UNIVERSITATEA POLITEHNICA BUCUREȘTI

Facultatea de Automatică și Calculatoare

Teză de doctorat

David-Traian IANCU

DETECȚIE, SEGMENTARE ȘI PREDICȚIE IN CONDUCERE AUTONOMĂ

COMISIA DE EVALUARE

Prof. Dr. Ing. Florin Pop University Politehnica of Bucharest	Președinte
Prof. Dr. Ing. Adina Magda Florea University Politehnica of Bucharest	Conducător de doctorat
Prof. Dr. Ing. Vasile Manta Technical University Gheorghe Asachi Iasi	Membru
Prof. Dr. Ing. Viorel Negru West University of Timișoara	Membru
Prof. Dr. Ing. Mariana Mocanu University Politehnica of Bucharest	Membru

BUCUREȘTI 2022

Mulțumiri

Aș dori să mulțumesc în primul rând coordonatorului meu de teză, care a făcut posibilă această cercetare, propunându-mi subiectul de cercetare și oferindu-mi îndrumările necesare de-a lungul tuturor acestor ani. De asemenea, aș dori să mulțumesc Universității Politehnica din București, care a oferit infrastructura necesară pentru realizarea unora dintre experimentele prezentate în această teză și, de asemenea, pentru că mi-a oferit un loc de muncă în cadrul echipei de cercetare. Aș dori să mulțumesc echipei de la Laboratorul AIMAS, care a contribuit la unele dintre lucrările de cercetare publicate pentru această cercetare și m-a ajutat cu camerele video și seturile de date. Actuala teză a fost posibilă datorită granturilor obținute de universitate în proiectele de cercetare asociate.

Rezumat

Conducerea autonomă a devenit una dintre cele mai importante provocări în ceea ce privește cercetarea de astăzi în viziunea computerizată și inteligența artificială. În dezvoltarea sa, companiile private au beneficiat de ajutorul comunității de cercetare pentru a dezvolta algoritmi mai buni. O mașină autonomă va avea avantaje evidente în viața de zi cu zi, dar pentru moment mai sunt multe de făcut pentru a dezvolta o mașină complet autonomă, funcțională și sigură. O mașină autonomă tipică are o mulțime de componente în ceea ce privește înțelegerea scenei, luarea deciziilor și controlul vehiculului. Această teză studiază în detaliu înțelegerea scenei cu privire la mașinile autonome și se concentrează pe unele dintre cele mai relevante sarcini privind înțelegerea scenei: detectarea obiectelor, urmărirea obiectelor, segmentarea semantică și a instanțelor și estimarea adâncimii. La limita dintre înțelegerea scenei și luarea deciziilor se află sarcina de predicție a traiectoriei mașinilor din jur, care se bazează pe înțelegerea scenei, dar este utilă în procesul de luare a deciziilor. Scopul tezei este de a analiza sarcinile de înțelegere a scenei și de a le folosi pentru a proiecta un nou algoritm de predicție a traiectoriei. Noutatea este că sarcina de predicție a traiectoriei este realizată folosind generarea video - o abordare niciodată întâlnită în literatură.

Teza analizează sarcinile de detectare și urmărire a obiectelor, segmentare semantică și a instanțelor, estimarea adâncimii și predicția traiectoriei, făcând o analiză cuprinzătoare a celor mai importante lucrări și seturi de date din acest moment. Pentru fiecare sarcină, au fost testate unele dintre cele mai bune arhitecturi existente. De asemenea, pentru fiecare dintre aceste sarcini au fost efectuate multiple experimente pe noi seturi de date înregistrate în campusul Universității Politehnica București, ținând cont de mai mulți parametri precum dimensiunea mașinilor sau ora din zi, pentru a vedea care dintre cele testate. arhitecturile funcționează cel mai bine într-un scenariu din viața reală. Scopul final este de a găsi cele mai bune arhitecturi care pot fi combinate pentru sarcina de predicție a traiectoriei folosind generarea video. În final, pentru sarcina de predicție a traiectoriei, teza propune o nouă metodă bazată pe detectarea obiectelor, segmentarea drumului, estimarea adâncimii și generarea video. De asemenea, propune trei noi variante de arhitectură de generare video cu rezultate mai bune pentru sarcina de predicție a traiectoriei. Cel mai mare avantaj este că, chiar dacă sarcina de generare video este mai complexă, elimină necesitatea unei traiectorii adnotate manual, care poate fi o sarcină

foarte laborioasă. În schimb, un algoritm de generare video ar putea fi antrenat cu orice videoclip posibil de conducere, ceea ce ar putea duce la predicții mai bune de traiectorie în viitor, cu date suficiente de antrenament.

Cuprins

Capitolul 1

Introducere

Conducerea autonomă a fost una dintre cele mai provocatoare sarcini din ultimii ani atât în industrie, cât și în mediul academic și a fost în mintea cercetătorilor și a producătorilor de automobile în ultimul deceniu. Avantajele unei mașini autonome sunt evidente, începând de la motive de siguranță și financiare și terminând cu confortul oamenilor. Dacă se va discuta despre siguranță, automatizarea mașinilor va duce la o lume mai bună dacă toate mașinile vor încorpora sisteme aproape perfecte care nu vor greși și vor respecta regulile de circulație. Majoritatea accidentelor din ziua de azi se produc din cauză ca soferii se angajează in depasiri periculoase, nu respectă semafoarele sau regulile de circulatie, așa că un sistem perfect va depăși aceste aspecte. Doar un mic procent din accidente au loc acum din cauza problemelor mașinii sau a vremii, așa că dacă se va proiecta un algoritm perfect pentru conducerea autonomă, numărul de accidente va fi substanțial mai mic. În ceea ce privește siguranța, potrivit unui studiu, peste 94% dintre accidente sunt produse din greșeală umană, restul accidentelor fiind legate de defecțiunea mașinii, drum, vreme, sau chiar motive necunoscute. Nu numai că conducerea autonomă va duce la un mediu mai sigur, ci și unul mai bun și mai ieftin - oamenii ar putea împărți mașinile, chiar și taxiurile, pentru a merge la destinație, fiind mai ieftine și totodată mai ecologice. Mai puține mașini pe stradă vor duce la timpi mai rapidi pentru a ajunge la destinație. Cu toate acestea, în prezent nu există o mașină autonomă perfectă. O mașină autonomă implică o mulțime de componente și o mulțime de algoritmi de vârf, combinând viziunea computerizată, inteligența artificială, învățarea automată și știința datelor. Această teză își propune să abordeze unele dintre cele mai importante părți ale unei mașini autonome și să propună un nou sistem de predicție a traiectoriei pentru o mașină autonomă, bazat pe generarea video, detectarea obiectelor, segmentarea semantică și predicția adâncimii.

1.1 Conducere autonomă

O mașină autonomă constă dintr-o mulțime de componente diferite. Există un strat de achiziție, cu niște senzori (camere foto, senzor GPS, LIDAR, RADAR, IMU etc). După aceea, sistemul auto are un strat de percepție, în care mașina recunoaște mediul înconjurător, poziția acestuia, vehiculele din jur, inclusiv urmărirea acestora, drumul, distanța până la vehiculele din jur etc. După stratul de percepție, există un strat de decizie, care încorporează date de la senzori sau chiar de la celelalte mașini (într-un scenariu în care sunt multe mașini autonome pe drum). În stratul de decizie, există multe componente – planificare locală și globală a rutei, planificare comportamentală (manevra care trebuie făcută, de exemplu traversarea unei benzi, observarea și ceea ce fac ceilalți participanți) și planificarea mișcării/urmărirea traseului, pentru a urma acțiunea dorită. După toate aceste straturi, stratul final comandă mașina – având în vedere un unghi de virare și un procent de accelerație sau frânare, actuatorii trebuie să emită acele comenzi.

Când se discută și se analizează conducerea autonomă, există 5 niveluri de automatizare care au fost standardizate cu doar câțiva ani în urmă. Primul nivel conține mașinile care pot menține aceeași viteză singure sau chiar pot face o frânare de urgență. Cu toate acestea, șoferul uman încă controlează mașina. Al doilea nivel implică faptul că software-ul va prelua controlul asupra mașinii, fără nicio intervenție din partea șoferului, dar șoferul trebuie să fie foarte atent pentru a relua controlul mașinii. Al treilea nivel ar presupune că șoferul nu ar trebui să urmărească neapărat drumul, ci ar trebui să conducă mașina atunci când este nevoie (în zonele interzise, de exemplu). Al patrulea nivel va presupune că omul nu va trebui să conducă deloc, dar această facilitate ar putea fi permisă doar în anumite spații sau circumstanțe. Al cincilea și ultimul nivel va implica că nu este nevoie de un volan, deoarece software-ul controlează totul. Cu toate acestea, din 2022, doar Honda (funcțională doar pe autostradă) și Mercedes au lansat o mașină autonomă de nivel trei, alte câteva companii așteaptă să lanseze mașini cu un grad similar de automatizare. Din păcate, ultimele două niveluri nu vor fi văzute curând pe piață, ceea ce conferă cercetătorilor responsabilitatea dezvoltării unor algoritmi mai buni, pentru a realiza o mașină complet autonomă. Zona de cercetare privind conducerea autonomă este departe de a fi depășită, noi arhitecturi fiind proiectate în fiecare lună.

1.2 Scopul tezei de doctorat

Scopul acestei cercetări este de a analiza conceptele de bază privind conducerea autonomă - detectarea și urmărirea obiectelor, segmentarea semantică și a instanțelor, estimarea adâncimii și predicția traiectoriei. De asemenea, pentru predicția traiectoriei este analizat domeniul generării video, care este legat de cercetarea condusului autonom. Teza este axată pe patru sarcini specifice - detectarea obiectelor, segmentarea semantică (în special pentru segmentarea semantică a drumurilor), estimarea adâncimii și predicția traiectoriei. Pentru fiecare sarcină sunt descrise cele mai relevante lucrări în domeniu și, de asemenea, studiile care s-au făcut cu privire la domeniul în sine, dar și pentru domenii conexe, cum ar fi generarea video. Această cercetare este, de asemenea, relevantă pentru compararea studiilor de ultimă generație cu privire la aceste subiecte și sunt descrise alte articole și studii de revizuire, pentru a arăta ce aduce nou această teză în ceea ce privește recenziile actuale. Pentru fiecare sarcină se analizează seturile de date existente și se realizează, de asemenea, seturi de date noi, înregistrate și adnotate manual de către echipa Universității Politehnica și sunt comparate cu cele existente. Seturile de date au fost înregistrate în campusul universitar și au luat în considerare factori precum ora din zi și dimensiunea mașinilor. Construcția setului de date este, de asemenea, o altă contribuție importantă în această teză, prin aducerea de noi seturi de date comunității de cercetare. Pentru fiecare sarcină se fac unele experimente folosind cele mai bune rețele disponibile în acest moment. Experimentele au fost realizate cu imaginile din seturile de date înregistrate și s-au realizat diferite statistici privind calitatea rezultatelor, timpul de inferență și posibilitatea utilizării algoritmilor într-o aplicație în timp real. Pentru sarcina finală, predicția traiectoriei, multe dintre rezultatele anterioare au fost incluse în vederea realizării unui nou algoritm de predicție a traiectoriei, care să ia în considerare segmentarea semantică, predicția adâncimii și detectarea obiectului și să aibă la bază ideea de video. generație – o abordare care nu a fost văzută anterior în literatura de specialitate. De asemenea, sunt propuse trei noi variante de arhitectură prin modificarea unei rețele populare de generare video, unele dintre ele obținând rezultate mai bune pentru sarcina de predicție a traiectoriei decât modelul de bază.

1.3 Conținutul tezei de doctorat

Teza este structurată în 8 capitole. Capitolul 1 constă în introducerea și prezentarea domeniului condusului autonom și a tezei. Restul tezei este structurat după cum urmează. În Capitolul 2 sunt detaliate câteva informații utile de bază pentru ca orice inginer cu suficiente cunoștințe de informatică să înțeleagă teza - este descris conceptul de rețele neuronale, care este folosit în toate experimentele realizate. Rețelele neuronale sunt arhitecturi noi care încearcă să simuleze creierul uman și sunt acum utilizate pe scară largă pentru sarcini de viziune pe computer. Acesta este motivul pentru care o scurtă prezentare pentru rețelele neuronale poate fi văzută ca fiind obligatorie, pentru ca această teză să fie mai bine înțeleasă. În Capitolul 3, este analizată munca aferentă fiecăreia dintre cele patru sarcini care sunt discutate în teză - detectarea obiectelor, segmentarea semantică, predicția traiectoriei și estimarea adâncimii. Sunt incluse și câteva informații succinte privind urmărirea obiectelor și generarea video, care este conceptul de bază al modelului de predicție a traiectoriei. În această cercetare este prezentat un nou model care prezice traiectoria utilizând un model de generare video, lucru care nu a fost încercat anterior în literatura de specialitate. Sunt prezentate și analizate atât studiile conexe, cât și recenziile referitoare la cele mai bune arhitecturi, pentru a arăta ce aduce nou teza actuală în raport cu celelalte recenzii. Următoarele patru capitole prezintă experimentele privind sarcina de studii. În Capitolul 4 sunt descrise rezultatele având în vedere sarcina de detectare a obiectelor, în Capitolul 5 sunt descrise rezultatele privind sarcina de segmentare semantică a drumului, Capitolul 6 constă în descrierea experimentelor efectuate pentru sarcina de estimare a adâncimii și, în final, Capitolul 7 constă în experimentele efectuate pentru sarcina de predicție a traiectoriei și, de asemenea, descrie modul de lucru propus și prezintă trei arhitecturi noi modificate bazate pe o arhitectură populară de generare video. Fiecare dintre aceste patru capitole are o structură asemănătoare, conținând informații cu privire la cele mai relevante seturi de date și, de asemenea, cu privire la setul de date propus, experimentele efectuate, metricile implicate și, de asemenea, prezentarea rezultatelor. Capitolul 7 are și o altă secțiune referitoare la noile arhitecturi propuse, rezultatele acestora luând în considerare aceeași sarcină și configurație. În final, concluziile și lucrările viitoare sunt discutate în capitolul 8.

Capitolul 2

Informații generale

În această teză, majoritatea arhitecturilor prezentate sunt rețele neuronale artificiale (ANN). Rețelele neuronale au fost dezvoltate în anii 1960, încercând să realizeze un model simplificat al creierului uman – există niște noduri numite neuroni care sunt conectate între ele – o arhitectură foarte simplă copiată de funcționarea creierului uman. Chiar dacă există multe straturi de neuroni, doar două sunt vizibile - stratul de intrare și stratul de ieșire, iar restul straturilor, care au numit straturi ascunse din motive evidente și sunt folosite pentru a calcula greutățile stratului final. . Conexiunile dintre neuroni sunt etichetate drept margini. În modelul ANN fiecare neuron și fiecare margine au, în general, o valoare, numită greutate. Greutatea va determina cât de importantă este acea componentă specifică în ceea ce privește rezultatul final. Pe măsură ce modelul învață luând mai multe date pentru modelare, ponderile anumitor neuroni se pot schimba în valoare - pot crește sau scădea. Chiar dacă modelele nu sunt atât de cunoscute, utilizarea ANN-urilor în învățarea automată, viziunea computerizată și procesarea limbajului natural a crescut doar în ultimii 20 de ani. În zilele noastre, multe dintre sarcinile de viziune computerizată, sarcinile de procesare a limbajului natural și, de asemenea, alte sarcini de învățare automată (de exemplu, găsirea căii în roboți, recunoașterea acțiunilor etc.) sunt realizate folosind aproape exclusiv rețele neuronale. Există multe tipuri de rețele neuronale.

2.1 Rețele neurale

2.1.1 Rețele neurale înainte

Cel mai simplu model, care va fi explicat pe scurt aici, este rețeaua neuronală feedforward (înainte). Dacă conține straturi ascunse, se mai numește și perceptronul nultilayer (MLP).

Acest model de bază poate fi îmbunătățit în continuare prin preluarea datelor în loturi și, de asemenea, prin efectuarea diferitelor modificări în ceea ce privește modificările pentru ponderi și părtinire. Există diferiți algoritmi de optimizare precum RMSProp, Momentum, Adam.

Rețeaua neuronală feedforward arată elementele de bază ale oricărei rețele neuronale – numărul de straturi, funcțiile de activare, funcția de ieșire, funcția de cost, metoda de optimizare. Acești parametri pot fi variați și pot forma un număr infinit de rețele neuronale artificiale. În general, nu există o metodă cunoscută pentru obținerea celei mai bune rețele. Cel mai important factor care duce la creșterea ANN-urilor și la rezultate mai bune este variația experimentelor. Rețelele actuale sunt în general inspirate de cele anterioare – se adaugă noi straturi și se fac noi modificări ale structurii, atâta timp cât rezultatele sunt mai bune.

