# Auxiliary-task learning for geographic data with autoregressive embeddings

Konstantin Klemmer
University of Warwick & New York University
UK
k.klemmer@warwick.ac.uk

Daniel B. Neill
New York University
USA
daniel.neill@nyu.edu

## ABSTRACT

Machine learning is gaining popularity in a broad range of areas working with geographic data. Here, data often exhibit spatial effects, which can be difficult to learn for neural networks. We propose SXL, a method for embedding information on the autoregressive nature of spatial data directly into the learning process using auxiliary tasks. We utilize the local Moran's I, a measure of local spatial autocorrelation, to "nudge" the model to learn the direction and magnitude of local spatial effects, complementing learning of the primary task. We further introduce a novel expansion of Moran's I to multiple resolutions, capturing spatial interactions over longer and shorter distances simultaneously. The novel multi-resolution Moran's I can be constructed easily and offers seamless integration into existing machine learning frameworks. Over a range of experiments using real-world data, we highlight how our method consistently improves the training of neural networks in unsupervised and supervised learning tasks. In generative spatial modeling experiments, we propose a novel loss for auxiliary task GANs utilizing task uncertainty weights. SXL outperforms domain-specific spatial interpolation benchmarks, highlighting its potential for downstream applications.

## CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; *Image representations*; • **Applied computing** → *Earth and atmospheric sciences.*

## KEYWORDS

Auxiliary Task Learning, Spatial Autocorrelation, GAN, Spatial Interpolation, GIS

**ACM Reference Format:**
Konstantin Klemmer and Daniel B. Neill. 2021. Auxiliary-task learning for geographic data with autoregressive embeddings. In *29th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '21), November 2–5, 2021, Beijing, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3474717.3483922


*SIGSPATIAL '21, November 2–5, 2021, Beijing, China*

ACM ISBN 978-1-4503-8664-7/21/11…$15.00
https://doi.org/10.1145/3474717.3483922

## 1 INTRODUCTION

When monitoring the physical environment, the gathered data are often geographic in nature and follow some spatial process: data values depend on their spatial locations. This violates a key assumption of many statistical learning frameworks, that data are independent and identically distributed (*iid*). The complexities of geographic data concern researchers in many academic fields and are a key focus of the geographic information sciences (GIS). The GIS community has a long tradition of analyzing spatial phenomena, developing metrics to measure spatial effects and deploying models to account for spatial dependencies. With the growing popularity of deep neural networks, applications to geospatial data domains have become more and more common. Nevertheless, these applications have only rarely inspired methodological innovation in neural networks. Recently, [15] specifically pointed to a lack of deep learning methods tailored to geospatial and spatio-temporal data in the context of earth system science. In light of this call-to-action, we propose `SXL`, a novel neural network method for geospatial data domains, that is inspired by domain expertise from GIS and explicitly learns spatial dependencies contained within the data. This is facilitated through a multi-task learning process, where a spatial embedding capturing local autoregressive effects at each data point is learned as an auxiliary task. We integrate one of the most prominent and widely used metrics in GIS into the model: the Moran's I measure of local spatial autocorrelation [1]. To also account for longer-distance spatial relations, we propose a novel multi-resolution local Moran's I by gradually coarsening the input data. We evaluate `SXL` on both generative and predictive spatial modeling tasks, providing empirical evidence for consistent and robust performance gains across multiple synthetic and real-world experimental settings.

## 2 RELATED WORK

Ideas from GIS and spatial statistics have inspired popular approaches in modern machine learning, from Gaussian Processes (GPs) to spatial scan statistics. Nevertheless, since the emergence of the era of deep neural networks, the relationship between the GIS and machine learning communities has been mostly defined through applications of existing neural network models to geographic data. Recent advances in machine learning that have been particularly useful to the GIS community include scalable GPs [6] and graph neural networks [11] for geospatial data.

