Gravitational lensing Super-resolution

Koo Ho Yin
HKUST
20951114
hykoo@connect.ust.hk

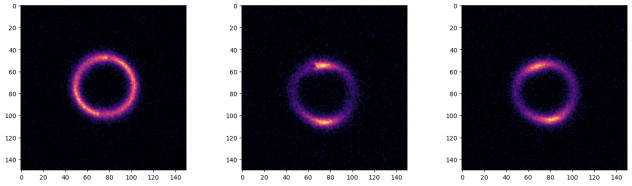
Syed Momin Ahmed Rizvi
HKUST
21017135
smarizvi@connect.ust.hk

Abstract

In this Project, we successfully upscale both the simulated and observed Gravitational lensing data provided by Google Summer of Code 2025. We have tried 3 models, namely RCAN [12], SRResNet and SRGAN. We used 3 evaluation metric, namely MSE, SSIM and PSNR. We also used bilinear interpolation as a baseline to see the advantages of using a neural network. The results show that RCAN [12] and SRResNet outperform the traditional interpolation approach in MSE, PSNR and SSIM. We also notice that the use of transfer learning greatly improved the training of models on observed data. We also verified that super-resolved images will improve classification accuracy compared to the lower-resolved image.

1. Introduction

Ever since astronomer Fritz Zwicky introduced the concept of dark matter, physicists have been searching through the entire universe for clues that could unveil the nature of these mysterious substances. The candidate of dark matter in the Beyond Standard model, e.g. WIMPS [8] and axion particles [5], have not yet been observed in the particle collider. The only evidence of dark matter comes from its gravitational attraction. This motivates the thorough study of the dark matter substructure found in strong gravitational lenses, such as Cold Dark Matter(CDM), Axion Dark Matter(Axion) and No-Substructure Dark Matter (NSDM/NO). [1]



(a) Axion (b) CDM (c) NSDM

Figure 1. Different GLS formed by different dark matter

Gravitational lensing(GLs) occurs when a massive celestial body — such as a galaxy cluster — causes a sufficient curvature of spacetime for the path of light around it to be visibly bent, as if by a lens. These lenses are located more than 5 billion light years away from Earth [2], causing its image to be blurry. This hinders physicists from observing the structure of GLs in order to correctly identify the type of dark matter that creates it.

Our aim with this project is to build a deep learning model to upscale the resolution of these images for better clarity and classification of structures.

The number of high-resolution observed GLs have been scarce, which places a large difficulty for deep learning purposes. Therefore, we first need to train our model to upscale the simulated data. Then we used our pre-trained model to upscale the observed data. Then we can use the upscaled image to do classification to see what is the potential dark matter that forms it.

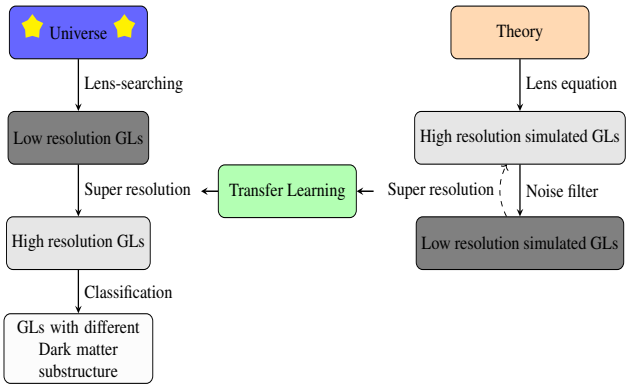


Figure 2. General Scheme in the search of Dark matter

Since the actual Gls has no labels, the classification CNN is trained on simulated Gls.

2. Related Work

2.1. Classification CNN on simulated GLs

There has been huge success in using a convolutional neural network) [1,3], such as Resnet-18,AlexNet and VGG to classify the type of dark matter that creates the gravitational lenses.

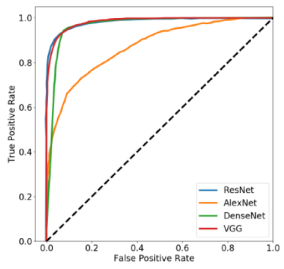


Figure 3. Success on classification task

2.2. Traditional SR neural network

There are experiments on using different traditional SR neural network such as RCAN, RDN and SRResnets [4]. Due to the simplicity of their implementations, it will be used as our base models.

