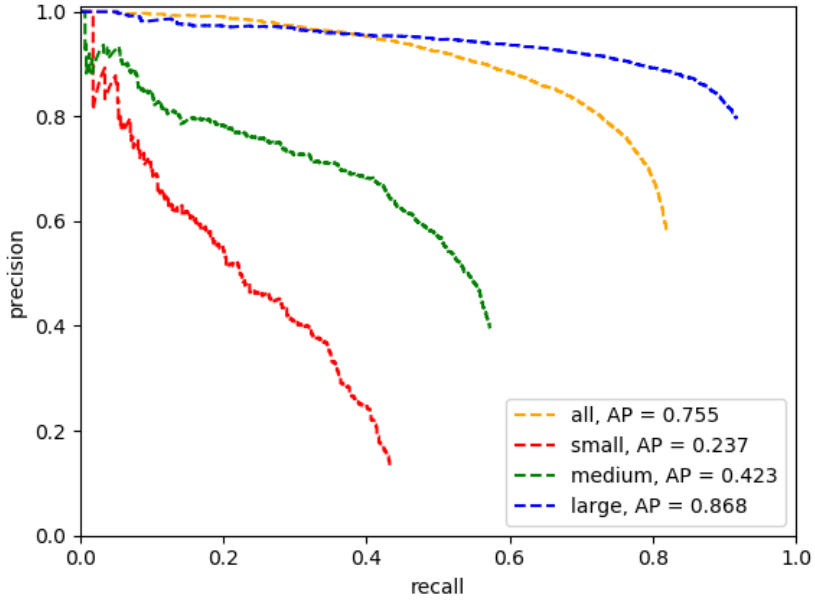
V Analytics - samouczący się system analizy ruchu osób na podstawie obrazów z kamer w warunkach ograniczonych zasobów obliczeniowych

Projekt realizowany
w ramach Regionalnego Programu Operacyjnego Województwa Śląskiego na lata 2014-2020
(Europejski Fundusz Rozwoju Regionalnego)
dla Osi Priorytetowej: I. Nowoczesna gospodarka
dla działania: 1.2. Badania, rozwój i innowacje w przedsiębiorstwach

Okres realizacji projektu:	01.09.2019 – 31.08.2021
Wydatki kwalifikowane:	3 688 296,76 PLN
Wnioskowane dofinansowanie:	2 632 460,14 PLN

Projekt badawczy miał na celu opracowanie systemu analizy obrazu umożliwiającego detekcję i śledzenie obiektów w warunkach ograniczonej mocy obliczeniowej. W ramach projektu przeprowadzono zarówno badania przemysłowe, jak i prace rozwojowe, które miały na celu stworzenie hybrydowych rozwiązań łączących głębokie sieci neuronowe z mniej złożonymi algorytmami klasycznymi, co pozwoli na bardziej efektywną detekcję i śledzenie obiektów. System miał również zawierać fazę adaptacji do specyfiki danej lokalizacji, która mogłaby przebiegać automatycznie lub półautomatycznie. Rezultatem projektu miało być rozwiązanie (V Analytics) o cechach nowatorskich w skali globalnej, stanowiące innowację na rynku docelowym, zgodnie ze stanem wiedzy na 2018 rok.
Bezpośrednim celem projektu było opracowanie przez IOT sp. z o.o. nowego produktu w postaci samouczącego się systemu do detekcji i śledzenia osób na podstawie obrazów z kamer, działającego w warunkach ograniczonych zasobów obliczeniowych (V Analytics). System ten oparty na zaawansowanych algorytmach uczenia maszynowego, miał wyróżniać się wysokim poziomem innowacyjności, oferując autorskie rozwiązania algorytmiczne, które dotychczas nie były stosowane ani znane na rynku globalnym.

Etap 1: Opracowanie architektury i trening sieci neuronowej TRL III
W ramach tego etapu stworzono architekturę oraz wytrenowano sieć neuronową do detekcji. Jej architektura ewoluowała w trakcie trwania projektu, od podobnej do detektora YOLO (https://pjreddie.com/darknet/yolo/), przez podobną do połączenia detektorów YOLO i SSD (https://arxiv.org/abs/1512.02325), aż po finalną wersję, którą można podsumować jako połączenie architektury ResNet (https://arxiv.org/abs/1512.03385) z detektorem SSD. Zmiany wynikają po części z prac nad poprawieniem skuteczności, ale też z zwiększaniem szybkości i ułatwieniem późniejszej implementacji w warunkach ograniczonej mocy obliczeniowej.
W pierwszym podejściu do optymalizacji sieci została ona zmniejszona. Wypróbowano wiele wariantów by wybrać taki, który wiąże się z jak najmniejszym spadkiem skuteczności. Sprawność detekcji (mAP - mean average precision) była mierzona na obrazach pochodzących z archiwalnych nagrań z monitoringu udostępnionych przez klienta. Na tak przygotowanych danych wyniosła: mAP=73,7% (żółta linia, opisana jako “all”) dla dużej wersji sieci i mAP=71,5% dla zmniejszonej, co spełnia założenia projektu.



