

Implementácia metódy čiastočne riadeného učenia Uni-Match do metódy Frame Field Learning pre úlohu extrakcie budov z leteckých snímok

Patrik Sabol¹, Bibiána Lajčinová²

¹Geodeticca Vision s.r.o., Floriánska 19, 044 01 Košice, Slovenská republika

²Národné Superpočítačové Centrum, Dúbravská cesta 3484/9, 84104 Bratislava-Karlová Ves, Slovenská republika

Úvod

Extrakcia budov v Geografických informačných systémoch (GIS) je kľúčová pre urbanistické plánovanie, environmentálne štúdie a riadenie infraštruktúry, pretože umožňuje presné mapovanie stavieb, vrátane odhaľovania nelegálnych stavieb za účelom dodržiavania právnych predpisov, alebo efektívnejšieho vyberania daní. Integrácia extrahovaných údajov o budovách s inými geopriestorovými vrstvami zlepšuje pochopenie dynamiky miest a priestorových vzťahov. Vzhľadom na rozsah a zložitosť týchto úloh rastie potreba automatizovať extrakciu budov pomocou techník hlbokého učenia, ktoré ponúkajú vyššiu presnosť a efektívnosť pri spracovaní veľkých geopriestorových dát.

V súčasnosti, väčšina najmodernejších segmentačných modelov hlbokého učenia poskytuje výstup iba v rastrovej forme. Avšak GIS často potrebujú dáta vo vektorovej forme. Jednou z metód, ktorá dokáže generovať dáta vo vektorovej forme, je Frame Field learning. Táto metóda generuje okrem segmentačnej masky aj frame field pole, ktoré obsahuje štrukturálne informácie o objektoch, ktoré sa následne využívajú v procese vektorizácie.

Modely Frame Field learningu sú tréňované metódou “s učiteľom” (z angl. “supervised learning”), ktorá potrebuje veľké množstvo anotovaných dát. Na získanie takéhoto množstva kvalitných dát je potrebná manuálna ľudská práca, ktorá však môže byť zdĺhavá a nákladná. Jednou z metód, ktorá môže znížiť závislosť od anotovaných dát, je “učenie s čiastočným učiteľom”, resp. “čiastočne riadené učenie” (z angl. “semi-supervised learning”). Tento prístup učenia využíva nielen anotované dáta, ale aj množinu neanotovaných dát.

Cieľom tejto spolupráce medzi Národným kompetenčným centrom pre HPC a Geodeticca Vision s.r.o. bolo identifikovať, implementovať a vyhodnotiť vhodnú metódu učenia s čiastočným učiteľom pre Frame Field learning.

Metódy

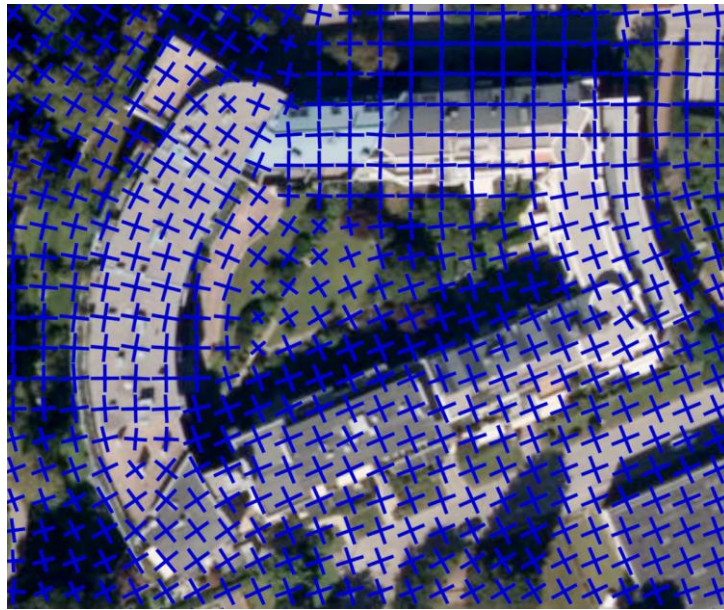
Frame Field learning

Hlavnou myšlienkou Frame Field learning [1] metódy je pomôcť vektorizačnému algoritmu vyriešiť nejednoznačné prípady pri vektorizácii, ktoré sú spôsobené diskretnou pravdepodobnostnou segmentačnou mapou (výstup zo segmentačného modelu), a to pridaním tzv. frame fields poľa (viď.

Obrázok 1) ako ďalšieho výstupu z neurónovej siete, reprezentujúceho geometrické charakteristiky budov.

