

Metodika vývoje sw. nástrojů pro automatizované rozpoznávání sakrálních staveb

Vypracovali: Jan Fesl, Jiří Jelínek, Michal Konopa, Marie Feslová a Kateřina Horníčková.

ABSTRAKT

Spolehlivá identifikace konkrétních míst, předmětů, lidí, atd. z fotografií je v praxi velmi častá úloha, která není jednoduchá a přináší řadu výzev. V případě velkého množství fotografií, je manuální identifikace objektů na nich zobrazených z časových důvodů téměř nemožná a je nutné ji automatizovat prostřednictvím softwarových systémů. V praxi se ukazuje, že vytvořit jediný univerzální identifikační systém, který dokáže na fotografiích např. lidí, zvířat či budov určit konkrétního člověka, zvíře či budovu, je bez další přidané expertní znalosti velmi složité. Situace je ještě obtížnější v případě, že fotografií, které slouží jako referenční, je jen velmi malé množství. Metody, které lze pro identifikaci objektů použít, mají sice obecnou platnost, ale musí se jejich použití adaptovat pro konkrétní scénář. Náš výzkumný tým vytvořil softwarový systém, který dokáže z fotografií úzce-specifických objektů (sakrálních staveb - kostelů či kaplí), resp. jejich interiérů nebo exteriérů spolehlivě identifikovat konkrétní objekt s velmi dobrou úspěšností. Námi zvolený přístup se jeví jako slibný pro námi vybranou oblast a je snadno přenositelný i na další oblasti.

KLÍČOVÁ SLOVA

Sacral objects, identification, machine learning, Siamese networks, neural networks, expert system.

1. ÚVOD A MOTIVACE

V rámci našeho projektu nazvaného “Informační systém pro středověké památky” jsme řešili výzkumný úkol, který si klade za úkol usnadnit uživatelům identifikaci konkrétních míst vyobrazených na fotografiích na základě shody s obrazovou databází. Zásadním předpokladem, díky kterému není tato úloha jednoduchá, je to, že v praxi k určitým objektům existuje je velmi malé množství obrazových vzorů a proto nelze použít klasické metody zpracování obrazu založených hlubokých neuronových sítí, neboť tyto vyžadují velké množství příkladů. Dalším předpokladem, který úlohu identifikace opět stěžuje, je to, že žádná fotografie v obrazové databázi není zcela stejná, jako analyzovaná.

Inspirací pro řešení tohoto úkolu, je to, že člověk dokáže určitým způsobem z jediné referenční fotografie, kterou srovnává s jinými fotografiemi, někdy určit, zda se jedná o shodné objekty. Obecně vzato, identifikace nikdy neprobíhá na základě celku, ale pouze na vybraných specifických detailech. Tohoto předpokladu jsme využili při návrhu našeho systému.

2. TEORETICKÝ ÚVOD

Vytvoření systému pro identifikaci sakrálních objektů vyžaduje spolupráci znalostí expertů z více oblastí a sice: historie, dějin umění a kultury, architektury (souhrnně umění a historie) a pokročilých metod strojového učení resp. počítačového vidění (souhrnně umělá inteligence - AI). Experti z oblasti umění a historie vědí, které specifické komponenty (jako např. oltáře, presbytáře či křtitelnice) jsou na fotografiích sakrálních objektů jedinečné a mohou tedy přispět k jednoznačné identifikaci konkrétního objektu. Experti z oblasti AI znají metody, které umožňují automatické nalezení specifických objektů na fotografiích a další metody, které dokáží nalezené komponenty porovnat se vzory v obrazové databázi.

