

RAPORT ȘTIINȚIFIC FINAL

Proiect: Modelarea, Estimarea și Gestionarea Situațiilor Periculoase prin Analiza Continuă a Sistemului Sofer-Vehicul-Mediu (MEDALS)

Site web: <https://cv.utcluj.ro/medals>

1. Obiective prevăzute și realizarea acestora

Obiectivul general al acestui proiect este proiectarea unui sistem pentru estimarea continuă a potențialului de pericolozitate al unor situații din trafic, prin concentrarea nu doar pe starea exterioară a traficului, și nu doar pe starea șoferului, ci pe sinergia acestor factori. În scopul atingerii obiectivului general, s-au definit obiective concrete, a căror realizare este descrisă în continuare.

O1. Percepția din vehicul a contextului de pericol exterior

Acest obiectiv a fost abordat în toate cele trei etape ale proiectului, deoarece este esențial ca mediul exterior să fie perceput cât mai bine, cât mai rapid, și în cât mai multe moduri posibile, pentru redundanță.

În **Etapa 1** au fost proiectate și implementate soluții privind achiziția și adnotarea datelor: aplicații pe telefoane mobile, un sistem bazat pe camera termică IR, și sistem bazat pe o placă embedded (Jetson Nano). Fiecare soluție are meritele ei: soluția bazată pe telefoane mobile este ușor de instalat, soluția bazată pe IR are sensibilitate sporită în detecția pietonilor, care sunt una din cele mai importante surse de pericol în trafic, iar soluția embedded este cea care se va apropia cel mai mult de sistemul final care va putea procesa datele în timp real la bordul autovehiculului.

Achiziția folosind telefoane mobile are avantajul că poate utiliza atât camera video a telefonului, cât și diferiți senzori (senzori de măsurare a inerției și de poziționare prin satelit). Astfel, procesul de achiziție va utiliza camera, giroscopul, accelerometrul, senzorul de poziționare prin satelit (GPS sau GLONASS) și senzorul de orientare (magnetometru). Acest sistem este extrem de portabil și ușor de utilizat, fiind necesar un suport de mobil de parbriz pentru a amplasa dispozitivul cu camera principală orientată spre scena de trafic din față.

Senzorul termic folosit pentru capturarea secvențelor este un senzor furnizat de compania FLIR (PathFindIR), care conține un microbolometru VOX capabil să detecteze radiațiile cu lungimea de undă între 8-14 μm . Acest senzor oferă imagini cu o rezoluție de 320 × 240 pixeli, și este echipat cu o lentilă cu distanța focală de 19mm care conferă un câmp vizual de 36 grade pe orizontală și 27 grade pe verticală. Camera funcționează în orice condiții și este protejată de mediile în care există praf sau de ploaie cu o cutie închisă ermetic, ce nu permite elementelor nedorite să ajungă la senzor.

Un alt sistem de achiziție a fost bazat pe o placă Jetson Nano, echipată cu o cameră video, un senzor inerțial IMU și un receptor GPS. Aceste date vor fi salvate pe un drive sau pe un card micro SDXC. Un calculator portabil poate monitoriza datele conectându-se la sistem via o conexiune WiFi, cu un terminal de tip Secure Shell (SSH). Evident, datele pot fi procesate și direct pe placa de dezvoltare.

Tot în **Etapa 1** au fost studiate posibilitățile de utilizare a bazelor de date publice existente, pentru antrenarea sistemelor. Bazele de date publice sunt disponibile gratuit pentru cercetători, și au avantajul de a oferi obiecte adnotate, segmentate, cu informații despre distanțe sau alte date relevante. Bazele de date deja disponibile au fost create de grupuri de

cercetare diverse sau de companii private și de aceea au un format diferit al imaginilor, cu dimensiuni diferite și cu adnotări stocate în formate variate:

Baza de date CityScapes a fost realizată de Daimler AG și Universitatea Tehnică din Darmstadt, și conține secvențe de imagini din diverse orașe germane. Există 50 secvențe cu un total de 5000 de imagini etichetate, în scenarii de trafic variate și în momente diferite ale zilei.

KITTI este o bază de date dezvoltată de institutul de cercetare tehnologică din Karlsruhe, ce conține aproape 15.000 imagini, din 22 de secvențe achiziționate în condiții meteo similare, pe timp de zi, în orașul Karlsruhe din Germania.

Mapillary Vistas este una dintre cele mai nou apărute baze de date și oferă informații din scenarii de trafic variate, achiziționate de utilizatorii aplicației Mapillary. Baza de date oferă un total de 25.000 imagini din trafic adnotate.

Berkeley Deep Drive este o bază de date din cadrul grupului de cercetare în inteligență artificială al Universității Berkeley, ce conține 100.000 de secvențe de minim 40 de secunde din traficul rutier, având un total de 120 milioane de imagini. Pentru antrenarea rețelelor artificiale, există 10.000 de imagini adnotate cu informații relevante pentru segmentare semantică și de instanțe de obiect și 100.000 de obiecte identificate prin chenare de încadrare (en. "bounding box"). Aceasta este cea mai complexă bază de date în momentul de față.

JAAD [1] este un set de date pentru studierea atenției comune în contextul conducerii autonome. Accentul se pune pe comportamentul pietonilor și al șoferului în punctul de trecere a străzii și pe factorii care îi influențează.

În afara utilizării adnotărilor din bazele de date internaționale, s-a efectuat și adnotarea de imagini proprii. Pentru a ne asigura că algoritmi dezvoltați pentru partea de percepție a mediului sunt suficient de robuști am achiziționat secvențe în condiții variate ale vremii, și în diverse condiții de luminozitate și temperatură. Secvențele capturate conțin scene de zii, noapte, ploaie, soare, zapada, ceata etc). Pentru adnotarea secvențelor achiziționate cu camera termală s-a utilizat instrumentul de adnotare Computer Vision Annotation Tool CVAT. Adnotările pentru pietoni sunt prezente pentru 86 de secvențe de diverse lungimi surprinse în timpul nopții sau zilei în orașul Cluj-Napoca, România. Pentru a putea antrena modele de rețele neuronale capabile să clasifice participanții la trafic și nu numai, s-a realizat și o adnotare la nivel de pixel. Adnotarea a fost realizată cu ajutorul programului Adobe Photoshop 2021.

Datele achiziționate de noi vor oferi informații similare, dar avantajul este că avem control complet asupra procesului de achiziție, ceea ce înseamnă că putem lua date noi pentru a acoperi scenarii diferite: în momente diferite ale zilei, diferite anotimpuri, oraș, autostradă, etc.

Încă din prima etapă a proiectului am obținut rezultate semnificative în domeniul detecției obiectelor relevante din imagini. Pentru imaginile din spectrul vizual, monoculare, a fost propus un sistem ce combină o rețea neuronală convoluțională pentru segmentare semantică cu elemente de procesare bazată pe proprietățile geometrice ale obiectelor, pentru segmentarea instanțelor obiectelor din scenă. În acest fel s-au obținut rezultate comparabile cu rețele neuronale convoluționale dedicate ce oferă predicții ale instanței de obiect (en. "instance segmentation"), dar fără a avea complexitatea și dependența de

resurse a acestora. Rețeaua artificială propusă de noi va oferi patru ieșiri (predicții) despre obiecte folosind segmentarea semantică a imaginii de intrare. Cele patru ieșiri corespund celor patru zone care alcătuiesc obiectele individuale din imaginea de intrare, astfel fiecare pixel va fi clasificat ca fiind parte din stânga sus, dreapta sus, stânga jos, dreapta jos a regiunii obiectului. Pe aceste rezultate am aplicat o post-procesare pentru a grupa sferturile în obiecte individuale, similar cu un algoritm de grupare. Unele obiecte care sunt parțial blocate vor avea pixeli etichetați corect, chiar dacă sferturile generate de rețea nu sunt complet vizibile.

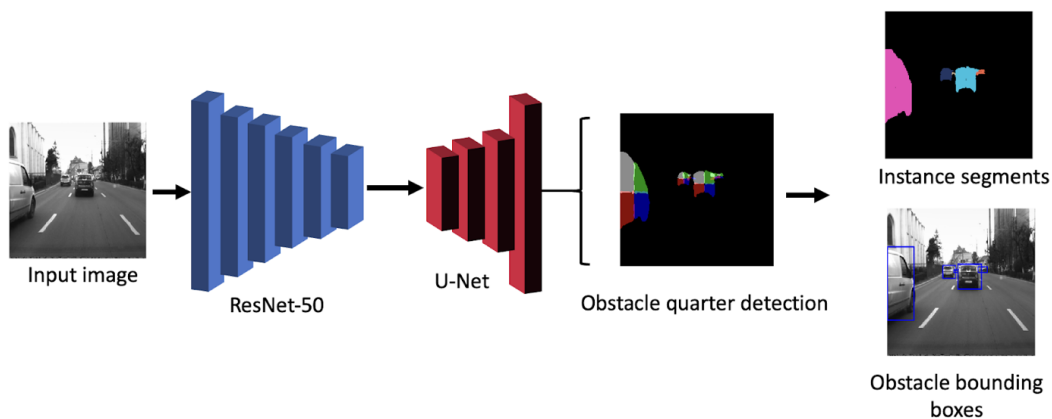


Figura 1. Identificarea obstacolelor combinând rețeaua neuronală cu post-procesarea geometrică

Una dintre problemele importante ale unui sistem de viziune bazat pe camere video este extragerea informației 3D, care nu este o informație intrinsecă acestui tip de senzor. Stereoviziunea ar fi o soluție, dar ea necesită o montare, o sincronizare și o calibrare precisă a camerelor, ceea ce limitează portabilitatea acestei soluții. Dacă utilizăm doar o singură cameră, pentru a infera cu succes distanța până la obiecte cea mai promițătoare soluție este utilizarea unei rețele neuronale. Arhitectura de bază a rețelei neuronale și modul de distribuire a straturilor sunt inspirate de o arhitectură des întâlnită în practică, VGG, la care se aduc mici modificări. VGG are la bază folosirea multor filtre de dimensiuni mici (3x3) și creșterea adâncimii per pixel în fiecare strat succesiv. Întreaga arhitectură poate fi asemănată cu o arhitectură de tipul encoder – decoder, unde partea de encoder reprezintă scăderea progresivă a înălțimii, a lungimii și creșterea adâncimii (downsampling) a hărților de trăsături, iar partea de decoder reprezintă operația inversă: creșterea înălțimii, a lungimii și descreșterea adâncimii (upsampling).



Figura 2. Disparități estimate din imagini monoculare.

A fost abordată de asemenea detecția și urmărirea pietonilor în imagini termale (IR – infraroșu). Aceste tipuri de imagini sunt valoroase pentru detecția pietonilor, deoarece aceștia au caracteristici termice specifice față de mediul înconjurător. Detectorul de pietoni se bazează pe o rețea de tip YOLO utilizată pentru determinarea poziției pietonilor în imagini infraroșu. Rețeaua a fost modificată și antrenată pe setul de imagini termale achiziționate cu

camera FLIR. În situațiile reale, detecțiile provenite de la detectorul de obiecte pot fii influențate de acuratețea detectorului, de probleme ale sensorului sau de ocluzii. Pentru îmbunătățirea performanței detectorului de obiecte, s-a creat un modul de asociere de date și urmărire. Pașii esențiali ai componentei asociere și de urmărire (tracking) sunt: Calculul scorului de similitudine și asocierea de date, Selectarea track-urilor, Actualizarea informațiilor track-urilor folosind datele provenite de la detectii, Rafinarea rezultatelor.

Soluția de tracking implementată este o metodă de urmărire prin detecție, unde mai întâi un detector de obiecte identifică obiectele relevante din scenă, iar trackerul urmărește elementele identificate.

