

ATTĒLA SEGMENTĀCIJAS PRECIZITĀTE ATKARĪBĀ NO U-NET MODEĻA DZIĻUMA

IMAGE SEGMENTATION ACCURACY DEPENDING ON THE DEPTH OF U-NET MODEL

Autors: **Jevgēnijs RIEKSTIŅŠ**, e-pasts: jevgenijsriekstins@gmail.com
Zinātniskā darba vadītājs: Dr.sc.ing., docents **Sergejs KODORS**,
Rēzeknes Tehnoloģiju akadēmija,
Atbrīvošanas aleja 115, Rēzekne, Latvija

Abstract. *The aim of this work is to obtain information about impact of the depth of U-Net architecture model into segmentation accuracy. Experiment was completed using dataset of DSM images. Neural networks were trained to recognize building locations. Experiment considered to decrease the number of U-Net filter blokes to measure impact on result accuracy.*

Keywords: *accuracy, image segmentation, machine learning, neural network.*

Ievads

Attēla atpazīšana cilvēkiem un dzīvniekiem ir dabiska, taču datoriem tas ir ārkārtīgi grūts uzdevums.

Cilvēka acs redz attēlu kā signālu kopu, ko interpretē smadzeņu redzes garoza. Rezultāts ir ainas pieredze, kas saistīta ar objektiem un jēdzieniem, kas saglabāti atmiņā. Mašīnredze (*computer vision*) atdarina šo procesu. Datori "redz" attēlu kā vektoru kopu (ar krāsu apzīmējumiem daudzstūrī) vai kā matricu (rastra attēli ar diskrētām krāsu vērtībām).

Izstrādājot sistēmu ar neironu tīklu, tā tiek apmācīta ar iepriekš sagatavotu mācību materiālu, lai izpildītu noteiktu uzdevumu. Piemēram, attēlu atpazīšanas neironu tīkls var iemācīties identificēt attēlus, kas satur ēkas, analizējot attēlus, kas ir manuāli anotēti ar vērtībām "ir ēka" vai "nav ēkas".

Apmācītais neirontīkls spēj identificēt ēkas arī citos attēlos ar pietiekoši augstu precizitāti. Mākslīgie neironu tīkli to dara bez jebkādam iepriekšējam zināšanām par ēkām, piemēram, par to, no kāda materiāla tie ir uzcelti, ka tiem var būt dažāds stāvu skaits, utt. Tā vietā, neironu tīkli automātiski ģenerē identificējošas pazīmes no mācību materiāla, ko tie apstrādā (no statistikas). Jo vairāk slāņu tiek izmantots, jo vairāk īpašību var iegaumēt neironu tīkls.

Neironu tīkla attēlu atpazīšanas algoritmi balstās uz datu kopas kvalitātes - attēliem, ko izmanto modeļa apmācībai un pārbaudei. Šeit ir daži svarīgi datu kopas sagatavošanas parametri un apsvērumi.

- Attēla lielums - augstākas kvalitātes attēls modelim sniedz vairāk informācijas, bet apstrādei ir nepieciešams vairāk neironu un vairāk skaitļošanas jaudas.
- Attēlu skaits - jo lielāka datu kopa, jo precīzāks būs apmācīts neironu tīkls, bet ir nepieciešams pārliecināties, ka apmācību kopa atspoguļo reālo populāciju.
- Kanālu skaits – pelēktoņu attēlam ir 1 kanāls (melnbalts) un krāsainiem attēliem parasti ir 3 krāsu kanāli (*RGB* krāsu modelis: sarkana, zaļa un zila krāsa).
- Attēlu izmērs – nepieciešams pārliecināties, ka attēliem ir vienādi izmēri. Parasti neironu tīkla modeļos tiek pielietoti attēli ar kvadrāta formu. [3]

Mūsdienās mašīnmācība kļuva populāra pateicoties gataviem rīkiem kā *TensorFlow*, *Keras*, *Caffe*, u.c., kas piedāvā gatavus blokus neirontīklu konstruēšanai un apmācībai bez īpašām iepriekšējām zināšanām vai matemātisko bāzi. Tagad, daudzās nozarēs sāk pielietot mašīnmācību automatizēšanas un datu analīzes uzdevumiem. Katru gadu tirgū parādās vairāk un vairāk viedo risinājumu balstīti uz mašīnmācības pielietošanas, piemēram, inteliģents palīgs kā *Google Assistant* vai valkājami fitnesa izsekotāji kā *Fitbit*, u.c.

