Metody śledzenia wzroku na urządzeniach mobilnych

Architektury estymacji kierunku spojrzenia, kalibracja, fuzja sensorów, wyniki porównawcze, analiza zdarzeń okulomotorycznych i walidacja

Streszczenie

Raport przedstawia szczegółowy przegląd metod szacowania punktu spojrzenia na ekranie (2D point of gaze, PoG) i kierunku patrzenia w przestrzeni 3D na smartfonach oraz tabletach. Omówiono rozwiązania model-based, feature-based, appearance-based i hybrydowe, algorytmy CNN, LSTM, GRU i transformer, a także fuzję kamer RGB, czujników głębi, podczerwieni i IMU. Zachowano ilościowe wyniki prac iTracker, Valliappan, DynamicGaze, RGBDGaze, EasyGaze i MAC-Gaze wraz z kontekstem pomiarów, ograniczeniami i obciążeniem obliczeniowym. Przedstawiono metodologię personalizacji, detekcji zdarzeń okulomotorycznych, przetwarzania edge oraz proponowany protokół walidacji między użytkownikami. Wykazano, że heterogeniczność metryk i protokołów nie pozwala na rzetelne połączenie wszystkich wyników w jedną ważoną metaanalizę; uprawnione są natomiast porównania wewnątrz tego samego badania i jawnie oznaczone syntezy opisowe.

Słowa kluczowe: eye tracking; gaze estimation; point of gaze; iTracker; RGBDGaze; DynamicGaze; CNN–LSTM; kalibracja; fiksacja; sakkada; I-DT; I-VT; Edge AI.

1. Cel, zakres i rodzaj syntezy

Analiza obejmuje urządzenia ręczne z kamerą przednią oraz systemy badawcze wykorzystywane do walidacji mobilnej estymacji spojrzenia. Przez punkt spojrzenia (PoG) rozumie się pozycję na ekranie w zdefiniowanym układzie współrzędnych; przez kierunek spojrzenia – orientację wektora w przestrzeni, wyrażaną zwykle w kącie. To odmienne wielkości fizyczne i nie są bezpośrednio zamienne bez geometrii obserwacji.

W analizie rozróżnia się publikowane wyniki benchmarkowe, dane urządzeń podawane przez producentów oraz cele inżynierskie. Nie przeprowadzono wspólnej reanalizy indywidualnych obserwacji ze wszystkich prac. Z tego powodu opracowanie ma charakter krytycznego przeglądu porównawczego oraz projektu protokołu, a nie pełnej metaanalizy efektów z estymacją wariancji między badaniami.

2. Rozwój mobilnego eye trackingu

Okres	Dominujące rozwiązania / postęp
2002–2010	Wczesne systemy handheld, często wymagające zewnętrznych kamer lub komputerów.
2016	GazeCapture i iTracker: estymacja obrazu → spojrzenie z sieci CNN, około 10–15 FPS.
2018	Przeglądy HCI analizujące interoperacyjność i praktyczne scenariusze mobilne.
2020	Personalizacja mobilnej estymacji w badaniach okulomotorycznych; redukcja błędu po krótkiej kalibracji.
2022	RGBDGaze: łączenie obrazu i głębi w zmiennych pozycjach użytkownika.
2023	Przeglądy architektur, kalibracji i aplikacji gaze na urządzeniach ręcznych.
2024–2025	DynamicGaze, modele temporalne, benchmarking TFLite, prace nad optymalizacją edge.
2025	MAC-Gaze: adaptacja i rekalkibracja wyzwalana ruchem IMU.

3. Model i geometria pomiaru spojrzenia

Model-based wykorzystuje geometryczny opis oka, centrum źrenicy i odbicia rogówkowe (glints), często generowane źródłem bliskiej podczerwieni. W technice pupil-center/corneal-reflection (PCCR) geometria jest stabilizowana przez położenie kamery i iluminatora; użycie na niezmodyfikowanym telefonie bywa trudne z uwagi na ruch urządzenia i ograniczony dostęp do NIR [8].

Feature-based korzysta z punktów charakterystycznych oczu i twarzy: kącików powiek, źrenicy i orientacji głowy (w zależności od implementacji) oraz relacji przestrzennych. Zaletą jest możliwość interpretacji wyników regresji, lecz detekcja drobnych elementów oka przy niskiej rozdzielczości jest wrażliwa na zasłonięcia i światło.

Appearance-based estymuje zależność między obrazem twarzy lub oczu a PoG, zwykle przez CNN lub transformer. Zwykle wymaga dużego, różnorodnego zbioru treningowego i normalizacji wejścia. Najczęściej spotykany potok to detekcja twarzy, wycięcie oczu, ujednolicenie rozmiaru, predykcja i kalibracja [1,2,7].

Hybrydy łączą wyuczone cechy obrazu z relacjami geometrycznymi; przykładem jest EasyGaze. Mogą łagodzić część problemów interpretowalności, ale wymagają niezawodnych punktów charakterystycznych. Metody temporalne łączą kolejne obserwacje, co ma znaczenie przy bodźcach ruchomych, choć zwiększa koszt i potencjalne opóźnienie.

