

Rekonstrukcja zdjęć twarzy z wykorzystaniem systemu uczenia maszynowego

Streszczenie. W pracy przedstawiono możliwości zastosowania autorskiego modelu systemu uczenia maszynowego do rekonstrukcji i rozpoznawania zniekształconych, względnie uszkodzonych wzorców, a w szczególności obrazów twarzy zakrytych częściowo maskami. Prezentowane rozważania nie stanowią końcowych rozwiązań na poziomie aplikacyjnym dla systemu przetwarzania obrazów. Należy je raczej traktować jako potwierdzenie zasady (proof-of-concept) i uzasadnienie dalszych badań nad wykorzystaniem systemów opartych na superpozycji wzorców.

Abstract. The paper presents the possibilities for using a proprietary model of a machine learning system for the reconstruction and recognition of distorted or damaged patterns, in particular the images of faces partially covered with masks. The presented considerations do not constitute final solutions at the application level for an image processing system. Rather, they should be treated as a proof-of-concept and a justification for developing solutions using systems based on pattern superposition. (Face image reconstructions by using machine learning system)

Słowa kluczowe: sztuczna inteligencja, uczenie maszynowe, rekonstrukcja i rozpoznawanie obrazów.

Keywords: artificial intelligence, machine learning, image reconstruction and recognition.

Wstęp

Uczenie maszynowe wchodzące w skład szerszego obszaru wiedzy nazywanego sztuczną inteligencją zajmuje się algorytmami, które na podstawie przykładowych danych, zwanych zbiorami uczącymi, budują modele matematyczne mogące automatycznie podejmować decyzje lub prognozować. Problem uczenia stanowi klucz do zrozumienia inteligencji zarówno w biologicznych strukturach mózgowych jak i w maszynach. Celem uczenia maszynowego jest tworzenie odwzorowań $y = F(x)$, $y \in R^m, x \in R^n$, generowanych przez zbiory treningowe: $S = \{x_i, y_i\}_{i=1}^N$, których wektory stanowią węzły aproksymacji. W punktach treningowych zachodzi zatem

$$(1) \quad y_i = F(x_i), i = 1, \dots, N$$

We wcześniejszych pracach opisany został autorski model uczenia maszynowego [1]. Model ten bazuje na analizie spektralnej, wykorzystując transformacje biortogonalne i ortogonalne zrealizowane w oparciu o macierze Hurwitza-Radona. Należy podkreślić, że charakteryzuje się on uniwersalnością, pozwalającą na realizację podstawowych funkcji systemów uczących się, takich jak kojarzenie (Pattern Association), rozpoznawanie i klasyfikacja wzorców (Pattern Recognition) oraz modelowanie odwrotne (Inverse Modelling). Jedną ze wspomnianych własności tego modelu jest rozpoznawanie i rekonstrukcja wzorców będących obrazami. W pracy [1] zamieszczono przykład gdzie obiektem rekonstrukcji był niepełny obraz Leny. Wcześniejsze prace z zawartymi w nich przykładami rekonstrukcji dają przesłanki do stworzenia systemu rozpoznawania osób noszących maski, w oparciu o w/w model nauczania maszynowego. System taki wydaje się być użyteczny w obecnej sytuacji społecznej. Warto zauważyć, że wyżej wymieniany autorski model uczenia maszynowego stanowi alternatywę dla klasycznych systemów rekonstrukcji/restauracji obrazów (image restoration), które wykorzystują takie narzędzia przetwarzania jak modelowanie odwrotne, filtry Wienera oraz analizę PCA (Principal Component Analysis). Klasyczne systemy są aktualnie intensywnie uzupełniane i zastępowane przez struktury różnego typu algorytmów neuronowych, szczególnie w zastosowaniach medycznych i w mikroskopii. Na uwagę zasługują tutaj struktury głębokiego uczenia (Deep Learning Neural Networks-DLNN) a szczególnie struktury spłotowego przetwarzania obrazów (Convolutional DLNN) [2], [3], [4].