Model	MSE	PSNR	SSIM
RCAN	0.00089	30.50028	0.56995
Residual Dense Network (RDN)	0.0009	30.49815	0.57196
SRResNet (18 Blocks)	0.0009	30.49482	0.57325
EDSR	0.0009	30.49347	0.57424
FSRCNN	0.0009	30.45184	0.56641
Hybrid FSRCNN (Feature Extraction Layers - Equivariant, Upscaling Layer - Convolutional)	0.00091	30.42472	0.57249
Bilinear Interpolation (Baseline)	0.00333	24.7818	0.30323
Equivariant FSRCNN	0.08281	19.35347	0.54386

Figure 4. SR task using traditional neural network

2.3. Physic-informed neural network

There are also models that use Physics-informed neural networks(PINNS) [7], that calculates the deflection angle to reconstruct the lens.

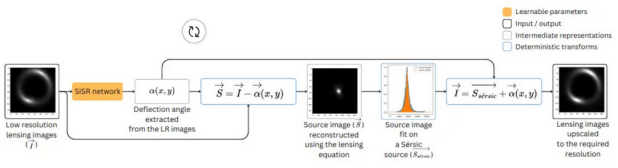


Figure 5. SR task using PINN

However, all of the above models only use simulated data to check for accuracy. The noise kernels and downsampling filters imposed on the images are known and simple. For actual observed images, the noise are unknown and more complicated. Therefore, we would like to see its effect if it is applied on observed GLs.

3. Data

For the super-resolution task, We will be using the following two simulated and observed GLs datasets provided by 2025 GSoC DeepLens evaluation test:

- Simulated data
- Observed Data

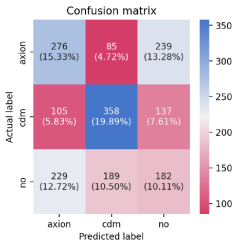
There are 10000 simuated greyscale 150×150 high resolution (HR) images and their corresponding 75×75 low resolution (LR) images. We also have 300 observed data with the same HR,LR pair.

For the classification task, we use the simulation from `lenstronomy`. The training data can be downloaded from `Github` [9].

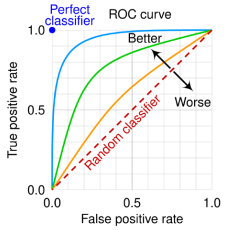
We prepared 9000 training data and 1800 of size 64×64 validation data of simulated GLs of 3 category, Axion,CDM and NSDM aforementioned.

4. Evaluation metric

For the super-resolution task, we will use the following evaluation metrics MSE (Mean Squared Error), SSIM (Structural Similarity Index), PSNR (Peak Signal-to-Noise Ratio). A higher PSNR and a lower MSE means that the difference between 2 images is small. While both PSNR and MSE capture the pixel-wise difference in image, SSIM is useful in capturing more nuanced changes in image quality, including local variations and textures. The closer SSIM to 1 means the image appear more natural to human eyes.



(a) Axion



(b) CDM

For the classification task, we will one vs rest ROC curve, which is a graph of the true positive rate (TPR) against the false positive rate (FPR) for the different possible thresholds of a diagnostic test. AUC, the Area under ROC curve(AUC) will show how accuracte the model is, the closer to 1, the more accurate the model is. The confusion matrix, on the hand, shows that which 2 classes our models will usually confuse with each others.

5. Methods

5.1. SR Model

We compared the performance of 3 models.

- **Residual Channel Attention Network**(RCAN): is a deep learning model that uses stacked residual groups with long skip connections to enable very deep networks (e.g., hundreds of layers) while easing gradient flow - each residual group contains multiple Residual Channel Attention Blocks (RCABs) [10].

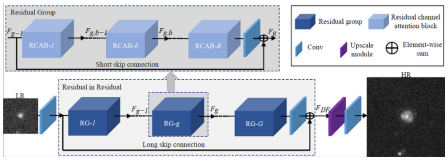


Figure 7. RCan

- **Super-Resolution Residual Network**(SRResnet): SRResNet builds upon the ResNet (Residual Network) framework, utilizing skip connections to enable training of very deep networks by alleviating the vanishing gradient problem [6].

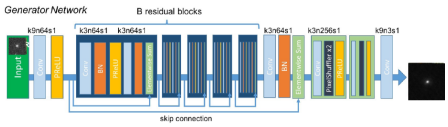


Figure 8. SRResnet

- **Super-Resolution Generative Adversarial Networks**(SRGANS): SRGAN uses a generator to upscale LR images and a discriminator to distinguish between real HR images and generated super-resolved images. The adversarial training encourages the generator to produce more realistic textures [6].

