

Universitatea de Vest din Timișoara  
Facultatea de Matematică și Informatică  
Programul de studii  
Inginerie Software

**Algoritmi metaeuristici**

# **Image Segmentation using U-net**

Titular curs:  
Conf. Dr. Flavia MICOTA

Studenti:  
Ceban Cristian  
Savilencu Nicolae

1. Introducerea	3
1.1.Sinteză	3
2. U-Net	3
3. Aplicația	5
3.1.Datele	5
3.2.Implementarea	5
3.3.Concluzie	9
4. Bibliografia	10

# 1. Introducerea

Scopul cercetării este cel de a demonstra segmentarea imaginilor pe un set de date relativ mic, care constituie din mașini, folosind o rețea neuronală convulsivă U-net, și de a ajunge la concluzia dacă astfel de algoritmi pot fi folosiți pentru optimizarea problemelor de segmentarea imaginilor.

## 1.1. Sinteza

Rețelele codificator-decodificator sunt abordări de ultimă oră pentru segmentarea imaginilor biomedicale, dar prezintă două probleme: operațiile de grupare utilizate pe scară largă pot elimina informații spațiale și prin urmare, se pierde semantica de nivel scăzut. Metodele de fuziune a caracteristicilor pot atenua aceste probleme, dar caracteristicile hărțile de caracteristici la scări diferite nu pot fi fuzionate cu ușurință, deoarece reducerea și creșterea eșantionării modifică rezoluția spațială a hărții de caracteristici.

Implementarea de segmentare de imagini utilizând arhitectura U-Net a devenit populară în ultimii ani datorită capacității sale de a obține rezultate precise în diverse domenii, cum ar fi medicina, agricultura și analiza de date geospațiale. U-Net a fost introdusă pentru prima dată în 2015 de către Olaf Ronneberger et al. și este bazată pe o arhitectură de encoder-decoder cu conexiuni skip între straturile encoderului și decoderului. Aceasta permite modelului să păstreze informații detaliate din imaginea de intrare în timp ce se reduce dimensiunea acesteia.

Studiile recente au arătat că U-Net poate fi îmbunătățit prin adăugarea de module de atenție, utilizarea de date suplimentare pentru antrenare și combinarea cu alte metode de segmentare. De exemplu, unii cercetători au combinat U-Net cu metode de învățare profundă, cum ar fi RetinaNet, pentru a obține o precizie îmbunătățită a segmentării.

## 2. U-Net

U-Net este o arhitectură de rețea neuronală convoluțională pentru segmentarea imaginilor. A fost dezvoltată de Olaf Ronneberger, Philipp Fischer și Thomas Brox în 2015. Forma de "U" a rețelei este formată din calea de eșantionare downsampling (partea stângă a "U"-ului) și calea de eșantionare upsampling (partea dreaptă a "U"-ului).

În segmentarea imaginilor se utilizează metode de micșorare și de supraeșantionare pentru a controla rezoluția spațială a hărților de caracteristici.

