

SEGMENTAREA SEMANTICĂ A IMAGINILOR DIGITALE DE DIFERITE DIMESIUNI

Alexandrina SOBOL^{1*}

Sergiu SCROB²

Inga LISNIC²

¹Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Informatică și Microelectronică, Departamentul Ingineria Software și Automatică, grupa TI-172, Chișinău, Republica Moldova

²Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Informatică și Microelectronică, Departamentul Ingineria Software și Automatică, Doctorand, Chișinău, Republica Moldova

*Autorul corespondent: Sobol Alexandrina, e-mail: alexandrina.sobol@isa.utm.md

Rezumat. În zilele noastre, segmentarea semantică este una dintre problemele cheie în domeniul viziunii computerului. Privind imaginea în întregime, segmentarea semantică este una dintre sarcinile de nivel înalt care deschide calea către înțelegerea completă a scenei. Importanța înțelegerii scenelor ca problemă principală a viziunii computerului este evidențiată de faptul că un număr tot mai mare de aplicații se alimentează din deducerea cunoștințelor din imagini. Unele dintre aceste aplicații includ interacțiunea om-calculator, realitatea virtuală, vehicule cu autovehicul, etc. În această epocă în continuă dezvoltare a inteligenței artificiale (IA), învățarea profundă va rezolva multe probleme de segmentare semantică, care sunt abordate folosind arhitecturi profunde. Cel mai adesea rețele neuronale convoluționale, care depășesc alte abordări de o marjă mare din punct de vedere al preciziei și eficienței.

Cuvinte cheie: procesarea imaginii, strat, pixel, rețele.

Introducere

Sarcina de detectare și clasificare automată a obiectelor este una dintre cele mai interesante sarcini ale calculatorului modern. În problemele de clasificare este necesar să se determine numai tipul obiectului descris, în problemele de detectare –construirea casetei de delimitare pentru toate obiectele de un anumit tip, în sarcina de segmentare semantică este necesar nu numai de a detecta și a clasifica obiectele, ci și a determina limitele acestora. Segmentarea semantică reprezintă gruparea imaginii pe mai multe straturi. Aceste straturi se creează cu ajutorul predicției la nivel de pixeli. Acești pixeli se izolează la nivel de clasă indiferent de numărul de instanțe. Astfel, sarcina segmentării semantice este cea mai dificilă sarcină a procesării imaginilor.

Dificultatea procesării este completată de variabilitatea ridicată a obiectelor din cadrul unei clase și asemănarea ridicată a elementelor obiectelor din clase diferite.

Majoritatea abordărilor pentru construirea algoritmilor de segmentare semantică includ următorii pași:

1. Pre-procesarea datelor
2. Segmentarea preliminară
3. Descrierea caracteristică
4. Instruirea programului și clasificarea
5. Post-procesarea bazată pe context

Se poate remarca faptul că algoritmi au o structură modulară, permit selectarea diferitelor metode în fiecare etapă și combinarea acestora.

Metodele clasice ale segmentării

Înainte de epoca învățării profunde, segmentarea folosea o mare varietate de tehnici de procesare a imaginilor, în funcție de zona de interes. Câteva metode populare ale vremii sunt:

- Segmentarea după nivelul de gri

Cel mai simplu mod de a segmenta semantic este de a codifica manual regulile sau proprietățile pe care trebuie să le îndeplinească o regiune pentru a i se atribui o etichetă specifică. Aceste reguli pot fi proiectate ca proprietăți de pixeli, de exemplu, intensitatea griului. Una dintre metodele care utilizează această tehnică se numește algoritmul Split and Merge. Acest algoritm împarte recursiv imaginile în subdomenii până când le este atribuită o etichetă specifică, apoi subdomeniile adiacente cu aceleași etichete sunt combinate.

Problema acestei metode este că regulile trebuie să fie specificate manual în cod. Cu toate acestea, uneori este incredibil de dificil de a descrie clase complexe, cum ar fi „oameni”, folosind doar informații despre intensitatea cenușiei. În consecință, în lucrul cu astfel de obiecte complexe, pentru predarea corectă a reprezentărilor, sunt necesare metode de extragere a caracterelor și de optimizare

- *Câmpuri aleatorii condiționate*

Luați în considerare segmentarea imaginii prin instruirea modelului pentru a atribui o clasă fiecărui pixel. Dacă modelul nu este perfect, puteți obține rezultate cu o segmentare zgomotoasă, ceea ce este adesea imposibil în natură (de exemplu, pixelii pisicilor sunt amestecate cu pixelii câinilor, așa cum se arată în imagine).

