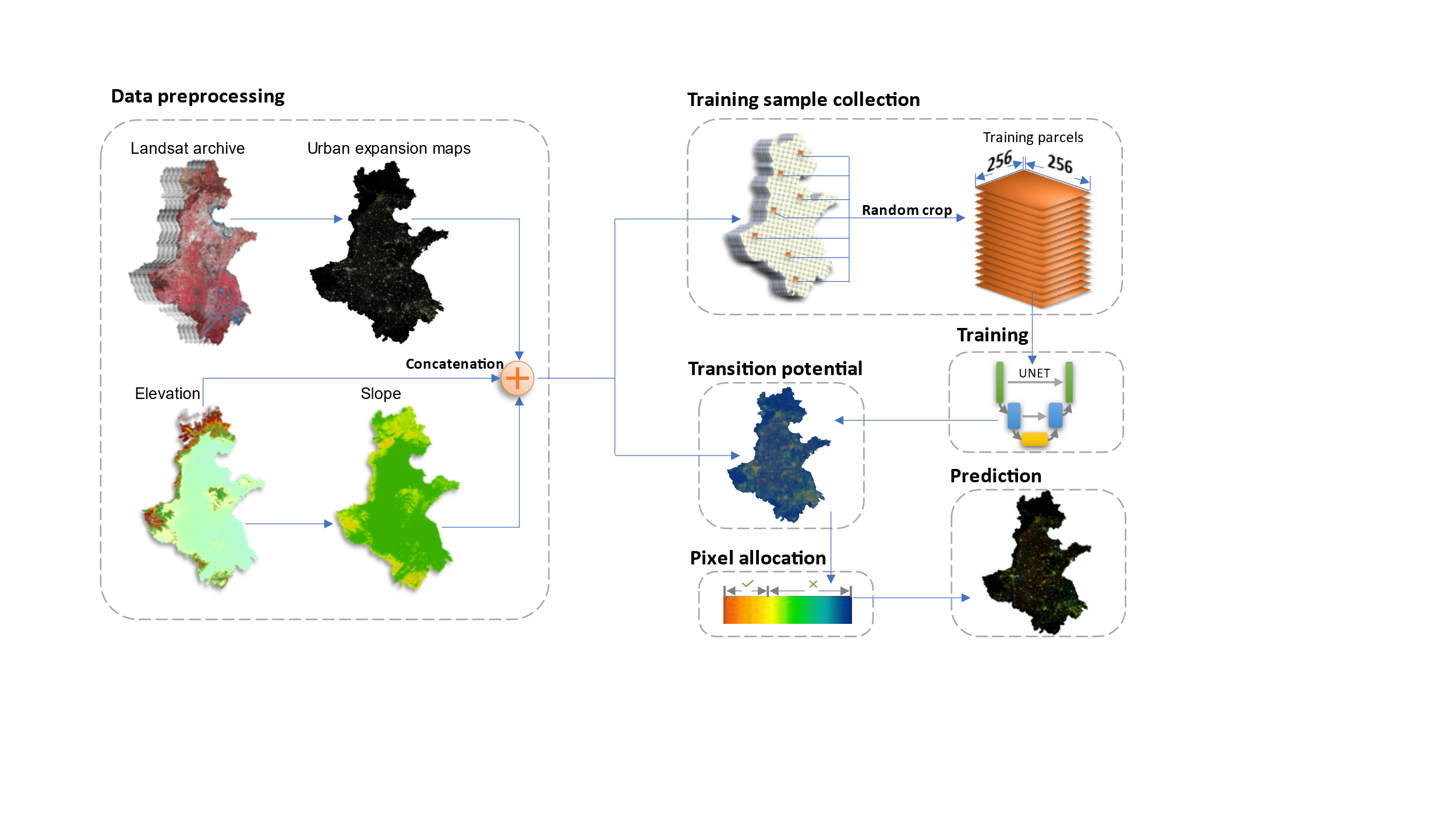
# 2. Data and Method

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). Road networks, Digital Elevation Model (DEM), the derived distance to road, and slope were used as auxiliary data to improve the projection performance. All data were resampled to 30m resolution and then concatenated to a multiband image 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). After sample collection, the UNET was used to train on these samples and produce the transition potential map with the input image. Lastly, a threshold derived from the historical urban area trajectory and determined the pixel with a greater value than this threshold as the urban expansion predictions.



