

UNIVERSITATEA BABEȘ-BOLYAI
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ



Dezvoltarea metodelor eficiente de învățare profundă în viziunea computerizată

Rezumatul tezei de doctorat

Student doctorand: Tudor Alexandru G. ILENI
Supraveghetor științific: Prof. dr. Anca M. ANDREICA

2024

Cuvinte cheie: viziune computerizată, învățare profundă, învățare automată, căutare arhitectură neuronală, rulare dinamică, distilare a cunoștințelor

Conținut

Cuprinsul tezei	2
Lista publicațiilor	4
1 Introducere	6
1.1 Motivație și ipoteze	6
1.2 Obiective	6
1.3 Contribuții originale	7
2 Contextul general și Lucrări similare	10
2.1 Metode tradiționale de procesare a imaginilor și tehnici de învățare profundă în viz- iunea computerizată	10
2.2 Abordări comune în reconstrucția 3D de față și calibrarea camerei	11
2.3 Prezentare generală a căutării arhitecturii neuronale automate	11
2.4 Optimizarea mecanismelor de învățare a ansamblului	12
2.5 Comprimarea modelului prin distilare de cunoștințe bazată pe atenție	12
3 Extracția și analiza caracteristicilor faciale	13
3.1 Descrierea metodei de analiză a feței	13
3.2 Rezultatele numerice ale analizei feței	14
4 Dezvoltarea Automată și Dinamică a Rețelelor de Învățare Profundă	15
4.1 Căutarea Arhitecturii Neuronale pentru Rețea Specializată Optimă	15
4.2 Execuția Dinamică în Ansambluri folosind Mecanismul de Gating	16
4.3 Distilarea Cunoștințelor prin Învățare Bazată pe Atenție	16
5 Reconstrucția Scenei 3D și a Feței	18
5.1 Calibrarea Camerei folosind o Abordare Evolutivă	18
5.2 Modelarea 3D a Feței Folosind Modele de Bază Deformabile	18
6 Concluzii	20
6.1 Abordarea unificată a analizei feței în spațiul 2D	20
6.2 Dincolo de analiza 2D folosind reconstrucția feței 3D	21
6.3 Îmbunătățiri căutarea arhitecturii, instruirea și inferența în modelele de învățare pro- fundă	22
Bibliography	24

Cuprinsul tezei

Conținut	ix
Lista de figuri	xi
Lista de tabele	xix
Nomenclatură	xxii
1 Introducere	1
2 Fundamente teoretice si rezultate din literatură	11
2.1 Metode tradiționale de procesare a imaginilor în viziunea computerizată	11
2.2 Tehnici de învățare profundă și rețele neuronale convoluționale	18
2.3 Optimizarea construcției arhitecturii prin căutarea arhitecturii neuronale	24
2.4 Optimizarea mecanismelor de învățare și de acces la ansamblu de învățare	28
2.4.1 Ansamblu de învățare	28
2.4.2 Mecanismul de gating în învățarea prin ansamblu	31
2.5 Comprimarea modelului prin distilare de cunoștințe bazată pe atenție	34
2.5.1 Online KD	35
2.5.2 KD bazat pe atenție	37
2.6 Abordări comune în reconstrucția facială 3D	38
2.6.1 Calibrarea camerei și sisteme de vedere stereo	39
2.6.2 Învățare profundă și modele statistice pentru reconstrucția feței 3D	42
3 Extracția și analiza caracteristicilor faciale	47
3.1 Metode de procesare a imaginilor: noi tehnici pentru segmentarea ochilor	47
3.1.1 Noua tehnică de detectare a irisului folosind pragul dinamic	50
3.1.2 Rezultatele experimentale de detectare a irisului	52
3.2 Abordări de învățare profundă: analiza feței și a părului	55
3.2.1 Noi implementări ale segmentării părului folosind FCN	55
3.2.2 Extensie ANN pentru detectarea cheliei	57
3.2.3 Noi abordări ML pentru recunoașterea culorii părului	57
3.2.4 Noi baze de date și studiu comparativ pentru analiza părului	59
3.2.5 Concluzii și lucrări viitoare	69
4 Dezvoltarea automată și dinamică a rețelelor de învățare profundă	71
4.1 Căutare arhitectură neuronală pentru o rețea specializată optimă	71
4.1.1 Noi abordări pentru generarea de celule utilizând rețeaua neuronală recurentă	71

4.1.2	Configurare experimentală și rezultate numerice	77
4.2	Inferență dinamică în ansambluri folosind mecanismul de gating	83
4.2.1	Metodă nouă de ansamblu dinamic: DynK-Hydra	83
4.2.2	Rezultatele comparative ale DynK-Hydra și discuții	88
4.3	Distilarea cunoștințelor prin învățarea bazată pe atenție	96
4.3.1	Nouă rețea de supraveghere utilizată în distilare online	97
4.3.2	Rezultatele experimentale de distilare a cunoștințelor	102
4.4	Concluzii și lucrări viitoare	110
5	Scena 3D și reconstrucția feței	113
5.1	Nouă metodă de calibrare a camerei folosind o abordare evolutivă	113
5.1.1	Calibrare - Formulare teoretică	114
5.1.2	Descrierea soluției și dimensiunea spațiului de căutare	114
5.1.3	Rezultatele experimentale pentru calibrare	118
5.2	Modelarea feței 3D	122
5.2.1	Reconstruiți fețe 3D utilizând modelul generic de înaltă densitate	122
5.2.2	Validarea și rezultatele reconstrucției feței 3D	126
5.2.3	Concluzii și lucrări viitoare	132
6	Concluzii și lucrări viitoare	135
6.1	Abordarea unificată a analizei feței în spațiul 2D	135
6.2	Dincolo de analiza 2D folosind reconstrucția feței 3D	136
6.3	Îmbunătățiri în modelele de învățare profundă	137
	Referințe	141

Lista publicațiilor

Clasamentul publicațiilor a fost realizat conform (Consiliul Național de Atestare a Titlurilor, Diplomelor și Certificatelor Universitare) aplicabil studenților doctoranzi înscriși după 1 octombrie 2018. Toate clasamentele sunt listate în funcție de clasificarea revistelor¹ și conferințe² în Informatică.

Publicații în Web of Science -Index de citare științifică extins

- [IDBM22] T.A. Ileni, A.S. Darabant, D.L. Borza, A.I. Marinescu. *DynK-hydra: improved dynamic architecture ensembling for efficient inference*. Complex and Intelligent Systems, pp.1-12, 2022
Rang Q1, 2022 IF=35, 4 puncte.
- [BDIM22] D.L. Borza, A.S. Darabant, T.A. Ileni, A.I. Marinescu. *Effective Online Knowledge Distillation via Attention-Based Model Ensembling*. Mathematics, 10(22), p.4285, 2022
Rang Q1, 2022 IF=2, 4 puncte.
- [NID20] S.C. Nistor, T.A. Ileni, A.S. Dărăbant. *Automatic development of deep learning architectures for image segmentation*. Sustainability, 12(22), p.9707, 2020
Rang Q2, 2020 IF=53, 2 puncte.

Publicații în Web of Science, Index de citare a lucrărilor conferinței

- [MID19] A.I. Marinescu, T.A. Ileni, A.S. Darabant. *A versatile 3d face reconstruction from multiple images for face shape classification*. In 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), pp. 1-6, IEEE, 2019.
Conferință-B, 4 puncte.
- [IBD19] T.A. Ileni, D.L. Borza, A.S. Darabant. *Fast In-the-Wild Hair Segmentation and Color Classification*. In VISIGRAPP (4: VISAPP), pp. 59-66, 2019.
Conferință-B, 4 puncte.
- [BID18] D. Borza, T.A. Ileni, A. Darabant. *A deep learning approach to hair segmentation and color extraction from facial images*. In Advanced Concepts for Intelligent Vision Systems: 19th

¹<https://uefiscdi.ro/premierea-rezultatelor-cercetarii-articole>

²<http://portal.core.edu.au/conf-ranks/>

International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19, pp. 438-449, Springer International Publishing.

Conferință-B, 4 puncte.

- [MDI20] A.I. Marinescu, A.S. Darabant, T.A. Ileni. *A Fast and Robust, Forehead-Augmented 3D Face Reconstruction from Multiple Images using Geometrical Methods*. In 2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), pp. 1-6, IEEE, 2020.

Conferință-C, 2 puncte.

- [MDI21] A.I. Marinescu, A.S. Darabant, T.A. Ileni. *Optimal Stereo Camera Calibration via Genetic Algorithms*. In IJCAI 2021 - International Joint Conference on Artificial Intelligence, Workshop, 2021. ISBN:978-0-9992411-9-6

workshop-A, 1 puncte.

- [TA20] T.A. Ileni. *Efficient iris segmentation and pupil detection for visagisme applications*. In 2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 000123-000128, 2020.

Conferință-C, 1 puncte.

Scorul publicațiilor: 26 puncte.

