Boletim da Sociedade Astronômica Brasileira, **34**, no. 1, 145-149

PSF-free galaxy shape measurement methods for the J-PAS survey

Leonardo V. Costa[1], Eduardo S. Cypriano[1], Clecio R. de Bom[2], Bernardo M. O. Fraga[2], Andre Z. Vitorelli[3] & +J-PAS collaboration

[1] IAG, Universidade de São Paulo. e-mail: leonardo.vieira.costa@usp.br, eduardo.cypriano@iag.usp.br
[2] Centro Brasileiro de Pesquisas Físicas. e-mail: clecio@debom.com.br, bernardo@cbpf.br
[3] CosmoStat, CEA, Université Paris-Saclay. e-mail: andrezvitorelli@gmail.com

Abstract. We present an optimization of the PSF modeling using the *softwares* SExtractor and PSFEx, plus a series of quality tests of this model. We use two methods to compute the shear values from the mini J-PAS survey data, KSB with regaussianization and convolutional neural networks, whose Pearson correlation coefficients when compared with the CFHTLenS survey are ~ 0.86 and ~ 0.88, respectively

Resumo. Nós apresentamos uma otimização do modelo da PSF usando os *softwares* SExtractor e PSFEx, junto a uma série de testes de qualidade deste modelo. Nós usamos dois métodos para calcular valores de cisalhamento a partir dos dados do levantamento mini J-PAS, KSB com regaussianização e redes neurais convolucionais, cujo os coeficientes de correlação de Pearson ao serem comparados com o levantamento CFHTLenS são ~ 0.86 e ~ 0.88, respectivamente.

Keywords. Galaxies: general – Gravitational lensing: weak – Techniques: miscellaneous – Catalogs

1. Introduction

Normally the average of galaxies' ellipticities is randomly orientated ($\langle e_i \rangle \approx 0$), once we don't have a preferred direction in the universe, however, the presence of weak gravitational lensing introduces a bias on this average ($\langle e_i \rangle \approx \gamma$), which is called shear and can be decomposed into two ellipticities components γ_1 and γ_2. Fig. 1 illustrates the generated distortions on a circular source by each combination of γ.

In the case of galaxy clusters, these distortions are perpendicularly oriented to the center of the cluster. We can use these ellipticities to calibrate the optical mass proxies by stacked weak lensing and determine cosmological parameters using galaxies cluster counts (Weinberg et al. 2013), which are one of the main cosmological probes for the Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS). In this context, the principal source of systematic errors is the shear estimation from the shapes of faint background galaxies, which are strongly affected by the image's Point Spread Function (PSF).

Because of that, we present an optimization of the PSF modeling using the *softwares* SExtractor and PSFEx, plus a series of PSF model quality tests, such as the Rowe statistics and other null tests, which attest to the validity of our model.

Using the GalSim HSM module, we choose the KSB ReGaussianization method (Hirata & Seljak 2003) to compute the shear values from the mini J-PAS survey data, which is the first public data of the J-PAS. In addition to this method, we use convolutional neural network (CNN) techniques (Ribli, Dobos, & Csabai 2019) to obtain PSF-corrected shear values from given galaxy images. The chosen model was an EfficientNet (Tan & Le 2019). We compare our shear results with a deeper survey, the CFHTLenS. These two methods are faster alternatives to the state of art forward-fitting methods (Huff & Mandelbaum 2017; Sheldon & Huff 2017; Zuntz et al. 2014) and once we expect about $4 \sim 6 \times 10^7$ galaxies in the full J-PAS catalog, the CPU time becomes another parameter to be optimized.

