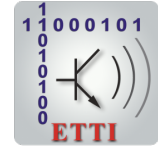




**NATIONAL UNIVERSITY FOR  
SCIENCE AND TECHNOLOGY  
POLITEHNICA BUCURESTI**



**Doctoral School of Electronics, Telecommunications  
and Information Technology**

**Decision No. 118 from 26/10/2023**

**Ph.D. THESIS  
SUMMARY**

**Saqib Nazir**

---

**OBȚINEREA HĂRȚILOR DE ADÂNCIME DIN IMAGINI  
DEFOCALIZATE FOLOSIND REȚELE NEURALE PROFUNDE**

**DEEP DEPTH FROM DEFOCUS FOR NEAR RANGE AND IN-SITU  
3D EXPLORATION**

---

**THESIS COMMITTEE**

<b>Prof. Dr. Ing. Gheorghe BREZEANU</b> UNSTPB	President
<b>Prof. Dr. Ing. Daniela COLȚUC</b> UNSTPB	PhD Supervisor
<b>Prof. Dr. Ing. Víctor Manuel Brea SANCHEZ</b> University of Santiago de Compostela (USC), Spain	Referee
<b>Dr. Ing. Habil. Miguel HEREDIA CONDE</b> University of Siegen, Germany	Referee
<b>Prof. Dr. Ing. Mihai DATCU</b> UNSTPB	Referee

**BUCHAREST 2021**

---

# Abstract

O imagine realizată cu o cameră convențională este o proiecție 2D a unei scene 3D. În timpul procesului de imagistică, informațiile celei de-a treia dimensiuni, adică adâncimea scenei, se pierde. Cu toate acestea, poate fi recuperată prin calcul, din setul de indicii vizuale prezente în imagini. Acest lucru a creat premisele pentru a percepe lumea în 3D prin intermediul camerelor convenționale.

Teza abordează provocarea estimării adâncimii monoculare folosind o singură imagine defocalizată, o sarcină esențială în viziunea computerizată, având în vedere aria largă de aplicații, de la robotică la mașini autonome. Progresele recente în învățarea profundă au revoluționat domeniul viziunii computerizate și, în special, al estimării profunzimii. Cu toate acestea, tehnicile anterioare bazate pe învățarea profundă neglijează adesea potențialul estomparii defocalizării, care este un indiciu important pentru estimarea adâncimii. Soluțiile existente folosesc mai multe imagini sau stive focale ale aceleiași scene și rareori imagini individuale. Pentru a umple acest gol, propunem o arhitectură nouă numită 2HDED:NET, care abordează atât estimarea adâncimii, cât și deblurring-ul imaginii dintr-o singură imagine defocalizată. Din cauza lipsei de seturi de date cu imagini defocalizate în mod natural și adevărul la adâncime, rețele precum 2HDED:NET sunt, în general, instruite pe date sintetice, fapt care le reduce performanțele pe date reale. Pentru a antrena 2HDED:NET la întregul său potențial, am construit un nou set de date numit iDFD care conține imagini defocalizate în mod natural, adnotate dublu cu imaginea cu focalizare completă și harta de adâncime a timpului zborului.

În cele din urmă, teza explorează domeniul promițător al învățării auto-supravegheate, prin conversia 2HDED:NET în estimarea hărții defocalizate în absența adâncimii adevărate la sol pentru antrenament. Pentru a face acest lucru, 2HDED:NET este îmbunătățit cu un modul de simulare a defocalizării care reconstruiește imaginea defocalizată din cea all-in-focus și harta de defocalizare estimată. Competențele modelului în estimarea hărții defocalizate este la egalitate cu cea a modelelor supravegheate de ultimă generație care utilizează mai multe imagini. Experimente cuprinzătoare efectuate pe diverse seturi de date reale sau sintetice au validat eficacitatea abordărilor propuse pentru estimarea hărții de adâncime sau defocalizare în diferite setări, cuprinzând medii interioare și exterioare.

# Table of contents

<b>Abstract</b>	<b>ii</b>
<b>1 Introducere</b>	<b>1</b>
1.1 Prezentarea domeniului tezei de doctorat . . . . .	1
1.2 Scope of the doctoral thesis . . . . .	1
1.3 Conținutul tezei de doctorat . . . . .	2
<b>2 Estimarea adâncimii monoculare folosind Deep Learning</b>	<b>3</b>
2.1 Regularizare de netezire . . . . .	3
2.1.1 Valori pentru claritatea marginilor . . . . .	4
2.2 Rezultate experimentale . . . . .	5
<b>3 Estimarea adâncimii din imagini defocalizate</b>	<b>7</b>
3.1 2HDED:NET Arhitectură . . . . .	7
3.2 Rezultate experimentale . . . . .	10
<b>4 iDFD: un set de date pentru estimarea adâncimii și defocalizarea</b>	<b>13</b>
4.1 Setul de date iDFD pentru DFD și Deblurring imagini . . . . .	13
4.2 Rezultate experimentale . . . . .	14
<b>5 Învățare auto-supravegheată pentru estimarea hărții defocalizate</b>	<b>19</b>
5.1 Autosupravegheat 2HDED:NET . . . . .	19
5.2 Rezultate experimentale . . . . .	21
<b>6 Concluzii</b>	<b>23</b>
6.1 Contribuții originale . . . . .	24
6.2 List of original publications . . . . .	25
6.2.1 Journal Paper . . . . .	25
6.2.2 Conference Papers . . . . .	25
6.3 Perspective pentru evoluții ulterioare . . . . .	26
<b>References</b>	<b>28</b>

# Chapter 1

## Introducere

### 1.1 Prezentarea domeniului tezei de doctorat

Viziunea computerizată (CV) este un domeniu al inteligenței artificiale (AI), care permite computerelor să extragă informații semnificative din imaginile digitale. Această teză se încadrează în categoria CV precoce, abordând problema provocatoare și fundamentală, numită Estimarea adâncimii monoculare (MDE). Cu modelele Deep Learning (DL), este posibil să extrageți caracteristici profunde dintr-o imagine RGB cu o singură captură, care poate oferi rezultate mai bune decât metodele analitice tradiționale.

### 1.2 Scope of the doctoral thesis

MDE a obținut îmbunătățiri semnificative odată cu dezvoltarea rețelelor neuronale profunde (DNN) în ultimii ani, dar importanța estompării defocalizării nu a fost încă exploatată. Prin urmare, scopul acestei teze cuprinde studiul și dezvoltarea unei noi abordări pentru estimarea adâncimii folosind estomparea defocalizării ca indiciu. Accentul principal este investigarea relației dintre neclaritatea defocalizării și informațiile de profunzime din imaginile monoculare. Un DNN numit 2HDED:NET este proiectat pentru a realiza estimarea adâncimii de la neclaritatea defocalizării, în timp ce abordează simultan deneclaritatea imaginii. Teza implică dezvoltarea unui set de date numit iDFD, constând din imagini defocalizate din lumea reală, omologii lor All-in-Focus (AiF) și hărți de adâncime corespunzătoare. Acest set de date servește ca o resursă valoroasă pentru instruirea și evaluarea modelului DNN propus. În plus, am îmbunătățit 2HDED:NET, pentru a permite învățarea auto-supravegheată pentru estimarea hărții defocalizate. Rețeaua este concepută pentru a învăța și estima hărțile defocalizate într-o manieră complet auto-supravegheată, eliminând nevoia de adnotare etichetată sau de supraveghere explicită.

### 1.3 Conținutul tezei de doctorat

Al doilea capitol prezintă evoluțiile recente în MDE folosind DL. Există diferite opțiuni pentru arhitecturi, funcții de pierdere, regularizare de netezire și setări experimentale propuse în literatură. Prin urmare, este dificil de stabilit influența lor asupra performanțelor. Am făcut o comparație a diferitelor regularizări de netezire propuse în literatură atât în metodele supravegheate, cât și în cele autosupravegheate. În plus, am folosit un termen de regularizare de netezire folosit într-un mod auto-supravegheat pentru a lucra într-o manieră supravegheată și am arătat că tehnicile propuse depășesc ceilalți termeni de netezire. Părți din acest capitol au fost publicate în [12].

Al treilea capitol se concentrează pe explorarea neclarității defocalizării ca indiciu pentru estimarea adâncimii. Estimarea adâncimii și deblurring-ul imaginii sunt două sarcini fundamentale și strâns legate. Efectuarea oricăreia dintre ele bazându-se pe o singură imagine este o problemă prost pusă. În ciuda acestui fapt, majoritatea modelelor existente le tratează separat. În acest capitol, am propus Rețeaua de estimare și deblurring cu două capete (2HDED:NET), care extinde o rețea convențională Depth from Defocus (DFD) cu o ramură de deblurring care împarte același codificator ca și ramura de adâncime. Părți din acest capitol au fost publicate în [14, 15].

Multe seturi de date au fost propuse în literatura de specialitate în scopul estimării adâncimii, NYU-Depth v2 și KiTTi fiind cele mai cunoscute. Dar când vine vorba de DFD, nu există un astfel de set de date cu defocalizarea reală. Prin urmare, majoritatea metodelor anterioare își antrenează DNN-ul în seturi de date sintetice. Pentru a depăși limitarea unui set de date cu neclaritatea reală a defocalizării, am propus Indoor Depth from Defocus (iDFD), un set de date Annotat Depth And Defocus, care conține imagini AiF defocalizate în mod natural și hărți dense de adâncime ale mediilor interioare. Părți din acest capitol au fost publicate în [13].

În acest capitol, propunem un model de învățare auto-supravegheat pentru Estimarea Hărții Defocalizate (DME) dintr-o singură imagine defocalizată. Metoda noastră se bazează pe un DNN propus recent, numit 2HDED:NET, pe care îl completăm cu un modul de simulare a defocalizării. Arătăm că rețeaua noastră auto-supravegheată ocolește nevoia de defocalizare Ground Truth (GT) sau hărți de adâncime. Pe lângă DME, rețeaua noastră reconstruiește imaginea AiF prin învățare supravegheată. Testăm rețeaua pe benchmark-uri sintetice și realiste și demonstrăm că este o soluție eficientă pentru DME și deblocarea imaginii atunci când este disponibilă o singură imagine defocalizată.

## Chapter 2

# Estimarea adâncimii monoculare folosind Deep Learning

În ultimii ani, au fost propuse mai multe arhitecturi profunde diferite, iar alegerea arhitecturii, a funcției de pierdere, a hiper-parametrilor și a modelelor preinstruite adecvate este importantă. În acest capitol, am analizat cele mai performante modele împreună cu alegerea rețelelor, etc. În plus, am analizat diferiți termeni de regularizare a netezimii și i-am selectat pe cei mai eficienți utilizați în studiile anterioare.

