

Mariusz FRĄCKIEWICZ, Filip HABRYN, Henryk PALUS
Politechnika Śląska

ANONIMIZACJA DANYCH WIDEO NA POTRZEBY SYSTEMÓW ADAS

Streszczenie. W pracy analizowane są metody anonimizacji twarzy w nagraniach wykorzystywanych przez systemy ADAS (ang. *Advanced Driver Assistance System*) w kontekście przepisów RODO. Przetestowano 12 konfiguracji łączących cztery detektory twarzy i trzy techniki anonimizacji na publicznie dostępnym zbiorze danych PP4AV. Wyniki wykazały wysoką skuteczność detektorów YOLOv8-Face i YuNet oraz niewystarczającą ochronę zapewnianą przez metodę pikselizacji. Najlepszy kompromis między skutecznością detekcji, wydajnością anonimizacji i ochroną prywatności osiągnęła konfiguracja będąca połączeniem detektora YuNet z rozmyciem Gaussowskim.

VIDEO DATA ANONYMIZATION FOR ADAS SYSTEMS

Summary. This paper analyzes face anonymization methods for video recordings used in Advanced Driver Assistance Systems (ADAS) in the context of General Data Protection Regulation (GDPR) requirements. Twelve configurations combining four face detectors and three anonymization techniques were evaluated using the publicly available PP4AV dataset. The results demonstrated the high effectiveness of the YOLOv8-Face and YuNet detectors, while pixelation provided insufficient privacy protection. The best trade-off between detection performance, anonymization efficiency, and privacy preservation was achieved by the configuration combining the YuNet detector with Gaussian blurring.

1. Wprowadzenie

Rozwój sztucznej inteligencji i uczenia maszynowego znacząco przyspieszył rozwój systemów wspomagania kierowcy ADAS, które stanowią kluczowy krok w kierunku autonomicznej mobilności oraz poprawy bezpieczeństwa ruchu drogowego [4, 5]. Systemy te opierają się na analizie danych pochodzących z kamer, radarów i lidarów, a ich rozwój wymaga gromadzenia ogromnych zbiorów nagrań obejmujących miliony kilometrów przejazdów [6, 1].

Masowa akwizycja danych wideo rodzi istotne wyzwania związane z ochroną prywatności. Wizerunek twarzy stanowi daną biometryczną podlegającą szczególnej ochronie prawnej na mocy przepisów RODO [15], co wymusza stosowanie skutecznych metod anonimizacji. Rozwiązania te muszą jednocześnie zapewniać wysoką skuteczność ochrony prywatności, odpowiednią wydajność obliczeniową oraz możliwość wdrożenia w systemach o ograniczonych zasobach sprzętowych.

W niniejszym artykule dokonano analizy porównawczej metod anonimizacji twa-

rzy przeznaczonych potencjalnie do zastosowań motoryzacyjnych. Oceniono cztery detektory twarzy: Viola–Jonesa [22], MTCNN [25], YuNet [23] oraz YOLOv8-Face [10], a także trzy techniki anonimizacji: rozmycie gaussowskie, pikselizację i maskowanie. Badania przeprowadzono na zbiorze PP4AV [13], koncentrując się na kompromisie pomiędzy skutecznością detekcji, wydajnością obliczeniową oraz poziomem ochrony prywatności. Zaimplementowano szeroki zestaw metryk, obejmujący m.in. precyzję, czułość, średnią liczbę klatek na sekundę FPS (ang. *Frames Per Second*), wskaźnik podobieństwa strukturalnego (SSIM, ang. *Structural Similarity Index Measure*) oraz miarę prywatności Privacy-TPR opartą na modelu FaceNet [18]. Dzięki temu możliwe było wielowymiarowe spojrzenie na badane zjawisko i sformułowanie wniosków praktycznych. Uzupełnieniem badań ilościowych na zbiorze PP4AV była analiza jakościowa przeprowadzona na danych zewnętrznych, udostępnionych do badań przez partnera z przemysłu. Brak adnotacji w tych danych uniemożliwiał formalne obliczanie miar, jednak pozwolił na praktyczną ocenę proponowanych rozwiązań w kontekście realistycznych scen miejskich, istotnie różniących się oświetleniem, kontrastem i perspektywą nagrania.

2. Zbiory danych i benchmarki

Skuteczność detekcji twarzy, jak i późniejszej anonimizacji jest ściśle zależna od jakości zbiorów danych wykorzystywanych do uczenia i oceny algorytmów. Poniżej przedstawiono popularne w literaturze zbiory benchmarkowe, których kluczowe parametry zestawiono w tabeli 1.

Tabela 1

Wybrane zbiory danych stosowane w badaniach nad detekcją i anonimizacją twarzy

Zbiór	Rok	# twarzy / klatek	Kontekst	Rozdzielczość*	Licencja
WIDER FACE [24]	2016	~393 000	Zdjęcia uliczne; duża zmienność skali i okluzji	$\geq 640 \times 480$	Academic use only
LFW [7]	2008	13 233	Portrety celebrytów w warunkach niekontrolowanych	250×250	BSD
FaceForensics++ [17]	2019	~1 000	Wysokiej jakości wideo; detekcja <i>deep-fake</i>	720×1280	Non-commercial (academic only)
PP4AV [13]	2023	17 880	Kamery pojazdów (ADAS) – warunki dzienne/nocne	1920×1080	CC BY-NC-ND 4.0

*Rozdzielczość podano orientacyjnie; niektóre ujęcia mogą być skalowane.

2.1. Zbiory ogólnego przeznaczenia

WIDER FACE [24] stanowi de-facto standard w ocenie detektorów twarzy. Kolekcja obejmuje 32 203 obrazy z 61 kategorii zdarzeń, co gwarantuje wysoką różnorodność skali, pozy i okluzji. Z uwagi na szczegółowe adnotacje (ramki oraz pięć punktów charakterystycznych) zbiór ten jest powszechnie wykorzystywany do uczenia i walidacji modeli, takich jak MTCNN czy RetinaFace. Mimo, że nie prezentuje perspektywy typowej dla kamer samochodowych, pozostaje punktem odniesienia dla oceny jakości detektorów na podstawie miary średniej precyzji mAP (ang. *mean Average Precision*), co czyni go kluczowym benchmarkiem w badaniach porównawczych.

Labeled Faces in the Wild (LFW) [7] pełni rolę klasycznego benchmarku do weryfikacji tożsamości osób. Zbiór jest niewielki i stosunkowo prosty w porównaniu z nowszymi bazami, jednak dzięki swojej długiej historii pozwala na porównywanie wyników z szerokim spektrum wcześniejszych prac, co ułatwia analizę trendów. Bywa używany do badania wpływu anonimizacji na zdolność rozpoznawania twarzy przez algorytmy *ArcFace* [2] czy *FaceNet* [18].

2.2. Zbiory ukierunkowane na przetwarzanie wideo

FaceForensics++ [17] dostarcza zbiór wideo wraz z wersjami zmanipulowanymi (*deep-fake*). Choć głównym celem jest detekcja manipulacji, wysoka rozdzielczość oraz realistyczne ruchy kamery czynią zbiór wartościowym dla testów anonimizacji w warunkach dynamicznych. Miary takie jak klasyczny PSNR czy nowoczesny LPIPS [26], obliczane pomiędzy oryginałem a wariantami po anonimizacji, stanowią istotny punkt odniesienia dla oceny użyteczności wizualnej oraz jakości zachowanych szczegółów obrazu, co czyni ten zbiór wartościowym także poza pierwotnym kontekstem detekcji manipulacji.