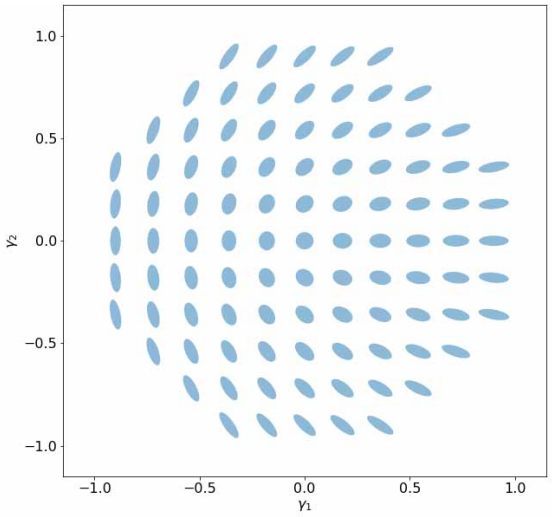

Figure 1. The shear distortion on a circular source. Image obtained from Meneghetti (2021).

2. Quadrupole moment

Is from the quadrupole moments that KSB-Regauss method does the shape and size measurements. Here we present some base equations needed for this work. The second-moment tensor of the light distribution $I(\theta)$ is defined as (Blandford et al. 1991),

$$Q_{ij} = \frac{\int I(\theta) w[I(\theta)](\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j) d^2\theta}{\int I(\theta) w[I(\theta)] d^2\theta}, \quad i, j \in 1, 2\,. \tag{1}$$

which we can extract two important size parameters, the trace,

$$T = Q_{11} + Q_{22}, \tag{2}$$

and the characteristic size σ,

$$\sigma^2 = \sqrt{\det(Q)} = \sqrt{(Q_{11}Q_{22} - Q_{12}^2)}, \tag{3}$$

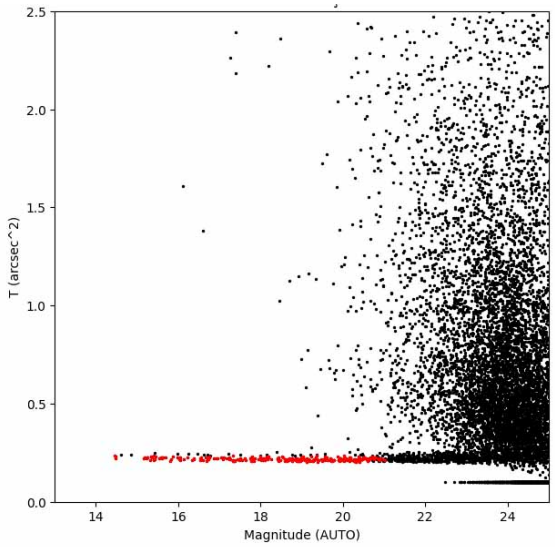

Figure 2. Example of our star selection in the magnitude and T size space parameter for AEGIS 1 field in the r-band. Black dots are all objects and the red ones are the selected stars.

these two quantities are proportional to the area of the object and can be related as, $T = 2\sigma^2$. We can calculate ellipticities in two ways using these size measurements,

$$\epsilon = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22} + 2\sqrt{Q_{11}Q_{22} - Q_{12}^2}} = \epsilon_1 + i\epsilon_2, \quad (4)$$

$$\chi = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}} = \chi_1 + i\chi_2, \quad (5)$$

both differ only in absolute value. Equation 4 is the way that CFHTLenS do its measurements, and Equation 5 is how KSB-Regauss does. They can be converted into each other as follows,

$$\epsilon = \frac{\chi}{1 + (1 - |\chi|^2)^{1/2}}; \quad \chi = \frac{2\epsilon}{1 + |\epsilon|^2}\,. \quad (6)$$

Our shear catalog uses the Equation 4 definition for ellipticities.

3. Star Selection and PSF optimization

Once stars are point sources, their images are very correlated with the PSF in their respective position. Using a star sample we can model the PSF, but we need to set some constraints in star selection to avoid bias in the PSF model.

We use the following criteria for star selection: we choose those with a magnitude between 16 $\sim$ 20, signal-to-noise ratio > 50 and SExtractor's stelar classification parameter CLASS STAR > 0.9. In addition, we make cuts using the trace T to select the stellar locus (Fig. 2). In this work, we use data from the AEGIS fields 1 to 4 in the g'r'i' bands.