But core concepts from GIS and geography are also gradually attracting more attention in neural networks research: For example, [12] propose vector embeddings to capture spatial context. While there exist approaches for capturing spatial autocorrelation in neural networks [19], the Moran's I metric has previously only been

used once in an explicit machine learning setting, as a stopping criterion for training generative adversarial networks [9].

In this paper, we use auxiliary learning, an approach using multi-task learning to improve performance on a primary task, originally conceptualized by [16]. The authors propose to give learners "hints" related to the original task throughout training in order to improve training speed and model performance. This can be understood as forcing the learner (e.g., a neural network) to focus its attention on certain patterns in the data, highlighted by the auxiliary objective. Auxiliary learning has been particularly successful in deep reinforcement learning [5] and has also been deployed for generative adversarial nets (GANs) [13]. However, to the best of our knowledge, measures of spatial autocorrelation such as Moran's I have never previously been used in any kind of multi-task learning setting, either for generative or predictive spatial modeling.

## 3 METHODOLOGY

### 3.1 Multi-resolution local Moran's I

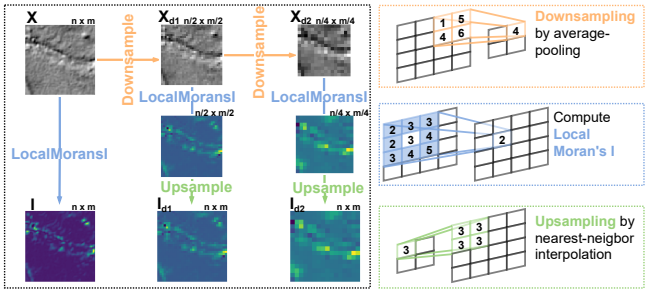


**Figure 1: Multi-resolution local Moran's I calculation with an example input at three different resolutions: Original input size ($n \times m$), downsampled by factor 2 and downsampled by factor 4.**

Working with geospatial data requires a careful assessment of and accounting for potential autoregressive effects–an intuition which neural networks traditionally do not provide. One of the most prominent measures, capturing spatial autocorrelation at the point-level, is the local Moran's I metric [1]. Local Moran's I measures the direction and extent of similarity between each observation and its local spatial neighbourhood. As such, it provides an indication for both local spatial clusters and spatial outliers.

Formally, let $X^{n \times m}$ be a 2-$d$ spatial matrix (image) and the vector $\mathbf{x} = vec(X)$ consists of $n_x = nm$ real-valued observations $x_i$, referenced by an index set $N_x = \{1, 2, ..., n_x\}$. We define the *spatial neighbourhood* of observation $i$ to be $\mathcal{N}_{x_i} = \{j \in N_x : w_{i,j} > 0\}$. Here, $w_{i,j}$ corresponds to a binary spatial weight matrix, indicating whether any observation $j$ is a neighbor of $i$. Throughout this study, we utilize queen contiguity (i.e. all adjacent grid cells, including diagonals, are neighbors), but the approach generalizes to arbitrary neighborhood definitions. We can compute the local Moran's I statistic, $I_i$, of observation $x_i$ with the mean over all observations $\bar{x}$ as:

$$I_i = (n_x - 1)\frac{x_i - \bar{x}}{\sum_{j=1}^{n_x}(x_j - \bar{x})^2} \sum_{j=1, j \neq i}^{n_x} w_{i,j}(x_j - \bar{x}) \tag{1}$$

$I_i$ can take positive or negative values: a positive value suggests that a data point is similar to its neighbors, which could indicate latent cluster structure. A negative value suggests that the data point is distinctly different from neighboring data points, which could indicate a changepoint or edge.

One of the main limitations of the local Moran's I metric is scale sensitivity [18]: its reliance on immediate neighbors can cause longer-range dependencies to be lost. Here, we propose a novel, multi-resolution representation of the local Moran's I by increasingly coarsening the input data for the Moran's I computation and then upsampling the output back to the original data shape. The coarsening step here is analogous to a 2-$d$ average pooling operation. The coarsened Moran's I is then upsampled again to the original input size $n \times m$ using nearest-neighbor interpolation. This whole process can be repeated several times to compute the local Moran's I at increasingly coarse resolutions. The local Moran's I values at different resolutions can then be stacked on top of one another, much like a multi-channel image (e.g. RGB image). As such, tensors provide an ideal data structure for our metric. We illustrate this with an example in Figure 1.

