Building the World - Could AI build our Synthetic Environments?

Graham Long

Thales UK

Crawley, West Sussex

Graham.Long@uk.thalesgroup.com

ABSTRACT

Constructing Synthetic Environments that satisfy the content, fidelity, functional and performance requirements and expectations of a widening range of applications, users and client systems, in a cost effective and timely manner, challenges todays Synthetic Environment (SE) production capabilities.

Rapidly expanding data availability, increasing processing and rendering performance, a heightened awareness and familiarity with digital, virtual environments in everyday life, sets an expectation and demand that Synthetic Environments will provide high quality, content rich, authentic representations of the real world.

Synthetic Environment production pipelines have evolved to incorporate a range of processes and activities that reflect the progressive convergence of traditional Modelling & Simulation (M&S) with geospatial, gaming and other relevant domains. But despite employing efficiency enhancing techniques, such as procedural content generation, these production pipelines typically still require significant effort and time to accomplish the task.

This paper will explore how emerging Artificial Intelligence (AI), machine learning (ML) and deep learning (DL) could be applied to a typical SE production pipeline, based on the recently approved Simulation Interoperability Standards Organization (SISO) Reuse and Interoperation of Environmental Data and Processes (RIEDP) standard.

Beyond M&S, AI is already being applied to the types of task and activity that crossover into SE production – geospatial data processing, construction of gaming assets and environments, computer vision, autonomous vehicles, behavioral modelling. The paper will identify potential AI techniques and technologies that may be applicable to each stage of the production process, consider the feasibility of applying AI across the entire pipeline, from design through to build of the complete SE, and the opportunities AI can offer to SE generation and SE users.

ABOUT THE AUTHORS

Graham Long has worked in the simulation industry for over 30 years. Starting his career as a synthetic environment developer, he has managed engineering teams engaged in the deployment of image generator and display systems, synthetic environment production and the development of synthetic environment generation and authoring tools, delivering solutions to over 20 civil and military flight simulation projects. His responsibilities at Training Solutions, Thales (UK), incorporate synthetic environment development, generation technologies, solutions, tools and interoperability standards.

Building the World - Could AI build our Synthetic Environments?

Graham Long

Thales UK

Crawley, West Sussex

Graham.Long@uk.thalesgroup.com

INTRODUCTION

Synthetic environments (SE) provide an artificial representation of the real, physical world with which simulation applications, systems and users interact. They can incorporate aspects of the natural environment, including terrain relief, natural and man-made cultural features, mobile entities (humans, animals, vehicles, weapons), ocean, weather and atmospheric models. Features will be encoded with the supplementary information (e.g. texture, material properties, topology) required to support synthetic environment client systems such as multi-spectral visualization (visible, infra-red, Night Vision), Computer Generated Forces (CGF), radar, or digital maps.

A deluge of data is now available from which to construct these environments. The proliferation of collection platforms, from satellites to drones and location enabled, connected devices is driving an exponential data growth. This goes beyond traditional geospatial sources, such as satellite and aerial imagery, to include point cloud data and even social media. The emergence of volunteered geographic information (VGI), like OpenStreetMap, has added another dimension of open data. In general, data has never been more abundant, affordable and available in a wide variety of forms, at high standards of quality and resolution.

At the same time, improvements in computing performance, big data and cloud technologies provide systems with the capability to access, process and render this data as a highly detailed, feature-rich virtual environment.

These environments must now support the needs of an increasingly complex, dynamic and uncertain operational environment, in a Live, Virtual and Constructive (LVC) context, with interoperable air and ground platforms conducting missions in a shared or common synthetic environment.

Simultaneously exploiting data, system performance and satisfying the multi-domain and interoperability requirements of a shared environment poses significant challenges for synthetic environment production. Not only to meet these requirements but to accomplish this within acceptable timeframes at an affordable cost.

There is no universal SE ground truth against which the efficiency of the pipeline can be measured. Despite the convergence toward a more holistic, common environment, requirements and content tend to remain domain and application specific. For example, driving simulation, ground troop or military transport aircraft operations all have different primary requirements for how the same city should be represented to fulfil their principal mission. The priorities of each domain also impose different demands on the construction process. Generating a highly detailed city with building apertures and interiors for ground operations presents a different construction problem to that of building a global environment with 1000's of airports and worldwide representative 3D content. But despite these differences, it is likely that environment production timeframes across all of these domains will be measured in 100's to 1000's of hours.

SYNTHETIC ENVIRONMENT GENERATION PIPELINE

Synthetic Environment requirements and content will vary and necessitate specific production tasks to be conducted, but fundamental generation pipelines and processes are similar. Figure 1 summarizes a generalized production pipeline based on the Simulation Interoperability Standards Organization (SISO), Reuse and Interoperation of Environmental Data and Processes (RIEDP) Reference Process Model (RPM). This defines the major steps of a typical pipeline, from requirements, through data collection, content creation, data processing (cleaning, aligning, intensification, specialization) up to point the run-time version is generated for the target system. The particular tools, methods and processes employed within each vendor-specific pipeline will vary, but the major production phases will be very similar.

