PERCEPTUAL LOSS FOR SUPERPIXEL-LEVEL MULTISPECTRAL AND PANCHROMATIC IMAGE CLASSIFICATION

Cheng Shi and Chi-Man Pun

Department of Computer and Information Sciences, University of Macau, Macau, China
email: cmpun@umac.mo

ABSTRACT

Convolutional neural networks (CNNs) have proven to be an effective way for deep feature extraction. However, multispectral and panchromatic images are susceptible to illumination unevenness and noise, and the default cross entropy loss function consider only the local information, resulting in misclassification. In this paper, we propose a novel superpixel-level deep neural networks for multispectral and panchromatic images classification, and define a novel perceptual loss function via non-local spectral and structure similarity to suppress the interference of unbalanced light and noise. We also propose the corresponding iteration optimization algorithm in this paper. Experimental results show that the proposed method performs better than the state-of-the-art methods.

Index Terms— Perceptual loss function, Convolutional neural networks, Spectral similarity constraint, Structure similarity constraint

1. INTRODUCTION

In recent years, very high-resolution satellites were launched frequently [1]. General remote sensing satellites carry both panchromatic and multispectral sensors. Panchromatic image has a high spatial resolution with only one spectral band. Multispectral image usually has four or eight bands, but the spatial resolution is four times smaller than panchromatic image. To better understand the objects, multispectral and panchromatic images are usually combined together for classification [2]-[5].

There are usually two ways used for multispectral and panchromatic images classification. The first one is to pan-sharping the multispectral and panchromatic images first, and the fused images are used for classification (P-to-C) [6 - 9]. It is expected that the pan-sharpened image can improve the classification accuracy, however, some researchers pointed out that the spectral and spatial artifacts in the pan-sharpened image has an inevitable impact on classification accuracy [8]. Another line of work is to extract the features from multispectral and panchromatic images first, and then fuse these features for classification [9-14] (C-to-F). In [10], a graph cut approach was combined with the linear mixture model to capture the relationships between the data at different resolutions iteratively. In [11], convolutional neural networks were applied on multi-local regions of multispectral to exploit the structure information. And then panchromatic image was used to fine-tune the classification map. Robinson et al. [12] compared the effect of these two lines on the classification results. P-to-C methods achieved lower classification accuracy because the results were affected by the pan-sharpened image, such as the spectral distortion problem. The method proposed in this paper belongs to the C-to-F methods.

We note from the above methods that pixel-based methods always lead to the noisy classification results, and superpixel-level classification provides a solution to this problem [11]. Hence, in this paper, the samples are generated based on the superpixels to avoid noisy classification results, and superpixels are the basic classification units. And then, deep neural networks are utilized to extract the features from multispectral and panchromatic images, respectively. Recently, deep learning has proven to be effective in feature learning [13-17]. Convolutional neural networks (CNNs) are one kind of deep neural networks, which have a two-dimensional (2D) form, and can better extract the spatial information. Hence, 2D CNNs [14] are used to extract the spatial feature from panchromatic image in our method. Multispectral image is a 3D spectral-spatial cube, 2D CNNs cannot exploit the space cube structure adequately, and therefore 3D CNNs [18] are applied on multispectral image to extract the spatial-spectral feature. Finally, the features extracted from multispectral and panchromatic images are fused together by the designed fusion rule.

Specially, in our previous work [19], superpixel-based 3D CNNs are used for hyperspectral image classification. In this work, traditional cross-entropy loss was used to compare the difference between the output and ground-truth. Hence, only the local information was considered for classification. Multispectral and panchromatic images are satellite images with higher spatial-resolution, and therefore the local information may not represent the characteristics of the objects accurately. However, global spatial dependency of the image can represent more complex local relationship [20]. Actually, there are many similar patches in multispect-