2.1 Pohled experta z oblasti dějin umění

Střídání uměleckých stylů charakterizuje vývoj umění. Styl je ucelený umělecký projev epochy charakterizovaný souvislou kombinací specifických forem a rysů typických pro dané období, které se uplatňují v tvorbě současných umělců [4] [5]. Styl neboli sloh představuje rozčleněný a navenek rozpoznatelný soubor vizuálních znaků, v nichž je objekt vytvořen, a tvoří formální jazyk díla, podobně jako grafémy v psaném textu [5]. Ty zkoumáme prostřednictvím formální analýzy. Některé z těchto prvků jsou snadno rozpoznatelné, zatímco jiné je obtížnější identifikovat, protože se skládají ze současného výskytu a vzájemného působení více specifických prvků [6].

V architektuře se standardní prvky projevují jak v dobově konvenčním přístupu k utváření prostoru, tak ve formálních detailech stavby, které ji identifikují. Soubor těchto prvků pomáhá odhalit stavební vývoj dané budovy v průběhu historie. Pozorování a určování odborníka je založeno na podobnosti a identifikaci těchto charakteristických prvků a jejich kombinací [8], které se postupně učí pozorováním během vzdělávání.

Kostely v českém prostředí prošly složitým architektonickým vývojem a mnoha úpravami. Výsledná stavba je často směsicí detailů z různých období; středověké (románské a gotické) tvořící základ zkoumaných staveb v jihočeském regionu, charakteristická sdružená románská okna, protáhlá gotická okna, gotický polygonální závěr, lomený gotický triumfální oblouk nebo sanktuář, sanktusníková věž, hranolová věž atd.

Ve středověku a raném novověku nedocházelo k masové výrobě. Stavební továrny vytvářely prvky jedinečné pro dané místo, odpovídající morfologii tehdejšího slohu, avšak s přispěním invence kameníků [9]. Ačkoli je každý detail stavby individuální, neboť vznikl jako jedinečné dílo, nese obecné rysy v podobě typické pro středoevropskou architekturu a region. Umožňuje nám tedy identifikovat a zařadit prvek k epoše a geografickému regionu nebo k řadě staveb, ale také k této konkrétní stavbě podle jejich jedinečných rysů a uspořádání (Wölfflin, 1915, 149-150) [4].

2.2 Aktuální trendy v oblasti počítačového vidění

Jak už bylo naznačeno, zpracování dat vychází z postupů pro detekci objektů v obraze. Tyto metody lze obecně rozdělit do dvou skupin na metodiky jednokrokové a vícekové. U vícekových metodik jsou samostatně řešeny úlohy identifikace potenciálních oblastí (regionů) vymezujících možné objekty, výběr regionů pro další zpracování a identifikace kategorie objektu v zadaném regionu. Typickými modely s vícekovým zpracováním jsou například sítě typu RPN,

RCNN, FastRCNN, FasterCNN či MaskCNN. Modely z této skupiny pak vynikají vysokou kvalitou prováděné identifikace, ale bohužel také nižší rychlostí zpracování vstupních dat, která omezuje jejich užití v reálném čase.

Do skupiny modelů využívajících jednokrokové zpracování pak řadíme zejména modely typu YOLO. Tyto modely řeší výše uvedené jednotlivé fáze v rámci komplexního zpracování. Tyto modely jsou výrazně rychlejší a lze je využít i pro online detekci objektů. Na druhé straně jsou však charakteristické poněkud nižší kvalitou detekce a některými dalšími problémy (například potížemi s detekcí malých objektů).

Pro první fázi zpracování vstupních obrazových dat byl pro tento projekt zvolen model typu YOLO, konkrétně pak verze YOLOv5, který byl v době přípravy aplikace posledním ve vývojové řadě. V současnosti se můžeme setkat rovněž s označením Yolov6 a Yolov7. Verze 6 používá způsob interface shodný s verzí 5, proto se lze setkat s využitím skriptů modelu verze 5 s váhami označovanými jako váhy verze 6, přičemž hlavním vylepšením je zde zejména vyšší rozlišení vstupního obrázku. U verze 7 panují určité nejasnosti ohledně tohoto názvu (existují dvě konkurenční a nekompatibilní verze) a i struktura a interface modelu se liší. Tato verze však má přinést významná vylepšení, zejména pak možnost přesného vymezení identifikovaného objektu (která však není pro naši aplikaci nutná).

