Improving Semi-supervised Neural Networks for Scene Understanding by Learning the Neighbourhood Graph

Jiwoong Im and Graham W. Taylor   School of Engineering, University of Guelph

Introduction

- The challenges of scene classification based on remote sensing images are:
  i) The relative scarsity of labeled data for training and evaluation
  ii) Tremendous noise in the data
- Our objective is to enhance scene classification based on semi-supervised neural networks (SSNN) [1] by improving the neighbourhood graph (NG).
- SSNN requires an explicitly constructed NG. It is based on the smoothness assumption that examples that are similar in input space should also be similar in label space.
- NG is constructed based on a k nearest neighbour criterion. A limitation with this approach is that Euclidean distance may not be an ideal measure of semantic similarity.

Background

The SSNN loss function consists of two terms: an error term (labeled information), and a regularizer term (unlabeled information):

$$\mathcal{L} = \frac{1}{l}\sum_{i=1}^{l} V(\mathbf{x}_i, y_i, f) + \lambda_u \frac{1}{l+u}\sum_{i,j=1}^{l+u} L(f_i, f_j, w_{ij})$$

where V is the hinge loss function, W is the edge values from the NG and f is a neural network. Two regularizer term are considered and they are analogous to the LapSVM[3] and TSVM [2]:

$$L = \begin{cases} ||f_i - f_j||^2, & \text{if } W_{ij} = 1 \\ \max(0, m - ||f_i - f_j||)^2, & \text{if } W_{ij} = 0 \end{cases}$$

$$L = \begin{cases} \eta^{(+)} V(\mathbf{x}_i, f(\mathbf{x}_i), c), & c = \text{sign}(f_i + f_j) \\ & \text{if } W_{ij} = 1 \\ -\eta^{(-)} V(\mathbf{x}_i, f(\mathbf{x}_i), c), & c = \text{sign}(f_j) \\ & \text{if } W_{ij} = 0 \end{cases}$$

Method

IDEA: using only labeled examples, we train a classifier on pairs of pixels and predict whether they are from the same class or not. The real-valued classifier confidence is used to form the weights in the NG.

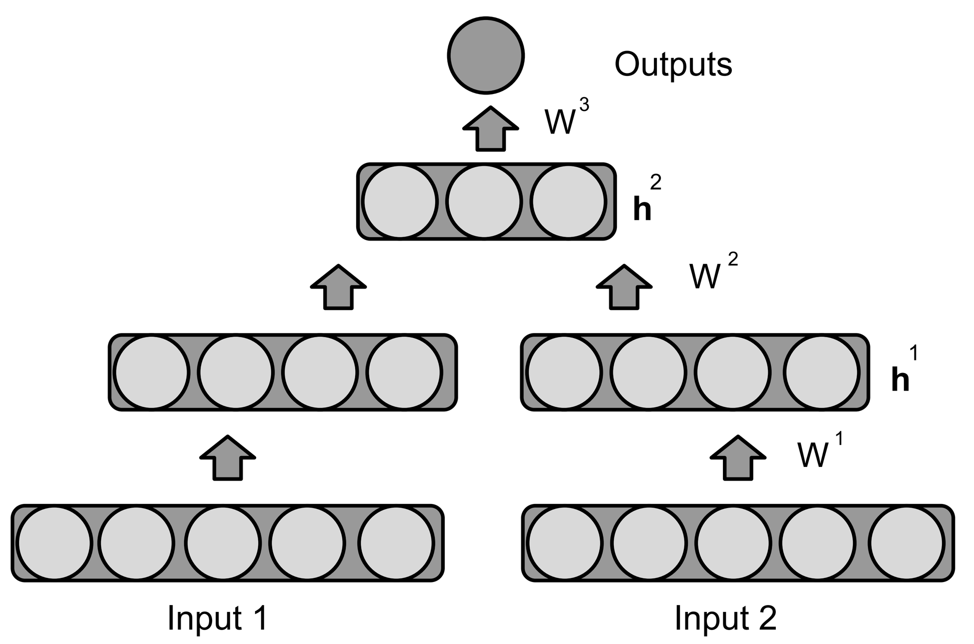

Figure 3: Classical net for neighborhood graph learning (NGL).

