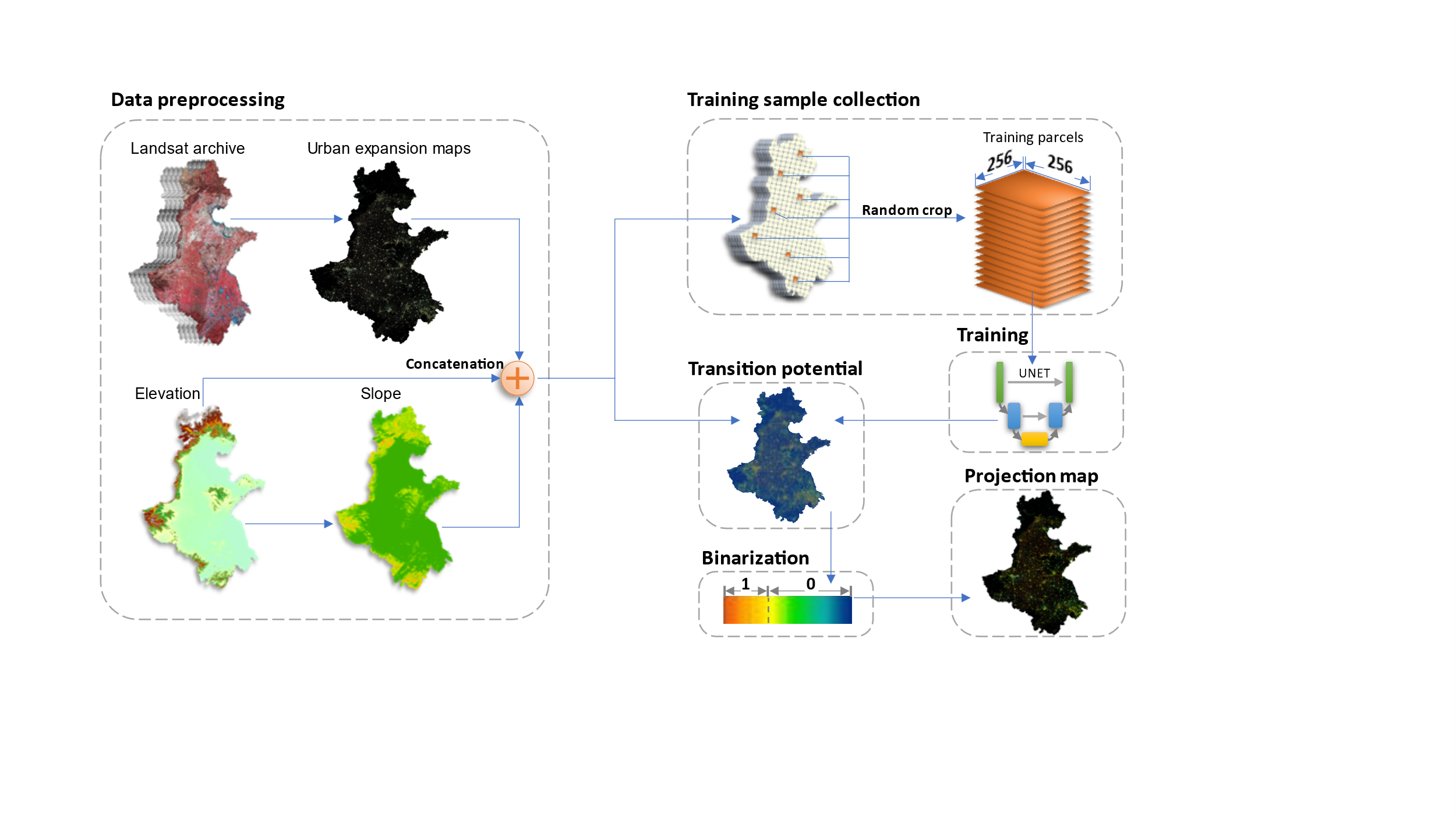
# 2. Data and Method

The research workflow is summarized in Fig. The Landsat archive was used to mapping the urban development from 1990 to 2019 with a 3-year interval (paper-1). Digital Elevation Model (DEM) and slope were used to improve the projection performance. All raster data had the same 30m resolution and were concatenated as the input image. The training samples were randomly cropped from the input image with a size of 256\*256 (i.e., the width and heights of the parcel were 256 pixels). The UNET was trained on these samples and then coupled with the input image to produce the transition potential map. Lastly, a threshold was determined to binarize the transition potential map into a projection map where pixels with value 1 as the urban land and 0 as non-urban land.



## 2.1 Data preprocessing

The raster data used in this study were summarized in table. The urban dynamics maps were obtained from (paper-1). This product spanned from 1990 to 2019 with three-year interval and had a consistent high accuracy of over 94% across years because the use of temporal features successfully delineates urban land from fallow cropland and a consistent check removed unreasonable urban developments. The digital elevation model (DEM) was obtained from the Shuttle Radar Topography Mission (SRTM) and the slope data was derived from the DEM. The DEM and slope were used to assist the machine to learn to adjust urban development patterns in complicated topology conditions, which have been proven effective in previous studies (ref).

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Type** | **Source** | **Resolution** | **Time Span** |
| Urban dynamics | Wang et. al. | 30m | 1990-2019 |
| DEM | SRTM | 30m | 2000 |
| Slope | Derived from DEM | 30m | 2000 |