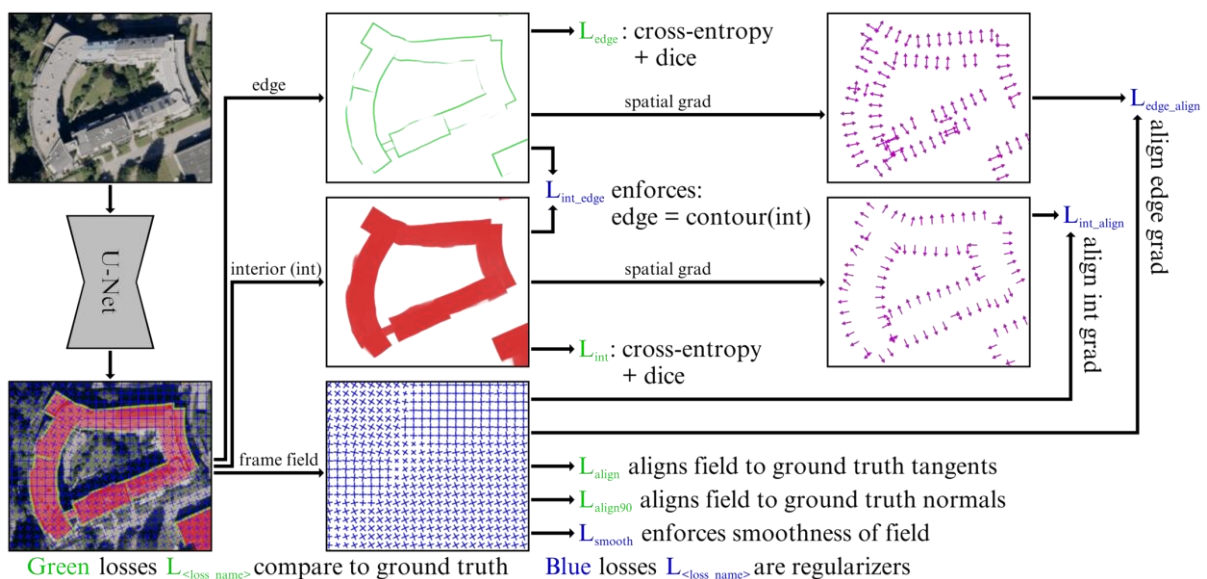
Frame field pole

Frame field je vektorové pole rádu 4, čo znamená, že každému bodu v rovine priradí 4 smerové vektory. Protiľahlé vektory majú rovnakú hodnotu, ale s opačným znamienkom, takže každému bodu v rovine je priradený vektor $\{u, -u, v, -v\}$. Tieto vektory postačujú na definovanie tvaru budov, ktoré sú z veľkej časti pravidelného tvaru s pravouhlými rohmi.



Obrázok 1: Ukážka frame field poľa definovaného pre budovu z trénovacej sady [1].

Frame Field learning



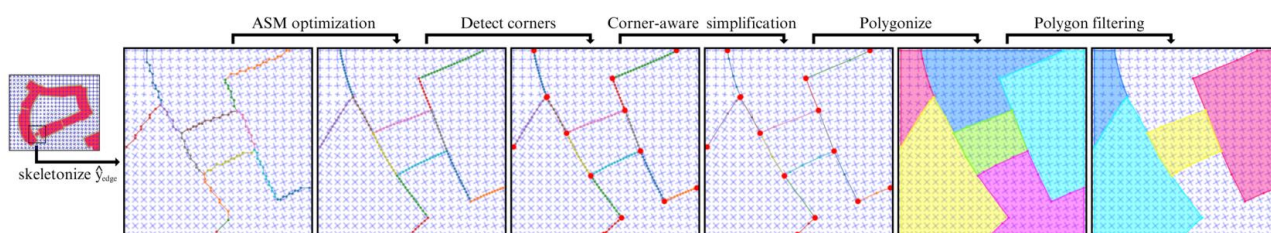
Obrázok 2: Diagram procesu Frame Field learning [1].

Proces metódy Frame Field learning môžeme zosumarizovať nasledovne:

1. Vstupom do neurónovej siete je RGB obraz o veľkosti $3 \times V \times \mathcal{S}$.

2. Na generovanie mapy príznakov (z angl. “feature map”) je možné využiť rôzne segmentačné architektúry, napr. U-Net.
3. Učenie je supervizované (patrí medzi metódy učenia s učiteľom), pričom pre učenie segmentačných masiek sa využívajú označené rastrované polygóny pre interiér a hranice budov. Ako stratová funkcia sa využíva lineárna kombinácia funkcií cross-entropy a Dice loss.
4. Pre učenie samotného Frame Field poľa sa využívajú vektory polygónov označených budov, kde konzistentnosť a presnosť Frame Field poľa zabezpečujú tri stratové funkcie:
 1. L_{align} stratová funkcia riadi správne natočenie Frame Field poľa na smery dotyčnice vektoru polygónu.
 2. $L_{align90}$ stratová funkcia zabraňuje, aby sa Frame Field pole degradovalo na priamkové pole.
 3. L_{smooth} zabezpečuje hladký priebeh Frame Field poľa.
5. Pre zachovanie konzistentnosti medzi segmentačnou pravdepodobnostnou mapou a Frame Field výstupom sú definované regularizačné stratové funkcie, ktoré zarovnávajú Frame Field pole s gradientmi segmentačnej mapy.

Vektorizácia



Obrázok 3: Vizualizácia procesu vektorizácie [1].

Proces vektorizácie transformuje výstup z natrénovanej neurónovej siete do topologicky čistých vektorov pomocou algoritmu Active Skeleton Model (ASM). Princíp algoritmu spočíva v iteratívnom posúvaní vrcholov skeletového grafu do ich ideálnej pozície. Skeletový graf je vygenerovaný pomocou morfolologickej operácie “thinning” z gradientu segmentačnej mapy. Iteratívny posun je riadený gradientovou optimalizačnou metódou, ktorej cieľom je minimalizovať energetickú funkciu, ktorá má nasledujúce zložky:

1. $E_{probability}$ – riadi prispôbenie skeletového grafu kontúram pravdepodobnostnej mapy budovy na konkrétnu hodnotu pravdepodobnosti (napr. 0.5)
2. $E_{frame\ field\ align}$ - riadi zarovnanie každej hrany skeletového grafu na Frame Field pole.
3. E_{length} – zaisťuje homogénnu distribúciu vrcholov skeletového grafu.

