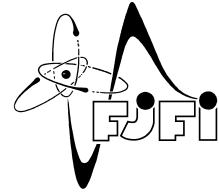




ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta jaderná a fyzikálně inženýrská



Sémantická segmentace plodin z časových řad snímků Sentinel-2

Semantic segmentation of crops from time series of Sentinel-2 images

Diplomová práce

Autor: **Bc. Emanuel Frátrik**
Vedúcí práce: **RNDr. Michal Šorel, Ph.D.**
Konzultant: **Ing. Jiří Franc, Ph.D.**
Akademický rok: 2023/2024

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student:	Bc. Emanuel Frátrik
Studijní program:	Aplikované matematicko-stochastické metody
Název práce (česky):	Sémantická segmentace plodin z časových řad snímků Sentinel-2
Název práce (anglicky):	Semantic segmentation of crops from time series of Sentinel-2 images
Jazyk práce:	slovenština

Pokyny pro vypracování:

1. Prostudujte dostupnou literaturu k problému sémantické segmentace plodin z družicových dat.
2. Seznamte se s daty obsaženými ve veřejném registru půdy LPIS a dalšími daty relevantními k rozpoznávání plodin, které poskytne vedoucí práce.
3. Navrhněte a otestujte vhodné rozšíření neurálních modelů pro sémantickou segmentaci na časové řady snímků.
4. Srovnejte navrženou metodu s metodami dostupnými v literatuře a diskutujte přínosy vašeho přístupu.

Doporučená literatura:

1. D. Sykas, et al., A Sentinel-2 Multiyear, Multicountry Benchmark Dataset for Crop Classification and Segmentation With Deep Learning. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15, 2022, 3323-3339
2. M. Belgiu, O. Csillik, Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, Remote Sensing of Environment 204, 2018, 509-523
3. M. O. Turkoglu, et al. Crop mapping from image time series: Deep learning with multi-scale label hierarchies, Remote Sensing of Environment 264, 2021, 112603
4. R. Güldenring, L. Nalpantidis, Self-supervised contrastive learning on agricultural images, Computers and Electronics in Agriculture 191, 2021, 106510

Jméno a pracoviště vedoucí diplomové práce:

RNDr. Michal Šorel, Ph.D.

Ústav teorie informace a automatizace AV ČR, Pod Vodárenskou věží 4, 182 08 Praha 8

Jméno a pracoviště konzultanta:

Ing. Jiří Franc, Ph.D.

Katedra matematiky FJFI ČVUT v Praze

Trojanova 13

120 00 Praha 2

Datum zadání diplomové práce: 31.10.2022

Datum odevzdání diplomové práce: 3.5.2023

Doba platnosti zadání je dva roky od data zadání.

V Praze dne 31.10.2022


.....
garant oboru


.....
vedoucí katedry




.....
děkan

Pod'akovanie:

Na tomto mieste by som sa rád pod'akoval predovšetkým svojmu školiteľovi RNDr. Michalovi Šorelovi, Ph.D. za ochotu, množstvo cenných rád a odborné vedenie mojej diplomovej práce. V podobnom duchu patrí moje pod'akovanie aj môjmu konzultantovi Ing. Jiřímu Francovi, Ph.D. V neposlednom rade sa chcem pod'akovať svojej rodine za podporu, ktorú mi dávali počas celého štúdia.

Čestné prehlásenie:

Prehlasujem, že som túto prácu vypracoval samostatne a uviedol som všetku použitú literatúru.

V Prahe dňa 8. januára 2024

Emanuel Frátrik

Frátrik

Názov práce:

Sémantická segmentace plodín z časových řad snímků Sentinel-2

Autor: Emanuel Frátrik

Program: Aplikované matematicko-stochastické metody

Druh práce: Diplomová práce

Vedúci práce: RNDr. Michal Šorel, Ph.D., Ústav teorie informace a automatizace AV ČR, Pod Vodárenskou věží 4, 182 08 Praha 8

Konzultant: Ing. Jiří Franc, Ph.D., Katedra matematiky FJFI ČVUT v Praze, Trojanova 13, 120 00 Praha 2

Abstrakt: Satelity misie Sentinel-2 sú v súčasnej dobe veľmi cenným zdrojom dát pre automatizáciu mapovania pokrytia krajiny. V prípade sémantickej segmentácie typov plodín, ktorou sa zaoberáme v našej práci, je esenciálne aby na maximalizáciu výslednej presnosti bolo efektívne využité vysoké časové rozlíšenie snímkov Sentinel-2, umožňujúce lepšie zachytiť temporálnu dynamiku rastu plodín. V rámci práce preto najprv zhrnieme možnosti klasifikácie plodín vo forme rešerše. Následne na ňu naviažeme návrhom vlastných vylepšení neurálnych modelov, schopných vykonať sémantickú segmentáciu plodín z časových rád snímkov a predstavíme vlastný dataset typov plodín pre Českú republiku pozostávajúci z časových rád snímkov Sentinel-2. Navrhnuté modely porovnáme s alternatívnymi prístupmi a prácu zakončíme praktickým návrhom aplikácie schopnej automaticky získať satelitné snímky a vykonať nad nimi sémantickú segmentáciu typov plodín v rámci Českej republiky.

Kľúčové slová: časová rada, hlboké učenie, klasifikácia plodín, satelitné snímky, Sentinel-2, sémantická segmentácia

Title:

Semantic segmentation of crops from time series of Sentinel-2 images

Author: Emanuel Frátrik

Abstract: Satellites of the Sentinel-2 mission are currently very valuable data sources for the automation of land cover mapping. In the case of the semantic segmentation of crop types, which we deal with in our work, it is essential to effectively use the high temporal resolution of Sentinel-2 images to maximize the resulting accuracy, enabling a better capture of the temporal dynamics of crop growth. As part of the work, we will therefore first summarize the possibilities of crop classification in the form of research. Subsequently, we will build on it by proposing our own improvements to neural models, capable of performing semantic segmentation of crops from time series of images, and we will present our own dataset of crop types for the Czech Republic consisting of time series of Sentinel-2 images. We will compare the proposed models with alternative approaches and conclude the work with a practical design of an application capable of automatically acquiring satellite images and performing semantic segmentation of crop types within the Czech Republic.

Key words: time series, deep learning, crop classification, satellite imagery, Sentinel-2, semantic segmentation

Obsah

Použité skratky	11
Úvod	12
1 Klasifikácia plodín v satelitných snímkach	14
2 Dátové zdroje a datasety	23
2.1 Satelitné snímky	23
2.2 Dataset PASTIS	24
2.3 Dataset S2TSCZCrops	25
2.3.1 Oblasť záujmu	25
2.3.2 Referenčné dáta	25
2.3.3 Temporálne charakteristiky	27
2.3.4 Implementačné detaily	28
3 Metodika	31
3.1 Teoretické pozadie	31
3.1.1 Attention mechanizmus	31
3.1.2 Model U-TAE	33
3.2 Návrh metódy	33
3.2.1 Pozičné kódovanie	33
3.2.2 TimeU-Net	35
3.2.3 W-TAE	35
3.3 Alternatívne prístupy	35
3.4 Regularizácia	38
3.4.1 Label smoothing	38
3.4.2 Temporálny dropout	39
3.4.3 Augmentácia datasetu	39
3.5 Post-processing	39
3.6 Problematika hraničných pixelov	40
3.7 Hodnotenie výsledkov	42
3.8 Implementačné detaily	43
4 Výsledky	45
4.1 Segmentácia pomocou navrhnutých modelov	45
4.1.1 U-TAE	45
4.1.2 W-TAE	45
4.1.3 TimeU-Net	46

4.1.4	Kvalitatívne výsledky	48
4.2	Hraničné pixely	51
4.3	Homogenizácia predikcie	54
4.4	Závislosť na dĺžke časovej rady	59
5	Diskusia	62
5.1	Navrhnutá metóda	62
5.2	Dataset S2TSCZCrops	63
5.3	Hraničné pixely	64
5.4	Post-processing	65
5.5	Závislosť na dĺžke časovej rady	66
5.6	Porovnanie s alternatívnymi prístupmi	66
6	Návrh aplikácie	69
6.1	Vstupy aplikácie	69
6.2	Spracovanie dát	69
6.3	Výstupy aplikácie	71
	Záver	73
	Literatúra	82
	Prílohy	83
	Charakteristiky použitých dát	83

Použité skratky

Skratka	Popis
NDVI	Normalized difference vegetation index
EVI	Enhanced vegetation index
NIR	Near Infrared Radiation
SWIR	Shortwave Infrared Radiation
UTM	Universal Transverse Mercator
WGS84	World Geodetic System 1984
LULC	Land Use Land Cover
OA	Overall accuracy
TOA	Top of atmosphere
BOA	Bottom of atmosphere
mIoU	mean Intersection over Union
SITS	Satellite image time series
S1	Sentinel-1
S2	Sentinel-2
PB	Půdní blok
DPB	Díl půdního bloku
SOTA	State of the art
DOY	Day of year
TAE	Temporal attention encoder [70]
L-TAE	Lightweight TAE [67]
ML	Maximum likelihood
MLP	Multi Layer Perceptron
SVM	Support vector machine
RF	Random forest
DT	Decision tree
CRF	Conditional random field

Úvod

Určenie typu pestovaných plodín na nejakom území vo forme plodinových máp je bezpochyby dôležitou súčasťou monitorovania poľnohospodárskych aktivít a stavu životného prostredia. Plodinové mapy sú predpokladom pre množstvo nadväzujúcich analýz a rozhodovacích procesov ako napr. predikcia žatvy, odhad výnosov, monitorovanie úrodnosti pôdy s čím súvisí aj stanovenie vhodnej stratégie rotácie plodín na poliach, ďalej tvorba národnej a medzinárodnej poľnohospodárskej politiky ako aj kontrola ich plnenia [3, 1].

Nakoľko sú poľnohospodárske aktivity charakterizované sezónnou premenlivosťou spôsobenou predovšetkým rotáciou pestovaných plodín, vzniká potreba pravidelnej aktualizácie plodinových máp typicky na ročnej báze. Spomínaná potreba aktualizácie ako aj náročnosť manuálnej tvorby plodinových máp sú dôvody prečo je nutné skúmať možnosti automatizácie procesu ich generovania. Dáta zo vzdialeného prieskumu Zeme, predovšetkým satelitné snímky v spojení s využitím metód strojového učenia sa ukazujú byť veľmi cenným prostriedkom ako danú automatizáciu uviesť do praxe [95].

V tejto práci sa venujeme aplikácii pokročilých metód hlbokého učenia pre sémantickú segmentáciu plodín v časových radách satelitných snímok. Ako zdroj satelitných snímok slúži misia Sentinel-2 poskytujúca multispektrálne obrazové dáta s až 13 spektrálnymi pásmami, čo z nej robí unikát v porovnaní s podobnými misiami akými sú napríklad Landsat a SPOT. Snímky, ktoré satelity misie Sentinel-2 ponúkajú, majú okrem postačujúceho priestorového rozlíšenia aj vysoké časové rozlíšenie v rádoch jednotiek dní, ktoré je ich hlavnou devízou a umožňuje tak využiť závislosť spektrálnej odozvy na fenologickej fáze [81] v ktorej sa rastlina nachádza, čo je obzvlášť vhodné pre zvyšovanie presnosti plodinových máp [95].

Hlavné ciele našej práce sa dajú zhrnúť do nasledujúcich štyroch bodov:

1. Vykonať rešerš metód pre sémantickú segmentáciu plodín z družicových dát
2. Na základe poskytnutých dát o type plodín pre Českú republiku vyhotoviť dataset pozostávajúci z časových rád snímok
3. Navrhnuť rozšírenie modelov hlbokého učenia tak aby boli schopné vykonávať sémantickú segmentáciu z časových rád satelitných snímok
4. Výsledný model otestovať, porovnať s dostupnými alternatívami a diskutovať jeho prínosy

V rámci práce sme si stanovili aj dodatočné ciele. Jedným z nich je preskúmať možnosti ako prispieť k zlepšeniu reprezentácie hraničných pixelov vo výsledných segmentačných mapách. Ďalej by sme chceli posunúť získaný model do praktickej roviny v zmysle návrhu a implementácie aplikácie schopnej dáta samostatne získať, klasifikovať a následne užívateľovi vizuálne prezentovať. Zdrojové kódy súvisiace s tréňovaním a testovaním modelov ako aj so samotnou prípravou resp. spracovaním dát a finálnou implementáciou aplikácie sú dostupné na adrese <https://github.com/Many98/Crop2Seg>.

Celkovo je práca rozdelená do šiestich kapitol. V 1. kapitole priblížime problematiku klasifikácie resp. mapovania plodín v satelitných snímkach vo forme rešerše existujúcich metód a prístupov k jej riešeniu. Následne v 2. kapitole zhrnieme charakteristiky satelitných snímok a popíšeme benchmarkový dataset typov plodín zvaný PASTIS. Taktiež zhrnieme postup tvorby a charakteristiky nami vytvoreného datasetu S2TSCZCrop. Návrh nášho modelu, stručný popis alternatívnych prístupov slúžiacich pre porovnanie ako aj zhrnutie metodiky experimentovania bude vykonané v kapitole 3. Kapitola 4 prinesie kvalitatívny a kvantitatívny prehľad výsledkov jednotlivých vykonaných experimentov, pričom prínosy a nedostatky navrhnutých postupov vzhľadom k celkovej presnosti sémantickej segmentácie plodín ako aj porovnanie s dostupnými alternatívami budeme diskutovať v nasledujúcej kapitole 5. Možný praktický prínos našej práce ukážeme v poslednej kapitole 6 kde na základe získaného modelu navrhne aplikáciu schopnú automaticky získať časovú radu satelitných snímok a vykonať nad nimi sémantickú segmentáciu typov plodín v Českej republike.

Kapitola 1

Klasifikácia plodín v satelitných snímkach

Využitie dát z diaľkového prieskumu Zeme pre účely mapovania resp. klasifikácie plodín je vítanou alternatívou ku klasickému terénnemu prieskumu, ktorý je síce často presnejší, ale časovo a finančne náročný. Satelitné snímky v tomto ohľade ponúkajú hneď niekoľko výhod. Jednak je to prívetivý a v niektorých prípadoch úplne voľný prístup, pokrytie veľkého územia, ale aj pravidelnosť snímania zemského povrchu. Predovšetkým misie Sentinel-1, Sentinel-2 a Landsat ponúkajúce radarové a multi-spektrálne snímky v strednom priestorovom rozlíšení sa tešia v súčasnej dobe veľkému zájmu v oblasti mapovania a analýzy plodín [47, 37, 22]. Nepopierateľnou devízou misii Sentinel-1 a Sentinel-2 je vysoké časové rozlíšenie snímok v jednotkách dní, ktoré umožňuje tvoriť časové rady s veľkým počtom pozorovaní, čo prináša lepšie možnosti modelovania temporálnej dynamiky rastu plodín a ich výslednej klasifikácie. Schopnosť penetrovať oblačnosť a nezávislosť na poveternostných podmienkach je hlavnou výhodnou využitia radarových snímok oproti snímkam v spektrálnej doméne. Z tohto dôvodu niektorí autori pre účely mapovania plodín pristúpili k použitiu práve radarových snímok [58, 75, 89, 15, 21, 101]. Vo viacerých prípadoch [68, 28, 34, 18] bola skúmaná fúzia radarovej a spektrálnej domény pričom závery poukazujú na vhodnosť takéhoto riešenia oproti prístupu len s jednou doménou. Na druhú stranu použitie výlučne spektrálnych snímok resp. príznakov sa zdá mať výraznejší vplyv na presnosť mapovania plodín v porovnaní s použitím radarových dát. Napríklad v [68] autori pri riešení sémantickej segmentácie plodín dosiahli vhodnou fúziou Sentinel-2 a Sentinel-1 snímok hodnotu metriky mIoU 66.3% oproti 63.6 % pri použití len Sentinel-2 snímok. V prípade spracovania radarových snímok bola dosiahnutá výrazne nižšia hodnota mIoU a síce 54.7 %. Klasifikácia na úrovni parciel vykazovala rovnaký trend.

Neodmysliteľnou súčasťou klasifikácie plodín v satelitných snímkach je temporálna dynamika ich spektrálnej odozvy súvisiaca s fyziologickými zmenami rastliny v rôznych štádiách rastu. Využitie a spracovanie časových rád satelitných snímok je z tohto hľadiska bezpochyby kľúčovým pre tvorbu presných plodinových máp [86, 95, 22]. Na druhú stranu aj keď použitie časových rád snímok prináša typicky zlepšenie výslednej presnosti klasifikácie tak takýto prístup nemusí byť z nejakého dôvodu možný. Či už je to kvôli výpočtovým alebo pamäťovým nárokom ich spracovania, nadmernému znečisteniu oblačnosťou [51] alebo jednoducho plodinová mapa musí byť doručená už v začiatkoch vegetačného obdobia plodín [40, 95]. V takýchto prípadoch autori volili buď prístup s mono-temporálnymi snímkami [46, 51, 87, 99, 54] v hlavnom vegetačnom období študovaných plodín, prípadne sa snažili z dostupnej časovej rady vybrať najvhodnejšie

snímky [95, 40]. Niektorí autori kompenzovali nedostatok informácie v temporálnej doméne snímkami s vysokým priestorovým rozlíšením [99, 54] alebo použitím hyperspektrálnych snímok [51].

Každopádne väčšina metód používaných na mapovanie plodín s ktorými sa stretne v literatúre nejakým spôsobom využíva vo svoj prospech aj časovú doménu [103, 36, 31, 71, 45, 95, 42, 22, 61, 64, 83]. Klasifikácia časových rád spektrálnych profilov prípadne extrahovaných spektrálno-temporálnych príznakov je spoločnou charakteristikou väčšiny klasických prístupov k mapovaniu plodín. V tomto prípade buď každý pixel reprezentuje nezávislé pozorovanie resp. najmenšiu jednotku, ktorá má byť klasifikovaná alebo sa pristúpi k predsegmentovaniu obrazu kedy sa pixely s podobnými priestorovými a spektrálnymi charakteristikami zlúčia dokopy a takto vytvoria objekt, ktorý je následne klasifikovaný podľa príznakov, ktoré ho charakterizujú ako celok [49, 99, 43, 52, 20, 17]. Alternatívou k predsegmentovaniu obrazu je použitie externej informácie o hraniciach poľa obsahujúceho jeden typ plodiny. Nevýhodou prvého zo spomínaných prístupov, tzv. analýzy obrazu na úrovni pixelov, je možnosť, že výsledné mapy budú obsahovať zašumené predikcie. Takýto salt and pepper šum vzniká predovšetkým kvôli novej vnútro-triednej variabilite spektrálnej odozvy pixelov reprezentujúcich plodinu na konkrétnom poli. Práve druhý spôsob, analýza na úrovni objektov, zamedzuje problému so salt and pepper šumom [49]. Objektové metódy navyše umožňujú pracovať aj s geometrickými a texturálnymi príznakmi študovaného povrchu napr. vo forme výpočtu plochy, shape indexu alebo hodnôt homogenity, entropie a kontrastu ko-okurenej matice [20, 99, 41, 91]. Vo všeobecnosti sa ukazuje, že na objektoch založená klasifikácia vykazuje v porovnaní na pixeloch založených metódach lepšie klasifikačné výsledky [17, 52, 49]. Na drúhu stranu nevýhoda objektových klasifikačných metód spočíva v nutnosti správneho predsegmentovania, čo je problém sám o sebe, prípadne v nutnosti disponovať vhodnými vektorovými dátami, ktoré dané objekty definujú. V tomto zmysle sú teda na pixeloch založené metódy priamočiare a zrejme aj kvôli tomu stále často využívané v oblasti mapovania plodín.

Oblíbenou a často používanou metódou v prípade klasifikácie plodín v časových radoch snímok je spracovanie pôvodných spektrálnych pásiem do formy tzv. spektrálnych indexov, ktoré sú špecificky navrhnuté tak, aby zvýraznili konkrétnu charakteristiku sledovaného povrchu. Medzi oblíbené indexy v oblasti mapovania plodín patrí napr. normalized difference vegetation index (NDVI) a enhanced vegetation index (EVI), ktoré slúžia na zvýraznenie a monitorovanie stavu vegetácie [44, 49, 17, 33].

K samotnej klasifikácii časových rád pristúpili autori rôzne. Jedným z priamočiarych prístupov je použitie rôznych prahovacích pravidiel definovaných na základe expertnej znalosti fenologických vzorcov jednotlivých plodín [27, 44, 58]. Takýto knowledge/rule-based klasifikačný prístup pozostáva zo stanovenia benchmarkových temporálnych profilov jednotlivých plodín a následnom určení vhodných charakteristík a práhov súvisiacich s typickými fenologickými stavmi plodiny voči ktorým budú porovnávané klasifikované vzorky. Výhodou takéhoto prístupu je to, že na zostavenie spomínaných benchmarkových temporálnych profilov nie je potrebná rozsiahla tréningová sada. Práve spomenutým spôsobom autori v [58] klasifikovali časovú radu radarových snímok pričom na vyhotovenie práhov použili charakteristiky ako napr. trend v časových radoch pásiem VV, VH a podielu VH/VV. Li et al. [44] extrahovali z časových rád NDVI tri fenologické stavy plodín a to konkrétne dátum výsadby plodiny, dátum prechodu z vegetačného do reprodukčného stavu a nakoniec dátum zberu plodiny. Dané tri charakteristické dátumy súviseli s hodnotami minima, maxima a inflexnými bodmi v rámci temporálneho profilu NDVI indexu. Porovnanie s random forest algoritmom ukázalo, že oba prístupy dosahujú porovnateľnú presnosť.

Analýzou temporálneho profilu NDVI sa zaoberali aj autori v [17]. Na klasifikáciu využili me-

tódu time-weighted dynamic time warping (TWDTW), ktorá slúži na meranie podobnosti medzi dvomi časovými radami. Jej výhodou je invariancia voči posunom v temporálnych resp. fenologických profiloch plodín spôsobenými meteorologickými zmenami naprieč rokmi prípadne kvôli výpadkom pozorovaní z dôvodu nepriaznivého počasia, nakoľko v procese výpočtu hľadá optimálne zarovnanie analyzovaných časových rád. Aj keď sa porovnaním s algoritmom random forest ukázalo, že random forest dosahuje lepšie klasifikačné výsledky vzhľadom k metrike overall accuracy, tak tento záver neplatil, keď boli použité na tréning iba tri vzorky časových rád. V prípade nedostatku tréningových dát sa ukázala byť metóda TWDTW robustnejšia. Vysoká výpočetná náročnosť metódy je však prekážkou jej aplikácie na veľké objemy dát zo vzdialeného prieskumu Zeme.

Iný prístup zvolili autori v [16] kedy za pomoci pravdepodobnostného modelu zvaného conditional random field (CRF) modelovali medziročnú rotáciu plodín na poliach od roku 2010 na základe deklarácií farmárov z LPIS systému. Odhad matice prechodu bol vykonaný za predpokladu temporálno-priestorovej homogenity parciel pričom vplyv samotných pozorovaní (spektrálnych a radarových) bol modelovaný algoritmom random forest.

Supervizované algoritmy ako maximal likelihood (ML), support vector machine (SVM), decision tree (DT) alebo random forest (RF) [18, 80, 33, 43, 20, 99, 30] boli taktiež nemálokrát úspešne použité na klasifikáciu plodín. Blickensdörfer et al. [18] klasifikovali 24 tried rôznych druhov plodín v Nemecku pomocou algoritmu random forest. Ako príznaky im slúžili spektrálne, radarové ale aj environmentálne dáta súvisiace s topológiou, teplotou a množstvom zrážok. Experimenty potvrdili prínos kombinácie spomínaných zdrojov dát. Li et al. [43] zvolili na klasifikáciu rozhodovací strom pričom ako príznaky využili všetkých 26 členov časovej rady NDVI indexu a navyše zadefinovali päť dodatočných tzv. time series indexov, ktoré mali za úlohu podchytiť špecifické fázy rastu plodín. Podobne Matton et al. [49] extrahovali z temporálnych profilov NDVI a červeného spektrálneho pásma päť charakteristických fáz rastu plodín a síce maximálnu hodnotu červeného spektrálneho pásma, maximálny pozitívny a negatívny sklon NDVI a nakoniec maximum a minimum NDVI profilu. Finálne príznaky korešpondovali s konkrétnymi hodnotami spektrálneho pásma, pozorovanými v spomínaných piatich fázach rastu. V rámci klasifikácie autori použili jednak k-means algoritmus ako aj ML klasifikátor pričom tento dosiahol lepší výsledok. Nakoniec spomeňme ešte prácu od Hu et al. [33] kde navrhujú metódu klasifikácie plodín pomocou algoritmu SVM, ktorú predchádza výber optimálnych fenologických príznakov teda tých, ktoré prinášajú dostatočnú separabilitu tried plodín. Výberový príznakový priestor bol pritom tvorený všetkými prvkami časových rád piatich vegetačných spektrálnych indexov.

Rozvoj výpočetnej techniky a pokroky v oblasti hlbokého učenia umožnili výskum mapovania plodín za pomoci pokročilých architektúr neuronových sietí schopných samostatne extrahovať príznaky z dát a učiť sa ich užitočné reprezentácie. V oblasti počítačového videnia to boli predovšetkým konvolučné neurónové siete pracujúce s lokálnou kontextuálnou informáciou, ktoré priniesli revolúciu v oblasti spracovania obrazových dát. Efektívnosť konvolučných vrstiev spočíva hlavne v zdieľaní parametrov a extrakcii lokálnych vzorov a štruktúr v obraze pričom ich hierarchickým spojením za sebou sa docieli zväčšenie receptívneho poľa a extrakcia hierarchických príznakov v rámci priestorovej domény obrázka.

Práve schopnosť spracovávať spektrálnu informáciu od susedných pixelov a efektívne tak extrahovať príznaky zo spektrálno-priestorovej dimenzie snímok umožňuje získavať plodinové mapy na úrovni pixelov s typicky vyššou kvalitou v porovnaní s klasickými prístupmi opísanými vyššie. Podotknime, že aj keď je označenie klasifikácia na úrovni pixelov často používaná zámenne s pojmom sémantická segmentácia tak my v našej práci označujeme za sémantickú seg-

mentáciu prípad kedy je do klasifikácie na úrovni pixelov explicitne zahrnuté aj spracovanie priestorovej dimenzie snímok ako je to napr. v prípade použitia konvolučných sietí.

Medzi architektúry úspešne aplikované na problematiku sémantickej segmentácie a spracovania satelitných snímok patrí napr. plne konvolučná neurónová sieť U-Net [59] a jej modifikácie [96, 26, 79, 31, 29, 60, 87, 74]. Samotná sieť bola pôvodne vytvorená pre segmentáciu v medicínskych snímkach pričom je charakteristická dvomi symetrickými ramenami zvanými encoder a decoder. Encoder slúži na extrakciu hierarchických príznakov zo vstupných dát pomocou konvolučných vrstiev, ktoré sa následne v časti decoder používajú na postupnú rekonštrukciu pôvodných priestorových rozmerov. Na zabránenie úniku informácie pri hierarchickom znižovaní priestorových rozmerov v časti encoder sú navyše použité tzv. skip spojenia kde sa feature mapy z encoder ramena spájajú s korešpondujúcimi feature mapami z decoder časti. Takéto skip spojenia umožňujú zlepšenie spracovania detailných štruktúr v obrázku, čo má za následok zvýšenie kvality segmentačnej mapy.

V predchádzajúcom sme poukázali na to, že klasické metódy pristupujú k mapovaniu plodín na úrovni pixelov v podstate výlučne ako ku klasifikácii časových rád, čo v praxi znamená, že každý pixel reprezentuje nezávislé pozorovanie charakterizované len spektrálno-temporálnou dimenziou. Nie je prekvapivé, že takýto prístup k mapovaniu plodín skúmali aj autori za pomoci špecializovaných architektúr neuronových sietí [85, 67, 100, 56, 65]. Rekurentné siete RNN ako aj ich vylepšené varianty ako napr. Long Short-Term Memory (LSTM) [32] a Gated Recurrent Unit (GRU) [23] explicitne vytvorené na spracovanie dát vykazujúcich sekvenčný charakter ako je napr. text v prípade spracovania prirodzeného jazyka, sú v tomto ohľade prirodzenou voľbou nakoľko umožňujú modelovať temporálnu dynamiku rastu plodín [69, 93, 62]. Z tohto dôvodu napríklad v [62] autori študovali možnosti rekurentných sietí na klasifikáciu 18 tried plodín. Konkrétne išlo o porovnanie klasického mono-temporálneho prístupu s konvolučnou sieťou a multi-temporálnej klasifikácie pomocou RNN a LSTM siete. Konvolučná sieť priniesla lepšie klasifikačné výsledky zhruba v prvých 50 dňoch roku kedy pravdepodobne ešte nebolo dostatok pozorovaní pre modelovanie temporálnej dynamiky vegetácie pomocou RNN a LSTM. Zhruba od dňa 80 kvalita klasifikácie RNN a LSTM prevýšila kvalitu CNN siete a postupne sa navyšovala pri použití dodatočných pozorovaní až do dňa 200 kedy sa saturovala na stabilnej hodnote. Na druhú stranu kvalita klasifikácie CNN siete bola po celý rok zhruba rovnaká s malými fluktuáciami.

Postupný sekvenčný prístup k spracovaniu časovej rady kedy sa kontextuálna informácia z celej časovej rady postupne agreguje v rámci tzv. skrytého stavu (hidden state), je ústredným problémom aplikácie rekurentných sietí, nakoľko takýto prístup jednak zamedzuje efektívne využívať paralelizáciu ponúkanú grafickými akcelerátormi, ale taktiež zhoršuje modelovanie dlhodobých kontextuálnych závislostí. Self-attention mechanizmus navrhnutý v [84] ako hlavný stavebný blok architektúry Transformer priniesol revolúciu v oblasti spracovania prirodzeného jazyka. Možnosť dynamicky generovať attention skóre a to paralelne pre všetky prvky sekvencie zároveň, malo za následok významné zlepšenie modelovania dlhodobých kontextuálnych závislostí oproti dovtedy používaným rekurentným sieťam.

