Dr inż. Artur Wilkowski
Instytut Automatyki i Informatyki Stosowanej
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska
ORCID: 0000-0002-6814-7645
e-mail: artur.wilkowski@pw.edu.pl

METODY ROZPOZNAWANIA AKTYWNOŚCI O CHARAKTERZE PORNOGRAFICZNYM W SEKWENCJACH WIDEO PRZY UŻYCIU NISKOPOZIOMOWYCH CECH OBRAZU

Streszczenie

Duża dostępność wszelkiego rodzaju materiałów publikowanych w Internecie, stwarza potrzebę mechanizmów kontroli treści, tak by trafiały one tylko do osób uprawnionych oraz będących chętnymi odbiorcami takich treści. Jednym ze szczególnie wrażliwych rodzajów treści są materiały wideo o charakterze pornograficznym, do których dostęp powinien mieć bardzo selektywny charakter. Do praktycznej realizacji tego celu niezbędne jest wypracowanie metod automatycznej klasyfikacji takich treści. Rozpoznawanie pornograficznego charakteru materiałów wideo jest przypadkiem szczególnym szerszego problemu rozpoznawania aktywności (HAR – Human Activity Recognition). Artykuł podejmuje się zadania przedstawienia technologii informatycznych umożliwiających klasyfikację materiałów wideo ze szczególnym uwzględnieniem danych pornograficznych. Przedstawione są metody klasyczne oraz najnowsze metody wykorzystujące uczenie głębokie. Artykuł skupia się na rozwiązaniach wykorzystujących niskopoziomowe (w niewielkim stopniu przetworzone) cechy obrazu.

Słowa kluczowe: Rozpoznawanie treści pornograficznych w wideo, Rozpoznawanie aktywności ludzkiej, Przetwarzanie wideo, Sieci Neuronowe.

WSTĘP

Ze względu na intensywny rozwój serwisów udostępniających dane multimedialne oraz serwisów społecznościowych, pojawia się potrzeba tworzenia automatycznych systemów rozpoznających i klasyfikujących obrazy oraz sekwencje wideo zawierające treści seksualne. Systemy takie znajdują swoje zastosowanie w narzędziach kontroli rodzicielskiej, narzędziach nadzorczych portali społecznościowych, kontroli treści wysyłanych przez komunikatory (sexting) i ogólnie tworzeniu filtrów treści przeznaczonych dla poszczególnych grup społecznych (praca, szkoła) oraz wiekowych (dzieci, nastolatki). Systemy rozpoznające treści pornograficzne mają również swoje zastosowanie jako jeden z etapów rozpoznawania treści o charakterze nielegalnym (np. typu CSAM – Child Sexual Abusive Material).

Użycie dowolnych metod uczenia maszynowego w celu ewaluacji tego typu materiałów zasadniczo wymaga ekstrakcji z poszczególnych klatek oraz ich sekwencji interesujących informacji, tzw. cech, które są istotne z punktu widzenia realizowanego zadania. Cechy takie mogą być ekstrahowane na podstawie wiedzy eksperckiej (hand-crafted features), ale mogą też być realizowane w sposób automatyczny (np. cechy sieci splotowych - CNN features). Na przykład w przypadku rozpoznawania aktywności o charakterze seksualnym istotnymi cechami wydaje się być kolor (odpowiadający kolorowi skóry) oraz powtarzalność ruchów. Rozwiązania agregujące takie proste cechy w ramach obrazów lub sekwencji osiągają dosyć dobrą skuteczność na danych pornograficznych, jednak stanowią spore uproszczenie problemu i zawodzą w przypadku trudniejszych przykładów (np. zapasy – gdzie kolor skóry dominuje lub skakanie na skakance – gdzie mamy do czynienia z powtarzalną aktywnością).

Należy mieć na uwadze, że rozpoznawanie aktywności seksualnych stanowi pewien przypadek szczególny uniwersalnego problemu rozpoznawania aktywności, dla którego proponowane są dosyć skomplikowane i efektywne wektory cech, które mogą być również z powodzeniem stosowane w tym węższym problemie. W niniejszej pracy zrealizowany zostanie przegląd efektywnych cech, które są wykorzystywane w rozpoznawaniu aktywności w sekwencjach wideo ze szczególnym uwzględnieniem cech używanych do rozpoznawania aktywności o charakterze seksualnym. Badane będą zastosowania cech o charakterze niskopoziomowym, tzn. nie analizującym w sposób jawny pozy sylwetki ludzkiej.

