

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Gašper Spagnolo

Lokalizacija brezpilotnih letalnikov

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM
PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Luka Čehovin Zajc
SOMENTOR: asist. Matej Dobrevski

Ljubljana, 2023

To delo je ponujeno pod licenco *Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 2.5 Slovenija* (ali novejšo različico). To pomeni, da se tako besedilo, slike, grafi in druge sestavine dela kot tudi rezultati diplomskega dela lahko prosto distribuira, reproducirajo, uporabljajo, priobčujejo javnosti in predelujejo, pod pogojem, da se jasno in vidno navede avtorja in naslov tega dela in da se v primeru spremembe, preoblikovanja ali uporabe tega dela v svojem delu, lahko distribuira predelava le pod licenco, ki je enaka tej. Podrobnosti licence so dostopne na spletni strani creativecommons.si ali na Inštitutu za intelektualno lastnino, Streliška 1, 1000 Ljubljana.



Izvorna koda diplomskega dela, njeni rezultati in v ta namen razvita programska oprema je ponujena pod licenco GNU General Public License, različica 3 (ali novejša). To pomeni, da se lahko prosto distribuira in/ali predeluje pod njenimi pogoji. Podrobnosti licence so dostopne na spletni strani <http://www.gnu.org/licenses/>.

Besedilo je oblikovano z urejevalnikom besedil L^AT_EX.

Kandidat: Gašper Spagnolo

Naslov: Lokalizacija brezpilotnih letalnikov

Vrsta naloge: Diplomaska naloga na univerzitetnem programu prve stopnje
Računalništvo in informatika

Mentor: doc. dr. Luka Cehovin Zajc

Somentor: asist. Matej Dobrevski

Opis:

V zadnjem času postaja uporaba brezpilotnih letalnikov vse bolj razširjena in se uporablja v različnih področjih, kot so agrikultura, kartiranje, vojaške operacije in še mnogo drugih. Kljub njihovi vsestranskosti pa se poraja ključno vprašanje: kako se brezpilotni letalniki obnašajo, ko izgubijo stik z GPS sistemom? Diplomaska naloga se osredotoča na to tematiko in predlaga metodo za lokalizacijo brezpilotnih letalnikov ob izgubi GPS signala.

Title: UAV localization

Description:

In recent times, the use of unmanned aerial vehicles (UAVs) has become increasingly prevalent, finding applications in various fields such as agriculture, mapping, military operations, and many others. Despite their versatility, a critical question arises: how do drones behave when they lose connection to the GPS system? This thesis focuses on this issue and proposes a method for localizing UAVs in the event of a GPS signal loss.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Metodologija	5
2.1	Konvolucijske nevronske mreže	6
2.2	Transformerska arhitektura	7
2.3	Zgradba transformerja	8
2.4	Vision Transformer (ViT)	13
2.5	Piramidni ViT (PVT)	14
2.6	Piramidni ViT z uporabo lokalnih značilnosti (PCPVT)	16
2.7	Siamska nevronska mreža za primerjavo vzorcev	17
3	Podatkovna množica	21
3.1	Slike brezpilotnega letalnika	22
3.2	Satelitske slike	26
3.3	Oznake	27
4	Rezultati	33
4.1	Implementacija	34
4.2	Učenje modela	40
4.3	Izbira kriterijske funkcije	41
4.4	Učenje s stratificiranim vzorčenjem	48

4.5	Vpliv velikosti Hanningovega okna	49
4.6	Regularizacija	50
4.7	Uporaba prednaučene mreže	52
5	Sklepne ugotovitve	55
A	Primeri izračuna RDS	59
B	Primerjava toplotnih kart	61
	Literatura	63

Seznam uporabljenih kratic

kratica	angleško	slovensko
UAV	unmanned aerial vehicle	brezpilotni letalnik
SAT	satellite	satelit
CNN	convolutional neural network	konvolucijska nevrnska mreza
RDS	relative distance score	ocena relativne razdalje
MSE	mean squared error	srednja kvadratna napaka
FPI	finding point in an image	iskanje tocke v sliki
WAMF	Weight-Adaptive Multi Feature fusion	Uteženo združevanje več značilk
PCPVT	Pyramid Vision Transformer with Conditional Positional encodings	Piramidni vision transformer s pogojnimi pozicijskimi kodiranj
HANN	Hanning loss function	Hanningova kriterjska funkcija
GWMSE	Gaussian Weighted Mean Squared Error	Gaussovo utežena srednja kvadratna napaka
HWMSE	Hanning Weighted Mean Squared Error	Hanningovo utežena srednja kvadratna napaka
CWMSE	Cross-Weighted Mean Squared Error	Križno utežena srednja kvadratna napaka

Povzetek

Naslov: Lokalizacija brezpilotnih letalnikov

Avtor: Gašper Spagnolo

Diplomsko delo se osredotoča na uporabo naprednih nevronske mreže za lokalizacijo brezpilotnih letalnikov s pomočjo satelitskih slik. V uvodu je predstavljena osnovna terminologija in koncepti s področja nevronske mreže. V metodološkem delu so podrobno razložene konvolucijske nevronske mreže, transformerska arhitektura ter njeni derivati, kot sta Vision Transformer (ViT) in Piramidni vision transformer (PVT). Posebno pozornost je namenjena Siamski nevronske mreži, ki je ključna za primerjavo vzorcev med satelitskimi in slikami iz brezpilotnega letalnika. Podatkovna množica, uporabljena za učenje in testiranje, vključuje slike brezpilotnega letalnika in satelitske slike. V razdelku rezultatov so predstavljene ključne faze implementacije, učenja modela, izbire kriterijske funkcije ter različne optimizacijske strategije, kot je uporaba stratificiranega vzorčenja. Poudarjena je tudi vloga Hanningovega okna in regularizacijskih tehnik, kot je izpuščanje nevronov. Delo zaključuje s sklepnimi ugotovitvami, ki poudarjajo potencial in učinkovitost predlagane metode za natančno lokalizacijo brezpilotnih letalnikov.

Ključne besede: Lokalizacija UAV, geo-lokalizacija, globoko učenje, transformer.

Abstract

Title: UAV localization

Author: Gašper Spagnolo

The thesis focuses on the use of advanced neural networks for the localization of drones using satellite imagery. In the introduction, foundational terminology and concepts from the realm of neural networks are presented. The methodology section provides a detailed exploration of convolutional neural networks, transformer architecture, and its derivatives like the Vision Transformer (ViT) and the Pyramid Vision Transformer (PVT). Particular emphasis is given to the Siamese Neural Network, which is pivotal for comparing patterns between satellite and drone images. The dataset used for training and testing encompasses drone images and satellite photos. In the results section, key stages of implementation, model training, criterion function selection, and various optimization strategies, such as Stratified Sampling, are discussed. The role of the Hanning window and regularization techniques like dropout is also highlighted. The work concludes with final observations that underscore the potential and efficiency of the proposed method for precise drone localization.

Keywords: UAV localization, geo-localization, deep learning, transformer.

Poglavje 1

Uvod

Brezpilotni letalniki so postali nepogrešljivo orodje v številnih sektorjih, od vojaških operacij do kmetijskega nadzora. Kljub njihovi široki uporabi pa se soočajo z izzivi pri avtonomni navigaciji, še posebej v okoljih, kjer je letalo omejen ali nezanesljiv. V idealnih razmerah brezpilotni letalniki za svojo navigacijo uporabljajo GPS signale, vendar pa lahko te signale motijo naravne in človeške ovire, kot so visoke stavbe, gorske formacije ali celo elektronske motnje. Izguba GPS signala lahko postane kritična, še posebej v tistih trenutkih, ko je natančna lokacija letalnika ključna za njegovo nalogo, zato je iskanje alternativne metode za lokalizacijo brezpilotnih letalnikov nujno.

Zgodnje metode, kot so navedene v virih [8], [20], [19] in [13], so se osredotočale predvsem na uporabo ročno izdelanih značilnosti. To pomeni, da so raziskovalci uporabljali specifične, predhodno definirane vzorce iz slik za določanje lokacije. Čeprav so te metode predstavljale pomemben začetek, so bile omejene v svoji natančnosti in prilagodljivosti.

S prihodom globokih konvolucijskih nevronske mreže (CNN) in njihove dokazane sposobnosti v obdelavi vizualnih podatkov so raziskovalci začeli avtomatsko pridobivati kompleksne in prilagodljive značilnosti neposredno iz podatkov med učenjem mreže. Raziskave v [27] so bile med prvimi, ki so se lotile tega področja z izvlečkom značilnosti za izziv geolokalizacije s pomočjo različnih pogledov, uporabljajoč vnaprej naučen CNN. Ugotovljeno je bilo,

da visokonivojske plasti v CNN vsebujejo bogate semantične informacije, ki lahko pripomorejo k boljši geolokalizaciji. Nadaljnje raziskave v [18] so razširile ta koncept z natančnim prilagajanjem predhodno naučenih mrež, da bi zmanjšali razdaljo značilnosti med satelitskimi slikami in slikami iz brezpilotnega letalnika.

V [21] je bil predstavljen pristop z uporabo modificirane siamske mreže. Ta pristop uporablja kontrastno izgubo za optimizacijo parametrov mreže, kar omogoča boljše razlikovanje med podobnimi in različnimi lokacijami. V [17] so bile predstavljene metode, ki so optimizirale opise slik, da so postale odporne na masivne spremembe perspektive, kot je pogled iz zraka proti tlem ali obratno. V [28] so predstavljene inovacije, ki uporabljajo prostorske informacije za izboljšanje globalnega koraka agregacije pri izvlečku značilnosti. Z uporabo mehanizma prostorske pozornosti so še dodatno izboljšali natančnost geolokalizacije.

Tradicionalne metode prepoznavanja slik se v kontekstu lokalizacije brezpilotnih letalnikov zdijo kot obetavna alternativa [2] in [30], vendar pa se ob njihovi uporabi pojavi cela paleta izzivov. Prvič, potrebujemo ogromno slikovno bazo, ki vključuje kompresirane satelitske slike območij nad katerimi letalnik leti. Velikost in obseg te baze lahko povzročita precejšnje računske in pomnilniške zahteve, kar lahko oteži njeno integracijo v realnočasovnih sistemih, kot so brezpilotni letalniki. Drugič, vsaka posodobitev ali sprememba v osnovni nevronske mreži, ki se uporablja za prepoznavanje slik zahteva ponovno obdelavo celotne slikovne baze. To ne le da je časovno potratno, ampak tudi zviša stroške, saj morajo vse slike ponovno potekati skozi postopek predprocesiranja in razpoznavanja. Tretjič, ko brezpilotni letalnik zajame sliko za primerjavo, mora ta slika biti primerjana z vsako sliko v bazi, da se ugotovi najboljše ujemanje. V praksi to pomeni da, ko imamo bazo sestavljeno iz milijonov slik, bo vsaka nova poizvedovalna slika potrebovala milijone primerjav, kar je zelo časovno potratno in računsko intenzivno.

V luči omejitev tradicionalnih metod prepoznavanja slik so raziskovalci razvili pristop, imenovan FPI (Finding Point with Image) [5]. FPI sprejme

dva vhodna podatka: sliko posneto z brezpilotnim letalnikom in pripadajočo satelitsko sliko. V kontekstu te satelitske slike je mesto, kjer je bila slika iz brezpilotnega letalnika posneta. Za obdelavo vsake slike se uporablja posebna nevronska mreža, kjer vsaka mreža obdeluje svoj nabor podatkov brez deljenja uteži z drugo. Ko sta sliki obdelani in njihove značilke izluščene, se med njima izvede operacija korelacije. Ta mera podobnosti se predstavi v obliki toplotne karte, ki prikazuje stopnjo ujemanja med sliko brezpilotnega letalnika in satelitsko sliko. Najvišja vrednost na toplotni karti natančno označuje mesto, kjer je brezpilotni letalnik posnel svojo sliko na večji satelitski sliki. Informacija se nato neposredno prevede v natančno lokalizacijo brezpilotnega letalnika na satelitski sliki.

Inovacije v znanstvenem raziskovanju pogosto vodijo do nadaljnjih metodoloških izboljšav. Nadgradnja metode FPI, znana kot WAMF-FPI, je dodatno izboljšala natančnost in učinkovitost lokalizacije brezpilotnih letalnikov [5]. Ta pristop je integriral koncepte iz območja sledenja objektov za potrebe lokalizacije ob soočanju z izzivi, ki jih predstavljajo razlike med slikami zajetimi z brezpilotnim letalnikom in satelitskimi slikami. Z uporabo dveh različnih uteži za izveček značilnosti iz slik posnetih z brezpilotnim letalnikom in satelitskih slik, WAMF-FPI omogoča natančnejše in bolj zanesljivo ujemanje slik. Dodatna optimizacija je bila dosežena z vključitvijo WAMF modula in uporabo Hanningove kriterijske funkcije, ki sta povečala učinkovitost modela.

WAMF-FPI je evolucija osnovne metode FPI. Ključna prednost WAMF-FPI je njegova napredna piramidna struktura izluščenja značilk, ki omogoča bolj natančno in raznoliko analizo vhodnih podatkov. Z uporabo te piramidne strukture se značilke izluščijo na več različnih ravneh, nato pa se skalirajo in medsebojno primerjajo, kar pridobi bolj robusten in natančen sklop informacij. Poleg tega WAMF-FPI optimizira kompresijske zmogljivosti, kar pripomore k hitrejšemu in učinkovitejšemu procesiranju podatkov. Medtem ko je v osnovni FPI metodi končna velikost značilk bila stisnjena na 16-krat manjšo od izvirne satelitske slike, v WAMF-FPI ta kompresijski faktor znaša

samo 4-krat manjšo velikost. To omogoča WAMF-FPI-ju, da ohrani več informacij ter pridobi boljšo lokalizacijsko natančnost ob hkratnem zmanjšanju računske obremenitve.

Kljub številnim obstoječim zbirkam, kot so CVUSA [11], CVACT [10] in University-1652 [29], večina ne zajema vseh realnih situacij s katerimi se srečuje brezpilotni letalnik. Na primer, zbirka CVUSA [11] je osredotočena predvsem na zgradbe. Zbirka University-1652 [29] uporablja posnetke univerz, vendar nima dovolj raznolikih posnetkov, saj je omejena le na univerzitetna okolja. V našem delu smo se soočili z odsotnostjo specifične javno dostopne podatkovne zbirke UL14, katero so uporabili avtorji uporabili v [5]. Zaradi te pomanjkljivosti smo se odločili za ustvarjanje lastne zbirke s pomočjo Google Earth Studio [9]. Naša zbirka obsega 11 evropskih mest. Glavni cilj izdelave te zbirke je bil zagotoviti raznolike podatke, ki bi služili kot robustna osnova za testiranje in validacijo pristopov. S tem smo želeli zagotoviti, da naša implementacija lahko obravnava različne scenarije, ki jih lahko sreča brezpilotni letalnik v realnem svetu. Cilj izdelave zbirke je bil zagotoviti raznolike podatke, ki bi lahko služili kot robustna osnova za testiranje in validacijo naše implementacije WAMF-FPI.