Motivovaní úspechmi architektúry transformera zvaného BERT [25] v oblasti spracovania prirodzeného jazyka, predovšetkým s dôrazom na prístup so self-supervizovaným predtrénovaním autori v [97] navrhli jeho použitie aj na klasifikáciu v časových radoch satelitných snímok. Hlavným cieľom tohto prístupu je aby sa model naučil užitočné spektrálno-temporálne reprezentácie obsiahnuté v časových radoch snímok a to bez prístupu k datasetu s ground truth. Následným dotrénovaním už klasickým supervizovaným prístupom na typicky menšom datasete, za využitia získaných váh, sa v druhej fáze zvanej fine-tuning model špecializuje na konkrétnu cieľovú úlohu ako napr. klasifikácia plodín. Ako tzv. pretext úlohu autorili navrhli predikciu vždy niekoľkých

náhodne vybraných prvkov časovej rady, ktoré boli kontaminované šumom. Experimentálne výsledky poukazujú na značné zlepšenie klasifikačných výsledkov v prípade aplikovania paradigmy self-supervizovaného predtrénovania. Na druhú stranu, keď spomínané predtrénovanie nebolo aplikované tak model dosahoval výsledky len mierne lepšie prípadne porovnateľné s výsledkami algoritmov RF a SVM, ktoré slúžili ako baseline modely. Ďalším zistením bolo, že alternatívne siete založené na 1D konvolúcii a LSTM bunke nedosahovali ani kvalitu klasifikácie algoritmov RF a SVM. Okrem iného autori skúmali možnosti agregácie reprezentácii prvkov časovej rady, ktoré boli výstupom transformera tak aby sa získala len jedna reprezentácia pre časovú radu ako celok. Navrhovaným bolo použitie tzv. CLS tokenu ale aj max a average pooling pri čom vo výsledku sa ukázal byť prístup s max poolingom najlepší v zmysle metriky overall accuracy.

Možnosti aplikácie Transformer architektúry na problematiku klasifikácie plodín v časových radách boli skúmané aj v [65] kde autori porovnávali viacero modelov medzi nimi napr. spomínaného Transformera, LSTM, TempCNN [56], hybridný DuPLO [35] model ale aj reziduálnu architektúru prispôsobenú na klasifikáciu časových rád [88]. Testy boli vykonané v dvoch nastaveniach. V prvom sa použili predspracované dáta Sentinel-2, kedy boli aplikované atmosferické korekcie a maskovanie oblačnosti. V druhom nastavení sa predspracovanie vynechalo. Porovnaním architektúr v spomínaných dvoch nastaveniach sa ukázala dominancia Transformer architektúry v prípade kedy nedošlo k predspracovaniu dát. Dodatočné štúdium spätnej propagácie gradientu a attention skóre potvrdilo domnienku, že Transformer ale aj rekurentný LSTM model sú robustné voči šumu v dátach, čo sa prejavuje ich schopnosťou automaticky odfiltrovať pozorovania znehodnotenú napr. oblačnosťou a zamerať sa len na kľúčové časti časovej rady na rozdiel od ostatných spomínaných modelov. Na druhú stranu v prípade predspracovania satelitných snímok dosahoval najlepšie výsledky model DuPLO pričom podľa autorov v tomto prípade nie je voľba metódy kritickým predpokladom pre dosiahnutie klasifikácie dostatočnej kvality.

V [70] Garnot et al. na riešenie úlohy klasifikácie plodín na poľnohospodárskych poliach z časových rád snímok S2 navrhli tzv. temporal attention encoder (TAE), ktorý vylepšoval klasický attention modul viacerými spôsobmi. Predovšetkým išlo o použitie tzv. master query, ktorá vznikla ako temporálny priemer query vektorov všetkých prvkov sekvencie čím sa zabezpečilo aby modul generoval jednotnú reprezentáciu pre celú časovú radu. Ide teda o alternatívu napr. k už spomínanej agregácii pomocou CLS tokenu prípadne pooling operácie. Okrem toho autori na zakódovanie pozícií použili hodnotu rozdielu dátumu snímky a vopred definovaného referenčného dátumu. Daný prístup priniesol zlepšenie jednak voči alternatívam založených na rekurentnej sieti ale aj voči metóde založenej na architektúre transformer použitej v [65]. Autori ďalej v [67] prišli s dodatočnými modifikáciami modelu a vytvorili tak verziu zvanú L-TAE, ktorá v každom ohľade prekonala TAE. Model dosahoval lepšie klasifikačné výsledky s výrazne nižším počtom parametrov a redukovala sa aj jeho výpočetná náročnosť v zmysle počtu floating point operácií (FLOPs). Modifikácie, ktoré priniesli úspech pozostávali z úplného nahradenie výpočtu master query ako temporálneho priemeru za master query, ktorá je priamo parametrom modelu ale aj použitia tzv. channel grouping stratégie. Zhang et al. [100] následne nadviazali na model L-TAE a obohatili ho o využitie dynamickej konvolúcie [92] na extrakciu lokálnych temporálnych vzorov, pracujúcej paralelne s L-TAE modulom. Daný prístup priniesol zlepšenie zhruba 2% v metrike mIoU pri klasifikácii LULC tried, ale len marginálne zlepšenie pri klasifikácii v rámci plodínového datasetu [90].

Na modelovanie čisto temporálnej dynamiky v dátach boli taktiež skúmané metódy využívajúce konvolúciu. Napr. v práci od Pelletier et al. [56] autori úspešne aplikovali 1D konvolúciu v časovom rozmere na klasifikáciu časových rád satelitných snímok. Ich architektúra zvaná TempCNN sa ukázala byť vhodnejšou na klasifikáciu časových rád snímok ako dovtedy používané rekurentné

siete. K podobným záverom ohľadom vhodnosti 1D CNN dospeli Zhong et al. [103] nakoľko v ich práci dosahoval prístup s 1D konvolyčnými vrstvami a inception modulom lepšie výsledky klasifikácie časových rád EVI indexov oproti prístupu s LSTM rekurentnou sieťou. Výkonnosť 1D CNN, LSTM, GRU a RF modelov na klasifikáciu piatich druhov plodín v časových radoch radarových dát skúmali aj v [101]. Model využívajúci 1D CNN dosiahol nepatrne lepšie celkové skóre oproti ostatným modelom v prípade využitia všetkých dostupných pozorovaní naprieč rokom. Experiment skúmajúci vplyv dĺžky časovej rady na celkovú presnosť modelov odhalil významnejší rozdiel vo výkonnosti a síce 1D CNN a RF dosahovali viditeľne vyššie hodnoty klasifikačných metrick oproti rekurentným modelom v prípade použitia len pozorovaní zo začiatku vegetačného obdobia. Nakoniec spomeňme ešte prístup spájajúci rekurentnú a konvolyčnú sieť do jedného celku s cieľom efektívne modelovať fenologické fázy plodín [50]. Daný model, zvaný Pixel R-CNN, operuje na úrovni pixelov, kedy najprv rekurentný encoder extrahuje temporálne vzorce obsiahnuté v časovej rade, ktoré sú následne hierarchicky spracované konvolyčnými vrstvami.

Prirodzene očakávanými rozšíreniami spracovania časových rád satelitných snímok, ktoré potenciálne zlepšia kvalitu sémantickej segmentácie plodín, sú architektúry schopné učiť sa súčasne temporálne a priestorové reprezentácie dát. V literatúre nájdeme priamočiare návrhy akým je napr. zlúčenie temporálnej a spektrálnej dimenzie časovej rady snímok [74] alebo zámena 2D konvolyčných vrstiev za ich viacrozmerné alternatívy [60, 36, 31]. Napr. Ji et al. [36] navrhli na segmentáciu plodín v časových radoch snímok vylepšiť topológiu VGG [73] o použitie 3D konvolúcie kedy sa filter posúva súčasne v priestorovej a časovej dimenzii. Výsledky ich práce poukázali na to, že takéto spracovanie časových rád snímok je z hľadiska výslednej presnosti klasifikácie plodín vhodnejšie v porovnaní s použitím 2D konvolúcie len v priestorovej doméne alebo 3D konvolúcie v priestorovej a spektrálnej doméne. V rámci článku [31] autori dokonca skúmali použitie 4D konvolúcie v architektúre U-Net na problém klasifikácie LULC v časových radoch snímok a porovnávali ho voči prístupu s použitím konvolúcie nižšieho rádu. Použitie 4D konvolúcie prinieslo najlepší výsledok za cenu zvýšených nárokov na výpočetný výkon. Je nutné podotknúť, že autori v prípade použitia 2D a 3D konvolúcie pracovali so snímkami tak, že zpriemerovali tú dimenziu, ktorá neposkytovala aktívnu informáciu. Napr. v prípade 2D prístupu autori agregovali časovú radu multispektrálnych snímok v časovej a spektrálnej dimenzii.

Hojne zastúpené sú aj pokusy o tvorbu akýchsi hybridných architektúr, typicky spájajúcich konvolyčné a rekurentné vrstvy [35, 64, 21, 63, 48, 83, 60, 102]. Napríklad už spomínaná architektúra DuPLO [35] pozostáva z dvoch paralelných častí resp. ramien kde prvé rameno tvorí konvolyčnú sieť zodpovednú za extrakciu priestorových príznakov a následne v druhom ramene operuje rekurentná bunka GRU s cieľom modelovať temporálnu dynamiku. Získané príznaky z oboch ramien sa následne spoja a spracujú klasifikačnou vrstvou. Autori RCNN [21] zvolili iný prístup a síce snímky v každom časovom kroku najskôr predspracuje konvolyčná sieť a následne je jej výstup použitý ako vstupný vektor príznakov do rekurentnej siete GRU.

Nahradením klasických lineárnych vrstiev za konvolyčné v rámci LSTM buniek umožnilo sieť zvanú convLSTM [72] lepšie modelovať priestorovo-temporálne charakteristiky v dátach a efektívnejšie tak predpovedať intenzitu zrážok. Takýto prístup našiel hneď svoje uplatnenie aj v oblasti mapovania LULC ako napr. v [64] kde bola na problematiku sémantickej segmentácie 17 tried plodín úspešne aplikovaná obojsmerná verzia (bidirectional) convGRU siete. Obojsmerný charakter mal eliminovať možný bias voči pozorovaniam vyskytujúcim sa neskôr v rámci časovej rady. Zaujímavým dôsledkom aplikácie convGRU resp. convLSTM siete na klasifikáciu v časových radoch satelitných snímok je možnosť, že malá časť rekurentných buniek, je schopná špecializovať sa na filtráciu oblačnosťou alebo inak znehodnotených pozorovaní [64, 63]. Takúto

možnosť podporujú aj výsledky z [102] kde autori skúmali vplyv chýbajúcich dát na výkon 5 modelov medzi ktorými boli 1D CNN, rekurentné siete a hybridné konvolučne-rekurentné siete. Ich experimenty odhaľujú, že aj pri 43.5% miere chýbajúcich pozorovaní v rámci časovej rady dosahujú 1D CNN, LSTM-CNN a GRU-CNN prijateľnú kvalitu klasifikácie s hodnotou metriky overall accuracy nad 85%. Ďalším prípadom kedy bola využitá convLSTM sieť na rozpoznávanie plodín je práca [60]. V tomto prípade autori navrhli použiť U-Net architektúru pričom temporálna dimenzia časovej rady snímok bola spracovaná práve spomínaným convLSTM encoderom. Model vo väčšine prípadov prekonal kvalitou segmentácie alternatívu v podobe 3D U-Net. Turkoglu et al. použili model convSTAR (convolutional STACKable Recurrent cell) [82] na hierarchickú segmentáciu plodín v troch úrovňach nomenklatury [83]. Použitie convSTAR modelu podľa autorov jednak prispelo k rýchlejšiemu a stabilnejšiemu trénovaniu na rozdiel od alternatív ako convLSTM a convGRU, ktoré výrazne trpia problémom miznúceho gradientu (vanishing gradient), ale zároveň postupný, hierarchický prístup ku klasifikácii mal podľa autorov zlepšiť mapovanie predovšetkým zriedkavo zastúpených tried plodín. Martinez et al. [48] sa snažili adresovať problém segmentácie plodín v tropických oblastiach, kde sa v rámci roka vystrieda niekoľko typov plodín. Z tohto dôvodu nadviazali na prácu [64] a navrhli použiť obojsmerný convLSTM model spolu s dodatočnou plne-konvolučnou sieťou ako napr. U-Net, zabezpečujúcou spracovanie priestorovej domény snímok. Takýto prístup umožnil tvoriť klasifikačnú mapu pre každý prvok časovej rady s použitím jedného modelu. Štúdiom klasifikácie plodín v časových radách snímok S2 na úrovni parciel pomocou konvolučnej, rekurentnej a hybridných konvolučne-rekurentných architektúr s rovnakým počtom parametrov v [69] autori empiricky ukázali, že najlepšie klasifikačné výsledky boli dosiahnuté pri súčasnom spracovaní časovej a priestorovej dimenzie snímok. Dôležitejším zistením bolo, že v prípade snímok S2 je vhodné alokovať väčšinu kapacity siete vo forme trénovateľných parametrov (autori uvádzajú 90%) na spracovanie temporálnej dynamiky obsiahnutej v dátach.

Ako už bolo spomenuté, prístupy založené na attention mechanizme vykazujú vo všeobecnosti lepšie schopnosti modelovania sekvencií v porovnaní s rekurentnými prístupmi. Aj z tohto dôvodu sa v niektorých prípadoch namiesto rekurentného modelu zvolila metóda založená práve na spomínanom attention module. Napr. sieť U-TAE [29] konkrétne spája U-Net a už spomínaný L-TAE modul [67], výpočetne menej náročnú verziu klasického attention modulu. Celkovo model dodržiava encoder-decoder prístup siete U-Net pričom temporálny encoder je aplikovaný v najnižšom priestorovom rozlíšení, nezávisle pre každú pixelovú pozíciu. Verzia L-TAE na rozdiel od klasického attention modulu je typu N-to-1 teda mapuje sekvenciu na jednu reprezentáciu, ktorá je spoločná pre celú časovú radu. Aj vďaka tejto vlastnosti L-TAE vyriešili autori problém skip spojení elegantne a to tak, že attention masku vypočítanú v najnižšom priestorovom rozlíšení upsamplingujú bilineárnou interpoláciou do každej vyššej úrovne, kde sa táto maska použije na temporálne priemerovanie časovej rady feature máp. Autori prišli aj s vlastným benchmarkovým datasetom PASTIS [29] pre panoptickú a sémantickú segmentáciu plodín z časových rád snímok S2 pričom na tomto datasete dosahuje model U-TAE hodnotu metriky mIoU 63.1%. Model prekonal alternatívy ako napr. 3D-Unet [60] ale aj ekvivalentné prístupy s U-Net a konvolučne-rekurentným temporálnym encoderom [60, 48]. SITS-Former [98] spája taktiež konvolučnú architektúru využívajúcu 3D konvolúcie, ktorá slúži na generovanie priestorovo-spektrálnych latentných reprezentácií vstupných patchov, ktoré sú následne spracované Transformerom. Samotný model bol navrhnutý s predpokladom využitia self-supervizovaného predtrénovania a následného fine-tuningu pričom predtrénovacia pretext úloha pozostáva z doplnenia centrálného pixela 5x5 patchov, ktoré boli odmaskované, na základe informácie zo zvyšných členov časovej rady. Výsledky opäť poukazujú na priaznivý vplyv takejto predtrénovacej procedúry. Hlavnou nevýhodou prístupu je nutnosť spracovať každý individuálny pixel v čase predikcie, čo je značne limitujúci faktor v prípade praktického

použitia tohto modelu.

Architektúra TSViT [78] teda Temporo-Spatial Vision Transformer je jedinou, ktorá pozostáva len z transformer blokov operujúcich ako v časovej tak aj priestorovej doméne snímok. Autori zistili, že v prípade spracovania časových rád snímok je vhodnejším zostaviť model explicitne tak aby bola najskôr spracovaná a extrahovaná temporálna informácia a až následne priestorová nakoľko v prípade zmeny poradia faktorizácie modelu došlo k výraznej degradácii kvality segmentácie a to až o 30% v prípade metriky mIoU. Toto zistenie podporuje domnienku, že priestorový kontext je v prípade snímok so stredným a nízkym rozlíšením menej dôležitý pre efektívnu klasifikáciu plodín na rozdiel od temporálnej dimenzie. Okrem iného, prospešným sa ukázalo byť použitie K CLS tokenov kde K reprezentuje počet tried plodín, čo modelu pravdepodobne umožní zameriavať sa na príznaky špecificky spojené s konkrétnou triedou plodiny. Dodajme, že hlavnou nevýhodou TSVit je ako v predchádzajúcom prípade vysoká výpočetná a pamäťová náročnosť v čase predikcie.

Cai et al. [19] prišli s novým a uceleným frameworkom pre učenie reprezentácii z časových rád snímok. V ich prístupe konceptuálne oddeľujú spracovanie časovej a priestorovej dimenzie snímok pričom proces učenia temporálnych reprezentácii je rozdelený do troch špecifických krokov založených na využití tzv. group propagation bloku (GP Block) [94]. V tomto procese autori uvažujú N tzv. temporálnych kontextuálnych zhhlukov (temporal context clusters), čo sú akési parametre modelu do ktorých sa v priebehu učenia modelu agreguje relevantná informácia súvisiaca s fenológiou plodín. Konkrétne táto agregácia prebehne v prvom kroku zvanom COLLECT a to na základe cross-attention operácie medzi vstupnou časovou radou a temporálnymi kontextuálnymi zhhlukmi. Ďalší krok (UPDATE) pozostáva zo spracovania extrahovanej informácie v rámci temporálnych kontextuálnych zhhlukov pomocou série MLP vrstiev. V poslednom kroku (DISTRIBUTE) dôjde k spätnej propagácii informácie zo zhhlukov do pôvodnej časovej rady opäť pomocou cross-attention operácie. Experimentálne výsledky dokazujú, že explicitné rozdelenie do spomínaných troch krokov je efektívnejšou formou extrakcie temporálnych reprezentácii z časových rád snímok. Výhodou daného prístupu je jednak možnosť explicitného predtrénovania temporálneho encoderu zvaného Exchanger napr. na pamäťovo nenáročnom pixel-set formáte snímok [70] ale aj to, že na extrakciu priestorových príznakov a samotnú segmentáciu je možné použiť ľubovoľný segmentačný model. Podľa našich vedomostí ide v dobe písania tejto práce o SOTA metódu čo sa týka sémantickej segmentácie plodín z časových rád snímok Sentinel-2 na datasete PASTIS [29] pričom model dosahuje hodnotu mIoU až 67.9%.

Mapovanie typov plodín za pomoci strojového učenia sa nezaobíde bez prístupu k tréningovému datasetu. V nedávnej dobe sa vďaka prístupu k dátam zo systému pre identifikáciu parciel LPIS v rámci EÚ zvýšil počet benchmarkových datasetov typov plodín. Samotné dáta LPIS vznikajú deklaráciami farmárov o type plodiny na korešpondujúcich parcelách a preto môžu obsahovať chyby. Aj napriek tomuto nedostatku patrí systém LPIS medzi veľmi cenné zdroje dát pre tvorbu plodinových datasetov. V nasledujúcich bodoch zhrnieme niektoré datasety využívajúce pre tvorbu ground truth informácie z národného LPIS systému.

- TimeSen2Crop [90] je dataset typov plodín pre Rakúsko pozostávajúci z časových rád nezávislých pixelov zo snímok Sentinel-2 od septembra 2017 do augusta 2018. Každý pixel v rámci časovej rady reprezentuje medián v danom mesiaci. Dataset obsahuje celkovo 16 typov plodín.
- BreizhCrops [66] obsahuje hodnoty časových rád priemerných spektrálnych odoziev pásiem Sentinel-2 pre poľnohospodárske polia z časti Francúzska naprieč rokmi 2017 a 2018.

Ground truth popisuje 9 typov plodín.

- DENETHOR [39] je ďalším datasetom typov plodín tentoraz pre severnú časť Nemecka popisujúci 9 typov plodín pre roky 2018 a 2019. Okrem časovej rady snímok Sentinel-2 sú obsiahnuté aj časové rady radarových snímok Sentinel-1 a komerčných snímok v 3-metrovom rozlíšení nazývaných Planet.
- Sen4AgriNet [76] je viac-ročný dataset pokrývajúci regióny vo Francúzsko a Španielsku pre roky 2016 až 2020. Obsiahnuté sú jednak patche snímok Sentinel-2 vhodné pre sémantickú segmentáciu ako aj agregované dáta na úrovni parcel. Nomenklatúra datasetu uvažuje až 168 tried.
- ZueriCrop [83] pozostáva z časovej rady patchov snímok Sentinel-2 rozmerov 24x24 pixelov nad Švajčiarskom pre rok 2019. Zahrnuté sú triedy typov plodín v troch hierarchických úrovňach pričom najdetailnejšia úroveň obsahuje 48 tried.
- S2-Agri [70] dataset obsahuje časové rady parcel extrahovaných zo Sentinel-2 dlaždice T31TFM nad južnou časťou Francúzska. Časové rady obsahujú 24 záznamov od januára do októbra 2017. Samotný formát je dvojakého druhu. Jednak vo forme patchov kedy bolo každé poľnohospodárske pole vyrezané do patchu rozmerov 32x32 pixelov ale aj vo forme tzv. pixel setu kedy sa z každého poľa navzorkovalo N reprezentatívnych pixelov. Nomenklatúra datasetu uvažuje 20 tried plodín.
- PASTIS [29] je benchmarkový dataset slúžiaci pre sémantickú a panoptickú segmentáciu obsahujúci 18 typov plodín z niekoľkých častí Francúzska pre rok 2019 pričom pozostáva z časových rád patchov snímok Sentinel-2. Novšia verzia tzv. PASTIS-R [68] obsahuje navyše aj časové rady patchov radarových snímok Sentinel-1.
- AgriSen-COG [71] podobne ako Sen4AgriNet obsahuje časovú radu patchov snímok Sentinel-2 ale aj časovú radu agregovaných dát pre jednotlivé parcely. Namiesto priemeru a štandardnej odchýlky autori využívajú na agregáciu metódu DBA [57]. Snímky pokrývajú až 5 krajín EÚ pre roky 2019-2020. Kvôli zvýšeniu kvality datasetu autori aplikovali metódu na detekciu anomálií pre zistenie možných chýb v tvorbe ground truth.

Kapitola 2

Dátové zdroje a datasety

Jedným z cieľov našej práce je vytvoriť dataset typov plodín pre ČR, nazvaný S2TSCZCrops (Sentinel-2 Time Series for semantic segmentation of Czech Crops), pozostávajúci z časových rád satelitných snímok. Daný dataset ako aj dátové zdroje použité na jeho tvorbu v tejto kapitole priblížime. Okrem toho popíšeme charakteristiky datasetu PASTIS, ktorý slúžil ako inšpirácia pre tvorbu S2TSCZCrops ale aj ako dodatočný benchmarkový dataset pre porovnanie so SOTA metódami.

2.1 Satelitné snímky

Ako už bolo spomenuté, v našej práci sme ako podkladové dáta vybrali snímky z misie Sentinel-2 [12], ktorá je súčasťou programu Copernicus a pozostáva z konštelácie dvoch satelitov na helio-synchrónnej orbite. Hlavnou úlohou tejto misie je systematické monitorovanie pevniny. Optický senzor MSI [13] na palube satelitov zaznamenáva až 13 spektrálnych pásiem pričom pásma B5, B6 a B7 reprezentujú tzv. vegetačný červený okraj, oblasť elektromagnetického spektra charakteristickú náhlym skokom odrazivosti (reflectance) zelenej vegetácie. Z tohto dôvodu sú multispektrálne snímky S2 obzvlášť vhodné na monitorovanie stavu vegetácie. Okrem toho snímky S2 vynikajú aj v časovom rozlíšení teda v čase za ktorý družica nasníma rovnaké územie, čo umožňuje efektívne sledovať temporálnu dynamiku spektrálnej odozvy plodín spojenej s ich fenológiou. Konkrétne konštelácia oboch satelitov umožňuje dosahovať časové rozlíšenie 5 dní pričom v stredných zemepisných šírkach sa doba skrakuje len na 2-3 dni [12]. Charakteristiky spektrálnych pásiem S2 sú uvedené v tab. 2.1.

Dáta z MSI senzorov prechádzajú viacerými úrovňami predspracovaniami ako napr. radiometrické, geometrické a atmosferické korekcie vyúsťujúce do niekoľkých nadvazujúcich produktov [13]. Pre koncových užívateľov sú najrelevantnejšie produkty v úrovni spracovania Level-1C (L1C) a Level-2A (L2A), ktoré sú rozložené do tzv. dlaždíc (tile) s pevne určenou polohou v pravidelnej mriežke. Každá dlaždica je reprezentovaná štvorcovým ortorektifikovaným rastrom v projekcii UTM/WGS84 zaberajúcim 110 x 110 km². Samotné dlaždice majú na okraji prekryv s vedľajšími dlaždicami zhruba 10 km. Jednotlivé produkty L1C a L2A sa líšia v úrovni atmosferickej korekcie. Produkt L1C reprezentuje tzv. top of atmosphere (TOA) odrazivosť (reflectance). Po aplikácii viacerých atmosferických korekcií vzniká L2A produkt poskytujúci tzv. bottom of atmosphere (BOA) odrazivosť. Všetky produkty poskytované misiou Sentinel-2 sú kompaktné uložené v súborovom kontajnery formátu SAFE. Okrem samotných obrazových dát vo formáte JPEG2000 pre jednotlivé spektrálne pásma sú v kontajnery obsiahnuté aj doplnkové dáta z družicovej tele-

Pásmo	Centrálna λ nm	Šírka - $\Delta\lambda$ nm	Rozlíšenie - m	Poznámka
B1	443	20	60	Detekcia rozptylu aerosólu
B2	490	65	10	Modrá
B3	560	35	10	Zelená
B4	665	30	10	Červená
B5	705	15	20	Vegetačný červený okraj
B6	740	15	20	Vegetačný červený okraj
B7	783	20	20	Vegetačný červený okraj
B8	842	115	10	Blízka infračervená (NIR)
B8A	865	20	20	Úzka blízka infračervená
B9	945	20	60	NIR - Absorpcia vodnej pary
B10	1375	30	60	SWIR - Detekcia tenkého cirrusu
B11	1610	90	20	Krátko-vlnná infračervená (SWIR)
B12	2190	180	20	SWIR

Tabuľka 2.1: Charakteristika spektrálnych pásiem vzhľadom k spektrálnemu a priestorovému rozlíšeniu získavanými inštrumentom MSI na palube družíc Sentinel-2 [2].

metrie, pomocné metadáta ale aj indikátory kvality ako masky oblačnosti a defektívnych pixelov. Produkt L2A obsahuje navyše raster s klasifikáciou scény do 12 tried [13].

Prístup k spomínaným produktom je možný po bezplatnej registrácii napr. cez dátový portál <https://scihub.copernicus.eu/dhus> prípadne v rámci dátového skladu prevádzkovaného CESNET-om <https://dhr1.cesnet.cz>. Okrem toho je možné využiť aplikačné programovateľné rozhrania (API) [11] pre účely automatizácie sťahovania.

2.2 Dataset PASTIS

Dataset PASTIS [29] pozostáva z časových rád multispektrálnych snímok Sentinel-2 a pokrýva rozličné časti Francúzska. Dataset má slúžiť ako benchmark pre úlohu panoptickú a sémantickú segmentáciu typov plodín. Celkovo je v datase 2433 časových rád snímok resp. tzv. patchov pričom rozmery každého patchu v časovej rade sú 128×128 pixelov. Použitých je 10 z celkovo 13 spektrálnych pásiem Sentinel-2 pričom vynechané sú pásma B01, B09, B10 slúžiace na monitorovanie atmosferických charakteristík. Jednotlivé snímky sú v 10m rozlíšení a korešpondujú s pozorovaniami BOA. Snímky v časovej rade boli získané medzi septembrom 2018 až novembrom 2019. Každá časová rada obsahuje medzi 38 - 61 pozorovaniami. Časové rozloženie jednotlivých pozorovaní v časových radách je nerovnomerné pričom mediánový interval je približne 5 dní. Pri zbere snímok bol aplikovaný filter na oblačnosť, ale niektoré pozorovania stále obsahujú čiastočné alebo úplné pokrytie oblačnosťou.

Každému pixelu v datase je priradená jedna z 18 možných tried typov plodín pričom navyše je trieda Background a Void. Informácie o type plodiny boli získané z francúzskeho národného systému na identifikáciu parciel LPIS. Každý pixel, ktorý neprislúcha triede Background má ďalej priradený kód resp. označenie inštancie nakoľko dataset slúži aj pre úlohu panoptickú segmentáciu. Všetkých 2433 časových rád je rozdelených do 5 foldov pre prípadné testovanie metódou krížovej validácie.

Označenie a kód	Poznámka
Background 0	Pozadie
Meadow 1	Lúka
Soft winter wheat 2	Pšenica letná (ozimná)
Corn 3	Kukurica
Winter barley 4	Jačmeň ozimný
Winter rapeseed 5	Repka olejná (ozimná)
Spring barley 6	Jačmeň jarný
Sunflower 7	Slničnica
Grapevine 8	Vinič
Beet 9	Repa
Winter triticale 10	Tritikale ozimné
Winter durum wheat 11	Pšenica tvrdá (ozimná)
Fruits, vegetables, flowers 12	Ovocie, zelenia, kvety
Potatoes 13	Zemiaky
Leguminous fodder 14	Strukovinové krmne plodiny
Soybeans 15	Sójové bôby
Orchard 16	Ovocný sad
Mixed cereal 17	Zmiešané obilniny
Sorghum 18	Cirok
Void 19	Parcely väčšinou mimo patch

Tabuľka 2.2: Popis nomenklatury datasetu PASTIS.

2.3 Dataset S2TSCZCrops

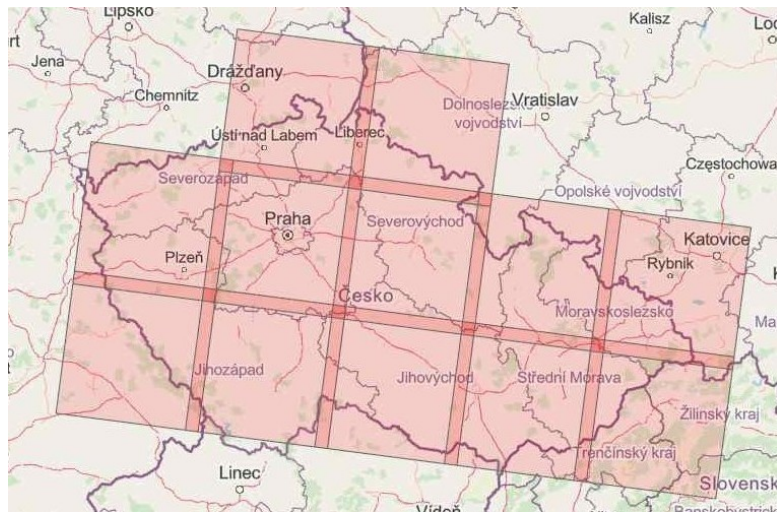
S2TSCZCrops (Sentinel-2 Time Series for semantic segmentation of Czech Crops) je dataset slúžiaci na tréovanie a porovnávanie modelov určených na sémantickú segmentáciu plodín zo snímok Sentinel-2 pre Českú republiku. Pri jeho tvorbe sme sa inšpirovali štruktúrou datasetu PASTIS. Celkovo obsahuje 6945 časových rád snímok rozdelených do tréovacej, validačnej a testovacej podmnožiny pričom jednotlivé charakteristiky tohto datasetu budú priblížené v nasledujúcich sekciách.