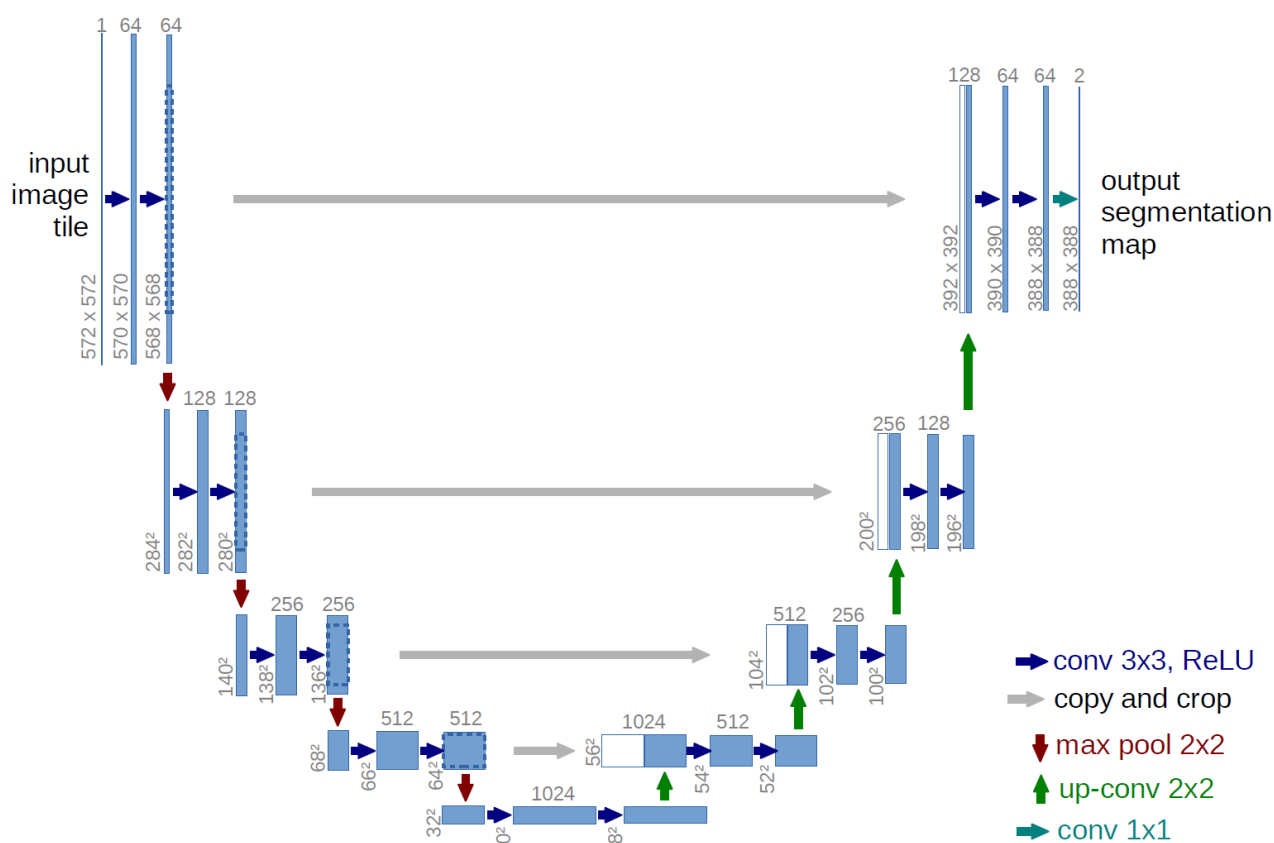
Downsampling, cunoscută și sub denumirea de "pooling", reduce rezoluția spațială a hărților de caracteristici prin combinarea informațiilor de la mai mulți pixeli adiacenți. Acest lucru are ca efect reducerea numărului de parametri și de calcule în rețea, crescând în același timp dimensiunea câmpului receptiv și făcând rețeaua mai rezistentă la mici translații ale datelor de intrare. Cu toate acestea, prin reducerea rezoluției spațiale, rețeaua poate pierde detalii fine care sunt importante pentru o segmentare precisă.

Upsampling, cunoscută și sub denumirea de convoluție transpusă, crește rezoluția spațială a hărților de caracteristici prin adăugarea de mai mulți pixeli. Acest lucru permite rețelei să recupereze detaliile fine care au fost pierdute în timpul procesului de reducere a eșantionării. Cu toate acestea, crește, de asemenea, numărul de parametri și de calcule în rețea.

Prin utilizarea unei combinații de eșantionare descendentă și ascendentă în arhitectura U-Net, rețeaua este capabilă să mențină un echilibru între eficiența de calcul și capacitatea de a păstra detaliile cu granulație fină. Calea de eșantionare descendentă este utilizată pentru a extrage caracteristicile de nivel înalt și pentru a reduce rezoluția spațială, în timp ce calea de eșantionare ascendentă este utilizată pentru a recupera detaliile de grad fin și pentru a crește rezoluția spațială. În plus, concatenarea hărților de caracteristici din stratul corespunzător în calea de eșantionare descendentă ajută rețeaua să utilizeze atât caracteristicile de nivel înalt, cât și cele cu granulație fină pentru o segmentare precisă.

Calea de supraeșantionare concatenează, de asemenea, hărțile caracteristice din stratul corespunzător din calea de subeșantionare, ceea ce ajută la păstrarea detaliilor fine în segmentarea finală. Arhitectura U-Net a avut succes într-o varietate de sarcini de segmentare a imaginilor, cum ar fi segmentarea imaginilor medicale, și este, de asemenea, utilizată pe scară largă în diverse aplicații, cum ar fi denaturarea imaginilor, generarea de imagini și alte sarcini de viziune computerizată.

