

Towards Storytelling with Animated Pictorial Map Objects – An Experiment with Convolutional Neural Networks

Raimund Schnürer^{a,*}, Cengiz Öztireli^b, René Sieber^a, Lorenz Hurni^a

^a Institute of Cartography and Geoinformation, ETH Zurich, {schnuerer/siebert/lhurni}@ethz.ch

^b Computer Graphics Laboratory, ETH Zurich, cengizo@inf.ethz.ch

* Corresponding author

Keywords: Convolutional Neural Networks, Pictorial Maps, Storytelling, Animation

Abstract:

Storytelling is a popular technique applied in many fields including cartography. On the one hand, stories can be told intrinsically by map elements per se. An often quoted example in this regard is Minard's map of Napoleon's Russian Campaign (e.g. Denil 2017) which depicts the loss of troops in a spatio-temporally aligned Sankey diagram. On the other hand, stories can be conveyed extrinsically by multimedia elements aside the map. For instance, the travel route of a soldier during the First World War can be shown on a temporally navigable map and accompanied with photos, videos, diary entries, and military forms (Cartwright & Field 2015). In this experiment, we follow a mixed approach where human figures on the map will be animated and address the map reader via speech bubbles. As source data, we consider pictorial maps from digital map libraries (e.g. the David Rumsey Map Collection¹) and social media websites (e.g. Pinterest²). These maps contain realistically drawn representations which are in our opinion very suitable for communicating personal narratives.

We present a workflow with convolutional neural networks (CNNs), a type of artificial neural network primarily used for image recognition, to detect human figures in pictorial maps. In particular, we use Mask R-CNN (He et al. 2017) for identifying bounding boxes and silhouettes of figures. For the segmentation of body parts (i.e. head, torso, arms, hands, legs, feet) and the detection of joints (i.e. nose, thorax, shoulders, elbows, wrists, hip, knees, ankles), we combine the U-Net architecture (Ronneberger et al. 2015) with a ResNet (He et al. 2015). In a final step, we implement a simple 2D animation of waving and walking characters and add speech bubbles near head positions. As a first training dataset, we created parametric SVG character models with different postures originating from the MPII Human Pose Dataset³. The second training dataset contains real image human body parts from the PASCAL-Part Dataset⁴. Humans from both datasets are placed randomly on pictorial maps without any other figures. Preliminary results show that the validation accuracy is the highest when synthetic and real training datasets are combined. We implemented the CNNs with TensorFlow's keras API⁵, whereas training data and animations are generated with the web browser.

Our approach enables giving storytellers a physical presence and anchoring them spatially within the map. By animating characters, we can gain the map reader's attention and guide him/her to special and possibly hidden places (e.g. in touristic maps). By telling personal stories, we may raise the interest of people to explore the maps (e.g. in museums) and give a better understanding of the often abstractly encoded information in maps (e.g. in atlases). When a certain aesthetic value has been reached, pictorial objects may also generate positive emotions so that anxieties about the complexity of data may become secondary (e.g. in education). Overall, the goal of our work is to engage map readers, give them valuable support while studying a map, and create long-lasting memories of the map content.

¹ <https://www.davidrumsey.com/blog/2015/3/31/over-2-000-pictorial-maps-in-online-collection>

² <https://www.pinterest.ch/search/pins/?q=pictorial%20map>

³ <http://human-pose.mpi-inf.mpg.de/>

⁴ http://www.stat.ucla.edu/~xianjie.chen/pascal_part_dataset/pascal_part.html

⁵ <https://www.tensorflow.org/guide/keras>



Figure 1. Bounding box and silhouette detection of human figures in a pictorial map



Figure 2. Body part segmentation (right) on pictorial validation data (left)



Figure 3. Joint detection on synthetic validation data



Figure 4. Anticipated waving animation and speech bubble placement