Łódź, 30.08.2022

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

Recenzja rozprawy doktorskiej Pani mgr inż.

Eweliny Bartuzi-Trokielewicz

pt.

Presentation attack-resistant palm recognition for mobile devices in unconstrained conditions

1. Tematyka i cele rozprawy

Przedstawiona do recenzji rozprawa doktorska dotyczy tematyki biometrycznej ochrony dostępu do zasobów aplikacji działających na urządzeniach mobilnych, gdzie źródłem informacji biometrycznej są obrazy linii papilarnych spodu dłoni. Prace nad rozwojem metod rozpoznawania biometrycznego, które są dedykowane implementacjom w urządzeniach mobilnych, są niezwykle **aktualne** z uwagi na dążenie do zapewnienia wysokiego poziomu bezpieczeństwa korzystania z powszechnie dostępnych usług, wymagających uwierzytelniania użytkowników. Założenie nienadzorowanego kontekstu akwizycji danych biometrycznych stanowi oczywiste wyzwanie dla algorytmów analizy biometrycznej z uwagi na konieczność zastosowania skutecznych metod ekstrakcji informacji biometrycznej z obrazów o potencjalnie bogatej treści, nieoptymalnej prezentacji biometryk oraz możliwości dokonywania prób oszukania systemu analizy przez podstawienie spreparowanych danych, nazywane w literaturze ‘atakami prezentacji’. W konsekwencji, podjęty w rozprawie obszar tematyczny jest **trudny**, a sformułowanie oryginalnych metod pozwalających na uzyskanie poprawy procesu analizy biometrycznej może stanowić istotne osiągnięcie w obszarze dyscypliny naukowej **informatyka techniczna i telekomunikacja**, w której prowadzony jest przewód Doktorantki.

Wybór biometrii dłoni jako podstawy rozpoznawania osób stanowi uzasadnioną alternatywę dla częściej stosowanego podejścia, wykorzystującego biometrię twarzy. Na marginesie warto stwierdzić, że podjęte przez Doktorantkę prace mają w części charakter ogólniejszy niż sugerowany tytułem rozprawy, wykraczając poza obszar identyfikacji użytkowników urządzeń mobilnych – detekcja w obrazach obszarów zawierających linie papilarnych dłoni, a następnie, identyfikacja na tej podstawie osób, ma istotne znaczenie w kryminalistyce jako narzędzie automatyzacji analizy materiału dowodowego.

Celem prac Doktorantki było opracowanie funkcjonalnego, kompleksowego algorytmu analizy obrazów dłoni prezentowanych kamerze urządzenia mobilnego, obejmującego trzy komponenty: ekstrakcję obszaru

spodu dłoni, przeprowadzenie procedury rozpoznawania na bazie informacji zawartej w wydzielonym obszarze oraz zapewnienie dodatkowej detekcji prób ataków prezentacji (ang. Presentation Attack Detection – PAD). W obszarze każdego z wymienionych komponentów procedury, Doktorantka podjęła próby sformułowania oryginalnych metod pozwalających na uzyskanie lepszych wyników odpowiednich analiz: zwiększenia dokładności wydzielania obszaru zainteresowania, zwiększenia poprawności rozpoznawania oraz zwiększenia skuteczności detekcji ataków prezentacji.

2. Struktura i tezy rozprawy

Rozprawa została napisana w języku angielskim, który w większości oceniam za poprawny, umożliwiający bezproblemowe śledzenie przekazywanych treści. Kompozycja pracy nie budzi zastrzeżeń – po prezentacji kontekstu prac, identyfikacji istniejących wyzwań i problemów oraz określeniu głównego i szczegółowych celów podjętych prac, przedstawionej we Wprowadzeniu, Doktorantka w Rozdziale 2 obszernie omawia zbiory danych, używane w pracach jako podstawa do tworzenia i testowania algorytmów, a następnie przechodzi do prezentacji swoich pomysłów i wyników ich ewaluacji. W kolejnych trzech rozdziałach opisane są badania Doktorantki dotyczące trzech wymienionych wcześniej aspektów biometrycznej analizy obrazów dłoni:

- poszukiwania metod ekstrakcji obszaru dłoni z obrazów pozyskiwanych kamerą urządzenia mobilnego, bez wprowadzania ograniczeń na rodzaj tła
- poszukiwania metod reprezentacji tekstury linii papilarnych spodu dłoni, pozwalających na uzyskanie wysokiej poprawności rozpoznawania
- opracowania skutecznych metod detekcji ataków prezentacji, a więc ataków na procedurę przedstawiania systemowi analizy biometrycznej spreparowanych danych wejściowych

Kompozycja każdego z modułów nie budzi zastrzeżeń – Doktorantka poprzedza prezentację swoich pomysłów przeglądem stanu wiedzy w podejmowanym obszarze tematycznym, uzupełnionym oceną wybranych, uznanych przez Nią jako reprezentatywne metod referencyjnych, po czym próbuje (ze zróżnicowanym stopniem szczegółowości) wyjaśnić istotę oraz szczegóły proponowanych przez nią metod, poddawanych w ostatniej części ewaluacji i konfrontacji z istniejącymi podejściami.

Doktorantka formułuje cztery stwierdzenia (‘statements’), wykazanie słuszności których jest przedmiotem jej prac:

1. Palm recognition in unconstrained conditions is more difficult than in laboratory, and the recognition accuracy of existing algorithms is insufficient.
2. The proposed method of hand segmentation allows for generating precise, consistent binary mask predictions for different conditions of data acquisition.
3. The proposed palmprint feature extraction method, based on the Siamese neural network, increases the accuracy of recognition in unconstrained conditions.

4. Presentation attack detection method, which was proposed in this doctoral study is resistant to common types of presentation attack instruments.

3. Merytoryczna ocena pracy

Prace Doktorantki mają bardzo silny wymiar praktyczny: opracowane przez Nią procedury są z powodzeniem implementowane i dają bardzo dobre wyniki w konfrontacji z rzeczywistymi danymi pozyskiwanym z użyciem urządzeń mobilnych, co bez wątpienia zasługuje na uznanie. Jednakże, w odniesieniu do wartości naukowej przedstawianych przez Nią koncepcji, moja opinia jest zróżnicowana – część z zaprezentowanych przez Doktorantkę pomysłów jest oryginalna i wartościowa, natomiast część nie stanowi w mojej opinii zauważalnego wkładu do dziedziny i wymaga przeprowadzenia dodatkowych, obszernych prac w celu potwierdzenia słuszności formułowanych tez. Struktura dalszej części oceny odpowiada trzem tematycznie odrębnym wątkom przedstawionym w rozprawie, wyrażonym postawionymi przez Doktorantkę tezami.