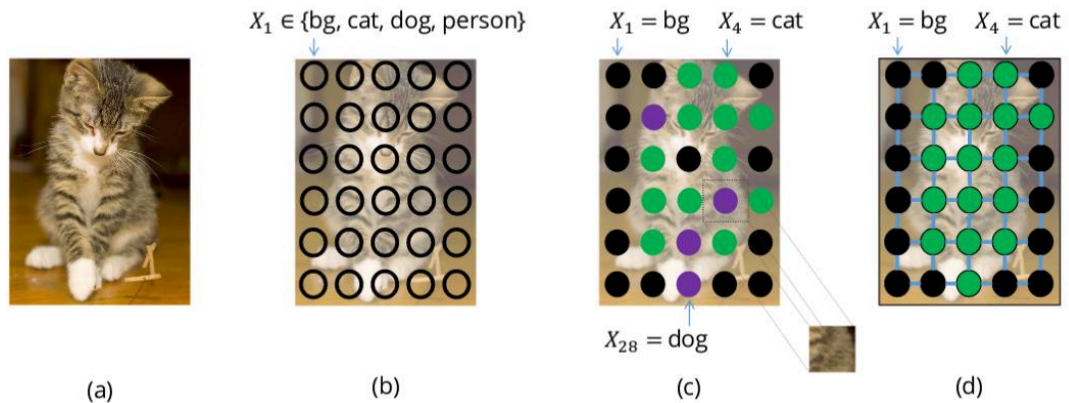


Figura 1. Modelul pentru metoda câmpurilor aleatorii condiționate [1]

Acest lucru poate fi evitat prin analizarea relațiilor anterioare dintre pixeli. Ideea este că obiectele sunt continue și, prin urmare, pixelii din apropiere trebuie să aibă aceleași etichete. Pentru modelarea acestei relații, se folosesc câmpuri aleatorii condiționate (USP).

USP-urile aparțin clasei de metode de modelare statică utilizate pentru predicții structurate. USP-urile iau în considerare contextul, adică relația dintre pixeli. Din această cauză, modelul este considerat un candidat ideal pentru segmentarea semantică a imaginilor.

Fiecare pixel din imagine este asociat cu un set finit de stări posibile. În cazul nostru, etichetele țintă vor fi un set de stări posibile. Costurile alocării unui stat (eticheta u) unui singur pixel (x) se numesc costuri unare. Pentru a modela relațiile dintre pixeli, avem în vedere și costurile, cunoscute sub numele de perechi, de atribuire a unei perechi de etichete (u, v) unei perechi de pixeli (x, y). Putem lua în considerare perechi de pixeli care sunt vecini imediați (USP grilă) sau putem să lucrăm cu toate perechile de pixeli de imagine (USP dens).

- *Metoda de învățare profundă*

Învățarea profundă a simplificat foarte mult etapele pentru segmentarea semantică, arătând în același timp o calitate impresionantă.

Una dintre cele mai simple și mai populare arhitecturi utilizate pentru segmentarea semantică este Rețeaua Complet Convoluțională (FCN) [2]. FCN este folosit pentru a transforma inițial imaginea de intrare la o dimensiune mai mică (în timp ce crește numărul de canale) printr-o serie de convoluții. Acest set de operații de convoluție este denumit în mod obișnuit un codificator. Apoi, ieșirea este decodată fie prin interpolare bilineală, fie printr-o serie de convoluții transpuse, care se numește decodificator.

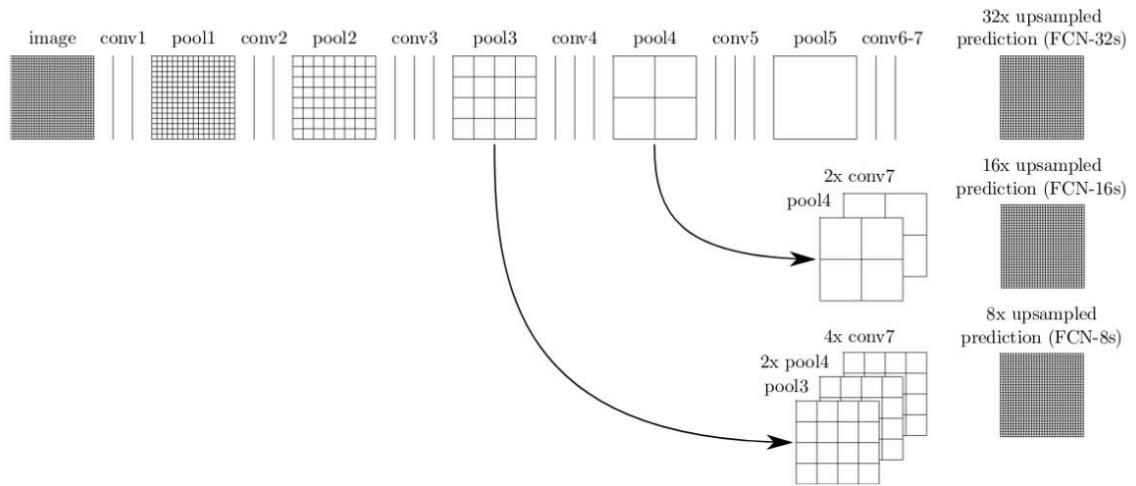


Figura 2. Conversia directă și inversă [1]

O astfel de arhitectură de bază, în ciuda eficacității sale, prezintă mai multe dezavantaje. Unul dintre acestea este prezența artefactelor, aranjate în ordine de șah, asociate cu suprapuneri neuniforme a ieșirilor într-o operație de convoluție transpusă.

