Yet Another Study on Active Learning and Human Pose Estimation

2nd ICML Workshop on Human in the Loop Learning – 2020

Sinan Kaplan[1], **Lasse Lensu**[1]

[1] Computer Vision and Pattern Recognition Laboratory, Lappeenranta-Lahti University of Technology LUT, Lappeenranta, Finland

Abstract: Active learning has not received much attention in the field of human pose estimation compared to image classification and object detection tasks. In this paper, a practical active learning strategy, ***currently under testing in an industrial online environment****, is proposed. An overview of the implemented strategy is presented along with initial results*

Motivation

- To achieve a reasonable level of accuracy for the pose-detection model with as little data as possible,
- To reduce sampling time by using **approximate nearest neighbors** instead of exhaustive search methods,
- To provide an analysis of practical challenges in an online environment

Proposed Active Learning Cycle

Th implemented AL strategy is illustrated below.

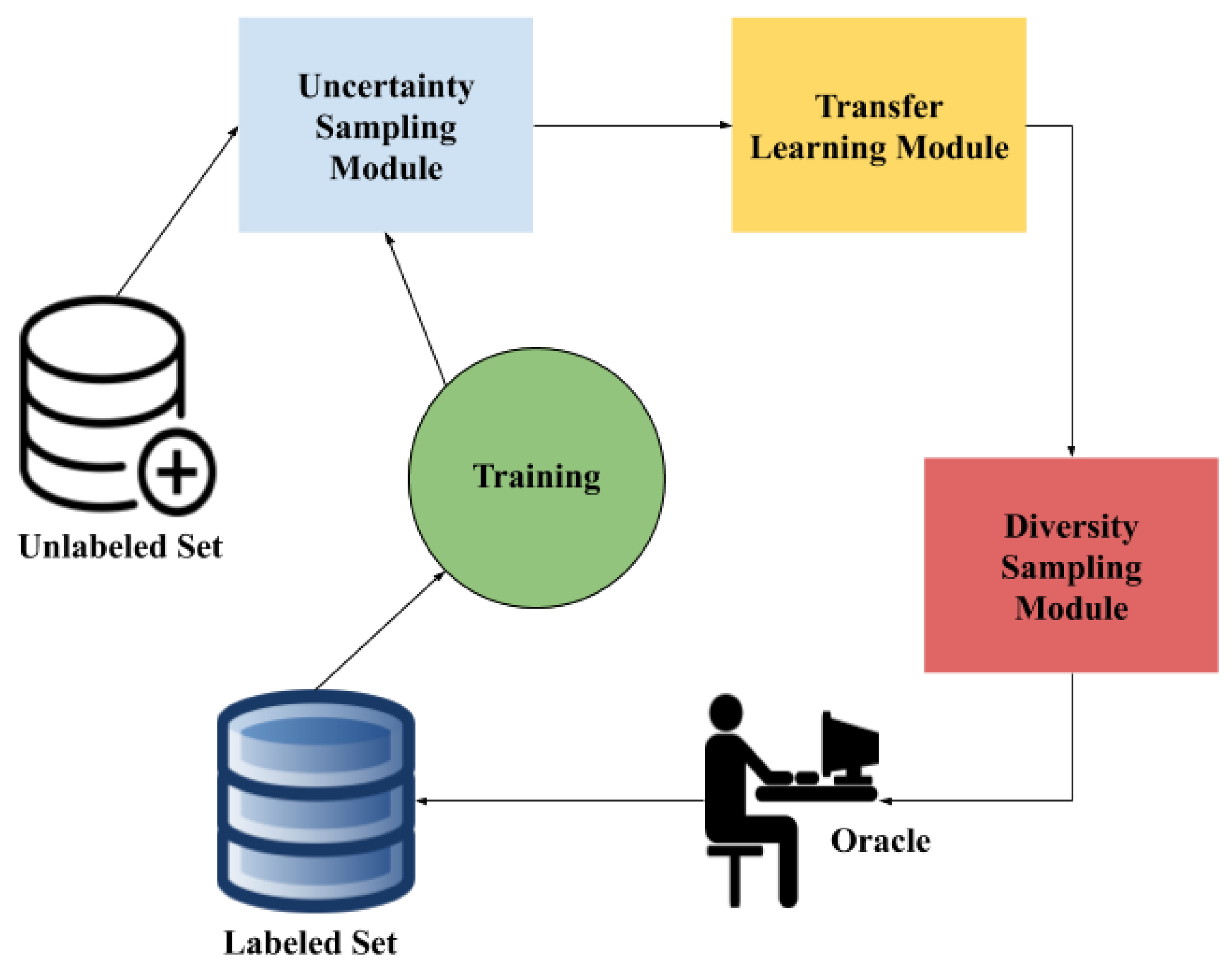

Figure 1. An overview of the active learning procedure

- ***Diversity Sampling Module:*** filtering data samples based on heatmap activation values.
- ***Transfer Learning Module:*** computing embedding features from a Resnet50 model trained on ImageNet data.
- ***Diversity Sampling Module:*** constructing approximate nearest neighbor tree (Spotify, 2020) to filter data samples diversely.