Należy mieć na uwadze, że skuteczność zależy od rozmiaru wykrywanych obiektów. Dla dużych obiektów (czyli takich jakie najczęściej się pojawiają na obrazach z kamer) skuteczność jest nieco większa. Dodatkowo stworzona została sieć neuronowa wykrywającą samochody, której skuteczność wynosi mAP=75,5%.

Opracowane śledzenie obiektów jest wzorowane na tzw. sieci syjamskiej (https://en.wikipedia.org/wiki/Siamese_neural_network). Została ona połączona z detektorem MKnet w celu zwiększenia szybkości. Sieć MKnet oprócz wykrywania obiektów dla każdego z nich wylicza także pewien wektor cech. Wektory cech dla dwóch obiektów są ze sobą porównywane w celu wyliczenia prawdopodobieństwa, że te dwa obiekty to tak naprawdę ten sam obiekt. Takie prawdopodobieństwo jest następnie używane podczas śledzenia do lepszego łączenia trajektorii. Cechy wyliczane przez sieć nie zostały ręcznie dobrane, są to abstrakcyjne liczby, dobrane podczas procedury treningowej tak by umożliwiały jak najlepsze odróżnianie obiektów od siebie. Taki algorytm jest uniwersalny, może zostać zastosowany zarówno do śledzenia ludzi jak i samochodów bez żadnych modyfikacji (oprócz konieczności wytrenowania odpowiedniej sieci neuronowej). Rozwiązanie przetestowano na zestawie danych OTB i otrzymano wynik 76% dla śledzenia ludzi, oraz 74% dla śledzenia samochodów.

Zastosowanie trackera LT1 w warunkach ograniczonej mocy obliczeniowej opierało się na fakcie, że kamery monitoringu pozostają nieruchome. W związku z tym, dwie kolejne klatki z nagrania różnią się jedynie w niewielkim stopniu. Nie ma zatem potrzeby wykonywania obliczeń dla całego obrazu za pomocą sieci neuronowej, wystarczy skupić się na fragmentach, które uległy zmianie. Aby zweryfikować, jaka część obrazu zmienia się między dwoma klatkami, przygotowano kilka filmów testowych. Obraz z kamery został podzielony na kafelki, a na przetworzonych filmach podświetlono te kafelki, które w istotny sposób

IoT System Sp. z o.o., z siedzibą w Katowicach przy ul. Zofii Nałkowskiej 5/2; 40-425 Katowice, NIP: 9542763070, REGON: 363890923, wpisana do Krajowego Rejestru Sądowego w Sądzie Rejonowym w Katowicach nr KRS 0000606263, kapitał zakładowy 5 000,00 zł.

zmieniły się z klatki na klatkę. W lewym górnym rogu wyświetlono średnią i chwilową zmianę. Filmy z benchmarku OTB nie są optymalnym przykładem ze względu na ich krótki czas trwania, przez co początkowy koszt obliczenia całej sieci neuronowej nie zdąży się zamortyzować w czasie. Natomiast na nagraniach z rzeczywistych lokalizacji udostępnionych przez firmę Netizens, średnio zmienia się od 2% do 20% obrazu, nawet w warunkach dużego zatłoczenia. Oznacza to, że możemy oczekiwać przyspieszenia inferencji sieci neuronowych od 5-krotnego (1/0.2) do 50-krotnego (1/0.02).