### 2.1 Regularizare de netezire

Regularizarea de netezire este utilizată pentru netezirea zonelor omogene ale obiectelor dintr-o scenă fără a degrada marginile unei adâncimi prezise. Având în vedere că locația marginilor abrupte din imaginile naturale este necunoscută, acestea trebuie identificate în același timp cu reconstrucția obiectului. Pentru a încuraja netezirea în interiorul unei regiuni plane și a descuraja netezirea pe margini, în studiile anterioare au fost propuse diferite soluții [5, 18, 9, 20, 6, 7]. Acest capitol face o comparație între diferiți termeni de regularizare utilizați fie în metodele de învățare supravegheată, fie în cele de învățare autosupravegheată. În plus, termenul de regularizare utilizat în prezent în metodele autosupravegheate a fost modificat, cum ar fi lucrul în mod supravegheat. Rezultatele experimentale pe NYU-Depth v2 au arătat că regularizarea bazată pe norma L1 a gradientului este cea mai bună, iar cea modificată în mod autosupravegheat le depășește pe celelalte. În cele din urmă, mai degrabă decât să se bazeze pe metrici de evaluare comune, se utilizează o măsură suplimentară de precizie bazată pe piramida dirijabilă și pe divergența Kullback-Leibler (KLD) pentru precizia marginilor adâncimilor estimate, care sunt mai sensibile la erorile de poziționare a marginilor.

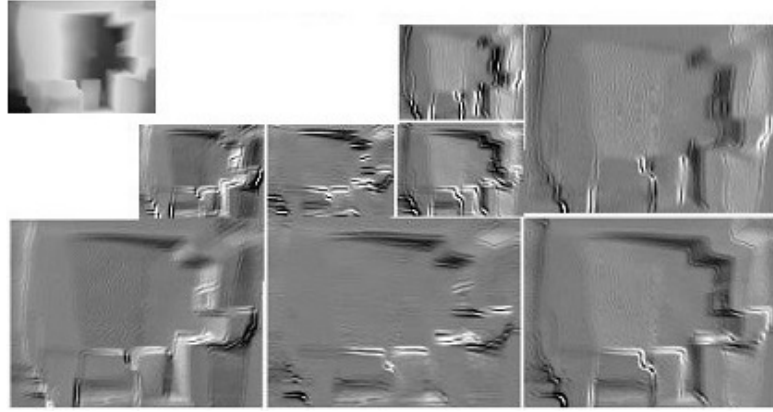


Figure 2.1 Descompunerea piramidei orientabile.

### 2.1.1 Valori pentru claritatea marginilor

Funcțiile de bază ale transformării wavelet dirijabile sunt derivatele direcționale de ordinul  $K_{th}$ , care se prezintă în mai multe scări și 4 orientări. Un exemplu de descompunere a hărții de adâncime estimate este prezentat în figura 2.1. Această piramidă dirijabilă specială conține  $4 \text{ subbenzideorientare}$ , la 2 scări. Motivul selectării primelor două scări este că acestea conțin cea mai mare frecvență spațială a imaginii. Având în vedere că degradarea marginilor cauzată de netezire este vizibilă în primul rând în domeniul frecvențelor înalte și ținând cont de dimensiunile imaginilor de test, nu s-a urmărit generarea unor scări mai mici din cauza impactului nesemnificativ pe care l-ar fi avut asupra evaluării KLD.

Pentru a evalua marginile, atât GT, cât și adâncimea estimată sunt descompuse pe 2 scări și 4 orientări, rezultând 8 sub-benzi, apoi KLD este calculat pe sub-benzi. Pentru a estima KLD, se creează histograme ale fiecărei sub-benzi, iar divergența sub-benzilor corespunzătoare se calculează așa cum este descris în [19]:

$$d(h_m||h) = \sum_{i=1}^L h_m(i) \log \frac{h_m(i)}{h(i)} \quad (2.1)$$

unde  $h_m(i)$  și  $h(i)$  sunt înălțimile normalizate ale histogramelor  $i_{th}$  ale adâncimilor GT și, respectiv, ale adâncimilor estimate, iar  $L$  este numărul de bini din histograme.

În cele din urmă, KLD global între adâncimea estimată și adâncimea GT se obține astfel:

$$D = \log_2 \left( 1 + \frac{1}{D_o} \sum_{K=1}^K |d^k(h_m^k||h^k)| \right) \quad (2.2)$$

unde  $K$  este numărul de sub-benzi, iar  $D_o$  este constanta folosită pentru a controla scara măsurării distorsiunii, în cazul nostru  $k = 8$  și  $D_o = 10$ .

Table 2.1 Comparație cantitativă cu diferite regularizări de netezire. Pentru RMSE, REL, Log 10 și KLD, mai mică este mai bună. Pentru  $\delta_1$ ,  $\delta_2$  și  $\delta_3$ , mai mare este mai bine.

Method	RMSE	REL	Log 10	$\delta_1$	$\delta_2$	$\delta_3$	KLD
Non-Regularization	0.619	0.159	0.064	0.786	0.950	0.986	9.0091e+03
Hu[9]	<b>0.555</b>	0.126	0.054	0.841	0.967	0.991	7.2385e+03
Eigen[5]	0.598	0.143	0.060	0.812	0.956	0.988	7.3216e+03
Ummenhofer[18]	0.582	0.134	0.058	0.822	0.963	0.991	7.2817e+03
Xian[20]	0.559	<b>0.125</b>	<b>0.054</b>	<b>0.844</b>	<b>0.968</b>	<b>0.991</b>	<b>7.1009e+03</b>
Godard[7] (Modified)	0.575	0.142	0.057	0.825	0.961	0.988	7.1929e+03

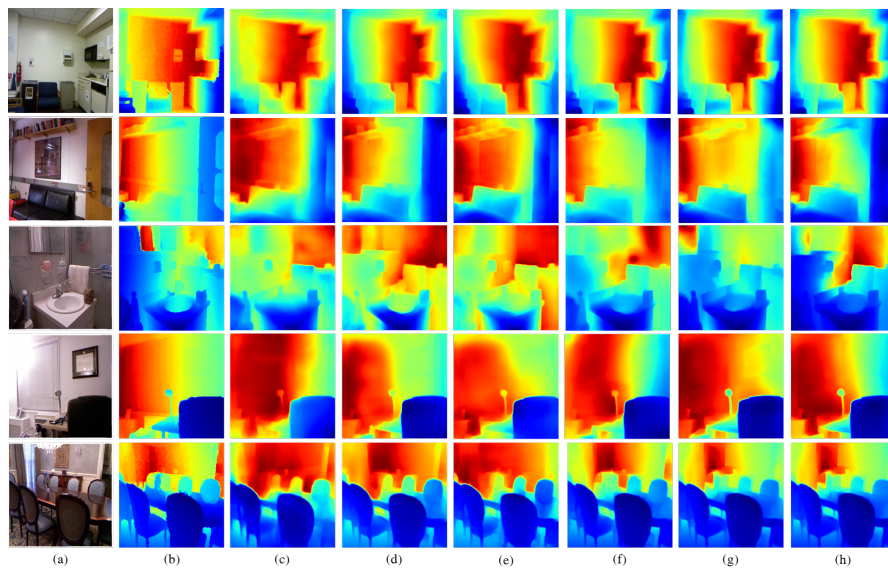


Figure 2.2 Hărți de adâncime estimate pentru 6 scene și 5 regularizări de netezire diferite. De la primul la ultimul rând; (a) imagini RGB de intrare, (b) harta de adâncime GT, (c) rezultate fără termen de regularizare, (d) Hu [9], (e) Eigen [5], (f) Ummenhofer [18], (g) Xian [20] și (h) funcția de pierdere Godard [7] (modificată).

## 2.2 Rezultate experimentale

Tabelul 2.1 prezintă rezultatele obținute pentru 5 regularizări diferite și pentru cazul fără nicio regularizare. Se poate observa că cele mai proaste rezultate sunt obținute fără regularizarea de netezire, ceea ce demonstrează încă o dată utilitatea introducerii unei astfel de regularizări. Cele mai bune rezultate sunt obținute pentru Hu [9] și Xian [20], acesta din urmă fiind ușor avantajat, cu excepția RMSE. Godard [7] depășește Eigen [5], deși provine din metodele autosupravegheate.

Rezultatele *KLD* pentru acuratețea marginilor sunt prezentate în ultima coloană a tabelului reftab2.3. Deși metodele supravegheate au produs performanțe bune în ceea ce privește indicatorii de evaluare comuni, precizia lor nu este apreciabilă în ceea ce privește



precizia marginilor, cu excepția lui Xian [20], care este de departe cea mai bună. În mod surprinzător, al doilea cel mai bun rezultat este oferit de soluția modificată provenită din metodele autosupravegheate. Explicația constă în utilizarea aceleiași norme L1. Se știe că norma L1 favorizează erorile mai rarefiate. În regularizarea de netezire, norma L1 este aplicată gradientului imaginilor, ceea ce dezvăluie dispersia.

Figura 2.2 prezintă hărțile de adâncime estimate prin utilizarea diferitelor regularizări de netezire, precum și imaginile RGB și GT corespunzătoare. Se poate observa că rezultatele fără termenul de regularizare suferă de o distorsiune puternică a formelor. Deși Hu și Eigen [9, 5] au reușit să producă imagini cu limite clare, acestea prezintă totuși mai multe margini moi incorecte, de exemplu, limitele lămpii de lumină de pe birou în cazul Hu [9] și ale obiectelor de pe chiuvetă în cazul Eigen [5]. Termenul de regularizare modificat în Godard [7] prezintă performanțe considerabile, producând cu acuratețe marginile obiectelor și structurile minuscule, cum ar fi obiectele din jurul chiuvetei și lampa de lumină de pe birou. De asemenea, în imaginea cu canapeaua, pot fi observate limite clare. Aici, Xian [20] nu a reușit să detecteze limitele exacte ale acestei imagini.

## Chapter 3

# Estimarea adâncimii din imagini defocalizate

Acest capitol prezintă cadrul DNN propus de noi pentru abordarea comună a sarcinilor de DFD și de debirozare a imaginii folosind o singură imagine defocalizată. Deși aceste sarcini sunt legate, modelele existente le tratează de obicei ca probleme separate. Metodele recente bazate pe DNN care au ca scop rezolvarea simultană a acestor sarcini estimează mai întâi harta de adâncime sau de defocalizare și apoi reconstruiesc imaginea focalizată pe baza acestei estimări [1]. În abordarea noastră, este introdusă o rețea de estimare a adâncimii și debirozare cu două capete (2HDED:NET), care extinde o rețea DFD convențională prin încorporarea unei ramuri de debirozare care împarte același codificator ca și ramura de adâncime.