V tej diplomski nalogi smo se odločili za implementacijo WAMF-FPI, kot je predstavljeno v izvornem članku [25], saj metoda velja za najnaprednejšo na področju. Implementirali smo vse, kakor je v članku opisano, z namenom dobiti objektivno sliko o učinkovitosti in natančnosti metode.

Diplomska naloga je razdeljena na pet osnovnih poglavij. V poglavju 1 so predstavljena temeljna izhodišča in namen raziskave. Poglavje 2 obsega podroben pregled uporabljenih tehnik, vključno s konvolucijskimi nevronskimi mrežami in različnimi oblikami Vision Transformerja. Poglavje 3 obravnava izbrane podatkovne vire, predvsem slike brezpilotnih letalnikov in satelitske slike. V poglavju 4 so predstavljeni rezultati implementacije, optimizacija in pristopi učenja modela. V poglavju 5 so povzete ključne ugotovitve naloge, delo pa se zaključuje s seznamom relevantne literature.

Poglavje 2

Metodologija

V tem poglavju bomo predstavili osnovne komponente, ki jih uporabljamo v našem modelu. Začeli bomo s konvolucijskimi nevronskimi mrežami, ki so temeljni gradnik večine modelov za obdelavo slik in nudijo močno orodje za izluščenje značilnosti iz vizualnih podatkov. Nadaljevali bomo s predstavitvijo transformerske arhitekture, ki je revolucionirala področje obdelave naravnega jezika in se v zadnjem času vedno bolj uporablja tudi v računalniškem vidu. Podrobneje se bomo osredotočili na zgradbo transformerja in njegove ključne komponente. V nadaljevanju se bomo posvetili Vision Transformerju (ViT) in njegovi razširjeni verziji - Piramidnem Vision Transformerju (PVT). Posebno pozornost bomo posvetili prilagojeni različici PVT, imenovani PCPVT, saj njeni deskriptorji zagotavljajo prostorsko skladnost in natančno poravnavo. Zaključili bomo s siamskimi nevronskimi mrežami, ki predstavljajo ključno komponento pri primerjavi vzorcev. Te mreže so še posebej pomembne, ko želimo primerjati dva ali več podobnih vzorcev in ugotoviti, ali med njimi obstajajo razlike.

Z vključitvijo vseh teh komponent in tehnik v naš model WAMF-FPI želimo razviti robusten in natančen sistem za lokalizacijo točk na slikah. V nadaljevanju poglavja bomo vsako od teh komponent podrobno raziskali, da bi bolje razumeli njihove lastnosti in kako prispevajo k celotnemu modelu.

2.1 Konvolucijske nevronske mreže

Konvolucijske nevronske mreže (ang. Convolutional Neural Networks - CNN) so metoda globokega učenja, specializirana za obdelavo vizualnih podatkov, zasnovana tako, da avtomatsko in adaptivno izvaja izvleček značilnosti iz slik.

2.1.1 Struktura in delovanje

Osnovni gradniki CNN obsegajo štiri glavne vrste plasti: konvolucijsko, aktivacijsko, združevalno (ang. pooling) in polno povezano.

1. **Konvolucijska plast:** Vsak nevron v tej plasti je povezan le z majhnim območjem v prejšnji plasti, namesto da bi bil povezan z vsemi nevroni, kot je to v običajnih nevronskih mrežah. Ko se filter (ali jedro) premika preko slike, izvaja konvolucijsko operacijo:

$$(I * K)(x, y) = \sum_m \sum_n I(m, n) \cdot K(x - m, y - n) \quad (2.1)$$

2. **Aktivacijska funkcija:** Po konvolucijski operaciji se uporabi aktivacijska funkcija za vsak izhod. Najpogosteje se uporablja funkcija ReLU:

$$\text{ReLU}(x) = \max(0, x) \quad (2.2)$$

3. **Združevalna plast:** Po konvolucijski in aktivacijski operaciji sledi združevalna plast, ki zmanjšuje dimenzije slike z uporabo operacij, kot je "max pooling":

$$P(i, j) = \max_{m, n \in R} I(i + m, j + n) \quad (2.3)$$

4. **Polno povezane plasti:** Delujejo kot klasične plasti v običajnih nevronskih mrežah. Vsak nevron je povezan z vsemi izhodi prejšnje plasti.

$$O_i = \sum_j W_{ij} \cdot I_j + b_i \quad (2.4)$$

Kjer je O_i izhod, W_{ij} matrika uteži, I_j vhod in b_i pristranskost (ang. bias).

2.2 Transformerska arhitektura

V tem podpoglavju bomo obravnavali razvoj in lastnosti transformerske arhitekture. Predstavljeno bo ozadje, vključno s prejšnjimi mehanizmi, kot so rekurentne nevronske mreže, ter podrobnosti o njihovi zgradbi in delovanju.

2.2.1 Predhodni mehanizmi

Praden so obstajali transformerji, so bile najpogostejše metode za obvladovanje zaporedij v jezikovnih modelih rekurentne nevronske mreže (ang. Recurrent Neural Networks - RNN) in njihove različice, kot so dolgokratni kratkotrajni spomini (ang. Long Short-Term Memory - LSTM) in obogatene RNN (ang. Gated Recurrent Units - GRU). Najpogostejša uporaba teh modelov v kontekstu strojnega prevajanja ali drugih nalog pretvarjanja zaporedja v zaporedje je bila uporaba strukture kodirnik-dekodirnik. V tej strukturi je bilo zaporedje vhodnih besed ali kodirano v latentni prostor z uporabo RNN (kodirnik), ta latentni vektor pa je bil nato uporabljen za generiranje zaporedja izhodnih besed ali žetonov z uporabo drugega RNN (dekodirnik). Problem s to strukturo je bil, da je bil latentni prostor omejen na velikost fiksne dolžine in je moral vsebovati vse informacije iz izvirnega zaporedja, ki so potrebne za generiranje ciljnega zaporedja. To je omejevalo model pri obvladovanju dolgih zaporedij, saj je bilo težko ohraniti informacije iz zgodnjega dela zaporedja do konca. Da bi to težavo rešili, so raziskovalci vključili mehanizem pozornosti, ki je omogočil dekodirniku, da se osredotoči na različne dele izvirnega zaporedja na različnih stopnjah generiranja ciljnega zaporedja. To je bil velik napredek, ki je omogočil boljše obvladovanje dolgih zaporedij [1].

2.2.2 Razlaga RNN kodirnik-dekodirnik arhitekture

Definirajmo problem strojnega prevajanja kot iskanje najboljše ciljne sekvence $\vec{E} = (e_0, e_1, \dots, e_m)$ glede na dane izvirne besede $\vec{F} = (f_0, f_1, \dots, f_n)$. Ta problem lahko izrazimo kot optimizacijo pogojne verjetnosti $P(\vec{E}|\vec{F})$.

Začnimo z opisom RNN-kodirnik-dekodirnik arhitekture. Imamo dva RNN modela, kodirnik RNNenc in dekodirnik RNNdec. Kodirnik z zaporedjem vektorjev \vec{F} proizvede skrito stanje h_n :

$$h_n = \text{RNNenc}(f_n, h_{n-1}) \quad (2.5)$$

Začetno stanje h_0 je pogosto postavljeno na nič ali se ga mreža nauči. Dekodirnik nato uporablja to skrito stanje, da generira ciljno zaporedje \vec{E} :

$$e_t = \text{RNNdec}(e_{t-1}, h_{t-1}) \quad (2.6)$$

Opomba: pri učenju se za e_{t-1} pogosto uporablja dejanska vrednost iz ciljnega zaporedja (ne izhod modela), kar je znano kot "teacher forcing" [22]. Izvorna zaporedja besed \vec{F} se tako vnašajo v kodirnik, ki generira skrita stanja za vsako besedo:

$$\vec{H} = \text{Kodirnik}(\vec{F}) \quad (2.7)$$

Za vsako besedo v ciljnem zaporedju \vec{E} se potem izračuna utežena vsota skritih stanj iz kodirnika:

$$\vec{at} = \text{Pozornost}(\vec{H}, e_{t-1}) \quad (2.8)$$

Potem se ta vektor uporabi za napoved ciljne besede:

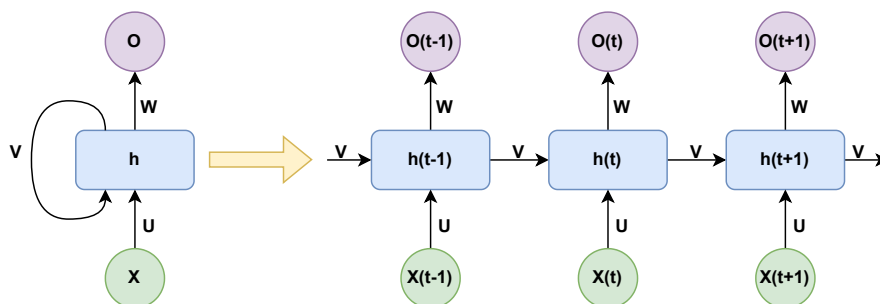
$$e_t = \text{Dekodirnik}(\vec{at}, e_{t-1}) \quad (2.9)$$

Ta pristop omogoča, da dekodirnik upošteva vse besede v izvornem zaporedju, ne samo prejšnje besede v ciljnem zaporedju, kar izboljša kakovost prevoda. Vendar je to zgolj matematična formulacija koncepta. Dejanski detajli, kot so vrste in struktura kodirnika in dekodirnika, so odvisni od specifičnega modela, ki ga uporabljamo.

Na sliki 2.1 je prikazana skica RNN modela.

2.3 Zgradba transformerja

Avtorji v članku [24] so predstavili novo arhitekturo za strojno prevajanje, ki se osredotoča na mehanizme pozornosti in se izogiba omejitvam RNN. Glavna

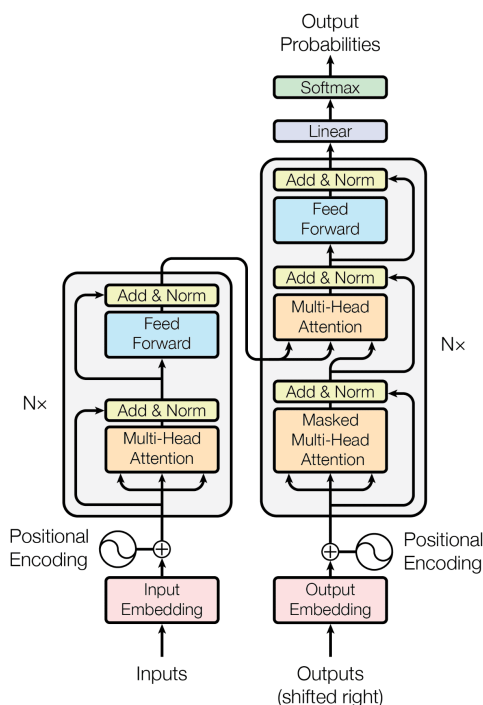


Slika 2.1: Skica RNN modela

inovacija je zamenjava RNN in njihovih skritih stanj z bolj učinkovitimi operacijami na osnovi pozornosti. Transformer model je model kodirnika-dekodirnika. Kodirnik sestavljajo N blokov na levi, dekodirnik pa N blokov na desni, vidno na sliki 2.2.

Med učenjem se vhodne besede $\vec{F} = (f_0, \dots, f_n)$ hkrati prenesejo v prvi blok kodirnika, izhod tega bloka pa se nato prenese v njegovega naslednika. Postopek se ponavlja, dokler vseh N blokov kodirnika ni obdelalo vhoda. Vsak blok ima dve komponenti: plast večglave samopozornosti (ang. Multi-Head Self-Attention), ki ji sledi polno povezana plast z aktivacijami ReLU, ki obdeluje vsak element vhodne sekvence vzporedno. Tako večglav sloj pozornosti kot polno povezana plast sledita koraku *Dodaj in Normiraj* - *dodaj* se nanaša na residualno povezavo, ki doda vhod vsake plasti na izhod, *normiraj* pa se nanaša na normalizacijo plasti. Ko je vhod prešel skozi vse bloke kodiranja, ostane kodirana predstavitev \vec{F} .

Dekodirnik pa sestoji iz treh korakov: maske večglave samopozornosti, večglave plasti pozornosti, ki povezuje kodirano izvorno predstavitev z dekodirnikom, in polno povezane plasti z aktivacijami ReLU. Tako kot v kodirniku, vsaki plasti sledi plast *Dodaj in Normiraj*. Dekodirnik sprejme vse ciljne besede $\vec{E} = (e_0, \dots, e_m)$ kot vhod. V procesu napovedovanja besede e_i ima dekodirnik dostop do prej generiranih besed. Ne more pa imeti dostopa do besed, ki sledijo e_i , saj te še niso bile generirane. Obstaja nekaj ključnih razlik od kodirnika - ena je, da so vhodi v prvo operacijo pozornosti v blokih dekodirnika maskirani, zato ime plasti. To pomeni, da se lahko katera koli



Slika 2.2: Izgled transformerja, iz članka "Attention is all you need" [24].

beseda v ciljnim izhodu nanaša samo na besede, ki so prišle pred njo. Razlog za to je preprost: med sklepanjem generiramo predvideni prevod \vec{E} besedo za besedo z uporabo izvirnega stavka \vec{F} .

Druga razlika od kodirnika je druga večglava plast pozornosti, ki se imenuje tudi plast pozornosti kodirnika-dekodirnika. Za razliko od plasti pozornosti na začetku blokov kodirnika in dekodirnika ta plast ni plast samopozornosti.

2.3.1 Utežena točkovna produktna pozornost

Utežena točkovna produktna pozornost (ang. Scaled Dot-Product Attention) se uporablja v vseh plasteh pozornosti v transformerju. Scaled Dot-Product Attention je skoraj identičen Dot-Product Attention-u, omenjenem prej pri Luongu [1].

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (2.10)$$

Edina razlika je, da je vhod v softmax skaliran s faktorjem $\frac{1}{\sqrt{d_k}}$. Avtorji pozornosti omenjajo, da delijo vhode v softmax funkcijo z $\sqrt{d_k}$, da bi ublažili učinke velikih vhodnih vrednosti, ki bi vodile do majhnih gradientov med učenjem [24].

V članku [24] in predhodni literaturi [1] se vrstice $Q \in \mathbb{R}^{m \times d_k}$ imenujejo poizvedbe, vrstice $K \in \mathbb{R}^{n \times d_k}$ ključ, in vrstice $V \in \mathbb{R}^{n \times d_v}$ vrednosti. Upoštevati je potrebno, da se za izvedbo mora število ključev in vrednosti n ujemati, vendar se lahko število poizvedb m razlikuje. Prav tako se mora dimenzionalnost ključev in poizvedb ujemati, vendar se lahko dimenzionalnost vrednosti razlikuje.

Postopek izračuna utežena točkovne produktne pozornosti je naslednji:

1. Izračunamo produkt med matrikama poizvedb Q in ključev K^T .
2. Produkt normaliziramo z delitvijo z $\sqrt{d_k}$, kjer d_k predstavlja dimenzijo ključev.
3. Na dobljen rezultat uporabimo funkcijo *softmax*, da pridobimo matriko uteži pozornosti.
4. Matriko uteži pomnožimo z matriko vrednosti V , da pridobimo končni izhod.

