

Zastosowanie sieci konwolucyjnej do wykrywania wybranych symptomów zmęczenia kierowcy

Streszczenie. Artykuł prezentuje wyniki badań mających na celu ocenę możliwości wykorzystania konwolucyjnej sieci neuronowej na potrzeby wykrycia objawów zmęczenia kierowcy w obrazie jego twarzy. Badania przeprowadzono z wykorzystaniem własnej bazy obrazów monochromatycznych zarejestrowanych w zakresie bliskiej podczerwieni. Uzyskane wyniki wskazują na przydatność proponowanego podejścia w budowie systemu poprawy bezpieczeństwa kierowania pojazdem.

Abstract. The paper presents the results of research aimed at the assessment of the possibilities of using convolutional neural networks to detect the symptoms of driver fatigue in a face image. The research was conducted with the use of own data pool which consisted of monochrome images acquired in the near infrared region. The results show that the proposed approach seems to be useful when implemented in the systems improving the safety of driving. (**Application of convolutional neural network to the problem of detecting selected symptoms of driver fatigue**).

Słowa kluczowe: przetwarzanie obrazów, monitoring kierowcy, konwolucyjne sieci neuronowe (CNN), uczenie głębokie.

Keywords: image processing, driver monitoring, convolutional neural networks (CNN), deep learning.

Wprowadzenie

Stanowiący potencjalne zagrożenie w ruchu komunikacyjnym stan przeciążenia psychofizycznego kierowców, którego detekcja jest przedmiotem prezentowanej pracy, należy do kategorii zmęczenia ogólnego. Może być ono efektem zarówno braku dostatecznej ilości snu jak również konieczności skupienia uwagi przez dłuższy czas w warunkach wzmożonego stresu wywołanego znaczącym natężeniem ruchu lub jazdy w trudnych warunkach atmosferycznych albo też monotonii pojawiającej się w trakcie spokojnej, ale długotrwałej jazdy. Stan ten przyjmuje formę zmęczenia, którego klasycznymi symptomami mogą być senność, obniżenie koncentracji i czujności, spowolnienie reakcji i zaburzenia percepcji [1], [2].

Wykorzystanie znanych i skutecznych metod analizy biosygnatów (jak EEG, EKG i GSR) do oceny powyższych symptomów jest niepraktyczne z uwagi na problemy z ich rejestracją w typowych warunkach eksploatacji pojazdu. Niewątpliwie kłopotliwe jest umieszczanie czujników na ciele kierowcy, np. czepka z zestawem elektrod do EEG, które wykazuje statystycznie istotne zmiany w rytmach mózgowych alfa i theta w trakcie spadku czujności [3]. Podejściem znacznie lepszym pod względem prostoty pobierania informacji diagnostycznej jest podejście bazujące na wykorzystaniu symptomów przeciążenia pojawiających się na twarzy badanej osoby, które można analizować metodami widzenia maszynowego.

Ten rodzaj wspomagania procesu rozpoznawania objawów zmęczenia kierowcy stanowi przedmiot zainteresowania dziedziny widzenia maszynowego od ponad 25 lat. Wtedy to w literaturze pojawiły się pierwsze prace, w których opisano zastosowanie metodyki opartej na aktywności okoruchowej i wykorzystującej do detekcji zmęczenia miarę PERCLOS (ang. *PERcentage of eyelid CLOSure over time*) [4], [5], [6], [7]. Niezależnie od pojawiających się w kolejnych latach innowacji, metodyka postępowania poddawała się standardowemu łańcuchowi przetwarzania w którym zasadniczym elementem była „inżynieria cech” wymagająca eksperckiego opisu obrazu za pomocą wektora parametrów, na bazie którego buduje się detektor.

