

Klasyfikacja dziewięciu typów materiałów pornograficznych za pomocą modelu sAI 0.3. Kontynuacja badań pilotażowych

Classification of nine types of pornographic materials using the sAI 0.3 model. Continuation of the pilot study

Wojciech Oronowicz-Jaśkowiak¹, Adam Siwiak², Krzysztof Róg³,
Agnieszka Oronowicz-Jaśkowiak⁴

¹ Polsko-Japońska Akademia Technik Komputerowych, Wydział Informatyki

² Uniwersytet Jagielloński, Instytut Psychologii

³ Uniwersytet Humanistycznospołeczny SWPS, Wydział Psychologii

⁴ Narodowy Instytut Onkologii im. Marii Skłodowskiej-Curie Państwowy Instytut Badawczy,
Zakład Radiologii I

Summary

Aim. Legal pornographic materials are a heterogenous group of audiovisual materials that depict one or more person over the age of eighteen engaging in sexual activities. The aim of this study was to train a model that could classify given types of pornographic materials.

Method. Materials included in the training set (3,600 materials) and the validation set (900 materials) were manually classified and tagged by psychologists-sexologists. Then, a deep neural network was trained on the dataset. Six models based on different architectures of convolutional neural networks were included in the study (ResNet152, ResNet101, VGG19, VGG16, Squeezenet 1.1, Squeezenet 1.0). Each model was trained on the same group of photographs, and fast.ai library was used for the training process.

Results. The final model allows for the classification of more types of pornographic materials with greater efficiency than the pilot model, and thanks to the manual labelling of individual photographs, the limitations of the classification are known.

Conclusions. The possible applications of the model in clinical sexology and psychiatry are discussed. The application of deep neural networks in sexology seems to be particularly promising for at least two reasons. Firstly, a tool for automated detection of pornographic materials involving minors can be developed and used during criminal proceeding. Secondly, after retraining the presented model on photographs of men and women not engaging in sexual activity the model could be used to filter content that is inappropriate for minors.

Słowa klucze: materiały pornograficzne, sieci neuronowe

Key words: pornographic materials, neural networks

Wprowadzenie

Wykrywanie i klasyfikowanie materiałów pornograficznych z użyciem modeli uczenia maszynowego, w tym głębokich sieci neuronowych, jest nowym zagadnieniem. Short i wsp. [1] dokonali przeglądu publikacji, których problematyka dotyczyła charakterystyki pornografii internetowej pojawiającej się w ciągu poprzednich dziesięciu lat. Podstawę analizy stanowiło 46 publikacji, przy czym w połowie z nich badania prowadzone były online, 43% badań wykonano z użyciem ankiet w wersji papier-olówek, w kilku zastosowano obie wspomniane metody, a jedno badanie przeprowadzono telefonicznie. W publikacjach analizowanych przez autorów pojawia się wiele różnych definicji pornografii, jednak aż w 84% z nich nie przedstawiono jej osobom badanym. Z tego powodu Short i wsp. [1] uważają, że niezbędne jest precyzyjne zdefiniowanie pojęcia pornografii w każdej przyszłej publikacji, gdyż kwestia definicji ma bezpośredni wpływ na wyniki oraz porównanie wyników uzyskanych przez różnych autorów. Takie zaniedbanie w konsekwencji prowadzi do tego, że jedni autorzy uważają, iż uwidocznienie narządów płciowych jest wystarczającym powodem, aby sklasyfikować treści jako pornograficzne, inni zaś za pornografię uznają jedynie pokazanie pełnego aktu seksualnego. Short i wsp. [1] twierdzą, że dobrze sformalizowana definicja powinna zawierać informację o rodzaju pornografii i sposobie, w jaki określony materiał ma oddziaływać na osobę, która taki materiał ogląda. Ich zdaniem za pornografię należy uznać każdy materiał o charakterze jednoznacznie seksualnym, prezentowany w celu wywołania podniecenia seksualnego lub seksualnych fantazji.

Ponad dwie dekady temu Scott [2] wyróżniał jedynie dwa typy pornografii:

- (1) łagodną (*soft-core*) – nie występują tu zachowania dewiacyjne, materiały tego typu przedstawiają aktywność seksualną za obopólną zgodą (w tej kategorii umieszczono również materiały „neutralne”, wykorzystywane do edukacji seksualnej);
- (2) twardą (*hard-core*) – występują tu zachowania dewiacyjne, przemoc seksualna, przykładowo akty pedofilne czy zoofilne.

Kilkanaście lat później Hald [3] prezentował osobom badanym 31 różnych kategorii materiałów pornograficznych, spośród których wyodrębnił osiem najczęściej wybieranych przez osoby badane. Badania Halda [3] ilustrują, w jak znaczący sposób rozwój nowych technologii może wpłynąć na różnorodność materiałów pornograficznych. Współcześnie, ze względu na powszechność internetu, materiały pornograficzne są szeroko dostępne.

