

2. félévi beszámoló

Kunsági-Máté Sándor

E-mail: kunsagisandor@gmail.com

PhD program: Részecskefizika és csillagászat

Témavezető: Dobos László

A dolgozat címe: Gépi tanulás a csillagászatban

2019. június 15.

1. Bevezetés

A csillagászatban manapság leginkább a vöröseltolódást becslő eljárásoknál használnak Deep Learning és gépi tanulási módszereket. A photo-z eljárás során a galaxisképekből kiredukált színindexek alapján adunk becslést a fotometrikus vöröseltolódásra.

A gyakorlatban sokszor előfordul azonban, hogy a galaxisok látóirányába esnek tejútrendszerbeli csillagok (előtércsillagok) is. Emiatt a meghatározott színindexek nem csak az adott galaxis-hoz fognak tartozni, továbbá maga a galaxis sem egyetlen színkomponensből tevődik össze (pl. HII régiók), ezért az integrált magnitúdók használata hibát eredményezhet a fotometrikus vöröseltolódás számításában.

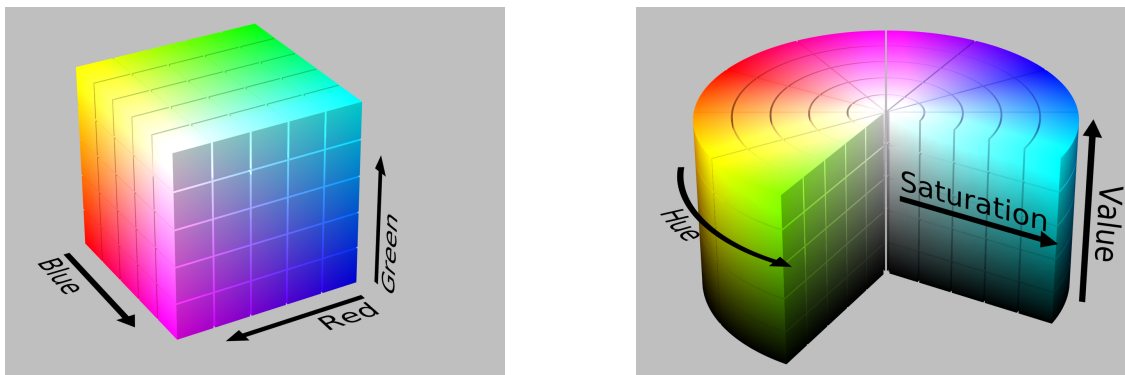
2. Kutatás a második félévben

Kutatásom célja, hogy a galaxisokat különböző színű régiókra bontsuk, majd ezekre külön-külön futtassunk photo-z becslést. A galaxisképek szegmentálásához korábban a Shi&Malik [1] algoritmust használtam, viszont ez egyrészt inkább a DES képeinél kontrasztosabb felvételekre működik megfelelően, másrészt pedig túl nagy futásidővel rendelkezett.

Emiatt a K-means klaszterezési eljárásához fordultam. Itt a legfontosabb feladat az egyes pixelek színét legjobban jellemző "tulajdonság" vektorok előállításának. Fontos szempont, hogy olyan koordinátarendszert használjunk, mely nem tartalmazza a pixelek fényességét. Erre azért van szükségünk, mivel az azonos színű, de eltérő méretű csillagpopulációk fényességben különbözhetnek egymástól.

A digitális fotófeldolgozásban használt RGB és HSV színmodellek mintájára a galaxisok szuperpixeleinek jellemzésére is bevezethetünk egy magasabb dimenziós általánosított

