



# Posudek oponenta závěrečné práce

**Oponent práce:** Ing. Magda Friedjungová, Ph.D.  
**Student:** Bc. Mark Sobolev  
**Název práce:** Srovnání Generativních adversariálních sítí a využití na medicínských datech  
**Obor / specializace:** Znalostní inženýrství  
**Vytvořeno dne:** 8. června 2021

## Hodnotící kritéria

### 1. Splnění zadání

[1] zadání splněno

► [2] zadání splněno s menšími výhradami

[3] zadání splněno s většími výhradami

[4] zadání nesplněno

Ke splnění zadání mám menší výhrady: Poslední bod dle mého nebyl splněn - student nezveřejnil svůj kód, resp. odkaz na něj se nevyskytuje v práci. Dále mi v práci chybí diskuze a srovnání architektur je neúplné.

### 2. Písemná část práce

50/100 (E)

Shrnutí:

Práci nepovažuji za přehlednou, některé sekce teoretické části jsou informačně nevyvážené a ne vždy navazují na praktickou část. Některé sekce jsou téměř prázdné, stejně tak některé tabulky působí nedokončeně. Praktická část je nedostatečně popsána, resp. není jasný scénář experimentů, text nedodrhuje logickou posloupnost a informace jsou podávány neuspořádaně. V práci je pro mě složité se orientovat. Dle mého student klade důraz na věci, které pro práci nejsou stěžejní (např. vyčerpávající popis VGG16, což je jen opisem již napsaného) a naopak stěžejním věcem nevěnuje dostatečný prostor. Přestože student provedl několik experimentů na několika veřejně dostupných datasetech, jejich popis je zmatečný a není mi tak jasné, zda student dané úloze správně porozuměl.

Odůvodnění:

Kapitola 1 je pouze na 1 stranu a obsahuje nepřesné, ne-li nepravdivé informace (U "unsupervised": "The goal is to divide the data into clusters."). Další popisy jsou nejednoznačné, chybí definice úlohy a explicitní vymezení se na obrazová data (je to podáno, jako by GANy s ničím jiným pracovat nemohly). Kapitola 2 se zabývá popisem různých GANů, nehledě na zadání práce - augmentace medicínských dat. Student

nezmiňuje GANy, které byly vytvořeny vyloženě pro "data augmentation" (rozumím, že na to lze použít téměř jakýkoliv GAN, avšak v práci není zmíněna motivace daného výběru), stejně tak mi v rešerši chybí rozbor článků, které se věnují primárně medicínským datům (student však pět článků cituje, přitom by si zasloužily větší pozornost, např. [13]). Nerozumím, proč je rozebrán Pix2Pix, ale není už ve stejném detailu popsán U-Net (přitom je v kapitole 4 používán) a PatchGAN. V sekci 2.6 je zmíněn "multi-scale structural similarity index measure", jeho popis však chybí. Kapitola 3 je popsána poměrně zmatečně. U datasetů by se dalo explicitně zmínit, zda mají labels. Dále bych čekala samostatnou sekci s popisem datasetů, a pak až výsledky použitých modelů. Co se týče čísel v textu, hodila by se pro přehlednost tabulka - dělení dat, počty dat po augmentaci apod. Pro výsledky modelu v sekci 3.1.1 CycleGAN je poskytnuta tabulka s výsledky, ale pro sekci 3.1.2 FastGAN už není (výsledky jsou alespoň v textu). Pro dataset 3.1 jsou zmíněné použité modely a výsledky, ale pro dataset 3.2 nic takového není a je zde jen popis datasetu a "trénování" FastGANu - vše je pak v tabulce 3.2. Dále je použit pouze FastGAN, ale chybí zde argumentace, proč není použit i CycleGAN. V sekci 3.4 je následující argumentace: "It is decided to use 1000 generated images and segmentation masks to augment the dataset due to their good visual quality." U modelů v sekci 3.1 ale argumentaci pro počet vygenerovaných dat nevidím a v 3.2 a 3.3 se počty vygenerovaných dat vůbec nevyskytují. V 3.1 chybí ukázky "real" a "fake" obrázků jako je tomu u ostatních datasetů. Závěr práce začíná odstavcem: "In this work is the goal to evaluate capabilities of a newly proposed generative adversarial network architecture and to compare it with selected GAN models currently used for medical visual data augmentation." což se podle mě v práci nestalo. Student zvolil několik GANů, kdy ani jeden nebyl původně určen pro augmentaci dat, a nezmiňuje GANy, které by se konkrétně používali pro generování medicínských dat - chybí reference. V práci chybí argumentace pro výběr CycleGANu pro augmentaci medicínských dat, přitom mezi top výsledky v této oblasti patří článek Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks, 2019 (<https://www.nature.com/articles/s41598-019-52737-x.pdf>). Takových článků existuje poměrně dost, neboť medicínská data neustále zápasí s málo záznamy pro trénování. Takže nerozumím, proč se student nevěnoval této oblasti v rešerši. V sekci 4.1 je 3x model CycleGANu, ale pouze 1x FastGAN, přičemž v sekci 3.1.2 byly natrénovány dva FastGANy. Článek zmíněný u 4.2 a na který experiment navazuje, se měl vyskytnout v rešerši. V tabulce 4.3 chybí výsledky pro "real datasets" - možná by taky v popisku tabulky měl být zmíněn název datasetu. V kapitole 4 pak dochází k záměně testovací, validační a trénovací sady (např. sekce 4.3.2). V sekci 4.3.3 není jasné, co student plánuje dělat. Podsekce se jmenuje "Data augmentation", což je hlavní náplní práce a nic zde není (str. 28). V sekci 4.3.3.1 bod 4 je prázdný, vůbec formátování této sekce je podivné. Chybí popis generátoru. V 4.3.4. "Classification accuracy on synthesized abnormal x-rays is 0.02" je zřejmě chyba. Text 4.2 je shodný s textem 4.1 (první dva odstavce). Tabulka 4.1 je nezajímavá vzhledem k obsahu. Tabulka 4.4 s popisem nedává smysl, to je na reálných a syntetických datech dohromady? Odkazy na tabulky 4.4 a 4.5 chybí v textu a jejich popisky jsou zavádějící/nejsou samovysvětlující. Tabulka 4.6 uvádí šest modelů (1-6, a zrovna tohle by stačilo napsat slovy, bez tabulky), tabulka 4.7 uvádí modelů pět a to 1-4 a 7. Opět na ně není z textu odkazováno. V sekci 4.4.4 se objevuje nová metrika (validation mean IoU), která není definována a ani neobsahuje referenci. Není jasné, co je Aug4 v tabulce 4.9. V 4.5.3 se poprvé vyskytuje "is trained without random jitter", což není nikde v práci vysvětleno ani zde není reference. Tabulka 4.10 neobsahuje výsledky. Závěr práce začíná odstavcem: "In this work is the goal to evaluate capabilities of a newly proposed generative adversarial network architecture and to compare it with selected GAN models currently used for medical visual data augmentation." což se podle mě v práci nestalo.

