

Prezentacja przedstawia rezultaty prac badawczych prowadzonych w ramach pracy doktorskiej nt. **„Przetwarzanie sekwencji wideo przy zastosowaniu podejścia deep learning do automatycznego generowania instrukcji stanowiskowych”**. Pracę prowadzono zgodnie z przyjętymi celami badawczymi. W rezultacie opracowano nową metodę przetwarzania sekwencji wideo przy zastosowaniu podejścia deep learning (CNN, CNN + SVM, CNN + R-CNN/YOLO) do automatycznego generowania instrukcji stanowiskowych w postaci filmu animowanego, pozwalającą na jednoczesną weryfikację wykonywanych czynności serwisowych. Metodę opracowano dla działu utrzymania ruchu w przedsiębiorstwie, w którym prowadzone są szkolenia dla osób podejmujących zatrudnienie na danym stanowisku pracy, w rozważanym przypadku na stanowisku serwisanta kotła na paliwo stałe.

Przeprowadzono 3 eksperymenty badawcze z wykorzystaniem 8 nagrań wideo przedstawiających czynności serwisowe kotła na paliwo stałe. Eksperymenty badawcze przeprowadzono z wykorzystaniem: Programu Matlab R2019a, konwolucyjnych sieci neuronowych (bazujących na architekturach Cifar10Net, Resnet18, Alexnet) z klasyfikatorami Softmax oraz SVM, detektora obiektów R-CNN działającego w oparciu o sieć konwolucyjną.

**Eksperyment 1:** Wyodrębniono cechy referencyjnych i testowych klatek materiału wideo (Algorytm 2). Na podstawie cech klatek referencyjnych wyznaczono klatki testowe odpowiadające klatkom wzorcowym (Algorytm 4). Do identyfikacji obiektów (Algorytm 3) na obrazie użyto sieci konwolucyjnej z dwoma typami klasyfikatorów. Zastosowanie CNN z SVM pozwoliło na identyfikację 3 z 6 rodzajów obiektów znajdujących się na klatkach obrazu, CNN z klasyfikatorem Softmax umożliwił wykrycie tylko 1 obiektu.

**Eksperyment 2:** Wyznaczono etapy czynności serwisowej (Algorytm 1) na podstawie których wyodrębniono cechy referencyjnych klatek materiału wideo (Algorytm 2). Na podstawie cech klatek referencyjnych wyznaczono klatki testowe odpowiadające klatkom wzorcowym (Algorytm 4). Zidentyfikowano obiekty znajdujące się na klatkach obrazu (Algorytm 3). Porównano cechy klatek obrazu a na podstawie otrzymanych danych wyznaczono klatki o podobnej „zawartości”. Skuteczność wyniosła 94,009%. Wygenerowano scenariusz postępowania na podstawie kolejnych etapów czynności serwisowej (Algorytm 5).

**Eksperyment 3:** Wyznaczono etapy czynności serwisowej (Algorytm 1). Wyodrębniono cechy referencyjnych klatek materiału wideo (Algorytm 2). Na podstawie cech klatek referencyjnych wyznaczono klatki testowe odpowiadające klatkom wzorcowym (Algorytm 4). Skuteczność: Resnet18 94%, Alexnet 87,51%. Zidentyfikowano obiekty znajdujące się na klatkach obrazu (Algorytm 3) z wykorzystaniem detektora R-CNN oraz YOLOv3-COCO. Porównano zidentyfikowane obiekty znajdujące się na klatkach referencyjnych i testowych (Algorytm 4). Skuteczność detekcji obiektów Cifar10Net 61,66%, Alexnet 48,33%, YOLO 73,15%. Wygenerowano scenariusz postępowania na podstawie kolejnych etapów czynności serwisowej (Algorytm 5). Zaimplementowano element kontroli czasu realizowanego etapu czynności (Algorytm 6).

W końcowym etapie prac zbudowano system informatyczny wspomagający proces szkoleń pracowniczych działający w oparciu o zaproponowany model. Rozwiązanie działa w oparciu o utworzony ręcznie dataset zawierający 3440 plików graficznych zawierających w sumie zaznaczenia obiektów takich jak: osoba, ręka, piec, sterownik, łopata, wiadro, kuweta, ślimak, zasobnik, motoreduktor, klucz (w sumie 6619 zaznaczonych obiektów dla 11 klas). Podczas treningu sieci 80% datasetu zakwalifikowano jako dane treningowe a 20% jako dane walidacyjne. Realizacja zadań związanych z systemem informatycznym służącym do analizy obrazu zakładała wykorzystanie do tego celu architektury sieci Yolov3, jednak szybkość działania systemu informatycznego ograniczona została przez możliwości techniczne maszyny na której aplikacja została uruchomiona (szybkość pracy systemu detekcji czynności dla modelu z Yolov3 po implementacji autorskich algorytmów wynosiła ok 5 klatek/sekundę). W związku z powyższym przyjęto architekturę sieci opartą na Yolov3 Tiny która pozwalała uzyskać około 26-27 klatek/sekundę.