3.1. Segmentacja obszaru dłoni

Prace Doktorantki ukierunkowane na wykazanie słuszności drugiej z przedstawionych tez rozprawy, poświęcone poszukiwaniu nowych metod ekstrakcji obszaru dłoni w obrazach rejestrowanych kamerą urządzeń mobilnych, a więc w obrazach charakteryzujących się dowolnie skomplikowanym tłem, zostały zamieszczone w Rozdziale 3 rozprawy. Autorka rozpoczyna prezentację treści przeglądem podstawowych metod stosowanych w segmentacji obrazów: punktowych - bazujących na analizie jasności lub koloru oraz obszarowych – wykorzystujących deskryptory statystyczne tekstury obrazu. W przedstawionym przeglądzie, pojawia się jedynie wzmianka o segmentacji semantycznej, dokonywanej z użyciem głębokich sieci neuronowych – Autorka zamieszcza referencje do trzech pozycji [39-41] opisujących wykorzystanie sieci konwolucyjnych do segmentacji obszaru dłoni. Zaprezentowany przegląd stanu wiedzy w obszarze segmentacji semantycznej wykorzystującej sieci głębokie nie jest kompletny i pomija publikacje prezentujące podobną jak przyjęta przez Nią metodykę analizy (np. wykorzystanie architektury RefineNet do segmentacji obrazów dłoni fotografowanych na dowolnym tle bez narzucania ograniczeń co do sposobu prezentacji dłoni [1], czy sieci U-Net do realizacji tego samego zadania [2]). Co więcej, segmentacja obszaru dłoni nie jest wcale jedynym pomysłem na precyzyjne określenie regionu zainteresowania, niezbędnego dla potrzeb biometrycznej analizy linii papilarnych. Doskonale rokującą alternatywą dla realizacji tego zadania jest wykorzystanie metod bazujących na identyfikacji lokalnych punktów charakterystycznych. Sztandarowym przykładem skutecznej detekcji punktów charakterystycznych dłoni, pozwalające również na wyznaczenie obszaru spodu dłoni dla potrzeb analizy biometrycznej jest, wykorzystująca głębokie klasyfikatory neuronowe, metoda zaimplementowana w powszechnie stosowanymi obecnie pakiecie MediaPipe [3]. Pominięcie alternatywnych pomysłów na realizację kompletnej ścieżki rozpoznawania biometrycznego, bazującego na analizie linii papilarnych spodu dłoni, zubaża tło prezentowanych rozważań.

Istotą podejścia zaproponowanego przez Doktorantkę do realizacji zadania ekstrakcji obszaru dłoni z obrazów jest półautomatyczna procedura przygotowania zbioru danych uczących, w oparciu o które trenowana ma być docelowa, głęboka sieć neuronowa o architekturze adekwatnej dla zadania segmentacji semantycznej. Celem Doktorantki jest uzyskanie zbioru dokładnych masek, stanowiących oczekiwane

wyniki segmentacji. Dla osiągnięcia tego celu wykorzystuje wyniki segmentacji obrazów uzyskiwane przez cztery różne pretrenowane głębokie sieci neuronowe (FCN, DeepLab, SegNet i U-Net), z których wybiera subiektywnie najlepsze jako materiał do ponownego treningu wybranych sieci 'bazowych' (DeepLabv3 i SegNet). Następnie, nauczony na 'dobrych' danych klasyfikator jest aplikowany do segmentacji obrazów, które wcześniej sprawiały problemy, a uzyskane wyniki są poddawane manualnym korektom, z wykorzystaniem napisanego przez Doktorantkę narzędzia do edycji obrazów. Prawdopodobnie, jest to robione po to, by wytworzyć dodatkowy podzbiór treningowy obrazów 'trudnych', na którym douczane są docelowe klasyfikatory, ale w tekście nie znalazłem takiej informacji, więc jest to tylko moja spekulacja. Przedstawiona przez Doktorantkę procedura pozwala na zwiększenie skuteczności uczenia poprzez zwiększenie liczebności zbioru treningowego, chociaż osiągnięcie takiego rezultatu wymaga zaangażowania eksperta w procesie subiektywnej oceny poprawności segmentacji oraz korygowania błędów segmentacji.

Ocena przydatności zaproponowanego przez Doktorantkę podejścia jest dokonywana w drodze weryfikacji eksperymentalnej i konfrontacji z konkurencyjnymi podejściami. Jako referencję dla opracowanej przez siebie metody Doktorantka przyjmuje trzy metody segmentacji punktowej i czwartą, stanowiącą kombinację metody punktowej i obszarowej, używającej cech lokalnych wyznaczonych na podstawie wyników filtracji filtrami Gabora. Niestety, wbrew temu co twierdzi Doktorantka, żadna z tych metod nie zasługuje na miano metody 'state of the art', więc sformułowane na podstawie porównania efektów segmentacji, mocno tryumfalne wnioski o supremacji przedstawionej przez Nią metody nad aktualnie istniejącymi rozwiązaniami są nieuprawnione. Najskuteczniejsze obecnie metody ekstrakcji dłoni w warunkach dowolności tła, dowolności ułożenia dłoni i dowolności warunków akwizycji, wykorzystują głębokie architektury do segmentacji semantycznej, podobne do używanych przez Doktorantkę, jako komponenty składowe przeprowadzanej przez Nią procedury. Przekonywującym potwierdzeniem korzyści płynących z zastosowania Jej procedury byłaby konfrontacja uzyskanych za jej pomocą wyników segmentacji z wynikami osiąganymi przez te elementarne architektury segmentacji semantycznej, a nie z wynikami uzyskiwanymi dla metod segmentacji punktowej i obszarowej. Niestety nie wiadomo, jak można by powiązać ze sobą informacje zawarte w Tabeli 3.2 (segmentacja w sieciach składowych, trenowanych bez dodatkowych zabiegów) i Tabeli 3.4 (segmentacje w sieci wytrenowanej metodą zaproponowaną przez Doktorantkę).

