



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

**INTERAKTIVNÍ SEGMENTACE 3D CT DAT
S VYUŽITÍM HLUBOKÉHO UČENÍ**

INTERACTIVE 3D CT DATA SEGMENTATION BASED ON DEEP LEARNING

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. KATEŘINA TRÁVNÍČKOVÁ

VEDOUcí PRÁCE

SUPERVISOR

Ing. OLDŘICH KODYM

BRNO 2020

Zadání diplomové práce



Studentka: **Trávníčková Kateřina, Bc.**

Program: Informační technologie Obor: Počítačová grafika a multimedia

Název: **Interaktivní segmentace 3D CT dat s využitím hlubokého učení**
Interactive 3D CT Data Segmentation Based on Deep Learning

Kategorie: Zpracování obrazu

Zadání:

1. Prostudujte základy konvolučních sítí a jejich aplikací na segmentaci 3D CT dat.
2. Vytvořte si přehled o současných metodách interaktivní segmentace.
3. Vyberte nebo navrhnete metodu interaktivní segmentace CT dat s využitím konvolučních neuronových sítí.
4. Obstarejte si databázi vhodnou pro experimenty.
5. Implementujte navrženou metodu a proveďte experimenty nad 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:

- Ronneberger, Olaf, Philipp Fischer and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." MICCAI (2015).
- Bredell, Gustav, Christine Tanner and Ender Konukoglu. "Iterative Interaction Training for Segmentation Editing Networks." MLMI (2018).

Při obhajobě semestrální části projektu je požadováno:

- Body 1 až 3.

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

Vedoucí práce: **Kodym Oldřich, Ing.**

Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.

Datum zadání: 1. listopadu 2019

Datum odevzdání: 3. června 2020

Datum schválení: 20. listopadu 2019

Abstrakt

Tato práce se zaměřuje na možnosti využití neuronových sítí pro segmentaci CT dat s omezenými možnostmi použití anotovaných dat. Hlavním prostředkem pro zlepšení kvality segmentace kostí pomocí modelu trénovaného na malé datové sadě je zde přidání uživatelské interakce. Dále jsou zkoumány možnosti využití transfer learningu v podobě předtrénování na interaktivní segmentaci na jiné než cílové datové sadě a v podobě předtrénování na restauraci dat pomocí cílové datové sady. Všechny zkoumané metody přinášejí určité zlepšení oproti baseline metodě, kterou je použití datově specifického automatického segmentačního modelu. Při trénování s velmi malými trénovacími množinami dochází ke zvýšení hodnoty Dice skóre až o desítky procent. Praktické uplatnění těchto metod může být například v jejich použití coby nástroje pro urychlení tvorby nového segmentačního datasetu.

Abstract

This thesis deals with CT data segmentation using convolutional neural nets and describes the problem of training with limited training sets. User interaction is suggested as means of improving segmentation quality for the models trained on small training sets and the possibility of using transfer learning is also considered. All of the chosen methods help improve the segmentation quality in comparison with the baseline method, which is the use of automatic data specific segmentation model. The segmentation has improved by tens of percents in Dice score when trained with very small datasets. These methods can be used, for example, to simplify the creation of a new segmentation dataset.

Klíčová slova

Zpracování obrazu, segmentace, hluboké učení, umělá inteligence, konvoluční neuronové sítě, transfer learning, medicínská data, CT data, objemová data, shlukování, extrakce příznaků, rekonstrukce obrazu, interaktivní segmentace, benchmarking, omezená trénovací množina.

Keywords

Image processing, segmentation, deep learning, artificial intelligence, convolutional neural nets, transfer learning, medical data, CT data, volume data, clustering, feature extraction, image restoration, interactive segmentation, benchmarking, limited training set.

Citace

TRÁVNÍČKOVÁ, Kateřina. *Interaktivní segmentace 3D CT dat s využitím hlubokého učení*. Brno, 2020. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Oldřich Kodým

Interaktivní segmentace 3D CT dat s využitím hlubokého učení

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně pod vedením pana Ing. Oldřicha Kodyma. Uvedla jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpala.

.....
Kateřina Trávníčková
25. června 2020

Poděkování

Děkuji vedoucímu této práce, Ing. Oldřichu Kodymovi, za jeho trpělivost a ochotu. Rovněž děkuji firmě TESCAN 3Dim za poskytnutí datasetů a lidem stojícím za virtuální organizací MetaCentrum za možnost využití výpočetního gridu.

Obsah

1 Úvod	2
2 Segmentace a metody hlubokého učení	4
2.1 Vymezení úlohy segmentace a možné přístupy k jejímu řešení	4
2.2 Hluboké učení v současné době	8
2.3 Problém s nedostatkem dat	11
2.4 Segmentace pomocí shlukování a CNN	16
3 Vlastní řešení segmentace CT dat	18
3.1 Společná konfigurace metod a trénovacího procesu	18
3.2 Vybraná rozšíření metod při použití malých datasetů	19
3.3 Dataset	24
3.4 Sada skupin modelů pro benchmarking metod	25
3.5 Implementace skriptů	27
4 Experimenty a vyhodnocení	30
4.1 Porovnání základních variant zvolených metod s automatickým modelem . .	30
4.2 Kombinace interaktivity a předtrénování na restauraci	39
4.3 Předtrénování na zdrojových datech	40
4.4 Rozšířená interaktivní metoda	41
4.5 Diskuze	42
5 Závěr	45
Literatura	47
A Numerické výsledky metod	51
B Vývoj Dice skóre při trénování	53
C Plakát	56

Kapitola 1

Úvod

Tato diplomová práce se zabývá problematikou analýzy snímků počítačové tomografie (dále zkracováno jako CT z angl. *computed tomography*) s využitím současných poznatků z oboru umělé inteligence. Konkrétněji se specializuje na medicínská data v podobě snímků lidského těla a na možnosti částečně asistované sémantické segmentace, která je zde chápána jako nalezení pixelů patřících jednomu anatomickému celku. Práce se zaměřuje na segmentaci kostí, avšak metody zde zkoumané mohou najít uplatnění i při segmentaci jiných celků.

Obor umělé inteligence, respektive obor hlubokého učení, zažívá několik posledních let rapidní rozvoj a zasahuje do spousty odvětví zpracování dat. Zpracování medicínských dat je důležitou oblastí, jejíž zdokonalení může mít výrazný přínos pro usnadnění práce odborníkům z řad lékařů a vést tak nepřímo i ke zlepšení kvality lékařské péče. Využití nejmodernějších přístupů, včetně poznatků z oboru hlubokého učení, je tedy na místě. Právě úloha segmentace patří mezi nejdiskutovanější a z hlediska praxe nejzajímavější z úloh zpracování medicínských dat, díky čemuž dlouhodobě probíhá v této oblasti intenzivní výzkum. Metody využívající umělé inteligence, zejména pak hlubokého učení, přitom dosahují nejlepších výsledků [29, 46].

Již svou bakalářskou práci jsem věnovala seznámení se s možnostmi využití hlubokého učení pro analýzu medicínských dat a tato činnost mě následně nasměřovala k pracovní činnosti zaměřené na vývoj tematicky související aplikace používané v praxi (mimo jiné) pro návrh lebečních implantátů. Moje motivace pro volbu výše popsaného tématu tedy vychází z předchozího studia, z mého současného pracovního zaměření a konečně také z osobního zájmu o dané téma a zájmu o prohloubení znalostí této problematiky.

V dnešní době již existují ověřené postupy segmentace medicínských dat s využitím hlubokého učení. Nejúspěšnější řešení zpravidla využívají učení s rozsáhlou profesionálně anotovanou datovou sadou. Tato práce si klade za cíl prověřit možnosti řešení situací, kdy je použití trénovacích dat omezené. Tímto omezením může být buď nedostatek/nepřítomnost anotací nebo celkově malé množství dat. Zkoumáno je použití interaktivního modelu jako alternativy k modelu automatickému a možnost předtrénování modelu na jiných datech či jiné úloze.

Kapitola 2 popisuje související teoretický podklad pro vymezení kontextu v kterém se tato práce pohybuje. Je zde přiblížena úloha segmentace a popsány různé přístupy, které lze pro její řešení využít. Uvedeny jsou i příklady metod používaných v praxi. Lze zde najít

úvod do problematiky hlubokého učení a přehled zajímavých řešení z oblasti zpracování obrazových, objemových a zejména medicínských dat. Kapitola 3 popisuje krátce metody navržené v rámci semestrálního projektu a blíže jsou zde vysvětleny konkrétní postupy a metody, které byly vybrány na základě poznatků zmíněných v první kapitole. V kapitole 4 jsou potom popsány výsledky experimentů s vybranými a navrženými metodami a je zde popsán pravděpodobný další možný vývoj experimentů.

Kapitola 2

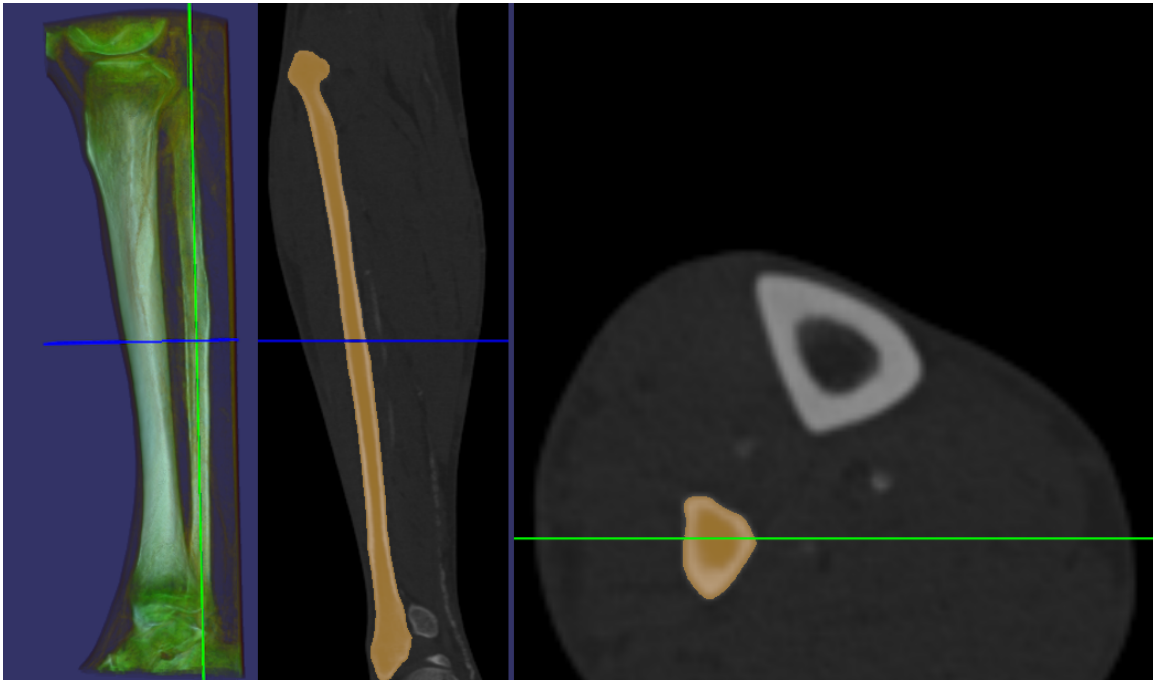
Segmentace a metody hlubokého učení

Tato kapitola slouží jako úvod do problematiky segmentace v kontextu zpracování obrazových (objemových) dat, zejména pak dat z oblasti medicíny. Je zde definována úloha segmentace tak, jak je dále chápána v této práci, a související metody, které se používají nebo v minulosti používaly k jejímu řešení. Součástí kapitoly je také stručný úvod do problematiky hlubokého učení a popis prostředků hlubokého učení, které jsou využívány k řešení úlohy segmentace nebo mohou být v kontextu této práce relevantní pro její řešení. Zmíněny jsou některé významné architektury použité k segmentaci obrazu případně medicínských dat. Související je i problematika transfer learningu, jejíž popis je zde rovněž obsažen. Stručně jsou zde také popsána specifika analýzy medicínských dat.

2.1 Vymezení úlohy segmentace a možné přístupy k jejímu řešení

Úlohou segmentace obrazu se v základu rozumí rozdělení obrazu do oblastí (regionů, segmentů...) podle zvoleného kritéria. Takovým kritériem může být například podobnost barevných/intenzitních hodnot samotných pixelů, či lokální informace ve zvoleném okolí pixelu (průměrná hodnota, histogram hodnot, histogram orientovaných gradientů...) [10]. Segmentaci lze také chápat jako klasifikaci každého pixelu (dále se zde pro tento způsob zpracování používá pojem "pixelwise") do jedné z daného počtu tříd. Tato definice problému je blízká zejména oblasti segmentace pomocí neuronových sítí, kde se pro klasifikaci celých obrazů a pixelwise klasifikaci používají podobné prostředky. S tímto pohledem také souvisí požadavek na segmentační metody, které umožní rozdělení obrazu na segmenty odpovídající logickým objektům reálného světa. Tento přístup je označován pojmem sémantická segmentace.

Zvolený přístup k segmentaci, a tedy i volba konkrétní metody, se odvíjí od účelu, za kterým je segmentace prováděna. Segmentace může sloužit jako předzpracování vstupu pro další úlohy zpracování obrazu. Obraz je rozdělen do regionů, pro které je charakteristická nějaká vlastnost a informace, která není pro další zpracování důležitá, může být zanedbána. Binární segmentace, tedy segmentace, kdy jsou odlišeny pouze dva druhy segmentů, zpravidla to bývá objekt a pozadí, bývá využívána k výběru objektu zájmu a jeho odlišení od okolí, které pro další zpracování není podstatné. Tento druh segmentace je často používán právě v medicíně, kdy objektem zájmu může být například konkrétní zdravý orgán



Obrázek 2.1: Ukázkový CT snímek lidské nohy s oranžově vyznačenou segmentační maskou pro kost fibula. Snímek je zde zobrazen jako náhled 3D vizualizace a dva možné pohledy v podobě řezů objemem.

ale také anomálie spojená s patologickými změnami typu nádorových onemocnění. Zde je požadavek na takovou segmentační metodu, která umožní co nejpřesnější odlišení jednoho či více regionů tak, aby výběr regionu odpovídal anatomii lidského těla. Konkrétní aplikací segmentace CT dat je například návrh kosterních implantátů firmou TESCOAN Medical¹.

Analýza CT dat

V následujících podkapitolách jsou záměrně zmiňovány především práce a výzkumy zaměřující se na analýzu CT dat. Z množství těchto prací vyplývá, že segmentace CT dat je důležitou oblastí, časté jsou především úlohy segmentace konkrétních kostí, srdce, mozku, cév nebo anomálií v podobě nádorů. Protože jsou CT data objektem zkoumání i v této práci, je potřeba připomenout několik skutečností a omezení, které se s jejich analýzou pojí. Přestože se situace stále zlepšuje a přibývá (i veřejně dostupných) datasetů z této oblasti [31], problém nedostupnosti dostatečného množství dat ještě zdaleka není zanedbatelný. Pro využití metod hlubokého učení je nutné co největší množství dat, což nemusí být velký problém u aplikací pracujících s obrazy okolního světa (detekce poznávacích značek, segmentace obrazu ulice pro automaticky řízená vozidla...). Avšak pořízení kvalitních CT snímků v dostatečném množství může být samo o sobě problematické. Při vytvoření CT snímku dochází k ozáření snímané osoby, což může mít negativní dopad na její zdraví a i v lékařské praxi je snaha frekvenci a trvání CT vyšetření redukovat na nezbytné minimum. Dostupnost těchto snímků také může podléhat omezením díky tomu, že se jedná o potenciálně citlivá data ze strany pacientů. Většinou jsou také CT data pořizována u více či méně patologických případů a množství CT skenů zdravých tkání je díky tomuto faktoru omezené.

¹<https://www.tescan-medical.com>

A konečně uvážíme-li konkrétně úlohu segmentace těchto dat, je pro natrénování specializované segmentační neuronové sítě potřeba velké množství anotací, které pro zachování požadované přesnosti musejí být vytvořeny odborníky. Právě z tohoto důvodu je důležité zkoumat možnosti využití neanotovaných dat pro zlepšení výsledků segmentace.

CT data jsou oproti běžným fotografiím naopak zvýhodněna tím, že zachycují souvislý 3D prostor a lze tedy trojrozměrných prostorových vazeb využít i při segmentaci, ať už použitím 3D konvolucí při zpracování neuronovými sítěmi, nebo pro zobrazení různých pohledů při vytváření uživatelského vstupu pro interaktivní segmentační metody. Ukázkou CT skenu lidské nohy s vyznačenou segmentační maskou představuje obrázek 2.1.

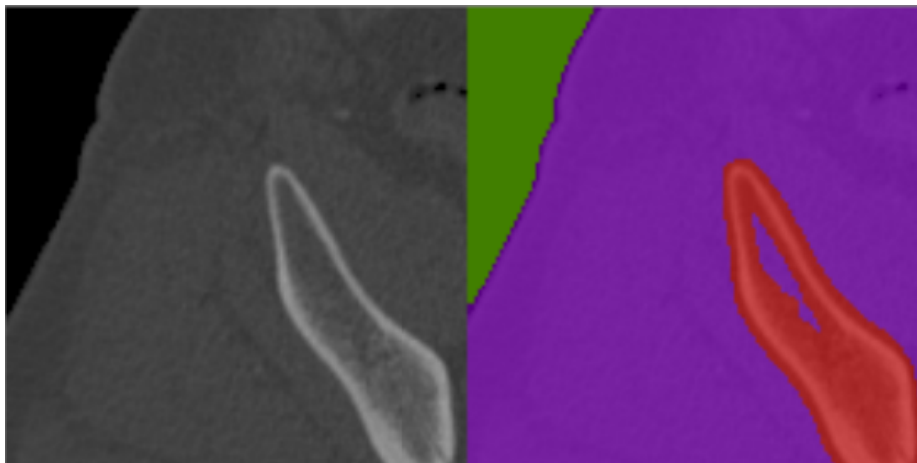
Klasické metody segmentace

V dnešní době patří hluboké učení mezi segmentační metody, které dosahují nejvyšší úspěšnosti. Existuje ale také množství metod, které stále mohou najít využití tam, kde z nějakého důvodu nelze neuronové sítě aplikovat nebo to pro jednoduchost úlohy není potřeba. Tyto metody pracují buďto se samotnými intenzitami pixelů (např. **prahování histogramu**) nebo se sekundární informací v podobě příznaků extrahovaných (například hrany oddělující různé regiony v obraze) nebo kombinují oba přístupy. Některé metody lze také použít ke zlepšení kvality výstupu segmentační neuronové sítě.

Aktivní kontury jsou metodou pracující právě na základě myšlenky, že objekty v obraze jsou odděleny hranami a mají kompaktní tvar. Algoritmus umožňuje pomocí správného nastavení několika parametrů (resp. sil tvarujících konturu) nalézt hranice objektu a vymežit jej tak vůči pozadí. S hranami pracuje také metoda **Level-set**, která umožňuje segmentaci složitějších tvarů. Detailnější popis některých metod lze najít v přehledu metod[20].

Z algoritmů založených na regionech lze uvést šíření oblasti (**region growth**). Metoda je nejprve inicializována rozdělením obrazu do menších regionů (například kvantováním intenzit nebo použitím stromové reprezentace obrazu) a následně dochází ke spojování menších oblastí ve větší podle zvolených kritérií. Autoři [35] ukazují, že variantu metody šíření oblastí společně s prahováním lze použít pro segmentaci jater v CT snímcích. Data jsou nejprve předsegmentována pomocí prahování na základě znalosti hodnot typických pro tkáň jater. Potom je na takto předpřipravená data použita metoda šíření oblastí, jejíž výsledky jsou následně ještě zlepšeny pomocí morfologických operací. Zde je použita metoda s počátečním bodem (seed), od kterého se postupně region šíří. Jako kritérium pro přidání bodu do regionu je použito srovnání intenzity hodnoceného pixelu se střední hodnotou v rámci regionu.

Shlukování lze také využít pro segmentaci. Shlukování obecně slouží k nalezení shluků dat v N-rozměrném prostoru podle **míry podobnosti** jednotlivých elementů. Je to metoda využívaná ke klasifikaci napříč různými vědními obory (psychologie, biologie, data mining...). Z hlediska klasifikace patří shlukování mezi klasifikační metody **bez učitele** (unsupervised), což znamená, že na rozdíl od učení **s učitelem** (supervised) nemáme k dispozici trénovací datovou sadu se známým přiřazením objektů do tříd [12]. Pro segmentaci je úloha formulována jako shlukování pixelů podobných vlastností resp. klasifikace pixelu do jedné z daného počtu tříd. Pixely jsou rozděleny do shluků buď podle jejich číselné reprezentace v použitém barevném modelu nebo pomocí sekundárních vlastností (příznaků, features...) odvozených zvoleným algoritmem pro extrakci příznaků. Shlukování potom probíhá v prostoru příznaků (feature space), kde každý pixel je popsán svým vektorem



Obrázek 2.2: Ukázka segmentace výřezu CT skenu lidské pánve pomocí metody aglomerativního shlukování do tří tříd na základě intenzitních hodnot.

příznaků (feature vector) udávajícím jeho pozici v prostoru příznaků. Návrh vhodných příznaků je složitá problematika sama o sobě a je specifická konkrétním datům, na kterých se segmentace provádí. Jiné příznaky budeme chtít extrahovat pro segmentaci rozsáhlých homogenních oblastí, jiné pro segmentaci oblastí se složitou texturou, u sémantické segmentace potom chceme příznaky, které budou co nejlépe vystihovat konkrétní objekt zájmu. Obecně lze říci, že je potřeba navrhnout takové příznaky, aby pixely patřící do jednoho regionu (shluku) ohodnocovaly podobně, zatímco pixely z různých regionů rozdílně [20]. Segmentaci pomocí aglomerativního shlukování můžeme vidět na obrázku 2.2. Jsou zde zřejmé nedostatky tohoto algoritmu. Vnitřní část kosti je reprezentována podobnými hodnotami jako měkká tkáň a algoritmus selhává při jejím správném přiřazení do třídy kosti, ač by to bylo pro další použití segmentace žádoucí.

K-means patří mezi nejčastěji zmiňované shlukovací metody. Metodu k-means použili autoři [23] k segmentaci obrazových dat. Data jsou nejprve předzpracována, což spočívá v úpravě kontrastu dat jejich přemapováním do zvoleného rozsahu hodnot. Druhým krokem je použití metody subtraktivního shlukování, která na základě hustoty dat dokáže odhadnout vhodné počáteční pozice centroidů, které jsou použity pro inicializaci algoritmu k-means. Pro vyhlazení výsledné segmentace je použit mediánový filtr. Autoři [14] použili metodu k-means k segmentaci jater (resp. jaterních cyst). Je provedeno srovnání výstupu klasické metody k-means s výstupy metody šíření oblastí a navrženo zlepšení metody k-means, které spočívá v následné úpravě výstupu základní metody (manuálně označeného regionu) pomocí morfologických operací. Na základě zmíněných výzkumů lze tvrdit, že metoda k-means je jednoduchá a nedosahuje dokonalých výsledků, avšak vhodnou úpravou výstupu této metody lze získat zajímavé a použitelné výsledky.

