



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

**DETEKCE VÝZNAMNÝCH BODŮ V MEDICÍNSKÝCH
OBRAZECH POMOCÍ HLUBOKÝCH NEURONOVÝCH
SÍTÍ**

LANDMARK DETECTION IN MEDICAL IMAGES USING DEEP NEURAL NETWORKS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

JURAJ ŠKANDERA

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. OLDŘICH KODYM

BRNO 2018

Vysoké učení technické v Brně - Fakulta informačních technologií

Ústav počítačové grafiky a multimédií

Akademický rok 2017/2018

Zadání bakalářské práce

Řešitel: **Škandera Juraj**

Obor: Informační technologie

Téma: **Detekce významných bodů v medicínských obrazech pomocí hlubokých neuronových sítí**

Landmark Detection in Medical Images Using Deep Neural Networks

Kategorie: Zpracování obrazu

Pokyny:

1. Seznamte se s problematikou hlubokých neuronových sítí a jejich učení.
2. Zorientujte se v metodách detekce anatomických bodů (též významných bodů nebo landmark points) v medicínských obrazových datech (CT data, RTG snímky, apod.) s využitím hlubokých neuronových sítí.
3. Vyberte vhodnou metodu použitelnou pro řešení zvoleného problému detekce významných bodů v medicínském obraze.
4. Implementujte navrženou metodu s využitím existujících nástrojů pro trénování hlubokých neuronových sítí.
5. Provedte experimenty nad připravenou datovou sadou.
6. Porovnejte dosažené výsledky a diskutujte možnosti budoucího vývoje.
7. Vytvořte stručný plakát nebo video prezentující vaši práci, její cíle a výsledky.

Literatura:

- Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3.

Podrobné závazné pokyny pro vypracování bakalářské práce naleznete na adrese <http://www.fit.vutbr.cz/info/szz/>

Technická zpráva bakalářské práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap (20 až 30% celkového rozsahu technické zprávy).

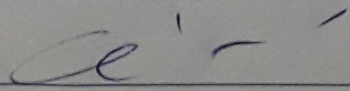
Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Kodym Oldřich, Ing.**, UPGM FIT VUT

Datum zadání: 1. listopadu 2017

Datum odevzdání: 16. května 2018

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav počítačové grafiky a multimédií
602 00 Brno, Lazetěcheva 2


doc. Dr. Ing. Jan Černocký
vedoucí ústavu

Abstrakt

Táto práca sa zaoberá detekciou anatomických bodov z cefalometrických RTG snímok pomocou konvolučných neurónových sietí. Program pracuje nad verejne dostupným datasetom, ktorý obsahuje bočné RTG snímky lebky. V rámci práce sú navrhnuté dve architektúry konvolučných neurónových sietí. Najlepšia architektúra dosahuje úspešnosť 73.22% pre detekciu do 5 mm. Program je tvorený v jazyku Python s využitím knižnice Tensorflow.

Abstract

This thesis deals with detection of anatomical landmarks from cephalometric X-ray images using convolutional neural networks. Program works with public available dataset, which consists of side X-ray images of skull. There are two architectures of convolutional neural networks proposed in this thesis. The best architecture achieves accuracy of 73.22% for detection within 5 mm. Program is created in Python language with use of Tensorflow framework.

Kľúčové slová

cefalometrické landmarky, konvolučné neurónové siete, detekcia landmarkov, hlboké učenie, cefalogram

Keywords

cephalometric landmarks, convolutional neural networks, landmark detection, deep learning, cephalogram

Citácia

ŠKANDERA, Juraj. *Detekce významných bodů v medicínských obrazech pomocí hlubokých neuronových sítí*. Brno, 2018. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Oldřich Kodým

Detekce významných bodů v medicínských obrazech pomocí hlubokých neuronových sítí

Prehlásenie

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne pod vedením Ing. Oldřicha Kodyma. Další informace mi poskytl Ing. Michal Španěl PhD. Uviedol som všetky publikácie a literárne pramene, z ktorých som čerpal

.....
Juraj Škandera
16. mája 2018

Podakovanie

Chcel by som poďakovať pánovi Ing. Oldřichovi Kodymovi za pripomenky, ochotu a trpezlivosť v priebehu vedenia mojej práce a pánovi Ing. Michalovi Španělovi PhD. za prvotné nasmerovanie pre vypracovanie tejto práce.

Obsah

1	Úvod	2
2	Dopredné neurónové siete	3
2.1	Neurón	3
2.2	Aktivačná funkcia	4
2.3	Architektúra siete	6
2.4	Učenie neurónových sietí	6
2.5	Pretrénovanie	8
2.6	Konvolučné neurónové siete	9
3	Detekcia Landmarkov	12
3.1	Detekcia landmarkov tváre	12
3.2	Detekcia landmarkov pre určenie pózy človeka	13
3.3	Detekcia landmarkov z medicínskych snímok	14
4	Cefalometrické Landmarky	16
4.1	Použitie teórie hier	18
4.2	Hybridný prístup	18
4.3	Konvolučné neurónové siete pre detekciu cefalometrických landmarkov . . .	20
5	Realizácia	22
5.1	Dataset	22
5.2	Použité technológie	23
5.3	Použité prístupy	23
5.4	Klasifikačný model	23
5.5	Prístup s použitím Stacked Hourglass Modelu	26
6	Výsledky	29
6.1	Klasifikačný model	29
6.2	Stacked Hourglass Model	31
6.3	Vyhodnotenie metód	34
7	Záver	36
	Literatúra	37
	Prílohy	39
A	Plagát	41

Kapitola 1

Úvod

V dnešnej dobe je v oblasti informačných technológií možné sledovať čoraz väčšie nadšenie pre oblasť umelej inteligencie. Je to spôsobené tým, že sa zisťuje potenciálny prínos algoritmov z tejto oblasti pre mnoho rôznych odvetví ako napríklad bankovníctvo alebo medicína. Jednou skupinou takýchto algoritmov, ktorá flexibilitou použitia prevyšuje všetky ostatné sú umelé neurónové siete. Aj keď umelý neurón ako stavebná jednotka týchto sietí bol definovaný už v 50. rokoch minulého storočia, je to až posledných pár rokov čo sa používanie neurónových sietí začalo rozmáhať. Jedným z dôvodov je výpočetná kapacita, ktorá v dobe definovania neurónu nebola dostupná. Ďalším dôvodom bola praktická ukážka možností neurónových sietí na súťaži ILSVRC 2012 [18] kedy neurónové siete v klasifikácii obrázkov prekonal tradičné algoritmy strojového učenia rozdielom triedy.

Táto práca sa zaoberá vytvorením takýchto sietí pre detekciu anatomických bodov v cefalometrickom RTG snímku. Dnes sa tieto body detekujú hlavne ručne a vyriešenie tejto úlohy by prinieslo ušetrenie množstva času pre lekárov, keďže anotovanie jedného snímku trvá aj skúsenému lekárovi okolo 20 minút. Riešením tejto úlohy sa napríklad zaoberala súťaž ISBI 2015 Challenge [21] avšak ani jeden z použitých algoritmov na tejto súťaži nebola neurónová sieť.

V práci sú navrhnuté 2 metódy riešenia tohto problému pomocou konvolučných neurónových sietí. Jedna z metód používa klasifikačný prístup a architektúra siete nie je plne konvolučná. Ďalšia metóda používa detekovanie pomocou pravdepodobnostných máp. Architektúra siete pri tejto metóde vychádza z tzv. hourglass modelu a je plne konvolučná. Najlepšie dosiahnuté výsledky boli plne konvolučnou metódou a to úspešnosť detekcie 73.22% do 5 mm.

V druhej kapitole definujeme niektoré pojmy z oblasti dopredných neurónových sietí o ktoré sa budú opierať nasledujúce kapitoly. Tretia kapitola oboznamuje o probléme detekcie landmarkov. V kapitole štyri sú rozobraté cefalometrické landmarky a spôsoby ich detekcie. Kapitola päť popisuje navrhnuté prístupy a v kapitole šesť sú prezentované ich výsledky.

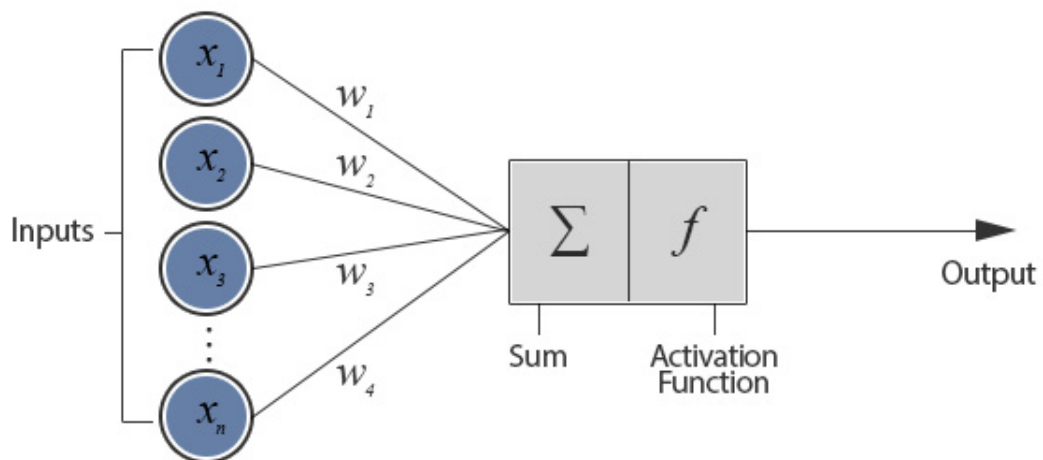
Kapitola 2

Dopredné neurónové siete

Dopredné neurónové siete sú základným modelom používaným v oblasti hlbokého učenia. Ich úlohou je aproximácia hľadanej funkcie. To znamená, že neurónovou sieťou chceme namapovať vstup x na výstup y . Následujúca kapitola popisuje teoretický základ v oblasti dopredných neurónových sietí.

2.1 Neurón

Neuróny sú základnými stavebnými jednotkami doprednej neurónovej siete. Podľa obrázka



Obr. 2.1: Príklad umelého neurónu. Prevzaté z [3]

vidíme, že výstupom neurónu je jednoduchá transformácia vstupov. Matematicky môžeme výstup neurónu vyjadriť pomocou vzorca

$$y = f\left(\sum_{n=1} i_n w_n + b\right) \quad (2.1)$$

kde:

- i_n je n -tý vstup neurónu
- w_n je váha pre n -tý vstup

- b_n je posunutie
- f je aktivačná funkcia

2.2 Aktivačná funkcia

Neurón bez aktivačnej funkcie si môžeme predstaviť ako lineárny separátor, kde váhy menia sklon priamky a posunutie nám túto priamku posúva po osi x . Keďže väčšinu funkcií reálneho sveta nie je možné klasifikovať lineárne, bola k neurónu pridaná aktivačná funkcia, ktorá nám zaisťuje nelinearitu. Uvedieme zopár príkladov aktivačných funkcií.

- **ReLU - Rectified Linear Unit**

Táto aktivačná funkcia je najbežnejšia pri dopredných neurónových sieťach. Zaisťuje nelinearitu výstupu avšak má blízko k lineárnej funkcii. Tým, že je skoro lineárna nám zachováva niektoré žiadané vlastnosti lineárnych modelov a to napríklad jednoduchosť optimalizácie pomocou gradientových metód.

$$g(z) = \max\{0, z\} \quad (2.2)$$

- **Sigmoida a Hyperbolický Tangens**

Tieto dve funkcie sú si na prvý pohľad na graf podobné s rozdielom rozsahu oboru hodnôt. Používajú sa väčšinou vo výstupných neurónoch pre určenie pravdepodobností dvoch hodnôt s vhodne zvolenou cenovou funkciou. Ich použitie v skrytých vrstvách siete sa nedoporučuje keďže ich saturácia na väčšine definičného oboru sťažuje učenie siete.

Sigmoida je definovaná vzorcom:

$$g(z) = \sigma(z) \quad (2.3)$$

Hyperbolický tangens zase:

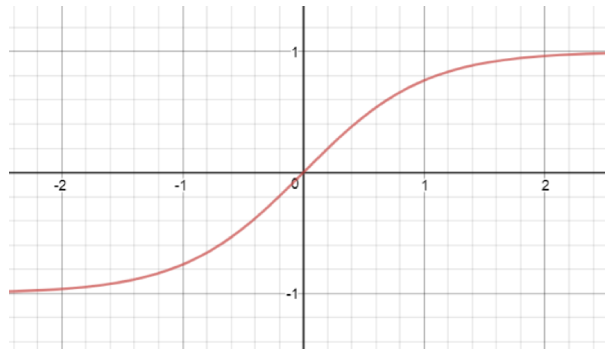
$$g(z) = \tanh(z) \quad (2.4)$$

Z pohľadu na vzorce to nevidíme ale tieto dve funkcie sú spolu späté keďže

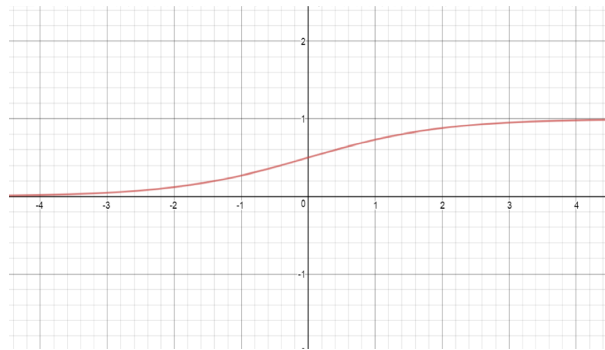
$$\tanh(z) = 2\sigma(2z) - 1 \quad (2.5)$$

- **Softmax**

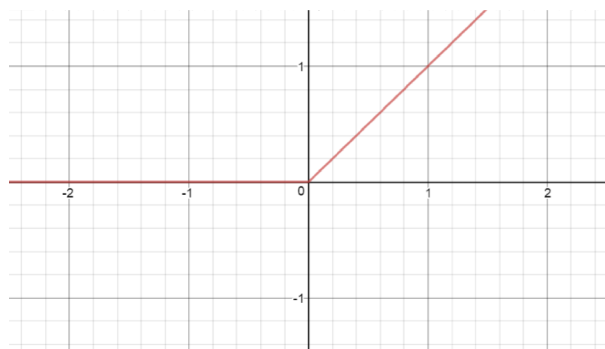
Ďalšou aktivačnou funkciou, ktorá sa používa vo výstupných vrstvách siete v klasifikačných problémoch je softmax. Podobne ako sigmoida tak výstupom softmaxu je hodnota medzi 0 a 1 s tým rozdielom, že hodnota softmaxu pre všetky výstupné neuróny dáva v súčte 1. Čo znamená že softmax udáva pravdepodobnosť náležitosti vzorku do triedy, ktorú reprezentuje daný neurón.



