UNIWERSYTET JAGIELLOŃSKI
W KRAKOWIE

Uniwersytet Jagielloński w Krakowie

Wydział Biochemii, Biofizyki i Biotechnologii

Uczenie maszynowe w biotechnologii i medycynie na przykładzie detekcji anomalii piersi w obrazowaniu mammograficznym

Michał Kowalski

Praca magisterska na kierunku *Biotechnologia Molekularna*

wykonana pod opieką prof. dr hab. Marty Pasenkiewicz-Gieruli

w Zakładzie Biofizyki Obliczeniowej i Bioinformatyki

Praca została wykonana z wykorzystaniem Infrastruktury PLGrid

Grant *neuralcbisddsm*

Kraków 2019

Spis treści

Wykaz skrótów i pojęć obcych

CBIS-DDSM – Curated Breast Imaging Subset of DDSM

DDSM – Digital Database for Screening Mammography

Feature extractor – Część sieci neuronowej odpowiadająca za obliczanie cech kluczowych obrazu

ROI – Region of Interest (interesujący rejon obrazu

YOLO – algorytm You Only Look Once

Streszczenie

Streszczenie w języku polskim

Użyteczność systemów opierających się na tzw. „sztucznej inteligencji” staje się bezdyskusyjna w coraz większej liczbie dziedzin życia i nauki. Nauki o życiu takie jak biologia, biotechnologia, czy medycyna z powodzeniem wykorzystują podstawowe technologie uczenia maszynowego względem problemów o niskiej złożoności, jednak stosowanie rozwiązań wykorzystujących „głębokie nauczanie maszynowe” nie cieszy się popularnością. Spowodowane jest to głównie niezadawalającą czystością danych dotyczących problemów o wysokiej złożoności oraz brakiem dyscypliny naukowej łączącej danologię z naukami o życiu. Na przykładzie detekcji anomalii na obrazach mammograficznych ze zbioru *Digital Database for Screening Mammography* przedstawione zostało potencjalne wykorzystanie metod „głębokiego uczenia maszynowego” w ujęciu biologiczno – medycznym oraz problemy dotyczące standardów jakości i organizacji danych. W tym celu wykorzystano algorytm *Tiny-YOLO* służący do detekcji obiektów. Wykazano, że praktyczne jego zastosowanie jest jednak niemożliwe do czasu ustalenia standardów zbierania i katalogowania danych oraz ich przetwarzania, o czym świadczy bardzo niska liczba poprawnych detekcji miejsc zmienionych chorobowo, zupełny brak rozróżniania typów anomalii oraz ogromne różnice pomiędzy właściwościami obrazów mammograficznych w zbiorach danych.

Słowa kluczowe: uczenie maszynowe, mammografia, detekcja obiektów, czystość danych, detekcja anomalii.

Streszczenie w języku angielskim

Overall usefullness of systems based on “Artificial Intelligence” technology has become indisputable in many areas of life and science. Life-Sciences such as biology, biotechnology or medicine has been successfully using basic machine learning techniques to solve problems with low complexity, nevertheless solutions based on deep learning are not quite as popular. The main cause behind that state is that problems with high complexity have unsatisfying cleanliness of data, also there is no academic discipline that could link knowledge from data science with knowledge from life sciences. By the example of detection of abnormalities on mammograms from *Digital Database for Screening Mammography* dataset, potential usage of deep learning methods in biological and medical concepts has been portrayed as well as problems concerning standards of data quality and orderliness. For this task an algorithm called *Tiny-YOLO* was used. It has been demonstrated that practical use of this algorithm remains impossible before determining the standards of collecting, cataloging and preprocessing of data as witnessed by unsignificant amount of correct detections of anomalies, absence of differentiability between types of anomalies and also huge differences between properties of mammogram images in the datasets.

Key words: machine learning, mammography, object detection, data cleanliness, anomaly detection.

Wstęp

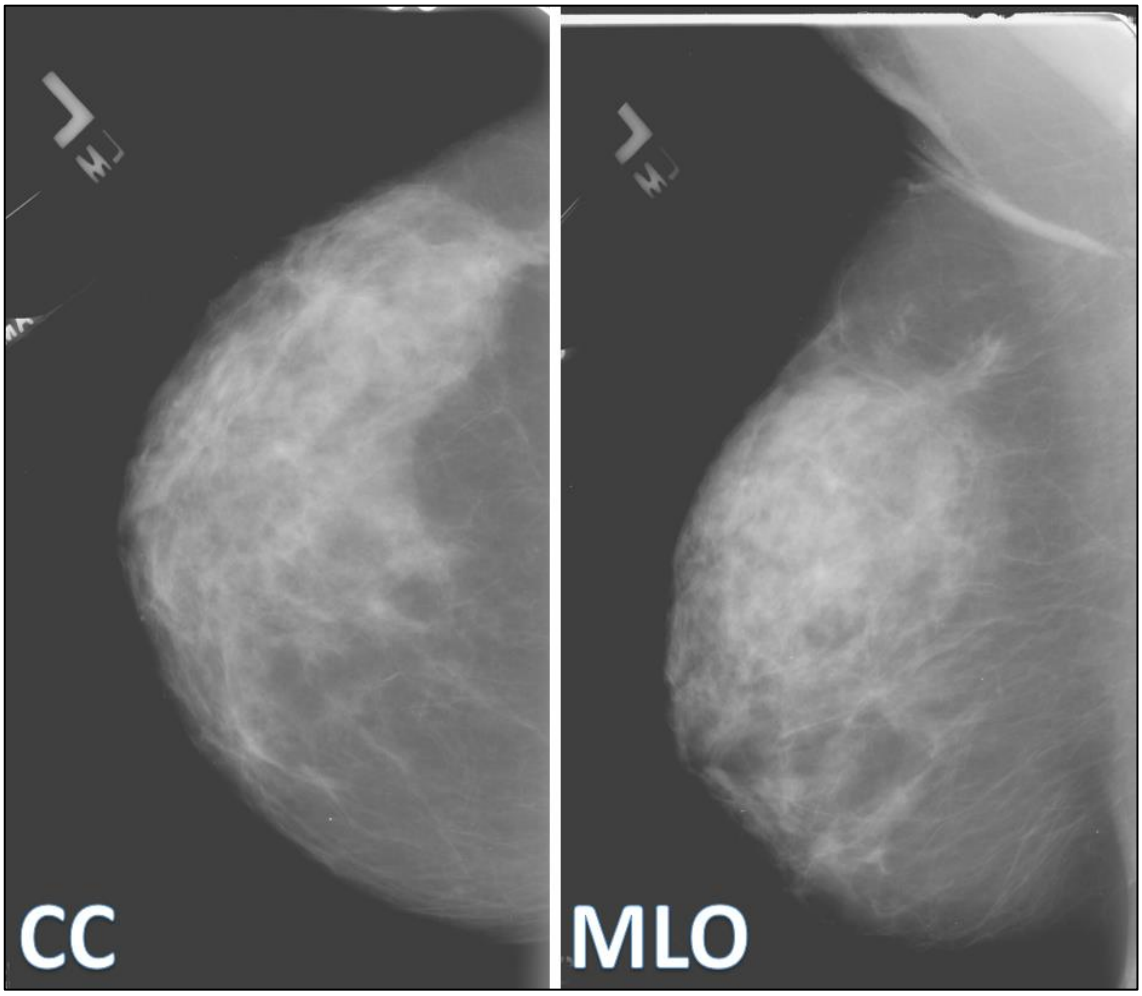
Dane

Uczenie maszynowe jest coraz powszechniej wykorzystywane w biologii i medycynie, zwłaszcza przez prywatne przedsiębiorstwa i koncerny farmaceutyczne. Istnieje bardzo dużo zbiorów danych dotyczących różnych zagadnień biologicznych, niemniej jednak operują one głównie na tzw. „niskim poziomie abstrakcji" - dane dotyczą ściśle sprecyzowanych zagadnień dotyczących jednego, dokładnie określonego problemu badawczego - i są wysoce wyspecjalizowane. Obecnie rozwój „głębokiego uczenia maszynowego" (ang. *deep learning*) otwiera zupełnie nowe ścieżki wykorzystywania informatyki i nauki o danych nawet jako metod analizy wyników badań klinicznych, wspomagania ich przeprowadzania czy jako narzędzi diagnostycznych. Obecne zbiory danych „wysokiego poziomu abstrakcji" (danych dotyczących problemów kompleksowych, często wielowymiarowych) pozostawiają jednak wiele do życzenia, gdyż najczęściej są nieustrukturyzowane, pozyskiwane z użyciem różnych procedur oraz nie posiadają odpowiednich metadanych pozwalających na ich normalizację. Przez właśnie ten brak standaryzacji tworzenia publicznych zbiorów danych (prywatne zbiory bywają budowane w sposób wzorowy, lecz nie są one dostępne do użytku publicznego) umożliwienie zastosowania technik „głębokiego uczenia maszynowego" jak i innych aspektów danologii (ang. *data science*). Niemniej jednak istnieją medyczne zbiory danych o dosyć dużym stopniu uporządkowania, publicznie dostępne w sieci. Jednym z tych zbiorów jest *Digital Database for Screening Mammography* – skompletowany w 2000 roku, zawierający dokładne opisy dla zanonimizowanych pacjentów, metadane anomalii oraz informacje dotyczące aparatury pomiarowej, celem normalizacji zbioru [1].

Nieprawidłowości zobrazowane w tym zbiorze, dzielą się na zwapnienia (klasa *calcification*) oraz guzy (klasa *mass*). Występuje również podzbiór zdjęć pochodzących od pacjentek zdrowych (klasa *normal*). Zestaw mammogramów dla każdej pacjentki składa się z dwóch rzutów (projekcji) dla obu piersi:

MLO - Mediolateral oblique view („projekcja skośna", „projekcja przyśrodkowo-boczna"), będąca podstawą oceny gruczołu piersiowego [2].

CC - Craniocaudal view („projekcja kranio-kaudalna", „projekcja góra-dół"), będąca uzupełnieniem tej oceny) [3].



Rycina 1 - Przedstawienie obu typów projekcji mammograficznej u tej samej pacjentki z nowotworem. CC – projekcja kranio-kaudalna, stanowiąca źródło informacji uzupełniających diagnozę, MLO – projekcja przyśrodkowo – boczna, stanowiąca źródło informacji podstawowych dla diagnozy.

Wykorzystanie uczenia maszynowego

Uczenie maszynowe mające korzenie wspólne ze statystyką (np. regresja liniowa, rachunek prawdopodobieństwa wraz z twierdzeniem Bayesa) posiada wiele gałęzi i zastosowań. Ogólny podział metod uczenia maszynowego można przedstawić w trzech paradygmatach:

Uczenie nadzorowane (ang. *supervised learning)* - wykorzystane w tej pracy, służące do rozwiązywania problemów takich jak regresja, klasyfikacja czy detekcja anomalii. Polega ono na tworzeniu algorytmu do ściśle określonego celu z wykorzystaniem zbioru treningowego zawierającego opis danych oraz jasno określonej funkcji kosztu. Przykładem tego typu uczenia maszynowego są systemy rozpoznawania twarzy [4].