Jednocześnie wzrasta liczba osób poszukujących pomocy z powodu problematycznego korzystania z pornografii. Dane pochodzące z samopomocowych grup 12-krokowych w Polsce świadczą, że pomiędzy rokiem 2009 a 2012 liczba członków grup związanych z nałogowymi zachowaniami seksualnymi wzrosła o 340% [4]. Jednym z częstych problemów jest wykształcenie się uwarunkowań masturbacyjnych

w odpowiedzi na specyficzny typ aktu seksualnego, który może nie być możliwy do osiągnięcia w rzeczywistości. Samo podejmowanie interwencji terapeutycznych w obszarze uzależnienia od pornografii jest utrudnione, zważywszy na to, że użytkownicy pornografii w bardzo niewielkim stopniu są świadomi negatywnych skutków częstego oglądania treści pornograficznych [5]. Podkreśla się również, że materiały pornograficzne mogą w niekorzystny sposób rzutować na rozwój dzieci i adolescentów [6]. Wśród zagrożeń wymienia się m.in. tworzenie nierealistycznych oczekiwań co do aktu seksualnego (w tym występującego często z wykorzystaniem elementów fetyszystycznych) czy seksualizację zachowania.

Podsumowując, nie tylko kategoryzowanie, co jest pornografią, a co nią nie jest, ale też klasyfikowanie, z jakiego typu pornografią mamy do czynienia, jest zadaniem, które staje się coraz trudniejsze wraz z poszerzaniem się zasobów tego rodzaju treści dostępnych przez internet.

Na potrzeby niniejszej pracy przyjęto, że legalne materiały pornograficzne stanowią niejednorodną grupę materiałów audiowizualnych, które przedstawiają jedną lub większą liczbę osób powyżej osiemnastego roku życia podczas podejmowania czynności o charakterze seksualnym.

Zastosowanie sieci neuronowych w seksuologii

Głębokie sieci neuronowe odwzorowują funkcjonowanie ludzkiego układu nerwowego z zastosowaniem struktur matematyczno-statystycznych, które można porównać do sztucznych neuronów. Składniki sieci stanowią siatkę sztucznych neuronów i ich połączeń, które mogą być dwójakiego rodzaju – ze sprzężeniem zwrotnym (wtedy musi upłynąć pewien czas, zanim uzyska się odpowiedź) lub bez sprzężenia zwrotnego (wtedy reakcja na pobudzenie jest natychmiastowa) [7]. Modele wykorzystujące sieci neuronowe wykonują złożone zadania matematyczne i na ich podstawie formułowane są predykcje lub deskrypcje określonych baz danych. Postner i Rothbart [8] już w 2007 roku wykazali, że sieci neuronowe mogą być wykorzystane w naukach psychologicznych do przewidywania i zrozumienia różnych form ludzkich zachowań. Chociaż pierwsze badania z zastosowaniem sieci neuronowych w naukach medycznych i naukach społecznych prowadzono już kilkanaście lat temu – na przykład do analizy związku pewnych zachowań z określonymi typami osobowości [9] – w ciągu kilku ostatnich lat zastosowania te zyskały na popularności.

Wedle prognoz szczególnie w dziedzinie seksuologii sieci neuronowe mogą stać się przydatne w pracy klinicznej. Przykładem rozpoznania, które może być w przyszłości wsparte przez modele uczenia maszynowego, jest diagnoza zaburzeń parafilicznych w wypadku spraw karnych związanych z uzyskiwaniem dostępu do materiałów pornograficznych z udziałem małoletnich.

Diagnozowanie zaburzeń preferencji seksualnych u sprawcy przestępstwa polegającego na uzyskiwaniu dostępu do materiałów pornograficznych z udziałem małoletnich jest wymagającym zadaniem, często wiążącym się z koniecznością analizy całego materiału pobranego przez sprawcę przestępstwa [10]. Ewentualne rozpoznanie zaburzenia pedofilnego u sprawcy określonego przestępstwa seksualnego ma dla niego

konsekwencje prawne, w tym w zakresie warunków i przebiegu odbywania kary, oraz konsekwencje terapeutyczne, czyli poddanie się specjalistycznej terapii. Z tego powodu osoby z zaburzeniami preferencji seksualnych często ukrywają określone fantazje seksualne¹, a na znaczeniu zyskuje obiektywna ocena materiałów pornograficznych. Trudnością jest natomiast przeprowadzenie rzetelnej analizy często kilkunastu tysięcy materiałów pornograficznych.

