Evaluation of the Operational Contribution of the Use of Neural Networks on Hyperspectral Images for the Benefit of Airborne Surveillance

C. Caubet
THALES DMS France
75-77 Avenue Marcel Dassault
33701 MERIGNAC Cedex
France +33 5 24 44 57 54

G. Guerrini
THALES DMS France
75-77 Avenue Marcel Dassault
33701 MERIGNAC Cedex
France +33 5 24 44 57 54

christophe.caubet@fr.thalesgroup.com
gilles.guerrini@fr.thalesgroup.com

P. Desbarats
Univ. Bordeaux, CNRS,
Bordeaux INP, LaBRI, UMR 5800
351, crs de la Libération
33405 TALENCE Cedex France
+ 33 5 40 00 69 18
pascal.desbarats@labri.fr

JP Domenger
Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800
351, crs de la Libération
33405 TALENCE Cedex France
+ 33 5 40 00 69 28
jean-philippe.domenger@u-bordeaux.fr

ABSTRACT

The usage of MultiSpectral and HyperSpectral Imaging (MSI and HSI) has developed a lot during the last decades. The development is possible due to the quantity of information contained in the spectral bands.

Hyperspectral imaging, with its combination of spatial and spectral aspects, allows the localization of materials in a geographical area. The main uses of HSI are pixel classification in remote sensing, especially for the discrimination of multiple types of materials. However, the operational contribution of this kind of data has never been proceeded in frame of airborne surveillance due to weak resolution of the imagery, the heterogeneity of the information….

We have carried out the acquisition of hyperspectral images of vehicles at different conditions, in the visible and very-near infrared (400-1000nm) spectral domains. We conducted a pre-study aiming at dimensioning the use of neural networks by using a simple neural network on the acquired hyperspectral dataset with airborne acquisition conditions.

After segmentation / labelization of data to define the different classes, we have conducted an analysis on the neural network parameters after the training process to understand the criteria used for detection / classification. We run part of the layers on every pixel of every training image.

The network resulted on 55% accuracy for vehicle detection and 30% for vehicle classification with improvement of only a few percents when adding spatial information. The network weights analysis showed that the characteristics learned during training were limited to the average quantity of light in the scene and vegetation.

We concluded that the visible and very-near infrared spectral domains were limited for the spectral analysis of vehicles. In future work, we will focus on the study of vehicles and surveillance materials in the near and short-wave infrared (1000-2500nm) domains.

We have also observed that calibration and external acquisition conditions have a major role for the quality of the information. Future works have to be focused on the topic to improve hyperspectral images

proceeding, identification and classification of materials and objects.

Finally, for future works, using more complex network architecture permitting to analyze simultaneously spatial and spectral contribution in hyperspectral image classification would allow the extraction of new information and possibly make it comparable to state-of-the-art color image segmentation, which benefit highly from spatial information.

1.0 INTRODUCTION

The usage of MultiSpectral and HyperSpectral Imaging (MSI or HSI) has developed a lot during the last decades [1–6]. This development is possible due to the quantity of information contained in the spectral bands. This information, benefiting from a more precise decomposition of wavelengths, allows both the detection of problems that are imperceptible with other imaging methods and an easier isolation of anomalies. The first large-scale use of this decomposition was in 1972 with the launch of the Landsat 1 satellite boarded with a multispectral sensor using four bands: one green, one red, two Near-InfraRed (NIR)[1] [3]. This satellite allowed the acquisition of multispectral data over large areas for the first time. The conducted analyses, particularly on land masses and environment [1, 7], demonstrated the appeal of using MSI. Since the launch of Landsat 1, sensors have developed and some are now able to acquire hundreds of bands simultaneously, resulting in hyperspectral imaging.

As of today, the main uses of HSI are pixel classification in remote sensing, especially for the discrimination of multiple types of materials [3, 18], and medicine [4, 8]. Research is also done in other fields [6], such as tracking and face recognition [9–11]. In [12], the 30 spectral bands of a hyperspectral video are interpolated to restore other bands and frames of the video. In [13], HSI is used to differentiate multiple inks with visually similar colors and much better classification results are obtained than using color (RGB) images of these inks.

However, one of the main obstacles to the use of HSI is the curse of dimensionality. The high dimensionality of HSI heavily increases the number of data to analyze and the computing time. Some methods tackle this issue by empirically removing bands [11, 14, 15] and others by extracting features [2, 3, 16, 17]. However, despite the lesser relevance of the lost information, such as the least used bands in a given classification task, this still leads to decreases in the performances. For this reason, despite a higher cost in time and processed data, other methods use this abundance of information to increase their classification accuracy [18, 19]. These methods mostly use neural networks, which are less impacted by the curse of dimensionality.

Works on neural networks began in 1943 [20], but their popularity increased a lot since the late 90s with [21] and especially this last decade with AlexNet [22], substituting the very popular support vector machines. These gains in popularity are partly due to the arrival of high computing power, optimization techniques, and the availability of numerous and large datasets.

Well-known types of neural networks are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). CNNs use convolutional layers to consider local information rather than only global information. The use of this local information, especially present in natural images, allows the extraction of contours, shapes, and other features [23, 24] in neural networks as would be possible with non-neural descriptors like in [25]. RNNs see their input data as sequences of smaller pieces of data and, as the sequence is being processed, memorize information from the earlier data of the sequence to improve the results on the later data. Uses of RNNs include mostly temporal data, more specifically speech [26] and video processing [27–29]. From the combination of both of these networks emerged the CRNNs [30, 31], which benefit from the advantages of both CNNs and RNNs.

[1] https://landsat.gsfc.nasa.gov/satellites/landsat-1/, last visited 2022-04-25.

Hyperspectral image classification has been addressed with a multitude of trends. Until a few years ago, approaches mostly used Support Vector Machines [32–35] which then evolved into comparison methods with the arrival of neural networks. The prospects of using neural networks to classify hyperspectral images are high since very little is known about the features to extract from hyperspectral images. Some of the most used deep learning approaches to hyperspectral imaging are CNNs [6, 17], which benefit greatly from the spatial dimensions of HSI, and RNNs [6, 18, 19], processing the spectral bands as a sequence of data. Some approaches also use CRNNs [17, 18, 36], benefiting from both local spatial information and sequential spectral information.

If HSI has already been used for tracking in [9], little research has been done about HSI in surveillance. Furthermore, despite its popular use for material discrimination in remote sensing [17, 18, 32, 33], to our knowledge, no research has been done on material discrimination for non-remote sensing applications. In this work, we present three major contributions. First, we design a spectral neural network to extract spectral information from hyperspectral images. Second, after creating and labeling a database of non-remote sensing hyperspectral images of vehicles, we classify these types of vehicles through the designed neural network. Finally, we analyze the networks to extract the most effective features used for the classification.

This paper is organized as follows. Section 2 explains the structure of our network and its parameters. In section 3, we describe the camera we used and the acquisition process. The pre-processing and labeling of the acquired data are explained in section 4. These labeled data are then used in multiple models of the previously described neural network in section 5 and these models are compared with each other. In section 6, analyses are conducted on three models of the neural network. To conclude, a short discussion about this work and further developments is presented in section 7.