Uczenie nienadzorowane (ang. *unsupervised learning*) - służące najczęściej w celach eksploracji danych, wyszukiwania zależności pomiędzy wartościami oraz katalogowania danych nieuporządkowanych. Przykładem są algorytmy klastrujące takie jak algorytm k-średnich (ang. “k-means”), najpopularniejszy algorytm analizy skupisk [4].

Uczenie ze wzmacnianiem (ang. *reinforcement learning*) - utworzone pod wpływem inspiracji psychologii behawioralnej. Wykorzystywane w problemach, o niejasnej strukturze, których nie sposób rozwiązać analitycznie. Wykorzystuje ono “środowisko” - będące meritum problemu, “agenta” - strukturę matematyczna podejmującą działania w środowisku oraz system nagrody i kary dla agenta (system penalizacyjny) za działania w środowisku [5]. Ten typ uczenia wykorzystywany jest najczęściej do tworzenia “sztucznej inteligencji” w grach komputerowych [6] lecz z powodzeniem wykorzystywany jest również w cheminformatyce czy naukach o życiu [7].

Sztuczne sieci neuronowe

Termin “sztuczna sieć neuronowa” powstał poprzez skojarzenie wykorzystywanych struktur matematycznych ze strukturą sieci, jaką tworzą neurony zwierząt. Tradycyjnie sztuczne sieci neuronowe, wykorzystywane głównie w uczeniu nadzorowanym, składają się z warstw perceptronów – jednostek wykonujących obliczenia, które wyrazić można poniższym wyrażeniem matematycznym:

$$a = g(W \times x^{(i)} + b)$$

Gdzie:

a - wartość wyjściowa perceptronu

g - funkcja aktywacyjna (np. funkcja sigmoidalna)

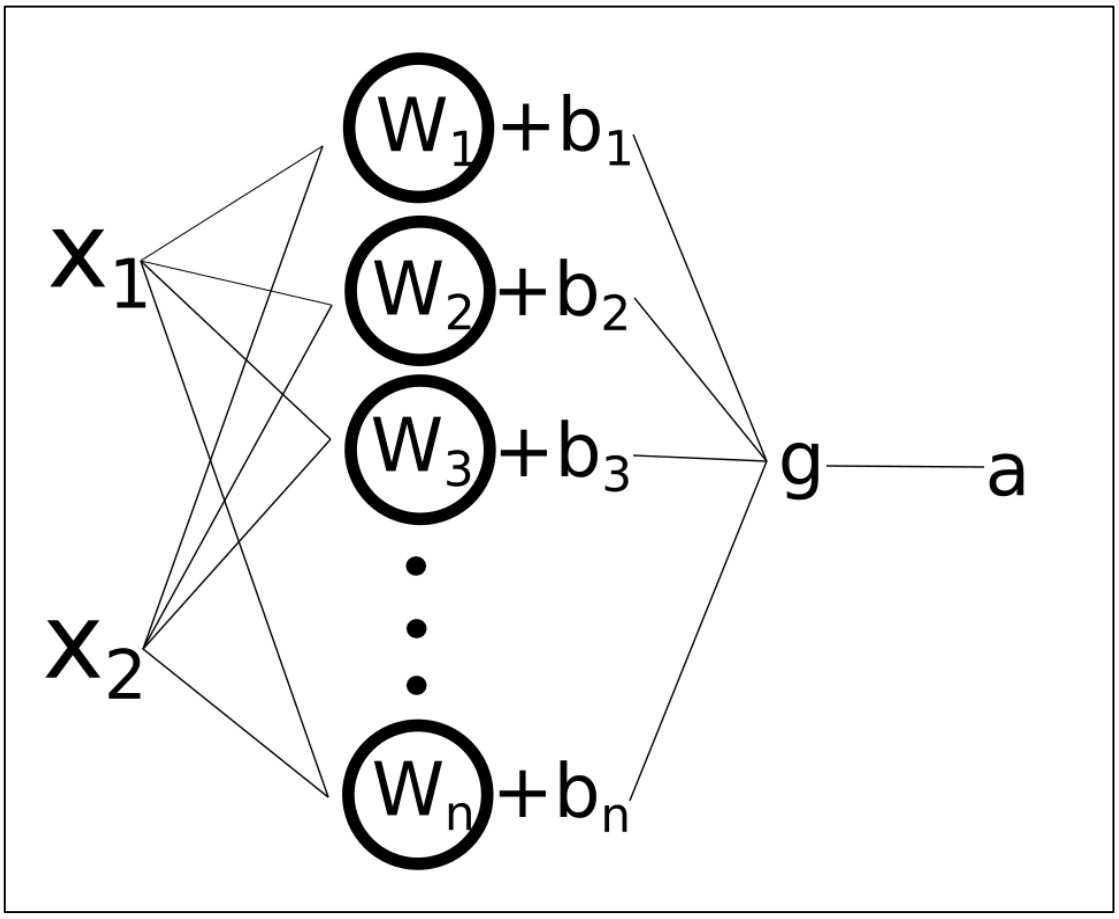
W - nauczalny wektor wag dla warstwy wejściowej

x - wartość wejściowa sieci (np. pojedyncza liczba rzeczywista, ich macierz lub wektor)

i - i-ty egzemplarz ze zbioru wartości wejściowych

b - nauczalna wartość przesunięcia progu aktywacji określana terminem “bias”

Funkcja aktywacyjna ma za zadanie przepuścić dalej wartości wyższe lub równe progowi aktywacji, zaś odrzucić (sprowadzić do wartości zerowej) wartości od niego niższe [8].



Rycina 2 - Graficzna przedstawienie sieci neuronowej z jedną warstwą perceptronów oraz dwiema wartościami wejściowymi. X – wartości wejściowe sieci, W – wagi , b – bias , g – funkcja aktywacyjna, a – wartość wyjściowa perceptronów (wynik działania sieci), n – numer perceptronu w warstwie.

Dla więcej niż jednej wartości wejściowej oraz więcej niż jednej warstwy, formuła matematyczna wygląda następująco:

$$a_j^{[l]} = g^{[l]}(\sum_k w_{jk}^{[l]} \times a_k^{[l-1]} + b_j^{[l]})$$

Gdzie:

a - wartość wyjściowa perceptronu

g - funkcja aktywacyjna (np. funkcja sigmoidalna)

l - numer warstwy

w - nauczalna waga dla warstwy wejściowej

$a^{[l-1]}$ - wartość wyjściowa poprzedniej warstwy sieci (lub wartości wejściowej w przypadku pierwszej warstwy)

j - j-ty element warstwy

k - k-ta wartość zbioru wartości wyjściowych poprzedniej warstwy sieci (lub wartości wejściowej w przypadku pierwszej warstwy)

b - nauczalna wartość przesunięcia progu aktywacji określana terminem "bias"

Sztuczne sieci neuronowe w znacznej większości bazują na algorytmie propagacji przedniej oraz propagacji wstecznej a przepływ informacji w sieci dla każdego z tych algorytmów jest jednostronny.

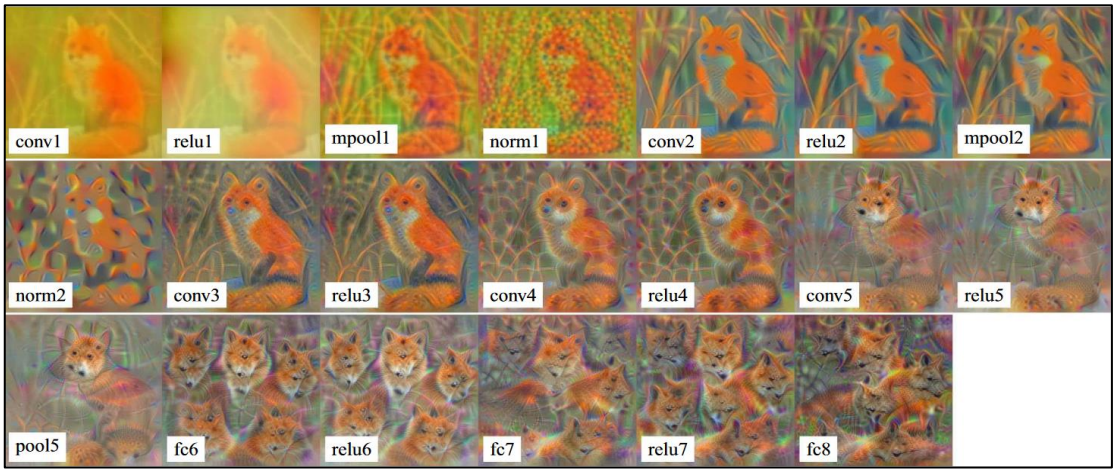
Algorytm propagacji przedniej jest algorytmem przetwarzania i propagowania w kolejnych warstwach informacji wejściowej podawanej sieci neuronowej. Jego rezultatem jest faktyczny wynik działania sieci neuronowej – wartość końcowa sieci.

Algorytm propagacji wstecznej jest algorytmem kluczowym dla zdolności "uczenia się" sieci. Polega on na obliczaniu pochodnych po każdym nauczalnym parametrze sieci oraz każdych wartości wyjściowych z warstw sieci wykorzystując regułę łańcuchową oraz wynik funkcji kosztu (odległości pomiędzy wartością predykcyjną a prawdziwą). Pochodne te nazywane gradientem są później wykorzystywane przez funkcje optymalizacyjne do obliczenia nowych nauczalnych parametrów sieci. Cały jeden etap treningu wykorzystujący propagację przednią i wsteczną nazywany jest epoką treningową (zwyczajowo czas trwania treningu wyraża się właśnie w epokach) i jest on powtarzany wielokrotnie, do czasu uzyskania satysfakcjonującego wyniku funkcji kosztu, lub uzyskania satysfakcjonującej wartości metryki sieci (np. czułości i swoistości - miar zdolności modeli używanych z powodzeniem również w analizie testów diagnostycznych w medycynie) [9].

Splotowe sztuczne sieci neuronowe

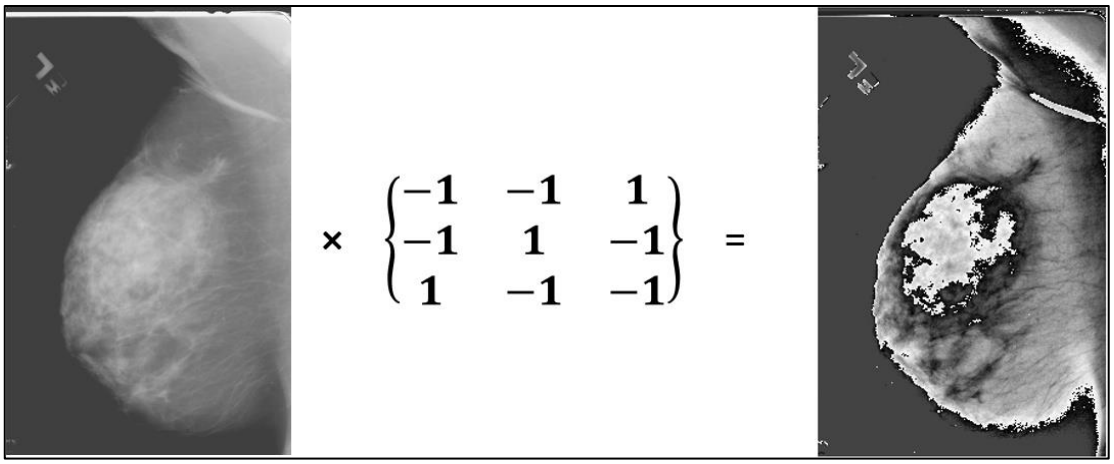
Komputerowa analiza obrazu z użyciem sztucznych sieci neuronowych, nazywana „widzeniem komputerowym” (ang. *computer vision*), odbiega bardzo mocno od klasycznych sieci typu „każda jednostka z każdą” (ang. *fully connected neural networks*). Do tego celu wykorzystywane są architektury zawierające warstwy „splotowe” (ang. *convolutional layers*), bazujące na idei mnożenia splotowego.

