VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

STEREO REKONSTRUKCE MRAČNA BODŮ POMOCÍ HLUBOKÝCH NEURONOVÝCH SÍTÍ

STEREO RECONSTRUCTION WITH DEEP NEURAL NETWORKS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE **Bc. RICHARD LETANEC**
AUTHOR

VEDOUCÍ PRÁCE **Ing. MICHAL ŠPANĚL, Ph.D.**
SUPERVISOR

BRNO 2024

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Zadání diplomové práce

156420

Ústav:	Ústav počítačové grafiky a multimédií (UPGM)
Student:	**Letanec Richard, Bc.**
Program:	Informační technologie a umělá inteligence
Specializace:	Strojové učení
Název:	**Stereo rekonstrukce mračna bodů pomocí hlubokých neuronových sítí**
Kategorie:	Zpracování obrazu
Akademický rok:	2023/24

Zadání:

1. Seznamte se s problematikou hlubokých neuronových sítí a jejich učení.
2. Zorientujte se v současných metodách stereo rekonstrukce, které využívají hlubokých neuronových sítí.
3. Vytvořte datovou sadu pro vlastní experimenty. Využít můžete i veřejně dostupná data.
4. Vyberte vhodné metody a navrhněte architekturu neuronové sítě pro danou úlohu.
5. Experimentujte s vaší implementací a případně navrhněte vlastní modifikace metod.
6. Porovnejte dosažené výsledky a diskutujte možnosti budoucího vývoje.
7. Vytvořte stručný plakát nebo krátké video prezentující vaši práci, její cíle a výsledky.

Literatura:

- Laga, Hamid & Jospin, Laurent & Boussaid, Farid & Bennamoun, Mohammed: A Survey on Deep Learning Techniques for Stereo-Based Depth Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, https://arxiv.org/abs/2006.02535.
- Huang, Zhengyu & Norris, Theodore & Wang, Panqu: ES-Net: An Efficient Stereo Matching Network, 2021, https://arxiv.org/pdf/2103.03922.pdf.

Podrobné závazné pokyny pro vypracování práce viz https://www.fit.vut.cz/study/theses/

Vedoucí práce:	**Španěl Michal, Ing., Ph.D.**
Vedoucí ústavu:	Černocký Jan, prof. Dr. Ing.
Datum zadání:	1.11.2023
Termín pro odevzdání:	17.5.2024
Datum schválení:	9.11.2023

Abstrakt

Cieľ tejto diplomovej práce je navrhnúť a natrénovať model neurónovej siete schopný odhadovať disparitnú mapu z dvojice obrázkov. Z odhadnutej disparitnej mapy bude následne možné vytvoriť hĺbkovú mapu a mračno bodov. Takýto proces sa nazýva stereo rekonštrukcia. Riešenie tejto úlohy pozostáva z dvoch krokov – výberu vhodnej dátovej sady a výberu vhodnej architektúry neurónovej siete.

V práci som porovnal dve architektúry neurónových sietí, ktoré som natrénoval na dátovej sade DrivingStereo, pozostával z párových obrázkov vyfotografovaných zo strechy auta a dotrénoval a vyhodnotil na dátovej sade KITTI 2015, pozostával z obrázkov rovnakého typu.

Ako prvú architektúru neurónovej siete som zvolil ES-Net, ktorý využíva prístup založený na sekvencii reziduálnych blokov a konvolučných vrstiev. Ako druhú architektúru som zvolil CREStereo, ktorá na predikciu disparitnej mapy využíva iteratívny prístup založený na rekurentných vrstvách. Vo všetkých porovnávacích testoch dosahuje lepšiu presnosť predikcie architektúra CREStereo.

Abstract

The aim of this thesis is to design and train a neural network model capable of estimating a disparity map from a pair of images. It will then be possible to create a depth map and point cloud from the estimated disparity map. Such a process is called stereo reconstruction. Solving this task consists of two steps – choosing a suitable dataset and choosing a suitable neural network architecture.

In my work, I compared two neural network architectures that I trained on the DrivingStereo dataset, consisting of paired images photographed from the roof of a car, and retrained and evaluated on the KITTI 2015 dataset, consisting of images of the same type.

As the first neural network architecture, I chose ES-Net, which uses an approach based on a sequence of residual blocks and convolutional layers. As the second architecture, I chose CREStereo, which uses an iterative approach based on recurrent layers to predict the disparity map. In all benchmark tests, the CREStereo architecture achieves better accuracy.

Kľúčové slová

stereo rekonštrukcia, disparitná mapa, hĺbková mapa, husté mračno bodov, hlboké neurónové siete, pytorch, ES-Net, CREStereo

Keywords

stereo reconstruction, disparity map, depth map, point cloud, deep neural networks, pytorch, ES-Net, CREStereo

Citácia

LETANEC, Richard. *Stereo rekonstrukce mračna bodů pomocí hlubokých neuronových sítí.* Brno, 2024. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Michal Španěl, Ph.D.

Stereo rekonstrukce mračna bodů pomocí hlubokých neuronových sítí

Prehlásenie

Prehlasujem, že som túto diplomovú prácu vypracoval samostatne pod vedením pána Ing. Michala Španěla Ph.D.. Uviedol som všetky literárne pramene, publikácie a ďalšie zdroje z ktorých som čerpal.

.....
Richard Letanec
17. mája 2024

Poďakovanie

Chcel by som poďakovať vedúcemu diplomovej práce pánovi Ing. Michalovi Španělovi Ph.D., za jeho pomoc a rady pri riešení tejto práce.

Obsah

Zoznam obrázkov

Kapitola 1

Úvod

Stereo rekonštrukcia je zložitý problém, ktorý sa dlhodobo rieši v oblasti počítačového videnia a spracovania obrazu. Stereo rekonštrukcia má aplikácie v rôznych oblastiach ako sú autonómne riadené autá, medicína, augmentovaná realita alebo 3D rozpoznávanie objektov [3, 16]. Keďže sa pri prevode scény z reálneho sveta na 2D plochu – fotografiu – stratí informácia o hĺbke, nie je jednoduché previesť 2D fotografiu naspäť do 3D priestoru. S rozmachom hlbokého učenia sa na riešenie tohoto problému začali využívať hlboké neurónové siete.

Cieľom tejto práce je navrhnúť a natrénovať hlbokú neurónovú sieť, ktorá by sa dala využiť na stereo rekonštrukciu z páru obrázkov. V rámci práce som sa rozhodol zvoliť a porovnať dve architektúry neurónových sietí. Prvou vybranou architektúrou je architektúra ES-Net, využívajúca prístup založený na reziduálnych blokoch a konvolučných vrstvách v rámci celého procesu odhadu disparitnej mapy [12]. Druhou vybranou architektúrou je architektúra CREStereo, ktorá okrem reziduálnych blokov a konvolučných vrstiev na extrakciu príznakov využíva aj attention bloky pri párovaní príznakov a iteratívny prístup s rekurentnými vrstvami na predikciu disparitných máp [18].

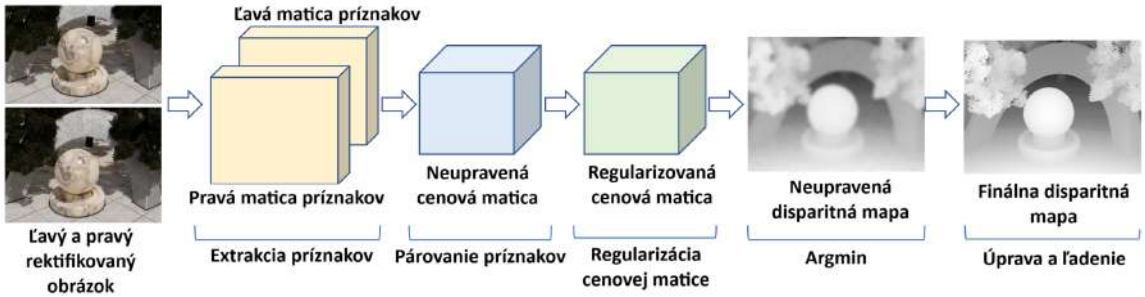
Modely som trénoval na dátovej sade DrivingStereo [42], pozostávajúcej z párov fotografií vyfotografovaných zo strechy auta. Modely som ďalej do-trénoval a vyhodnotil na dátovej sade KITTI 2015 [22]. Mnou natrénované modely dosiahli state-of-the-art úroveň presnosti na porovnávacích testoch KITTI. Model CREStereo dosiahol hodnoty chyby D1-all 2.16% a model ES-Net dosiahol hodnoty chyby D1-all 3.18%.

Táto technická správa je členená na kapitoly, pričom každá kapitola sa venuje inej oblasti riešenia daného problému. V kapitole 2 popisujem základné prístupy k stereo rekonštrukcií hĺbky z obrazu pomocou neurónových sietí. V kapitole 3 bližšie popisujem niektoré moderné state-of-the-art metódy hlbokého učenia, ktoré riešia problém stereo rekonštrukcie hĺbky z obrazu. V kapitole 4 píšem o najbežnejších dátových sadách, ktoré sa v dnešnej dobe využívajú na trénovanie a evaluáciu týchto metód. V kapitole 5 potom navrhujem svoje riešenie stereo rekonštrukcie hustého mračna bodov, ktoré vychádza z naštudovaných princípov a metód zhrnutých v kapitole 2, 3 a 4. V 6 popisujem implementačné výstupy mojej práce. V kapitole 7 popisujem experimenty a porovnávam moje výsledky trénovania modelov architektúr ES-Net a CREStereo. V kapitole 8 zhrňujem dosiahnuté výstupy a navrhujem možnosti ďalšieho postupu pri riešení úlohy stereo rekonštrukcie.

Kapitola 2

Stereo rekonštrukcia hĺbky z obrazu

Stereo rekonštrukcia hĺbky z obrazu predstavuje proces odhadovania jednej alebo viacerých hĺbkových máp z množiny dvojrozmerných snímok, ktoré zachytávajú tú istú trojrozmernú scénu. Tieto snímky môžu pochádzať z jednej kamery – monokulárna rekonštrukcia, dvoch kamier – stereo rekonštrukcia alebo viacerých kamier – multi-view stereo rekonštrukcia. Pri rekonštrukcií môžu byť vonkajšie a vnútorné parametre kamery buď známe alebo sa použijú ich predpokladané hodnoty. Výsledná hĺbková mapa môže byť odhadnutá z pôvodného pozorovacieho bodu alebo z novo-definovaného pozorovacieho bodu [16].



Obr. 2.1: Kroky pre stereo odhad hĺbky z obrázku. Prevzaté a preložené z [16].

Tradičné prístupy stereo rekonštrukcie hĺbky využívajú štyri kroky pre odhad hĺbky [26, 10, 4, 25]:

1. extrakcia príznakov
2. párovanie príznakov medzi obrázkami
3. výpočet mapy disparity
4. vyladenie mapy disparity

Tieto prístupy môžu v špecifických prípadoch dosahovať dobré výsledky, avšak nie, ak sa použijú na obrázky s bezpríznakovými oblasťami alebo oblasťami s opakujúcim sa vzorom.

V poslednom období sa na riešenie problému rekonštrukcie hĺbky z obrazu začali používať metódy hlbokého učenia. Tieto metódy sa vo všeobecnosti snažia minimalizovať chybovú funkciu 2.1 [16]:

$$\mathcal{L}(\mathbf{I}) = d(f_\theta(\mathbf{I}), D) \tag{2.1}$$

kde:

- $\mathcal{L}(\mathbf{I})$ je chybová funkcia
- $\mathbf{I}$ je množina obrázkov
- f_θ je prediktor, ktorý zo vstupných obrázkov zrekonštruuje disparitnú mapu
- $f_\theta(\mathbf{I})$ je zrekonštruovaná disparitná mapa
- D je skutočná disparitná mapa
- $d(.,.)$ je funkcia merajúca rozdiel medzi skutočnou a zrekonštruovanou disparitnou mapou

V tejto kapitole sa budem ďalej venovať prístupom k stereo rekonštrukcií hĺbky s využitím hlbokých neurónových sietí.

2.1 Stereo párovacie prístupy využívajúce neurónové siete

Metódy využívajúce stereo párovacie prístupy sa priamo učia párovať zhodné pixely medzi dvoma vstupnými obrázkami. Takéto párovanie sa potom dá previesť na disparitnú mapu alebo optický tok a následne ďalej na hĺbkovú mapu pre vstupný obrázok. Pri tomto prístupe sa prediktor f skladá z troch častí, ktoré sú trénované nezávisle na sebe [16]:

1. extraktor príznakov
2. párovač príznakov a agregátor ceny
3. odhadovač disparity/hĺbky

Výstupom takýchto metód je disparitná mapa D, ktorá minimalizuje funkciu energie $E(D)$ [16]:

$$E(D) = \sum_x C(x, d_x) + \sum_x \sum_{y \in \mathcal{N}_x} E_s(d_x, d_y) \tag{2.2}$$

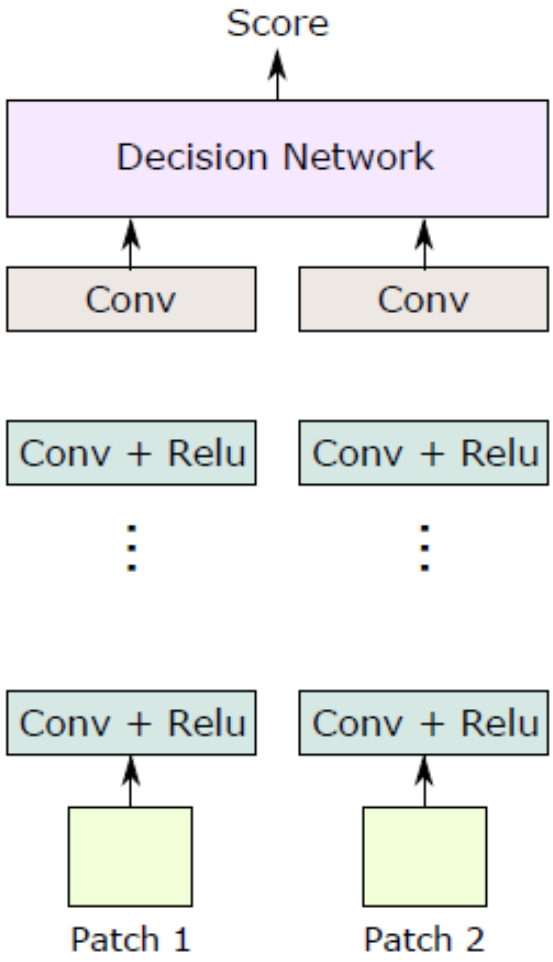
kde:

- x, y sú pixely v ľavom a pravom obrázku
- d_x je disparita na pixely x
- $\mathcal{N}_x$ je okolie pixelu x
- $\sum_x C(x, d_x)$ je párovacia cena, keď je rozsah disparity diskretizovaný do n_d úrovní disparity, vznikne z nej 3D cenová matica s rozmermi W x H x n_d, pričom W a H sú rozmery obrázku [16]
- $\sum_x \sum_{y \in \mathcal{N}_x} E_s(d_x, d_y)$ je regularizačný člen

Z disparitnej mapy sa nakoniec vypočíta hĺbka pomocou triangulácie s využitím vnútorných parametrov kamery, ktoré sú známe [16].

Extrakcia a párovanie príznakov

V základných typoch architektúr neurónových sietí sa vektory príznakov počítajú z výsekov z ľavého a pravého obrázku so stredom v bode x a y s využitím konvolučnej neurónovej siete. Následne sa vektory príznakov spárujú a vypočíta sa skóre podobnosti – $C(x, d)$ – s použitím buď štandardných metrík podobnosti L1, L2 a korelačných metrík alebo pomocou neurónovej siete. Jednotlivé komponenty sa pri tomto prístupe môžu trénovať buď samostatne alebo spolu [16]. Príkladom takejto architektúry je MC-CNN od Zbontara a LeCuna [46], zobrazená na obrázku 2.2.



Obr. 2.2: MC-CNN architektúra na učenie sa príznakov. Prevzaté z [16, 46].

Disparita a regularizácia

Po výpočte cenovej matice C je možné disparitu vypočítať z rovnice 2.2 zahodením regularizačného člena a aplikovaním funkcie argmin. Takáto disparitná mapa však môže byť zašumená a preto niektoré metódy využívajú regularizačný člen E_s, napríklad 2.3 [16]:

$$E_s(d_x, d_y) = \alpha_1 \delta(|d_{xy} = 1) + \alpha_2 \delta(|d_{xy} > 1) \tag{2.3}$$

kde:

- $d_{xy} = d_x - d_y$
- α_1, α_2 sú dopredu zvolené váhy také, že $\alpha_1 > \alpha_2 > 0$
- δ je Kroneckerovo delta

Výsledná disparita pre pixel x sa potom vypočíta pomocou rovnice 2.4 [16]:

$$d_x = \arg\min_d \sum_s E_s(x, d) \tag{2.4}$$

2.2 End-to-end prístupy

End-to-end metódy stereo rekonštrukcie hĺbky sa rozdeľujú na dva druhy. Tie, ktoré tento problém formulujú ako regresnú úlohu a tie, ktoré napodobňujú tradičné prístupy, kde jednotlivé kroky postupu rieši iná časť neurónovej siete.

Regresné metódy odvodzujú hĺbkovú mapu priamo zo vstupných obrázkov bez explicitnej extrakcie príznakov. Takéto metódy sú rýchle za behu, avšak vyžadujú obrovské množstvo trénovacích dát.