Pod označením Yolov5 se však neskryvá pouze jediný model, ale celá skupina modelů lišících se strukturou, typy a počtem použitých vrstev v rámci modelu.

Modely typ YOLO lze trénovat klasickým způsobem, tedy od začátku s náhodnými váhami. Tento proces je však velmi pomalý a je výhodnější využít techniky tzv. transfer learningu a použít model již s natrénovanými váhami. K dispozici jsou váhy trénované na datasetu COCO. Právě ty byly použity v našem projektu, avšak bylo nutné zvolený typ modelu dotrénovat. Důvodem byl fakt, že náš projekt pracuje se specifickými sakrálními objekty, které jsou v datasetu COCO zastoupeny velmi málo. Zvažována byla i varianta využití jiného standardizovaného datasetu, který by byl k dispozici. Nicméně autoři nenalezli datovou sadu vhodnou pro klasifikaci tohoto, typu objektů. Pro trénování modelů byla tedy použit dataset, který je zmiňován výše v tomto příspěvku.

První fáze zpracování dat o historických památkách byla popsána výše. Jejím základem je identifikace kategorií objektů rozpoznaných na vstupních obrazech. Výstupem tedy je seznam objektů identifikovaných v obraze včetně jejich klasifikace a umístění, umožňujícího vytvořit pro ně příslušné výřezy z původního obrazu. Tyto výřezy a informace o klasifikaci jsou pak vstupy do dalšího zpracování.

Druhá fáze zpracování dat v rámci tohoto projektu navazuje na právě popsanou fázi první, tedy detekci objektu v obraze a jeho kategorizaci. Pro dosažení cílů projektu je nutné nejen zařadit objekt do kategorie, ale pokusit se o jeho konkrétní identifikaci jako instance v rámci kategorie.

V rámci fáze 2 se tedy snažíme v rámci dané kategorie identifikovat, zda dva zadané obrazy zachycují tentýž objekt. Základem činnosti je zkoumání podobnosti dvou vstupních obrazů zachycujících objekty stejného typu. Pro tento úkol existuje celá řada technik zabývajících se analýzou obrazu, dedikovaných i dalších, původně určených pro jiné typy úloh. Tyto techniky původně vycházely z klasických postupů zpracování obrazových dat a nevyužívaly postupy

založené na strojovém učení. Tyto techniky lze rozdělit do dvou základních skupin: area-based a feature-based.

U area-based technik je základem analýza obrazu po jeho jednotlivých částech či pixelech. Podobnost obrazů zde může být založena např. na informacích o barevnosti částí obrazů (porovnávání barevných histogramů). Jen připomeňme, že z první fáze máme k dispozici již pouze výřezy zachycující identifikované objekty a uvedená technika může být i proto úspěšná. Je však značně citlivá na rozdíly v jasnosti obrazů a rozdílné barevné podání.

Feature based techniky se zaměřují na extrakci příznaků z obrazu a jejich následné zpracování. Jednou z metod je identifikace klíčových bodů v obraze ve stupních šedi (jednoznačně identifikovatelných) a následné porovnání nalezených bodů v obou obrazech z hlediska jejich umístění a vzájemné polohy. Technika je označována jako scale-invariant feature transformation (SIFT) a je schopna ve výstupu nalézt příslušné zobrazení popisující, jak jeden obraz transformovat, abychom získali druhý. Primární určení je pro aplikace vyžadující právě uvedenou transformaci, tedy např. panoramatické spojování obrazů. Technika je schopna rozpoznat podobnost i mezi obrazy výrazně se odlišujícími jasně, polohou a záběrem pozorovatele, otočením, zoomem atd. Její nasazení je náročnější na hardwarové prostředky.