Obr. 2.2: Graf hyperbolického tangensu



Obr. 2.3: Graf sigmoidy



Obr. 2.4: Graf ReLU

2.3 Architektúra siete

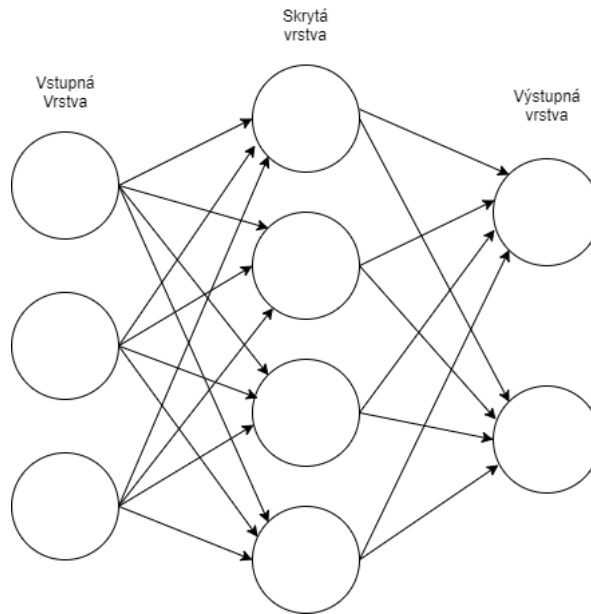
K aproximácii hladanej funkcie nám často nestačí jeden neurón, keďže dokáže modelovať len lineárne separovateľné problémy a preto ich serializovane skladáme do viacerých vrstiev. Výsledná štruktúra má podobu acyklického grafu a je schopná riešiť lineárne neseparovateľné problémy. Táto štruktúra sa nazýva neurónová sieť. Vrstvu siete si môžeme predstaviť ako vektor neurónov, ktoré majú rovnaký vstup. Výstupom vrstvy je vektor hodnôt s veľkosťou podľa počtu neurónov danej vrstvy. Z hľadiska funkcie vrstiev ich môžeme rozdeliť na:

- **vstupnú vrstvu**, ktorej vstupom je vstup neurónovej siete
- **skryté vrstvy**, ktoré sa nachádzajú medzi vstupnou a výstupnou vrstvou
- **výstupnú vrstvu**, ktorej výstupom je požadovaný výstup siete

Výstup základnej siete s tromi vrstvami a vstupom môžeme popísať ako:

$$f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x))) \quad (2.6)$$

kde x reprezentuje vstup siete a tým pádom je funkcia $f^{(1)}$ funkciou vstupnej vrstvy, $f^{(2)}$ funkciou skrytej vrstvy a $f^{(3)}$ je funkcia výstupnej vrstvy. Počet vrstiev, neurónov v nich a ich prepojenia nazývame architektúrou siete.



Obr. 2.5: Príklad jednoduchšej neurónovej siete s architektúrou 3-4-2.

2.4 Učenie neurónových sietí

Namiesto chovania, ktoré je špecifikované pomocou programu, neurónová sieť sa svoje chovanie učí na dátach, ktoré do siete prúdia v procese učenia. Učením rozumieme správny výber váh a posunutí na jednotlivých neurónoch s cieľom účinného vykonávania zadanej úlohy. Učenie je iteratívny proces, kedy prezentujeme naše dáta sieti a tá prispôsobuje

svoje parametry a tým pádom minimalizuje svoju chybu. Cieľom učenia je nájsť funkciu, ktorá najlepšie aproximuje priradenia x na y prezentované v datasete. Nevýhodou takejto špecifikácie chovania je hlavne black box prístup a výpočetná náročnosť učenia.

Cenová funkcia

Cenová funkcia, niekedy nazývaná ako chybová funkcia je funkcia, ktorou ohodnocujeme chovanie modelu. Toto ohodnotenie spočíva v porovnaní očakávaného výstupu pre konkrétnu vzorku z datasetu, s výstupom, ktorý nám dala sieť. Niektoré najpoužívanejšie cenové funkcie sú:

- **Stredná chyba**

$$ME(y, \hat{y}) = 1/n \sum_{i=1}^n \hat{y}_i - y_i \quad (2.7)$$

- **Stredná chyba**

$$MSE(y, \hat{y}) = 1/n \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2.8)$$

Rozdielom v použití strednej kvadratickej chyby oproti strednej chybe je prísnejšia penalizácia veľkých odchýliek.

Gradientový zostup

Táto metóda trénovania siete patrí medzi iteratívne metódy. Trénovaním siete rozumieme nájdenie váh a posunutí takých, pre ktoré je hodnota cenovej funkcie minimálna. Táto metóda je vhodná pre jej iteratívny charakter najmä preto, lebo kvôli nelinearite siete sú priebehy cenových funkcií nekonvexné. Narozdiel od metód používaných pre lineárne modely, v ktorých je zaručená konvergencia hodnoty funkcie do globálneho minima sú konvexné optimalizačné metódy nepoužiteľné. S pomocou gradietového zostupu sa vieme dostať k veľmi malej hodnote cenovej funkcie, ktorá nemusí byť minimálna.

Pri gradientovom zostupe postupujeme v cenovej funkcii opačným smerom ako gradient v dĺžke určitého kroku. Gradientom funkcie rozumieme smer najstrmšieho rastu funkcie a v neurónovej sieti sa vyráta pomocou algoritmu zpetnej propagácie. Pri výpočte gradientu treba dbať na to, že zmena vo váhach jedného neurónu (a tým pádom aj v jeho aktivácii) dokáže ovplyvniť celú sieť. Tento efekt musí byť kombinovaný pre všetky neuróny. Spomínaný algoritmus slúži k zaisteniu tohto javu. Vypočítaný gradient následne slúži vo vzorci optimalizačného algoritmu. Jedna iterácia tohto algoritmu vyzerá takto:

$$x_{i+1} = x_i - \alpha \nabla_{x_i} f(x_i) \quad (2.9)$$

kde:

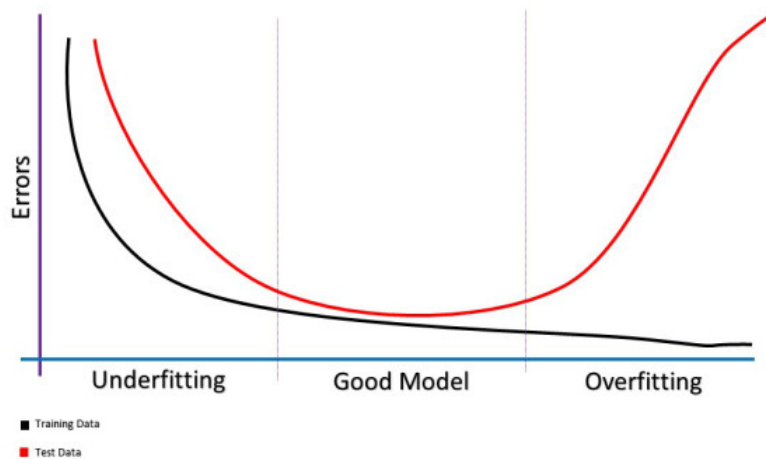
- \mathbf{x} je vektor parametrov modelu
- ∇_{x_i} je gradient funkcie pre tieto parametre
- $f(\mathbf{x})$ je hodnota cenovej funkcie
- α je koeficient učenia

2.5 Pretrénovanie

V priebehu tréovania siete je bežným javom, že chyba na tréovacích dátach klesá. Avšak pri tréovaní siete chceme minimalizovať chybu na testovacích dátach. Táto chyba nemusí klesať s chybou tréovacou a preto je treba proces tréovania zastaviť v ideálnom momente. Ak by tréovanie pokračovalo s tým, že chyba na testovacích dátach by stúpala a chyba na tréovacích dátach by klesala tak hovoríme, že dochádza k pretrénovaniu siete. Jedná sa o to, že sieť sa priveľmi prispôsobila tréovaciemu datasetu a stráca generalizačné schopnosti na iných dátach. Tento jav nastáva často pri tréovaní na nedostatočne veľkom datasete alebo pri výbere tréovacieho datasetu, ktorý dostatočne nereprezentuje variabilitu dát z množiny, pre ktorú sieť tréujeme.

Pretrénovanie vzniká aj kvôli tréovaniu komplexnejšej architektúry ako potrebujeme. Na-trénovaná neurónová sieť by mala byť aproximáciou hľadanej funkcie a keď je hľadaná funkcia pomerne jednoducho aproximovateľná, tak nepotrebujeme model, ktorý je zložitý lebo v tom prípade bude citlivejší na anomálie v datasete, ktoré nemusia zodpovedať všeobecnej podobe datasetu.

Pri tréovaní siete sa samozrejme snažíme zabrániť pretrénovaniu siete a nájsť najlepší okamih kedy je treba tréovanie zastaviť. Jedným zo spôsobov je kontrola chyby na testovacím datasete po každej epoche tréovania a ak sa táto chyba zmenila len minimálne tak sme pravdepodobne našli bod, kedy je vhodné zastaviť tréovanie siete. V lekárskejších dátach je pretrénovanie siete rozšíreným problémom keďže takéto datasety zväčša nie sú veľké kvôli náročnosti získania a spracovania takýchto dát pre použitie v oblasti hlbokého učenia.



Obr. 2.6: Závislosť chyby od doby tréovania. Prevzaté z [4]

Pre zlepšenie generalizácie siete na nových dátach sa v oblasti neurónových sietí často používajú stratégie, ktoré sa označujú ako regularizácia. Ich cieľom je zlepšiť schopnosť siete generalizovať, často na úkor výkonu na tréovacích dátach.

Augmentácia datasetu je jednou z možností ako regularizovať model. Jedná sa o pridanie „falošných“ dát do datasetu, ktoré sú generované z pôvodného datasetu rôznymi formami deformácie pôvodného obrazu ako napríklad rotácia, scaling alebo posun. Táto technika je výhodná najmä v oblasti klasifikácie obrazu alebo detekcie reči. V oblasti detekcie reči sa augmentácia môže vykonávať napríklad vloženíím náhodného šumu.

Batch normalizácia je technika ktorá vychádza z toho, že vstup siete býva často normalizovaný do určitého rozsahu. Pri batch normalizácii normalizujeme vstup do jednotlivých vrstiev siete z cieľom odstrániť extrémne hodnoty aktivácií a tak zlepšiť reakciu siete na rôzne anomálie, ktoré po batch normalizácii nemajú taký dosah na výstup. Cieľom normalizácie je zrýchlenie tréovania a zlepšenie výkonu siete. Zrýchlenie je dôsledkom rýchlejšej konvergencie siete napriek pridaniu normalizačnej vrstvy. Pri použití normalizačnej vrstvy môžeme takisto používať väčšiu hodnotu koeficientu učenia.

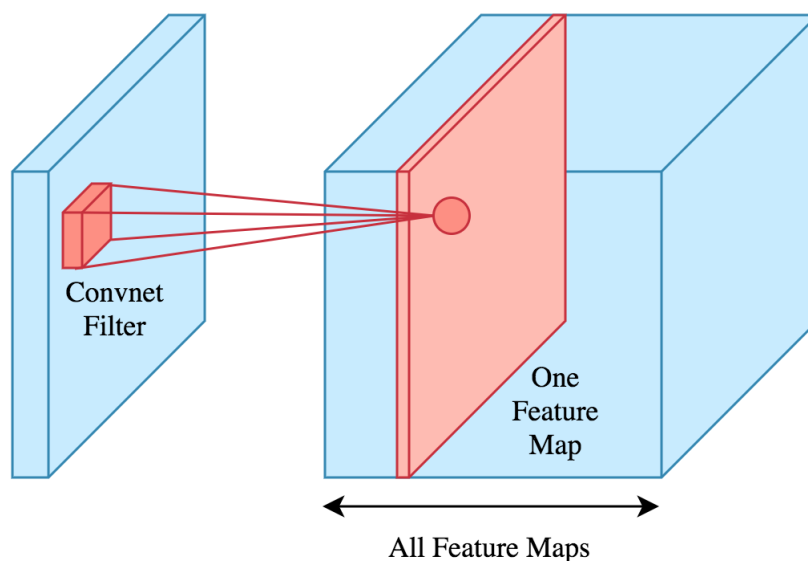
Dropout je ďalšou z výpočtetne nenáročných regularizačných techník. Dropout vrstva slúži na vynásobenie niektorých hodnôt predchádzajúcej vrstvy nulou. Dôsledkom tohto vypustenia je väčšia nezávislosť neurónov od výstupu predchádzajúcich vrstiev. Aplikovaním dropoutu na každú vrstvu siete tak môžeme o tréovaní jednej siete uvažovať ako o tréovaní podsietí, ktoré zdieľajú váhy parametrov. Pri dropout vrstvách je vhodné pre tréovanie používať menšie batche aby sa nepretréovala podsieť, ktorá je zrovna aktívna pre tento batch.

2.6 Konvolučné neurónové siete

Konvolučné neurónové siete, sú špeciálnym druhom neurónových sietí, ktoré sa používajú na spracovanie dát s tzv. mriežkovou topológiou. Príkladom takýchto dát môžeme chápať napríklad zvukové dáta kde mriežkou rozumieme jednotlivé vzorky dát rozdelené v časovom intervale alebo obrázky, ktoré sú reprezentované 2-D mriežkou. Tieto siete sú takisto nasadené vo veľkom množstve praktických aplikácií. Ich výhodou je napríklad redukcia počtu váh pre jednotlivé vrstvy siete. Hlavným rozdielom oproti bežným sieťam je v použití konvolúcie ako hlavnej operácie pri doprednom priechode namiesto maticového násobenia. Jednoduchá konvolučná sieť pozostáva z dvoch druhov vrstiev a to konvolučných a poolingových vrstiev. Pri sieťach, ktoré nie sú plne konvolučné sa používa aj plne prepojená vrstva.

