

Absztrakt

Az első és egyben legelterjedtebb galaxis osztályozást Hubble alkotta meg 1926-ban, aki négy fő típust határozott meg: spirális, elliptikus, irreguláris és lentikuláris. Ezek a fő osztályok további al osztályokra bonthatók, melyek közül kifejezetten érdekesek és nehezen detektálhatók a gyűrűs galaxisok. A dolgozat keretein belül két módszer került kifejlesztésre a gyűrűs galaxisok detektálására: az első megközelítés egy módosított Faster R-CNN-t használt, míg a második a DBSCAN – CNN alapú hálózat volt, melyek a Sloan Digital Sky Survey program égbolt képeivel kerültek tanításra és tesztelésre. Az eredmények tekintetében a DBSCAN - CNN alapú megoldás bizonyult jobbnak, ami a fő osztályokat 99,7%-os, a gyűrűs galaxisokat 94,9%-os pontossággal határozta meg.

Kulcsszavak

galaxis, neurális hálózat, DBSCAN, Faster R-CNN

Motiváció

A csillagászok számára kulcsfontosságú a galaxis típusának ismerete, mivel egy adott kutatáshoz csak egy bizonyos fajtájú galaxisok felelnek meg. Jelenleg körülbelül 2 trillió galaxis létezéséről tudunk, amit teleszkópjaink évente milliárdnyi galaxissal bővítenek, melyek azonosítást tűztek ki célul a 2007-ben indult Galaxy Zoo projekt, aminek hála már több, mint egymilliárd galaxis került osztályozásra.

Megfigyelhető, hogy minél távolabbi és halványabb egy galaxis, annál nehezebb felismerni az apró jellegzetességeket, ami különösen igaz, ha a spirális és az elliptikus galaxisokat szeretnénk szétválasztani. Így tehát adott a kérdés, hogy lehetséges-e a gyűrűs galaxis felismerése valamilyen neurális hálózat alapú, a már létező megoldásoknál pontosabb osztályozó rendszert készíteni?

Hasonló fejlesztések

- ▶ A 70-es években, talán az elsők között született a Faint Object Classification and Analysis System módszer, mely a képen lévő képpontok körüli pontok átlagfényességéből számolva több, mint 90% pontossággal volt képes megkülönböztetni galaxisokat a csillagoktól.
- ▶ Később, a 90-es évek közepén az SExtractor projekt már neurális hálózattal különítette el a galaxisokat a csillagoktól, mely azonos fényerősségű képterületeket használt bemenetként és a fényesebbek esetén (<20 magnitúdó) közel 95% pontosságot produkált.
- ▶ Végül a 2000-es években a Sloan Digital Sky Survey program keretein belül készült neurális hálózat és SVM alapú módszer már több, mint 90% pontosságot ért el.