2.3.1 Oblasť záujmu

Plodinové ground truth dáta, ktoré máme k dispozícii pokrývajú práve celú Českú republiku, ktorá z tohto dôvodu reprezentuje nami zvolenú oblasť záujmu. Snímky sme získavali vo forme dlaždíc Sentinel-2 v súradnicovom referenčnom systéme UTM 33N pričom pokrytie oblasti záujmu použitými dlaždiciami je zobrazené na obr. 2.1. Dlaždice boli získané z časového okna od začiatku septembra 2018 až po koniec októbra 2019. Prekryv temporálneho okna s rokom 2018 bol vykonaný nakoľko v datasete uvažujeme aj ozimné plodiny, ktorých výsadba sa typicky vykonáva už na jeseň predchádzajúceho roku.

2.3.2 Referenčné dáta

Referenčné informácie o type plodín na poliach v rámci ČR boli poskytnuté Štátnym poľnohospodárskym intervenčným fondom ČR (SZIF) a vznikali na základe klasifikácie typov plodín na



Obr. 2.1: Pokrytie oblasti záujmu dlaždicami S2.

poliach extrahovaných zo systému na identifikáciu parciel LPIS [3]. Dáta LPIS v rámci ČR sú od roku 2022 verejne dostupné a obsahujú geopriestorové atribúty vo vektorovom formáte (.shp) popisujúce hranice poľnohospodársky využívanej pôdy ako aj jej využitie deklarované jednotlivými farmármi uchádzajúcimi sa o dotácie. Základnou jednotkou LPIS je tzv. pôdny blok (PB) reprezentujúci súvislú poľnohospodársky využívanú plochu viditeľne oddelenú od iných terénnych prvkov. Evidujú sa u neho štyri všeobecné typy využitia a to *Zemědělsky obhospodařovaná půda*, *Ekologicky významný prvek*, *Zalesněná půda* a *Vodní plocha*. Pôdny blok sa ďalej delí na menšie úseky tzv. diely pôdneho bloku (DPB), ktoré sú už najmenšou jednotkou na ktorých je pestovaná samostatná plodina [4]. Podotknime, že delenie poľnohospodárskej pôdy v LPIS nesúvisí s členením parciel v rámci katastra. Vo všetkých prípadoch kedy používame slovo parcela máme preto na mysli DPB. Aj keď je DPB najmenšou jednotkou LPIS tak v nej bohužiaľ nie je zaznačený typ pestovanej plodiny, ale len tzv. kultúra ako môžeme vidieť v prílohe 6.3 v tab. 6.1. Z tohto dôvodu sa pri tvorbe referenčných dát pre ČR pristúpilo ku dodatočnej klasifikácii pomocou metódy Sen2Agri [6], ktorá vznikla ako súčasť projektu Sen4CAP (Sentinels for Common Agriculture Policy) [7], kde bola Česká republika jednou z pilotných testovacích krajín. Sen2Agri systém používa algoritmus random forest pre klasifikovanie typov plodín na úrovni objektov (parciel). Ako príznaky sú využité časové rady agregovaných štatistík z radarových a multispektrálnych snímok S1 a S2 [9, 6]. V rámci pilotného testovania boli plodinové mapy vzniknuté klasifikáciou, validované SZIFom na základe in-situ dát s výsledkom, že presnosť (overall accuracy) výslednej klasifikácie dosahuje v ČR pre rok 2019 hodnotu viac ako 90% [10, 8]. Okrem toho validačný report obsahuje navyše aj niekoľko odporúčaní na reklasifikáciu resp. zlúčenie niektorých typov plodín kvôli častej zámennosti a podobnej fenológii.

Výsledná plodinová mapa pre ČR v roku 2019, ktorú máme k dispozícii uvažuje vo svojej nomenklatúre 14 tried plodín. Danú nomenklatúru preberáme do datasetu S2TSCZCrops pričom jej celkové zhrnutie je v tabuľke 2.3. Podotknime, že trieda Not Classified popisuje neklasifikovanú ornú pôdu predovšetkým kvôli malým rozmerom DPB. V istom zmysle problematickou je aj trieda Other crops nakoľko reprezentuje minoritné plodiny, ktorých typ nie je známy. Názvy tried v rámci poskytnutej plodinovej mapy, počty parciel a ich spoločná výmera sú uvedené v prílohe 6.3 v tabuľke 6.2.

Označenie a kód	Poznámka
Background 0	Pozadie
Grassland 1	Trávnaté oblasti
Fruit & Vegetables 2	Ovocie a zelenina
Summer cereals 3	Jarné obilniny
Winter cereals 4	Ozimné obilniny
Rapeseed 5	Repka olejná (ozimná)
Maize 6	Kukurica
Forage crops 7	Kŕmne plodiny
Sugar beet 8	Cukrová repa
Flax & Hemp 9	L'an a Konope
Permanent fruit 10	Trvalé ovocné sady
Hopyards 11	Chmelnice
Vineyards 12	Vinice
Other crops 13	Ostatné plodiny
Not classified 14	Neklasifikované pixely

Tabuľka 2.3: Popis nomenklatúry datasetu S2TSCZCrops.

2.3.3 Temporálne charakteristiky

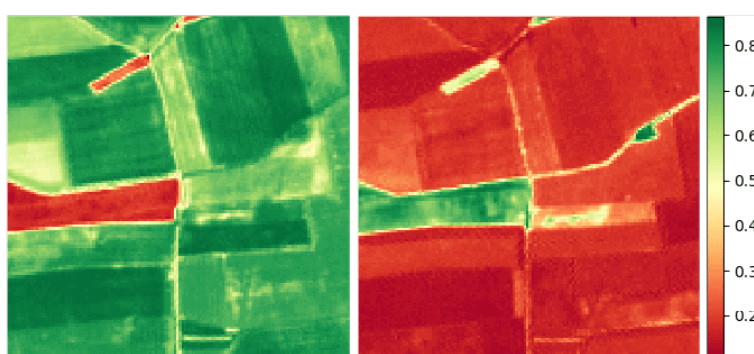
Jedinečnosť fenológie jednotlivých plodín je dôvod prečo pri probléme klasifikácie uvažujeme aj temporálny rozmer snímok. Častým spôsobom ako reprezentovať a zachytiť jednotlivé fenologické fázy plodín je graf časovej rady priemernej hodnoty spektrálneho indexu NDVI charakterizujúceho množstvo chlorofylu obsiahnutého v rastlinách. NDVI je normalizované do intervalu $(-1, 1)$ pričom hodnoty ≤ 0.1 reprezentujú povrch bez zelených rastlín ako napr. skaly, sneh, prípadne vodné plochy. Hodnoty v intervale $(0.1, 0.4]$ korešpondujú s povrchom s nízkym zastúpením zelených rastlín. Môže ísť o trávou pokryté oblasti, prípadne o plodiny v rannej fáze rastu. Vyššie hodnoty NDVI napr. v intervale $(0.4, 1.0]$ charakterizujú povrchy s vysokým zastúpením zelene resp. zelenej vegetácie. Na obrázku 2.2 sú zobrazené snímky toho istého územia v dvoch rozličných časoch a ich korešpondujúce NDVI indexy. Na snímkach je viditeľý kontrast medzi poľom v strednej časti na ktorom je vysadená cukrová repa a zvyškom poľnohospodárskej oblasti kde sú prevažne ozimné obilniny a repka olejná (ozimná). Rozdielnosť v hodnotách NDVI súvisí s posunom vegetačných období ozimných plodín a cukrovej repy (vid' obr. 2.3).

Pre účely analýzy temporálnej dynamiky tried plodín sme využili parcely v rámci dlaždice T33UVR pričom do výpočtu priemernej hodnoty NDVI boli započítané len pixely z vnútorných oblastí parciel aby sa zabránilo znečisteniu spektrálnej odozvy z hraničných pixelov. Kvôli prehľadnosti sú na grafe zobrazené len štyri temporálne profily. Z grafu 2.3 je zrejmé, že ozimné obilniny a repka olejná sú koncom mája takmer na vrchole vegetačného obdobia (vysoká hodnota NDVI) pričom cukrová repa len v jeho začiatkoch (nízka hodnota NDVI). Na druhú stranu koncom júla dochádza k zberu ozimných plodín pričom cukrová repa pre zmenu dosahuje vrchol svojho vegetačného obdobia.

Temporálne profily zvyšných plodín sú dostupné na https://github.com/Many98/Crop2Seg/tree/main/data/temporal_profiles. Z vizuálnej inšpekcie daných grafov vyplýva jedinečnosť fenológie väčšiny uvažovaných tried plodín. Niektoré triedy ako napr. Permanent fruit a Vineyards vykazujú podobné trendy v temporálnych profiloch. Rozdiel v krivkách NDVI týchto plodín je len v miernom posune nahor vo väčšine vegetačného obdobia, čo súvisí pravdepodobne



(a) Sentinel-2 snímky



(b) NDVI

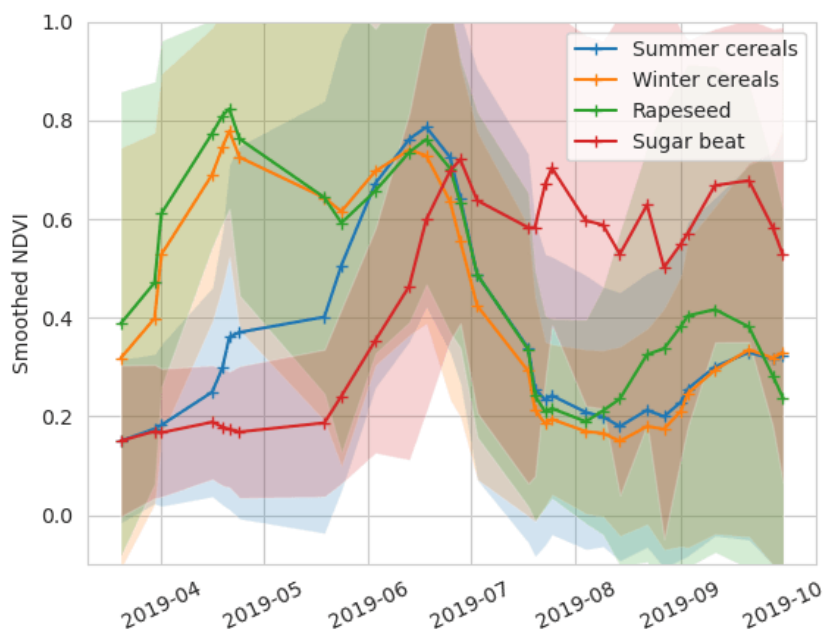
Obr. 2.2: Na obrázku a) vidíme zľava snímky z 24.5.2019 a 23.7.2019 v prirodzených farbách. Ďalej na obrázku b) sú vyobrazené korešpondujúce NDVI heatmapy.

s väčším objemom zelene v prípade ovocných sádov. Takmer identické temporálne charakteristiky vykazujú aj trieda Grassland a Forage Crops, čo môže znamenať problematickú separabilitu týchto tried. Taktiež trieda Other crops zdieľa podobné fenologické charakteristiky s triedami Flax & Hemp a Annual fruit & vegetable.

2.3.4 Implementačné detaily

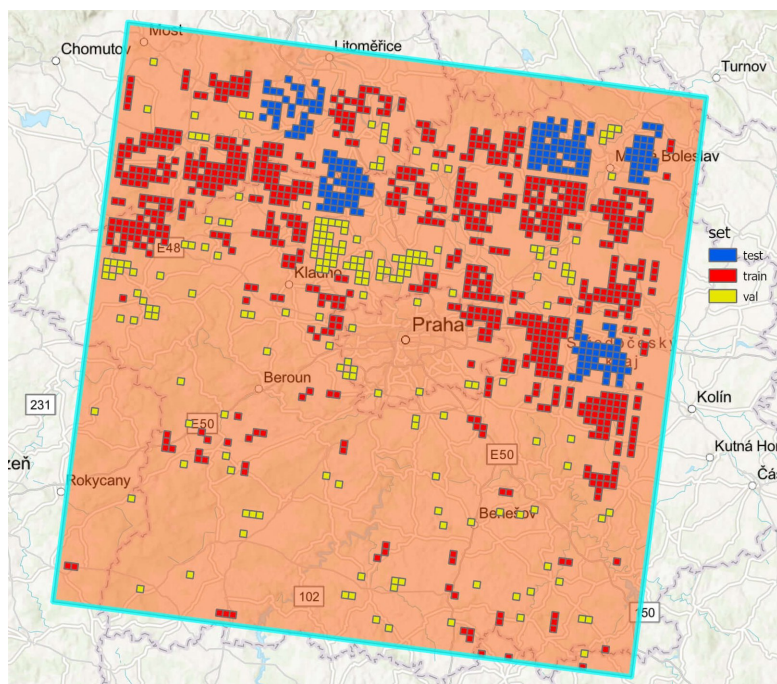
Jednotlivé pozorovania Sentinel-2 (dlaždice rozmerov 10980x10980 pixelov) boli získavané nepravidelne od začiatku septembra 2018 až po koniec októbra 2019 a to v závislosti na dostupnosti snímok. Počet pozorovaní na jednu dlaždicu bol obmedzený na maximálne 5 za mesiac. Okrem toho boli dlaždice s vysokým zastúpením oblačnosti prípadne snehu vylúčené z ďalšieho spracovania. Celkovo sme získali časové rady, ktoré obsahovali 27-61 pozorovaní S2.

Ako už bolo spomenuté tak pri tvorbe sme sa inšpirovali štruktúrou datasetu PASTIS. Z tohto dôvodu sme samotné dlaždice rozdelili na menšie tzv. patche rozmerov 128 x 128 pixelov s 10 spektrálnymi pásmami. Vylúčili sme spektrálne pásma B01, B09, B10. Snímky sú v 10m rozlíšení a korešpondujú s BOA odrazivosťou. Iný typ predspracovania snímok nebol využitý. Podotkneme ešte, že nakoľko časť dlaždíc bola čiastočne pokrytá oblačnosťou tak tento typ šumu sa preniesol aj na vzniknuté patche a z tohto dôvodu sú niektoré patche v rámci časových rád úplne pokryté oblačnosťou. Ako uvidíme ďalej model bol schopný naučiť sa takéto pozorovania ignorovať.

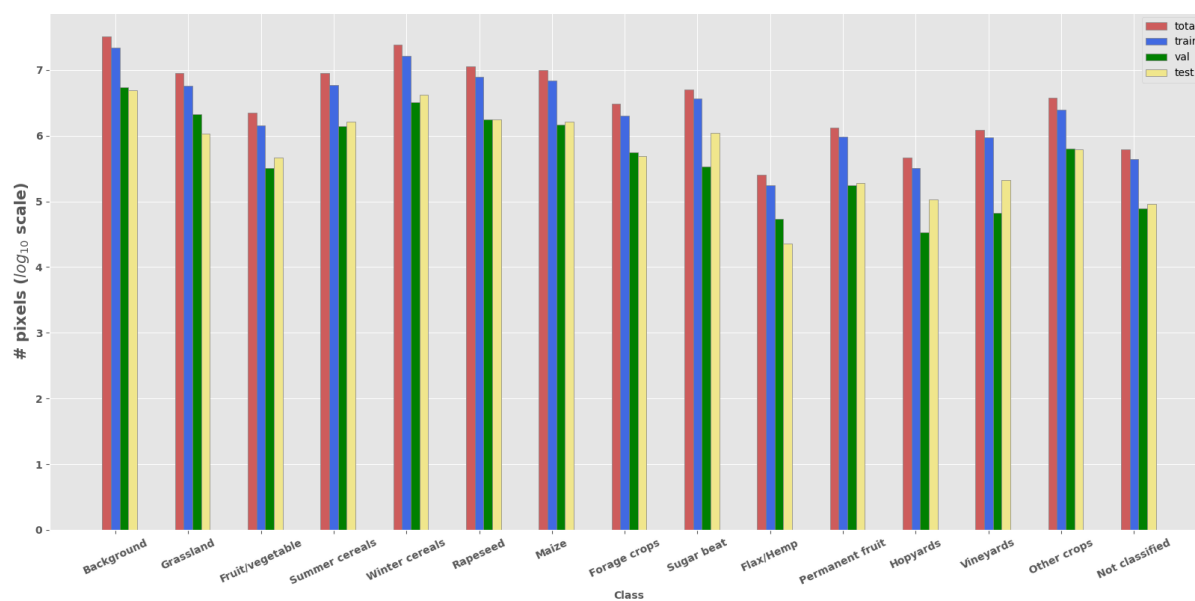


Obr. 2.3: Temporálne profily tried Summer cereals, Winter cereals, Rapeseed a Sugar beet.

V prípade datasetu S2TSCZCrops sme zvolili štandardný spôsob delenia na tri časti v pomere zhruba 70:15:15 na tréningový, validačný a testovací. Rovnaký pomer bol použitý pre každú z uvažovaných S2 dlaždíc. Aby sa zabránilo znečisteniu validačného a testovacieho datasetu dátami z tréningového datasetu vyžadovali sme aby pri vzorkovaní boli príslušné patche zaradené vždy do rovnakého pod-datasetu. Ďalej sme z datasetu vylúčili patche, ktoré boli úplne pokryté majoritnými triedami Background a Grassland. Konkrétne rozdelenie patchov v rámci dlaždice T33UVR do jednotlivých pod-datasetov je zobrazené na obrázku 2.4. V konečnom dôsledku dataset pozostáva zo 6945 patchov časových rád S2 a z toho je 4734 tréningových, 1081 validačných a 1130 testovacích. Distribúcia pixelov v jednotlivých pod-datasetoch je zobrazená v 2.5. Najnižšie zastúpenie má trieda Flax & Hemp s rádovo 10^5 pixelmi. Background trieda ma naopak najvyššie zastúpenie v datasete s rádovo 10^7 pixelmi.



Obr. 2.4: Rozdelenie patchov do trérovacieho, validačného a testovacieho datasetu v rámci dlaždice T33UVR



Obr. 2.5: Distribúcia počtu pixelov pre triedy plodín v datasete S2TSCZCrops.

Kapitola 3

Metodika

V tejto kapitole bližšie uvedieme detaily výpočtu attention mechanizmu a architektúry U-TAE [29] z ktorých pri návrhu modelov vychádzame. Ďalej zhrnieme špecifiká architekturných prvkov nami navrhnutých modelov ako aj metód dosahujúcich SOTA výsledky v oblasti mapovania plodín z časových rád snímok. Uvedené budú taktiež implementačné detaily ako aj metodika experimentovania.

3.1 Teoretické pozadie

Aj keď v našej práci pracujeme len s pozmenenou verziou attention mechanizmu, menovite L-TAE [67], ktorá umožňuje efektívne extrahovať temporálne príznaky z dát a nie priamo s verziou, ktorá je základným pilierom architektúry Transformer [84] tak považujeme za vhodné na tomto mieste, čitateľovi priblížiť základné princípy spojené s jej výpočtom.

Vstupné dáta t.j. časové rady snímok Sentinel-2 uvažujeme vo forme tenzorov $\mathbf{X} \in \mathbb{R}^{t \times C \times H \times W}$ kde t je počet prvkov časovej rady, C reprezentuje počet spektrálnych kanálov a následne H a W sú postupne výška a šírka snímky. Všeobecne ale možno uvažovať vstupnú sekvenciu dát vo forme matíc $\mathbf{X} \in \mathbb{R}^{t \times d}$ kde t je opäť počet prvkov sekvencie a d korešponduje s dimenziou prvkov sekvencie.

3.1.1 Attention mechanizmus

Ako názov napovedá, môžeme attention mechanizmus neformálne popísať ako metódu, ktorá umožňuje modelu sústrediť sa na špecifické alebo v istom ohľade dôležité charakteristiky vstupných dát. Spomínaná schopnosť „sústrediť sa“ je typicky realizovaná priradením váh jednotlivým prvkom a ich následnou kombináciou na základe priradených váh. Prejdime teraz k formálnejšiemu popisu attention.

Uvažujme sekvenciu t vektorov dimenzie d_v zvanú *values* a usporiadanú do matice

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t]^T \in \mathbb{R}^{t \times d_v} \quad (3.1)$$

Ďalej uvažujme sady tzv. *query* a *key* vektorov dimenzii d_q taktiež usporiadaných do matíc

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_t]^T \in \mathbb{R}^{t \times d_q} \quad (3.2)$$

$$\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_t]^T \in \mathbb{R}^{t \times d_q} \quad (3.3)$$

Následne výpočet novej reprezentácie pre všetkých t prvkov vstupnej sekvencie resp. tzv. embeddingu $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_t]^T$ pomocou attention mechanizmu pozostáva z váženej sumy *values* vektorov \mathbf{v}_i pričom ako váhy slúži attention skóre $\mathbf{a}_i = f(\mathbf{q}_i, \mathbf{K}) \in [0, 1]^T$ kde $i \in 1, \dots, t$, ktoré meria podobnosť medzi *query* a *key* vektormi. Celkovo je teda výpočet attention možné zhrnúť v nasledujúcom maticovom zápise

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{E} = \mathbf{A}^T \mathbf{V} \in \mathbb{R}^{t \times d_v} \quad (3.4)$$

kde $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_t]^T \in \mathbb{R}^{t \times t}$ pričom platí $\sum_n^T \mathbf{a}_{i,n} = 1$.

Rozličné typy attention mechanizmu sa líšia predovšetkým vo forme výpočtu matice attention skóre \mathbf{A} prípadne matíc \mathbf{Q}, \mathbf{K} a \mathbf{V} . Autori v [84] rozlišujú dva typy attention mechanizmu. Prvým z nich je self-attention kedy je spomínaná *query* matica \mathbf{Q} získaná zo vstupnej sekvencie $\mathbf{X} \in \mathbb{R}^{t \times d}$ lineárnou transformáciou $\mathbf{Q} = \mathbf{X} \mathbf{W}_Q$ kde $\mathbf{W}_Q \in \mathbb{R}^{d \times d_q}$. Ďalším typom je tzv. cross-attention kedy sa *query* matica získa lineárnou transformáciou z inej ako vstupnej sekvencie. Podobne ako \mathbf{Q} v prípade self-attention sú získané aj matice \mathbf{V} a \mathbf{K} lineárnou transformáciou vstupnej sekvencie \mathbf{X} . Konkrétne $\mathbf{K} = \mathbf{X} \mathbf{W}_K$ kde $\mathbf{W}_K \in \mathbb{R}^{d \times d_q}$ a $\mathbf{V} = \mathbf{X} \mathbf{W}_V$ kde $\mathbf{W}_V \in \mathbb{R}^{d \times d_v}$. Attention skóre je následne vypočítané ako *softmax* škálovaného skalárneho súčinu medzi \mathbf{Q} a \mathbf{K}

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_q}}\right) \in \mathbb{R}^{t \times t} \quad (3.5)$$

kde d_q je už spomínaná dimenzia *query* a *key* vektorov.

Aby sa v procese učenia pomocou attention mechanizmu umožnilo modelu špecializovať sa na rôzne aspekty vstupnej sekvencie tak je v praxi výpočet attention vykonaný paralelne a nezávisle h -krát pričom každý výpočet má vlastnú podskupinu parametrov s dimenziami $d_v = d_q = \frac{d}{h}$. Každá podskupina parametrov resp. každý nezávislý výpočet sa označuje ako hlava a celkovo je tento prístup nazývaný ako Multi-Head Attention [84]. Výsledné reprezentácie z každej hlavy sú spojené do jednej (pomocou operácie spojenia resp. concatenation) a transformované lineárnou vrstvou.

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{concat}[\text{head}_1, \dots, \text{head}_h] \mathbf{W}_{out} \quad (3.6)$$

kde $\text{head}_i = \text{Attention}(\mathbf{X} \mathbf{W}_Q^i, \mathbf{X} \mathbf{W}_K^i, \mathbf{X} \mathbf{W}_V^i)$ a $\mathbf{W}_{out} \in \mathbb{R}^{hd_v \times d}$.

Dôležitou vlastnosťou self-attention mechanizmu je permutačná equivariancia. Inak povedané zámena poradia prvkov vstupnej sekvencie nezmení výstupné reprezentácie, ale prejaví sa len korešpondujúcou zmenou poradia prvkov výstupnej sekvencie. V prípade spracovania prirodzeného jazyka alebo iných dát so sekvenčným charakterom je potrebné do modelu nejakým spôsobom zapracovať aj znalosť relatívnej či absolútnej pozície prvkov sekvencie. V [84] sa tento problém vyriešil pričítaním vektora pozičných kódov \mathbf{pe}_i ku každému vektoru zo vstupnej sekvencie \mathbf{x}_i . Jednotlivé komponenty v rámci pozičného vektora korešpondujúceho s i -tým prvkom vstupnej sekvencie majú hodnoty dané ako

$$[\mathbf{pe}_i]_{dim=1}^d = \begin{cases} \sin\left(\frac{i}{\tau \frac{2dim}{d}}\right) & \text{ak } dim = 2k \\ \cos\left(\frac{i}{\tau \frac{2dim}{d}}\right) & \text{inak} \end{cases} \quad (3.7)$$

kde $k \in \mathbb{N}$, d je dimenzia vektorov vstupnej sekvencie, $\tau = 10000$ a $i \in \{1, 2, \dots, t\}$.

3.1.2 Model U-TAE

Neurónová sieť U-TAE [29] je model z ktorého pri návrhu metódy vychádzame. Okrem toho v dobe začiatku písania tejto práce dosahoval SOTA výsledky na úlohe sémantickej segmentácie plodín z časových rád snímok Sentinel-2 na benchmarkovom datase PASTIS. Na obr. 3.1 môžeme vidieť jeho schématické znázornenie z ktorého je zrejmé, že ide o modifikáciu siete U-Net.

Podobne ako v U-Net, prvé rameno zabezpečuje priestorový downsampling a extrakciu priestorových príznakov s tým rozdielom, že v rámci U-TAE je aplikované pre všetky snímky zo vstupnej časovej rady súčasne a nezávisle.

V najnižšom priestorovom rozlíšení feature máp je použitý nezávislý model, zvaný L-TAE [67], slúžiaci ako temporálny encoder, ktorý je zodpovedný za efektívnu extrakciu temporálnych príznakov pomocou modifikovaného attention mechanizmu. Najdôležitejšou charakteristikou L-TAE je, že namiesto výpočtu *query* vektorov lineárnou transformáciou zo vstupnej sekvencie, model používa jeden tzv. *master query* vektor, ktorý je jeho trénovateľným parametrom. Výpočet attention len s jedným *query* vektorom umožňuje mapovať sekvenciu vektorov na práve jeden vektor a zakódovať tak relevantnú temporálnu informáciu obsiahnutú v časovej rade do jednej reprezentácie. Podobne ani výpočet *values* vektorov neprebíha lineárnou transformáciou vstupnej časovej rady, ale tieto vektory sú priamo rovné vstupnej časovej rade t.j. $V=X$.

Temporálny encoder L-TAE využíva pozičné kódovanie vo forme 3.7 až na to, že parameter $\tau = 1000$ a jednotlivé pozície i reprezentujú počet dní uplynutých od pevne stanoveného referenčného dátumu. Podotknime ešte, že L-TAE je aplikovaný pre každý pixel nezávisle. Výstupom L-TAE je pixelová maska rozmerov $\mathbf{a}_4^{pixel} = [h \times T \times \frac{H}{8} \times \frac{W}{8}]$ obsahujúcich attention skóre, kde dimenzia h korešponduje s jednotlivými hlavami vo výpočte attention. V každej úrovni spracovania $l \in \{1, 2, 3, 4\}$ sa attention masky získané temporálnym encoderom upsamplujú do požadovaného priestorového rozlíšenia pomocou bilinéarnej interpolácie a použijú sa na temporálne váženú sumu časovej rady feature máp. Dôležitým detailom výpočtu temporálnej váženej sumy je, že je vykonaná po blokoch. V každej úrovni spracovania l sa teda sekvencia feature máp rozdelí na h blokov rozmerov $T \times \frac{C_l}{h} \times H_l \times W_l$ kde $l \in \{1, 2, 3, 4\}$, $H_4 = \frac{H}{8}$, $W_4 = \frac{W}{8}$ a vážená suma sekvencie feature máp sa vypočíta v každej z h podskupín samostatne na základe attention skóre príslušnej hlavy pixelovej masky \mathbf{a}_l^{pixel} . Nakoniec samotné decoder rameno už svojou štruktúrou a funkciou kopíruje model U-Net.

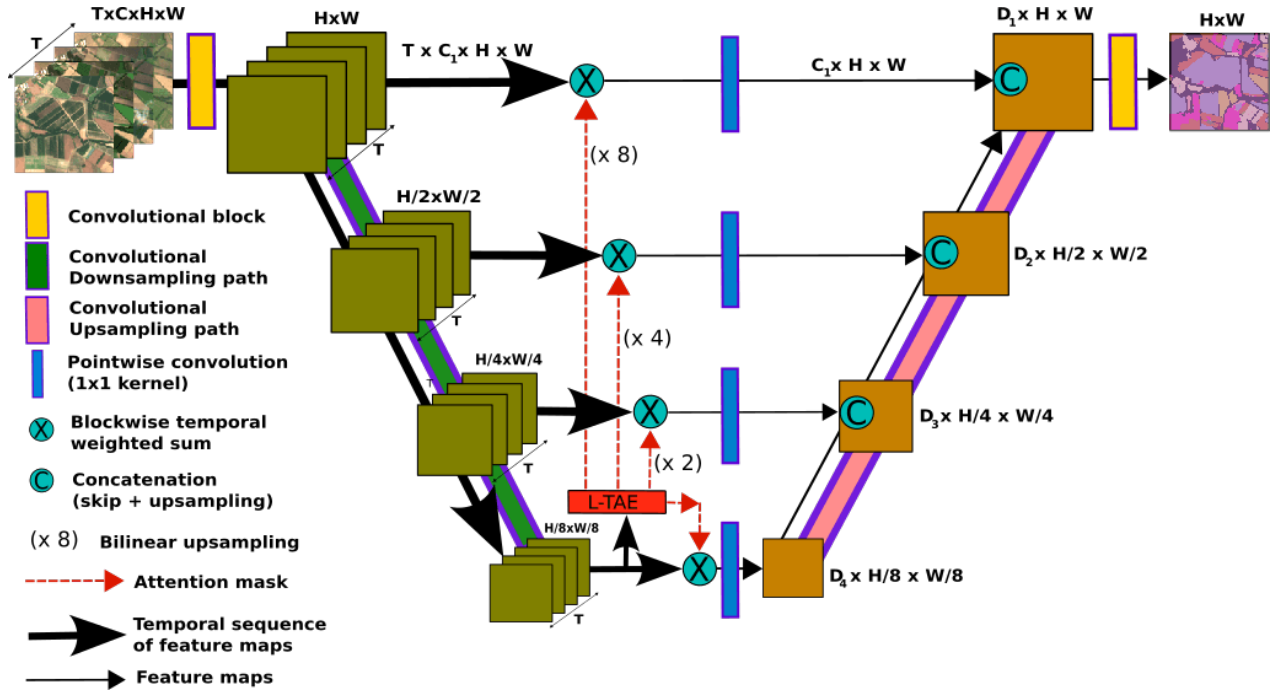
3.2 Návrh metódy

Nami navrhované vylepšenia pozostávajú jednak zo zmien v pozičnom kódovaní, ale taktiež zo zmien v poradí spracovania časovej rady snímok, čo vyúsťuje do návrhu dvoch nových architektur, menovite TimeU-Net a W-TAE.

