

# Efikasnost modela dubokog učenja za segmentaciju retinalnih krvnih sudova pri ograničenim resursima: Eksperimentalna studija na DRIVE skupu

Jovana Mitrović

Fakultet tehničkih nauka, Univerzitet u Novom Sadu

mitrovicj1012@gmail.com

**Apstrakt**—U ovom radu se istražuje efikasnost različitih strategija obučavanja modela za segmentaciju krvnih sudova na retini, u uslovima ograničenih podataka i resursa. Eksperimentima na DRIVE skupu analizirani su uticaji hiperparametara, arhitektonskih modifikacija U-Net modela, upotreba SegFormer-B0, kao i efekti augmentacija i maskiranja. Evaluacija se bazira na metrikama Dice, IoU, preciznosti i odzivu, ograničenim na vidno polje. Rezultati pokazuju da SegFormer nadmašuje U-Net u kvantitativnim metrikama i stabilnosti performansi, ali uz gubitak vaskularnog kontinuiteta. U-Net bolje čuva morfološku cjelovitost vaskularnih struktura. Zaključak je da se izbor modela mora, pored standardnih metrika, temeljiti i na specifičnim kliničkim zahtjevima.

**Ključne riječi**—segmentacija krvnih sudova, U-Net, SegFormer, DRIVE skup, augmentacija podataka, FOV maskiranje, ograničeni resursi

## I. UVOD

Automatska segmentacija retinalnih krvnih sudova ima ključnu ulogu u savremenoj oftalmologiji jer omogućava rano otkrivanje vaskularnih promjena povezanih sa sistemskim oboljenjima poput dijabetesa i hipertenzije. Ručna segmentacija je vremenski zahtjevna i subjektivna, te automatski pristupi omogućavaju brzu, pouzdanu i reproducibilnu analizu velikog broja snimaka. Precizna segmentacija je i preduslov za dalje analize, poput mjerjenja širine sudova i detekcije bifurkacija, što čini ovaj zadatak centralnim za računarsku dijagnostiku u oftalmologiji [2].

Ovaj zadatak je tehnički izazovan: krvni sudovi su često široki svega 1–2 piksela i razgranati u kompleksne mreže, snimci su niskog kontrasta sa varijabilnim osvjetljenjem, a patološke promjene mogu zaklanjati relevantne strukture. Dodatno, izrazit disbalans klasa otežava obučavanje modela i vodi ka pristrasnosti prema pozadinskoj klasi [3].

Dosadašnja istraživanja su pokazala značajan napredak u tačnosti segmentacije, ali većina njih prepostavlja dostupnost snažnog hardvera i velikih skupova podataka, što ograničava njihovu primjenu u obrazovnim i istraživačkim institucijama sa skromnim resursima. Ovaj rad se bavi navedenim jazom u literaturi kroz eksperimentalnu studiju efikasnosti različitih arhitektura u uslovima ograničenih resursa.

### A. Struktura rada

Rad je organizovan na sljedeći način: U drugom poglavlju pruža se pregled relevantne literature, sa posebnim fokusom

na kritičku analizu ograničenja postojećih pristupa u kontekstu ograničenih resursa. Treće poglavlje detaljno opisuje DRIVE skup podataka, tehnike predobrade i eksperimentalnu metodologiju, uključujući strategiju podjele podataka i kriterijume evaluacije. Četvrto poglavlje predstavlja arhitekture U-Net i SegFormer modela, sa detaljnim opisom implementacijskih modifikacija i strategija obuke prilagođenih ograničenim resursima. Peto poglavlje sadrži kvantitativne i kvalitativne rezultate eksperimenata, uključujući analizu dinamike obuke, varijabilnosti performansi i uticaja različitih konfiguracija. Šesto poglavlje donosi diskusiju rezultata sa razmatranjem ograničenja istraživanja i preporukama za praktičnu primjenu. Konačno, sedmo poglavlje sumira ključne zaključke i doprinos istraživanja u kontekstu efikasne implementacije modela za segmentaciju u ograničenim uslovima.

## II. PREGLED LITERATURE

Segmentacija krvnih sudova na retini je aktivno područje istraživanja sa značajnim napretkom u posljednjoj deceniji. Pristupi se mogu podijeliti u tri glavne kategorije: tradicionalne metode, pristupi zasnovani na konvolutivnim neuron-skim mrežama i novije transformerske arhitekture.

### A. Tradicionalni pristupi

Raniji radovi u segmentaciji retinalnih krvnih sudova oslanjali su se na tehnike obrade slike poput filtriranja, morfoloških operacija i algoritama praćenja putanje [2]. Staal i saradnici su razvili pristup zasnovan na detekciji grebena (ridge detection), koji je dugo bio referantan za DRIVE skup podataka. Iako su ovi pristupi postizali razumne rezultate, imali su ograničenu prilagodljivost: osjetljivi su na varijabilno osvjetljenje, patološke promjene i šum, te često proizvode fragmentovane segmentacije bez kontinuiteta tankih krvnih sudova.

### B. Pristupi zasnovani na konvolutivnim neuron-skim mrežama

Uvođenje dubokog učenja je označilo revoluciju u polju medicinske segmentacije, pri čemu je U-Net arhitektura [3] postala standard, zahvaljujući koder-dekoder strukturi sa preskočnim vezama koje omogućavaju kombinovanje detalja niskog nivoa sa semantičkim informacijama visokog nivoa.