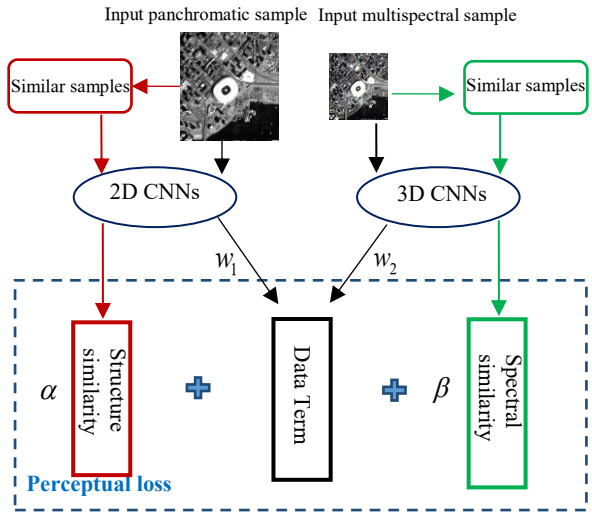


Fig.1. The framework of the proposed networks. We train the structure feature and spectral feature using the 2D CNNs and 3D CNNs, respectively. A perceptual loss function is defined by the similar samples in feature space to fully exploit the spectral and spatial information of multispectral and panchromatic images.

ral and panchromatic images. These similar samples can provide more efficient constraints for local samples. The nonlocal similarity constrain was usually used in image denoising [21] and image restoration [22], and ideal performances were obtained. In this paper, we train the proposed classification framework by an optimization perceptual loss function. If several samples are similar in spatial and spectral domain, they should also have similar characteristics in feature domain [23]. For each sample, we construct the spectral similarity constraint from multispectral image and structure similarity constraint from panchromatic image in higher-level feature space. These constraints can force the similar samples to have the same label, and increase the robustly of the proposed multispectral and panchromatic images classification neural networks (MPCNNs).

The rest of the paper is organized as follows: section 2 shows the proposed model in detail. The experimental results are shown in section 3. Section 4 concludes this paper.

2. PROPOSED MODEL

2.1. Genetate superpixel-level samples

Since pixel-level classification approaches always leads to noisy classification results, especially for high-resolution remote sensing data. Hence we take superpixels as the basic classification units to maintain the spatial consistency in the classification map. Panchromatic has a high spatial resolution, and hence in this paper, superpixel algorithm, entropy rate segmentation (ERS), is applied on panchromatic images to generate the superpixels. And then, we map the superpix-el map of the panchromatic image to multispectral image to obtain the low-resolution superpixel map. We take each superpixel as the central of a spatial window, and map the window to the panchromatic and multispectral images to extract the panchromatic and multispectral samples respectively.

2.2. Image classification via optimized perceptual loss function

In this section, we will introduce the two processes of MPCNNs: forward propagation and back propagation processes.

2.2.1. Forward propagation

Take the 3D multispectral sample x and 2D panchromatic sample y as the inputs of MPCNNs. Since multispectral sample has a 3D form, we apply the 3D CNNs on the 3D multispectral sample x to capture the deep spectral-spatial information better.

$$F_{MS} = G_{3D}(x) \tag{1}$$

where $G_{3D}(\bullet)$ means the forward propagation process of 3D CNNs, and F_{MS} is the deep spatial-spectral feature of multispectral 3D sample.

Panchromatic image has a 2D form, and hence 2D CNNs are applied on the panchromatic sample to extract more detailed spatial feature.

$$F_{PAN} = G_{2D}(y) \tag{2}$$

where $G_{2D}(\bullet)$ is the forward propagation process of 2D CNNs, and F_{PAN} is the deep spatial feature of panchromatic 2D sample, which has the same dimension with F_{MS}.

We set two connected weight matrixes w_1 and w_2 to combine the 3D spectral-spatial features and 2D detailed spatial feature. And then input the combined features into softmax classifier.

$$F = w_1 F_{MS} + w_2 F_{PAN} \tag{3}$$

$$\hat{c}_{i,a} = \frac{e^{\theta_a^T F^i}}{\sum_{c \in Class} e^{\theta_c^T F^i}} \tag{4}$$

where θ is the trainable parameters of softmax classifier, and $\hat{c}_{i,a}$ means the probability that the $i-th$ sample belongs to the class a. The position where the class with the maximal probability is set as the output label.

2.2.2. Back propagation

To train the MPCNNs, we define a novel loss function via nonlocal similar sample. Spatially similar samples should remain similar in the feature space, and hence, we design the spectral and spatial similar constraints on loss function to improve the classification accuracy. For each sample, we search its spectral similar samples from the multispectral image and spatial similar samples from the panchromatic image, respectively.