Doktorantka do oceny wyników eksperymentów używa, obok powszechnie stosowanej metryki ilościowej 'IoU', przede wszystkim subiektywnych miar 'poprawności segmentacji', bazujących na opinii eksperta. Taki sposób podsumowania wyników daje jedynie poglądowe oszacowanie efektów analizy i nie może stanowić obiektywnej podstawy do formułowania zbyt daleko idących wniosków. Zamiast stosowania subiektywnej miary porównawczej jakości wydzielania obszaru dłoni, Doktorantka mogła dokonać pośredniej, ale obiektywnej oceny efektów segmentacji, porównując wyniki pełnej biometrycznej analizy obrazu. Możliwym scenariuszem takiego postępowania byłoby przeprowadzenie najpierw segmentacji obrazu dłoni za pomocą metody Doktorantki i metod referencyjnych, a następnie, zastosowanie identycznego algorytmu dalszej analizy, obejmującego wybór obszaru zainteresowania i klasyfikację (a więc metodykę opisaną w Rozdziale 4). Jeżeli poprawna segmentacja dłoni jest rzeczywiście istotnym czynnikiem dla powodzenia biometrycznej analizy dłoni, uzyskiwane wyniki klasyfikacji dawałyby ilościową podstawę oceny korzyści użycia Jej podejścia w porównaniu do stosowania prostych metod segmentacji semantycznej.

Doktorantka proponuje wykorzystanie zastosowanego przez siebie schematu budowy zbioru treningowego do realizacji zadań segmentacji obrazów dwóch innych modalności biometrycznych: wydzielania tęczówki oka i wydzielania obszaru palca. Niestety, wątki te są przedstawione w pracy w sposób niezwykle skąpy, co z jednej strony jest zrozumiałe, bowiem odbiegają od tematyki biometrycznej analizy dłoni, z drugiej jednak strony, wprowadzają sporo zamieszania i niejasności. Przykładowo, zupełnie nie rozumiem informacji zawartej w Tabelach 3.4 i 3.5, odnoszącej się do 'trzeciego' scenariusza eksperymentalnego (zgodnie z wyjaśnieniem, ten scenariusz to 'rozszerzenie' wariantu 'drugiego' o dwie dodatkowe modalności biometryczne, co już wymaga szerszego komentarza, bo nie bardzo wiem, jak segmentacja tęczówki może być utrudniana przez obecność '47 kategorii tekstur tła'). Co oznaczają dane przedstawione w ostatniej kolumnie obydwu Tabel? Czy zaproponowana przez Doktorantkę procedura segmentacji obrazów dłoni działa lepiej, gdy sieć trenuje się również na obrazach oka? Taki wniosek wydaje się być absurdalny, ale jest on uprawniony na podstawie przedstawionego tekstu. Czy może wyniki z ostatniej kolumny to jakaś agregacja różnych eksperymentów, więc nie można ich porównywać z danymi z wcześniejszych kolumn?

Podsumowując ocenę merytoryczną pierwszego z zaprezentowanych w rozprawie wątków, chcę stwierdzić, że Doktorantka sformułowała oryginalną, półautomatyczną koncepcję zwiększenia poprawności realizacji segmentacji semantycznej. Szkoda, że ewaluacja metody została dokonana w sposób niekompletny, co utrudnia ocenę jej znaczenia z punktu widzenia wkładu do rozważanej dyscypliny naukowej.