Sieci „splotowe” z początku wykorzystywane były głównie do problemu klasyfikacji a ich nauczalna funkcja jądra (ang. *kernel*) korzysta z korelacji wzajemnej - korelacji krzyżowej (ang. *corss-correlation*). Używanie tego typu warstw przyspiesza proces trenowania sieci poprzez możliwość nauczenia jej „wzorców”. Macierz pikseli jest segmentowana na wiele mniejszych macierzy o wymiarach równych wymiarom okna funkcji jądra, które są traktowane jako jedna wartość wejściowa warstwy, co redukuje ilość operacji matematycznych wykonywanych przez sieć. Z każdą kolejną warstwą sieci splotowe uczą się coraz bardziej złożonych cech obrazu, co pozwala ograniczyć głębokość sieci neuronowej czyli ilość warstw z których się składa.



Rycina 3 - Wizualizacja przetwarzania zdjęcia lisa przez kolejne warstwy splotowej sieci neuronowej; conv - warstwy splotowe, pool - warstwy zbierające (redukcja rozmiarów wartości wyjściowej poprzedniej warstwy), norm - warstwy normalizujące, relu – warstwy aktywacyjne wykorzystujące funkcję „Rectified Linear Unit”, będącą obecnie złotym standardem funkcjiaktywacyjnych, fc – warstwy „Fully Connected” (każdy perceptron jednej warstwy połączony jest z każdym perceptronem warstwy następnej). [10]

Inspiracją dla tych warstw była kora wzrokowa oraz to jak neurony korowe odpowiadają wyłącznie na ściśle określone bodźce wzrokowe związane z rozpoznawaniem wzorców. Podobnie jak w biologicznym procesie, filtry (funkcje jądra) warstw splotowych odpowiedzialne są za przesiew informacji matematycznej zawartej w analizowanym obrazie na podstawie której obliczana jest wartość wyjściowa sieci. Sieci splotowe również podlegają klasycznemu procesowi uczenia składającego się z propagacji przedniej (ang. *forward propagation*), obliczaniem błędu sieci (ang. *loss function*), polegającej na wyliczeniu funkcji kosztu oraz propagacji wstecznej (ang. *back propagation*) [11].

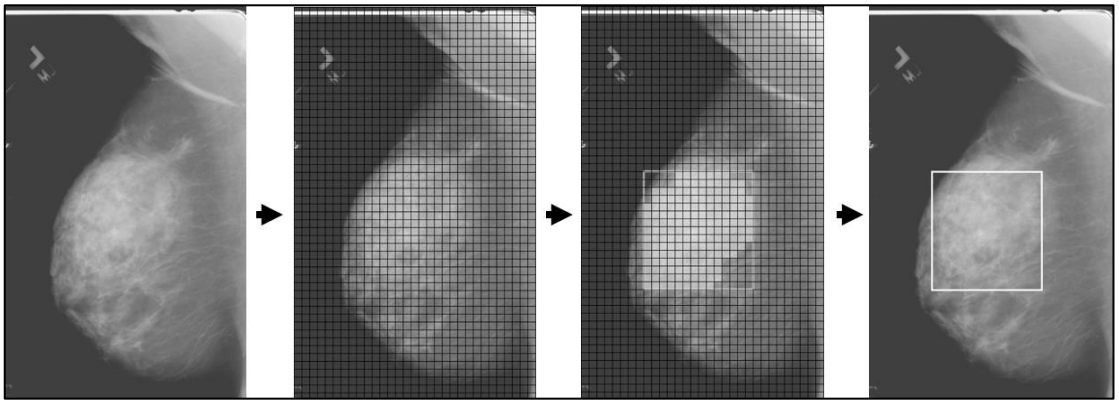


Rycina 4 – Przedstawienie działania funkcji jądra warstw splotowych na przykładzie wyszukiwania diagonalnych na obrazie mammograficznym zawierającym nowotwór. Okno funkcji wyszukującej diagonalne skanuje cały obraz (w tym przypadku przesuwając się w prawo lub w dół o jeden piksel – w zależności od położenia okna) czego rezultatem jest przetworzony obraz po prawej stronie.

Algorytm YOLO

Dla zagadnienia segmentacji lub identyfikacji ze wskazaniem położenia, pożądanego w kontekście radiologii i mammografii, niezbędne jest wykorzystanie najnowocześniejszych dostępnych architektur. W niniejszej pracy został wykorzystany algorytm *You Only Look Once* w wersji trzeciej (w skrócie *YOLOv3*) autorstwa Josepha Redmona [12], zaimplementowany w języku *Python*, przy użyciu pakietu *Tensorflow* [13]. Umożliwia wskazanie położenia regionu, w którym zawarty jest interesujący obszar, oraz jego klasyfikację.

Bazuje on na predykcji prostokątnej bryły brzegowej (ang. *bounding box*) oraz predykcji klasy i przynależności bryły brzegowej do danej klasy. Architektura składa się z części pełniącej funkcję ekstrakcyjną cech obrazu kluczowych dla predykcji klasy - *feature extractor,* oraz zespołu warstw pełniących funkcję obliczającą położenie regionu danej klasy. Działanie algorytmu rozpoczyna się poprzez utworzenie mapy cech kluczowych obrazu. Kolejnym krokiem jest jej segmentacja z użyciem siatki o znanych własnościach geometrycznych (podział obrazu na obszary o jednakowych wymiarach). Używając kombinacji mapy cech kluczowych oraz segmentacji obrazu, możliwym jest określenie współrzędnych (szerokość, wysokość, centrum) obszarów, w których znajdują się wyznaczone klasy (obszarów posiadających cechy tych klas). Następnie mapa cech poddawana jest procesowi o nazwie *upsampling* – polegającemu na zwielokrotnieniu informacji wejściowej w procesie interpolacji. Powiększenie skali mapy umożliwia dokładniejsze zlokalizowanie interesujących obszarów. W zależności od wersji architektury oraz oczekiwanej dokładności lokalizacji, proces ten może być powtórzony kilkukrotnie.



Rycina 5 - Graficzne przedstawienie działania warstw przeszukujących algorytmu YOLO. Od lewej: informacja wejściowa sieci neuronowej, graficzne przedstawienie siatki, na podstawie której obliczane będzie położenie miejsca zmienionego chorobowo, graficzne przedstawienie mapy przynależności do klasy na siatce wraz z zaznaczeniem bryły brzegowej, wyjściowy obraz wraz z zaznaczoną bryłą brzegową wskazującą miejsce zmienione chorobowo.

Funkcja kosztu algorytmu *YOLO*, przedstawiana jest równaniem:

$$\lambda coord \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} = [(xi - \hat{x}i)^2 + (yi - \hat{y}i)^2] +$$

$$\lambda coord \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\widehat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \right] +$$

$$\sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left(Ci - \hat{C}i\right)^2 +$$

$$\lambda noobj \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{noobj} \left(Ci - \hat{C}i\right)^2 +$$

$$\sum_{i=0}^{S^2} 1_{i}^{obj} \sum_{c \epsilon classes} (pi(c) - \widehat{pi}(c))^2$$

Gdzie:

xi, yi – przewidziane koordynaty centroidu bryły brzegowej

w_i, h_i – przewidziana szerokość i wysokość bryły brzegowej

Ci – punktacja stopnia pewności dla obiektu zainteresowań w obszarze

$pi(c)$ – „koszt" klasyfikacji

1_{ij}^{obj} – konstrukcja warunkowa – 1 gdy obiekt znajduje się w polu wiążącym, 0 gdy nie

1_{ij}^{noobj} – konstrukcja warunkowa – 1 gdy obiekt nie znajduje się w polu wiążącym, 0 gdy tak

$\sum_{j=0}^{B}$ - suma z wszystkich obszarów wiążących

$\sum_{i=0}^{S^2}$ - suma z wszystkich komórek obszarów wiążących

i – numer komórki obszaru wiążącego

j – numer obszaru wiążącego

$\hat{x}i, \hat{y}i$ – poprawne koordynaty centroidu bryły brzegowej

$\widehat{w}_i, \hat{h}_i$ - poprawna szerokość i wysokość bryły brzegowej

Ci – punktacja stopnia pewności dla przewidzianej bryły brzegowej

$\hat{C}i$ - punktacja stopnia pewności dla poprawnej bryły brzegowej

$\hat{p}i(c)$ – „koszt" poprawnej klasy (penalizacja tylko w przypadku obecności obiektu w komórce)

$pi(c)$ - "koszt" przewidzianej klasy (penalizacja tylko w przypadku obecności obiektu w komórce)

S^2 – dwuwymiarowa komórka siatki (np. S = 13 pikseli)
B – bryła brzegowa składająca się z dwuwymiarowych komórek (S^2)

$\lambda noobj$ – stała regularyzacji w przypadku braku obiektu w komórce siatki zawartej w bryle brzegowej

$\lambda coord$ – stała regularyzacji dla koordynatów bryły brzegowej

Dla elementów równania związanych z koordynatami bryły brzegowej oraz obecności obiektu, funkcją kosztu jest średni błąd kwadratowy. W przypadku klasyfikacji jest to entropia krzyżowa.

Biotechnologia

Badania biotechnologiczne, aby móc sprostać wymaganiom zleceniodawców oraz zapotrzebowaniem rozwijającego się świata z powodzeniem mogłyby wykorzystywać podobne algorytmy w „inteligentnej mikroskopii" (ang. *smart microscopy*) [14]. Przykładem może być automatyczna lokalizacja oocytu i klasyfikacja jego stadium dojrzewania przy badaniach dotyczących płodności czy też rozpoznawanie i klasyfikacja wzorców białek na obrazach mikroskopowych w celach diagnostycznych [15]. Stosowanie tego typu systemów mogłoby znacznie obniżać nakład czasu potrzebnego na poprawne wykonanie eksperymentu jak i zwiększyć ogólną wydajność laboratoryjną. Szerokie zastosowanie sztucznych sieci neuronowych jak i technik stosowanych w naukach o danych jest przyszłością dokładnej i dobrze uprawianej nauki. Wymaga to jednak wypracowania standardowych metod pozyskiwania danych, umożliwiających ich wykorzystanie niezależnie od czynników takich jak różnice w sprzęcie laboratoryjnym, protokołach wewnętrznych czy różnice w sposobie segregacji i zapisu danych.

