1. Streszczenie

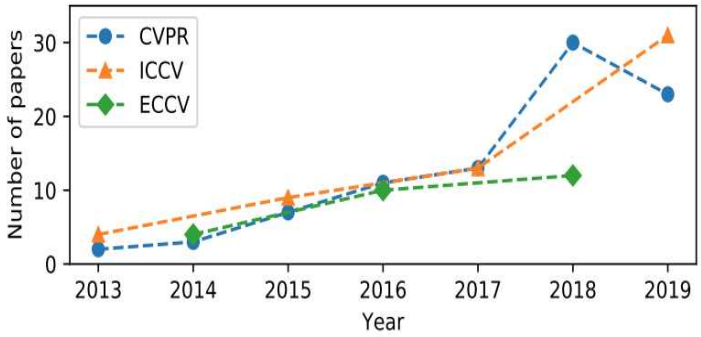
Praca zawiera opis oraz definicję problemu reidentyfikacji obrazów. W dokumencie przedstawiono motywację do wyboru powszechnie wykorzystywanego systemu do testowania własnych rozwiązań problemu reidentyfikacji. Wykorzystanie narzędzia do wytestowania architektury EfficientNet oraz autorskich modyfikacji. W pracy przedstawiono przykład zastosowania systemu w połączeniu z systemem detekcji osób. Wypróbowano wykorzystać do śledzenia obiektów na własnym datasetach. Zaproponowano dodatkową augumetnację danych oraz przedstawiono wyniki na podtawie którch wyciągęto wnioski dotyczących dalszych badań.

Praca zawiera opis oraz definicję problemu reidentyfikacji obrazów. W dokumencie przedstawiono motywację do podjęcia się tego tematu oraz wymieniono przykładowe obecnie istniejące rozwiązania tego zagadnienia. W dalszej kolejności zaproponowano własne modyfikację do istniejącego rozwiązania, w tym użycie sieci EfficientNet jako sieci bazowej. Wykonano szereg testów oraz przestawiono ich wyniki oraz wnioski z rekomendacjami do dalszych badań.

2. Wstęp

2.1 Motywacja

Problem rozpoznawania osób i obiektów na różnych ujeciach z wielu kamer stał się w ostatnim czasie jednym z najczęściej badanych zagadnień. Potwierdza to cytac z pracy [4] Torchreid: Library for Deep Learning Person Re_Identyfication in Pytoch "Driven by the growing demands for intelligent surveillance and forensic applications, person re-identification (re-ID) has become a topical research area in computer vision." w wolnym tłumaczeniu "Rozwiązanie umożliwiające identyfikację osób z różnych kamer oraz ujęć stało się jednym z najczęściej badanych zagadnień. Stało się tak za sprawą rosnącego zainteresowania służ wykorzstaniem aplikacyjnym takiego rozwiązania". Na potwierdzenie tego wniosku przedstawiono wykres obrazujący ilość publikacji traktujących o tej tematyce na przestrzeni lat.



Wykres 1: Zestawienie ilości publikacji dotyczących reidentyfikacji w latach. [6]

Potencjał rozwiązania tego zagadnienia wykracza jednak poza użycie jakim są zainteresowane służby bezpieczeństwa. Jest zdecydowanie bardziej ogólnym problemem dającym możliwość przypisywania identyfikatora dla obiektów tej samej klasy. Uogólnia zatem temat detekcji oraz rozszerza dziedzinę jej rozwiązań o klasy pośrednie w stusunku do wykorzystanych w procesie uczenia.

Warto również zaznaczyć, że zgodnie z [6] znaczący postęp w tej tematycie dokonał się właśnie dzięki GSN (głębokim siecią neuronowym). Odnosząć się do cytatu "Person re-identification (ReID) with deep neural networks has made progress and achieved high performance in recent years. However, many state-of-the-arts methods design complex network structure and concatenate multibranch features. In the literature, some effective training tricks or refinements are briefly appeared in several papers or source codes" należy zauważyć, że zgodnie z przytoczonym fragmentem rozwój w tej dziedzinie jest bardzo dynamiczny i istnieje przestrzeń na łaczenie wielu proponowanych rozwiązń.

