

# HyperConv: spatio-spectral classification of hyperspectral images with deep convolutional neural networks

Seyoon Ko<sup>a</sup> · Goo Jun<sup>b</sup> · Joong-Ho Won<sup>a,1</sup>

<sup>a</sup>Department of Statistics, Seoul National University;

<sup>b</sup>Department of Epidemiology, Human Genetics & Environmental Sciences,  
The University of Texas Health Science Center at Houston

(Received May 1, 2016; Revised June 24, 2016; Accepted June 30, 2016)

---

## Abstract

Land cover classification is an important tool for preventing natural disasters, collecting environmental information, and monitoring natural resources. Hyperspectral imaging is widely used for this task thanks to sufficient spectral information. However, the curse of dimensionality, spatiotemporal variability, and lack of labeled data make it difficult to classify the land cover correctly. We propose a novel classification framework for land cover classification of hyperspectral data based on convolutional neural networks. The proposed framework naturally incorporates full spectral features with the information from neighboring pixels and has advantages over existing methods that require additional feature extraction or pre-processing steps. Empirical evaluation results show that the proposed framework provides good generalization power with classification accuracies better than (or comparable to) the most advanced existing classifiers.

Keywords: hyperspectral image, land cover classification, convolutional neural network, deep learning

---

## 1. 서론

지구상의 넓은 지리적 영역의 시간적인 특성을 이해하기 위해 원격 감지를 통한 토지 피복의 분류가 필수적인 도구로 사용되고 있다. 초분광(hyperspectral) 영상은 한 픽셀에 수백 개의 스펙트럼 밴드의 정보를 포함하므로 다양한 토지 피복의 종류에 대해 풍부한 정보를 시의적절하게 얻을 수 있다.

토지 피복의 정확한 분류에는 몇 가지 난점이 있다. 첫째, 초고차원의 데이터에 따르는 차원의 저주가 있다. 둘째, 생태적, 지리적 변화에 따른 공간적인 변동성과 계절, 기후, 또는 인간의 개발에 의한 시간적인 변동성이 있다. 셋째, 기분류된(labeled) 데이터의 양이 현저히 제한되어 있다.

앞서 언급된 어려움을 극복하기 위해 많은 최신 기계 학습 방법들이 시도되어 왔다 (Camps-Valls 등, 2014). 이들 방법론에서는 고차원 데이터를 강건(robust)하게 다루기 위한 차원 축소(dimensionality

---

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIP) (Nos. 2013R1A1A1057949 and 2014R1A4A1007895).

<sup>1</sup>Corresponding author: Department of Statistics, Seoul National University, 1 Gwanak-ro, Seoul 08826, Korea. E-mail: [wonj@stats.snu.ac.kr](mailto:wonj@stats.snu.ac.kr)

reduction)나 특징 추출(feature selection) 방법 등을 많이 사용해 왔다. 최근에는 범용성과 좋은 일반화 성능을 갖고 있는 심층 신경망(deep neural networks)을 이용한 접근법이 주목받고 있다 (Bengio, 2009). 심층 신경망은 분류기 훈련과 동시에 차원 축소가 가능하여 별도의 차원 추출 단계를 필요로 하지 않으며 공간적인 정보를 추가하기 편리한 구조를 갖고 있다. 최근의 연구에서 Chen 등 (2014)은 주성분 분석(principal component analysis; PCA)과 오토인코더에 기반한 심층 학습(deep learning) 알고리즘을 초분광 영상을 통한 토지 피복 분류 연구에 성공적으로 적용하였다. Chen 등 (2014)은 공간적 상관관계를 설명하기 위해 인접한 픽셀들로 이루어진 조각(patch)에 주성분 분석을 활용하여 차원을 축소한 뒤, 여기에 여러 층의 오토인코더와 소프트맥스 분류기를 활용하였다.

이 논문에서 우리는 초분광 영상 데이터 분석을 위해 컨볼루션 신경망(convolutional neural network; CNN)을 활용하여 더 직접적인 심층 학습 구조를 제안한다. 우리 방법에서는 컨볼루션 네트워크를 활용해서 공간적인 정보를 직접 다루게 되어 별도의 전처리를 통한 차원 축소 단계가 필요치 않게 된다. 여기서는 공개된 데이터인 케네디 우주센터(Kennedy Space Center; KSC) 초분광 영상 데이터를 활용하여 제안한 구조를 평가하였다. 그 결과 컨볼루션 신경망이 초분광 영상 데이터에 대해서도 좋은 분류 능력을 갖고 있으며, 공간 정보를 포함하는 것이 풍부한 스펙트럼 정보에 의한 이득을 유지하면서 분류기의 성능을 향상시킴을 볼 수 있었다.

## 2. 심층 학습과 컨볼루션 신경망

### 2.1. 심층 학습(deep learning)

심층 학습은 패턴 인식과 분류 문제에 사용되는 지도 학습과 비지도 학습, 준지도 학습에 광범위하게 사용될 수 있는 알고리즘의 종류이다. 심층 학습 알고리즘은 인공 신경망 구조에 그 기원을 두고 있는데, 이는 특징에 여러 단계의 비선형 변환을 계층적으로 적용하는 것을 포함한다. 다층 퍼셉트론(perceptron) 등의 다층 인공 신경망 구조는 지난 수십 년 동안 알려져 있었다. 그러나 모형의 복잡도가 층 수에 지수적으로 증가하기 때문에 이러한 구조를 훈련하는 것은 어려운 일이었고, 전통적인 훈련 알고리즘을 이러한 구조에 적용했을 때의 성능은 충분히 좋지 않았다. 2006년에 힌튼(Hinton)은 이러한 “심층” 모형을 훈련하기 위해 간단하지만 매우 효과적인 아이디어를 제시하였다 (Hinton과 Salakhutdinov, 2006). 이 방법에서는 계층적인 네트워크 모형을 비지도 학습 방법을 활용하여 각 층마다 차례대로 초기화한 후, 지도학습 기법을 활용하여 모형을 미세 조정하였다. 상위 층의 특징은 하위 층에서 학습되고, 학습된 특징들은 상위 층으로 갈수록 점점 더 복잡하고 추상화된다.