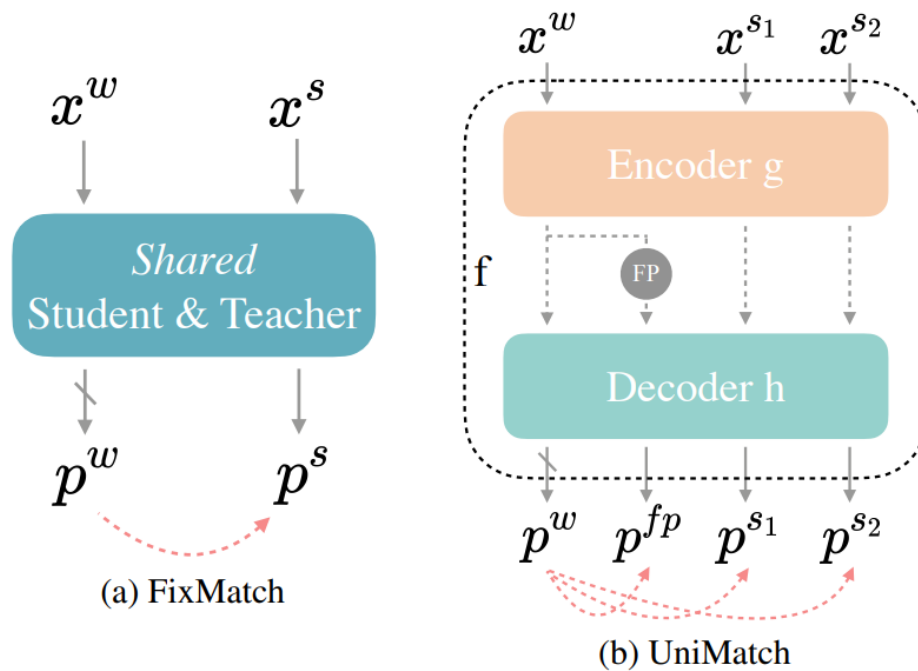
UniMatch metóda čiastočne riadeného učenia

UniMatch [2], pokročilá metóda učenia s čiastočným učiteľom z kategórie regulátorov konzistentnosti, stavia na základných princípoch vytvorených metódou FixMatch [3], ktorá je základnou metódou v tejto kategórii algoritmov. Funguje na princípe pseudo-označovania (z angl. “pseudo-labeling”) v kombinácii s reguláciou konzistentnosti.

Základný princíp metódy FixMatch spočíva v generovaní pseudo-označení (anotácií) pre neanotované dáta, pomocou predikcií neurónovej siete. To znamená, že pre slabo perturbovaný neanotovaný vstup x^w sa vygeneruje predikcia p^w , ktorá slúži ako pseudo-označenie pre predikciu silne perturbovaného vstupu x^s . Následne sa vypočíta hodnota chybovej funkcie, napr. cross-entropy(p^w, p^s), pričom do úvahy sa berú iba tie oblasti z p^w , ktoré majú hodnotu pravdepodobnosti väčšiu ako daný prah, napr. >0.95 .

Rozšírenie metódy UniMatch oproti metóde FixMatch spočíva v dvoch princípoch:

1. UniPerb (Unified Perturbations for Images and Features) - aplikácia perturbácie na úrovni príznakov (z angl. "feature perturbation"). V praxi to znamená, že na výstup (teda príznak - feature) z encoder vrstvy neurónovej siete sa aplikuje dropout funkcia, ktorá náhodne vynuluje niektoré príznaky. Takto upravený výstup z encoder vrstvy následne vstupuje do decoder časti siete, ktorá vygeneruje p^{fp} .
2. DusPerb (Dual-Stream Perturbations) – namiesto jednej silnej perturbácie sa využívajú dve silné perturbácie x^{s1} a x^{s2} .



Obrázok 4: (a) FixMatch základná metóda, (b) použitá UniMatch metóda. FP je označenie pre feature perturbation (perturbácia príznakov), w ako weak (slabá) a s ako strong (silná) perturbácia [2].

V konečnom dôsledku máme tri stratové funkcie - cross-entropy(p^w, p^{fp}), cross-entropy(p^w, p^{s1}), cross-entropy(p^w, p^{s2}). Tie sa nakoniec lineárne kombinujú so supervizovanou stratovou funkciou.

Táto metóda v súčasnosti patrí medzi state-of-the-art metódy učenia s čiastočným učiteľom. Hlavnou výhodou tejto metódy je jej jednoduchosť pri implementácii a nevýhodou je jej citlivosť na výber vhodnej slabej a silnej perturbácie.

Integrácia UniMatch metódy do Frame Field učenia

Implementácia UniMatch do Frame Field learning frameworku

Aby sme mohli implementovať UniMatch metódu do Frame Field learning štruktúry, potrebovali sme najprv definovať slabú a silnú perturbáciu v kontexte leteckých snímok. Ako slabé perturbácie sme zvolili základné priestorové transformácie obrazu, vrátane rotácie, zrkadlenia a vertikálneho/horizontálneho prevrátenia. Všetky tieto transformácie sú oprávnené pre letecké snímky.

V prípade silných perturbácií sme použili fotometrické transformácie. Tie zahŕňajú úpravy odtieňa, farby, či jas obrazu. Poskytujú výraznejšie zmeny snímok než s použitím priestorových transformácií.