### 3.2 Auxiliary learning of spatial autoregressive structures

Auxiliary task learning shares the benefits of multi-task learning: auxiliary tasks hint at specific patterns in the data for the model to focus attention on. Further, they introduce a representation bias, whereas the model prefers latent representations of the data that work for both primary and auxiliary tasks, thus helping with generalization. Lastly, auxiliary tasks can work as regularizers by introducing inductive bias and decreasing the risk of overfitting the model. Here, we want to use the local Moran's I embedding as auxiliary tasks. The main motivation for any auxiliary tasks is "relatedness" to the primary task: spatial theory characterizes a spatial pattern as a reflection of underlying spatial processes. Accordingly, [3] concludes that "[...] the capability of generalizing and quantifying spatial patterns is a prerequisite to understanding the complicated processes governing the distribution of spatial phenomena."–explicitly mentioning the power of the Moran's I metric to capture these effects. Recent research further highlights the importance of learning at multiple resolutions to support a comprehensive understanding of spatial processes [17].

With our experiments, we focus on two distinct settings: generative spatial modeling using GANs [7], and predictive spatial modeling in the form of spatial interpolation. To outline the application of our proposed auxiliary task approach, we introduce the GAN example in detail–the predictive modeling formulation follows from this directly. GANs are comprised of two neural networks, a Generator **G** that produces fake data and a Discriminator **D** that seeks to distinguish between real and fake data. These two networks are agents in a two-player game, where **G** learns to produce synthetic data samples that are faithful to the true data generating process, and **D** learns to separate real from fake samples. The standard GAN loss function is thus given as:

$$\min_G \max_D \mathcal{L}_{GAN}(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})}\big[\log D(\mathbf{x})\big] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})}\big[\log(1 - D(G(\mathbf{z})))\big], \tag{2}$$

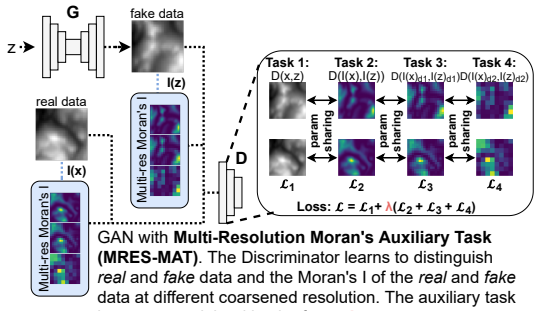


**Figure 2: GAN with spatially explicit auxiliary tasks using multi-resolution Moran's I.**

We propose two options to integrate our auxiliary tasks:

*(1)* Adding $N$ auxiliary tasks (single- or multi-resolution) with losses $\mathcal{L}^{(D)}_{AT_i}$ using fixed loss weights $\lambda$ (see Figure 2):

$$\min_G \max_D \mathcal{L}^{\lambda}_{MRES-MAT}(D,G) = \mathcal{L}_{GAN}(D,G) + \lambda(\mathcal{L}^{(D)}_{AT_1} + \cdots + \mathcal{L}^{(D)}_{AT_N}). \tag{3}$$

*(2)* Automatically learning the weights of main and auxiliary loss using each task's homoskedastic uncertainty to inform the weight. We expand the loss proposed by [4] to the GAN setting:

$$\min_G \max_D \mathcal{L}^{UW}_{MRES-MAT}(D,G) = \mathcal{L}^{(G)} + \left(\frac{1}{2\sigma_1^2}\mathcal{L}^{(D)}_{MT} + \frac{1}{2\sigma_2^2}\mathcal{L}^{(D)}_{AT_1} + \cdots + \frac{1}{2\sigma^2_{N+1}}\mathcal{L}^{(D)}_{AT_N} + \sum_{i=1}^{N+1}\log\sigma_i\right), \tag{4}$$

where $\sigma_1, \ldots, \sigma_{N+1}$ are the model noise parameters, $\mathcal{L}^{(G)}$ is the generator loss and $\mathcal{L}^{(D)}_{MT}$ the main task discriminator loss. This constitutes the first adaption of the uncertainty task balancing principles to the multi-task GAN family. The adaptation of *SXL* to predictive spatial models follows intuitively (analogous to the "prediction" task of the GAN discriminator).