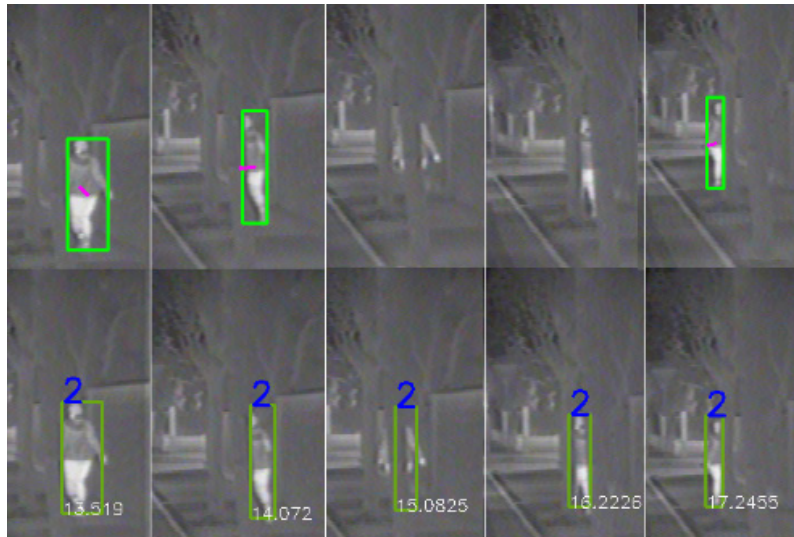


Figura 3. Urmărirea unui pieton în caz de ocluzie datorată unui copac.

În cadrul **Etapei 2**, a fost continuat procesul de adnotare în imaginile termale (IR – InfraRed), imagini în care nivelul de intensitate este proporțional cu temperatura obiectului reprezentat. Deoarece pietonii au o temperatură a corpului caracteristică, ei sunt în mod special vizați de analiza în domeniul infraroșu. În urma adnotării din etapa 1 a 1000 de imagini, s-a constatat că acest set nu este suficient, astfel că a fost adnotat încă un set de 1000 de imagini, ce acoperă multiple scenarii dificile în condiții de luminozitate diferită, temperatură crescută, și diverse obiecte, pietoni sau animale prezente pe carosabil. De asemenea, setul nou de date conține imagini termale din perspective diferite, necesare sistemelor de percepție a mediului în momentul în care se dorește execuția unor manevre variate, precum întoarcerile. Spre exemplu, în Figura 4 este ilustrat un scenariu achiziționat vara pe timp de noapte. Din cauza faptului că pietonul este mult mai cald decât carosabilul, se pot observa chiar și urmele deplasării acestuia pe drum.

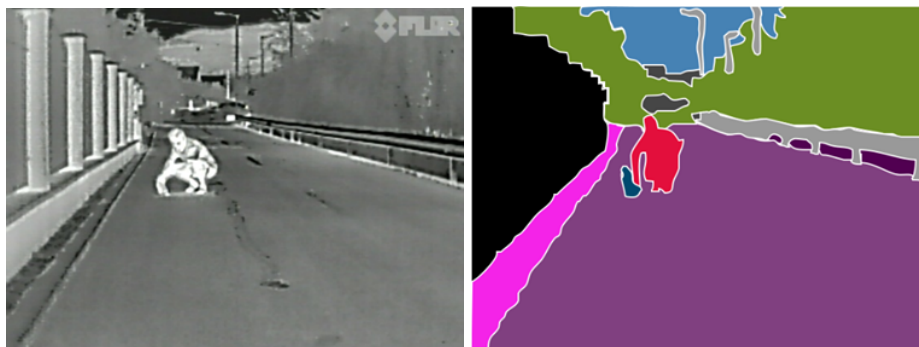


Figura 4. Persoană în mijlocul drumului, cu animalul de companie: imaginea termală și imaginea adnotată.

Tot în cadrul **etapei 2** a fost abordată o soluție pentru adnotarea automată a imaginilor folosind senzorii activi. Adnotarea imaginilor în mod manual implică un efort considerabil, mai ales dacă această adnotare se realizează la nivel de pixel, așa cum este cazul adnotării pentru segmentarea semantică sau pentru segmentarea panoptică. O soluție alternativă este utilizarea de senzori suplimentari, calibrați împreună cu camerele video, dar care pot oferi în mod direct informație 3D, informație care poate fi procesată în mod direct pentru a determina dacă zona observată, și implicit un anumit grup de pixeli, reprezintă o zonă de drum sau de obstacol. Senzorii LiDAR sunt capabili să estimeze distanțele către obiectele din jur cu o precizie ridicată. De asemenea, acești senzori sunt capabili să funcționeze atât ziua cât și noaptea, fără a avea pierdere de performanță. Senzorii LiDAR oferă informația 3D sub formă de nor de puncte, care trebuie procesate pentru a identifica obiectele individuale, dar și suprafața drumului.

Pentru detecția suprafeței drumului este utilizată o versiune paralelă a algoritmului RANSAC, care poate profita de arhitectura paralelă GPU. După detecția punctelor de drum, se va continua cu identificarea obstacolelor, prin gruparea punctelor 3D care nu aparțin drumului. Pentru gruparea punctelor 3D, folosim o abordare care se bazează pe densitatea norului de puncte și joacă un rol important în identificarea structurilor neliniare. Această abordare este cunoscută în mod obișnuit sub numele grupare spațială bazată pe densitate a aplicațiilor cu zgomot (density-based spatial clustering of applications with noise) sau DBSCAN. În figura 5 putem observa rezultatele grupării obiectelor împreună cu punctele de drum.

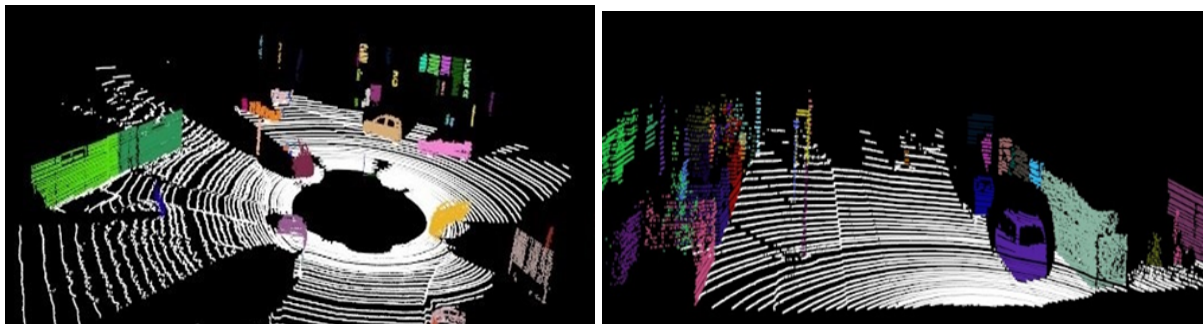


Figura 5. Obiectele identificate (culori diferite), și punctele de drum (alb) folosind LIDAR.

Pentru percepția mediului din imagini monoculare, tot în cadrul etapei 2 a fost realizat un sistem capabil să realizeze cel mai complex tip de segmentare, segmentarea panoptică, care combină segmentarea semantică cu identificarea instanțelor (deci cu identificarea fiecărui obiect în parte). Sistemul apoi aplică un algoritm de urmărire (tracking) atât pe cuboidele care conțin obiectele, cât și pe măștile (mulțimile de pixeli) ce aparțin obiectelor.

Sistemul utilizează trei modele de rețele neuronale pentru realizarea segmentării panoptice și pentru validarea acestora: ERFNet pentru segmentarea semantică, Yolact și YOLOv5 pentru identificarea instanțelor obiectelor. Pentru validarea detecțiilor se verifică dacă clasele semantice ale obiectelor YOLOv5 corespund cu clasa semantică a cel puțin unuia dintre cele două obiecte YOLACT. Dacă clasele corespund, obiectele sunt considerate valide. În caz contrar, se verifică clasa dominantă din chenarul de încadrare generat de YOLOv5, folosind imaginea de segmentare semantică.

După etapa de validare, toate obiectele identificate sunt urmărite pentru a menține identitatea lor în cadre succesive. Fiecare obiect identificat va avea ca atribute un cuboid (chenar), o mască de pixeli, și histograma claselor semantice în regiunea de interes definită de chenarul obiectului. Pe lângă aceste trăsături, se mai extrag alți descriptori, care vor fi utilizați în etapa de asociere de date. În Figura 6 sunt ilustrate rezultatele pe o scenă complexă, unde sunt multe autovehicule și pietoni.

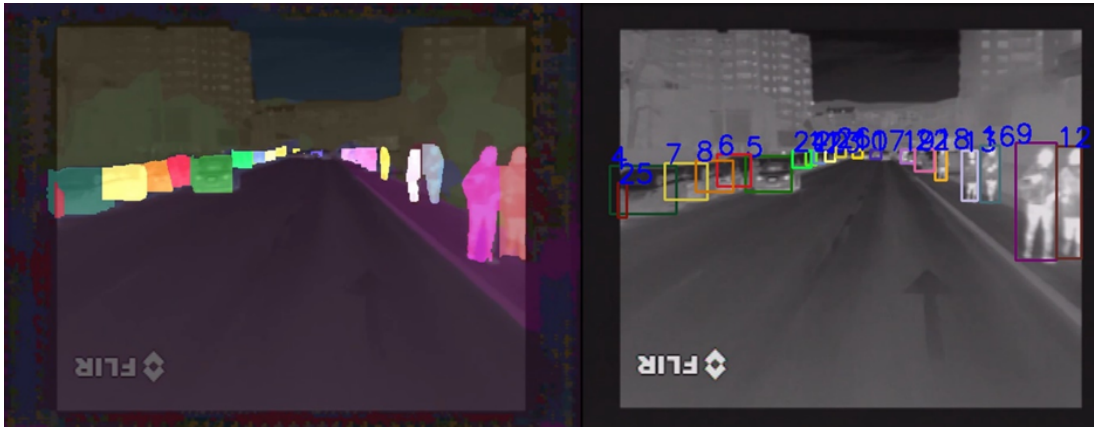


Figura 6. Rezultatele algoritmului pe o scenă mai complexă, cu multe obiecte.

În cadrul **Etapei 3** a fost continuat efortul privind detecția și urmărirea obiectelor folosind segmentarea semantică. Acum sistemul folosește trei abordări diferite bazate pe învățare pentru metoda de validare și o abordare bazată pe ingineria trăsăturilor și model matematic pentru partea de asociere și urmărire a datelor. Rezultatele abordărilor bazate pe date și pe modele sunt fuzionate pentru a obține rezultatele finale atât la nivel de mască cât și la nivel de cutie de încadrare a obiectelor. O diagramă care arată o imagine intuitivă a fluxului de procesare propus este prezentată în Figura 7. Abordările bazate pe învățare utilizate pentru validarea clasei semantice sunt o hartă de segmentare semantică obținută folosind modelul ERFNet care a fost modificat pentru a putea fi antrenat pe un set de date termale personalizat care conține 25 de clase semantice, un detector YoloV5 antrenat pe FLIR ADAS și un detector YOLACT, care produce măști de instanță și casete de încadrare, care a fost antrenat pe un set de date termale propriu.

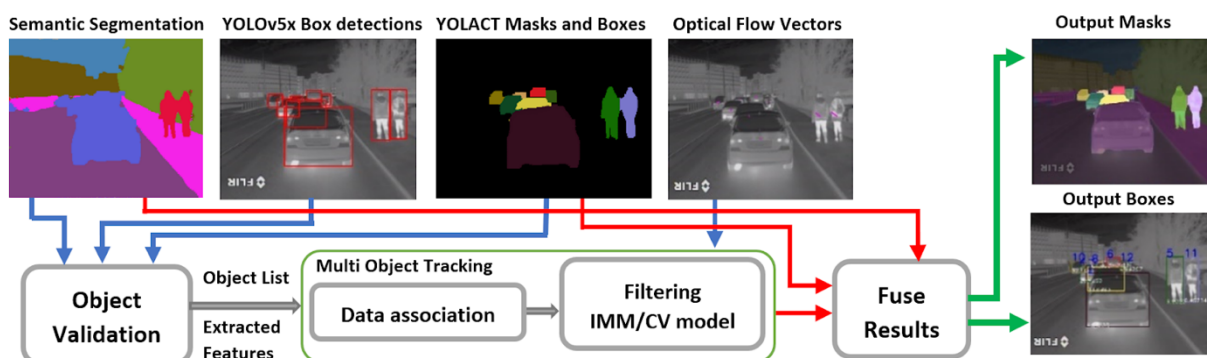


Figura 7. Fluxul de procesare al soluției de urmărire, segmentare și validare în contextul obiectelor multiple în domeniul termal.