Wydajność przetwarzania na zastosowanym komputerze wyposażonym w procesor Intel i7-3770 wynosi około 100 klatek na sekundę (fps). Wydajność ta jest zmienna i zależy od chwilowego natężenia ruchu, stąd średnia wartość 100 fps została uzyskana z całego nagrania użytego do testów. Nowoczesne akceleratory VPU, takie jak Intel Movidius, mogą być używane jako uzupełnienie mocy obliczeniowej. Testy przeprowadzone przez zespół wskazują na wydajność na poziomie 30 fps dla sieci MKnet. Jednak ta wydajność jest na chwilę obecną zbyt niska, aby VPU stanowiły poważną alternatywę. Pomimo to, wyniki są wystarczająco obiecujące, aby obserwować rozwój rynku akceleratorów VPU, ponieważ w przyszłości prawdopodobnie pojawią się szybsze warianty tych urządzeń.
Zastosowanie architektury klient-serwer, mającej na celu odciążenie lokalnych zasobów obliczeniowych, wydaje się być dobrą alternatywą dla przetwarzania obrazu lokalnie. Jednak przeprowadzona analiza techniczna i biznesowa wykazała, że klienci, którzy mogliby najwięcej zyskać dzięki temu rozwiązaniu, ze względu na niską wydajność ich komputerów, mają jednocześnie stosunkowo niską przepustowość łącza internetowego. Transmisja filmów na zewnętrzny serwer wiązałaby się dla nich z nieakceptowalnym obciążeniem łącza, dlatego nie planuje się kontynuacji prac w tym kierunku.
Na potrzeby ręcznego tworzenia danych treningowych opracowano specjalny edytor, który umożliwia obrysowywanie ludzi prostokątami. Mimo że zrezygnowano z fazy adaptacyjnej podczas instalacji systemu w nowej lokalizacji, edytor ten będzie nadal wykorzystywany do zbierania większej ilości danych w przyszłości.

Etap 2: Integracja technologii i optymalizacja TRL IV
Na tym etapie głównym celem prac było zwiększenie gotowości technologicznej systemu śledzenia, co obejmowało rozpoczęcie prób laboratoryjnych oraz połączenie wszystkich komponentów w spójną całość. Równocześnie, prace skupiły się na inżynierii oprogramowania, aby zapewnić kompatybilność zarówno na poziomie danych, jak i architektury całego systemu. Oba te cele zostały osiągnięte. W instalacji laboratoryjnej połączono dane z rejestracji trajektorii metodą radiową z danymi pochodzącymi z systemu śledzenia obrazu wideo. Zgodność trajektorii okazała się bardzo wysoka, co umożliwiło obliczenie istotnych parametrów, takich jak liczba osób wchodzących do określonego obszaru.
Ponadto uzyskano bardzo obiecujące wyniki wstępnej optymalizacji ewaluatora sieci, który w rzeczywistych warunkach będzie działał z ograniczoną dostępnością do mocy obliczeniowej. Zweryfikowano działanie komponentów takich jak tracker, detektor oraz tracker LT1 w warunkach operacyjnych, wykorzystując archiwalne nagrania z kamer z lokalizacji, w których system miałby działać. Skuteczność liczenia liczby osób przechodzących przez dane miejsce wyniosła 75% dla systemu działającego bez ograniczeń mocy obliczeniowej oraz około 72%

IoT System Sp. z o.o., z siedzibą w Katowicach przy ul. Zofii Nałkowskiej 5/2; 40-425 Katowice, NIP: 9542763070, REGON: 363890923, wpisana do Krajowego Rejestru Sądowego w Sądzie Rejonowym w Katowicach nr KRS 0000606263, kapitał zakładowy 5 000,00 zł.

przy ograniczeniu mocy obliczeniowej do pojedynczego rdzenia CPU, bez żadnej fazy adaptacyjnej.

Etap 3: Testowanie systemu w warunkach operacyjnych TRL V
Projekt zakładał wprowadzenie fazy treningowej lub adaptacyjnej dla systemu działającego w warunkach ograniczonej mocy obliczeniowej. Planowano, że poprzez dostosowanie systemu do konkretnej lokalizacji można by zmniejszyć stopień generalizacji układów analizujących, co z kolei miałoby na celu ograniczenie ilości obliczeń. Jednak w toku prac badawczych zdecydowaliśmy się zrezygnować z tego podejścia, argumentując to kilkoma istotnymi powodami. Po pierwsze, wydajność systemu działającego na CPU okazała się na tyle wysoka, że nie było potrzeby dodatkowego zmniejszania liczby obliczeń. Po drugie, skuteczność detekcji na CPU była równa lub zbliżona do tej osiąganej przez oryginalną implementację na GPU, a podczas testów nie zaobserwowano żadnego pogorszenia jakości detekcji. Po trzecie, wprowadzenie fazy adaptacyjnej wymagałoby znacznych nakładów pracy człowieka, związanego z koniecznością zebrania i oznaczenia danych treningowych dla każdej nowej lokalizacji. Z punktu widzenia przyszłych zastosowań systemu, bardzo pożądane było zminimalizowanie tej pracy, najlepiej do zera, nawet jeśli miałoby to nastąpić kosztem niewielkiego spadku skuteczności systemu.