## 2.1 Data preprocessing

### 2.1.1 Urban dynamic mapping

Spectral, temporal, meteorology, and elevation predictors were used to map the urban dynamics of the study area from 1990 to 2019. The spectral predictors were the median value taken from the image collection of each 3-year Landsat archive. The temporal predictors were the coefficients from the Fourier Transformation fitting on the image collection. The meteorology and elevation predictors were used to improve the classification under different climatic and topological conditions. This urban dynamic product has the highest overall accuracy comparing to other products in the study area because the temporal predictors removed the errors classifications from fallow farmland, and a mapping correction algorithm further improved the mapping consistency.

|  |  |  |
| --- | --- | --- |
| **Data Type** | **Source** | **Time Span** |
| Road network | Open Street Map | 2014-now |
| DEM | NASA | 2000 |
| Slope | Derived from DEM | 2000 |
| Urban dynamic map | Wang et. al. | 1990-2019 |

### 2.1.2 Auxiliary data processing

Road network and DEM were used to improve the urban projection. Because the projection model only accepts raster data, the road network was converted to raster data where each pixel indicates the shortest distance to the nearest road. The slope data was derived from the DEM and included in the auxiliary data.

## 2.2 Train the Deep learning model

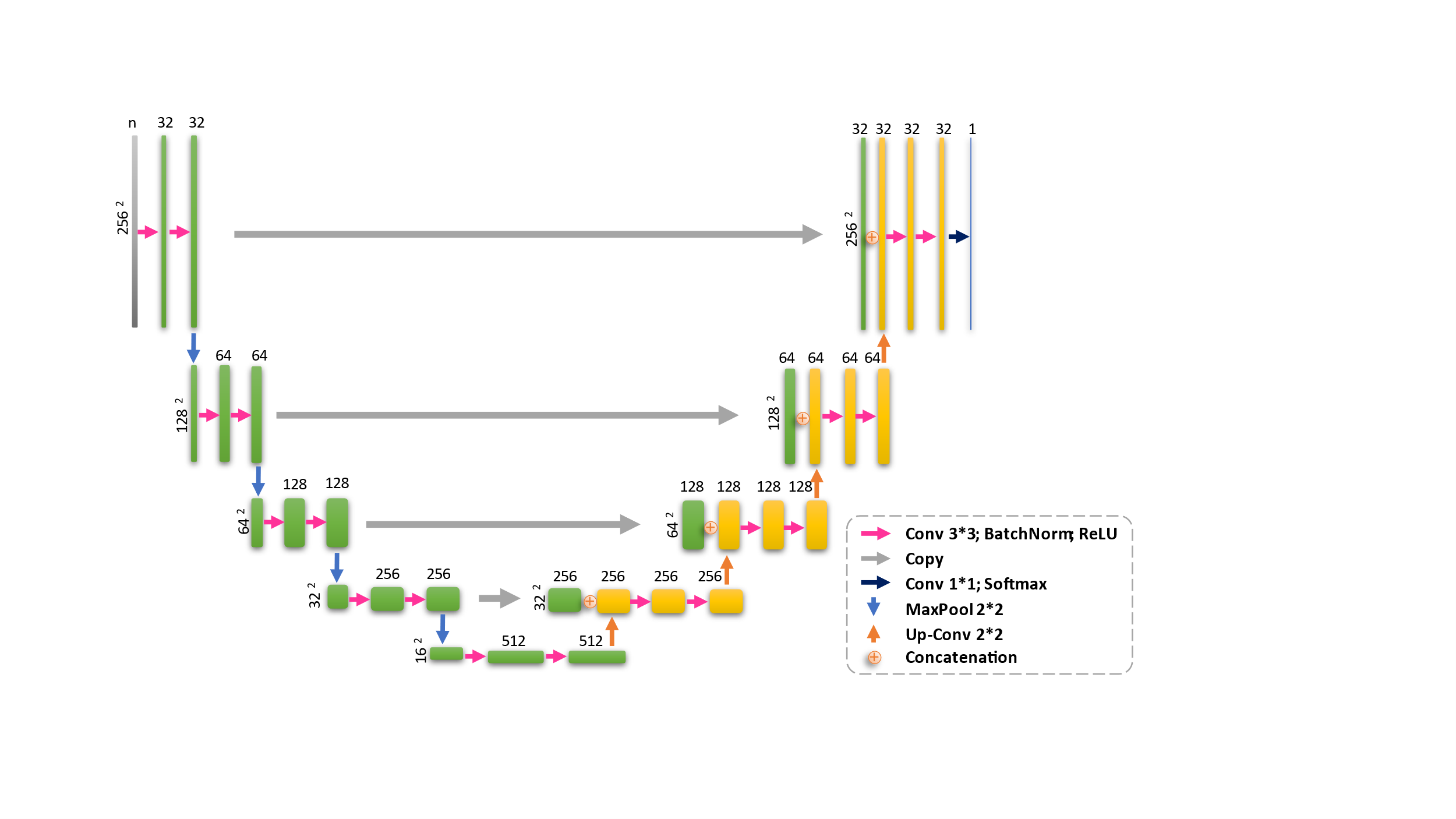
### 2.2.1 Control sample collection

All data were concatenated into a multiband image to improve the efficiency of control sample collection. The *neighborhoodToArray* module of the Google Earth Engine platform was used to turn the neighborhood of each pixel of the multiband input image into a 3-dimensional matrix (i.e., a multiband image parcel). We set the neighborhood size to 256 following the common data science practices and thus the height and width of control samples were 256 pixels. A total of 25,000 sample parcels were randomly collected among which 20,000 were used for training and 5,000 for evaluating purpose.