4. Validation tests

This section presents a series of validation tests that we did to attest to the quality of the PSF model and our shear catalog.

Table 1. The average relative errors of the σ sizes and ellipticities between stars and PSF in all fields and bands.

	$\langle f_{\delta\sigma}\rangle$	$\langle f_{\delta e_1}\rangle$	$\langle f_{\delta e_2}\rangle$
Model	-0.0001 ± 0.0005	-0.0001 ± 0.0001	0.0001 ± 0.0001
Test	-0.0002 ± 0.0006	-0.0002 ± 0.0002	0.0001 ± 0.0002

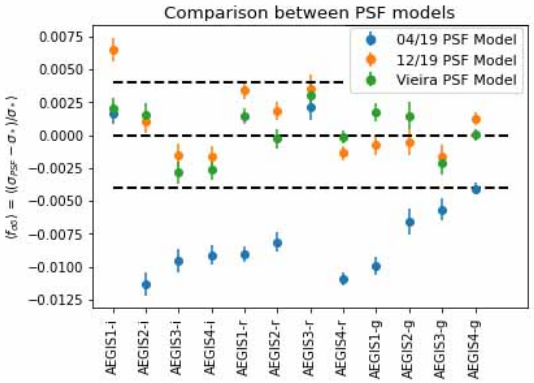

Figure 3. The average of relative residuals of σ per field and band. The traced lines represent the weak lensing thresholds. The average of all fields/bands is $\langle f_{\delta\sigma}\rangle_{04/19} = -0.0067 \pm 0.0013$, $\langle f_{\delta\sigma}\rangle_{12/19} = 0.00085 \pm 0.00072$, $\langle f_{\delta\sigma}\rangle_{V} = 0.00026 \pm 0.00052$.

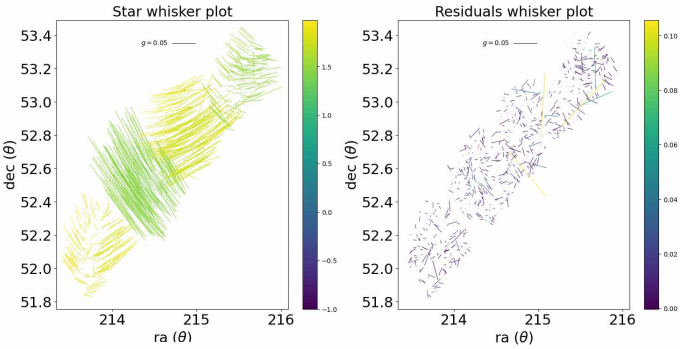

Figure 4. An example of star and residuals whisker plot in r-band. The other bands have similar behavior.

4.1. PSF Model

We separate the selected stars into two samples, one to be used in the PSF model (2/3) and the other for validation tests to check possible model bias (1/3). Since the stars and the PSF are highly correlated, we expect their relative error concerning σ and e_i to be closer to zero. In this case our results show no bias (Tab 1). We also compared these results with the mini J-PAS PSF models. In the case of ellipticities, we didn't detect substantial differences between the models. Fig. 3 shows the obtained values for σ sizes, in average our PSF model performs better.

We also computed the whisker plots, which are ellipticities maps. These maps are useful to detect possible spatial patterns of the PSF, especially when comparing the residual between stars and PSF ellipticities (Fig. 4). In our analysis, we didn't detect spacial bias in the PSF model.