2.2 Definicja problemu

2.3 Cel pracy

Dalsza cześć pracy wymaga sformułowania założeń. Te natomiast sprowadzją się do postawienia celi jakie adresować będzie ta praca. Zawierto je w dwóch punktach:

- Modyfikacja istniejącego rozwiązania z wytestowaniem nowewej dla frameworku architektury sieci.
- Wykorzystnie rozwiązani jako cześci systemu śledzenia obiektu na kolejnych ujęciach z nagrania.

Założenia czynione w dalszej części pracy będą wynikały z powyższych punktów.

Pierwszy z nich odnosi sie do wykorzystania wybranego frameworku jako bazowego rozwiązania, będącego wstępem do modyfikacji. Porównania wyników jakie uzyskano oraz próbie modyfikacji modeli w celu poprawy wyników dla wyspecyfikowanego typu zagadnienia. Próby będą dotyczyły zmiany architektury sieci, trybów uczenia, zmiany funkcji aktywacji oraz modyfikacji optymalizowanej funkcji strat. Porównane zostaną frameworki
[1]reid-strong-baseline
oraz
[4]Torchreid: Library for Deep Learning Person Re_Identyfication in Pytoch
Opisane zostaną powody dla, któych wybrano jeden z nich oraz sposób wykorzystania do finetuningu pretrenowanych sieci.

Drugie założenie specyfikuje problem. Nadaje kierunek w jakim ma podążać optymalizaja wyników oraz cech sieci. W tym dostosowanie sieci do pracy w trybie "real time" co wymusza skracanie czasu obliczeń i zmiejszanie sieci przy zachowaniu tego samego poziomu dokładności. Powody wybieranych rozwiązań upatrywać należy właśnie w spełnieniu tego założenia.
Praca zawiera opis oraz definicję problemu reidentyfikacji obrazów. W dokumencie przedstawiono motywację do wyboru powszechnie wykorzystywanego systemu do testowania własnych rozwiązań problemu reidentyfikacji. Wykorzystanie narzędzia do wytestowania architektury EfficientNet oraz autorskich modyfikacji. W pracy przedstawiono przykład zastosowania systemu w połączeniu z systemem detekcji osób. Wypróbowano wykorzystać do śledzenia obiektów na własnym datasetach.

3. Wybór systemów do reidentyfikacji

W badaniach rozważano wykorzystanie dwóch powszechnie używanych framework-ów:

- [1] reid-strong-baseline

Framework z szeroko opisanymi usprawnieniami w procesie trenowania. Zawiera szeroką gamę zbiarów danych oraz wynik zaprezentowanych przez twórców jest obecnie najlepszym opublikowanym wynikiem problemu reidentyfikacji.

- [5] Torchreid: Library for Deep Learning Person Re_Identyfication in Pytoch

Jest to framework napisany z użyciem PyTorch i jest kompleksowym narzędziem do tworzenia i porównywania sieci stworzonych do reidentyfikacji sieci. Zawiera wiele baz danych.
Ostatecznie zdecydowano o wykorzystaniu [1] reif-strong-baseline. Argumentami przeważajacymi na wyborze tego systemu były:

- dokładny opis wykorzystanych usprawnień w procesie trenowania.
- duża elastyczność rozwiązania umożliwiająca znaczące modyfikacje.

Poniżej zaprezentowano wyniki jakie zawarto w repozytorium github framework-u reid-strong-baseline:

Results (rank1/mAP)

Model	Market1501	DukeMTMC-reID
Standard baseline	87.7 (74.0)	79.7 (63.8)
+Warmup	88.7 (75.2)	80.6(65.1)
+Random erasing augmentation	91.3 (79.3)	81.5 (68.3)
+Label smoothing	91.4 (80.3)	82.4 (69.3)
+Last stride=1	92.0 (81.7)	82.6 (70.6)
+BNNeck	94.1 (85.7)	86.2 (75.9)
+Center loss	94.5 (85.9)	86.4 (76.4)
+Reranking	95.4 (94.2)	90.3 (89.1)

