

---

# **ANALIZA OBRAZÓW I WIDEO - projekt**

## **Zliczanie pojazdów na podstawie obrazu z kamery umieszczonej nad jezdnią**

---

**Marcel Cielinski**

Wydział Informatyki i Zarządzania  
Politechnika Wrocławskiego  
236747@student.pwr.edu.pl

**Weronika Pawlak**

Wydział Informatyki i Zarządzania  
Politechnika Wrocławskiego  
236740@student.pwr.edu.pl

### **1 Opis problemu**

Podjęte zagadnienie dotyczy zliczania pojazdów na podstawie obrazu z kamery umieszczonej nad jezdnią. Porównane zostało działanie rozwiązania neuronowego i bezneuronowego. Natomiast problematykę zadania podzielono na dwa podproblemy - detekcja i zliczanie. Detekcja może opierać się o rozpoznanie konkretnych obiektów (rozwiązanie neuronowe) lub wykrycie ruchu (rozwiązanie bezneuronowe). Problem zliczania polega na odpowiednim sumowaniu wykrytych obiektów. Jego zasadniczą trudnością jest stwierdzenie, czy rozpatrywana detekcja została już wcześniej zliczona i tym samym, czy należy dodać ją do sumy.

### **2 Przyjęte założenia i ograniczenia**

Zostały przyjęte następujące założenia upraszczające:

- Kamera pozostaje nieruchoma - dopuszczalne są niewielkie drgania, wynikające np. z wiatrów atmosferycznych. Możliwe jest również wykorzystanie nagrani, w których występują stosunkowo niewielkie problemy ze stabilnością, np. wykonanych telefonem komórkowym bez dodatkowego osprzętu.
- Przetwarzany jest cały strumień nagrania wideo (dotyczy metody tradycyjnej) - w metodzie tradycyjnej wykonywany jest post-processing, mający na celu wyeliminowanie z detekcji części szumów. Metoda ta nie jest zatem przystosowana do odczytu w czasie rzeczywistym (w trakcie przetwarzania).
- Wyznaczana jest strefa detekcji - obszar, w którym pojazdy są zliczane uzależniony jest od nagrania i obiektu zainteresowań. W przypadku wielu nagrani kadr obejmuje częściowo pas ruchu w przeciwnym kierunku. Jeśli celem jest zliczenie pojazdów poruszających się tylko jednym pasem ruchu, konieczne jest odpowiednie obcięcie strefy detekcji.
- Brak klasyfikowania obiektów detekcji jako pojazdy (dotyczy metody tradycyjnej) - metoda bezneuronowa bazuje na detekcji ruchu na przestrzeni kolejnych klatek, wykrycie pojazdów jest uzależnione od doboru parametru dotyczącego minimalnej wielkości konturów.

### **3 Założony zakres rozwiązania**

Założona została implementacja następujących składowych:

- Podejście neuronowe - wykorzystanie sieci YOLOv3 w celu detekcji obiektów.
- Podejście tradycyjne - wykorzystanie metody różnicy klatek (ang. frame differencing), w celu detekcji poruszających się obiektów.
- Zliczanie - moduł zliczania wspólny dla obu rozwiązań.
- Ulepszenia działania rozwiązań na nagraniach w trudniejszych warunkach np. po zmroku.

### **4 Opis przyjętej metody**

#### **4.1 Rozwiązanie neuronowe**

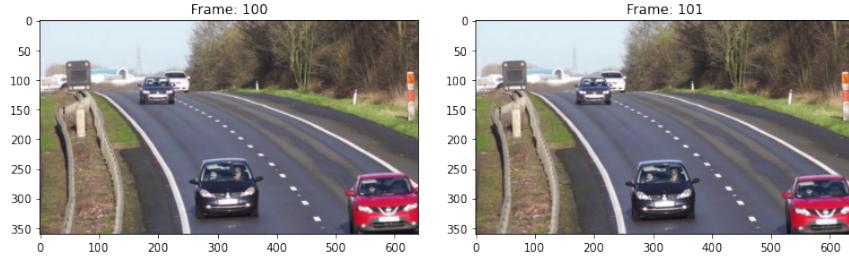
W rozwiąaniu neuronowym można wyróżnić następujące kroki:

- Wyznaczenie obszaru detekcji i ewentualne dostrojenie parametrów. Domyślna wielkość obszaru detekcji wynosi 40% wysokości obrazu. Istnieje również możliwość manipulowania parametrem progu pewności sieci, od którego zależy nadanie etykiety. Domyślana wartość 0.5 (prawdopodobieństwo) sprawdza się dobrze.
- Ekstrakcja klatek wejściowego strumienia wideo.
- Wykorzystanie algorytmu YOLOv3 w celu detekcji i klasyfikacji obiektów.
- Filtrowanie obiektów - brane są pod uwagę tylko obiekty konkretnych klas, odzwierciedlające pojazdy. Zbiór etykiet klasyfikowanych jako pojazd: [bicycle, car, motorbike, bus, truck, train].
- Zliczanie obiektów - algorytm opisany w sekcji 4.3.
- Złączenie przetworzonych klatek do wyjściowego strumienia wideo.