Cel

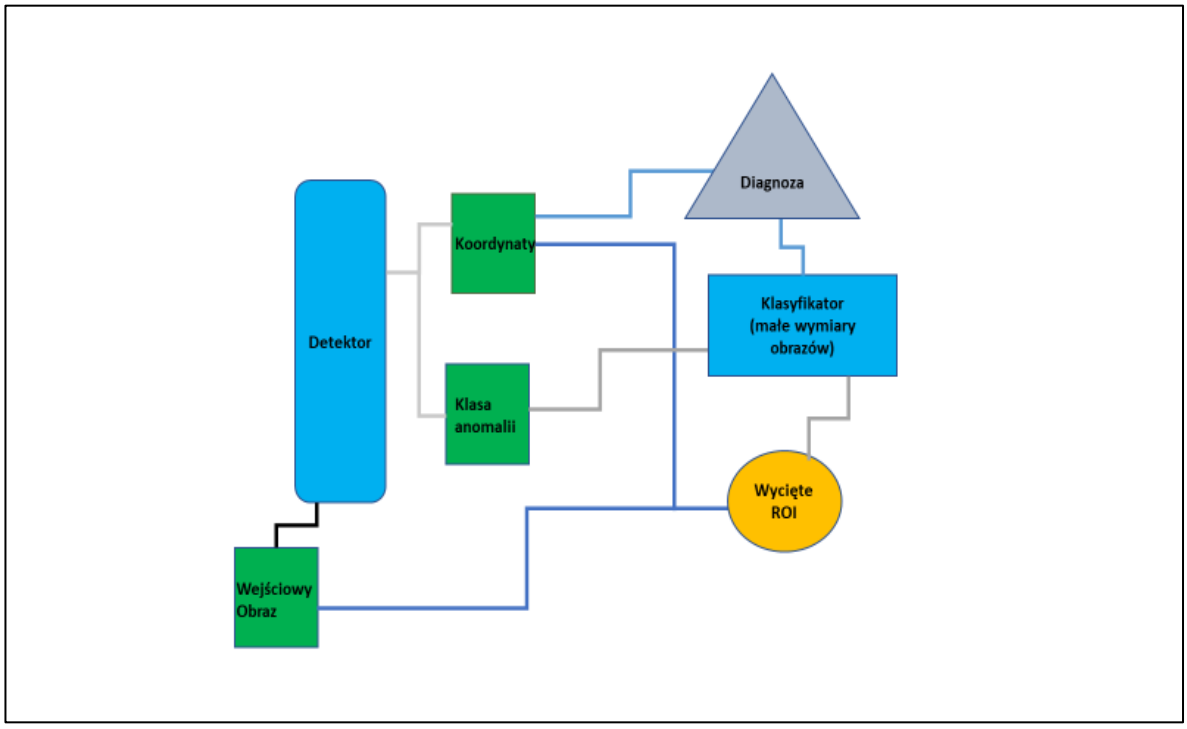
Inteligentny system diagnostyki piersi jest wyzwaniem przez lata podejmowanym przez wiele grup badawczych. Większość z nich skupia się na klasyfikacji mammogramów względem sześciu kategorii przynależności skali BI-RADS [16] lub klasyfikacji binarnej „Malignant vs Benign" (klasyfikacja zmian złośliwych lub łagodnych). Grupa dr. Krzysztofa Gerasa z New York University: School of Medicine wykorzystała potencjał uniwersyteckich danych aby podejść do tego zagadnienia z niespotykaną dotąd głębokością analizy i precyzją predykcji tworząc rewolucyjny system orzekania o złośliwości lub łagodności zmian, wraz z możliwością określenia ich lokalizacji. Nie mniej jednak system ten nie jest pozbawiony pewnych niedociągnięć. Przystosowany jest głównie pod nowoczesną aparaturę pomiarową oraz jako informację wejściową przyjmuje cztery mammogramy (dwie projekcje dla każdej piersi) - przez co jest bardzo kosztowny obliczeniowo. Algorytm nie odróżnia również guzów od zwapnień. System ten został wytrenowany na niedostępnym publicznie zbiorze danych co uniemożliwia jego ponowne wytrenowanie z dodatkiem własnych danych, czy możliwość skorzystania z nich celem jego rozwinięcia i modyfikacji [17] [18].

Korzystając z jednego z nielicznych dostępnych w sieci zbiorów danych w dziedzinie nauk o życiu, skupiającym się na „detekcji obiektów" (ang. *object detection*) będącego wysoce rozwiniętym problemem klasyfikacji – postanowiono stworzyć podobny system dystrybuowany na zasadach otwartego kodu źródłowego. Udostępniony w ten sposób system mógłby być dowolnie modyfikowany i ulepszany przez wszystkich zainteresowanych.

Celem pracy dyplomowej było przetestowanie algorytmu wykrywania obiektów pod kątem predyspozycji do utworzenia z niego rdzenia uniwersalnego, składającego się z sekwencji wysoce wyspecjalizowanych w pojedynczych zagadnieniach modeli systemu detekcji i klasyfikacji rodzaju anomalii na obrazach mammograficznych oraz eksploracja jakości i struktury wykorzystanego do tego celu zbioru danych. Uniwersalność systemu oznacza traktowanie każdego typu projekcji mammograficznej jako jedynego źródła informacji tzn. bez uzupełniania się.

Eksploracja danych jako element danologii stanowi podstawę wyciągania wniosków dotyczących wydajności modelu.

Rdzeń systemu dokonujący detekcji i wstępnego rozpoznania rodzaju miejsc zmienionych chorobowo stanowiłby podstawę dalszej analizy obrazu, przeprowadzanej tylko i wyłącznie na miejscach wyznaczonych przez niego. Określenie złośliwości i rodzaju zmiany wraz ze wskazaniem miejsc zmienionych chorobowo stanowiłoby uzupełnienie pracy lekarskiej, podstawę diagnozy i mogłoby przyspieszyć proces diagnostyczny czy też pozwolić na eliminację części błędów przypadkowych.



Rycina 6 - Schemat działania proponowanego systemu diagnostycznego . Wejściowy obraz jest przetwarzany przez detektor (YOLO) z rezultatem uzyskania informacji o rodzaju anomalii oraz o koordynatach miejsca zmienionego chorobowo na obrazie. Na podstawie tej informacji z obrazu wejściowego wycinany jest jego element zawierający wyłącznie miejsce zmienione chorobowo po czym wraz z informacją odnośnie rodzaju anomalii przetwarzany jest przez sieć neuronową pełniącą rolę klasyfikacji złośliwości zmiany. Wynik działania tego klasyfikatora wraz z informacjami dotyczącymi położenia miejsca zmienionego chorobowo jest prezentowany lekarzowi jako uzupełnienie informacji potrzebnych do do postawienia diagnozy.

Metody

Przygotowanie zbioru danych

Publiczny zbiór danych *Digital Database for Screening Mammography* (*DDSM*) pobrano z oficjalnego repozytorium.

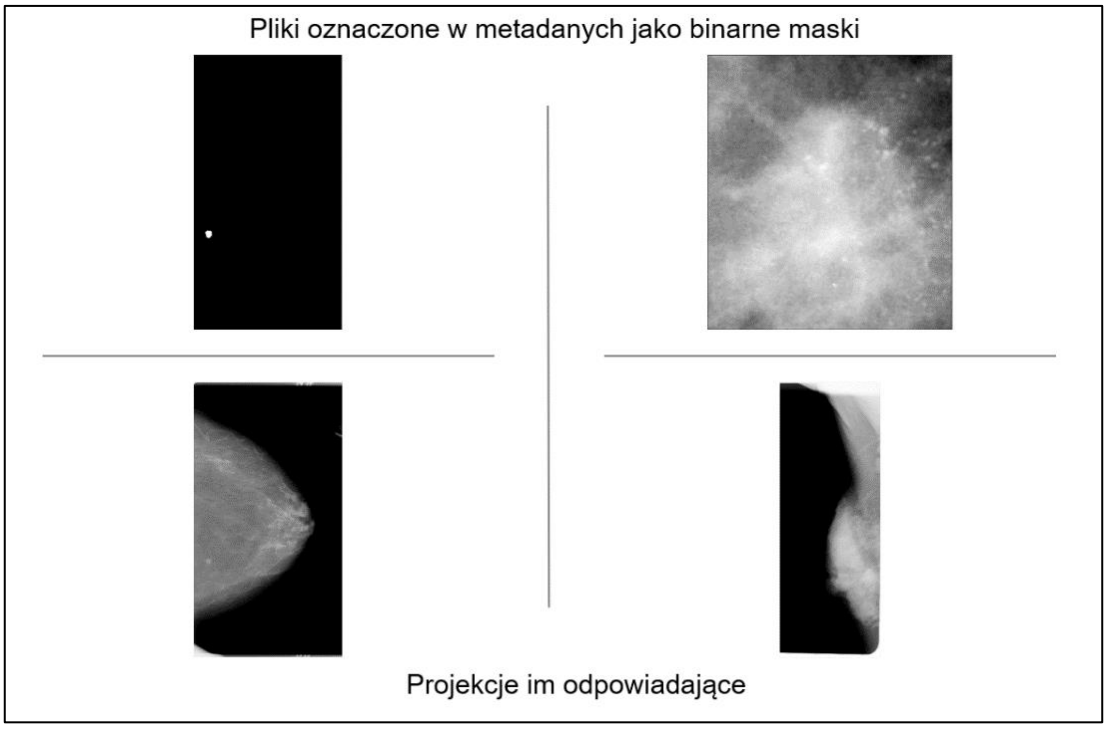
Wykonano selekcję obrazów z uwzględnieniem metadanych dotyczących struktury katalogowej zbioru danych, tak by podzbiory do dalszego wykorzystania składały się z klas „Calcification" – obrazy mammograficzne zawierające zwapnienia, „Mass" – obrazy mammograficzne zawierające guzy, oraz „Normal" – prawidłowe obrazy mammograficzne. Metadane te opisują wyłącznie przypadki, w których wykryte są anomalie, co umożliwiło utworzenie dwóch podzbiorów: podzbioru przeznaczonego do detekcji anomalii wraz z ich klasyfikacją oraz podzbioru przeznaczonego do samej detekcji anomalii.

Trzeci podzbiór przeznaczony do detekcji anomalii wraz z ich klasyfikacją utworzono wraz z przypadkami zdrowymi tworząc sztuczne opisy miejsc zdrowych w obrębie piersi. Klasa zdrowa pełni w nim funkcję klasy negatywnej. Pełny zbiór *DDSM* podzielony jest na podstawie kryterium złośliwości nowotworów i występują w nim również obrazy należące do *benign without callback* – zmiany łagodne bez potwierdzenia badaniem histopatologicznym – oraz *unsure*, których diagnoza budzi wątpliwości. Obrazy te pominięto w przygotowywaniu podzbiorów

Nie wykonano rozdzielenia obrazów z uwzględnieniem typu projekcji, celem próby utworzenia narzędzia uniwersalnego. Procedura normalizacji obrazów została wykonana z wykorzystaniem narzędzi utworzonych w języku *Python* przez Francisco Gimenez z zespołu dr. Rebecci Sawyer Lee z uniwersytetu Stanforda, w celu utworzenia nowej, wyczyszczonej wersji zbioru *DDSM (CBIS-DDSM)* [19].

Przygotowany przez ten sam zespół podzbiór *Curated Breast Imaging Subset of DDSM* niestety nie nadawał się do użytku czy modyfikacji ze względu na jego niedokładne

przygotowanie tj. metadane nie opisują tych plików, które powinny (występują stochastyczne błędy w segregacji).