Jak wykazano wcześniej, istnieje wiele rodzajów materiałów pornograficznych, a ich klasyfikacja jest utrudniona. Poszukiwane są zatem coraz bardziej dokładne techniki klasyfikowania obrazów, takie jak sieci neuronowe. W badaniu Wanga i Kosińskiego [12] porównano skuteczność rozpoznawania orientacji seksualnej przez człowieka oraz przez model. Człowiek poprawnie rozróżniał mężczyzn heteroseksualnych od homoseksualnych w 61% przypadków. Użycie sieci neuronowych pozwalało na rozróżnienie 81% przypadków, a po prezentacji pięciu obrazów tej samej twarzy dokładność modelu wzrosła do 91% [12]².

Ciekawe połączenie analizowania obrazu i ruchu zaproponowali Perez i wsp. [13]. Nagość utrwalona na obrazach, które można spotkać w internecie, nie zawsze ma kontekst seksualny. Aby polepszyć skuteczność kategoryzowania materiałów, autorzy zaproponowali możliwość analizowania także ruchu. Łącząc statyczny obraz (zdjęcie) z obrazem ruchu (wideo), uzyskali model cechujący się dokładnością na poziomie 97,9% oraz redukcję błędów o 65,4%.

Cel oraz uzasadnienie badań

Model sAI 0.1 dostosowanej do materiałów pornograficznych zaproponował ostatnio Oronowicz-Jaśkowiak [10]. Wersja pilotażowa wytrenowana została do różnicowania między siedmioma typami pornografii. Dokładność modelu wynosi 70%. Wykorzystana metodologia doboru materiału treningowego nie była usystematyzowana (por. uzasadnienie w dalszej części artykułu), ponadto model cechował się niską dokładnością w wypadku klasyfikacji materiałów spoza zbioru treningowego. W związku z tym wytrenowano nowy model sieci neuronowej, oparty na modelu sAI 0.1 [10] oraz bazie sexACT [14]. Dalsze prowadzenie badań w tym kierunku przez rozwijanie obecnego modelu jest istotne z kilku powodów.

Po pierwsze, ze względu na podobną treść wizualną materiałów pornograficznych z udziałem osób dorosłych oraz materiałów pornograficznych z udziałem małoletnich w przyszłości na podstawie tego modelu będzie można stworzyć kolejne, które będą użyteczne w kontekście opiniowania sędowo-seksuologicznego [15].

Po drugie, stworzony model można już wykorzystywać do prowadzenia badań naukowych, na przykład do dokonania automatycznej klasyfikacji oraz przydzielania

¹ Dodatkowym utrudnieniem jest to, że obraz parafilii oraz kryteria diagnostyczne są tworzone głównie na podstawie informacji pochodzących od osób, które weszły w konflikt z prawem lub same zgłosiły się po pomoc [11].

² Dyskusji wymaga etyka zastosowania praktycznego tego modelu, w tym również celowość jego zastosowania.

danych zdjęć do odpowiednich warunków badawczych, jak choćby w zadaniu Stroopa w warunku emocjonalnym [16].

Metodologia

Materiał badawczy stanowiło łącznie 4500 zdjęć – 3600 z nich zostało użytych do wytrenowania sieci, a pozostałych 900 do jej późniejszej walidacji. Proces tworzenia modelu, który różnicuje materiały pornograficzne z udziałem małoletnich, składał się z kilku etapów.

Etap 1. Przygotowanie materiału treningowego i walidacyjnego

W celu wytrenowania modelu do automatycznej klasyfikacji typów materiałów pornograficznych pobierano oraz opisywano zdjęcia zawierające pornografię z udziałem osób pełnoletnich, przyporządkowaną do odpowiednich kategorii materiałów pornograficznych.

Można odnaleźć setki kategorii tematycznych odnoszących się do materiałów pornograficznych z udziałem osób dorosłych. Kategorie te mogą dotyczyć różnej aktywności seksualnej, jej kontekstu bądź występowania specyficznego obiektu (fetyszu). Na użytek niniejszych badań kierowaliśmy się przeglądami literatury przedmiotu [1, 3] i na tej podstawie wyodrębniliśmy kilka grup tematycznych materiałów pornograficznych.

Zastosowane przez nas kryterium wyboru grup tematycznych materiałów pornograficznych zdecydowanie nie wyczerpuje złożoności tego zagadnienia. Przyjęto kryterium priorytetyzacji materiałów pornograficznych występujących najczęściej, natomiast nie wszystkie z włączonych grup tematycznych materiałów pornograficznych występują najczęściej. Drugim kryterium wyboru było znaczenie kliniczne włączonych grup tematycznych, jak w wypadku materiałów pornograficznych określanych jako „AB/DL” (ang. „Adult Baby / Diaper Lover”, pol. „dorosłe dziecko / sympatyk pieluch”). Zdecydowaliśmy również o włączeniu do modelu kategorii materiałów pornograficznych z udziałem osób, u których widoczny był tatuaż³, co uzasadniamy możliwością względnie łatwego różnicowania między materiałami tego typu a innymi kategoriami materiałów pornograficznych.

