, co czyni ją niezwykle skuteczną w segmentacji obrazów, gdzie istotna jest precyzyjna lokalizacja oraz kontekst globalny.



Rys. 2.9: Rysunek poglądowy architektury sieci U-Net. Źródło: opracowanie własne.

Architektura sieci U-Net zbudowana jest z ścieżki kontraktacyjnej (enkoder) i ścieżki ekspansywnej (dekoder). Ścieżka kontraktacyjna składa się z kilku bloków, gdzie każdy blok zawiera dwie operacje splotowe 3×3 (z paddingiem, aby zachować rozmiar) oraz operację maksymalnego łączenia 2×2 , z przesunięciem o 2 piksele (ang. stride), która zmniejsza rozdzielcość o połowę. Ścieżka ekspansywna łączy cechy i informacje przestrzenne poprzez sekwencję splotów w górę 2×2 (transponowanych), konkatenację z cechami o wysokiej rozdzielcości ze ścieżki kontraktacyjnej oraz dwukrotny splot 3×3 z aktywacją ReLU. Po ostatnim etapie dekodowania, stosowana jest pojedynczy splot 1×1 , który redukuje liczbę kanałów do liczby klas (zwykle 2 dla segmentacji binarnej). Przykładowa architektura sieci U-Net została przedstawiona na schemacie 2.9.

Główną innowacją U-Net są połączenia skrótu (ang. skip connections) pomiędzy odpowiadającymi sobie warstwami z części kontraktacyjnej i ekspansywnej. Informacje z wcześniejszych etapów enkodera są przekazywane bezpośrednio do dekodera, co pomaga w dokładnej lokalizacji krawędzi oraz lepszym odwzorowaniu obiektów o różnych rozmiarach.

Jak to zostało wcześniej wspomniane, sieci splotowe używane są w dwóch głównych problemach: klasyfikacji i regresji. Wracając do definicji segmentacji neuronowej, jest to swego rodzaju klasyfikacja, lecz nie całego obrazu ale poszczególnych jego pikseli. W wyniku czego, na wyjściu sieci dla każdego piksela otrzymujemy etykietę klasy, do której został on przyporządkowany. Etykiety te są następnie wizualizowane w postaci masek segmentacyjnych, które

obrazują struktury zidentyfikowane w analizowanym obrazie.

2.2.3 Miary jakości

Aby zweryfikować dokładność dokonanej segmentacji posiadając wzorcowe oznaczenia danych struktur biomedycznych na obrazach wykorzystuje się miary jakości.

W celu wprowadzenia definicji wybranych miar przyjęto następujące oznaczenia. Jako Z określono zbiór wokseli przewidzianych jako należących do danej klas, zaś jako Y zdefiniowano zbiór wokseli rzeczywiście określających daną klasę. Zbiór Z_C określa woksele leżące na powierzchni segmentu zdefiniowanego poprzez zbiór Z i analogicznie Y_C określa zbiór wokseli znajdujących się na obszarze zdefiniowanym przez zbiór Y .

Do najczęściej w literaturze wykorzystywanych miar jakości segmentacji należą następujące metody:

1. współczynnik podobieństwa Dice'a (DSC) — miara ta określa stopień pokrycia się przewidzianego wyniku z rzeczywistymi danymi. Przyjmuje wartości z zakresu $[0; 1]$, im wyższa wartość tym segmentacja przebiegła dokładniej. Współczynnik ten oblicza się go za pomocą następującego wzoru:

$$DSC = \frac{2|Y \cap Z|}{|Y| + |Z|} \quad (20)$$

2. maksymalny odstęp Hausdorffa (HD) — określa największą odległość pomiędzy przewidywanym a rzeczywistym obszarem danej klasy. Miarę tę wyraża się w milimetrach, a celem jest uzyskanie jak najmniejsze wartości. Maksymalny odstęp Hausdorffa oblicza się używając poniższej formuły.

$$HD = \max\{d(Y, Z), d(Z, Y)\}, \quad (21)$$

gdzie:

$$d(Z, Y) = \max_{z \in Z} \min_{y \in Y} d(z, y) \quad (22)$$

oraz $d(z, y)$ oznacza odległość euklidesową między dwoma wokselami z i y wyrażona w milimetrach.

3. iloraz objętości (VR) — miara ta określa stosunek przewidywanej do rzeczywistej objętości danej klasy. Uzyskana wartość 1 oznacza, że przewidywana struktura odpowiada tej rzeczywistej. Wartości powyżej progu 1 określają, że predykowana struktura jest większa od wzorcowego oznaczenia, zaś poniżej miara ta informuje, że model przewidział zbyt małą objętość mapy. Aby go obliczyć stosuje się poniższy wzór:

$$VR = \frac{|Z|}{|Y|} \quad (23)$$

4. średni odstęp pomiędzy powierzchniami (MSD) — miara ta, jak nazwa wskazuje, wyraża średni odstęp pomiędzy przewidywaną a rzeczywistą powierzchnią segmentu. Wartości MSD podaje się milimetrach, małe wartości tej miary informują o wysokiej dokładności segmentacji. Oblicza się ją następująco:

$$MSD = \frac{1}{2}(d_M(Y_C, Z_C) + d_M(Z_C, Y_C)), \quad (24)$$

gdzie

$$d_M(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} d(a, b) \quad (25)$$

Dodatkowo wybrano miarę precyzji (ang. precision) oraz czułości (ang. recall), które pozwalają na określenie jakości działania segmentacji pod innym kątem. Miary te zazwyczaj stosowane są w zadaniach klasyfikacji, niemniej jednak w przypadku małej struktury LAD mogą stanowić istotną informację na temat przebiegu procesu segmentacji.

Dysponując dwoma binarnymi maskami, rzeczywistą i przewidywaną, można obliczyć miary jakości obrazujące, w jakim stopniu jedna maska jest pokryta przez drugą oraz jak dokładnie udało się umieścić maskę w danym obrazie.

Miara precyzji (ang. Positive Predicted Value, PPV) określa jaką część stanowią poprawnie oznaczone jako segment woksele do liczby wszystkich wokseli mapy struktury. Miara ta przyjmuje wartości z przedziału $[0; 1]$, gdzie maksymalna wartość określa bezbłędną segmentację. Wzór wygląda następująco:

$$PPV = \frac{TP}{TP + FP}, \quad (26)$$

gdzie TP to ilość przypadków wokseli prawdziwie pozytywnych, czyli takich które zostały oznaczone jako należące do maski i faktycznie do niej powinny należeć; FP to z kolei liczba obiektów fałszywie pozytywne — woksele uznane jako szukana struktura, lecz w rzeczywistości obrazują one inne tkanki.

Z kolei miara czułości (ang. True Positive Rate, TPR), pozwala na ocenienie jaki udział prawidłowo rozpoznanych wokseli wśród wszystkich stanowią te, które powinny zostać oznaczone jako maska struktury. Przyjmuje ona wartości z przedziału $[0; 1]$, analogicznie jak w przypadku precyzji, im wyższa wartość czułości tym segmentacja przebiegła dokładniej.

$$TPR = \frac{TP}{TP + FN}, \quad (27)$$

gdzie TN to ilość wokseli prawdziwie negatywnych, czyli takich które nie powinny znaleźć się w masce struktury i tak też się stało.

2.3 Przegląd literatury

Temat segmentacji struktur anatomicznych nie stanowi niszy, aczkolwiek publikacji skupiających się na niewielkich rozmiarów obiektach, takich jak tętnice wieńcowe nie jest dużo. Problem wyodrębnienia LAD bardzo często spotykany jest jako proces segmentacji, ale nie z wyników tomografii komputerowej, lecz angiografii czy koronarografii. Przykładem tego są artykuły, pierwszy z nich porusza problem klasyfikacji naczyń wieńcowych [Kaba et al., 2023], zaś drugi skupia się stricte na ich segmentacji [Su et al., 2019]. W obu tych publikacjach autorzy do segmentacji stosują splotowe sieci neuronowe i wyniku czego uzyskują dokładność identyfikacji naczyń rzędu min. 70%. Jednakże, w porównaniu zdjęć tomografii komputerowej do angiografii, badanie to wykonywane jest w obrębie samego serca i skupia się na analizie przepływu krwi przez żyły. Stąd też pacjentowi przed badaniem podawany jest środek kontrastujący. W rezultacie pozwala to na łatwiejszą identyfikację przebiegu tętnic wieńcowych. Niestety, w przypadku tomografii komputerowej bez środka kontrastowego granice struktur naczyniowych są trudno definiowalne.

Studium literatury przedmiotu wykazało, że automatyczna segmentacja naczyń wieńcowych LAD na podstawie skanów tomografii komputerowej jest tematem rzadko podejmowanym przez badaczy. Niemniej jednak pewien zespół badawczy podjął się tego problemu, w wyniku czego powstał artykuł prezentujący skuteczne rozwiązanie automatycznej segmentacji struktur kardiologicznych z obrazów tomografii komputerowej [Chen et al., 2024]. Celem autorów było wyodrębnienie ze skanów tomografii komputerowej 19 struktur, w tym komór serca oraz LAD. Aby to osiągnąć, utworzono i zbadano dwa podejścia bazujące na splotowych sieciach neuronowych, odpowiednio 3D U-Net i 3D nnU-Net oraz podejście wieloatlasowego, które opisane zostało w oddzielnej publikacji [Zhou et al., 2017].

Model 3D U-Net składał się z etapu zgrubnej lokalizacji, który określał umiejscowienie i pola receptive wszystkich podstruktur serca oraz udoskonalonego etapu segmentacji, który skupiał się na lokalnej segmentacji struktur wchodzących w skład danej grupy struktur. W rezultacie zostały wytrenowane cztery modele segmentacji 3D U-Net dla każdej grupy podkonstrukcji. W kolejnym etapie oddano treningowi model niezależnie, dla każdej indywidualnej struktury zwiększając przy tym rozdzielcość. W przypadku modelu 3D nnU-Net trenowano cztery wieloklasowe modele o pełnej rozdzielcości dla każdej grupy struktur, które połączono w ujednolicony model. Autorzy dzięki zastosowaniu tych modeli uzyskali wartości miary DSC dla U-Net, nnU-Net i multi-atlas odpowiednio 0,57, 0,65 i 0,23. Wartości te informują, że w tym przypadku atlasowe podejście było mało skuteczne, zaś neuronowe modele działały z zadowalającą dokładnością. Należy również wspomnieć, że autorzy korzystali z zarówno obrazów tomografii wykonywanej bez, jak i z kontrastem.

Kolejną publikacją poruszającą temat segmentacji lewej przedniej zastępującej tętnicy wieńcowej wykorzystuje metodę geometryczną [van den Bogaard et al., 2019]. Praca ta skupia się wyłącznie na wyodrębnieniu struktury LAD ze skanów bezkontrastowej tomografii komputero-

wej pacjentek z nowotworem piersi. W pierwszym kroku prezentowanej metody zlokalizowano lewą i prawą komorę serca za pomocą podejścia atlasowego, które zostało stworzone z wykorzystaniem zewnętrznego narzędzia Mirada RTx. LAD jest największą tętnicą wieńcową, biegnącą w bruździe międzykomorowej przedniej, rozciągając się od podstawy serca do wierzchołka, stąd też obranie koncepcji bazującej na punktach orientacyjnych serca może stanowić dobry punkt wyjścia do automatycznej segmentacji LAD. To rozwiązanie w pierwszej kolejności rozszerza wcześniej ustalone maski komór, po czym wybierana jest ich część wspólna. Następnie dla punktów, uzyskanych poprzez redukcję obszaru styku komór, stosowana jest regresja liniowa w celu dopasowania linii. Finalnie przekształcano powstały odcinek w rozmiar odpowiadający LAD — dokonano odpowiedniego przycięcia i rozszerzenia po grubości. W ten sposób uzyskano wartość miary DSC równą 0,15, na podstawie której segmentację można uznać za mało skuteczną. Dokładność tej metody można określić jako niską, jednak należy zwrócić uwagę, że rozwiązanie to jest pierwszą próbą zastosowania geometrycznej metody lokalizacji tętnicy.

Jednym z nielicznych, którzy podjęli się wykorzystać jednocześnie metodę atlasową i neuronową jest Finnegan [Finnegan et al., 2023]. Przedstawione rozwiązanie składa się z trzech modułów, które za zadanie mają wyodrębnić serce, struktury wchodzące w jego skład, jak i leżące w sąsiedztwie tętnice. Do segmentacji całego serca wykorzystano sieć nnU-Net, zaś do identyfikacji jam serca, bazy aorty, tętnicy płucnej i żyły głównej posłużyło podejście bazujące na atlasach. Zastosowano w tym celu transformację afinyczną, poprzedzającą elastyczną. Trzeci moduł segmentujący zastawki i tętnice wieńcowe, w tym LAD, bazuje na wynikach segmentacji z poprzednich modułów i wykorzystuje w tym celu metodę geometryczną. Dodatkowo rozwiązanie to korzysta z informacji o położeniu szukanych struktur względem znanych. Wyniki segmentacji LAD wynosiły mniej więcej 0,19 wartości miary Dice'a, co można interpretować jako segmentację obarczoną dużym błędem.

Zainspirowany publikacją Finnegana autor pracy naukowej Sipiński [Sipiński, 2023], dotyczącej wsparcia procesu planowania radioterapii metodami neuronowymi i atlasowymi, podjął się zbudować rozwiązanie również będące hybrydą sieci neuronowej i atlasów. W tym przypadku wykorzystano sieć w pełni splotową oraz zdefiniowano odmienny niż dotychczas sposób łączenia tych podejść. Finnegan używa różnych algorytmów do segmentacji samodzielnie, z kolei Sipiński wykorzystał sieć do wstępnej segmentacji, której wyniki następnie stanowią wejście do metody atlasowej. Nowe rozwiązanie dokładnie lokalizuje serce oraz jego komory, ale niestety w przypadku lewej przedniej zastępującej tętnicy wieńcowej uzyskane wyniki nie są dobre, osiągnięto wartość średniej miary DSC równą 3%.

Publikacji na temat segmentacji w obszarze gruczołów piersiowych, z uwagi na popularność tematu i większe rozmiary struktury, jest więcej. Zgromadzone w literaturze badania prezentują akceptowalną jakość segmentacji. Wyodrębnienie gruczołu piersiowego nie jest jednak trywialne, z racji różnorodności dostępnych anatomii pacjentek. W artykule Ciardo [Ciardo et al., 2017] zaprezentowano metodę bazującą na atlasach, stosowaną w radioterapii

raka piersi. Naukowcy podzieli wstępnie zbiór danych względem obwodu klatki piersiowej oraz rodzaju operacji jaką przeszła dana pacjentka. Atlasy zostały utworzone przez zewnętrze oprogramowanie MIM. Wynik dokładności działania metody dla jednej z bibliotek wynosił ok. 87%, co można uznać za dobrze przeprowadzoną segmentację.

Podejście oparte o sieci neuronowe prezentowane w artykule Liu [Liu et al., 2021] pozwala na identyfikację ze zdjęć tomografii komputerowej między innymi gruczołów piersiowych. Wykorzystano sieć U-ResNet, wywodzącą się z modelu 2D U-Net, składającego się z kodera i dekodera. Takie podejście do segmentacji osiągnęło dokładność identyfikacji gruczołów piersiowych w wysokości ok. 94%, co niemalże jest blisko idealnemu dopasowaniu masek segmentacji z oznaczeniami eksperta.

Natomiast segmentacja loży po guzie nowotworowym w porównaniu do gruczołów piersiowych, czy LAD wymaga, aby dane pacjentek były zgodne pod kątem umiejscowienia loży (odpowiedni kwadrant piersi) oraz spójne pod względem objętości tego miejsca. Badania prowadzone przez Xie [Xie et al., 2023] wykorzystywały informacje o umiejscowieniu i wielkości nowotworów oraz ilości wycinanych tkanek w procesie ich usuwania. Proponowane podejście bazuje na sieci 3D U-Net, która powszechnie jest używana w segmentacji struktur biomedycznych. Sieć przyjmowała dane w dwóch kanałach: wzmacniona przedoperacyjna i pooperacyjna tomografia komputerowa. Dodatkowo wykorzystano pięciokrotne krzyżowanie danych. Uzyskany w ten sposób model dokonał segmentacji skutecznie, ponieważ miara Dice'a osiągnęła średnią wartość 0,81.

Z przeprowadzonego przeglądu literatury wynika, że na przestrzeni ostatnich lat tematy segmentacji LAD, gruczołów piersiowych czy loży zyskują coraz większą popularność. Uzyskane wyniki pozostawiają pole do ulepszeń i poszukiwań bardziej precyzyjnych rozwiązań. W przypadku segmentacji medycznej dokładność wyodrębnienia lokalizacji danej struktury jest niezwykle istotna, gdyż na podstawie tych wyników zależy Twoje zdrowie, a nawet życie.

Najskuteczniejsza z metod segmentacji LAD uzyskała wartość miary Dice'a ok. 65% [Chen et al., 2024], wynik ten określa tą metodę jako przeciętnie skuteczną. Segmentacja loży po guzie wymaga posiadania zbioru danych pacjentek posiadających zbliżoną lokalizację guza, co stanowi wyzwanie. W rezultacie, temat segmentacji miejsca po guzie nowotworowym piersi w literaturze jest rzadko podejmowany. W publikacji Xie [Xie et al., 2023] segmentowano lewostronne loże po guzie i uzyskano najlepszą wartość miary Dice'a ok. 81%. Pomimo zadowalającego wyniku rozwiązanie to ma potencjał do wprowadzenia innowacji.

2.4 Przegląd istniejących narzędzi

Automatyczna segmentacja struktur biomedycznych jest zagadnieniem leżącym na pograniczu medycyny i informatyki. Narzędzia do tego typu zadań są dedykowane medykom, a dostęp do

nich jest zwykle bardzo ograniczony. W przeglądzie literatury przewijają się nazwy oprogramowań do segmentacji atlasowej, niestety wszystkie z wymienionych są narzędziami płatnymi, trudno dostępnymi.

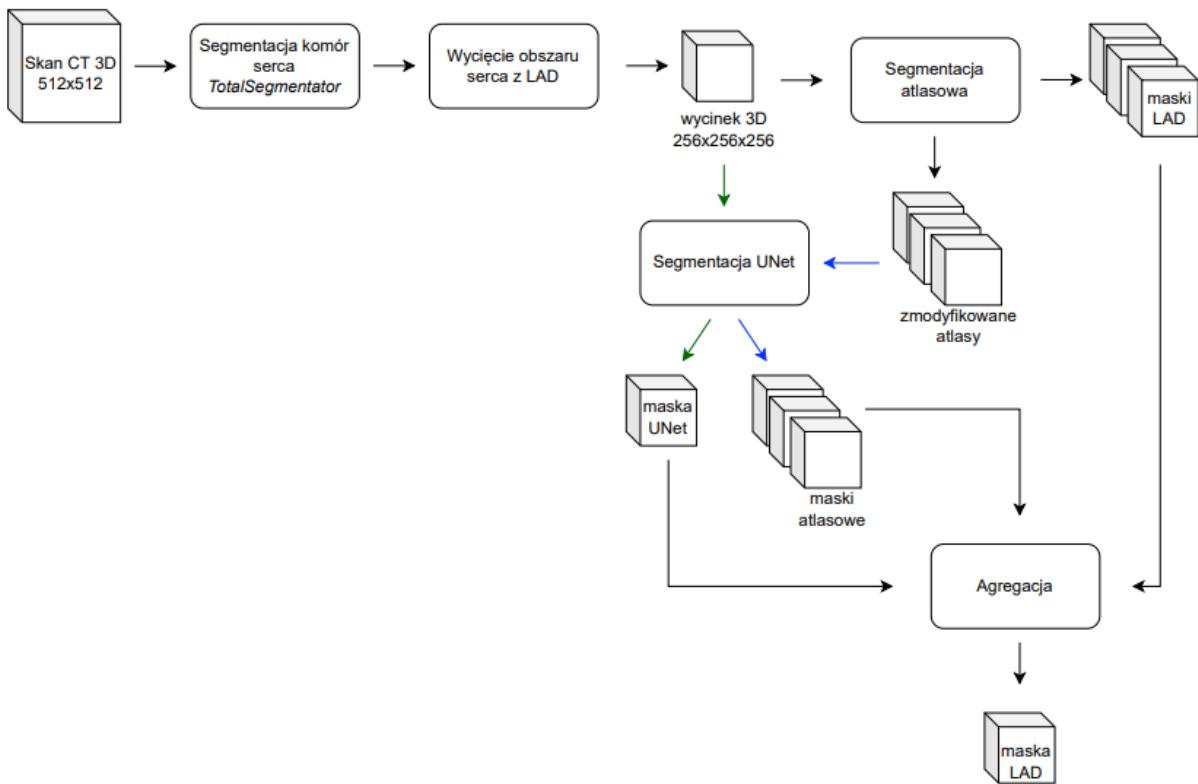
Jednym z niewielu bezpłatnych narzędzi obsługujących segmentację bazującą na atlasach jest Greedy [Yushkevich et al., 2016] [Yushkevich, 2024]. Umożliwia ono rejestrację obrazów medycznych, jak i dokonywanie transformacji w celu dopasowania etykiet. Oprogramowanie to obsługuje różne funkcje celu, transformacje, formaty plików i co więcej, pozwala na konfigurację parametrów wejściowych. Kilka ważniejszych funkcjonalności tego narzędzia to między innymi możliwość rejestracji elastycznej bazującej na krzywych B-sklejanych (ang. B-spline), iteracyjna rejestracja w różnych rozdzielczościach obrazów czy rejestracja dwu- lub trójwymiarowa.

