

Generative Adversarial Urban Growth Prediction

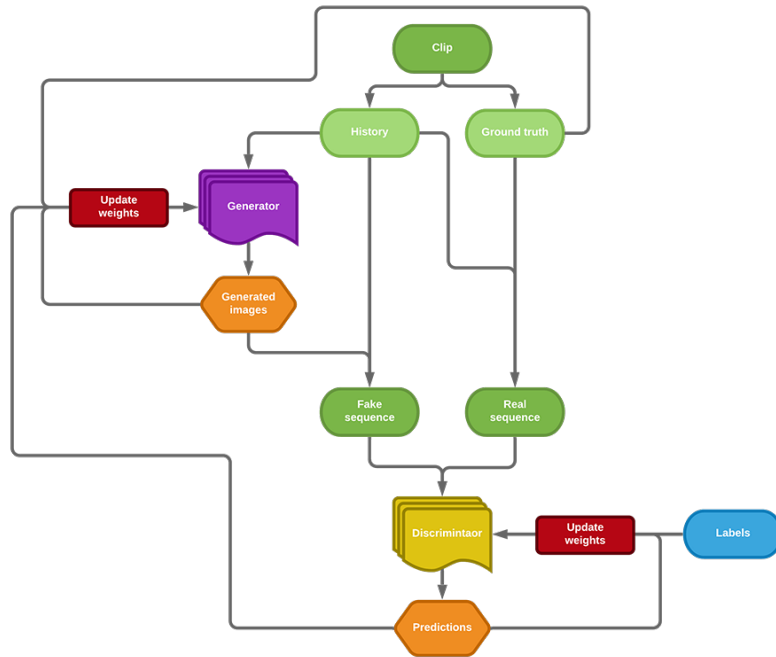
Ziad Khattab

1 Network model

Prediction of urban growth from a sequence of satellite images requires not only the identification of growth patterns, but the ability to generate images of future maps with decently realistic quality and good enough definition to visually identify urban and non-urban areas. For this, a generative adversarial network (GAN) is used.

In a GAN, a generator network receives the history frames and attempts to provide a realistic continuation to the clip, and a discriminator network attempts to determine whether the clips it receives are real or fake, assigning a probability to the image between 0 and 1 of how likely it is to be real. The two networks compete against one another, with the generator attempting to fool the discriminator into thinking that the generated output images are areal, while the discriminator attempts to pick apart real and fake images more accurately.