Pierwsze działania były ukierunkowane na półautomatyczne pobranie zdjęć z serwisów internetowych zawierających materiały pornograficzne różnego rodzaju. W dalszej kolejności zespół trzech psychologów seksuologów dokonywał oceny jakości danego zdjęcia oraz przyporządkowywał je do odpowiedniej kategorii. W tabeli 1 przedstawiono kryteria jakościowe włączenia danego zdjęcia.

³ Wyodrębnienie tej kategorii, choć interesujące z punktu widzenia przetwarzania obrazów przez sieć neuronową, niekoniecznie ma duże znaczenie kliniczne ze względu na to, że fetyszyzm związany z obecnością tatuaży rzadko jest przedmiotem interwencji klinicznych.

Tabela 1. Kryteria jakościowe materiału badawczego

Kryteria włączające	Kryteria wyłączające
1. Dobra widoczność całego ciała lub fragmentu istotnego dla danego typu materiału pornograficznego. 2. Rozdzielczość zdjęcia wynosząca co najmniej 800 x 600 px. 3. Zdjęcia w formacie .jpeg lub .jpg.	1. Niewyraźny obraz uniemożliwiający identyfikację typów obiektów przedstawionych na zdjęciach. 2. Rozdzielczość zdjęcia mniejsza niż 800 x 600 px. 3. Zdjęcia w innych formatach, w szczególności grafiki wektorowe.

Każdy materiał pornograficzny został opisany przez psychologów seksuologów pod kątem właściwości danego zdjęcia (zob. tab. 2), ponadto każde zdjęcie otrzymało unikalny numer identyfikacyjny.

Tabela 2. Obecność obiektów klasyfikowanych na zdjęciach przez psychologów seksuologów⁴

Kategoria	Czynniki identyfikacyjne
Wspólne dla wszystkich kategorii	a) numer ID danego zdjęcia; b) liczba osób; c) liczba kobiet; d) liczba mężczyzn; e) brak danych o płci; f) obecność pochwy; g) obecność penisa; h) obecność pośladków; i) obecność odbytu; j) obecność piersi; k) penetracja; l) aktywność seksualna; m) elementy ubioru
Kategoria „AB/DL”	a) strój dziecięcy; b) butelka; c) smoczek; d) zabawki – zabawki pluszowe; e) zabawki – inne; f) pielucha; g) dziecięce upięcie włosów
Kategoria „Akrotomofilia”	a) brak kończyny górnej; b) brak obydwu kończyn górnych; c) brak kończyny dolnej; d) brak obydwu kończyn dolnych; e) kule; f) wózek inwalidzki; g) proteza kończyny dolnej; h) protezy obydwu kończyn dolnych; i) deformacje
Kategoria „BDSM”	a) maszyna; b) lina; c) łańcuch; d) kajdanki; e) obroża; f) pejcz; g) opaska na oczy; h) maska; i) packa; j) szpicruta; k) ostroga; l) klamry na sutki; m) naciągacz jąder; n) pas cnoty; o) knebel
Kategoria „Fetyszyzm – stopy”	a) jedna stopa; b) dwie stopy; c) stopa lub stopy uczestniczące w aktywności seksualnej
Kategoria „Fetyszyzm – szpilki”	a) buty na obcasach
Kategoria „Fetyszyzm – lateks”	a) ubranie lateksowe
Kategoria „Fetyszyzm – tatuaż”	a) tatuaże lub tatuaż
Kategoria „Fetyszyzm – podkolanówki”	a) podkolanówki; b) zakolanówki
Kategoria „Fetyszyzm – grawiditofilia”	a) wyraźnie widoczne oznaki ciąży

⁴ Podstawowa charakterystyka materiałów pornograficznych zostanie przedstawiona w odrębnym artykule ze względu na ilość danych oraz ich znaczenie poznawcze dla zrozumienia kluczowych cech obiektów przedstawionych na materiałach pornograficznych.

Stworzona w ten sposób baza danych obejmowała zatem 4500 materiałów, z których każdy został w dokładny sposób opisany oraz zweryfikowany pod kątem jakości.

