Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 004.9

DOI: https://doi.org/10.24867/24BE33Kljajic

SEGMENTACIJA LEZIJA U MAMOGRAFSKIM SLIKAMA KORIŠĆENJEM SEGFORMER MREŽNE ARHITEKTURE

LESION SEGMENTATION IN MAMMOGRAPHIC IMAGES USING THE SEGFORMER NETWORK ARCHITECTURE

Jovana Kljajić, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – *Detekcija promena u tkivu dojke, koje su često vezane za razvoj malignih oboljenja, omogućava postavljanje rane dijagnoze, što je jedan od ključnih faktora za povećanje uspešnosti lečenja raka dojke. Radi toga potrebno je pronaći algoritme koji će na brz i precizan način detektovati postojanje lezija. Zbog svoje jednostavnosti i efikasnosti, primena transformera u segmentaciji slike tokom poslednjih godina postaje sve popularnija. U ovom radu korišćena je Segformer mrežna arhitektura za segmentaciju lezija u okviru mamografskih snimaka koji postoje u INbreast bazi. Dobijeni rezultati pokazuju značajan potencijal za primenu ove arhitekture u realnim uslovima.*

Ključne reči: *Segformer, segmentacija, lezija, mamografski snimci*

Abstract – *Changes in breast tissue can indicate early stages of malignant diseases, in addition an early diagnosis is one of the key factors in increasing the success of the treatment. Therefore, it is necessary to find algorithms that will efficiently and precisely detect the presence of lesions. Due to their simplicity and efficiency, the application of transformers in image segmentation has gained a lot attention in recent years. In this paper, the Segformer network architecture was used for lesion segmentation in mammography images. The obtained results demonstrate significant potential for the application of this architecture in real-world settings.*

Keywords: *Segformer, segmentation, lesion, mammography images*

1. UVOD

Semantička segmentacija je jedan od osnovnih zadataka kojima se bavi kompjuterska vizija, a podrazumeva da se vrši klasifikacija pojedinačnih piksela koji sačinjavaju sliku. S obzirom na sličnost između klasifikacije i semantičke segmentacije, veliki broj arhitektura koje su kreirane za klasifikaciju slika se uspešno koriste za njihovu semantičku segmentaciju.

Jedan takav model, koji će ovde biti predstavljen i prilagođen za segmentaciju lezija u okviru mamografskih snimaka, jeste *Segformer* [1].

NAPOMENA:
Ovaj rad proistekao je iz master rada čiji je mentor dr Nikša Jakovljević, vanr. prof.

Ovaj model se zasniva na transformerima, i veoma je efikasan, tačan i robustan [1]. Novine koje on uvodi su: 1) ne zahteva poziciono kodovanje, 2) koristi obeležja različite rezolucije koje generiše hijerarhijski koder zasnovan na transformerima 3) računski efikasan All-MLP (eng. *All – Multilayer Perceptron*) dekoder.

Semantička segmentacija lezija na mamografskim snimcima je veoma važna jer može značajno da olakša i ubrza doktorima davanje dijagnoze, čime se povećavaju šanse pacijenta za izlečenje, a pošto su u pitanju ljudski životi ovakvi modeli moraju imati veoma visoku tačnost.

U nastavku će biti detaljno opisana arhitektura korišćenog modela u najvećoj meri zasnovana na radu [1]. Opis baze podataka koja je korišćena za treniranje i testiranje modela je data u delu 3. Nakon toga sledi prikaz rezultata uz odgovarajuću diskusiju i na kraju su dati najvažniji zaključci.