Mnoge varijante U-Net-a razvijene su specifično za retinalnu segmentaciju. Na primjer, R2U-Net [8] uči rekurentne blokove radi boljeg modelovanja sekvencijalnih zavisnosti,

dok Attention U-Net [9] koristi mehanizme pažnje za fokusiranje na relevantne regije slike. Dodatne modifikacije, kao što su DU-Net [13], U-Net++ [15] i MultiResU-Net [16], su unaprijedile performanse kroz različite strukturne inovacije. Ipak, arhitekture konvolutivnih neuronskih mreža imaju relativno lokalni receptivni prostor, što otežava modelovanje dugoročnih zavisnosti, a dublje varijante postaju zahtjevne za obuku na malim medicinskim skupovima i sa ograničenim hardverom.

### C. Transformerski pristupi

Sa pojavom Vision Transformer-a (ViT) [5], otvorene su nove mogućnosti za globalno modelovanje konteksta. U medicinskoj segmentaciji, SegFormer [4], TransU-Net [10], Swin-U-Net [11] i FAT-Net [14] su pokazali superiore rezultate zahvaljujući self-attention mehanizmima i hijerarhijskim strukturama. Prednost transformera je što svaki piksel može „vidjeti“ cijelu sliku, ali ova arhitektura je znatno zahtjevija: kvadratna složenost mehanizma pažnje, potreba za velikim količinama podataka i često neadekvatna prilagodljivost prethodno obučenim modelima otežavaju njihovu primjenu u ograničenim uslovima.

### D. Izazovi u praktičnoj primjeni

Većina postojećih radova fokusira se na maksimizaciju performansi na standardnim referentnim skupovima (DRIVE, STARE, CHASE\_DB1), uz korištenje snažnih GPU sistema i dugotrajne obuke [4], [9]. Ova praksa donosi nekoliko ključnih pitanja. Prvo, optimizacija hiperparametara pri malim veličinama paketa predstavlja izazov jer mala veličina paketa povećava varijansu gradijenata, što otežava stabilnu i brzu konvergenciju modela. Zatim, efikasnost transformer arhitektura na malim medicinskim skupovima podataka nije dovoljno istražena, te nije jasno da li se prednosti globalnog konteksta u ovim modelima zaista održavaju u uslovima ograničenih podataka i hardverskih resursa. Dalje, postoji problem balansiranja između kvantitativnih metrika kao što su Dice i IoU i stvarne kliničke korisnosti; fokus na ove standardne metrike može zanemariti morfološku konzistentnost segmentacije krvnih sudova [7].

### E. Pozicioniranje ovog istraživanja

Za razliku od literature koja prepostavlja optimalne uslove obuke, ovaj rad ispituje praktične aspekte implementacije konvolutivnih neuronskih mreža i transformerskih arhitektura u edukativnim i istraživačkim okruženjima sa strogo ograničenim resursima. Analizira se kompromis između performansi, stabilnosti obuke i morfološke konzistentnosti, da bi se doprinijelo boljem razumijevanju efikasnih strategija segmentacije u realnim uslovima.

## III. MATERIJAL I METODOLOGIJA

### A. Skup podataka

DRIVE (Digital Retinal Images for Vessel Extraction) se sastoji od 40 slika, u koje spada 7 slučajeva abnormalne patologije.

Svaka slika ima dimenzije 584×565 piksela u RGB formatu. Originalna podjela od 20 slika za obuku i 20 za testiranje je reorganizovana na 30 slika za treniranje, 5 za validaciju i 5 za testiranje, radi efikasnije upotrebe podataka.

Jedna od odlika DRIVE skupa je izražen disbalans klasa, budući da krvni sudovi zauzimaju svega oko 10% ukupnih piksela. Ovaj disbalans uslovljava pažljiv izbor funkcije greške i metoda evaluacije.

### B. Predobrada i augmentacija

Za poboljšanje sposobnosti generalizacije modela u uslovima ograničenih podataka primijenjen je skup augmentacionih transformacija korišćenjem Albumentations biblioteke [6]:

- Normalize: Normalizacija na ImageNet statistike
- Resize: Promjena dimenzija na 256×256 za U-Net, odnosno 512×512 za SegFormer
- HorizontalFlip, VerticalFlip: Proizvoljna horizontalna i vertikalna rotacija
- RandomRotate90: Rotacije za 90°, 180° i 270°
- ShiftScaleRotate: Male translacije, skaliranja i rotacije do ±15°
- RandomBrightnessContrast: Nasumične promjene u osvjetljenju i kontrastu
- GridDistortion: Nelinearne geometrijske deformacije

### C. FOV maskiranje

FOV (Field-of-View) maska definiše kružno područje mrežnjače obuhvaćeno snimkom. Pikseli van ove oblasti sadrže šum i regione sa malo informacija. Sve funkcije greške i evaluacione metrike su ograničene isključivo na FOV regiju, čime se eliminiše uticaj irelevantnih piksela.