Vektorji poizvedbe in ključev se med seboj primerjajo preko skalarnega produkta. Ta produkt nam pove, koliko pozornosti naj določen ključ nameni določeni poizvedbi. Utežena vsota vektorskih vrednosti določa, koliko informacij iz vsakega ključa se upošteva v končnem izhodu. V tem postopku so uporabljene le matrične in vektorske operacije, brez dodatnih učljivih parametrov.

2.3.2 Večglava pozornost

Večglava pozornost (ang. Multi-Head Attention) je razširitev mehanizma pozornosti Scaled Dot-Product Attention. V večglavi pozornosti se vhodni podatki (poizvedbe, ključi in vrednosti) najprej transformirajo v več različnih prostorov z uporabo linearnih preslikav. Nato se za vsak niz izračuna funkcija pozornosti Scaled Dot-Product Attention. Rezultati teh funkcij pozornosti se nato združijo skupaj v eno matriko. Končno se ta matrika preslika nazaj v izviren prostor z uporabo druge linearne preslikave, da se pridobi končni rezultat večglave pozornosti. Avtorji to izrazijo v spodnji obliki:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (2.11)$$

Vsak head_i je rezultat izvajanja Scaled Dot-Product Attention na i -tem nizu transformiranih poizvedb, ključev in vrednosti:

$$\text{head}_i = \text{Attention}(QW_{Q_i}, KW_{K_i}, VW_{V_i}) \quad (2.12)$$

kjer so $Q \in \mathbb{R}^{m \times d_{\text{model}}}$, $K \in \mathbb{R}^{n \times d_{\text{model}}}$, in $V \in \mathbb{R}^{n \times d_{\text{model}}}$. Poleg tega, ob upoštevanju hiperparametra h , ki označuje število glav pozornosti, velja: $W_{Q_i} \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_{K_i} \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_{V_i} \in \mathbb{R}^{d_{\text{model}} \times d_v}$, in $W_O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

Vsak izračun glave ima drugačno linearno preslikavo za matrike ključev, poizvedb in vrednosti. Vsaka od teh preslikav se nauči med učenjem.

2.3.3 Pozornost kodirnik-dekodirnik

Tretja in zadnja uporaba pozornosti v članku [24] je pozornost kodirnik-dekodirnik, ki se uporablja v blokih dekodirnika neposredno po sloju maske večglave pozornosti, da se povežejo izvirne in ciljne sekvence. Medtem ko so pri samopozornosti vsi trije vhodi enaka matrika, to tukaj ne velja.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O, \quad \text{head}_i = f(Q, K, V)$$

Ko govorimo o pozornosti med kodirnikom in dekodirnikom, je edina razlika od prej v tem, da Q izhaja iz sloja maske večglave pozornosti, medtem ko

sta K in V kodirani predstavitvi \vec{F} . Lahko bi razmišljali o tem tako, da model zastavlja vprašanje o tem, kako se vsak položaj v ciljni sekvenci nanaša na izvor, in pridobiva predstavitve izvora za uporabo pri generiranju naslednje besede v cilju. Pomembno je poudariti, da vsi bloki dekodirnika prejmejo enake podatke od kodirnika. Od prvega do N -tega bloka dekodirnika vsak uporablja kodirano izvorno sekvenco kot ključe in vrednosti.

2.4 Vision Transformer (ViT)

Transformerji so prvotno bili omejeni na obdelavo zaporedij, kar je idealno za jezik, vendar ne nujno za slike, ki so običajno dvodimenzionalne. To se je spremenilo z razvojem Vision Transformerja (ViT) s strani Google-a [7]. Namesto da bi slike obdelovali kot dvodimenzionalne mreže pikslov (kot to počnejo konvolucijske nevronske mreže), Vision Transformer slike obravnava kot zaporedje majhnih kvadratov ali zaplat. To omogoča uporabo istih tehnik samo-pozornosti, ki so bile učinkovite v jezikovnih modelih, tudi za obdelavo slik. Ta pristop je pokazal obetavne rezultate, saj je Vision Transformer dosegel ali presegel učinkovitost konvolucijskih nevronskih mrež na številnih nalogah računalniškega vida [7].

2.4.1 Arhitektura ViT

Arhitektura ViT obravnava slike dimenzij $H \times W \times C$ tako, da jih razdeli na zaplate dimenzij $P \times P$. Pri tem sta H in W višina in širina slike, C je število barvnih kanalov, P pa predstavlja dimenzijo zaplate. Kot rezultat tega postopka dobimo $(H \cdot W)/P^2$ zaplat, ki se vsaka zravna v 1D vektor dolžine $P^2 \cdot C$.

Vsak 1D vektor x se nato prenese skozi linearni model:

$$z = Wx + b \tag{2.13}$$

Ker transformerji ne vsebujejo inherentne informacije o poziciji vložkov v zaporedju, je treba dodati pozicijske vložke:

$$e_i = z_i + p_i \quad (2.14)$$

Zaporedje vložkov se nato prenese skozi bloke transformerja, ki vsebujejo večglavo samopozornost in feed-forward mreže:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (2.15)$$

Za končno klasifikacijo slike se uporabi klasifikacijska glava:

$$y = \text{softmax}(W_2 \text{ReLU}(W_1 e)) \quad (2.16)$$

2.5 Piramidni ViT (PVT)

Piramidni ViT (PVT) [26] je bil razvit z namenom vključitve piramidne strukture v okviru Transformerja. Arhitektura PVT je razdeljena na štiri stopnje. Vsaka od teh stopenj je sestavljena iz plasti za vdelavo zaplat (ang. patch embedding) in iz več plasti Transformerskega kodirnika. Značilnost te arhitekture je, da se izstopna ločljivost štirih stopenj postopoma zmanjšuje, kar sledi piramidni strukturi. Na najvišji stopnji je ločljivost značilnostne mape največja, medtem ko se na najnižji stopnji zmanjša.

Za boljše razumevanje si pogledjmo podrobneje prvo stopnjo: Vhodna slika velikosti $H \times W \times 3$ je razdeljena na zaplate velikosti $4 \times 4 \times 3$. To pomeni, da je število zaplat enako $HW/4^2$. Vsaka zaplata je nato sploščena in prenesena v linearno projekcijo, kar rezultira v vdelavi zaplat velikosti $HW/4^2 \times C1$. Te

vdelane zaplate, skupaj z dodano vdelavo položaja, prehajajo skozi Transformerski kodirnik z $L1$ plastmi. Izhod iz tega kodirnika je nato preoblikovan v značilnostno mapo $F1$ velikosti $H/4 \times W/4 \times C1$.

Matematično to lahko izrazimo kot:

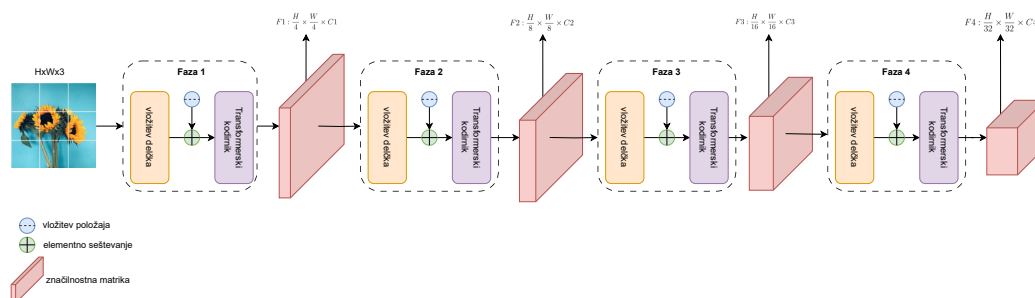
$$F1 = \frac{H}{4} \times \frac{W}{4} \times C1 \quad (2.17)$$

Naslednje stopnje PVT sledijo podobnemu pristopu, vendar z različnimi ločljivostmi in dimenzijami. Na primer, značilnostne mape $F2$, $F3$ in $F4$ so pridobljene z različnimi koraki, ki so 8, 16 in 32 slikovnih pik glede na vhodno sliko.

Ena izmed ključnih inovacij v PVT je uporaba pozornosti za zmanjšanje prostorskega obsega (ang. Spatial Reduction Attention - SRA) namesto tradicionalne večglave pozornostne plasti (ang. Multi Headed Attention - MHA). Ta pristop omogoča PVT-ju, da učinkovito obdela značilnostne mape visoke ločljivosti.

V primerjavi z ViT, PVT prinaša večjo prilagodljivost, saj lahko generira značilnostne mape različnih meril/kanalov v različnih fazah. Poleg tega je bolj vsestranski, saj se lahko enostavno vključi in uporabi v večini modelov za spodnje naloge. Prav tako je bolj prijazen do računalniških virov in spomina, saj lahko obdela značilnostne mape višje ločljivosti.

Na sliki 2.3 je prikazana skica PVT modela.



Slika 2.3: Skica PVT modela

2.6 Piramidni ViT z uporabo lokalnih značilnosti (PCPVT)

Twins-PCPVT [4] je zasnovan na osnovi PVT in CPVT [3]. Glavna razlika med Twins-PCPVT in PVT je v načinu uporabe pozicijskih kodiranj. V PVT so uporabljena absolutna pozicijska kodiranja, medtem ko Twins-PCPVT uporablja pogojna pozicijska kodiranja (ang. Conditional Positional Encoding - CPE), ki so bila predlagana v CPVT.

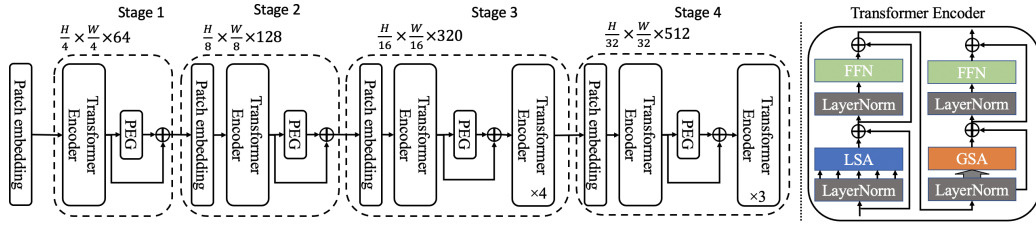
PVT je uvedel piramidno večstopenjsko strukturo z namenom boljšega obravnavanja nalog goste napovedi, kot so zaznavanje objektov in semantična segmentacija. Vendar je bilo ugotovljeno, da je manjša učinkovitost PVT-ja v veliki meri posledica uporabe absolutnih pozicijskih kodiranj. Absolutna pozicijska kodiranja se soočajo s težavami pri obdelavi vhodov različnih velikosti, kar je pogosto v nalogah goste napovedi.

V Twins-PCPVT so absolutna pozicijska kodiranja nadomeščena s pogojnimi pozicijskimi kodiranjmi (CPE), ki so odvisna od vhodov in se tako lahko naravno izognejo zgoraj omenjenim težavam. Generator pozicijskega kodiranja (ang. Positional Encoding Generator - PEG), ki generira CPE, je postavljen za prvim kodirnim blokom vsake stopnje. Uporablja najpreprostejšo obliko PEG, tj. 2D globinsko konvolucijo brez normalizacije serij.

$$CPE = f(PEG(E_1, E_2, \dots, E_n)) \quad (2.18)$$

Kjer je CPE pogojno pozicijsko kodiranje, f je funkcija, ki generira kodiranje na podlagi vhodnih značilnosti, in E_i so značilnosti iz različnih stopenj kodirnika. Twins-PCPVT združuje prednosti tako PVT-ja kot CPVT-ja, kar ga naredi enostavnega za učinkovito implementacijo. Eksperimentalni rezultati so pokazali, da ta preprosta zasnova lahko doseže zmogljivost nedavno predlaganega Swin transformerja [12].

Na sliki 2.4 je prikazana skica PCPVT modela.



Slika 2.4: Skica PCPVT modela, iz članka o modelu Twins [4]

2.7 Siamska nevronska mreža za primerjavo vzorcev

Siamske nevronske mreže predstavljajo sodoben pristop v domeni primerjave vzorcev v računalniškem vidu. Z zmožnostjo učinkovite primerjave med paroma slik so siamske mreže pridobile pozornost v številnih aplikacijah, kjer je ključnega pomena zanesljiva ocena podobnosti. V tem podpoglavju bomo obravnavali osnovno arhitekturo siamske mreže, metodologijo njenega učenja ter aplikacije in prednosti, ki jih prinaša v prakso.

2.7.1 Osnovna arhitektura siamske mreže za primerjavo vzorcev

Klasična siamska mreža za primerjavo vzorcev sestoji iz dveh identičnih podmrež, ki delijo enake uteži. Vsaka podmreža prejme sliko: ena je ciljna slika, druga pa je iskana slika. Oba vhoda se preoblikujeta v značilnostne vektorje prek teh podmrež. Nato se izračuna razdalja med obema vektorjema, običajno z evklidsko razdaljo, da se ugotovi, kako podobni sta sliki.

Matematično, za dve sliki x_1 in x_2 , podmreži proizvedeta predstavitve $f(x_1; \theta)$ in $f(x_2; \theta)$. Razdalja D med tema dvema predstavitvama je določena kot:

$$D(f(x_1; \theta), f(x_2; \theta)) = |f(x_1; \theta) - f(x_2; \theta)|_2 \quad (2.19)$$

2.7.2 Učenje siamske mreže za primerjavo vzorcev

Da bi siamsko mrežo usposobili za učinkovito primerjavo vzorcev, potrebujemo nabor učnih podatkov, ki vsebuje pare podobnih in različnih slik. Med učenjem je cilj zmanjšati razdaljo med podobnimi slikami in povečati razdaljo med različnimi slikami. Kriterijska funkcija, običajno uporabljena pri učenju siamskih mrež za primerjavo vzorcev, je kontrastna kriterijska funkcija, definirana kot:

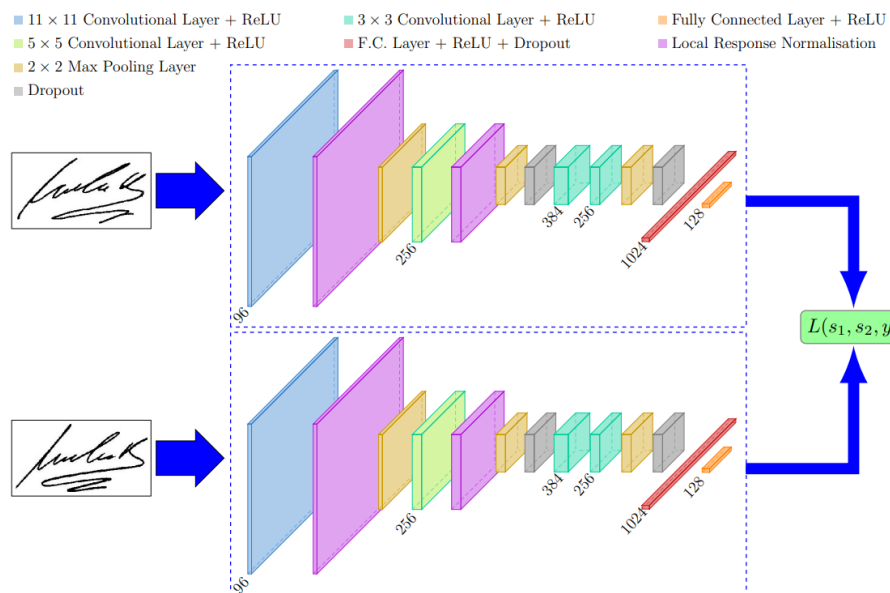
$$L(y, D(f(x_1; \theta), f(x_2; \theta))) = y \cdot \frac{1}{2} D^2 + (1 - y) \cdot \frac{1}{2} \max(0, m - D)^2 \quad (2.20)$$

Kjer y označuje oznako podobnosti (1 za podobne in 0 za različne), m pa je prag, ki določa mejo med podobnimi in različnimi slikami.

