Łódź, 28.02.2023

Prof. dr hab. inż. **Krzysztof Ślot**
Instytut Informatyki Stosowanej
Politechnika Łódzka

Recenzja rozprawy doktorskiej Pani mgr inż.
Weroniki Gutfeter
pt.

Identyfikacja twarzy na podstawie obrazów wieloujęciowych z zastosowaniem głębokich sieci agregujących

1. Tematyka i struktura rozprawy

Przedstawiona do recenzji rozprawa doktorska dotyczy tematyki biometrycznego rozpoznawania twarzy, a szczególnym obszarem tej obszernej dziedziny, któremu poświęcone są prace Doktorantki jest problem optymalnego wykorzystania informacji zawartej w tzw. obrazach 'wieloujęciowych' - różniących się pozą zdjęciach tej osoby, będących w dyspozycji algorytmu analizy. Podjęty przez Doktorantkę wątek prac jest ważny i aktualny – poprawna identyfikacja lub weryfikacja twarzy ma ogromne znaczenie praktyczne, a istniejące podejścia do realizacji tego zadania w żaden usystematyzowany sposób nie próbują wykorzystać informacji zawartej w wieloelementowych zbiorach przykładów danej kategorii. Podjęty przez Doktorantkę problem jest zarazem nietrywialny, a jego rozwiązanie wymaga wykazania się znajomością zaawansowanych metod analizy danych, posiadaniem eksperckiego rozeznania w obszarze tematycznym biometrii twarzy oraz pomysłowością i kreatywnością. Obszar merytoryczny badań wymaganych do osiągnięcia zakładanych celów rozprawy lokuje się bez wątpienia w obrębie dyscypliny naukowej informatyka techniczna i telekomunikacja, w której prowadzony jest przewód Doktorantki.

Zawartość rozprawy odzwierciedla chronologię prac Doktorantki w podjętym obszarze. Obszerna jej część zawiera bardzo kompetentny, szeroki i wnikliwy przegląd rozmaitych podejść do problemu uwzględnienia wpływu pozy na poprawność rozpoznawania, uzupełniony prezentacją dobrze dobranego materiału eksperymentalnego (baz danych) stanowiącego podstawę dla weryfikacji efektów prac, a także wskazaniem metryk, używanych w ilościowej ocenie wyników. Aby uzyskać obiektywny pogląd co do istotności różnych zaproponowanych koncepcji, Doktorantka dokonuje doświadczalnej weryfikacji jakości wybranych metod, reprezentatywnych dla różnych nurtów prac, korzystając z dostępnych procedur lub dokonując własnych implementacji tych procedur. Metody te są uszeregowane według rosnącego poziomu złożoności, począwszy od

narzucających się prostych koncepcji na rozwiązanie postawionego zadania, do dużo bardziej skomplikowanych algorytmów. Ostatnia część pracy poświęcona jest prezentacji wkładu własnego Doktorantki, która przedstawia odpowiednie adaptacje metod opracowanych dla rozwiązywania innych kategorii problemów, uzyskując satysfakcjonujące rezultaty przeprowadzanych analiz. Metodyka prezentowanych prac nie budzi zastrzeżeń – proponowane koncepcje są dobrze uzasadnione, podlegają starannej weryfikacji eksperymentalnej, podsumowanej z wykorzystaniem właściwych miar ilościowych i zwieńczonych jakościową dyskusją otrzymanych rezultatów.

2. Cele i tezy rozprawy

Celem badań Doktorantki jest opracowanie strategii rozpoznawania biometrycznego, pozwalającej na redukcję wrażliwości analizy na potencjalnie dowolną pozę rejestrowanej twarzy poprzez optymalne wykorzystanie informacji zawartych w dostępnych zbiorach ujęć twarzy danej osoby. Powodem podjęcia przez Nią prac w rozważanym obszarze były względy praktyczne – sformułowanie odpowiedniej metody pozwoliłoby na zwiększenie poprawności rozpoznawania osób w systemach policyjnej analizy materiału pozyskanego z monitoringu przestrzeni publicznych, dla których to osób dysponuje się zestawem kilku ujęć twarzy, różniących się pozą (kątem obserwacji twarzy przez kamerę o osi optycznej skierowanej prostopadle do osi twarzy), określanych w pracy terminem ‘obrazów wieloujęciowych’. Konsekwencje osiągnięcia założonego celu wykraczają oczywiście poza wspomniane konkretne zastosowanie i stanowią rozwiązanie o ogólnej przydatności.