Metoda **graph cut** [7] za pomoci ruční nebo automatizované anotace objektu a pozadí hledá minimální řez grafem a rozděljuje tak obraz (v základní podobě algoritmu) na dvě třídy segmentů. Dalšími metodami používanými pro segmentaci jsou například algoritmy založené na **random forests** [34].

2.2 Hluboké učení v současné době

Moderním přístupem, který překonává většinu klasických algoritmů pro klasifikaci či segmentaci, jsou v dnešní době neuronové sítě. Pro zpracování obrazu se jedná hlavně o konvoluční neuronové sítě, které jsou popsány dále v této kapitole. Lze předpokládat, že obecný princip neuronových sítí a jejich historické počátky již není v dnešní době potřeba detailně rozepisovat a následující popis se tedy zaměřuje spíše na výšeúrovňový pohled na problematiku. (Pokud by měl čtenář zájem o detailnější popis problematiky neuronových sítí, lze jej odkázat například na knihu [6].) Jsou zde stručně popsány druhy vrstev, které se typicky používají při zpracování obrazu, a uvedeny příklady různých řešení problémů z oblasti zpracování obrazových a medicínských dat. Zajímavým směrem, který je v dnešní době také zkoumán v souvislosti s neuronovými sítěmi, je učení bez učitele, resp. možnosti využití trénování na neanotovaných datech, případně využití principu transfer learningu.

Obecná struktura neuronových sítí

(Umělou) neuronovou síť se rozumí výpočetní graf bez smyček. Jako synonymum pro pojem **neuronová síť** je v tomto kontextu používán také obecnější pojem **model**, který bude dále používán v tomto významu. Konkrétní topologie modelu je pro zjednodušení popisována pomocí **vrstev**. Skupina nejčastěji používaných vrstev se již stala standardem a existuje pro ně podpora v mnoha různých frameworkích. Podle architektury sítě (tj. konfigurace vrstev), případně podle výskytu standardizovaných bloků, lze také rozdělit síť do kategorií, například plně propojené sítě, konvoluční sítě, čistě konvoluční sítě, residuální sítě a další. Síť lze také rozdělit do kategorií podle úkolu, na který jsou natrénovány, například na síť segmentační, klasifikační nebo detekční.

Plně propojená vrstva je historicky základní jednotkou neuronových sítí. Skládá se z neuronů, přičemž operace pro vstup x o velikosti m neuronu i a výstup y_i je dán rovnicí

$$y_i = w_1x_1 + \dots + w_mx_m + b$$

kde w a b jsou parametry získané trénováním sítě. Plně propojená vrstva je tvořena typicky velkým množstvím takovýchto neuronů, jejichž počet udává rozměr výstupu vrstvy. Protože výše popsaná operace je sama o sobě lineární, jsou výstupy plně propojené vrstvy zpracovány ještě nelineární aktivační funkcí. Použití plně propojených vrstev při zpracování obrazu spočívá často v jejich umístění na konec klasifikační sítě, kde slouží pro převod příznaků extrahovaných konvoluční částí na vektor pravděpodobnosti zařazení vstupního obrazu do konkrétních tříd [37].

Konvoluční vrstva představuje základní stavební blok většiny sítí, které pracují s obrazem nebo s jinak prostorově svázanými informacemi (zvuk, video, objemová data...). Je tvořena sadou konvolučních **filtrů** (kernelů, jader). Filtr je rovněž tvořen parametry w , které jsou získané trénováním sítě. Vstupem konvoluční vrstvy je typicky obraz o šířce x , výšce y a počtu kanálů ch . Počet kanálů filtru je stejný jako počet kanálů vstupu vrstvy a výsledkem konvoluce jednoho vícekanálového filtru s celým vstupem je jeden kanál výstupu vrstvy. Rozměr výstupu v dimenzích x a y je tedy dán rozměrem vstupu a rozestupem jednotlivých přiložení filtru, tj. **krokem** (stride). Konvoluční vrstva bývá doplněna o **padding**, což je orámování vstupu nulovými hodnotami v takovém rozsahu, aby velikost původního vstupu vrstvy a výstupu vrstvy byla zachována. Výstup konvoluční vrstvy bývá označován

také jako **aktivace** nebo **aktivační mapa**. Snaha o hlubší porozumění významu těchto aktivací vedla autory [44] k analýze jednotlivých vrstev konvoluční sítě a jejich vizualizaci. Na základě vizualizací je patrné, že vrstvy blíže vstupu sítě se naučily detekovat základní příznaky, jakými jsou například hrany, či "bloby", zatímco více zanořené vrstvy detekují složitější příznaky, které odpovídají charakteristickým texturám či tvarům. Na skryté vrstvy konvoluční neuronové sítě lze tedy pohlížet jako na hierarchický extraktor příznaků, který je navíc optimalizovaný pro konkrétní data, na která byl natrénován, což je jeden z důvodů, proč neuronové sítě dosahují v oblasti zpracování obrazu mimořádných úspěchů.

Vrstva označovaná jako **maxpooling** je další častou složkou zejména sítí označovaných jako **enkodér** (bude popsáno dále). Jedná se o vrstvu sloužící k redukci rozlišení vstupu se zachováním nejsilnějších aktivací. Pro tuto vrstvu se nastavuje okno, které je přikládáno na vstup a v rámci kterého se vybírá vždy maximální hodnota. Okno je posouváno v x a y dimenzích s daným krokem, čímž je dána velikost výstupu. Typicky se používá okno 2×2 s krokem 2, čímž dochází k redukci rozměrů vstupu na polovinu. Maxpooling není jediná poolingová vrstva, která se používá, lze se setkat také například s variantou **average pooling**. Jako komplement k poolingovým vrstvám se v architekturách typu enkodér-dekodér používají dekonvoluční vrstvy, které naopak slouží jako prostředek ke zvětšení rozlišení výstupu vrstvy oproti rozlišení vstupnímu.

Výše byla zmíněna nutnost použití nelinearit, typicky za každou vrstvou neuronové sítě. Nelinearity bývají označovány jako **aktivační funkce**. V dnešní době nejčastější aktivační funkcí je funkce **rectifier linear unit**, neboli **ReLU**. Jedná se o jednoduchou operaci výběru $\max(0, x)$, kdy tedy dochází k potlačení záporných hodnot a předání kladných hodnot další vrstvě sítě. Funkce je vyhodnocena pro každý element vstupu vrstvy a pro každý produkuje samostatný výstup. Je možné použít i jiné druhy aktivačních funkcí. Někdy k tomu přímo vede požadavek na omezení výstupu sítě do určitého rozsahu. Za tímto účelem se používá například funkce **softmax** a to pro svou vlastnost normalizace vstupních hodnot do rozsahu $(0, 1)$. Výstupní hodnoty softmax vrstvy lze tedy interpretovat jako pravděpodobnosti, čehož se využívá například u klasifikačních sítí, kdy tak dochází k odhadu pravděpodobnosti náležitosti klasifikovaného vzorku do každé třídy.

Proces učení neuronových sítí

Proces učení neuronové sítě je iterativní optimalizační proces. Je vybrána **chybová funkce** (loss function, cost function), která provádí srovnání skutečného výstupu sítě v aktuální iteraci a výstupu požadovaného. Volba konkrétní chybové funkce se odvíjí od řešené úlohy. Pro úlohu segmentace lze použít například chybové funkce Cross Entropy nebo Dice loss. Výstup chybové funkce (tedy chyba celé sítě v aktuální iteraci) je zpětně zpropagována celou sítí a na základě gradientů jsou upraveny hodnoty daných parametrů.

Historicky bylo běžné použití algoritmu **stochastic gradient descend**, který funguje velice jednoduše, avšak jehož úspěch či neúspěch je silně ovlivněn správným nastavením **learning rate**. Proto se od tohoto algoritmu v dnešní době spíše upouští a rozšířenějšími jsou dnes různé varianty adaptivních optimalizačních algoritmů. Tyto algoritmy jsou méně citlivé na nastavení learning rate a interně jeho velikost automaticky přizpůsobují na základě heuristik. **Adam** [15] je jedním z těchto adaptivních algoritmů.

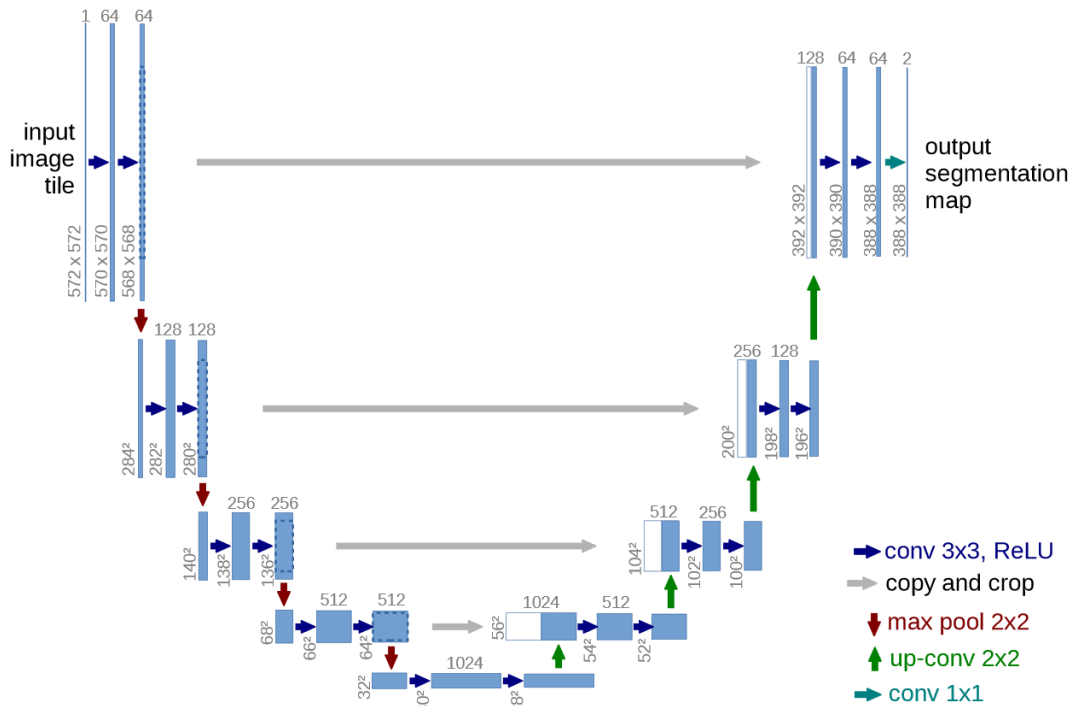
Architektury pro segmentaci obrazových dat

V průběhu posledních let bylo ukázáno že metody hlubokého učení jsou aplikovatelné na široké spektrum úloh zpracování obrazu. Běžné je použití konvolučních neuronových sítí pro úlohy detekce, klasifikace, segmentace, sledování objektu, ale také zlepšení kvality obrazových dat, redukce šumu, komprese, případně generování syntetických obrazových dat. Mezi významné zástupce úspěšných architektur neuronových sítí pro zpracování obrazu patří například AlexNet[16], VGG[37], GoogLeNet[38], ResNet [13] nebo FCN [36].

Konvoluční neuronové sítě jsou silným nástrojem pro extrakci vysokoúrovňových příznaků. Jistým způsobem lze i segmentační masku považovat za vysokoúrovňový příznak a na základě této myšlenky skutečně lze použít jednoduchou síť skládající se pouze z několika konvolučních vrstev pro extrakci segmentační masky. Tento přístup ale není příliš efektivní a kvalita segmentace není tak vysoká, jak by mohla být při použití složitější architektury. Postupem času došlo k vyčlenění typu architektury označované jako **enkodér-dekodér**, případně specifitější varianta **autoenkodér** [3]. Jak už název napovídá, tato architektura se skládá ze dvou částí. První část, **enkodér**, slouží k extrakci charakteristické informace ze vstupu. Dochází přitom ke snižování rozlišení v osách x a y vstupu a současněmu navyšování počtu filtrů konvolučních vrstev. Toto má za následek potlačení nedůležité informace a posílení rysů charakteristických pro konkrétní vstupní data. Druhou částí architektury je **dekodér**, který má za úkol navyšování rozlišení rozšíření charakteristické informace zpět do (prostorového) kontextu vstupu. Navyšování rozlišení přitom může být provedeno různými způsoby, například obyčejnou interpolací nebo použitím výše zmíněných dekonvolučních vrstev. Způsob dekodování extrahované informace odpovídá účelu použití sítě. U autoenkodérů je úkolem sítě transformovat vstup na výstup tak, aby přitom došlo k co nejmenší degradaci kvality dat [5].

Mezi prvními, kdo použili dekonvoluci k vytvoření segmentační masky v rámci plně konvoluční sítě, jsou autoři Long et al. [36], kteří tímto rozšířili možnosti použití úspěšných klasifikačních sítí taky pro úlohu segmentace. Upravenou architekturu typu enkodér-dekodér použili autoři architektury SegNet [4]. Enkodér se v tomto případě skládá z třinácti konvolučních vrstev, kterým zrcadlově odpovídá třináct vrstev dekodéru. Architektura je navíc vylepšena přidáním dodatečných spojení mezi enkodérem a dekodérem, která slouží k lepšímu umístění extrahované informace v dekodérové části. Přesněji, v enkodérové části jsou uloženy indexy maximálních hodnot nalezených v max pooling vrstvách a tyto indexy jsou předány odpovídajícím vrstvám dekodéru, kde je provedeno nadzorkování právě podle těchto indexů a nadzorkované aktivační mapy jsou zpracovány konvolučními vrstvami dekodérové části. Podle autorů je tento způsob předání informace o prostorovém kontextu významným rozšířením architektury typu enkodér-dekodér, přičemž předání pouhých indexů vede k úsporné implementaci. Podobný přístup použili také autoři architektury DeconvNet [24].

Další modifikaci této architektury použili autoři sítě U-net [29] navržené přímo pro segmentaci medicínských dat. Podobně jako SegNet používá U-net alternativní cesty mezi enkodérovou a dekodérovou částí sítě pro předání informace o prostorovém kontextu. V případě U-net jsou ale předávány celé aktivační mapy enkodérových konvolucí předcházejících poolingovou operaci. Tyto aktivační mapy jsou konkatenovány s odpovídajícími výstupy dekonvolucí dekodérové části a tenzor vzniklý touto konkatenací je předán jako vstup další



Obrázek 2.3: Schéma originální architektury sítě U-net. Převzato z původní publikace [29].

konvoluční vrstvě. Dekonvoluce v tomto případě sama zajišťuje nadvzorkování. Schéma U-net architektury ukazuje obrázek 2.3 převzatý z originální publikace. Novější variantou této architektury je její rozšíření v podobě 3D konvolucí [46] pro segmentaci objemových dat. Lepší využití prostorové informace přidáním třetí dimenze vstupu je důležitým zlepšením také pro segmentaci CT dat. V tomto případě autoři také adresují problém časové náročnosti vytvoření pixelwise anotací pro objemová data, a to pomocí navrženého postupu trénování z řídkce anotovaného datasetu. V daném objemu je vždy anotováno jen několik řezů. Vstupem sítě je celý objem, na výpočtu chybové funkce se ale podílejí jen ty řezy, které byly anotovány. Přesněji řečeno je použita váhovaná chybová funkce, která zajišťuje jednak rovnoměrný podíl pozadí a segmentovaného objektu na chybě i při jejich nevyváženém zastoupení v datasetu, a zároveň umožňuje nastavení váhy na nulovou hodnotu pro neanotované řezy.

2.3 Problém s nedostatkem dat

Datasety pro trénování neuronových sítí obsahují typicky tisíce trénovacích vzorků. Jak už bylo zmíněno, nejen v oblasti medicínských dat může být vytvoření dostatečně velkého datasetu problematické či časově náročné, například kvůli nutnosti vytvoření anotací experty nebo kvůli zdravotním rizikům která se pojí s pořizováním CT snímků. Kromě toho může být kvůli odlišnému rozložení anatomie problém s přenositelností trénovacích datasetů mezi geografickými regiony. Dalším omezením může být velmi malý výskyt segmentovaného objektu v populaci v případě segmentace patologických útvarů. Použití malého datasetu je problematické například proto, že snadněji dochází k přetrénování modelu, který potom

ztrácí schopnost dobře generalizovat na jiná než trénovací data, tedy se stává méně robustním. Kromě jiných regularizačních technik lze k řešení tohoto problému použít například augmentovaná data, vytvořená z dostupného datasetu.

Augmentace trénovacích dat

Augmentace dat je technika používaná jako metoda regularizace při trénování neuronových sítí a slouží tedy jako prostředek pro zlepšení schopnosti modelů generalizovat. Review na téma augmentace dat pro segmentaci nádorů mozku vytvořili autoři Nalepa et al. [22]. Popsané metody byly vybrány z dvaceti publikací BraTS 2018 Challenge a otestovány na datasetu BraTS za použití trénovacích dat 570 pacientů. Augmentace jsou rozděleny do několika kategorií. První zmíněnou kategorií jsou afinní transformace, které jsou snadno implementovatelné a díky tomu i často používané. Zavedením afinních transformací lze znatelně navýšit počet trénovacích vzorků a vytvořit tak robustnější model. Některé afinní transformace ale mohou vést na anatomicky nesprávná data. Použití transformace posunutí může být výhodné jen u prostorově závislých modelů a pro plně konvoluční sítě tedy nepřináší žádné zlepšení. Geometrickou transformaci v podobě elastických deformací používají například autoři původní publikace o segmentaci se sítí U-net [29].

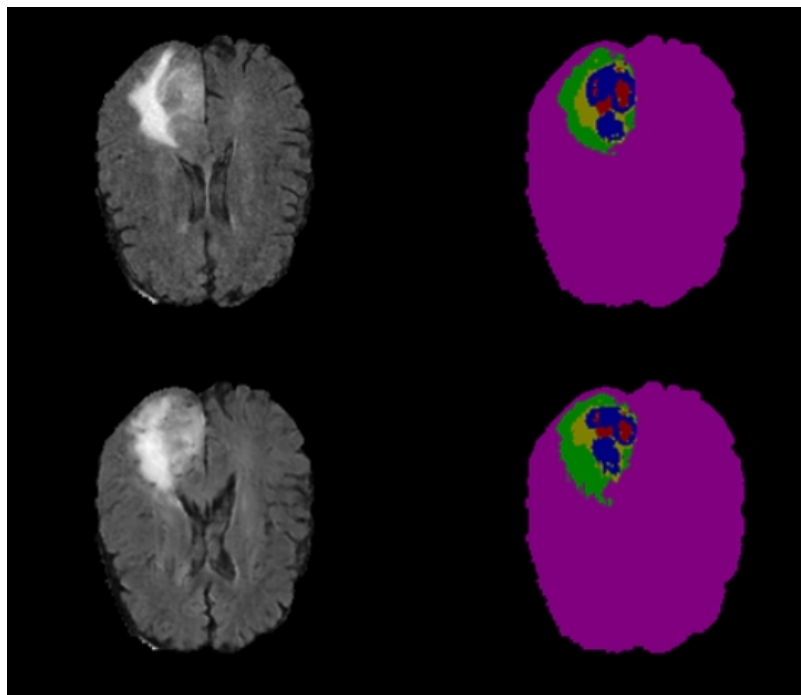
Kromě geometrických transformací se používají také transformace na úrovni pixelů, tedy změny intenzitních hodnot. Použití takovýchto augmentací je relevantní i pro medicínská data, protože intenzity představující různé tkáně se mohou lišit ať už díky variabilitě pacientů nebo díky rozdílnosti snímacích zařízení. Kromě nelineárních změn intenzity sem lze zařadit také rozmazání nebo přidání šumu. Detailnější použití výše zmíněných augmentačních technik je popsáno v práci Nalepy et al. [22].

Generativní adversariální sítě a syntetická data

Dalším přístupem pro navýšení velikosti datasetu může být použití generativních adversariálních sítí (GANs) pro vytvoření syntetických trénovacích dat. Při použití GAN je zaveden **generátor** a **diskriminátor**, přičemž generátor má za úkol generovat data, která budou pro diskriminátor nerozlišitelná od dat skutečných. Synteticky generovaná data mají uplatnění například při segmentaci nádorů, které mohou být vytvořeny na základě modelů předpovídajících nádorový růst.

Mok et al. [21] používají GANs jako metodu pro augmentaci dat pro navýšení objemu trénovacího datasetu. Elastická deformace je místo na originální MRI skeny aplikována na label mapu, tj. snímek s anotacemi typů tkání mozku. Z takto upravených label map jsou potom generována data syntetická. Obrázek 2.4 ukazuje příklad takto vygenerovaného snímku. Použití augmentovaných trénovacích dat přináší zlepšení v řádu jednotek procent Dice skóre při trénování na datasetu BRATS15.

CycleGANs pro augmentaci nekontrastních CT dat používají autoři Sandfort et al. [33]. Veřejné CT datasety pro segmentaci břišních orgánů obsahují zpravidla kontrastní snímky, avšak v praxi nelze vždy kontrastní látku při vytváření skenu použít. Autoři pomocí GAN vytvářejí syntetická nekontrastní data z dat kontrastních, pro lepší natrénování segmentace skutečných nekontrastních dat. GANs lze také použít jiným způsobem, přímo jako součást

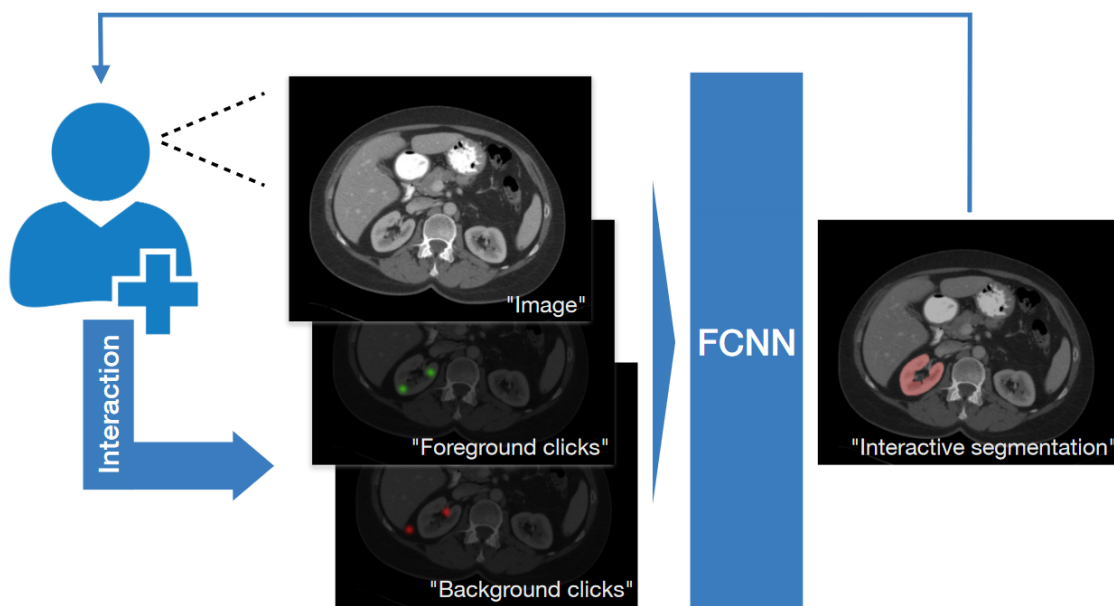


Obrázek 2.4: Využití GANs pro augmentaci datasetu nádorových onemocnění. V prvním řádku je ukázka originálního snímku mozku a groundtruth anotace rozlišující tkáň mozku a různé druhy tkání nádoru. Druhý řádek ukazuje syntetická data vygenerovaná z groundtruth snímku, který byl nejprve deformován. Převzato z původní publikace [21] a upraveno.

