بنام خدا

Observing Dark World

مرتضی صادقی 9130202

**خلاصه**

مساله ای که امسال در سایت kaggle در آن شرکت کردیم مساله ObservingDarkWorld است. در این مساله خواسته شده بکمک مدلسازی، مکان سیاهچاله را در آسمانها پیدا کنیم. تعدادی نمونه حل شده وجود دارند که در آنها مکان سیاهچاله و تاثیر حضور آن بر کهکشان های اطرافش محاسبه شده اند. حال از ما خواسته شده بکمک یک مجموعه آموزشی train مدلی بسازیم که جواب را درمجموعه test پیداکنیم. مساله برای کمک به ما سه نمونه راه حل ارائه داده که هیچیک دقت کافی و لازم را ندارند. من هم سعی کردم بکمک گسترش یکی از همان ها مساله را حل کنم. البته نتوانستم مساله را به خوبی حل کنم، علت این بود که تا روزهای پایانی متوجه این مساله نبودم که توزیع آسمانها در مجموعه train و test متفاوت است و تلاش خود را بر بالابردن دقت جواب روی دسته ای از آسمانها متمرکز کردم که در مجموعه train اکثریت آسمان ها بودند ولی اکثریت آسمان های test شبیه آنها نبودند. البته یکبار هم سعی کردم بکمک هوش مصنوعی مساله را حل کنم اما نتوانستم چون اطلاعات و مهارت لازم در این زمینه را نداشته ام. درکل بهترین جوابی که بدست آورده ام دارای امتیاز 0.86 شد که رتبه 64 را بدست آورد.

**توضیح درباره بخش ها**

در بخش 1 مساله را تعریف میکنم به همراه یک نمونه از روش های حل آن که خود مساله به ما داده است. دربخش 2 روشهایی که خود دنبال کردم را میاورم، در بخش 3 بررسی میکنم که چرا روش ها خوب جواب ندادند، در بخش 4 شبیه سازی ها و در بخش 5 امتیاز سایت kaggle را نشان میدهم.

**1 - تعریف مساله**

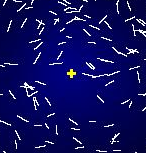
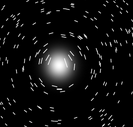
مساله این است که قرارگرفتن یک سیاهچاله در آسمان باعث میشود تصاویر دریافتی از آسمان کمی تغییر کنند، چون سیاهچاله اثر گرانشی بر نوروارد میکند و این باعث میشود نور از مسیر اصلی خود کمی منحرف شود و بنابراین تصویری که ما میبینیم دستخوش تغییر شود. مکان برخی از سیاهچاله ها برای ما نشاخته شده هستند، اما برخی سیاهچاله ها هم هستند که مکان آنها را نمیدانیم اما میدانیم که در آسمان یک سیاهچاله ای هست. اینجا میخواهیم که مکان دقیق سیاهچاله در آسمانها پیداشود، البته این کار را بکمک نمونه های حل شده انجام میدهیم. یعنی بکمک آسمان هایی که مکان سیاهچاله آنها مشخص است، مدلی میسازیم که آن بتواند روی آسمان های با مکان مجهول سیاهچاله ها هم جواب دهد. مساله از ما انتظار دارد که این کار را بکمک تاثیری که حضور سیاهچاله بر ظاهر کهکشانها در تصویر میگذارد انجام دهیم، و تاثیر مدنظر ماسله، یکی کشیدگی کهکشان است و یکی تغییر زاویه آن در تصویر.

تعدادی آسمان با مکان مشخص سیاهچاله وجود دارند که بخشی از آنها بعنوان train و بخشی بعنوان test به ما داده شده تا مدلسازی ما را ارزیابی کند. بدین منظور تعداد 300 آسمان با مکان مشخص سیاهچاله بعنوان train برای مدلسازی داده شده و تعداد 120 آسمان با مکان نامشخص سیاهچاله ها بعنوان test داده شده، هردو مجموعه به سه دسته تقسیم میشوند : آسمان های شامل یک سیاهچاله، دوسیاهچاله، و سه سیاهچاله. هر آسمان شامل تعدادی کهکشان است که هرکهکشان دارای مشخصه های مختصات x,y و دو عدد e1 و e2 است که نرمالیزه شده اند و بین -1 و 1 قرار گرفته اند. عدد e1 کشیدگی کهکشان و عدد e2 زاویه کهکشان را نشان میدهد. این دو عدد تاثیر حضور سیاهچاله را بر کهکشانها نشان میدهند. درکنار اینها به ما سه نمونه حل مساله را نیز داده اند. یک نمونه بطور رندوم جوابهایی میسازد، یک نمونه بکمک ابزار lenstool جواب میسازد و یک نمونه هم بنام gridded benchmark همه آسمان را جستجو میکند تا ببیند در کدام محل بیشتر حضور سیاهچاله محتمل هست. این روش در ادامه توضیح داده میشود.

