Two-Stage GAN for Sparse Galaxy Image Generation

Department of Computer Science ETH Zurich, Switzerland
Group: Gantastic 4

Siméone de Fremond Carlos Gomes Andra-Maria Ilieş Andrej Ivanov

***Abstract*—The rapid growth of astronomical data in recent years has prompted interest in building generative models of cosmological images. This paper focuses on 2 related tasks in this area: 1) generating realistic images and 2) scoring how realistic an image is. We begin by formalizing the structure of the images in terms of galaxies and their properties, and show that simplistic baselines are not sufficient for both tasks. We then present a novel generative approach consisting of a two-stage sampling generative adversarial neural network (GAN), capturing the position of cosmological objects and their representation separately, which produces diverse and realistic looking images. Furthermore, we develop an accurate scoring method based on Power Spectrum features. Using this scoring method we empirically show that our two-stage generative model significantly outperforms both the baselines and a standard GAN approach. Finally, we provide experiments towards understanding of the most important factors in our model.**

I. Introduction

With the rapidly growing quantity of astronomical images [1] [2] [3], there has been an increasing interest in using machine learning for tasks such as galaxy processing, classification, segmentation, deblending and building generative models of cosmological images and objects. In this work, we are given a dataset of cosmological images with which we tackle two related tasks: 1) **generation of realistic images**, 2) **scoring how realistic an image is**.

Having an accurate but very complex generative model of galaxy images can help astronomers derive more interpretable factors describing the structure of the Cosmos [4]. Furthermore, synthetic data is necessary for training models to perform a wide range of processing tasks on these images [5] [6]. As for the second task, a good scoring model can be used to automatically detect noisy and corrupted images, and to quantitatively evaluate generative models.

This generative task is particularly difficult, because the factors controlling the formation and properties of stars and galaxies are complex and currently not well understood. A complete generative model would be required to capture aspects ranging from large-scale galaxy formation priors to individual galaxy properties and optical artefacts of the imaging system [7].

The current approaches in the related literature focus on modelling individual galaxies and stars, or modelling images densely populated with cosmological objects. Certain works attempt to model the images using hard-coded domain knowledge [6] [8]. Alternatively, approaches employing deep learning generative models, such as GANs [9] [10] [11] and VAEs [12] have been gaining popularity in research.

Similarly to these works, we attempt to build a completely machine-learning-based generative model, without relying on expert knowledge in astronomy. However, the nature of our data is slightly different than that of previous papers. To the best of our knowledge, this is the first work that models sparse cosmological images, where the cosmological objects are rare and with a very small resolution. Our main contributions are:

- An analysis and breakdown of the most important factors in the images.
- Baseline models: intuitive and decently performing baselines for both tasks.
- Generative model: a two stage sampling model where we break down the task into generating the location of galaxies in the image, followed by generating individual galaxies. We leverage a separate GAN for each of these tasks and produce realistic images, both qualitatively and quantitatively.
- Scoring model: a Fourier transform based model that scores how realistic images are with high accuracy.

II. Models and methods

A. Modelling of the problem

On a high level, the images in our dataset are of a high resolution (1000x1000 pixels), but most of the image represents a constant dark background populated with a number of very distant, and hence tiny, celestial objects.

Background: in all real images, the background is of an almost fixed intensity of less than 3. In 80% of the images, the background intensity is strictly 0. In the other 20%, the background is of intensity of 1 or 2. Our hypothesis is that the second mode of the data is an optical artifact. For the rest of this work, we will focus on modelling images with background of intensity 0, and we assume the other mode of the data can be modelled analogously.

Objects: we consider any non-black cluster of pixels a galaxy. Visually, a galaxy is defined by 4 main properties: **location** in the image, **size**, **intensity** and **shape**.

