

Utilizarea rețelelor neuronale profunde în predicția deplasării participanților la traficul rutier

Teză de doctorat – Rezumat

Pentru obținerea titlului științific de doctor la
Universitatea Politehnica Timișoara

În domeniul de doctorat Inginerie Electronică, Telecomunicații și Tehnologia Informației

autor ing. Bogdan Ilie SIGHENEA

conducător științific Prof. Univ. Dr. Habil. Ing. Cătălin Daniel CĂLEANU
Septembrie 2023

Teza de doctorat este structurată în 7 capitole. Capitolul 1 prezintă o scurtă introducere în tematica utilizării rețelelor neuronale profunde în vederea prezicerii traiectoriei pietonilor, motivând necesitatea unor astfel de sisteme. Capitolul 2 se referă la elemente de inteligență artificială cu accent pe conceptul de rețea neuronală artificială. Sunt prezentate concepte legate de arhitectura și modalitatea de instruire aferente rețelelor neuronale profunde, în special modele folosite în problema predicției traiectoriei pietonilor (PTP). Analiza stadiului actual al realizărilor aferente PTP este efectuată pe parcursul Capitolului 3. Rezolvarea cu succes a problemei PTP depinde de informația senzorială disponibilă sistemului de predicție. Din acest motiv a fost ales să se analizeze caracteristicile și performanțele diverselor tipuri de senzori în cadrul Capitolului 4 al tezei. Soluțiile propuse sunt prezentate împreună cu rezultatele experimentale aferente în cadrul capitolelor 5 respectiv 6. Lucrarea se încheie cu un capitol destinat concluziilor și posibilelor direcții de dezvoltare ulterioară. De asemenea lucrarea conține o secțiune de referințe bibliografice.

În figura de mai jos se poate urmări întreaga structură a lucrării împărțită pe capitole și subcapitole.

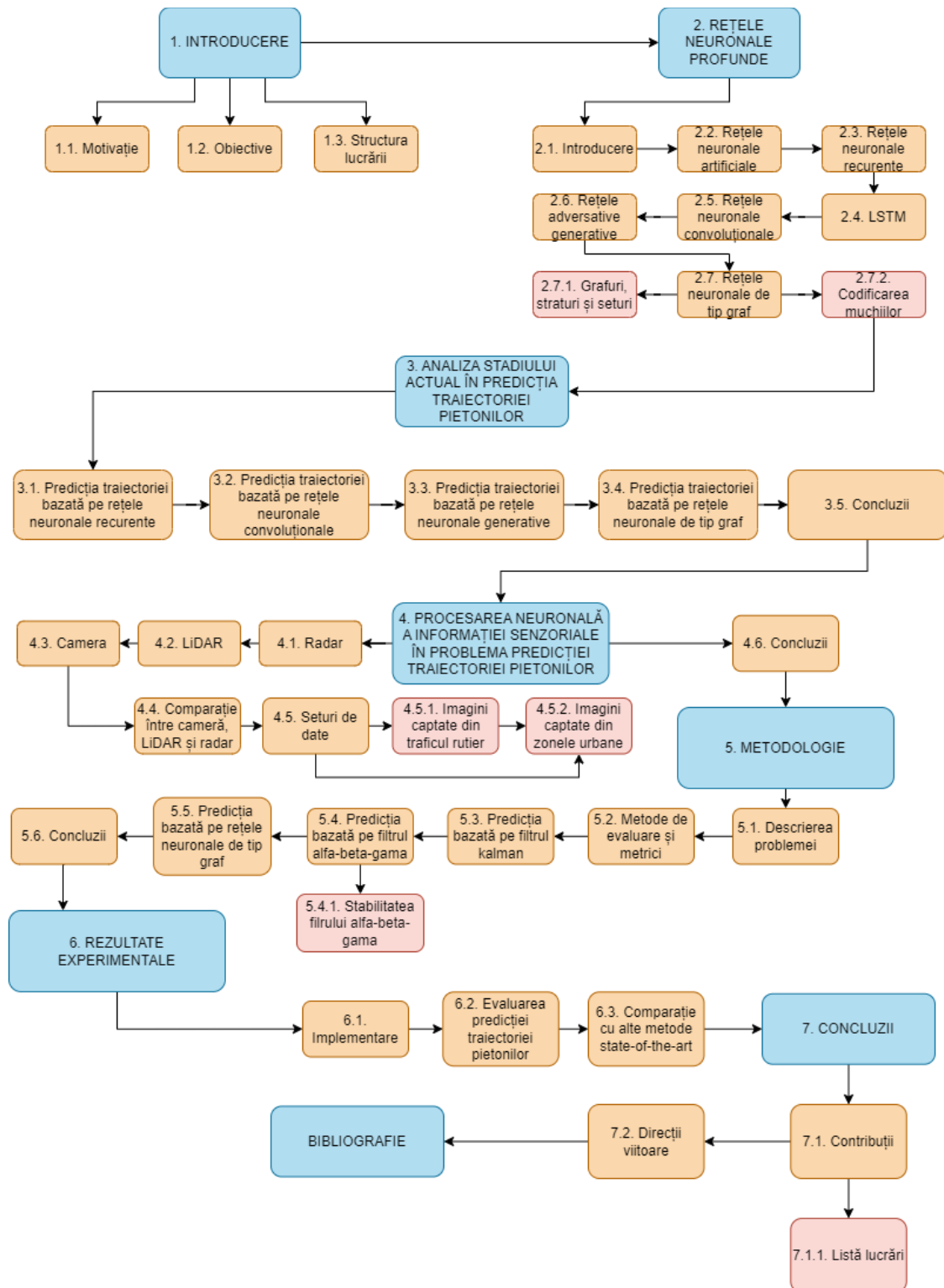


Figura 1. Structura lucrării pe subcapitole.

1. Introducere

În ultimele decenii, producătorii de automobile au lucrat în mod constant la îmbunătățirea experienței de conducere și a face vehiculele rutiere mai sigure prin dezvoltarea tehnologiilor de asistență pentru șofer. Pentru a evalua amploarea progreselor în tehnologia de asistență a șoferului, au fost definite șase niveluri de autonomie de către „Society of Automotive Engineers” (SAE). Aceste niveluri variază de la 0, care corespunde condusului complet manual, până la 5 complet autonom, care este scopul final al cercetărilor recente realizate atât de industria auto, cât și de lumea academică.