Doktorantka formułuje trzy tezy, wykazanie słuszności których określa zakres jej prac:

1. Możliwe jest zbudowanie systemu identyfikacji twarzy, który będzie uwzględniał informacje o zmienności przestrzennej twarzy i różnicy w wyglądzie w zależności od pozy bez konieczności zastosowania specjalizowanych urządzeń do akwizycji, jak skanery 3D. Można pokazać, że system, w którym wzorce twarzy są rejestrowane na podstawie zbioru obrazów wieloujęciowych będzie lepszy niż system oparty na zdjęciach frontalnych.

2. Zbiory wieloujęciowe obrazów twarzy wymagają dedykowanych rozwiązań, a składanie informacji z poszczególnych próbek można zrealizować za pomocą modułu agregującego zintegrowanego w strukturze sieci neuronowej w postaci tzw. modelu agregującego multi-view. System identyfikacji wykorzystujący modele multi-view powinien wykazać wskaźniki jakości identyfikacji nie gorsze niż system oparty o modele jednoujęciowe tzw. single-view

3. Modele agregujące stosowane dotąd do rozpoznawania obiektów 3D należących do zamkniętego słownika obiektów (takie jak MVCNN i RotationNet) można dostosować do zadania identyfikacji biometrycznej twarzy.

3. Merytoryczna ocena pracy

Przedstawiona rozprawa ma charakter eksperymentalny – Doktorantka formułuje heurystyczne hipotezy, które następnie, korzystając z adekwatnych zbiorów danych, poddaje weryfikacji eksperymentalnej. Punktem wyjścia dla prac Doktorantki jest ocena poprawności rozpoznawania twarzy w warunkach nienadzorowanej akwizycji obrazów, z użyciem najlepiej spisujących się obecnie algorytmów bazujących na głębokich sieciach konwolucyjncych i z zastosowaniem standardowej strategii uczenia, gdzie obrazy twarzy dla docelowych kategorii osób są używane w

procedurze dostrajania parametrów wstępnie wytrenowanej architektury. Jako narzędzie przeprowadzenia analizy stosuje pretrenowaną na bazie zdjęć VGGFace2 głęboką sieć neuronową ResNet50, powszechnie uznawaną za jedną z najlepszych architektur do klasyfikacji obrazów. Aby umożliwić klasyfikację próbek kategorii nieznanych algorytmowi w fazie uczenia, a więc, przygotować się do budowy algorytmu klasyfikacji działającego na zbiorze ‘otwartym’, Doktorantka przyjmuje jako podstawę klasyfikacji podobieństwo kosinusowe wektorów generowanych przez przedostatnią warstwę sieci ResNet. W efekcie przeprowadzonych, bardzo wartościowych w mojej opinii eksperymentów, Doktorantka wskazuje na istotne upośledzenie wyników analizy spowodowane zmianami pozy obserwowanej twarzy. Dodatkowo, Doktorantka wskazuje na możliwość osiągnięcia nieznacznej poprawy wyników poprzez włączenie do zbiorów treningowego i testowego obrazów wieloujęciowych, wykazując tym samym potrzebę podjęcia prac nad próbą optymalnego wykorzystania zawartych w nich informacji. Sieć wykorzystana przez Doktorantkę w celu oceny wpływu pozy na jakość rozpoznawania została przez Nią użyta w dalszych pracach jako model odniesienia dla analiz porównawczych oraz jako komponent proponowanych przez Nią, bardziej złożonych architektur, służący do ekstrakcji reprezentacji obrazu twarzy.