### 3.1 2HDED:NET Arhitectură

Figura 3.1 descrie arhitectura 2HDED:NET. Având în vedere o singură imagine defocalizată  $I$ , scopul rețelei noastre este de a estima harta de adâncime  $\hat{I}^{depth}$  și de a restaura imaginea AiF  $\hat{I}^{aif}$ . După cum se arată în figura 3.1, 2HDED:NET este alcătuită dintr-un codificator și două decodificatoare care produc în paralel harta de adâncime și imaginea AiF. Prin utilizarea caracteristicilor învățate de către același codificator, ambele capete pot beneficia reciproc unul de celălalt. 2HDED:NET este o metodă supravegheată care necesită adâncimea GT, precum și imaginile AiF pentru instruire.

Pentru rețeaua de codificare, se utilizează rețeaua DenseNet-121 [2]. După cum sugerează și numele său, DenseNet constă în straturi conectate dens. La fel ca în cazul [2], stratul max-pooling este înlocuit cu un strat convoluțional de  $4 \times 4$  pentru a reduce rezoluția, crescând în același timp numărul de hărți de canale de caracteristici. Conexiunile de săritură sunt utilizate între părțile de codificare și decodificare pentru a simplifica învățarea. Codificatorul ajută la obținerea de caracteristici cu rezoluție

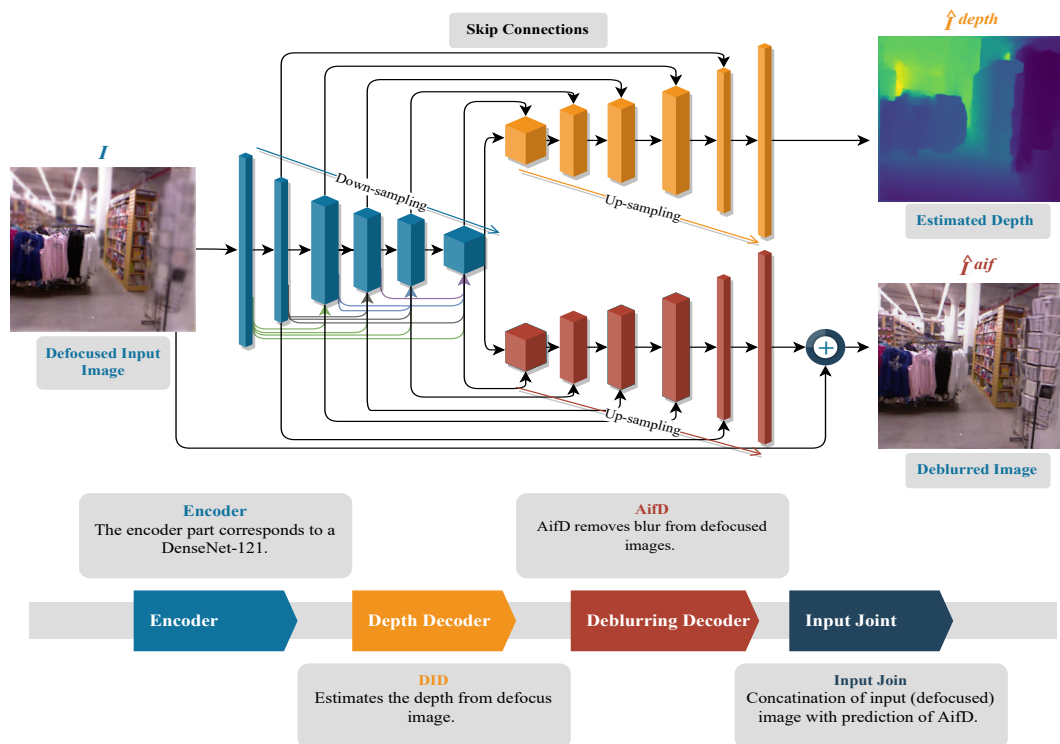


Figure 3.1 Arhitectura 2HDED:NET constă dintr-un codificator și două decodificatoare care funcționează în paralel. Capul superior estimează harta de adâncime, iar cel inferior imaginea AiF. Rețeaua este alimentată cu imagini RGB defocalizate.

multiplă din imaginea de intrare, care sunt utile pentru cele două sarcini pe care le îndeplinește 2HDED:NET.

Decodicatorul de estimare a adâncimii (DED) este inspirat de [2]. Acesta este format din 5 straturi de decodare, fiecare cu 4  $ori4$  care mărește rezoluția hărții caracteristicilor, urmată de o convoluție de  $3 \times 3$  care reduce efectul de aliasing al suprașantionării. Funcțiile de normalizare a loturilor și ReLU sunt incluse după fiecare strat de convoluție pentru a face învățarea mai stabilă și pentru a permite reprezentarea neliniarităților.

Decodorul de deblurring este denumit decodor AiF (AifD). Spre deosebire de DED, ieșirea lui AifD este o imagine RGB cu trei canale. Un strat comun de intrare este utilizat pentru a agrega imaginea de intrare defocalizată cu ieșirea AifD pentru predicția finală. Conținutul imaginii defocalizate și predicția corespunzătoare din AifD sunt încorporate în stratul comun de intrare, oferind acestui cap o orientare mai detaliată pentru învățarea debirografierii. Spre deosebire de metodele care utilizează procesarea în conductă, în care mai întâi se prezice harta de profunzime sau de defocalizare și apoi se recuperează imaginea AiF, capul nostru de debirozare nu se bazează pe astfel de estimări, evitând dependența de hărți de profunzime insuficiente în unele cazuri.

O caracteristică importantă a soluției propuse este că, odată ce 2HDED:NET este complet antrenat, o sarcină poate fi efectuată și atunci când celălalt cap este îndepărtat, de exemplu, DFD poate fi efectuată fără capul AifD și viceversa.

Instruirea 2HDED:NET este supravegheată simultan de hărțile de adâncime GT și de imaginile AiF. Pentru a lua în considerare aceste informații duble, se propune o funcție de pierdere cu doi termeni, unul care ia în considerare pierderea de adâncime și altul pentru imaginea deblurată. Aceste două componente sunt echilibrate pentru a avea contribuții aproximativ egale.

Majoritatea metodelor DL propuse pentru estimarea adâncimii au fost antrenate cu funcții de pierdere bazate pe regresie la nivel de pixel, calculate ca medie a diferențelor absolute (norma  $L1$ ), diferențe pătrate (norma  $L2$ ) sau combinații ale acestora [2].

Se recurge la norma  $L1$  ca funcție de pierdere pentru estimarea adâncimii, recunoscută pentru capacitatea sa de a estima soluții rare, așa cum se observă în cazul hărților de adâncime [2, 14]:

$$L_1^{Depth} = \frac{1}{n} \sum_{i=1}^n |\hat{I}_i^{depth} - I_i^{depth}| \quad (3.1)$$

unde  $\hat{I}^{profunditate}$  este adâncimea estimată,  $I^{profunditate}$  este GT,  $i$  este pixelul curent și  $n$  este numărul de pixeli.

Adesea, această pierdere este completată de un termen de regularizare de netezire care are rolul de a elimina structurile de amplitudine redusă din harta de adâncime, în timp ce ascute marginile principale [7, 8, 20]. În cazul rețelei propuse, precizia adâncimii este îmbunătățită prin combinarea normei  $L1$  cu termenul de netezire utilizat în mod obișnuit în învățarea supravegheată și definit ca [20]:

$$L_{grad} = \frac{1}{n} \sum_i |\Delta_x R_i| + |\Delta_y R_i| \quad (3.2)$$

unde  $R_i = \hat{I}_i^{profunditate} - I_i^{profunditate}$  și  $\Delta_x$  și  $\Delta_y$  sunt derivatele spațiale în raport cu axa  $x$  și axa  $y$ . Ca urmare, funcția globală de pierdere a adâncimii este definită sub forma (3):

$$L_{depth} = L_1^{Depth} + \mu L_{grad} \quad (3.3)$$

unde  $\mu$  este un coeficient de ponderare stabilit la 0,001.

Au fost propuse diferite funcții de pierdere pentru a antrena DNN-urile pentru debirozare a imaginilor. Funcțiile de pierdere a conținutului în funcție de pixel, cum ar fi  $L1$  și  $L2$  normă, sunt cele mai comune. Pentru antrenarea 2HDED:NET, au fost testate norma  $L1$  și funcția de pierdere Charbonnier [3], care este versiunea netezită a lui  $L1$ . Pierderea Charbonnier este calculată ca o eroare pătratică între imaginea deblurată estimată  $\hat{I}^{aif}$  și imaginea GT AiF  $I^{aif}$ :

$$L_{charb} = \frac{1}{n} \sum_{i=1}^W \sum_{j=1}^H \sqrt{(\hat{I}_{i,j}^{aif} - I_{i,j}^{aif})^2 + \epsilon^2} \quad (3.4)$$

unde  $\epsilon$  este un hiperparametru stabilit la  $1e-3$ . Acest hiperparametru acționează ca o pierdere pseudo-Huber și netezește erorile mai mici decât  $\epsilon$ .

Funcția de pierdere definită fie ca normă Charbonnier, fie ca normă  $L1$ , este îmbunătățită prin solicitarea unui SSIM ridicat. Acest lucru duce la adăugarea termenului de regularizare:

$$L_{SSIM} = 1 - SSIM(\hat{I}_{i,j}^{aif}, I_{i,j}^{aif}) \quad (3.5)$$

ceea ce face ca funcția de pierdere completă de deblurare să fie:

$$L_{deblur} = L_{charb} + \Psi L_{SSIM} \quad (3.6)$$

unde  $\Psi$  este o pondere stabilită la 4.