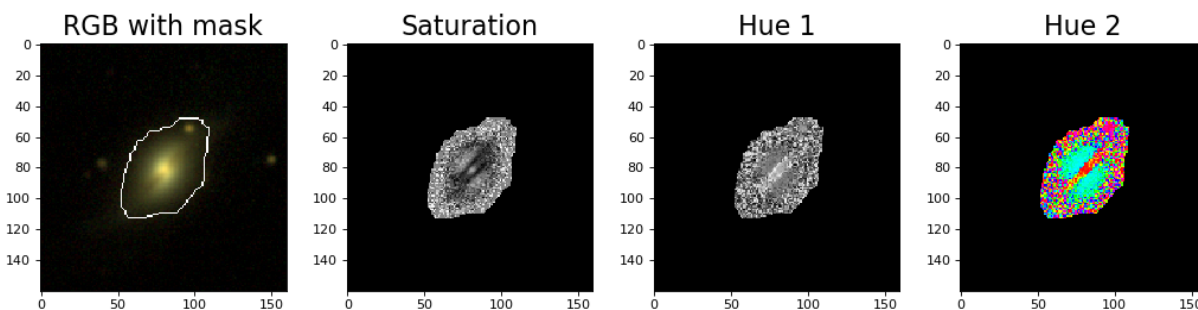
színmodellt. A HSV színtérmodell legfőbb erőssége, hogy a fényességet és a színkomponenseket külön kezeli, ellentétben az RGB modellel (ld. 1. ábra).



1. ábra. RGB és HSV színtérmodellek (forrás: <https://en.wikipedia.org/>)

A HSV modellt legegyszerűbben úgy kaphatjuk meg, ha az RGB kockára a főátló (*fényesség*) irányából nézünk rá. Ekkor a *szaturáció* (saturation) a főátlótól mért távolságot, a *színárnyalat* (hue) pedig az irányszöveget jelenti (polár koordináták). A szegmentáláshoz használt négy (G, R, I, Z) színszűrő egy négydimenziós színteret fog kifeszíteni. Az RGB-HSV transzformáció analógiájára vetítsük le a színtér pontjait a négydimenziós kocka főátlójára (fényesség) merőleges síkra. Mivel a galaxisok képein jelentős fényesség gradiens tapasztalható, ezért egy egyszerű főkomponens analízissel megállapítható a fényesség komponens, illetve az erre merőleges irányok.

A levetítés után egy háromdimenziós alteret kapunk, amely pontjait fejtsük ki gömbi koordinátarendszerben. Ekkor – a megszokott jelölésrendszert követve – r a szaturációnak, $\theta \in [0, \pi]$ és $\phi \in [0, 2\pi]$ pedig két "hue" szögnek feleltethető meg. Az alábbi képsorozaton (2. ábra) egy poros galaxis – G, R és I szűrőkből készült – RGB felvételét, valamint a háromdimenziós altérre vetített pontok gömbi koordinátáit láthatjuk.