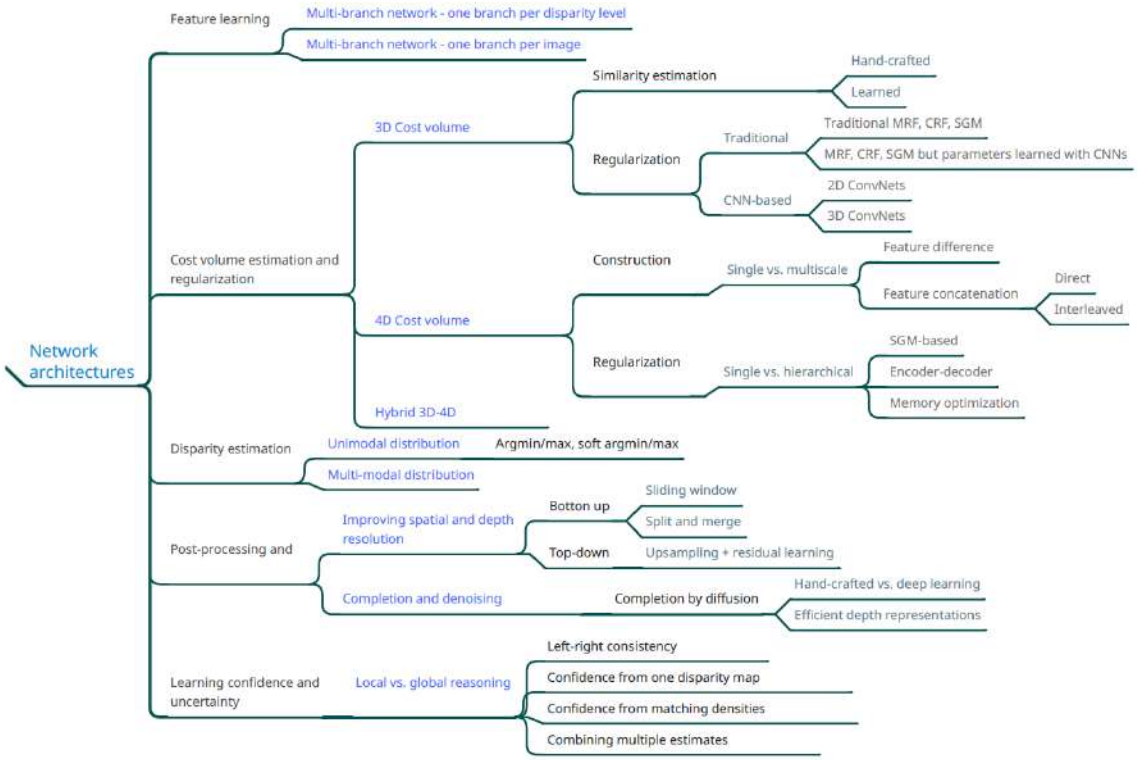
Metódy druhého typu opäť rozdeľujú problém na jednotlivé kroky [16]:

1. učenie sa príznakov
2. odhad cenovej matice a regularizácia
3. výpočet mapy disparity
4. vyladenie mapy disparity
5. učenie sa mapy spoľahlivosti

Na rozdiel od stereo párovacích prístupov, ktoré na výpočet príznakov používajú výseky zo vstupných obrázkov, end-to-end prístupy počítajú príznaky z celého vstupného obrázku naraz. Existujú dva spôsoby výpočtu príznakov, pričom oba využívajú viac-vetvové siete, kde pri prvom spôsobe každá vetve reprezentuje jeden vstupný obrázok a pri druhom každá vetva reprezentuje jednu úroveň disparity.

Po získaní príznakov nasleduje výpočet párovacieho skóre – cenovej matice. Cenová matica môže byť trojdimenzionálna, kde tretia dimenzia je úroveň disparity, štvordimenzionálna, kde tretí rozmer je dimenzia príznakov a štvrtý rozmer je úroveň disparity a nakoniec hybridná cenová matica, ktorá kombinuje vlastnosti predchádzajúcich dvoch typov [16].

Jednotlivé kroky a ich varianty pre odhad hĺbky end-to-end prístupmi je zobrazený na obrázku 2.3.



Obr. 2.3: Kroky end-to-end stereo odhadu hĺbky a ich rôzne varianty. Prevzaté a upravené z [16].

Kapitola 3

Moderné metódy odhadu hĺbky z obrazu

V tejto kapitole som popísal vybrané state-of-the-art end-to-end metódy hlbokého učenia na stereo rekonštrukciu hĺbky z obrazu. Keďže budem v tejto kapitole ukazovať výsledky popisovaných metód na porovnávacích testoch, uvediem najskôr metriky, ktoré tieto testy využívajú na porovnanie efektivity metód.

3.1 Metriky evaluácie odhadu disparitnej mapy používané v porovnávacích testoch

Porovnávacie testy od inštitúcií KITTI, Middlebury a ETH3D používajú niekoľko rôznych metrík na evaluáciu odhadu disparitnej mapy.

Základnou metrikou v porovnávacom teste KITTI je metrika D1, čo je percento odľahlých hodnôt v ľavej snímke. Za odľahlú hodnotu sa považuje pixel s koncovou chybou (EPE – end-point error) väčšou alebo rovnou ako 3 alebo väčšou alebo rovnou ako 5%, rovnica 3.1. KITTI počíta túto metriku zvlášť pre pixely na pozadí, zvlášť pre pixely v popredí a tiež aj pre všetky pixely na obrázku [22].

$$D1 = \frac{1}{P} \sum_{p}^{P} \begin{cases} 1 & \text{ak } err_p(d_{gt}, d_{pr}) >= 3 \text{ alebo } \frac{err_p(dgt_p, dpr_p)}{|dgt_p|} >= 0.05 \\ 0 & \text{inak} \end{cases} \tag{3.1}$$

kde:

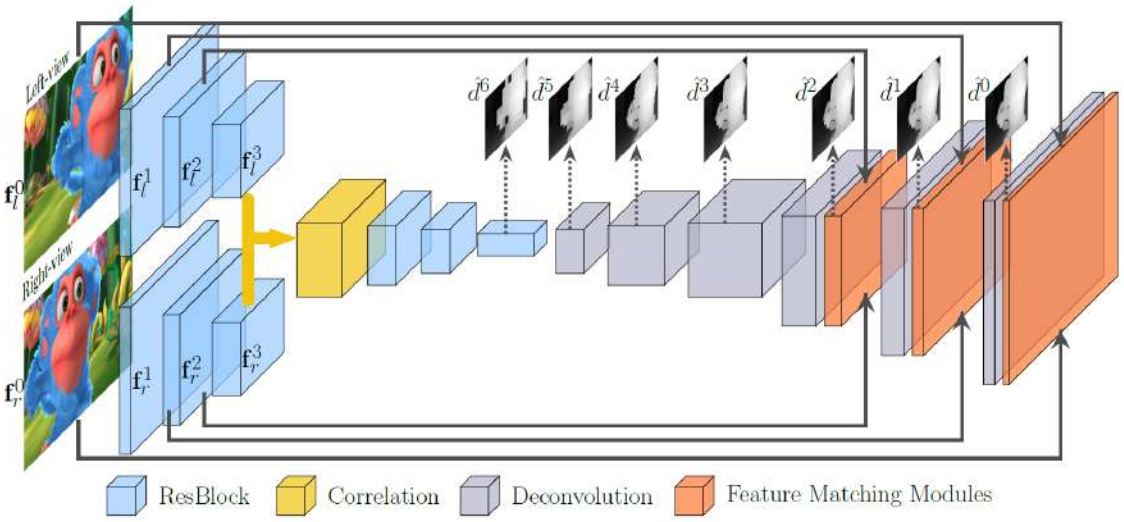
- P je počet pixelov v disparitnej mape
- dgt_p je skutočná disparita na pixely p
- dpr_p je odhadnutá disparita na pixely p
- $err_p(d_{gt}, d_{pr})$ je koncová chyba na pixely p, priemerný absolútny rozdiel medzi hodnotami pixelov disparitných máp

Porovnávací test Middlebury vypočítava až 10 metrík presnosti odhadu disparitnej mapy. Hlavnou metrikou v ich testoch je metrika bad2.0, to znamená percento pixelov s chybou väčšou ako 2.0. Metrika je teda podobná metrike D1 v porovnávacích testoch KITTI 3.1 ale nezapočítava pixely s chybou väčšou ako 5%. Ďalšie metriky sú bad0.5,

bad1.0 a bad4.0, ktoré počítajú pixely s chybou väčšou ako 0.5, 1.0 a 4.0, ďalej priemerná absolútna chyba, RMS – smerodajnú odchýlku chyby, a kvantilové chyby na kvantiloch A50 – mediánová chyba, A90, A95 a A99 [27].

Porovnávací test ETH3D vypočítava metriky Bad 1.0, Bad 0.5, priemernú absolútnu chybu a RMSE – smerodajnú odchýlku chyby rovnako ako porovnávací test Middlebury.

3.2 Metóda Efficient Stereo Matching Network



Obr. 3.1: Schéma architektúry ES-Net. Prevzaté z [12].

Metóda Efficient Stereo Matching Network [12] – ES-Net – si kladie za cieľ dosiahnuť vysokú rýchlosť predikcie pri zachovanú presnosti v aplikáciach z reálneho sveta. Toto dosahuje vďaka použitiu 2D konvolúcie na rozdiel od väčšiny ostatných metód využívajúcich 3D konvolúciu. Ďalej predstavuje výpočet viac-škálovej cenovej matice s deformáciou obrazu, metódu ES-Net-M s vylepšeným modelovaním prekrytia a v procese trénovania zavádzajú predtrénovanie bez učiteľa a vlastné plánovanie poradia dátových sád, ktoré vylepšuje presnosť modelu [12].

Viac-škálová cenová matica s deformáciou obrazu

ES-Net sa pri výpočte viac škálovej cenovej matice inšpirovalo architektúrou PWC-Net [32], ktorá využíva deformáciu zdrojového obrázku na cieľový, čím rieši problém posunu pixelov medzi obrázkami, ktorý je väčší ako maximálna nastavená disparita [12].

ES-Net vypočítava disparitu na siedmych úrovniach škálovania ($\hat{d}^s : s = 0, ..., 6$) a následne vypočíta cenová matica pre úrovne $s = 0, 1, 2$. Pred výpočtom cenovej matice sa mapa príznakov pre pravý obrázok najskôr deformuje s využitím funkcie 3.2 [12]:

$$\tilde{f}_r^s(x, y) = f_r^s(x + up_2(\hat{d}^{s+1})(x, y), y) \tag{3.2}$$

kde:

- $\tilde{f}_r^s$ je výsledná pokrivená mapa príznakov pre pravý obrázok

- s je úroveň škálovania
- $hatd^{s+1}$ je hrubá bilineárne nadvzorkovaná disparita na škále $s+1$
- up_2 bilineárna interpolácia

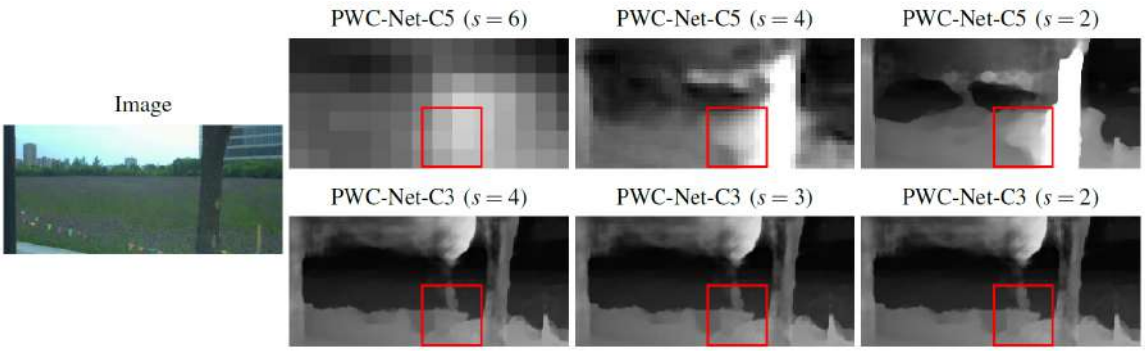
a následne sa vypočíta cenová matica pomocou rovnice 3.3 [12]:

$$c(x_1, y_1, d) = \frac{1}{N_c}\langle f_l(x_1, y_1), f_r(x_1 - d, y_1)\rangle \tag{3.3}$$

kde:

- $c(x_1, y_1, d)$ je cenová matica pre pixel na pozícií (x_1, y_1)
- $\frac{1}{N_c}$ je veľkosť príznakového kanálu
- $d, d \in \langle -2, 2\rangle$ sú úrovne disparity
- f_l a f_r sú mapy príznakov pre ľavý a pravý obrázok
- $\langle ...\rangle$ je skalárny súčin

Tým, že ES-Net vypočítava viac-škálová cenová matica len na vyšších rozlíšeniach (škálovacia úroveň s = 2, 1, 0) na rozdiel od architektúry PWC-Net, ktorá vypočítava cenu už od nízkeho rozlíšenia (škálovacia úroveň s = 6, 5, 4, 3, 2, 1), ES-Net redukuje chybu rozmazaných hrán, ktorá vzniká na nízkych rozlíšeniach a propaguje sa aj do vyšších rozlíšení. Príklad rozmazania hrán je možné vidieť na obrázku 3.2. Okrem toho je ES-Net vďaka zníženému množstvu potrebných výpočtov efektívnejšia [12].



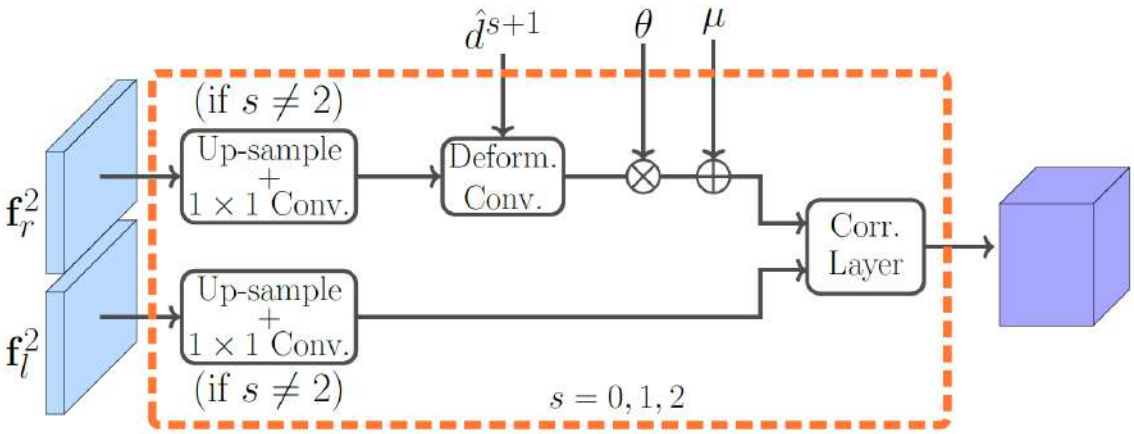
Obr. 3.2: Porovnanie rozmazania hrán architektúrou PWC-Net na rôznych úrovniach škálovania. Prevzaté z [12].

Modelovanie prekrytia

Autori ES-Net vytvorili aj modifikáciu pôvodnej architektúry, nazvanú ES-Net-M, ktorá rieši problém duplikácie prekrytých objektov pri aplikovaní deformácie, ktorý znižuje presnosť vypočítanej cenovej matice, pomocou vylepšeného modelovania prekrytia.

ES-Net-M využíva na každej úrovni škálovania mapy príznakov f_l^2 a f_r^2, ktoré na úrovni škálovania $s < 2$ upraví bilineárnym nad-vzorkovaním a 1x1 konvolúciou do správnych rozmerov. Následne deformuje pravý obrázok, podobne ako ES-Net, prenásobí ho maskou

prekryvu θ a pripočíta k nemu kompenzačný člen μ, čím sa zníži efekt duplikácie a zlepší sa kvalita vypočítanej cenovej matice. Maska prekryvu θ a kompenzačný člen μ sa odvodia z výstupov konvolúcie na úrovni škálovania $s + 1$ [12]. Schéma modelovania prekrytia architektúrou ES-Net-M je zobrazená na obrázku 3.3.



Obr. 3.3: Schéma modelovania prekrytia ES-Net-M. Prevzaté z [12].

Pred-trénovanie bez učiteľa

Pred-trénovanie bez učiteľa je bežný spôsob inicializácie váh pri trénovaní hlbokých neurónových sieti v úlohách klasifikácie obrázkov, odhadu toku alebo detekcie objektov.

Autori ES-Net ako prví zavádzajú pred-trénovanie bez učiteľa pri riešení problému stereo párovania. Model pred-trénujú bez učiteľa s použitím fotometrickej reprojekčnej chybovej funkcie, na rovnakej dátovej sade, na ktorej ho následne trénujú s učiteľom. Tento prístup zvyšuje presnosť odhadu disparity modelom ES-Net aj ES-Net-M [12].

Poradie trénovacích dátových sád

Bežný postup pri trénovaní stereo párovacích modelov, je najskôr trénovať na veľkej a synteticky vygenerovanej dátovej sade SceneFlow [21] a následne model dotrénovať na cieľovej dátovej sade pre daný porovnávací test, ako je napríklad dátová sada KITTI [22].

Autori ES-Net do svojho trénovacieho plánu pridali ešte jednu veľkú dátovú sadu s fotografiami z reálneho sveta, DrivingStereo [42], ktorá je podobná cieľovej dátovej sade KITTI. Pritom zistili, že poradie dátových sád má vplyv na presnosť výsledného modelu. Pri trénovaní v poradí SceneFlow, DrivingStereo, KITTI bol výsledný model presnejší, ako pri trénovaní v poradí DrivingStereo, SceneFlow, KITTI, čo naznačuje, že je lepšie trénovať na dátovej sade podobnej cieľovej sade neskôr. Výsledky trénovania v rôznych poradiach dátových sád je zobrazený v tabuľke 3.1 [12].

