Universitatea Babeş-Bolyai
Facultatea de Matematică şi Informatică, Departamentul de Informatică
Institut National des Sciences Appliquees de Rouen
Laboratoire d'Informatique de Traitement de l'Information et des Systemes

Învatarea Profunda Incrucisata Pentru Estimarea Riscului Pietonilor

Teza de doctorat
Rezumat

Doctorand:
Ing. Dănuţ Ovidiu POP

Conducatori de doctorat:

Dr. Horia F POP	Profesor	UBB Cluj-Napoca
Dr. Abdelaziz BENSRHAIR	Profesor	INSA de Rouen

Consilieri de doctorat:

Dr. Fawzi NASHASHIBI	Doctor HDR	INRIA Paris
Dr. Alexandrina ROGOZAN	Profesor Asociat	INSA de Rouen

8 Noiembrie 2019

Cuprins
Teza

Cuprins
Rezumat

Cuvinte Cheie

- Detectie de Pietoni;
- Recunoastere de Actiuni;
- Predictie de Actiuni;
- Clasificare de Pietoni;
- Cross-Modality Learning;
- Deep Learning;
- Convolutional Neural Network

Introducere

Acesta teza de doctorat este rezultatul activității mele de cercetare în domeniul transportului inteligent pentru a dezvolta/investiga un sistem de protecție pentru pietoni cu mai multe functionalitati (PPS), care sa includ nu numai clasificarea, detectarea și urmărirea pietonilor, ci și clasificarea și predicția de acțiune ale pietoniloe si in final sa estimeze riscul pietonilor. Scopul cercetării noastre este de a dezvolta un sistem inteligent de protecție pentru pietoni bazat doar pe un singur sistem de viziune stereo folosind o modalitate încrucișată (deep learning) de învățare profundă. Sistemul trebuie să fie capabil să nu detecteze doar toți pietonii cu o precizie inalta, dar și sa fie capabil sa urmărieasca toate căile pietonale, pentru a clasifica acțiune pietonilor și pentru a prezice următoarele lui acțiuni și, în final, pentru a estima riscul pietonilor, timpul de traversare pentru fiecare pieton. Pentru acest lucru noi am dezvoltat uratorul plan:

1. In primul rand, investigăm componenta de clasificare unde analizam cum învățarea reprezentărilor dintr-o modalitate ar permite recunoașterea repezentarilor in alta modalitate/modalitati utilizand diferite abordări de învățare profundă, denumită modalitate de învăţare încrucișată.

   Schema de fuziune tarzie conectata cu învățarea CNN-lui este profund investigată pentru recunoașterea pietonilor bazată pe setul de date stereo viziune Daimler. Prin urmare, un CNN independent pentru fiecare modalitate imagistică (Intensitate, Adâncime și Optical Flow) este utilizat înainte de fuzionarea scorurilor de ieșire probabilistică ale CNN cu un Perceptron cu mai multe straturi care oferă decizia finala de recunoaștere a pietonilor.

   Noi propunem patru modele diferite de învățare ale rețelelor neuronale convoluționale bazate pe modalitatea profundă de invatare încrucișara:

   (a) o învățare particulară a modalităților încrucișate;

   (b) o învățare separată a modalităților încrucișate;

   (c) o învățare corelată cu modalități încrucișate;

   (d) un model incremental de învățare cu modalități încrucișate.

   Mai mult, proiectăm și o nouă arhitectură CNN, numită LeNet +, care inbunatateste performanța clasificării, nu numai pentru fiecare clasificator de modalități, dar și pentru schema multifuncțională de fuziune tarzie. În cele din urmă, propunem să utilizam modelul LeNet + cu modelul abordării incrementale a modalității încrucișate utilizând setări optime de învățare, obținute cu un model de validare încrucișată de K-fold.

   Această metodă depășește clasificatorul furnizat de setul de date Daimler pentru ambele seturi de date: pietoni nonocluzi și pietoni parțial ocluzi.