Konvolučná vrstva

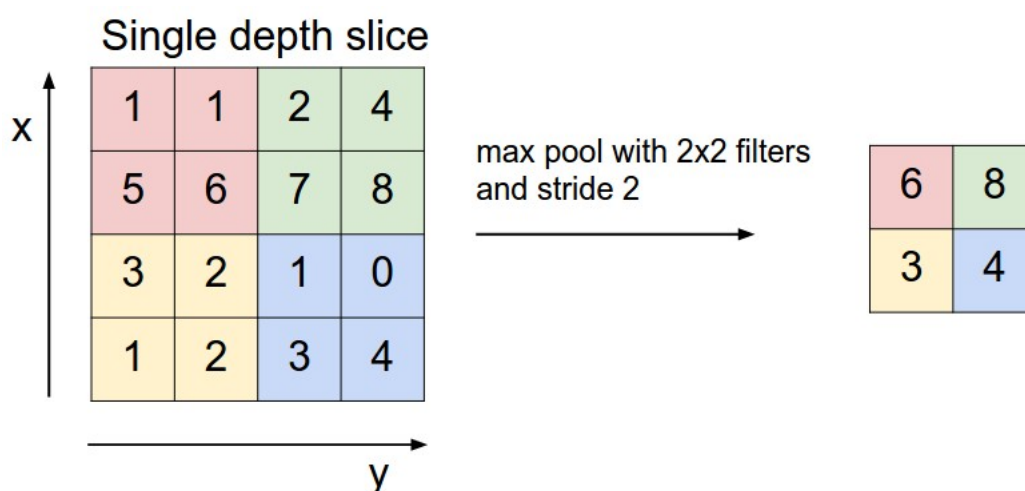
Vstupom do konvolučnej vrstvy býva multidimenzionálne pole, na ktoré je následne aplikovaná operácia konvolúcie. Výstupom vrstvy je takisto multidimenzionálne pole. Prvky tohto poľa nazývame feature mapy. Takáto feature mapa je výstupom jedného konvolučného filtru aplikovaného na vstup vrstvy. Počet týchto filtrov určuje veľkosť výstupného poľa. Veľkosť filtru určuje počet váh, ktoré sa sieť musí vzhľadom na tento filter naučiť. Celkový počet váh je pre jednu vrstvu potom *počet filtrov * veľkosť filtru*.



Obr. 2.7: Spracovanie obrazu pomocou konvolučnej vrstvy. Prevzaté z [1]

Poolingová vrstva

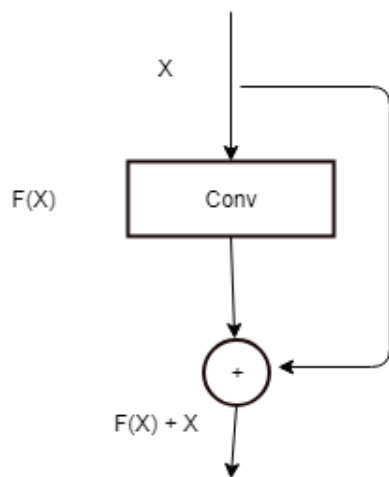
Funkciou poolingovej vrstvy je redukcia priestoru s cieľom obmedziť počet parametrov, ktoré sa sieť musí naučiť a taktiež pomáha vytvoriť reprezentáciu obrazu, ktorá je invariantná voči malým transláciám vstupu. Pooling je aplikovaný nezávisle na každú z feature máp a následuje za vrstvou konvolučnou. Väčšinou funguje na princípe výberu maxima z priestoru so zvolenou veľkosťou a toto maximum je ďalej propagované na výstup. Pri poolingovej vrstve nás zaujímajú dva parametre a to veľkosť priestoru, z ktorého vyberáme maximum a posun tohto priestoru po feature mape. Na veľkosti týchto dvoch parametrov závisí schopnosť siete zaznamenávať detaily obrazu. Čím menšie je poolingové jadro tak tým väčšie detaily je schopná sieť zaznamenávať.



Obr. 2.8: Príklad poolingovej vrstvy s priestorom 2x2 a posunom 2. Prevzaté z [2]

Reziduálne spoje

Reziduálne spoje sú technika, ktorá sa používa pri hlbokých sieťach. Jej cieľom je zabrániť gradientu aby sa rapídne znižoval počas spätnej propagácie, keďže znižovanie gradientu obmedzí schopnosť sa učiť skorším vrstvám. Reziduálny blok obsahuje klasické vrstvy siete avšak výstup tohto bloku sa sčíta z jeho vstupom. Toto robí vrstvy siete nezávislejšie na výstupoch predchádzajúcich vrstiev.



Obr. 2.9: Fungovanie reziduálneho spoju

Kapitola 3

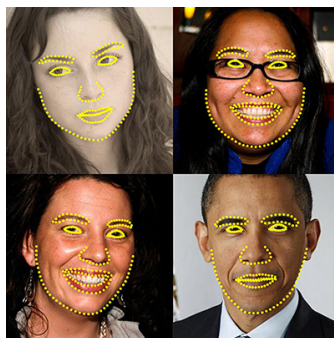
Detekcia Landmarkov

V počítačovom videní je detekcia landmarkov častou úlohou. Jedná sa o nájdenie určitých bodov v obraze, za pomoci ktorých vieme určiť komplikovanejšie vlastnosti obrazu. Príkladom je napríklad určenie polohy kĺbov pri obraze osoby s cieľom určiť jej pózu alebo nájdenie landmarkov na obraze tváre s cieľom určiť pohlavie alebo totožnosť osoby. Landmarkami, ktorých poloha sa zisťuje z tváre môžu byť napríklad špička nosa, okraje očí alebo okraje pusu. V dnešnej dobe sa pre detekciu landmarkov hojne využívajú konvolučné neurónové siete, ktoré nahradili ostatné metódy pre detekciu landmarkov pre svoju väčšiu úspešnosť na štandardných benchmarkoch.

3.1 Detekcia landmarkov tváre

Detekcia landmarkov tváre je jeden s kľúčových úkonov, ktorý musí byť vykonaný aby sme mohli ďalej analyzovať obraz a taktiež s ním pracovať. Analýzou obrazu môžeme napríklad rozumieť rozpoznanie osoby. Pri rozpoznaní osoby z obrazu je obvykle prvým krokom nájdenie tváre a umiestnením bounding boxu okolo nej. Ak sme z obrazu získali tvár tak landmarky z nej môžeme získať napríklad použitím konvolučných neurónových sietí, ktoré v dnešnej dobe pre tento úkon dosahujú strednú chybu okolo 8%. Niektorými metódami pre detekciu landmarkov tváre sú napríklad:

- **FacemarkKazemi** [11] - Táto metóda je implementovaná v knižnici OpenCV [9] a slúži na rýchlu real-time detekciu landmarkov tváre. Metóda funguje pomocou prediktívneho modelu, ktorý využíva viacero regresných stromov na odhad umiestnenia landmarku pomocou intenzít pixelov.
- **Tweaked Convolutional Neural Networks** [23] – Tento prístup funguje na princípe natrénovania jednoduchej konvolučnej siete a následnom spresňovaní výstupov jednotlivých konvolučných vrstiev.



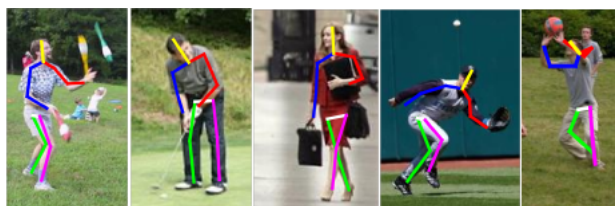
Obr. 3.1: Príklad detekovaných landmarkov na tvári. Prevzaté z [11]

Oproti cefalometrickej detekcie landmarkov je detekcia landmarkov tváre rozdielna v tom, že pri rôznych polohách hlavy sa štruktúra, ktorú landmarky vytvárajú dost líši. Takisto je treba rátať s tým, že niektoré landmarky môžu byť prekryté. V cefalogramoch sme väčšinou schopní určiť všetky landmarky, ktoré vytvoria model podobný modelu pri inom snímku. To však vyžaduje aj väčšiu presnosť pri ich určovaní keďže chyba okolo jedného centimetra pri jednom landmarku cefalogramu môže spôsobiť problém pri výslednej analýze kdežto závislosť landmarkov tváre nie je tak veľká a chyba pri jednom landmarku nemusí znamenať neschopnosť určiť celkový model tváre.

3.2 Detekcia landmarkov pre určenie pózy človeka

Podobne ako pri detekcii landmarkov z tváre je detekcia landmarkov z celého tela kľúčovým úkonom pre zistenie toho, v akej polohe sa telo nachádza. Je žiadúce aby sa postava človeka najprv extrahovala z obrazu pomocou bounding boxu a z tohto obrazu ďalej určovali polohy kĺbov, hlavy, atď. Z landmarkov je ďalej vytvorený grafický model, ktorý zobrazuje polohu tela a na základe týchto závislostí prvkov tohto modelu vieme ďalej určiť polohu, ktorej sa telo nachádza a poprípadne aj úkon, ktorý telo vykonáva. Metódy, ktoré sa zaoberajú určením polohy človeka sú napríklad:

- **DeepPose** [19] - DeepPose určuje polohu landmarkov pomocou konvolučnej neurónovej siete, kde v prvej fáze regresná časť siete odhadne hrubú polohu landmarku a následne je okolo tejto hrubej pozície vytvorený bounding box, ktorý je vstupom do ďalšej časti siete, kde sa táto pozícia doladuje.
- **Stacked Hourglass Model** [16] – Podobne ako v DeepPose, sú hlavným prvkom stacked hourglass modelu konvolučné neurónové siete. Model je podrobne opísaný v nasledujúcej sekcii.

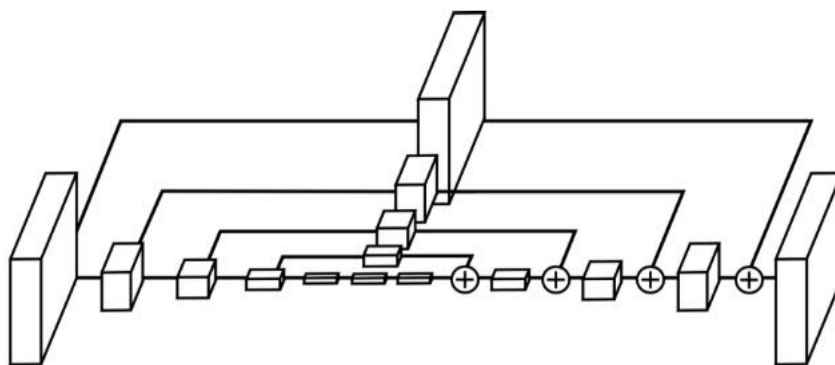


Obr. 3.2: Príklad detekcie pózy človeka. Prevzaté z [8]

Stacked Hourglass Model

Jednou z metód použiteľnou aj pre detekciu landmarkov je stacked hourglass model [16]. Je to metóda založená na použití plne konvolučných neurónových sietí. Architektúra vychádza z jednoduchého hourglass modelu taktiež nazývaného encoder-decoder sieť. Využitie hourglass modelu bolo predvedené napríklad pri úlohe segmentácie architektúrou SegNet [7].

Každý hourglass modul má enkoderovú časť a dekodérovú časť. Enkoderová časť pozostáva z konvolúcií a downsamplingu na určité rozlíšenie. Týmto sa z obrazu vytiahnu podstatné informácie, z ktorých následne môžeme rekonštruovať obraz v dekodérovej časti. Samotná rekonštrukcia neprebíha len z finálneho výstupu enkodéru a jeho upsamplingu ale aj z výstupov jeho jednotlivých konvolučných vrstiev, ktoré sú následne kombinované z výstupmi upsamplingu. Toto je možné vykonať napríklad s pomocou reziduálneho spojenia. Z dekodéru potom dostaneme obraz s pôvodným rozlíšením. Architektúru takejto siete nazývame plne konvolučnou architektúrou keďže nevyužívame plne prepojené vrstvy.



Obr. 3.3: Vizualizovaný hourglass modul. Prevzaté z [16]

Stacked hourglass model sa od bežného hourglass modelu líši tým, že je v ňom umiestnených viac takýchto modulov za sebou a na výstup každého z nich je aplikovaná loss funkcia. To znamená že výstup každého modelu sa dá použiť ako finálna predikcia. Spracovaním výstupov predchádzajúceho modulu je možné ďalej zdokonaľovať tieto predikcie. Pri optimalizácii modelu sú aplikované chybové funkcie aj na čiastkové výstupy jednotlivých modulov. Váhy medzi modulmi však nie sú zdieľané.

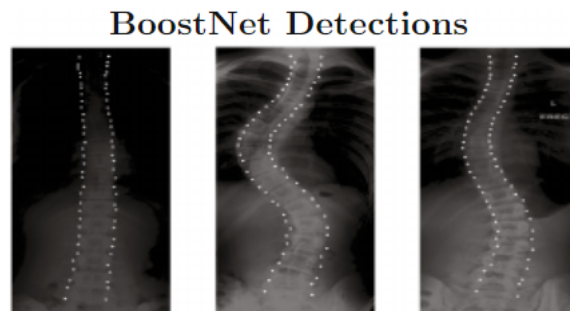
3.3 Detekcia landmarkov z medicínskych snímok

Okrem vyššie zmienených prístupov sa automatická detekcia landmarkov využíva aj v lekárskom prostredí. Detekovaním landmarkov z medicínskych dát sa zaoberá aj táto práca. Okrem cefalometrických landmarkov sa detekujú aj landmarky napr. prostaty, mozgu, [24] alebo chrbtice [22]. Problémom pri vytváraní modelov pre detekciu landmarkov z medicínskych dát je často ich nekvalita a nedostatok. Okrem augmentácie sa tento problém rieši napríklad vytváraním syntetických dát. Takto vytvorené automatické systémy dokážu uľahčiť prácu lekárom a zlepšiť presnosť detekcie.

Detekcia landmarkov chrbtice pre AIS

Jedným zo systémov pre detekciu landmarkov z medicínskym obrazov je napríklad BoostNet [22]. Tento model rieši detekciu landmarkov chrbtice pri chorobe AIS (Adolescent Idiopathic Scoliosis). Táto choroba spôsobuje abnormálne zakrivenie chrbtice. Užitočnosť tohto systému spočíva v tom, že včasnou detekciou symptómov tejto choroby je možné zastaviť jej pokračovanie. Po detekcii určitých landmarkov chrbtice sme schopný určiť uhly medzi stavcami, ktoré následne ukazujú fázu skoliózy.

Samotný systém pozostáva y konvolučných vrstiev, ktoré slúžia na extrakciu príznakov, boost vrstvy, ktorá odstraňuje vplyv nepodstatných extrahovaných príznakov ako napríklad šum, chyby snímku či rôzne anomálie. Boost vrstva je encoder-decoder modul, kde sa snažíme príznaky rekonštruovať tak, aby sme čo najviac zmenšili chybu oproti priemerým hodnotám príznakov. Príznaky, ktoré sa snažíme upraviť sú definované ako príznaky, ktorých hodnoty sú vyššie ako určená hranica od priemeru na datasete. Následne sa z upravených príznakov vyhodnotia pozície landmarkov vo výstupnej vrstve z ohľadom na závislosti landmarkov.