### D. Tehnička ograničenja okruženja

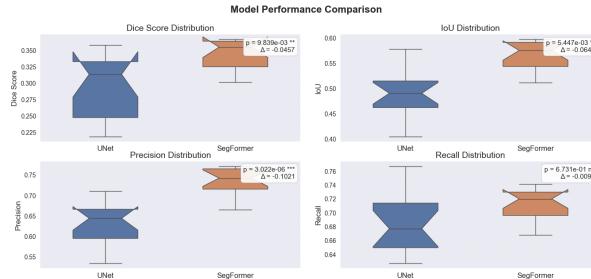
Važno je naglasiti da je ovo istraživanje sprovedeno kao edukativni eksperiment u okviru besplatnog Google Colab okruženja, sa ciljem razumijevanja uticaja različitih modifikacija na proces obuke. Ograničenja u vidu kratkog vremena izvršavanja sesija, ograničene GPU memorije i periodičnih prekida rada značajno utiču na ostvarene rezultate - stoga se rezultati moraju tumačiti u kontekstu ovih ograničenja.

## IV. EKSPERIMENTALNA POSTAVKA

### A. Dizajn eksperimenta

Istraživanje je sprovedeno kroz tri glavna eksperimentalna bloka.

1) *Optimizacija hiperparametara:* Za oba modela (U-Net [3] i SegFormer-B0 [4]) je sprovedena automatska optimizacija hiperparametara koristeći Optuna okvir sa semplerom zasnovanim na strukturiranoj Parzenovoj estimaciji (Tree-structured Parzen Estimator — TPE). Prostor hiperparametara obuhvatao je sljedeće varijable: veličina paketa uz vrijednosti 1, 2, 4 i 8; stopa učenja iz log-uniformne distribucije u opsegu  $[5 \times 10^{-5}, 5 \times 10^{-3}]$ ; raspadanje težina (weight decay) takođe



Slika 1: Statističko poređenje performansi modela kutijastim dijagramom

iz log-uniformne distribucije u opsegu  $[1 \times 10^{-7}, 1 \times 10^{-3}]$ ; te broj akumulacionih koraka sa vrijednostima 1, 2 i 4.

Eksperimentalni parametri su uključivali maksimalno 15 proba po modelu, sa po 25 epoha po probi. Za rano zauzavljanje definisano je strpljenje od 10 epoha, dok je kriterijum optimizacije maksimizacija IoU metrike na validacionom skupu.

2) *Evaluacija uticaja augmentacija i FOV maskiranja:* Uticaj augmentacija i FOV (Field of View) maskiranja evaluiran je kroz četiri različite konfiguracije za svaki model: bez FOV maskiranja i bez augmentacija; bez FOV maskiranja sa augmentacijama; sa FOV maskiranjem bez augmentacija; i sa oba pristupa, FOV maskiranjem i augmentacijama. Za svaku od ovih konfiguracija korišćeni su optimalni hiperparametri identifikovani u fazi optimizacije.

#### B. Kriterijumi evaluacije

1) *Kvantitativne metrike:* Za evaluaciju su korišćene standardne metrike binarne segmentacije. Dice koeficijent se definiše kao

$$Dice = \frac{2|P \cap T|}{|P| + |T|}$$

gdje  $P$  predstavlja predikciju modela, a  $T$  stvarnu oznaku. Presjek po uniji (IoU) računa se kao

$$IoU = \frac{|P \cap T|}{|P \cup T|}$$

Takođe su mjereni preciznost i odziv, definisani kao

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

pri čemu  $TP$ ,  $FP$  i  $FN$  označavaju broj stvarno pozitivnih, lažno pozitivnih i lažno negativnih piksela. Kada je FOV maskiranje bilo aktivno, sve metrike su računate unutar FOV regiona, u skladu sa preporukama prethodnih studija [1], [12].

2) *Statistička analiza:* Za analizu varijabilnosti performansi korišćene su srednje vrijednosti i standardne devijacije kroz sve probe, dok su različite vrste grafičkih prikaza, uključujući kutijaste dijagrame, dijagrame rasipanja i violinske dijagrame, korišćene za vizualizaciju rezultata i identifikaciju korelacija među hiperparametrima.

#### C. Implementacijske specifikacije

Eksperimenti su sprovedeni u Google Colab Pro okruženju sa Tesla T4 GPU-om koji raspolaže sa 16 GB VRAM-a.

Optuna okvir (verzija 3.0+) je korišćen za optimizaciju hiperparametara. Albumentations biblioteka [6] je služila za augmentaciju podataka, dok su se za modele koristile implementacije iz segmentation\_models\_pytorch (za U-Net) i HuggingFace Transformers biblioteke (za SegFormer) [4].

Kako bi se osigurala reproduktivnost eksperimenta, fiksirano je sjeme (seed) 42 za sve faze eksperimenta. Takođe su primijenjene determinističke CUDA operacije gdje je to bilo moguće, kao i isto sjeme za učitavače podataka (data loaders) i augmentacije.

#### D. Ograničenja eksperimentalne postavke

Resursna ograničenja podrazumijevala su maksimalnu veličinu paketa ograničenu na 8 zbog dostupne GPU memorije, kao i kratko trajanje sesija koje je ograničilo broj proba na 15 po modelu. Povremeni prekidi u radu uticali su na kontinuitet eksperimenta.

Sa metodološke strane, mali broj pokušaja mogao je biti nedovoljan za potpunu eksploraciju prostora hiperparametara, dok zbog vremenskih ograničenja nije korišćena unakrsna validacija. Optimizacija je fokusirana isključivo na IoU metriku.

Uprkos navedenim ograničenjima, eksperimentalna postavka omogućava validno poređenje performansi različitih modela i konfiguracija u uslovima ograničenih resursa, što je praktičan scenario za mnoge istraživačke institucije [7].