PP4AV (ang. *Privacy-Preserving for Autonomous Vehicles*) [13] to pierwszy publicznie dostępny zbiór klatek z perspektywy pojazdu, zawierający jednocześnie adnotacje twarzy pieszych i tablic rejestracyjnych. Metadane obejmują m.in. kategorię oświetlenia (dzień/noc), warunki pogodowe oraz parametry trajektorii jazdy, co umożliwia testowanie algorytmów w ekstremalnych przypadkach ADAS (np. przy słabym oświetleniu, deszczu, dynamicznych zmianach prędkości). Zbiór jest udostępniany na licencji CC BY-NC-ND 4.0, co pozwala na jego swobodne wykorzystanie akademickie, lecz wyklucza zastosowania komercyjne.

3. Wybrane metody detekcji twarzy

Niezbędnym krokiem w procesie anonimizacji jest precyzyjna lokalizacja twarzy. Błąd typu fałszywie negatywnego (FN, ang. *False Negative*) skutkuje krytycznym naruszeniem prywatności. Poniżej omówiono ewolucję kluczowych architektur dla detekcji twarzy, od metod klasycznych po nowoczesne, zoptymalizowane sieci neuronowe.

3.1. Detektor twarzy Viola-Jonesa

Algorytm Viola-Jonesa jest jedną z pierwszych metod umożliwiających skuteczną detekcję twarzy w czasie rzeczywistym. Algorytm wykorzystuje cechy Haara obliczane z użyciem tzw. obrzu całkowego oraz kaskadę klasyfikatorów tworzonych za pomocą algorytmu AdaBoost. Zastosowanie struktury kaskadowej pozwala na szybkie odrzucanie obszarów niezawierających twarzy, co zapewnia wysoką wydajność obliczeniową. Mimo przełomowego znaczenia i szerokiego zastosowania we wczesnych systemach detekcji twarzy, metoda jest wrażliwa na zmiany oświetlenia oraz rotację twarzy.

3.2. MTCNN

Architektura MTCNN (ang. *Multi-task Cascaded Convolutional Network*) realizuje detekcję w sposób „od ogółu do szczegółu” przy pomocy kaskady trzech następujących po sobie sieci neuronowych: P-Net (ang. *Proposal Network*), R-Net (ang. *Refine Network*) oraz O-Net (ang. *Output Network*). Takie podejście pozwala na stopniowe filtrowanie kandydatów na twarze oraz precyzyjne dopasowywanie ramek otaczających. Co istotne, w ramach jednego procesu lokalizowana jest nie tylko twarz, ale również

jej kluczowe punkty charakterystyczne (oczy, nos, usta), co jest przydatne w dalszych etapach analizy. Mankamentem MTCNN jest jednak relatywnie wysoki koszt obliczeniowy – działanie kaskady trzech sieci zwiększa opóźnienia, co utrudnia zastosowanie w systemach czasu rzeczywistego, szczególnie w urządzeniach wbudowanych bez dedykowanego procesora graficznego (GPU).

Proces detekcji przebiega kaskadowo w trzech etapach, po uprzednim przeskalowaniu obrazu wejściowego do różnych rozmiarów (tzw. piramida obrazów):

1. **P-Net (ang. *Proposal Network*):** pierwsza, w pełni konwolucyjna sieć, operuje na wejściu o niskiej rozdzielczości (12×12 pikseli). Szybko skanuje cały obraz, generując dużą liczbę wstępnych okien-kandydatów wraz z wektorami do korekty ich położenia,
2. **R-Net (ang. *Refinement Network*):** kandydaci o wysokim prawdopodobieństwie z sieci P-Net są wycinani i skalowani do rozmiaru 24×24 pikseli. Bardziej złożona architektura R-Net, zawierająca już warstwę w pełni połączoną, odrzuca znaczną część fałszywych detekcji i dalej doprecyzowuje współrzędne ramek,
3. **O-Net (ang. *Output Network*):** ostatnia i najbardziej rozbudowana sieć w kaskadzie przyjmuje na wejście pozostałe okna w rozdzielczości 48×48 . Dokonuje ona ostatecznej weryfikacji, a jej wyjście stanowi ostateczna ramka otaczająca oraz precyzyjnie zlokalizowane pozycje pięciu kluczowych punktów twarzy.

3.3. YOLO

Rodzina detektorów YOLO (ang. *You Only Look Once*) zrewolucjonizowała podejście do detekcji obiektów, traktując ją jako pojedynczy problem regresji [14]. Obraz jest dzielony na siatkę, a każda komórka przewiduje ramki otaczające oraz prawdopodobieństwa klas. Takie jednoetapowe podejście zapewnia bardzo wysoką prędkość przetwarzania. Nowoczesne implementacje, takie jak YOLOv8, dzięki licznym udoskonaleniom architektury i procesu uczenia, stanowią branżowy standard w wielu zastosowaniach wymagających analizy w czasie rzeczywistym [20].

Architektura współczesnych detektorów YOLO, jest standardowo dzielona na trzy kluczowe komponenty [3]:

- **Backbone (rdzeń, ekstrakcja cech):** Odpowiada za ekstrakcję cech z obrazu wejściowego na różnych poziomach rozdzielczości. Składa się z szeregu warstw konwolucyjnych. W nowszych wersjach, takich jak YOLOv8, stosuje się zoptymalizowane bloki C2f (wariant Cross Stage Partial) oraz moduł SPPF (ang. *Spatial Pyramid Pooling Fast*), który agreguje cechy z różnych pól recepcyjnych,
- **Neck (część pośrednia, agregacja cech):** Agreguje mapy cech z różnych etapów szkieletu. Najczęściej stosowana jest architektura PANet (ang. *Path Aggregation Network*), która poprzez operacje próbkowania w górę (ang. *upsampling*) i konkatenacji łączy głębokie cechy semantyczne z płytszymi cechami o wysokiej rozdzielczości,
- **Head (warstwa wyjściowa, predykcja):** Końcowy moduł odpowiedzialny za predykcję. Operuje na mapach cech i generuje współrzędne ramek otaczających, ocenę pewności detekcji oraz prawdopodobieństwa klas. Zastosowanie wielu głowic na różnych poziomach pozwala na skuteczne wykrywanie obiektów o różnych rozmiarach.

W przypadku YOLOv8-Face mamy do czynienia z wariantem *fine-tuned*, specjalnie wytrenowanym do detekcji twarzy, co zwiększa jego użyteczność w kontekście anonimizacji. Warto dodać, że YOLOv8 występuje w kilku wariantach różniących się wielkością modelu (od *n* – nano, po *x* – extra large), co pozwala znaleźć kompromis pomiędzy skutecznością a wydajnością obliczeniową.