3.2.1 Pozičné kódovanie

Ako už bolo spomenuté, v našej metóde využívame temporálny encoder L-TAE, ktorý vychádza a v istých ohľadoch modifikuje klasický attention modul. Z nášho pohľadu navrhujeme jeho úpravu len v oblasti pozičného kódovania. Konkrétne ide o dve alternatívy ku kódovaniu použitom v základnom nastavení L-TAE (viď. vzťah 3.7):

pe^{F+L} (fixované + učené). V tomto prípade navrhujeme obohatiť fixované (sinusoidálne) pozičné kódovanie dní uplynutých od referenčného dátumu pomocou učného kódovania absolútnych pozícií dní v roku vo forme day of year (DOY), nakoľko predpokladáme, že tieto umožnia



Obr. 3.1: Znáznornenie architektúry U-TAE [29], ktorá je základom pre návrh metódy. Model je úzko spätý s dizajnom modelu U-Net, ktorý vylepšuje o možnosť spracovania časových rád snímok na základe temporálneho encodera L-TAE.

modelu sa naučiť, lepšie podchytiť, špecifické charakteristiky fenologického cyklu plodín. Učenie pozíčných embeddingov je zabezpečené lineárnou transformáciou one-hot vektoru DOY pozícií, ktoré možno formálne zhrnúť ako

$$pe^L = DW^L \in \mathbb{R}^{t \times d_{out}} \quad (3.8)$$

kde $D \in \mathbb{R}^{t \times 365}$ je matica reprezentujúca one-hot vektory DOY pozícií časovej rady s t pozorovaniami a $W^L \in \mathbb{R}^{365 \times d_{out}}$ je učená matica lineárnej transformácie. Celkovo je predpis navrhovaného pozíčného kódovania daný

$$pe^{F+L} = pe^F + pe^L \in \mathbb{R}^{t \times d_{out}} \quad (3.9)$$

kde pe^F je fixované kódovanie počtu uplynutých dní od referenčného dátumu pomocou predpisu 3.7 pričom $\tau = 1000$. Použitie fixného kódovania pe^F sa môže zdať nadbytočné keď že model má možnosť sa z dát naučiť potrebné reprezentácie pe^L . Ich prítomnosť je ale dôležitá, pretože učené pozíčné kódovanie pe^L samostatne, nedokáže generalizovať na pozície, ktoré sa nevyskytovali v tréningových dátach počas učenia [53]. Práve pri spracovaní nerovnomerne rozložených pozorovaní v rámci časových rád Sentinel-2 snímok, môže nastať prípad kedy niektoré pozície nemusia byť v čase tréningovania k dispozícii.

pe^E (vylepšené). Ďalší návrh opäť vychádza z fixovaného kódovania podľa 3.7 kde nahradzujeme pozície v zmysle dní uplynutých od referenčného dátumu za formát DOY. Modifikácia spočíva v následnej aplikácii (učenej) lineárnej transformácie, ktorá opäť zvýši flexibilitu fixného kódovania a umožní modelu doučiť sa prípadné vylepšenia pozíčných embeddingov. Tento prístup

označujeme ako vylepšený (enhanced) a formálne ho možno zapísať takto

$$pe^E = pe_{DOY}^F W^E \in \mathbb{R}^{t \times d_{out}} \quad (3.10)$$

kde $pe_{DOY}^F \in \mathbb{R}^{t \times d_{out}}$ je opäť fixované kódovanie, ale tenzor DOY pozícií získaných podľa predpisu 3.7 pričom $\tau = 1000$ a $W^E \in \mathbb{R}^{d_{out} \times d_{out}}$ je opäť učená transformačná matica.

3.2.2 TimeU-Net

V rámci architektúry U-TAE je aplikovaný temporálny encoder (L-TAE) v najnižšom priestorovom rozlíšení a následne sa pixelové masky reprezentujúce attention skóre bilineárne up-samplujú do vyšších priestorových rozlíšení aby boli použité na výpočet temporálne váženej sumy vstupných sekvencií. Takýto krok autori odôvodňujú predovšetkým znížením pamäťových a výpočetných nárokov nakoľko temporálny encoder musí spracovať každý pixel nezávisle.

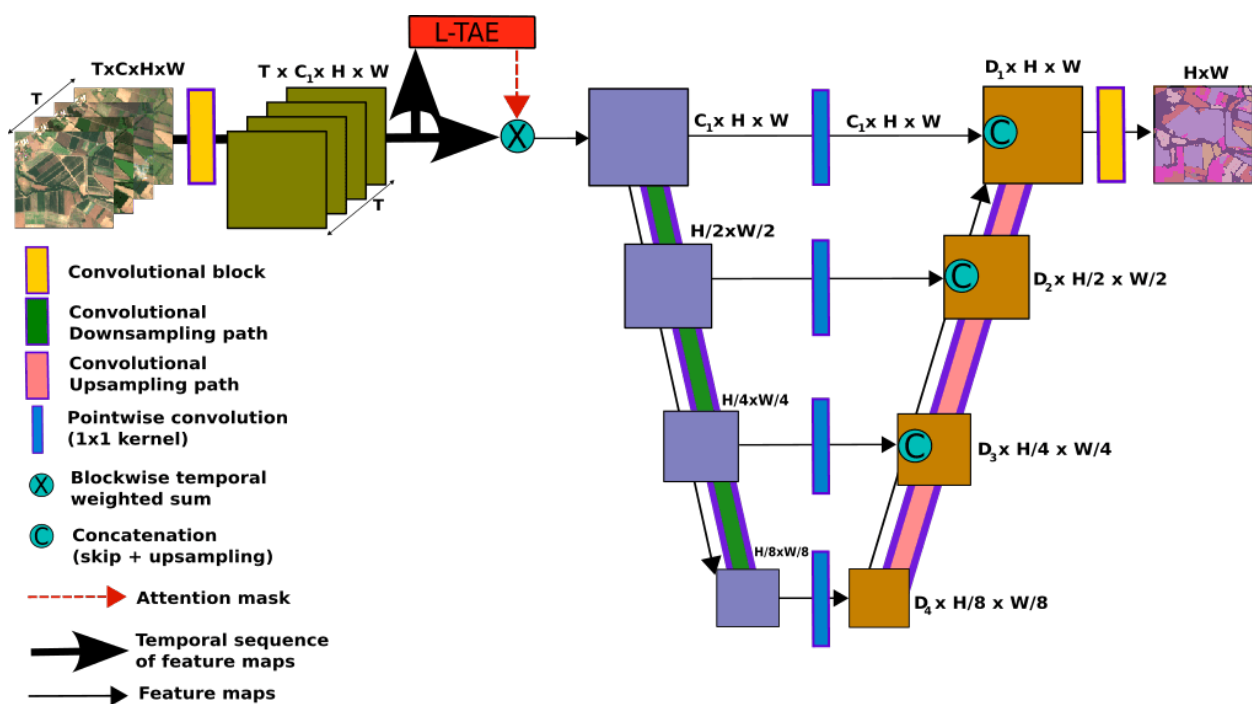
Na druhú stranu empirické výsledky [19, 78] poukazujú na to, že pre zvyšovanie kvality sémantickej segmentácie plodín z časových rád Sentinel-2 snímok je vhodné, primárne extrahovať temporálne reprezentácie dát a až dodatočne riešiť modelovanie priestorovej domény snímok. Z tohto dôvodu navrhujeme priamočiaru modifikáciu architektúry U-TAE v tom zmysle, že sa L-TAE použije v najvyššej úrovni t.j. na časovú radu feature máp v pôvodnom rozlíšení snímok. Takto modifikovanú architektúru nazývame TimeU-Net, nakoľko po agregácii temporálnych príznakov pokračuje spracovanie modelom U-Net. Model je znázornený na obr. 3.2. Podotknime, že takáto zmena poradia spracovania temporálnej a priestorovej domény snímok zvýšila flexibilitu modelu oproti U-TAE, nakoľko TimeU-Net už nie je úzko spätá s konkrétnym segmentačným modelom (v našom prípade U-Net), ale umožňuje jeho nahradenie za iný, výkonnejší segmentačný model podobne ako v [19].

3.2.3 W-TAE

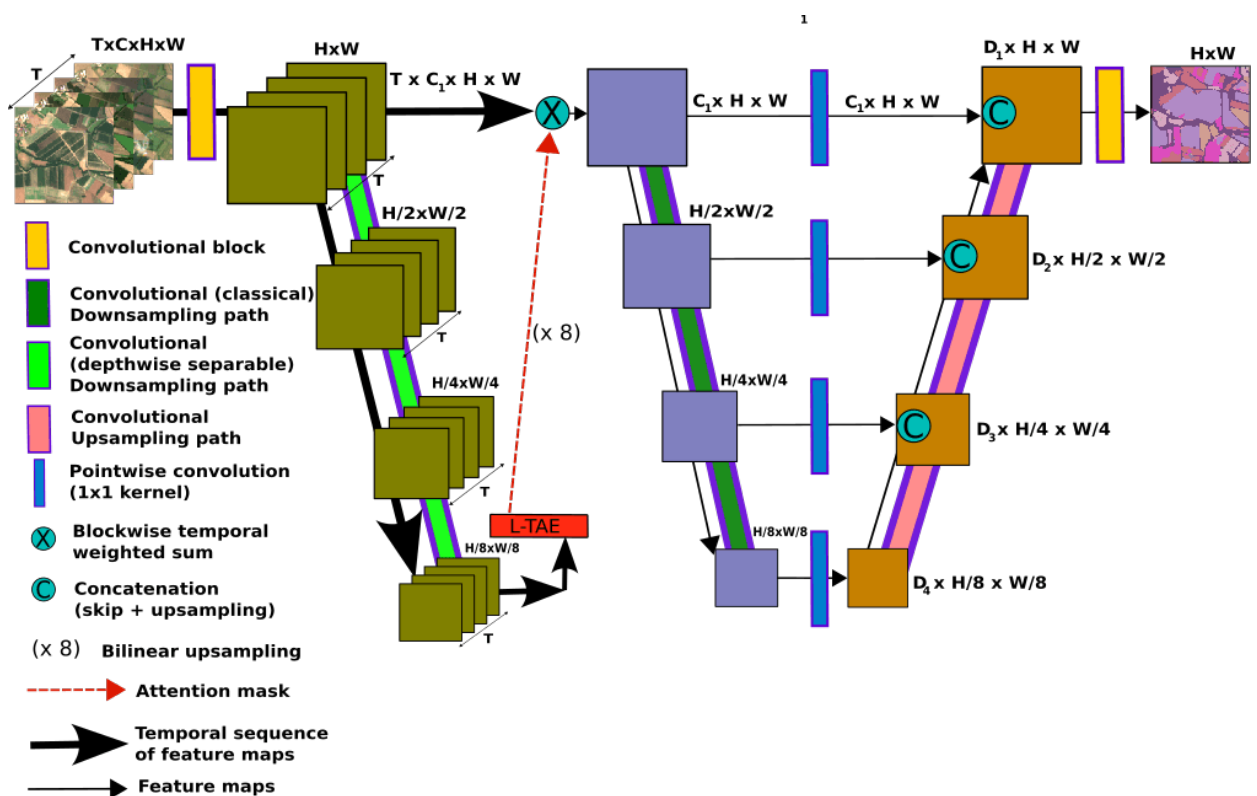
Nadväzujúcim návrhom je architektúra W-TAE, ktorú možno vidieť na obrázku 3.3. Opäť ide o modifikáciu U-TAE v niekoľkých ohľadoch. Za prvé zamieňame výpočetne náročnú 2D konvolúciu za depthwise-separable konvolúciu [24] v (temporálne) zdieľanom priestorovom encoderi. Následne po spracovaní sekvencie priestorovým encoderom sa aplikuje temporálny encoder L-TAE. Pixelové attention skóre masky sú následne up-samplované priamo do pôvodného priestorového rozlíšenia snímok kde sa použijú na výpočet temporálne váženej sumy vstupných snímok. Spracovanie ďalej pokračuje segmentačným modelom U-Net. Celkovo sa dá pozerieť na W-TAE ako na zjednodušenú verziu TimeU-Net, popísanú v predchádzajúcej sekcii, nakoľko je jej architektúra opäť explicitne rozdelená na dve časti kde jedna zabezpečuje extrakciu temporálnych príznakov pomocou akejsi priestorovej aproximácie výpočtu attention, modulom L-TAE a druhá slúži na spracovanie priestorovej dimenzie snímok a generovanie výslednej segmentácie. Za cenu mierného navýšenia počtu parametrov je takýto model dokonca výpočetne menej náročný ako architektúra U-TAE (viď. tab. 5.1) a podobne ako v prípade TimeU-Net naviac získal flexibilitu v zmysle možnosti nahradenia segmentačného modelu U-Net za výkonnejší.

3.3 Alternatívne prístupy

Pre účely porovnania s našim navrhnutým prístupom v tejto sekcii uvedieme niekoľko alternatívnych architektúr neurónových sietí umožňujúcich vykonať sémantickú segmentáciu z časových rád satelitných snímok.



Obr. 3.2: Schématické znázornenie architektúry TimeU-Net. Temporálny encoding je v prípade tohto modelu aplikovaný priamo v najvyššom rozlíšení a teda nedochádza k priestorovej aproximácii výpočtu attention. Po extrakcii temporálnych príznakov model pokračuje klasickým, U-Netu podobným modelom, ktorý zabezpečí sémantickú segmentáciu.



Obr. 3.3: Schématické znázornenie architektúry W-TAE. Model je akýmsi hybridom medzi U-TAE a TimeU-Net. Podobne ako v U-TAE je temporálny encoding vykonaný v najnižšom priestorovom rozlíšení a teda získané attention skóre je len priestorovou aproximáciou skutočného attention skóre pre každý pixel. Po vykonaní temporálneho encodingu a bilineárneho upsamplingu,, architektúra pokračuje nezávislým modelom U-Net, čo prináša možnú flexibilitu podobne ako v prípade TimeU-Net.

U-Net Ako baseline model sme použili zjednodušenú verziu siete U-Net [59] kde spracovanie časovej dimenzie snímok riešime priamočiarým, ale naivným spôsobom a síce zlúčením spektrálnej a temporálnej dimenzie do jednej novej dimenzie. Rozmery snímok sa teda zmenia z $[T \times C \times H \times W]$ na $[TC \times H \times W]$. Sekvenčný charakter časovej rady snímok v tomto prípade nie je využitý.

3D U-Net Nahradením 2D konvolučných vrstiev, kedy sa filter posúva v dvoch priestorových smeroch snímky teda $[H \times W]$ a extrahuje tak výlučne priestorovo-spektrálne príznaky, 3D konvolučnými vrstvami kedy sa filter posúva jednak v spomínaných priestorových rozmeroch, ale tentoraz aj v časovom rozmere snímok t.j. $[T \times H \times W]$, čo má za následok simultánne spracovanie priestorových, temporálnych a spektrálnych vzorov časovej rady snímok, získame 3D U-Net [60].

ConvLSTM Ide o hybridnú architektúru [72], ktorá vznikne nahradením lineárnych vrstiev 2D konvolučnými vrstvami v rámci rekurentnej siete LSTM [32]. Model je tak schopný pracovať v temporálnej a priestorovej dimenzii snímok súčasne.

U-ConvLSTM Priamočiarou alternatívou k modelu U-TAE je tzv. U-ConvLSTM [60] kedy sa namiesto temporálneho encodera L-TAE použije hybridná sieť ConvLSTM. Vychádzajúc z [29] nahradzujeme batch normalizáciu za group normalizáciu.

TSViT Model TSViT [78] je jedinou architektúrou založenou výlučne na attention module, ktorý je použitý na spracovanie časovej ako aj priestorovej dimenzie časovej rady Sentinel-2 snímok. Práve kvôli tejto vlastnosti je architektúra značne výpočetne a pamäťovo náročná.

Exchanger+U-Net/Mask2Former Exchanger [19] rozdelí učeň proces učenia temporálnych reprezentácií do troch špecifických krokov založených na využití tzv. group propagation bloku (GP Block) [94]. Samotný model Exchanger funguje len ako temporálny encoder podobne ako napr. L-TAE a preto vyžaduje dodatočné použitie ľubovoľného segmentačného modelu ako napr. U-Net alebo Mask2Former. Podľa našich vedomostí ide v dobe písania tejto práce o SOTA metódu (prekonávajúcu tak predošlé SOTA metódy TSViT a U-TAE) čo sa týka sémantickej segmentácie plodín z časových rád snímok Sentinel-2 na datasete PASTIS [29].

3.4 Regularizácia

Akúkoľvek modifikáciu procesu učenia, ktorej cieľom je zlepšenie generalizačných schopností modelu môžeme označiť za regularizačnú techniku. V našom prístupe sme uvažovali niekoľko takýchto techník, menovite label smoothing, temporálny dropout a augmentáciu datasetu, ktoré čitateľovi v nasledujúcej sekcii priblížime.

3.4.1 Label smoothing

Metóda zvaná label smoothing [77] má za cieľ zabrániť pretrénovaniu (overfitting) modelu vnesením šumu do ground truth odozvy, ktorá je v prípade klasifikácie vzorky (x_i, y_i) do K tried s anotáciami z množiny $y_i \in \{1, 2, \dots, K\}$ typicky reprezentovaná one-hot vektorom pravdepodobností vo forme

$$p(y|x_i) = \begin{cases} 1 & \text{ak } y = y_i \\ 0 & \text{inak} \end{cases} \quad (3.11)$$

Aplikáciou label smoothingu nahradíme pôvodnú reprezentáciu ground truth 3.11 novou distribúciou s predpisom

$$p^{LS}(y|x_i) = \begin{cases} 1 - \alpha + \frac{\alpha}{K} & \text{ak } y = y_i \\ \frac{\alpha}{K} & \text{inak} \end{cases} \quad (3.12)$$

kde K je počet uvažovaných tried a $\alpha \in (0, 1]$ je paramter metódy. Takýmto vyhladením ground truth distribúcie sa docieľi aby model negeneroval predikcie s prehnane vysokou istotou, čo môže mať pozitívny vplyv na generalizačné schopnosti modelu.

3.4.2 Temporálny dropout

Inšpirovaní klasickou regularizačnou technikou zvanou Dropout autori v [68] navrhli náhodne vylúčiť každé pozorovanie s pravdepodobnosťou p zo vstupnej časovej rady snímok. Danú techniku nazvali temporálny dropout pričom jeho cieľom je zabezpečiť aby sa model nadmerne nespoľiehal na prítomnosť konkrétnych pozorovaní vo vstupnej sekvencii, ale bol schopný extrahovať potrebné temporálne vzorce zo zvyšných snímok a tak dosiahnúť lepšie generalizačné vlastnosti. Spomínaný postup je aplikovaný len v tréningovej fáze. V testovacej resp. validačnej fáze má model prístup ku všetkým dostupným pozorovaniam. Nakoľko podľa výsledkov autorov takýto prístup zlepšil výslednú schopnosť segmentácie siete tak sme sa rozhodli danú regularizáciu vyskúšať aj v našich experimentoch.

3.4.3 Augmentácia datasetu

Augmentácia datasetu slúži na zväčšenie počtu tréningových vzoriek bez explicitného získavania nových anotácií. Umelé navýšenie veľkosti tréningového datasetu je docieľené duplikáciou už existujúcich dát pomocou vhodne definovaných transformácií. Takýmto spôsobom sa zabezpečí aby sa model učil reprezentácie dát, ktoré sú invariantné voči použitým transformáciám.

V našom prípade sme ako transformácie volili náhodnú rotáciu a orezanie obrázka na veľkosť 64x64 pixelov z pôvodných 128x128 pixelov. Taktiež sme použili vertikálne a horizontálne prevrátenie. Spomínané transformácie boli aplikované na každý člen časovej rady v rámci tréningového datasetu.

3.5 Post-processing

Ako možný post-processing navrhujeme zjednotiť top-1 predikciu (tzv. hard predikcia) modelu prípadne iné ukazatele ako napr. istotu predikcie vyjadrenú predpovedanou distribúciou (tzv. soft predikcia) na úrovni parcel a to buď na základe majoritného hlasovania alebo určením maximálnej pravdepodobnosti z agregovaných pixelových soft predikcií. Samotné vektorové dáta určujúce polohu a hranice DPB by boli získané buď z databázy LPIS alebo na základe predikcie triedy reprezentujúcej hranicu, naším modelom. Formálne je zjednotenie hard predikcie voči DPB na základe majoritného hlasovania dané ako

$$pred_{max,hard}^{DPB} \in \arg \max_{i \in \{1,2,\dots,K\}} \{c_i\} \quad (3.13)$$



Obr. 3.4: Navrhované zjednotenie top-1 predikcie na základe vektorových dát LPIS. Prvý obrázok zľava reprezentuje vektorovú masku parceli resp. DPB, ďalším je samotná top-1 predpoveď modelu pričom ich spojením získame zjednotenú predikciu na úrovni parceli.

kde c_i je počet pixelov v rámci analyzovaného DPB kedy model predpovedal triedu i . Grafické znázornenie zjednotenia top-1 predikcie je na obr. 3.4. Podobne môžeme zapísať zjednotenie hard predikcii zistením indexu triedy, pre ktorú má vektor agregovaných pixelových soft-predikcii maximálnu hodnotu

$$pred_{max,soft}^{DPB} \in \arg \max_{i \in \{1,2,\dots,K\}} \left\{ \sum_{j \in DPB} p_{ji} \right\} \quad (3.14)$$

kde tenzor p_{ji} reprezentuje predikovanú pravdepodobnosť triedy i pre j -ty pixel v rámci DPB.

Podotkneme, že uvažovaný post-processing nemá za cieľ zlepšiť kvalitu predikcie resp. zvýšiť (pixelové) metriky modelu aj keď takáto možnosť nie je vylúčená ako môžeme vidieť v [49]. Našou motiváciou je praktickosť získaného výstupu pričom pri návrhu post-processingu vychádzame z predpokladu, že väčšina rozhodnutí ohľadom poľnohospodárskej politiky súvisiacej s určovaním typu plodiny sa vzťahuje k jednotlivým parcelám resp. DPB a nie k jednému pixelu ako priestorovej jednotke záujmu. Použitím sémantickej segmentácie si tak jednak v čase inferencie udržujeme flexibilitu v podobe získania soft predikcii pre jednotlivé pixely, čo umožňuje analyzovať istotu predikcie a možné nezrovnalosti v pestovaní plodín na častiach parceli voči deklaráciám farmárov, ale zároveň vykonaním popísaného zjednotenia voči DPB umožníme potenciálnemu používateľovi urobiť rozhodnutie vzhľadom k parcele ako celku.

3.6 Problematika hraničných pixelov

Hranice poľí sú oblasťou kde často dochádza k výraznej degradácii presnosti segmentácie. Takáto degradácia úzko súvisí s miešaním spektrálnej odozvy susediacich typov povrchov v snímkach s nízkym alebo stredným priestorovým rozlíšením akými sú aj snímky Sentinel-2 (viď. obr. 3.5). Okrem toho, pri rasterizácii vektorových dát určujúcich hranice sa typicky postupuje tak, že hraničnému pixelu sa priradí trieda korešpondujúca s geometriou, ktorá má maximálny prienik s daným pixelom. Z tohto dôvodu je principiálne nemožné jednoznačne priradiť jednu konkrétnu triedu hraničnému pixelu.

Podľa nášho názoru je v prípade hraničných pixelov omnoho prirodzenejšie pracovať s distribúciou pravdepodobností resp. soft predikcii prípadne explicitne priradiť takýmto pixelom novú, špecifickú anotáciu. V nasledujúcich odstavcoch priblížime dva prístupy, ktoré sme využili v našej práci.



Obr. 3.5: Miešanie spektrálnej odozvy pixelov na hraniciach sémanticky rovnorodých oblastí. Snímka pochádza z 25.7.2019.

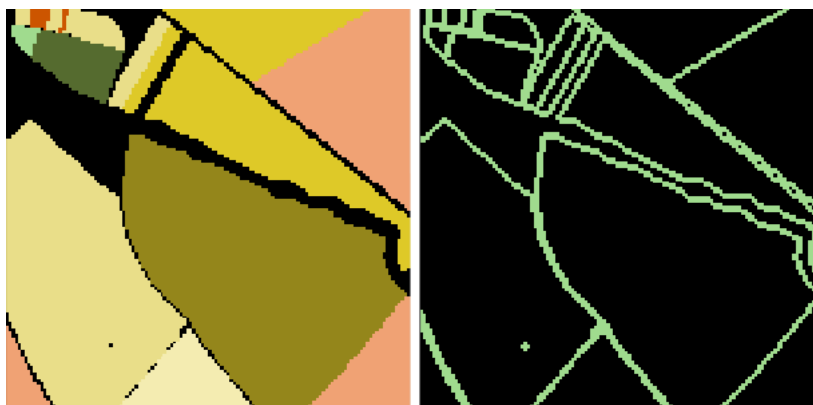
Smooth loss Ide o modifikáciu label smoothingu [77] v rámci CrossEntropy loss funkcie, špeciicky navrhnutú pre dodatočné vyhladenie ground truth pravdepodobností hraničných pixelov. Upravený vektor pravdepodobností je daný ako

$$p_{Boundary}^{LS}(y|x_i) = \begin{cases} \frac{1-\alpha}{|\chi|} + \frac{\alpha}{K} & \text{ak } y \in \chi \\ \frac{\alpha}{K} & \text{inak} \end{cases} \quad (3.15)$$

kde $\alpha \in (0, 1]$ je parametrom metódy, χ je množina tried hraničných pixelov voči i -temu analyzovanému pixelu a $|\chi|$ je počet prvkov χ , pričom $|\chi| \geq 1$ nakoľko trieda y_i patrí za každých okolností do χ . Inak povedané, pixely, ktoré nie sú hraničnými, sú vyhladené podľa vzťahu 3.12, ktorý je špeciálnym prípadom 3.15 pri $|\chi|=1$. Podotknime, že tento prístup nemá za cieľ vylepšiť výsledné metriky klasifikácie, ale vniesť do predikcie hraničných pixelov neurčitost', ktorá vyplýva z podstaty problému a potenciálne tak umožniť modelu vierohodnejšie reprezentovať distribúciu pravdepodobností tried na hraniciach sémanticky rovnorodých oblastí.

Hraničné pixely ako nová trieda Druhou možnosťou je uvažovať hranicu ako dodatočnú triedu a snažiť sa ju predpovedať. V našom prípade sme sa rozhodli umožniť architektúre neurónovej siete aby súčasne predpovedala segmentáciu plodín ale aj binárnu segmentáciu hraníc sémanticky rovnorodých oblastí. V tomto prípade má teda model dva zdroje supervízie. Opäť nie je cieľom vylepšiť výsledné metriky, ale tentoraz doplniť plodinovú segmentáciu o explicitnú predpoveď hraničných pixelov.

Nakoľko sme v priebehu tvorby datasetu explicitne neukladali informácie ohľadom hraníc polí tak túto informáciu musíme extrahovať z ground truth masiek. Extrakcia prebieha pomocou vytvorenia one-hot zakódovanej reprezentácie ground truth masky a následnej aplikácie morfologickej dilatácie so štruktúrym elementom tvaru kríža 3×3 . Dilatácia umožní rozšírenie oblastí jednotlivých tried pričom následnou detekciou prekrývajúcich sa oblastí zistíme hranice polí. Nevýhodou spomínaného prístupu je šírka extrahovaných hraníc, ktorá je minimálne 2 pixely. Na druhú stranu, extrakcia hraníc za behu má nižšie pamäťové nároky keďže nemusíme informáciu o hranici explicitne ukladať. Obrázok 3.6 zobrazuje príklad ground truth a korešpondujúce extrahované hranice polí.



Obr. 3.6: Extrakcia hraníc polí pomocou dilatácie. Naľavo vidíme ground truth masku a napravo extrahované hranice.

3.7 Hodnotenie výsledkov

V tejto práci hodnotíme výsledky segmentácie pomocou štandardných ukazateľov akými sú konfúzna matica, celková accuracy ale aj pre úlohu sémantickej segmentácie vhodnejšou metrikou v podobe intersection over union (IoU) resp. jej priemernou hodnotou mIoU. V nasledujúcich odstavcoch priblížime detaily výpočtov jednotlivých uvažovaných metrik.

Confusion matrix reprezentuje kompaktný záznam výsledkov klasifikácie na základe kumulácie informácie o počte úspešných a neúspešných výsledkov predikcie modelu. Takýto záznam sa hodí pre jednoduchšiu vizuálnu inšpekciu kvality modelu. Samotná konfúzna matica teda nie je metrika a v prípade binárnej klasifikácie nadobúda tvar

	1	0	Skutočné
1	TP	FN	
0	FP	TN	
Predpovedané			

Číslo 1 indikuje pozitívnu triedu a 0 negatívnu. TP znamená počet true positive výsledkov, podobne TN je počet true negative výsledkov, FP počet false positive a FN false negative výsledkov. Rozšírenie pre klasifikáciu s viacerými triedami je priamočiare a preto ho nebudeme uvádzať.

Accuracy udáva podiel počtu správnych predpovedí klasifikátora ku počtu všetkých jeho predpovedí. V prípade binárnej klasifikácie je definovaná ako

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (3.16)$$

Výhodou tejto metriky je jej interpretovateľnosť no na druhú stranu prináša neuspokojivý odhad kvality klasifikátora pri nevyváženom datasete nakoľko nadhodnocuje jeho kvalitu. Túto metriku je teda vhodné použiť ak sú distribúcie tried porovnateľné.

