# SYNPIPE: SYNTHETIC OBJECT PIPELINE OF HYPER SUPRIME-CAM STRATEGY SURVEY PROGRAM

SONG HUANG [1] CLAIRE LACKNER [1,2] ALEXIE LEAUTHAUD [1], RYOMA MURATA [1,3], HSC SOFTWARE TEAM [1,4,5], AND MANY OTHERS.

*Draft version 0*

ABSTRACT

The Subaru Strategy Survey Program (SSP) is an ambitious multiband imaging survey using the Hyper Suprime-Cam (HSC). The HSC collaboration has developed sophisticated data reduction pipeline that will also be used by the Large Synoptic Survey Telescope. Here we introduce the SynPipe: a flexible framework to test the performance of HSC pipeline using synthetic stellar and extedned objects. Through injecting synthetic objects to images from individual exposures, SynPipe creates realistic coadd images and photometric measurements for them In this work, we demonstrate the basic photometric performance of HSC pipeline using synthetic stars and galaxies close to the expected detection limit of SSP Wide survey. We showed that . . . SynPipe is also being used to test the selection of background galaxies for weak lensing analysis, the detection of high redshift dropouts, and photometry of low surface brightness objects.

**TODO: Summarize the tests have been done**

*Subject headings:* methods: data analysis, methods: statistical surveys

## 1. INTRODUCTION

**TODO: SSP/HSC background**

- Why large imaging survey? Basic scientific application: cosmology, galaxy evolution, galactic structure, and trasient objects.
- Brief background about HSC and SSP. Could borrow a few paragraph from the camera and survey design paper, rewrite in slightly different way.
- Accurate photometry becomes very challenging, hence requires careful systematic tests.
- Briefly mentions similar efforts from DES collaboration. e.g. `Balrog` ((Suchyta et al. 2016)) and there is another one. . .

## 2. HSC PIPELINE OVERVIEW

**TODO: Brief introduction of the HSC pipeline.**

- Basic workflow of the HSC pipeline, can borrow something from Jim's paper, and rewrite in our own words.
- Emphasize the new algorithms used by HSC pipeline, and the improvements compared with other surveys.
- Explain the most important features that are worth testing:
  1. Detection of sources down to the survey limit;
  2. PSF magnitude and colors for stellar objects;
  3. cModel photometry for galaxies;
  4. Impact of "blendedness" on photometry;
  5. Impact of proximity to bright stars on photometry (?)
  6. Star-Galaxy separation (?)

## 3. SYNPIPE OVERVIEW

The following paragraph is prepared for the weak lensing shape catalog paper, only show here as a place holder . . .

- Basic design of SynPipe:
  1. Flexible, highly integrated with HSC pipeline;
  2. Why we want to inject synthetic objects onto single exposures instead of directly putting them on coadd?
- Since we simulate the real data reduction process as much as possible, SynPipe will generate the same amount of data as the real HSC pipeline. We do not plan to apply SynPipe to the entire HSC survey, but use it to test the performance of HSC pipeline on a representative subset of data.

**TODO: Need a flow-chart to visualize the procedures of SynPipe**

To evaluate the performence of `hscPipe`, including the photometry and shape measurements that are related to weak lensing analysis, we design a flexible framework (`SynPipe`[6]) to conduct systematic tests by injecting simulated images of galaxies on to the HSC images. `SynPipe` read positions and detailed model information of the fake objects from an input catalog. By default, it provides the functions to help you randomly inject objects on a single `Tract` according to required number density, or put fake objects on an evenly separated grid given expected distance between objects.

`SynPipe` uses `GalSim v1.4` (Rowe et al. 2015) as backend to reliably simulate images of fakes objects. So far, the fake object can be either star or galaxies that are described

[1]Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo Institutes for Advanced Study, the University of Tokyo (Kavli IPMU, WPI), Kashiwa 277–8583, Japan
[2]**TODO: Correct affiliation**
[3]**TODO: Correct affiliation for Tokyo University**
[4]**TODO: Correct affiliation for NAOJ**
[5]**TODO: Correct affiliation for Princeton University**
[6]https://github.com/clackner2007/fake-sources

by single- or double-Sérsic models. In stead of providing the model parameters in the input catalog, we also support the parametric models of $I_{F814W} \leq 25.2$ COSMOS galaxies provided by `GalSim.COSMOSCatalog()`, which is very similar to the input catalog of the `GREAT3` challenge (Mandelbaum et al. 2013). Using model identifications in the input catalog, `GalSim` will generate images of the best parametric models (either single-Sérsic or Sérsic-bulge plus exponential disk). Also, external shear can be added to fake object in the form of $g1$ and $g2$. For model with high Sérsic index, the simulated image can truncate after certain times of the effective radius by request to improve the speed.

