frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 04 June 2021
doi: 10.3389/frai.2021.673062

Emulation of Cosmological Mass Maps with Conditional Generative Adversarial Networks

Nathanaël Perraudin[1], *Sandro Marcon*[2], *Aurelien Lucchi*[2] *and Tomasz Kacprzak*[3]*

[1]Swiss Data Science Center, ETH Zurich, Zurich, Switzerland, [2]Institute for Machine Learning, ETH Zurich, Zurich, Switzerland, [3]Institute for Particle Physics and Astrophysics, ETH Zurich, Zurich, Switzerland

Weak gravitational lensing mass maps play a crucial role in understanding the evolution of structures in the Universe and our ability to constrain cosmological models. The prediction of these mass maps is based on expensive N-body simulations, which can create a computational bottleneck for cosmological analyses. Simulation-based emulators of map summary statistics, such as the matter power spectrum and its covariance, are starting to play increasingly important role, as the analytical predictions are expected to reach their precision limits for upcoming experiments. Creating an emulator of the cosmological mass maps themselves, rather than their summary statistics, is a more challenging task. Modern deep generative models, such as Generative Adversarial Networks (GAN), have demonstrated their potential to achieve this goal. Most existing GAN approaches produce simulations for a fixed value of the cosmological parameters, which limits their practical applicability. We propose a novel conditional GAN model that is able to generate mass maps for any pair of matter density Ω_m and matter clustering strength σ_8, parameters which have the largest impact on the evolution of structures in the Universe, for a given source galaxy redshift distribution $n(z)$. Our results show that our conditional GAN can interpolate efficiently within the space of simulated cosmologies, and generate maps anywhere inside this space with good visual quality high statistical accuracy. We perform an extensive quantitative comparison of the N-body and GAN -generated maps using a range of metrics: the pixel histograms, peak counts, power spectra, bispectra, Minkowski functionals, correlation matrices of the power spectra, the Multi-Scale Structural Similarity Index (MS-SSIM) and our equivalent of the Fréchet Inception Distance. We find a very good agreement on these metrics, with typical differences are <5% at the center of the simulation grid, and slightly worse for cosmologies at the grid edges. The agreement for the bispectrum is slightly worse, on the <20% level. This contribution is a step toward building emulators of mass maps directly, capturing both the cosmological signal and its variability. We make the code[1] and the data[2] publicly available.

Keywords: generative models, cosmological simulations, cosmological emulators, N-body simulations, mass maps, generative adversarial network, fast cosmic web simulations, conditional GAN

[1]https://renkulab.io/gitlab/nathanael.perraudin/darkmattergan
[2]https://zenodo.org/record/4646764

OPEN ACCESS

Edited by:
Maria Han Veiga,
University of Michigan, United States

Reviewed by:
Bowei Chen,
University of Glasgow,
United Kingdom
Riccardo Zese,
University of Ferrara, Italy

****Correspondence:***
Tomasz Kacprzak
tomaszk@phys.ethz.ch

Specialty section:
This article was submitted to Machine Learning and Artificial Intelligence, a section of the journal Frontiers in Artificial Intelligence

Received: *26 February 2021*
Accepted: *06 May 2021*
Published: *04 June 2021*

Citation:
Perraudin N, Marcon S, Lucchi A and Kacprzak T (2021) Emulation of Cosmological Mass Maps with Conditional Generative Adversarial Networks. Front. Artif. Intell. 4:673062. doi: 10.3389/frai.2021.673062

1 INTRODUCTION

The N-body technique simulates the evolution of the Universe from soon after the big bang, where the mass distribution was approximately a Gaussian random field, to today, where, under the action of gravity, it becomes highly non-Gaussian. The result of an N-body simulation consists of a 3D volume where the positions of particles represent the density of matter in specific regions. This 3-dimensional representation can then be projected in 2 dimensions by integrating the mass along the line of sight with a lensing kernel. The resulting images are called *sky convergence maps*, often referred to simply as the *cosmological mass maps*. These maps can be compared with real observations with the purpose of estimating the cosmological parameters and testing cosmological models. Their simulation, however, is a very challenging task: a single large N-body simulation can take from a few hours to several weeks on a supercomputer (Springel et al., 2005; Potter et al., 2017; Collaboration et al., 2019; Sgier et al., 2019).

One approach to overcome this challenge is to use simulation-based emulators of summary statistics of the maps. Emulators have so far focused on: (a) the power spectrum, which is commonly used in cosmology (Knabenhans et al., 2019; Heitmann et al., 2016; Knabenhans et al., 2020; Angulo et al., 2020), (b) covariance matrices of 2-pt functions (Sgier et al., 2019; Taylor et al., 2013; Sato et al., 2011), and (c) non-Gaussian statistics of mass maps, which can be a source of significant additional cosmological information (Pires et al., 2009; Petri et al., 2013; Zürcher et al., 2020; Fluri et al., 2018). These approaches, however, always considered a specific summary statistic, which limits the type of analysis that can be performed using the mass-map data. They typically do not simultaneously capture both the signal and its variation: the emulators interpolate the power spectrum across the cosmological parameter space, without considering the change in its covariance matrix, which is typically taken from the fiducial cosmology parameter set. This is a known source of potential error in the analysis (Eifler et al., 2009) and was shown to have a large impact on the deep learning-based constraints (Fluri et al., 2018). The solution proposed in this work address these problems simultaneously. We construct a map-level probabilistic emulator that generates the mass maps directly, and can accurately capture the signal and its variability. This emulator, built for a specific target survey dataset, would be of great practical use for innovative map-based cosmological analyses, additionally capturing the variation of the maps across the cosmological parameter space.

With a similar goal, multiple contributions have leveraged the recent advances in the field of deep learning to aid the generation of cosmological simulations. In particular, recent works (Mustafa et al., 2017; Rodriguez et al., 2018; Nathanaël et al., 2019; Tröster et al., 2019) have demonstrated the potential of Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) for production of N-body simulations. The work of (Mustafa et al., 2017; Rodriguez et al., 2018; Nathanaël et al., 2019; Tröster et al., 2019; Giusarma et al., 2019; He et al., 2019) has shown deep generative models that can accurately model dark matter distributions and other related cosmological signals. However, a practical application of these approaches in an end-to-end cosmological analysis is yet to be demonstrated. In this work, we take an essential step toward the practical use of generative models by creating the first emulator of weak lensing mass maps as function of cosmological parameters. This step allows the generate mass maps with any parameters without the need to retrain the generative model. Our conditional GAN model generates convergence maps dependent on values of two parameters that have the largest impact on the evolution of the Large Scale Structure (LSS) of the Universe: Ω_m, which controls the matter density as a fraction of total density, and $\sigma 8$, which controls the strength of matter density fluctuations (see (Refregier, 2003; Kilbinger, 2015) for reviews). Those are the only two parameters that can be effectively measured using the convergence maps data. After training, the conditional model can then interpolate to unseen values of σ_8 and Ω_m by varying the distribution of the input latent variable. Other works (Tamosiunas et al., 2020; Villaescusa-Navarro et al., 2020) have since also explored such models, although with the emphasis on generating various cosmological fields themselves, either in ^{2}D or 3D.

To assess that the GAN-generated maps are statistically very close to the originals, we perform an extensive quantitative comparison. We evaluate our GAN using both cosmological and image processing metrics: the power spectral density, mass map histogram, peak histogram, the bispectrum, Minkowski functionals, Multi-Scale Structural Similarity (MS-SSIM) (Wang et al., 2003), and an adaptation of the Fréchet Inception Distance (FID) (Heusel et al., 2017). We also compare the statistical consistency of a batch of generated maps by computing the correlation matrices of power spectra. Moreover, we assess the agreement as a function of cosmological parameters. This set of comparisons is the most exhaustive presentation of the capacity of generative models to learn the dark matter maps, to date. In this work we use the data generated by (Fluri et al., 2019).

We build a sky convergence map dataset made of 57 different cosmologies (set of parameters) divided into a training set and a test set. The test set consists of 11 cosmological parameters sets was used to asses the capacity of the GAN to interpolate to unseen cosmologies.

This paper is structured as follows. In **Section 2** we present a new type of generative adversarial network whose generated output can be conditioned on a set of parameters in the form of continuous values. **Section 3** describes the simulation dataset used in this work. In **Section 4** we describe the metrics used to evaluate the quality of the generative model. **Section 5** shows the maps generated by our machine learning model, as well as compares its results to the original, simulated data. We summarize our findings and discuss the future prospects in **Section 6**. **Appendix A** contains the architectures of the neural networks used in this work.

2 CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

A GAN consists of two neural networks, *D* and *G*, competing against each other in a zero-sum game. The task of the

discriminator D is to distinguish real (training) data from fake (generated) data. Meanwhile, the *generator G* produces samples with the goal of deceiving the discriminator into believing that the generated data is real. Both networks are trained simultaneously and if the optimization process is carried out successfully, the generator will learn to produce the data distribution (Goodfellow et al., 2014). Learning the optimal parameters of the discriminator and generator networks can be formulated as optimizing a min-max objective. Optimizing a GAN is a challenging task due to the fact that it consists of two networks competing against each other. In practice, one often observes unstable training behaviors which can be mitigated by relying on various types of regularization methods (Roth et al., 2017; Gulrajani et al., 2017). In this paper, we rely on Wasserstein GANs (Arjovsky et al., 2017) with the regularization approach suggested in (Gulrajani et al., 2017). The model we use conditions both the generator and the discriminator on a given random variable y, yielding the following objective function,

$$\min_G \max_D \mathop{\mathbb{E}}_{(x,y)\sim\mathbb{P}_r}[D(x,y)] - \mathop{\mathbb{E}}_{z\sim\mathbb{P}_z, y\sim\mathbb{P}_y}[D(G(z,y))] + \lambda \mathop{\mathbb{E}}_{(x,y)\sim\mathbb{P}_r\cup\mathbb{P}_g}\left[\left(\|\nabla_x D(x,y)\|_2 - 1\right)^2\right] \quad (1)$$

where $\mathbb{P}_r$ and $\mathbb{P}_z$ are the data and latent variable distributions. The parameter $\lambda \geq 0$ is the penalty coefficient of the regularization term that ensures that the gradient norm of the discriminator is close to 1. This ensures that the discriminator is 1-Lipschitz, which is a requirement for optimizing the Wasserstein distance (Gulrajani et al., 2017; Arjovsky et al., 2017). The prior distribution of the latent variable, e.g., a uniform or a Gaussian distribution, defines implicitly the generator distribution $\mathbb{P}_g$ by $(x,y) = G(z,y), z \sim \mathbb{P}_z, y \sim \mathbb{P}_y$.