Recall (sensitivity) udáva podiel správne detekovaných relevantných vzoriek voči počtu všetkých relevantných vzoriek. Jeho nízka hodnota indikuje veľký počet falošne negatívnych predpovedí. Vzťah pre výpočet je nasledujúci:

$$recall = \frac{TP}{TP + FN} \quad (3.17)$$

Precision je metrika, ktorá udáva mieru dôveryhodnosti predpovede daným klasifikátorom. Nízka hodnota tejto metriky poukazuje na prevahu falošne pozitívnych predpovedí a je definovaná ako:

$$precision = \frac{TP}{TP + FP} \quad (3.18)$$

F1 score (dice coefficient) sa definuje ako harmonický priemer metrick recall a precision:

$$F_1 = 2 \frac{precision * recall}{precision + recall} = 2 \frac{TP}{2TP + FN + FP} \quad (3.19)$$

F1 score teda agreguje informáciu, ktorú získame z metrick precision a recall a poslúži tak ako lepšia miera kvality klasifikátora pri nevyváženom datasete.

Intersection over Union (jaccard index) tiež známy ako koeficient podobnosti meria podobnosť medzi ground truth a predpoveďou klasifikátora. Táto metrika je definovaná nasledovne:

$$IoU = \frac{TP}{TP + FN + FP} \quad (3.20)$$

TP môžeme považovať za prienik množín resp. segmentačných oblastí ground truth a predpovede klasifikátora a následne súčet TP, FN a FP reprezentuje zjednotenie ground truth a samotnej predpovede klasifikátora. Použitie IoU býva časté v úlohách ako detekcia objektov a sémantická segmentácia pretože dáva lepší vzhľad o kvalite klasifikátora oproti metrikám ako Accuracy a podobne. Ďalej v práci budeme používať priemernú hodnotu IoU značenú mIoU a počítanú ako

$$mIoU = \frac{1}{|C|} \sum_{i \in C} IoU_i \quad (3.21)$$

kde C značí množinu tried.

3.8 Implementačné detaily

Implementácia jednotlivých modelov neurónových sietí bola založená na frameworku PyTorch [55]. Celkovo je logika pre tréovanie a testovanie modelov ovládaná pomocou skriptu `train.py`¹, ktorý poskytuje CLI rozhranie. Na optimalizáciu sme využívali vo všetkých prípadoch algoritmus Adam [38] v základnom nastavení s learning rate-om 0.001. Ako loss funkcia bola použitá Cross Entropia. Veľkosť mini-batch-u, ktorá sa ukázala byť optimálna bola 4 v prípade datasetu PAS-TIS a 8 v prípade datasetu S2TSCZCrops. Jednotlivé modely boli tréované po dobu 70 epoch bez early-stopping stratégie pričom model, ktorý dosahoval najvyššiu hodnotu metriky mIoU vzhľadom k validačnému datasetu bol vybraný ako najlepší a použitý v testovacej fáze. Na akceleráciu učenia neurónových sietí bola použitá grafická karta NVIDIA Tesla A100. Celkovo boli výpočetné kapacity zastrešené Ústavom teórie informácie a automatizácie AV ČR.

¹<https://github.com/Many98/Crop2Seg/blob/main/train.py>

Poznámka k reprodukovateľnosti experimentov Na zaistenie reprodukovateľnosti výsledkov naprieč behmi sme postupovali štandardným spôsobom a síce explicitným nastavením náhodného stavu v rámci generátora náhodných čísel. Nanešťastie prostredie PyTorch negarantuje úplnú reprodukovateľnosť výsledkov [5] keďže niektoré operácie ako napr. bilineárny up-sampling, ktorý nami navrhnuté architektúry využívajú, majú pre GPU akcelerátory len stochastické implementácie výpočtu derivácie. Z tohto dôvodu je možné, že pri replikácii nami uvedených experimentov, nebude možné dospieť k numericky zhodným výsledkom na požadovanej úrovni presnosti.

Kapitola 4

Výsledky

V nasledujúcich sekciách postupne zhrnieme kvantitatívne výsledky experimentov súvisiace s navrhnutými architekturnými zmenami, riešením problematiky hraničných pixelov ale aj post-processingu predstavenými v kapitole 3. Následne uvedieme kvalitatívne výsledky v podobe ukážok segmentácie plodín najlepšimi získanými modelmi.

Vzhľadom k stochastickému charakteru niektorých operácií v prostredí PyTorch (viď. odstavec 3.8) by bolo vhodné vykonať každý experiment niekoľkokrát a následne uvádzať priemerné skóre s korešpondujúcim rozptylom. Takýto prístup sme si, ale z časového hľadiska nemohli dovoliť a preto uvádzame výsledky len po jednom behu experimentov.

4.1 Segmentácia pomocou navrhnutých modelov

V rámci experimentovania s navrhnutými modelmi sme rozčlenili experimenty do niekoľkých skupín. V prvej sme testovali vplyv zmien v pozičnom kódovaní pričom experiment s označením baseline korešponduje s architektúrou v základnom nastavení, bez použitia regularizácii a zmien v pozičnom kódovaní t.j. používame fixované kódovanie dní uplynutých od referenčného dátumu. Následne sme v ďalších skupinách testovali vplyv temporálneho dropoutu a label smoothingu na získané modely, prípadne vplyv augmentácie datasetu.

4.1.1 U-TAE

Aj keď model U-TAE v základnom nastavení nie je naším návrhom tak ho podrobujeme experimentom a vylepšeniam ohľadom pozičného kódovania a použitia regularizačných techník. Tabuľka 4.1 popisuje výsledky vykonaných experimentov s týmto modelom. Na datasete PASTIS bola dosiahnutá najvyššia hodnota mIoU 63.66 % s použitím pozičného kódovania daného predpisom 3.9 a temporálneho dropoutu s pravdepodobnosťou $p=0.1$ zatiaľ čo v prípade S2TSCZCrops datasetu bolo dosiahnuté najvyššie mIoU v prípade použitia vylepšeného pozičného kódovania 3.10 a label smoothingu s parametrom $\alpha=0.1$.

4.1.2 W-TAE

K experimentom s pozičným kódovaním, použitím temporálneho dropoutu a label smoothingu sme pridali v prípade modelu W-TAE aj dva experimenty kedy sme pristúpili k augmentácii datasetov. Z prehl'adu výsledkov v tab. 4.2 sa zdá, že takéto použitie augmentácie neprineslo žiadané

Experiment	S2TSCZCrops			PASTIS		
	mIoU	F1 score	Acc	mIoU	F1 score	Acc
baseline	72.14	81.92	91.03	63.17	75.73	83.11
pe^{F+L}	72.54	82.06	91.28	62.55	75.25	83.32
pe^E	72.18	82.11	91.05	61.58	74.30	82.90
baseline + temporal dropout (0.1)	70.67	79.21	91.15	62.48	75.30	83.27
pe^{F+L} + temporal dropout (0.1)	72.15	81.55	91.20	63.66	76.17	83.53
pe^E + temporal dropout (0.1)	72.33	81.94	91.20	63.18	75.80	83.44
baseline + label smoothing (0.1)	72.01	81.77	91.09	63.25	75.97	83.49
pe^{F+L} + label smoothing (0.1)	72.29	81.80	91.20	62.33	75.08	83.40
pe^E + label smoothing (0.1)	72.78	82.39	91.24	61.39	74.22	82.70

Tabuľka 4.1: Výsledky experimentov modelu U-TAE na datasetoch PASTIS a S2TSCZCrops (uvedené v %). Výsledky na datasete PASTIS sú vzhľadom k 5 testovacím foldom.

zlepšenie. Model dosiahol najlepšie segmentačné výsledky pri použití vylepšeného pozičného kódovania 3.10 pričom tentoraz bol nápomocný temporálny dropout na datasete S2TSCZCrops a label smoothing na PASTISe. Všimnime si taktiež, že na datasete PASTIS nedosahuje W-TAE v základnom nastavení (baseline) lepší segmentačný výsledok ako model U-TAE.

4.1.3 TimeU-Net

Model TimeU-Net dosahuje na oboch datasetoch najvyššie hodnoty metriky mIoU spomedzi nami navrhnutých resp. modifikovaných modelov ako možno vidieť v tab. 4.3. V prípade oboch datasetov išlo o baseline model pričom na PASTISe boli navyše použité obe regularizačné techniky a síce label smoothing a temporálny dropout. Ďalej uvádzame aj priebeh učenia najlepšieho modelu (baseline) na datasete S2TSCZCrops (viď. obr. 4.1). Najlepšia epocha bola stanovená ako tá kedy neuronová sieť dosiahla najvyššiu hodnotu mIoU na validačnej podmnožine datasetu. V našom prípade išlo o poslednú epochu 70. Z grafov vidíme saturáciu validačnej loss funkcie ako aj validačnej accuracy metriky no na druhú stranu sa zdá, že validačná hodnota mIoU by pri dodatočom tréovaní ešte pokračovala v mierne rastúcom trende.

Na obrázku 4.2 sú ďalej zobrazené výsledné hodnoty metrick pre každú sledovanú plodinu resp. triedu na základe najlepšieho modelu TimeU-Net (vzhľadom k mIoU) na datasete S2TSCZCrops. Triedy Flax & Hemp, Forage crops a Permanent fruit dosiahli celkovo najnižšie hodnoty mIoU. Naopak triedy Background, Summer cereals, Winter cereals, Rapeseed, Maize, Sugar beet a Hopyards sú klasifikované s relatívne vysokou presnosťou. Celkový prehľad správne a chybné klasifikovaných pixelov pre spomínaný model (baseline) je obsiahnutý v konfúznej matici na obr. 4.3. Z neho vidíme napr., že trieda Forage crops bola až v 19% prípadov zamenená za triedu Grassland. Taktiež trieda Flax & Hemp bola vo vysokej miere modelom zamieňaná za triedu Grassland alebo Other crops. Nakoniec spomeňme ešte triedu Permanent fruit, ktorá bola často zamieňaná s triedou Background, ale čiastočne aj s triedou Grassland.

Experiment	S2TSCZCrops			PASTIS		
	mIoU	F1 score	Acc	mIoU	F1 score	Acc
baseline	72.53	82.26	91.04	61.66	74.23	83.20
pe^{F+L}	72.96	82.37	91.43	61.49	74.62	83.25
pe^E	72.58	82.27	91.16	63.32	75.95	83.63
baseline + temporal dropout (0.1)	72.21	81.81	91.24	63.48	76.05	83.55
pe^{F+L} + temporal dropout (0.1)	72.95	82.62	91.10	61.80	74.57	83.31
pe^E + temporal dropout (0.1)	73.23	82.84	91.37	60.08	73.12	83.14
pe^E + temporal dropout (0.05)	-	-	-	63.42	76.04	83.72
pe^E + temporal dropout (0.3)	-	-	-	63.38	76.02	83.70
baseline + label smoothing (0.1)	72.23	82.04	91.13	62.44	75.19	83.66
pe^{F+L} + label smoothing (0.1)	72.99	82.49	91.39	62.97	75.86	83.73
pe^E + label smoothing (0.1)	72.77	82.27	91.29	63.54	76.05	83.77
pe^E + label smoothing (0.1) + temporal dropout (0.1)	72.31	81.81	91.29	61.77	74.56	83.38
pe^{F+L} + augmentation	70.28	80.22	90.25	-	-	-
pe^E + augmentation	-	-	-	60.90	73.94	82.14

Tabuľka 4.2: Výsledky experimentov modelu W-TAE na datasetoch PASTIS a S2TSCZCrops (uvedené v %). Výsledky na datasete PASTIS sú vzhľadom k 5 testovacím foldom.

Experiment	S2TSCZCrops			PASTIS		
	mIoU	F1 score	Acc	mIoU	F1 score	Acc
baseline	73.48	83.10	91.19	62.40	75.16	83.19
pe^{F+L} *	72.05	81.79	91.02	62.30	75.09	83.29
pe^E	71.86	81.61	90.94	62.23	75.01	83.09
baseline + temporal dropout (0.1)	73.32	82.87	91.47	61.50	74.75	83.00
baseline + label smoothing (0.1)	72.02	81.79	91.00	62.72	75.39	83.21
baseline + label smoothing (0.1) + temporal dropout (0.1)	73.26	82.94	91.17	64.84	77.31	84.07

Tabuľka 4.3: Výsledky experimentov modelu TimeU-Net na datasetoch PASTIS a S2TSCZCrops (uvedené v %). Výsledky na datasete PASTIS sú vzhľadom k 5 testovacím foldom. * znamená, že experiment bol vykonaný s batch size 4 kvôli vysokým pamäťovým nárokom.

Nakoniec môžeme na obr. 4.4 pozorovať porovnanie modelov U-TAE a TimeU-Net v zmysle metriky mIoU dosiahnutej pre jednotlivé triedy plodín na datasete PASTIS. V prípade tried plodín Winter barley, Spring barley, Beet a Potatoes dosahuje model U-TAE o niečo lepšie výsledky. Ostatné triedy sú lepšie detekované modelom TimeU-Net. Najväčší rozdiel pozorujeme v prípade plodín Winter triticales, Mixed cereals a Sorghum.

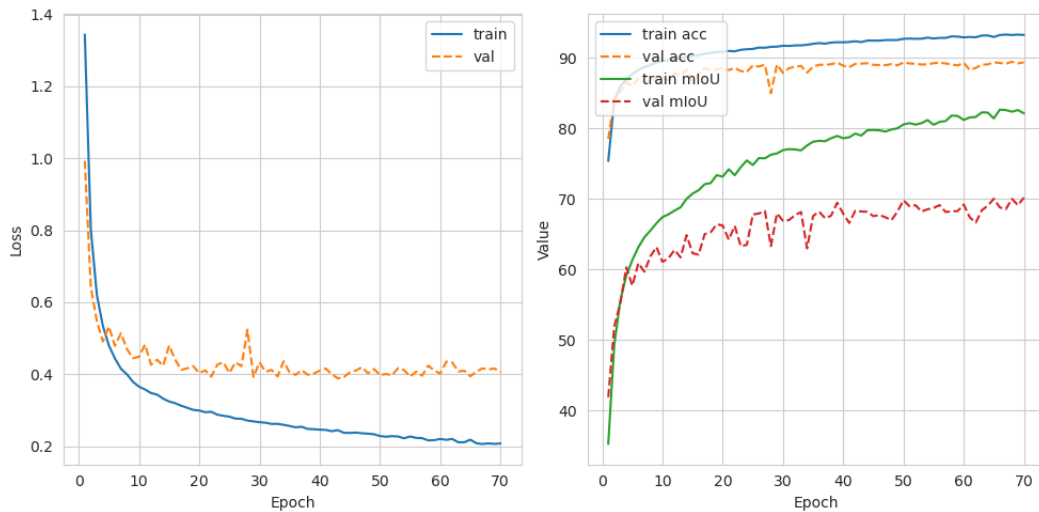
4.1.4 Kvalitatívne výsledky

Obrázok 4.5 pre zaujímavosť porovnáva pixelové attention skóre masky generované modelmi W-TAE a TimeU-Net. Ide o štyri z možných šestnástich masiek prislúchajúcim jednotlivým hlavám temporálneho encodera L-TAE. Na prvý pohľad je vidieť značný rozdiel v miere detailu medzi prezentovanými attention skóre maskami pre W-TAE a TimeU-Net. Tento rozdiel spočíva v spôsobe spracovania nakoľko v prípade W-TAE najprv dochádza k priestorovému downsamplingu až po ktorom dochádza ku generovaniu attention skóre masiek temporálnym encoderom, ktoré sú následne bilineárne upsamplované do pôvodného rozlíšenia snímky. Model TimeU-Net naopak generuje attention skóre masky priamo v najvyššom priestorovom rozlíšení snímok a teda nedochádza k žiadnej priestorovej aproximácii.

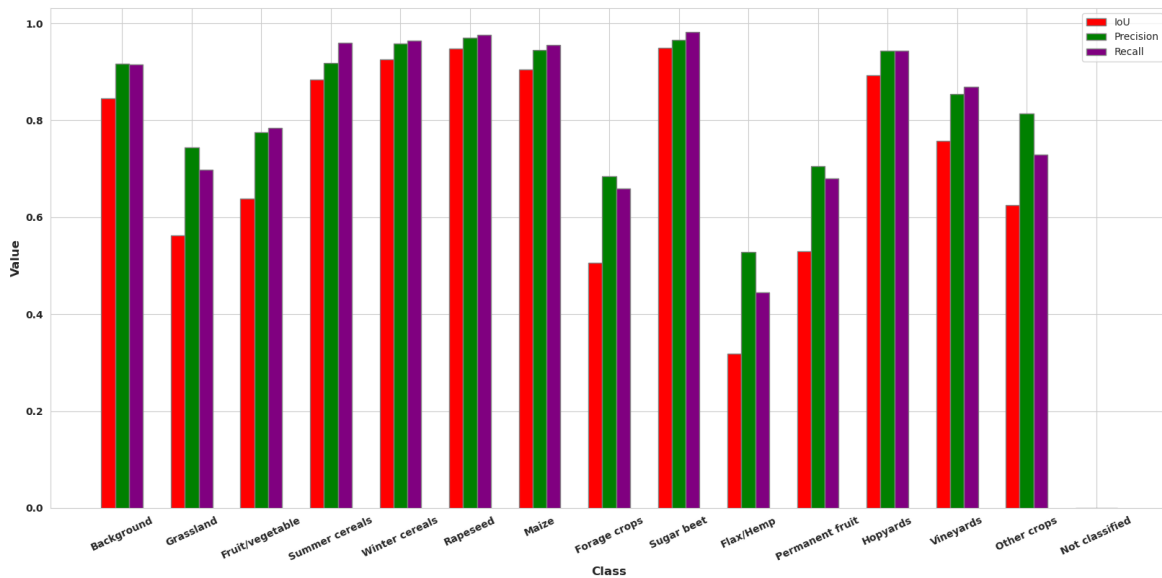
Následne môžeme pozorovať na obr. 4.6 porovnanie predikcii modelov U-TAE, W-TAE a TimeU-Net. Top-1 predikcie modelov sa zdajú byť medzi sebou podobné len s občasnými odlišnosťami. Nakoniec na obrázku 4.7 je vyobrazená istota predikcie korešpondujúca s predpoveďami na obr. 4.6. Zjavný pokles jasu v grafe istoty predikcie pre U-TAE v porovnaní s W-TAE a TimeU-Net je spôsobená použitím label smoothingu s $p = 0.1$ ¹.

Kvalitatívny prehľad predpovedí modelu U-TAE v základnom nastavení (baseline) v porovnaní s TimeU-Net na datasete PASTIS je zobrazený na obr. 4.8. Celkovo sa opäť zdá, že segmentačné masky oboch modelov sú zhruba porovnateľné. Model TimeU-Net, ale dokázal lepšie lokalizovať a segmentovať triedy Sorghum (prvý riadok z vrchu), Spring barley (druhý riadok od hora), Orchard (tretí riadok od spodu), prípadne Mixed cereals (druhý riadok od spodu). Na druhú

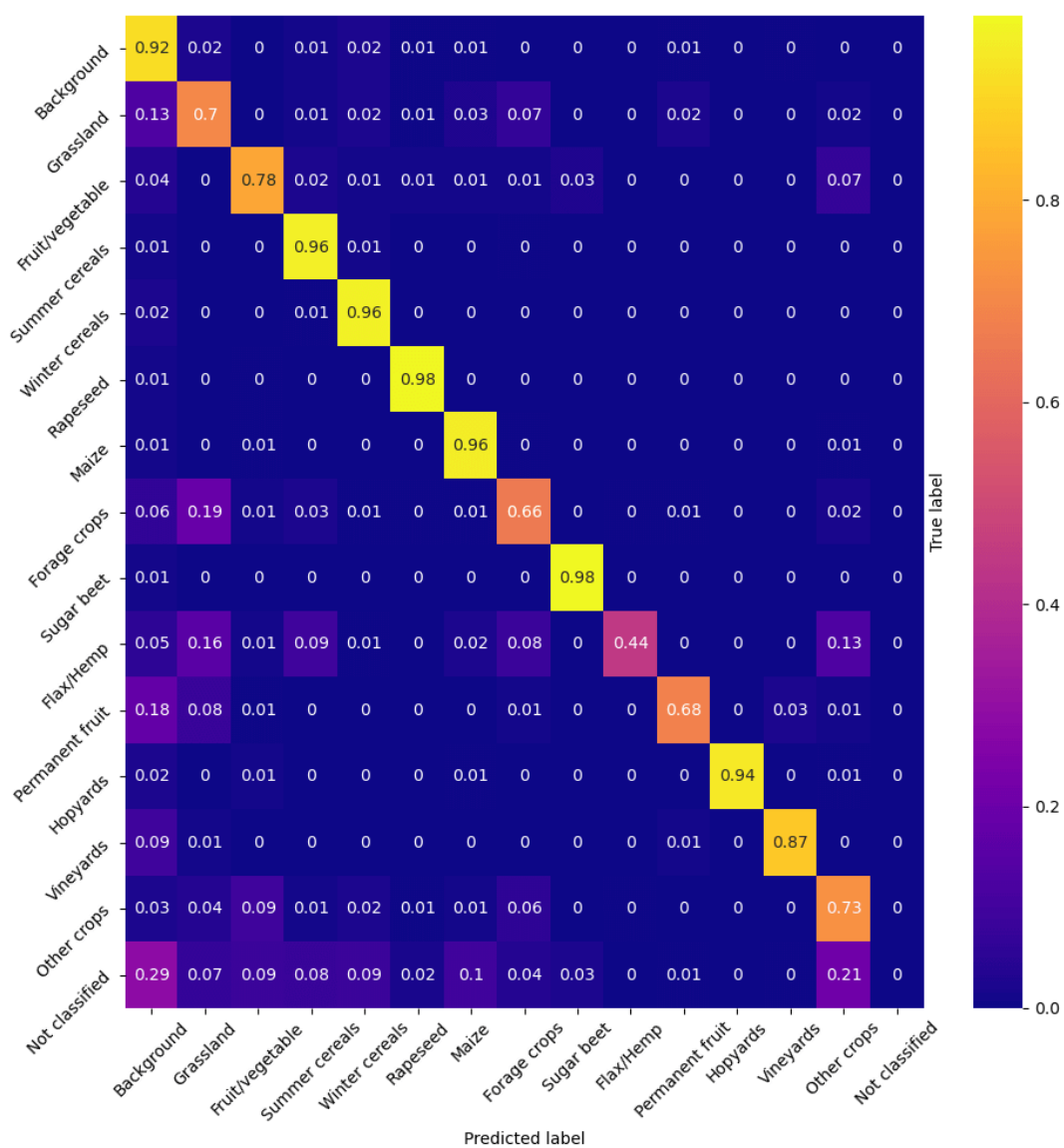
¹Najlepší model U-TAE na datasete S2TSCZCrops používal vylepšené pozičné kódovanie a label smoothing (viď. posledný riadok tab. 4.1)



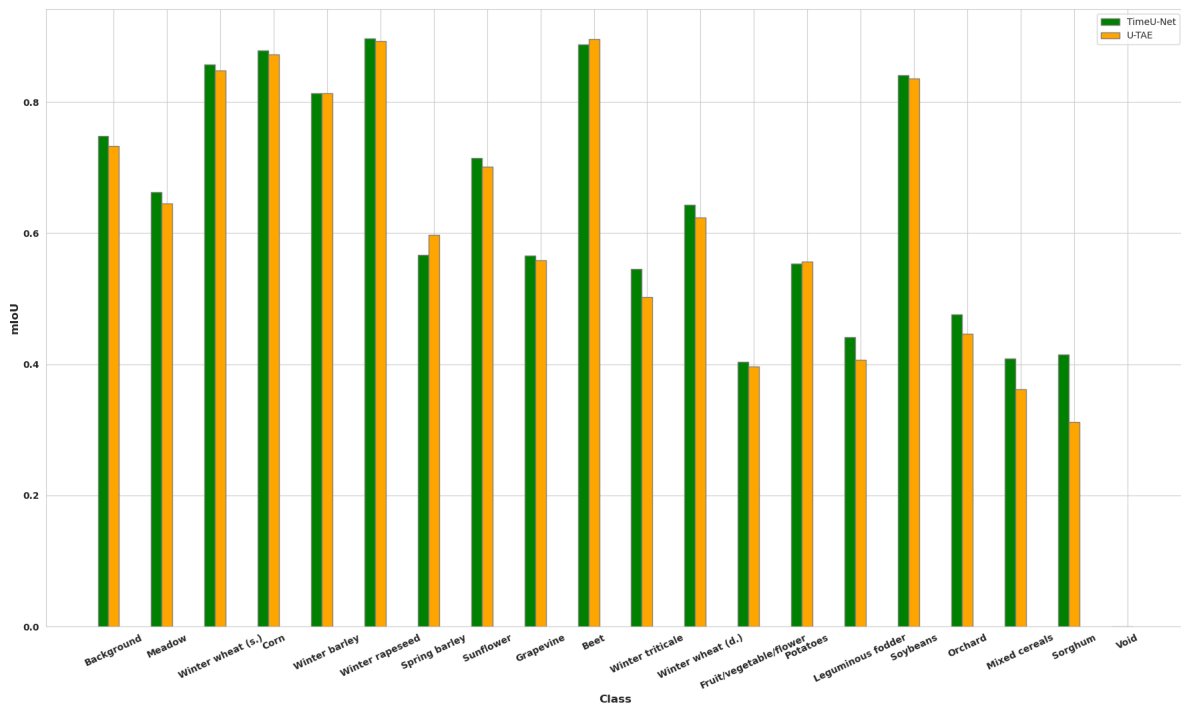
Obr. 4.1: Priebeh učenia modelu TimeU-Net (baseline) na datasete S2TSCZCrops. Naľavo sú znázornené loss funkcie (raw hodnota) a napravo metriky accuracy a mIoU (hodnoty v %).



Obr. 4.2: Hodnoty testovacích metrick modelu TimeU-Net (baseline) pre jednotlivé triedy plodín na datasete S2TSCZCrops.



Obr. 4.3: Konfúzna matica modelu TimeU-Net (baseline) na datsete S2TSCZCrops. Podotkname, že ide o normalizovanú verziu a preto diagonála reprezentuje metriku recall t.j. podiel správne detekovaných pixelov danej triedy na testovacej podmnožine datsetu S2TSCZCrops.



Obr. 4.4: Porovnanie modelov TimeU-Net (najlepší model) a U-TAE (pôvodný návrh [29]) vzhľadom k hodnotám mIoU pre jednotlivé triedy plodín na datase PASTIS.

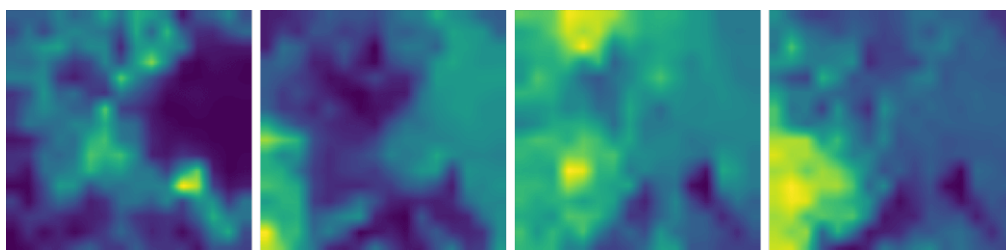
stranu model U-TAE je schopný lepšie detekovať triedu Winter barley (druhý riadok od hora), Meadow (tretí riadok od hora) a Leguminous fodder (prvý riadok od dola).

4.2 Hraničné pixely

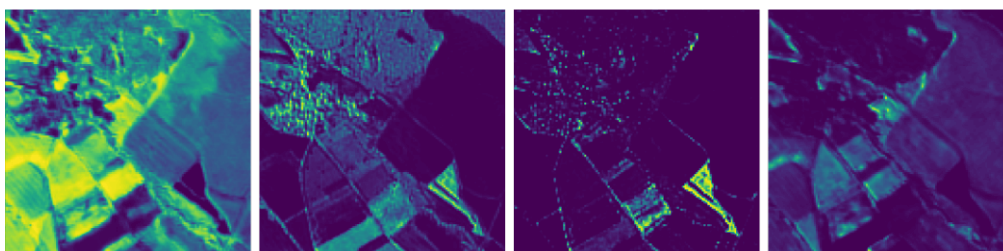
Jedným z problémov ktorým sme sa okrajovo zaoberali bolo zlepšenie reprezentácie hraničných pixelov. Pre tento účel sme navrhli dva prístupy a síce tzv. Smooth loss, čo je modifikácia label smoothingu v rámci Cross Entropy funkcie a predikciu hraničných pixelov ako dodatočnej triedy. Oba prístupy sme testovali za použitia modelu W-TAE s pozičným kódovaním pe^{F+L} (predpis 3.9) na datase S2TSCZCrops. Výsledky sú zhrnuté v tab. 4.4. Všimnime si, že pri dodatočnej predikcii hranice dosahuje model W-TAE zatiaľ najvyššiu hodnotu mIoU (73.31%) spomedzi ostatných experimentov s týmto modelom (porovnaj s tab. 4.2).

Ďalej môžeme na obrázku 4.9 pozorovať distribúcie počtu pixelových predpovedí s konkrétnou istotou predpovede. V prípade experimentu bez použitia Smooth Loss funkcie (podobrázok a)) si všimnime, že s väčšinou (top-1) predpovedí si je model veľmi istý. Až v 80% prípadov pixelov z vnútorných oblastí generuje model predikciu s istotou viac ako 93%. Na druhú stranu pri hraničných pixeloch si je istý s pravdepodobnosťou viac ako 93% len v zhruba 30% prípadoch. Na podobrázku b) vidíme zastúpenie po použití Smooth Loss funkcie. Predovšetkým v prípade hraničných pixelov došlo k výraznej zmene distribúcie, ktorá dosahuje svoj modus v intervale istoty predikcie 46%-53%.