Porovnanie výsledkov

Metódy ES-Net a ES-Net-M dosahujú jedny z najvyšších presností pri porovnaní s inými state-of-the-art metódami zameranými na rýchlosť predikcie. Pri porovnaní s metódami, ktoré nie sú zamerané na rýchlosť, dosahuje ES-Net podobné alebo len o pár desatín per-

Model	Plán trénovania	D1-all
ES-Net	SF+K	2.80%
	DS+K	2.14%
	SF+DS+K	2.11%
	SF+DS+K	**2.02%**
ES-Net-M	SF+K	2.38%
	DS+K	2.36%
	DS+SF+K	2.03%
	SF+DS+K	**1.85%**

Tabuľka 3.1: Porovnanie chyby modelu ES-Net a ES-Net-M pri rôznych spôsoboch plánovania dátových sád. Prevzaté a preložené z [12].

centa horšie výsledky pričom doba spracovania je niekoľko násobne menšia. Prevzaté a preložené z [12].

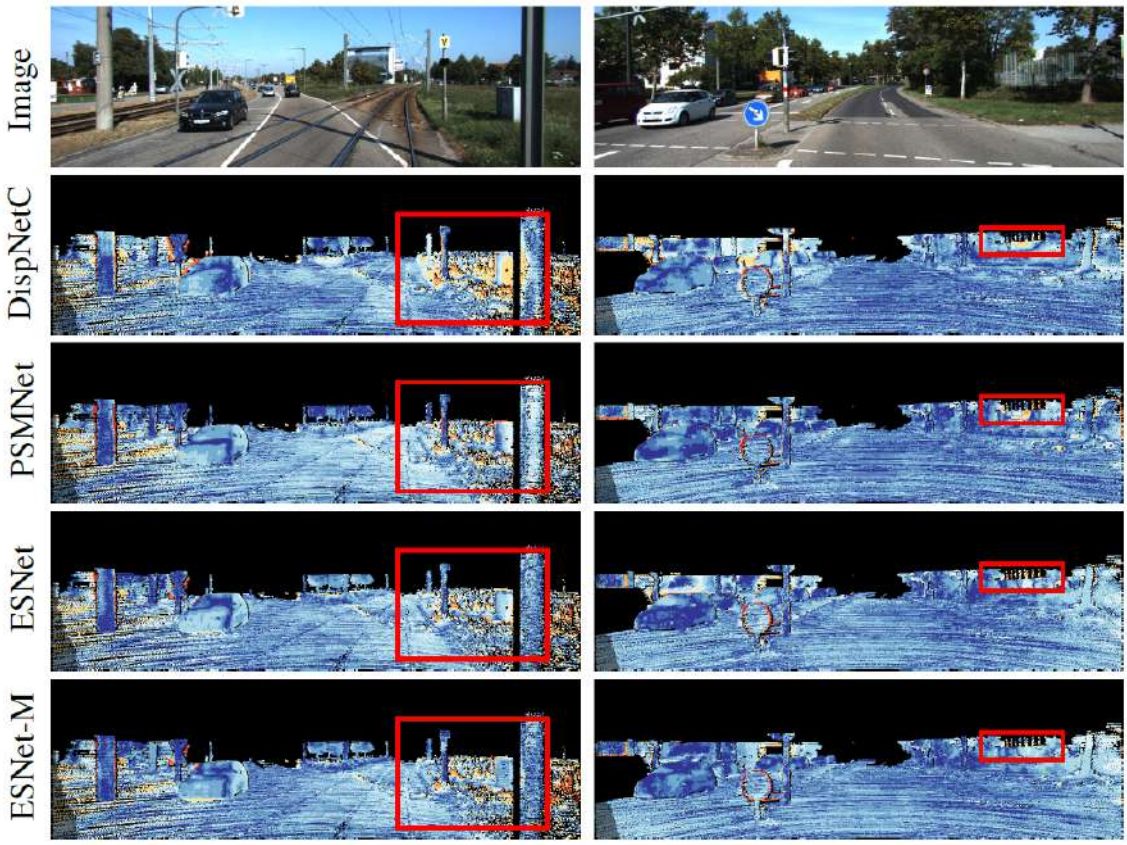
Porovnanie s inými state-of-the-art metódami je zobrazené v tabuľke 3.2. Vizualizácia chyby predikcie disparity je zobrazená na obrázku 3.4.

Tabuľka 3.2: Porovnanie State-of-the-art metód s metódou ES-Net [12].

Model	SceneFlow	KITTI 2015			Čas (ms)
	EPE	D1-bg	D1-fg	D1-all	
Čas <= 100ms					
MADNet [35]	-	3.75%	9.20%	4.66%	14
DispNetC [21]	1.68	4.32%	4.41%	4.34%	20
RTSNet [17]	-	2.86%	6.19%	3.41%	20
Fast DS-CS [43]	2.01	2.83%	4.31%	3.08%	20
FADNet [37]	**0.83**	2.68%	**3.5%**	2.82%	44
StereoNet [15]	1.10	4.30%	7.45%	4.83%	49
ESMNet [8]	0.84	2.57%	4.86%	2.95%	67
DeepPruner-Fast [7]	0.97	2.32%	3.91%	2.59%	74
ES-Net [12]	0.95	2.29%	4.17%	2.60%	28
ES-Net-M [12]	**0.84**	**2.15%**	3.74%	2.42%	42
Čas >100ms					
HSM [41]	-	1.80%	3.85%	2.14%	**135**
iResNet-i2 [19]	-	2.25%	**3.40%**	2.44%	137
DeepPruner-Best [7]	0.86	1.87%	3.56%	2.15%	178
PSMNet [5]	1.09	1.86%	4.62%	2.32%	520
GC-Net [14]	2.5	2.21%	6.16%	2.87%	1030
GA-Net [44]	**0.84**	**1.48%**	3.46%	**1.81%**	2700

3.3 Metóda Cascaded Recurrent Stereo Matching Network

Metóda Cascaded Recurrent Stereo Matching Network [18] – CREStereo – predstavuje niekoľko návrhov na zlepšenie stereo párovania pri riešení problémov z reálneho sveta. Navrhuje hierarchickú neurónovú sieť s rekurentným vylepšovaním, ktoré postupne vylepšuje disparitnú mapu od hrubej predikcie až po jemnú. Ďalej predstavuje adaptívnu skupinovú



Obr. 3.4: Porovnanie chyby predikcie disparity s metódou ES-Net. Červenejšie pixely predstavujú väčšiu chybu. Prevzaté z [12].

korelačnú vrstvu, ktorá zmenšuje dôsledky chybnej rektifikácie obrázkov. Nakoniec autori tejto metódy vytvorili vlastnú syntetickú dátovú sadu, ktorá má pomáhať pri generalizácií pri fotografiách z reálneho sveta. Vďaka tomu dosahuje CREStereo state-of-the art výsledky [18].

Adaptívna skupinová korelačná vrstva

Pri fotografiách z reálneho sveta niekedy nie je možné dosiahnuť dokonalú rektifikáciu obrazu, kvôli nepresnostiam kamier alebo ich nepresnému umiestneniu, čo môže spôsobiť, že zodpovedajúce body neležia v obrázkoch na tej istej horizontálnej úrovni a nedajú sa ľahko spárovať [18].

CREStereo počíta koreláciu len v lokálnom okne, rovnica 3.4, čo znižuje pamäťové a výpočtové nároky. Výsledná cenová matica je teda oveľa menšia ako pri iných metódach. Na riešenie problému nesprávne rektifikovaného obrazu využíva CREStereo 2D-1D striedavé lokálne prehľadávanie. Pri 1D prehľadávaní sa nastaví $g(d) = 0$ a $f(d) = [-r, r], r = 4$. Pri 2D prehľadávaní sa na koreláciu využíva okno o veľkosti $k \times k$, kde $k = \sqrt{2r + 1}$. Keďže statické prehľadávacie okno nefunguje dobre v bezprízнakových oblastiach, CREStereo využíva adaptívne okno s ofsetom dx, dy. Nakoniec CREStereo rozdeľuje mapu príznakov na G skupín, pre každú skupinu vypočíta koreláciu samostatne a na konci ich spája [18].

$$Corr(x, y, d) = \frac{1}{C} \sum_{i=1}^{C} F_1(i, x, y) F_2(i, x'', y'') \tag{3.4}$$

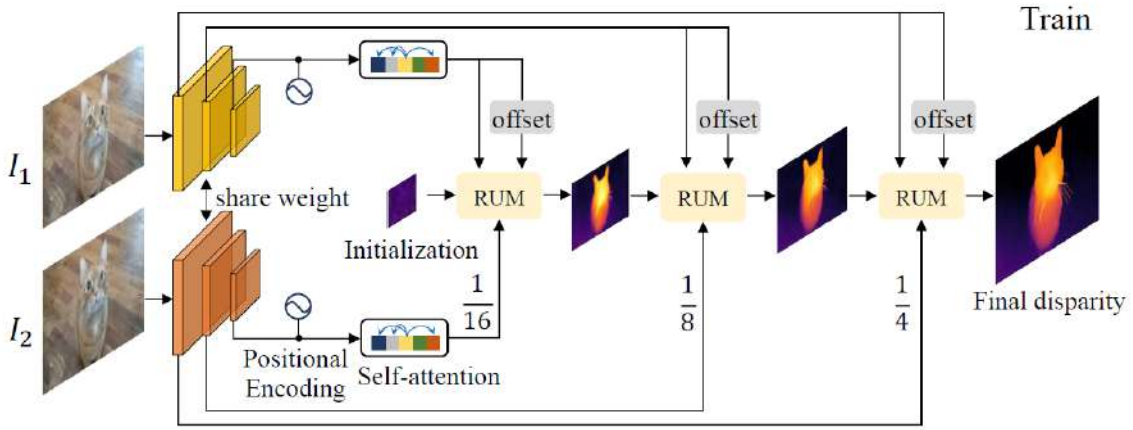
kde:

- F_1, F_2 sú mapy príznakov
- x, y je pozícia pixelu
- $x'' = x + f(d) + dx, y'' = y + g(d) + dy$
- $Corr(x, y, d)$ je párovacia cena d-teho korelačného páru, kde $d \in [0, D-1]$
- C je počet príznakových kanálov
- $f(d), g(d)$ fixný ofset súčasného pixelu v horizontálnom a vertikálnom smere

Kaskádová rekurentná neurónová sieť

V oblastiach obrázku bez príznakov alebo s opakujúcou sa textúrou sú párovacie metódy najrobustnejšie pri použití mapy príznakov s nízkym rozlíšením, čím sa ale môžu stratiť jemné detaily. CREStereo na riešenie tohto problému predstavuje kaskádové rekurentné vylepšovanie pre výpočet korelácie a aktualizáciu disparity [18].

Táto neurónová sieť využíva rekurentný aktualizačný modul – RUM – založený na GRU blokoch a adaptívnej skupinovej korelačnej vrstve, ktorý počíta koreláciu nezávisle pre každú mapu príznakov na rôznych kaskádových úrovniach a vylepšuje predikovanú disparitu v niekoľkých iteráciách. Prvá úroveň kaskády začína na 1/16 rozlíšenia pôvodného obrázku a je inicializovaná s nulovou disparitnou mapou. Ďalšie úrovne potom preberajú predikovanú disparitnú mapu z predchádzajúcej úrovne. Všetky RUM moduly zdieľajú rovnaké váhy [18]. Schéma architektúry je zobrazené na obrázku 3.5.



Obr. 3.5: Schéma architektúry CREStereo. Prevzaté z [18].

Syntetická trénovacia dátová sada

Autori CESNet vytvorili pomocou nástroja Blender syntetickú trénovaciu dátovú sadu, ktorá má simulovať prípady scén z reálneho sveta, s ktorými majú stereo párovacie metódy zvyčajne problémy. Dátová sada pozostáva z párov obrázkov a prislúchajúcej disparitnej mape. Ukážka na obrázku 3.6.

Autori pri generovaní dátovej sady využili vyše 40000 3D objektov z dátovej sady ShapeNet a ďalšie objekty z nástroja Blender aby vytvorili náročné scény s rôznymi dierami v objektoch. Do scén náhodne vložili svetelné zdroje s rôznou farbou a svietivosťou. Objekty obalili textúrami z fotografií z reálneho sveta, pričom využili hlavne textúry bez príznakov alebo s opakujúcimi sa vzormi. Výsledná dátová sada obsahuje obrázky s vysokou škálou rozloženia disparity [18].

Tabuľka 3.3: Výsledky metódy CREStereo v porovnávacích testoch Middlebury. Prevzaté a preložené z [18].

Metóda	Bad 2.0	Bad 1.0	AvgErr	RMS	A95
CREStereo [18]	**3.71**	**8.25**	**1.15**	**7.70**	**1.58**
RAFT-Stereo [20]	4.74	9.37	1.27	8.41	2.29
LocalExp [33]	5.43	13.9	2.24	13.4	4.81
HITNet [34]	6.46	13.3	1.71	9.97	4.26
LEAStereo [6]	7.15	20.8	1.43	8.11	2.65
SDR [40]	7.69	18.8	2.94	15.4	7.13
MC-CNN-acrt [46]	8.08	17.1	3.82	21.3	14.1
CFNet [29]	10.1	19.6	3.49	15.4	16.4
HSMNet [36]	10.2	24.6	2.07	10.3	4.32
AdaStereo [30]	13.7	29.5	2.22	10.2	5.67
AANet++ [39]	15.4	25.5	6.37	23.5	48.8

Tabuľka 3.4: Výsledky metódy CREStereo v porovnávacích testoch ETH3D. Prevzaté a preložené z [18].

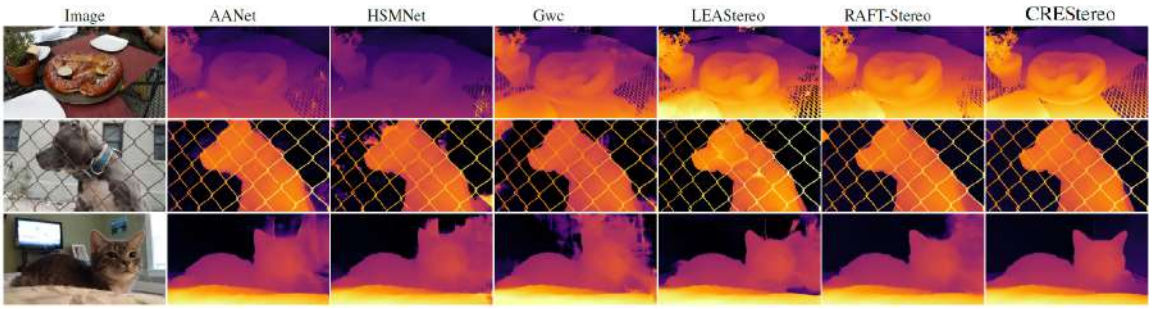
Metóda	Bad 1.0	Bad 0.5	AvgErr	RMSE
CREStereo [18]	**0.98**	**3.58**	**0.13**	**0.28**
RAFT-Stereo [20]	2.44	7.04	0.18	0.36
HITNet [34]	2.79	7.83	0.20	0.46
AdaStereo [30]	3.09	10.22	0.24	0.44
CFNet [29]	3.31	9.87	0.24	0.51
GwcNet [9]	3.66	12.04	0.29	0.67
iResNet [19]	3.68	10.26	0.24	0.51
HSMNet [36]	4.00	11.33	0.28	0.62
AANet [39]	5.01	13.16	0.31	0.68
GANet [15]	6.56	25.41	0.43	0.75

Porovnanie výsledkov

Autori metódu CREStereo overili na rôznych porovnávacích testoch, pričom v každom dosiahli state-of-the-art výsledky. Metóda CREStereo skončila na prvom mieste v porovnávacích testoch Middlebury [27], tabuľka 3.3, a ETH3D, tabuľka 3.4, [28].

Obr. 3.6: Ukážka obrázkov a prislúchajúcich disparitných máp zo synteticky vygenerovanej dátovej sady pre metódu CREStereo. Prevzaté z [18].

Porovnanie metódy na dátovej sade z reálneho sveta Holopix50K [11], ukazuje, že CREStereo je lepší ako ostatné metódy v oblastiach obrázku s jemnými detailami alebo v oblastiach bez príznakov. Ukážka výsledkov je na obrázku 3.7.



Obr. 3.7: Porovnanie predikcie disparity s modelom CREStereo. Prevzaté z [18].

3.4 Metóda Attention Concatenation Volume Network

Autori metódy Attention Concatenation Volume Network [38] – ACVNet – predstavujú novú metódu výpočtu cenovej matice, ktorá generuje váhy pre attention bloky, ktoré potlačujú redundantné informácie a zlepšujú dôležité informácie vo výslednej cenovej matici. Takúto cenovú maticu nazývajú Attention Concatenation Volume – ACV, ktorá sa dá podľa nich ľahko implementovať aj do iných neurónových sietí. Ďalej predstavujú neurónovú sieť ACVNet, ktorá dosahuje state-of-the-art presnosť v rôznych porovnávacích testoch [38].

Attention Concatenation Volume

Proces konštrukcie attention zreťazenej cenovej matice pozostáva z troch krokov: konštrukcia zreťazenej cenovej matice, generovanie váh pre attention bloky a filtrovanie attention.