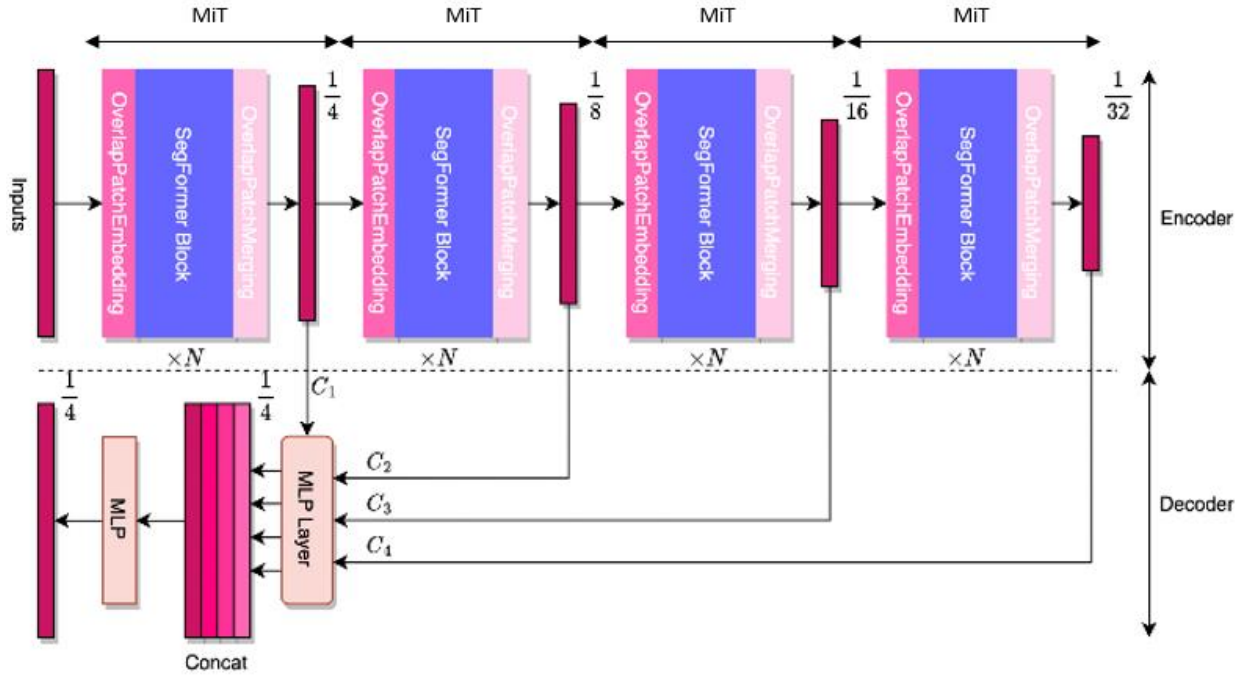
2. OPIS ARHITEKTURE *SEGFORMER*-A

Arhitektura *Segformera* je prikazana na sl. 1 gde se uočava da se sastoji iz dva funkcionalna dela. Prvi je hijerarhijski koder zasnovan na transformerima (na slici označen sa *encoder*), koji generiše obeležja visoke i niske rezolucije. Drugi deo je All-MLP dekoder koji služi za kombinovanje ovih obeležja dobijenih u više nivoa, kako bi proizveo finalnu masku za semantičku segmentaciju.

Kao ulaz u model očekuje se slika dimenzija $H \times W \times 3$, konkretno, modeli koji su ovde korišćeni očekuju kao ulaz mamografsku sliku dimenzija $512 \times 512 \times 3$. Ulazna slika se deli na segmente dimenzija 4×4, koji se zatim koriste kao ulazi u hijerarhijski koder koji na izlazu daje pomenuta obeležja sa različitim rezolucijama [1].

2.1 Hijerarhijski koder zasnovan na transformerima

Prvi funkcionalni deo, koder, služi za određivanje obeležja iz ulaznih slika koja na najbolji način prave razliku između željenih klasa (piksel jeste ili nije lezija). On na izlazu daje obeležja različitih rezolucija, koja će posle, u okviru dekodera, biti kombinovana radi donošenja odluke. U okviru *Segformera* stepeni kodera, kojih ima 4, su označavani sa *Mix Transformer encoders*, ili skaćeno MiT, tako da će i u nastavku teksta biti tako označeni, kao i na slici 1. U nastavku će biti detaljnije opisan svaki deo od kog se sastoji MiT [1] i to *overlap patch embedding*, *segformer* blok i *overlap patch merging*.