O altă îmbunătățire față de algoritmul realizat în Etapa 2 este selectarea modelului de mișcare pe baza fluxului optic. Urmărirea obiectelor folosind filtrul Kalman implică modelarea mișcării acestora, iar modelele de mișcare pot fi simple, precum mișcarea cu viteză constantă, sau mai complexe, precum modelele interactive de mișcare ce includ

combinarea unui model de viteză constantă cu un model accelerat. Selecția modelului de mișcare adecvat este esențială pentru îmbunătățirea calității procesului de urmărire, iar una dintre sursele de informație care ne pot ghida spre modelul cel mai bun este fluxul optic – mișcarea aparentă a punctelor obiectului în spațiul imagine.

Chiar dacă traiectoriile individuale obținute din fluxul optic pot fi eronate, prin agregarea rezultatelor de la mai multe traiectorii în interiorul unei regiuni de interes putem obține indicii cu privire la mișcarea obiectelor în cadre succesive. Pentru a agrega datele din fluxul optic, stocăm valorile fluxului optic în 36 de acumulate, unde, folosind datele unghiului fluxului, fiecare vector de flux votează în interiorul unui acumulator. Valorile medii pentru magnitudinea și unghiul fluxului optic sunt calculate pentru fiecare acumulator după ce toți vectorii de flux optic din regiunea de interes au votat. În cele din urmă, mărimea medie și unghiul corespunzător casetei în care au fost exprimate majoritatea voturilor sunt selectate ca principali parametri de flux pentru regiunea de interes.

Soluția propusă a fost implementată în C++, iar modelele de rețele neuronale utilizate au fost antrenate în Pytorch și portate cu ajutorul framework-ului Libtorch în C++. OpenCV a fost folosit pentru desenarea, afișarea și fuzionarea rezultatelor într-o singură imagine. CUDA și OpenMP au fost folosite pentru a obține o performanță în timp real. Timpul total de rulare este de 100 ms și din punct de vedere calitativ segmentarea semantică are un IoU de 62,03%, YOLOv5 obține un scor mAP de 55,04%, Yolact are un scor de 34,6% mAP (31,2% mAP fără rafinare). Metoda de urmărire propusă a fost evaluată pe benchmark-ul online PTB-TIR și rezultatele pot fi observate în Tabelul 1.

Tabel 1. Evaluarea soluției de urmărire MOTSV pe benchmark-ul PTB-TIR

| | Scorul de succes urmărire | Scorul de precizie urmărire |
|-----------------------|---------------------------|-----------------------------|
| MDNet | 63.5% | 79.3% |
| Metoda propusă | 63.1% | 82.6% |
| VITAL | 62.2% | 81.0% |
| TADT | 58.8% | 73.4% |
| MLSSNet | 51.4% | 70.6% |

O2. Percepția din vehicul a condiției șoferului

În **Etapă 1** a fost explorată posibilitatea de a analiza starea șoferului pe baza timpului de reacție față de manevrele autovehiculelor din față. A fost experimentat un detector preexistent, capabil să ruleze în timp real pe placa Jetson, o arhitectură GPU paralelă dar totuși cu resurse limitate de calcul. Detectorul este o aplicație în C++, *detectnet*, ce poate fi configurată cu mai multe modele preantrenate care sunt disponibile, sau un model nou care trebuie să adere la anumite specificații. Modelul pe care l-am utilizat este modelul SSD-MobileNet-V2, care a fost lansat în 2019, și care suportă 91 de clase de obiecte diferite, incluzând autovehicule, biciclete, pietoni, și poate determina chenarele delimitatoare ale acestora în imagini. Având un detector pentru a estima în timp real starea obstacolelor din calea noastră, chiar dacă nu toate acestea sunt detectate, iar detecția poate fi ocazional eronată, se poate observa reacția șoferului față de acestea. Pentru estimarea reacției șoferului se monitorizează întregul ansamblu șofer/autovehicul folosind senzorul inerțial, IMU. Acesta ne poate transmite, în timp real, accelerațiile pe cele trei axe, precum și rotațiile autovehiculului, deci putem identifica în mod corect dacă autovehiculul nostru execută o

manevră de accelerare, una de frânare, sau dacă execută un viraj. Vom putea corela manevrele proprii cu manevrele autovehiculului din față, și putem estima care este timpul de reacție tipic, și dacă reacția este mai rapidă ca de obicei (indicând o stare de agitație), sau dacă este mai lentă, indicând oboseală sau neatenție.

În aceeași etapă a fost experimentată o altă neinvazivă pentru estimarea stării șoferului, monitorizarea distribuției presiunii asupra scaunului. Un șofer agitat va avea mișcări dese, cu variații mari ale presiunii, iar un șofer obosit se va mișca mai lent. O postură neobișnuită a șoferului, cu presiunea distribuită inegal, poate indica o durere sau un alt fel de posibilă incapacitate. Pentru a monitoriza aceste situații am proiectat un dispozitiv care va monitoriza presiunea în nouă puncte, distribuite într-o matrice de 3x3 celule. Acest dispozitiv de tip pernă a fost gândit cu intenția de a fi plasat pe scaunul șoferului. Elementul central al dispozitivului este folia conductivă cu rezistența variabilă la presiune 3M Velostat. Atunci când se aplică o forță asupra foliei, valoarea rezistenței electrice în acel punct scade. Pentru realizarea sistemului am suprapus două folii Velostat și am adăugat pe fiecare față câte trei benzi conductoare, organizate în rânduri și coloane (benzile de deasupra sunt perpendiculare pe cele de dedesubt). Acest ansamblu a fost cusut între două straturi de cauciuc, pentru a le ține rigide și pentru a proteja integritatea sistemului în momentul operării.

Din păcate, metodele experimentate în Etapa 1 nu au dus la rezultatele preconizate. Metoda bazată pe timpul de reacție a fost limitată de necesitatea de antrenare în situații de trafic real, unde chiar și simularea unei reacții inadecvate duce la punerea în pericol a șoferului și a celorlalți participanți la trafic, iar metoda bazată pe determinarea presiunii pe scaunul șoferului a fost limitată de dese erori electrice și mecanice. Din acest motiv, începând cu etapa 2 ne-am concentrat eforturile spre determinarea stării șoferului folosind secvențe de imagini capturate cu camere orientate spre șofer, plasate în interiorul autovehiculului.

În **Etapa 2** au fost studiate și implementate metode de recunoaștere a orientării feței șoferului, metode de recunoașterea stării de somnolență a șoferului, cât și modele de recunoaștere a stării de ebrietate. Toate modelele utilizează doar imagini capturate cu o cameră video color montată în interiorul autovehiculului, orientată spre fața șoferului.

Acțiunile șoferului au fost clasificate în patru categorii, în funcție de orientarea feței, și anume: privind la drum, privind în dreapta sau în stânga, și privind înapoi. Pentru a determina orientarea actuală a feței, a fost utilizată o implementare BiSeNet care a fost antrenată cu setul de date CelebAMask-HQ. Analizând patru trăsături faciale principale, inclusiv ochii, nasul, gura, urechile, este posibil să se estimeze orientarea feței șoferului.

Starea de somnolență este analizată prin prisma acțiunii ochilor șoferului și a cășcatului. Primul pas constă în detecția fețelor persoanelor care sunt în cadru folosind un detector al bibliotecii *dlib*. În regiunea de interes în care este detectată fața se extrage un șablon de repere pentru punctele cheie ale feței. O parte din punctele cheie sunt utilizate pentru extragerea conturului ochilor. Șablonul este definit de 68 de puncte cheie pe conturul feței, conturul ochilor, al gurii și al nasului, după cum se observă în figura 8, unde persoana este cu ochii deschiși și apoi cu ochii închiși. O parte din punctele din șablon sunt utilizate pentru a determina starea ochilor deschiși, respectiv închiși. Pentru aceasta se calculează factorul de aspect al fiecărui ochi. În urma stabilirii stării ochilor individului, în caz de ochi închiși se va porni un cronometru ce are menirea de a monitoriza durata acestei stări. Cronometrul se resetează la detectarea stării de ochi deschiși.

Starea șoferului este monitorizată continuu și dacă se colectează un minim de cinci stări cu ochii închiși se rulează un algoritm de predicție de tipul Long Short Term Memory (LSTM), care are rolul de a prezice cât timp va sta persoana cu ochii închiși la următoarea acțiune de clipire.

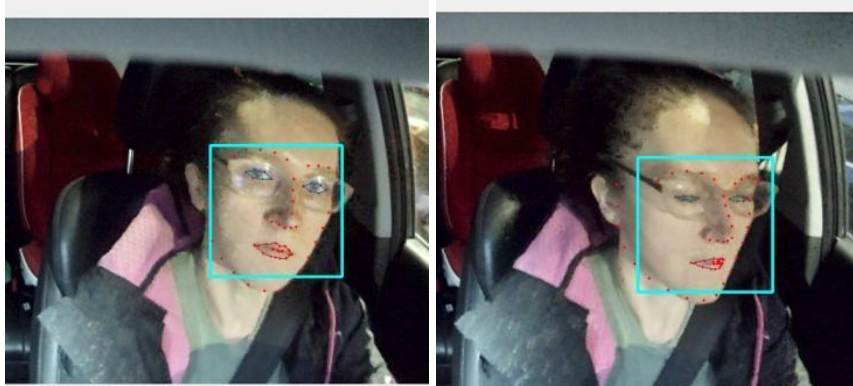


Figura 8: Șoferul clipind. În imaginea din stânga, ochii sunt deschiși, iar în dreapta ochii sunt închiși.

În aceeași etapă a fost abordată și recunoașterea stării de ebrietate, care este un factor extrem de important în riscul de accident rutier. Au fost studiate, implementate și comparate mai multe metode de recunoaștere a stării de ebrietate din imagini monoculare color. Datorită lipsei unui set de date de dimensiune mare pentru astfel de imagini, dezvoltarea s-a axat pe două direcții: modele de tip CNN (Convolutional Neural Networks) și modele convenționale de învățare bazate pe trăsături.

Pentru realizarea setului de date de antrenare și testare s-a pornit de o bază de date publică, ce conține un număr de doar 212 imagini, dintre care 75% dintre acestea aveau persoane în stare de ebrietate. Deoarece acest set de date a fost insuficient pentru antrenarea algoritmilor de învățare profundă, s-au colectat mai multe imagini digitale din mediul online în care au fost surprinși indivizi care se încadrează în cele două clase de interes. Pentru fiecare clasă s-au colectat imagini în care erau prezenți unul sau mai mulți indivizi, fie cu simptome de ebrietate, fie fără. S-a obținut în final un set de imagini cu 1300 de persoane în stare de ebrietate și 1010 cu persoane care nu sunt sub influența alcoolului. Acest set de date a fost augmentat pentru antrenarea modelelor de învățare profundă. Setul a fost împărțit în date de antrenare, testare și validare, în proporție de 70%-15%-15%. Modelele de tip CNN considerate au fost ResNet50, InceptionNetV3 și EfficientNetB0. Structura lor a fost adaptată pentru a clasifica imagini de cele două tipuri (ebrietate și non-ebrietate). Modelele au fost antrenate pentru cel puțin 100 de epoci. S-au realizat augmentări ale datelor (de tipul rotații, translații, oglindire orizontală). Rezultatele cele mai bune din punctul de vedere al metricii de acuratețe sunt prezentate în tabelul 2.

