

WERYFIKACJA AUTENTYCZNOŚCI KOLORÓW NA ZDJĘCIACH WYKONANYCH W TECHNICIE ANALOGOWEJ

Piotr SOKOŁOWSKI

Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki
Katedra Systemów Multimedialnych
tel.: (58) 347-13-01, e-mail: psokolow@multimed.org

Streszczenie: W artykule opisano zagadnienie odróżniania historycznych fotografii pomiędzy oryginalnie kolorowe a koloryzowane. Rozważono problem doboru zdjęć pod względem technologii, w jakiej zostały wykonane. Następnie wykorzystując sieci neuronowe już w części wyuczone na innych zbiorach danych, sprawdzono ich efektywność w rozwiązywaniu badanego problemu. Rozważono wpływ rozmiaru obrazu podanego na wejściu, architektury zastosowanej sieci, a także zestawu danych użytego do uczenia sieci i wyodrębniania cech. W rezultacie potwierdzono przydatność opracowanego zbioru do treningu sieci, a także zaobserwowano, że zwiększanie rozmiaru sieci nie przynosi dodatkowych korzyści. Uzyskana trafność rozróżniania sięgnęła ponad 92 %.

Słowa kluczowe: zbiory danych, kolorowa fotografia historyczna, uczenie maszynowe, uwiaryzalnianie.

1. WPROWADZENIE

Informacje o kolorystyce obiektów, zanim trafią za pośrednictwem zdjęć analogowych do zbiorów, takich jak opisane w literaturze [1][2][3][4][5][6], przebywają wieloetapową drogę, na której są narażone na przekłamania. Już na samym początku, podczas fotografowania, zapis, który zostanie utrwalony na kliszy zależy od minimum dwóch czynników, czyli od zastosowanej technologii, jak też od oświetlenia. Następnie skanuje się wywołane fotografie, lub w rzadkich przypadkach wprowadza się do komputera negatywy, które cyfrowo przekształcają się w pozytywy. W obu przypadkach występują dwa dodatkowe źródła nieporządkanych zmian. Pierwszy rodzaj przekłamań wynika z trwania przez analogowe nośniki swoich właściwości w miarę upływu czasu, gdyż zwykle wiele lat dzieli moment wykonania zdjęcia od jego cyfryzacji. Drugi problem wiąże się z niedoskonałościami skanera, takimi jak brak dokładnej kalibracji. Na koniec, nawet w domenie cyfrowej dane nie są w pełni bezpieczne, na przykład z powodu działania kompresji stratnej. W efekcie dwa zdjęcia dokładnie tego samego obiektu, wykonane w na tyle krótkim odstępie czasowym, że sam przedmiot nie uległ zmianie, mogą prezentować zupełnie inne kolory. W szerszym zakresie zagadnienia przetwarzania obrazów i ich cyfrowego kodowania zostały opisane m.in. w książce [7].

Ponadto nie wszystkie przekłamania powstają samorzutnie, niektóre są wprowadzane intencjonalnie. Dzieje się tak na przykład w przypadku skanowania fotografii monochromatycznych, które poddaje się koloryzacji. W taki sposób naniiesione dodatkowe informacje nie prezentują sobą

żadnych wartości dokumentacyjnych, a ze względu na liczne przekłamania w oryginalnych fotografiach kolorowych nie są łatwe do wykrycia, nawet przy posiadaniu kilku autentycznych kolorowych zdjęć referencyjnych. Bardzo często zdarza się jednak, że obiekt był przemalowywany, stąd inne ujęcia nie mogą stanowić punktu odniesienia.

Najpewniejszą metodą pozyskania dokładnych informacji o historycznej kolorystyce obiektu jest jego analiza, tak jak pokazano to w literaturze [8]. Niestety, jest to zwykle trudna i kosztowna metoda, ponadto niemożliwa do zastosowania, gdy sam przedmiot lub jego malowanie nie przetrwały. Z tych powodów fotografie są często jedynym punktem odniesienia przy odtwarzaniu oryginalnej kolorystyki historycznego obiektu.