2. În al doilea rând, studiem modul în care învățarea modalităților încrucișate îmbunătățește un model end-to-end de detectare a pietonilor ai actinuilor acestor. Studiem modul în care învățarea modalităților încrucișate îmbunătățește îmbunătățește un model end-to-end de detectare a pietonilor ai actinuilor acestor. Ne concentrăm atât pe detectarea pietonilor cât și pe recunoastrea acțiunilor pietonilor bazată setdul de date Joint Attention for Autonomous Driving (JAAD), aplicând abordări de învățare profundă.

   Obiectivul principal al acestei abordări este de a afla dacă trece un pieton sau dacă acțiunea pietonului prezintă o situație critică. Cel mai crucial caz pentru pieton și șoferi este atunci când pietonul trece strada print fața vehiculului și soferul nu opri mașina sau evita in timp pietonul.

   Introduce o componentă de detectare a pietonilor unificată bazată pe învățare profundă care recunoaște, de asemenea, diferite acțiuni ale pietonilor; acest lucru este în contrast cu obișnuite metode de detectare a pietonilor, care deosebeste doar pietonii și nepietoni printre ceilalți participanți la traficul rutier.

   Pentru acest lucru, definim patru acțiuni principale pentru pietoni, pentru a afla dacă actinuea pietonul prezintă o situație riscantă.

   (a) pietonul se pregătește să traverseze strada;

   (b) pietonul traversează strada;

   (c) pietonul urmează să traverseze strada;

   (d) intenția pietonului este ambiguă.

3. În al treilea rând, analizăm predicția acțiunii pietonale și estimarea timpului de traversa a strazi.

   Sistemul de detectare a pietonilor este una dintre componentele vitale ale sistemul avansat de asistență a șoferului, deoarece contribuie la siguranța rutieră. În acest caz, securitatea participanților la trafic ar putea fi îmbunătățită în mod semnificativ daca sistemul ar putea recunoaște și prezice acțiuni pietonale sau chiar poate estima timpul de trecere pentru fiecare pieton. În acest capitol, ne concentrăm pe predicția acțiunilor pietonale și estimăm peste cat timp pietonul va traversa strada. Am utilizat setul de date Joint Attention for Autonomous Driving (JAAD), aplicând abordări de învățare profundă.

   Noi propunem:

   (a) o predicție a acțiunilor pietonilor folosind o rețea de învățare profundă recurentă pentru a prezice următoarele acțiuni ale pietonului pe scurt (T + 1,T + 2, T + 3, T + 4, T + 5), mediu (T + 14) și timp îndelungat (T + 40);

   (b) o estimare a timpului pentru a traversa un singur și mai mulți pietoni, aplicatie care utilizează rețea de învățare profundă recurentă.

   Folosim o retea recurenta numita Long Short-Term Memory (LSTM) pentru a estima actinuea pietonul folosind 5, 14 și, respectiv, 40 de cadre anterioare. Demonstram ca integrarea mai multor etichete pietonale pentru detectare, combinată cu LSTM, poate sa obțina o performanță semnificativă.

Rezumat

Detectarea pietonilor este o problemă extrem de dezbătută în comunitatea științifică, datorita importanței sale majore utilizat in multe aplicații, în special în domeniile de siguranță auto, robotică și supraveghere. În ciuda diferitelor metode dezvoltate în ultimii ani, detectarea pietonilor este încă o provocare deschisă a cărei precizia și robustetea trebuie îmbunătățite.

Un sistem de detectare a pietonilor are trei componente principale: senzorii utilizați pentru a captura datele vizuale, modalitatea - componentele de procesare a imaginilor și componente de clasificare. În general, toate aceste componente sunt procesate și dezvoltate împreună pentru a obține o performanță ridicată de detectare, dar uneori fiecare elementar putea fi investigat separat în funcție de solicitarile aplicatiei.

