FERMILAB-CONF-21-535-SCD

Using Mask R-CNN to detect and mask ghosting and scattered-light artifacts in astronomical images

Dimitrios Tanoglidis
dtanoglidis@uchicago.edu
Department of Astronomy and Astrophysics
Kavli Institute for Cosmological Physics
University of Chicago

Aleksandra Ćiprijanović
aleksand@fnal.gov
Fermi National Accelerator Laboratory

Alex Drlica-Wagner
kadrlica@fnal.gov
Fermi National Accelerator Laboratory
Kavli Institute for Cosmological Physics
Department of Astronomy and Astrophysics
University of Chicago

Brian Nord
nord@fnal.gov
Fermi National Accelerator Laboratory
Kavli Institute for Cosmological Physics
Department of Astronomy and Astrophysics
University of Chicago

Michael H. L. S. Wang
mwang@fnal.gov
Fermi National Accelerator Laboratory

Ariel Jacob Amsellem
ajamsellem@uchicago.edu
Kavli Institute for Cosmological Physics
University of Chicago

Kathryn Downey
kdowney@uchicago.edu
Department of Astronomy and Astrophysics
University of Chicago

Sydney Jenkins
sydneyjenkins@uchicago.edu
Department of Astronomy and Astrophysics
University of Chicago

Diana Kafkes
dkafkes@fnal.gov
Fermi National Accelerator Laboratory

Zhuoqi Zhang
zhuoqizhang@uchicago.edu
Department of Astronomy and Astrophysics
University of Chicago

Abstract

Wide-field astronomical surveys are often affected by the presence of undesirable reflections (often known as "ghosting artifacts" or "ghosts") and scattered-light artifacts. The identification and mitigation of these artifacts is important for rigorous astronomical analyses of faint and low-surface-brightness systems. In this work, we use images from the Dark Energy Survey (DES) to train, validate, and test a deep neural network (Mask R-CNN) to detect and localize ghosts and scattered-light artifacts. We find that the ability of the Mask R-CNN model to identify affected regions is superior to that of conventional algorithms that model the physical processes that lead to such artifacts, thus providing a powerful technique for the automated detection of ghosting and scattered-light artifacts in current and near-future surveys.

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

1 Introduction

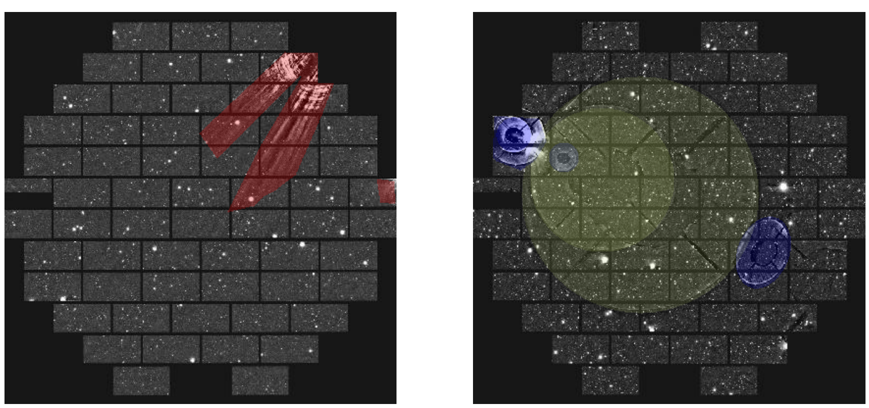
Wide-field imaging surveys at optical and near-infrared wavelengths, that map large parts of the sky, have provided a wealth of astronomical information that has enabled a better understanding of the processes that govern the growth and evolution of the Universe and its contents. Near-future surveys, such as the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST; [12]), will further expand our knowledge of the Universe by extending measurements to unprecedentedly faint astronomical systems. Such surveys will produce terabytes of data each night and measure tens of billions of stars and galaxies.

Images collected by galaxy surveys often contain artifacts caused by scattered and reflected light (commonly known as "ghosting artifacts" or "ghosts") from bright astronomical sources. The effective mitigation of these artifacts, and the spurious brightness variations they introduce, is important for the detection and precise measurement of faint astronomical systems, a major goal of current and upcoming surveys [3,9,14,17]. Modern telescopes introduce light baffles and anti-reflective coatings on key optical surfaces, that greatly reduce the occurrence of ghosts and scattered-light events.