Practically, there exist many techniques and architectures to condition the generator and the discriminator (Gauthier, 2014; Perarnau et al., 2016; Reed et al., 2016; Odena et al., 2017; Miyato and Koyama, 2018). However, all the architectures in these works are conditioning on discrete parameters. We instead propose a different design that works specifically for continuous parameters and will be shown to have good performance in practice. We note that our conditioning technique could be used with other architectures as well. For simplicity we describe the case of a single parameter, but our technique was implemented for the case of two parameters. Our idea is to adapt the distribution of the latent vector according to the conditioning parameters using the function $\hat{z} = f(z,y)$. Specifically, the function f simply rescales the norm of the latent vector according to the parameter y. Given the range $y \in [a,b]$, f reads:

$$\hat{z} = f(z,y) = \left(l_0 + \frac{l_1 - l_0}{b-a}(y-a)\right)\frac{z}{||z||_2}. \quad (2)$$

Using this function, the length of the z vector is mapped to the interval $[l_0, l_1]$. In our case, we used $l_0 = 0.1\sqrt{n}$ and $l_1 = \sqrt{n}$, where n is the size of the latent vector. For the discriminator, the parameters are concatenated directly after the convolutional layers as in (Reed et al., 2016). The relation between the features extracted from the convolutional layers and the parameters might in general be non-local. We therefore increase the complexity of the mapping functions of the discriminator and generator by adding some linear layers (as in a multi-layer perceptron) at the end of the discriminator and the beginning of the generator. The proposed model is sketched in **Figure 1** and the architecture is described in more details in **Appendix A**. Specific parameters can be found in **Table A1**.

3 SKY CONVERGENCE MAPS DATASET

The data used in this work is the non-tomographic training and testing set introduced in (Fluri et al., 2019), without noise and intrinsic alignments. The simulation grid consists of 57 different cosmologies in the standard cosmological model: a flat Universe with cold dark matter (ΛCDM) (Lahav and Liddle, 2019). Each of these 57 configurations was run with different values of Ω_m and σ_8, resulting in the parameter grid shown in **Figure 2**. The output of the simulator consists of the particle positions in 3D space. The mass maps are obtained by the *gravitational lensing* technique (see (Bartelmann, 2010) for review). It consists of a tomographic projection of the particle densities along the radial (redshift) direction against the *lensing kernel*. This kernel is dependent on the relative distances between the observer and the lensed galaxies that are used to create the mass maps. The source galaxy redshift distribution $n(z)$ used in this work is the non-tomographic distribution from (Fluri et al., 2019). The projected matter distribution is pixelized into images of size 128 px × 128 px, which corresponds to $5^\circ \times 5^\circ$ of the sky. Eventually, the resulting dataset consists of 57 sets of 12,000 sky convergence maps for a total of 684,000 samples. At training time, we randomly rotate and flip the input image to augment the dataset.

The dataset is split into a training and test set in the following way: 11 cosmologies (132,000 samples) are selected for the test set, and the remaining 46 cosmologies (552,000 samples) are assigned to the training set, as depicted in **Figure 2**. This split is used to ensure that the model could interpolate to unseen cosmologies. At evaluation time, we use the cosmologies from the test set to validate the interpolation ability of our network. In the following sections, we show detailed summary statistics for the cosmologies marked with letters A, B, C, and D. We make the dataset publicly available.[3]

4 QUANTITATIVE COMPARISON METRICS

We make a quantitative assessment of the quality of the generated maps using both cosmological summary statistics and similarity metrics used in computer vision. We focus on the following statistics:

1. the power spectral density C_ℓ, which describes how strongly the maps are correlated as a function of pixel separation ℓ,

[3]https://zenodo.org/record/4646764

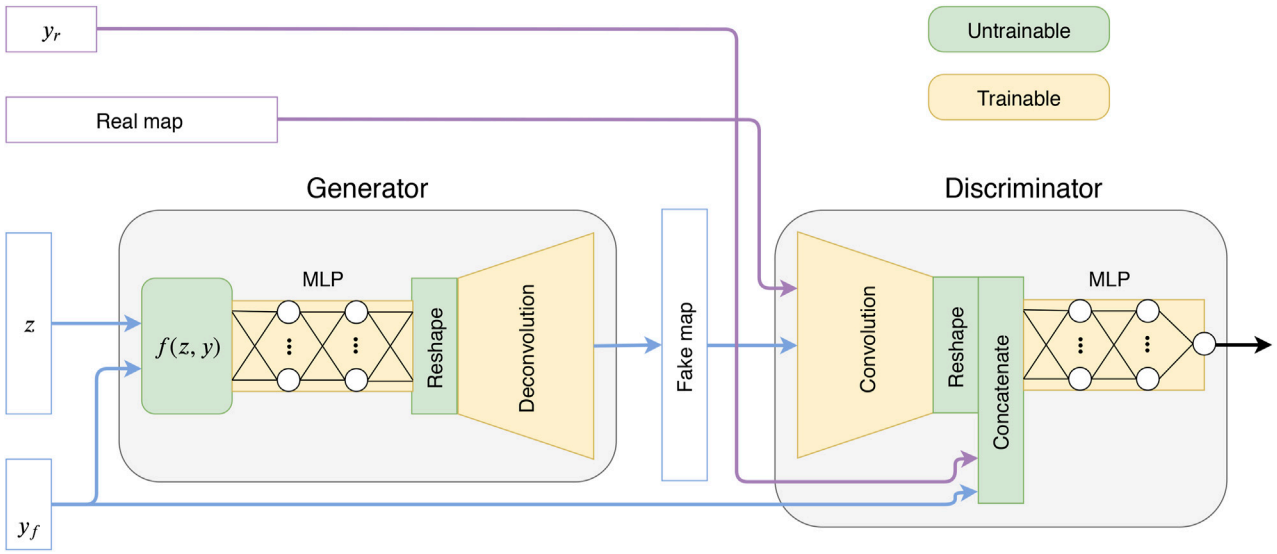

FIGURE 1 | Sketch of the proposed GAN model, where z is a latent variable and y is a parameter vector (y_r = real, y_f = fake).

2. the distribution of mass map pixels N_{pixels}, compared using Wasserstein-1 distance and histograms,
3. the distribution of mass map peaks N_{peaks}, which describes the distribution of values at the local maxima of the map, compared also compared using Wasserstein-1 distance and histograms,
4. the bispectrum B_ℓ, which describes the three-point correlation of the folded triangles of different size,
5. Minkowski functionals, which are morphological measures of the map, and consist of three functions: V_0, which describes the area of the islands after thresholding of the map at some density level, V_1, their perimeter, and V_2, their Euler characteristic (their number count minus the number of holes),
6. the Pearson's correlation matrices $R_{\ell\ell'}$ between the C_ℓ of maps at different cosmologies,
7. the Multi-Scale Structural Similarity Index (MS-SSIM) (Wang et al., 2003; Odena et al., 2017), which is an image similarity measure commonly used in computer vision,
8. the Fréchet Distance between the output of a CNN regressor trained to predict Ω_m, σ_8, similarly to the Fréchet Inception Distance calculated using the Google Inception v3 network (Heusel et al., 2017).

The mass map histograms and the peak counts are simple statistics used to compare the maps and constrain cosmological models (see Gatti et al., 2020; Kacprzak et al., 2016 for examples). These metrics, however, ignore the spatial information in the maps. The angular power spectrum C_ℓ or its real-space equivalent, the angular correlation function, is the most common statistic for constraining cosmology with LSS (see (Kilbinger, 2015) for review). The 2-pt functions capture only the Gaussian part of the fluctuations. The 3-pt correlation function, or the bispectrum, probes higher order information and has also been used for constraining cosmological models

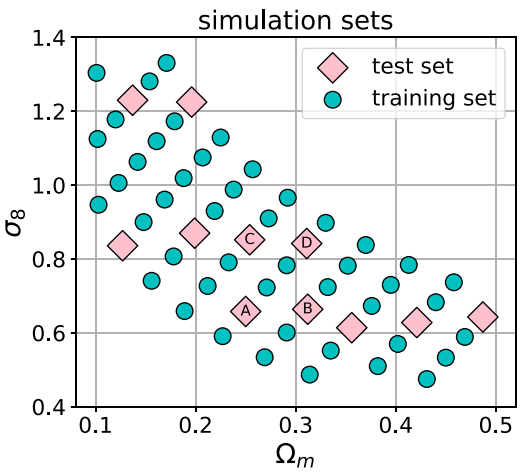

FIGURE 2 | The cosmological parameter grid used in this work, from (Fluri et al., 2019). The circles and diamonds show the training and the test sets, respectively. The total number of models was 57, of which 46 were used as the training set and 11 as the test set. The models labeled A, B, C, and D are investigated in more detail in **Section 5**.

(Takada and Jain, 2003; Fu et al., 2014). Similarly, the Minkowski functionals have also been used for cosmological measurements (Petri et al., 2015) as an alternative statistic that extracts topological information from the maps.