심층 신경망의 구성 단위가 되는 단층 모형, 혹은 “얕은” 모형은 제한된 볼츠만 머신(restricted Boltzmann machine) (Hinton, 2010), 풀링 장치(pooling unit) (LeCun 등, 1989), 컨볼루션 신경망 (LeCun 등, 1998), 오토인코더와 잡음 제거 오토인코더 (Vincent 등, 2008) 등을 포함한다. 심층 신뢰 네트워크(deep belief network) (Hinton 등, 2006), 심층 볼츠만 머신(deep Boltzmann machine), 적층 오토인코더(stacked autoencoder), 적층 잡음 제거 오토인코더(stacked denoising autoencoder) (Vincent 등, 2010), 적층 컨볼루션 오토인코더(stacked convolutional autoencoder) (Masci 등, 2011)와 같은 다양한 중요한 심층 학습 구조들은 이러한 “얕은” 구조를 쌓는 방식으로 만들어진다. 심층 학습에 대한 보다 자세한 설명은 Goodfellow 등 (2016)에서 읽을 수 있다. 이 논문에서는 컨볼루션 신경망을 핵심 단위로 하여 초분광 영상 데이터의 분류를 위한 심층 구조를 구현한다.

### 2.2. 컨볼루션 인공 신경망(convolutional neural network)

컨볼루션 신경망은 LeCun 등 (1998)에 의해 제시된 피드 포워드(feed-forward) 인공 신경망의 일종

으로, 특히 영상 데이터 분석에 적합한 모형이다. 이름에서 보듯, 컨볼루션 신경망은 국소적인 수용장(receptive field)를 갖는 컨볼루션 층을 사용하여 사물을 인식할 때 필요한 병진 불변성(translational invariance)을 네트워크의 구조에 반영하게 된다. 컨볼루션 층에서는 기본적으로 2차원의 이동 평균(moving average) 필터를 사용하여 병진 불변인 특징을 여럿 추출한다. 각 필터의 출력은 비선형 활성화 함수를 거친 뒤 풀링 층을 거치게 된다. 풀링 층에서는 데이터를 축소하며 입력의 작은 변화에 내성을 갖도록 해 준다. 컨볼루션 신경망에서 풀링 방법으로는 최댓값 풀링(max-pooling) (Yang 등, 2009)이 가장 많이 사용된다. 풀링 층의 출력은 이어서 다층 퍼셉트론과 같은 완전히 연결된 층의 입력으로 사용된다.

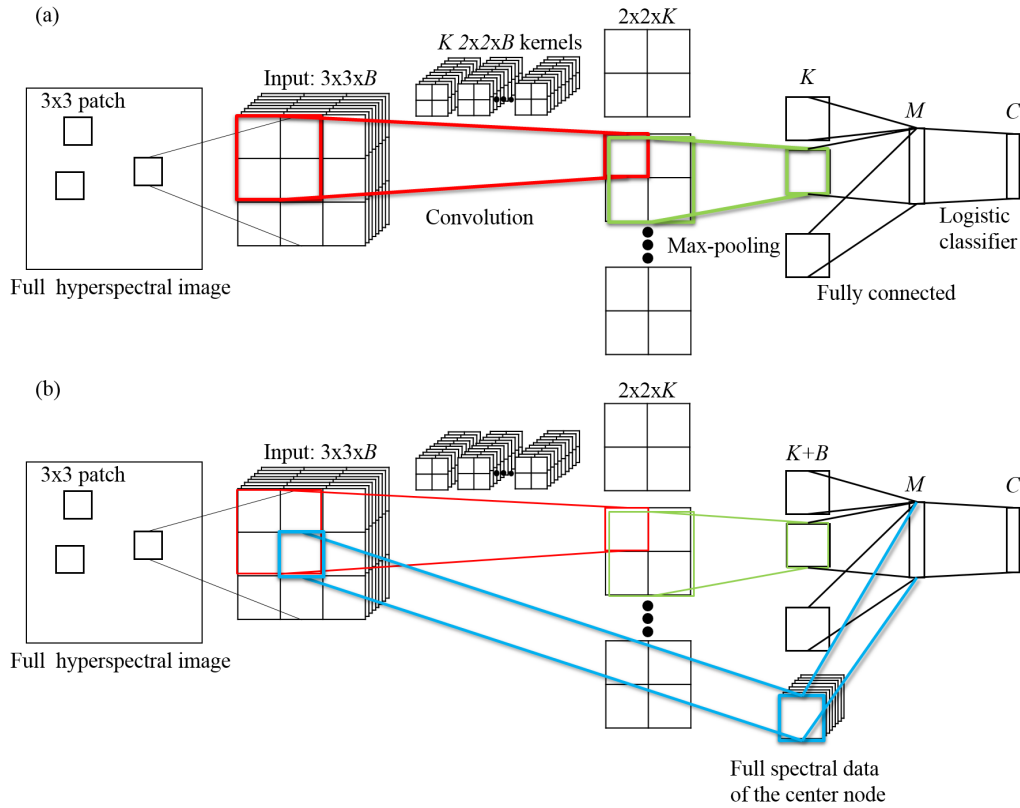
컨볼루션 신경망은 사물 인식 문제에서 가장 중요하게 사용되는 심층 신경망 모형이다. LeCun 등 (1998)의 모형은 컨볼루션 층과 최댓값 풀링 층이 번갈아가며 여러 번 쌓인 구조로 구성되어 있는, 그 자체로 “깊은” 신경망이었다. 이 모형에서는 가중치 공유(weight sharing)를 사용하여 파라미터의 수가 상대적으로 적어진다. 때문에 이름이 붙은 훈련 데이터가 많더라도 컨볼루션 신경망의 훈련은 그래픽 처리 장치(graphical processing unit; GPU)를 통해 쉽게 병렬화할 수 있다. 특히, 2012년 ImageNet 영상분류 경진대회에서 컨볼루션 신경망을 사용한 모형이 다른 경쟁작을 제치고 우승을 차지하면서 컨볼루션 신경망의 인기가 폭발적으로 늘어났다. AlexNet이라고 불리는 이 우승작 (Krizhevsky 등, 2012)은 GPU 여러 대를 사용한 병렬화를 통해 컨볼루션 신경망에 기반을 둔 방법을 보다 큰 모형으로, 더 큰 크기의 훈련 데이터에 적용할 수 있었다. 이후 컨볼루션 신경망의 다양한 변형들이 여러 분야에서 널리 사용되고 있다.