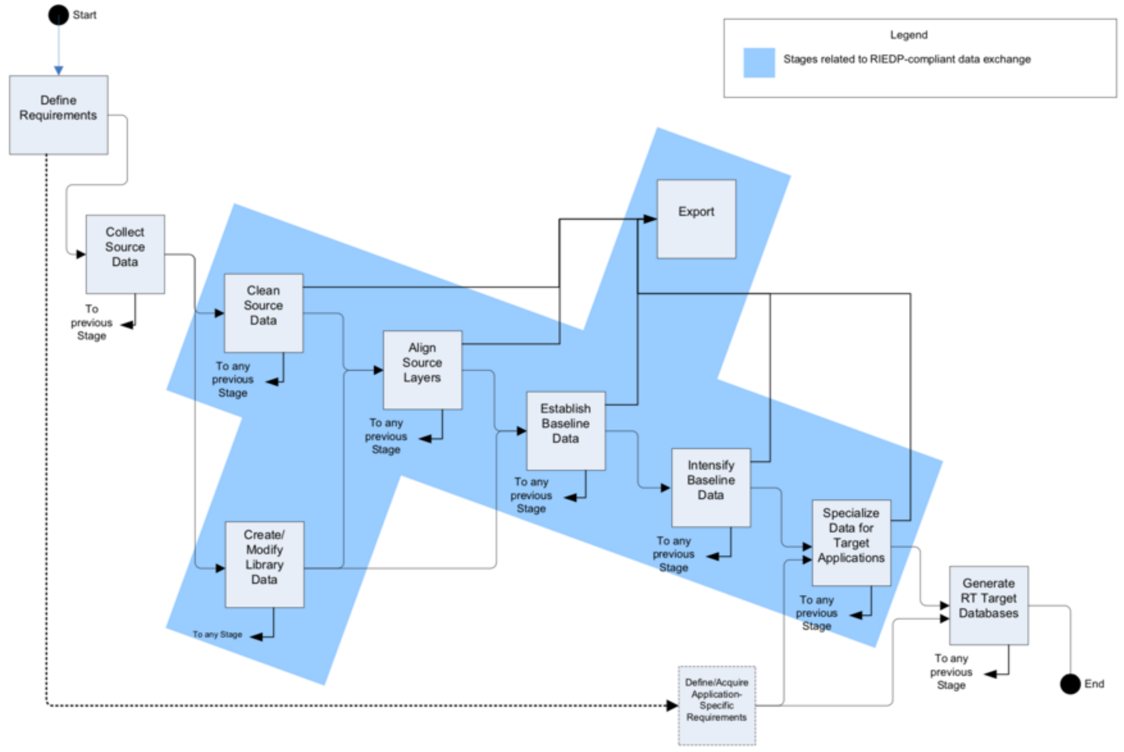

Figure 1. RIEDP Reference Process Model (RPM)

80% of the effort to construct a Synthetic Environment is devoted to activities concerned with data collection, library content creation and data production phases (highlighted in blue in Figure 1). As the fidelity and complexity of both the data and the synthetic environments increases, pipelines must adapt and improve. This is necessary to ensure they can meet evolving requirements, at affordable cost, in acceptable timeframes and take advantage of the opportunities presented by data and technology evolution.

These challenges are not unique to Modelling & Simulation (M&S). The gaming industry is engaged in a similar struggle to control its content creation costs and timescales, as the capabilities of its systems and expectations of its customers increase. The explosion in geospatial data is driving innovation in the fields of earth observation and remote sensing to exploit this data. AI is beginning to make inroads into these and many other sectors to help address some of these challenges but has very made little penetration into M&S.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI) is transforming many fields, from earth observation to finance, healthcare, security, gaming and entertainment - and has the potential to bring a similarly disruptive change to M&S and SE's. Many definitions of AI exist, but in essence it refers to the capability of a machine to imitate intelligent human behavior.

From the inception of the term in 1955, "artificial intelligence" has been subject to periods of advancement and stagnation, constrained by lack of available data and computing power. From the 1990's, Central Processing Unit (CPU) performance improvements and web driven data proliferation accelerated AI development and gave rise to

Machine Learning. Machine Learning is a sub-category of AI that learns by utilizing algorithms to parse data, learn from that data, and then apply what they've learned to make informed decisions (Zendesk 2017).

Deep Learning is the next evolutionary step of AI. A subset of Machine Learning, Deep Learning employs Artificial Neural Networks that are computational models inspired by the human brain. A neural network has a large number of processors operating in parallel, arranged in an input layer, several hidden layers and an output layer. Each layer is made up of many highly interconnected nodes; each connection has a weight, which adjusts the signal between layers. Learning is achieved by transmitting data through the layers. A process of backpropagation then adjusts the network weights until the error of its prediction is reduced to the point that the prediction is acceptable.

Neural networks are effective because they are incredibly adaptive and learn very quickly. Unlike traditional machine learning algorithms, neural networks are designed to learn, grow and improve with more data and more usage. But because input data is essential to how neural networks learn, they require a lot of training data. Their architecture is also computationally expensive.

2012 was a major milestone in Deep Learning advancement when Alex Krizhevsky, Sutskever and Hinton applied the parallelization of GPU's to deep learning processing to develop the "AlexNet" model and win the 2012 ImageNet challenge. (Large Scale Visual Recognition Challenge) (Russakovsky, Deng, Su, et al. 2015). Since then, the adoption and large scale use of GPU's for most deep learning research, and the appearance in 2016 of dedicated processing chips, such as Google's Tensor Processing Unit (TPU), have been a critical driver of AI progress. An OpenAI study (OpenAI 2018) concluded that the amount of compute used to train AI models has been increasing exponentially with a 3.5 month doubling time since 2012. Now cloud computing and large data processing centers are making this computing power more accessible and available.

