# Formatting the Landscape: Spatial conditional GAN for varying population in satellite imagery

**Tomas Langer** *
Intuition Machines Inc

**Natalia Fedorova**
Max Planck Institute for Evolutionary Anthropology

**Ron Hagensieker**
osir.io

## Abstract

Climate change is expected to reshuffle the settlement landscape: forcing people in affected areas to migrate, to change their lifeways, and continuing to affect demographic change throughout the world. Changes to the geographic distribution of population will have dramatic impacts on land use and land cover and thus constitute one of the major challenges of planning for climate change scenarios. In this paper, we explore a generative model framework for generating satellite imagery conditional on gridded population distributions. We make additions to the existing ALAE [30] architecture, creating a spatially conditional version: SCALAE. This method allows us to explicitly disentangle population from the model's latent space and thus input custom population forecasts into the generated imagery. We postulate that such imagery could then be directly used for land cover and land use change estimation using existing frameworks, as well as for realistic visualisation of expected local change. We evaluate the model by comparing pixel and semantic reconstructions, as well as calculate the standard FID metric. The results suggest the model captures population distributions accurately and delivers a controllable method to generate realistic satellite imagery.

## 1 Introduction

Human beings are not actionless pawns in the face of climate change, they adapt to both direct and indirect pressures brought about by an increasingly unstable climate. People can either choose to leave problematic areas, migrating both locally and internationally to adapt to their circumstances, or they can stay and change their lifeways. In either case, and on top of expected demographic change, human adaptation to climate change thus reshuffles the settlement landscape [9]. As local populations ebb and flow, land use and land cover change in response. Due to this high mobility of populations, particularly given recent work on climate induced migration [32], one of the major challenges for planning for climate change scenarios is thinking about where people will be, and how this will change the landscape. State-of-the-art work on gridded population forecasts shows us the value of a greater geographic resolution and border oblivious approach [32, 20]. However, such forecasts still require analytic processing to evaluate their consequences for local landscapes.

In this paper, we explore the potential for generating satellite imagery conditional on population change as an in-between step for analysis, and as a means of directly visualizing forecasts in a realistic way. To do so, we employ the latest generative models from the field of unsupervised machine learning, namely generative adversarial networks [13]. GANs are a state-of-the-art technique for high resolution image generation, as they can generate images from random noise, also called the

---

*contact author: langer.tomas@yahoo.com

latent space. We refer the reader to a review article [39] for additional details. Generative models have been successfully applied to various high resolution datasets, mainly faces (e.g. CelebAHQ [23], FFHQ [21]), objects (e.g. Imagenet [8]), and scenes (e.g. LSUN [40], Cityscapes [7]). However, generative models in the high resolution earth observation domain are relatively under-explored [33].

On top of image generation, one might wish to edit a real image via a trained generative model. For this, a mapping from image space to latent space is required. This is a natural feature of autoencoders [25] or flow based generative models [24]. However, one can learn the mapping for a trained GAN model as well, which is called projection [1, 22]. In order to facilitate fast and accurate mapping, we decided to train a hybrid model based on Adversarial Latent AutoEncoder (ALAE) [30], which combines a GAN and an AE by training a generator and an encoder end-to-end.

Furthermore, to explicitly control the generated population, we add a conditional input to the generator (more information in appendix 5.1), where the input label is a pixel-level population map.

Our main contributions are:

- Adding spatial conditioning module to the ALAE architecture
- Training the model on satellite + population data
- Visualizing population in generated images
- Evaluating the generative model's performance in terms of quality and the population change effect

## 2 Methods

### 2.1 Data

Two data sets are utilized over a study site of continental Central America including Mexico. We focus on Central America here as it has been identified as a region already experiencing high levels of internal and international migration due to climate change, and has been the focus of prior modelling efforts [20]. However, this approach could of course be applied to any geographic region.

Image data is derived from surface reflectance data from ESA's Sentinel-2 mission [26]. Sentinel-2 is a constellation of two satellites, which collect images at a 5 day revisit. The second data set is the Global Human Settlement population data (GHS-POP) for the years 2000 and 2015 from the European Commission's Joint Research Centre [34, 11]. Both datasets are publicly available, and the details of how we sample the data are available in appendix 5.3.
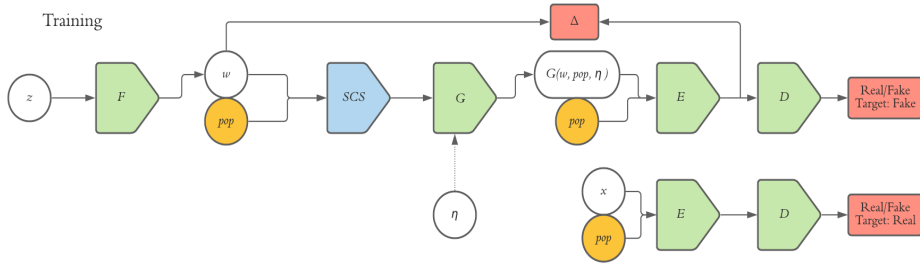
### 2.2 Model architecture



Figure 1: **SCALAE model training** where white circles represent original ALAE inputs, yellow circles are our additional inputs, green boxes are original ALAE modules, blue boxes are our modified modules, and red boxes are the losses.