Z páru vstupných obrázkov sa získajú mapy príznakov a z nich sa skonštruuje zreťazená cenová matica pomocou rovnice 3.5 [38].

$$\mathbf{C}_{concat}(., d, x, y) = Concat\{\mathbf{f_l}(x, y), \mathbf{f_r}(x - d, y)\} \tag{3.5}$$

kde:

- f_l, f_r je ľavá a pravá mapa príznakov
- x, y je pozícia pixelu
- d je úroveň disparity
- *Concat* je operácia zreťazenia

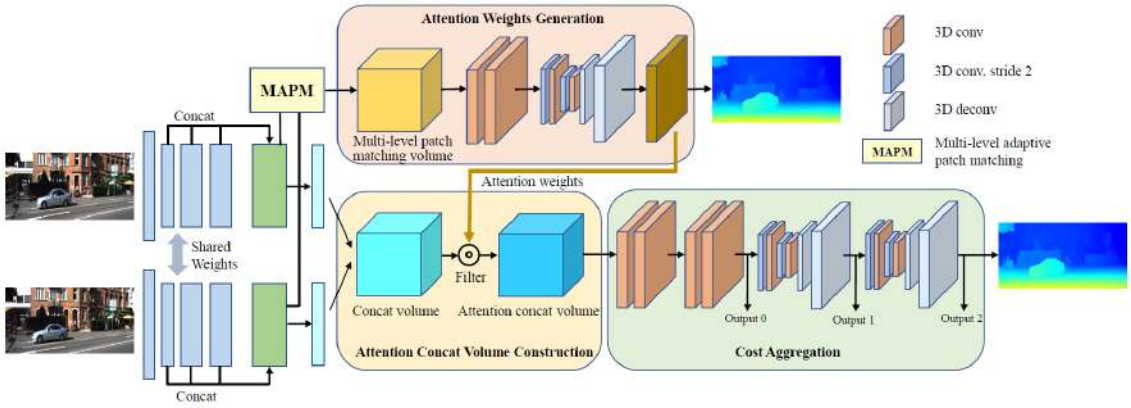
Váhy attention blokov **A** sa vygenerujú extrakciou geometrických informácií z korelácie medzi pármi vstupných obrázkov. Výsledná attention zreťazená cenová matica sa vypočíta pomocou rovnice 3.6 [38].

$$\mathbf{C}_{ACV}(i) = \mathbf{A} \odot \mathbf{C}_{concat}(i) \tag{3.6}$$

Architektúra ACVNet

Architektura ACVNet sa skladá zo štyroch modulov: extraktor príznakov, konštruktor attention zreťazenej cenovej matice, agregátor ceny a prediktor disparity.

Extraktor príznakov je podobný architektúre ResNet a extrahuje tri mapy príznakov, ktoré sa následne zreťazia do 320-kanálovej mapy príznakov, ktorá je použitá na generovanie váh attention blokov. Táto mapa sa následne skomprimuje pomocou konvolučných vrstiev do 32-kanálovej mapy príznakov z ktorej sa skonštruuje zreťazená cenová matica a následne konštruktor attention zreťazenej cenovej matice vytvorí ACV. ACV sa agreguje do troch výstupov z ktorých sa odhadnú tri disparitné mapy [38]. Schéma architektúry ACVNet je zobrazená na obrázku 3.8.



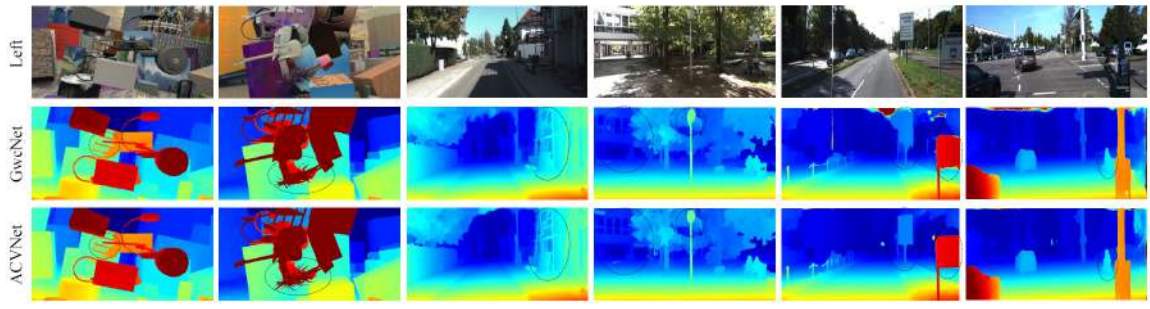
Obr. 3.8: Schéma architektúry ACVNet. Prevzaté z [38].

Porovnanie výsledkov

Metóda ACVNet končí v porovnávacích testoch KITTI 2012, 2015 [22] a SceneFlow [21] na druhom mieste, pričom v prípade KITTI testu dosahuje najlepší čas výpočtu. Porovnanie výsledkov je zobrazené v tabuľke 3.5. V prípade porovnávacieho testu ETH3D [28] končí metóda ACVNet až na prvom mieste [38]. Porovnanie výslednej predikovanej disparity je zobrazené na obrázku 3.9.

Tabuľka 3.5: Výsledky porovnávacieho testu KITTI 2012 a KITTI 2015 pre metódu ACV-Net. V porovnávacích testoch KITTI 2012, x-noc znamená chybu na pixely väčšiu ako x percent v oblastiach obrázka bez oklúzie a x-all znamená chybu na pixely väčšiu ako x percent v celom obrázku. Prevzaté a preložené z [38].

	KITTI 2012						KITTI 2015			
Metóda	2-noc	2-all	3-noc	3-all	EPE noc	EPE all	D1-bg	D1-fg	D1-all	Čas (ms)
GC-Net [14]	2.71	3.46	1.77	2.30	0.6	0.7	2.21%	6.16%	2.87%	900
PSMNet [5]	2.44	3.01	1.49	1.89	0.5	0.6	1.86%	4.62%	2.32%	410
EdgeStereo [31]	2.32	2.88	1.46	1.83	**0.4**	**0.5**	1.84%	3.30%	2.08%	320
GwcNet [9]	2.16	2.71	1.32	1.70	0.5	**0.5**	1.74%	3.93%	2.11%	320
GANet-deep [44]	1.89	2.50	1.19	1.60	**0.4**	**0.5**	1.48%	3.46%	1.81%	1800
AcfNet [45]	**1.83**	**2.35**	1.17	1.54	0.5	**0.5**	1.51%	3.80%	1.89%	480
HITNet [34]	2.00	2.65	1.41	1.89	**0.4**	**0.5**	1.74%	3.20%	1.98%	20
CFNet [29]	1.90	2.43	1.23	1.58	**0.4**	**0.5**	1.54%	3.56%	1.88%	180
LEAStereo [6]	1.90	2.39	**1.13**	**1.45**	0.5	**0.5**	1.40%	**2.91%**	**1.65%**	300
ACVNet [38]	**1.83**	**2.35**	**1.13**	1.47	**0.4**	**0.5**	**1.37%**	3.07%	**1.65%**	200



Obr. 3.9: Výsledky porovnávacieho testu KITTI 2012 a KITTI 2015 pre metódu ACVNet. Prevzaté z [38].

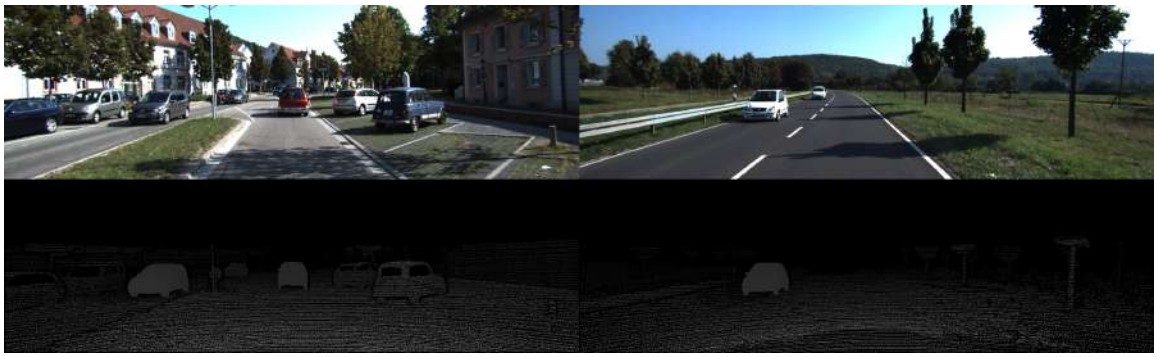
Kapitola 4

Dátové sady používané na trénovanie a evaluáciu stereo párovacích metód

V tejto kapitole popíšem niektoré dátové sady, ktoré sa bežne používajú pri trénovaní a evaluácií metód stereo rekonštrukcie.

4.1 Dátová sada KITTI

Dátová sada KITTI 2015 [22], projekt inštitútu technológií Karlsruhe a Technologického inštitútu Toyota v Chicagu, pozostáva z 200 trénovacích a 200 testovacích párov rektifikovaných fotografií, vyfotografovaných dvomi kamerami zo strechy auta pri jazde mestom aj mimo mesta. Pre každý pár trénovacích fotografií je k dispozícií obrázok skutočnej mapy disparity, vytvorenej pomocou LiDAR kamery. Všetky obrázky sú vo formáte .png v rozlíšení 1241x376. Okrem dátovej sady je priložený aj kalibračný súbor s vnútornými a vonkajšími parametrami kamier. Táto dátová sada je využívaná v state-of-the-art metódach pri evaluácií v KITTI Vision Benchmark – porovnávacích testoch [22]. Ukážka dátovej sady je zobrazená na obrázku 4.1.

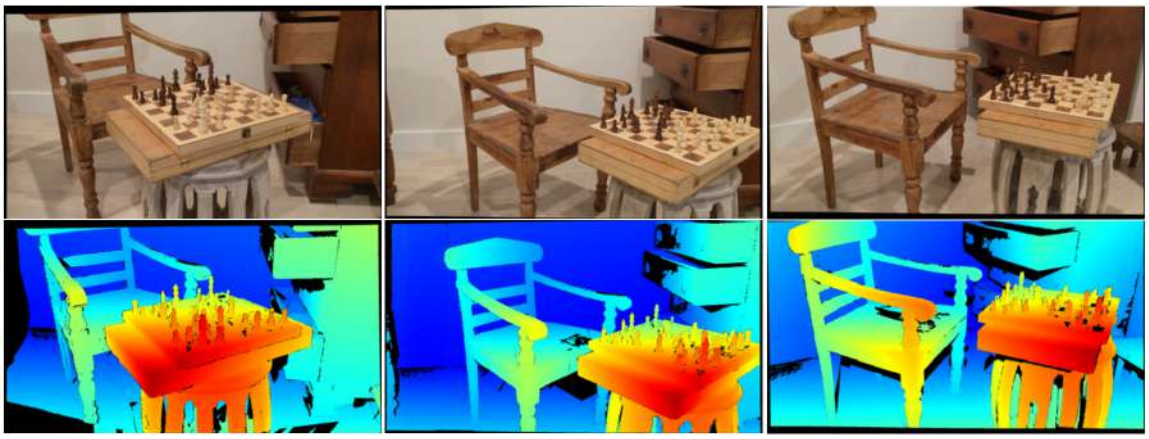


Obr. 4.1: Ukážka dátovej sady KITTI 2015 [22]. Prvý riadok – ľavý obrázok, druhý riadok – skutočná disparita pre ľavý obrázok

4.2 Dátová sada Middlebury

Dátová sada Middlebury 2021 Mobile [27] je zložená z jedenástich scén vyfotografovaných v miestnostiach. V každej scéne sú rôzne objekty rozložené po miestnosti. Každá scéna je vyfotografovaná z rôznych uhlov a za rôznych svetelných podmienok. Dátová sada obsahuje 24 párov rektifikovaných fotografií vo formáte .png s rozlíšením 1920x1080 a pre každú fotografiu je priložená skutočná mapa disparity vo formáte .pfm, vytvorená pomocou systému štruktúrovaných svetiel, so sub-pixlovými a samo-kalibračnými komponentami, predstavenom v [27].

Táto dátová sada je využívaná v state-of-the-art metódach pri evaluácií v Middlebury Stereo Evaluation – porovnávacích testoch. Ukážka dátovej sady Middlebury 2021 Mobile je zobrazená na obrázku 4.2.



Obr. 4.2: Ukážka dátovej sady Middlebury 2021 Mobile. Prevzaté z [27]. Prvý riadok – ľavý obrázok, druhý riadok – skutočná disparita pre ľavý obrázok

4.3 Dátová sada DrivingStereo

Dátová sada DrivingStereo [42] obsahuje 182188 párov fotografií vo formáte .jpg, vyfotografovaných zo strechy auta, podobne ako dátová sada KITTI, z toho 174437 párov je trénovacích a 7751 testovacích. Rozlíšenie fotografií je v priemere 881x400. Pre každý pár fotografií je v dátovej sade priložená skutočná disparitná a hĺbková mapa vo formáte .png, vytvorená pomocou LiDAR kamery [42].

Táto dátová sada je využívaná v state-of-the-art metódach pri trénovaní modelov. Ukážka dátovej sady DrivingStereo je zobrazená na obrázku 4.3.

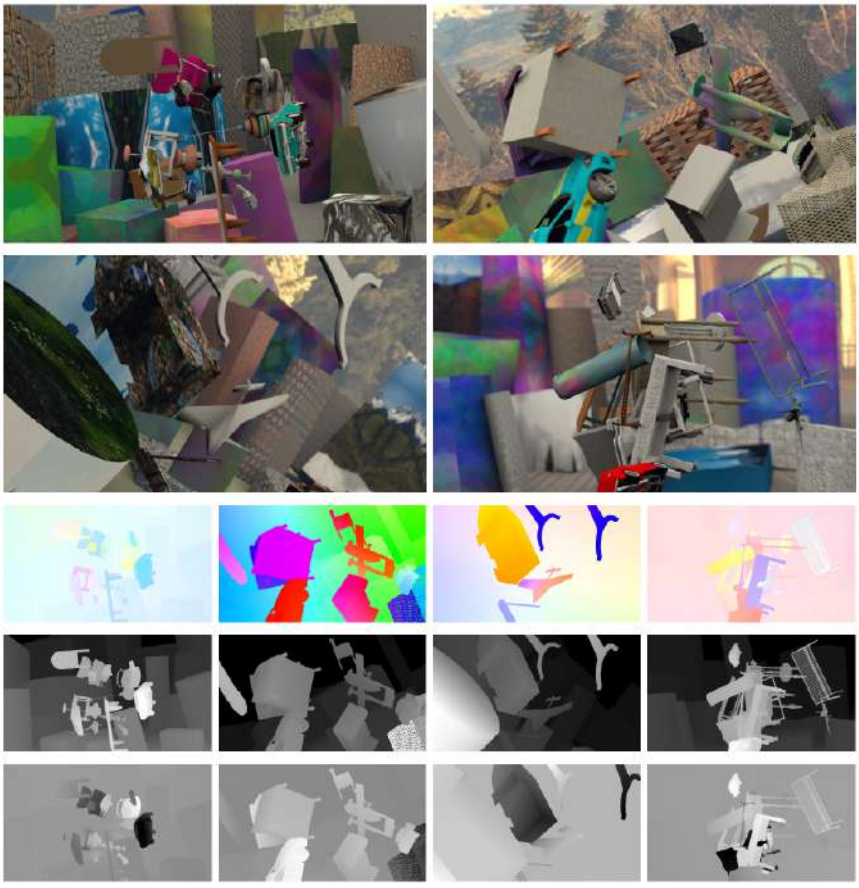
4.4 Dátová sada SceneFlow

Dátová sada SceneFlow [21] obsahuje vyše 39000 synteticky vygenerovaných párov obrázkov v rozlíšení 960x540 vo formáte .png. SceneFlow obsahuje tri podsady, FlyingThings3D s obrázkami rôznych lietajúcich objektov, Monkaa s obrázkami opice z krátkeho animovaného filmu Monkaa a Driving s obrázkami z pohľadu z ulice podobným fotografiám v dátovej sade KITTI. Pre každý pár obrázkov je k dispozícií skutočná disparitná mapa vo formáte

Obr. 4.3: Ukážka dátovej sady DrivingStereo. Prevzaté z [42].

.pfm. Okrem disparitnej mapy dátová sada obsahuje aj obrázky optického toku a zmeny disparity [21].

Táto dátová sada je využívaná v state-of-the-art metódach pri trénovaní modelov. Ukážka dátovej sady SceneFlow FlyingThings3D je zobrazená na obrázku 4.4.



Obr. 4.4: Ukážka dátovej sady SceneFlow FlyingThings3D. Prevzaté z [21].

Kapitola 5

Návrh riešenia rekonštrukcie mračna bodov

Cieľom tejto práce je vytvoriť a natrénovať model hlbokej neurónovej siete, ktorý bude schopný odhadovať disparitnú mapu z dvojice obrázkov. Z odhadnutej disparitnej mapy následne bude možné vypočítať hĺbkovú mapu a zrekonštruovať husté mračno bodov, znázorňujúce danú scénu v troch rozmeroch. Rozhodol som sa vybrať dve z metód, ktoré som skúmal v kapitole 3, a porovnať ich.

Odhad hĺbky a stereo rekonštrukcia má aplikácie v rôznych oblastiach počítačového videnia a spracovania obrazu. Príkladmi aplikácií sú augmentovaná realita, kedy stereo rekonštrukcia pomáha pri vkladaní virtuálnych objektov správne do priestoru a tým zlepšuje používateľskú skúsenosť. V medicíne umožňuje stereo rekonštrukcia vytvárať presné 3D modely orgánov, čím pomáha pri diagnostike ochorení a plánovaní operácií [3, 16]. V robotike a v autonómnych systémoch, ako sú napríklad samo-riadiace automobily, sa pomocou stereo rekonštrukcie vytvárajú modely okolitého priestoru, čo je dôležité na detekciu prekážok a navigáciu [16].

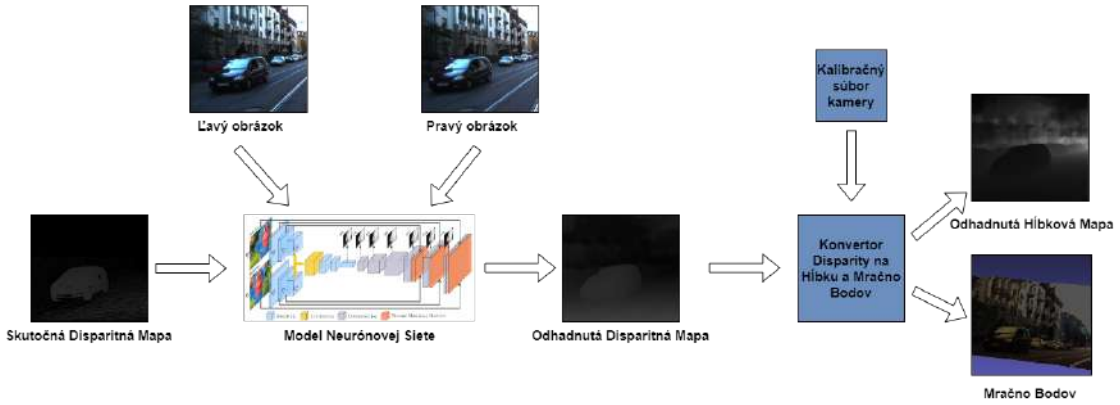
V tejto kapitole sa venujem dielčím krokom návrhu a realizácie riešenia:

- Výber vhodných dátových sád na trénovanie modelov hlbokých neurónových sietí a na evaluáciu výsledkov natrénovaných modelov
- Výber architektúr hlbokých neurónových sietí, konkrétne som zvolil architektúru ES-Net a CRES-Stereo, ktoré budem ďalej v experimentoch porovnávať
- Princíp transformácia disparitnej mapy na hĺbkovú mapu a mračno bodov

V tejto kapitole som popísal výber dátových sád pre trénovanie a evaluáciu modelov neurónových sietí, výber architektúry neurónovej siete na odhad disparitnej mapy a postup pri vytváraní hĺbkovej mapy a mračna bodov z disparitnej mapy.