3.2. Identyfikacja biometryczna na podstawie linii papilarnych dłoni

Zrealizowane przez Doktorantkę prace nad biometryczną analizą linii papilarnych dłoni, opisane w Rozdziale 4 rozprawy, stanowią najciekawszy w mojej opinii fragment pracy, odnoszący się przede wszystkim do trzeciej ze sformułowanych we wstępie pracy tez (teza pierwsza zawiera dość oczywiste przypuszczenie o zwiększeniu stopnia trudności rozpoznawania w warunkach braku możliwości narzucenia ograniczeń na proces akwizycji obrazu). Doktorantka prezentuje dwuelementową procedurę, której pierwszym etapem jest autorska metoda ekstrakcji obszaru zainteresowania, a drugim – wyznaczenie deskryptora tego obszaru i jego klasyfikacja. O ile zaproponowana przez Doktorantkę procedura ekstrakcji obszaru analizy, mimo wykazanej w eksperymentach skuteczności, ma czysto inżynierski charakter, o tyle zaproponowana, oryginalna metoda klasyfikacji udowadnia Jej doskonałą orientację w obszarze zaawansowanych metod uczenia maszynowego, kreatywność i intuicję badawczą.

Doktorantka trafnie identyfikuje dwa podstawowe źródła problemów rozpoznawania linii papilarnych dłoni: nieliniowe deformacje treści obrazu (zależną od napięcia mięśni, zmienną strukturę wewnętrznej powierzchni dłoni) oraz niewielką liczbę dysponowanych dla klas przykładów, stanowiącą fundamentalny problem dla opracowania skutecznych metod uczenia głębokiego. Aby zmierzyć się z pierwszym problemem, Doktorantka decyduje się na wybór metody analizy, pozwalającej na zapewnienie elastyczności wyboru obszarów istotnych dla podejmowania decyzji, oferowanej w przypadku głębokich sieci neuronowych przez mechanizm skupiania uwagi ('attention'). Możliwość kontekstowej dywersyfikacji znaczenia różnych fragmentów informacji podawanej na wejście sieci stanowi, jak wykazuje Doktorantka, skuteczny pomysł radzenia sobie z analizą struktur podlegających nieliniowym deformacjom. W przypadku drugiego z przedstawionych problemów – praktycznego braku możliwości 'wyuczenia' standardowego klasyfikatora głębokiego w obliczu posiadania skrajnie ograniczonych liczebności danych reprezentujących dane klasy, uwaga Doktorantki skupia się na architekturze syjamskiej, stanowiącej rozwiązanie dedykowane dla budowy dyskryminatywnej reprezentacji klas w

warunkach rozważanego ograniczenia liczby dostępnych przykładów. Efektem obydwu pomysłów jest zaproponowana przez Doktorantkę architektura głęboka, trenowana z użyciem dwuskładnikowej funkcji celu, maksymalizującej, poprzez zastosowanie schematu uczenia sieci syjamskiej na parach próbek zgodnych i różnych, dyskryminatywność reprezentacji obszaru zainteresowania i jednocześnie wymuszającej dla przetwarzanych par obrazów zgodność przestrzennego rozkładu informacji uznawanej za kluczową w podejmowaniu decyzji.

Przeprowadzona przez Doktorantkę weryfikacja eksperymentalna metody nie budzi zastrzeżeń, a uzyskane wyniki konfrontowane są z rezultatami rozpoznawania uzyskiwanymi za pomocą dobrze zidentyfikowanych metod referencyjnych, stanowiących aktualnie najbardziej skuteczne podejścia do realizacji rozważanego problemu.