2.0 NEURAL NETWORK

2.1 Spectral neural network

The spatial correlation of images is being widely studied and a lot of methods [25, 37, 38], including convolutional neural networks [21, 23, 39], have already been developed to benefit from spatial information. For this reason, in this paper, we focus our study on the spectral correlation of the images.

The designed network, our “spectral neural network”, is a fully-connected dense neural network. A multi-class version of our network, in which three labels Background, Car, and Truck are considered, is depicted in fig. 1. Our network consists of two layers, i.e. one intermediate and one output layer. A single layer is not enough to draw out the full advantage of the non-linearity of neural networks. However, adding more layers, despite being able to extract deeper features, increases the chances of the network overfitting due to the small size of the input data. The spectral information of each spatial pixel is given as an input to the network. This information is then processed through the two layers of the network and the output of the network is the class calculated by the network for the input pixel. For an input hyperspectral image, of which the structure is described in section 3, the network is applied to all spatial pixels of this image. This results in a 2D image that has the size of the spatial dimensions of the input image, in which the value of each pixel is one of the three defined classes.

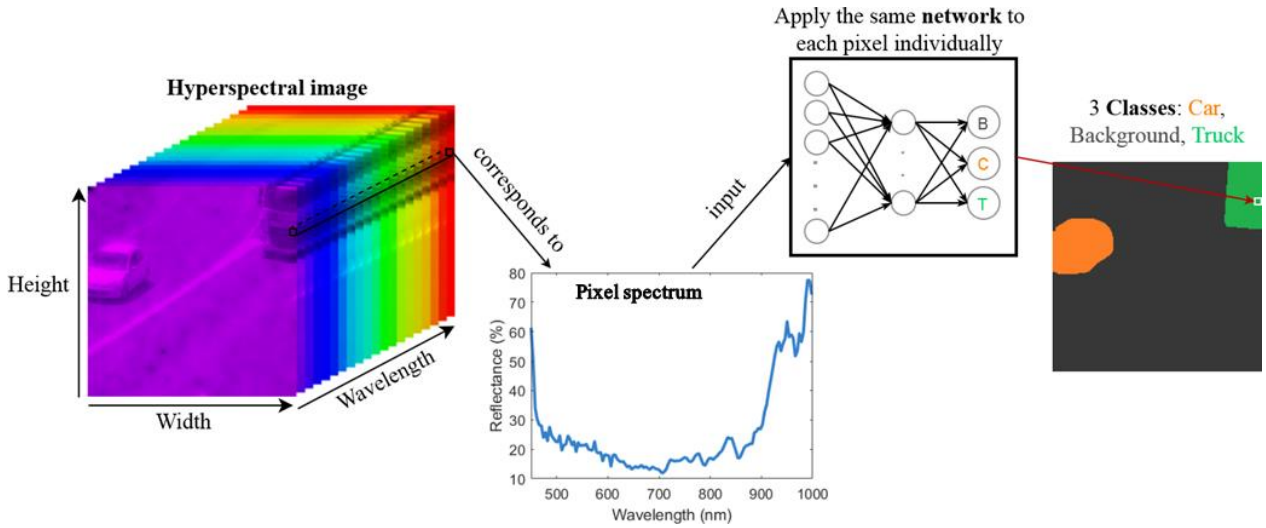

Figure 1: Pixel-by-pixel processing of a hyperspectral image through our spectral neural network.

2.2 Models and parameters

In these experiments, from 400 labeled images, 350 images (87.5%) were chosen for the training set and 50 images (12.5%) for the validation set, containing 200*200 pixels each. Because our network is only compared through variations of its own hyperparameters (cf. Sec. 4) and not with other methods, no test set is considered in our experiments. The 350 training images are pre-loaded before running the networks, because of the size of the images and because the time required to load the images would considerably slow down the training phase if done at each epoch. The networks were designed using PyTorch in Python.

Two independent versions of the network described in Section 2.1 are created. One is a "detection" network that acts as a binary classifier: the three vehicle classes are combined and compared against the background (or non-vehicle) class. The other one, or the "classification" network, is a multi-class network in which the car and truck classes are compared against each other and the background, and the class "others" is ignored because it is severely underrepresented.

All hyperparameters, except the activation functions, have been set for each version of the network. Both versions of the network have two layers. For both versions, we created sixteen models each using one activation function on the intermediate layer and one on the output layer. The activation functions used are ReLU, Leaky-ReLU (LReLU), sigmoid, and tanh. The models are denoted according to their activation functions, e.g. the model named "tanh-LReLU" uses tanh in the intermediate layer and Leaky-ReLU in the output layer.

These models are compared over the Accuracy of the vehicle classes, calculated with True Positives, False Positives, and False Negatives as follows: Accuracy = TP/(TP+FP+FN).

For both of these networks, if the accuracy obtained is sufficient, these networks will prove that the spectral information is pertinent and can be used to improve current detection and classification methods used in surveillance. Further analysis of these networks will also be conducted to understand the features extracted from the hyperspectral images.

3.0 DATA

The images were acquired using the camera FireflEYE S185 from Cubert GmbH, which allows the acquisition of hyperspectral snapshots. This camera acquires instantaneously 138 bands from 450 to 998nm (from blue to NIR). However, the first band (450nm) and the last twelve bands (954-998nm) are noisy and might introduce errors when analyzed. The camera has a spatial resolution of 50x50 pixels, which can be increased by pan-sharpening using the panchromatic 1000x1000 channel of the camera. This processing is needed because labeling is not efficient with 50x50 pixels images. However, because the increase in resolution by pan-sharpening is virtual, it can introduce artifacts and distortions depending on the ratio of the increase in resolution. To allow more efficient labeling while limiting the number of artifacts introduced, we increase the spatial resolution of the images to 200x200 pixels. We acquired images of moving vehicles. These images have been acquired with various azimuth and inclination angles, distance, and lighting conditions. These variations aim to diversify the dataset for it to be more flexible in the neural networks. All images were acquired from above the vehicles, up to an inclination of 70 degrees.

4.0 DATA PROCESSING

To facilitate the labeling processing and to have a better visual representation of the content of the images, color (RGB) images are reconstructed from the hyperspectral data.

Bands over 780nm are removed because they belong to the infrared domain. Red, green, and blue channels are interpolated from the remaining bands, depending on their wavelengths. Due to the unequal representation of red, green, and blue in hyperspectral images and due to the camera not acquiring bands part of the blue wavelengths, the three channels are then normalized separately. The risk of normalizing each channel separately is the over-representation of some of the channels. On another side, only applying a combined normalization leads to a sub-representation of some of the channels.

Examples of reconstructed images are shown in Fig. 2. The sky was visually brighter than the gray tone that can be seen, due to the lack of blue wavelengths (380-450nm), in the reconstruction in Fig. 2right.