또한, GPU를 활용한 훈련 과정의 속도를 높이기 위해 정류된 선형 단위(rectified linear unit; ReLU) (Nair과 Hinton, 2010)가 도입되었으며 컨볼루션 신경망의 성능을 높이기 위한 정규화(regularization) 방법의 일환으로 드롭아웃(dropout) (Hinton 등, 2012)이 도입되었다.

### 3. 초분광 영상 분류를 위한 컨볼루션 신경망

#### 3.1. 제안된 모형의 구조

일반적인 영상 분류와 초분광 영상 분류의 핵심적인 차이는 픽셀 하나하나가 전체 스펙트럼 정보에 대응되는 풍부한 정보를 보유하고 있으나 전체 영상 중 구분류된 픽셀의 수가 매우 적다는 점이다. 따라서, 초분광 영상을 위한 컨볼루션 신경망 구조는 이러한 특성을 반영해야 한다. 여기서는 각 픽셀을 중심으로 하여  $3 \times 3$  크기의 조각을 입력으로 받기로 한다. 따라서 구분류된 픽셀을 중심으로 한 같은 크기의 조각들이 훈련 데이터를 이룬다. 이 크기를 선택한 이유는 분류기가 공간적인 상관관계를 활용하면서 입력 차원을 제한하여 전체 데이터가 GPU의 메모리에 들어가게 하기 위해서이다.  $B$ 를 스펙트럼 밴드의 수라고 하면 입력 차원은  $3 \times 3 \times B$ 가 된다. 일반적인 초분광 데이터에서  $B$ 의 값은 100을 넘는다. 입력 층 다음에는  $K$ 개의 필터를 갖는 컨볼루션 층이 오게 된다. 각각의 필터는  $2 \times 2$  크기의 수용장을 갖게 하였다. 이렇게 하면 각 필터는  $2 \times 2 \times B$ 개의 가중치를 갖게 되며, 각 노드에 대응되는 바이어스 값을 갖게 된다. 여기에 대응되는 출력의 차원은  $2 \times 2 \times K$ 이다. 그 다음에는 비선형 변환으로 이루어진 활성화 함수를 거치게 되며, 이들의 최댓값을 선택하는 최댓값 풀링 층을 거치게 된다. 최댓값 풀링 층은  $K$ 개의 단위를 갖고 있으며, 각각  $2 \times 2$  크기의 입력을 갖게 된다. 여기에 대응되는  $K$ 개의 출력은  $M$ 개의 출력을 갖는 인공 신경망을 거치며, 이후  $C$ 개의 출력을 갖는 소프트맥스(다항 로지스틱) 다중 분류기를 거치게 된다. 여기서  $C$ 는 분류하고자 하는 종류의 수를 의미한다. 제시된 심층 구조는 Figure 3.1(a)에 그림으로 표현되어 있다. 각 층마다 필요한 파라미터(가중치와 바이어스)의 수는 Table 3.1에 주어져 있다. 앞으로 이 모형을 HYPERCONV라고 부르기로 하자.



**Figure 3.1.** Structure of the convolutional neural network for hyperspectral image classification ((a) A simple model comprised of a convolutional layer, a max-pooling layer, a fully-connected layer and a logistic regression layer (HYPERCONV). (b) Full spectral information of the center pixel is also directly fed into the fully connected layer (HYPERCONV.CF)).

**Table 3.1.** Number of parameters for training

	HYPERCONV	HYPERCONV.CF
Convolution	$(4B + 1)K$	$(4B + 1)K$
Max-pooling	None	None
Fully connected	$(K + 1)M$	$(K + B + 1)M$
Logistic classifier	$(M + 1)C$	$(M + 1)C$

The number includes parameters for weight terms and bias terms.

앞서 설명한 구조에서 픽셀 사이의 공간적인 상관 관계는 병진 불변성을 가진 컨볼루션 층에 의해 효과적으로 다루어지지만, 초분광 영상은 다른 영상에 비해 훨씬 많은 스펙트럼 영역 정보를 가지고 있다. 이러한 스펙트럼 정보를 보다 명시적으로 다루기 위해, Chen 등 (2014)의 방법과 같이 최댓값 풀링 층의 출력과 함께 중심 픽셀의 전체 스펙트럼 정보를 완전히 연결된 은닉층의 입력으로 넣을 수 있다. 이렇게 하면 공간 정보를 고려하여 추출한 특징 뿐만 아니라 원하는 픽셀의 스펙트럼 정보가 분류 작업에 기여하게 된다. 이러한 접근법은 Figure 3.1(b)에 묘사되어 있다. 앞으로 이 모형을 HYPERCONV.CF라고 하자.

### 3.2. 모형의 훈련

이 절에서는 제안된 모형을 훈련시키는 방법에 대해 설명한다. 여기서 가중치와 바이어스 파라미터는 각각 회귀 모형에서의 회귀 계수와 절편 항과 같은 의미로 사용되었으며, 컨볼루션, 풀링, 피드포워드, 역전파와 같은 심층 학습 용어에 대한 보다 자세한 설명은 Goodfellow 등 (2016)에서 확인할 수 있다. 3.1절에서 설명하였듯이, HYPERCONV 모형은 다섯 층으로 이루어져 있다. 각각의 층을 차례대로  $l = 1, 2, 3, 4, 5$ 로 나타내기로 하자. 각각은 입력 층, 컨볼루션 층, 최댓값 풀링 층, 완전 연결 층, 그리고 소프트맥스 분류기 층에 대응된다. 더불어 각 층  $l = 2, 3, 4, 5$ 에 대해 대응되는 가중치 파라미터로 구성된 행렬 또는 다차원 배열을  $W^{(l)}$ 로, 바이어스 파라미터를  $b^{(l)}$ 로 나타내자.

$W^{(2)}, \dots, W^{(5)}$ 는 무작위로 초기화하고, 바이어스 파라미터  $b^{(2)}, \dots, b^{(5)}$ 는 0으로 초기화하였다. 초기화가 된 상태에서 HYPERCONV를 훈련시키는 과정은 피드포워드 단계와 역전파 단계의 두 단계로 이루어진다. 먼저 피드포워드 단계에서는 이전 층의 출력을 다음 단계의 입력으로 받아들이는 일을 반복한다. 모형을 시험할 때에도 같은 방법을 사용한다. 역전파 단계에서는 마지막 층에서부터 첫 층까지 확률적 경사 강하법(stochastic gradient descent)을 통해 모형의 비용 함수(cost function)를 최소화한다.