Arhitectura rețelei este ilustrată în Figura 1. Ea constă dintr-un drum de contractare (partea stângă) și un drum de extindere (partea dreaptă). Drumul de contractare urmează arhitectura tipică a unei rețele convoluționale. Acesta constă din aplicarea repetată a două convoluții  $3 \times 3$  (convoluții ne-suprapuse), fiecare urmată de o unitate lineară rectificată (ReLU) și o operație de max pooling  $2 \times 2$  cu pasul 2 pentru downsampling. La fiecare pas de downsampling dublăm numărul de canale de caracteristici. Fiecare pas în drumul de extindere constă dintr-un upsampling al hărții de caracteristici urmat de o convoluție  $2 \times 2$  ("up-convoluție") care reduce la jumătate numărul de canale de caracteristici, o concatenare cu harta de caracteristici corespunzător tăiată din drumul de contractare și două convoluții  $3 \times 3$ , fiecare urmată de o ReLU. Tăierea este necesară din cauza pierderii pixelilor de la margine în fiecare



convoluție. La stratul final, o convoluție 1x1 este utilizată pentru a mapa fiecare vector de caracteristici cu 64 de componente la numărul dorit de clase. În total, rețeaua are 23 de straturi convoluționale.

Fig 1. Arhitectura U-net<sup>1</sup>

### 3. Aplicația

#### 3.1. Datele

Setul de date conține un număr mare de imagini de mașini (sub formă de fișiere .jpg). Fiecare mașină are exact 16 imagini, fiecare dintre ele fiind realizată din unghiuri diferite. Fiecare mașină are un id unic, iar imaginile sunt denumite în funcție de id\_01.jpg, id\_02.jpg ... id\_16.jpg. În plus față de imagini, se oferă și câteva metadata de bază despre marca, modelul, anul și ornamentul mașinii.

Pentru setul de antrenament, se furnizează un fișier .gif care conține masca decupată manual pentru fiecare imagine. Sarcina constă în segmentarea automată a mașinilor din imaginile din dosarul setului de testare.