Figure 2: Examples of the color reconstruction of images, taken from multiple angles, of (left) a car, (center) a truck, and (right) multiple vehicles.

The pixels of the images of this dataset have been classified into four classes: background, car, truck, and “others”. The last class includes all under-represented vehicles. The pixels that do not contain part of a vehicle (i.e. neither car, truck, or “others”) are labeled as background.

The labeling was performed in two steps. First, a pre-trained neural network, Mask-RCNN[2] is used to segment images automatically according to the type of vehicle. This method makes large errors in the shape and labels of the vehicles but greatly reduces the amount of manual labeling to be done. The second step consists of a manual classification to correct the errors induced by Mask-RCNN.

The result of this labeling is a 2D ground-truth classification map with the same resolutions as those of the spatial resolutions of the input hyperspectral image. The classes background, car, truck, and "others" represent respectively 79%, 6%, 14%, and under 1% of the pixels of the labeled images. The main reason for the difference between the representation of cars and trucks is the fact that trucks are covering larger portions of the images.

5.0 LEARNING

5.1 Detection

As described in Section 2, the networks consist of one intermediate layer and one output layer. For the detection network, the number of neurons of the intermediate layer was arbitrarily set to 15 and the output layer contains 1 neuron. The optimization algorithm used is the *stochastic gradient descent* with a *learning rate* of 0.005 and a *momentum* of 0.9. Previous experiments showed that networks converged too slowly with a *learning rate* of 0.001 and the momentum is set arbitrarily to prevent the networks from getting stuck in a local minimum. The loss function is the Mean-Squared Error (MSE).

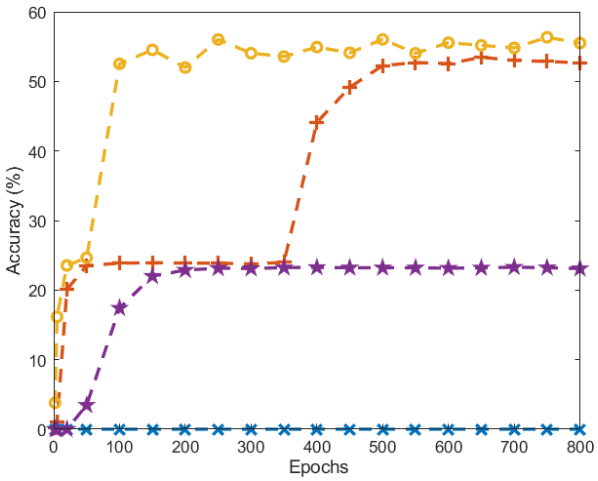

Figure 3: Trend curves of the sixteen vehicle detection models over 800 epochs. Each color corresponds to a different model.

In Fig. 3, the evolution over epochs follows four different trends. First, three models suffer from the vanishing gradient problem and classify all pixels as background. Second, three other models manage to reach around 23% accuracy on vehicles. However, after reaching this score, the models fail to increase more. Third, five models first reach 23% accuracy on vehicles quickly, then stabilize for hundreds of epochs before suddenly increasing again to reach over 50%. Finally, the five other models reach this 50% in the first 100 epochs.

In these trends, we can notice two thresholds in the vehicle class, the first being 23%, and the second 50%. These thresholds mean that there are large amounts of pixels with a similar spectral intensity being shifted

[2] "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow", https://github.com/matterport/Mask_RCNN, last visited 2022-04-25.

from one class to another. The second threshold will be analyzed in section 6.2. Due to the loss function being the MSE, the accuracy often decreases as the network tries to minimize the variations rather than only the error rate.

Although that these models could be improved further given more epochs and a better fine-tuning of the hyperparameters, these results are already sufficient to be able to analyze the features extracted by the network. Because the performances of the models have high variations, it is not possible to identify a distinct "best" network here. As such, the model we decided to analyze is the one that obtains the best accuracy (55.51%) on the vehicle class at 800 epochs.

5.2 Classification

For the classification network, the number of neurons of the intermediate layer was arbitrarily fixed to 23 and the output layer contains 3 neurons, one for each class, the "others" class being ignored due to its sub-representation. The *argmax* function is applied to the output of the network so that the output class is the class associated with the output neuron holding the highest value. The optimization algorithm chosen for this network was also the *stochastic gradient descent* with a *learning rate* of 0.005 and a *momentum* of 0.9. Previous networks converged too slowly with a *learning rate* of 0.001 and the *momentum* is set arbitrarily to prevent the networks from getting stuck in a local minimum. The three classes being unequally represented, the chosen loss function is the Cross-Entropy, which enables the association of a weight with each class. The effect on the classification is that the underrepresented classes are less neglected.

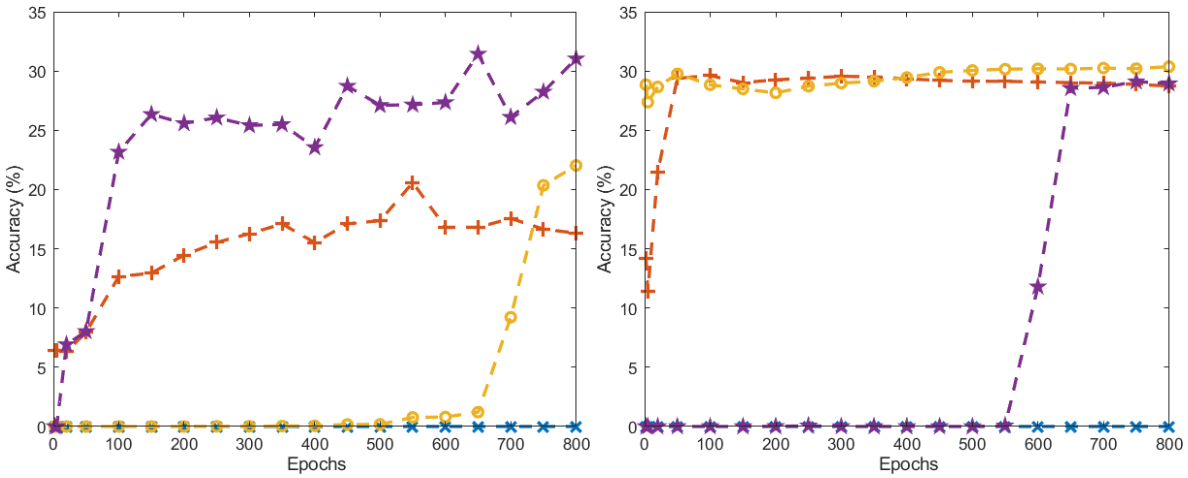

Figure 4: Trend curves of the sixteen (left) car and (right) truck classification models over 800 epochs. Each color of each subfigure corresponds to a different model.