Tabel 2. Performanța diferitelor modele de CNN pentru determinarea stării de ebrietate

| Model | Acuratețe |
|----------------|-----------|
| ResNet50 | 74 % |
| InceptionNetV3 | 75 % |
| EfficientNetB0 | 72 % |

Au fost abordate, în vederea determinării stării de ebrietate, și metode non-neuronale, bazate pe trăsături extrase din imagini și clasificatori convenționali ("clasici"). Pentru

extragerea trăsăturilor a fost utilizată biblioteca Dlib, capabilă să detecteze fața și să extragă puncte cheie din ale acesteia.

Având în vedere faptul că persoanele în stare de ebrietate prezintă roșeață în jurul obrazilor, nasului și frunții din cauza circulației sângelui afectată de consumul de alcool, din aceste zone a fost extrasă media de culoare a canalului care indică culoarea roșie. În zonele obrazilor și a frunții se calculează media acestui canal, iar în zona nasului este extrasă valoarea canalului întâi din imaginea digitală RGB din cadrul punctelor identificate. Astfel din aceste zone sunt extrase 5 caracteristici. Celelalte trăsături sunt obținute din zona ochilor și a cavității bucale. Pentru a nu extrage o mulțime de trăsături care ar putea afecta procesul de antrenare a modelului de învățare, în zona ochilor s-au prelucrat distanțe în funcție de coordonatele primite, pentru a observa starea ochilor dacă aceștia sunt închiși sau deschiși, astfel că pentru fiecare ochi s-au extras 4 puncte la extremitățile ochiului și s-a calculat distanța normalizată dintre ele, rezultând două trăsături pentru fiecare ochi. Pentru zona cavității bucale s-au considerat punctele de interes și distanța dintre ele. Toate trăsăturile descrise mai sus au fost folosite ca date de intrare pentru algoritmi de clasificare clasici bazați pe trăsături: RandomForest, SupportVector Machine, Decision Tree. Acuratețea obținută este prezentată în tabelul 3.

Tabel 3. Performanța diferiților clasificatori convenționali în detectarea stării de ebrietate.

| Model | Acuratețe |
|------------------------|-----------|
| Random Forest | 80 % |
| Support Vector Machine | 85 % |
| Decision tree | 78 % |

Performanța cea mai bună este obținută cu un clasificator de tip SVM. Modelul antrenat poate fi integrat în faza următoare a proiectului în modulul de răspuns la pericol.

În **Etapă 3** a fost continuat efortul privind determinarea gradului de atenție al șoferului. Cercetarea s-a axat în principal pe utilizarea rețelelor neuronale convoluționale pentru a determina elementele care pot indica un grad de neatenție din partea șoferului, și pe combinarea acestor elemente pentru a produce un scor de atenție. Deoarece abordările bazate pe rețele neuronale necesită date pentru antrenare, s-a depus efort și în direcția adnotării acestor date. Pornind de la mai multe seturi de date publice utilizate la monitorizarea comportamentului șoferilor, s-a extins adnotarea acestor baze de date cu gradul de atenție. În contextul proiectului, am creat propriile adnotări ale gradului de atenție al șoferilor pentru un set de date numit "DMD – Driver Monitoring Dataset". Înregistrările din acest set de date au fost realizate din trei perspective diferite cu trei camere poziționate strategic pentru a surprinde fața, mâinile și corpul șoferului. Fiecare cameră prezintă trei canale care includ informații color, infraroșu și de adâncime. La realizarea videoclipurilor au participat 37 de voluntari și au fost înregistrate 40 de ore și 45 de minute de material video. 10 dintre participanți purtau ochelari, unii au fost înregistrați cu și fără a-i purta. Unele activități au fost înregistrate cu mașina în mișcare, cu mașina oprită pe loc și într-un simulator. Setul de date a fost planificat să conțină 13 activități legate de conducerea distrasă și anume: conducere în siguranță, scriere de mesaje, convorbire telefonică, utilizarea radioului mașinii, băutul de suc/apa, vorbitul cu alți pasageri, pieptănatul. Activitățile au fost apoi adnotate la intervale de cadru, limitele fiecărui cadru fiind definite, ceea ce înseamnă că începutul și sfârșitul fiecărei activități sunt stabilite pe baza unor criterii clare. Acest set a fost augmentat cu gradul de atenție al șoferului. În cadrul proiectului am

considerat doua tipuri de imagini: a) imagini înregistrate doar cu camera care surprinde volanul și mâinile șoferului și b) imagini înregistrate cu camera care surprinde fața șoferului. Imaginile au fost sincronizate pe bază de timestamp (informație care există în setul DMD).

Setul de date original conține aproximativ 150 de videoclipuri pentru fiecare dintre aceste camere, iar în cadrul proiectului am selectat 50 dintre ele, fiecare fiind diferite în ceea ce privește condițiile de filmare, șoferul și obiectele prezente. Peste 8.000 de imagini au fost adnotate pentru fiecare set de date, iar datele au fost împărțite folosind următoarea abordare: 70% pentru antrenare, 20% pentru validare și 10% pentru testare. Pentru imaginile care surprind mainile șoferului și volanul am stabilit adnotari ale următoarelor clase: ținerea ambelor mâini pe volan (hand on wheel), ținerea unei singure mâini pe volan(hand not on wheel), o mână este implicată în utilizarea telefonului(hand on phone), o mână este utilizată pentru a ține un obiect (o sticlă, un pieptăn). Aceste etichete au fost alese deoarece sunt cele mai importante pentru a identifica gradul de implicare al șoferului în gestionarea volanului și pentru a identifica comportamentele care pot afecta nivelul de atenție și siguranța la volan. Pentru imaginile care surprind fata șoferului etichetele claselor stabilite sunt legate de privirea șoferului și atenția acestuia asupra drumului. Acestea includ următoarele clase: privirea îndreptată spre drum(looking at the road), privirea nu este îndreptată spre drum sau în oglinzi (not looking at the road), prezența unor alte obiecte (object), prezența telefonului (phone).

Gradul de atenție al șoferului este adnotat ca un număr real în intervalul 0 ... 1, unde 1 înseamnă că șoferul este extrem de atent, în timp ce 0 înseamnă că șoferul este foarte distras. Procesul de adnotare a fost realizat în doua etape: a) etapa automata și b) etapa manuala. În etapa automată se inferează fiecărei imagini dintr-o secvență un grad de atenția în mod automat, pe baza scorului de încredere returnat de un detector de obiecte de tip YOLO. Pentru imaginile de la camera care surprinde mâinile scorul este stabilit conform algoritmului din figura 9 (coloana stângă): astfel, dacă se detectează un obiect (telefon, sticlă) sau faptul ca mâinile șoferului nu sunt pe volan, atunci scorul de atenție este mic. Dacă ambele mâini sunt pe volan scorul de atenție este mare. Pentru imaginile provenite de la camera care captează fața șoferului, scorul de atenție este stabilit conform algoritmului din figura 9, coloana dreaptă. Dacă șoferul se uita la drum atunci gradul de atenție este mare, dacă șoferul nu se uita la drum (are capul întors in lateral sau privește in jos) sau dacă apare un obiect atunci gradul de atenție este mai mic.

Data: Object detection score O_s ,
hands category id: hc
 $hc = 0 \rightarrow$ One Hand On Wheel;
 $hc = 1 \rightarrow$ Hand Not On Wheel;
 $hc = 2 \rightarrow$ Hand On Bottle;
 $hc = 3 \rightarrow$ Hands On Wheels;
 $hc = 4 \rightarrow$ Hand On Phone;
 $hc = 5 \rightarrow$ Hand On Hair Comb.
Result: Attention score: A_s
 $A_s \leftarrow -1$;
if $O_s > 0.7$ **then** $p \leftarrow 0$;
if $O_s \in [0.3, 0.7]$ **then** $p \leftarrow 0.1$;
if $O_s < 0.3$ **then** $p \leftarrow 0.2$;
if $hc == 0$ **then** $A_s = 0.8 - p$;
if $hc \in [1, 2, 5]$ **then** $A_s = 0.6 - p$;
if $hc == 3$ **then** $A_s = 1 - p$;
if $hc == 4$ **then** $A_s = 0.1 + p$;
if $hc == -1$ **then** $A_s = 0.5$;

Data: Object detection score O_s ,
face category id: fc
 $fc = 0 \rightarrow$ Looking at the road;
 $fc = 1 \rightarrow$ Not looking at the road;
 $fc = 2 \rightarrow$ object is detected;
 $fc = 3 \rightarrow$ phone is detected
Result: Attention score: A_s
 $A_s \leftarrow -1$;
if $O_s > 0.7$ **then** $p \leftarrow 0$;
if $O_s \in [0.3, 0.7]$ **then** $p \leftarrow 0.2$;
if $O_s < 0.3$ **then** $p \leftarrow 0.2$;
if $fc == 0$ **then** $A_s = 1.0 - p$;
if $fc == 1$ **then** $A_s = 0.1 + p$;
if $fc == 2$ **then** $A_s = 0.4 + p$;
if $fc == 3$ **then** $A_s = 0.2 + p$;
if $fc == -1$ **then** $A_s = 0.5$;

Figura 9. Stabilirea gradului de atenție pe baza indiciilor de distractori observați din camera video.

În faza manuală de adnotare sunt analizate scorurile inferate automat în etapa anterioară de adnotare a gradului de atenție. Astfel utilizatorii pot să crească sau să scadă gradul de atenție în funcție de percepția proprie și în funcție de istoricul acțiunilor (dacă șoferul vorbește de mult timp la telefon gradul de atenție este foarte mic, dacă șoferul privește scurt în stânga sau în dreapta gradul de atenție este mai mare).

Sistemul propus pentru prezicerea gradului de atenție al șoferului este prezentat în figura 10. Modulele principale sunt următoarele:

1. Modulul de procesare al informațiilor de la camera frontală. Acest modul utilizează algoritmi de detecție a orientării feței (șoferul privește sau nu la drum) și de detecție a obiectelor bazat pe arhitectura Yolo.
2. Modulul de procesare al informațiilor de la camera care vede mâinile și volanul. Acesta este bazat tot pe un detector de tip Yolo antrenat pentru a recunoaște mâinile și volanul, precum și obiecte precum telefon, sticla de apă / suc.
3. Modulul de fuziune a informațiilor și de extragere a trăsăturilor are ca rol construirea unui vector de trăsături pe baza cărui se va face predicția gradului de atenție. Vectorul de trăsături conține poziția obiectelor detectate și scorul de încredere al acestora.
4. Modulul de predicție a scorului de atenție folosește o arhitectură de tip LSTM (Long Short Term Memory). Au fost construite trei modele diferite: un model este antrenat utilizând doar trăsăturile de la camera frontală, al doilea model este antrenat doar pe trăsăturile calculate pe imaginile de la camera care vede volanul, și al treilea model este antrenat pe trăsături calculate pe imaginile de la ambele camere.

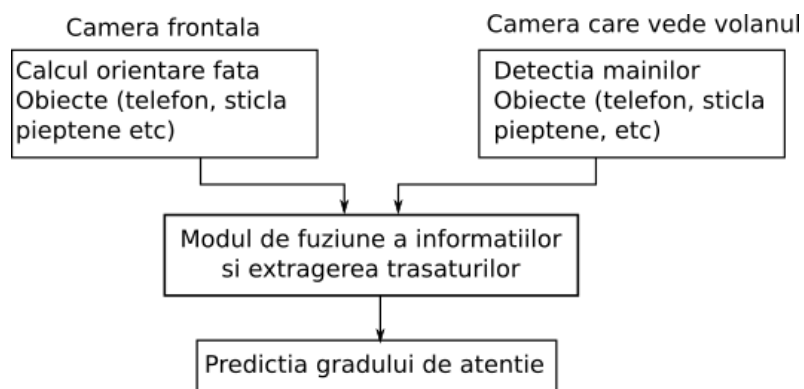


Figura 10. Modulele utilizate pentru predicția gradului de atenție al șoferului.