3.4. YuNet

YuNet to minimalistyczny model konwolucyjnej sieci neuronowej (CNN), zaprojektowany od podstaw z myślą o ekstremalnej wydajności na urządzeniach brzegowych o ograniczonych zasobach. Jego architektura bezkotwicowa (ang. *anchor-free*), posiadająca mniej niż 76 tysięcy parametrów, osiąga wynik mAP na poziomie 0,91 na zbiorze WIDER FACE (*easy set*) przy czasie inferencji wynoszącym zaledwie 1,6 ms na standardowym procesorze laptopowym. Czyni go to praktycznym wyborem do zastosowań w systemach wbudowanych. W porównaniu do bardziej złożonych modeli, takich jak YOLOv8-Face, YuNet oferuje konkurencyjną skuteczność przy znacznie mniejszych wymaganiach obliczeniowych. Jest to szczególnie istotne w kontekście elektronicznych jednostek sterujących (ECU), gdzie kluczowe jest zachowanie wysokiej przepustowości (FPS) przy minimalnym zużyciu energii. Dzięki temu YuNet jest atrakcyjnym rozwiązaniem dla aplikacji motoryzacyjnych, w których niezbędne jest równoczesne zapewnienie efektywności i oszczędności energetycznej.

3.5. Podsumowanie i porównanie metod

Powyższe metody zostały syntetycznie zestawione w tabeli 2, która uwypukla kluczowe różnice w ich charakterystykach wydajnościowych i rozmiarowych. Wybór konkretnego detektora w praktyce jest zawsze kompromisem pomiędzy wymaganą skutecznością a dostępnymi zasobami obliczeniowymi platformy docelowej.

Tabela 2
Porównanie detektorów twarzy

Metoda	Rok	Wydajność (orientacyjna)	Parametry [mln]
Viola–Jones	2001	~15 FPS @ CPU [22]	^a
MTCNN	2016	~10 FPS GPU; wolniej na CPU [25]	^b
YOLOv8n-Face	2023	~12 FPS @ CPU [21]	3,2 [21]
YuNet	2023	>600 FPS @ CPU [23]	0,076 [23]

^a Metoda klasyczna (kaskada cech) – liczba parametrów nie ma sensu.

^b Różne implementacje MTCNN – brak spójnych danych o liczbie parametrów.

4. Anonimizacja

Ewolucja technik anonimizacji odzwierciedla zmianę paradygmatu — od „niszczenia informacji” do „zastępowania informacji przy zachowaniu istotnych atrybutów”. Wybór konkretnej metody zależy od kompromisu pomiędzy siłą ochrony prywatności, użytecznością danych a złożonością obliczeniową.

4.1. Techniki zniekształcające (obfuskacja)

Najprostszą i historycznie najwcześniejszą grupą metod są techniki zniekształcania obrazu (ang. *obfuscation*), które modyfikują region obrazu zawierający twarz na

poziomie pikseli, aby uniemożliwić jej rozpoznanie. Do najpopularniejszych należą rozmycie gaussowskie, pikselizacja oraz maskowanie (całkowite zasłonięcie twarzy, np. czarnym prostokątem). Ich podstawową wadą jest nie tylko niszczenie użytecznych atrybutów (np. mimiki), ale również rosnące ryzyko odwracalności operacji, co szczegółowo analizowano w pracach takich jak [16]. Zaletą jest natomiast bardzo niska złożoność obliczeniowa. Metody te, choć niezwykle szybkie i łatwe w implementacji, praktycznie nie nadają się do scenariuszy wymagających zachowania użytecznych atrybutów obrazu (np. kierunku spojrzenia czy mimiki), co czyni je niekompatybilnymi z wymaganiami systemów monitorowania kierowcy DMS (ang. *Driver Monitoring System*).

4.2. Metody oparte na uczeniu maszynowym (generatywne)

W odpowiedzi na ograniczenia prostych filtrów, rozwinęły się metody oparte na głębokim uczeniu maszynowym, których celem jest zastąpienie oryginalnej twarzy nową, syntetyczną, wygenerowaną przez model:

- **Generatywne Sieci Przeciwwstawne (GAN):** Architektury takie jak DeepPrivacy [8] składają się z generatora i dyskryminatora. Generator uczy się tworzyć realistyczne twarze pasujące do kontekstu sceny, podczas gdy dyskryminator ocenia ich autentyczność. Zaletą jest wysoka jakość, a wadą potencjalna niestabilność treningu i duża złożoność obliczeniowa,
- **Autoenkodery (AE) i Autoenkodery Wariacyjne (VAE):** Modele te, oparte na architekturze koder-dekoder, uczą się kompresować dane wejściowe do niskowymiarowej reprezentacji w przestrzeni latentnej. Kluczową ideą jest tu **rozplątanie cech** (ang. *feature disentanglement*), czyli nauczenie modelu takiej reprezentacji, w której różne wymiary odpowiadają za różne atrybuty obrazu (np. tożsamość, mimika). Pozwala to na modyfikację jedynie wymiarów kodujących tożsamość [12],
- **Modele Dyfuzyjne:** Najnowsza i obecnie najpotężniejsza klasa modeli generatywnych [11]. Działają na zasadzie stopniowego dodawania szumu do obrazu, a następnie uczenia sieci neuronowej odwracania tego procesu. Oferują stabilny trening i generowanie obrazów o bezprecedensowej jakości. Ich głównym ograniczeniem w kontekście zastosowań w systemach ADAS pozostaje jednak bardzo wysoki koszt obliczeniowy oraz duże zapotrzebowanie na pamięć, co istotnie utrudnia implementację na platformach wbudowanych działających w czasie rzeczywistym.

4.3. Techniki zakłócające (Cloaking)

Jeszcze inne podejście, którego przykładem jest Fawkes [19], polega na dodawaniu do obrazu twarzy subtelnych, niemal niewidocznych dla człowieka zakłóceń. Zakłócenia te są tak zaprojektowane, aby powodować, że modele rozpoznawania twarzy klasyfikują obraz jako należący do zupełnie innej tożsamości, skutecznie chroniąc prywatność osoby na zdjęciu. Należy jednak podkreślić, że skuteczność tego podejścia jest mocno zależna od architektury modelu używanego do rozpoznawania twarzy. Perturbacje zaprojektowane dla jednego algorytmu mogą być mniej skuteczne w przypadku innych sieci neuronowych, co ogranicza ich uniwersalność.

5. Ewaluacja wyników

W celu kompleksowej oceny, zdefiniowano zestaw metryk pogrupowanych w trzy kategorie:

- **Jakość detekcji:** Do oceny skuteczności lokalizacji twarzy wykorzystano standardowe miary stosowane w zadaniach detekcji obiektów: *Precyzję* (1)), która określa odsetek poprawnych detekcji, oraz *Czułość* (2), wskazującą na odsetek wykrytych prawdziwych twarzy. Do oceny dokładności dopasowania ramek służy miara *IoU* (ang. *Intersection over Union* (3)).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

gdzie:

TP (ang. *True Positives*) – liczba poprawnie wykrytych twarzy, FP (ang. *False Positives*) – liczba błędnie wykrytych obiektów, które w rzeczywistości nie są twarzami.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

gdzie:

FN (ang. *False Negatives*) oznacza liczbę twarzy, które nie zostały wykryte przez detektor.

$$\text{IoU} = \frac{\text{Area}(B_{\text{predicted}} \cap B_{\text{ground_truth}})}{\text{Area}(B_{\text{predicted}} \cup B_{\text{ground_truth}})} \quad (3)$$

gdzie:

gdzie: $B_{\text{predicted}}$ jest ramką predykcji wygenerowaną przez detektor, $B_{\text{ground_truth}}$ jest ramką referencyjną, $\cap$ oznacza część wspólną zbiorów, $\cup$ ich sumę, natomiast $\text{Area}(\cdot)$ oznacza pole powierzchni danego obszaru.