Klasa	Wejście i model	Zalety	Ograniczenia
Model-based	Źrenica / NIR / PCCR / geometria	Fizyczna interpretacja	Sprzęt i kalibracja geometryczna.
Feature-based	Landmarki oka, twarzy, głowy	Lekkość i czytelność	Błąd detekcji punktów i perspektywa.
Appearance-based	Obrazy RGB + CNN/ViT	Skalowalność, brak specjalnej kamery	Bias danych i duży koszt obliczeń.
Hybrydowa	CNN + cechy geometryczne	Łączenie wiedzy i uczenia	Zależność od obu torów.
Temporalna	CNN + GRU/LSTM/Transformer	Wykorzystanie dynamiki	Latencja i stabilność strumieniowania.

4. Dokładność i parametry głównych implementacji

4.1. iTracker / GazeCapture (Krafka i in., 2016)

Metoda	CNN appearance-based z wieloma wejściami (twarz, oczy, face grid)
Sensor / kontekst	RGB; telefony i tablety; opcjonalna kalibracja
Dokładność	Bez kalibracji: 1,71 cm telefon, 2,53 cm tablet; z kalibracją: 1,34 cm i 2,12 cm
Czas / koszt	Prędkość około 10–15 FPS w ówczesnych urządzeniach mobilnych.
Ograniczenie	Wyniki dotyczą zadania i epoki sprzętowej publikacji; ważne odniesienie metodologiczne, nie współczesna granica wydajności.

4.2. Valliappan i in. (2020)

Metoda	Mobilny model wzrokowy z personalizacją
Sensor / kontekst	Przednia kamera RGB / landmarki; Pixel 2 XL i zadania okulomotoryczne
Dokładność	1,92 ± 0,20 cm przed personalizacją; 0,46 ± 0,03 cm po; percentyle [5,95] = [0,31;0,72] cm
Czas / koszt	Kalibracja około 30 s i ~100 klatek; raportowany model około 170 tys. parametrów.
Ograniczenie	Wynik po kalibracji uzyskano w określonym protokole; nie oznacza równie małego błędu przy chodzeniu i zmianach oświetlenia.

4.3. DynamicGaze CNN (Gunawardena i in.)

Metoda	CNN dla pojedynczej klatki
Sensor / kontekst	RGB; DynamicGaze, bodziec dynamiczny
Dokładność	RMSE 1,468 cm
Czas / koszt	255,67 ms inferencji modelu na Samsung S22 w TFLite.
Ograniczenie	Punkt odniesienia dla modeli rekurencyjnych w tym samym badaniu.

4.4. DynamicGaze CNN+GRU

Metoda	CNN i warstwa rekurencyjna GRU
Sensor / kontekst	RGB; DynamicGaze
Dokładność	RMSE 1,091 cm, poprawa 25,7% wobec CNN
Czas / koszt	414,81 ms inferencji modelu na Samsung S22.
Ograniczenie	Poprawa dokładności kosztem dłuższego czasu generowania; nie jest to pomiar pełnego toru.

4.5. DynamicGaze CNN+LSTM

Metoda	CNN i rekurencyjne LSTM
Sensor / kontekst	RGB; DynamicGaze
Dokładność	RMSE 0,955 cm, poprawa 34,9% wobec CNN
Czas / koszt	426,18 ms na Samsung S22; w porównaniu statystycznym wariant najlepszy w danym protokole.
Ograniczenie	ANOVA $F(3,396)=254,42$, $p<0,001$; nie uogólniać poza to badanie bez niezależnej replikacji.

4.6. DynamicGaze iTracker (baseline)

Metoda	Odtworzenie iTracker w benchmarku temporalnym
Sensor / kontekst	RGB; DynamicGaze
Dokładność	RMSE 1,499 cm
Czas / koszt	742,22 ms inferencji modelu na Samsung S22.
Ograniczenie	Nie utożsamiać z pierwotnymi wartościami błędu GazeCapture dla telefonów i tabletów.

4.7. RGBDGaze (Arakawa i in., 2022)

Metoda	Fuzja cech RGB i Depth
Sensor / kontekst	RGB + czujnik TrueDepth; konteksty stania, siedzenia, chodzenia i leżenia
Dokładność	1,89 cm błąd średni wobec 2,26 cm dla RGB-only; 16,3% relatywnej redukcji
Czas / koszt	W zestawieniu nie podano porównywalnego czasu inferencji.
Ograniczenie	Efekt zależy od konstrukcji zbioru i obecności działającej kamery głębi.

4.8. EasyGaze (Cheng i in., 2022)

Metoda	Hybrydowa analiza obrazu oczu i geometrii
Sensor / kontekst	RGB; badanie rozdzielczości i oświetlenia
Dokładność	Średni błąd około 1,93° przy obrazie 96×48 px i oświetleniu frontalnym
Czas / koszt	Brak jednolitej informacji o FPS.
Ograniczenie	Błąd kątowy nie jest tożsamy z błędem centymetrowym na ekranie.

