

MRI szegmentáció model ensemble módszerrel - Model Mavericks

Készítette:

Egyedi Zsolt - I9D6EJ

Rimai Dániel - BR2BUJ

Csató Erik - IRKR10

1. Feladatkíírás

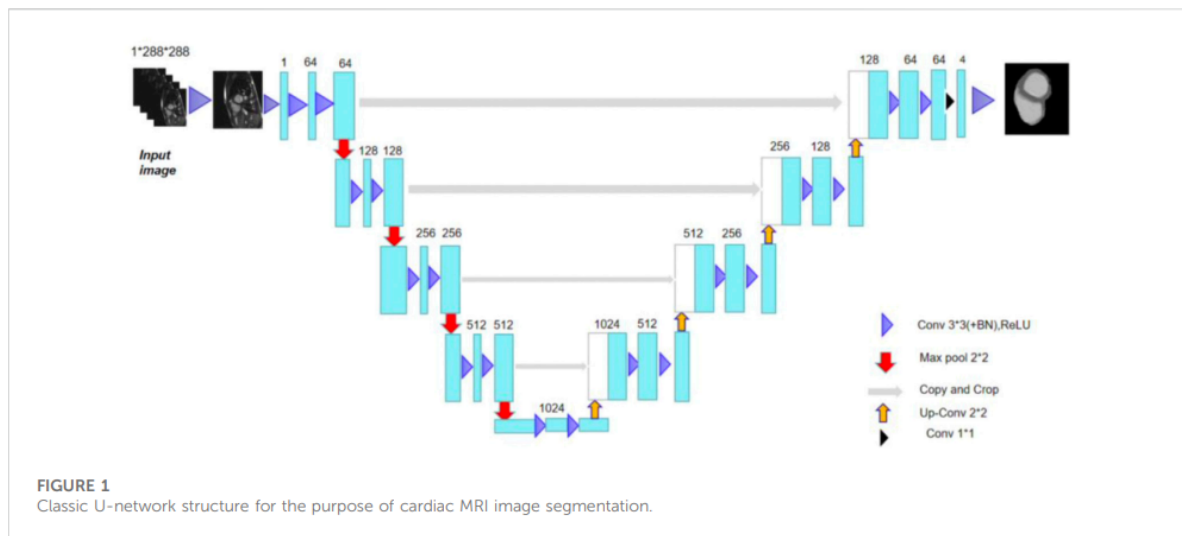
A szív kamráinak és a szívizomfalak térfogatának ismeretében bizonyos betegségek diagnosztizálhatóak. Ezek a régiók mellkasi MRI képeken jól látszanak. A pontos diagnózis felállítása érdekében elengedhetetlen a jobb kamrai endocardium, illetve a bal kamrai endocardium és epicardium pontos szegmentációja, így pontosan meghatározva a diagnosztikához szükséges térfogatokat. Feladatunk a fent említett három szerv szegmentálása az MRI felvételeken model-ensemble módszert használva. A model-ensemble módszer egy olyan módszer, mely során több különböző neurális hálót használnak a predikcióhoz. Előnye, hogy a több különböző modell a képeknek más-más tulajdonságaira tud rátanulni, így a több feature-ből előállított predikció pontosabb, mint az ensemble-t alkotó modellek predikciója külön-külön. Ennek az ára az, hogy így több modellt kell tanítani, illetve nagyobb számítási kapacitást igényel. Ennek ellenére gyakran használják a pontosabb predikció miatt.

2. Megoldás

2.1 Feladathoz használt architektúrák

A gyakorlatban a szív különböző szegmentációihoz nagyon sok különböző architektúrát használnak, ahogyan az a [1] cikkben is látható az első táblázatban. A táblázatban az is jól látszik, hogy a feladatunkhoz - ami a "ventricle segmentation" -

leggyakrabban használt architektúráinak egyike az ún. U-net architektúra. Ezt az architektúrát számos cikkben említik [2] [3] különböző felhasználással.