Vehiculele autonome (VA) au potențialul de a transforma lumea așa cum o cunoaștem, revoluționând transportul, făcându-l mai rapid, mai sigur și mai puțin intensiv în lucru. Este important ca sistemele AV să perceapă cu acuratețe și să reacționeze în siguranță la diverse scenarii de conducere din lumea reală. Acest lucru necesită ca sistemele de percepție AV să înțeleagă comportamentul participanților la trafic din jur (de exemplu, vehicule, pietoni și bicicliști) și să prezică cu exactitate traiectoriile și comportamentele lor viitoare [1]. Pentru predicții pe termen scurt, poate fi acceptabil să se utilizeze abordări bazate pe fizică pură. Cu toate acestea, deoarece scenariile viitoare sunt necunoscute, un sistem de predicție pe termen lung este esențial pentru a permite nu numai modelarea interacțiunii între diferiți agenți, ci și pentru a identifica regiunile traversabile definite de traseele rutiere și de conformitatea cu regulile de circulație.

Conform ultimului raport pentru accidentele rutiere publicat de către „European Road Safety Observatory” [2], aproape 24 000 de oameni au murit în accidentele rutiere din UE în anul 2019, dintre care 4668 sunt pietoni. Din acest total, 729 de pietoni care au murit în accidentele rutiere sunt din România. În ceea ce privește statisticile la nivel mondial, datele sunt mai îngrijorătoare. Raportul Global privind Siguranța Rutieră [3] publicat de Organizația Mondială a Sănătății (OMS), indică faptul că peste 1,3 milioane de oameni au murit în accidente rutiere în întreaga lume în anul 2017. Aproximativ 23% dintre aceste decese este reprezentat de către pietoni. Pe de altă parte, în țările Organizației pentru Cooperare și Dezvoltare Economică (OCDE), peste 20 000 de pietoni își pierd viața anual. Decesele pietonilor reprezintă între 8% și 37% din totalul deceselor din trafic, în funcție de țară și an [4]. Cele mai multe dintre aceste accidente tragice au loc în zone aglomerate de la trecerile de pietoni, cu vizibilitate redusă din cauza atenției scăzute a șoferului și/sau a oboselii.

2. Motivație

Potrivit [5], numărul accidentelor cu pietoni în vârstă este influențat de factori multipli din mediul localităților. Ca rezultat, reducerea (sau eliminarea) acestor coliziuni este o preocupare importantă de siguranță. În aceste situații, a ajuta șoferul include prezicerea comportamentului pietonilor. Acest lucru ajută la reducerea efectului diverșilor factori care ar putea afecta negativ siguranța în trafic (cum ar fi oboseala, vizibilitatea slabă, distracția cognitivă accidentală etc.).

Într-un eventual impact, pietonii nu au practic nici-o protecție. Prin urmare, reducerea (eliminarea) acestor accidente este o problemă importantă de siguranță. Ajutarea șoferului în astfel de condiții include prezicerea traiectoriei și/sau comportamentul pietonului și atenuarea erorilor consecvente ale șoferului (de exemplu, oboseală, gândirea cognitivă) și include dezvoltarea de noi tehnologii pentru a reduce numărul de accidente (cu până la 93,5%, conform [6]).

Noi, ca oameni, luăm decizii intuitive importante bazate pe secvențele de acțiuni și interacțiuni cu alte persoane din scenă pentru a obține o navigare sigură și lină. Această intuiție permite mișcări care sunt foarte dinamice, deoarece putem decide ce traseu să luăm într-o manieră foarte dinamică. Această informație simplă, dar valoroasă, este crucială pentru a decide următorul pas care trebuie făcut. Cu toate acestea, odată cu apariția "Deep Learning", sunt dezvoltati algoritmi avansați care "citesc" intențiile pietonilor și permit acțiuni în consecință. Sunt explorate diferite metode, de la predicția traiectoriei la analiza comportamentală [7]. În această teză de doctorat, investigăm ideea utilizării imaginilor monocular RGB ca informații de bază pentru a prezice traiectoria pietonilor.

3. Obiective

Această teză are ca și obiectiv predicția traiectoriei pietonilor în diferite scenarii folosind rețele neuronale profunde. Acest domeniu fiind de mare succes dar și foarte studiat de către cercetători datorită dezvoltării senzorilor optici (exemplu camera RGB, Radar, LiDAR etc.) dar și apariția de noi arhitecturi specifice învățării profunde ("deep learning"). Pentru atingerea obiectivului propus, au fost definite mai multe sarcini de lucru:

- Analiza stadiului curent în domeniu predicției traiectoriei a pietonilor, precum și analiza tuturor arhitecturilor existente folosind rețele neuronale profunde.
- Identificarea și implementarea de noi soluții "deep learning" (hibrid) care aduc îmbunătățiri substanțiale la soluțiile existente.
- Aplicarea modelelor dezvoltate asupra celor mai cunoscute baze de date pentru acest domeniu, și depășirea soluțiilor "state-of-the-art" la nivel de performanță și acuratețe.
- Testarea fezabilității și limitele metodelor propuse într-un mod extins în condiții ideale, utilizând baze de date din lumea reală.
- Măsurarea influenței modelării a patru dinamici pietonale diferite, adică statul pe loc, pornirea, oprirea și mersul pe jos. Aceste dinamici permit definirea adecvată a schimbărilor efectuate de pietoni în scenarii reale.
- Dezvoltarea unei metode de predicție a traseul pietonilor, aplicând noi modele neuronale (ex. arhitecturi neuronale de tip CNN, LSTM și GNN).

4. Rezumatul tezei pe capitole

4.1. Rețele neuronale profunde

Pentru a rezolva problema PTP, în ultimii ani, au fost propuse în literatura de specialitate mai multe metode bazate pe învățarea profundă. Această secțiune detaliază cele mai utilizate metode din această zonă, clasificate în funcție de tipul arhitecturii profunde ("Deep Neural Network", DNN). Metodele identificate de predicție a traiectoriei pietonale bazate pe învățarea profundă au utilizat în mare parte trei structuri arhitecturale, după cum urmează.

Cu costuri mai reduse de calcul și comunicare mai rapidă, care oferă acces nelimitat la informație și o mai bună înțelegere a lumii fizice din jurul nostru, luarea deciziilor puternic automate, cum ar fi inteligența artificială, devine tehnologia principală a secolului XXI. Astfel, inteligența artificială a fost elementul de bază al multor aplicații, cum ar fi mașinile autonome,

asistenții digitali și imagistica medicală, doar pentru a enumera câteva. Cu toate acestea, pare să existe o lipsă de înțelegere corectă a acestei tehnologii critice. În plus, datorită adopției largi a acestei tehnologii, există multe neînțelegeri în terminologie. Această interpretare greșită poate fi observată în principal atunci când termenii inteligență artificială, învățare automată și învățare profundă sunt interschimbați. Deși termenii par echivalenți, sensul fiecărui termen variază, iar aceasta secțiune își propune să articuleze clar diferențele dintre inteligența artificială, învățarea automată și învățarea profundă.