Struktura modelu uczenia maszynowego do przetwarzania obrazów

W niniejszej pracy rozważa się zbiór M obrazów czarno-białych reprezentowanych przez m wierszy i n kolumn, jest to zatem zbiór $(m \cdot n)$ pikseli o różnych odcieniach szarości. Do analizy wektorowej, każdy z obrazów transformuje się przez konkatenację m wierszy do postaci wektora kolumnowego $u_i (m \cdot n \times 1), i = 1, \dots, M$. Tak więc zbiór M obrazów jest reprezentowany przez macierz:

$$(2) \quad U = [u_1, u_2, \dots, u_M], \dim u_i = m \cdot n = 2^k, k = 3, 4, \dots,$$

gdzie:

$$(3) \quad M < \frac{1}{2}(n \cdot m)$$

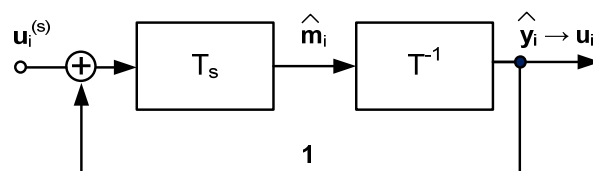
Zbiór obrazów zniekształconych dany jest przez macierz:

$$(4) \quad U^{(s)} = [u_1^{(s)}, u_2^{(s)}, \dots, u_M^{(s)}]$$

Nietrudno zauważyć, że zbiór treningowy ma postać:

$$(5) \quad S = \{u_i, u_i^{(s)}\}_{i=1}^M$$

Korzystając z opublikowanej [1] ogólnej struktury modelu uczenia maszynowego, rekonstrukcja obrazów $u_i^{(s)}$ dokonywana jest z zastosowaniem transformacji reprezentowanych blokowo na rysunku 1.



Rys. 1. Struktura modelu uczenia maszynowego do przetwarzania obrazów

Transformacja ortogonalna $T(\cdot)$, wykorzystująca rodzinę macierzy Hurwitza-Radona, pozwala na wyznaczenie widma Haar'a-Fouriera wektorów systemowych $u_i, i = 1, \dots, M$:

$$(6) \quad m_i = \frac{1}{2}(W_{2^k} + 1)u_i = T(u_i)$$

gdzie: W_{2^k} – macierz antysymetryczna ortogonalna, $W_{2^k}^2 = -\mathbf{1}$; \mathbf{m}_i – widma wektorów $\mathbf{u}_i, i = 1, \dots, M$; $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_M\}$ – macierz wektorów widma; $\mathbf{1}$ – macierz jednostkowa.

Blok T^{-1} systemu z rysunku 1 realizuje operację:

$$(7) \quad \mathbf{u}_i = (-W_{2^k} + \mathbf{1})\mathbf{m}_i = T^{-1}(\mathbf{m}_i)$$

Natomiast blok T_s transformację biortogonalną:

$$(8) \quad \mathbf{m}_i = (2 \cdot \mathbf{1} - W_S - W_{2^k})^{-1} \mathbf{u}_i$$

Stąd

$$(9) \quad T_s(\cdot) = (2 \cdot \mathbf{1} - W_S - W_{2^k})^{-1}$$

gdzie: $W_S = \mathbf{M}(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T$ – macierz symetryczna

