

Segmentacja - badanie skuteczności wybranych metod wstępnego przetwarzania obrazu

Ignacy Stępka

3 października 2021

1 Wprowadzenie

Celem badania jest optymalizowanie jakości modelu za pomocą różnych metod przetwarzania wstępnego obrazu (preprocessing). W problemach opartych o analizę obrazu w dziedzinach mikroskopii i medycyny często mamy do czynienia z małą ilością danych. Za pomocą przetwarzania wstępnego jesteśmy w stanie między innymi uzyskać ze zdjęć konkretne przydatne informacje, normalizować i przekształcać rozkład kolorów, generować bardziej ogólne zestawy danych poprzez stosowanie augmentacji.

Problemem, na którym przeprowadzane były eksperymenty jest zadanie 1 z konkursu ISIC2018 dotyczące segmentacji mikroskopowych zdjęć nowotworowych zmian skórnych.

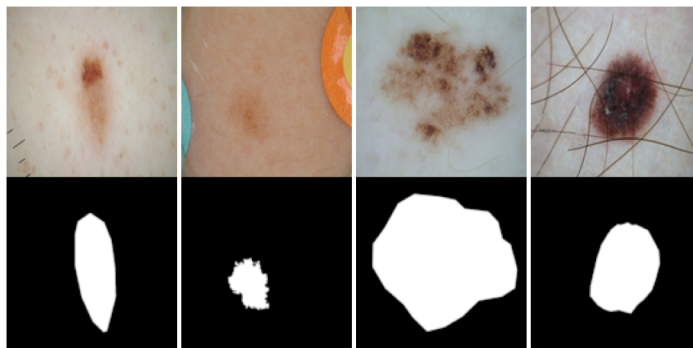
Po selekcji najlepiej działającego modelu bazowego, eksperymenty opierały się o badanie metod przetwarzania wstępnego. Warto dodać, że dobór hiperparametrów podlegał tylko powierzchownej optymalizacji z powodu skupienia badań na preprocessingu. Zatem istnieje szansa, że model mógłby osiągać lepsze wyniki przy dalszym rozwoju i optymalizowaniu swojej architektury oraz hiperparametrów.

Całość kodu, z którego wykorzystaniem generowane były doświadczenia dostępna jest pod:
<https://github.com/Iggyyy/Lesion-boundary-segmentation>

2 Dane

Dane zostały pobrane z oficjalnej strony konkursu ISIC2018. Jest to około 2500 mikroskopowych zdjęć przedstawiających zmiany skórne, które są charakterystyczne dla różnych rodzajów nowotworów. Do każdego zdjęcia dołączona jest odpowiadająca mu binarna maska, zaznaczająca pole, które jest wspomnianą zmianą skórną. Według, [1] zdjęcia są wynikiem ręcznie wycinanych i obrabianych próbek, których maski (ground truth) zostały wyznaczone przez specjalistów z tej dziedziny.

Na zdjęciach mamy do czynienia z różnymi problemami, które utrudniają prostą segmentację. Są nimi ślady markerów na skórze, włosy, efekt vignette wokół zdjęcia, okrągłe znaczniki kontrastowe, linijki, różne jasności, wielkości zbliżenia oraz kolory skóry.



Rysunek 1: Zdjęcia zmian skórnych i ich binarne maski.

3 Architektura

Architekturą modelu jest UNet [2] z uproszczoną funkcją energii, którą jest binarna entropia krzyżowa. UNet pozbawiony jest także wczytywania mapy wag odnoszącej się do podkreślania granic pomiędzy różnymi klasami komórek. Pominięcie tego czynnika jest spowodowane faktem, że problem segmentacji binarnej taką granicę ma tylko jedną.

Dodatkowo, dane wejściowe są przeskalowane do stałej rozdzielczości, którą jest 256x256. Obniżenie rozdzielczości z 572 pikseli z oryginalnej wersji UNet do 256 spowodowane było uproszczeniem złożoności obliczeniowej problemu, oraz tym, aby możliwe było zastosowanie wszystkich badanych metod przetwarzania wstępnego, a w szczególności tych bardzo kosztownych obliczeniowo jak np. ZCA.

Z powodu obniżenia ilości pikseli o około połowę, liczba kanałów na poszczególnych poziomach UNetu została również dwukrotnie obniżona w stosunku do tej w oryginalnym artykule, zatem ich liczbach na kolejnych poziomach modelu to 32,64,128,256,512.