Przedstawiona przez Doktorantkę prezentacja prac nad rozpoznawaniem linii papilarnych jest w większości przejrzysta i ciekawa, niestety, zawiera również liczne niedopowiedzenia. Doktorantka pozostawia domyślności czytelnika ogólną strukturę algorytmu identyfikacji dłoni: pisze o tym, jak wytrenować model wyznaczający dyskryminatywną reprezentację linii papilarnych dłoni, jednak nie informuje, jak przeprowadza analizę w wytrenowanej sieci. Jeżeli dobrze rozumiem, podstawą klasyfikacji obrazu jest wektor cech wyznaczany przez wytrenowaną sieć, który jest następnie poddawany porównaniu (nie wiadomo według jakiej zasady) z wektorami reprezentującymi klasy (nie wiadomo jak określonymi – czy z każdym z wektorów zbioru treningowego, czy klasa ma jednego reprezentanta). Podobnie, bardzo wartościowa analiza możliwości redukcji rozmiaru obrazów, jakie mają być poddawane analizie, pozwalająca na uproszczenie architektury sieci używanej do realizacji zadania, jest komunikowana w zdawkowy sposób. Doktorantka jako kryterium używa współczynnika korelacji wzajemnej, szkoda że nie podaje precyzyjnie, jak go definiuje (obrazy mają różne rozmiary). Co więcej, jeżeli celem analizy jest zachowanie jak największej ilości informacji oryginalnej, dlaczego nie używa jako kryterium informacji wzajemnej? Wreszcie, na podsumowującym analizy rysunku 4.6 Doktorantka nie uznaje za stosowne podać, w jakich jednostkach mierzy rozmiar obrazu (jeśli są to piksele, to oznacza, że mozaika o rozmiarze 7 x 7 punktów zapewnia informację wystarczającą dla przeprowadzenia klasyfikacji dla obrazów z dwóch baz danych, w co szczerze wątpię).

Podsumowując prace Doktorantki w obszarze poszukiwania nowych metod biometrycznej identyfikacji linii papilarnych spodu dłoni chcę jednoznacznie stwierdzić, że są one wartościowe i oryginalne, stanowiąc w mojej opinii zauważalny wkład naukowy do rozważanej dziedziny.

3.3. Detekcja ataków prezentacji

Przedstawione w Rozdziale 5 efekty prac Doktorantki w obszarze detekcji ataków prezentacji na biometrię linii papilarnych spodu dłoni uważam za słabszą część rozprawy, z uwagi na niezwykle zdawkową komunikację prezentowanych treści oraz powierzchowność merytorycznej analizy podnoszonych problemów i proponowanych rozwiązań.

Ogólna struktura prezentacji materiału jest przejrzysta – Doktorantka najpierw przedstawia problem, kompetentnie informuje o przyjętej taksonomii ataków prezentacji i stanie wiedzy w zakresie detekcji ataków prezentacji, następnie omawia rozważone przez Nią metody ataków oraz sposoby ochrony przed atakami, podsumowując tekst wynikami weryfikacji przeprowadzonych eksperymentów. Niestety, na

każdym z etapów prezentacji, pojawia się deficyt wyjaśnień szczegółowych, utrudniający lub uniemożliwiający śledzenie przedstawianych treści.

Pierwszym obszarem prac Doktorantki była generacja syntetycznych obrazów dłoni, które mają stanowić najbardziej ‘wyrafinowaną’ formę ataków prezentacji na system biometrycznej analizy dłoni. Niestety, nie rozumiem (nie zostało to wyjaśnione) dlaczego wygenerowane sztucznie obrazy dłoni z fałszywą, transferowaną teksturą linii papilarnych, miałyby stanowić większe wyzwanie dla algorytmu detekcji ataku prezentacji niż zdjęcie (aby utworzyć obraz syntetyczny według metodyki przyjętej przez Doktorantkę, konieczne jest posiadanie zdjęcia ‘atakowanej’ dłoni). Ponieważ prezentacja wygenerowanego obrazu jest dokonywana z użyciem wyświetlacza czy monitora, wydaje się że skuteczność proponowanego ataku, nie powinna różnić się od ataku przedstawienia zdjęcia dłoni.