4.9. MAC-Gaze (Lei i in., 2025)

Metoda	Model wizualny i ciągła kalibracja inicjowana ruchem
Sensor / kontekst	RGB / RGB-D + IMU; RGBDGaze i MotionGaze
Dokładność	RGBDGaze: 1,73 → 1,41 cm; MotionGaze: 2,81 → 1,92 cm

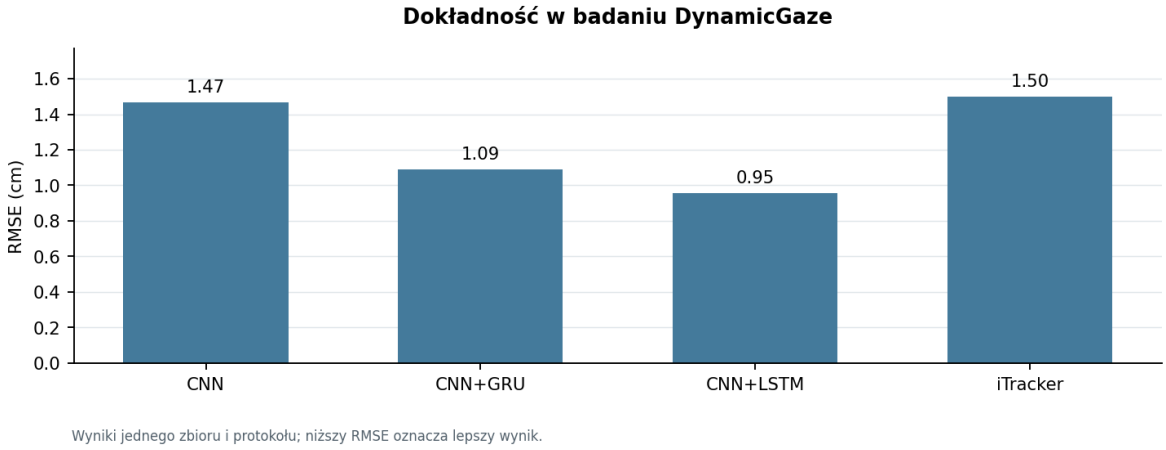
Czas / koszt	Nie podano wspólnej latencji porównywalnej z serią DynamicGaze.
Ograniczenie	Redukcja odpowiednio ~18,5% i ~31,7%; konieczne jest raportowanie kosztu rekalkibracji.

4.10. Apple ARKit (odniesienie w RGBDGaze)

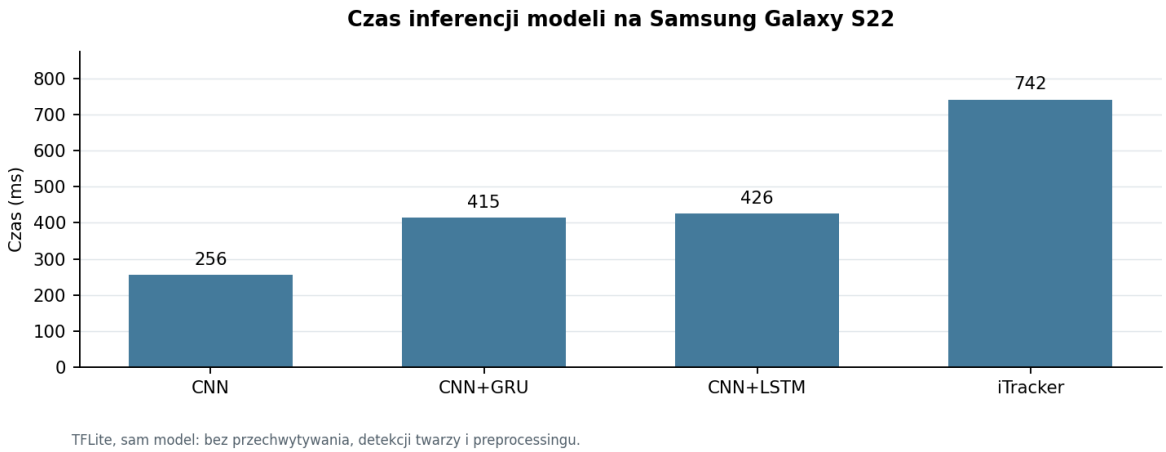
Metoda	Wbudowany pipeline estymacji spojrzenia
Sensor / kontekst	TrueDepth / elementy NIR zależne od modelu urządzenia
Dokładność	6,38 cm w określonym porównaniu RGBDGaze
Czas / koszt	Brak bezpośrednio porównywalnej szybkości.
Ograniczenie	Wartość benchmarkowa dla określonego zastosowania, a nie uniwersalna dokładność ARKit.

5. Porównanie ilościowe w ramach DynamicGaze

W badaniu DynamicGaze cztery warianty estymacji oceniano w ramach jednego zadania. Redukcja RMSE CNN+LSTM względem CNN wyniosła (1,468–0,955)/1,468 = 34,9%, a CNN+GRU względem CNN (1,468–1,091)/1,468 = 25,7%. Autorzy raportują ANOVA F(3,396)=254,42, p<0,001 oraz porównania post hoc Tukey, w których CNN+LSTM uzyskał najlepszy rezultat w danym ustawieniu. Nie świadczy to o statystycznej dominacji tej architektury we wszystkich innych scenariuszach.



Rysunek 1. RMSE modeli z tego samego benchmarku DynamicGaze; błąd wyrażony w centymetrach.



Rysunek 2. Czas inferencji TFLite na Samsung Galaxy S22 dla tej samej serii modeli.