W artykule zaproponowano metodę sprawdzania autentyczności kolorystyki zdjęcia historycznego, wykorzystującą uczenie maszynowe. W związku z tym opracowano zbiór przykładów zdjęć zarówno kolorowych, jak też koloryzowanych, które oryginalnie były wykonane w technice analogowej, a następnie uruchomiono na nim kilka znanych algorytmów w celu automatycznego wykrycia charakterystycznych cech odróżniających oba typy fotografii i dobrania odpowiednich parametrów dla klasyfikatora. Na końcu zweryfikowano trafność przygotowanej w taki sposób klasyfikacji dla kilku grup zdjęć.

2. ZESTAW DANYCH

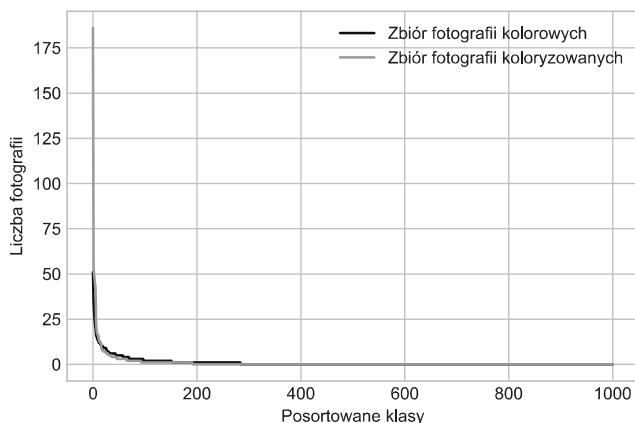
Wszystkie wykorzystane fotografie pobrano z Internetu, z serwisu Flickr. Zebrano łącznie 1 921 zdjęć, 962 kolorowe i 959 koloryzowane ręcznie przez artystów. W celu osiągnięcia generalizacji, a więc otrzymania prawidłowych rezultatów dla dowolnych fotografii analogowych, zdjęcia kolorowe były wykonane w dwóch, możliwie różniących się między sobą, technologiach Autochrome [9] oraz Kodachrome [10].

Wspomniane różnice technologiczne wynikają z kilku aspektów. Ze względu na zastosowane rozwiązania, analogowe technologie fotografii kolorowej można podzielić na te wykorzystujące addytywną oraz subtraktywną syntezę barw. Do pierwszej grupy należy praktycznie tylko Autochrome, natomiast w drugiej można wydzielić kolejne dwie podgrupy, rozróżnialne względem miejsca umieszczenia składowych barwnych. Kodachrome i Fujicolor zawierały je w wywoływaczu, natomiast w pozostałych rozwiązaniach umieszczono je w warstwie światłoczułej. Tą ostatnią grupę można ponownie rozdzielić

według sposobu, dzięki któremu udało się uzyskać dyfuzję cząsteczek odpowiadających poszczególnym składowym barwnym. W technologiach Agfacolor, Gevacolor, Ferraniacolor, Ektachrome, Ektacolor, Fotocwiel i Fotoncolor osiągnięto to poprzez rozbudowanie cząsteczki składowej, natomiast w technologii Kodacolor osłonięto składowe substancją nierozpuszczalną w wodzie.

Kolejnym problemem jest tematyka zdjęć. Sieci neuronowe stosowane do klasyfikacji w zawodach takich jak ImageNet [1] są uczone rozróżniania i przyporządkowywania obrazów ze względu na znajdujące się na nich obiekty. Na przykład na potrzeby zawodów ImageNet zebrano zestaw danych w którym znajduje się 14 197 122 zdjęć podzielonych na 21 841 kategorii tematycznych, przyporządkowanych według obiektów znajdujących się na fotografiach. Z całej tej liczby wydzielono 1 431 167 fotografii zorganizowanych w 1000 kategorii, które tworzą zbiór ILSVRC-2012-CLS. Zawody polegają na skonstruowaniu i wyuczeniu sieci na wspomnianym zbiorze w taki sposób, aby osiągnęła najlepszą możliwą celność klasyfikacji. Zdjęcia testowe, na podstawie których ocenia się sieci pochodzą spoza zbioru ImageNet.