řešení segmentační úlohy. Autoři Rezaei et al. [28] používají conditional GAN (cGAN) pro segmentaci nádorů v multimodálních snímcích mozku.

Možností, která částečně souvisí se zapojením GANs, je i generování syntetického datasetu. Kromě GANs lze k tomuto použít metodu randomizace domény. Parametry pro vytváření syntetických dat (nasvícení, poloha, textura...) jsou generovány náhodně a navíc nerealistickým způsobem, aby byl model nucen naučit se konkrétní rysy analyzovaných dat [39]. Tento přístup je relativně snadno uplatnitelný v jiných než medicínských odvětvích. Například Ward et al. [41] natrénovali model pro segmentaci rostlinných listů ve fotografických snímcích s využitím smíšeného datasetu, který obsahoval jak reálná, tak syntetická data. Autoři využívají přístupu randomizace domény a vytvářejí 3D modely listů, které jsou náhodně deformovány a texturovány. Render takto vytvořeného 3D modelu rostliny je umístěn do fotosnímku květináče a takto vytvořená data jsou přidána do trénovací množiny.

Co se týče medicínských dat, je použití syntetických datasetů omezené vysokou složitostí dat. V některých případech lze uplatnit reverzní adaptaci domény [18], kde jsou cílová data jsou s využitím adversiální sítě upravována tak, aby byla podobná datům syntetickým. Model, vykonávající cílovou úlohu, lze potom trénovat na datech syntetických. Pokud ale budeme uvažovat data složitostí odpovídající 3D CT snímkům, je použití výše zmíněných přístupů značně obtížné.



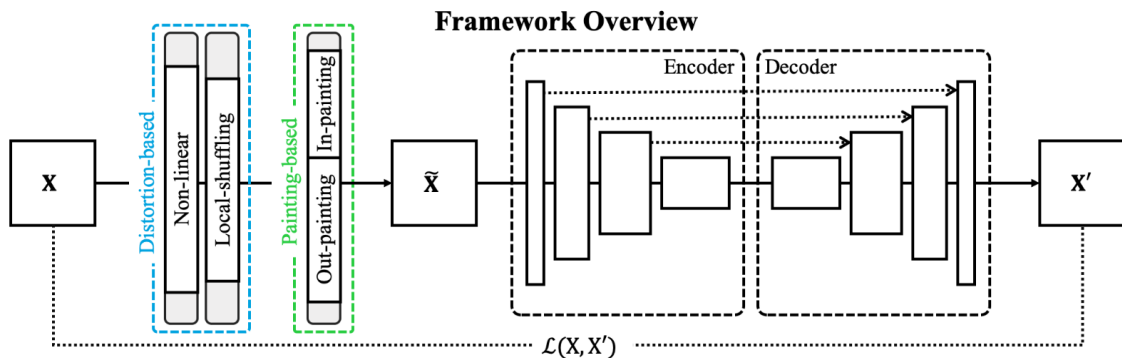
Obrázek 2.5: Schéma interaktivní segmentace s využitím plně konvoluční neuronové sítě. Převzato z původní publikace [32] a upraveno.

Využití slabých anotací a uživatelské interakce

Další možností zlepšení výsledků trénování (nejen) s malými datasety je přidání rozšiřující informace na vstup sítě. Takovouto informací může být například uživatelská interakce. Přidání uživatelské interakce může pomoci jak lepší lokalizaci segmentovaného objektu, tak vyrovnání drobných rozdílů v datové doméně, které by mohly být pro automatický model problematické. Protože se většina experimentů, provedených v rámci této práce, zaměřuje právě na interaktivní segmentaci, jsou zde zmíněny některé práce, které se touto, a příbuznou, problematikou zabývají.

Roth et al. [30] používají slabé anotace v podobě vyznačení extrémních bodů v kombinaci s algoritmem random walker [11] pro natrénování segmentačního modelu. Motivací je v tomto případě právě vytvoření anotačního nástroje pro usnadnění vytváření medicínských segmentačních datasetů. Jiným způsobem používají extrémní body pro segmentaci také autoři Maninis et al. [19]. Pokusy o trénování čistě pomocí slabých anotací v podobě črtů provedli autoři Can et al. [9].

Pace et al. [25] popisují iterativní segmentaci s omezenými datasety a demonstrují tento přístup na datasetu s dvaceti snímky magnetické rezonance. Segmentace lidského srdce je v tomto případě generována rekurentním modelem iterativně z výchozího bodu stanoveného uživatelem. Bredell et al. [8] popisují interaktivní segmentační síť se zpětnou vazbou. Nejprve je pomocí jiného segmentačního algoritmu vytvořena predikce segmentace, která je společně s uživatelským vstupem předána konvoluční síti. Výsledek segmentace může být upraven iterativně, přičemž síti je předložena jak uživatelská interakce, tak její vlastní výstup z předchozí iterace. Další práce rovněž používají uživatelský vstup ve formě črtů [2, 40], kliknutí [17, 42] nebo jejich kombinaci [1].



Obrázek 2.6: Schéma frameworku pro unsupervised trénování restaurace CT dat. Převzato z původní publikace [45].

Experimenty, které byly provedeny v rámci této práce čerpají nejvíce z originální publikace autorů Sakinis et al. [32]. Autoři navrhli metodu využití map uživatelských vstupů v podobě kliknutí do vstupního snímku. Vytvořeny jsou dvě mapy, a to mapa pro označení objektu a obdobná mapa pro pozadí. Tyto mapy jsou přidány na vstup modelu jako další vstupní kanály. Model je trénován na datasetu s více třídami, přičemž pro jeden vstup je vždy za segmentovaný objekt zvolena pouze jedna vybraná třída. Autoři tvrdí, že takovýto způsob trénování jednak zlepšuje výsledky segmentace, ale umožňuje i do jisté míry generalizovat na anatomicky jiná data, než na kterých byl model trénován, například nádorové oblasti. Schéma metody ukazuje obrázek 2.5.

Využití transfer learningu v hlubokém učení

Protože metody hlubokého učení dosahují tím lepších výsledků, čím větší je množství (anotovaných) trénovacích dat, existuje v poslední době také snaha o nalezení způsobu znovupoužití datasetů i pro zdánlivě nesouvisející úlohy. Lze předpokládat, že v některých případech lze využít znalosti získané trénováním na datasetu pokrývající jednu oblast pro zlepšení výsledků trénování v oblasti pokryté jiným datasetem. Je tedy požadavek na **přenos znalosti** (transfer of knowledge) a obor postupů umožňujících splnění požadavku na přenos znalosti lze označit souhrnně pojmem **transfer learning**.

Motivace pro výzkum oblasti transfer learningu vychází ze způsobu, jakým se učí lidský mozek. Přesněji z předpokladu, že lidský mozek dokáže využít podobnosti mezi známými a neznámými daty pro zařazení dat neznámých na základě podobnosti rysů s daty známými, resp. zkušenosti získané řešením známého úkolu pro řešení úkolu nového. Z pohledu strojového učení lze pro lepší vymezení problematiky definovat několik následujících pojmů (podle [26]). **Doména** \mathcal{D} je dvojice $(\mathcal{X}, P(X))$, kde $X = x_1, \dots, x_n \in \mathcal{X}$, přičemž \mathcal{X} je prostor všech vektorů příznaků dané domény a $P(X)$ je marginální rozložení pravděpodobnosti. Různé domény se mohou lišit v prostorech příznaků a/nebo v marginálním rozložení pravděpodobnosti. **Úkol** \mathcal{T} je dvojice $(\mathcal{Y}, f(\cdot))$, kde \mathcal{Y} je prostor anotací a $f(\cdot)$ je funkce mapující $x \in X$ na $y \in \mathcal{Y}$, která je získána trénováním. **Transfer learning** lze potom chápat jako využití zdrojové domény D_S a zdrojového úkolu T_S pro zlepšení procesu učení prediktivní funkce $f_T(\cdot)$ v cílovém úkolu T_T , přičemž $D_S \neq D_T$ a $T_S \neq T_T$. Je vhodné zmínit, že klasický přístup (hlubokého) učení, kdy se využívá trénovacích dat pro získání statistického

modelu odpovídajícího datům testovacím předpokládá, že $D_S = D_T$ a $T_S = T_T$.

Techniky transfer learningu mohou být rozděleny do kategorií podle toho jaká znalost je přenášena resp. podle toho v čem se odlišují zdrojová a cílová doména či úkol. Detaily lze nalézt v [26]. V oblasti hlubokého učení se poměrně běžně využívá transfer learningu v podobě inicializace vah trénované sítě pomocí vah extrahovaných ze sítě natrénované na souvisejícím úkolu. Takto inicializovaná síť může být kompletně přizpůsobena nové úloze, nebo může dojít ke "zmrazení" některých (zejména) počátečních vrstev a modifikaci pouze zanořenějších vrstev, které extrahují informace charakterističtější pro konkrétní $P(X)$ zdrojové domény. Zpravidla se přitom dosahuje vyšší úspěšnosti v řešení cílového úkolu než při použití náhodné inicializace. Průzkum v oblasti přenositelnosti natrénovaných modelů provedli Yosinski et al. [43]. V rámci výzkumu se autoři zabývají mimo jiné otázkou kde přesně a jak prudce dochází v rámci sítě k "překlopení" mezi obecnými a úkolu/doméně specifickými příznaky. Zajímavé je také zjištění vazeb mezi méně a více zanořenými vrstvami sítě. Pokud je použito větší množství (zde 4, 5, 6) zmražených vrstev, je obtížnější natrénovat hlubší vrstvy pro řešení cílového úkolu. Tyto vzájemné vazby jsou charakteristické pro vrstvy nacházející se ve středu sítě a u počátečních nebo koncových vrstev nejsou tolik patrné. Výsledky také potvrzují výše zmiňované tvrzení o specifitě vrstev blízkých konci sítě.

Využití metod transfer learningu pro unsupervised trénování obecných modelů pro analýzu CT dat navrhli autoři Zhou et al. Ve své publikaci [45] popisují proces trénování několika **autodidaktických** modelů, které lze následně využít jako inicializační pro řešení konkrétnějších úloh, například segmentaci či klasifikaci. Tyto modely, nazvané Models Genesis, jsou trénovány pomocí zdrojových úloh, které jsou navrženy tak, aby vstupní data pro trénování sítě bylo možné generovat automaticky a jako anotace sloužila data původní. Přesněji se jedná o úkoly z oblasti restaurace obrazových dat a to **nelineární transformace intenzity** pro zachycení tvarů a rozložení intenzity konkrétních orgánů lidského těla, **lokální přemístění jednotlivých pixelů** pro naučení lokálních hranic a textury a **vložení náhodných dlaždic** (inpainting a outpainting) do obrazu pro natrénování globální geometrie a prostorové umístění orgánů. Schéma navrženého frameworku je na obrázku 2.6. Podle autorů modely inicializované pomocí generických modelů dosahují výrazně vyšší úspěšnosti v cílových úlohách než modely vycházející z náhodných inicializací.

Využití restaurace obrazu pro natrénování modelu schopného lépe extrahovat příznaky se věnují také autoři Pathak et al. [27]. Podobně jako autoři Models Genesis využívají úlohy **inpainting** pro zachycení globálního kontextu a takto natrénovaný model označují jako **context encoder**. Navíc zde autoři ještě experimentují s použitím adversariální chybové funkce, která se používá u generativních sítí a nabízí možnost detailnějšího dokreslení poškozené oblasti než použití jednoduchých chybových funkcí typu $L1$ nebo $L2$.

2.4 Segmentace pomocí shlukování a CNN

V rámci semestrálního projektu předcházejícího této práci byla blíže zkoumána možnost využití u-net modelu trénovaného na restauraci dat (obdobně jako Models Genesis [45]), coby extraktoru příznaků pro segmentaci pomocí shlukování. Protože na neuronové sítě lze pohlížet také jako na hierarchické extraktory příznaků, lze předpokládat, že model trénovaný na konkrétní obrazová data se naučí extrahovat relevantní příznaky, které lze potom využít právě pro segmentaci. Byly provedeny experimenty s modelem trénovaných na restauraci

dat tak, jak je popsáno výše. Bylo ale zjištěno, že u takto trénovaného modelu dochází k mizení aktivací v dekodérové části, pokud jsou na vstupu nepoškozená data a tedy není možné využít vysokoúrovňové mapy příznaků. Při použití enkodér-dekodér architektury bez konkatenáčních spojů byly výsledky o něco málo zajímavější, ale pořád nepřesvědčivé. Z tohoto důvodu bylo od použití shlukování v rámci této práce úplně upuštěno.

Kapitola 3

Vlastní řešení segmentace CT dat

Předchozí kapitola shrnuje současný stav problematiky segmentace obrazových a objemových dat, se zaměřením na segmentaci dat z oblasti medicíny. V této kapitole jsou popsány metody, které byly na základě poznatků z předchozí kapitoly vybrány pro řešení segmentace kostí v CT skenech lidských pacientů.

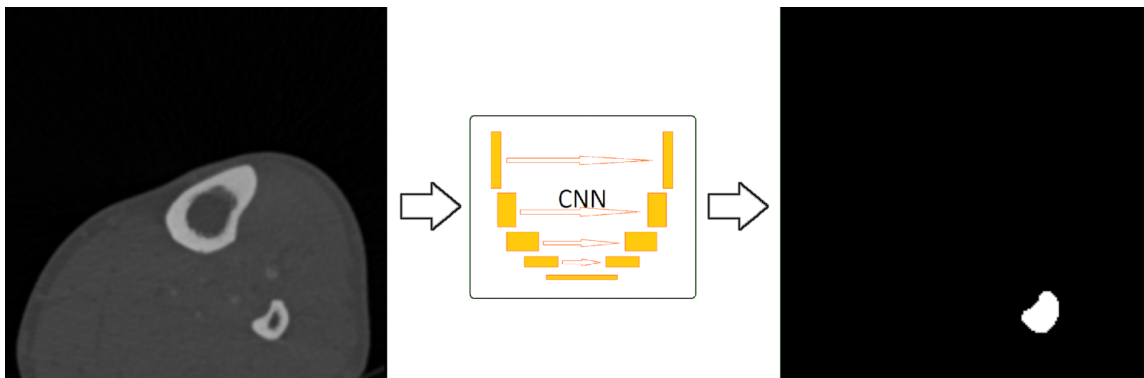
Jak už bylo řečeno, pro natrénování segmentační neuronové sítě, která dokáže segmentovat s vysokou přesností a dobře generalizovat na nová data, je zpravidla nutné použití velkého množství anotovaných trénovacích dat. Protože ale vytvoření velkého trénovacího datasetu pro segmentaci medicínských dat je časově náročné, byly vybrány některé metody pro zlepšení kvality segmentace i při použití malého trénovacího datasetu. Použity jsou metody transfer learningu, konkrétně předtrénování na zástupné úloze restaurace obrazu případně předtrénování na segmentaci anatomicky jiného druhu dat. V rámci řešení byl v kontextu omezeného množství trénovacích dat otestován také interaktivní přístup jako nástroj pro vytvoření větších trénovacích datasetů s menší časovou investicí. Je zde zaveden interaktivní segmentační model, řízený uživatelskými anotacemi v podobě kliknutí.

Jako baseline metoda, tedy metoda, se kterou jsou ostatní zkoumané přístupy v této práci porovnány, byl zvolen standardní model architektury U-net trénovaný na úlohu automatické segmentace vybrané kosti. Schéma metody ukazuje obrázek 3.1. Srovnání je provedeno pomocí vytvoření a optimalizace instancí modelu s různě velkými trénovacími podmnožinami datasetu. Pro každou metodu byla natrénována skupina instancí modelů s různě velkými trénovacími podmnožinami. Nejmenší trénovací podmnožina obsahuje data pouze jednoho pacienta (reps. pouze jeden 3D CT sken).

3.1 Společná konfigurace metod a trénovacího procesu

Architektura všech použitých modelů je obdobná jako u originálního 2D U-net modelu [29] co se týče počtu konvolučních filtrů v každé konvoluční vrstvě. Každou konvoluční vrstvu následuje ReLU nelinearita, případně sigmoida v případě poslední konvoluční vrstvy. Počet bloků sestávajících z konvoluční vrstvy a nelinearity, je rovněž zachován, včetně konkateinačních propojí na různých úrovních sítě.

Vstupem sítě jsou, v případě automatických modelů, 2D výřezy o rozměru 96×96 , které jsou samplovány z originálních CT dat. V případě interaktivních modelů jsou přidány další kanály představující uživatelský vstup, případně zpětnou vazbu z předchozí iterace.



Obrázek 3.1: Segmentace kosti v řezu CT dat pomocí automatického modelu. Tento postup byl vybrán jako baseline metoda pro tuto práci.

Interaktivní vstupy budou blíže popsány v následujících částech práce.

Vstupem chybové funkce je dvojice tvořená výstupem sítě v podobě mapy představující predikci objektu a odpovídajícím výřezem z groundtruth segmentace. Jako chybová funkce byla použita Cross Entropy.

Za optimalizační algoritmus byl zvolen Adam [15] s learning rate 10^{-5} . Vzorky v minibatchi jsou vybírány náhodným generátorem bez opakování. Řez je znovu vybrán do minibatche až po vyčerpání kapacity celé trénovací množiny. Velikost minibatche byla nastavena na 32 řezů pro všechny experimenty. Vzhledem k trénování s různě velkými trénovacími množinami jsme upustili od konceptu trénování po epochách a veškerá srovnání konvergence modelů jsou vyhodnocena vzhledem k počtu trénovacích iterací. Modely byly trénovány po dobu 65 000 iterací, pokud není dále uvedeno jinak.

Ačkoliv byla výše zmíněna možnost použití regularizací pro zlepšení výsledků segmentace a omezení přetrénování, rozhodli jsme se je v experimentech nepoužívat. Je žádoucí, aby na vytvořeném benchmarku byl dobře pozorovatelný přínos jednotlivých testovaných metod. Vzhledem k poměrně vysoké přesnosti segmentace dlouhých kostí základním U-net modelem by zavedení regularizací mohlo vést na hůře rozlišitelné výsledky. Okrajově byly také provedeny experimenty s Dice loss, která vede k lepším výsledkům. Protože celkově lepší hodnoty vedou také na méně patrné rozdíly mezi výsledky zkoumaných metod, byly experimenty vyhodnoceny pouze s použitím Cross Entropy.

Zde popsané základní nastavení je společné pro všechny modely a případné změny tohoto nastavení jsou konkrétně uvedeny dále.

3.2 Vybraná rozšíření metod při použití malých datasetů

Znatelné zlepšení kvality segmentace už při použití jediného 3D skenu může být v praxi důležitým krokem při návrhu metody pro rychlejší vytvoření nového trénovacího datasetu, který bude sloužit pro natrénování přesnějšího automatického segmentačního modelu. Pro zlepšení kvality segmentace, zejména při použití velmi malého počtu trénovacích vzorků, byly vybrány dva přístupy - interaktivní segmentace a transfer learning. V následujícím

Tabulka 3.1: Souhrnná tabulka označení metod použitých v této práci.

Označení	Popis
Seg	Automatický model s náhodnou inicializací. Vstup je 2D výřez CT skenu.
SegPTaskfhit	Automatický model předtrénovaný na restauraci. Vstup je 2D výřez CT skenu.
SegI	Interaktivní model s náhodou inicializací. Vstup je 2D výřez CT skenu a dva kanály s uživatelskou interakcí.
SegIPTaskshift	Interaktivní model předtrénovaný na restauraci. Vstup je 2D výřez CT skenu a dva kanály s uživatelskou interakcí.
SegIPDatashift	Interaktivní model předtrénovaný na odlišných datech. Vstup je 2D výřez CT skenu a dva kanály s uživatelskou interakcí.
SegIPlus	Interaktivní model s náhodou inicializací. Vstup je 2D výřez CT skenu, dva kanály s uživatelskou interakcí a vstupní predikce segmentace.

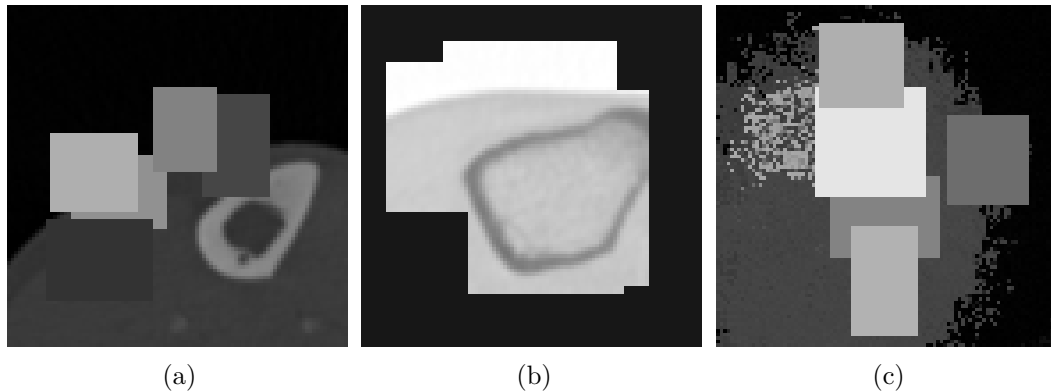
textu jsou tyto přístupy vysvětleny a výčet konkrétních metod využívajících tyto přístupy či jejich kombinace, je uveden v tabulce 3.1.

Unsupervised předtrénování na restauraci

Obdobně jako v originální publikaci [45], bylo i zde využito možnosti natrénovat model na úloze restaurace obrazu. V takovém případě je trénovací dvojice tvořena vstupem sítě v podobě poškozených dat a groundtruth v podobě dat původních. Distorze může být prováděna automatizovaně a úloha se tak stává plně unsupervised. Model natrénovaný na restauraci obrazu sice nelze přímo použít pro segmentaci, lze nicméně předpokládat, že model takto získává schopnost extrahovat příznaky z použitých dat a že tato schopnost do jisté míry přetrvává i po trénování na segmentační úlohu, kde vede ke zlepšení výsledné kvality oproti modelu, který byl inicializován náhodně.

Metody distorze jsou převzaty z výzkumu [45] a použity obdobným způsobem jako v původní práci. Konkrétně to jsou metody inpainting, outpainting, nelineární změna intenzity a lokální shuffling pixelů. Ukázkové výřezy CT dat, na kterých byla provedena distorze, jsou na obrázku 3.2. Konkrétní uplatnění distorzních metod odpovídá původnímu frameworku (obrázek 2.6) a je následující:

Nejprve je vybráno, zda bude použit originální výřez nebo výřez ve kterém byl proveden shuffling pixelů. Vzhledem k časové náročnosti operace shuffling byl předgenerován sekundární dataset, kde byla tato operace již provedena a při samotném trénování se pouze přepíná mezi originálním a sekundárním datasetem na základně náhodné volby. Pravděpodobnost výběru primárního nebo sekundárního datasetu je 50 % pro každý. Je zřejmé, že tak dochází k částečnému snížení variability trénovacích dat, tato změna ale není pro výsledky práce zásadní. I kdyby toto vedlo k částečnému omezení schopnosti restaurovat data, konkrétní vliv kvality restaurace na výsledný segmentační model není již součástí zde provedených experimentů.



