

Detekcja obiektów

Michał Ciesielski 145325

1. Opis problemu

Poruszony problem w tym referacie polegał na wykryciu charakterystycznych obiektów (w tym przypadku ludzi) na tle określonego środowiska. Podejście do rozwiązania tego problemu znalazło swoje zastosowanie podczas rozwoju pandemii COVID-19 w 2020 roku. Wykorzystywanym wejściem do tego problemu został więc obraz z kamery (raspberry pi4) przedstawiający pewien obszar w topologii miasta, w którym może znajdować się większe zgromadzenie ludzi. Rozpatrywany system musiał rozpoznawać 2 kluczowe elementy podczas społecznej izolacji w trakcie trwania pandemii, jakimi był odstęp 2 metrów oraz noszenie maski przez zgromadzonych. Wykrycie złamania reguł społecznej izolacji zostało w ten sposób zautomatyzowane i ograniczało w ten sposób interwencje służb porządkowych.

2. Wcześniej wykorzystywane metody w detekcji obiektów

Przed rozwojem głębokich sieci neuronowych, w detekcji obiektów stosowano głównie metody oparte na deskryptorach obrazów i klasyfikacji obiektów. Jedną z popularnych metod było wykorzystanie algorytmu Viola-Jonesa, który opierał się na kaskadzie klasyfikatorów cech Haar'a. Algorytm ten wykorzystywał okna o różnych rozmiarach do wykrycia obiektów na obrazie. Następnie, na podstawie wykrytych cech Haar'a, klasyfikator decydował, czy dany fragment obrazu zawiera obiekt czy nie.

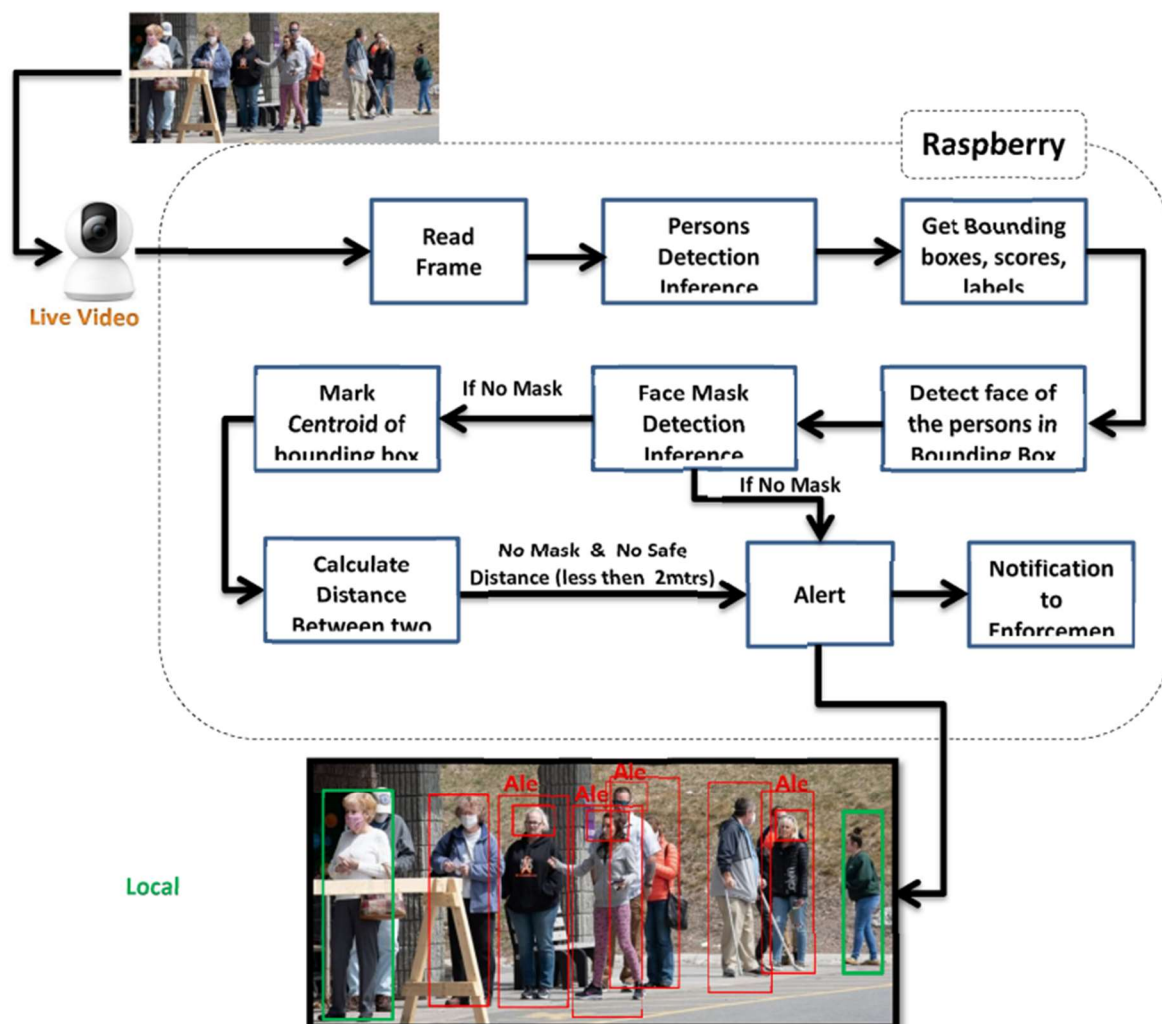
Inną metodą było wykorzystanie deskryptorów cech lokalnych, takich jak SIFT, SURF, czy ORB. Algorytmy te pozwalają na wykrycie charakterystycznych punktów na obrazie i opisanie ich w sposób umożliwiający późniejszą identyfikację. Następnie, na podstawie dopasowania punktów charakterystycznych, można było określić położenie i kształt obiektów na obrazie.