برای محاسبه کارایی مدل مان، یک کد metric به زبان python نوشته شده که اختلاف جواب های ما را با جواب های آسمان های train مقایسه میکند. از این کد بعنوان معیار در اجراهایم استفاده کرده ام.

**1-2 - Gridded benchmark**

یکی از روش های یافتن سیاهچاله برپایه این فرض است که حضور یک سیاهچاله در محلی باعث میشود کهکشان های اطراف آن بر آن مماس شوند و ازنظر اندازه نیز کشیده شوند. شکل زیر یک نمونه حضور سیاهچاله را شنان میدهد که در ان علامت زرد محل سیاهچاله است :

شکل 1 - تاثیر حضور سیاهچاله بر شکل کهکشان های اطراف آن. تصویر **چپ** مربوط به نمونه سایت kaggle است. در این تصویر نقطه سفید مکان سیاهچاله است. تاثیر سیاهچاله بر کشیدگی و مماس شدن کهکشان های اطراف را میبینید. تصویر سمت **راست** یک نمونه شبیه سازی شده مجموعه train مربوط به آسمان 41 است. نقطه زرد مرکز سیاهچاله را نشان میدهد. در این تصویر برخی کهشکان ها از قاعده تصویر قبلی پیروی نمیکنند که آنها درواقع نویز هستند.

البته بنظر میرسد برخی کهکشانها از این قاعده پیروی نمیکنند، درحقیقت آنها کهکشانهایی هستند که اصلا نزدیک سیاهچاله و تحت تاثیر آن نیستند، بلکه در عکس دوبعدی ای که ما گرفته ایم آن کهکشانها نزدیک سیاهچاله قرارگرفته اند درحالیکه فاصله زیادی با آن دارند و تحت تاثیر آن قرار ندارند. این داده ها درواقع نویز حساب میشوند.

روش gridded benchmark به ازای هر نقطه از آسمان همه کهکشانها را درنظر میگیرد، و برای هرکهکشان بکمک دو عدد e1 و e2 حساب میکند که آن نقطه چقدر مستعد حضور سیاهچاله است. درواقع اگر آن نقطه مستعد حضور سیاهچاله باشد باید کهکشانها بر آن مماس باشند و کشیده باشند، این تاثیر را در فرمول زیر تحت عنوان نیروی مماسی میتوان بیان کرد :



که phi زاویه کهکشان نسبت به نقطه موردنظر است. وقتی مجموعه نیرو های مماسی وارد شده توسط کهکشان ها بر یک نقطه از بقیه نقاط بیشتر باشد، آن نقطه بیشتر مستعد حضور سیاهچاله است، و آن را بعنوان مکان سیاهچاله درنظر میگیریم. در شبیه سازی زیر که مربوط به آسمان شماره 11 از مجموعه train است، رنگ آبی نشان دهنده شدت نیروی مماسی است که این مقادیر بین 0 تا 255 نرمالیزه شده اند، سیاهچاله در نقطه ای قرار دارد که شدت نیروی مماسی در آنجا بیشینه است.



شکل 2 - شدت نیروی مماسی در آسمان شماره 11 از محموعه train . سیاهچاله در جایی قرار گرفته که شدت این نیرو بیشینه است.

درواقع اجراها نشان میدهند که این روش خوب جواب میدهد، اما دقت آن کافی نیست. و این دلایلی دارد، صرفنظر از نویز و داده های خراب، علت اصلی آن است که نمیتوانیم همه نقاطه آسمان را وجب به وجب چک کنیم، چون برای هر نقطه مجبوریم همه کهکشانها را وارد محاسبه کنیم که در برخی آسمانها نزدیک 700 تا هستند و این زمان زیادی برای اجرا میبرد، بنابراین مجبوریم دقت را کمی پایین ببریم و مثلا 5% نقاط را بررسی کنیم. البته این روش ممکن است برخی جاها جواب ندهد –همانطور که نداده- و در مورد برخی آسمان خاص در مجموعه train –آسمان های 20و 32و 60و 75و 82- توزیع نیروی مماسی بگونه ای است که بیش از یک منطقه در آسمان اکسترموم هستند و روش ما در اطراف یکی از این بیشینه ها به دام میفتد که آنجا مکان سیاهچاله نبوده. درواقع بیشتر آسمان های test از این نوع هستند که در ادامه این نکته به تفضیل توضیح داده میشود. روشی که خود مساله بررسی کرده، با دقتbin معادل 15\*15 نقطه بوده که در زمان کمتر از 1 دقیقه روی همه آسمان های train انجام میشود. یعنی هر نقطه نماینده محدوده ای به حدود 4200/15 طول و عرض بوده است.