## 4 EXPERIMENTS

| Model | Toy | PetrelGrid | DEM | TreeCanopy |
|---|---|---|---|---|
| GAN [7] | 0.0934 | 0.4106 | **0.1120** | 0.1138 |
| GAN-MAT UW | 0.1077 | 0.4860 | 0.1814 | 0.1132 |
| GAN-MRES-MAT UW | **0.0917** | **0.4014** | 0.1180 | **0.1038** |
| DCGAN [14] | 0.1534 | 0.2993 | 0.0591 | 0.0654 |
| DCGAN-MAT UW | 0.2319 | 0.3049 | **0.0591** | 0.1009 |
| DCGAN-MRES-MAT UW | **0.0938** | **0.2793** | 0.0612 | **0.0635** |
| EDGAN [20] | 0.0269 | **0.2909** | 0.0499 | 0.0322 |
| EDGAN-MAT UW | 0.0276 | 0.3061 | 0.0481 | 0.0316 |
| EDGAN-MRES-MAT UW | **0.0241** | 0.2971 | **0.0469** | **0.0314** |

**Table 1: Test MMD scores (lower is better) of different GAN configurations with uncertainty weights. We compare synthetic samples from these generators to held-out test data to compute the scores.**

*Generative modeling.* In our first experiment, we want to examine whether our proposed method can improve the learning of spatial data generating processes with GANs. To assess model quality, we use the Maximum Mean Discrepancy (MMD) metric [2]. A lower MMD score between samples of real and synthetic data indicates higher quality of the synthetic samples. As GAN training is notoriously difficult, we train ten cycles of each tested GAN using the best generator (based on validation scores) from each cycle to compute test scores. We select four datasets for our experiments: *(1)* toy dataset of a Gaussian peak mirroring a Gaussian dip, *(2)* a *PetrelGrid* seabed relief dataset [10], *(3)* a Digital Elevation Model (*DEM*) dataset obtained via the `elevatr` *R* package, and *(4) Tree canopy* data from the "Global Tree Change" project [8]. These datasets are chosen to represent a range of different geospatial patterns occurring in real-world physical environments and relate to important modeling challenges in the earth sciences, ecology, or geography. The modularity of our proposed auxiliary task learning method allows us to test it on a range of different GAN models. We chose the original GAN implementation, denoted here as *GAN* [7], the *DCGAN* [14] and lastly an Encoder-Decoder GAN (*EDGAN*) architecture recently proposed by [20] and explicitly designed for geospatial applications. Table 1 shows results from our experiments. We can see that the auxiliary task settings improve performance for most experiments, agnostic of the underlying GAN architecture, by usually 3-10%. However, the optimal loss weight setting for applying the auxiliary task appears to vary. The **MRES MAT UW** strategy seems to be the safest bet, as it does not require further, manual weight parameter tweaking.

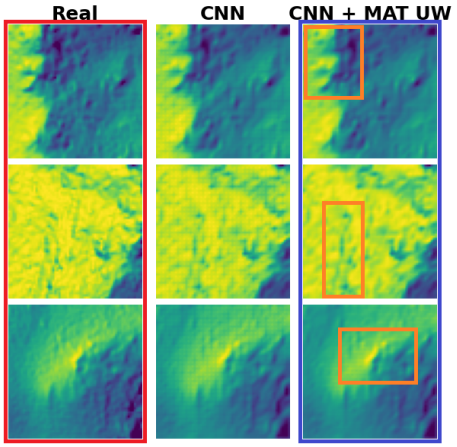


**Figure 3: Interpolation results on samples from the test set, across the different benchmark models, presenting our CNN + MAT UW model. The orange boxes highlight areas where the improvement over the benchmark models becomes visually apparent.**