2.1.2 Rețele neuronale recurente

Pe lângă rețeaua neuronală feedforward, acum există și alte rețele neuronale care sunt utilizate. Rețeaua neuronală feedforward are doar conexiuni de la un strat la următorul. Următorul pas a fost introducerea conexiunilor între același strat sau chiar între un strat și cel anterior. Cel mai simplu exemplu este o rețea neuronală complet recurentă (RNN), în care fiecare neuron din rețea este conectat la fiecare alt neuron. RNN de bază ia în considerare doar stările ascunse anterioare prin concatenarea stărilor ascunse într-un vector mai mare. Cu toate acestea, cele mai utilizate rețele neuronale recurente folosesc unități Long Short-Term Memory (LSTM) sau Gated Recurrent Units (GRU). Cele mai comune rețele folosesc LSTM, care au reușit să abordeze unele probleme pe care le au rețelele neuronale feedforward, de exemplu problema gradientului de dispariție sau a gradientului de explozie, motiv pentru care LSTM-urile sunt folosite pentru sarcini care necesită timp – pt. exemplu predicția traiectoriei unei mașini. GRU-urile sunt un model simplificat al LSTM și sunt folosite mai mult în procesarea limbajului natural, dar au și utilizarea lor în viziunea computerizată.

2.1.3 Rețele neuronale convoluționale

Un alt tip de rețele neuronale sunt rețelele neuronale convoluționale (CNN). Acest tip de ANN-uri au obținut cele mai bune rezultate pentru sarcinile referitoare la imagini, datorită proprietăților lor de manipulare a spațiului. Se poate afirma că RNN-urile sunt cele mai bune atunci când trebuie luat în considerare timpul și CNN-urile sunt cele mai bune dacă spațiul trebuie încorporat. Unele rețele neuronale moderne au atât componente recurente, cât și convoluționale. O rețea neuronală convoluțională este practic o rețea neuronală feedforward care are cel puțin unele straturi care efectuează convoluții. Convoluția este o operație specială care reduce practic dimensiunea spațiului de intrare (de exemplu, o imagine).

2.1.4 Rețele cu autoencodare

Pe lângă RNN-uri și CNN-uri, există și alte tipuri de rețele neuronale artificiale. Nu o rețea în sine, ci o arhitectură utilă, este modelul decodorului de codificator. Se aplică în special în ceea ce privește rețelele neuronale recurente și constă din două părți – un codificator, care preia intrarea și o transformă într-o stare de dimensiune fixă (un tensor multidimensional), și un decodor, care va prelua tensorul și va încerca să aplica din nou transformarea. Cu cât decodorul este mai bun înseamnă că encoderul reușește să reprezinte componentele relevante ale intrării. Schema de decodor al codificatorului este utilizată în viziunea computerizată, procesarea limbajului natural și alte sarcini de învățare automată.

Un tip de rețea neuronală care este foarte utilizat în ultimii ani este Variational Autoencoder (VAE). Autoencoderul însuși este utilizat pentru reducerea dimensionalității, prin utilizarea schemei de decodor al codificatorului. Cu toate acestea, autoencoderul variațional are o altă utilizare importantă - încearcă să obțină proprietăți bune pentru rezultatul codificatorului, care se numește spațiu latent, pentru a preleva mostre din spațiul latent și a le utiliza pentru sarcina de generare video, de exemplu .

2.1.5 Rețele neuronale generative

Ultima rețea neuronală prezentată în această secțiune este Generative Adversarial Network (GAN). Această rețea este concepută special pentru generare – fie imagini noi, fie text. În această teză vor fi exploatate beneficiile GAN pentru generarea video și predicția video. Ideea din spatele GAN este foarte simplă – constă din două rețele diferite – un generator și un discriminator.

Există și alte tipuri de rețele neuronale, de exemplu rețelele de transformatoare, care sunt utilizate în procesarea limbajului natural, sau harta auto-organizată (SOM), o altă rețea pentru reducerea dimensionalității. Cu toate acestea, utilizarea lor în înțelegerea scenei și conducerea autonomă este încă limitată, motiv pentru care alte arhitecturi nu sunt mai detaliate.

Capitolul 3

Probleme conexe

În acest capitol sunt descrise arhitecturile de ultimă generație privind cele mai importante subiecte pentru această cercetare și, de asemenea, pentru conducerea autonomă - detectarea obiectelor, urmărirea obiectelor, segmentarea semantică, de instanță și panoptică, predicția traiectoriei și generarea video. Pentru fiecare dintre aceste subiecte este prezentată evoluția arhitecturilor de-a lungul istoriei, sunt realizate câteva categorii privind abordările, sunt menționate cele mai bune arhitecturi în ceea ce privește rezultatele lor în aplicații din viața reală și, de asemenea, sunt discutate unele dintre cele mai bune articole de recenzie referitoare la subiect. În loc să menționăm doar unele dintre cele mai bune arhitecturi utilizate, această secțiune poate fi văzută ca o trecere în revistă și o comparație a celor mai bune tehnologii cu privire la unele dintre cele mai importante subiecte în conducerea autonomă în ceea ce privește anul 2022 și ar trebui să fie considerată drept unul dintre cele mai importante contribuțiile tezei.

3.1 Provocări în conducerea autoonomă

3.1.1 Detecție de obiecte

Abordările de deep learning pot fi împărțite în trei clase, având în vedere aspectul lor istoric. Există detectoare care funcționează într-un proces în două etape și detectoare care au o singură etapă. Toate folosesc ancore pentru a detecta obiecte. Cele mai noi rețele nu folosesc deloc ancore, așa că pot fi grupate într-o altă categorie, detectoare fără ancore.

3.1.2 Urmărirea obiectelor

Sarcina de urmărire a obiectelor implică identificarea aceluiași obiect în cadre diferite - după detectarea obiectului, urmărirea obiectului ar trebui să lege același obiect între cadre. Un avantaj al urmăririi obiectelor este că se poate deduce poziția unui obiect (o persoană sau o mașină) dacă se știe cu siguranță că acesta apare într-un anumit cadru, dar

detectarea va eșua, din diverse motive (ocluzie, estomparea imaginii cauzată de mișcare, variații de iluminare, rezoluție, scară etc.).

3.1.3 Segmentare

În secțiunea următoare, se va analiza stadiul tehnicii în ceea ce privește segmentarea semantică și a instanțelor, precum și câteva studii conexe care abordează problema segmentării semantice, ce a fost prezentată și ce este îmbunătățit în această lucrare, în ceea ce privește segmentarea semantică și segmentarea instanțelor. problemă. Prima discuție este despre unele dintre cele mai utilizate arhitecturi pentru clasificarea obiectelor, care sunt folosite pentru extragerea de caracteristici în rețelele de segmentare semantică, apoi se analizează sarcina de segmentare a instanțelor, unde sunt detectate obiectele, fiecare instanță este identificată și fiecare pixel. al obiectului este clasificat, dar fundalul nu este clasificat, precum și segmentarea semantică, unde fiecare pixel al imaginii este clasificat, fără a ține cont de instanțe individuale. De asemenea, va fi analizată o nouă abordare, segmentarea panoptică, care combină cele două metode. La final, vor fi analizate articolele de recenzie aferente și ce aduce nou această teză studiului de segmentare.

3.1.4 Segmentarea instanțelor

Sarcina de segmentare a instanțelor se bazează, în general, pe detectarea obiectelor.

3.1.5 Segmentare semantică

Această subsecțiune analizează arhitecturile de segmentare semantică. Ele sunt foarte importante pentru conducerea autonomă deoarece pot detecta și clasifica pixelii de fundal, inclusiv drumul, care este una dintre cele mai importante sarcini în conducerea autonomă și, de asemenea, subiectul actualului studiu. Există multe arhitecturi, majoritatea bazate pe rețele neuronale convoluționale, cu arhitecturi și optimizări diferite.

3.1.6 Segmentare panoptică

În această subsecțiune sunt descrise câteva arhitecturi care realizează atât segmentarea instanței, cât și sarcina de segmentare semantică. Această abordare se numește segmentare panoptică și este o modalitate foarte recentă de a face segmentarea și, de asemenea, este utilă pentru conducerea autonomă.

3.1.7 Estimarea adâncimii

În această secțiune sunt analizate și comparate cele mai importante rețele de estimare a adâncimii, precum și cele mai relevante lucrări de revizuire referitoare la rețelele de adâncime. Prima diviziune privind studiile de estimare a adâncimii se poate face în ceea ce privește numărul de camere care sunt utilizate. Rețelele de adâncime stereo folosesc două camere și au o precizie mai bună, dar au nevoie și de un sistem mai complex, astfel încât rețelele de adâncime monoculară au propriile avantaje. În experimentele efectuate s-au folosit doar rețele de estimare a adâncimii monoculare, care au avantajul că sunt mai ușor de antrenat și testat și au nevoie de o infrastructură mai ieftină.

3.1.8 Predicția traiectoriei

În această secțiune sunt descrise cele mai relevante rețele în ceea ce privește sarcinile aferente experimentelor realizate pentru predicția traiectoriei. Cele mai relevante rețele folosesc arhitecturi LSTM, RNN, GAN, LSTM-CNN sau CNN.

3.1.9 Generare de video

Sarcina de a realiza cadre care formează un videoclip poate fi împărțită în două categorii diferite – există rețele care încearcă să genereze cadre aleatorii care ar putea fi considerate un videoclip real (fără nicio legătură cu unul real) și, de asemenea, rețele care încearcă să prezică noi cadre cu un videoclip original. Sarcinile de generare video și predicție video sunt, totuși, legate și în multe cazuri cei doi termeni pot fi folosiți într-un mod interschimbabil. Cele mai relevante rețele utilizează arhitecturi LSTM, RNN, GAN, LSTM-CNN sau VAE.

Capitolul 4

Evaluarea detecției de obiecte pentru mașini autonome

În acest capitol sunt descrise cele mai importante rezultate privind detectarea obiectelor și sarcina aferentă acesteia, urmărirea obiectelor. În prima secțiune sunt descrise câteva seturi de date clasice pentru sarcina de detectare a obiectelor și, de asemenea, setul de date utilizat în experimentele curente. În secțiunea următoare sunt descrise experimentele și metricile utilizate pentru sarcina de detectare a obiectelor. În ultima secțiune rezultatele sunt descrise și analizate.

4.1 Setul de date pentru detecția de obiecte din UPB

Unele dintre imaginile din setul de date POLI pot fi găsite în Figura 4.1.

Fig. 4.1 Imagini ale setului de date POLI

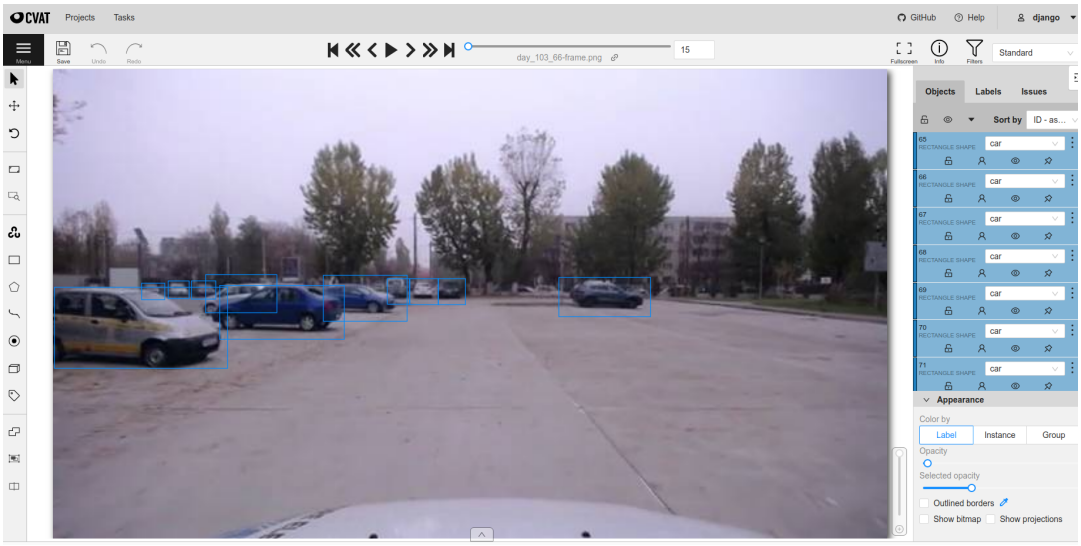

Fig. 4.2 Instrumentul de adnotare CVAT

Setul de date a fost înregistrat pe străzile campusului, trecând pe lângă multe mașini și studenți. Setul de date a fost etichetat manual folosind un instrument online, CVAT, care simplifică sarcina de adnotare, prin interpolarea casetelor de delimitare din cadre între două cadre adnotate, apoi casetele rezultate pot fi ajustate manual pentru a se potrivi perfect cu adevărul de bază. Adnotarea este una dintre sarcinile cele mai consumatoare de timp în ceea ce privește viziunea computerizată, cu multă muncă manuală care trebuie făcută pentru ca casetele de delimitare să fie cât mai aproape de cea ideală. O captură de ecran cu capabilitățile instrumentului CVAT poate fi văzută în Figura 4.2.

Setul de date este format din 13001 imagini, care conțin 60227 obiecte, din care 90% sunt fie mașini, fie persoane (41064 obiecte sunt mașini, 14576 sunt persoane). În setul de date există și semne de circulație și biciclete. Setul de date este unul greu, deoarece există cadre cu multe mașini adnotate într-un loc de parcare unde se vede și drumul aglomerat care se află în fața universității, cu mai multe mașini traversând în fiecare secundă. Detectoarele de ultimă generație au avut dificultăți în ceea ce privește acest set de date, așa cum se poate observa în Capitolul 5, rechemarea fiind mai mică decât un set de date clasic, mai puțin aglomerat.

4.2 Experimente și rezultate

În această secțiune sunt analizate experimentele și metricile realizate pentru sarcina de detectare a obiectelor. Pentru a rezuma, au fost utilizate 4 rețele – YOLO v3, RetinaNet, Faster R-CNN și SSD și le-au testat pe setul de date BDD100K, pentru a vedea performanțele acestora în ceea ce privește retragerea și acuratețea și modul în care rezultatele variază în funcție de ora zilei (ziua, amurgul sau zorii și noaptea) și, de asemenea, s-a făcut o statistică diferită cu privire la dimensiunea obiectelor. Aceleași arhitecturi au

Tabelul 4.1 Precizia pe setul de date BDD100K

	DC	DO	DR	DS	DSC	DSO	DSR	DSS	NC	NO	NR	NS
YOLO AP@.50IOU	0.66	0.65	0.64	0.65	0.65	0.65	0.62	0.65	0.63	0.64	0.61	0.64
SSD AP@.50IOU	0.92	0.93	0.92	0.92	0.94	0.93	0.93	0.93	0.90	0.93	0.90	0.91
Faster R-CNN AP@.50IOU	0.84	0.86	0.86	0.88	0.86	0.86	0.87	0.89	0.82	0.86	0.82	0.83
RetinaNet AP@.50IOU	0.28	0.27	0.32	0.32	0.27	0.26	0.29	0.30	0.31	0.33	0.32	0.34
YOLO MAP	0.56	0.55	0.55	0.55	0.55	0.55	0.53	0.55	0.54	0.55	0.53	0.55
SSD MAP	0.79	0.80	0.79	0.79	0.80	0.79	0.79	0.80	0.74	0.79	0.75	0.75
Faster R-CNN MAP	0.67	0.69	0.68	0.70	0.68	0.68	0.68	0.71	0.64	0.68	0.63	0.65
RetinaNet MAP	0.34	0.34	0.37	0.36	0.34	0.33	0.35	0.36	0.37	0.39	0.37	0.38

Tabelul 4.2 Rechemarea pe setul de date BDD100K

	DC	DO	DR	DS	DSC	DSO	DSR	DSS	NC	NO	NR	NS
YOLO AR@.50IOU	0.36	0.37	0.36	0.36	0.35	0.37	0.34	0.35	0.20	0.33	0.18	0.22
SSD AR@.50IOU	0.17	0.17	0.19	0.19	0.16	0.17	0.18	0.19	0.12	0.16	0.11	0.14
Faster R-CNN AR@.50IOU	0.14	0.14	0.13	0.14	0.12	0.13	0.12	0.13	0.05	0.11	0.04	0.06
RetinaNet AR@.50IOU	0.09	0.08	0.10	0.10	0.08	0.08	0.09	0.09	0.06	0.09	0.06	0.07
YOLO MAR	0.30	0.31	0.30	0.30	0.29	0.31	0.29	0.30	0.17	0.28	0.15	0.19
SSD MAR	0.14	0.15	0.16	0.16	0.14	0.14	0.15	0.16	0.10	0.14	0.09	0.11
Faster R-CNN MAR	0.11	0.11	0.10	0.11	0.09	0.10	0.09	0.10	0.04	0.09	0.03	0.05
RetinaNet MAR	0.11	0.11	0.11	0.12	0.10	0.10	0.10	0.11	0.07	0.11	0.07	0.08

fost testate față de setul de date propus în Campusul Universității Politehnica, adnotat manual. De asemenea, s-au făcut aceleași statistici, pentru a vedea care dintre rețele au rezultate mai bune și, de asemenea, cum se vor adapta rezultatele de la un set de date mare la unul mai mic, necunoscut, care nu a fost folosit pentru reglaj fin. Aceasta a fost realizată pentru a vedea performanțele reale ale arhitecturilor, fără a avea de-a face cu supraadaptarea.

4.2.1 Rezultate

Rezultatele pentru precizie pot fi văzute în Tabelul 4.1 și Tabelul 4.3 (doar pentru mașină și persoană), iar rezultatele pentru rechemare pot fi văzute în Tabelul 4.2 și Tabelul **??** (doar pentru mașină și persoană). Rezultatele pentru setul de date POLI sunt prezentate în tabelul 4.5. Unele diagrame privind precizia și reamintirea cu privire la dimensiunea mașinii pot fi văzute în Figura 4.3 și în Figura 4.4.