**2 - روش های پیشنهادی من**

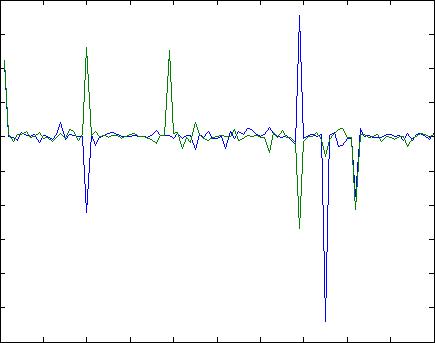
روش هایی که من دنبال کردم سه دسته ایده تقسیم میشوند، یک دسته افزایش دقت و بهینه سازی روی روش gridde benchmark است، یک روش استفاده از classification است و یک روش هم درواقع بیان دیگری از نیروی مماسی است. درکنار این ها از visualization هایی هم استفاده کرده ام که همه اینها را در ادامه توضیح میدهم. همه کارهای انجام شده در نرم افزار Matlab R 2012 انجام شده اند، و بدلیل کمبود امکانات دیالوگ سازی در این محیط، دیالوگی برای انجام محاسبات طراحی نکرده ام و همه کارها ازطریق اجرای اسکریپت ها و فانکشن ها انجام میشوند.

**2-1 - ایده اول، افزایش دقت و بهینه سازی روی gridded benchmark**

روش gridded benchmark روی آسمان های تک سیاهچاله اجرا میشود. پس اول درمورد آسمان های تک سیاهچاله ای صحبت میکنم. سپس سعی میکنم دو سیاهچاله را بهمین ترتسی حل کنم.

در این ایده، ابتدا سعی کردم تعداد نقطه های مورد محاسبه را افزایش دهم، این کار البته باعث میشد زمان محاسبه بیتشر شود، این پیچیدگی بنظر میرسد نمایی باشد; برای دقت 15\*15 تا 200\*200 زمان محاسبه از 5 دقیقه تجاوز نمیکند درحالیکه برای دقت 400\*400 زمان محاسبه به 1 ساعت افزایش میابد. ولی درعوض جواب با دقت بهتری بدست می آمد و بهترین جواب را تعداد 800\*800 نقطه بدست آورده ام که حدود 12 ساعت برای 100 آسمان train اجرا شد. در [این جدول](idea1_reports.docx) زیر برخی از این اجراها آورده شده اند. در ردیف 1 با دقت 400 و ردیف 2 با دقت 800 اجراشده است که درنتیجه امتیاز metric آنها از 1.96 به 0.98 رسیده است و خطای اقلیدوسی جواب های من نسبت به train از 196 به 108 رسیده و همچنین زمان اجرا از 1 ساعت به 4 ساعت رسیده است. این بهترین جوابی است که با این دقت بدست میاید و با افزایش دقت، امتیاز metric از این حد تجاوز نمیکند.

بنابراین مجبورم یکسری بهینه سازی هایی انجام بدهم. بهینه سازی اول من روی کد gridded benchmark بوده که چند مقدار بصورت تکراری در یک حلقه محاسبه میشدند که با بیرون کشیدن آنها، زمان اجرای 12 ساعت برای دقت 800، به 4 ساعت کاهش یافت! نمودار زیر مقدارخطا برای 100 آسمان train را نشان میدهد.

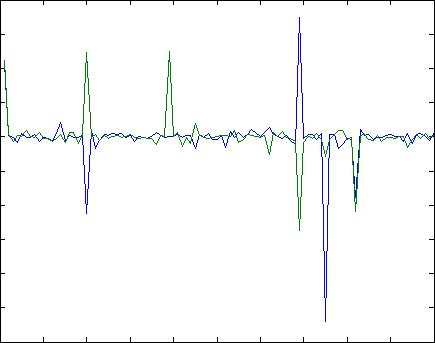


شکل 3 - مقدارخطا برای آسمان های train با دقت 800 در plan001 – خطا در آسمان های 1و20و38و69و75و82 زیاد است که درمورد همه روش هایی که امتحان کرده ام این خطاها بهمین ترتیب زیاد اند.

بهینه سازی بعدی شرکت کمتر ستاره ها در محاسبه است. همانطور که در شکل 1 اشاره شد بنظر میرسد برخی داده ها اظافی و نویز باشند، و ما این ستاره ها را نمیتوانیم تشخیص دهیم، بنابراین با انتخاب تصادفی ستاره ها میتوانیم از حجم این نویز بکاهیم. اجراهای ردیف 3 و4 در [جدول یادشده](idea1_reports.docx) این کار را انجام داده اند و یک پنجاهم ستاره ها را با دقت 15 و 100 محاسبه کرده اند که امتیاز 1.2 و خطای اقلیدوسی زیاد 1283 را داشته اند. چون این خطا نسبت به خطای روش قبلی با همان دقت بسیار بیشتر بود، از ادامه آن صرفنظر کردم. البته تاثیر بسیار خوبی روی زمان اجرا داشته است.