#### 3.2. Implementarea

<sup>1</sup> <https://arxiv.org/pdf/1505.04597.pdf>



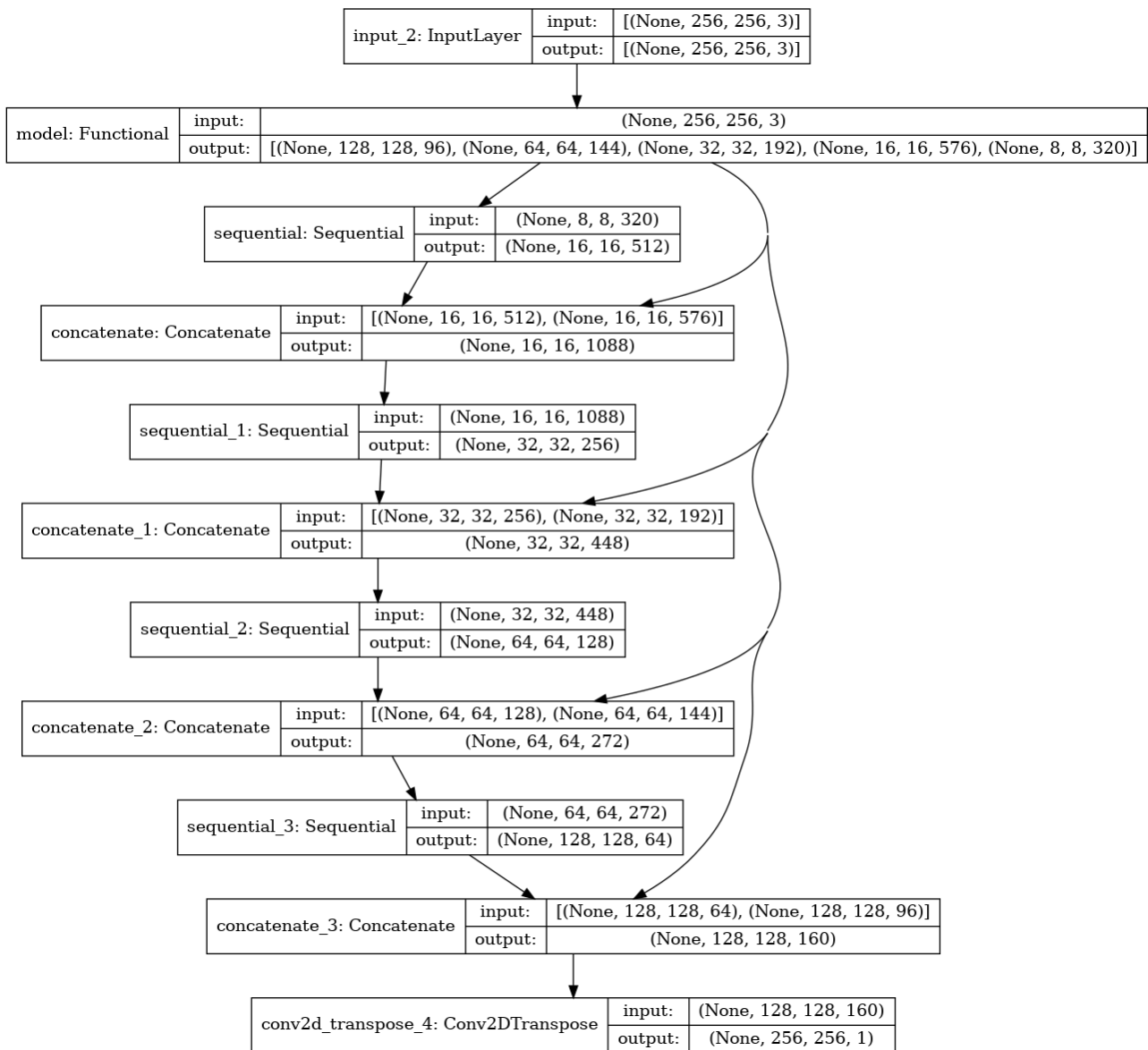
Fig 2. Exemplu de imagine

Pentru implementarea s-a folosit tensorflow, în date există imaginea inițială, și masca automobilului.

Se va folosi modelul U-net, pentru encoder (downsampler) v-a fi folosit MobileNetV2 ce este un model pre-antrenat pe alte date și valabil în `tf.keras.applications`. Arhitectura finală care va fi folosită poate fi văzută în figura 3. Se poate vedea clar câteva trepte de downsampling, ce constă în aplicarea repetată a două convoluții, fiecare urmată de o unitate lineară rectificată (ReLU) și o operație de Max pooling 2x2 cu pasul 2 pentru downsampling. Fiecare pas în drumul de extindere constă dintr-un upsampling al hărții de caracteristici urmat de o convoluție 2x2 (“up-convoluție”) care reduce la jumătate numărul de canale de caracteristici, o concatenare cu harta de caracteristici corespunzător tăiată din drumul de contractare și două convoluții 3x3, fiecare urmată de o ReLU. Tăierea este necesară din cauza pierderii pixelilor de la margine în fiecare convoluție. La stratul final, o convoluție 1x1 este utilizată pentru a mapa fiecare vector de caracteristici cu 64 de componente la numărul dorit de clase. În total, rețeaua are 23 de straturi convoluționale.

Dacă încercăm să facem o prezicere cu rețeaua (fără să fie antrenată pe datele noastre), rezultatul se poate vedea în figura 4. Se poate vedea clar că rețeaua identifică niște frontiere a obiectelor, dar nu este axată pe a identifica mașini încă. Scorul rețelei folosit este DICE COEFFICIENT<sup>2</sup>, care este foarte puternic pentru problemele de a compara agreement în funcție de pixel între o segmentare prezisă și adevărul de bază corespunzător. Se vor folosi relativ puține epoci (15) pentru a demonstra rezultatul, și se va folosi un early stop pentru a împiedica overfitting pentru datele de test.