To make the image simulation process as realistic as possible, `SynPipe` injects fake objects directly onto images of individual CCD at the single exposure level, instead of putting them to the already stacked, final image products (e.g. Suchyta et al. 2015). In this way, the fake objects from different single exposures will go through the exactly same detecting, stacking, and measuring procedures as real ones. Hence, the pipeline results on fake objects can share very similar systematical effects as real ones, which is important for tests related to weak lensing analysis.

Given certain input fake object, `SynPipe` converts the model magnitude into flux in requested HSC filter using the available photometric calibration. For fake galaxy, we convolve the simulated image with PSF model estimated by `hscPipe` at the desired coordinate; and we model fake star by simply scaling the PSF model to the input flux. Then, we add the PSF-convolved snapshot to all single CCD images that overlaps with it according to available WCS information. At the edge of the CCD, the simulated image will be clipped before adding to it. Meanwhile, `SynPipe` adds a new mask plane (`FAKE`) on the existed one to highlight the regions affected by fake objects. After these fake-object-added images being reduced by `hscPipe`, we match the input coordinates of fake objects with the output catalogs within certain radius. Cases of unmatched objects and multiple matches can be saved too when necessary.

Using the infrastructures provided by `hscPipe`, the above process (e.g. the injection of fake objects) can be easily parallelized. Hence, the overall efficiency of `SynPipe` is almost identical with the reduction processes of real data.

## 4. TEST DESIGN

**TODO: A lot to discuss here....**

### 4.1. *Selection of HSC Data*

**TODO: So far, all the tests have been done on WIDE tract 8766 and 8767, but we should be more careful in this selection. The following is my proposal:**

- In this work, we will focus on the WIDE survey. The results should also be instructive for the DEEP and UDEEP survey.
- Pick three WIDE tracts with median, best, and worst seeing in i-band;
  If we want to maximum the number of useful fake objects, we can choose the relative "clean" ones; If we want to test the impact of bright stars on photometry of nearby objects, we can choose one tract with several very bright stars.
- I think we want to show multiband results, but may not necessary to use all five bands. For instance, we can only use g, r, and i band data.
  Meanwhile, if there is concern about the performance in Y-band, we can test Y-band separately, or just include all five bands.

### 4.2. *Synthetic Stellar Object*

- We simulate stellar object using the reconstructed PSF model given the desired coordinates at individual exposure level. The spatial variantion of seeing, and seeing difference among all exposures are naturally taken into account.
- **Magnitude distribution**:
  1: If we just want to test the PSF magnitude, we can use very simple distribution of magnitude, e.g flat distribution between 21 to 27 magnitude.
  2: If we want to also focus on the detection limits and star-galaxy separation at faint end, we can inject more faint stars. (e.g. Claire used to just inject large number of 26, 27 magnitude stars, and see how many are recovered).
- **Color distribution**:
  1: We can just make all color equal 0.0. This way, we can still test the PSF magnitude in each band independently.
  2: Magnitude-dependent color distribution from "real" stars. This could be taken from other studies, or from the observed HSC stars. The only issue is that at very faint end, do we have reliable information about the color distribution ? (and does it matter?)
- **Number of Stars**:
  For point sources, we can afford adding a large amount into one Tract. (e.g. 500-1000 per CCD). SH will check papers from other projects to see how many is considered large enough