W swoich pracach Doktorantka wyróżnia trzy metody generacji sztucznych obrazów – dwie z nich są gotowymi implementacjami, pozyskanymi z Internetu, zaś trzecią wskazuje Ona jako metodę własną. Niestety, nie wyjaśnia na czym polega autorski wymiar metody – opis przedstawiony w części 5.3.3 nie prezentuje żadnego oryginalnego pomysłu, a jedynie przytacza dwa wzory i ilustracje pochodzące ze źródeł. Co więcej, w prezentowanym opisie jest wiele nieścisłości. Wzór (5.1), wbrew zapowiedzi: ‘The content of the image ... can be described by the formula:’ nie kwantyfikuje ‘zawartości obrazu’, ale wyraża miarę błędu oceniającą różnicę między wynikami przetwarzania obrazów: oryginalnego i wygenerowanego, dla określonej warstwy sieci. Doktorantka przytacza wyrażenie (5.1) jako komponent funkcji celu stosowany do oceny podobieństwa treści, tymczasem w materiale źródłowym jest on wskazany jako komponent używany do oceny rekonstrukcji komponentu ‘stylu’ (liczba ‘4’ w mianowniku wzoru (5.1) wynika z iloczynowego charakteru komponentów budujących wyrażenie oceniające styl).

W części dotyczącej detekcji ataków prezentacji, Doktorantka informuje o dwóch zastosowanych przez Nią, różnych podejściach: analizie statystycznej lokalnej tekstury obrazu z użyciem trzech różnych deskryptorów (LBP, BSIF i częstotliwościowej) oraz analizie z użyciem klasyfikatorów konwolucyjnych. Moim podstawowym zastrzeżeniem wobec przedstawionej przez Doktorantkę metodyki detekcji ataków prezentacji jest całkowite pominięcie kwestii kompresji obrazów i ich ewentualnego wpływu na uzyskiwane wyniki detekcji. Kompresja wprowadza do mikrostruktury obrazu powtarzalne artefakty, które być może, stanowią kluczowy czynnik w podejmowaniu decyzji o prawdziwości lub fałszywości prezentowanego obrazu. Dlatego też, przeprowadzając ataki prezentacji, należałoby co najmniej podzielić dysponowane obrazy na dwie kategorie: poddawanych kompresji stratnej (najpowszechniejsza to oczywiście format ‘jpg’), należących do wskazanej przez Doktorantkę kategorii najmniej zaawansowanych sposobów ataku, oraz niepoddawanych kompresji lub poddawanych kompresji bezstratnej, co odpowiada bardziej zaawansowanemu atakowi. Pomijając wątek wpływu kompresji, Doktorantka naraża się na wyciąganie z przeprowadzanych przez siebie eksperymentów fałszywych wniosków – na przykład, o wysokiej skuteczności ataków przeprowadzanych z użyciem obrazów generowanych metodą transferu stylu, wynikających być może z faktu braku artefaktów kompresji metodą jpg.

Przedstawiona przez Doktorantkę weryfikacja eksperymentalna sprawdzanych przez Nią metod detekcji mogłaby być obszerniejsza – jako miarę oceny przedstawia współczynnik fałszywego odrzucenia próbki dla jednego (wybranego arbitralnie?) poziomu współczynnika fałszywej akceptacji (10%). Bazując na

uzyskanych wynikach, przedstawia w Podsumowaniu pracy zupełnie nieuprawnioną, w świetle przytoczonych rezultatów, informację o osiągnięciu za pomocą swojej metody poprawności detekcji ataku prezentacji na poziomie 99%. Nie mam pojęcia, jak Doktorantka policzyła ten wskaźnik, chyba że dla przedstawionego przypadku współczynnik fałszywej akceptacji był bliski zeru (dlaczego więc nie przedstawiła ilościowego podsumowania tak znakomitych efektów w postaci macierzy pomyłek?).