Na sliki 2.5 je prikazana skica siamske mreže uporabljene za primerjavo podpisov.

2.7.3 Aplikacije in prednosti

Siamske mreže za primerjavo vzorcev so se izkazale za izjemno koristne v številnih aplikacijah, kot so prepoznavanje in sledenje objektom, biometrija ter varnost in nadzor. V primerjavi s tradicionalnimi metodami imajo siamske mreže večjo odpornost na variacije v svetlobi, rotaciji, lestvici in drugih deformacijah. Zaradi globlje hierarhične predstavitve slike so sposobne zaznati in primerjati kompleksne značilnosti, ki jih manj kompleksne metode morda ne bi opazile.



Slika 2.5: Skica siamske mreže, model SigNet [6]

Poglavje 3

Podatkovna množica

V svetu raziskovanja je podatkovna množica ključnega pomena za razvoj, testiranje in validacijo modelov. Kljub obstoju številnih zbirk, kot so CVUSA [11], CVACT [10] in University-1652 [29], večina ne zajema vseh realnih situacij, s katerimi se srečuje brezpilotni letalnik. Konkretno, CVUSA se osredotoča na zgradbe, medtem ko University-1652 predstavlja predvsem univerzitetna okolja. Zaradi pomanjkljivosti obstoječih zbirk in odsotnosti javno dostopne zbirke UL14 iz [5], smo imeli izziv z zbiranjem ustreznih podatkov za analizo.

Da bi premostili to vrzel, smo se odločili za ustvarjanje lastne zbirke. Za pridobivanje slik iz brezpilotnega letalnika smo uporabili orodje Google Earth Studio [9] in pridobili slike iz 11 evropskih mest. Te slike odražajo raznolikost terena, vključno z zgradbami, parki, zelenimi in vodnimi površinami. Dodatno smo uporabili Mapbox API [14] za pridobitev pripadajočih satelitskih slik.

Skupno je naša podatkovna množica obsežna in vključuje več kot 11.000 slik. V članku [5, 25] so avtorji uporabili podatkovno množico UL14, ki vključuje 6.768 slik za učenje in 2.331 slik za validacijo. Ta zbirka se osredotoča večinoma na slike stavb večjih kitajskih univerz. V nasprotju s tem naša zbirka ponuja širši spekter značilnosti za analizo in bolje odraža realne okoliščine. Cilj izdelave naše zbirke je bil zagotoviti raznolike podatke, ki bi

lahko služili kot robustna osnova za testiranje in validacijo naše implementacije WAMF-FPI. Poleg tega smo želeli, da so slike posnete iz zgornjega pogleda, osredotočene na pogled brezpilotnega letalnika. Želimo se prepričati, da je naš pristop robusten in da lahko obravnava različne scenarije, ki jih lahko sreča brezpilotni letalnik v realnem svetu.

3.1 Slike brezpilotnega letalnika

Nabor podatkov, ki ga predstavljamo, je bil zasnovan z namenom raziskovanja in analize lokalizacije brezpilotnih letalnikov v različnih mestnih scenarijih. Osredotoča se na dve ključni območji:

1. gosto pozidana mestna območja z zgradbami in
2. odprte zelene površine, kot so parki in travniki.

Zajem slik je bil izveden na naključnih poteh po mestu, kar omogoča širok spekter scenarijev. V mestnih območjih je poudarek na razumevanju, kako se brezpilotni letalniki lokalizirajo in navigirajo med visokimi zgradbami, kjer so lahko GPS signali zmanjšani ali moteni. V zelenih območjih je cilj razumeti, kako se brezpilotni letalniki obnašajo v okoljih, kjer so vizualni vzorci manj unikatni. V naboru podatkov za učenje je 10.000 slik iz desetih mest, pri čemer vsako mesto prispeva 1.000 slik. Vsaka slika je opremljena z oznakami lokacije kamere v sistemu ECEF. Sistem ECEF (ang. Earth Centered, Earth Fixed) je globalni koordinatni sistem z izhodiščem v središču Zemlje. Brezpilotni letalniki so bili kalibrirani na višini 150 metrov nad navedeno nadmorsko višino mesta. Kamere na brezpilotnih letalnikih imajo vidno polje 80 stopinj in so usmerjene pravokotno na središče Zemlje. Vse slike so bile ustvarjene z uporabo orodja Google Earth Studio [9].

V naboru so mesta s tipično evropsko arhitekturo, kombinacijo zelenja in stavb. Mesta, vključena v učni nabor podatkov, so:

- **Maribor:** Nadmorska višina: 272m, Višina brezpilotnega letalnika: 150m, Skupaj: 422m nad morsko gladino.

- **Trst:** Nadmorska višina: 23m, Višina brezpilotnega letalnika: 150m, Skupaj: 173m nad morsko gladino.
- **Zagreb:** Nadmorska višina: 158m, Višina brezpilotnega letalnika: 150m, Skupaj: 308m nad morsko gladino.
- **Gradec:** Nadmorska višina: 353m, Višina brezpilotnega letalnika: 150m, Skupaj: 503m nad morsko gladino.
- **Celovec:** Nadmorska višina: 446m, Višina brezpilotnega letalnika: 150m, Skupaj: 596m nad morsko gladino.
- **Videm:** Nadmorska višina: 113m, Višina brezpilotnega letalnika: 150m, Skupaj: 263m nad morsko gladino.
- **Pula:** Nadmorska višina: 17m, Višina brezpilotnega letalnika: 150m, Skupaj: 167m nad morsko gladino.
- **Pordenone:** Nadmorska višina: 24m, Višina brezpilotnega letalnika: 150m, Skupaj: 174m nad morsko gladino.
- **Szombathely:** Nadmorska višina: 212m, Višina brezpilotnega letalnika: 150m, Skupaj: 362m nad morsko gladino.
- **Benetke:** Nadmorska višina: -1m, Višina brezpilotnega letalnika: 150m, Skupaj: 149m nad morsko gladino.

Dodatno je bil v nabor dodan tudi testni nabor podatkov za Ljubljano, ki vključuje 1.000 slik.

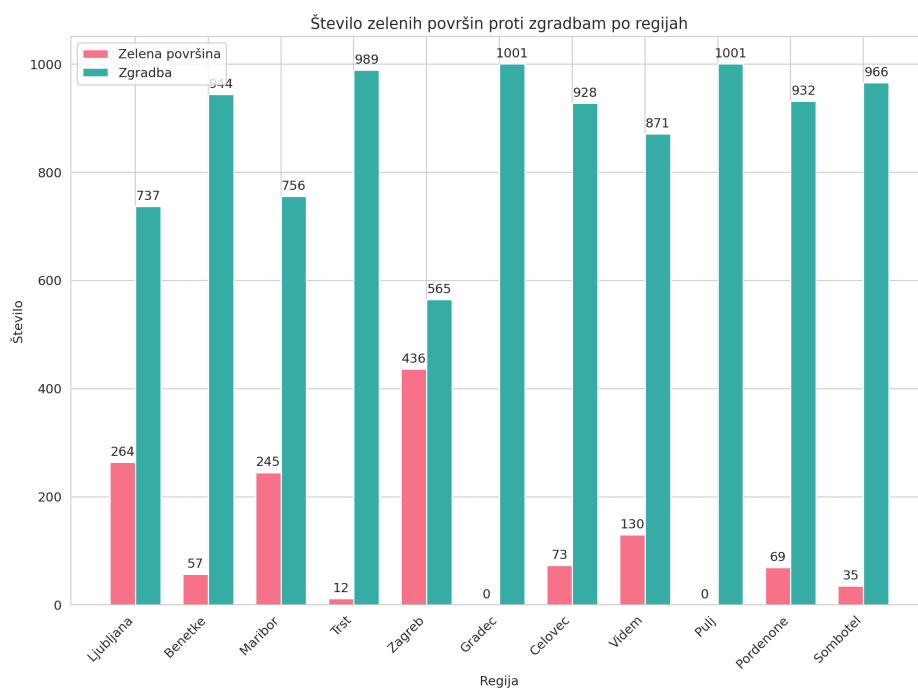
Na Sliki 3.2 je prikazana vizualna razdelitev zelenih površin in stavb za različna mesta, temelječa na analizi slik, ki smo jih zajeli v našem podatkovnem naboru. Vsako mesto razkriva svojo edinstveno strukturo in raven urbanizacije. Te razlike so ključnega pomena pri razumevanju izzivov, s katerimi se srečujejo brezpilotni letalniki pri lokalizaciji in navigaciji v različnih mestnih okoljih.



Slika 3.1: Slika prikazuje lokacije mest, ki so vključena v nabor podatkov.

Nekatera mesta, kot sta Gradec in Pula, kažejo višjo stopnjo urbanizacije z minimalno prisotnostjo zelenih površin. To pomeni, da bodo brezpilotni letalniki v teh okoljih večinoma navigirali med zgradbami. Na drugi strani pa mesta, kot je Zagreb, predstavljajo večjo mešanico zgradb in zelenih površin. Takšne razlike lahko vplivajo na algoritme lokalizacije in navigacije brezpilotnih letalnikov, saj se morajo prilagajati različnim scenarijem in oviram.

Na slikah 3.3 in 3.4 so prikazani raznoliki primeri zajeti z brezpilotnim letalnikom.



Slika 3.2: Graf prikazuje razmerje med zelenimi površinami in stavbami za vsako mesto.



Slika 3.3: Raznoliki primeri slik zajetih z brezpilotnim letalnikom.



Slika 3.4: Raznoliki primeri slik zajetih z brezpilotnim letalnikom.

3.2 Satelitske slike

Za vsako sliko posneto z brezpilotnim letalnikom smo poiskali ustrezno satelitsko zaplato. Ta korak je bil ključnega pomena, saj je zagotovil, da so satelitske slike popolnoma usklajene z slikami posnetimi iz brezpilotnega letalnika v smislu geografske lokacije. Ko smo identificirali ustrezno satelitsko zaplato, smo jo prenesli neposredno iz Mapbox API-ja, vira za visokokakovostne satelitske slike. Da bi zagotovili dodatno globino in kontekst za vsako lokacijo, nismo prenesli samo osrednje zaplate, temveč tudi vse njene sosednje zaplate. Te sosednje zaplate smo nato združili z osrednjo zaplato za ustvarjanje enotne TIFF datoteke.

Ko govorimo o zaplatah v kontekstu kartografije in GIS (Geografski informacijski sistem), se običajno nanašamo na kvadratne segmente, ki pokrivajo Zemljo in se uporabljajo za hitrejše in učinkovitejše prikazovanje zemljevidov na spletu. Sistem zaplat je zelo priljubljen v spletnih kartografskih aplikacijah, kot je Google Maps.

Za pretvorbo geografskih koordinat (latitudo in longitudo) v zaplatne koordinate (x, y) na določeni ravni povečave z uporabo Mercatorjeve projekcije, lahko izrazimo:

- Pretvorba geografskih koordinat v radiane:

$$\begin{aligned} \text{lat}_{\text{rad}} &= \text{latitude} \times \frac{\pi}{180}, \\ \text{lon}_{\text{rad}} &= \text{longitude} \times \frac{\pi}{180} \end{aligned}$$

- Pretvorba radianov v normalizirane koordinate Mercatorja:

$$\begin{aligned} x &= \frac{\text{lon}_{\text{rad}} = +\pi}{2\pi}, \\ y &= \frac{\pi - \log(\tan(\frac{\pi}{4} + \frac{\text{lat}_{\text{rad}}}{2}))}{2\pi} \end{aligned}$$

- Pretvorba normaliziranih koordinat v zaplatne koordinate:

$$\begin{aligned} \text{tile}_x &= \text{floor}(x \times 2^z), \\ \text{tile}_y &= \text{floor}(y \times 2^z) \end{aligned}$$

Na slikah 3.5 in 3.6 so prikazani primeri pripadajočih satelitskih slik za slike zajete z brezpilotnim letalnikom.

3.3 Oznake

V okviru raziskave smo iz visokoločljivostnih satelitskih TIFF datotek naključno izrezali regije velikosti 400x400 pikslov. Pri vsaki iteraciji je bil izrez drugačen, s poudarkom na vključevanju referenčne točke lokalizacije v izrez. Ta pristop zagotavlja izpostavljenost modela različnim scenarijem ob ohranjanju natančnosti lokalizacijskih podatkov. Slike, pridobljene z brezpilotnimi letalniki, so bile obdelane s tehniko *osrednjega izreza* in različnimi stopnjami povečave, združujoč detajlnost teh slik z obsežnostjo satelitskih posnetkov. Spodaj na slikah 3.7, 3.9, 3.8, 3.10 in 3.11 je prikazanih nekaj primerov takšnih izrezov. Na vsaki sliki je s pomočjo rdečega krogeca



Slika 3.5: Primer pripadajočih satelitskih slik.

označen center izreza, ki predstavlja referenčno točko lokalizacije iz brezpi-
lotnega letalnika, s čimer je omogočeno lažje prepoznavanje osredotočenosti
izreza.



Slika 3.6: Primer pripadajočih satelitskih slik.



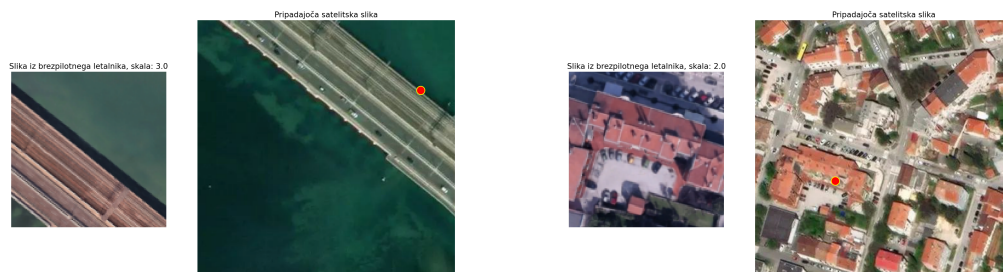
Slika 3.7: Leva slika prikazuje Gradec s 2-kratno povečavo slike iz brezpilotnega letalnika, desna pa Trst s 2,5-kratno povečavo slike iz brezpilotnega letalnika.