Wśród stosowanych metod ekstrakcji cech w monitoringu zmęczenia najczęściej stosowane są podejścia geometryczne, w których informacja o stanie twarzy bazuje na lokalizacji określonych jej kształtów (np. krawędzi ust przy ziewaniu) i punktów charakterystycznych [8], [9], [10]. Cechy geometryczne mogą obejmować odległości

między takimi punktami, kąty jakie tworzą linie je łączące lub odległości pomiędzy punktami przy określonym wyrazie twarzy a punktami obrazu odniesienia z neutralnym wyrazem twarzy. Cechy te są atrakcyjne z uwagi na ich łatwą interpretację i szybkość przeprowadzenia obliczeń (o ile zostaną w obrazie wykryte). Alternatywnym podejściem jest wykorzystanie cech oddających wygląd obszaru twarzy, które zawierają informację o intensywności pikseli. Przykładem jest podejście holistyczne traktujące cały obraz jako macierz pikseli, z której metodami redukcji wymiaru danych znanych z biometrii jak np. technika twarzy własnych (ang. *eigenfaces*), bazująca na analizie składowych głównych PCA, wytwarzane są wektory cech liczbowych opisujące wyraz twarzy [11]. Inne podejście stosuje obliczanie cech bazujących na pikselach po podziale twarzy na obszary charakterystyczne dla jej wyrazów. Przykładem jest tu detektor teksturowy LBP (ang. *local binary pattern*), który jako wektor cech wykorzystuje histogram obrazu powstałego na drodze progowanego opisu sąsiedztwa pojedynczego piksela za pomocą zer i jedynek [12]. Do generacji cech stosowana jest również klasyczna filtracja obrazów z użyciem predefiniowanych masek o różnych skalach i orientacjach, np. masek filtrów Gabora [13].

Problem badawczy

Wobec dużej zmienności osobniczej potencjalnych użytkowników systemu wykrywania zmęczenia, spowodowanej m.in. wiekiem, wyrazem twarzy, wąsami itp., arbitralny opis ekspercki wynikający ze wspomnianej „inżynierii cech” dostarcza danych charakteryzujących się stosunkowo dużą niepewnością. Prezentowany w artykule problem badawczy dotyka oceny możliwości wykorzystania w przedmiotowej tematyce zupełnie innego podejścia do przetwarzania obrazów a mianowicie automatycznego poszukiwania dystynktywnej informacji za pomocą konwolucyjnej sieci neuronowej. Podejście takie wpisuje się w problematykę tzw. głębokiego uczenia (ang. *deep learning*), które stanowi jedno z najnowocześniejszych postępowań w dziedzinie współczesnego uczenia maszynowego [14].

W literaturze naukowej dostępne są publikacje dotyczące zastosowania sieci konwolucyjnych w wizyjnych systemach monitorowania stanu kierowcy. Sieci te są wykorzystywane zarówno w zadaniu automatycznej detekcji twarzy na obrazie [15], [16], [17] jak i do rozpoznawania stanu zmęczenia lub poszczególnych jego objawów a także

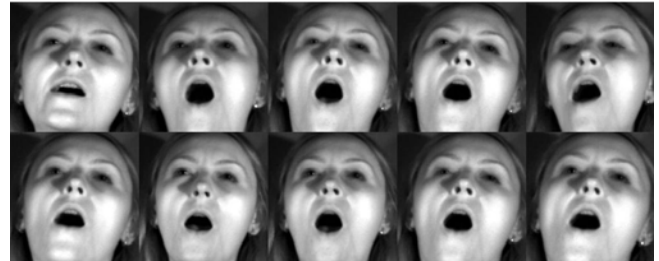
nieuwagi kierowcy. Popularne w literaturze jest zastosowanie sieci konwolucyjnych do określenia kątów obrotu twarzy na obrazie w różnych płaszczyznach a tym samym kierunku patrzenia, co pozwala na detekcję nieuwagi kierowcy lub jego zaśnięcia [16], [17], [18], [19]. Wykorzystanie tego typu podejścia umożliwia również wykrycie ziewania [19] oraz zamkniętych oczu [19], [20] na podstawie fragmentów obrazów twarzy odpowiadających obszarom odpowiednio ust lub oczu. Dodatkowo w pracy [20] zaproponowano wykorzystanie sieci konwolucyjnej do detekcji czynności takich jak jedzenie czy rozmawianie przez telefon komórkowy, które mogą rozpraszać kierowcę w trakcie jazdy. Nieco inne podejście zaproponowali autorzy prac [21], [22], [23], w których zadaniem sieci było rozróżnienie pomiędzy stanem zmęczenia a jego brakiem zamiast detekcji poszczególnych objawów. W pracy [21] wykorzystywana jest jedna sieć konwolucyjna, której dane wejściowe stanowi pojedynczy obraz, a decyzja końcowa podejmowana jest w oparciu o wyniki uzyskane dla kolejnych 60 ramek obrazu z kamery. Autorzy pracy [22] zaproponowali ocenę zmęczenia kierowcy za pomocą równoległe działających trzech konwolucyjnych sieci neuronowych o różnych strukturach przetwarzających ten sam obraz. Wynik klasyfikacji jest ustalany na podstawie integracji danych wyjściowych poszczególnych sieci. W pracy [23] zaproponowano wykorzystanie trójwymiarowej konwolucyjnej sieci neuronowej, która przetwarza nie pojedyncze obrazy a sekwencje wielu obrazów jednocześnie. Takie podejście umożliwia ocenę stanu zmęczenia na podstawie cech wypracowanych dla całej sekwencji ramek obrazu z kamery, co niewątpliwie wydaje się być bardziej wiarygodnym rozwiązaniem ale na pewno kosztowniejszym obliczeniowo w porównaniu z metodami wykorzystującymi sieci dwuwymiarowe.