În **Capitolul 1** ne preocupăm de îmbunătățirea sarcinii de clasificare, care este partea centrală a detectorului pietonal. Schema de fuziune tarzie conectata de reteaua CNN este profund investigată pentru recunoașterea pietonilor bazată pe setul de date stereo viziune Daimler. Astfel, un CNN independent pentru fiecare modalitate imagistică (intensitate, adâncime și flux optic) este utilizat înainte de fuziunea scorurilor de ieșire probabilistică ale CNN cu un strat multiplu Perceptron care oferă decizia de recunoaștere. Pentru a atinge acest obiectiv, dezvoltăm următoarea metodologie bazată pe patru CNN-uri

1. Lenet [LBBH98], deoarece este o arhitectură simplă și care permite sa funcționeaze mai bine chiar și pe un procesor (CPU) (folosind dimensiuni mici de imagine, implicit este 32x32pixeli);

2. Lenet +, care noi am propuso si care este o variație a Lenet și îmbunătățește clasificarea pietonilor pentru fiecare clasificator de modalități utilizat;

3. AlexNet pentru impactul său incontestabil asupra învățării, datorită unui echilibru dintre arhitectură și performanță compactă;

4. VGG-16 [SZ14] datorită performanței ridicate obținute cu o arhitectură vastă în mod obișnuit utilizat în detectarea pietonilor

Pentru aceasta, am urmat procedura de mai jos, bazându-ne pe o abordare de învățare profundă:

- Analizam performanța retelelor CNN AlexNet și LeNet utilizand setul de date Caltech in care folosim doar casetele de delimitare a pietonilor (bounding

box-BB) care sunt mai mari de 50 pixeli. Toate BB au fost redimensiuni la dimensiuni patratice (64 x 64 pixeli);

- Combinarea a trei modalități de imagine (intensitate, adâncime și flux optic) pentru a antrena o rețea neurală convoluțională (CNN), folosind o metodă de fuziune timpurie și fuzionarea rezultatelor a trei CNN-uri independente, cu o metodă de fuziunea târzie;

- Evaluarea arhitecturii LeNet cu diverși algoritmi de învățare și politicilor ratei de învățare folosind metoda clasică de învățare;

- Propunem o metodă particulară de învățare a modalităților încrucișate în care un CNN este instruit și validat pe aceeași modalitate de imagine, dar testat pe alta modalitate imagistica;

- Propunem o metoda de învățare separată a modalității încrucișate care folosește o modalitate de imagine diferită pentru antrenament decât pentru validare;

- Propunem o metodă de învățare corelată cu modalități încrucișate unde se învață un CNN unic (antrenare și validat) cu imagini de intensitate, adâncime și, respectiv, fluxuri optice pentru fiecare cadru;

- Propunem o metoda de invatare incrusisata incrementala (Incremental Cross-Modality learning) în care un CNN este învățat cu primele cadre de modalitate de imagini, apoi un al doilea CNN, inițializat cu informatiile sinapselor de la primul CNN, este învățat pe al doilea set de imagine, modalitate de imagini diferite, iar în final un al treilea CNN inițiat cu inforatiile sinapselor de al doilea CNN, este învățat pe ultimele cadre de modalitate de imagine;

- Propunem sa îmbunătățind învățarea încrucișată print intermediul unei noi arhitecturi CNN (Lenet +) pe care am propus-o, antrenat si cu metoda K-fold-Validation;

- Antrenam retele AlexNet și VGG-16 folosind setările implicite cu metoda de învățare clasică și, respectiv, cu metoda de învățare profundă a modalității încrucișate incrementala (incremental cross modality deep learning);

- Optimizam hiperparametrii CNN-lui (pas de convoluție, dimensiunea nucleului, numărul de convoluții de ieșiri, ponderile straturilor complet conectate) pentru metoda de învățare clasică și, respectiv, pentru metoda de învățare profundă a modalității încrucișate;

- Implementam schema de fuziune tarzie folosind un Multi-Layer Perceptron (MLP) atât pentru metodele clasice, cât și pentru cele incrementale amintite mai sus.