بهینه سازی بعدی بر این فرض استوار است که تاثیر سیاهچاله بر ستاره های نزدیک بیشتر است تا ستاره های دور. البته این فرض اشتباه است، چون تصویری که ما دراختیار داریم یک عکس دوبعدی از آسمان سه بعدی است و فاصله ها در آن ، آنطور که به نظر میرسند نیستند. البته این بهینه سازی تاثیری بر زمان نداشته و درواقع تاثیر آن روی جواب و حذف داده های نویزی مدنظر بوده. بهرحال این روش هم توانسته تاثیر خوبی داشته باشد. در این روش فقط ستاره های داخل bin درنظرگرفته شده اند. اجراهای 5و6و7 در [جدول بالا](idea1_reports.docx) با دقت های مختلف این روش را آزمایش کرده اند، دقت 400 با خطای 500 امتیاز 1.0096 آورده که بهترین امتیاز این روش است. دراقع افزایش دقت بیشتر از 400 مقدور نیست چون درآن صورت حضور ستاره ها خیلی کمتر میشود. در ردیف های 8و9 هم درصد حضور ستاره ها –روش قبلی- با این روش ترکیب شده که نتیجه بهتری نداشته. البته ازنظر زمانی برای دقت 400 زمان به 10 ساعت افزایش یافته است. این روش بهرحال خطای خیلی زیادی تولید کرده، اما هنوز امید برای بهترکردن آن هست.

در بهینه سازی بعدی تاثیر ستاره های نزدیک بطور میانیگن درنظر گرفته شده. اینهم ازنظر زمانی مثل روش قبل پیشرفتی نداشته، اما جوابها را بهتر کرده. درواقع این منطقی هم هست که از میانگین استفاده کنیم، یعنی برای هرنقطه، میانگین نیروی وارده توسط ستاره های داخل bin آن را درنظر بگیرم. این روش تاثیر خوبی روی زمان اجرا و امتیاز نسبت به روش قبلی داشته و امتیاز آن نزدیک به روش اول است. برای دقت 400 مدت زمان 6 ساعت سپری شده و خطا 108 بوده و امتیاز 0.98 شده که خیلی به روش اول نزدیک است. [در جدول](idea1_reports.docx) ردیف های 10 تا 12 که از plan009 اجراشده اند این نتایج را نشان میدهند. در ردیف های 13و14 بازهم درصد حضور ستاره ها را کمترکردن ولی امتیاز حضور 0.75 ستاره های نزدیک چیزی نزدیک 0.98 بوده ولی درمورد حضور 0.5 انها جواب چندان خوب نبوده. بدلیل زمان محاسباتی، از بالابردن دقت صرفنظر کردم چون بنظر هم نمیرسد با بیشترکردن دقت به نتیجه لازم برسم. شکل زیر مقدار خطا را برای این روش برای 100 آسمان اول train نشان میدهند:



شکل 4 - مقدار خطا برای روش plan009 برای 100 آسمان اول train. خطاها مشابه اجرای plan001 هستند.

و اما آخرین بهنیه سازی در این مورد، استفاده از روش quad-tree است. منظور این است که ابتدا با دقت کم مثلا 400 بیشترین نقطه را پیدا کنم سپس درهمسایگی اطراف آن نقطه دقت را بیشتر کنم. این روش منطقا باید خیلی بهینه باشد و واقعا هم ازنظر زمانی بهینه است ولی ازنظر امتیاز پیشرفتی نمیکند، چون یک نکته را درنظر نگرفته و آن نکته ای است که در بخش gridded benchmark گفته بودم : ممکن است روش در یک ماکس محلی به دام بیفتد درحالیکه جواب در ماکس محلی دیگری قرار دارد. این یعنی مقدار نیروی مماسی همیشه جای سیاهچاله را به ما نمیگوید.این فرض اساس روشهای بعدی قرار میگیرد. درمورد این روش یکبار با زوم 5 مرحله ای و دقت اولیه 100، و سه بار با زوم دو مرحله ای با دقت اولیه 400 اجرا کردم که بهترین جواب مربوط به اجرای 18 در جدول از plan0012 است که با دقت اولیه 400 و سپس چک کردن همه نقاط در همسایگی جواب اولیه بدست آمده که امتیاز 1.061 داشته و این درحالی است که اگر یک سوم نقاط را بررسی میکردم امتیاز 1.062 بدست میامد و این نشان میدهد این روش بهتر از این به جواب نمیرسد. اجرا درمورد دقت های اولیه کمتر، افتضاح بوده که در جدول نیاوردم. ردیف های 15تا18 [جدول](idea1_reports.docx) مربوط به این روش است.

حتی برای محاسبه محدوده اطراف از روش هیستوگرام تجمعی نیز استفاده کرده بودم، بدین ترتیب که بکمک هیستوگرام اطراف پیک نیروی مماسی را درنظرگرفتم و دوباره در آن محدوده ها دقت را بالابردم اما بدلایل ذکرشده جواب نداد.