Capitolul 1

Introducere

1.1 Motivație și ipoteze

Această Teză urmărește două direcții de probleme. Pe de o parte, abordăm sarcina analizei feței, atât în spațiul 2D cât și 3D. Pe de altă parte, abordăm sarcina optimizării inferenței prin tăierea dinamică a ansamblului, distilarea cunoștințelor și căutarea arhitecturii neuronale. Analiza feței umane are aplicații în domenii precum biometria (detectarea feței și re-identificarea), medicina (diagnosticarea pe baza evoluției culorii ochilor), industria modei (asortarea accesoriilor) și interacțiunea om-mașină. Abordările moderne pentru analiza feței și, în general, Viziunea Computerizată, folosesc modele de învățare profundă care necesită seturi mari de date de antrenament și costuri computaționale ridicate. Prin urmare, crearea de noi seturi de date și reducerea costurilor de inferență sunt de mare interes. Pe scurt, am dorit să validăm următoarele ipoteze din perspective teoretice și aplicative:

- Este posibil să antrenăm un model de analiză a feței robust și rapid (segmentarea părului/feței, detectarea pupilei, clasificarea culorii părului) și cât de multe date avem nevoie pentru a atinge o acuratețe adecvată?
- Pornind de la un model generic de față 3D, putem să-l deformăm pentru a crea fața unui subiect specific?
- Cum putem modela un algoritm genetic pentru a rezolva problema selecției imaginilor pentru calibrarea camerei?
- Prin ce mijloace putem genera automat arhitecturi de rețele neuronale adaptate pentru sarcini specifice?
- În ce măsură putem reduce costul de inferență într-o rețea mare folosind tăierea dinamică?
- Cât de mult poate învăța un model de dimensiuni mici prin distilarea cunoștințelor ?

1.2 Obiective

Pornind de la motivații și ipoteze teoretice și aplicative, stabilim următoarele obiective pentru această Teză.

- Integrarea procesării imaginilor și învățării profunde pentru sarcinile de analiză a feței. Combinarea ambele abordări de procesare a imaginilor și învățare profundă pentru analiza feței.

Compară beneficiile și dezavantajele metodelor de procesare a imaginilor cu procesele de învățare profundă în contextul analizei feței.

- Dezvoltarea de modele robuste și eficiente pentru analiza feței. Antrenarea unui model rapid și robust de analiză a feței pentru sarcini precum segmentarea părului/feței, detectarea pupilei și clasificarea culorii părului și investigarea impactului dimensiunii setului de date asupra acurateții modelelor de analiză a feței.
- Integrarea modelelor ML în diverse unelte și aplicații. Aplicarea de modele de analiză a feței în domenii specifice precum biometria, medicina și industria modei. Explorarea procesului de deformarea a unui model generic de față 3D pentru a crea o față specifică a unui individ.
- Reducerea costurilor de inferență în modelele de învățare profundă. Explorarea de tehnici pentru reducerea costurilor de inferență în rețele mari folosind tăierea dinamică. Investigarea în ce măsură modelele de dimensiuni mici pot învăța de la rețele mai mari prin distilarea cunoștințelor.
- Optimizarea costurilor de căutare și computaționale. Implementarea un controler de rețea recurentă pentru a genera automat blocuri de arhitectură de învățare profundă. Modelarea unui algoritm genetic pentru a îmbunătăți procesul de calibrare a camerei.

1.3 Contribuții originale

Mai specific, în domeniul analizei feței, am dezvoltat un instrument de analiză a părului care include segmentarea părului și clasificarea culorii, un detector de pupile și iris și un flux de lucru pentru reconstrucția feței 3D. Pentru a optimiza costurile computaționale studiem și dezvoltăm un ansamblu de rețele tăiate dinamic bazate pe un *gater*, un mecanism de distilare online pentru a crea *studenti* performanți și preciși și o rețea recurentă care prezice arhitectura blocului potrivit pentru o sarcină specifică.

O mare parte din cercetare s-a concentrat pe analiza feței folosind imagini 2D (planare) [Kel70, VJ01, KKA⁺20]. Indiferent de metoda utilizată, unele probleme sunt prezente în contextul capturilor 2D *in-the-wild* cum ar fi lipsa datelor de adâncime și condițiile de iluminare precare.

Sarcinile de analiză a feței sunt o subsecțiune a Viziunii Computerizate (CV), în care algoritmi de clasificare sunt adaptați pentru fețele umane, pentru extragerea sau clasificarea mai multor atribute precum culoarea părului, forma feței, culoarea ochilor, ID-ul feței, poziția capului, identitatea etc.

Pentru lucrarea de analiză a feței, atât în spațiul 2D cât și 3D, contribuțiile originale ale Tezei sunt:

- Anotarea manuală și disponibilizarea publică a 3.5k de imagini care conțin masca feței și a părului, 20k de anotări ale culorii părului și 300 de adnotări ale conturului ochiului [BID18, IBD19, TA20].
- Evaluarea rețelelor neuronale profunde de segmentare a feței/părului, analizarea rezultatelor folosind mai mulți clasificatori și compararea performanțelor cu modelele de ultimă oră [TA20].
- Dezvoltarea și ajustarea unui algoritm de detectare a irisului și a pupilei, care se ocupă de mai multe tipuri de zgomot, folosind atât tehnici de Procesare a Imaginilor cât și de învățare automată, și compararea rezultatelor cu metode similare din literatură.

- Antrenarea unui clasificator de culoare a părului, folosind doar zona părului, pe un set de date echilibrat cu următoarea taxonomie: roșu, maro, gri, negru și blond. Analiza impactului spațiului de culoare utilizat și compararea rezultatelor cu starea actuală a cercetării, obținând rezultate comparabile sau mai bune [BID18, IBD19].
- Dezvoltarea unui model de flux de lucru pentru reconstrucția feței 3D care reconstruiește întreaga zonă a feței, nu limitată la frunte, ca majoritatea modelelor actuale [MDI20, MID19].

În continuare, explicăm mai detaliat contribuțiile. Pentru sarcina de analiză a feței, mai întâi antrenăm o Rețea Neuronală Profundă de segmentare a capului, care inferă zona feței și a părului. Deoarece în comunitatea de cercetare lipsește un set de date anotate de dimensiuni suficient de mare, noi am reușit să anotăm manual mai mult de 3.5k de imagini. Pentru sarcina de segmentare, am obținut o acuratețe medie a pixelilor de 92.1% pentru imaginile feței frontale, care este mai mare decât lucrările similare (în momentul publicării 91.5% pentru setul de date Figaro1k [MSLB18]).

Folosind zona părului antrenăm un alt predictor, de data aceasta o Rețea Neuronală Artificială (ANN) cu trei straturi pentru a clasifica culoarea părului subiectului. Cele mai bune rezultate au fost obținute când am antrenat o rețea neuronală cu histogramele de culoare calculate pe spațiul de culoare LAB. Pentru fiecare dintre cele trei canale de intrare, calculăm o histogramă de dimensiune $256/8=32$, care este concatenată și normalizată. Taxonomia părului include clasele negru, maro, blond, gri și roșu. Pentru a mări setul de date disponibil de culoare a părului, anotăm mai mult de 20k de imagini. Pentru clasificarea culorii părului, acuratețea noastră (de 99.1%) este mult superioară față de abordări similare care au 88.6% la momentul publicării. Ca inconvenient, majoritatea lucrărilor similare folosesc seturi de date neechilibrate.

În timp ce analizăm fața, suntem, de asemenea, interesați de zona ochilor. Pentru aceasta, dezvoltăm și ajustăm un algoritm de detectare pentru iris și pupilă. Fluxul de lucru începe cu un regresor (antrenat cu 300 de mostre anotate manual) care prezice colțurile ochilor, pupila, pleoapele și un punct pe marginea irisului. Folosind aceste puncte, ajustăm un algoritm de Procesare a Imaginilor care identifică pupila ca fiind cea mai întunecată zonă a ochiului și irisul folosind o imagine binară bazată pe prag.

Pentru a atenua problemele menționate anterior în spațiul 2D, cum ar fi lipsa datelor de adâncime și condițiile de iluminare, cercetătorii și aplicațiile industriale adoptă modele de față 3D în analiza lor. În contextul mai larg al capturării 3D, următoarele tehnici sunt utilizate: sisteme de viziune stereo (care necesită camere calibrate sau necalibrate) [ARL⁺10], camere RGB-D (care recuperează, de asemenea, informații despre adâncime), scanări laser 3D [LTW95], metode 3D-din-2D (de obicei în scenarii de camere necalibrate). În lucrarea noastră, pentru reconstrucția feței 3D, propunem un algoritm de potrivire a modelului statistic bazat pe Structura din Mișcare și un model deformabil.

Fluxul de lucru de reconstrucție începe cu detectarea conturului feței folosind un regresor de aliniere a feței, la care adăugăm încă cinci puncte din zona superioară a frunții (extrasă folosind rețeaua noastră de segmentare a feței). Majoritatea cercetătorilor limitează reconstrucția feței 3D până la linia sprâncenelor sau mijlocul frunții. Folosind mai multe achiziții ale subiectului construim un nor de puncte 3D al reperelor feței (folosind Structura din Mișcare) și apoi deformăm un model de bază de față 3D pentru a obține rezultatul final. Comparativ cu alte lucrări obținem rezultate similare, având avantajul unei reconstrucții complete a feței 3D.

În continuare, în domeniul modelării 3D, dezvoltăm un algoritm genetic paralel care are ca scop îmbunătățirea procesului de calibrare a camerei prin selectarea celui mai potrivit subset de achiziții ale unui model (de obicei o tablă de șah) necesar ca intrare pentru algoritmul de calibrare.

În această încercare, modelăm cromozomul ca un tablou binar având o lungime egală cu numărul de achiziții. Dacă un element este setat pe *True*, achiziția corespunzătoare este utilizată ca intrare în

algoritmul de calibrare. În timpul evoluției populației, folosim operatori de mutație și încrucișare.