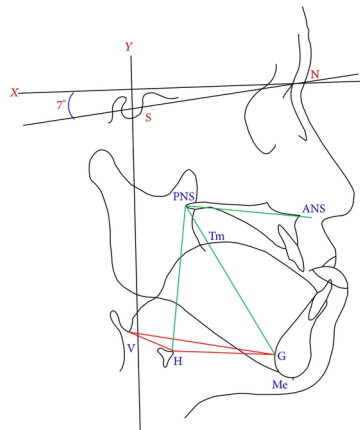


Obr. 3.4: Príklad chrbtíc a landmarkov detekovaných pomocou BoostNet. Prevzaté z [22]

Kapitola 4

Cefalometrické Landmarky

Detekcia cefalometrických landmarkov je prvým krokom pri cefalometrickej analýze, ktorá má významnú úlohu v stomatológii alebo plastickej chirurgii. Okrem toho je jej úlohou napríklad aj diagnóza problémov pri vývine hlavy a ohodnotenie liečby takýchto problémov alebo plánovanie chirurgických zákrokov. Zaoberá sa vzťahmi medzi landmarkami. Príkladom takéhoto vzťahu môžu byť napríklad uhly alebo vzdialenosti medzi landmarkami. Táto analýza sa vykonáva na cefalometrických rádiologických snímkoch.



Obr. 4.1: Niektoré analyzované vzťahy medzi landmarkami. Prevzaté z [20]

Po dlhú dobu boli tieto landmarky detekované manuálne expertom. Toto riešenie je stále často využívané. Nevýhodou manuálnej detekcie je hlavne jej časová náročnosť a požiadavky na schopnosti človeka vykonávajúceho túto detekciu. Vzhľadom na tieto problémy boli navrhnuté automatické metódy, ktoré majú analyzátorom pomôcť. Prvotné metódy neboli príliš efektívne, avšak neskôr boli navrhnuté efektívnejšie prístupy. Tieto prístupy môžeme klasifikovať do viacerých kategórií:

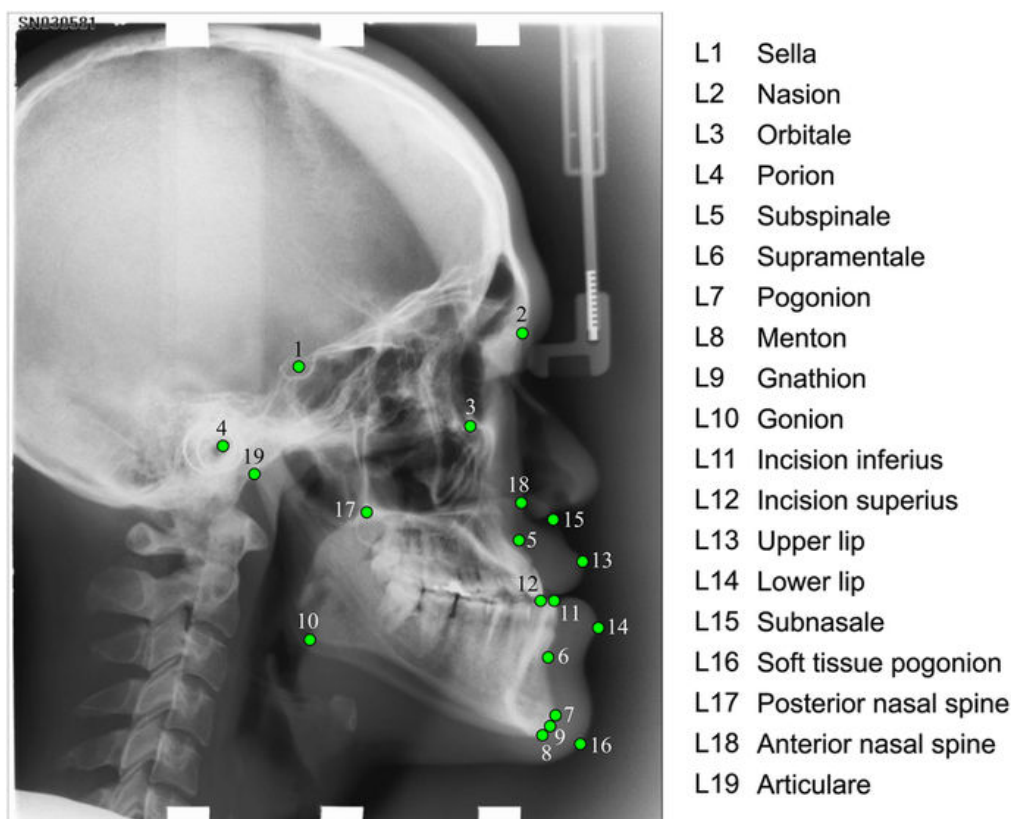
- **Metódy využívajúce spracovanie obrazu**
- **Štatistické modely**
- **Soft computing metódy**
- **Hybridné metódy**

Prvý druh metód je stále veľmi používaný. Obvykle sa pri tomto prístupe na cefalogram aplikujú rôzne metódy na detekciu hrán. S hranami a vedomosťou o štruktúre lebky vieme určiť polohu niektorých landmarkov. Napriek jednoduchosti a dostupnosti tohto prístupu existujú nevýhody a to napríklad závislosť na kvalite obrazu a neschopnosť určiť polohu landmarkov, ktoré sa nenachádzajú na žiadnej významnej hrane. Tieto metódy ale stále hojne využívajú napríklad aj v kombinácii s ostatnými prístupmi.

Štatistické modely fungujú na princípe vytvorenia priestorového modelu pre lokácie landmarkov na základe šablony, získanej z tréningových dát, ktorá určuje pravdepodobnosť výskytu landmarku na určitých pozíciách. Stanovenie presnej lokácie landmarkov je ďalej odvíjané od podmienenej pravdepodobnosti pre jednotlivé landmarky modelu vytvoreného podľa šablony. Výhodou týchto modelov je hlavne nezávislosť voči transláciám a rotáciám. Jednou z nevýhod je citlivosť na šum v obraze a neschopnosť detekovať anomálie.

Najdôležitejšou a najviac používanou soft computing metódou v oblasti detekcie landmarkov sú umelé neurónové siete. V nasledujúcich kapitolách sú predložené prístupy k detekcii landmarkov pomocou neurónových sietí a takisto diskutované ich výhody a nevýhody.

V dnešnej dobe sa pre detekciu cefalometrických landmarkov používajú hybridné metódy. Ich výhodou je možnosť kombinovania vyššie spomenutých postupov pre rôzne landmarky kvôli ich rozdielnej podobe a umiestnení v cefalograme.



Obr. 4.2: Zoznam detekovaných landmarkov. Prevzaté z [14]

Na obrázku 4.2 môžeme vidieť príklad anotovaného cefalogramu. Na cefalometrickú analýzu sa obvykle používa prvých 16 landmarkov. Každý landmark je umiestnený v rozdielnych regiónoch lebky kde môžeme rozlišovať či sa landmark nachádza v mäkkom tkanive, na hranici kosti alebo napríklad priamo na kosti. Takisto nervy a krvné kapiláry môžu určovať rôzne anatomické vlastnosti rôznych oblastí, v ktorých sa landmarky nachádzajú.

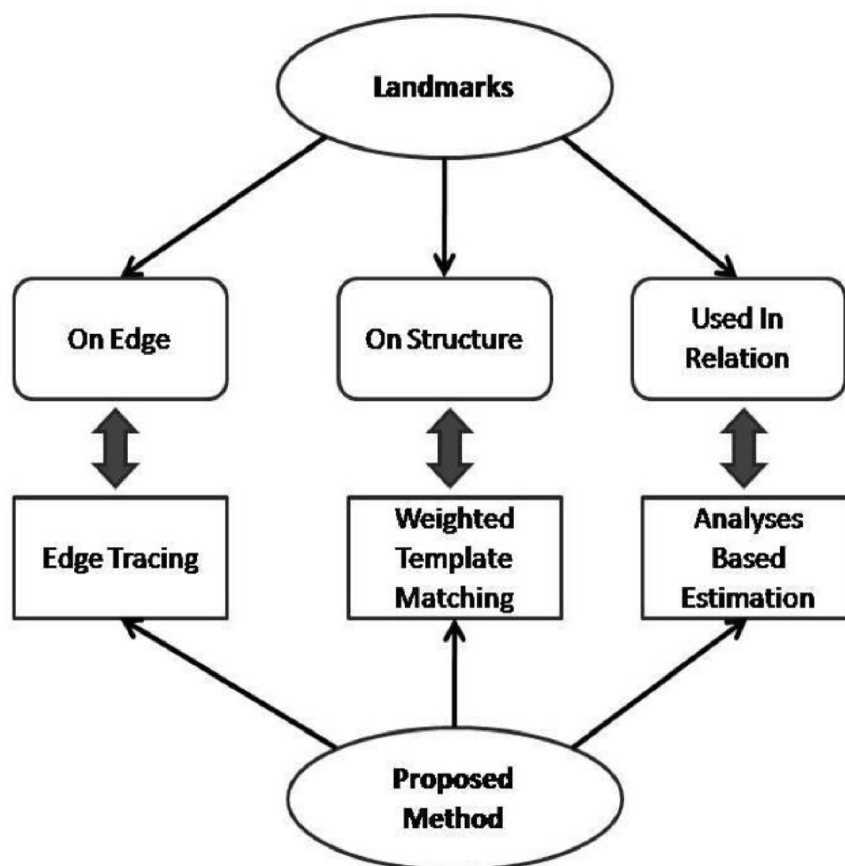
4.1 Použitie teórie hier

Tento algoritmus [10] prezentovali B. Ibragimov et al. na súťaži IEEE ISBI Challenge[21] v úlohe detekcie cephalometrických landmarkov. Jeho princíp je extrakcia intenzít a priestorových závislostí jednotlivých landmarkov z tréningového datasetu. Podoby landmarkov sú extrahované pomocou Haarových príznakov z okolia pixelu landmarku v tréningovom datasete. Na základe nich vieme určiť pravdepodobnosť výskytu landmarku na danom pixeli z testovacieho datasetu. Pixely z najväčšou pravdepodobnosťou výskytu následne zaradíme do poľa kandidátov.

Okrem pravdepodobností sa taktiež berie do úvahy aj priestorová závislosť landmarkov. Túto získame tak, že z tréningového datasetu zistíme uhly a vzdialenosti medzi každými dvoma landmarkami pre každý snímok. Zo všetkých získaných hodnôt sa následne vytvorí priestorový model pre každé dva landmarky. Pre zaistenie flexibility sa z výpočtu priestorového modelu vylúčia kombinácie z vysokou variabilitou uhlu a vzdialenosti. Samotná teória hier prichádza na rad pri rozhodovaní pri výbere landmarku z poľa kandidátov. Ako hráči sú reprezentované landmarky, body z poľa kandidátov sú stratégie a pravdepodobnosti, že landmark prislúcha bodu z poľa kandidátov plus ohodnotenie priestorového modelu ako ohodnotenie ťahov.

4.2 Hybridný prístup

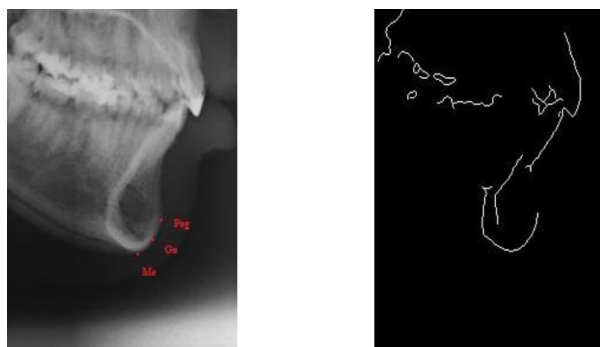
Na základe nepodobnosti jasových okolí určitých skupín landmarkov a efektívnosti rôznych prístupov na tieto skupiny, uverejnili Mahsid Majd a Farzaneh Shoeleh článok[15], ktorý popisuje hybridný prístup k detekcii landmarkov. V tomto prístupe sú landmarky rozdelené do troch skupín a na každú skupinu je použitá iná metóda. Týmito metódami sú sledovanie hrán, vážená zhoda podľa šablóny a odhad podľa analýzy. Sledovanie hrán je použité na landmarky nachádzajúce sa na hranách, zhoda je pre landmarky v okolí viditeľných štruktúr a posledná metóda odhaduje priame vzťahy medzi niektorými landmarkami keďže v cefalometrickej analýze niekedy nie je potreba poznať presnú pozíciu ale stačia nám jeho vzťahy s ostatnými landmarkami. Prvým krokom, ktorý je v tomto postupe vykonaný je extrakcia regiónu okolo landmarku. Táto extrakcia je na základe znalosti ľudskej lebky tak, že v extrahovanom regióne sa landmark nachádza so 100% pravdepodobnosťou. V ďalších fázach detekcie pozície sa pracuje s týmto regiónom.



Obr. 4.3: Princíp rozdelenia spracovania pri hybridnej metóde. Prevzaté z [15]

Edge Tracing

Edge tracing je rýchly spôsob ako detekovať landmarky nachádzajúce sa na ľahko extrahovateľných hranách. V tejto metóde je na región aplikovaný detektor hrán a z výstupu detektoru môžeme určiť polohu landmarku.



Obr. 4.4: Oblasť okolo landmarku pred a po použití detektoru hrán. Prevzaté z [15]

Weighted Template Matching

Tento mechanizmus je používaný na detekciu landmarkov, ktoré sa nachádzajú v špecifickej štruktúre viditeľnej a extrahovateľnej z regiónu. Prvým krokom je umelé vytvorenie šablóny z tréningových dát, ktorá najlepšie aproximuje danú štruktúru. Táto šablóna sa ďalej porovnáva so štruktúrou na cefalogramoch, na ktorom šablónu posúvame a hľadáme oblasť, kde nám šablóna sedí. Po nájdení oblasti, ktorá nám najlepšie sedí so štruktúrou sme schopní anotovať landmark podľa šablóny.

Analysis Based Estimation

Detekcia niektorých landmarkov, ktoré majú vo svojom okolí nervové alebo cievne štruktúry je pomocou vyššie spomenutých metód náročná. Tieto landmarky sa v hybridnom prístupe anotujú na základe vedomostí o ostatných landmarkoch, ktoré sú detekovateľné edge tracingom alebo s pomocou WTM. Príkladom môže byť landmark č. 3, ktorý sa obvykle nachádza v strede úsečky spájajúcej landmark č. 2 a landmark č. 17. Takisto je vhodné spomenúť, že pri cefalometrickej analýze nás zaujímajú tvary priestorových modelov, ktoré landmarky vytvárajú a preto si môžeme dovoliť vynechať presné pozície niektorých landmarkov ak poznáme vzťahy, v ktorých tieto landmarky vystupujú.

4.3 Konvolučné neurónové siete pre detekciu cefalometrických landmarkov

Okrem využitia pri detekcii landmarkov tváre alebo landmarkov potrebných pri určovaní pozície človeka sa konvolučné neurónové siete začali používať aj pre spracovanie medicínskych obrazových dát. Úlohou siete pri takejto úlohe je naučiť sa rozpoznávať vizuálne vzory, ktoré sa nachádzajú pri rôznych landmarkoch a následne tieto vzory správne použiť pri nových dátach. To, ako dobre bude sieť fungovať závisí hlavne na tom, či sa v tréningovom datasete nachádza každý možný vzor, ktorý sa potom môže nachádzať v reálnych dátach. Ďalej je to schopnosť siete generalizovať tieto vzory na nové dáta.