- **Jakość obrazu (użyteczność danych):** miary te kwantyfikują stopień degradacji obrazu po zastosowaniu anonimizacji. Wykorzystano PSNR (ang. *Peak Signal-to-Noise Ratio*) (4) oraz SSIM (ang. *Structural Similarity Index*) (5), które porównują obraz zanonimizowany z oryginałem odpowiednio na poziomie pikseli i struktury obrazu:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (4)$$

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (5)$$

- **"Siła" anonimizacji (prywatność):** Do oceny, jak skutecznie anonimizacja usuwa tożsamość osoby, zaimplementowano miarę adversarialną opartą na modelu rozpoznawania twarzy. Wykorzystano model *FaceNet* [18] (w implementacji z *facenet-pytorch*, wytrenowany na zbiorze VGGFace2) do ekstrakcji wektorów cech (ang. *embeddings*) z oryginalnych i zanonimizowanych wizerunków. Siła

ochrony jest mierzona jako *Privacy-TPR*, czyli odsetek twarzy, dla których odległość euklidesowa między wektorem oryginalnym a zanonimizowanym jest mniejsza od zadanego progu, co oznacza, że tożsamość nie została skutecznie ukryta. Niższa wartość tej miary świadczy o silniejszej ochronie prywatności. W przeciwieństwie do klasycznych miar detekcji, gdzie wyższe wartości są pożądane, w przypadku *Privacy-TPR* celem jest minimalizacja wyniku,

- **Jakość detekcji po anonimizacji (mAP-post):** W celu oceny, jak bardzo dana metoda anonimizacji degraduje użyteczność obrazu dla algorytmów wizyjnych, wprowadzono miarę *mAP-post*. Jest to standardowa miara średniej precyzji (ang. *mean Average Precision*, mAP), obliczana dla wyników detekcji uzyskanych na obrazie po zastosowaniu anonimizacji. Jako dane referencyjne (ang. *ground truth*) do obliczenia tej metryki służą oryginalne, niezmienione adnotacje. Wartość mAP-post wyrażona w procentach wskazuje, w jakim stopniu detektor jest w stanie poprawnie zlokalizować twarze, mimo nałożonych zniekształceń. Wysoka wartość świadczy o niewielkim wpływie anonimizacji na zdolność do re-detekcji.

6. Eksperymenty badawcze i uzyskane wyniki

Przeprowadzone badania miały charakter kompleksowej analizy porównawczej. Zdefiniowano łącznie 12 unikalnych scenariuszy eksperymentalnych, które powstały z połączenia każdego z czterech wybranych detektorów twarzy z każdą z trzech podstawowych technik anonimizacji. Kombinacje te przedstawiono w tabeli 3. Dla każdego z 12 scenariuszy przeprowadzono pełny cykl przetwarzania na całym zbiorze danych PP4AV, składającym się z ponad 3447 obrazów. Z punktu widzenia liczby pikseli zajmowanych przez twarz dokonano rozróżnienia na twarze małe (rozmiar poniżej 40x40 pikseli) i twarze duże.

Tabela 3
Zestawienie 12 scenariuszy eksperymentalnych

Detektor	Rozmycie gaussowskie	Pikselizacja	Maskowanie
Violi-Jonesa	✓	✓	✓
MTCNN	✓	✓	✓
YuNet	✓	✓	✓
YOLOv8-Face	✓	✓	✓

6.1. Środowisko badawcze

- **Środowisko programistyczne:** Oprogramowanie zaimplementowano w języku *Python 3.8* w ramach środowiska wirtualnego *Conda*. Interfejsem deweloperskim był *Jupyter Notebook*. Do kluczowych zadań wykorzystano biblioteki: *OpenCV* do przetwarzania obrazu, *PyTorch 2.4.1* ze środowiskiem *CUDA 12.1* do akceleracji modeli głębokich, *Pandas* do analizy danych oraz *Matplotlib* i *Seaborn* do wizualizacji wyników.
- **Platforma sprzętowa:** Badania przeprowadzono na komputerze z procesorem *AMD Ryzen 7 7800X3D*, *32 GB* pamięci *RAM DDR5* i kartą graficzną *NVIDIA GeForce RTX 4070 Ti SUPER*.

6.2. Analiza skuteczności detektorów

Pierwszym krokiem analizy ilościowej była ocena skuteczności czterech zaimplementowanych detektorów twarzy. W tym celu, dla każdego detektora, zagregowano wyniki detekcji ze wszystkich 3447 obrazów ze zbioru PP4AV, a następnie obliczono miary Precyzji i Czułości (Recall). Wyniki te zostały przedstawione w tabeli 4. De-

Tabela 4

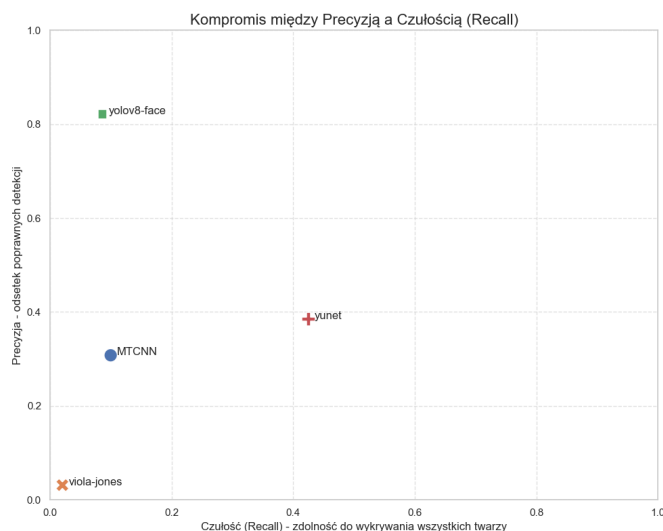
Zagregowane wyniki skuteczności dla detektorów

Detektor	Precyzja	Czułość (Recall)		
		Ogólna	Małe twarze	Duże twarze
MTCNN	0,308	0,100	0,088	0,453
Violi-Jonesa	0,031	0,020	0,013	0,244
YOLOv8-Face	0,821	0,086	0,069	0,628
YuNet	0,385	0,425	0,417	0,686

tektor *YOLOv8-Face* osiągnął zdecydowanie najwyższą precyzję (0,821), co oznacza, że jego detekcje były w przeważającej mierze poprawne i generował on bardzo niewiele fałszywych alarmów. Z kolei najwyższą ogólną czułość (0,425) uzyskał model *YuNet*, co świadczy o jego zdolności do wykrywania największego odsetka wszystkich twarzy obecnych na obrazach. Wyniki te wskazują na istnienie fundamentalnego kompromisu: *YOLOv8-Face* jest bardziej „ostrożny” i precyzyjny, natomiast *YuNet* bardziej „czuły”, lecz kosztem niższej precyzji. Klasyczny algorytm Violi-Jonesa uzyskał bardzo niskie wyniki w obu metrykach, co potwierdza jego historyczny charakter.