4.2. ρ-statistics

In addition to the relative residuals $\langle f_{\delta\sigma}\rangle$, we can use a series of two-point correlation functions, known as ρ-statistics, to check

for error propagations of the PSF in the shear. The ρ-statistics were initially presented in Rowe 2010 and later expanded in Jarvis et al. 2016 and are defined as follows

$$\begin{aligned}
\rho_1(\theta) &= \langle \delta e^*_{PSF}(x)\delta e_{PSF}(x+\theta)\rangle;\\
\rho_2(\theta) &= \langle e^*_{PSF}(x)\delta e_{PSF}(x+\theta)\rangle;\\
\rho_3(\theta) &= \langle \left(e^*_{PSF}\frac{\delta T_{PSF}}{T_{PSF}}\right)(x)\left(e_{PSF}\frac{\delta T_{PSF}}{T_{PSF}}\right)(x+\theta)\rangle;\\
\rho_4(\theta) &= \langle \delta e^*_{PSF}(x)\left(e_{PSF}\frac{\delta T_{PSF}}{T_{PSF}}\right)(x+\theta)\rangle;\\
\rho_5(\theta) &= \langle e^*_{PSF}(x)\left(e_{PSF}\frac{\delta T_{PSF}}{T_{PSF}}\right)(x+\theta)\rangle;
\end{aligned} \tag{7}$$

with e_{PSF} being the ellipticity of the PSF, δe_{PSF} its residue with regard the stars, T_{PSF} is the size described by Equation 2 and δT_{PSF} the residual associated with this measure. Preferably we expect $\rho_i \approx 0$. The upper limits of each component are given by Jarvis et al. 2016,

$$\begin{aligned}
|\rho_{1,3,4}(\theta)| &< \left\langle \frac{T_{PSF}}{T_{gal}} \right\rangle^{-2} \delta\xi_+^{max}(\theta),\\
|\rho_{2,5}(\theta)| &< |\alpha|^{-1}\left\langle \frac{T_{PSF}}{T_{gal}} \right\rangle^{-1} \delta\xi_+^{max}(\theta),
\end{aligned} \tag{8}$$

where the parameter α represents the "leakage" of the PSF in the galaxies shapes and $\delta\xi_+^{max}$ is the sensitivity of $\xi_+(\theta)$ with respect to cosmology (Guinot et al. 2022). The α parameter can be computed by the cross-correlation between PSF and galaxies ellipticities described by the following equation,

$$\alpha = \frac{\xi_+^{gp}(\theta) - \langle e_{gal}\rangle^*\langle e_{PSF}\rangle}{\xi_+^{pp}(\theta) - |\langle e_{PSF}\rangle|^2}, \tag{9}$$

where ξ_+^{gp} is the correlation function between galaxies and PSF ellipticities and ξ_+^{pp} is the PSF ellipticities autocorrelation function. Our results for these statistics are illustrated in Figs. 5 and 6. Part of the magnitude of our uncertainties and the sensitivity of the limits is due to the reduced area of the mini J-PAS, with the full catalog, we expect more smoothness on these limits. The thresholds used were designed for cosmic shear science (for more information about cosmic shear we recommend Kilbinger 2015), which demands more restrictive constraints. In the case of cluster counts, we can be more permissive with these limits. However, for our case, being close to these limits is still a good indicator of quality.

Our PSF leakage results are in Fig. 7. Once we share similar criteria in the uncertainties of cosmological parameter σ_8 with Jarvis et al. 2016, we can use the same threshold for α: 3% for a strict limit, and 10% for a softer limit.

4.3. Shear catalog null tests

After the PSF model tests, we create a shear catalog using KSB-Regauss. We compare the mean galaxy shear with the PSF ellipticities, T sizes of the PSF and galaxies, and with the signal-to-noise ratio S/N. The Figs. 8 to 10 exemplifies these correlations for r-band. All results for the slope m and the intercept c are in Tab 2. We expect that these correlations are compatible with zero, otherwise, we would have a leakage of these components in the ellipticity of the galaxies, in which case new corrections would be necessary.