W kolejnych sekcjach przedstawione zostaną specyficzne cechy, które były z powodzeniem używane w problemie rozpoznawania sekwencji obrazów o charakterze pornograficznym oraz opierające się na nich metody rozpoznawania.

ROZPOZNAWANIE TREŚCI PORNOGRAFICZNYCH W MATERIAŁACH WIDEO

Rozpoznawanie filmów o charakterze pornograficznym stanowi przypadek szczególny bardziej ogólnego problemu rozpoznawania aktywności (Human Activity Recognition). W zasadzie wszystkie cechy, które są wykorzystywane w problemie rozpoznania aktywności, mogą być użyte w rozpoznawaniu treści pornograficznych. Ze względu na charakter treści, niekiedy używane są cechy specjalizowane np. koncentrujące się na obszarach skóry lub ruchu cyklicznym.

ECHY OPARTE O WIEDZĘ EKSPERCKĄ

Kolor skóry oraz cykliczna charakterystyka ruchu to istotne cechy, które mogą być wykorzystane do rozpoznawania treści o charakterze pornograficznym[1]. W cytowanej pracy jest wykorzystywany bardzo szeroki zestaw cech opartych w dużej mierze na rozpoznawaniu koloru skóry. W artykule proponowane są trzy kategorie cech o charakterze strukturalnym:

1. Cechy ekstrahowane z pojedynczej klatki obrazu, w których stosowane są cechy geometryczne wykrytych obszarów skóry. Cechy ekstrahowane są dla

[1] A. Behrad, M. Salehpour, M. Saeidi, M., Obscene Video Recognition Using Fuzzy SVM and New Sets of Features, International Journal of Advanced Robotic Systems, tom 10 (2), 2013, doi:10.5772/55517.

wszystkich klatek, dlatego dodatkowo stosowane są metody redukcji rozmiaru danych (PCA/LDA).
2. Cechy ekstrahowane z sekwencji obrazów. W ramach tych cech tworzone są przestrzenno-czasowe wolumeny obszarów skóry. Wolumeny tworzone są za pomocą algorytmów grafowych. Ekstrahowane są 2 rodzaje cech: cechy geometryczne wolumenu przestrzenno-czasowego oraz cechy geometryczne rzutu wolumenu na płaszczyznę.
3. Cechy opisujące ruch cykliczny. Uwzględniane są cechy ruchu cyklicznego w ramach obszarów skóry, po wyeliminowanie sekwencji, w których ruch wynika z ruchu kamery. Cykliczność ruchu wykrywana jest za pomocą metody korelacji z ramką referencyjną i analizą cech częstotliwościowych (obliczonych przy pomocy transformaty Fouriera).

Uzyskane wektory cech analizowane są przy pomocy ważonego klasyfikatora SVM.

Okazuje się, że cechy wykorzystywane do wykrywania treści pornograficznych niekoniecznie muszą odzwierciedlać zawartość materiału, ale mogą opierać się na analizie montażu filmu[2]. Autorzy cytowanej pracy badają częstość wystąpień zmian sceny, przyjmując założenie, że profesjonalne materiały filmowe charakteryzują się częstszymi cięciami i zmianami sceny, podczas gdy filmy pornograficzne (częściej niskobudżetowe) charakteryzują się dłuższymi scenami i mniejszą liczbą cięć.

CECHY O DESKRYPTORY LOKALNE

Klasyczne metody rozpoznawania aktywności oparte np. o Bag-of-Visual-Works (BoVW)[3] dobrze sprawdzają się również w wykrywaniu nagości w wideo. W przywołanym rozwiązaniu, w celu redukcji rozmiaru danych i zmniejszenia redundancji danych, w sekwencji wykrywane są najpierw klatki kluczowe. Realizowane jest to przez analizę koloru (komponent Hue przestrzenie HSV), a dokładniej porównanie histogramów tej cechy za pomocą narzędzia „przecięcia histogramów". Po wyborze klatek kluczowych stosowana jest metoda ekstrakcji cech BoVW oparta o kilka rodzajów deskryptorów (SURF, ORB, Opponent-SIFT, SIFT, Hue SIFT, BRIEF) oraz klasyfikator SVM.

