

SkeletonNet: Shape Pixel to Skeleton Pixel

Sabari Nathan
Couger Inc.
sabari@couger.co.jp

Priya Kansal
Couger Inc.
priya@couger.co.jp

Abstract

Deep Learning for Geometric Shape Understanding has organized a challenge for extracting different kinds of skeletons from the images of different objects. This competition is organized in association with CVPR 2019. There are three different tracks of this competition. The present manuscript describes the method used to train the model for the dataset provided in the first track. The first track aims to extract skeleton pixels from the shape pixels of 89 different objects. For the purpose of extracting the skeleton, a U-net model which is comprised of an encoder-decoder structure has been used. In our proposed architecture, unlike the plain decoder in the traditional U net, we have designed the decoder in the format of HED architecture, wherein we have introduced 4 side layers and fused them to one dilation convolutional layer to connect the broken links of the skeleton. Our proposed architecture achieved the F1 score of 0.77 on test data.

1. Introduction

The extracted skeletons from the images are widely used in various areas like computer vision and image processing for optical character recognition [17], fingerprint recognition [28], motion detection [14], object tracking [13], etc. Skeletons are also widely used in life sciences for plant morphology [4]. Deep Learning for Geometric Shape Understanding at CVPR 2019 has organized SkelNetOn challenge. In this challenge, a pre-segmented image dataset with the corresponding skeleton representations in three tracks is provided [25]. The first track has posed the challenge of extracting the skeleton pixels from the given pre-segmented images [25][16][19][24]. We have approached this challenge as an edge detection problem and introduced a version of HED architecture in the decoder part of our proposed architecture. The rest of the sections of this manuscript describe the dataset, related work, methodology and results of the model used to secure 3rd place in the challenge.

2. Related Work

Skeleton extraction is a widely researched area in the last 10 years. However, the most recent works are mainly focused on the extracting skeleton from the RGB images [22][11], which involves segmentation or detection of the objects and extract the skeleton at the same time. Also, an extensive research is done either on edge detection [8][3][27][23] or segmentation [27][10] individually. These kinds of works do not suit fully to the present task. Some initial works are done on the extracting skeleton from the pre-segmented images [2][1][5] [9] which is similar to our task. However, most of these works are focused on the skeleton pruning to remove the unwanted branches rather than skeleton extraction. In the work done by [7], the authors introduced the boundary noise to avoid the uninformative branch creations. [15] used skeleton strength maps (SSM) which are calculated by the isotropic diffusion of the Euclidian distance transformation of binary images and their gradient. After calculating the SSM, they connected all the local maxima points of SSM with the shortest possible line to extract the skeletons. [6] approached the task of skeleton extraction as image generation model and used the generative adversarial network to extract the skeletons.

We have approached the present task as an edge detection problem and hence our work is more inspired by Holistically-nested Edge Detection (HED) model [26]. Similar to HED architecture, we have also fused the side layers into the final output layer. But to improve the performance of HED, instead of taking the output of convolution layers as side layers, we have introduced CS-SE layers at the end of each up-sampling layer and have considered the output of CS-SE layers as side layers. The detail of our approach is presented in section 3.1.

3. Dataset

The challenge is organized in two phases. In the development phase, 1219 images with their ground-truth for training and 242 images without ground-truth for validation are provided. In the final phase, a total of 266 test images are given. Participants are asked to submit their prediction for validation images in the development phase