Z metody SIFT vychází další z technik označovaná jako speeded up robust features (SURF), obsahující stejné globální kroky, ale lišící se od SIFT v jejich konkrétní realizaci. Použití SURF je méně náročné na zdroje a výpočet je tedy rychlejší.

Výše zmíněné techniky nevyužívají principů strojového učení. To však do oblasti zpracování grafiky přineslo zcela nové podněty a techniky. Důraz je zde kladen na učení s učitelem umožňující lépe customizovat model pro konkrétní úlohu. Nutné je však mít k dispozici dostatečně velkou a anotovanou trénovací množinu.

Jako vhodná architektura se pro porovnávání detekovaných objektů s referenční databází se jeví siamské sítě, které mají na vstupu dva porovnávané obrázky a na výstupu jsou schopny poskytnout míru jejich vzájemné shody. Tyto sítě patří do feature-based kategorie nástrojů. I v tomto případě je nutné si uvědomit, že pracujeme s velmi specifickou skupinou objektů na obrazech, tedy objekty sakrálními. Proto je opět nutné předtrénovaný model dotrénovat na specifických datech vzniklých v rámci datasetu popsaného výše.

Základní struktura siamských sítí je na následující. Síť má 2 vstupy (obrazy se stejným rozlišením a barevnou hloubkou), přičemž tyto vstupy jsou nejprve zpracovány prvním stupněm modelu založeným na konvolučních vrstvách, které mají nejlepší výsledky při zpracování grafické informace a zachycení základních vzorů (příznaků), které se v obrazech vyskytují. Tyto vrstvy jsou doplněny o nastavbu několika dense vrstev a tato celá struktura formuje podmodel označovaný dále jako *tower*. Tower je se stejnými vahami použit pro zpracování obou vstupních obrazů, jeho váhy jsou tedy sdílené. Podmodel tower je možno opět vytvořit od počátku a trénovat jej pouze na námi dodaných vstupních datech. Využít však lze i již známé postupy a struktury zejména pro prvotní zpracování obrazových dat pomocí CNN vrstev. V úvahu zde přicházejí zejména předtrénované modely VGG-16, Inception (v3), ResNet50 nebo EfficientNet. S výhodou je tak možné využít transfer learning neboli převzetí nastavení těchto modelů vytvořeného na některém ze známých existujících datasetů. V prezentovaném řešení byl použit model VGG16 (bez nastavbových vrstev) s nastavením vah vytvořeným na datasetu ImageNet. Na tuto strukturu

následně navazuje několikavrstvá struktura složená z plně propojených vrstev. Na výstupu zde dochází k významné dimenzionální redukci oproti vstupu, který např. pro VGG16 má dimenzi 224 x 224 x 3.

Výstupem stupně tower jsou dva vektory o stejné dimenzi, pro každý obrázek jeden. Tyto vektory tvoří vstup do druhého stupně modelu, který slouží k porovnání obrázků a stanovení míry jejich podobnosti. Podmodel je tvořen pouze několika vrstvami. Role první z nich je zásadní, neboť počítá euklidovskou vzdálenost mezi předloženými vektory. Následující vrstvy pak tuto hodnotu pouze normalizují do rozsahu (0, 1) a převádějí do jediného výstupu udávajícího míru podobnosti předložených obrazů. Nastavení ztrátové funkce celého modelu je rovněž jedním z klíčových faktorů úspěšnosti modelu. V předloženém modelu využíváme ztrátovou funkci převzatou z (https://keras.io/examples/vision/siamese_contrastive/), která se využívá pro zdůraznění rozdílů mezi obrázky.

