

uzupełnić słowa kluczowe

UNIwersytet Gdański
Wydział Matematyki, Fizyki i Informatyki

Dariusz Adrychowski

nr albumu: 251682

Wizualna inwentaryzacja znaków drogowych

Uzupełnić tytuł po uzgodnieniu

Praca magisterska na kierunku:

INFORMATYKA

Promotor:

dr Jakub Neumann

Gdańsk 2018/2019

Streszczenie

Dodać streszczenie

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Słowa kluczowe

słowa kluczowe

Spis treści

Wprowadzenie	9
1. Bazowe definicje	11
1.1. Sieci neuronowe i algorytmy	11
2. Istniejące rozwiązania i historia rozwoju klasyfikacji wideo	13
2.1. Rozwiązania i narzędzia	16
2.1.1. Algorytm Viola—Jones	16
2.1.2. Dalal i Triggs - HOG	16
2.1.3. CNN	16
2.1.4. R—CNN	16
2.1.5. YOLOv3	16
2.1.6. SSD	17
2.1.7. Capsule Networks	17
2.1.8. Tensorflow	17
2.1.9. Detectron	17
2.1.10. Darknet	17
2.1.11. Mask RCNN	17
2.1.12. OpenVINO toolkit	17
2.1.13. OpenCV	17
2.1.14. Python	17
2.1.15. OpenStreet Map	17
2.1.16. GIS Server	17
2.2. Istniejące rozwiązania	17
2.2.1. Historia	17
2.2.2. NanoNets[5]	17
3. Pre—procesing danych	19
3.1. Pozyskanie danych	19
3.2. Klasyfikacja	20

3.3. Uczenie sieci i weryfikacja rezultatów	20
4. Realizacja	21
4.1. Pozyskanie danych wideo	21
4.2. Rozpoznawanie obiektów	21
4.2.1. Porównanie skuteczności algorytmów	21
4.2.2. Wybór środowiska	21
4.2.3. Wybór hardware	21
4.2.4. Wybór software	21
4.2.5. Opis realizacji softwarowej	21
5. Analiza wyników	23
5.1. Dokładność działania algorytmu	23
5.2. Możliwe udoskonalenia	23
5.3. Wnioski końcowe	23
6. Zakończenie	25
6.1. Ciąg dalszy — perspektywy rozwoju	25
6.2. Co dalej...	25
7. Brudnopis — do USUNIECIA w pracy finalnej	27
7.1. Wykrywanie obiektów OpenCV[6]	27
7.2. Umieszczenie obiektów na mapie	27
A. Tytuł załącznika jeden	29
B. Tytuł załącznika dwa	31
Bibliografia	33
Spis tabel	35
Spis rysunków	37
Oświadczenie	39

Todo list

■ uzupełnić słowa kluczowe	1
■ Uzupełnić tytuł po uzgodnieniu	1
■ Dodać streszczenie	3
■ dodać źródło: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive	
■ Biblioteka i narzędzia do segmentacji obrazu	17
■ https://www.mobileye.com/	17
■ Dokonczyc rownania	19
■ pozyskanie danych do nauki sieci, oprogramowanie do klasyfikacji	20
■ Jak uczymy, jaki zbiór danych, jaka sieć, opis testów i dokładności	20

Wprowadzenie

Celem niniejszej pracy było stworzenie oprogramowania umożliwiającego automatyczną akwizycję obrazu, jego segmentację i wykrywanie zadanych wcześniej, zdefiniowanych obiektów — znaków drogowych. Następnie korzystając z informacji o ich geolokalizacji, określenie ich położenia i jeśli okaże się to możliwe umieszczenia na mapie. Całość oprogramowania z założenia będzie realizowana przy użyciu kamery i wytrenowanej sieci neuronowej, która wykona detekcję i lokalizację na podstawie danych z kamery. Z założenia rozwiązanie nie musi działać online, dane geolokalizacji umieszczone będą wewnątrz zapisu wideo. Przykładowym zastosowaniem niniejszego oprogramowania może być automatyczna inwentaryzacja znaków drogowych.

ROZDZIAŁ 1

Bazowe definicje

Poniższy rozdział zawiera opis podstawowych pojęć używanych w niniejszej pracy — ich definicję, skróconą historię i krótki opis.

1.1. Sieci neuronowe i algorytmy

ROZDZIAŁ 2

Istniejące rozwiązania i historia rozwoju klasyfikacji wideo

Poniższy rozdział zawiera opis istniejących rozwiązań komercyjnych i niekomercyjnych, z którymi autor miał okazję zapoznać. Dodatkowo dołączono krótką historię rozwoju i opisano sposób wykorzystania ich w trakcie opracowywania rozwiązania będącego tematem niniejszej pracy.