Deep Learning requires large datasets for the training of its models. The web and new data capture devices have proliferated data, but the emergence of specialized, open datasets, such as ImageNet (Russakovsky et al. 2015) with over 10 million tagged images, has stimulated rapid development of Deep Learning models. There are many other training datasets (CityScapes, ShapeNet, MNIST, COCO). Many are available to support open competitions to create the best deep learning models on platforms like Kaggle.

The creation and training of deep learning models has been greatly facilitated by the appearance of open source software frameworks, typically available through GitHub and developed with Python. Among the most popular are TensorFlow (originally developed the Google Brain group), Keras and PyTorch. Cloud providers, such as Amazon, Microsoft, Google and IBM also now offer Artificial Intelligence services through the Cloud that can be integrated with internal applications.

The AI research culture is very open, dynamic and fast moving, with many researchers publishing results in databases such as Cornell University's arXiv, while many also make their associated software available as open source.

Transfer learning is a powerful deep learning technique that focuses on applying knowledge gained while solving one problem by leveraging pre-trained models and adapting them to solve a different but related problem. An existing model, trained on a very large training dataset (like ImageNet) can be adapted to a new problem, and re-trained with a much smaller dataset. For example, a model trained to identify images of cars can be re-trained to identify trucks. This is possible because most layers in an object recognition neural network are performing core object recognition tasks, and only the final output layer needs to be adapted to "learn" the new task. It can be likened to standing on the shoulders of giants – new models do not need to be developed from scratch, they capitalize on the work already done to build a large training dataset, and the time to train the model. Computer Vision and Natural Language Processing are two areas that lend themselves to transfer learning techniques, and many of their deep learning models are openly shared for re-use (VGG-16, VGG-19, Word2Vec, GloVe).

The availability of parallel computing, open training datasets and source frameworks, an open research culture, re-usable models and transfer learning have combined to accelerate deep learning since 2012 and make it a much more accessible and feasible technology to exploit and develop. It does not alleviate the need for Deep Learning expertise and knowledge but does help accelerate the development process.

APPLICATION OF AI TO THE SE PIPELINE

There are many types of artificial neural network architectures (at least 29, according to www.asimovinstitute.org) that operate in different ways to achieve different outcomes. Mainly due to the prominent role of imagery in SE production, and the nature of the tasks to be performed, Convolution Neural Networks (CNN's) and Generative Adversarial Networks (GAN's) are two of the most common architectures employed in the deep learning solutions described in the following sections.

CNN's are well suited to the field of computer vision, performing image classification, object detection, object instance detection and image segmentation.

Proposed only 5 years ago, GAN's are perhaps the hottest topic in deep learning. Their uniqueness and strength derive from using two competing neural networks, a generator and discriminator that effectively learn from each other. The generator network tries to create an output, a fake picture of a dog, for example, the discriminator tries to determine if the output is real or fake. The networks compete against each other to achieve high levels of competence in their respective tasks. GAN's have achieved impressive results when applied to aspects of "computational creativity" - painting like Picasso, composing music or text. The adversarial, self-learning nature of GAN's also reduces their need for large training datasets.

The following sections will use the RIEDP RPM phases as the basis for exploring where and how AI can or has the potential to be applied to the SE pipeline. AI is a vast, wide-ranging technology, that is moving very fast and offers many possibilities, the examples within each RPM phase indicate just some of these possibilities, but there are certainly many more.

Collect Source Data

Obtaining suitable source data is fundamental to synthetic environment construction. It involves research, data search and discovery of repositories, catalogues, authoritative data providers, data retrieval, evaluation, cataloguing, and classification.

Finding and retrieving suitable images is a common data task, at initial data collection and during production. The unstructured nature of images makes them difficult data to search. Semantic metadata can be added as labels and tags to describe and classify image content and enable searches on this metadata. But such searches can be compromised by inaccurate, inconsistent or missing metadata. A more intuitive alternative is to apply a Deep Learning visual search that uses images themselves as the query to direct a search and find similar images.

Visual search is in use from e-commerce to medicine, and extensively by the likes of Google, Amazon, Bing, and Pinterest ("A lot of the future of search is going to be about pictures instead of keywords," Ben Silbermann, Pinterest CEO). Visual search solutions now enable consumers to find products of interest simply by using an image as their search criteria, for example, providing an image of a dress to find similar or identical products in a clothing vendors product catalogue.

The specific neural network architectures of these systems will vary but are based on the image classification and object detection capabilities of CNN's. For example, Pinterest (Zhai et al. 2017) has provided several visual search and recommendation services, (Related Pins, Similar Looks, Flashlight and Lens). With billions of images and hundreds of millions of users, they have explored network architectures to optimize performance using existing models, such as AlexNet, as well evaluating Faster R-CNN and Single Shot Detection (SDD) to improve object detection.

The web provides the biggest data repository available. Providing a sample image of a building with a particular architectural style to a Google, Bing or TinEye image search, is fast and has a high chance of success. But visual search can be developed and applied at an enterprise level too (Ameisen 2018). It offers a mechanism to obtain more value from asset repositories that may not be very metadata rich and to search original source imagery as well as texture libraries.