Oprócz klasycznej agregacji BoVW stosowane są też jej rozwinięcia np. BossaNova[4]. Podczas, gdy w klasycznej metodzie BoVW dla każdego obrazu określana jest liczba przyporządkowań punktów charakterystycznych do danego słowa kodowego (ze słownika), tutaj dla każdego słowa kodowego jest tworzony odrębny histogram, do którego „wpadają" deskryptory znajdujące się w określonym przedziale odległości od słowa kodowego. W ten sposób uzyskuje się znacznie bogatszą reprezentację obrazu. W dyskutowanym rozwiązaniu dane są też

[2] R. Mustafa, D. Zhu, A novel method for sensing obscene videos using scene change detection, Telkomnika Indonesian J. Electr. Eng. 13(2), s. 300–304, 2015.

[3] H. Yatawatte, A. Dharmaratne, Content Based Video Retrieval for Obscene Adult Content Detection, Neural Information Processing. ICONIP 2015, Lecture Notes in Computer Science, tom. 9492. Springer, Cham., ed. S. Arik, T. Huang, W. Lai, Q. Liu, 2015, https://doi.org/10.1007/978-3-319-26561-2_46

[4] S. Avila, N. Thome, M. Cord, E. Valle, A. d. A. Araújo, Pooling in image representation: The visual codeword point of view, Computer Vision and Image Understanding, tom 117(5), s. 453-465, 2013, https://doi.org/10.1016/j.cviu.2012.09.007

wzbogacane o „klasyczne" cechy BoVW oraz normalizowane. Metoda wykorzystuje lokalne deskryptory HueSIFT (SIFT wzbogacony o dane o kolorze) w procesie agregacji oraz klasyfikator SVM działający na klatkach kluczowych ekstrahowanych z materiału wideo. Metoda uzyskuje bardzo dobre wyniki na zaprezentowanym zbiorze danych pornograficznych „NPDI Pornography Database" obejmującym 800 nagrań wideo (skuteczność 89.5%+-1%).

Koncepcja agregacji BossaNova została również zaadoptowana do szerszej grupy deskryptorów (BRIEF, ORB, BRISK, FREAK, BinBoost16)[5]. W cytowanej pracy dodatkowo głosowanie większościowe, realizowane dla klatek wideo, jest zastąpione przez ocenę pojedynczego deskryptora obliczonego na podstawie deskryptorów poszczególnych klatek za pomocą mediany. Umożliwia to dalsze zwiększenie skuteczności do 90.9%+-1% (dla wersji deskryptora BinBoost16) na zbiorze „NPDI Pornography Database". Dodatkowo powiększenie słownika słów kodowych BinBoost16 umożliwiło uzyskanie skuteczności 92.4%+-2% dla deskryptora BoVW oraz 92.0%+-1% dla deskryptora BossaNova w przypadku użycia agregacji deskryptora w czasie za pomocą mediany[6].

W zagadnieniu rozpoznawania materiałów pornograficznych porównywane też są rożne konfiguracje cech czasowo-przestrzenno-skalowych[7]. W cytowanej pracy oprócz porównania istniejących detektorów/deskryptorów takich cech: STIP, DTRACK (czyli gęste trajektorie) oraz referencyjnej metody SURF dla klatek kluczowych, proponuje się nowy, szybki detektor/deskryptor TRoF charakteryzujący się obniżoną sygnaturą pamięci. Detektor TRoF stanowi dosyć wierną adaptację znanego detektora SURF do przestrzeni czasowo-przestrzennej, deskryptor natomiast w istocie wykorzystuje 3 deskryptory SURF osadzone w centrum wykrytego owalu (blob) i działające wzdłuż 3 głównych płaszczyzn układu współrzędnych. Dodatkowo badane jest połączenie detektora DTRACK z dekskryptorem TRoF (czyli DTRoF). Deksryptory lokalne są agregowane w czasie i przestrzeni za pomocą wektorów Fishera i klasyfikowane przy pomocy klasyfiktatora SVM. Działanie algorytmów było weryfikowane na rozszerzonej bazie danych „NPDI Pornography Database" obejmującej 2000 filmów określanej jako zbiór danych „Pornography-2k". Najlepsze skuteczności na badanym zbiorze osiągnęła metoda gęstych trajektorii (DTRACK) – 95.76% oraz metoda DTRoF – 95.58%.