### 4.3. *Synthetic Galaxy*

**TODO: There are multiple decisions to be made here!!**

- **Types of Synthetic Galaxies**:
  1. Only single-Sérsic object or also want to try bulge+disk model? 2. Besides the single-Sérsic model, which is "more realistic", do we want to test basic Exp and deV models. The Exp and deV models can be used to sanity-check the cModel photometry. I have done this before, and the results look quite reasonable for `hscPipe v3.8.5`.
- **Separate the low-Sérsic and high-Sérsic index ones?**
  In the tests conducted by SH, the accuracy of cModel photometry shows clear dependence on Sérsic index. In some of the tests, SH has separated the sample into two groups using Sérsic index lower/higher than 2.5 to highlight such difference. If we will only single-Sérsic model, SH prefers to use this approach; If we choose to test Exp and deV models as while, it may not be necessary to do so; we can still show the dependece of systematical error on Sérsic index.
- **Magnitude Distribution and Limit**:
  Since we will be using Claire's catalog from COSMSO galaxies, we can simply randomly sample the given mangitude distribution. But, the amount of galaxies that

are brighter than 22.0 mag is quite small, if we want to have reliable statistics at the bright end, we can also test the bright and faint galaxies separately. SH has done this before.

- **Structural Parameters**:
  1. Sérsic index should be between 0.5 and 5.0. There are galaxies that appear to be more extended than that, but SH thinks single-Sérsic is no longer appropriate model for them. 2. Should have limit on effective radius (actually, the average surface brightness within effective radius) to exclude too "diffuse" objects. There are interesting diffuse galaxies, but in Claire's catalog, most of the diffuse models are due to problematic fitting results, especially at the faint end. SH thinks that they should be removed.

- **Color Distribution**:
  1. Should mention that color gradient is not considered here. (which is possible if we change the size and/or Sérsic index in different band, or use different colors for bulge and disk. 2. Again, we could assume all color equal 0.0, and focus on the photometry in each band separately. 3. Or we can use naive, but "reasonable" color: e.g. bluer color for low-Sérsic objects; redder color for high-Sérsic objects, which is used by SH for previous test. 4. In principle, we can use the rich photometric data in COSMOS to come up with a realistic multiband catalog for Claire's model.

- **Blendedness of Galaxies**:
  1. We can avoid blending between two fake galaxies by designing the input catalog more carefully; or inject fake galaxies on a grid with large separation. 2. To test the impact of blendedness on the photometry, SH used to separate the galaxies with large blendedness from the pipeline. Meanwhile, we can only add fake galaxies near real objects, or adding pairs of fake objects.

### 4.4. *Running SynPipe Test*

Briefly explain how the SynPipe runs are conducted. **TODO: Do we want to run afterburner.py, and test the color using PSF-matched aperture photometry?**

- Prepare input catalogs, avoid adding fake objects on pixels affected by saturated or interpolated masks.

- Run stack.py and multiband.py using the default configuration of HSC pipeline; Do we also want to test the cModel photometry using per-pixel variance information? Or should we try to turn on the garbage suppression?

- Briefly explain the matching processes:
  1. Match the input catalog with the pipeline outputs using 2 pixel matching radius; In case of multiple matches, all matches are kept in the catalog, but only the closest one is used for comparison.
  2. Match the input catalog with the pipeline results from data without synthetic objects using a slightly larger matching radius. This is mainly to see how many synthetic objects are heavily "contaminated" with the real objects.

## 5. RESULTS

### 5.1. *Detection Limits of Point Sources (Optional)*

It seems to SH that, if we seriously want to discuss this topic, we need to design the tests very carefully; and the detection limits of point sources can be tested using the "sky object". But, we can still say a few words here. Excluding the synthetic stars affected by problematic pixels and real objects, we can simply summarize the fraction of detected point sources at different magnitude and seeing condition. We can also show a few examples of undetected stars.

### 5.2. *Star-Galaxy Separation (Optional)*

Again, this topic itself should deserve its own paper, and I know that Jose has made a lot of improvements. However, it is quite straightforward for SH to briefly mention something:
1. Fraction of stars that are classified as galaxies; Their magnitude and blendedness distributions.
2. Fraction of galaxies that are classified as stars; Their magnitude, blendedness, and structural parameter distributions. (In SH's previous test, faint, exponential-like galaxies are slightly more likely to be classified as stars)

### 5.3. *Photometric Performance*

#### 5.3.1. *General Behaviours (Optional)*

- What fraction of stars have failed PSF magnitude (should be very low)

- What fraction of galaxies have failed (1) Kron, (2) cModel photometry; Their magnitude, blendedness, structural parameter distributions; (In SH's previous tests, there are still a few percents of fake galaxies that have failed cModel; and it is slightly more likely for galaxy with high Sérsic index to have failed Kron or cModel.)