피드포워드 단계에서는 입력 층에서부터 출력 층까지 각 층의 중간 출력을 계산하고, 이를 다음 층의 입력으로 사용한다.  $x^{(l)}$ 을  $l$ 번째 층, 즉  $l - 1$ 번째 층의 출력( $l \geq 2$ )으로 놓자. 컨볼루션 층( $l = 2$ )은  $k$ 번째 필터  $W^{(2,k)}$ 가  $W^{(2,k)} = (W_{i,j,b}^{(2,k)})_{i=1,2,j=1,2;b=1,\dots,B}$ 로 주어질 때  $W^{(2)} = [W^{(2,1)}, \dots, W^{(2,K)}]$ 로 이루어진  $2 \times 2 \times B \times K$ 차원의 가중치를 가지고  $b^{(1)} = (b_1^{(1)}, \dots, b_K^{(1)})$ 로 주어지는  $K$ 개의 바이어스를 갖는다. 이 층의 입력은  $x^{(1)} = (x_{i,j,b}^{(1)})_{i=1,2,j=1,2;b=1,\dots,B}$ 로 주어지는  $3 \times 3 \times B$ 의 차원의 값을 갖는다. 따라서 이 층의 출력은  $x^{(2)} = (x_{ijk}^{(2)})_{i=1,2;j=1,2;k=1,\dots,K}$ 로 나타낼 수 있으며, 그 값은

$$x_{ijk}^{(2)} = \tanh \left( \sum_{i',j'=1}^2 \sum_{b=1}^B W_{i',j',b}^{(2,k)} x_{i+i'-1,j+j'-1,b}^{(1)} + b_k^{(2)} \right) \quad (3.1)$$

로 주어진다.

최댓값 풀링 층에서 출력  $x^{(3)}$ 는  $K$ 개의 성분을 가지며, 각각은

$$x_k^{(3)} = \max_{1 \leq i,j \leq 2} x_{ijk}^{(2)} \quad (3.2)$$

로 주어진다.

다음 층( $l = 4$ )은 완전 연결층으로,  $K \times M$  차원의 가중치  $W^{(4)} = (W_{mk}^{(4)})_{m=1,\dots,M;k=1,\dots,K}$ 와  $M$ 차원의 바이어스  $b^{(4)} = (b_1^{(4)}, \dots, b_M^{(4)})$ 를 갖는다. 이에 대응되는 출력  $x^{(4)}$ 의  $m$ 번째 성분은

$$x_m^{(4)} = \tanh \left( \sum_{k=1}^K W_{mk}^{(4)} x_k^{(3)} + b_m^{(4)} \right), \quad m = 1, \dots, M \quad (3.3)$$

으로 주어진다.

만약 중심 픽셀을 함께 입력으로 사용한다면 이 층은  $K$ 차원 대신  $K + B$ 차원의 입력을 가지며, 각 성분은

$$x_m^{(4)} = \tanh \left( \sum_{k=1}^K W_{mk}^{(4)} x_k^{(3)} + \sum_{k=K+1}^{K+B} W_{mk}^{(4)} x_{22}^{(1)} + b_m^{(4)} \right)$$

으로 나타난다.

소프트맥스 분류기 층( $l=5$ )은 역시 완전 연결층의 일종이며, 가중치  $W^{(5)} = (W_{cm}^{(5)})_{c=1, \dots, C; m=1, \dots, M}$ 와 바이어스  $b^{(5)} = (b_1^{(5)}, \dots, b_C^{(5)})$ 를 갖는다. 따라서 출력은

$$x^{(5)} = \sigma \left( W^{(5)} x^{(4)} + b^{(5)} \right) \quad (3.4)$$

와 같이 주어진다. 단,  $(\sigma(z))_c = e^{z_c} / (\sum_{c'=1}^C e^{z_{c'}})$ 이다.  $x_c^{(5)}$ 은 입력이 토지 피복 종류  $c$ 에 대응될 확률로 해석할 수 있다.

마지막으로, 입력의 분류는

$$\hat{y} = \operatorname{argmax}_{1 \leq c \leq C} x_c^{(5)}$$

를 계산함으로써 주어지게 된다.

앞서 언급된 다중 분류 모형을 위해 사용되는 위험 함수는 다음과 같이 주어지는 크로스 엔트로피 손실 함수(다항분포 가능도 함수)의 기대값이다.

$$L(x^{(1)}, y) = - \sum_{c=1}^C 1\{y=c\} \log(x_c^{(5)}). \quad (3.5)$$

따라서, 크기  $N$ 의 표본  $\{(x_i^{(1)}, y_i)\}_{i=1}^N$  ( $y_i$ 는 훈련 라벨)이 주어졌을 때 모형의 훈련을 위해 최소화할 비용 함수는 (3.5)에 대한 실험적 위험 함수로 주어진다.

$$J(x_1^{(1)}, \dots, x_N^{(1)}; y_1, \dots, y_N) = - \left( \frac{1}{N} \right) \sum_{i=1}^N \sum_{c=1}^C 1\{y_i=c\} \log(x_{i,c}^{(5)}). \quad (3.6)$$

파라미터들은 확률적 경사 강하법(stochastic gradient descent)을 이용하여 적합하게 된다. 이 방법에서는 훈련 표본을 무작위로 섞어 같은 크기  $L$ 의 배치(batch)으로 나눈다. 각 집단  $t$ 마다 비용 함수 (3.6)의 경사(gradient)는 아래에 설명된 방법으로 계산되며, 파라미터는 학습 속도(learning rate)  $\eta$ 로 갱신된다. 즉,  $\theta$ 를 파라미터  $\{(W^{(l)}, b^{(l)})\}_{l=2}^5$ 를 모은 것이라고 하면, 다음 식을 따라 파라미터를 갱신하게 된다.

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_{\theta} J \left( x_{j_1}^{(1)}, \dots, x_{j_L}^{(1)}; y_{j_1}^{(t)}, \dots, y_{j_L}^{(t)} \right), \quad (3.7)$$

여기서  $t = 1, \dots, T$ 이고,  $T$ 는 표본 배치의 수이다.  $T$ 개의 표본 집단을 한 번씩 거치는 과정을 하나의 에폭(epoch)이라고 하며, 이 과정은 미리 정해진 수의 에폭만큼 반복된다. 이 배치 확률적 경사 강하 알고리즘은 데이터 수준에서 병렬화하기 쉬운 GPU를 활용해 효과적으로 계산할 수 있다.