The evolution over 800 epochs of the trends of the sixteen classification models can be seen in Fig. 4. As with vehicle detection, some models fail to classify any pixel as either a car or a truck. Another model originally suffers from the same problem but it manages to classify correctly pixels as a car after more than 550 epochs. Some models fail to recognize the truck class and yet other models fail to recognize the car class. Another model slowly learns to recognize cars but does not manage to exceed 3% accuracy in 800 epochs. Finally, six of the sixteen models manage to reach at least 20% on the car class and 30% on the truck class under 100 epochs. The model considered the "best" classification model as of 800 epochs is the one reaching 31.04% accuracy on cars and 33.51% on trucks, surpassing all the other models by over 1.5% accuracy on both classes.

6.0 DATA ANALYSIS

Because the models analyze the spectral dimension, the data of the intermediate and output layers can be visualized as 2D spatial images. We first run part of the layers on every pixel of an image. For each pixel, we normalize the values of all the neurons of a layer to the [0, 1] interval. Finally, the values are converted to the jet colormap; hence the lower values are visualized as a dark blue and the highest as a dark red, going through other colors including cyan, green and yellow.

Using this, 16 images can be obtained on the detection models and 26 on the classification models: 15 or 23 images corresponding to the activation of the neurons of the intermediate layer, 1 or 3 images to the activation of the output layer.

The intermediate layer images show the features extracted from the hyperspectral image, while the output layer images show the assertiveness of the network about its classification of each pixel. The weights between the layers have also been extracted to analyze the impact of each neuron on the classification. Due to the excessive number of weights between the input and the intermediate layers, only the weights between the intermediate and output layers are disclosed in this paper.

6.1 Detection - Best model

We first analyzed the “best” detection model selected in Sec. 5.1, the LReLU-tanh. The color reconstruction and the activation images of two test hyperspectral images can be found in Figs. 5 and 6.

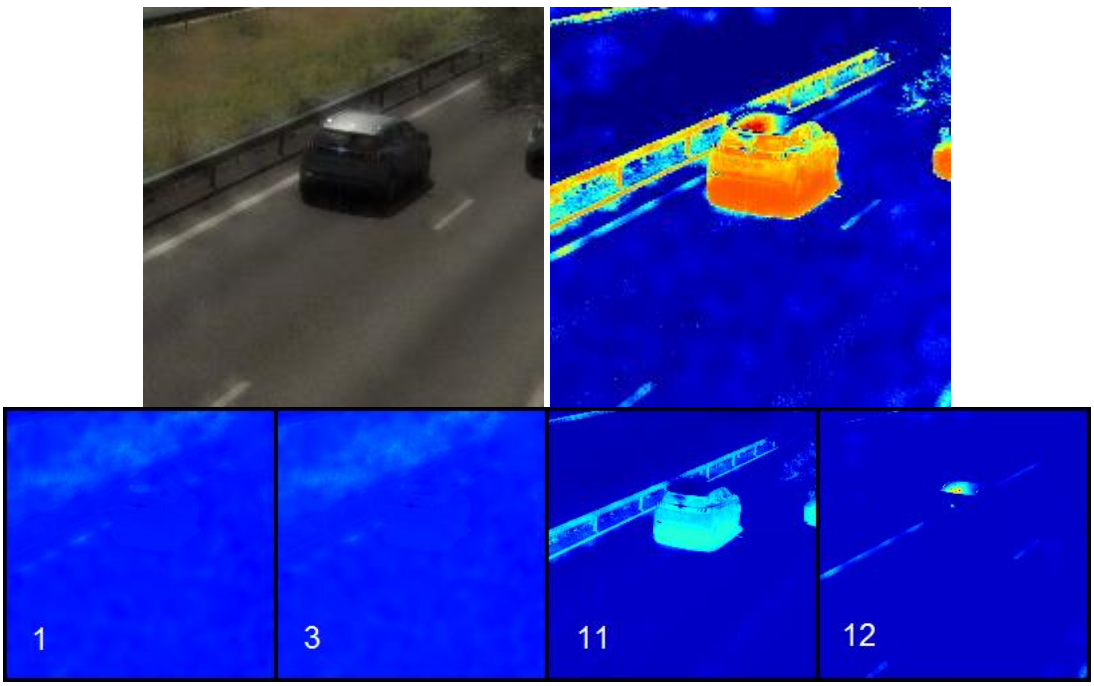

Figure 5. Best detection model car example: (up left) Color reconstruction. (up right) Activation of the output neuron. (down) Activation of four of the 15 intermediate neurons.

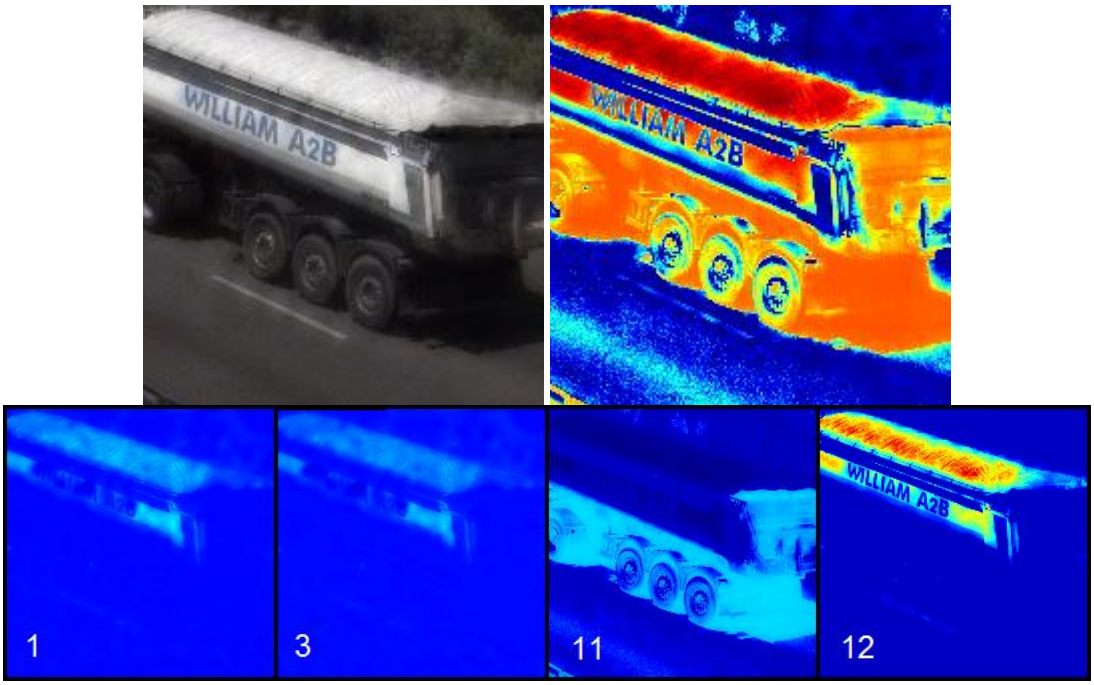

Figure 5. Best detection model truck example: (up left) Color reconstruction. (up right) Activation of the output neuron. (down) Activation of four of the 15 intermediate neurons.

Neuron	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-0.8	-0.2	-0.8	0.4	0.5	0.0	-0.1	0.3	0.1	0.0	3.4	2.1	-0.7	0.6	0.2

Table 1: Weights between the intermediate and output layers of the best detection model.