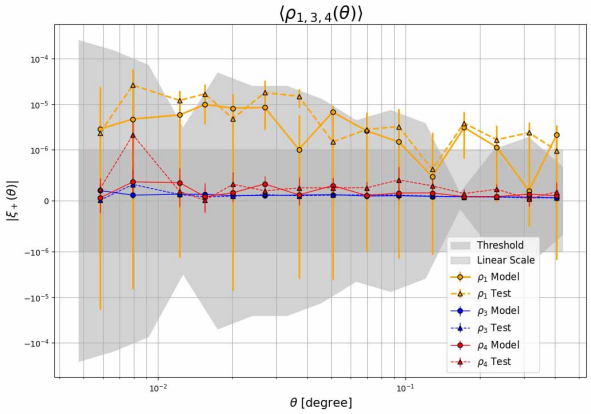

Figure 5. The ρ-statistics for ρ_1, ρ_3 and ρ_4. We used a symmetric logarithmic scale because of the uncertainties, where the dark grey area represents the linear part. The light grey represents the thresholds (Equation 8). The continuous and traced lines represent, respectively, the model and test star samples.

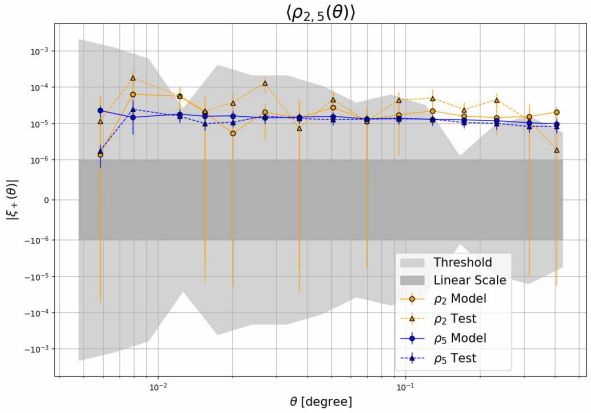

Figure 6. The same as the previous figure, but for ρ_2 and ρ_5.

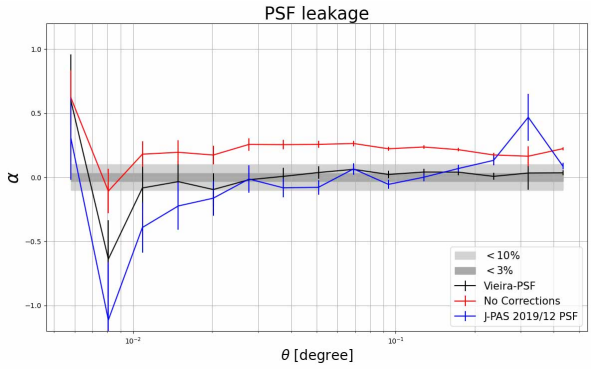

Figure 7. The α values of PSF Leakage, here we compare two PSF models, one made by us and another available by J-PAS. We also compare the case where there are no PSF corrections. The grey area represents the soft (dark) and hard (light) limits.

5. Shape measurements

We present our results of shape measurements by KSB-Regaus and CNN. As mentioned, the CFHTLenS will be used as validation data, especilly to compute the Pearson coefficient ρ.