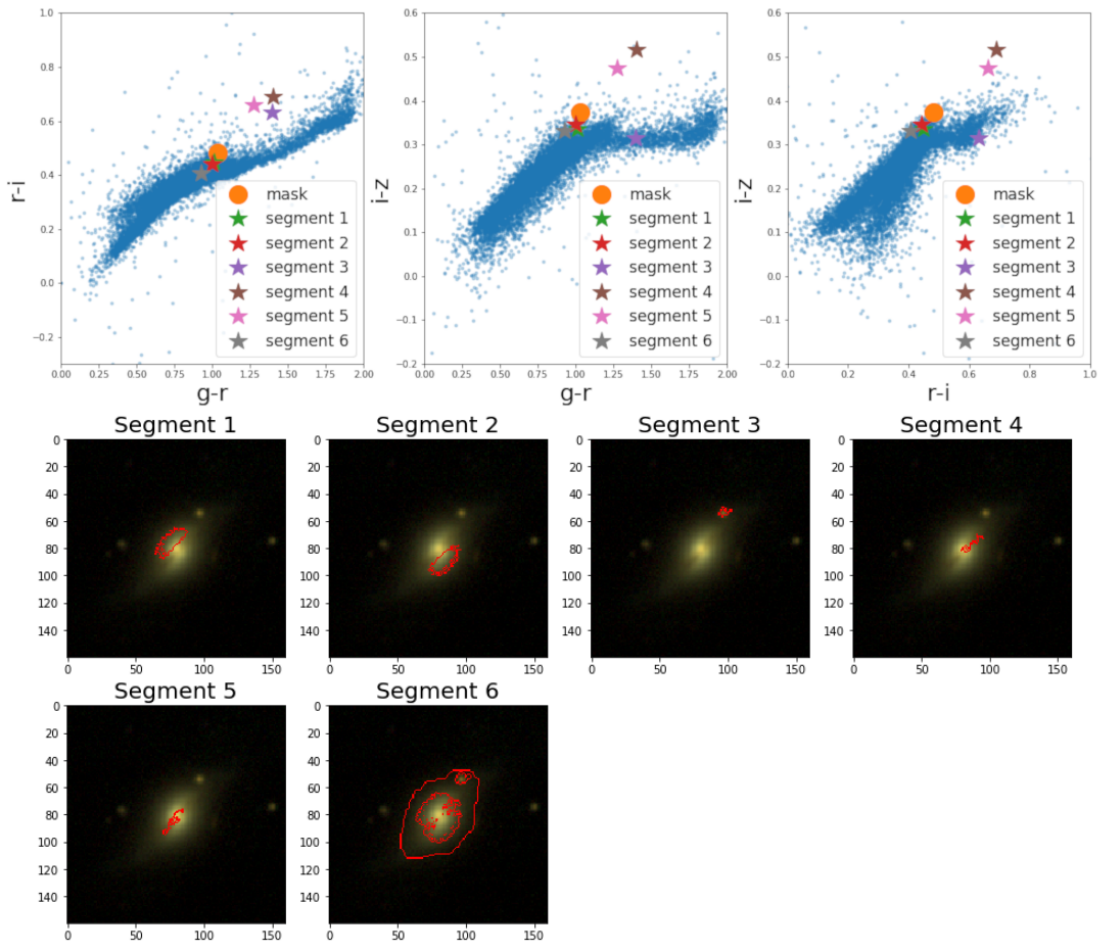


2. ábra. Poros galaxisról készült RGB kép, valamint a PCA fényességre ortogonális három főkomponense. Megj.: a "Hue 2" ($\phi \in [0, 2\pi]$) periodikus koordináta, ezért az ábrázoláshoz színkódolást alkalmaztam.

A galaxisképek színtérbeli szegmentációjához a két hue koordinátát használtam. Az algoritmus főbb lépései:

1. a galaxis kiterjedésének megfelelő maszk meghatározása
2. PCA alkalmazása a G,R,I,Z koordinátákra a maszkon belül
3. *Hue 1* és *Hue 2* meghatározása
4. a pixelek csoportosítása K-means módszerrel (Megj.: az általam használt `sklearn.cluster.KMeans()` függvény csak euklideszi távolságot tud alkalmazni, ezért a periodikusság figyelembe vételéhez a *Hue 1*, *Hue 2* szögeket egységömbre vetítettem.)
5. összefüggő komponensek keresése, valamint a PSF-nél (pontterületi függvény) kisebb szegmensek eldobása

A 3. ábrán azt szemléltetem, hogy a szegmentáló algoritmus valóban képes megfelelően elkülöníteni a különböző színindexű területeket. A legfelső sorban kék pontokkal jelöltem a vizsgálatához használt kb. 11000 galaxis maszkra kiszámolt színindexeit, narancssárga körrel az adott galaxis maszkjára illetve csillagokkal a szegmenseire kiszámolt színindexeket. Az egyes szegmenseket az alsó sorokban található képsorozaton tudjuk beazonosítani (piros határoló vonal).