Slika 1. *Arhitektura Segformer modela [2].*

2.1.1 *Overlap Patch Embedding*

Na početku svakog MiT bloka, nalazi se modul koji radi *overlap patch embedding.* Ovaj proces se sastoji iz toga da se slika izdeli na segmente (eng. *patch*) koji se zatim spajaju u jednu sekvencu. Ovde se preklapanja obezbeđuju time što je postavljeno da je veličina segmenta veća od pomeraja (eng. *stride*) što dovodi do toga da se informacije dele između segmenata.

2.1.2 *Segformer* blok

Na slici 2. dat je detaljan prikaz modula od kojih se sastoji *Segformer* blok. Sekvenca koja pristiže na njegov ulaz prvo prolazi kroz *efficient self-attention* blok, a zatim kroz *Mix Feed Forward* mrežu (Mix-FFN).

Efficient Self-Attention:

Ovaj modul unosi najveću računsku kompleksnost u model, i samim tim usporava proces obuke modela. On predstavlja malu modifikaciju originalnog *multi-head self-attention* procesa. Svaka od glava (eng. *head*) Q, K, i V ima istu dimenziju $N \times C$, gde je $N = H \times W$ i predstavlja dužinu sekvence, C je broj mapa obeležja. Vrednost *self-attention* modula se procenjuje na osnovu sledeće formule:

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^{\text{T}}}{\sqrt{d_{head}}}\right)V \qquad (1)$$

Računska kompleksnost ovog procesa je $O(N^2)$, što je veoma nepraktično u slučaju da su slike koje se dovode na ulaz modela velike rezolucije. Zbog toga se radi proces redukcije sekvence, koji je prvi put predstavljen u [3]. U okviru ovog procesa koristi se faktor redukcije R koji na sledeći način redukuje dužinu sekvence:

$$\widehat{K} = \text{Reshape}(\frac{N}{R}, C \cdot R)(K) \qquad (2)$$

$$K = \text{Linear}(C \cdot R, C)\left(\widehat{K}\right) \qquad (3)$$

gde je K sekvenca koju je potrebno redukovati, $\text{Reshape}(N/R, C \cdot R)(K)$ predstavlja promenu dimenzija sekvence K sa dimenzija $N \times C$ na dimenzije $(N/R) \times (C \cdot R)$, a $\text{Linear}(C_{in}, C_{out})(\cdot)$ se odnosi na linearni sloj koji kao ulaz prima tenzor dimenzije C_{in}, a na izlazu daje tenzor dimenzije C_{out}. Novo K koje se dobija na izlazu modula za redukciju, ima dimenzije $(N/R) \times C$. Kao rezultat dobija se smanjenje kompleksnosti modula pažnje (*self-attention*) sa $O(N^2)$ na $O(N^2/R)$.

Mix-FFN*:*

U okviru ovih modela, kako bi se uzela u obzir informacija o lokacijama objekata u okviru slika, koristi se Mix-FFN. On implementira konvolucioni sloj po dubini, u nastavku DWC (eng. *Depthwise Convolution*) koji koristi kernele dimenzije 3×3 i primenjuje se direktno nad izlazom iz *feed-forward* mreže (FFN). Ono što ovaj tip konvolucije razlikuje od uobičajene, jeste to što se koristi zaseban kernel za svaki od kanala ulaznog segmenta.

Ovaj modul se može definisati pomoću sledeće formule:

$$\mathbf{x}_{out} = \text{MLP}\left(\text{GELU}\left(\text{DWC}_{3\times3}\left(\text{MLP}(\mathbf{x}_{in})\right)\right)\right) + \mathbf{x}_{in} \qquad (4)$$

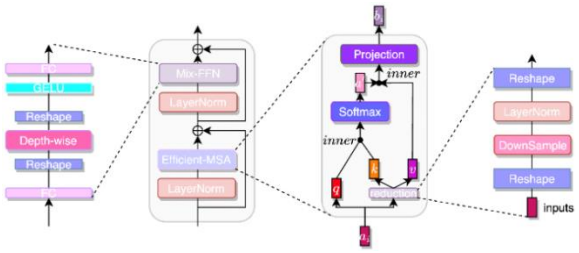
gde $\mathbf{x}_{in}$ predstavljaju obeležja iz *self-attention* modula koja se dovode na ulaz ovog modula.