## V. ARHITEKTURA MODELA

#### A. U-Net arhitektura

U-Net je konvolutivna neuronska mreža razvijena za segmentaciju, karakteristična po simetričnoj koder-dekoder strukturi sa preskočnim vezama. Korišćena je implementacija iz segmentation\_models\_pytorch biblioteke sa sljedećim modifikacijama:

- Početni broj konvolutivnih filtera smanjen je na 32 u odnosu na standardnih 64, kako bi se smanjila ukupna memorijska zahtjevnost modela i omogućila obuka pri malim veličinama paketa.
- InstanceNorm2d umjesto BatchNorm, radi stabilnosti sa malim veličinama paketa
- Dropout u dekoderskim slojevima sa linearnim opadanjem stopce od 0.3 do 0.0

## B. SegFormer-B0 arhitektura

SegFormer-B0 arhitektura predstavlja kombinaciju transformera i jednostavnog MLP dekodera, a razvijen je prvenstveno za semantičku segmentaciju prirodnih scena.

U ovom istraživanju korišćen je prethodno trenirani model nvidia/segformer-b0-finetuned-ade-512-512 iz HuggingFace Transformers biblioteke, uz sljedeće adaptacije kako bi bio prilagođen zadatku binarne segmentacije:

- Originalni izlazni sloj višeklasne segmentacije je zamjenjen novim konvolutivnim slojem sa jednim izlaznim kanalom, namijenjenim binarnoj klasifikaciji
- Novi sloj je inicijalizovan Xavier uniformnom metodom, što obezbeđuje stabilan početak obuke
- Primijenjena je strategija zagrijavanja, u kojoj je koder zamrznut tokom prve 3 epoha, omogućavajući dekoderu da se adaptira bez uticaja prethodne obuke
- Pomoćna grana (auxiliary head), koja je dio originalne arhitekture, reinicijalizovana je i uključena u obuku sa ciljem da poboljša tok gradijenata tokom ranih faza obuke.

Iako SegFormer nije dizajniran za medicinski domen, njegove sposobnosti modelovanja globalnog konteksta su se pokazale kao korisne u segmentaciji vaskularnih struktura. Međutim, mijenjanje zadatka modela iz domena prirodnih slika u oblast medicinske slike iziskuje dodatne regularizacione i adaptacione mehanizme, koji nisu dostupni u ovom radu zbog resursnih ograničenja. SegFormer-B0 je ipak pokazao bolje performanse nego pristup zasnovan na konvolutivnim neuronskim mrežama.

## C. Strategija obuke

Funkcije greške u segmentaciji mogu se grubo podijeliti u tri kategorije:

- **Zasnovane na raspodjeli** (npr. unakrsna entropija): fokusirane na piksel-po-piksel klasifikaciju
- **Zasnovane na regijama** (npr. Dice): fokusirane na cjelokupan preklop između predikcije i stvarne maske
- **Mješovite** (npr. BCE+Dice): kombinacije koje balansiraju lokalnu i globalnu tačnost

Kombinovana funkcija greške, koja spaja binarnu unakrsnu entropiju i Dice koeficijent, je odabrana zbog specifičnih okolnosti zadatka segmentacije krvnih sudova. BCE komponenta se fokusira na tačnost klasifikacije pojedinačnih piksela, što osigurava stabilnost gradijenata tokom obuke i koristi se za precizno učenje graničnih slučajeva. Međutim, BCE ne uzima u obzir prostornu koherenciju niti problem neravnoteže klase (krvni sudovi zauzimaju samo mali dio slike). Dice greška optimizuje preklapanje između predikcije i maske, favorizujući tačnije obuhvatanje tankih, povezanih struktura.

Kombinovanjem obje komponente:

$$\text{Loss} = 0.6 \times \text{BCE} + 0.4 \times \text{Dice} \quad (1)$$

postiže se ravnoteža između lokalne preciznosti (BCE) i globalne segmentacione konzistentnosti (Dice). Ovaj pristup

je potvrđen kao efikasan u brojnim radovima iz oblasti medicinske segmentacije, gdje je prikazano da kombinovana funkcija greške poboljšava performanse modela u zadacima sa izrazitom neravnotežom klasa. [7]

Korišćen je Adam optimizator sa ReduceLROnPlateau (faktor 0.5, strpljenje (patience) = 5) i strategijom ranog zauzavljanja (strpljenje = 10 epoha bez poboljšanja validacionog IoU rezultata).

## VI. EKSPERIMENTALNI REZULTATI

### A. Kvantitativna evaluacija

Evaluacija je sprovedena korišćenjem standardnih metrika za binarnu segmentaciju, opisanih u poglavljvu IV-B1.

Tabela I: Performanse modela (srednje vrijednosti  $\pm$  standardna devijacija)

Model	Dice	IoU	Preciznost	Odziv	Najbolji Dice	Proba
U-Net	0.292 $\pm$ 0.049	0.490 $\pm$ 0.052	0.631	0.686	0.359	15
SegFormer	0.338 $\pm$ 0.041	0.554 $\pm$ 0.064	0.734	0.695	0.367	15

Tabela I i slika 1 jasno pokazuju superiornost SegFormer modela u svim ključnim metrikama. SegFormer postiže višu prosječnu Dice vrijednost u poređenju sa U-Net-om, kao i prednost u IoU metriči, što ukazuje na to da SegFormer ima bolju sposobnost da precizno definiše granice krvnih sudova.