W żadnym z czterech przytoczonych wariantów czas wykonania samego modelu na telefonie nie odpowiada budżetowi 33,3 ms na klatkę typowemu dla 30 Hz. Aby mówić o aplikacji czasu rzeczywistego, należy ponadto doliczyć czas akwizycji,

detekcji twarzy, przetwarzania wstępnego, filtracji, renderowania i aktualizacji interfejsu. Stwierdzenie „model ma wysoką dokładność" nie jest zatem równoznaczne z „system reaguje w czasie rzeczywistym".

6. Architektury temporalne, kwantyzacja i modele brzegowe

CNN+GRU i CNN+LSTM wykorzystują zależności czasowe między obrazami, których nie może odtworzyć klasyfikacja pojedynczej klatki. Dla wideo i interakcji z ruchomymi bodźcami dostęp do historii obserwacji pozwala częściowo kompensować krótkotrwałe zakłócenia. W modelach przyczynowych szczególne znaczenie ma sposób inicjalizacji stanu po utracie twarzy, rekalibracji i zmianie orientacji telefonu.

Pruning, post-training quantization, knowledge distillation i redukcja rozdzielczości umożliwiają zmniejszanie kosztów obliczeniowych. W jednym eksperymencie dla CNN+LSTM na urządzeniu edge raportowano RMSE 0,955 cm dla oryginalnego wariantu, 1,192 cm po kwantyzacji i 1,366 cm po pruning. Dla porównania są to odpowiednio wzrosty błędu około 24,8% i 43,0% względem punktu odniesienia; nie wolno ich traktować jako uniwersalnego kosztu tych technik [5].

Koszt pamięci zależy zarówno od liczby parametrów i precyzji wag, jak i od tensora wejściowego, buforów kamery, aktywacji, stanu sieci rekurencyjnej oraz implementacji runtime. Model zoptymalizowany przez TensorFlow Lite lub Core ML powinien być profilowany na reprezentatywnych urządzeniach z włączonym oszczędzaniem energii i po rozgrzaniu termicznym. Nie wolno ekstrapolować opóźnienia samej sieci do jakości interakcji.

7. Fuzja RGB, głębi, NIR/IR oraz IMU

RGB. Podstawowy sensor mobilny, dostępny w szerokiej klasie smartfonów. Czuły na oświetlenie, zasłonięcia i różnice właściwości kamer.

RGB–D. Umożliwia oddzielne uwzględnienie odległości i orientacji twarzy. Przykład RGBDGaze wskazuje spadek średniego błędu z 2,26 do 1,89 cm w badanym kontekście [4].

NIR/IR. Wspiera obserwację oka przy zdefiniowanym oświetleniu podczerwonym. Przy danych zależnych od platformy trzeba rozróżnić dostępność fizycznego sensora od możliwości wykorzystania jego obrazu przez API.

IMU. Akcelerometr i żyroskop mogą wykrywać zmianę położenia urządzenia, podnosić odporność filtrowania albo inicjować ponowną kalibrację. Rozwiązania typu MAC-Gaze wymagają dodatkowego testu, czy częsta adaptacja nie powoduje dryfu.

8. Kalibracja i personalizacja

Kalibracja jawna wymaga świadomego wskazywania punktów na ekranie, co pozwala budować relację PoG–pozycja ekranu dla konkretnego użytkownika i konfiguracji telefonu. Personalizacja może poprawiać dokładność, czego przykładem jest redukcja błędu średniego z 1,92 ± 0,20 cm do 0,46 ± 0,03 cm, lecz dodaje koszt czasowy około 30 s i wymaga odrębnego zbioru weryfikacyjnego [3].

Kalibracja niejawna wykorzystuje kontekst interakcji, na przykład cel kliknięcia, wymaga jednak ochrony przed samowzmacniającym się błędem. Kalibracja ciągła może aktualizować parametry po wykryciu zmiany pozycji telefonu, ale bez niezależnego sygnału odniesienia istnieje ryzyko dryfu. Wynik po personalizacji powinien być podawany osobno od wyniku bez kalibracji oraz od wyniku po nieznanej zmianie użytkownika.

Warunek	Co mierzyć	Błąd metodologiczny
Brak kalibracji	Generalizacja na nowych użytkowników	Mieszanie z wynikami po personalizacji.
Kalibracja jawna	Czas, liczba punktów, test hold-out	Test na tych samych punktach co kalibracja.
Kalibracja niejawna	Stabilność podczas zwykłego UX	Traktowanie kliknięcia jako prawdy bez kontroli.
Kalibracja ciągła	Częstość, dryf, koszt energii	Adaptacja na danych testowych bez deklaracji.
Motion-aware	Przed/po zmianie pozycji	Brak kontroli zmian oświetlenia i osoby.

9. Narzędzia, biblioteki i sprzęt referencyjny

GazeCapture / iTracker

Dane oraz praca referencyjna do odtwarzania regresji spojrzenia na RGB i analizy kalibracji.

RGBDGaze

Implementacja i benchmark z czujnikiem głębi w scenariuszach mobilnych.

OpenGaze

Narzędzia HCI i reprezentacje appearance-based do eksperymentów z interakcją.

WebGazer.js

Przeglądarkowy eye tracking oparty na kamerze, korzystający z interakcji użytkownika do kalibracji. Wymaga szczególnie ostrożnej oceny na telefonach.