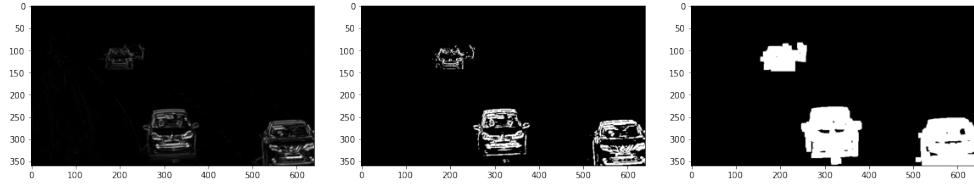
#### **4.2 Rozwiązanie bezneuronowe**

W rozwiąaniu bezneuronowym można wyróżnić następujące kroki:

- Wyznaczenie obszaru detekcji i dostrojenie parametrów. Domyślna wielkość obszaru detekcji wynosi 40% wysokości obrazu. Ponadto należy określić graniczną wartość wielkości konturu, powyżej której obiekt jest klasyfikowany jako pojazd. Jest to konieczne, ponieważ ustawienie tego parametru zależy od poziomu oddalenia kamery.
- Ekstrakcja klatek wejściowego strumienia wideo.
- Konwersja klatek do skali szarości.
- Różnicowanie klatek - kolejne pary.
- Binaryzacja różnicowanych klatek - progowanie.
- Dylatacja różnicowanych klatek.
- Odnajdywanie konturów obiektów znajdujących się w polu detekcji oraz wygenerowanie na ich podstawie prostokątnych ramek ograniczających (ang. bounding box).
- Zliczanie obiektów - algorytm opisany w sekcji 4.3.
- Walidacja i filtrowanie detekcji - post-processing, w obrębie którego wykluczone zostają detekcje, które z dużą dozą pewności nie są pojazdem (np. pojawiające się jedynie na 2 klatkach szумy).
- Złączenie przetworzonych klatek do wyjściowego strumienia wideo.



Rysunek 1: Dwie kolejne klatki z nagrania wideo



Rysunek 2: Klatki z Rysunku 1 - różnicowanie - binaryzacja - dylatacja

### 4.3 Algorytm zliczania

Zasadniczą częścią algorytmu zliczalnia jest stwierdzenie, czy wykryty obiekt jest nowy. Odbywa się to w następujący sposób:

---

#### **Algorithm 1:** Is new detection

---

**Input :** *detections\_in\_previous\_frames*  
*current\_box*  
*number\_of\_frames\_to\_check*  $\leftarrow$  *FPS/4*

**Output:** *True OR False*

*distance*  $\leftarrow \infty$

**for** *i*  $\leftarrow 0$  **to** *number\_of\_frames\_to\_check* **do**

*tempdist, index* =  
**SHORTEST\_DISTANCE\_TO\_DETECTIONS**(*detections\_in\_previous\_frames*[*i*],  
*current\_box*)

**if** *tempdist* < *distance* **then**  
 | *distance* = *tempdist*

**if** *dist* > *max(current\_box[width], current\_box[height]) / 2* **then**  
 | **return** *True*

**else**  
 | *current\_box[index]*  $\leftarrow$  *index*  
 | **return** *False*

---

Wyżej opisana metoda jest wykonywana na obiektach znajdujących się w polu detekcji. Jeśli obiekt zostanie oznaczony jako już istniejący w liście detekcji wcześniejszych klatek, zostaje mu przywrócony odpowiedni identyfikator. W przeciwnym przypadku, przyporządkowywany jest nowy indeks oraz doliczany jest do aktualnej sumy wszystkich wykrytych obiektów.

