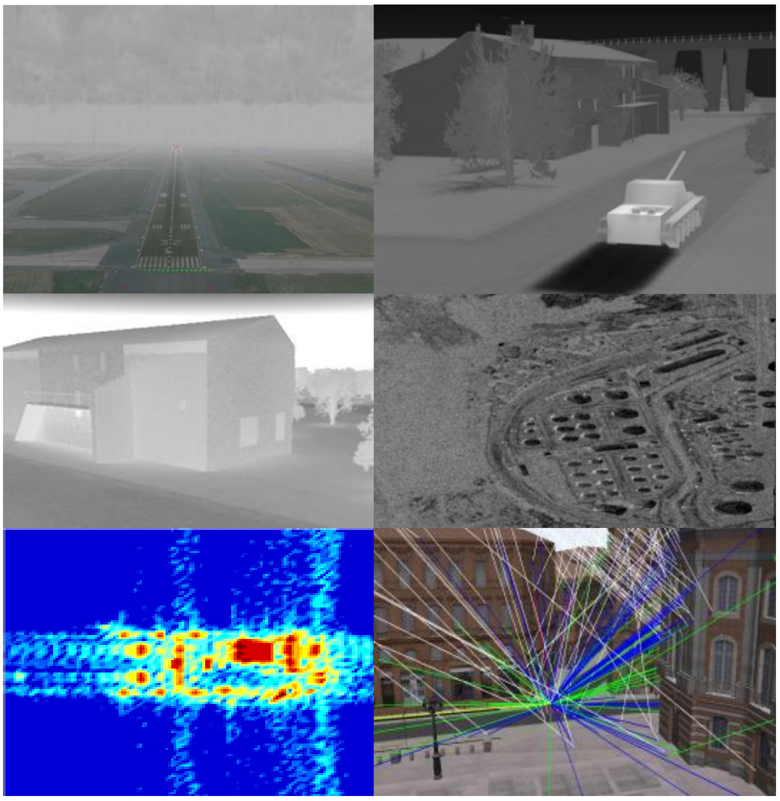
Whitepaper – OKTAL-SE
Sensor simulation for deep learning



1 - Synthetic image generation for Deep Learning image processing algorithms

1.1 - Current situation

Artificial Intelligence and deep learning have been studying for a while by various universities and IT firms, but is has recently made significant improvements, thanks to great technical progress on 3 different areas:

- Hardware computation capacities, thanks to GPU, cloud-based infrastructure…
- Deep-learning Algorithms, mostly open-source, driven by GAFA massive investment in this domain.
- The explosion in the amount of available data, and its exploitation capability through big data technologies.

Deep Learning has then become an accessible technology in the computer vision area, from a technical and financial perspective. However, even though data availability has become massive, it is still a challenge to find relevant data for some kind of applications like the object detection and classification. Well-known constraints such as confidentiality or multi-sensor data availability are examples of bottlenecks for the efficient deployment of Deep Learning process in all domains.
In this context, with its capacity to create “missing” real data by synthetic ones, simulation software can be game changer. But there are still some challenges to be faced, to make this approach operational.

1.2 - What for?

In various domains, deep learning algorithms are explored to process images data (sensors images in the visible, IR or radar bands...):

- Earth observation from space
- Security surveillance (civilian safety)
- Guidance systems
- Vision based navigation systems (automotive, aviation, UAV, maritime and railway industries)
- Detection, Recognition and Identification (DRI) systems, in the defense industry

In these applications, the access to relevant training database is not ensured. In the Defense domain, the sensitivity of information related to some targets leads to a real scarcity of real learning data. The recent increase of sensor fusion systems merging visible, infrared and even radar images to enhance the detection capability makes things harder when comes the time to gather learning data covering all the sensors images of the same area at the same time.
The huge amount of images data available in world databases like COCO or ImageNet are most of the time in the visible domain only and cannot be considered as a solution for the critical applications mentioned above.