B. Generative models

1) Baselines: We built 2 intuitive baselines used as a benchmark for the difficulty of the problem. Both of them make the simplifying assumption that the galaxies' properties

$$p(image) = p(l_1)p(l_2|l_1)p(l_3|l_2, l_1)... \prod_g p(g) \quad (1)$$

Fig. 1: Equation representing the unconditional sampling model, where l stands for galaxy location and g for galaxy

are mutually independent. Furthermore, both baselines model the position of a galaxy simply as a uniform distribution over the pixels, and given the intensity and the size, the shape of the galaxy is a fixed diamond-like shape. The first baseline is a naive approach where the size and intensity of a galaxy are simply sampled from a normal distribution. The second baseline is informed by the actual statistics of the data, and introduces 2 improvements. Firstly, the size and intensity of galaxies are sampled from a geometric distribution, to match the insight that most galaxies are small and faint. Secondly, it models large and small galaxies separately, in an attempt to indirectly and partially capture the mutual interaction of sizes and intensities between galaxies.

2) End-to-end GAN: As a first attempt of a fully data driven solution, we attempted to model the galaxy images using a GAN. GANs are unsupervised implicit generative models that have been shown to be able to model very complex distributions and produce realistic images in numerous domains. [13].

In theory, this method should be able to model all the dependencies between the features of the galaxies present in the image.

3) Two-stage sampling GAN: Since a traditional GAN seemed to struggle to capture full galaxy images, we moved to a two stage approach: a first stage where we sample the positions of galaxies, followed by a second part where we sample a galaxy at each location.

Our motivation for this approach is the fact that there are very strong priors on the structure of the images, and with an end-to-end GAN we're unnecessarily forcing the network to learn these from scratch. Firstly, we know that most of the image is a completely black background. Thus, it is unnecessary to model the entire image, when most of it carries no information. Secondly, we know that the objects in the image follow the same distribution, a trait not present in traditional image generation. Furthermore, we know from prior works [10] that GANs have been able to successfully model individual galaxy images at smaller scales (32x32 to 64x64).

Unconditional GAN: There are two stages to the process, illustrated in Fig. 2:

1) Position sampling: For each image in our dataset, we divide it into a 32x32 grid. Cells in this grid with galaxies are marked with a 1, and those without are marked with a 0. The first GAN, which we will refer to as the **position GAN**, learns these as 32x32 images, and is able to output such a grid representing the locations of galaxies in the image.
2) Galaxy sampling: we use a second GAN, which we will call the **galaxy GAN**, to generate images of individual galaxies. The training data is made up of extracted 32x32 galaxy patches from our original dataset. Specifically, we found success with the BEGAN [14] architecture.

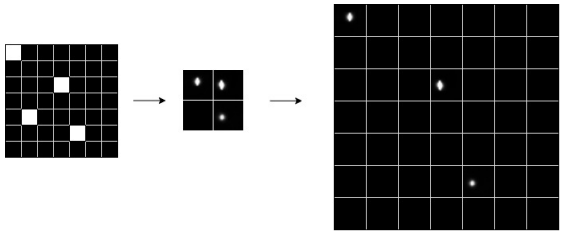

Fig. 2: The position GAN generates a grid, and the galaxy GAN produces the necessary galaxies. A generated galaxy is placed wherever a cell is 1, and 0s wherever it is 0.

$$p(image) = p(s_1, l_1)p(s_2, l_2|s_1, l_1)... \prod_g p(g|s_g) \quad (2)$$

Fig. 3: Equation representing the conditional sampling model, where l stands for galaxy location, s for galaxy size class and g for galaxy

There are 2 assumptions present in this model, and captured through Fig.1. The first one is that a galaxy's features are independent from its location. Thus, galaxies are placed at locations in the order they are sampled. The second assumption is that there is no mutual dependence between each galaxy present in an image, other than their location. While this assumption is more clearly flawed (an image with only very large and bright galaxies would not look very realistic), it allows us to greatly simplify the training and generation process.