GazeML

Zestaw implementacji metod wizyjnych, przydatny dla badań replikacyjnych.

MobileEye

Kod i modele powiązane z badaniami nad estymacją na smartfonach z bodźcami dynamicznymi.

Tobii Pro Glasses 3

Nagłowny system do badań mobilnych, oparty na PCCR; deklarowane rejestrowanie surowych danych do 100 Hz. Nie jest algorytmem samodzielnie uruchomionym na smartfonie.

Pupil Labs Core / Neon

Systemy nagłowne oraz narzędzia i raporty walidacyjne, potencjalnie przydatne jako sensor odniesienia.

Systemy nagłowne i smartfonowe realizują różne konfiguracje pomiarowe. Porównanie ich dokładności jest uprawnione po zdefiniowaniu wspólnego układu odniesienia, procedury kalibracji oraz sposobu dopasowania czasowego. Same deklaracje częstotliwości próbkowania urządzeń nie gwarantują skuteczności detekcji zdarzeń okulomotorycznych.

10. Zdarzenia okulomotoryczne i analiza sekwencji

Fiksacja odpowiada odcinkowi względnie stabilnego spojrzenia na obszarze zainteresowania; sakkada to szybkie przesunięcie kierunku wzroku między kolejnymi obszarami. Sekwencja fiksacji i sakkad tworzy scanpath, przydatny do opisu strategii przeszukiwania sceny. Analiza mrugnięć i średnicy źrenicy może dostarczać dodatkowych informacji, ale jakość mobilnych kamer RGB często nie wystarcza do ich wiarygodnego pomiaru na poziomie znanym z dedykowanych trackerów.

I-DT (dispersion threshold). Próbki obejmujące krótki przedział czasu klasyfikuje się jako fiksację, jeżeli dyspersja ich położeń pozostaje poniżej zadanego progu. Wyniki zależą od skali i stabilności układu współrzędnych oraz długości minimalnego okna.

I-VT (velocity threshold). Prędkości zmian kierunku spojrzenia porównuje się z progiem, aby rozdzielać wolne i szybkie zdarzenia. W smartfonowym badaniu Valliappan i in. stosowano około 22°/s na danych rzędu 30 Hz [3].

HMM / modele probabilistyczne. Modelują rozkład obserwacji i prawdopodobieństwa przejść pomiędzy stanami, lecz wymagają kalibracji do częstotliwości i jakości danych; złożoność nie musi zapewniać poprawy w każdych warunkach.

Mikrosakkady. Algorytm Engberta i Kliegla wykorzystuje adaptacyjny próg prędkości. Użycie przy 30–60 Hz i wysokim szumie obrazu jest ograniczone; do pomiarów mikrozdarzeń z reguły potrzebny jest znacznie szybszy tracker i odpowiedni sygnał odniesienia [11].

Algorytm	Parametr krytyczny	Potencjalny problem

I-DT	Dyspersja i czas minimalny	Dryf, szum i mobilny ruch głowy.
I-VT	Próg prędkości i odstęp próbek	Różna częstość próbkowania i kwantyzacja.
HMM	Parametry emisji i przejść	Niejawne założenia o częstości stanów.
Mikrosakkady	Wysoka częstotliwość, niski szum	Niewystarczająca rozdzielczość mobilna.

11. Metryki i wymogi porównywalności

Błąd euklidesowy PoG w centymetrach na płaszczyźnie ekranu zależy od fizycznego rozmiaru urządzenia i sposobu mapowania pikseli na jednostki długości. Błąd kątowy w stopniach zależy od kierunku obserwacji i nie może być zsumowany w tej samej tabeli średnich z błędem centymetrowym. Dla przybliżeń małych kątów zależność geometryczna wymaga znajomości odległości oka od ekranu, zatem pozornie proste przeliczenie może wprowadzać istotny błąd.

Należy jawnie podawać, czy miarą jest średni błąd euklidesowy, mediana błędu, RMSE pozycji, MAE czy błąd kątowy. Nie można zastępować jednej drugą bez rozkładu obserwacji. Szczególnie ważne są definicje odrzucania pierwszych próbek po skoku bodźca oraz traktowania przerw w śledzeniu, ponieważ usunięcie trudnych odcinków może poprawić pozorny wynik.

Grupa miar	Minimalne raportowanie
PoG 2D	Średnia, mediana, RMSE/MAE w cm; kalibracja i rozmiar ekranu.
Gaze 3D	Błąd w °; odległość od urządzenia i parametry geometrii.
Dynamika	FPS efektywny, jitter interwału, dropout, opóźnienie P95.
Kalibracja	Liczba punktów, czas, zakres i wynik hold-out.
Zasoby	RAM, pamięć GPU/NPU, zużycie energii i throttling.
Generalizacja	Nowe osoby, urządzenia, pozycje i oświetlenie; przedziały ufności.
Zdarzenia	F1/precision/recall detektora fiksacji-sakkad i parametry progów.

12. Protokół eksperymentalny dla urządzeń mobilnych

Projekt danych. Minimum trzy osie zmienności: osoby, urządzenia i kontekst postawy. Osobno testować siedzenie, stanie, chodzenie, zmianę kąta telefonu, odległości i oświetlenia. Każdy pomiar powinien zawierać odcinki referencyjne oraz okresy swobodnego użycia.