We use ALAE [30] for the basis of our model, which is in turn based on the StyleGAN [21] architecture. We refer the reader to these papers for in depth description of the architecture, and focus here instead on our modifications. We adapt the ALAE codebase [31] for our training and evaluation, and our additional code and trained models can be accessed here [2].

---

[2]https://github.com/LendelTheGreat/SCALAE

Our training model differs from ALAE in 2 ways: the added *Spatial Conditional Style* (SCS) module, and the modified encoder input. Hence, the equation 2 from the ALAE paper can be modified to equation 1. All training modifications are also shown in figure 1.

$$\text{G} = G \circ SCS \circ F, \text{ and D} = D \circ E \tag{1}$$

The SCS module is used to feed conditional information (population map) to the generator. The style input *w* and appropriately resized population map *pop* are combined together by a learned function that maps both to the same number of channels, then summed together, and finally fed into the adaptive instance normalization layers of the generator. This process can be seen in figure 5 in the appendix 5.5.

The discriminator gets the conditional information (population map) as an input so that it provides a learning signal based on how the population map and the generated image fit together. This is done via a simple concatenation of the RGB and population channels. Because the ALAE discriminator is partly an encoder, the population map is concatenated with the RGB channels of either real or fake image and fed into the encoder to predict the style *w*, which is then fed into the discriminator head.

### 2.3 Reconstruction

After model training, we reconstruct a real satellite image by mapping it to the latent space of the model. This is trivial thanks to the autoencoder architecture, visualized in figure 6 in appendix 5.6.

Since the population map serves as an input to the generator, we can feed in a custom population map and control the population in the reconstructed output. This process can be used to visualize how real world places would look like with an alternative population distribution, for example, a climate change induced population change.

## 3 Results

### 3.1 Generation model quality

To evaluate quality of generations we calculate the Fréchet Inception Distance (FID) [17], more information in appendix 5.8. The overall model FID is $45.13$ (lower is better) for random generations. Our work would benefit from general benchmarking in the field, as at the moment, FID scores are only useful as a comparison across different models evaluated on the same data set. We thus report the FID score with the hopes of stimulating further comparison in the future.

In the absence of a direct comparison, we visually check the images, and as portrayed by figure 8 in appendix 5.10.1 we confirm that generations strongly resemble realistic satellite imagery with sufficient diversity.

### 3.2 Reconstruction quality

To further evaluate model quality we focus on reconstructing real satellite imagery. In this case, we give the model a real reference image as an additional input, which, as expected, improves the FID to $35.97$. Examples of reconstructions can be seen in figure 2 and more in appendix 9.

By generating reconstructions of real images, we create matching pairs that can be used for evaluation. We calculate the difference between the pairs of each real vs reconstructed image. The difference measure can be done on the pixel level, and on the semantic level using pretrained Inception features (same features as used in the FID score calculation). We use a standard l2 distance to compute the difference measures.

For the pixel level, the mean pixel l2 distance between pairs is $277.40$, but a frequency plot of the distribution of pixel distances for $9436$ image comparisons has a long tail, so much worse images in terms of pixel reconstruction are also possible, shown in figure 7a in appendix 5.9. The mean value of the semantic distance is $15.15$ and here a frequency plot of the distances is much more normally distributed, meaning better and worse images are equally likely (see figure 7b in appendix 5.9). The extreme tails of both of the above measure are visualized in figure 12 in appendix 5.10.5.

Figure 2: *left*: original input image, *center*: input population map, *right*: generated image

### 3.3  Population effect

To evaluate the effect of population on reconstructed images, and thus show the efficacy of our model (i.e. how well the population conditioning works), we produce generations for varying population inputs and visualize the pixel difference between them. In figure 3, we calculate the pixel difference averaged over 20 samples of style vectors $w$, showing the high consistency of the model reconstruction. Thus highlighting that the population conditioning is spatially consistent. An example of this process is visualized in figure 10 in appendix 5.10.3 and in figure 11 in appendix 5.10.4.
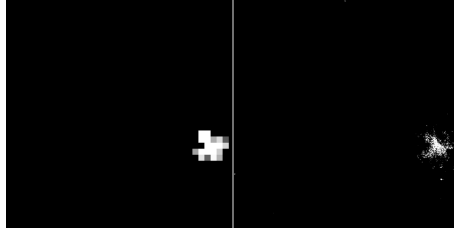


Figure 3: *left*: population input, *right*: pixel difference averaged over 20 styles

## 4  Discussion and Concluding remarks

We have created a model architecture that makes it possible to spatially condition style-based generative methods and thus to explicitly disentangle the latent space from a spatial label. We show that the population in the generated images can be manually controlled in a fine-grained manner, giving the user the ability to change population in specific parts of an image. Moreover, the encoder of our network can be used to map real images to the latent space, making it possible to not only edit fake, but also real, imagery. We believe this model could be useful for visualizing climate change related population forecasts such as those modelled in [20], as it allows practitioners and researchers to generate imagery flexibly, concretely, and with a means to characterize uncertainty.