Celem podjętych prac, których wyniki zostały przedstawione w niniejszym artykule, było opracowanie komputerowej metody przetwarzania obrazów twarzy osób, które podejmowały wysiłki symulowania ustalonych z góry wybranych objawów, uznanych na podstawie literatury za atrybuty zmęczenia, sygnalizowanego określonymi reakcjami mimicznymi twarzy. Należały do nich: ziewanie, marszczenie brwi będące efektem wzmoczonego skupiania uwagi oraz przymykanie powiek będące skutkiem senności. Zadaniem przetwarzania było wykrycie powyższych objawów. Należy zaznaczyć, że zrealizowane prace miały charakter *stricte* techniczny i nie obejmowały zagadnień dotyczących problematyki o charakterze fizjologicznym i psychologicznym, jak np. zadania wypracowania miary faktycznego zmęczenia na podstawie częstotliwości występowania objawów, dodatkowo z uwzględnieniem braku ich występowania u niektórych osób.

Materiał

Materiał do badań stanowiły obrazy zarejestrowane w zakresie bliskiej podczerwieni za pomocą kamery Basler acA2000-165umNIR wyposażonej w obiektyw dedykowany do rejestracji obrazów w zakresie światła widzialnego i bliskiej podczerwieni oraz filtr wąskopasmowy dopasowany do długości fali oświetlacza. Wybór tego zakresu spektralnego jak i samego zestawu rejestrującego, czyli kamery, filtra i oświetlacza został podyktowany koniecznością monitorowania kierowcy w różnych warunkach oświetleniowych (jazda zarówno w warunkach dziennych jak i nocnych). W efekcie rejestracji powstał uporządkowany zbiór danych symptomów zmęczenia stanowiących podstawę do opracowania i weryfikacji działania algorytmu. W zbiorze tym znalazły się obrazy twarzy łącznie 19 osób. Liczebność podzbiorów osób symulujących ziewanie oraz marszczenie brwi wynosiła 11.

W przypadku przymykania powiek było to 13 osób. Rozmiary obrazów przyjętych do badań zostały ustalone na 200×200 pikseli. Ich przykłady, reprezentujące zakładane objawy, zostały zamieszczone na rysunkach od 1 do 3.



Rys. 1. Przykłady obrazów przyjętych do badań i reprezentujących objaw ziewania



Rys. 2. Przykłady obrazów przyjętych do badań i reprezentujących objaw marszczenia brwi



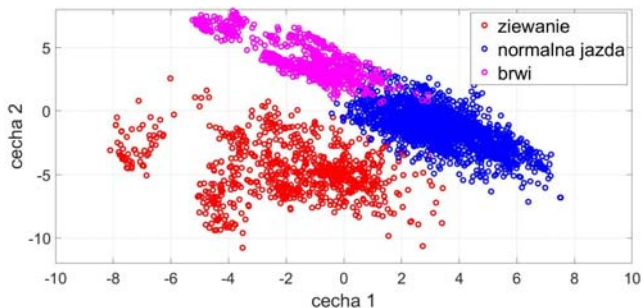
Rys. 3. Przykłady obrazów przyjętych do badań i reprezentujących objaw przymykania powiek