W zadaniu sprawdzania autentyczności kolorów treść nie powinna mieć wpływu na rezultaty, toteż aby uniknąć ryzyka skojarzenia faktu koloryzacji z jakimś konkretnym obiektem przeanalizowano rozkład tematyczny fotografii. Na rysunku 1 pokazano wykres ilości zdjęć w poszczególnych kategoriach tematycznych, natomiast tablica 1 i 2 pokazują liczbę zdjęć w 10 najpopularniejszych kategoriach tematycznych w obu klasach. Do klasyfikacji wykorzystano sieć MobileNetV2 [11] wyuczoną na zbiorze ILSVRC-2012-CLS [1].



Rys. 1. Rozłożenie fotografii pomiędzy kategorie tematyczne

Tablica 1. Dziesięć najpopularniejszych klas tematycznych w zbiorze zdjęć kolorowych

Nazwa klasy	Liczba zdjęć
mundur wojskowy	51
zamek	34
klasztór	29
brzeg jeziora	25
więzienie	21
kościół	20
stodoła	16
szopa	16
pałac	14
ławka parkowa	14

Najbardziej popularną klasą w obu zbiorach jest mundur wojskowy, niemniej na pozostałych, niższych miejscach wszystkie tematy różnią się. W zbiorze zdjęć kolorowych dominuje architektura, natomiast w zbiorze fotografii koloryzowanych przeważają elementy garderoby. Wynika to z faktu obecności dużej liczby portretów.

Tablica 2. Dziesięć najpopularniejszych klas tematycznych w zbiorze zdjęć koloryzowanych

Nazwa klasy	Liczba zdjęć
mundur wojskowy	186
muszka	54
garnitur	49
futro	48
suknia	43
peruka	43
pan młody	22
bluza	18
ornat	17
stanik	16

Do większości z rozróżnianych 1001 klas tematycznych w obu zbiorach nie została przyporządkowana żadna fotografia. Wśród kolorowych zdjęć jest 717 pustych klas, natomiast wśród koloryzowanych jest ich 809, stąd większą różnorodność tematyczną prezentują zdjęcia kolorowe. Do dziesięciu najpopularniejszych klas przynależą 240 zdjęć kolorowych i 496 zdjęć koloryzowanych, co stanowi odpowiednio 25% i 51,9% wszystkich fotografii w danej grupie. 103 klasy tematyczne wystąpiły zarówno wśród zdjęć kolorowych jak też koloryzowanych. Należy do nich łącznie 1260 obrazów. Wynika z tego, że na podstawie tematyki jednoznacznie można przyporządkować 661 zdjęć stanowiących jedynie 34,4% ogółu. W takiej sytuacji można postawić tezę, że jeśli trafność wyuczonej sieci będzie wynosiła ponad 34,4%, to faktycznie sieć nauczyła się odróżniać zdjęcia kolorowe od koloryzowanych. W przeciwnym wypadku, tj. jeśli trafność będzie na poziomie wspomnianej wartości, sieć w rzeczywistości będzie rozróżniać tylko tematykę.



Rys. 2. Przykładowy efekt losowania 12 zdjęć z przypisanymi klasami tematycznymi przez sieć MobileNetV2

3. UCZENIE MASZYNOWE

Na przygotowanym zbiorze danych wyuczono kilka klasyfikatorów różnych sieci neuronowych odróżniania fotografii pomiędzy kolorowe a koloryzowane. Do tego celu

wykorzystano już wyćwiczone sieci na zbiorze ILSVRC-2012-CLS, o nazwach ResNetV2 [12], MobileNetV2 [11], InceptionV3 [13]. Jedną sieć tj. InceptionV3 była dodatkowo przetestowana w wariancie wyuczonym na zbiorze iNaturalist [2]. Z każdej z tych sieci podmieniono ostatnią warstwę odpowiadającą za przypisanie do klasy, natomiast zastawiono w całości część wyliczającą wektor cech opisujący zdjęcie. Zastosowana metoda uczenia została szeroko opisana w literaturze, m. in. w poz. [14]. Trening zawsze trwał 4000 epok. Po każdej epoce, odbywało się sprawdzenie trafności klasyfikacji dokonywanych przez sieć. W tabelicy 3 podano liczbę cech, na podstawie której każda z sieci dokonywała klasyfikacji.