3. METHODOLOGIE

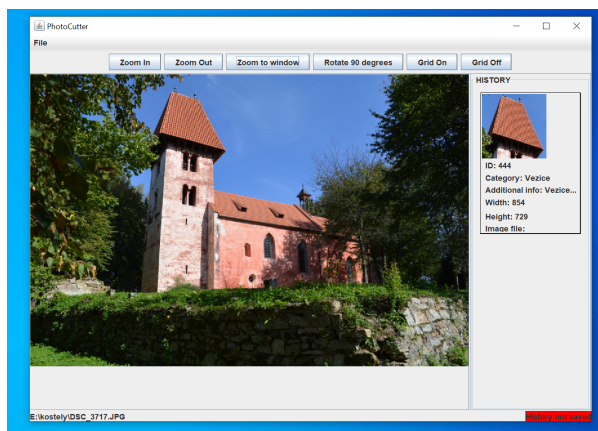
Pro trénování Yolo či Siamských neurálních sítí je nezbytné vlastnit kvalitní datovou sadu, kterou jsme ovšem v tomto případě museli zhotovit od základu. Na základě doporučení expertů jsme vybrali několik základních komponent, které se běžně nachází v sakrálních stavbách a mohou tedy sloužit k identifikaci konkrétních objektů. Dále jsme navrhli dvou-stupňové schéma systému pro identifikaci objektů, které jsme implementovali v našem systému.

3.1 Vytvoření datové sady

Jelikož jsme pro naše účely 6000 fotografií reálných sakrálních staveb, které jsme manuálně zkontrolovali a identifikovali příslušné komponenty. Celková velikost našeho datasetu byla cca. 94 GB. Abychom byli schopni vytvořit danou klasifikaci komponent, bylo nutné vytvořit pro tento účel speciální aplikaci, kterou jsme nazvali PhotoCutter. Stručný popis aplikace je uveden níže.

3.1.1 Aplikace PhotoCutter

Aplikace PhotoCutter slouží k vytváření výřezů hledaných objektů z fotografií a jejich následnému ukládání do souborů. PhotoCutter je Java aplikace s grafickým uživatelským rozhraním založeným na knihovně Java Swing.



Obr. 1: PhotoCutter - hlavní okno

Na obrázku uvedeném níže je možné vidět jakým způsobem probíhala klasifikace jednotlivých fotografií sakrálních objektů, resp. vyhledání jednotlivých komponent uvedených in Table 1.

3.2 Struktura datové sady

Do testovací sady jsme na základě našich odborných zkušeností vybrali některé zásadní detaily a stavební prvky, např. sochu, fresku, oltář, kazatelnu, věž, okno, věžičku/zvonici atd. Zásadní je také jejich kombinace - teprve jejich spojení tvoří jedinečnou identifikaci dané stavby a základ pro časové, slohové a formální zařazení a hodnocení stavby i pro vlastní identifikaci stavby. Nástěnné malby jsou díky své formě snadno identifikovatelné, proto byly mezi prvky zařazeny i současné kostelní malby. Kostel lze identifikovat podle stylové úrovně a ikonografie maleb v interiéru.

Kompletní seznam kategorií včetně jejich příkladové vizualizace je uveden v tabulce 1. Grafy v pravém sloupci představují histogram rozložení obrazů seřazený podle plochy obrazu vyjádřeně v kilopixelech (Kpix).

4. VÝSLEDKY A MĚŘENÍ

4.1 Detekce objektů

Data použitá pro trénování modelu obsahovala celkem 1097 obrázků s 5164 vyznačenými objekty. Testovací data se pak skládala z 239 obrázků s 1240 objekty. Testovací data nebyla nijak využita v procesu učení jednotlivých částí modelu.