Pentru a doua direcție (metode de optimizare în învățarea profundă) a cercetării noastre abordăm următoarele probleme: căutarea arhitecturii neuronale, distilarea cunoștințelor și tăierea dinamică. Listăm contribuțiile originale în domeniu:

- Proiectarea și antrenarea unei arhitecturi de învățare profundă bazată pe *dynamic pruning* care reduce costurile de inferență cu mai mult de 2x comparativ cu lucrările similare și cu mai mult de 5x comparativ cu rețelele profunde clasice [IDBM22].
- În domeniul Căutării Arhitecturii Neuronale implementăm un controler de rețea recurentă care prezice blocuri de arhitectură pentru sarcini specifice de Viziune Computerizată. Experimentele au fost efectuate pentru a studia impactul consumului de energie în procesul de căutare a unei arhitecturi potrivite. [NID20].
- Îmbunătățirea performanțelor rețelelor ușoare prin procesul de Distilare a Cunoștințelor Online, folosind un cohort de studenți antrenați în paralel. Obținem un câștig în acuratețe de mai mult de 2% menținând același număr de parametri [BDIM22].

Propunem un controler de căutare a arhitecturii neuronale (NAS) pentru a crea automat blocuri de arhitectură de învățare profundă, pentru a atenua problema alegerii manuale a unui model potrivit. Comunitatea de cercetare a propus mai multe moduri de a aborda aceste probleme, inclusiv combinarea mai multor învățători slabi (boosting [DCJ+94], bagging [Bre96], algoritmul de stivuire), antrenarea mai multor modele pe același set de date și combinarea/selectarea predicțiilor lor pentru a obține un rezultat mai bun [HLP+17, BWHY05, FHL19] (adică Învățare în Ansamblu), antrenarea unui model mai mare și tăierea lui pentru inferență specializată [CGW+19], inferență dinamică folosind o parte a rețelei [MMSF18], crearea dinamică a unei arhitecturi specializate pe sarcină (adică Căutarea Arhitecturii Neuronale) [EMH18, BGNR16, ZL16], distilarea cunoștințelor pentru a crea un model de dimensiuni mici [HVD15, LZG18, GYMT20].

Am găsit abordarea Căutării Arhitecturii Neuronale (NAS) foarte promițătoare. În acest sens, am dezvoltat un predictor recurent pentru a genera celule (grup de operații interconectate) care sunt apoi stivuite în șabloane mai mari de rețele. Testăm fluxul de lucru de descoperire a rețelei pentru o sarcină de segmentare a instanței și obținem un scor IoU ridicat.

Optimizarea costului computațional (în special pentru inferență) este un subiect de cercetare de mare interes în prezent. Cu compromisul de a necesita mai multe resurse în timpul antrenării (timp și memorie), sarcini precum Distilarea Cunoștințelor, rețelele neuronale dinamice sau un Amestec de Experti încearcă să atingă acest obiectiv. Abordarea noastră este o distilare bazată pe atenție online, unde un cohort de studenți sunt antrenați împreună și ieșirea lor este combinată dinamic pentru a crea un predictor mai puternic. Această ieșire este ulterior distilată componentelor individuale.

Rezultatele (exprimate în câștig de acuratețe) depășesc sau sunt cel puțin comparabile cu abordările actuale.

Pe de altă parte, rețelele dinamice (sau tăierea dinamică) [HHS+21] sunt o îmbunătățire a rețelelor neuronale artificiale clasice, unde doar o parte a rețelei se execută, bazată pe anumite condiții de oprire/score, sau doar o parte a intrării este considerată. În această lucrare, creăm un ansamblu, compus din mai multe ramuri (rețea neuronală convoluțională), având un trunchi comun și un gater, care decide dinamic ce ramuri să activeze pentru fiecare probă de intrare. În acest mod, numărul de operații (flops) de inferență este redus considerabil.

În acest proces, împărțim spațiul de intrare în mai multe partiții și atribuim un model de ramură (din ansamblu) fiecărei partiții. Această *atribuire* se face printr-o funcție de eroare care încurajează fiecare model de ramură să fie specializat pe spațiul de intrare specific.

Capitolul 2

Contextul general și Lucrări similare

În acest capitol, oferim contextul general și fundalul noțiunilor pe care le-am abordat în Teză. Discuția noastră este despre metodele tradiționale și de învățare în adâncime utilizate în viziunea computerizată, în special în sarcina analizei feței și optimizarea pentru învățare profundă, cum ar fi căutarea de arhitectură neuronală, distilare a cunoștințelor și tăiere dinamică.

2.1 Metode tradiționale de procesare a imaginilor și tehnici de învățare profundă în viziunea computerizată

Algoritmii tradiționali de procesare a imaginilor în analiza feței sunt dezvoltati pentru diverse sarcini precum detectarea feței, recunoașterea, segmentarea părului și detectarea irisului. Aceste metode implică utilizarea unor algoritmi specifici pentru fiecare sarcină. De exemplu, în recunoașterea facială [Kel70], sunt folosite metode precum eliminarea fundalului, potrivirea șablonului, netezirea dinamică bazată pe praguri și detectarea marginilor. Clasificatorii în cascadă bazați pe caracteristici Haar [VJ04] sunt un exemplu notabil, care combină tehnicile de procesare a imaginilor și învățarea automată. Această metodă construiește imagini integrale, în care valoarea fiecărui pixel este suma tuturor pixelilor de deasupra și din stânga acestuia, pentru a calcula rapid caracteristici precum caracteristicile Haar. Selectarea caracteristicilor se realizează de obicei prin algoritmi de învățare automată precum AdaBoost [Sch13], care selectează funcții relevante dintr-un set mai mare pentru a crea cursanți slabi care lucrează colectiv pentru sarcini precum detectarea feței.

Abordările de învățare profundă abordează unele limitări ale metodelor tradiționale de procesare a imaginilor cum ar fi nevoia de predefinire și reglare a algoritmilor pentru sarcini specifice și constrângeri legate de locația imaginii și cerințele hardware externe. Metodele de învățare profundă implică de obicei selectarea unui model matematic, de obicei o rețea neuronală convoluțională (CNN), colectarea datelor etichetate și rularea algoritmilor de optimizare ghidați de o funcție de pierdere. Aceste metode încorporează automat constrângeri precum distribuția uniformă a texturii sau nevoia de dispozitive specializate. De exemplu, metodele de învățare profundă pentru segmentarea părului [PN17, LCP⁺18] și clasificarea coafurii în medii neconstrânse utilizează CNN-urile pentru a construi hărți de probabilitate a părului din pete de imagini, clasificate după clasificatori precum Random Forest. Alte aplicații de învățare profundă în segmentarea capului, clasificarea culorii părului și segmentarea irisului [WZL⁺19] valorifică capacitatea CNN-urilor de a învăța caracteristici din date fără programare explicită pentru fiecare sarcină. Această abordare oferă mai multă flexibilitate și adaptabilitate în medii diverse și dinamice.

2.2 Abordări comune în reconstrucția 3D de față și calibrarea camerei

În reconstrucția feței 3D sunt folosite mai multe tehnici, inclusiv sisteme de vedere stereo [ARL⁺10], camere RGB-D, scanări laser 3D [LTW95] și metode 3D-din-2D. Aceste metode pot fi clasificate în continuare în potrivirea modelelor statistice, sisteme fotometrice și abordări de învățare profundă [MPS21].

Cea mai comună tehnică în ajustarea modelelor statistice este utilizarea modelelor 3D Morphable (3DMM). În această abordare, un model de față generic, compus din geometrie (vârfurile), albedo (proprietatea de reflexie) și textura (culoare), este modificat pe baza caracteristicilor faciale ale subiectului pentru a reconstrui o nouă față 3D care se potrivește cel mai bine fotografiilor date. Această metodă necesită un model de față 3D predefinit, care este adaptat pentru a se potrivi trăsăturilor faciale individuale.

Modelele fotometrice presupun un model de reflexie lambertian, în care luminozitatea unei suprafețe rămâne constantă indiferent de punctul de vedere al observatorului. Aceste modele deconstruiesc o imagine sau un set de imagini în normale, albedo și culoare.

Modelele de învățare profundă pentru reconstrucția 3D din 2D folosesc o singură fotografie facială pentru a reconstrui fața. Aceste modele folosesc un model volumetric de învățare antrenat pentru a deduce caracteristicile de profunzime ale subiectului. Deși sunt eficiente și directe, provocările apar atunci când părți ale feței nu sunt prezente în imagine din cauza ocluziilor, unghiurilor nefrontale ale feței, iluminării slabe etc. Modelul trebuie apoi să ”inventeze” informațiile lipsă cum ar fi geometria și textura.

Calibrarea camerei este un proces fundamental în viziunea computerizată, mai ales în contextul reconstrucției 3D. Implică determinarea parametrilor interni (intrinseci) și externi (extrinseci) ai camerei pentru a surprinde cu precizie geometria mediului.

Printre primele metode de calibrare a camerei [Bro71, Fai75] a folosit un obiect având mai multe plane ortogonale 3D, care sunt reprezentate cu acuratețe într-un spațiu 3D (digital). Pentru a elimina necesitatea dispozitivelor de calibrare scumpe, au apărut tehnici de autocalibrare. Acest cadru necesită o cameră în mișcare în jurul unei scene statice. Fiecare cadru capturat oferă două constrângeri asupra parametrilor intrinseci ai camerei. Cel puțin trei imagini sunt suficiente pentru a recupera atât parametrii intrinseci, cât și cei extrinseci [Har94, LF97].