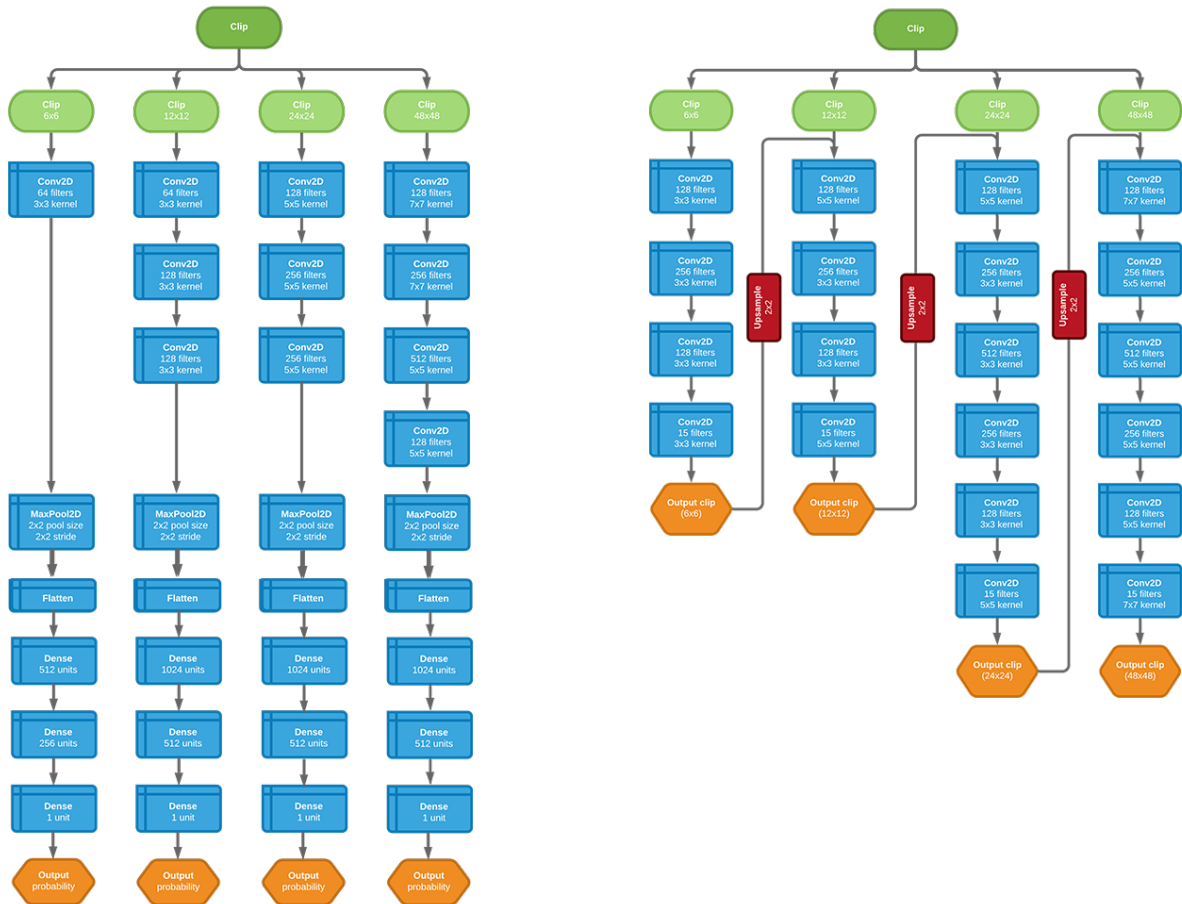
Figure 1: graph of the generative adversarial model



The discriminator is a convolutional neural network image classifier that runs the clips through convolutional layers activated with ReLU, and then to fully connected layers to obtain a scalar prediction between 0 and 1. It operates on a multi-scale model, meaning that a 48 by 48 pixel square clip is downsampled to 6 by 6, 12 by 12, 24 by 24, and the original clip, and a prediction generated for each scale.

The generator is a fully convolutional image generator model that also operates with the same four-scale model. However, a significant difference from the discriminator is that the generator concatenates the upsampled output of each scale to the next one to strengthen the time dependency.

Figure 2: graph of the discriminator and generator models



2 Loss functions

To mathematically define the loss functions used in training, we must first define the generator and discriminator models as functions,

$$\begin{aligned}\text{gen} &: \text{images} \rightarrow \text{images} \\ \text{disc} &: \text{images} \rightarrow [0, 1]\end{aligned}$$

2.1 ℓ_p loss

We define the ℓ_p loss as,

$$\ell_p(x, y) = |x - y|^p, \quad p \in \{1, 2\}$$

This loss represents either the absolute difference (if $p = 1$), or the absolute squared difference (if $p = 2$) between the generated images x and the ground truth images y . This is the simplest metric for the accuracy of the generated images.

2.2 Adversarial loss

First, define the binary crossentropy loss as,

$$\text{bce}(y, y') = - \sum_i y_i \log(y'_i) + (1 - y_i) \log(1 - y'_i)$$

where y represents the predicted labels given as output of the discriminator, and y' represents the true labels assigned to the data.

Next, we define the adversarial loss,

$$\text{adv}(x, y) = \text{bce}(\text{disc}(x, y), 1) + \text{bce}(\text{disc}(x, \text{gen}(x)), 0)$$