Mašīnmācību var pielietot šādiem uzdevumiem:

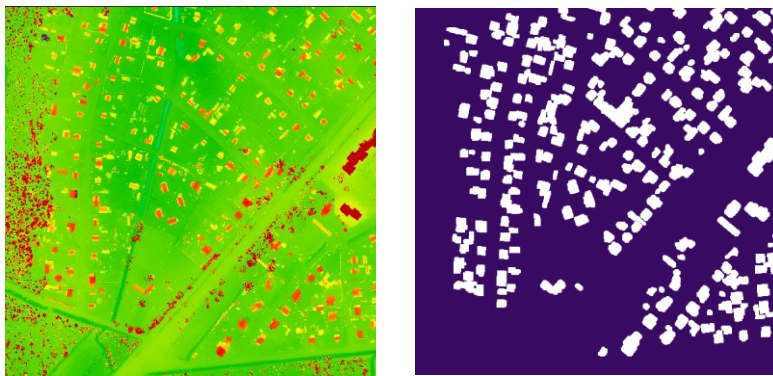
- Prognozēšana: kad pieejami vēsturiski dati tiek pielietoti prognozējot nākamus notikumus vai vērtības. Lai izpildītu prognozi ilgākam laika posmam, vēsturiski dati tiek papildināti ar prognozes rezultātiem un algoritms tiek atkārtots. [1]
- Attēla atpazīšana – piemēram, sejas atpazīšana attēlā vai video. Šim mērķim jā sagatavo datu bāze, kur katra persona veido atsevišķu kategoriju.
- Runas atpazīšana – kad audio ieraksts tiek pārveidots teksta formātā. To izmanto meklēšanas uzdevumiem, telefona numuru ievadei vai ierīces vadībai, kad komandas tiek atdotas ar balss palīdzību. To var izmantot arī datu ievadīšanas un dokumentu sagatavošanas uzdevumos.
- Medicīniskās diagnozes – piemēram, eksistē risinājumi, kas tiek pielietoti vēža audu atpazīšanai. [2]

Mašīnmācības sistēmu pielietošanas pamatā ir precizitāte, ar kādu šīs sistēmas spēj veikt darbu.

Pētījuma mērķis ir izpētīt kā *U-Net* neironu tīkla modeļa dziļums ietekmē uz segmentācijas precizitāti.

Materiāli un metodes

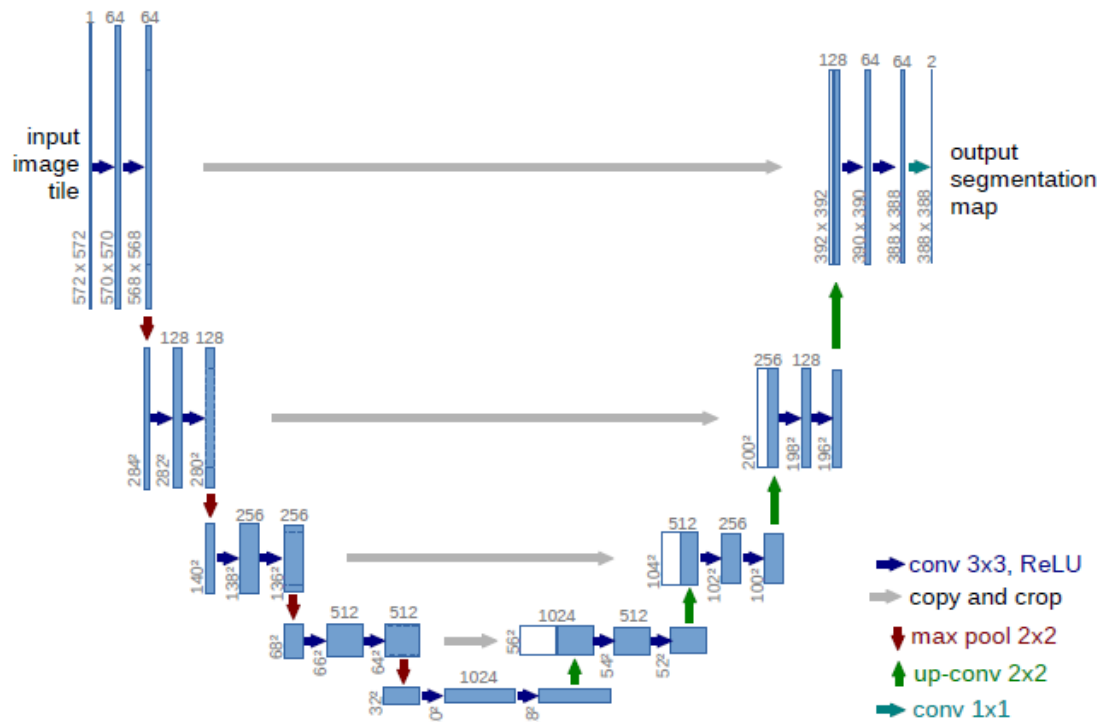
Datu kopa: izmantojot neironu tīklu ar *U-Net* arhitektūru, uzdevums ir apmācīt to atpazīt būves, pielietojot zemes virsmas attēlus (*DSM images*, skat. 1. attēlu) ar izmēru 572x572. Datu kopa sastāv no 160 attēliem, kur katram attēlam ir dota gaidāma izvades maska (skat. 1. attēlu). Lai novērtētu neironu tīkla precizitāti, tiek pielietoti papildus 40 attēli.



