ICIP 2019

STAMPNET: UNSUPERVISED MULTI-CLASS OBJECT DISCOVERY

Joost Visser, Alessandro Corbetta, Vlado Menkovski, Federico Toschi

EINDHOVEN UNIVERSITY OF TECHNOLOGY

ABSTRACT

Unsupervised object discovery in images involves uncovering recurring patterns that define objects and discriminates them against the background.

In this work, we propose StampNet, a novel autoencoding neural network that localizes shapes (objects) over a simple background in images and categorizes them simultaneously.

StampNet consists of a discrete latent space that is used to categorize objects and to determine the location of the objects. The object categories are formed during the training, resulting in the discovery of a fixed set of objects.

METHOD

The encoder is a multi-layered convolutional neural network that develops a representation of the image. The decoder works in two stages (Fig 1.): a **selection and localization (SL) layer** and a **stamp layer**. The role of the SL layer is to assign specific stamps to particular locations based on the encoder representation (Fig 2.). The stamp layer is a simple convolutional layer that produces the reconstructed image. The SL layer's produces an activation map that is 'one-hot' (i.e. activates a single stamp on a single location) for each shape that we aim to detect (Fig. 3). This results in the last layer learning filters that act as stamps.

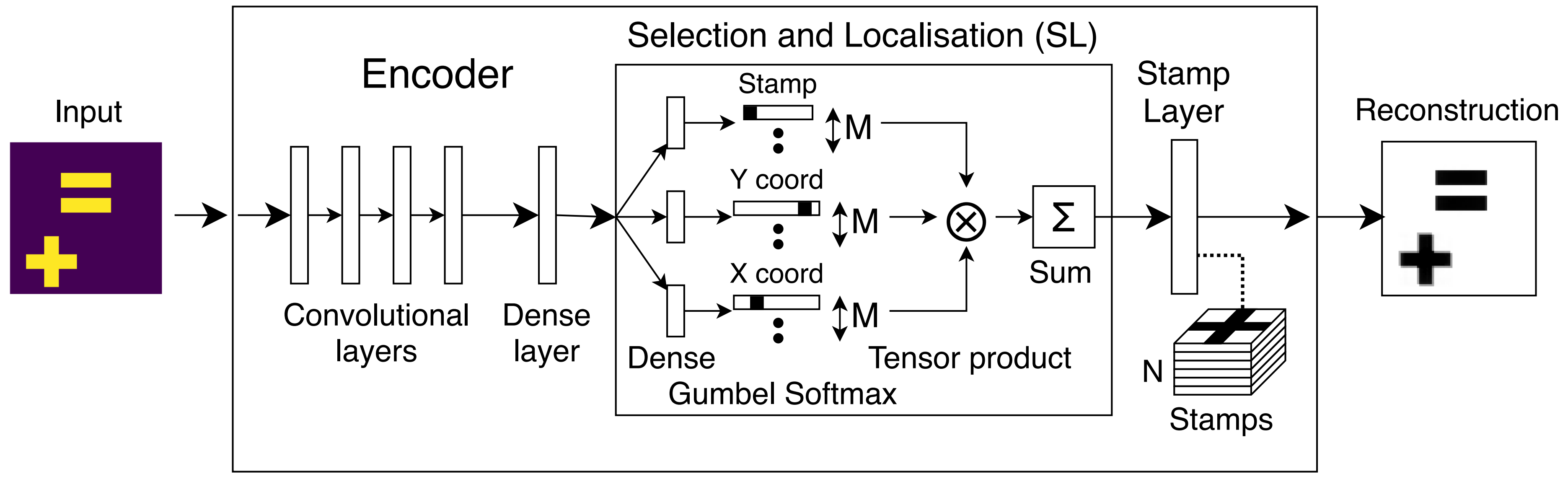

Fig.1. The StampNet architecture

The advantage of the StampNet architecture is that we can directly interpret the model's inferred objects by visualizing the stamp layer. We can also directly observe the localization of these objects directly from the SL activation map.