Cu pierderile de adâncime și de deblurare definite ca în ecuația 3.3 și 3.6, pierderea totală pentru formarea 2HDED:NET este următoarea:

$$L_{2HDED} = L_{depth} + \lambda L_{deblur} \quad (3.7)$$

## 3.2 Rezultate experimentale

Imaginile defocalizate sintetic au fost folosite în multe lucrări recente [1, 2] dedicate fie inferenței de adâncime, fie restaurării imaginilor. Pentru experimente, folosim reperele NYU-Depth-V2, Make3D, DFD și DIOD. Tabelul 3.1 prezintă, în jumătatea stângă, rezultatele pentru estimarea adâncimii obținute cu rețele antrenate pe setul de date NYU-Depth-V2. Pentru setul de date NYU-Depth-V2, cea mai bună acuratețe în termeni de RMSE este obținută de Carvalho et al. [2]. Din aceeași categorie de rețele care utilizează imagini defocalizate, există [1]. În medie, acuratețea hărților de adâncime a [1] este mai slabă cu 0,2 în RMSE, comparativ cu cel mai bun rezultat obținut de [2]. 2HDED:NET se află la jumătatea distanței dintre [2] și [1] cu un RMSE de 0,244. În categoria rețelelor care gestionează atât imagini de adâncime, cât și imagini deblurate, 2HDED:NET este cea mai bună în toate parametrii.

După cum a fost deja afirmat de [1, 2], utilizarea imaginilor defocalizate, spre deosebire de imaginile AiF, duce la o îmbunătățire semnificativă a preciziei estimării adâncimii. Rezultatele calitative pe setul de date NYU-Depth-V2 pentru ambele cazuri, AiF și out-of-focus, sunt prezentate în figura 3.2. În toate exemplele, îmbunătățirea este evidentă atunci când se utilizează imagini nefocalizate pentru instruire. În cazul

Table 3.1 Compararea 2HDED:Net cu metodele SoA pentru estimarea adâncimii și deblurarea imaginilor pe seturile de date NYU-Depth-V2.

Method			Depth Estimation		Deblurring	
	Depth	Deblur	RMSE ↓	Abs. rel ↓	PSNR ↑	SSIM ↑
Carvalho et al. [2]	✓	×	<b>0.144</b>	0.036	–	–
Anwar et al. [1]	✓	✓	0.347	0.094	34.21	–
2HDED:Net	✓	✓	0.244	<b>0.029</b>	<b>34.85</b>	0.99

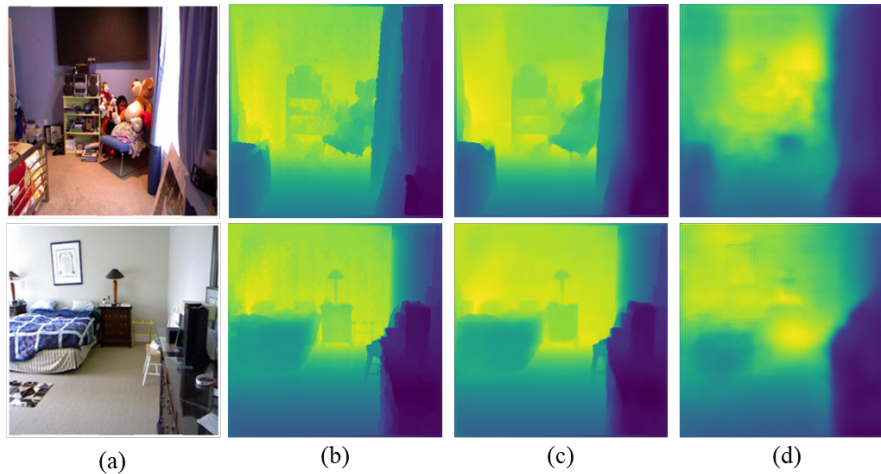


Figure 3.2 Exemple de estimare a adâncimii cu 2HDED:NET (cap de deblorare eliminat) din testele de referință NYU-Depth-V2 după antrenarea cu imagini defocalizate sau AiF. (a) Imagini AiF RGB, (b) Adâncimea GT, (c) Adâncimea estimată cu imagini defocalizate, (d) Adâncimea estimată cu imagini AiF.

imaginilor AiF, deși 2HDED:NET reușește să estimeze regiunile apropiate și îndepărtate, conținutul scenei rămâne indiscernabil. Conținutul devine evident doar atunci când hărțile de adâncime sunt estimate din imagini defocalizate.

În evaluarea 2HDED:NET, un accent deosebit este pus pe compararea acesteia cu rețeaua propusă de Anwar et al. [1]. Deși ambele rețele furnizează hărți de adâncime și imagini deblurate, rețeaua lui Anwar utilizează o abordare de procesare în conductă, ceea ce o face un punct de comparație adecvat cu rețeaua noastră.

În Figura 3.3, sunt prezentate rezultatele de deblorare a imaginilor pe setul de date NYU-Depth-V2, restaurate atât de [1] cât și de 2HDED:NET. Metoda noastră atinge un PSNR de 34,85 dB, care este cu aproape 0,7 dB mai mare decât cel al lui [1] pentru aceeași imagine. Eliminarea neclarității poate fi bine observată în zonele delimitate de dreptunghiurile roșii: lumina de pe tavan și marginile mobilierului.

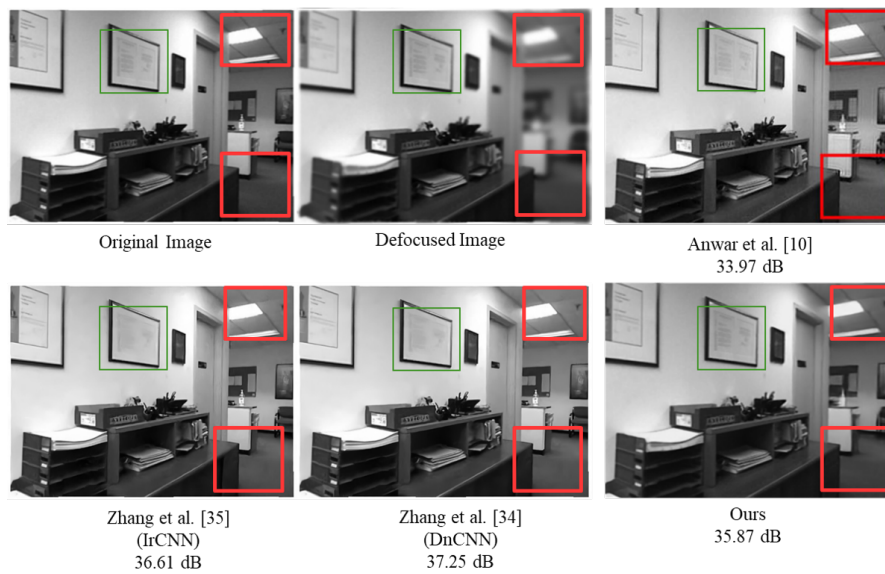


Figure 3.3 Compararea 2HDED:NET cu soluția de tip pipeline a lui Anwar et al. [1], și două metode generale pentru restaurarea imaginilor [21, 22]: un exemplu din setul de date NYU-Depth-V2.

# Chapter 4

## iDFD: un set de date pentru estimarea adâncimii și defocalizarea

În acest capitol, se propune un set de date pentru estimarea adâncimii și debirocratizarea defocalizării pe baza imaginilor defocalizate în mod natural. Setul de date iDFD propus în acest capitol este un set de date la scară largă cu 764 triple de imagini constând din imagini AiF, omologii defocalizate și hărțile de adâncime corespunzătoare.

În ultimii ani, au apărut numeroase tehnici bazate pe DL pentru MDE. Dintre acestea, metodele supravegheate au demonstrat în mod constant cele mai impresionante rezultate. Cu toate acestea, succesul metodelor DL supravegheate depinde de existența unor repere de instruire mari și diverse, cum ar fi KITTI, Cityscapes sau Make3D, care conțin imagini RGB și GT-ul lor corespunzător, constând fie în hărți de adâncime dense, fie în nori de puncte.

### 4.1 Setul de date iDFD pentru DFD și Deblurring imagini

În capitolul anterior, este propusă o rețea MTL numită 2HDED:NET [14], în care DFD și deblurring se sprijină reciproc pentru a îndeplini mai bine aceste sarcini. 2HDED:NET este antrenată pe un set de date defocalizate sintetic, fără a confirma eficiența sa pe imagini defocalizate natural din cauza lipsei unui astfel de set de date [14].

Un set de date consistent cu imagini defocalizate, adnotate pentru adâncime și defocalizare, ar stimula metodele DL supravegheate pentru MDE și debirocratizarea imaginilor și ar îmbunătăți acuratețea acestora. Pentru a umple această lacună, este introdus iDFD, un set de date adnotat pentru adâncime și defocalizare, cu imagini defocalizate în mod natural și adâncime și AiF GT corespunzătoare. iDFD este un set de date la scară largă cu diverse scene de interior. Setul de date a fost achiziționat cu o cameră MS-Kinect și o cameră Digital Single-Lens Reflex (DSLR) strâns cuplate între ele. Camera DSLR a fost adoptată pentru a capta imaginile defocalizate în mod natural,



Table 4.1 iDFD Prezentare generală.

Sensors	Scenes	Range	Data	Total no. Images	Depth Maps	Out-of-Focus DSLR Settings
DLSR-Nikon/MS-KINECT	Indoor	0 – 10m	RGB/Depth	764	Raw/In-painted	Lens: 14mm $f/2.8$

deoarece camera RGB a MS-Kinect nu are posibilitatea de a lucra cu diferite diafragme. Tabelul 4.1 prezintă o prezentare generală a setului de date iDFD. Setul de date conține o varietate de scene de interior capturate în 5 medii diferite: dormitor, cameră de zi, laboratoare, bucătărie și muzeu. În total, sunt capturate 764 scene, majoritatea provenind din laboratoare. Figura 4.1b arată distribuția scenelor în mediile de 5. Unele scene, cum ar fi cele din dormitoare și camere de zi, sunt mult mai simple ca și conținut decât scenele provenite din laboratoare și muzee. Scopul a fost de a achiziționa trei imagini pentru fiecare scenă: defocalizată, AiF și de adâncime.

O cameră DSLR-Nikon cu un obiectiv de 14mm la două diafragme diferite este utilizată pentru a capta imagini AiF și defocalizate. Pentru imaginile AiF se folosește o diafragmă îngustă cu  $f/10$ , iar pentru cele defocalizate o diafragmă mai largă cu  $f/2,8$ . Deschiderea largă creează o adâncime de câmp puțin adâncă, rezultând o imagine neclară.