Wszechstronne narzędzie 3DSlicer [Fedorov et al., 2012] typu open source pozwala na wizualizacje, przetwarzanie, segmentację, rejestrację oraz analizę obrazów i siatek trójwymiarowych. Ogromną zaletą 3DSlicera jest możliwość zainstalowania zewnętrznych wtyczek, które dodatkowo rozszerzają funkcjonalność aplikacji. Jedną z nich jest wtyczka zawierająca model Total-Segmentator [Wasserthal, 2024a], który został przedstawiony w artykule [Wasserthal et al., 2023]. Autorzy stworzyli narzędzie pozwalające na segmentację aż 104 struktur anatomicznych, w czego skład wchodzą zarówno narządy, kości, mięśnie i naczynia krwionośne. Rozwiążanie to nie jest jednak dostosowane, aby lokalizować LAD, skupia się jedynie na większych tętnicach i żyłach w obrębie serca. Natomiast, dokładnie segmentuje obszary, na których model został wyuczony (ok. 0,94 wartości miary Dice'a), jednak trzeba w tym celu odpowiednio ustawić kontrast w obrazach CT. Model bazuje na sieci nnU-Net, który jest również dostępny poprzez api [Wasserthal, 2024b] dedykowane językowi programowania Python.

Na rynku komercyjnym jest niewiele narzędzi do segmentacji medycznej. Programy o największej skuteczności i najbogatszej ofercie funkcjonalności nie umożliwiają wyodrębnienia LAD, gruczołów piersiowych oraz loży po guzie. Przegląd istniejących oprogramowań wykazał, że brakuje instrumentów dedykowanych do wykrywania niewielkich rozmiarów struktur, zmian nowotworowych czy loży po guzach.

3 Opis rozwiązania

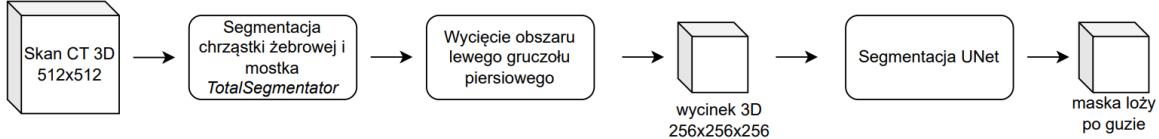
W celu wyodrębnienia z obrazów skanów tomografii komputerowej lewej przedniej zastępującej tętnicy wieńcowej (LAD) stworzono hybrydowe podejście do segmentacji semantycznej. Metoda ta łączy algorytm bazujący na atlasach z podejściem opartym o sieć splotową. Opracowany algorytm składa się z kilku kroków, z czego pierwszym jest odpowiednie przygotowanie danych wejściowych. Na wstępnie przetworzonych skanach CT przeprowadzana jest segmentacja metodą wielo-atlasową, w wyniku czego otrzymuje się zestaw map segmentacji struktury oraz zmodyfikowane obrazy atlasów. Niezależnie, na tych samych danych sieć splotowa poddawana jest procesowi nauki. Następnie, po określeniu parametrów sieci, generowana jest dla danego pacjenta maska tętnicy. Punktem łączącym obie metody jest przetworzenie przez wyuczoną sieć obrazów dopasowanych atlasów do przypadków testowych. Uzyskane z poszczególnych kroków maski danego pacjenta są ze sobą łączone, otrzymując w ten sposób pojedynczą mapę tętnicy LAD. Poniżej zamieszczono schemat opisanego rozwiązania.



Rys. 3.1: Schemat modułu segmentacji tętnicy LAD. Źródło: opracowanie własne.

Podejście to jednak nie może być stosowane do segmentacji loży po guzie w gruczole pierśiowym, z racji dużej zmienności anatomicznej. W zależności od pacjentki miejsce to znajduje się w różnych częściach gruczołu i posiada odmienne wymiary, gdyż nie jest to struktura wywodząca się z anatomii ciała człowieka. Podejście atlasowe ma sens jedynie w przypadku narządów anatomicznych, charakteryzujących się większą spójnością pod względem budowy i lokalizacji. Nawet w różnorodnej grupie chorych różniących się np. masą ciała, serce będzie

łatwiej zdefiniować niż obszar, w którym znajdował się guz nowotworowy. Loża po guzie jest obszarem będącym skutkiem leczenia, stąd też stosowanie metody opartej na atlasach obarczone jest dużą losowością i brakiem spójności w wyznaczeniu tego obszaru. W tej sytuacji algorytm sprowadza się wyłącznie do zastosowania sieci splotowej.



Rys. 3.2: Schemat modułu segmentacji loży po guzie. Źródło: opracowanie własne.

3.1 Wstępne przetwarzanie

Przed przystąpieniem do segmentowania poszczególnych obrazów pacjentów, należy je wstępnie przetworzyć. Proces ten dokonywany jest dwojako: jako wyodrębnienie wycinków rejonów zawierający badany organ oraz przetworzenie skanów zawierających całe ciało pacjenta. W wyniku tego procesu powstają dwa obrazy trójwymiarowe, wraz z zestawem map wzorcowej segmentacji, gotowe do przekazania na wejście poszczególnych metod.

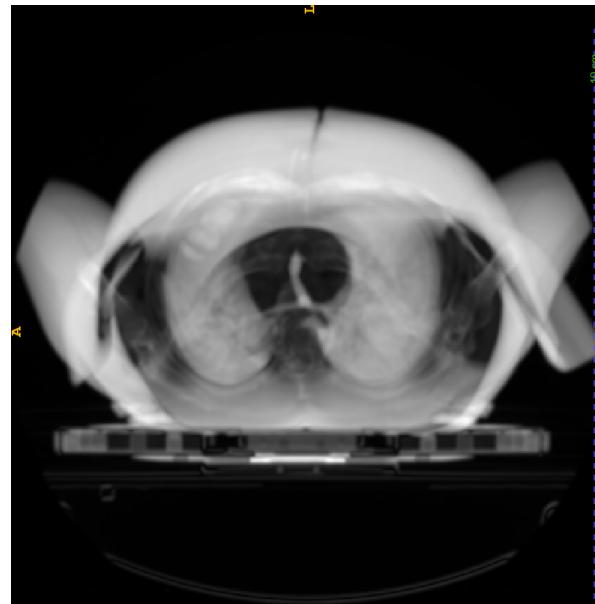
Preprocessing obrazów dla całego pacjenta został odwzorowany analogicznie jak to zostało opisane w pracy Sipińskiego [Sipiński, 2023]. Proces ten składa się z trzech głównych etapów:

- generowania map segmentacji na podstawie konturów struktur biomedycznych,
- normalizacji wartości obrazów do wartości z przedziału $[0; 1]$,
- przycinania i skalowania.

Aby uzyskać ekspercką mapę segmentacji struktur, początkowo tworzone są puste obrazy o rozmiarach danego skanu, po czym uzupełniane są one wartościami w pikselach znajdujących się wewnętrz wielokątów będących konturami danej struktury. W przypadku normalizacji, najpierw dane spoza zakresu $[-1000; 1000]$ przycinane są do wartości granicznych, a następnie skalowane do zakresu $[0; 1]$. Z kolei, aby wyznaczyć okno przycinania, obliczane jest odchylenie standardowe dla każdego piksela pomiędzy przekrojami. Na podstawie tych wartości dobierane jest minimalne okno tak, aby poza nim znajdowały się tylko wartości, które nie przekraczają progu 0,2. Dla danego pacjenta wszystkie przekroje przycinane są tym samym oknem, a następnie są rozciągane do rozmiaru 512×512 pikseli (na rysunku 3.3 przedstawiono poszczególne etapy procesu kadrowania). Tak przygotowane przekroje następnie są składane ponownie w obraz trójwymiarowy.



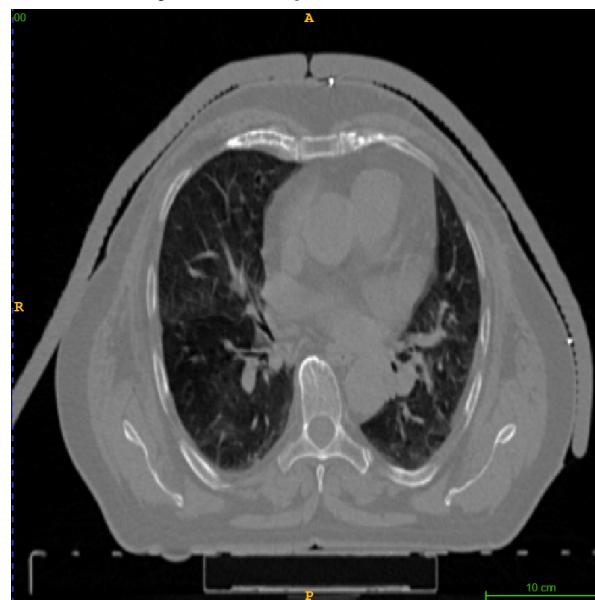
Przykładowy przekrój skanu pacjenta ze znormalizowanymi wartościami do przedziału $[0, 1]$.



Odchylenie standardowe wokseli pomiędzy przekrojami w skanie. Jaśniejszy kolor oznacza większe odchylenie.



Zastosowanie progowania dolnego z wartością 0,2 oraz najmniejszy prostokąt ograniczający, który zawiera wszystkie niezerowe wartości



Efekt końcowy po przycięciu do prostokąta ograniczającego i rozciągnięcia do oryginalnego rozmiaru.

Rys. 3.3: Ilustracja przedstawiająca poszczególne kroki procesu kadrowania przekrojów całego ciała pacjenta ze skanów CT. Źródło: opracowanie własne.

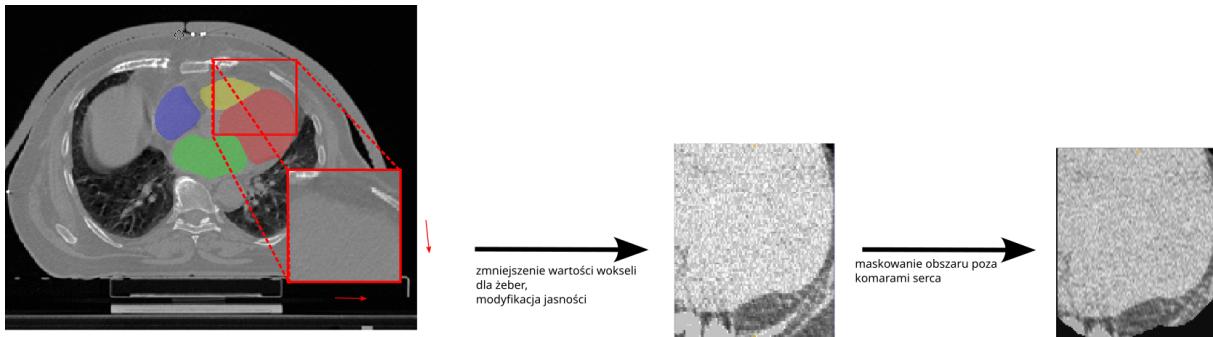
Wiedząc, że tętnica wieńcowa LAD anatomicznie biegnie w bruzdzie międzykomorowej przedniej ku dołowi aż do wcięcia koniuszka serca oraz loże po guzie dotyczą wyłącznie pacjentek chorych na lewostronny nowotwór piersi, obrano koncepcje zawężenia obszaru segmentacji do obszaru zawierającego tylko badane struktury.

Aby wyodrębnić obszar minimalny, w obrębie którego znajduje się badana struktura medyczna, skorzystano z narzędzia TotalSegmentator [Wasserthal, 2024a]. Oprogramowanie to pozwala na zlokalizowanie narządów wewnętrznych znajdujących się w sąsiedztwie rozważanej struktury. W celu segmentacji LAD, zewnętrzny moduł uruchamiany był w trybie segmentacji komór serca w wysokiej rozdzielczości. Jako wynik oprócz samych komór generowane były maski przedzionków, mięśnia sercowego oraz tętnicy płucnej. Segmentacje naczyń krwionośnych oraz obydwu przedzionków zostały usunięte, a mięsień sercowy został wcielony w maskę komory lewej. Aby segmentować lożę po guzie, narzędzie to uruchamiane było w trybie *total*, segmentując tym samym wszystkie dostępne struktury. Spośród nich wybrano maski mostka oraz chrząstki żebrowej, które posłużyły w dalszym etapie do zawężenia obszaru poszukiwań miejsc, w których przed usunięciem znajdowały się guzy.

Kolejnym etapem przetwarzania jest wycięcie z trójwymiarowego obrazu tomografii komputerowej minimalnego prostopadłościanu zawierającego badaną strukturę. Określenie tego obszaru dokonywane jest w oparciu o wyniki segmentacji TotalSegmentatora.

Współrzędne wycinanego prostopadłościanu danych zostały dobrane w sposób eksperymentalny tak, aby obejmował przestrzeń LAD dla pacjentów w zbiorze danych. Aby określić optymalne współrzędne wykrawanego obszaru z LAD, najpierw scalono maski komór w jedną. Uzyskana w ten sposób maska struktur nie pokrywała się ze skanem pacjenta, dlatego też obrazy CT zostały przekształcone poprzez odbicie względem osi wzdłuż kręgosłupa a następnie zgodnie z osią poprzeczną do niego, aby dopasować maskę do obrazu. Następnie ograniczono skan tylko do obszaru wyznaczonego przez tę maskę. Na podstawie najdalej wysuniętych punktów krańcowych mapy komór powstała bryła (prostopadłościan), której koordynaty posłużyły do wycięcia ze skanu tomografii komputerowej obszaru komór. Powstały w ten sposób wycinek CT poddano modyfikacjom, odpowiednio zmieniając jego współrzędne. Finalnie dobrane parametry przekształcenia zapewniły przesunięcie tylnej ściany prostopadłościanu, położonej najbliżej kręgosłupa, o 30 wokseli w kierunku mostka. Ściana przednia, położona najbliżej żeber, zmieniła swoją lokalizację przesuwając się o 20 wokseli w przód. Lewa boczna i górna ściana pozostały bez zmian. Z kolei prawa boczna została przemieszczona o 20 wokseli w lewo. Dolna ściana bryły obniżyła się o 20 wokseli.

Dodatkowo w przypadku modułu dedykowanego LAD, w powstałym wycinku dokonano zmniejszenia wartości wokseli reprezentujących kości żebrowe (1000 HU), które to pierwotnie znaczco odznaczały się od pozostałych struktur. Wszystkie wartości wokseli powyżej progu 50 HU zostały zredukowane do zera, wartości te zostały dobrane eksperymentalnie. Dopasowano również jasność obrazów tak, aby szukany obszar był łatwiejszy do zlokalizowania gołym okiem. W tym celu skorzystano z korekcji gamma, która jest techniką używaną aby zmodyfiko-



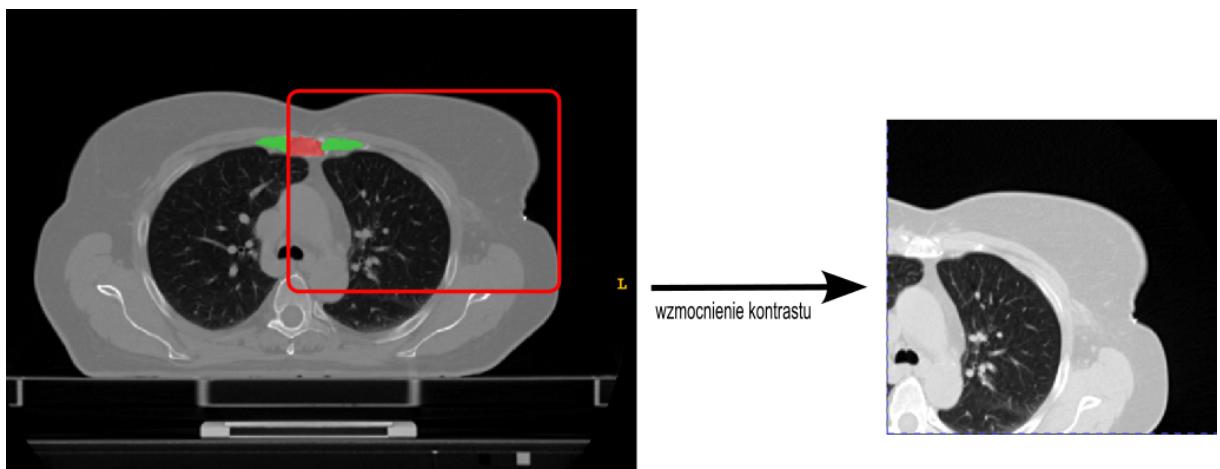
Rys. 3.4: Schemat wycinania obszaru LAD na przykładzie jednego z przekrojów skanu CT z zaznaczonymi maskami komór i przedsięwziętymi przez narzędzie TotalSegmentator. Źródło: opracowanie własne.

wać jasność obrazu bez zmiany jego kontrastu. Metoda modyfikuje poziomy jasności w obrazie poprzez zastosowanie potęgi liczbowej (gamma) do wartości wokseli. Parametr gamma został ustawiony na wartość 5,5 co przyczyniło się do przyciemnienia obrazu. Co więcej, usunięto struktury znajdujące się poza obszarem wyznaczonym przez maski komór serca, powiększonej o 20 wokseli w każdym kierunku. Ostatecznie wycinki rozciągnięto do ujednoliconego rozmiaru 256×256 pikseli.

W przypadku loży po guzie sytuacja jest nieco inna, gdyż w zależności od pacjentki guz znajdował się w różnych częściach lewego gruczołu piersiowego. Dlatego też, dysponując maską chrząstki żebrowej oraz mostka, przeprowadzono proces wykrawania tego obszaru. W tym przypadku, obróceniu poddane zostały maski wygenerowane przez narzędzie TotalSegmentator, a nie skany jak to ma miejsce w procesie wycinania obszarów z LAD. Maski te zostały odbite względem osi wzdłuż kręgosłupa a kolejno zgodnie z osią prostopadłą do niego. Wycinanie polegało na ograniczeniu po szerokości skanu (współrzędne x) od prawej krawędzi mostka do lewej strony ciała w kierunku wzdłuż kręgosłupa. Wzdłuż ciała wycinano od punktu chrząstki żebrowej położonego najbliżej głowy obniżonego o 20 wokseli, ku najniżej położonemu punktowi mostka przesuniętego o 30 wokseli w dół (współrzędne z). Zaś w poprzek (współrzędne y) skan ograniczono od przodu ciała pacjentki do płaszczyzny chrząstki, plus 50 wokseli w głąb ciała. Co więcej, modyfikacji został poddany kontrast wycinka skanu. Przekształcenie to polegało na rozciągnięciu wartości wokseli obrazu tak, aby pokryły szerszy zakres intensywności. Metoda ta bazuje na 2 i 98 percentylu wartości wokseli z obrazu, które wyznaczają minimalne i maksymalne wartości intensywności. Modyfikacja wartości wokseli ze skanu dokonywana jest w oparciu o wzór:

$$I = \frac{I - p2}{p98 - p2}, \quad (28)$$

gdzie I określa wartości wokseli z obrazu, zaś $p2, p98$ to wartości 2 i 98 percentyla danych. Zmiana ta miała na celu wyróżnienie tkanek gruczołu piersiowego, tak aby odróżniały się na tle innych struktur.



Rys. 3.5: Schemat wycinania obszaru loży po guzie na przykładzie jednego z przekrojów skanu CT z zaznaczonymi maskami mostka i chrząstki żebrowej. Źródło: opracowanie własne.

Analogicznie, jak to zostało przedstawione w przypadku wstępniego przetwarzania skanów całego ciała pacjenta, wygenerowano eksperckie mapy segmentacji zapisane w plikach DICOM. Następnie, otrzymane w ten sposób maski, dopasowane do oryginalnego skanu pacjenta, poddano procesowi wycinania, który przebiegał identycznie jak w przypadku skanów. W wyniku czego otrzymano maski w rozmiarze 256×256 odpowiadające strukturom z obrazu wycinka.

Tak przygotowane obrazy stanowiły dane wejściowe metody atlasowej. Natomiast przed przetworzeniem obrazów przez sieć splotową, zostały one poddane transformacji obejmującej ujednolicenie rozmiarów wycinków, jak i obrazów całych ciał pacjentów, do rozmiaru $256 \times 256 \times 256$ oraz skalowanie intensywności obrazów, w wyniku czego dane obrazu zostają znormalizowane do przedziału $[0; 1]$. Co więcej, obrazy zostały obrócone o 180° w prawą stronę (zgodnie z ruchem wskazówek zegara). Transformacje te dotyczą zarówno przetwarzania dla segmentacji LAD jak i loży po guzie. Stosowanie ich było konieczne z racji korzystania z przetrenowanych wag MedicalNet [Chen et al., 2019], która to dokonywała analogicznych przekształceń.

