HEXAGON

White paper

Building detection from aerial and satellite images using semantic segmentation

Identifying and analyzing footprints of buildings in aerial and satellite data is an important first step in many applications, including updating maps, modeling cities, analyzing urban growth and monitoring informal settlements. But manually identifying and collecting information about buildings from single or stereo imagery is very tedious and costly.

To solve this problem, various automatic building methods, which can be broadly categorized as pixel-based or object-based, have been introduced. Height data in the form of a digital surface model, either generated through stereo matching or directly acquired by a laser scanner, is also used as input to increase the accuracy of the detection.

In general, the pixel-based method processes and assigns classes to individual pixels in the image. Well known pixel-based classification methods include minimum-distance/nearest neighbor, parallelepiped and maximum likelihood classifiers. These methods rely mainly on the radiometric characteristic of the image, often resulting in salt and pepper classification with many small regions.

On the other hand, object-based image analysis tools, such as IMAGINE Objective, start by segmenting the image into groups of similar pixels and then classify and label the segments based on other cues, like shape, size and texture. While the object-based image analysis methods have proven to get better results than the pixel-based classifiers, they are generally complex and need a subject matter expert to perform the classification.

Recently, advances in computing power have made deep learning-based classification methods viable. Several deep learning-based classification and object detection methods have been implemented in ERDAS IMAGINE.

In this paper, we review the general categories of deep learning-based classifiers and then look at how semantic segmentation can be used for detecting buildings.

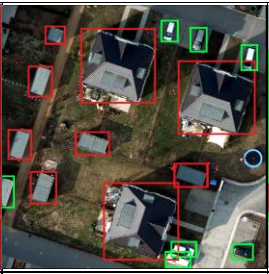
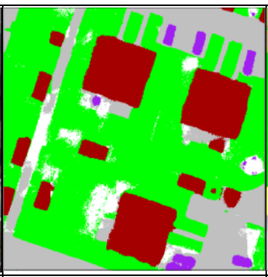
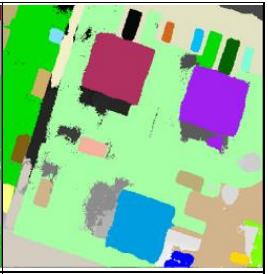
In general, classification tasks in deep learning can be grouped in four broad categories:

Image labeling: In image labeling, an image is assigned a single class. For a large image, the image can be virtually divided into regular tiles and each tile assigned to a single class.

Object detection: In object detection, the goal is to detect objects within an image along with the bounding boxes of the objects. That means the class, position and size of each object is predicted.

Semantic segmentation: Semantic segmentation groups together parts of an image that belong to the same object class. The goal is to label each pixel of an image with a class telling what is being represented.

Instance segmentation: Instance segmentation goes a step further from semantic segmentation to give a unique ID to every instance of a particular object identified in the image.

				
Original image chip	Image labeling	Object detection	Semantic segmentation	Instance segmentation
	The image chip is assigned to a single class	Bounding boxes for the objects are generated	Each pixel in the chip is assigned to a class	Each object is identified separately

Semantic segmentation

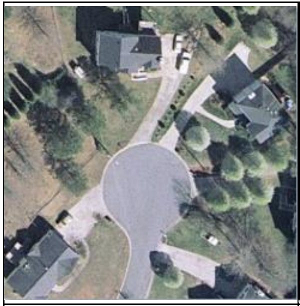
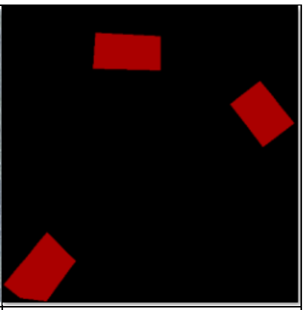
The goal of semantic image segmentation is to label each pixel of an image with a class of what is being represented. This is generally done in two steps. First, a convolutional neural network (CNN) model is used to extract objects by segmenting images at higher pyramid levels. The extracted objects are then propagated down through the pyramid layers to the full resolution image to better refine their boundaries. The features are then used as masks to label the pixels of the image.