Visual image search can also be used to search for a specific instance of an object, such as the Eiffel Tower, and retrieve all other instances of that object from a catalogue of reference images. To encourage the development of models that can provide this capability, Google has recently launched the Landmark Recognition and Retrieval 2019 challenges (Google 2019a) with a new Google Landmark Boxes training dataset of 86k images (Figure 2).

Figure 2. Google Landmark Boxes dataset

Create/Modify Library Data

Environments require content - buildings, vegetation, airports, vehicles, animals, human characters. Models now contain thousands of polygons, multiple texture maps, sophisticated articulation and motion. As the complexity of content has increased, so has the task of modelling the 3D objects, designing the textures and rigging the animations.

3D Models

Creating 3D models that provide detailed and realistic representations of vehicles, buildings or characters is a creatively skilled, largely manual process, performed using 3D authoring tools. Automating 3D model generation brings time and cost benefits and reduces the dependency on skilled 3D modelling resources. Neural Network architectures offer two potential approaches to automated generation of 3D models from just a single input image.

One approach utilizes a Convolutional Neural Network (CNN) to generate geometry represented as a voxel grid from an input image. Because of the cubic growth of the volume with increasing resolution, voxel grids have been limited to a coarse 32^3 resolution, which does not capture model surfaces well. Using higher resolutions very quickly becomes computationally infeasible. Häne, Tulsiani & Malik (2017) have proposed a Hierarchical Surface Prediction (Figure 3) to improve the model to a higher resolution of 256^3 voxels.

Figure 3. Input image, left, voxel model, right.

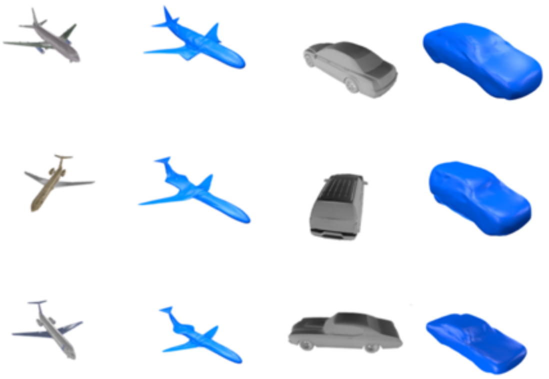

Figure 4. Generated mesh model (blue)

Alternatively, a graph convolutional neural (GCN) network can be used as the basis to generate a 3D shape as a triangle mesh from a color image (see Figure 4). GEOMetrics (Smith, Fujimoto, Romero, Meger, 2019) uses a CNN for feature extraction from the image, GCN's for mesh deformation, then applies an adaptive face splitting procedure to enhance local complexity when necessary.

Even at the higher resolution, the voxel-based solution still has many shape anomalies; it also requires the voxels to be converted to a mesh for them to be used in authoring tools and renderers. In these examples, the mesh-based solution produces better results, but it too still has shape anomalies. While offering promise as a method of generating 3D models from minimal input, the current quality and fidelity of these particular neural net approaches is unacceptably low for Synthetic Environment applications.

Texture generation

Exploiting the multi-texturing and Physics Based Rendering (PBR) capabilities of modern GPU's requires creation of a surface color (diffuse) map, as well as specular, normal, height, roughness, metallic, ambient occlusion, transparency, light and detail maps. As with 3D modelling, the process is creatively skilled, and is performed by a texture artist using image processing and texture creation tools.

There are several possibilities for applying neural networks to the texture creation process, one of which is to use a neural network to generate entirely synthesized images. GANs have made remarkable progress in image-to-image translation and style transfer solutions that can use different types of input image to generate a synthesized image.

For example, using a relatively simple semantically labelled image of a subject, in this case, a building image labelled with the areas of the doors, windows and walls (see Figure 5), a GAN can learn to generate a synthesized photo-realistic image of the building containing the required features, in the desired style (Isola, Zhu, Zhou, Efros, 2018). This process can be applied in reverse, to generate a labelled map from an image of a façade, (Harley, Wie, Saragih, Fragkiadaki, 2019). The segmentation of doors, windows, walls in the label map could then be applied to support sensor simulation.

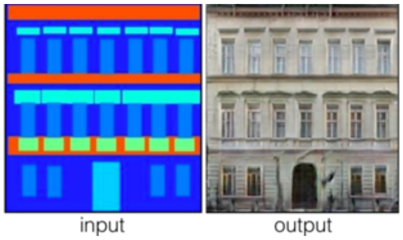

Figure 5. Label to facade

Image synthesis solutions have been improving rapidly, progressing to higher quality and HD resolution. The latest research is now providing user capability to creatively direct the generation and synthesis process. Nvidia Research (Park, Liu, Wang, Zhu, 2018) have developed a GAN model (See Figure 6) that allows user control by simply drawing areas designating the features of interest, in this case, tree, cloud, sky, mountain, sea or grass, in the label map (top row). By choosing an external style image, the user controls the global appearance (first column) of the output image, and the model then renders the corresponding landscape images (grid of 12 images). This capability is provided to the user through an interactive app "GauGAN" (see Figure 7).

