

Dirbtinio intelekto metodų tyrimas astrofizikinių objektų klasifikavimui ir/arba svarbių savybių nustatymui

Doktorantas: Tomas Mūžas

Darbo vadovas: prof. dr. Tadas Meškauskas

Darbo konsultantas: dr. Andrius Vytautas Misiukas Misiūnas

Doktorantūros studijų laikotarpis: 2022 – 2026 m.

Studijų planas ir jo vykdymo suvestinė

Metai	Egzaminai		Dalyvavimas konferencijose				Publikacijos		
			Tarptautinės		Nacionalinės		Su citav. rodikl.		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būsena
I (2022/2023)	2	2	-	1	-	-	-	-	-
II (2023/2024)	2	-	-	-	-	-	-	-	-
III (2024/2025)	-	-	1	-	-	-	1	-	-
IV (2025/2026)	-	-	1	-	-	-	1	-	-
Iš viso	4	2	2	1	0	0	2	0	-

2022/2023 m.m. pavasario semestro darbo planas ir jo vykdymas (1)

Egzaminai

Planas	Įvykdyta	Būklė
Vaizdų ir duomenų analizė (FF) (2023-06-14)	2023-06-14	Išlaikyta
Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika (2023-06-22)	2023-06-22	Išlaikyta

2022/2023 m.m. pavasario semestro darbo planas ir jo vykdymas (2)

Dalyvavimas konferencijose

Planas	Įvykdyta	Tipas
ICCSA (2023-07-02 – 07-06), Atėnai, Graikija	Large Scale Study of Binary Galaxy Image Classification and the Impact of Image Augmentation Techniques, <i>T. Mūžas, A. V. Misiukas Misiūnas, T. Meškauskas,</i> Computational Science and Its Applications – ICCSA 2023, Springer, Lecture Notes of Computer Science vol. 13957, DOI: 10.1007/978-3-031-36808-0_27	Tarptautinė

Tyrimo objektas

- Astrofizikinių objektų klasifikavimas ir/ar savybių nustatymas
- profesionalų parengti arba savanorių balsavimu grįsti duomenų rinkiniai
- antžeminių ir kosmoso teleskopų nuotraukos

Tyrimo tikslas

Pasiūlyti naują arba patobulintą astrofizikinių objektų klasifikavimo modelį, paremtą dirbtinio intelekto metodais

Tyrimo uždaviniai

1. Sukurti astrofizikinių objektų duomenų analizės metodiką, apibrėžti kriterijus objektų klasifikavimo ir/ar savybių patikimumui įvertinti. Taip pat parengti metodiką sintetinių duomenų generavimui.
2. Atrinkti patikimų duomenų rinkinius, skirtus modelių apmokymui ir validavimui.
3. Apmokyti skirtingus dirbtinio intelekto modelius naudojant parengtą apmokymo duomenų rinkinį. Skaičiavimus atlikti naudojant paskirstyto skaičiavimo resursus bei grafinius procesorius.
4. Išskirti efektyviausius modelius bei jų parametrus, optimizuoti jų taikymą.