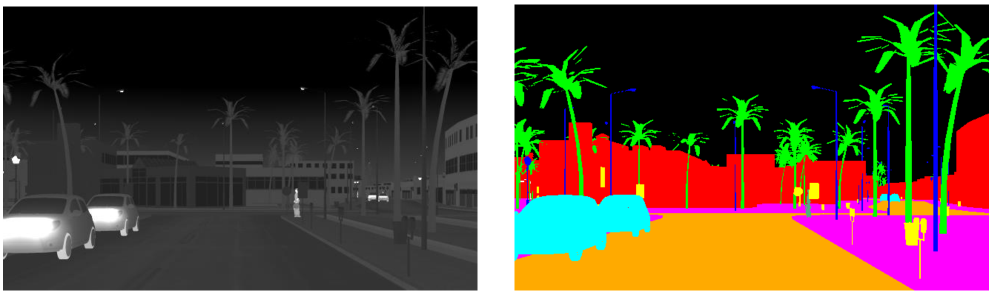



1.3 - Why Synthetic images?

Dedicated approaches are currently investigated to tackle the lack of relevant real data. For instance, neural style transfer could help to create infrared data starting from other electro-optic spectral data. But this comes all the much harder when considering Radar images (like Synthetic Aperture Radar) where there is no obvious correlation is some signals received in the radio-frequency domain (and reconstructed by the RADAR) and a visible image seen by a camera.
Frugal learning techniques are also investigated but still rely on the availability of some few real labeled data, that could be really challenging when addressing confidential areas or targets in the Defense domain.
Therefore, creating learning data, as we would expect the real ones to be, appears very attractive.

Synthetic data provides users with several key advantages:

- Knowledge of the ground truth (labelling is straightforward)
- Ability to consider any environment (background, weather, lightning conditions…), and any "confidential" areas/objects
- Ability to carry out parametric studies
- Ability to consider all type of acquisition mode (any type of sensors) as long as the simulation mean is qualified for this physics based approach.



A synthetic infrared image (left hand-side) and its segmentation map (right hand-side), both generated by OKTAL-SE software

2 - Return on experience and challenges to address

2.1 - Lessons learned from experiments

OKTAL-SE has made experiments through several projects, such as the DGA (French Mod) led project, Man Machine Teaming, in partnership with DASSAULT AVIATION, and THALES companies : https://man-machine-teaming.com/oktal-se-magellium-et-le-projet-constitution-de-bases-de-donnees-de-scenes-synthetiques-eo-ir-rf/

The experiments made in these projects stress out 2 main points:

- The synthetic data have to be realistic, deep-learning algorithm wise
- The training process and the CNN configuration have to be adapted to synthetic data

The following paragraphs describe the main challenges to face, in order to obtain good results when a deep learning algorithm is trained with synthetic data.

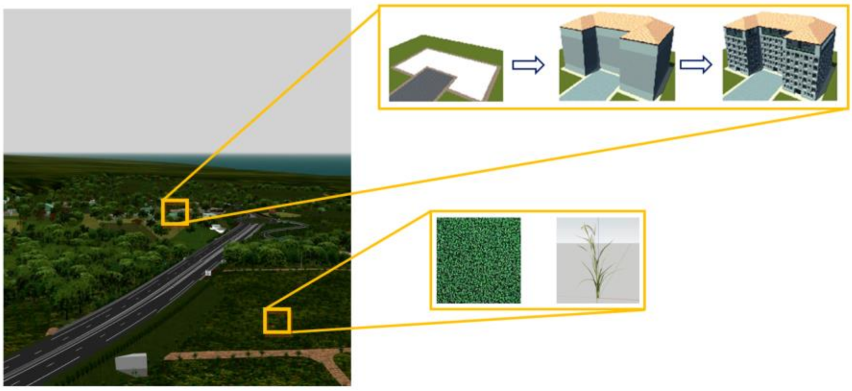
2.2 - Simulation technical challenges to address

2.2.1 - Realism of synthetic world

Synthetic data are a virtual representation of the real world. To be realistic, these data require a sufficient level of details, in order to obtain populated scenes, and to avoid too smooth and poor images. A classical solution is to enrich the images by adding details and entropy (noise). The main advantage is the simplicity of the approach. The drawback is that it neglects the interaction between the added elements and the surrounding environment of the 3D scene (interaction due to cast shadows, electromagnetic coupling,…). Then, it appears more realistic to work directly at the source level - meaning at the 3D mock up level - to increase the level of detail. Of course, using this technique involves other challenges, in particular the memory size of the 3D synthetic environment.