1. attēls. Datu kopas parauga piemērs: pa kreisi *DSM* attēls, pa labi – gaidāma maska

Mašīnmācības inženieri uzskata, ka sekmīgai neironu tīklu apmācībai ir nepieciešami daudzi tūkstoši anotētu mācību paraugi. Lai samazinātu attēlu skaitu, tiek pielietota papildināšanas metode (*augmentation*). Šajā eksperimentā tiek pielietota papildināšanas metode, kad attēli tiek pagriezti 90° un katram attēlam tiek iegūts spoguļa attēls.

***U-Net* arhitektūra** (skat. 2. attēlu) tiek izmantota, lai izpildītu attēlu semantisku segmentāciju un anotāciju. Arhitektūra sastāv no saslēgta ceļa priekš konteksta uztveršanas un simetriski paplašināma ceļa, kas nodrošina objektu lokalizāciju.



2. attēls. *U-Net* arhitektūra [5]

Arhitektūrā ir divi ceļi (skat. 2. attēlu). Pirmais ceļš ir kodēšanas ceļš, ko izmanto konteksta uztveršanai attēlā. Kodētājs sastāv no konvolūcijas slāņiem (*convolution layer*) un maksimālā signāla izvēles slāņa (*max pooling*). Otrais ceļš ir simetriski paplašinošais ceļš (saukts arī par dekodētāju), ko izmanto, lai nodrošinātu precīzu objektu lokalizāciju attēlā, izmantojot dekonvolūcijas slāņus (*deconvolution layer*). Katrs konvolūciju slānis (zila rūtiņa 2. attēlā) satur vairākus filtrus (*filters*). Filtru skaits ir norādīts virs slāņa. Apstrādājama attēla izmērs ir norādīts kreisajā pusē. Baltās rūtiņas apzīmē kā izvades matricas tiek apvienotas. Bultas apzīmē dažādas operācijas.

Kodētāja ceļš seko tipiskai konvolūciju tīkla arhitektūrai. Tas sastāv no vairākiem blokiem ar diviem konvolūcijas slāņiem ar kodola logu 3x3 (*kernel window*), kuriem seko *ReLU* aktivācijas slāņi un maksimāla signāla izvēles slāņi ar logu 2x2. Katra nākamā blokā filtru skaits tiek dubultots. Katrs dekodētāja ceļa posms sastāv no blokiem ar 2x2 dekonvolūcijas slāņiem un *ReLU* slāņiem. Katram dekodētāja blokam filtru skaits tiek samazināts.

Rīki: Neironu tīkla apmācība notiek, izmantojot *TensorFlow Keras* un programmēšanas valodu *Python*. Lai atbalstītu *Python* programmēšanas valodu, tiek pielietota izplatīšanas programma *Anaconda*, kas paredzēta veikt skaitļošanu zinātniskiem mērķiem (datu zinātne, mašīnmācība, utt.). Papildus tiek izmantots *Jupyter Notebook*, kas izstrādāts, lai atbalstītu interaktīvo datu zinātņu un zinātnisko skaitļošanu.