2.3 Prezentare generală a căutării arhitecturii neuronale automate

Neural Architecture Search (NAS) (în lb. română căutarea arhitecturii neuronale automate) este esențială în evoluția algoritmilor de învățare automată, mai ales pe măsură ce datele devin din ce în ce mai variate și abundente. Algoritmii tradiționali de învățare automată [KSH12, SZ14] sunt de obicei dezvoltați și ajustați de oameni pentru sarcini specifice. Cu toate acestea, pe măsură ce nevoia de a aplica acești algoritmi în diferite domenii crește, există o nevoie presantă de un cadru automatizat capabil să creeze și să construiască acești algoritmi. Aici NAS joacă un rol crucial.

NAS implică trei elemente fundamentale: spațiul de căutare, strategia de căutare și strategia de estimare a performanței. Spațiul de căutare include toate arhitecturile posibile pe care metoda NAS le ia în considerare. De exemplu, arhitecturile convenționale de rețea neuronală convoluțională (CNN) sunt construite prin stratificarea straturilor convoluționale și gruparea pentru a obține dimensiunile dorite, echilibrând latența și acuratețea. Cu toate acestea, abordările moderne NAS se concentrează pe crearea de module sau celule mai mici [LCY13, SLJ⁺15] care servesc drept blocuri pentru întreaga arhitectură CNN, adaptate sarcinilor specifice.

Spațiul de căutare NAS poate deveni extins, cuprinzând diferite operațiuni sau straturi și conexiunile acestora. Parcurgerea eficientă a acestui spațiu este crucială. Căutările aleatorii pot consuma mult timp și pot să nu ia în considerare informațiile obținute anterior. Pentru a face față acestor provocări, au fost propuși mai mulți algoritmi NAS, inclusiv algoritmi evolutivi, învățare prin consolidare, metode complet diferențiabile și metode concepute pentru inferență rapidă, unele luând în considerare latența dispozitivului real.

2.4 Optimizarea mecanismelor de învățare a ansamblului

Învățarea prin ansamblu și-a dovedit superioritatea față de predictorii unici, atât în aspecte teoretice, cât și empirice, pentru diferite sarcini, inclusiv regresie și clasificare. Conceptul de diversitate a ansamblurilor este crucial, cu ansambluri clasificate în patru niveluri [SS97] în funcție de numărul de erori. Aceste niveluri variază de la erori fără coincidență (Nivelul 1), unde votul majoritar este întotdeauna corect, până la situații în care niciun membru nu este corect (Nivelul 4). Diversitatea în cadrul unui ansamblu este esențială, iar informațiile sugerează că membrii unui ansamblu de Nivel 2 sau Nivel 3 pot fi ajustați pentru a îmbunătăți acuratețea clasificării generale.

2.5 Comprimarea modelului prin distilare de cunoștințe bazată pe atenție

Distilarea cunoștințelor (KD) implică un model mai mic (elevul) care învață de la un model mai mare (profesorul) [BCNM06]. Elevul este instruit și penalizat în funcție de modul în care rezultatele sale diferă de cele ale profesorului. Un progres cheie în KD a fost introducerea unui proces de ”înmuiere” în rezultatul profesorului, folosind un factor de scalare a temperaturii pe activarea softmax [HVD15]. Această abordare permite elevului să învețe relația dintre clasa actuală și cele similare. Scopul este de a distila ”cunoștințele dark” de la profesor la elev, asigurându-se că modelul elevului învață relații complicate și nuanțe capturate de modelul mai mare.

Distilarea online a cunoștințelor diferă de distilarea offline prin faptul că nu există un model de profesor pre-format [ZXHL18, LZG18]. În schimb, o cohortă de studenți antrenează și împărtășește simultan cunoștințele, fiecare acționând atât ca elev, cât și ca profesor pentru ceilalți. Această metodă implică mai mulți cursanți (studenți) împărtășirea și distilarea cunoștințelor, ceea ce adaugă complexitate, dar poate duce la rezultate de învățare mai solide.

Mecanismele de atenție în distilarea cunoștințelor sunt inspirate de capacitatea sistemului senzorial uman de a se concentra asupra elementelor esențiale. În contextul KD, mecanismele de atenție sunt folosite pentru a redistribui ponderile unei hărți de caracteristici. Această abordare pune accent pe cele mai semnificative caracteristici și canale, îndreptând învățarea modelului elevului către aceste aspecte cruciale. Este deosebit de util în sarcinile de viziune computerizată, unde concentrarea pe caracteristici sau zone specifice poate îmbunătăți semnificativ performanța.

Capitolul 3

Extracția și analiza caracteristicilor faciale

3.1 Descrierea metodei de analiză a feței

Acest capitol este structurat în două secțiuni principale: Metode de procesare a imaginilor și abordări de învățare profundă.

Această secțiune detaliază tehnicile pentru segmentarea ochilor, inclusiv detectarea pupilei, recunoașterea pleoapelor și a colțului ochilor și segmentarea irisului. Acesta abordează provocări precum acoperirea pleoapelor, culoarea neuniformă a irisului, dilatarea pupilei și zgomotul datorat genelor și reflexelor luminii.

Abordări de învățare profundă: Această parte acoperă metode avansate de segmentare a părului și clasificare a culorilor folosind învățarea automată. Introduce rețele neuronale complet convoluționale (FCN) inspirate de arhitecturile U-Net și VGG pentru segmentarea părului și metode de clasificare a culorii părului folosind atât rețele neuronale artificiale (ANN) cât și clasificatoare aleatorii forestiere. Această secțiune discută și despre detectarea cheliei ca o extensie a segmentării părului.

În prima secțiune, detaliem o abordare inovatoare pentru segmentarea ochilor, concentrându-ne în special pe detectarea și segmentarea pupilei, irisului, pleoapelor și colțurilor ochilor. Tehnicile folosesc o combinație de metode tradiționale de procesare a imaginilor și algoritmi de învățare automată. Pașii cheie includ detectarea inițială a feței și a regiunii ochilor, detectarea reperelor pentru trăsăturile ochilor și rafinarea segmentării pupilei și irisului. Capitolul abordează, de asemenea, provocări precum acoperirea variată a pleoapelor, culoarea neuniformă a irisului, dilatarea pupilei și interferența de la gene și reflexiile luminii. Aceste metode sunt semnificative pentru aplicațiile în analiza caracteristicilor faciale, în special în zonele care necesită detectarea precisă a caracteristicilor oculare. Secțiunea subliniază importanța acurateții și robusteții în aceste tehnici datorită gamei lor largi de aplicații, de la identificarea biometrică până la analiza emoțională.

A doua secțiune se concentrează pe două metode pentru segmentarea părului și clasificarea culorii părului în imaginile faciale folosind tehnici de învățare automată. Introduce două metode principale de învățare profundă pentru segmentarea părului, explorând rețele neuronale complet convoluționale inspirate de arhitecturile U-Net și VGG. În plus, secțiunea acoperă clasificarea culorii părului, prezentând abordări care implică utilizarea histogramelor de culoare a părului cu rețele neuronale artificiale și clasificarea super-pixeli cu Random Forest Classifiers. Prin extinderea arhitecturii de segmentare a părului cu o ramură complet conectată, efectuăm și clasificarea cheliei. Această secțiune este crucială pentru înțelegerea aplicațiilor avansate de învățare profundă în analiza trăsăturilor faciale, vizând în special atributele părului.

3.2 Rezultatele numerice ale analizei feței

În această secțiune, oferim un rezumat al rezultatelor obținute la sarcinile de analiză facială.

Metoda utilizată pentru detectarea irisului și a pupilei a implicat adnotarea manuală a 200 de imagini din setul de date Chicago. Rezultatele evaluării au fost raportate în termeni de eroare pătrată medie și eroare absolută medie pentru fiecare reper, combinând puncte atât pentru ochiul stâng, cât și pentru ochiul drept. Pentru detectarea conturului ochilor obținem următoarele erori exprimate în termeni de eroare medie absolută Pupila: 1.24 Raza: 2.46 Colțul exterior: 2.46, colțul interior: 3.11, vârful ochiului: 3.65, fundul ochiului: 2.68. Având în vedere acuratețea detecțiilor punctuale, calculată folosind un prag de 0,05 pentru cea mai gravă eroare cumulată, avem o valoare peste 95%. Aceasta înseamnă că eroarea este mai mică decât dimensiunea pupilei.

Pentru segmentarea părului, cele două metode de învățare profundă pe rețele neuronale complet convoluționale (FCN) au fost inspirate de arhitecturile U-Net și VGG. Pentru segmentarea părului folosind FCN, a fost obținută o acuratețe a pixelilor de 84,75% pe un subset al bazei de date Figaro 1k și o acuratețe medie a pixelilor de 90,77% cu un interval de încredere (C.I.) de [86,43, 95,11], care este la cel puțin la egalitate cu metoda similară din literatura de specialitate [MSLB18]. Pentru a instrui și a evalua modelele de segmentare, adnotăm manual peste 3.5k imagini și le facem disponibile publicului.

A fost propus un model probabilistic pentru detectarea cheliei și ajustat pe peste 6500 de imagini din setul de date CelebA. Precizia generală a modelului pentru sarcina de detectare a cheliei a fost raportată la 93,33%, cu alte valori precum precizia, reamintirea și scorul f1, toate fiind de 93,50. Intervalul de încredere la un nivel de semnificație de 95% este [91,3, 95,3].