Correlation	e_i	g-band	r-band	i-band
Shear T_{PSF}	e_1	$m = 0.015 \pm 0.027$ $c = -0.014 \pm 0.018$	$m = -0.031 \pm 0.038$ $c = -0.015 \pm 0.020$	$m = -0.002 \pm 0.019$ $c = 0.000 \pm 0.012$
	e_2	$m = -0.055 \pm 0.051$* $c = 0.041 \pm 0.034$*	$m = -0.13 \pm 0.05$* $c = 0.066 \pm 0.027$*	$m = 0.010 \pm 0.018$ $c = -0.010 \pm 0.011$
Shear T_{Gal}	e_1	$m = -0.007 \pm 0.018$ $c = 0.005 \pm 0.023$	$m = 0.003 \pm 0.014$ $c = -0.002 \pm 0.014$	$m = 0.012 \pm 0.009$* $c = -0.013 \pm 0.010$*
	e_2	$m = -0.0140 \pm 0.0069$* $c = 0.0222 \pm 0.0089$*	$m = 0.0077 \pm 0.0086$ $c = -0.0074 \pm 0.0086$	$m = 0.003 \pm 0.011$ $c = -0.006 \pm 0.011$
Shear e_1-PSF	e_1	$m = -0.48 \pm 0.27$* $c = 0.0004 \pm 0.0038$	$m = -0.17 \pm 0.21$ $c = 0.0022 \pm 0.0032$	$m = -0.04 \pm 0.15$ $c = -0.0007 \pm 0.0054$
	e_2	$m = -0.26 \pm 0.49$ $c = 0.0060 \pm 0.0071$	$m = -0.04 \pm 0.14$ $c = -0.0004 \pm 0.0022$	$m = 0.04 \pm 0.11$ $c = -0.0034 \pm 0.0040$
Shear e_2-PSF	e_1	$m = -0.016 \pm 0.14$* $c = -0.0000 \pm 0.0023$	$m = -0.019 \pm 0.075$ $c = 0.0019 \pm 0.0024$	$m = 0.02 \pm 0.15$ $c = 0.0004 \pm 0.0035$
	e_2	$m = -0.36 \pm 0.25$* $c = 0.0023 \pm 0.0042$	$m = -0.06 \pm 0.11$ $c = -0.0016 \pm 0.0036$	$m = 0.01 \pm 0.14$ $c = -0.0031 \pm 0.0032$
Shear S/N	e_1	$m = 0.00092 \pm 0.00087$* $c = -0.020 \pm 0.016$*	$m = -0.00015 \pm 0.00028$ $c = 0.0052 \pm 0.0072$	$m = 0.00024 \pm 0.00020$* $c = -0.0056 \pm 0.0054$*
	e_2	$m = 0.0002 \pm 0.0011$ $c = 0.001 \pm 0.021$	$m = -0.00006 \pm 0.00021$ $c = 0.0006 \pm 0.0053$	$m = 0.00003 \pm 0.00021$ $c = -0.0038 \pm 0.0057$

Table 2. Values of m e c for each correlation. * indicates measurements which zero compatibility is bigger than 1σ, although all values are within 3σ compatibility.

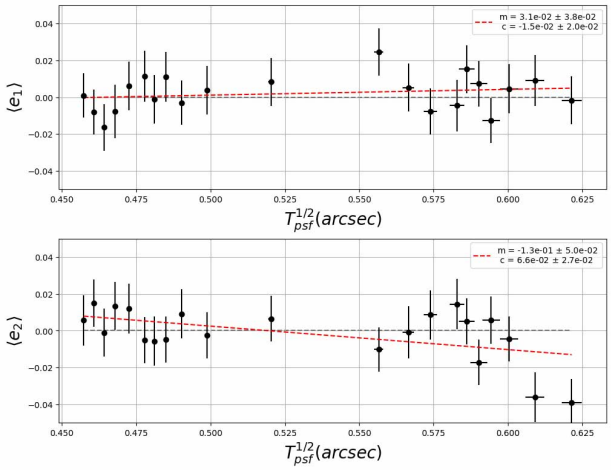

Figure 8. The mean galaxy shear in function of T_{PSF}.

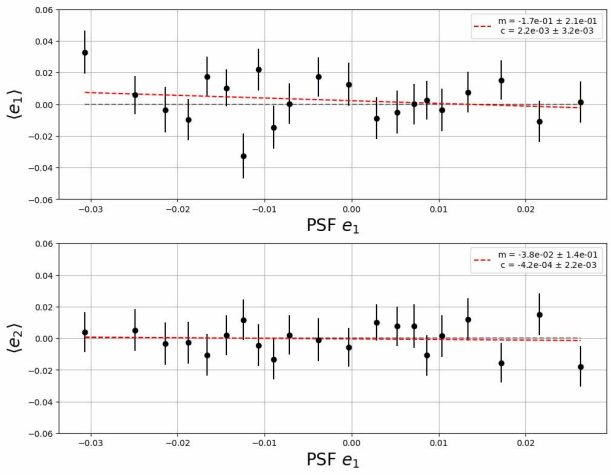

Figure 9. The mean galaxy shear in function of e_{1-PSF}.