5.1 Výber dátovej sady

Prvým krokom pri riešení úlohy je výber vhodnej dátovej sady na trénovanie a testovanie modelov hlbokých neurónových sietí určených na odhad disparitnej mapy z páru obrázkov. Z dátových sád popísaných v kapitole 4, som sa rozhodol použiť na trénovanie modelov dátovú sadu DrivingStereo [42], ktorá obsahuje najväčší počet obrázkov zo všetkých skúmaných dátových sád. Ako testovaciu sadu som sa rozhodol použiť dátovú sadu KITTI 2015 [22],



Obr. 5.1: Postup rekonštrukcie hustého mračna bodov.

keďže sa bežne používa na evaluáciu modelov na odhad disparitnej mapy s využitím ich porovnávacích testov a tiež preto, že rovnako ako dátová sada DrivingStereo, obsahuje snímky ulice z pohľadu auta. Keďže dátová sada KITTI 2015 sprístupňuje skutočnú mapu disparity len pre ich trénovacie dáta, rozhodol som sa rozdeliť ich trénovaciu sadu na 150 trénovacích vzoriek a 50 vyhodnocovacích vzoriek, na ktorých budem môcť vyhodnotiť natrénované modely na poskytnutých porovnávacích testoch KITTI Vision Benchmark.

Augmentácia dátovej sady

Dôležitým krokom, pre zlepšenie presnosti a schopnosti generalizovať pri modeloch hlbokých neurónových sietí je augmentácia dátovej sady. Hoci je dátová sada DrivingStereo pomerne rozsiahla, inšpiroval som sa inými prácami a rozhodol sa na dáta aplikovať nasledujúce augmentácie [12, 18]:

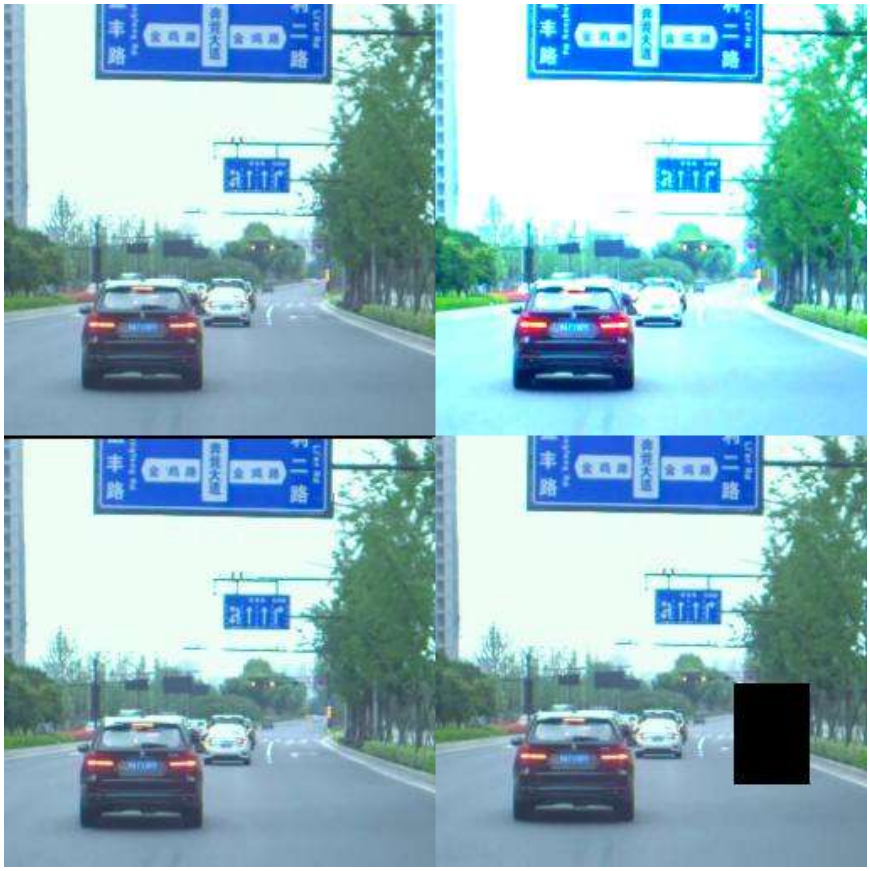
- Normalizácia dátovej sady s použitím stredných hodnôt a štandardných odchýlok vypočítaných na daných dátových sadách s použitím normalizačnej rovnice 5.1
- Náhodné orezanie vstupných obrázkov do veľkosti 512x256 pixelov
- Zmena jasu, kontrastu a saturácie vstupných obrázkov s náhodným faktorom v rozmedzí $\langle 0.6, 1.4 \rangle$
- Náhodné pootočenie v uhle $\langle -0.1, 0.1 \rangle$ stupňov, náhodné vertikálne posunutie o $\langle -0.015 * H, 0.015 * H \rangle$ pixelov, kde H je výška obrázka a náhodná zmena veľkosti pravého vstupného obrázka do $\langle 0.99, 1.01 \rangle$ násobku pôvodnej veľkosti – simulácia nedokonalej rektifikácie obrázkov
- Náhodné vynulovanie obdĺžnikovej oblasti v pravom vstupnom obrázku – simulácia oklúzie – ktorá má pomer obsahu v rozmedzí $\langle 0.01, 0.1 \rangle$ k obsahu obrázku

Augmentácie budem na obrázky aplikovať dynamicky pri načítaní dátových sád. Všetky augmentácie – okrem normalizácie a náhodného orezania, ktoré sa aplikujú vždy – sa na každú vstupnú vzorku aplikujú s pravdepodobnosťou $p = 0.5$. Augmentácie sú znázornené na obrázku 5.2. Obrázky budú normalizované pomocou Z-score normalizácie – tzv. štandardizácie, rovnica 5.1:

$$\mathbf{I_N} = \frac{\mathbf{I} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \tag{5.1}$$

kde

- $\mathbf{I_N}$ je normalizovaný obrázok
- $\mathbf{I}$ je pôvodný obrázok
- $\boldsymbol{\mu}$ je vektor stredných hodnôt pre každý farebný kanál obrázkov z dátovej sady
- $\boldsymbol{\sigma}$ je vektor štandardných odchýlok pre každý farebný kanál obrázkov z dátovej sady



Obr. 5.2: Ukážka augmentácií aplikovaných na pôvodný obrázok po náhodnom orezaní do veľkosti 256x256 pixelov. Pre jednotlivé augmentácie v ukážkach boli použité hraničné hodnoty popísané vyššie; vľavo hore – pôvodný obrázok, vpravo hore – zmena jasu, kontrastu a saturácie, vľavo dole – pootočenie, vertikálny posun a zmena veľkosti, vpravo dole – vloženie oklúzie

5.2 Architektúra neurónovej siete ES-Net

Ako prvú architektúru neurónovej siete som zvolil metódu ES-Net [12], ktorú som bližšie popísal v kapitole 3.2. Metódu som vybral kvôli tomu, že dosahuje relatívne vysokú presnosť odhadu disparitnej mapy pri vysokej rýchlosti výpočtu pričom využíva štandardný

typ architektúry, pozostávajúci zo sekvencie reziduálnych blokov a konvolučných a dekonvolučných vrstiev v rámci celého procesu odhadu hĺbky z obrazu.

Výber chybovej funkcie pre metódu ES-Net

Na trénovanie modelu ES-Net budem používať chybovú funkciu MultiscaleLoss použitú v metóde ES-Net [12]. Metóda ES-Net predikuje disparitné mapy na siedmich úrovniach škálovania, chybová funkcia MultiscaleLoss teda iteratívne zmenšuje skutočnú disparitnú mapu na úroveň škály predikovanej disparitnej mapy, ktorú porovnáva. Chybová funkcia MultiscaleLoss sa dá popísať rovnicou 5.2 :

$$L = \sum_{s=0}^{7} \omega_s l(\hat{d}_s, d_{gts}) \tag{5.2}$$

pričom

- L je výsledná vypočítaná chyba
- s je úroveň škálovania
- ω_s sú váhy pre predikovanú disparitnú mapu na úrovni škálovania s
- $\hat{d}_s$ je predikovaná disparitná mapa na úrovni škálovania s
- d_{gts} je skutočná disparitná mapa zmenšená na úroveň škálovania s
- $l(\hat{d}_s, d_{gtgts})$ je kritérium, ktoré meria rozdiel medzi zadanými disparitnými mapami

Nastavenie váh v chybovej funkcií MultiscaleLoss popisuje rovnica 5.3, pričom hodnotu γ som zvolil 0.8.

$$\omega_s = \gamma^s, s \in \langle 0, 6 \rangle \tag{5.3}$$

Ako kritérium $l(\hat{d}_s, d_{gts})$ je v chybovej funkcií MultiscaleLoss zvolená funkcia smoothL1Loss 5.4 [23, 12].

$$l(x, y) = \frac{1}{N} \sum_{n=0}^{N} (\begin{cases} 0.5(x_n - y_n)^2 & \text{ak } |x_n - y_n| < 1.0 \\ |x_n - y_n| - 0.5 & \text{inak} \end{cases}) \tag{5.4}$$

kde

- N je počet vzoriek
- x_n je predikovaná disparitná mapa
- y_n je skutočná disparitná mapa

5.3 Architektúra neurónovej siete CREStereo

Ako druhú architektúru neurónovej siete som zvolil metódu CREStereo [18], ktorú som bližšie popísal v kapitole 3.3. Táto metóda využíva, okrem štandardných reziduálnych blokov a konvolučných vrstiev na extrakciu príznakov, aj rekurentné vrstvy na odhad disparitnej mapy s iteratívnym vylepšovaním odhadu.

Výber chybovej funkcie pre metódu CREStereo

Na trénovanie modelu CREStereo budem používať chybovú funkciu SequenceLoss z metódy CREStereo [18]. Metóda CREStereo na rozdiel od metódy ES-Net predikuje disparitné mapy v pôvodnom rozlíšení, chybová funkcia SequenceLoss teda na rozdiel od MultiscaleLoss porovnáva predikované disparitné mapy s neupravenou skutočnou disparitnou mapou. Chybová funkcia SequenceLoss sa dá popísať rovnicou 5.5:

$$L = \sum_{p=0}^{P} \omega_p l(d_p, d_{gt}) \tag{5.5}$$

pričom

- L je výsledná vypočítaná chyba
- P je počet predikovaných disparitných máp
- ω_p sú váhy pre predikovanú disparitnú mapu p
- d_p je predikovaná disparitná mapa p
- d_{gt} je skutočná disparitná mapa
- $l(d_p, d_{gt})$ je kritérium, ktoré meria rozdiel medzi zadanými disparitnými mapami

Nastavenie váh v chybovej funkcií SequenceLoss popisuje rovnica 5.6, pričom hodnotu γ som zvolil 0.8 a P je počet predikovaných disparitných máp [18].

$$\omega_p = \gamma^{P-p-1}, p \in \langle 0, P) \tag{5.6}$$

Za kritérium $l(d_p, d_{gt})$ je v chybovej funkcií SequenceLoss zvolená funkcia L1Loss 5.7 [23, 18].

$$l(x, y) = \frac{1}{N} \sum_{n=0}^{N} (|x_n - y_n|) \tag{5.7}$$

kde

- N je počet vzoriek
- x_n je predikovaná disparitná mapa
- y_n je skutočná disparitná mapa

5.4 Výpočet hĺbkovej mapy a rekonštrukcia mračna bodov

Posledným krokom po odhadnutí disparitnej mapy, je výpočet hĺbkovej mapy a rekonštrukcia mračna bodov z disparitnej mapy. K tomu je potrebné poznať vnútorné parametre kamier, ktoré vyfotografovali vstupné snímky. Dátové sady DrivingStereo aj KITTI majú ku každému vstupnému páru obrázkov priložený kalibračný súbor, z ktorého sa dajú vnútorné parametre získať a u oboch dátových sád vyzerajú kalibračné súbory rovnako. Pre každú z kamier použitých pre vytvorenie dátovej sady sa v kalibračnom súbore nachádzajú, každá na samostatnom riadku ako séria desatinných čísiel, nasledovné matice [22, 42]:

- S: vektor veľkosti 1x2 reprezentujúci rozmery obrázka pred rektifikáciou
- K: kalibračná matica kamery veľkosti 3x3 pred rektifikáciou
- D: vektor skreslenia kamery veľkosti 1x5 pred rektifikáciou
- R: rotačná matica kamery veľkosti 3x3
- T: translačný vektor kamery veľkosti 3x1
- S_rect: vektor veľkosti 1x2 reprezentujúci rozmery obrázka po rektifikácií
- R_rect: rektifikačná matica veľkosti 3x3
- P_rect: projekčná matica veľkosti 3x4 po rektifikácií

Pre účely výpočtu hĺbkovej mapy a rekonštrukcie mračna bodov z disparitnej mapy odhadnutej nad ľavým obrázkom, je podstatná matica P_rect ľavej, rovnica 5.8, a pravej, rovnica 5.9, kamery, z ktorých je možné vytvoriť reprojekčnú maticu Q s rozmermi 4x4, rovnica 5.10. Maticou Q sa následne transformuje disparitná mapa do troj-kanálového obrázka, predstavujúceho scénu v troch rozmeroch, rovnica 5.11 [2].

Výsledný obrázok sa nakoniec dá uložiť do súboru vo formáte .png ako hĺbková mapa alebo do súboru vo formáte .ply ako mračno bodov . Ukážka vytvorenia mračna bodov na obrázku 5.3.

$$P_{L}\_rect = \begin{bmatrix} f & 0 & cx_L & 0 \\ 0 & f & cy_L & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5.8}$$

$$P_{R}\_rect = \begin{bmatrix} f & 0 & cx_R & Tx * f \\ 0 & f & cy_R & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5.9}$$

$$Q = \begin{bmatrix} 1 & 0 & 0 & -cx_L \\ 0 & 1 & 0 & -cy_L \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/Tx & (cx_L - cx_R)/Tx \end{bmatrix} \tag{5.10}$$

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = Q * \begin{bmatrix} x \\ y \\ disp(x,y) \\ z \end{bmatrix} \tag{5.11}$$

kde:

- f: je ohnisková vzdialenosť kamier
- (cx_L, cy_L) a (cx_R, cy_R): sú koordináty hlavného bodu ľavej a pravej kamery
- Tx: je horizontálny posun medzi hlavnými bodmi kamier
- $disp(x, y)$: je disparita v bode (x, y)
- Z: je vypočítaná hĺbka v bode (x, y)

Obr. 5.3: Ukážka vytvorenia mračna bodov z mapy disparity. Prvý obrázok – pôvodný obrázok, druhý obrázok – ofarbená mapa disparity, tretí obrázok - ofarbená hĺbková mapa, štvrtý obrázok – vytvorené mračno bodov s namapovaným pôvodným obrázkom.

Kapitola 6

Implementácia

Riešenie som implementoval v jazyku Python, vo verzií 3.10.12 s použitím frameworku Pytorch [23], vo verzií 2.1.2+cu121.

V rámci riešenia som implementoval niekoľko skriptov:

- skript *train.py* obsahuje trénovací cyklus pre modely neurónových sietí, konfigurácia trénovania je uložená v súbore $config.yaml$
- skript *test.py* generuje predikované disparitné mapy vo formáte .png, pomocou zadaného natrénovaného modelu neurónovej siete
- skript $generate3D.py$ transformuje disparitné mapy vo formáte .png na hĺbkové mapy vo formáte .png a mračná bodov vo formáte .ply, vnútorné parametre kamery som získal z kalibračného súboru dátovej sady, tak ako je to popísané v kapitole 5.4. Mapu disparity som previedol na hĺbkovú mapu a mračno bodov pomocou frameworku OpenCV [2]
- skript *stat.py* vypočítava strednú hodnotu a štandardnú odchýlku pre zadanú dátovú sadu

6.1 Načítanie dátových sád

Pre každú použitú dátovú sadu som implementoval triedu zabezpečujúcu načítanie a augmentáciu danej dátovej sady v súbore $dataset\_loaders.py$. Každá trieda umožňuje zvoliť cestu k dátovej sade, vybrať fázu trénovania modelu a zvoliť, či sa má vykonať pokročilá augmentáciu. Augmentácia dátovej sady sa vykonáva dynamicky pri načítaní dát pomocou balíka $torchvision.transforms$.

Načítanie dátovej sady DrivingStereo zabezpečuje trieda $DrivingStereoDataset$, načítanie dátovej sady KITTI 2015 zabezpečuje trieda $KITTIDataset$ a načítanie dátovej sady Middlebury zabezpečuje trieda $MiddleburyDataset$, ktoré sú podtriedami triedy $Dataset$ balíka Pytorch.

6.2 Architektúry neurónových sietí

Architektúru siete ESNet som prevzal z pôvodnej metódy ES-Net [12], popísanej v kapitole 3.2. Architektúra ES-Net využíva aj vrstvy neurónovej siete z balíkov *channelnorm* a $resample2d$ z pytorch implementácie architektúry FlowNet2 [13], Flownet2-pytorch [24],

keďže tieto vrstvy neboli vo frameworku Pytorch implementované. V architektúre som nahradil niektoré zastarané funkcie použité v pôvodnej verzií.