Performanța modulului de clasificare a culorii părului bazat pe ANN pentru diferite spații de culoare și dimensiuni ale vectorului de caracteristici a fost raportată, cele mai bune rezultate obținute folosind spațiul de culoare LAB cu 8 dimensiuni de bin. De exemplu, procentele de acuratețe pentru diferite spații de culoare cu dimensiunea bin 111 au fost: RGB - 87,80%, HSV - 88,10% și LAB - 88,30%. Într-un scenariu fără păr negru, metoda a atins o acuratețe de recunoaștere a culorii părului de 89,6% cu un interval de încredere de [87,16%, 92,04%], depășind alte lucrări din literatură [KPR⁺14]. Pentru această sarcină, adnotăm manual imagini de 20 k pentru antrenament și imagine de 2 k pentru testare.

Capitolul 4

Dezvoltarea Automată și Dinamică a Rețelelor de Învățare Profundă

În acest capitol, prezentăm cercetarea noastră fundamentală privind optimizarea învățării profunde. Metodele propuse atacă diferite etape ale lucrului cu modelele DL, de la căutarea arhitecturii până la antrenament și inferență. Rezultatele numerice sunt promițătoare, depășind abordările actuale din literatură.

4.1 Căutarea Arhitecturii Neuronale pentru Rețea Specializată Optimă

NAS (Neural Architecture Search) este un proces automatizat pentru proiectarea arhitecturilor de rețea neuronală. Obiectivul principal al NAS în acest context este de a identifica cea mai eficientă și eficace arhitectură pentru o sarcină dată. Această abordare este deosebit de relevantă pentru sarcini care necesită configurații de rețea specializate. Scopul este de a optimiza arhitectura pentru a îmbunătăți metrici de performanță precum acuratețea, luând în considerare și eficiența computațională.

Această secțiune introduce o nouă metodă pentru căutarea și optimizarea automată a arhitecturilor CNN pentru sarcini specifice, cum ar fi segmentarea ochelarilor. Abordarea folosește o Rețea Neuronale Recurentă (RNN) bazată pe învățare prin întărire pentru a genera configurații eficiente de celule de CNN-uri. Aceste celule formează blocurile de construcție ale arhitecturii finale CNN. Structura acestor celule este reprezentată ca un graf aciclic dirijat, optimizat pentru performanță cu operații variate precum convoluția și convoluția dilatată.

Procesul implică antrenarea unei rețele neuronale recurente (RNR) cu învățare prin întărire pentru a maximiza o funcție de recompensă, care este în general legată de performanța arhitecturilor generate pe un set de date de validare. În timp, RNR-ul învață să propună arhitecturi din ce în ce mai eficiente. După procesul de căutare, arhitectura cu cea mai bună performanță este antrenată de la zero pentru a evalua performanța sa pe sarcina țintă.

Principalul avantaj al NAS este capacitatea sa de a automatiza proiectarea arhitecturilor de rețea neuronală, care poate fi o sarcină extrem de complexă și consumatoare de timp atunci când este realizată manual. De exemplu, pentru 8 operații posibile și o dimensiune a celulei de 5 noduri, am putea avea 327.680 de combinații posibile pentru a crea o celulă. Automatizând acest proces, NAS poate descoperi potențial arhitecturi noi și extrem de eficiente care s-ar putea să nu fie evidente prin explorare manuală. În plus, NAS poate fi adaptat pentru a optimiza rețele nu doar pentru acuratețe, ci și pentru alte constrângeri, cum ar fi eficiența computațională, făcându-l potrivit pentru implementare în medii cu resurse limitate. Pentru o sarcină specifică de segmentare a ochelarilor, celula noastră generată, instantiată într-un șablon de rețea mai mare, performează la 0.96 IoU.

4.2 Execuția Dinamică în Ansambluri folosind Mecanismul de Gating

Execuția dinamică în ansambluri este o tehnică avansată în optimizarea rețelelor neuronale. Scopul său principal este de a îmbunătăți eficiența computațională și acuratețea modelelor de învățare profundă. Acest lucru se realizează prin activarea selectivă doar a părților relevante dintr-un ansamblu de rețele neuronale. Tehnica este deosebit de utilă în situații în care diferite ramuri ale rețelei sunt specializate pentru diferite sub-sarcini.

Teza propune frameworkul DynK-Hydra, care îmbunătățește rețelele neuronale profunde pentru sarcini specifice prin împărțirea claselor în clustere și specializarea ramurilor rețelei pentru aceste sub-sarcini. Un mecanism de gating alege dinamic cele mai relevante ramuri pentru fiecare intrare, îmbunătățind atât acuratețea cât și viteza. Acest cadru este conceput pentru a alege dinamic cea mai relevantă ramură pentru fiecare intrare în timpul inferenței, reducând astfel calculele inutile care sunt tipice în modelele mari, monolitice. Acest mecanism este testat pe mai multe seturi de date, arătând rezultate promițătoare în ceea ce privește eficiența și acuratețea.

La baza acestei abordări de execuție dinamică se află mecanismul de gating. Acest mecanism funcționează ca un comutator inteligent care determină ce ramuri ale ansamblului ar trebui activate pentru o anumită intrare. Decizia se bazează pe caracteristicile datelor de intrare, asigurând că doar cele mai relevante și specializate segmente ale rețelei sunt utilizate. Acest proces duce la o reducere semnificativă a suprasolicitării computaționale, deoarece doar o fracțiune din întregul ansamblu este activă în orice moment dat.

Rezultate numerice evidențiază câștigurile de eficiență în ceea ce privește viteza și cerințele computaționale reduse, împreună cu îmbunătățiri sau menținerea unei acurateți ridicate în performanța sarcinii. De exemplu, pentru CIFAR-100 [KH⁺09] DynK-Hydra a atins aproximativ 74% acuratețe cu 139M FLOPs, demonstrând o îmbunătățire de 2.7x ori față de HydraRes [MMSF18], care a necesitat 378M FLOPs pentru o acuratețe similară. În comparație, cea mai mare arhitectură ResNet a atins 73.56% acuratețe cu 767M FLOPs, făcând DynK-Hydra aproximativ de 5,5 ori mai eficientă în ceea ce privește timpul de inferență. Pe setul de date TinyImageNet [LY15], a fost obținut un câștig de 1,43% în precizie folosind cadrul propus. Lățimile intervalului de încredere pentru experimentele cu vanilie și KD au fost [51,94, 53,90] și, respectiv, [53,37, 55,33].

4.3 Distilarea Cunoștințelor prin Învățare Bazată pe Atenție

Această secțiune introduce o abordare inovatoare pentru optimizarea rețelelor neuronale profunde. Această abordare își propune să creeze un model de mici dimensiuni care imită comportamentul și performanța unui model mai mare. Utilizează un mecanism de distilare online în care supervisorul este o combinație ponderată de studenți (rețele neuronale), iar cunoștințele (ieșirea supervisorului) sunt distilate înapoi studenților. După antrenament, se folosește un singur student (mai precis), obținându-se astfel o precizie ridicată, menținând în același timp o amprentă de memorie și un timp de execuție scăzut.

Mecanismul online KD tratează mai multe modele ca studenți, învățând unul din predicțiile celuilalt. Atenția ponderează dinamic rezultatul fiecărui elev în ansamblul final. Experimentele au fost efectuate folosind diferite arhitecturi de rețea binecunoscute și pe mai multe seturi de date de clasificare a imaginilor. Pierderea clasică de entropie încrucișată a fost utilizată pentru antrenarea modelelor individuale "vanila" înainte de antrenamentul aceleiași arhitecturi de rețea cu cadrul KD.

Cadrul utilizează Modulul Convolutional Block Attention (CBAM) pentru a calcula hărți de atenție pe canal și dimensiuni spațiale. Acest mecanism se concentrează pe caracteristicile cele mai discriminatorii ale modelului folosind operațiuni de pooling și un perceptron multi-strat partajat.

Mecanismul de atenție a canalului calculează ponderile utilizate în asamblarea predicțiilor elevilor.

Cadrul KD a fost evaluat pe baza benchmark-urilor de clasificare a imaginilor CIFAR-10 [HP18], CIFAR-100, [KH⁺09] și TinyImageNet [LY15]. Versiunea ”vanila” a modelului a fost mai întâi antrenată și evaluată independent, apoi a fost folosită aceeași arhitectură în ansamblu, iar cel mai bun student a fost selectat după antrenament. Îmbunătățirea acurateței (adică câștigul) a fost o măsură critică pentru evaluarea eficienței procesului de distilare a cunoștințelor. O comparație numerică, folosind arhitectura ResNet-32 [HZRS16], cu abordări similare din literatură este prezentată în Tabelul 4.1 pentru CIFAR-10 și în Tabelul 4.2 pentru CIFAR-100.

Table 4.1: Compararea modelului de KD cu modele similare din literatură pe baza de date CIFAR-10 ResNet-32 [HZRS16] folosind arhitectura ResNet-32 [HZRS16] ca și rețea student.

Method	Vanilla	KD	KD Gain
ONE [LZG18]	93.07%	94.01%	0.94%
CLCNN [SC18]	93.17%	94.14%	0.97%
OKDDip [CMW ⁺ 20] net.	93.66%	94.38%	0.72%
OKDDip [CMW ⁺ 20] br.	93.66%	94.42%	0.76%
PCL [WG21]	93.26%	94.33%	1.07%
Proposed	92.77%	93.88%	1.11%

Table 4.2: Compararea modelului de KD cu modele similare din literatură pe baza de date CIFAR-100 ResNet-32 [HZRS16] folosind arhitectura ResNet-32 [HZRS16] ca și rețea student.