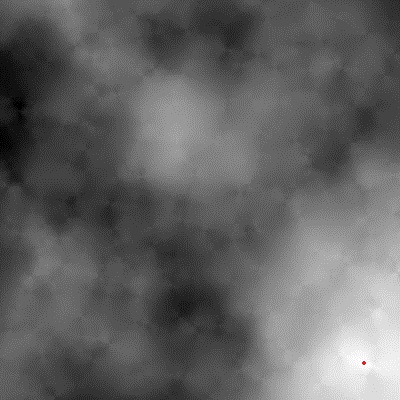
در تحلیل نتایج فوق میتوان گفت که روش نیروی مماسی که در gridded benchmark آورده شده بهتر از 0.96 امتیاز نمیاورد، و برای امتیاز بهتر باید یا روش را تغییر داد یا بسراغ حل آسمان های با 2 و 3 سیاهچاله رفت.

**2-1-2 - حل دو سیاهچاله بکمک gridded benchmark**

یک مساله ای هست : سیاهچاله های دوم و سوم روی ماکس های دوم و سوم نمیفتند. این بار از یک فرض شروع میکنم که البته فرض درستی هم هست : با حذف اثر یک سیاهچاله، میشود سیاهچاله دوم را پیداکرد. و با حذف اثر سیاهچاله ای که روی ماکس نیروی مماسی قرار دارد، و سپس محاسبه مجدد نیروهای مماسی میتوان سیاهچاله بعدی را روی ماکس جدید پیداکرد. حالا اینکه چطور بشود اثر یک سیاهچاله را حذف کرد، باید روش ها را امتحان کنم.

یک ایده این است که ستاره های نزدیک سیاهچاله را حذف کنم، اجراهای روی تک سیاهچاله نشان داده اند که این ستاره ها بیشترین تاثیر را از سیاهچاله میپذیرند. ولی با حذف ستاره های از شعاع 100 تا 4000 امتیاز در حدود 1.08 تا 1.04 تغییر کرد و بهتر از آن نشد. همچنین خطاهای اقلیدوسی آنها حدود 2000 بوده که یعنی این روش اصلا جواب نمیدهد. [درجدول دوم](idea1_2ndhalo_reports.docx)، ردیف های 5 تا 9 این نتایج را نشان میدهند.

ایده دیگر این است که ستاره هایی که بیشترین نیروی مماسی را بر مرکط سیاهچاله اول وارد کرده اند -که میتوانند در فواصل دور هم باشند- را حذف کنم. [درجدول دوم](idea1_2ndhalo_reports.docx) در ردیف اول با دقت 50 و حذف 0.75 قویترین ستاره ها امتیاز 1.016 شد و با افزایش دقت bin به 400 جواب فرقی نکرد. بهترین مقدر استانه همین 0.75 بوده. البته خطاها بقدری زیادند حدود1200 که میتوان گفت این روش هم جواب نمیدهد. در شکل زیر نشان داده شده که تغییر مقدار آستانه از 0.8 به 0.7 چقدر شکل نیروی مماسی را تغییر داده است:

شکل 5 - تاثیر حذف ستاره های با بیشترین تاثیر نیروی مماسی روی مرکز سیاهچاله. سمت راست 0.8 ستاره های با بیتشرین نیرو را حذف کرده و سمت چپ 0.7. این نشان میدهد که با حذف ستاره های قوی، شکل نیروی مماسی تغییر زیادی میکند و میتوان سیاهچاله دوم را مجددا حساب کرد.

در ایده بعدی، این دو ایده را با هم ترکیب میکنم یعنی هم شعاع و هم قدرت ستاره ها را تاثیر میدهم. همچنین تاثیر همه ستاره ها یا ستاره های نزدیک را هم بررسی میکنم. اما نتایج بازهم بهتر نشدند. ردیف های 10تا 13 از [جدول دوم](idea1_2ndhalo_reports.docx) این نتایج را نشان میدهند.

همچنین در plan020 سعی کردم مقادیر e1 و e2 را پس از حذف سیاهچاله اول کمترکنم مثلا در 0.75 ضرب کنم که این هم نتیجه ای نداشت.

**2-2 - ایده دوم، استفاده از classification**

روشی که در gridded benchmark آورده شده بود درحقیقت یک معیار چشمی برای تشخیص مرکز سیاهچاله بود، میتوان این معیار چشمی رابکمک تکنیک های image processing تشخیص داد، البته گرفتن تصویر از اطراف سیاهچاله نمیتواند مفید باشد چون تصاویر باهم زیاد فرقی ندارند، ولی میتوان از ویژگی های دیگری برای classification استفاده کرد. من مقادیر e1 و e2 را به هشت دسته تقسیم کردم، سپس با الهام از روش ویژگی های sift که مرسوم است، برای هر منطقه از آسمان تراکم مقادیر e1 و e2 را در آن 8 بازه ثبت کردم، و سپس به آنها برچسب 1 برای حضور سیاهچاله در آن منطقه و برچسب -1 برای عدم حضور سیاهچاله در آن منطقه را درنظر گرفتم، و با یک پرسپترون سه لایه و با حرکت پنجره های همپوشان بهاندازه 200\*200 روی تصویر، کار آموزش را انجام دادم، هرچند حدود 10 ساعت اجرا برای پرسپترون 20 لایه زمان برد، ولی بازهم به جواب خوبی نرسید و میشود گفت که این نوع انتخاب ویژگی اصلا خوب نبوده است.