The agreement between the pixel and peak values of N-body and GAN-generated images is quantified using the Wasserstein-1 distance $W_1(P, Q)$. This distance corresponds to the *optimal transport* of probability mass to turn the distribution P into Q. As it is scale-dependent, we calculate it after normalizing the pixel values: we subtract the mean and divide by the standard deviation. We use mean and standard deviation of all N-body generated images for a given cosmology, for both samples. This

way, the W_1 distance is easily interpretable: for a Gaussian with $\mu = 0, \sigma = 1$, a 1σ shift of the mean corresponds to $W_1 = 1$, and scaling its variance by $\times 2$ lead to $W_1 \approx 0.8$.

For B_ℓ, C_ℓ, and $V_{0,1,2}$ we calculate the simple fractional difference between the original and generated samples, defined as $f_x = (x_{\text{GAN}}/x_{\text{N-body}})/x_{\text{N-body}}$. We quantify the agreement between correlation matrices by comparing their Frobenious norms $\|\cdot\|_F$. For a N-dimensional, diagonal covariance matrix with elements σ_i^2, the Frobenious norm scales linearly with $\sigma/\sqrt{N}$. This way, it can be interpreted as a linear proxy for information content. We define the fractional difference between the Frobenious norm of GAN and N-body correlation matrices as:

$$f_{\text{R}} = \frac{\left\|\text{R}^{\text{GAN}}\right\|_F - \left\|\text{R}^{\text{N-body}}\right\|_F}{\left\|\text{R}^{\text{N-body}}\right\|_F}. \tag{3}$$

The Multi-Scale Structural Similarity Index (MS-SSIM) is useful in order to detect the problem commonly known as mode collapse, where the generator produces only a small subset of the training data distribution. Detecting this undesirable behavior is non-trivial as summary statistics can still agree during mode collapse. Taking inspiration from (Odena et al., 2017), one solution is to leverage the MS-SSIM score from (Wang et al., 2003) to quantify this effect. This metric was first proposed for prediction of similarity in human perception of images. Taking two images as inputs, it returns a value between 0 and 1, where 1 means "identical" and 0 means "completely different." As the mass maps are stochastic and only similar in a statistical way, we are not interested in the similarity between a pair of specific images, but in the average similarity of a large set of images. We calculate the significance of the difference in the SSIM measures in the following way:

$$s_{\text{SSIM}} = \frac{\langle \text{SSIM}^{\text{GAN}} \rangle - \langle \text{SSIM}^{\text{N-body}} \rangle}{(\sigma[\text{SSIM}^{\text{GAN}}] + \sigma[\text{SSIM}^{\text{N-body}}])/2} \tag{4}$$

where $\langle \text{SSIM} \rangle$ is the mean score, and $\sigma[\text{SSIM}]$ is the standard deviation. Large differences in the SSIM score indicate a significant difference in the samples generated by the GAN, thus pointing out to potential problems with the quality of the generated samples. On the other hand, a small difference will be an indicator that the generative model preserve the data statistics.

Finally, we calculate an adaptation of the Fréchet Inception Distance (FID) (Heusel et al., 2017) between N-body and GAN -generated images. The Inception Score (IS) (Salimans et al., 2016) and FID have become standard measures for GANs. The idea consists to compare statistics of the output of the Google Inception-v3 network (Szegedy et al., 2016) for the ImageNet dataset (Deng et al., 2009). This has proven to be well correlated with human score. As the reference Inception network used for the FID was trained with the ImageNet dataset, its output statistics are meaningless for cosmological mass maps. To solve this challenge, we create our own reference network that is well suited for cosmological mass maps. This network is a CNN trained to perform a regression task and predict the true σ_8, Ω_m parameters, similarly to (Fluri et al., 2018; Schmelzle et al., 2017; Gupta et al., 2018). Its parameters and detailed explanations of its construction can be found in **Table A2** and in **Appendix B**. The adapted FID score is obtained by comparing the regressor outputs for the N-body and GAN images. As regressor is composed of seven layers, this comparison depends on high order moments. Naturally, we expect that a well working conditional GAN should generate samples with similar output distribution to the one of the real samples. To estimate the distance between the two statistics distributions, we first approximate the network predictions with a normal distributions μ_r, Σ_r and μ_g, Σ_g, for the N-body and GAN -generated input, respectively. The FID is then calculated as:

$$\text{FID} = \left\|\mu_r - \mu_g\right\|^2 + \left\|\Sigma_r^{1/2} - \Sigma_g^{1/2}\right\|^2. \tag{5}$$

Note that this formula also correspond to the Wasserstein-1 distance between the two Gaussian distributions (Dowson and Landau, 1982). Eventually, before calculating FID, we normalize the network outputs for each true cosmology: we subtract the mean and divide by the standard deviation of the N-body sample. For the ease of interpretation, we report the square root of FID. This way, a 1σ difference in the mean CNN predictions will correspond to $\text{FID}^{1/2} = 1$. Similarly, a change of 1σ in the covariance matrix also leads to $\text{FID}^{1/2} = 1$.

5 RESULTS

We trained the GAN model described in **Section 2** and **Appendix A**. We used RMSPROP as an optimizer with an initial learning rate of 10^{-5} and a batch size of 64. The discriminator was updated 5 times more than the generator. The gradient penalty was set to 10 and the negative slope of the LeakyRelu $\alpha = 0.2$. It took a week to train the model for 40 epochs on a GeForce GTX 1080 GPU. Similar to (Reed et al., 2016; Odena et al., 2017; Miyato and Koyama, 2018; Mirza and Osindero, 2014), we use batches composed of samples from different parameter sets. Note that the batches were composed of samples from different cosmologies from the *training* set. The summary statistics are computed using 5,000 real and fake samples for every pair of parameters of the *test* set. The peaks are extracted by searching for all pixels greater than their 5×5 patch neighborhood, i.e. their 24 neighbors. Then, the histogram of the extracted peaks values is computed. We rely on LENSTOOLS (Petri, 2016) to compute the power spectra, bispectra and the Minkowski functionals. For the bispectrum, we use the *folded* configuration, with the ratio between one of the triangle sides and the base is set to the default value of 0.5. The SSIM is computed using the SCIKIT-IMAGE packge[4] (Van der Walt et al., 2014). We make our code is publicly available.[5]

Figure 3 shows images generated by the conditional GAN and as well as original ones, for several values of Ω_m and σ_8 parameters. They are visually indistinguishable. Furthermore, the image structure evolves similarly with respect of the

[4]https://scikit-image.org/

[5]https://renkulab.io/gitlab/nathanael.perraudin/darkmattergan

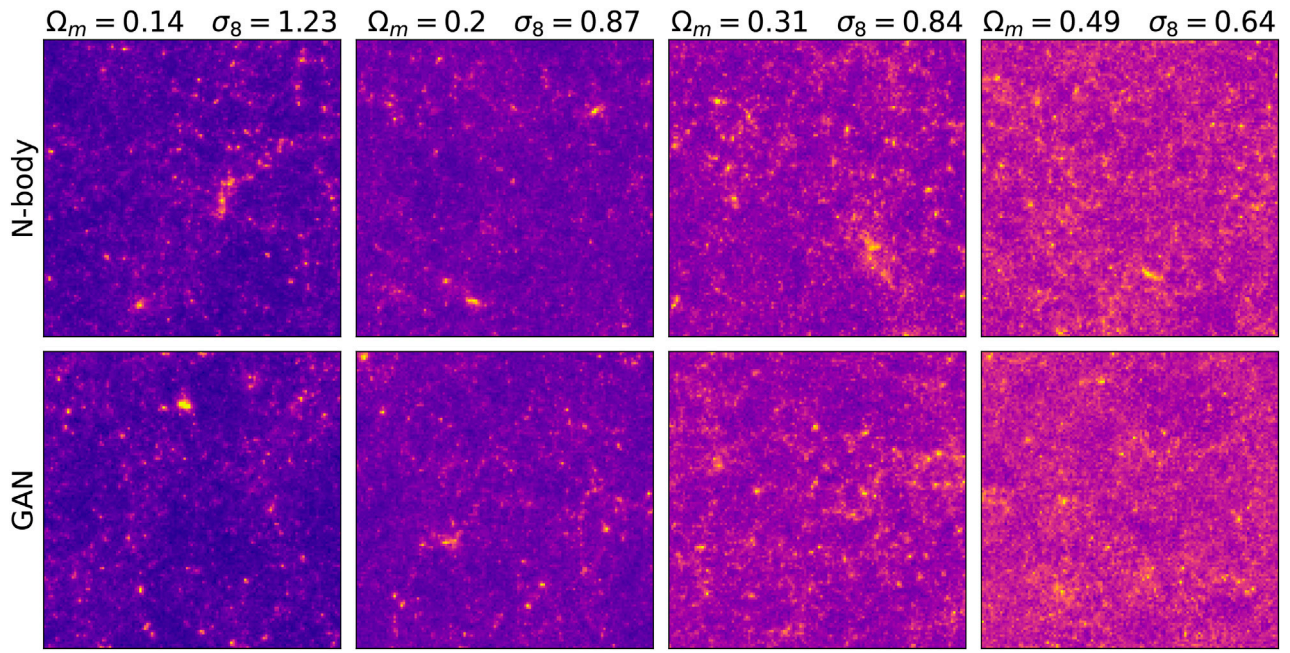

FIGURE 3 | The original N-body images and GAN-generated images for four cosmological parameter sets.

cosmological parameters change. As predicted by the theory, increasing Ω_m results in convergence maps with additional mass and increasing σ_8 in images with higher variance in pixel intensities. In **Figure 4** the same latent variable z is used to generate different cosmologies. The smooth transition from low to high mass density hints that the latent variable control the overall mass distribution and the conditioning parameter its two cosmological properties σ_8, Ω_m.