CECHY GENEROWANE PRZEZ SIECI GŁĘBOKIE SIECI NEURONOWE

W zadaniu rozpoznawania pornograficznych materiałów wideo, mogą mieć zastosowanie podstawowe architektury splotowych sieci neuronowych, które wcześniej były z powodzeniem używane w rozpoznawaniu obrazów[8]. Badane

[5] C. Caetano, S. Avila, S. Guimarães, A. d. A. Araújo, Pornography detection using BossaNova video descriptor, 2014 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, pp. 1681-1685, 2014.

[6] C. Caetano, S. AvilA, W.R. Schwartz, S.J.F. Guimarães, A. d. A. Araújo, A mid-level video representation based on binary descriptors: A case study for pornography detection, Neurocomputing, Tom 213, s. 102-114, 2016, https://doi.org/10.1016/j.neucom.2016.03.099

[7] D. Moreira, S. Avila, M. Perez, D. MoraeS, V. Testoni, E. Valle, S. Goldenstein, A. Rocha, Pornography classification: The hidden clues in video space–time, Forensic Science International, Tom 268, s. 46-61, 2016, https://doi.org/10.1016/j.forsciint.2016.09.010

[8] M. Moustafa, Applying deep learning to classify pornographic images and videos, Pacific Rim Symposium on Image and Video Technology (PSIVT), 2015.

w przywołanej pracy sieci to ANet (oparty o sieć AlexNet), GNet (oparty o sieć GoogLeNet) oraz AGNet – realizujący późną fuzję wyników dwu poprzednich sieci (używając agregacji średniej i maksimum). W celu polepszenia wyników zastosowano transfer wiedzy ze zbioru ImageNet. Raportowana skuteczność dla zbioru „NPDI Pornography Database" sięga 94.1%+-2% dla sieci używającej późnej fuzji z funkcją maksimum.

Sposób agregacji różnego rodzaju cech dostępnych w wideo jest wyzwaniem dyskutowanym w literaturze[9]. W cytowanej pracy mamy do czynienia z porównaniem kilku zastawów cech wejściowych sieci GoogLeNet oraz różnych strategii agregacji przetworzonych cech. Dodatkowo, oceniana jest użyteczność wykorzystania transferu wiedzy z bazy danych ImageNet. Badane są cechy RGB, skali szarości, cechy ruchu – generowane za pomocą przepływu optycznego oraz wektorów ruchu obecnych w kompresji filmów FMPEG. Modele fuzji obejmują:

- fuzję wczesną – poszczególne rodzaje cech są sklejane przed wysłaniem ich do sieci neuronowej
- fuzję pośrednią – poszczególne rodzaje cech są niezależnie przetwarzane przez warstwy splotowe sieci, następnie agregowane w ramach sekwencji oraz klasyfikowane
- fuzję późną - poszczególne rodzaje cech są niezależnie przetwarzane przez warstwy splotowe sieci, klasyfikowane, a agregacja w ramach sekwencji obejmuje już wyniki działania klasyfikatora (oceny poszczególnych klatek).

Klatki w sekwencjach próbkowane są z częstotliwością 1 Hz. W większości przypadków (oprócz wczesnej fuzji RGB+flow) wykorzystywany jest transfer wiedzy, a do klasyfikacji liniowy klasyfikator SVM. Eksperymenty pokazały, że najskuteczniejszą architekturą jest kombinacja informacji RGB i informacji o ruchu wykorzystująca mechanizm późnej fuzji oraz transfer wiedzy (skuteczność 96.4% na zbiorze „Pornography-2k"). Porównywalne wyniki osiągane były jednak również przez pozostałe modele fuzji.