2.1.3 *Overlap Patch Merging*

Zadatak ovog dela je da obezbedi smanjenje dimenzije mapa obeležja (na sl. 1 predstavljenih bordo pravougaonicima) sa dimenzije $F_1(\frac{H}{4} \times \frac{W}{4} \times C_1)$ na $F_2(\frac{H}{8} \times \frac{W}{8} \times C_2)$ po uzoru na ideju datu u ViT modelu [4].

Kako bi se obezbedilo da nakon primene ovog algoritma ostane očuvana lokalna informacija između ovih segmenata, postoji preklapanje među njima, zbog čega se čitav proces i naziva *overlapped patch merging.* Radi sprovođenja ovog postupka, neophodno je definisati nekoliko parametara. Prvi je K koji predstavlja veličinu segmenta. Sledeći parametar jeste S, koji predstavlja pomeraj između susednih segmenata, i poslednji para-

metar je P koji ukazuje na broj piksela kojima se proširuje (eng. *padding*) ulazna mapa obeležja. U okviru modela *Segformer*-a, u zavisnosti od kompleksnosti strukture koja se koristi, ovi parametri su ili $K = 7, S = 4, P = 3$ ili $K = 3, S = 2, P = 1$.



Slika 2. *Prikaz dijagrama detaljne arhitekture Segformer bloka [2].*

2.2 All-MLP dekoder

Dekoder koji se koristi u okviru *Segformer*-a sastoji se isključivo od MLP (eng. *Multilayer Perceptrons*) slojeva kako bi se izbegli dekoderi koji se inače koriste u okviru drugih metoda, a računski su veoma zahtevni. Ključni element koji omogućava korišćenje ovakvog jednostavan dekodera jeste hijerarhijski koder zasnovan na transformerima jer obezbeđuje veće efektivno receptivno polje (eng. *effective receptive field*) od CNN (eng. *Convolutional Neural Network*) kodera. MLP dekoder koji se koristi sastoji se od četiri osnovna sloja. U prvom sloju, mape obeležja F_i koje su generisane u od svakog pojedinačnog MiT bloka u okviru kodera, prolaze kroz linearni sloj kako bi se uskladile dimenzije kanala.

$$\widehat{F_i} = \text{Linear}(C_i, C)(F_i), \forall i \tag{5}$$

Drugi sloj usklađuje dimenzije mapa obeležja tako da sve imaju istu dimenziju $\frac{H}{4} \times \frac{W}{4} \times C$, nakon čega se vrši njihova konkatenacija. Mape obeležja iz prvog MiT bloka već imaju ove dimenzije, te se prosleđuju direktno u deo za konkatenaciju, dok se dimenzije ostalih povećavaju.

$$\widehat{F_i} = \text{Upsample}\left(\frac{H}{4} \times \frac{W}{4}\right)(\widehat{F_i}), \forall i \tag{6}$$

U narednom linearni sloj stapa konkatenirane mape obeležja F.

$$F = \text{Linear}(4C, C)\big(\text{Concat}(\widehat{F_i})\big), \forall i \tag{7}$$

Na kraju se mapa obeležja F prosleđuje na ulaz još jednog linearnog sloja kako bi on napravio predikciju segmentacione maske koja ima sledeću rezoluciju $\frac{H}{4} \times \frac{W}{4} \times N_{cls}$, gde je N_{cls} broj kategorija. Formula kojom se definiše dolazak do ove segmentacione maske je:

$$M = \text{Linear}(C, N_{cls})(\text{F}) \tag{8}$$

3. BAZA PODATAKA

U okviru ove studije korišćena je baza podataka INbreast koja sadrži mamografske snimke i odgovarajuće segmentacione maske. Postoje snimci bolesnih i zdravih dojki [5]. Ukupno ima 107 snimaka bolesnih pacijenata u celoj bazi. Svi snimci su na početku iz DICOM formata prebačeni u jpg format. Snimcima su promenjene dimenzije, tako da budu $512 \times 512 \times 3$ jer je to očekivani ulaz mreža koje su korišćene. Nad svim slikama je primenjena ekvalizacija histograma, dok je nad slikama iz trening skupa pored ekvalizacije primenjena još i rotacija i CLAHE (adaptivna ekvalizacija histograma) kako bi se povećao skup slika za obuku.