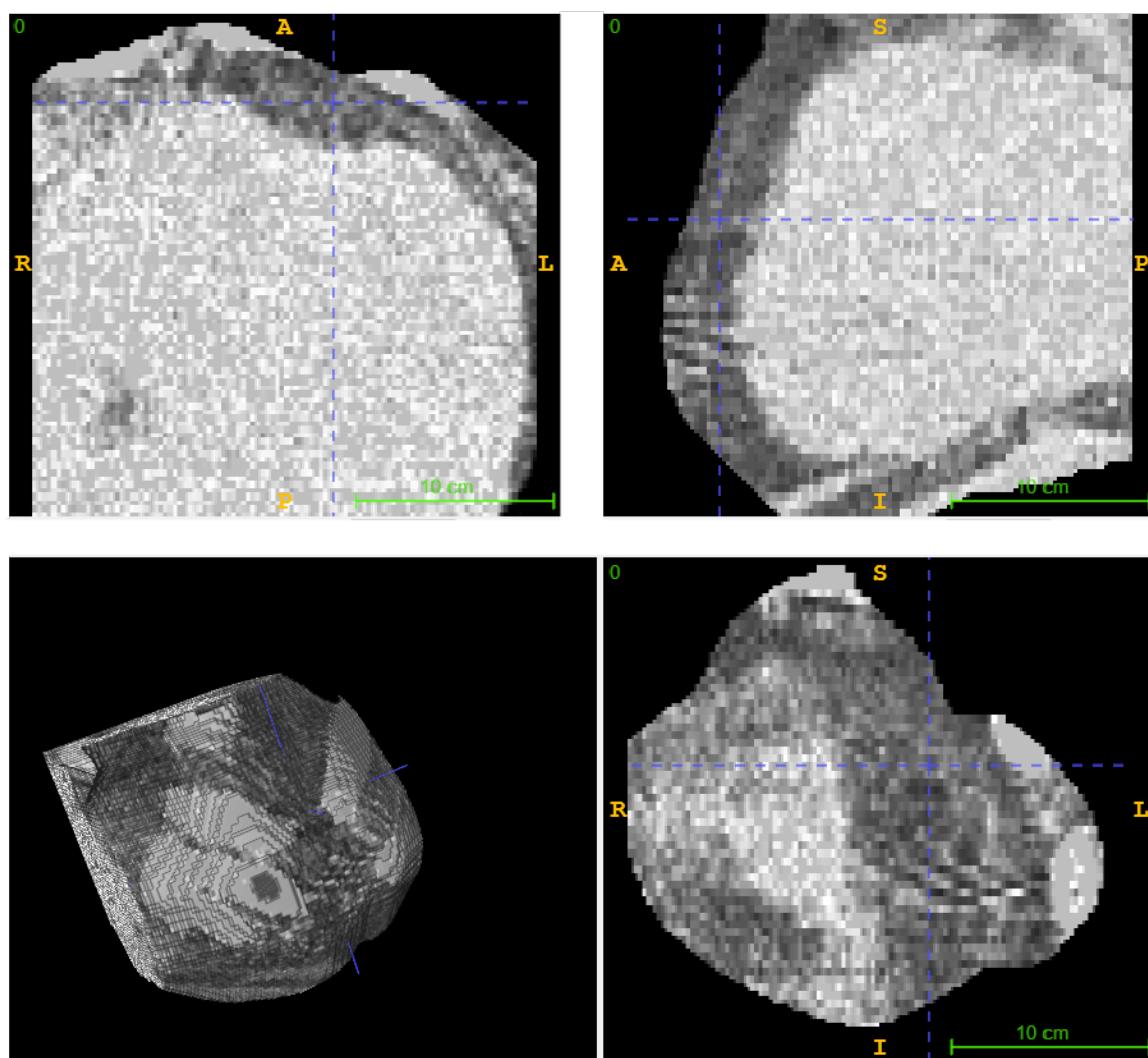
Dla uzyskania większej czytelności spisano wyżej opisane operacje dla lewej przedniej za-

stępującej tętnicy wieńcowej LAD w postaci listy kroków:

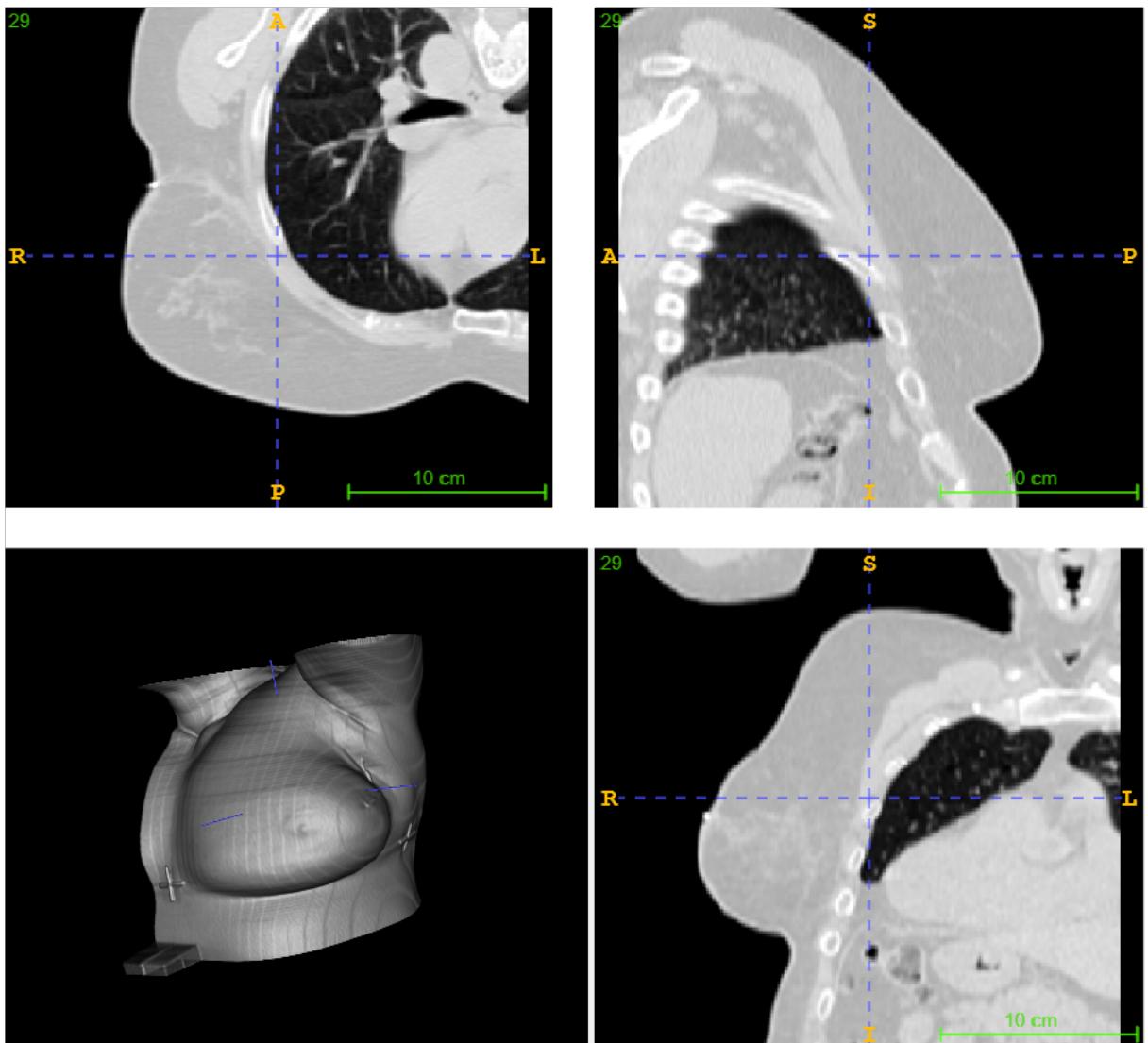
1. segmentacja TotalSegmentator w trybie komór serca w wysokiej rozdzielczości,
2. modyfikacje uzyskanych masek (usunięcie tętnicy płucnej i przedśionków, scalenie mięśnia sercowego z komorą lewą), scalenie komór w jedną mapę,
3. odbicie skanu tomografii komputerowej względem osi wzdłuż kręgosłupa, a następnie zgodnie z osią poprzeczną do niego,

4. wycinanie skanu oraz wzorcowej maski struktury bazując na współrzędnych maski komór serca,
5. redukcja do zera wartości w wycinku powyżej 50 HU (ujednolicenie żeber),
6. modyfikacja jasności obrazu,
7. rozciągnięcie do rozmiaru 256×256 ,
8. usunięcie danych poza obszarem komór serca powiększonego o 20 wokseli w każdym kierunku,
9. skalowanie intensywności, normalizacja, zmiana rozmiaru do $256 \times 256 \times 256$, obrót o 180° w prawą stronę (transformacje do przetwarzania przez sieć).

Wynik powyższych operacji został przedstawiony na rysunku 3.6.



Rys. 3.6: Przykładowy wycinek skanu CT obszaru serca po procesie wstępnego przetwarzania. Źródło: opracowanie własne.



Rys. 3.7: Przykładowy wycinek skanu CT obszaru lewego gruczołu piersiowego po procesie wstępniego przetwarzania. Źródło: opracowanie własne.

Transformacje do stworzenia wycinka lewej strony ciała pacjentki wykorzystywanej do segmentacji loży po guzie prezentują się następująco:

1. segmentacja TotalSegmentator w trybie *total*,
2. usunięcie wszystkich map struktur, z wyjątkiem maski mostka i chrząstki żebrowej,
3. odbicie maski mostka i chrząstki żebrowej względem osi wzdłuż kręgosłupa, a następnie zgodnie z osią poprzeczną do niego,
4. wycinanie skanu oraz wzorcowej maski struktury, bazując na współrzędnych wyznaczonych przez maski mostka i chrząstki żebrowej,
5. modyfikacja kontrastu,
6. rozcięgnięcie do rozmiaru 256×256 ,

7. skalowanie intensywności, normalizacja, zmiana rozmiaru do $256 \times 256 \times 256$, obrót o 180° w prawą stronę (transformacje do przetwarzania przez sieć).

Rysunek 3.7 przedstawia wycinek skanu tomografii komputerowej powstały poprzez wykonanie powyżej opisanych operacji.

3.2 Segmentacja właściwa

Metoda atlasowa

Punktem odniesienia w przypadku atlasowej segmentacji obrazu jest metoda i parametry zaprezentowane w pracy Sipińskiego [Sipiński, 2023]. Stosowano w tym celu narzędzie Greedy [Yushkevich, 2024] [Yushkevich et al., 2016], z wykorzystaniem funkcji kosztu sumy kwadratów różnic (SDD). Proces składał się z trzech operacji: dopasowanie obrazów według momentów bezwładności, elastyczne rejestrowanie oraz zastosowania odkształceń do obrazów i map segmentacji atlasu.

Dopasowanie według momentów jest słuszną strategią, gdy dwa obrazy mają podobną zawartość, ale są tak rozbieżne, że tryby afierny i sztywny zawodzą. Dopasowanie to jest szczególnie przydatne w przypadku obiektów binarnych. Na tym etapie, w zastosowanej wersji, zostają wyrównane środki masy dwóch obrazów. Jako wynik tejże operacji otrzymuje się macierz przekształceń, analogicznie jak w przypadku rejestracji afiernej czy sztywnej.

Do dokonania rejestracji odkształceń wykorzystano macierz z poprzedniego etapu, uruchamiając krok ten w trybie wielu rozdzielczości. Wynikiem jest trójwymiarowy obraz, w którym każdy woksel zawiera informacje składowych x, y, z wektora przemieszczenia. Obraz odkształcenia określa przemieszczenie dla każdego woksela, które mapuje środek tego woksela na odpowiadającą mu lokalizację w skanie badanego pacjenta.

Posiadając powyższe informacje, dokonywana jest deformacja atlasu i maski segmentacji, w wyniku czego uzyskano mapę szukanej struktury oraz przekształcone atlasy.

Za atlasy uznano kilkadziesiąt skanów pacjentów ze zbioru treningowego, w wyniku czego jako rezultat działania metody dla danego pacjenta otrzymano kilkadziesiąt masek segmentacji LAD.

Metoda neuronowa

Do celów segmentacji za pomocą sztucznej sieci splotowej, zastosowano popularną sieć 3D U-Net stosowaną powszechnie do tego typu problemów. Wykorzystano implementację udostępnioną poprzez bibliotekę MONAI [Cardoso et al., 2022], dedykowaną głębokiemu uczeniu w medycynie. Wybrana eksperymentalnie architektura zawiera enkoder zbudowany ze szkieletem ResNet-34 do segmentacji LAD, zaś ResNet-50 do loży po guzie. U-Net został poddany procesowi dotrenowania, douczając wstępnie wytrenowane wagi MedicalNet [Chen et al., 2019]

w całej architekturze sieci. MedicalNet stanowi zbiór pretrenowanych modeli głębokiego uczenia, które zostały zaprojektowane specjalnie do analizy danych medycznych 3D, takich jak tomografia komputerowa. Stosowanie wag przetrenowanych wymaga korzystania wyłącznie z jednego kanału wejściowego. Sieć ta segmentuje trójwymiarowe obrazy, w wyniku czego uzyskujemy również przestrzenną maskę badanej struktury. Pozostałe parametry sieci stosowane w wersji domyślnej.

Do treningu zastosowano optymalizator *Adam* z wartością współczynnika nauki 0,001, zaś jako funkcję kosztu stosowano kombinację *CrossEntropyLoss* i *DiceLoss*, udostępnioną poprzez funkcję *DiceCELoss*. Aby zbalansować liczebność klas pikseli, stosowano ważoną funkcję straty, gdyż woksele obszaru badanej struktury stanowią mniejszość w przetwarzanym obrazie. Trening odbywał się na przestrzeni 20 epok, korzystając z techniki wcześniego zatrzymywania (ang. early stopping), aby nie doprowadzić do przeuczenia sieci oraz techniki normalizacji wsadowej. Liczba epok została wybrana eksperymentalnie, pozwalając na przeprowadzenie procesu nauki w akceptowalnym czasie.

Metoda atlasowo-neuronowa

Dysponując modelem sieci splotowej, douczonym na danych treningowych oraz zbiorem obrazów dopasowanych atlasów do skanów pacjentów testowych, dokonano fuzji metody atlasowej i neuronowej. Połączenie to polega na przetworzeniu przez sieć obrazów zmodyfikowanych atlasów. Skany te na wstępie poddawane są transformacjom wykonywanym na obrazach przed przekazaniem ich do sieci (rozszerzenie do rozmiaru $256 \times 256 \times 256$, obrócenie oraz normalizacja). W rezultacie działania metody dla danego pacjenta otrzymujemy tyle map struktury, ile było atlasów stosowanych w segmentacji atlasowej. Maski te, następnie poddawane zostają procesowi głosowania — agregacja dokonywana jest woksel po wokselu, dla każdego z nich wybierana wartość 1 (miejscem struktury), gdy co najmniej dwie maski są jednomyślne, w innym przypadku woksel traktowany jest jako tło (wartość 0). Metoda głosowania została dobrana na podstawie przeprowadzonych eksperymentów opisanych w kolejnym rozdziale.

Agregacja metody atlasowej, neuronowej i atlasowo-neuronowej

Posiadając wyniki segmentacji uzyskane przez zastosowanie metody atlasowej, neuronowej i atlasowo-neuronowej, dokonano próby ich połączenia. Aby to uczynić wymagane jest, żeby wszystkie mapy struktury znajdowały się w tej samej przestrzeni, dlatego też wynikowe maski z metody atlasowej poddano transformacji wykonywanej na danych przed przetworzeniem ich przez sieć U-Net (rozszerzenie do rozmiaru $256 \times 256 \times 256$ i obrócenie o 180° w prawą stronę). W przypadku metody atlasowo-neuronowej, w tym podejściu do segmentacji pomijany jest krok łączenia masek, czyli przetwarzany zostaje zestaw map uzyskany poprzez predykcję sieci U-Net na atlasach dopasowanych do obrazów testowych. Ogólnie z wszystkich trzech metod, dla danego pacjenta, uzyskuje się $2 \cdot g + 1$ masek, gdzie g określa liczbę używanych atlasów.

Dysponując maskami tętnicy danego pacjenta osadzonych w tej samej przestrzeni, wykonywana jest agregacja polegająca na głosowaniu, w wyniku którego zgodność minimum dwóch masek decyduje o istnieniu struktury w danym wokselu. Łączenie wykonywane jest analogicznie jak to miało miejsce w metodzie atlasowo-neuronowej.

Następnie maskę, będącą wynikiem agregacji, poddano modyfikacji usuwając oznaczenia struktury tam, gdzie obraz skanu nie zawiera żadnej tkanki. Na podstawie znormalizowanego obrazu CT wybierane są współrzędne wokseli, których wartość jest mniejsza niż próg 0,2 (wartość dobrana eksperymentalnie). W wyniku czego otrzymano mapę określającą czy dany woksel przedstawia ciało, oznaczone wartością 0, czy też powietrze — wartość 1. Na bazie tej mapy, z maski struktury LAD usuwane zostają oznaczenia tętnicy w wokselach, których odpowiadająca wartość w skanie określała powietrze.

4 Wstępne eksperymenty

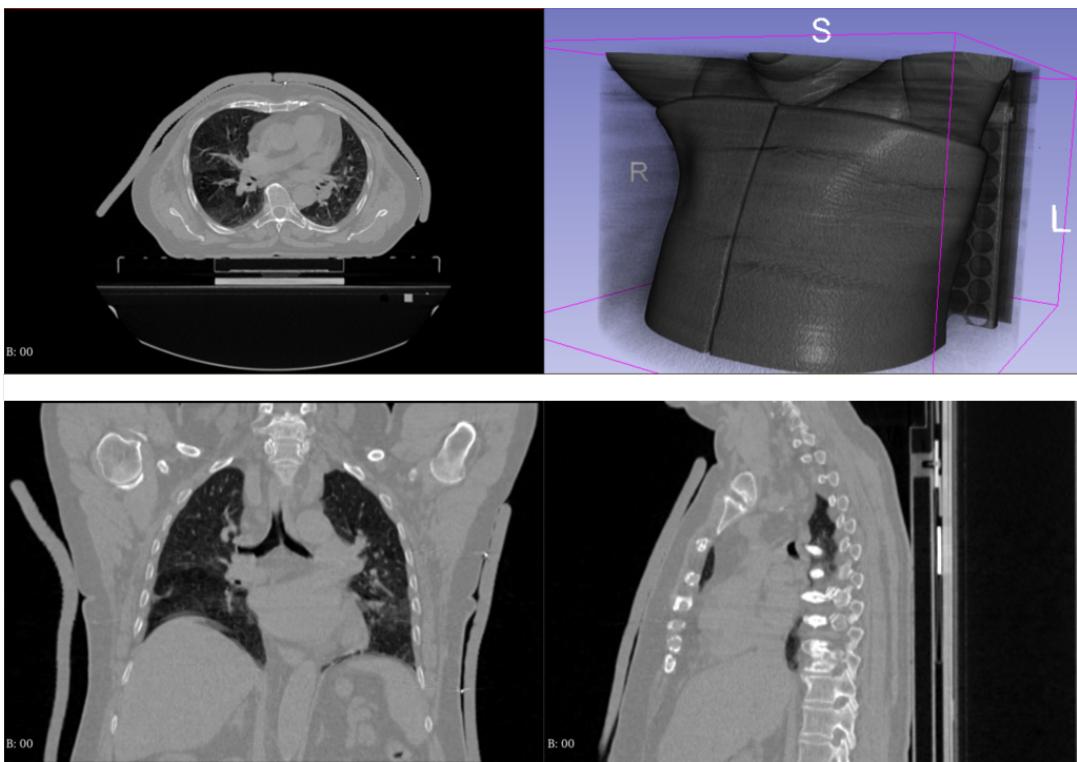
Tworzenie architektury rozwiązania było złożonym procesem, podczas którego koncepcja wielokrotnie poddawana była modyfikacji tak, aby finalnie otrzymać model pozwalający na naj-dokładniejsze segmentowanie. Przeprowadzono szereg eksperymentów, które na celu miały wybranie wartości tylko niektórych argumentów poszczególnych metod segmentacji. Używane implementacje algorytmów posiadają wiele parametrów, stąd też skupiono się na tych, które potencjalnie mają największy wpływ na działanie metody. Większa ich część została ustalona empirycznie, z reguły posiada domyślne wartości, zaś wartości kilku z nich (ze względu na niewielką liczbę kombinacji) dopasowywane są na podstawie wyników eksperymentów tak, aby znaleźć najlepsze rozwiązanie.

4.1 Zbiór danych

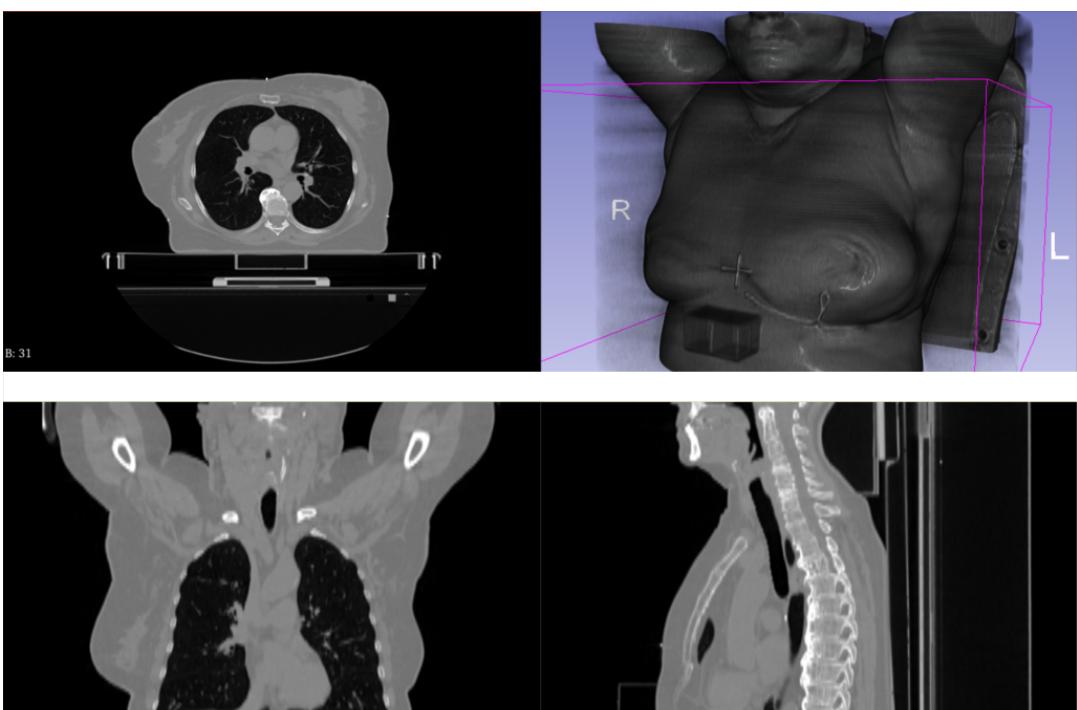
Baza danych powstała dzięki udostępnionych przez Wojewódzkie Wielospecjalistyczne Centrum Onkologii i Traumatologii imienia Mikołaja Kopernika w Łodzi wyników badań tomografii komputerowej (CT) tułowia pacjentów. Zbiór ten składa się 62 zestawów tomografii, zapisanych w formacie DICOM. Format ten jest jednym z formatów plików medycznych, pozwalających na przechowywanie nie tylko samych obrazów, ale również zawiera dane wrażliwe pacjenta jak i metadane samych skanów. Dane zostały dostarczone z ośrodka w zanonimizowanej postaci, niezawierającej imion i nazwisk. W zależności od pacjenta, wyniki CT opisane są na różnej ilości skanów, od 86 do 308. Każdy z nich jest rozmiaru 512×512 pikseli. Występują też różne odstępy między przekrojami chorych, które zawierają się w przedziale 1,5 mm do 5 mm. Z kolei odstępy pomiędzy środkami sąsiadujących pikseli w przekroju skanu wynoszą ok. 0,98 mm, 1,27 mm lub 1,60 mm w zależności od pacjenta.