W literaturze skuteczność detektorów twarzy najczęściej wyraża się za pomocą *TPR* (ang. *True Positive Rate*), czyli miary równoważnej czułości (*Recall*). Do ewaluacji powszechnie stosuje się zbiór Fddb [9], zawierający głównie obrazy twarzy frontalnych, oraz zbiór WIDER FACE [24], który jest znacznie trudniejszy, ponieważ obejmuje małe, częściowo zasłonięte oraz niefrontalne twarze. Zbiór PP4AV [13] jest jeszcze bardziej wymagający, gdyż składa się ze scen miejskich typowych dla systemów ADAS, w których dominują małe i niefrontalne twarze oraz występują dodatkowe zakłócenia środowiskowe.

Wzajemną relację między Precyzją a Czułością dla każdego z detektorów przedstawiono w formie wykresu punktowego na rysunku 1.



Rys. 1. Zależność pomiędzy Precyzją a Czułością (ogólną) dla czterech testowanych detektorów. Idealny detektor znajdowałby się w prawym górnym rogu

6.3. Analiza wydajności obliczeniowej

Kluczowym aspektem w ocenie systemów przeznaczonych do zastosowań w motoryzacji jest ich wydajność obliczeniowa, wyrażona w klatkach na sekundę (FPS). Dla systemów czasu rzeczywistego, takich jak ADAS, wartość ta powinna być możliwie jak najwyższa i stabilna. Średnią wydajność dla każdego z 12 testowanych scenariuszy, obliczoną na podstawie czasów przetwarzania całego zbioru danych, przedstawiono w tabeli 5.

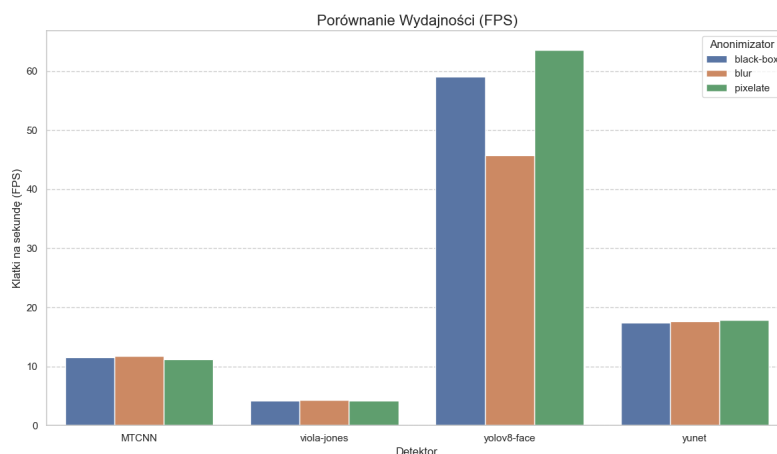
Tabela 5

Porównanie średniej wydajności (FPS) dla 12 scenariuszy badawczych

Detektor	Anonimizator	Średnia wydajność (FPS)
MTCNN	maskowanie (black-box)	6,4
	rozmycie (blur)	6,3
	pikselizacja	4,7
Viola–Jones	maskowanie (black-box)	4,2
	rozmycie (blur)	4,2
	pikselizacja	4,3
YOLOv8-Face	maskowanie (black-box)	67,1
	rozmycie (blur)	66,1
	pikselizacja	69,3
YuNet	maskowanie (black-box)	17,2
	rozmycie (blur)	17,1
	pikselizacja	17,1

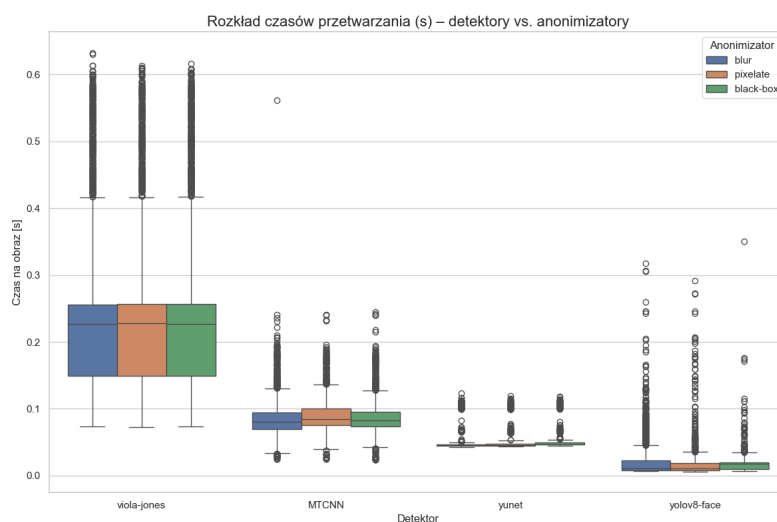
Analiza powyższych wyników jednoznacznie wskazuje, że detektor YOLOv8-Face jest zdecydowanie najwydajniejszym z testowanych modeli, osiągając średnio 66-69 FPS. Taka wydajność z niewielką awaryjnością spełnia wymagania systemów czasu rzeczywistego.

Na drugim miejscu uplasował się YuNet, który ze średnim wynikiem ok. 17 FPS również przekracza próg 15 FPS, często uznawany za minimum dla płynnego przetwarzania wideo. Znacznie słabiej wypadły modele MTCNN (ok. 4-6 FPS) oraz Viola-Jonesa (ok. 4 FPS). Co niezwykle istotne, wyniki pokazują, że wybór metody anonimizacji ma marginalny wpływ na ogólną wydajność systemu. Czas obliczeniowy jest niemal w całości zdominowany przez proces detekcji. Graficzne porównanie tych wartości przedstawiono na rysunku 2. Oprócz średniej wydajności, istotna jest również jej stabilność i przewidy-



Rys. 2. Porównanie wydajności (FPS) dla poszczególnych kombinacji detektora i anonimizatora

walność. Rozkład czasów przetwarzania dla pojedynczych obrazów zbadano za pomocą wykresów pudełkowych, przedstawionych na rysunku 3. Należy podkreślić, że prezen-



Rys. 3. Rozkład czasów przetwarzania (w sekundach) dla poszczególnych scenariuszy. Wykres pudełkowy ilustruje medianę, kwartyle oraz wartości odstające

owane wartości odnoszą się wyłącznie do *inferencji*, czyli działania już wytrenowanych modeli.

Wykresy pudełkowe potwierdzają wnioski z analizy średniej wydajności i dostarczają dodatkowych informacji. Modele YOLOv8-Face oraz YuNet charakteryzują

się nie tylko najkrótszym, ale również najbardziej stabilnym i przewidywalnym czasem przetwarzania. Rozstęp międzykwartylowy jest dla nich bardzo wąski, a wartości odstające nieliczne. W przeciwieństwie do nich, detektor MTCNN wykazuje bardzo dużą wariację czasów przetwarzania, z licznymi wartościami odstającymi przekraczającymi nawet 1,5 sekundy. Oznacza to, że choć jego średnia wydajność jest niska, w niektórych przypadkach potrafi on działać wielokrotnie wolniej, co czyni go nieprzewidywalnym i nieodpowiednim dla systemów krytycznych pod względem bezpieczeństwa.