Możliwości i charakterystyka narzędzia

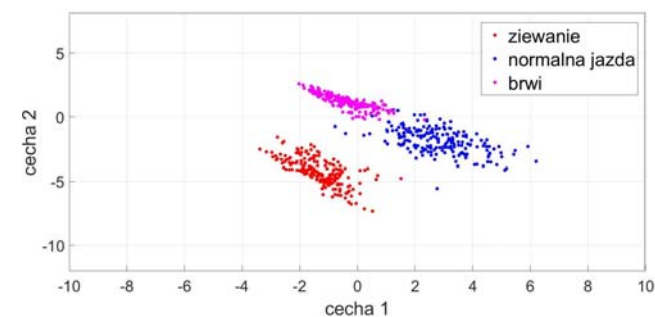
Wstępne wyniki uzyskane w zadaniu rozpoznawania obrazów reprezentujących stan normalnego wyrazu twarzy, stan wyrazu twarzy pojawiający się przy marszczeniu brwi oraz stan ziewania pokazują, że odpowiednio dobrana sieć konwolucyjna jest w stanie w sposób automatyczny, jedynie na podstawie dostarczonych obrazów, wygenerować z nich cechy, które mogą stać się podstawą podejmowania decyzji. Na rysunku 4 przedstawiono wykresy rozprożeń takich cech w przestrzeni 2D, wypracowanych dla wyróżnionych stanów mimicznych na etapie uczenia. W uczeniu brały udział całe obrazy twarzy wycięte za pomocą procedury Viola-Jonesa i pochodzące od podzbioru osób biorących udział w rejestracjach. Dopełnieniem tego podzbioru były obrazy jednej z osób, które utworzyły dane do weryfikacji. Wyniki przeprowadzonej weryfikacji zilustrowano na rysunku 5, na którym przedstawiony został wykres rozprożeń dla cech powstałych w wyniku podawania na wejście nauczonej sieci obrazów ze zbioru testującego. Obrazy te pochodziły od jednej osoby i zawierały reakcje mimiczne należące do każdej z 3 wyróżnionych kategorii. Zwraca uwagę właściwe wpisywanie się danych testujących w obszary zajęte przez wzorce uczące.

Podstawą konstrukcji sieci konwolucyjnej są warstwy przyjęte w ilości i rozmiarze zależnym od stopnia złożoności zadania rozpoznawania [14], [24]. Z uwagi na charakter danych wejściowych (dane obrazowe) oraz stosowane

współcześnie metody uczenia, powszechne jest stosowanie warstw, których wyszczególnienie, zastosowane w przykładowym zadaniu rozpoznawania w obrębie 3 klas (stan normalny, ziewanie i marszczenie brwi), ilustruje rysunek 6. Warstwa wejściowa jest zdefiniowana rozmiarami obrazów poddawanych przetwarzaniu i otrzymuje surowe wartości ich pikseli, które w procesie uczenia są eksponowane wraz z informacją o przynależności do klas.



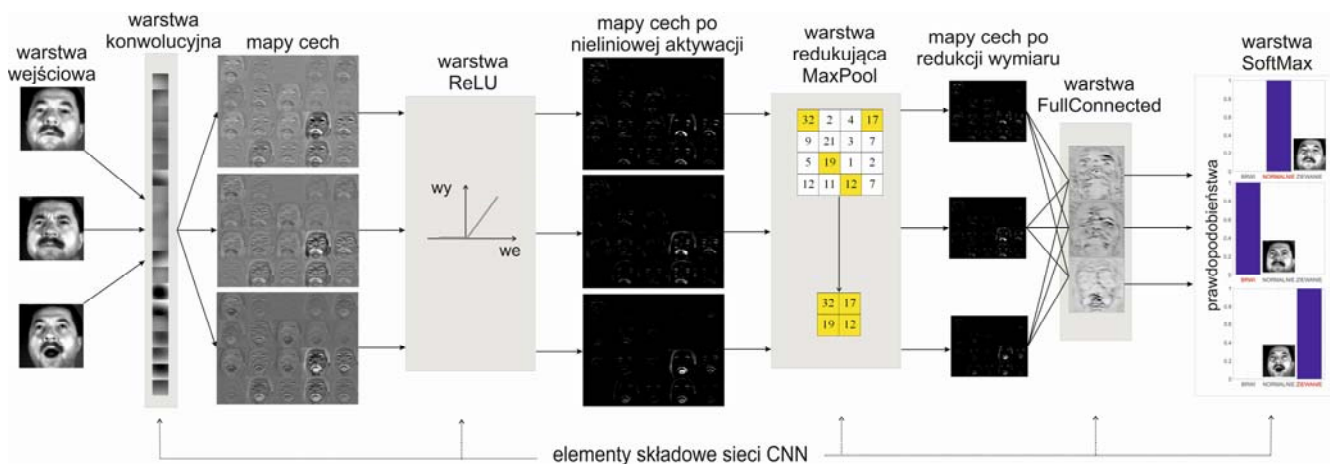
Rys. 4. Wykres rozproszony cech wyselekcjonowanych przez sieć konwolucyjną dla danych ze zbioru uczącego