Sieci neuronowe splotowe mogą być wspomagane przez sieć rekurencyjną w celu lepszego uchwycenia zależności czasowych w sekwencjach[10]. W cytowanej pracy analizowane są warianty sieci splotowej ResNet oraz GoogLeNet, uczonych na bazie ImageNet. Nie jest stosowany tuning tej części sieci. Wstępnie przetworzone cechy są następnie analizowane przez sieć rekurencyjną LSTM, która dokonuje klasyfikacji krótkiej podsekwencji obrazów i jest uczona na danych uczących. Raportowane wyniki wskazują na skuteczność rzędu 95.6%+-1% dla architektury sieci splotowej ResNet-101 na „starszym" zbiorze „NPDI Pornography Database" obejmującym 800 sekwencji filmowych.

W nowszych architekturach sieci można również spotkać mechanizmy atencyjne[11]. Może to obejmować wykorzystanie modułu atencyjnego w celu lepszego

[9] M. Perez, S. Avila, D. Moreira, V. Testoni, E. Valle, S. Goldenstein, A. Rocha, Video pornography detection through deep learning techniques and motion information, Neurocomputing, tom 230, s. 279-293, 2017, https://doi.org/10.1016/j.neucom.2016.12.017

[10] J. Wehrmann, G.S. Simões, R.C. Barros, V.F. Cavalcante, Adult content detection in videos with convolutional and recurrent neural networks, Neurocomputing, tom 272, s. 432-438, 2018, https://doi.org/10.1016/j.neucom.2017.07.012

[11] A. Gangwar, V. González-Castro, E. Alegre, E. Fidalgo, AttM-CNN: Attention and metric learning based CNN for pornography, age and Child Sexual Abuse (CSA) Detection in images, Neurocomputing, tom 445, s. 81-104, 2021, ,https://doi.org/10.1016/j.neucom.2021.02.056

skupienia uwagi sieci na obszarach kluczowych w rozpoznaniu treści o charakterze seksualnym. Proponowana w cytowanej publikacji sieć operuje na pojedynczych klatkach obrazu. W sieci wykorzystane są znane z literatury moduły Inception, InceptionResnet oraz InceptionReduction. Oprócz tego proponowane jest użycie dwóch modułów atencyjnych. W obu przypadkach realizowana jest korelacja wyjściowego (wysokoprzetworzonego) wektora sieci z elementami przestrzennymi wolumenu z warstw wcześniejszych. Wyniki takiej korelacji tworzą mapę atencji, która następnie wykorzystana jest do nadania wag wybranym (najbardziej interesującym) obszarom sieci. Sieć jest uczona na nowym zbiorze Pornography-2M – obejmującym 2 miliony obrazów. Testy na zbiorze „Pornography-2K" wskazują na skuteczność rzędu 96.45% bez tuningu na zbiorze „Pornography-2K" oraz 97.1% po tuningu.

Model DVRGNet[12] jest jednym z najbardziej skutecznych w badanym zastosowaniu modeli łączących cechy sieci splotowych i rekurencyjnych. Wykorzystuje jednocześnie kilka znanych architektur sieci splotowych DenseNet, VGG, ResNet, GoogLeNet, wspieranych przez moduły Bi-LSTM. Poszczególne sieci uczone są niezależnie w sposób przypominający Boosting, a poszczególnym architekturom i przykładom uczącym przypisywane są wagi określające odpowiednio ich skuteczność lub trudność. Sieć działa na danych RGB oraz danych o ruchu (obliczone wektory ruchu dla makrobloków obrazu) i uzyskuje znakomitą skuteczność rzędu 99.42% na bazie „Pornography-2k".

Konkurencyjny model DeepHSAR[13] rozwiązuje bardziej złożone zadanie polegające na etykietowaniu wideo z uwzględnieniem wielu rodzajów aktywności seksualnej. System jest uczony częściowo na podstawie danych etykietowanych, częściowo nieetykietowanych. Wykorzystuje dwa potoki przetwarzania, pierwszy oparty o analizę obrazu jako całości (przy użyciu sieci ResNet z modułem atencyjnym) oraz potok bazujący na drobnoziarnistej analizie obrazu, wykorzystujący nienadzorowaną ekstrakcję fragmentów, ich klasteryzację i etykietowanie - takie etykiety służą następnie do rozwiązania podstawowego problemu nadzorowanego uczenia. Oba potoki działają równolegle i ostateczny wynik jest określany za pomocą fuzji danych. Metoda zastosowana na zbiorze „Pornography-2K" (po przemapowaniu etykiet) uzyskała bardzo dobre wynik 99.85% skuteczności.