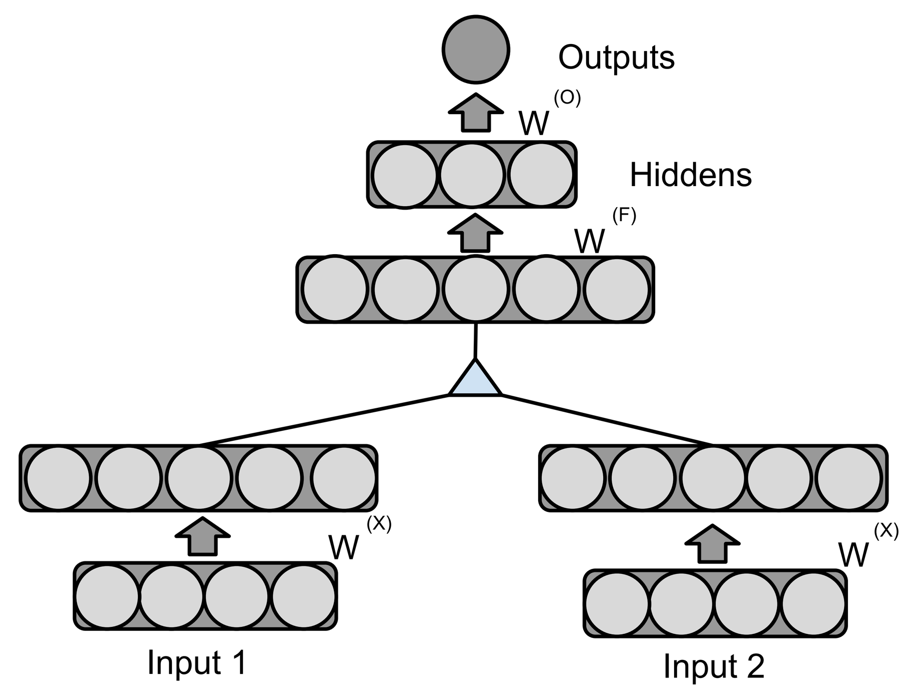

Figure 4: Multiplicative net for neighborhood graph learning (MNGL).

- The first architecture consists of two representation pathways joined by a common hidden layer.
- The first two sets of weights are tied between pathways.
- The second architecture is based on a third-order model which permits multiplicative interactions between representation pathways.
- The third order model learns relationships between transformed pixel intensities.

Experiments

# Labeled Examples	5	10	15	30
TSSNN	29.9 ± 16.13	25.9 ± 7.21	23.8 ± 3.14	21.2 ± 1.92
LapSSNN	**27.3 ± 6.46**	24.0 ± 1.08	23.4 ± 1.58	21.6 ± 1.79
TSSNN+NGL	30.7 ± 11.33	26.7 ± 2.73	25.0 ± 2.38	21.5 ± 2.86
LapSSNN+NGL	29.4 ± 12.62	26.8 ± 3.65	25.9 ± 7.41	20.1 ± 4.52
TSSNN+MNGL	31.0 ± 10.11	27.4 ± 2.59	25.2 ± 2.45	21.0 ± 2.06
LapSSNN+MNGL	30.1 ± 18.26	**23.7** ± **1.78**	**22.6** ± **4.49**	**18.9** ±**1.62**

Table 1: IP6 test set - classification error rates (%).

# Labeled Examples	5	10	15	30
TSSNN	26.8 ± 2.79	24.7 ± 2.09	23.7 ± 5.24	21.6 ± 1.46
LapSSNN	27.3 ± 2.82	25.9 ± 1.35	25.1 ± 4.07	24.2 ± 1.28
TSSNN+NGL	27.5 ± 3.84	24.7 ± 1.23	23.0 ± 3.08	20.8 ± 2.38
LapSSNN+NGL	**23.2** ± **1.45**	**22.2** ± **1.84**	**20.5** ± **2.48**	**19.7** ± **0.87**
TSSNN+MNGL	23.6 ± 4.65	22.2 ± 3.15	21.3 ± 6.62	21.3 ± 1.02
LapSSNN+MNGL	26.4 ± 5.01	25.1 ± 1.38	23.7 ± 3.33	21.0 ± 4.15

Table 2: PaviaU test set - classification error rates (%).