Although multispectral image has low spatial resolution, it has high spectral information. For each multispectral sample, we search its similar samples by calculating their spectral similarities.

$$S_1^*=\{i+|\mathrm{Corr}(\mathrm{Mean}(x_i)-\mathrm{Mean}(x_{i+}))<\mathrm{threshould}\} \quad (5)$$

where S_1^* is a set containing the spectral similar samples to x_i, $i+$ is the subscript of the sample, and $Corr(\bullet)$ is the correlation coefficient operator. For each sample, we calculate the spatial mean value in each spectral dimension, and compare the mean values of the other samples in the whole image by the correlation coefficient. If they have small differences in mean value, they are consider as spectral similar samples.

Panchromatic image has more spatial information, and hence we use structural similarity (SSIM) index to evaluate the spatial similarity of two samples.

$$S_2^*=\{i+|(\mathrm{SSIM}(y_i,y_{i+}))<\mathrm{threshould}\} \quad (6)$$

where S_2^* is a set containing the spatial similar samples to y_i .

According to the similar samples and ground truth of the training sample, the loss function is defined in Eq.(7).

$$\begin{aligned}\xi=&\sum_{i\in S}\sum_{a\in Class}-c_{i,a}\ln(\hat{c}_{i,a})\\&+\alpha\|G_{3D}(x_{i,a})-\sum_{i+\in S_1^*}w_{i+,a}^1G_{3D}(x_{i+,a})\|_2^2\\&+\beta\|G_{2D}(y_{i,a})-\sum_{i+\in S_2^*}w_{i+,a}^2G_{2D}(y_{i+,a})\|_2^2\end{aligned} \quad (7)$$

where the first term is data term, and $c_{i,a}$ is the ground truth label. The second term is the spectral similarity constraint, which means the multispectral sample and its linear weighted spectral similar samples are similar in 3D CNNs feature space, and $w_{i+,a}^1=G_{3D}(x_{i+,a})/\sum_{i+\in S_1^*}G_{3D}(x_{i+,a})$. The third term is the spatial structure similarity constraint, that means the panchromatic sample and its linear weighted spatial structure similar samples are similar in 2D CNNs feature space, and $w_{i+,a}^2=G_{2D}(y_{i+,a})/\sum_{i+\in S_2^*}G_{2D}(y_{i+,a})$. α and β is intense parameters. And then the loss function is minimized using gradient descent method to update the parameters in MPCNNs.

To minimize the loss function we propose an alternative optimization method. First, we fix the parameters in 2D CNNs, and minimize the first term and the second term in Eq. (7) to update the parameters in softmax classifier and 3D CNNs.

$$\begin{aligned}\min_{W,G_{3D}}\xi_1=&\min_{W,G_{3D}}\sum_{i\in S}\sum_{a\in Class}-c_{i,a}\ln(\hat{c}_{i,a})\\&+\alpha\|G_{3D}(x_{i,a})-\sum_{i+\in S_1^*}w_{i+,a}^1G_{3D}(x_{i+,a})\|_2^2\end{aligned} \quad (8)$$

Second, we fix the parameters in 3D CNNs, and minimize the first term and the third term in Eq.(7) to update the parameter in softmax classifier and 2D CNNs.

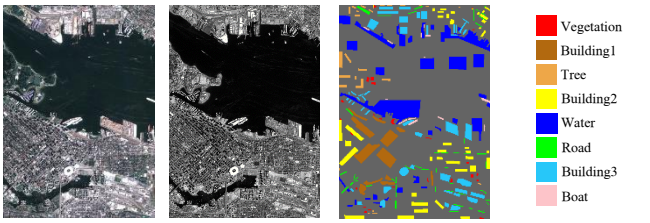
$$\begin{aligned}\min_{W,G_{3D}}\xi_1=&\min_{W,G_{3D}}\sum_{i\in S}\sum_{a\in Class}-c_{i,a}\ln(\hat{c}_{i,a})\\&+\beta\|G_{2D}(y_{i,a})-\sum_{i+\in S_2^*}w_{i+,a}^2G_{2D}(y_{i+,a})\|_2^2\end{aligned} \quad (9)$$