Reszta architektury modelu UNet pozostaje niezmienna względem oryginalnej architektury zaproponowanej w [2]. Taka konfiguracja posiada 7.85 mln parametrów podlegających uczeniu.

Dane zostały podzielone losowo na 3 zbiory (treningowy, walidacyjny, testowy), których podział procentowy to kolejno 70,15,15.

4 Metryki i parametry

Model jak i wszystkie funkcje zbudowane były za pomocą biblioteki tensorflow i keras.

4.1 Metryki

Dla wszystkich metryk poza indeksem Jaccarda wartości predykcji były binaryzowane przez wartość progową wysokości 0.5.

- Accuracy

Stosunek poprawnie zaklasyfikowanych pikseli do liczby pikseli w zdjęciu. Wartość wyliczana binarnie pojedynczo dla każdego piksela.

- Jaccard index

$$J(A, B) = \frac{A \cap B}{A \cup B} \quad (1)$$

Miara podobieństwa pomiędzy dwoma zbiorami danych polegająca na wyliczeniu wartości części wspólnej zbiorów pikseli należących do interesującej nas klasy i podzieleniu przez ich sumę.

- AUC - pole pod krzywą

Metryka Area under Curve obliczana na podstawie ROC curve, czyli bazując na stosunku liczby pikseli true positive do false positive.

- Precision

Podobnie jak Accuracy stosowana jest wartość progowa, aby wyliczyć sumę binarnie poprawnie zaklasyfikowanych pikseli w stosunku do tych, które były błędnie pozytywnie zaklasyfikowane.

Wiodącą metryką, która zaproponowana została w opisie zadania na ISIC2018 jest indeks Jaccarda, który bardzo dobrze oddaje ogólną jakość modelu ponieważ bierze pod uwagę tylko regiony zainteresowania i na ich podstawie oblicza dokładność predykcji.

4.2 Parametry

Uczenie modelu oparte było o funkcję celu w postaci binarnej entropii krzyżowej optymalizowanej przez algorytm optymalizacyjny Adam (lr=0.001).

Dla wyznaczania optymalnej liczby iteracji algorytmu uczenia zastosowany został mechanizm EarlyStopping, który obserwował wartość funkcji straty na zbiorze walidacyjnym. Uczenie było zatrzymywane gdy podczas 15 kolejnych iteracji model nie był w stanie osiągnąć lepszego wyniku. Po przerwaniu model wracał do zapisanego punktu kontrolnego, w którym osiągnął dotychczasowo najniższy loss.

Znaczącymi parametrami są również: maksymalna liczba iteracji uczenia (epoch=100) oraz wielkość batch (batch-size=8).

5 Wyniki eksperymentalne

W badaniach brały udział następujące metody przetwarzania wstępnego: ‘Gaussian Blur’, ‘Per-channel mean normalization’, ‘Histogram Cutoff Equalization’, ‘Augumentacje’, ‘Connected Components’, ‘ZCA’. Wstępnie wszystkie techniki były porównywane za pomocą metryk ‘Test Accuracy’ oraz ‘Jaccard index’, na których podstawie projektowane były kolejne eksperymenty. ‘ZCA’ jak i ‘Connected Components’ okazały się bardzo mało efektywne, a ich stosowanie obniżało poziom wyników modelu poniżej poziomu ‘Baseline’.

Augumentacje oraz ZCA były aplikowane z wykorzystaniem generatora z tensorflow: ImageDataGenerator. Connected Components wyznaczane były za pomocą biblioteki opencv, a pozostałe metody zaimplementowane były przy pomocy Python Image Library (PIL).

5.1 Badane techniki

- Gaussian Blur - zastosowanie gausowskiego filtra 3x3 w celu rozmycia zdjęcia i zmniejszenia wpływu losowych artefaktów na przewidywaną maskę.
- Per-channel mean normalization - odejmowanie średniej jasności zdjęcia od każdego piksela, w wyniku czego średnia jasności całego zdjęcia zbliżona jest do zera.
- Histogram Cutoff Equalization - znane również jako Contrast Enchancing, które polega na przedstawieniu wszystkich jasności pikseli w postaci histogramu i odcięciu 2% skrajnych wartości i przeskalowaniu pozostałego wykres tak, aby spowrotem zajmował całą przestrzeń histogramu.
- Augumentacje - przekształcenia zdjęć aplikowane losowo przed każdą iteracją procesu uczenia. Dzięki tej technice, nawet przy małej ilości danych wejściowych, jesteśmy w stanie dobrze generalizować w celu uogólnienia uczenia modelu.