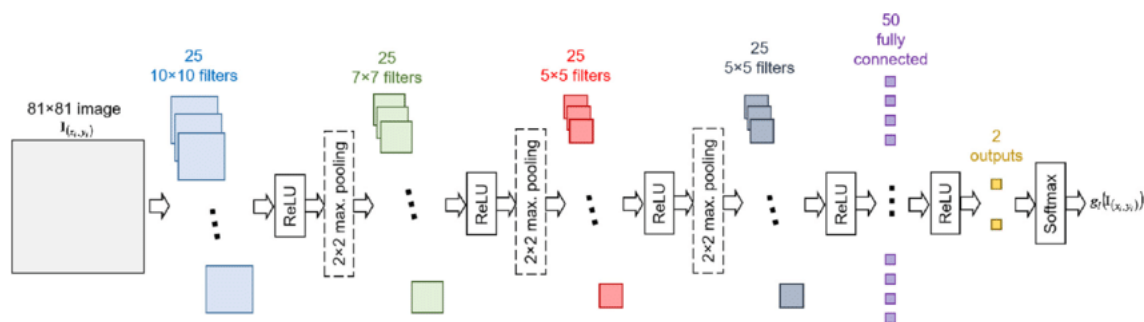
Avšak aj pri použití nástrojov ako konvolučné neurónové siete nemusíme mať istotu, že na medicínskych dátach budú podávať lepšie výsledky ako napríklad štatistické modely. Jedným z problémov prečo nemusia byť konvolučné neurónové siete vhodné pre každú úlohu s medicínskymi dátami je napríklad veľkosť datasetu. Oproti datasetom potrebným na iné úlohy je veľkosť medicínskych datasetov malá. Konvolučné neurónové siete často potrebujú veľké množstvo dát aby sa dokázali naučiť rôzne vzory, ktoré sa v týchto dátach nachádzajú a tým pádom aj lepšie generalizovať na reálne dáta. Tento problém môžeme riešiť napríklad regularizáciami, ktoré boli spomenuté v kapitole 2.3.

CNN v kombinácii s priestorovým modelom

Príkladom využitia konvolučných neurónových sietí je napríklad systém navrhnutý Serkan O. Aricom a Bulat Ibragimovom [6], ktorý pre detekciu landmarkov používa prvotný odhad pomocou konvolučných neurónových sietí a tento odhad ďalej spresňuje pomocou priestorového modelu.

V tomto prístupe sa používa jedna natrénovaná sieť pre detekciu každého landmarku. Vstu-

pom siete je okolie pixelu, o ktorom má byť rozhodnuté či sa jedná o landmark alebo nie. Výstupom je pravdepodobnosť toho, či je daný pixel landmarkom. Keďže v použítom datase [21] je landmarkom len 1 pixel, ktorý má dĺžku 0.1 mm a požadovaná maximálna chyba je rádovo väčšia, je ako landmark označených aj viacero pixelov okolo toho, ktorý je anotovaný. Snímky datasetu sú taktiež 3x downsamplované. Ako trénovací dataset je z každého landmarku vytvorených 25 pozitívnych a 500 negatívnych okolí. V každom okolí je sieťou klasifikovaný pixel v strede. Toto okolie má veľkosť 80x80 pixelov. V rámci jedného trénovania sa sieť učí na okoliach každého landmarku. Po nájdení pixelov, ktoré boli sieťou detekované ako landmarky ich umiestníme do poľa kandidátov a pokračujeme ako v systéme popísanom v kapitole 4.1 [10].



Obr. 4.5: Sieť použitá v tomto prístupe. Prevzaté z [6]

V rámci state of the art prístupov k detekcii cefalometrických landmarkov bol tento systém lepší ako doteraz používané systémy. To dokazuje úspešnosť konvolučných sietí aj pre úlohy s malým datasetom. Nevýhodou tohto modelu je nutnosť znalosti apriornej informácie o tom, kde približne sa v snímku landmark nachádza.

Plne konvolučná sieť

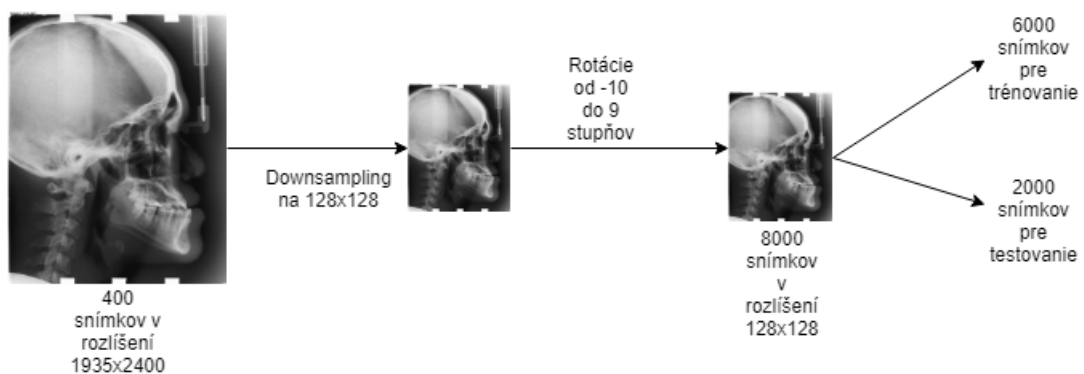
Tento prístup [17] používa na detekciu landmarkov plne konvolučnú sieť v kombinácii z reziduálnymi spojmi. Jedná sa o jeden modul hourglass modelu pre detekciu pózy človeka. Podobne ako v predchádzajúcom použití konvolučnej siete sú jej vstupom výrezy snímku okolo landmarku. Výstupom je heatmapa. Narozdiel od predchádzajúcej metódy je v tomto systéme natrénovaná pre každý landmark jedna sieť. Vstupné výrezy sú získavané z obrázku spôsobom, kedy je sú z trénovacieho datasetu získané pozície landmarkov a najväčšie a najmenšie súradnice určitého landmarku sú ďalej použité ako súradnice okna, ktoré je vstupom siete. Veľkosť a poloha okna sa v priebehu trénovania a testovania nemení.

Kapitola 5

Realizácia

5.1 Dataset

Prvou úlohou pri návrhu a experimentovaní s neurónovými sieťami je použitie vhodného datasetu pre túto úlohu. V rámci tejto práce bol použitý dataset vytvorený pre súťaž v detekcii cefalometrických landmarkov [21]. Dataset obsahuje 400 cefalogramov od rôznych ľudí. Pre každý cefalogram existujú 2 anotácie prevedené a kontrolované expertami na detekovanie cefalometrických landmarkov. Pri experimentovaní bola kvôli konzistentnosti použitá len jedna sada anotácií od jedného experta. Cefalogramy majú veľkosť 1935x2400 pixelov kde pomer pixelov voči milimetrom je $1 \text{ px} = 0.1 \text{ mm}$. Veľkosť tohto datasetu je pre trénovanie hlbokých neurónových sietí nedostatočná a preto bol dataset pri rôznych experimentoch augmentovaný. Rozdelenie datasetu na trénovaciu a testovaciu sadu bolo urobené v pomere 3:1. Následne bola prevedená augmentácia na obidvoch sadách a to prostredníctvom rotácie snímkov. Snímky boli rotované o -10 až 9 stupňov takže z každého snímku bolo vytvorených ďalších 19 snímkov. Bohužiaľ snímky s rozmermi 1935x2400 sú dosť veľké a spracovanie takýchto snímkov by trvalo priveľa času. Snímky sú teda zmenšené na rozmer 128x128 čo je priateľná veľkosť pre spracovanie a pomer je $1 \text{ px} = 1.5 \text{ mm}$ v osi x a $1 \text{ px} = 1.875 \text{ mm}$ v osi y. Výsledný dataset má teda 8000 snímkov kde trénovacia sada má veľkosť 6000 snímkov.



Obr. 5.1: Vytvorenie datasetu

5.2 Použité technológie

Na vytvorenie a trénovanie sietí použitých v rámci práce bola použitá knižnica Tensorflow [5], ktorá je momentálne najrozšírenejšou knižnicou pre vysokovýkonné numerické výpočty a taktiež poskytuje množstvo funkcií pre vytváranie a trénovanie neurónových sietí ako napríklad chybové funkcie, optimalizátory alebo jednotlivé vrstvy siete. Jej charakteristikou je reprezentácia premenných pomocou tenzorov. Výhodou je taktiež možnosť používať grafickú kartu ako hardware pre tieto výpočty keďže grafická karta je na takéto výpočty stavaná. Neurónové siete použité v tejto práci boli trénované na grafickej karte NVIDIA GeForce GTX 850M s pamäťou o veľkosti 2GB. Pamäť 2GB neumožňuje trénovať väčšie modely s veľkým množstvom parametrov avšak jej výhodou je podpora CUDA (platforma pre paralelné výpočty) a CuDNN (Deep Neural Network library), ktoré násobne zmenšili dobu potrebnú na natrénovanie siete. Problémy pri trénovaní na tomto GPU však vznikali práve už pri spomínanom trénovaní modelov s veľkým množstvom parametrov kedy pamäť grafickej karty nestačila na uchovanie všetkých tenzorov. Tento problém sa dal čiastočne vyriešiť zmenšením veľkosti batchu.

Okrem Tensorflow sa takisto používal numpy a PIL pre prácu so snímkami. Ako jazyk, v ktorom sa siete programovali bol použitý Python3.4 pod operačným systémom ElementaryOS Freya.

5.3 Použité prístupy

Pre detekciu cefalometrických landmarkov pomocou konvolučných neurónových sietí boli v rámci tejto práce vyskúšané 2 prístupy. Ďalej sa experimentovalo s týmito prístupmi v ich použití. To znamená architektúru siete, použité hyperparametry, vstupy a výstupy. Prvým prístupom je prístup klasifikačný, ktorého architektúra sa pri experimentoch skladala z jednoduchých konvolučných a poolingových vrstiev nasledovanými plne prepojenou vrstvou a vrstvami pre klasifikáciu. Druhý prístup je plne konvolučná neurónová sieť, ktorá je postavená na princípe stacked encoder-decoder (hourglass) siete.

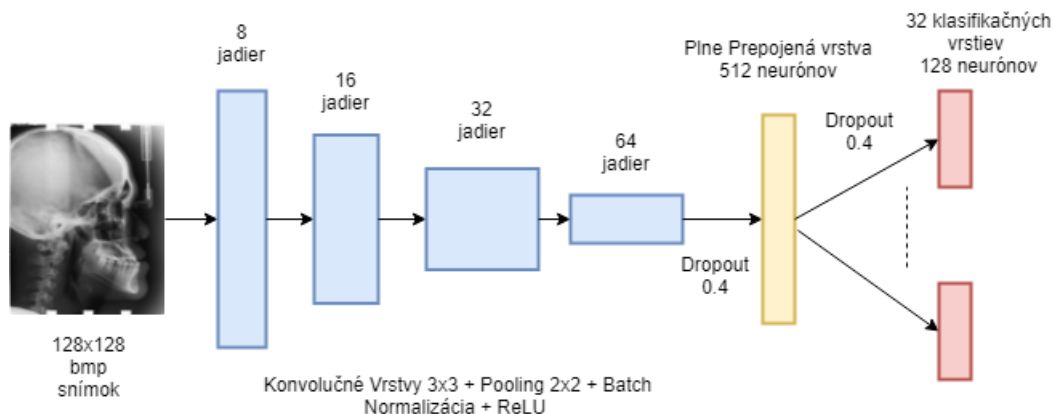
5.4 Klasifikačný model

Pri klasifikačnom prístupe pristupujeme nezávisle k obidvom osiam cefalogramu. Vstupom modelu sú dáta a ich labely. Dátami sú snímky popísané v kapitole 5.1 a labely pre jeden landmark sú tvorené dvomi one hot poľami, kde veľkosť poľa je určená rozsahom dimenzie, pre ktorú je pole určené. Keďže snímky majú veľkosť 128x128 tak dĺžka poľa je 128. Prvkami poľa sú samé nuly okrem prvku na indexe s číslom, ktoré je rovnaké ako číslo pixelu kde sa nachádza landmark. Pomocou klasifikačnej metódy anotujeme 16 landmarkov, ktoré sú najdôležitejšie pre cefalometrickú analýzu namiesto pôvodných 19-tich. To znamená, že keď neberieme do úvahy veľkosť batchu tak vstupom siete je 128x128x3 tenzor, ktorý reprezentuje obrázok a jeho labelom a teda výstupom modelu je 16x2x128 tenzor, kde prvá dimenzia predstavuje počet landmarkov, druhá dimenzia reprezentuje osu x a osu y a tretia dimenzia rozsah týchto osí a teda aj rozsah one hot poľa.

Architektúra

Základom modelu je jednoduchá konvolučná časť. Z výstupu tejto časti je ďalej vytvorený jednorozmerný vektor. Tento vektor je vstupom do nekonvolučnej časti modelu kde je jedna

plne prepojená vrstva, na ktorú je pri tréovaní aplikovaný dropout. Výstup tejto vrstvy je vstupom pre 32 klasifikačných vrstiev – jedna vrstva pre každú z dvoch dimenzií obrazu.



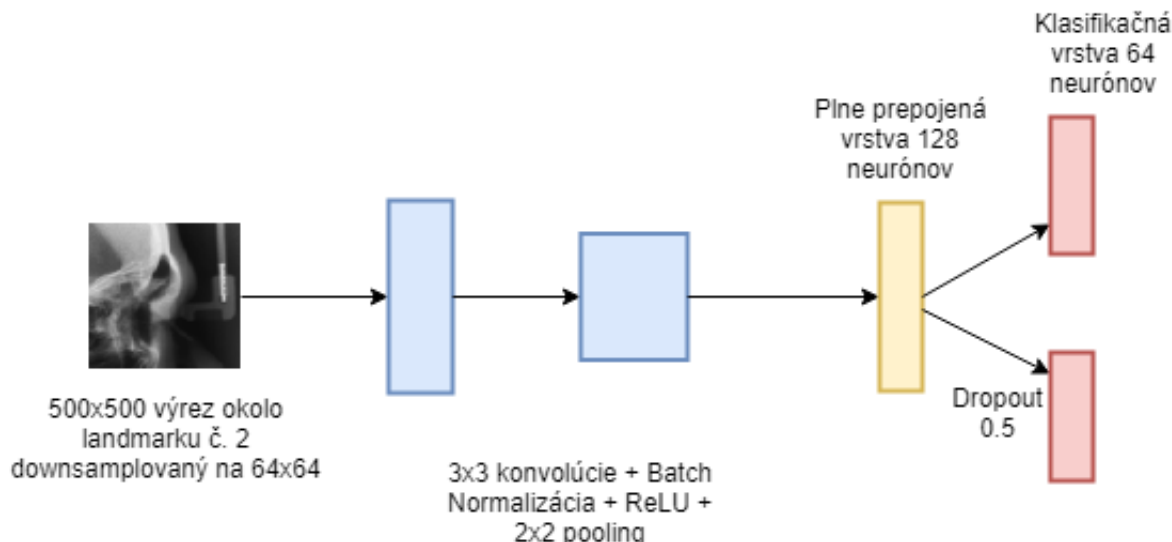
Obr. 5.2: Architektúra klasifikačného modelu s najlepšimi výsledkami

Pri konvolučných vrstvách je ako aktivačná funkcia použitá ReLU. Pred samotným zavedením nelinearity pomocou ReLU je na výstupy aplikovaná batch normalizácia. V klasifikačnej časti je ako aktivačná funkcia použitý softmax. Výsledná trieda je získaná ako index maxima pre dané pole. Ako chybová funkcia je použitá cross entropia a optimalizátor je Adam [12]. Získanie finálnych súradníc landmarkov prebieha tak, že z výstupu každej z klasifikačných vrstiev je zobrať index maxima. Index potom reprezentuje súradnicu pixelu, na ktorom sa v danej osi nachádza landmark. Kombináciou týchto indexov získame finálnu súradnicu landmarku.