Furthermore, the ability to map real images to the latent space opens up several image editing possibilities. We can continuously perform latent space arithmetic to create meaningful changes in the generated images, following previous GAN examples (e.g. [21, 22, 16]). Moreover, combining latent space arithmetic with explicit population conditioning delivers more control over exactly what is generated and where. Importantly, this can be done continuously, not just to generate a static outcome, but also to interpolate between or visualize a distribution of possible outcomes.

Likewise, it is difficult to evaluate the climate change effect on population on real imagery without reference longitudinal data. This will become more possible as longitudinal satellite data collections with matching population grids become available for longer time spans. Finally, the imagery we generate can be fed directly into existing frameworks for land use and land cover analysis, without further retraining or adaptation.

## Acknowledgments and Disclosure of Funding

## References

[1] Rameen Abdal, Yipeng Qin, and Peter Wonka. *Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?* 2019. arXiv: 1904.03189 [cs.CV].

[2] Rameen Abdal et al. *StyleFlow: Attribute-conditioned Exploration of StyleGAN-Generated Images using Conditional Continuous Normalizing Flows*. 2020. arXiv: 2008.02401 [cs.CV].

[3] Andrew Brock, Jeff Donahue, and Karen Simonyan. *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. 2019. arXiv: 1809.11096 [cs.LG].

[4] Ting Chen et al. *Self-Supervised GANs via Auxiliary Rotation Loss*. 2019. arXiv: 1811.11212 [cs.LG].

[5] Xi Chen et al. *InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets*. 2016. arXiv: 1606.03657 [cs.LG].

[6] Mehdi Cherti. "Deep generative neural networks for novelty generation : a foundational framework, metrics and experiments". PhD thesis. Jan. 2018.

[7] Marius Cordts et al. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. 2016. arXiv: 1604.01685 [cs.CV].

[8] J. Deng et al. "ImageNet: A Large-Scale Hierarchical Image Database". In: *CVPR09*. 2009.

[9] Catherine Devitt. "Climate Change and Population Displacement". In: September (2015), pp. 27–32.

[10] Adji B. Dieng et al. *Prescribed Generative Adversarial Networks*. 2019. arXiv: 1910.04302 [stat.ML].

[11] A.J. Florczyk et al. *GHSL Data Package 2019 - Technical report by the Joint Research Centre (JRC), European Union*. 2019, p. 38. ISBN: 9789276131861. DOI: 10.2760/0726. URL: https://ghsl.jrc.ec.europa.eu/documents/GHSL%7B%5C_%7DData%7B%5C_%7DPackage%7B%5C_%7D2019.pdf?t=1478q532234372.

[12] Robert Geirhos et al. *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness*. 2019. arXiv: 1811.12231 [cs.CV].

[13] Ian J. Goodfellow et al. "Generative adversarial nets". In: *Advances in Neural Information Processing Systems* 3.January (2014), pp. 2672–2680. ISSN: 10495258. arXiv: arXiv:1406.2661v1.

[14] Noel Gorelick et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone". In: *Remote sensing of Environment* 202 (2017), pp. 18–27.

[15] Shuyang Gu et al. *GIQA: Generated Image Quality Assessment*. 2020. arXiv: 2003.08932 [eess.IV].

[16] Erik Härkönen et al. *GANSpace: Discovering Interpretable GAN Controls*. 2020. arXiv: 2004.02546 [cs.CV].

[17] Martin Heusel et al. *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. 2018. arXiv: 1706.08500 [cs.LG].

[18] Xun Huang and Serge Belongie. *Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization*. 2017. arXiv: 1703.06868 [cs.CV].

[19] Phillip Isola et al. *Image-to-Image Translation with Conditional Adversarial Networks*. 2018. arXiv: 1611.07004 [cs.CV].

[20] Bryan Jones. *Modeling Climate Change-Induced Migration in Central America & Mexico Methodological Report*. Tech. rep. ProPublica, 2020.

[21] Tero Karras, Samuli Laine, and Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Networks*. 2019. arXiv: 1812.04948 [cs.NE].

[22] Tero Karras et al. *Analyzing and Improving the Image Quality of StyleGAN*. 2020. arXiv: `1912.04958 [cs.CV]`.

[23] Tero Karras et al. *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. 2018. arXiv: `1710.10196 [cs.NE]`.

[24] Diederik P. Kingma and Prafulla Dhariwal. *Glow: Generative Flow with Invertible 1x1 Convolutions*. 2018. arXiv: `1807.03039 [stat.ML]`.

[25] Diederik P. Kingma and Max Welling. "Auto-encoding variational bayes". In: *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* Ml (2014), pp. 1–14. arXiv: `1312.6114`.

[26] Jérôme Louis et al. "Sentinel-2 Sen2Cor: L2A processor for users". In: *Proceedings Living Planet Symposium 2016*. Spacebooks Online. 2016, pp. 1–8.

[27] Mehdi Mirza and Simon Osindero. "Conditional Generative Adversarial Nets". In: (2014), pp. 1–7. arXiv: `1411.1784`. URL: `http://arxiv.org/abs/1411.1784`.