Architektúru siete CREStereo, ktorá bola pôvodne implementovaná vo frameworku MegEngine [1], som previedol do frameworku Pytorch [23].

6.3 Chybové funkcie

Chybové funkcie som implementoval v súbore *criterions.py*. Chybová funkcia MultiscaleLoss pre architektúru ES-Net je implementovaná v triede MultiscaleLoss. Pôvodná verzia tejto chybovej funkcie používa ručne nastavené váhy, nastavené na trénovanie na štyroch kolách po siedmych epochách [12]. Keďže som svoje modely trénoval len jedno kolo, použil som nastavenie váh z metódy CREStereo [18], tak ako je popísané v kapitole 5.2.

Chybová funkcia SequenceLoss pre architektúru CREStereo je implementovaná v triede SequenceLoss. Funkciu som previedol do frameworku Pytorch podľa pôvodnej funkcie SequenceLoss použitej v metóde CREStereo [18].

Kapitola 7

Experimenty a výsledky

V tejto kapitole popíšem experimenty, ktoré som vykonal na modeloch architektúr ES-Net a CREStereo a porovnám ich výsledky.

Modely som trénoval na počítači s grafickou kartou Nvidia RTX 4060 s 8 GB pamäte VRAM a 16 GB pamäte RAM.

Na trénovanie všetkých modelov som použil optimalizačný algoritmus Adam s rýchlosťou učenia nastavenou na $1e-4$ pri prvotnom trénovaní na všetkých dátových sadách, rovnako ako v pôvodných metódach ES-Net a CREStereo [12, 18]. Model ES-Net som trénoval s chybovou funkciou MultiscaleLoss a model CREStereo s chybovou funkciou SequenceLoss.

Implementoval som tiež skoré zastavenie, ktoré pri trénovaní na dátovej sade DrivingStereo aktivuje, ak sa 2 epochy za sebou nezlepší testovacia presnosť na modely CREStereo a 5 epoch na modely ES-Net. Na dátovej sade KITTI 2015 sa skoré zastavenie aktivuje, ak sa 5 epoch testovacia presnosť nezlepší na modely CREStereo a 10 epoch na modely ES-Net. Pri prvej aktivácií skorého zastavenia sa nastaví rýchlosť učenia na $1e-5$ a tréning pokračuje ďalej. Pri druhej aktivácií skorého zastavenia tréning skončí.

Dátovú sadu KITTI 2015 som rozdelil na trénovaciu, testovaciu a evaluačnú sadu tak, že prvých 50 vzoriek som použil na evaluáciu, zo zvyšných 150 vzoriek som vybral 15 náhodných vzoriek na testovanie a zvyšných 135 vzoriek som použil na trénovanie. Pri dátovej sade DrivingStereo som použil pôvodné rozdelenie na trénovaciu sadu so 174431 vzorkami a testovaciu sadu so 7751 vzorkami. Veľkosť mini-batchu som nastavil u oboch modelov na 2 vzorky.

Vo všetkých experimentoch som po každej epoche modely validoval na testovacej dátovej sade a modely s najlepšou presnosťou na testovacej sade som ukladal.

7.1 Trénovanie na dátovej sade KITTI 2015 bez augmentácie

Ako prvý experiment som sa rozhodol vykonať trénovanie oboch modelov na menšej dátovej sade KITTI 2015 s použitím len základnej augmentácie – Z-score normalizácie a náhodného orezania vstupných obrázkov. Cieľom bolo overiť funkčnosť vybraných modelov a oboznámenie sa s nimi. Oba modely sa trénovali približne 50 epoch, pričom modelu ES-Net trvala jedna epocha na trénovacom stroji v priemere 110ms a modelu CREStereo v priemere 1.01s.

Napriek tomu, že išlo o malú trénovaciu sadu, metóda CREStereo dosiahla aj tak pomerne dobré výsledky, ktoré sú takmer zrovnateľné s niektorými metódami popísanými v kapitole 2.2. Výsledky z porovnávacieho testu KITTI sú v tabuľke 7.1.

Tabuľka 7.1: Výsledky na porovnávacom teste KITTI, pri trénovaní bez augmentácie na dátovej sade KITTI 2015.

	KITTI 2015		
	D1-bg	**D1-fg**	**D1-all**
ES-Net	19.02%	25.3%	20.06%
CREStereo	5.35%	10.85%	6.26%

Porovnanie výstupov z porovnávacích testov je zobrazené na obrázku 7.1. Hoci metóda CREStereo dosahuje dobré výsledky v samotnom teste, predikovaná disparitná mapa ešte nie je úplne ostrá. V porovnaní s metódou ES-Net je však metóda CREStereo lepšie schopná rozoznávať vzdialené objekty. Metóda ES-Net má okrem toho tiež problém rozoznať blízke objekty, ktoré sú tmavšie, napríklad v tieni a predikuje ich ako vzdialené.

7.2 Trénovanie na dátovej sade KITTI 2015 s augmentáciou

Cieľom druhého experimentu bolo pre mňa implementovať a otestovať augmentácie dát na dátovej sade KITTI 2015. Po naštudovaní ostatných metód, som sa rozhodol skombinovať augmentácie použité pri trénovaní pôvodných modelov ES-Net a CREStereo.

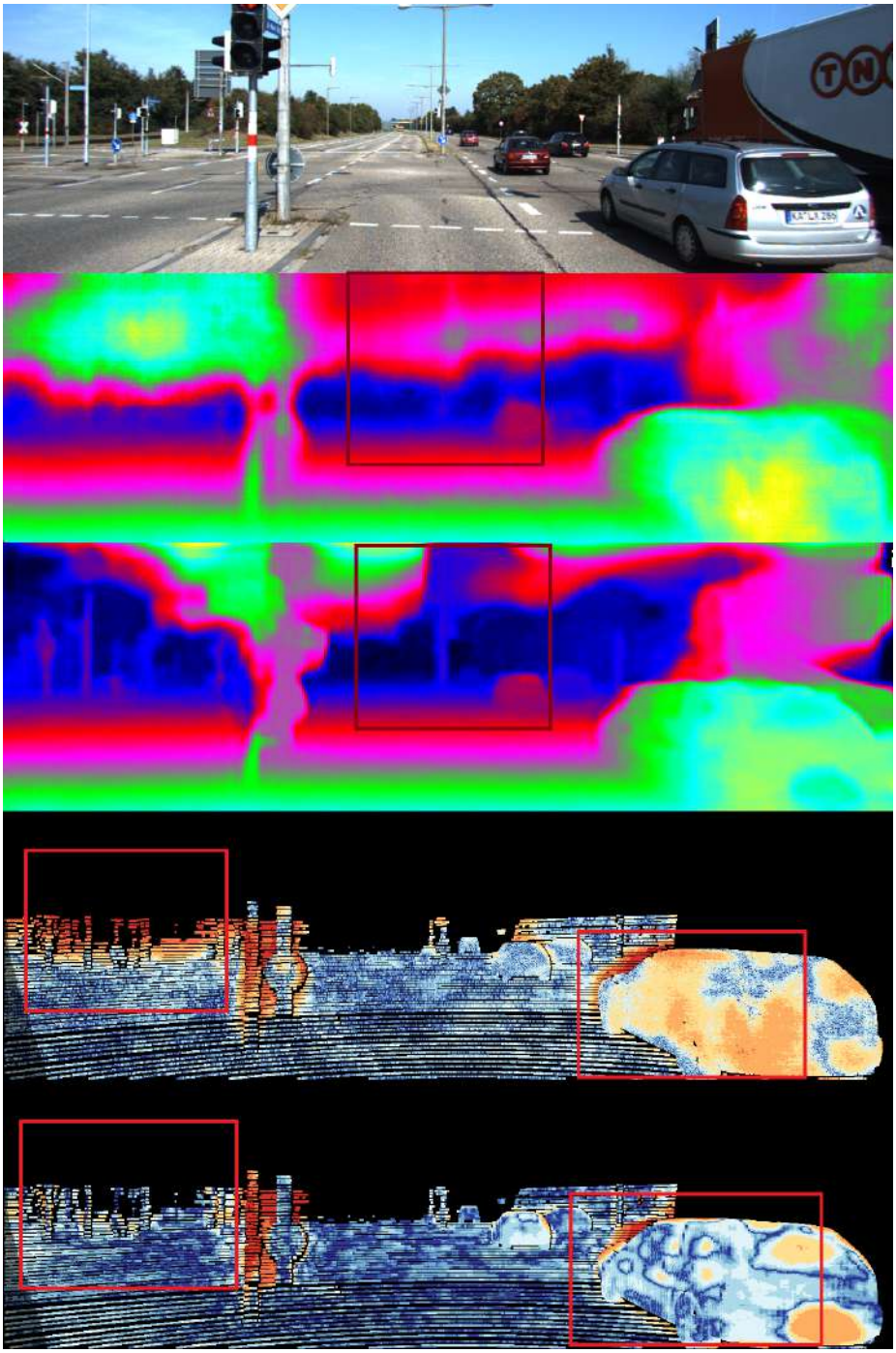
Po implementácií všetkých augmentácií popísaných v kapitole 5.1 a natrénovaní modelov na dátovej sade KITTI 2015, som zistil, že výsledná presnosť modelov sa zhoršila, pri metóde ES-Net až približne o 5% oproti modelu trénovanému bez augmentácií. Po ďalšom testovaní som zistil, že príčinou tohto zhoršenia sú augmentácie náhodné posunutie, pootočenie a zmena veľkosti pravého obrázka a náhodné vloženie oklúzie do pravého obrázka. Po odstránení jednej z týchto augmentácií, sa síce presnosť modelov mierne zlepšila, nie však až tak, ako pri odstránení oboch. Nakoniec som teda ďalej dátové sady augmentoval len pomocou Z-Score normalizácie, náhodným orezaním oboch vstupných obrázkov a náhodnou zmenou jasu, kontrastu a saturácie. Pri takejto augmentácií sa presnosť predikcie na dátovej sade KITTI 2015 zlepšila na modely ES-Net o viac ako 5% a na modely CREStereo o približne 1%. Výsledky trénovania s augmentáciami sú zobrazené v tabuľke 7.2.

Tabuľka 7.2: Výsledky modelov na porovnávacom teste KITTI, pri trénovaní s úplnou augmentáciou a len s augmentáciou farby.

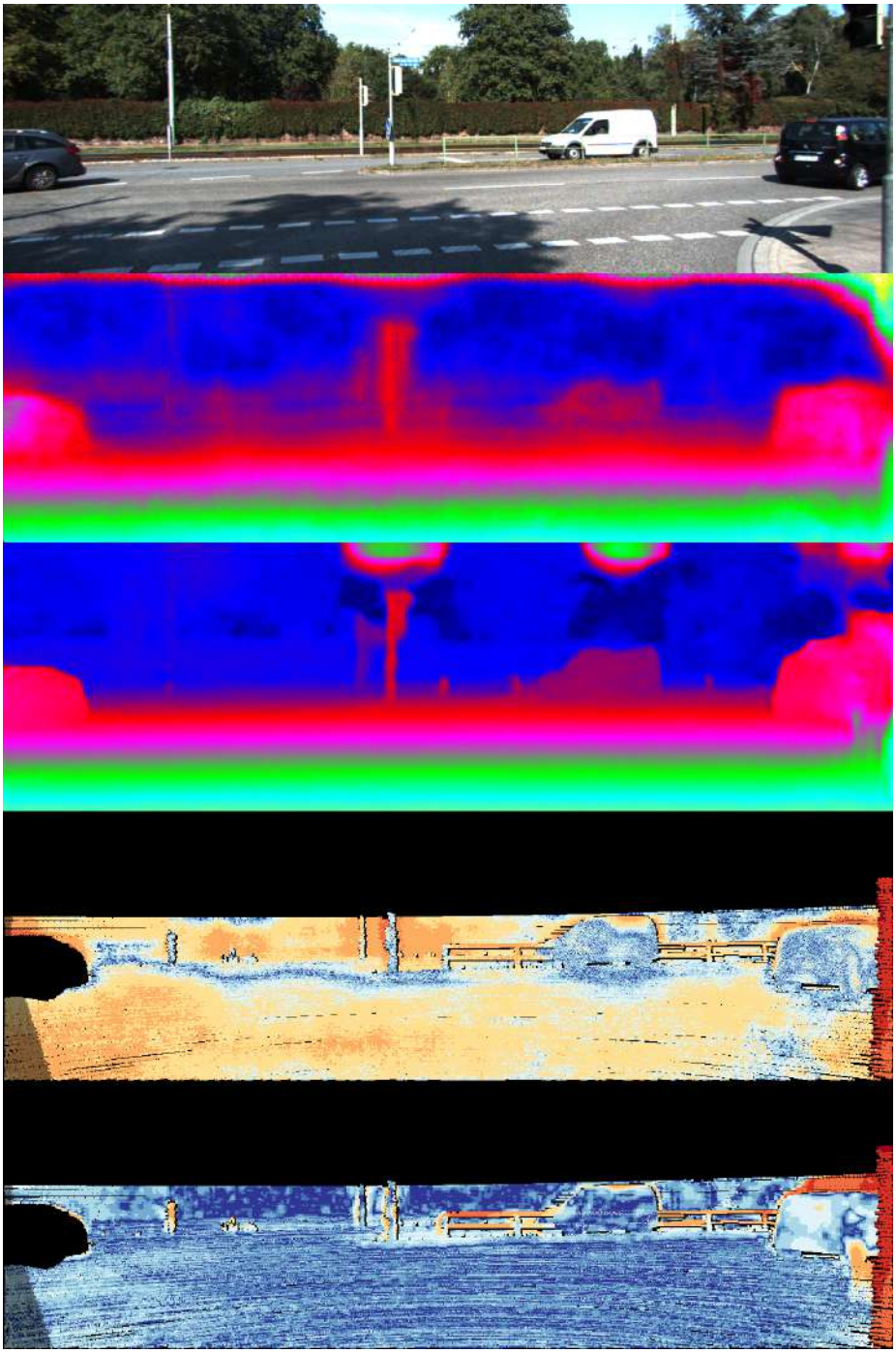
	KITTI 2015 - Úplná aug.			**KITTI 2015 - Aug. farby**		
	D1-bg	**D1-fg**	**D1-all**	**D1-bg**	**D1-fg**	**D1-all**
ES-Net	24.34%	29.36%	25.17%	12.35%	23.42%	14.19%
CREStereo	5.66%	9.78%	6.35%	4.61%	9.05%	5.35%

Na obrázku 7.2 sú zobrazené výstupy trénovania modelov s úplnou augmentáciou. Z chyby predikcie je jasne vidieť, že model ES-Net nezvláda trénovanie s úplnou augmentáciou. Ďalšou príčinou môže byť aj to, že na vstupnom obrázku sú objekty vo väčšej diaľke a na okrajoch obrázka, na väčšej časti obrázka je tiež zachytený tieň. Model CREStereo s daným vstupným obrázkom problém nemá, okrem stĺpu na pravom okraji obrázka, ktorý zle predikuje.

Pri porovnaní s obrázkom 7.3, ktorý zobrazuje tú istú vstupnú vzorku, ale trénovanie prebiehalo len s augmnetáciou farby je vidieť, že metóda ES-Net dosahuje o približne 45% lepšiu presnosť ako s úplnou augmentáciou. Metóda CREStereo je tiež o pár percent lepšia.



Obr. 7.1: Porovnanie výstupov trénovania modelov na dátovej sade KITTI 2015 bez augmentácie. Ukážka zachytáva výstup s najväčším rozdielom chyby predikcie medzi metódami. Metóda ES-Net dosiahla na tomto výstupe chybu D1-all 38.81% a metóda CREStereo dosiahla chybu D1-all 12.24%. Prvý obrázok – vstupný ľavý, druhý obrázok – disparitná mapa odhadnutá metódou ES-Net, tretí obrázok – disparitná mapa odhadnutá metódou CREStereou, štvrtý obrázok – chyba predikcie disparity metódou ES-Net, piaty obrázok – chyba predikcie disparity metódou CREStereo.



Obr. 7.2: Porovnanie výstupov trénovania modelov na dátovej sade KITTI 2015 s úplnou augmentáciou tak, ako som ju popísal v kapitole 5.1. Ukážka zachytáva výstup s najväčším rozdielom chyby predikcie medzi metódami. Metóda ES-Net dosiahla na tomto výstupe chybu D1-all 56.82% a metóda CREStereo dosiahla chybu D1-all 4.14%. Prvý obrázok – vstupný ľavý, druhý obrázok – disparitná mapa odhadnutá metódou ES-Net, tretí obrázok – disparitná mapa odhadnutá metódou CREStereou, štvrtý obrázok – chyba predikcie disparity metódou ES-Net, piaty obrázok – chyba predikcie disparity metódou CREStereo.

Na obrázku 7.4 je zachytený výstup z trénovania na ktorom dosiahli obe metódy najlepší výsledok pri trénovaní na KITTI dátovej sade s augmentáciou farby. Z obrázku je vidieť, že metóda ES-Net má menšie problémy s predikciou vzdialenejších objektov a opakujúcich sa textúr ako je tráva. Hoci metóda CREStereo mala podľa disparitnej mapy problém s odhadom vzdialenosti oblohy, táto časť obrázku nebola braná do úvahy v porovnávacom teste a metóda dosiahla chyby predikcie 1.54%.