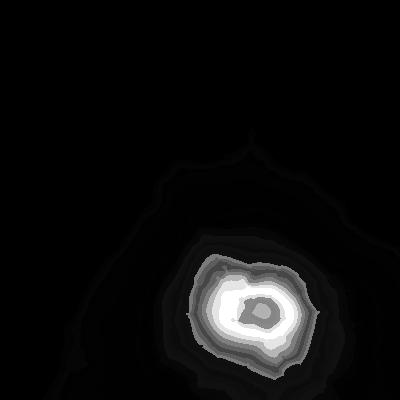
**2-3- ایده سوم، استفاده از تکنیک یافتن مرکز دایره**

اینبار از روشی که برای پیداکردن مرکز دایره پیدامیشد استفاده کردم، در این روش از هر پیکسلی یک خط در جهت زاویه گرادیان آن رسم میشد و جایی که خطها بیشتر تراکم داشتند آنجا مرکز دایره میشود. من اینبار رابطه نیروی مماسی را به این شکل درآوردم و از هر کهکشانی خطی در مرکز آن و در امتداد زاویه آن رسم میکنم، بنابراین خطهای زیادی روی هم میفتند و جاییکه سیاهچاله هست آنجا حتما خطهای بیشتری روی هم افتاده اند ولی لزوما اینطور نمیشود و شاید سیاهچاله روی ماکزیموم چندم افتاده باشد، پس جاهایی که بیشترین تراکم را دارند (درحدود0.9 ماکس) را پیداکرده و نیروی مماسی آنهارا محاسبه میکنم و بزرگترین را بعنوان جواب برمیگردانم.

[درجدول 3](idea3_report.docx) اینکار را با مقدار آستانه های مختلف آزمودم و متوجه شدم بهترین مقدار برای آستانه 0.6 است. سپس مقدار دقت bin را برای آن آزمودم و بهترین مقادیر را 800 و 900 داشتند که امتیاز را به 0.98 و 0.97 رساندن، یعنی حدود همان بهترین جواب که در plan001 و plan009 بدست آمده بودند.

**2-4- سایر ایده ها**

در plan019 فکرکردم اگر مقدار ماکس نیروی مماسی همیشه مکان سیاهچاله را نشان نمیدهد، پس حتما مقادیر پایینتر از ماکس هستند که بیشتر میشود روی آنها حساب کرد. مثلا شاید ماکس اکثر اوقات روی محدوده 0.9 تا 0.92 ماکس میفتد! بدین منظور یک آمار از همه آسمان های train گرفتم که ببینم هرکدام از مقادیر 0 تا 255 چقدر احتمال پذیرش مرکز سیاهچاله را دارند، سپس این مقادیر آماری را در شکل های نیرویهای مماسی هر آسمان ضرب کردم تا ببینم کجا احتما حضور سیاهچاله بیشتر است، ولی شبیه سازی نتایج نشان میدهد که این روش همچنان ضعف دارد و دیگر آنرا اصلاح نکردم.

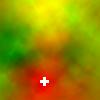
 

شکل 6 - نتیجه روش آماری برای محاسبه بهتر مرکز سیاهچاله. **سمت چپ** ضرب مقادیر آماری را نشان میدهد که احتمال حضور سیاهچاله را نشان میدهد، **سمت راست** : پس از کمی اصلاح این روش را نشان میدهد، لکه سیاه جایی را نشان میدهد که احتمال حضور سیاهچاله در آن بیشتر است، ونشانه قرمز مرکز واقعی را نشان میدهد. درمورد این شکل این مرکز درست محاسبه شده.

شکل 7 - همان آزمایش روی آسمان شماره 39 از train که نشان میدهد این روش خوب جواب نمیدهد.

یکبار دیگر فکر کردم شاید سیاهچاله فقط روی ماکس e1 میفتد یا e2 یا اینکه به یکی از آنها نزدیکتر است. شبیه سازی زیر نشان میدهد که چنین نیست.



شکل 8 - در این شبیه سازی، مقادیر e1 با رنگ سبز و مقادیر e2 با رنگ قرمز نشان داده شده اند و زنگ زرد جایی است که e1 و e2 باهم زیادهستند. سیاهچاله روی ماکس e2 قرار دارد ولی به e1 چندان ربطی ندارد.

شکل 9 - سمت راست e1 و سمت چپ e2 جداگانه حساب شده. این همان نتیجه شبیه سازی بالا را نشان میدهد.