For each step, gradient descent method is applied to solve the minimize problem. When the training process reach the preset iteration times, the forward propagation process get the final classification results.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Datasets

In this section, the dataset *grss_dfc_2016* [24] is used to evaluate the proposed model. *grss_dfc_2016* dataset was provided by The 2016 IRRR GRSS Data Fusion Contest. The multispectral and panchromatic images were acquired by DEIMOS-2 satellite on March 31, 2015 and May 30, 2015 over Vancouver, Canada. The multispectral image contains four spectral bands (red, green, blue, and NIR bands) at 4-m spatial resolution, and panchromatic image contain one band at 1-m spatial resolution. The level 1C image with eight classes is used for classification experiment. The size of the multispectral image is 1311×873, and the size of the panchromatic image is 5244×3492. The false color multispectral image and panchromatic image are shown in Fig.2 (a) and (b). The ground-truth map is shown in Fig.2 (c) and the classes label is shown in Fig.2 (d).



(a) (b) (c) (d)

Fig.2. False color multispectral and panchromatic images. **(a)** False color multispectral image of level 1C image. **(b)** Panchromatic image of level 1C image. **(c)** Ground-truth map. **(d)** Classes label.

3.2. Experimental results of level 1C image

For level 1C image, we use ERS to generate 60000 superpixels. 15% of the labeled superpixels are randomly selected for training, and the remaining labeled superpixels are used for testing. Ten independent experiments are conducted, and the average values of overall accuracy (OA) and kappa coefficient are used to evaluate the classification results. In the experiment, the spatial window size is set as 31×31. The 3D CNNs contain two convolutional layers, two max-pooling layers, and one full connection layer. The first convolutional layer has 20 filters with size $6\times6\times3$, and the second convolutional layer has 40 filters with size

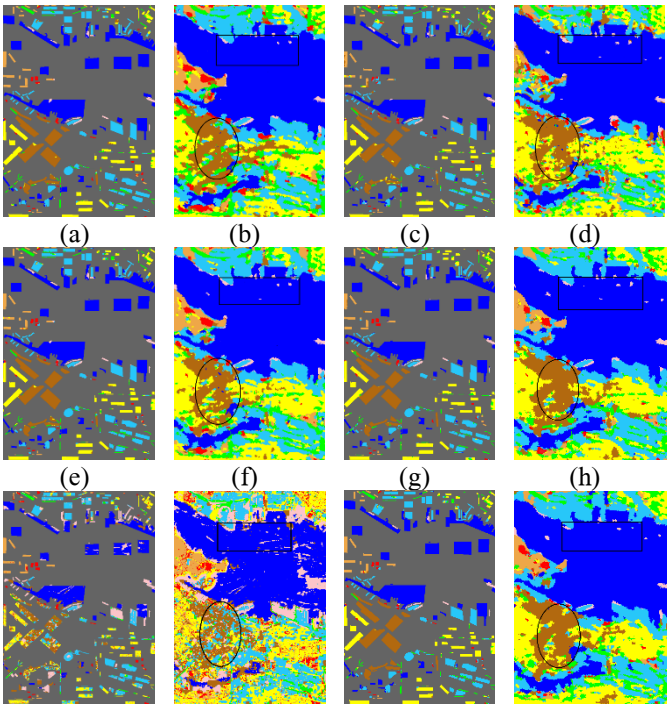


Fig.3. Classification map of level 1C image. **(a)** and **(b)** Multispectral image classification map. **(c)** and **(d)** Panchromatic image classification map. **(e)** and **(f)** Cross entropy loss function classification map. **(g)** and **(h)** Perceptual loss classification map (proposed). **(i)** and **(j)** Classification map of reference [10]. **(k)** and **(l)** Classification map of [11].