6.4. Analiza "siły" anonimizacji

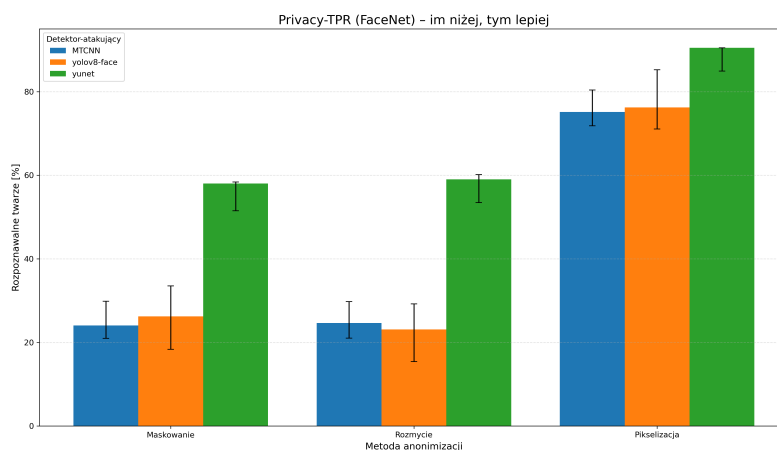
Ostatnim etapem analizy ilościowej była ocena faktycznej "siły" zaimplementowanych technik anonimizacji. W tym celu zbadano dwa aspekty: zdolność detektora do ponownego odnalezienia twarzy na obrazie już zanonimizowanym oraz, co kluczowe, zdolność dedykowanego modelu rozpoznawania twarzy do odzyskania tożsamości osoby z takiego obrazu. Szczegółowe wyniki dla trzech "najsilniejszych" detektorów (MTCNN, YOLOv8-Face, YuNet) przedstawiono w tabeli 6. Wyniki dla klasycznego algorytmu Viola-Jonesa pominięto z uwagi na skrajnie niską skuteczność detekcji i brak wartości poznawczej w analizie jakości anonimizacji. W kolumnie Wskaźnik Przetrwania zapisano odsetek twarzy ponownie wykrytych po anonimizacji. Natomiast Privacy-TPR to odsetek twarzy rozpoznanych przez model FaceNet, gdzie mniejsza wartość oznacza lepszy wynik.

Tabela 6
Wyniki dla "siły" anonimizacji

Detektor	Metoda	Wskaźnik Przetrwania (%)	Privacy-TPR (%)	CI 95%
MTCNN	black-box	0,8	23,9	[21,0 - 29,7]
	blur	1,2	24,5	[20,5 - 29,2]
	pixelate	48,9	74,8	[71,8 - 80,3]
YOLOv8-Face	black-box	1,3	26,3	[18,4 - 33,3]
	blur	21,3	23,1	[15,6 - 30,2]
	pixelate	88,8	75,6	[70,5 - 84,5]
YuNet	black-box	3,7	58,1	[51,5 - 58,4]
	blur	32,3	59,0	[53,4 - 60,2]
	pixelate	77,7	90,5	[84,9 - 89,6]

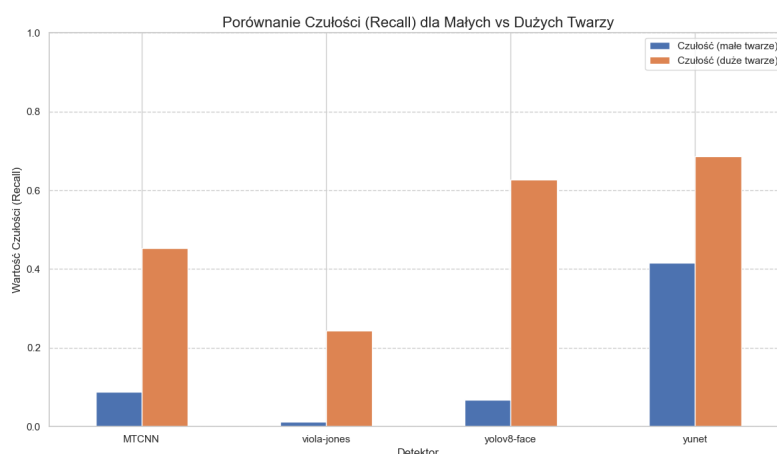
Pierwszą obserwacją jest fakt, że technika *pikselizacji* w bardzo niewielkim stopniu utrudnia ponowną detekcję twarzy – w zależności od detektora, od 49% do aż 89% twarzy jest ponownie lokalizowanych. Znacznie skuteczniejsze pod tym względem są metody rozmycia i maskowania, choć i w ich przypadku detektor YuNet wykazuje pewną odporność, odnajdując 32% twarzy po rozmyciu.

Kluczową miarą jest jednak *Privacy-TPR*, która ocenia, czy tożsamość osoby może zostać odzyskana przez dedykowany model rozpoznawania twarzy. Wyniki tej miary, wraz z 95-procentowymi przedziałami ufności CI (ang. *Confidence Interval*), zwizualizowano na rysunku 4.



Rys. 4. Porównanie metryki Privacy-TPR (im niższa wartość, tym silniejsza ochrona prywatności). Słupki błędów oznaczają 95-procentowy przedział ufności

Jak pokazano na rysunku 4 technika *pixelate* jest zdecydowanie najslabszą metodą anonimizacji. Niezależnie od użytego detektora, pozwala ona na odtworzenie tożsamości w ponad 75% przypadków, a dla detektora YuNet wskaźnik ten sięga aż 90%. Oznacza to, że pikselizacja, mimo popularności, nie zapewnia skutecznej ochrony prywatności w kontekście nowoczesnych modeli rozpoznawania twarzy.



Rys. 5. Porównanie Czułości (*Recall*) detektorów dla twarzy małych (poniżej 40x40 pikseli) oraz dużych

Metody *blur* i *black-box* oferują znacznie lepszy poziom ochrony, redukując rozpoznawalność do ok. 23-26% w przypadku detektorów MTCNN i YOLOv8-Face. Ciekawą obserwacją jest anomalia dla detektora YuNet – gdy jest on używany do wstępnej lokalizacji twarzy, skuteczność anonimizacji drastycznie spada, a Privacy-TPR wzrasta do ok. 58-59%. Można to wyjaśnić jego wysoką skutecznością w wykrywaniu małych twarzy (rys. 5). Twarze te, mające niską rozdzielczość, nawet po zanonimizowaniu mogą być na tyle podobne do oryginału (w przestrzeni cech FaceNet), że odległość euklidesowa między ich wektorami cech nie przekracza progu rozpoznania. Wskazuje to na istotne wyzwanie: skuteczna anonimizacja małych, odległych twarzy jest znacznie trudniejsza niż w przypadku dużych wizerunków.