The process involves three major steps:

- Collecting training data
- Creating an initialized (trained) machine intellect
- Segmentation of buildings

Gathering training data

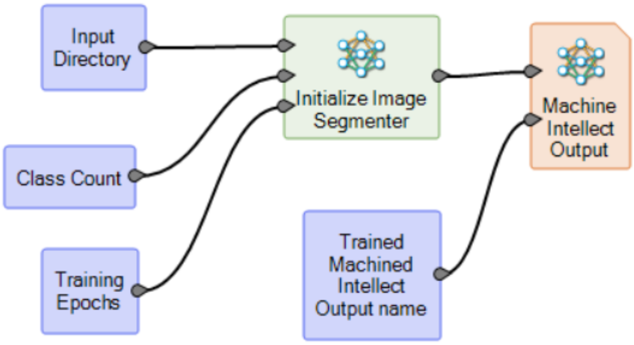
One of the most important tasks in deep learning-based classification and object detection processes is collecting training data. For semantic segmentation, the training data are image chips that are labeled at the pixel level. This means each pixel of the image chip is annotated with the class (or classes) it represents, such as buildings and not-buildings.

	
Training image chip	Label image identifying buildings in the image chip

Creating an initialized (trained) machine intellect

The next step is to create an initialized (trained) machine intellect based on the training image chips and labels.

In ERDAS IMAGINE, this is done using the Spatial Modeler environment, which has an operator named Initialize Image Segmenter that performs the initialization. The operator uses a U-Net fully convolutional network to build a machine intellect for performing image segmentation.



During initialization, the network is trained based on the training image chips and their associated labels. The resulting machine intellect can then be used to segment new images that are similar to the data used during initialization. During this step, the accuracy of the initialized machine intellect is verified using an independent dataset.

Segmentation of buildings

The last step of the workflow is segmenting buildings in other images using the machine intellect created in the initialization step. This is also done in the Spatial Modeler environment by constructing a model that uses the initialized machine intellect and the image to be segmented. This automatically generates an image with buildings segmented from the rest of the image.

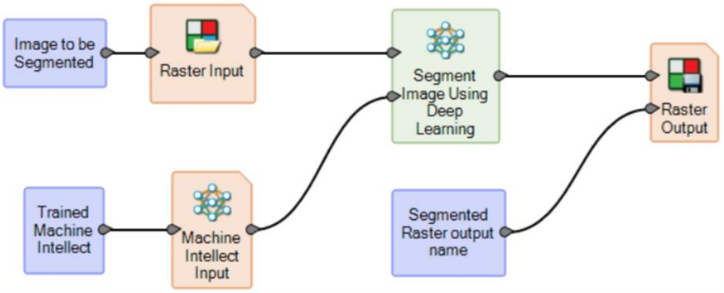


Image to be Segmented
Raster Input
Segment Image Using Deep Learning
Raster Output
Trained Machine Intellect
Machine Intellect Input
Segmented Raster output name

For a single class segmentation, the pixels that are not identified as buildings will be assigned as background pixels. The output image from the segmentation will be pixels that are assigned either to the building class or the background.

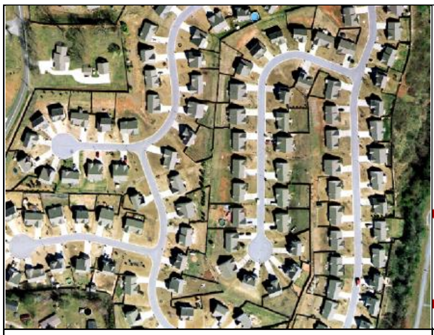
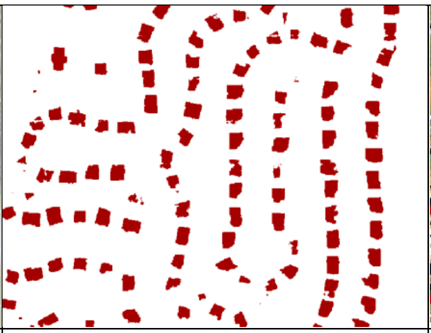
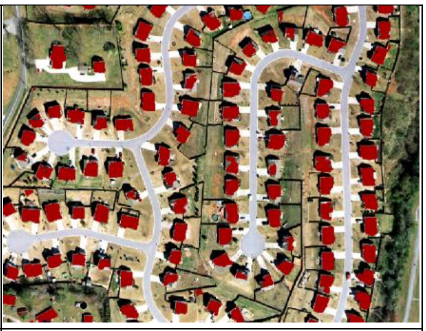
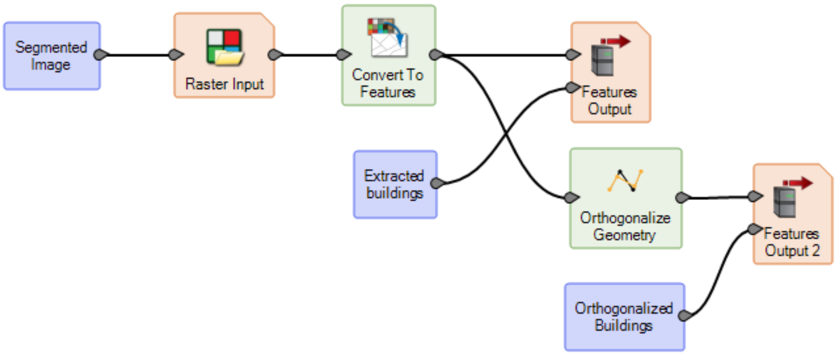
		