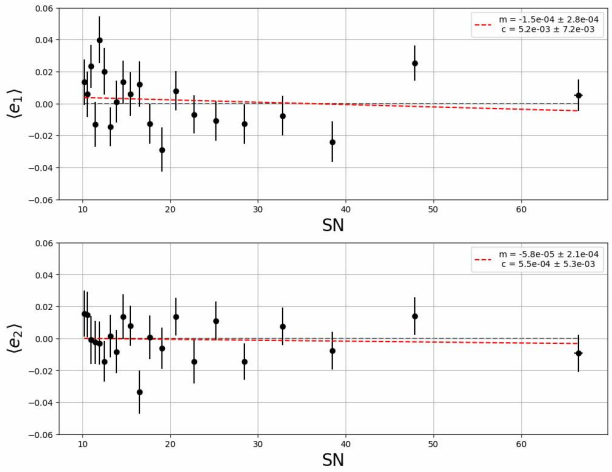

Figure 10. The mean galaxy shear in function of S/N.

5.1. KSB-Regauss

Our tests demonstrated that the r-band has more objects passing through our cuts and it also has better Pearson coefficients in all S/N in comparison to the other bands (Fig. 11). Furthermore, the HSM module can't calculate some galaxies' ellipticities, in which case it registers the measurement as $e_i = NaN$. These losses for bands g'r'i' are respectively 2.4%, 1.5%, 2.3% of the total number of galaxies measured, and again r-band performs better. Therefore, the r-band was chosen to make our shear catalogue. We can also correct the offset present in Fig. 12 using its residuals, removing the remaining multiplicative and additive bias. By doing this, we obtain $\rho = 0.86$.

For the optical mass proxies by stacked weak lensing we need that the uncertainties associated with the slope m and mass M be equal to $\Delta m < \Delta \ln M/1.5$, our calculations lead to $\Delta \ln M \approx 3\%$, since the obtained $\Delta m \sim 1\%$, this condition is satisfied.

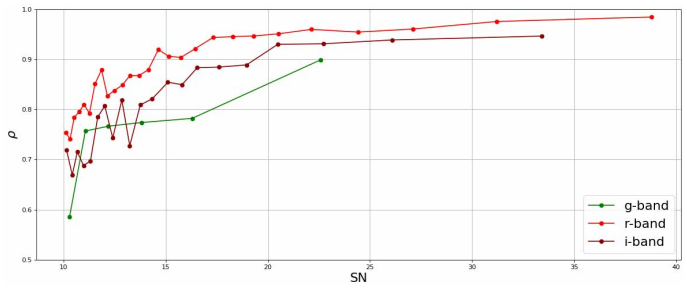

Figure 11. The Pearson coefficient in the function of S/N for g'r'i' bands. Each point represents the average of a group of 200 measurements of galaxy ellipticities.

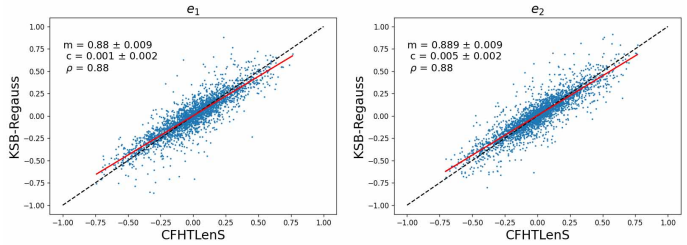

Figure 12. Comparison between KSB-Regauss and CFHTLenS ellipticities measurements.

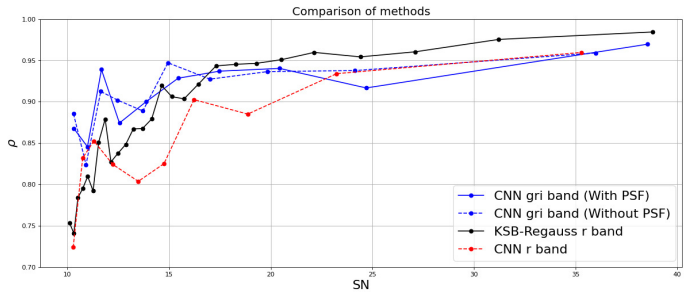

Figure 13. The Pearson coefficient in function of S/N for the CNN and KSB-Regauss.