W celu analizy charakterystyki przedmiotów oraz sytuacji występujących na zdjęciach zaprezentowano wybrane statystyki opisowe dla poszczególnych typów materiałów pornograficznych. Część charakterystyk była wspólna dla różnych kategorii materiałów pornograficznych (zob. tab. 2), natomiast część była unikalna dla poszczególnych klas. Szczegółowa analiza informacji oraz obiektów przedstawionych na materiałach pornograficznych wykracza poza ramy niniejszego artykułu, w tabeli 3 przedstawiono zatem podstawowe statystyki opisowe dla wspólnych klas.

Tabela 3. Podstawowe statystyki opisowe dotyczące liczby osób przedstawionych na zdjęciach

	n	Minimum	Maksimum	Średnia	Odchylenie standardowe
Liczba osób	4500	1	10	1,1831	0,48105
Liczba kobiet	4500	0	10	0,9671	0,41293
Liczba mężczyzn	4500	0	2	0,1089	0,31929

Etap 2. Wybór odpowiedniego modelu pretrenowanego

W celu wyboru modelu pretrenowanego, który cechowałby się najbardziej właściwymi parametrami procesu trenowania, przeprowadzono serie badań, testując wybrane modele.

Do badań zaklasyfikowano sześć modeli opartych na różnych architekturach konwolucyjnych sieci neuronowych. Każdy z nich został wytrenowany na tej samej grupie zdjęć, a proces treningu odbywał się z wykorzystaniem biblioteki fast.ai [17]. Zdjęcia zostały poddane standardowej augmentacji zaimplementowanej w bibliotekę fast.ai, w tym zakresie żaden z hiperparametrów nie był modyfikowany. Każda seria badawcza (przeznaczona dla danego modelu pretrenowanego) składała się z 50 cykli treningu modelu. W tabeli 4 przedstawiono końcowe informacje dotyczące straty trenowania, straty walidacji oraz dokładności klasyfikacji⁵.

Tabela 4. Przegląd charakterystyki procesu trenowania z uwzględnieniem modeli pretrenowanych

Architektura	Strata trenowania	Strata walidacji	Dokładność klasyfikacji
ResNet152 [18]	0,044511	0,390854	89,33%
ResNet101 [19]	0,058082	0,478085	87,33%
VGG19 [20]	0,180486	0,476862	85,22%
VGG16 [20]	0,167459	0,502113	85,00%

⁵ Gdy obserwowano oznaki przetrenowania, prezentowano dane odnoszące się do ostatniego cyklu uczenia się, który poprzedzał przetrenowanie.

Squeezenet 1.1 [21]	0,406253	0,840177	73,22%
Squeezenet 1.0 [21]	0,398491	0,799043	74,11%

Na podstawie wyników zaprezentowanych w tabeli 4 za najlepszy model można uznać ResNet152, wytypowano go zatem do dalszego etapu prac.

Etap 3. Wyłączenie wybranych kategorii materiału badawczego

W celu określenia, czy poszczególne kategorie wpływają negatywnie na dokładność modelu, przeprowadzono serię treningów sieci z wykorzystaniem modelu wskazanego empirycznie w poprzednim etapie, z wyłączeniem poszczególnych kategorii materiałów pornograficznych. W tabeli 5 zaprezentowano informacje końcowe dotyczące tego eksperymentu.

Tabela 5. Proces treningu sieci z wykorzystaniem modelu ResNet152 bez poszczególnych kategorii

Kategorie usunięte z modelu na czas treningu	Strata trenowania	Strata walidacji	Dokładność klasyfikacji
„AB/DL”	0,04375	0,444737	88,22%
„Akrotomofilia”	0,03379	0,357016	91,10%
„BDSM”	0,04756	0,329997	90,72%
„Fetyszyzm – stopy”	0,03771	0,445882	89,09%
„Fetyszyzm – grawiditofilia”	0,03210	0,433339	89,47%
„Fetyszyzm – szpilki”	0,03824	0,376584	90,10%
„Fetyszyzm – podkolanówki”	0,03755	0,342921	90,97%
„Fetyszyzm – lateks”	0,04520	0,535365	88,00%
„Fetyszyzm – tatuaże”	0,03572	0,504898	88,47%