Tabelul 4.3 Precizia pe setul de date BDD100K - doar mașini și persoane

	DC	DO	DR	DS	DSC	DSO	DSR	DSS	NC	NO	NR	NS
YOLO AP@.50IOU	0.72	0.72	0.74	0.72	0.72	0.72	0.73	0.73	0.72	0.74	0.73	0.76
SSD AP@.50IOU	0.92	0.93	0.92	0.92	0.94	0.93	0.93	0.93	0.90	0.93	0.90	0.91
Faster R-CNN AP@.50IOU	0.87	0.87	0.87	0.88	0.88	0.87	0.89	0.89	0.85	0.88	0.85	0.85
RetinaNet AP@.50IOU	0.28	0.27	0.32	0.32	0.27	0.26	0.29	0.30	0.31	0.33	0.32	0.34
YOLO MAP	0.60	0.60	0.62	0.60	0.60	0.61	0.61	0.61	0.60	0.62	0.60	0.62
SSD MAP	0.79	0.80	0.79	0.79	0.80	0.79	0.79	0.80	0.74	0.79	0.75	0.75
Faster R-CNN MAP	0.69	0.70	0.69	0.70	0.70	0.69	0.69	0.71	0.66	0.70	0.66	0.67
RetinaNet MAP	0.34	0.34	0.37	0.36	0.34	0.33	0.35	0.36	0.37	0.39	0.37	0.38

Tabelul 4.4 Rechemarea pe setul de date BDD100k - doar mașini și persoane

	DC	DO	DR	DS	DSC	DSO	DSR	DSS	NC	NO	NR	NS
YOLO AR@.50IOU	0.48	0.50	0.50	0.50	0.47	0.50	0.47	0.50	0.30	0.46	0.27	0.34
SSD AR@.50IOU	0.17	0.17	0.19	0.19	0.16	0.17	0.18	0.19	0.12	0.16	0.11	0.14
Faster R-CNN AR@.50IOU	0.20	0.21	0.20	0.21	0.17	0.19	0.17	0.21	0.09	0.16	0.08	0.10
RetinaNet AR@.50IOU	0.09	0.08	0.10	0.10	0.08	0.08	0.09	0.09	0.06	0.09	0.06	0.07
YOLO MAR	0.40	0.43	0.42	0.42	0.39	0.42	0.39	0.42	0.25	0.38	0.22	0.28
SSD MAR	0.14	0.15	0.16	0.16	0.14	0.14	0.15	0.16	0.10	0.14	0.09	0.11
Faster R-CNN MAR	0.16	0.17	0.16	0.17	0.14	0.15	0.14	0.17	0.07	0.13	0.06	0.08
RetinaNet MAR	0.11	0.11	0.11	0.12	0.10	0.10	0.10	0.11	0.07	0.11	0.07	0.08

Tabelul 4.5 Rezultate pe setul de date POLI

	AP@0.50IOU	MAP	AR@0.50IOU	MAR
YOLO	0.69	0.55	0.59	0.47
SSD	0.79	0.65	0.12	0.10
Faster R-CNN	0.68	0.54	0.16	0.13
RetinaNet	0.17	0.26	0.06	0.09
YOLO (car and person)	0.71	0.57	0.62	0.50
SSD (car and person)	0.90	0.74	0.13	0.10
Faster R-CNN (car and person)	0.80	0.63	0.17	0.13
RetinaNet (car and person)	0.20	0.30	0.06	0.10

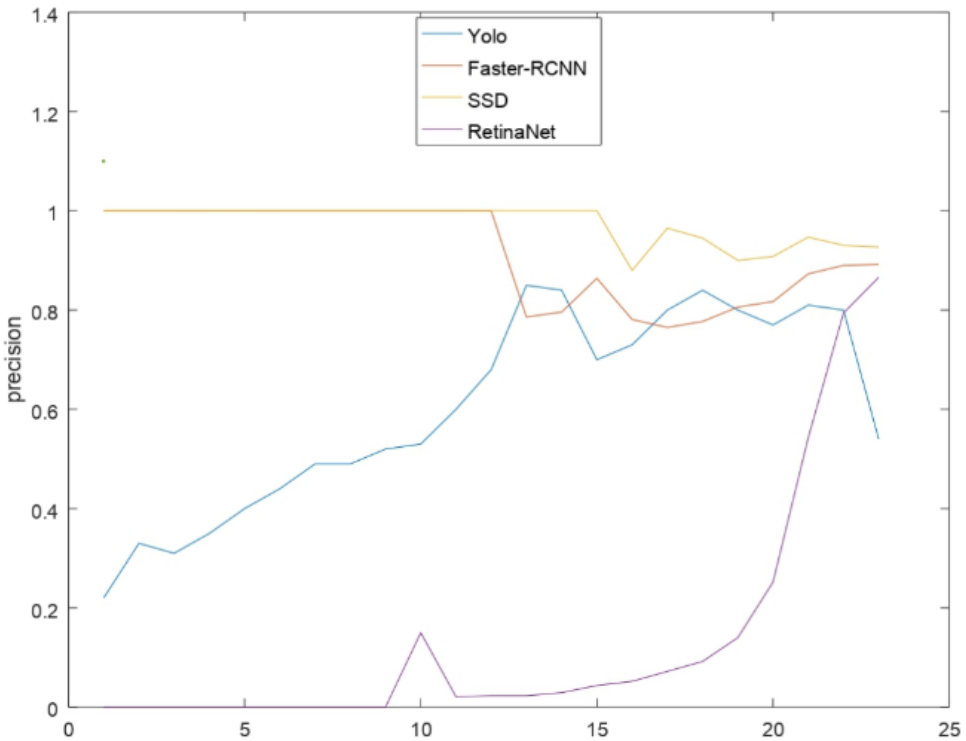

Fig. 4.3 Precizia raportată la mărimea obiectelor

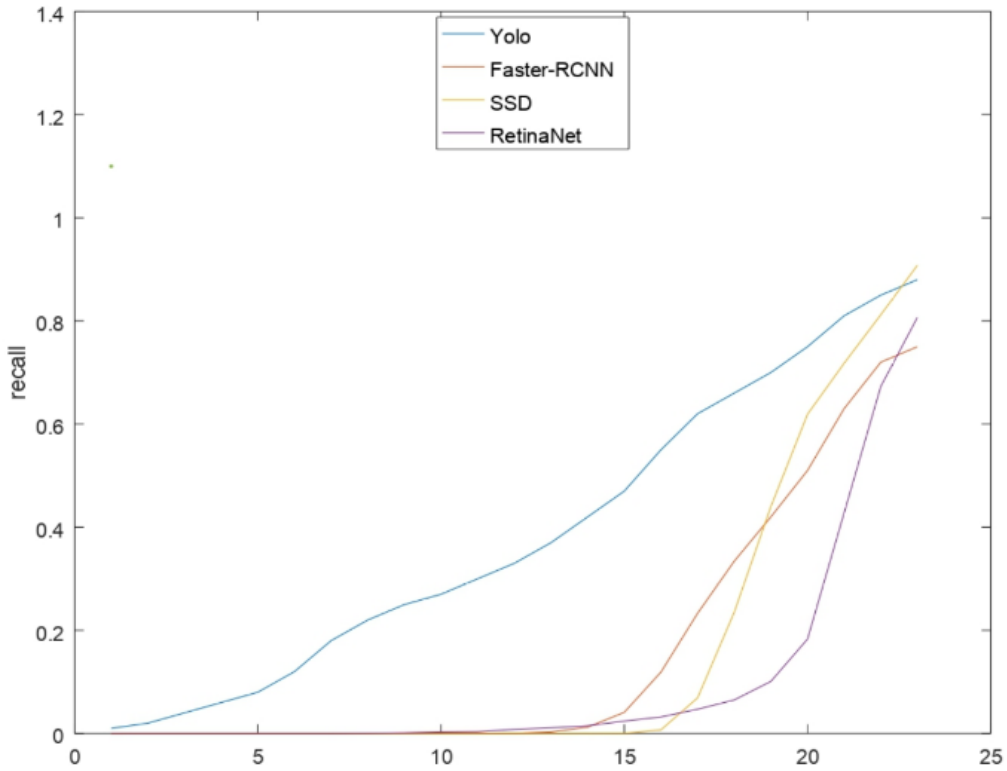

Fig. 4.4 Rechemarea raportată la mărimea obiectelor

Capitolul 5

Evaluarea segmentării pentru mașini autonome

În acest capitol sunt prezentate experimentele efectuate cu privire la sarcina de segmentare semantică, cu accent pe segmentarea semantică rutieră. Prima secțiune a acestui capitol descrie unele dintre cele mai relevante seturi de date de segmentare utilizate în literatură, dar și setul de date de segmentare POLI, setul de date utilizate pentru experimentele curente. Următoarea secțiune descrie cele mai importante experimente efectuate și, de asemenea, metricile pentru evaluarea calității experimentelor. Ultima secțiune propune și face o interpretare cu privire la datele prezentate.

5.1 Setul de date pentru segmentare din UPB

Pentru sarcina de segmentare, în campusul universitar Politehnica au fost înregistrate 138 de filme în diferite scenarii – în timpul zilei, în timpul nopții și, de asemenea, la răsărit sau apus, care a fost etichetat ca amurg, similar adnotărilor BDD100k. Unele dintre imaginile au fost înregistrate în zona de parcare, din fața clădirii Automatic and Control Science, cu vederea mașinilor în mișcare pe bulevardul foarte încununat de lângă facultate, cu multe mașini care ar putea fi eventual depistate. În Figura 5.1 poate fi văzută zona de parcare în diferite setări de lumină - primele două în timpul zilei, a treia în timpul amurgului și ultima în timpul nopții.

Imaginile sunt preluate de pe străzile din incinta Universității Politehnica din campusul București. Înregistrările au fost pe vreme senină (fără ploaie sau ninsoare). Fiecare film conține câteva cadre, însumând aproximativ 20.000 de cadre adnotate manual, care conțin segmentarea pentru drum. Cadrele au adnotate manual drumul pentru utilizarea CVAT, un instrument online generarea adevărului la sol, care permite utilizatorilor să facă segmentarea realizând poligoane foarte complexe. Lucrarea de a nota a fost foarte complexă, având în vedere că adnotarea drumului nu este la fel de simplă adnotarea unui obiect, care poate fi desenat doar două puncte (stânga sus și dreapta jos). Segmentarea

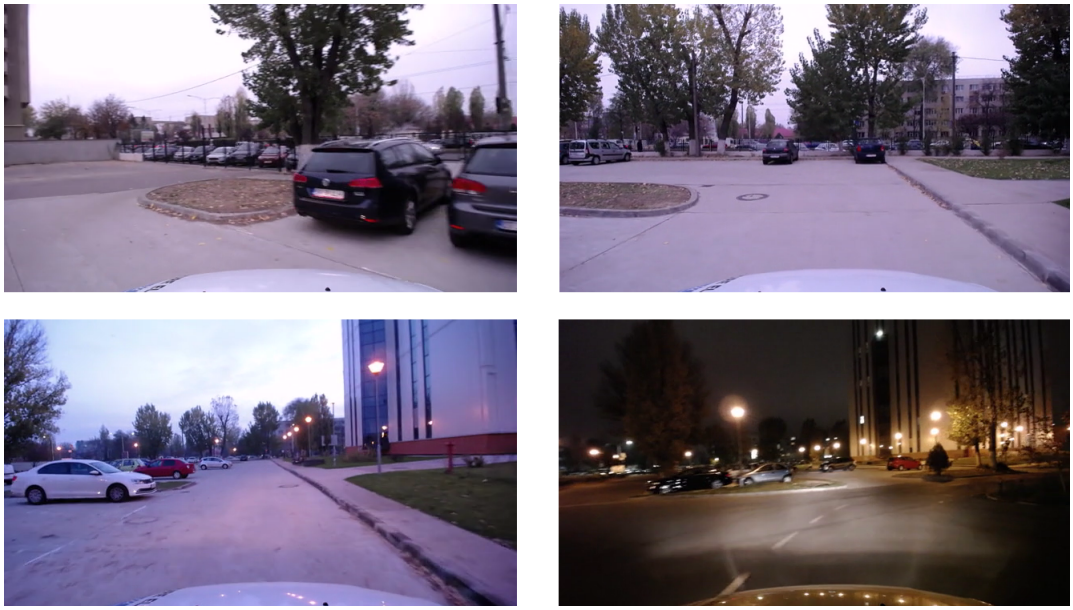

Fig. 5.1 Setul de date POLI - zona de parcare

drumului presupune foarte multe puncte, avand in vedere ca drumul va fi estimat ca un poligon cu mai multe laturi. lucru este deosebit de dificil atunci când acest lucru de-a face cu forme rotunde, care vor necesita o mulțime de puncte pentru a simula forma circulară. Un exemplu de segmentare finală a drumului în toate condițiile de lumină naturală poate fi văzut în Figura 5.2. De asemenea, un exemplu de adnotare a drumului folosit CVAT poate fi văzut în Figura 5.3.

Pentru experimente nu au fost folosite toate imaginile deoarece timpul de segmentare a fost destul de mare pentru unele rețele. În schimb, a fost selectat doar un cadru din 20 pentru a avea un set reprezentativ bun și, de asemenea, pentru a diminua timpul de inferență pentru date. În ceea ce privește tipul pozelor, sunt 735 de imagini făcute în timpul zilei, 133 în amurg și 165 în noapte.

5.2 Segmentarea drumului - experimente și rezultate

În această secțiune sunt descrise experimentele și metricile realizate cu privire la segmentarea semantică. Detectarea drumului este importantă din motive evidente – pentru a evita depășirea limitelor drumului și pentru a păstra direcția bună. Benzile rutiere ar putea fi un instrument important, de asemenea, dar nu toate drumurile il au, asa ca un bun sistem de conducere autonoma ar trebui sa poata folosi doar informatiile referitoare la pozitia drumului. Experimentele constau în detectarea drumului în setul de date propus folosind unele dintre cele mai bune rețele existente pentru segmentare și compararea rezultatelor luând în considerare ora din zi.

Fig. 5.2 Exemple de seturi de date de segmentare POLI

5.2.1 Rezultate

În această subsecțiune sunt prezentate rezultatele privind segmentarea semantică rutieră. Pentru fiecare dintre rețelele care a fost utilizată există experimente și rezultate pentru setul de date propus privind TP, FP, acuratețe și IoU în ziua, amurg, noapte și, de asemenea, o medie pentru întregul set de date. Pentru scenariul de conducere autonomă, cea mai importantă măsură este IoU. Ieșirea arhitecturii FCN pe unele cadre și, de asemenea, drumul adnotat manual pot fi văzute în Figura 5.5. Rezultatele pentru toate

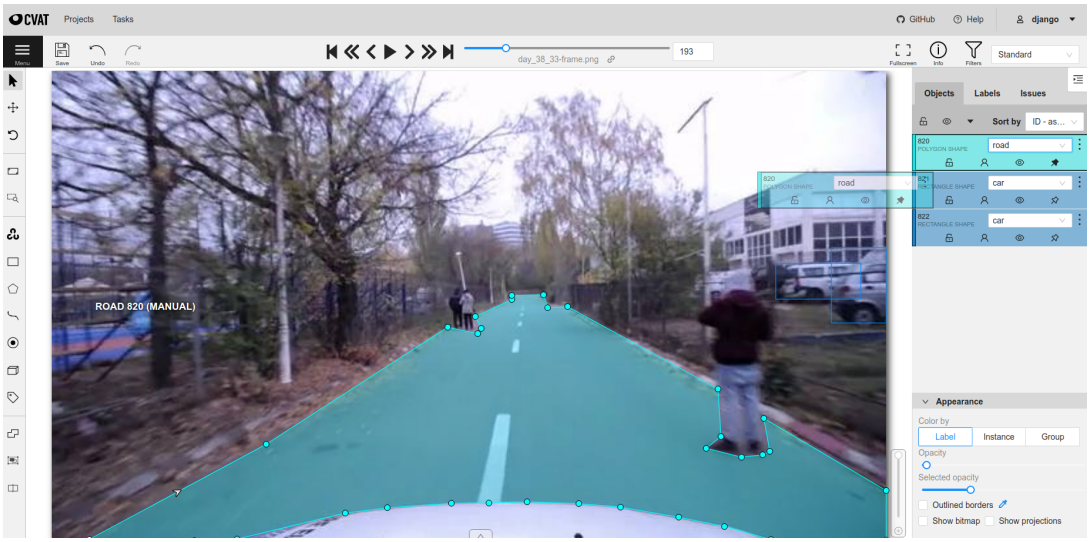

Fig. 5.3 Segmentarea drumului folosind CVAT

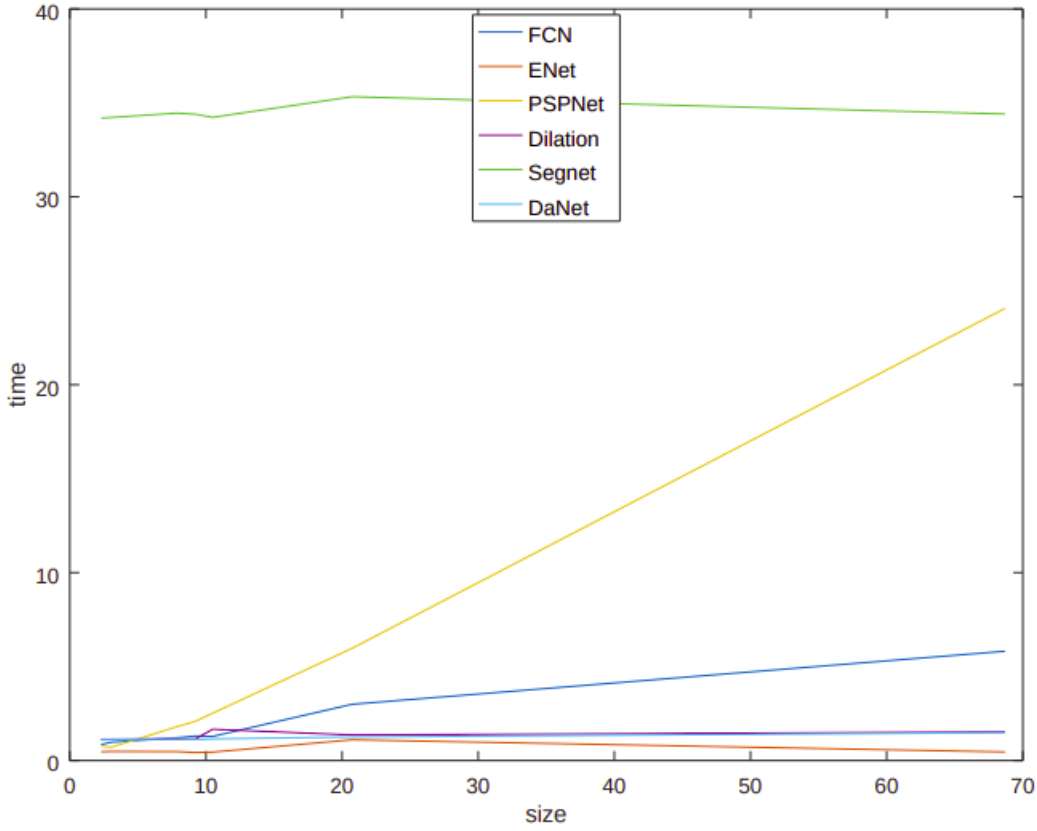

Fig. 5.4 Timpul de segmentare

rețelele testate pot fi văzute în tabelul 5.1. De asemenea, timpul de segmentare poate fi văzut în 5.4.