Intuitively, the position GAN captures global features of the image. More concretely, it models the joint distribution of galaxy positions. The galaxy GAN captures the galaxy property (shape, size and intensity) priors. A parallel can be drawn between this method and progressive GANs [15], where the layers in the network are gradually added throughout the training process. This shares the same intuition, learning coarser image features first and gradually incorporating finer details. We take this and make it completely explicit with our division into 2 separate networks.

Conditional GAN: In our final approach, we also model the dependencies between the properties of different galaxies, in addition to their location, as described by Fig. 3. Furthermore, this method provides us with explicit control over the properties we identified in the introduction.

We do this by transforming the galaxy GAN into a conditional GAN [16]. We use the size of the galaxy as the conditioning property because we observe that it is correlated with both the intensity and the shape. We heuristically define size thresholds that allow us to assign a class to each galaxy in the dataset.

The position GAN now produces a class label for each position, which defines the type of galaxy present there. The galaxy CGAN then takes this label as an additional input.

This method allows for a good amount of control in the images produced, enabling the position GAN to capture the aforementioned dependencies between properties of galaxies, and the galaxy CGAN to use this information, while still producing diverse galaxies within the different classes.

C. Scoring models

This section presents various classes of features considered for the scoring task. For each set of features (except the CNN), we trained a random forest and a gradient boosting model.

1) Baseline 1 - Simple galaxy features: In this model, we manually engineer features based on basic galaxy properties. More specifically, the features represent common statistics of the size, intensity and coordinates of the galaxies in the image. Additionally, we attempt to capture interactions between the properties, with features such as the the number of galaxies which are both ”large” and ”bright”.

2) Baseline 2 - CNN: This second baseline was developed in two phases. We firstly tried to apply a small CNN with only 3 convolutional layers and maxpooling, where images are resized to 150x150 pixels. Later, we drastically increased the size of the network, going for a ResNet50 [17] architecture and making use of transfer learning with a model trained on ImageNet.

3) Histogram: The histogram is a 256 feature vector representing distribution of pixel intensities in the image. The intuition behind this choice of features is to capture the high contrast between background and galaxies in real images.

4) HOG and LBP: The histogram of gradients [18] and local binary patterns [19] are features that complement each other, which is why they are frequently used together [20]. The main advantage of HOG and LBP, compared to a histogram, is that the attribute extraction is performed on pixels' neighbourhood, providing a local characteristic.

5) Frequency Spectrum: The Fourier transform decomposes the image into its sine and cosine components [21], converting the image to frequency domain. Applying it on the cosmological images results in a 1000x1000 representation, each point describing the contribution the wave. We observed that the 250th row contains a diverse mix of both low frequency and high frequency data, so we used as features its second half, due to its symmetry property.

6) Power Spectrum: The power spectrum [22] can be derived from the Fourier representation of the image. We used the azymuthally summed power spectrum representation of the image, as a method to reduce the dimensionality of the data to 1D [23]. This is calculated by adding the values of the points that are at the same distance from the center, thus providing a radial profile. To reduce the amplitude of the discontinuities in this dimension, we apply a windowing function to smooth it.

III. Results

A. Generative models

The dataset used to train the end-to-end GAN, as well the two-stage models, was created by joining the images labeled as real with images scored above 3. All models were trained on a Tesla K80 GPU.

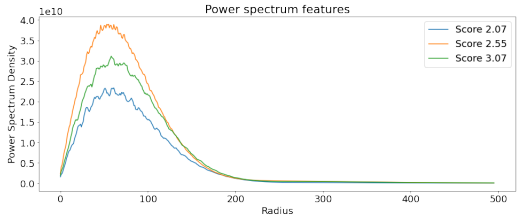

Fig. 4: Comparison of features for images with different scores

TABLE I: Scoring models' performance. Thicker lines separate the different approaches. Top to bottom: baseline 1 (simple galaxy features), baseline 2 (CNN), Histrogram + HOG + LBP, Fourier-based.