역전파 알고리즘은 마지막 층부터 첫 층까지 층마다 연쇄 법칙(chain rule)을 이용하여 경사를 계산한다. 예를 들어, 마지막 소프트맥스 분류 층( $l=5$ )에서는  $z_c^{(5)} = \sum_{m=1}^M W_{cm}^{(5)} x_m^{(4)} + b_m^{(5)}$  여기서  $m = 1, \dots, M$ 을 정의하여  $c = 1, \dots, C$ 에 대해

$$\delta_c^{(5)} = \frac{\partial J}{\partial z_c^{(5)}} = \left( \frac{1\{y_i=c\}}{x_{c,i}^{(5)}} \sigma' \left( z_c^{(5)} \right) \right)$$

을 계산한다. 이와 같이, 완전 연결 은닉층에서는  $z_m^{(4)} = \sum_{k=1}^K W_{mk}^{(4)} x_m^{(3)} + b_k^{(4)}$ 을 정의하여

$$\delta_m^{(4)} = \frac{\partial J}{\partial z_m^{(4)}} = \left( \sum_{c=1}^C W_{cm}^{(5)} \delta_c^{(5)} \right) \tanh' \left( z_m^{(4)} \right)$$

을 계산한다.

그러면 완전 연결층을 위한 갱신 식은  $l = 4, 5$ 에 대해 다음과 같이 주어진다.

$$\frac{\partial J}{\partial W_{ij}^{(l)}} = \delta_i^{(l)} x_j^{(l-1)}, \quad (3.8)$$

$$\frac{\partial J}{\partial b_i^{(l)}} = \delta_i^{(l)}. \quad (3.9)$$

최댓값 풀링과 컨볼루션 층에서는 최댓값의 선택과 컨볼루션 구조를 각각 고려하여 위와 비슷한 갱신 식을 얻을 수 있으며, THEANO (Bergstra 등, 2010)와 같은 심층 학습 패키지에서는 이것을 자동으로 계산해 준다.

전체 훈련 알고리즘은 알고리즘 1으로 요약할 수 있다.

---

**Algorithm 1** Training HYPERCONV
 

---

**Require:** input data  $x_1$ , training label  $Y$ , number of filters for CNN  $K$ , number of hidden units in the fully connected neural network layer  $M$ , learning rate  $\eta$ , batch size  $m$ , number of batches  $T$ , training epochs  $te$ , total number of classes  $C$

- 1: initialize weight matrices  $W_i$  randomly, biases  $b_i$  as zeros
  - 2: **for** every training epoch **do**
  - 3:   **for**  $t$  in  $1 \dots T$  **do**
  - 4:     Feedforward step: compute  $x_{j_1}^{(5)}, \dots, x_{j_L}^{(5)}$  through (3.1)–(3.4).
  - 5:     Compute the cost  $J$  for the  $t^{th}$  batch as (3.6)
  - 6:     Backpropagation step: update parameters from top layer to the bottom with learning rate  $\eta$  following (3.7)–(3.9)
  - 7:   **end for**
  - 8: **end for**
  - 9: **return**  $\theta$
- 

## 4. 실험

### 4.1. 데이터 설명

제안된 방법을 시험하기 위해서는 케네디 우주센터를 찍은 초분광 데이터를 사용하였다. 이 데이터는 미국 국립 항공 우주국(National Aeronautics and Space Administration; NASA)의 공중 수송 가시광선/적외선 영상 분광계(Airborne Visible/Infrared Imaging Spectrometer; AVIRIS)를 이용하여 1996년 3월 23일에 촬영한 데이터이다. AVIRIS 기기는 400nm에서 2500nm까지의 파장 영역에서 약 10nm 간격으로 224개 밴드의 반사율 정보를 수집한다. 이 데이터는  $512 \times 614$ 픽셀 크기이며, 공간 해상도는 18미터이다. 물에 흡수되는 파장 영역과 신호 대 잡음비(signal-to-noise ratio; SNR)가 작은 파장 영역을 제외한 176개 밴드의 정보가 공개되어 있으며 이를 분류에 사용하였다. 전체 314,368 (=  $512 \times 614$ )개의 픽셀 중 5,211개만이 분류되어 있으며, 분류 라벨은 Table 4.1과 같이 총 13종의 토지 피복으로 이루어져 있다. 각 밴드의 반사율 수준은 영상 전체에 걸쳐 표준화되었다. Chen 등 (2014)과 비슷하게, 기본류된 픽셀들을 (훈련 집합):(검증 집합):(테스트 집합) = 6:2:2로 나누었다. 즉, 라벨이 있는 픽셀들 전체의 60%를 훈련 집합, 20%를 검증 집합, 그리고 나머지를 테스트 집합으로 잡았다.

**Table 4.1.** Land cover classes and numbers of pixels in KSC dataset

Class no.	Class name	Training set	Validation set	Test set
1	Scrub	467	153	132
2	Willow swamp	140	50	49
3	Cabbage palm hammock	147	60	47
4	Cabbage palm / Oak hammock	149	49	49
5	Slash pine	87	33	41
6	Oak / Broadleaf hammock	129	52	46
7	Hardwood swamp	61	26	17
8	Graminoid marsh	256	88	84
9	Spartina marsh	318	87	110
10	Cattail marsh	275	70	57
11	Salt marsh	250	82	82
12	Mud flats	288	97	113
13	Water	553	173	193
Total		3120	1020	1020

#### 4.2. 실험 환경

모든 모형은 파이썬(Python)에 기반한 심층 학습 라이브러리인 THEANO (Bergstra 등, 2010) 버전 0.6에 기반하여 구현되었다. THEANO는 심층 학습 응용에서 자주 나타나는 다차원 배열을 쉽게 정의하고 최적화하는 기능을 제공한다. 구현된 코드는 <https://github.com/kose-y/HyperConv>에 공개되어 있다. 모형은 Nvidia GTX 780 GPU가 장착된 Intel Core i7-4790K @ 4.00GHz CPU에서 훈련하였다. GPU는 3GB의 메모리를 가지고 있으며 케플러 구조로 설계되었다. 훈련 속도  $\eta$ 는 0.01로 고정되었고, 확률적 경사 강하법의 훈련 에폭 수는 20,000회로 하였다. 배치 크기  $L$ 은 60이었다.