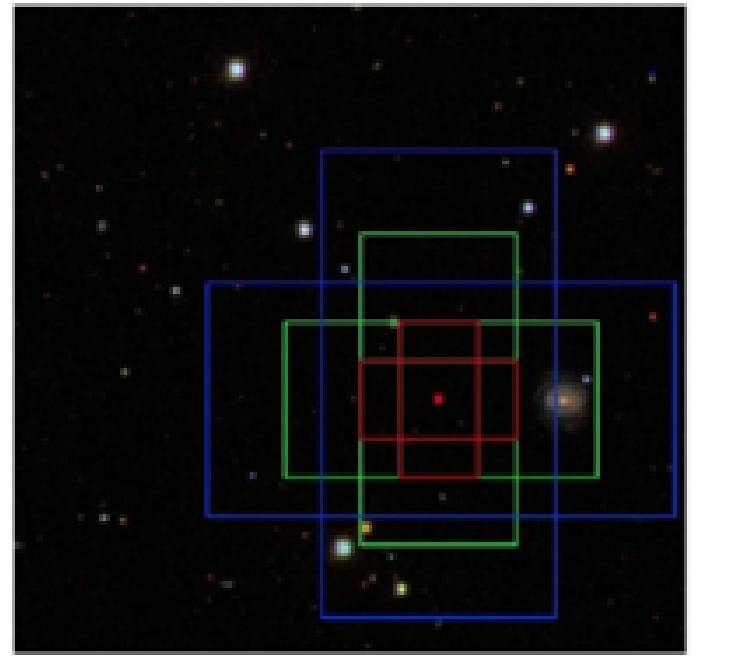
Sikerek

- ▶ Kari TDK 1. helyezett
- ▶ OTDK különdíj
- ▶ 1 konferenciacikk

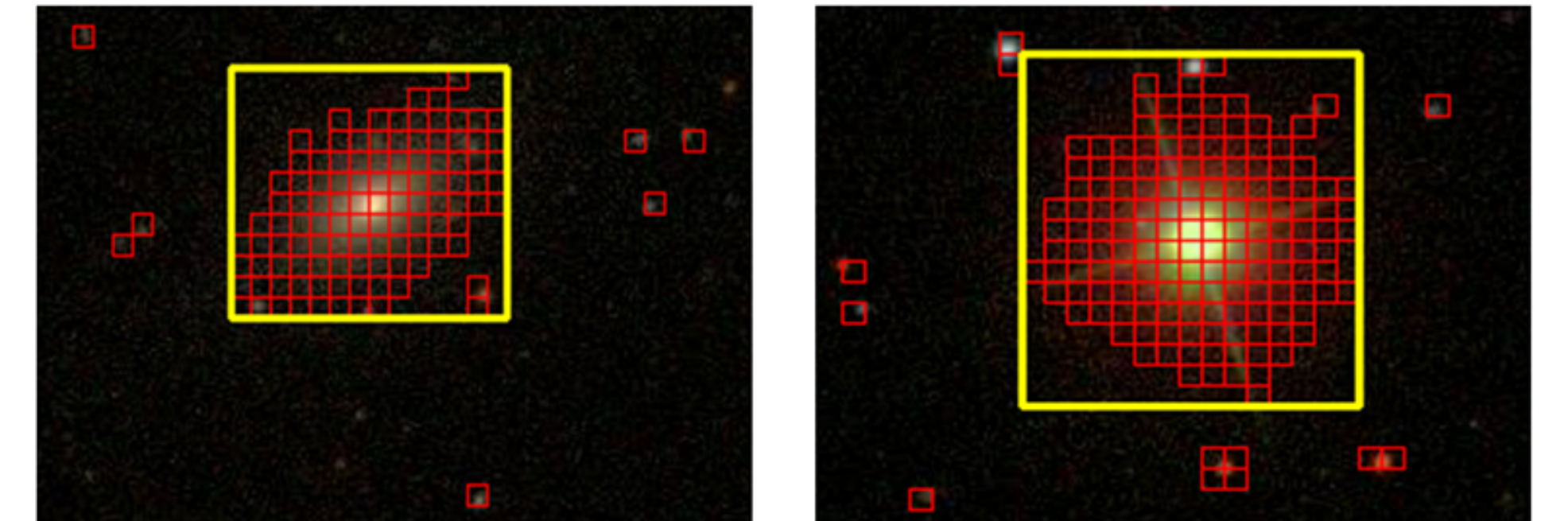
Saját módszer bemutatása

A munka megkezdésének első lépéseként az adathalmazt kellett meghatározni, ami a Galaxy Zoo Catalogue 9. verziója került felhasználásra a jobb generalizációja végett.

Ezt követően elkészült a Faster R-CNN detektáláson alapuló hálózat, ami két modulból áll: egy RPN-ből (Region Proposal Network) és egy Fast R-CNN-ből, illetve a kettő között egy osztott konvolúciós rétegből. Az RPN egy adott képen kijelöli azokat a régiókat (jobb oldali ábra), ahol az általunk keresett objektum valamelyike megtalálható és egy valószínűségi pontszámot rendel hozzájuk. Majd ezután a Fast R-CNN végzi el a konkrét objektum osztályozását.