Igor Aizenberg și colegii săi au inventat pentru prima dată termenul de învățare profundă în anul 2000 [8]. Algoritmii de învățare profundă sunt o subcategorie a algoritmilor de învățare automată care imită procesul de învățare al oamenilor prin învățarea mașinii prin intermediul exemplelor. Algoritmii de învățare profundă utilizează structuri de învățare complexe cu mai multe straturi cunoscute sub numele de rețele neurale, care învață o reprezentare implicite a datelor brute în mod autonom pentru a produce rezultatul dorit. Cu alte cuvinte, pentru a face ca algoritmii tradiționali de învățare automată să funcționeze, este necesară o etapă esențială, dar extrem de complicată, cunoscută sub numele de extragerea de caracteristici, care trebuie realizată manual de către experții din domeniul pentru care algoritmii să funcționeze. Pe de altă parte, algoritmii de învățare profundă învață aceste caracteristici extrase automat în timp ce structurile de învățare din acești algoritmi se optimizează pentru a obține cea mai bună reprezentare abstractă posibilă a datelor de intrare.

Diferitele modele de traiectorie de mișcare (cu mai multe origini și destinații) și interacțiunea umană dinamică sunt cheia pentru un model de predicție a traiectoriei sub circumstanțe complexe. Cele mai multe metode existente bazate pe învățarea profundă depind puternic de scenarii specifice, deoarece efectuează predicția traiectoriei folosind coordonate absolute. În realitate, traiectoria de mișcare este o mișcare relativă care coincide cu timpul, iar interacțiunea umană este o mișcare relativă între pietoni. Acest lucru motivează construirea unui model de predicție a traiectoriei pentru mișcarea relativă atât a traiectoriei de mișcare, cât și a interacțiunii umane.

Diferite arhitecturi de învățare profundă au fost proiectate pentru a aborda sarcini specifice în diferite domenii. De exemplu, RNA recurente și mecanismele de atenție au fost utilizate în principal pentru modelarea limbajului în procesarea naturală a limbajului. Iar rețelele convoluționale sunt utilizate extensiv pentru a rezolva problema clasificării imaginilor și recunoașterea obiectelor în domeniul viziunii artificiale. Aceste arhitecturi au fost, de asemenea, extinse și adaptate în alte domenii, cum ar fi previziunile financiare și înțelegerea climei. Ele sunt, de asemenea, utilizate în domeniul condusului autonom și în mod specific, pentru prognozarea mișcării.

4.2. Analiza stadiului actual în predicția traiectoriei pietonilor

Acest capitol prezintă stadiul evoluției studiilor și metodelor de lucru legate de problematica de predicție a traiectoriei pietonilor. Am propus o taxonomie a metodelor pentru o mai bună înțelegere și clasificare a diferitelor abordări ale problemei [9].

Pentru a rezolva problema PTP, în ultimii ani au fost introduse în literatura de specialitate mai multe metode bazate pe o învățare aprofundată. Acest capitol detaliază cele mai utilizate metode din domeniu, clasificate în funcție de tipul arhitecturii DNN. Metodele identificate de

predicție a traiectoriei pietonilor, bazate pe metode aprofundate de studiu au folosit în principal patru structuri arhitecturale. Acestea vor fi descrise în cele ce urmează. O corespondență între arhitecturile de rețele neuronale folosite în literatură și problema PTP se poate observa în Figura 2.

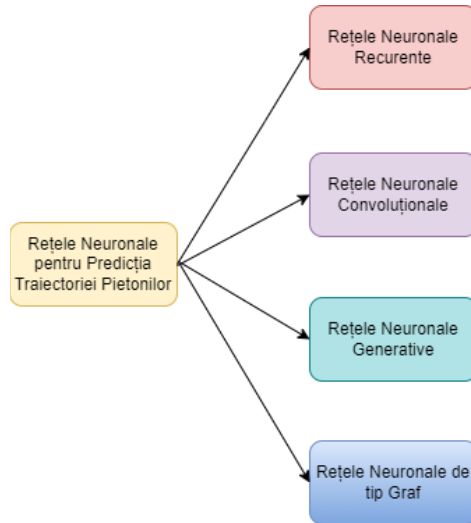


Figura 2. Cele mai utilizate rețele neuronale profunde pentru predicția traiectoriei pietonilor.

Arhitecturile utilizate la predicția traiectoriei pietonilor includ rețelele neuronale recurente, cele convoluționale, cele generative dar și cele de tip graf.

4.3. Procesarea neuronală a informației senzoriale în problema predicției traiectoriei pietonilor

În cadrul acestui capitol sunt examinate lucrările reprezentative din domeniu (“State-of-the-art”, SOTA), cu precădere metodele de predicție care folosesc rețele neuronale profunde împreună cu cei mai utilizați senzori (camera, LiDAR și radar) pentru conducerea autonomă a vehiculelor. S-au prezentat comparativ performanțele fiecărui senzor auto (radar, LiDAR și cameră) prin evidențierea avantajelor și dezavantajelor acestora în diferite sarcini, de exemplu detecția și clasificarea obiectelor, estimarea distanței, detecția marcajelor rutiere și a limitelor drumului etc. Au fost identificate cele mai importante seturi de date disponibile public, folosite de către cercetători în implementarea soluțiilor de tip PTP.

Un vehicul autonom (VA) reprezintă un ansamblu echipat cu senzori (LiDAR, radare, camere etc.) și sisteme capabile să-l controleze automat astfel încât să poată rula pe drum fără intervenție umană. Pentru a îndeplini astfel de sarcini, sistemul de control trebuie să poată detecta obiecte în apropierea vehiculului și să estimeze traiectoriile lor viitoare și parametri cinematici ai mișcării acestora [10].

Senzorii și sistemele care procesează informațiile primite oferă asistență șoferilor, semnalizându-le diferite circumstanțe pentru minimizarea riscul de expunere. Acestea pot chiar automatiza sarcinile de conducere pentru a elimina erorile umane [11]. Pentru a colecta informații din mediul exterior, VA utilizează senzori numiți exteroceptivi. Procesarea informațiilor primite duce la recunoașterea altor participanți la trafic (pietoni, vehicule, etc.) și a obiectelor din

apropiere. În ultimii ani, senzorii externi ai VA au câștigat importanță, în special datorită dezvoltării procesării imaginilor și camerelor, deoarece aceste sisteme permit o gamă largă de aplicații [12].