Punkt odniesienia. Kalibrowany sensor zewnętrzny o zweryfikowanej jakości lub kontrolowany bodziec ekranowy; konieczna synchronizacja zegarów i jawna procedura rozróżnienia reakcji sakkadycznej od pozycji prezentowanego bodźca.

Rozdział prób. Podział group split według osób i dodatkowy hold-out urządzeń. Okna z jednej sesji nie mogą trafiać losowo do treningu i testu przy ocenie uogólniania między użytkownikami.

Modele bazowe. iTracker/CNN, model temporalny, RGB–D (tam, gdzie sensor dostępny), IMU-triggered calibration oraz model z kwantyzacją. Testy muszą mieć spójny preprocessing i wspólny protokół.

Przetwarzanie online. Raportować pełny czas capture → face/eye detect → preprocess → model → kalibracja → filtracja → API. Analizować P50/P95/P99, temperaturę, rozładowanie baterii i stabilność przez dłuższą sesję.

Ocena statystyczna. Bootstrap grupowy po osobach, porównania sparowane warunków, CI dla różnicy median/średnich błędów, analiza osób odstających i zmiany dokładności w czasie.

13. Bezpieczeństwo danych i zakres zastosowania

Obraz twarzy i oczu jest szczególnie wrażliwy na wtórne wykorzystanie do identyfikacji. System badawczy powinien domyślnie przetwarzać dane lokalnie, ograniczać retencję surowych ramek, a na potrzeby oceny jakości zapisywać metadane,

estymaty spojrzenia, poziom ufności i niezbędne znaczniki czasu. Zgoda uczestnika, analiza podstaw prawnych i plan udostępniania zbioru muszą odpowiadać celowi projektu oraz charakterowi danych.

Mobilny eye tracking może służyć do oceny interakcji, dostępności, percepcji wizualnej lub badań okulomotorycznych. Nie należy utożsamiać zmiennej wzrokowej z intencją, prawdomównością, zaangażowaniem czy diagnozą medyczną. Do zastosowań klinicznych niezbędna jest osobna walidacja kryterialna i wymogi właściwe dla wyrobu lub procedury medycznej.

14. Główne ograniczenia literatury

- Heterogeniczność raportowania cm vs °, średni błąd vs RMSE, oraz odmienność protokołów kalibracji uniemożliwiają wspólną ilościową metaanalizę bez dodatkowych danych.
- Błąd w ruchu („real-world motion gap”) pozostaje ograniczeniem: modele trenowane w pozycji stabilnej mogą tracić dokładność przy zmianie głowy i urządzenia.
- Czas przetwarzania całego potoku może być znacznie większy od czasu inferencji samej sieci; w raportach konieczne są także pomiary energii i pamięci.
- Personalizacja zmniejsza błąd w niektórych protokołach, ale zwiększa koszt interakcji i może nie przenosić się między sesjami.
- Metody detekcji zdarzeń potrzebują niezależnej walidacji na danych o rzeczywistej częstotliwości i szumie; mikrosakkady są szczególnie problematyczne.
- Brak danych surowych, wartości wariancji i wspólnych splitów ogranicza formalną ocenę wielkości efektu między publikacjami.

15. Wnioski końcowe

Dla estymacji mobilnej bez dodatkowego wyposażenia podstawowy kierunek stanowią modele appearance-based i ich odmiany temporalne. W warunkach zmiennej postawy istotne są RGB–D oraz mechanizmy IMU-triggered calibration. Dane porównawcze wskazują, że poprawa RMSE może być znaczna w obrębie określonego benchmarku, lecz na urządzeniu mobilnym kosztem większego czasu inferencji. Dlatego przy projektowaniu systemu należy optymalizować nie pojedynczą sieć, lecz cały proces pomiarowy. Najbardziej wartościowe przyszłe eksperymenty powinny łączyć odtwarzalną kalibrację, rzeczywiste scenariusze ruchu, miary błędu PoG, analizę jakości detekcji zdarzeń i pełny profil energetyczno-czasowy.

16. Zależności geometryczne i precyzyjna definicja błędu

16.1. Błąd PoG 2D

Niech $p_i = (x_i, y_i)$ oznacza położenie referencyjne punktu na ekranie, a $\hat{p}_i = (\hat{x}_i, \hat{y}_i)$ – wynik modelu po kalibracji, wyrażony w centymetrach. Dla pojedynczej próbki błąd euklidesowy wynosi $e_i = \sqrt{[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]}$. Średni błąd pozycyjny jest średnią e_i, natomiast RMSE powinno zostać zdefiniowane przez badacza, np. $\sqrt{[(1/N)\Sigma e_i^2]}$. Są to odmienne statystyki i mogą istotnie różnić się przy dużych błędach jednostkowych.

Dla ekranów o różnych rozmiarach współrzędne pikselowe należy przekształcić na jednostki fizyczne na podstawie rozdzielczości i fizycznej wielkości aktywnego obszaru. Należy uwzględnić orientację ekranu i obrót urządzenia. Procent błędu względem przekątnej ekranu może służyć uzupełniająco, ale nie powinien zastępować fizycznej miary PoG ani informacji o odległości wzroku.