Analizăm diferiți algoritmi si rate de învățare folosind arhitectura LeNet. Demonstram că clasificatorul care utilzeaza fuziune tarzie depășește nu numai toate modalitățile unice, ci și clasificatorul de fuziune timpurie.

De asemenea, examinăm toate metode noatre de clasificare cu cea de învățare clasică în care fiecare CNN este învățat și evaluat pe aceeași modalitate de imagine. De asemenea, comparăm toate aceste modele de învățare cu abordările clasice

de învățare din cadrul MoE propus în [EESG10, EG11] și deep Boltzmann-Machine [OW12] pentru recunoașterea atât a pietonilor parțial oclusi, cât și a celor neoclusi.

În **Capitolul 2**, ne concentrăm atât pe detectarea pietonilor, cât și pe recunoașterea acțiunilor pietonale, bazată pe setul de date Joint Attention for Autonomous Driving (JAAD) [KRT16], aplicând abordări de învățare profundă (deep learning). Obiectivul principal al acestei abordări este de a afla dacă trece un pieton sau dacă acțiunea pietonului prezintă o situație critică. Cazul cel mai crucial pentru pieton și șoferi este atunci când pietonul traversează strada din fața vehiculului, iar soferul nu poate opri masinsa sau sa evita la timp pietonul.

Vom introduce o componentă de detectare a pietonilor bazată pe învățare profundă, care recunoaște, de asemenea, diferite acțiuni pietonale; acest lucru este în contrast cu metodele obișnuite de detectare a pietonilor, care discriminează doar pietonii și ne-pietonii dintre ceilalți participanți la trafic.

Pentru acest lucru, definim patru acțiuni principale pentru pietoni pentru a afla dacă acțiunea pietonului prezintă o situație riscantă:

1. Pietonul se pregătește să traverseze strada (Pedestrian is Preparing to Cross the street; PPC), unde pietonul se plimbă/stă în picioare, acordă atenție sau nu traficului rutier, isi schimbă sau nu comportamentul înainte de traversare. În acest caz, acțiunile ar putea fi: se mișca, priveste, sta, da din cap, face cu mana, încetineste și în final traversarea străzii. Luăm în considerare toate acțiunile până la evenimentul cand prietonul incepe sa treaca strada ca si clasa PPC. În acest caz, pietonul trece cu siguranță strada după aceste acțiuni.

2. Pietonul traverseaza strada (Pedestrian is Crossing the street; PC), unde pietonul este observat din punctul in care incepe sa traverseze până in momentul cand pietonul a traversat drumul. În acest caz, este obligatorie acțiunea de traversare în timpul acestui eveniment, dar nu este obligatoriu să existe un eveniment specific înainte de evenimentul încrucișat. Există secvențe video în care pietonii sunt adnotați doar din punctul de a traversa strada. Pietonul ar putea manifesta chiar și alte acțiuni precum privirea, face cu mana, accelerarea, a da din cao, încetinirea, privire in stanga sau dreapta în timpul acestui eveniment.

3. Pietonul este pe punctul de a traversa strada (Pedestrian is About to Cross the stree; PAC), unde pietonul este pe punctul de a traversa și acordă atenție trafficului rutier și răspunde în funcție de eveniment. În acest caz, acțiunile ar putea fi: mișcarea, privirea sper autovehicol, sta, face cu mana, da din cap, încetineste, dar nu vor traversa strada. Pietonul in acest caz nu traversează strada după aceste acțiuni.

4. Intenția pietonului este ambiguă (Pedestrian intention is Ambiguous; PA), unde pietonul se plimbă/stă, iar intenția sa este ambiguă. În acest caz, acțiunile ar putea fi: mișcarea, privire spre masina, sta, accelerarea. Avem în vedere toate acțiunile după ce pietonul traversează strada. În acest caz, pietonul a traversat drumul sau manifesta alte actiuni care nu prezintă o situație de risc.