Zastosowanymi technikami augumentacji były:

- Rotacja, w dowolnym kierunku, maksymalnie do 30 stopni.
- Pionowe i poziome odbicie zdjęcia.
- Pionowe i poziome przesunięcie zdjęcia, maksymalnie do 20% szerokości oryginalnego zdjęcia.
- Zoom do 10% więcej niż przybliżenie oryginalnego zdjęcia .
- W wyniku powyższych przekształceń mogą się pojawić obszary nie zawierające żadnych pikseli z oryginalnego zdjęcia. Rozwiązaniem na takie przypadki jest wypełnianie pustych pól kolorem czarnym. Wybór takiego sposobu jest nie przypadkowy ponieważ jego efekt pokrywa się z efektem vignette, który znajduje się na części zdjęć źródłowych.

Podczas badań parametry dla wymienionych wyżej technik testowane były w różnych konfiguracjach i wielkościach nasilenia. (np. moc rotacji czy przesunięcie w augumentacji) Z wytycznych jakie otrzymano, badania odbywały się z naciskiem na eksplorację technik augumentacji.

5.2 Porównanie wyników badań

Badania wykonywane były na tym samym ‘seedzie’, a każda próbka jest wynikiem uśrednienia z trzech niezależnych prób uczenia. Wyniki oparte są o ewaluację na testowych zbiorach. Tabela 1 przedstawia otrzymane wyniki dla 2 tysięcy danych wejściowych i 250 danych testowych. Tabela 2 zestawia eksperymenty dla 1.6 tysiąca danych wejściowych i 600 danych testowych.

Name	Lesion melanoma ISIC2018		Bacteria detection	
	Accuracy	Jaccard index	Accuracy	Jaccard index
Bs	94.45 \pm 0.15	77.98 \pm 0.19	96.85 \pm 0.26	82.04 \pm 0.79
Bs+Aug	94.63 \pm 0.27	78.75 \pm 0.72	97.20 \pm 0.16	83.42 \pm 0.99
Bs+His	94.49 \pm 0.12	78.65 \pm 0.63	96.93 \pm 0.25	82.66 \pm 1.08
Bs+Pcn	94.53 \pm 0.03	78.73 \pm 0.19	97.01 \pm 0.23	83.05 \pm 1.06
Bs+Gaus	94.27 \pm 0.23	76.92 \pm 1.19	96.75 \pm 0.13	81.40 \pm 0.58
Bs+Aug+His	94.56 \pm 0.53	78.79 \pm 1.22	97.13 \pm 0.21	83.50 \pm 0.89
Bs+Aug+Pcn	94.67 \pm 0.21	78.77 \pm 1.10	97.18 \pm 0.13	83.41 \pm 0.89
Bs+Aug+Gaus	94.37 \pm 0.23	77.38 \pm 1.42	96.97 \pm 0.25	82.66 \pm 0.95
Bs+Aug+His+Pcn	94.82 \pm 0.42	79.12 \pm 1.00	97.27 \pm 0.13	83.61 \pm 0.99

Tabela 1: Zestawienie wyników dla podziału danych 70-15-15

Name	Lesion melanoma ISIC2018		Bacteria detection	
	Accuracy	Jaccard index	Accuracy	Jaccard index
Bs	92.79 \pm 0.82	71.33 \pm 1.35	97.11 \pm 0.14	80.35 \pm 0.86
Bs+Aug	93.88 \pm 0.14	75.18 \pm 0.61	97.25 \pm 0.04	81.39 \pm 0.22
Bs+His	93.57 \pm 0.18	74.52 \pm 0.40	97.20 \pm 0.13	81.08 \pm 0.84
Bs+Pcn	93.56 \pm 0.10	73.86 \pm 0.47	97.18 \pm 0.18	81.21 \pm 1.07
Bs+Gaus	92.56 \pm 0.41	70.62 \pm 1.55	96.82 \pm 0.22	79.25 \pm 1.12
Bs+Aug+His	94.16 \pm 0.18	76.18 \pm 0.52	97.28 \pm 0.19	81.90 \pm 1.02
Bs+Aug+Pcn	94.10 \pm 0.24	76.06 \pm 1.22	97.30 \pm 0.08	81.92 \pm 0.57
Bs+Aug+Gaus	93.61 \pm 0.54	74.30 \pm 1.65	96.96 \pm 0.15	79.69 \pm 0.69
Bs+Aug+His+Pcn	94.18 \pm 0.19	76.25 \pm 0.61	97.29 \pm 0.17	81.94 \pm 0.90