Nowoczesne modele typu Transformer również stopniowo znajdują zastosowanie w rozpoznawaniu pornografii w filmach[14] [15]. W pierwszej cytowanej pracy stosowane są warstwy transformer wspomagane przez warstwę LSTM oraz warstwę w pełni połączoną. W badaniach na bazie „Pornography-2K" rozwiązanie uzyskuje skuteczność 99.6%. Z kolei w drugim artykule wykorzystano pre-

[12] K. Rautela, D. Sharma, V. Kumar, D. Kumar, DVRGNet: an efficient network for extracting obscenity from multimedia content, Multimed Tools Appl, tom 83, s. 28807–28825, 2024, https://doi.org/10.1007/s11042-023-16619-9

[13] A. Gangwar, V. González-Castro, E. Alegre, E. Fidalgo, A. Martínez-Mendoza, DeepHSAR: Semi-supervised fine-grained learning for multi-label human sexual activity recognition, Information Processing & Management, tom 61(5), 2024, https://doi.org/10.1016/j.ipm.2024.103800

[14] K. Rautela, D. Sharma, V. Kumar, D. Kumar, „Obscenity detection transformer for detecting inappropriate contents from videos", Multimed Tools Appl, tom 83, s. 10799–10814, 2024, https://doi.org/10.1007/s11042-023-16078-2

[15] D. Zhu, X. Shan, C. Wu, K. Yung, A.W. Ip, „Multi Frame Obscene Video Detection With ViT: An Effective for Detecting Inappropriate Content", Int. J. Semant. Web Inf. Syst., tom 20(1), s. 1–18, 2024, https://doi.org/10.4018/IJSWIS.359768

trenowany model ViT do analizy wideo na podstawie kilku sąsiednich klatek. W celu wykrycia interakcji pomiędzy klatkami jeden z tokenów (odpowiadający klasie) jest wymieniany pomiędzy modułami atencyjnymi (MHSA) w kolejnych etapach przetwarzania, a przed etapem finalnej klasyfikacji. Model uzyskuje stosunkowo dobrą skuteczność 96.2% na bazie danych „NPDI Pornography Database".

PODSUMOWANIE

Zrealizowany przegląd pokazuje główne obszary rozwoju algorytmów ekstrakcji cech nakierowanych na rozpoznawanie materiałów pornograficznych. Proste cechy wskazywane przez ekspertów (np. oparte o kolor skóry, czy ruch powtarzalny), jak również uniwersalne cechy obrazu i sekwencji oparte o deskryptory punktów charakterystycznych, zastępowane są przez zaawansowane metody analizy obrazu oparte o splotowe sieci neuronowe, a ostatnio o mechanizmy atencji. Rozwiązania tego typu, w szczególności wykorzystujące architekturę transformer, były wcześniej z powodzeniem używane do rozwiązywania uniwersalnego problemu rozpoznawania aktywności, w oparciu o wciąż rosnącą moc obliczeniową i dostępność coraz obszerniejszych zbiorów danych. Wdrożenie takich rozwiązań na grunt rozpoznawania materiałów pornograficznych jest naturalną konsekwencją rozwoju algorytmów.

Pytanie, czy użycie dodatkowych informacji o obecności koloru skóry, czy ruchu cyklicznego w obrazie wciąż może zwiększać skuteczność rozpoznania, poprzez umożliwienie sieci skupienie się na istotnych elementach obrazu lub sekwencji pozostaje jednak otwarte. Takie badania wydają się interesującym kierunkiem rozwoju istniejących algorytmów.

Bibliografia

Avila S., Thome N., Cord M., Valle E., Araújo A. d. A., „Pooling in image representation: The visual codeword point of view", Computer Vision and Image Understanding, tom 117(5), 2013, https://doi.org/10.1016/j.cviu.2012.09.007

Behrad A., Salehpour M., Saeidi M., Barati M. „Obscene Video Recognition Using Fuzzy SVM and New Sets of Features", International Journal of Advanced Robotic Systems, tom 10(2), 2013, doi:10.5772/55517

Caetano C., Avila S., Guimarães S., Araújo A. d. A., „Pornography detection using BossaNova video descriptor," 2014 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal.

