

RECENZJA

rozprawy doktorskiej mgr inż. Tomasza Gandora
pt. Detekcja obiektów w obrazach cyfrowych z użyciem uczenia głębokiego w
warunkach stratnej kompresji obrazu
wykonana na zlecenie Polsko-Japońskiej Akademii Technik Komputerowych

1. Zakres, cel i tezy Rozprawy

Rozprawa dotyczy wpływu stratnej kompresji na skuteczność detekcji obiektów na obrazach, realizowanej za pomocą modeli opartych o uczenie maszynowe głębokie w aspektach jakościowym oraz ilościowym. Potrzeba badań w zakresie przedstawionym w rozprawie wynikała z realnego problemu zidentyfikowanego we wdrażanym systemie CityEye, służącym do inteligentnej analizy sekwencji wideo pochodzących z kamer monitoringu miejskiego. Uzyskana w ten sposób wiedza znajdzie zastosowanie w tym systemie.

Cel podjęcia tematu ma aspekt praktyczny, jak i teoretyczny. **Wiedza praktyczna** uzyskana w badaniach umożliwi specjalistom od systemów detekcji obiektów pracującym z dużą liczbą dużych obrazów, kiedy jest stosowana ich stratna kompresja, otrzymanie wskazówek, jak dobierać parametry kompresji w sposób optymalny z punktu widzenia skuteczności działania systemu, oszczędności przestrzeni dyskowej i efektywnego wykorzystania energii. **W zakresie teoretycznym** Doktorant oczekuje poszerzenie wiedzy o głębokich modelach detekcji obiektów w warunkach stratnej kompresji obrazu, a także o możliwości przeciwdziałania pogorszeniu skuteczności detekcji w wielu różnych zastosowaniach tych systemów.

Doktorant stawia następujące trzy tezy:

Teza 1. Kompresja JPEG nie powoduje spadku precyzji detekcji dla parametru Q z przedziału [30, 100]. Przyczyną spadku skuteczności detekcji jest w tym przypadku pogorszenie czułości detekcji.

Teza 2. Spadek skuteczności detekcji obiektów dla głębokich modeli, mierzony za pomocą miar AP, jest podobny zarówno w modelach jednoetapowych, jak i dwuetapowych dla parametru Q z przedziału [30, 100].

Teza 3. Trening detektorów obiektów z użyciem obrazów silnie ($Q=20$) lub średnio ($Q=40$) skompresowanych poprawia skuteczność detekcji na obrazach poddanych silnej kompresji, dla parametru Q z przedziału [5, 30). Źródłem poprawy jest zwiększenie czułości detekcji dla tych obrazów.

2. Struktura Rozprawy

Rozprawa została zredagowana na 168 stronicach w 8-miu rozdziałach, z Bibliografią liczącą 111 pozycji w zdecydowanej większości (106 pozycji) wywołanych w treści Rozprawy. Jest reprezentatywna dla zagadnień rozwiązywanych w Rozprawie, choć nie obejmuje zagadnień ogólnego rozpoznawania i rozumienia otoczenia. W przeglądzie literatury nie ma wzmianki o ośrodkach i badaczach krajowych z osiągnięciami w przedmiocie Rozprawy. Doktorant jest współautorem 2 publikacji (anglojęzycznych) i nie ma publikacji samodzielnej. Całość jest poprzedzona Streszczeniem (po angielsku), Spisem treści,

Spisem symboli, Spisem tabel i zakończona Bibliografią oraz 4-ma dodatkami A, B C i D, które zawierają (w kolejności): Wykaz skrótowców i symboli, Miary skuteczności modeli pre-treningowych, Miary skuteczności modeli wytrenowanych oraz Kody źródłowe programów.

Bibliografia nie została uporządkowana ani chronologicznie, ani tematycznie, ani alfabetycznie, co utrudnia jej śledzenie i ocenę jej potencjalnej ważności dla zawartości Rozprawy.

2.1. Zawartość Rozprawy

Zawartość kolejnych rozdziałów Rozprawy jest następująca:

1. Rozdział 1 pt. **Wprowadzenie** omawia zagadnienia tematyczne Rozprawy, w szczególności:

- detekcję obiektów na obrazach,
- uczenie głębokie,
- miary skuteczności detekcji obiektów,
- kompresję obrazu.

a także zawiera: cel, zakres i tezy Rozprawy.

2. Rozdział 2 pt. **Przegląd literatury** jest omówieniem literatury na którą Doktorant powołuje się analizując zależności między jakością obrazu a klasyfikacją oraz niezawodnością sieci konwolucyjnych. Odwołania do literatury są też w kolejnych rozdziałach.