Figure 5 shows the histograms of pixels (top) and peaks (bottom) of the original maps simulated using N-body simulations (blue), and their GAN-generated equivalents (red), for the four models A,B,C,D shown in **Figure 2**. The peak counts were selected as maxima of the surrounding 24 neighbors. The solid line corresponds to the median of the histograms from 5,000 realisations, and the bands to 32% and 68% percentiles. The bottom part of each panel shows the fractional difference between the statistics, defined as $f_x = (x_{\mathrm{GAN}}/x_{\mathrm{N-body}})$. The normlized Wasserstein-1 distance of the pixel values distribution (see **Section 4**) is: $W_1^{\mathrm{pixel}} = 0.04, 0.02, 0.02, 0.02$, for models A, B, C, and D, respectively. That indicates that the histograms differ on the level of <5%. Similarly, the Wasserstein-1 distances of the peak value distribution for models A, B, C, and D is: $W_1^{\mathrm{peak}} = 0.04, 0.03, 0.01, 0.02$. The agreement here is also very good, on <5% level.

The 2-pt and 3-pt statistics are shown in **Figure 6**. The power spectra C_ℓ overlap almost perfectly for all the cosmologies lying inside the parameter grid used for training. Again, the agreement is better than 5%. The agreement for the bispectrum B_ℓ is good for models B and C, but worse for A and B; the GAN model seems to underestimate the strength of the 3-pt correlations for these models, which differ by ≈20%. We note that the 3-pt signal is very weak and has a large variation, which may be difficult to model for the GANs.

The Minkowski functionals are presented in **Figure 7**. They were calculated using LENSTOOLS (Petri, 2016). The functional V_0 (first line) corresponds to the area of the emerging "islands," V_1 (second line) to their circumference, and V_2 (third line) to their Euler characteristic (their number count minus the number of holes in them). The value of threshold κ, above which the functional values, i.e. the "islands" are calculated, is shown on the x-axis. Here the agreement is typically better than 10%, with some model D agreeing much better, to ≈2%. The large differences in the fractional difference plots are due to instability close to value of $V = 0$. The confidence limits of the summary statistics shown in these figures overlap very well, which indicates that the variability of these statistics is also captured very well by the GAN system.

The Pearson's correlation matrices **R** of the power spectra are shown in **Figure 8**. Those correlations were created from a coarsely-binned C_ℓ in range $\ell \in [300, 3000]$. The upper and lower triangular parts of the matrix show the original N-body correlations and the GAN correlations, respectively. We calculate the Frobenious norms of these matrices and compare their ratios using **Eq. 3**. For the models A, B, C, D this difference is: $f_{\mathrm{R}} = 0.02, 0.14, 0.06, 0.06$. This agreement is overall very good, with model B being slightly worse. As the precision requirements for covariance matrices are not as strict as for the summary statistics, this level of agreement can be considered satisfactory for upcoming applications (Taylor et al., 2013).

We calculate the mean and standard deviation of MS-SSIM score between 5,000 randomly selected images for each cosmology, both for GAN and original N-body maps. We test

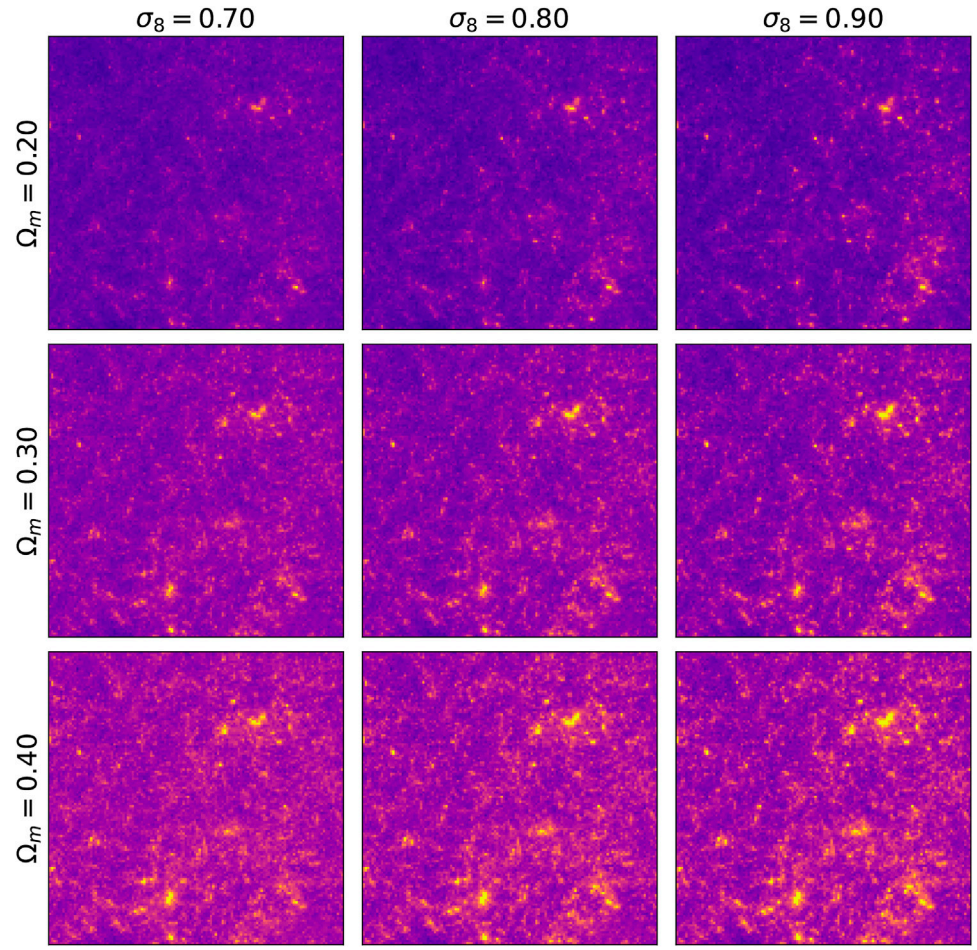

FIGURE 4 | Images generated with the same random seed but with different input cosmology parameters Ω_m and σ_8.

if the mean SSIM score is consistent between the N-body and GAN data using **Eq. 4**. The SSIM difference significance for the four models A, B, C, and D are: $s_{\text{SSIM}} = 0.08, 0.23, -0.27, 0.42,$. This indicates very good statistical agreement for these models.

Figure 9 shows the prediction of a regressor CNN trained on the N-body images with true σ_8, Ω_m values. For each category, we make the prediction with 500 randomly selected maps. The shaded areas show the 68 and 95% probability contours for the N-body image input (blue) and the GAN image input (red). The agreement is relatively good, but differences in the spread of these distributions is noticeable. The Fréchet Distance (FID) computed using the reference cosmological CNN, as described in **Section 4**, is: $\text{FID}^{1/2} = 1.26, 1.44, 1.00, 1.19,$ for models A,B,C, and D. This indicates a slight difference according to this metric and agree with the distributions in **Figure 9**.

We compare the summary statistics as a function of cosmological parameters for both the training and the test set. We used the training and test sets displayed in **Figure 2**. **Figure 10** shows the six quantities as a function of cosmological parameters:

top left: significance of the difference s_{SSIM} in Multi-Scale Structural Similarity Index (**Eq. 4**),

top center: Fréchet distance using a CNN regressor (**Eq. 5**). Note that for a Gaussian distribution, a difference of $\text{FID}^{1/2} = 1$ corresponds to either a 1σ shift in the mean or 1σ difference in standard deviation,

top right: normalized Wasserstein-1 distance of the pixel value distribution. For a Gaussian distribution, a 1σ change in the mean corresponds to $W_1 = 1$, and an increase in standard deviation of $\times 2$ to $W_1 \approx 0.8$,

bottom left: average fractional differences in the power spectrum f_{C_ℓ},

bottom center: fractional differences in the power bispectrum f_{B_ℓ},

bottom left: fractional difference in the Frobenious norm of the correlation matrices f_R (**Eq. 3**).

Overall, we notice that the agreement between the N-body simulations and GAN-generated maps is the best in the center of the grid for both the training and test set. The fact that the differences in neighboring cosmologies are similar indicates that the GAN system can efficiently learn the latent interpolation of

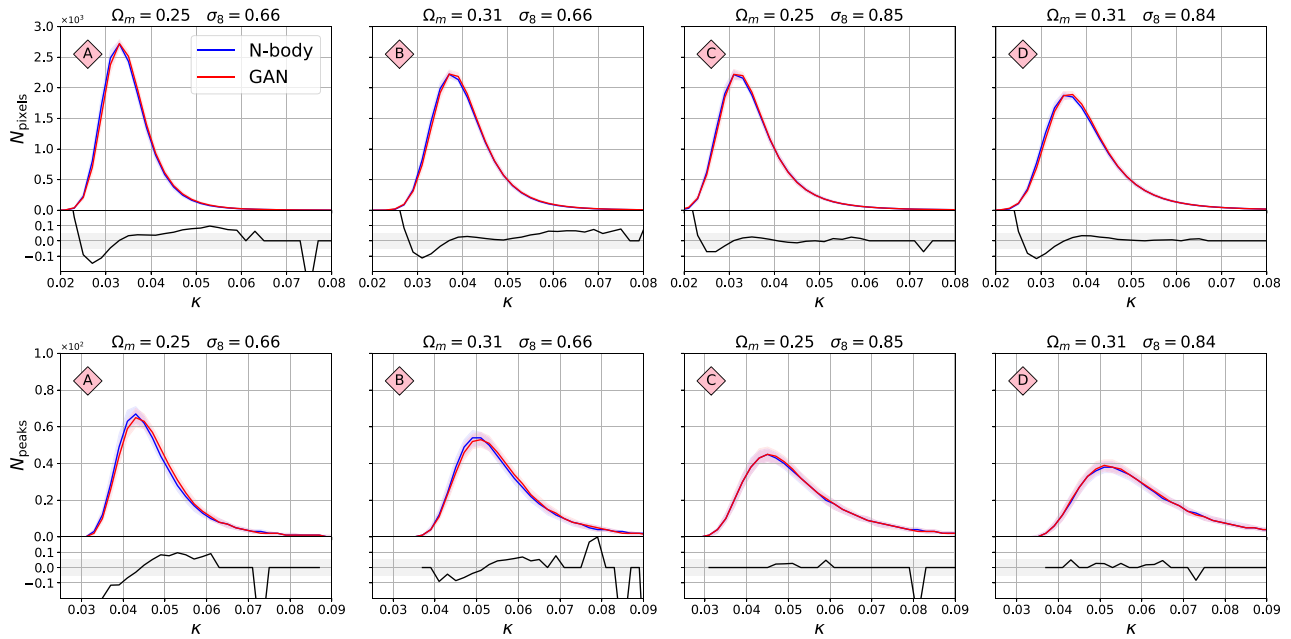

FIGURE 5 | Comparison of histogram of pixel values **(top)** and peaks **(bottom)** between the original N-body and the GAN-generated maps. Models A, B, C, D correspond to the ones marked in **Figure 2**. The x-axis value κ is the map pixel intensity. The solid line is the median histogram from 5,000 randomly selected maps. The bands correspond to 32 and 68% confidence limits of the ensemble of histograms. The lower panels show the fractional difference between the median histograms.