A U-net architektúra vázlatos rajza [4]

A U-net architektúra egy olyan tisztán konvolúciós neurális hálózat, amelynek lényege egy olyan enkóder-dekóder pár, melyben az azonos szinteken lévő enkóder és dekóder párok egy-egy skip-connection segítségével össze vannak kötve, ahogy ez az 1. ábrán is látható. Ezek a skip-connectionök segítik a gradiensek terjedését a mélyebben fekvő rétegek irányába. A hálózat enkóder része a képet a dimenziói mentén csökkenti, míg mélységében egyre növeli az adatmennyiséget. A bottleneck után található dekóder rész a mélység csökkentése közben állítja vissza az eredeti képméretet. A háló rétegei megtanulják a kép bizonyos jellemzőit, melyek segítségével utána a szegmentációs maszkot (vagy egyes esetekben jellemzőtérképeket (feature map)) elkészíti.

A szakirodalomban U-net architektúrából több különböző, ensemble módszerrel felépített modellt használnak. Vannak olyan megoldások, ahol a különböző U-net modellek által adott maszkok átlaga alapján készítik el a végső maszkot [2], míg másik megoldásnál egy úgynevezett különbözőség-támogató rendszerbe rakják, amelynél az egyes modelleket az ensemble rendszerbe egyesével rakják bele olyan sorrendben, hogy mindig a bentlévőktől legjobban különböző modellt adják hozzá [3]. Van egy úgynevezett Tri-U-net megoldás is, amelyben két U-net segítségével jellemzőtérképeket állítanak elő, majd később egy harmadik U-net segítségével ezekből a jellemzőtérképekből állítják elő a végső maszkot [2]. Mi is ez a Tri-U-net alapján készítettük el a megoldásunkat.

Ezeknek a U-net modelleknek a pontosságát legtöbb helyen a Dice érték segítségével értékelik ki, amely egy olyan metrika, ami a neurális hálózat által eltalált értékek arányát veszi figyelembe az összes pixelhez képest. A Dice értékeket szokták súlyozni az osztályok gyakoriságával, ha nem egyenletes az eloszlásuk, illetve figyelmen kívül hagyhatják a háttér osztályhoz tartozó pixelek helyes

csoportosítását. A különböző cikkekben található modellekkel általában 0.9 feletti, legtöbbször 0.92 és 0.95 közötti Dice értékeket érnek el. Ezen adatok alapján a projektünkben 0.85-ös Dice értéket szeretnénk megcélózni.

2.2 Az adatbázis

Az adatbázis a Dijoni kórház (Franciaország) 150 páciensének MRI felvételei [5]. A pácienseket úgy választották ki, hogy a náluk előforduló 4 féle szívbetegség (szívinfarktus, dilatív vagy tágulatos kardiomiopátia, hipertrófiás kardiomiopátia és abnormális jobb kamra működés) egyikével rendelkező páciensek képeiből, illetve egészséges páciensek képeiből 30-30 álljon rendelkezésre. Ezek a képek szét vannak osztva tanító (20-20 kép az egyes kategóriákból), illetve tesztelő (10-10 kép az egyes kategóriákból) adatsorra.

Az adatok NIFTI típusú három (illetve négy) dimenziós orvosi képek, amelyek MRI segítségével lettek rögzítve. A modellek tanításához kétféle dimenziós képeket akartunk használni, ezért mindegyik MRI felvételt a harmadik tengelye mentén feldaraboltuk. A képek dimenziói eltérő hosszúságúak, ezért ezeket egységes formára kellett hoznunk. Mindegyik képet kibővítettük annyi fekete képponttal, hogy az egyes dimenziók akkorák legyenek, mint az adatbázisban található legnagyobb dimenzió mérete. Végül, mindegyik képet normalizáltuk, hogy a pixelek nulla és egy közötti értékeket vegyenek fel.

2.3 Baseline model

Az baseline modell egy előre tanított ResNet hálózatot használ képi szegmentációs feladatokra. A modellt úgy módosítottuk, hogy az első konvolúciós réteg bemenetként 1 csatornát is elfogadjon (miután fekete-fehér képekkel dolgozunk), és a végső fully connected réteget a kimeneti osztályok számának megfelelően állítottuk be. A modellt 4 osztály megkülönböztetésére állítottuk be az egyes képtömbök szélességének és magasságának megfelelően. Dice Loss-t alkalmazunk a képekhez tartozó kimenetek predikciójának és a valós kimenetek átfedésének kiértékelésére. A baseline modellt végül nehézségek miatt nem sikerült túltanítani, de tervezzük annak továbbfejlesztését a közeljövőben.