Obrázek 3.2: Tři různé trénovací výřezy pro úlohu restaurace obrazu. Inpainting (3.2a), outpainting s nelineární transformací intenzity (3.2b) a inpainting s lokálním shufflingem pixelů (3.2c).

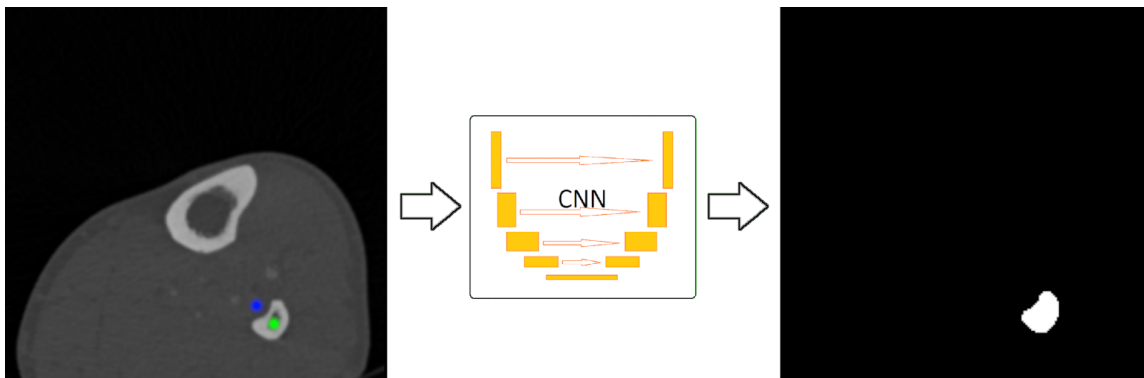
Následuje operace lineární transformace intenzit, která je uplatněna s 50% pravděpodobností. Jako poslední změna dat je vložení bloků náhodné intenzity. To může být provedeno buďto v inpainting variantě, kdy jsou bloky rozmístěny kolem středu výřezu, nebo outpainting, kdy je oblast kolem středu ponechána beze změny a okraj výřezu je orámován vloženými bloky. Výběr mezi variantou (inpainting/outpainting) je rovněž náhodný s pravděpodobností 50 % pro každou variantu. Maximální počet vložených bloků při inpaintingu je 5.

Jako chybová funkce byla použita funkce L1. Je vhodné upozornit, že použití této funkce vede k trénování modelu tak, jako by existovala pouze jedna správná podoba rekonstruovaných dat, což ale plně neodpovídá realitě. Čím větší je inpaintingem zakrytá oblast, tím je anatomicky správné řešení variabilnější. Stejně tak denzita tkání se může lišit mezi různými pacienty a nelze tak předpokládat jedno správné řešení restaurace intenzitních změn. Pokud by cílovou úlohou skutečně byla restaurace CT dat, bylo by vhodné zvážit použití jiné chybové funkce, která bude tento problém zohledňovat, například adversariální chybovou funkci [27]. Pro tuto práci je funkce L1 dostačující.

Během trénovacího procesu restaurační sítě jsou ukládány checkpointy modelů s nejlepší hodnotou chybové funkce. Checkpoint následně použitý jako inicializační pro segmentační sítě byl trénován po 43 000 iterací s hodnotou chybové funkce $L1 = 0,0177$. Maximální počet iterací během jedné trénovací úlohy je omezen časovým limitem pro běh úlohy na výpočetním gridu. Tento limit je 24 hodin, což odpovídá přibližně 60 000 iteracím trénování restaurační sítě.

Označení objektu a pozadí uživatelem

Druhým přístupem pro trénování úlohy segmentace s malým datasetem je zavedení uživatelské interakce. Obdobně jako v práci autorů Sakinis et al. [32] jsou přidány dva typy interakce, kterou může uživatel provést, a to označení objektu a označení pozadí kliknutím (obrázek 3.3).



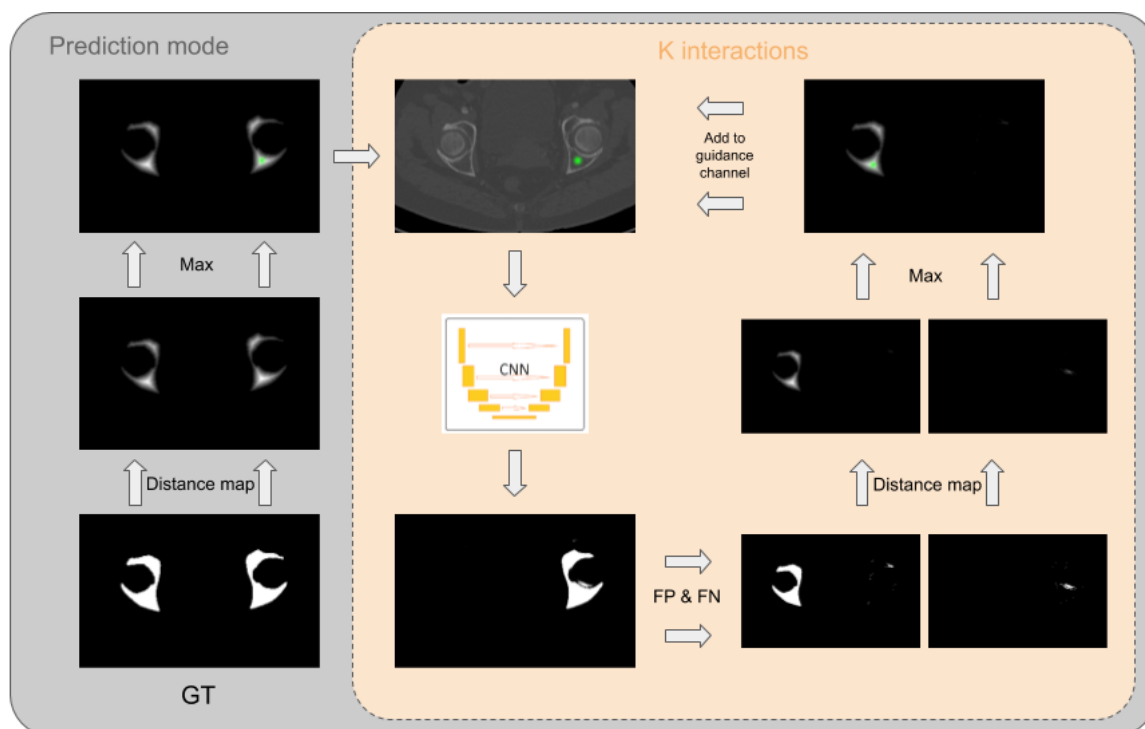
Obrázek 3.3: Interaktivní model. Na vstup jsou přidány dva kanály s uživatelskou interakcí. Zeleně je vyznačen kanál interakce pro objekt, modře pro pozadí.

Vstupem automatické segmentační sítě je pouze 2D výřez originálních dat. Pro základní interaktivní síť jsou přidány na vstup další dva kanály, které odpovídají mapám kliknutí. Jedna mapa odpovídá kliknutím označující pozadí, druhá mapa reprezentuje označení objektu. Kromě této úpravy vstupu je architektura interaktivního modelu stejná, jako architektura modelu automatického.

Protože není reálné, aby během trénování označoval objekt či pozadí skutečný uživatel, byl použit obdobný způsob automatického generování interakcí jako v původní práci. Každá iterace trénování interaktivní sítě se skládá ze dvou částí. První částí je část **prediktivní**, kdy je automaticky vygenerováno K umělých uživatelských interakcí, aniž by docházelo k updatu vah modelu. Z těchto K iterací je potom vybráno $t \in \langle 1, K \rangle$ interakcí, které jsou použity pro vstup sítě v části, kdy dochází k updatu vah. Hodnota t je pro každý řez v minibatchi jiná.

Sítě v původní práci byly trénovány na datasetu s větším počtem tříd. Tato práce pracuje ale také s daty, kde pro jeden řez CT skenu existuje anotace jen pro jednu třídu. Aby se síť neučila pouze vybírat známý objekt bez ohledu na přítomnost uživatelské interakce, byly během trénování některé řezy náhodně předloženy síti bez jakékoliv uživatelské interakce a groundtruth anotace pro tyto řezy byla upravena tak, aby obsahovala pouze nulové hodnoty, tedy hodnoty reprezentující pozadí.

Samotné generování interakcí, které ilustruje obrázek 3.4, probíhá následovně: první interakcí pro každý řez je vždy kliknutí do nejzraněnějšího bodu oblasti objektu. Tento bod je určen z groundtruth anotací pomocí výpočtu Euklidovské vzdálenosti každého bodu objektu od rozhraní objekt-pozadí. Je proveden dopředný průchod modelem, čímž dojde k vytvoření prvotní predikce. Z tohoto výstupu sítě jsou vypočteny falešně pozitivní a falešně negativní oblasti a následující interakce je umístěna do oblasti s největší rozlohou, ať už je to oblast pozadí (odvozeno z rozlohy falešně pozitivní oblasti) nebo oblast objektu (podle falešně negativní oblasti). Tímto způsobem jsou přidány další interakce, s opakovaným znovuvypočtením predikce mezi každou interakcí, až do maximálního počtu K , který byl v této práci nastaven na $K = 5$. Lze předpokládat, že obdobně by provedl interakci skutečný uživatel při snaze označit nejvíce chybová místa. Zároveň použitím variabilního počtu interakcí je síť nucena nespolehat vždy na maximální počet interakcí, což by nebylo žádoucí



Obrázek 3.4: Automatické generování interakcí při trénování interaktivního modelu. První interakce je odvozena z groundtruth, další interakce jsou zvoleny na základě falešně negativních a falešně pozitivních segmentací.

chování, protože by docházelo k vynucování příliš vysokého počtu interakcí i při použití sítě skutečným uživatelem.

Pro trénování segmentačních sítí byla použita chybová funkce Cross Entropy, avšak pro ukládání nejlepších checkpointů je použita jiná metrika. Maximální počet trénovacích iterací pro interaktivní modely je 65 000 iterací, přičemž jako finální je uložen checkpoint s maximální hodnotou Dice skóre průběžně vyhodnocovanou na validačních datech po každých 1000 trénovacích iteracích. Tento postup byl zvolen, protože Dice skóre vyjadřuje kvalitu segmentační masky lépe než samotná chybová funkce. Interaktivní modely byly trénovány na dvou typech dat, a to na datasetu kosti lýtkové (fibula) a pánevní (pelvis). Trénovací data jsou blíže popsána v další sekci.

Kromě základní varianty interaktivní metody byla implementována ještě rozšířená varianta s přidáním zpětné vazby. U této metody je kromě jednoho kanálu s řezem CT dat a dvou kanálů s uživatelskými interakcemi přidán ještě kanál čtvrtý, který obsahuje vstupní predikci. Tato predikce v praxi může být výstupem jiné metody nebo, jak je tomu v případě trénování modelu, může představovat zpětnou vazbu z předchozí interakce. Generování uživatelských interakcí v prediktivní části je obdobné jako u základní interaktivní metody kromě toho, že je na vstup vložen také předchozí výstup sítě. Tento výstup je použit bez jakýchkoliv úprav. Lze předpokládat, že se model takto naučí lépe opravovat své vlastní chyby predikce na základě uživatelského vstupu.

3.3 Dataset

Pro trénování modelů byl použit dataset CT skenů částí lidského těla poskytnutý firmou TESCAN 3Dim. Z datasetu pokrývajícího končetiny, oblast pánve a hlavy byly vybrána některá data tak, aby bylo možné zkoumat chování vybraných metod při segmentaci různě složitých dat. Byly připraveny tři samostatné datasety: **Fibula**, **Pelvis** a **Longbones**.

Dataset Fibula pro základní sadu experimentů

Dataset Fibula je považován za základní dataset a byly na něm provedeny všechny experimenty v této práci, přičemž některé z nich byly následně vyzkoušeny i s datasetem Pelvis. Dataset Fibula se skládá z CT snímků, které obsahují vždy spodní část jedné nohy, od kotníku po přibližně polovinu kolene. V datasetu jsou rovnoměrně zahrnuta data levé i pravé nohy a je zde obsažen i případ pacienta s kloubní náhradou kolene. Groundtruth anotace jsou rovněž součástí datasetu a obsahují pixelwise označení přítomnosti kosti lýtkové (fibula). Kost fibula má (v axiálních řezech) poměrně jednoduchý tvar, který v případě zdravé kosti není příliš variabilní a segmentace tak není příliš náročná, s výjimkou oblastí hlavic kosti. Co naopak dělá úlohu složitější je přítomnost druhé kosti, kosti holenní (tibia), která s kostí lýtkovou sdílí jistou podobnost a model se tedy musí naučit buďto mezi kostmi rozlišovat, nebo využívat uživatelské interakce při označení objektu.

Dataset Longbones a předtrénování na odlišných datech

Dataset Longbones je smíšený dataset a obsahuje několik různých druhů dlouhých kostí a byl vytvořen pro experimenty s transfer learningem označované v této práci jako **datashift**. Na rozdíl od předtrénování na restaurační úlohu (**taskshift**) je v tomto případě síť nejprve předtrénována na segmentaci zdrojových dat a teprve následně dotrénována na segmentaci cílových dat. V případě dotrénování už zdrojová data nejsou použita. Tento způsob přetrénování může být použit v případě, kdy již máme k dispozici anotovanou datovou sadu, která sdílí jistou podobnost s novou, neanotovanou datovou sadou. V případě segmentace fibuly lze tedy použít předtrénování na datasetu Longbones, který obsahuje skeny stehenní kosti (femur), pažní kosti (humerus) a holenní kosti (tibia). Dataset obsahuje celkem 205 snímků, a to 87 snímků kosti stehenní, 89 snímků kosti holenní a 29 snímků kosti pažní.

V případě, že lze při použití datashift předtrénování dosáhnout ucházející přesnosti segmentace už při použití několika málo skenů cílových dat, lze tento způsob trénování využít například pro tvorbu nového datasetu a to následujícím způsobem: nejprve manuálně vytvořit dvě počáteční anotace, tyto použít k natrénování slabého interaktivního modelu s využitím předtrénování na zdrojových datech a slabý model potom použít k vytvoření dalších anotací. Iterativně lze takto vytvořit dataset nových dat, který již může být použit pro natrénování automatického modelu.

Dataset Pelvis pro datové rozšíření experimentů

Poslední použitý dataset je dataset Pelvis, na kterém byly provedeny některé vybrané experimenty pro ověření použitelnosti metod pro segmentaci jiného typu dat, než jsou dlouhé

Tabulka 3.2: .

Dataset	Trénovací	Validační	Evaluační
Fibula	70	9	10
Pelvis	37	5	5
Longbones: Femur	64	9	9
Longbones: Tibia	65	9	9
Longbones: Humerus	21	3	3

kosti. Dataset pelvis obsahuje 50 snímků oblasti pánve. Kromě pánevní kosti jsou v této oblasti přítomny kosti z okolí pánve (např. kost sedací) a v krajních snímcích skenů také části kostí stehenních. Je vhodné si uvědomit, že jeden řez z datasetu Pelvis pokrývá levou i pravou polovinu těla na rozdíl od datasetů Longbones a Fibula, kde se v jednom řezu nachází buď pravá nebo levá varianta segmentované kosti. Rozdíl je taky v počtu celistvých regionů v rámci jednoho řezu, kdy řezy dlouhými kostmi obsahují typicky jeden celistvý region segmentovaného objektu, zatímco pánev se v 2D řezu vyskytuje v několika oddělených regionech. Z tohoto důvodu je pro interaktivní segmentaci pánve nutno použít víc uživatelských interakcí na jeden řez, než u dat s dlouhými kostmi.

Rozdělení podmnožin pro trénování a vyhodnocení

Všechny datasety byly náhodně rozděleny do čtyř podmnožin: trénovací, validační, evaluační a záložní. Trénovací data byla použita v různém množství u různých modelů, což bude popsáno dále. Validací data byla použita při procesu trénování pro ověření konvergence. Během trénování je na validačních datech měřeno Dice skóre a na základě jeho hodnoty je ukládán nejlepší model. Validací data nejsou přímo použita pro update parametrů modelu. Kvalita finálního modelu, vybraného na základě validačního měření, je vyhodnocena pomocí evaluačních dat. Pro evaluaci je měřeno Dice skóre a Hausdorfova vzdálenost. Implementace metrik měřených nad celým objemem byla převzata z DeepMind repozitáře ¹. Záložní data byla ponechána nepoužita pro případ, že by bylo z jakéhokoliv důvodu potřeba nahradit některý ze skenů v ostatních podmnožinách.

Konkrétní číselné (počet 3D skenů resp. počet pacientů) rozdělení datasetů do trénovacích, validačních a evaluačních podmnožin je uvedeno v tabulce 3.2:

3.4 Sada skupin modelů pro benchmarking metod

Výše popsané techniky byly vybrány pro řešení problému nutnosti velkého množství anotovaných dat pro trénování segmentačních modelů. Pro ověření použitelnosti jednotlivých technik (a jejich kombinací) bylo natrénováno značné množství modelů s různým nastavením metod a různým množstvím trénovacích dat. Pro každou metodu nebo kombinaci metod byla vytvořena skupina modelů trénovaných na různém množství skenů z datasetu Fibula a pro vybrané metody byla vytvořena skupina modelů trénovaných obdobně na datasetu Pelvis. Skupiny pro dataset Fibula obsahují vždy čtrnáct modelů pro každou metodu

¹<https://github.com/deepmind/surface-distance>

a skupiny pro dataset Pelvis modelů pět. Každý model byl trénován na daném počtu trénovacích 3D skenů. Počty 3D skenů v trénovacích podmnožinách byly navrženy tak, aby bylo možné pozorovat vliv počtu dat na kvalitu segmentace pomocí dané metody. Pro dataset Fibula byly tyto počty 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70 a pro dataset Pelvis potom 1, 2, 10, 30, 37.

Lze očekávat, že trénování modelů s minimálním počtem trénovacích dat nemusí nutně vést k výsledkům, které lze v praxi použít jako finální segmentaci. Extrémní případy jsou ale zajímavé kvůli sledování trendu zlepšení kvality segmentace pomocí různých metod. Praktické využití modelů trénovaných na minimálním počtu pacientů je coby anotační nástroj, který sice neposkytuje velmi kvalitní výsledky, ale může usnadnit manuální vytvoření groundtruth segmentace pro pozdější natrénování automatického modelu.

Tři základní skupiny

První experiment se zaměřuje na srovnání tří základních variant navržených metod, kterými jsou Seg, SegPTaskshift, SegI. Byly vytvořeny tři skupiny modelů odpovídající vybraným metodám. Seg skupina představuje automatické modely trénované s náhodou inicializací a bez interaktivního vstupu. SegPTaskshift je skupina automatických modelů předtrénovaných na úloze restaurace obrazu a SegI je skupina základní interaktivní metody s náhodnou inicializací. Skupiny Seg a SegI byly natrénovány jak pro dataset Fibula, tak pro dataset Pelvis. Skupina SegPTaskshift nebyla natrénována pro dataset Pelvis, jelikož se jedná o neinteraktivní metodu, která není z hlediska této práce, zaměřené spíše na interaktivní segmentaci, tolik zásadní.

Taskshift a interakce

Druhým experimentem je kombinace obou základních přístupů. SegIPTaskshift je interaktivní model předtrénovaný na restauraci obrazu. Předtrénování probíhá stejným způsobem jako u automatického modelu. Ve fázi, kdy se trénuje restaurace obrazu, jsou kanály pro uživatelský vstup nevyužity a data jsou do nich vkládána až při trénování segmentace. Metoda byla vyzkoušena jak na datasetu Fibula, tak na datasetu Pelvis.

Datashift

U skupiny SegIPDatashift byl využit dataset Longbones pro předtrénování interaktivního modelu. Předpokládá se, že pro zvýšení přesnosti segmentace lze využít podobnosti mezi zdrojovými a cílovými daty, v tomto případě tedy podobnosti ve tvaru (zejména u fibuly coby typu cílových dat), konzistenci a intenzitě segmentované tkáně. Modely v této skupině jsou nejprve trénovány na interaktivní segmentaci smíšeného datasetu Longbones a následně stejným způsobem přetrénovány na segmentaci cílových dat, již bez přítomnosti dat z datasetu Longbones. Vytvořena byla skupina pro oba cílové datasety, Pelvis i Fibula.

Interakce a zpětná vazba

Poslední skupinou modelů je skupina SegIPlus. Jedná se o rozšířenou interaktivní metodu. Jak již bylo popsáno výše, na vstup modelu je přidán čtvrtý kanál, který obsahuje hrubou segmentaci, resp. v případě trénování segmentaci z předchozí interakce. Jedná se tedy o

jakousi zpětnou vazbu, která má modelu pomoci lépe adresovat místa, která byla s předchozími interakcemi segmentována nesprávně. Skupina byla natrénována pouze pro dataset Fibula.

3.5 Implementace skriptů

Pro implementaci trénování neuronových sítí byl použit skriptovací jazyk Python 3 a framework Tensorflow 1.13, resp. jeho funkcionální API. Dále byly využity moduly pro manipulaci s daty, zejména knihovna pytables umožňující použití strukturovaných souborů **hdf5**. Tento formát lze použít na zapouzdření celého datasetu, přičemž lze využít i funkcionality "in-memory" souborů, které umožňují rychlý přístup k datům, která jsou v průběhu trénování udržována v operační paměti, což výrazně urychluje přípravu minibatche při trénování. Pro další manipulaci s daty byla použita především knihovna numpy. Distorzní metody pro trénování restaurace obrazu byly převzaty z repozitáře Models Genesis [45] a upraveny. Trénování sítí probíhalo na výpočetním gridu MetaCentrum s využitím GPU Tesla T4 s Cuda compute capability 7.1. Skripty, které jsou součástí elektronické části této práce, jsou popsány dále v této sekci.

Skripty k manipulaci s daty

Bylo vytvořeno několik jednoduchých skriptů pro manipulaci s CT daty. Vzhledem k tomu, že dataset obsahuje data ve dvou formátech (DICOM a numpy), byla data nejprve sjednocena, a to převedením DICOM dat do společného formátu numpy pomocí skriptu **dicomToNpy.py**. Sjednocená data jsou ve formátu numpy a to vždy ve třech souborech pro každý jeden 3D sken. Tyto soubory jsou **orig.npy** obsahující originální CT data, **seg.npy** obsahující groundtruth anotaci a **ID.xml** obsahující informaci o reálných rozměrech voxelů (kde ID je označení unikátní název 3D skenu).

Pro získání rychlé představy o podobě dat obsažených v datasetu lze použít skript **generateSUM.py**, který vytvoří sumační průměty jednotlivých 3D skenů a uloží je jako obrazové náhledy ve formátu jpg.

K samotnému trénování byla použita data ve formátu hdf5 v implementaci pytables. K převedení datasetů do tohoto formátu slouží skripty **NpyToHdf5Simple.py**, **NpyToHdf5Mixed.py** a **spacingToHdf5.py**. První skript slouží k převedení jednoduchého datasetu s jedním typem dat a byl použit pro vytvoření datasetu Fibula a Pelvis. Druhý skript byl potom použit pro dataset Longbones a umožňuje vytvoření smíšeného datasetu se zachováním hierarchie. Data tak mohou být v datasetu Longbones uložena ve skupinách podle typu kosti. Oba skripty umožňují také pixel shuffling pro vytvoření předgenerovaných dat pro trénování restaurace obrazu. Skript **spacingToHdf5.py** slouží k převedení xml souborů s údaji o rozměrech voxelů. Data pro každý dataset jsou ve výsledku uložena ve čtyřech hdf5 souborech, kterými jsou soubor obsahující originální ct data, soubor s daty na kterých byl proveden pixel shuffling, soubor s groundtruch anotacemi a soubor s rozměry voxelů.

