Team Name: IPIU

Authors: Lin Cheng, Ge Li, Xu Liu, Hao Zhu, Wei Zhao, Jing Chen, Xiaoyan Yang, Shubo Wei, Qian Wang, Licheng Jiao.

Approach: Two stage fusion network with multiple neighbor feature ensemble.

In the contest, we proposed a two stage fusion network with multiple neighbor feature ensemble for Data Fusion Classification Challenge. We introduce our method from three aspects i.e. data preprocessing, the main model and post-pocessing.

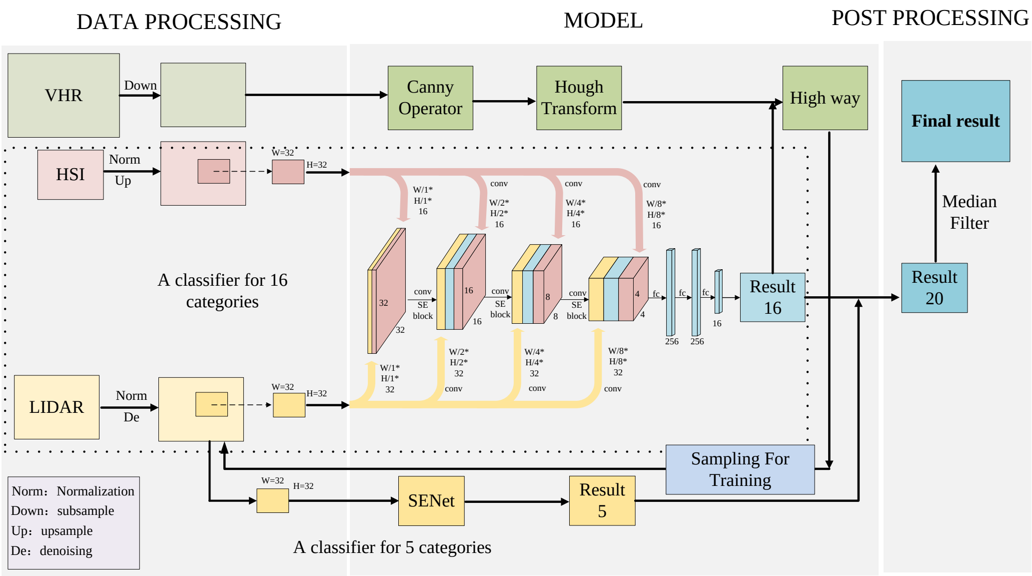

Fig.1 The overall framework

I. Data Preprocessing

For HIS data, we convert the original data into the format of '.tif' by the software of ENVI5.1. To guarantee the pixel-to-pixel alignment between the HSI data and the ground truth, the bilinear interpolation method is utilized to resize the original HSI data with a size of 2404*8344*48, and the 48 channels HSI data is normalized by channels in min-max normalization.

For LiDAR data, the noisy points in the geotiff rasters are replaced with the pixel value next to it, the object height information is obtained by subtracting DEM from DSM.

For VHR RGB data, the original data is resized to the size of 2404*8344*3, which is in accordance with the resolution of ground truth.

II. Model

Stage 1--Multiple Neighbor Feature Ensemble

The main motivation of this framework is that not all objects have the same scale, such as the ninth class non-residential buildings and the twentieth class cars. Generally, the convolutional neural network is used to classify different classes by an image patch with a fixed scale to represent a class of a pixel included in the patch. Therefore, a fusion framework is proposed to combine LiDAR and HSI data, besides, four scales (32*32, 16*16, 8*8, 4*4) data are used to facilitate this classification issue. In addition, to stress the different classes feature information, the SENet [1-2] block is added to the fusion network to learn different weights among various feature map channels.

Stage 2--Road Extraction and Classification

Because of the difference between the train data and the test data, the classification result of roads (the 10^{th}, 11^{th}, 12^{th}, 13^{th}, 14^{th} classes) cannot get a high accuracy in deep convolution neural network, especially the misclassification between the highway and other road categories. By analyzing the VHR RGB data, we find that the width of highway is larger than other roads. Therefore, the edge detection method, called Canny operator, is used to detect the edge of the highway in terms of the VHR RGB data, then we use the Hough transformation to detect lines on the binary skeleton, so distances between all lines are obtained.

According to the fusion framework result obtained above, from which the five kinds of roads (the 10^{th}, 11^{th}, 12^{th}, 13^{th}, 14^{th} classes) are regarded as one category (i.e., road), while the width between two lines can be calculated. So if one pixel is classified as the road, and the width between the two lines is greater than a threshold, we identify that the pixels between the two lines are highways (i.e., the 14^{th} classes). Finally, we add these samples to the fusion network to retrain the network.

Instead of trying to find the best fusion model, A five categories output SENet is trained to overcome the mistake classification between the roads, including the category from 10^{th} to 15^{th},

III. Result Post-processing

The result of stage one can be fused by the classification map from stage two by replacing the category of road. The fusion result is filtering by median filter to get the final result.

IV. Conclusions

The classification map for the whole test data is achieved by combining the three sub-strategies above, and the median filter is adopted to smooth the classification map as the final classification result map. Eventually, we submit the final classification result map and the overall accuracy is 79.92%.

V. Reference

[1] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[J]. arXiv preprint arXiv:1709.01507, 2017.

[2] https://github.com/farmingyard/SENet