Method	Vanilla	KD	KD Gain
DML [ZXHL18]	68.99%	71.19%	2.20%
KDCL [GWW ⁺ 20] ¹	71.28%	73.76%	2.48%
OKDDip net. [CMW ⁺ 20]	71.24%	74.60%	3.36%
OKDDip br. [CMW ⁺ 20]	71.24%	74.37%	3.13%
SAD [JHP21]	75.32%	77.47%	2.15%
PCL [WG21]	71.28%	74.14%	2.86%
Proposed	69.6%	72.76%	3.16%

Capitolul 5

Reconstrucția Scenei 3D și a Feței

Acest Capitol discută beneficiile obținute folosind reconstrucția facială 3D, concentrându-se pe limitările impuse de analiza imaginilor 2D, în special în imagini cu fețe nefrontale și în lipsa informații de adâncime. Capitolul introduce o optimizare procesului de calibrare a camerei folosind un algoritm genetic, un model de detectare a punctelor de față înculzând marginea superioară a frunții pentru o geometrie completă a feței și două modele de reconstrucție 3D Morphable Model (3DMM) utilizând 73 de repere faciale.

5.1 Calibrarea Camerei folosind o Abordare Evolutivă

Propunem o abordare inovatoare folosind un algoritm genetic pentru optimizarea calibrării camerei. Calibrarea camerei este procesul de găsire a parametrilor camerei intrinseci și extrinseci. Parametrii intrinseci se referă la caracteristicile interne ale camerei, cum ar fi lungimea focală, punctul principal și raportul de aspect, în timp ce parametrii extrinseci descriu poziția și orientarea camerei în spațiu. În această lucrare nu am abordat direct algoritmul de calibrare a camerei, ci am folosit implementarea bine stabilită din librăria OpenCV [?]. Metoda noastră implică selectarea celor mai bune capturi dintr-un set de imagini pentru a minimiza erorile de calibrare.

Procesul de calibrare folosește tehnici evolutive pentru a gestiona eficient un spațiu mare de soluții, folosind operatori de mutație și încrucișare. Metodologia este testată prin diverse experimente, demonstrând superioritatea sa în acuratețe și robustețe comparativ cu metodele tradiționale, în special în calibrarea stereo pentru măsurători 3D.

Secțiunea de rezultate numerice și comparații detaliază rezultatele experimentale ale procesului de calibrare a camerei. Include rezultate de calibrare stereo și mono, comparând abordarea propusă bazată pe algoritm genetic cu metodele stabilite. Experimentele demonstrează o acuratețe sub-milimetrică în măsurătorile 3D pentru perechi de puncte stereo. Pentru calibrarea mono, metoda propusă depășește tehnicile existente, cum ar fi *CalWiz* [PS19], cu o marjă semnificativă în ceea ce privește eroarea medie pătratică. În cel mai bun experiment al nostru, obținem o eroare de 0.357px comparativ cu cel mai bun experiment *CalWiz* de 0.631px.

5.2 Modelarea 3D a Feței Folosind Modele de Bază Deformabile

Acest Capitol introduce două abordări inovatoare pentru reconstrucția fețelor 3D. Ambele utilizează tehnici de tipul Structure From Motion (SFM) și Radial Basis Function (RBF) pe modele de înaltă densitate, pornind de la diverse imagini faciale în diferite poziții, în mare parte frontale. Prima abordare folosește un model 3D generic reantrenat cu repere suplimentare ale frunții, în timp ce a doua

abordare, folosind un model deformabil, integrează estimarea poziției feței și punctele frunții inferate dintr-o mască a părului.

Analizând abordările curente bazate pe Învățare Profundă [FWS⁺18], acestea reconstruiesc modelul 3D al feței dintr-o singură captură de cameră. Sistemul lor se bazează pe un model statistic, care a fost antrenat pe setul de date disponibil public. Aceste metode sunt susceptibile la outliers, deoarece acei subiecți sunt departe de modelul statistic prestabilit al feței. În experimentele noastre, subliniem îmbunătățirea în captarea detaliilor faciale, în special în zona frunții, și conturăm procesul de la detectarea reperelor până la reconstrucția 3D detaliată.

Procesul nostru începe cu un model 3D generic, texturat, al unei fețe umane care reprezintă un punct intermediar între caracteristicile masculine și feminine. Un pas crucial este stabilirea unei mapări între cele 73 de repere 2D și vârfurile 3D corespunzătoare pe acest model. Pentru a realiza acest lucru, redăm modelul 3D generic într-o imagine 2D și apoi aplicăm detectorul de repere faciale. Ulterior, folosim o tehnică de ray-casting pornind de la punctul de origine al sistemului către reperele detectate. Această metodă este folosită pentru a mapa fiecare reper 2D la un index specific de vârf pe modelul 3D. Maparea se realizează pe baza proximității triunghi-vârf cel mai apropiat, asigurând o aliniere precisă între reprezentările 2D și 3D.

Prima abordare valorifică modelul BU-3DFE 3D [YWS⁺06a] și un regresor DLIB extins [Kin09] pentru predicția reperelor frunții. Extensia furnizată include cele 5 puncte superioare ale frunții. A doua abordare folosește modelul Basel [GFB⁺18] și inferă punctele frunții dintr-o mască a părului, integrând o rețea de estimare a poziției. Ambele abordări implică utilizarea Structure From Motion (SFM) și Radial Basis Function (RBF) pe un model de vârfuri de înaltă densitate.

Pentru a măsura acuratețea metodei propuse, am calculat procentajul de repere, detectate pe randa feței din 3D, care se încadrează într-un $\epsilon = 10^{-2}$. Folosim modelele de față 3D din Baza de Date BU-3DFE [YWS⁺06a], care conține 100 de subiecți (56 de femei și 44 de bărbați). Obținem o acuratețe de 87.34% pe setul de date feminin și 91.25% pe setul de date masculin. Acuratețea generală peste conturul feței doar (fără caracteristicile de adâncime ale feței interne) este 98%.

Capitolul 6

Concluzii

În această teză, am contribuit la domeniul multifacetat al viziunii computerizate, concentrându-ne pe îmbunătățirea analizei faciale prin metode avansate de segmentare a părului și clasificare a culorii, detectarea pupilei și irisului, tăierea dinamică, distilarea cunoștințelor de ansamblu bazată pe atenție și reconstrucția feței 3D. Munca noastră îmbină tehnicile tradiționale de viziune pe computer cu abordări moderne de învățare automată, ceea ce duce la soluții inovatoare și progrese semnificative în domeniu.

Fiecare segment al acestei lucrări nu este o contribuție de sine stătătoare, ci o parte a unui întreg integrat, care urmărește să depășească limitele modului în care înțelegem și procesăm informațiile vizuale în era digitală. Mai precis, pornim de la detaliile complexe ale analizei feței 2D, avansăm în spațiul 3D expansiv pentru o înțelegere mai holistică și culminăm cu perfecționarea modelelor de învățare profundă prin căutarea arhitecturii inovatoare și optimizarea inferenței.

6.1 Abordarea unificată a analizei feței în spațiul 2D

Explorarea noastră în analiza feței 2D, care implică detectarea pupilei, irisului [TA20], segmentarea părului și clasificarea culorilor [BID18, IBD19] folosește o combinație de arbori de regresie, operații morfologice și clasificare bazată pe rețele neuronale. Metodele dezvoltate au arătat o acuratețe și eficiență ridicate, demonstrând aplicabilitatea lor practică.

Metodele noastre nu sunt limitate de niciun dispozitiv hardware și pot segmenta și detecta cu precizie regiunea vizibilă a irisului, feței și părului. Pupila și irisul sunt detectate cu precizie prin utilizarea unui algoritm iterativ care determină partea cea mai întunecată a ochiului. Rezultatele experimentale arată că cea mai mare eroare cumulată este mai mică de 0.05, ceea ce înseamnă că locația estimată a pupilei este în interiorul pupilei reale.

Mergând mai departe în analiza feței, am propus o serie de algoritmi începând cu segmentarea părului și a feței și apoi clasificarea culorilor. În cele din urmă, trecem dincolo de spațiul 2D propunând un proces de reconstrucție a feței 3D. Mai precis, zonele de păr și față sunt determinate folosind un CNN de ultimă generație. Operatori morfologici suplimentari sunt aplicați pe masca de par pentru a umple eventualele goluri din zona parului. Pixelii de păr sunt analizați în continuare fie printr-un random Forest Classifier bazat pe caracteristicile superpixeli [BDD16]), fie printr-o rețea neuronală artificială clasică [IBD19] pentru a determina culoarea părului. Pentru a antrena și a testa algoritmul propus, am adnotat peste 4000 de imagini dintr-o bază de date existentă cu culoarea părului.

Teza prezintă, de asemenea, studii comparative ale diferitelor tehnici utilizate în analiza feței. Folosind rezultatele numerice, se poate decide care spațiu de culoare sau arhitectura de rețea se potrivește cel mai bine pentru sarcini similare și oferă o bună intuiție pentru estimarea compromisului

dintre latență și precizie.

În lucrările viitoare, intenționăm să perfecționăm abordările actuale făcându-le mai rezistente la condițiile de iluminare și poziția feței. De asemenea, intenționăm să extindem sarcinile de analiză a feței pentru a include forma feței și detectarea culorii ochilor. Un subiect fierbinte în zilele noastre este generarea de imagini sintetice (realiste) folosind metode generative și utilizarea lor pentru a antrena modelele supravegheate. În plus, vom investiga direcția în care intrarea pentru clasificarea culorii părului este întreaga imagine. Pentru aceasta, vom folosi o arhitectură de învățare profundă și, potențial, mai multe date de intrare.