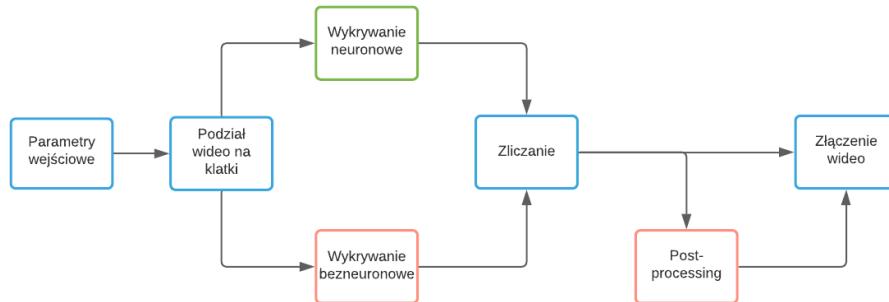
## 5 Wykorzystane komponenty obce

W celu zaimplementowania rozwiązania wykorzystane zostały następujące komponenty:

- Przetrenowana na zbiorze *COCO* sieć *YOLOv3* wykorzystana do detekcji obiektów:
  - zbiór klas coco.names zawierający 80 etykiet
  - plik konfiguracyjny yolov3.cfg
  - przetrenowane wagi yolov3.weights
- Biblioteka *opencv-python*:
  - możliwości dotyczące przetwarzania wideo (dzielenie na klatki, zapis wyjściowego strumienia, pobieranie informacji itp.)
  - możliwości wspomagające implementację metod tradycyjnych takie jak: różnicowanie, binaryzacja, dylatacja, odnajdywanie konturów
  - możliwości dotyczące rysowania (oznaczenia wykrytych obiektów i ich punktów centralnych)

## 6 Architektura rozwiązania

Na poniższym schemacie przedstawiono model rozwiązania. Kolorem niebieskim oznaczono elementy wspólne, kolorem zielonym elementy rozwiązania neuronowego, kolorem pomarańczowym elementy rozwiązania bezneuronowego.



Rysunek 3: Model rozwiązania

## 7 Wkład własny

Do wkładu własnego można zaliczyć:

- implementację rozwiązania tradycyjnego
- implementację redukcji szumów w rozwiąaniu tradycyjnym
- implementację mechanizmu zliczania
- analizę porównawczą obu metod

## 8 Opis zgromadzonych danych

Dane zostały pozyskane ze źródeł internetowych. Dobór nagrani do badań miał na celu przetestowanie działania obu rozwiązań na różnorodnych przypadkach. Wybrane przykłady różnią się kątem nachylenia kamery, warunkami, charakterem drogi (miasto, autostrada).

### 8.1 Nagrania podstawowe

Pierwsza grupa nagrani posłużyła do sprawdzenia poprawności implementacji. Zawiera proste, dobrej jakości nagrania, głównie na trasach szybkiego ruchu.



Rysunek 4: Przykładowe klatki z prostych nagrań

### 8.2 Nagrania ze wzmożonym natężeniem ruchu

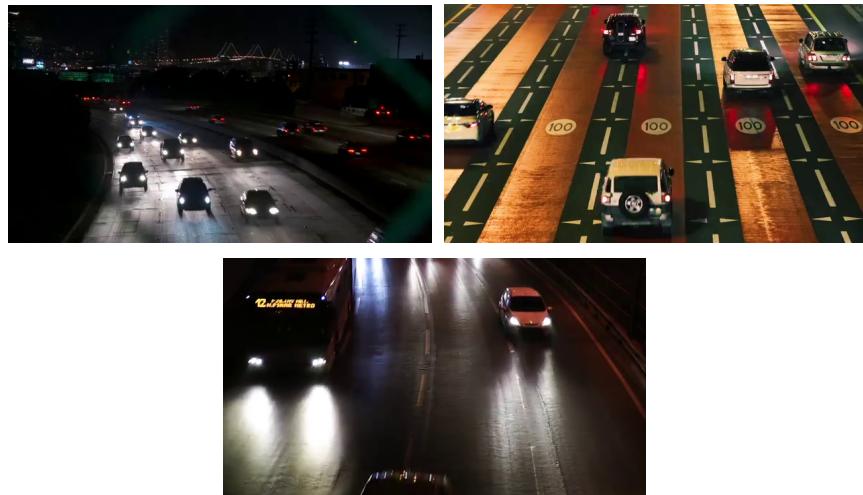
Przetestowane zostały również trudniejsze przypadki, ruchu w mieście, na których samochody poruszają się znacznie wolniej, pojawia się wiele jednośladów, występują liczne szумy.



