

Üzleti Intelligencia

8. Előadás: Objektum detekció

Kuknyó Dániel
Budapesti Gazdasági Egyetem

2023/24
1. félév

1 Bevezetés

2 Objektum detekció

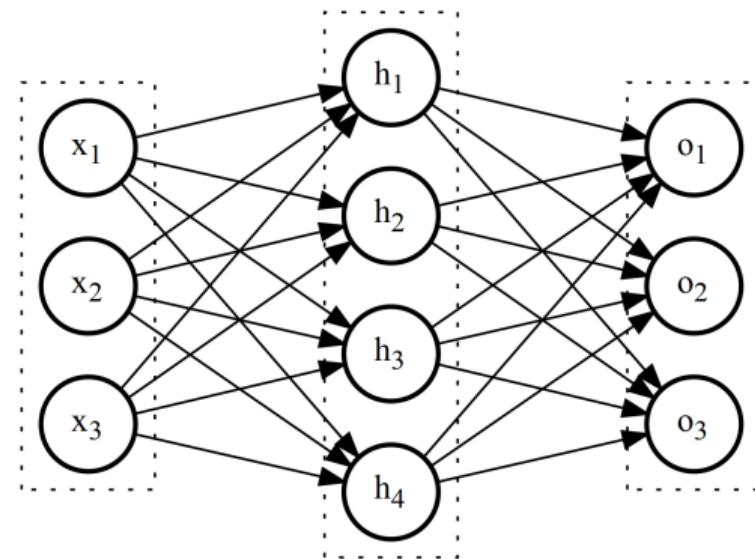
1 Bevezetés

2 Objektum detekció

Osztályozás

Osztályozás esetén a neurális hálózat output rétegében minden lehetséges osztályhoz egy neuron tartozik. A neuronok azt a valószínűséget becsülik meg, hogy az adott mintaegyed mekkora valószínűsséggel esik a neuronhoz tartozó kategóriába.

Ez a softmax réteg, ahol az egyes neuronok aktivációs függvénye a softmax függvény. A softmax rétegen lévő neuronok outputjainak összege mindig 1.



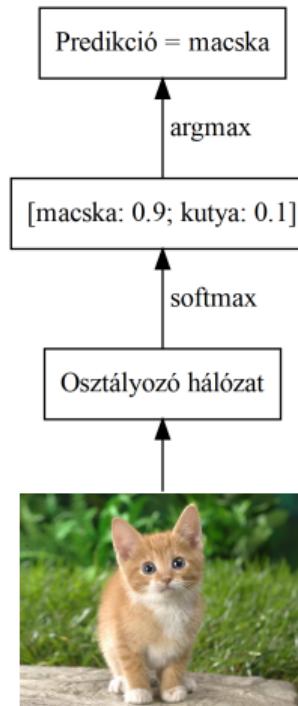
Predikció osztályozás esetén

A legegyszerűbb módja az osztályozásnak, ha a hálózat outputként **összesen egy címkét** ad outputként, ami a predikciója az input képre (vagy adatra) vonatkozóan.

Ebben az esetben a hálózat softmax rétege azt becsüli meg, hogy **mekkora valószínűséggel tartozik a mintaegyed a tanító osztályok valamelyikébe**. Például:

$$[\text{macska} : 0.9; \text{kutya} : 0.1]$$

Ezután az output úgy áll elő, hogy a hálózat **kiválasztja a legnagyobb valószínűségű osztályt** az *argmax* operátorral, majd visszaadja a legnagyobb valószínűséghez tartozó címkét.



Lokalizáció

Az objektum lokalizáció feladata az osztályozás egy speciális esete. A lokalizáció problémájában nem csak azt kell megbecsülnie a neurális hálózatnak, hogy milyen objektum található egy képen, hanem meg is kell adnia az **objektum pozícióját azzal, hogy megadja a kereteződobozának koordinátáit**.

