

System do przechowywania i rozpoznawania obrazów wideolaryngoskopowych krtani

Tworzenie klasyfikatora (etap drugi) - raport

Ada Kołodziejczak 193450, Jakub Bronowski 193208, Piotr Trybisz 193557, Igor Józefowicz 193257

Spis treści

Wstęp	1
Wykorzystywane biblioteki	1
Wykorzystane dane	1
Osiągnięte wyniki	1
Inne nowości w systemie	2
Podsumowanie	3

Wstęp

Celem poniższego raportu jest opisanie wyników działania klasyfikatora. Dokument ten stanowi jedynie o osiągniętych wynikach. Opisy techniczne systemu zawierają się w innych dokumentach. Ponadto, nie jest to jeszcze ostateczna wersja klasyfikatora. W przyszłości planowane jest dalsze trenowanie na podstawie (wciąż do nas wpływających) nowych danych, modelu w ramach dyplomu inżynierskiego.

Wykorzystywane biblioteki

W projekcie wykorzystano szereg bibliotek i narzędzi wspomagających przetwarzanie danych, budowę modelu głębokiego uczenia oraz ocenę jego wydajności. Do analizy danych i operacji pomocniczych użyto bibliotek takich jak *pandas*, *numpy* oraz *os*. Przetwarzanie i pobieranie obrazów wspomagały *cv2* oraz *requests*, natomiast do utrzymania powtarzalności wyników zastosowano mechanizmy ustalania ziarna losowości. Kluczową rolę w budowie i trenowaniu modelu odegrała biblioteka *TensorFlow* wraz z modułem *Keras*, w ramach której wykorzystano różne pretrenowane architektury jako bazę do transfer learningu. Ewaluację klasyfikatora przeprowadzono z użyciem metryk z biblioteki *sklearn*, co pozwoliło na dokładne określenie skuteczności predykcji.

Wykorzystane dane

Na moment pisania tego raportu, zbiór zawiera 859 obrazów krtani oraz 1943 oznaczeń. Ze względu na ciągłe zwiększanie się zbioru danych, klasyfikator jest wciąż w fazie rozwoju. W związku z tym, poniższe wyniki mogą ulegać zmianom w miarę dodawania nowych danych i dalszego trenowania modelu. Niestety nastąpiło pominięcie i ilość obrazów, które zostały wykorzystane do trenowania poszczególnych klasyfikatorów, nie zostało odnotowane. Ta informacja zostanie archiwizowana w przyszłości, gdyż jest ona istotna z perspektywy potencjału modeli. Część opisywanych wyników została osiągnięta na podstawie mniejszej liczby obrazów. Oznacza to, że te modele mają potencjał na dalsze poprawianie wyników, gdy zbiór danych zostanie rozszerzony. Ich trenowanie zostanie powtórzone. Dalsze prace nad klasyfikatorem będą kontynuowane w ramach dyplomu inżynierskiego.

Osiągnięte wyniki

W ramach przeprowadzonych eksperymentów porównano kilka architektur głębokich sieci neuronowych w zadaniu binarnej klasyfikacji obrazów. Poza poniżej wymienionymi modelami, testowano również inne architektury, takie jak **MobileNetV2**, **ResNet50**, jednak nie przyniosły one lepszych rezultatów niż te opisane poniżej.

Najlepsze rezultaty osiągnięto przy użyciu modelu **EfficientNetB0**, który uzyskał najwyższą dokładność (Acc = 71,9%) oraz najwyższą wartość średniego zbalansowanego F1-score (macro = 0,7191), przy stosunkowo niewielkiej liczbie parametrów (4,2 mln) i krótkim czasie treningu (ok. 235 s). Warto zaznaczyć, że model ten

utrzymał wysoką jakość predykcji zarówno dla klasy „healthy” ($F1 = 0.7222$), jak i „sick” ($F1 = 0.7159$), co świadczy o jego zrównoważonej wydajności.

Kolejne miejsca zajęły modele **DenseNet121** i **EfficientNetB1**, które osiągnęły bardzo zbliżone wartości dokładności ($Acc \approx 69,5\%$) oraz F1-score macro. Mimo większej liczby parametrów (6,95 mln dla B1 i 7,1 mln dla DenseNet121) i zauważalnie dłuższego czasu trenowania, nie udało się im wyraźnie poprawić wyników względem EfficientNetB0.

Model **DenseNet169**, mimo największej liczby parametrów (ponad 12,8 mln) i wydłużonego czasu treningu (1 170 s), uzyskał najslabsze rezultaty. Osiągnięta dokładność ($Acc = 56,7\%$) i niska wartość F1-score dla klasy „healthy” (0.3529) wskazują na istotne problemy z generalizacją i wyraźną nierównowagę predykcji między klasami. Podejrzewamy, że przyczyną tego stanu rzeczy może być zbyt duża liczba parametrów w stosunku do dostępnych danych treningowych, co prowadzi do przeuczenia modelu.

Poniższa tabela przedstawia najlepsze osiągnięte wyniki dla poszczególnych testowanych modeli:

Model	Acc.	F1 macro	F1 healthy	F1 sick	Param.	Res. (px)	Czas (s)	Uwagi
EfficientNetB0	0.7191	0.7191	0.7222	0.7159	4.2 mln	224×224	235	Najlepszy kompromis skuteczności i wydajności
EfficientNetB1	0.6966	0.6942	0.6667	0.7216	7.0 mln	240×240	1224	Większy model, ale słabszy niż B0
DenseNet121	0.6966	0.6965	0.7033	0.6897	7.1 mln	224×224	490	Podobny do B1, średni czas
DenseNet169	0.5674	0.5140	0.3529	0.6751	12.9 mln	224×224	1168	Ze względu na dużą ilość parametrów i nie odpowiednią ilość danych, model przeucza się
MobileNetV2	0.5800	0.5800	-	-	2.3 mln	224×224	180	Lekki i szybki, ale niska jakość
ResNet50	0.5200	-	-	-	23.5 mln	224×224	-	Duża pojemność, fatalne wyniki, możliwe przeuczenie

Inne nowości w systemie

Poza tworzeniem klasyfikatora, na prośbę interesariuszy w systemie wprowadzono również inne funkcjonalności. W celu poprawy doświadczenia użytkowników dodano możliwość ustawienia tagów w dowolnej kolejności, co umożliwia lepszą optymalizację pracy lekarzy. Dodatkowo, zmieniono sposób wyświetlania opisów tagów - zamiast wyświetlania ich pod nazwą tagu, teraz są one wyświetlane w formie dymków po najechaniu kursorem na nazwę tagu. Zmniejsza to rozmiar listy tagów, co ułatwia pracę lekarzy, eliminując konieczność przewijania listy w celu znalezienia odpowiedniego tagu.

Podsumowanie

Najlepsze wyniki klasyfikacji osiągnięto modelem EfficientNetB0. Pozostałe modele nie przewyższyły jego wyników. System jest wciąż rozwijany - dane są sukcesywnie rozszerzane, a dalsze prace nad modelem będą kontynuowane w ramach pracy dyplomowej. Wprowadzono też usprawnienia interfejsu, poprawiające wygodę użytkownika.