Student zvolil několik GANů, kdy ani jeden nebyl původně určen pro augmentaci dat, a nezmiňuje GANy, které by se konkrétně používali pro generování medicínských dat - chybí reference.

Práce dále obsahuje několik překlepů a chyb, přetékané tabulky, nekvalitní obrázky, chybná matematická značení a vzorce (např. 2.1), v některých tabulkách chybí data apod.

### 3. Nepísemná část, přílohy

65 /100 (D)

Student v systému KOS odevzdal tři pomocné skripty a zbývající mi nasdílel - OK. V zadání práce slovo "publish" vnímám jako veřejný repozitář (např. GitHub, GitLab) s readme souborem, jak daný obsah spustit a experimenty reprodukovat. Tento popis se nenachází ani v písemné části práce. Odevzdané notebooky z Google Colab nejsou na první dobrou spustitelné (chybí drive.mount a cesta k souborům, které jsou na disku studenta), ale nic nestandardního jsem v kódu neobjevila. Přestože nejsou kódy komentované, zasvěcený tuší, co student provádí.

### 4. Hodnocení výsledků, jejich využitelnost

55 /100 (E)

Hlavním problémem práce je, že student jasně nedefinuje co a proč bude dělat. Dle mého se student snažil pokrýt příliš široké téma, a tak jednotlivé části nejsou odpovídajícím způsobem zpracovány ani prezentovány. Z textu není jasné, zda student dané problematice porozuměl. Student avizoval, že bude srovnáno více modelů, to se však stalo pouze v sekci 4.1 (FastGAN vs. CycleGAN). Práce působí nedokončeným dojmem, chybí mi srovnání a diskuze výsledků, takže i jejich posouzení je pro mě nejasné.

### Celkové hodnocení

55 /100 (E)

Práci navrhuji hodnotit stupněm E na základě výše zmíněných připomínek.

### Otázky k obhajobě

Byla bych ráda, kdyby student připravil přehlednou prezentaci, kde stručně vysvětlí návrh jednotlivých experimentů a provede stručnou diskuzi dosažených výsledků včetně srovnání architektur.

## **Instrukce**

### **Splnění zadání**

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

### **Písemná část práce**

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

### **Nepísemná část, přílohy**

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

### **Hodnocení výsledků, jejich využitelnost**

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

### **Celkové hodnocení**

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.