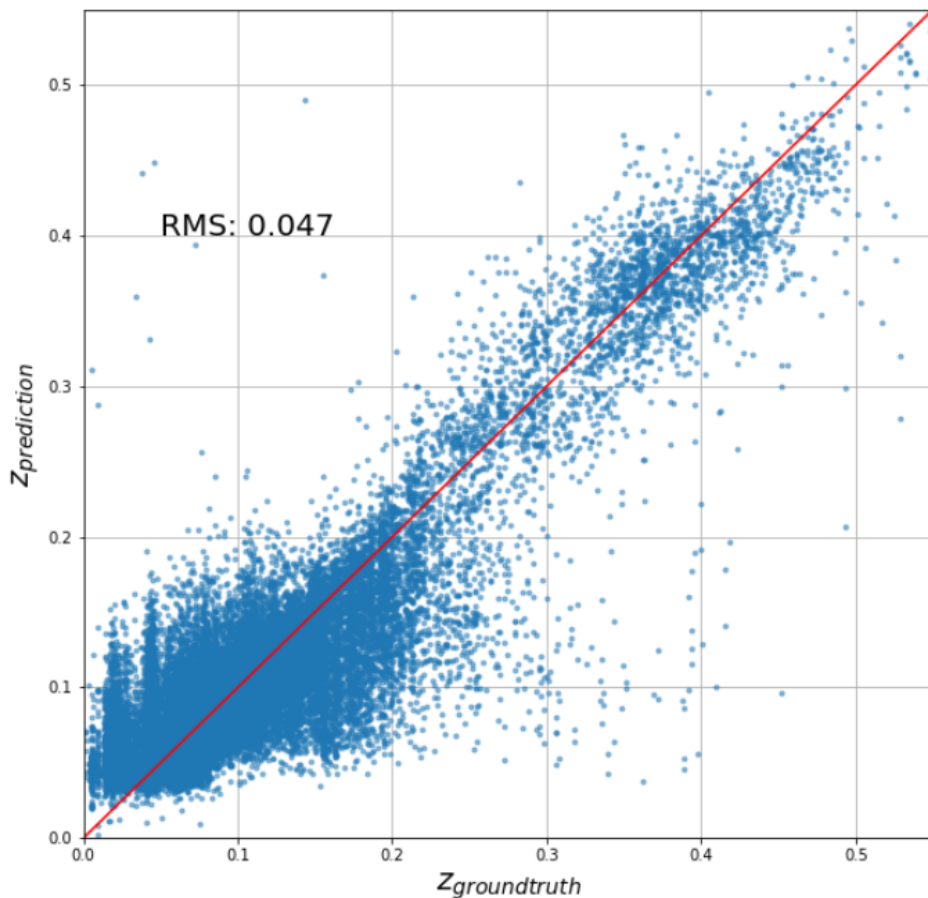


3. ábra

Két fontos megállapítást tehetünk:

- az előtércsillag (3-as szegmens), valamint a porsávhoz tartozó két szegmens (4-es és 5-ös) szemmel láthatóan eltér a galaxisok többsége által kijelölt "nyomvonalról" a szín-index diagramokon
- ezek a kilógó pontok pedig jelentősen befolyásolják a galaxis maszkon kiszámolt színindexeit, észrevehetően elhúzzák azt a fő nyomvonalról

A photo-z becsléshez a legfontosabb feladat, hogy a tanítóhalmazba ne kerüljenek bele az előtércsillagokhoz, illetve poros régiókhoz tartozó szegmensek. Ennek eléréséhez a szegmensek színindexeinek a maszktól mért távolsága alapján végeztem szűrést az adathalmazon. Ezt követően lokális lineáris regressziót [2] alkalmaztam a kb. 27000 szegmens színindexeire. A korrelációs grafikont az alábbiakban ábrázoltam (ld. 4. ábra):



4. ábra. Galaxisszegmensek színindexeire alkalmazott lokális lineáris regresszió szerint becsült fotometrikus vöröseltolódás korrelációs ábrája.

Az irodalomban (pl.: [3]) leggyakrabban megtalálható RMS érték 0.02 körül mozog, azaz a módszert tovább kell pontosítani a most elért 0.047-es érték csökkentéséhez.

3. Tervezett feladatok

A további szükséges lépéseket az alábbi pontokban foglaltam össze:

- megfelelő modell kiválasztása (alkalmas paraméterekkel) a photo-z becsléshez
- hibaszámolás a szegmensek színindexeire, majd ennek figyelembe vétele a photo-z becslés során
- a PSF pontosabb kezelése a szegmentálás fázisában, mivel jelenleg néhol egyetlen pixel szélességű szuperpixelek is előfordulnak, mely nem fizikai eredmény
- publikáció készítés

4. Elvégzett kurzusok a félévben

- Adatbányászat a csillagászatban (FIZ/2/119), oktató: Dobos László
- Csillag- és galaxispopulációk (FIZ/2/120), oktató Dobos László
- Csillagaktivitás - aktív csillagok 1. (FIZ/2/033E:2), oktatók: Vida Krisztián, Kriskovics Levente
- Naprendszerbeli plazmák fizikája (FIZ/2/082E), oktató: Németh Zoltán

Hivatkozások

- [1] J. Shi and J. Malik. Normalized cuts and image segmentation. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, pages 731-737, 1997.
- [2] https://en.wikipedia.org/wiki/Local_regression
- [3] A. Molino et. al. The ALHAMBRA Survey: Bayesian photometric redshifts with 23 bands for 3 deg, MNRAS, pages 2891-2922, 2014.