Metody te miały jednak pewne ograniczenia, takie jak złożoność obliczeniowa czy trudność w wykryciu obiektów o różnych rozmiarach i orientacjach. Dlatego też, wraz z rozwojem głębokich sieci neuronowych, metody te zostały zastąpione bardziej zaawansowanymi i skutecznymi metodami detekcji obiektów.

Swego rodzaju rewolucją w problemie detekcji obiektów okazało się wykorzystanie wspomnianych głębokich sieci neuronowych. Modele wykorzystujące głębokie sieci neuronowe do wykrywania ludzi można podzielić na dwie rodziny. Pierwszą z nich jest rodzina 2-etapowych detektorów, takich jak R-CNN, Fast R-CNN i Faster-CNN, drugą z kolei jest rodzina 1-etapowych detektorów np. SSD, YOLO. R-CNN (Region-based Convolutional Neural Network) to pierwszy algorytm, który wprowadził pojęcie regionów zainteresowania (ROI) do wykrywania obiektów. Algorytm dzielił obraz na wiele mniejszych obszarów (ROI) i dla każdego z tych obszarów stosował sieć CNN w celu wykrycia obiektów. Kolejne wersje tego typu algorytmów wprowadzały liczne usprawnienia w celu poprawienia szybkości działania np. dla algorytmu Fast R-CNN zamiast przetwarzać każdy ROI z osobna, obraz był przetwarzany tylko raz przez sieć CNN, a następnie dla każdego ROI wykorzystywano jedną z warstw sieci do ekstrakcji cech.

Ze względu na ograniczenia mocy obliczeniowej i wykorzystanie raspberry pi4 konieczne było wykorzystanie sprawniejszych algorytmów. W tym celu wykorzystano algorytm SSD przetwarzający obraz jednym przejściem sieci neuronowej.

3. Opis zastosowanej metody w artykule



Zdj. 1 Diagram architektury systemu

Zaproponowany w artykule system wykorzystywał podejście transfer learningu do optymalizacji wydajności. Kamera dostarcza z wideorejestratora ramki w odcieniach szarości do dalszego przetwarzania. Dalej wykorzystywana jest architektura MobileNetV2, w tym wspomniany wcześniej algorytm SSD do znajdowania ludzi na obrazie. Model został już wytrenowany na dużym zbiorze obrazów takich jak ImageNet i PascalVOC, zgodnie z ideą transfer learningu. W trakcie kolejnego etapu, właściwego strojenia odpowiednich parametrów, wykorzystany został niski współczynnik uczenia, aby zapobiec dużemu odchyleniu na już zastosowanych filtrach konwolucyjnych. Metody te testowano korzystając z modułów OpenCV i TensorFlow. Algorytm wykrywania ludzi w czasie rzeczywistym SSD przy użyciu modułu MobileNetV2 i OpenCV osiągnął 91,2% mAP, co oznaczało wynik zdecydowanie lepszy od porównywanego algorytmu „state-of-the-art” –

Faster R-CNN. Wykrywani ludzie zostali oznaczeni za pomocą prostokątnej ramki. Aby obliczyć odległość między poszczególnymi osobami na nagraniu wykorzystano podobieństwo trójkątów. Najpierw wyznaczono ogniskową obiektywu kamery, potem odległość człowieka od kamery, a na końcu odległość euklidesową pomiędzy środkowymi punktami 2 zaznaczonych prostokątnych obszarów. Z kolei do wykrycia twarzy znowu wykorzystano architekturę MobileNetV2, wytrenowaną wcześniej na podstawie zbioru zawierającego 3165 zdjęć twarzy z założoną maską lub nie. Podobnie jak w przypadku całego ciała człowieka, twarz została zaznaczona prostokątnym obszarem. Zaproponowany system monitorował miejsca publiczne i jeśli znalazła się w kontrolowanym obszarze osoba bez maski, wysyłany był sygnał do lokalnych władz z twarzą danej osoby. W tym samym czasie mierzony był dystans pomiędzy poszczególnymi osobami. Jeśli więcej niż 20 osób naruszało w danym momencie obowiązujący dystans społeczny przez określony czas, sygnał był wysyłany do lokalnego oddziału policji.