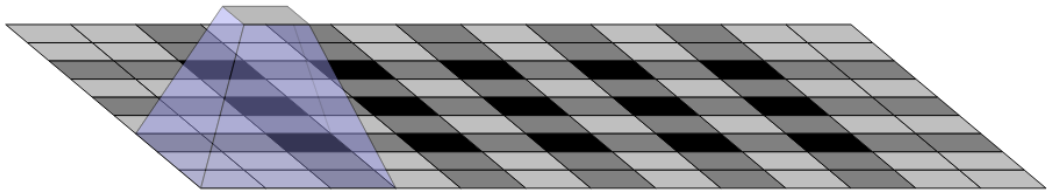


Figura 3. Formarea artefactelor, aranjate în ordine de șah [1]

Un alt dezavantaj este asociat cu rezoluția scăzută la margini din cauza pierderilor de informații în procesul de codificare. Pentru a crește calitatea modelului FCN de bază, au fost propuse mai multe soluții, cum ar fi: U-Net, modelul tiramisu, metoda pe mai multe nivele, metode hibride.

Învățarea profundă (Deep Learning)

- *Rețele neuronale convoluționale (CNN)* Segmentarea imaginii cu CNN implică alimentarea segmentelor unei imagini ca intrare într-o rețea neuronală convoluțională, care etichetează pixelii. CNN nu poate procesa întreaga imagine dintr-o dată. Scanează imaginea, uitându-se la un mic „filtru” de câțiva pixeli de fiecare dată până când a mapat întreaga imagine.

- *Rețele complet convolutive (FCN-uri)* CNN-urile tradiționale au straturi complet conectate, care nu pot gestiona diferite dimensiuni de intrare. FCN-urile folosesc straturi convoluționale pentru a procesa diferite dimensiuni de intrare și pot funcționa mai rapid. Stratul final de ieșire are un câmp receptiv mare și corespunde înălțimii și lățimii imaginii, în timp ce numărul de canale corespunde numărului de clase. Straturile convoluționale clasifică fiecare pixel pentru a determina contextul imaginii, inclusiv locația obiectelor.

Precizia pe pixel

Această măsurătoare este auto-explicativă, deoarece produce o precizie de predicție a clasei pe pixel, precum în Ec. (1) [3].

$$acc(P, GT) = \frac{|pixels\ correctly\ predicted|}{|total\ nb\ of\ pixels|} \quad (1)$$

Jaccard (intersecție asupra Uniunii)

Această măsurătoare de evaluare este adesea folosită pentru segmentarea imaginii, deoarece este mai structurată Jaccard-ul este o metrică de evaluare pe clasă, care calculează

numărul de pixeli din intersecția dintre hărțile de segmentare prezise și cele adevărate pentru o anumită clasă, împărțită la numărul de pixeli din uniunea dintre cele două hărți de segmentare, de asemenea pentru acea clasă, precum în Ec. (2) [3].

$$jacc(P(class), GT(class)) = \frac{|P(class) \cap GT(class)|}{|P(class) \cup GT(class)|} \quad (2)$$

unde P este harta de segmentare prevăzută și GT este harta de segmentare adevăratului. P (clasa) este apoi masca binară care indică dacă fiecare pixel este prevăzut ca clasă sau nu. În general, cu cât este mai aproape de 1, cu atât mai bine.

Concluzii

Algoritmii de învățare profundă au rezolvat mai multe sarcini de viziune pe calculator cu un nivel din ce în ce mai mare de dificultate, tehnologia de învățare profundă a realizat, de asemenea, mari realizări în segmentarea semantică a imaginilor. Pe baza rețelei neuronale complet revoluționare, mulți savanți continuă să încerce să optimizeze rețeaua, să învețe caracteristicile ierarhice și să o facă mare.

Segmentarea semantică a imaginii este o provocare abordată recent de rețelele neuronale profunde de la capăt la capăt. Una dintre problemele principale dintre toate arhitecturile este de a ține cont de contextul vizual global al intrării pentru a îmbunătăți predicția segmentării. Modelele de ultimă generație folosesc arhitecturi care încearcă să lege diferite părți ale imaginii pentru a înțelege relațiile dintre obiecte.

Conceptele de aplicație prezentate în articol nu sunt cuprinzătoare, deoarece comunitatea de cercetare se străduiește constant să aprofundeze și să extindă cunoștințele în acest domeniu.

Referințe:

1. Семантическая сегментация: краткое руководство. [online] [accesat 27.02.2020]. Disponibil: <https://neurohive.io/ru/osnovy-data-science/semantic-segmentation/?fbclid=IwAR3KmFY16iyhiOE9wIXtFRxWjs8PhZaCLCKfeFDw7PVp6O3DNEh7aPDxLY>
2. Semantic Segmentation — Popular Architectures. [online] [accesat 27.02.2020]. Disponibil: <https://towardsdatascience.com/semantic-segmentation-popular-architectures-dff0a75f39d0>
3. Fully Convolutional Networks (FCN) for 2D segmentation. [online] [accesat 27.02.2020]. Disponibil: http://deeplearning.net/tutorial/fcn_2D_seg.html