*Predictive spatial modeling.* We now tackle spatial interpolation, obtaining high-resolution spatial data from a low-resolution input. It is a regression task and can be evaluated using the root mean squared error (RMSE) between real and predicted high-resolution output. We again train 10 models per strategy and compare their performance when no model selection is used (final model used for prediction on test set) and when model selection on a validation set is applied, saving the 10 best models, one from each run. We use hillshades of DEM data from the National Ecological Observatory Network (NEOS), a common use case in geography, for the

| Model / Task | RMSE 32 → 64 (no model selection) | (model selection) |
|---|---|---|
| BicInt | 0.0667 | – |
| IDW | 0.0693 | – |
| OK | 0.0801 | – |
| UK | 0.0796 | – |
| CNN | 0.0678(±0.0128) | 0.0503(±0.0008) |
| CNN + MAT UW | **0.0649(±0.0119)** | **0.0496(±0.0006)** |
| CNN + MAT $\lambda = 0.1$ | 0.0665(±0.0118) | 0.0516(±0.0018) |
| CNN + MAT $\lambda = 0.01$ | 0.0666(±0.0178) | 0.0532(±0.0033) |

**Table 2: Model RMSE scores and their standard deviation (over 10 runs) on held-out test data.**

interpolation task. Spatial statistics provides a range of tools to tackle interpolation problems as benchmarks: (1) Bicubic interpolation (*BicInt*), (2) Inverse Distance Weighting (*IDW*), (3) Ordinary Kriging (*OK*), and (4) Universal Kriging (*UK*). We compare these established methods to a simple CNN implementation with two hidden layers (5). The modularity of SXL allows us to simply plug-in our **MAT**. In this setting, we do not use the **MRES MAT**, as our experiments show that further coarsening the already-reduced image is counterproductive. The CNN model main tasks are optimised using MSE loss, while the auxiliary tasks use $\ell 1$ loss. The results of this experiment are presented in Table 2 and Figure 3. We can again see a positive effect of the **MAT** on the performance of the CNN model—outperforming all other benchmarks. If no model selection is deployed, both hard loss weights and uncertainty weights produce models that outperform the naively trained CNN. **MAT UW** models provide the best average performance increase, of around 5%. If model selection is utilized, the **MAT UW** strategy outperforms the naive CNN by about 1.5%. Both of these performance increases are statistically significant, according to a paired t-test of the mean prediction scores. As compared to the use of fixed loss weights, the **MAT UW** strategy prevails in all interpolation experiments, whether model selection is applied or not.

## 5 CONCLUSION

With SXL we propose the use of single- and multi-resolution measures of local spatial autocorrelation for improving the learning of geospatial processes. We introduce a novel, flexible multi-resolution version of the local Moran's I statistic using coarsened inputs. We demonstrate its integration as an auxiliary task into generative and predictive neural network models, using both hard (static) and task uncertainty (automatically learned) loss weights. We empirically show robust, consistent and significant performance gains of up to 10% for generative spatial modeling and up to 5% for predictive spatial modeling when using this strategy.

## ACKNOWLEDGMENTS

This work is supported in part by the UK Engineering and Physical Sciences Research Council under Grant No. EP/L016400/1.