Etap 4: Walidacja systemu w symulowanych warunkach operacyjnych TRL VI
Celem tego etapu była walidacja systemu oraz przeprowadzenie testów w symulowanych warunkach operacyjnych. Za warunki operacyjne uznano zarówno ograniczenie mocy obliczeniowej, np. do pojedynczego CPU, jak i korzystanie z obrazu wideo zarejestrowanego przez kamery działające w nieoptymalnych warunkach, przy różnych zakłóceniach, takich jak zmienne oświetlenie. Podczas testów w tych warunkach zbadano:

- Wydajność symulowanych serwerów do przetwarzania i analizy obrazu w systemie o ograniczonej mocy obliczeniowej była identyczna z wynikami uzyskanymi w warunkach laboratoryjnych. Nie spodziewano się tutaj żadnych niespodzianek.
- Na wcześniejszym etapie zrezygnowano z konieczności przeprowadzania dodatkowej fazy treningowej dla systemu działającego w warunkach ograniczonej mocy obliczeniowej.
- Choć akceleratory VPU mogą być obiecującym rozwiązaniem w przyszłości, obecnie ich wydajność nie jest jeszcze wystarczająca.
- Na poprzednim etapie stwierdzono, że stosowanie architektury klient-serwer, z biznesowego punktu widzenia, nie jest możliwe.

Testy w symulowanych warunkach operacyjnych potwierdziły skuteczność i gotowość systemu, co było zgodne z wynikami uzyskanymi podczas wcześniejszych testów laboratoryjnych.

Etap 5: Testy funkcjonalne systemu TRL VII
Ewaluacja trackera polegała na podłączeniu go do systemu MovStat w celu przetestowania skuteczności wyliczania liczby osób przechodzących przez wybrane miejsce, z wykorzystaniem systemu do wyrywkowego testowania skuteczności. Choć instalacja systemu smart-tagów w warunkach operacyjnych nie była możliwa, zespół przeprowadził wszystkie

IoT System Sp. z o.o., z siedzibą w Katowicach przy ul. Zofii Nałkowskiej 5/2; 40-425 Katowice, NIP: 9542763070, REGON: 363890923, wpisana do Krajowego Rejestru Sądowego w Sądzie Rejonowym w Katowicach nr KRS 0000606263, kapitał zakładowy 5 000,00 zł.

przygotowania i testy przy użyciu instalacji zamontowanej w biurze firmy. W przyszłości ewentualna instalacja u klientów końcowych nie powinna napotkać żadnych trudności. Zaprojektowano dwa systemy pozyskiwania danych do oceny skuteczności: jeden bazujący na smart-tagach, a drugi na ręcznym oznaczaniu przejść przez dane miejsce. Z perspektywy pozostałych komponentów trackera, nie ma znaczenia, z którego z tych dwóch źródeł pochodzą dane. Ponieważ wcześniejsze badania i testy wykazały, że smart-tagi działają ze 100% skutecznością, podobnie jak ręczne przygotowanie danych, podane wcześniej wyniki skuteczności są niezależne od metody generowania danych testowych.

Etap 6: Testy operacyjne w warunkach rzeczywistych TRL VIII
W ramach prac związanych z VIII poziomem gotowości technologicznej przeprowadzono ocenę systemów w celu potwierdzenia zgodności z założeniami projektowymi.