Pentru a achiziționa adâncimea GT, se utilizează camera MS-Kinect în modul de adâncime NFOV (Narrow Field-of-View), cu un interval de operare de 0,5 - 3,8 m și un câmp de interes (FoI) maxim de  $75^\circ$  ori  $65^\circ$ . Adâncimea poate fi furnizată și în afara intervalului de operare, dar aceasta depinde în întregime de reflectivitatea obiectelor. În selectarea scenelor, au fost evitate zonele în care informațiile de adâncime ar putea să nu fie captate cu acuratețe, cum ar fi umbrele. Configurația experimentală este prezentată în figura 4.1a. Camera Kinect a fost montată pe un DSLR-Nikon cu ajutorul unui adaptor cu sabot, iar configurația noastră experimentală este similară cu cea a lui Carvalho et al. [2]. Pentru preprocesarea datelor, a fost urmată procedura lui Qiu et al. [16], care constă în denoising, inpainting (opțional), înregistrarea imaginii, normalizare și decupare.

## 4.2 Rezultate experimentale

În această secțiune, setul de date iDFD este evaluat pe 2HDED:NET propus în capitolul anterior. Pentru a sublinia importanța utilizării de date reale, rețeaua este reeducată în aceleași condiții pe setul de date NYU-Depth-V2, completat cu imagini defocalizate sintetice. Rezultatele obținute pe cele două repere sunt comparate în ceea ce privește acuratețea adâncimii și deblurring-ul.

Tabelul 4.2 prezintă erorile de estimare a adâncimii pentru ambele repere NYU-Depth-V2 și iDFD. Atunci când se efectuează instruirea pe setul de date NYU-Depth-V2, RMSE este mai mic decât cel al adâncimilor brute din setul de date iDFD. Cifrele arată

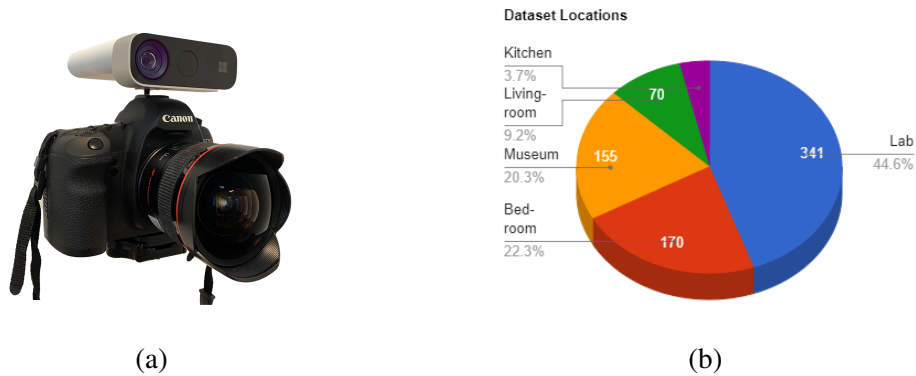


Figure 4.1 (a). Configurație experimentală: Camera DSLR-Nikon este cuplată la camera MS-Kinect cu ajutorul unui adaptor pentru montare pe sabot. Cele două camere sunt fixate pe un trepied pentru a evita mișcarea. DSLR-ul este utilizat pentru a capta imagini AiF și Defocus (RGB), iar Kinect este utilizat pentru a capta adâncimea GT pentru imaginile RGB. (b). Locațiile de achiziție a datelor și numărul total de imagini pe locație.

o precizie mai bună atunci când se utilizează adâncimea pictată. Aceste rezultate trebuie luate sub rezerva unui GT care nu este măsurat în proporție de 100%. Cele mai slabe rezultate sunt obținute pentru rețeaua antrenată pe NYU-Depth-V2 și testată pe setul de date iDFD. Figurile 4.2 și 4.3 prezintă trei exemple de hărți de adâncime obținute după antrenarea cu adâncimea brută și, respectiv, cu adâncimea pictată. Inspectia vizuală arată că, atunci când se utilizează imagini brute, rețeaua tinde să șteargă regiuni mici, cum ar fi spațiile goale de pe rafturi din cel de-al doilea exemplu. Aceste detalii sunt păstrate de hărțile de adâncime obținute după instruirea pe adâncimea pictată. Acest lucru sugerează că vopsirea în interior ar trebui efectuată înainte de instruire, apoi după instruirea cu adâncimea brută.

Tabelul 4.3 ilustrează rezultatele de deblurring ale 2HDED:NET pe reperele NYU-Depth-V2 și iDFD. 2HDED:NET a arătat rezultate promițătoare pentru deblurring-ul imaginilor defocalizate sintetic din NYU-Depth-V2, unde PSNR-ul a fost îmbunătățit de la 26,09 dB la 32,68 dB (Tabelul 4.3). Pe imaginile defocalizate natural iDFD, 2HDED:NET este capabil să îmbunătățească PSNR de la 25,83 dB la 36,25 dB, ceea ce înseamnă un câștig de 10,43 dB, semnificativ mai mare decât cel de 6,59 dB obținut în

Table 4.2 Estimarea profunzimii de către 2HDED:NET pe reperele NYU-Depth-V2 și iDFD..

Training/Testing dataset	$RMSE \downarrow$	$Abs. rel \downarrow$	$\delta 1 \uparrow$	$\delta 2 \uparrow$	$\delta 3 \uparrow$
NYU-Depth-V2/NYU-Depth-V2	0.281	0.266	0.877	0.942	0.958
NYU-Depth-V2/iDFD	0.401	0.284	0.759	0.813	0.839
iDFD/iDFD (raw depth)	0.312	0.194	0.727	0.793	0.807
iDFD/iDFD (in-painted depth)	0.248	0.145	0.799	0.854	0.959

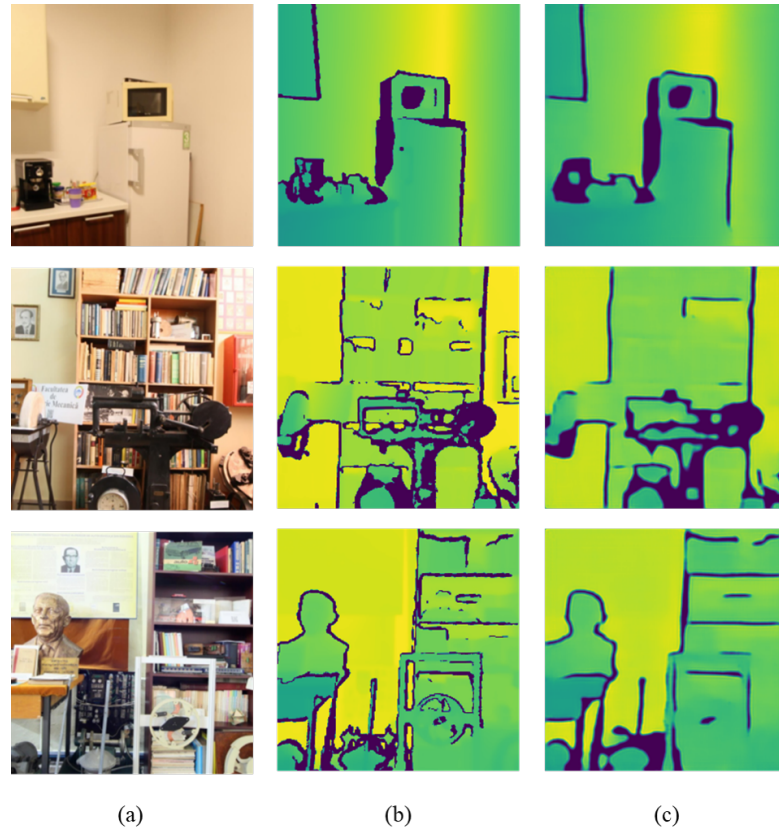


Figure 4.2 Rezultate cu 2HDED:NET antrenat pe adâncimea brută: a) Imagine RGB, b) Adâncimea brută GT, c) Adâncimea estimată.

cazul NYU-Depth-V2. PSNR-ul inițial de 25,83 dB a fost calculat luând ca referință imaginea AiF obținută cu deschiderea  $f/10$ .

În cele din urmă, rețeaua antrenată pe NYU-Depth-V2 a fost testată pe iDFD. Deblurring-ul a fost mai modest, cu un PSNR final de numai 30,09 dB.

Acest rezultat arată că o rețea antrenată pe un set de date cu defocalizare naturală și imagini RGB de înaltă calitate poate fi mai eficientă pentru debirocratizarea imaginilor decât aceeași rețea antrenată pe un set de date defocalizat sintetic de mari dimensiuni, cum ar fi cel obținut din NYU-Depth-V2.

Table 4.3 Rezultatele de debirocratizării obținute cu 2HDED:NET antrenat sau cu reperle iDFD și NYU-Depth-V2.

Dataset	PSNR $\uparrow$	SSIM $\uparrow$
NYU-Depth-V2 after deblurring (training on NYU-Depth-V2)	32.68	0.91
NYU-Depth-V2 defocused	26.09	0.49
Gain	6.59	0.42
iDFD after deblurring (training on iDFD)	36.25	0.99
iDFD defocused	25.83	0.51
Gain	10.43	0.43
iDFD after deblurring (training on NYU-Depth-V2)	30.09	0.86

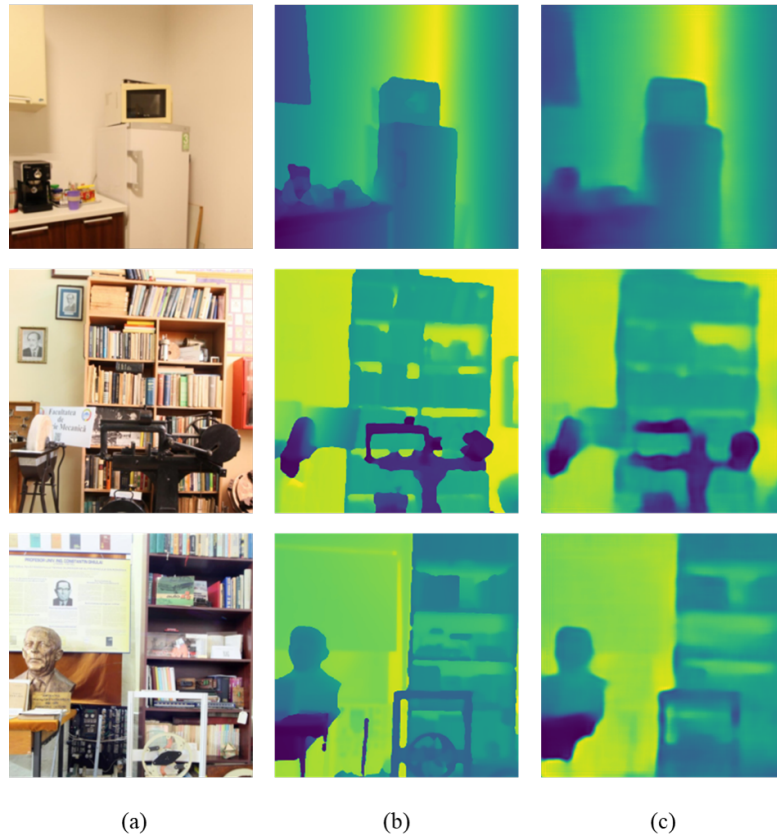


Figure 4.3 Rezultate cu 2HDED:NET antrenat pe adâncimea pictată: a) imagine RGB, b) adâncime GT, c) adâncime estimată.