Pentru sarcina de detectare a obiectelor, detectarea ar trebui să fie îmbunătățită în viitor pentru a avea încredere într-o rețea pentru o aplicație de mașină cu conducere autonomă, iar studiul actual poate fi continuat prin propunerea unei noi arhitecturi de detectare a obiectelor. Chiar dacă precizia are valori decente, rechemarea ar trebui

îmbunătățită pentru a avea mai multe obiecte detectate – un obiect care nu este detectat este o posibilă cauză a unui accident, deci este important să existe o rechemare mai bună în rețelele viitoare. De asemenea, timpul poate fi îmbunătățit, deoarece în afară de detecție există și alte componente care trebuie să ruleze între două cadre (segmentare, vehicul și predicție de adâncime etc.), deci este nevoie și de un timp de inferență mai bun. Pe viitor, ar fi o idee bună ca aceste modele să fie reglate fin pentru setul de date Politehnica, pentru a vedea cum se vor ajusta parametrii atunci când rețelele sunt antrenate pe același set de date. De asemenea, setul de date ar putea fi mărit în ceea ce privește numărul obiectelor și diversitatea acestora.

Pentru segmentarea semantică, studiul actual ar trebui să extindă setul de date cu mai mult de o clasă, pentru a vedea rezultatele segmentării cel puțin pentru vehicule și oameni. De asemenea, mai multe rețele ar trebui ajustate fin în aplicațiile viitoare ale acestui studiu, în special pentru sarcina de segmentare a drumurilor (pentru a scoate doar două categorii, drum sau nu), pentru a vedea cum se vor îmbunătăți rezultatele.

Pentru estimarea adâncimii, studiul actual ar trebui să realizeze un set de date mai bun cu senzori LiDAR realizati pentru a estima mai bine eroarea de estimare. Erorile ar putea fi, de asemenea, diminuate dacă rețelele ar fi antrenate folosind setul de date dorit, dar în acest scop setul de date ar trebui înregistrat cu camere stereo, deoarece majoritatea rețelelor sunt antrenate cu seturi de date stereo și testate cu cele monoculare.

Pentru sarcina de predicție a traiectoriei, aplicațiile viitoare ale acestui studiu ar trebui să dezvolte și să antreneze în mod specific un model de generare video, luând în considerare în special sarcina de predicție a traiectoriei. De asemenea, la fiecare pas (detecție, segmentare, estimare a adâncimii) modelele corespunzătoare ar putea fi reglate fin pentru a funcționa mai bine pentru un anumit set de date și pentru sarcina de predicție a traiectoriei în conducerea autonomă.

Tabelul 5.1 Rezultate pentru segmentarea drumului

	zi	amurg	noapte	Avg
	FCN			
TP	0.782	0.742	0.572	0.743
FP	0.057	0.051	0.070	0.058
Acc	0.897	0.893	0.835	0.887
IoU	0.673	0.647	0.476	0.638
	ENET			
TP	0.781	0.725	0.711	0.763
FP	0.218	0.448	0.320	0.264
Acc	0.780	0.597	0.686	0.741
IoU	0.527	0.352	0.393	0.483
	PSPNet			
TP	0.940	0.947	0.836	0.924
FP	0.048	0.058	0.089	0.056
Acc	0.948	0.943	0.889	0.938
IoU	0.831	0.820	0.676	0.805
	Dilation			
TP	0.508	0.431	0.958	0.571
FP	0.248	0.176	0.941	0.350
Acc	0.687	0.718	0.299	0.629
IoU	0.338	0.313	0.267	0.324
	SegNet			
TP	0.955	0.903	0.454	0.868
FP	0.157	0.233	0.028	0.146
Acc	0.872	0.803	0.829	0.856
IoU	0.673	0.558	0.414	0.617
	Danet 512			
TP	0.651	0.759	0.673	0.668
FP	0.442	0.604	0.312	0.442
Acc	0.584	0.497	0.681	0.589
IoU	0.307	0.294	0.367	0.315
	Danet 768			
TP	0.770	0.825	0.713	0.768
FP	0.574	0.744	0.325	0.556
Acc	0.518	0.413	0.684	0.532
IoU	0.303	0.274	0.382	0.312

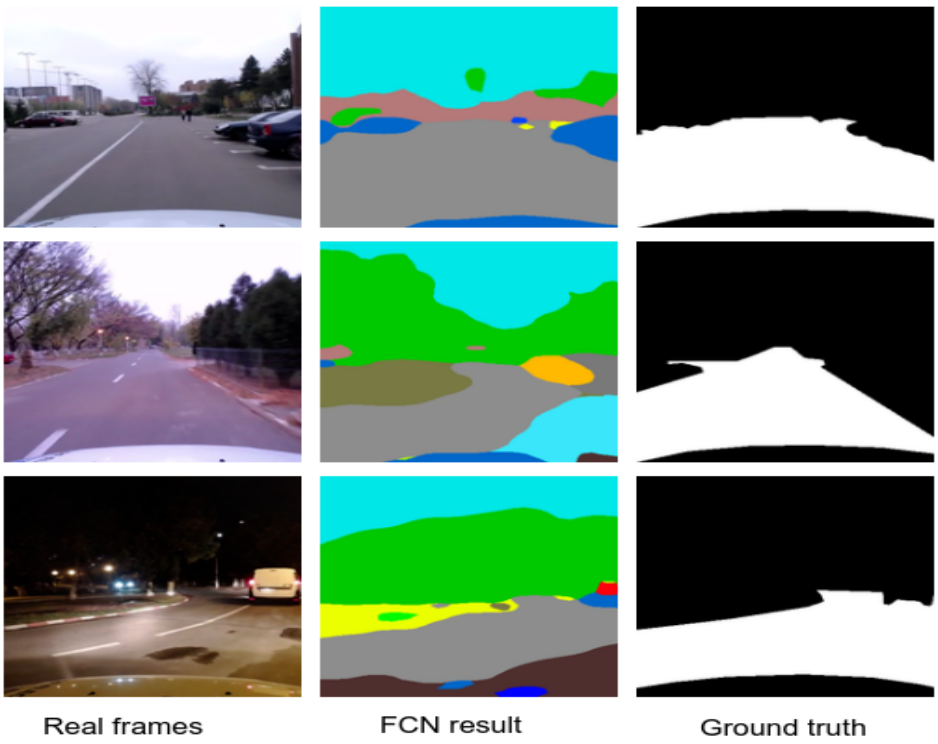

Fig. 5.5 Rezultate pentru FCN vs rezultate reale

Capitolul 6

Evaluarea adâncimii pentru mașini autonome

Acest capitol trece la analiza sarcinii de estimare a adâncimii, care este crucială în analiza distanței de la vehicul la mașinile din jur. Urmând structura capitolelor anterioare, prima secțiune analizează cele mai importante seturi de date pentru estimarea adâncimii și, de asemenea, seturile de date utilizate pentru experimentele curente, înregistrate manual în campusul Universității Politehnica din București. Următoarea secțiune descrie experimentele efectuate pentru estimarea adâncimii cu privire la mașinile din jur și, de asemenea, metricile utilizate pentru evaluarea calității rezultatelor. Secțiunea finală arată rezultatele și face, de asemenea, o interpretare a rezultatelor.

6.1 Setul de date din UPB pentru evaluarea adâncimii

Setul de date de adâncime numit setul de date de adâncime POLI este realizat pentru estimarea adâncimii și a fost înregistrat folosind hardware specializat pentru adâncime - o cameră Intel RealSense RGB-D. Urmând ideea din setul de date anterior realizat, setul de date de adâncime POLI este împărțit în trei înregistrări diferite – imagini din zi, imagini din amurg/zori și imagini din noapte. Această diviziune a ajutat la o mai bună înțelegere a problemelor de estimare a adâncimii având în vedere diferite tipuri de setări de lumină.

Înregistrarea setului de date a avut unele provocări, deoarece nu întotdeauna harta de adâncime avea aceeași dimensiune ca și imaginile. Uneori s-au pierdut unele imagini de adâncime, ceea ce a necesitat mai multe înregistrări și, de asemenea, o ajustare între cadrele hărții de adâncime și cadrele reale. O altă problemă a fost că camera avea o precizie mai mică în timpul nopții.

Camera RealSense este un model D435, cu cadre înregistrate la calitate HD – 1280x720 pixeli. Cu toate acestea, harta de adâncime a fost înregistrată doar la o

rezoluție de 848x480, ceea ce a necesitat un pas suplimentar de preprocesare pentru a se potrivi cu dimensiunile. De asemenea, unele dintre modelele de estimare a adâncimii au necesitat o rezoluție mai mică pentru imagini, iar imaginile au trebuit să fie reduse. Imaginile înregistrate au fost obținute cu camera montată în centrul parbrizului mașinii. Chiar dacă o singură locație a fost luată în considerare pentru înregistrări, având mai multe setări de lumină, setul de date POLI este un set de date bun pentru testare.

Una dintre provocările referitoare la un set de date de adâncime este senzorul utilizat - chiar dacă RealSense are rezultate foarte bune la testarea în timpul zilei, rezultatele pe timpul nopții sunt departe de a fi precise, ceea ce a adăugat ceva zgomot în rezultate. Cu toate acestea, camera este aproape de două ori mai ieftină decât versiunea mai bună, D415, iar un LiDAR complet este mult mai scump. Având în vedere acest motiv financiar, senzorul D435 a fost un compromis bun și a permis efectuarea experimentelor. Unele dintre imaginile de adâncime care se află în setul de date de adâncime POLI pot fi văzute în Figura 6.1.

[!t]

Setul de date final constă din 516 imagini înregistrate în timpul zilei, 1039 în timpul amurgului/zorii și alte 637 imagini înregistrate în timpul nopții. Setul de date final a fost obținut prin eșantionarea setului de date înregistrat pentru a obține cadre diferite din momente diferite, în loc de a avea cadre consecutive cu structură similară.

Pe lângă setul de date de profunzime, pentru experimentele realizate a fost folosit setul de date anterior, realizat pentru segmentare semantică (segment de date POLI), fără un adevăr de fond corespunzător, pentru a vedea performanța relativă a unui model față de celelalte. Performanța absolută poate fi observată folosind setul de date adnotat cu adâncime, apoi performanța relativă este măsurată pe celălalt set de date.

6.2 Rezultate și experimente

În această secțiune sunt descrise experimentele realizate cu modelele de referință și cu seturile de date propuse privind sarcina de estimare a adâncimii. După cum sa menționat mai devreme, au fost utilizate două seturi de date diferite - un set de date este adnotat cu adevărul de la sol și înregistrat cu o cameră RGB-D, iar celălalt este setul de date de segmentare POLI și nu are un adevăr de la sol pentru estimarea adâncimii. Cu toate acestea, adevărul de bază este luat în considerare luând rezultatul celor mai bune rețele din setul de date anterior. Acest lucru ajută la studierea performanței relative a rețelelor.

6.2.1 Rezultate

În această subsecțiune sunt prezentate toate rezultatele obținute în experimentele curente. Sunt analizate cele șase modele care au fost testate pe două seturi de date, toate conținând RMSE pentru zi, amurg, noapte și, de asemenea, RMSE mediu. După cum sa menționat, primul set de date conține și adevărul de la sol înregistrat cu camera Intel RealSense RGB-D, iar cel de-al doilea nu conține adevărul de la sol, dar pentru ambele seturi de date adevărul de la sol a variat ca fiind unul dintre rezultatele celor mai bune. retelelor. S-a măsurat și RMSE luând în considerare doar mașinile din imagini, ceea ce este mai relevant pentru sarcina de conducere autonomă. Pentru adevărul de bază au fost luate în considerare rezultatele de la camera de adâncime pentru primul set de date dar și rezultatele de la DenseDepth, Megadepth și Monodepth. Pe lângă aceasta, a fost măsurat și timpul de inferență în ceea ce privește dimensiunea imaginii și RMSE în ceea ce privește dimensiunea obiectului. În tabelul 6.1 sunt prezentate rezultatele pentru primul set de date, luând în considerare diferite rezultate de adevăr de bază. Rezultatele pentru al doilea set pot fi văzute în tabelul 6.2.

Cele mai relevante experimente pot fi văzute în Tabelul 6.3 și Tabelul 6.4, unde sunt prezentate rezultatele pentru RMSE privind doar mașina, care este mai precisă pentru conducerea autonomă, unde scopul este estimarea distanța de la mașina ego-ului până la vehiculele din jur. Mașinile au fost adnotate manual și pe baza adnotărilor a fost calculat RMSE. În tabelul 6.3 sunt afișate rezultatele pentru primul set și în tabelul 6.4 sunt afișate rezultatele pentru al doilea set.

Experimentele finale se referă la RMSE în ceea ce privește dimensiunea mașinii și viteza în ceea ce privește dimensiunea imaginii. Dimensiunile mașinilor au fost împărțite în 14 categorii și s-a măsurat RMSE pentru fiecare dimensiune. În figura 6.1 se găsesc rezultatele pentru RMSE privind dimensiunea mașinii, în figura 6.2 se vede viteza în ceea ce privește dimensiunea imaginii și, în final, în figura 6.3 unele rezultate de predicție pentru estimarea adâncimii folosind rețeaua Monodepth.