Arhitectura LSTM folosită de modulul de predicție a scorului de atenție conține două straturi recurente (recurrent layers) care au fiecare câte 16 unități ascunse, urmate de un strat liniar care are o singură ieșire și este folosit pentru a prezice un număr real din intervalul $[0,1]$ care reprezintă gradul de atenție. Fiecare din cele trei modele de tip LSTM au fost antrenate pentru 30 de epoci, folosind un batch size de 256, o funcție de loss de tip eroare medie patrată și un optimizator de tip Adam.

Sistemul a fost testat atât pentru performanța detecției individuale a obiectelor distractorie, unde a fost obținută o precizie mare pentru majoritatea claselor care surprind poziția mâinii în raport cu volanul și orientarea feței în raport cu drumul. O precizie mai mică este

înregistrată pentru detecția situațiilor când mana este pe un pieptene, datorită numărului mic de imagini de acest fel din secvențe.

Rezultatul modelelor de predicție a gradului de atenție sunt prezentate în Tabelul 4. Se poate observa din tabel că toate cele 3 modele au o eroare medie pătratică foarte mică atunci când prezic gradul de atenție pentru cadrul următor (decalaj/offset = 1). Din tabelul 4 se poate observa o performanță bună pentru un decalaj mic al cadrelor și cu cât secvența considerată este mai lungă cu atât predicția este mai bună.

Tabel 4. Metrice de evaluare a modelelor de predicție a gradului de atenție pentru diferite lungimi de secvență (sequence length) și deplasament al cadrelor (frame offset).

| (Frame offset Sequence Length) | R^2 hands | MAE hands | MSE hands | R^2 face | MAE face | MSE face | R^2 combined | MAE combined | MSE combined |
|-----------------------------------|----------------|--------------|--------------|---------------|-------------|-------------|-------------------|-----------------|-----------------|
| 1, 1 | 0.592 | 0.098 | 0.066 | 0.866 | 0.037 | 0.008 | 0.761 | 0.055 | 0.015 |
| 5, 1 | 0.386 | 0.130 | 0.100 | 0.545 | 0.086 | 0.028 | 0.571 | 0.086 | 0.027 |
| 10, 1 | 0.345 | 0.149 | 0.107 | 0.263 | 0.125 | 0.046 | 0.417 | 0.110 | 0.037 |
| 15, 1 | 0.266 | 0.164 | 0.119 | 0.151 | 0.143 | 0.053 | 0.330 | 0.122 | 0.042 |
| 20, 1 | 0.216 | 0.174 | 0.128 | 0.128 | 0.149 | 0.054 | 0.291 | 0.129 | 0.045 |
| 25, 1 | 0.179 | 0.182 | 0.134 | 0.122 | 0.151 | 0.054 | 0.260 | 0.133 | 0.047 |
| 30, 1 | 0.145 | 0.188 | 0.139 | 0.116 | 0.152 | 0.055 | 0.232 | 0.136 | 0.049 |
| 1, 5 | 0.628 | 0.097 | 0.061 | 0.857 | 0.039 | 0.009 | 0.818 | 0.050 | 0.012 |
| 5, 5 | 0.523 | 0.123 | 0.078 | 0.539 | 0.087 | 0.029 | 0.616 | 0.084 | 0.024 |
| 10, 5 | 0.379 | 0.146 | 0.101 | 0.260 | 0.125 | 0.046 | 0.439 | 0.109 | 0.036 |
| 15, 5 | 0.280 | 0.161 | 0.117 | 0.153 | 0.144 | 0.053 | 0.361 | 0.121 | 0.040 |
| 20, 5 | 0.223 | 0.172 | 0.126 | 0.126 | 0.150 | 0.054 | 0.309 | 0.129 | 0.044 |
| 25, 5 | 0.191 | 0.180 | 0.132 | 0.116 | 0.152 | 0.055 | 0.266 | 0.133 | 0.047 |
| 30, 5 | 0.166 | 0.187 | 0.136 | 0.115 | 0.153 | 0.055 | 0.237 | 0.137 | 0.048 |
| 1, 10 | 0.649 | 0.095 | 0.057 | 0.861 | 0.038 | 0.009 | 0.825 | 0.049 | 0.011 |
| 5, 10 | 0.509 | 0.124 | 0.080 | 0.552 | 0.083 | 0.028 | 0.628 | 0.083 | 0.024 |
| 10, 10 | 0.370 | 0.146 | 0.102 | 0.287 | 0.122 | 0.044 | 0.461 | 0.108 | 0.034 |
| 15, 10 | 0.293 | 0.161 | 0.115 | 0.153 | 0.144 | 0.053 | 0.366 | 0.122 | 0.040 |
| 20, 10 | 0.248 | 0.172 | 0.122 | 0.128 | 0.149 | 0.054 | 0.311 | 0.129 | 0.044 |
| 25, 10 | 0.188 | 0.179 | 0.132 | 0.112 | 0.152 | 0.055 | 0.270 | 0.133 | 0.046 |
| 30, 10 | 0.167 | 0.187 | 0.136 | 0.115 | 0.153 | 0.055 | 0.239 | 0.137 | 0.048 |
| 1, 20 | 0.64 | 0.095 | 0.059 | 0.871 | 0.036 | 0.008 | 0.760 | 0.057 | 0.015 |
| 5, 20 | 0.487 | 0.126 | 0.084 | 0.564 | 0.083 | 0.027 | 0.579 | 0.087 | 0.027 |
| 10, 20 | 0.381 | 0.146 | 0.101 | 0.284 | 0.124 | 0.044 | 0.425 | 0.111 | 0.036 |
| 15, 20 | 0.299 | 0.161 | 0.114 | 0.149 | 0.144 | 0.053 | 0.324 | 0.122 | 0.041 |
| 20, 20 | 0.23 | 0.171 | 0.125 | 0.120 | 0.150 | 0.055 | 0.293 | 0.130 | 0.045 |
| 25, 20 | 0.198 | 0.18 | 0.131 | 0.118 | 0.151 | 0.055 | 0.271 | 0.134 | 0.047 |
| 30, 20 | 0.173 | 0.187 | 0.135 | 0.131 | 0.150 | 0.054 | 0.238 | 0.136 | 0.048 |

Din predicția gradului de atenție se poate extrage și evoluția în timp a nivelului de distragere al șoferului de la un mod de condus atent. În figura 11 se pot observa grafice ale evoluției gradului de atenție în timp. În partea superioară a figurii observăm un profil temporal cu atenție medie spre ridicată, iar în partea inferioară observăm un profil de condus cu atenție mică spre medie, cu multe variații ale gradului de atenție în timp.

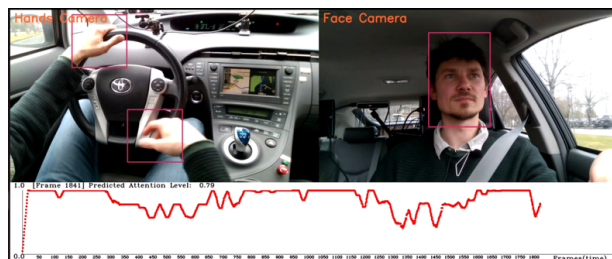




Figura 11. Evoluția temporală a gradului de atenție pentru diferite secvențe.

O3. Analiza stării vehiculului

Acest obiectiv este un obiectiv tehnic, ce oferă suport pentru celelalte obiective, de percepție și de analiză a răspunsului la pericol. Starea vehiculului este importantă în procesul de urmărire, pentru a putea prezice poziția următoare a obiectelor, și în procesul de estimare a necesității de a acționa pentru prevenirea unui accident, în cazul în care viteza noastră nu este adaptată la situația observată.

Pentru perceperea stării autovehiculului au fost experimentate două abordări independente. O abordare implică utilizarea de senzori externi, precum senzorul inerțial, care va determina accelerația sau rata de rotație (yaw rate), sau receptorul GPS care va determina viteza.

Datele citite de la senzorul IMU ICM-20948 ajung la CPU prin două canale I2C diferite, un canal trimițând date legate de giroscop și accelerometru, iar celălalt date de la busola digitală (magnetometru). Analizând aceste valori putem identifica manevrele bruște, cum ar fi rotirea rapidă a volanului, și pot fi corelate cu imaginile interne/exterioare pentru determinarea unei surse de pericol.

Modulul receptor GPS NEO-8M trimite mai multe mesaje prin UART, raportând totodată timpul precis sincronizat cu timpul global UTC. Diferitele propoziții NMEA sunt analizate, iar viteza curentă în km/h este citită din mesajul GPVTG.

Dezavantajul acestor senzori este că ei vor da aceste informații cu precizie redusă, sau, în cazul GPS, viteza va fi o viteză medie, cu o inerție a măsurătorii semnificativă. Avantajul lor este că sunt independenți de tipul de autovehicul utilizat, și sunt portabili, putând fi montați ușor pe orice vehicul, oricând.

O abordare mai invazivă este citirea stării parametrilor autovehiculului, precum viteza, accelerația, unghiul volanului, poziția pedalei de accelerație sau a pedalei de frână, de pe magistrala CAN (Control Area Network). Accesul la CAN se poate face prin portul OBD, portul de diagnostic care se găsește pe aproape orice vehicul modern. Acest port poate furniza, în timp real, date despre viteză, despre turația motorului, sau despre starea unor indicatori de avarie ai autovehiculului. Pentru citirea acestor informații se utilizează un adaptor de la OBD la interfața serială UART, apoi datele preluate din interfața serială vor fi procesate de un microcontroller Arduino, care apoi le va transmite către dispozitivele noastre mobile folosind o interfață Bluetooth de tip HC-05.

Deși datele preluate de pe magistrala CAN sunt precise și sunt lipsite de întârziere, în final s-a ales avantajul portabilității și a independenței de tipul de vehicul. Astfel, am ales să ne concentrăm pe utilizarea senzorilor inerțiali și GPS, care pot fi transportați ușor dintr-un vehicul în altul, putând utiliza mai multe vehicule fără riscul de a le afecta funcționarea prin intervenții pe magistrala CAN.

O4. Proiectarea unui model de pericol multidimensional

Acest obiectiv complex se bazează pe rezultatele percepției (interioare și exterioare autovehiculului), dar și a stării acestuia. În primele etape ale proiectului, acest model a fost intenționat a fi proiectat prin analiza individuală a diferitelor elemente din trafic. În **etapa 1** am dedicat acest efort estimării intenției pietonilor de a traversa prin fața autovehiculului. Postura pietonilor în secvențe de imagini termale sau color a fost extrasă pe baza unui algoritm de estimare regională de poziție pentru mai multe persoane, AlphaPose. Extragerea trasaturilor scheletului se realizează pe detecțiile furnizate de detectorul de pietoni. Aceste detecții sunt analizate de o rețea de transformări spațiale și simetrice și un estimator de poziție, care furnizează punctele cheie ale scheletului.

AlphaPose pune la dispoziție mai multe modele pre-antrenate pe mai multe seturi de date, precum COCO și Halpe. Pentru modelele antrenate pe setul de date COCO se obțin 17 puncte cheie corespunzătoare părților corpului, precum: nas, ochi, urechi, umeri, coate, încheieturi, solduri, genunchi și glezne. Coordonatele (x, y) din imagine ale punctelor corespunzătoare șoldurilor, gleznelor și gleznelor unui pieton detectat și urmărit în imaginea în infraroșu reprezintă o parte din vectorul de trăsături folosit la recunoașterea acțiunii de a traversa. Exemple de puncte cheie și schelet pentru o imagine din setul JAAD sunt prezentate în Figura 12 (stânga), și pentru o imagine din setul CROSSIR în Figura 12 (dreapta).