[28] Cedric Oeldorf and Gerasimos Spanakis. "LoGANv2: Conditional style-based logo generation with generative adversarial networks". In: *Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019* (2019), pp. 462–468. DOI: `10.1109/ICMLA.2019.00086`. arXiv: `1909.09974`.

[29] Taesung Park et al. "Semantic image synthesis with spatially-adaptive normalization". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019-June (2019), pp. 2332–2341. ISSN: 10636919. DOI: `10.1109/CVPR.2019.00244`. arXiv: `1903.07291`.

[30] Stanislav Pidhorskyi, Donald A. Adjeroh, and Gianfranco Doretto. "Adversarial Latent Autoencoders". In: C (2020), pp. 14092–14101. DOI: `10.1109/cvpr42600.2020.01411`. arXiv: `2004.04467`.

[31] Stanislav Pidhorskyi, Donald Adjeroh, and Gianfranco Doretto. *Adversarial Latent Autoencoders*. 2020. URL: `https://github.com/podgorskiy/ALAE`.

[32] Kanta Kumari Rigaud et al. "Groundswell: Preparing for internal climate migration". In: (2018). URL: `https://openknowledge.worldbank.org/handle/10986/29461`.

[33] David Rolnick et al. "Tackling Climate Change with Machine Learning". In: (2019). arXiv: `1906.05433`. URL: `http://arxiv.org/abs/1906.05433`.

[34] Marcello Schiavina, Sergio Freire, and Kytt MacManus. "GHS population grid multitemporal (1975, 1990, 2000, 2015) R2019A". In: *Eur. Comm. JRC* (2019).

[35] Yujun Shen et al. *InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs*. 2020. arXiv: `2005.09635 [cs.CV]`.

[36] Akash Srivastava et al. *VEEGAN: Reducing Mode Collapse in GANs using Implicit Variational Learning*. 2017. arXiv: `1705.07761 [stat.ML]`.

[37] Christian Szegedy et al. *Going Deeper with Convolutions*. 2014. arXiv: `1409.4842 [cs.CV]`.

[38] Ting-Chun Wang et al. *High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs*. 2018. arXiv: `1711.11585 [cs.CV]`.

[39] Zhengwei Wang, Qi She, and Tomas E. Ward. *Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy*. 2020. arXiv: `1906.01529 [cs.LG]`.

[40] Fisher Yu et al. "LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop". In: *arXiv preprint arXiv:1506.03365* (2015).

[41] Ning Yu et al. *Inclusive GAN: Improving Data and Minority Coverage in Generative Models*. 2020. arXiv: `2004.03355 [cs.CV]`.

[42] Jun-Yan Zhu et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. 2020. arXiv: `1703.10593 [cs.CV]`.

[43] Peihao Zhu et al. "SEAN: Image Synthesis With Semantic Region-Adaptive Normalization". In: (2020), pp. 5103–5112. DOI: `10.1109/cvpr42600.2020.00515`. arXiv: `1911.12861`.

## Broader Impact

We envision that our research here could be of benefit for both local and international organizations who are committed to integrating AI methodologies into their planning and policy workflows. However, given the complexity of the models and the infrastructure and financing required for

training, there is an obvious gap in who is able to actually use these models leading to concerns of centralization.

In terms of biases, we identified several sources of bias in our method, that could lead to undesired outcomes, and should be explicitly taken into account before any direct application of this method. First of all, there is the model bias of GANs themselves. GANs are know to suffer from mode dropping, which could results in uncommon features of the dataset being ignored by the generative model, even if they are presents in the data. This bias can be approximately quantified by measuring diversity of generations using recent methods [15] and visualized, which means it can be evaluated in relation to particular use cases. Moreover, recent methods have made substantial improvements to mitigate this bias [36, 10, 41]. Note that our method SCALAE is a hybrid that includes an autoencoder in the latent space, which partly reduces the mode dropping problem, however, proper evaluation of this phenomena is ongoing research and is left for future work.

Secondly, there are several sources of bias in the data itself. On one hand, it is the data collection bias. For example, because of cloud cover, satellite data collection usually focuses on the dry season only. On the other hand, there is some concern for leveraging path dependency biases, coming from the fact that the generated images can only reflect patterns that were observed before. It thus cannot capture new developmental trajectories. Nonetheless, there is interesting new research being done in the direction of novelty generation [6].

Finally, we do not anticipate, or recommend, that a generative approach be used in isolation for policy and planning - it is a tool to aid field professionals and ideally should be linked directly with theoretically and locally informed behavioral models, in fact, we postulate that the conditioning approach we develop here makes this imminently possible.

# 5 Appendices

## 5.1 Additional background on conditional GANs

In addition to image generation, GANs and other methods have shown promising results on controlling the generation process. This can be done when the latent space is sufficiently disentangled, in which case, adjusting a section of the latent space results in a meaningful semantic change in the generated image. A well disentangled latent space can be learned in a completely unsupervised fashion [5], however, providing an explicit signal to the model, if available, results in more precise control over the generated image [4]. Therefore, our method focuses on so called conditional image generation. Conditioning, even though it requires additional labels, helps disentanglement and allows explicit content generation.