Rycina 7 - Przedstawienie braku uporządkowania w zbiorze CBIS-DDSM. Binarne maski, na podstawie których zostałyby przygotowane koordynaty brył brzegowych dla YOLO, niejednokrotnie były zastąpione przez wycięte „Regions of Interests" (nie koniecznie pochodzące od tej samej pacjentki) lub nie miały tych samych wymiarów co pełny obraz mammograficzny.

Na podstawie wyżej wspomnianych narzędzi skonstruowano własny skrypt normalizujący, sortujący i ekstrahujący metadane z podzbioru. Kod udostępniono w repozytorium projektu (https://github.com/michalkowalski94/MSADnn).

Przygotowany podzbiór podzielono na kolejne podzbiory treningowe (stanowiące 80% obrazów) podzbiory walidacyjne (stanowiące 10% obrazów) oraz podzbiory testowe (stanowiące 10% obrazów). Podzbiory przygotowano pod problemy badawcze klasyfikacji typów anomalii oraz detekcji miejsc zmienionych chorobowo.

Zasoby obliczeniowe

W pracy magisterskiej wykorzystano zasoby obliczeniowe udostępnione przez *ACK Cyfronet AGH* na podstawie grantu obliczeniowego *neuralcbisDDSM* z wykorzystaniem infrastruktury *PLGrid*. Obliczenia zostały wykonane na klastrze *Prometheus*.

Trening algorytmu

Dla przedstawionego celu badawczego, wybrano najszybszą i najbardziej zaawansowaną obecnie architekturę *YOLO* w wersjach *YOLOv3* oraz *Tiny-YOLO*. Wersje te różnią się szkieletem warstw ekstrahującej cechy kluczowe (*feature extractor*) oraz skalami przeszukiwania obrazu (warstw mapujących, przeszukujących). Utworzono w języku *Python* skrypt przygotowujący plik z adnotacjami treningowymi pochodzącymi z metadanych uzyskanych podczas przygotowywania podzbioru *DDSM*. Na podstawie gotowej implementacji, utworzono w języku *Python* w pakiecie Keras na silniku Tensorflow skrypty treningowe dla modeli *YOLOv3* oraz *Tiny-YOLO* z możliwością zmian rozmiarów obrazów dla warstwy wejściowej (z zachowaniem ich proporcji i marginesami zawierającymi piksele o wartości 0).

Trening sieci przeprowadzano formie rozdystrybuowanej, z użyciem dwóch kart graficznych *nvidia Tesla K40d* o wbudowanej pamięci 11 GB na jednym węźle klastra. Architektury rozdzielano funkcjonalnością warstw, *feature extractor* na jednej karcie, warstwy mapujące na drugiej. Wykonano serię treningów algorytmu dla różnych rozmiarów obrazów
i problemu klasyfikacji na prawidłowe, ze zwapnieniem oraz z guzem. Metodą prób i błędów utworzono kilka wersji modeli gotowych do interpretacji. Trening przeprowadzano z użyciem podzbiorów treningowych oraz walidacyjnych.

Po przeanalizowaniu wyników funkcji kosztu zauważono, że zasoby obliczeniowe dla modeli YOLOv3 okazały się być niewystarczające - największe możliwe wymiary obrazów mieszczące się w pamięci kart graficznych (1760 pikseli na 1760 pikseli) powodowały utratę znacznej części informacji z mammogramów podczas ich skalowania, dlatego architekturę tę odrzucono.

Utworzono sześć modeli *Tiny-YOLO* dla rozmiarów 4800 pikseli na 4800 pikseli oraz 3424 piksele na 2432 piksele (z uwagi na dużą ilość obrazów o proporcji wysokości obrazu do szerokości wynoszącej blisko 1.40:1) dla trzech problemów klasyfikacji:

- klasyfikacja obszarów normalnych, ze zwapnieniem oraz z guzem

- klasyfikacja obszarów ze zwapnieniem oraz z guzem

- wyszukiwanie obszarów z widoczną anomalią

Trening każdego z modeli kończono w momencie widocznego plateau funkcji kosztu. Z uwagi na bardzo duże wymiary obrazów, jeden trening trwał od 9 do 14 dni, ze średnim czasem trwania jednej epoki wynoszącej jeden dzień.

Końcowe modele przetestowano pod kątem:

- Czułości - dla klasyfikacji, stanowiącej stosunek wyników prawdziwie dodatnich do sumy prawdziwie dodatnich i fałszywie ujemnych. Jest to miara zdolności poprawnej klasyfikacji pozytywnej.
- Specyficzności (swoistości) - stosunek wyników prawdziwie ujemnych do sumy prawdziwie ujemnych i fałszywie dodatnich. Jest to miara zdolności poprawnej klasyfikacji negatywnej.
- F1 - średnia harmoniczna czułości i specyficzności.
- Wartości predykcyjnej dodatniej - będącej proporcją prawdziwie pozytywnych wyników wśród wszystkich wyników pozytywnych. Wyraża ona prawdopodobieństwo prawdziwości pozytywnego wyniku.
- Wartości predykcyjnej ujemnej - będącej proporcją prawdziwie negatywnych wyników wśród wszystkich wyników negatywnych. Wyraża ona prawdopodobieństwo prawdziwości negatywności wyniku.
- Indeksu Jaccarda (ang. *Intersection over Union*, *IoU*) - mierzącego podobieństwo między dwoma zbiorami. Definiowany jest jako iloraz części wspólnej zbiorów do sumy zbiorów. Używany jest on jako miara poprawnej detekcji brył brzegowych.
- Prawidłowości klasyfikacji (precyzji) – wyrażaną poprzez procent poprawnie zaklasyfikowanych przypadków

Analizę statystyczną wyników wykonano z użyciem języka programowania R.

Wyniki

Używając programu napisanego w języku *Python* przeprowadzono porównanie wydajności modeli na podzbiorze testowym. Mając na uwadze możliwość więcej niż jednej detekcji na obraz oraz informację o tylko jednej lokalizacji miejsca zmienionego chorobowo na mammogram, w finałowym zestawieniu brano pod uwagę najlepsze wyniki z prób.

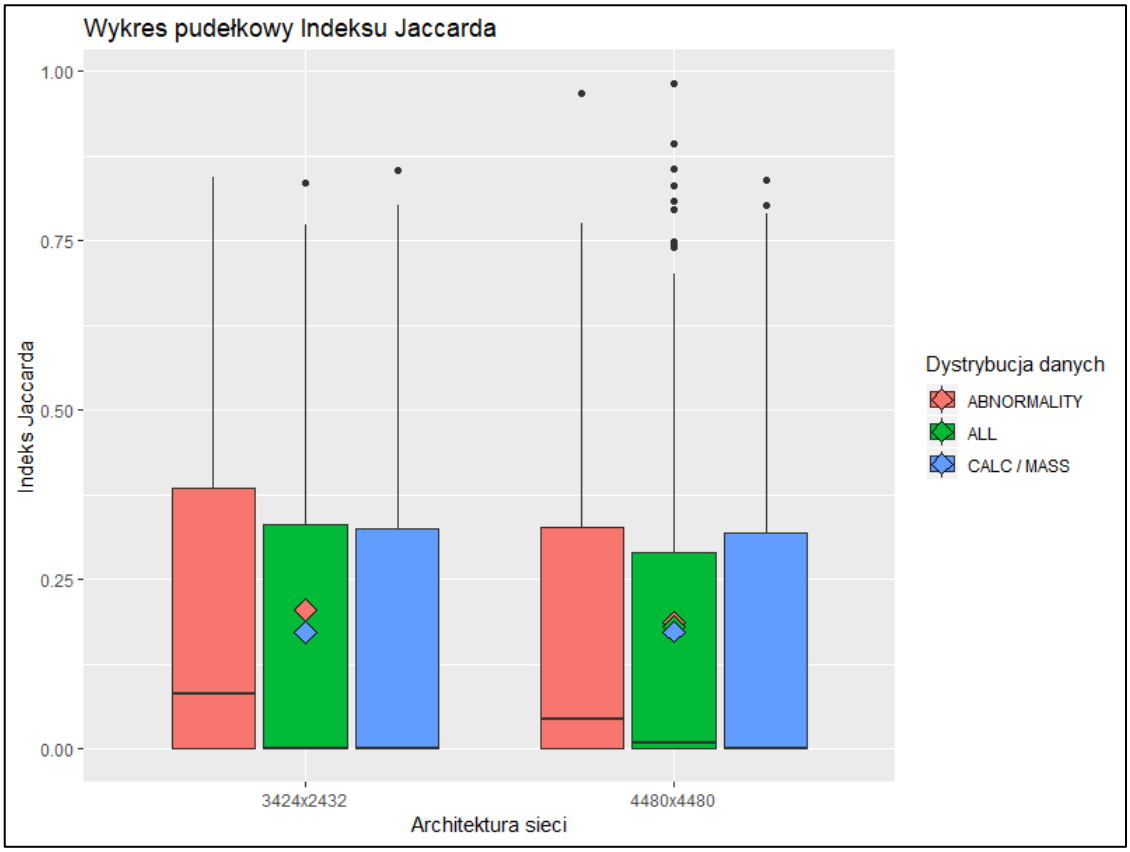
Detekcja anomalii

Wyniki indeksu Jaccarda w populacjach (jako populacja traktowane były zbiorcze wyniki pochodzące od jednego algorytmu) zostały poddane analizie testem Shapiro-Wilka w celu sprawdzenia normalności rozkładu. Następnie przeprowadzono na nich analizę porównawczą zdolności poprawnego identyfikowania miejsc zmienionych chorobowo używając testu Kruskala-Wallisa.

Tabela 1 - Wyniki zbiorcze indeksu Jaccarda dla sześciu wariantów architektury Tiny YOLO.

Typ:Wymiary	**Ilość klisz**	**Indeks Jaccarda (średnia [SD])**
ABNORMALITY:3424x2432	**200**	**0.20 (0.25)**
ALL:3424x2432	186	0.17 (0.24)
CALC / MASS:3424x2432	200	0.17 (0.24)
ABNORMALITY:4480x4480	200	0.19 (0.24)
ALL:4480x4480	186	0.18 (0.25)
CALC / MASS:4480x4480	200	0.17 (0.25)
p		**0.190**
test		Kruskal-Wallis

Różnice pomiędzy wariantami modeli oraz modelami okazały się być nieistotne statystycznie. Używając pakietu *ggplot2* wykonano wykres pudełkowy dla indeksu Jaccarda wszystkich modeli, celem łatwiejszej interpretacji.



Wykres 1 – Wykres pudełkowy indeksu Jaccarda dla sześciu najlepiej lokalizujących rejony zainteresowań modeli. Dystrybucja danych określa podzbiór wykorzystywany dla danego modelu oraz zadanie jakie wykonywał: ABNORMALITY – wyłącznie detekcja anomalii, ALL – detekcja i klasyfikacja wszystkich klas, CALC/MASS – detekcja i klasyfikacja wyłącznie guzów i zwapnień. Wertykalnymi liniami zaznaczone są wartości powyżej trzeciego kwartylu, liniami horyzontalnymi zaznaczone są mediany, romboidalne kształty odpowiadają oznaczeniu średnich arytmetycznych zaś kropkami zaznaczone są wartości odstające. Pudełka są reprezentacjami rozstępów kwartylnych.