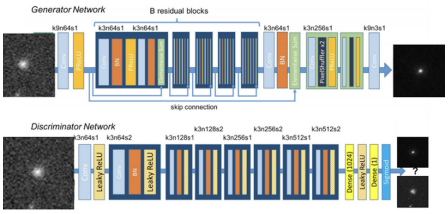


Figure 9. SRRGan

Experiments show that RCAN and SRresnet outperforms SRGAN.

5.2. Transfer learning

We tried 2 types of transfer learning methods, namely fine-tuning and adversive discriminative domain adaptation.

1. Fine-tuning

Figure 10. Fine-tuning

After obtaining our pre-trained model on simulated GLs, we can perform fine-tuning. It first freezes the feature extraction layer of the model, and trains the upsampling layer on observed GLs. This assumes that the simulated GLs and observed GLs have simular feature but different generation method. Then we unfreeze all layers and train with a smaller learning rate on observed GLs. This ensures that all layers can fit better in the observed data.

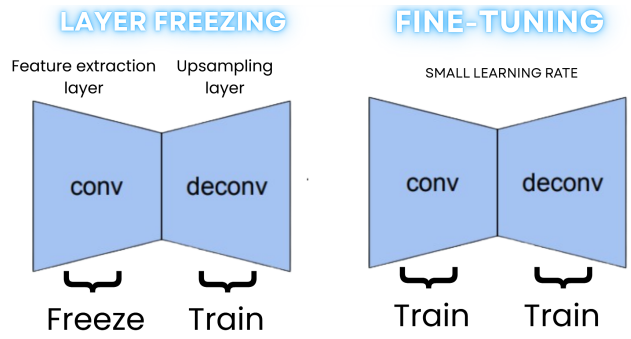


Figure 11. Fine tunning

2. Adversive Discriminative Domain adaptation(ADDA)

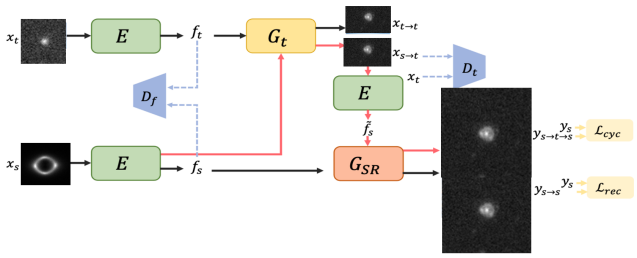


Figure 12. Fine tunning

The observed GLs(target domain) and simulated GLs(source domain) are 2 different domains, but both require up-sampling.

To perform domain adaptation using ADDA [11], we want the encorder E to encode the target feature into the source, so that the SR model can up-sample it. This can be done by confusing the discriminator.

However, since the size of our observed data is small, we fail to train the discriminator to distinguish between 2 domains, which pulled down the performance of encoder as a whole.

	Fine-tunning	ADDA
MSE	0.004609	0.009760
SSIM	0.7388	0.4262
PSNR	29.99	23.56

Figure 13. Comparison of ADDA and fine-tunning using SRresnet as SR model

Therefore, we settled with fine-tunning.

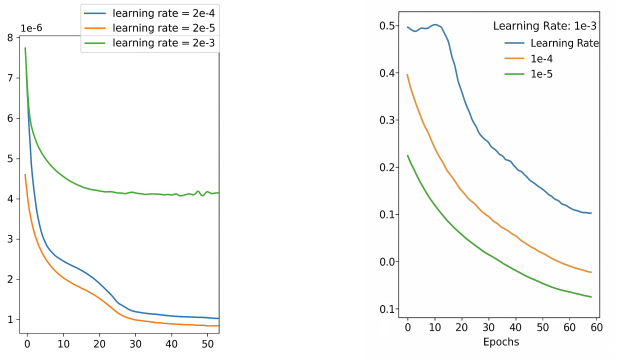
6. Experiment

Step I: SR on simulated GLs

	Bilinear (baseline)	RCAN	SRResnet	SRGAN
MSE	0.00006921	0.00006035	0.00005981	0.002784
SSIM	0.9756	0.9772	0.9765	0.5664
PSNR	41.64	42.23	42.27	26.52

We fit the training data using MSE loss and evaluate the test data using all 3 metrics.