**Kritična analiza varijabilnosti:** Kutijasti dijagram (slika 1) pokazuje da SegFormer, pored boljih srednjih performansi, demonstrira i značajno manju varijabilnost rezultata, sa užim interkvartilnim rasponom i manje izraženim odstupajućim vrijednostima. U-Net pokazuje veći broj odstupajućih vrijednosti, što ukazuje na nepredvidivost performansi u različitim eksperimentalnim uslovima.

Radarski dijagram (slika 3) vizualizuje balans performansi između različitih metrika. SegFormer redovno pokazuje bolje performanse kroz sve dimenzije evaluacije.

Najbolje konfiguracije hiperparametara identifikovane kroz Optuna optimizaciju:

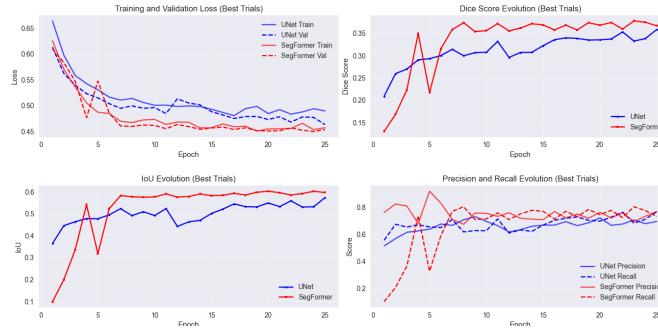
**U-Net:**  $\text{LR} = 6.34 \times 10^{-4}$ ,  $\text{BS} = 1$ ,  $\text{WD} = 2.18 \times 10^{-4}$

**SegF:**  $\text{LR} = 4.84 \times 10^{-4}$ ,  $\text{BS} = 1$ ,  $\text{WD} = 1.15 \times 10^{-6}$

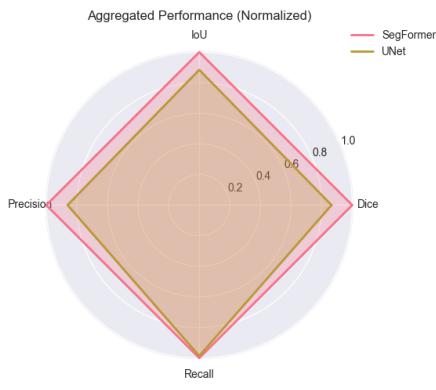
### B. Analiza dinamike obučavanja

Analiza krive učenja (slika 2) pokazuje fundamentalne razlike u dinamici obuke između dva modela. SegFormer konvergira brže i dostiže stabilan nivo performansi već oko 15. epoha, sa konzistentnim razmakom između metrika obuke i validacionih metrika. U-Net, s druge strane, ima sporiju konvergenciju sa izraženim oscilacijama u validacionim metrikama nakon 20. epoha, pri čemu se razmak između performansi na skupovima za obuku i validaciju progresivno povećava – jasna indikacija preprilagođavanja.

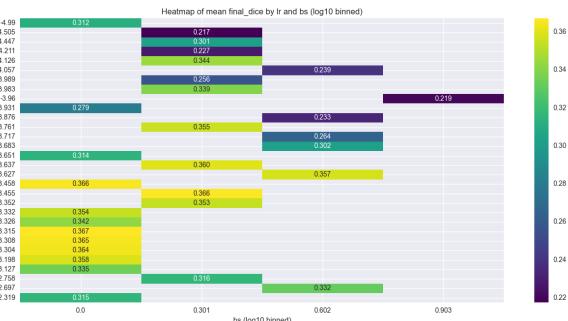
Standardne devijacije Dice metrika tokom epoha dodatno potvrđuju veću stabilnost SegFormer-a – prosečna varijabilnost tokom obuke je oko 30% manja u poređenju sa U-Net-om. SegFormer redovno ima manje standardne devijacije kroz cijeli proces obuke (prosječno 30% manja varijabilnost).



Slika 2: Evolucija metrika obuke za najbolje probe U-Net i SegFormer modela



Slika 3: Radarski dijagram normalizovanih performansi oba modela



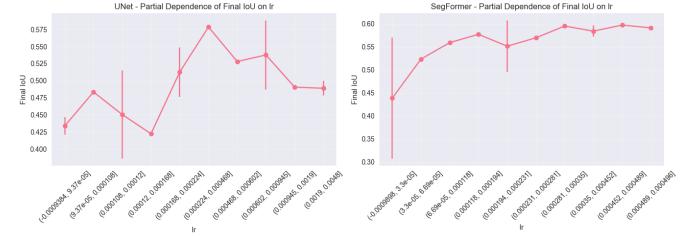
Slika 4: Dijagram rasipanja - analiza uticaja ključnih hiperparametara na Dice rezultat

### C. Analiza hiperparametara

Dijagram rasipanja (slika 4) prikazuje interakcije između stope učenja, veličine paketa i raspadanja težina. SegFormer ima klaster dobrih rezultata oko LR=5e-4, dok U-Net ima fragmentovanu distribuciju, što čini optimizaciju hiperparametara značajno težom. Za oba modela, optimalna stopa učenja se nalazi u opsegu  $4 - 7 \cdot 10^{-4}$ . Ako posmatramo raspadanje težina, U-Net ima najbolje rezultate pri umjerenim vrijednostima ( $\sim 10^{-4}$ ), dok SegFormer postiže najbolje rezultate pri

vrlo malim vrijednostima ( $\sim 10^{-6}$ ), tolerajući širok spektar vrijednosti bez degradacije performansi.