Tabela 2: Zestawienie wyników dla podziału danych 60-20-20

W przedstawionych tabelach można dobrze uwypuklić dwie zależności. Przede wszystkim, liczba danych, na których odbywa się uczenie ma kluczowe znaczenie pod kątem ogólnej jakości modelu. Zmniejszenie liczby danych wejściowych o 20% skutkuje wynikami niższymi o średnio 4 pp. W przypadku mniejszego zbioru danych stosowanie metod przetwarzania wstępnego w bardzo dużym stopniu wpływa na osiągane wyniki, a kluczową techniką dającą najwięcej zysku jest augmentacja. Dla większego zbioru danych te metody również przynoszą pozytywne efekty, jednak ich wpływ na poprawę modelu jest dużo niższy.

Na podstawie przeprowadzonych eksperymentów można zauważyć zależność, zastosowanie augmentacji zawsze poprawia wynik w stosunku do kombinacji, które jej nie zawierają. Dla zbioru ISIC2018 rozszerzenie uczenia o augmentację podwyższa sprawowanie modelu w problemie z 2 tys. danych o średnio 0.6% i w problemie z 1.6tys danych o średnio 3.9%. Obie te wartości w kontekście odległości od siebie wyników pomiędzy poszczególnymi modelami są bardzo znaczące, ponieważ odpowiadają za średnio ponad 50 procent całej zmiany.

Podobne zależności przedstawiają ‘Histogram Equalization’ oraz ‘Per-channel mean normalization’. Jediną różnicą w stosunku do augmentacji jest to, że ich wpływ na poprawę sprawowania modelu jest niższy, ale również znaczący.

6 Podsumowanie

Na podstawie przedstawionych badań, można wnioskować, że przetwarzanie wstępne obrazu dla problemu segmentacji mikroskopowych zdjęć poprawia sprawowanie modelu. Wyniki badań sugerują stosowanie augmentacji, która wykazuje się znaczną poprawą wyników mającą duże znaczenie w sprawowaniu się modeli, w szczególności gdy dostępnych danych wejściowych jest mało. Augmentacja znakomicie sprawdza się w przypadku zbiorów danych, które zawierają niedużą ilość pozycji, ponieważ pozwala znakomicie generalizować zdjęcia i zapobiegać przeuczaniu.

Techniki ‘Histogram Equalization’ i ‘Per-channel mean normalization’ w przeprowadzonych testach wykazują się poprawą wyniku względem ‘Baseline’, zatem również warto je stosować. Są to proste techniki wymagające niedużo obliczeń, a pozwalają nieznacznie podwyższyć uzyskiwane wyniki.

‘Connected Components’ prawdopodobnie nie działało dobrze z powodu specyfikacji zdjęć (np. efekt vignette, okrągłe znaczniki) gdzie nie dało się łatwo określić co jest tłem a co zmianą skórną, stąd kolejność dodawania odpowiednich kanałów do zdjęcia była niejednolita.

ZCA jest bardzo kosztowne pod kątem wykorzystania pamięci i czasu działania dla tak dużych zdjęć, stąd została ona odrzucona po wstępnych testach gdzie ta technika nie wykazała żadnej poprawy wyników predykcji względem podstawowego modelu.

Z racji na wysokość odchylenia standardowego jednoznaczne określenie najlepszego zestawu technik jest trudne, jednak z przeprowadzonych badań i uśrednionych wyników najlepszą kombinacją metod przetwarzania wstępnego niewątpliwie jest augmentacja poprzedzona ‘Histogram Equalization’ oraz ‘Per-channel mean normalization’.

Literatura

- [1] Philipp Tschandl. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data* vol. 5 180161., 2018.
- [2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [3] H.R. Tizhoosh Sara Ross-Howe. The effects of image pre- and post-processing, wavelet decomposition, and local binary patterns on u-nets for skin lesion segmentation. Kimia Lab, University of Waterloo, Waterloo, Ontario, Canada.