2.4 Veszteség függvény

Az adathalmazban található képek esetében jóval nagyobb arányban vannak 0-as osztályok, mint 1-es, 2-es, illetve 3-as kimeneti értékek. Felmerül a kérdés, hogy az ilyen esetekben az osztályegyensúly hogyan kezelhető. Gyakori probléma

ilyenkor, ha az egyik osztály jelentősen nagyobb arányban van jelen a többihez képest, akkor a tanítás során egy egyszerűbb veszteségfüggvény (pl. Cross-Entropy) a többségi osztály felé hajlik, ami a kisebbségi osztályok teljesítményének romlásához vezethet. Például, ha a modell minden értéket 0-nak jósol (és nem tudja megtalálni a kisebbségi osztályokat, tehát az 1-es, 2-es és 3-as értékeket), akkor CE-vel közel tökéletes veszteség érhető el, ezáltal a modell megtanulja figyelmen kívül hagyni a kisebbségi osztályokat. A Focal Loss ezt a problémát úgy kezeli, hogy a jól osztályozott példák veszteségét csökkenti, lehetővé téve a modell számára, hogy a nehezen osztályozható példákra koncentráljon. A Focal Loss képlete az alábbi ábrán látható. Fontos megjegyezni, hogy a jobb eredmények elérése érdekében kísérletezni kell α és γ hiperparaméterek különböző értékeivel, és figyelni a modell teljesítményét a finomhangolás során. Az α finomhangolása segíthet az osztályegyensúly kezelésében, mivel megfelelő súlyokat rendelhet különböző osztályokhoz. Például a kisebbségi osztály esetében magasabb értékeket lehet megadni α -nak, hogy nagyobb fontosságot tulajdonítson neki a tanítás során. A γ választása szignifikánsan befolyásolja a veszteségfüggvény viselkedését. Magasabb γ értékek kiemelik a nehéz példák helyes osztályozását, míg alacsonyabb értékek veszteséget hasonlóvá teszik CE veszteségfüggvényhez. A γ optimális értéke a probléma konkrét jellegétől függően változhat.

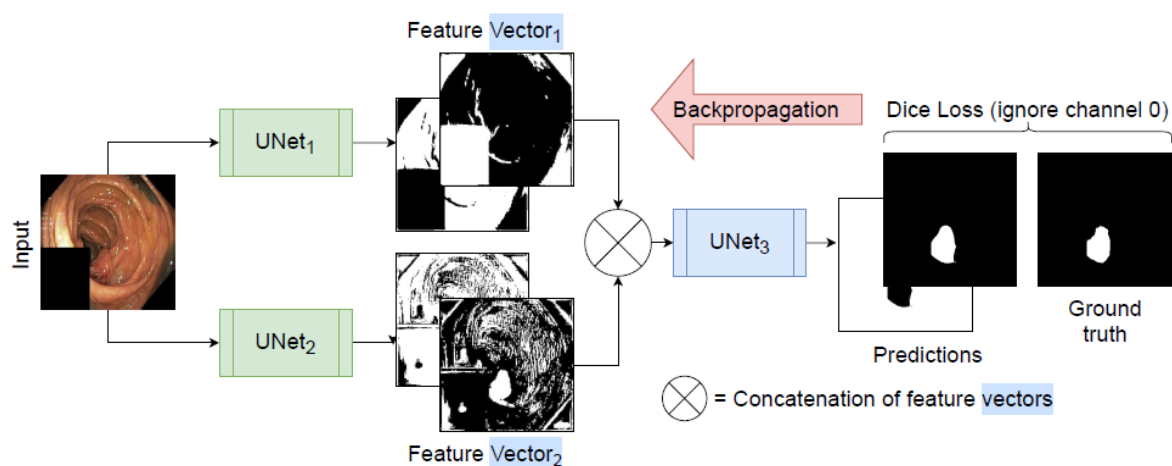
$$FL(p) = \begin{cases} -\alpha(1-p)^\gamma \log(p), & \text{if } y = 1 \\ -\alpha p^\gamma \log(1-p), & \text{otherwise} \end{cases}$$