Figura 4.4 prezintă două imagini cu un conținut complex, restaurate de 2HDED:NET antrenat pe setul de date iDFD. Culturile reprezentând microscopul din prima imagine sau perna și tabloul din cea de-a doua imagine, sunt mărite pentru a arăta cum detaliile se scufundă în neclaritate și apar după restaurarea cu 2HDED:NET.

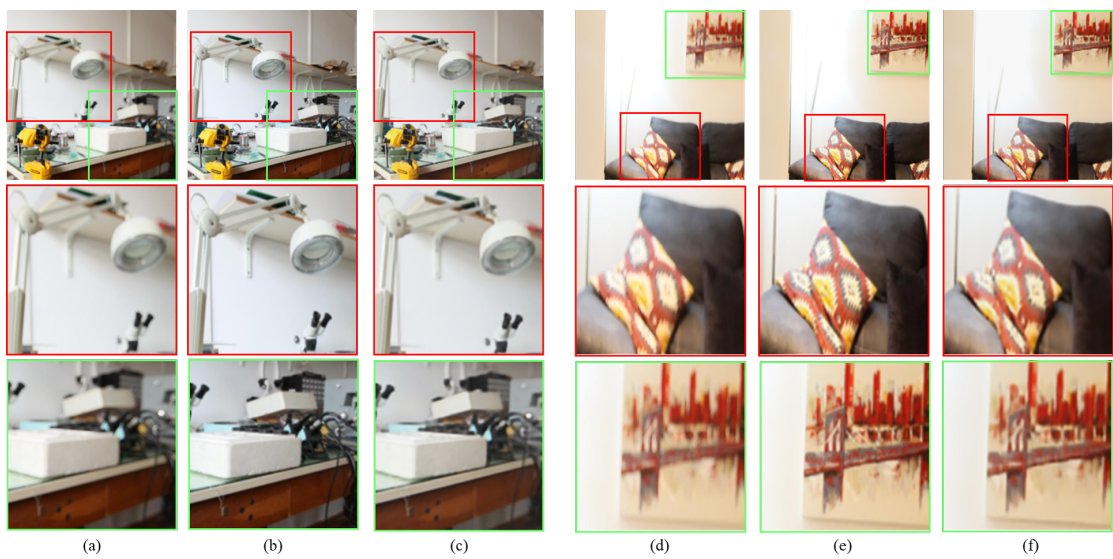


Figure 4.4 Rezultatele 2HDED:NET pentru debirozare a imaginii: de la stânga la dreapta în primul rând a) Defocalizare (intrare), b) AiF (GT) și c) Debirozare. În mod similar, d) Defocus (intrare), e) AiF (GT) și f) Deblurred. Al doilea rând prezintă regiunile mărite pentru a ilustra rezultatele debirografierii. Două exemple sunt luate din laboratorul de electronică și din camera de zi.

## Chapter 5

# Învățare auto-supravegheată pentru estimarea hărții defocalizate

În acest capitol, se propune un DNN auto-supervizat de la un capăt la altul, conceput pentru estimarea hărții de defocalizare (DME) dintr-o singură imagine RGB defocalizată. Metoda propusă este construită pe bazele 2HDED:NET introdusă în capitolul 3, dar rețeaua este îmbunătățită cu un modul de simulare a defocalizării care permite instruirea autosupravegheată pentru DME. În plus față de harta de defocalizare, rețeaua noastră reconstruiește, de asemenea, imaginea AiF utilizând învățarea supravegheată. Rețeaua este testată pe criterii de referință sintetice și reale, demonstrând eficiența sa pentru DME și debirocratizarea imaginii atunci când este disponibilă o singură imagine defocalizată.

### 5.1 Autosupravegheat 2HDED:NET

Întregul proces de instruire este rezumat în figura. 5.1. 2HDED:NET este o rețea DNN MTL propusă în capitolul 3, rețeaua este concepută pentru a rezolva problema DFD și a reconstrui o imagine AiF dintr-o singură imagine defocalizată într-o manieră supravegheată.

În acest capitol, utilizarea neclarității de defocalizare este valorificată. Abordarea implică antrenarea 2HDED:NET ca o rețea autosupravegheată. Pentru a realiza acest lucru, 2HDED:NET este completată cu un modul de simulare a defocalizării, care permite generarea unei imagini reblurate în timpul antrenamentului. Această imagine reblurată specifică este derivată din harta de defocalizare estimată și din imaginea GT AiF. Pentru a menține consecvența, aceiași parametri utilizați la generarea seturilor de date de formare sunt utilizați la reblurirea imaginilor în timpul formării.

Pentru a antrena 2HDED:NET autosupravegheat, se utilizează o funcție de pierdere similară cu cea utilizată în [7], care încorporează norma  $L1$  pentru a evalua diferența absolută dintre imaginea RGB reblurată sintetic  $I'$  și imaginea defocalizată de intrare  $I^b$ .

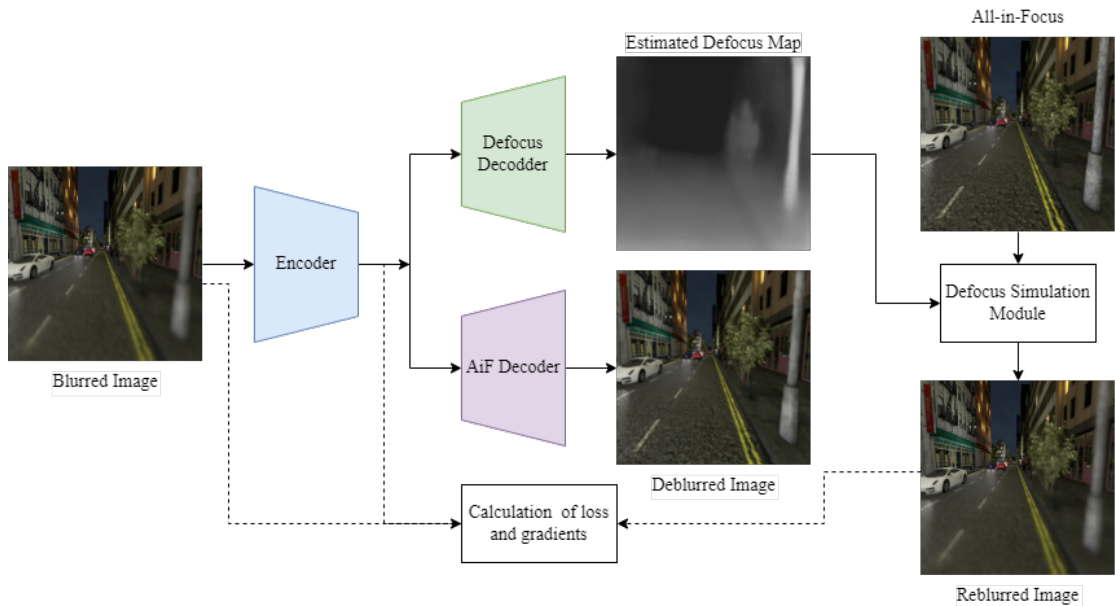


Figure 5.1 Arhitectura 2HDED:NET constă dintr-un codificator și două decodificatoare care funcționează în paralel. Ramura superioară estimează harta de defocalizare, iar cea inferioară imaginea AiF. Rețeaua este alimentată cu o imagine RGB defocalizată. Modulul de simulare a defocalizării reface imaginile cu ajutorul hărții de defocalizare estimate și a imaginilor AiF în timpul antrenamentului.

Funcția de pierdere este:

$$L_{re} = \frac{1}{n} \sum_{t=1}^n \frac{\alpha}{2} (1 - SSIM(I^r - I^b)) + (1 - \alpha) (|I^r - I^b|) \quad (5.1)$$

unde  $|\cdot|$  reprezintă norma  $L1$ , iar SSIM este indicele de similaritate structurală.  $\alpha = 0,85$  este o constantă.

Pentru a preveni schimbările drastice ale hărții de defocalizare prezise în regiunile omogene, în funcția de pierdere se introduce constrângerea de regularizare a netezimii. Constrângeri similare de regularizare a netezimii au fost utilizate în aproape toate metodele de estimare a hărților de adâncime supravegheate sau autosupravegheate [17, 7, 12]. Această constrângere sporește, de asemenea, coerența dintre AiF și harta de defocalizare estimată. Funcția de pierdere este:

$$L_{smooth} = \frac{1}{n} \sum_{t=1}^n |\delta_x I^d| e^{-\beta \delta_x I^{AiF}} + |\delta_y I^d| e^{-\beta \delta_y I^{AiF}} \quad (5.2)$$

unde  $I^d$  este harta de defocalizare estimată,  $I^{AiF}$  este imaginea GT AiF, iar  $\delta$  este gradientul calculat pe axa x și respectiv y.  $\beta$  setat la 2,5 este factorul de scară pentru sensibilitatea marginilor [17].

Au fost propuse diferite funcții de pierdere pentru a antrena DNN-urile pentru debirozarea imaginilor. Pentru a antrena 2HDED:NET, în capitolul 3 se utilizează funcția de pierdere Charbonnier [15], care este versiunea netezită a lui  $L1$ . Pierderea

Table 5.1 Analiză cantitativă pentru estimarea hărților de defocalizare și debirozare a imaginilor pe diferite seturi de date, cele mai bune rezultate sunt prezentate cu litere îngroșate.

	Trained on DFD & and tested on RTF dataset				Trained & tested on SYNDOF dataset	Trained & tested on iDFD dataset
	[4]	[10]	[11]	Ours	Ours	Ours
Evaluarea estimării hărților de defocalizare în termeni de MSE și MAE						
MSE	<b>0.094</b>	0.172	0.098	0.205	0.374	0.350
MAE	<b>0.196</b>	0.311	0.239	0.358	0.509	0.457
Comparație de debirozare a imaginii în termeni de PSNR și SSIM						
PSNR	25.42	24.03	<b>26.08</b>	25.08	31.74	34.08
SSIM	<b>0.850</b>	0.761	0.823	0.797	0.812	0.908

Charbonnier este calculată ca eroare pătratică între imaginea deblurată estimată  $\hat{I}^{AiF}$  și imaginea GT AiF  $I^{AiF}$ :

$$L_{deblur} = \frac{1}{n} \sum_{i=1}^W \sum_{j=1}^H \sqrt{(\hat{I}_{i,j}^{AiF} - I_{i,j}^{AiF})^2 + \varepsilon^2} \quad (5.3)$$

unde  $\varepsilon$  este un hiperparametru stabilit la  $1e-3$ . Acest hiperparametru acționează ca o pierdere pseudo-Huber și netezește erorile mai mici decât  $\varepsilon$ .