Ebben az esetben a hálózat outputja nem csak egy címke, hanem a dobozt meghatározó négy koordináta is. Ezek a leggyakrabban:

- x : A doboz bal felső sarkának x koordinátája.
- y : A doboz bal felső sarkának y koordinátája.
- w : A doboz szélessége (width).
- h : A doboz magassága (height).

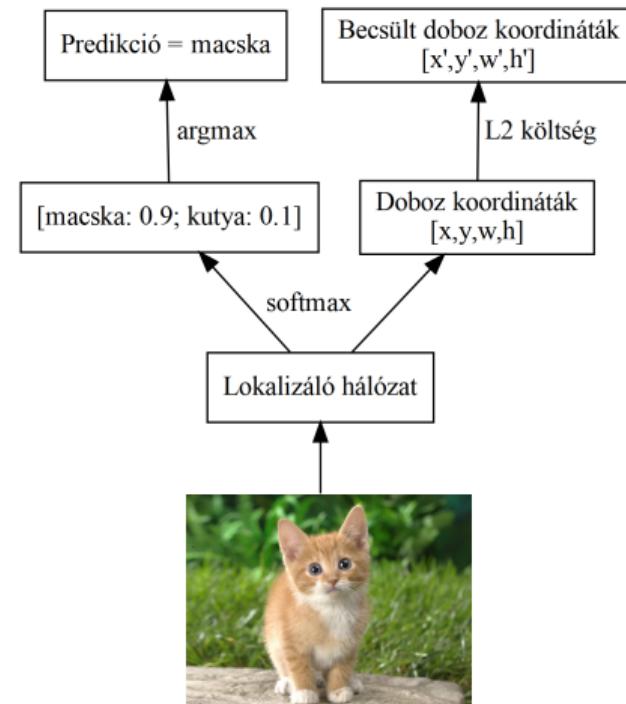


Predikció lokalizáció esetén

Lokalizáció esetén a neurális hálózat két részre ágazik, mivel két különböző feladatot kell elvégeznie:

Osztályozás: ez megegyezik az osztályozó hálózat által végzett feladattal. A hálózat megbecsül egy valószínűségeszlást, majd ebből kiválasztja a legnagyobb valószínűséghez tartozó osztályt, amit visszaad mint predikció.

Kereteződoboz predikciók: a lokalizáció feladatához a neurális hálózatnak meg kell becsülnie az érdekelt régiók dobozainak koordinátáit. A neurális hálózat kiválasztja a becsült kereteződobozok közül azt, amelyik a **legnagyobb valószínűséggel tartalmazza a keresett objektumot**.



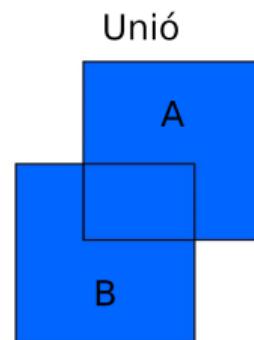
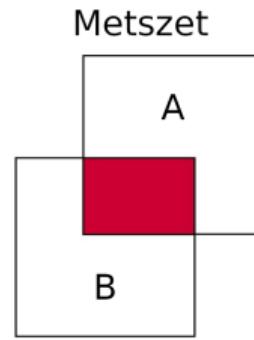
Pontosság mérése kereteződobozokkal

Mivel az objektum detekció egy felügyelt tanulási algoritmus, **szüksége van címkékre**. Ebben az esetben a címkék a valós osztályt és a kereteződoboz x, y, w, h koordinátáit tartalmazzák.

A jóságfüggvény A valós és B becsült kereteződoboz koordinátáit hasonlítja össze az IoU (metszet/unió) mérőszámmal:

$$IoU = \frac{A \cap B}{A \cup B}$$

Tehát minél jobban illeszkedik a becsült doboz a valósra, annál közelebb van a mérőszám 1-hez. Ha a két doboznak nincs közös területe, $IoU = 0$.