4. OBUKA MODELA

Baza podataka je podeljena na pet približno jednakih delova kako bi se mogli obučiti pet različitih modela, gde se za svaki selektuju različiti skupovi za test i validaciju, a ostatak se koristi za trening. Pri podeli baze se vodilo računa da se svi podaci jednog pacijenta nalaze u okviru istog dela. Na kraju se za procenu performansi modela koristi prosečna vrednost korišćenih mera tačnosti radi dobijanja pouzdanijih rezultata.

Postoji nekoliko modela *Segformer*-a koji se razlikuju po složenosti arhitekture i broju slojeva koje imaju. Najjednostavniji je MiT-B0 a najsloženiji je MiT-B5. Jedina izmena u odnosu na polaznu arhitekturu je ta da je prilagođen broj klasa na 2 (1 lezija, 0 nije lezija).

4.1. Korišćeni parametri i mere uspešnosti

Svi modeli su obučavani 100 epoha (što je na osnovu krive funkcije cene i Jaccardovog indeksa na validacionom skupu u zavisnosti od broja epoha, pokazalo više nego dovoljno). Za funkciju cene izabrana je binarna uzajamna entropija. Optimizator koji je korišćen je *Adam* sa brzinom učenja 6×10^{-5}. Mere tačnosti koje su korišćene pri selekciji najboljeg modela su *Jaccard Index* i *Dice Score*. Definicije ovih metrika date su sledećim formulama:

$$\text{Jaccard} = \frac{A \cap B}{A \cup B} \tag{9}$$

$$\text{Dice} = \frac{2 \cdot (A \cap B)}{|A| + |B|} \tag{10}$$

gde smatramo da je A maska koja je predviđena od strane modela, a B predstavlja masku koja je istinita.

U okviru svakog procesa obuke na kraju je izabran kao konačni model onaj koji je dao najbolju *Jaccard Index* meru uspešnosti na validacionom skupu.

5. REZULTATI

Za sve nivoe složenosti arhitekture, trenirano je po pet različitih modela, kao što je ranije navedeno. U tabeli 1, dat je pregled vrednosti prosečnih mera uspešnosti, za sve modele koji su trenirani, kao i broj parametara svakog modela. Vrednosti koje su prikazane odnose se na uspešnost klasifikacije piksela koji pripadaju klasi lezija.

Analizirajući dobijene rezultate, uočavamo da se prosečne vrednosti dobijenih metrika povećavaju do MiT-B3 modela, dok nakon njega, za modele MiT-B4 i MiT-B5, se smanjuju. Do toga verovatno dolazi usled toga što komplikovanije arhitekture imaju mogućnost da zapaze specifične karakteristike koje ne oslikavaju celu klasu, i samim tim je lošija generalizacija. Najlošije rezultate daje MiT-B0, koji ima najjednostavniju arhitekturu i najmanje parametara.

Na slici 3, prikazana je originalna slika, konture očekivane segmentacije zelenom bojom, i segmentacije predviđene od strane modela, crvenom bojom. Uočava se da je model pored lezije segmentovao i još jedan deo tkiva. Ovo je greška koja se uočava na određenom broju slika, koju verovatno uzrokuje sličnost u strukturi tkiva. Ono što je dobro u ovakvim slučajevima jeste to što tumor jeste detektovan i to sa velikom preciznošću.

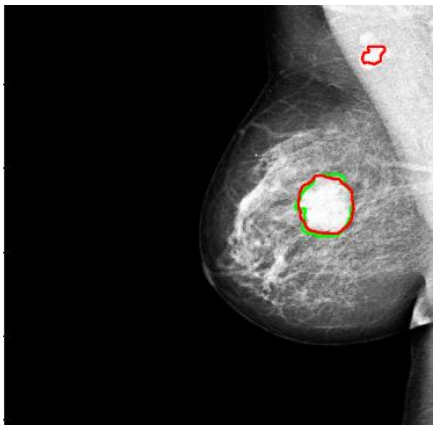
Tabela 1. *Prikaz dobijenih rezultata i mera uspešnosti za svaki od modela i broj parametara modela*