#### 4.3. 분류 성능

컨볼루션 커널 수  $K$ 와 완전 연결 은닉층의 노드 수  $M$ 을 여러 값으로 변경시키며 성능 실험을 수행하였다. 각 모형의 분류 정확도와 실행 시간은 Table 4.2에 주어져 있다. HYPERCONV의 정확도는 0.9765–0.9873 범위에 들어갔으며, HYPERCONV.CF의 정확도는 0.9843–0.9873 범위에 들어가서 HYPERCONV보다 약간 높았다. 양 모형에서 최고의 정확도는 모두 0.9873이었지만 HYPERCONV.CF가 해당 정확도를 얻기 위해 더 적은 수의 커널과 훈련 시간을 사용하였다. 더불어, HYPERCONV.CF가 여러  $(K, M)$  설정에 대해 일관되게 이 정확도를 얻었다. 3,000번째 에폭까지의 학습 곡선은 Figure 4.1에 주어져 있다. 이를 통해 HYPERCONV.CF의 정확도가 HYPERCONV보다 먼저 안정되는 것을 볼 수 있었다. 중심 픽셀의 전체 스펙트럼 정보를 은닉층에 입력으로 넣는 것이 분류기가 더 수월하게 안정되게 하는 것으로 보인다.

몇몇 성능이 가장 좋은 모형의 컨볼루션 층의 학습된 필터의 가중치를 Figure 4.2에 나타내었다. 그림은 네 개의 영역으로 나누어져 있는데, 각각은  $2 \times 2$  크기의 필터의 각 위치를 나타낸다. 각 영역은  $176 \times K$ 개의 픽셀로 이루어져 있으며, 각 행은 하나의 밴드, 각 열은 하나의 필터 커널에 대응된다. 선명한 세로 무늬는 커널 가중치 사이에 높은 분광학적 상관 관계가 있음을 의미한다. 또한, 간헐적인 가로 줄무늬는 일부 파장 영역이 분류에 필요한 정보를 더 풍부하게 갖고 있음을 의미한다. 따라서 제안된 모형이 매우 복잡한 초분광 영상에서 구별을 잘 나타내는 특징을 추출할 수 있음을 알 수 있다.

훈련 시간은 커널의 수  $K$ 에 크게 의존했으며 은닉층의 노드 수  $M$ 에는 영향을 적게 받았다. 예를 들



**Table 4.2.** Accuracies of HYPERCONV and HYPERCONV.CF models with various  $K$  and  $M$  values

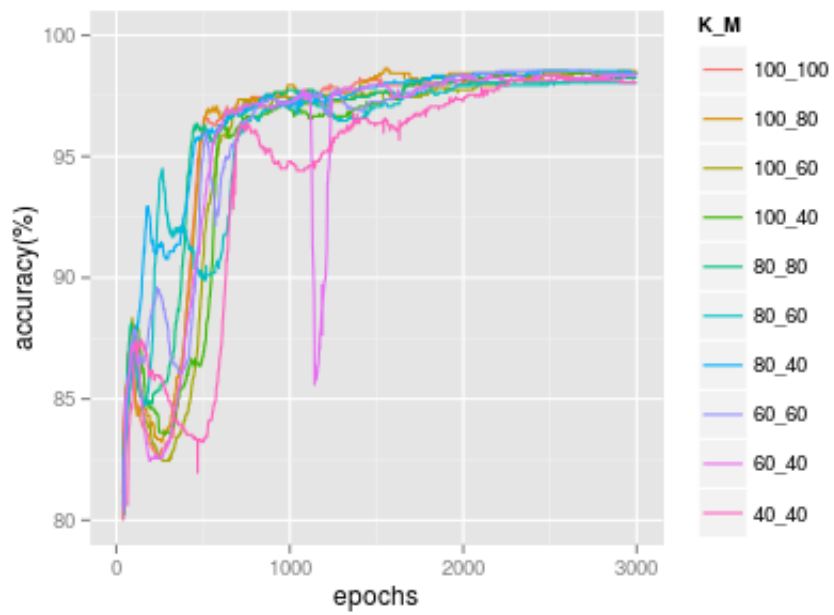
	$K$	$M$	Test accuracy	Training time (min)
HYPERCONV	100	100	0.9843	254.23
	100	80	0.9833	254.11
	100	60	0.9824	254.15
	100	40	<b>0.9873</b>	255.56
	80	80	0.9765	203.31
	80	60	0.9843	202.35
	80	40	0.9843	202.95
	60	60	0.9843	157.83
	60	40	0.9853	157.48
	40	40	0.9863	111.89
HYPERCONV.CF	100	100	0.9853	264.65
	100	80	0.9863	268.56
	100	60	0.9843	265.93
	100	40	0.9853	270.51
	80	80	<b>0.9873</b>	219.54
	80	60	0.9863	215.58
	80	40	<b>0.9873</b>	213.57
	60	60	<b>0.9873</b>	162.24
	60	40	0.9843	161.89
	40	40	0.9863	113.66

어, HYPERCONV.CF 모형을 고정된  $K = 100$ 으로 훈련하는 것은  $M = 40, 60, 80, 100$  모두 260분보다 오래 걸렸으며, 분류 성능에는 큰 차이가 없었다. 훈련 시간은  $K$ 에 대해 거의 선형적으로 증가하였다.  $K$ 가 주어져 있을 때, 여러 값의  $M$ 에 대한 훈련 시간의 차이는 3% 이내였다. HYPERCONV.CF의 훈련 시간은 HYPERCONV보다 평균적으로 4.7% 더 오래 걸렸다. Table 4.2에 나타난 훈련 시간은 전체 훈련 과정에 걸린 시간을 나타내며, 피드포워드 단계, 파라미터 갱신 단계, 검증과 테스트 단계 모두를 포함한다.