Sformułowanie własnych koncepcji rozpoznawania biometrycznego twarzy o zredukowanej wrażliwości na zmienność pozy zostało poprzedzone, jak wcześniej wspomniano, eksperymentalną oceną jakości działania wybranych, zaproponowanych w literaturze strategii wykorzystania materiału wieloujęciowego, uzupełnioną kilkoma autorskimi pomysłami i podsumowaną ciekawą dyskusją uzyskanych wyników. Efektem tych prac jest wskazanie najbardziej obiecującego kierunku badawczego – modyfikacji metod zakładających użycie głębokich modeli agregujących informacje pochodzące z wielu widoków twarzy tej samej osoby, który stał się głównym obszarem zainteresowania Doktorantki. W ostatniej części rozprawy Doktorantka kolejno opisuje swoje pomysły na adaptację trzech metod rozpoznawania, z których dwie: ‘wielowidokowa’ sieć konwolucyjna (Multi-View Convolutional Neural Network – MCNN) i sieć ‘rotacyjna’ (RotationNet), stanowiły architektury zaproponowane do rozpoznawania obiektów innych niż twarze, na podstawie informacji zawartej w kilku dostępnych widokach tych obiektów, zaś trzecia stanowi adaptację metody ‘transformerów’ opracowanej dla przetwarzania języka naturalnego.

3.1. Weryfikacja istniejących podejść do klasyfikacji zdjęć wieloujęciowych

Pierwszy grupa metod zmierzających do zwiększenia poprawności rozpoznawania twarzy o dowolnej pozie, której reprezentatywny przykład został wdrożony i zweryfikowany przez Doktorantkę, bazuje na ciekawej koncepcji wykorzystania posiadanego zbioru wieloujęciowego do budowy kompletnego modelu 3D, z którego następnie można generować dowolny widok referencyjny, dopasowany do pozy analizowanego zdjęcia. W efekcie przeprowadzenia szeregu wartościowych eksperymentów, Doktorantka negatywnie weryfikuje biometryczną przydatność tego podejścia, jednocześnie formułując szereg istotnych wskazówek, których uwzględnienie w konstrukcji algorytmu może poprawić jego atrakcyjność dla innych potencjalnych zastosowań.

Kolejną testowaną przez Nią koncepcją jest pomysł agregacji deskryptorów obrazów różnych widoków twarzy, zmierzającej do uzyskania reprezentacji integrującej wiedzę o niezależnym od kąta obserwacji wyglądzie twarzy. Doktorantka rozważa dwa zaproponowane podejścia do agregacji: dokonywaną na poziomie deskryptorów widoków i dokonywaną na poziomie wyników indywidualnych porównań widoków z nieznaną próbką. Doktorantka rozważa kilka prostych metod

agregacji (uśrednienie, mediana, a w przypadku fuzji dla wyznaczanych odległości – średnia, odległość minimalna/maksymalna, soft-min). Doktorantka podsumowuje swoje doświadczenia konkluzją o niewielkim lub żadnym zysku płynącym z przedstawionych metod agregacji dokonywanej w odniesieniu do próbek galeryjnych (tworzących bazę wiedzy o klasach). Ten wniosek nie wydaje się być niezgodny z intuicją – raczej hipoteza, że uśrednianie deskryptorów (podobnie jak selekcja wartości minimalnych lub maksymalnych dla odpowiadających sobie pozycji) lub wyników porównań powinno zwiększyć poprawność klasyfikacji, jest w moim odczuciu niepoprawna. Dobry deskryptor odzwierciedla wiele aspektów treści zawartej w danych wejściowych, ale założenie, że informacje o tych aspektach są ‘rozplątane’ i mają ciągłe i skupione rozkłady dla każdej z cech (wtedy proponowane agregacje mają sens), jest w moim odczuciu kompletnie nieuzasadnione. Oprócz agregacji dokonywanej po stronie próbek galeryjnych, Doktorantka sprawdza również efekty agregacji próbek zapytań (w przypadku nagrań z monitoringu, prawdopodobne jest posiadanie wielu ujęć nieznanej osoby, więc taki zabieg jest jak najbardziej zasadny). Tym razem, zgodnie z intuicją uzyskuje znaczącą poprawę wskaźników rozpoznawania, przy czym przeprowadzone testy dotyczyły jedynie jednego scenariusza, w którym klasa była reprezentowana przez reprezentację odpowiadającą zdjęciu frontalnemu, a zapytanie było uśrednionym wektorem dla różnych ujęć danej twarzy ze zbioru testowego. Wyniki eksperymentów wykorzystujących wiele ujęć osób ze zbioru testowego były bardzo obiecujące, szkoda, że nie przeprowadzono sprawdzeń dla innych scenariuszy – być może okazałoby się, że dokonywanie rozważanych przez Doktorantkę sposobów agregacji nie przynosi poprawy w porównaniu z klasyfikacją w schemacie ‘najbliższych sąsiadów’, gdzie najbliższej pary próbek poszukiwano by w zbiorze galeryjnym i zbiorze zapytań.