Rys. 5. Wykres rozproszony odpowiedzi sieci na wymuszenie danymi ze zbioru testującego

Wymiar warstwy to wysokość i szerokość obrazów wejściowych, które w ramach wstępnego przetwarzania muszą być ujednoczone. Warstwę konwolucyjną tworzy generalnie pewna liczba zbiorów neuronów. Wejścia każdego z neuronów połączone są z pewną liczbą pikseli obrazu wejściowego tworzących lokalne, prostokątne pola recepcyjne o zadanych wymiarach lub z pewną liczbą elementów warstwy poprzedzającej, jeśli sieć posiada kilka warstw konwolucyjnych. Dla każdego ze zbiorów neuronów na etapie uczenia wyznaczane są wagi, przy czym każdy z neuronów należących do określonego zbioru ma te wagi jednakowe. Przetwarzanie obrazu przez nauczoną warstwę konwolucyjną może być więc ekwiwalentnie zastąpione

jego filtracją zrealizowaną za pomocą splotu (konwolucji) z wykorzystaniem neuronów-filtrów o ustalonych wagach-współczynnikach. Zbiór wszystkich powstałych w wyniku splatania obrazów tworzy mapy cech. Liczba 20 map cech przedstawionych na rysunku 6 odpowiada 20 neuronom (filtrów) w warstwie konwolucyjnej. Wynik działania warstwy konwolucyjnej może zawierać wartości ujemne, które są eliminowane przez nieliniową funkcję aktywacji o nazwie ReLU (ang. *rectified linear unit*). Funkcja ta wartości pikseli, które w mapach cech są dodatnie pozostawia bez zmian a ujemne zamienia na zerowe. Jej istnienie w sieciach głębokiego uczenia wynika z konieczności minimalizacji obliczeń wartości klasycznych funkcji aktywacji, jak to ma miejsce np. w przypadku funkcji sigmoidalnej. Sieci wykorzystujące funkcję aktywacji typu ReLU uczą się kilkakrotnie szybciej niż sieci, w których występują funkcje klasyczne [25]. Operację łączenia wielu wyników pochodzących z sąsiedztwa w mapach cech w jedną cechę realizuje warstwa redukująca tzw. warstwa pooling. Warstwa ta realizuje filtrację statystyczną w obrębie maski o zadanych rozmiarach wyznaczając wybraną statystykę, np. wartość maksymalną (tzw. MaxPooling przedstawiony na rys. 6) lub wartość średnią. Maski przetwarzające obraz w ten sposób nie pokrywają się. W warstwie tej nie występuje proces uczenia, ale dzięki możliwości wybrania odpowiednio dużego kroku przy przesuwaniu maski otrzymuje się redukcję wymiarów map cech (pomniejszone mapy cech), co sprzyja poprawie szybkości uczenia. Poza tym, w przypadku stosowania po niej kolejnej warstwy konwolucyjnej, warstwa redukująca daje jej neuronom możliwość objęcia większej powierzchni obrazu i tym samym kształtowanie innych rodzajów cech. Sieć uzyskuje również pewną odporność na przesunięcia wzorców obrazów. Ostatnia warstwa redukująca staje się warstwą wejściową dla jednej lub kilku warstw neuronów, z których każdy jest połączony ze wszystkimi neuronami (wyjściami) warstwy poprzedzającej – stąd nazwa: warstwa pełnego połączenia, tzw. warstwa FullConnected. Ostatnia tego typu warstwa dokonuje agregacji informacji ze wszystkich warstw poprzednich i zamyka strukturę automatycznego generatora cech. Liczba jej węzłów wyjściowych jest równa liczbie rozpoznawanych klas. Celem wypracowania decyzji klasyfikacyjnej, wyjściowa warstwa typu FullConnected jest poddawana działaniu aktywacji za pomocą funkcji, którą formalnie utożsamia się z istnieniem oddzielnej warstwy o nazwie Softmax. Funkcja typu softmax oblicza dla każdego neuronu z ostatniej warstwy typu FullConnected prawdopodobieństwo przynależności obrazu wejściowego do określonej kategorii.