Am examinat partea de detecție prin aplicarea unui detector de obiecte generic si public RetineNet [LGG+17] și Faster R CNN. Am gestionat arhitecturile Resnet50 [HZRS15] și Inception V2 CNN pentru sarcina de clasificare cu folosing codul public

Keras descris în [LGG+17] și Tensorflow descris în [LGG+17]. Tot procesul de instruire se bazează pe setul de date JAAD [RKT17].

Setul de date Jaad oferă doar modalitatea de imagine RGB. Pentru a aplica InCML, trebuie să extragem modalitatea de imagine Adâncime și Flux optic și apoi să aplicăm un MLP cu pas de fuziune târzie.

Pentru aceasta problema, dezvoltăm următoarea metodologie bazându-ne pe o abordare de învățare profundă:

- Antrenam setul de date JAAD cu tagul folosing doar clasa pieton cu RetinaNet [LGG+17] pentru detectarea pietonilor;
- Împărțim setul de date (JAAD) [KRT16] în patru clase: pietonul se pregătește să traverseze strada, pietonul traversează strada, pietonul este pe punctul de a traversa strada, iar intenția pietonilor este ambiguă .
- Extragem fluxul optic și adâncimea din setul de date JAAD.
- Antrenam RetinaNet cu toate eșantioanele pietonilor folosind clasele de acțiuni ale pietonilor menționate mai sus folosind modaliatea imagistica RGB, fluxul optic și adâncimea pentru detectarea și clasificarea acțiunilor;
- Antrenam RetinaNet impreuna cu metoda InCML pentru detectarea pietonilor și recunoașterea acțiunilor pietonilor;

InCML a depășit abordarea clasică de detectare pe toate modalitățile, dar performanța sa este semnificativă statistic numai pentru modalitatea de imagine RGB. Am observat că performanța detectorului InCML este direct proporțională cu realizările fiecărei acțiuni de detectare a pietonilor.

Capitolul 3 se referă la rezolvarea predicției acțiunii pietonilor și estimarea timpului de trecere pentru mai multi pietoni utilizand un de model de învățare profundă (deep learning).

Estimarea timpul de trecere a pietonilor este mai dificil decât prezicerea acțiunii pietonilor, deoarece necesită o analiză fină a întregii scene, precum și o analiză fină a mișcării pietonilor. Subliniem că această sarcină este o provocare chiar și pentru ființele umane.

Dificultatea de a rezolva această problemă vine din lipsa bazelor de date publică adnotate. Prin urmare, există putine baze de date publica adnotată cu timpul de trecere a pietonilor, în timp ce există mai multe baze de date uriașe interesante de detectare a pietonilor (Kitti, Caltech, printre altele). Problema este că aceste baze de date nu oferă nici o etichetă de acțiune pentru pietoni. Singura informație publică setată cu etichete de acțiune pentru pietoni în traficul din mediu urban este JAAD [KRT16]. Deoarece acest set de date nu oferă în mod direct adnotările pentru trecerea pietonilor, acesta limitare ne oblica sa îl determinăm noi pentru fiecare traiectorie pietonală (secvențe cadru). Prin urma am selectat câteva indicii din setul de date publice JAAD [KRT16] pentru a rezolva această problemă, apoi am făcut adnotarea TTC pentru pietele video.

Abordarea convențională pentru rezolvarea dificultății predicției comportamentului pietonal este de a utiliza cel puțin unul dintre elementele dinamice care contribuie la percepția situațiilor de comportament pietonal, cum ar fi traiectoria [HTDD18] sau viteza [SG13] sau să anticipezi destinația finală a pietonilor [RWLS18]. Mai mult, pentru a obține o performanță ridicată de predicție a mișcării pietonilor, este

obligatoriu să se țină seama de informațiile de context temporal pentru a ajuta la predicția comportamentului pietonului.

Problema de predicție este grupată în două categorii:

1. Scenarii de coliziune, scenari de evitare (modelare pe termen scurt), în care obiectivul este să reacționeze cu manevrele de urgență pentru obiecte. Orizontul de predicție este aici maxim. 1-2 secunde [**?**, RRL$^+$18].