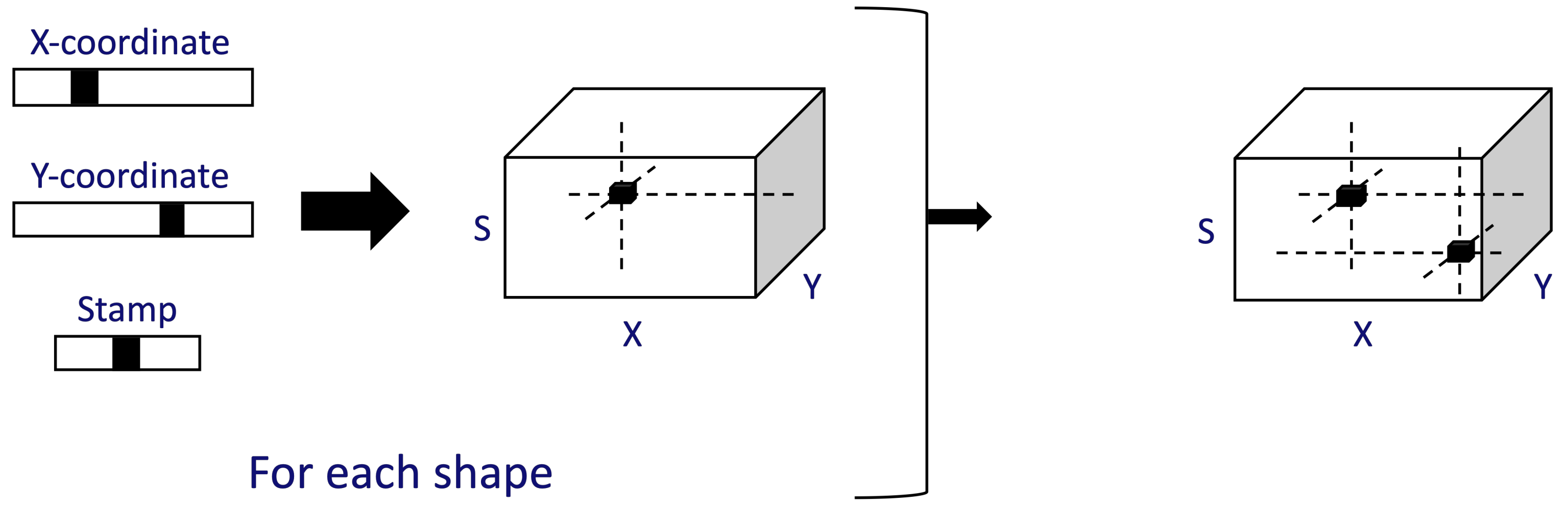

Fig.2. The selection and localization layer

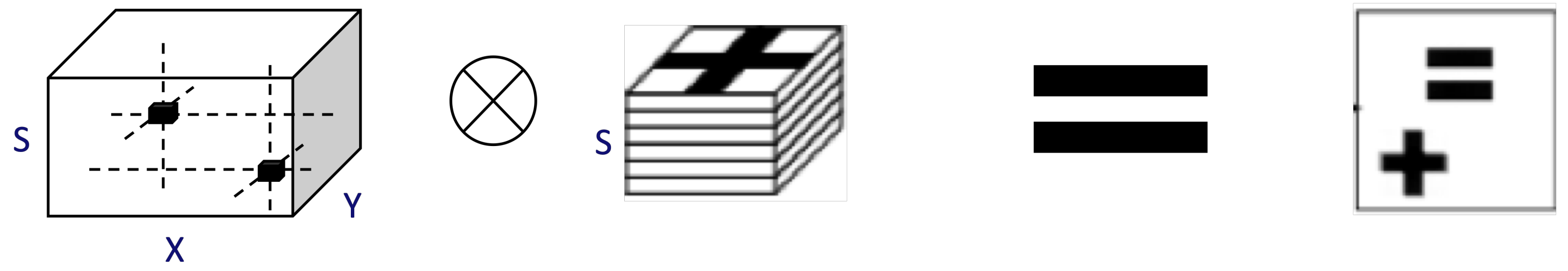

Fig.3 The SL action map and stamp layer generating the output

RESULTS

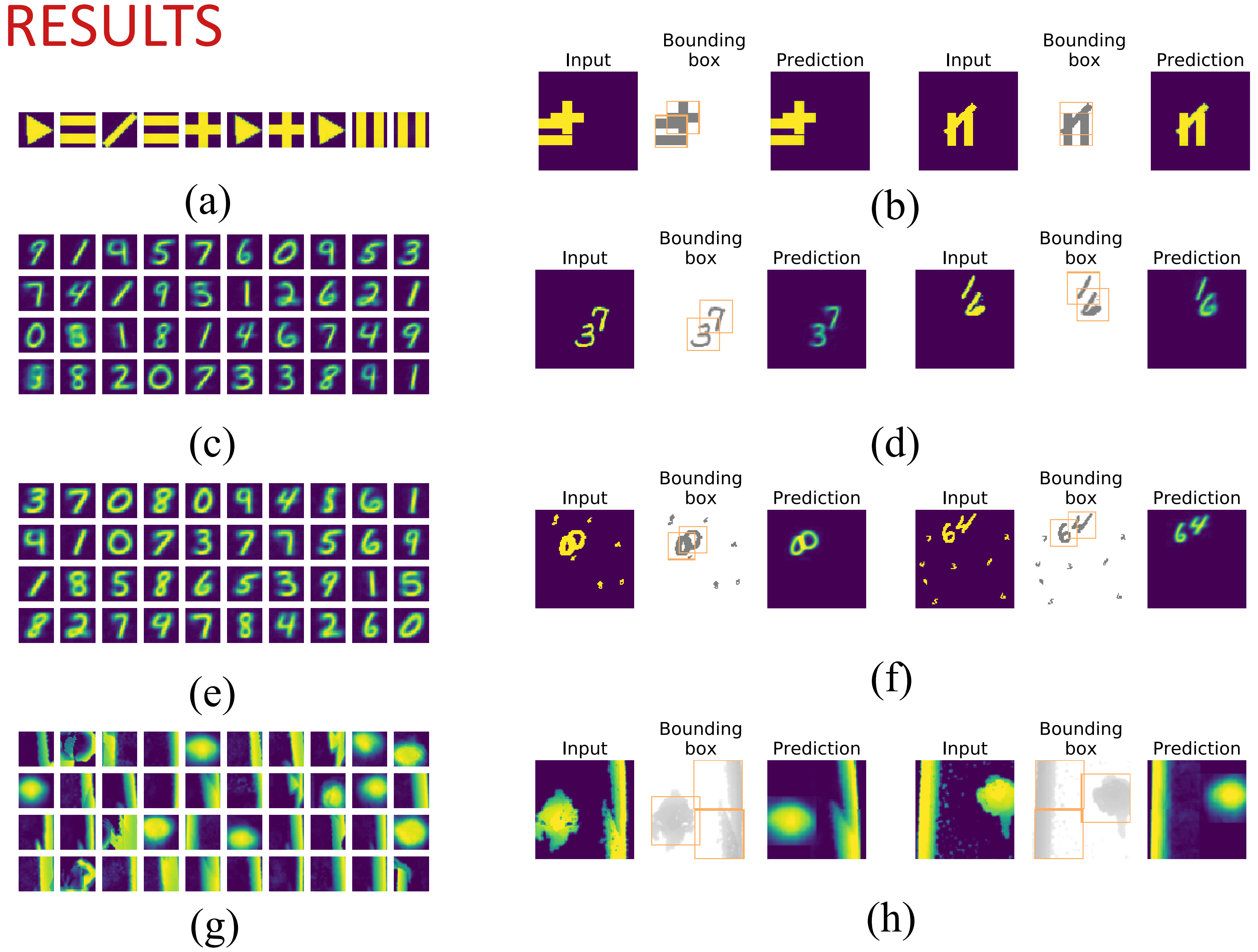

Fig.4 Results on four datasets: *Simple Shapes* (a,b), *Translated MNIST* (c,d), *Cluttered Translated MNIST* (e,f) and *Pedestrian Tracking* (g,h). Left: (a,c,e,g) stamps learned by our model. Right: (b,d,f,h) samples of each dataset.

CONCLUSIONS

The results in Figure 4d and 4f show that StampNet is able to detect and localize overlapping MNIST digits without the need for any labels. Furthermore, the network clusters the shapes in the dataset as stamps (Figure 4e). We demonstrate an example of the value of this in an application of pedestrian tracking in overhead images.