Klasyfikacja typów anomalii

Korzystając z pakietu *caret* przeprowadzono analizę statystyczną zdolności poprawnej klasyfikacji anomalii u sieci dla trzech klas („normal” - klasa sztuczna, „mass” - guzy, „calcification”- zwapnienia) oraz u modeli dla dwóch klas („mass”, „calcification”).

Tabela 2 - Statystyki dotyczące wydajności modeli wobec klasyfikacji trój-klasowej. Puste komórki są wynikiem braku możliwości wyliczenia pewnych statystyk (głównie dotyczy to klasy sztucznej gdyż nie została ani razu wykryta). Klasy: „norm” – klasa zdrowa (sztuczna), „calc” – zwapnienie, „mass” – guz.

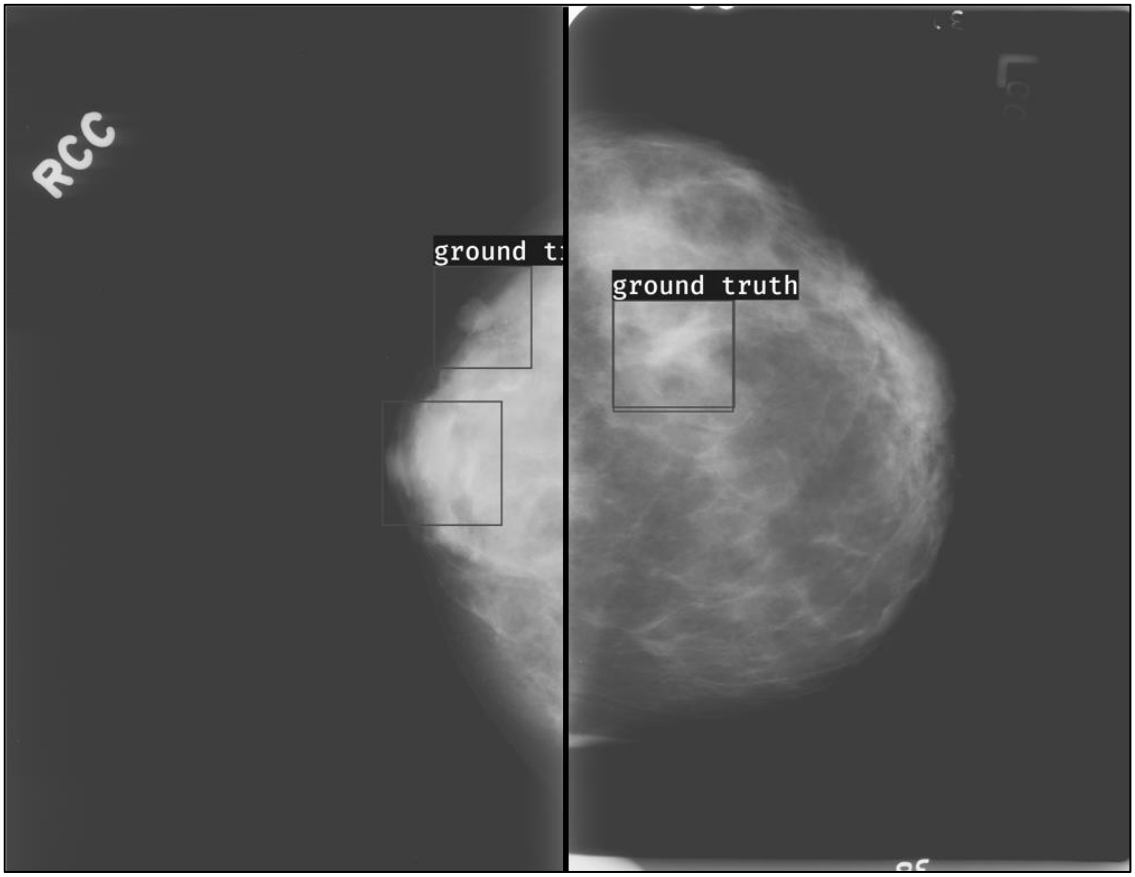
Statystyki	**3424 norm (sztuczna)**	**3424 mass**	**3424 calc**	**4480 norm (sztuczna)**	**4480 mass**	**4480 calc**
n	153	102	80	153	102	80
Czułość		**0.981**	**0.098**		1.000	0.000
Specyficzność	1.000	0.098	0.981	1.000	0.000	1.000
Wartość predykcyjna dodatnia		0.580	0.800		0.559	
Wartość predykcyjna ujemna		0.800	0.580			0.559
F1		0.729	0.174		0.717	
Prawidłowość		0.539	0.539		0.500	0.500

Tabela 3 - Statystyki dotyczące wydajności modeli wobec klasyfikacji dwu-klasowej. Klasy: „mass" – guz, „calc" – zwapnienie.

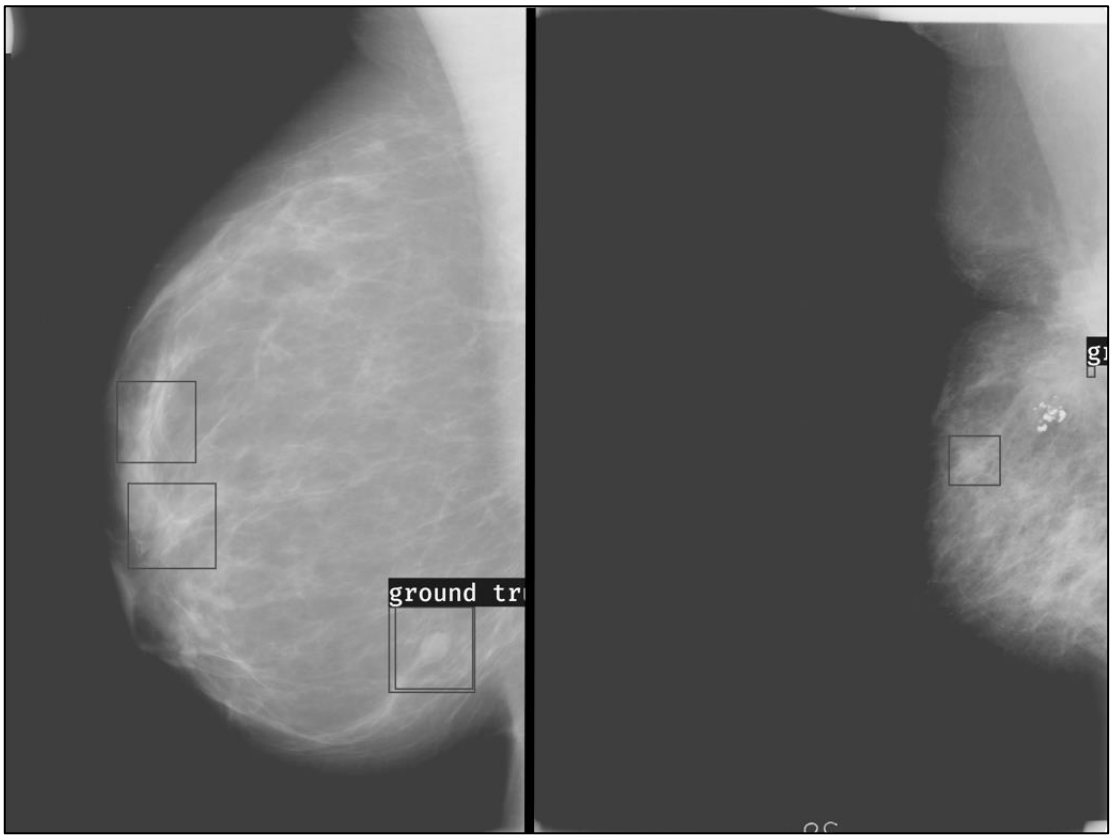
Statystyki	**3424 mass**	**3424 calc**	**4480 mass**	**4480 calc**
n	116	78	116	78
Czułość	0.975	0.013	0.975	0.051
Specyficzność	0.013	0.975	0.051	0.975
Wartość predykcyjna dodatnia	0.602	0.250	0.611	0.571
Wartość predykcyjna ujemna	0.250	0.602	0.571	0.611
F1	0.744	0.024	0.752	0.093
Prawidłowość	0.494	0.494	0.513	0.513

Analiza wizualna

Mając na względzie główne zadanie detektora – detekcję obszarów zmienionych chorobowo, zwizualizowano wyniki modelu z najwyższym średnim indeksem Jaccarda (tj. 3423 piksele na 2432 piksele – wykrywanie anomalii bez ich klasyfikacji) dla obu rodzajów projekcji.

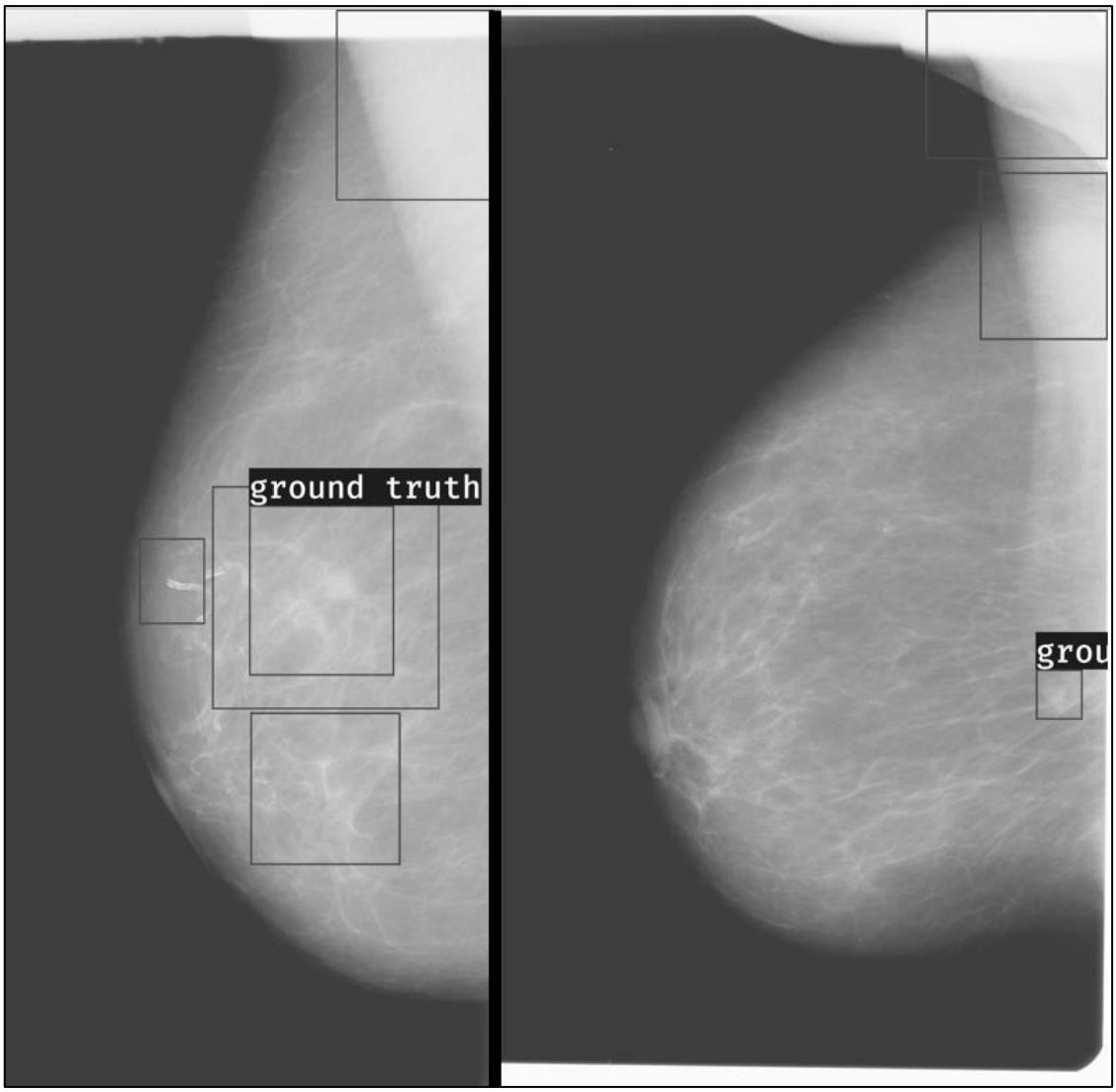


Rycina 8 - Porównanie najgorszego (lewa strona) i najlepszego (prawa strona) wyniku predykcji na projekcji CC, modelu o najwyższym średnim indeksie Jaccarda. Lewa strona: C_0336_1.RIGHT_CC. Prawa strona: B_3492_1.LEFT_CC. Obszar podpisany jako „ground truth" przedstawia miejsce oznaczone przez lekarza jako zmienione chorobowo. Ucięcie podpisu w centrum ryciny spowodowane jest jego częściowym wygenerowaniem poza granicami obrazu.