Tabela 7

Zestawienie uzyskanych wyników wraz z oceną stopnia degradacji obrazu

Detektor	Anon.	Prec.	Rec.	mAP-post	FPS	PSNR	SSIM	Priv. TPR
MTCNN	black-box	0,308	0,100	0,0	11,5	8,92	0,0027	23,94
	blur	0,308	0,100	0,0	11,8	18,05	0,3465	24,55
	pixelate	0,308	0,100	6,9	11,2	23,89	0,7000	74,85
YOLOv8-Face	black-box	0,821	0,086	0,1	59,0	9,21	0,0008	26,25
	blur	0,821	0,086	2,2	45,7	18,78	0,2364	23,13
	pixelate	0,821	0,086	7,7	63,6	27,20	0,8549	75,63
YuNet	black-box	0,385	0,425	0,2	17,4	9,22	0,0013	58,10
	blur	0,385	0,425	11,3	17,7	18,89	0,2167	58,95
	pixelate	0,385	0,425	26,2	17,8	26,63	0,8385	90,50

6.5. Zniekształcenie obrazu w wyniku anonimizacji

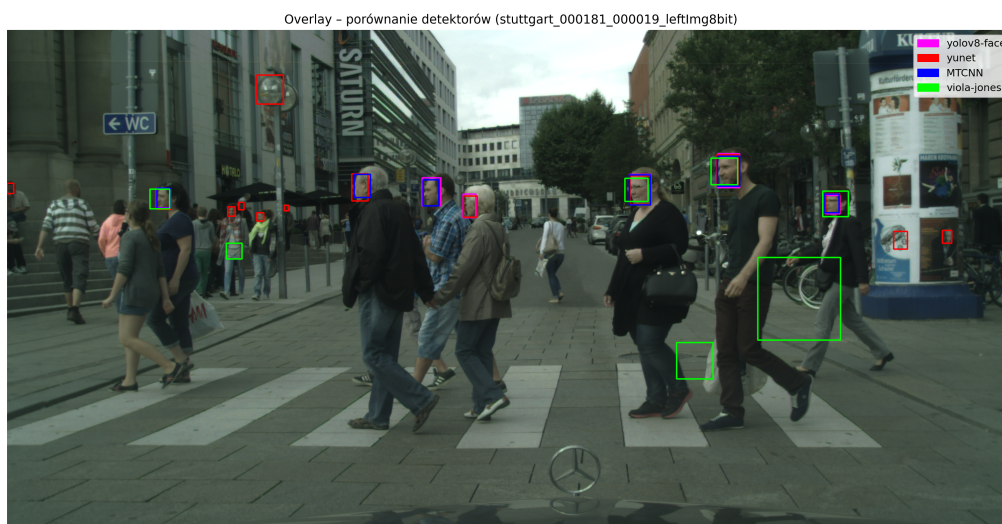
Dla oceny stopnia degradacji obrazu w wyniku przeprowadzonej anonimizacji zastosowano miary: PSNR oraz SSIM, które odnoszą się wyłącznie do regionów zainteresowania ROI (ang. *Region of Interest*), czyli obszarów twarzy poddanych anonimizacji. Dane przedstawiono w zbiorczej tabeli 7 gdzie dodatkowo wartości Precision, Recall oraz mAP-post zostały wyrażone w postaci ułamka (0–1), natomiast Privacy-TPR w procentach. Analogicznie, wartości FPS oznaczają liczbę przetworzonych obrazów na sekundę. Privacy-TPR została zdefiniowana jako skuteczność modelu reidentyfikacyjnego FaceNet w odróżnianiu tożsamości osób na obrazach po anonimizacji. Wysoka wartość oznacza, że model nadal poprawnie rozpoznaje dużą część osób (słaba ochrona prywatności), podczas gdy niska wartość świadczy o skutecznym zatarciu tożsamości.

Zastosowanie anonimizacji z założenia zmienia obraz w obszarze znalezionej twarzy stąd zdecydowano się na liczenie miar jakości tylko dla tych obszarów: PSNR oraz SSIM. Wyniki pokazują, że np. anonimizacja metodą *pixelate* daje stosunkowo wysokie wartości wskaźników jakości, co oznacza, że struktura obrazu w ROI jest wciąż dobrze zachowana. Sytuacja ta koreluje z wysokim poziomem ponownej rozpoznawalności twarzy (Privacy-TPR $\approx$ 75–90%). Z kolei metoda *black-box* prowadzi do ekstremalnej degradacji ROI (bardzo niskie SSIM), zapewniając znacznie lepszą ochronę prywatności (niższy Privacy-TPR), ale kosztem spadku użyteczności wizualnej. W przypadku wykorzystania danych po anonimizacji np. przez systemu monitorujące kierowcę DMS zagadnienie to pokazuje że konieczne jest szukanie kompromisu pomiędzy "siłą" anonimizacji a stratą jakości obrazu.

6.6. Przykładowe wyniki

Przykładowe wyniki działania detektorów twarzy na zbiorze danych PP4AV przedstawiono na rysunku 6, natomiast rezultaty anonimizacji przedstawia rysunek 7. Proponowane rozwiązania zostały przetestowane także na danych pochodzących od partnera z przemysłu związanego z systemami ADAS. Uzyskiwane wyniki są obiecujące, co można zauważyć na rysunku 8. Zbiór danych pochodzący od partnera był nagrywany w bardzo różnych warunkach i często jest gorszej jakości niż zbiór publicz-

ny PP4AV. Występują zatem sytuacje gdzie proponowane rozwiązanie nie radzi sobie całkowicie dobrze. Rozwiązaniem tego problemu może być dotrenowanie modeli na dodatkowych danych pochodzących od partnera, co wymaga długotrwałego procesu adaptacji.

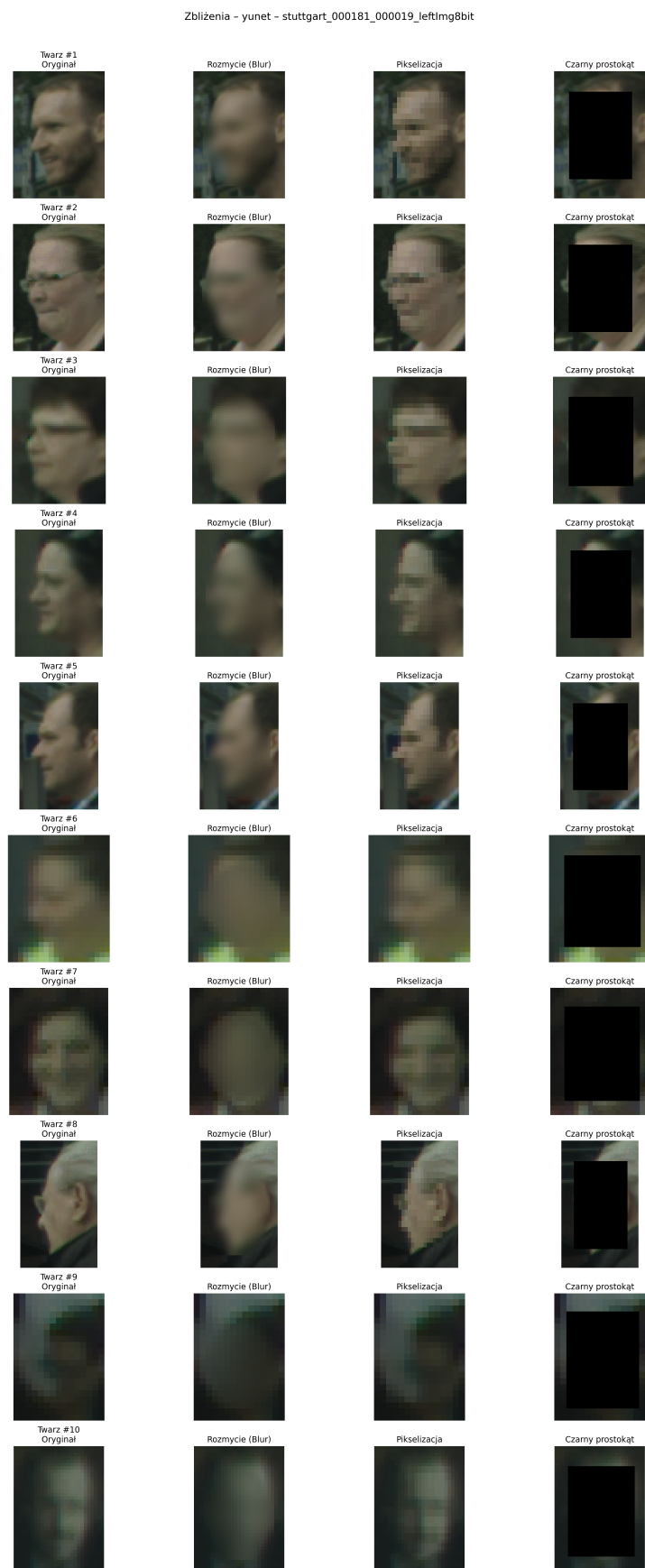


Rys. 6. Porównanie skuteczności czterech badanych detektorów (Violi–Jonesa , MTCNN, YOLOv8-Face, YuNet) na przykładzie obrazu *Stuttgart_000181*