The weights associated with each intermediate neuron toward the output neuron are referenced in Tab. 1. As there is only one output neuron, the pixels with an output value closest to 0 are classified as background and closest to 1 as a vehicle. Hence, if the value of an intermediate neuron is positive, a negative weight will make it shift toward the background class and a positive weight toward the vehicle class. This relation is reversed if the value of the intermediate neuron is negative. However, the activation of the first layer is LReLU, meaning that these negative values are overall smaller than the positive ones. Here, neurons 11, 12, 1, and 3 possess the highest weights.

Neurons 1 and 3, which both have negative weights toward the output layer, distinguish part of the vegetation and mostly ignore both vehicles and roads. According to the weights between the input and the intermediate layer, these neurons mostly use the bands around 550nm (green) and 750nm (red/NIR) while rejecting the bands around 950nm (NIR). However, these neurons also respond strongly to the truck, due to its highly reflective white paint.

Neuron 11 distinguishes particularly zones of low intensity. As such, this neuron is adapted for the detection of shadows, which in these images mostly appear on vehicles. For this reason, this neuron is misguiding the classification of background shadow elements, such as roads or security fences.

Neuron 12 reacts highly to the 600-700nm (red) bands and close to 950nm (NIR) and rejects other bands. The areas appearing the most on this neuron are the ones with the highest luminosity: the white paint of the truck, the white lines of the road, and the areas reflecting the sun strongly on the car.

Six other neurons appear almost empty on test images and are poorly weighted by the output neuron. Finally, no useful information has been extracted from the other neurons, despite the medium weight associated with most of them.

6.2 Detection - Threshold

To understand the origin of this threshold in the classification of the vehicle class, we analyzed the network LReLU-sigmoid both before (300 epochs) and after (800 epochs) overcoming this threshold.

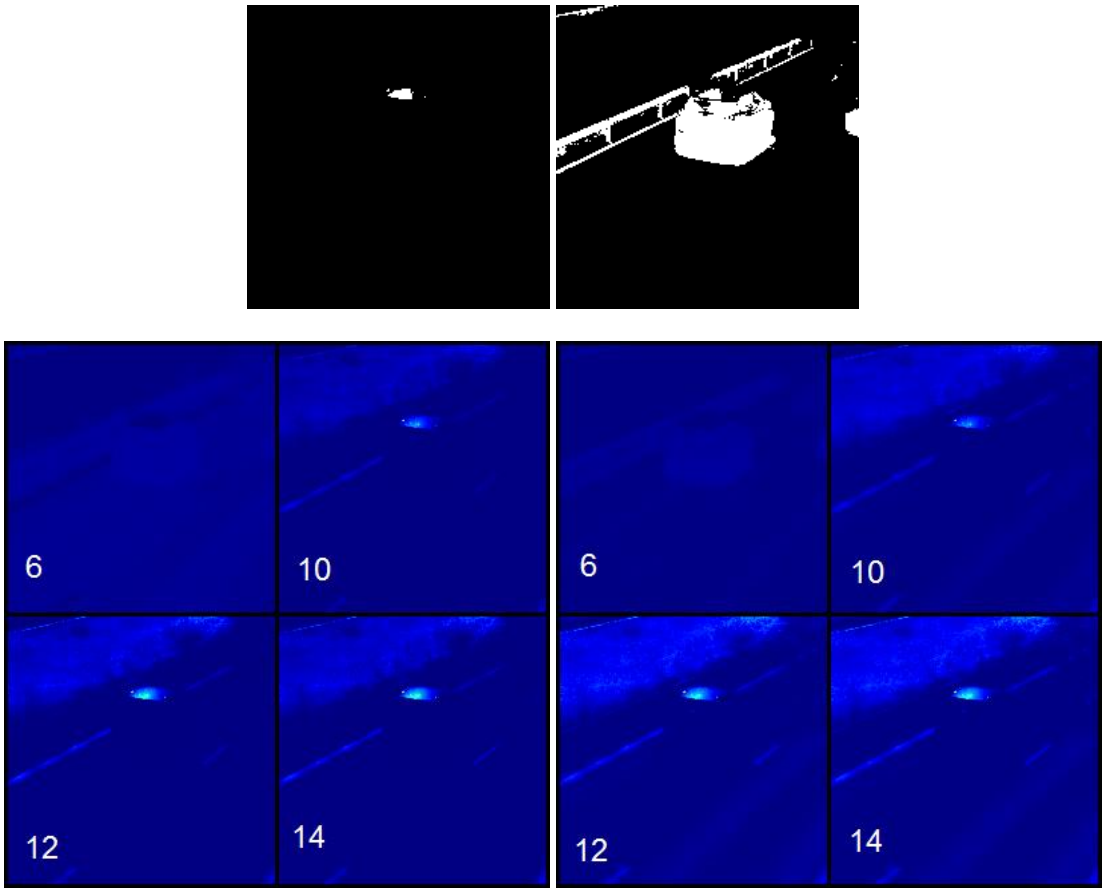

Figure 7. Threshold detection model car example: (up) Output classifications of the model after (left) 300 and (right) 800 epochs, vehicle in white and background in black. (down) Activation of four of the 15 intermediate neurons after (left) 300 and (right) 800 epochs.