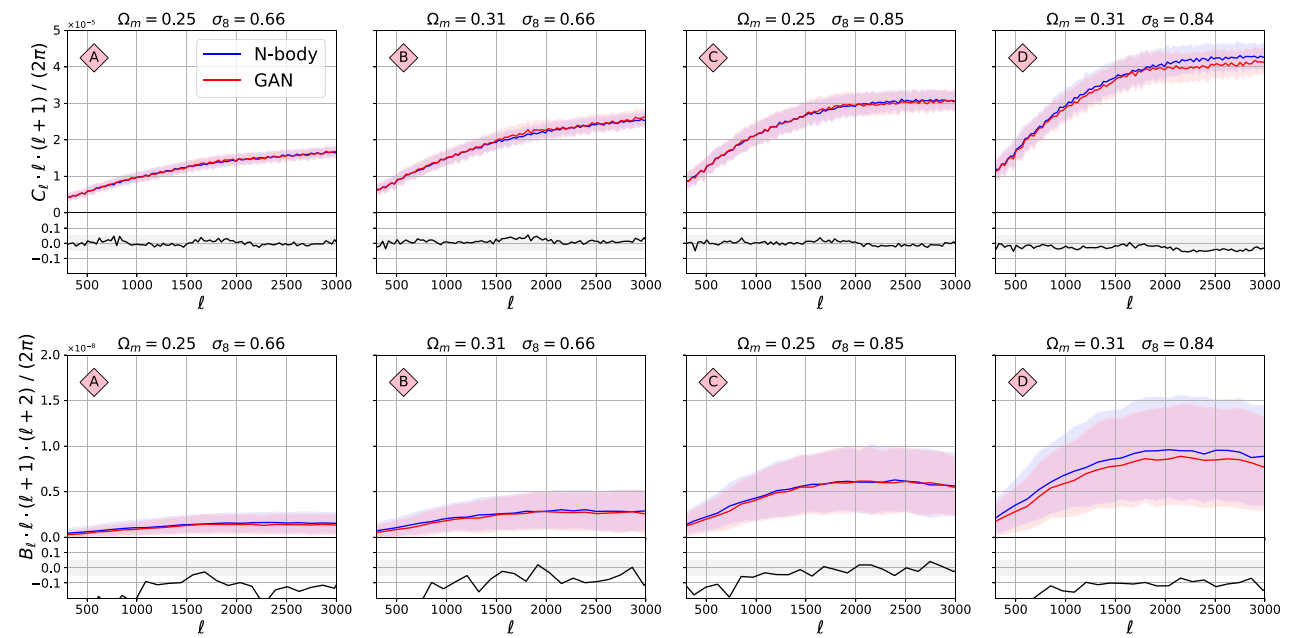

FIGURE 6 | Comparison of the 2-pt and 3-pt functions between the original N-body maps and GAN-generated maps. The structure of this figure is the same as for **Figure 5**.

the maps. The agreement worsens at the edges of the grid. We observe the biggest deterioration in the realism of the GAN model for the high Ω_m and low σ_8 parameters. This is most prominent for the correlation matrix and the SSIM differences. Conversely, the biggest difference for the bispectrum is present for low Ω_m and high σ_8.

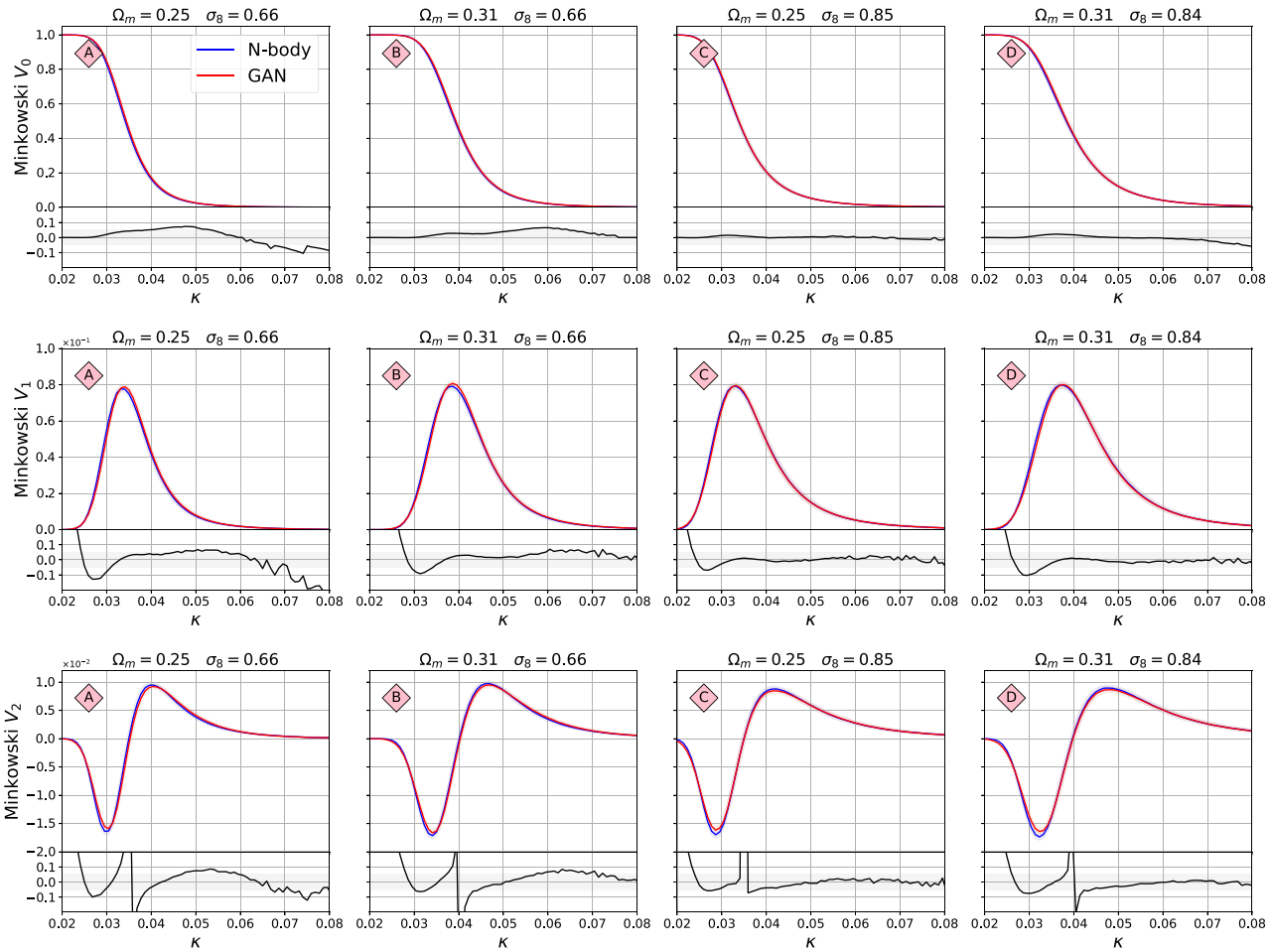

FIGURE 7 | Comparison of the Minkowski functional between the original N-body maps and GAN-generated maps. The value of threshold κ, above which the functional value is calculated, is shown on the x-axis. The functional V_0 corresponds to the area of the emerging "islands," V_1 to their circumference, and V_2 to their Euler characteristic (their number count minus the number of holes in them). The structure of this figure is the same as for **Figure 5**.

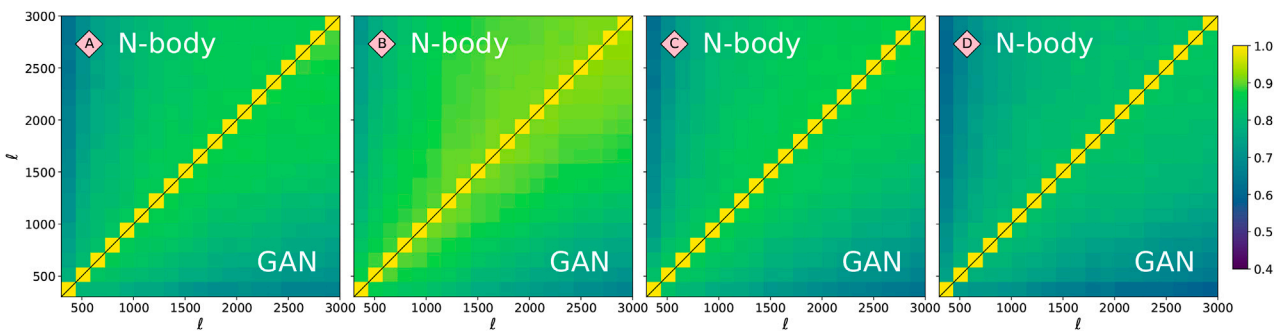

FIGURE 8 | Pearson's correlation matrices for the four models highlighted in **Figure 2**. The upper triangular corresponds to the original N-body power spectra, while the lower triangular to the power spectra of GAN-generated images. The fractional difference of the Frobenious norms (**Eq. 3**) of these matrices is $f_R = 0.02$, 0.14, 0.06, 0.06, for models A, B, C, and D, respectively.