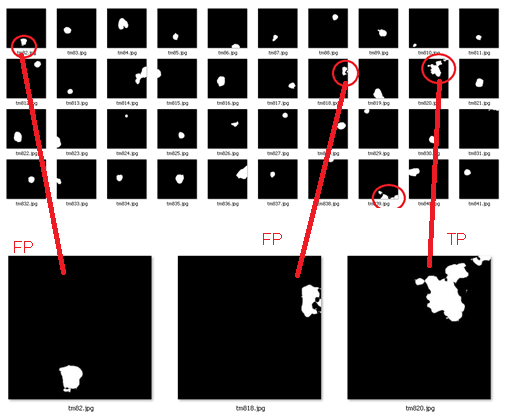
شکل 10 - همان شبیه سازی اینبار بجای e1 از منفی e1 استفاده کردم. سیاهچاله جایی قرار دارد که جمع منفی e1 و e2 بیشینه باشد واین همان چیزی است که رابطه نیروی مماسی در gridded benchmark میگفت.

**3- چرا روش ها جواب نمیدهند؟**

این سوال برایم پیش آمد که چرا بهترین جواب های من درمورد آسمانهای خاصی خطاهای زیادی دارند.

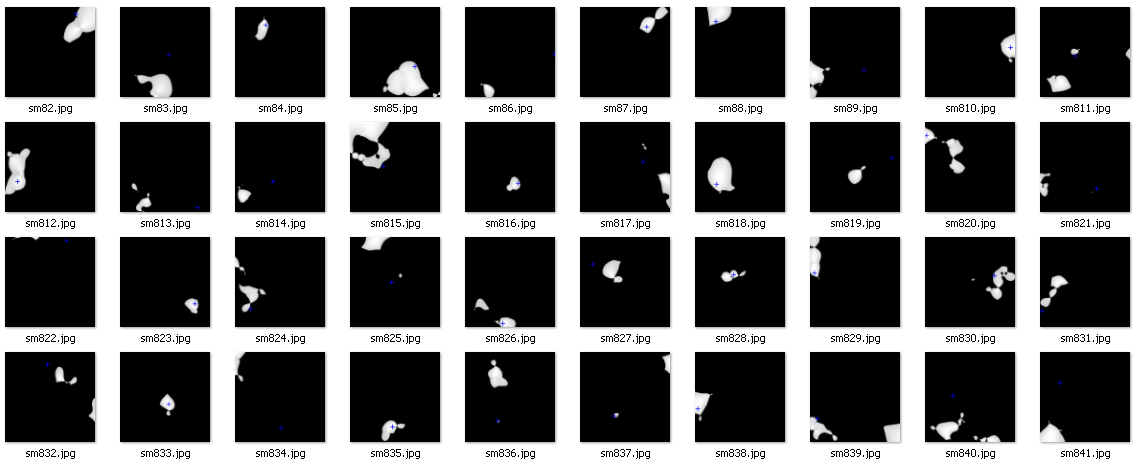
در plan021-idea1 سعی کردم آن آسمانهارا مطالعه کنم، ابتدا بکمک مقادیر e1 و e2 این کار را کردم. از همه آسمانها 6 ویژگی محاسبه کردم : مقادیر max و min و range برای e1 و e2 .سپس با بررسی این مقادیر توانستم یک قانون پیدا کنم : باید مقدار بازه e1 در حدود 0.72 بزرگترین بازه e1 بین همه ستاره ها قرار گیرد و از 0.73 آن کمتر باشد. تعداد آسمانهایی که در این شرط حضور دارند 18 تا است. پس دنبال قانون دیگری رفتم، و شرط گذاشتم محدوده های 6 ویژگی یادشده نباید از min محدوده های 6 ویژگی برای آن 8 آسمان خاص کمتر و نباید از max شان بیشتر باشند. در این شرط دو آسمان اظافی هم واردشدند بعنوان FP . اما همین شرط را که روی آسمان های test بردم، هیچ آسمانی چنان شرطی نداشت.

پس در plan021-idea2 دنبال روش دیگری رفتم. متوجه شدم وقتی در آسمان های مشکل دار فقط بخش های با مقدار بیشتر از 0.8 ماکس نیروی مماسی را نشان میدهم، تصویر شامل دو کامپوننت مجزا میشود که جواب من روی یکی از این کامپوننت ها واقع میشد درحالیکه جواب اصلی روی کامپوننت دیگر واقع شده بود. البته 4 آسمان دیگر هم بعنوان FP واردشدند که نتوانستم با morphology آنهارا ازبین ببرم چون دراینصورت آسمان 69 که بیشترین خطارا داشت ازبین میرفت.



شکل 11 - آسمانهای مشکل دار دارای دو کامپوننت مجزاهستند. البته با این فرض تعدادی FP هم داریم. درشکل بالا 4 آسمان با این فرض پیداشده اند که دوتای آنها FP هستند.