Tabela 3. Długości wektorów cech w poszczególnych sieciach

Nazwa sieci	Liczba cech
ResNetV2	2048
MobileNetV2	1280
InceptionV3	2048

Wszystkie wykorzystane sieci są szeroko dostępne i można je uruchamiać na różnych rozwiązaniach sprzętowych a także programowych, gdyż istnieją ich realizacje wykonane w różnych technologiach. Przykładowo są dostępne ich wersje działające zarówno na komputerach osobistych, telefonach jak też i serwerach. Umożliwiają one także regulację głębokości i rozmiaru wejścia w celu dostosowania do różnych zadań i maszyn obliczeniowych różnej mocy. Stąd, zależnie od komplikacji zadania i potrzeb, wykorzystując każdą z wspomnianych sieci można zarówno skonstruować usługę działającą zdalnie na serwerze lub lokalnie na terminalu użytkownika.

Do celów uczenia wszystkie zgromadzone zdjęcia zostały rozdzielone na trzy podzbiory. Zbiór 1709 w pełni automatycznie wylosowanych zdjęć został użyty do treningu. Kolejnych 100 wydzielono do testu który odbywał się na końcu każdej epoki. Ostatnie 112 wykorzystano tylko raz, na zakończenie całego procesu uczenia, do sprawdzenia sieci po przejściu wszystkich epok. Na rysunku 2 przedstawiono przykładowy rezultat losowania 30 zdjęć.

4. WYNIKI

W trakcie eksperymentu sprawdzono wpływ rozmiaru obrazka na wejściu sieci, wpływ zwiększenia głębokości sieci przy pozostawieniu stałymi pozostałych parametrów oraz wpływ zbioru danych, który został użyty do wyuczenia części wyodrębniającej wektor cech ze zdjęcia na trafność klasyfikacji w omawianym zadaniu rozdzielania historycznych fotografii na kolorowe i koloryzowane.

4.1. Wpływ rozmiaru obrazka

Tabela 4. Trafność klasyfikacji w zależności od wielkości zdjęcia na wejściu

Rozmiar obrazka (w pikselach)	Trafność (%)
244x244	92.0%
160x160	91.5%
128x128	92.6%
96x96	90.5%

Wpływ zmiany wielkości obrazka na wejściu sieci na trafność sprawdzono na sieci MobileNetV2. Sieć wyodrębnia 1280 cech. Tabela 4 prezentuje uzyskane

wyniki dla poszczególnych rozmiarów. W zbadanym zakresie nie zaobserwowano żadnych znaczących różnic.

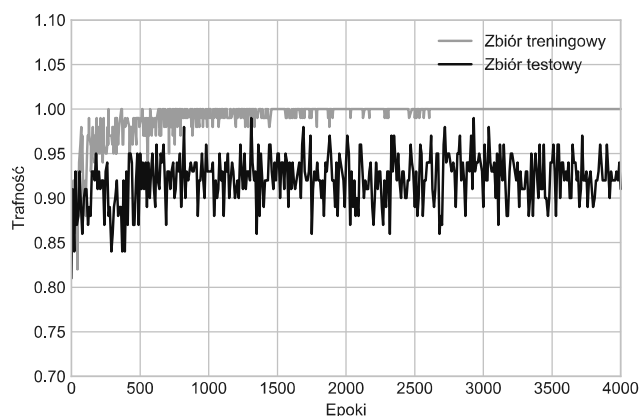
4.2. Wpływ głębokości sieci

Wpływ zmiany głębokości sieci na trafność zbadano korzystając z sieci ResNetV2. Sieć przyjmuje na wejściu zdjęcia o wymiarach 244 na 244 piksele i wyodrębnia z nich 2018 cech. Tabela 5 prezentuje uzyskane wyniki dla poszczególnych głębokości. Podobnie jak w poprzednim przypadku nie zaobserwowano żadnych znaczących różnic.