Wśród plików formatu DICOM zawarte są również informacje o konturach wybranych struktur wymaganych do zaplanowania radioterapii w postaci wielokątów umieszczonych na przekrojach obrazów. W przypadku nowotworów umiejscowionych w górnej części tułowia jest to przykładowo serce, tarczyca czy rdzeń kręgowy. Aby uzyskać maskę segmentacji danego organu należy określić dla każdego przekroju, które piksele mają się znaleźć wewnętrz tych wielokątów.

Zbiór ten można podzielić na dane pacjentów zawierające segmentację serca, komór sercowych oraz tętnicy LAD. Byli to z reguły pacjenci posiadający nowotwór płuc (29 przypadków). Drugi podzbiór skanów to wyniki badania pacjentek z lewostronnym nowotworem piersi. Obrazy te zawierają segmentację ekspercką serca, tętnicy LAD, gruczołów piersiowych oraz loży po guzie (33 pacjentki). Wszystkie skany w zbiorze danych zawierają wyłącznie wyniki z tomografii komputerowej, bez podanego środka kontrastowego. Na rysunku 4.1 oraz 4.2 zamieszczono przykładowe skany pacjentów.



Rys. 4.1: Przykładowy skan tomografii komputerowej ze zbioru danych pacjentki z nowotworem piersi po operacji mastektomii. Źródło: opracowanie własne.



Rys. 4.2: Przykładowy skan tomografii komputerowej ze zbioru danych pacjentki z nowotworem piersi po operacji oszczędzającej gruczoł piersiowy. Źródło: opracowanie własne.

4.2 Lewa przednia zastępująca tętnica wieńcowa (LAD)

Zbiór danych, składający się z wyników tomografii komputerowej 62 pacjentów, został podzielony na podzbiory treningowy, walidacyjny i testowy w stosunku 80 : 10 : 10, co stanowiło 50 pacjentów treningowych, 6 walidacyjnych i tyle samo testowych. Podział ten jest stały, niezmienny podczas wykonywania eksperymentów. W przypadku metody atlasowej zbiór walidacyjny dodatkowo posłużył jako dane testowe, zaś zbiór treningowy stanowił atlasy.

Preprocessing

Eksperymenty w kontekście przygotowania skanów przed procesem segmentowania dotyczyły głównie doboru rozmiaru wycinanego prostopadłościanu. Podczas kolejnych prób dostosowania wielkości obszaru serca weryfikowano skuteczność kadrowania za pomocą metody atlasowej. Każdorazowo dla danego przypadku badanego rozmiaru wycinka, dla jednej pary atlasu i obrazu pacjenta testowego, szukano transformacji atlasowej dającej lepszy od poprzedniego rezultat. Metoda atlasowa uruchamiana była zgodnie z parametrami wykorzystywanymi w pracy Sipińskiego [Sipiński, 2023] (więcej w sekcji dedykowanej metodzie atlasowej), a wyniki sprawdzane były wizualnie.

Pierwszym podejściem było surowe wycięcie bazujące na wysegmentowanych, przez narzędzie TotalSegmentator, komorach serca. Polegało ono na wyodrębnieniu ze skanu obszaru zawierający tylko rejon określony przez maski komór. Rozwiążanie to niestety nie zawierało LAD, gdyż tętnica ta leży przy ścianie, na styku komór serca. Dlatego też, w kolejnych próbach modyfikowano wycinanie zmieniając wielkość prostopadłościanu. Początkowo bazowano na rozmiarach serca danego pacjenta, odpowiednio przemnażając współrzędne komór przez stałe zależne od lokalizacji serca. Podejście to wydawało się najbardziej elastycznym z możliwych, jednakże trudnością było dobranie współczynników tak, aby obszar ten był poprawnie wyodrębniony dla wszystkich pacjentów. W zbiorze skanów znajdują się przypadki ciał o bardzo różnych rozmiarach serc, w wyniku czego dla kilku skanów problematyczne okazało się dostosowanie parametrów tak, aby LAD znajdował się wewnątrz wycinanego obszaru. Testowano również koncepcje bazującą na sztywnym wycinaniu, najpierw wyodrębniając obszar komór, a następnie odpowiednio przesuwając krawędzie wycinanej bryły. W tym przypadku wynikowe obrazy dla każdego z pacjentów zawierały w całości poszukiwaną tętnicę.

Po określeniu tym sposobem, dla danego pacjenta, wymiarów prostopadłościanu zawierającego część komór serca oraz miejsce, w którym znajduje się LAD, maska rzeczywiście lokalizacji tętnicy wieńcowej została analogicznie przycięta. Tak przygotowane obrazy (przykład przedstawiony na rysunku 4.3) posłużyły jako wejście dla metody atlasowej. W wyniku jej działania, co prawda uzyskano niezerową maskę LAD, lecz jej lokalizacja znajdowała się gdzieś w okolicy żeber. Obszar ten, czyli kości, na skanie tomografii komputerowej jest o wiele jaśniejszy niż otaczające tkanki, co najprawdopodobniej stanowiło przyczynę takiego rezultatu.

Aby zmusić metodę do segmentowania tętnicy, a nie kości, obszar kości żebrowych w wy-



Rys. 4.3: Przekrój wycinka serca bez obróbki.



Rys. 4.4: Przekrój wycinka serca z maskowaniem żeber.



Rys. 4.5: Przekrój wycinka serca ze zmodyfikowaną jasnością i zamaskowanymi żebrami.

Źródło: opracowania własne.

cinkach został ujednolicony — dla wartości wokseli większych od progu 50 zostały one zmniejszone do zera. Graniczna wartość została dobrana eksperymentalnie, wizualnie sprawdzając, przy jakim progu zdecydowana powierzchnia kości zostanie ujednoliciona. Operacja ta przyczyniła się do ogólnego rozjaśnienia obrazów, co przedstawia rysunek 4.4, jednocześnie tracąc zdolność do naocznej lokalizacji miejsca LAD. Natomiast metoda atlasowa, dla tak przygotowanego skanu, oznaczyła jako tętnicę obszar pomiędzy skórą, a żebrami. Rezultat ten znaczająco się różni od rzeczywistej lokalizacji tętnicy.

Kolejna modyfikacja stanowiła modyfikację jasności wycinka, gdzie żebra zostały zamaskowane, jako wynik uzyskano obraz przedstawiony na rysunku 4.5. W tej próbie zastosowanie metody atlasowej doprowadziło do zlokalizowania LAD poza obszarem serca. W konsekwencji obszar niebędący sercem został zamaskowany, traktując go jak powietrze (zmieniając wartość wokseli poza sercem na -1000 HU) z uwzględnieniem marginesu 20 wokseli. Finalnie, otrzymano wycinek, w którym jaśniejsze miejsca określają komory serca, zaś ciemniejsze stanowią przestrzeń poza nim gdzie znajduje się szukana tętnica.

Sprawdzano też kadrowanie zdjęć z uwzględnieniem zakresu Hounsfielda dla krwi, wybierając z wycinka tylko dane zawierające wartości z przedziału $[30, 70] \text{ HU}$. Uzyskano w ten sposób wycinek zawierający tylko miejscowo uzupełnione dane, nieuwzględniające obszaru LAD, co spowodowało odrzucenie tej koncepcji.

Ostatecznie wykorzystywano sztywne wycinanie, bazując na współrzędnych prostopadłoscianu wyznaczonych przez maskę komór serca. Przekształcenia prezentują się następująco:

- po grubości ciała (współrzędne y) przesunięto o 30 wokseli granicę komór od strony pleców w kierunku przodu ciała, zaś o 20 wokseli od ściany komór w kierunku żeber,
- po szerokości ciała (współrzędne x) przesunięto z prawej strony krawędź prostopadłoscianu o 20 wokseli w kierunku środka komór, z lewej strony granicę bryły określa najdalej wysunięty punkt ściany komór,

- po długości ciała (współrzędne z) przesunięto podstawę prostopadłościanu o 20 wokseli w dół, granicę górnej ściany określa najwyższy wysunięty punkt ściany komór.

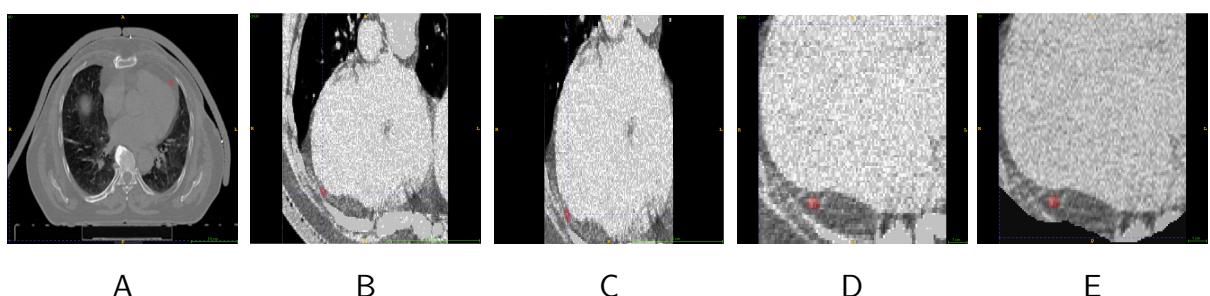
Zastosowano również maskowanie żeber, z jednoczesnym dopasowaniem jasności (rysunek 4.5) oraz usunięty został obszar skanu znajdujący się poza maską komór, poszerzoną o 20 wokseli w każdym kierunku.

Metoda atlasowa

Bazując na wiedzy dotyczącej segmentowania atlasami zawartej w pracy Sipińskiego [Sipiński, 2023], jako kolejny etap badania, przeprowadzono analizę wpływu wybranych parametrów na dokładności segmentacji LAD.

Jako pierwsze sprawdzano, w jaki sposób zawężenie obszaru poszukiwań ze skanów tomografii komputerowej wpłynie na precyzję lokalizowania tętnicy. Do każdego pomiaru wykorzystywano ten sam zestaw parametrów wejściowych narzędzia Greedy [Yushkevich, 2024] stosowanych uprzednio we wcześniej wspomnianej pracy. Zmianie został poddawany zakres danych ze skanów tomografii pacjentów, będące danymi wejściowymi do metody — jako atlas oraz obraz, dla którego celem jest znalezienie umiejscowienia LAD.

Obrazy skanów przed kadrowaniem zostały poddane przetwarzaniu zgodnie z założeniami preprocessingu rozwiązania. Badano 5 różnych wielkości trójwymiarowego obrazu — obraz całego ciała pacjenta opisanego na przekrojach o rozmiarze 512×512 pikseli (rysunek 4.6 A) oraz 4 różnych podejść do wykrawania obszaru LAD (256×256). Obraz 4.6 B powstał poprzez wycięcie obszaru komór serca z marginesem 30 wokseli, natomiast rysunek 4.6 C przedstawia wyniki wycięcia z marginesem 10 wokseli. Z kolei obrazy 4.6 D i E powstały według opisu preprocessingu w dwóch wariatach: z maskowaniem obszaru znajdującego się poza rozszerzoną maską komór serca o 20 wokseli w każdym z kierunków (rysunek 4.6 E) oraz bez niego (rysunek 4.6 D). Analogicznie dla kadrowania skanu danego przypadku badawczego, modyfikowano wzorcowe maski tętnicy pacjentów, tak aby były one zgodne ze zmodyfikowanym skanem.



Rys. 4.6: Obrazy przekroju skanu pacjenta w danym formatowaniu: A - całe ciało; B - wycinek obszaru komór z 30 wokselowym marginesem; C - wycinek obszaru komór z 10 wokselowym marginesem; D - wycinek proponowanego rozwiązania bez maskowania; E - wycinek proponowanego rozwiązania z maskowaniem wraz ze wzorcową maską LAD.
Źródło: opracowanie własne.

Dla tak przygotowanych obrazów dokonywano dopasowania momentów bezwładności oraz rejestrowano odkształcenia. Stosując następnie te transformację do obrazu atlasu i jego maski, w rezultacie dla każdego z pacjentów otrzymano 50 masek struktury. Wyniki średnich wartości miar jakości segmentacji dla wszystkich przypadków testowych zebrane w tabeli 1.

formatowanie	DSC	HD [mm]	VR	MSD [mm]	PPV	TPR
A	$0,04 \pm 0,07$	$30,02 \pm 13,74$	$0,93 \pm 0,51$	$8,72 \pm 5,89$	$0,04 \pm 0,08$	$0,05 \pm 0,08$
B	$0,13 \pm 0,10$	$24,81 \pm 13,91$	$0,74 \pm 0,41$	$4,09 \pm 2,41$	$0,15 \pm 0,12$	$0,12 \pm 0,11$
C	$0,13 \pm 0,12$	$54,41 \pm 32,03$	$0,75 \pm 0,37$	$7,99 \pm 6,82$	$0,14 \pm 0,12$	$0,13 \pm 0,13$
D	$0,11 \pm 0,11$	$62,40 \pm 33,41$	$0,93 \pm 0,40$	$8,57 \pm 6,69$	$0,11 \pm 0,11$	$0,12 \pm 0,12$
E	$0,09 \pm 0,10$	$72,87 \pm 42,26$	$0,81 \pm 0,40$	$11,39 \pm 9,79$	$0,11 \pm 0,11$	$0,09 \pm 0,11$

Tabela 1: Średnie wartości miar jakości segmentacji LAD dla różnych zakresów obszaru poszukiwań metody atlasowej. Źródło: opracowanie własne.

Z zamieszczonych wyników, jednoznacznie metoda bazująca na atlasach wyodrębniała najlepiej strukturę z obszaru komór serca powiększonego o 30 wokseli w każdym kierunku (rysunek 4.6 B). Miary podobieństwa uzyskały najwyższe ze wszystkich rezultatów wartości, z jednocześnie otrzymaniem najmniejszych odległości pomiędzy strukturami. Segmentacja w obrębie całego skanu wypadła najsłabiej, dając dokładność według Dice'a rzędu kilku procent. Kolejno zmniejszanie obszaru do pewnego momentu, czyli zakresu komór z marginesem 10 wokseli (rysunek 4.6 C), powodowało zwiększenie skuteczności metody. Dla proponowanego sposobu kadrowania skanu bez maskowania obszaru poza sercem zaobserwować można pogorszenie wyodrębniania struktury. Natomiast samo maskowanie danych niewchodzących w skład serca również nie przyniosło oczekiwanej polepszenia działania metody. Niemniej jednak wszystkie eksperymenty cechuje ogólnie słaba dokładność segmentacji, maksymalnie algorytm zlokalizował LAD ze skutecznością 13%, co stanowi tylko niewielką część poprawnie oznaczonej struktury. Należy również tutaj wspomnieć, że narzędzie Greedy prowadzi obliczenia niedeterministycznie, w wyniku czego wielokrotne uruchamianie procedury segmentacji daje różne wyniki.

Jako drugi badany parametr metody atlasowej wybrano współczynnik epsilon. Parametr ten używany jest w trybie rejestracji deformacji i określa rozmiar kroku czasowego używanego do aktualizacji odkształcenia w każdej iteracji. Jego większe wartości mogą przyspieszyć rejestrację, ale mogą również spowodować, że deformacja stanie się niedyfemorficzna (nie posiada odwrotnego przekształcenia). Wartość domyślna epsilon to 1.0, a zazwyczaj wartości tego parametru mieszczą się w zakresie [0.25, 0.5]. Dla zmniejszej wartości tego parametru uruchamiano metodę atlasową na minimalnych wycinkach skanów pacjentów powstały zgodnie z prezentowanym rozwiązaniem (rysunek 4.6 E).

W przypadku eksperymentów z parametrem epsilon udało się delikatnie polepszyć wyniki segmentacji metody bazującej na atlasach. Wraz ze wzrostem parametru epsilon do wartości 5, wzrosła również dokładność wyodrębniania tętnicy z obszaru komór serca. Zaś dla

epsilon	DSC	HD [mm]	VR	MSD [mm]	PPV	TPR
0,1	0,05 ± 0,08	75,58 ± 36,85	0,77 ± 0,46	12,72 ± 8,65	0,06 ± 0,09	0,05 ± 0,08
0,5	0,08 ± 0,10	68,92 ± 38,69	0,87 ± 0,45	10,42 ± 8,16	0,09 ± 0,10	0,08 ± 0,10
1	0,09 ± 0,10	72,87 ± 42,26	0,81 ± 0,40	11,39 ± 9,79	0,11 ± 0,11	0,09 ± 0,11
2	0,10 ± 0,10	66,20 ± 37,08	0,82 ± 0,42	9,71 ± 8,03	0,11 ± 0,11	0,11 ± 0,11
5	0,11 ± 0,11	69,75 ± 43,28	0,80 ± 0,49	11,55 ± 12,59	0,11 ± 0,11	0,11 ± 0,12
10	0,10 ± 0,11	78,23 ± 52,17	0,88 ± 0,65	15,03 ± 21,26	0,10 ± 0,11	0,11 ± 0,13

Tabela 2: Średnie wartości miar jakości segmentacji LAD dla różnych rozmiarów wartości parametru epsilon metody atlasowej. Źródło: opracowanie własne.

wartości mniejszych od domyślnej segmentacja przebiegała mniej skutecznie uzyskując tym samym większe odległości pomiędzy wynikiem a wzorcową maską. Niemniej jednak otrzymane rezultaty pozwalają stwierdzić, że wartość parametru nie ma większego wpływu na skuteczność działania metody bazującej na atlasach. Poprzez eksperymentalny dobór wartości tego parametru, uzyskano wzrost dokładności Dice'a o zaledwie 3%, tym samym nie zaobserwowano zmian czułości i precyzji algorytmu. Finalnie, segmentacja jedynie metodą atlasową z wycinków skanów nie przyniosła zadowalających efektów. Dlatego też, w dalszej części pracy używano domyślnej wartości parametru epsilon.

Metoda neuronowa

W przypadku części algorytmu opartej o splotową sieć neuronową początkowo próbowano udoskonalić podejście wykorzystujące w pełni splotową sieć neuronową (FCN) [Sipiński, 2023] ze wstępnie wytrenowanym szkieletem ResNet-101 [He et al., 2016]. Modyfikowano ilość epok treningowych, współczynnik nauki oraz dodano zbiór walidacyjny do modelu, lecz te zabiegi nie wpłynęły pozytywnie na proces uczenia się sieci. Na tym etapie, dla segmentacji LAD z wycinków przekrojów uzyskano średnio wyniki miary Dice'a na poziomie kilkunastu procent. Jednakże powtarzany proces treningu nie dawał w każdej próbie zbliżonych wartości. Z reguły dla większości przypadków uczących się ta nie była w stanie odnaleźć tętnicy, generując puste maski. Wynik ten już na tym etapie udało się nieco polepszyć względem rezultatów Sipińskiego (3% miary Dice'a) poprzez przetwarzanie wycinków serca, a nie całego skanu oraz wydzielenie zbioru walidacyjnego. Mimo to pojawił się pomysł na wykorzystanie innej architektury sieci dedykowanej zadaniom segmentacji — sieci U-Net [Ronneberger et al., 2015]. Implementację tej sieci można znaleźć w bibliotece *segmentation_models_pytorch*, umożliwiającą uruchomienie z przetrenowanymi wagami na zbiorze ImageNet [Deng et al., 2009]. W tym przypadku pojawiła się trudność w dopasowaniu transformacji danych pacjentów tak, aby odpowiadały one wejściu na etapie pretreningu. W wyniku działania sieci otrzymywano maskę w postaci szumu pikseli nieskorelowanych z danymi obrazu.

Kolejną próbą było dopasowanie implementacji sieci U-Net z biblioteki MONAI [Cardoso et al., 2022], tym razem dla obrazów 3D zapisanych w formacie nii (NIfTI). Sieć ta wykorzystuje splot trójwymiarowy, co pozwala na przewidywanie lokalizacji tętnicy w kontek-

ście całego skanu pacjenta czy wycinku, a nie w ramach danego przekroju jak miało to miejsce we wcześniejszych eksperymentach. Udostępniona implementacja sieci, dedykowana segmentacji struktur biomedycznych, była testowana z zastosowaniem wstępnie wytrenowanych wag MedicalNet [Chen et al., 2019]. Stosowanie wag nielosowych wymusza korzystanie tylko z jednego kanału wejściowego sieci. W tym przypadku dostrojenie parametrów sieci poprzez naukę w całej architekturze dało wizualnie zadowalające efekty. W celu poprawy wyników dodatkowo testowano dotrenowanie dla konkretnych elementów architektury, w różnych kombinacjach. Jednakże żadna z prób nie polepszyła wyników otrzymanych poprzez modyfikację wag w całej sieci.

Sprawdzano również jaki wpływ na skuteczność nauki ma stosowanie ważonej funkcji straty. Dla wycinków serca nie stanowiło to dużej różnicy, zaś dla obrazów całego ciała pacjenta sieć bez zwiększenia istotności obszaru LAD nie dawała żadnych masek wyjściowych. Finalnie zastosowano ważoną funkcję straty, z wagą 10 dla wokseli określających strukturę.