Experiments

We show the proposed sampling strategy on the COCO validation set to qualitatively assess whether the strategy picks diverse samples .The OpenPose-plus model (Tensorlayer, 2020) is used as human pose estimation model. Uncertainty sampling is carried with the trained model on the first batch of data provided by the company. In Figure 2, the samples with low and high activation from the model outputs are shown.

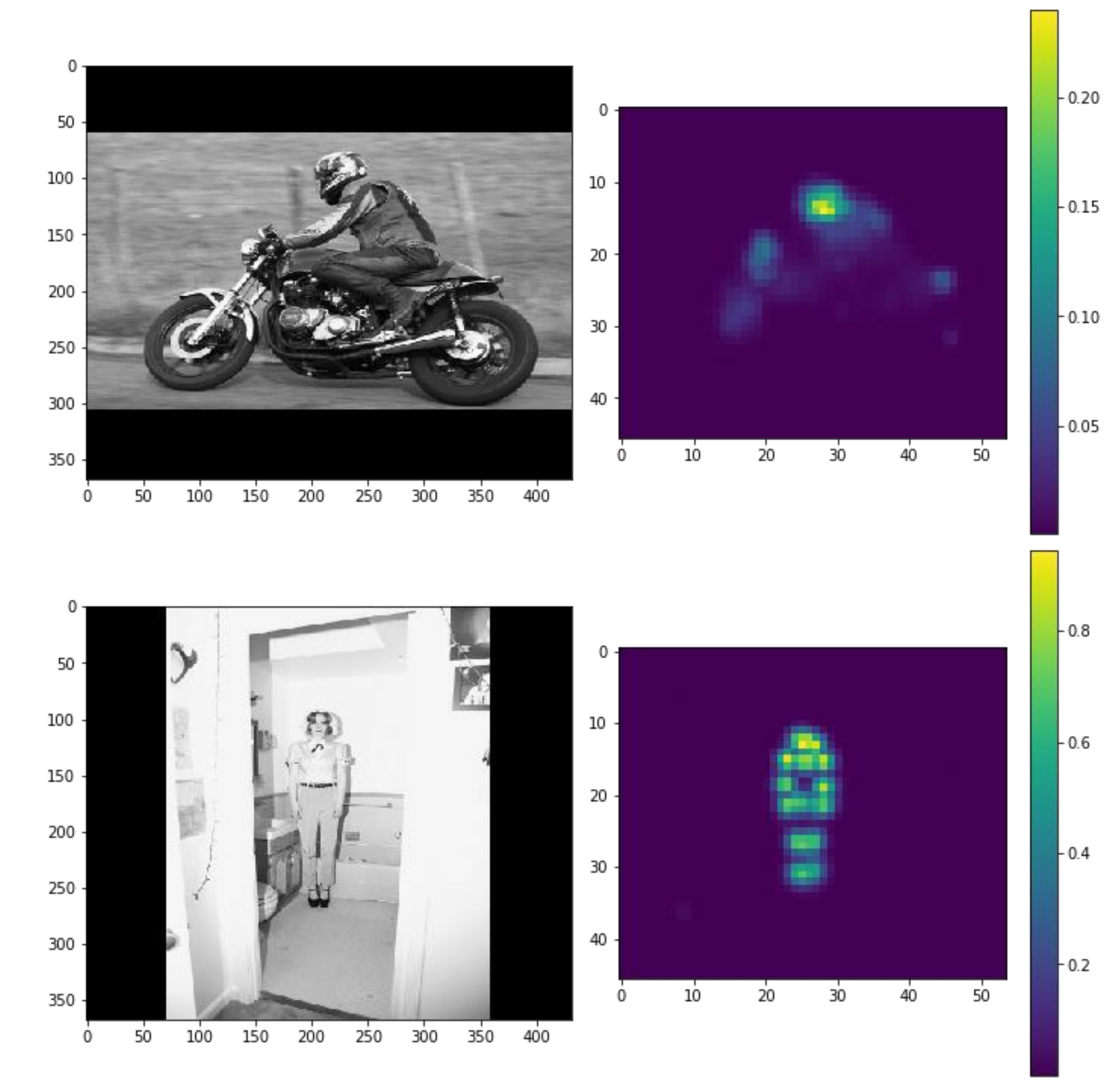

Figure 2. Examples of low and high heatmap activations. The first row shows an image with a corresponding low activation values and the second row demonstrates an image with a high activation. The first image will be filtered by uncertainty module for annotation.

The diverse samples are identified by the diversity sampling module. Here, 20% of the samples are chosen to be annotated. The chosen diverse samples from the COCO validation set is demonstrated in Figure 3.