Backbone	Market1501	DukeMTMC-reID
ResNet18	91.7 (77.8)	82.5 (68.8)
ResNet34	92.7 (82.7)	86.4(73.6)
ResNet50	94.5 (85.9)	86.4 (76.4)
ResNet101	94.5 (87.1)	87.6 (77.6)
ResNet152	80.9 (59.0)	87.5 (78.0)
SeResNet50	94.4 (86.3)	86.4 (76.5)
SeResNet101	94.6 (87.3)	87.5 (78.0)
SeResNeXt50	94.9 (87.6)	88.0 (78.3)
SeResNeXt101	95.0 (88.0)	88.4 (79.0)
IBN-Net50-a	95.0 (88.2)	90.1 (79.1)

Tabela 1: Zestawienie wyników uzyskanych w pracy [6]

Wskazują one na wyniki jakie otrzymano wykorzystując każdą z kolejnych modyfikacji treningów jakie opisane zostaną w punkcie 6 tej pracy. Rozróżniono w niej wyniki wykonane na podstawie dwóch różnych datasetów:

- Market1501
- DukeMTMC-reID

Pozwala to na estymację jak każda ze zmian w procesie trenowania może poprawić wyniki dla nowo trenowanej sieci.
Zaprezentowano również wyniki w zależności od użytej sieci bazowej.

4. Data sets

Lista dostępnych datasetów obsługiwanych przez - [5] Trchroeid: Library for Deep Learning Person Re_Identyfication in Pytoch jest bogata i zawiera:

a. Datasety zawierajace pojedyńcze ujęcia:
- Market1501 [7] Market1501 - pdf
- CUHK03
- DukeMTMC-reID
- MSMT17
- VIPeR
- GRID
- CUHK01
- SenseReID
- QMUL-iLIDS
- PRID

b. Datasety ze nagraniami video:
- MARS
- iLIDS-VID
- PRID2011
- DukeMTMC-VideoReID

Dwa najczęściej wykorzystywane w tego typu zadaniach datasety to:
- Market1501 [7] Market1501 - pdf
- DukeMTMC-reID

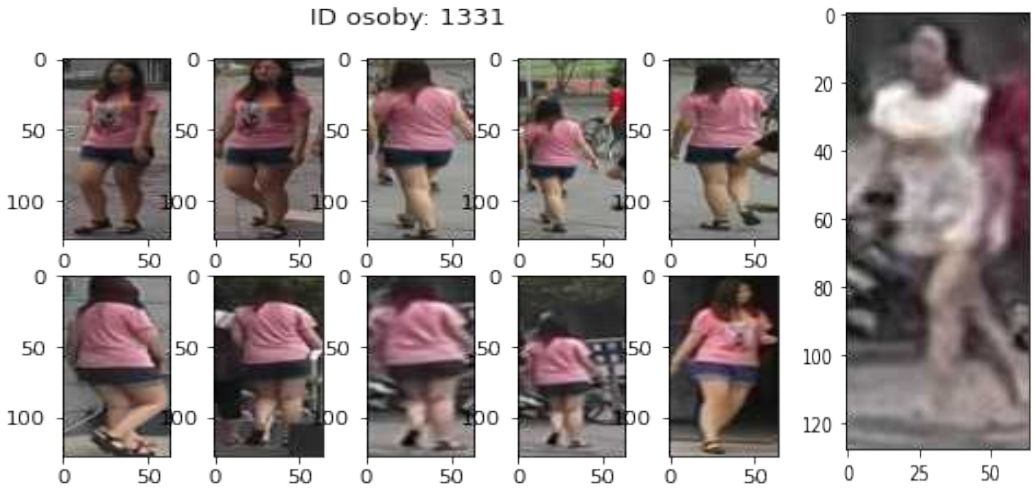
4.1 Market1501

To jeden z najczęściej wykorzystywanych zbiór danych do trenowania systemów reidentyfikacji osób. Jak twierdzą twórcy: "To nowy wysokiej jakości dataset do reidetnyfikacji osób nazwanym 'Market-1501'. Ogólenie dotychczasowe zbory:

- są ograniczone w skalowaniu
- posiadają ręcznie zaznaczone obramowania, które nie są dostępne w realnych ustawieniach
- posiadają jedynie jeden obraz wzorcowy dla każdej z osób z zestawu instancji do wyszukiwania