Rys. 6. Ilustracja struktury sieci CNN w problemie rozpoznawania 3 stanów wyrazu twarzy kierowcy

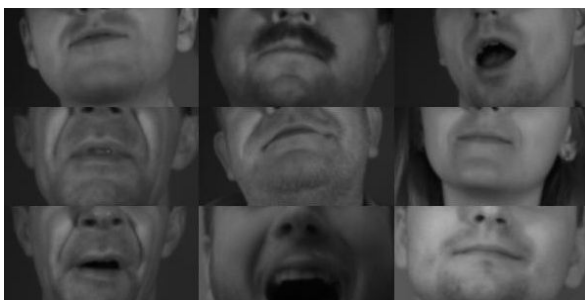
Metoda badawcza

Celem uwzględnienia możliwości występowania kilku atrybutów zmęczenia jednocześnie, finalne rozwiązanie przyjęło jednak nieco inną postać. Zaproponowano równoległe użycie 3 niezależnych detektorów, z których każdy podejmuje decyzję binarną w ramach symptomu, na który ma reagować, tzn. wykrywa dany symptom lub stwierdza stan normalny. Ograniczenie liczby klas do dwu stanowi znaczące ułatwienie dla procesu uczenia i umożliwia konstrukcję prostej struktury sieci neuronowej, która dzięki temu w trybie rozpoznawania nowych przypadków będzie działać szybko spełniając wymóg implementacji algorytmu w czasie rzeczywistym. W ramach podjętych prac zrealizowano próbę nauczania i jednocześnie weryfikacji przyjętej struktury sieci w sposób sekwencyjny. Obrazy każdej z pojedynczych osób stanowiły zbiór testowy względem zbioru uczącego utworzonego z obrazów pozostałych osób. Tym sposobem uzyskano niezależne wyniki detekcji dla wszystkich osób symulujących zakładane objawy, przy czym obrazy żadnej z nich nie wystąpiły w zbiorze uczącym.

Oceny stopnia wytrenowania sieci oraz jej zdolności do generalizacji przy przetwarzaniu danych testujących dokonano za pomocą macierzy błędów, zwanej też macierzą pomyłek (ang. *confusion matrix*). Macierz błędów jest macierzą kwadratową, w której wiersze odpowiadają zadanym klasom a kolumny decyzjom podejmowanym przez klasyfikator. Innymi słowy liczba znajdująca się na przecięciu i -tego wiersza i j -tej kolumny w macierzy błędów to liczba przypadków z klasy i , które zostały przypisane do klasy j . Elementy tej macierzy mogą być wyrażone w wartościach bezwzględnych lub w procentach. Podejście takie, w przeciwieństwie do miar globalnych, w sposób rzetelny uwzględniło uzyskiwanie przypadków mylenia wyróżnionych objawów ze stanem normalnym.

Wyniki

Celem uzyskania maksymalnie dużej szybkości działania na etapie rozpoznawania nowych przypadków, dla każdego z detektorów przyjęto najprostsze struktury sieci obejmujące wyszczególnione na rysunku 6 warstwy oraz ograniczone obszar analizy do fragmentów twarzy zawierających dany objaw. Przykład takiego ograniczenia przedstawiony jest na rysunku 7.



Rys. 7. Przykłady obrazów ze zbioru uczącego wykorzystanych w detekcji ziewania po ograniczeniu obszaru analizy

Wobec obserwowanego zjawiska przeuczenia się sieci zaprogramowanej do pracy przy dużej liczbie iteracji, proces uczenia zorganizowany został tak, aby uległ samoczynnemu przerwaniu po osiągnięciu średniej liczby poprawnych rozpoznań w ostatnich 20 iteracjach na poziomie co najmniej 95%. Wobec faktu, że uczenie i testowanie sieci odbywało się w sposób sekwencyjny, liczba macierzy błędów przy rozpoznawaniu każdego z objawów była równa liczbie osób symulujących dany objaw. Celem wypracowania miar charakteryzujących wykrycie objawu zastosowano uśrednianie macierzy błędów

uzyskiwanych dla każdej z osób. Uzyskane tym sposobem wyniki oceniające poszczególne detektory przedstawione zostały w tabelach od 1 do 3.