2. Modelare pe termen lung, în care scopul este de a avea un comportament de conducere mai confortabil. Orizontul de predicție este de 2+ secunde, în funcție de viteza și mediul vehiculului [RK15].

Ne concentrăm pe abordarea de predicție pe termen lung și pe termen scurt atât a poziției pietonale, cât și a acțiunii folosind un LSTM (folosind cadrele următoare, T + 1, T + 2, T + 3, T + 4, T + 5, T + 14, T +40) pentru a ține cont de informațiile de context temporal (cadre anterioare din T-5, T-14 și T-40). Intrările LSTM sunt coordonatele casetei de legătură 2D (bounding box; BB) furnizate de componenta de detecție menționată mai sus.

Ori de câte ori se aplică metoda de detectare a pietonilor, datele de intrare LSTM sunt etichetele pietonale (clasa etichetă) și coordonatele BB care anticipează următoarele cadre după coordonatele BB ale pietonilor și comportamentul acesteia.

Estimarea timpului de trecere pentru fiecare pieton este esențială pentru sistemele ADAS, deoarece ar putea prezice dacă și când ar putea exista o situație riscantă.

Din punct de vedere al învățării automate, estimarea TTC poate fi considerată o problemă de regresie, unde ne propunem să estimăm un număr întreg sau o valoare reală (indiferent dacă avem în vedere un număr de cadre sau un timp în secundă) pentru fiecare cadru al unui videoclip. Deoarece dinamica semnalului este esențială pentru a estima eficient TTC, am apelat în mod natural la utilizarea unei rețele neuronale recurente pentru a capta contextul mișcării. Printre modelele recurente, am ales să folosim LSTM-uri care și-au arătat eficiența în multe probleme de analiză a secvenței.

Pentru a prezice timpul de trecere a pietonilor, am propus două abordări:

- estimare individuală pentru fiecare secvență BB pietonală furnizată de detectorul pietonal (folosind doar eșantioane PPC)

- estimări multiple pentru toți pietonii detectați (folosind toate eșantioanele).

Subliniem faptul că componentele de detectare și predicție sunt învățate independent.

Etapa de detectare se bazează pe RetinaNet și are ca intrare toate imaginile RGB și returnează caseta de delimitare corespunzătoare pietonului și eticheta ei de acțiune.

Modelul de predicție se bazează pe LSTM și are coordonatele casetei de delimitare 2D (BB) ca date de intrare furnizate de componenta de detectare. Ieșirea constă în estimarea timpului până la momentul trecerii a strazii pentru fiecare pieton și se informeaza peste câte cadre va trece pietonul. Luăm în considerare informațiile de context temporal pentru cadrele anterioare din T-5, T-14 și T-40.

Prima metodă TTC returnează o performanță mai bună decât cea de-a doua, dar o considerăm pe cea de-a doua promisă, deoarece este mai realistă.

Concluzii

În această teză, ne-am concentrat pe dezvoltarea unui sistem de protecție pietonală cu multiple sarcini (PPS), care este o funcție esențială a sistemelor avansate de asistență la șoferi (ADAS), deoarece poate reduce accidentele de circulație prin asistarea șoferului și chiar oprirea vehiculului pentru a preveni accidentele iminente. Sistemul nostru PPS include nu numai clasificarea, detectarea și urmărirea pietonilor, dar și clasificarea și predicția unităților de acțiune pentru pietoni și, în final, estimeaza riscului pentru pietoni (timpul de trecere). Această problemă particulară a fost rezolvată prin utilizarea abordărilor originale de învățare profundă a modalității încrucișate.

În capitolul 1, am introdus diferite metode de învățare bazate pe învățarea profundă a modalității încrucișate a rețelelor neuronale convolutive (CNN):

În capitolul 2, am abordat mai multe probleme legat de detectia pietonilor:

- Am aplicat învățarea profundă a modalității încrucișate incrementale pe metoda de detectare (InCML);
- Am aflat dacă trece un pieton și dacă acțiunea pietonului prezintă o situație critică, unde am definit patru acțiuni principale pentru pietoni:
    1. pietonul se pregătește să traverseze strada;
    2. pietonul traversează strada;
    3. pietonul este pe punctul de a traversa strada;
    4. intenția pietonului este ambiguă;
- Am introdus o componentă de detectare pietonală unificată bazată pe învățare profundă a modalității încrucișate, care, de asemenea, recunoaște diferite acțiuni pietonale.