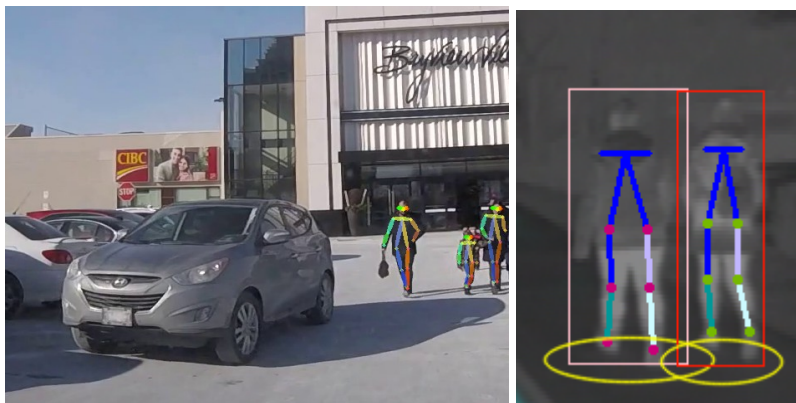


Figura 12. Puncte cheie și schelet extrase din imagini.

Recunoașterea acțiunii de a traversa strada: Componenta care analizează acțiunea de a traversa strada cuprinde o rețea de tipul Long Short Term Memory (LSTM), care utilizează un vector temporal de trăsături constituit din: trăsături de postură corespunzătoare picioarelor și trunchiului pietonului, contextul semantic al drumului extras pe baza segmentării semantice, informație de mișcare furnizată de modulul de urmărire al pietonilor.

În **Etapa 2** s-a definit un model de identificare a pericolului folosind rezultatele detecției pietonilor și a vehiculelor. Modelul se bazează pe clasificatori care primesc ca intrare o serie de trăsături sintetice, formate din:

a) Trăsături pentru pietoni:

- Poziția pietonilor în scena care este furnizată de detectorul de obiecte sub formă de coordonate x, y, lățime, înălțime a pietonului în pixeli.
- Distanța pietonului față de mașina proprie, distanța pietonului față de drum și distanța pietonului față de celelalte mașini din trafic.
- Acțiuni ale pietonului: trece sau nu trece strada, privește spre ego-vehicul sau nu.

b) Trăsături pentru vehicule:

- Tipul de mișcare: vehicul oprit, vehicul care se mișcă încet, accelerează, frânează, sau se mișcă repede. Pentru setul de date JAAD aceste trăsături sunt inferate din adnotări, iar pentru setul de date construit cu modulul de achiziție al acestui proiect, aceste informații sunt date de senzori.

În cadrul componentei de clasificare s-au experimentat mai multe metode de învățare convenționale, bazate pe trăsături, cum ar fi AdaBoost, Random Forest, Arbori de Decizie, Support Vector Machines (SVM), dar și modele bazate pe rețele neuronale, precum Long Short Term Memory (LSTM). Deoarece setul de date este de dimensiune medie, modelele convenționale, care utilizează trăsături, sunt potrivite pentru a obține o acuratețe satisfăcătoare, mai bună decât a modelului bazat pe rețele neuronale.

S-au implementat, testat și comparat mai multe modele de clasificare a gradului de pericol. Evaluarea modelelor s-a realizat pe setul standard JAAD, dar și pe setul de imagini construit în cadrul acestui proiect. Rezultatele obținute pe setul JAAD sunt prezentate în tabelul 5, iar rezultatele pe imaginile adnotate în cadrul acestui proiect sunt prezentate în tabelul 6.

Tabel 5. Rezultate clasificare grad pericol, set de date JAAD.

| Clasificator | Acuratețe | Senzitivitate | Precizie | Scor F1 |
|-------------------|-----------|---------------|----------|---------|
| Arbore de decizie | 94 % | 93% | 95% | 94% |
| Random Forest | 97% | 97% | 97% | 97% |
| AdaBoost | 82% | 81% | 80% | 81% |
| SVM | 81% | 78% | 81% | 80% |
| LSTM | 77% | 73% | 75% | 74% |

Tabel 6. Rezultate clasificare grad pericol, set date proprii.

| Clasificator | Acuratețe | Senzitivitate | Precizie | Scor F1 |
|-------------------|-----------|---------------|----------|---------|
| Arbore de decizie | 62% | 52% | 57% | 54% |
| Random Forest | 80% | 89% | 86% | 87% |
| AdaBoost | 78% | 74% | 80% | 77% |
| SVM | 67% | 63% | 68% | 65% |
| LSTM | 72% | 69% | 71% | 70% |

După cum se poate observa din tabelul care arată rezultatele experimentale, clasificatorii de tip Random Forest furnizează cele mai bune rezultate. Acești clasificatori agregă mai mulți clasificatori de tip arbori de decizie. Totodată, Random Forest sunt potriviți și pentru clasificarea multi-clasă, cum este cazul de față, deoarece avem de recunoscut trei tipuri de pericol: mic, mediu, ridicat. Clasificatorii de tip LSTM au o performanță mai redusă deoarece, prin structura lor, au nevoie de mult mai multe date de antrenare decât clasificatorii convenționali, pentru a obține rezultate bune.

În **etapa 3** am abordat o metodă nouă, inspirată de tendințele moderne de a implementa sisteme end to end pentru asistența condusului bazate pe rețele neuronale convoluționale. Sistemele end to end implică integrarea percepției și a acțiunii într-un singur bloc, antrenând sistemul să acționeze pe baza fluxului de date senzoriale, fără a necesita o interpretare și o reprezentare intermediară, explicabilă pentru utilizatorul uman. Deoarece abordarea integrează detecția pericolului cu acțiunea de evitare acestuia, o vom descrie în secțiunea următoare, dedicată proiectării modelului de răspuns.

O5. Proiectarea unui model de răspuns la pericol

Acest important obiectiv a fost abordat începând cu **etapa 2**. Ideea inițială a fost de a genera, folosind simulatorul de trafic CARLA, situații periculoase împreună cu percepția senzorială corespunzătoare acestora, urmând să antrenăm o rețea neuronală recurentă pentru a recomanda acțiunea optimă.

Abordarea a fost schimbată total în **etapa 3**. Inspirați de tendințele actuale de a avea o soluție end to end bazată pe rețele neuronale, am explorat posibilitatea de a genera un răspuns direct pe baza unei secvențe de imagini combinată cu datele de stare ale autovehiculului propriu. Inițial am încercat o rețea recurentă, a cărei intrare urma să fie imaginea curentă percepută de cameră și viteza autovehiculului, dar rezultatele nu au fost satisfăcătoare. Abordarea care a dus la rezultate bune a fost o rețea convoluțională clasică, în care o secvență de imagini a fost introdusă ca o singură imagine de adâncime mare.

Rețeaua neuronală oferă predicția frânării de urgență pe baza fluxului de imagini și a vitezei vehiculului propriu. Intrarea rețelei este o secvență de 20 de imagini și viteza vehiculului propriu asociată cu fiecare imagine. Valorificând informațiile spațiale și temporale capturate de rețeaua neuronală, modelul prezentat de noi e capabil să anticipeze cu precizie necesitatea acțiunilor imediate de frânare.

Arhitectura este prezentată în figura 13.

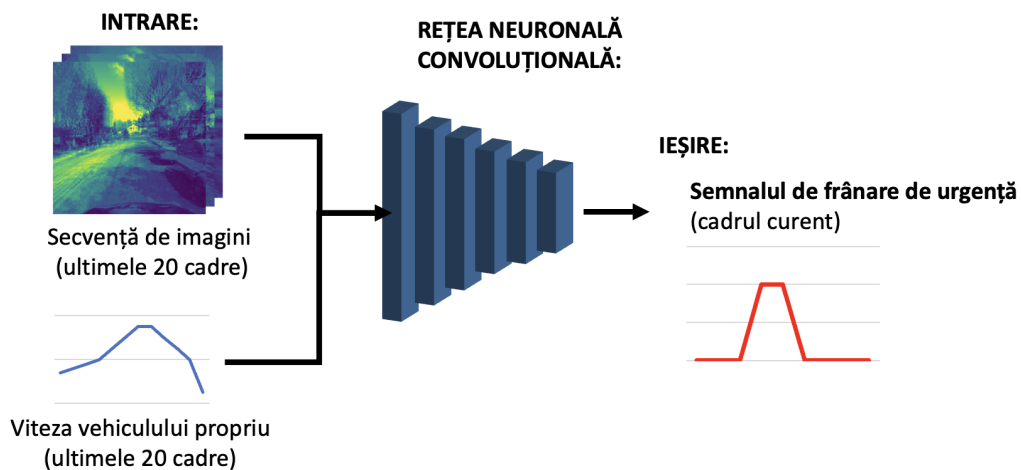


Figura 13. Sistemul propus pentru detecția situațiilor periculoase folosind imagini și date senzoriale (viteza vehiculului propriu).

Modelul primește două intrări (o secvență de 20 de imagini consecutive din trecut ale scenei de trafic rutier și 20 de valori reprezentând viteza vehiculului propriu pentru fiecare imagine) și are o singură ieșire (semnalul de frânare de urgență a vehiculului propriu pentru momentul actual de timp). Secvența de 20 de imagini este prezentată rețelei sub forma unei imagini cu dimensiunea de 300x300 pixeli, fiecare pixel având adâncimea de 20 de canale, fiecare canal fiind intensitatea (grayscale) unei imagini din cele 20 din secvență. Arhitectura rețelei începe cu cinci straturi convoluționale pentru a procesa intrarea de 20 de imagini: primele 3 straturi convoluționale au câte 24, 36 și 48 de filtre de dimensiune 5x5, urmate fiecare de funcția de activare de tip ReLU pentru a introduce non-linearitate.

A doua intrare o reprezintă viteza vehiculului propriu. Este un semnal uni-dimensional cu o lungime de 20, capturând informații despre viteză. Ieșirea din ultimul strat convoluțional și a doua intrare (viteza) sunt apoi concatenate. După concatenare, rețeaua continuă cu o serie de 3 straturi dense cu câte 100, 50 și 10 unități, fiecare urmat de câte o funcție de activare ReLU.

În final, se creează un strat de ieșire cu o singură unitate și funcție de activare liniară pentru a produce ieșirea finală a modelului. Ieșirea reprezintă valoarea semnalului pentru frânarea de urgență, prezis pe baza informațiilor combinate din imagini și viteza vehiculului.

Pentru a antrena această rețea, a fost nevoie să folosim baze de date publice din care să extragem secvențele de câte 20 de imagini și 20 de viteze asociate, dar și să avem informații referitoare la situații de frânare, sau frânarea de urgență. Am folosit setul de date Honda Deep Drive (HDD), care este o bază de date utilizată în domeniul vehiculelor autonome și cercetării în domeniul viziunii artificiale. Include date colectate de la o varietate de senzori, cum ar fi camerele, senzori de tip LIDAR (Light Detection and Ranging), și GPS (Global Positioning System), pentru a captura informații despre mediul înconjurător al vehiculului. De asemenea, dispune de date și din rețeaua CAN (en. "Controller Area Network") a vehiculului pentru a obține date direct de la mașină, cum ar fi viteza, accelerația, sau chiar și date despre presiunea aplicată pe pedala de frână și pe pedala de accelerație. Datele vizuale captate cu camere constau în imagini color cu o rezoluție de 1280x720 de pixeli, capturate la o frecvență de 30 de cadre pe secundă.

Deoarece baza de date nu include adnotare referitoare la gradul de pericolozitate al unei situații din trafic, ne-am folosit de datele referitoare la pedala de frână, asociate cu fiecare imagine. Analizând derivata acestui semnal (presiunea exprimată în kPa (kilopascal) asupra pedalei de frână), am putut identifica valorile mari ale acestui semnal și am folosit presupunerea că o derivată cu valoare mare implică o acționare de urgență a frânei, deci un răspuns la o situație de pericol. Maximele locale ale derivatei semnalului de frână, filtrate pentru a elimina zgomotele, sunt selectate pentru a indica pericolul, după cum este ilustrat în figura 14.