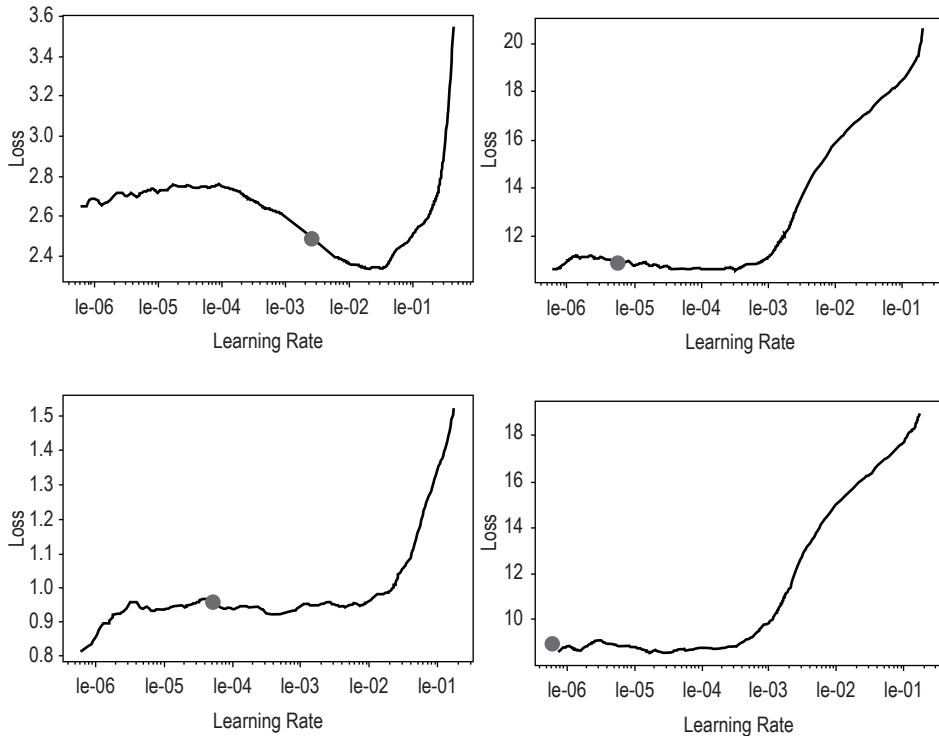
Jak wynika z analizy wyników przedstawionych w tabeli 5, można stwierdzić, że wyłączenie poszczególnych kategorii nie wpływa znacznie na dokładność różnicowania między danymi typami materiałów pornograficznych. Nie ma zatem podstaw do usuwania ze zbioru któregośkolwiek z nich, ponieważ żaden nie obniża znacznie jakości modelu.

Etap 4. Dostosowywanie hiperparametrów ostatecznego modelu

W celu dostrojenia parametrów uczenia się sieci przeprowadzono obliczenia mierzące do wyznaczenia optymalnego współczynnika uczenia *learning rate* [22]. Procedurę tę powtarzano po każdej serii pięciu cykli treningu sieci. Przeprowadzono zatem łącznie cztery etapy, a każdy z etapów składał się z pięciu cykli trenowania.

Na rysunku zamieszczono wykresy przedstawiające optymalną wartość współczynnika uczenia, którą wyznaczono automatycznie w miejscu najbardziej stromego

gradientu (zaznaczony punkt na wykresie; odpowiednio: $1e-03$, $1e-05$, $1e-04$, $1e-06$). Po czwartym etapie przzerwano proces treningu sieci, ponieważ zaobserwowano oznaki przetrenowania.



Rysunek. Wykresy przedstawiające optymalny współczynnik learning rate

Współczynnik uczenia (ang. *learning rate*) wskazuje optymalny kierunek aktualizacji wag, połączeń między sztucznymi neuronami, tak aby ilość błędów popełnianych przez sieć (ang. *loss*) była możliwie najniższa.

Tabela 6. Końcowe wartości treningu sieci po dostrojeniu hiperparametrów

Etapy	Strata trenowania	Strata walidacji	Dokładność klasyfikacji
Etap 1	1,084223	0,448046	85,89%
Etap 2	0,895491	0,357969	90,11%
Etap 3	0,865549	0,322195	91,00%
Etap 4	0,848440	0,304644	91,88%

Dyskusja

Celem badań było wytrenowanie modelu sieci neuronowej do klasyfikacji zadanych typów materiałów pornograficznych. Ukazano proces doboru materiału badawczego, przedstawiono empiryczne uzasadnienie dla wyboru modelu pretrenowanego Res-Net152 oraz wskazano, że zasadne jest wykorzystanie wszystkich kategorii typów materiałów pornograficznych. W ostatnim etapie dokonano modyfikacji hiperparametrów modelu. Ostatecznie uzyskano model zdolny do różnicowania dziewięciu typów materiałów pornograficznych z dokładnością bliską 92%, co można uznać za wysoki wynik.

W modelu pilotażowym [10] materiał zdjęciowy wybierano przez wyszukiwarke Google, posługując się odpowiednimi słowami kluczowymi, które dotyczyły określonych wcześniej rodzajów materiałów pornograficznych. Z tak utworzonych plików ręcznie wybrano zdjęcia, które posłużyły do wytrenowania modelu. Model umożliwiał różnicowanie między siedmioma typami pornografii z blisko 70% dokładnością. Znaczna liczba błędów w klasyfikacji wynikała z błędnego przypisania danego zdjęcia do odpowiedniej kategorii [10].