„Deep Learning” koncentruje się na czterech podstawowych domenach — klasyfikacja obrazów, rozpoznawanie mowy, semantyczna klasyfikacja tekstów i rozpoznawanie/ klasyfikacji video. Z punktu widzenia formalnego video jest tylko sekwencją obrazów, co zostało dobrze udokumentowane i przedstawione chociażby przez Andej’a Karpathy — obecnie [styczeń 2019] zatrudnionego jako Director of AI at Tesla. Jego praca „Large-scale Video Classification with Convolutional Neural Networks”

dodac zdrojlo: <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42455.pdf>

opisuje sposób klasyfikacji obrazu wideo podobnie do klasyfikacji CNN¹ dla modelowego obrazu. Praca ta jest jakby punktem przełomowym w dziedzinie przetwarzania obrazu — wcześniejsze rozwiązania bazowały na opisywaniu sklasyfikowanego obrazu zestawem słów go reprezentującym² i decydowaniu przy użyciu algorytmu k-means oraz SVM³. Artykuł przedstawia podstawy integracji wszystkich wcześniejszych technik w jeden model CNN. Podejście do analizy obrazu wideo poprzez podzielenie jej na trzy komponenty miało wpływ na wszystkie późniejsze algorytmy przetwarzania obrazu:

¹CNN — Convolutional Neural Networks

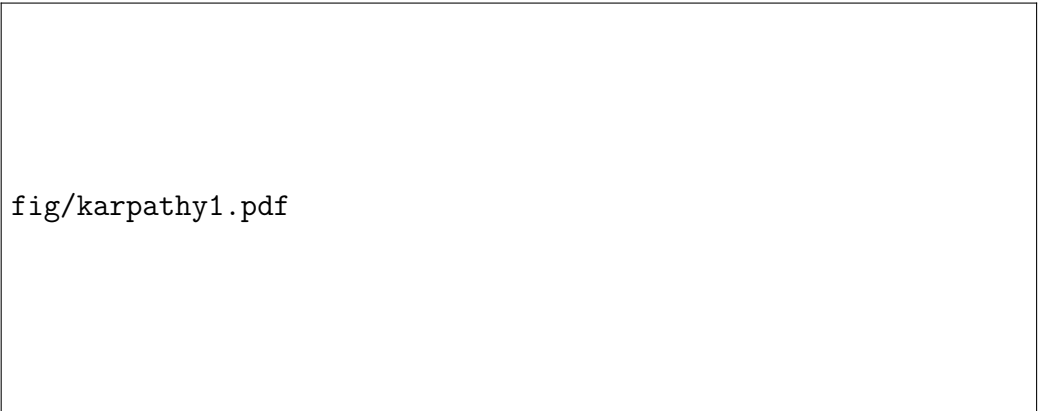
²bag of words

³SVM — Support Vector Machines

- połączenie elementów w dziedzinie czasu (interakcje, zmiany i przekształcenia)
- adaptatywny rozmiar analizowanego regionu
- transfer learning

Aby zmniejszyć ilość potrzebnej pamięci do budowania modelu, Karpathy ze swoim zespołem jednocześnie przetwarzał tylko jeden film, dodatkowo wstępna obróbka była przeprowadzana na różnych maszynach. Aby zapewnić jednakową długość danych wejściowych klipy pochodzące z YouTube podzielono na półsekundowe sekwencje. Agregacja przewidywanych półsekundowych klipów przypomina testowanie klasyfikacji w dziedzinie czasu - w przypadku obrazu klatka z początku klipu nawet przedstawiając ten sam obraz po pół sekundzie będzie przedstawiała ten sam obraz po pewnej transformacji. Obiekt na obu klatkach może zostać odkształcony w wyniku przemieszczenia się czy to samego obiektu czy też kamery (przekształcenia izomorficzne), jak również mogą ulec zmianie warunki w których wykonywane jest nagranie (wiatr, deszcz, nasłonecznienie itp.). Karparhy et al. przedstawiają również nowe podejście do dziedziny czasu (w momencie publikacji) — aplikują zasady opisujące sieci konwulsyjne do dziedziny czasu i zależności czasowych w wideo. Grupa ramek jest składana razem (grupowana) i staje się obrazem wejściowym do sieci CNN Standardowa sieć CNN jako wejście otrzymuje macierz danych — wysokość, szerokość i kolor (wartości w każdym z trzech składowych). Karpathy i jego zespół rozszerzyli parametry wejściowe i wcześniejszą ramkę umieścili na obecnej — zachowując rozmiary przekazu podwójną informację o kolorze — dla obu ramek osobno. Ze względu na scalanie ramek, Karpathy et al. zaproponowali różne strategie realizacji przy założeniu klasyfikowaniu jednej ramki (kombinowanej) jednocześnie. W zależności od tego, które ramki są składane, zespół osiągnął inne rezultaty. Poniższy rysunek przedstawia przyjęte strategie.