## REFERENCES

[1] Luc Anselin. 1995. Local Indicators of Spatial Association—LISA. *Geographical Analysis* 27, 2 (sep 1995), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x arXiv:1011.1669

[2] Karsten M. Borgwardt, Arthur Gretton, Malte J. Rasch, Hans Peter Kriegel, Bernhard Schölkopf, and Alex J. Smola. 2006. Integrating structured biological data by Kernel Maximum Mean Discrepancy. In *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btl242

[3] Yue Hong Chou. 1995. Spatial pattern and spatial autocorrelation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 988. Springer Verlag, 365–376. https://doi.org/10.1007/3-540-60392-1_24

[4] Roberto Cipolla, Yarin Gal, and Alex Kendall. 2018. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00781 arXiv:1705.07115

[5] Yannis Flet-Berliac and Philippe Preux. 2019. MERL: Multi-Head Reinforcement Learning. (sep 2019). arXiv:1909.11939 http://arxiv.org/abs/1909.11939

[6] Jacob R. Gardner, Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. 2018. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1809.11165 http://papers.nips.cc/paper/7985-gpytorch-blackbox-matrix-matrix-gaussian-process-inference-with-gpu-accelerationhttp://arxiv.org/abs/1809.11165

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems (NeurIPS)*. 2672–2680. http://papers.nips.cc/paper/5423-generative-adversarial-nets

[8] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R.G. Townshend. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342, 6160 (nov 2013), 850–853. https://doi.org/10.1126/science.1244693

[9] Konstantin Klemmer, Adriano Koshiyama, and Sebastian Flennerhag. 2019. Augmenting correlation structures in spatial data using deep generative models. *arXiv:1905.09796* (2019). arXiv:1905.09796 http://arxiv.org/abs/1905.09796

[10] J. Li. 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. In *Proceedings - 20th International Congress on Modelling and Simulation, MODSIM 2013*. https://doi.org/10.36334/modsim.2013.a9.li

[11] Mike Li, Elija Perrier, and Chang Xu. 2019. Deep Hierarchical Graph Convolution for Election Prediction from Geospatial Census Data. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (jul 2019), 647–654. https://doi.org/10.1609/aaai.v33i01.3301647

[12] Gengchen Mai, Krzysztof Janowicz, Bo Yan, Rui Zhu, Ling Cai, and Ni Lao. 2020. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells. In *International Conference on Learning Representations (ICLR)*. arXiv:2003.00824 http://arxiv.org/abs/2003.00824

[13] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2017. Conditional image synthesis with auxiliary classifier gans. In *34th International Conference on Machine Learning, ICML 2017*, Vol. 6. 4043–4055. arXiv:1610.09585 http://proceedings.mlr.press/v70/odena17a.html

[14] Alec Radford, Luke Metz, and Soumith Chintala. 2016. Unsupervised representation learning with deep convolutional generative adversarial networks. In *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*. arXiv:1511.06434

[15] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 7743 (feb 2019), 195–204. https://doi.org/10.1038/s41586-019-0912-1

[16] S. C. Suddarth and Y. L. Kergosien. 1990. Rule-injection hints as a means of improving network performance and learning time. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 412 LNCS. Springer Verlag, 120–129. https://doi.org/10.1007/3-540-52255-7_33

[17] Jianpeng Xu, Xi Liu, Tyler Wilson, Pang Ning Tan, Pouyan Hatami, and Lifeng Luo. 2018. Muscat: Multi-scale spatio-temporal learning with application to climate modeling. In *IJCAI International Joint Conference on Artificial Intelligence*. https://doi.org/10.24963/ijcai.2018/404

[18] Na Zhang and Hongyan Zhang. 2011. Scale variance analysis coupled with Moran's I scalogram to identify hierarchy and characteristic scale. *International Journal of Geographical Information Science* 25, 9 (sep 2011), 1525–1543. https://doi.org/10.1080/13658816.2010.532134

[19] Yunchao Zhang, Yanjie Fu, Pengyang Wang, Xiaolin Li, and Yu Zheng. 2019. Unifying inter-region autocorrelation and intra-region structures for spatial embedding via collective adversarial learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3292500.3330972

[20] Di Zhu, Ximeng Cheng, Fan Zhang, Xin Yao, Yong Gao, and Yu Liu. 2019. Spatial interpolation using conditional generative adversarial neural networks. *International Journal of Geographical Information Science* (apr 2019), 1–24. https://doi.org/10.1080/1365881YYxxxxxxxx