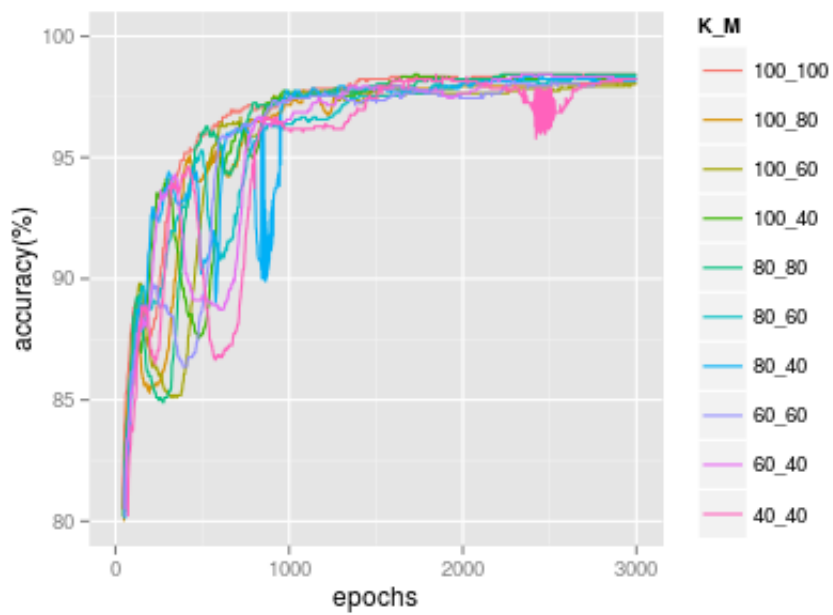
#### 4.4. 다른 방법과의 비교

서포트 벡터 머신(support vector machine)은 초분광 영상을 통한 토지 피복 분류 문제에서 매우 뛰어난 성능을 보이는 것으로 알려져 있다 (Melgani와 Lorenzo, 2004; Gualtieri와 Chettri, 2000). 앞서 언급한 바와 같이 Chen 등 (2014)은 최근에 이 문제에 심층 학습에 기반한 구조인 적층 오토인코더를 사용하는 방식을 제안하였다. 공간 정보를 활용하기 위해 이 방법에서는 PCA를 통해 스펙트럼 차원을 축소한 뒤 축소된 영상의  $3 \times 3$  조각을 적층 오토인코더의 입력으로 사용하였다(PCA-SAE). 더불어, 해당 연구에서는 중심 픽셀의 전체 스펙트럼 정보를 컨볼루션 신경망을 통해 축소된  $3 \times 3$  영상과 이어붙여 스펙트럼 정보를 강조하는 방법을 사용하였다(PCA-CENTERFEED-SAE). 적층 오토인코더를 통해 사전학습된 특징에 다층 퍼셉트론 구조를 이용하여 미세 조정(fine-tuning)하는 방식으로 0.9876의 정확도를 얻었다. 이와 유사한 실험 환경에서 우리는 심층 학습의 간단한 단위만을 사용한 심층 컨볼루션 신경망을 이용하여 해당 연구 결과와 거의 비슷한 0.9873의 정확도를 얻을 수 있었다.

제안한 방식을 훈련 시간의 측면에서 Chen 등 (2014)과 비교해 보자. Chen 등 (2014)에서는 실제 훈련 시간에 대해 직접 언급하지 않고 다층 오토인코더를 훈련하는 데 필요한 시간만 그림으로 보여 주었

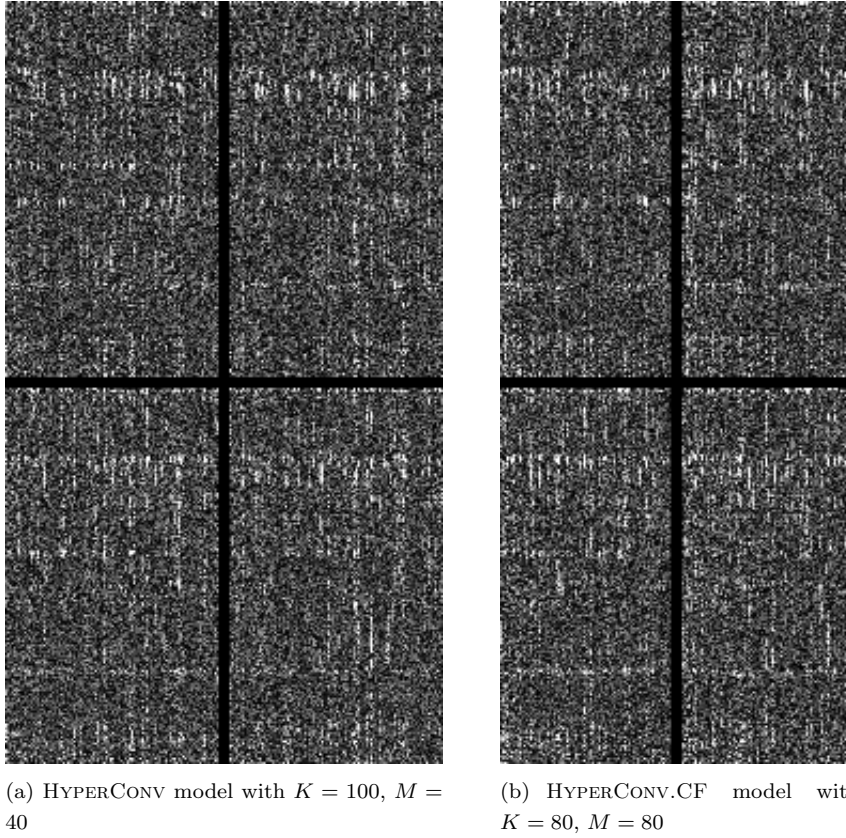


(a) HYPERCONV



(b) HYPERCONV.CF

**Figure 4.1.** Learning curve of the models for (a) HYPERCONV (b) HYPERCONV.CF up to 3000 epochs.



**Figure 4.2.** Visualized filter weights of the first convolutional layer of (a) HYPERCONV model with  $K = 100$ ,  $M = 40$ , and (b) HYPERCONV.CF model with  $K = 80$ ,  $M = 80$ . The intensity of each pixel represents the absolute value of the weight.

**Table 4.3.** Accuracies of various land cover classification methods. The methods that used full spectral data of the central pixel along with spatial information were the best. Our CNN-based method could attain accuracy similar to PCA-based approach of Chen *et al.* (2014), but the proposed structure requires less training time.