However, a complete elimination of those artifacts is often impossible. Furthermore, given the large datasets produced by astronomical surveys, rejection by visual inspection is not feasible and automated methods should be developed. The Dark Energy Survey (DES; [6,7]), for example, uses a Ray-Tracing algorithm [16] that models the physical processes that lead to ghosting and scattered-light artifacts (configuration of the telescope, positions of known bright stars, etc.) to predict the presence and location of artifacts.

In this work we use an object detection and segmentation algorithm, namely Mask Region-Based Convolutional Neural Network (Mask R-CNN; [12]) to predict the location of ghosts and scattered-light artifacts in DES images. We use a set of manually annotated images and masks for training, validation, and testing. We show that this method performs better in locating affected areas in astronomical images compared to conventional methods, like the Ray-Tracing algorithm mentioned above. Our code and data are available at: `https://github.com/dtanoglidis/DeepGhostBusters`

2 Data



(a) (b)

Figure 1: Examples of full-focal-plane DECam images containing ghosts and scattered-light artifacts, and the corresponding ground-truth masks created through manual annotation. In panel (a) we have artifacts of type 'Rays' (in red), while in panel (b) we have artifacts of type 'Bright' (in blue) and 'Faint' (in yellow).

In this work we use images that come from the full six years of DES observations, obtained with the 570-megapixel Dark Energy Camera (DECam; [9]), mounted on the 4m Blanco Telescope in Chile. The DES data cover ~ 5000 deg^2 of the southern sky in five photometric filters, $grizY$, to a depth of $i \sim 24$ mag [1].

Specifically, we use 2000 images that cover the full DECam focal plane and are known to contain ghosts and scattered-light artifacts. These are part of the dataset used in Chang et al. [5] to train a standard CNN classifier to distinguish between images with and without ghosts, and is publicly available at `https://des.ncsa.illinois.edu/releases/other/paper-data`. We refer the interested reader to that paper for details on the data preprocessing. The focal plane of DECam consists of 62 charge-coupled devices (CCDs), as can be seen on Figure 1. During training each sample is the full focal plane image and not individual CCDs; however, as we discuss in §4, we evaluate the performance of our model on a per-CCD basis, i.e. how well Mask R-CNN is able to flag the affected CCDs.

Training the Mask R-CNN algorithm requires both images and ground-truth segmentation masks identifying objects of interest in each image. To create these masks, we used the VGG Image Annotator (VIA; https://www.robots.ox.ac.uk/ vgg/software/via/), a simple manual annotation software for images, audio and video. During manual annotation, we categorized the artifacts into three distinct morphological categories: 'Rays' (scattered-light artifacts originating from the light of off-axis stars scattering off of the DECam filter changer), 'Bright' (high-surface-brightness and relatively compact ghosts that come from multiple reflections off the DECam focal plane and the telescope lenses) and 'Faint' (lower-surface-brightness and more diffuse ghosting artifacts, that also originate from multiple reflections between the focal plane and the telescope lenses or filters, or internal reflections off of the faces of the lenses). We note, however, that the distinction between bright and faint ghosts is quite arbitrary, and does not come with a specific surface-brightness threshold.

In Figure 1 we present examples of DECam focal-plane images and ground-truth masks of artifacts: in panel (a) we can see examples of 'Rays' (in red the mask), while in panel (b) examples of 'Bright' (in blue) and 'Faint' (in yellow) artifacts.

3 Method

We train Mask R-CNN [12], a state-of-the art instance segmentation (a combination of object detection and image segmentation) algorithm, to detect and mask ghost and scattered-light artifacts. Previous applications of Mask R-CNN in astronomy have focused on deblending and classifying galaxies from imaging data [4,9]. In the first stage of the model, the input images are fed into a pre-trained deep CNN, also known as the *backbone* network, with its fully connected layers removed. This produces a feature map that is passed into the Region Proposal Network (RPN), to produce a limited number of candidate Regions of Interest (RoIs). However, each of the proposed RoIs can have a different size, and the RoIAlign method is used to perform a bilinear interpolation on the feature maps, to produce fixed-size feature maps of the candidate regions that pass to main part of the algorithm.