Figure 6. User control of semantics & style

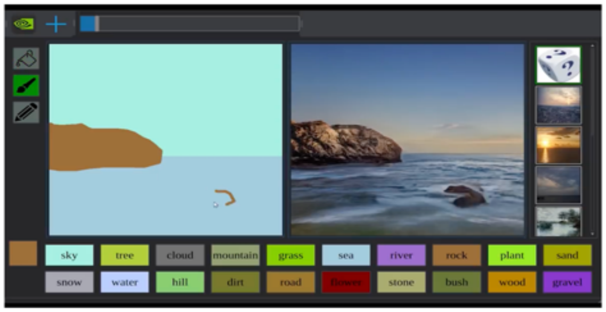

Figure 7. "GauGAN" interactive app

User control of the image synthesis process has also emerged from other research (Bau et al. 2019) and (Wang et al. 2018a), which both provide interactive editing and painting apps to modify the operation of the GAN and its synthesised output image.

Image to image translation techniques are also able to generate synthesised images from the edges in a line sketch. In Figure 8, (Huang, Liu, Belongie, Kautz 2018), three different variants of a shoe image have been synthesised from a simple line sketch.

Figure 8. Edge to image synthesis

Image synthesis techniques provide the underlying capability to generate images of any features. In a SE context, this technique has the potential to simplify and automate some texture creation tasks. For example, it could be used to automate the generation of a set of geo-representative building facades in a particular architectural style; or to quickly create a texture from a sketch or label map when other data is lacking, and without requiring texture creation skills.

The synthesis techniques described so far have created the surface color map, but generating specular, normal, height, roughness, metallic, ambient occlusion, transparency, light or detail maps is a non-trivial task, particularly when applied to many textures. Automatically creating maps from an input image offers significant productivity benefits. Figure 9 (Deschaintre, Aittala, Durand, Drettakis, Bousseau, 2018), illustrates a CNN based approach capable of recovering per-pixel normal, diffuse albedo, specular albedo and specular roughness from a single picture of a flat surface. (But this does require the input to be lit by a hand-held flash).

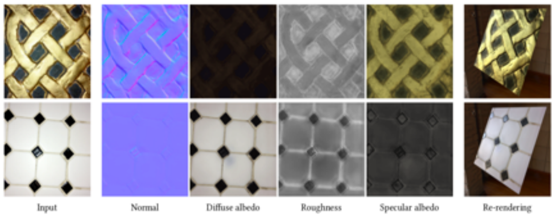

Figure 9. Texture maps recovered from input image

GAN's also provide other image processing capabilities that can be applied to texture creation, including super-resolution that can be used to re-master textures, seam removal, colorization and in-painting. Many of these capabilities are now appearing in commercial image editing tools – some examples are Adobe Sensei (Luan, Paris, Shechtman, Bala 2017) and Artomatix ArtEngine (see artomatix.com).

Data Cleaning / Alignment / Intensification / Specialization

As it becomes necessary for SE to support collective and joint scenarios, with interoperating air and ground platforms while correlating across live, virtual and constructive systems, they become increasingly reliant on high-fidelity, multi-resolution, authoritative geospatial data. This places great emphasis on efficient and effective geospatial data processes, and their capacity to cope with greater data volume, fidelity, heterogeneity and rate of change.

Geospatial data processing can entail data alignment, cleaning, classification, enrichment and intensification, for any raster (imagery, elevation, point cloud) or vector (cultural feature) data. In practice, the specific data tasks performed are dependent on the state of the available data and SE requirements. There is enormous potential to extract information and value from the rapidly expanding pool of geospatial data to support the construction of the rich SE's required to support today's operational requirements. But this data can only be effectively exploited through the application of automated processes. Driven by at least 35 different ML related geospatial challenges (e.g. SpaceNet, DeepGlobe), significant progress is being made by applying deep learning to longstanding geospatial data issues such as extracting buildings and roads, classifying land use and objects from satellite or aerial imagery.

Many of these have direct relevance to M&S. The USSOCOM 3D Urban Challenge serves to illustrate the growing convergence between M&S and geospatial deep learning solutions. The following examples illustrate some of the possible applications of Deep Learning geospatial processing to aspects of Synthetic Environment production.

Synthetic Map Generation

Creating maps from Synthetic Environment data may be required for Plan View Displays, or when real mapping data is lacking, or when the environment is entirely artificial and has no real map data. Figure 10 (Ganguli, Garzon, Glaser, 2019) illustrates how, using a conditional GAN architecture, image to image translation can create a synthesized map of the features present in an input aerial image.

Figure 10. GeoGAN – maps from imagery

Synthetic imagery generation

There will be circumstances when no satellite or aerial imagery is available at all. In this scenario, a GAN (Xu, Zhao, 2018) can be used to generate a synthesized aerial image from a base map defining the layout of roads, buildings, trees etc. The GAN uses a real sample image of the required geographic style to perform a style transfer and output an image that reflects the feature layout of the base map, with feature styles from the sample image. In this way, it becomes feasible to create realistic imagery to compensate for missing data or as an alternative to real imagery. This creates entirely new possibilities. In Figure 11, a single base map of Corvallis has been used to create three images, with the same geographical feature layout, but with three different styles applied from New York, Seattle and Beijing.