16.2. Błąd kierunku 3D

Dla znormalizowanych wektorów kierunku spojrzenia u i û błąd kątowy można zdefiniować jako $\theta = \arccos[\text{clip}(u\cdot\hat{u}, -1, 1)]$, z wynikiem w stopniach. Jeżeli przyjmie się stałą odległość od ekranu d i uproszczoną geometrię małego kąta, wówczas przemieszczenie na ekranie można przybliżyć przez $d\cdot\tan(\theta)$. To przybliżenie nie uwzględnia wszystkich zmian orientacji i położenia głowy; dlatego raportowane błędy stopniowe i centymetrowe muszą pozostać odrębne, chyba że zastosowano tę samą geometrię przeliczenia i jawnie określono niepewność.

17. Jakość danych i synchronizacja w badaniach ruchomych

Dla danych mobilnych konieczne jest zbieranie znaczników monotonicznych dla kamery, IMU, wyświetlanego bodźca i wyniku modelu. Sama liczba klatek na sekundę nie informuje o regularności próbkowania; do obliczania prędkości spojrzenia powinien być wykorzystywany rzeczywisty odstęp czasowy między poprawnymi obserwacjami. Nieprawidłowe przyjęcie stałego interwału 1/30 s podczas przejściowych spadków FPS może spowodować sztuczne przekroczenie progu detektora sakkad.

Okna ze stratą detekcji twarzy należy oznaczać, a nie usuwać bez raportowania. Ważnymi wskaźnikami są: odsetek prawidłowo estymowanych próbek, liczba i długość przerw, odsetek klatek z zasłonięciem oczu, zmiany rozdzielczości wejścia oraz błędy po wznowieniu śledzenia. W badaniu z ruchem telefonu trzeba osobno rejestrować jego orientację, przyspieszenie i przybliżoną odległość od twarzy, jeśli sensor ją udostępnia.

18. Plan eksperymentów czynnikowych dla eye trackingu mobilnego

Czynnik	Poziomy do zweryfikowania	Możliwy wpływ
Postawa	Siedzenie, stanie, chodzenie	Ruch głowy i telefonu; dyspersja obserwacji.
Światło	Stałe / zmienne / boczne	Kontrast oka i jakość landmarków.
Urządzenie	Co najmniej dwie klasy kamery	Parametry optyki i akceleratora.
Kalibracja	Brak / jawna / ciągła	Błąd bazowy i dryf.
Modalność	RGB / RGB–D / RGB+IMU	Odporność na ruch i koszt przetwarzania.
Model	CNN / GRU / LSTM / hybryda	Jakość i latencja.
Kompresja	FP32 / INT8 / pruning	Jakość, RAM, FPS, energia.
Bodziec	Punkty stałe / ruchome / naturalna treść	Wpływ sakkad, fiksacji i kontekstu.

Projekt czynnikowy powinien obejmować kolejne sesje tej samej osoby, aby można było rozdzielić wariancję wewnątrzosobniczą i międzyosobniczą. Przy planowaniu wielkości próby należy użyć przewidywanej różnicy błędu i zmienności między uczestnikami, a nie mnożyć liczebność przez liczbę mocno skorelowanych klatek. Randomizacja kolejności warunków i kontrbalansowanie zadań ograniczają wpływ uczenia się kalibracji i zmęczenia.

19. Walidacja algorytmów detekcji zdarzeń

Ocena fiksacji i sakkad wymaga oddzielnego zbioru odniesienia anotowanego na podstawie szybszego trackera lub zweryfikowanych danych eksperymentalnych. Należy wyznaczyć tolerancję czasową dla początku i końca zdarzenia, ponieważ nawet zgodność co do liczby sakkad może maskować błędy lokalizacji granic. Dla predykcji I-VT podawać próg prędkości, sposób obliczania pochodnej i filtr wygładzający; dla I-DT – maksymalną dyspersję i minimalny czas trwania. Algorytmy probabilistyczne należy porównywać przy tych samych danych i z podobną liczbą hiperparametrów dostrajanych na walidacji.

Detekcja mrugnięć i mikrosakkad wymaga odrębnej analizy możliwości sensora. Wartości pochodzące ze strumienia 30 Hz nie odtwarzają zjawisk w zakresie czasowym obserwowanym przez kamery o wysokiej częstotliwości. Żaden model interpolacyjny nie zastępuje niezależnej informacji o przebiegu niezarejestrowanych mikrozdarzeń; ewentualne szacunki powinny być wyraźnie oznaczone jako inferencja, nie bezpośredni pomiar.

20. Specyfikacja raportowania kosztu czasu rzeczywistego

Etap	Pomiar	Sposób raportowania
Akwizycja	Czas ekspozycji i odbioru klatki	Średnia, P95, utrata klatek.
Detekcja twarzy	Czas detektora i odsetek trafień	P50/P95 i failure rate.
Normalizacja	Crop/resize/rectification	Czas CPU/GPU i rozdzielczość.
Model	Inferencja PoG lub gaze 3D	P50/P95/P99, pamięć i runtime.
Kalibracja	Aktualizacja parametrów	Czas, częstotliwość i odchylenie.