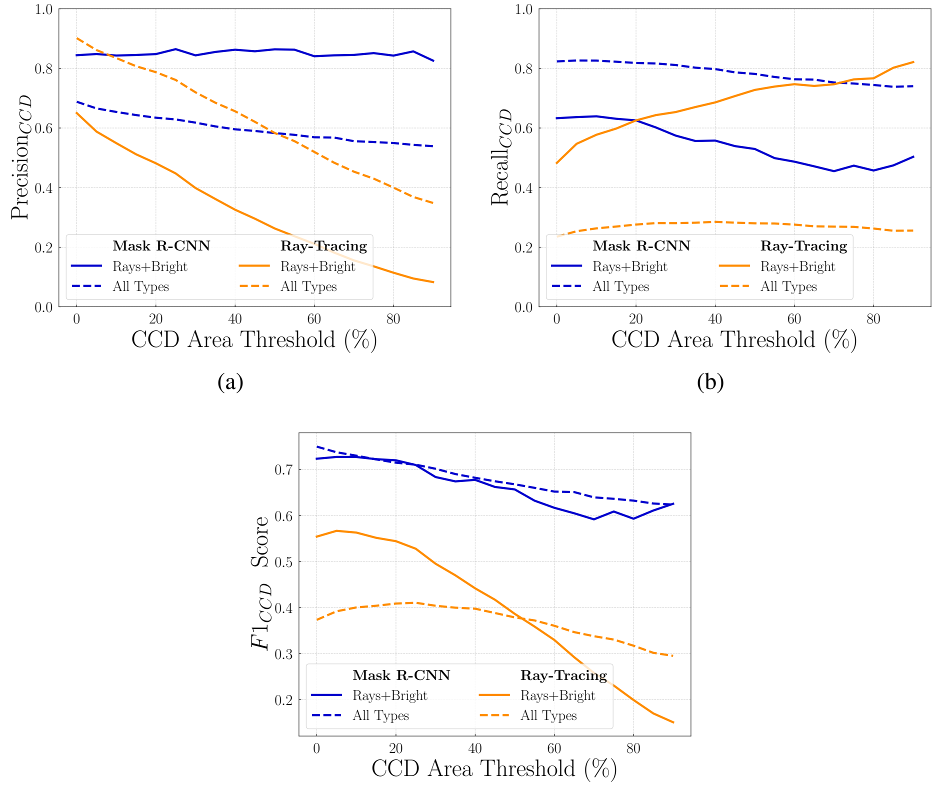
The main part of Mask R-CNN performs three tasks in parallel. A softmax classifier learns to predict the class of the object within the RoI (L_{cls} loss). A regressor learns the best bounding box coordinates (L_{bbox} loss). Finally, a Fully Convolutional Network (FCN) performs semantic segmentation (L_{mask} loss), i.e. a per-pixel classification, that creates the masks. The total loss is thus $L_{\text{tot}} = L_{\text{cls}} + L_{\text{bbox}} + L_{\text{mask}}$.

We use the Mask R-CNN implementation by Abullah [2], that is written in Python using the high-level `Keras` library on a `TensorFlow` backend. We use the default 101-layer deep residual network (ResNet-101; [13]) as the backbone CNN architecture.

Before training, we randomly split the full dataset of 2000 images (and the corresponding ground-truth masks) described in §2 into a training set (1400 images), a validation set (300 images), and a test set (300 images). We initialize the learning procedure with the weights learned on the Microsoft Common Objects in Context (MS COCO) dataset [17], which consists of $\sim$ 330k images of 91 classes of everyday objects.

The training is performed in different stages with progressively smaller learning rates (one order of magnitude less each time), from $\eta = 4 \times 10^{-3}$ for the first 15 epochs, down to $\eta = 4 \times 10^{-6}$ (20 epochs in each stage), for a total of 75 epochs. This allows for deeper training with less risk of overfitting. We utilized the 25 GB high-RAM Nvidia P100 GPUs available through the Google Colaboratory (Pro version). The training took ~ 4 hours to complete. The inference time is ~ 0.34s per image to predict.

4 Results



(a)

(b)

(c)

Figure 2: CCD-based (a) precision, (b) recall, and (c) $F1$ score of the Mask R-CNN model (blue lines) and the Ray-Tracing algorithm (orange lines). We consider both the combination of all types of artifacts (solid lines) and the combination of 'Rays'+'Bright' (dashed lines). The CCD area threshold is defined as the fraction of the CCD area that must be covered for it to be classified as affected.

The conventional Ray-Tracing algorithm used by DES flags affected focal plane images on a CCD-by-CCD basis — i.e., if a CCD contains a ghost or scattered-light artifact, the entire CCD is removed from processing. To compare the performance of the Mask R-CNN to the conventional algorithm, we use metrics that are based on whether a CCD contains a ghost or scattered-light artifact. Specifically, we consider each image as a 1D array of length 62 with entries 0 and 1, where 0 corresponds to CCDs that do not contain an artifact, and 1 corresponds to those that do contain one. Then, we define the CCD-based precision, recall, and $F1$ score (harmonic mean of the previous two) based on the number of CCDs in a batch of images (for example, in the test set presented here) that were correctly classified as containing an artifact.