Stopa učenja ima najjači uticaj na performanse kod oba modela, pri čemu optimalni opseg varira između arhitektura. Raspadanje težina pokazuje slabiju, ali učestalu negativnu korelaciju sa performansama kod U-Net-a (-0.3). Kod SegFormera ovaj uticaj je zanemarljiv (-0.05), što potvrđuje robusnost SegFormer arhitekture na regularizacione parametre.



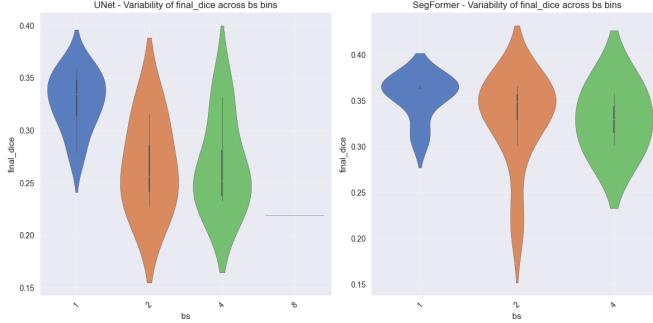
Slika 5: Parcijalna zavisnost IoU rezultata od stope učenja (sa binovanim vrijednostima)

Analiza parcijalne zavisnosti (Slika 5) potvrđuje da optimalna stopa učenja za oba modela leži u srednjem opsegu testiranih vrijednosti, sa SegFormer-om koji pokazuje užu optimalnu zonu, ali veće maksimalne performanse.

### D. Varijabilnost performansi

SegFormer ima konzistentniji opseg rezultata sa manjom varijabilnošću (standardna devijacija 0.041 vs 0.049), što je dobro za praktičnu primjenu u uslovima gdje je pouzdanost kritična. Individualne tačke pokazuju da SegFormer ima manje "katastrofalnih" rezultata (minimum Dice 0.25 vs 0.18 kod U-Net-a), što sugerira stabilniju arhitekturu u nepovoljnim scenarijima.

Violinski dijagram (slika 6) dodatno potvrđuje da SegFormer postiže stabilnije performanse kroz različite konfiguracije hiperparametara, sa uže definisanim optimalnim opsegom vrijednosti stope učenja. Analiza distribucijskih oblika pokazuje da SegFormer ima unimodalnu distribuciju sa definisanim pikom, dok U-Net pokazuje bimodalnu tendenciju, što objašnjava poteškoće u optimizaciji hiperparametara.



Slika 6: Violinski dijagrami distribucije Dice rezultata kroz vrijednosti stope učenja

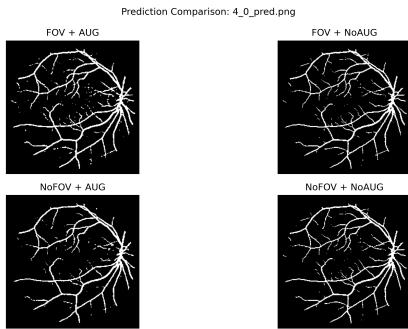
Tabela II: Uticaj augmentacija i FOV maskiranja na performanse

Konfiguracija	U-Net Dice	U-Net IoU	SegFormer Dice	SegFormer IoU
Bez FOV, bez augmentacija	0.369	0.574	0.389	0.611
Bez FOV, sa augmentacijama	0.349	0.551	0.378	0.604
Sa FOV, bez augmentacija	0.369	0.574	0.384	0.611
Sa FOV, sa augmentacijama	0.355	0.564	0.380	0.607

#### E. Analiza uticaja augmentacija i FOV maskiranja

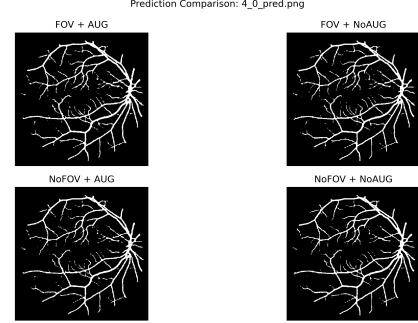
Tabela II prikazuje rezultate eksperimenta sa različitim kombinacijama augmentacija i FOV maskiranja. Zanimljivo je da augmentacije nisu dovele do poboljšanja performansi. To može biti posljedica postojće varijabilnosti u DRIVE skupu ili destabilizacije obuke pri malim veličinama paketa. SegFormer konzistentno nadmašuje U-Net u kvantitativnim metrikama kroz sve konfiguracije, sa posebno izraženom prednošću u IoU skorovima.

Analiza varijansi između različitih konfiguracija otkriva da U-Net pokazuje veću osjetljivost na augmentacije (pad Dice koeficijenta za 0.020 sa augmentacijama), dok SegFormer pokazuje stabilniju toleranciju (pad od 0.011). FOV maskiranje ima blag pozitivan uticaj kod oba modela.