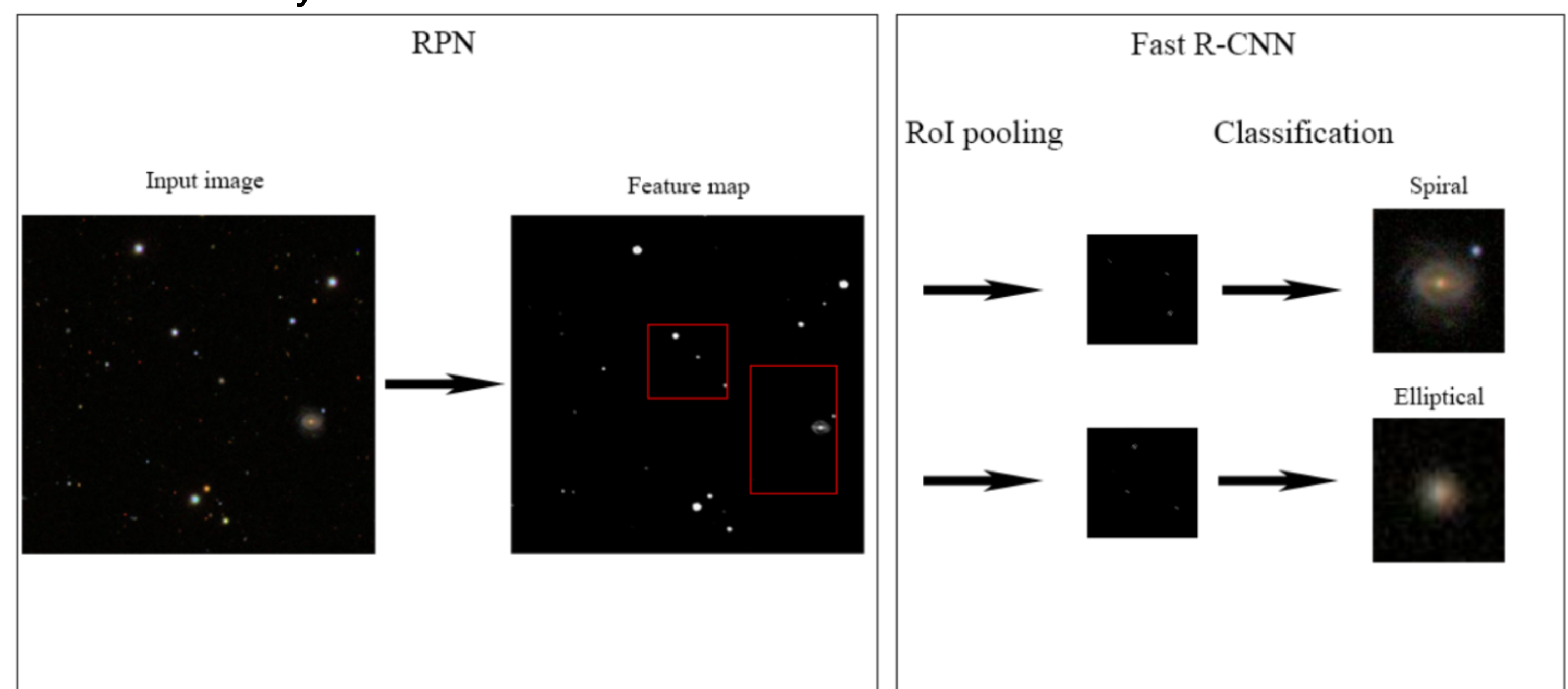


Bár a Faster R-CNN korai eredményei ígéretesek voltak, egy másik módszer is kifejlesztésre került az eredmények összehasonlítására. Ebben a megoldásban a detektálás egy sűrűség alapú klaszterező algoritmus, a DBSCAN segítségével történik. Ennek az algoritmusnak 3 hiperparamétere van: intenzitás határértéke, epsilon és minPts. Az intenzitás alapján a képpontok nullákra és egyesekre kerülnek felosztásra, amit követően az epsilon határozza meg azt a sugarat, amelyben legalább a minPts-nek megfelelő pontnak kell lennie, hogy azt a területet a klaszterezés részének tekinthessük (jobb oldali ábra).



Végezetül az osztályozása történt meg a megtalált galaxisoknak ahol a fő kategória megállapításához veszteségfüggvénynek a kategórikus keresztentropia került felhasználásra, míg gyűrűsség megállapítására szigmoid aktivációs függvény és bináris keresztentropia került alkalmazásra.

A detektálást és az osztályozást az alábbi ábra szemlélteti.