The difference between unconditional and conditional image generation is that the former generates realistic images from random noise, whereas the latter generates realistic images from a given input label. This label can come in many forms, usually image-level [27, 3, 28] or pixel-level [19, 29, 43] attributes, or even images from a different domain [38, 42]. In our case, the input label is a pixel-level population map.

## 5.2 Related works

A popular section of generative models relevant to our problem are the domain transfer models [19, 38, 42]. These methods produce a function that translate images between 2 or more domains (e.g. horse to zebra, day to night, etc). In comparison with our method, they do not allow for sampling random images. Thus, our method has the advantage of being used outside of its current use case without retraining, an important point when we consider the financial and environmental costs. However, most domain transfer methods contain direct skip connections between the reference real image and the generated output, which is known to reproduce better pixel-wise reconstructions.

Another side of work focuses on discovering meaningful latent space directions after the model was fully trained [16, 35, 2]. This is complimentary to our model and the exploration of these is left for future work.

## 5.3 Data: Sampling information

Sampling sites for Sentinel-2 imagery were determined based on largest population increase in this time period, and a set of $9436$ tiles with extents of $1024$ by $1024$ pixels was extracted. The selected sites are shown in appendix 5.4. Surface reflectance data of the dry season (Jan-March) was cloud-masked and averaged utilizing Google Earth Engine [14]. Finally, GHS-POP data was reprojected into the corresponding UTM-zones of the Sentinel-2 tiles. Given the illustration purposes of this study we have not included the non-visible bands of Sentinel-2, and focus on the RGB channels only, as those are easily interpretable to the human eye. However, our method is invariant to the number of channels used and could be trivially retrained with the full multi-spectral depth of Sentinel-2.

## 5.4 Sites for image acquisition
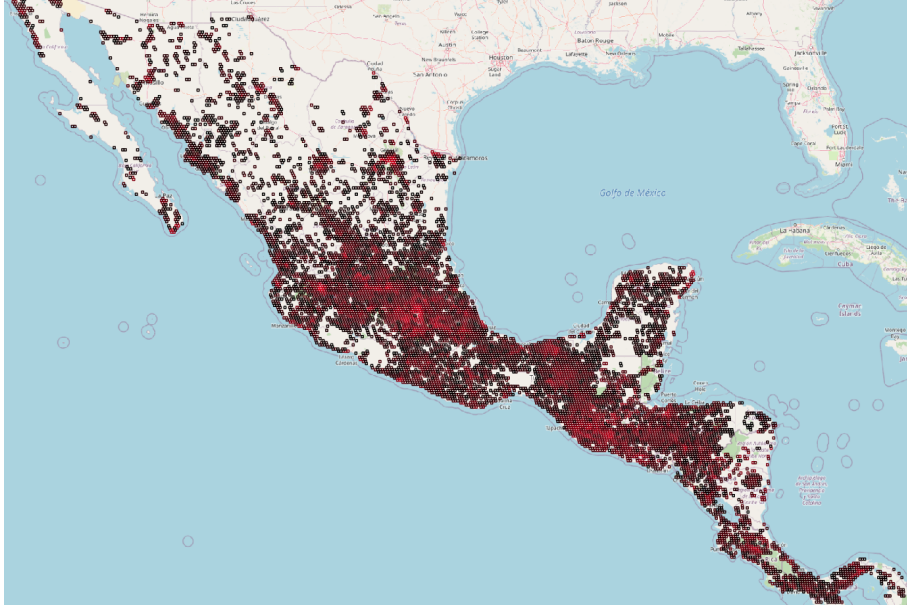


Figure 4: **Sites for image acquisition** Image acquisition based on population growth between 2000 and 2015; brighter color indicates higher population growth.
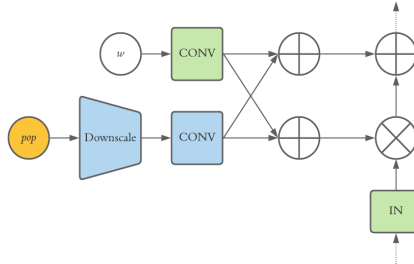
## 5.5 SCS module



Figure 5: **SCS module** To the best of our knowledge, adding spatial conditioning to a StyleGAN-like architecture has not been explored before. [28] successfully add a non-spatial class conditioning into a StyleGAN by feeding a class label $c$ together with noise $z$ into the F network and subsequently into the G network. This approach however, does not take spatial dimensions into account. We take inspiration from [29, 43] as they use the same AdaIN [18] layers as StyleGAN, and introduce the SCS module.
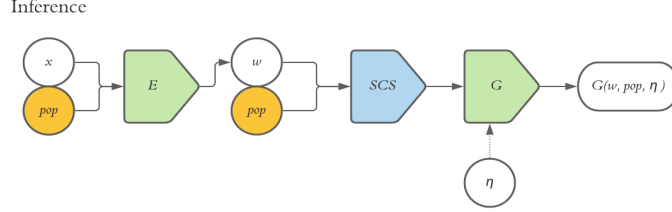
## 5.6 SCALAE reconstruction method



Figure 6: **SCALAE reconstruction method** where white circles represent original ALAE inputs, yellow circles are our additional inputs, green boxes are original ALAE modules, and blue boxes are our modified modules.