Trénovací skripty

Skripty pro řízení trénovacího procesu jsou celkem čtyři a odpovídají vybraným metodám. Trénovací skripty mají obdobnou kostru, liší se však ve způsobu přípravy minibatche a způsobu validace. Skripty umožňují trénování s vytvořeným nového modelu stejně jako použití modelu předtrénovaného. Toho lze využít buď k transfer learningu nebo k rozdělení trénování na několik postupných úloh, což bylo nutné kvůli časovému omezení trvání jedné úlohy na výpočetním gridu.

Skripty je potřeba pro správné fungování nastavit, což lze udělat úpravou části skriptu označená jako `settings`. V této části lze nastavit místo uložení logů a dalších výstupů, počet iterací, počty trénovacích a validačních dat, rozměry trénovacích výřezů a další. Některé atributy lze nastavit také pomocí argumentů. Takto lze nastavit například počet trénovacích dat, což bylo použito k jednoduššímu hromadnému spuštění celé trénovací skupiny modelů s různým počtem trénovacích skenů.

Během trénování jsou ukládány ukázkové výstupy sítě v podobě obrazových dat ve formátu png. U interaktivních modelů jsou také ukládány ukázkové mapy interakcí, které byly použity pro trénování daného modelu. Nakonec je také uložen soubor `log.txt`, který obsahuje údaje o vývoji chybové funkce a Dice skóre během trénování a stejné informace jsou pro rychlejší získání přehledu o vývoji trénování ukládány v podobě grafů.

Konkrétní použití trénovacích skriptů je následující:

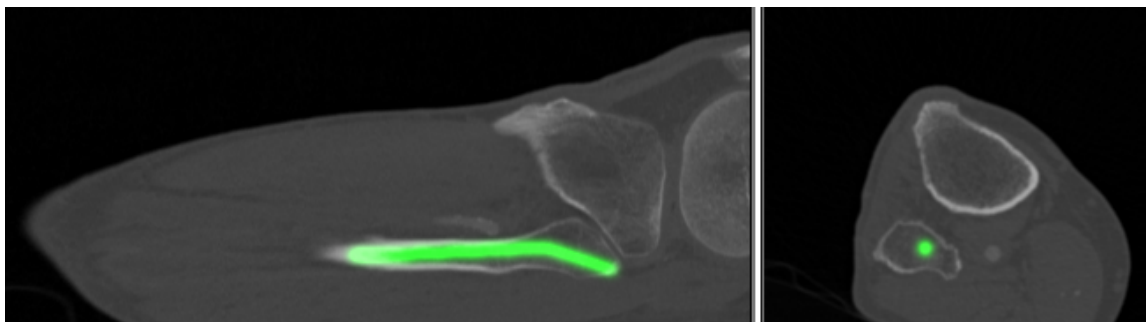
- **Resto.py**: restaurace obrazu, předtrénování modelů pro další úlohy,
- **Seg.py**: automatická segmentace datasetů Fibula a Pelvis, skupiny Seg, SegPTaskshift,
- **SegI.py**: interaktivní segmentace datasetů Fibula a Pelvis, skupiny SegI, SegIPDatashift, SegIPTaskshift,
- **SegIPlus**: interaktivní segmentace rozšířená o zpětnou vazbu, skupina SegIPlus.

Evaluační skripty

Pro automatizované vyhodnocení kvality segmentace finálních modelů byly vytvořeny tři evaluační skripty: `evaluateSeg.py`, `evaluateSegI.py` a `evaluateSegIplus.py`. Skripty opět fungují obdobně a liší se hlavně manipulací s daty. Kostra, společná pro všechny tři skripty, se skládá z načtení zvoleného modelu, CT dat a groundtruth anotace a postupného vyhodnocení kvality segmentace na zvoleném rozsahu dat. Vyhodnocení kvality je provedeno pomocí metrik Haunsdorfova vzdálenost a Dice skóre, jejich implementace byla rovněž převzata z DeepMind repozitáře. Tyto metriky jsou měřeny na celém segmentovaném 3D objemu po tom, co byly získány výstupy pro jednotlivé řezy. Pro lepší čitelnost výsledků jsou v textu této práce uvedeny pouze některé výstupy evaluace, avšak kompletní měření je k dispozici v elektronické části a to v souboru **Results.xlsx**.

Nástroj pro interaktivní segmentaci

Posledním vytvořeným skriptem je skript `segmentationTool.py`. Skript umožňuje jednoduché manuální použití interaktivních sítí. Implementováno je prohlížení 3D objemu ve dvou



Obrázek 3.5: Příklad anotací fibuly vytvořených interaktivním nástrojem v laterálním pohledu (vlevo) a v axiálním řezu (vpravo).

různých náhledech (axiální a laterální) a umístění uživatelských interakcí v podobě kliknutí přímo do obrazu. Pro správné fungování sítě trénované po 2D řezech je nutné označení v každém řezu 3D objemu, práce vyžadována od uživatele může být ale značně redukována vhodnou implementací anotačního nástroje.

Zde lze zapnout funkci propojování kliknutí. Vytvoření anotací pro interaktivní segmentaci pak může probíhat následovně. Uživatel prochází vybraný objekt po jednotlivých řezech a umístí kliknutí vždy jednou za několik řezů. Pokud je zapnuta funkce propojování kliknutí, dojde k propojení dvou po sobě následujících kliknutí, čímž dojde k označení objektu ve všech vrstvách, které leží mezi řezy, do kterých byla kliknutí umístěna. Tímto způsobem lze označit dlouhou kost typu fibula během několika minut až sekund.

Druhým způsobem označení je označení v laterálním pohledu. Zde lze také použít funkci propojování kliknutí. Kost typu fibula typicky stačí označit pomocí několika málo propojených kliknutí v jednom laterálním řezu a následně provést korekci krajních částí kosti v řezu axiálním. Do budoucna lze rovněž uvažovat implementaci heuristik pro průchod objektem na základě density, což by umožnilo automatické propojení extrémních bodů objektu a tedy další výrazné urychlení anotace. Příklad anotací vytvořených interaktivním nástrojem lze vidět na obrázku 3.5.

Kapitola 4

Experimenty a vyhodnocení

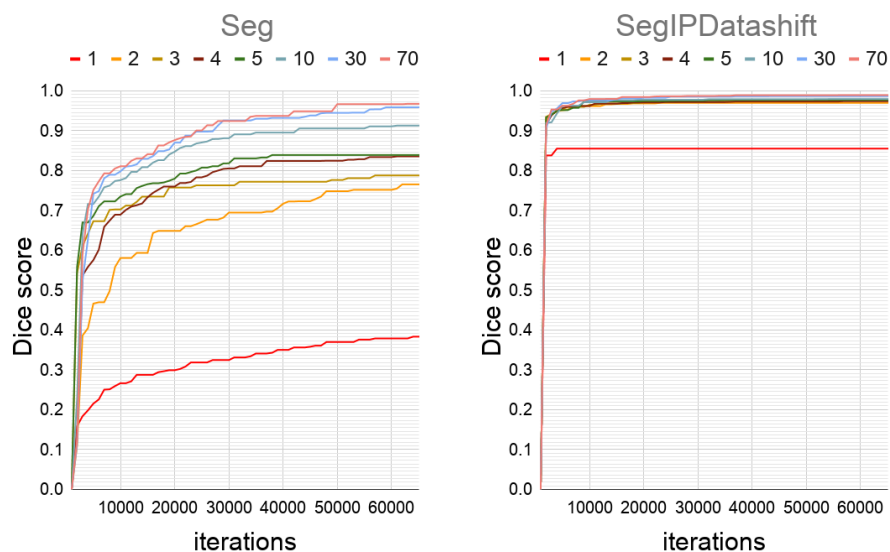
V předchozí kapitole byly popsány vybrané metody pro řešení úlohy segmentace CT dat. Tato kapitola popisuje jednotlivé experimenty, provedené s výše popsanými metodami. Pro každou metodu byla natrénována sada modelů lišících se v rámci skupiny množstvím trénovacích dat, které bylo k jejich trénování použito. Vzhledem k tomu, že bylo provedeno poměrně velké množství experimentů, byly pro lepší čitelnost z grafů a tabulek vynechány informace o některých modelech v rámci skupiny. Trendy ve výsledcích jednotlivých metod jsou patrné i s vynecháním některých dat. Kompletní numerické výsledky jsou uvedeny v elektronické části této práce.

Nejprve je uvedeno srovnání dvou zvolených přístupů, tedy předtrénování na zdrojové úloze (SegPTaskshift) a použití interaktivní metody (SegI), s plně automatickým modelem (Seg). Dále jsou popsány rozšiřující experimenty, kterými jsou transfer learning v podobě kombinace interaktivního přístupu a předtrénování na zdrojové úloze (SegIPTaskshift), kombinace interaktivního přístupu a předtrénování na zdrojových datech (SegIPDatashift), interaktivní segmentace s přidanou zpětnou vazbou (SegIPlus).

Vyhodnocení metod bylo provedeno na základě tří různých typů dat, která byla naměřena a shromážděna. Modely jsou sledovány pomocí záznamu hodnot chybové funkce a případně Dice skóre již během procesu trénování, což umožňuje srovnat rychlost konvergence, která je, jak je dále patrné, u některých modelů výrazně vyšší než u baseline modelu (Seg), a která může být z praktického hlediska také důležitým parametrem. Druhým typem numerických výsledků je Dice skóre měřené pro finální verze modelů s použitím evaluačních dat. Tímto je zajištěna větší objektivita pro srovnání modelů, na rozdíl od srovnání výsledků dosažených během trénování na datech validačních. Posledním způsobem hodnocení je vizuální porovnání výstupů sítí a v případě interaktivních modelů také testování pomocí implementovaného nástroje pro interaktivní segmentaci.

4.1 Porovnání základních variant zvolených metod s automatickým modelem

Základním experimentem, který byl proveden jako první, je srovnání dvou zvolených rozšíření s baseline modelem, kterým je náhodně inicializovaná a plně automatická segmentační síť. Experiment ověřuje myšlenku, zda je možné použít transfer learning, či případně in-

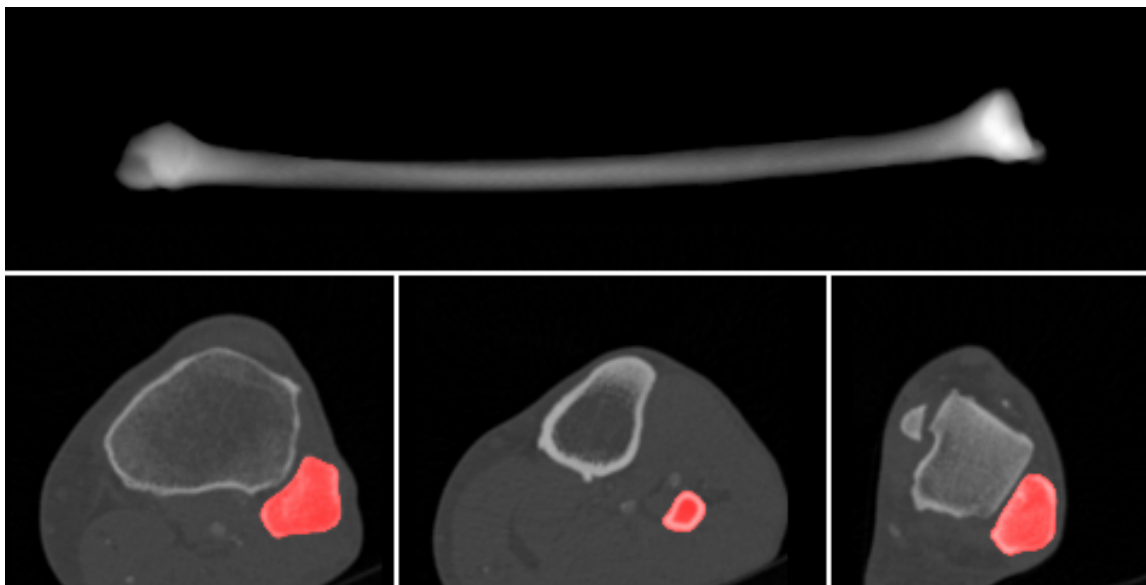


Obrázek 4.1: Grafy konvergence pro metody Seg a SegIPDatashift. Pro každou iteraci je uvedena vždy nejlepší dosažená hodnota Dice skóre na validačních datech. Je vidět, že interaktivní model předtrénovaný na zdrojových datech konverguje při dotrénování velmi rychle ve srovnání s baseline modelem.

teraktivní uživatelský vstup, pro zlepšení kvality segmentace. Přínos zvolených rozšíření je zvláště patrný u modelů s nízkým počtem trénovacích 3D skenů. Tento experiment byl také zpracován jako samostatný článek pro studentskou konferenci Excel 2020.

Úspěšnost baseline metody

První vyhodnocení baseline metody bylo provedeno s datasetem **Fibula**. Z grafu vývoje Dice skóre v průběhu trénování (graf 4.1) je vidět, že modely ze skupiny Seg po počátečním skoku konvergují relativně pomalu v průběhu 65 000 iterací. Základní automatický model dosahuje při plném počtu trénovacích skenů (tj. 70 skenů) relativně vysoké hodnoty Dice skóre na validačních datech v průběhu trénování a obdobně i na evaluačních datech při automatické evaluaci konečného modelu. Segmentace s přesností pohybující se kolem hodnoty 0,98 Dice skóre může být považována za kvalitní a z praktického hlediska použitelnou. Příklad takové segmentace je na obrázku 4.2. Pro vytvoření velmi dobrých segmentací, u nichž bude nutnost manuální úpravy minimální, požadujeme, aby většina segmentací dosahovala Dice skóre aspoň 0,98 a aby ani ojedinělé případy neklesaly pod hodnotu 0,97. Pro srovnání je na obrázku 4.3 ukázka segmentace s Dice skóre 0.900. Takováto segmentace již pro praktické použití vyžaduje větší úpravy, při segmentaci fibuly zpravidla v oblasti kloubních hlavic. Prvním modelem ze skupiny Seg, který dosahuje takto ideálních hodnot Dice skóre, je model Seg30. Naopak, jak lze očekávat, modely s nízkým počtem trénovacích skenů dosahují daleko menších hodnot Dice skóre. Zvláště nízká je hodnota u modelu Seg1, tedy modelu trénovaného pouze s jedním trénovacím skenem. Kromě horšího odhadu správného tvaru segmentační masky pro segmentovanou kost je nízká hodnota Dice skóre u tohoto modelu dána také špatnou lokalizací segmentovaného objektu v případě, že se jedná o opačnou (levou/pravou) variantu než s jakou byl model trénován. (Tento nedostatek by



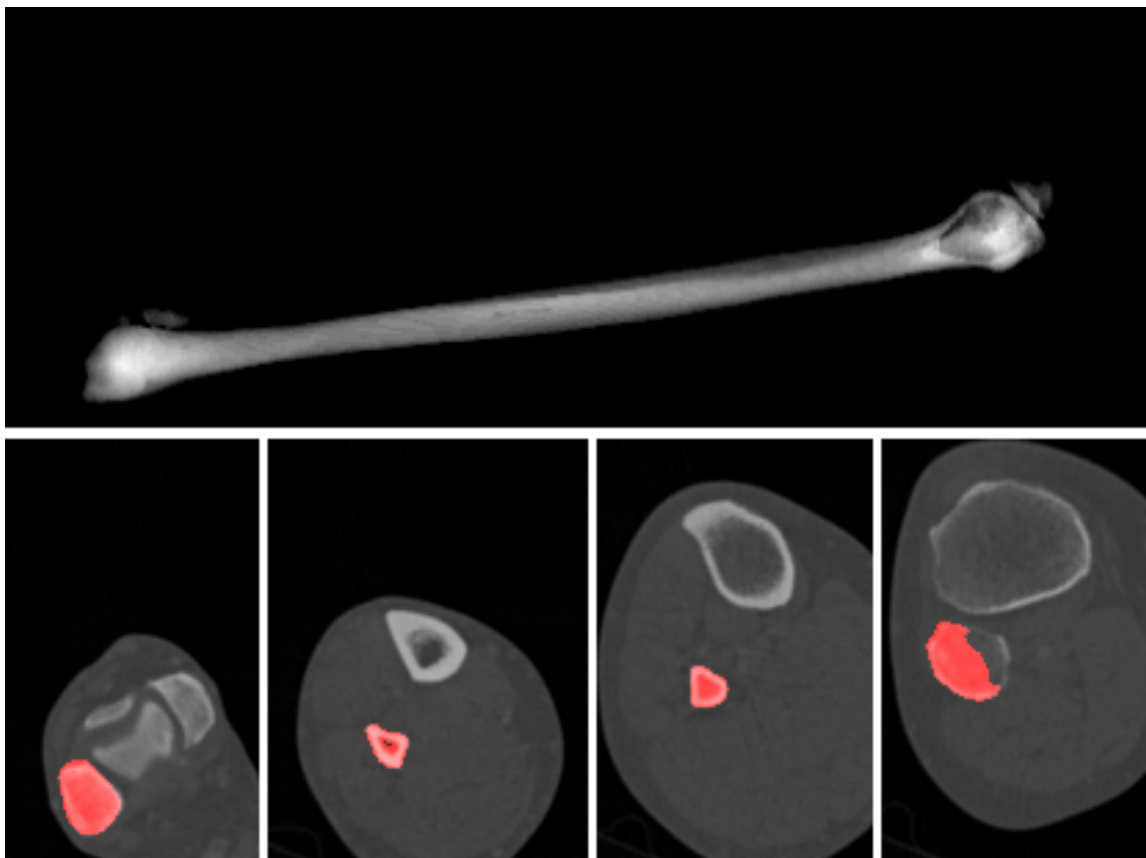
Obrázek 4.2: Příklad segmentace Dice skóre s hodnotou 0.989. Tři různé řezy CT snímkem fibuly a průmět celé 3D segmentační masky.

byl částečně řešitelný také zavedením regularizace datasetu v podobě transformací, které by zahrnovaly převrácení obrazu podle osy.)

Při vizuálním zhodnocení výsledků automatické evaluace modelu Seg1 si lze všimnout, že lokalizace stranově opačné kosti u modelu s jedním trénovacím skenem naprosto selhává. Segmentace známé strany je často neúplná, selhává segmentace zejména vnitřní části kosti. Dochází také poměrně často k falešně pozitivní segmentaci sousedních kostí, zejména v oblasti kloubních hlavic. K falešně pozitivní segmentaci dochází také u stranově neznámých dat, kde je jako kost označena chybně i měkká tkáň přiléhající k sousedící tibii (obrázek 4.4). Model Seg70 je oproti tomu přesnější, většina problémů se zde nevyskytuje nebo se vyskytuje v menší míře. Zřídka dochází k falešně pozitivním segmentacím nesprávných objektů v řezech, kde se fibula vůbec nevyskytuje. Tvar segmentované kosti je občas nepatrně odlišný od groundtruth segmentace, což může být ale způsobeno také limitovanou přesností člověkem vytvořených anotací.

Při trénování s datasetem **Pelvis** dosahuje model Seg vyššího Dice skóre, než je tomu u datasetu Fibula. To je překvapivé, protože pánevní kost je tvarově složitější než kost lýtková a dalo by se tedy čekat, že bude její segmentace náročnější. Vývoj Dice skóre na validačních datech je z hlediska rychlosti konvergence obdobný jako u modelů trénovaných na Fibule. Je vhodné upozornit, že na rozdíl od Fibuly obsahuje jeden 3D sken datasetu Pelvis obě anatomické části (levá/pravá) pánevní kosti. Je tedy vhodné srovnávat model Seg1 Pelvis raději s modelem Seg2 Fibula. Přesto jsou ale výsledky pro segmentaci pánve lepší. Vysvětlením může být například lepší jednotnost dat u datasetu Pelvis, daná menší škálou použitých snímacích zařízení. Tato teorie ale nebyla vyvrácena ani potvrzena.

Při trénování na segmentaci **pánve** jsou vizuálně zhodnoceny výstupy modelů Seg1 a Seg37, kde Seg37 je model s největší trénovací množinou pro pánev. Jsou zde podobné rozdíly mezi výsledky, jako i datasetu Fibula. Model Seg1 má problém jak s falešně pozitivní

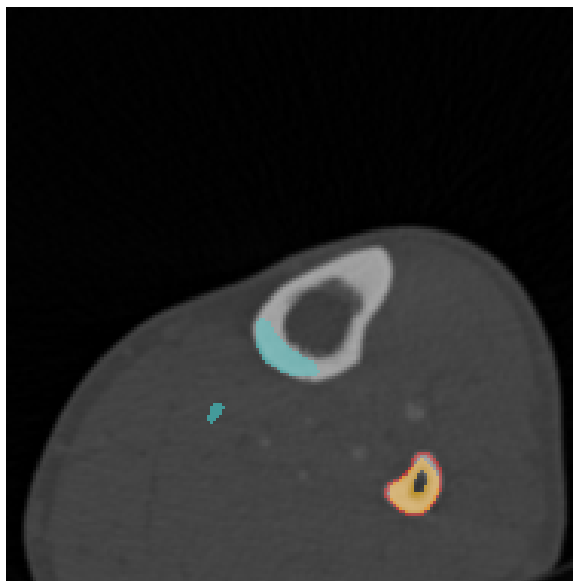


Obrázek 4.3: Příklad segmentace Dice skóre s hodnotou 0.900. Tři různé řezy CT snímkem fibuly a průmět celé 3D segmentační masky.

tak falešně negativní segmentací. K falešně pozitivní segmentaci dochází v některých případech v částech 3D skenu, kde se ještě pánev nenachází a jsou zde chybně za regiony pánve pokládány zde přítomné femury. Kromě toho má model mnohdy velmi zřetelný problém s falešně negativní segmentací různých částí pánve. Problematické oblasti jsou přitom různé a nejedná se o tak jednoznačný problém jako je například segmentace vnitřku kosti u modelů trénovaných na Fibulu. Oproti tomu model SegI37 s falešně negativní segmentací uvnitř kosti nemá problém téměř nikdy a dochází zde většinou k chybám nesprávné segmentace přilehlého okolí pánve či některých výběžků. Některé příklady suboptimální segmentace jsou vidět na obrázku 4.5.