7.3 Trénovanie na dátovej sade DrivingStereo a KITTI 2015 s augmentáciou farby

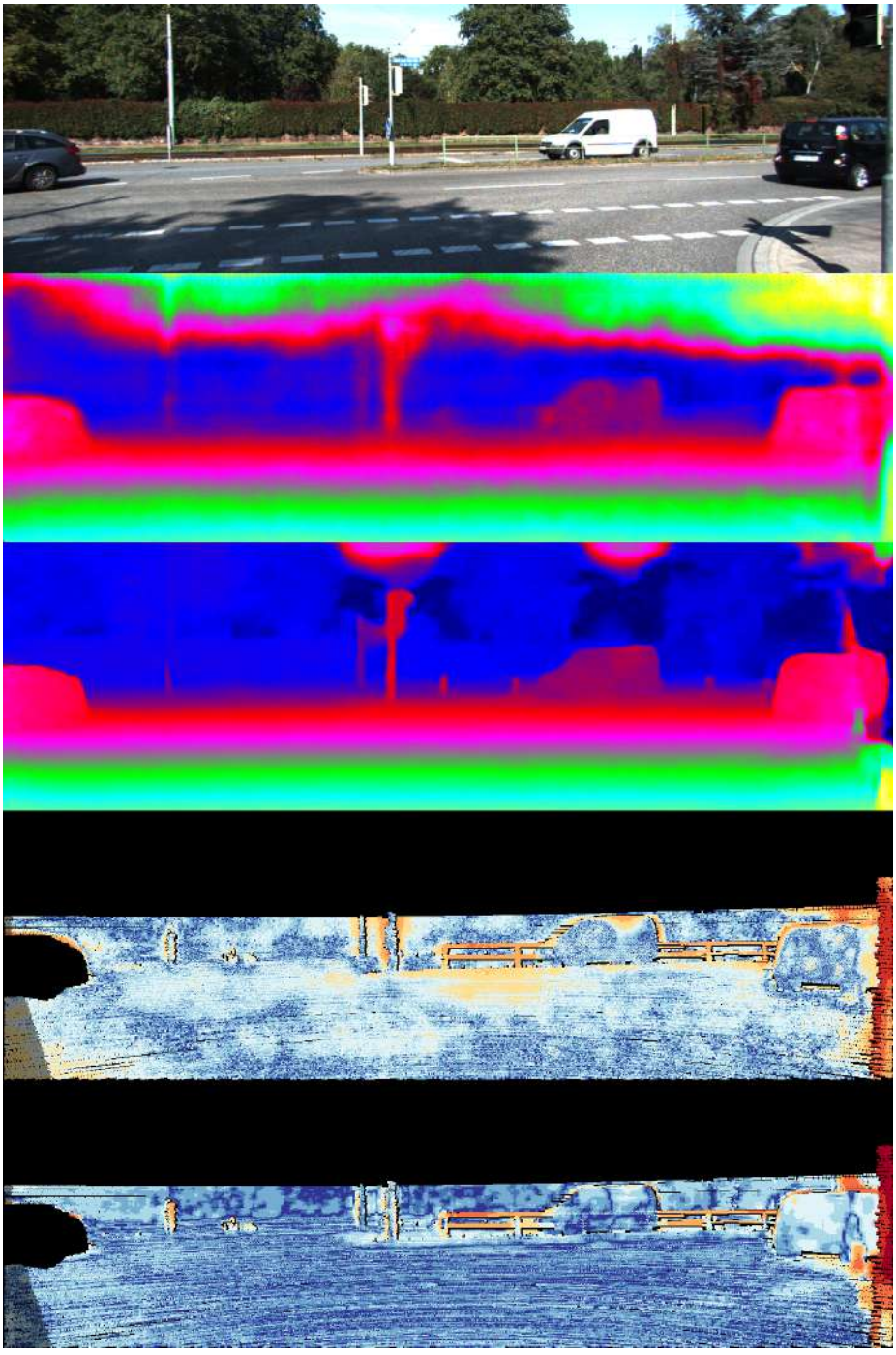
V poslednom experimente bol môj cieľ natrénovať modely na veľkej dátovej sade DrivingStereo s použitím augmentácie a následne modely dotrénovať opäť na dátovej sade KITTI 2015. Po zisteniach z predchádzajúceho experimentu, som sa rozhodol použiť len augmentáciu farby aj pri trénovaní na veľkej dátovej sade.

Vďaka trénovaniu na veľkej dátovej sade sa modelu ES-Net podarilo dosiahnuť porovnateľné výsledky s modelom CREStereo, avšak po do-trénovaní na dátovej sade KITTI 2015 model ES-Net opäť o približne 1% predbehol, z čoho vyplýva, že model CREStereo sa dokáže lepšie natrénovať aj na malých dátových sadách. Mnou natrénované modely CRES-Net aj ES-Net dosiahli výsledky porovnateľné so state-of-the-art modelmi. Výsledky na porovnávacom teste KITTI sú zobrazené v tabuľke 7.3.

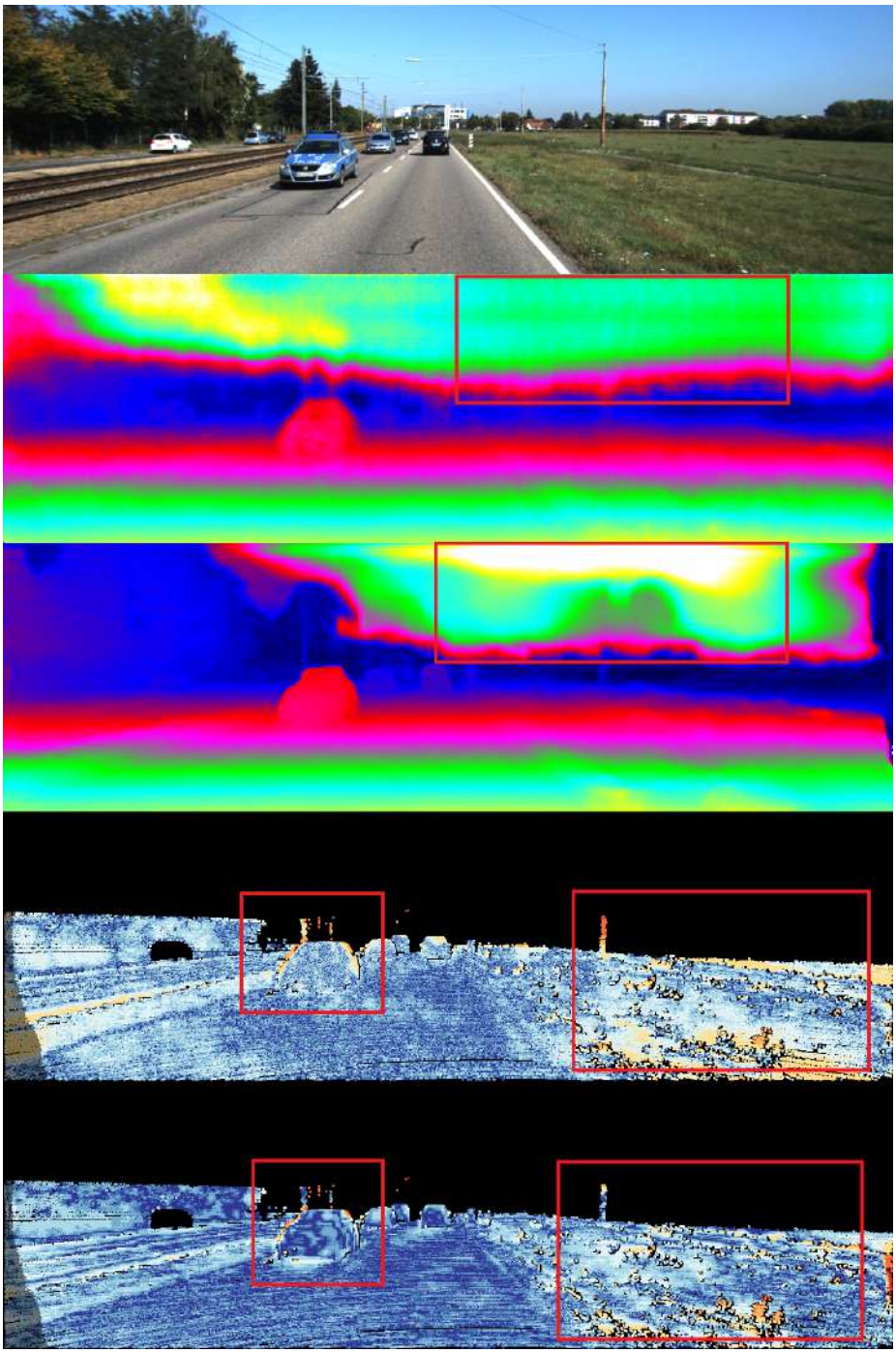
Tabuľka 7.3: Výsledky modelov na porovnávacom teste KITTI, pri trénovaní na dátovej sade DrivingStereo a pri do-trénovaní tých istých modelov na dátovej sade KITTI 2015.

	DrivingStereo			**DrivingStereo + KITTI 2015**		
	D1-bg	**D1-fg**	**D1-all**	**D1-bg**	**D1-fg**	**D1-all**
ES-Net	3.58%	6.92%	4.13%	3.1%	3.6%	3.18%
CREStereo	3.31%	7.39%	3.99%	2.12%	2.35%	2.16%

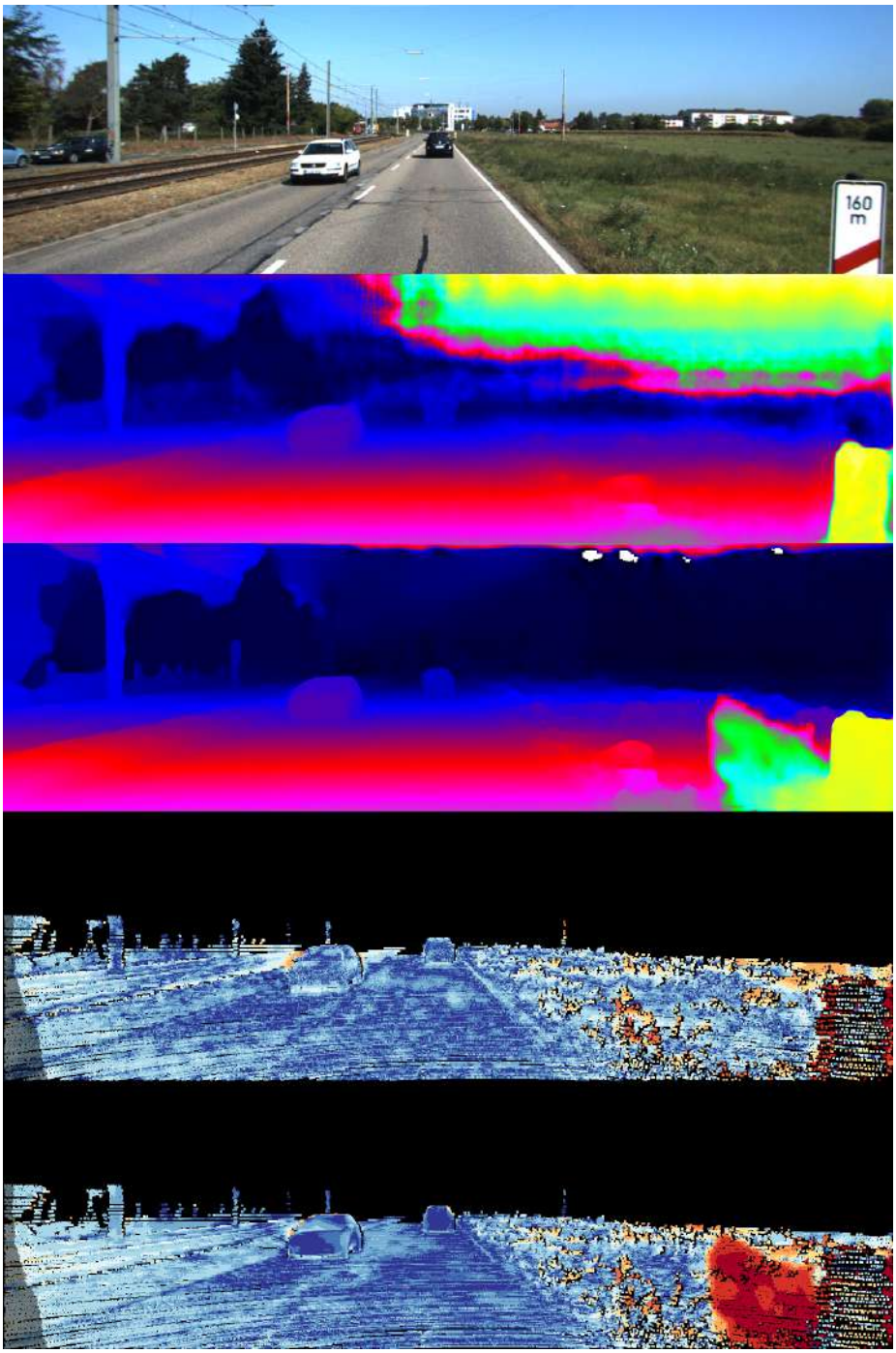
Na obrázku 7.5 je vidieť, že veľká dátová sada pomohla modelu ES-Net sa dostať na úroveň modelu CREStereo, kedy ho prekonal o viac ako 5%. Ďalej môžeme vidieť, že metóda CREStereo má problémy správne predikovať vzdialenosť objektov blízko pri okraji obrázku. Ukážka hĺbkovej mapy a mračna bodov je na obrázku 7.6.



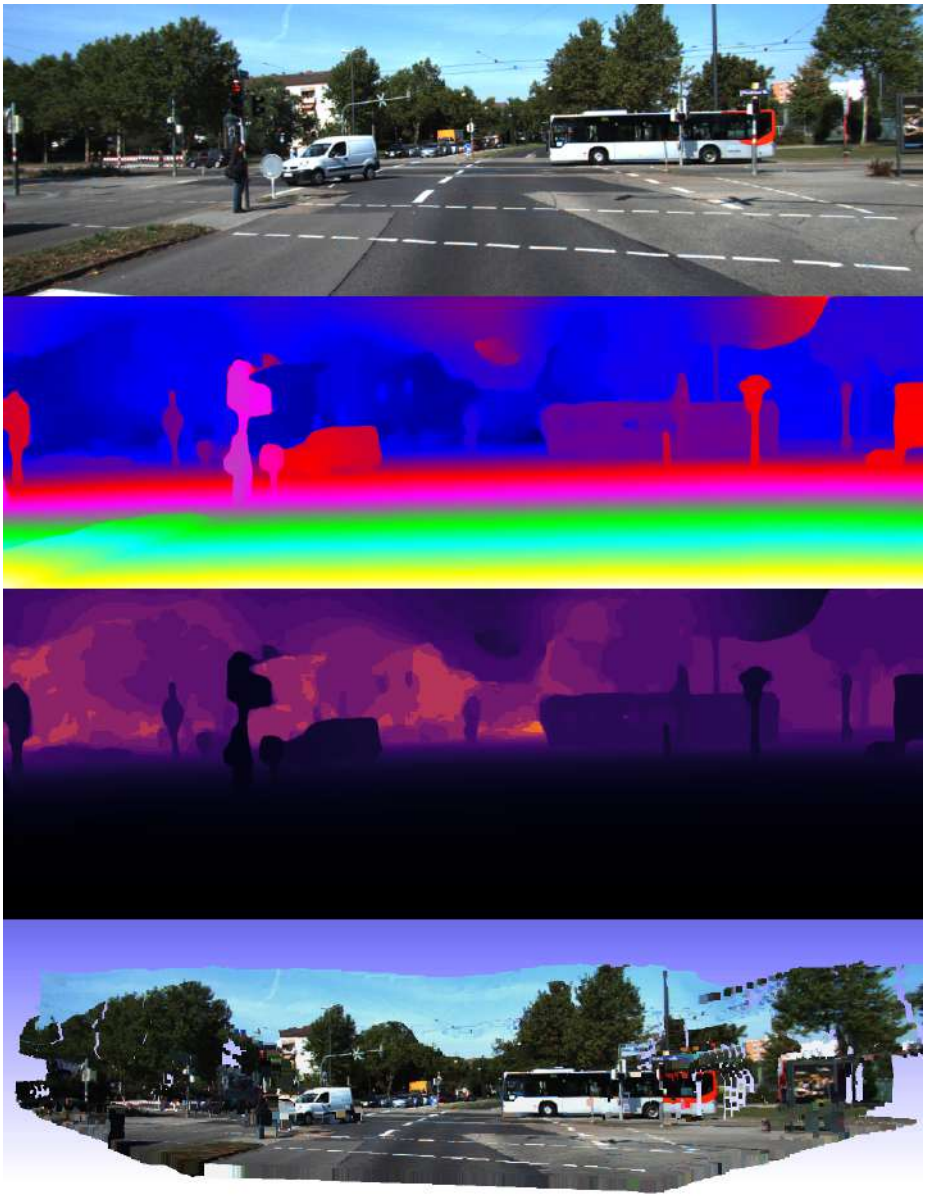
Obr. 7.3: Porovnanie výstupov trénovania modelov na dátovej sade KITTI 2015 s augmentáciou farby. Ukážka zachytáva rovnakú vzorku ako je na obrázku 7.2, pre porovnanie výsledkov po odstránení afinných augmentácií a vkladania oklúzie. Metóda ES-Net dosiahla na tomto výstupe chybu D1-all 11.86% a metóda CREStereo dosiahla chybu D1-all 3.58%. Prvý obrázok – vstupný ľavý, druhý obrázok – disparitná mapa odhadnutá metódou ES-Net, tretí obrázok – disparitná mapa odhadnutá metódou CREStereou, štvrtý obrázok – chyba predikcie disparity metódou ES-Net, piaty obrázok – chyba predikcie disparity metódou CREStereo.



Obr. 7.4: Porovnanie výstupov trénovania modelov na dátovej sade KITTI 2015 s augmentáciou farby. Ukážka zachytáva výstup s najmenšou chybou predikcie u oboch metód. Metóda ES-Net dosiahla na tomto výstupe chybu D1-all 5.45% a metóda CREStereo dosiahla chybu D1-all 1.54%. Prvý obrázok – vstupný ľavý, druhý obrázok – disparitná mapa odhadnutá metódou ES-Net, tretí obrázok – disparitná mapa odhadnutá metódou CREStereou, štvrtý obrázok – chyba predikcie disparity metódou ES-Net, piaty obrázok – chyba predikcie disparity metódou CREStereo.