1 Bevezetés

2 Objektum detekció

Objektum detekció

Az objektum detekció a lokalizáció általánosítása több objektumra egyetlen képen belül. Ebben az esetben a neurális hálózat feladata **minden ismert objektum lokalizálása a képen.**

Ebben az esetben a hálózat outputja minden egyes ismert objektumra:

- Az objektum becsült címkéje.
- Az objektum kereteződobozának x, y, w, h koordinátái.

Mivel az érdekelt objektumok száma képenként eltérő lehet, az **objektum detekciós hálózatok outputjainak számossága is képenként különbözik!**



Kezdeti detektor implementáció

Az első objektum detektor modellek implementációi meglehetősen naívak voltak. Az eljárás szerint a modell végigcsúsztat egy előre meghatározott méretű ablakot a képen, és minden egyes általa meghatározott képről eldönti, hogy háttér-e (*bgr*) vagy valamelyik keresett osztályba tartozik-e:

$$output = [bgr : 0.8, kutya : 0.1, macska : 0.1]$$

Ebben az esetben $\hat{y} = \underset{p}{\operatorname{argmax}}(output) = bgr$, és az algoritmus eldobja a kereteződoboz predikcióját.



Kezdeti detektor implementáció

Az első objektum detektor modellek implementációi meglehetősen naívak voltak. Az eljárás szerint a modell végigcsúsztat egy előre meghatározott méretű ablakot a képen, és minden egyes általa meghatározott képről eldönti, hogy háttér-e (*bgr*) vagy valamelyik keresett osztályba tartozik-e:

$$output = [bgr : 0.0, \text{kutya} : 0.01, \text{macska} : 0.99]$$

Ebben az esetben $\hat{y} = \underset{p}{\operatorname{argmax}}(output) = \text{macska}$, és a becsült kereteződoboz az, ahol a legnagyobb a keresett osztály valószínűsége.



Miért nem lehetséges ez az implementáció?

Hány lehetséges kereteződoboz van egy $H \cdot W$ méretű képen?

Ha van egy $h \cdot w$ méretű kereteződoboz a lehetséges x koordináták száma $W - w + 1$, és a lehetséges y koordináták száma $H - h + 1$. A lehetséges megoldások száma egyetlen kereteződoboz méretre: $(W - w + 1) \cdot (H - h + 1)$.

Az összes lehetséges kereteződoboz méretre pedig:

$$\sum_{h=1}^H \sum_{w=1}^W (W - w + 1) \cdot (H - h + 1) = \frac{H(H+1)}{2} \frac{W(W+1)}{2}$$

Ez azt jelenti, hogy egy $800 \cdot 600$ méretű képre ~ 58 millió lehetséges kereteződoboz létezik.

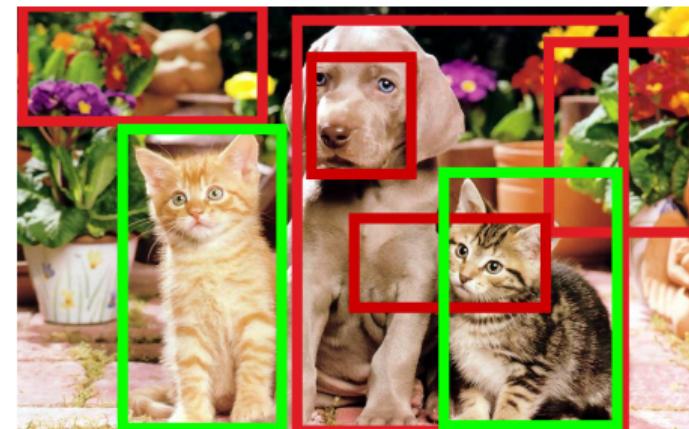


Területi javaslatok

Az eljárás ami meggyorsította a folyamatot a területi javaslatok algoritmusa volt.