Alternatívne možnosti

Okrem prístupu keď sieť tréujeme na klasifikáciu všetkých 16 landmarkov bola v rámci experimentovania vyskúšaná aj možnosť natréovať sieť s podobnou architektúrou práve na jeden landmark. Rozdiel oproti vyššie uvedenej architektúre spočíva v tom, že klasifikačné vrstvy pre tréovanie na jednom landmarku sú len dve. Tento spôsob je podľa vykonaných experimentov menej účinný keďže sieť sa lepšie učí pri pozorovaní komplexnejších štruktúr, ktoré sú vytvorené závislosťami jednotlivých landmarkov kdežto pri tréovaní len pre jeden landmark sieť pozoruje len štruktúru okolo tohto landmarku.

Ďalším z alternatívnych prístupov vyskúšaným v rámci hľadania najlepšej siete bolo vytvorenie klasifikačného modelu, ktorému bola ako informácia poskytovaná len časť obrázka okolo landmarku. V rámci experimentovania boli vytvárané výrezy s rozmerom 500x500 pixelov s tým, že landmark bol náhodne umiestnený v rozmedzí 400x400 pixelov v tomto výreze. Výhodou bola možnosť takmer neobmedzenej augmentácie datasetu pomocou posuvov v rámci výrezu, v ktorom sa landmark nachádzal. Nevýhodou vkladania takýchto výrezov by vo výslednom systéme bola nutnosť informácie o tom, v ktorom výreze v rámci celého snímku sa landmark nachádza. Na získanie tejto informácie by sme buď potrebovali ďalšiu sieť alebo iný model, ktorý by tento výrez vedel získať.



Obr. 5.3: Architektúra siete pri odhade z výrezov

Tento prístup sa však neosvedčil slubnými výsledkami pravdepodobne kvôli nedostatku informácie vo výreze a veľkou variabilitou pozície landmarku v tomto výreze. Jednou z možností ako tento problém riešiť by bolo do tohto modelu dávať výrezy s menšou variabilitou pozície landmarku kde by tieto výrezy boli z modelu, ktorý by hrubo predikoval súradnice landmarku.

Problém nebol v modeli, ktorý by mal byť dostačujúci ale vo vytvorenom datasete pre vstup do modelu. Sieť sa teda nevedela rozpoznať štruktúru keďže vstupy boli v malom rozlíšení (64x64) čo by až tak nevyhovovalo avšak hlavným problémom bola spomínaná veľká variabilita umiestnenia landmarku v rámci výrezu. V prvých pokusoch bola veľkosť výrezu 400x400 pixelov, čo odpovedá 4x4 cm. Chybou pri vytváraní týchto výrezov bolo, že posuv bol náhodný do 1.9 cm v oboch smeroch takže sme mohli získať výrezy kde by bol landmark umiestnený na pixely 390, 390 ale aj 10, 10 vkladáním takýchto výrezov sa síce sieť učila avšak predikcia nebola dostatočná. Ďalším pokusom bolo vkladanie výrezu 500x500 s takými istými možnosťami posuvov ako pri 400x400. Výrez by tak zachytával väčšiu priestorovú informáciu a bola by obmedzená variabilita umiestnenia landmarku. Výsledky tohto pokusu boli lepšie ako pokusu predchádzajúceho avšak stále nedosahovali nejakých výraznejších kvalít. Jednou z možností by bolo ešte zmenšiť variabilitu umiestnenia tým, že by bol výrez vybraný podľa dopredu poznanej informácie. Napríklad s pomocou najvyšších a najnižších súradníc landmarku v rámci datasetu, ktoré by vytvárali bounding box ohraničujúci možné výskyty landmarkov a rozšírením tohoto bounding boxu podobne ako v prípade príklade plne konvolučnej siete [17].

Takéto vyberanie výrezu má však dve hlavné nevýhody:

- Prvou nevýhodou je, že hranice okna vytvoreného podľa dostupného datasetu by nemuseli zachytiť landmark na nových dátach. Dá sa to však riešiť nájdením vhodnej, dostatočne veľkej hodnoty, o ktorú by sa zväčšil bounding box vytvorený pomocou najvyšších a najnižších súradníc landmarkov v datasete. Ak by nové dáta neobsahovali nejakú extra anomáliu (čo však obsahovať môžu), tak by táto nevýhoda neexistovala.

- Ďalšia nevýhoda vyplýva z rôznej variability umiestnenia landmarkov. Bounding box vytvorený pomocou najvyšších a najnižších súradníc môže a určite sa nájdu také landmarky kde aj má rozdielnu veľkosť. Riešiť sa to dá fixnou veľkosťou výrezu pri všetkých landmarkoch, čo však spôsobí problém, kedy variabilita umiestnení landmarkov vo výreze bude rôzna. Toto môže spôsobiť, že jeden landmark bude detekovateľný ľahšie ako druhý. Problém s rôznou variabilitou umiestnenia je aj pri vkladaní celého obrázka ako vstupu avšak pri výrezoch spôsobuje väčšie problémy. Ďalším riešením je rôzna veľkosť vstupov podobne ako v encoder-decoder sieti[17] avšak z tohto riešenia vyplýva ďalší problém - rozdielna veľkosť vstupu. V kombinácii CNN a shape modelu [6] riešia tento problém fixným oknom, masívnou augmentáciou datasetu a pridaním negatívnych výrezov, v ktorých sa landmark nenachádza. Dostatočne zložitý model a veľký počet dát celkovo riešia problém variability umiestnenia.

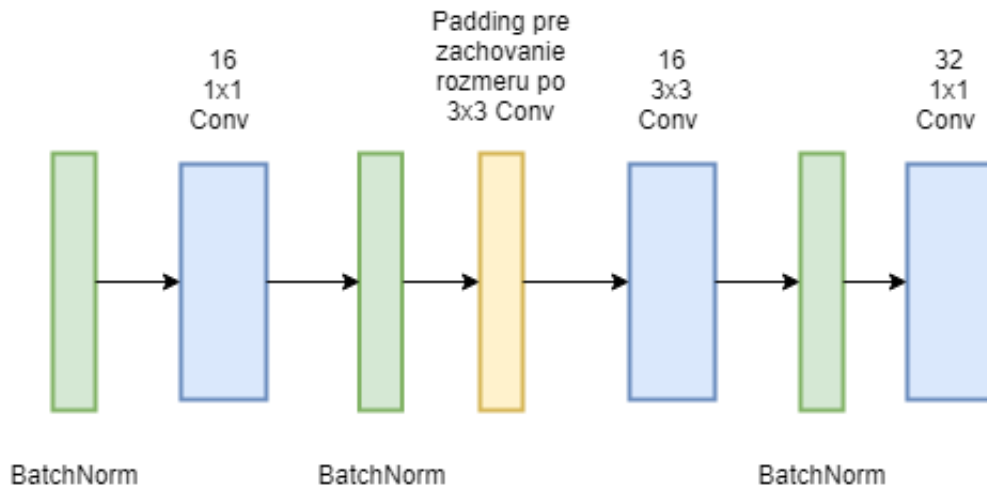
5.5 Prístup s použitím Stacked Hourglass Modelu

Druhý prístup, ktorého výsledky boli sľubné, vychádza zo stacked hourglass modelu popísaného v kapitole 3.3. Podobne ako pri jeho použití pri určovaní pózy človeka je model implementovaný ako plne konvolučná sieť, ktorej výstupy sú heatmappy pre každý landmark znázorňujúce predpokladanú pozíciu landmarku. Vstupom modelu sú podobne ako pri klasifikačnom prístupe snímky s rozlíšením 128x128. Labelom pre každý snímok je sada heatmáp s rozmermi 128x128. Každá heatmapa je labelom pre jeden landmark. Okolo pixelu, na ktorom sa landmark nachádza sú hodnoty dvojrozmerného Gaussovského rozloženia so stredom v tomto pixelu. Veľkosť oblasti, v ktorej sa tieto hodnoty nachádzajú je 10x10 pixelov. Pri tomto prístupe bolo detekovaných všetkých 19 landmarkov s dostupnou anotáciou namiesto 16, ktoré boli detekované v klasifikačnom prístupe.

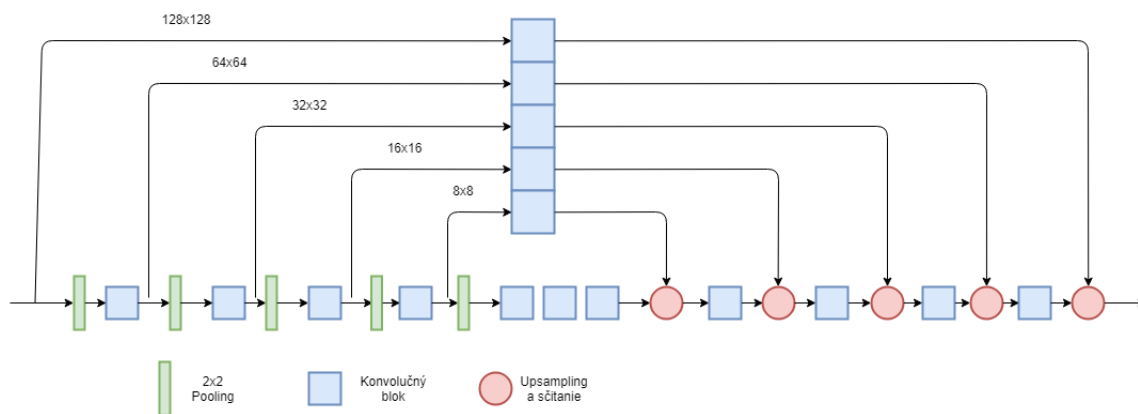
Architektúra

Výsledná architektúra vychádza zo stacked hourglass modelu pre určenie pózy človeka [16] avšak je jednoduchšia keďže sú použité len 2 hourglass moduly. Podľa výsledkov pri odhadovaní pózy človeka však väčší počet modulov ako 2, má väčšiu presnosť len minimálne avšak 2 moduly majú znateľne lepšiu presnosť ako 1. V rámci zlepšenia presnosti siete je aplikovaná loss funkcia aj čiastkový výstup a to výstup prvého hourglass modulu. Výstup tohto modulu vytvára prvotný odhad, ktorý je ďalej spracovávaný druhým modulom. Konvolúcie použité v architektúre majú buď jadrá o veľkosti 3x3 alebo 1x1. 1x1 konvolučné vrstvy sa v modeli používajú pre redukciu alebo zväčšenie dimenzionality výstupov predchádzajúcich konvolučných vrstiev. Tieto konvolúcie sú využívané najmä pre reziduálne spoje kedy je potrebné kombinovať informácie z minulých vrstiev a tieto nemusia mať rovnakú dimenzionalitu.

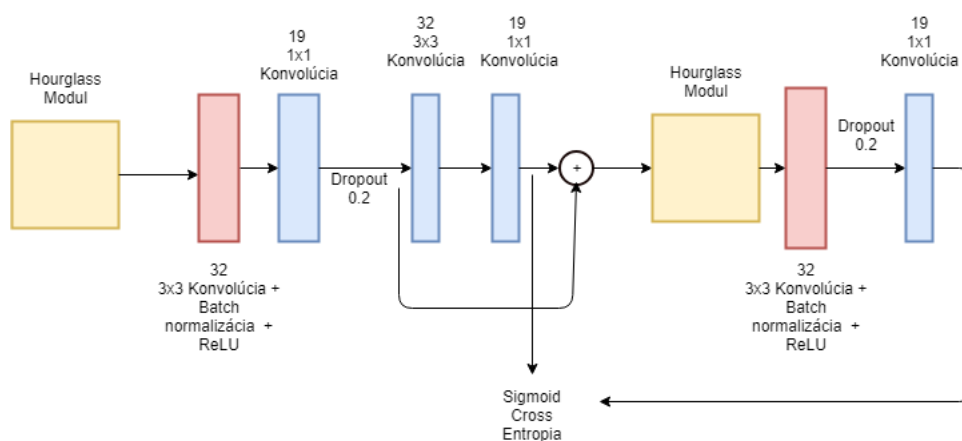
V samotnom module prebieha downsampling v štyroch krokoch na rozlíšenie 8x8, počet príznakových máp je stále rovnaký a to 32. Následne prebieha upsampling a kombinácia príznakov z downsamplingovej časti. Rozdiel oproti ostatným podobným plne konvolučným sieťam je v tom, že upsampling je namiesto dekonvolučných vrstiev vykonávaný metódou najbližšieho suseda. K výstupom týchto upsamplingov sú potom prirátané príznaky v rovnakom rozlíšení z konvolučnej časti. Výstup prvého hourglass modulu je ešte spracovaný niekoľkými konvolúciami. Ako chybová funkcia je použitá cross entropia a optimalizátor je Adam [12].



Obr. 5.4: Architektúra konvolučného bloku používaného v hourglass module



Obr. 5.5: Architektúra hourglass modulu



Obr. 5.6: Architektúra celého modelu

Okrem celej architektúry bolo treba aj experimentovať z vhodnou podobou labelu. Label v podobe one hot mapy by pre túto úlohu nebol vhodný keďže relatívne presné pred-

ikcie by boli striktné penalizované. To znamená, že každá predikcia v okolí landmarku by bola nesprávna. Spôsob riešenia labelu v práci je vytvorenie Gaussovského rozloženia okolo landmarku. Toto zlepšuje možnosť siete dokonvergovať k správnejmu odhadu. Najvhodnejšie sa ukázalo Gaussovské rozloženie 10x10. Boli vyskúšané aj menšie veľkosti avšak výsledky s nimi boli horšie. Bohužiaľ sa do modelu, ktorého výsledky sú prezentované nepodarilo zakomponovať iný výber z výslednej heatmapy ako výber indexu maxima z tejto heatmapy. Keďže výsledná heatmapa mala približne podobu Gaussovky, bolo by zaujímavé vybrať výslednú súradnicu ako stred Gaussovky preloženéj cez túto heatmapu avšak toto riešenie sa nepodarilo implementovať.