7. Podsumowanie

Uzyskane wyniki pozwalają stwierdzić, że model YOLOv8-Face oferuje najwyższą wydajność i precyzję (tab. 5), co czyni go idealnym do zadań wymagających bezbłędnej identyfikacji dobrze widocznych obiektów. Jego skuteczność drastycznie spada jednak w przypadku detekcji małych twarzy, co w kontekście bezpieczeństwa i prywatności jest wadą krytyczną. W tym scenariuszu znacznie lepsze wyniki osiągał detektor YuNet, który cechuje się najwyższą czułością (tab. 4), zwłaszcza dla małych twarzy, co jest kluczowe z perspektywy kompletności procesu anonimizacji – minimalizuje ryzyko pominięcia jakiegokolwiek twarzy. W zakresie analizy technik anonimizacji pokazano, że popularna metoda pikselizacji zapewnia niewystarczający poziom ochrony prywatności, pozwalając na odtworzenie tożsamości w ponad 75% przypadków (tab. 6). Stwarza to fałszywe poczucie bezpieczeństwa i dyskwalifikuje ją jako wiarygodne rozwiązanie. Skuteczniejsze okazały się rozmycie gaussowskie i proste maskowanie, które redukowały rozpoznawalność twarzy do znacznie niższego poziomu.

Za konfigurację oferującą najlepszy kompromis między kompletnością detekcji, wydajnością a siłą ochrony prywatności uznano połączenie detektora YuNet z metodą anonimizacji *blur* opartą na rozmyciu gaussowskim. Wybór detektora YuNet podyktowany jest jego najwyższą czułością, która minimalizuje ryzyko pominięcia jakiegokolwiek twarzy wymagającej anonimizacji, co stanowi nadrzędny priorytet. Z kolei rozmycie gaussowskie zapewnia skuteczny poziom ochrony przy znikomym narzucie obliczeniowym i zachowaniu dobrej jakości obrazu.



Rys. 7. Przykładowe twarze wykryte przez YuNet oraz efekty zastosowania trzech metod anonimizacji



Rys. 8. Przykład detekcji (YuNet) i anonimizacji twarzy (rozmycie) na danych pochodzących od partnera

LITERATURA

1. Colossus . When will robots go mainstream? <https://joincolossus.com/article/when-will-robots-go-mainstream/> 2025. Dostęp: 2025-09-03.
2. Deng J., Guo J., Xue N., Zafeiriou S.: ArcFace: Additive angular margin loss for deep face recognition, In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, p. 4690–4699.
3. Elhanashi A., Dini P., Saponara S., Zheng Q.: TeleStroke: real-time stroke detection with federated learning and YOLOv8 on edge devices, Journal of Real-Time Image Processing, 2024, vol. 21(4) p. 121.
4. European Commission . Road safety thematic report: Advanced driver assistance systems. Technical report European Commission, Directorate General for Transport Brussels 2021.
5. European New Car Assessment Programme . Euro NCAP 2025 Roadmap: In Pursuit of Vision Zero 2017.
6. Herger M. Tesla investor day 2023. <https://thelastdriverlicenseholder.com/2023/03/01/tesla-investor-day-2023/> 2023. Dostęp: 2025-09-03.
7. Huang G. B., Mattar M., Berg T., Learned-Miller E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07–49 University of Massachusetts, Amherst 2007.

8. Hukkelås H., Mester R., Lindseth F.: DeepPrivacy: A generative adversarial network for face anonymization, In: International Symposium on Visual Computing (ISVC), 2019, p. 566–578.
9. Jain A. K. Learned-Miller E. G.: Fddb: A benchmark for face detection in unconstrained settings. Technical Report 07–49 University of Massachusetts, Amherst 2010.
10. Jocher G., Chaurasia A., Qiu J.: Ultralytics YOLOv8. <https://github.com/ultralytics/ultralytics> 2023. Dostęp: 2026-06-034.
11. Kung H.-W., Varanka T., Saha S., Sim T., Sebe N.: Face anonymization made simple, In: Proceedings of the Winter Conference on Applications of Computer Vision (WACV), February 2025, p. 1040–1050.
12. Li T. Lin L.: Anonymousnet: Natural face de-identification with measurable privacy, In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2019,.
13. Linh K. Do H.: PP4AV: Privacy-Preserving for Autonomous Vehicles Dataset. <https://github.com/khaclinh/pp4av> 2023. Dostęp: 2025-07-25.
14. Redmon J., Divvala S., Girshick R., Farhadi A.: You only look once: Unified, real-time object detection, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, p. 779–788.
15. Regulation P. Regulation (EU) 2016/679 of the European Parliament and of the Council, Regulation (eu), vol. 679(2016) p. 10–3.
16. Ren Z., Lee Y. J., Ryoo M. S.: Learning to anonymize faces for privacy preserving action detection, In: Proceedings of the European Conference on Computer Vision (ECCV), 2018, p. 620–635.
17. Rössler A., Cozzolino D., Verdoliva L., Riess C., Thies J., Nießner M.: Faceforensics++: Learning to detect manipulated facial images, In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, p. 1–11.
18. Schroff F., Kalenichenko D., Philbin J.: Facenet: A unified embedding for face recognition and clustering, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, p. 815–823.
19. Shan S., Wenger E., Zhang J., Li H., Zheng H., Zhao B. Y.: Fawkes: Protecting privacy against unauthorized deep learning models, In: 29th USENIX Security Symposium (USENIX Security 20), 2020, p. 1589–1604.
20. Ultralytics . Explore Ultralytics YOLOv8. <https://docs.ultralytics.com/models/yolov8/> 2023. Dostęp: 2026-06-25.
21. Ultralytics . Model Comparison: YOLOv8 vs YOLO11. <https://docs.ultralytics.com/compare/yolov8-vs-yolo11/> 2024. Dostęp: 2026-06-25.
22. Viola P. Jones M.: Rapid object detection using a boosted cascade of simple features, In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) volume 1, 2001, p. 511–518.

23. Wu W., Peng H., Yu S.: Yunet: A tiny millisecond-level face detector, *Machine Intelligence Research*, 2023, vol. 20(5) p. 656–665.
24. Yang S., Luo P., Loy C. C., Tang X.: WIDER FACE: A face detection benchmark, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, p. 5525–5533.
25. Zhang K., Zhang Z., Li Z., Qiao Y.: Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Processing Letters*, 2016, vol. 23(10) p. 1499–1503.
26. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O.: The unreasonable effectiveness of deep features as a perceptual metric, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, p. 586–595.