W celu wyeliminowania tych problemów zbór Market-1501 posiada trzy cechy. Pierwsza z nich to 32 000 zidentyfikowanych obramowań, ponad to ponad 500 tysięcy obrazów, które tworzą największy zbiór osób do trenowania modeli z zakresu zadań reidentyfikacji. Druga to stworzenie obrazów został stworzony przez 'Deformable Part Model' (DPM) jako

detektor pieszych. Trzecia to stowrzenie zbioru w otwartym systemie dzięki któremu, któremu każda osoba zawiera wiele ujęć z wielu kamer."
Przykład ze zbioru Markets1501 zaprezentowano poniżej:



Rys 1: Przykład danych ze zbioru Markets1501

4.2 Duke MTMC

"Duke MTMC (Multi-Target, Multi-Camera) to zbiór danych z nagrań wideo z monitoringu zrobionych na kampusie Duke University w 2014 roku i jest używany do badań i rozwoju systemów śledzenia wideo, ponownej identyfikacji osób i rozpoznawania twarzy o niskiej rozdzielczości.
Zestaw danych zawiera ponad 14 godzin zsynchronizowanego obrazu wideo z 8 kamer przy 1080p i 60 FPS, z ponad 2 milionami klatek 2000 uczniów idących do i z zajęć. Osiem kamer monitorujących rozmieszczonych na terenie kampusu zostało specjalnie skonfigurowanych do rejestrowania uczniów „w okresach między wykładami, kiedy ruch pieszy jest duży” to opis zbioru z pracy [10] Duke MTMC dataset

4.3 Własny zbiór danych

W celu weryfikacji przydatności modeli do wykorzystania w zadaniu śledzenia osób na nagraniach wideo, stworzono własny zbiór danych. Zbór ten pełnił rolę zbioru porównawczego do oceny jakość tworzenia embedingów z nagrań kamery 360. Posłużono się modelem kamery powszechnie stosowanym w punktach sprzedaży lokalizowanych w centrach handlowych oraz samodzielnych salonach obsługi klientów.
Zbiór wygenrowano z nagrania