Intermediate Supervision

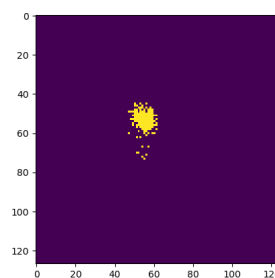
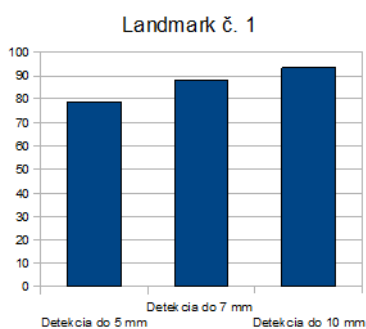
Najväčšou výhodou stacked hourglass modelu oproti jednoduchému hourglass modelu je zaradenie viacerých hourglass modulov za seba. Toto umožňuje sieti iteratívne spracovávať príznaky zo všetkých rozlíšení kdežto pri bežnej sieti sú príznaky vyššej úrovne pri prechode sieťou stratené. Pri každej takejto iterácii, kde jednu iteráciu reprezentuje jeden hourglass modul je navyše možné aplikovať loss funkciu na výstup a tým pádom aplikovať loss funkciu voči ground truth nielen na celkový výstup ale aj na tieto čiastkové výstupy. Takéto spracovanie umožňuje sieti lepšie sa naučiť priestorové závislosti landmarkov, ktoré sú dôležité pri finálnom stanovení súradníc landmarku.

Kapitola 6

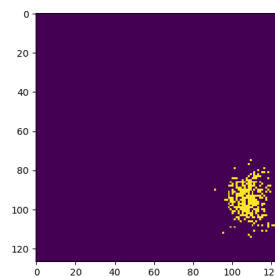
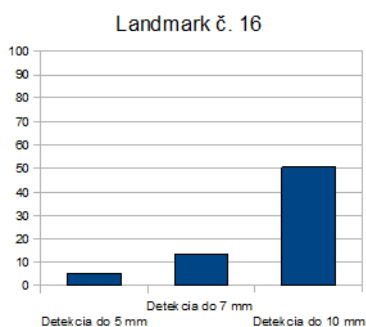
Výsledky

6.1 Klasifikačný model

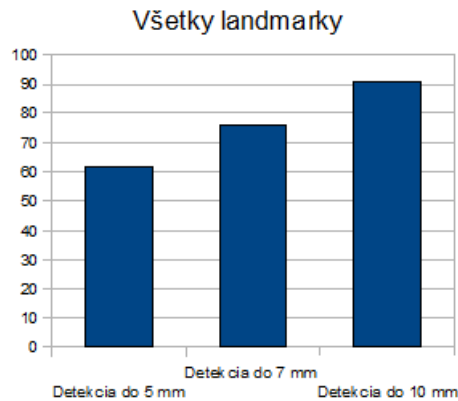
Výsledky klasifikačného modelu sú získané z najlepšej nájdenej konfigurácie siete pre presnosť na testovacích dátach. Trénovanie bolo ukončované, keď výsledky piatich po sebe idúcich validácií boli horšie ako predchádzajúca validácia. Validácia prebiehala po každých trinástich epochách. Grafy sú uvedené pre landmarky s najlepšou a najhoršou detekciou do 5 mm. Takisto je vytvorený priemer výsledkov tohto modelu.



Obr. 6.1: Úspešnosť siete na landmarku č. 1 Obr. 6.2: Variabilita umiestnenia landmarku č. 1



Obr. 6.3: Úspešnosť siete na landmarku č. 16 Obr. 6.4: Variabilita umiestnenia landmarku č. 16



Obr. 6.5: Úspešnosť siete na všetkých landmarkoch

Landmark	<5 mm (%)	<7 mm (%)	<10 mm (%)	AVG v osi x (mm)	AVG v osi y (mm)
Sella	78.8	88.3	93.2	4.56	4.48
Nasion	71.6	83.9	93	4.59	5.31
Orbitale	51.25	73.65	91.9	5.56	7.39
Porion	64.65	79.6	92.95	5.63	5.46
Subspinale	74.45	88.25	95.9	4.07	5.32
Supramentale	33.8	51.35	84.8	5.16	10.13
Pogonion	66.85	81	94.6	5	5.12
Menton	65.9	79.15	93.75	5.24	5.05
Gnathion	64.7	79	94.6	5.24	5.02
Gonion	60.45	77	91.95	5.09	6.61
Incision inferius	77.3	89.05	96.2	4.16	4.51
Incision superius	73.5	85.4	95.9	4.69	4.68
Vrchná pera	59.7	76.3	93.35	4.37	6.97
Spodná pera	64.5	78.65	92.65	4.1	6.3
Subnasale	76.05	88.15	94.9	4.36	4.89
Pogonion MT	5.25	13.55	50.5	6.54	17.29
Priemer	61.8	75.77	90.63	4.9	6.53

Tabuľka 6.1: Úspešnosť siete na landmarkoch. Zelenou sú zvýraznené landmarky, ktoré majú úspešnosť väčšiu ako 70%. Červenou tie pod 65%.

Zhodnotenie

Okrem výslednej architektúry a hyperparametrov, pre ktorú sú uvedené výsledky boli takisto vyskúšané architektúry s menším počtom konvolučných vrstiev avšak tie nedosahovali také dobré výsledky. Ukázalo sa, že čím hlbšia sieť je použitá pri tomto prístupe, tým je presnejšia. Výsledky boli získané s veľkosťou batchu 50, koeficientom učenia nastaveným na 0.05. Okrem softmaxu aplikovaného na výstup bola takisto vyskúšaná sigmoida avšak výsledky softmaxu boli lepšie.

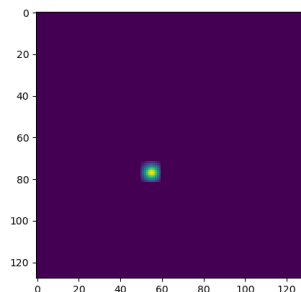
Z výsledkov vidíme, že rozdiel pri odhadovaní landmarku č. 1 oproti landmarku č. 16 je priepastný. Landmark č. 16 je *pogonion* čo znamená najvonkajší bod brady. Je to spôsobené rozličnou charakteristikou oblastí okolo landmarkov a taktiež kvôli veľkej variabilite umiestnenia landmarku oproti ostatným landmarkom. To znamená že sieť sa lepšie naučila rozpoznávať štruktúru, ktorá sa nachádza okolo landmarku 1 ako okolo landmarku 16. Vo výsledkoch vidíme, že sieť robilo problém pri landmarku 16 odhadovať súradnicu y. Zlá schopnosť detekovať landmark č. 16 platila aj pre nasledujúci model. Taktiež je možné povedať, že landmark 16 je najproblémovejším landmarkom pre detekciu aj u expertov keďže v dostupnom datasete bol rozdiel medzi dvoma expertami pri detekcii landmarku 16 2.53 mm v osi x a 6 mm v osi y. V rámci 16 landmarkov dosiahla sieť na 5 mm vzdialenosť úspešnosť 61.8%.

Okrem architektúry, ktorá odhadovala všetky pozície naraz bola vyskúšaná aj architektúra, ktorá odhadovala pozíciu len jedného landmarku. Natrénovanie takejto architektúry bolo vykonané pre 3 landmarky, konkrétne 4, 8, 16. Výsledky tejto architektúry boli horšie. Na landmarku 8 sa odhad pre vzdialenosť menšiu ako 5 mm líšil zanedbateľne avšak na landmarku 4 bol odhad horší až o okolo 16%. Na landmarku 16 bol odhad rovnako zlý.

6.2 Stacked Hourglass Model

Výsledky Stacked Hourglass Modelu boli takisto získané z najlepšej nájdennej konfigurácie. Veľkosť batchu pri tejto konfigurácii bola 16. Trénovanie bolo zastavené ak tri po sebe idúce validácie mali výsledky horšie ako validácia pred nimi. Validácie prebiehaly každých 1000 batchov. Koeficient učenia mal hodnotu 0.005 v prvých 2000 batchoch, následne sa jeho hodnota znížila na 0.0005. Táto zmena bola vykonaná kvôli zvýšeniu rýchlosti tréovania a osvedčila sa keďže tréovanie pri koeficiente 0.0005 trvalo približne 8 hodín a nebolo vidieť výrazné skoky v presnosti medzi validáciami a pri koeficiente 0.005 sa model rýchlo natrénoval na presnosť okolo 40% a pri prvej validácii s koeficientom 0.0005 sa táto presnosť znateľne zvýšila. Veľkosť batchu bola limitovaná na 16 kvôli neschopnosti GPU spracovať väčší batch.

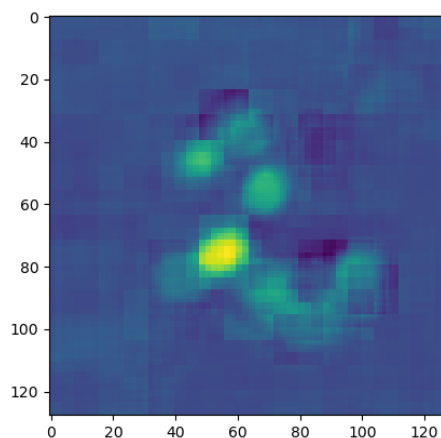
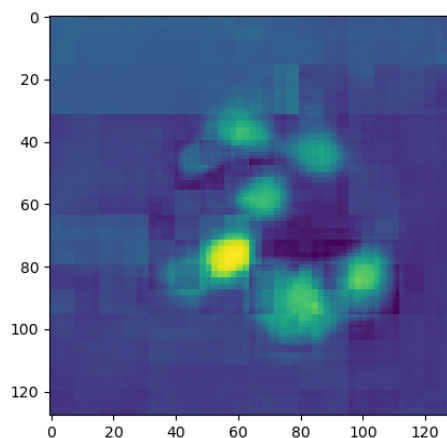
Počas tréovania bolo zaujímavé sledovať ak vyzerá výstup modelu, ktorým boli heat-mapy. Pri prvých iteráciách tréovania sa výstupy pre všetky landmarky dosť podobali a až neskôr bolo v heatmapách jasne vidieť landmark pre ktorý je heatmapa určená.



Obr. 6.6: Ground truth pre snímok 6000 a landmark 4

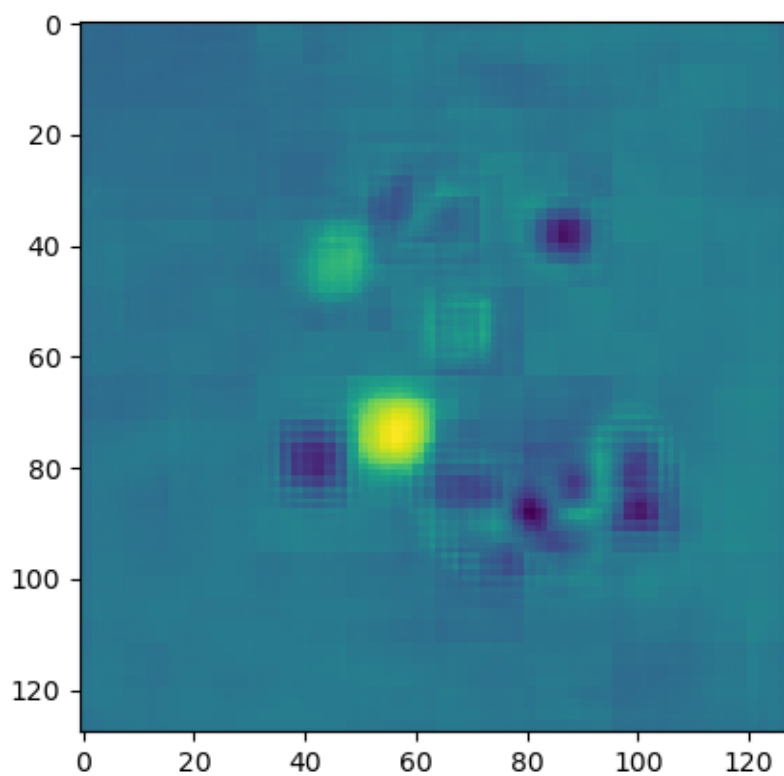


Obr. 6.7: Snímok 6000



Obr. 6.8: Heatmapa po 500 batchoch pre snímok 6000, landmark 4

Obr. 6.9: Heatmapa po 1000 batchoch pre snímok 6000, landmark 4

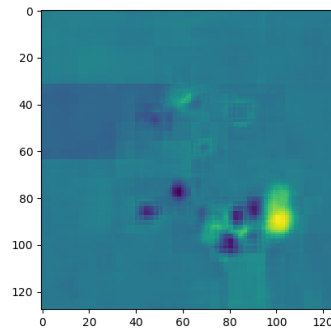
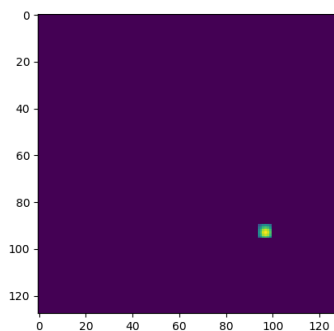


Obr. 6.10: Heatmapa pri konci trénovania (5000) pre snímok 6000, landmark 4

Z obrázkov môžeme vidieť vývoj heatmáp kedy na prvej heatmape vidíme, že sú vysoké hodnoty na miestach, kde sieť odhaduje iné landmarky a hodnoty blízke nule tam, kde sieť neodhaduje žiadny landmark. Tento jav bol vidieť najmä pri skorších fázach tréningu pri koeficiente učenia 0.0005 kde po prvej validácii vyzerali heatmapy pre všetky landmarky rovnako. Avšak tu už je vidieť, že si sieť istá, že landmark 4 sa nachádza v určitom priestore. Tento priestor je ale stále dosť veľký a najvyššie hodnoty, ktoré sa v ňom vyskytujú sú podobné.

Na ďalšej heatmape vidíme, že trend pri tréningu je znižovanie hodnôt tam, kde sa predpokladajú iné landmarky ako ten, voči ktorému je heatmapa tréningovaná. Okrem toho priestor, kde sa predpokladá pozícia landmarku 4 má po jeho okrajoch menšie hodnoty ako pri prvej heatmape a počet blízkych vysokých hodnôt sa zmenšil a tieto hodnoty sa presunuli do stredu priestoru. Okrem toho sa zväčšujú hodnoty v miestach, kde sieť nepredpokladá pozíciu žiadneho landmarku.