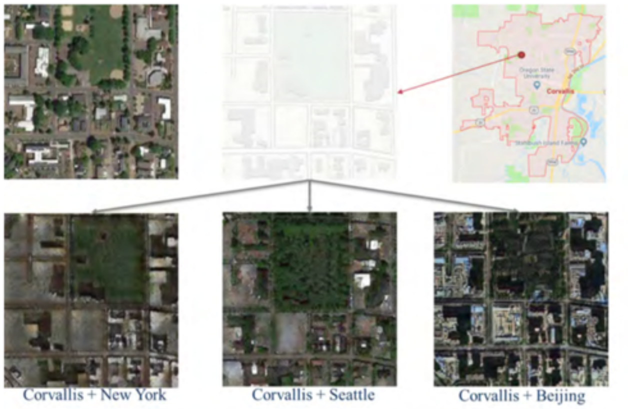

Figure 11. “Spoof” satellite images in three styles

Feature Extraction

Constructing a synthetic version of the real-world requires knowledge of the type, location and characteristics of real natural and man-made features, this is fundamental to generating content - 3D feature representation and placement, topological road networks, airport layouts, 3D urban environments, land-use classification - for all Synthetic Environment client systems. But this data is not reliably available in suitable forms, quality, coverage, currency or content. The explosion in geospatial data now offers many data products that contains information of value (for example, high-resolution imagery (HRI), hyperspectral imagery (HSI), point cloud data, GPS track data). The application of Deep Learning is now making it feasible to extract and derive information using automated processes.

Roads are one such feature but extracting them from imagery is challenging due to the complexity and variability of road structures and their partial occlusion by shadows or obstacles. Figure 12 illustrates three different approaches using CNN architectures. In Figure 13a (Gao, Song, Dai & Chen, 2019), the roads have been extracted from high-resolution imagery alone. Figure 13b (Sun, Di, Che, Liu, & Wang, 2019) has combined extraction from imagery with open GPS track data from vehicles and merged this data to generate the road network. In Figure 13c (Bonafilia, Yang, Gill, Basu, 2019), the CNN has been trained using OpenStreetMap data to improve its capability to extract road features from imagery.

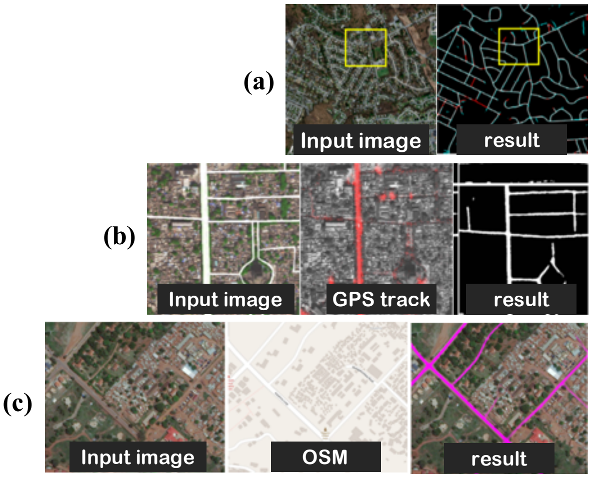

Figure 12. Road Extraction

Building feature extraction is important to SE construction, but also urban planning and mapping and is receiving a lot of research attention through challenges like the SpaceNet Building Extractor Challenge. CNN architectures are used to extract footprints Figure 13a (Lin, Jing, 2019), and accurate building orientation in Figure 13b (Chen, 2019).


(a) (b)

Figure 13. Building Extraction

Image classification provides a semantic view of features to identify land use, vegetation, buildings and water bodies. In Figure 14a (Zhong, Li, Clausi & Wong, 2019), a GAN architecture has been used to classify Hyperspectral Imagery into land use feature classes on 30m resolution data. In Figure 14b (Lv, Ming, Lu, Zhou, Wang, Bao, 2018), a CNN solution has been applied to classify very high (0.8m) resolution imagery for agricultural and urban land use features.

(a) **(b)** **(c)**

Figure 14. Image Classification

Semantically identifying features and extracting building footprints are important enablers to the reconstruction of 3D buildings from imagery, which combines this information with height data, such as a Digital Surface Model (DSM). However, occlusion by trees, atmospheric effects or matching errors create irregularities and noise in the DSM. This produces noisy or partly missing building structures. Using CNN's and a cGAN, Bittner, Körner, Fraundorfer & Reinartz, (2019) (see Figure 15) proposed a methodology that can refine a huge variety of building shapes in DSMs automatically, making them more rectangular and, at the same time, produce improved roof classification maps.

Figure 15. DSM Refinement

The capacity to automatically extract features and semantic information from geospatial data is a significant benefit to SE's. Deep Learning can provide the automated mechanism to reliably obtain feature data, alleviate dependency on existing data, ensure currency, correlation and data provenance and create a semantically rich data layer that can be exploited for many purposes - visualization, sensors, reasoning systems.

Synthetic Environment Run-time

The run-time version of the Synthetic Environment is beyond the scope of the RIEDP RPM, but deep learning and neural network solutions have begun to make interesting impacts in the rendering pipeline. GANs have great potential to change how content will be produced going forward in ways that go beyond the examples described in the previous sections. Using the same principles of semantically labelled input and a GAN based generation of a synthesized output, research is now applying this approach to the real-time rendering pipeline.