Tabelul 6.1 Rezultate pe primul set de date

Model	zi	amurg	noapte	Avg
Adevăr de bază - camera de adâncime				
Megadepth	128.12	140.99	139.38	137.59
DORN	72.51	98.46	55.39	82.00
LKVOLearner	98.36	109.63	97.29	103.56
SfMLearner	113.91	126.74	109.32	118.92
Monodepth	122.81	135.66	120.28	128.37
DenseDepth	82.96	83.85	87.65	84.77

Adevăr de bază - DenseDepth				
Megadepth	105.37	109.40	90.15	103.19
DORN	90.96	93.09	85.26	90.38
LKVOLearner	80.78	79.69	64.99	75.98
SfMLearner	96.37	100.18	84.74	95.03
Monodepth	111.97	111.97	100.42	108.74
Adevăr de bază - Megadepth				
DORN	125.66	130.95	139.30	132.23
LKVOLearner	47.89	54.03	57.44	53.69
SfMLearner	45.13	52.83	60.04	53.38
Monodepth	55.49	64.44	70.30	64.26
DenseDepth	105.37	109.40	90.15	103.19
Adevăr de bază - Monodepth				
Megadepth	55.49	64.44	70.30	64.26
DORN	104.46	100.40	114.93	105.76
LKVOLearner	44.49	46.07	51.81	47.46
SfMLearner	42.98	42.74	47.87	44.35
DenseDepth	111.97	111.97	100.42	108.74

Tabelul 6.2 Rezultate pentru al doilea set de date

Model	zi	amurg	noapte	Avg
Adevăr de bază - DenseDepth				
Megadepth	52.72	59.24	44.25	52.27
DORN	60.60	67.36	62.29	61.75
LKVOLearner	51.14	58.69	43.84	51.01
SfMLearner	59.87	67.36	46.28	58.84
Monodepth	64.97	74.54	56.72	64.95
Adevăr de bază - Megadepth				
DORN	20.12	21.49	25.10	52.27
LKVOLearner	12.13	13.45	13.38	61.75
SfMLearner	15.30	17.41	16.05	51.01
Monodepth	19.95	22.34	20.21	58.84
DenseDepth	52.72	59.24	44.25	64.95
Adevăr de bază - Monodepth				
Megadepth	19.95	22.34	20.21	20.30
DORN	13.15	16.75	11.72	13.42
LKVOLearner	15.10	16.81	16.35	15.53
SfMLearner	7.53	9.08	15.33	9.45
DenseDepth	64.97	74.5	56.72	64.95

Tabelul 6.3 Rezultate pentru primul set de date (doar mașini)

Model	zi	amurg	noapte	Avg
Adevăr de bază - camera de adâncime				
Megadepth	47.51	71.74	68.75	65.66
DORN	58.31	84.97	34.84	70.68
LKVOLearner	47.68	69.96	48.25	60.73
SfMLearner	51.33	73.68	41.18	62.75
Monodepth	56.77	84.92	42.34	71.17
DenseDepth	62.80	59.18	71.31	62.79
Adevăr de bază - DenseDepth				
Megadepth	52.76	60.23	29.84	53.35
DORN	74.62	75.19	65.95	73.24
LKVOLearner	54.25	56.34	40.38	52.89
SfMLearner	60.48	61.40	50.82	59.13
Monodepth	71.12	75.73	57.89	71.20

	zi	amurg	noapte	Avg
Adevăr de bază - Megadepth				
DORN	45.89	30.01	61.68	42.54
LKVOLearner	24.39	17.72	31.53	22.95
SfMLearner	33.71	19.60	43.24	29.74
Monodepth	38.05	27.08	48.93	35.49
DenseDepth	52.76	60.23	29.84	53.35
Adevăr de bază - Monodepth				
Megadepth	38.05	27.08	48.93	35.49
DORN	19.61	15.08	25.23	18.77
LKVOLearner	22.75	22.41	24.00	22.83
SfMLearner	20.27	18.65	21.29	19.64
DenseDepth	71.12	75.73	57.89	71.20

Tabelul 6.4 Rezultate pentru al doilea set de date (doar mașini)

Model	zi	amurg	noapte	Avg
Adevăr de bază - DenseDepth				
Megadepth	99.52	101.37	104.43	100.56
DORN	91.49	92.99	86.44	90.90
LKVOLearner	81.77	82.27	72.82	80.47
SfMLearner	101.46	101.56	88.81	99.56
Monodepth	114.31	115.04	105.41	113.03
Adevăr de bază - Megadepth				
DORN	114.81	120.19	126.21	117.40
LKVOLearner	40.10	46.43	51.42	42.94
SfMLearner	38.57	44.07	41.18	39.74
Monodepth	45.79	52.42	59.19	49.04
DenseDepth	99.52	101.37	104.43	100.56
Adevăr de bază - Monodepth				
Megadepth	45.79	52.42	59.19	49.04
DORN	104.57	104.59	100.99	104.01
LKVOLearner	39.32	40.10	45.96	40.55
SfMLearner	27.72	28.66	44.76	31.18
DenseDepth	114.31	115.04	105.41	113.03

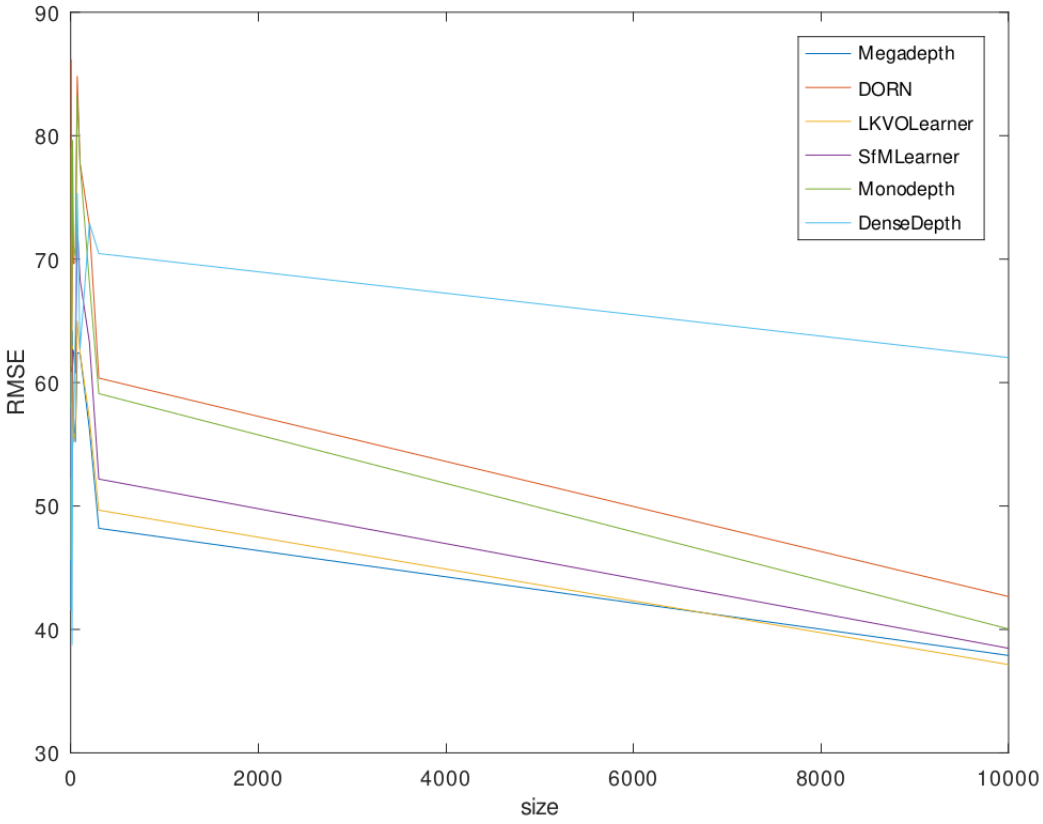

Fig. 6.1 RMSE raportat la mărimea mașinilor

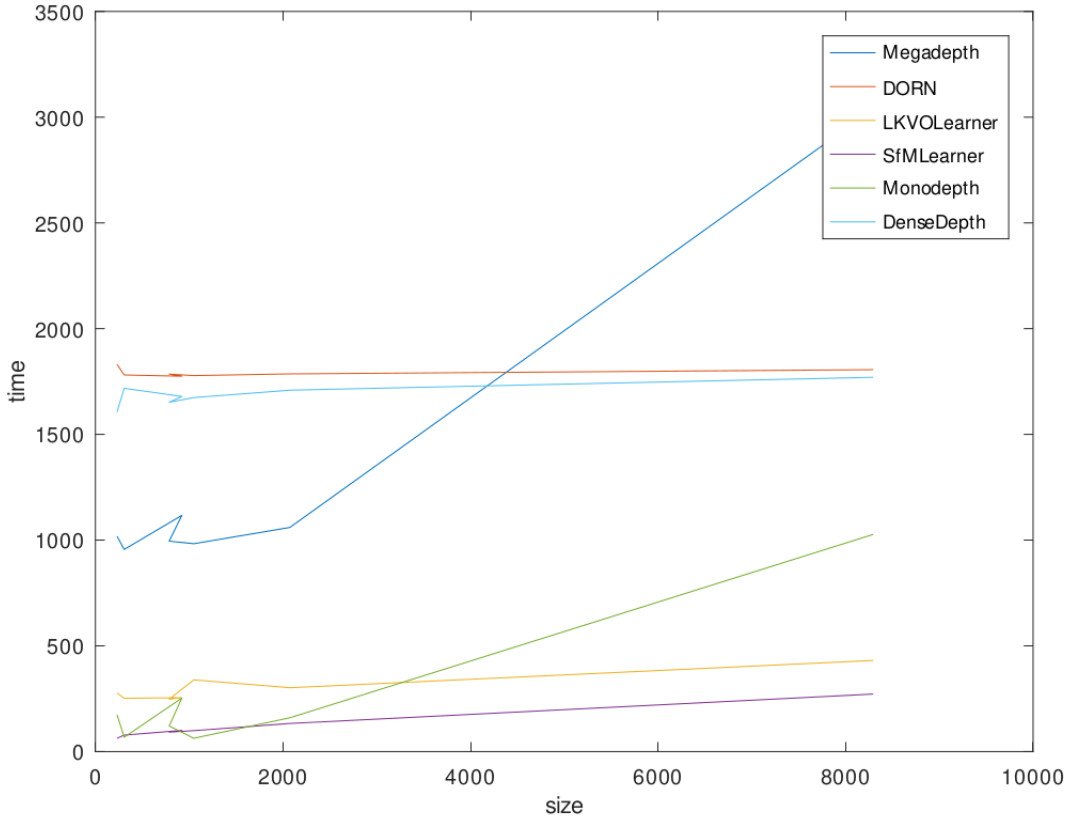

Fig. 6.2 Viteza raportată la dimensiunea imaginilor

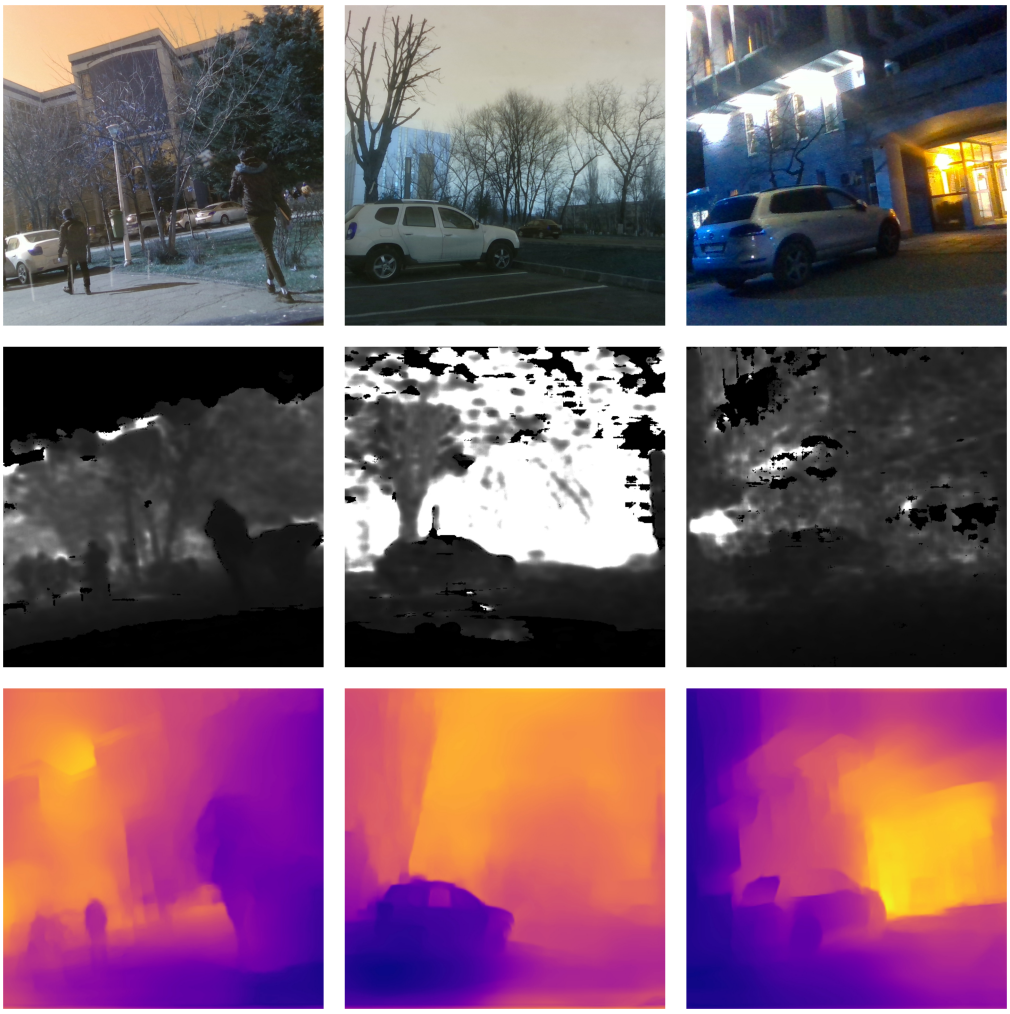

Fig. 6.3 Rezultate pentru estimarea adâncimii

Capitolul 7

Predicția traiectoriei - arhitectură și implementare

Acest capitol descrie cea mai importantă sarcină a tezei - predicția traiectoriei. Pentru sarcina de predicție a traiectoriei se propune o nouă arhitectură, bazată pe predicția video, detectarea obiectelor și segmentarea semantică. Arhitectura utilizată pentru această sarcină poate fi văzută ca o legătură între toate celelalte componente ale acestei teze și poate fi văzută, de asemenea, ca scop principal - de a proiecta un nou model de predicție a traiectoriei care poate fi antrenat fără a fi nevoie de date adnotate, folosind un arhitectura de predicție video ca rețea de bază. Capitolul este organizat astfel. Prima secțiune descrie cele mai importante seturi de date utilizate pentru sarcina de predicție a traiectoriei și, de asemenea, descrie setul de date utilizat pentru experimentele curente. Următoarea secțiune descrie experimentele făcute pentru această sarcină și, de asemenea, metricile utilizate în evaluarea rezultatelor. Următoarea secțiune prezintă arhitectura propusă pentru această sarcină și variațiile acesteia în ceea ce privește experimentele curente. Secțiunea finală a capitolului prezintă și analizează rezultatele.

7.1 Setul de date pentru predicția traiectoriei

După cum sa menționat mai devreme, pentru această sarcină se utilizează un set de date înregistrat în campusul universitar Politehnica și anume setul de date de segmentare POLI. Setul de date constă dintr-un înregistrator de filme scurte în timpul zilei, amurgului și nopții. Setul de date original a fost folosit pentru sarcina de segmentare a mașinii, așa cum a fost deja menționat în una dintre secțiunile anterioare. Au fost alese cele mai relevante cadre, cu cel puțin un vehicul în el, pentru a testa sistemul de predicție a traiectoriei pe aceste imagini. Pentru toate imaginile existente s-au format scurtmetraje care conțin 35 de cadre. Din acest număr, 30 sunt considerate a fi cunoscute și utilizate pentru a alimenta rețelele neuronale și 5 urmează a fi detectate. Sunt 106 videoclipuri care au fost înregistrate în timpul zilei, 36 înregistrate în amurg sau în zori și 47 în

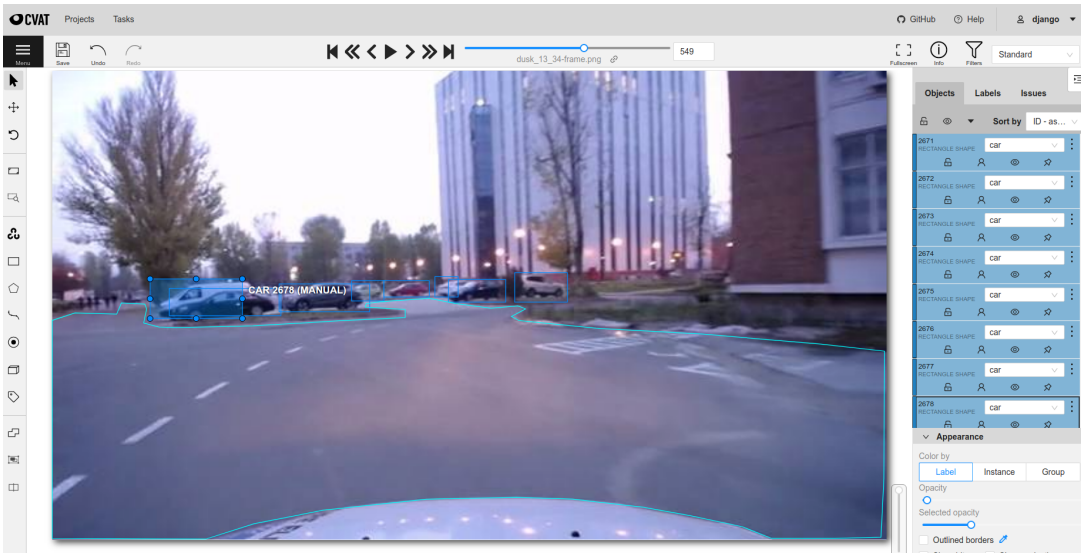

Fig. 7.1 Adnotarea drumului și a mașinilor pentru predicția traiectoriei

timpul nopții, deci un total de 189 de videoclipuri scurte, care conțin 6675 de cadre. Toate cadrele sunt adnotate pentru segmentarea drumului și detectarea mașinilor - segmentarea a fost păstrată din setul de date de segmentare POLI și adnotarea mașinilor a fost făcută special pentru această sarcină, folosind instrumentul CVAT. Un exemplu de segmentare a drumurilor și de adnotare auto folosind CVAT poate fi văzut în figura 7.1.

Cadrele conțin aproximativ 4000 de mașini adnotate în imaginile reale doar în cadrele care trebuie prevăzute (în aproximativ 945 de cadre), pe lângă celelalte mașini care au fost în primele cadre cunoscute și nu trebuiau adnotate.