5.2. Convolutional neural network

Our CNN receives vignettes of 51 × 51. The input used was the galaxy images and their respective PSF images, and the CFHTLenS ellipticities, which the CNN will try to reproduce. We trained models with the bands individually and with the composition of the three. To train our model, we use ∼ 2.7k galaxies, where 400 were separated to test. Using Data Augmentation we improve our training dataset in 8 times. Fig. 13 compares the CNN results with KSB-Regauss in r-band. Since the three bands compositions shows better results, it will be our main model for the CNN. In this case we have obtained $\rho = 0.88 \sim 0.90$ and $\Delta m \sim 2.5$, so for the CNN our necessary condition is not satisfied (< 2%). In part this occurs due to the reduced number of objects for the linear fit (400) in comparison with the KSB-Regauss method (∼ 2.7k).

As Fig. 13 shows, the CNN with 3 bands performs quite better in comparison to the KSB-Regauss in the r-band. Especially in $10 < S/N < 15$, which is the range that contains the faintest galaxies, for bigger S/N the CNN performs equally or less, but we have fewer objects as S/N increases. Something of note is that using PSF images does not affect the CNN measurements quality. This flag that our CNN is unable to do PSF corrections using PSF images.

6. Conclusion

We hope to demonstrate with this work that KSB-Regauss and CNN methods could be used in the J-PAS context to create reliable shear catalogs as an alternative to the standard methods. Our tests with the forward-fitting methods consumed about 7 CPU seconds per galaxy, for the KSB-Regauss and CNN, the CPU seconds per galaxy were 0.019 and 0.006 respectively. This shows how these two methods could reduce substantially the time consumed in shape measurements. Once the surveys become larger in area and depth, consequently they will have more detected objects. Reducing the CPU time needed is a crucial task.

Another key result is that the CNNs can achieve the same or better results of well know methods like the KSB-Regauss. Our CNN model is ∼ 3 times faster than KSB-Regauss and computes better the ellipticities for the faintest galaxies. Despite that, there is still space for improvement. Starting with the uncertainties in the multiplicative biases of the CNN measures. The KSB-Regauss has more outputs parameters, like σ size, resolution factor and errors, while our CNN only has the ellipticities as output. Additionally, our CNN model is unable to correct the PSF effects. One reason of this problem could be that the PSF information is not in the image parameter space, which the CNN usually obtain information, but in the convolution process (Wang et al. 2022). Other approaches are needed to consider their effects.

References

Meneghetti M., 2021, LNP. doi:10.1007/978-3-030-73582-1
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, PhR, 530, 87
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Ribli D., Dobos L., Csabai I., 2019, MNRAS, 489, 4847
Tan M., Le Q. V., 2019, arXiv, arXiv:1905.11946
Huff E., Mandelbaum R., 2017, arXiv, arXiv:1702.02600
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2014, ascl.soft. ascl:1409.013
Blandford R. D., Saust A. B., Brainerd T. G., Villumsen J. V., 1991, MNRAS, 251, 600
Rowe B., 2010, MNRAS, 404, 350. doi:10.1111/j.1365-2966.2010.16277.x
Jarvis M., Sheldon E., Zuntz J., Kacprzak T., Bridle S. L., Amara A., Armstrong R., et al., 2016, MNRAS, 460, 2245
Guinot A., Kilbinger M., Farrens S., Peel A., Pujol A., Schmitz M., Starck J.-L., et al., 2022, A&A, 666, A162
Kilbinger M., 2015, RPPh, 78, 086901
Wang H., Sreejith S., Lin Y., Ramachandra N., Slosar A., Yoo S., 2022, arXiv, arXiv:2210.01666