Rycina 9 - Porównanie najgorszego (lewa strona) i najlepszego (prawa strona) wyniku predykcji na projekcji MLO, modelu o najwyższym średnim indeksie Jaccarda. Lewa strona: A_1121_1.LEFT_MLO. Prawa strona: B_3025_1.RIGHT_MLO. Obszar podpisany jako „ground truth" przedstawia miejsce oznakowane przez lekarza jako zmienione chorobowo. Ucięcie podpisu na rycinach spowodowane jest ich częściowym wygenerowaniem poza granicami obrazu.

Podczas eksploracji wyników zauważono, że algorytm czasem zaznaczał artefakty lub mięśnie pacjentek jako obszar zmieniony chorobowo.



Rycina 10 - Predykcje na mięśniach. Zaznaczanie mięśni (obrazów o wysokiej gęstości optycznej) jako miejsc zmienionych chorobowo, świadczy o tym, że sieć nauczyła się rozpoznawać klasę głównie na podstawie tego parametru. Ucięcie podpisu po prawej stronie ryciny spowodowane jest jego częściowym wygenerowaniem poza granicami obrazu.

Wyniki dla wszystkich pacjentek ze zbiorów testowych zestawiono w formie tabeli dostępnej na stronie repozytorium.

Dyskusja

Ogólna wydajność wszystkich modeli względem postawionego problemu badawczego jest bardzo słaba. Jest to wynik wielu czynników dotyczących jakości zbioru danych jak i samego wykorzystania architektury.

Interpretacja wyników

Wszystkie przedstawione w wynikach modele wykazały się wydajnością nie wystarczającą do wykorzystania jako rdzenia systemu diagnostycznego. Najwyższy uzyskany średni indeks Jaccarda (model 3424piksele na 2432 piksele dla samego zagadnienia lokalizacji) jest i tak nie wystarczający, gdyż jako zadowalający wynik przyjmuje się średnią większą niż 0.5.

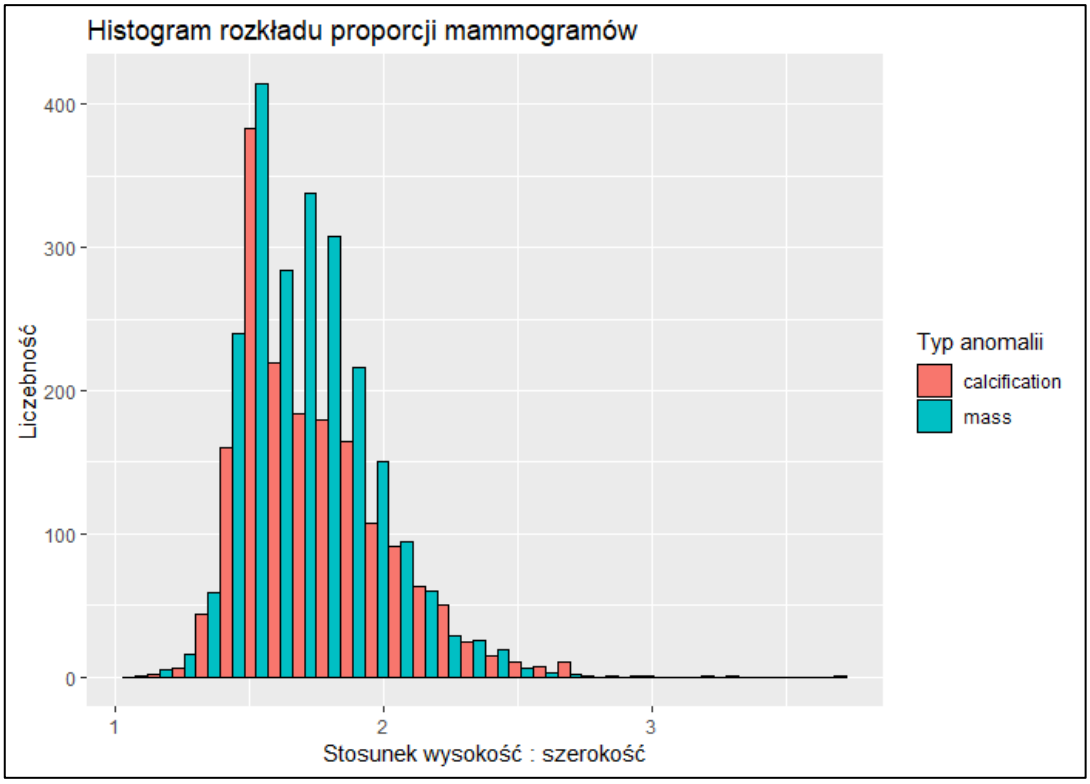
Rozwiązanie zagadnienia klasyfikacji praktycznie nie istnieje, gdyż uzyskane modele w niemal wszystkich przypadkach przewidują obecność wyłącznie guza. Argumentem przemawiającym na niekorzyść dalszego stosowania mammografii jako testu diagnostycznego jest częsty brak możliwości rozróżnienia pomiędzy tkankami gruczołowymi a faktycznymi guzami u kobiet poniżej pięćdziesiątego roku życia – informacja przekazana ustnie przez lekarza radiologa.

Powyższe problemy są bezpośrednio powiązane z architekturą *Tiny-YOLO*, gdyż oryginalna wersja przystosowana była do obrazów o wielkości 416 pikseli na 416 pikseli (ilość warstw oraz rozmiar funkcji jądra) jak i z jakością użytych do wytrenowania sieci neuronowych danych.

Jakość zbioru danych

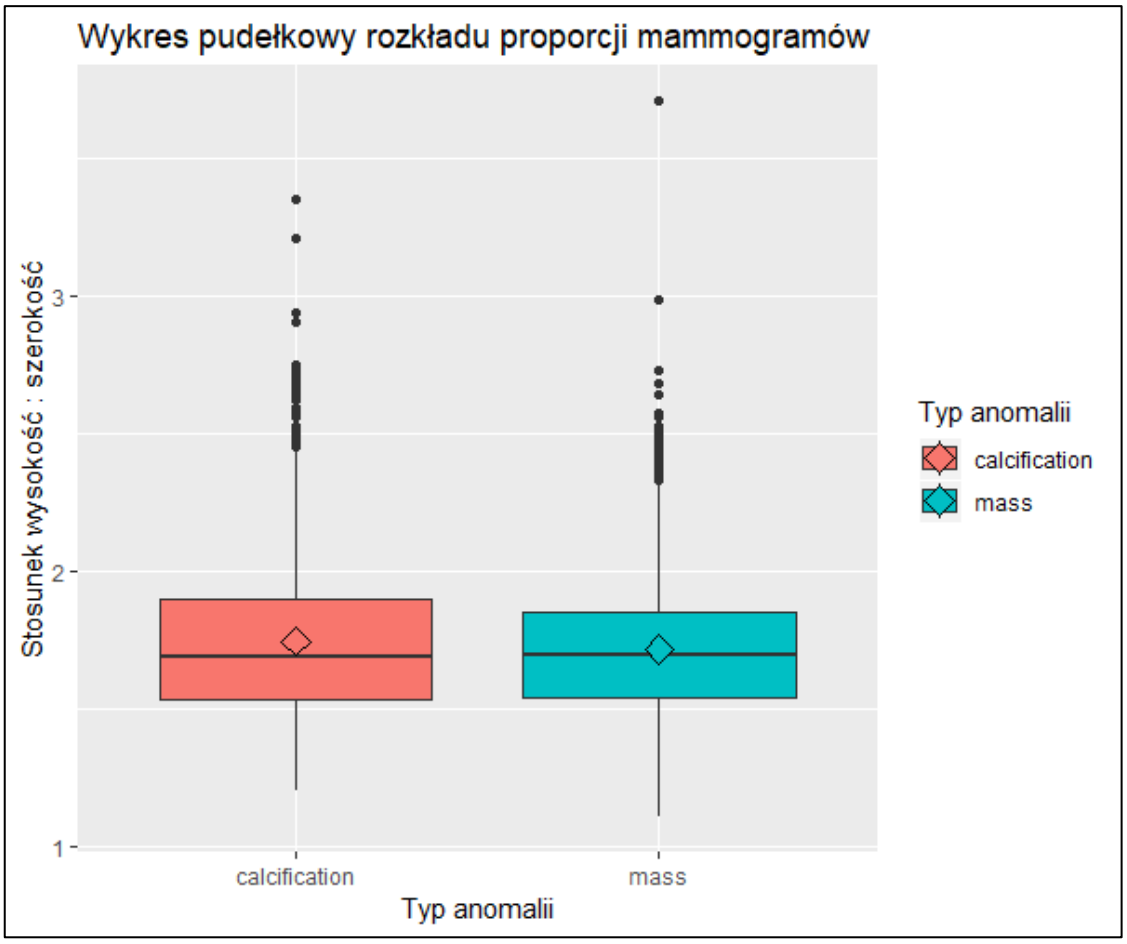
Mammogramy zawarte w zbiorze nie były pozyskiwane według ustandaryzowanej procedury. Obrazy te są kliszami faktycznego wyniku obrazowania, skanowanymi na różnych urządzeniach w różnej rozdzielczości. Obrazy te nie mają ustandaryzowanych proporcji czy rozmiarów, zaś normalizacja ich była wykonywana mając na względzie metadane dotyczące aparatury pomiarowej używanej dla danych partii. Proporcje wysokości do szerokości obrazów wahają się od niemal 1:1 do ponad 3.5:1 co skutkuje utratą danych na poziomie przygotowywania obrazów do warstwy wejściowej sieci, gdyż podczas skalowania, wartości oryginalne pikseli są zastępowane

przez nowe, dopasowane do skali, utworzone tak by jedynie wizualnie były zbliżone do pierwowzoru.
W przypadku jednej z pierwszych sieci (przyjmujących obrazy o wymiarach 1760 pikseli na 1760 pikseli), guz niejednokrotnie sprowadzany był do formy pojedynczego piksela, a zwapnienia znikały z obrazów.