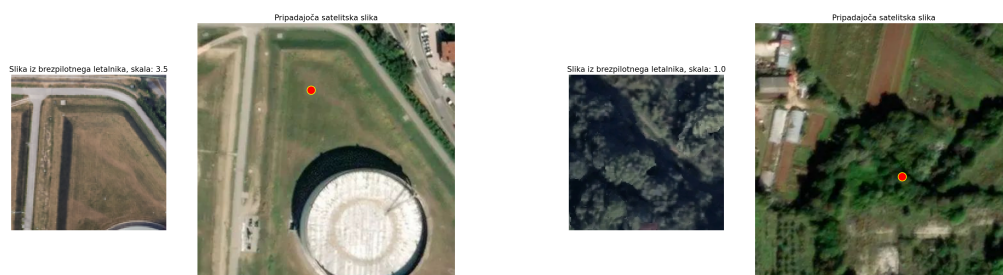
Slika 3.8: Leva slika prikazuje mesto Szombathely s 1,5-kratno povečavo slike iz brezpilotnega letalnika, desna pa z 2-kratno povečavo slike iz brezpilotnega letalnika.



Slika 3.9: Leva slika prikazuje Zagreb s 1,5-kratno povečavo slike iz brezpilotnega letalnika, desna pa mesto Szombathely s 2,5-kratno povečavo slike iz brezpilotnega letalnika.



Slika 3.10: Leva slika prikazuje Benetke s 3-kratno povečavo slike iz brezpilotnega letalnika, desna pa mesto Pula s 2-kratno povečavo slike iz brezpilotnega letalnika.



Slika 3.11: Leva slika prikazuje Trst s 3,5-kratno povečavo slike iz brezpilotnega letalnika, desna pa mesto Pula z 1-kratno povečavo slike iz brezpilotnega letalnika.

Poglavje 4

Rezultati

V tem poglavju so podrobno predstavljeni rezultati, doseženi v različnih fazah implementacije in optimizacije modela WAMF-FPI. Naš izhodiščni korak je bil zagotoviti stabilno osnovo, kar smo dosegli z implementacijo modela skladno z metodologijo, opisano v izvirnem članku. Ta pristop nam je zagotovil referenčno točko, od katere smo izvajali nadaljnje optimizacije in izboljšave.

Med optimizacijo modela smo se posvetili iskanju optimalne kriterijske funkcije. Da bi bolje razumeli, katera funkcija bi lahko prinesla najboljše rezultate v našem primeru, smo izvedli serijo eksperimentov z različnimi funkcijami ter jih evalvirali glede na njihovo učinkovitost in zanesljivost. Kot naslednji korak smo preučili stratificirano vzorčenje, tehniko, ki bi lahko pripomogla k izboljšanju natančnosti in robustnosti modela z zagotavljanjem bolj uravnoteženega učnega nabora. Pregledali smo tudi vpliv Hanningovega okna ter analizirali, kako različne velikosti tega okna vplivajo na končne rezultate modela.

V zaključni fazi naših eksperimentov smo se osredotočili na regularizacijo, predvsem na tehniko izpuščanja nevronov. Zaradi kompleksnosti modelov globokega učenja smo želeli razumeti, kako bi regularizacija lahko pomagala preprečiti prekomerno prilagajanje ter izboljšala splošno učinkovitost modela. Vsako od teh področij je v nadaljevanju podrobno obravnavano, pri čemer so podane analize, interpretacije in ključne ugotovitve, ki smo jih pridobili v

tem procesu.

4.1 Implementacija

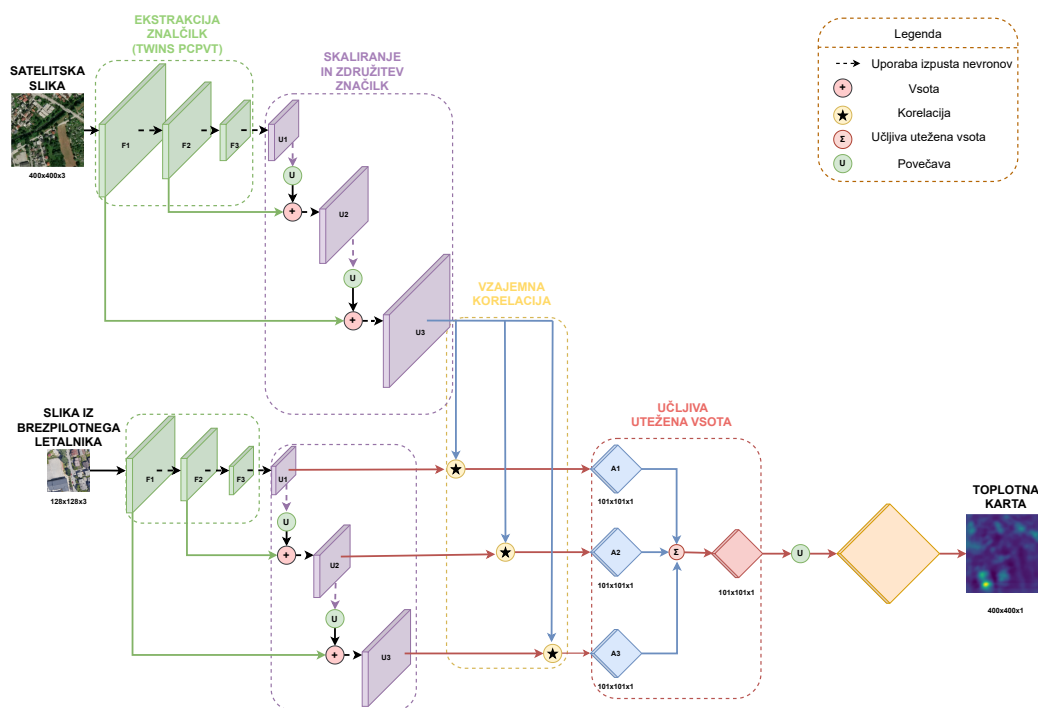
Sledenje objektov v okviru računalniškega vida običajno temelji na izračunu podobnosti med referenčno in iskalno podobo v trenutnem okviru. Medtem ko temeljna metoda za iskanje točk znotraj slike izhaja iz metodologije sledenja objektov, je prva v primerjavi z drugo bolj zapletena. To je posledica različnih perspektiv med predlogo (sliko posneto z brezpilotnim letalnikom) in iskalno sliko (satelitsko sliko), ki povzročajo veliko variacijo.

Metoda iskanja točk uporablja satelitsko sliko kot referenčno in sliko iz brezpilotnega letalnika kot poizvedbeno. Obe sliki – posneto z brezpilotnim letalnikom in satelitsko sliko relevantnega območja – se nato prenesejo v end-to-end mrežo. Po obdelavi je rezultat toplotna karta, kjer točka z najvišjo vrednostjo predstavlja lokacijo brezpilotnega letalnika, kot jo predvideva model. Lokacijo nato preslikamo na satelitsko sliko, pri čemer položaj brezpilotnega letalnika določimo na podlagi geografske širine in dolžine, ki jih vsebuje satelitska slika. V FPI [5] avtorji kot modul za izluščenje značilnosti uporabljajo dva Deit-S [23] brez deljenih uteži za vertikalne poglede slike brezpilotnega letalnika in satelitske slike. Izluščene značilnosti nato uporabimo za izračun podobnosti in izdelavo toplotne karte. Lokacijo z najvišjo vrednostjo toplotne karte nato preslikamo na satelitsko sliko, da določimo lokacijo brezpilotnega letalnika.

V FPI je za izračun podobnosti uporabljena zadnja plast zemljevidnih značilnosti [5]. Zaradi tega, ker je izhodna toplotna karta 16-krat manjša od vhodne satelitske slike, model izgubi veliko prostorskih informacij, kar vodi v znatno izgubo natančnosti pozicioniranja.

Da bi izboljšali lokalizacijske sposobnosti modela, smo uporabili strukturo piramidnih značilnosti (Twins-PCPVT) in modul utežno prilagodljivega združevanja večznačilnostnih lastnosti (WAMF). K osnovnemu modelu so bile dodane izboljšave z vključitvijo dveh močnejših PCPVT-S modulov za

izluščenje značilnosti iz slik brezpilotnega letalnika in satelitskih slik. Da bi boljše zajeli informacije na različnih ločljivostih in ohranili več prostorskih informacij, so bile prvotno izluščene značilnosti poslane v omrežje značilnostne piramide za nadaljnjo obdelavo. Modul WAMF je bil nato uporabljen za izračun podobnosti in združevanje različnih značilnosti. Končne združene značilnosti so bile razširjene za izdelavo končne izhodne napovedne mape. Rezultat je toplotna karta iste velikosti kot vhodna satelitska slika v modelu WAMF-FPI. Na sliki 4.1 je prikazana skica arhitekture modela WAMF-FPI.



Slika 4.1: Skica arhitekture modela

4.1.1 Modul za izluščenje značilnosti

Model WAMF-FPI temelji na strukturi, ki je podobna siamski arhitekturi, vendar se od tradicionalnih metod sledenja objektom loči v določenih ključnih točkah, ki jih bomo v tem razdelku opisali. Zaradi občutne razlike med sa-

telitskimi slikami in slikami brezpilotnega letalnika, ki izvirajo iz različnih naprav, veji modela WAMF-FPI za vsako od teh vrst slik ne uporabljata metode deljenja uteži. WAMF-FPI kot vhod uporablja satelitske slike dimenzij $400 \times 400 \times 3$ in slike brezpilotnega letalnika dimenzij $128 \times 128 \times 3$. Značilnosti obeh vrst slik so izluščene s pomočjo PCPVT-S.

V modelu smo odstranili zadnjo stopnjo PCPVT-S in uporabili samo prve tri stopnje za izluščene značilnosti. Pri dimenzijah vhodnih slik $400 \times 400 \times 3$ in $128 \times 128 \times 3$ oba pristopa pridobita značilnostne mape z obliko $25 \times 25 \times 256$ in $8 \times 8 \times 320$ oziroma.

V primerjavi z Deit-S [23], ki je bil uporabljen v FPI [5], ima PCPVT-S piramidno strukturo. Ta struktura je bolj prilagodljiva za naloge goste napovedi. Pravzaprav uporaba piramidne strukture zagotavlja osnovo za kasnejšo integracijo modula WAMF. Poleg tega Mreža s piramidno strukturo lahko zmanjša obseg potrebnih izračunov in s tem izboljša hitrost procesiranja, kar je ključno za učinkovito uporabo metode v praksi.

Po izluščanju informacij iz slike s pomočjo PCPVT-S se podobnost neposredno izračuna na zadnjih značilnostnih mapah. Kljub temu je končni izhod stisnjen samo za faktor štiri v primerjavi z vhomom, kar je potem s bikubično interpolacijo povečano nazaj na velikost vhodne satelitske slike.

Pristranskost, ki je posledica nizke ločljivosti značilnostne mape, je bila odstranjena že na samem začetku. Ker značilnostna mapa z visoko ločljivostjo vsebuje več prostorskih informacij, je bila združena z globoko značilnostno mapo, bogato s semantičnimi informacijami, preko lateralne povezovalne strukture.

Prva faza obdelave v metodi WAMF-FPI je uporaba konvolucijskega jedra velikosti ena, ki prilagodi kanalsko dimenzijo tri-stopnjske značilnostne mape, pridobljene s pomočjo PCPVT-S. Število izhodnih kanalov je bilo nastavljeno na 64, kar zagotavlja kompaktno in učinkovito zastopanje značilnosti. Po tej fazi sledi upsampling operacija na značilnostnih mapah zadnjih dveh stopenj, ki poveča njihovo ločljivost in s tem omogoča bolj precizno lokalizacijo. Te mape se nato kombinirajo z značilnostnimi mapami istega merila iz osnov-

nega modela.

Končno, značilnosti se dodatno izluščene s pomočjo konvolucijskega jedra velikosti 3, kar modelu omogoča izluščanje bolj kompleksnih značilnosti iz združenih map. Rezultat je združena značilnostna mapa, ki združuje plitve (prostorske) in globoke (semantične) informacije. Ta bogata kombinacija modelu omogoča učinkovito prepoznavanje in lokalizacijo objektov na vhodnih slikah.

4.1.2 Arhitektura utežno-prilagodljivega združevanja večznačilnostnih lastnosti (WAMF)

Modul za združevanje značilnosti je zasnovan tako, da združuje informacije iz dveh ločenih vhodnih tokov, v tem primeru iz UAV (brezпилotnega letalnika) in SAT (satelita). Ta modul uporablja piramido značilnosti iz obeh in izračuna korelacije med njimi, da jih združi v en sam izhod.

Za začetek se izvedejo konvolucijske operacije na značilnostnih mapah UAV in SAT. Konvolucijske operacije so izvedene s konvolucijskimi jedri velikosti 1×1 , kar omogoča prilagoditev kanalskih dimenzij značilnostnih map.

Za UAV značilnostne mape:

$$U1_{UAV} = \text{Conv1UAV}(s3_{UAV}) \quad (4.1)$$

$$U2_{UAV} = \text{Povečava}(U1_{UAV}) + \text{Conv2UAV}(s2_{UAV}) \quad (4.2)$$

$$U3_{UAV} = \text{Povečava}(U2_{UAV}) + \text{Conv3UAV}(s1_{UAV}) \quad (4.3)$$

Za SAT značilnostne mape:

$$U1_{SAT} = \text{Conv1SAT}(s3_{SAT}) \quad (4.4)$$

$$U2_{SAT} = \text{Povečava}(U1_{SAT}) + \text{Conv2SAT}(s2_{SAT}) \quad (4.5)$$

$$U3_{SAT} = \text{Povečava}(U2_{SAT}) + \text{Conv3SAT}(s1_{SAT}) \quad (4.6)$$

Kjer je Povečava funkcija, ki poveča prostorsko resolucijo značilnostne mape z uporabo bikubične interpolacije.

$$A1 = \text{corr}(U1_{\text{UAV}}, U3_{\text{SAT}}) \quad (4.7)$$

$$A2 = \text{corr}(U2_{\text{UAV}}, U3_{\text{SAT}}) \quad (4.8)$$

$$A3 = \text{corr}(U3_{\text{UAV}}, U3_{\text{SAT}}) \quad (4.9)$$

Kjer je corr funkcija za izračun korelacije med dvema značilnostnima mapama.

Korelacija v kontekstu obdelave slik je postopek izračuna podobnosti med dvema slikama ali značilnostnima mapama. V osnovi ena značilnostna mapa (poimenovana poizvedba) drsi čez drugo značilnostno mapo (poimenovana iskalna regija) in izračuna podobnost med njima na vsaki lokaciji. Rezultat tega postopka je nova značilnostna mapa, imenovana korelacijska mapa, kjer vsaka vrednost predstavlja stopnjo podobnosti med poizvedbo in delom iskalne mape na določeni lokaciji.

Matematično je korelacija med dvema funkcijama f in g definirana kot:

$$(f \star g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t + \tau)d\tau \quad (4.10)$$

V kontekstu diskretnih signalov, kot so slike ali značilnostne mape, je korelacija definirana kot:

$$(f \star g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n + m] \quad (4.11)$$

Nazadnje se izvede uteženo združevanje teh treh koreliranih značilnostnih map s pomočjo naučljivih uteži:

$$\text{združena_mapa} = w_1 \cdot A1 + w_2 \cdot A2 + w_3 \cdot A3 \quad (4.12)$$

Za dokončanje postopka se uporabi bikubična interpolacija, da se združena mapa poveča na velikost vhodne satelitske slike. Na izhodu dobimo toplotno karto iste velikosti kot vhodna satelitska slika v WAMF-FPI.