Dôležitým krokom bola implementácia perturbácie na úrovni príznakov (feature perturbation). Túto perturbáciu sme implementovali ako dropout mechanizmus vo vrstve medzi encoder a decoder časťami architektúry U-Net. Tento mechanizmus zahodí (nastaví na nulu) náhodne vybrané hodnoty príznakov (výstup z encoder vrstvy). Takto upravené hodnoty výstupu z encoder časti siete vstupujú ďalej do decoder časti U-Net architektúry.

V prípade dual-stream perturbácií sme prispôbili Frame Field framework tak, aby využíval dve silné perturbácie. Predikcia pre slabú perturbáciu sa použila ako pseudo-označenie pre dve silné perturbácie (preto označenie dual-stream). Dve silné perturbácie prispievajú k celkovej robustnosti a efektívnosti modelu.

Prostredníctvom týchto úprav bola UniMatch metóda úspešne integrovaná do Frame Field learning algoritmu, čím sa zvýšila jeho schopnosť efektívne spracúvať a učiť sa z anotovaných a hlavne neanotovaných dát.

Experimenty

Dáta

Anotované dáta

Anotované dáta použité v štúdiu pochádzajú z troch rôznych zdrojov, detaily sú uvedené v Tabuľke 1.

Názov	Typ dát. sady	Pokrytá rozloha (km ²)	Počet budov	Ortofotografické rozlíšenie (m)	Rastrová dlaždica (px)
Geodeticca-Buildings	Súkromná	198.18	50 354	0.25	12500x10000
INRIA [4]	Verejná	810.00	206 679	0.30	5000x5000
Landcover.ai [5]	Verejná	216.27	12 354	0.25/0.50	9000x9500 / 4200x4700

Tabuľka 1: Prehľad troch zdrojov anotovaných dát použitých na tréning modelov.

Neannotované dáta

Neannotované dáta (verejne dostupné vysoko kvalitné letecké snímky) pochádzajú z Geodetického a kartografického ústavu (GKÚ) [6]. Pri výbere sme sa zamerali na oblasť s rozlohou 7 000 km², čím bola zaistená diverzita rôznych povrchov krajín a mestských prostredí.

Spracovanie dát: Patching

Anotované aj neanotované snímky boli spracované pomocou metódy “patching”, ktorá obraz rozdeľuje na malé časti veľkosti 320x320px. Táto veľkosť bola špecificky vybraná tak, aby vyhovovala požiadavkám pre vstup zvolenej neurónovej siete. Takýmto spôsobom vzniklo z anotovaných dát približne 55 000 malých častí a z neanotovaných dát okolo 244 000 častí.

Trénovanie

Architektúra modelu

Použitý model sme navrhli s pomocou U-Net architektúry s EfficientNet-B4 základom. Táto kombinácia poskytuje dobrú rovnováhu presnosti a efektívnosti, čo je veľmi dôležité pri práci s komplexnými segmentačnými úlohami. EfficientNet-B4 ako základ neurónovej siete bol vybraný pre optimálnu rovnováhu medzi spotrebou pamäte a výkonom. V metóde Frame Field learning sa U-Net architektúra ukázala byť vysoko efektívna, o čom svedčia výsledky použitia tejto siete v rôznych štúdiách.

Trénovací proces

Na trénovanie sme použili AdamW optimalizátor, ktorý kombinuje výhody Adam optimalizácie s regularizačnou metódou “weight decay”, čím pomáha modelu lepšie generalizovať. Aby sme sa vyhli pretrénovaniu modelu, použili sme L2 regularizáciu a taktiež bola použitá metóda ReduceLROnPlateau na optimalizáciu parametra rýchlosti učenia. Táto metóda upravuje parameter rýchlosti učenia na základe validačnej straty.

Úpravy potrebné pre implementáciu učenia s čiastočným učiteľom

Kľúčovým aspektom nášho tréovania bolo nastavenia podielu anotovaných a neanotovaných obrázkov. Experimentovali sme s pomermi od 1:1 do 1:5 (počet anotovaných : počet neanotovaných). Takýmto spôsobom sme zisťovali, ako rôzne množstvá neanotovaných dát ovplyvňujú trénovací proces. Identifikovali sme optimálny pomer pre trénovanie nášho modelu tak, aby bolo zachované efektívne učenie s využitím metódy učenia s čiastočným učiteľom.

Vyhodnotenie modelu

Na vyhodnotenie nášho modelu na extrakciu budov sme zvolili metriky, ktoré precízne merajú ako presne sa predikcie zhodujú so skutočnými štruktúrami.

Intersection over Union (IoU)

Kľúčovou metrikou, ktorú sme využívali je metrika s názvom Intersection over Union (IoU). Počíta zhodu medzi predikciami modelu a skutočným tvarom budov. Hodnota skóre IoU blízka 1 znamená, že naše predikcie sú podobné skutočným budovám. Táto metrika je nevyhnutná na posúdenie geometrickej presnosti pre segmentované oblasti, pretože odráža presnosť vytýčenia hraníc budov. Okrem toho, vyhodnotením pomeru správne predikovanej oblasti ku kombinovanej oblasti (zjednotenie oblasti predikcie a skutočnej oblasti), nám IoU poskytuje jasnú mieru efektivity modelu v zachytávaní skutočného kontextu a tvaru budov v komplexnej mestskej krajine.