At this stage, we also experimented to approximately obtain a good learning rate.



(a) SRResnet (b) RCAN

Figure 14. Validation loss using different learning rates

From the evaluation, it seems that both RCAN and SRResnet perform equally well on the super-resolution, while SRGAN does not do a good job. It may be due to the failure to convergence during training. We can also see that RCAN and SRResnet outperforms bilinear interpolation, this shows the edge of using a neural network.

Additionally, we can also observe that if we let RCAN train for longer, we can potentially achieve better results since the loss is still decreasing greatly. Due to time constraints with limited resources, we did not train for longer time period.

Step II: SR on observed GLs

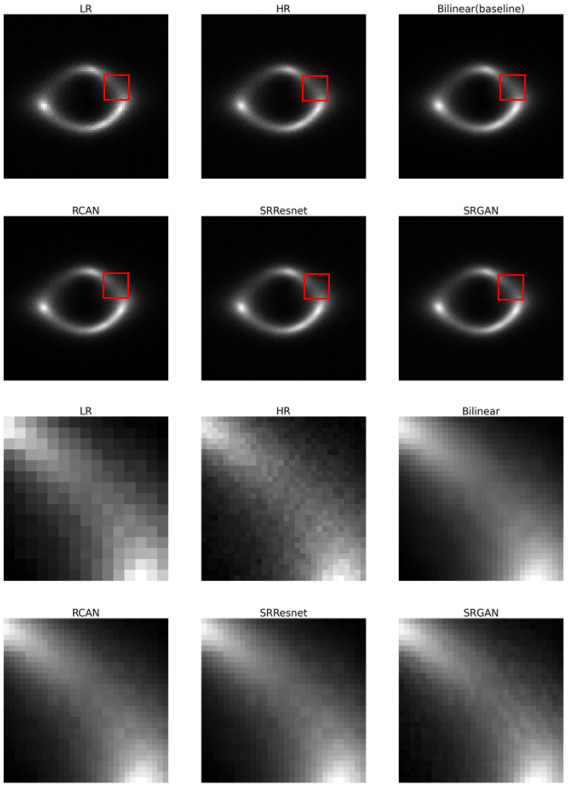
	Bilinear (baseline)	RCAN	SRResnet	Unpretrained RCAN
MSE	0.01010	0.004488	0.004609	0.008310
SSIM	0.4467	0.7756	0.7388	0.3280
PSNR	24.05	32.57	29.99	24.58

Due to the poor performance of SRGAN, we only compare the effect of SRresnt and RCAN on observed GLs. From Figure 15(b)sample(2) we can see that since real data contain much more noise than observed data, bilinear interpolation fails to remove the noise in the image, causing its SSIM to be much lower than RCAN and SRResnet.The neural network model SResnet and RCAN on the other hand, are very good at denoising. We can also notice that if we do not apply fine-tunning on RCAN, the performance of RCAN will be even worse on SSIM than bilinear interpolation. From Figure 15 sample (1,2), we can see that pre-trained RCAN turns the image blank to lower MSE. This shows that transfer learning is crucial for RCAN to upsample the observed data.

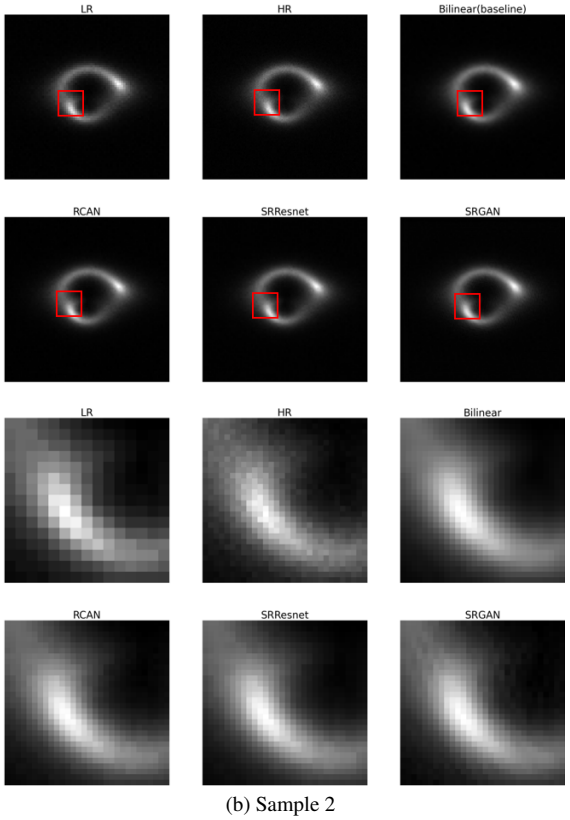
However, we can also see that the models generally perform worse on observed Gls than simulated Gls from the evaluation metrics. From the image, many fine details, such as the brim of ring in figure(15a) in LR is disappeared in SRresnet, although RCAN is able to slightly keep it. Nevertheless, no models are able to see that the bright dot in figure(15a) is actually coming from 2 light sources. This raises a concerns whether the deterministic feature of a GLs, for determining its type of dark matter, will be lost or fabricated by the models. To confront our worries, we will test our models on the accuracy of classification in the stepIII.