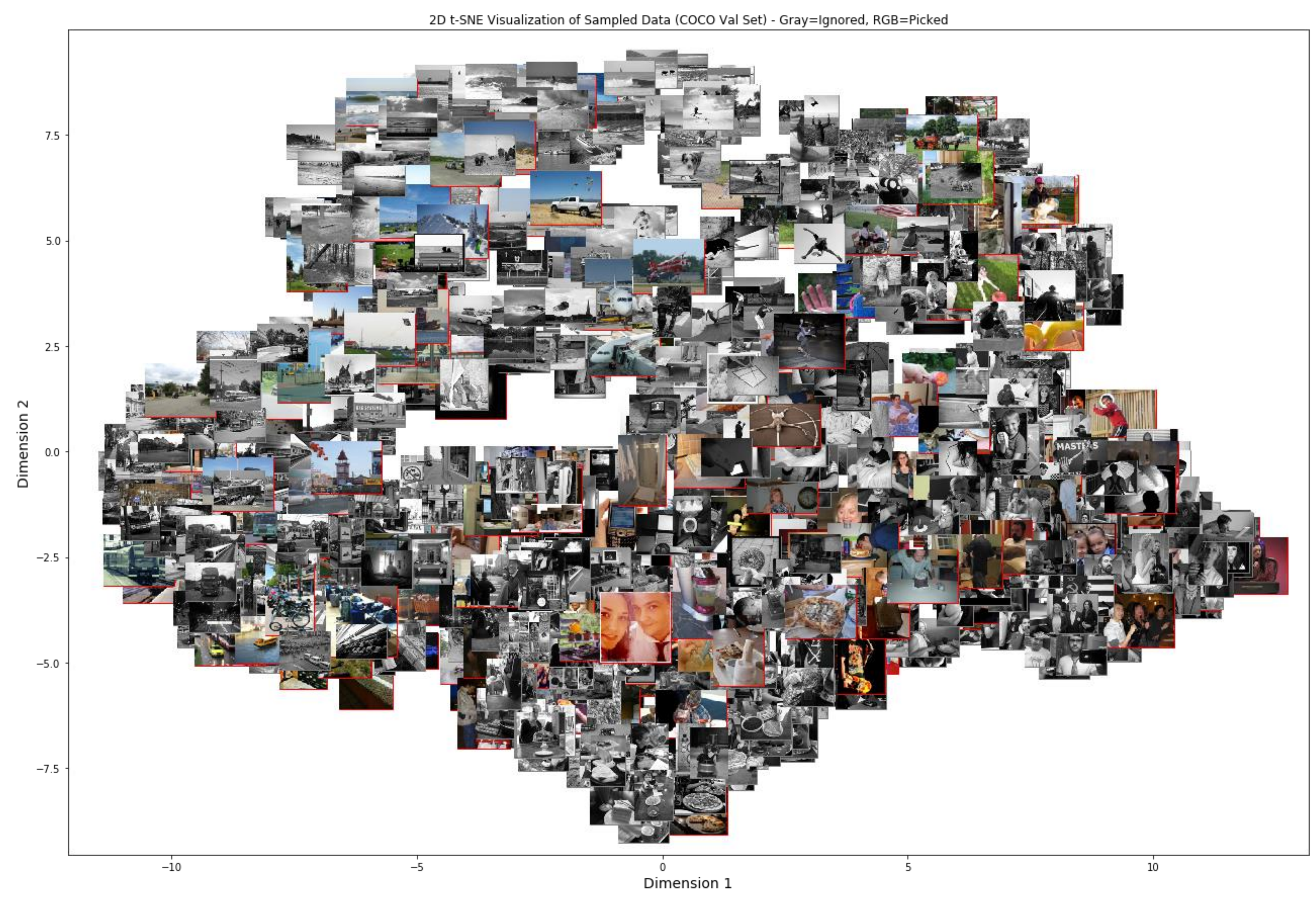

Figure 3. The results from diversity sampling module on the COCO validation set. The images shown in color are the ones chosen for the annotation based on approximate nearest neighbors search.

Discussion

Based on the initial quantitative assessments on COCO set, we can say that our approach is able to select diverse samples successfully. We believe more diverse samples enable us to get a more accurate pose model. To validate this claim further, we will report **ongoing test results** accordingly.

Reference(s)

Spotify. spotify/annoy, May 2020. URL https:// github.com/spotify/annoy.

Tensorlayer. tensorlayer/hyperpose, Jun 2020. URL https://github.com/tensorlayer/ hyperpose.

Sinan Kaplan
Computer Vision and Pattern Recognition Laboratory, Lappeenranta-Lahti University of Technology LUT, P.O. Box 20, 53850
Lappeenranta, Finland
sinan.kaplan@lut.fi

Lasse Lensu
Computer Vision and Pattern Recognition Laboratory, Lappeenranta-Lahti University of Technology LUT, P.O. Box 20, 53850
Lappeenranta, Finland
lasse.lensu@lut.fi