## 5.7 Training details

We train the SCALAE model end-to-end on the paired Sentinel-2 and GHS-POP data sets on 4xV100 GPUs, following the default training parameters from the ALAE codebase [31]. The Sentinel-2 imagery RGB channels are scaled between -1 and 1. The population map is log transformed and likewise scaled between -1 and 1. We train with progressive growing for 200 epochs, switching to higher resolution every 16 epochs. Our base learning rate is 0.002 and batch size 512, both adjusted accordingly with the default progressive growing schedule. The training losses remain unchanged.

Our code and all training parameters will be publicly released upon publication.

## 5.8 FID details

The FID score uses features from an Inception network [37] trained on the Imagenet [8] data set, which consists of natural photos of various objects. However, it does not include any images from satellite or other earth observation domain. We note that the domain shift in this case is not problematic because Imagenet-trained networks mostly focus on textures [12], which are the main features we are trying to quantify.
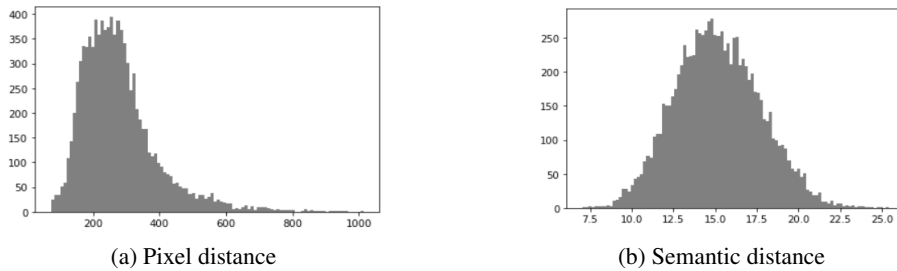
## 5.9 Reconstruction: distance histograms



(a) Pixel distance

(b) Semantic distance

Figure 7: **Histograms of distances**

9

## 5.10 Additional visualisations

### 5.10.1 Random generations



Figure 8: **Random generated samples**
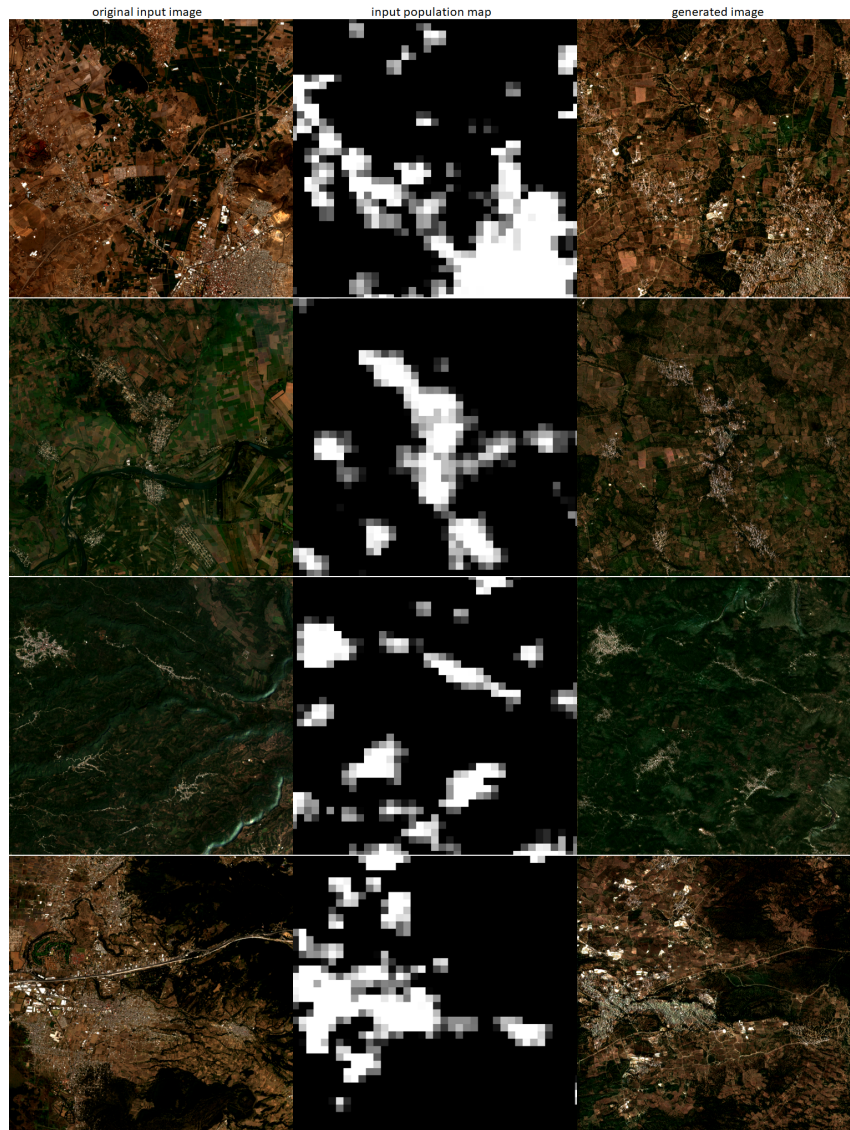
### 5.10.2 Reconstructions