3. Rozdział 3 pt. **Uczenie maszynowe głębokie** zawiera omówienie metod uczenia głębokiego w zakresie niezbędnym do dalszych rozważań na temat detekcji obiektów. W szczególności, Rozdział zawiera:

- relacje uczenia maszynowego i AI,
- uczenie nadzorowane i nienadzorowane,
- transfer uczenia,
- uczenie ze wzmocnieniem.

4. Rozdział 4 pt. **Detakcja obiektów** zawiera opis zadania detekcji obiektów. Na początku charakterystykę metod starszych, opartych o widzenie maszynowe, po czym metod aktualnych, bazujących na uczeniu głębokim. Detekcję obiektów opisano określając dane wejściowe i postacie spodziewanych wyników, skrótowo prezentując wcześniejsze podejścia, nie obejmujące głębokiego uczenia maszynowego. Szczegółowo przedstawiono metody detekcji obiektów zawierające uczenie głębokie z budową sieci neuronowych i podział na elementy funkcjonalne. Następnie dokonano przeglądu literaturowych modeli detekcji obiektów. W szczególności poruszano następujące zagadnienia:

- zadanie detekcji obiektów,
- widzenie maszynowe;
- uczenie głębokie w detekcji obiektów, a w nim:
 - przepływ danych w detektorze obiektów,
 - kręgosłup detektora obiektów,
 - detektor jednoetapowy,
 - detektor dwuetapowy,
 - trening detektorów obiektów;
- ocenę skuteczności detekcji obiektów, a w niej:
 - ocenę pojedynczego wyniku,
 - miary zależne od progu ufności,
 - miary średniej precyzji (AP).

5. Rozdział 5 pt. **Kompresja obrazów** zawiera omówienie następujących zagadnień:

- rodzaje kompresji obrazów (bezstratne i stratne - całkowicie bezbłędne w odwzorowaniu danych obrazu po dekompresji oraz odwzorowanie ze stratami informacyjnymi),
- kompresję JPEG (algorytm JPEG jako powszechnie wykorzystywana metoda stratnej kompresji obrazów kolorowych),
- sterowanie jakością w algorytmie JPEG (parametr jakości Q),
- miary jakości obrazu (referencyjne i bezreferencyjne).

6. Rozdział 6 pt. **Badania empiryczne** stanowi główna część Rozprawy i zawiera opis przeprowadzonych przez Doktoranta badań. Dotyczą one w szczególności:

- odtwarzalności badań,
- zestawu eksperymentalnego w etapie I,
- wyników bazowych modeli pre-trenowanych,
- skuteczności detekcji w funkcji parametru Q,
- zestawu eksperymentalnego w etapie II,
- wyników dla modeli wytrenowanych.

7. Rozdział 7 pt. **Omówienie rezultatów i wnioski** zawiera omówienie wyników eksperymentów i wnioski sformułowane na podstawie treści w Rozdziale 6. w szczególności:

- wnioski i wskazania praktyczne,
- ulepszenia treningu detektorów,
- ocenę detekcji obiektów.

Część z wniosków ma charakter praktycznych zaleceń, mogących być pomocnymi w inżynierii systemów monitoringu wideo wykorzystujących stratną kompresję. Znajdują się tu też odniesienia do publikacji, które umożliwiają poszerzenie zakresu badań opisywanych w Rozprawie. Niektóre z przytoczonych prac zawierają odkrycia i obserwacje, które potwierdzają lub są zbieżne z tymi, które sformułowano na podstawie badań wykonanych w Rozprawie.

Pierwsza obserwacja dotyczy działania kompresji JPEG i wynikowej jakości obrazu mierzonej za pomocą podobieństwa strukturalnego (SSIM). Wniosek praktyczny jest taki, że stosowanie skrajnych wartości parametru Q nie ma sensu – za wysokie wartości to duży rozmiar plików wyjściowych, a za niskie to gwałtowny spadek jakości, w obu przypadkach z nikłymi korzyściami. Wybór wartości w środkowym, szerokim przedziale (ok. 20–90), może być podyktowany wymaganiami konkretnego zastosowania. Na przykładzie zbioru COCO, który używa głównie wartości $Q = 96$, widać, że praktyka w budowaniu zbiorów OiM do detekcji obiektów była zgodna z tą obserwacją, a uzyskane tu wyniki dodatkowo potwierdzają jej trafność.

2.2. Zakresy badań empirycznych

Badania empiryczne składały się z dwóch etapów realizacji celów Rozprawy:

1. **Pierwszy etap** dotyczył badania wpływu kompresji stratnej na skuteczność detekcji obiektów z użyciem uczenia głębokiego. Przedstawiono w nim przyjęte w Rozprawie praktyki, które zapewniają odtwarzalność przeprowadzonych badań: użycie ogólnodostępnych zbiorów danych, publiczne udostępnienie danych wynikowych oraz oprogramowania użytego w realizacji badań - zarówno autorskiego, napisanego specjalnie w tym celu, jak i udostępnionego przez innych autorów.