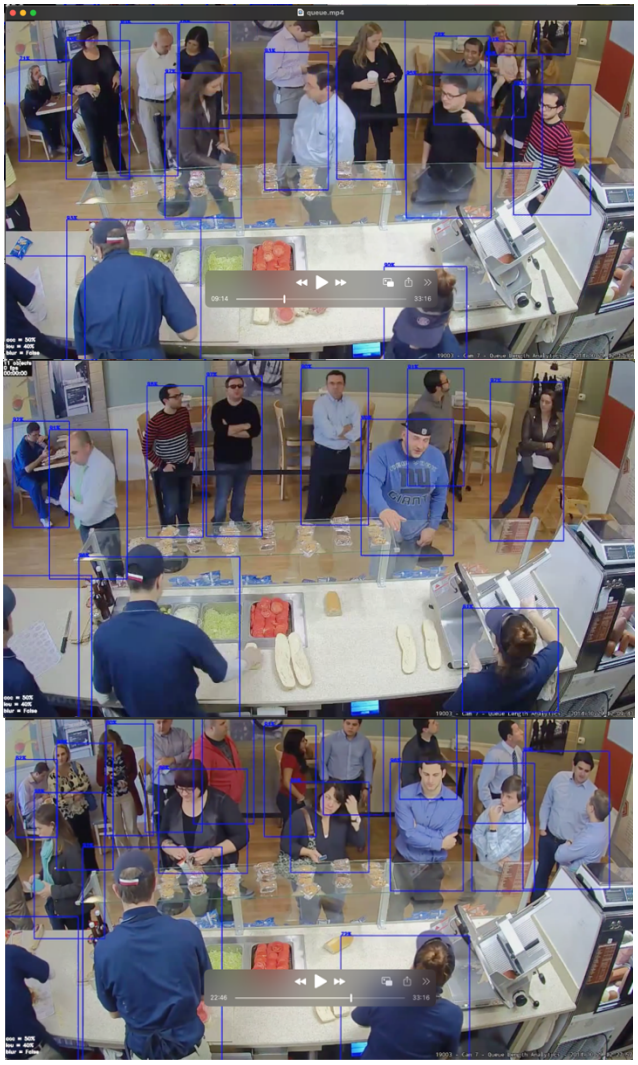
1. Dostosowano moduł analiz statystycznych trajektorii do możliwości systemu śledzenia. W zależności od jakości i liczby kamer oraz zdolności wykorzystania wszystkich danych przez moduł śledzenia, możliwe będzie uzyskanie mniej lub bardziej szczegółowych statystyk. Komentarz: Jeśli interesujący obszar dla klienta jest dobrze widoczny na danej kamerze, możliwe jest liczenie każdej statystyki. Nie ma innych ograniczeń.
2. Określono procedurę adaptacji systemu do nowej lokalizacji, która obejmuje szkolenie osoby wdrażającej oraz sposób instalacji oprogramowania.
3. Zaimplementowano narzędzie do adaptacji systemu w nowych lokalizacjach, które zawiera funkcje konfiguracji systemu dla analizowanych obszarów statystycznych. Komentarz: Zgodnie z wcześniejszą analizą zrezygnowano z fazy adaptacyjnej.
4. Opracowano narzędzie do wyrywkowej weryfikacji modułu śledzenia na podstawie zdefiniowanych nagrań podczas działania systemu w rzeczywistych warunkach.
5. Opracowano architekturę sprzętową, oprogramowanie oraz narzędzia do automatyzacji procesu testowania podczas adaptacji systemu w nowych lokalizacjach. Komentarz: Jak wskazano w punktach 2 i 3, choć system V Analytics nie wymaga fazy adaptacyjnej, system do wyrywkowej weryfikacji został przygotowany i zintegrowany z podobnym systemem dla programu MovStat.

Etap 7: Ostateczna integracja i walidacja TRL IX
Opracowany moduł śledzenia zostanie przetestowany w rzeczywistych warunkach poprzez jego podłączenie do funkcjonującego w firmie Netizens systemu MovStat, na podstawie zawartego porozumienia.

1. Opracowany zostanie interfejs komunikacyjny pomiędzy obecną wersją systemu a modułem śledzenia. Komentarz: System V Analytics od początku był integrowany z MovStatem w sposób ciągły, dlatego nie ma potrzeby tworzenia osobnego interfejsu komunikacyjnego. Interfejs został już wykonany w ramach wcześniejszych etapów prac.
2. Opracowany zostanie sposób zapisu danych oraz ich agregacji na potrzeby prezentacji danych dla klienta. Komentarz: System V Analytics został zintegrowany z istniejącym już w MovStacie systemem zapisu i prezentacji danych. Dodano jedynie nowe funkcje, które zostały opisane w raportach z prac.

IoT System Sp. z o.o., z siedzibą w Katowicach przy ul. Zofii Nałkowskiej 5/2; 40-425 Katowice, NIP: 9542763070, REGON: 363890923, wpisana do Krajowego Rejestru Sądowego w Sądzie Rejonowym w Katowicach nr KRS 0000606263, kapitał zakładowy 5 000,00 zł.

3. Opracowany zostanie mechanizm kolekcji danych w celu późniejszej predykcji ruchu na badanych obszarach oraz wykrywania niestandardowych zachowań. Komentarz: Jak wyżej, mechanizm został włączony do systemu.
4. Dostosowanie opracowanego rozwiązania pod działanie na systemach operacyjnych Windows, Linux, MacOS. Komentarz: W wyniku analizy systemów stosowanych u potencjalnych klientów nie stwierdzono zastosowań systemów MacOS do obsługi CCTV lub innych systemów wizyjnych, co sprawia, że przygotowanie implementacji dla MacOS nie jest zasadne.



IoT System Sp. z o.o., z siedzibą w Katowicach przy ul. Zofii Nałkowskiej 5/2; 40-425 Katowice, NIP: 9542763070, REGON: 363890923, wpisana do Krajowego Rejestru Sądowego w Sądzie Rejonowym w Katowicach nr KRS 0000606263, kapitał zakładowy 5 000,00 zł.