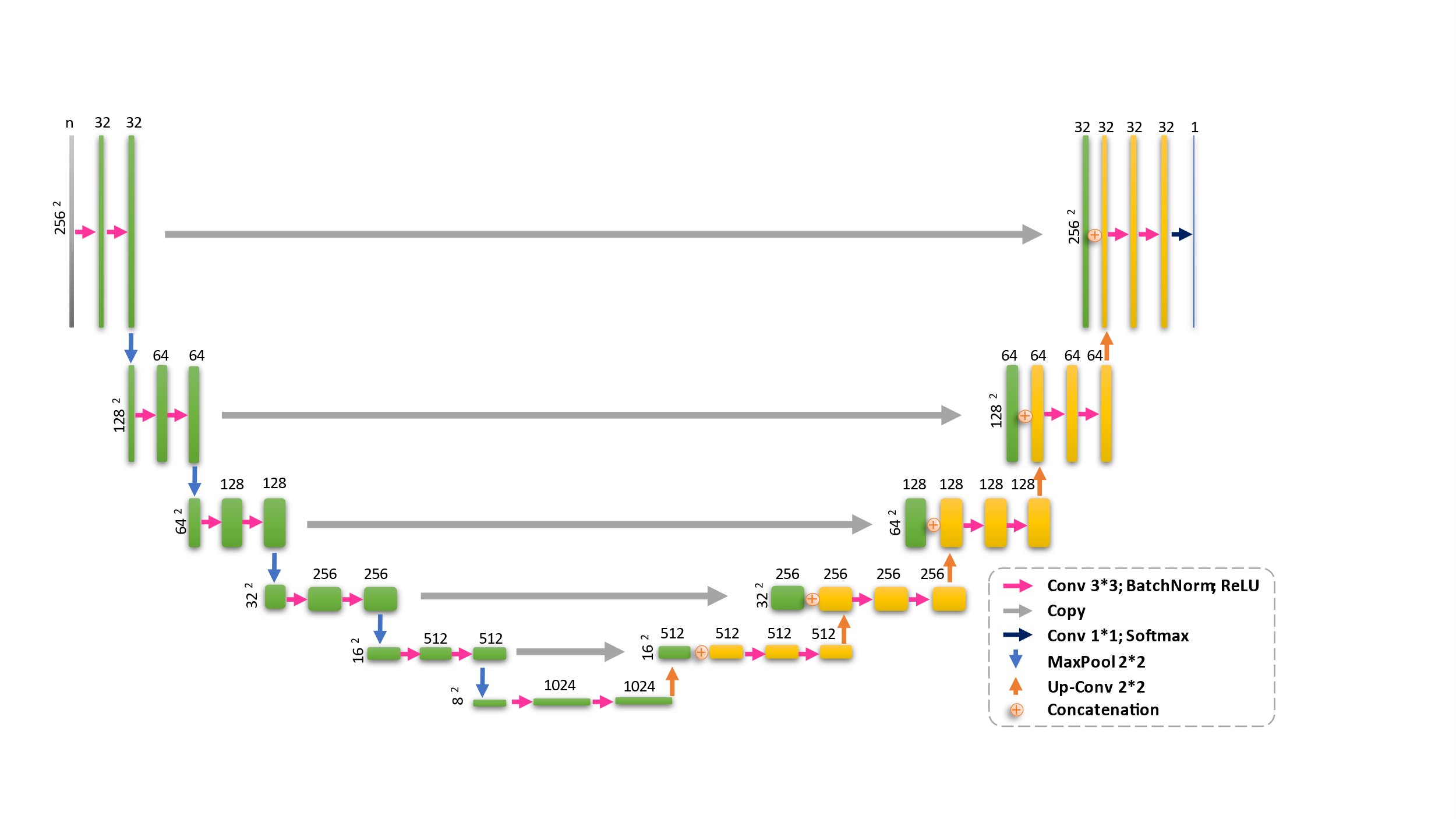
## 2.2 Control sample collection

Control samples were collected using Google Earth Engine for its parallel computation advantages (ref). The *neighborhoodToArray* module was used to randomly crop parcel samples from the input multiband image. We set the neighborhood size to 256 following a common data science practices (ref) and collected 25,000 parcel samples, among which 20,000 were used for the model training and the rest 5,000 for validation.

## 2.3 Model training

UNET is a deep convolutional network that commonly used for image segmentation (ref). Down-sampling, up-sampling and skip-connections are the primary components for UNET (ref). The down-sampling enables UNET to extract the general context of the input image, the up-sampling refined the context pattern to precise shapes, and the skip-connections balances the generalization of down-sampling and the localization of up-sampling (ref). The output of the UNET is an image that has the same size to the input data, where each pixel indicating the probability belonging to a preset category.