W dalszej części etapu I opisano konfigurację zestawu eksperymentalnego dla etapu I eksperymentów - testowy zbiór OiM, sposób i rezultaty degradacji jakości poprzez kompresję obrazu, wykorzystane miary skuteczności detekcji obiektów oraz zestawienie testowanych modeli. Następnie

przedstawiono wyniki pierwszego etapu eksperymentów, który dotyczył badania dziewięciu pre-trenowanych modeli detekcji obiektów na reprezentatywnym zbiorze testowym w całym zakresie wartości parametru kompresji Q .

2. **Drugi etap** dotyczył badań możliwości poprawy skuteczności detekcji na obrazach z pogorszoną w skutek kompresji jakością. Przedstawiono w nim prowadzony trening modeli detekcji obiektów, poprzedzony opisem zestawu eksperymentalnego użytego w tym etapie, ze wskazaniem elementów wspólnych z etapem I. Opisano różnice dotyczące próbkowania wartości Q , architektur modeli i procedury ich treningu, a także konfiguracji sprzętowej stanowisk, na których był on przeprowadzony.

3. Uzyskane wyniki

Przedstawiona Rozprawa dotyczy problemów wpływu kompresji stratnej na detekcję obiektów wykorzystującą uczenie głębokie. W celu szczegółowego przebadania tego wpływu zebrano niezbędną wiedzę dostępną w literaturze, która dotyczy działania detektorów obiektów opartych o głębokie konwolucyjne sieci neuronowe oraz wybrano algorytm kompresji JPEG jako reprezentatywny przykład kompresji stratnej. Zgromadzone informacje zostały włączone do teoretycznej części Rozprawy. Na ich podstawie zostały zaprojektowane dwa etapy eksperymentów:

1. w etapie I analizowano miary skuteczności detekcji obiektów reprezentatywnej grupy dziewięciu pre-trenowanych modeli detekcji na obszernym zbiorze zróżnicowanych obrazów COCO val2017 (5 tys. obrazów),

Wyniki uzyskane w etapie I eksperymentów umożliwiają sformułowanie następujących wskazań dla projektantów i operatorów systemów monitoringu stosujących detekcję obiektów z użyciem sieci głębokich:

- Bezwzględnie nie stosować parametru $Q < 20$, a niewskazanie jest $Q < 35$. Stosowna wartość zależy od rodzaju obiektu: im większy obiekt, tym skuteczne wykrywanie może być przy niższym Q .
- Jeśli chodzi o wysoką precyzję, a dopuszczalne są niewykryte obiekty (FN), to wartość Q może być z przedziału 35–100. W praktyce można przyjąć wartość precyzji jako stałą.
- Jeśli celem detekcji jest wysoka czułość, to wymagania względem jakości obrazu są wyższe. Zalecane wartości parametru Q są w przedziale 70–94.

2. w etapie II przeprowadzono treningi ośmiu modeli podzielonych na dwie architektury i zbadano ich skuteczność detekcji analogicznie jak w etapie I.

Wyniki uzyskane w etapie II znacznie poszerzają - względem wyników etapu I - informacje na temat badanych architektur, bowiem zbadano już nie pojedyncze modele, a przeprowadzono eksperymenty na dwóch architekturach, poddając je zróżnicowanym procesom treningu. Dało to możliwość zaobserwowania, że miara PPV nie tylko jest w dużym stopniu odporna na stratną kompresję obrazów wejściowych, ale że jest ona odporna również na kompresję obrazów treningowych. **Ten wynik uzupełnia Tezę 1 Rozprawy.** Wpływ kompresji obrazu na TPR w modelach trenowanych w sposób standardowy (w tym modele pre-trenowane) jest inny niż wpływ na TPR modeli trenowanych lub dostrajanych z użyciem obrazów silnie skompresowanych. Dla tych drugich, powyżej pewnej wartości parametru Q występuje pozioma linia (podobnie jak dla miary AP). W modelach pre-trenowanych i trenowanych w przedziale $Q \ni [30, 96]$ standardowo występował jednostajny spadek czułości rzutując na wynik miary AP. Modele trenowane na obrazach o średniej kompresji wykazują podobieństwo do obu pozostałych grup.

Na podstawie części wyników etapu II eksperymentów przygotowano rozdział w monografii „Artificial Intelligence and Data Processing”.