Focal Loss képlete

A Focal Loss kipróbálása után, mégis a Dice Loss használata mellett döntöttünk, miután több a területen releváns cikkben is ezt a veszteség függvényt javasolták, továbbá ez egy gyakran használt metrika kép szegmentációs feladatokban. Ennek segítségével értékelhető a predikció és az elvárt kimenet közötti átfedés. Az együttható kiszámolja a két halmaz hasonlóságát azok metszetének és uniójának figyelembevételével. A Dice Loss használata előnyösebb a Focal Loss-al szemben bizonyos helyzetekben, mert beépítetten kezeli az osztályegyensúly problémát azzal, hogy kihangsúlyozza a kisebbségi osztályokat, melyek könnyen háttérbe szorulhatnak a Focal Loss pontatlan használatakor.

2.5. A Tri-U-net architektúra

A szegmentáció megoldásához a Tri-U-Net architektúrát választottuk, melynek vázlatos szerkezete a 3. ábrán látható. Ennek a modellnek a lényege, hogy az első két U-net struktúra a bejövő képek alapján elkészít 8-8 jellemzőtérképet, amelyek alapján egy harmadik U-net létrehozza a megfelelő maszkot. Azért ezt a modellt választottuk, mert az ensemble segítségével a különböző modellekkel pontosabb predikciót tudunk készíteni, ugyanakkor csak 3 modellből áll a rendszer, így nem lesz sokkal nagyobb az erőforrás igénye, mintha csak 1 modell adna predikciót.



A Tri-U-net architektúra vázlatos felépítése [2]

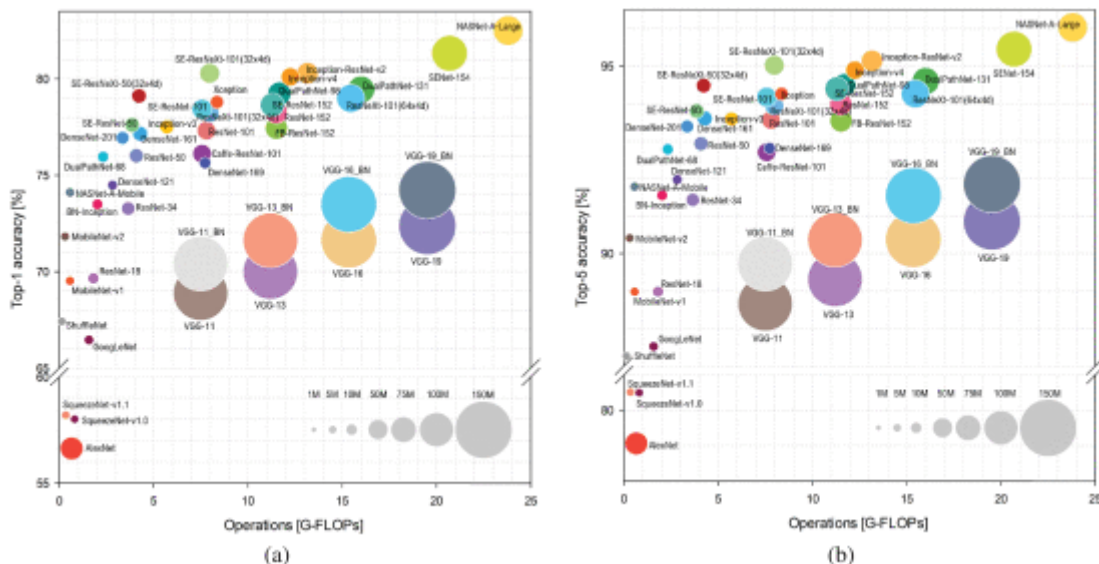
A pontosabb predikció elérése érdekében úgy terveztük, hogy az első két U-net enkódere egy-egy előtanított háló legyen. A két előtanított hálót a 4. ábra segítségével választottuk ki azon szempontok szerint, hogy 1) melyekre hivatkoztak cikkekben, illetve 2) olyan előtanított modelleket kerestünk, amelyek viszonylag kis számítás igényrel jó pontosságot értek el az ImageNet-1k validációs adathalmazon. Így esett a választásunk az [1] cikkben említett DenseNet-re (DenseNet201-re), valamint a [3] cikkben említett ResNet-re (a cikkben ResNet34 volt említve, viszont a 3. ábrán látható módon a ResNet50 egy kis plusz erőforrás igényért cserébe jelentősen nagyobb pontosságot ér el).