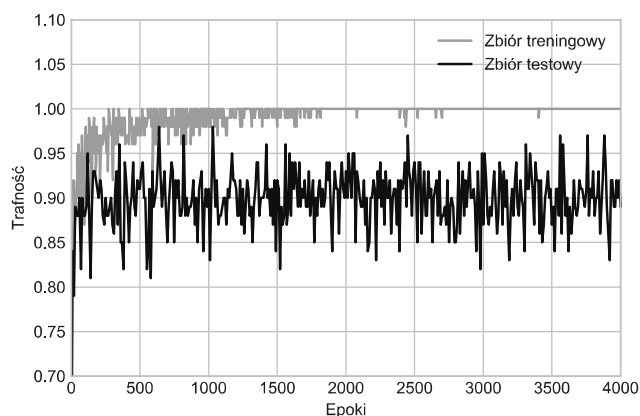
Tabela 5. Trafność klasyfikacji w zależności od głębokości sieci

Głębokość sieci (w warstwach)	Trafność (%)
50	90,5%
152	92,1%

Rysunki 3 i 4 zawierają wykresy krzywych uczenia się obu sieci neuronowych. Widać na nich, jaką trafność klasyfikacji uzyskała sieć po każdej z epok, na obu zbiorach, treningowym i testowym. Obie sieci osiągnęły nasycenie, tj. po epoce nr 1000 trafność praktycznie pozostaje na tym samym poziomie, tak więc dalsze uczenie nie jest konieczne. Z tego powodu możliwe jest także zmniejszenie liczby epok do okolic wspomnianej wartości.



Rys. 3. Krzywe uczenia się sieci ResNetV2 złożonej z 50 warstw



Rys. 4. Krzywe uczenia się sieci ResNetV2 złożonej z 152 warstw

4.3. Wpływ zestawu danych użytego do uczenia wyodrębniania cech

Do tego eksperymentu użyto sieci InceptionV3 która pracuje na obrazkach o rozmiarze 299 na 299 pikseli i wyróżnia 2048 cech, wyuczonej na dwóch zbiorach ILSVRC-2012-CLS i iNaturalist. Pierwszy jest zbiorem ogólnego przeznaczenia, a drugi z nich jest zbiorem

drobnoziarnistym, złożonym ze zdjęć gatunków roślin i zwierząt. Tablica 6 prezentuje uzyskane wyniki. Tak samo jak w poprzednich przypadkach nie zaobserwowano znaczących różnic.

Tablica 6. Trafność klasyfikacji w zależności od zestawu danych użytego do uczenia wyodrębniania cech

Zbiór uczący	Trafność (%)
ILSVRC-2012-CLS	91,0%
iNaturalist	90,1%

5. WNIOSKI KOŃCOWE

Wszystkie sieci zostały poprawnie nauczone rozróżniania zdjęć kolorowych od koloryzowanych. Osiągnięta trafność na poziomie co najmniej 90% jest dalece wyższa niż ta, której można było oczekiwać na podstawie klasyfikacji tematycznej zdjęć. Potwierdza to też przydatność skonstruowanego zestawu zdjęć do treningu sieci. Ponadto zaobserwowano jedynie marginalne różnice w trafności, w zależności od zastosowanej sieci. Najprawdopodobniej wynikają one z faktu, że zbiór użyty do weryfikacji był każdorazowo losowany od nowa, stąd czasem trafiały się przypadki nieco trudniejsze, a niekiedy łatwiejsze. Wynika z tego, że do rozwiązania opisanego zadania wystarczają z powodzeniem najprostsze i najmniejsze sieci, które da się stosować na urządzeniach przenośnych, dlatego jest możliwe stworzenie stosownej aplikacji na urządzenia przenośne, która ułatwiłaby pracę modelarzom i rekonstruktorom poszukującym autentycznych fotografii.