4.1.3 RDS metrika

Da bi lahko ovrednotili in primerjali zmogljivost našega modela, uporabljamo metriko RDS [25]. Zaradi različnih meril podatkov v naboru podatkov vsak piksel v različnih satelitskih slikah predstavlja različno razdaljo. Čeprav model morda najde točko, ki je na satelitski sliki blizu dejanske lokacije, lahko v resničnem prostoru povzroči veliko napako. Da bi se izognili težavam zaradi spremembe merila, RDS izračuna relativno razdaljo na ravni pikselov med napovedano in dejansko točko.

Enačba za izračun RDS je naslednja:

$$RDS = e^{-k \times \frac{\sqrt{\left(\frac{dx}{w}\right)^2 + \left(\frac{dy}{h}\right)^2}}{2}} \quad (4.13)$$

Kjer so:

- w širina v pikslih satelitske slike,
- h višina v pikslih satelitske slike,
- dx pikselska razdalja med vodoravnimi koordinatami napovedane pozicije in dejanske pozicije,
- dy pikselska razdalja med navpičnimi koordinatami napovedane pozicije in dejanske pozicije,
- k je faktor merila, ki je v tem delu postavljen na 10.

Za lažje razumevanje delovanja RDS metrike smo dodali dodatek A k diplomskemu delu, ki vsebuje tri primere izračuna metrike.

4.2 Učenje modela

Model smo učili na računalniški konfiguraciji, opremljeni s procesorjem Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60GHz z 12 jedri. Dodatno je računalnik vseboval grafično kartico NVIDIA GeForce RTX 3060 z 12 GB pomnilnika. Naš razvoj je temeljil na platformi Ubuntu z uporabo Python [15] knjižnice PyTorch [16]. V času učenja našega modela ni prišlo do povečane energetske porabe, saj je bil naš računalnik neprestano napajen iz lokalne sončne elektrarne. To pomeni, da je bil celoten postopek učenja izveden na okolju prijazen način, brez dodatnega obremenjevanja električnega omrežja ali uporabe fosilnih goriv.

Za doseg optimalnih rezultatov smo uporabili specifične hiperparametre in nastavitve:

Hitrost učenja: Uporabljena sta bila dva različna parametra: $lr_fusion = 0.0004$ za združevanje in $lr_backbone = 0.0001$ za osnovno arhitekturo.

Prilagajanje hitrosti učenja: $gamma = 0.2$ z mejniki na epochah 9, 13 in 15.

Delovni procesi: Skupno 24 hkratnih delovnih procesov (`num_workers = 24`).

Epoh: Model je bil učen skozi 24 epoh.

Velikost serije: `batch_size = 16`.

Mešanje podatkov: Podatki so bili premešani pred vsako epoho.

Funkcija izgube: Uporabljena je bila `hanning` funkcija.

Vizualizacija: Vključena za spremljanje napredka učenja.

Za vsako iteracijo učenja smo iz vsake satelitske TIFF datoteke naključno izrezali regijo velikosti 400x400 pikslov. Ključnega pomena je bilo, da se je točka lokalizacije vedno nahajala nekje znotraj te izrezane regije. Ta metoda

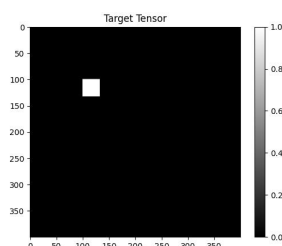
nam je zagotovila, da je bil model izpostavljen širokemu naboru scenarijev in kontekstov, hkrati pa smo ohranili natančnost in relevantnost lokalizacijskih podatkov. S tem pristopom smo uspešno sestavili nabor podatkov, ki združuje najboljše iz obeh svetov: detajlnost slik posnetih z brezpilotnim letalnikom in širino satelitskih slik, kar omogoča poglobljeno analizo in učinkovito učenje.

4.3 Izbira kriterijske funkcije

Zanimalo nas je, kako se bo model obnesel, ko izbiramo različne kriterijske funkcije.

4.3.1 Hanningova kriterijska funkcija

V članku WAMF-FPI [25] so avtorji predlagali uporabo Hanningove kriterijske funkcije. Prvi pomemben vidik te funkcije izgube je dodelitev uteži vzorcem. Namesto enakega pomena vseh pozitivnih vzorcev, kriterijska funkcija Hanning dodeli različne uteži glede na lokacijo vzorca.



Slika 4.2: Primer vzorca, središče je točka lokacije vzorca.

To je zato, ker je pomembnost središčnega položaja veliko večja kot pomembnost robovih položajev, kar v kontekstu satelitskih slik logično smiselno. Za normalizacijo teh pozitivnih uteži se uporablja Hanningovo okno, za normalizacijo negativnih uteži pa $1/\#\text{negativnih vzorcev}$. Uteži so dodeljene tako, da je vsota uteži pozitivnih in negativnih vzorcev enaka 1. Toda

ker je število negativnih vzorcev običajno večje od števila pozitivnih vzorcev, postane utež negativnih vzorcev manjša. Da bi to popravili, se uvede hiperparameter, imenovan Negativna utež (NG), ki prilagodi utež negativnih vzorcev.

Hanningova funkcija:

$$\text{Hanning}(n) = \begin{cases} 0.5 + 0.5 \cos\left(\frac{2\pi n}{M-1}\right) & \text{za } 0 \leq n \leq M-1 \\ 0 & \text{sicer} \end{cases} \quad (4.14)$$

Utezi primerov:

- Utež negativnih vzorcev:

$$w_{pos} = NG / (NN(NW + 1))$$

- Utež pozitivnih vzorcev:

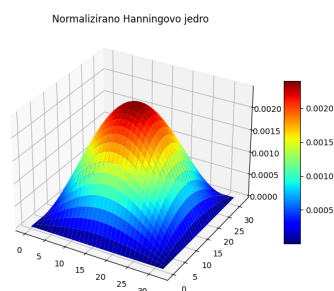
$$w_{neg} = HN(n) / (NW + 1)$$

Kjer je:

- **NG** je Negativna utež
- **NN** je število vseh vzorcev
- **NW** je normalizacijski faktor
- **HN(n)** je vrednost Hanningove funkcije na lokaciji n

4.3.2 Gaussovo utežena srednja kvadratna napaka

Gaussovo utežena srednja kvadratna napaka (ang. Gaussian Weighted Mean Squared Error - GWMSE) je spremenjena funkcija izgube, namenjena izboljšanju modelov, ki obravnavajo podatke, kot so satelitske slike. Glavna značilnost GWMSE je dodeljevanje uteži vzorcem na zelo podoben način

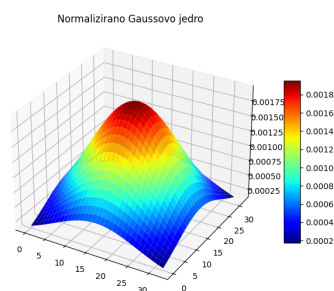


Slika 4.3: Normalizirano Hanningovo jedro

kot pri Hanningovi funkciji izgube. Namesto enakega pomena vseh pozitivnih vzorcev, GW MSE različnim vzorcem dodeljuje različne uteži glede na njihovo lokacijo. Za normalizacijo teh uteži se uporablja Gaussova funkcija.

Gaussova funkcija:

$$\text{Gauss}(n) = \begin{cases} \exp\left(-\frac{(n-\mu)^2}{2\sigma^2}\right) & \text{za } 0 \leq n \leq M-1 \\ 0 & \text{sicer} \end{cases} \quad (4.15)$$



Slika 4.4: Normalizirano Gaussovo jedro

4.3.3 Hanningovo utežena srednja kvadratna napaka

Hanningovo utežena srednja kvadratna napaka (ang. Hanning Weighted Mean Squared Error - HWMSE) je spremenjena funkcija izgube, namenjena

izboljšanju modelov, ki obravnavajo podatke, kot so satelitske slike. Glavna značilnost HWMSE je dodeljevanje uteži vzorcem na zelo podoben način kot pri Gaussovi funkciji izgube. Namesto enakega pomena vseh pozitivnih vzorcev, HWMSE različnim vzorcem dodeljuje različne uteži glede na njihovo lokacijo. Za normalizacijo teh uteži se uporablja Hanningovo okno.

Hanningova funkcija je podana kot:

$$\text{Hanning}(n) = \begin{cases} 0.5 + 0.5 \cos\left(\frac{2\pi n}{M-1}\right) & \text{za } 0 \leq n \leq M-1 \\ 0 & \text{sicer} \end{cases} \quad (4.16)$$

4.3.4 Križno utežena srednja kvadratna napaka

Funkcija izgube križno utežena srednja kvadratna napaka (ang. Cross-Weighted Mean Squared Error - CWMSE) je različica standardne srednje kvadratne napake (Mean Squared Error - MSE), ki vključuje uteževanje dveh različnih skupin vzorcev: tistih, katerih resnična vrednost je večja od 0 (t.i. "resničnih" vzorcev) in tistih, katerih resnična vrednost je manjša ali enaka 0 (t.i. "ne-resničnih" vzorcev). Končna funkcija izgube se izračuna kot utežena kombinacija srednjih kvadratnih napak za "resnične" in "ne-resnične" vzorce, pri čemer se uteži vzorcev različnih skupin prekrizajo. Ta pristop se formalno izraža z naslednjo enačbo:

$$\text{loss} = \frac{\text{true_weight} \cdot N_{\text{true}} \cdot \text{MSE}_{\text{false}} + \text{false_weight} \cdot N_{\text{false}} \cdot \text{MSE}_{\text{true}}}{N_{\text{all}}} \quad (4.17)$$

- N_{true} : število vzorcev, katerih resnična vrednost je večja od 0.
- N_{false} : število vzorcev, katerih resnična vrednost je enaka ali manjša od 0.
- N_{all} : skupno število vzorcev.
- $\text{MSE}_{\text{true}} = \frac{1}{N_{\text{true}}} \sum_{i=1}^{N_{\text{true}}} (y_i - \hat{y}_i)^2$ za vzorce, katerih resnična vrednost je večja od 0.

- $MSE_{\text{false}} = \frac{1}{N_{\text{false}}} \sum_{i=1}^{N_{\text{false}}} (y_i - \hat{y}_i)^2$ za vzorce, katerih resnična vrednost je enaka ali manjša od 0.
- `true_weight` in `false_weight`: uteži, dodeljene skupinama *true* in *false*.

4.3.5 Primerjava rezultatov

V kontekstu geolokalizacije brezpilotnih letalnikov v modelu WAMF-FPI je Hanningova kriterijska funkcija izkazala izjemno učinkovitost glede na vrednosti RDS. Kot je razvidno iz Tabele 4.1, razmerje RDS_{train} za Hanningovo kriterijsko funkcijo je 0.893, kar kaže na visoko natančnost pri učenju modela. Čeprav se razmerje RDS_{val} zmanjša na 0.709, je še vedno precej višje v primerjavi z drugimi preučevanimi kriterijskimi funkcijami. V primerjavi s Hanningovo kriterijsko funkcijo so druge kriterijske funkcije praktično neuporabne, kar potrjuje, da je Hanningova kriterijska funkcija optimalna izbira za geolokalizacijo brezpilotnih letalnikov v obravnavanem modelu [25].

Kriterijska funkcija	vrednost	RDS_{train}	RDS_{val}	$\overline{\Delta}_m$ [m]
HANN	8.49	0.893	0.709	43.42
GWMSE	0.001	0.077	0.074	234.48
HWMSE	4.04e-06	0.061	0.059	232.55
CWMSE	0.007	0.07	0.06	242.70

Tabela 4.1: Rezultati ob uporabi različnih kriterijskih funkcij. Kjer je $\overline{\Delta}_m$ povprečna napaka v metrih.

4.3.6 Analiza rezultatov

Hanningova kriterijska funkcija, ki je značilna po dodeljevanju uteži vzorcem glede na njihovo lokacijo, je na učni množici dosegla skupno vrednost 8.49 in RDS_{train} vrednost 0.893. Kljub temu, da je na validacijski množici dosegla nekoliko nižjo RDS_{val} vrednost 0.709, to kaže, da se je znanje dobro preneslo

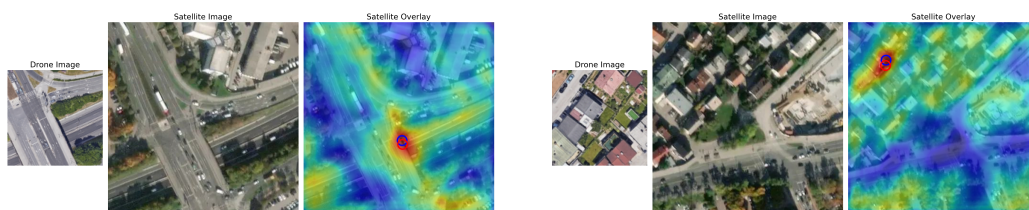
na validacijsko množico. Zaradi teh pozitivnih rezultatov smo Hanningovo kriterijsko funkcijo uporabljali v nadaljnjem testiranju.

V spodnjem razdelku so na slikah 4.5, 4.6, 4.7, 4.8, 4.9 in 4.10 predstavljeni primeri lokalizacije z modelom WAMF-FPI na vzorcu iz validacijske množice za Ljubljano. Ti izbrani primeri osvetlujejo uspešnost in pomanjkljivosti modela pri obvladovanju kompleksnih scenarijev lokalizacije. S pomočjo teh primerov lahko podrobneje razumemo zmogljivosti in omejitve uporabljenega modela v praksi. Na vsaki sliki je z rdečim krogcem označena dejanska lokacija (ang. ground truth), medtem ko je z modrim krogcem označena predikcija modela, ki predstavlja najvišjo točko v toplotni karti.

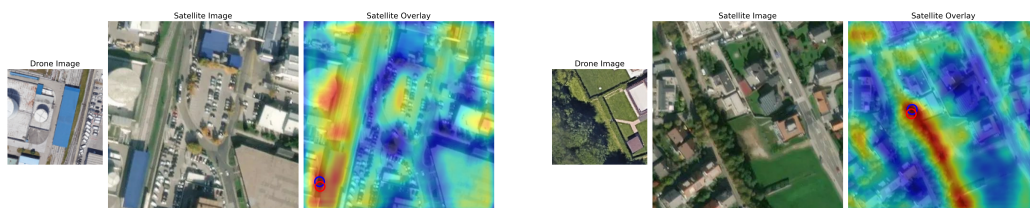
Primeri dobre lokalizacije



Slika 4.5: Leva slika prikazuje stanovanjsko hišo v stanovanjskem naselju z napako 6.93m in RDS vrednostjo 0.90. Na desni vidimo parkirišče z monotono okolico. Napaka je 3.77m, RDS pa 0.94.