Samotná predpoveď modelov je vyobrazená v 4.10 a 4.11. Konkrétne na obrázku 4.10 sú zobrazené okrem top-1 predpovede (podobrázok a)) modelu aj predpovede hranice (podobrázok b)). Celkovo je predpoveď hranice veľmi podobná ground truth maske hranice s príležitostnými



(a) W-TAE

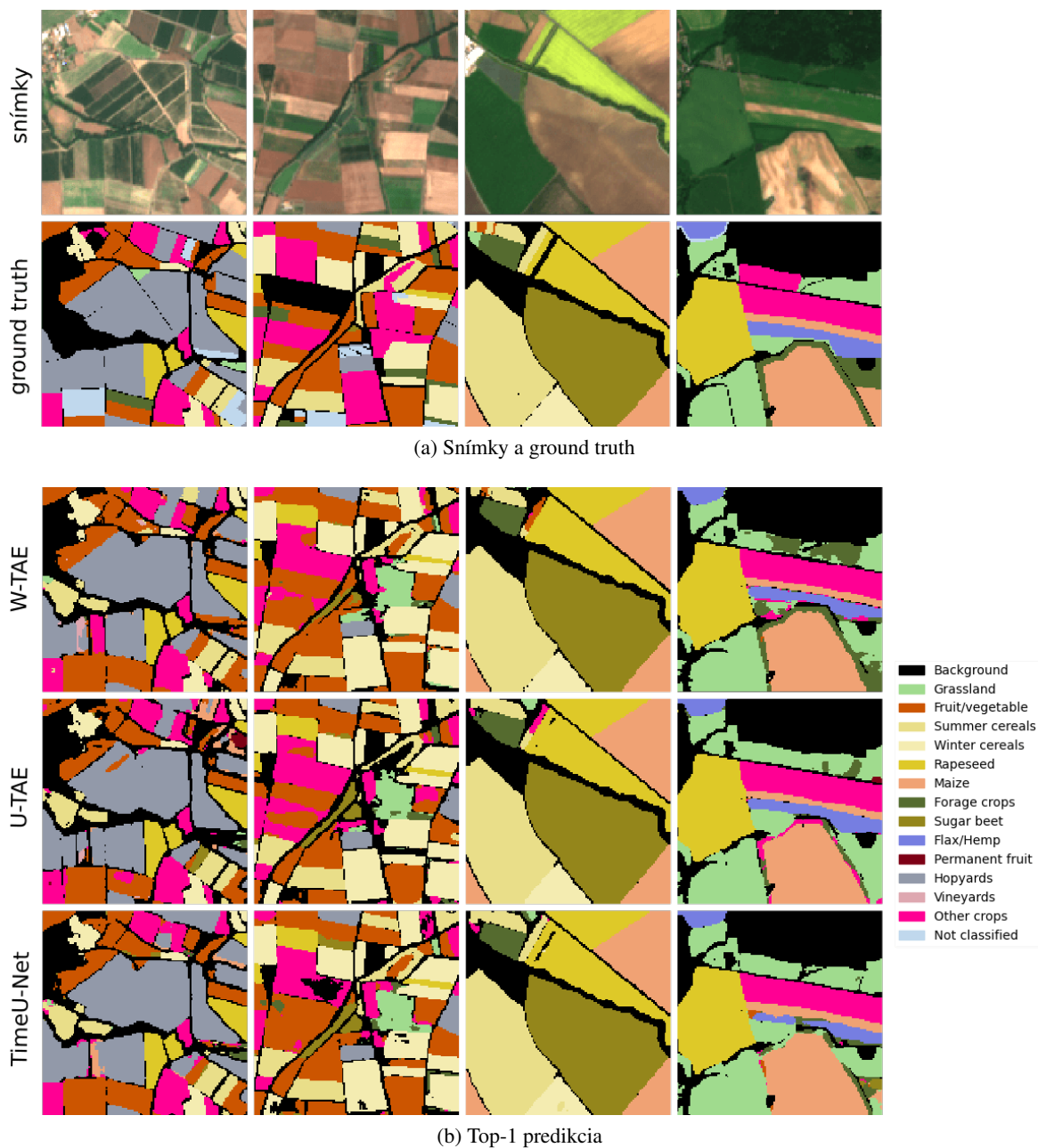


(b) TimeU-Net

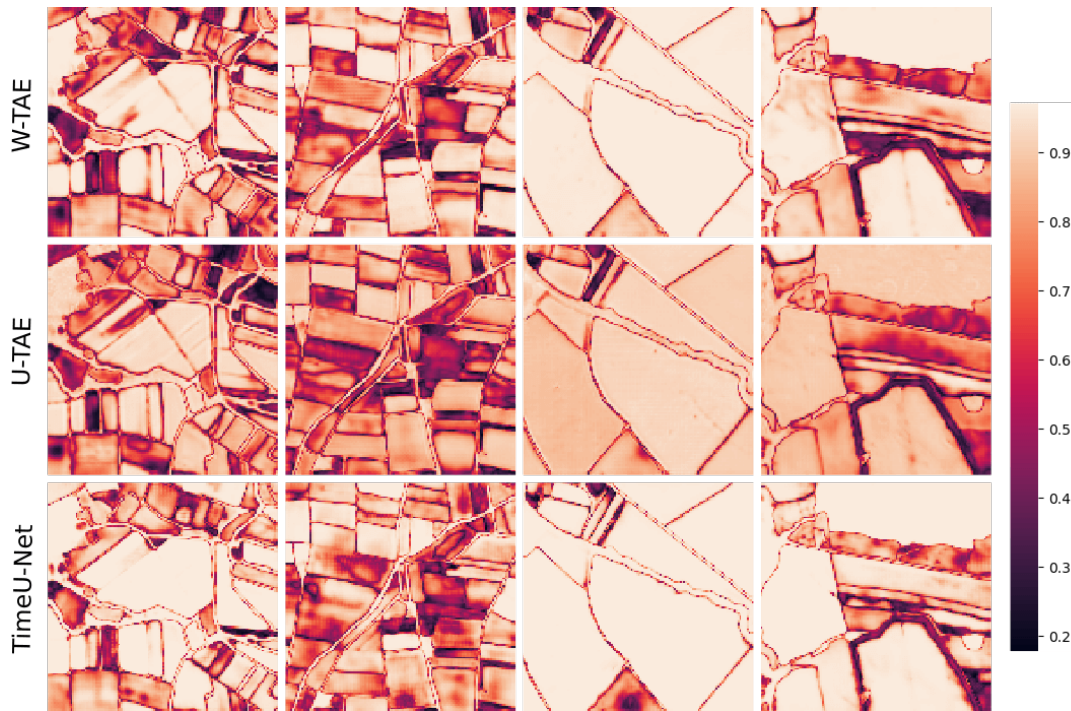


(c) Ground truth a Sentinel-2 snímka

Obr. 4.5: Porovnanie pixelových attention skóre masiek pre 4 zo 16 možných hláv: a) model W-TAE b) model TimeU-Net c) ground truth maska a korešpondujúca snímka v pravých farbách z 18.6.2019. Upozorníme, že attention masky v podobrázku a) a b) spolu nesúvisia a ide nám len o zvýraznenie kvalitatívnej rozdielnosti v miere detailu, ktorú jednotlivé attention skóre masky ponúkajú. Attention skóre generované modelom U-TAE je kvalitatívne zhodné so skóre generovaným sieťou W-TAE a preto ho explicitne neuvádzame.



Obr. 4.6: Prehľad predpovedí modelov na datasete S2TSCZCrops. Predpovede sú získané za použitia najlepších model vzhľadom k metrike mIoU.



Obr. 4.7: Pokračovanie obr. 4.6. Prehľad istoty predikcie t.j. pravdepodobnosť top-1 predpovede modelu.

nedokonalosťami (viď. červené obdĺžniky). Na druhú stranu v niektorých prípadoch model detekoval dodatočné rozčlenenie parciel, ktoré nie je zaznačené v ground truth (viď. modrý obdĺžnik). Následne na obrázku 4.11 možno pozorovať predikciu po použití Smooth Loss prístupu. Všimnime si na tomto obrázku hlavne istotu predikcie, ktorá je pre hraničné pixely výrazne tmavšia v porovnaní s predikciou hranice v podobrázku a) v 4.10.

4.3 Homogenizácia predikcie

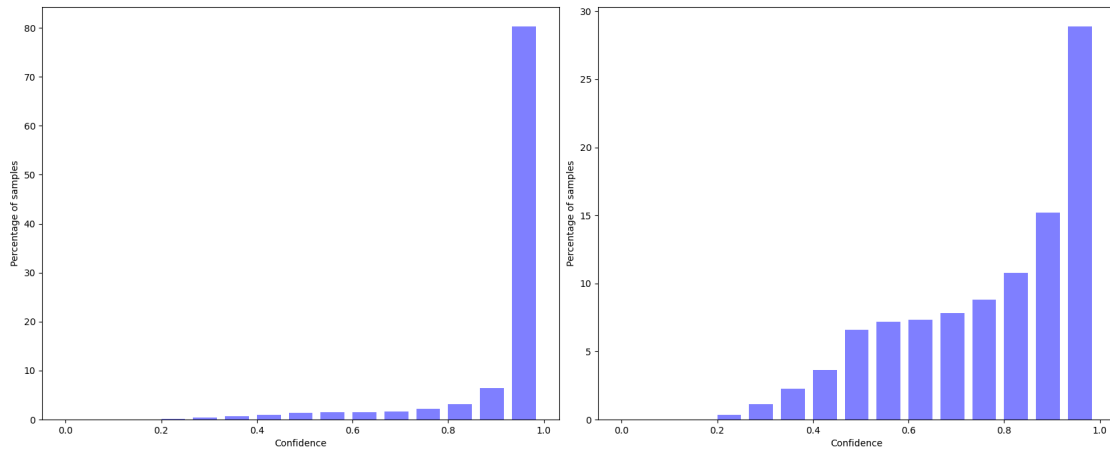
Navrhnutý post-processing v zmysle zjednotenia (homogenizácie) pixelovej predikcie na úrovni parciel sme taktiež podrobili experimentu. Zjednotenie bolo aplikované na výstupy natrénovaného modelu W-TAE s pozičným kódovaním pe^{F+L} a label smoothingom, ktorý súčasne predpovedal aj

Experiment	S2TSCZCrops		
	mIoU	F1 score	Acc
model + label smoothing (0.1) + predikcia hranice	73.31	82.78	91.59
model + smooth loss	71.54	81.51	90.04

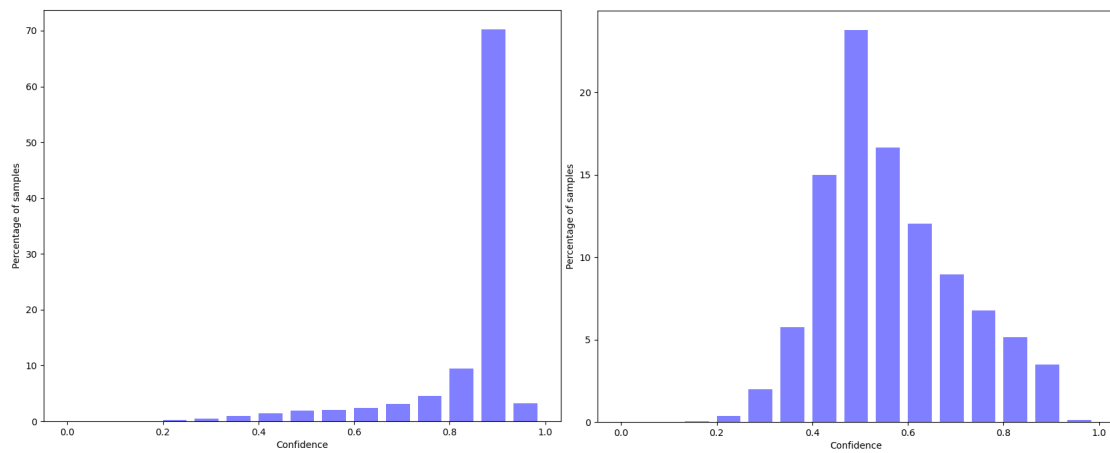
Tabuľka 4.4: Výsledky experimentov ohľadom spracovania hraničných pixelov (uvedené v %). Experimenty boli vykonané za použitia modelu W-TAE s pozičným kódovaním pe^{F+L} na datase S2TSCZCrops. Podotkneme, že prezentovaná hodnota mIoU v prípade predikcie hranice vyjadruje segmentačnú kvalitu sémantickej segmentácie plodín. Segmentácia hranice bola dosiahnutá s mIoU 81%.



Obr. 4.8: Kvalitatívne výsledky predikcie modelov U-TAE (pôvodný návrh - baseline [29]) a TimeU-Net (najlepší model) na datasete PASTIS.

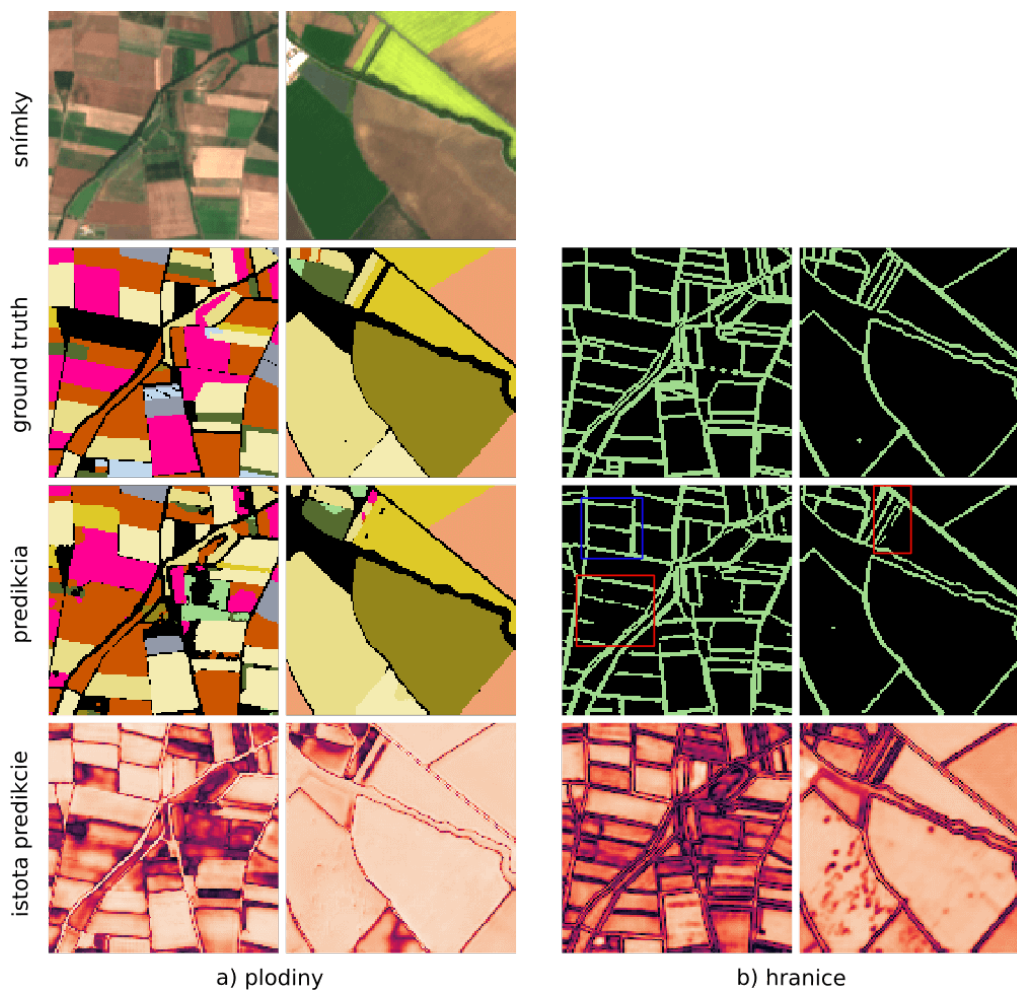


(a) Bez Smooth Loss prístupom

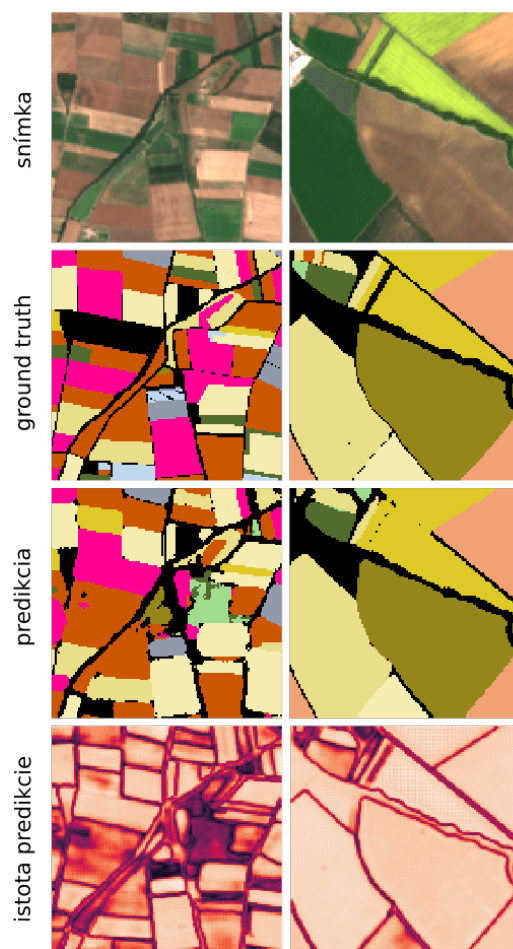


(b) So Smooth Loss prístupom

Obr. 4.9: Distribúcia počtu pixelových predpovedí s konkrétnou istotou predpovede. Obrázok a) zľava zobrazuje distribúciu pre pixely z vnútorných oblastí a následne z hraničných oblastí pre základný model W-TAE b) zobrazuje zľava opäť distribúciu najprv pre vnútorné oblasti polí a následne pre hranice a to pri použití Smooth Loss prístupu.



Obr. 4.10: Prehľad predikcie modelu W-TAE, ktorý predpovedá súčasne plodinovú masku ale aj hranice sémanticky rovnorodých oblastí. Červené obdĺžniky v predikcii hranice označujú nedokonalosti v označení pixelov, ktoré by mali byť hraničnými. V niektorých prípadoch (modrý obdĺžnik) model detekoval dodatočné rozčlenenie parciel aj keď nebolo vyznačené v ground truth.



Obr. 4.11: Prehľad predikcie modelu W-TAE po použití Smooth Loss funkcie. Istota predikcie hraničných pixelov sa zdá byť nižšia v porovnaní napr. s obr. 4.10, čo je reprezentované znížením jasů hraničných pixelov.

Experiment	S2TSCZCrops		
	mIoU	F1 score	Acc
model	73.31	82.78	91.59
model + soft homogenizácia	80.47	87.69	95.85
model + hard homogenizácia	80.18	87.50	95.63

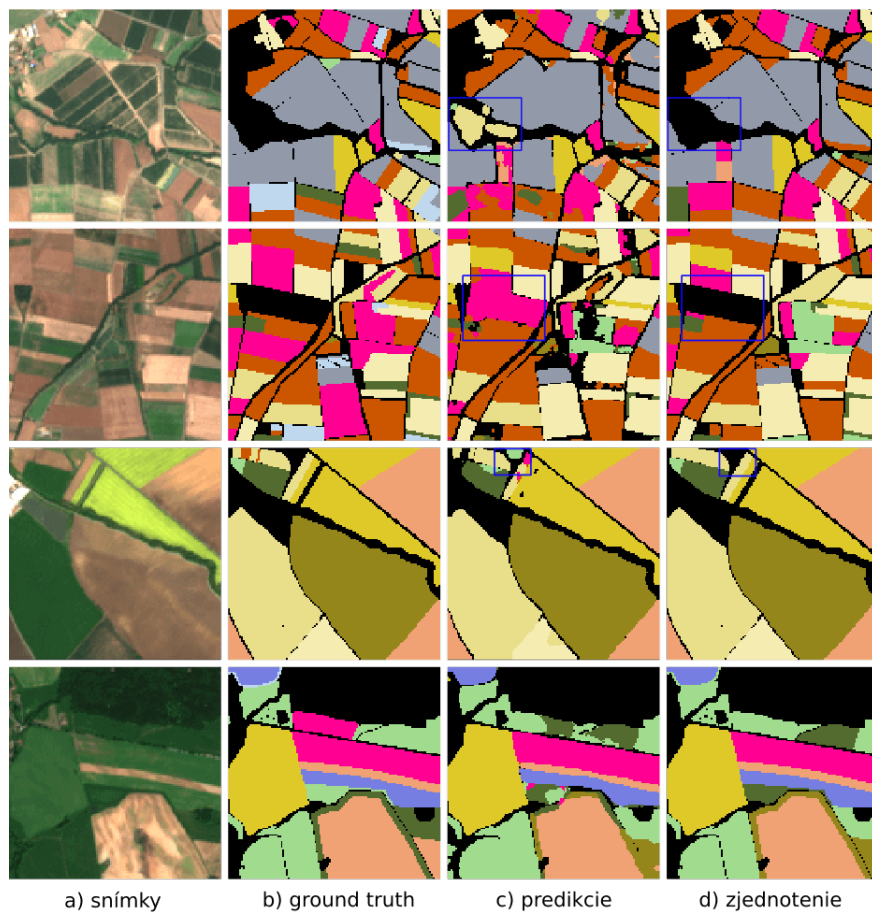
Tabuľka 4.5: Porovnanie výsledkov modelu (W-TAE) vs. toho istého modelu po aplikácii post-processingu t.j. zjednotenia pixelových predikcií na úrovni DPB (uvedené v %). Oba prezentované typy zjednotenia (soft a hard) dosahujú porovnateľnú kvalitu segmentácie.

hranicu polí (vid'. predchádzajúcu sekciu 4.2). Tabuľka 4.5 zaznamenáva výsledok pri použití soft verzie (soft predpovede) ale aj hard verzie (majoritné hlasovanie). Oba prístupy zvýšili segmentačnú kvalitu zhruba o 7 percentuálnych bodov čo sa týka metriky mIoU. Kvalitatívne porovnanie predikovaných segmentačných masiek a výstupu po aplikácii soft zjednotenia na úrovni parciel nájdeme na obr. 4.12.

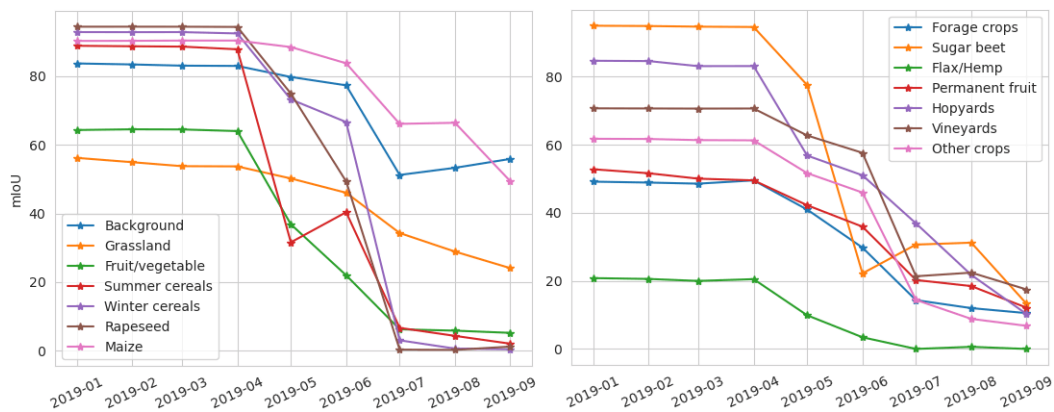
4.4 Závislosť na dĺžke časovej rady

Pre overenie závislosti výslednej segmentačnej kvality modelov na dĺžke použitej časovej rady sme podrobili model W-TAE vyhodnocovaniu na testovacom datase s premenlivými dĺžkami časových rád snímok. Celkovo sme vykonali dva experimenty. V prvom experimente sme stanovili konkrétny počiatočný dátum z ktorého môže snímka pochádzať a použili všetky pozorovania počnúc počiatočným dátumom až po 1.11.2019. Ďalej sme počiatočný dátum menili pričom sa celkový interval možných pozorovaní v rámci časovej rady skracoval (zľava smerom doprava na časovej ose). Výsledok pre jednotlivé triedy je zobrazený na podobrázku a) v 4.13. Druhý experiment bol podobný pričom v tomto prípade sme interval možných pozorovaní predlžovali zľava smerom doprava (na časovej ose) t.j. na vyhodnotenie boli použité všetky snímky od 1.9.2018 až po konkrétny terminálny dátum. Výsledok tohto experimentu možno vidieť na podobrázku b) v 4.13.

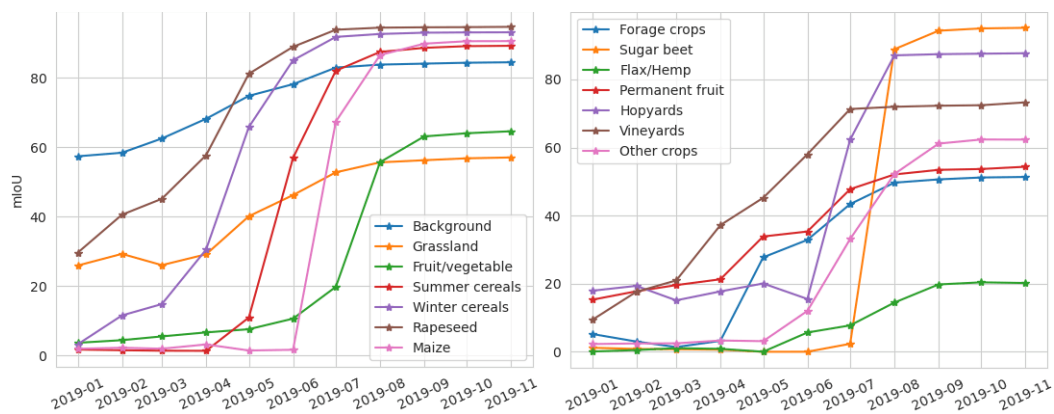
Na oboch podobrázkoch je vidno zjavný trend a síce v a) vidíme, že až v intervale od 1.4.2019 do 1.5.2019 dochádza u väčšiny plodín k významnejšiemu poklesu mIoU, čo naznačuje malú dôležitosť pozorovaní spred 1.4.2019. Podobne krivky v b) naznačujú, že pre všetky plodiny stačia pozorovania zhruba do 1.9.2019 pričom snímky po tomto dátume majú už len minimálny vplyv na zlepšenie segmentačnej kvality.



Obr. 4.12: Kvalitatívne výsledky po aplikácii zjednotenia predikcie pomocou soft distribúcie na základe LPIS dát. Modré obdĺžniky určujú oblasti kde model detekoval plodinu aj keď v ground truth nebola (vid'. c)) explicitne vyznačená. Nakoľko je, ale zjednotenie predikcie vykonané na základe LPIS dát, ktoré vo vyznačených častiach obrázkov predpokladajú inú ako poľnohospodársku pôdu, tak dochádza k ich „eliminácii“ (vid'. d)).



(a)



(b)

Obr. 4.13: Na podobrázku a) vidíme hodnoty mIoU pre jednotlivé plodiny získané pri použití všetkých pozorovaní počnúc dátumom korešpondujúcim so znakom hviezdčky až po 1.11.2019. t.j. napr. hodnoty pre 2019-01 znamenajú, že boli využité všetky snímky od 1.1.2019 až po 1.11.2019 prípadne hodnoty pre 2019-09 reprezentujú výsledok po využití pozorovaní len z intervalu 1.9.2019-1.11.2019. Podobne je na obrázku b) vyobrazená kumulatívna hodnota mIoU kedy boli použité pozorovania vždy od 1.9.2018 až po dátumy korešpondujúce s jednotlivými bodmi (označené hviezdčkou) pričom tentoraz sa interval možných pozorovaní predlžoval. V tomto prípade teda hodnoty pre bod 2019-01 udávajú mIoU pri použití snímok z intervalu 1.9.2018 až po 1.1.2019. Kvôli prehľadnosti sú grafy rozdelené vždy na dva.

Kapitola 5

Diskusia

V tejto kapitole sa pokúsime kriticky zhodnotiť dosiahnuté výsledky a to jednak v zmysle nami navrhnutých architekturných zmien a modelov ale aj vytvoreného datasetu plodín pre Českú republiku - S2TSCZCrops. Ďalej uvedieme porovnanie s alternatívnymi prístupmi k riešeniu problematiky sémantickej segmentácie plodín.

5.1 Navrhnutá metóda

Na základe dostupnej literatúry na začiatku písania tejto práce sme zvolili model U-TAE [29] ako najlepšie východisko pre návrh metódy a ďalší výskum nakoľko model v danom čase dosahoval SOTA výsledky v oblasti sémantickej segmentácie plodín z časových rád snímok Sentinel-2. Sériou experimentov, ktoré mali za cieľ odhaliť vhodné vylepšenie modelu U-TAE sme prišli s návrhom zmien v pozičnom kódovaní temporálneho encoderu L-TAE ale aj s návrhom zmeny poradia spracovania dát, čo vyústilo do dvoch architekturných nazvaných W-TAE a TimeU-Net. Celkovo sme na základe navrhnutých vylepšení a architekturných prípadne za pomoci regularizačných metód boli schopný dosiahnuť vyššiu kvalitu segmentácie ako so samotným modelom U-TAE.

V ďalšej sekcii zhodnotíme vytvorený dataset S2TSCZCrops, kde popíšeme okrem iného aj problém so šumom v referenčných dátach, ktorý nám znemožňuje robiť závery o kvalite modelov resp. použitých modifikácii a preto postavíme závery iba vzhľadom k výsledkom na datasete PASTIS.

Výsledky prezentované v tabuľkách 4.1, 4.2 a 4.3 nenaznačujú významný prínos zmien v pozičnom kódovaní samých o sebe s výnimkou modelu W-TAE, ktorý dosiahol porovnateľnú hodnotu mIoU (63.32%) s modelom U-TAE v základnom nastavení (63.17%) až po využití vylepšeného pozičného kódovania pe^E . Vyššie hodnoty mIoU boli typicky dosiahnuté až pri súčasnej aplikácii zmeny v pozičnom kódovaní a konkrétnej regularizačnej technike. V prípade modelu U-TAE sa zdá, že prínosným bolo len nastavenie kedy sme využili fixované a učené kódovanie spolu temporálnym dropoutom. Výsledná hodnota mIoU 63.66% je porovnateľná s výsledkom prezentovaným v [68] kde autori použili temporálny dropout na model U-TAE v základnom nastavení. Model W-TAE vykazuje čiastočné zlepšenie taktiež pri použití temporálneho dropoutu pričom najlepší klasifikačný výsledok bol dosiahnutý pri spojení vylepšeného pozičného kódovania a label smoothingu. Experimenty s aplikáciou augmentácie datasetu popísanou v sekcii 3.4.3 neprieniesli očakávané zlepšenie ani napriek zvýšenému počtu tréningových epôch z pôvodných 70 na 100.

Celkovo je zrejmé, že modely U-TAE a W-TAE sú veľmi podobné čo sa týka kvality sémantickej segmentácie. Takýto záver nie je prekvapivý ak vezmeme do úvahy, že oba modely generujú

temporálne reprezentácie časových rád snímok veľmi podobným spôsobom. Čo sa týka modelu TimeU-Net tak tento dosahuje celkovo najlepší segmentačný výsledok spomedzi nami navrhnutých prístupov pričom ide o model v základnom nastavení s použitím label smoothingu ako aj temporálneho dropoutu. Zvýšenú kvalitu sémantickej segmentácie si vysvetlí ujdeme predovšetkým absenciou priestorovej aproximácie výpočtu attention skóre masiek čo umožňuje modelu lokalizovať oblasti s relevantnými plodinami s vyššou mierou detailu. Na druhú stranu zlepšenie kvality segmentácie nie je v prípade tohto modelu jednoznačné keďže výrazne lepší výsledok oproti ďalším uvažovaným úpravám bol dosiahnutý len v jednom prípade. Každopádne veríme, že výsledok pri najmenšom prezentuje potenciál modelu TimeU-Net priniesť plodinové mapy s vysokou presnosťou.