6.2 Dincolo de analiza 2D folosind reconstrucția feței 3D

Analiza feței 2D are limitările sale, în special pentru sarcinile în care subiectul ar trebui să fie într-o vedere frontală (detectia formei feței) sau când sunt necesare măsuri legate de adâncime pe față. Recunoscând limitările analizei 2D, munca noastră în reconstrucția feței 3D marchează un pas semnificativ înainte. În contextul spațiului 3D, am făcut următoarele contribuții: am proiectat un algoritm genetic pentru îmbunătățirea calibrării camerei (care selectează un subset mai bun de achiziții de intrare) și am dezvoltat două conducte pentru reconstrucția feței 3DMM, inclusiv zona frunții.

În ceea ce privește procedura de calibrare a camerei, am introdus o metodă bazată pe algoritmi genetici pentru a alege automat cele mai potrivite imagini de calibrare stereo dintr-o colecție mai mare de achiziții. Spre deosebire de abordarea comparată, conducta noastră este complet integrată și autonomă, eliminând necesitatea oricărei implicări a utilizatorului, în afară de capturarea mai multor instanțane ale unui model de calibrare din diferite unghiuri distincte. Considerăm că această abordare este mai potrivită pentru calibrarea camerelor din mai multe motive. În primul rând, folosește toate datele disponibile într-o abordare evolutivă, mai degrabă decât într-o manieră *greedy*, cum ar fi în lucrările similare. În al doilea rând, sistemul este plug-and-play, rulează folosind calculul paralel și poate realiza o calibrare excelentă pentru orice număr de camere, inclusiv calibrare mono și stereo.

Pentru reconstrucția facială 3D, propunem o conductă compusă din Structure From Motion urmată de funcții Radial Basis aplicate unui model de vârf de înaltă densitate pentru a obține o aproximare adecvată a feței subiectului în trei dimensiuni. Primul pas este folosit pentru a crea un nor de 73 de puncte de repere 3D, în timp ce al doilea pas deformează un model 3D generic pentru a realiza reconstrucția finală. În cele două abordări propuse, folosim modele generice diferite, cum ar fi Basel [GFB⁺18] și BU-3DFE [YWS⁺06b]. Extragem cele 68 de repere faciale 2D folosind regresorul DLIB [Kin09], în timp ce reperele superioare ale frunții sunt extrase folosind o versiune reantrenată a regresorului [Kin09] sau dintr-o mască de față, pe care am detectat-o anterior.

Metodele noastre au fost testate în scenarii din lumea reală și au reconstruit cu succes modele faciale 3D ale unor indivizi aleatoriu. În timp ce abordarea alternativă a reconstrucției one-shot folosind viziune profundă poate fi tentantă, are limitarea că zonele ocluse ar trebui "inventate" de rețea.

În ceea ce privește lucrările viitoare, intenționăm să creștem acuratetea modelului prin extinderea numărului de repere faciale utilizate. O abordare este de a genera fețe sintetice cu poziții binecunoscute ale reperelor faciale. În plus, ne vom concentra pe subiectul sensibil al clasificării formei feței și vom determina diverse metrice faciale interne (cum ar fi distanța interpupilară, dimensiunea ochilor etc.) în speranța de a dezvolta un sistem precis de recomandare a ochelarilor care să țină cont de valorile 3D îmbunătățite.

6.3 Îmbunătățiți căutarea arhitecturii, instruirea și inferența în modelele de învățare profundă

După cum am menționat anterior, cercetările și experimentele noastre se bazează în mare parte pe modele de învățare profundă. Când încercăm să implementăm modelele pentru a se potrivi scenariilor din lumea reală, întâmpinăm probleme în ceea ce privește latența și utilizarea memoriei. Pentru a rezolva aceste probleme, luăm în considerare următoarele soluții Căutare automată a arhitecturii neuronale, tăierea dinamică și distilarea cunoștințelor.

Proiectarea și implementarea unui algoritm de uz general care prelucrează toate aceste date poate fi o provocare și, prin urmare, trebuie create diferite tipuri și variații de algoritmi. Rezolvarea fiecărei probleme individual necesită resurse umane semnificative și poate dura mai mult timp. O soluție este utilizarea algoritmilor de căutare a arhitecturii neuronale pentru a găsi algoritmi necesari, ceea ce reduce nevoia de resurse umane. Am dezvoltat o abordare de căutare a arhitecturii neuronale care generează o rețea neuronală convoluțională folosind o rețea neuronală recurentă ca propunere. Cadrul nostru generează celule (grupuri de straturi) care sunt evaluate individual pe un set de date mai mic. Cele mai bune celule sunt instanțiate în continuare în șabloane mai mari pentru a le evalua pe setul de date țintă.

Am ales să testăm celula noastră generată pe două sarcini de segmentare semantică. Primul experiment a fost efectuat pe un set de date de ochelari de vedere, iar al doilea pe setul de date public Pascal VOC 2012 [EVGW+10]. Intersecția medie peste unire pe setul de testare al setului nostru de date este 0,9683, iar pentru setul de date general Pascal VOC, este 0,3289. Celula utilizată pentru evaluare a fost descoperită pe un set de date mai mic pentru ochelari, iar acesta poate fi un motiv pentru performanța scăzută a setului de date Pascal VOC de uz general. Pentru lucrări viitoare, intenționăm să descoperim o celulă specifică pentru fiecare sarcină.

Pentru optimizarea rețelei, am propus două metode (1) antrenarea rețelelor dinamice (adică ansambluri de porți) și (2) stimularea unei rețele mici folosind distilare de cunoștințe.

Pentru prima abordare de optimizare, ansamblul nostru dinamic numit *DynK-Hydra* are ca scop reducerea timpului de inferență al sarcinilor de clasificare cu un număr mediu până la mare de clase, păstrând în același timp acuratețea generală. Arătăm îmbunătățiri ale timpului de inferență de ordinul de 2-5,5 ori în comparație cu rețelele ResNet de bază [HZRS16] și o îmbunătățire marginală a preciziei de 1,2% față de lucrări similare, cum ar fi HydraRes [MMSF18]. Aplicăm o execuție rară dinamică și arătăm că o reducere semnificativă a timpului de inferență este încă posibilă în comparație cu HydraRes. Întrucât procesul propus implică doar faza de instruire (făcând-o mai complexă), considerăm că justifică câștigurile de timp de inferență.

Acest din urmă mecanism de optimizare implică o cohortă de studenți care sunt instruiți simultan. Ieșirea lor este concatenată dinamic ca o sumă ponderată realizată de un mecanism bazat pe atenție. Rezultatul final reprezintă cunoștințele *profesorului*, care sunt distilate înapoi la componente. Metoda propusă a fost testată pe mai multe seturi de date de referință, folosind arhitecturi de învățare profundă binecunoscute. În toate scenariile de antrenament, studentul distilat în cunoștințe este mai puternic decât un student instruit independent *vanilla*. În comparație cu abordări similare de ultimă generație, mecanismul nostru obține un câștig de precizie mai bun sau este cel puțin comparabil.

Cu toate acestea, există multe procese de optimizare a inferenței și toate implică, de obicei, o suprasolicitare în timpul antrenamentului. Ca o privire de ansamblu, dacă se folosesc seturi de date foarte mari și sistemul beneficiază de o capacitate mare de memorie, un ansamblu dinamic este potrivit. Dacă sarcina la îndemână nu este generică și se caută o arhitectură adecvată, relativ mică, se poate folosi o căutare a arhitecturii neuronale. Dacă dorim să folosim o arhitectură mică, dar să beneficiăm de o capacitate mare de pregătire, se pot aduna mulți studenți ușoare și să le sporească

performanța formându-i în distilare online de cunoștințe.

Generarea unei arhitecturi de rețele neuronale profunde personalizată și reducerea costurilor de inferență (atât amprenta memoriei, cât și latența) sunt de mare interes pentru noi. În lucrările viitoare, intenționăm să îmbunătățim capacitățile tuturor metodelor propuse: cadru generativ, ansamblu dinamic și proces de distilare a cunoștințelor. Pentru a îmbunătăți cadrul generativ (arhitectură) și a-l face și mai rapid, intenționăm să introducem o funcție de evaluare surogat care poate evalua celulele intermediare fără a le antrena. Pentru procesul de distilare a cunoștințelor, căutăm modalități de a distila cunoștințele la nivelul hărților de caracteristici. Mai mult, investigăm diverse caracteristici suplimentare care vor fi utilizate pentru mecanismul bazat pe atenție pentru a ne asigura că distilăm doar cunoștințele relevante. În ansamblul dinamic, ne propunem să explorăm peisajul costurilor (atât vizual, cât și numeric), în special pentru ramurile individuale, pentru a înțelege și ghida mai bine procesul de antrenament. În plus, ne vom concentra pe îmbunătățirea procedurii de antrenament pentru a reduce numărul de ramuri activate la 1 (în prezent, numărul mediu de activare este mai mare de 2,5).