Pri konci tréningu už môžeme vidieť veľmi nízke hodnoty v miestach iných landmarkov, zväčšené hodnoty v miestach kde nie je žiaden landmark a hlavne to, že priestor, kde sa predpokladá pozícia landmarku 4 sa zmenšil a pripomína ground truth heatmapu.



Obr. 6.11: Ground truth pre snímok 6000, Obr. 6.12: Heatmapa pri konci tréningu landmark 16 (5000) pre snímok 6000, a landmark 16

Podobne ako pri klasifikačnom prístupe bol najväčší problém s detekovaním landmarku č. 16. Výsledky tohoto landmarku boli vo vzdialenosti do 5 mm 10.5% čo je lepšie ako 5.25% v klasifikačnom prístupe ale stále boli rádovo horšie ako výsledky ostatných landmarkov. Problém bol opäť v určovaní súradnice y kedy priemerná vzdialenosť predikcie od reálu bola 12.14 mm. Vo výslednej heatmape siete je skutočne vidieť vysokú variabilitu v ose y aj po dlhom tréningu.

Landmark	<5 mm (%)	<7 mm (%)	<10 mm (%)	AVG v osi x (mm)	AVG v osi y (mm)
Sella	89.5	94.15	97.35	2.68	2.75
Nasion	75.85	88.4	95.55	3.37	4.3
Orbitale	54.75	77.45	94.95	4.85	5.6
Porion	75.4	86.35	93.75	4.07	3.93
Subspinale	81.4	93.15	96.8	3.16	3.89
Supramentale	47.75	68.6	91.5	4.21	6.84
Pogonion	80.95	91.55	97.3	3.7	3.52
Menton	81.8	89.8	97.15	3.94	3.31
Gnathion	85.45	91.35	96.65	3.75	3.36
Gonion	63.85	80.4	94.85	4.38	4.83
Incision inferius	87.35	94.4	97.65	3.22	3.29
Incision superius	82.5	92.65	97.5	3.77	3.45
Vrchná pera	74.4	88.75	96	3.33	4.68
Spodná pera	72.7	85.05	95.4	3.54	4.64
Subnasale	82.75	91.85	97.15	3.45	3.52
Pogonion MT	10.5	22.8	64.3	6.45	12.14
Posterior nasal spine	86.1	91.85	96.65	3.67	3.08
Anterior nasal spine	80	89.3	95.65	4.12	3.5
Articulare	79.8	91.35	96.45	3.46	3.63
Priemer 16-tich LM	71.58	83.53	93.99	3.87	4.62
Priemer	73.22	84.69	94.35	3.85	4.43

Tabuľka 6.2: Úspešnosť siete na landmarkoch. Zelenou sú zvýraznené landmarky, ktoré majú úspešnosť väčšiu ako cca 70%. Červenou tie pod 65%.

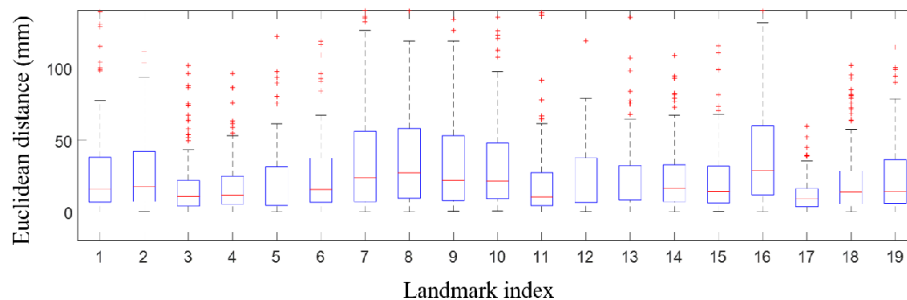
6.3 Vyhodnotenie metód

Z výsledkov vidíme, že metóda s pomocou plne konvolučnej siete dosahovala väčšiu presnosť na všetkých landmarkoch. Táto sieť mala však aj väčšie množstvo parametrov a z toho vyplývala aj rádoovo dlhšia doba natrénovania tejto siete.

Úspešnosti sietí dosahovali lepšie výsledky ako niektoré metódy [13] avšak oproti state of the art prístupom zaostávali.

Porovnanie so state of the art metódami

Pri porovnávaní state of the art modelov nasaditeľných v reálnych podmienkach sa väčšinou používa metrika, kedy je landmark detekovaný správne, ak je chyba menšia ako 2 mm a akceptovateľný ak je chyba menšia ako 4 mm. Pri 4 mm je model vhodný na detekovanie tohto landmarku ak pri jeho detekovaní dosahuje úspešnosť aspoň 80%. Bohužiaľ podľa tejto metriky nie sú prezentované výsledky dostatočné keďže na 4 mm dosahuje najlepší model priemernú úspešnosť 56.16% a na najlepšie detekovanom landmarku 78.3% (landmark 1). Napríklad kombinácia konvolučnej siete a shape modelu [6] dosahovala pri detekcii do dvoch milimetrov úspešnosť na všetkých landmarkoch 75.58% čo je viac ako vytvorený model dosahoval pri detekcii do piatich milimetrov.



Obr. 6.13: Úspešnosť metódy založenej na natrénovaní jedného modelu pre jeden landmark. Prevzaté z [13].

Z obrázku vidíme, že obidve navrhnuté metódy prekonal metódu založenú na regre-
sii pozície každého landmarku zvlášť. Avšak treba dodať, že pri tejto metóde bol rozmer
obrázka len 64x64 a dataset nebol augmentovaný.

Reálne chyby pri detekovaní

Index landmarku	Priemerný rozdiel odha- dov v datasete v ose X	Priemerný rozdiel odha- dov v datasete v ose Y
1	0.43	0.74
2	0.78	1.19
3	2.28	1.66
4	1.5	2.55
5	1.32	2.39
6	1.27	2.62
7	0.34	1.16
8	1.24	0.3
9	0.71	0.63
10	1.92	1.91
11	0.42	0.43
12	0.33	0.29
13	1.02	2.4
14	1.46	1.45
15	0.85	1.15
16	2.53	6
17	1.22	0.54
18	1.68	0.49
19	0.83	1.81

Tabuľka 6.3: Rozdiely medzi dvoma odhadmi expertami

Z tabuľky môžeme vidieť, že rozdiely odhadu medzi expertami sú menšie ako chyby pri
odhade sieťou avšak taktiež môžeme vidieť, že pomer chyby medzi dvoma landmarkami
odhadovanými expertami kopíruje pomer chyby medzi landmarkami odhadovanými sieťou.
Takisto je vidieť, že na umiestnení landmarku 16 robí expertom problém sa zhodnúť a
priemerný rozdiel v súradnici y je veľmi veľký v porovnaní z ostatnými rozdielmi. Opačne
je to pri landmarku 1 kde je rozdiel najmenší, čo znamená, že sa ľahko detekuje.

Kapitola 7

Záver

Cielom práce bolo experimentovanie z rôznymi architektúrami konvolučných neurónových sietí a ich nastavením pri úlohe detekcie cefalometrických landmarkov. Pre úlohu takejto detekcie má z použitých metód najväčší potenciál stacked hourglass model. Okrem experimentovania bolo cieľom získať poznatky o detekcii cefalometrických landmarkov a state of the art metódach, ktoré sa pre túto úlohu používajú a naučiť sa pracovať s knižnicou Tensorflow [5].

Okrem spomínaného stacked hourglass modelu sa taktiež experimentovalo s využitím klasifikačného prístupu avšak oproti výsledkom stacked hourglass modelu tento prístup značne zaostával. Úspešnosť stacked hourglass modelu bola 73.22% pri detekcii do piatich milimetrov. Úspešnosť klasifikačnej architektúry bola 61.8% pri detekcii do piatich milimetrov. Klasifikácia toho, či sa na danom pixely nachádza landmark bola úspešná najmä v kombinácii so shape modelmi [6] takže aj tento prístup k detekcii landmarkov sa používa v state of the art metódach.

Do budúcnosti by bolo vhodné vyskúšať použitie viac hourglass modulov ako dva a taktiež zväčšiť vstupné rozlíšenie obrázku. Takisto experimentovanie s výberom výslednej hodnoty výstupu siete by bolo vhodné. Okrem toho sa v niektorých state of the art metódach osvedčilo ako vstup siete dávať len výrez okolo detekovaného landmarku čo by umožnilo aj väčšie možnosti pri augmentácii datasetu.

Literatúra

- [1] Convolutional Neural Network. URL: <https://brilliant.org/wiki/convolutional-neural-network/>.
- [2] CS231n Convolutional Neural Networks for Visual Recognition. URL: <http://cs231n.github.io/convolutional-networks/>.
- [3] For Dummies - The Introduction to Neural Networks we all need ! (Part 1). URL: <https://medium.com/technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-c50f6012d5eb>.
- [4] Overfitting / Underfitting – How Well Does Your Model Fit? URL: <https://meditationsonbianddatascience.com/2017/05/11/overfitting-underfitting-how-well-does-your-model-fit/>.
- [5] Abadi, M.; Barham, P.; Chen, J.; et al.: TensorFlow: A System for Large-scale Machine Learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, Berkeley, CA, USA: USENIX Association, 2016, ISBN 978-1-931971-33-1, s. 265–283. URL <http://dl.acm.org/citation.cfm?id=3026877.3026899>
- [6] Arik SÖ, X. L., Ibragimov B: Fully automated quantitative cephalometry using convolutional neural networks. *Journal of Medical Imaging*, 1 2017.
- [7] Badrinarayanan, V.; Kendall, A.; Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *ArXiv e-prints*, November 2015, [1511.00561](https://arxiv.org/abs/1511.00561).
- [8] Belagiannis, V.; Rupprecht, C.; Carneiro, G.; et al.: Robust Optimization for Deep Regression. 05 2015.
- [9] Bradski, G.: The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [10] Ibragimov, B.; Likar, B.; Pernus, F.; et al.: A Game-Theoretic Framework for Landmark-Based Image Segmentation. *IEEE Transactions on Medical Imaging*, ročník 31, č. 9, Sept 2012: s. 1761–1776, ISSN 0278-0062, doi:10.1109/TMI.2012.2202915.
- [11] Kazemi, V.; Sullivan, J.: One Millisecond Face Alignment with an Ensemble of Regression Trees. In *CVPR*, 2014.
- [12] Kingma, D. P.; Ba, J.: Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, December 2014, [1412.6980](https://arxiv.org/abs/1412.6980).

- [13] Lee, H.; Park, M.; Kim, J.: Cephalometric landmark detection in dental x-ray images using convolutional neural networks. 03 2017.
- [14] Lindner, C.; Wang, C.-W.; Huang, C.-T.; aj.: Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. ročník 6, 09 2016: str. 33581.
- [15] Majd, M.; Shoeleh, F.: A Novel Hybrid Approach for Cephalometric Landmark Detection. *ArXiv e-prints*, Jún 2015, [1506.03936](#).
- [16] Newell, A.; Yang, K.; Deng, J.: Stacked Hourglass Networks for Human Pose Estimation. *ArXiv e-prints*, Marec 2016, [1603.06937](#).
- [17] Park, S.: Cephalometric Landmarks Detection using Fully Convolutional Networks. 08 2017.
- [18] Russakovsky, O.; Deng, J.; Su, H.; aj.: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, ročník 115, č. 3, 2015: s. 211–252, doi:10.1007/s11263-015-0816-y.
- [19] Toshev, A.; Szegedy, C.: DeepPose: Human Pose Estimation via Deep Neural Networks. *ArXiv e-prints*, December 2013, [1312.4659](#).
- [20] Tseng, Y.-C.; Lai, S.; Lee, H.-E.; aj.: Are Hyoid Bone and Tongue the Risk Factors Contributing to Postoperative Relapse for Mandibular Prognathism? ročník 2016, 03 2016: s. 1–7.
- [21] Wang, C. W.; Huang, C. T.; Hsieh, M. C.; aj.: Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge. *IEEE Transactions on Medical Imaging*, ročník 34, č. 9, Sept 2015: s. 1890–1900, ISSN 0278-0062, doi:10.1109/TMI.2015.2412951.
- [22] Wu, H.; Bailey, C.; Rasoulinejad, P.; aj.: Automatic Landmark Estimation for Adolescent Idiopathic Scoliosis Assessment Using BoostNet. 09 2017.
- [23] Wu, Y.; Hassner, T.; Kim, K.; aj.: Facial Landmark Detection with Tweaked Convolutional Neural Networks. *ArXiv e-prints*, November 2015, [1511.04031](#).
- [24] Zhang, J.; Liu, M.; Shen, D.: Detecting Anatomical Landmarks From Limited Medical Imaging Data Using Two-Stage Task-Oriented Deep Neural Networks. *IEEE Transactions on Image Processing*, ročník 26, č. 10, Oct 2017: s. 4753–4764, ISSN 1057-7149, doi:10.1109/TIP.2017.2721106.

Prílohy

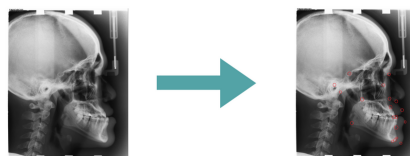
Príloha A

Plagát

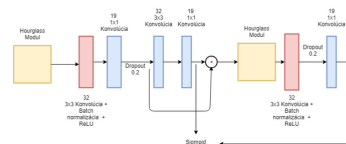


Detekce významných bodů v medicínských obrazech pomocí hlubokých neuronových sítí

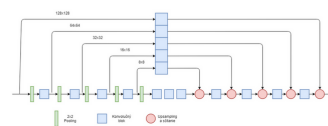
Autor: J. Škandera, xskand01@stud.fit.vutbr.cz
Vedúci: O. Kodým, ikodym@fit.vutbr.cz



Úlohou je detekcia 19 landmarkov v cefalometrických snímkoch. Najlepšie výsledky boli dosiahnuté plne konvolučnou architektúrou s využitím reziduálnych spojov, zaradením viacerých encoder-decoder modulov za seba a aplikovaním chybovej funkcie na každý e-d modul.

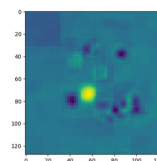


Architektúra modelu kde hourglass modul je jeden encoder-decoder blok.



Architektúra jedného encoder-decoder (hourglass) bloku.

Trénovanie prebiehalo s datasetom augmentovaným rotáciami. Vstupom siete je 128x128 snímok a výstupom je sada heatmáp.



Úspešnosť detekcie na 5 mm vzdialenosť bola 73.22%. Priemerná vzdialenosť landmarku od odhadu bola 3.85 mm v smere osi x a 4.43 mm v smere osi y. 10 landmarkov sa podarilo detekovať do 5 mm s úspešnosťou väčšou ako 80%. Landmarky detekované s úspešnosťou menšou ako 70% boli 4.