Precision, Recall (senzitivita) a F1 skóre

Metrika nazývaná precision vyjadruje podiel správne identifikovaných budov zo všetkých identifikovaných budov. Senzitivita (angl. “recall”) ilustruje schopnosť modelu zachytiť všetky skutočné budovy. Vysoká hodnota tejto metriky poukazuje na citlivosť modelu pri detekcii budov. F1

skóre kombinuje precision a senzitivitu do jednej metriky, poskytujúc vyvážený obraz výkonu modelu.

Complexity Aware IoU (cIoU)

Ďalšou použitou metrikou bola Complexity Aware IoU (cIoU) [7]. Táto metrika rieši nedostatky IoU tým, že vyvažuje presnosť segmentácie a komplexnosť tvarov polygónov. Zatiaľ čo IoU môže viesť model k vytváraniu veľmi komplexných polygónov, cIoU zaručuje, že komplexnosť polygónov (počet ich vrcholov) je zachovaná realistická, čím odráža skutočný tvar budov, ktoré sú obvykle málo komplexné.

N Ratio Metrika

Metrika N ratio je doplnkovým komponentom v našej vyhodnocovacej stratégii. Porovnáva počet vrcholov v našich predpovedaných tvaroch s tými v skutočných budovách [7]. Tým nám metrika pomáha porozumieť, ako presne náš model replikuje detailnú štruktúru budov.

Max Tangent Angle Error (MTAE)

Na zaistenie čistej geometrie pri extrakcii budov, je dôležité presné meranie pravidelnosti kontúr. Chyba maximálneho uhla dotyčníc (resp. Max Tangent Angle Error (MTAE)) [1] je metrika navrhnutá presne pre tieto potreby, a je doplnením Intersection over Union (IoU) metriky. Špecificky cieľi na nedostatok IoU metriky, ktorým je to, že segmentácia s okrúhlymi rohmi môže dosiahnuť vyššie skóre než segmentácia s presnejšími (ostrejšími) rohmi. Vyhodnocovaním zhody okrajov budov cez porovnávanie uhlov dotyčníc vo vybraných bodoch predikovaných a skutočných kontúr, MTAE efektívne penalizuje nepresnosti v orientácii okrajov. Toto zameranie na presnosť okrajov je dôležité pre produkovanie čistých vektorových reprezentácií budov, zdôrazňujúc dôležitosť presného vymedzenia hraníc v segmentačných úlohách.

Vyhodnotenie

Natréňované modely boli testované na veľkej dátovej množne leteckých snímok v plnej veľkosti (namiesto malých častí, pomocou ktorých bola sieť trénovaná). Takéto testovanie poskytuje presnejšie zobrazenie reálnych použití takýchto modelov. Na extrakciu budov zo snímok v plnej veľkosti sme použili techniku posuvného okna, čím boli vytvorené predikcie po jednotlivých segmentoch obrázku. Na okraje prekrývajúcich sa segmentov bola použitá pokročilá priemerovacia technika, dôležitá pre minimalizáciu nežiadúcich efektov a zachovanie konzistentnosti v rámci predikčnej mapy. Výstupná predikčná mapa v plnej veľkosti bola následne vektorizovaná do presných vektorových polygónov s použitím algoritmu Active Skeleton Model (ASM).

Výsledky

Metóda tréovania	Podiel (anotované: neanotované)	IoU (%)	Precision (%)	Senzitivita (%)	F1 Skóre (%)	N Ratio	cIoU (%)	Priem. MTAE(°)
Učenie s učiteľom	-	80.50	85.75	94.27	89.81	2.33	48.89	18.60
Učenie s čiast. učiteľom	1:1	83.88	87.66	93.41	90.44	1.94	56.98	20.47
Učenie s čiast. učiteľom	1:3	85.35	90.04	94.25	92.10	1.76	61.91	18.92
Učenie s čiast. učiteľom	1:5	85.77	90.04	94.76	92.34	1.65	64.75	17.45

Tabuľka 2: Výsledky tréovania modelov pre základný prístup (učenie s učiteľom) a prístupy učenia s čiastočným učiteľom s rôznymi podielmi použitých anotovaných a neanotovaných obrázkov.

Výsledky z experimentov, odrážajúce výkon segmentačného modelu natrénovaného s rôznymi nastaveniami, odhalili zaujímavé zistenia (viď. Tabuľka 2). Vyhodnotili sme výkon základného modelu (len supervizovaný prístup) a výkon modelov tréovaných metódami učenia s čiastočným učiteľom s použitím rôznych podielov anotovaných a neanotovaných dát (1:1, 1:3, a 1:5).