Conform [13], asistența la conducere autonomă se bazează în principal pe sisteme legate de procesarea imaginilor provenite de la camere. LiDAR reprezintă cel mai necesar senzor în sistemele auto. În contrast cu camerele, acesta este caracterizat de o detecție omnidirecțională și nu este afectat de condițiile de lumină. Un sfert din piața senzorilor auto este reprezentat de senzorii ultrasonici și radar, în timp ce alți senzori exteroceptivi, cum ar fi microfoanele, acumulează 18% din piață (vezi Figura 3).

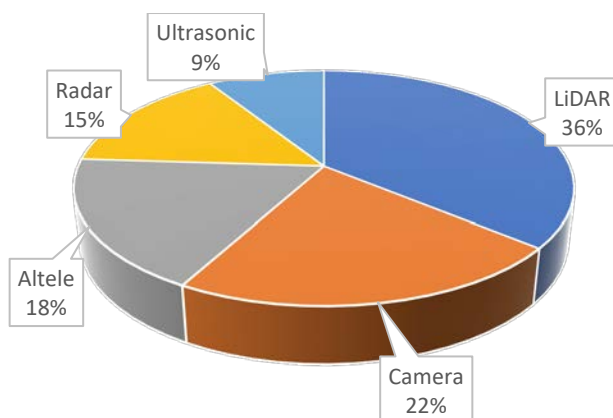


Figura 3. Predicția creșterii pieței senzorilor auto (rata de creștere anuală compusă (CAGR), 2017-2022) pe senzorii exteroceptivi, conform [79].

Un set de senzori amplasați pe vehicule captează date care sunt ulterior procesate pentru a obține o reprezentare digitală a mediului înconjurător. Această informație este folosită pentru evaluarea riscurilor precum și planificarea și prezicerea traiectoriilor, ceea ce conduce la controlul mișcării vehiculului.

Percepția oferă mijloacele prin care vehiculul poate obține informații referitoare la ce se întâmplă în mediul său de operare. Într-un scenariu de conducere autonomă, există numeroși participanți la trafic în jurul vehiculului: pietoni, bicicliști, motocicliști, alte vehicule, etc. O provocare majoră constă în a percepe toate aceste elemente într-un mod continuu și precis, fără a avea alarme false (raportarea de obstacole inexistente) sau lipsa de detecție (adică omisiunea de obstacole reale). În acest scop, se pot utiliza mai mulți senzori exteroceptivi activi și pasivi, printre care diferite tipuri de camere video, LiDAR și RADAR. Algoritmii de detecție de ultimă generație oferă și clasificare obiectelor percepute.

Cu toate acestea, nu există o tehnologie de senzori unică capabilă să ofere informații spațio-temporale precise și complete despre tot ceea ce înconjoară vehiculul, fiecare având propriile avantaje și dezavantaje [14]. Soluțiile aplicate în prezent combină astfel diferite tipuri de senzori prin intermediul unui proces de interconectare a mai multor senzori (fuziune).

Pentru a testa sistemele de predicție a traiectoriei pietonale, cercetătorii utilizează în mod obișnuit mai multe seturi de date. Acestea furnizează imagini ale pietonilor din diferite scenarii (promenadă, treceri de pietoni, trotuare, etc.). În acestea oamenii se mișcă în diferite direcții.

Seturile de date mari sunt resurse importante pentru antrenarea și testarea modelelor DNN însă date de acest tip necesită etichetare. Într-un scenariu posibil, pietonul poate fi instruit să

efectueze acțiuni predefinite (oprire, traversare pe trecere de pietoni, traversare, etc.), dar nu este limitat la acestea. Din păcate, numărul de date colectate este limitat. Se pot identifica două provocări majore în seturile de date disponibile: deoarece sunt predefinite, acestea nu cuprind toate informațiile (pietonul înregistrat este un "actor", variabilitatea acțiunilor sale este inexistentă). O a doua provocare este reprezentată de faptul că în viața reală, acțiunile pietonilor sunt variabile, ele fiind determinate de evenimente aleatorii (sosirea autobuzului, semaforul își schimbă culoarea, etc.). Scenariile din viața reală au fost utilizate pentru modelele de nivel scăzut, cum ar fi detecția și urmărirea. Din păcate, ele nu oferă datele necesare pentru modelele de nivel superior (de exemplu, interacțiunile sociale).

4.4. Metodologie

În cadrul acestui capitol au fost descrise metodele folosite de-a lungul cercetării pentru predicția traiectoriei pietonilor. În legătură cu tematica prezentului capitol am publicat lucrările [15], [16] pentru predicția bazată pe filtrul alfa-beta-gama. S-a efectuat o analogie între filtrul Kalman și filtrul alfa-beta-gama cu scopul de a identifica acurateței metodei. Pentru predicția bazată pe rețele neuronale de tip graf am publicat lucrările [17], [18]. Aici au fost prezentate metodele și arhitectura soluțiilor de predicție folosind rețeaua neurală de convoluție spațio-temporală pe graf (ST-GCNN) și rețeaua neurală de convoluție cu extrapolare temporală (TXP-CNN). Pentru evaluarea metodelor și comparația acestora cu alte metode din literatură, au fost prezentate cele mai reprezentative metrice din domeniu.

Obiectivul principal al prezicerii traiectoriei este de a estima pozițiile viitoare ale unui grup de N pietoni într-o scenă reală. Acest lucru se bazează pe pozițiile lor anterioare și curențe pe o reprezentare de tip hartă pe parcursul unui interval de timp, începând de la momentul T_o (așa cum se poate observa în Figura 4) după T_p iterații în timp. Poziția actuală a fiecărui pieton într-o scenă este reprezentată de coordonata reală $X = (x, y)$ la iterația i aceasta fiind notată cu $X_t^i = (x_t^i, y_t^i)$, cu $t \in \{1, \dots, T_o\}$. Aici (x_t^i, y_t^i) sunt variabile aleatoare care descriu distribuția de probabilitate a poziției pietonului n la momentul t în cadrul scenei (n este dependent de scenă). Traiectoria viitoare reală/adevărată este notat $Traj_{obs} = \{Y_t^i = (x_t^i, y_t^i)\}$, unde $i \in \{1, \dots, n\}$, $T_o + 1 \leq t \leq T_p$.