7.2 Arhitectura propusă

7.2.1 Modelul generic

Arhitectura propusă pentru predicția traiectoriei implică multe rețele, câte una pentru fiecare dintre sarcinile descrise mai devreme – generarea video, estimarea adâncimii, segmentarea semantică și detectarea obiectelor. Intrarea arhitecturii constă în videoclipuri mici, care conțin aproximativ 35 de cadre – primele 30 sunt considerate a fi cunoscute și sunt folosite pentru a alimenta rețelele de generare video, iar ultimele 5 trebuie prezise. Prima rețea implicată în acest proces este rețeaua de generare video. Rețeaua de generare video a variat în experimentele efectuate, așa cum se poate observa în capitolul următor, care descrie rezultatele. După pasul de generare a videoclipului, s-au obținut câteva cadre prezise pentru scena dată. Predicția a fost făcută luând în considerare doar ultimele 5 cadre prezise. Cadrele sunt prezise din aceleași imagini – primele 30 de cadre sunt folosite pentru a prezice ultimele 5. Unele rețele adaugă următorul cadru prezis la

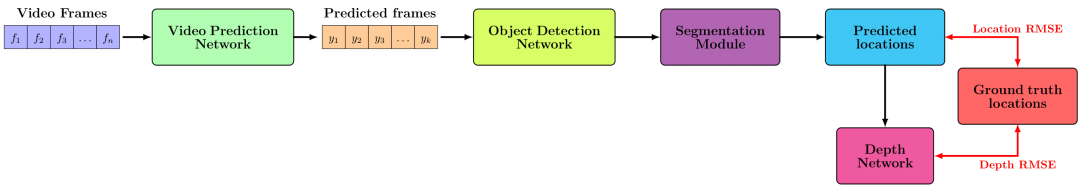

Fig. 7.2 Arhitectura generică propusă pentru predicția traiectoriei

cadrele reale pentru a prezice altul. Odată ce imaginile prezise sunt obținute, cadrele trec printr-o rețea de detectare a obiectelor – în cazul actual, YOLO v4. Mașinile au fost și adnotate manual pentru o mai bună estimare a generației și, de asemenea, pentru a înlătura greșelile făcute de rețeaua de detecție. Unele dintre cadrele generate au o calitate scăzută și au trebuit să fie redimensionate la o rezoluție mai mică. Chiar dacă un om ar putea identifica în continuare poziția majorității mașinilor într-o imagine de rezoluție scăzută, rețeaua de detectare a obiectelor va avea probleme în ceea ce privește detectarea mașinilor, motiv pentru care au fost testate atât detectarea adevărului la sol, cât și rezultatele YOLO. După acest pas, dar independent de acesta, imaginile trec printr-o rețea de segmentare semantică – FCN, în cazul de față. Acest pas este necesar pentru a utiliza segmentarea drumului ca informație suplimentară pentru viitoarele mașini. Există un alt modul care folosește segmentarea drumului și învață cel mai bun mod de a reajusta poziția mașinii față de drum. Poziția relativă a mașinii față de drum se calculează cu poziția veche și, de asemenea, cu noua poziție, apoi se obține poziția finală ca medie ponderată a celor două poziții. Chiar daca imbunatatirea nu este mare, s-au obtinut unele imbunatatiri avand in vedere rezultatele fara a tine cont de segmentarea drumului.

Ultimul pas înainte de a calcula valorile reale este de a pune toate cadrele prezise și cadrele reale ca intrare pentru o rețea de estimare a adâncimii. Cu aceste rezultate – pozițiile mașinilor, obținute luând în considerare segmentarea drumului, dar și adâncimea cadrelor, pot fi calculate două metrici relevante – RMSE pentru locații și, de asemenea, RMSE pentru adâncimea în ceea ce privește acele locații.

În acest fel, pot fi măsurate atât distanța dintre locația reală și locația prezisă, cât și distanța dintre adâncimea locației reale a mașinii și locația estimată a mașinii. Acest lucru ar putea fi mai precis decât luarea în considerare doar a distanței în pixeli. Mai multe detalii despre valorile vor fi oferite în secțiunea următoare. Arhitectura generică poate fi văzută în Figura 7.2.

7.2.2 Modele specifice

Arhitectura anterioară este un model generic, fără a ține cont de rețelele utilizate. Scopul principal al tezei este de a studia cele mai importante arhitecturi privind fiecare dintre sub-sarcinile existente din modelul principal și de a folosi cele mai bune modele pentru

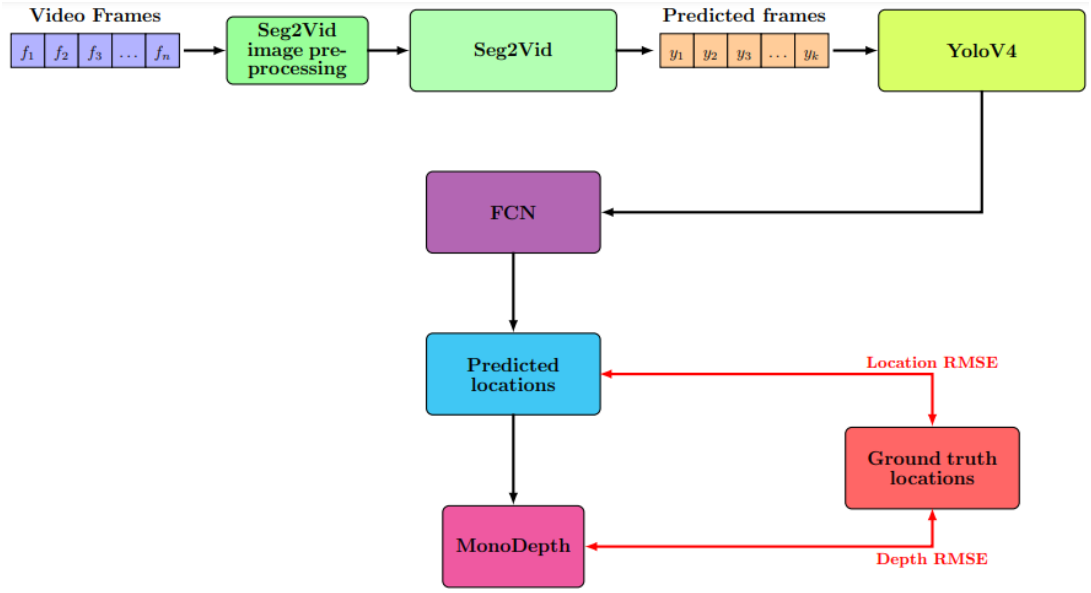

Fig. 7.3 Arhitectura propusă pentru SAVP

a avea o arhitectură competitivă.

Pentru experimentele curente, YOLO v4 a fost selectată ca cea mai bună arhitectură în ceea ce privește detectarea obiectelor, FCN ca cea mai versatilă arhitectură pentru segmentarea semantică și Monodepth ca cea mai bună arhitectură în ceea ce privește nevoia experimentelor. În ceea ce privește arhitectura de generare video utilizată, sunt trei modele propuse și testate.

Arhitectura modelului bazat pe SAVP poate fi văzută în 7.3. Diferența dintre modelul generic este că rețeaua specifică utilizată sunt reprezentate și în această figură - SAVP, YOLO, MonoDepth și FCN. Arhitectura pentru Segnet poate fi văzută în 7.4, iar arhitectura pentru PredNet poate fi văzută în 7.5. Fiecare dintre aceste arhitecturi a fost testată folosind setul de date descris în acest capitol, iar rezultatele sunt analizate în secțiunea următoare. Arhitecturile prezentate reprezintă una dintre cele mai importante contribuții ale tezei.

7.3 Experimente și rezultate

În această secțiune sunt descrise experimentele utilizate pentru sarcina de predicție a traiectoriei, cum au fost calculate metricile.

7.3.1 Rezultate

În această secțiune sunt prezentate rezultatele obținute la toate experimentele efectuate privind predicția traiectoriei și generarea video, implicând RMSE pentru locația prezisă, adâncimea prezisă și, de asemenea, RMSE privind dimensiunea mașinii și tim-

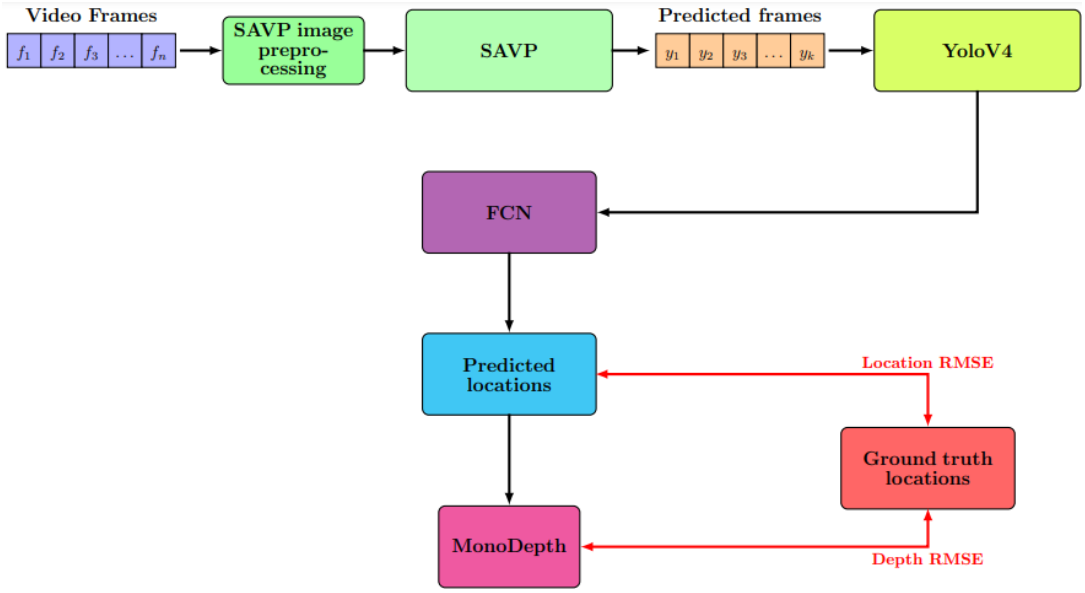

Fig. 7.4 Arhitectura propusă pentru Seg2Vid

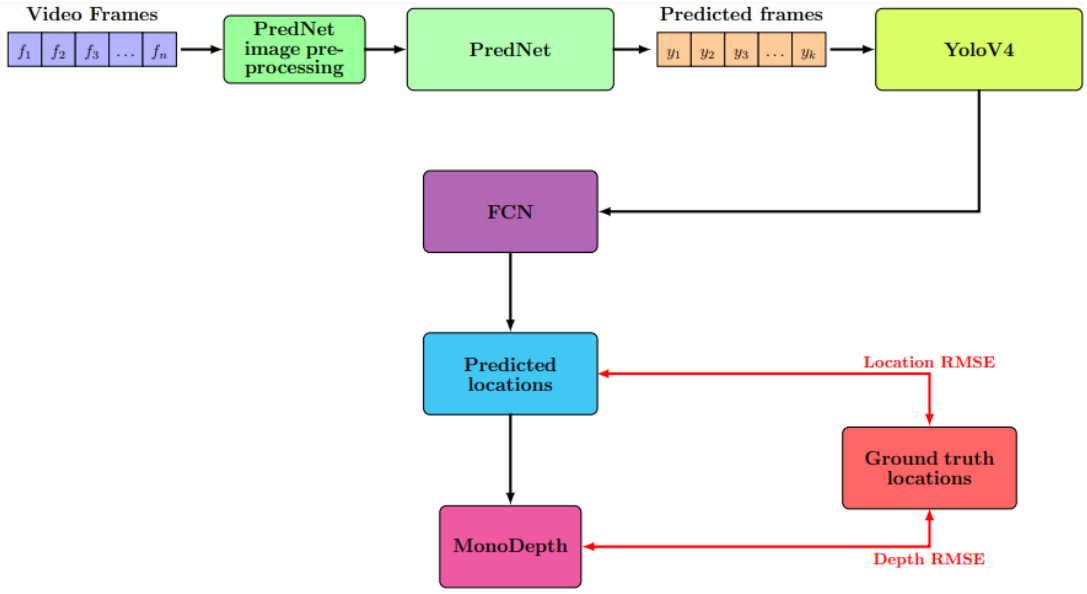

Fig. 7.5 Arhitectura propusă pentru PredNet

pul de inferență privind dimensiunea imaginii. În figura 7.6 sunt prezentate unele dintre predicții. Există o predicție pentru fiecare dintre cele trei rețele utilizate (PredNet, SAVP și Segnet) împreună cu adevărul de la sol pentru fiecare scenariu implicat (zi, amurg și noapte). În Tabelul 7.1 și Tabelul 7.2 sunt afișate rezultatele pentru locație, iar în Tabelul 7.3 și Tabelul 7.4 sunt afișate rezultatele pentru adâncime. Fiecare rezultat este analizat în paragrafele următoare.

Fig. 7.6 Rezultatele generării video

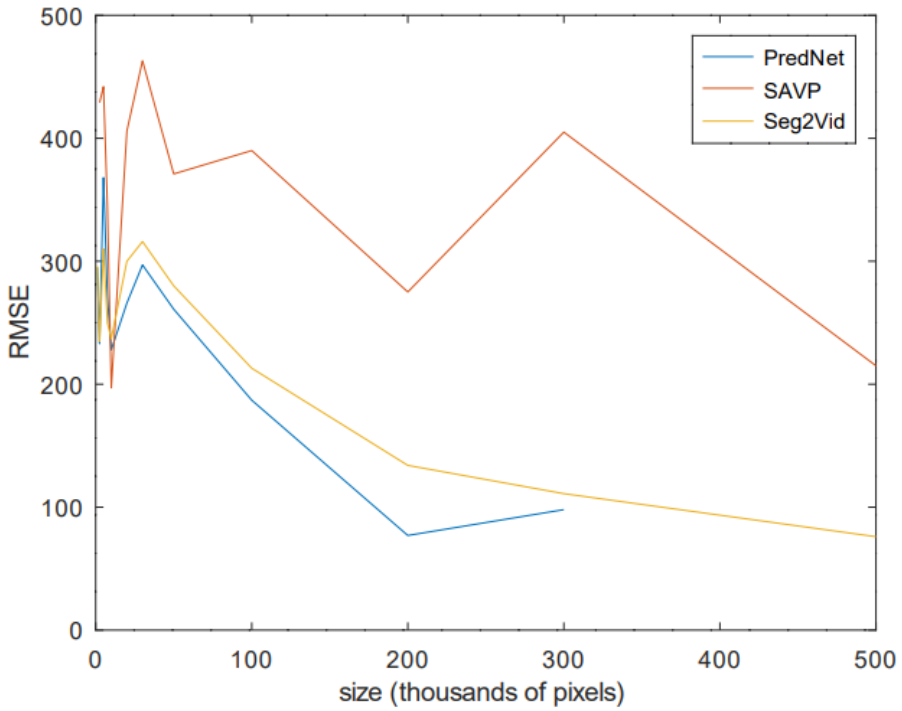

Fig. 7.7 RMSE raportat la mărimea mașinii

7.4 Un model îmbunătățit pentru predicția traiectoriei

În această secțiune sunt descrise trei modele noi propuse ca contribuție originală în această teză, având în vedere sarcina de predicție a traiectoriei. Rețeaua PredNet este mai bine analizată și de asemenea sunt propuse trei modificări în această teză având în

Tabelul 7.1 RMSE pentru locație (detecție reală)

	zi	amurg	noapte	Avg
PredNet				
No segmentation	318.20	71.73	94.28	247.91
GT segmentation	317.20	71.10	93.70	247.11
FCN segmentation	317.12	71.24	93.49	247.00
SAVP				
No segmentation	294.46	128.88	116.83	233.71
GT segmentation	294.12	128.84	116.34	233.41
FCN segmentation	294.12	128.85	116.15	233.41
Seg2Vid				
No segmentation	321.20	148.91	115.27	265.85
GT segmentation	316.06	145.65	113.58	261.67
FCN segmentation	320.19	147.16	114.63	264.91
TraPHic	83.78	54.60	114.27	86.29

Tabelul 7.2 RMSE pentru locație (detecție YOLO)

	zi	amurg	noapte	Avg
PredNet				
No segmentation	280.01	193.69	46.45	269.98
GT segmentation	275.45	157.91	46.40	263.69
FCN segmentation	275.45	158.58	46.40	263.69
SAVP				
No segmentation	615.78	568.45	543.74	561.84
GT segmentation	494.61	480.95	322.51	390.35
FCN segmentation	497.73	478.46	323.96	393.00
Seg2Vid				
No segmentation	320.88	278.37	103.58	310.92
GT segmentation	306.16	191.44	98.29	287.88
FCN segmentation	306.36	196.98	102.15	289.27

vedere modelul de bază, cu rezultate mai bune în ceea ce privește sarcina de predicție a traiectoriei.