- 1. IoU:** hodnota IoU metriky bola pre základný model na hodnote 80.50%. S prínosom neanotovaných dát do tréovacieho procesu pozorujeme stabilný nárast, dosahujúc až 85.77%, s použitím pomeru 1:5 anotovaných k neanotovaným obrázkom.
- 2. Precision, senzitivita a F1 skóre:** Hodnota metriky precision sa zlepšila z hodnoty 85.75% pre základný model na hodnotu 90.04% pre model s použitým podielom 1:5. Podobne senzitivita sa zľahka zvýšila z hodnoty 94.27% na 94.76%. F1 skóre taktiež narástlo z hodnoty 89.81% na 92.34%. Tieto zlepšenia naznačujú, že zakomponovaním metódy s čiastočným učiteľom sa model stal presnejším a spoľahlivejším v predikciách.
- 3. N Ratio a cIoU:** Výsledky ukazujú znateľný pokles v hodnote metriky N Ratio z hodnoty 2.33 pre základný model, na hodnotu 1.65 pre model s 1:5 podielom (anotované : neanotované), čo indikuje, že učenie s čiastočným učiteľom produkuje jednoduchšie, ale presnejšie vektorové tvary, ktoré viac pripomínajú skutočné štruktúry budov. Toto zjednodušenie tvarov pravdepodobne prispieva k zvýšenej použiteľnosti výstupu v praktických GIS aplikáciách. Súbežne, hodnoty metriky (cIoU) sa signifikantne zlepšili z hodnoty 48.89% pre základný model, na hodnotu 64.75% pre model s 1:5 podielom. Preto sa zdá, že učenie s čiastočným učiteľom nezlepšuje len zhodu predikovaných stôp budov a skutočných stôp budov, ale tiež generuje jednoduchšie vektorové tvary, ktoré sú bližšie reálnym geometrickým tvarom budov.
- 4. Priemerná MTAE:** Redukcia metriky MTAE z 18.60° na 17.45° pri použití učenia s čiastočným učiteľom predstavuje zlepšenie v geometrickej presnosti predikcií modelu. To naznačuje, že táto metóda učenia je lepšia pri zachytávaní architektonických prvkov budov s presnejšie definovanými uhlami, čo prispieva k produkcii topologicky jednoduchších a čistejších vektorových polygónov.

Trénovanie na HPC

HPC konfigurácia

Trénovanie bolo realizované na HPC klastri Devana vybavenom dostatočnými výpočtovými zdrojmi. HPC klaster Devana disponuje 8 GPU uzlami. Každý GPU uzol obsahuje 4 GPU karty NVIDIA A100 s kapacitou VRAM 40GB, 64 jadier CPU a 256GB kapacity RAM. Plánovanie úloh zabezpečuje systém Slurm.

PyTorch Lightning knižnica

Na paralelizáciu sme použili knižnicu PyTorch Lightning, ktoré poskytuje užívateľsky priateľské prostredie pre prácu s viacerými GPU. Táto knižnica umožňuje užívateľovi špecifikovať počet GPU, počet výpočtových uzlov, poskytuje rôzne distribuované stratégie a možnosť mixed-precision trénovania.

Slurm a PyTorch Lightning nastavenie

Pri trénovaní pomocou 1 GPU vyzerala naša Slurm konfigurácia nasledovne:

```
#SBATCH --partition=ngpu
#SBATCH --gres=gpu:1
#SBATCH --cpus-per-task=16
#SBATCH --mem=64000
```

A nastavenie PyTorch Lightning pre *Trainer*:

```
trainer = Trainer(accelerator="gpu", devices=1)
```

Takto sme alokovali jednu GPU kartu zo štyroch dostupných na danom uzle, a 16 CPU zo 64 dostupných, následkom čoho máme 16 workerov pre data loadery. Keďže učenie s čiastočným učiteľom využíva dva data loadery, (jeden pre anotované a ďalší pre neanotované dáta), alokovali sme 8 workerov pre každý z nich. Je dôležité zaručiť, aby celkový počet jadier pre data loadery nepresiahol počet dostupných jadier, pretože trénovanie môže zlyhať.

Distribuované dátovo-paralelné (DDP) trénovanie

S použitím PyTorch Lightning distribuovaného dátovo-paralelného trénovania (DDP) sme dosiahli, že každá použitá GPU bola operovaná nezávisle:

- Každá GPU spracovala časť dátovej sady.
- Všetky procesy inicializovali model nezávisle.
- Všetky procesy vykonali dopredné a spätné šírenie paralelne.
- Gradienty boli synchronizované a spriemerované medzi procesmi.
- Každý proces aktualizoval svoj optimalizátor individuálne.

S týmto prístupom vypočítame počet data loaderov nasledovne: pre učenie s čiastočným učiteľom v prostredí jedného uzla so 4 GPU kartami a dvoma typmi data loaderov, máme 8 data loaderov, pričom každý má 8 workerov – dohromady 64 workerov.

Na plné využitie jedného uzla so 4 GPU sme použili nasledovnú konfiguráciu:

```
#SBATCH --partition=ngpu
#SBATCH --gres=gpu:4
```

```
#SBATCH --exclusive
#SBATCH --cpus-per-task=64
#SBATCH --mem=256000
```

Kľúčové slovo „--exclusive“ znamená, že daný výpočtový uzol nebude súčasne poskytnutý inému používateľovi. Špecifikácie „--cpus-per-task=64“ a „--mem=256000“ sú v danom nastavení redundantné, nakoľko sa použijú všetky výpočtové zdroje daného uzla.

PyTorch Lightning *Trainer*, nastavíme nasledovne:

```
trainer = Trainer(accelerator="gpu", devices=4, strategy="ddp")
```

Využitie viacerých výpočtových uzlov

S použitím PyTorch Lightning knižnice je tiež možné využiť viacero výpočtových uzlov v HPC systéme. Napríklad, využitie 4 uzlov so 4 GPU kartami na každom uzle (dohromady 16 GPU) bolo konfigurované:

```
trainer = Trainer(accelerator="gpu", devices=4, strategy="ddp", num_nodes=4)
```

Analogicky, Slurm konfigurácia bola nastavená takto:

```
#SBATCH --nodes=4
#SBATCH --ntasks-per-node=4
#SBATCH --gres=gpu:4
```

Tieto nastavenia a výsledky zdôrazňujú škálovateľnosť a flexibilitu komplexného tréningového procesu modelov strojového učenia v HPC prostredí, najmä pre úlohy, ktoré vyžadujú významné výpočtové zdroje, ako je napríklad naša úloha využívajúca učenie s čiastočným učiteľom v geopriestorovej dátovej analýze.