Prin urmare, pierderea totală  $L_{total}$  devine o sumă ponderată a ambelor pierderi, după cum urmează:

$$L_{total} = \lambda_1 L_{re} + \lambda_2 L_{smooth} + \lambda_3 L_{deblur} \quad (5.4)$$

Aici, ponderile sunt stabilite ca fiind  $\lambda_1 = 2$ ,  $\lambda_2 = 0,5$  și  $\lambda_3 = 0,01$ .

## 5.2 Rezultate experimentale

Comparația cantitativă a abordării noastre DME este prezentată în jumătatea superioară a tabelului 5.1.

În cazul setului de date DED, calcularea erorilor după testele efectuate pe acest set de date nu a putut fi efectuată din cauza indisponibilității hărților de defocalizare GT și a imaginilor AiF GT. Cu toate acestea, rezultatele vizuale sunt prezentate în figura 5.2. Metoda noastră autosupravegheată reușește să recupereze o cantitate destul de mare de caracteristici, estimările fiind mai apropiate de hărțile GT defocus, în special în comparație cu [4].

Pentru defocalizarea defocalizării, rezultatele comparative sunt furnizate pentru setul de date RTF în jumătatea inferioară a tabelului 5.1. Parametrii de evaluare utilizați includ PSNR și SSIM. Metodele din [4, 11] dau rezultate mai bune, dar metoda noastră este



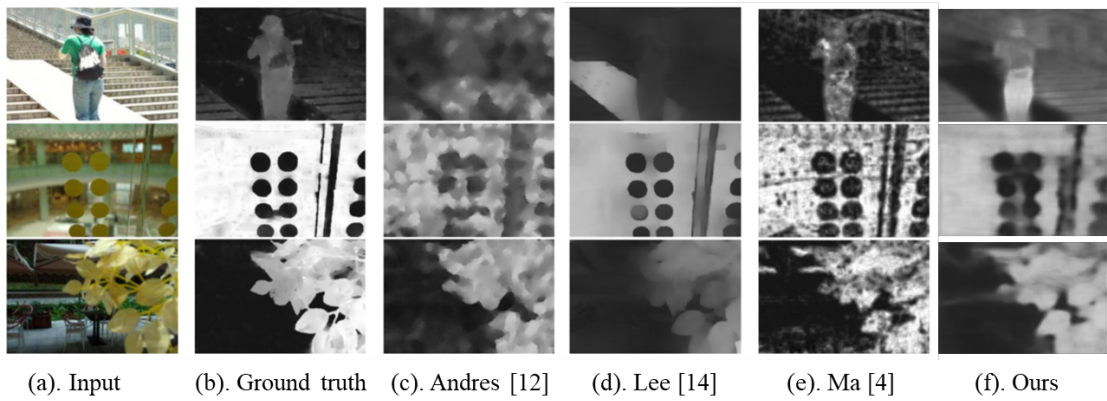


Figure 5.2 Compararea estimării hărților de defocalizare cu [4], [10] și [11] pe setul de date DED.

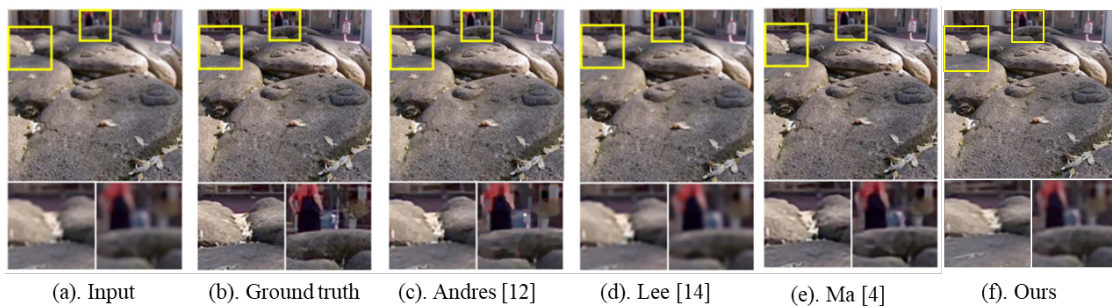


Figure 5.3 Comparație cu [4], [10] și [11] pe RTF în timp ce a fost antrenat pe setul de date DED.

mai precisă decât cea din [10]. În figura 5.3 este prezentat un exemplu de deblurare a defocalizării pe setul de date RTF. Imaginile din setul de date RTF sunt decupate și mărite în zone mici pentru o mai bună vizualizare și comparație. În timp ce [11] a obținut rezultate impresionante în debirocratizarea imaginilor, metoda noastră depășește [10] în ceea ce privește păstrarea detaliilor fine.

# Chapter 6

## Concluzii

În această teză, am dezvoltat o abordare nouă pentru estimarea adâncimii dintr-o singură imagine defocalizată folosind DNN, explorând neclaritatea defocalizării ca indiciu de adâncime. Inițial, am efectuat o analiză amănunțită a DNN-urilor în MDE, cuprinzând diverse arhitecturi profunde, modele preinstruite și funcții de pierdere. Pentru a îmbunătăți calitatea hărților de profunzime, am studiat mai multe soluții pentru regularizarea de netezire utilizată în instruirea rețelei. Am comparat eficacitatea acestora atât în cazul DNN-urilor supravegheate, cât și în cazul celor autosupravegheate. În plus, am adaptat termenul de netezire într-o manieră autosupravegheată pentru a funcționa într-o abordare supravegheată. Pentru experimente, am folosit ca suport un DNN supravegheat bazat pe o rețea Encoder-Decoder. Cu ajutorul experimentelor noastre, arătăm că metodele care utilizează termeni de regularizare bazați pe norma L1 obțin cea mai bună acurătate. În loc să ne bazăm pe metrici de evaluare comune, calitatea marginilor a fost măsurată cu ajutorul unor metrici dedicate constând în distanța Kullback-Leibler aplicată pe descompunerea piramidei directe a hărților de adâncime. Investigațiile noastre în acest sens s-au bazat pe imagini AiF pentru instruire.

Ulterior, am exploatat potențialul defocalizării neclare, o caracteristică inerentă a fenomen inerent în imaginile realizate cu o cameră cu obiectiv, și am dezvoltat un DNN multitask numit 2HDED:NET, pentru estimarea adâncimii și deblurirea imaginii. Acest model nou este construit pe baza arhitecturii codificator-decodificator, prin grefarea unui al doilea decodor, astfel încât să realizeze în paralel aceste două sarcini. În comparație cu metodele anterioare care se bazează pe arhitecturi de tip pipeline cu hărți de adâncime ca rezultate intermediare, 2HDED:NET a demonstrat performanțe bune, producând hărți de adâncime precise și recuperând imagini AiF din imagini unice defocalizate. Am validat pe larg modelul propus pe seturi de date de referință atât în interior, cât și în exterior, demonstrând superioritatea sa față de metoda de referință [1].

Există o limitare critică atunci când se lucrează cu neclarități sintetice în imaginile de intrare, așa cum este cazul pentru testele de referință existente care includ doar imagini AiF și adâncime GT. Această diferență de domeniu între seturile de date reale și sintetice poate duce la o scădere semnificativă a performanțelor DNN. Pentru a rezolva

această problemă, am introdus setul de date iDFD, care conține imagini defocalizate în mod natural împreună cu perechile AiF și hărțile de adâncime corespunzătoare. Experimentele noastre au demonstrat importanța instruirii și testării DNN-urilor pe scene naturale pentru a reduce această diferență de domeniu. În plus, am antrenat/testat diferite metode de estimare a adâncimii SoA pe setul de date iDFD și am comparat rezultatele cu 2HDED:NET.

Inițial, 2HDED:NET a fost operat într-o manieră supravegheată, necesitând hărți de adâncime GT și imagini AiF. Pentru a depăși provocarea de a obține un astfel de GT care necesită forță de muncă, am transformat 2HDED:NET într-o rețea de antrenament auto-supravegheată prin introducerea unui modul de simulare a defocalizării care regenerează imaginile re-neclarate sintetic. Această arhitectură nouă permite estimarea hărților defocalizate în absența adâncimii sau a defocalizării GT în timpul antrenamentului.

## 6.1 Contribuții originale

În această teză, am folosit estomparea defocalizării ca un indiciu pentru îmbunătățirea calității estimării adâncimii din imagini individuale. În acest scop, propunem o abordare nouă pentru a efectua estimarea adâncimii și deblocarea imaginii folosind o singură imagine nefocalizată ca intrare. Principalele contribuții ale acestei teze sunt:

- Am propus o nouă arhitectură bazată pe învățarea profundă numită 2HDED:NET cu două decodoare paralele care estimează adâncimea și recuperează imagini AiF dintr-o singură imagine nefocalizată. Arhitectura cu două capete distinge rețeaua noastră de metodele existente care utilizează procesarea pipeline. Paralelizarea sarcinilor reduce complexitatea rețelei, totul menținând în același timp performanțe de estimare în profunzime și de estompere comparabile cu abordările SoA. Cu experimente ample, arătăm că performanța metodei propuse este competitivă cu abordările SoA atât pentru estimarea adâncimii, cât și pentru deblocarea imaginii.
- Am propus un set de date iDFD, care este o colecție de scene de interior defocalizate în mod natural, care are noutatea de a fi suplimentată atât de adâncime, cât și de AiF GT. Testele pe iDFD cu rețeaua multi-task 2HDED:NET, care estimează simultan adâncimea și estompează imaginea, au dovedit că antrenarea unei rețele pe date reale, mai degrabă decât pe date simulate, cum ar fi NYU-Depth V2 defocalizate sintetic, este mult mai eficient. Acest set de date este util și pentru rețelele de antrenament dedicate deblurring-ului imaginilor.
- De asemenea, am propus o metodă nouă, care estimează împreună o hartă defocalizată și reconstruiește o imagine AiF dintr-o singură imagine defocalizată. Îndrumarea oferită de Modulul de simulare a defocalizării permite rețelei să estimeze harta de focalizare într-un mod auto-supravegheat, pe de altă parte, este

supravegheată decolorarea imaginii. Experimentele pe diferite seturi de date realiste și sintetice arată că metoda noastră autosupravegheată propusă obține rezultate promițătoare comparând metodele supravegheate de SoA atât pentru estimarea hărții defocalizate, cât și pentru sarcinile de defocalizare a imaginii defocalizate, atât din punct de vedere cantitativ, cât și calitativ.

