

U-Net based convolutional neural network for skeleton extraction

Oleg Panichev
Ciklum
Ukraine, Kyiv, Amosova str., 12
pol e@ci kl um. com

Alona Voloshyna
Ciklum
Ukraine, Kyiv, Amosova str., 12
avol @ci kl um. com

Abstract

Skeletonization is a process aimed to extract a line-like object shape representation, skeleton, which is of great interest for optical character recognition, shape-based object matching, recognition, biomedical image analysis, etc.. Existing methods for skeleton extraction are typically based on topological, morphological or distance transform and are known to be sensitive to the noise on the boundary and require post-processing procedure for redundant branches pruning. In this work, we introduce U-net based approach for direct skeleton extraction of the object within Pixel Skel-NetOn - CVPR 2019 challenge, inspired by CNNs success in skeleton extraction from real images task. The main idea of our approach is to consistently edit a skeleton mask by feature propagation through different scale layers. It opposes final skeleton generation from different scale object shape representations as occurs in approaches with deep supervision for skeleton extraction from the real image.

Our U-net based model showed 0.75 F1-score on the validation set and the ensemble of eight identical models, trained on different data subsets, got 0.7846 F1-score on the test data.

1. Introduction

Skeletonization is a process aimed to extract a line-like object shape representation, skeleton, allowing the reconstruction of original object shape. [18] gives such skeleton definition: “The skeleton S is a geometric graph, which means that S can be decomposed into a finite number of connected arcs, called skeleton branches, composed of points of degree two, and the branches meet at skeleton joints (or bifurcation points) that are points of degree three or higher.” Skeletonization is used for optical character recognition [16], object matching and recognition [23], biomedical image analysis: vessel system geometrical and structural analysis and surgery planning [14, 17], lungs tree analysis [1], etc.

A Skeleton-based object descriptor aggregates geome-

try, symmetry and topology of its shape [9, 10]. Skeleton should contain the centers of maximal disks (medial axis points) lying inside of the object and touching boundaries at least in two points, which is used for object shape reconstruction. This shape descriptor is required to be invariant to translation, scale and rotation, since these transformations do not change the shape of the object. Demir et al. in [3] pointed out the main challenges of object skeletonization: dimensionality reduction while transforming the shape to the skeleton; the transition to continuous domain to get the best skeletal representation; the trade-of between skeleton simplicity and shape representational power.

There are three classical main ways of skeleton extraction: morphological thinning based on iterative boundary removal, geometric methods based on Voronoi diagram, distance transform based methods [4]. A good survey on skeletonization methods is provided by [16]. But such methods are sensitive to the noise on the boundary and require post-processing procedure for redundant branches pruning [16, 18].

According to the recent great success of convolutional neural networks in different computer vision tasks: classification, segmentation, object detection etc., their ability to represent data in the latent feature space could be used for direct skeleton extraction of the object without further pruning.

2. Related work

There are plenty of mathematical methods for skeleton extraction from the object shape [16], and a very few devoted to skeleton extraction based on Neural Networks. Skeletonization problem could be considered as a per-pixel classification problem known as semantic segmentation. This idea was adopted by Holistically-Nested Edge Detection (HED) method [22] – combination of fully convolutional network (FCN) and deep supervision. Authors referred to the preference of multiple scale predictions combination for final edge map generation. Most of the existing CNN-based architectures for real images skeletonization are based on HED architecture.