Ponieważ przedstawione przez Doktorantkę wątki porównań wektorów (bez lub z agregacją różnymi sposobami) to użycie prostych i znanych idei klasyfikacji minimalnoodległościowej, naturalnym rozszerzeniem listy scenariuszy klasyfikacji może być klasyfikacja w schemacie k-NN lub klasyfikacja metodą najbliższej średniej, ale z uwzględnieniem odległości Mahalanobisa, czyli oceną rozrzutów w obrębie klasy galeryjnej (aby zachować prostotę obliczeniową, z użyciem np. diagonalnej macierzy kowariancji). Doktorantka nie podejmuje jednak tego tropu – być może z uwagi na intuicyjnie bardziej obiecujące możliwości oferowane przez metody agregacji bazującej na treningu, oferowanej przez modele agregujące.

Modele agregujące (ich istotą jest integracja uczenia modułu agregacji z procesem treningu klasyfikatora) rozważone w pracy Doktorantki – MVCNN i RotationNet, zostały zidentyfikowane przez Nią jako najbardziej obiecująca ścieżka klasyfikacji obrazów wieloujęciowych. Znowu, jest to zgodne z intuicją – zastąpienie ‘ręcznie’ dobieranych reguł wyznaczania reprezentacji danych przez metody uczenia stoi u podstaw sukcesów uczenia głębokiego. Weryfikacja poprawności rozpoznawania obrazów wieloujęciowych, która wymagała od Doktorantki przygotowania odpowiedniego materiału eksperymentalnego, została przeprowadzona, podobnie jak wszystkie wcześniejsze eksperymenty, w sposób poprawny metodycznie i podsumowana sformułowaniem uprawnionych wniosków.

3.2. Prace Doktorantki w zakresie adaptacji architektur MVCNN i RotationNet

Osiągnięcia wskazane przez Doktorantkę jako oryginalne, własne koncepcje w obszarze biometrycznej analizy twarzy to głębokie modele agregujące, stanowiące modyfikacje oryginalnych koncepcji przetwarzania obrazów wieloujęciowych. Dwa pierwsze z nich, to architektury

obliczeniowe stanowiące modyfikacje sieci MVCNN i RotationNet, nazwane przez Doktorantkę odpowiednio: MVCNN-Sygnalityka i RotationNet-Sygnalityka. Pierwszym elementem nowości, który Doktorantka wprowadza do bazowych architektur, jest ich przystosowanie do radzenia sobie w warunkach klasyfikacji na zbiorach ‘otwartych’, czyli zawierających klasy, które nie były używane w treningu klasyfikatora. Taki scenariusz klasyfikacji warunkuje praktyczne znaczenie metody, dlatego też przydatność zaproponowanego rozwiązania jest oczywista, chociaż zastosowana przez Doktorantkę strategia osiągnięcia tego celu: oparcie klasyfikacji na podobieństwie wektorów referencyjnego i badanego, generowanego przez wybrane warstwy gęstej sieci, nie jest nowa [1]. Druga modyfikacja, wprowadzona do metody MVCNN, dotyczy zastąpienia agregacji deskryptorów różnych widoków, dokonywanej w metodzie oryginalnej przez wybór maksymalnych wartości odpowiadających sobie komponentów wektorów składowych (max-pooling), przez uśrednienie tych elementów (average pooling). Chociaż zaproponowana, alternatywna strategia daje lepsze efekty, brak jakiejkolwiek analizy uzasadniającej to podejście (z pełną świadomością, że takiej analizy być może nie da się przeprowadzić), obniża rangę zaproponowanego pomysłu. Wreszcie, trzecim elementem autorskiej modyfikacji algorytmu treningu jest wykorzystanie modułu ekstraktora cech obrazu z parametrami pretrenowanymi na bazie twarzy (Doktorantka pokazuje, że daje to lepsze wyniki niż inicjalizacja losowa lub inicjalizacja w wyniku treningu na bazie zawierających inne niż twarze kategorie obrazów). Wprowadzenie rozważanego ulepszenia wydaje się być wskazaniem pewnej dobrej praktyki niż istotnym wkładem do budowy algorytmów klasyfikacji. W odniesieniu do drugiego z modeli: RotationNet-Sygnalityka, Doktorantka proponuje dopuszczenie losowej kolejności prezentacji sieci widoków, co jest nowością, ale wiąże się ze zwiększeniem złożoności obliczeniowej procedury.