Tabela 1. Macierze błędów w procesie detekcji objawu ziewania (uśrednione wartości procentowe)

objaw	decyzja - uczenie		decyzja - testowanie	
	normalnie	ziewanie	normalnie	ziewanie
normalnie	99	1	96	4
ziewanie	4	96	5	95

Tabela 2. Macierze błędów w procesie detekcji objawu marszczenia brwi (uśrednione wartości procentowe)

objaw	decyzja - uczenie		decyzja - testowanie	
	normalnie	brwi	normalnie	brwi
normalnie	99	1	97	3
brwi	4	96	10	90

Tabela 3. Macierze błędów w procesie detekcji objawu przymykania powiek (uśrednione wartości procentowe)

objaw	decyzja - uczenie		decyzja - testowanie	
	normalnie	powieki	normalnie	powieki
normalnie	96	4	88	12
powieki	6	94	17	83

Podsumowanie

Przedstawione wyniki badań wskazują na potencjalne możliwości wykorzystania konwolucyjnych sieci neuronowych w rozpoznawaniu symptomów zmęczenia kierowcy. Wartości średnie błędów rozpoznawania objawu ziewania i marszczenia brwi w zbiorze danych obrazowych nie przekroczyły 10%, co można uznać za zadowalającą zdolność uogólniania, zwłaszcza przy uwzględnieniu faktu, że obrazy osób, które brały udział w testowaniu nie zasilają zbioru uczącego. Gorsze wyniki (ok. 15% błędów) zostały zaobserwowane w zakresie objawu przymykania oczu. Mogło to być spowodowane trudnością określenia przez osoby, których obrazy wykorzystano, ścisłej granicy pomiędzy oczami całkowicie otwartymi, oczami przymkniętymi i oczami całkowicie zamkniętymi. Opracowane detektory zostały zaimplementowane do pracy na stanowisku demonstracyjnym, którego zadaniem była detekcja i wskazywanie wyróżnionych objawów w trybie *on-line*. Zaobserwowano efekt krótkotrwałych fałszywych alarmów pojawiających się w związku z naturalną mimiką twarzy występującą w trakcie normalnej jazdy, rozmowy, rozmyślenia itp. Mimika ta może w sposób chwilowy dawać objawy podobne do wykrywanych, ale eliminacja jej wpływu jest możliwa dzięki filtracji ciągu prawdopodobieństw wyznaczanych dla kolejnych klatek przez warstwę SoftMax.

Praca współfinansowana ze środków NCBiR w ramach projektu No. PBS3/B9/29/2015.

Autorzy: dr hab. inż. Jacek Jakubowski, Wojskowa Akademia Techniczna, Wydział Elektroniki, ul. Kaliskiego 2, 00908 Warszawa, e-mail: jacek.jakubowski@wat.edu.pl;
mgr inż. Jolanta Chmielińska, Wojskowa Akademia Techniczna, Wydział Elektroniki, ul. Kaliskiego 2, 00908 Warszawa, e-mail: jolanta.pacan@wat.edu.pl.

LITERATURA

- [1] Otręba P., Paluch R, Obciążenie psychiczne personelu sterowania ruchem kolejowym, *Zeszyty Naukowe Wyższej Szkoły Zarządzania Ochroną Pracy w Katowicach*, Nr 1(4)/2008, ss. 17-32.
- [2] Krueger G. P., Sustained Work, Fatigue, Sleep Loss and Performance: A Review of the Issues, *Work and Stress*, *An International Journal of Work, Health and Organisations*, Volume 3, 1989 - Issue 2.