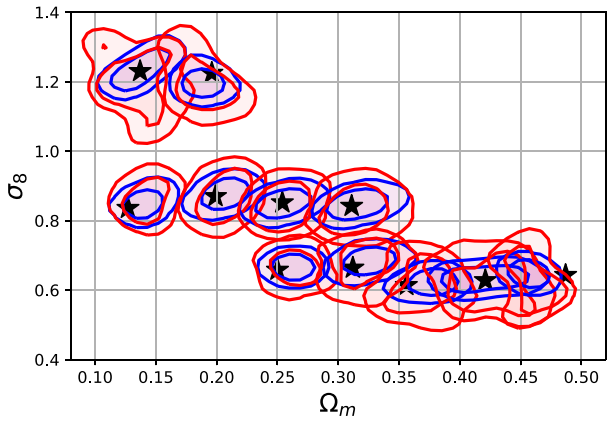

FIGURE 9 | Predictions of a regressor CNN trained to predict the Ω_m, σ_8 from input images. The details of this experiment are described in **Section 4** and the network architecture in **Table A1** in **Appedix A**. The contours encircle the 68 and 95% samples for the N-body maps (blue) and GAN-generated maps from the test set. The black stars show the true values of the test set parameters.

6 CONCLUSION

We proposed a new conditional GAN model for continuous parameters where conditioning is done in the latent space. We demonstrated the ability of this model to generate sky convergence maps when conditioning on the cosmological parameters Ω_m and σ_8. Our model is able to produce samples that resemble samples from the test set with good statistical accuracy, which demonstrates its generalization abilities. The agreement of the low order summary statistics (pixel and peak histograms and power spectrum) is very good, typically on the <5% level. Higher order statistics (Minkowski functionals, bispectrum) agree well, but with larger differences, generally around ≈10%, and in some cases ≈20%. The comparison of the Multi-Scale Structural Similarity Index (MS-SSIM) shows a good agreement in this metric, with the exception of the low σ_8 and high Ω_m edge of the grid. Moreover, the GAN model is able to capture the variability in the conditioned dataset: we observe that the scatter of the summary statistics computed from an ensemble is very similar

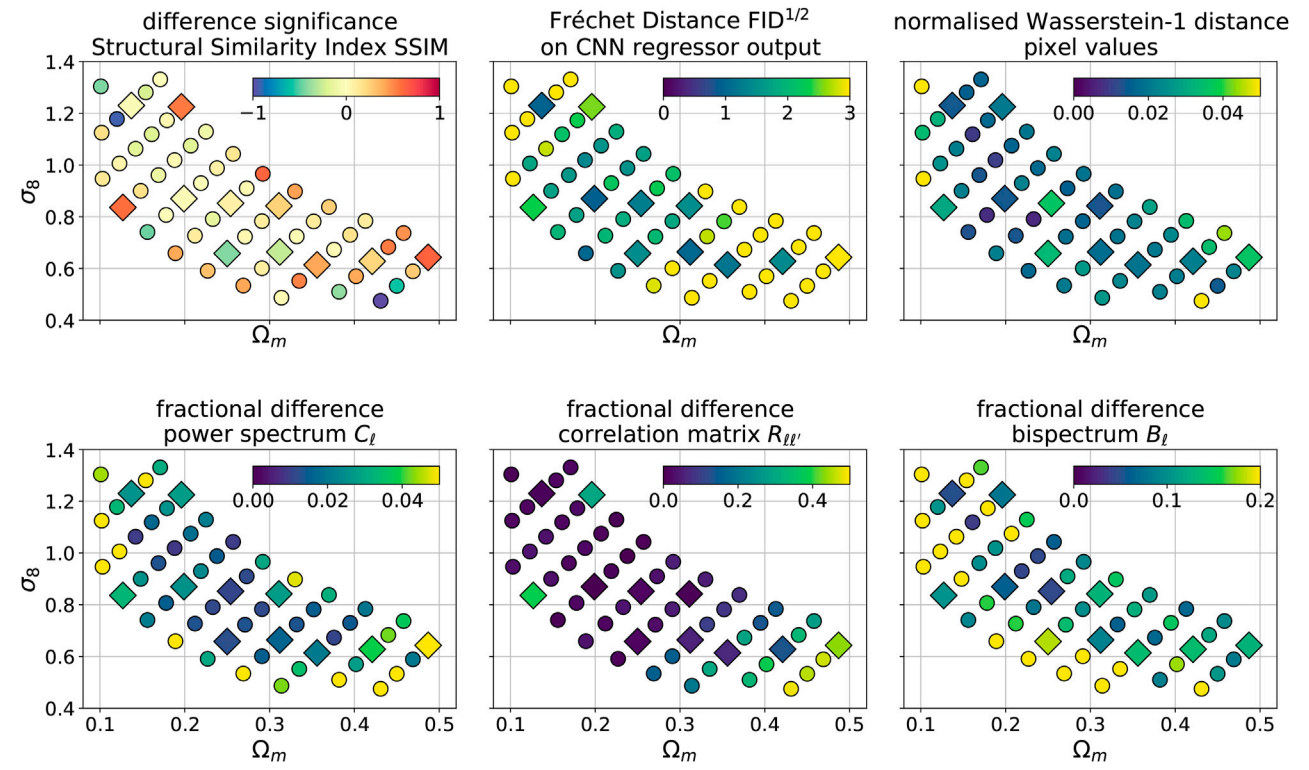

FIGURE 10 | Differences between summary statistics of the original N-body and the GAN-generated images. The left panel shows the significance of the difference in Multi-Scale Structural Similarity Index (MS-SSIM), defined in **Eq. 4**. The upper middle panel presents the Fréchet Distance, computed using a regressor CNN and **Eq. 5** (see **Section 4**). The upper right panel shows the normalized Wasserstein-1 distance in the pixel value distributions (see **Section 4**). The lower left panel shows the mean absolute fractional difference of the power spectra C_ℓ. The lower middle panel shows the fractional difference in the Frobenius norms of correlation matrices, defined in **Eq. 3**. The lower right panel present the mean absolute fractional difference of the bispectrum B_ℓ. The circles and squares indicates parameters from the training and the test sets.

between the original and generated images. The investigation of the correlation matrices of the power spectra also shows a good agreement, with a quality deteriorating close to the edges of the grid, especially for low σ_8 and high Ω_m. This is not unexpected, as the training set contains less information near the edges of the grid. More investigation is needed to more close inspect the behavior of the generative model in these areas. As generative models are rapidly growing in popularity in machine learning, we anticipate to be able to solve these problems in the near future.

Our results offer good prospects for GAN-based conditional models to be used as emulators of cosmology-dependent mass maps. As these models efficiently capture both the signal and its variability, the map-level emulators could potentially be used for cosmological analyses. They can accurately predict the power spectrum and its covariance, which is often unattainable in standard cosmological analyses (Eifler et al., 2009). It can also be used for non-Gaussian analyses of lensing mass maps, such as, for example, in (Zürcher et al., 2020; Parroni et al., 2020). Further experiments will be needed, however, to bring the generative models to a level where they can be of practical use in a full, end-to-end cosmological analysis.

In this paper, we have demonstrated the ability of generative AI models to serve as emulators of cosmological mass maps for a given redshift distribution of source galaxies $n(z)$. Generative models have also been shown to work directly on the full or sliced 3D matter density distributions (Nathanaël et al., 2019; Tröster et al., 2019; Villaescusa-Navarro et al., 2020). The three dimensional generation of cosmological fields proves to be particularly difficult. As most of the survey experiments publish their lensing catalogs and their corresponding redshift distributions, the generation of projected maps, as shown in this work, could be of direct practical use. Another challenge will be posed by the large sky area of the upcoming surveys and their spherical geometry. Spherical convolutional neural networks architectures have been proposed (Perraudin et al., 2018; Krachmalnicoff and Tomasi, 2019; McEwen et al., 2021). These architectures are expected to be easy to implement with generative models, which offers good prospect for the development of spherical mass map emulators.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://renkulab.io/projects/nathanael.perraudin/darkmattergan, https://zenodo.org/record/4646764

AUTHOR CONTRIBUTIONS

NP experiment lead, experiment execution, code maintenance, paper writing SM experiment execution, implementation of conditional GANs AL machine learning advice, consulting, paper writing TK dataset creation, cosmology advice, consulting, paper writing, plotting, method evaluation.

FUNDING

This work was supported by a grant from the Swiss Data Science Center (SDCS) under project "DLOC: Deep Learning for Observational Cosmology" and Grant Number 200021 169130 from the Swiss National Science Foundation (SNSF).

ACKNOWLEDGMENTS

This work was supported by a Grant from the Swiss Data Science Center (SDCS) under project *DLOC: Deep Learning for Observational Cosmology* and Grant Number 200021_169130 from the Swiss National Science Foundation (SNSF). We thank Thomas Hofmann, Alexandre Réfrégier, and Fernando Perez-Cruz for advice and helpful discussions. We thank the Cosmology Research Group of ETHZ and particularly Janis Fluri for giving us access to the dataset. Finally, we thank the two anonymous reviewers who provided extensive feedback that greatly improved the quality of this paper.

