



Prof. dr hab. Ewa Grabska
Zakład Projektowania i Grafiki Komputerowej
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytet Jagielloński

Kraków, 29. 11. 2022

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Detekcja obiektów w obrazach cyfrowych z użyciem uczenia głębokiego w warunkach stratnej kompresji obrazu

Autor rozprawy: mgr inż. Tomasz Gandor

Promotor rozprawy: dr hab inż. Henryk Josiński

Promotor pomocniczy: dr hab inż. Jakub Nalepa

Informacje ogólne

Oceniana rozprawa doktorska mgr inż. Tomasza Gandora zatytułowana „Detekcja obiektów w obrazach cyfrowych z użyciem uczenia głębokiego w warunkach stratnej kompresji obrazu” stanowi opracowanie o objętości 155 stron, zawierające 32 rysunki, 16 tabel, spis literatury obejmujący 111 pozycji, cztery załączniki oraz elektroniczną wersję rozprawy w formacie PDF na płycie CD.

I. O przedmiocie rozprawy

Rozprawa doktorska mgr inż. Tomasza Gandora dotyczy problemów trudnych do formalnego opisanego z dziedziny sztucznej inteligencji, w rozwiązaniu których ludzie najczęściej kierują się inteligencją i/lub intuicją. Do tego typu problemów należą przykładowo rozpoznawanie wypowiedzianych słów czy wykry-

wanie twarzy na obrazach. Umożliwienie automatycznego rozwiązywanie problemów, które wydają się subiektywnym podejmowaniem decyzji związanym ze znajomością świata rzeczywistego, wymaga wyposażenia komputerowych systemów sztucznej inteligencji w zdolność zdobywania własnej wiedzy poprzez wydobywanie wzorców z surowych danych, Taka zdolność zwana jest uczeniem maszynowym. Rozwinięciem uczenia maszynowego jest forma głębokiego uczenia stwarzająca warunki systemom sztucznej inteligencji zarówno rozumienia świata w kategoriach hierarchii pojęć, jak również uczenia się złożonych koncepcji budując je z prostszych. Graficzna reprezentacja tego typu hierarchii charakteryzuje się wielowarstwową głębią.

Tematem niniejszej pracy doktorskiej jest wykrywanie obiektów w obrazach cyfrowych z użyciem głębokiego uczenia maszynowego w warunkach stratnej kompresji obrazów. *Wykrywanie obiektów* jest to metoda związana z komputerową wizją i przetwarzaniem obrazów, która zajmuje się ujawnianiem obiektów należących do określonej semantycznie klasy obiektów, przykładowo: ludzi, budowli architektonicznych, czy też pojazdów. Dobrze zbadane dziedziny wykrywania obiektów obejmują zarówno wykrywanie twarzy, jak i pieszych. *Kompresja obrazów* jest rodzajem kompresji danych stosowanych do obrazów cyfrowych w celu zmniejszenia ich kosztów przechowywania lub przesyłania. W porównaniu z ogólnymi metodami kompresji danych, używane algorytmy w kompresji obrazów mogą wykorzystywać percepcję wizualną oraz statystyczne własności danych obrazów. *Stratna kompresja* obrazów obejmuje klasę metod, która wykorzystuje niedokładne przybliżenia i częściowe odrzucanie danych do reprezentowania obrazów. Możliwość redukcji danych przy użyciu kompresji stratnej jest znacznie wyższa niż przy użyciu technik bezstratnych. Kompresja stratna jest najczęściej stosowana do kompresji danych multimedialnych, takich jak audio, wideo oraz obrazy.

Należy zwrócić uwagę, że doktorant proponując tematykę przedstawioną w rozprawie podejmuje się dokonania analizy wybranych zagadnień percepcji wizualnej obrazów z punktu widzenia człowieka oraz modeli uczenia maszynowego. Człowiek postrzegając obraz korzysta z doświadczeń, nastawienia do otaczającego świata oraz własnej jego interpretacji. Okazuje się, że sama ludzka percepcja jest poszukiwaniem znaczenia, w którym często przekracza się zamierzenia stymulując twórcze rozwiązania. Widzenie wizualne nie jest percepcją bierną. Jest procesem złożonym z akcji tworzących konstrukcję postrzeganego obrazu. Takie rozumowanie prowadzi do koncepcji dynamicznej percepcji, która opiera się zarówno na zewnętrznej informacji graficznej, jak i na konstruktywnym procesie mentalnym. Doktorant podejmuje próbę znalezienia analogii pomiędzy percepcją ludzką a percepcją inteligencji obliczeniowej w postaci modeli głębokiego uczenia. Pomimo, że jego badania ograniczają się do wybranych aspektów, które wyznacza tematyka rozprawy to jednak dotyczą najistotniejszych problemów współczesnych, biorąc pod uwagę, że większość z nas pozostaje pod wpływem różnego rodzaju obrazów przekazywanych za pomocą telewizji, komputerów, czy smartfonów.