<sup>2</sup> <https://www.kaggle.com/code/yerramvarun/understanding-dice-coefficient>



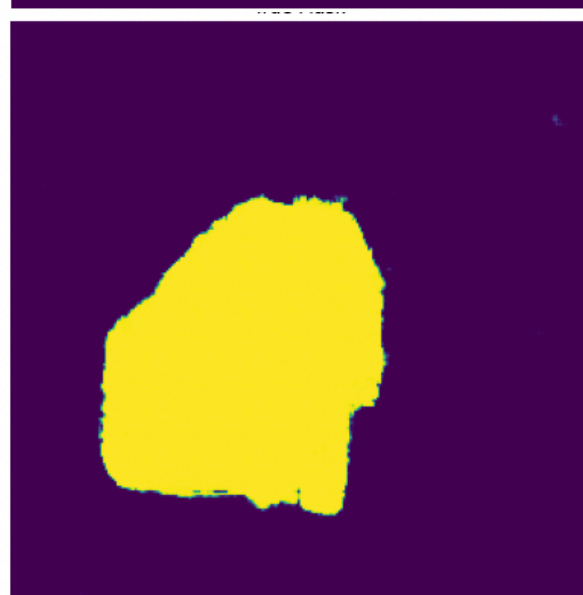
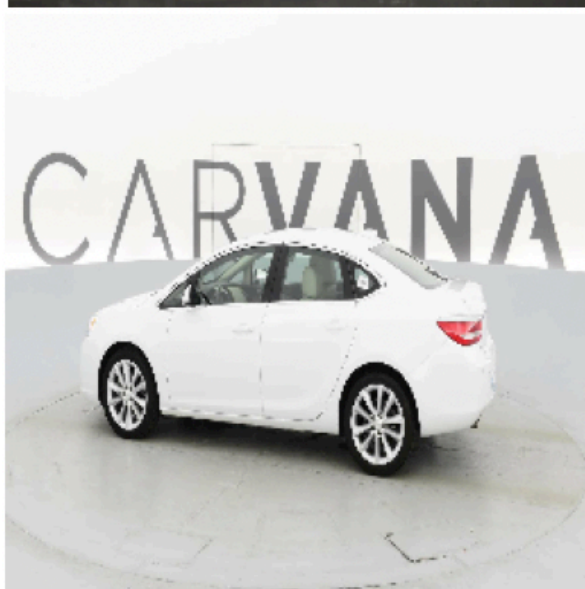