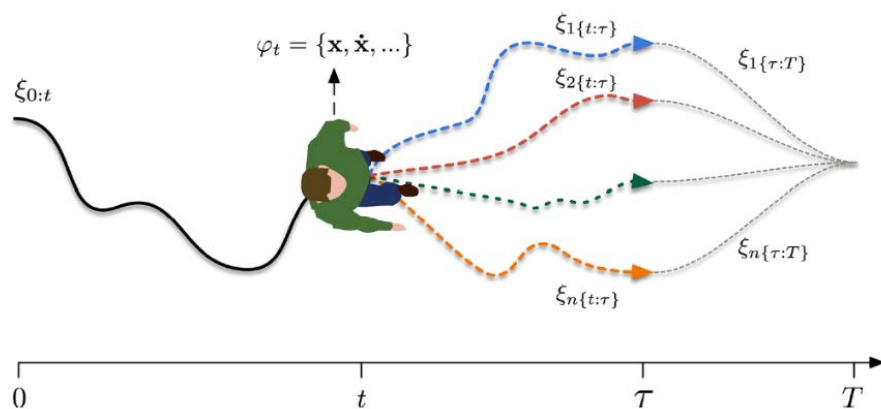


Figura 4. Distribuția temporală a poziției fiecărui pieton începând de la momentul T_o până la T_p . Aici se pot menționa trei tipuri de poziții: observate, reale-viitoare și prezise [17].

Pozițiile prezise, notate $Traj_{pred} = \{\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i)\}$, $i \in \{1, \dots, n\}$, $T_o + 1 \leq t \leq T_p$ reprezintă o serie de variabile aleatoare. Acestea sunt rezultatul unei presupunerii că poziția pietonului i la momentul t urmează o distribuție gaussiană bidimensională, notată $\hat{Y}_t^i \sim \mathcal{N}(\mu_t^i, \sigma_t^i, \rho_t^i)$. În cadrul acestei distribuții $\mu_t^i = (\mu_x, \mu_y)_t^i$ reprezintă centrul grupului de pietoni la momentul t . Deviația standard a distribuției este reprezentată de $\sigma_t^i = (\sigma_x, \sigma_y)_t^i$, iar coeficientul de corelație este notat ρ_t^i . Pentru a obține predicția traiectoriei, arhitectura propusă prezice parametrii distribuției gaussiene $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)_t^i$.

Pentru a estima pozițiile la momentele $T_o + 1 \leq t \leq T_p$ pentru fiecare participant la trafic se folosesc locațiile observate la momentele de timp $1 \leq t \leq T_o$. Pentru estimarea traiectoriei se folosește o arhitectură de rețea neuronală. Pentru a antrena parametrii modelului, se utilizează funcția de pierdere a logaritmului negativ al probabilității, așa cum este prezentată în Ecuația (1).

$$L^i(W) = - \sum_{t=1}^{T_p} \log(f(x_t^i, y_t^i | \mu_x, \mu_y, \sigma_x, \sigma_y, \rho)) \quad (1)$$

Unde W reprezintă parametrii rețelei antrenate. Valoarea pierderii este minimizată pentru a obține performanțele optime ale rețelei.

4.4.1. Metode de evaluare

În cercetarea în domeniul vederii artificiale, traiectoriile sunt adesea descrise prin statistici de mișcare, cum ar fi numărul de coliziuni, accelerația medie, viteza medie și distanța totală parcursă [19]. Pentru fiecare participant la trafic, există opt etape de observare, de exemplu de durată 3,2 secunde. Pentru toate seturile de date, se folosesc douăsprezece etape (de exemplu de durată 4,8 secunde) pentru a reprezenta pozițiile reale viitoare. Pentru evaluarea diferenței poziției dintre traiectoria reală și cea estimată se folosesc normele euclidiene L2.

ADE și FDE sunt cele două metrici principale utilizate pentru evaluarea regresorilor determinați. În timp ce aceste metrici sunt naturale pentru obiectivul propus, ușor de implementat și interpretabile.

Figura 5 (a) ilustrează una dintre cele mai comune modalități de a compara direct traiectoriile alăturate, adică de a măsura cât de departe este pentru fiecare t și apoi de a calcula media acestor distanțe pentru a obține eroarea medie pe durata prognozei. Aceasta este cunoscută în mod obișnuit ca Eroarea Medie de Deplasare („Average Displacement Error”, ADE) și este raportată de obicei în unități de lungime, de exemplu metri:

$$ADE = \frac{\sum_{n=0}^N \sum_{t=0}^{T_p} \|\hat{p}_t^n - p_t^n\|_2}{N \times T_p} \quad (2)$$

De multe ori, este posibil ca interesul să fie reprezentat de eroarea punctului final al traiectoriei subiectului, ilustrată în Figura 5 (b) (în special, se compară doar punctele \hat{p}_3 și p_3). Aceasta furnizează o măsură a erorii metodei la sfârșitul orizontului de predicție și este denumită

frecvent eroarea de deplasare finală („Final Displacement Error” - FDE). Si ea este de obicei raportată în unități de lungime:

$$FDE = \frac{\sum_{n=0}^N \|\hat{p}_t^n - p_t^n\|_2}{N}, t = T_p, \quad (3)$$

unde N reprezintă numărul de pietoni, T_p numărul pașilor de timp prezis, iar p_t^n și \hat{p}_t^n sunt rezultatul real și rezultatul prezis la pasul de timp t .

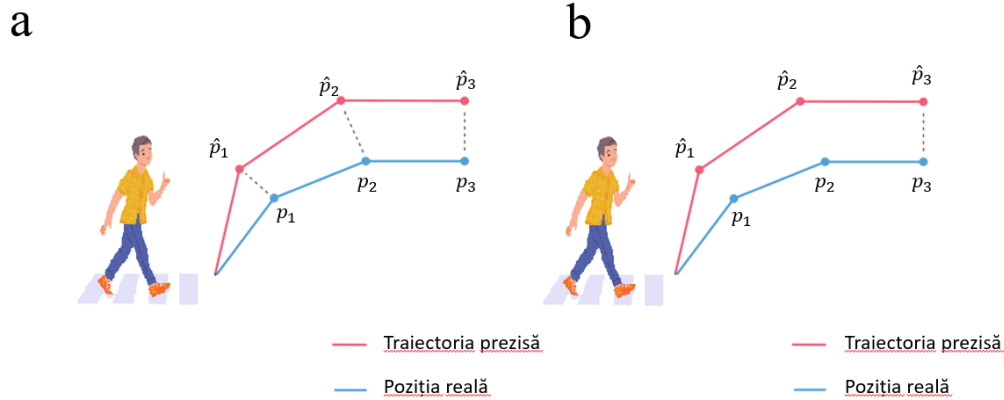


Figura 5. Ilustrații ale metricilor. (a) Eroare de deplasare medie (ADE), (b) Eroare deplasare finală (FDE).

4.4.2. Predicția bazată pe rețele neuronale de tip graf

În cadrul acestui subcapitol este furnizată o descriere a aspectelor cheie ale metodei de estimare propuse. Ulterior, va fi discutată în detaliu concepția fiecărui modul, inclusiv crearea generală a modelului. Scopul principal este de a prezice traiectoriile viitoare pentru numeroși factori care interacționează folosind istoricul pozițiilor și informațiile de context. Metoda de predicție este, de asemenea, extensibilă la cadrele de urmărire multi-țintă [17].