REFERENCES

Angulo, R. E., Zennaro, M., Contreras, S., Aricò, G., Pellejero-Ibañez, M., and Stücker, J. (2020). The BACCO Simulation Project: Exploiting the Full Power of Large-Scale Structure for Cosmology. arXiv e-prints, arXiv:2004.06245.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein Generative Adversarial Networks. *Int. Conf. Mach. Learn.* 14, 214–223. doi:10.1109/icpr.2018.8546264

Bartelmann, M. (2010). Gravitational Lensing. *Class. Quan. Grav.* 27, 233001. doi:10.1088/0264-9381/27/23/233001

Collaboration, E., Knabenhans, M., Stadel, J., Marelli, S., Potter, D., Teyssier, R., et al. (2019). Euclid Preparation: Ii. The Euclidemulator–A Tool to Compute the Cosmology Dependence of the Nonlinear Matter Power Spectrum. *Monthly Notices R. Astronomical Soc.* 484, 5509–5529.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). Imagenet: A Large-Scale Hierarchical Image Database. *IEEE Conf. Computer Vis. Pattern Recognit.* 33, 248–255. doi:10.1109/cvprw.2009.5206848

Dowson, D. C., and Landau, B. V. (1982). The Fréchet Distance between Multivariate normal Distributions. *J. Multivariate Anal.* 12, 450–455. doi:10.1016/0047-259X(82)90077-X

Eifler, T., Schneider, P., and Hartlap, J. (2009). Dependence of Cosmic Shear Covariances on Cosmology. *A&A* 502, 721–731. doi:10.1051/0004-6361/200811276

Fluri, J., Kacprzak, T., Lucchi, A., Refregier, A., Amara, A., and Hofmann, T. (2018). Cosmological Constraints from Noisy Convergence Maps through Deep Learning. arXiv preprint arXiv:1807.08732.

Fluri, J., Kacprzak, T., Lucchi, A., Refregier, A., Amara, A., Hofmann, T., et al. (2019). Cosmological Constraints with Deep Learning from KiDS-450 Weak Lensing Maps. arXiv e-prints, arXiv:1906.03156.

Fu, L., Kilbinger, M., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., et al. (2014). CFHTLenS: Cosmological Constraints from a Combination of Cosmic Shear Two-point and Three-point Correlations. *Monthly Notices R. Astronomical Soc.* 441, 2725–2743. doi:10.1093/mnras/stu754

Gatti, M., Chang, C., Friedrich, O., Jain, B., Bacon, D., Crocce, M., et al. (2020). Dark Energy Survey Year 3 Results: Cosmology with Moments of Weak Lensing

Mass Maps - Validation on Simulations. *MNRAS* 498, 4060–4087. doi:10.1093/mnras/staa2680

Gauthier, J. (2014). "Conditional Generative Adversarial Nets for Convolutional Face Generation," in *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition* (London: Winter Semester), 2.

Giusarma, E., Reyes Hurtado, M., Villaescusa-Navarro, F., He, S., Ho, S., and Hahn, C. (2019). Learning Neutrino Effects in Cosmology with Convolutional Neural Networks. arXiv e-prints, arXiv:1910.04255.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Nets. *Adv. Neural Inform. Process. Syst.* 11, 2672–2680. doi:10.3156/jsoft.29.5_177_2

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved Training of Wasserstein Gans. *Adv. Neural Inform. Process. Syst.* 12, 5767–5777. doi:10.1109/ismsit.2019.8932868

Gupta, A., Matilla, J. M. Z., Hsu, D., and Haiman, Z. (2018). Non-gaussian Information from Weak Lensing Data via Deep Learning. *Phys. Rev. D* 97, 103515. doi:10.1103/physrevd.97.103515

He, S., Li, Y., Feng, Y., Ho, S., Ravanbakhsh, S., Chen, W., et al. (2019). Learning to Predict the Cosmological Structure Formation. *Proc. Natl. Acad. Sci. USA* 116, 13825–13832. doi:10.1073/pnas.1821458116

Heitmann, K., Bingham, D., Lawrence, E., Bergner, S., Habib, S., Higdon, D., et al. (2016). The Mira-Titan Universe: Precision Predictions for Dark Energy Surveys. *ApJ* 820, 108. doi:10.3847/0004-637X/820/2/108

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017a). Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Adv. Neural Inform. Process. Syst.* 7, 6626–6637. doi:10.1057/9781137294678.0453

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017b). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv e-prints, arXiv:1706.08500.

Kacprzak, T., Kirk, D., Friedrich, O., Amara, A., Refregier, A., Marian, L., et al. (2016). Cosmology Constraints from Shear Peak Statistics in Dark Energy Survey Science Verification Data. *Mon. Not. R. Astron. Soc.* 463, 3653–3673. doi:10.1093/mnras/stw2070

Kilbinger, M. (2015). Cosmology with Cosmic Shear Observations: a Review. *Rep. Prog. Phys.* 78, 086901. doi:10.1088/0034-4885/78/8/086901

Kingma, D. P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980

Knabenhans, M., Stadel, J., Marelli, S., Potter, D., Teyssier, R., Legrand, L., et al. (2019). Euclid Preparation: II. The EUCLIDEMULATOR—a Tool to Compute the Cosmology Dependence of the Nonlinear Matter Power Spectrum. *Monthly Notices R. Astronomical Soc.* 484, 5509–5529. doi:10.1093/mnras/stz197

Knabenhans, M., Stadel, J., Potter, D., Dakin, J., Hannestad, S., Tram, T., et al. (2020). Euclid Preparation: IX. EuclidEmulator2 – Power Spectrum Emulation with Massive Neutrinos and Self-Consistent Dark Energy Perturbations. arXiv e-prints, arXiv:2010.11288.

Krachmalnicoff, N., and Tomasi, M. (2019). Convolutional Neural Networks on the HEALPix Sphere: a Pixel-Based Algorithm and its Application to CMB Data Analysis. *A&A* 628, A129. doi:10.1051/0004-6361/201935211

Lahav, O., and Liddle, A. R. (2019). The Cosmological Parameters. arXiv e-prints, arXiv:1912.03687.

Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proc. Icml (Citeseer)* 30, 3. doi:10.21437/interspeech.2016-1230

McEwen, J. D., Wallis, C. G. R., and Mavor-Parker, A. N. (2021). Scattering Networks on the Sphere for Scalable and Rotationally Equivariant Spherical CNNs. arXiv e-prints, arXiv:2102.02828.

Mirza, M., and Osindero, S. (2014). Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784.

Miyato, T., and Koyama, M. (2018). "cGANs with Projection Discriminator," in International Conference on Learning Representations.

Mustafa, M., Bard, D., Bhimji, W., Al-Rfou, R., and Lukić, Z. (2017). Creating Virtual Universes Using Generative Adversarial Networks. arXiv preprint arXiv:1706.02390.

Nathanaël, P., Ankit, S., Kacprzak, T., Lucchi, A., Hofmann, T., and Réfrégier, A. (2019). Cosmological N-Body Simulations: A challenge for Scalable Generative Models. arXiv preprint arXiv:1908.05519.

Odena, A., Olah, C., and Shlens, J. (2017). "Conditional Image Synthesis with Auxiliary Classifier gans," in Proceedings of the 34th International Conference on Machine Learning, 70, 2642–2651.

Parroni, C., Cardone, V. F., Maoli, R., and Scaramella, R. (2020). Going Deep with Minkowski Functionals of Convergence Maps. *A&A* 633, A71. doi:10.1051/0004-6361/201935988

Perarnau, G., Van De Weijer, J., Raducanu, B., and Álvarez, J. M. (2016). Invertible Conditional gans for Image Editing. arXiv preprint arXiv:1611.06355.

Perraudin, N., Defferrard, M., Kacprzak, T., and Sgier, R. (2018). Deepsphere: Efficient Spherical Convolutional Neural Network with Healpix Sampling for Cosmological Applications. arXiv preprint arXiv:1810.12186.

Petri, A., Haiman, Z., Hui, L., May, M., and Kratochvil, J. M. (2013). Cosmology with Minkowski Functionals and Moments of the Weak Lensing Convergence Field. *Phys. Rev. D* 88, 123002. doi:10.1103/PhysRevD.88.123002

Petri, A., Liu, J., Haiman, Z., May, M., Hui, L., and Kratochvil, J. M. (2015). Emulating the CFHTLenS Weak Lensing Data: Cosmological Constraints from Moments and Minkowski Functionals. *Phys. Rev. D* 91, 103511. doi:10.1103/PhysRevD.91.103511

Petri, A. (2016). Mocking the Weak Lensing Universe: The LensTools Python Computing Package. *Astron. Comput.* 17, 73–79. doi:10.1016/j.ascom.2016.06.001

Pires, S., Starck, J.-L., Amara, A., Réfrégier, A., and Teyssier, R. (2009). Cosmological Model Discrimination with Weak Lensing. *A&A* 505, 969–979. doi:10.1051/0004-6361/200811459

Potter, D., Stadel, J., and Teyssier, R. (2017). PKDGRAV3: beyond Trillion Particle Cosmological Simulations for the Next Era of Galaxy Surveys. *Comput. Astrophys.* 4, 2. doi:10.1186/s40668-017-0021-1

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. (2016). Generative Adversarial Text to Image Synthesis. arXiv preprint arXiv:1605.05396.

Refregier, A. (2003). Weak Gravitational Lensing by Large-Scale Structure. *Annu. Rev. Astron. Astrophys.* 41, 645–668. doi:10.1146/annurev.astro.41.111302.102207

Rodriguez, A. C., Kacprzak, T., Lucchi, A., Amara, A., Sgier, R., Fluri, J., et al. (2018). Fast Cosmic Web Simulations with Generative Adversarial Networks. arXiv preprint arXiv:1801.09070.