7.4.1 Arhitectura propusă

La un nivel superior, PredNet poate fi văzut ca mai multe straturi convoluționale recurente, a căror ieșire trece printr-o activare a unei unități liniare rectificate (ReLU) și un strat de max-pooling cu pas 2. Acum, în ceea ce privește straturile recurente convoluționale, acestea constau din patru diferite. straturi de circumvoluții. Primul este un strat de reprezentare, care este un strat recurent care face o predicție pe baza intrării de reprezentare curentă. Intrarea și predicția reprezintă alte două straturi de convoluții. Ultimul strat este un strat de eroare care este calculat pe baza intrării și a predicției și devine următorul strat de intrare. Stratul de reprezentare la un pas dat se bazează pe

Tabelul 7.3 RMSE pentru adâncime (detecție reală)

	zi	amurg	noapte	Avg
PredNet				
No segmentation	142.08	8.63	23.60	104.78
No segmentation (pred)	142.98	14.02	25.74	105.68
GT segmentation	145.13	8.36	22.89	111.06
GT segmentation (pred)	145.63	13.18	24.13	111.60
FCN segmentation	145.58	8.54	23.46	111.37
FCN segmentation (pred)	146.54	13.92	25.48	112.16
SAVP				
No segmentation	141.49	13.34	25.95	103.33
No segmentation (pred)	126.14	18.58	31.47	93.01
GT segmentation	141.97	13.24	25.23	106.49
GT segmentation (pred)	126.38	17.41	31.29	95.68
FCN segmentation	141.98	13.24	25.70	106.49
FCN segmentation (pred)	126.88	18.50	31.30	95.96
Seg2Vid				
No segmentation	134.10	18.68	27.52	103.76
No segmentation (pred)	129.25	23.52	32.47	100.62
GT segmentation	138.29	18.57	27.41	110.01
GT segmentation (pred)	132.85	22.54	30.79	106.21
FCN segmentation	138.28	18.66	27.41	110.01
FCN segmentation (pred)	132.84	23.36	32.23	106.20

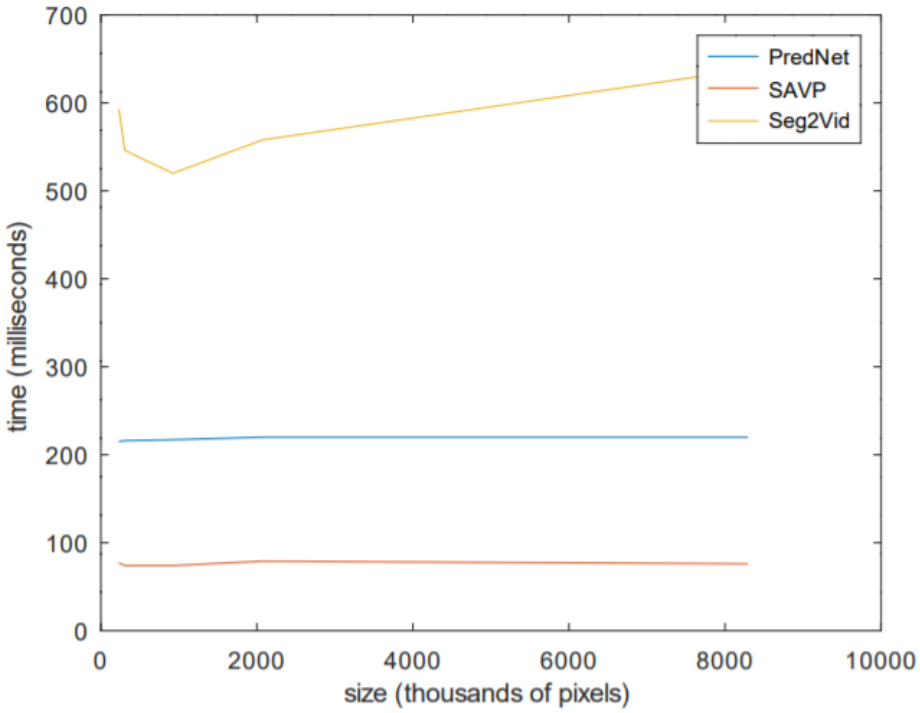

Fig. 7.8 RMSE raportat la mărimea imaginii

Tabelul 7.4 RMSE pentru adâncime (detecție YOLO)

	zi	amurg	noapte	Avg
PredNet				
No segmentation	69.69	20.21	6.87	65.08
No segmentation (pred)	168.77	24.93	0.64	156.86
GT segmentation	62.19	19.58	6.51	58.89
GT segmentation (pred)	165.66	16.27	0.55	154.70
FCN segmentation	65.27	19.58	5.94	61.76
FCN segmentation (pred)	159.33	19.52	0.55	149.27
SAVP				
No segmentation	38.66	29.47	32.69	33.45
No segmentation (pred)	81.66	69.23	44.97	57.21
GT segmentation	34.76	31.80	36.41	36.09
GT segmentation (pred)	78.15	49.28	41.71	52.57
FCN segmentation	34.95	31.80	36.41	36.09
FCN segmentation (pred)	70.73	48.20	42.04	52.57
Seg2Vid				
No segmentation	62.51	25.72	20.26	56.57
No segmentation (pred)	137.39	40.86	19.79	122.98
GT segmentation	64.13	25.91	19.60	58.77
GT segmentation (pred)	136.17	31.78	14.89	123.16
FCN segmentation	64.10	25.91	19.60	58.75
FCN segmentation (pred)	134.93	32.61	19.78	122.65

stratul de reprezentare la pasul anterior, stratul de eroare la pasul anterior și, de asemenea, pe stratul de reprezentare la pasul următor (care poate fi obținut inițial prin utilizarea supraeșantionării). Rețeaua principală este realizată pentru a prezice doar un singur cadru viitor, având în vedere un videoclip de intrare, totuși rețeaua poate fi, de asemenea, reglată pentru a prezice până la cinci cadre în viitor. Pentru experimentele actuale, arhitecturile au fost, de asemenea, reglate fin pentru a prezice cinci cadre viitoare, având în vedere doar videoclipul inițial.

Această cercetare propune trei versiuni diferite ale reprezentării interne a straturilor convoluționale. Versiunea standard folosește un model cu patru straturi cu convoluții 3x3 pentru predicția imaginilor de conducere, așa cum se poate vedea în depozitul lor Git. Modelele propuse sunt urmatoarele:

P_5_5 înlocuiește pur și simplu convoluțiile 3x3 cu convoluții 5x5, fără a adăuga straturi suplimentare.

P_3_5 este un model cu 6 straturi cu două straturi convoluționale suplimentare 3x3, având în vedere modelul anterior, P_5_5. De asemenea, înlocuiește activarea ReLU cu

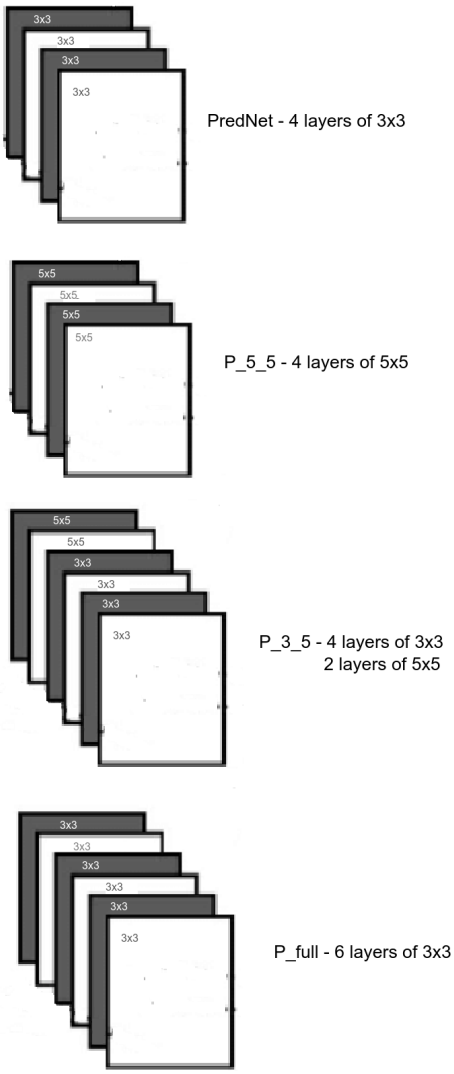

Fig. 7.9 Structuri convoluționale propuse

PReLU, care în loc să pună la zero valori negative învață un parametru care se înmulțește cu valoarea răspunsului, conform următoarei ecuații:

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ a * x & \text{otherwise} \end{cases} \tag{7.1}$$

În cele din urmă, P_full este, de asemenea, un model cu 6 straturi format din doar 3x3 straturi convoluționale și care utilizează, de asemenea, funcția de activare PRELU.

Modificările pot fi văzute mai bine în Figura 7.9.

Fluxul de lucru privind predicția traiectoriei este similar cu cel prezentat în 7.5, cu singura diferență că arhitectura PredNet este înlocuită cu noile variații, P_3_5, P_5_5 și P_full.

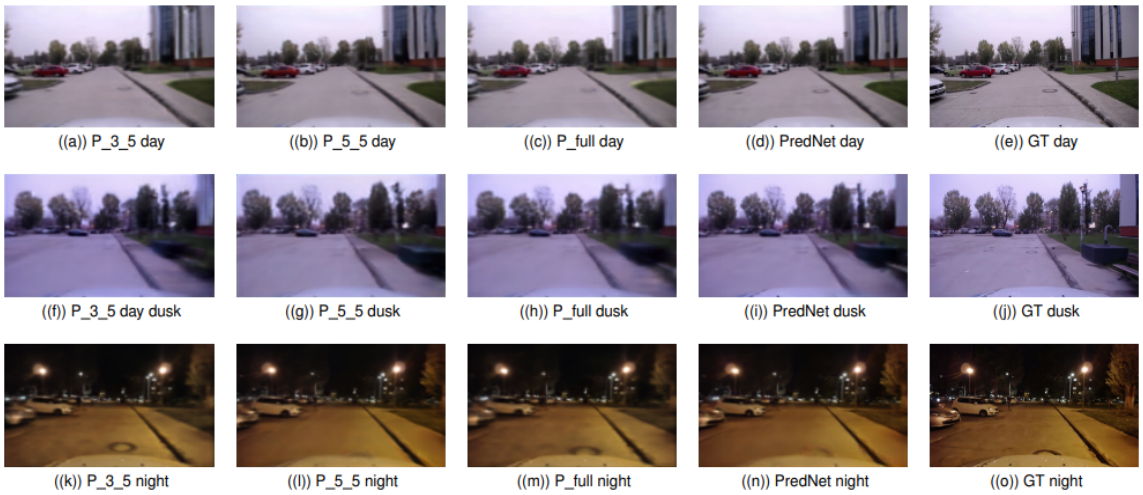

Fig. 7.10 Rezultate predicție - arhitecturi modificate

Rezultatele pot fi găsite în Tabelul 7.5, Tabelul 7.6 și, de asemenea, în Figura 7.10 și în Figura 7.11. În figura 7.10 există câteva imagini cu al doilea cadru prezis în diferite scenarii din fiecare arhitectură, inclusiv arhitectura originală Prednet și adevărul de bază. Tabelul 7.5 conține NRMSE cu privire la locația pentru fiecare dintre cele optsprezece setări și cu privire la ora din zi. Tabelul 7.6 conține NRMSE cu privire la adâncimea pentru aceleași setări ca în primul tabel. În cele din urmă, în Figura 7.11 se poate vedea o diagramă pentru NRMSE ținând cont de dimensiunea mașinii, luând în considerare ultima configurație - detecțiile de la YOLO și segmentarea din rețeaua FCN. În tabele tipul de segmentare este descris prin „fără segment", „segm GT" sau „segm FCN", având în vedere că nu a fost utilizată nicio segmentare pentru îmbunătățirea rezultatelor, segmentarea a fost utilizată folosind adevărul de bază, iar segmentarea a fost utilizată folosind Rețeaua FCN pentru imaginile prezise. Detecția utilizată este descrisă de „GT det" și „YOLO det", adică poziția mașinilor a fost luată în considerare prin datele adnotate manual și, respectiv, prin rezultatele YOLO. În final, în tabelul 7.6, dacă adâncimea a fost calculată luând în considerare încadrarea prezisă, apare abrevierea „pred D".

Tabelul 7.6 NRMSE pentru adâncime

	zi	amurg	noapte	Avg
PredNet				
No segm, GT det	0.555	0.034	0.092	0.409
No segm, pred D, GT det	0.559	0.055	0.101	0.413
GT segm, GT det	0.567	0.033	0.089	0.434
GT segm, pred D, GT det	0.569	0.051	0.094	0.436
FCN segm, GT det	0.569	0.033	0.092	0.435
FCN segm, pred D, GT det	0.572	0.054	0.100	0.438

	zi	amurg	noapte	Avg
No segm, YOLO det	0.272	0.079	0.027	0.254
No segm, pred D, YOLO det	0.659	0.097	0.003	0.613
GT segm, YOLO det	0.243	0.076	0.025	0.230
GT segm, pred D, YOLO det	0.647	0.064	0.002	0.604
FCN segm, YOLO det	0.255	0.076	0.023	0.241
FCN segm, pred D, YOLO det	0.622	0.076	0.002	0.583
P_3_5				
No segm, GT det	0.555	0.041	0.083	0.418
No segm, pred D, GT det	0.552	0.046	0.076	0.415
GT segm, GT det	0.559	0.040	0.083	0.436
GT segm, pred D, GT det	0.556	0.045	0.076	0.433
FCN segm, GT det	0.557	0.041	0.083	0.434
FCN segm, pred D, GT det	0.555	0.046	0.076	0.432
No segm, YOLO det	0.293	0.110	0.055	0.263
No segm, pred D, YOLO det	0.699	0.168	0.087	0.623
GT segm, YOLO det	0.264	0.130	0.065	0.246
GT segm, pred D, YOLO det	0.682	0.083	0.066	0.614
FCN segm, YOLO det	0.241	0.116	0.050	0.241
FCN segm, pred D, YOLO det	2.401	0.073	0.070	0.577
P_5_5				
No segm, GT det	0.561	0.046	0.063	0.416
No segm, pred D, GT det	0.561	0.048	0.072	0.416
GT segm, GT det	0.565	0.044	0.063	0.432
GT segm, pred D, GT det	0.562	0.046	0.072	0.430
FCN segm, GT det	0.568	0.046	0.063	0.433
FCN segm, pred D, GT det	0.566	0.048	0.072	0.432
No segm, YOLO det	0.271	0.096	0.045	0.246
No segm, pred D, YOLO det	0.633	0.092	0.052	0.568
GT segm, YOLO det	0.254	0.095	0.041	0.235
GT segm, pred D, YOLO det	0.630	0.059	0.052	0.567
FCN segm, YOLO det	0.257	0.095	0.041	0.237
FCN segm, pred D, YOLO det	0.593	0.063	0.052	0.536
P_full				
No segm, GT det	0.544	0.036	0.071	0.410
No segm, pred D, GT det	0.541	0.048	0.089	0.409
GT segm, GT det	0.557	0.035	0.069	0.434
GT segm, pred D, GT det	0.553	0.045	0.083	0.432

	zi	amurg	noapte	Avg
FCN segm, GT det	0.555	0.036	0.071	0.433
FCN segm, pred D, GT det	0.552	0.048	0.087	0.432
No segm, YOLO det	0.293	0.110	0.055	0.263
No segm, pred D, YOLO det	0.660	0.129	0.048	0.584
GT segm, YOLO det	0.272	0.116	0.051	0.250
GT segm, pred D, YOLO det	0.680	0.083	0.054	0.610
FCN segm, YOLO det	0.237	0.112	0.046	0.237
FCN segm, pred D, YOLO det	0.640	0.069	0.066	0.574

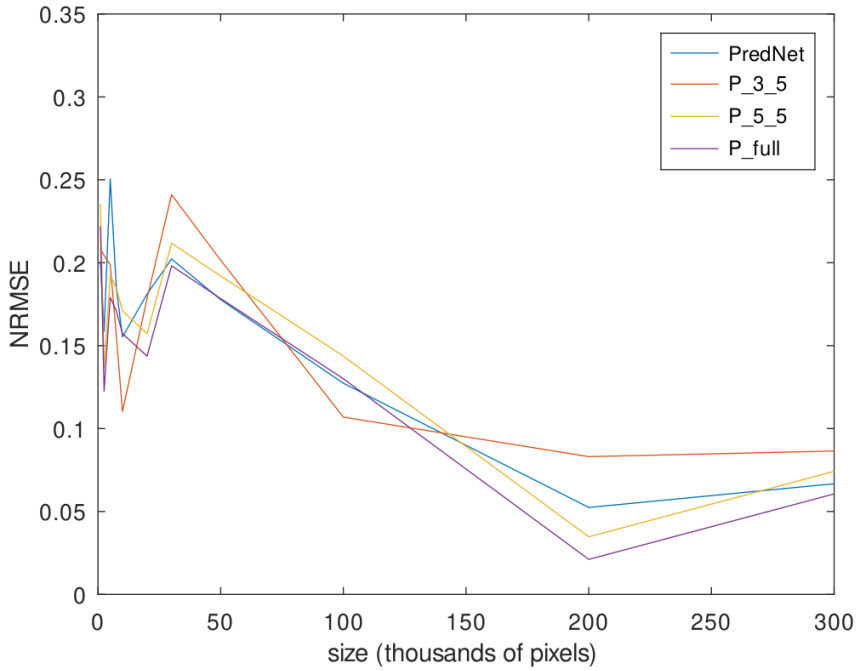

Fig. 7.11 NRMSE raportat la dimensiunea mașinilor

Tabelul 7.5 NRMSE pentru locație

	zi	amurg	noapte	Avg
PredNet				
No segm, GT det	0.217	0.049	0.064	0.169
GT segm, GT det	0.216	0.048	0.064	0.168
FCN segm, GT det	0.216	0.049	0.064	0.168
tNo segm, YOLO det	0.191	0.132	0.032	0.184
GT segm,YOLO det	0.188	0.108	0.032	0.180
FCN segm, YOLO det	0.188	0.108	0.032	0.180
P_3_5				
No segm, GT det	0.208	0.056	0.061	0.164
GT segm, GT det	0.207	0.056	0.061	0.170
FCN segm, GT det	0.207	0.056	0.061	0.163
No segm, YOLO det	0.218	0.238	0.063	0.216
GT segm,YOLO det	0.215	0.157	0.065	0.210
FCN segm, YOLO det	0.211	0.156	0.062	0.206
P_5_5				
No segm, GT det	0.204	0.054	0.046	0.158
GT segm, GT det	0.204	0.053	0.046	0.158
FCN segm, GT det	0.204	0.054	0.046	0.158
No segm, YOLO det	0.225	0.136	0.050	0.210
GT segm,YOLO det	0.208	0.103	0.047	0.193
FCN segm, YOLO det	0.208	0.101	0.050	0.193
P_full				
No segm, GT det	0.209	0.057	0.050	0.164
GT segm, GT det	0.209	0.057	0.049	0.164
FCN segm, GT det	0.209	0.057	0.050	0.164
No segm, YOLO det	0.187	0.207	0.033	0.186
‘ GT segm,YOLO det	0.177	0.121	0.028	0.179
FCN segm, YOLO det	0.177	0.122	0.028	0.172
Traffic	0.057	0.037	0.078	0.059