Model Single Frame z rysunku 2.1 reprezentuje wcześniejszy sposób przetwarzania/ klasyfikacji obrazu wideo — pojedyncza ramka jest interpretowana jak statyczny obraz, informacja temporalna nie jest uwzględniana.



fig/karpathy1.pdf

Rysunek 2.1. Strategie przetwarzania ramek zaproponowane przez Karpathy et al. źródło: "Large-scale Video Classification with Convolutional Neural Networks"


Late Fusion grupuje i scala ramki skraje z zadanego przedziału czasu — na ostatnią ramkę nakłada się ramkę pierwszą.

Early Fusion grupuje kilka ramek z ciągłego przedziału czasu.

Slow Fusion jest najbardziej skomplikowanym modelem — cztery częściowo nakładające się na siebie grupy ramek (overlapping równy jest dwie ramki) wybrane z ciągłego przedziału czasu są przekształcane przez sieć.

W wyniku eksperyment stwierdzono, że Slow Fusion dostarcza najbardziej użyteczną informację o obrazie, jednak jakość ta nie była dużo większa od jakości informacji otrzymanej z modelu Single Frame. Najlepsze wyniki dawało uśrednienie informacji pochodzącej ze wszystkich rozpatrywanych modeli (Single + Early + Late + Slow). W omawianej pracy przedstawiono również koncept przetwarzania obrazu sieci CNN wielu rozdzielczości.

Zasada działania przedstawiona na rysunku 2.2 — dwa odrębne wejścia przetwarzają obraz przez odrębne sieci.



fig/karpathy1.pdf

Rysunek 2.2. Strategia klasyfikowania obrazu przez sieci CNN o wielu rozdzielczościach źródło: "Large-scale Video Classification with Convolutional Neural Networks"

2.1. Rozwiązania i narzędzia

2.1.1. Algorytm Viola—Jones

2.1.2. Dalal i Triggs - HOG

HOG — Histograms of Oriented Gradients (2005?)/todo[color=yellow]Sprawdzić, źródło

2.1.3. CNN

Deep Learning 2012

2.1.4. R—CNN

2.1.5. YOLOv3

YOLO - You Only Look Once [7]

2.1.6. SSD

2.1.7. Capsule Networks

2.1.8. Tensorflow

2.1.9. Detectron

2.1.10. Darknet

2.1.11. Mask RCNN

2.1.12. OpenVINO toolkit

2.1.13. OpenCV

Biblioteka i narzędzia do segmentacji obrazu

2.1.14. Python

2.1.15. OpenStreet Map

2.1.16. GIS Server

2.2. Istniejące rozwiązania

2.2.1. Historia

Historycznie rzecz biorąc pierwszym systemem rozpoznającym znaki drogowe był produkt powstały przy współpracy MobileEye

<https://www.mobileye.com/>

z Continental AG na potrzeby firmy BMW — system rozpoznawania znaków drogowych dla BMW serii 7. System rozpoznawania znaków od tego czasu

2.2.2. NanoNets[5]

ROZDZIAŁ 3

Pre—procesing danych

W poniższym rozdziale omówiony zostanie proces pozyskania danych, uczenia sieci neuronowej i ogólny zarys procesu powstawania.

3.1. Pozyskanie danych

Podstawowym źródłem danych dla niniejszej pracy jest kamera wideo z możliwością zakodowania pozycji geograficznej, w której znajdowała się kamera w momencie filmowania. Kamera GoPRO Hero[1] od wersji 5 poza obrazem video rejestruje dane m.in. bieżącej prędkości, lokalizacji, kierunku ruchu, przeciążenia itp. Dane te zapisywane są wraz z obrazem w strumieniu video, w formacie mp4. Korzystając z zewnętrznych narzędzi[2][3] możliwa jest ekstrakcja tych danych do jednego z otwartych formatów danych.

Obraz video kodowany jest przy użyciu kodeka w standardzie MPEG-4 Part 14 i zapisywany jest z rozszerzeniem mp4.