The following is the result of the SR image.

The red square shows local feature that is magnified in the image below.

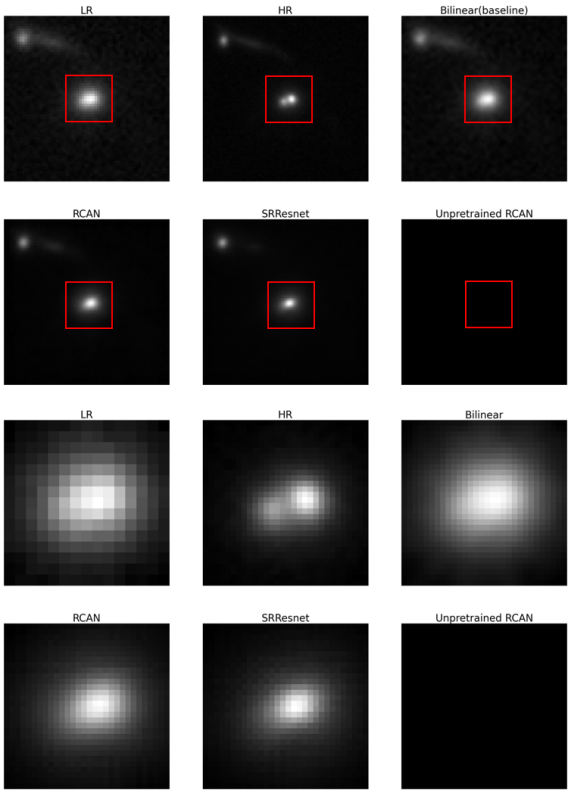


(a) Sample 1

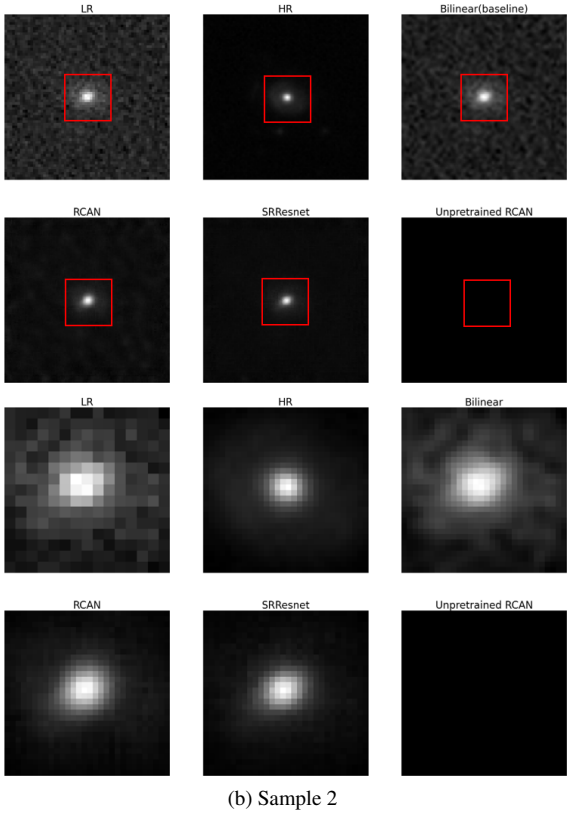


(b) Sample 2

Figure 15. Comparison of SR results on simulated SRs



(a) Sample 1



(b) Sample 2

Figure 16. Comparison of super-resolution results on observed GLs

Step III: Classification on simulated GLs

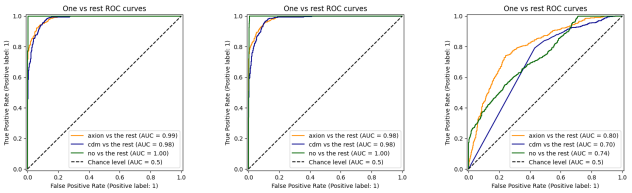
To verify our assumption, a high resolved image will increase classification accuracy than low-resolved image.