Eksperiments sastāv no trīs posmiem:

- 1) 10 reizes apmācīt neironu tīklu ar 5 *U-Net* slāņu blokiem;
- 2) 10 reizes – ar 4 blokiem;
- 3) 10 reizes – ar 3 blokiem, lai novērtētu precizitātes kritumu atkarībā no neirontīkla dziļuma.

Rezultāti un diskusija

Kopumā eksperiments tika izpildīts 30 reizes, katru reizi mainot apmācīšanas slāņu skaitu. Pirms eksperimenta izpildīšanas pašā neironu tīklā tika nokonfigurēti attēla parametri (skat. 3. attēlu). Tiek izmantoti kvadrātiski attēli ar malas garumu 572 pikseļi. Ievades attēli ir melnbalti un sastāv no 2 kanāliem (*bands*). Viens kanāls satur virszemes normalizētu virszemes

augstumu pēc minimāla un maksimāla punkta attēla, otrais kanāls satur informāciju par tukšiem pikseliem.

```
# Net parameters

image_size = 572
output_size = 388
image_channels = 2
```

3. attēls. Attēla parametru iestatīšana

Neironu tīkla apmācība tika veikta, izmantojot grafisko procesoru (*GPU*), jo tas spējīgs apstrādāt neirontīklus gandrīz 10 reizēs ātrāk nekā centrālais procesors. 4. attēlā ir doti *GPU* parametri.

```
# Device configuration, GPU is used automatically
print(device_lib.list_local_devices())

[name: "/device:CPU:0"
 device_type: "CPU"
 memory_limit: 268435456
 locality {
 }
 incarnation: 11095711417240101368
 , name: "/device:GPU:0"
 device_type: "GPU"
 memory_limit: 2621100032
 locality {
   bus_id: 1
   links {
 }
 }
 incarnation: 6496245097394903381
 physical_device_desc: "device: 0, name: GeForce GTX 960, pci bus id: 0000:01:00.0, compute capability: 5.2"
]
```

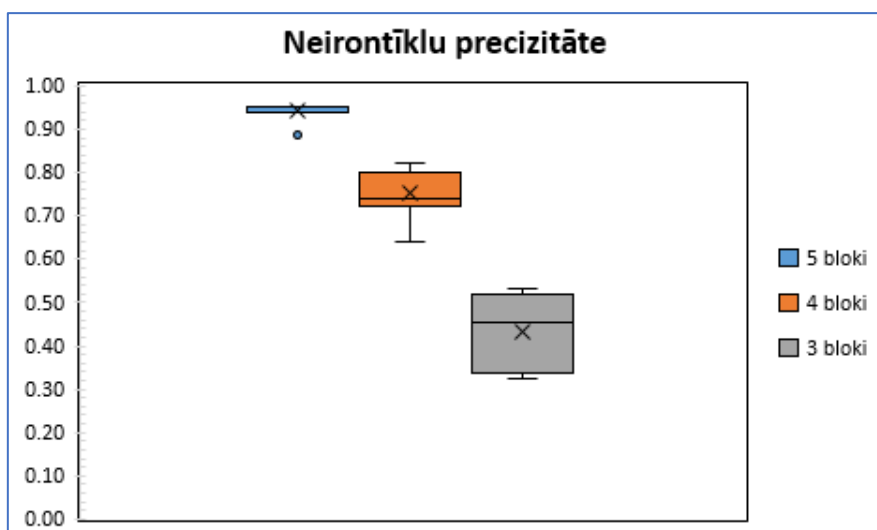
4. attēls. *GPU* parametri

Attēla atpazīšanas precizitātei jābūt pēc iespējas lielākai, tas ir tieši atkarīgs no neironu tīkla apmācīšanas procesa. Katrs neirona tīkla apmācības eksperiments sastāv no 160 posmiem (*epoch*), kad neirontīkls tiek apmācīts un tā precizitāte tiek pārbaudīta, aprēķinot vidēju precizitāti no 40 testa attēliem. Izpildot eksperimentu var secināt, ka neirontīkla dziļums jeb bloku skaits ietekmē uz rezultāta precizitāti (skat. 1. tabulu un 5. attēlu).