Based on a video-to-video synthesis technique (Wang et al. 2018b), Nvidia Research (Nvidia 2018) have trained a model to use a virtual environment built in Unreal Engine that is segmented to describe object location and general characteristics (left image, Figure 14), such as which part of the image contains a car or a building, or where the edges of an object are. The network, trained on real videos, then uses this semantic information to render, in real-time, a photo-realistic version of this virtual environment. (right image, Figure 14).

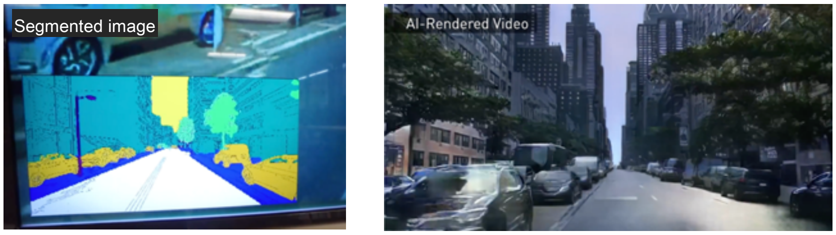

Figure 14. Video-to-video synthesis

The Google Stadia game streaming service is using machine learning style transfer (Google 2019b), to enable the appearance of a virtual world to be changed in real-time by applying a user selected art style to the environment. Because in both of these cases the output is synthetically generated, a scene can be easily changed, either by modifying the visual style and appearance of the scene, or in the case of the Nvidia example, even removing, modifying, or adding objects. These are very embryonic features but provide a glimpse of how construction and rendering of virtual environments may evolve. It can be imagined how applying this type of solution to M&S may open new possibilities - to easily reconfigure environments to emerging needs for example. But such approaches are certain to present challenges when applied to authoritative, real-world environments shared between actors in a joint, collective context.

Synthetic Environment Design

The potential to apply ML to many of the tasks that span the SE production pipeline poses the question of if and how such capabilities could be feasibly combined into a design and production process that is wholly or in part directed by some form of AI. Research has been conducted in the Games industry (Giacomello, Lanzi, Loiacono, 2018) to assess the use of deep learning in Game level design or to assist a human Game designer. Although game design is concerned with generating largely artificial environments, compared to the mainly real, geographically specific environments required for M&S, this research demonstrated the feasibility of training a GAN to generate basic game levels. This may provide a rudimentary indicator of their potential for application to more complex environments.

Given the rapid advancement in applying deep learning to computationally creative tasks through the implementation of GAN architectures in particular, it is not inconceivable that a GAN based solution can be trained to generate real environments. From inputs such as geo-location and satellite imagery, it can be taught how to identify particular land use, learn the characteristic features - vegetation, architectural style - which cultural features to include, and be trained on real data and existing synthetic environments.

CONCLUSIONS

This paper has illustrated the potential to apply AI, and particularly Deep Learning techniques, to tasks throughout the SE pipeline. With the exception of established web image search services, these examples indicate the art of the possible, rather than ready-made solutions. But AI and Deep Learning are evolving very quickly, across many fields, creating new and novel possibilities to apply these technologies to SE production. Developing solutions is not trivial, and AI's need for data and computational power cannot be overlooked. But the pace of AI research and greater access to computing and data resources has stimulated the emergence of open-source frameworks, re-usable models, training datasets and transfer learning techniques, to make AI a more accessible technology. There is significant opportunity to begin to apply Deep Learning to SE production. GAN architectures are relatively immature and challenging to develop but have great potential to transform creative activities. Deep Learning computer vision is heavily researched and widely used and is perhaps the most readily applicable to SE tasks today. Inevitably in such a dynamic field, progress is uneven and is more readily adopted in some cases than others. Of the major phases of the SE pipeline, capitalizing on the progress underway in other fields to direct these technologies to SE geospatial data processing tasks will deliver significant benefits. Developing deep learning capabilities to automatically transform the mass of real-world geospatial data into a synthetic world, or augment the data when necessary, will be transformative to the SE process, increasing realism, accuracy, agility, reducing generation timescales and cost. It will also lay the foundation to extend the exploitation of this technology into new and novel areas, that may even include the design of the SE itself.

REFERENCES

Bau D, Zhu J, Strobelt H, Zhou B, Tenenbaum J, Freeman W, Torralba A (2019). Gan Dissection: Visualizing & Understanding Generative Adversarial Networks. Retrieved From: https://arxiv.org/abs/1811.10597v2

Bittner K, Körner M, Fraundorfer F, Reinartz P. (2019). Multi-Task cGAN for Simultaneous Spaceborne DSM Refinement and Roof-Type Classification. Recovered From: https://doi.org/10.3390/rs11111262

Bonafilia D, Yang D, Gill J, Basu S. (2019). Building High Resolution Maps for Humanitarian Aid and Development with Weakly- and Semi-Supervised Learning. Recovered From: https://research.fb.com/publications/building-high-resolution-maps-for-humanitarian-aid-and-development-with-weakly-and-semi-supervised-learning/

Chen Y. (2019). Learning Orientation-Estimation Convolutional Neural Network for Building Detection in Optical Remote Sensing Image. Recovered From: https://arxiv.org/abs/1903.05862v1