3.2. Wykorzystanie transformera w analizie obrazów wieloujęciowych

Najciekawszym i najbardziej wartościowym, z punktu widzenia oceny dorobku Doktorantki, fragmentem rozprawy jest jej ostatnia część, która prezentuje koncepcję wykorzystania modułu kodującego transformera jako narzędzia realizacji zadania klasyfikacji obrazów wieloujęciowych.

Sednem pomysłu Doktorantki jest dostrzeżenie analogii występujących między problem analizy obrazów wieloujęciowych, których istotą jest reprezentacja trójwymiarowego obiektu poprzez agregację informacji cząstkowych, zawartych w dwuwymiarowych rzutach, z problemem analizy języka naturalnego, gdzie treść zdania wynika z agregacji jego komponentów. W efekcie, dla rozwiązania zadania rozpoznawania twarzy decyduje się użyć wspomnianej wcześniej koncepcji transformera, którą adaptuje do specyfiki rozwiązywanego zadania. Adaptacja obejmuje określenie organizacji procesu przetwarzania, niezbędnego dla realizacji postawionego celu, jak również nowe propozycje rozwiązań szczegółowych. Jako architekturę zapewniającą generację dyskryminatywnej, zagregowanej informacji o wyglądzie twarzy Doktorantka wskazuje część kodującą algorytmu transformera. Jako efekt transformacji danych wejściowych przez proponowany algorytm, Doktorantka przyjmuje dwie alternatywne reprezentacje. Pierwsza (nazywana przez Doktorankę SygnaT-token), to używany również w oryginalnej koncepcji wynik transformacji ‘tokenu’ pomocniczego (określanego typowo jako ‘CLS’), uzyskiwany w wyniku liniowej agregacji informacji o przetwarzanych obrazach, poddanej następnie nieliniowemu przekształceniu w wielowarstwowej sieci gęstej. Druga, to autorski pomysł wykorzystania reprezentacji generowanych w wyniku transformacji tokenów odpowiadających elementom sekwencji wejściowej, które po uśrednieniu są również argumentem nieliniowej transformacji w sieci gęstej (podejście, nazywane przez Doktorantkę SygnaT-avg). Aby nauczyć zaproponowaną

architekturę poprawnej reprezentacji informacji o wyglądzie osoby, Doktorantka dokonuje wstępnego treningu inspirowanego algorytmem BERT, sterowanego kryteriami poprawności klasyfikacji i poprawności predykcji reprezentacji brakujących komponentów zbioru wieloujęciowego. W zaprezentowanych scenariuszach uczenia (wstępnego i zasadniczego), Doktorantka proponuje potraktowanie sekwencji testowej jako następnika sekwencji galeryjnej (oddzielonego standardowym tokenem wejściowym ‘SEP’).

Zaproponowana metoda rozpoznawania wieloujęciowych zdjęć twarzy z użyciem transformera stanowi najskuteczniejsze (ze wszystkich rozważanych przez Doktorantkę sposobów) rozwiązanie problemu, a osiągane efekty są zdecydowanie konkurencyjne względem istniejących standardów dziedziny. Na szczególne podkreślenie zasługuje skuteczność autorskiej propozycji reprezentacji wyniku (‘SygnaT-avg’), przewyższająca efekty zastosowania podejścia wykorzystującego mechanizm oryginalnej koncepcji.