These metrics depend on whether a CCD contains an artifact or not. In most cases, the mask of an artifact only partially covers a CCD. Thus, we define a threshold for the fraction of the CCD area that must be covered for the CCD to be classified as affected. In Figure 2 we present the CCD-based (a) precision, (b) recall, and (c) $F1$ score of the Mask R-CNN and Ray-Tracing algorithms, as a function of that area threshold (the CCDs masked as "bad" by the Ray-Tracing were obtained from `https://des-ops.fnal.gov:8082/exclude/`).

We consider two cases: when all types of artifacts are combined (solid lines) and when only the combination 'Rays'+'Bright' is considered (dashed lines). We do that for a fair comparison, since Ray-Tracing is not optimized to detect very faint ghosts. In Table 1 we present the values of these metrics for the two models, at a 0% area threshold (single pixel of an artifact within a CCD needs to be present, for the CCD to be counted as "bad"), which is the most conservative approach for "bad" CCD rejection.

Table 1: CCD-based evaluation metrics (precision, recall, $F1$ score) for the Mask R-CNN and Ray-Tracing algorithms, at 0% CCD area threshold.

	Mask R-CNN		Ray-Tracing	
	Rays+Bright	Rays+Bright+Faint	Rays+Bright	Rays+Bright+Faint
Precision	84.3%	68.7%	64.7%	89.9%
Recall	63.6%	82.5%	48.4%	23.5%
$F1$ score	72.5 %	75.0%	55.4%	37.3%

As indicated by the $F1$ score plot in Figure 2, which is a combination of the precision and recall, Mask R-CNN performs better for both artifact type combinations and across all area threshold levels. From the Table 1 values (at the single pixel area threshold), we see that for the combination 'Rays+Bright' R-CNN performs significantly better than the Ray-Tracing algorithm as indicated by all three metrics. For the combination 'Rays+Bright+Faint' Ray-Tracing achieves higher precision, but significantly lower recall compared to the Mask R-CNN model; recall can be seen as more important metric if the target is to detect the highest possible number of affected CCDs — in other words to minimize the remaining artifacts in the data.

As we can see on Figure 2, recall tends to decrease with the area threshold for the Mask R-CNN model.This is mainly driven by the 'Bright' ghosts that are relatively small, only partially covering the CCDs that contain them. As we increase the area threshold, only a few such ghosts can pass it. Generally, precision decreases, while recall increases as a function of the CCD area threshold for both artifact combinations. As we increase the threshold, fewer CCDs are labeled as containing artifacts and thus the purity decreases while the completeness increases.

For practical applications of the Mask R-CNN model for astronomical images analyses, the applied threshold can be set based on whether we favor a higher purity (precision) or completeness (recall), which in turn depends on the specific science case.

5 Conclusion

In this work, we applied a state-of-the art object detection and segmentation algorithm, Mask R-CNN, to the problem of finding and masking ghosts and scattered-light artifacts in astronomical images from DECam. The effective detection and mitigation of these artifacts is especially important for low-surface-brightness science, an important target of future surveys.

We compare the ability of Mask R-CNN in masking affected CCDs in ghost-containing images with that of the Ray-Tracing algorithm. We find that the Mask R-CNN model has superior performance, as measured by the $F1$ score. These results hold across different CCD area thresholds and for the two combinations of the morphological classes discussed in this work — 'Bright'+'Rays' and 'Bright'+ 'Rays'+'Faint'.

The results presented here highlight the promise of object detection and segmentation methods in tackling the identification of ghosts and scattered-light artifacts. Such automated techniques can facilitate the efficient separation of artifacts from scientifically useful data in upcoming surveys like the LSST.

Broader Impact

In this work we used a popular object detection/image segmentation algorithm in an astrophysical context. Object detection algorithms are known to have significant societal impact, especially when they are applied to humans (face recognition). The application of such algorithms to the physical sciences is limited so far. By using them in science and better understanding cases where they fail (produce false positives/negatives) we can alleviate some of the biases that are inherent to the training of these algorithms.

Acknowledgments and Disclosure of Funding

We would like to thank Colin Burke, Chihway Chang, Tom Diehl, Brenna Flaugher, and Steve Kent for useful discussions and suggestions.

A. Ćiprijanović is partially supported by the High Velocity Artificial Intelligence grant as part of the Department of Energy High Energy Physics Computational HEP sessions program.