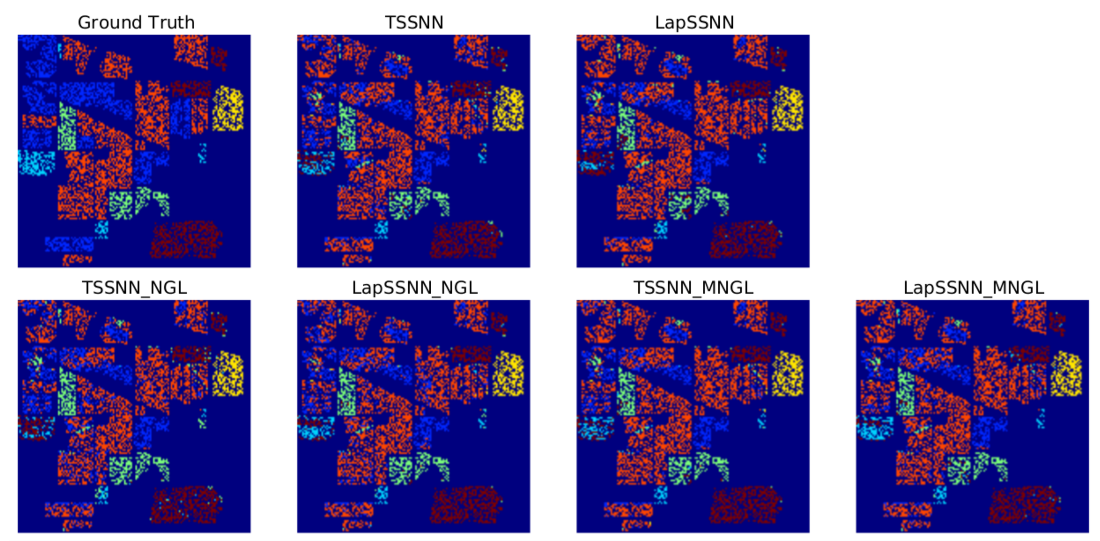

Figure 1: The top left scene mask is the ground truth of the IP6 dataset, and the next six images are the predicted scene masks using various types of SSNNs trained with 30 labeled examples.

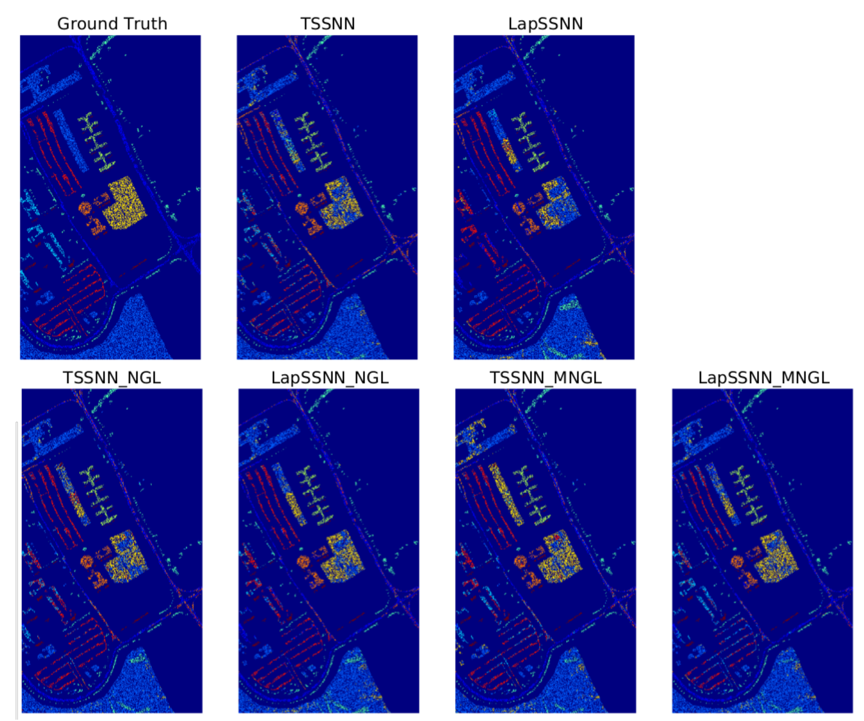

Figure 2: The top left scene mask is the ground truth of PaviaU, and the next six images are the predicted scene masks using various types of SSNNs trained with 30 labeled examples.

Future Research

- Experimenting with a single rather than two-stage architecture for jointly learning the NG and the HSI classifier
- Investigating unsupervised representation learning for building the NG

References

[1] Frédéric Ratle, Gustavo Camps-Valls, Jason Weston. Semisupervised Neural Networks for Efficient Hyperspectral Image Classification, *Geoscience and Remote Sensing Letters, IEEE.*

[2] Lorenzo Bruzzone, Mingmin Chi, Mattia Marconcini. A novel transductive SVM for semisupervised classification of remote-sensing images *Geoscience and Remote Sensing, IEEE Transactions on.*

[3] Luis Gomez-Chova, Gustavo Camps-Valls, Jordi Muñoz-Marì, Javier Calpe-Maravilla, Semi-supervised image classification with Laplacian support vector machines. *IEEE Signal Processing Magazine* .