Slika 7: Predikcije U-Net modela pri različitim uslovima obuke

Vizuelna analiza prikazana na slikama 7 i 8 pokazuje razliku u strategijama segmentacije između dva modela, koja nije u potpunosti izražena kroz kvantitativne metrike. Iako SegFormer ostvaruje više vrijednosti Dice i IoU metrika, vizuelnom inspekcijom predikcija uočava se da taj rezultat potiče od agresivnog smanjenja lažno pozitivnih regiona, pri čemu model često zanemaruje tanje sudove i gubi kontinuitet

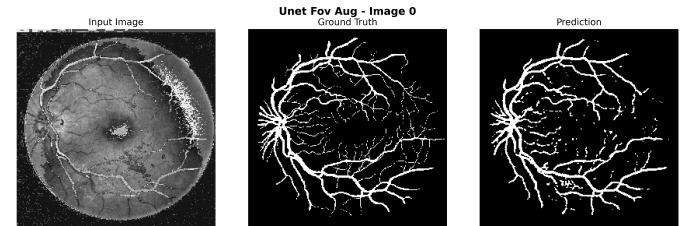


Slika 8: SegFormer-B0 predikcije pri različitim uslovima obuke

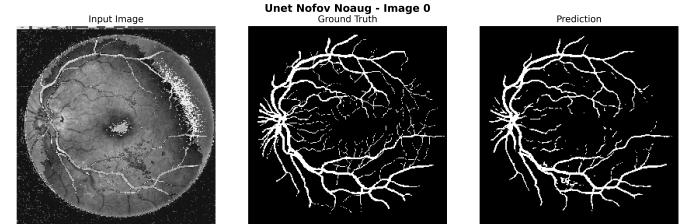
segmentacije, naročito u perifernim zonama retine. Rezultat su vizuelno čiste ali fragmentovane segmentacije, sa prekidima u oblasti bifurkacija i suženja.

U-Net održava vaskularni kontinuitet i u uslovima šuma i niskog kontrasta, uspješno detektujući i slabo izražene grane vaskularne mreže. Iako vodi do većeg broja lažno pozitivnih piksela i slabijih kvantitativnih rezultata, očuvanje morfološke cjelovitosti segmentiranih struktura, u kontekstu sličnih kliničkih scenarija, predstavlja prednost.

Ovi rezultati ukazuju da metrički superioriji model ne mora nužno biti klinički pogodniji, kao i da evaluacija mora obuhvatiti i topološku i funkcionalnu verifikaciju rezultata, a ne samo numeričku tačnost.

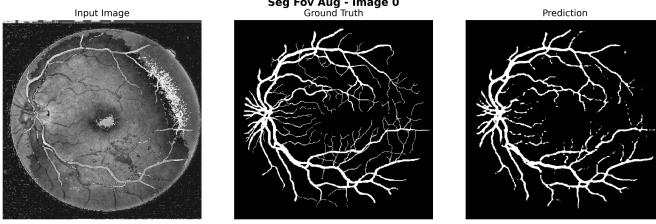


Slika 9: U-Net predikcija sa FOV maskiranjem i augmentacijama

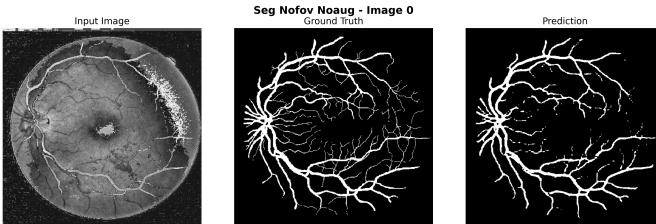


Slika 10: U-Net predikcija bez FOV maskiranja i augmentacija

Detaljno vizuelno poređenje (slike 9-12) otkriva kompleksnu dinamiku između različitih konfiguracija obuke. FOV maskiranje konzistentno ima pozitivan uticaj na kvalitet segmentacije kod oba modela, eliminirajući šum van oblasti mrežnjače i omogućavajući modelima da se fokusiraju na relevantne strukture. Posebno je vidljivo da FOV maskiranje



Slika 11: SegFormer-B0 predikcija sa FOV maskiranjem i augmentacijama



Slika 12: SegFormer-B0 predikcija bez FOV maskiranja i augmentacija

povećava kontrast vaskularne segmentacije u centralnim regionima kod oba modela.

Važna opservacija je da SegFormer ima tendenciju "prorjeđavanja" vaskularnih struktura, što rezultuje segmentacijama koje, iako preciznije u smislu eliminacije lažno pozitivnih regiona, gube kritične informacije o kompletnosti vaskularne mreže. U kliničkim scenarijima gdje je potpuna vizualizacija vaskularnog stabla neophodna za dijagnostiku, ova karakteristika može biti problematična.

Augmentacije pokazuju neočekivan negativan ili neutralan uticaj kod oba modela, što može biti posljedica nekoliko faktora: male veličine paketa koja destabilizuje efekat augmentacija, već postojeće varijabilnosti u DRIVE skupu podataka (koja čini dodatne augmentacije redundantnim) ili specifičnosti medicinskih slika retine koja zahtjeva precizniju kalibraciju augmentacijskih parametara.

Rezultati sugerisu da izbor modela treba da bude vođen specifičnim kliničkim zahtjevima: SegFormer za zadatke gdje je prioritet minimizacija lažno pozitivnih detekcija i preciznost granica, a U-Net za zadatke gdje je kritična cjelovitost i kontinuitet vaskularne segmentacije.