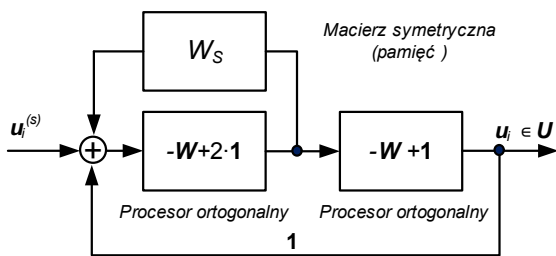
Struktura z rysunku 1 pełni funkcję estymatora widma $\{\hat{\mathbf{m}}_i\}$:

$$(10) \quad \hat{\mathbf{m}}_i = T_s(\mathbf{u}_i^{(s)}), i = 1, \dots, M$$

W systemie, ze względu na iteracyjne działanie pętli sprzężenia zwrotnego, uzyskuje się konwergencję wektorów:

$$(11) \quad \hat{\mathbf{m}}_i \rightarrow \mathbf{m}_i,$$

$$(12) \quad \hat{\mathbf{y}}_i \rightarrow \mathbf{u}_i, i = 1, \dots, M$$



Rys. 2. Schemat blokowy aproksymatora ze skupioną pamięcią

Strukturę z rysunku 1 można także przedstawić jako model ze skupioną pamięcią zamieszczony na rysunku 2.



Rys. 3. Zapamiętane obrazy twarzy (źródło: <https://pixabay.com/pl>)

Przykłady rekonstrukcji twarzy i rozpoznawania osób

Opisany w poprzednim rozdziale system maszynowego przetwarzania twarzy 9 osób ($M = 9$) zapisane w formie macierzy (64×64) określającej stopień szarości poszczególnych pikseli obrazów. Zapamiętane obrazy twarzy są przedstawione na

rysunku 3. Na potrzeby analizy wektorowej każdy z obrazów został przekształcony przez konkatencję 64 wierszy do postaci wektora kolumnowego $\mathbf{u}_i (64 \cdot 64 \times 1), i = 1, \dots, 9$. Po takim przekształceniu zbiór 9 obrazów jest reprezentowany przez macierz $\mathbf{U} (4096 \times 9)$:

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_9], \dim \mathbf{u}_i = 64 \cdot 64 = 4096 = 2^k, k = 12$$



cji i klasyfikacji



Rys. 4. Rekonstrukcja obrazów twarzy osób noszących maski

W każdym wektorze na ostatniej pozycji (4096) zastąpiono daną o stopniu zaciemnienia skrajnego piksela obrazu numerem identyfikacyjnym osoby, do której jest przypisany ten obraz. W przedstawionych doświadczeniach numerem identyfikacyjnym był numerem kolejnym obrazu 1, 2, ..., 9. Tak utworzone wektory $u_i, i = 1, 2, \dots, 9$ posłużyły do syntezy systemu uczenia maszynowego zgodnie z procedurą opisaną w poprzednim rozdziale. Przykłady rekonstrukcji zdjęć osób noszących maski przedstawiono na rysunku 4. Przypisanie zdjęcia do konkretnej osoby następowało na podstawie numeru identyfikacyjnego zapisanego na ostatnich pozycjach wektorów $u_i, i = 1, 2, \dots, 9$.

W tabeli 1 przedstawiono wartości nominalne „indeksu rozpoznania” oraz uzyskane jego wartości po 100 iteracjach.

Tabela 1. Wartości „indeksu rozpoznania” osoby

Numer zdjęcia	Wartość nominalna indeksu	Wartość indeksu po 100 iteracjach
1	1,0	0,8622
2	2,0	1,6240
3	3,0	2,3660
4	4,0	3,9983
5	5,0	5,1259
6	6,0	5,8842
7	7,0	6,7262
8	8,0	8,0466
9	9,0	8,9576

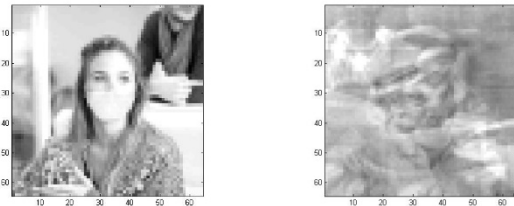
Tabela 2. Zbieżność procesu iteracyjnego

Liczba iteracji	Wartość indeksu	Liczba iteracji	Wartość indeksu
1	-0,0813	7	1,4607
2	0,1758	8	1,5394
3	0,5332	9	1,5843
4	0,8703	10	1,6078
5	1,1394	12	1,6233
6	1,3327	100	1,6240