První fáze zpracování dat v modelu je zastoupena sítí YOLOv5. Jak je uvedeno výše, sítí tohoto typu existuje více, testovány byly následující varianty:

Model	Resolution
-------	------------

yolov5m	640
yolov5x	640
yolov5s6	1280
yolov5m6	1280
yolov5n	640
yolov5n	640

Modely byly trénovány po dobu 400 epoch na trénovací množině. Následně byly nastavené modely evaluovány pomocí testovací množiny. Hodnoty klíčových parametrů charakterizujících trénování i testování jsou shrnuty v následující tabulce:

	Train	Test
Model	P R mAP@.5 mAP@	P R mAP@.5 mAP@
yolov5m	0.983 0.857 0.917 0.856	0.953 0.882 0.899 0.814
yolov5x	0.986 0.91 0.934 0.907	0.974 0.878 0.896 0.838
yolov5s6	0.978 0.865 0.926 0.869	0.949 0.877 0.9 0.812
yolov5m6	0.985 0.912 0.934 0.908	0.976 0.888 0.898 0.836
yolov5n	0.981 0.844 0.949 0.762	0.953 0.852 0.891 0.685
yolov5n	0.956 0.822 0.872 0.68	0.923 0.845 0.881 0.67

Pro další zpracování obrázků byl vybrán model „yolov5x“, a to na základě dosažené hodnoty mAP@ při evaluaci na testovací množině. Model se skládal z 444 vrstev a měl celkem 86328181 parametrů.

V následující tabulce jsou uvedeny klíčové údaje z trénování vybraného modelu podle jednotlivých kategorií objektů:

Class	Images	Labels	P	R	mAP@.5	mAP@
all	1097	5164	0.986	0.91	0.934	0.907
Detail ostění/přípory/konzoly	1097	389	1	0.992	0.995	0.945

Freska	1097	1352	0.999	0.996	0.995	0.978
Kaple	1097	36	0.993	1	0.995	0.986
Kazatelna	1097	41	0.994	1	0.995	0.985
Klenba	1097	190	0.999	1	0.995	0.989
Kostel	1097	241	1	0.994	0.995	0.993
Kruchta/Empora	1097	3	0.807	1	0.995	0.995
Křtitelnice	1097	19	0.978	1	0.995	0.964
Obraz	1097	120	0.995	1	0.995	0.97
Okno	1097	1393	0.999	0.991	0.995	0.95
Oltář	1097	129	0.998	1	0.995	0.987
Pilíř/Sloup	1097	339	0.995	0.997	0.995	0.961
Portál	1097	143	0.998	1	0.995	0.985
Průčelí	1097	1	0.965	1	0.995	0.995
Římsa	1097	4	1	0	0.25	0.15
Sanktuář/Pastoforium	1097	25	1	0.976	0.995	0.931
Socha	1097	141	1	1	0.995	0.955
Svorník	1097	210	1	0.999	0.995	0.915
Varhany	1097	19	0.987	1	0.995	0.98
Věžice/Zvonice/Sanktusník na střeše	1097	109	0.992	1	0.995	0.957
Vítězný oblouk	1097	7	0.981	1	0.995	0.975
Věž	1097	252	0.999	0.992	0.995	0.986
Závěr kostela/Presbytář	1097	1	1	0	0.332	0.332

K dispozici je i výstup testování pro jednotlivé kategorie objektů:

Class	Images	Labels	P	R	mAP@.5	mAP@
all	239	1240	0.974	0.878	0.896	0.838
Detail ostění/přípory/konzoly	239	109	0.96	0.817	0.905	0.807
Freska	239	294	0.989	0.932	0.96	0.9
Kaple	239	3	0.983	1	0.995	0.897
Kazatelna	239	12	0.995	1	0.995	0.964
Klenba	239	39	1	0.815	0.895	0.813
Kostel	239	61	1	0.938	0.985	0.949
Kruchta/Empora	239	2	0.838	0.5	0.507	0.507
Křtitelnice	239	3	0.98	1	0.995	0.995
Obraz	239	21	0.898	0.857	0.851	0.826
Okno	239	313	0.993	0.915	0.983	0.888
Oltář	239	40	1	0.954	0.982	0.926
Pilíř/Sloup	239	109	0.97	0.898	0.921	0.865
Portál	239	32	0.997	1	0.995	0.949
Římsa	239	1	1	0	0	0
Sanktuář/Pastoforium	239	4	0.983	1	0.995	0.971
Socha	239	37	0.998	0.946	0.951	0.904
Svorník	239	67	0.956	0.967	0.943	0.772
Varhany	239	2	0.962	1	0.995	0.995