Model	Features	Score (MAE)
XGBoost	3 basic features only	0.54360
XGBoost	all individual properties	0.34519
XGBoost	individual + interaction properties	0.34491
CNN (resnet50)	-	0.31
CNN (basic)	-	1.07
RF	Histogram	0.30469
XGBoost	Histogram	0.26923
RF	Histogram + HOG + LBP	0.28677
RF	Fourier Transform	0.32029
RF	Power Spectrum 1D	0.12954
XGBoost	Histogram + Power Spectrum 1D	**0.09396**

Fig. 6 in the appendix shows samples from each of our methods compared to real images from the dataset. Images from the GAN that was trained on full images are clearly mode collapsing, with grid-like patterns.

For the rest of the methods, it is quite difficult to visually distinguish the generated images from real ones, or even determine how realistic they look. For a better comparison, Fig. 5 shows only the small generated patches of cosmological objects for each of these methods.

The results for both methods are very good. We see that the

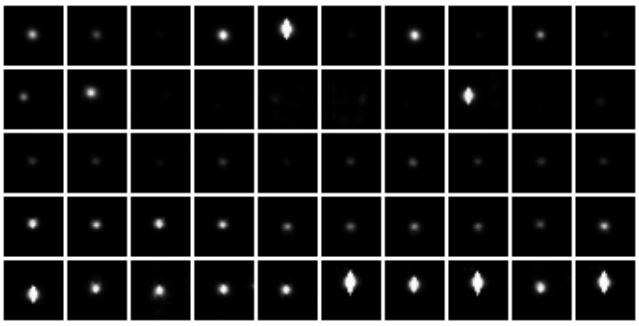

Fig. 5: Galaxies generated by each of the methods. Row 1: Real images; Row 2: Unconditional GAN; Row 3: Conditional GAN, Class small; Row 4: Conditional GAN, Class medium; Row 5: Conditional GAN, Class large. The galaxies are sampled randomly from the latent space.

unconditional galaxy GAN is able to capture the diversity in galaxies of the real images, while still producing high quality samples.

TABLE II: Generative models' performance

Model	Mean Score	Train time	Execution Time
Real Images	1.93	-	-
Baseline 1	0.60	-	6s
Baseline 2	0.89	-	13s
End-to-end GAN	0.85	40m	18s
Unconditional GAN	1.85	40m	10s
Conditional GAN	**1.90**	40m	6s

As for the galaxy CGAN, we can observe that it has learned separate galaxy classes, as was our goal, producing small, medium, and large galaxies, showing diversity in terms of shape and intensity within each class.

For a more holistic evaluation, we turn to Table II, where 1000 images were generated per model and evaluated by the best preforming scorer. While the baseline and end-to-end GAN achieve very poor scores, confirming our qualitative evaluation, both two-stage sampling methods achieve similarly high scores. Somewhat surprisingly, we find only a very marginal improvement in switching from the unconditional to the conditional approach, a result which we discuss later. We consider these quite satisfactory when compared to the scores obtained by images from the labeled dataset, which serves as an intuitive benchmark.

B. Scoring

The regressors are trained on a dataset of 9600 1000x1000 pixels scored images, with scores from 0 to 8, where higher represents more realistic images. They are then evaluated on a different set of images, and we use mean absolute error as our evaluation metric. As we can see from Table II, manually extracted features coupled with gradient boosting perform, by far, the best; more notably the power spectrum features.

C. Experiments

The following experiments study the influential factors on the score of an image for a two-stage model.

1) Random positions instead of position GAN: To evaluate the necessity of the position GAN, we compare its performance with a model which assigns positions randomly. Both methods result in **similar performance** in terms of average score, even for CGAN. This could either mean that the position GAN did not manage to learn the position distribution properly, or that the position distribution is fundamentally irrelevant to this problem.