3.3. Podsumowanie oceny merytorycznej prac Doktorantki

Niewątpliwą zaletą przedstawionych w rozprawie prac jest gruntowna weryfikacja stosowanych obecnie podejść do problemu klasyfikacji zdjęć wieloujęciowych. Implementacja metod, obfitość scenariuszy eksperymentalnych i obszerność podsumowania wyników stanowią bardzo wartościowy efekt prac Doktorantki. Kluczowym elementem w ocenie wartości prac z perspektywy wymagań przewodu jest jednak ocena nowatorskiego wkładu Doktorantki do dziedziny. O ile w odniesieniu do zaproponowanych przez Nią ulepszeń algorytmów MVCNN i MV-RotationNet analizy obrazów wieloujęciowych, jej wkład w mojej opinii nie jest znaczący, o tyle zdecydowanie wartościowym, oryginalnym efektem jej badań jest propozycja metody analizy bazująca na koncepcji transformera, pozwalająca na uzyskanie poprawności analizy przewyższającej istniejące podejścia, co stanowi osiągnięcie o istotnym znaczeniu naukowym i praktycznym.

Na zakończenie opinii, chciałbym podjąć wątek alternatywnej i nie podjętej przez Doktorantkę strategii uczenia (określanej przez Nią jako ‘uczenie metryczne’, choć w literaturze znanej raczej jako tzw. meta-learning). Schemat ‘meta-learningu’ wydaje się lepiej pasować do podjętego problemu (zadanie to dopasować próbkę galeryjną do próbki testowej), niż zastosowany przez Doktorantkę schemat uczenia modeli konkretnych kategorii i oczekiwanie, że wyuczona reprezentacja będzie dyskryminatywna dla próbek wcześniej nieznanych. W pracy zabrakło mi próby konfrontacji obydwu podejść: proponowanego przez Doktorantkę oraz meta-learningu, dedykowanego dla przypadku testowania na zbiorach otwartych. Echa problemu, który w mojej opinii jest konsekwencją wybranej przez Doktorantkę drogi, odzywają się w wynikach eksperymentu podsumowanego na rys. 4.2d (wraz z postępem uczenia maleje poprawność klasyfikacji na zbiorze otwartym). Nie jest to niezrozumiałe – sieć podczas treningu zaczyna powoli specjalizować się w rozpoznawaniu klas pochodzących z zamkniętego zbioru treningowego, co pogarsza poprawność rozpoznawania dla klas nieznanych. Mimo, że jako przyczynę problemu, Doktorantka wskazuje zbyt dużą liczbę modyfikowanych parametrów w stosunku do posiadanego zbioru przykładów i proponuje użycie typowej dla transferu wiedzy metody „zamrażania” części wag, co przynosi częściową poprawę, wydaje mi się, że głównym źródłem problemu jest wybrana strategia uczenia. Wydaje się, (co wymagałoby oczywiście sprawdzenia), że zaobserwowane zjawisko mogłoby nie mieć miejsca gdyby użyto schematu ‘meta-learningu’, koncentrującego się na wyszukiwaniu różnic i podobieństw między niewielkimi zbiorami galeryjnymi i zapytaniami.

4. Uwagi szczegółowe

Praca jest napisana starannie, ale znajduje się w niej pewna liczba fragmentów, które są niejasne i wymagają doprecyzowania, które są polemiczne, wreszcie, które są w mojej opinii niepoprawne. Poniżej prezentuję ich listę, precyzując w każdym przypadku rodzaj formułowanego zastrzeżenia.

str. 35: „Przez głębokie sieci neuronowe najczęściej rozumiemy perceptrony wielowarstwowe, …: nie rozumiem

str. 36: „Zestawienie ze sobą warstw splotowych i skalujących przyczynia się do niewrażliwości modelu na niewielkie przesunięcia obiektów w obrazie.” - wyjaśnienie wprowadza w błąd: warstwa konwolucyjna zapewnia inwariantność względem translacji. To o co chodziło Autorce to zapewne selekcja najlepszego dopasowania filtru do treści obrazu w regionie określonym przez rozmiar okna decymacji.

str. 36: „dobrze sprawdzają się przy przetwarzaniu danych o strukturze siatki ...” to jest żargon, nie ma czegoś takiego, są dane określone na dyskretnej przestrzeni 2D

str. 37: Po co zwrot ‘pewnego rodzaju’ w zdaniu „Sam algorytm uczenia opiera się na pewnego rodzaju optymalizacji.”?