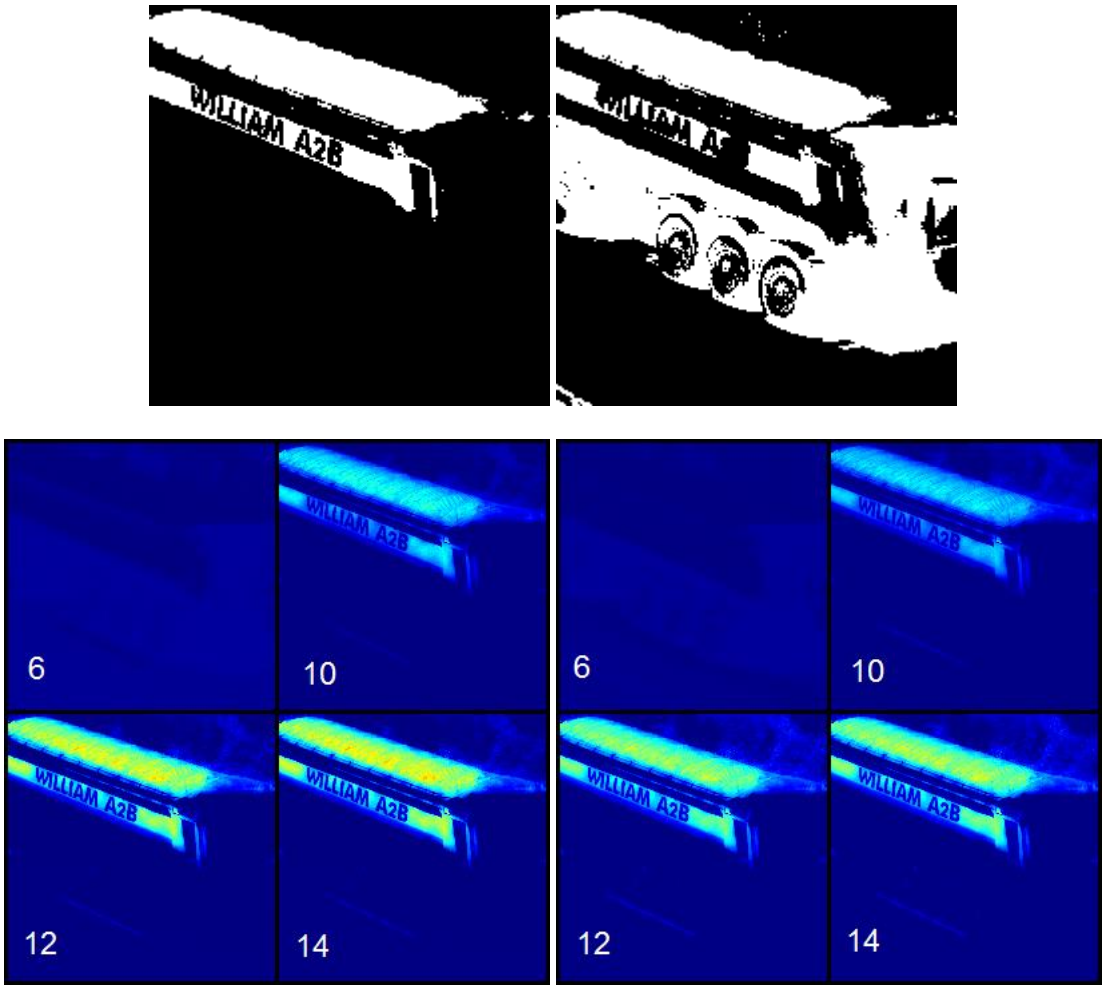

Figure 8. Threshold detection model truck example: (up) Output classifications of the model after (left) 300 and (right) 800 epochs, vehicle in white and background in black. (down) Activation of four of the 15 intermediate neurons after (left) 300 and (right) 800 epochs.

Neuron	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight (ep 300)	-0.6	-0.1	0.0	0.0	-0.2	0.5	-2.6	0.0	0.5	1.4	-0.1	2.0	0.2	2.1	0.2
Weight (ep 800)	-1.3	-0.1	0.0	0.0	-0.3	2.2	-5.2	0.0	0.8	2.4	-0.1	3.6	-0.2	3.8	0.2

Table 2: Weights between the intermediate and output layers of the detection model LReLU-sigmoid trained over 300 and 800 epochs.

The weights associated with each intermediate neuron toward the output neuron are referenced in Tab. 2.

In both Figs. 7 and 8, there are few differences between neuron activation at 300 and 800 epochs. At 800 epochs, the activation of neuron 7 is weaker on vegetation but, on the contrary, the activation of neurons 10, 12, and 14 is stronger on vegetation and weaker on the truck.

However, neuron 6 distinguishes itself from the others. This neuron doesn't differentiate roads from shadows

at 300 epochs but does at 800 epochs. The weights between the input layer and this neuron are either zero or slightly negative at 300 epochs, but greatly negative at 800 epochs. The weight between this neuron and the output layer has been multiplied by 4, while the weights of the other neurons have been multiplied by 2 at most, showing that this neuron has become much more significant in the classification. These criteria are similar to those of neuron 11 of the model analyzed in Sec. 5.1 and show that despite the models learning differently, both found that the shadows are valuable information for the classification.

We conclude that the main differences between the output classifications of Figs. 7 and 8 (up) are the shadow areas, which are detected at 800 epochs but not at 300 epochs. The threshold trend curve in Fig. 3 is due to the shift of the classification of these shadow areas from background to vehicle.

6.3 Classification - Best model

We then analyzed the best classification network, the tanh-LReLU. The images extracted from the neurons of this model with two examples are shown in Figs. 9 and 10. In the output classification, a background pixel is shown as black, car as gray, and truck as white. The weights between the intermediate and output layers are shown in Tab. 3.

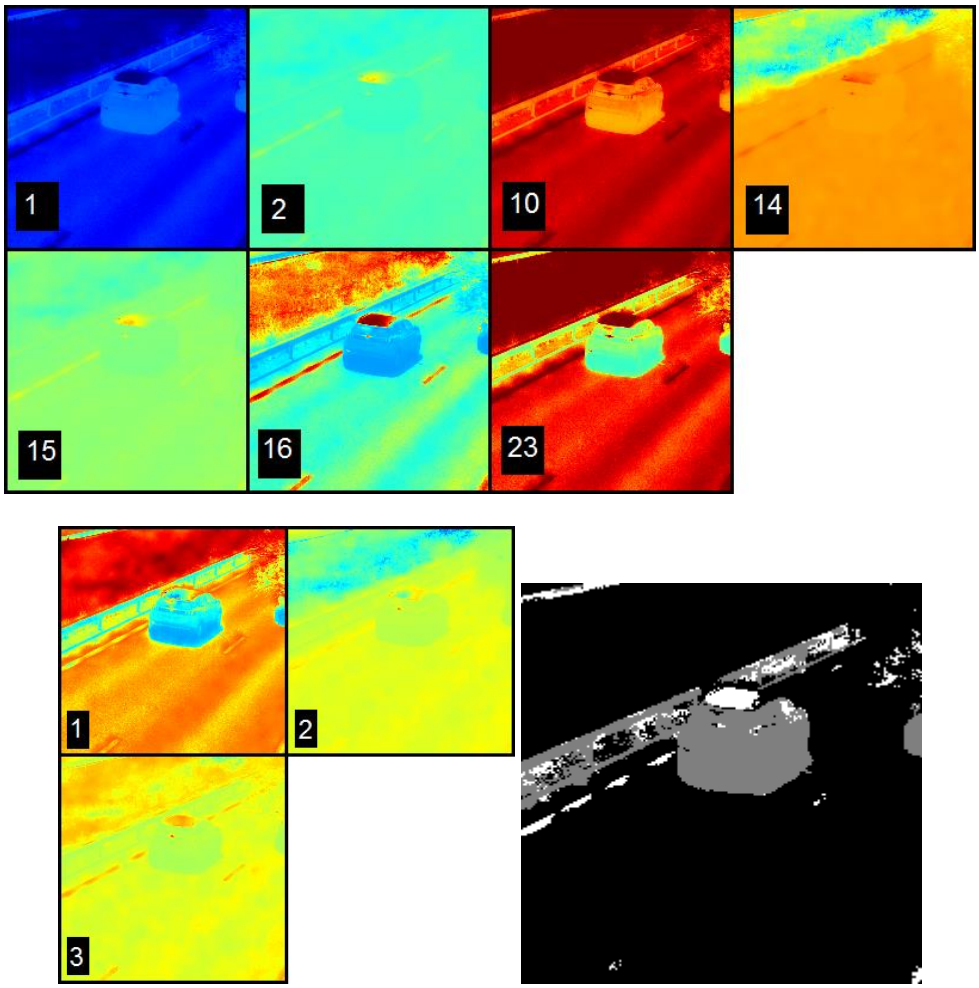

Figure 9. Best classification model car example: (up) Activation of seven of the 23 intermediate neurons. (down left) Activation of the 3 output neurons. (down right) Output classification, car in gray, truck in white, and background in black.