Implementacja MONAI oferuje architekturę sieci 3D U-Net z wykorzystaniem różnych wariantów szkieletu sieci w pełni splotowej. Dlatego też, sprawdzano jaki wpływ ma zastosowana architektura enkodera na dokładność segmentacji przy pozostałych niezmienionych parametrach. Wyniki zebrano poniżej:

szkielet	DSC	HD [mm]	VR	MSD [mm]	PPV	TPR
EfficientNet-b0	$0,22 \pm 0,21$	$76,89 \pm 61,85$	$0,49 \pm 0,70$	$9,22 \pm 8,52$	$0,68 \pm 0,27$	$0,18 \pm 0,20$
ResNet-34	$0,44 \pm 0,10$	$102,27 \pm 50,11$	$1,91 \pm 1,34$	$13,45 \pm 9,48$	$0,40 \pm 0,08$	$0,51 \pm 0,17$
ResNet-50	$0,38 \pm 0,19$	$66,96 \pm 36,09$	$0,97 \pm 0,75$	$5,62 \pm 4,03$	$0,45 \pm 0,15$	$0,37 \pm 0,18$

Tabela 3: Średnie wartości miar jakości segmentacji LAD dla zbioru testowego sieci U-Net trenowanej z wykorzystaniem różnych szkieletów enkodera. Źródło: opracowanie własne.

Z powodu ograniczeń sprzętowych zbadano jedynie trzy szkielety. Przy bardziej złożonej architekturze enkodera dochodziło do przekroczenia pamięci, co uniemożliwiło przeprowadzenie nauki sieci.

Szkielet ResNet-34 enkodera dla sieci wyodrębniającej LAD z wycinków skanów najdokładniej zlokalizował strukturę, tym samym uzyskując średnią wartość współczynnika Dice'a o wartości ok. 44%. Wyniki w tym przypadku obejmują z reguły większą przestrzeń, niż zajmuje faktycznie tętnica o czym świadczą wysokie wartości miar odległości i wskaźnika czułości. Z kolei ResNet-50, korzystający z bloku wąskiego gardła, nieco gorzej wyodrębnił strukturę. Uzyskano dla tej architektury niszą wartość DSC, lecz jednocześnie odległość Hausdorffa jak i średnia odległość między strukturami wykazują na mniejsze wartości. Otrzymane maski są wielkościowo bardziej zbliżone do masek oczekiwanych, o czym również świadczy nieco wyższa wartość precyzji sieci. Natomiast, dla odmiennej architektury szkieletu EfficientNet wartość miary Dice'a spadła o połowę w odniesieniu do wartości uzyskanej przez zastosowanie ResNet-34. Jednakże to dla tego przypadku otrzymano największą dokładność metody, na poziomie ok. 68%.

Jak widać z powyżej zamieszczonych rezultatów, odpowiedni dobór szkieletu enkodera sieci ma duży wpływ na skuteczność działania metody. Napotkane trudności z uruchomieniem procesu nauki dla sieci o architekturach wymagających większej liczby parametrów niż ResNet-50 ograniczyły zakres badań. Mimo to udało się uzyskać wyniki wizualnie akceptowalne, choć od oryginalnej struktury nieco nadal odbiegają. Najdokładniej wysegmentowała LAD sieć złożona z enkodera o szkielecie ResNet-34 i dlatego też w dalszych badaniach wykorzystywano tę wartość parametru.

Podejście do agregacji wyników

Dla kroku algorytmu łączącego wyniki z poszczególnych metod początkowo wybrano klasyczne głosowanie większościowe. Wybór wartości danego woksela w masce był zgodny z tym, który najwięcej razy się powtórzył. Podejście to w praktyce dla zbioru 101 masek struktury pacjenta generowało zerowe mapy LAD. Przyczyną tego był brak zgodności większości map segmentacji dla każdego woksela maski i w konsekwencji cały obraz został zinterpretowany jako tło, tym samym nie wykrywając tętnicy. Wynik ten doprowadził do podjęcia prób modyfikacji techniki głosowania.

W kolejnym podejściu testowano różne warianty głosowania, różniące się od siebie liczbą zgodnych masek pozwalających stwierdzić istnienie LAD w danym punkcie. I tak dla głosowania, w którym co najmniej 2 mapy uznawały obecność tętnicy, otrzymywano szeroko zakrojoną strukturę pokrywającą w dużej części wzorcowe oznaczenie. Sprawdzano również inne ilości zgodnych masek, lecz dla każdego z przypadków obserwowano spadek czułości algorytmu. Chęć maksymalizacji miary czułości przyczynił się do wyboru głosowania opartego o zgodność dwóch struktur.

Wnioski

Jak to zostało opisane w eksperymentach dotyczących wstępnego przetwarzania, odpowiednie przygotowanie skanów pod metodę atlasową jest ważną czynnością. Gdy obszar LAD nie wyróżniał się znaczco na tle innych tkanek serca, atlasowe podejście działało źle lokalizując obszar poza bruzdą międzykomorową. Samo zawężanie obszaru poszukiwań dla algorytmu bazującego na atlasach może wpłynąć na skuteczność segmentowania, choć uzyskane wyniki nie stanowią reguły, z racji niedeterministycznego charakteru narzędzia Greedy. Natomiast przebadany parametr epsilon znaczco nie zmienia działania metody, dlatego też w dalszych badaniach stosowano jego domyślną wartość.

Porównując wyniki segmentacji 3D U-Net do sieci FCN, operującej na przekrojach skanów, zaobserwowano wzrost dokładności według miary Dice'a (z kilkunastu do 44%). Przyczyny tego można szukać w stosowaniu pretrenowanych wag, które to dla sieci U-Net zostały ustalone poprzez przetwarzanie trójwymiarowych obrazów medycznych, w odróżnieniu do sieci FCN, której baza danych ImageNet nie skupia się na zdjęciach diagnostycznych. Dlatego

też, wybrano architekturę U-Net jako tą wykorzystywaną w neuronowej metodzie segmentacji. Przeszkodą okazały się ograniczenia sprzętowe wymuszające korzystanie ze szkieletów enkodera sieci U-Net o relatywnie małych ilościach parametrów, w wyniku czego stosowano w dalszych badaniach szkielet ResNet-34. Sam wybór szkieletu jest istotny, dla różnych architektur uzyskano niejednakowe dokładności, co potwierdza tezę, że parametr ten wpływa na skuteczność segmentacji. W przypadku tak małej struktury jaką jest LAD, również stosowanie ważonej funkcji straty przyczyniło się do poprawy działania metody, dlatego też parametr ten był stosowany w dalszej części pracy.

Z kolei eksperymenty dotyczące agregacji wyników z poszczególnych metod, potwierdzają trudność segmentacji LAD z racji odmiенноśc rezultatów w zależności od stosowanego algorytmu. Ostatecznie wybrano podejście określające dane w wokselu na podstawie tylko dwóch masek. Decyzja ta została podjęta z racji chęci uwzględnienia jak największej liczby danych z każdej mapy oraz maksymalizacji czołości wyznaczając obszar, który w największym stopniu zawiera w sobie wzorcowe oznaczenia.

Z powyższych rozważań wynika, że pomimo przeprowadzonych eksperymentów wybrane parametry metod nie zawsze wpływają znacząco na dokładność segmentacji. Przebadane podejścia posiadają znaczną liczbę parametrów pozwalających na modyfikację działania algorytmu. Możliwe, że wśród nich istnieją takie, dla których odpowiednie dobranie wartości przyczyni się do znacznej poprawy działania metod, co wiąże się z poszerzeniem przeprowadzonych eksperymentów.

4.3 Loża po guzie

Dla problemu segmentacji loży po guzie dysponowano nieco mniejszym zbiorem danych, w którego skład wchodziły 33 zestawy skanów tomografii komputerowej. Zbiór ten, analogicznie jak w przypadku tętnicy LAD, został podzielony na podzbiory w stosunku 80 : 10 : 10, co stanowi 27 przypadków treningowych oraz po 3 przypadki testowe i walidacyjne.

Algorytm, w tym przypadku upraszcza się do zastosowania neuronowej metody segmentacji. W poszukiwaniu najdokładniejszego rozwiązania sprawdzano różne szkielety enkodera sieci przy pozostałych stałych parametrach, identycznych jak dla sieci dedykowanej segmentacji LAD. We wszystkich badaniach na wejście sieci przekazywane były trójwymiarowe obrazy wycinków lewych stron ciała pacjentek. Średnie wyniki miar jakości zamieszczono w tabeli 4.

W przypadku szkieletu ResNet-34, na zbiorze treningowym udało się otrzymać niezerowe wartości metryk tylko dla 5 pacjentów. Średnia miara DSC tych wartości wynosi ok. 28%, lecz dla pozostałych przypadków, w tym tych ze zbioru walidacyjnego i testowego, sieć nie wygenerowała żadnych masek segmentacji loży. Natomiast, dla szkieletu EfficientNet-b0 oraz EfficientNet-b2 nauka skończyła się przedwcześnie z powodu wzrastającego błędu na zbiorze walidacyjnym. W konsekwencji sieć dla wszystkich pacjentów ze zbioru nie przewidziała lokalizacji loży. Jedynie architektura sieci U-Net zbudowana ze szkieletu ResNet-50 pozwoliła na

szkielet	DSC	HD [mm]	VR	MSD [mm]
EfficientNet-b0	0	∞	0	∞
EfficientNet-b2	0	∞	0	∞
ResNet-34	0	∞	0	∞
ResNet-50	$0,61 \pm 0,06$	$59,49 \pm 30,19$	$3,66 \pm 1,67$	$3,32 \pm 1,60$

Tabela 4: Średnie wartości miar jakości segmentacji loży po guzie dla zbioru testowego sieci U-Net, trenowanej z wykorzystaniem różnych szkieletów enkoder. Źródło: opracowanie własne.

przeprowadzenie procesu nauki, w którym dla każdego przypadku treningowego jak i walidacyjnego udało się zlokalizować LAD. Dla pacjentów testowych to rozwiązanie uzyskało przyzwoitą dokładność na poziomie ok. 61% miary Dice'a.

Stosowanie sieci U-Net ze szkieletem enkodera o architekturze posiadającej ogromną liczbę parametrów taka jak np. ResNet-101 czy EfficientNet-b7, podobnie jak to miało miejsce w przypadku LAD, powodowało przekroczenie pamięci maszyny. Niemniej jednak, zastosowanie bardziej złożonej architektury szkieletu może przyczynić się do poprawy dokładności segmentacji.

Wnioski

Wnioski płynące z przeprowadzono do tej pory badania sugerują, że segmentacja loży po guzie jest zadaniem trudnym z racji niejednorodności guzów pacjentek, pod względem wielkości i umiejscowienia. Zbiór danych, według eksperta, zawiera wyłącznie przypadki zbliżone do siebie, niemniej jednak miejsca po guzie w zależności od pacjentki wyglądają inaczej. Co więcej, obszar loży nie wyróżnia się ze struktury gruczołu piersiowego — lekarze, aby oznaczyć ten obszar korzystają z markerów umieszczonych w gruczole piersiowym podczas zabiegu usuwania zmiany nowotworowej. Sieć splotowa różnicuje tkanki na podstawie intensywności wokseli, dlatego też brak widocznych granic pomiędzy strukturami stanowi dodatkowe utrudnienie. Natomiast, odpowiedni dobór szkieletu sieci niewątpliwie ma ogromne znaczenie na proces segmentacji. W przypadku architektury prostszej niż ResNet-50 nie udało się nauczyć sieci segmentowania loży.

5 Wyniki

5.1 Segmentacja LAD

Rozdział ten stanowi zestawienie wyników poszczególnych metod: bazującej na atlasach, neuronowej, atlasowo-neuronowej oraz agregacji tych trzech rozwiązań. Dokładności działania metod weryfikowane była poprzez obliczanie poszczególnych miar jakości. Z racji małych rozmiarów struktury do oceny wyodrębnienia LAD dodatkowo wykorzystano miary precyzji i czułości, z reguły nie stosowane do walidacji segmentacji. W przypadku lewej przedniej zastępującej tętnicy wieńcowej, celem jest maksymalizacja precyzji oraz osiągnięcie idealnej czułości (wartości 1). Przeprowadzono obliczenia zarówno dla skanów obejmujących całe ciało pacjenta, jak i dla wycinków obszaru z szukaną strukturą. Wszystkie metody bazowały na tym samym podziale zbioru, opisany w sekcji eksperymentów. Poniżej zebrano wyniki dla wycinków skanów.

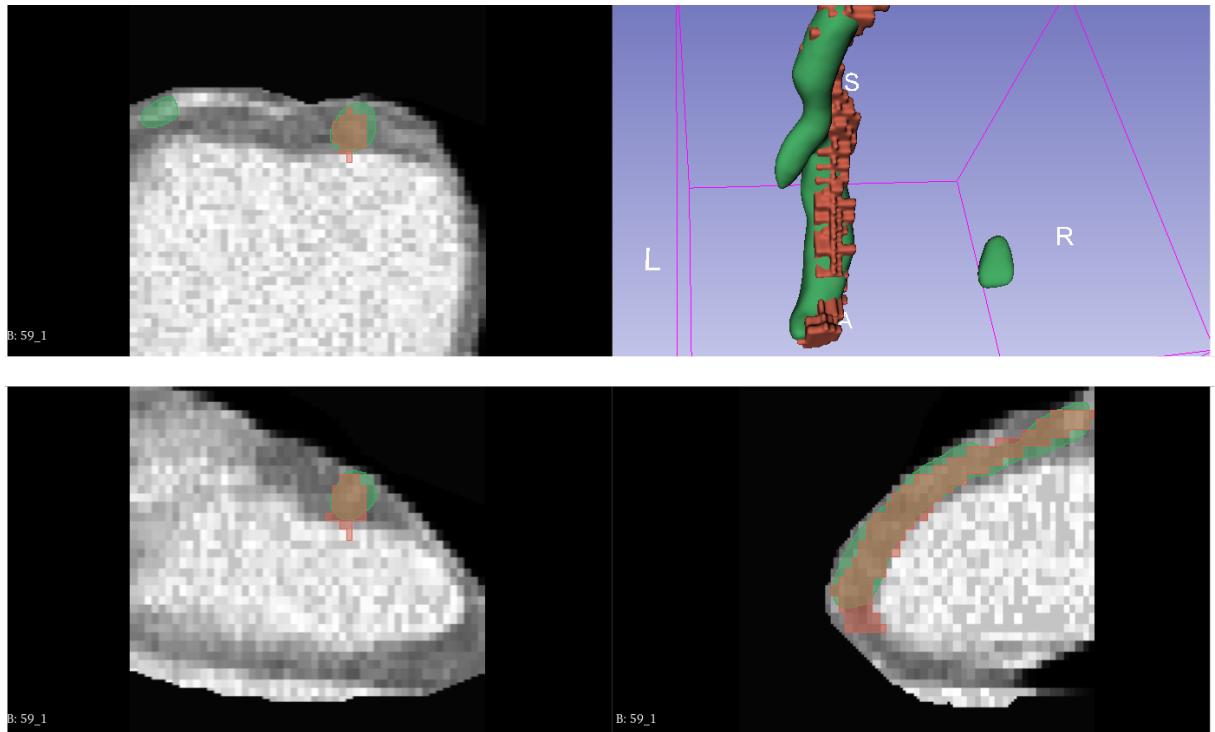
metoda	DSC	HD [mm]	VR	MSD [mm]	PPV	TPR
neuronowa	$0,44 \pm 0,10$	$102,27 \pm 50,11$	$1,91 \pm 1,34$	$13,45 \pm 9,48$	$0,40 \pm 0,08$	$0,51 \pm 0,17$
atlasowa	$0,09 \pm 0,10$	$72,87 \pm 42,26$	$0,81 \pm 0,40$	$11,39 \pm 9,79$	$0,11 \pm 0,11$	$0,09 \pm 0,11$
atlasowo-neuronowa	$0,15 \pm 0,12$	$81,70 \pm 24,95$	$1,344 \pm 1,09$	$16,51 \pm 15,34$	$0,20 \pm 0,14$	$0,12 \pm 0,11$
kombinacja powyższych	$0,17 \pm 0,05$	$107,69 \pm 11,90$	$25,90 \pm 7,50$	$19,15 \pm 7,03$	$0,10 \pm 0,03$	$0,83 \pm 0,10$

Tabela 5: Średnie wartości miar jakości segmentacji LAD z wycinków skanów tomografii komputerowej. Źródło: opracowanie własne.

Zestawiając miary jakości wszystkich czterech podejść do wycinania skanów, najdokładniej wysegmentowała LAD metoda oparta o sieć splotową. Uzyskano w tym przypadku średnią wartość miary Dice'a na poziomie ok. 44%, co niestety można uznać za słabą skuteczność. Sieć wygenerowała maski, których największe odległości pomiędzy wzorcową segmentacją osiągnęły bardzo wysoką wartość, bo aż 102 mm. Stało się tak z racji dodatkowego oznaczenia obszaru poza miejscem położenia LAD (przykład na rysunku 5.1). Z reguły wszystkie maski utworzone przez sieć obejmowały obszar lekko poza tętnicą. Średni odstęp pomiędzy strukturą przewidywaną a oczekiwana jest wysoki, co informuje o sporej rozbieżności pomiędzy nimi. Wartość miary precyzji informuje, że mniej niż połowa wokseli tętnicy została poprawnie uwzględniona w masce. Natomiast czułość modelu jest na dobrym poziomie, a zatem otrzymane maski w większości zawierały w sobie wzorcowe oznaczenia. Wizualnie, segmentacja przebiegła zadowalająco, przykładem tego jest wynik metody umieszczony na rysunku 5.1. Przewidywana maska na całej długości zawierała w sobie maskę wzorcową, jednakże dodatkowo jako tętnica potraktowany został niewielki obszar obok LAD.

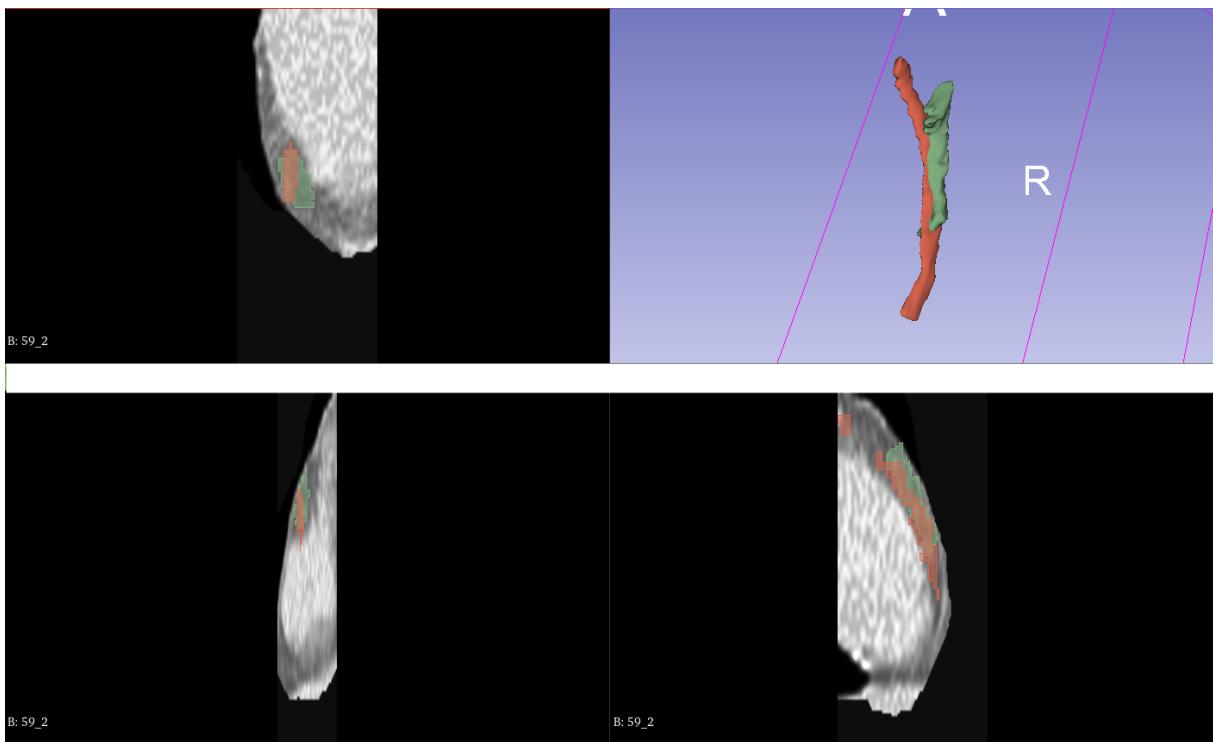
Najgorzej z wyodrębnieniem LAD poradziło sobie podejście bazujące na atlasach. Pomimo bardzo małej wartości miary DSC, na poziomie kilku procent, odległości Hausdorffa dla poszczególnych masek nie były tak duże jak w przypadku metody wykorzystującej sieć U-Net. Przyczyną tego można szukać w doborze niedokładnej transformacji atlasów w obraz badanego pacjenta. W większości przypadków, w kontekście danego pacjenta, zdeformowane atlasy zauważalnie różniły się od siebie, jak i od testowego skanu. Niekiedy przekształcenia te tworzyły

z atlasów obraz z dziurami, to znaczy z obszarami z wartościami 0. W konsekwencji dopasowanie atlasowych masek nie przebiegło skutecznie. Należy również wspomnieć, że dysponowano zbiorem 50 atlasów, z których w kontekście danego pacjenta otrzymywano bardzo różne dokładności segmentacji. Na rysunku 5.2 segmentacja atlasowa jest zdecydowanie mniejsza w stosunku do rzeczywistej struktury. Jednakże na odcinku, w którym obie maski występują, różnica nie jest rażąca.