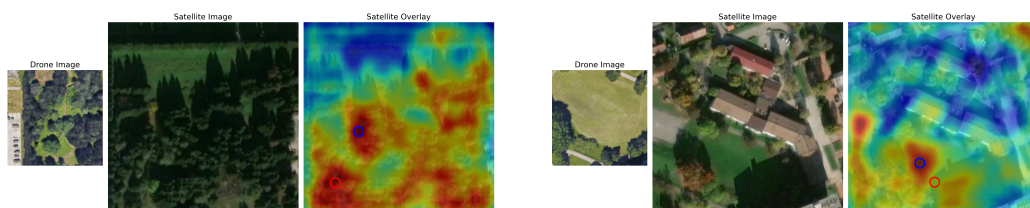


Slika 4.6: Leva slika prikazuje izsek avtoceste z napako 6.55m in RDS vrednostjo 0.91. Desna slika okolico stanovanjske hiše v naselju z napako 5.73m in RDS vrednostjo 0.90.

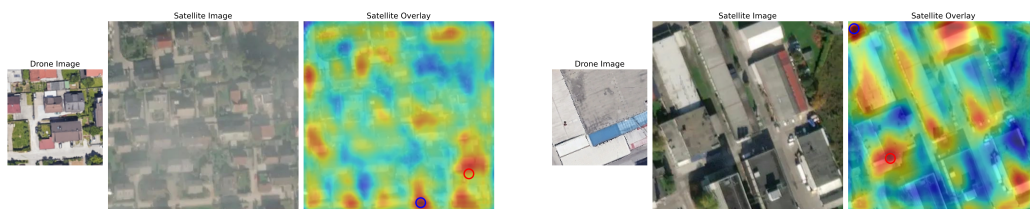


Slika 4.7: Na levi je industrijska stavba v industrijski coni z napako 3.99m in RDS vrednostjo 0.95. Desna slika pa prikazuje travnik ob stanovanjskih hišah v naselju z napako 2.07m in RDS vrednostjo 0.97.

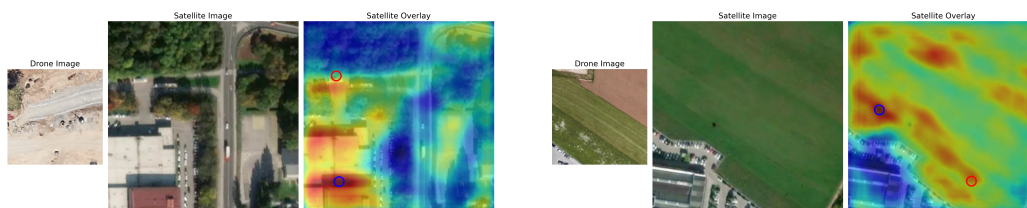
Primeri slabe lokalizacije



Slika 4.8: Obe sliki prikazujeta travnike in drevesa. Napaka na levi sliki je 50.74m z RDS vrednostjo 0.31, na desni pa 30.16m z RDS vrednostjo 0.42.



Slika 4.9: Leva slika prikazuje stanovanjsko hišo, kjer satelitsko sliko prekriva oblak, otežujoč izveleček značilnosti. Napaka je 54.24m in RDS je 0.30. Desna slika prikazuje industrijsko stavbo v monotoni industrijski coni z napako 150.47m in RDS vrednostjo 0.10.



Slika 4.10: Leva slika prikazuje gradbšče prisotno na sliki iz brezpilotnega letalnika, odsotno na satelitski sliki z napako 169.43m in RDS vrednostjo 0.09. Desna slika pa polje v monotoni okolici z napako 155.47m in RDS vrednostjo 0.10.

4.4 Učenje s stratificiranim vzorčenjem

Stratificirano vzorčenje igra ključno vlogo pri ocenjevanju kakovosti modela v heterogenih podatkovnih zbirkah. V tem podpoglavju bomo raziskali kaj so prednosti in slabosti stratificiranega vzorčenja.

4.4.1 Stratificirano vzorčenje

Stratificirano vzorčenje zagotavlja, da so vse podkategorije v populaciji ustrezno zastopane v vzorcu, kar izboljša natančnost ocen. Vendar pa model morda ni pripravljen na povsem nove, nevidene podatke in se lahko preveč prilagodi specifični distribuciji podatkov.

4.4.2 Rezultati

Način	HANN_{val}	$\text{RDS}_{\text{train}}$	RDS_{val}	$\overline{\Delta}_m$ [m]
Originalno učenje	8.49	0.893	0.709	43.42
Učenje s strat. vzorčenjem	3.17	0.750	0.731	17.89

Tabela 4.2: Rezultati ob uporabi stratificiranega vzorčenja. Kjer je $\overline{\Delta}_m$ povprečna napaka v metrih.

Iz rezultatov 4.2 je razvidno, da stratificirano vzorčenje pozitivno vpliva na natančnost modela. Model, naučen s to metodo, je dosegel rahlo višjo uspešnost na validacijski množici in boljšo generalizacijo. Kljub temu je treba upoštevati omejitve stratificiranega vzorčenja, kot so omejena generalizacija in težave pri podatkih, ki se močno razlikujejo od originalne distribucije.

4.5 Vpliv velikosti Hanningovega okna

Hanningovo okno, ključni element za določanje uteži vzorcev v satelitskih slikah, se prilagaja glede na svojo velikost. Spreminjanje velikosti okna neposredno vpliva na razporeditev in obliko uteži, kar ima posledično vpliv na kakovost rezultatov.

4.5.1 Dinamika različnih velikosti Hanningovih oken

Majhna velikost okna omejuje območje vzorcev, ki ga zajema. Takšna omejitev lahko zmanjša učinkovitost povratnega razširjanja med učenjem modela, saj kriterijska funkcija nima dovolj širokega vpliva na celotno mrežo. Nasprotnje predstavlja preveliko okno, ki zajema široko paleto vzorcev. Kljub širšemu zajemu, lahko detajli v sliki postanejo manj opazni, kar zmanjšuje natančnost predikcij.

4.5.2 Eksperimentalni rezultati

V eksperimentu smo vsak model posebej natrenirali z različnimi velikostmi Hanningovega okna, da bi ocenili vpliv velikosti oken na modelovo natančnost. Za testiranje smo uporabili kombinacijo slike iz brezpilotnega letalnika in satelitske slike, zagotavljajoč enake vhodne podatke za vse modele. Referenčni sliki za testiranje sta prikazani na sliki 4.11.

Eksperimenti so bili izvedeni z različnimi velikostmi oken, da bi ugotovili njihov vpliv na uspešnost modela. Primeri so prikazani na slikah B.1. Podatki kažejo na optimalno ravnovesje med velikostjo oken in natančnostjo



Slika 4.11: Primer referenčnih slik, ki smo jih uporabili za testiranje.

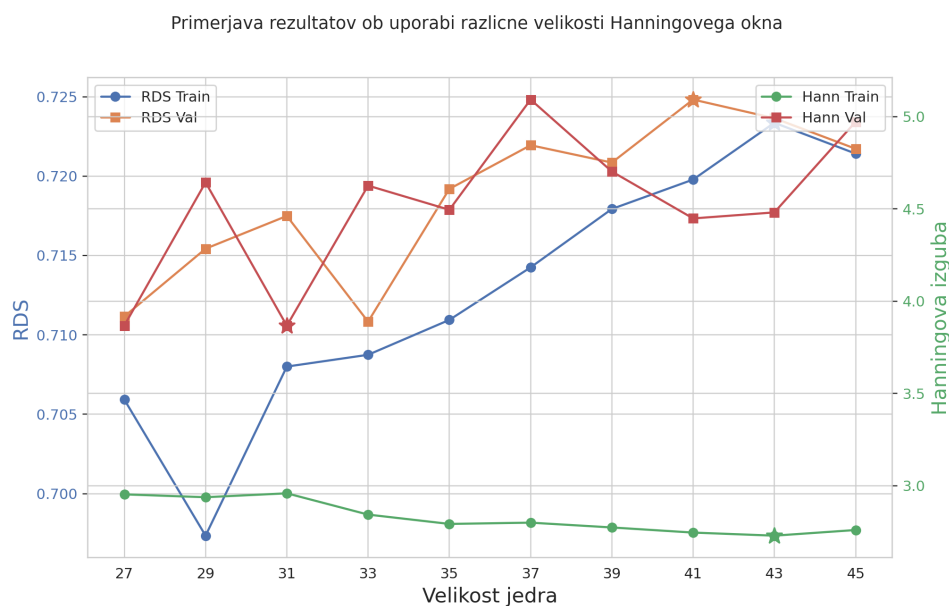
modela. Najboljše uspešnosti so bile dosežene z okni velikosti 31 in 33. Te velikosti sovpadajo s priporočili iz literature, kjer je bila optimalna velikost okna določena na 33 [25].

Čeprav imajo nekatera druga okna boljšo vrednost kriterijske funkcije (vidno v dodatku B), je analiza slik pokazala, da je najmanj šuma prav pri oknih velikosti 31 in 33. Okna, ki imajo manjše ali večje jedro od teh velikosti, začnejo vnašati šum na različnih lokacijah, kar vodi do zmanjšane natančnosti pri lokalizaciji. Ta šum lahko moti interpretacijo satelitskih slik in zmanjša zanesljivost modela.

Zaključimo lahko, da je izbira prave velikosti Hanningovega okna ključna za doseganje optimalnih rezultatov.

4.6 Regularizacija

V tem podpoglavju raziskujemo tehniko izpuščanja nevronov kot sredstvo regularizacije v nevronskih mrežah. Ocenjujemo njen vpliv na model Twins, predstavimo pa tudi, kako različni parametri te tehnike vplivajo na uspešnost



Slika 4.12: Primerjava rezultatov ob uporabi različnih velikosti Hanningovega okna, na celotni validacijski množici.

modela.

4.6.1 Izpuščanje nevronov

V svetu strojnega učenja je regularizacija ključna tehnika, ki se uporablja za preprečevanje prekomernega prilagajanja modela. Prekomerno prilagajanje se pojavi, ko model postane preveč specifičen za učni nabor podatkov, kar pomeni, da se "preveč nauči" podrobnosti in šuma v učnih podatkih, kar vodi v slabo zmogljivost na novih, nevidenih podatkih. Med različnimi tehnikami regularizacije je "izpuščanje nevronov" (ang. dropout) ena izmed najbolj priljubljenih in učinkovitih metod za nevronske mreže. Koncept izpuščanja nevronov je preprost, a močan: med učenjem se določen odstotek nevronov v mreži naključno "izklopi" ali izpusti. To pomeni, da se med posameznim prehodom naprej določeni nevroni (in njihove povezave) začasno odstranijo iz mreže.

V modelu smo uporabili izpuščanje nevronov na več ključnih mestih:

1. **v modelu Twins:** Izpuščanje nevronov je bilo uporabljeno za regulacijo različnih komponent modela, vključno z deli, kot so `attn_drop`, `proj_drop`, `head_drop`, `mlp_drop1`, `mlp_drop2` in `pos_drops`. Vsaka od teh komponent ima svojo specifično vlogo v arhitekturi modela. Z dodajanjem izpuščanja nevronov na te komponente smo dodali dodatno raven regularizacije, ki pomaga preprečiti prekomerno prilagajanje.
2. **v modulu za Združevanje Značilnosti:** Po vsaki konvolucijski operaciji v fuzijskem delu modela smo dodali izpuščanje nevronov. Konvolucijske plasti lahko hitro postanejo kompleksne in se prekomerno prilagodijo podatkom, zlasti ko delujejo na visokodimenzionalnih značilnostih. Z dodajanjem izpuščanja nevronov po vsaki konvolucijski plasti smo zmanjšali to tveganje in povečali robustnost modela.

Izpuščanje nevronov je ena izmed najbolj učinkovitih tehnik regularizacije za nevronske mreže. Z njegovo uporabo v modelu smo zagotovili, da je model bolj robusten in manj nagnjen k prekomernemu prilagajanju na učne podatke. V kompleksnih modelih, kot je Twins, kjer je veliko komponent, ki se lahko prekomerno prilagodijo podatkom, je uporaba izpuščanja nevronov ključnega pomena za zagotavljanje natančnih in zanesljivih rezultatov.

4.6.2 Rezultati

4.7 Uporaba prednaučene mreže

V tem podpoglavju raziskujemo vpliv uporabe prednaučene mreže Twins za izluščenje značilnosti pred združitvijo v modulu za združevanje značilnosti. Cilj je oceniti, kako uporaba prednaučene mreže vpliva na uspešnost modela WAMF-FPI.

Parameter	UAV	Satelit	Združevanje
dropout	0.1	0.1	0.1
attn_drop	0.1	0.1	-
proj_drop	0.1	0.1	-
head_drop	0.1	0.1	-
mlp_drop1	0.1	0.1	-
mlp_drop2	0.1	0.1	-
pos_drops	0.05	0.05	-

Tabela 4.3: Parametri z uravnovanim izpustom nevronov.

Parameter	UAV	Satelit	Združevanje
dropout	0.15	0.05	0.05
attn_drop	0.15	0.05	-
proj_drop	0.15	0.05	-
head_drop	0.15	0.05	-
mlp_drop1	0.15	0.05	-
mlp_drop2	0.15	0.05	-
pos_drops	0.1	0.05	-

Tabela 4.4: Parametri z neuravnovešenim izpuščanjem nevronov.

4.7.1 Prednaučena mreža za izluščenje značilnosti

Uporaba prednaučenih modelov v strojnem učenju omogoča izkoristek že obstoječega znanja modela za pospešitev in izboljšanje učenja na novem naboru podatkov. Zlasti v kontekstu globokih nevronske mreže so prednaučeni modeli dragoceni, saj lahko pomagajo modelom hitreje konvergirati in v nekaterih primerih doseči boljše rezultate.

4.7.2 Rezultati

Naši rezultati kažejo, da je uporaba prednaučene mreže Twins privedla do boljših rezultatov v primerjavi z modelom, ki ni uporabljal prednaučene

Način	HANN _{val}	RDS _{train}	RDS _{val}	$\overline{\Delta}_m$ [m]
Brez izpuščanja	8.49	0.893	0.709	43.42
Uravnovešeno izpuščanje	5.49	0.725	0.690	21.67
Neuravnovešeno izpuščanje	5.42	0.725	0.719	18.11

Tabela 4.5: Rezultati ob uporabi različnih izpuščanj nevronov. Kjer je $\overline{\Delta}_m$ povprečna napaka v metrih.

mreže. To poudarja prednost prenosa znanja iz prednaučenih modelov na specifične naloge.

Način	HANN _{val}	RDS _{train}	RDS _{val}	$\overline{\Delta}_m$ [m]
Prednaučena mreža	8.49	0.893	0.709	43.42
Brez prednaučene mreže	8.21	0.627	0.630	60.23

Tabela 4.6: Rezultati ob uporabi različnih izpustov. Kjer je $\overline{\Delta}_m$ povprečna napaka v metrih.

Uporaba prednaučene mreže Twins je omogočila boljše zajemanje in interpretacijo značilnosti iz našega nabora podatkov, kar je vodilo k izboljšanim rezultatom. To potrjuje, da so prednaučeni modeli lahko zelo koristni v nekaterih scenarijih, še posebej, ko želimo izkoristiti že obstoječe znanje za izboljšanje uspešnosti na novih nalogah.