### 2.2.2 The UNET model for spatial feature extraction

UNET is a deep convolutional network that typically used for image segmentation. Down-sampling, up-sample, and skip-connections are the primary components for UNET. 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 and localization between down-sampling and up-sampling. The output of the UNET is an image that has the same size to the input data, and each pixel of the output image indicates the probability belonging to the target category.

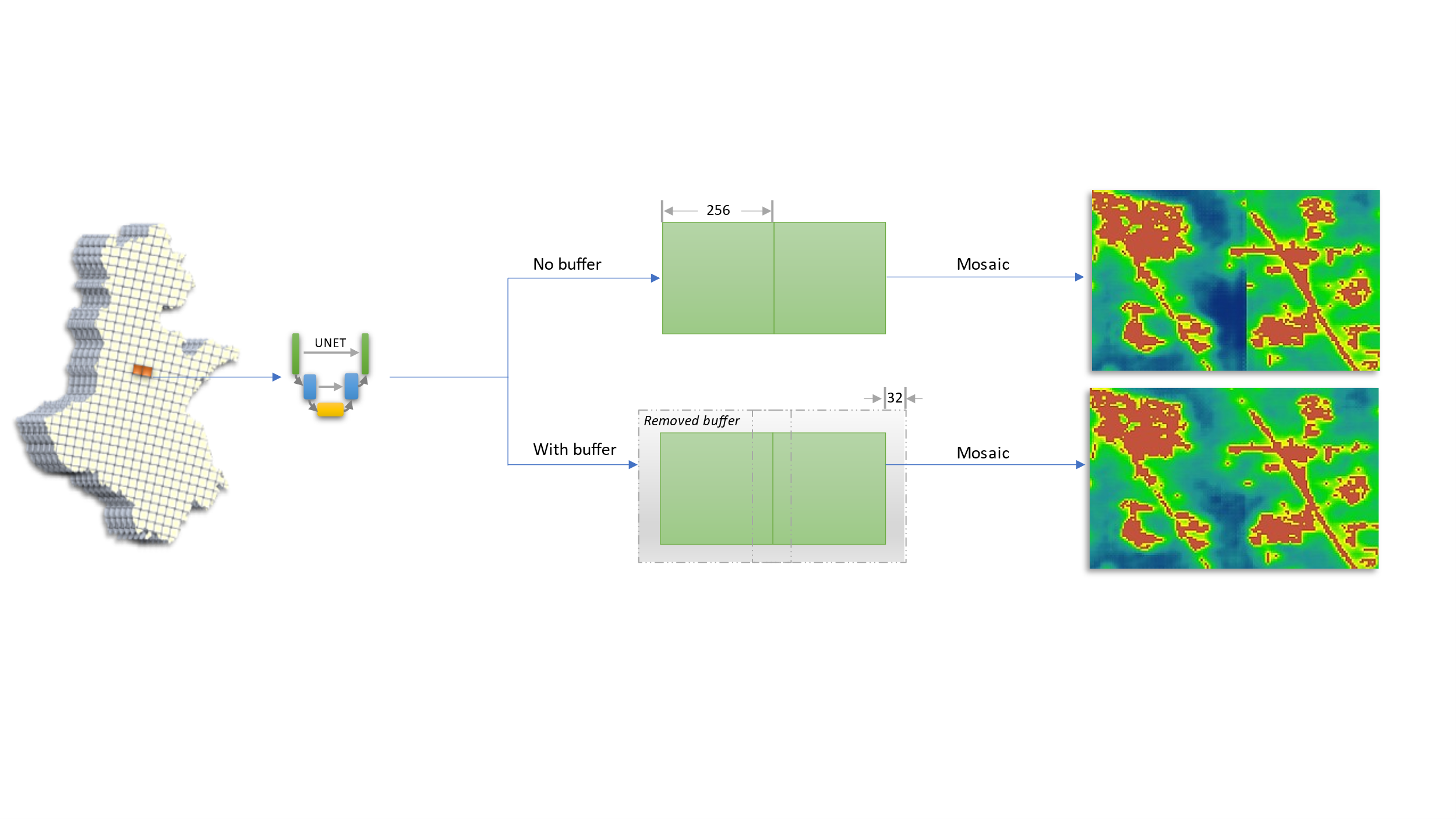
In this study, the UNET was trained on preprocessed image parcels. The input image parcels were comprised by two historical urban maps and the auxiliary data. The output image is a single band image where each pixel indicating a probability to be an urban pixel in the projection data. During the training process, a target image in a later data was used to calibrate the model by minimizing the difference between the predicted image with the target image.



## 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.