5.2 Dataset S2TSCZCrops

V sekcii zaoberajúcej sa charakteristikou datasetu S2TSCZCrops 2.3 sme uviedli, že referenčné dáta, ktoré máme k dispozícii sú výstupom iného modelu, konkrétne algoritmu Sen2Agri [6] o ktorom máme znalosť, že by mal dosahovať presnosť (accuracy) na úrovni parciel viac ako 90%. Je zrejmé, že z tohto dôvodu je v anotáciách použitých referenčných dát prítomný v istej miere šum, čo nám znemožňuje robiť jednoznačné závery ohľadom vhodnosti architekturných zmien a návrhov vylepšení uvedených v kapitole 3. Aj napriek tomuto problému považujeme dataset S2TSCZCrops za dôležitý výstup našej práce keďže podľa našej znalosti v súčasnosti nejestvuje pre územie Českej republiky iný dataset plodín pozostávajúci z časových rád snímok Sentinel-2.

Jednou z možností ako odhadnúť vplyv šumu v referenčných dátach na kvalitu výsledných modelov je použiť iný dataset, ktorý je do istej miery podobný S2TSCZCrops a obsahuje anotácie s vysokou mierou presnosti. V našom prípade by mohlo ísť napr. o PASTIS [29] o ktorom vieme, že jeho presnosť by mala presahovať 98%. Následne by bolo možné umelo vkladať v procese tréovania vopred určené množstvo šumu a sledovať jeho vplyv na úpadok kvality tréovaného modelu voči presnej anotácii. Nakoniec by sa zo získaného výsledku odvodila možná hodnota presnosti našich modelov voči skutočnému ground truth. Na vykonanie popísaného odhadu sme si bohužiaľ nenašli čas a preto sa aspoň pokúsime na základe znalosti presnosti podkladového algoritmu Sen2Agri [6] voči reálnemu ground truth a presnosti nami získaných modelov (voči ground truth generovanému pomocou Sen2Agri) odhadnúť hornú a dolnú hranicu v ktorej sa pohybuje reálna presnosť našej metódy voči skutočnému ground truth.

Označme chybu modelu Sen2Agri voči skutočnému ground truth ako $e_{Sen2Agri}$ o ktorej vieme, že platí $e_{Sen2Agri} \leq 0.1$. Označením „skutočné ground truth“ máme na mysli bezchybné anotácie. Ďalej potrebujeme poznať odhad chyby resp. metriky accuracy našej metódy, ktorá bola tréovaná voči dátam generovaným modelom Sen2Agri. Túto hodnotu sme dokázali odhadnúť po aplikácii post-processingu t.j. zjednotenia pixelových predikcii pomocou LPIS dát. Accuracy modelu WTAE na úrovni parciel sme odhadli na 83%. Jeho chyba je teda zrejme $e_{model} = 0.17$. Prejdime k samotným odhadom.

Hypoteticky sa môže v najlepšom prípade stať, že náš model sa naučí robiť chyby v klasifikácii voči ground truth generovanému modelom Sen2Agri tak, že tieto chyby budú presne kontradikovať chyby, ktoré vykonáva model Sen2Agri voči skutočnému ground truth a teda dôjde k pomyselnéj eliminácii chýb. Tento poznatok môžeme popísať ako $|e_{Sen2Agri} - e_{model}|$ pričom absolútna hodnota v predpise slúži na ošetrenie prípadu kedy chybovosť nášho modelu prevýši chybovosť

modelu Sen2Agri t.j. $e_{Sen2Agri} \leq e_{model}$. Celkovo je horný odhad daný ako

$$accuracy_{skutočná} \leq 1 - |e_{Sen2Agri} - e_{model}| \quad (5.1)$$

Podobne v najhoršom možnom scenári sa stane to, že náš model sa naučí robiť tie isté resp. ekvivalentné chyby ako model Sen2Agri a navyše vykoná $e_{model} \times 100\%$ chýb voči referenčným dátam generovaným algoritmom Sen2Agri. V takomto prípade získame sčítaním spomínaných dvoch zdrojov chýb dolný odhad presnosti našej metódy voči skutočnému ground truth a síce

$$accuracy_{skutočná} \geq 1 - (e_{Sen2Agri} + e_{model}) \quad (5.2)$$

Keďže sme chybu modelu W-TAE odhadli na $e_{model} = 0.17$ tak môžeme dané odhady kvantifikovať na

$$accuracy_{skutočná} = 0.83 \pm 0.1 \quad (5.3)$$

čo znamená, že spodná hranica presnosti na úrovni parciel našej metódy voči skutočnému ground truth je 73%. Šírka intervalu odhadu s hodnotou 20% nám ako sme už avízovali, znemožňuje robiť akékoľvek závery o vhodnosti navrhnutých metód na základe datasetu S2TSCZCrops.

Vyjadriť sa ďalej ku kvalite klasifikačných výsledkov jednotlivých uvažovaných tried plodín datasetu S2TSCZCrops. Najmenej úspešný bol model v prípade triedy Flax & Hemp, ktorá je zároveň najnižšie zastúpenou triedou v rámci datasetu, čo môže byť práve príčina nízkej presnosti jej klasifikácie. Na druhú stranu plodiny reprezentované triedami Summer cereals, Winter cereals, Rapeseed, Maize a Sugar beet boli klasifikované s vysokou presnosťou za čo podľa nás vďaka ačtia jednak ich relatívne vysokému zastúpeniu v rámci datasetu, ale samozrejme aj špecifickým temporálnym profilom, ktoré tieto triedy plodín vykazujú. Vysoké zastúpenie triedy v rámci datasetu, ale nie je až také podstatné za predpokladu výrazne odlišného temporálneho vývoja spektrálnej charakteristiky rastliny. Toto tvrdenie podporuje vysoká hodnota metriky mIoU triedy Hopyards, ktorá je druhou najnižšie zastúpenou plodinou v datasete S2TSCZCrops pričom jej temporálny profil NDVI¹ je odlišný od všetkých ostatných plodín. Ďalej môžeme poznamenať, že podobnosti v temporálnych profiloch dvojíc tried Grassland, Forage crops a Fruit & vegetables, Other crops sa prejavujú v ich častej vzájomnej zámene. Triedy Grassland, Vineyards a Permanent fruit boli často zamieňané s triedou reprezentujúcou pozadie (Background). Takáto zámena je, ale očakávaná keďže vinice či prípadne ovocné sady sú často pestované v bezprostrednej blízkosti alebo priamo v zastavaných oblastiach, ktoré sú súčasťou triedy Background. Podobne trieda reprezentujúca ovocné sady (Permanent fruit) bola v niektorých prípadoch zamenená za triedu Grassland, čo je vysvetliteľné tým, že povrch v okolí sadov môže byť pokrytý trávou a teda dochádza k znečisteniu spektrálnej odozvy triedy Permanent fruit. Spomínané dôvody chýb v klasifikácii t.j. podobnosť v temporálnych profiloch a nerovnomernosť zastúpenia pixelov v rámci tréningového datasetu sú prirodzenou súčasťou riešeného problému a boli preto očakávané. Celkovo považujeme získané výsledky za uspokojivé.

5.3 Hraničné pixely

Argumentujúc principiálnou nemožnosťou jednoznačne určiť práve jednu anotáciu pixelu, ktorý sa vyskytuje na hranici dvoch resp. viacerých sémanticky rovnorodých oblastí sme navrhli dve možnosti ako zlepšiť reprezentáciu hraničných pixelov a to jednak v zmysle pravdepodobnostnej predpovede modelu s použitím Smooth Loss funkcie ale aj v zmysle dodatočnej binárnej sémantickej segmentácie hranice.

¹Temporálny profil triedy Hopyards je dostupný na https://github.com/Many98/Crop2Seg/blob/main/data/temporal_profiles/hopyards.png

Vhodnosť návrhu Smooth Loss funkcie je v istom zmysle diskutabilná. Je zrejmé, že v rámci každého hraničného pixelu je miera ktorou daný povrch prispieva do spektrálnej zmesi odlišná a pravdepodobne je úmerná ploche prieniku pixelu a hranice sémanticky rovnorodnej oblasti. V našom návrhu, ale používame rovnomerné prerozdelenie medzi hraničné pixely a nepočítame s plochami prieniku. V tomto zmysle je úprava ground truth distribúcie pomocou predpisu 3.15 rovnako nevhodná ako prístup s one-hot distribúciou 3.11. Aj keď nie je náš zjednodušený prístup úplne korektný tak ho považujeme za krok bližšie k lepšej reprezentácii predpovedanej distribúcie hraničných pixelov. Keďže nemáme k dispozícii nijakú vhodnú mieru ktorou môžeme toto tvrdenie podporiť tak si musíme vystačiť s obrázkom 4.9 kde sú zobrazené distribúcie počtu predpovedí s konkrétnou istotou. Upresníme, že istotou predpovede myslíme pravdepodobnosť (výstup po aplikácii softmax) top-1 predikcie modelu. Ako bolo očakávané tak model sa naučil predpovedať hraničné pixely s výrazne nižšou istotou pričom distribúcia je koncentrovaná okolo intervalu 46-53% (vid' obr. 4.9 b) napravo). V budúcnosti by bolo vhodné upraviť metódu 3.15 tak aby vzala do úvahy napr. už spomínanú plochu prieniku pixelu a hranice sémanticky rovnorodnej oblasti.

Druhý prístup, ktorý spočíva v explicitnej predpovedi resp. označení hraničných pixelov nie je ničím výnimočný. Ide o binárnu sémantickú segmentáciu hranice s ktorou si model podľa nášho názoru poradil veľmi dobre (vid'. obr. 4.10) pričom dosiahol metriku mIoU až 81%. Výsledok v tab. 4.4 taktiež naznačuje priaznivý vplyv súčasnej predpovede segmentácie plodín ako aj hranice, čo si vysvetlíme tým, že model mal k dispozícii dva zdroje supervízie. Problematickým v tomto ohľade vidíme len postup akým je určená hranica. V snahe ušetriť úložný priestor ako aj čas sme sa totiž rozhodli generovať informáciu o hranici na základe už existujúcej segmentačnej masky plodín a to za behu programu pomocou morfolologickej operácie dilatácia. Týmto spôsobom, ale nedokážeme od seba odlíšiť dve hraničné parcely resp. DPB s rovnakou plodinou. Z tohto dôvodu pracujeme len so sémanticky rovnorodými oblasťami, ktoré vznikli po rasterizácii vektorových dát. Ďalším problémom vyplývajúcim so spôsobu akým generujeme informáciu o hranici je, že hranica má typicky šírku dva pixely, čo samozrejme nie je optimálne. V ďalšom výskume by teda bolo vhodné použiť presnejšiu informáciu o hraniciach získanú priamo z vektorových dát LPIS.

5.4 Post-processing

Post-processing v zmysle zjednotenia pixelových predpovedí na úrovni DPB (vid'. obr. 4.12) je zamýšľaný ako praktický krok, ktorý prinesie užívateľovi navyše predpoveď typu plodín na úrovni parciel. Takýto prístup by sa mohol zdať zbytočný argumentujúc, že metóda mohla byť navrhnutá už od začiatku ako klasifikácia na úrovni parciel. Podľa nášho názoru by sme sa, ale týmto spôsobom obrali o flexibilitu pixelovej predikcie, ktorá má potenciál detekovať nezrovnalosti v pestovaní plodín na konkrétnej parcele. Navrhnutý post-processing je teda podľa nás prijateľný kompromis.

Aj keď výsledky zjednotenia (vid'. tab. 4.5) poukazujú na značné zlepšenie metriky mIoU oproti predpovedi samotného modelu W-TAE tak ich nepovažujeme za preukazné kvôli už spomínanému problému so šumom v referenčných dátach datasetu S2TSCZCrops. Nechávame teda na čitateľovi aby si urobil závery sám. Na druhú stranu, nám zjednotenie predikcie umožnilo vypočítať accuracy na úrovni parciel resp. DPB a tým vyčíslit' spodný odhad možnej accuracy našej metódy voči skutočnému ground truth. Accuracy na úrovni objektov sme určili na 83% voči poskytnutým referenčným dátam. To znamená, že pri najhoršom je presnosť našej metódy na úrovni objektov voči skutočnému ground truth 73%.

5.5 Závislosť na dĺžke časovej rady

Časová rada snímok Sentinel-2, ktorú používame pozostáva z pozorovaní od 1.9.2018 až po 1.11.2019. Je zrejmé, že nie všetky pozorovania budú mať rovnaký prínos na výslednú segmentačnú kvalitu modelu. Aby sme preskúmali vplyv zmien dĺžky časovej rady snímok na výslednú hodnotu metriky mIoU modelu W-TAE vykonali sme dva experimenty.

V prvom sme postupne skracovali časovú radu zľava (na časovej ose). V tomto prípade model dostával stále menej a menej pozorovaní a teda typicky dochádzalo k úpadku mIoU ako je zobrazené na podobrázku a) v 4.13. Výsledok poukazuje na minimálny vplyv pozorovaní spred 1.4.2019 za predpokladu použitia všetkých snímok až po 1.11.2019 nakoľko na grafe vidíme len minimálny pokles do 1.4.2019. Aj keď takýto záver nie je veľmi prekvapivý keďže hlavné vegetačné obdobie rastlín resp. fenologická fáza v ktorej dochádza k rastu nad úrovňou zeme začína typicky koncom marca, tak sme očakávali výraznejšiu zmenu minimálne v prípade ozimných plodín (Winter cereals a Rapeseed). Z praktického hľadiska možno túto znalosť využiť na zníženie pamäťových nárokov spojených so spracovaním alebo už samotným získaním dlaždíc Sentinel-2. Pre predstavu predpokladajme, že použijeme 2 dlaždice za mesiac. Veľkosť dlaždice Sentinel-2 v L2A spracovaní je typicky viac ako 1 GB. Zmenšením dĺžky požadovanej časovej rady v zmysle časového okna od 1.9.2018 do 1.11.2019 na okno od 1.4.2019 do 1.11.2019 teda ušetríme aspoň 14 GB úložného priestoru.

V prípade druhého experimentu sme sledovali zmeny v mIoU pri predlžovaní použitej časovej rady opäť zľava (na časovej ose) t.j. použili sme všetky snímky od 1.9.2018 až po uvažovaný dátum. Vo všeobecnosti kvalita segmentácie vzhľadom k mIoU postupne narastala. Výsledok daného experimentu môžeme pozorovať na podobrázku b) 4.13, kde je opäť zjavný trend a síce pri použití všetkých snímok od 1.9.2018 má použitie snímok po 1.9.2019 zanedbateľný vplyv na kvalitu segmentácie. Toto zistenie korešponduje so zistením v prvom experimente. Praktický rozmer tohto experimentu spočíva v zistení doby v rámci roku kedy je možné dosiahnuť pre jednotlivé plodiny segmentáciu s postačujúcou kvalitou napr. z dôvodu potreby získať skorú predpoveď v strede vegetačného obdobia plodín. Napr. pre triedu Maize reprezentujúcu plodinu kukurica pozorujeme, že model nie je schopný túto plodinu odlíšiť od ostatných dokým nemá k dispozícii pozorovania po 1.6.2019 nakoľko v tomto období dochádza k výraznému skoku hodnoty mIoU. Toto pozorovanie korešponduje s temporálnym profilom kukurice², kde vidíme náhly nárast hodnoty NDVI v mesiaci jún, čo značí náhly nárast množstva zelenej hmoty. Podobne trieda Sugar beet je modelom takmer nedetekovaná pri pozorovaniach do 1.7.2019 pričom náhly vzrast mIoU zaznamenávame v siedmom mesiaci. Ozimné plodiny ako Winter cereals a Rapeseed je model schopný odlíšiť s vysokou presnosťou pri použití snímok do 1.5.2019. Repka olejná (Rapeseed) je dokonca detekovaná s hodnotou mIoU viac ako 50% pri použití snímok len do 1.4.2019. V prípade repky, ale takáto kvalita segmentácie nie je prekvapivá ak vezmeme do úvahy fázu jej kvitnutia (apríl a máj), ktorá je typická jasne žltými kvetmi³.

5.6 Porovnanie s alternatívnymi prístupmi

Pre porovnanie navrhnutého prístupu s modelmi popísanými v sekcii 3.3 sme zvolili použitie benchmarkového datasetu PASTIS. Výsledky sú zhrnuté v tabuľke 5.1. Podotkneme, že z časového hľadiska sme nedokázali všetky alternatívne metódy experimentálne replikovať. Konkrétne, nižšie uvedené výsledky segmentačnej kvality modelov TSViT, Exchanger s U-Net a Exchanger s

²Temporálny profil je dostupný na adrese https://github.com/Many98/Crop2Seg/blob/main/data/temporal_profiles/maize.png

³Snímka obsahujúca kvitnúcu repku je zobrazená na obr. 4.6 (tretí stápec zľava)

Mask2Former modelom vzhľadom k metrike mIoU sme prebrali z [78] a [19]. Ďalej implementácie ako aj nastavenie hyperparametrov ConvLSTM, U-ConvLSTM, 3D U-Net a U-TAE (pôvodný model) sme prebrali z [29].

Naivný prístup s modelom U-Net v ktorom zlučujeme temporálnu a spektrálnu dimenziu snímok sa ukázal byť najmenej úspešný. Samozrejme toto bolo očakávané keďže model nijakým spôsobom nevyužíval vo svoj prospech temporálne poradie snímok. Jeho výhodou je len nízky inferenčný čas. Na druhú stranu 3D U-Net model je schopný dosiahnuť prakticky porovnateľný inferenčný čas a pritom vyššiu kvalitu segmentácie.

Modely využívajúce hybridnú konvolučne-rekurentnú architektúru ConvLSTM t.j. samotná ConvLSTM a U-ConvLSTM dosahujú porovnateľné výsledky jednak vzhľadom ku kvalite segmentácie ale aj hodnoty inferenčného času. Celkovo ale zaostávajú napr. za modelom 3D U-Net a modelmi založenými na attention mechanizme. Výrazne vysoká hodnota FLOPs v prípade ConvLSTM súvisí s nastavením počtu konvolyčných filtrov skrytého stavu ConvLSTM bunky a so sekvenčným charakterom spracovania dát t.j. aplikácie rovnakej ConvLSTM bunky T-krát kde T je počet pozorovaní v rámci časovej rady.

Replikáciou modelu U-TAE sme dospeli k hodnote mIoU rovnej 63.2%. Táto hodnota sa líši od hodnoty reportovanej v [29] o 0.1%, čo je pravdepodobne čiastočne spôsobené stochastickým charakterom niektorých operácií v PyTorch, ale čiastočne aj spôsobom normalizácie snímok. Autori v [29] totiž používali na normalizáciu testovacích resp. validačných podmnožín dát štatistiky vypočítané vzhľadom k týmto testovacím a validačným setom. Tento prístup nie je úplne korektný nakoľko pri ňom jednak dochádza k úniku informácie, ale zároveň nie je použiteľný na inferenciu nad jednou vzorkou. Zmenou na normalizáciu pomocou štatistík generovaných tréningovým datasetom sme dosiahli prezentovanú hodnotu 63.2%. Použitím (fixného + učeného) pozičného kódovania 3.9 spolu s temporálnym dropoutom sme umožnili modelu ešte čiastočne navýšiť kvalitu segmentácie v zmysle celkového skóre mIoU a to zhruba o 0.5%.

Aj keď upravená verzia U-TAE dosahuje celkovú metriku mIoU vyššiu ako najlepší model W-TAE (vylepšené pozičné kódovanie a label smoothing) tak tento model stále považujeme za relevantný keďže ide o akúsi zovšeobecnenú a flexibilnejšiu verziu U-TAE. Ako sme už v rámci popisu modelu W-TAE zdôrazňovali, model je navrhnutý tak aby nebol nutne závislý na použití segmentačného modelu U-Net, ale umožňoval využiť výkonnejší segmentačný model. Rovnaký záver ohľadom flexibility platí aj na model TimeU-Net, ktorý naviac dosiahol lepšie mIoU skóre zhruba o 1% oproti upravenej verzii U-TAE. Zlepšenie kvality sémantickej segmentácie jednotlivých plodín modelu TimeU-Net oproti U-TAE podporuje aj stĺpcový graf 4.4 kde vidíme, že v prípade väčšiny uvažovaných plodín datasetu PASTIS dosahuje metrika mIoU vyššiu hodnotu v prospech modelu TimeU-Net. Takéto zlepšenie pravdepodobne súvisí s výpočtom temporálnej reprezentácie (attention skóre) pre každý pixel nezávisle pričom nedochádza k žiadnej priestorovej aproximácii (pre porovnanie vid'. 4.5).

Celkovo môžeme povedať, že nami navrhnuté resp. pozmené architektúry dosahujú vyššiu hodnotu mIoU ako pôvodný návrh U-TAE z ktorého sme vychádzali. Jednomyseľne musíme dodať, že ich kvalita stále zaostáva za modelom TSViT a za súčasnými SOTA architektúrami využívajúcimi Exchanger model. Na druhú stranu sú modely W-TAE a TimeU-Net ľahšie jednak v zmysle počtu parametrov ale aj počtu floating point operácií a môžu naviac ponúknuť nižší inferenčný čas.

Názov	mIoU	# parametrov	FLOPs	Inferenčný čas
U-Net (TxC)	52.0	14.58 M	48 G	7 ms
ConvLSTM [63, 72]	57.5	1.01 M	482 G	26 ms
U-ConvLSTM [60]	57.6	1.52 M	50 G	25 ms
3D U-Net [60]	58.4	1.55 M	93 G	8 ms
U-TAE (pôvodná) [29]	63.2	1.09 M	47 G	22 ms
W-TAE (navrhnutá metóda)	63.5	1.18 M	27 G	22 ms
U-TAE (upravená)	63.7	1.09 M	47 G	22 ms
TimeU-Net (navrhnutá metóda)	64.8	1.05 M	40 G	70 ms
TSViT [78]	65.4	2.16 M	582 G	213 ms
Exchanger+U-Net [19]	66.8	8.08 M	300 G	136 ms
Exchanger+Mask2Former [19]	67.9	24.59 M	329 G	-

Tabuľka 5.1: Porovnanie modelov vzhľadom k dosiahnutému mIoU (v %) na datasete PASTIS a vzhľadom k počtu parametrov, počtu floating point operácii (FLOPs) a inferenčnému času. Hodnoty FLOPs a inferenčný čas sú odhadované za použitia časovej rady rozmerov $[T \times C \times H \times W] = [30 \times 10 \times 128 \times 128]$. Hodnoty mIoU sú vypočítané na základe všetkých piatich foldov v rámci datasetu PASTIS. Inferenčný čas vyjadruje čas potrebný na jeden beh forward fázy modelu. Všetky výpočty boli vykonané na grafickej karte NVIDIA Tesla A100.

Kapitola 6

Návrh aplikácie

Praktický rozmer našej práce je realizovaný návrhom a implementáciou aplikácie schopnej automaticky získať potrebné Sentinel-2 dáta a na základe modelu s najlepšími klasifikačnými schopnosťami vykonať sémantickú segmentáciu plodín nad územím Českej republiky. Implementácia prebiehala v jazyku Python a je riešená formou webového rozhrania postaveného na frameworku Streamlit [14]. Zdrojové kódy súvisiace s touto prácou sú verejne dostupné na <https://github.com/Many98/Crop2Seg> kde je samozrejme popísaný postup spustenia aplikácie¹. Schématický postup spracovania dát je zobrazený na obr. 6.1.

6.1 Vstupy aplikácie

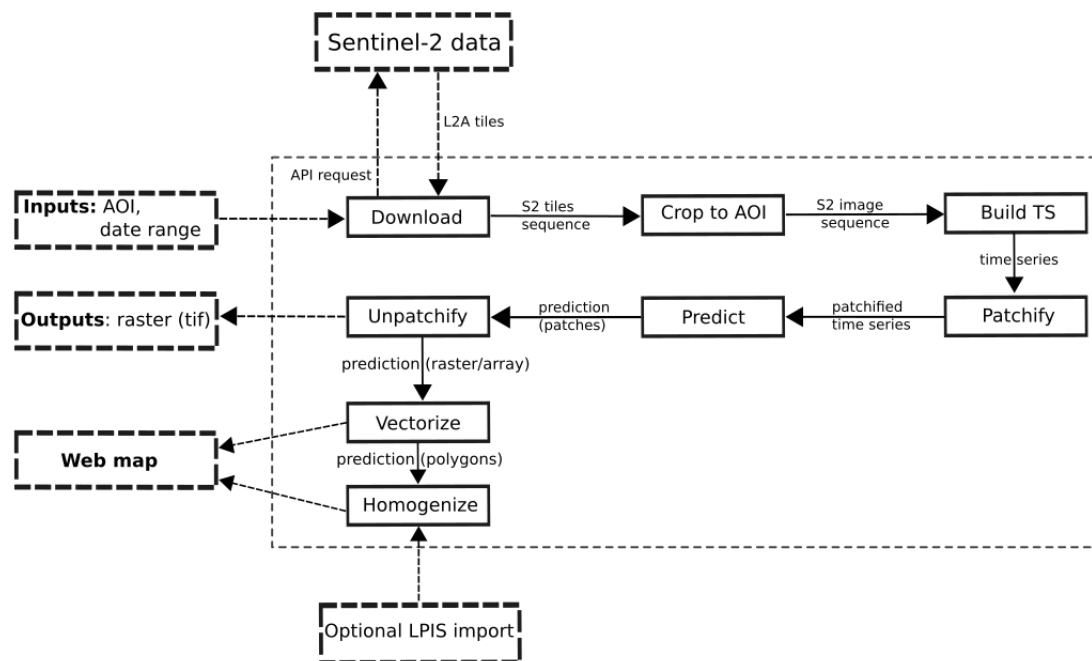
Navrhnutá aplikácia požaduje na vstupe zadanie oblasti záujmu a časového okna na základe ktorého bude stiahnutá množina snímok Sentinel-2. V súčasnej dobe riešime zadanie oblasti záujmu len formou kliknutia na vopred pripravenú štvorcovú mriežku polí rozprestierajúcu sa naprieč Českou republikou. Časové okno je riešené jednak zadaním konkrétneho roku a následne definovaním rozpätia s krokom jeden mesiac. Užívateľ má taktiež možnosť určiť či chce povoliť použitie dát LPIS (pre vybraný rok) a taktiež určiť povolenie na cachovanie stiahnutých dát. Kliknutím na tlačidlo Predict crops sa spustí automatické spracovanie. Znázornenie úvodnej strany aplikácie je na obr. 6.2.

6.2 Spracovanie dát

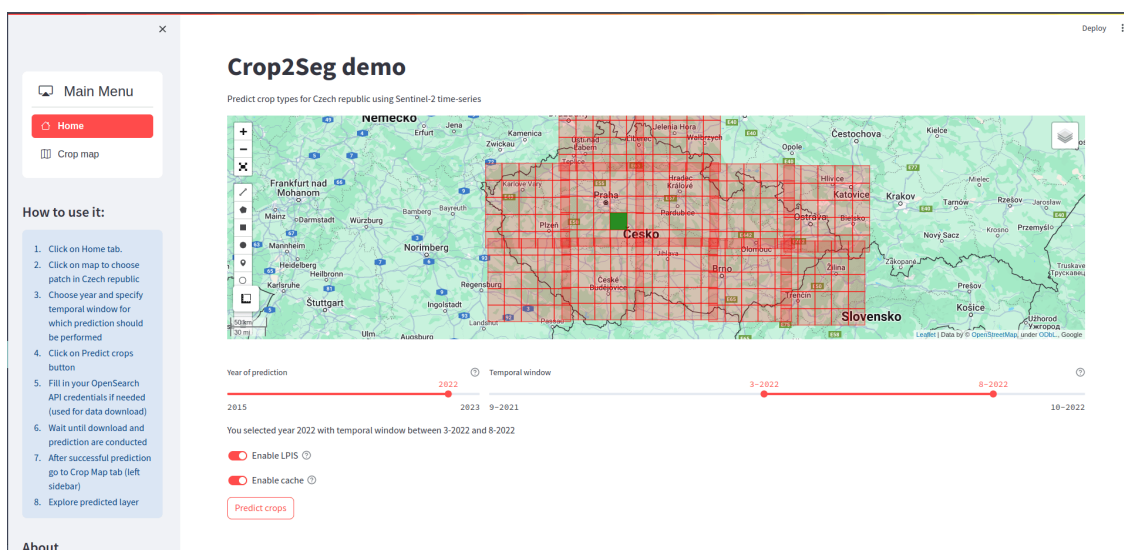
Popíšme v nasledujúcom jednotlivé kroky spracovania dát do detailu.

Získanie dát (Download) Stiahnutie potrebných snímok Sentinel-2 prebieha pomocou API požiadavky na <https://dhr1.cesnet.cz/>. Ak neboli určené prístupové údaje na stránku <https://dhr1.cesnet.cz/> v rámci config súboru tak bude program explicitne požadovať zadanie týchto prístupových údajov nakoľko stiahnutie Sentinel-2 snímok cez OpenSearchAPI vyžaduje prihlásenie. Samotné sťahovanie môže byť v závislosti na rýchlosti internetového pripojenia značne časovo náročné a preto sú stiahnuté dáta na základe povolenia užívateľa a cachované kvôli zrýchleniu budúcich behov programu. Okrem toho podotknime, že sťahované sú celé dlaždice S2

¹Po naklonovaní repozitára a nainštalovaní potrebných balíčkov (requirements.txt) je lokálne spustenie aplikácie možné volaním `streamlit run crop2seg.py`



Obr. 6.1: Schématický návrh automatického spracovania Sentinel-2 dát s cieľom priniest' užívateľovi sémantickú segmentáciu plodín. Podotknime, že zobrazené sú len hlavné kroky postupu.



Obr. 6.2: Úvodná strana aplikácie s možnosťou zadania vstupov a spustenia automatického spracovania.

čo navyšuje taktiež pamäťové nároky aplikácie. Do budúcnosti bude teda vhodné prísť s riešením ako stiahnuť len potrebnú podmnožinu snímky S2.

Orezanie (Crop to AOI) Keďže predpokladáme, že požadovaná oblasť záujmu bude menšia ako rozmer samotnej dlaždice tak dôjde k jej orezaniu na menšie rozmery. V súčasnej verzii tento krok nie je implementovaný a namiesto neho využívame explicitné rozdelenie územia ČR do štvorcových dlaždíc, ktoré slúžia ako oblasť záujmu.

Predspracovanie a zostavenie časovej rady (Build TS) Všetky snímky boli doteraz spracované nezávisle a preto je v tomto kroku potrebné z množiny snímok zostaviť časovú radu aby model mohol využiť informáciu o temporálnom zoradení pozorovaní.