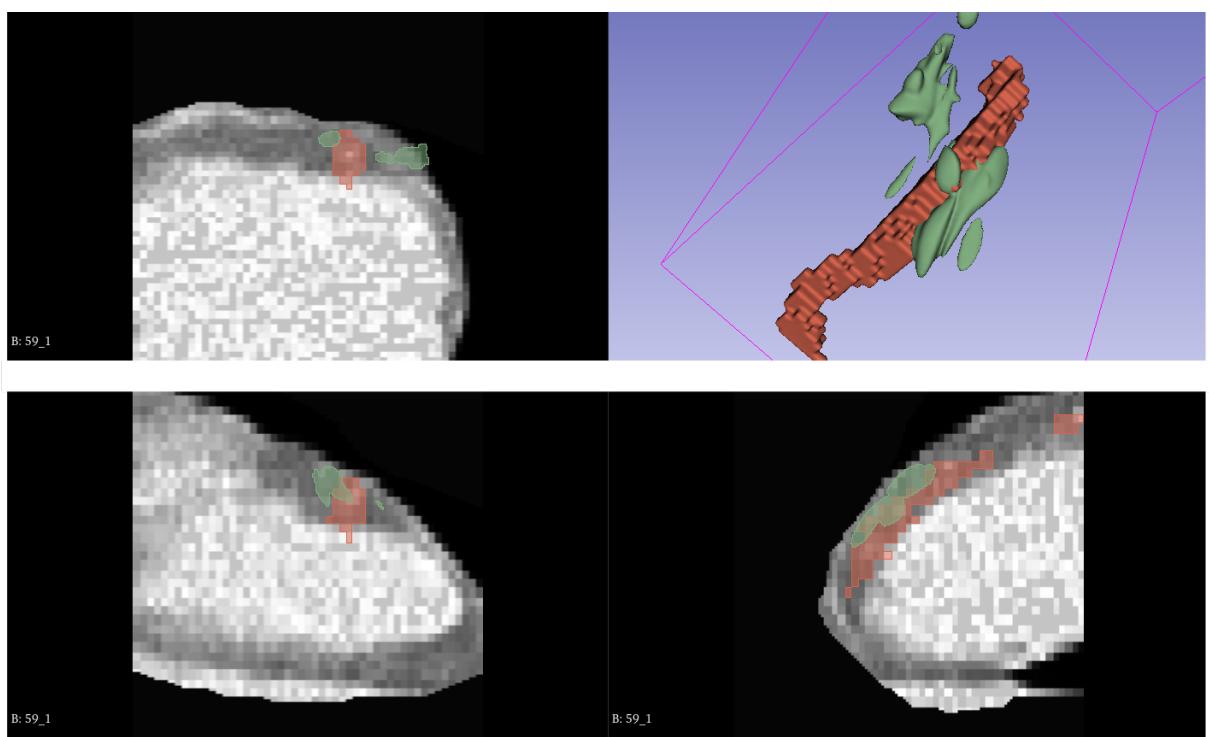


Rys. 5.1: Przykładowy wynik segmentacji LAD z wycinka skanu CT pacjenta nr 59 metodą neuronową (zielona maska przewidywana, czerwona wzorcowa). Wartość DSC wyniku: 61%. Źródło: opracowanie własne.

Podejście do agregacji algorytmu bazującego na atlasach i sieci neuronowej opierało się na przetworzeniu przez wyuczoną sieć obrazów dopasowanych atlasów do przypadków testowych. Zaobserwowano poprawę segmentacji, w porównaniu do czystego algorytmu atlasowego. Jednakże segmentacja nadal jest na poziomie kilkunastu procent dokładności według miary Dice'a. Z drugiej strony, uzyskane w ten sposób mapy LAD charakteryzują się statystycznie dużą wartością odległości Hausdorffa, większą o ok. 30 mm niż w przypadku wyników atlasowych. Metoda ta, wygenerowała maski struktury, które z reguły oddalone są od rzeczywistej lokalizacji LAD i mają bardzo mało wokseli wspólnych z oczekiwana maską. Przewidywana maska struktury przedstawiona na obrazie 5.3 stanowi zbiór malutkich fragmentów, które w większości nie są z sobą połączone. Maska ta występuje tylko na niewielkim odcinku wzorcowego oznaczenia i zawiera elementy znajdujące się w dużym oddaleniu od niego.

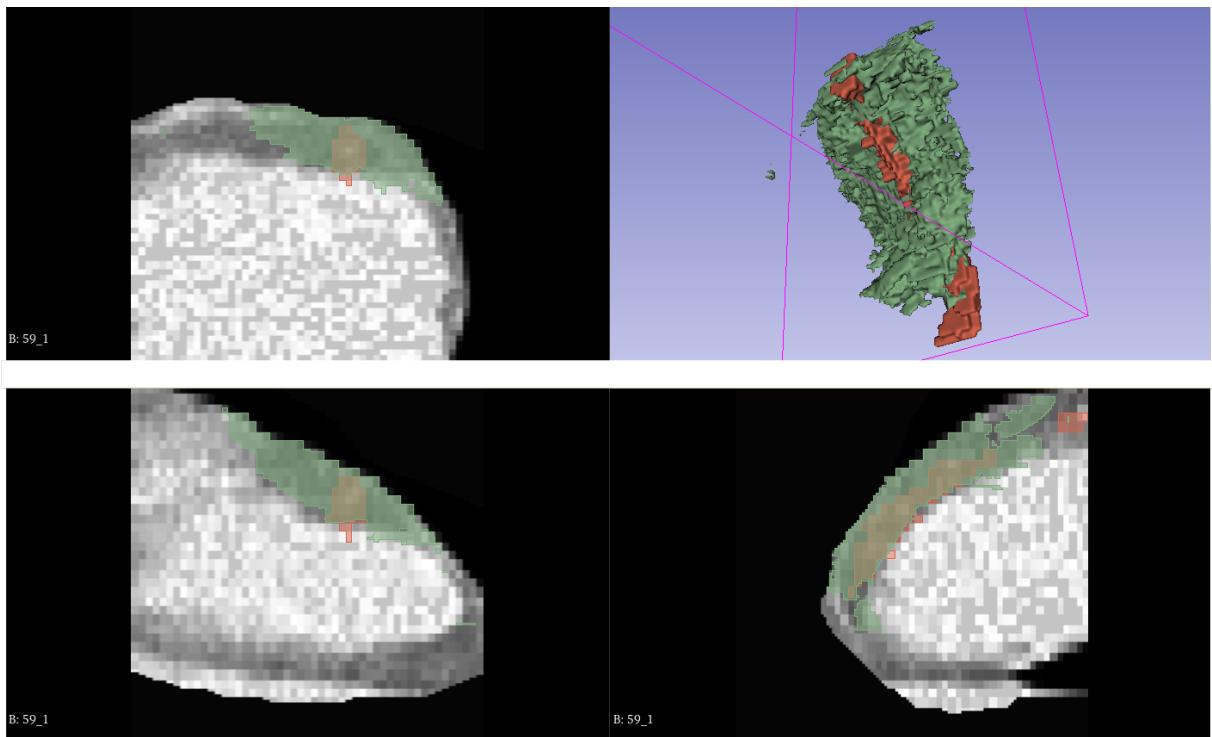


Rys. 5.2: Przykładowy wynik segmentacji LAD z wycinka skanu CT pacjenta nr 59 metodą atlasową. Atlasem w tym przypadku jest wycinek skanu pacjenta nr 4 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 32%.
Źródło: opracowanie własne.



Rys. 5.3: Przykładowy wynik segmentacji LAD z wycinka skanu CT metody atlasowo-neuronowej dla pacjenta nr 59 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 23%. Źródło: opracowanie własne.

Z kolei, dla kombinacji wyników wszystkich trzech wyżej analizowanych podejść, segmentacja według uśrednionej miary DSC była na poziomie 17%. Agregacji dla danego pacjenta poddawano 1 wynik z czystej metody neuronowej, 50 masek z metody atlasowej oraz 50 masek z metody atlasowo-neuronowej. W rezultacie otrzymano mapę pokrywającą zdecydowanie większą powierzchnię niż tę wyznaczoną przez eksperta. Niemniej jednak, wzorcowa maska LAD zawierała się w przewidywanej strukturze, o czym świadczy wysoka wartość miary czułości (83%). Celem było uzyskanie maksymalnej czułości oraz jak największej precyzji algorytmu, dlatego też osiągnięte wartości tych metryk pozwalają stwierdzić, iż segmentacja przebiegła dokładnie. Problematyczna, na przykładzie wyników pacjenta nr 59 (rysunek 5.4), okazała się być wygięta część tętnicy, której nie uwzględniała wygenerowana maska. Aczkolwiek, na pozostałe długości struktury wzorcowe oznaczenie zawierało się w tym przewidywanym.



Rys. 5.4: Przykładowy wynik segmentacji LAD z wycinka skanu CT agregacji metody atlasowej, neuronowej i ich fuzji, dla pacjenta nr 59 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 23%. Źródło: opracowanie własne.

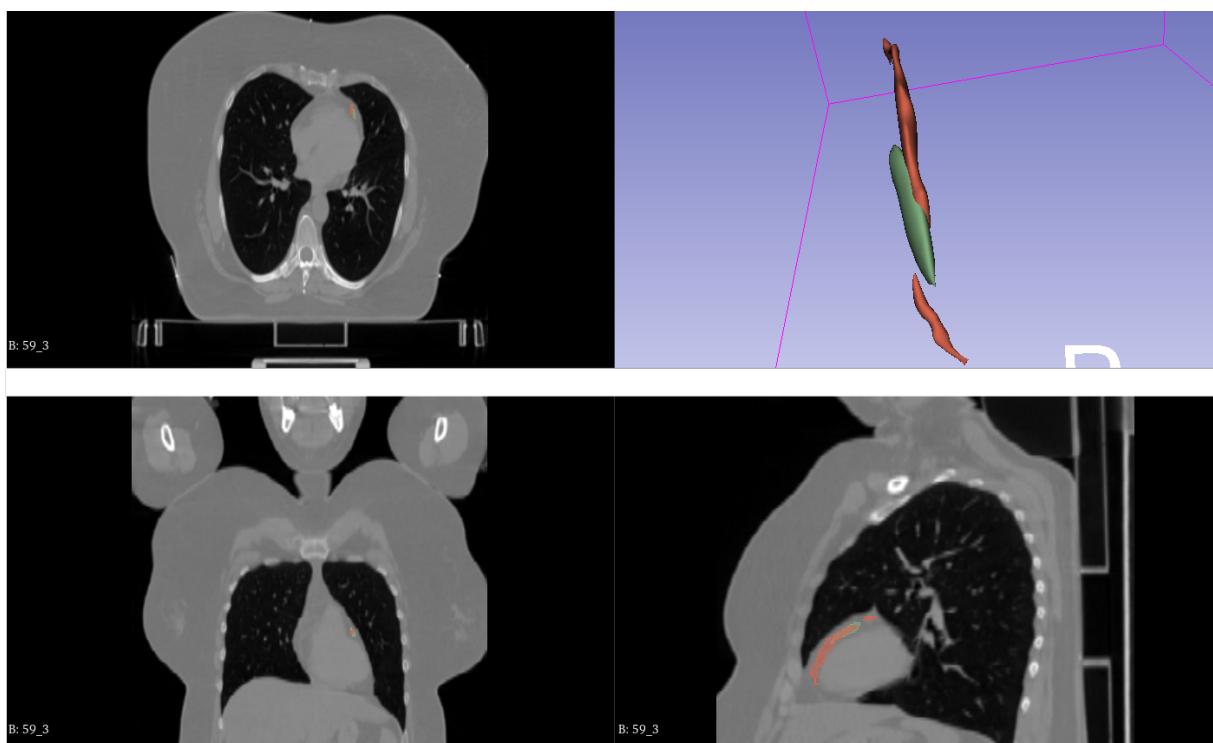
Powtarzając obliczenia tym razem dla skanów obejmujących całe ciała pacjentów, analogicznie obliczono średnie wartości miar jakości wraz z odchyleniem standardowym i umieszczone je w tabeli poniżej.

Z uwagi na niewielkie rozmiary struktury, jak się spodziewano, segmentacja ze skanów całego tułowia pacjentów przyniosła gorsze rezultaty, w porównaniu do przetwarzania wycinków. Z badanych metod, to sieć U-Net najlepiej poradziła sobie z segmentacją LAD. Pomimo niskiej dokładności według Dice'a jednocześnie uzyskano względnie małą średnią odległość Hausdorffa oraz średni odstęp pomiędzy strukturami. Oznacza to tyle, że maski są w niewielu wokselach

metoda	DSC	HD [mm]	VR	MSD [mm]	PPV	TPR
neuronowa	$0,17 \pm 0,10$	$18,33 \pm 5,81$	$1,84 \pm 2,22$	$4,93 \pm 2,90$	$0,17 \pm 0,11$	$0,26 \pm 0,29$
atlasowa	$0,04 \pm 0,07$	$30,02 \pm 13,74$	$0,93 \pm 0,51$	$8,72 \pm 5,89$	$0,04 \pm 0,08$	$0,05 \pm 0,08$
atlasowo-neuronowa	$0,001 \pm 0,002$	$35,65 \pm 12,00$	$2,46 \pm 3,66$	$17,58 \pm 9,47$	$0,001 \pm 0,001$	$0,002 \pm 0,007$
kombinacja powyższych	$0,01 \pm 0,01$	$48,79 \pm 11,48$	$18,15 \pm 10,91$	$17,83 \pm 7,34$	$0,002 \pm 0,004$	$0,03 \pm 0,06$

Tabela 6: Średnie wartości miar jakości segmentacji LAD ze skanów całego ciała tomografii komputerowej. Źródło: opracowanie własne.

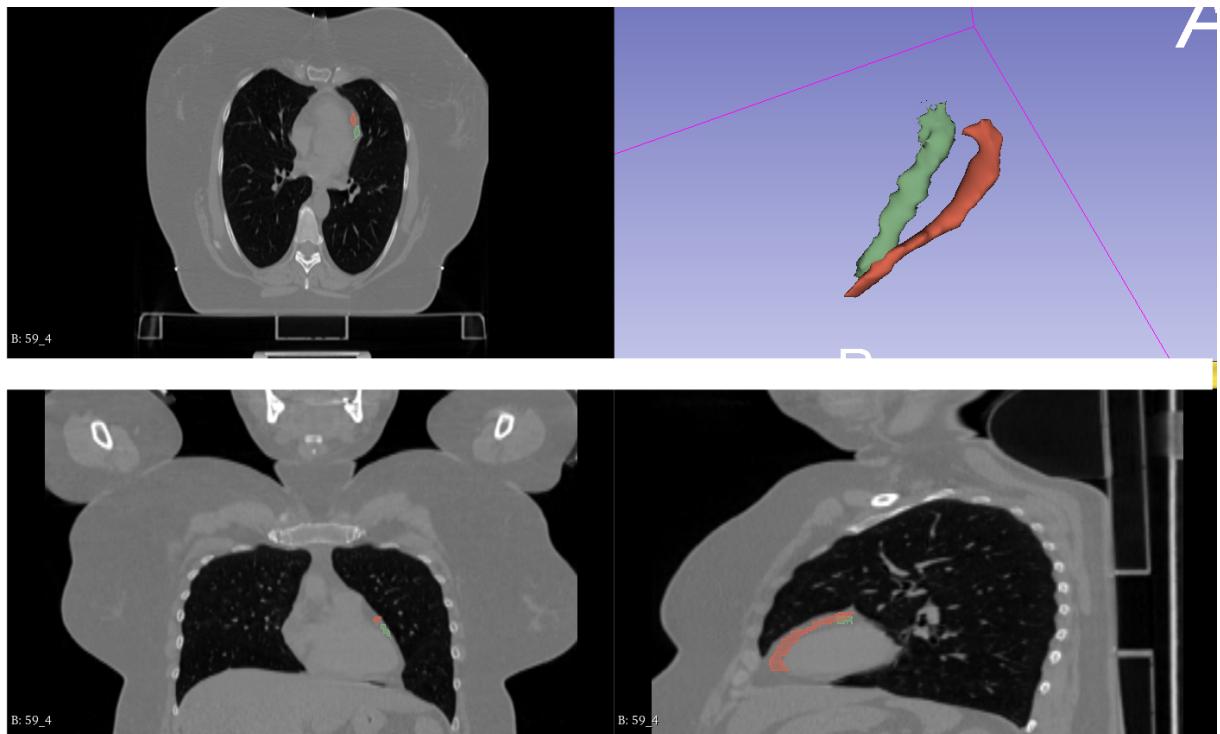
zgodne, lecz jak już tak jest to błąd dopasowania maski jest mały. Dla prezentowanego wyniku segmentacji (rysunek 5.5) przewidywana maska jest zdecydowanie krótsza w porównaniu do rzeczywistych rozmiarów tętnicy. Jednak, na odcinku wyznaczonym przez maskę z sieci, struktura zbytnio nie odbiega od wzorcowej segmentacji. Zaobserwować w tym przypadku można nieciągłość oznaczenia eksperta, co w naturze nie ma miejsca. Dlatego też, do analizy wyników należy podchodzić z pewnym dystansem, gdyż nawet lekarz może popełnić błąd oznakowania struktury.



Rys. 5.5: Przykładowy wynik segmentacji LAD z całego skanu CT metody neuronowej dla pacjenta nr 59 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 17%. Źródło: opracowanie własne.

Algorytm bazujący na atlasach dla około połowy zestawu par atlas-pacjent nie wskazał wokseli wchodzących w skład tętnicy. Zaś dla pozostałych, dokładność według DSC wahała się na poziomie kilku/kilkunastu procent. Powodem tego najprawdopodobniej jest nieprecyzyjne dopasowanie atlasu, jak i jego maski do obrazu badanego pacjenta. Transformacja atlasowa dla segmentacji tak maleńkiej struktury, z tak szerokiego obszaru poszukiwań musi być bardzo dokładna. Maski dla wybranego pacjenta przedstawione na rysunku nr 5.6, kształtem są do

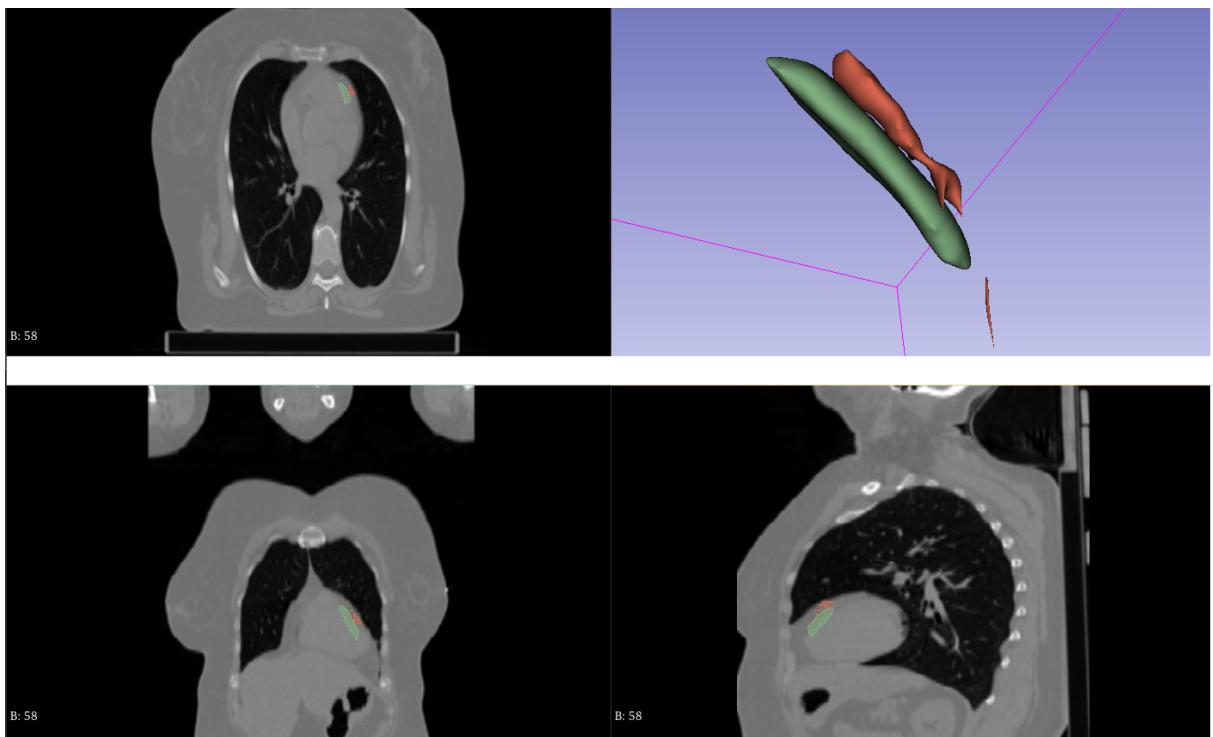
siebie bardzo podobne. Niestety, tylko dla niewielkiej ilości wokseli posiadają część wspólną.



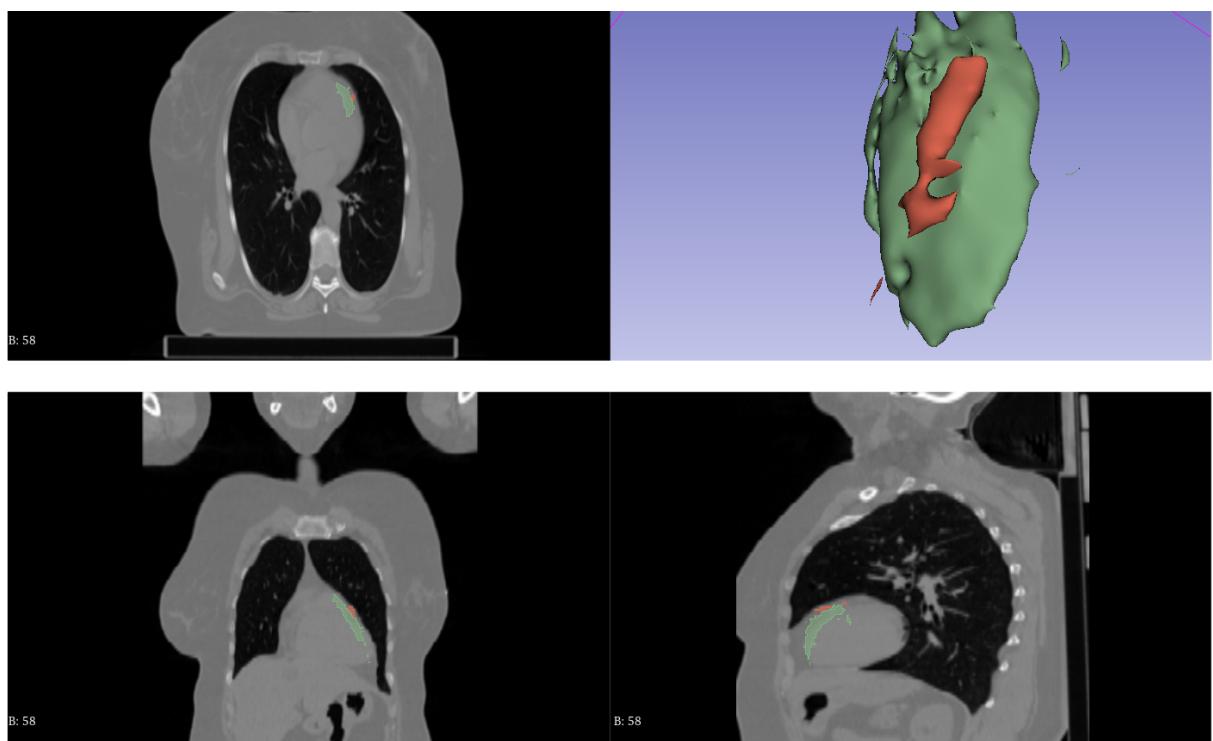
Rys. 5.6: Przykładowy wynik segmentacji LAD z całego skanu CT metody atlasowej dla pacjenta nr 59 i atlasu nr 4 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 4%. Źródło: opracowanie własne.