Bibliografie

- [ARL⁺10] Oleg Alexander, Mike Rogers, William Lambeth, Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul Debevec. The digital emily project: Achieving a photo-realistic digital actor. *IEEE Computer Graphics and Applications*, 30(4):20–31, 2010.
- [BCNM06] Cristian Buciluă, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 535–541, 2006.
- [BDD16] Diana Borza, Adrian Darabant, and Radu Danescu. Real-time detection and measurement of eye features from color images. *Sensors*, 16(7):1105, 2016.
- [BDIM22] Diana-Laura Borza, Adrian Sergiu Darabant, Tudor Alexandru Ileni, and Alexandru-Ion Marinescu. Effective online knowledge distillation via attention-based model ensembling. *Mathematics*, 10(22):4285, 2022.
- [BGNR16] Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. Designing neural network architectures using reinforcement learning. *arXiv preprint arXiv:1611.02167*, 2016.
- [BID18] Diana Borza, Tudor Ileni, and Adrian Darabant. A deep learning approach to hair segmentation and color extraction from facial images. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 438–449. Springer, 2018.
- [Bre96] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- [Bro71] D. C. Brown. Close-Range Camera Calibration. *Photogrammetric Engineering and Remote Sensing*, 37(8):855–866, 1971.
- [BWHY05] Gavin Brown, Jeremy Wyatt, Rachel Harris, and Xin Yao. Diversity creation methods: a survey and categorisation. *Information fusion*, 6(1):5–20, 2005.
- [CGW⁺19] Han Cai, Chuang Gan, Tianzhe Wang, Zhekai Zhang, and Song Han. Once-for-all: Train one network and specialize it for efficient deployment. *arXiv preprint arXiv:1908.09791*, 2019.
- [CMW⁺20] Defang Chen, Jian-Ping Mei, Can Wang, Yan Feng, and Chun Chen. Online knowledge distillation with diverse peers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34(04), pages 3430–3437, 2020.
- [DCJ⁺94] Harris Drucker, Corinna Cortes, Lawrence D Jackel, Yann LeCun, and Vladimir Vapnik. Boosting and other ensemble methods. *Neural Computation*, 6(6):1289–1301, 1994.

- [EMH18] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *arXiv preprint arXiv:1808.05377*, 2018.
- [EVGW⁺10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.
- [Fai75] W. Faig. Close-Range Camera Calibration:Mathematical Formulation. *Photogrammetric Engineering and Remote Sensing*, 41(12):1479–1486, 1975.
- [FHL19] Stanislav Fort, Huiyi Hu, and Balaji Lakshminarayanan. Deep ensembles: A loss landscape perspective. *arXiv preprint arXiv:1912.02757*, 2019.
- [FWS⁺18] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In *ECCV*, 2018.
- [GFB⁺18] Thomas Gerig, Andreas Forster, Clemens Blumer, Bernhard Egger, Marcel Lüthi, Sandro Schönborn, and Thomas Vetter. Morphable face models - an open framework. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 75–82, 2018.
- [GWW⁺20] Qiushan Guo, Xinjiang Wang, Yichao Wu, Zhipeng Yu, Ding Liang, Xiaolin Hu, and Ping Luo. Online knowledge distillation via collaborative learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11020–11029, 2020.
- [GYMT20] Jianping Gou, Baosheng Yu, Stephen John Maybank, and Dacheng Tao. Knowledge distillation: A survey. *arXiv preprint arXiv:2006.05525*, 2020.
- [Har94] Hartley. An algorithm for self calibration from several views. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 908–912, 1994.
- [HHS⁺21] Yizeng Han, Gao Huang, Shiji Song, Le Yang, Honghui Wang, and Yulin Wang. Dynamic neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [HLP⁺17] Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Liu, John E Hopcroft, and Kilian Q Weinberger. Snapshot ensembles: Train 1, get m for free. *arXiv preprint arXiv:1704.00109*, 2017.
- [HP18] Tien Ho-Phuoc. Cifar10 to compare visual recognition performance between deep neural networks and humans. *arXiv preprint arXiv:1811.07270*, 2018.
- [HVD15] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
- [HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [IBD19] Tudor Alexandru Ileni, Diana Laura Borza, and Adrian Sergiu Darabant. Fast in-the-wild hair segmentation and color classification. In *VISIGRAPP (4: VISAPP)*, pages 59–66, 2019.

- [IDBM22] Tudor Alexandru Ileni, Adrian Sergiu Darabant, Diana Laura Borza, and Alexandru Ion Marinescu. Dynk-: improved dynamic architecture ensembling for efficient inference. *Complex & Intelligent Systems*, pages 1–12, 2022.
- [JHP21] Mingi Ji, Byeongho Heo, and Sungrae Park. Show, attend and distill: Knowledge distillation via attention-based feature matching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7945–7952, 2021.
- [Kel70] Michael David Kelly. *Visual identification of people by computer*. Number 130 in -. Department of Computer Science, Stanford University., 1970.
- [KH⁺09] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. *arXiv*, 2009.
- [Kin09] Davis E King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(Jul):1755–1758, 2009.
- [KKA⁺20] Khalil Khan, Rehan Ullah Khan, Kashif Ahmad, Farman Ali, and Kyung-Sup Kwak. Face segmentation: A journey from classical to deep learning paradigm, approaches, trends, and directions. *IEEE Access*, 8:58683–58699, 2020.
- [KPR⁺14] A Krupka, J Prinosil, K Riha, J Minar, and M Dutta. Hair segmentation for color estimation in surveillance systems. In *Proc. 6th Int. Conf. Adv. Multimedia*, pages 102–107, 2014.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [LCP⁺18] Alex Levinshtein, Cheng Chang, Edmund Phung, Irina Kezele, Wenzhangzhi Guo, and Parham Aarabi. Real-time deep hair matting on mobile devices. In *2018 15th Conference on Computer and Robot Vision (CRV)*, pages 1–7. IEEE, 2018.
- [LCY13] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.
- [LF97] Q.T. Luong and O. D. Faugeras. Self-Calibration of a Moving Camera from Point-Correspondences and Fundamental Matrices. *Int. Journal of Computer Vision*, 22(3):261–289, 1997.
- [LTW95] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Realistic modeling for facial animation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 55–62, 1995.
- [LY15] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7:7, 2015.
- [LZG18] Xu Lan, Xiatian Zhu, and Shaogang Gong. Knowledge distillation by on-the-fly native ensemble. *arXiv preprint arXiv:1806.04606*, 2018.
- [MDI20] Alexandru Ion Marinescu, Adrian Sergiu Darabant, and Tudor Alexandru Ileni. A fast and robust, forehead-augmented 3d face reconstruction from multiple images using geometrical methods. In *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6. IEEE, 2020.

- [MDI21] Alexandru Ion Marinescu, Adrian Sergiu Darabant, and Tudor Alexandru Ileni. Optimal stereo camera calibration via genetic algorithms. In *2021 Workshop of Artificial Intelligence for Autonomous Driving (IJCAI)*, pages 1–1. IEEE, 2021.
- [MID19] Alexandru Ion Marinescu, Tudor Alexandru Ileni, and Adrian Sergiu Darabant. A versatile 3d face reconstruction from multiple images for face shape classification. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6. IEEE, 2019.
- [MMSF18] Ravi Teja Mullapudi, William R Mark, Noam Shazeer, and Kayvon Fatahalian. Hydranets: Specialized dynamic architectures for efficient inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8080–8089, 2018.
- [MPS21] Araceli Morales, Gemma Piella, and Federico M Sukno. Survey on 3d face reconstruction from uncalibrated images. *Computer Science Review*, 40:100400, 2021.
- [MSLB18] Umar Riaz Muhammad, Michele Svanera, Riccardo Leonardi, and Sergio Benini. Hair detection, segmentation, and hairstyle classification in the wild. *Image and Vision Computing*, 71:25–37, 2018.
- [NID20] Sergiu Cosmin Nistor, Tudor Alexandru Ileni, and Adrian Sergiu Dărăbant. Automatic development of deep learning architectures for image segmentation. *Sustainability*, 12(22):9707, 2020.
- [PN17] Hugo Proença and João C Neves. Soft biometrics: Globally coherent solutions for hair segmentation and style recognition based on hierarchical mrfs. *IEEE Transactions on Information Forensics and Security*, 12(7):1637–1645, 2017.
- [PS19] Songyou Peng and Peter Sturm. Calibration Wizard: A guidance system for camera calibration based on modelling geometric and corner uncertainty. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1497–1505, 2019.
- [SC18] Guocong Song and Wei Chai. Collaborative learning for deep neural networks. *arXiv preprint arXiv:1805.11761*, 2018.
- [Sch13] Robert E Schapire. Explaining adaboost. In *Empirical inference*, pages 37–52. Springer, 2013.
- [SLJ⁺15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [SS97] Amanda JC Sharkey and Noel E Sharkey. Combining diverse neural nets. *The Knowledge Engineering Review*, 12(3):231–247, 1997.
- [SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [TA20] Ileni Tudor-Alexandru. Efficient iris segmentation and pupil detection for visagisme applications. In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 000123–000128. IEEE, 2020.

- [VJ01] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee, 2001.
- [VJ04] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.
- [WG21] Guile Wu and Shaogang Gong. Peer collaborative learning for online knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35(12), pages 10302–10310, 2021.
- [WZL⁺19] Caiyong Wang, Yuhao Zhu, Yunfan Liu, Ran He, and Zhenan Sun. Joint iris segmentation and localization using deep multi-task learning framework. *arXiv preprint arXiv:1901.11195*, 2019.
- [YWS⁺06a] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato. A 3d facial expression database for facial behavior research. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, FGR '06*, page 211–216, USA, 2006. IEEE Computer Society.
- [YWS⁺06b] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato. A 3d facial expression database for facial behavior research. *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 211–216, 2006.
- [ZL16] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016.
- [ZXHL18] Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. Deep mutual learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4320–4328, 2018.