Table 1. Classification evaluation of level 1C image

Classes	MS	PAN	Corss-entropy	Perceptual loss	[10]	[11]
1	82.12	47.20	68.26	85.63	**91.27**	71.39
2	97.67	95.51	97.77	97.39	52.44	**98.94**
3	96.39	77.26	88.98	**97.14**	86.35	89.94
4	91.38	94.06	92.57	**95.41**	62.61	94.71
5	98.2	98.94	98.77	98.82	80.23	**98.96**
6	70.05	67.83	74.5	**77.29**	35.60	69.26
7	96.29	91.15	96.7	97.53	64.64	**98.19**
8	79.43	80.19	78.95	**82.16**	83.79	78.49
OA	94.80	92.67	94.93	**96.41**	68.94	95.62
kappa	93.16	90.34	93.32	**95.25**	60.99	94.23

$6\times6\times1$. The full connection layer has 300 units. The 2D CNNs contains two convolutional layers with 20 and 40 filters. The size of each filter is 6×6. The parameters of α and β are set as 0.05 and 0.03, respectively. The threshold in Eq. (5) and Eq.(6) is set as 0.98. If the object does not have its similar samples, the second term and the third term in Eq. (7) will be equal to 0, and the classification only relies on the object itself. For the compared methods, the parameters are selected as default in their papers. The classification results are shown in Fig.3. First, we only use single multispectral or panchromatic image for classification, and the results are shown in Fig.3 (a)-(d). Fig.3 (a) and (c) are the classification maps of the ground-truth area, and Fig.3 (b) and (d) are the classification maps of the whole image. Multispectral image has low spatial resolution, but high spectral resolution. Hence its classification map has better spatial consistency (circle area), but some details is lost (square area). In contrast, panchromatic image contains more detail information, but too much details lead to noises in the classification map. Consider the advantages of multispectral and panchromatic images, we combine them by traditional cross-entropy loss and the proposed perceptual loss for classification, which are shown in Fig.3 (e)-(h). Both loss functions can better combine the advantages of multispectral and panchromatic images, and however, perceptual loss function keeps more detailed information, meanwhile, obtains a smoother classification map. We also compare the proposed model with references [10] and [11]. Reference [10] proposed a graph cut method to Markovian energy minimization to generate classification map on the highest resolution image, and effectively established the relationship between different resolutions. However, this paper belongs to the pixel-level non-deep method, and hence, the classification map shows the noisy classification result in the regions where buildings are clustered (circle region in Fig. 3 (j)). Reference [11] proposed multiple local CNN model for classification, which is the deep-based classification method. This model captures the information of multiple local regions, and uses deep learning to extract more robust features, thereby improving the regional consistency of the classification map. However, due to the global information is not considered in the classification process, some small independent objects are misclassified (rectangular region in Fig. 3 (l)). Table 1 shows the classification evaluates of the compared methods. It is shown that the proposed model achieves higher performance in most classes.

4. CONCLUSION

In this paper, we propose a superpixel-level multispectral and panchromatic images classification framework. Meanwhile, a perceptual loss function is defined to capture the spectral and structure similarities. An iteration optimization algorithm is proposed to solve the perceptual loss. The experimental results show that the proposed model can effectively classify high-resolution remote sensing images with higher accuracies and better regional consistency.

Acknowledgement

This work was supported in part by the Research Committee of the University of Macau under Grants MYRG2015-00011-FST and MYRG2015-00012-FST, and the Science and Technology Development Fund of Macau SAR under Grants 093/2014/A2 and 041/2017/A1. The authors would like to thank Deimos Imaging for acquiring and providing the data used in this paper, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee.

5. REFERENCES

[1] L. Alparone, B. Aiazzi, S. Baronti S, et al. "Multispectral and Panchromatic Data Fusion Assessment Without Reference," Photogrammetric Engineering and Remote Sensing, vol. 74, no. 4, pp. 193-200, 2015.

[2] S. Yang, M. Wang, L. Jiao, "Fusion of Multispectral and Panchromatic Images based on Support Value Transform and Adaptive Principal Component Analysis," Information Fusion, vol. 13, no. 3, pp. 177-184, 2012.

[3] A. G. Mahyari, M. Yazdi, "Panchromatic and Multispectral Image Fusion Based on Maximization of Both Spectral and Spatial Similarities," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 6, pp. 1976-1985, 2011.

[4] W. Huang, L. Xiao, Z. Wei, et al. "A New Pan-Sharpening Method With Deep Neural Networks," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 5, pp. 1037-1041, 2017.