Capitolul 8

Concluzii și lucrări viitoare

Prezenta teză abordează cele mai importante sarcini privind conducerea autonomă, de la înțelegerea scenei până la predicția traiectoriei pentru mașinile din jur. Teza a urmărit îndeaproape patru probleme diferite - detectarea obiectelor, segmentarea semantică a drumurilor, estimarea adâncimii și predicția traiectoriei. Pe lângă acestea, s-a mai discutat despre urmărirea obiectelor, segmentarea instanțelor și panoptice și generarea video. Pentru fiecare dintre aceste sarcini au fost descrise cele mai importante lucrări de cercetare și, de asemenea, articole de recenzie și seturile de date corespunzătoare pentru fiecare sarcină. De asemenea, unele dintre cele mai bune rețele au fost analizate pe niște seturi de date adnotate la campusul Universității Politehnica București, ținând cont de ora din zi, dimensiunea mașinilor, timpul de inferență și alte statistici, pentru a vedea cum se poate. să se obțină cea mai bună performanță într-o aplicație reală pentru conducerea autonomă. Scopul final al tezei este de a propune și implementa o arhitectură pentru predicția traiectoriei, bazată pe rezultatele evaluării pentru unele dintre cele mai bune arhitecturi din literatura de specialitate pentru detectarea obiectelor, segmentarea semantică, estimarea adâncimii și predicția video. Această abordare ar putea schimba jocul pentru sarcina de predicție a traiectoriei, deoarece nu implică date adnotate și poate fi antrenată folosind orice posibil videoclip de conducere de pe internet. De asemenea, fiecare sarcină este susținută de cel puțin o lucrare de cercetare, care conține unele dintre rezultatele prezentate în teză. Rezultatele sunt detaliate cu privire la fiecare dintre aceste patru sarcini individuale.

8.1 Rezultate notabile

Pentru sarcina de detectare a obiectelor au fost testate Yolo, Faster R-CNN, SSD și Retina Net. În experimentele efectuate, Yolo are cea mai bună reamintire medie, dar SSD are cea mai bună precizie medie medie. De asemenea, după cum era de așteptat, rezultatele sunt mai bune atunci când au fost testate doar două categorii – mașină și persoană, rechemarea este mai bună ziua decât noaptea, dar precizia nu are variații

importante în ceea ce privește ora zilei. Pentru sarcina de detectare a obiectelor a fost înregistrat un nou set de date, care a fost adnotat manual, în campusul Universității Politehnica din București. În ceea ce privește setul de date realizat în Universitatea Politehnica, rechemarea este mai bună decât pentru setul de date BDD100k, deoarece sunt mai puține obiecte implicate, dar precizia a rămas în același interval. Acest rezultat arată că, în general, detectoarele sunt robuste și au același comportament atât pe setul de date POLI, cât și pe setul de date BDD100k. În ceea ce privește dimensiunea obiectului, obiectele mai mari au fost detectate în general, există o creștere a retragerii, dar clasa detectată tinde să fie greșită în multe cazuri decât pentru obiectele mai mici, care au fost detectate slab, dar în cazul unei detectări clasa a fost atribuit corect obiectului. În ceea ce privește timpul necesar pentru o detecție, Yolo a avut cele mai bune rezultate și a putut performa într-un scenariu în timp real (mai ales în ceea ce privește modelul lor minuscul, care poate funcționa în timp real pentru 30 fps), cu mențiunea că este nevoie de un GPU puternic care rulează în pentru a atinge aceste rezultate.

Pentru sarcina de segmentare semantică au fost testate SegNet, ENet, PSPNet, Dilation și DaNet. Cele mai bune rețele au fost PSPNet și FCN, care au procente foarte bune pentru metricile afișate și pot fi folosite în aplicații reale, iar cea mai proastă rețea testată a fost DaNet, care nu poate fi de încredere într-o segmentare semantică rutieră fără o reglare ulterioară a rețeaua special pentru această sarcină. Cu toate acestea, modelul se limitează la segmentarea drumului și nu includea alte categorii, cum ar fi segmentarea mașinilor. Un set de date special conceput pentru această sarcină, Poli segmentation dataset, a fost înregistrat în campusul universitar. După cum era de așteptat, rezultatele au fost mai bune ziua și mai rele noaptea. Din păcate, în ceea ce privește timpul de procesare, rețelele nu pot funcționa în scenarii de viață reală, în care ar putea fi necesară o segmentare pentru 30 de cadre pe secundă – cel mult ar putea procesa doar câteva imagini pe secundă. Mai este loc de îmbunătățire în ceea ce privește acuratețea, dar cea mai mare problemă acum este timpul de inferență. Cu o singură excepție, timpul de evaluare a unei imagini nu a avut variații în ceea ce privește dimensiunea imaginilor, ceea ce este un aspect pozitiv.

Pentru sarcina de estimare a adâncimii s-au folosit două seturi de date – primul set de date a fost cel înregistrat cu camera Intel RealSense, iar al doilea este setul de date realizat anterior în campusul universitar Politehnica și utilizat pentru segmentarea semantică, care a fost folosit în vederea testării. rețelele una împotriva celeilalte, pentru a vedea care sunt avantajele unei astfel de informații în ceea ce privește o paradigmă de învățare nesupravegheată. Toate rezultatele au fost analizate privind ziua, amurgul și noaptea. RMSE a fost calculat pentru fiecare din această categorie și, de asemenea, o eroare medie pentru toate trei. S-a analizat viteza rețelelor pentru diferite dimensiuni de imagine și influența dimensiunii mașinii în ceea ce privește RMSE și, de asemenea,

s-a analizat atât eroarea de adâncime pentru toți pixelii din imagine, cât și eroarea de adâncime numai pentru mașină. Deloc surprinzător, eroarea luând în considerare doar mașinile a fost semnificativ mai mică. Majoritatea rețelelor au funcționat în perioade similare, chiar dacă dimensiunea imaginilor a fost diferită, ceea ce este un rezultat bun ținând cont de un scenariu de viață reală cu imagini full HD sau chiar 4K. De asemenea, se poate observa că precizia estimării adâncimii a fost mai bună pentru mașinile mai mari, ceea ce era și un rezultat așteptat. Din experimentele efectuate, cele mai bune rețele au fost SfMLearner și LKVOLearner, dar și Monodepth 2 au avut rezultate bune, având în vedere și o rafinare manuală a calității rezultatelor. Având în vedere că camera RGB-D nu este la fel de precisă ca un scaner LIDAR, o estimare calitativă a imaginilor rezultate a arătat că trebuie luată în considerare și Monodepth 2. Cu toate acestea, deoarece experimentele au fost foarte diferite pentru aproape fiecare rețea, există cel puțin un scenariu în care rețeaua a funcționat mai bine decât restul celorlalte, ceea ce ar putea fi explicat prin adevărul imprecis de teren și, de asemenea, rezultate imprecise ale rețelelor, în general. Vestea bună este că rețelele ar putea fi folosite în aplicații din viața reală, datorită vitezei lor, iar un sistem autonom ar putea beneficia de adâncimea lor estimată pentru mașinile din jur. Cu toate acestea, fără a folosi senzori scumpi, rezultatele estimate ar putea diferi mult pentru adâncimea reală, mai ales în amurg sau noaptea. Rezultatele variază în funcție de rețeaua utilizată și de adevărul de la sol, dar în general cele mai bune rezultate au fost obținute în timpul zilei.

Pentru sarcina finală de predicție a traiectoriei, au fost testate PredNet, Seg2Vid și SAVP, iar rezultatele lor sunt comparate cu TraPHic. Aici a fost folosit și setul de date POLI Segmentation. Cea mai relevantă rețea testată a fost PredNet când s-a luat în considerare RMSE pentru locație cu previziunile YOLO pentru mașini, totuși SAVP a avut și rezultate bune. Au fost dezvoltate trei noi variante ale modelelor PredNet, cu modelul final, PredNet_full, obținând rezultate mai bune decât versiunea de bază în aproape toate experimentele. Deși rezultatele nu sunt încă la fel de bune ca o rețea specializată de predicție a traiectoriei, cel mai mare avantaj este că o rețea de generare video poate fi antrenată pe orice videoclip, facilitând procesul de antrenament. Cu o rețea de generație foarte bună, traiectoriile ar putea fi deduse cu ușurință. De asemenea, rezultatele au arătat că segmentarea drumului ar putea îmbunătăți ușor detectarea simplă a mașinilor în cadrele prevăzute.

8.2 Contribuții originale

Această teză are câteva contribuții noi cu privire la sarcinile de detectare a obiectelor, segmentare semantică, estimare a adâncimii și predicție a traiectoriei, care ar putea ajuta comunitatea de cercetare să dezvolte algoritmi mai buni de conducere autonomă.

Prima și cea mai importantă contribuție este un nou model de predicție a traiectoriei, care se bazează pe generarea video și utilizează detectarea obiectelor, segmentarea semantică și estimarea adâncimii. Fiecare dintre sarcinile particulare care sunt legate de arhitectura finală a fost analizată profund pentru a utiliza cele mai bune rețele existente în modelul final. Chiar dacă o predicție de traiectorie dedicată obține, în acest moment, rezultate mai bune, un model bazat pe generarea video are avantajul că nu necesită date de antrenament adnotate manual și poate fi antrenat folosind orice videoclip de condus care există pe internet. Au fost propuse și testate trei arhitecturi diferite bazate pe trei rețele diferite, iar rezultatele au fost prezentate în teză.

A doua contribuție constă în propunerea a trei modele diferite de predicție video, prin modificarea arhitecturii originale PredNet privind straturile convoluționale și funcția de activare. Două dintre modelele propuse au în general rezultate mai bune decât rețeaua originală, iar ultimul model propus, P_full, depășește rețeaua originală în aproape orice experiment.

A treia contribuție constă în trei noi seturi de date diferite înregistrate în campusul Universității Politehnica din București și adnotate manual. Primul set de date a fost folosit pentru detectarea obiectelor, al doilea set de date a fost realizat pentru segmentarea semantică a drumurilor și, de asemenea, pentru predicția traiectoriei, conținând adnotări ale drumului și ale mașinilor, iar ultimul set de date a fost colectat folosind o cameră Intel RGB-D și a fost folosit pentru sarcina de estimare a adâncimii.

A patra contribuție constă într-o testare amplă a diferitelor arhitecturi pentru conducerea autonomă luând în considerare lumina și timpul zilei pentru diferite sarcini (detecția obiectelor, segmentarea semantică, estimarea adâncimii, predicția traiectoriei), abordare care nu este foarte des luată în considerare în literatura de specialitate. . Deoarece, în general, rețelele sunt instruite pe anumite seturi de date specifice, rezultatele pot varia foarte mult dacă experimentele sunt făcute în timpul zilei, în zori sau amurg sau în timpul nopții. Din experimentele actuale se poate observa că rezultatele tind să fie semnificativ mai rele în timpul nopții. Acest lucru se datorează lipsei de pregătire în condiții de zori sau noapte și ar trebui luat în considerare în dezvoltarea altor seturi de date și în pregătirea arhitecturilor viitoare.

A cincea contribuție care poate fi evidențiată este o trecere în revistă la zi a arhitecturilor de ultimă generație privind sarcinile enumerate și o analiză comparativă. Recenziile recenziilor discută aspecte arhitecturale și includ teste diferite pentru unele dintre cele mai bune arhitecturi cu statistici diferite, cum ar fi ora din zi, dimensiunea mașinii, timpul de inferență.

8.3 Lucrări viitoare

Pentru sarcina de detectare a obiectelor, detectarea ar trebui să fie îmbunătățită în viitor pentru a avea încredere într-o rețea pentru o aplicație de mașină cu conducere autonomă, iar studiul actual poate fi continuat prin propunerea unei noi arhitecturi de detectare a obiectelor. Chiar dacă precizia are valori decente, rechemarea ar trebui îmbunătățită pentru a avea mai multe obiecte detectate – un obiect care nu este detectat este o posibilă cauză a unui accident, deci este important să existe o rechemare mai bună în rețelele viitoare. De asemenea, timpul poate fi îmbunătățit, deoarece în afară de detecție există și alte componente care trebuie să ruleze între două cadre (segmentare, vehicul și predicție de adâncime etc.), deci este nevoie și de un timp de inferență mai bun. Pe viitor, ar fi o idee bună ca aceste modele să fie reglate fin pentru setul de date Politehnica, pentru a vedea cum se vor ajusta parametrii atunci când rețelele sunt antrenate pe același set de date. De asemenea, setul de date ar putea fi mărit în ceea ce privește numărul obiectelor și diversitatea acestora.

Pentru segmentarea semantică, studiul actual ar trebui să extindă setul de date cu mai mult de o clasă, pentru a vedea rezultatele segmentării cel puțin pentru vehicule și oameni. De asemenea, mai multe rețele ar trebui ajustate fin în aplicațiile viitoare ale acestui studiu, în special pentru sarcina de segmentare a drumurilor (pentru a scoate doar două categorii, drum sau nu), pentru a vedea cum se vor îmbunătăți rezultatele.

Pentru estimarea adâncimii, studiul actual ar trebui să realizeze un set de date mai bun cu senzori LiDAR realizati pentru a estima mai bine eroarea de estimare. Erorile ar putea fi, de asemenea, diminuate dacă rețelele ar fi antrenate folosind setul de date dorit, dar în acest scop setul de date ar trebui înregistrat cu camere stereo, deoarece majoritatea rețelelor sunt antrenate cu seturi de date stereo și testate cu cele monoculare.

Pentru sarcina de predicție a traiectoriei, aplicațiile viitoare ale acestui studiu ar trebui să dezvolte și să antreneze în mod specific un model de generare video, luând în considerare în special sarcina de predicție a traiectoriei. De asemenea, la fiecare pas (detecție, segmentare, estimare a adâncimii) modelele corespunzătoare ar putea fi reglate fin pentru a funcționa mai bine pentru un anumit set de date și pentru sarcina de predicție a traiectoriei în conducerea autonomă.

Articole publlicate

D. T. Iancu, A. M. Florea, “An improved vehicle trajectory prediction model based on video generation”, Studies in Informatics and Control.Acceptat pentru publicare, vol. 32, no 1, 2023

D. T. Iancu, M. Nan, A. S. Ghita și A. M. Florea, “Trajectory Prediction using Video Generation in Autonomous Driving,” Studies in Informatics and Control, vol. 31, no. 1, pp. 37–48, 2022. WOS:000779783700004, Impact Factor 2.18

A. S. Ghita, A. M. Florea, M. Nan și D. T. Iancu, “People Trajectory Prediction applied on Social Robotics Scenarios,” UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science, 2022. În curs de publicare.

D. T. Iancu, M. Nan, A. S. Ghita și A. M. Florea, “Vehicle Depth Estimation for Autonomous Driving”, UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science, pp. 3–20, 2021. WOS:000692196300001

V. Radu, M. Nan, M. Trascau, D. T. Iancu, A. S. Ghita și A. M. Florea,“Car crash detection in videos”, in 2021 23rd International Conference on Control Systems and Computer Science (CSCS), pp. 127–132, IEEE, 2021. În curs de indexare WOS.

A. S. Ghita, M. Nan, D. T. Iancu și A. M. Florea, “Top-level Scene Information Extraction from Eye-level View Images”, in 2021 23rd International Conference on Control Systems and Computer Science (CSCS), pp. 133–137, IEEE, 2021. În curs de indexare WOS.

D.T. Iancu, A. Sorici, A.M.Florea, “Neural road semantic segmentation in driving scenarios, International Conference on Electronics”, Computers and Artificial Intelligence (ECAI) (pp. 1-6), 2020 IEEE. WOS:000627393500073

D.T. Iancu, A. Sorici, A.M.Florea, “Object detection in autonomous driving - from large to small datasets”, International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-6), 2019 IEEE. WOS:000569985400026

Participare în granturi de cercetare

A. M. Florea et al. PETRA - People Detection and Tracking for Social Robots and Autonomous Cars, PN-III-P2-2.1-PED-2019-4995, 2020-2022.

A. M. Florea et al. ROBIN - Robots and Society: Cognitive Systems for Personal Robots and Autonomous Vehicles, complex project Nr. 72 PCCDI, 2018-2020.