W toku eksperymentów obliczeniowych obrazy ze zbioru val2017 zostały poddane kompresji algorytmem JPEG dla wszystkich możliwych wartości parametru Q, tj. liczb całkowitych od 1 do 100. Na podstawie przeprowadzonych eksperymentów scharakteryzowano zachowanie modeli detekcji i potwierdzono tezy Rozprawy:

1. **Tezę 1** o treści: Kompresja JPEG nie powoduje spadku precyzji detekcji dla parametru Q z przedziału [30, 100]. Przyczyną spadku skuteczności detekcji jest w tym przypadku pogorszenie czułości detekcji.

Potwierdzenie Tezy 1 przynoszą wykresy miar czułości i precyzji jako funkcji parametru Q, a także wartości miar PPV i TPR, przeanalizowane i zestawione w tabelach dostępnych w załączniku B. Dodatkowym potwierdzeniem są wykresy z II etapu eksperymentów.

2. **Tezę 2** o treści: Spadek skuteczności detekcji obiektów dla głębokich modeli, mierzony za pomocą miar AP dla parametru Q z przedziału [30, 100], jest podobny zarówno w modelach jednoetapowych, jak i dwuetapowych.

Tezę 2 potwierdzają wyniki pokazujące proporcjonalny spadek skuteczności we wszystkich rozpatrywanych modelach. Wyniki pierwszego etapu realizacji celów Rozprawy opublikowano w prestiżowym czasopiśmie MDPI Sensors (z IF 3.847).

3. **Tezę 3** o treści: Trening detektorów obiektów z użyciem obrazów silnie ($Q=20$) lub średnio ($Q=40$) skompresowanych poprawia skuteczność detekcji na obrazach poddanych silnej kompresji, dla parametru Q z przedziału [5, 30]. Przyczyną poprawy jest zwiększenie czułości detekcji dla tych obrazów.

Potwierdzenie Tezy 3 dają wyniki etapu II realizacji celów Rozprawy. Uzyskane dane demonstrują osiągniętą poprawę miar skuteczności detekcji dla obrazów poddanych nie tylko silnej ($Q \ni [5, 30]$) – co zawiera Teza 3), ale też średniej kompresji. Modele trenowane na obrazach poddanych kompresji z parametrem $Q = 40$ osiągały wartości miary AP nie gorsze niż wyniki bazowe dla niektórych modeli nawet w przedziale $Q \ni [5, 75]$.

Dodatkowym dorobkiem związanym z Rozprawą są:

1. Prezentacja wczesnych wyników na warsztatach dla doktorantów Polsko-Japońskiej Akademii Technik Komputerowych WDSiT, 2018.
2. Rozszerzony abstrakt i video-prezentacja, warsztaty EEML, 2020.
3. Opublikowanie repozytoriów danych etapu I oraz II eksperymentów, (dostęp: 27 września 2022).
4. Udostępnienie kodu źródłowego oprogramowania użytego do badań na platformie GitHub, (dostęp: 27 września 2022).

Doktorant wskazuje też kierunki, na które planuje rozszerzyć przyszłe badania: wpływ innych algorytmów kompresji na detekcję obiektów, np. popularny algorytmu WebP1, który też korzysta z transformacji DCT, a także:

1. użycie dodatkowych miar skuteczności detekcji, które niosą informacje np. o zmianach dokładności lokalizacji obiektów,
2. trening detektorów obiektów na większej liczbie zbiorów uczących i testowanie na innych zbiorach testowych, z uwzględnieniem postępów w dziedzinie treningu detektorów obiektów,
3. zbadanie większej liczby architektur i modeli detekcji.

4. Podsumowanie Recenzji

Na uwagę zasługuje mnogość przeprowadzonych obliczeń i eksperymentów, których wyniki mogą być wskazówkami dla specjalistów od systemów detekcji obiektów jak dobierać parametry kompresji w sposób optymalny z punktu widzenia skuteczności działania systemu.

W zakresie teoretycznym Doktorant istotnie poszerzył wiedzę o głębokich modelach detekcji obiektów w warunkach stratnej kompresji obrazu, a także o możliwości przeciwdziałania pogorszeniu skuteczności detekcji w różnych zastosowaniach tych systemów.

Nieliczne niedociągnięcia i krytyczne uwagi nie mają istotnego wpływu na zawartość i zrozumienie tekstu Rozprawy. Nie podważają też olbrzymiego dorobku Doktoranta zawartego w całej Rozprawie. Nie można też niedoceniać wiedzy i biegłości Doktoranta w zakresie omawianych w Rozprawie systemów. Zatem, Rozprawa stanowi istotny i niezaprzeczalny Jego dorobek w pełni realizujący postawione w Celu Rozprawy zadanie badawcze.

Sądzę, że przedstawiona mi do rezensji Rozprawa odpowiada z nadmiarem wymogom stawianym rozprawom doktorskim. **Wnoszę o dopuszczenie jej Autora do kolejnych etapów przewodu doktorskiego.**

W. Mokrzycki