Pagrindiniai galaktikų tipai

Spiralinė

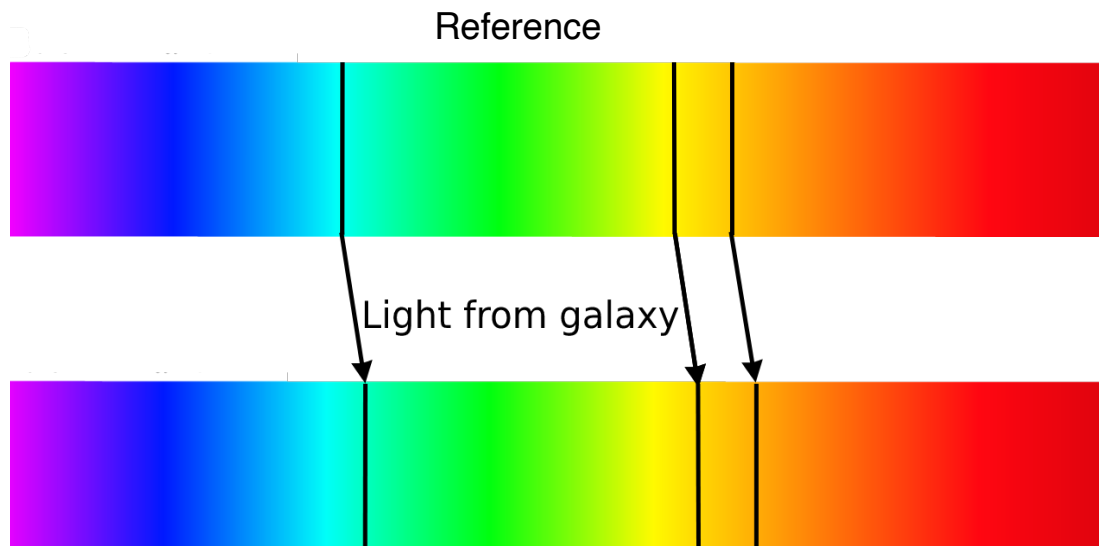


Eliptinė



Raudonasis poslinkis

Raudonasis poslinkis (žymima z) – galaktikos skleidžiamos šviesos bangos dažnio sumažėjimas jai sklindant kosmosu.



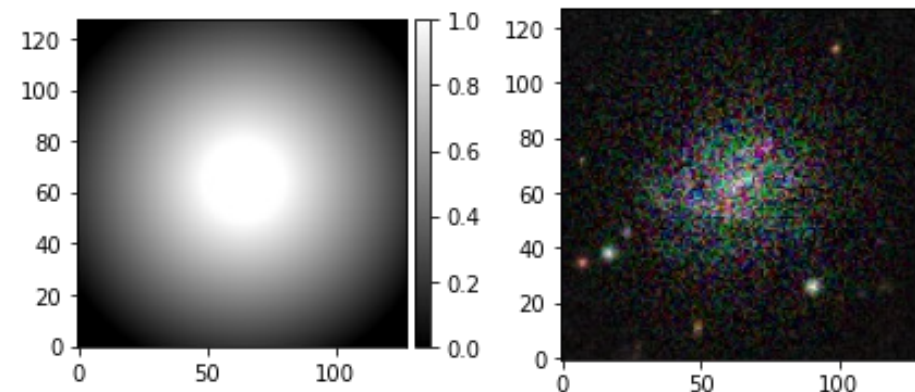
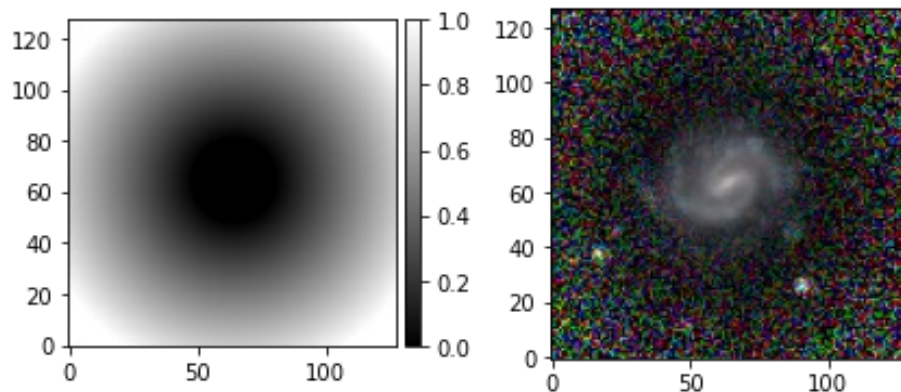
Kuo didesnė raudonojo poslinkio reikšmė, tuo galaktika yra toliau.