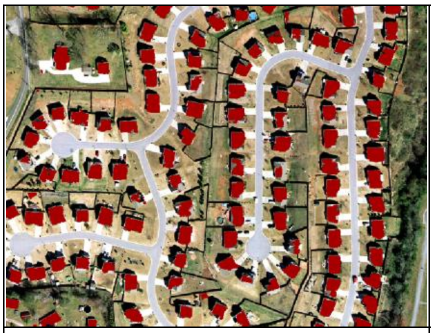
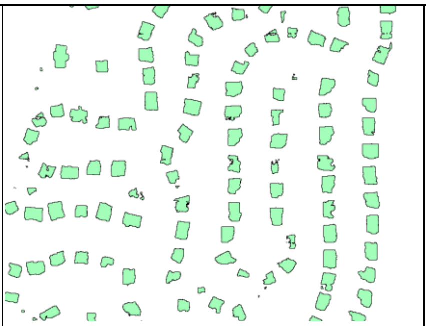
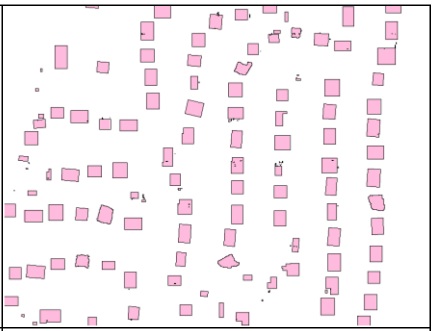
Image to be segmented	Segmented image	Segmented image superimposed on the input image

Next steps

The segmented image can be used as an input in other applications, such as change detection (identification of new buildings) or city delineation. For applications that need the detected buildings in vector format, a spatial model can be constructed in the Spatial Modeler environment that takes the segmented image and generates a vectorized format.



The model takes the segmented image and generates polygon features from the pixels labeled as buildings. The extracted polygons are simplified by making them orthogonal. Further processing can also be applied to the features, such as filtering polygons based on minimum area criteria to remove false positives.

		
Segmented image	Extracted features	Orthogonalized features

Semantic segmentation in ERDAS IMAGINE is a useful method for identifying buildings in the process of image classification. With machine learning-based methodology, it reduces the need for expert users and is highly accurate. The ease and flexibility of semantic segmentation make it a valuable tool in the process of building extraction.

HEXAGON

Hexagon is a global leader in digital reality solutions, combining sensor, software and autonomous technologies. We are putting data to work to boost efficiency, productivity, quality and safety across industrial, manufacturing, infrastructure, public sector, and mobility applications. Our technologies are shaping production and people-related ecosystems to become increasingly connected and autonomous – ensuring a scalable, sustainable future.

Hexagon’s Safety, Infrastructure & Geospatial division improves the resilience and sustainability of the world’s critical services and infrastructure. Our solutions turn complex data about people, places and assets into meaningful information and capabilities for better, faster decision-making in public safety, utilities, defense, transportation and government.

Hexagon (Nasdaq Stockholm: HEXA B) has approximately 21,000 employees in 50 countries and net sales of approximately 3.8bn EUR. Learn more at hexagon.com and follow us @HexagonAB.