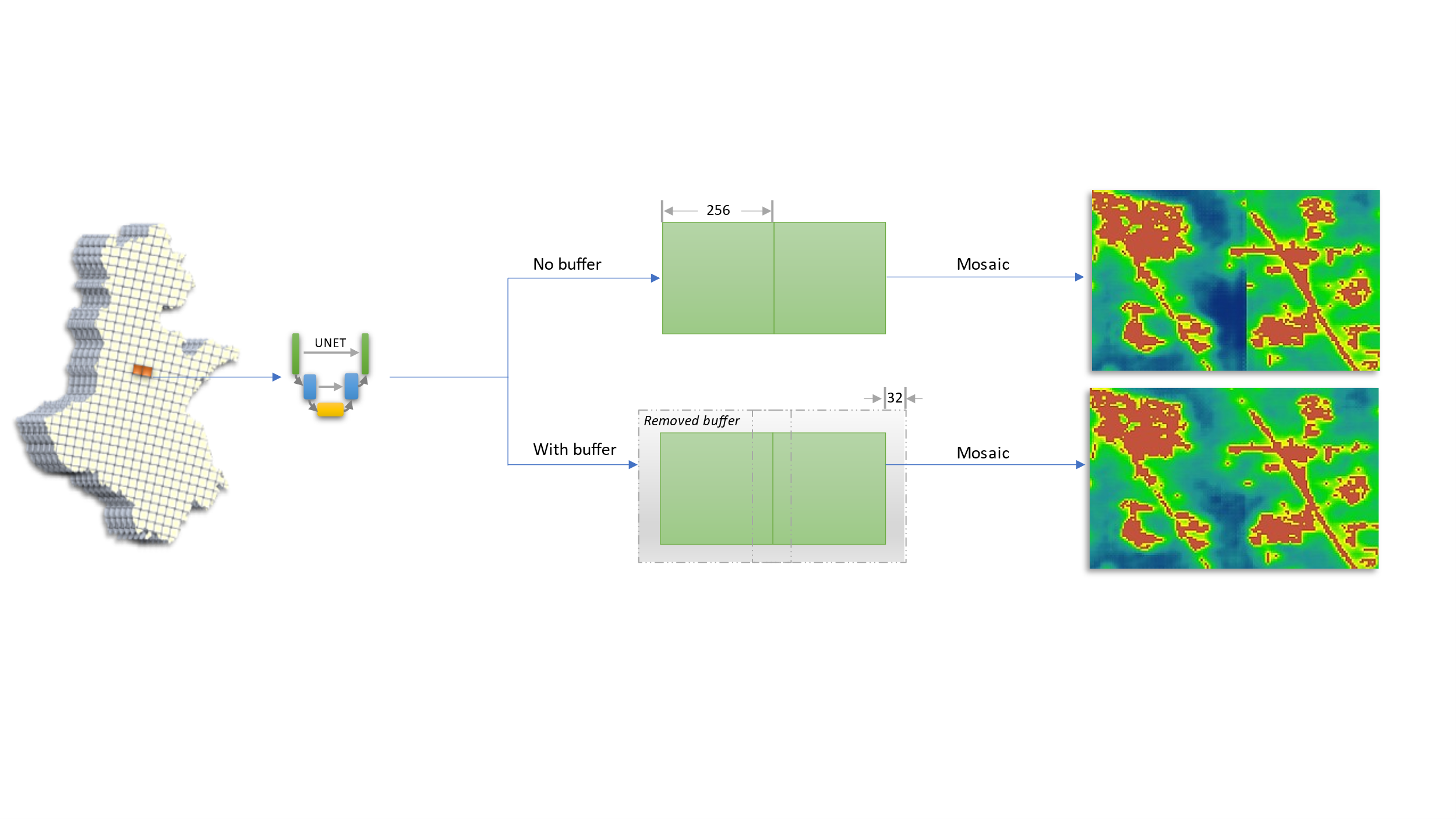
In this study, the UNET was trained on parcel images with the size of 256 × 256 pixels. Each training image parcel include a 3-layer input data (an early data urban map, a DEM and a slope map) and a single-layer target data (a later data urban map indicating the true urban expansion). The 3-layer input data, during training, will be transformed to the size of 8 × 8 pixels after 5 down-sampling processes, and then converted to a single-layer output image with the original size of 256 × 256 pixels over another 5 up-sampling operations. The mean squared error (MSE) was used to measure the difference between the output and the target image. A total of > 31 millions of parameters were involved in the training process and the derivative of each parameter to the MSE difference was calculated via a backpropagation process. The parameters were updated according to the derivatives: if increasing a parameter would increase the MSE, then this parameter will be decreased, and vice versa. The validation samples had the same structure to the training sample but not updating the parameters, which were used to assess the UNET on unsee data.



## 2.3 Producing the projection map

### 2.3.1 Applying the trained UNET to the input multiband image

The input multiband image was split into image parcels, then supplied to the trained UNET to produce transition potential image parcels, and lastly mosaiced into a single band image. The split image parcels were buffered with 32 additional pixels compared to training image parcels, and the buffer pixels were removed before the mosaicking process to reduce the “edge effect.”



### 2.3.2 Allocating the expansion pixels to projection map

The number of the expansion pixels *n* was determined by a regression on the historical urban areas. Then the pixels with a value greater than the *n*th highest were determined as the simulated urban expansion pixels.

