The model consists in a CNN trained to learn first a series of downsamplings (a standard CNN) and then a series of nonlinear upsamplings by means of deconvolutions, up to the original input size. The CNN is trained on independent patches, allowing more flexible training criteria, including both color and normalized DSM as stacked input channels. Inference with this model simply consists in a single forward pass of the arbitrarily sized test image and a map with exactly the same resolution as the original image is obtained, very efficiently. This allow to better preserve geometric structures and to take into account the relationships between semantic classes in the input patch, learning de facto a locally structured segmentation model.