## VII. DISKUSIJA I OGRANIČENJA

### A. Analiza performansi

Postignuti rezultati (Dice 0.29-0.34, IoU 0.49-0.55) zaoštaju za najboljim vrijednostima na DRIVE skupu [17], što je očekivano s obzirom na eksperimentalna ograničenja, uključujući resurse i obim podataka.

#### Ograničenja resursa:

- ograničeno vrijeme sesije ograničava dubinu optimizacije hiperparametara
- Tesla T4 GPU sa 16GB VRAM ograničava veličine paketa na 1-4 uzorka

- periodični prekidi rada narušavaju kontinuitet eksperimenta
- ograničen pristup najsvremenijim modelima

**Ograničenost skupa podataka:** DRIVE skup sa svega 40 slika je ekstremno ograničen resurs za obučavanje dubokih modela i iziskuje napredne tehnike regularizacije i prelaznog učenja koje nisu implementirane u ovom radu.

**Simplifikovane konfiguracije:** Zbog pomenutih ograničenja, korišćene su relativno jednostavne arhitektonске modifikacije i strategije obučavanja, bez naprednih tehniki poput ansamblskih metoda [1] ili kompleksnih augmentacija [6].

### B. Ključni nalazi

SegFormer-B0 je pokazao superiornost u odnosu na U-Net model, što može biti pripisano upotrebi težina iz ADE20K skupa, boljoj sposobnosti modelovanja globalnog konteksta ili efikasnijoj arhitekturi za rad sa ograničenim podacima.

FOV maskiranje se pokazalo kao bitna komponenta za postizanje relevantnih rezultata, dok se augmentacije, neočekivano, nisu pokazale kao korisne.

### C. Preporuke za buduće istraživanje

Na osnovu dobijenih rezultata i uočenih ograničenja, mogu se predložiti brojni pravci za unapređenje istraživanja u budućem radu. Upotreba naprednjih prethodno obučenih modela, kao što su Swin Transformer, Segment Anything Model (SAM) ili domenski prilagođeni modeli prethodno obučeni na medicinskim skupovima, mogla bi da omogući efikasnije iskorišćavanje reprezentacija u kontekstu sličnih zadataka.

Dalji korak u unapređenju se odnosi na upotrebu sofisticiranih funkcija greške koje su prilagođene problemu sa disbalansom klasa. Funkcije poput Tverski greške, fokalne ili ujedinjenje fokalne greške (Unified Focal Loss) omogućavaju efikasnije učenje na neuravnoteženim pikselima - ove funkcije eksplicitno favorizuju učenje na rijetkim ili teško procjenjivim pikselima.

Dalje, budući rad bi mogao uključiti ansamblske pristupe, gdje se više modela trenira nezavisno, a njihove predikcije se kombinuju u cilju poboljšanja robusnosti i smanjenja varijabilnosti rezultata.

Neophodno je sprovesti i unakrsnu validaciju, kako bi se izbjegla prevelika zavisnost od konkretne podjele skupova za obuku i testiranje. S obzirom da je DRIVE skup mali, evaluacija na samo jednoj podjeli može dovesti do nesigurnih zaključaka.

Na kraju, važno je razmotriti razvoj i primjenu hibridnih evaluacionih metrika koje bi istovremeno kvantifikovale tačnost i morfološku konzistentnost segmentacije. Standardne metrike poput Dice i IoU nisu uvjek dovoljne za evaluaciju kvaliteta vaskularnih struktura, kako ne uzimaju u obzir kontinuitet i topološku verodostojnost segmentacije.

## VIII. ZAKLJUČAK

Ovaj rad istražuje kompromis između kvantitativne tačnosti i morfološke konzistentnosti u segmentaciji krvnih su-

dova na retini, pod uslovima ograničenih podataka. Transformer arhitektura SegFormer-B0 ima prednost u standardnim metrikama, dok vizuelna inspekcija pokazuje da U-Net bolje čuva kontinuitet vaskularne mreže. Dakle, evaluacija modela u medicinskom domenu mora nadilaziti numeričke metrike i obuhvatiti domenski značaj struktura. Buduće istraživanje treba da integrise topološke metrike i kompleksnije adaptacione strategije, kako bi se ostvarila ravnoteža između preciznosti i kliničke korisnosti.

## REFERENCES

- [1] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.
- [2] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.
- [4] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.
- [5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [6] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albuminations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.
- [7] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.
- [8] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *arXiv preprint arXiv:1802.06955*, 2018.
- [9] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- [10] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransU-Net: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
- [11] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-U-Net: U-Net-like pure transformer for medical image segmentation," *European Conference on Computer Vision*, pp. 205–218, Springer, 2022.
- [12] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-U-Net: Spatial attention U-Net for retinal vessel segmentation," *International Conference on Pattern Recognition*, pp. 1236–1242, IEEE, 2021.
- [13] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DU-Net: A deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.
- [14] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and T. Wen, "FAT-Net: Feature adaptive transformers for automated skin lesion segmentation," *Medical Image Analysis*, vol. 76, p. 102327, 2022.
- [15] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "U-Net++: A nested U-Net architecture for medical image segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11, Springer, 2018.
- [16] N. Ibtehaz and M. S. Rahman, "MultiResU-Net: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [17] G. Kovács and A. Fazekas, "How (not) to evaluate retinal vessel segmentation methods," *arXiv preprint arXiv:2111.03853*, 2021.