Z kolei, metoda atlasowo-neuronowa, tylko dla jednego przypadku testowego wygenerowała nieupustą maskę (uzyskano wartość DSC ok. 1%). Stało się tak, z racji dużej rozbieżności map składowych — nie istniały minimum dwie mapy zgodne w danym wokselu. Uzyskane wyniki miar jakości są gorsze niż w przypadku oddziennie stosowanej metody atlasowej i neuronowej. Dla tego rozwiązania uzyskano znikome wartości miar precyzji, czułości oraz miary DSC, zaś dużą wartość współczynnika objętości informującą, że przewidywana maska jest większa od tej oczekiwanej. Na rysunku 5.7 przedstawiono wynik działania tej metody, maski są do siebie podobne, jednakże przewidywana segmentacja jest znacznie przesunięta względem wzorcowego rozwiązania. Niemniej jednak, wielkości i kształt struktur są do siebie zbliżone.

Podobne wyniki do fuzji metod otrzymano dla agregacji wszystkich trzech rozwiązań, gdyż w tym przypadku wygenerowano tylko dwie niezerowe maski segmentacji (DSC 1% i 2%). Połączenie pojedynczych niedokładnych wyników z poszczególnych metod nie przyniosło zamierzonych rezultatów, o czym świadczą znikome wartości miary Dice'a, precyzji i czułości oraz duże wartości miar odległości. Sytuacja jest podobna, jak to miało miejsce powyżej, maski są względem siebie przesunięte, na co wskazują duże wartości miar odległości oraz bardzo niski współczynnik podobieństwa. W tym przypadku (rysunek 5.8), struktura wzorcowa znajduje się bliżej ściany osierdzia, zaś przewidywana maska określa obszar bardziej w głąb serca. Część wspólna obu map segmentacji stanowi bardzo małą przestrzeń.



Rys. 5.7: Przykładowy wynik segmentacji LAD z całego skanu CT metod atlasowo-neuronowej dla pacjenta nr 58 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 2%. Źródło: opracowanie własne.



Rys. 5.8: Przykładowy wynik segmentacji LAD z całego skanu CT agregacji metody atlasowej, neuronowej i ich fuzji, dla pacjenta nr 58 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 1%. Źródło: opracowanie własne.

Z powyższych interpretacji wyników można dojść do wniosku, że segmentacja LAD została wykonana skuteczniej poprzez przetwarzanie wyodrębnionych obszarów zawierających strukturę niż całości skanów CT. Najdokładniej zlokalizowała tętnicę metoda neuronowa zarówno w przypadku analizy wycinków, jak i całości skanów. Tętnica LAD jest niewielką strukturą, dlatego też obrano niestandardowe podejście do walidacji wyników, polegające na stosowaniu miary czułości i precyzji. W przypadku wycinków dla metody agregującej wyniki z poszczególnych algorytmów uzyskano wartość czułości na poziomie 83%, co można uznać za dobrze przeprowadzoną segmentację. Za cel postawiono sobie nie tyle precyzyjne znalezienie LAD, co było by najbardziej oczekiwany wynikiem, lecz obszar w którym to naczynie krwionośne się znajduje i to udało się zrobić.

Warto tutaj wspomnieć o czasie działania metody atlasowej dla skanów całego tułowia i wycinków. Dla mniejszych obrazów czas trwania obliczeń wynosił mniej więcej 20 minut, zaś aby przetworzyć skany całego ciała narzędzie Greedy [Yushkevich, 2024] potrzebowało ponad 36 godzin. Jak widać różnica jest ogromna.

Segmentacja struktury lewej przedniej zastępującej tętnicy wieńcowej nie jest zadaniem łatwym. Oznaczanie tej struktury na wynikach tomografii komputerowej przez danego lekarza bardzo zależy od jego wiedzy i doświadczenia, nie jest to sam wynik analizy obrazów. Bez znajomości budowy ciała człowieka, LAD jest praktycznie niewidoczny na skanach tułowia pacjenta tomografii komputerowej wykonywanej bez kontrastu. Dokonywując oznaczeń eksperci kierują się przeważnie anatomią, dlatego też znalezienie bruzdy między komorami, w której znajduje się tętnica dla medyka stanowi już duże ułatwienie.

5.2 Segmentacja loży

Przeprowadzone obliczenia dla segmentacji loży dotyczyły zarówno skanów obejmujących całe ciało pacjenta, jak i dla zauważonego obszaru poszukiwań. Metoda neuronowa bazowała na tym samym podziale zbioru, opisany w sekcji eksperymentów. Wyniki poszczególnych uśrednionych miar jakości prezentują się następująco:

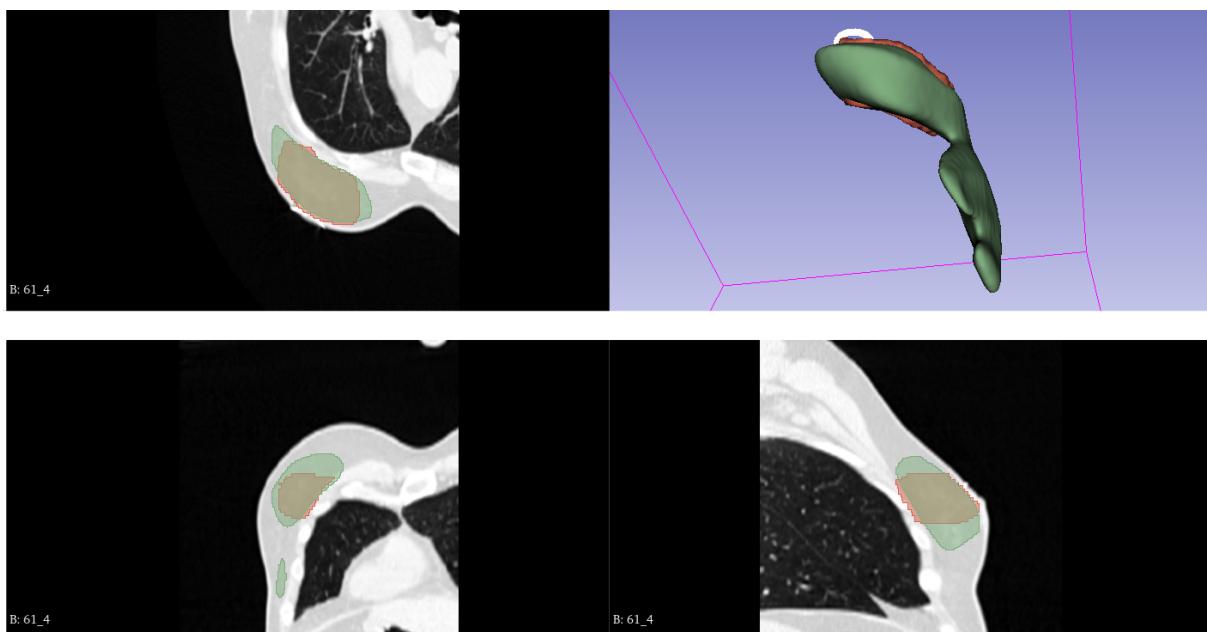
obszar	DSC	HD [mm]	VR	MSD [mm]	PPV	TPR
wycinek	$0,61 \pm 0,06$	$59,49 \pm 30,19$	$3,66 \pm 1,67$	$3,32 \pm 1,60$	$0,46 \pm 0,07$	$0,90 \pm 0,04$
całość	$0,69 \pm 0,02$	$17,21 \pm 10,37$	$0,53 \pm 0,11$	$1,21 \pm 0,42$	$0,91 \pm 0,11$	$0,56 \pm 0,02$

Tabela 7: Średnie wartości miar jakości segmentacji loży po guzie metodą neuronową z wycinków, jak i całości skanów tomografii komputerowej. Źródło: opracowanie własne.

W przypadku segmentacji z wycinka lewej strony ciała pacjentek metoda według współczynnika podobieństwa poradziła sobie z zadaniem dobrze, bo na poziomie ok. 61%. Z drugiej strony wyniki te charakteryzują duże wartości miar odległości od wzorcowej struktury. Również współczynnik objętości jest zgodny co do tego, że maski przewidywane obejmują znacznie

większą przestrzeń niż te oznaczone przez eksperta. Wartość miary precyzji potwierdza tę tezę. Natomiast uzyskano bardzo wysoką czułość — 90%, co świadczy o tym, że zdecydowana większość struktury loży zawiera się w wygenerowanej masce.

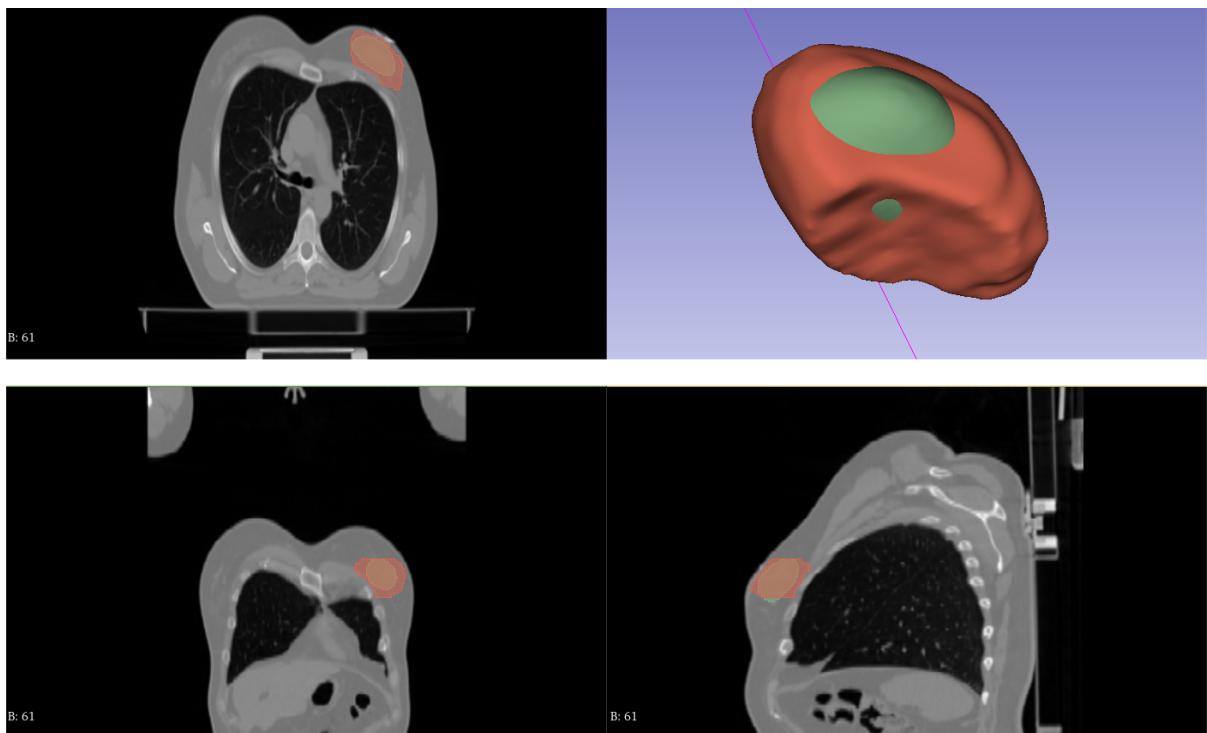
Według zamieszczonego przykładu wyniku działania metody (rysunek 5.9) oprócz miejsca loży dodatkowo sieć uznała tkanki przy żebrych, jako też wchodzące w skład szukanej struktury. Niemniej jednak, segmentacja nie wygląda źle, większa część rzeczywistej struktury zawiera się w masce wygenerowanej przez sieć U-Net. Wynik ten obrazuje wartość miary Dice'a na poziomie ok. 54%, za to miara odstępu Hausdorffa (79,9 mm) przejęła wysoką wartość, z racji wyznaczenia o wiele większej maski niż zamierzano.



Rys. 5.9: Przykładowy wynik segmentacji loży po guzie (kolor zielony) metodą neuronową wraz z maską oczekiwana (kolor czerwony) dla wycinka skanu pacjentki nr 61. Wartość miary DSC: 54%. Źródło: opracowanie własne.

Dla przykładu segmentacji loży ze skanu całego ciała pacjentki, metoda zlokalizowała loże nieco dokładniej. Podejście oparte o sieć neuronową dla tak przygotowanych obrazów z reguły wyodrębniła obszar, który znajdował się wewnętrz wzorcowej struktury. Maski te charakteryzują niskie wartości miar odległości oraz bardzo wysoka precyzja. Współczynnik podobieństwa również jest na akceptowalnym poziomie, co świadczy o dokładnie wykonanej segmentacji.

Reprezentatywny przykład wyniku dla tego podejścia (rysunek 5.10) prezentuje dokładność segmentacji według miary Dice'a o wartości 71%. W tym przypadku wygenerowana struktura prawie w całości zawarta jest wewnętrz oczekiwanej maski, jednakże istnieje niezerowa odległość pomiędzy nimi. Mimo to otrzymany wynik jest zadowalający, pozwalając uznać segmentację za pomyślnie wykonaną.



Rys. 5.10: Przykładowy wynik segmentacji loży po guzie (kolor zielony) metodą neuronową wraz z maską oczekiwana (kolor czerwony) dla całego skanu pacjentki nr 61. Źródło: opracowanie własne.

Z powyższych rezultatów wynika, że w przypadku wycinka ta sama sieć nieco gorzej poradziła sobie z segmentacją miejsca po guzie, jednak nadal osiągając akceptowalne rezultaty. Problem segmentacji loży jest z reguły trudnym zadaniem, wymagającym posiadania zbioru tomografii komputerowych pacjentów będących w podobnym stadium choroby, posiadających guzy w zbliżonych miejscach, o podobnych rozmiarach. Dlatego też, przygotowanie reprezentatywnego zestawu skanów zazwyczaj stanowi wyzwanie.

Eksperci, aby wyznaczyć obszar, w którym znajdował się guz posiadały informację przed operacją pacjentki, takie jak wielkość guza, ilość wyciętej tkanki oraz dokładna jego lokalizacja. Dane te z pewnością mogły posłużyć do ulepszenia prezentowanego rozwiązania. Lekarze, dysponując znacznikami umiejscowionymi w gruczołku piersiowym, dokonują oznaczenia struktury. Znaczniki te widoczne są na skanach jako jaśniejsze niepołączone ze sobą punkty. Sieć spłotowa różnicując natężenia wokseli zadowalała sobie z segmentacją loży, co można uznać jako sukces.

5.3 Analiza porównawcza wyników

Aby skonfrontować wyniki dla proponowanego rozwiązania segmentacji lewej zastępującej tętnicy wieńcowej LAD, zgromadzono udostępnione w literaturze rezultaty różnych podejść do tego problemu. W poniższej tabeli zawarte zostały wyniki segmentacji z przytoczony w przeglądzie literatury publikacji.

metoda	DSC	HD [mm]	VR	MSD [mm]
atlas-based + geometria [van den Bogaard et al., 2019]	0,15	-	-	-
U-Net [Chen et al., 2024]	0,57	-	0,84	1,70
nnU-Net [Chen et al., 2024]	0,65	-	1,19	1,40
multi-atlas [Chen et al., 2024]	0,23	-	0,51	7,30
nnU-Net + atlas-based + geometrii [Olloni et al., 2024]	0,09	-	-	4,20
FCN + atlas-based [Sipiński, 2023]	$0,03 \pm 0,06$	$53,43 \pm 29,66$	$0,21 \pm 0,17$	$19,55 \pm 14,73$
agregacja metod dla całości skanu	$0,01 \pm 0,01$	$48,79 \pm 11,48$	$18,15 \pm 10,91$	$17,83 \pm 7,34$
agregacja metod dla wycinka skanu	$0,17 \pm 0,05$	$202,65 \pm 79,39$	$25,90 \pm 7,50$	$19,15 \pm 7,03$

Tabela 8: Zestawienie średnich wartości miar jakości segmentacji LAD z literatury z wynikami proponowanej metody. Źródło: opracowanie własne.

Problem wyodrębnienia LAD spotyka się w literaturze z różnymi wynikami, zazwyczaj jednak odzwierciedlającą one trudność w zlokalizowaniu tej struktury. Najdokładniej wysegmentowało tętnicę rozwiązań z artykułu Chena [Chen et al., 2024], które to prezentuje zestawienie wyników dla trzech metod. Każda z nich miała dostęp do skanów zarówno tomografii wykonywanej bez, jak i z kontrastem. Kontrast podawany dożylnie uwydatnia naczynia krwionośne na obrazach CT, pozwalając tym samym na łatwiejsze zlokalizowanie LAD. Dlatego też, można wnioskować, że ta wiedza pozwoliła na uzyskanie dość dobrych rezultatów jak na tak niewielką strukturę (uzyskano wartość miary DSC na poziomie 65%). Z kolei, dla utworzenia atlasów badacze korzystali z narzędzia MACS, tworzącego zbiór przekształconych obrazów atlasów do obrazów testowych. Następnie, wykorzystywali algorytm STAPLE (głosowanie ważone) do łączenia wyników z poszczególnych atlasów. W kontrze do wartości miar jakości metod neuronowych prezentowane podejście wieloatlasowe uzyskało zdecydowanie słabszą dokładność. Najskuteczniej działała metoda oparta o sieć nnU-Net, mimo że operowała ona na przekrojach skanów, a nie na trójwymiarowym obrazie pacjenta.

Wyniki pochodzące z pozostałych publikacji nie są już tak zadowalające. Sipiński [Sipiński, 2023] korzysta z fuzji metody neuronowej i atlasowej, polegającej na wstępny wyodrębnieniu i rozjaśnieniu struktury przez sieć FCN, które kolejno przetwarzane zostały przez metodę atlasową. Agregacja atlasowych map dokonywana była poprzez głosowanie większościowe. Podejście to niestety dało znikome rezultaty (tylko 3% miary Dice'a) dla tomografii bezkontrastowej, otrzymując tym samym strukturę znacznie oddaloną od wzorcowej i mającą z nią mało części wspólnych.

Nieco lepiej, na poziomie 9% współczynnika podobieństwa, segmentacji dokonała metoda oparta na geometrii serca [Olloni et al., 2024]. Podejście to poszukiwało bruzdy międzykomorowej, przez którą następnie interpolowało prostą stanowiącą tętnicę LAD. We wcześniejszych etapach algorytmu stosowano nnU-Net do zlokalizowania serca, z którego następnie metodą atlasową lokalizowano struktury wchodzące w jego skład. Należy również wspomnieć, że badacze w tym przypadku pracowali na skanach utworzonych z wykorzystaniem kontrastu.

Podejście również wykorzystujące geometrię serca, prezentowane w artykule Bogaarda [van den Bogaard et al., 2019], prezentuje dokładność ok. 15% miary DSC. Rozwiązanie to zo-

stało przebadane na zbiorze skanów wykonywanych bezkontrastowo. Metoda polega na zlokalizowaniu komór serca poprzez metodę atlasową, które w kolejnym kroku zostały powiększone o 2 woksele. Z przecięcia się masek komór wyodrębniona została część wspólna, obszar ten następnie został zredukowany do zbioru punktów. Kolejno, wykorzystując regresję liniową, dopasowano prostą do otrzymanych wcześniej punktów, którą poddano ekstrapolacji, wydłużając ją aż do przecięcia ściany serca. Ograniczono następnie wynik do obszaru serca i poddano modyfikacji prostą, zwiększając odpowiednio jej grubość. Uzyskane w ten sposób rezultaty charakteryzują się niską wartością współczynnika podobieństwa, przeważnie znajdując LAD lekko przesunięty względem wzorcowego oznaczenia.

Prezentowane w tej pracy podejście wyróżnia się koncepcją znalezienia obszaru bruzdy między komorami. Powyżej wymienione rozwiązania za zadanie stawiają wyodrębnienie dokładnie samej struktury, co jest najbardziej oczekiwany rezultatem. Jednoznacznie prace pokazują, że metoda bazująca na atlasach gorzej radzi sobie z segmentacją w porównaniu do algorytmów opartych o sieć neuronową. Prezentują one różne podejścia do tematu, jednakże żadne z nich nie korzysta z koncepcji zawężania obszaru poszukiwań poprzez wycinanie. Co do liczbowych wyników, podejście Chena [Chen et al., 2024] wyróżniają się o wiele lepszą dokładnością, niemniej jednak rezultaty te nie bazowały wyłącznie na skanach wykonywanych bez kontrastu. Ograniczając się jedynie do miary Dice'a wyliczanej przez wszystkich badaczy, uzyskane wyniki agregacji metody atlasowej, neuronowej i atlasowo-neuronowej dla wycinków prezentują nie najgorszą dokładność na tle innych rozwiązań. Skupiając się tylko na wynikach opartych na tomografii wykonywanej bez kontrastu ([Sipiński, 2023], [van den Bogaard et al., 2019]), przedstawione w tej pracy rozwiązanie najskuteczniej poradziło sobie z zadaniem segmentacji tętnicy LAD. Niemniej jednak, wynik 17% DSC pokazuje, że nadal istnieje pole do poprawy działania metody.

metoda	DSC	HD [mm]
U-Net [Xie et al., 2023]	$0,81 \pm 0,07$	$19,25 \pm 6,01$
wycinek	$0,61 \pm 0,06$	$59,48 \pm 30,19$
całość	$0,69 \pm 0,02$	$17,21 \pm 10,37$

Tabela 9: Zestawienie średnich wartości miar jakości segmentacji loży z literatury z wynikami proponowanej metody. Źródło: opracowanie własne.