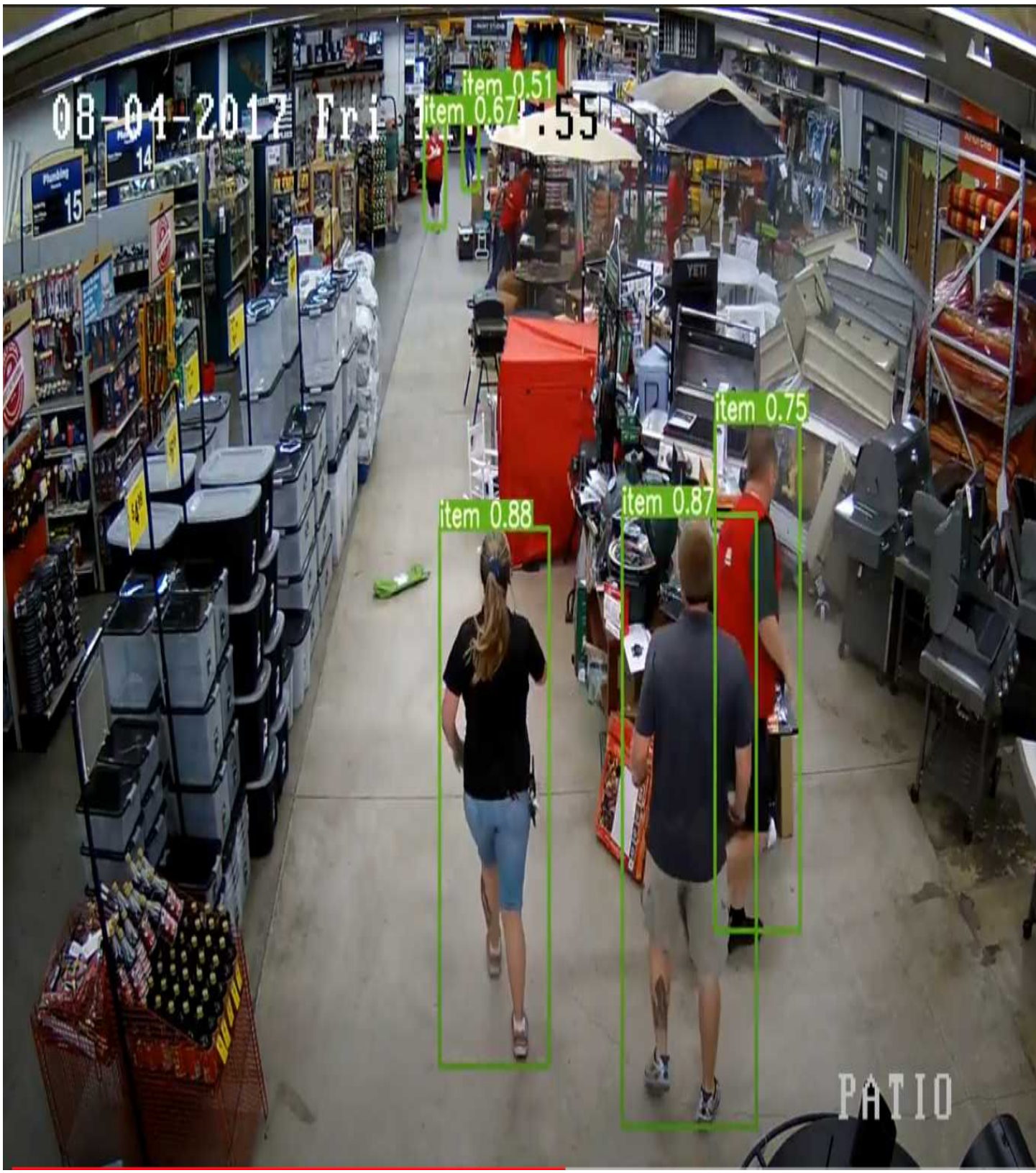
Rys 2. Przykład ujęcia z autorskiego nagrania kamerą 360
Link do autorskiego nagrania

Do wyodrębnienia obiektów na nagraniu wykorzystano framework YOLO5 [11] z obektów wyekstrachowanych z nagrania wybrano jedną postać ludzką. Z nagrania o długości **1min26s** uzyskano **1203** wycięte fragmenty z tą samą postacią w różnych pozach. Przykładowe wycięte obrazy:

Rys 3. Przykłady wyfiltrowanych obiektów z nagrania kamerą 360

4.4 Zbiór wygenerowany z nagrania pobranego z YouTube

Zbiór wygenerowano z nagrania umieszczonego poniżej.



Rys 5. Przykład ujęcia z nagrania pobranego z portalu YouTube

Link do nagrania

Korzystając z frameworku YOLO5 zmodyfikowanym na potrzeby tej pracy, jedynie do detekcji postaci ludzkich, wygenerowano dataset z wyciętymi osobami z nagrania. Posłużą one do stworzenia embedingów i wyszukania tej samej osoby z kolejnych klatek nagrania.

4.6 Podsumowanie wyboru dataset-u

W tej pracy wykorzystany zostanie jedynie zewnętrzny dataset Maret1501 oraz własny dataset porównawczy. Zdecydowano o użyciu jednego zbioru danych z powodu ograniczeń sprzętowych oraz by w procesie porównawczym wyeliminować wpływ doboru danych uczących, które przy wielu zbiorach danych w połaczeniu z ograniczoną ilością epoko uczenia znacząco wpływałaby na ostateczny wynik modeli.

Szczegółwy opis zbioru danych znajduje się w pracy [7] Market1501 - pdf. Analiz zbioru danych zawarta jest w notebooku ...

Własny dataset zostanie wykorzystany jedynie w celu wyciągnięcia wniosków o jakości modeli w dwóch wybranych kryteriach:

- zbierzności embedinów dla datasetu złożonego z jednej postaci
- czasu przetwarzania.

Parameter czasu przetwarzania jest szczególnie istotny pod względem wykorzystania w systemie śledzenia i identyfikacji rozpoznanych sylwetek ludzkich