Figure 9: **Targeted reconstructions** Additional examples
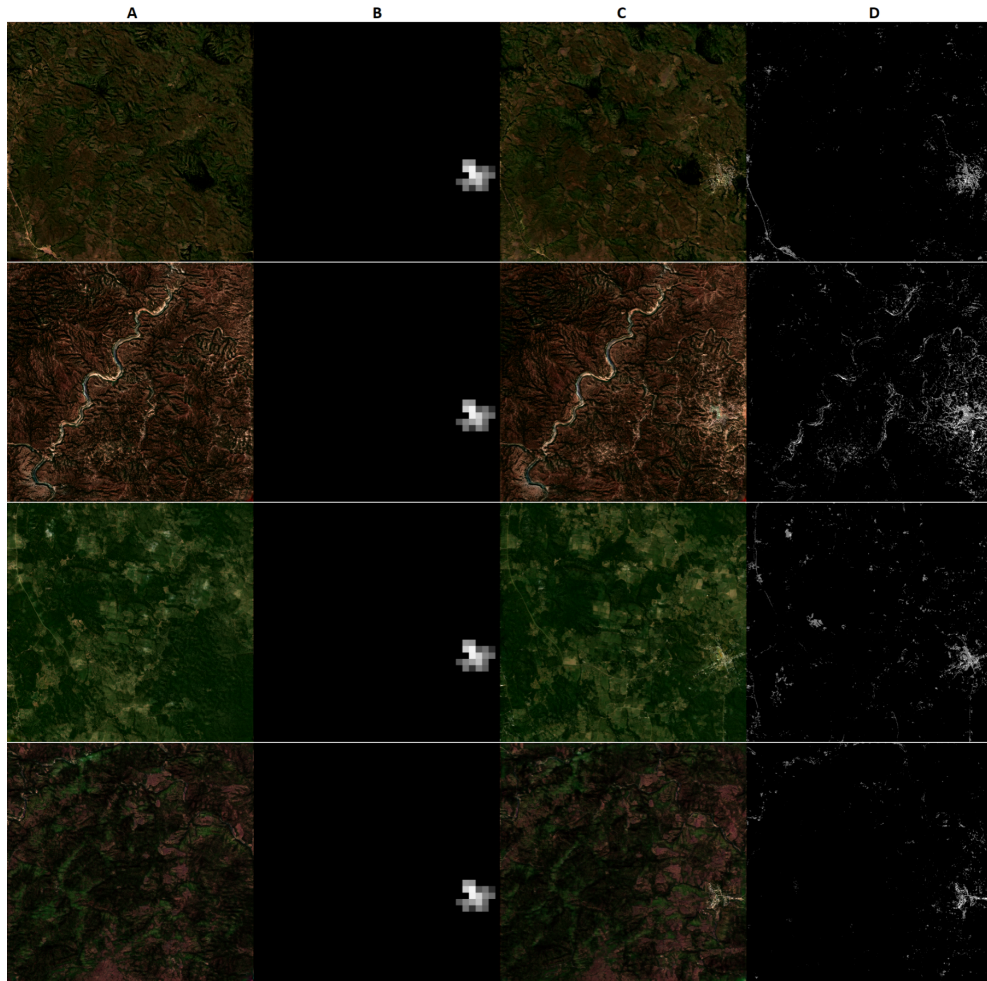
### 5.10.3 Population pixel difference



Figure 10: **Population pixel difference** Panel D contains the pixel difference between the generated images (panels A and C) and clearly reproduces the input population (panel B). Still, speckles of small pixel differences appear throughout the image. This shows that changing the population has some impact globally, suggesting some level of entanglement between the population map and the latent style *w*.
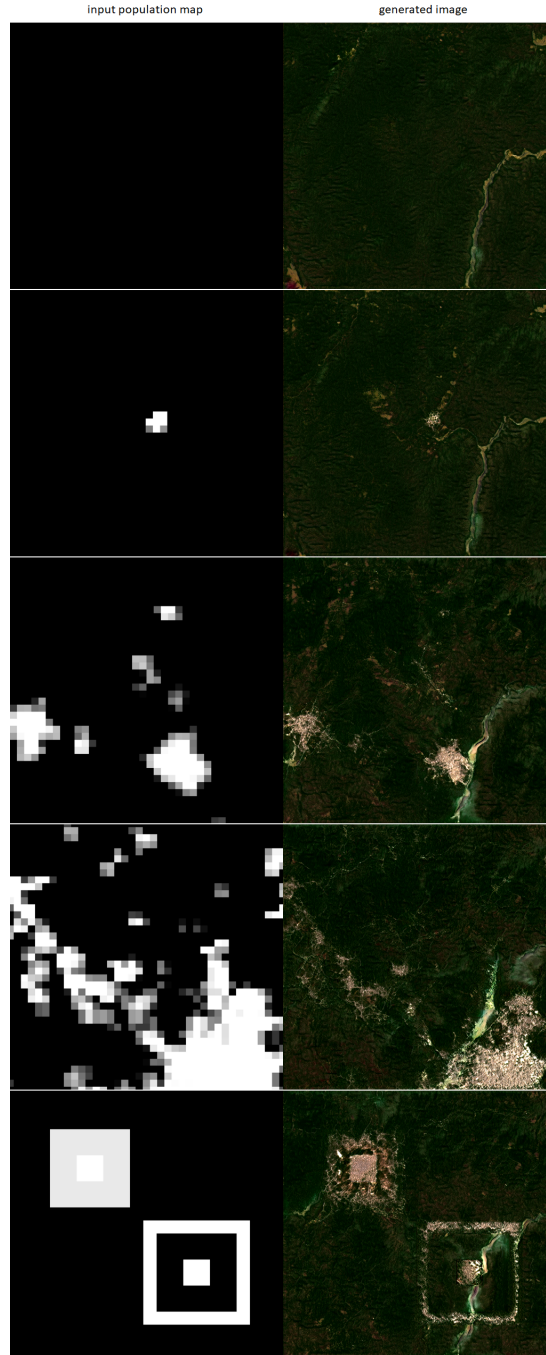
### 5.10.4 Population manipulation