Z tabeli wynika, że wartość indeksu po zaokrągleniu do wartości całkowitej w większości przypadków odpowiada wartości nominalnej. Zatem system dokonał poprawnej identyfikacji osoby. Jedynie w przypadku zdjęcia numer 3 nastąpiło błędne rozpoznanie. Dalsze zwiększanie liczby iteracji nie wpływało na zmianę wskaźnika gdyż proces szybko zbiega się do wartości końcowej. Zbieżność procesu iteracyjnego ilustruje tabela 2 w której przedstawiono wartości indeksu po kolejnych iteracjach. Doświadczenie przeprowadzono dla zdjęcia numer 2 z nominalną wartością współczynnika 2.0.

Istotny rezultat potwierdzający zasadę działania

rozważanego tutaj systemu uzyskano podstawiając na jego wejście niezapisane w nim zdjęcie. Odpowiedź widoczna na rysunku 5 jest superpozycją zapisanych w systemie zdjęć.



Rys. 5. Próba rozpoznania niezapisanego zdjęcia

Dla przeprowadzonych wcześniej rekonstrukcji obrazów obliczono błędy średniokwadratowe (MSE). Wyniki zamieszczono w tabeli 3. W drugiej kolumnie tabeli znajduje się błąd średniokwadratowy opisujący różnicę pomiędzy zdjęciem oryginalnym a zdjęciem „zamaskowanym”. Natomiast w trzeciej kolumnie błąd opisujący różnicę pomiędzy zdjęciem oryginalnym a zdjęciem po procesie rekonstrukcji. W każdym przypadku gdy analizie jest poddawany zapisany w systemie obraz następuje zmniejszenie błędu średniokwadratowego. W przypadku próby rekonstrukcji przedstawionej na rysunku 5, zdjęcia niezapisanego w systemie, błąd MSE wynosił 2950,90.

Tabela 3. Błąd średniokwadratowy rekonstrukcji

Numer zdjęcia	MSE (zdjęcie oryginalne - zdjęcie w masce)	MSE (zdjęcie oryginalne - zdjęcie po rekonstrukcji)
1	366,49	105,06
2	595,96	176,38
3	1573,00	570,95
4	398,00	37,58
5	552,04	114,55
6	675,67	112,13
7	828,53	221,09
8	171,52	26,05
9	327,06	40,75

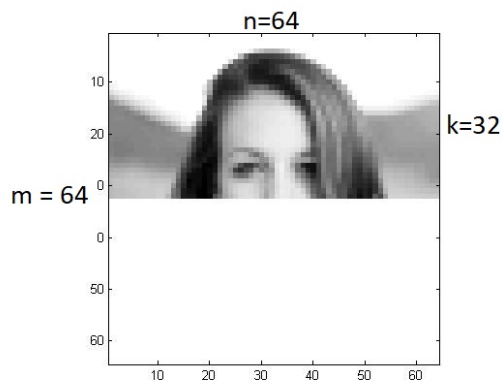
Ułamkowa wartość indeksu z tabeli 1 jest rezultatem mechanizmu działania systemu jest to ważona kombinacja liczb 1, ..., 9.

Uwaga:

W przypadku innego sposobu maskowania zilustrowanego przykładowo na rysunku 6, zbiór obrazów zniekształconych dany jest zgodnie z zależnością (4) przez macierz:

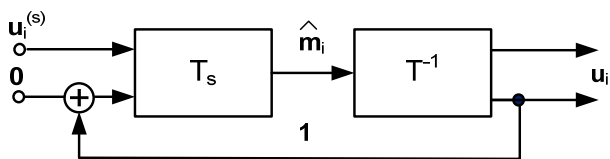
$$(13) \quad U^{(s)} = [u_1^{(s)}, u_2^{(s)}, \dots, u_M^{(s)}]$$

gdzie: $\dim u_i^{(s)} = (k \cdot n \times 1), i = 1, \dots, M$
 $0,1m < k < m$



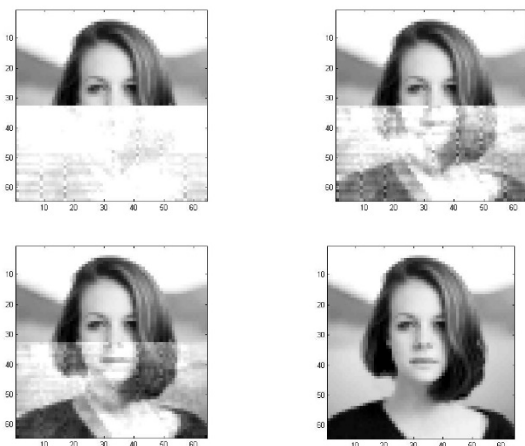
Rys. 6. Zamaskowany obraz twarzy numer 9

Struktura modelu do rekonstrukcji tak zniekształconych obrazów ma postać jak na rysunku 7.



Rys. 7. Struktura systemu do rekonstrukcji z utrzymaniem na wejściu fragmentu obrazu (k-wierszy)

Proces rekonstrukcji obrazu z rysunku 6 zilustrowano na rysunku 8 na którym pokazano uzyskane rezultaty po 1, 5, 10 i 100 iteracjach. Po 100 iteracjach indeks MSE rekonstrukcji wynosi 0 a „wskaźnik identyfikacji” 9.0. Warto porównać powyższe wartości z danymi dla zdjęcia numer 9 zamieszczonymi w tabelach 1 i 3.



Rys. 8. Proces rekonstrukcji obrazu z rysunku 6 (rezultat po 1, 5, 10 i 100 iteracjach)

Podsumowanie

Celem pracy było pokazanie możliwości zastosowania modelu uczenia maszynowego, o strukturze jak na rysunku 1, do rekonstrukcji i rozpoznawania zniekształconych, względnie uszkodzonych wzorców a w szczególności obrazów twarzy z założonymi maskami. Prezentowane w pracy rozważania nie stanowią jednak końcowych rozwiązań o poziomie aplikacyjnym dla systemu przetwarzania obrazów. Należy je raczej traktować jako potwierdzenie zasady (proof-of-concept) i uzasadnienie rozwijania aplikacyjnych rozwiązań z wykorzystaniem systemów opartych na superpozycji wzorców.

Autorzy: dr inż. Wiesław Citko, Uniwersytet Morski Gdynia, Wydział Elektryczny, ul. Morska 81-87, 81-225 Gdynia, E-mail: w.citko@we.umg.edu.pl; inż. Adam Trzebiatowski, Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, ul. G.Narutowicza 11/12, 80-233 Gdańsk, E-mail: atrzeb@wp.pl; dr hab. inż. Wiesław Sienko, Uniwersytet Morski Gdynia, Wydział Elektryczny, ul. Morska 81-87, 81-225 Gdynia, E-mail: w.sienko@we.umg.edu.pl.

LITERATURA

- [1] Citko W., Sienko W., Hamiltonian and Q-Inspired Neural Network-Based Machine Learning, *IEEE Access*, vol.8 (2020), 220437-220449, doi: 10.1109/ACCESS.2020.3043035
- [2] Ravishankar S., Ye J. C., Fessler J. A., Image Reconstruction: From Sparsity to Data-Adaptive Methods and Machine Learning, *Proceedings of the IEEE*, vol.108, no.1 -(2020), 86-109, doi: 10.1109/JPROC.2019.2936204
- [3] Kaderuppan S. S., Wong W. W. L., Sharma A., Woo W. L., Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy, *IEEE Access*, vol.8 (2020), 214801-214831, doi: 10.1109/ACCESS.2020.3040319
- [4] Gonzales R. C., Woods R. E., Digital Image Processing, Pearson International Edition, 2008.