Obr. 7.5: Porovnanie výstupov trénovania modelov na dátovej sade DrivingStereo s augmentáciou farby. Ukážka zachytáva výstup s najväčším rozdielom chyby predikcie medzi metódami. Metóda ES-Net dosiahla na tomto výstupe chybu D1-all 3.79% a metóda CREStereo dosiahla chybu D1-all 9.22%. Prvý obrázok – vstupný ľavý, druhý obrázok – disparitná mapa odhadnutá metódou ES-Net, tretí obrázok – disparitná mapa odhadnutá metódou CREStereou, štvrtý obrázok – chyba predikcie disparity metódou ES-Net, piaty obrázok – chyba predikcie disparity metódou CREStereo.



Obr. 7.6: Ukážka hĺbkovej mapy a mračna bodov vytvoreného z disparitnej mapy predikovanej model CREStereo natrénovanom na dátovej sade DrivingStereo a KITTI 2015. Prvý obrázok – vstupný ľavý, druhý obrázok – disparitná mapa odhadnutá metódou ES-Net, tretí obrázok – disparitná mapa odhadnutá metódou CREStereou, štvrtý obrázok – chyba predikcie disparity metódou ES-Net, piaty obrázok – chyba predikcie disparity metódou CREStereo.

Kapitola 8

Záver

V tejto práci som sa venoval rekonštrukcií hĺbky z obrazu pomocou hlbokých neurónových sietí. Oboznámil som sa s metódami a princípmi, ktoré sa používajú pri riešení stereo rekonštrukcie, či už ide o klasické princípy založené na extrakcií príznakov a ich párovaní medzi obrázkami, až po neurónové siete, ktoré nahradzujú jednotlivé klasické kroky stereo rekonštrukcie až po metódy hlbokého učenia, ktoré sa stávajú čím ďalej tým viac kľúčovejšími pri riešení tohoto problému.

V rámci riešenia tejto práce som sa zameral na porovnanie niektorých metód hlbokého učenia, určených na odhad disparity z párov obrázkov. Konkrétne som si vybral metódy ES-Net a CREStereo, ktoré pristupujú k tomuto problému mierne odlišne. ES-Net využíva dnes už štandardné princípy spracovania obrazu neurónovými sieťami, reziduálne bloky a konvolučné vrstvy. CREStereo okrem toho používa iteratívne princípy využívajúce rekurentné vrstvy pre tvorbu a vylaďovanie predikcií hĺbky.

Natrénoval som, otestoval a porovnal som niekoľko modelov týchto architektúr na rôznych dátových sadách a s rozličnou mierou augmentácie dát. Niektoré augmentačné metódy ako je affínne transformácie obrazu a vkladanie oklúzií do obrazu sa pri trénovaní neosvedčili a modely s nimi dosahovali horšie výsledky ako bez augmentácie.

Z experimentov vyplynulo, že metóda CREStereo dosahuje takmer vo všetkých prípadoch lepšie výsledky ako metóda ES-net, a tiež sa dokáže efektívnejšie natrénovať na malej dátovej sade. Avšak metóda ES-Net je taktiež schopná state-of-the-art výsledkov pri trénovaní na veľkej dátovej sade, okrem toho v porovnaní s metódou CREStereo je niekoľkonásobne rýchlejšia pri inferencií a samotnom trénovacom cykle.

Literatúra

[1] *MegEngine:A fast, scalable and easy-to-use deep learning framework*. Dostupné z: https://github.com/MegEngine/MegEngine.

[2] *The OpenCV Reference Manual*. 4.7.0. OpenCV, December 2022 [cit. 10-5-2023]. Dostupné z: https://docs.opencv.org/4.7.0/d9/d0c/group__calib3d.html.

[3] Bernhardt, S., Abi Nahed, J. a Abugharbieh, R. Robust Dense Endoscopic Stereo Reconstruction for Minimally Invasive Surgery. In: Menze, B. H., Langs, G., Lu, L., Montillo, A., Tu, Z. et al., ed. *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, s. 254–262. ISBN 978-3-642-36620-8.

[4] Birchfield, S. a Tomasi, C. Depth discontinuities by pixel-to-pixel stereo. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Bombay, India: IEEE Computer Society, Jan 1998, s. 1073–1080. DOI: 10.1109/ICCV.1998.710850. ISBN 0-8186-6265-4. Dostupné z: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=710850.

[5] Chang, J.-R. a Chen, Y.-S. Pyramid Stereo Matching Network. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2018, s. 5410–5418. DOI: 10.1109/CVPR.2018.00567. Dostupné z: https://arxiv.org/pdf/1803.08669.pdf.

[6] Cheng, X., Zhong, Y., Harandi, M., Dai, Y., Chang, X. et al. Hierarchical Neural Architecture Search for Deep Stereo Matching. In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. 2020. Dostupné z: https://arxiv.org/pdf/2010.13501.pdf.

[7] Duggal, S., Wang, S., Ma, W., Hu, R. a Urtasun, R. DeepPruner: Learning Efficient Stereo Matching via Differentiable PatchMatch. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, Nov 2019, s. 4383–4392. DOI: 10.1109/ICCV.2019.00448. Dostupné z: https://arxiv.org/pdf/1909.05845.pdf.

[8] Guo, C., Chen, D. a Huang, Z. Learning Efficient Stereo Matching Network With Depth Discontinuity Aware Super-Resolution. *IEEE Access*. IEEE Computer Society. nov 2019, zv. 7, s. 159712–159723. DOI: 10.1109/ACCESS.2019.2950924. ISSN 2169-3536. Dostupné z: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8890688.

[9] Guo, X., Yang, K., Yang, W., Wang, X. a Li, H. Group-Wise Correlation Stereo Network. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2019, s. 3268–3277. DOI: 10.1109/CVPR.2019.00339. Dostupné z: https://arxiv.org/pdf/1903.04025.pdf.

[10] Hannah, M. J. *Computer matching of areas in stereo images.* Stanford, CA, USA, 1974. Dizertačná práca. Stanford University, CA, Department of computer science. AAI7427032. Dostupné z: https://apps.dtic.mil/sti/pdfs/AD0786720.pdf.

[11] Hua, Y., Kohli, P., Uplavikar, P., Ravi, A., Gunaseelan, S. et al. Holopix50k: A Large-Scale In-the-wild Stereo Image Dataset. In: *CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Seattle, WA, 2020.* June 2020. Dostupné z: https://arxiv.org/pdf/2003.11172.

[12] Huang, Z., Norris, T. B. a Wang, P. ES-Net: An Efficient Stereo Matching Network. *ArXiv e-prints*. 2021. DOI: 10.48550/ARXIV.2103.03922. Dostupné z: https://arxiv.org/abs/2103.03922.

[13] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A. et al. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jul 2017, s. 1647–1655. DOI: 10.1109/CVPR.2017.179. ISSN 1063-6919. Dostupné z: https://arxiv.org/pdf/1612.01925.

[14] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R. et al. End-to-End Learning of Geometry and Context for Deep Stereo Regression. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct 2017, s. 66–75. DOI: 10.1109/ICCV.2017.17. ISSN 2380-7504. Dostupné z: https://arxiv.org/pdf/1703.04309.pdf.

[15] Khamis, S., Fanello, S., Rhemann, C., Kowdle, A., Valentin, J. et al. StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction. In: Ferrari, V., Hebert, M., Sminchisescu, C. a Weiss, Y., ed. *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, 2018, s. 596–613. ISBN 978-3-030-01267-0. Dostupné z: https://arxiv.org/pdf/1807.08865.pdf.

[16] Laga, H., Jospin, L. V., Boussaid, F. a Bennamoun, M. A Survey on Deep Learning Techniques for Stereo-Based Depth Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Los Alamitos, CA, USA: IEEE Computer Society. apr 2022, zv. 44, č. 4, s. 1738–1764. DOI: 10.1109/TPAMI.2020.3032602. ISSN 1939-3539. Dostupné z: https://arxiv.org/pdf/2006.02535.

[17] Lee, H. a Shin, Y. Real-Time Stereo Matching Network with High Accuracy. In: *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei, Taiwan: IEEE Computer Society, Sep 2019, s. 4280–4284. DOI: 10.1109/ICIP.2019.8803514. ISBN 978-1-5386-6249-6. Dostupné z: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8803514.

[18] Li, J., Wang, P., Xiong, P., Cai, T., Yan, Z. et al. Practical Stereo Matching via Cascaded Recurrent Network with Adaptive Correlation. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2022, s. 16242–16251. DOI: 10.1109/CVPR52688.2022.01578. Dostupné z: https://arxiv.org/pdf/2203.11483.

[19] Liang, Z., Feng, Y., Guo, Y., Liu, H., Chen, W. et al. Learning for Disparity Estimation Through Feature Constancy. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2018, s. 2811–2820. DOI: 10.1109/CVPR.2018.00297. Dostupné z: https://arxiv.org/pdf/1712.01039.pdf.

[20] Lipson, L., Teed, Z. a Deng, J. RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching. In: *2021 International Conference on 3D Vision (3DV)*. Los Alamitos, CA, USA: IEEE Computer Society, Dec 2021, s. 218–227. DOI: 10.1109/3DV53792.2021.00032. Dostupné z: https://arxiv.org/pdf/2109.07547.pdf.

[21] Mayer, N., Ilg, E., Häusser, P., Fischer, P., Cremers, D. et al. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2016, s. 4040–4048. DOI: 10.1109/CVPR.2016.438. ISSN 1063-6919. Dostupné z: https://arxiv.org/pdf/1512.02134.

[22] Menze, M. a Geiger, A. Object scene flow for autonomous vehicles. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2015, s. 3061–3070. DOI: 10.1109/CVPR.2015.7298925. ISSN 1063-6919. Dostupné z: https://www.cvlibs.net/publications/Menze2015CVPR.pdf.

[23] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, s. 8024–8035. Dostupné z: https://arxiv.org/pdf/1912.01703.

[24] Reda, F., Pottorff, R., Barker, J. a Catanzaro, B. *Flownet2-pytorch: Pytorch implementation of FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks*. GitHub, 2017. Dostupné z: https://github.com/NVIDIA/flownet2-pytorch.

[25] Scharstein, D. Matching images by comparing their gradient fields. In: *Proceedings of 12th International Conference on Pattern Recognition*. Jerusalem, Israel: IEEE Computer Society, Oct 1994, sv. 1, s. 572–575 vol.1. DOI: 10.1109/ICPR.1994.576363. ISBN 81-7319-221-9. Dostupné z: https://www.cs.middlebury.edu/~schar/papers/gradient.pdf.

[26] Scharstein, D., Szeliski, R. a Zabih, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. Kauai, HI, USA: IEEE Computer Society, Dec 2001, s. 131–140. DOI: 10.1109/SMBV.2001.988771. ISBN 0-7695-1327-1. Dostupné z: https://vision.middlebury.edu/stereo/taxonomy-IJCV.pdf.

[27] Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N. et al. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. In: Jiang, X., Hornegger, J. a Koch, R., ed. *Pattern Recognition*. Cham: Springer International Publishing, 2014, s. 31–42. ISBN 978-3-319-11752-2. Dostupné z: https://www.cs.middlebury.edu/~schar/papers/datasets-gcpr2014.pdf.

[28] Schöps, T., Schönberger, J. L., Galliani, S., Sattler, T., Schindler, K. et al. A Multi-view Stereo Benchmark with High-Resolution Images and Multi-camera Videos. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jul 2017, s. 2538–2547. DOI: 10.1109/CVPR.2017.272. ISSN 1063-6919. Dostupné z: https://www.cvlibs.net/publications/Schoeps2017CVPR.pdf.

[29] Shen, Z., Dai, Y. a Rao, Z. CFNet: Cascade and Fused Cost Volume for Robust Stereo Matching. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2021, s. 13901–13910. DOI: 10.1109/CVPR46437.2021.01369. Dostupné z: https://arxiv.org/pdf/2104.04314.pdf.

[30] Song, X., Yang, G., Zhu, X., Zhou, H., Wang, Z. et al. AdaStereo: A Simple and Efficient Approach for Adaptive Stereo Matching. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2021, s. 10323–10332. DOI: 10.1109/CVPR46437.2021.01019. Dostupné z: https://arxiv.org/pdf/2004.04627.pdf.

[31] Song, X., Zhao, X., Hu, H. et al. Edgestereo: A context integrated residual pyramid network for stereo matching. In: Springer. *Asian Conference on Computer Vision*. 2018, s. 20–35. Dostupné z: https://arxiv.org/pdf/1803.05196.pdf.

[32] Sun, D., Yang, X., Liu, M.-Y. a Kautz, J. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2018, s. 8934–8943. DOI: 10.1109/CVPR.2018.00931. Dostupné z: https://arxiv.org/pdf/1709.02371.

[33] Taniai, T., Matsushita, Y., Sato, Y. a Naemura, T. Continuous 3D Label Stereo Matching Using Local Expansion Moves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE Computer Society. oct 2018, zv. 40, č. 11, s. 2725–2739. DOI: 10.1109/TPAMI.2017.2766072. ISSN 1939-3539. Dostupné z: https://arxiv.org/pdf/1603.08328.pdf.

[34] Tankovich, V., Häne, C., Zhang, Y., Kowdle, A., Fanello, S. et al. HITNet: Hierarchical Iterative Tile Refinement Network for Real-time Stereo Matching. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2021, s. 14357–14367. DOI: 10.1109/CVPR46437.2021.01413. Dostupné z: https://arxiv.org/pdf/2007.12140.pdf.

[35] Tonioni, A., Tosi, F., Poggi, M., Mattoccia, S. a Stefano, L. D. Real-Time Self-Adaptive Deep Stereo. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun

2019, s. 195–204. DOI: 10.1109/CVPR.2019.00028. Dostupné z: https://arxiv.org/pdf/1810.05424.pdf.

[36] Wang, J., Jampani, V., Sun, D. et al. Improving deep stereo network generalization with geometric priors. *ArXiv preprint arXiv:2008.11098*. 2020. Dostupné z: https://arxiv.org/pdf/2008.11098.pdf.

[37] Wang, Q., Shi, S., Zheng, S., Zhao, K. a Chu, X. FADNet: A Fast and Accurate Network for Disparity Estimation. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Paris, France: IEEE Computer Society, May-sep 2020, s. 101–107. DOI: 10.1109/ICRA40945.2020.9197031. ISBN 978-1-7281-7395-5. Dostupné z: https://arxiv.org/pdf/2003.10758.pdf.

[38] Xu, G., Cheng, J., Guo, P. a Yang, X. Attention Concatenation Volume for Accurate and Efficient Stereo Matching. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2022, s. 12971–12980. DOI: 10.1109/CVPR52688.2022.01264. Dostupné z: https://arxiv.org/pdf/2203.02146.

[39] Xu, H. a Zhang, J. AANet: Adaptive Aggregation Network for Efficient Stereo Matching. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, s. 1956–1965. DOI: 10.1109/CVPR42600.2020.00203. Dostupné z: https://arxiv.org/pdf/2004.09548.pdf.

[40] Yan, T., Gan, Y., Xia, Z. a Zhao, Q. Segment-Based Disparity Refinement With Occlusion Handling for Stereo Matching. *IEEE Transactions on Image Processing*. Los Alamitos, CA, USA: IEEE Computer Society. jun 2019, zv. 28, č. 8, s. 3885–3897. DOI: 10.1109/TIP.2019.2903318. Dostupné z: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8661596.

[41] Yang, G., Manela, J., Happold, M. a Ramanan, D. Hierarchical Deep Stereo Matching on High-Resolution Images. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE Computer Society, June 2019, s. 5510–5519. DOI: 10.1109/CVPR.2019.00566. ISBN 978-1-7281-3293-8. Dostupné z: https://arxiv.org/pdf/1912.06704.pdf.

[42] Yang, G., Song, X., Huang, C., Deng, Z., Shi, J. et al. DrivingStereo: A Large-Scale Dataset for Stereo Matching in Autonomous Driving Scenarios. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2019, s. 899–908. DOI: 10.1109/CVPR.2019.00099. Dostupné z: https://openaccess.thecvf.com/content_CVPR_2019/papers/Yang_DrivingStereo_A_Large-Scale_Dataset_for_Stereo_Matching_in_Autonomous_Driving_CVPR_2019_paper.pdf.

[43] Yee, K. a Chakrabarti, A. Fast Deep Stereo with 2D Convolutional Processing of Cost Signatures. In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Los Alamitos, CA, USA: IEEE Computer Society, Mar 2020, s. 183–191. DOI: 10.1109/WACV45572.2020.9093273. Dostupné z: https://arxiv.org/pdf/1903.04939.pdf.

[44] Zhang, F., Prisacariu, V., Yang, R. a Torr, P. S. GA-Net: Guided Aggregation Net for End-To-End Stereo Matching. In: *2019 IEEE/CVF Conference on Computer*

Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, Jun 2019, s. 185–194. DOI: 10.1109/CVPR.2019.00027. Dostupné z: https://arxiv.org/pdf/1904.06587.pdf.

[45] Zhang, Y., Chen, Y., Bai, X. et al. Adaptive unimodal cost volume filtering for deep stereo matching. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, sv. 34, s. 12926–12934. Dostupné z: https://arxiv.org/pdf/1909.03751.pdf.

[46] Žbontar, J. a LeCun, Y. Computing the stereo matching cost with a convolutional neural network. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE Computer Society, June 2015, s. 1592–1599. DOI: 10.1109/CVPR.2015.7298767. ISBN 978-1-4673-6964-0. Dostupné z: https://arxiv.org/pdf/1409.4326.

Príloha A

Obsah přiloženého paměťového média

- **/src** Zdrojové súbory a skripty pre trénovanie a evaluáciu modelov
- **/datasets** Dátové sady používané pri trénovaní a evaluácií modelov
- **/results** Výstupy z trénovania a evaluácie, vytvorené hĺbkové mapy a mračná bodov
- **plagat.pdf** Plagát k projektu
- **/doc_src** Zdrojové kódy k tejto dokumentácií
- **dokumentacia.pdf** Táto dokumentácia k diplomovej práci