Interaktivní přístup

Interaktivní přístup přináší výrazné zlepšení číselných výsledků automatické evaluace segmentace pro modely s nízkým počtem trénovacích skenů, zejména u datasetu **Fibula**. Důvodem je vliv přidání uživatelského vstupu, který pomáhá k lepší lokalizaci segmentovaného objektu oproti baseline metodě. Příklad, kdy interaktivní segmentace zafungovala mnohem lépe, než baseline model, je možné vidět na obrázku 4.6. Při porovnávání výsledků je vhodné mít na paměti, že interaktivní model je oproti baseline modelu zvýhodněn přidanou informací v podobě uživatelské interakce, která je u interaktivního modelu vždy přítomna. Zdá



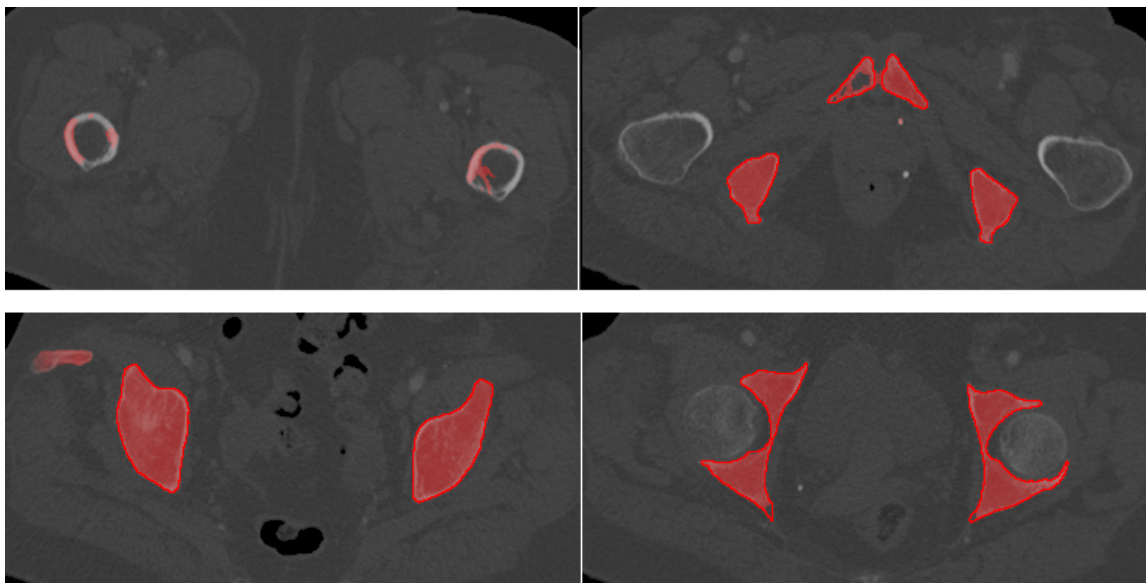
Obrázek 4.4: Ukázka segmentace kosti, jejíž stranová varianta (levá/pravá) nebyla použita při trénování modelu Seg1 (modrá) a SegI1 (oranžová). Groundtruth anotace je vyznačena červeným obrysem.

se také, že interaktivní modely při trénování konvergují rychleji než baseline modely.

Při bližším prozkoumání výsledků automatické evaluace SegI1 je vidět, že i s použitím jednoho kliknutí na řez dochází k lepší lokalizaci segmentované fibuly než u modelu Seg1. Oproti Seg1 ubývá případů chybně segmentovaných regionů v oblastech, kde není přítomna hledaná kost. Model už je částečně schopný segmentovat i stranově opačnou kost, zejména ve středové oblasti kosti (viz obrázek 4.4). Problém s špatně vysegmentovaným vnitřkem kosti je ale stále patrný. Při automatické evaluaci s maximálním počtem uživatelských interakcí $K = 5$ dochází jen k velmi malému zlepšení výsledků. Model reaguje nejvíce na první interakci, další interakce už ale pomáhají jen v omezené míře.

Lepších výsledků lze dosáhnout již s modelem SegI2. Při manuální anotaci s jedním kliknutím na každý řez dosáhl model v konkrétním případě Dice skóre 0,8723. Přidáním anotací v druhém průchodu došlo ke zlepšení Dice skóre na hodnotu 0,9501. Oblast okolo hlavice kosti stále obsahuje chybně segmentovaná místa. Je ale možné, že s použitím regularizací, včetně transformace převrácení obrazu během trénování, by bylo možné dosáhnout z praktického hlediska zajímavých výsledků již s modelem Seg1.

Model SegI70 méně chybuje v některých okrajových oblastech kosti než model Seg70. Nedochozí k segmentaci v řezech bez přítomnosti fibuly, protože se model správně naučil segmentovat pouze označený objekt. V okrajových částech, kde dochází k falešně pozitivní segmentaci sousední kosti, je možné tuto chybu dobře odstranit přidáním dalších interakcí. Při manuálním testování modelu Seg70 se podařilo dodatečnou anotací v podobě dvou propojených kliknutí doplnit část vnitřku kosti prve chybně označenou jako pozadí a rovněž několika dalšími kliknutími s anotací pozadí odstranit falešně pozitivní segmentaci v okolí hlavice kosti. V tomto konkrétním případě došlo ke zlepšení Dice skóre dodatečnou anotací z 0,9731 na 0,9796, což není mnoho. Vizualně je ale patrné (obrázek 4.7), že přidáváním



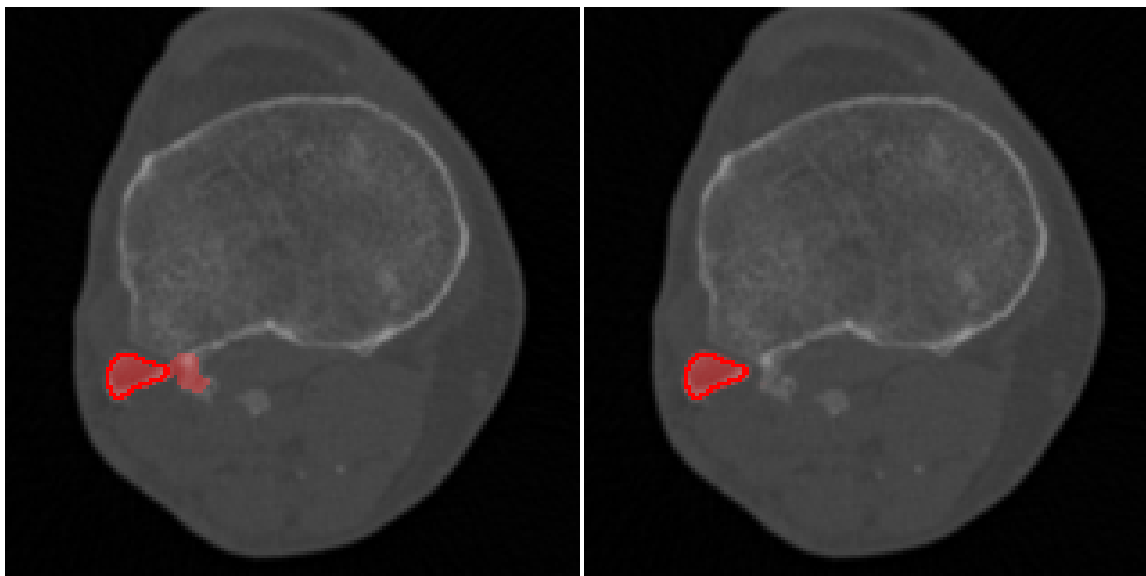
Obrázek 4.5: Ukázka některých problémů, které mají modely Seg1 (horní řádek) a Seg37 (dolní řádek) při segmentaci pánve. (Groundtruth vyznačeno červenou linkou.) Model Seg1 chybně vysegmentoval i femury přítomné v jednom z krajních řezů.

interakcí může dojít k dobrému odstranění chyb v segmentaci.

Z hlediska numerických výsledků automatické evaluace nedochází u datasetu **Pelvis** k tak výraznému zlepšení jako u datasetu **Fibula**, což je ale částečně dané tím, že hodnoty u baseline modelu jsou pro segmentaci pánve již relativně vysoké a není zde tedy takový prostor pro zlepšení. Pokud není přítomen dostatečný počet interakcí, je výsledek segmentace horší než baseline metoda. Co se týče konvergence, dochází i zde k urychlení.

Při srovnání výstupů interaktivního modelu a automatického modelu pro pánev lze vidět, že interaktivní vstup pomáhá odstranit chybu v podobě falešně pozitivní segmentace femurů, která se vyskytovala u modelu Seg1. I když u modelu SegI1 stále dochází k nežádoucí přítomnosti děr v segmentované pánvi (i při velkém počtu anotací objektu), obecně je tento problém u interaktivního modelu menší. Podle automatické evaluace s jednou interakcí se zdá, že u modelu SegI37 má oblast ovlivněná jedním uživatelským vstupem větší plochu než u modelu SegI1. (Model SegI37 dokáže na základě jednoho kliknutí správně segmentovat větší oblast než model SegI1.) Rozdílný dosah uživatelské interakce ukazuje obrázek 4.8. Při evaluaci modelu SegI37 s pěti interakcemi už je kvalita výstupní segmentace velmi vysoká. Model dobře řeší problém falešně pozitivních i falešně negativních segmentací a chyby jsou při zběžné vizuální kontrole téměř nepatrné.

Modely SegI1 a SegI37, které byly natrénovány na segmentaci **pánve**, byly rovněž vyhodnoceny pomocí manuální evaluace. Obrázek 4.9 ukazuje, jak model SegI37 pracuje s postupně přidávaným uživatelským vstupem. Pro správné vysegmentování všech regionů náležících pánvi je vhodné umístit do každého řezu až čtyři interakce anotující objekt. Model Seg37 dokáže v některých případech správně segmentovat i neoznačené oblasti, na základě přítomnosti anotací v sousedních regionech. Lze tedy tvrdit, že přítomnost uživatelem označeného bodu nemá vliv pouze na region, v kterém se nachází a model se nejspíš



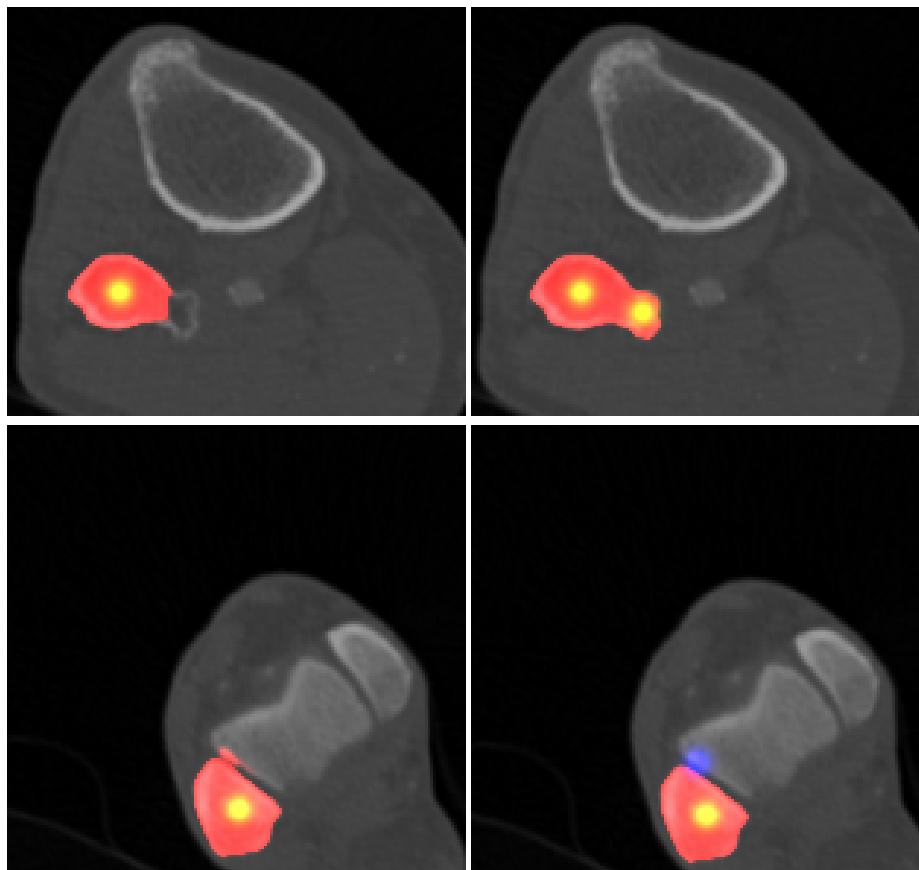
Obrázek 4.6: Příklad, kdy interaktivní segmentace fibuly zafungovala lépe než automatický model. Vlevo výstup modelu Seg70, vpravo SegI70.

částečně naučil logickou vazbu mezi oddělenými regiony pánve. Při použití maximálně tří interakcí na řez dosahuje model SegI37 Dice skóre 0,9494. Při přidání čtvrté interakce do řezů s oddělenými regiony se Dice skóre zvýší na 0,9802. Dalšími interakcemi už nedochází k výraznému zlepšení (při označení chyb zlepšení na 0,9803), ale vzhledem ke kvalitě výsledku není další zlepšení nezbytně nutné. Model SegI1 segmentuje sousedící oblasti s menší jistotou. (Tzn. odezvy pro tyto oblasti jsou přítomné, ale jsou slabší.) Při použití až čtyř interakcí dosahuje model v konkrétním případě Dice skóre 0,9504. Tuto hodnotu lze zlepšit přidáním dalších interakcí na 0,9583, avšak v některých případech již model nereaguje na přidání interakce tak, jak by bylo žádoucí.

Předtrénování na restauraci

Restaurace obrazu je zajímavá úloha sama o sobě. Jak už bylo zmíněno v předchozí kapitole, pro restaurační model byly uloženy dva checkpointy, a to pro iterace 43 000 a 104 000. Mezi těmito checkpointy už nedošlo k výraznému zlepšení chybové funkce (z 0,177 na 0,138), avšak při vizuálním srovnání výstupů sítě je patrné, že pozdější model dokáže, na rozdíl od modelu dřívějšího, překvapivě lépe rekonstruovat detaily přítomné kosti.

Model trénovaný po 43 000 iterací dokáže ve většině případů dobře reagovat na nelineární změny intenzity. Do jisté míry funguje i restaurace dat, na kterých byl proveden pixel shuffling, avšak výsledky jsou lepší ve středových oblastech kosti. V oblastech kolem hlavic kloubů je restaurovaná kost méně intenzivní, než by měla být a tvarově je vzdálena od správné anatomie. Při řešení outpainting dochází v některých případech k hrubým anatomickým chybám při rozhraní tkáně a okolního prostředí tvořeného vzduchem. Vznikají zde anatomicky nesprávné laloky, a to tím spíš, čím je oblast zakrytá outpaintingem větší. Jednodušší (resp. méně rozsáhlé) případy outpaintingu jsou řešeny uspokojivě. Velkým problémem je rekonstrukce kosti při outpaintingu, který často zakrývá velké úseky fibuly nebo



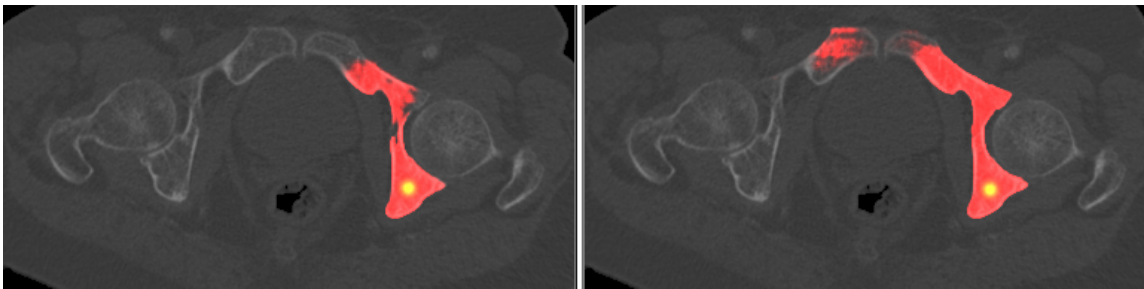
Obrázek 4.7: Přidáním dalších interakcí dochází u modelu SegI70 k odstranění některých chyb v segmentaci fibuly.

tibie. Ve většině případů je zakrytá kost nebo její část ignorována.

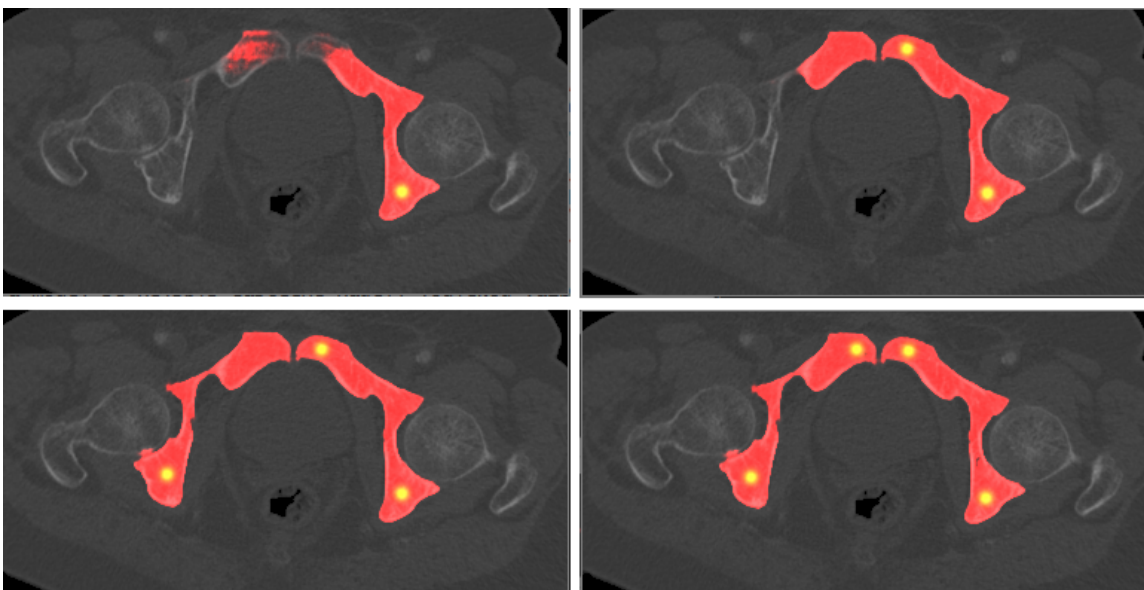
Oproti tomu model trénovaný po 104 000 iterací dosahuje tvarově jistějších výsledků při řešení pixel shufflingu a nedopouští se tak hrubých chyb při řešení outpaintingu. Největší rozdíl je ale v rekonstrukci kostí postižených inpaintingem. Pozdější model dokáže částečně rekonstruovat zasaženou kost. Přestože rekonstrukce stále není anatomicky přesná, je zde mnohem výraznější náznak správného tvaru. Z toho lze usuzovat, že se model lépe naučil extrahovat příznaky odpovídající anatomickým strukturám. Ukázka některých výstupu rekonstrukčních modelu je na obrázku 4.10.

Numerické výsledky naznačují, že použití modelu předtrénovaného unsupervised způsobem na restauraci přináší jisté zlepšení kvality segmentace. Největší rozdíl je, dle očekávání, u modelu s nejméně trénovacími daty. Důvodem je, kromě velkého prostoru pro zlepšení u modelů trénovaných na méně datech, také to, že maximální počet trénovacích skenů pro segmentaci se rovná počtu skenů, na kterém byl předtrénován restaurační model. Je možné, že při použití většího množství neanotovaných skenů pro trénování restaurace, by došlo k většímu zlepšení i u modelu Seg70.

Pozitivní segmentace modelu SegPTaskshift1 se zdají být v mnoha případech prostоровě rozsáhlejší a odezvy na hrany kosti v krajních řezech fibulou jsou mnohdy silnější



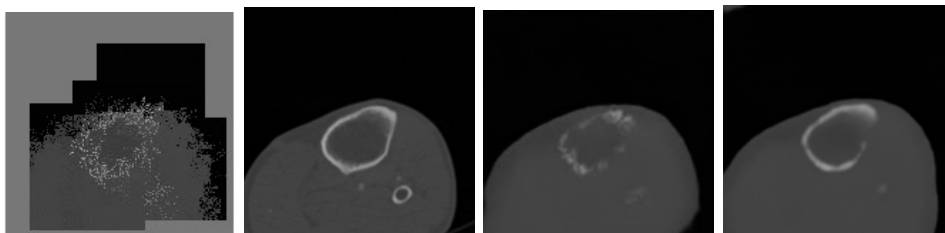
Obrázek 4.8: Dosah uživatelské interakce je odlišná u modelu SegI1 (vlevo) a SegI37 (vpravo).



Obrázek 4.9: Reakce modelu SegI37 na uživatelskou interakci při segmentaci pánve.

než u nepředtrénované varianty modelu. Přestože jsou číselné výsledky lepší u předtrénovaného modelu, vizuálně se rozdíly těžko hodnotí. U modelu SegPTaskshift70 je patrný menší výskyt falešně pozitivních segmentací, zejména proto, že to je vizuálně nejvýraznější nepřesnost u modelu Seg70.

Výrazným rozdílem mezi baseline modely a modely ze skupiny SegPTaskshift je také rychlost konvergence. Zatímco modely ze skupiny Seg pozvolna konvergují v průběhu 65 000 iterací, modely SegPTaskshift mají mnohem rychlejší nástup a přijatelných hodnot Dice skóre dosahují už po 30 000 iteracích. Je možné, že u baseline modelů by při přidání dalších iterací ještě docházelo ke zlepšení, což může vést k drobné nepřesnosti při srovnání konečných checkpointů skupin Seg a SegPTaskshift. Zejména u modelu se sedmdesáti trénovacími skeny může být drobné zlepšení způsobeno rychlejší konvergencí a tedy dosažení lepšího konečného stavu, protože zde není žádný rozdíl v počtu dat při předtrénování a při trénování segmentace. Vliv použití předtrénovaných modelů je ale zjevný, alespoň u ostatních modelů. Do budoucna by bylo zajímavé vyzkoušet předtrénování na rozsáhlejších datasetech bez anotací, se zahrnutím různých typů tkání. Přestože by byly zahrnuty i anatomické celky, které se v cílových datech nenacházejí, podobnost tkání by mohla pomoci



Obrázek 4.10: Ukázka výstupu restauračních modelů. Zleva: vstupní výřez, výřez před distorzí, výstup modelu po 43 000 iteracích a po 104 000 iteracích.

segmentačnímu modelu při analýze cílových dat.

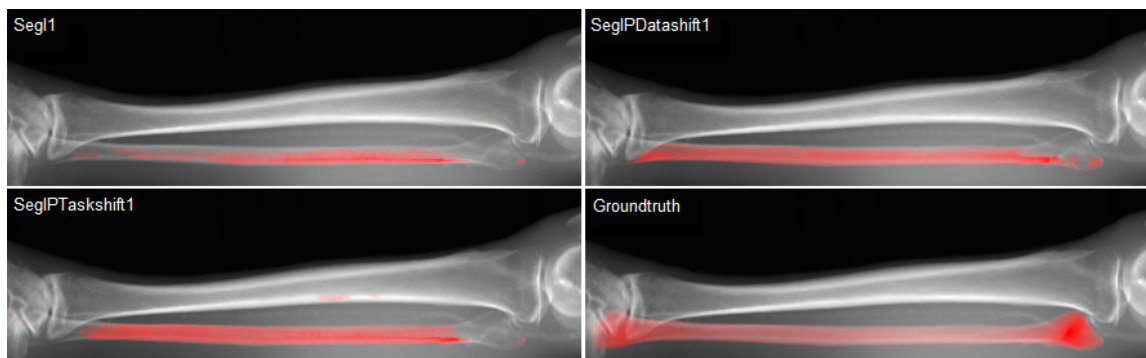
Tato metoda byla vyhodnocena pouze pro dataset Fibula.