Problemos

- Atlikus literatūros apžvalgą paaiškėjo, kad nors ir kiti autoriai pasiekia puikų tikslumą (95-99%), tačiau **naudojami tik nedideli duomenų rinkiniai (< 100,000), sudaryti tik iš artimiausių ir ryškiausių galaktikų ($z \sim 0.1$).**
- Be to, **visi autoriai pateikia skirtingas rezultatų vertinimo metrikas,** dėl ko sudėtinga palyginti straipsnių rezultatus tarpusavyje.
- Autoriai naudoja nuotraukų augmentacijos technikas, tačiau **tik Cheng et al. vertina augmentacijų poveikį.**

Atlikti eksperimentai

- Naudojant duomenų rinkinį iš **315,942 įvaraus atstumo ir ryškumo galaktikų ($z \leq 0.28$)**, buvo apmokyti 4 architektūrų modeliai
- Buvo vertinamos šios metrikos: accuracy, precision, recall, true negative rate (TNR), F1.
- Buvo vertinama, kokią įtaką metrikoms turi augmentacijos
- **Išbandyti 2 nauji augmentacijos būdai**



Atliktų eksperimentų esminiai rezultatai (1)

Supaprastintos Dieleman et al. architektūros modelis (siekiant įvertinti augmentacijų įtaką nepakankamai apsimokančiam (*angl. underfitting*) modeliui)

Experiment	Accuracy (σ)	F1 (σ)
Baseline (no augmentations)	94.72% (0.24)	90.93% (0.41)
Standard augmentations (rotation, zoom and flips)	94.56% (0.17)	90.67% (0.30)
Standard + random noise	94.47% (0.16)	90.49% (0.30)
Standard + centre noise	94.63% (0.20)	90.81% (0.32)
Standard + outside noise	94.72% (0.23)	90.95% (0.42)

Atliktų eksperimentų esminiai rezultatai (2)

ResNet50

Experiment	Acc. (σ)	TNR (σ)	F1 (σ)
Baseline (no augmentations)	93.53% (0.48)	95.21% (1.44)	88.66% (1.14)
Standard augmentations (rotation, zoom and flips)	95.21% (0.23)	97.22% (0.39)	91.82% (0.36)
Standard + random noise	95.38% (0.09)	97.06% (0.20)	92.05% (0.14)
Standard + centre noise	95.23% (0.28)	97.12% (0.42)	91.84% (0.49)
Standard + outside noise	95.07 (0.37)	97.08% (0.38)	91.58% (0.61)

Rezultatai

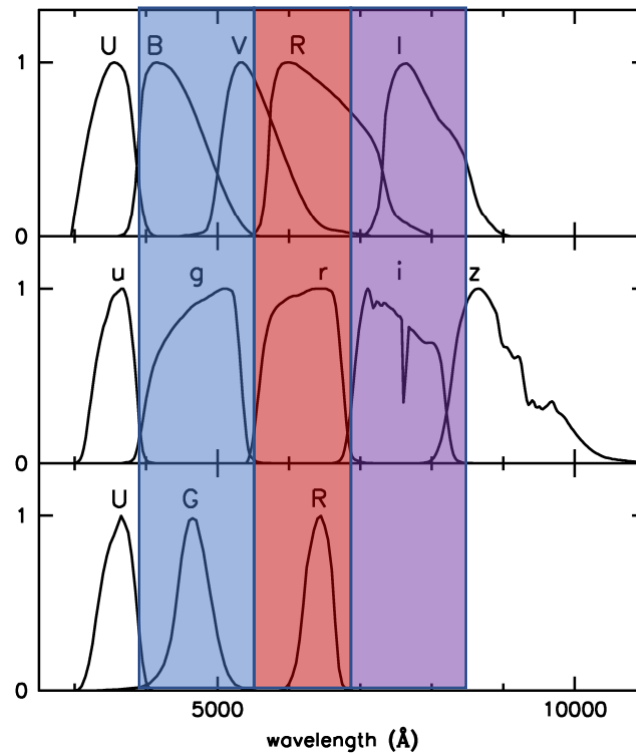
- Atlikta galaktikų klasifikavimo į spiralines ir eliptines literatūros apžvalga
- Sudarytas patikimai klasifikuotų GZ1 ir GZ2 galaktikų sąrašas ir sugneruotos jų nuotraukos
- Įvertinta standartinių augmentacijų (pasukimas, apvertimas ir priartinimas) įtaka modelių metrikoms
- Išbandyti du nauji, atsitiktiniu triukšmu paremti augmentacijos būdai