بهرحال همین آزمایش را روی آسمانهای test انجام دادم و متوجه شدم اکثریت آنها همین مشکل را دارند، مخصوصا 40 آسمان اول که تک سیاهچاله ای هستند. شکل زیر برخی آسمانهای 1 تا 40 را نشان میدهد :

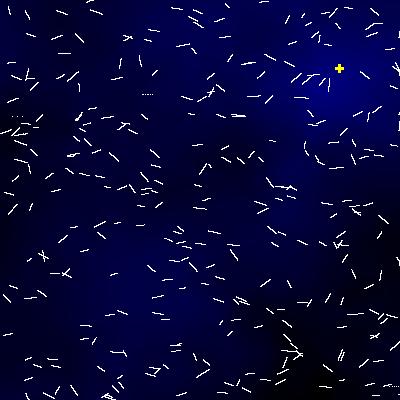


شکل 12 - آسمان های test اکثرا شامل بیشتر از دو کامپوننت هستند. درحالیکه آسمان های train اینطور نیستند.

این نشان میدهد که درحالیکه من روش خود را برای بدست آوردن جواب خوب روی اکثریت مجموعه train برده بودم از تعدادی خاص غفلت کرده بودم که داده های test از آن نوع بوده.

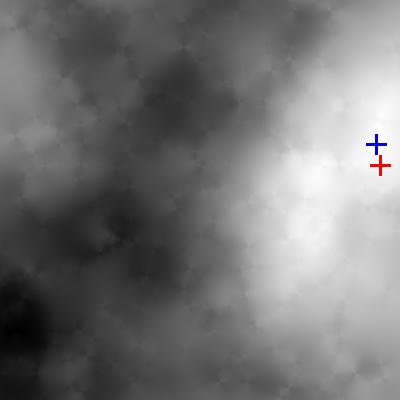
**4- شبیه سازی ها**

یک شبیه سازی انجام دادم که شکل ستاره ها و نیروی مماسی و مرکز سیاهچاله را در آسمان نشان میدهند. در این شبیه سازی میشود تاثیر سیاهچاله بر شکل ستاره ها ،و داده های نویزی را دید. البته برای مشخص بودن ستاره ها، آنها را بزرگتر از اندازه اصلی شان نشان دادم، حدود 40 برابر کشیده تر.

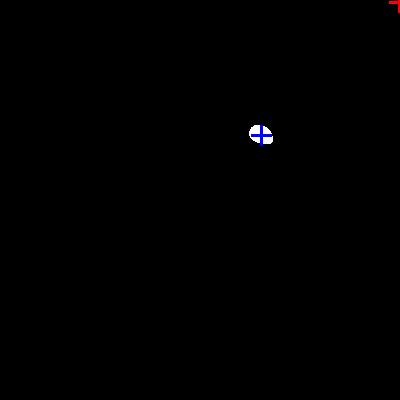


شکل 12 - یک نمونه از شبیه سازی ها روی یکی از آسمان های test. تاثیر سیاهچاله بر ستاه های اطراف، و همچنین داده های نویزی دیده میشوند (آنهایی که از سیاهچاله تاثیر نپذیرفته اند)

یک شبیه سازی هم فقط نیروی مماسی را نشان میدهد و سپس جواب اصلی را با رنگ قرمزو جوابی که من محاسبه کردم را با رنگ آبی نشان میدهد. این شبیه سازی برای اطمینان از آن است که من همیشه ماکس را انتخاب میکنم. البته از اینها نتیجه گرفتم که سیاهچاله همیشه در ماکس قرار ندارد.



شکل 13 - یک نمونه شبیه سازی نیروی مماسی آسمان 14 از train برای اطمینان از اینکه من نقطه ماکس را انتخاب میکنم و سیاهچاله هم در ماکس قرار دارد.



شکل 14 - همان شبیه سازی برای آسمان 20 از train ، ولی با نمایش مقادیر بزرگتر از 0.9 ماکس که نشان میدهد سیاهچاله همیشه روی ماکس قرار ندارد (در گوشه بالای راست)

دو نمونه دیگر شبیه سازی نیز قبلا در بین روش ها آورده شده بود که برای کاهش حجم، اینجا آنها را نمیاورم.

**5- جواب سایت kaggle**

من قبلا 5 یوزر با نامهای mortezamsp1 تا mortezamsp5 ساخته بودم. ولی زمانیکه اعلام شده بود یوزرهای تقلبی پاک میشوند من دوتای آنهارا خودم پاک کردم، و بقیه را فرصت نکردم باتوجه به سرعت اینترنت ولی خود سایت بقیه را پاک کرد و همان دوتایی که من خودم پاک کردم بودم، بطور کامل حذف نشدند و در نتایج شرکت کردند و mortezamsp4 رتبه 56 را آورد و mortezamsp3 رتبه 66 را آورد با امتیازهای 0.86 هرچند نمیشود الآن به داده های آن دسترسی داشت یا واردش شد یا دیدکه کدام آپلود بهترین نتیجه را داشته. نتیجه اکانت های پاک نشده :