Metoda constă în două componente: rețeaua neurală de convoluție spațio-temporală pe graf (ST-GCNN) și rețeaua neurală de convoluție cu extrapolare temporală (TXP-CNN). Prima componentă utilizează convoluția pentru a extrage caracteristici. Aceste caracteristici oferă o descriere concisă a istoricului traiectoriilor observate. Acestea sunt introduse în TXP-CNN, care le folosește pentru a prezice pozițiile tuturor pietonilor din grup și pentru a extrapola traiectoriile viitoare. Diagrama de ansamblu a acestei metode este ilustrată în Figura 6.

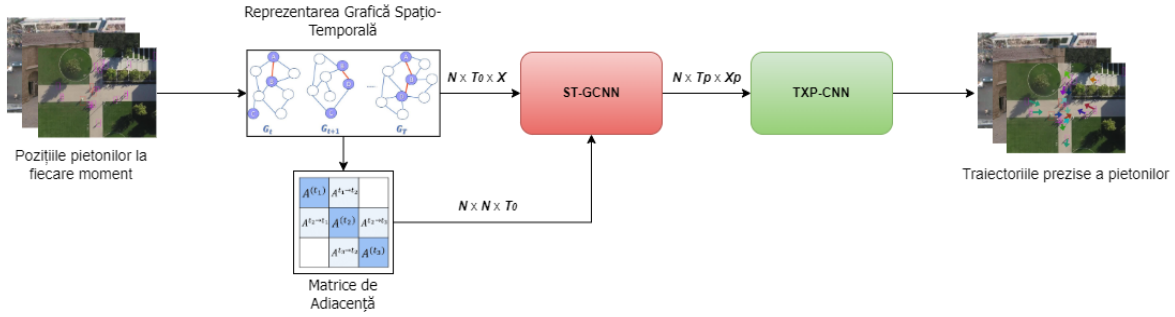


Figura 6. Arhitectura generală a metodei propuse. Prin optimizarea dimensiunii stratului, este posibilă obținerea unei precizii sporite în predicția traiectoriei.

Prima componentă a arhitecturii este ST-GCNN. Este folosită pentru a încorpora reprezentarea obiectului [20]. Ea are rolul de a extrage înglobarea spațio-temporală din graficul de intrare.

A doua componentă utilizată este TXP-CNN. Aceasta lucrează direct pe dimensiunea temporală a înglobării grafice V care crește pentru a satisface cerințele de precizie în predicție. Cea de a doua componentă are mai puțini parametri decât unitățile recurente, deoarece se bazează pe operații de convoluție în spațiul caracteristicilor. Este important de menționat că stratul TXP-CNN nu este invariant la permutare, deoarece variațiile în înglobarea grafică conduc la rezultate diferite. Cu toate acestea, predicțiile sunt invariante dacă ordinea pietonilor este alterată [18].

Întregul proces de predicție a traiectoriei pietonilor este ilustrat în Algoritmul 1.

Algoritmul 1 Rețele Convoluționale cu Graf Dinamic

Date de intrare: coordonatele pietonilor $X = (x, y)$;

Date de ieșire: metricele de evaluare, eroare medie de deplasare (ADE) și eroarea deplasării finale (FDE);

- 1: **for** $t \in [1, T_0]$ **do**
- 2: reprezintă traiectoriile drept un graf: $G_t = (V_t, E_t)$;
- 3: calculează traiectoriile viitoare; $Traj_{obs} = Y_t^i = (x_t^i, y_t^i)$
- 4: **end for**
- 5: creează distanță de distribuție $N \times N$ matrice de adiacență A_t^m utilizând ecuația (5.5.1);
- 6: generează matricea laplaciiană;
- 7: **for** fiecare $i \in 1, \dots, n$ **do**
- 8: **for all** $t \in [1, T_0]$ **do**
- 9: distribuția probabilității pentru traiectoria prezisă: $\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i) \sim \mathcal{N}(\mu_t^i, \sigma_t^i, \rho_t^i)$;
- 10: **end for**
- 11: **end for**
- 12: colectează toate locațiile prezise și cele reale pentru fiecare pieton;
- 13: calculează ADE și FDE;
- 14: **return** ADE și FDE

Figura 7 prezintă arhitectura generală a metodei optimizate. Mai exact, mai întâi este creat arborele de traiectorie pentru a furniza spațiul discret structurat P_{coarse} . Similar, interacțiunea spațială și traiectoria observată sunt codificate una după alta pentru a obține codificarea interacțiunii și codificarea observată.

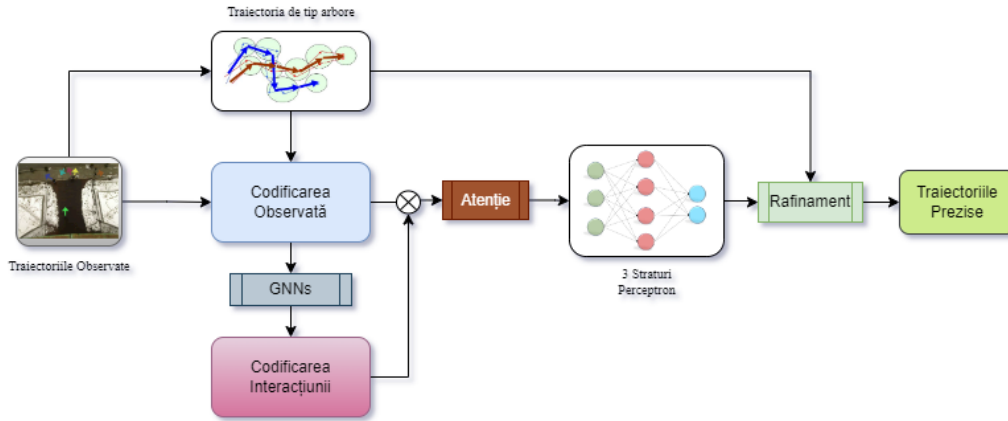


Figura 7. Arhitectura generală a metodei propuse [21].

O altă abordare utilizată în lucrarea [18] a fost implementarea unei arhitecturi de tip codificator-decodor așa cum se poate vedea în Figura 8. Numit “Spatio-Temporal Graph Convolutional Network” (ST-GCNN), este utilizată pentru modelarea interacțiunilor temporale ale traiectoriilor pietonale observate în timp. Aici, o rețea de grafice temporale este utilizată pentru a modela interacțiunile temporale ale unui singur pieton în pași de timp diverși. Graficul temporal al pietonului N este creat și reprezintă pozițiile relative ale acestuia la diferite eșantioane de timp.