Poglavje 5

Sklepne ugotovitve

Brepilotni letalniki predstavljajo revolucionarni korak v tehnologiji, ki je našel svojo uporabo v številnih sektorjih, od vojaških operacij do kmetijskega nadzora. Kljub njihovi široki uporabi pa se soočajo z več ključnimi izzivi, zlasti na področju avtonomne navigacije. V diplomski nalogi smo se osredotočili na raziskovanje in implementacijo metode WAMF-FPI za lokalizacijo brepilotnih letalnikov na podlagi slik. Spodaj so izpostavljene naše glavne ugotovitve in predlogi:

1. **Učinkovitost in natančnost metode:** Naša implementacija WAMF-FPI je pokazala, da je metoda obetavna. Ugotovili smo, da je sama arhitektura izredno učinkovita in da lahko zagotovi natančno lokalizacijo brepilotnih letalnikov tudi v zahtevnih pogojih.
2. **Potencial za izboljšave:** Kljub izjemni učinkovitosti metode WAMF-FPI smo identificirali nekaj ključnih področij, kjer bi se lahko izvedle izboljšave. Eden od predlogov je uporaba boljše nevronske mreže za izluščanje značilnosti. Morda bi bile konvolucijske nevronske mreže novejših generacij ali nekatere druge arhitekture bolj primerne za ta namen.
3. **Optimizacija združevanja značilnosti:** Med našo analizo smo opazili, da bi lahko del združevanja značilnosti optimizirali z uporabo me-

tode pozornosti, kar bi omogočilo še boljše ujemanje med slikami zajetimi z brezpilotnim letalnikom in satelitskimi slikami.

4. **Iskanje pripadajoče satelitske slike:** Ena od glavnih težav, s katerimi se metoda še vedno sooča, je identifikacija prave satelitske slike, ki ustreza sliki posneti z brezpilotnim letalnikom. To predstavlja izziv, še posebej v bazah z milijoni slik, in je eno od področij, ki zahteva nadaljnje raziskave. Poskusili smo to implementirati, vendar je na žalost zmanjkalo časa. Kljub temu obstaja potencial za integracijo te metode z obstoječimi senzorji na brezpilotnem letalniku za izboljšanje lokalizacije.
5. **Raziskava različnih kriterijskih funkcij:** V okviru naše analize smo preizkusili več kriterijskih funkcij, vključno s Hanningovo kriterijsko funkcijo, Gaussovo uteženo srednjo kvadratno napako, Hanningovo uteženo srednjo kvadratno napako ter križno uteženo srednjo kvadratno napako. Rezultati so pokazali, da je Hanningova kriterijska funkcija izstopala kot najbolj učinkovita med vsemi preizkušeni. Te ugotovitve so v skladu z implementacijo in rezultati, predstavljenimi v izbranem članku.
6. **Regularizacija in računske obremenitve:** Ugotovili smo, da ima regularizacija v modelu z uporabo izpuščanja nevronov pomembno vlogo pri preprečevanju prenaučenja. Vendar pa je treba skrbno uravnotežiti med računskimi obremenitvami in natančnostjo modela.
7. **Praktična uporaba:** Naša največja ambicija za prihodnost je preizkusiti metodo WAMF-FPI na dejanskem brezpilotnem letalniku. S tem bi lahko dobili boljšo predstavbo o realni učinkovitosti in uporabnosti metode v praksi.

Metoda WAMF-FPI predstavlja pomemben korak naprej v lokalizaciji brezpilotnih letalnikov, še posebej v okoljih, kjer je satelitski signal omejen ali nezanesljiv. Kljub obetavni učinkovitosti metode pa obstajajo še nekateri

izzivi in priložnosti za izboljšave. Naša raziskava je postavila trdne temelje za nadaljnji razvoj in implementacijo metode v realnih sistemih brezpilotnih letalnikov. Naslednji koraki bi vključevali nadaljnje optimizacije modela, razširitev podatkovnih zbirk, uporabo časovne informacije, vzpostavitev povratne zanke (ang. feedback loop) in končno implementacijo na dejanskih brezpilotnih letalnikih.

Dodatek A

Primeri izračuna RDS

Za boljše razumevanje, kako se RDS izračuna in kaj nam predstavlja, si oglejmo tri različne primere.

Primer 1: Za $w = 400px$, $h = 400px$, $dx = 0px$, $dy = 0px$ in $k = 10$ dobimo:

$$RDS_1 = e^{-10 \times \frac{\sqrt{\left(\frac{0}{400}\right)^2 + \left(\frac{0}{400}\right)^2}}{2}} = 1 \quad (\text{A.1})$$

Ker sta dx in dy oba 0, je RDS za ta primer enak 1 (kar pomeni, da je napovedana pozicija točno na dejanski poziciji).

Primer 2: Za $w = 400px$, $h = 400px$, $dx = 2px$, $dy = 0px$ in $k = 10$ dobimo:

$$RDS_2 = e^{-10 \times \frac{\sqrt{\left(\frac{2}{400}\right)^2 + \left(\frac{0}{400}\right)^2}}{2}} = 0.975 \quad (\text{A.2})$$

Tukaj je napovedana pozicija rahlo odmaknjena samo v vodoravni smeri za 2 piksla. $RDS = 0.975$ kaže na minimalno odstopanje napovedane pozicije od dejanske.

Primer 3: Za $w = 400px$, $h = 400px$, $dx = 10px$, $dy = 14px$ in $k = 10$ dobimo:

$$RDS_3 = e^{-10 \times \frac{\sqrt{\left(\frac{10}{400}\right)^2 + \left(\frac{14}{400}\right)^2}}{2}} = 0.806 \quad (\text{A.3})$$

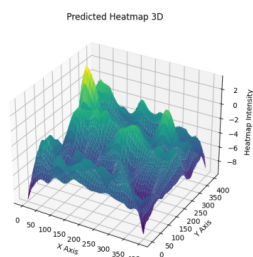
V tem primeru je napovedana pozicija odmaknjena tako v vodoravni kot navpični smeri. *RDS* vrednost 0.806 kaže na večjo relativno napako v primerjavi s prejšnjim primerom.

RDS metrika nam omogoča kvantitativno oceno natančnosti napovedane pozicije v primerjavi z dejansko pozicijo. Višja kot je vrednost *RDS*, bližje je napovedana točka dejanski točki. V obratnem primeru, nižja kot je vrednost *RDS*, večja je napaka med napovedano in dejansko točko.

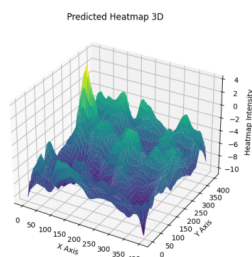
Dodatek B

Primerjava toplotnih kart

Velikost hanningovega okna: 27, RDS: 0.83

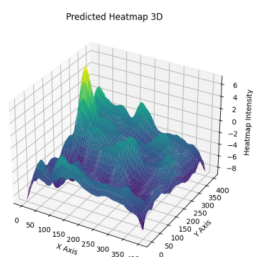


Velikost hanningovega okna: 29, RDS: 0.83

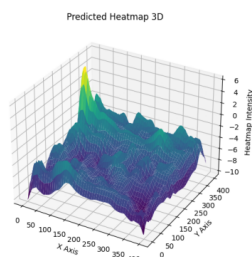


Slika B.1: Primerjava toplotnih map ob uporabi velikosti 27 in 29 Hanningovega okna.

Velikost hanningovega okna: 31, RDS: 0.84

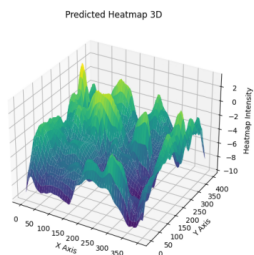


Velikost hanningovega okna: 33, RDS: 0.81

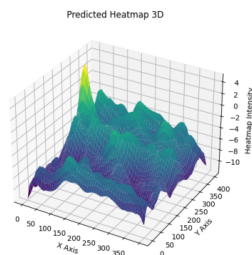


Slika B.2: Primerjava toplotnih map ob uporabi velikosti 31 in 33 Hanningovega okna.

Velikost hanningovega okna: 35, RDS: 0.85

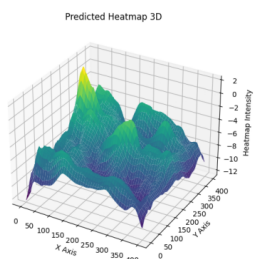


Velikost hanningovega okna: 37, RDS: 0.81

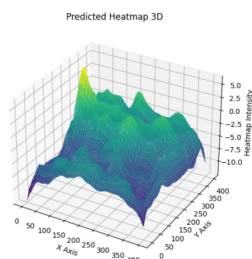


Slika B.3: Primerjava toplotnih map ob uporabi velikosti 35 in 37 Hanningovega okna.

Velikost hanningovega okna: 39, RDS: 0.80

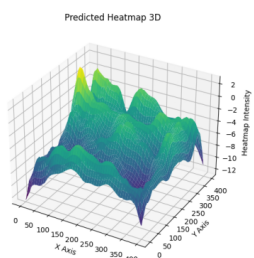


Velikost hanningovega okna: 41, RDS: 0.81

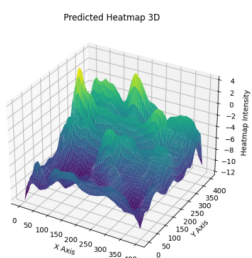


Slika B.4: Primerjava toplotnih map ob uporabi velikosti 39 in 41 Hanningovega okna.

Velikost hanningovega okna: 43, RDS: 0.81



Velikost hanningovega okna: 45, RDS: 0.79



Slika B.5: Primerjava toplotnih map ob uporabi velikosti 43 in 45 Hanningovega okna.

Literatura

- [1] Dzmitry Bahdanau, Kyunghyun Cho in Yoshua Bengio. “Neural Machine Translation by Jointly Learning to Align and Translate”. V: *arXiv preprint arXiv:1409.0473* (2015).
- [2] Mollie Bianchi in Timothy D. Barfoot. “UAV Localization Using Autoencoded Satellite Images”. V: *arXiv preprint arXiv:2102.05692* (2021). URL: <http://arxiv.org/abs/2102.05692v1>.
- [3] Xiangxiang Chu in sod. “Conditional positional encodings for vision transformers”. V: *arXiv preprint arXiv:2102.10882* (2021).
- [4] Xiangxiang Chu in sod. “Twins: Revisiting the design of spatial attention in vision transformers”. V: *Advances in Neural Information Processing Systems* 34 (2021), str. 9355–9366.
- [5] Ming Dai in sod. “Finding Point with Image: An End-to-End Benchmark for Vision-based UAV Localization”. V: *arXiv preprint arXiv:2208.06561* (2022).
- [6] Sounak Dey in sod. “SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification”. V: *arXiv preprint arXiv:1707.02131* (2017). URL: <http://arxiv.org/abs/1707.02131v2>.
- [7] Alexey Dosovitskiy in sod. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. V: *arXiv preprint arXiv:2010.11929* (2020).

- [8] A. Zamir F. Castaldo in sod. “Semantic cross-view matching”. V: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)* (2015). DOI: <https://doi.org/10.1109/iccvw.2015.137>.
- [9] Google. *Google Earth Studio*. Dostopano: 30.8.2023. 2021. URL: <https://www.google.com/earth/studio/>.
- [10] Liu Liu in Hongdong Li. “Lending Orientation to Neural Networks for Cross-view Geo-localization”. V: *arXiv preprint arXiv:1903.12351* (2019). URL: <http://arxiv.org/abs/1903.12351v1>.
- [11] Liu Liu in Hongdong Li. “Predicting Ground-Level Scene Layout from Aerial Imagery”. V: *arXiv preprint arXiv:1612.02709* (2016). URL: <http://arxiv.org/abs/1612.02709v1>.
- [12] Ze Liu in sod. “Swin transformer: Hierarchical vision transformer using shifted windows”. V: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, str. 10012–10022.
- [13] H. S. Sawhney M. Bansal in sod. “Geo-localization of street views with aerial image databases”. V: *Proceedings of the 19th ACM international conference on Multimedia - MM '11* (2011). DOI: <https://doi.org/10.1145/2072298.2071954>.
- [14] Mapbox. *Mapbox API Dokumentacija*. Dostopano: 30.8.2023. 2021. URL: <https://www.mapbox.com/api-documentation/>.
- [15] Python. *Python: Programski jezik*. Dostopano: 5.9.2023. 2021. URL: <https://www.python.org/>.
- [16] PyTorch. *PyTorch: Odprtokodni knjižnica za strojno učenje*. Dostopano: 30.8.2023. 2021. URL: <https://pytorch.org/>.
- [17] M. Feng S. Hu in sod. “CVM-net: Cross-view matching network for image-based ground-to-aerial geo-localization”. V: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018). DOI: <https://doi.org/10.1109/cvpr.2018.00758>.

- [18] R. Souvenir S. Workman in N. Jacobs. “Wide-area image geolocalization with aerial reference imagery”. V: *2015 IEEE International Conference on Computer Vision (ICCV)* (2015). DOI: <https://doi.org/10.1109/cvpr.2005.202>.
- [19] T. Senlet in A. Elgammal. “A framework for global vehicle localization using stereo images and satellite and road maps”. V: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (2011). DOI: <https://doi.org/10.1109/iccvw.2011.6130498>.
- [20] S. Belongie T.-Y. Lin in J. Hays. “Cross-view image geolocalization”. V: *2013 IEEE Conference on Computer Vision and Pattern Recognition* (2013). DOI: <https://doi.org/10.1109/cvpr.2013.120>.
- [21] Y. Cui T.-Y. Lin in sod. “Learning deep representations for ground-to-aerial geolocalization”. V: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015). DOI: <https://doi.org/10.1109/cvpr.2015.7299135>.
- [22] *Teacher forcing*. https://en.wikipedia.org/wiki/Teacher_forcing. Dostopano: 30.8.2023. 2023.
- [23] Hugo Touvron in sod. “Training data-efficient image transformers & distillation through attention”. V: *arXiv preprint arXiv:2012.12877* (2020). URL: <http://arxiv.org/abs/2012.12877v2>.
- [24] Ashish Vaswani in sod. “Attention is all you need”. V: *Advances in neural information processing systems* 30 (2017).
- [25] Guirong Wang in sod. “WAMF-FPI: A Weight-Adaptive Multi-Feature Fusion Network for UAV Localization”. V: *Remote Sensing* 15.4 (2023), str. 910.
- [26] Wenhai Wang in sod. “Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions”. V: *arXiv preprint arXiv:2102.12122* (2021).

-
- [27] S. Workman in N. Jacobs. “On the location dependence of convolutional neural network features”. V: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2015). DOI: <https://doi.org/10.1109/cvprw.2015.7301385>.
- [28] X. Yu Y. Shi in sod. “Optimal feature transport for cross-view image geo-localization”. V: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.07 (2020), str. 11 990–11 997. DOI: <https://doi.org/10.1609/aaai.v34i07.6875>.
- [29] Zhedong Zheng in sod. “University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization”. V: *arXiv preprint arXiv:2002.12186* (2020). URL: <http://arxiv.org/abs/2002.12186v1>.
- [30] Sijie Zhu, Mubarak Shah in Chen Chen. “TransGeo: Transformer Is All You Need for Cross-view Image Geo-localization”. V: *arXiv preprint arXiv:2204.00097* (2022). URL: <http://arxiv.org/abs/2204.00097v1>.