II. O postawionym w rozprawie celach

Cel jaki rozważa w rozprawie doktorant to analiza z użyciem miary skuteczności oddziaływania stratnej kompresji na wykrywanie obiektów w obrazach za pomocą maszynowego głębokiego uczenia.

Główne hipotezy Autora odnoszą się do:

1. wskazań dotyczących doboru parametrów kompresji stratnej obrazów w sposób optymalny, z punktu widzenia skuteczności działania systemów komputerowych, w których stan liczebny przetwarzanych obrazów jest duży,
oraz
2. przeciwdziałania pogorszeniu skuteczności detekcji obiektów w obrazach przez modyfikację zbioru obrazów użytego do treningu modelu.

III. Przegląd zawartości rozprawy

W **Rozdziale 1** Autor wyjaśnia na czym polega widzenie maszynowe, krótko charakteryzuje wykrywanie obiektów w obrazach oraz przedstawia liczne dziedziny zastosowań tego typu detekcji. Kluczowe znaczenie w detekcji obrazów poddanych odpowiedniej kompresji odgrywa maszynowe uczenie głębokie rozważane w niniejszej rozprawie, której cel i zakres badań opisany jest w dalszej części rozdziału.

Rozdział 2 ma charakter przeglądowy. Dotyczy jakości obrazów oraz metod ich klasyfikacji. Autor prezentuje wybrane metody kompresji obrazów dla których dokonywana jest detekcja zawartych w nich obiektów za pomocą głębokiego uczenia. Analiza materiałów przedstawianego przeglądu istniejących metod zainspirowała doktoranta do przeprowadzenia dodatkowych badań odnoszących się do doboru parametrów w algorytmie Joint Photographic Experts Group (JPEG) stratnej kompresji oraz do wzbogacenia zbioru uczącego.

W **Rozdziale 3** scharakteryzowane zostało uczenie maszynowe, które można potraktować jako jedno z następstw podejmowania prób zastosowania idei sztucznej inteligencji w praktyce. W sposób interesujący Autor zaprezentował taksonomię zadań uczenia maszynowego, jak również modele uczenia głębokiego wykorzystujące sztuczne sieci neuronowe z nadzorowaniem, które stanowią podwaliny metod zastosowanych w rozprawie.

W **Rozdziale 4** Autor przedstawił etapy procesu detekcji obiektów w kontekście widzenia maszynowego i podjął próbę sformalizowania danych wejściowych i wyjściowych dla systemu komputerowego dokonującego detekcji. Dane wyjściowe mogą różnić się reprezentacją prostokąta w którym znajduje się

wykrywany obiekt. Różne reprezentacje prostokąta zależą od rodzaju funkcji atrybutowania zdefiniowanej dla prostokąta. W ramach uczenia maszynowego etapy procesu detekcji są traktowane jako zadania i w tym wypadku mamy do czynienia z modelami wielozadaniowymi z kategorii klasyfikacji i regresji realizowanych w kontekście uczenia nadzorowanego. W dalszej części rozdziału scharakteryzowano dwa typy modeli detekcji: jedno- i dwuetapowe. Okazuje się, że detekcja obiektów realizowana za pomocą głębokiego uczenia maszynowego jest aktualnie najbardziej skuteczną metodą.

W **Rozdziale 5** zaprezentowano techniki kompresji obrazów ze szczególnym uwzględnieniem metody JPEG, którą Autor wykorzystał w eksperymentach obliczeniowych opisanych w kolejnych rozdziałach. Należy podkreślić, że problemy sterowania jakością w algorytmie JPEG zostały przedstawione z wielką starannością i widoczną znajomością tematu.

W **Rozdziale 6** Autor opisał w jaki sposób cele rozprawy doktorskiej zostały zweryfikowane praktycznie przez eksperymenty dotyczące wpływu kompresji stratnej na wywoływanie oczekiwanego skutku w detekcji obiektów z zastosowaniem uczenia głębokiego. W dalszym ciągu zbadał ewentualności poprawy skuteczności detekcji odnoszącej się do obrazów z gorszą jakością na skutek kompresji.