The following techniques can address these challenges

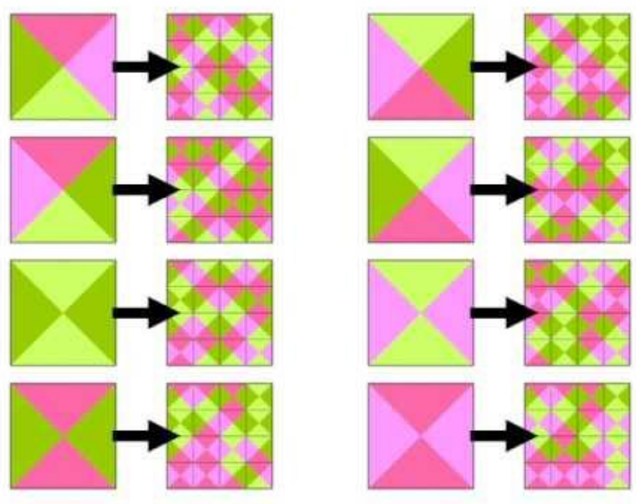
<u>Procedural generation</u>
The principle of procedural generation is to store a rule to generate an object, or to pick it from a library, rather than storing all objects of a scene. Apart from enriching the level of detail of a scene, this technique automates a part of the generation process, which leads to a cost reduction for generating a detailed synthetic environment.



OKTAL-SE procedural algorithms, used to automatically generate buildings and vegetation

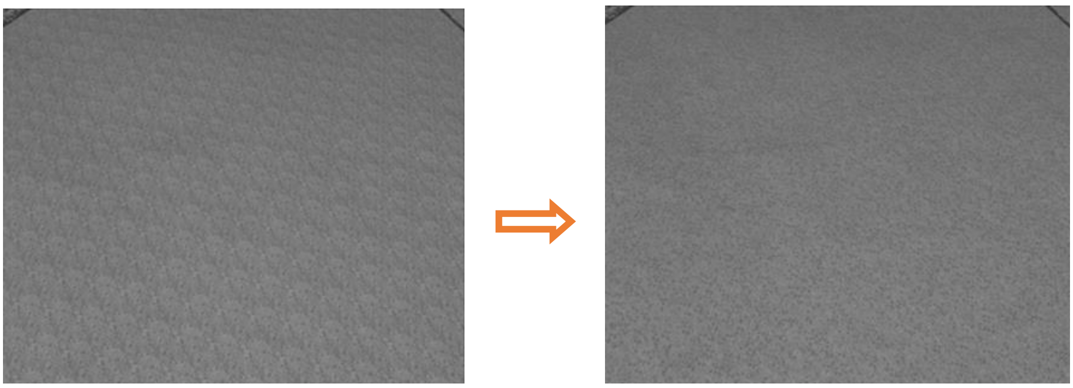
<u>Wang tiling</u>

Another problem in using synthetic data, is the apparition of repeatability patterns (especially in clutter areas such as fields, forest canopies…), due to the fact that the same texture patch is usually repeated on wide areas. This phenomenon, called the “ravioli effect” is to be banned when synthetic data is used to feed image processing algorithms, otherwise it will create artificial alarms in the image treatment process. For Deep Learning algorithm, this problem can become highly critical, as it can generate a noise training phenomenon, called overfitting.
For a 2D texture, Wang tiling principle, relies on a mathematical algorithm, applied to a limited number of tiles (patches). The algorithm selects and copies the tiles to automatically arrange them side by side with matching borders, *without* repeating them.



The scheme on the left displays Wang tiling principles, used to fight the “ravioli effect”.

This technique is applied to fight monotony on 2D textures.



A radiometric texture (IR) without (left) and with (right) OKTAL-SE Wang tiling algorithm

Recent study and development carried out by OKTAL-SE have adapted this mathematical process to 3D patches (for instance terrain tiles with 3D buildings) including the physical properties that are needed to address physics based sensor approach.