Caetano C., Avila S., Schwartz W.R., Guimarães S.J.F., Araújo A. d. A., „A mid-level video representation based on binary descriptors: A case study for pornography detection", Neurocomputing, Tom 213, 2016, https://doi.org/10.1016/j.neucom.2016.03.099

Gangwar A., González-Castro V., Alegre E., Fidalgo E., „AttM-CNN: Attention and metric learning based CNN for pornography, age and Child Sexual Abuse (CSA) Detection in images", Neurocomputing, tom 445. 2021, ,https://doi.org/10.1016/j.neucom.2021.02.056.

Gangwar A., González-Castro V., Alegre E., Fidalgo E., Martínez-Mendoza A., „DeepHSAR: Semi-supervised fine-grained learning for multi-label human sexual activity recognition", Information Processing & Management, tom 61(5), 2024, https://doi.org/10.1016/j.ipm.2024.103800

Moreira D., Avila S., Perez M., Moraes D., Testoni V., Valle E., Goldenstein S., Rocha A., „Pornography classification: The hidden clues in video space–time,", Forensic Science International, Tom 268, 2016, https://doi.org/10.1016/j.forsciint.2016.09.010

Moustafa M., „Applying deep learning to classify pornographic images and videos", Pacific Rim Symposium on Image and Video Technology (PSIVT), 2015.

Mustafa R., Zhu D., „A novel method for sensing obscene videos using scene change detection," TELKOMNIKA Indonesian J. Electr. Eng. 13(2), 2015.

Perez M., Avila S., Moreira D., Testoni V., Valle E., Goldenstein S., Rocha A., „Video pornography detection through deep learning techniques and motion information", Neurocomputing, tom 230, https://doi.org/10.1016/j.neucom.2016.12.017

Rautela K., Sharma D., Kumar V., Kumar D., „DVRGNet: an efficient network for extracting obscenity from multimedia content," Multimed Tools Appl, tom 83, 2024, https://doi.org/10.1007/s11042-023-16619-9

Rautela K., Sharma D., Kumar V., Kumar D. „Obscenity detection transformer for detecting inappropriate contents from videos", Multimed Tools Appl, tom 83, 2024, https://doi.org/10.1007/s11042-023-16078-2

Wehrmann J, Simões G.S., Barros R.C., Cavalcante V.F., „Adult content detection in videos with convolutional and recurrent neural networks" ,Neurocomputing, tom 272, s. 432-438, 2018, https://doi.org/10.1016/j.neucom.2017.07.012

Yatawatte H., Dharmaratne A., „Content Based Video Retrieval for Obscene Adult Content Detection", Neural Information Processing. ICONIP 2015, Lecture Notes in Computer Science, tom. 9492. Springer, Cham., ed. Arik S., Huang T., Lai W., Liu Q., 2015, https://doi.org/10.1007/978-3-319-26561-2_46

Zhu D., Shan X., Wu C., Yung K., Ip A.W., „Multi Frame Obscene Video Detection With ViT: An Effective for Detecting Inappropriate Content", Int. J. Semant. Web Inf. Syst., tom 20(1), https://doi.org/10.4018/IJSWIS.359768

METHODS FOR RECOGNIZING PORNOGRAPHIC ACTIVITY IN VIDEO SEQUENCES USING LOW-LEVEL IMAGE FEATURES

Abstract

The high availability of all kinds of material published on the Internet, creates the need for the mechanisms to control the content, so that it goes only to authorized persons and those who are willing recipients of such content. One particularly sensitive type of content is pornographic video material, access to which should be highly selective. For the practical realization of this goal, it is necessary to develop methods of automatic classification of such content. Recognizing the pornographic nature of video materials is a special case of the broader problem of Human Activity Recognition (HAR). The article undertakes the task of presenting information technologies that make it possible to classify video materials with a special focus on pornographic data. Classical methods and the latest methods using deep learning are presented. The article focuses on solutions using low-level (low-processed) image features.

Keywords: Recognition of pornographic content in video, Human Activity Recognition, Video processing, Neural networks