Decodorul conține un alt modul, și anume “Time-Extrapolator Convolutional Neural Network” (TXP-CNN), care este folosit pentru a prezice pașii următori. TXP-CNN lucrează direct la dimensiunea temporală a încorporării graficului și o mărește ca o necesitate pentru predicție. TXP-CNN are mai puțini parametri decât unitățile recurente, deoarece se bazează pe operații de convoluție asupra entităților. O caracteristică specială ce privește stratul TXP-CNN este că nu este un invariant la permutare, deoarece modificările în încorporarea graficului chiar înainte de TXP-CNN conduc la rezultate diferite.

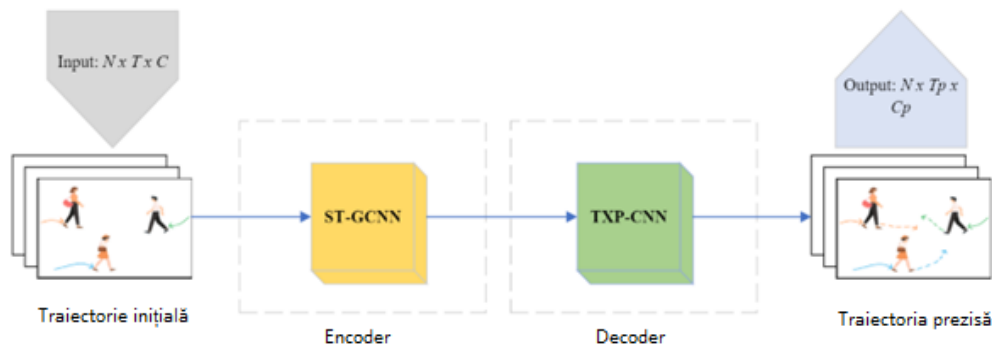


Figura 8. Arhitectura de tip codificator-decodor.

4.5. Rezultate Experimentale

În acest capitol, sunt prezentate experimentele și evaluările efectuate în ceea ce privește predicția traiectoriei pietonilor. Așa cum a fost menționat în Capitolul 5, experimentele au fost realizate folosind metode de predicție a traiectoriei pietonilor bazate în principal pe folosirea rețelelor neuronale profunde. Evaluăm performanța metodelor noastre în comparație cu metode similare de predicție a traiectoriei bazate pe rețele neuronale profunde.

Predicția traiectoriei joacă un rol în estimarea potențialului de risc al pietonilor, deoarece poate fi utilizată și pentru prezicerea traversării, ceea ce este necesar pentru prevenirea coliziunilor. Pentru a evalua aplicabilitatea metodelor de predicție a traiectoriei, evaluăm performanța metodelor pe care le-am propus, în mod cantitativ și calitativ, utilizând seturile de date ETH [22], UCY [23] și Stanford Drone [24], deoarece acestea sunt larg utilizate în literatură, disponibile public, și utilizează coordonate din lumea reală. Utilizarea metodelor bazate pe date necesită disponibilitatea unor date de calitate în cantitate suficientă. Mai exact, în predicția traiectoriei pietonilor, datele disponibile pot fi în două formate diferite: în coordonate de imagine sau în coordonate din lumea reală. Coordonatele de imagine înseamnă că fiecare pieton este reprezentat cu pixelii pe care îi ocupă în imaginea camerei, în timp ce coordonatele din lumea reală înseamnă că fiecare pieton este reprezentat prin poziția sa în metri, cu originea într-un punct arbitrar al lumii. Evaluările calitative sunt efectuate prin vizualizarea traiectoriilor generate.

S-a analizat calitativ modul în care metodele pe care le-am propus capturează interacțiunile sociale între pietoni și le iau în considerare atunci când prezic distribuțiile. Am prezentat cazuri în care D-STGCN [17], PTPCNN [18] și TreeGNN [21] prezic cu succes traiectorii fără coliziuni între pietoni care vin din unghiuri diferite, mențin mersul paralel și prezic corect rezultatul situațiilor în care o persoană se întâlnește cu un grup de pietoni.

4.6. Concluzii

Predicția traiectoriei pietonilor este unul dintre numeroasele domenii în care apariția rețelelor neuronale profunde au schimbat complet abordarea problemei. Înainte de 2016, modelele bazate pe dinamica fizică erau singura modalitate de a prezice cu exactitate traiectoriile viitoare ale pietonilor, dar în anul 2016 s-a demonstrat că învățarea profundă ar putea face acest lucru mai bine. De atunci, au fost propuse diverse arhitecturi noi cu rezultate din ce în ce mai bune. Modelele bazate pe date, cum ar fi modelele folosind rețele neuronale convoluționale, recurente sau de tip graf, sunt foarte sensibile la cantitatea și calitatea datelor de antrenament și la modul în care aceste date sunt apoi integrate modelului.

În această teză, au fost analizate mai multe aspecte, printre care date, senzori și metodele actuale de ultimă generație aplicate pentru problema de predicție a traiectoriei pietonilor. Chiar dacă procesul de dezvoltare a unor soluții fiabile a fost foarte laborios până acum, este totuși necesar să se depună un efort mai mare pentru a realiza un sistem care să asigure securitatea pietonilor pe străzi. Lucrările selectate au fost examinate cu privire la mai mulți factori comuni, pentru a asigura o comparație simplă pentru cei interesați de acest subiect dar și pentru cititorii acestei teze.

Prin utilizarea abordărilor de învățare profundă, sistemele actuale sunt capabile să rezolve mai bine problema PTP. Aceste metode presupun o serie de locații pentru pietoni în ultimele câteva secunde și produc o serie de locații viitoare.

4.6.1. Contribuții

În lucrarea [17] am abordată o nouă tehnică pentru predicția traiectoriei pietonilor, folosind rețele neuronale de tip GNN. Această soluție introduce o abordare bazată pe două componente: o rețea neuronală cu graf spațial (SGNN) pentru modelarea interacțiunii și o rețea neuronală cu graf temporal (TGNN) pentru extracția caracteristicilor de mișcare. SGNN utilizează o metodă de atenție pentru a colecta periodic interacțiuni spațiale între toți pietonii. TGNN folosește și o metodă de atenție, pentru a colecta modelul de mișcare temporală al fiecărui pieton. Cu o dimensiune variabilă mai mică (date și model) și o rată de predicție mai bună, D-STGCN este mai compactă și mai eficientă decât alte soluții SOTA, oferind rezultate experimentale mai bune luând în considerare valorile erorii medii de deplasare (ADE) și erorii finale de deplasare (FDE), (se poate consulta tabelul 1).