2) Fixed positions, new galaxies: In this analysis, we evaluate how the choice of galaxies influences the score. More specifically, for a fixed position matrix we consider 50 different sets of galaxies to fill the positions. Repeating this experiment for 1000 base position matrices, we find that the distribution of scores varies significantly as we vary the sets of galaxies. This result suggests that the **interactions of the properties between galaxies have a high impact on the score**, further motivating the need for a conditional model.

3) Fixed positions, shuffle patches: In this experiment, for a fixed position matrix and a fixed set of galaxy patches, we permute the galaxy-location pairings. The mean standard deviation of the score for 1000 images using 50 permutations is 0.09. The result shows that these permutations have little effect on the score, leading to the conclusion that the **properties of a galaxy are independent of its location**.

IV. Discussion

Generative models: The generative baselines produce images with very low scores, which indicates that naively modelling the galaxy property priors using generic probability distributions is not sufficient to accurate model real images.

As for the end-to-end GAN, in addition to the weaknesses that motivated our two-stage sampling approach, another possible factor for its poor performance could be the large resolution of the images. This resulted in the need to downscale, potentially occluding many galaxies.

Consequently, allowing the two-stage models to work with full resolution galaxy patches, while disregarding the visually unimportant low intensity portions of the image, contributes to the detailed results they are able to produce.

Experiment 2 supports the need for our conditional model, which can theoretically model all the dependencies between galaxies. However, in practice, we only managed to get a small improvement. This could indicate that our conditioning variable should be improved. Indeed, the slight overlap between adjacent classes shown in Fig. 5 further points to possible improvements for future work, using data-driven rather than manually engineered classes.

Scoring models: The hand-crafted features of baseline 1 are good at scoring fake images, but fail to accurately find the score of real images. Attempting to engineer features for galaxy interactions only marginally improves the score, which demonstrates the need for more involved methods to capture these nuances in real images. Power spectrum features are able to automatically capture frequency-domain characteristics of realistic cosmological images, which we empirically verify as suitable for this task.

V. Conclusion

In this work, we develop and compare different methods for cosmological image generation and scoring. We begin by formalizing the structure of the images in terms of galaxies and their properties, and show that simplistic baselines are not sufficient for both tasks. Our main novelty, a two-stage GAN model, produces outputs indistinguishable from real images. We also develop an accurate image scoring method, based on image power spectrum features, which we use to objectively establish the realism of our generated images. Finally, we present experiments that justify our methodology.

REFERENCES

[1] J. Allison, S. Anderson, J. Andrew, J. Angel, L. Armus, D. Arnett, S. Asztalos, T. Axelrod *et al.*, "Lsst science book, version 2.0," *arXiv preprint arXiv:0912.0201*, 2009.

[2] T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, J. Asorey, S. Avila, O. Ballester, M. Banerji, W. Barkhouse, and et al., "The dark energy survey: Data release 1," *The Astrophysical Journal Supplement Series*, vol. 239, no. 2, p. 18, Nov 2018. [Online]. Available: http://dx.doi.org/10.3847/1538-4365/aae9f0

[3] J. Green, P. Schechter, C. Baltay, R. Bean, D. Bennett, R. Brown, C. Conselice, M. Donahue, X. Fan, B. S. Gaudi, C. Hirata, J. Kalirai, T. Lauer, B. Nichol, N. Padmanabhan, S. Perlmutter, B. Rauscher, J. Rhodes, T. Roellig, D. Stern, T. Sumi, A. Tanner, Y. Wang, D. Weinberg, E. Wright, N. Gehrels, R. Sambruna, W. Traub, J. Anderson, K. Cook, P. Garnavich, L. Hillenbrand, Z. Ivezic, E. Kerins, J. Lunine, P. McDonald, M. Penny, M. Phillips, G. Rieke, A. Riess, R. van der Marel, R. K. Barry, E. Cheng, D. Content, R. Cutri, R. Goullioud, K. Grady, G. Helou, C. Jackson, J. Kruk, M. Melton, C. Peddie, N. Rioux, and M. Seiffert, "Wide-field infrared survey telescope (wfirst) final report," 2012.