#### 5.3.2. *PSF photometry of Stars*

- Magnitude v.s. $\Delta$(Magnitude) for different seeing conditions.

- Magnitude v.s. $\Delta$(Color); Color v.s. $\Delta$(Color) for different seeing conditions.

- Discuss the impact from "blendedness" and bright object masks.

#### 5.3.3. *Kron photometry of Galaxies*

- Magnitude v.s. $\Delta$(Magnitude) for different seeing conditions. Can also separate them into low- and high-Sérsic index groups.

- Magnitude v.s. $\Delta$(Color); Color v.s. $\Delta$(Color) for different seeing conditions; Can also separate them into low- and high-Sérsic index groups.

- Discuss the impact from "blendedness" and bright object masks.

#### 5.3.4. *cModel photometry of Galaxies*

- Magnitude v.s. $\Delta$(Magnitude) for different seeing conditions. Can also separate them into low- and high-Sérsic index groups.
- Magnitude v.s. $\Delta$(Color); Color v.s. $\Delta$(Color) for different seeing conditions; Can also separate them into low- and high-Sérsic index groups.
- Discuss the impact from "blendedness" and bright object masks.
- If we decide to test the Exp and deV models, we can show results for cmolde.exp and cmodel.dev photometry first

#### 5.3.5. *Color using PSF-matched Aperture Photometry (Optional)*

**TODO: Depends on whether we want to test the afterburner.**
Focus on whether the afterburner results can help improve the accuracy of color for galaxies, especially the ones with very high blendedness.

#### 5.3.6. *Shape and Structural Parameters of Galaxies (Optional)*

**TODO: Decide whether to include this.** Using the default pipeline configuration, cModel still shows serious biases in recovering the size and shape of galaxies. We can:
1. Do not show results here, but mention a few words in the Future Plans section below: "We are aware that the cModel and the configuration has some issues, we will use SynPipe to see if we can improve this."
2. Show the results, explain the main reasons for these problems, and use it as a warning to people who want to use these results.
3. Comparisons with the results from test using the per-pixel variance information for cModel. (SH's previous test during the Taipei meeting shows clear improvements)

Relationship between the fracDev and the Sérsic index of galaxies (or other parameters). In SH's previous test, fracDev is only useful to separate low- and high-Sérsic index galaxies when they are brighter than 21 mag in i-band.

## 6. CONCLUSIONS AND FUTURE PLANS

Summary of the paper

- Basically the information in abstract, but in different way.
- Summarize the basic performance of HSC pipeline into a few bullet points.

Other applications of the SynPipe (just mentioned a few here)

- Can be used as a standard QA tools for future data release and pipeline update; Can be used to tweak configuration of pipeline to improve the photometry of certain objects (useful for other HSC users).
- Ryoma Murata's tests and other tests related to weak lensing shape catalog.
- Ono-san's test on the detection of dropouts (have not contacted Ono-san to see if he want to be mentioned here)
- Johnny Greco and SH's test about the detection of ultra-diffuse galaxies (UDG).

Further improvements of the SynPipe

- More flexible, realistic models for galaxies.
- For many purposes, it should be Ok to directly inject synthetic objects onto the deepCoadd images. This can improve the efficiency.
- Only do photometric measurements on objects that overlap the FAKE mask plane. This can improve the efficiency of SynPipe.
- Right now, the SynPipe applies PSF convolution to synthetic object using the exactly "accurate" PSF model from HSC pipeline. We can simulate the typical PSF model uncertainty by reconstructing PSF model using a different location on the same Visit.
- Accuracy of background subtraction has not been taken into account

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.
**TODO: Full acknowledgement**

- Acknowledgements of the Python libraries

This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration, 2013).

## REFERENCES

Suchyta, E., Huff, E. M., Aleksić, J., et al. 2016, MNRAS, 457, 786