Filtrowanie	Usuwanie odstających próbek	Opóźnienie dodatkowe i jitter.
Interfejs	Publikacja punktu spojrzenia	End-to-end od kamery do ekranu.
Energia	Pobór podczas długiej sesji	Waty, bateria i profil termiczny.

Użyteczną metryką dla interfejsu jest czas od obserwacji fizycznego bodźca do aktualizacji widocznego obszaru lub odpowiedzi systemu, a nie tylko czas modelu. W sytuacji, gdy estymator działa z niejednakową częstotliwością, należy rozdzielić wiek aktualnej estymaty od czasu obsługi nowej klatki. Projekt powinien wprowadzić limit maksymalnego wieku PoG i oznaczać stan niepewności zamiast wyświetlać pozornie aktualny punkt po dłuższym zaniku sygnału.

21. Minimalny pakiet reprodukowalności

Dla każdego eksperymentu należy archiwizować: identyfikator wersji aplikacji i modelu, konfigurację detektora twarzy, definicję punktu referencyjnego, dane kalibracyjne, rozdzielczość i fizyczny rozmiar ekranu, wersję systemu operacyjnego, rodzaj akceleratora, sposób synchronizacji kamer i IMU oraz pełny skrypt ewaluacji. Dla publikacji porównawczych niezbędne są również zagregowane rozkłady błędu według osób i sytuacji, a nie jedynie jedna wartość średniej. Dopiero taka dokumentacja pozwala na wiarygodną analizę efektów, przeprowadzenie metaanalizy z wariancją wewnątrz- i międzybadaniową oraz ocenę rzeczywistego postępu technologicznego.

Bibliografia

[1] Lei Y, He S, Khamis M, Ye J. An End-to-End Review of Gaze Estimation and its Interactive Applications on Handheld Mobile Devices. ACM Computing Surveys. 2023. doi:10.1145/3606947.

[2] Krafka K, Khosla A, Kellnhofer P, et al. Eye Tracking for Everyone. Proceedings of CVPR. 2016.

[3] Valliappan N, Dai N, Steinberg E, et al. Accelerating eye movement research via accurate and affordable smartphone eye tracking. Nature Communications. 2020;11:4553.

[4] Arakawa R, Goel M, Harrison C, Ahuja K. RGBDGaze: Gaze Tracking on Smartphones with RGB and Depth Data. Proceedings of ICMI. 2022. doi:10.1145/3536221.3556568.

[5] Gunawardena N, Lui GY, Ginige JA, Javadi B. Smartphone-based Eye Tracking System using Edge Intelligence and Model Optimisation. Internet of Things. 2024/2025; arXiv:2408.12463.

[6] Khamis M, Alt F, Bulling A. The Past, Present, and Future of Gaze-enabled Handheld Mobile Devices: Survey and Lessons Learned. MobileHCI. 2018. doi:10.1145/3229434.3229452.

[7] Cheng Y, Wang H, Bao Y, Lu F. Appearance-based Gaze Estimation With Deep Learning: A Review and Benchmark. arXiv:2104.12668.

[8] Kar A, Corcoran P. A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms. IEEE Access. 2017. doi:10.1109/ACCESS.2017.2735633.

[9] Salvucci DD, Goldberg JH. Identifying Fixations and Saccades in Eye-Tracking Protocols. Proceedings of ETRA. 2000.

[10] Andersson R, Larsson L, Holmqvist K, Stridh M, Nyström M. One algorithm to rule them all? An evaluation and discussion of ten eye movement event detection algorithms. Behavior Research Methods. 2016. doi:10.3758/s13428-016-0738-9.

[11] Engbert R, Kliegl R. Microsaccades uncover the orientation of covert attention. Vision Research. 2003.

[12] Cheng S, Ping Q, Wang J, Chen Y. EasyGaze: Hybrid eye tracking approach for handheld mobile devices. Virtual Reality & Intelligent Hardware. 2022. doi:10.1016/j.vrih.2021.10.003.

[13] Lei Y, Zhao M, Wang Y, et al. MAC-Gaze: Motion-Aware Continual Calibration for Mobile Gaze Tracking. arXiv:2505.22769.

[14] Apple. ARFaceAnchor.lookAtPoint – Apple Developer Documentation.

[15] Tobii. Tobii Pro Glasses 3; Tobii Pro SDK Developer Documentation.

[16] Pupil Labs. Pupil Core open platform; Pupil Labs Neon Accuracy Test Report.

[17] Zhang X, Sugano Y, Bulling A. Evaluation of Appearance-Based Methods and Implications for Gaze-Based Applications (OpenGaze). arXiv:1901.10906.

[18] Huang J, White R, Buscher G. WebGazer: Scalable Webcam Eye Tracking Using User Interactions. IJCAI. 2016.

[19] TensorFlow. Post-training quantization; pruning in Keras. TensorFlow Model Optimization documentation.

[20] Gou J, Yu B, Maybank SJ, Tao D. Knowledge Distillation: A Survey. International Journal of Computer Vision. 2021.

[21] US20160282937A1. Gaze tracking for a mobile device. Google Patents.

[22] RGBDGaze: kod i dane https://github.com/FIGLAB/RGBDGaze ; Pupil Labs Pupil https://github.com/pupil-labs/pupil ; GazeML https://github.com/swook/GazeML.