Active Learning for Delineation of Curvilinear Structures

Agata Mosinska-Domanska  
CVLAB, EPFL  
CH-1015 Lausanne  
egata.mosinska@epfl.ch

Raphael Sznitman  
ARTORG, University of Bern  
CH-3010 Bern  
raphael.sznitman@artorg.unibe.ch

Pascal Fua  
CVLAB, EPFL  
CH-1015 Lausanne  
pascal.fua@epfl.ch

ABSTRACT
Many recent delineation techniques rely on supervised Machine Learning algorithms, which normally require extensive amounts of annotated ground-truth data. We propose an Active Learning approach that significantly speeds up the process of annotation of curvilinear structures. Contrary to conventional methods, it takes into account the geometric specificities of the problem, reducing the training set by up to 80% without compromising the final reconstruction quality.

CCS Concepts
• Theory of computation—Active learning  • Computing methodologies—Object detection.

Keywords
delineation; machine learning; active learning; biomedical imaging; aerial images

1. INTRODUCTION
Reconstructing complex curvilinear structures from images is a challenging task due to their different appearances dependent on the kind of structure and imaging modality. Recently, it has been shown that employing Machine Learning to distinguish structures of interests can greatly enhance the reconstruction results. While this approach proved to be effective, it requires significant amount of annotated training data, which can be troublesome to obtain especially in some domains such as biomedical imaging.

We propose an Active Learning (AL) approach, which considerably reduces the annotation time and effort. The key idea behind AL is that it is possible to select a small subset of the most informative samples and label only those. While most of the conventional techniques originate from fields different from Computer Vision and thus do not take into account the geometrical information during query, our method uses the assumption that paths to be annotated form a graph, which helps us better estimate informativeness of each sample.

2. METHOD
Samples which have to be annotated by the user correspond to edges of an overcomplete graph, such as the one presented in Figure 1. In Random Sampling (RS) each of them is assigned to a positive or negative class in a random order. After acquiring a new label we can retrain the classifier and access its predictions. In Uncertainty Sampling (US), at every iteration we choose to query the sample that is characterised by the largest probability entropy.

In our Density Probability Propagation (DPPS) [2] approach, after each iteration we regularise probabilities by propagating them in the neighbourhood of an overcomplete graph and computing entropy only after this step. This is expected to act as a better surrogate of informativeness. Moreover, we combine regularised entropy with density measures, which ensure that the selected samples are not only informative, but also diverse and representative for the underlying distribution.

3. RESULTS
We evaluated our method on four datasets including both natural and microscopic images and comparing it to state-of-the-art RALF [1] and ID [3] algorithms.

![Figure 1. Left: an overcomplete graph of positive and negative edge samples. Right: a learning curve for Neurons dataset](image)

As can be seen in Table 1, our approach outperforms state-of-the-art baselines. Moreover, the selected queries often coincide with ambiguous samples from the human point of view, such as connections between two overlapping branches or discontinuities. This results not only in improved classification performance, but also a better final reconstruction quality.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Roads</td>
<td>0.808</td>
<td>0.817</td>
<td>0.822</td>
<td>0.816</td>
<td>0.835</td>
</tr>
<tr>
<td>Axons</td>
<td>0.818</td>
<td>0.821</td>
<td>0.821</td>
<td>0.816</td>
<td>0.836</td>
</tr>
<tr>
<td>Neurons</td>
<td>0.625</td>
<td>0.653</td>
<td>0.665</td>
<td>0.658</td>
<td>0.697</td>
</tr>
<tr>
<td>Vessels</td>
<td>0.942</td>
<td>0.944</td>
<td>0.943</td>
<td>0.948</td>
<td>0.956</td>
</tr>
</tbody>
</table>

4. CONCLUSION
We propose an approach to including in AL the geometrical information combined with density measures, which increases the effectiveness of annotation process. Our method shows superior performance when compared to various state-of-the-art AL techniques and evaluated on wide range of datasets.

5. REFERENCES