Az első két U-net egy konvolúciós réteggel kezdődik, amely az egycsatornás fekete-fehér MRI képeket átalakítja 3 csatornás képekké, amelyeket a fent említett két előtanított háló igényel (mert RGB-s képeken voltak előtanítva). Ezeket a 3 csatornás képeket kapják meg az előtanított modellek, amelyek konvolúciós rétegek segítségével az adatokat tömörítik jellemzővektorokká. A DenseNet a jellemzővektorok előállításához a kép szélességét és magasságát 16-od részére nyomja össze, míg a csatornaszámot 1920-ra növeli. A ResNet ezzel szemben a

képet 32-ed részére nyomja össze, míg a csatornaszámot 2048-ra növeli. Az előtanított rétegeket egy-egy dekóder rész követi Conv2d és Conv2dTranspose rétegekből, amelyek visszaállítják a képek eredeti méretének megfelelő jellemzőterképeket. Az egyes U-netek 8-8 jellemzőterképet állítanak elő, melyet 1-1 nyolccsatornás tenzorban tárolnak. A harmadik U-net egy általunk összerakott U-net, mely ezeket a jellemzővektorokat enkódolja, majd végül az enkódolt vektorból a dekóder segítségével előállítja a szegmentációs maszkot.

A Tri-U-net modell egyben van tanítva, az egyes U-netek külön-külön nem voltak tanítva. A tanítást a Dice Loss szerint végeztük, a kiértékelés során viszont figyelembe vettük a confusion mátrixot is.

A Colab-on lévő erőforrás keretek miatt az előtanított hálókkal nem volt lehetőségünk kísérletezni, így készítettünk egy olyan Tri-U-net-et is, amely három, a 3. U-netként felhasznált saját U-netből áll. Jelen állapotban sajnos a baselinehoz hasonlóan ezt sem sikerült túltanítanunk egy batch-en, ezen szintén a közeljövőben még dolgozni tervezünk.



2.6. Felhasználói felület:

A felhasználói felület kettő fő panelből áll. A bal oldali felületet használva lehet MRI felvételt tartalmazó fájlt feltölteni. A fájl formátuma a következő kiterjesztések valamelyike lehet: nii, nii.gz. Az inputnak megadott fájlt felhasználva fogja a neurális háló szegmentálni a szív megfelelő tartományait. A prediktált területek a bemeneten megadott képen lesznek kiszínezve, majd a jobb oldali panelen fog megjelenni a szegmentált kép. A bal oldali gombot megnyomva a kiválasztott fájl törlődik és lehetőség van új fájl kiválasztására.

Megjegyzés: ChatGPT-t használtunk Python kód és komment generálásra. Az eredmények minden esetben ellenőrizve voltak és szükség esetén javítva lettek.

Források:

[1] [Deep Learning for Cardiac Image Segmentation: A Review - by Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai and Daniel Rueckert](#)

[2] [DivergentNets: Medical Image Segmentation by Network Ensemble - by Vajira Thambawita, Steven A. Hicks, Pål Halvorsen and Michael A. Riegler](#)

[3] [Diversity-Promoting Ensemble for Medical Image Segmentation - by Mariana-Iuliana Georgescu, Radu Tudor Ionescu and Andreea-Iuliana Miron](#)

[4] [Deep learning-based image segmentation model using an MRI-based convolutional neural network for physiological evaluation of the heart - by Wanni Xu, Jianshe Shi, Yunling Lin, Chao Liu, Weifang Xie, Huifang Liu, Siyu Huang, Daxin Zhu, Lianta Su, Yifeng Huang, Yuguang Ye and Jianlong Huang](#)

[5] [O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, et al. "Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved ?" in IEEE Transactions on Medical Imaging, vol. 37, no. 11, pp. 2514-2525, Nov. 2018](#)

[6] [Benchmark Analysis of Representative Deep Neural Network Architectures - by Simone Bianco, Remi Cadene, Luigi Celona, Paolo Napoletano, et al.](#)