Věžice/Zvonice/Sankt usník na střeše	239	22	1	0.959	0.995	0.939
Vítězný oblouk	239	2	0.967	1	0.995	0.798
Věž	239	67	0.983	0.94	0.973	0.94

V testovacích datech není zastoupena poslední kategorie z trénovací sady, proto není uvedena.

Rozdílná kvalita detekce objektů je ovlivněna jak počty příkladů v jednotlivých kategoriích, tak předchozím natrénování sítě YOLO, která byla na našich datech pouze dotrénována.

4.2 Identifikace objektů dle databáze vzorů

Ve druhé fázi byl použit model siamské sítě s konvoluční podčástí VGG 16 předtrénovanou na datasetu imagenet a pro modifikaci vah uzamčenou. Nadstavba byla tvořena dense vrstvou a dále několika vrstvami, jejichž úkolem bylo spojení výstupů z podmodelů „tower“ pro oba vstupní obrazy. Po prvních testech byla doplněna ještě druhá varianta struktury modelu, kde byla odstraněna poslední vrstva VGG16, která významně redukovala počet nastavitelných parametrů (modely označené jako „strip“). Počet těchto parametrů pak vzrostl z 69 796 na 3 215 524. Testovány byly také dvě varianty hodnotící loss funkce – mse a contrast loss (viz výše).

Vstupní trénovací data byla tvořena výřezy objektů z trénovací množiny doplněnými o rozšířené obrazy (modifikace polohy, rotace, jasu, míry detailu, ...) pro získání více příkladů shody obrazů a tím i pozitivních příkladů. Testovací data byla vytvořena obdobným postupem, avšak z původní testovací množiny.

Výsledky z trénovací i testovací fáze modelu jsou uvedeny v následující tabulce (uvedeny jsou pouze nejlepší výsledky pro danou strukturu modelu):

			Training		Testing	
Model	Batch size	Epochs training	Loss	Accuracy	Loss	Accuracy
strip_mse	20	600	0,0007	0,9994	0,0425	0,9481
strip_closs	20	600	0,0007	0,9994	0,0400	0,9484
mse	40	200	0,0041	0,9957	0,0561	0,9365
closs	20	400	0,0073	0,9933	0,0765	0,9104

Uvedené 4 varianty nastavení modelu byly následně použity i při zkoumání celkové schopnosti celé aplikace detekovat lokalitu dle zadaného obrazu.

4.3 Detekce specifických objektů

Při testování celkových schopností aplikace byly výřezy objektů na obrázcích trénovací množiny zpracovány podčástí „tower“ siamské sítě a jejich výsledný otisk uložen do databáze. Při předložení neznámého obrazu pak byly otisky z databáze porovnávány siamskou sítí s výřezy z obrazu získanými sítí YOLOv5. Neznámé obrazy byly brány z původní testovací množiny.

Pro každý předložený neznámý výřez tak byla získán výřez ze stejné kategorie s největší mírou podobnosti a míra této podobnosti. Sada těchto výřezů z trénovací množiny pak byla finálně zpracovaná tak, aby byla určena lokace neznámého obrazu. Pro toto zpracování byly testovány 4 různé metodiky:

- Volba lokace podle maximálního součtu měr podobnosti podle lokací (max_sum_qual)
- Volba lokace podle první absolutně nejvyšší míry podobnosti podle lokací (max_abs_qual_first)
- Volba lokace podle maximálního počtu podobností podle lokací (max_sum_cnt)
- Volba lokace podle maximálního součtu měr podobnosti podle lokací s respektováním vah kategorií objektů podle jejich četnosti (max_sum_qual_wtd)
- Volba lokace podle maximálního součtu měr podobnosti podle lokací s respektováním vah kategorií objektů podle inverze jejich četnosti (max_sum_qual_wtd_inv)