[4] S. Ravanbakhsh, F. Lanusse, R. Mandelbaum, J. G. Schneider, and B. Póczos, "Enabling dark energy science with deep generative models of galaxy images," in *AAAI*, 2017.

[5] R. B. Metcalf, M. Meneghetti, C. Avestruz, F. Bellagamba, C. R. Bom, E. Bertin, R. Cabanac, F. Courbin, A. Davies, E. Decencière, R. Flamary, R. Gavazzi, M. Geiger, P. Hartley, M. Huertas-Company, N. Jackson, E. Jullo, J.-P. Kneib, L. V. Koopmans, F. Lanusse, C.-L. Li, Q. Ma, M. Makler, N. Li, M. Lightman, C. E. Petrillo, S. Serjeant, C. Schafer, A. Sonnenfeld, A. Tagore, C. Tortora, D. Tuccillo, M. B. Valentin, S. Velasco-Forero, G. A. V. Kleijn, and G. Vernardos, "The strong gravitational lens finding challenge," *Astronomy and Astrophysics*, vol. 625, 2019.

[6] E. Bertin and S. Arnouts, "Sextractor: Software for source extraction," *Astronomy and Astrophysics Supplement Series*, vol. 117, no. 2, pp. 393–404, 1996.

[7] D. W. Hogg, "Theories of everything." [Online]. Available: http://videolectures.net/nipsworkshops2011_hogg_theories/

[8] L. Miller, T. D. Kitching, C. Heymans, A. F. Heavens, and L. Van Waerbeke, "Bayesian galaxy shape measurement for weak lensing surveys – I. Methodology and a fast-fitting algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 382, no. 1, pp. 315–324, 10 2007. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2007.12363.x

[9] M. Dia, E. Savary, M. Melchior, and F. Courbin, "Galaxy image simulation using progressive gans," *arXiv preprint arXiv:1909.12160*, 2019.

[10] L. Fussell and B. Moews, "Forging new worlds: high-resolution synthetic galaxies with chained generative adversarial networks," *ArXiv*, vol. abs/1811.03081, 2018.

[11] M. J. Smith and J. E. Geach, "Generative deep fields: arbitrarily sized, random synthetic astronomical images through deep learning," *arXiv: Instrumentation and Methods for Astrophysics*, 2019.

[12] J. Regier, A. C. Miller, J. McAuliffe, R. P. Adams, M. D. Hoffman, D. Lang, D. Schlegel, and Prabhat, "Celeste: Variational inference for a generative model of astronomical images," in *ICML*, 2015.

[13] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," 2014.

[14] D. Berthelot, T. Schumm, and L. Metz, "Began: Boundary equilibrium generative adversarial networks," 2017.

[15] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *CoRR*, vol. abs/1710.10196, 2017. [Online]. Available: http://arxiv.org/abs/1710.10196

[16] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," vol. 1, pp. 886–893 vol. 1, 2005.

[19] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," vol. 1, pp. 582–585 vol.1, 1994.

[20] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," pp. 32–39, 2009.

[21] Fourier transform. [Online]. Available: https://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm

[22] 2d spectrum characterization. [Online]. Available: https://www.cygres.com/OcnPageE/Glosry/SpecE.html

[23] 2d spectrum characterization. [Online]. Available: https://medium.com/tangibit-studios/2d-spectrum-characterization-e288f255cc59