str. 40: Doktorantka definiuje odległość kosinusową pomijając wyjaśnienia używanej notacji. Jak rozumiem, symbole Ya i Yb oznaczają wektory reprezentacji. Jeśli tak, to przedstawiony iloczyn to chyba iloczyn skalarny, ale jeden z argumentów powinien być tu transponowany. Co więcej, Doktorantka twierdzi, że wartość odległości będzie zawarta w przedziale [0..1], czego nie rozumiem: kosinus zmienia się w zakresie od -1 do 1, więc skąd ten przedział?

str.49: W nagłówku Tabeli 2.2 pojawia się niezdefiniowany akronim TAR, który prawdopodobnie powinien mieć postać TPR.

str. 56: nie rozumiem sposobu ‘grafowego’ uporządkowania metod redukcji wrażliwości klasyfikacji na zmiany pozy, przedstawionych na rys. 3.1 – czy poprzez zastosowanie połączeń Autorka chciała uwypuklić bliskość koncepcyjną par metod? Chyba nie, bo podane podejścia można by ułożyć w dość dowolnej kolejności.

str. 91. Doktorantka mogłaby pomóc w wyjaśnieniu istoty metody ‘RotationNet-Sygnalityka’ poprzez odpowiednie wsparcie opisu tekstowego dobrze wyjaśnionym materiałem graficznym. Pewnie rys. 4.2 ma taki potencjał, ale niestety, trzeba się mocno domyślać znaczenia użytych tam oznaczeń i stylów. Co oznaczają litery ‘z’ i litery ‘i’ z indeksami – czy to są wyniki predykcji dla hipotezy danego widoku?. Czy dwa zacienione w różny sposób prostokąty, stanowią symboliczną reprezentację wektorów hipotez uzyskanych dla różnych widoków, gdzie węższy prostokąt, odpowiadający symbolowi ‘i’ oznacza ‘incorrect view’?. Co podlega agregacji? - prawdopodobnie, różne widoki.

str. 95: niezrozumiałe zdanie: ‘Wybór sekwencji jest wynika z z minimalizacji wartości neuronów kodujących ujęcie”

str. 103 „oraz zestaw par wektorów kluczy K i wektorów wartości W w wektor wyjściowy.” Zamiast symbolu W powinien być symbol V.

str. 106: „W eksperymentach używano modułu MHA złożonego z ośmiu równoległych modułów atencji tzw. głów, a znaczy to, że każdy z tokenów wejściowych jest cięty przed wejściem do modułu na osiem równych części.” - to stwierdzenie jest zaskakujące. Token to reprezentacja elementu sekwencji, zaś istota ‘wielogłowości’ to próba znalezienia szeregu alternatywnych kontekstów dla danej sekwencji. ‘Cięcie’ tokenów na kawałki to rezygnacja z wielowymiarowej reprezentacji danych wejściowych. W pracach dotyczących transformerów tokeny wejściowe są podawane równolegle na wejścia każdej z przetwarzających głów, a to co jest decymowane w stopniu proporcjonalnym do liczby głów, to rozmiary wektorów V, z których składane są wektory stanu lub wektory kontekstu, generowane przez warstwy gęste (bo wynik jest konkatenacją wyjściowych wektorów kontekstu).

str. 108: „Dla sieci SygnaT (avg) wyniki są nominalnie wyższe, ale mieściły się w zakresie zmienności.” - co Doktorantka chciała przez to powiedzieć?

5. Wniosek końcowy

W podsumowaniu niniejszej recenzji chciałbym stwierdzić, że przedstawiona praca zawiera oryginalne i wartościowe koncepcje, stanowiące zauważalny wkład do dziedziny biometrycznego rozpoznawania twarzy z użyciem tzw. obrazów wieloujęciowych. W konsekwencji uważam, że rozprawa doktorska Pani magister inżynier Weroniki Gutfeter pt. „Identyfikacja twarzy na podstawie obrazów wieloujęciowych z zastosowaniem głębokich sieci agregujących” **spełnia**

wymagania określone w odnośnych przepisach i tym samym **wnioskuję o dopuszczenie Doktorantki do publicznej obrony**.

Signed by /
Podpisano przez:

Krzysztof Maciej
Ślot
Politechnika Łódzka

Date / Data:
2023-02-28 15:56

Literatura

[1] Wang M., Deng W., „Deep Face Recognition: A Survey, CoRR, vol. abs/1804.06655, 2018, http://arxiv.org/abs/1804.06655