Roth, K., Lucchi, A., Nowozin, S., and Hofmann, T. (2017). Stabilizing Training of Generative Adversarial Networks through Regularization. *Adv. Neural Inform. Process. Syst.* 13, 2018–2028. doi:10.21203/rs.2.22269/v1

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved Techniques for Training Gans. *Adv. Neural Inform. Process. Syst.* 4 2234–2242. doi:10.1117/12.2513139.6013937645001

Sato, M., Takada, M., Hamana, T., and Matsubara, T. (2011). Simulations of Wide-Field Weak-Lensing Surveys. II. Covariance Matrix of Real-Space Correlation Functions. *ApJ* 734, 76. doi:10.1088/0004-637X/734/2/76

Schmelzle, J., Lucchi, A., Kacprzak, T., Amara, A., Sgier, R., Réfrégier, A., et al. (2017). Cosmological Model Discrimination with Deep Learning. arXiv preprint arXiv:1707.05167.

Sgier, R. J., Réfrégier, A., Amara, A., and Nicola, A. (2019). Fast Generation of Covariance Matrices for Weak Lensing. *J. Cosmol. Astropart. Phys.* 2019, 044. doi:10.1088/1475-7516/2019/01/044

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., et al. (2005). Simulations of the Formation, Evolution and Clustering of Galaxies and Quasars. *Nature* 435, 629–636. doi:10.1038/nature03597

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the Inception Architecture for Computer Vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826.

Takada, M., and Jain, B. (2003). The Three-point Correlation Function in Cosmology. *MNRAS* 340, 580–608. doi:10.1046/j.1365-8711.2003.06321.x

Tamosiunas, A., Winther, H. A., Koyama, K., Bacon, D. J., Nichol, R. C., and Mawdsley, B. (2020). Investigating Cosmological GAN Emulators Using Latent Space Interpolation. arXiv e-prints, arXiv:2004.10223.

Taylor, A., Joachimi, B., and Kitching, T. (2013). Putting the Precision in Precision Cosmology: How Accurate Should Your Data Covariance Matrix Be?. *Month. Notices R. Astronomical Soc.* 432, 1928–1946. doi:10.1093/mnras/stt270

Tröster, T., Ferguson, C., Harnois-Déraps, J., and McCarthy, I. G. (2019). Painting with Baryons: Augmenting N-Body Simulations with Gas Using Deep Generative Models. *MNRAS* 487, L24–L29. doi:10.1093/mnrasl/slz075

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., et al. (2014). Scikit-Image: Image Processing in python. *PeerJ* 2, e453. doi:10.7717/peerj.453

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Spergel, D. N., Somerville, R. S., Dave, R., et al. (2020). The CAMELS Project: Cosmology and Astrophysics with MachinE Learning Simulations. arXiv e-prints, arXiv:2010.00619.

Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale Structural Similarity for Image Quality Assessment. *The Thrity-Seventh Asilomar Conference On Signals. Syst. Comput.* 2, 1398–1402. doi:10.1109/acssc.2003.1292181

Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., and Refregier, A. (2020). Cosmological Forecast for Non-gaussian Statistics in Large-Scale Weak Lensing Surveys. arXiv e-prints, arXiv:2006.12506.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Perraudin, Marcon, Lucchi and Kacprzak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX A. GENERATIVE ADVERSARIAL NETWORK ARCHITECTURE

Model **Table A1** summarizes the architecture of the GAN system, i.e. the generator and the discriminator. From the latent variable z and the cosmological parameters σ_8 and Ω_m, the generator starts by computing $\hat{z}$ using **Eq. 2**. As a second step, $\hat{z}$ is transformed with three linear layers. i.e. a Multi Layer Perceptron (MLP), that outputs a 32768 tensor (g_0 to g_2). The data is then reshaped to $8 \times 8 \times 512$ (h_3) and further transformed with four deconvolutional layers with stride 2 and kernel sizes of 3×3 or 5×5 (g_4 to g_7). The last generator layer consists of a deconvolution with stride 1 and kernel size 7×7 and it is intended to generate fine-grained details (g_8). The discriminator is symmetric to the generator with two exceptions. First the parameters σ_8 and Ω_m are concatenated in d_5 just before the first linear layer. Second, an extra linear layer is added at the end of the discriminator (d_9) in order to recover a single output. All layers are separated by a LeakyRelu activation function with the parameter $\alpha = 0.2$ (Maas et al., 2013).

Training The cosmological dataset described in **Section 3** is used to train the GAN, where the batches are composed of samples from different cosmologies. We select a Wasserstein loss, with a gradient penalty of 10 (Arjovsky et al., 2017). We use RMSProp as an optimizer with an initial learning rate of 10^{-5}, and a batch size of 64. The discriminator is updated 5 times more often than the generator. The model is trained for 10^{-5} epochs on a GeForce GTX 1080 GPU, which takes around 170 h.

APPENDIX TABLE A1 | Conditional GAN architecture.

Layer	Operation	Activation	Dimension
Generator			
$\hat{z}$	**Eq. 2**		$b \times 128$
g_0	Linear	Relu	$b \times 256$
g_1	Linear	Relu	$b \times 512$
g_2	Linear	Relu	$b \times 32768$
g_3	Reshape		$b \times 8 \times 8 \times 512$
g_4	Deconv ($k = 3 \times 3$, $s = 2$)	Relu	$b \times 16 \times 16 \times 256$
g_5	Deconv ($k = 5 \times 5$, $s = 2$)	Relu	$b \times 32 \times 32 \times 128$
g_6	Deconv ($k = 5 \times 5$, $s = 2$)	Relu	$b \times 64 \times 64 \times 64$
g_7	Deconv ($k = 5 \times 5$, $s = 2$)	Relu	$b \times 128 \times 128 \times 32$
g_8	Deconv ($k = 7 \times 7$, $s = 1$)	Relu	$b \times 128 \times 128 \times 1$
Discriminator			
X			$b \times 128 \times 128$
d_0	conv ($k = 7 \times 7$, $s = 1$)	LeakyRelu	$b \times 128 \times 128 \times 32$
d_1	conv ($k = 5 \times 5$, $s = 2$)	LeakyRelu	$b \times 64 \times 64 \times 64$
d_2	conv ($k = 5 \times 5$, $s = 2$)	LeakyRelu	$b \times 32 \times 32 \times 128$
d_3	conv ($k = 5 \times 5$, $s = 2$)	LeakyRelu	$b \times 16 \times 16 \times 256$
d_4	conv ($k = 3 \times 3$, $s = 2$)	LeakyRelu	$b \times 8 \times 8 \times 512$
d_5	Reshape + concatenate		$b \times 32770$
d_6	Linear	LeakyRelu	$b \times 512$
d_7	Linear	LeakyRelu	$b \times 256$
d_8	Linear	LeakyRelu	$b \times 128$
d_9	Linear	LeakyRelu	$b \times 1$

d_5 is a layer that reshapes the tensor to a vector and then concatenates the conditioning parameters to it. Here b is the batch size, k the convolutional kernel size and s the stride. The number of filters (convolution layer) and the number of neurons (linear layers) is shown in blue.

APPENDIX B. REGRESSOR TRAINING

Given real and generated images, the general idea of the Frechet Inception Distance (FID) is to compute the distance between some of their complex statistics. For natural images, these statistics are given by the last layer, i.e. the logits, of a pre-trained Inception-V3 network (Szegedy et al., 2016). As these statistics are meaningless for our cosmological data, we build new ones using a carefully designed regressor. Given an image, the regressor is trained to predict the two parameters Ω_m and σ_8. We provide the regressor weights with the code to make our FID metric reusable.

Data Naturally, we use the training dataset described in **Section 3**, i.e. 46 different cosmologies composed by 12000 images each. This training dataset is further randomly split into a regressor training set (80%) and a restressor test set (20%).

Model The architecture of the regressor is described in **Table A2**. It shares the same structure as the GAN discriminator. It consists of a four convolutional layers followed by three linear layers with leaky relu non-linearity. The last layer is a linear layer with two outputs and it is responsible for producing the predicted parameters. We select the LeakyRelu activation functions for better gradient propagation.

Training We use the mean squared error between the predicted and true parameters as a loss function. The model was trained for 20 epochs using an Adam (Kingma and Ba, 2014) optimizer with an initial learning rate of $3 \cdot 10^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$ and a batch size of 64. The mean squared error evaluated on the test set corresponds to $8.93e^{-5}$, which is low enough for the purpose of computing the FID.

APPENDIX TABLE A2 | Architecture of the regressor.

Layer	Operation	Activation	Dimension
X			$b \times 128 \times 128$
h_0	conv ($k = 7 \times 7$, $s = 1$)	LeakyRelu	$b \times 128 \times 128 \times 32$
h_1	conv ($k = 5 \times 5$, $s = 2$)	LeakyRelu	$b \times 64 \times 64 \times 64$
h_2	conv ($k = 5 \times 5$, $s = 2$)	LeakyRelu	$b \times 32 \times 32 \times 128$
h_3	conv ($k = 5 \times 5$, $s = 2$)	LeakyRelu	$b \times 16 \times 16 \times 256$
h_4	conv ($k = 3 \times 3$, $s = 2$)	LeakyRelu	$b \times 8 \times 8 \times 512$
h_5	Reshape		$b \times 32768$
h_6	Linear	LeakyRelu	$b \times 512$
h_7	Linear	LeakyRelu	$b \times 256$
h_8	Linear	LeakyRelu	$b \times 128$
h_9	Linear	Linear	$b \times 2$

Here b is the batch size, k the convolutional kernel size and s the stride. The number of filters (convolution layer) and the number of neurons (linear layers) is shown in blue. The LeakyRelu activation uses the parameter $\alpha = 0.2$.