Aby zminimalizować liczbę tego rodzaju błędów i podwyższyć dokładność modelu w niniejszym badaniu pobrany z internetu materiał zdjęciowy przyporządkowywany był do odpowiedniej kategorii przez specjalistów, którzy posługiwali się restrykcyjnymi kryteriami włączającymi i wyłączającymi w ocenie materiału badawczego. Każde zdjęcie było przez nich opisywane pod względem obecności klasyfikowanych obiektów. Następnie eksperymentalnie wybrano model pretrenowany, który pozwalał na największą dokładność klasyfikacji.

Rozwój prac nad głębokimi sieciami neuronowymi w seksuologii wydaje się szczególnie obiecujący z co najmniej dwóch powodów. Po pierwsze, możliwe jest stworzenie narzędzia przeznaczonego do automatycznego wykrywania materiałów pornograficznych z małoletnimi i zastosowanie go w trakcie postępowania sądowego. Prezentowany w niniejszej pracy model nie mógłby posłużyć (bez modyfikacji) do różnicowania między materiałami pornograficznymi z udziałem małoletnich a materiałami z udziałem osób dorosłych, jednak w przyszłości (z użyciem innego materiału treningowego) klasyfikacja taka byłaby możliwa. Taki model mógłby posłużyć także do szacowania wieku osób występujących w tego rodzaju materiałach [23] czy w opiniowaniu materiałów pornograficznych dla sądu [15]. W efekcie możliwe będzie ulepszenie modeli, które pozwolą na różnicowanie między materiałami pornograficznymi z udziałem dzieci a innym typem materiałów, co znajdzie swoje zastosowanie w sprawach karnych związanych z uzyskiwaniem dostępu do treści pornograficznych z udziałem dzieci.

Po drugie, po dotrenowaniu zaprezentowanego modelu o zdjęcia kobiet oraz mężczyzn, którzy nie podejmują aktywności seksualnej, czy też o zdjęcia przedmiotów niezwiązanych z ludźmi, można by wykorzystać model do filtrowania treści nieodpowiednich do oglądania przez małoletnich. Jest to istotne, ponieważ materiały pornograficzne są powszechnie dostępne w internecie [24], a dzieci mają z nimi często kontakt nieintencjonalnie [25]. Istnieją już systemy pozwalające na filtrowanie mate-

riałów pornograficznych z udziałem osób dorosłych w bezpiecznych przeglądarkach dla dzieci⁶, niemniej wartością dodaną niniejszego modelu jest możliwość różnicowania między materiałami fetyszystycznymi. Proste modele sieci neuronowych, które pozwalają na różnicowanie między podejmowaniem aktywności seksualnej lub nagością a brakiem tych stanów, mogłyby zostać uzupełnione o subtelniejszą analizę, obejmującą również takie materiały, które wydają się nieodpowiednie dla małoletnich, ale nie przedstawiają nagości lub aktywności seksualnej (materiały fetyszystyczne).

Po trzecie, model umożliwia nie tylko dokonywanie prostych klasyfikacji danych zdjęć do odpowiedniej kategorii, ale też pozwala jednocześnie na wyszukiwanie zdjęć, które nie odpowiadają poznanym kategoriom. Niewykluczone, że dzięki zastosowaniu takiego podejścia w przyszłości możliwe będzie wyszukanie nieopisanych jeszcze materiałów pornograficznych, przykładowo wiążących się z rzadkimi parafiliami. Wykorzystanie podejścia tego rodzaju będzie wymagało zmodyfikowania architektury sieci neuronowej do innego typu zadań – nie klasyfikacji, lecz wyszukiwania najbardziej zbliżonych wzorców w danym typie obrazów, co wydaje się interesującym kierunkiem badań.

Po czwarte, jak już wspomniano, tworzenie modeli tego rodzaju może pozwolić w przyszłości na opracowanie lepszego materiału bodźcowego dla zadania Stroopa w warunkach emocjonalnym [16]. Zadaniem pacjenta jest w nim jak najszybsze zareagowanie (najczęściej przez naciśnięcie odpowiedniego klawisza) w odpowiedzi na bodźce wizualne, które pojawiają się na ekranie. Tworzenie wielu warunków eksperymentalnych wymaga długotrwałego procesu doboru różnorodnych zdjęć, co mogłoby zostać zautomatyzowane. W konsekwencji mógłby zostać przygotowany taki materiał, który pozwoliłby osiągnąć lepsze rezultaty o znaczeniu klinicznym i wykorzystać je w powyższym zadaniu.

Adnotacja

Model sAI 0.3 wraz z dokumentacją został zarejestrowany w bazie OSF i jest dostępny pod następującym numerem DOI: 10.17605/OSF.IO/M2X6P.