Figure 11: **Custom population input** We fix the latent style input *w* and change population input. The population is accurately generated, even with unrealistic population distribution in the last row.
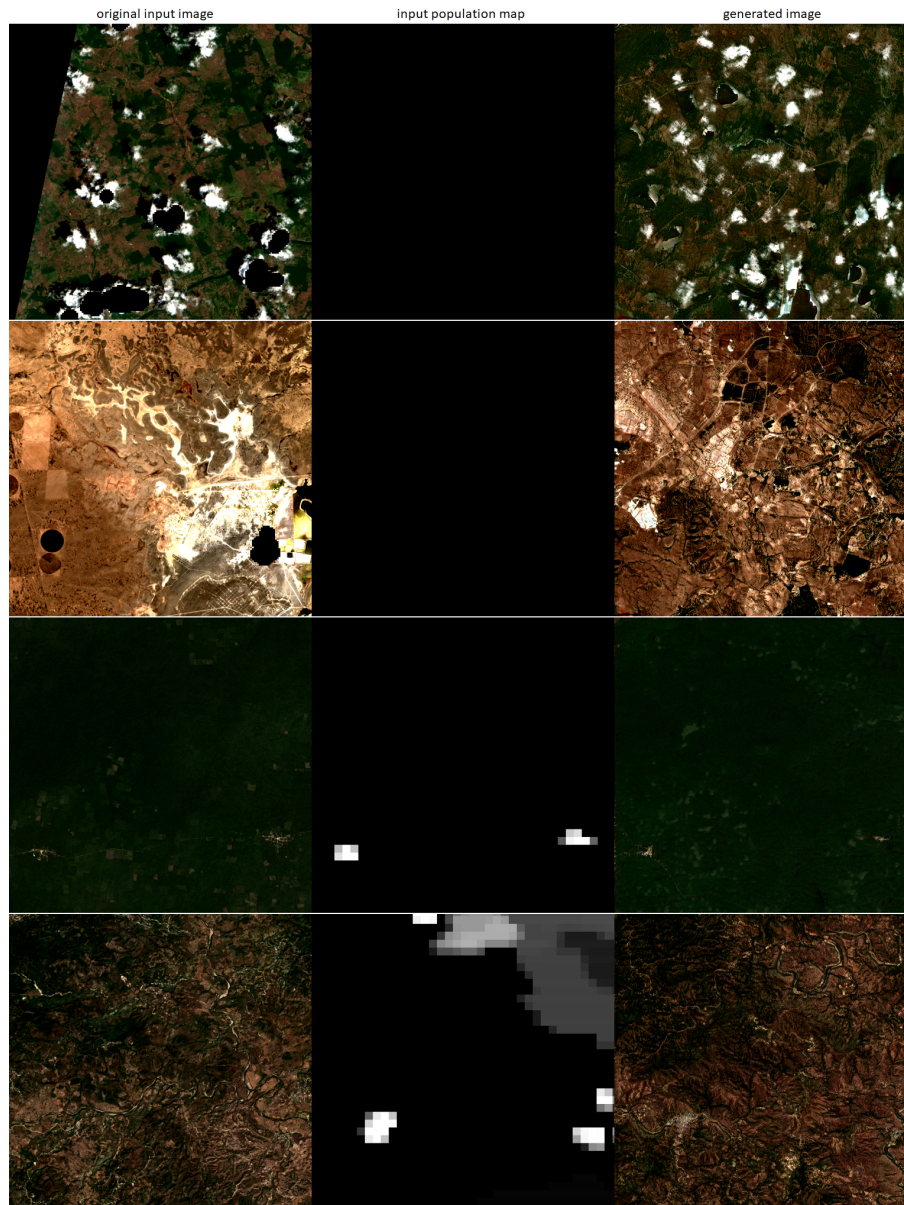
### 5.10.5 Best and worst reconstructions



Figure 12: **Best and worst generations** From top to bottom: semantic worst, pixel worst, pixel best, semantic best