Wykres 2 - Histogram rozkładu stosunku wysokości do szerokości obrazów w zbiorze DDSM. Wykres podzielony na typy anomalii: calcification - zwapnienie, mass - guz.

Skalowanie (z zachowaniem proporcji obrazów oraz użyciem marginesów pikseli o wartości zerowej) do rozmiaru o proporcji równej medianie ze stosunków proporcji (lub średniej, gdyż wartości te są zbliżone) mogłoby poprawić wydajność sieci. Stosowanie proporcji 1:1 i rozmiarów będących wielokrotnością 32 jest stosowane jako standard, gdyż skutecznie przyspiesza proces treningu sieci z uwagi na techniczne aspekty działania kart graficznych. W tym przypadku jednak prędkość jej działania jest nieistotna dopóki sieć nie będzie odnosić sukcesów w predykcji.



Wykres 3 – Wykres pudełkowy stosunku wysokości do szerokości obrazów w zbiorze DDSM. Wykres podzielony na typy anomalii: calcification - zwapnienie, mass - guz. Wertykalnymi liniami zaznaczone są wartości powyżej trzeciego kwartylu oraz poniżej pierwszego kwartylu, liniami horyzontalnymi zaznaczone są mediany, romboidalne kształty odpowiadają oznaczeniu średnich arytmetycznych zaś kropkami zaznaczone są wartości odstające. Pudełka są reprezentacjami rozstępów kwartylnych.

Ponadto w podgrupie *ROI* zbioru *DDSM* (wycięte interesujące rejony mammogramów do celów klasyfikacji) zawartych jest więcej obszarów aniżeli opisanych w metadanych (w których na jeden mammogram przypada tylko jeden obszar zmieniony chorobowo). Jest możliwe, że sieć wynajdywała w niektórych przypadkach miejsca zmienione chorobowo, o których poprawności nie miała prawa „wiedzieć", gdyż nie były one zawarte w pliku adnotacyjnym. Jednym z możliwych rozwiązań jest powtórne postawienie diagnoz na mammogramach i oznakowanie każdego z miejsc zmienionych chorobowo.

Użycie sztucznej klasy „normalnej” (obszarów zaznaczonych w okolicach centrum wysokości mammogramów od pacjentek zdrowych) okazało się nie być właściwą drogą na wdrożenie klasy niezmienionej chorobowo. Powinno ono nastąpić poprzez odręczne zaznaczenie obszarów o dużym gradiencie gęstości optycznej tkanki, tak by sieć nie mogła się nauczyć używać gęstości jako właściwości klas zmienionych chorobowo (mogłoby to pomóc w problemie klasyfikacji typów anomalii jak i stanowiłoby dobrą podstawę dla problemu ich samej detekcji – wprowadziłoby to klasę negatywną).

Architektura

Utworzenie *feature extractora* specyficznego do obranego zagadnienia i wielkości obrazów oraz obranie tylko jednej warstwy lokalizacyjnej wybranej dla średniej ich rozdzielczości (lub mediany, wartość należy ustalić eksperymentalnie) mogłoby znacznie poprawić wydajność sieci. Przy wykorzystywanym zbiorze danych leży to jednak na granicy możliwości, gdyż znaczna część obrazów po skalowaniu do obranego rozmiaru o ustalonej proporcji, i tak będzie niosła informację w różnych skalach. Dla jak najlepszego wykorzystania potencjału sieci, informacje powinny być ustandaryzowane (o jednakowych proporcjach i jednakowej skali). Możliwym rozwiązaniem tego problemu dla dalszego rozwoju badań jest ręczna obróbka wszystkich mammogramów do jednakowych wymiarów (co również niesie za sobą konieczność ponownego oznakowania miejsc zmienionych chorobowo – sprecyzowania nowych koordynatów dla *ROI*). Jednak każda tego typu ingerencja w zbiór danych musi być poprzedzona konsultacją lekarską (lub zostać przez lekarzy dokonana, gdyż osoba bez specjalistycznej wiedzy z dziedziny radiologii nie jest w stanie poprawnie interpretować tych obrazów) o ile system miałby faktycznie być wartościowy dla diagnostyki nowotworów piersi.

Wnioski końcowe

Eksperyment

W celu poprawienia jakości rdzenia systemu (wykluczając możliwość edycji danych), należy utworzyć architekturę opierającą się na zasadzie działania algorytmu *YOLO* lecz z użyciem *feature extractora* przystosowanego do obrazów o dużej rozdzielczości. Architektura ta powinna również być przystosowana do detekcji w jednej skali mapy cech kluczowych, tak by nie tracić mocy obliczeniowej na dwie lub trzy skale jak w przypadku testowanych wersji. Problem na którym powinna skupiać się siec to detekcja anomalii (należy zatem odręcznie zaznaczyć obszary klasy zdrowej o dużej gęstości optycznej, w ten sposób powinno dać się wykluczyć wykorzystanie jej jako jednego z głównych kryteriów klasyfikacji), aby klasyfikacja na rodzaje anomalii i złośliwość mogły zostać orzeczone jako wynik działania sieci operujących już na mniejszych obrazach – *ROI* wykrytych przez tę architekturę. Potencjalnym rozwiązaniem mogłoby również być wykorzystanie podczas treningu obrazów od pacjentek zdrowych jako klasy negatywnej ale bez podawania żadnej informacji dotyczącej miejsc o zwiększonej gęstości optycznej, w ten sposób sieć mogłaby nauczyć się odrzucać ten parametr w zupełności.

Dane w medycynie i biotechnologii

Dane ze zbioru *DDSM*, mimo iż z pozoru dokładnie skompletowane i skatalogowane, zawierają wiele wad, które nie są widoczne na pierwszy rzut oka. Zbiór ten został utworzony stosunkowo dawno, kiedy nie istniały jeszcze uniwersalne standardy dotyczące zbierania, obróbki czy segregacji danych. Obecnie badacze z dziedzin takich jak medycyna, biologia czy biotechnologia powinni znać podstawy „czystości danych", tworzyć plany dotyczące ich kompletowania oraz sięgać po opinię specjalistów, którzy na tych danych będą pracować.

Rodzi to jednak zagrożenie dysonansu postrzegania problemów badawczych i natury danych, naukowcy z dziedzin nauk o życiu nie posiadają tej samej wiedzy co naukowcy zajmujący się danymi
i informacjami (jak i odwrotnie).

Aby móc bez przeszkód korzystać z najnowszych technologii analizy informacji w celach biologiczno – medycznych (oraz by móc je rozwijać), powinna istnieć dyscyplina łącząca aspekty bioinformatyki, biologii oraz danologii. Z wykorzystaniem tej dziedziny, tworzenie czystych zbiorów danych, ich analiza czy konstruowanie dla nich modeli matematycznych lub systemów sztucznej inteligencji, stanowiłoby uzupełnienie biotechnologii oraz pozwalałoby na potencjalną redukcję kosztów poprzez lepszą automatyzację pracy, optymalizację jej oraz ograniczenie błędów przypadkowych.

Podziękowania

Chciałbym wyrazić wdzięczność wobec prof. dr hab. Marty Pasenkiewicz-Gieruli za umożliwienie rozwoju w dziedzinie danologii oraz za ciągłe zainteresowania tematem pracy i postępami. Wyrazy podziękowania kieruję również do mgr. Wojciecha Gałana za ogromny wkład merytoryczny, konstruktywną krytykę, czujne oko oraz umożliwienie wykonania tej pracy z wykorzystaniem infrastruktury *PLGrid*.

Zasoby

Grant obliczeniowy *neuralcbisddsm,* infrastruktura *PLGrid,* klaster *Prometheus*.
Język programowania *Python* w wersji 3.6.5 oraz biblioteki:

- *tensorflow-gpu* w wersji 1.12 oraz 1.6.
- *Keras* w wersji 2.1.5.
- Pillow w wersji 5.4.1.
- opencv-python w wersji 4.0.0.21.

- inne biblioteki wymagane przez wyżej wymienione.

Biblioteka *CUDA* w wesji 9.0.

Biblioteka *CUDnn* w wersji 7.3.1.

Język programowania R w wersji 3.5.3 oraz biblioteki:

- ggplot2 w wersji 3.1.
- caret w wersji 6.0.
- inne biblioteki wymagane przez wyżej wymienione.

Spis literatury

[1] U. o. S. Florida, "DDSM: Digital Database for Screening Mammography," University of South Florida, 2001. [Online]. Źródło: http://marathon.csee.usf.edu/Mammography/Database.html. [11.2018].

[2] C. Hacking and S. Pacifici, "Mediolateral oblique view," [Online]. Źródło: https://radiopaedia.org/articles/mediolateral-oblique-view. [27.05.2019].

[3] C. Hacking and S. Pacifici, "Craniocaudal view," [Online]. Źródło: https://radiopaedia.org/articles/craniocaudal-view. [27.05.2019].

[4] S. Raschka and V. Mirjalili, Python Machine Learning Second Edition, Birmingham: Packt Publishing Ltd., 2017.

[5] R. S. Sutton and A. G. Barto, "Introduction," in *Reinforcement Learning: An Introduction*, Cambridge, MA, MIT Press, 2018.

[6] OpenAI, "Dota 2," OpenAI, 11 04 2017. [Online]. Źródło: https://openai.com/blog/dota-2/. [27.05.2019].

[7] M. Olivecrona, T. Blaschke, O. Engkvist and H. Chen, "Molecular de-novo design through deep reinforcement learning," BMC Part of Springer Nature, 09.09.2017. [Online]. Źródło: https://jcheminf.biomedcentral.com/articles/10.1186/s13321-017-0235-x. [27.05.2019].

[8] A. Sharma, "Understanding Activation Functions in Neural Networks," A Medium Corporation (US), 30.03.2017. [Online]. Źródło: https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0. [27.05.2019].

[9] Leland Stanford Junior University, *Machine Learning*, Coursera, Inc. [Kurs Online]

[10] A. Mahendran and A. Vedaldi, "Visualizing Deep Convolutional Neural Networks Using Natural Pre-images," *International Journal of Computer Vision*, vol. 3, 2016.

[11] I. Goodfellow, Y. Bengio and A. Courville, "Convolutional Networks," *Deep Learning*, MIT Press, 2016.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv.org*, 08.04.2018.

[13] qqwweee, "qqwweee/keras-yolo3," GitHub, Inc. (US), 2018. [Online]. Źródło: https://github.com/qqwweee/keras-yolo3. [11.2018].

[14] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree and N. M. Rajpoot, "Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images," *IEEE Transactions on Medical Imaging*, vol. 5, 05 2016.

[15] Kaggle, "Human Protein Atlas Image Classification," 2018. [Online]. Źródło: https://www.kaggle.com/c/human-protein-atlas-image-classification. [27.05.2019].

[16] The American College of Radiology, "ACR® American College of Radiology," [Online]. Źródło: https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/BIRADS-Reference-Card.pdf. [27.05.2019].

[17] K. J. Geras and et al., "High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks".*arXiv.org, 2018.*

[18] K. J. Geras and et al., "Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening," *arXiv.org*, 2019.

[19] R. Sawyer Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, 2019.

Dodatki

Repozytorium projektu - https://github.com/michalkowalski94/MSADnn