4.2 Kombinace interaktivity a předtrénování na restauraci

Je vidět, že předtrénování na restauraci nepřináší samo o sobě takové zlepšení jako zavedení uživatelské interakce, nicméně jistá míra zlepšení zde je. Interaktivní model dosahuje zajímavých výsledků sám o sobě. Dalším experimentem bylo ověření, zda kombinace obou přístupů vede k dalšímu zlepšení výsledků segmentace.

U datasetu **Pelvis** dochází jen k minimálnímu zlepšení výsledků oproti nepředtrénovanému interaktivnímu modelu. Ke zlepšení zde navíc dochází jen u modelů s menším počtem trénovacích dat, pravděpodobně protože počet dat k předtrénování je celkově menší, takže pro modely s více trénovacími daty než deset 3D skenů nepřináší předtrénování takovou výhodu. Nelze také jednoznačně určit konkrétní vlastnost modelů SegIPTaskshift1, 2 a 10, která by se na tomto zlepšení podílela, protože výstupy jsou vizuálně velmi podobné jako u nepředtrénovaných interaktivních modelů. Výsledky segmentace s jednou interakcí jsou ale u interaktivní segmentace pánve horší než baseline metoda, jak u základní, tak u předtrénované varianty. K tomu dochází hlavně proto, že interaktivní síť segmentuje jen do určité vzdálenosti od uživatelské interakce a jedna interakce nestačí na anotování všech regionů pánve, které jsou rozmístěny na různých místech řezu. Při vizuální kontrole segmentace pánve s pěti interakcemi nejsou patrné zásadní rozdíly v kvalitě. U obou datasetů ale dochází k zrychlení konvergence oproti skupině SegI i oproti skupině SegPTaskshift.

Numerické výsledky ukazují další výrazné zlepšení segmentace datasetu **Fibula** u modelů s menšími trénovacími množinami. Model SegIPTaskshift1 dosahuje Dice skóre 0,7220, což je nejlepší výsledek pro Dataset Fibula při evaluaci s jednou interakcí a tato hodnota je zároveň o 38 % lepší než u baseline modelu Seg1. U modelů s většími množinami je zlepšení pouze velmi malé (v řádu setin až tisícín Dice skóre). Po vizuálním vyhodnocení výsledků automatické evaluace s datasetem **Fibula** lze říct, že segmentace modelem SegIPTaskshift70 už je velmi přesná a případné drobné chyby v koncových částech kosti lze odstranit již přidáním jedné interakce navíc. K falešně pozitivním segmentacím mimo přilehlé okolí segmentovaného objektu nedošlo v žádném ze sledovaných případů.



Obrázek 4.11: Úspěšnost různých interaktivních modelů při segmentaci jim neznámé stranové varianty kosti fibula se liší. Ukázány jsou průměty výstupů modelů SegI1, SegIPTaskshift1 a SegIPDatashift1.

Velký rozdíl mezi modely SegI1 a SegIPTakshift1 je ale v segmentaci stranově neznámé kosti (tj. stranově opačná varianta kosti, než byla použita při trénování). SegI1 dokáže omezeně segmentovat prostřední část kosti, ale v oblastech kloubních hlavic selhává. SegIPTaskshift reaguje o několik řezů blíže oblasti kloubních hlavic než nepředtrénovaný model (viz obrázek 4.11) a celková odezva je mnohem silnější. Z tohoto lze usoudit, že předtrénovaný model do jisté míry zachovává schopnost reagovat i na tu stranovou variantu kosti, která nebyla do segmentačních trénovacích dat zařazena, protože již měl možnost tato data zpracovat dříve. Proč nedochází k lepšímu využití původní znalosti může být kvůli horší možnosti využití původních filtrů z důvodu vazeb mezi vrstvami neuronových sítí popsaných v práci autorů Yosinski et al. [43].

4.3 Předtrénování na zdrojových datech

Druhým možným způsobem využití transfer learningu je, kromě předtrénování na zdrojové úloze, také předtrénování na zdrojových datech, která sdílejí s cílovými daty nějakou podobnost. V tomto případě byla za zdrojová data vybrány CT skeny dlouhých kostí jiných než fibula. Lze předpokládat, že model dokáže těžit z podobnosti mezi různými dlouhými kostmi a do jisté míry i z podobnosti kostní tkáně dlouhých kostí a pánve. Protože se práce zaměřuje především na interaktivní segmentaci, byl takto předtrénován právě interaktivní model. Interaktivní model předtrénovaný na dlouhých kostech je označován jako SegIPDatashift.

U datasetu **Fibula** dosahuje právě tento model průměrně nejlepších výsledků při evaluaci s jednou uživatelskou interakcí. Model SegIPDatashift1 dosahuje evaluační Dice skóre 0.7170, což je druhý nejlepší výsledek hned po modelu SegIPTaskshift1. Numerické výsledky skupin SegIPTaskshift a SegIPDatashift jsou relativně blízké. U většiny modelů z těchto dvou skupin dosahují lepších výsledků modely ze skupiny Datashift. Je také patrné, že model předtrénovaný na dlouhých kostech dokáže segmentovat stranově neznámou fibulu ještě lépe než model předtrénovaný na jiné úloze. Obdobně jako u metody Taskshift bylo zaznamenáno, jak model reaguje na stranově neznámou kost (viz obrázek 4.11). U modelu SegIPDatashift je viditelná reakce ještě blíže koncovým oblastem kosti a ve středových částech kosti dochází i k nalezení kompletně vyplněné masky, která tvarem dobře odpovídá

groundtruth segmentaci.

Modely dosahují vysoké rychlosti konvergence. U datasetu **Fibula**, kde je rozdíl mezi zdrojovými a cílovými daty minimální, dosahují modely dobrých výsledků už po 10 000 iteracích a do 20 000 iterací lze dosáhnout výsledků, které se již velmi nemění. U datasetu **Pelvis**, kde je podobnost zdrojových a cílových dat trochu menší, jsou tyto zajímavé body dány iteracemi číslo 20 000 a 30 000. Lze tedy tvrdit, že při takovém předtrénování lze obecný model pro segmentaci dlouhých kostí velmi rychle přizpůsobit na nový typ dat, je-li tento dostatečně podobný datům zdrojovým.

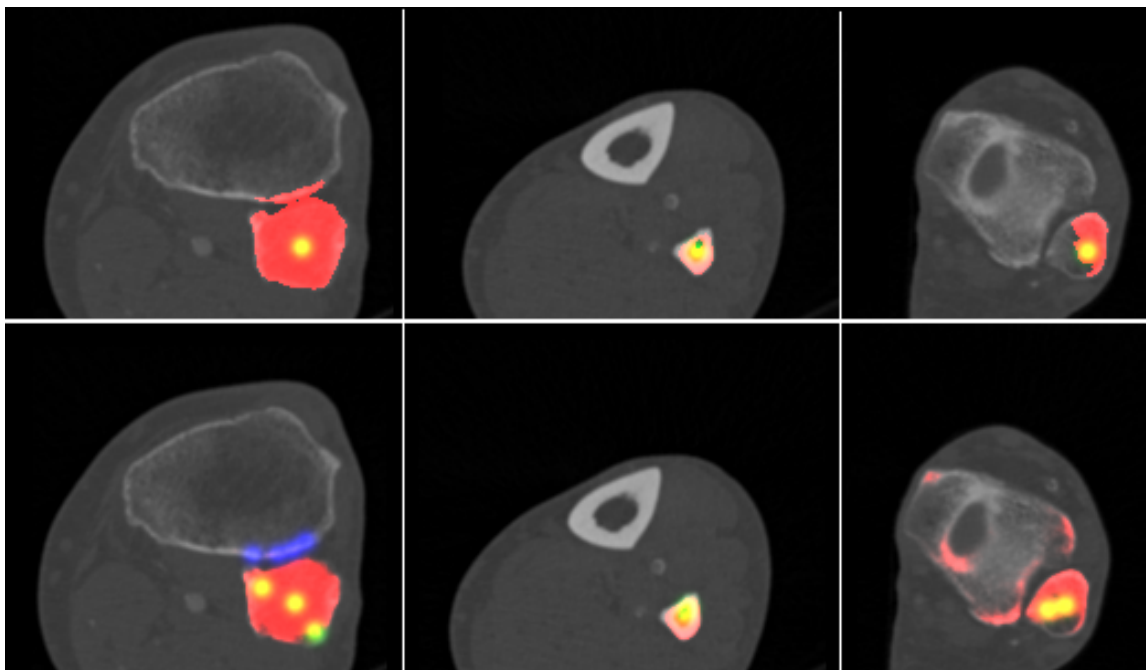
U datasetu **Pelvis** dosáhl model SegIPDatashift vůbec nejlepších výsledků při evaluaci s pěti interakcemi, a to Dice skóre 0,9905 u modelu SegIPDatashift37. Vizuálně jsou výstupy podobné jako u předchozích metod.

U modelu předtrénovaného na segmentaci dlouhých kostí bylo také manuálně otestováno, do jaké míry je možné jeho použití pro segmentaci neznámých dat bez dodatečného dotrénování na tato data. Pro dataset **Pelvis** je chování následující. U sledovaného ukázkového případu bylo možné při označení čtyř anotačních bodů objektu (pánve) na jeden řez dosáhnout Dice skóre 0,6622. Tuto hodnotu pak bylo možné dále zlepšit až na hodnotu 0,8134 při osmi anotacích objektu a čtyřech anotacích pozadí na řez. U datasetu **Fibula** byl rozsah dosaženého skóre od 0,7537 do 0,8283 při použití maximálně pěti anotací. Další anotace v tomto případě již nevedly na zásadní zlepšení výsledků. U obou datasetů lze pozorovat, že využití interakcí pro segmentaci neznámých dat je omezené (obrázek 4.12). Pokud je anotace pro segmentaci požadovaného (modelu neznámého) objektu blízko objektu, který byl použit při trénování, dochází k falešně pozitivní segmentaci známého objektu společně s objektem neznámým. Zároveň nelze dost dobře použít anotaci pozadí pro potlačení falešně pozitivní segmentace, protože reakce na tuto není dostatečně silná nebo naopak vede i ke snížení jistoty správné segmentace objektu.

4.4 Rozšířená interaktivní metoda

Posledním experimentem bylo rozšíření interaktivní metody o přidání zpětné vazby. Zpětná vazba je realizována pomocí čtvrtého vstupního kanálu, do kterého lze vložit predikci segmentace z předchozího kroku. Předpokládá se, že model se takto naučí lépe reagovat na uživatelský vstup při korekci chyb z předchozích kroků. Teoreticky lze také použít hrubou segmentaci vytvořenou nějakou jinou metodou.

Numerické výsledky rozšířené metody jsou pro modely s větším počtem trénovacích skenů srovnatelné se základním interaktivním modelem SegI. Pro evaluaci s jednou interakcí jsou výsledky dokonce horší než SegI, i když zde pořád dochází k jistému zlepšení oproti automatické baseline metodě. Horší výsledky pro první interakci jsou z toho důvodu, že model relativně silně přihlíží k hrubé predikci, která je mu předložena na vstup. Pokud provedeme několik iterací dopředného průchodu, aniž bychom přidali další uživatelské interakce, dochází ke zlepšení výsledku segmentace právě díky přidané zpětné vazbě (obrázek 4.13). Pokud vyzkoušíme předložit síti na vstup segmentační masku vytvořenou pomocí prahování, je vidět, že model zachovává do jisté míry původní podobu segmentace. Oblasti, kde je přítomna známá kost, jsou ale posíleny a naopak oblasti, kde byla prahováním vysegmen-



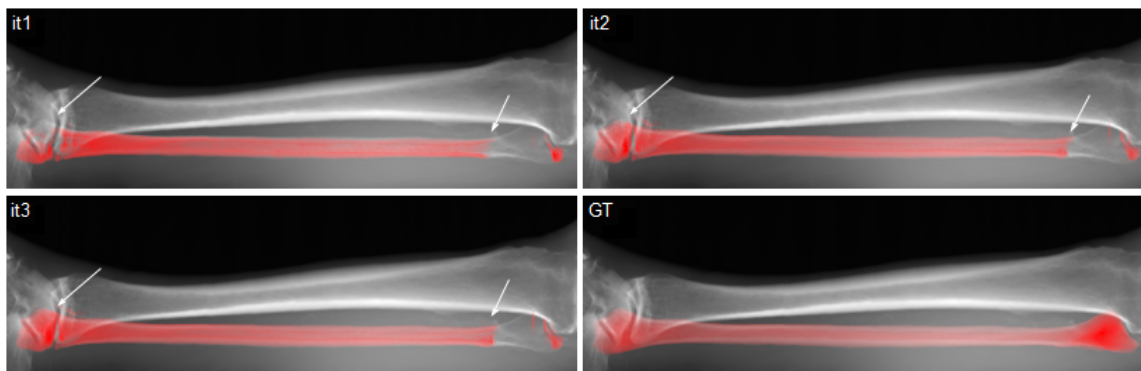
Obrázek 4.12: Pokus o segmentaci fibuly pomocí modelu trénovaném pouze na datasetu Longbones. Je vidět, že dodatečné interakce (druhý řádek) lze pro opravu segmentace neznámých dat využít jen do jisté míry a v některých případech dochází naopak ke ztrátě jistoty segmentací objektu. (Modrou barvou je zobrazena anotace pozadí, zelenou/žlutou anotace objektu.)

tována kost neznámá, jsou potlačeny. Nicméně vliv počáteční vstupní segmentace je patrný.

Konvergence je v tomto případě spíše pomalejší než u ostatních segmentačních modelů. Podle vývoje Dice skóre měřeného na validaci s jednou interakcí je dokonce možné usoudit, že by výsledky mohly být ještě zlepšeny dalšími trénovacími iteracemi, které už ale z časových důvodů nebyly provedeny.

4.5 Diskuze

U datasetu **Fibula** došlo u všech použitých rozšíření metod ke zlepšení Dice skóre. Největší problémy při segmentaci fibuly nastávají v okrajových částech kosti, kde dochází k falešně negativní segmentaci fibuly, někdy v kombinaci s falešně pozitivní segmentací sousedící kosti. Druhou obecně problematickou oblastí je vnitřek kosti podél celé její délky. Zavedením různých forem předtrénování a přidáním interakce dochází k potlačení těchto problémů. Neinteraktivní model předtrénovaný na restauraci dat má numericky lepší výsledky, vizuálně ale rozdíl není příliš patrný. Nejlepších výsledků dosahují modely SegIPTaskshift a SegIPDatashift, které využívají jak interaktivity, tak nějakého způsobu předtrénování. U datasetu **Pelvis** jsou číselné hodnoty lepší než u datasetu Fibula. U rozšířených metod stále dochází ke zlepšení, ale vzhledem k celkově dobrým výsledkům není míra zlepšení tak výrazná. Nejlepší výsledky zde má model SegIPDatahift. U obou datasetů dochází zavedením rozšíření metod také k výraznému urychlení konvergence modelů, přičemž nejrychleji



Obrázek 4.13: Interaktivní metoda rozšířená o zpětnou vazbu dosahuje zlepšení výsledků při dalších iteracích i bez přidání dalších uživatelských anotací.

se trénují modely SegIPDatahift.

Interaktivní modely pro segmentaci fibuly dosahují výrazně lepších výsledků než modely neinteraktivní, což je očekávatelné. Při manuálním zhodnocení interaktivní segmentace bylo zjištěno, že modely se naučily reagovat na uživatelský vstup a zlepšovat tím výsledky segmentace, ale pouze do jisté míry. Modely obecně nereagují na anotaci v podobě kliknutí jako na přesné označení segmentovaného regionu, ale spíše jako na přibližné označení místa, kde model hledá známý objekt. K segmentaci známého objektu tak do jisté míry dochází i v případě, je-li anotace objektu umístěna v jeho okolí. Míra reakce na anotaci je dána jistotou segmentace známého objektu v daném místě. Pokud je odezva modelu v daném místě silná, nelze potlačit segmentaci objektu ani umístěním anotací pro pozadí na toto místo. Pokud jsou odezvy slabší a jistota segmentace objektu je tedy menší, dokáže model využít dodatečné informace v podobě označení pozadí pro opravu chyby falešně pozitivní segmentace. Zdá se, že u modelů trénovaných s více daty jsou nesprávné segmentace méně jisté a model tedy dokáže lépe využít uživatelské interakce pro opravu vlastních chyb.

Fakt, že se modely naučily reagovat tak, jak je popsáno výše, komplikuje možnost využití interaktivních modelů jako jediného nástroje pro vytvoření nového datasetu. Pokud nemáme k dispozici žádné trénovací anotace pro nová data, je možnost jejich vytvoření interaktivním modelem trénovaným na zdrojová data omezená. I interaktivní modely hledají ve snímcích známý objekt a segmentace nového objektu nelze příliš dobře vynutit. Řešením by mohlo být natrénování modelu, který spíše než konkrétní kost, bude schopen segmentovat uzavřené kosterní regiony v závislosti spíše na uživatelské interakci než na znalosti konkrétního anatomického útvaru. Způsobem, jak tohoto dosáhnout, by mohlo být trénování na smíšeném datasetu, kde by groundtruth segmentace zahrnovala všechny kosterní regiony v daném řezu, ale při trénování by byl vždy označen pouze jeden uzavřený region. Otázky, jak velký by musel být takovýto dataset pro natrénování dobrého obecného modelu pro segmentaci kostí a zda by takovýto postup byl výhodný z hlediska úsilí vynaloženého k jeho vytvoření, jsou vhodné k dalšímu zamyšlení.

I přes výše zmíněné je ale stále možné použití interaktivních modelů pro tvorbu nového datasetu v kombinaci s jiným (interaktivním či manuálním) nástrojem, který by byl použit ke zlepšení výstupů neuronové sítě. Pokud bychom skutečně chtěli pomocí výše popsaných

metod vytvořit nový dataset, nejlepší volbou by bylo předtrénovat interaktivní model na zdrojových datech (metoda datashift). Tato metoda je vhodná, protože dosahuje poměrně dobrých výsledků a zároveň dotrénování na nová data konverguje rychleji než ostatní metody. Bylo by nutné vytvořit manuálně či jiným segmentačním nástrojem počáteční počet anotací a s těmi natrénovat první verzi modelu. Vhodný počet počátečních anotací se liší v závislosti na konkrétních datech. Pro dataset Fibula lze první datashift model natrénovat s již dvěma 3D CT skeny fibuly. Takový slabý model dosahuje průměrného Dice skóre 0,951 a dokáže již poskytnout anotace, u kterých by manuální úprava nebyla výrazně časově náročná. Úprava by rovněž mohla být provedena s pomocí jiného poloautomatického segmentačního algoritmu. Pokud bychom neměli k dispozici zdrojový dataset pro natrénování datashift modelu, druhou nejlepší volbou by bylo použití interaktivního modelu předtrénovaného na restauraci dat, tedy metodu taskshift.

Zajímavé by také bylo provedení dalších experimentů s metodou SegIPlus. Metodu by bylo možné vyzkoušet trénovat jako prostředek pro zlepšení hrubé segmentace poskytnuté jiným segmentačním algoritmem. Je možné, že síť se při zpětnovazebním trénování s omezeným datasetem učí příliš spoléhat na vlastní výsledky na trénovacích datech a reakce na vstupy při evaluaci či na predikci poskytnutou jiným algoritmem není tak efektivní. Právě použití predikcí vytvořených jinými algoritmy, či algoritmy, již při trénování by mohlo vést na lepší doplnění či úpravu hrubé predikce sítí s využitím uživatelských interakcí.

Je možné, že k praktické použitelnosti zkoumaných metod by mohly dopomoci také některé dále zmíněné úpravy. Podobně jako v práci [46], lze použít místo segmentace 2D výřezů segmentaci ve 3D s využitím 3D konvolucí. Zajímavou možností by bylo také provést předtrénování úlohy restaurace na mnohem větším datasetu. U interaktivní segmentace by možná docházelo k lepší interaktivitě, omezila-li by se velikost chybových oblastí, do kterých se při trénování ještě umísťují interakce. U současného způsobu trénování jsou generovány další interakce i v případě chyb o rozsahu jednotek pixelů, což může vést k potlačení odezvy na anotaci pozadí. (Výsledný model často nereaguje na anotaci pozadí dostatečně, je-li tato během trénování umístována v těsné blízkosti segmentovaného objektu.) Zvýšení počtu maximálních interakcí v průběhu trénování je rovněž možné.

V rámci vedlejších experimentů bylo taky vyzkoušeno použití Dice loss jako chybové funkce pro segmentaci, přičemž takto dosažené výsledky byly lepší než s chybovou funkcí v podobě Cross Entropy. Do budoucna by tedy bylo vhodnější používat Dice loss.

Kapitola 5

Závěr

Bylo natrénováno velké množství modelů pro důkladné zhodnocení použitelnosti vybraných metod pro segmentaci CT dat. Práce se soustředila zejména na využití uživatelské interakce jako nástroje pro zlepšení kvality segmentace a tímto sekundárně i jako nástroje pro redukci minimálního množství dat nutného k natrénování v praxi použitelného segmentačního modelu. Za účelem úspěšného trénování segmentačních modelů s malými datasety byly vybrány dva přístupy z oblasti transfer learningu, a to předtrénování na zdrojové úloze restaurace obrazu a na zdrojových datech v podobě anatomických objektů odlišných od dat cílových.

Konkrétně se jednalo o segmentaci lidských kostí a to kosti lýtkové (fibula) a pánevní (pelvis). Za účelem natrénování modelu byly připraveny tři různé datasety a to dva datasety pro cílovou segmentační úlohu, tedy dataset Fibula a dataset Pelvis, a jeden dataset pro předtrénování na cílových datech, označovaný jako Longbones. Dataset Longbones obsahoval směs několika různých dlouhých kostí.

Pro důkladné ověření vlivu vybraných rozšíření na kvalitu výsledných segmentací při různém množství trénovacích dat byl navržen benchmarkovací postup v podobě vytvoření a následné optimalizace skupiny modelů, kde každý model byl trénován s jiným množstvím trénovacích dat. Tyto skupiny byly natrénovány pro jednotlivé metody a následně bylo zhodnoceno, jak si jednotlivé metody vedly. Srovnání bylo provedeno z hlediska rychlosti konvergence trénování a kvality výsledných segmentací. Kvalita segmentací byla ohodnocena metrikou Dice skóre a blíže vizuálně zhodnocena jak u výstupů automatické evaluace, tak s využitím jednoduchého nástroje pro interaktivní segmentaci, který byl v rámci práce implementován.

Nejlepších výsledků dosáhly interaktivní varianty metod datashift a taskshift, tedy interaktivní model předtrénovaný na zdrojových datech či zdrojové úloze, přičemž u metody datashift dochází zároveň k výraznému urychlení konvergence při dotrénování na cílová data. Metoda SegIPDatashift dosahuje Dice skóre 0,951 již s dvěma trénovacími 3D CT skeny. Praktické využití těchto metod může být například v usnadnění tvorby nového datasetu, kdy lze natrénovat model použitelný pro vytvoření nových anotací již s počtem trénovacích 3D skenů v řádu jednotek.

Takovéto použití modelů je do jisté míry omezeno chováním modelů natrénovaných na velmi malém počtu dat (například pouze jeden 3D sken.). Interaktivní modely se naučily

reagovat na uživatelský vstup, avšak závislost na dobré znalosti segmentovaného objektu je patrná. Vzhledem k tomu, že metody nebyly plně vyladěny, je možné, že dalšími úpravami, jako například zavedením regularizací, by došlo ke zlepšení výsledků i s velmi malým počtem trénovacích dat, což by vedlo k lepší aplikovatelnosti metod v praxi.