Po uzyskaniu dwóch ujęć tego samego obiektu, znając dokładną lokalizację miejsca skąd wykonano oba ujęcia oraz kierunek, w którym należałoby się udać z obu miejsc aby dotrzeć do rozpoznanego obiektu (a ang: bearings) można obliczyć położenie obiektu. Jak podaje źródło[4]

Formuła— odległość kątowa pomiędzy p1 — p2

$$\delta_{12} = 2 \arcsin\left(\sqrt{\left(\sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos \varphi_1 * \cos \varphi_2 * \sin^2\left(\frac{\Delta\lambda}{2}\right)\right)}\right)$$

$$\theta_a = \arccos\left(\frac{(\sin\varphi_2 - \sin\varphi_1 * \cos \delta_{12})}{(\sin\theta_{12} * \cos \varphi_1)}\right)$$

$$\theta_b = \arccos\left(\frac{(\sin\varphi_1 - \sin\varphi_2 * \cos \delta_{12})}{(\sin\theta_{12} * \cos \varphi_2)}\right)$$

Dokonczyc rownania

3.2. Klasyfikacja

pozyskanie danych do nauki sieci, oprogramowanie do klasyfikacji

3.3. Uczenie sieci i weryfikacja rezultatów

Jak uczymy, jaki zbiór danych, jaka sieć, opis testów i dokładności

ROZDZIAŁ 4

Realizacja

4.1. Pozyskanie danych wideo

4.2. Rozpoznawanie obiektów

4.2.1. Porównanie skuteczności algorytmów

4.2.2. Wybór środowiska

4.2.3. Wybór hardware

4.2.4. Wybór software

4.2.5. Opis realizacji softwarowej

ROZDZIAŁ 5

Analiza wyników

5.1. Dokładność działania algorytmu

5.2. Możliwe udoskonalenia

5.3. Wnioski końcowe

ROZDZIAŁ 6

Zakończenie

6.1. Ciąg dalszy — perspektywy rozwoju

6.2. Co dalej...

ROZDZIAŁ 7

Brudnopis — do USUNIECIA w pracy finalnej

7.1. Wykrywanie obiektów OpenCV[6]

7.2. Umieszczenie obiektów na mapie

DODATEK A

Tytuł załącznika jeden

Treść załącznika jeden.

DODATEK B

Tytuł załącznika dwa

Treść załącznika dwa.

Bibliografia

- [1] . [https://community.gopro.com/t5/
GoPro-Metadata-Visualization/Extracting-the-metadata-in-a-useful-format/
gpm-p/40293](https://community.gopro.com/t5/GoPro-Metadata-Visualization/Extracting-the-metadata-in-a-useful-format/gpm-p/40293). [Online; dostęp 22 styczeń 2019].
- [2] . <https://github.com/KonradIT/gopro-utils>. [Online; dostęp 22
styczeń 2019].
- [3] . <https://github.com/JuanIrache/gopro-utils>. [Online; dostęp 22
styczeń 2019].
- [4] . <http://www.movable-type.co.uk/scripts/latlong.html>. [Online;
dostęp 22 styczeń 2019].
- [5] How to easily do Object Detection on Drone Image-
ry using Deep learning. [https://medium.com/nanonets/
how-we-flew-a-drone-to-monitor-construction-projects-in-africa-using-deep-learn-](https://medium.com/nanonets/how-we-flew-a-drone-to-monitor-construction-projects-in-africa-using-deep-learning-786e0c0f3000)
2018. [Online; dostęp 11 styczeń 2019].
- [6] V. Gupta. Face Detection – OpenCV, Dlib and
Deep Learning. [https://www.learnopencv.com/
face-detection-opencv-dlib-and-deep-learning-c-python,](https://www.learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/) Octo-
ber 22nd, 2018. [Online; dostęp 11 styczeń 2019].
- [7] S. Darkflow. <https://github.com/thtrieu/darkflow>. [Online; dostęp
6 styczeń 2019].

Spis tabel

Spis rysunków

2.1. Strategie przetwarzania ramek zaproponowane przez Karpa- thy et al. źródło: "Large-scale Video Classification with Co- nvolutional Neural Networks"	15
2.2. Strategia klasyfikowania obrazu przez sieci CNN o wielu roz- dzielczościach źródło: "Large-scale Video Classification with Convolutional Neural Networks"	16

Oświadczenie

Ja, niżej podpisany(a) oświadczam, iż przedłożona praca dyplomowa została wykonana przeze mnie samodzielnie, nie narusza praw autorskich, interesów prawnych i materialnych innych osób.

.....

data

.....

podpis