Dla zadania segmentacji loży w artykule Xie [Xie et al., 2023] porusza porównywalny problem, bo również bazujących na skanach pacjentek z lewostronnym rakiem piersi. Prezentuje on podejście wykorzystujące dodatkową wiedzę na temat ilości wyciętej tkanki, umiejscowienia oraz wielkości guza. Rozwiązanie to również bazuje na sieci 3D U-Net, która operowała na skanach przed i po operacji pacjentek, w wyniku czego uzyskano średnią wartość współczynnika DSC $0,81 \pm 0,07$ oraz maksymalną odległość Hausdorffa na poziomie $19,25 \pm 6,01$ mm.

W porównaniu do proponowanego w tej pracy rozwiązania, wyniki segmentacji z całego skanu tomografii komputerowej jak i wycinka wypadają słabiej. Uzyskano mniejszą dokładność Dice'a dla obu podejść, lecz dla przetwarzania skanów całego tułowia otrzymano mniejszą niż w artykule [Xie et al., 2023] średnią maksymalną odległość Hausdorffa. Oznacza to tyle, że maski struktury, choć mniej pokrywają się ze wzorcowym oznaczeniem, leżą bliżej tej struktury niż w przypadku referencyjnego rozwiązania.

Ogólnie rzecz ujmując, segmentacja loży po guzie jest zadaniem specyficznym, gdyż w tym przypadku nie mamy do czynienia ze strukturą anatomiczną. Próby zlokalizowania tej struktury są podejmowane wyłącznie poprzez zastosowanie sieci splotowej, a w literaturze jest to temat rzadko podejmowany. W wybranej publikacji autorzy korzystali z sieci wyuczonej od podstaw na badanym zbiorze danych, który w porównaniu do zestawu skanów wykorzystywanego w tej pracy, został wzbogacony o cenną wiedzę. W rezultacie otrzymali oni wyniki nieco lepsze, lecz bardziej odległe od wzorcowego oznaczenia.

6 Podsumowanie i perspektywy rozwoju

Głównym założeniem niniejszej pracy było zbadanie możliwości wykorzystania różnych technik do segmentacji semantycznej obrazów medycznych. W tym celu zbadano cztery podejścia: metodę atlasową, neuronową, atlasowo-neuronową oraz metodę będącą ich agregacją. Podczas pracy skupiono się na etapie preprocessingu danych, sprawdzając tym samym czy ograniczanie danych ze skanów CT do obszaru badanej struktury przyczyni się do polepszenia dokładności metod. Dokonano implementacji algorytmów oraz przeprowadzono szereg eksperymentów zarówno dla wycinków, jak i całości obrazów sprawdzając za pomocą miar jakości skuteczność działania poszczególnych metod.

W oparciu o wyniki wyodrębnienia loży po guzie metodą neuronową można stwierdzić, że prezentowane rozwiązanie dokonuje segmentacji w zadowalający sposób. W tym przypadku proces lokalizacji struktury dał dokładniejsze rezultaty dla analizy skanów całego tułowia pacjentek, w porównaniu do wyników dla wycinków skanów. Porównując się z literaturą, prezentowana metoda osiągnęła niższe wartości miar jakości, tym samym segmentując obszar mniej skutecznie.

Na podstawie zebranych wyników badań dla segmentacji tętnicy LAD, można wnioskować, że proponowane rozwiązanie pozwala na wyodrębnienie struktury poprawnie, w rezultacie otrzymując wyznaczony przez maskę obszar bruzdy międzymorowej. W porównaniu do metod z literatury bazujących jedynie na wynikach tomografii bezkontrastowej, otrzymano lepsze wyniki miar jakości i tym samym dokładniejszą segmentację.

Zbadane algorytmy we wszystkich przypadkach eksperymentalnych uzyskały wyniki, według miar jakości, dalekie od idealnego dopasowania do maski oczekiwanej. Dlatego też, w oparciu o wnioski płynące z przeprowadzonych badań można zaproponować następujące obszary do dalszych eksperymentów.

Zarówno dla analizy wycinków, jak i całości skanów, metoda bazująca na atlasach pod względem dokładności wypadła najgorzej. Przyczyna tego najprawdopodobniej leży w niedokładnej transformacji używanej do deformacji obrazów atlasów oraz ich masek. Narzędzie Gready oferuje szereg parametrów, z których wyróżnia się nietestowany współczynnik wygładzenia. Parametr ten stosowany jest na etapie rejestracji deformacji, którego wartość definiuje, w jakim stopniu pozwalamy na zginanie i rozciąganie skanu atlasu. Odpowiednia zmiana wartości domyślnej niesie nadzieję na ulepszenie skuteczności działania metody.

Z kolei, dla podejścia bazującego na sieci neuronowej, dla wycinków oraz całości skanów, posiadanie większego zbioru danych przyniosłoby lepsze efekty treningu. Rozważyć należy również dodanie augmentacji danych z posiadanego zbioru. Dodatkowo przeprowadzenie eksperymentów segmentacji zarówno dla loży, jak i LAD na bardziej rozbudowanych architekturach enkodera np. ResNet101 może przyczynić się do polepszenia rezultatów.

Co do wstępniego przetwarzania skanów dla segmentacji LAD wartym sprawdzenia byłby

sposób wycinania bruzdy między komorowej poprzez odpowiednie odjęcie z obszaru osierdzia, odczytanego z maski zapisanej w plikach DICOM, przestrzeni wyznaczonej przez maski komór TotalSegmentatora. W takim przypadku teoretycznie powinno otrzymać się obszar tkanek blisko przylegających do komór, w tym miejsce położenia LAD. Pomyślem dającym podobny efekt jest wzbogacenie proponowanego procesu wyodrębnienia obszaru tętnicy poprzez maskowanie wokseli wchodzących w skład komór serca. Efekty z obu tych podejść powinny być podobne.

Jak widać, zaprezentowane metody oferują szerokie pole modyfikacji, które w skutkach powinny objawić się jako jeszcze dokładniejsze wyodrębnienie struktury ze skanów. Tak więc niewykluczone, iż tematyka tejże pracy w przyszłości będzie dalej rozwijana.

Literatura

- [dic, 2024] (2024). Digital Imaging and Communications in Medicine (DICOM). URL: <https://www.dicomstandard.org> (data dostępu: 2024-08-31).
- [Cardoso et al., 2022] Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., Nath, V., He, Y., Xu, Z., Hatamizadeh, A., Myronenko, A., Zhu, W., Liu, Y., Zheng, M., Tang, Y., Yang, I., Zephyr, M., Hashemian, B., Alle, S., Darestani, M. Z., Budd, C., Modat, M., Vercauteren, T., Wang, G., Li, Y., Hu, Y., Fu, Y., Gorman, B., Johnson, H., Genereaux, B., Erdal, B. S., Gupta, V., Diaz-Pinto, A., Dourson, A., Maier-Hein, L., Jaeger, P. F., Baumgartner, M., Kalpathy-Cramer, J., Flores, M., Kirby, J., Cooper, L. A. D., Roth, H. R., Xu, D., Bericat, D., Floca, R., Zhou, S. K., Shuaib, H., Farahani, K., Maier-Hein, K. H., Aylward, S., Dogra, P., Ourselin, S., and Feng, A. (2022). MONAI: An open-source framework for deep learning in healthcare. doi:10.48550/arXiv.2211.02701.
- [Chen et al., 2019] Chen, S., Ma, K., and Zheng, C. (2019). Med3D: Transfer learning for 3D medical image analysis. *ArXiv*, abs/1904.00625. doi:10.48550/arXiv.1904.00625 .
- [Chen et al., 2024] Chen, X., Mumme, R. P., Corrigan, K. L., Mukai-Sasaki, Y., Koutroumpakis, E., Palaskas, N. L., Nguyen, C. M., Zhao, Y., Huang, K., Yu, C., Xu, T., Daniel, A., Balter, P. A., Zhang, X., Niedzielski, J. S., Shete, S. S., Deswal, A., Court, L. E., Liao, Z., and Yang, J. (2024). Deep learning-based automatic segmentation of cardiac substructures for lung cancers. *Radiotherapy and Oncology*, 191:110061. doi:10.1016/j.radonc.2023.110061.
- [Ciardo et al., 2017] Ciardo, D., Gerardi, M. A., Vigorito, S., Morra, A., Dell'acqua, V., Diaz, F. J., Cattani, F., Zaffino, P., Ricotti, R., Spadea, M. F., Riboldi, M., Orecchia, R., Baroni, G., Leonardi, M. C., and Jereczek-Fossa, B. A. (2017). Atlas-based segmentation in breast cancer radiotherapy: Evaluation of specific and generic-purpose atlases. *The Breast*, 32:44–52. doi:10.1016/j.breast.2016.12.010.
- [Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. ieee. doi:10.1109/CVPR.2009.5206848.
- [Fedorov et al., 2012] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J., Pieper, S., and Kikinis, R. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–41. doi:10.1016/j.mri.2012.05.001 URL: <https://www.slicer.org> (data dostępu: 2024-05-31).

[Finnegan et al., 2023] Finnegan, R., Chin, V., and Chlap, P. (2023). Open-source, fully-automated hybrid cardiac substructure segmentation: development and optimisation. *Phys Eng Sci Med*, 377-393(46). doi:10.1007/s13246-023-01231-w.

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. doi:10.1109/CVPR.2016.90.

[Hinton et al., 2012] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. doi:10.48550/arXiv.1207.0580.

[Ioffe and Szegedy, 2015] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 448–456. PMLR. doi:10.48550/arXiv.1502.03167 .

[Kaba et al., 2023] Kaba, S., Haci, H., Isin, A., Ilhan, A., and Conkbayir, C. (2023). The application of deep learning for the segmentation and classification of coronary arteries. *Diagnostics*, 13(13). doi:10.3390/diagnostics13132274.

[Liu et al., 2021] Liu, Z., Liu, F., Chen, W., Tao, Y., Liu, X., Zhang, F., Shen, J., Guan, H., Zhen, H., Wang, S., Chen, Q., Chen, Y., and Hou, X. (2021). Automatic segmentation of clinical target volume and organs-at-risk for breast conservative radiotherapy using a convolutional neural network. *Cancer management and research*, (13):8209–8217. doi:10.2147/CMAR.S330249.

[Olloni et al., 2024] Olloni, A., Lorenzen, E. L., Jeppesen, S. S., Diederichsen, A., Finnegan, R., Hoffmann, L., Kristiansen, C., Knap, M., Milo, M. L. H., Møller, D. S., Pøhl, M., Persson, G., Sand, H. M., Sarup, N., Thing, R. S., Brink, C., and Schytte, T. (2024). An open source auto-segmentation algorithm for delineating heart and substructures – development and validation within a multicenter lung cancer cohort. *Radiotherapy and Oncology*, 191:110065. doi:10.1016/j.radonc.2023.11006.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing. doi:10.1007/978-3-319-24574-4_28.

[Sipiński, 2023] Sipiński, M. (2023). Wsparcie planowania radioterapii z wykorzystaniem sieci neuronowych (praca magisterska z Politechniki Łódzkiej).

[Su et al., 2019] Su, Y., Jihoon, K., Jae-Hyung, R., Jae-Hwan, L., Heejun, K., Lae-Jeong, P., Dong, J. K., Hyeonkyeong, Y., Jaehee, H., Do-Yoon, K., Pil, H. L., Jung-Min, A., Soo-Jin, K., Duk-Woo, P., Seung-Whan, L., Young-Hak, K., Cheol Whan, L., Seong-Wook, P., and Seung-Jung, P. (2019). Deep learning segmentation of major vessels in x-ray coronary angiography. *Scientific Reports*, 9(16897 (2019)). doi:10.1038/s41598-019-53254-7.

[Tan and Le, 2019] Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR. URL: <https://proceedings.mlr.press/v97/tan19a.html> (data dostępu: 2024-05-31).

[van den Bogaard et al., 2019] van den Bogaard, V. A., van Dijk, L. V., Vliegenthart, R., Sijtsema, N. M., Langendijk, J. A., Maduro, J. H., and Crijns, A. P. (2019). Development and evaluation of an auto-segmentation tool for the left anterior descending coronary artery of breast cancer patients based on anatomical landmarks. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*, 136:15–20. doi:10.1016/j.radonc.2019.03.013.

[Wasserthal, 2024a] Wasserthal, J. (2024a). TotalSegmentator - 3D slicer extension for fully automatic whole body CT segmentation using "Totalsegmentator" AI model. URL: <https://github.com/lassoan/SlicerTotalSegmentator> (data dostępu: 2024-05-31).

[Wasserthal, 2024b] Wasserthal, J. (2024b). TotalSegmentator API GitHub: wasserth/Total-Segmentator. URL: <https://github.com/wasserth/TotalSegmentator> (data dostępu: 2024-05-31).

[Wasserthal et al., 2023] Wasserthal, J., Breit, H. C., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., Heye, T., Boll, D. T., Cyriac, J., Yang, S., Bach, M., and Segeroth, M. (2023). Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5):e230024. doi:10.1148/ryai.230024.

[Xie et al., 2023] Xie, X., Song, Y., and Ye, F. (2023). Prior information guided auto-segmentation of clinical target volume of tumor bed in postoperative breast cancer radiotherapy. *Radiat Oncol*, 170(18). doi:10.1186/s13014-023-02355-9.

[Yushkevich, 2024] Yushkevich, P. (2024). Greedy GitHub - pyushkevich/greedy: Very fast greedy diffeomorphic registration code. URL: <https://github.com/pyushkevich/greedy> (data dostępu: 2024-05-31).

[Yushkevich et al., 2016] Yushkevich, P., Pluta, J., Wang, H., Wisse, L., Das, S., and Wolk, D. (2016). Fast automatic segmentation of hippocampal subfields and medial temporal

lobe subregions in 3 Tesla and 7 Tesla MRI. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 12(7):126–127. doi:10.1016/j.jalz.2016.06.205.

[Zhou et al., 2017] Zhou, R., Liao, Z., Pan, T., Milgrom, S. A., Pinnix, C. C., Shi, A., Tang, L., Yang, J., Liu, Y., Gomez, D., Nguyen, Q.-N., Dabaja, B. S., Court, L., and Yang, J. (2017). Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiotherapy and Oncology*, 122(1):66–71. doi:10.1016/j.radonc.2016.11.016.

Spis rysunków

2.1	Para skanów atlasu i testowego pacjenta, używana do stworzenia poniższych transformacji. Źródło: opracowanie własne.	9
2.2	Transformacja afinicza zobrazowana na pojedynczym przekroju wybranego atlasu i obrazu pacjenta. Źródło: opracowanie własne.	9
2.3	Transformacja elastyczna zobrazowana na pojedynczym przekroju wybranego atlasu i obrazu pacjenta. Źródło: opracowanie własne.	10
2.4	Schemat sztucznego neuronu McCullocha-Pittsa. Źródło: opracowanie własne.	12
2.5	Schemat sieci wielowarstwowego perceptronu. Źródło: opracowanie własne.	13
2.6	Poglądowy schemat sieci splotowej. Źródło: opracowanie własne.	15
2.7	Uproszczona wizualizacja zasady działania operacji splotu. Źródło: opracowanie własne.	16
2.8	Dwie odmiany splotowych bloków rezydualnych: po lewej blok podstawowy, po prawej blok wąskiego gardła. Źródło: opracowanie własne.	18
2.9	Rysunek ogólny architektury sieci U-Net. Źródło: opracowanie własne.	21
3.1	Schemat modułu segmentacji tętnicy LAD. Źródło: opracowanie własne.	28
3.2	Schemat modułu segmentacji loży po guzie. Źródło: opracowanie własne.	29
3.3	Ilustracja przedstawiająca poszczególne kroki procesu kadrowania przekrojów całego ciała pacjenta ze skanów CT. Źródło: opracowanie własne.	30
3.4	Schemat wycinania obszaru LAD na przykładzie jednego z przekrojów skanu CT z zaznaczonymi maskami komór i przedsięwiorów, wygenerowanych przez narzędzie TotalSegmentator. Źródło: opracowanie własne.	32
3.5	Schemat wycinania obszaru loży po guzie na przykładzie jednego z przekrojów skanu CT z zaznaczonymi maskami mostka i chrząstki żebrowej. Źródło: opracowanie własne.	33
3.6	Przykładowy wycinek skanu CT obszaru serca po procesie wstępniego przetwarzania. Źródło: opracowanie własne.	34
3.7	Przykładowy wycinek skanu CT obszaru lewego gruczołu piersiowego po procesie wstępniego przetwarzania. Źródło: opracowanie własne.	35
4.1	Przykładowy skan tomografii komputerowej ze zbioru danych pacjentki z nowotworem piersi po operacji mastektomii. Źródło: opracowanie własne.	40
4.2	Przykładowy skan tomografii komputerowej ze zbioru danych pacjentki z nowotworem piersi po operacji oszczędzającej gruczoł piersiowy. Źródło: opracowanie własne.	40
4.3	Przekrój wycinka serca bez obróbki.	42
4.4	Przekrój wycinka serca z maskowaniem żeber.	42
4.5	Przekrój wycinka serca ze zmodyfikowaną jasnością i zamaskowanymi żebrami.	42

4.6	Obrazy przekroju skanu pacjenta w danym formatowaniu: A - całe ciało; B - wycinek obszaru komór z 30 wokselowym marginesem; C - wycinek obszaru komór z 10 wokselowym marginesem; D - wycinek proponowanego rozwiązania bez maskowania; E - wycinek proponowanego rozwiązania z maskowaniem wraz ze wzorcową maską LAD. Źródło: opracowanie własne.	43
5.1	Przykładowy wynik segmentacji LAD z wycinka skanu CT pacjenta nr 59 metodą neuronową (zielona maska przewidywana, czerwona wzorcowa). Wartość DSC wyniku: 61%. Źródło: opracowanie własne.	51
5.2	Przykładowy wynik segmentacji LAD z wycinka skanu CT pacjenta nr 59 metodą atlasową. Atlasem w tym przypadku jest wycinek skanu pacjenta nr 4 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 32%. Źródło: opracowanie własne.	52
5.3	Przykładowy wynik segmentacji LAD z wycinka skanu CT metody atlasowo-neuronowej dla pacjenta nr 59 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 23%. Źródło: opracowanie własne.	52
5.4	Przykładowy wynik segmentacji LAD z wycinka skanu CT agregacji metody atlasowej, neuronowej i ich fuzji, dla pacjenta nr 59 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 23%. Źródło: opracowanie własne.	53
5.5	Przykładowy wynik segmentacji LAD z całego skanu CT metody neuronowej dla pacjenta nr 59 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 17%. Źródło: opracowanie własne.	54
5.6	Przykładowy wynik segmentacji LAD z całego skanu CT metody atlasowej dla pacjenta nr 59 i atlasu nr 4 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 4%. Źródło: opracowanie własne. .	55
5.7	Przykładowy wynik segmentacji LAD z całego skanu CT metod atlasowo-neuronowej dla pacjenta nr 58 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 2%. Źródło: opracowanie własne.	56
5.8	Przykładowy wynik segmentacji LAD z całego skanu CT agregacji metody atlasowej, neuronowej i ich fuzji, dla pacjenta nr 58 (zielona maska przewidywana, czerwona wzorcowa). Uzyskano w tym przypadku DSC o wartości 1%. Źródło: opracowanie własne.	56
5.9	Przykładowy wynik segmentacji loży po guzie (kolor zielony) metodą neuronową wraz z maską oczekiwana (kolor czerwony) dla wycinka skanu pacjentki nr 61. Wartość miary DSC: 54%. Źródło: opracowanie własne.	58

5.10 Przykładowy wynik segmentacji loży po guzie (kolor zielony) metodą neuro-nową wraz z maską oczekiwana (kolor czerwony) dla całego skanu pacjentki nr 61. Źródło: opracowanie własne. 59

Spis tabel

1	Średnie wartości miar jakości segmentacji LAD dla różnych zakresów obszaru poszukiwań metody atlasowej. Źródło: opracowanie własne.	44
2	Średnie wartości miar jakości segmentacji LAD dla różnych rozmiarów wartości parametru epsilon metody atlasowej. Źródło: opracowanie własne.	45
3	Średnie wartości miar jakości segmentacji LAD dla zbioru testowego sieci U-Net trenowanej z wykorzystaniem różnych szkieletów enkodera. Źródło: opracowanie własne.	46
4	Średnie wartości miar jakości segmentacji loży po guzie dla zbioru testowego sieci U-Net, trenowanej z wykorzystaniem różnych szkieletów enkoder. Źródło: opracowanie własne.	49
5	Średnie wartości miar jakości segmentacji LAD z wycinków skanów tomografii komputerowej. Źródło: opracowanie własne.	50
6	Średnie wartości miar jakości segmentacji LAD ze skanów całego ciała tomografii komputerowej. Źródło: opracowanie własne.	54
7	Średnie wartości miar jakości segmentacji loży po guzie metodą neuronową z wycinków, jak i całości skanów tomografii komputerowej. Źródło: opracowanie własne.	57
8	Zestawienie średnich wartości miar jakości segmentacji LAD z literatury z wynikami proponowanej metody. Źródło: opracowanie własne.	60
9	Zestawienie średnich wartości miar jakości segmentacji loży z literatury z wynikami proponowanej metody. Źródło: opracowanie własne.	61