Methods	Accuracy
linear SVM (Chen <i>et al.</i> , 2014)	0.9552
logistic regression	0.9137
SAE (Chen <i>et al.</i> , 2014)	0.9673
PCA-SAE (Chen <i>et al.</i> , 2014)	0.9776
PCA-centerfeed-SAE (Chen <i>et al.</i> , 2014)	<b>0.9876</b>
HYPERCONV	<b>0.9873</b>
HYPERCONV.CF	<b>0.9873</b>

다 (Chen 등 2014, Figure 10). 이 그림에서 단층 오토인코더를 50,000 에폭 동안 훈련하는 데 필요한 시간은 10-20분 사이인 것으로 보인다. 이 논문에서 가장 좋은 결과인 “PCA-centerfeed-SAE”를 얻을 때는 앞의 8배인 400,000회의 에폭을 사용했고, 단층이 아닌 5층의 적층 오토인코더를 사용했으므로

80–160분의 수 배에 이르는 훈련 시간이 필요했을 것이다. 이는 우리 모형에서 가장 좋은 결과를 얻을 때 걸린 시간인 111–270분에 비해 긴 시간으로 생각된다.

더욱이, 우리의 접근 방법에서는 PCA(시간복잡도  $O(p^2n + p^3)$ ,  $p$ 는 주성분 수,  $n$ 은 픽셀의 수)와 같은 전처리를 통한 차원 축소를 필요로 하지 않는다. 다시 말해, 심층 컨볼루션 신경망을 사용한 우리 모형은 차원 축소와 특징 추출을 한꺼번에 할 수 있다. 다양한 분류 모형에 대한 성능 비교는 Table 4.3에 주어져 있다.

## 5. 결론

심층 인공 신경망은 초분광 영상 데이터의 분류를 위한 다목적 도구가 될 가능성이 있다. 본 논문에서 제안한 접근 방식에서는 전체 스펙트럼 특징과 공간적 상관관계 정보를 함께 사용하기 위해 컨볼루션 신경망을 사용하였다. 적층 오토인코더에 기반한 방식을 포함한 다른 최첨단 기술들과 비교할 때, 우리의 방식은 초분광 데이터 분석을 위해 더 직관적인 심층 신경망 구조를 사용한다. KSC 데이터를 이용한 실험 결과 이 방식이 기존의 고성능 알고리즘에 비해 더 높거나 비슷한 정확도를 갖는 것을 볼 수 있었다. 또한, 기본적인 컨볼루션 신경망 구조에 더해 중심 픽셀의 스펙트럼 정보를 강조하기 위해 컨볼루션 신경망의 출력과 중심 픽셀의 전체 스펙트럼 정보를 이어붙여 다항 분류기의 입력으로 사용하면 한층 더 좋은 성능의 분류기를 얻을 수 있음을 확인할 수 있었다. 이 구조를 보다 발전시키면 단순한 지도 학습 뿐만 아니라 전이 학습(transfer learning), 준지도 학습 등등의 복잡한 세팅에서의 토지 피복 분류에 중요한 역할을 할 수 있을 것이다.

## References

- Bengio, Y. (2009). Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, **2**, 1–127.
- Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., and Bengio, Y. (2010). Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*.
- Camps-Valls, G., Tuia, D., Bruzzone, L., and Atli Benediktsson, J. (2014). Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, *Signal Processing Magazine, IEEE*, **31**, 45–54.
- Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y. (2014). Deep learning-based classification of hyperspectral data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **7**, 2094–2107.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Book in preparation for MIT Press. Retrieved from <http://www.deeplearningbook.org>.
- Gualtieri, J. A. and Chettri, S. (2000). Support vector machines for classification of hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, **2**, 813–815.
- Hinton, G. (2010). A practical guide to training restricted Boltzmann machines, *Momentum*, **9**, 926.
- Hinton, G., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets, *Neural Computation*, **18**, 1527–1554.
- Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks, *Science*, **313**, 504–507.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors, *arXiv preprint arXiv:1207.0580*.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097–1105.

- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition, *Neural Computation*, **1**, 541–551.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, **86**, 2278–2324.
- Masci, J., Meier, U., Cireřan, D., and Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural Networks and Machine Learning–ICANN 2011* (pp. 52–59), Springer.
- Melgani, F. and Lorenzo, B. (2004). Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1778–1790.
- Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814.
- Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning* (pp. 1096–1103), ACM.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *The Journal of Machine Learning Research*, **11**, 3371–3408.
- Yang, J., Yu, K., Gong, Y., and Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 1794–1801).

# 심층 컨볼루션 신경망을 사용한 초분광 영상의 공간 분광학적 분류 기법

고세윤<sup>a</sup> · 전구<sup>b</sup> · 원중호<sup>a,1</sup>

<sup>a</sup>서울대학교 통계학과, <sup>b</sup>Department of Epidemiology, Human Genetics & Environmental Sciences,  
The University of Texas Health Science Center at Houston

(2016년 5월 1일 접수, 2016년 6월 24일 수정, 2016년 6월 30일 채택)

---

## 요약

초분광 영상 데이터는 픽셀마다 수백 개의 스펙트럼 밴드에 대한 정보가 주어지는 고차원 데이터로, 농업, 식품처리, 광물학, 물리학, 환경학, 지리학 등 광범위한 분야에 활용되고 있다. 그 중 하나는 토지 피복의 분류 문제인데, 이는 자연 재해 예방, 자연 자원 감시, 환경에 대한 정보 수집에 있어서 중요한 문제이다. 하지만 차원의 저주, 시공간적 변동성, 레이블된 데이터의 부족 때문에 토지 피복의 정확한 분류에는 어려움이 따른다. 이 논문에서는 이러한 문제를 해결하기 위해 컨볼루션 신경망에 기반한 새로운 심층 학습 구조를 제안한다. 제안된 구조는 원하는 지점 주변 픽셀의 정보를 컨볼루션 신경망을 통해 처리하고, 그 지점의 스펙트럼 정보를 강조하기 위해 컨볼루션 층의 출력과 스펙트럼 정보를 함께 소프트맥스 분류기의 입력으로 사용한다. 이 구조는 추가적인 특징 추출 과정을 필요로 하지 않고, 그래픽 처리 장치 등을 이용한 병렬화가 간편하다는 점에서 기존 방법들보다 유리하다. 실험 결과, 제안된 구조는 기존에 가장 좋은 성능을 보인 분류기와 비슷하거나 더 좋은 분류 정확도를 보여 좋은 일반화 성능을 보이는 것을 확인할 수 있었다.

주요용어: 초분광 영상, 토지 피복 분류, 컨볼루션 신경망, 심층 학습

---

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (Nos. 2013R1A1A1057949, 2014R1A4A1007895).

<sup>1</sup>교신저자: (08826) 서울특별시 관악구 관악로 1, 서울대학교 통계학과. E-mail: wonj@stats.snu.ac.kr