		Jaccard Index	Dice Score	Broj parametara
MiT-B0	**1**	54.96%	81.84%	3.8 M
	2	54.03%	82.29%	
	3	42.59%	76.31%	
	4	51.15%	80.93%	
	5	49.82%	79.07%	
Prosečna vrednost		**50.51%**	**80.09%**	
MiT-B1	**1**	51.31%	80.03%	13.7 M
	2	62.17%	86.20%	
	3	51.92%	81.05%	
	4	59.34%	85.45%	
	5	46.67%	77.43%	
Prosečna vrednost		**54.28%**	**82.03%**	
MiT-B2	**1**	58.07%	83.80%	27.5 M
	2	64.11%	86.94%	
	3	54.65%	82.27%	
	4	64.04%	87.59%	
	5	49.05%	78.46%	
Prosečna vrednost		**57.98%**	**83.81%**	
MiT-B3	**1**	58.94%	84.73%	47.3 M
	2	73.35%	91.99%	
	3	62.57%	86.27%	
	4	61.22%	86.16%	
	5	50.24%	78.70%	
Prosečna vrednost		**61.26%**	**85.57%**	
MiT-B4	**1**	58.00%	84.16%	64.1 M
	2	72.63%	90.76%	
	3	58.14%	83.16%	
	4	62.51%	86.79%	
	5	51.42%	79.45%	
Prosečna vrednost		**60.54%**	**84.86%**	
MiT-B5	**1**	61.11%	85.54%	84.7 M
	2	55.06%	81.87%	
	3	58.59%	83.51%	
	4	64.66%	87.64%	
	5	53.82%	80.11%	
Prosečna vrednost		**58.65%**	**83.73%**	

Iako MiT-B3 daje najbolje rezultate, treba napomenuti da on ima oko dvanaest puta više parametara od MiT-B0 što u značajnoj meri utiče na brzinu modela.

6. ZAKLJUČAK

U ovom radu je detaljno analizirana arhitektura *Segformer* modela i opisan je svaki od modula koji se u njoj nalazi. U nastavku je prikazano kako se ovaj model može prilagoditi na bazu mamografskih snimaka INBreast kako za potrebe segmentacije lezija. Na kraju su prikazani, opisani i upoređeni rezultati koji su dobijeni sa MiT B0-B5 modelima. Broj slika koje su bile dostupne za obuku modela je veoma mali, te bi sa povećanjem ovog skupa rezultati verovatno mogli biti značajno bolji. Rezultati koji su postignuti sa MiT-B3 bolji su od svih ostalih modela, dok MiT-B0 daje najlošije rezultate, koji se po *Jaccard Index*-u razlikuju od najboljeg modela za čak 10.75%.



Slika 3. *Primer gde je model pored lezije označio i još jedan deo tkiva lezijom. Očekivana segmentacija je označena zelenom bojom, predviđena crvenom. Predviđanje MiT-B3 modela.*

7. LITERATURA

[1] E. Xie, W. Wang et al., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *arXiv* (Cornell University), Dec. 2021, doi: https://doi.org/10.48550/arxiv.2105.15203

[2] K. H. Zhang0_0, "An Overview of Segformer and Details Description," GitHub, Apr. 15, 2023. https://github.com/ACSEkevin/An-Overview-of-Segformer-and-Details-Description (accessed Apr. 26, 2023).

[3] W. Wang et al., "Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions," *IEEE Xplore*, Oct. 01, 2021. https://ieeexplore.ieee.org/document/9711179

[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, et al. "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv*, 2020.

[5] I. C. Moreira, I. Amaral, I. Domingues et al. "INbreast: toward a full-field digital mammographic database," *Academic Radiology*. 2012 Feb;19(2):236-48. doi: 10.1016/j.acra.2011.09.014. Epub 2011 Nov 10. PMID: 22078258.

Kratka biografija:

Jovana Kljajić rođena je u Somboru 1999. god. Master rad na Fakultetu tehničkih nauka iz oblasti Elektrotehnike i računarstva – Komunikacione tehnologije i obrada signala odbranila je 2023.god.

kontakt: jovana.kljajic18@gmail.com