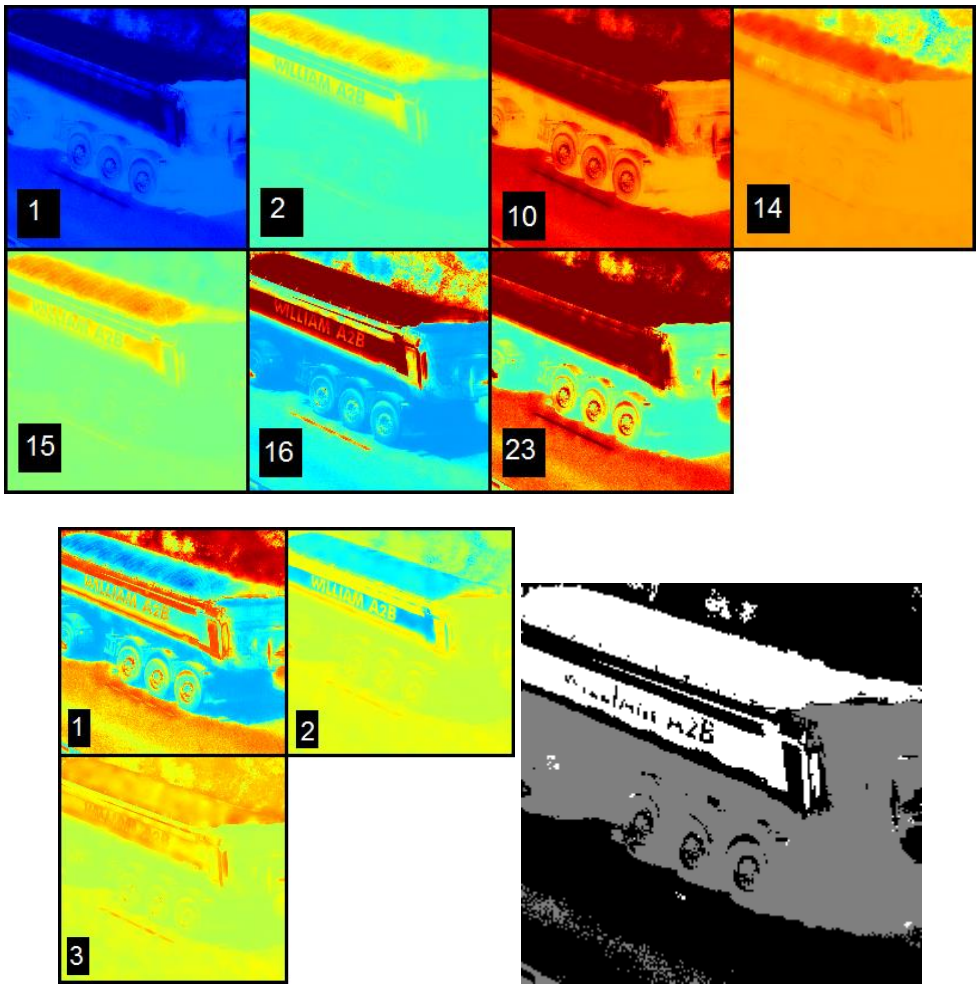

Figure 10. Best classification model truck example: (up) Activation of seven of the 23 intermediate neurons. (down left) Activation of the 3 output neurons. (down right) Output classification, car in gray, truck in white, and background in black.

Neuron	1	2	3	4	5	6	7	8	9	10	11	12
Background	-0.9	-1.6	0.9	0.4	0.9	-1.5	1.7	0.2	0.8	4.1	0.2	0.0
Car	-1.0	0.7	-0.5	-0.1	-0.3	0.7	-0.6	-0.6	0.3	0.3	-0.4	-0.1
Truck	-0.6	-0.1	0.2	0.8	0.1	-0.3	0.2	0.1	0.4	0.4	-0.1	-0.1
Neuron	13	14	15	16	17	18	19	20	21	22	23	
Background	0.6	-1.7	-1.2	-2.6	-0.9	0.9	0.5	0.4	-0.2	1.3	4.8	
Car	0.3	1.6	0.8	0.8	-2.0	-1.6	0.0	-1.6	-0.6	-1.5	-0.1	
Truck	0.8	0.5	0.0	-0.5	-1.3	-0.3	0.2	-0.4	0.2	-0.7	-0.2	

Table 3: Weights between the intermediate and output layers of the classification model tanh-LReLU trained over 800 epochs.

The first activation of this model is tanh, so the values of the intermediate layer neurons range from -1 to 1 and are centered on 0. This means that, with the jet colormap, the values obtained from the tanh that are closer to 0 are colored green, negative values shift the color toward blue and positive values toward red. Finally, the class associated with a pixel is the class corresponding to the neuron with the highest activation for this pixel, i.e. the color closest to dark red.

Therefore, neuron 1, despite its visual similarity with the shadow neurons of the previous models, associates stronger negative attention to vegetation and luminous elements. All three output neurons are weighting this neuron negatively, meaning that they are promoting high-intensity areas and vegetation. Surprisingly, the class giving the strongest weight to this neuron is the car class, though car regions contain mostly low-intensity pixels.

Neurons 2, 14, 15, and 16 are given high positive weights by the car class, are rejected by the background and ignored by the truck class. This allows the car class to value highly the luminous regions while not according too much importance to vegetation due to neuron 14. This effect is reversed for the background class which then neglects luminous regions.

The background class also associates high weights to neurons 10 and 23, detecting highly non-shadow areas (i.e. vegetation, high-, and mid-intensity areas).

The informations extracted by the neurons of the intermediate layer are low-, mid-, and high-intensity areas and vegetation. With this information, the model can obtain satisfying classification rates on both vehicle classes. On the background output neuron, most parts of the vehicles are rejected while vegetation and most road (or mid-intensity) areas have high values. However, this neuron also rejects part of the background elements such as security fences and white lines, due to their respectively very low and very high luminosity.

Due to the car and truck containing both mostly shadowy areas, these areas have similar values on car and truck classification. However, these areas are still considered to be a car by the network, and none of the models in Fig. 4 is able to classify shadow areas as a truck, because doing so would highly reduce the accuracy on the car class.