Here, the Wang Tiling technique has been applied to OKTAL-SE 2D source data

OKTAL-SE Wang Tiling algorithm is then applied to a 3D environment, to avoid repeatability on the automatic 3D building generation process

<u>Photorealistic techniques</u>

The photorealistic side of synthetic images, well known in the video game industry, is also important when these data are used for a deep learning algorithm, to minimize the "reality gap" (difference between real and synthetic data). Some techniques such as IBR (Imaged Based Rendering) or projection mapping, use real photo images to build realistic 3D scenes. Some other techniques, such as the use of billboards, or "theatre curtains", are used to create realistic images in a cost-effective manner. Then, the enhancement of geometric and texture resolutions is also used to create more realistic scenes.

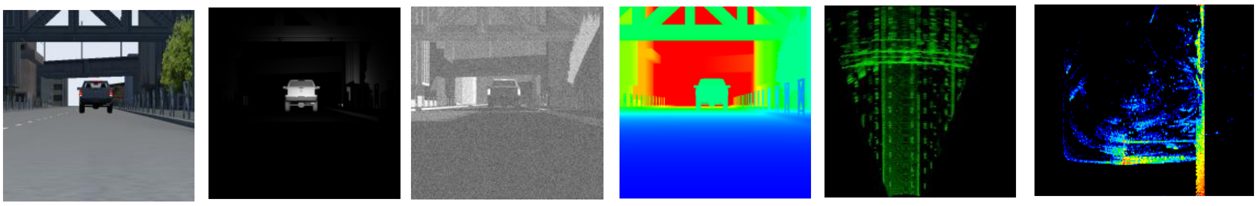
Here, projection mapping technique has been used by OKTAL-SE to build photorealistic scenes

These techniques can be helpful to enhance the photorealism of generated synthetic data, but the memory constraints have to be kept in mind, especially in the context of physics-based simulation, when software have to compute physics modelling in the same time than the image rendering process.

2.2.2 - Multi-sensors

As described on the above paragraphs, it can be a real struggle for users to access to specific sensor data, such as hyperspectral visible images, or IR and RADAR images. To build a synthetic sensor database, simulation software has to have the capacity to generate different kind of sensor data, preferably from the same Synthetic Environment (so there is one only reference environment).

This point is mandatory to help users to design sensor fusion algorithms (i.e., several IR bands fused with visible), or multi-sensors hybridization algorithm. These kinds of algorithms are used, for instance, in the frame of ADAS applications, where visible ad infrared cameras are used, as well as LiDAR, RADAR, and GNSS (GPS) sensors.



The above images, generated by OKTAL-SE sensor simulation software, display the same scene, generated form different simulated sensors, from the electro-optic to radiofrequency domains (car embedded sensors)

2.2.3 - Massive data generation

As deep learning algorithms required a large amount of data, synthetic dataset generation has to be massive. This aspect can then be a challenge, as massive computation capacity can be required. With the recent improvement in GPU performances, a solution can be to exploit the graphic-board to accelerate image generation computation. Using this mode, real time rendering engine can easily generate 60 images per second, or more.

Beyond the frame rate pf generation, monitoring aspects have to be taken into account. Different monitoring techniques and processes can allow users to automatically identify incorrect images, and to sort them apart from the generated training dataset. Some other techniques will allow users to keep a control on the generation process, and to avoid unfortunate crash, which could have happened in the middle of a long generation process.

2.2.4 - Mitigation of the dataset memory size

When generating a massive training dataset, representing all possible conditions, the final dataset size can be quite important. This aspect can be a problem for storing or manipulating data. It is then important for users to be able to establish a sharp specification of the required data simulation parameters, in order to remain efficient and to generate the right amount of data, saving a useless extra use of memory.

2.3 - Deep-Learning technical challenges

The CNN final performance not only is linked to the inherent quality of the synthetic data, but also to the way databases are constituted. Several best practices have to be followed, in order to obtain good results:

- **Specifying relevant synthetic data:** Neuronal Networks behaviors are hard to predict and very sensitive to their training data too. As such, it is a real challenge to specify, in advanced, the ideal content of the training database. An iterative empiric approach is usually the best way to draw conclusions from a dataset N, in order to specify the content of the dataset N+1. This process is to run continually, until the performance of the CNN has reached its maximum.

- **Synthetic and real data hybridization:** in most cases, a full synthetic database will lead to poor results. At the contrary, when real data are mixed with synthetic ones, results can be a way better than a 100% real database, especially for under-represented classes. The challenge is then to find the optimum ratio (usually named gamma) between real and synthetic data.