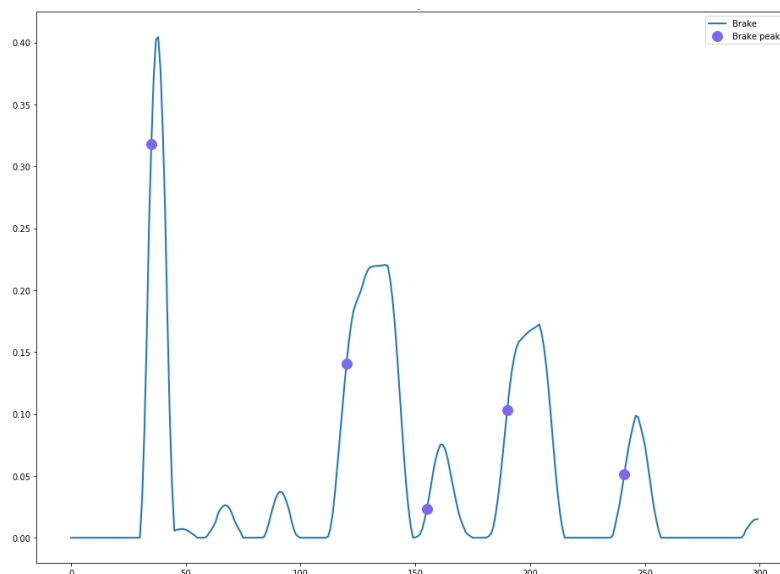


Figura 14. Semnalul pedalei de frână, cu valorile maxime ale derivatei de ordin 1 (gradient) filtrate.

Deoarece semnalul de frână este un răspuns la situația de pericol, întârziat de timpul de răspuns al șoferului, semnalul pe care îl vom considera că semnalează pericolul va fi anticipat față de maximul local al derivatei presiunii pedalei, după cum este ilustrat în figura 15. Astfel, am putut genera un semnal de referință (ground truth) pentru situații de pericol, asociat cu imagini reale dintr-o bază de date de dimensiuni mari, codificând percepția asupra pericolului a unui șofer real, în condiții de trafic real. În consecință am reușit adnotarea pentru gradul de pericolozitate a unei baze de date cu zeci de mii de imagini în mod automat.

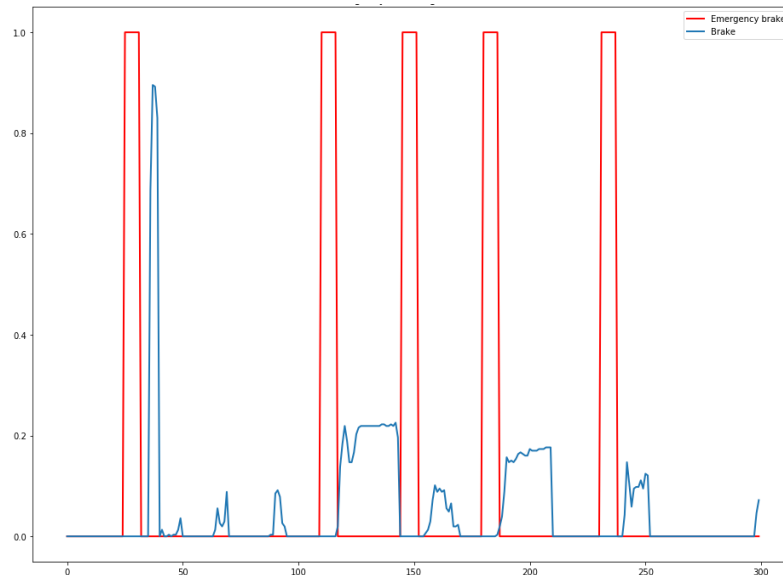


Figura 15. Crearea unui semnal de frânare de referință pentru situația de pericol (roșu), pe baza semnalului de frână (albastru), exploatând situațiile de frânare bruscă.

Având acest nou semnal de referință, și imaginile combinate cu viteza proprie din baza de HDD, am antrenat rețeaua neuronală convoluțională folosind funcția de cost MSE (en. “Mean Squared Error”) și optimizatorul Adam. Un exemplu de predicție este prezentat în figura 16.

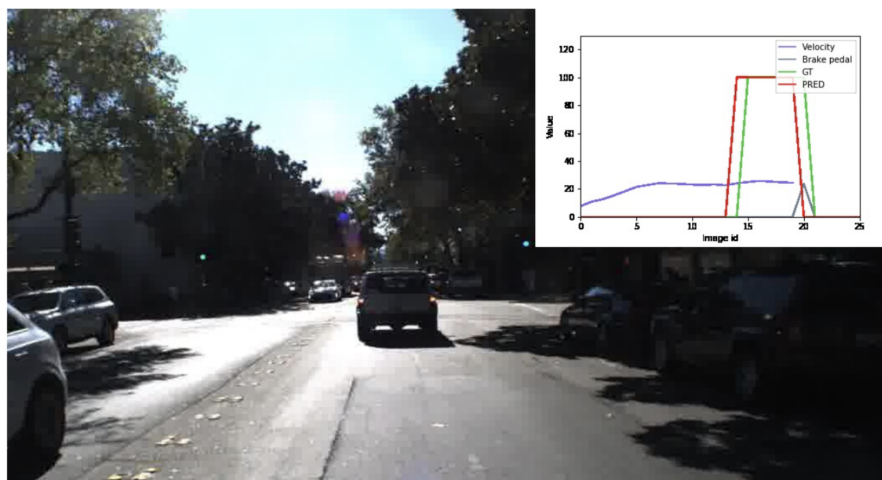


Figura 16. Exemplu de predicție a semnalului de frânare de urgență (roșu), comparat cu semnalul de referință (verde).

Pentru a evalua performanța modelului propus de a prezice semnalul de frânare de urgență pe baza secvențelor de imagini și a datelor de viteză, am efectuat o evaluare folosind diverse metrice de clasificare: TP (True Positive), FP (False Positive), TN (True Negative), FN (False Negative) și am calculat de asemenea acuratețea, precizia și rata de erori de tip TP (True

Positive Rate) și FP (False Positive Rate) și scorul F1. Rezultatele evaluării sunt prezentate în tabelul 7, unde evaluarea a fost efectuată pe un set distinct de 5 secvențe din baza de date de la Honda Research Institute.

Tabel 7. Performanța sistemului CNN de identificare a necesității de frânare de urgență.

| | Setul #1 | Setul #2 | Setul #3 | Setul #4 | Setul #5 |
|--|----------|----------|----------|----------|----------|
| Număr total de cadre din set | 5090 | 3603 | 11154 | 2372 | 6266 |
| Semnale de frânare de urgență reale din set (Ground Truth) | 58 | 60 | 68 | 28 | 114 |
| True Positive (TP) | 51 | 55 | 55 | 24 | 101 |
| False Positive (FP) | 40 | 34 | 26 | 1 | 19 |
| True Negative (TN) | 4992 | 3509 | 11060 | 2343 | 6133 |
| False Negative (FN) | 7 | 5 | 13 | 4 | 13 |
| True Positive Rate (TPR) | 0.879 | 0.916 | 0.808 | 0.857 | 0.885 |
| False Positive Rate (FPR) | 0.007 | 0.009 | 0.002 | 0.0004 | 0.003 |
| Acuratețea | 0.990 | 0.989 | 0.996 | 0.997 | 0.994 |
| Precizia | 0.560 | 0.617 | 0.679 | 0.960 | 0.841 |
| Scor F1 | 0.684 | 0.738 | 0.738 | 0.905 | 0.863 |

Evaluarea s-a efectuat prin numărarea tuturor semnalelor de frână de urgență cu lungimea egală sau mai mare de 10 cadre (considerate semnale puternice) atunci când s-au calculat falsurile pozitive (FP). În calculul "True Positive" (TP), s-a considerat orice semnal cu o lungime mai mare de 1 cadru.

06. Proiectarea și implementarea unei aplicații demonstrator

Ținta finală a proiectului este un sistem care să combine avantajele unui telefon mobil (portabil, echipat cu senzori) cu avantajele unui PC (posibilitate de extindere a capacităților senzoriale, putere de procesare mare). Un sistem care se apropie de aceste deziderate este placa de dezvoltare nVidia Jetson Nano, care oferă putere de calcul relativ mare, incluzând un GPU care poate rula soluții bazate pe rețele neuronale.

În **etapa 1** a fost realizat un sistem capabil de achiziție și procesare de date de la o cameră video, și date de orientare și accelerație de la un senzor inerțial, IMU. Aceste date sunt salvate pe un USB drive sau pe un card micro SDXC. Un calculator portabil poate monitoriza

datele conectându-se la sistem via o conexiune WiFi, cu un terminal de tip Secure Shell (SSH). Evident, datele pot fi procesate și direct pe placa de dezvoltare.



Figura 17. Sistemul bazat pe Jetson Nano.

Sistemul are următoarele componente cheie: Placa nVidia Jetson Nano Rev. A02, cu procesor ARM A57 Quad-Core, GPU Maxwell cu 128 de nuclee, 4 GB RAM, conector dedicat pentru cameră, 4 porturi USB 3.0, porturi GPIO, I2C, SPI, UART. Sistemul este de asemenea echipat cu un senzor inerțial (IMU) TDK MPU-6050. Acest senzor are un giroscop cu domeniul de măsurare de ± 250 , ± 500 , ± 1000 , și ± 2000 °/sec, și un accelerometru cu domeniul de măsurare de $\pm 2g$, $\pm 4g$, $\pm 8g$, și $\pm 16g$. Acest senzor este potrivit pentru estimarea mișcărilor rapide, dar și a celor lente. Este conectat la Jetson Nano folosind interfața I2C. Pentru conectivitate cu alte sisteme de calcul a fost instalat un adaptor wireless Intel Wi-Fi 6 (802.11 ax) AX200NGW.

În **etapa 2** s-a continuat dezvoltarea sistemului bazat pe placa nVidia Jetson Nano, care a fost echipată cu două camere video, una pentru monitorizarea traficului din fața autovehiculului, pentru detecția participanților la trafic și a pericolului posibil creat de aceștia, a doua cameră fiind pentru monitorizarea șoferului și identificarea posibilelor stări periculoase ale acestuia. Sistemul este alimentat de la priza de brichetă a autovehiculului, și poate achiziționa și procesa date în timpul mersului.



Figura 18. Sistemul de achiziție și procesare instalat în interiorul autovehiculului.

În aceeași etapă a început dezvoltarea arhitecturii software a aplicației demonstrator, centrată pe capacitatea plăcii nVidia Jetson Nano de a efectua în timp real calcul paralel

folosind nucleele GPU cu care este echipată, acest calcul paralel incluzând inferențe pe baza rețelelor neuronale convoluționale (CNN). Aplicația software utilizează patru fire de execuție, ce se execută în paralel: un fir este responsabil pentru citirea accelerometrului, giroscopul și a direcției magnetice de la un senzor inerțial IMU, prin magistrala I2C, al doilea fir este responsabil pentru citirea vitezei autovehiculului de la un senzor GPS prin interfața serială UART, iar celelalte fire sunt dedicate achiziționării și procesării imaginilor de la cele două camere.

Imaginile de la camera externă și cea internă sunt procesate în GPU pentru detectarea obiectelor. Imaginile de la camera orientată spre exterior sunt procesate printr-o rețea SSD MobileNet V2 care s-a optimizat cu mediul de lucru TensorRT pentru a rula mai rapid pe sistemul Nano suportat de CUDA. Această rețea acceptă 91 de clase de obiecte COCO. Imaginile de la camera internă sunt deduse printr-o rețea FaceNet care a fost optimizată în același mod cu TensorRT ca și cea anterioară, responsabilă cu detectarea feței șoferului.