Deschaintre V, Aittala M, Durand F, Drettakis G, Bousseau A. (2018). Single-Image SVBRDF Capture with a RenderingAware Deep Network. *ACM Trans. Graph*. 37, 4, Article 128 (August 2018), 15 pages. https://doi.org/10.1145/3197517.3201378

E. Ameisen. Jul 5, (2018). Building an image search service from scratch. Retrieved from: https://blog.insightdatascience.com/the-unreasonable-effectiveness-of-deep-learning-representations4ce83fc663cf

Ganguli S, Garzon P, Glaser N. (2019). GeoGAN: A Conditional GAN with Reconstruction and Style Loss to Generate Standard Layer of Maps from Satellite Images. Retrieved from: https://arxiv.org/abs/1902.05611v2

Gao L, Song W, Dai J, Chen Y. (2019). Road Extraction from High-Resolution Remote Sensing Imagery Using Refined Deep Residual Convolutional Neural Network. Recovered From: https://doi.org/10.3390/rs11050552

Giacomello E, Lanzi P, Loiacono D. (2018). DOOM Level Generation using Generative Adversarial Networks. Retrieved from: https://arxiv.org/abs/1804.09154v1

Google. (2019a). Google Landmark Recognition. Retrieved from: https://www.kaggle.com/c/landmark-recognition2019

Google. (2019b). Stadia GDC 2019 Gaming Announcement Retrieved from: https://youtu.be/nUih5C5rOrA

Hagerty P. (2017). (https://medium.com/the-downlinq/super-resolution-on-satellite-imagery-using-deep-learning-part-3-2e2f61eee1d3)

Häne C, Tulsiani S, Malik J. (2017). Hierarchical Surface Prediction for 3D Object Reconstruction. Retrieved from: https://arxiv.org/abs/1704.00710v2

Harley A, Wie S, Saragih J, Fragkiadaki K. (2019). Image Disentanglement and Uncooperative ReEntanglement for High-Fidelity Image-to-Image Translation. Retrieved from: https://arxiv.org/abs/1901.03628v1

Huang X, Liu M, Belongie S, Kautz J. Multimodal Unsupervised Image-to-Image Translation. Retrieved from**:** https://arxiv.org/abs/1804.04732v2

Isola P, Zhu J, Zhou T, Efros A. (2018). Image-to-Image Translation with Conditional Adversarial Networks. Retrieved from: https://arxiv.org/abs/1611.07004v3

Lin J, Jing W. (2019). ESFNet: Efficient Network for Building Extraction from High-Resolution Aerial Images. Recovered From: https://arxiv.org/abs/1903.12337v2

Luan F, Paris S, Shechtman E, Bala K. (2017) Deep Photo Style Transfer. Retrieved from: https://arxiv.org/abs/1703.07511v3

Lv X, Ming D, Lu T, Zhou K, Wang M, Bao H (2018). A New Method for Region-Based Majority Voting CNNs for Very High Resolution Image Classification. Recovered From: https://doi.org/10.3390/rs10121946

Nvidia. (2018). Research at NVIDIA: *The First Interactive AI Rendered Virtual World*. Retrieved from: https://youtu.be/ayPqjPekn7g

OpenAI. (2018) AI and Compute. Retrieved from: https://openai.com/blog/ai-and-compute/

Park T, Liu M, Wang T, Zhu J. (2018). Semantic Image Synthesis with Spatially-Adaptive Normalization. Retrieved from: https://arxiv.org/abs/1903.07291v1

Russakovsky, O., Deng, J., Su, H. et al. Int J Compute Vis (2015) 115: 211. Recovered From: https://doi.org/10.1007/s11263-015-0816-y

SISO. (2018). Reuse and Interoperation of Environmental Data and Processes (RIEDP) Data Model Foundations SISO-GUIDE-007-2018. p24. Retrieved from: https://www.sisostds.org/Default.aspx?tabid=105&EntryId=49215

Smith E, Fujimoto S, Romero A, Meger D.(2019). GEOMetrics: Exploiting Geometric Structure for Graph Encoded Objects. Recovered from: https://arxiv.org/abs/1901.11461v1

Sun T, Di Z, Che P, Liu C, Wang Y. (2019). Leveraging Crowdsourced GPS Data for Road Extraction from Aerial Imagery. Retrieved from: https://arxiv.org/abs/1905.01447v1

Wang T, Liu M, Zhu J, Tao A, Kautz J, Catanzaro B. (2018a). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. Retrieved from: https://arxiv.org/abs/1711.11585v2

Wang T, Liu M, Zhu J, Liu G, Tao A, Kautz J, Catanzaro B. (2018b). Video-to-Video Synthesis. Retrieved from: https://arxiv.org/abs/1808.06601v2

Xu, C., & Zhao, B. (2018). Satellite Image Spoofing: Creating Remote Sensing Dataset with Generative Adversarial Networks (Short Paper). *GIScience.*

Zendesk (2017) A simple way to understand machine learning vs deep learning https://www.zendesk.com

Zhai A, Kislyuk D, Jing Y, Feng M, Tzeng E, Donahue J, Du Y, Darrell T. (2017). Visual Discovery at Pinterest. Retrieved from: https://arxiv.org/abs/1702.04680v2

Zhong Z Li J, Clausi D, Wong A. (2019). Generative Adversarial Networks and Conditional Random Fields for Hyperspectral Image Classification. Recovered From: https://arxiv.org/abs/1905.04621v1