Tabelul 1. Rezultate cantitative ale metodelor de ultimă generație pentru seturile de date ETH, UCY și SDD definite în termenii metricilor ADE/FDE. Coloana AWG reprezintă rezultatele medii între scenele seturilor de date ETH-UCY

Metode	SDD	ETH	HOTEL	UNIV	ZARA1	ZARA2	AWG
Metodă propusă [17]	15.18/2 5.50	0.63/ 1.03	0.37/ 0.58	0.46/ 0.78	0.35/ 0.56	0.29/ 0.48	0.42/ 0.68
Metodă propusă [21]	15.84/2 5.17	0.64/ 1.07	2.29/ 1.85	0.41/ 0.67	0.27/ 0.43	0.24/ 0.38	0.77/ 0.88

O altă soluție pe care am publicat-o [21] prezintă o abordare bazată pe arbori GNN pentru a face față provocării de predicție multimodală. Arborele este proiectat pe baza datelor observate și este, de asemenea, utilizat pentru a prezice traiectoriile viitoare. În comparație cu abordările anterioare care utilizează variabile latente implicite pentru a descrie posibile traiectorii viitoare, comportamentele de mișcare pot fi reprezentate direct cu o abordare de tip arbore (de exemplu, mergeți drept și apoi viraj la stânga) și, astfel, sunt oferite de către model traiectorii mai potrivite din punct de vedere social.

Pentru a obține informații despre interacțiune dintre pietoni dar și dintre pietoni și mediul înconjurător am propusă o soluție [18] bazată pe două componente (codificator și decodificator). Această soluție a fost implementată folosind GNN + CNN și au fost obținute rezultate foarte bune în termenii ADE și FDE pentru PTP.

În cadrul acestui articol [25] au fost analizate cele mai recente soluții bazate pe învățare profundă pentru problema predicției traiectoriei pietonilor împreună cu senzorii utilizați (camera, LiDAR și radar) și metodologiile de procesare aferente. S-au prezentat comparativ performanțele

fiecărui senzor auto (radar, LiDAR și cameră) prin evidențierea avantajelor și dezavantajelor acestora în diferite sarcini, de exemplu detecția și clasificarea obiectelor, estimarea distanței, detecția marcajelor rutiere și a limitelor drumului etc. Au fost identificate cele mai importante seturi de date disponibile public, folosite de către cercetători în implementarea soluțiilor de tip PTP și a indicatorilor de performanță utilizați în procesul de evaluare.

În cadrul articolelor pe care le-am publicat [15], [16] au fost descrise metodele pentru predicția bazată pe filtrul alfa-beta-gama. S-a efectuat o analogie între filtrul Kalman și filtrul alfa-beta-gama cu scopul de a identifica acurateței metodei. Rezultatele au arătat că utilizarea unor astfel de arhitecturi pot oferi o soluție fiabilă pentru PTP.

Bibliografie selectivă

- [1] S. Lefèvre et al., „A survey on motion prediction and risk assessment for intelligent vehicles”, *Robomech J*, 2014.
- [2] European Road Safety Observatory, „Annual Accident Report”, 2020.
- [3] WHO, „Global Status Report on Road Safety”, 2018.
- [4] ITF, *Pedestrian Safety, Urban Space and Health*, 2012.
- [5] D. Gálvez-Pérez et al., “The Influence of Built Environment Factors on Elderly Pedestrian Road Safety in Cities: The Experience of Madrid”, *Int. J. Environ. Res. Public Health*, Vol. 19, 2022.
- [6] T. Winkle, „Safety benefits of automated vehicles: Extended findings from accident research for development, validation, and testing”, *Autonomous Driving*. Springer, 2016.
- [7] S. Ahmed et al., „Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey”, *Applied Sciences*, Vol. 9, 2019.
- [8] I. N. Aizenberg, N. N. Aizenberg, and J. P. Vandewalle, “Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications”., Kluwer Academic Publishers, 2000.
- [9] B. I. Sighencea, R. I. Stanciu and C. D. Căleanu, „A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction”. *Sensors*, 21(22):7543, 2021.
- [10] J. Ziegler, et al., „Making Bertha Drive—An Autonomous Journey on a Historic Route”. *IEEE Intell. Transp. Syst. Mag.*, 6, 8–20, 2014.
- [11] C. Guo, et al., „Cooperation between driver and automated driving system: Implementation and evaluation”. *Transp. Res. Part F Traffic Psychol. Behav.*, 61, 314–325, 2019.
- [12] F. M. Ortiz, et al., „Vehicle Telematics via Exteroceptive Sensors: A Survey”. *arXiv:2008.12632*, 2020.
- [13] Yole Developpement. *MEMS and Sensors for Automotive: Market & Technology Report*. 2017. Disponibil online: <https://bit.ly/2X5pL70> (accessed on 23 July 2021).
- [14] J. Kocic, et al., „Sensors and sensor fusion in autonomous vehicles”, 2018.
- [15] B. I. Sighencea, I. Rares Stanciu and C. Sorandaru, "Using the α - β - γ Filter to solve the Threshold Problem," *IEEE EUROCON 2021 - 19th International Conference on Smart Technologies*, pp. 45-50, 2021.
- [16] B. I. Sighencea, R. I. Stanciu, C. Șorândaru and C. D. Căleanu, „The Alpha-Beta Family of Filters to Solve the Threshold Problem: A Comparison”. *Mathematics*, 10, 880, 2022.

- [17] B. I. Sighencea, I. R. Stanciu and C. D. Căleanu, „D-STGCN: Dynamic Pedestrian Trajectory Prediction Using Spatio-Temporal Graph Convolutional Networks”. *Electronics*, 12, 611, 2023.
- [18] B. I. Sighencea, R. I. Stanciu and C. D. Căleanu, "Pedestrian Trajectory Prediction in Graph Representation Using Convolutional Neural Networks," 2022 IEEE 16th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 2022, pp. 000243-000248.
- [19] J. Amirian, et al., „Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories with GANs”. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2964–2972, 16–17 June 2019.
- [20] A. Mohamed, et al., „Social-STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction”. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14412–14420, 16–18 June 2020.
- [21] B. I. Sighencea, "Pedestrian Trajectory Prediction Based on Tree Method using Graph Neural Networks," 24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 245-249, 2022.
- [22] S. Pellegrini, „You’ll never walk alone: Modeling social behavior for multi-target tracking”. In *Proceedings of the IEEE 12th International Conference on Computer Vision*, pp. 261–268, 27 September–4 October 2009.
- [23] A. Lerner, „Crowds by example”. *Comput. Graph. Forum*, 26, 655–664, 2007.
- [24] A. Robicquet, et al. „Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes”. In *Computer Vision–ECCV*, Springer, Volume 9912, 2016.
- [25] B. I. Sighencea, R. I. Stanciu and C. D. Căleanu, „A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction”. *Sensors*, 21(22):7543, 2021.