Fig 5. Prezicerile făcute



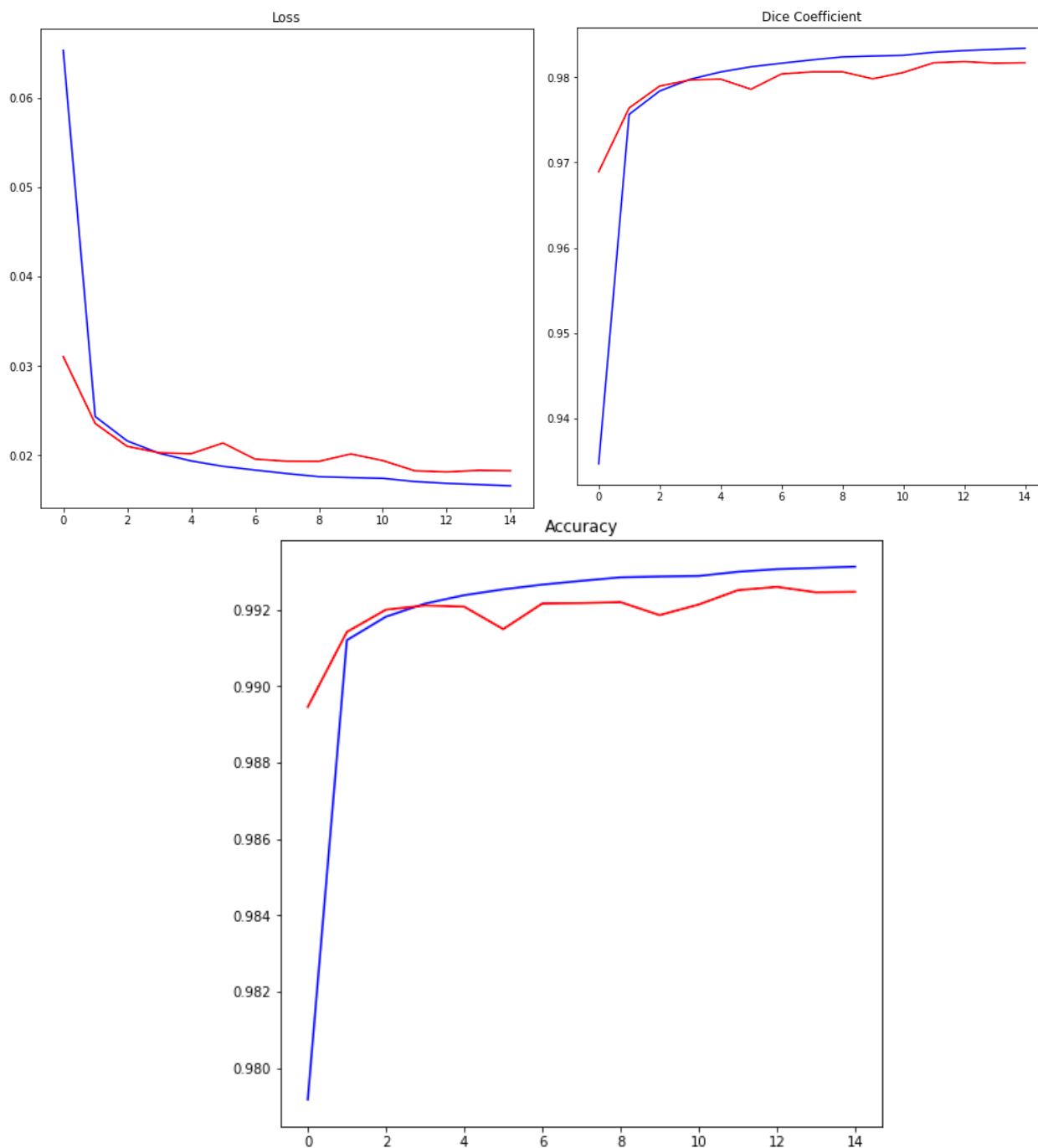


Fig 6. Câteva din metricile utilizate

### 3.3. Concluzie

Rețeaua are un scor mediu de 0.98 (dice coeff) și 0.99 (binary coeff), astfel poate prezice cu o exactitate ridicată dacă un pixel constituie o parte din mașină sau nu.

U-Net este o arhitectură puternică de rețea neuronală convoluțională care a fost adoptată pe scară largă în diverse domenii, în special în segmentarea imaginilor medicale, datorită capacității sale de a păstra detaliile cu granulație fină și de a lucra bine cu seturi de date mici. Aceasta a obținut rezultate de vârf în multe aplicații și a devenit o alegere populară în rândul cercetătorilor și practicienilor.

Unul dintre principalele avantaje ale U-Net este capacitatea sa de a gestiona cantități mari de date și poate fi antrenat pe seturi mari de date cu imagini de înaltă rezoluție, ceea ce este important pentru aplicațiile din lumea reală. U-Net oferă, de asemenea, un echilibru între eficiența computațională și capacitatea de a păstra detaliile de finețe, ceea ce este esențial pentru multe aplicații bazate pe imagini.

În ceea ce privește direcțiile viitoare, arhitectura U-Net ar putea fi extinsă pentru a lucra cu date 3D, ceea ce ar fi util în aplicațiile de imagistică medicală, cum ar fi CT și RMN. În plus, încorporarea mecanismelor de atenție și a altor tehnici avansate, cum ar fi rețelele Generative Adversarial Networks (GAN), ar putea contribui la îmbunătățirea performanțelor U-Net în diverse aplicații.

În general, U-Net este o arhitectură versatilă care s-a dovedit a fi eficientă în diverse aplicații din lumea reală și care are un mare potențial de dezvoltare și îmbunătățire în viitor.

## 4. Bibliografia

- <https://www.tensorflow.org/tutorials/images/segmentation>
- <https://www.kaggle.com/competitions/carvana-image-masking-challenge/data>
- <https://en.wikipedia.org/wiki/U-Net>
- <https://arxiv.org/pdf/1505.04597.pdf>
- <https://www.kaggle.com/competitions/uw-madison-gi-tract-image-segmentation/discussion/319035>
- <https://www.kaggle.com/code/mikolajbuchwald/3d-mri-brain-tumor-segmentation-u-net>
- <https://medium.com/projectpro/u-net-convolutional-networks-for-biomedical-image-segmentation-435699255d26>
- <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9330594>
- <https://arxiv.org/abs/1801.04381>