- [3] Eoh H. J., Chung M. K., Kim S. H., Electroencephalographic study of drowsiness in simulated driving with sleep deprivation, *International Journal of Industrial Ergonomics*, vol. 35, nr 4, 2005, ss. 307-320.
- [4] Grace R., i inni, A drowsy driver detection system for heavy vehicles, *Proceedings of the 17th Digital Avionics Systems Conference*, 1998, ss. 136 1-8.
- [5] Veeraraghavan H., Papanikolopoulos N. P., Detecting Driver Fatigue Through the Use of Advanced Face Monitoring Techniques, *Publikacja ITS Institute Research Report no. CTS 01-05*, 2001, University of Minnesota.
- [6] Tock D., Craw I., Tracking and measuring drivers' eyes, *Image and Vision Computing*, Volume 14, Issue 8, 1996, ss. 541-54.
- [7] Tock D., Craw I. Blink Rate Monitoring for a Driver Awareness System In: Hogg D., Boyle R. (eds) *BMVC92*. Springer (1992), London, ss. 518-527.
- [8] Yufeng L., Zengcai W., Detecting Driver Yawning in Successive Images, *The 1st International Conference on Bioinformatics and Biomedical Engineering*, 2007, ss. 581-583.
- [9] Alioua N. i inni, Driver's Fatigue Detection Based on Yawning Extraction, *International Journal of Vehicular Technology*, vol. 2014, Article ID 678786.
- [10] Tang X, i inni, Real-time image-based driver fatigue detection and monitoring system for monitoring driver vigilance, *35th Chinese Control Conference (CCC)*, 2016, ss. 4188-4193.
- [11] Bergasa L. M. i inni, Analysing Driver's Attention Level using Computer Vision, *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems*, Beijing, China, 2008, ss. 1149-1154.
- [12] Zhang Y., Hua C., Driver fatigue recognition based on facial expression analysis using local binary patterns, *Optik - International Journal for Light and Electron Optics*, Vol. 126, nr 23, 2015, ss. 4501-4505.
- [13] Fan X., i inni, Dynamic Human Fatigue Detection Using Feature-Level Fusion, *International Conference on Image and Signal Processing*, 2008, ss. 94-102.
- [14] Schmidhuber J., Deep Learning in Neural Networks: An Overview, *Technical Report IDSIA-03-14 / arXiv:1404.7828 v4*.
- [15] Zhang W., Murphey Y. L., Wang T., Driver yawning detection based on deep convolutional neural learning and robust nose tracking, *International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland 12-17 July 2015.
- [16] Yuen K., Martin S., Trivedi M. M., Looking at faces in a vehicle: A deep CNN based approach and evaluation, *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, 1-4 November 2016.
- [17] Yuen K., Martin S., Trivedi M. M., On looking at faces in an automobile: Issues, algorithms and evaluation on naturalistic driving dataset, *23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, 4-8 December 2016.
- [18] Vora S., Rangesh A., Trivedi M. M., On generalizing driver gaze zone estimation using convolutional neural networks, *IEEE Intelligent Vehicles Symposium (IV)*, Redondo Beach, CA, USA, 11-14 June 2017.
- [19] Ribarić S., Lovrenčić J., Pavešić N., A neural-network-based system for monitoring driver fatigue, *15th IEEE Mediterranean Electrotechnical Conference - MELECON 2010*, Valletta, Malta, 26-28 April 2010.
- [20] Yan Ch., Jiang H., Zhang B., Coenen F., Recognizing driver inattention by convolutional neural networks, *8th International Congress on Image and Signal Processing (CISP 2015)*, 2015.
- [21] Dwivedi K., Biswaranjan K., Sethi A., Drowsy driver detection using representation learning, *IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, 21-22 February 2014.
- [22] Park S., Pan F., Kang S., Yoo C.D. (2017) Driver drowsiness detection system based on feature representation learning using various deep networks. In: Chen CS., Lu J., Ma KK. (eds) *Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science*, vol 10118. Springer, Cham.
- [23] Huynh XP., Park SM., Kim YG. (2017) Detection of driver drowsiness using 3d deep neural network and semi-supervised gradient boosting machine. In: Chen CS., Lu J., Ma KK. (eds) *Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science*, vol 10118. Springer, Cham
- [24] Krizhevsky A., Sutskever I., Hinton G.E.,: Imagenet classification with deep convolutional neural networks, *NIPS Proceedings*, 2012, ss. 1106–1114.
- [25] Glorot X., Bordes A., Bengio Y., Deep Sparse Rectifier Neural Networks, *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, 2011, ss. 315-323.