

Streszczenie rozprawy doktorskiej

Tytuł rozprawy: Deep learning methods for computer vision-based industrial inspection applications

Tytuł spolszczony: Zastosowanie metod uczenia głębokiego w wybranych problemach widzenia komputerowego w kontekście przemysłowej oceny jakości

Autor: mgr inż. Przemysław Dolata

Widzenie komputerowe znajduje liczne zastosowania w zadaniach automatycznej inspekcji w różnych gałęziach przemysłu. Metody oparte o uczenie głębokie umożliwiły stosowanie inspekcji wizyjnej w scenariuszach, w których techniki klasyczne okazywały się niewystarczające. Idea uczenia przez transfer oraz rosnąca dostępność głębokich modeli wizyjnych o wysokiej jakości uczyniły wdrażanie złożonych systemów wizyjnych jeszcze łatwiejszym. Jednak pomimo postępu, standardowe modele nadal nie osiągają podobnie spektakularnych wyników w pewnych dziedzinach. Oznacza to, że aby z powodzeniem wdrażać tam systemy wizyjne oparte o modele głębokie, konieczne są dalsze badania.

Celem niniejszej rozprawy jest opracowanie metod uczenia głębokiego dla wybranych zastosowań inspekcji wizyjnej, w szczególności takich, gdzie standardowe podejście oparte o uczenie przez transfer nie okazało się skuteczne lub możliwe. Przykłady takich zagadnień pojawiają się w dziedzinach takich jak rolnictwo precyzyjne, przetwórstwie żywności, a także pewnych scenariuszach sterowania procesami przemysłowymi. Wyjątkowe uwarunkowania i wymagania stawiane systemom wizyjnym w tych dziedzinach sprawiają, że ogólne modele dostępne w literaturze albo nie transferują się do nich z powodzeniem, albo nie rozwiązują w całości problemu wizyjnego. Aby osiągnąć cel rozprawy, w jej toku postawione zostały trzy pytania badawcze, odpowiedzi na które znajdują się w cyklu sześciu publikacji naukowych.

Pierwsze pytanie badawcze dotyczyło opracowania klasyfikatorów opartych o konwolucyjne sieci neuronowe do szczególnego przypadku klasyfikacji *wielowidokowej*, w której próbkę podlegającą klasyfikacji opisuje kilka obrazów. Taką formę zadania przyjmuje często automatyczna inspekcja jakości, podczas której obiekty muszą być obrazowane z wielu stron. Przykłady tak sformułowanej inspekcji często spotykane są w przetwórstwie żywności. Struktura ogólnych modeli głębokich do klasyfikacji nie pozwala na użycie ich do wnioskowania z wielu obrazów bez znaczących modyfikacji. Odpowiedź na to pytanie znajduje się w Publikacjach 1 i 2, które omawiają budowę i uczenie dwustrumieniowych konwolucyjnych sieci neuronowych. W Publikacji 1 zbadane zostały metody łączenia informacji z wielu obrazów wejściowych, a w Publikacji 2 zaproponowany został szczególny przypadek modelu klasyfikacyjnego o specjalizowanych strumieniach ekstrakcji cech.

Drugie pytanie badawcze związane było z optymalizacją struktury sieci U-Net do zadań segmentacji semantycznej wymagających przetwarzania o niskim opóźnieniu. Niektóre zadania sterowania procesami przemysłowymi, na przykład monitorowanie procesów laserowych w obróbce addytywnej, mogłyby być automatyzowane za pomocą widzenia komputerowego. Jednak w zadaniach tych niezbędne jest przetwarzanie w czasie rzeczywistym, dlatego wymagają one modeli o bardzo dużej szybkości przetwarzania. U-Net, choć oferuje wysoką jakość segmentacji, cechuje wysoka złożoność obliczeniowa, sprawiając, że model ten jest niewystarczająco wydajny do tego celu. W Publikacji 3 zaproponowana została metoda optymalizacji architektury tego modelu, pozwalająca na znaczące skrócenie czasu inferencji

przy jednoczesnym zachowaniu jakości predykcji. Publikacja ta została nagrodzona nagrodą *Best Paper* na konferencji FedCSIS 2019.

Ostatnie pytanie badawcze dotyczyło zastosowań metod uczenia głębokiego do segmentacji instancyjnej dla celów estymacji wymiarów gęsto rozmieszczonych małych obiektów na podstawie pojedynczych obrazów. Niewielki rozmiar obiektów, ich duża liczba i bliskość, często powodująca powstawanie okluzji, w połączeniu z potrzebą dokładnego szacowania ich wymiarów, to typowe wyzwania w zadaniach inspekcji w takich gałęziach przemysłu jak rolnictwo precyzyjne czy przetwórstwo żywności. Ogólne modele głębokie wykazują niską skuteczność w przetwarzaniu obrazów powstałych w takich warunkach. Ze względu na złożoność tego zagadnienia, to pytanie badawcze zostało rozbite na trzy pytania pomocnicze, które ukierunkowały badania prowadzące do powstania trzech Publikacji (4, 5, oraz 6). Pierwsze pytanie pomocnicze, zaadresowane w Publikacji 4, dotyczyło problemu estymacji wymiarów niewielkich, trójwymiarowych obiektów obrazowanych za pomocą klasycznych kamer bez informacji o głębi. Drugie pytanie pomocnicze polegało na zaproponowaniu metody uczenia modeli *amodalnej* segmentacji instancyjnej, to znaczy modeli zdolnych do segmentacji obiektów częściowo przysłoniętych, bez amodalnej anotacji danych uczących; odpowiednie podejście zaproponowano w Publikacji 5. Trzecie pytanie pomocnicze związane było z poprawą jakości detekcji małych obiektów w obrazach o wysokiej gęstości. W Publikacji 6 zaprezentowana została nowa metoda detekcji obiektów specjalizowana do takich przypadków.

Niniejsza rozprawa podsumowuje rezultaty sześciu pierwszoautorskich publikacji naukowych. W chwili złożenia rozprawy, publikacje te były cytowane 37 razy, co potwierdza ich wpływ naukowy. Ponadto, cztery z tych publikacji zostały opracowane w drodze prac badawczych związanych z realizacją grantów badawczo-rozwojowych fundowanych przez Narodowe Centrum Badań i Rozwoju, co podkreśla ich praktyczną wartość w dziedzinie automatycznej inspekcji wizyjnej.

Przemysław Zieliński