4. Wyniki działania metody

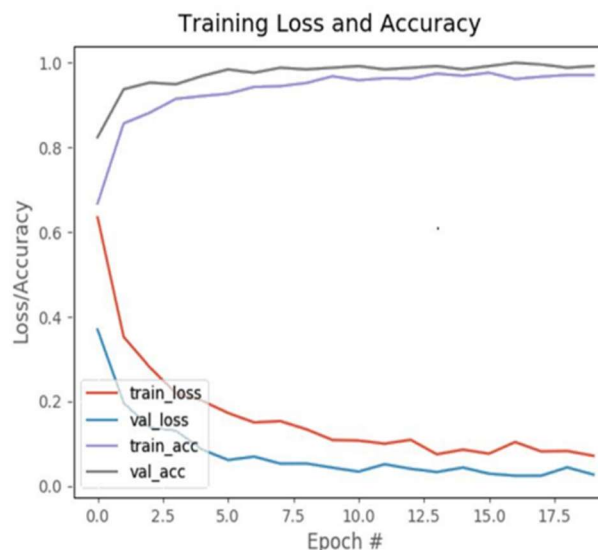
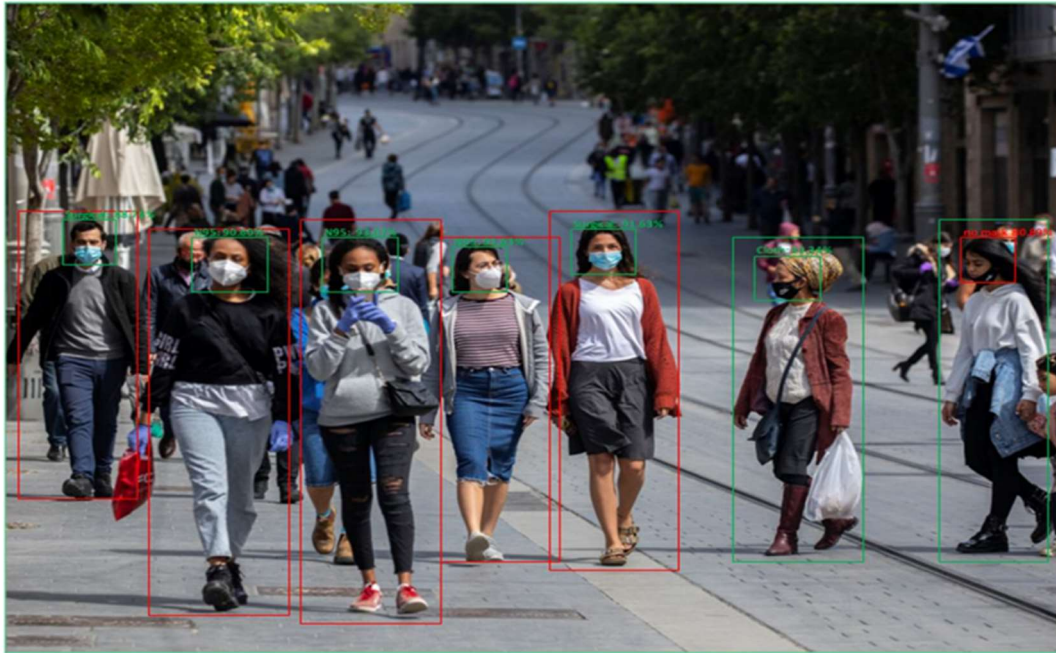


Fig 4: Model training accuracy/loss curves

Zdj. 2 Krzywe uczenia

W opisywanym podejściu 80% z 3165 obrazów wykorzystano w celu wyuczenia sieci wykorzystującej algorytm single shot detector. W pierwszym etapie wymiary obrazów były przekształcane do wymiaru 224x224 pikseli, przekonwertowane na tablice w NumPy i wzbogacone o odpowiednie etykiety do klasyfikacji. Model został wytrenowany na bazie 1000 kroków wykorzystując optymalizator ADAM. System wykrywa dystans społeczny i maski na twarzach z precision score równym 91,7% ze współczynnikiem ufności 0,7, precision value 0,91 i recall value 0,91 z FPS = 28,07.



Zdj. 3 Wynik testowy modelu

Rezultaty uzyskiwane przez ten system okazały się satysfakcjonujące. Oznacza to, że system mógł być z powodzeniem wykorzystywany w takich miejscach publicznych jak lotniska, stacje metra, galerie handlowe. Badane aspekty wykrywania niebezpieczeństw rozprzestrzeniania się chorób w trakcie pandemii lub sezonu grypowego mogą zostać wzbogacone o kolejne, takie jak na przykład wykrywanie kaszlu i kichania, badanie temperatury.

Link do artykułu:

https://www.researchgate.net/profile/Shashi-Yadav-8/publication/343346690_Deep_Learning_based_Safe_Social_Distancing_and_Face_Mask_Detection_in_Public_Areas_for_COVID-19_Safety_Guidelines_Adherence/links/5f2a488c299bf13404a24a8a/Deep-Learning-based-Safe-Social-Distancing-and-Face-Mask-Detection-in-Public-Areas-for-COVID-19-Safety-Guidelines-Adherence.pdf