Piśmiennictwo

1. Short MB, Black L, Smith AH, Wetterneck CT, Wells DE. *A review of internet pornography use research: Methodology and content from the past 10 years*. *Cyberpsychol. Behav. Soc. Netw.* 2012; 15(1): 13–23.
2. Scott AD. *Pornografia. Jej wpływ na rodzinę, społeczeństwo, kulturę*. Gdańsk: Human Life International; 1995.
3. Hald GM. *Gender differences in pornography consumption among young heterosexual Danish adults*. *Arch. Sex. Beh.* 2006; 35(5): 577–585.
4. Gola M. *Neuronalne mechanizmy nalogowych zachowań*. W: Habrat B red. *Zaburzenia uprawiania hazardu i tzw. nalogi behawioralne*. Warszawa; Wydawnictwo IPiN [w druku].

⁶ Takich przeglądarek internetowych, które przed wyświetleniem danej strony dziecka skanują ją w poszukiwaniu materiałów mogących być nieodpowiednimi dla dzieci.

5. Hald GM, Malamuth NM. *Self-perceived effects of pornography consumption*. Arch. Sex. Behav. 2008; 37(4): 614–625.
6. Quadara A, El-Murr A, Latham J. *The effects of pornography on children and young people*. Melbourne: Australian Institute of Family Studies; 2017.
7. Nawrocka M, Drozd M, Maszczyk A, Gołaś A. *Wprowadzenie do predykcji z wykorzystaniem sztucznych sieci neuronowych i metod statystycznych*. Ogrody Nauk i Sztuk 2015; 5: 203–211.
8. Posner MI, Rothbart MK. *Research on attention networks as a model for the integration of psychological science*. Ann. Rev. Psychol. 2007; 58(1): 1–23.
9. Read SJ, Monroe BM, Brownstein AL, Yang Y, Chopa G, Miller LC. *A neural network model of the structure and dynamics of human personality*. Psychol. Rev. 2010; 117(1): 61–92.
10. Oronowicz-Jaśkowiak W. *Pilotażowy model sieci neuronowej do zastosowań związanych z klasyfikacją siedmiu typów materiałów pornograficznych*. Przegląd Seksuologiczny 2019; 1(55): 21–31.
11. Lew-Starowicz Z. *Seksuologia sądowa*. Warszawa: Wydawnictwo Lekarskie PZWL; 2000.
12. Wang Y, Kosiński M. *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*. J. Pers. Soc. Psychol. 2018; 114(2): 246–257.
13. Perez M, Avila S, Moreira D, Moraes D, Testoni V, Valle E i wsp. *Video pornography detection through deep learning techniques and motion information*. Neurocomputing 2017; 230: 279–293.
14. Oronowicz-Jaśkowiak W. *SexACT database v. 0.1*. www.sexailab.pl (dostęp: 2.01.2020).
15. Oronowicz-Jaśkowiak W. *The application of neural networks in the work of forensic experts in child abuse cases*. Postępy Psychiatr. i Neurol. 2019; 4(28): 273–282.
16. Oronowicz-Jaśkowiak W, Lew-Starowicz M. *Polish adaptation of emotional Stroop test in assessment of pedophilia – A pilot study*. Psychiatr. Pol. 2021; 55(1): 85–100.
17. Jeremy H. *Fast.ai software library*. www.fast.ai (dostęp: 10.07.2019).
18. He K, Zhang X, Ren S, Sun J. *ResNet-152*. www.kaggle.com (dostęp: 10.07.2019).
19. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition*. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. S. 770–778.
20. Simonyan K, Zisserman A. *Very deep convolutional networks for large-scale image recognition*. ArXiv preprint 2014. www.arxiv.org (dostęp: 10.07.2019).
21. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size*. ArXiv preprint 2016. www.arxiv.org (dostęp: 10.07.2019).
22. Smith LN. *Cyclical learning rates for training neural networks*. IEEE Winter Conference on Applications of Computer Vision, 2017. S. 464–472.
23. Yang TY, Huang YH, Lin YY, Hsiu PC, Chuang YY. *SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation*. International Joint Conferences on Artificial Intelligence 2018; 5(6).
24. Wolak J, Mitchell K, Finkelhor D. *Unwanted and wanted exposure to online pornography in a national sample of youth Internet users*. Pediatrics 2007; 119(2): 247–257.
25. Makaruk K, Włodarczyk J, Michalski P. *Kontakt dzieci i młodzieży z pornografią*. Warszawa: Fundacja Dajemy Dzieciom Siłę; 2017.

Otrzymano: 10.03.2020

Zrecenzowano: 12.03.2021

Otrzymano po poprawie: 18.03.2021

Przyjęto do druku: 26.05.2021