Tato práce vznikla za podpory projektu LM2018140 e-Infrastruktura CZ (e-INFRA CZ) poskytnutého v rámci programu Velké infrastruktury pro výzkum, vývoj a inovace.

Literatura

- [1] AGUSTSSON, E., UIJLINGS, J. R. R. a FERRARI, V. Interactive Full Image Segmentation by Considering All Regions Jointly. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, s. 11614–11623.
- [2] AMREHN, M., GAUBE, S., UNBERATH, M., SCHEBESCH, F., HORZ, T. et al. UI-Net: Interactive Artificial Neural Networks for Iterative Image Segmentation Based on a User Model. In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. Září 2017.
- [3] ARRUDA, H., BENATTI, A., COMIN, C. a F. COSTA, L. da. *Learning Deep Learning (CDT-15)*. Září 2019.
- [4] BADRINARAYANAN, V., KENDALL, A. a CIPOLLA, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017, roč. 39, č. 12, s. 2481–2495.
- [5] BALDI, P. Autoencoders, Unsupervised Learning, and Deep Architectures. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Bellevue, Washington, USA: PMLR. 02 Jul 2012, roč. 27, s. 37–49.
- [6] BISHOP, C. M. *Pattern recognition and machine learning*. New York: Springer, 2006. ISBN 0-387-31073-8.
- [7] BOYKOV, Y. Y. a JOLLY, M. . Interactive graph cuts for optimal boundary region segmentation of objects in N-D images. In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. 2001, s. 105–112 vol.1.
- [8] BREDELL, G., TANNER, C. a KONUKOGLU, E. Iterative Interaction Training for Segmentation Editing Networks. In: *MLMI@MICCAI*. 2018.
- [9] CAN, Y. B., CHAITANYA, K., MUSTAFA, B., KOCH, L. M., KONUKOGLU, E. et al. Learning to Segment Medical Images with Scribble-Supervision Alone. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer International Publishing, 2018, s. 236–244. ISBN 978-3-030-00889-5.
- [10] D. FORSYTH, J. P. *Computer Vision: A Modern Approach*. 2. vyd. Boston: Pearson, 2012. ISBN 978-0136085928.
- [11] GRADY, L. Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006, roč. 28, č. 11, s. 1768–1783.

- [12] HANA ŘEZANKOVÁ, V. S. *Shluková analýza dat*. 2. vyd. Praha: Professional Publishing, 2009. ISBN 978-80-86946-81-8.
- [13] HE, K., ZHANG, X., REN, S. a SUN, J. Identity Mappings in Deep Residual Networks. *Computer Vision – ECCV 2016*. Říjen 2016, roč. 9908, s. 630–645.
- [14] KAUR, R., KAUR, L. a GUPTA, S. Enhanced K-Mean Clustering Algorithm for Liver Image Segmentation to Extract Cyst Region. *IJCA Special Issue on Novel Aspects of Digital Imaging Applications (DIA)*. 2011, č. 1, s. 59–66.
- [15] KINGMA, D. P. a BA, J. Adam: A Method for Stochastic Optimization. *CoRR*. 2014, abs/1412.6980.
- [16] KRIZHEVSKY, A., SUTSKEVER, I. a HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc. 2012, s. 1097–1105.
- [17] MAHADEVAN, S., VOIGTLAENDER, P. a LEIBE, B. Iteratively Trained Interactive Segmentation. In: *BMVC*. 2018.
- [18] MAHMOOD, F., CHEN, R. a DURR, N. J. Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training. *IEEE Transactions on Medical Imaging*. 2018, roč. 37, č. 12, s. 2572–2581.
- [19] MANINIS, K.-K., CAELLES, S., PONT TUSET, J. a GOOL, L. V. Deep Extreme Cut: From Extreme Points to Object Segmentation. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2017, s. 616–625.
- [20] MICHAL ŠPANĚL, V. B. *Obrazové segmentační techniky* [online]. [cit. 2019-12-22]. Dostupné z: <http://www.fit.vutbr.cz/~spanel/segmentace/>.
- [21] MOK, T. C. W. a CHUNG, A. C. S. Learning Data Augmentation for Brain Tumor Segmentation with Coarse-to-Fine Generative Adversarial Networks. *CoRR*. 2018, abs/1805.11291.
- [22] NALEPA, J., MARCINKIEWICZ, M. a KAWULOK, M. Data Augmentation for Brain-Tumor Segmentation: A Review. *Frontiers in Computational Neuroscience*. Prosinec 2019, roč. 13.
- [23] NAMEIRAKPAM, D., SINGH, K. a CHANU, Y. Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*. Prosinec 2015, roč. 54, s. 764–771.
- [24] NOH, H., HONG, S. a HAN, B. Learning Deconvolution Network for Semantic Segmentation. In: *Computer Vision (ICCV), 2015 IEEE International Conference on*. 2015.
- [25] PACE, D. F., DALCA, A. V., BROSCHE, T., GEVA, T., POWELL, A. J. et al. Iterative Segmentation from Limited Training Data: Applications to Congenital Heart Disease. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support*. 2018, roč. 11045, s. 334–342.

- [26] PAN, S. J. a YANG, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*. Oct 2010, roč. 22, č. 10, s. 1345–1359. ISSN 2326-3865.
- [27] PATHAK, D., KRAHENBUHL, P., DONAHUE, J., DARRELL, T. a EFROS, A. A. Context Encoders: Feature Learning by Inpainting. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, s. 2536–2544.
- [28] REZAEI, M., HARMUTH, K., GIERKE, W., KELLERMEIER, T., FISCHER, M. et al. Conditional Adversarial Network for Semantic Segmentation of Brain Tumor. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing, 2017, s. 241–252. ISBN 978-3-319-75238-9.
- [29] RONNEBERGER, O., FISCHER, P. a BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015, s. 234–241. ISBN 978-3-319-24574-4.
- [30] ROTH, H., ZHANG, L., YANG, D., MILLETARI, F., XU, Z. et al. Weakly Supervised Segmentation from Extreme Points. *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*. Springer International Publishing. 2019, s. 42–50. ISSN 1611-3349.
- [31] SAKELLAROPOULOS, G., KISTLER, M., BONARETTI, S., PFAHRER, M., NIKLAUS, R. et al. The virtual skeleton database: an open access repository for biomedical research and collaboration. *Journal of Medical Internet Research*. 2013, roč. 15.
- [32] SAKINIS, T., MILLETARI, F., ROTH, H., KORFIATIS, P., KOSTANDY, P. et al. Interactive segmentation of medical images through fully convolutional neural networks. *ArXiv*. 2019.
- [33] SANDFORT, V., YAN, K., PICKHARDT, P. a SUMMERS, R. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports*. Prosinec 2019, roč. 9.
- [34] SCHROFF, F., CRIMINISI, A. a ZISSERMAN, A. Object Class Segmentation using Random Forests. In: *BMVC*. Leden 2008.
- [35] SHAIMAA ELMORSY, M. A. A Region Growing Liver Segmentation Method with Advanced Morphological Enhancement. *2015 32nd National Radio Science Conference (NRSC)*. Březen 2015.
- [36] SHELHAMER, E., LONG, J. a DARRELL, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Květen 2016, roč. 39, s. 1–1.
- [37] SIMONYAN, K. a ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*. Zář 2015, abs/1409.1556.
- [38] SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S. et al. Going deeper with convolutions. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Červen 2015, s. 1–9.

- [39] TREMBLAY, J., PRAKASH, A., ACUNA, D., BROPHY, M., JAMPANI, V. et al. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2018, s. 1082–10828.
- [40] WANG, G., LI, W., ZULUAGA, M. A., PRATT, R., PATEL, P. A. et al. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*. 2018, roč. 37, č. 7, s. 1562–1573.
- [41] WARD, D., MOGHADAM, P. a HUDSON, N. Deep Leaf Segmentation Using Synthetic Data. In: *BMVC*. červenec 2018.
- [42] XU, N., PRICE, B., COHEN, S., YANG, J. a HUANG, T. Deep Interactive Object Selection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, s. 373–381.
- [43] YOSINSKI, J., CLUNE, J., BENGIO, Y. a LIPSON, H. How transferable are features in deep neural networks? *ArXiv*. 2014.
- [44] ZEILER, M. D. a FERGUS, R. Visualizing and Understanding Convolutional Networks. In: *Computer Vision – ECCV 2014*. Springer International Publishing, 2014, s. 818–833. ISBN 978-3-319-10590-1.
- [45] ZHOU, Z., SODHA, V., RAHMAN SIDDIQUEE, M. M., FENG, R., TAJBAKSH, N. et al. *Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis*. Říjen 2019.
- [46] ÇIÇEK Özgün, ABDULKADIR, A., LIENKAMP, S. S., BROX, T. a RONNEBERGER, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *ArXiv*. 2016.

Příloha A

Numerické výsledky metod

Zde jsou uvedeny výsledky evaluace vybraných modelů. Hodnoty odpovídají průměrnému Dice skóre dosaženému na evaluačních datech. Sloupec **Počet skenů** udává, kolik bylo použito trénovacích 3D skenů pro natrénování modelů, jejichž výsledky jsou uvedeny na daném řádku.

Tabulka A.1: Výsledky vybraných modelů pro datasetu Fibula. Interaktivní metody byly vyhodnoceny s jednou uživatelskou interakcí na řez.

Počet skenů	Seg	SegPTaskshift	SegI	SegIPTaskshift	SegIPDatashift	SegIPlus
1	0.348	0.392	0.588	0.722	0.717	0.422
2	0.749	0.830	0.929	0.932	0.951	0.902
3	0.795	0.855	0.949	0.958	0.962	0.948
5	0.819	0.858	0.943	0.963	0.967	0.952
10	0.831	0.904	0.948	0.972	0.970	0.958
30	0.974	0.977	0.977	0.979	0.980	0.978
70	0.981	0.984	0.983	0.984	0.982	0.978

Tabulka A.2: Tabulka vyhodnocení interaktivních modelů na datasetu Fibula za přítomnosti až pěti uživatelských interakcí na jeden řez.

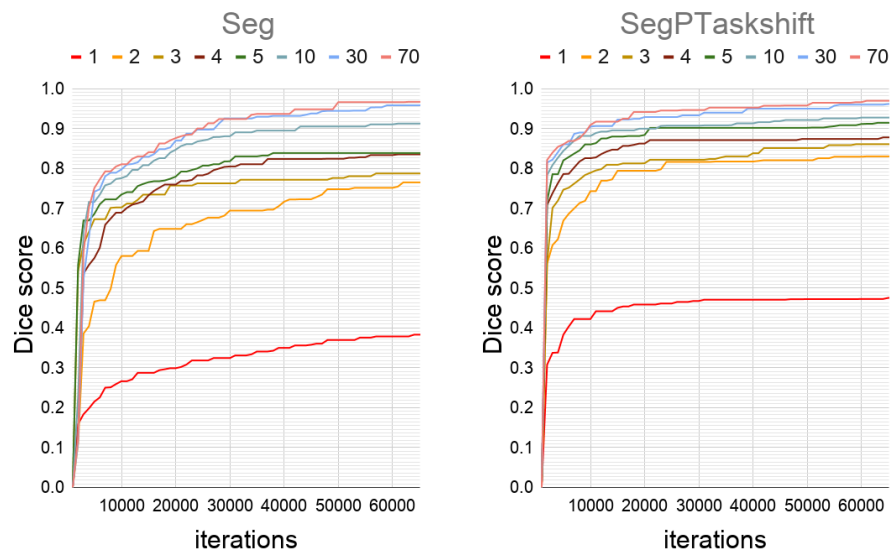
Počet skenů	SegI	SegIPTaskshift	SegIPDatashift	SegIPlus
1	0.554	0.734	0.751	0.460
2	0.932	0.946	0.961	0.954
3	0.959	0.970	0.969	0.966
5	0.963	0.971	0.973	0.973
10	0.975	0.978	0.976	0.973
30	0.983	0.983	0.983	0.983
70	0.985	0.986	0.986	0.985

Tabulka A.3: Tabulka vyhodnocení interaktivních modelů na datasetu Pelvis za přítomnosti až pěti uživatelských interakcí na jeden řez.

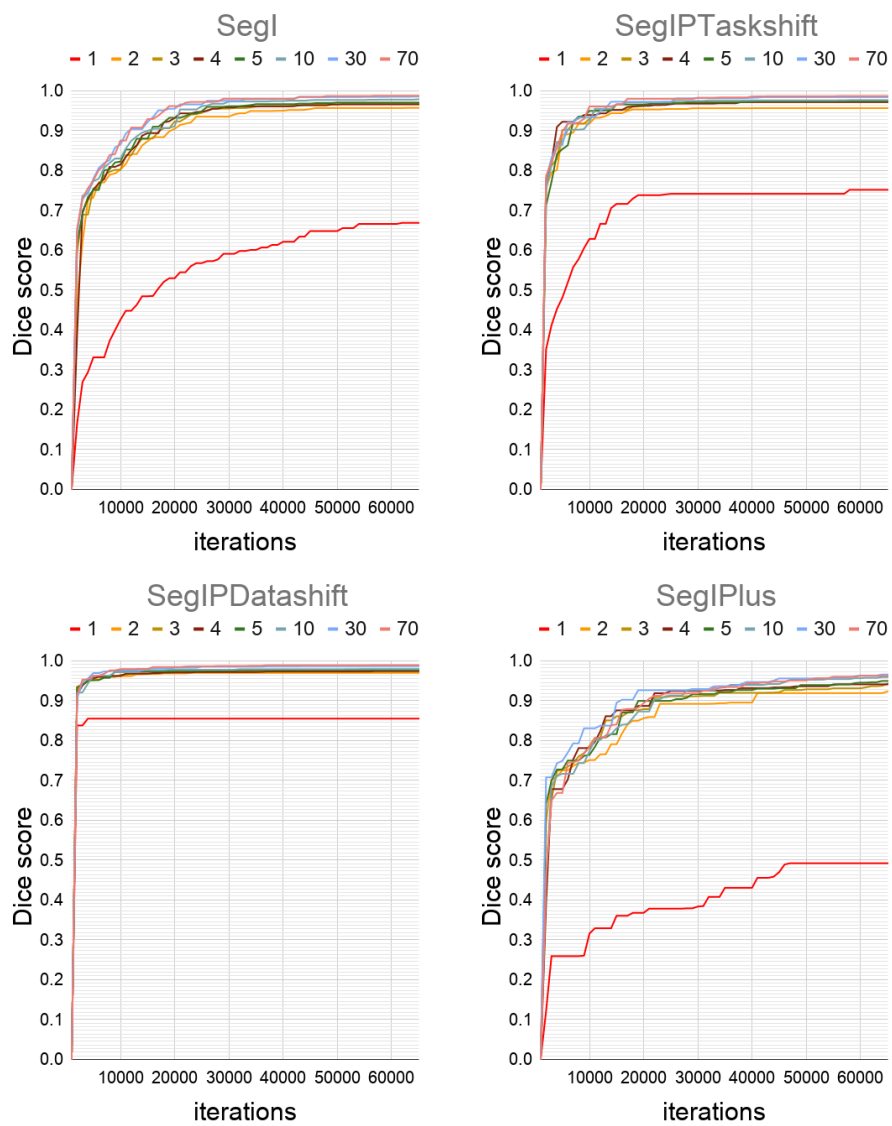
Počet skenů	Seg	SegI	SegIPTaskshift	SegIPDatashift
1	0.927	0.972	0.975	0.979
2	0.971	0.983	0.984	0.985
10	0.987	0.987	0.989	0.990
30	0.988	0.989	0.990	0.991
37	0.989	0.989	0.988	0.991

Příloha B

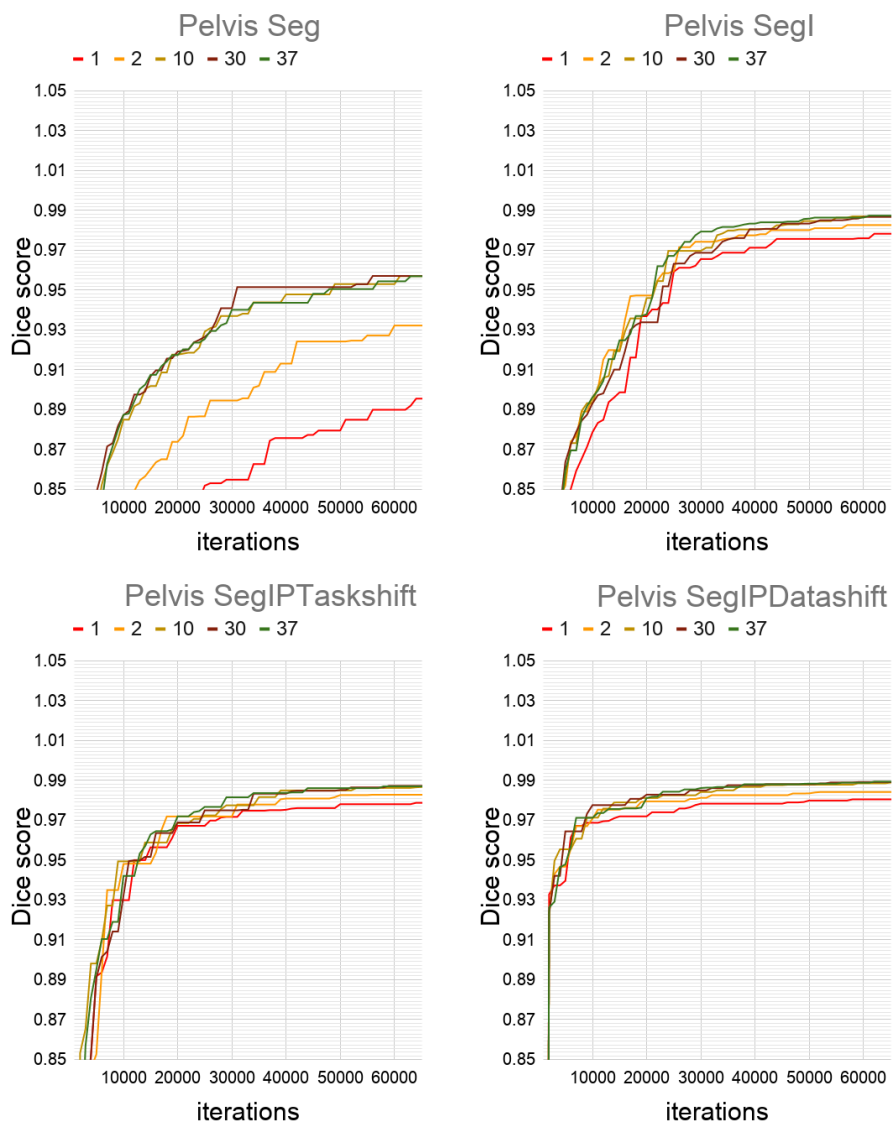
Vývoj Dice skóre při trénování



Obrázek B.1: Vývoj Dice skóre během trénování automatických modelů na datasetu Fibula. Baseline (Seg) a model předtrénovaný na restauraci (SegPTaskshift). Zaznamenána je vždy nejlepší dosažená hodnota.



Obrázek B.2: Vývoj Dice skóre během trénování interaktivních modelů na datasetu Fibula. Zaznamenána je vždy nejlepší dosažená hodnota.



Obrázek B.3: Vývoj Dice skóre během trénování modelů na datasetu Pelvis. Zaznamenána je vždy nejlepší dosažená hodnota. (Zobrazen je pouze specifický rozsah Dice skóre.)

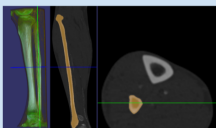
Příloha C

Plakát

Interactive 3D CT Data Segmentation Based on Deep Learning

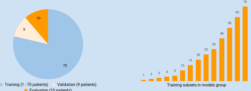
author: Kateřina Trávníčková, supervisor: Oldřich Kodym

What is it about?
Obtaining the sufficient amount of training data for deep learning based segmentation might be an issue, especially in the medical field. In this work, we focus on the task of longitudinal bone segmentation in human body CT scans. We designed three groups of segmentation models with varying amount of training data to examine the impact of two possible means of battling the lack of data problem.




Dataset and the training subsets

- Benchmarking of the segmentation methods on Fibula bone dataset.
- Groups of models for each method, each containing 14 models which are trained on variably sized subsets.




Baseline models group (Seg)

- Training pair consisting of 2D input patch and corresponding manual segmentation patch.
- U-net CNN architecture (shared among all 3 groups).
- Automatic approach with no other extensions.




Unsupervised pretraining (SegPretrained)

- Inspired by Models Genesis [1].
- Pretrain without manual annotations on image restoration.
- Finetune to the segmentation task.
- Improvement from **0.749** (baseline model) to **0.830** Dice score when trained on two patients.



User interaction (SegInter)

- Based on interactive segmentation model [2].
- Object and background clicks (2 additional channels).
- Better object localization than the baseline.
- Promising results - Dice score **0.929** with only two training patients.



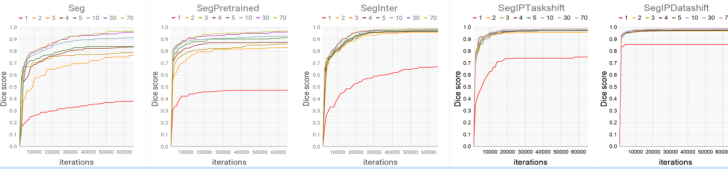
Interaction and pretraining (SegIPTaskshift & SegIPDatashift)

- Interactive approach in combination with two different types of pretraining.
- SegIPTaskshift is an interactive model pretrained on data restoration task.
- SegIPDatashift is pretrained on different CT dataset with various types of long human bones.
- The table of average Dice score on evaluation data shows how these two methods generally reach the best results.

# scans	Seg	SegIPTaskshift	Seg	SegIPTaskshift	SegIPDatashift
1	0.348	0.392	0.588	0.722	0.717
2	0.749	0.830	0.929	0.932	0.951
3	0.795	0.855	0.949	0.958	0.962
5	0.819	0.858	0.943	0.961	0.967
10	0.831	0.901	0.948	0.972	0.970
30	0.974	0.977	0.977	0.979	0.980
70	0.981	0.984	0.983	0.984	0.982

Comparison of Dice score during the training of different models

- Dice score is measured on the validation data during the training.
- In **Seg** group we can clearly see the tendency of quality loss with small amount of data.
- SegPretrained** group is slightly better than the baseline, most visibly on the 1 patient model. There is also a visible speed up in the convergence.
- The results of the **SegInter** group are promising. This way we can obtain considerably good segmentation quality with as few as two training patients.
- The best results are achieved by the **SegIPDatashift** and **SegIPTaskshift** models. There is also very significant improvement of the training convergence.
- Practical use** of the interactive method (possibly combined with pretraining) lies in speeding up the creation of new segmentation datasets. We are able to update the training set with new data created by the model with human expert assistance. This can be done iteratively until the required segmentation quality is reached.



[1] SHEN, Zhenqi, Xiangyang DING, and Mubai MA. BAYESIAN DEEP PRIORS. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 509–518. doi:10.1109/CVPR.2017.101. [2] HADJIIOANNIDIS, Konstantinos, and Demetrius Tsetos. Interactive segmentation of medical images through fully convolutional neural networks. CoRR, 2015, arXiv:1503.02625.

Obrázek C.1: Plakát shrnující tuto diplomovou práci.