7.0 CONCLUSION

In this paper, we created our own dataset of hyperspectral images of vehicles, the first to our knowledge. The results of the networks are satisfying, considering the fact that only the spectral dimension was used. However, it would be interesting to compare such networks with similar networks applied to color images to observe the contribution of hyperspectral images compared to color images. We analyzed and found the most discriminant features extracted by the neural networks: vegetation and high-, mid-, and low-intensity areas. These features, appearing respectively green/yellow, white, gray, and black, seem to be extractible from color images. However, though the networks still make errors on pixels that appear similar to the road, the tires are mostly classified as vehicles. Furthermore, the networks are able to identify the vegetation not only through green and red but also NIR wavelengths.

In the future, further training networks on more specific areas like tires and roads or through new labeling might allow the extraction of new features. Training the networks through more epochs and modifying more hyperparameters would allow to obtain better results and maybe extract new features. Finally, using networks to analyze simultaneously spatial and spectral contribution in hyperspectral image classification would allow the extraction of new information and possibly make it comparable to state-of-the-art color image segmentation systems which benefit highly from spatial information.

Furthermore, more focus should be given to the analysis of 1D spectra and 3D hyperspectral images for material discrimination and identification. The impact of illumination conditions on spectral characteristics should also be studied. Calibration methods should be proposed to improve the performance of material discrimination and identification under unstable conditions.

REFERENCES

[1] R. B. Morrison and M. E. Cooley, "Assessment of flood damage in Arizona by means of ERTS-1 imagery," *Goddard SFC Symp. on Significant Results obtained from the ERTS-1*, vol. 1, Sect. A and B, Denver, CO, USA, pp. 755-760, 1973.

[2] C.I. Chang, Q. Du, T. Sun and M.L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remo. Sens.*, vol.37, no.6, pp.2631–2641, Nov. 1999.

[3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[4] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, Jan. 2014, DOI. 10.1117/1.JBO.19.1.010901.

[5] L. Zhu, Y. Chen, P. Ghamisi and J.A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 56 (9) (2018), pp. 5046-5063.

[6] A. Signoroni, M. Savardi, A. Baronio and S. Benini, "Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review," *Journal of Imaging*, DOI. 10.3390/jimaging5050052, Apr. 2019.

[7] J. B. Peterson, F. E. Goodrick and W. N. Melhorn, "Delineation of the boundaries of a buried preglacial valley with Landsat-l data," *Proc. of NASA Earth Resources Survey Symp.*, vol. 1-A: Agr., Environment; pp. 97-103, 1975.

[8] H. Akbari et al., "Cancer detection using infrared hyperspectral imaging," *Cancer Sci.*, 102(4), pp. 852–857, 2011.

[9] H. V. Nguyen, A. Banerjee and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," *CVPR Workshop*, 2010.

[10] M. Uzair, A. Mahmood and A. Mian, "Hyperspectral face recognition using 3D-DCT and partial least squares," *BMVC*, pp. 57.1–57.10, 2013.

[11] M. Uzair, A. Mahmood and A. Mian "Hyperspectral face recognition with spatiospectral information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127–1137, Mar. 2015.

[12] A. Mian and R. Hartley "Hyperspectral video restoration using optical flow and sparse coding," *Optics Express*, vol. 20, no. 10, pp. 10658–10673, 2012.

[13] Z. Khan, F. Shafait and A. Mian, "Hyperspectral imaging for ink mismatch detection," *Proc. of the ICDAR*, pp. 877-881, DOI. 10.1109/ICDAR.2013.179, 2013.

[14] B. Guo, S. R. Gunn, R. I. Damper and J. D.B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Letters*, vol. 3, no. 4, pp. 522-526, Oct.

2006.

[15] S. Jia, Z. Ji, Y. Qian and L. Shen, "Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal," *IEEE Journal Applied Earth Observ. and Remote Sens.*, vol. 5, no. 2, pp. 531-543, 2012.

[16] W. Zhao and S. Du, "Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544-4554, 2016.

[17] H. Wu and S. Prasad, "Convolutional recurrent neural networks for hyperspectral data classification," *Remote Sens.*, vol.9, no.3, DOI. 10.3390/rs9030298, Mar. 2017.

[18] L. Mou, P. Ghamisi and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[19] Y. Guo, S. Han, H. Cao, Y. Zhang and Q. Wang, "Guided filter based Deep Recurrent Neural Networks for Hyperspectral Image Classification," *Procedia Computer Science*, vol. 129, pp. 219-223, 2018.

[20] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bulletin of Mathematical Biophysics, vol. 5, no. 4, pp. 115-133, Dec. 1943.

[21] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Object Recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[22] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.

[23] C. Dong, C. C. Loy and K. He, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Trans. PAMI, vol. 38, no. 2, pp. 295-307, June 2015.

[24] V. Jain and S. Seung "Natural image denoising with convolutional networks," NIPS, pp. 769-776, 2009.

[25] C. Harris and M. Stephens "A combined corner and edge detector," Alvey Vision Conference, pp. 147-151, 1988.

[26] T. Young, D. Hazarika, S. Poria and E. Cambria, "Recent trends in deep learning based natural language processing," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55-75, 2018.

[27] S. E. Kahou, V. Michalski, K. Konda, R. Memisevic and C. Pal, "Recurrent Neural Networks for Emotion Recognition in Video," Proc. International Conf. on Multimodal Interaction, pp. 467-474, Nov. 2015.

[28] H. Yu, J. Wang, Z. Huang, Y. Yang and W. Xu, "Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks," IEEE CVPR, pp. 4584-4593, 2016.

[29] N. Srivastava, E. Mansimov and R. Salakhutdinov "Unsupervised Learning of Video Representations using LSTMs," ICMI, 2015.

[30] R. Socher, B. Huval, B. Bhat, C. D. Manning and A. Y. Ng, "Convolutional-Recursive Deep Learning for 3D Object Classification," NIPS, 2012.

[31] Z. Zuo, B. Shuai, G. Wang, X. Liu, X. Wang, B. Wang and Y. Chen, " Convolutional Recurrent Neural

Networks: Learning Spatial Dependencies for Image Representation," IEEE CVPR, pp. 18-26, 2015.

[32] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778-1790.

[33] G. Camps-Valls and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification," IEEE Trans. Geosci. Remo. Sens., vol.43, no.6, pp.1351-1362, 2005.

[34] G. Mountrakis, J. Im and C. Ogole "Support vector machines in remote sensing: A review," ISPRS Journal Photo. Remo. Sens., vol.66, no.3, pp.247-259, May 2011.

[35] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," Proceedings of the IEEE, vol. 101, no. 3, pp. 652-675, Mar. 2013.

[36] L. Mou, L. Bruzzone and X.X. Zhu, "Learning spectral-spatial oral features via a recurrent convolutional neural network for change detection in multispectral imagery," IEEE Trans. Geosci. Remo. Sens., vol. 57, no. 2, pp. 924-935, Feb. 2019.

[37] S. Goferman, L. Zelnik-Manor and Ayellet Tal, "Context-Aware Saliency detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34, no. 10, pp. 1915-1926, Oct. 2012.

[38] Q. Yan, L. Xu, J. Shi and J. Jia, "Hierarchical Saliency Detection,", IEEE CVPR, pp. 1155-1162, 2013.

[39] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR, 2015.