Dokonania Doktorantki w obszarze rozwoju metod detekcji ataku prezentacji, mają w mojej ocenie, charakter implementacyjny – z powodzeniem wdraża metody generacji skomplikowanych treści oraz różne algorytmy binarnej klasyfikacji obrazów. W przedstawionym materiale nie znajduję jednak wystarczająco jasno opisanych pomysłów, które mógłbym uznać za naukowo znaczące z perspektywy rozwoju metod biometrycznej analizy danych.

4. Uwagi dodatkowe

W przedstawionej rozprawie znajduje się niewielka liczba dostrzeżonych przeze mnie usterek technicznych. Pierwsza z nich pojawia się w zdaniu definiującym cel pracy:

The aim of this doctoral study is to investigate factors that may influence existing palmprint recognition algorithms and to propose methods taking these changes into consideration and enabling correct recognition, including presentation attack detection.

gdzie prawdopodobnie słowo ‘factors’ zostało omyłkowo zastąpione słowem ‘changes’. W prezentacji stanu wiedzy Doktorantka wprowadza taksonomię metod segmentacji, wyróżniając dwie kategorie metod ‘klasycznych’: ‘bazujące na progu’ i ‘bazujące na teksturze’, co jest niepotrzebną próbą redefinicji istniejącej od dawna, powszechnie przyjętej systematyzacji podejść do segmentacji na metody ‘punktowe’ i ‘obszarowe’ (‘point-wise, region-wise’).

W prezentowanych przez Doktorantkę wzorach 3.2-3.4 brakuje indeksu G przy wartości średniej. W opisie metody segmentacji semantycznej: *„fully convolutional network (FCN) [27], that uses blocks of convolution and maxpooling layers to first decompress an image to 1/32th”*, Doktorantka omyłkowo używa słowa decompress zamiast compress. Ostatnie dwa akapity podsumowania Rozdziału 3 są zupełnie niepotrzebne – pierwszy powiela informacje ze wstępu, drugi – powiela informacje z wcześniejszej części podsumowania. Zamieszczona na stronie 49 referencja [53] jest wskazana jako źródło informacji o metodzie ekstrakcji obszaru zainteresowania dłoni, tymczasem dotyczy rozpoznawania tęczówki. Wreszcie, wydaje się, że wektory Mm* na rys. 4.9 są przedstawione w odwrotnej orientacji (rozumiem, że wartości przypisane wierszowi to maksymalna wartość mapy ‘atencji’, co jest sprzeczne z zawartością rysunku).

5. Wniosek końcowy

W podsumowaniu niniejszej recenzji chciałbym stwierdzić, że przedstawiona praca zawiera oryginalne i wartościowe koncepcje, stanowiące zauważalny wkład do dziedziny biometrycznej analizy danych, pozyskiwanych w realistycznych warunkach z użyciem urządzeń mobilnych. Zaproponowane i zweryfikowane przez Doktorantkę metody analizy obrazów, ukierunkowane na rozpoznawanie osób na

podstawie struktury linii papilarnych spodu dłoni, mają silny wymiar praktyczny i mogą stanowić elementy przydatne dla tworzenia zaawansowanych systemów biometrycznych.

Konkludując recenzję, chciałbym stwierdzić, że rozprawa doktorska Pani magister inżynier Eweliny Bartuzi-Trokielewicz pt. „Presentation attack-resistant palm recognition for mobile devices in unconstrained conditions” **spełnia** w moim przekonaniu wymagania określone w odnośnej ustawie o stopniach i tytule naukowym i tym samym **wnioskuję o jej dopuszczenie do publicznej obrony**.

Signed by /
Podpisano przez:

Krzysztof Maciej
Ślot
Politechnika Łódzka

Date / Data:
2022-09-02 12:28

Literatura

[1] A. Urooj, A. Borji, Analysis of hand segmentation in the wild, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4710–4719.

[2]] W. Wang, Y.u. Kaicheng, J. Hugonot, Pascal Fua and Mathieu Salzmann, Recurrent U-Net for Resource-Constrained Segmentation, ICCV, 2019

[3] https://google.github.io/mediapipe/solutions/hands