Získané výsledky celkové kvality (přesnosti) určení lokace pro jednotlivá nastavení modelu a různou finální metodiku jsou v následující tabulce:

Model	max_sum_qual	max_abs_qual_first	max_sum_cnt	max_sum_qual_wtd	max_sum_qual_wtd_inv
mse	0,4084	0,2939	0,4008	0,3779	0,3588
closs	0,5038	0,4160	0,5191	0,4809	0,4351
strip-mse	0,6031	0,5458	0,5916	0,5611	0,4809
strip-closs	0,6298	0,5725	0,5916	0,6107	0,5153

Ze získaných výsledků je zřejmé, že existuje určitý nesoulad mezi výsledky testů na samotném modelu siamské sítě a finálními výsledky celé aplikace. Důvodem v tomto případě může být nevhodná skladba validační množiny při testech aplikace. Použití originální testovací množiny mohlo způsobit, že byly testovány i obrázky z lokalit, které v originální trénovací množině nebyly zastoupeny. Vhodnější by pak bylo použít množinu obrázků z lokalit zastoupených v trénovací množině, ale pořízených zcela odděleně od této množiny.

Z tabulky je však zřejmé, že nejlepších výsledků bylo dosaženo na modelu „strip_closs“ a s použitím finální klasifikace podle maximálního součtu měr podobnosti podle lokací (max_sum_qual).



Cíl EÚS
Česká republika –
Svobodný stát Bavorsko
2014–2020



Přirodovědecká
fakulta
Faculty
of Science
Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice



Evropská unie
Evropský fond
pro regionální rozvoj

SEZNAM LITERATURY

- [1] L. Jose, Architectural Heritage Elements image Dataset, available online, <https://old.datahub.io/dataset/architectural-heritage-elements-image-dataset>, 2017.
- [2] J. Llamas, P.M. Lerones, R. Medina, E. Zalama, J. Gómez-García-Bermejo, Classification of architectural heritage images using deep learning techniques Appl. Sci., 7 (10) (2017), p. 992, 10.3390/app7100992
- View in ScopusGoogle ScholarL. Jose, Architectural Heritage Elements image Dataset, available online, <https://old.datahub.io/dataset/architectural-heritage-elements-image-dataset>, 2017.
- [3] J. Fesl, J. Jelínek, K. Horníčková, Z. Nevařilová, M. Konopa, M. Feslová, AI-based system for cultural heritage objects identification from real photos, 12th International Conference on Advanced Computer Information Technologies (2022), pp. 476-479, 10.1109/ACIT54803.2022.9912752, 2022.
- [4] H. Wölfflin Principles of art history, Eric Fernie, Art History and Its Methods. A Critical Anthology, Phaidon Press, New York - London (1915), pp. 137-140 (reprint) 2008.
- [5] H. Focillon. The Life of Forms in Art (1934), pp. 171-173.
- [6] E.H. Gombrich, StyleD. Preziosi (Ed.), The Art of Art History, A Critical Anthology, Oxford University Press, Oxford – New York (1998), pp. 150-163.
- [7] G. Morelli, Italian painters Eric Fernie, Art History and Its Methods. A Critical Anthology, Phaidon Press, New York - London (1890), p. 108, (reprint) 2008.
- [9] P. Chotěbor Stavební postupy svatovítské stavební huti ve středověku zjištěné při konzervaci katedrály sv. Víta na Pražském hradě, Staletá Praha, 22 (2016), pp. 122-129.

Metodika byla vytvořena v rámci projektu "Informační systém pro středověké památky v česko-bavorském příhraničí, č. 335, který je financován z Programu přeshraniční spolupráce Česká republika - Svobodný stát Bavorsko Cíl EÚS 2014 - 2020.