1. tabula

Precizitāte atkarībā no *U-Net* arhitektūras bloku skaita

Nr.	5 bloki	4 bloki	3 bloki
1	0,8885	0,7984	0,4841
2	0,9506	0,7226	0,5265
3	0,9406	0,7404	0,3983
4	0,9504	0,7256	0,532
5	0,9519	0,6403	0,3258
6	0,9508	0,8220	0,4531
7	0,9506	0,7885	0,5078
8	0,9514	0,7342	0,3494
9	0,9408	0,7987	0,3284
10	0,9507	0,7762	0,4282
VID	0,9427	0,7547	0,4334



5. attēls. Blokdigramma ar neirontīklu precizitātes

Var redzēt, ka precizitāte samazinās, samazinot neirona tīkla dziļumu (slāņu bloku skaitu). Samazinot bloku skaitu, samazinājās arī kopējais neironu tīkla apmācības laiks. Apmācot neirontīklu ar pieciem blokiem, katrs apmācības posms aizņēma 10 sekundes, tāpēc kopējais laiks, lai apmācītu vienu neirontīkla modeli, bija nepieciešamas 1600 sekundes jeb apmēram 26,6 minūtes. Apmācot, neirontīklu ar četriem blokiem, katrs apmācības posms aizņēma 9 sekundes, tāpēc kopējais laiks - 1440 sekundes jeb 24 minūtes. Neirontīkls ar trīs blokiem – 8 sekundes, kopējais laiks – 1280 sekundes jeb ~21,3 minūtes.

Secinājumi

Attēla atpazīšanas precizitāte ir svarīgs parametrs, kas noteic to praktisko pielietojumu. Eksperimentā bija noteikts, ka lielāks bloku skaits (lielāks neirontīkla dziļums) nodrošina labāku atpazīšanas precizitāti, bet tas pieprasa vairāk laika, lai apmācītu neironu tīklu. Tā kā pētījuma eksperimentā tika izmantoti attēli ar mazu izšķirtspēju (572x572) un 2 kanāliem, apmācības process bija diezgan ātrs, bet ja tiktu izmantoti attēli ar lielāko izšķirtspēju vai tiktu pielietotas vairāki krāsu kanāli, apmācības process aizņemtu daudz vairāk laika. Var redzēt, ka, izmantojot trīs *U-Net* slāņu blokus, vidēja precizitāte ir 43%, ar to var secināt, ka izmantot mazāku bloku skaitu nav lietderīgi, jo iegūtais precizitātes rezultāts būs ļoti neprecīzs. Izmantot vairāk par pieciem blokiem šī uzdevuma gadījumā arī nav lietderīgi, jo būs lieks resursu patēriņš nekā nepieciešams.

Summary

Accuracy is the most important parameter in the machine learning to identify practical application of neural network. It can be seen that the largest number of layers provides greater image recognition accuracy, but it takes more time to train the neural network. Because the study experiment used low-resolution images (572x572) and only two bands, the learning process was quite fast, but if the highest-resolution images were used or many bands were used, the learning process would take much longer. It can be seen that using three filter blocks of U-Net architecture, the average accuracy is 43%, that means that it will not make sense to use a smaller number of filter blocks, because the obtained accuracy result will be very inaccurate. Additionally, it can be concluded that the use of more than five filter blocks is not necessary for this task, it will be an unnecessary consumption of resources.

Literatūra

1. Forecasting: Principles and Practice [Tiešsaite]
Pieejams: <https://otexts.com/fpp2/nnetar.html> [Piekļuve 15.04.2020]
2. An Introduction to Machine Learning by Anmol Behl [Tiešsaite]
Pieejams: <https://becominghuman.ai/an-introduction-to-machine-learning-33a1b5d3a560> [Piekļuve 15.04.2020]
3. Neural Networks for Image Recognition: Methods, Best Practices, Applications, [Tiešsaite]
Pieejams: <https://missinglink.ai/guides/computer-vision/neural-networks-image-recognition-methods-best-practices-applications/> [Piekļuve 15.04.2020]
4. Understanding Semantic Segmentation with UNET
Pieejams: <https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47> [Piekļuve 17.04.2020]
5. U-Net: Convolutional Networks for Biomedical Image Segmentation by Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Computer Science Department and BIOS Centre for Biological Signalling Studies, University of Freiburg, Germany, 2015.