W zbadanym zakresie rozmiarów obrazów wprawdzie nie zaobserwowano żadnych różnic w trafności, niemniej jednak w obszarze dalszych badań leży problem do jakiego stopnia można zmniejszać fotografie na wejściu, żeby utrzymać podobny poziom skuteczności. Możliwe, że dalsze zwiększanie rozmiaru, a więc także dostarczanie kolejnych informacji o zdjęciu spowodowałyby w pewnym momencie znaczącą poprawę wyników. Być może także lekka zmiana parametrów w części już wyuczonej spowodowałyby poprawę skuteczności klasyfikacji.

Do zbadania także pozostaje problem ustalenia na podstawie jakich właściwości sieć nauczyła się rozróżniać fotografie. W tym celu potrzebna byłaby pogłębiona analiza kolorymetryczna, która pozwoliłaby użyć jeszcze mniejszych sieci lub nawet zrezygnować z uczenia maszynowego na rzecz prostszych rozwiązań.

6. BIBLIOGRAFIA

- O. Russakovsky i in., „ImageNet Large Scale Visual Recognition Challenge”, *Int. J. Comput. Vis.*, t. 115, nr 3, ss. 211–252, 2015.
- G. Van Horn i in., „The iNaturalist Species Classification and Detection Dataset - Supplementary Material”, *Comput. Vis. Found.*, nr July, ss. 4–6, 2017.
- S. Maji, E. Rahtu, J. Kannala, M. Blaschko, i A. Vedaldi, „Fine-Grained Visual Classification of Aircraft”, 2013.
- T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, i F. F. Li, „Fine-grained car detection for visual census estimation”, 31st AAAI Conf. Artif. Intell. AAAI 2017, ss. 4502–4508, 2017.
- J. Krause, M. Stark, J. Deng, i L. Fei-Fei, „3D object representations for fine-grained categorization”, w *Proceedings of the IEEE International Conference on Computer Vision*, 2013, ss. 554–561.
- L. Yang, P. Luo, C. C. Loy, i X. Tang, „A large-scale car dataset for fine-grained categorization and verification”, w *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, t. 07-12-June, ss. 3973–3981.
- M. Domański, *Obraz cyfrowy: reprezentacja, kompresja, podstawy przetwarzania: standardy JPEG i MPEG*. Wydawnictwa Komunikacji i Łączności, 2010.
- P. Tångeberg, „The use of colours on the seventeenth century royal warship Vasa”, *Endeavour*, t. 24, nr 4. ss. 147–151, 2000.
- C.-A. Francois-Franck, „Colour microphotography with autochrome plates by MM. A. et L. Lumiere.”, *COMPTES RENDUS Hebd. DES SEANCES L Acad. DES Sci.*, t. 144, ss. 1340–1342, 1907.
- L. D. Mannes i L. Godowsky, „The Kodachrome Process for Amateur Cinematography in Natural Colors”, *J. Soc. Motion Pict. Eng.*, 1935.
- M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, i L. C. Chen, „MobileNetV2: Inverted Residuals and Linear Bottlenecks”, w *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, ss. 4510–4520.
- K. He, X. Zhang, S. Ren, i J. Sun, „Identity mappings in deep residual networks”, w *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, t. 9908 LNCS, ss. 630–645.
- C. Szegedy, V. Vanhoucke, S. Ioffe, i J. Shlens, „Rethinking the Inception Architecture for Computer Vision”.
- J. Donahue i in., „DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition”.

CHECKING COLOR AUTHENTICITY IN ANALOG PHOTOS

The article describes a dataset designed to train neural networks distinguishing historical photographs between the ones that have original historic color and those which were contemporary colorized. The problem of choosing photos in terms of technology and content was considered. Using some of the pre-trained neural networks on other collections, their effectiveness in solving the studied issue was checked. The influence of the input image size, the depth of the neural network used as well as the data set used to train the network to extract features was investigated. As a result, the usefulness of the developed set for network training was confirmed, and it was observed that increasing the network did not bring any additional benefits. The reached accuracy is up to 92.6%.

Keywords: dataset, color historic photography, machine learning, authenticity.