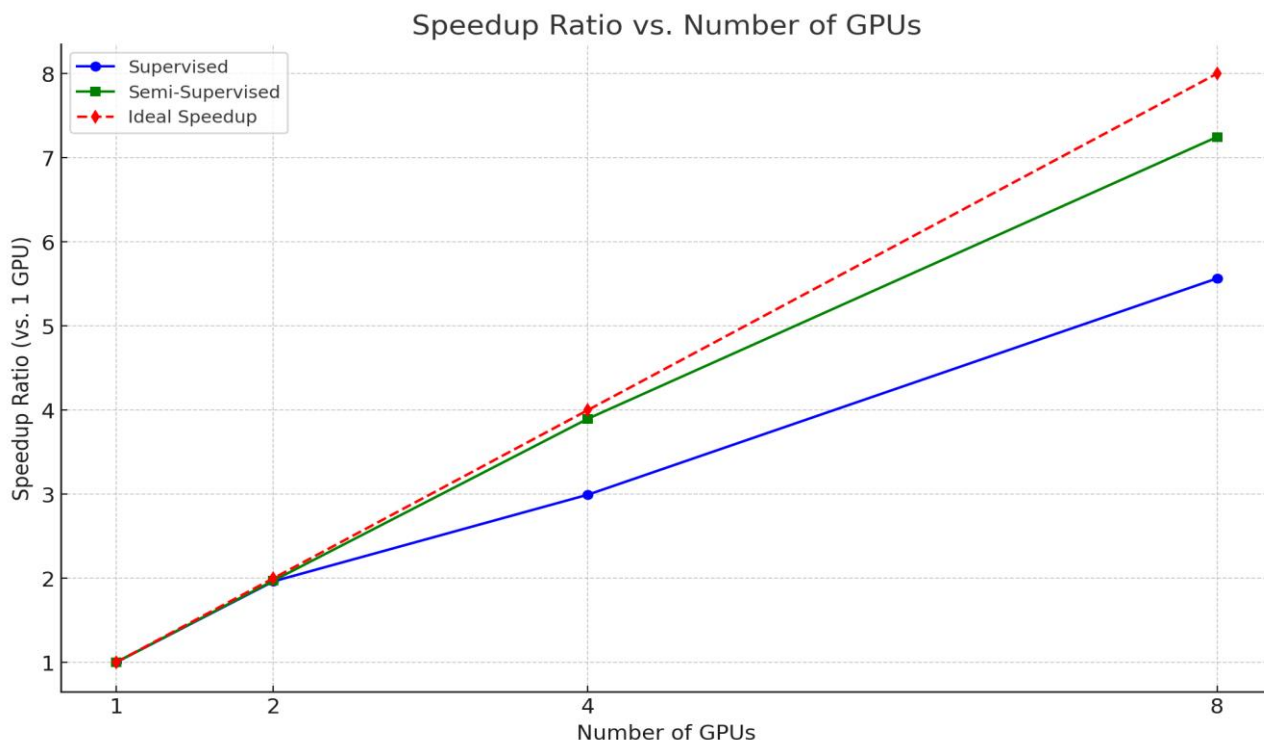
Analýza škálovateľnosti tréningovania

Typ tréningovania	# GPU	Čas na 1 epochu (min)	Pomer urýchlenia (vs. 1 GPU)
Učenie s učiteľom	8 (2 uzly)	1:25	5.56x
	4	2:38	2.99x
	2	4:01	1.96x
	1	7:53	1.00x
Učenie s čiastočným učiteľom, podiel dát 1:1	8 (2 uzly)	3:55	7.25x
	4	7:17	3.90x
	2	14:24	1.97x
	1	28:23	1.00x

Tabuľka 3: Výsledky tréningovania prístupov učenia s učiteľom a učenia s čiastočným učiteľom s 1, 2, 4 a 8 GPU. Pre každú konfiguráciu je uvedený čas na jednu epochu a pomer urýchlenia proti 1 GPU.

V analýze škálovateľnosti tréningovania sme dôkladne preskúmali vplyv rozširovania výpočtových zdrojov na efektívnosť tréningovania modelov s využitím knižnice PyTorch Lightning.

Tento prieskum zahŕňal metódy učenia s učiteľom aj čiastočným učiteľom s dôrazom na zvyšovanie počtu GPU kariet, vrátane prístupu využívajúceho 2 uzly (8 GPU).



Obrázok 5: Urýchlenie pre tréningovanie supervizovanou a nesupervizovanou metódou vzhľadom na počet použitých GPU. Pre porovnanie je uvedené aj ideálne (lineárne) urýchlenie. Učenie s čiastočným učiteľom je bližšie ideálnemu škálovaniu, t.j. efektívnejšie využíva výpočtové zdroje.

Kľúčovým zistením z tejto analýzy je, že nárast v pomeroch urýchlenia pre učenie s učiteľom nie je priamo úmerný počtu použitých GPU kariet. Ideálne, zdvojnásobenie počtu GPU kariet by malo zdvojnásobiť urýchlenie (t.j., napr. použitie 4 GPU kariet by malo mať za následok štvornásobné urýchlenie voči jednej GPU karte). Skutočné hodnoty urýchlenia boli nižšie než ideálne hodnoty. Tento nesúlad možno pripísať tzv. overhead-u (.j. nutnému navýšeniu operácií, ako transfer dát, I/O a pod. a tým pádom aj celkovému trvaniu výpočtu) asociovanému s manažovaním viacerých GPU kariet a výpočtových uzlov, obzvlášť synchronizácii dát cez všetky GPU karty, čo má za následok pokles efektívnosti.