- **Domain randomization:** The principles of this method is to generate several images, from one source image, by replicating it, and introducing little perturbations. This method will force the algorithm to learn more complex features and at the same time focus on specific information of the data, to select the highly informative pixels. The use of this method with synthetic data has still to be consolidated but, so far, research results are quite encouraging.

- **Transfer learning:** Transfer learning is a learning methodology which is based on the weights use of a different neural network. It could be useful to opt for the Transfer learning when a training *dataset A* has an insufficient amount of data, but representative of the operational case. With another complete *dataset B*, less representative of the operational case, a neural network is trained. This weight transfer allows to improve the neural network inference performance trained by *dataset A* thanks to the learning features acquired by the *dataset B*. In our case, the algorithm could be firstly trained on massive synthetic dataset (massive but less representative), and then, through transfer learning, the algorithm could be trained to a real dataset (with fewer data). So far, the experiments made on this domain are quite promising, although the level of maturity of such an approach need to be raised.

- **Neural style transfer:** Algorithms such as GAN (Generative Adversarial Networks) can learn the style of one set of images and transfer it to another set. This technique can be used to transfer the style of real sensor images to synthetic sensor images, in order to add some realistic grains. OKTAL-SE has participated to research works on this area, in the frame of MMT project : https://man-machine-teaming.com/irt-et-oktal-se-pour-le-projet-simulation-de-capteurs-augmentee-par-reseaux-de-neurones/. This technique is promising, but still requires research works to be mature enough to be used.

3 - Conclusion

Using physics-based sensor simulation software to generate synthetic datasets is a promising concept. This solution could be used to resolve challenges faced by critical industries, where the access to training data is not something obvious (multi-sensors images, sensitive targets detection….).

However, some challenges have to be faced to turn this concept from a promising idea to an operational solution, and make it a real game changer:

- A dedicated process, adapted to synthetic data, has to be developed for the training and the assessing phases of a CNN.
- The quality of the generated data has to be adapted (physics-based generation, non-uniformity of textures, rich level of details…).
- The simulation solution has to have a multi-sensors generation capability, in order to address data frugality environment.
- Not only users need to be simulation tools users, but also AI experts, as they need to master the required skills to specify the content of a hybrid training dataset, and to make this dataset adapted to a CNN (GAN, domain randomization, etc…).

4 - References

[1] *Victor-Emil Neagoe, Adrian-Dumitru Ciotec, Lorenzo Bruzzone*, **A Weakly-Supervised Change Detection Technique For SAR Images,** Published in IGARSS 2019 - *2019 IEEE International Geoscience and Remote Sensing Symposium*

[2] *Kevin Lee, David Moloney*, **Evaluation of Synthetic Data for Deep Learning Stereo Depth Algorithms on Embedded Platforms**, Publié en 2017 *4th International Conference on Systems and Informatics (ICSAI)*

[3] *Marc Flammante*, **History and uses of the newly published artificial images for neural networks patent,** Published in 2019 *on LinkedIn*

[4] *Sungho Kim, Woo-Jin Song et So-Hyun Kim*, **Double Weight-Based SAR and Infrared Sensor Fusion for Automatic Ground Target Recognition with Deep Learning**, Published in *2019 in Remote Sensing*

[5] *Chen-Chun Huang, Yi-Leh Wu, Cheng-Yuan Tang,* **Human Face Sentimenet Classification Using Synthetic Sentiment Images With Deep Convolutional Neural Networks,** Published in 2019**,** *Conference: 2019 International Conference on Machine Learning and Cybernetics (ICMLC)*

[6] *Pelin İrtem, Emre İrtem, Nesli Erdoğmuş*, **Impact of variations in synthetic training data on fingerprint classification**, Published in 2019 *International Conference of the Biometrics Special Interest Group (BIOSIG)*

[7**] Towards Learning 3d Object Detection and 6d Pose Estimation from Synthetic Data,** Published in 2019, *Conference: 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*

[8] *Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, Stan Birchfield ,***Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization,** Submitted in 2018

[9] *Slawomir Bak, Peter Carr, Jean-Francois Lalonde*, **Domain Adaptation through Synthesis for Unsupervised Person Re-identification,** Submitted in 2018

[10] *German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, Antonio M. Lopez*, **The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes**, Publié en 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*