În **etapa 3** a fost depus efort pentru a integra fiecare algoritm în aplicația demonstrator. Limitarea principală este nucleul GPU, care este esențial pentru inferența folosind rețelele neuronale. În acest moment avem următoarele rețele:

Hardware-ul trebuie să suporte mai multe tipuri diferite de rețele neuronale pentru scopuri diferite. În sistem trebuie integrate următoarele rețele neuronale:

- O variantă personalizată a unei rețele pentru determinarea orientării capului șoferului, implementată cu framework-ul PyTorch.
- O rețea SSD MobileNet v3 pentru detecția obiectelor din imagini, implementată cu framework-ul TensorFlow.
- O rețea personalizată pentru urmărirea multiplelor obiecte din imagini termice.
- O rețea personalizată pentru estimarea situațiilor de pericol iminent.

Aceste rețele sunt antrenate pe stații de lucru echipate cu nuclee GPU mult mai puternice decât Jetson Nano. Pentru a putea fi portate pe hardware-ul țintă, aceste rețele preantrenate trebuie convertite într-un format comun, apoi în cel pentru platforma țintă. Modelele PyTorch și TensorFlow sunt convertite în formatul ONNX - Open Neural Network eXchange. Din acest format se va genera modelul optimizat în formatul proprietar UFF pentru a putea fi integrat în proiectul CUDA pentru Jetson Nano, bazat pe C++.

Arhitectura software generică a aplicației a fost de asemenea actualizată. Cea mai importantă provocare este existența unui singur GPU pe placa Jetson Nano, ceea ce înseamnă că fluxul de procesare trebuie să țină cont de această resursă partajată. Deși GPU este capabil să facă mai multe calcule în paralel, nu este posibil să specificăm noi separat care parte din resurse se pot aloca pentru o anumită sarcină. Astfel, calculele diferite care trebuie făcute pentru a procesa imaginile de intrare de la camerele video trebuie specificate într-o ordine anume, și vor fi executate într-o manieră secvențială.

Fluxul de execuție actualizat este prezentat în figura 19.

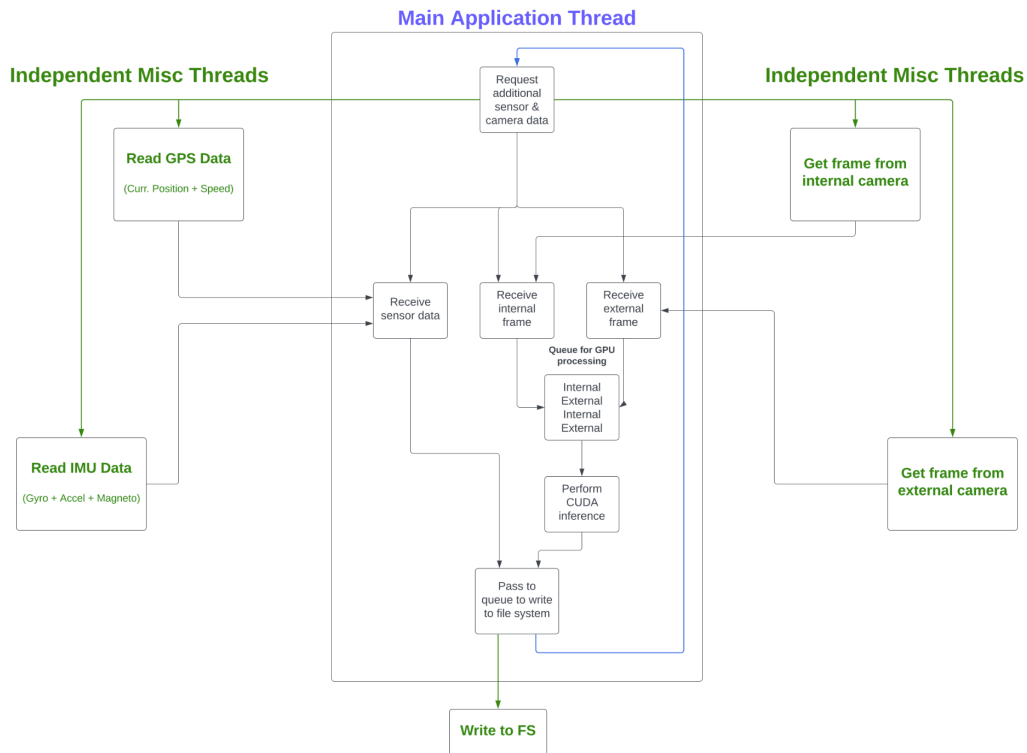


Figura 19. Fluxul actualizat al aplicației.

07. Diseminarea rezultatelor

Au fost publicate un total de 10 articole, din care două sunt în jurnal ISI, iar 8 la conferințe care sunt sau urmează să fie indexate ISI (la conferințe indexarea poate dura până la un an după publicare). Un al 11-lea articol este trimis spre publicare într-un jurnal ISI, dar este în acest moment în evaluare. De asemenea, prin implicarea studenților au fost realizate trei lucrări de diplomă din domeniul proiectului, iar un student a reușit să elaboreze o lucrare premiată la o conferință studențească. Nu în ultimul rând, un doctorand din echipa proiectului a reușit să finalizeze teza de doctorat.

Articole științifice:

Etapa 1

1. M. P. Muresan, S. Nedevschi, R. Danescu, "Robust Data Association using Fusion of Data-Driven and Engineered Features for Real Time Pedestrian Tracking in Thermal Images", Sensors, Volum 21, Numar 23, Numar articol 8005, 2021. [jurnal ISI - Zona rosie]
2. R. Itu and R. Danescu, "Object detection using part based semantic segmentation," 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2021, pp. 227-231.
3. R. D. Brehar, C. C. Vancea, M. P. Mureșan, S. Nedevschi and R. Dănescu, "Pose Based Pedestrian Street Cross Action Recognition in Infrared Images," 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2021, pp. 41-46.
4. M. P. Muresan, M. Raul, S. Nedevschi and R. Danescu, "Stereo and Mono Depth Estimation Fusion for an Improved and Fault Tolerant 3D Reconstruction," 2021 IEEE

17th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2021, pp. 233-24.

Etapa 2

5. R. Itu and R. Danescu, "Part-Based Obstacle Detection Using a Multiple Output Neural Network," *Sensors*, vol. 22, no. 12, număr articol 4312, 2022. **[jurnal ISI – Q1, zona roșie]**
6. R. D. Brehar, R. O. Băbuț, A. Fuzes and R. Dănescu, "Outdoor Traffic Scene Risk Estimation in the Context of Autonomous Driving," 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2022, pp. 129-134.
7. M. P. Muresan, R. Schlanger, R. Danescu, S. Nedevschi, "Real-Time Obstacle Detection using a Pillar-based Representation and a Parallel Architecture on the GPU from LiDAR Measurements", 2023 International Conference on Computer Vision Theory and Applications (VISAPP) , 2023, pp. 779-787.

Etapa 3

8. R. Brehar, G. Coblișan, A. Fuzes, R. Dănescu, "Driver Attention Estimation Based on Temporal Sequence Classification of Distracting Contexts", In Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics - Volume 1 (ICINCO), 2023, SciTePress, pp. 578-585.
9. R. Itu, R. Danescu, "On-Board Estimation of Vehicle Speed and the Need of Braking Using Convolutional Neural Networks", In Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics - Volume 1 (ICINCO), 2023, SciTePress, pp. 600-607.
10. R. Itu, R. Danescu, "Predicting Emergency Braking in Vehicles Using a CNN with Sequential Image and Velocity Data", 2023 IEEE 19th International Conference on Intelligent Computer Communication and Processing (ICCP), 2023.
11. R. Itu, R. Danescu, "Fully Convolutional Neural Network for Vehicle Speed and Emergency Brake Prediction", *Sensors*, in review. **[jurnal ISI]**

Implicarea studenților în activitatea de cercetare:

Articol prezentat la conferința științifică a studenților CSSC 2023:

Ioana Vîrnă, Conf. Dr. Ing. Raluca Didona Brehar: Facial Expression and Stress Level Detection System (premiul 2)

Lucrări de licență:

1. Student: George Coblișan, Coordonator științific: Conf. Dr. Ing. Raluca Didona Brehar, Titlu: Detectarea acțiunilor unui șofer și prezicerea gradului de atenție
2. Student: Ioana Vîrnă, Coordonator științific: Conf. Dr. Ing. Raluca Didona Brehar, Titlu: Sistem de detectare a expresiilor faciale și a nivelului de stres
3. Student: Mădălina-Alexandra Marc, Coordonator științific: Conf. Dr. Ing. Raluca Didona Brehar, Titlu: Predicția nivelului de oboseală în funcție de acțiunile feței

Teze de doctorat:

1. Mircea Paul Mureșan, "Multimodal Measurement Approaches for Autonomous Systems", îndrumător Prof. Sergiu Nedevschi, 2023.

2. Rezultate obținute

Urmând realizarea fiecărui obiectiv concret, au fost obținute rezultate care au fost descrise în secțiunile dedicate fiecărui obiectiv în parte. Au fost obținute rezultate originale, concretizate în publicații, pentru domeniul de percepție al scenei din imagini monoculare color și din imagini termale, incluzând detecția autovehiculelor și a pietonilor, și urmărirea acestora, dar și pentru monitorizarea șoferului pentru a determina dacă acesta este neatent sau dacă este în stare de ebrietate. De asemenea, a fost realizat un sistem pentru determinarea în mod generic a situației de pericol, folosind în mod direct imaginile capturate de la camera video, sub formă de secvență de 20 de imagini, împreună cu viteza autovehiculului propriu. A fost realizat un sistem hardware și software ușor de montat în orice autovehicul, echipat cu camere video, pentru achiziția și procesarea datelor la bordul autovehiculului, pentru transformarea algoritmilor realizați în aplicații concrete.

Considerăm că fiecare obiectiv definit a fost îndeplinit cu succes.

Indicatori de rezultat:

Articole ISI: 2 articole de jurnal în zona roșie, 1 articol de jurnal în evaluare

Articole conferințe ISI: 8 articole

Teză de doctorat: 1

Algoritmi, metode: 1 algoritm pentru percepția monoculară a obiectelor și identificarea instanțelor acestora, 1 algoritm pentru urmărirea obiectelor din imagini termale, 1 algoritm pentru segmentarea de instanțe din imagini termale, 1 algoritm pentru determinarea atenției șoferului, 1 algoritm pentru determinarea situațiilor de pericol și frânarea de urgență.

3. Impactul rezultatelor obținute

Rezultatele publicate au fost bine primite de comunitatea științifică internațională, acest lucru fiind dovedit prin citările obținute în acest scurt timp.

1. Articolul "Robust data association using fusion of data-driven and engineered features for real-time pedestrian tracking in thermal images", publicat în jurnalul Sensors în 2021 a fost **citat de 18 articole în jurnale ISI și în 4 articole de conferință**, fiind ales, în anul 2023, ca "**Editor's Choice Article**" pentru jurnalul Sensors.
2. Articolul "Stereo and mono depth estimation fusion for an improved and fault tolerant 3D reconstruction", publicat la conferința ICCP 2021, a fost citat de 3 articole în jurnale ISI, și un articol la conferința ICCV 2023, una dintre cele mai prestigioase conferințe din domeniul Computer Vision.
3. Articolul "Part-based obstacle detection using a multiple output neural network", publicat în jurnalul Sensors în anul 2022 a fost **citat de 3 articole în jurnale ISI**.

Cel mai important rezultat obținut: determinarea gradului de pericol și necesitatea de frânare folosind doar o secvență de imagini, introdusă ca o imagine multi-canal într-o rețea neuronală convoluțională, împreună cu informația de viteză. Succesul acestei abordări arată că nu este întotdeauna necesară interpretarea scenei în termeni de obiecte, pietoni, drum, pentru că o rețea neuronală poate infera condițiile de pericol în mod direct.

Director Proiect,

Prof. Dr. Ing. Radu Danescu