Učenie s čiastočným učiteľom ukázalo mierne iný trend, viac približujúci sa ideálnemu (lineárnemu) nárastu urýchlenia. Zdá sa, že komplexnosť a vyššie výpočtové nároky učenia s čiastočným učiteľom zmierňujú dopad overhead nákladov a tým umožňujú efektívnejšie využívanie viacerých GPU. Napriek výzvam spojeným so synchronizáciou dát cez viacero GPU kariet a výpočtových uzlov, vyššie výpočtové nároky učenia s čiastočným učiteľom umožňujú efektívnejšie škálovanie zdrojov, t.j. urýchlenie bližšie ideálnemu scenáru.

Záver

Výskum predstavený v tejto práci úspešne demonštruje efektívnosť integrácie metódy UniMatch, ktorá patrí medzi metódy učenia s čiastočným učiteľom, do Frame Field learning metódy, pre úlohy extrakcie budov z leteckých snímok. Táto integrácia primárne adresuje notorický nedostatok

anotovaných dát v aplikáciách hlbokého učenia v geografických informačných systémoch (GIS) a navyše, poskytuje škálovateľný a efektívny prístup z hľadiska úspory nákladov.

Výsledky sumarizované v tejto štúdií indikujú, že použitie učenia s čiastočným učiteľom významne zlepšuje výkon modelu vo viacerých kľúčových metrikách, vrátane Intersection over Union (IoU), presnosti pozitívnych predikcií, senzitivity, F1 skóre, N Ratio, complexity-aware IoU (cIoU), a priemernej chyby Max Tangent Angle Error (MTAE). Obzvlášť, zlepšenia v metrikách IoU a cIoU zdôrazňujú zvýšenú presnosť modelu vo vymedzovaní stôp budov a generovaní vektorových tvarov, ktoré vierohodne reprezentujú skutočné štruktúry. Tento výsledok je dôležitý pre aplikácie urbanistického plánovania, environmentálne štúdie a manažment infraštruktúry, kde sú precízne mapovanie a popis budov kľúčové.

Prezentovaná metodika, ktorá kombinuje Frame Field learning s inovatívnym UniMatch prístupom, preukázala, že je vysoko efektívna vo využívaní kombinácie anotovaných a neanotovaných dát. Táto stratégia nielen že zlepšuje geometrickú presnosť predikcií modelu, ale tiež zaručuje generovanie jednoduchších a topologicky presnejších vektorových polygónov. Navyše, škálovateľnosť a efektívnosť tréningu na HPC systéme Devana s použitím knižnice PyTorch Lightning a distribuovanej, dátovo-paralelnej stratégie (DDP) bola kľúčová pre zvládnutie tak výpočtovo náročných úloh, akým je učenie s čiastočným učiteľom nad príslušnými dátami, v časovom rozsahu rádovo desiatok minút, až hodín.

Práca zdôrazňuje potenciál učenia s čiastočným učiteľom v zlepšovaní automatickej extrakcie budov z leteckých snímok. Implementácia UniMatch do Frame Field learning metódy predstavuje významný krok vpred, poskytujúc robustné riešenie pre výzvy spojené s nedostatkom dát a potreby vysokej presnosti geopriestorovej dátovej analýzy. Tento prístup zlepšuje efektívnosť a presnosť extrakcie budov, a taktiež otvára nové možnosti pre aplikácie metód učenia s čiastočným učiteľom v GIS a príbuzných oblastiach.

Pod'akovanie

Výskum bol realizovaný s podporou Národného kompetenčného centra pre HPC, projektu EuroCC 2 a Národného Superpočítačového Centra na základe dohody o grante 101101903-EuroCC 2-DIGITALEUROHPC-JU-2022-NCC-01.

Časť výskumu bola realizovaná s využitím výpočtovej infraštruktúry obstaranej v projekte Národné kompetenčné centrum pre vysokovýkonné počítanie (kód projektu: 311070AKF2) financovaného z Európskeho fondu regionálneho rozvoja, Štrukturálnych fondov EU Informatizácia spoločnosti, operačného programu Integrovaná infraštruktúra 2014-2020.

Literatúra:

- [1] Nicolas Girard, Dmitriy Smirnov, Justin Solomon, and Yuliya Tarabalka. "Polygonal Building Extraction by Frame Field Learning". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2021), pp. 5891-5900.
- [2] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi. "Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation". In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2023), pp. 7236-7246. doi: 10.1109/CVPR52729.2023.00699.

- [3] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence". In: CoRR, vol. abs/2001.07685 (2020). Available: <https://arxiv.org/abs/2001.07685>.
- [4] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. "Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark". In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (2017). IEEE.
- [5] Adrian Boguszewski, Dominik Batorski, Natalia Ziemba-Jankowska, Tomasz Dziedzic, and Anna Zambrzycka. "LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads from Aerial Imagery". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2021), pp. 1102-1110.
- [6] "Ortofotomozaika." Geoportal SK. Accessed February 14, 2024. <https://www.geoportal.sk/sk/zbgis/ortofotomozaika/>.
- [7] Stefano Zorzi, Shabab Bazrafkan, Stefan Habenschuss, and Friedrich Fraundorfer. "PolyWorld: Polygonal Building Extraction with Graph Neural Networks in Satellite Images". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022), pp. 1848-1857.