Învățarea profundă incrementală încrucișata (InCML) a depășit abordarea clasică de detectare pe toate modalitățile, dar performanțele sale sunt semnificative statistic doar pentru modalitatea de imagine RGB. Am observat că performanța detectorului de învățare profundă a modalității încrucișate este direct proporțională cu realizările fiecărei detectări a acțiunilor pietonale.

Am extins componenta de detectare a pietonilor folosind învățarea profundă incrementala încrucișata (InCML), luând în considerare contextul temporal pentru a prezice următoarea acțiune pentru pietoni. Am analizat această problemă în cea

de-a treia parte a cercetării noastre fără a folosi învățarea profundă a modalității încrucișate.

În capitolul 3, am combinat componenta de detectare a pietonilor cu predicția de acțiune pietonală și estimarea timpului de traversare.

Am dezvoltat o predicție a acțiunilor pietonilor, utilizând o estimare a timpului pentru a traversa un singur și mai mulți pietoni folosind o memorie pe termen scurt (LSTM)

Am utilizat o memorie pe termen scurt (LSTM) [HS97] pentru a estima acțiunea intenției pietonilor folosind precedentele 5, 14 și, respectiv, 40 de cadre ca pași de timp. Am arătat că integrarea mai multor etichete pietonale pentru partea de detecție și îmbinarea cu LSTM, poate obține o performanță semnificativă.

Pentru lucrările viitoare, intenționăm să creăm un timp de estimare a detectorului de aplecare profundă a modalității încrucișate end-to-end, care să poată face toate funcționalitățile într-o singură etapă (detectare, recunoaștere a acțiunii, predicție de acțiune, estimare a timpului de traversat). În plus, intenționăm să aplicăm modelul incremental încrucișat pentru clasificarea și detectarea altor obiecte rutiere (semne de circulație și semafoare), precum și a utilizatorilor rutieri (vehicule, bicicliști).

Bibliography

[EESG10] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila. Multi-cue pedestrian classification with partial occlusion handling. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 990–997, June 2010. 5

[EG11] M. Enzweiler and D. M. Gavrila. A multilevel mixture-of-experts framework for pedestrian classification. *IEEE Transactions on Image Processing*, 20(10):2967–2979, Oct 2011. 5

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997. 10

[HTDD18] M. Hoy, Z. Tu, K. Dang, and J. Dauwels. Learning to predict pedestrian intention via variational tracking networks. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3132–3137, Nov 2018. 6

[HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 5

[KRT16] Iuliia Kotseruba, Amir Rasouli, and John K. Tsotsos. Joint attention in autonomous driving (JAAD). *CoRR*, abs/1609.04741, 2016. 5, 6

[LGG+17] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *CoRR*, abs/1708.02002, 2017. 5, 6

[OW12] W. Ouyang and X. Wang. A discriminative deep model for pedestrian detection with occlusion handling. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3258–3265, June 2012. 5

[RK15] E. Rehder and H. Kloeden. Goal-directed pedestrian prediction. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 139–147, Dec 2015. 7

[RKT17] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017. 6

[RRL+18] D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf. A literature review on the prediction of pedestrian behavior in urban scenarios. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3105–3112, Nov 2018. 7

[RWLS18] E. Rehder, F. Wirth, M. Lauer, and C. Stiller. Pedestrian prediction by planning using deep neural networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–5, May 2018. 6

[SG13] Nicolas Schneider and Dariu M. Gavrila. Pedestrian path prediction with recursive bayesian filters: A comparative study. In Joachim Weickert, Matthias Hein, and Bernt Schiele, editors, *Pattern Recognition*, pages 174–183, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 6