We acknowledge the Deep Skies Lab as a community of multi-domain experts and collaborators who've facilitated an environment of open discussion, idea-generation, and collaboration. This community was important for the development of this project.

This material is based upon work supported by the National Science Foundation under Grant No. AST-2006340. This work was supported by the University of Chicago and the Department of Energy under section H.44 of Department of Energy Contract No. DE-AC02-07CH11359 awarded to Fermi Research Alliance, LLC. This work was partially funded by Fermilab LDRD 2018-052.

This project used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

References

[1] Abbott, T. M. C., “The Dark Energy Survey Data Release 2”, *The Astrophysical Journal Supplement Series*, vol. 255, no. 2, 2021. doi:10.3847/1538-4365/ac00b3.

[2] Abdulla, W., 2017. "Mask r-cnn for object detection and instance segmentation on keras and tensorflow". $https://github.com/matterport/Mask_RCNN$.

[3] Brough, S., et al., “The Vera Rubin Observatory Legacy Survey of Space and Time and the Low Surface Brightness Universe”, *arXiv e-prints*, 2020.

[4] Burke, C. J., “Deblending and classifying astronomical sources with Mask R-CNN deep learning”, *Monthly Notices of the Royal Astronomical Society*, vol. 490, no. 3, pp. 3952–3965, 2019. doi:10.1093/mnras/stz2845.

[5] Chang, C., Drlica-Wagner, A., Kent, S. M., Nord, B., Wang, D. M., and Wang, M. H. L. S., “A machine learning approach to the detection of ghosting and scattered light artifacts in dark energy survey images”, *Astronomy and Computing*, vol. 36, 2021. doi:10.1016/j.ascom.2021.100474.

[6] Dark Energy Survey Collaboration, “The Dark Energy Survey”, *arXiv e-prints*, 2005.

[7] Dark Energy Survey Collaboration, “The Dark Energy Survey: more than dark energy - an overview”, *Monthly Notices of the Royal Astronomical Society*, vol. 460, no. 2, pp. 1270–1299, 2016. doi:10.1093/mnras/stw641.

[8] Dutta, A., Zisserman, A., 2019. The VIA annotation soft1179 ware for images, audio and video, in: Proceedings of the 27th ACM International Conference on Multimedia, ACM, New York, NY, USA. URL: $https://doi.org/10.1145/3343031.3350535, doi: 10.1145/3343031.3350535$

[9] Farias, H., Ortiz, D., Damke, G., Jaque Arancibia, M., and Solar, M., “Mask galaxy: Machine learning pipeline for morphological segmentation of galaxies”, *Astrophysics Source Code Library*, 2021. ascl:2101.007.

[10] Flaugher, B., “The Dark Energy Camera”, *The Astronomical Journal*, vol. 150, no. 5, 2015. doi:10.1088/0004-6256/150/5/150.

[11] Greco, J. P., “Illuminating Low Surface Brightness Galaxies with the Hyper Suprime-Cam Survey”, *The Astrophysical Journal*, vol. 857, no. 2, 2018. doi:10.3847/1538-4357/aab842.

[12] He, K., Gkioxari, G., Dollár, P., and Girshick, R., “Mask R-CNN”, *arXiv e-prints*, 2017.

[13] He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition”, *arXiv e-prints*, 2015.

[14] Ivezić, Ž., et al., “LSST: From Science Drivers to Reference Design and Anticipated Data Products”, *The Astrophysical Journal*, vol. 873, no. 2, 2019. doi:10.3847/1538-4357/ab042c.

[15] Kaviraj, S., “The low-surface-brightness Universe: a new frontier in the study of galaxy evolution”, *arXiv e-prints*, 2020.

[16] Kent, Stephen M. Ghost Images in DECam. United States: N. p., 1900. Web. doi:10.2172/1690257.

[17] Lin, T.-Y., et al., “Microsoft COCO: Common Objects in Context”, *arXiv e-prints*, 2014.

[18] Tanoglidis, D., Drlica-Wagner, A., Wei, K., et al., “Shadows in the Dark: Low-surface-brightness Galaxies Discovered in the Dark Energy Survey”, *The Astrophysical Journal Supplement Series*, vol. 252, no. 2, 2021. doi:10.3847/1538-4365/abca89.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Throughout the paper the assumptions made are being mentioned.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Broader Impact statement
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]

(b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Yes, in a github repo, see Sec. 1
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] see Sec. 3
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Sec. 3
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Sections 2, 3
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] In anonymous (for now) github repo, see Sec. 1
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]