Rysunek 5: Przykładowe klatki z trudniejszych nagrań

### 8.3 Nagrania nocne

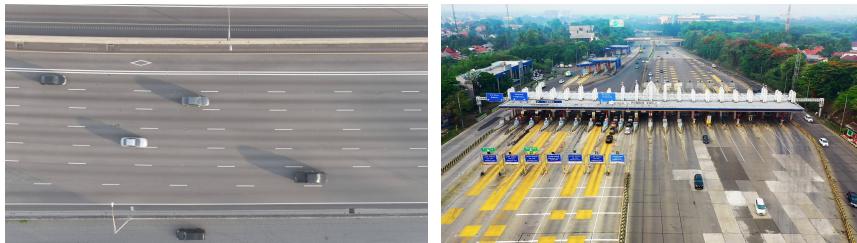
Działanie obu metod sprawdzone zostało także na nagraniach nocnych, które są znacznie trudniejsze, ponieważ samochody nie są tak widoczne, pojawiają się cienie, światła itp.



Rysunek 6: Przykładowe klatki z nagrani nocnych

#### 8.4 Nagrania z góry

Jako kolejną grupę wyodrębniono nagrania z góry. Ta grupa charakteryzuje się znaczącą rozbieżnością wyników testowanych metod.



Rysunek 7: Przykładowe klatki z nagrani z góry

#### 8.5 Charakterystyka nagrani wideo

Wideo	Typ	Długość
1	podstawowy	46 s
2	podstawowy	60 s
3	podstawowy	21 s
4	podstawowy	22 s
5	podstawowy	21 s
6	nocny	14 s
7	nocny	36 s
8	nocny	60 s
9	duży ruch	27 s
10	duży ruch + zmrok	37 s
11	widok z góry	40 s
12	widok z góry	22 s

Tabela 1: Charakter i długość nagrani wideo wykorzystanych do eksperymentów

## 9 Wynik eksperymentalne

### 9.1 Cel badań

Celem badań było porównanie zaimplementowanych metod na nagraniach różnego typu oraz wykrycie mocnych i słabych stron każdego z rozwiązań. Ponadto, chodziło o zweryfikowanie, czy oprogramowane narzędzia sprawdzą się w rzeczywistych warunkach.

### 9.2 Przypadki szczególnie

Każdą z grup : 8.2, 8.3, 8.4 można traktować jako inny przypadek szczególny. Badania na trudniejszych dla algorytmów nagraniach, pozwoliły na wychwycenie błędów i wprowadzenie usprawnień.

### 9.3 Miary poprawności

W podjętym zagadnieniu, metrykę dokładności dla typowego użytkownika mogłyby stanowić miara oparta na porównaniu - [rzeczywista liczba pojazdów : liczba pojazdów zwrócona przez algorytm]. Należy jednak pamiętać, że bez dodatkowych informacji na temat popełnionych błędów, tego typu miara nie daje informacji o jakości rozwiązania. Można również zastosować standardowe miary wykorzystywane w uczeniu maszynowym jak np. dokładność (ang. accuracy).

### 9.4 Wyniki liczbowe

Wideo	Prawidłowy wynik	Podejście tradycyjne	Podejście neuronowe	Wielokrotne wykrycia PT:PN	Pominięte wykrycia PT:PN
1	37	37	37	0:0	0:0
2	33	31	33	0:0	2:0
3	10	9	12	1:2	2:0
4	19	18	19	0:0	1:0
5	21	22	23	1:2	0:0
6	24	25	20	4:0	3:4
7	23	25	25	2:2	0:0
8	97	97	118	3:21	3:0
9	23	26	32	6:9	3:0
10	101	107	76	14:~15	8:~40
11	35	35	47	0:18	0:6
12	61	68	2	11:0	4:59

Tabela 2: Tabela przedstawia rzeczywistą i predykowaną liczbę samochodów na nagraniach oraz pominięte i wielokrotne wykrycia dla każdej z metod (PT - podejście tradycyjne, PN - podejście neuronowe). Numeracja jest zgodna z Tabelą 1.

Grupa nagrań	Średnie accuracy podejście tradycyjne	Średnie accuracy podejście neuronowe
podstawowe	0.91	0.95
nocne	0.87	0.86
duży ruch	0.75	0.63
widok z góry	0.9	0.29
ogółem	0.87	0.76

Tabela 3: Tabela przedstawia średnią wartość accuracy dla każdej z grup

## 9.5 Wyniki - nagrania

Nagrania wejściowe oraz przetworzone przez oba algorytmy, można znaleźć <pod linkiem>. Nazwy filmów odpowiadają tym przedstawionym w tabelach. Ewentualne rozbieżności w wynikach końcowych na nagraniach i tych przedstawionych w raporcie wynikają z ręcznej walidacji autorów. Przykładowo na nagraniu 8. część pasa ruchu w przeciwnym kierunku nie została obcięta - zliczone w tym obszarze samochody zostały ręcznie odjęte od sumy.