Rozdelenie na menšie patche (Patchify) Vzhľadom k pamäťovým nárokom behu modelu neurónovej siete je potrebné časovú radu snímok podľa potreby ešte rozdeliť na menšie štvorcové podobrázky zvané patche.

Sémantická segmentácia (Predict) Tento krok zahŕňa jednak predspracovanie vo forme normalizácie snímok a následnú inferenciu modelom.

Zostavenie predpovede pôvodných rozmerov (Unpatchify) Po úspešnej predpovedi modelom neurónovej siete aplikujeme spätné zostavenie patchov do rozmerov požadovanej oblasti záujmu. Získaný výstup sa následne prevedie na raster a exportuje sa do formátu tif.

Tvorba vektorových dát (Vectorize) Pre jednoduchšie zobrazenie predpovede modelu vo webovej mape sme pristúpili ku vektorizácii top-1 predikcie. V praxi to znamená, že homogénne oblasti kde model predpovedal triedu A budú označené za samostatný polygón.

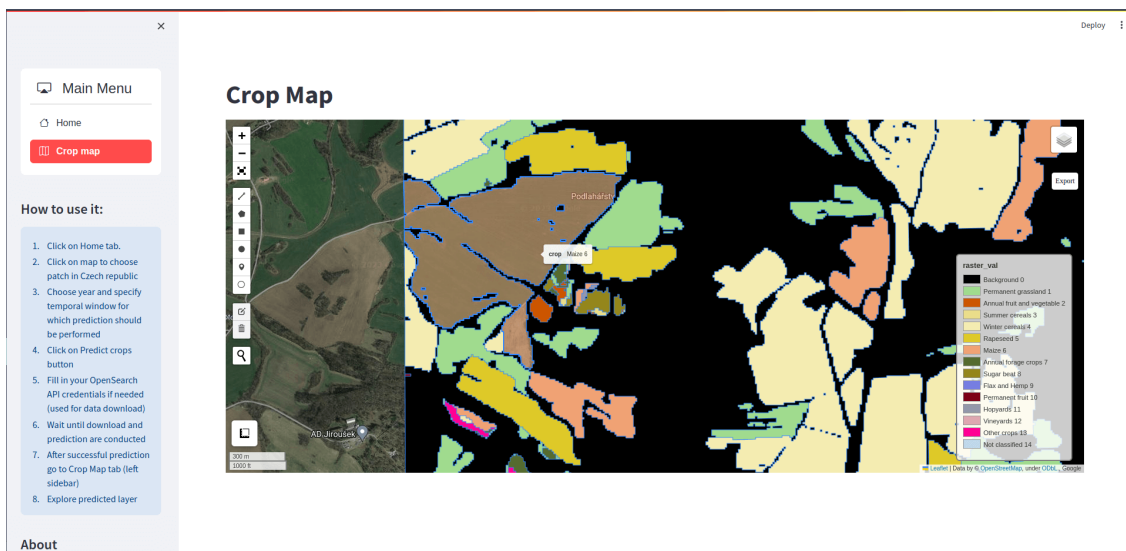
Post-processing - zjednotenie predpovede (Homogenize) Za predpokladu, že užívateľ umožnil použitie LPIS dát, dôjde k stiahnutiu LPIS vrstvy a následnému zjednoteniu pixelovej predikcie tak ako je navrhnuté v sekcii 3.5.

6.3 Výstupy aplikácie

Výstupy aplikácie sú dvojakého druhu. Za prvé ide o webovú mapu, ktorá je generovaná priamo v rámci grafického užívateľského rozhrania aplikácie (Záložka Crop map). Druhým výstupom je samotný raster predpovede vo formáte tif.

Webová mapa (Web map) V tejto mape je zobrazená predpoveď ako polygónová vrstva. Ak užívateľ umožnil použitie LPIS dát tak je vyexportovaná taktiež zjednotená predpoveď ako aj samotná LPIS vrstva. Znázornenie predikcie je na obr. 6.3.

Výstupy (Outputs) Raster vo formáte tif je zamýšľaný ako permanentný výstup s ktorým môže užívateľ neskôr pracovať v špecializovaných GIS nástrojoch ako napr. ArcGIS alebo QGIS. Raster obsahuje top-1 predpoveď modelu ako aj soft predikciu jednotlivých pixelov t.j. pravdepodobnostnú distribúciu jednotlivých tried plodín.



Obr. 6.3: Webová mapa s predikciou plodín.

Záver

Cieľom našej práce bolo preskúmať možnosti sémantickej segmentácie plodín z družicových dát, navrhnúť vhodné rozšírenie neurálnych modelov tak aby boli schopné spracovať časovú radu snímok Sentinel-2 a efektívne využiť temporálnu informáciu v nej obsiahnutú, ďalej využiť dáta z verejného registra pôdy LPIS pre zostavenie datasetu typov plodín pre Českú republiku a nakoniec navrhnutú metódu porovnať s alternatívnymi prístupmi a zhodnotiť jej prínosy.

Pre tento účel sme zvolili model U-TAE ako vhodné východisko a navrhli sme zmeny v pozíciom kódovaní temporálneho encoderu L-TAE ako aj zmeny poradia spracovania dát, čo vyústilo do dvoch architektúr nazvaných W-TAE a TimeU-Net. Celkovo sme na základe navrhnutých vylepšení a architektúr prípadne za pomoci regularizačných metód boli schopný dosiahnuť vyššiu kvalitu segmentácie ako so samotným modelom U-TAE. Ďalej sme pripravili dataset plodín pozostávajúci z časových rád snímok Sentinel-2, ktorý je podľa našej znalosti prvým takého druhu pre Českú republiku. Problém s použitými referenčnými dátami nám, ale znemožnil využiť ho na porovnanie s alternatívnymi prístupmi a preto sme dodatočne využili dataset PASTIS. Experimenty preukázali, že navrhnuté vylepšenia majú potenciál zlepšiť presnosť plodinových máp oproti prístupu z ktorého sme vychádzali. Aj keď naše modely nedosahujú segmentačné kvality ako súčasné SOTA prístupy tak sme priniesli modely, ktoré sú ľahšie v zmysle počtu parametrov ale aj počtu floating point operácií a môžu navyiac ponúknuť nižší inferenčný čas. Okrem toho sú oba modely W-TAE ako aj TimeU-Net navrhnuté tak aby bolo spracovanie temporálnej a priestorovej informácie explicitne oddelené do dvoch krokov resp. častí siete čo zaručuje, že modely sú flexibilnejšou verziou U-TAE a umožňujú nahradiť sieť U-Net za výkonnejší segmentačný model.

Okrajovo sme v rámci práce skúmali možnosti zlepšenia reprezentácie hraničných pixelov. Za prvé sme navrhli zmenu Cross Entropy loss funkcie pomocou label smoothingu s cieľom upraviť pravdepodobnostné rozdelenie pre hraničné pixely čo podľa nás lepšie popisuje nejednoznačnosť anotácie týchto pixelov. Ďalej sme umožnili modelu generovať plodinovú segmentáciu súčasne s dodatočnou binárnou sémantickou segmentáciou hranice. Týmto spôsobom model získal okrem iného ďalší zdroj supervízie čo pravdepodobne vysvetľuje zvýšenie kvality klasifikácie plodín v takomto nastavení.

Práci sme nakoniec dodali praktický rozmer návrhom a implementáciou aplikácie postavenej na získaných modeloch neurónových sietí, umožňujúcou na základe vstupov od užívateľ a stiahnuť potrebné Sentinel-2 snímky a nad zostavenou časovou radou vykonať inferenciu. Užívateľ má následne k dispozícii segmentačnú mapu plodín pre vybrané územie prezentovanú vo webovom rozhraní ale aj v rastrovom formáte tif.

Podotkneme, že aj keď máme k dispozícii modely schopné riešiť sémantickú segmentáciu plodín z časovej rady snímok pre rok 2019, je otáznou, či sú schopné rovnako kvalitne klasifikovať plodiny aj v iných rokoch. Podobne by mohla byť problematickou aj adaptácia na nové regióny ak by sme zamýšľali model použiť na iné oblasti ako ČR. Z tohto dôvodu je podľa nás výskum temporálnej či regionálnej adaptácie vhodným námetom pre ďalšiu prácu.

Literatúra

- [1] The common agricultural policy at a glance. https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-glance_en. online cit.: 2023-08-14.
- [2] Sentinel-2 Spectral Response Functions. https://sentinels.copernicus.eu/documents/247904/685211/S2-SRF_COPE-GSEG-EOPG-TN-15-0007_3.1.xlsx. online cit.: 2023-08-11.
- [3] O aplikaci Registr půdy. <https://eagri.cz/public/web/mze/farmer/LPIS/>, . online cit.: 2023-09-12.
- [4] Uživatelská dokumentace pLPIS - Veřejný Registr půdy. <https://eagri.cz/public/portal/-q335331---agE8ptNB/prirucka-lpis-pro-verejnost>, . online cit.: 2023-09-12.
- [5] Reproducibility in PyTorch. <https://pytorch.org/docs/stable/notes/randomness.html>. online cit.: 2023-11-10.
- [6] Sen2Agri-System. <https://github.com/Sen2Agri/Sen2Agri-System>, . online cit.: 2023-09-04.
- [7] Sen4CAP - Sentinels for Common Agriculture Policy. <http://esa-sen4cap.org/>, . online cit.: 2023-09-04.
- [8] Benefits of Sen4CAP project for SZIF (CZ Paying Agency). https://www.szif.cz/cs/CmDocument?rid=%2Fapa_anon%2Fcs%2Fdokumenty_ke_stazeni%2Fnepub%2F1548398803248%2F1575557477418.pdf, . online cit.: 2023-09-04.
- [9] Sen4CAP - Design Justification File for crop type mapping. http://esa-sen4cap.org/sites/default/files/06_Sen4CAP_DDF_v1.3_CropType.pdf, . online cit.: 2023-09-04.
- [10] Sen4CAP - Validation Report. http://esa-sen4cap.org/sites/default/files/14_Sen4CAP_VR_v1.2.pdf, . online cit.: 2023-09-04.
- [11] APIs Overview. <https://scihub.copernicus.eu/userguide/APIsOverview>, . online cit.: 2023-09-12.
- [12] Sentinel-2 Mission Guide. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2>, . online cit.: 2023-09-15.

- [13] Sentinel-2 MSI User Guide. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi>, . online cit.: 2023-09-15.
- [14] Streamlit. <https://streamlit.io/>. online cit.: 2023-12-14.
- [15] María Arias, Miguel Ángel Campo-Bescós, and Jesús Álvarez Mozos. Crop Classification Based on Temporal Signatures of Sentinel-1 Observations over Navarre Province, Spain. *Remote Sensing*, 12(2), 2020. ISSN 2072-4292. <https://doi.org/10.3390/rs12020278>.
- [16] S. Bailly et al. Crop-Rotation Structured Classification using Multi-Source Sentinel Images and LPIS for Crop Type Mapping. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1950–1953, 2018. <https://doi.org/10.1109/IGARSS.2018.8518427>.
- [17] Mariana Belgiu and Ovidiu Csillik. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204:509–523, 2018. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2017.10.005>.
- [18] Lukas Blickensdörfer et al. Mapping of crop types and crop sequences with combined time series of sentinel-1, sentinel-2 and landsat 8 data for germany. *Remote Sensing of Environment*, 269:112831, 2022. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2021.112831>.
- [19] Xin Cai et al. Revisiting the Encoding of Satellite Image Time Series. *arXiv e-print*, 2023. <https://doi.org/10.48550/arXiv.2305.02086>.
- [20] Isabel Luisa Castillejo-González et al. Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture*, 68(2):207–215, 2009. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2009.06.004>.
- [21] Jose Bermudez Castro, Raul Queiroz Feitosa, and Patrick Nigri Happ. An Hybrid Recurrent Convolutional Neural Network for Crop Type Recognition Based on Multitemporal Sar Image Sequences. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 3824–3827, 2018. <https://doi.org/10.1109/IGARSS.2018.8517280>.
- [22] Xinglu Cheng et al. Application of deep learning in multitemporal remote sensing image classification. *Remote Sensing*, 15(15), 2023. ISSN 2072-4292. <https://doi.org/10.3390/rs15153859>.
- [23] Kyunghyun Cho et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv e-print*, 2014. <https://doi.org/10.48550/arXiv.1406.1078>.
- [24] François Chollet. Xception: Deep learning with depthwise separable convolutions. *arXiv e-print*, 2017. <https://doi.org/10.48550/arXiv.1610.02357>.
- [25] Jacob Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-print*, 2019. <https://doi.org/10.48550/arXiv.1810.04805>.
- [26] Xiangsuo Fan et al. Improved U-Net Remote Sensing Classification Algorithm Fusing Attention and Multiscale Features. *Remote Sensing*, 14(15), 2022. ISSN 2072-4292. <https://doi.org/10.3390/rs14153591>.

- [27] Saskia Foerster et al. Crop type mapping using spectral–temporal profiles and phenological information. *Computers and Electronics in Agriculture*, 89:30–40, 2012. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2012.07.015>.
- [28] G. Fontanelli et al. Agricultural crop mapping using optical and SAR multi-temporal seasonal data: A case study in Lombardy region, Italy. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 1489–1492, 2014. <https://doi.org/10.1109/IGARSS.2014.6946719>.
- [29] V. Fare Garnot and L. Landrieu. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4852–4861, Los Alamitos, CA, USA, oct 2021. IEEE Computer Society. <https://doi.org/10.1109/ICCV48922.2021.00483>.
- [30] Babak Ghassemi et al. Designing a European-Wide Crop Type Mapping Approach Based on Machine Learning Algorithms Using LUCAS Field Survey and Sentinel-2 Data. *Remote Sensing*, 14(3), 2022. ISSN 2072-4292. <https://doi.org/10.3390/rs14030541>.
- [31] Michalis Giannopoulos, Grigorios Tsagkatakis, and Panagiotis Tsakalides. 4D U-Nets for Multi-Temporal Remote Sensing Data Classification. *Remote Sensing*, 14(3), 2022. ISSN 2072-4292. <https://doi.org/10.3390/rs14030634>.
- [32] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. ISSN 0899-7667. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [33] Qiong Hu et al. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *International Journal of Applied Earth Observation and Geoinformation*, 80:218–229, 2019. ISSN 1569-8432. <https://doi.org/10.1016/j.jag.2019.04.014>.
- [34] T. Idol, B. Haack, and R. Mahabir. Radar and optical remote sensing data evaluation and fusion; a case study for Washington, DC, USA. *International Journal of Image and Data Fusion*, 6:138 – 154, 2015. <https://doi.org/10.1080/19479832.2015.1017541>.
- [35] Roberto Interdonato et al. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:91–104, 2019. ISSN 0924-2716. <https://doi.org/10.1016/j.isprsjprs.2019.01.011>.
- [36] Shunping Ji et al. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sensing*, 10(1), 2018. ISSN 2072-4292. <https://doi.org/10.3390/rs10010075>.
- [37] Abhasha Joshi et al. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sensing*, 15(8), 2023. ISSN 2072-4292. <https://doi.org/10.3390/rs15082014>.
- [38] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv e-print*, 2017. <https://doi.org/10.48550/arXiv.1412.6980>.
- [39] Lukas Kondmann et al. DENETHOR: The DynamicEarthNET dataset for Harmonized, inter-Operable, analysis-Ready, daily crop monitoring from space. *NeurIPS Track on Datasets and Benchmarks*, 2021.

- [40] Guang Li et al. Crop type mapping using time-series Sentinel-2 imagery and U-Net in early growth periods in the Hetao irrigation district in China. *Computers and Electronics in Agriculture*, 203:107478, 2022. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2022.107478>.
- [41] Huapeng Li et al. Temporal Sequence Object-based CNN (TS-OCNN) for crop classification from fine resolution remote sensing image time-series. *The Crop Journal*, 10(5): 1507–1516, 2022. ISSN 2214-5141. <https://doi.org/10.1016/j.cj.2022.07.005>.
- [42] Qianjing Li, Jia Tian, and Qingjiu Tian. Deep learning application for crop classification via multi-temporal remote sensing images. *Agriculture*, 13(4), 2023. ISSN 2077-0472. <https://doi.org/10.3390/agriculture13040906>.
- [43] Qingting Li et al. Object-Based Crop Classification with Landsat-MODIS Enhanced Time-Series Data. *Remote Sensing*, 7(12):16091–16107, 2015. ISSN 2072-4292. <https://doi.org/10.3390/rs71215820>.
- [44] Ruiyuan Li et al. Phenology-based classification of crop species and rotation types using fused MODIS and Landsat data: The comparison of a random-forest-based model and a decision-rule-based model. *Soil and Tillage Research*, 206:104838, 2021. ISSN 0167-1987. <https://doi.org/10.1016/j.still.2020.104838>.
- [45] Zhengtao Li, Guokun Chen, and Tianxu Zhang. A CNN-Transformer Hybrid Approach for Crop Classification Using Multitemporal Multisensor Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:847–858, 2020. <https://doi.org/10.1109/JSTARS.2020.2971763>.
- [46] Tingyu Lu, Meixiang Gao, and Lei Wang. Crop classification in high-resolution remote sensing images based on multi-scale feature fusion semantic segmentation model. *Frontiers in Plant Science*, 14, 2023. ISSN 1664-462X. <https://doi.org/10.3389/fpls.2023.1196634>.
- [47] Mouad Alami Machichi et al. Crop mapping using supervised machine learning and deep learning: a systematic literature review. *International Journal of Remote Sensing*, 44(8): 2717–2753, 2023. <https://doi.org/10.1080/01431161.2023.2205984>.
- [48] J. A. Ch. Martinez et al. Fully convolutional recurrent networks for multivariate crop recognition from multitemporal image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171:188–201, 2021. ISSN 0924-2716. <https://doi.org/10.1016/j.isprsjprs.2020.11.007>.
- [49] Nicolas Matton et al. An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series. *Remote Sensing*, 7(10):13208–13232, 2015. ISSN 2072-4292. <https://doi.org/10.3390/rs71013208>.
- [50] Vittorio Mazzia, Aleem Khaliq, and Marcello Chiaberge. Improvement in Land Cover and Crop Classification based on Temporal Features Learning from Sentinel-2 Data Using Recurrent-Convolutional Neural Network (R-CNN). *Applied Sciences*, 10(1), 2020. ISSN 2076-3417. <https://doi.org/10.3390/app10010238>.

- [51] Shiyao Meng et al. Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. *Computers and Electronics in Agriculture*, 186:106188, 2021. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2021.106188>.
- [52] Soe W. Myint et al. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5):1145–1161, 2011. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2010.12.017>.
- [53] Joachim Nyborg, Charlotte Pelletier, and Ira Assent. Generalized classification of satellite image time series with thermal positional encoding. *arXiv e-print*, 2022. <https://doi.org/10.48550/arXiv.2203.09175>.
- [54] Asli Ozdarici-Ok, Ali Ozgun Ok, and Konrad Schindler. Mapping of Agricultural Crops from Single High-Resolution Multispectral Images—Data-Driven Smoothing vs. Parcel-Based Smoothing. *Remote Sensing*, 7(5):5611–5638, 2015. ISSN 2072-4292. <https://doi.org/10.3390/rs70505611>.
- [55] Adam Paszke et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [56] Charlotte Pelletier, Geoffrey I. Webb, and François Petitjean. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing*, 11(5), 2019. ISSN 2072-4292. <https://doi.org/10.3390/rs11050523>.
- [57] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011. ISSN 0031-3203. <https://doi.org/10.1016/j.patcog.2010.09.013>.
- [58] Carole Planque et al. National Crop Mapping Using Sentinel-1 Time Series: A Knowledge-Based Descriptive Algorithm. *Remote Sensing*, 13(5), 2021. ISSN 2072-4292. <https://doi.org/10.3390/rs13050846>.
- [59] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, 2015. <https://doi.org/10.48550/arXiv.1505.04597>.
- [60] Rose Rustowicz et al. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. In *CVPR Workshops*, 2019.
- [61] Marc Rußwurm and Marco Körner. MULTI-TEMPORAL LAND COVER CLASSIFICATION WITH LONG SHORT-TERM MEMORY NEURAL NETWORKS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1/W1:551–558, 05 2017. <https://doi.org/10.5194/isprs-archives-XLII-1-W1-551-2017>.
- [62] Marc Rußwurm and Marco Körner. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1496–1504, 2017. <https://doi.org/10.1109/CVPRW.2017.193>.

- [63] Marc Rußwurm and Marco Körner. Convolutional lstms for cloud-robust segmentation of remote sensing imagery. *arXiv e-print*, 2018. <https://doi.org/10.48550/arXiv.1811.02471>.
- [64] Marc Rußwurm and Marco Körner. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, 7(4), 2018. ISSN 2220-9964. <https://doi.org/10.3390/ijgi7040129>.
- [65] Marc Rußwurm and Marco Körner. Self-attention for raw optical Satellite Time Series Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:421–435, 2020. ISSN 0924-2716. <https://doi.org/10.1016/j.isprsjprs.2020.06.006>.
- [66] Marc Rußwurm et al. BreizhCrops: A Time Series Dataset for Crop Type Mapping. *arXiv e-print*, 2020. <https://doi.org/10.48550/arXiv.1905.11893>.
- [67] Vivien Sainte Fare Garnot and Loic Landrieu. Lightweight Temporal Self-attention for Classifying Satellite Images Time Series. In *Advanced Analytics and Learning on Temporal Data*, pages 171–181, 12 2020. ISBN 978-3-030-65741-3. https://doi.org/10.1007/978-3-030-65742-0_12.
- [68] Vivien Sainte Fare Garnot, Loic Landrieu, and Nesrine Chehata. Multi-modal temporal attention models for crop mapping from satellite time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187:294–305, 2022. ISSN 0924-2716. <https://doi.org/10.1016/j.isprsjprs.2022.03.012>.
- [69] Vivien Sainte Fare Garnot et al. Time-Space Tradeoff in Deep Learning Models for Crop Classification on Satellite Multi-Spectral Image Time Series. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 6247–6250, 07 2019. <https://doi.org/10.1109/IGARSS.2019.8900517>.
- [70] Vivien Sainte Fare Garnot et al. Satellite Image Time Series Classification With Pixel-Set Encoders and Temporal Self-Attention. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12322–12331, 06 2020. <https://doi.org/10.1109/CVPR42600.2020.01234>.
- [71] Teodora Selea. AgriSen-COG, a Multicountry, Multitemporal Large-Scale Sentinel-2 Benchmark Dataset for Crop Mapping Using Deep Learning. *Remote Sensing*, 15(12), 2023. ISSN 2072-4292. <https://doi.org/10.3390/rs15122980>.
- [72] Xingjian Shi et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. <https://doi.org/10.48550/arXiv.1506.04214>.
- [73] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-print*, 2015. <https://doi.org/10.48550/arXiv.1409.1556>.
- [74] Andrei Stoian et al. Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. *Remote Sensing*, 11(17), 2019. ISSN 2072-4292. <https://doi.org/10.3390/rs11171986>.

- [75] Yingwei Sun et al. Geo-parcel based Crops Classification with Sentinel-1 Time Series Data via Recurrent Neural Network. In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pages 1–5, 2019. <https://doi.org/10.1109/Agro-Geoinformatics.2019.8820218>.
- [76] Dimitrios Sykas et al. A Sentinel-2 Multiyear, Multicountry Benchmark Dataset for Crop Classification and Segmentation With Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3323–3339, 2022. <https://doi.org/10.1109/JSTARS.2022.3164771>.
- [77] Christian Szegedy et al. Rethinking the inception architecture for computer vision. *arXiv e-print*, 2015. <https://doi.org/10.48550/arXiv.1512.00567>.
- [78] Michail Tarasiou, Erik Chavez, and Stefanos Zafeiriou. ViTs for SITS: Vision Transformers for Satellite Image Time Series. *arXiv e-print*, 2023. <https://doi.org/10.48550/arXiv.2301.04944>.
- [79] Alireza Taravat et al. Advanced Fully Convolutional Networks for Agricultural Field Boundary Detection. *Remote Sensing*, 13(4), 2021. ISSN 2072-4292. <https://doi.org/10.3390/rs13040722>.
- [80] Kenichi Tatsumi et al. Crop classification of upland fields using random forest of time-series landsat 7 etm+ data. *Computers and Electronics in Agriculture*, 115:171–179, 2015. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2015.05.001>.
- [81] Feng Tian et al. Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe. *Remote Sensing of Environment*, 260:112456, 2021. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2021.112456>.
- [82] Mehmet Ozgur Turkoglu et al. Gating Revisited: Deep Multi-layer RNNs That Can Be Trained. *arXiv e-print*, 2021. <https://doi.org/10.48550/arXiv.1911.11033>.
- [83] Mehmet Ozgur Turkoglu et al. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, 264:112603, 2021. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2021.112603>.
- [84] Ashish Vaswani et al. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. <https://doi.org/10.48550/arXiv.1706.03762>.
- [85] Elliot Vincent, Jean Ponce, and Mathieu Aubry. Pixel-wise Agricultural Image Time Series Classification: Comparisons and a Deformable Prototype-based Approach. *arXiv e-print*, 2023. <https://doi.org/10.48550/arXiv.2303.12533>.
- [86] Francesco Vuolo et al. How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation*, 72:122–130, 2018. ISSN 1569-8432. <https://doi.org/10.1016/j.jag.2018.06.007>.
- [87] Lijun Wang et al. Deep segmentation and classification of complex crops using multi-feature satellite imagery. *Computers and Electronics in Agriculture*, 200:107249, 2022. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2022.107249>.

- [88] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1578–1585, 2017. <https://doi.org/10.1109/IJCNN.2017.7966039>.
- [89] Pengliang Wei et al. Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 174:198–214, 2021. ISSN 0924-2716. <https://doi.org/10.1016/j.isprsjprs.2021.02.011>.
- [90] Giulio Weikmann, Claudia Paris, and Lorenzo Bruzzone. TimeSen2Crop: A Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop-Type Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4699–4708, 2021. <https://doi.org/10.1109/JSTARS.2021.3073965>.
- [91] A. J. W. De Wit and J. G. P. W. Clevers. Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing*, 25(20):4091–4112, 2004. <https://doi.org/10.1080/01431160310001619580>.
- [92] Felix Wu et al. Pay Less Attention with Lightweight and Dynamic Convolutions. *arXiv e-print*, 2019. <https://doi.org/10.48550/arXiv.1901.10430>.
- [93] Jinfan Xu et al. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, 247:111946, 2020. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2020.111946>.
- [94] Chenhongyi Yang et al. GPViT: A High Resolution Non-Hierarchical Vision Transformer with Group Propagation. *arXiv e-print*, 2023. <https://doi.org/10.48550/arXiv.2212.06795>.
- [95] Zhiwei Yi, Li Jia, and Qiting Chen. Crop Classification Using Multi-Temporal Sentinel-2 Data in the Shiyang River Basin of China. *Remote Sensing*, 12(24), 2020. ISSN 2072-4292. <https://doi.org/10.3390/rs12244052>.
- [96] Mingyang Yu et al. AGs-Unet: Building Extraction Model for High Resolution Remote Sensing Images Based on Attention Gates U Network. *Sensors*, 22(8), 2022. ISSN 1424-8220. <https://doi.org/10.3390/s22082932>.
- [97] Y. Yuan and L. Lin. Self-supervised pre-training of transformers for satellite image time series classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pages 1–14, 2020. <https://doi.org/10.1109/JSTARS.2020.3036602>.
- [98] Y. Yuan et al. SITS-Former: A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification. *International Journal of Applied Earth Observation and Geoinformation*, 106:102651, 2022. ISSN 1569-8432. <https://doi.org/10.1016/j.jag.2021.102651>.
- [99] Peng Zhang et al. Parcel-level mapping of crops in a smallholder agricultural area: A case of central China using single-temporal VHSR imagery. *Computers and Electronics in Agriculture*, 175:105581, 2020. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2020.105581>.
- [100] Weixiong Zhang et al. Attention to Both Global and Local Features: A Novel Temporal Encoder for Satellite Image Time Series Classification. *Remote Sensing*, 15(3), 2023. <https://doi.org/10.3390/rs15030618>.

- [101] Hongwei Zhao et al. Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China. *Remote Sensing*, 11(22), 2019. ISSN 2072-4292. <https://doi.org/10.3390/rs11222673>.
- [102] Hongwei Zhao et al. Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information. *Remote Sensing*, 13(14), 2021. ISSN 2072-4292. <https://doi.org/10.3390/rs13142790>.
- [103] Liheng Zhong, Lina Hu, and Hang Zhou. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221:430–443, 2019. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2018.11.032>.

Prílohy

Charakteristiky použitých dát

Kultúra	Skratka
standardní orná půda	R
travní porost (na orné půdě)	G
úhor	U
trvalý travní porost	T
vinice	V
chmelnice	C
ovocný sad	S
školka	K
rychle rostoucí dřeviny	D
jiná trvalá kultura	J
zalesněná půda	L
rybník	B
mimoprodukční plocha	M
jiná kultura	O

Tabuľka 6.1: Prehľad zaznamenávaných kultúr v rámci dielu pôdneho bloku [4].

Plodina anglicky	Plodina slovensky	# parcel	Rozloha (ha)
Permanent grassland	Trvalá trávnatá oblasť	313300	1086468
Winter cereals	Ozimné obilniny	85699	981016
Not classified	Neklasifikované	55125	25685
Summer cereals	Jarné obilniny	32762	282639
Rapeseed	Repka olejná (ozimná)	28926	378681
Annual forage crops	Jednoročné kŕmne plodiny	26393	122299
Maize	Kukurica	26123	311354
Other crops	Ostatné plodiny	15309	137709
Vineyards	Vinice	11358	15206
Annual fruit and vegetables	Jednoročné ovocie a zelenina	10189	67287
Permanent fruit	Trvalé ovocie	5777	16239
Sugar beet	Cukrová repa	3420	59397
Hopyards	Chmelnice	2515	5326
Flax and hemp	L'an a konope	307	2591

Tabuľka 6.2: Počty parcel jednotlivých typov plodín v rámci poskytnutého datasetu.

■	Background
■	Meadow
■	Winter wheat (s.)
■	Corn
■	Winter barley
■	Winter rapeseed
■	Spring barley
■	Sunflower
■	Grapevine
■	Beet
■	Winter triticale
■	Winter wheat (d.)
■	Fruit/vegetable/flower
■	Potatoes
■	Leguminous fodder
■	Soybeans
■	Orchard
■	Mixed cereals
■	Sorghum
■	Void

a) PASTIS

■	Background
■	Grassland
■	Fruit/vegetable
■	Summer cereals
■	Winter cereals
■	Rapeseed
■	Maize
■	Forage crops
■	Sugar beet
■	Flax/Hemp
■	Permanent fruit
■	Hopyards
■	Vineyards
■	Other crops
■	Not classified

b) S2TSCZCrops

Obr. 6.4: Prehl'ad legend pre dataset PASTIS a S2TSCZCrops.