APPENDIX

A. Final Galaxy Images

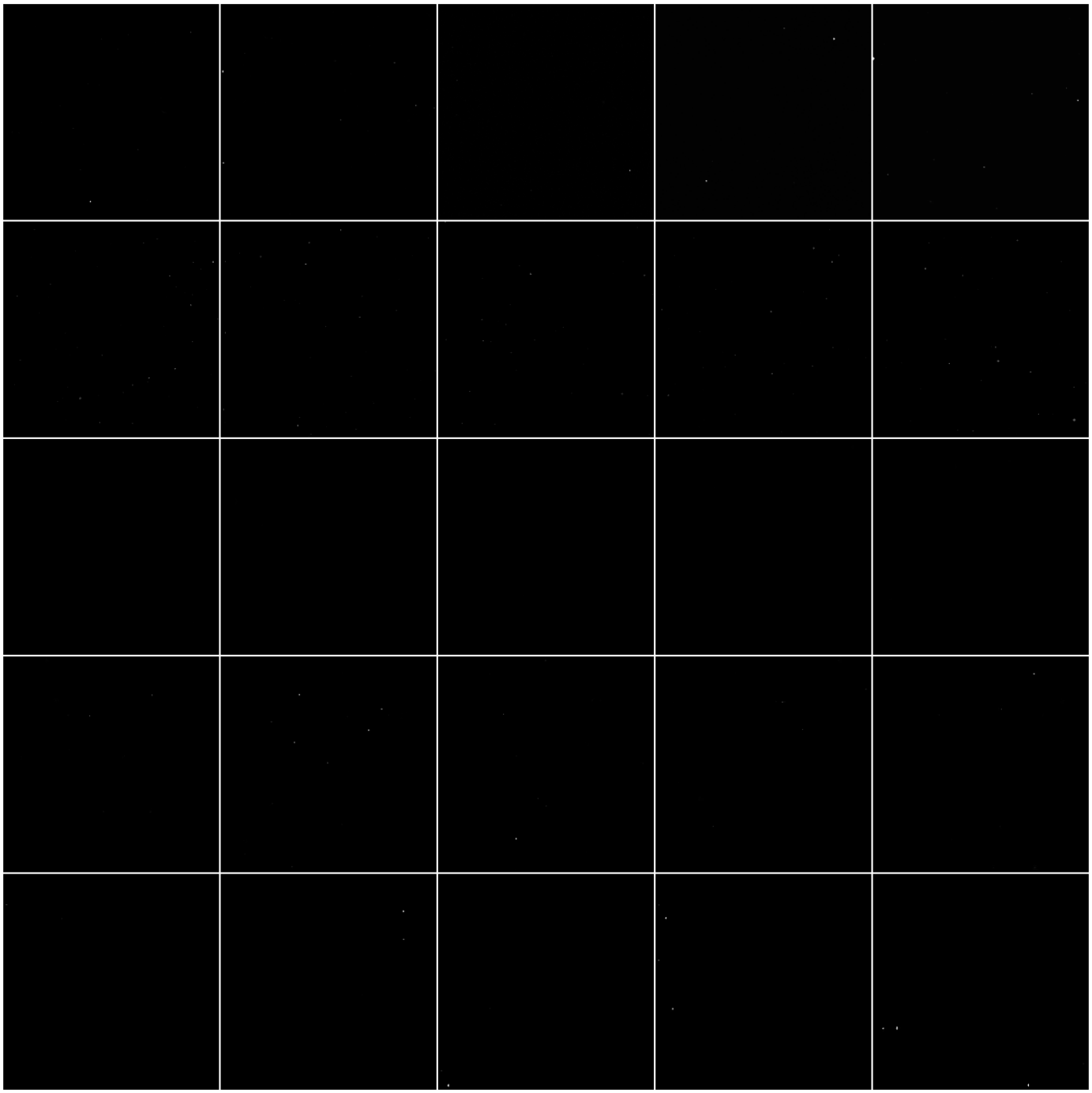

Fig. 6: Images generated by different methods compared with real images. Row 1: Real images; Row 2: Baseline model; Row 3: End-to-end GAN; Row 4: Unconditional GAN; Row 5: Conditional GAN.