Then we down-sample it to size 32×32 then up-sample it using our SR-model, which is pretrained using all 3 categories.

	HR	LR	SR
accuracy	0.9200	0.3417	0.9167

Figure 17. Accuracy of classification using all 3 category to train SR model

From the data, we can see that high-resolved image drastically improved classification using simulated data. And we can also see that the accuracy between HR and SR is close, which shows the SR model has low reconstruction loss.



(a) HR classification (b) SR classification (c) LR classification

Figure 18. Classification comparisons between high resolution (HR), super resolution (SR), and low resolution (LR) images.

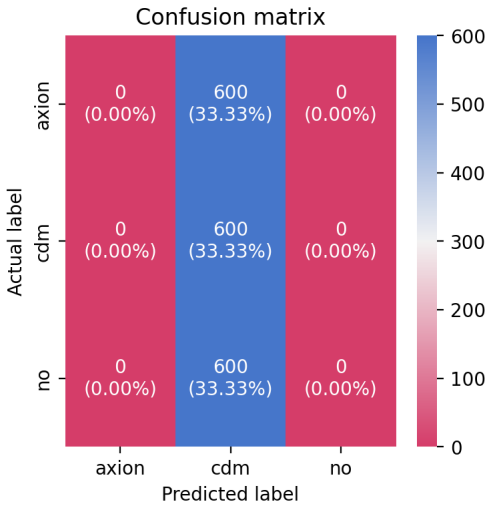
From the ROC curve of HR and SR, we can see that the model is very good at identifying NSDM, with exactly 1 AUC score. For LR, we can see that the AUC of all the 3 categories are not decent at all.

On the other hand, we can also notice that if the SR-model has not been trained on a certain type of gravitational lenses, it is very probable for the up-sampled SR image to be misclassified.

This time, we only allow our SR model to be trained on CDM catergory.

	HR	LR	SR
accuracy	0.9200	0.3417	0.3333

Figure 19. Accuracy of classification using only CDM to train SR model



This shows that SR-model will very likely to add in artificial features to the gravitational lenses that does not exist in LR image. If the actual dark matter is not made up of any 3 category of dark matter, it will not be detected.

7. Conclusion

In this project, we observe the power of neural networks on the super-resolving gravitational lenses. This type of technique can likely be used for other astronomical images such as black holes and galaxies. Although super resolution can improve classification accuracy if the dark matter are known by the SR model, it will fail if the dark matter are not known. In the future, we can investigate what type of features are learned by the model for classification by representation learning. This may help physicists to determine the property of dark matter with better ease.

8. Code

We are willing to share this report and code with others. The code for our experiments is available on the following github repository.

Additionally, please refer to the references below for further elaboration.

References

[1] Stephon Alexander et al. Decoding dark matter substructure without supervision. *arXiv preprint*, arXiv:2008.12731, 2020.

[2] NASA. Focusing in on gravitational lenses, 2023.

[3] Michelle Ntampaka et al. Deep learning the morphology of dark matter substructure. *arXiv preprint*, arXiv:1909.07346, 2019.

[4] pranath reddy. Super-resolution for strong gravitational lensing.

[5] J. Preskill, M. B. Wise, and F. Wilczek. Cosmology of the invisible axion. *Physics Letters B*, 120(1-3):127–132, 1983.

[6] Pranath Reddy. Super-resolution for strong gravitational lensing.

[7] Anirudh Shankar. Physics-informed unsupervised super-resolution of lensing images: Gsoc 2024 x ml4sci. *Medium*, 2024.

[8] G. Steigman and M. S. Turner. Cosmological constraints on the properties of weakly interacting massive particles. *Nuclear Physics B*, 253:375–386, 1985.

[9] M. Toomey. Deeplensesim.

[10] Sik-Ho Tsang. Review: Rcan — deep residual channel attention networks (super resolution).

[11] Wei Wang, Haochen Zhang, Zehuan Yuan, and Changhu Wang. Unsupervised real-world super-resolution: A domain adaptation perspective. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4298–4307, 2021.

[12] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. *arXiv preprint*, arXiv:1807.02758, 2018. Published in ECCV 2018.