Rozdział 7 zawiera analizę wyników eksperymentów przedstawionych przez Autora w **Rozdziale 6**. Wynikiem analizy są zalecenia, które mogą okazać się przydatne w systemach monitoringu wideo wykorzystujących modele detekcji obiektów należących do obrazów poddawanych stratnej kompresji.

Rozdział 8 zawiera podsumowanie oraz pokazuje wybrane kierunki kontynuacji badań.

V. Ocena wyników rozprawy

Trzeba stwierdzić, że cel rozprawy został osiągnięty. Przedstawione rozwiązanie problemu detekcji obiektów w obrazach cyfrowych z użyciem uczenia głębokiego w warunkach stratnej kompresji obrazu ma, moim zdaniem, zarówno wartość praktyczną, jak i wartość poznawczą. Na uwagę zasługuje fakt, że na etapie analizy uzyskanych wyników Autor dokonał przeglądu pozycji literatury, które pozwolą poszerzyć zakres tematyczny rozprawy. Doktorant najwyraźniej zna poruszany temat i jego literaturę, oraz ma własne pomysły, które umie zrealizować. Rozprawa jest spójna logicznie i jej układ jest bez zarzutu.

Mam nadzieję, że szerszy zakres tematyczny prowadzonych badań pozwoli Autorowi scharakteryzować relacje pomiędzy percepcją ludzką a percepcją inteligencji obliczeniowej w postaci modeli głębokiego uczenia. Z dotychczasowych wyników zaprezentowanych w rozprawie wynika, że w obu przypadkach percepcja ma charakter dynamiczny.

Rozprawa ma również i słabsze punkty.

Strona 30:

1. Autor podejmuje się formalizacji danych wejściowych oraz wyjściowych w detekcji obrazu. Opis formalny z definicji wymaga większej precyzji.

Uwaga ogólna: typy liczb, które są wprowadzane należy od razu zdefiniować.

Jest:

Dany jest zbiór N kategorii (klas obiektów), które należy wykrywać.

Powinno przykładowo być:

Niech $n \in \mathbb{N}$ będzie liczbą elementów zbioru kategorii obiektów, gdzie $n \geq 1$.

Elementy zbioru liczb naturalnych powinny być oznaczane małymi literami. Zastosowane oznaczenie N jest jedną z używanych notacji zbioru liczb naturalnych.

2. Powinien być podany typ liczb r_j , $r_j \in \mathbb{R}$ dla $j \in \{1, \dots, 4\}$
Czy lepszym rozwiązaniem nie byłby, np. element $v \in \mathbb{R}^4$?
3. Zamiast karkołomnej definicji zbioru indeksów $Z \cap [1, N]$, wystarczy użyć zbioru $\{1, \dots, n\}$. Czy indeksem zamiast c_i nie powinno być i ? Użycie c_i prowadzi do kolizji oznaczeń. Taka notacja związana jest z ufnością modelu.

Strona 46

Równości $P(X) = \{p(x_i)\}$ oraz $X = \{x_i\}$ wymagają wyjaśnienia ze względu na brak zdefiniowania zakresu indeksu i . Z przedstawionych zapisów nie jest wiadomo, czy rozważania dotyczą zmiennej losowej, która jest dyskretna o skończonej liczbie stanów. W jaki sposób należy interpretować sumowanie po elementach $x_i \in X$ jeżeli nie mamy żadnej informacji o zakresie indeksu i ?
W konsekwencji, wzory 4.5, 4,6 oraz 4.7 są niepoprawne.

Na poprawny zapis matematyczny zarówno notacji, jak i wzorów należałoby szczególnie zwrócić uwagę przy publikacji uzyskanych wyników.

VI. Konkluzja

W sumie nie mam jednak poważniejszych zastrzeżeń do rozprawy i oceniam ją zdecydowanie pozytywnie. Oceniana rozprawa doktorska spełnia wymagania, jakie ustawa o stopniach i o tytule naukowym przewiduje dla rozpraw doktorskich w dziedzinie nauk inżyniersko-technicznych w dyscyplinie: informatyka techniczna i telekomunikacja i na tej podstawie wnoszę o dopuszczenia jej Autora – mgr inż. Tomasz Gandora – do publicznej obrony.