Išvados

- Standartinės augmentacijos pagerina tikslumą ir F1 įvertį.
- Atsitiktinis triukšmas papildomai nepagerina jokių metrikų, tačiau sumažina standartinį nuokrypį (net iki 9 kartų ResNet50 atveju).
- Nei viena augmentacijos technika nepagerina precision, recall ir TNR metrikų.
- Jei modelis nepakankamai apsimoko (*angl. underfitting*), nei viena augmentacijos technika taip pat nepagerina jokių metrikų.

Numatomi darbai (1)

- Atrinkti patikimai klasifikuotas galaktikas visuose GZ duomenų rinkiniuose
- Teleskopų stebimų šviesos bangų ilgius suskirstyti į 3 režius:

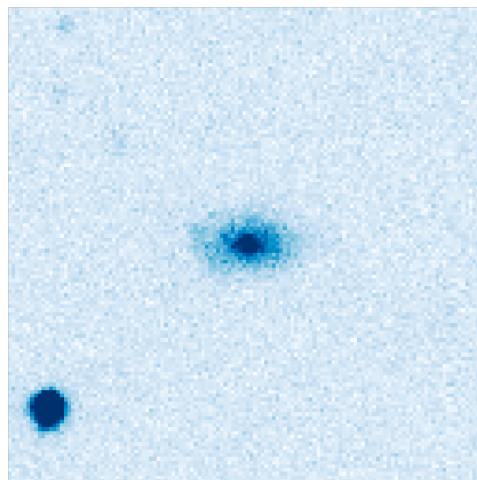


Numatomi darbai (2)

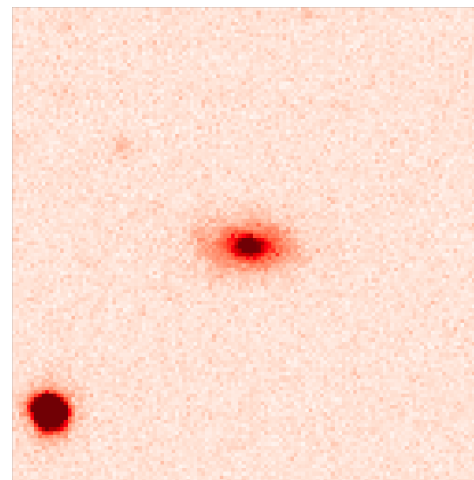
- Sugeneruoti galaktikos nuotrauką kiekviename iš režijų:



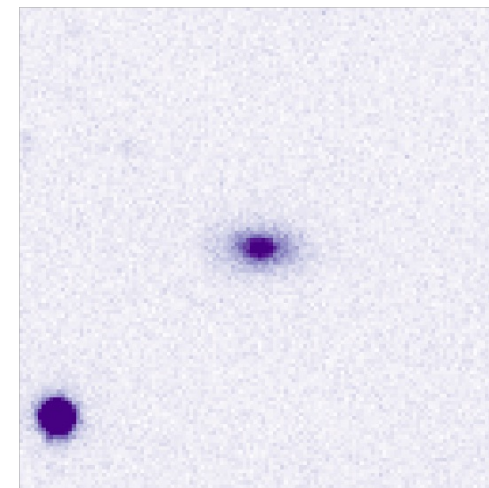
RGB nuotrauka



g režis



r režis



i režis

Numatomi darbai (3)

- Pasirinkti architektūras ir apmokyti modelius su kiekvienu iš režijų
- Patikrinama, ar kuris nors iš režijų neša daugiau informacijos (geresnis tikslumas ir kitos metrikos) nei kiti

Jei nėra vieno režio, nešančio žymiai daugiau informacijos:

- Kiekvienam režiiui parenkamas geriausias modelis
- Modeliai balsuoja dėl galutinės klasės
- Galimai atsižvelgiama į raudonojo poslinkio reikšmę, kiekvienam modeliui priskiriant atitinkamą balso svorį

Ačiū už dėmesį!