# 10 Ocena końcowa

## 10.1 Podejście neuronowe

### 10.1.1 Wyniki zadowalające

Najlepsze wyniki uzyskano na prostych nagraniach w dzień. Sieć bardzo dobrze radziła sobie z wykrywaniem obiektów w kolejnych klatech, nie było zatem problemów ze zliczaniem. Algorytm stosunkowo dobrze sprawdził się również na nagraniach nocnych. W tym przypadku pojawiały się drobne błędy w zliczaniu wynikające z chwilowych zniknięć detekcji.

### 10.1.2 Wyniki niezadowalające

Sporo błędów pojawiało się w przypadku nagrąć ze wzmożonym ruchem. Błędy te dotyczyły zarówno pomijania jak i wielokrotnego zliczania pojazdów. Najgorzej jednak sieć YOLO poradziła sobie z nagraniami z góry, na których skuteczność była bardzo niska, w jednym przypadku niemalże zerowa.

### 10.1.3 Problemy - przyczyny i możliwe rozwiązania

Problemy w zliczaniu, wynikające ze znikających detekcji można wyeliminować poprzez ulepszenie lub zmianę algorytmu śledzenia obiektu. Niestety jest to kwestia dość skomplikowana, ponieważ często powtórna detekcja pojawia się w odstępie znaczącej liczby klatek. Efektywność algorytmu YOLO na widoku z góry dałoby się poprawić dotrenowując sieć na tym szczególnym przypadku. Prawdopodobnie w zbiorze uczącym, dla klas odzwierciedlających pojazdy, znalazło się mało przykładów widoku z góry. Sieć nie radzi sobie z klasyfikacją obiektów, nawet po zmniejszeniu wymaganego progu pewności.

## **10.2 Podejście bezneuronowe**

### **10.2.1 Wyniki zadowalające**

Podobnie jak w przypadku rozwiązania neuronowego bardzo dobre wyniki uzyskano na zbiorze prostych nagrani, na którym skuteczność była bardzo wysoka. Po redukcji szumów, podejście tradycyjne zaskakującym dobrze poradziło sobie również na nagraniach nocnych, które początkowo były bardzo problematyczne. Światła i cienie powodowały, że ciężko było jednoznacznie określić kontury obiektów. Po zastosowaniu post-processingu można uznać wyniki za zadowalające. W przeciwieństwie do sieci neuronowej, w podejściu różnicy klatek, widok z góry nie stanowi problemu. Metoda ta opiera się na detekcji ruchu, zatem nie ma konieczności rozpoznania obiektu. Można powiedzieć, że widok z góry jest dla tego podejścia nawet lepszy - wyeliminowane zostaje ryzyko wzajemnego przysłaniania się pojazdów.

### **10.2.2 Wyniki niezadowalające**

Niezbyt zadowalające wyniki osiągnięto na nagraniach ruchu w mieście, na których samochody poruszają się wolniej lub/i znajdują się blisko siebie. Sporo błędów wynikało ze zliczenia grupy obiektów poruszających się w nieznacznej odległości jako jeden obiekt.

### **10.2.3 Problemy - przyczyny i możliwe rozwiązania**

Przyczyną problemów jest prostota metody. Detekcja ruchu nie gwarantuje separowania obiektów. Jeśli pojazdy znajdują się bardzo blisko ich rozdzielenie i poprawne zliczenie przy użyciu metody różnicy klatek jest niemalże niemożliwe.

## **10.3 Wnioski końcowe**

W początkowej fazie eksperymentów wydawało się, że tradycyjne podejście do podjętego zagadnienia nie ma szans konkurować z podejściem neuronowym. Co ciekawe okazało się, że wcale tak nie jest. Jasnym jest, że metoda różnicy klatek ma swoje wady, ponieważ opiera się na bardzo prostych założeniach. Warto jednak zauważyć, że wyniki osiągnięte przy pomocy tego algorytmu są dość stabilne - nie ma przypadków skrajnych, w których algorytm okazał się zupełnie nieskuteczny. Taka sytuacja miała miejsce w przypadku podejścia neuronowego i nagrania z góry. Ponadto na tak prostych metodach, jak różnica klatek, łatwiej jest testować różnego rodzaju ulepszenia / zmiany, ponieważ nie są tak wymagające obliczeniowo jak w przypadku podejścia neuronowego. Łatwiej jest także zidentyfikować przyczynę błędów.