Ahelyett, hogy egy előre meghatározott méretű ablakot csúsztatna végig egy képen, a területi javaslatok algoritmusa megpróbálja a kép **pixeleteit klaszterezni hasonlóság szerint**, ezzel területeket formázva.

Később ezen területeknek a kereteződobozai lesznek átadva egy osztályozó neurális hálózatnak ami eldönti, hogy a keresett objektumok valamelyike mekkora valószínűsséggel található meg a területen.



Területi javaslatok szelektív kereséssel

A szelektív keresés egy potenciális objektum területeket javasoló algoritmus, ami jóval hatékonyabban képes működni, mint a csúszóablakos algoritmus:

- ① **Kép szegmentálása:** Az input kép pixeleinek csoportosítása szín, textúra és egyéb vizuális tulajdonságok szerint.
- ② **Régiók hasonlósága:** A különböző régiók összeolvasztása valamilyen hasonlósági mérték szerint.
- ③ **Hierarchikus csoportosítás:** Az összeolvasztott régiók hierarchikus klaszterezése.
- ④ **Objektum javaslatok:** Objektum javaslatok halmazának elkészítése, amik területi javaslatként szolgálnak.
- ⑤ **Objektum detekció:** A területi javaslatok átadódnak a neurális hálózatnak osztályozásra.



Területi javaslatok szelektív kereséssel

A szelektív keresés egy potenciális objektum területeket javasoló algoritmus, ami jóval hatékonyabban képes működni, mint a csúszóablakos algoritmus:

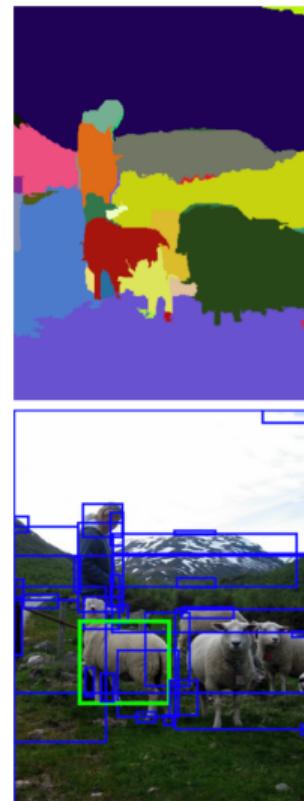
- 1 Kép szegmentálása:** Az input kép pixeleinek csoportosítása szín, textúra és egyéb vizuális tulajdonságok szerint.
- 2 Régiók hasonlósága:** A különböző régiók összeolvasztása valamilyen hasonlósági mérték szerint.
- 3 Hierarchikus csoportosítás:** Az összeolvasztott régiók hierarchikus klaszterezése.
- 4 Objektum javaslatok:** Objektum javaslatok halmazának elkészítése, amik területi javaslatként szolgálnak.
- 5 Objektum detekció:** A területi javaslatok átadódnak a neurális hálózatnak osztályozásra.



Területi javaslatok szelektív kereséssel

A szelektív keresés egy potenciális objektum területeket javasoló algoritmus, ami jóval hatékonyabban képes működni, mint a csúszóablakos algoritmus:

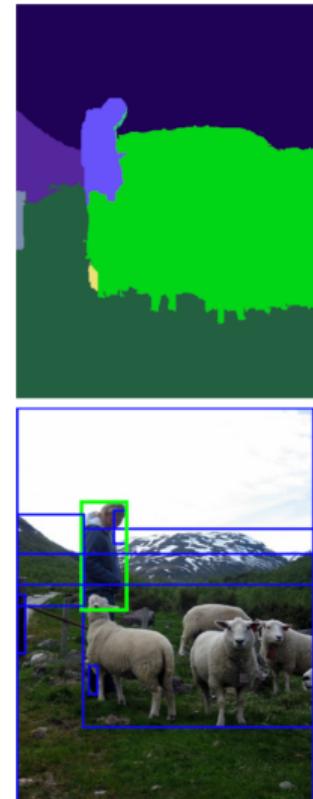
- ① **Kép szegmentálása:** Az input kép pixeleinek csoportosítása szín, textúra és egyéb vizuális tulajdonságok szerint.
- ② **Régiók hasonlósága:** A különböző régiók összeolvasztása valamilyen hasonlósági mérték szerint.
- ③ **Hierarchikus csoportosítás:** Az összeolvasztott régiók hierarchikus klaszterezése.
- ④ **Objektum javaslatok:** Objektum javaslatok halmazának elkészítése, amik területi javaslatként szolgálnak.
- ⑤ **Objektum detekció:** A területi javaslatok átadódnak a neurális hálózatnak osztályozásra.



Területi javaslatok szelektív kereséssel

A szelektív keresés egy potenciális objektum területeket javasoló algoritmus, ami jóval hatékonyabban képes működni, mint a csúszóablakos algoritmus:

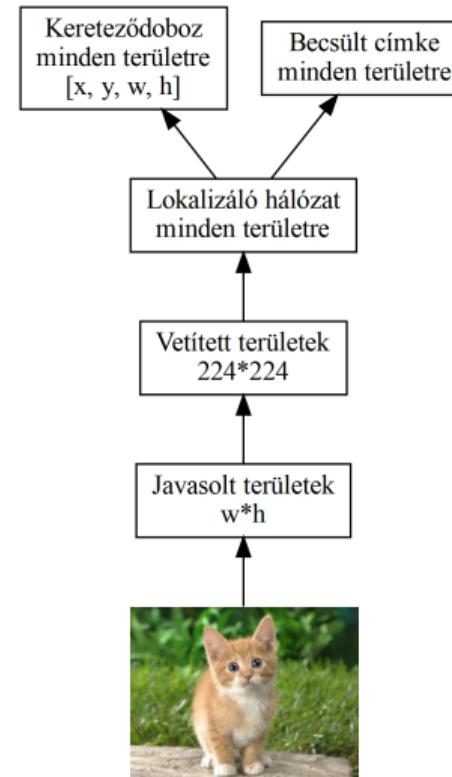
- 1 Kép szegmentálása:** Az input kép pixeleinek csoportosítása szín, textúra és egyéb vizuális tulajdonságok szerint.
- 2 Régiók hasonlósága:** A különböző régiók összeolvasztása valamilyen hasonlósági mérték szerint.
- 3 Hierarchikus csoportosítás:** Az összeolvasztott régiók hierarchikus klaszterezése.
- 4 Objektum javaslatok:** Objektum javaslatok halmazának elkészítése, amik területi javaslatként szolgálnak.
- 5 Objektum detekció:** A területi javaslatok átadódnak a neurális hálózatnak osztályozásra.



Predikció objektum detekció esetén

Az objektum detekció folyamata területjavasolással:

- ① A területjavasoló **kiválaszt érdekelt régiókat** a képről. Ezek lehetnek bármekkora méretűek.
- ② A javasolt régiók **fix méretűre vetítése**. Ez leggyakrabban $224 \cdot 224$ a gyors osztályozás miatt.
- ③ A lokalizációs hálózat **megbecsüli az osztályokat** és a **kereteződobozok koordinátáit** a vetített terület alapján



Predikció objektum detekció esetén

Az objektum detekció folyamata területjavasolással:

- ① A területjavasoló **kiválaszt** érdekelt **régiókat** a képről. Ezek lehetnek bármekkora méretűek.
- ② A javasolt régiók **fix méretűre vetítése**. Ez leggyakrabban $224 \cdot 224$ a gyors osztályozás miatt.
- ③ A lokalizációs hálózat **megbecsüli** az **osztályokat** és a **kereteződobozok koordinátáit** a vetített terület alapján