[5] M. A. Shaban, O. Dikshit, "Evaluation of the Merging of SPOT Multispectral and Panchromatic Data for Classification of an Urban Environment," International Journal of Remote Sensing, vol. 23, no. 2, pp. 249-262, 2002.

[6] N. Kosaka, "Forest Type Classification Using Data Fusion of Multispectral and Panchromatic High-resolution Satellite Imageries," IEEE International Geoscience and Remote Sensing Symposium, pp. 2980-2983, 2015.

[7] F. Palsson, J. R. Sveinsson, J. A. Benediktsson, and H. Aanaes, "Classification of Pan-sharpened Urban Satellite Images," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 1, pp. 281-297, 2012.

[8] L. Bruzzone, "Can Multiresolution Fusion Techniques Improve Classification Accuracy?" Image and Signal Processing for Remote Sensing, vol. 6365, 2006, doi: 10.1117/12.691208; http://dx.doi.org/10.1117/12.691208.

[9] T. Mao, H. Tang, J. Wu, W. Jiang, S. He, and Y. Shu, "A Generalized Metaphor of Chinese Restaurant Franchise to Fusing both Panchromatic and Multispectral Images for Unsupervised Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4594-4604, 2016.

[10] G. Moser, A. De Giorgi, and S. B. Serpico, "Multiresolution Supervised Classification of Panchromatic and Multispectral Images by Markov Random Fields and Graph Cuts," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 9, pp. 5054-5070, 2016.

[11] W. Zhao, L. Jiao, W. Ma, et al. "Superpixel-Based Multiple Local CNN for Panchromatic and Multispectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 7, pp. 4141-4156, 2017.

[12] G. D. Robinson, H. N. Gross, J. R. Schott, "Evaluation of Two Applications of Spectral Mixing Models to Image Fusion," Remote Sensing of Environment, vol. 71, no. 3, pp. 272-281, 2000.

[13] Y. Lecun, L. Bottou, "Gradient-based Learning Applied to Document Recognition," Proceedings of the IEEE, vol. 86, no.11, pp. 2278-2324, 1998.

[14] K. Nogueira, O. A. B.Penatti, J. A. D. Santos, "Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification," Pattern Recognition, vol. 61, pp. 539-556, 2016.

[15] X. Ren, Y. Zhou, J. He, et al. "A Convolutional Neural Network-Based Chinese Text Detection Algorithm via Text Structure Modeling," IEEE Transactions on Multimedia, DOI 10.1109/TMM.2016.2625259, 2017.

[16] M. Y. Jiu, H. C. Sahbi, "Deep Kernel Map Networks for Image Annotation, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1571-1575, 2016

[17] Y. R. Zhou, S. B. Song, N. M. Cheung, "On Classification of Distorted Images with Deep Convolutional Neural Networks," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1213-1217, 2017.

[18] S. W. Ji, W. Xu, M. Yang, et al. "3D Convolutional Neural Networks for Human Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, 2013.

[19] C. Shi, and C. M. Pun, "Superpixel-based 3D Deep Neural Networks for Hyperspectral Image Classification," Pattern Recognition, vol. 74, pp. 600-616, 2018.

[20] L. He, J. Li, C. Liu, et al. "Recent Advances on Spectral-Spatial Hyperspectral Image Classification: An Overview and New Guidelines," IEEE Transactions on Geoscience and Remote Sensing, 2017, DOI: 10.1109/TGRS.2017.2765364.

[21] Y. H. Chan, Y. H. Fung, "Two-Direction Nonlocal Model for Image Denoising," IEEE Transactions on Image Processing, vol. 22, no. 1, pp.408-412, 2013.

[22] W. Dong, G. Shi, X. Li, "Nonlocal Image Restoration with Bilateral Variance Estimation: A Low-Rank Approach," IEEE Transactions on Image Processing, vol. 22, no. 2, pp. 700-711, 2013.

[23] J. Johnson, A. Alahi, and F. F. Li, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," Computer Vision and Pattern Recognition, pp. 694-711, 2016.

[24] 2016 IEEE GRSS Data Fusion Contest. Online: http://www.grss-ieee.org/community/technical-committees/data fusion".