## 6.2 List of original publications

### 6.2.1 Journal Paper

Saqib Nazir<sup>a</sup>, Lorenzo Vaquero<sup>b</sup>, Manuel Mucientes<sup>b</sup>, Víctor M. Brea<sup>b</sup>, Daniela Coltuc<sup>a</sup>. *Depth Estimation and Image Restoration by DL From Defo*. In IEEE Transactions on Computational Imaging(TCI), volume. 9, pages. 607-619, 2023. ISSN: 1051-4651. DOI: 10.1109/TCI.2023.3288335.

<sup>a</sup> Research Center for Spatial Information (CEO SpaceTech), University POLITEHNICA of Bucharest (UPB), Romania.

<sup>b</sup> Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Spain.

### 6.2.2 Conference Papers

Saqib Nazir<sup>a</sup>, Cristian Damian<sup>a</sup>, Daniela Coltuc<sup>a</sup>. *Self-supervised Defocus Map Estimation and Auxiliary Image Deblurring Given a Single Defocused Image*. In IEEE Digital Image Computing: Techniques and Applications (DICTA), volume. –, pages. –, 2023. ISSN: –. DOI: [–submitted](#).

<sup>a</sup> Research Center for Spatial Information (CEO SpaceTech), University POLITEHNICA of Bucharest (UPB), Romania.

Saqib Nazir<sup>a</sup>, Zhouyan Qiu<sup>b</sup>, Daniela Coltuc<sup>a</sup>, Joaquin Martinez-Sanchez<sup>b</sup>, Pedro Arias<sup>b</sup>. *iDFD: A Dataset Annotated for Depth and Defocus*. In Springer Cham, Scandinavian Conference on Image Analysis (SCIA), volume. 13885, pages. 6-19, 2023. ISSN: 978-3-031-31435-3. DOI: <https://doi.org/10.1007/978-3-031-31435-3-5>.

<sup>a</sup> Research Center for Spatial Information (CEO SpaceTech), University POLITEHNICA of Bucharest (UPB), Romania.

<sup>b</sup> Centro de Investigación en Tecnologías, Energía y Procesos Industriales (CINTECX), Universidade de Vigo, Applied Geotechnology Group, Vigo, Spain.

Saqib Nazir<sup>a</sup>, Lorenzo Vaquero<sup>b</sup>, Manuel Mucientes<sup>b</sup>, Víctor M. Brea<sup>b</sup>, Daniela Coltuc<sup>a</sup>. *2HDED:Net for Joint Depth Estimation and Image Deblurring from a Single Out-of-Focus Image*. In IEEE International Conference on Image Processing (ICIP), pp. 2006-2010, Bordeaux (France), 2022. ISSN: 1051-4651. DOI: 10.1109/ICIP46576.2022.9897352.

<sup>a</sup> Research Center for Spatial Information (CEO SpaceTech), University POLITEHNICA of Bucharest (UPB), Romania.

<sup>b</sup> Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Spain.

Saqib Nazir<sup>a</sup>, Daniela Coltuc<sup>a</sup>. *Edge-preserving Smoothing Regularization for Monocular Depth Estimation*. In Proceedings of the 26th International Conference on Automation and Computing (ICAC), pp. 1–6, Portsmouth (United Kingdom), 2021. ISSN: 1051-4651. DOI: 10.23919/ICAC50006.2021.9594153.

<sup>a</sup> Research center for spatial information CEO SpaceTech, University POLITEHNICA of Bucharest (UPB), Romania.

### 6.3 Perspective pentru evoluții ulterioare

În această teză, am propus 2HDED:NET, care beneficiază de estomparea nefocalizată pentru estimarea hărților de adâncime. Arată rezultate promițătoare pentru estomparea nefocalizată pentru DFD; cu toate acestea, metoda nu este capabilă să estimeze hărțile de adâncime în prezența înceteșării în mișcare sau a mișcării camerei. Prin urmare, munca noastră viitoare va investiga estimarea adâncimii din tremurul camerei sau neclaritatea mișcării. Șeful AiFD al 2HDED:NET servește scopului deblocării nefocalizate; cu toate acestea, este limitat în gestionarea zgomotului prezent în imaginile defocalizate. Lucrările noastre viitoare vor folosi, de asemenea, tehnici care pot aborda reducerea zgomotului în timpul declarării imaginii.

Pe de o parte, estomparea defocalizării este un indiciu important pentru estimarea adâncimii, dar pe de altă parte, degradează imaginea și afectează calitatea imaginii. Cu 2HDED:NET auto-supravegheat, am reușit să generăm hărțile de focalizare fără informații despre adâncime/defocalizare GT, dar ne purtăm interesul către restaurarea imaginii într-o manieră complet auto-supravegheată. Deoarece există multe cazuri în lumea practică, în care o pereche clară de imagini înceteșate nu este disponibilă, prin urmare, evoluțiile viitoare ar putea include și dezvoltarea unui antrenament auto-supravegheat atât pentru ieșirile AiF, cât și pentru hărțile defocus. O opțiune promițătoare ar fi folosirea acestei abordări ca primă etapă a unui proces de formare care este perfecționat și mai mult printr-o etapă de reglare fină supravegheată.

Evoluțiile ulterioare se concentrează, de asemenea, pe progresul către o abordare auto-supravegheată pentru estimarea profunzimii prin adăugarea de informații suplimentare, astfel încât să elimine incertitudinea de focalizare.

Setul de date iDFD propus în această teză este proiectat pentru estimarea adâncimii și deblocarea imaginii. Deși acest set de date este cuprinzător și poate servi ca o resursă valoroasă pentru formarea DNN-urilor, în prezent cuprinde scene exclusiv din medii interioare. Cu toate acestea, în scenariile din lumea reală, întâlnim scene provocatoare atât în medii interioare, cât și în aer liber. În consecință, potențialele cercetări viitoare în această teză implică extinderea setului de date iDFD pentru a include mediile exterioare. Pentru a capta adâncimea GT din scenele în aer liber este necesar un senzor specializat, cum ar fi LiDAR. Această extindere ar spori aplicabilitatea setului de date și ar stimula dezvoltarea modelelor DNN capabile să abordeze scene diverse și complexe din lumea reală. În plus, este necesar să se stabilească un set de date DFD real, care să cuprindă atât setările interioare, cât și cele exterioare, complet cu informații despre adâncimea GT. Acest pas este crucial pentru a oferi studii viitoare în domeniul DFD și deblurring imagini cu repere cantitative bazate pe date mai realiste.

# References

- [1] Anwar, S., Hayder, Z., and Porikli, F. (2021). Deblur and deep depth from single defocus image. *Machine vision and applications*, 32(1):1–13.
- [2] Carvalho, M., Le Saux, B., Trouvé-Peloux, P., Almansa, A., and Champagnat, F. (2018). Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks? In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0.
- [3] Charbonnier, P., Blanc-Feraud, L., Aubert, G., and Barlaud, M. (1994). Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, volume 2, pages 168–172. IEEE.
- [4] D’Andrès, L., Salvador, J., Kochale, A., and Süsstrunk, S. (2016). Non-parametric blur map regression for depth of field extension. *IEEE Transactions on Image Processing*, 25(4):1660–1673.
- [5] Eigen, D., Puhrsch, C., and Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27.
- [6] Godard, C., Mac Aodha, O., and Brostow, G. J. (2017). Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 270–279.
- [7] Godard, C., Mac Aodha, O., Firman, M., and Brostow, G. J. (2019). Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3828–3838.
- [8] Gur, S. and Wolf, L. (2019). Single image depth estimation trained via depth from defocus cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7683–7692.
- [9] Hu, J., Ozay, M., Zhang, Y., and Okatani, T. (2019). Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 1043–1051. IEEE.
- [10] Lee, J., Son, H., Rim, J., Cho, S., and Lee, S. (2021). Iterative filter adaptive network for single image defocus deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2034–2042.
- [11] Ma, H., Liu, S., Liao, Q., Zhang, J., and Xue, J.-H. (2021). Defocus image deblurring network with defocus map estimation as auxiliary task. *IEEE Transactions on Image Processing*, 31:216–226.

- [12] Nazir, S. and Coltuc, D. (2021). Edge-preserving smoothing regularization for monocular depth estimation. In *2021 26th International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE.
- [13] Nazir, S., Qiu, Z., Coltuc, D., Martínez-Sánchez, J., and Arias, P. (2023a). idfd: A dataset annotated for depth and defocus. In *Scandinavian Conference on Image Analysis*, pages 67–83. Springer.
- [14] Nazir, S., Vaquero, L., Mucientes, M., Brea, V. M., and Coltuc, D. (2022). 2hded: Net for joint depth estimation and image deblurring from a single out-of-focus image. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 2006–2010. IEEE.
- [15] Nazir, S., Vaquero, L., Mucientes, M., Brea, V. M., and Coltuc, D. (2023b). Depth estimation and image restoration by deep learning from defocused images. *arXiv preprint arXiv:2302.10730*.
- [16] Qiu, Z., Martínez-Sánchez, J., Brea, V. M., López, P., and Arias, P. (2022). Low-cost mobile mapping system solution for traffic sign segmentation using azure kinect. *International Journal of Applied Earth Observation and Geoinformation*, 112:102895.
- [17] Si, H., Zhao, B., Wang, D., Gao, Y., Chen, M., Wang, Z., and Li, X. (2023). Fully self-supervised depth estimation from defocus clue. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9140–9149.
- [18] Ummenhofer, B., Zhou, H., Uhrig, J., Mayer, N., Ilg, E., Dosovitskiy, A., and Brox, T. (2017). Demon: Depth and motion network for learning monocular stereo. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5038–5047.
- [19] Wang, Z. and Simoncelli, E. P. (2005). Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human vision and electronic imaging X*, volume 5666, pages 149–159. SPIE.
- [20] Xian, K., Zhang, J., Wang, O., Mai, L., Lin, Z., and Cao, Z. (2020). Structure-guided ranking loss for single image depth prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 611–620.
- [21] Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017a). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155.
- [22] Zhang, K., Zuo, W., Gu, S., and Zhang, L. (2017b). Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938.