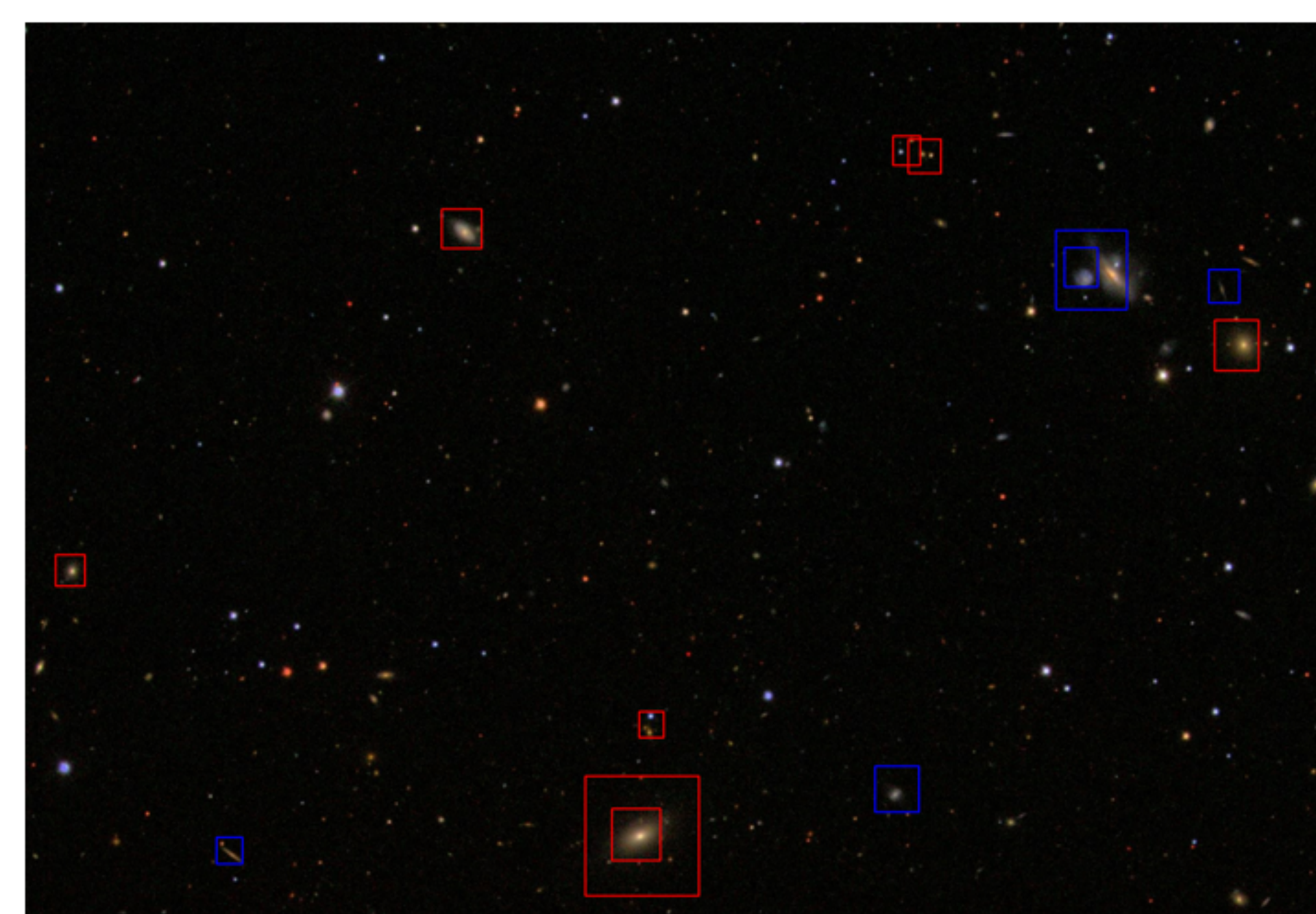


Eredmények értékelése

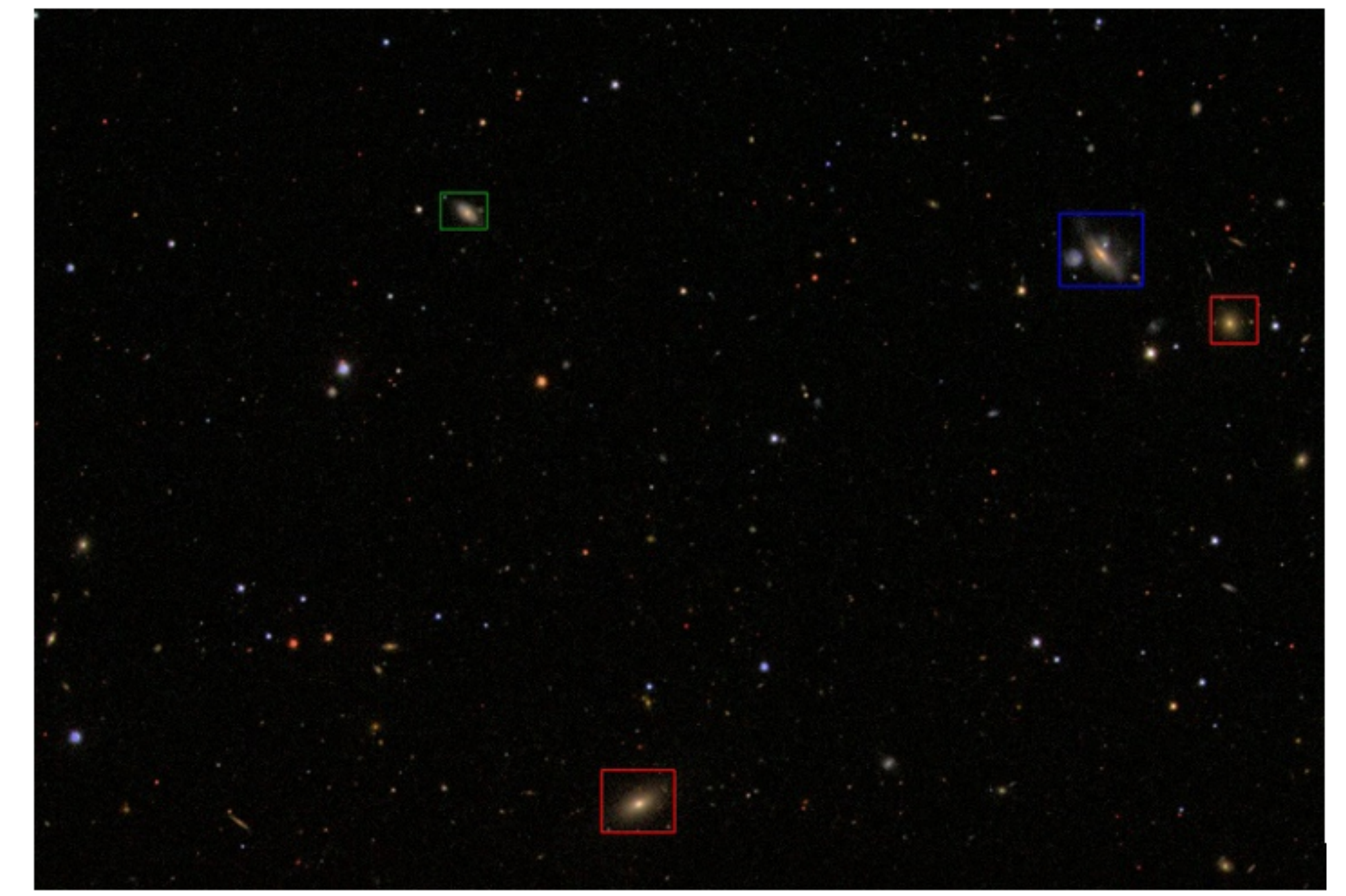
Ahogy az alábbi táblázatban látható, minden mérőszám azt mutatja, hogy a DBSCAN módszer jobb, mint bármely más megoldás. A Faster R-CNN-hez képest nőtt a fő osztály pontossága, és ami még fontosabb, jelentősen nőtt a gyűrűs alakú F1 pontszám. A két hálózat által észlelt galaxisok egy SDSS-képen az alábbi ábrákon láthatók. Bár a DBSCAN nem észlelt két kis galaxist (amit a választott méretkorlát okoz), kevesebb a hamis pozitív detektálás, és nincs átfedés a határoló dobozok között.

Model	Main class accuracy	Ring accuracy	Ring precision	Ring recall	Ring F1 score
DBSCAN	99.7%	94.9%	95.7%	94.0%	94.9%
Faster R-CNN	92.8%	89.3%	58.1%	74.3%	65.2%

A Faster R-CNN által megtalált és azonosított galaxisok



A DBSCAN által megtalált és azonosított galaxisok



Zárásként továbbfejlesztési lehetőségként merülhet fel a DBSCAN futásidejének javítása, mely a modell párhuzamosításával lenne elérhető, illetve az osztályozás további pontosítása, amihez más teleszkópok képeire lenne szükség.