The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan

UNSUPERVISED HYPERSPECTRAL IMAGE SEGMENTATION USING DEEP GLOBAL CLUSTERING

Yu-Tang Chang, Pin-Wei Chen, Shih-Fang Chen*

Department of Biomechatronics Engineering, National Taiwan University, Taiwan.

*Corresponding Author: sfchen@ntu.edu.tw

ABSTRACT

Hyperspectral imaging (HSI) combines rich spectral and spatial information, supporting field monitoring and crop assessment in precision agriculture. HSI scenes from one dataset usually share the same background and foreground classes, yet spectra from one region differ from those in another. Pixels that describe the same object therefore cluster together in spectral space; mapping these clusters back onto the image yields pseudo-segmentations that can stand in for class labels. However, processing large-scale HSI datasets poses significant computational challenges due to millions of pixels per frame. This study presents Deep Global Clustering (DGC), an end-to-end unsupervised method for HSI segmentation without requiring pre-training. The architecture integrates convolutional neural networks with learnable cluster modules to extract dataset-level spectral-spatial features. Training uses partially overlapping 64 × 64 crops from each HSI frame and optimizes four losses, including intra-cluster coherence, centroid repulsion, crop consistency, and balanced centroid usage. DGC was evaluated on 86 leaf HSI data, comprising healthy samples and leaves affected by brown blight disease, each roughly 1000 × 1000 pixels across 301 spectral bands from 400 nm to 1000 nm. With two clusters, DGC achieved background-tissue separation with IoU of 0.972; four-clusters segmentation yielded distinct groups for background texture, healthy lamina, and the lesion region; the blight area appeared as its own cluster despite some residual noisy pixels, demonstrating unsupervised detection of the disease. The DL-based method demonstrates scalable performance by maintaining constant memory usage through efficient HSI processing. These results establish DGC as a practical solution for unsupervised HSI analysis in precision agriculture applications.

Keywords: Hyperspectral imaging, Image segmentation, Unsupervised learning, Deep learning, Clustering.

INTRODUCTION

Hyperspectral imaging (HSI) captures hundreds of narrow spectral bands per pixel, adding spectral detail to spatial information and enabling field monitoring and crop assessment in precision agriculture. HSI is an image stack with many wavelength-specific channels (e.g., 400 nm, 402 nm, or 404 nm). However, computational challenges arise from the massive data volume: a 1000×1000 scene with hundreds of bands can exceed 1GB, raising significant computation bottlenecks.

In agriculture applications, HSI pixel spectra are annotated as entities under a unified rule shared across datasets. For example, background, leaf, and lesion areas are common across leaf HSI data (Fig. 1a). Because spectra add informative detail beyond RGB, pixels from the same entity typically lie close in spectral space and form unsupervised clusters. Projecting these clusters back onto the image yields pseudo-segmentations. Centroids, however, rarely transfer across scenes because illumination varies during acquisition. Data volume is another constraint: an HSI data is large; hence joint processing of many HSI data is often impractical.

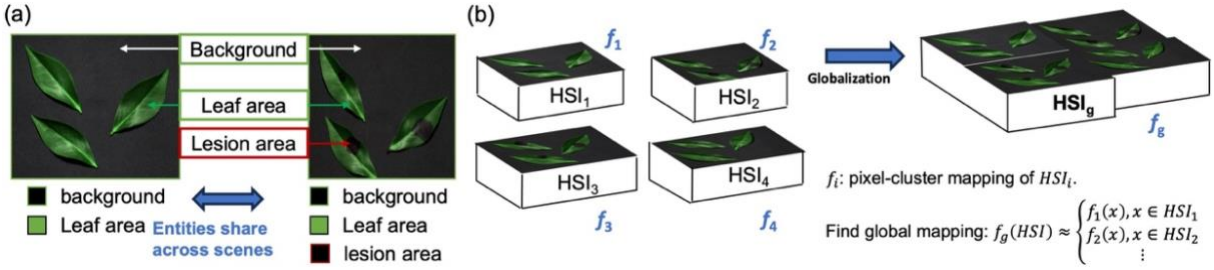

Fig. 1. Leaf dataset. (a) Entities. (b) Global versus per-scene (local) HSI pixel clustering.

Traditional clustering methods struggle with HSI's high dimensionality and computational requirements. To capture these clusters efficiently, this work introduces Deep Global Clustering (DGC), an end-to-end unsupervised method for HSI pixel clustering and pseudo-segmentation. DGC is a low-cost DL framework that approximates dataset-level global scene clustering by iteratively sampling local scenes from HSI data (Fig. 1b).

MATERIALS AND METHODS

HSI Leaf Dataset and Evaluation Metrics

HSI data were collected with a Hyperspec MV.X (Headwall Photonics Inc.; Bolton, MA, USA), each at 1000 × 1000 pixels and 301 bands from 400–1000 nm. Scenes contained a black conveyor (background), healthy leaves, and leaves with brown blight disease. The dataset included 86 samples: 64 with lesions and 22 healthy-only. Manual annotation provided ground truth for evaluation. The primary metric was Intersection over Union (IoU) on background–tissue (B–T) separation.

Deep Global Clustering (DGC)

DGC has two learnable parts: a hybrid 1D/2D CNN encoder and an unrolled mean-shift (UMS) module that stores global centroids (Comaniciu & Meer, 2002; Fig. 2a). The encoder first applies three 1D convolutions along the spectral axis (kernel size 9) to compress per-pixel

spectra to a latent feature space (dimension = 32). It then applies two 2D convolutions over the image plane (kernel size 3 × 3) to add local spatial context within a small receptive field. The UMS module maintains a bank of centroids and runs five Gaussian mean-shift iterations. To reduce memory, pixels are processed in 5 × 5 patches. After feature extraction, each pixel is assigned to the nearest centroid in the learned space, producing the HSI pseudo-segmentation.

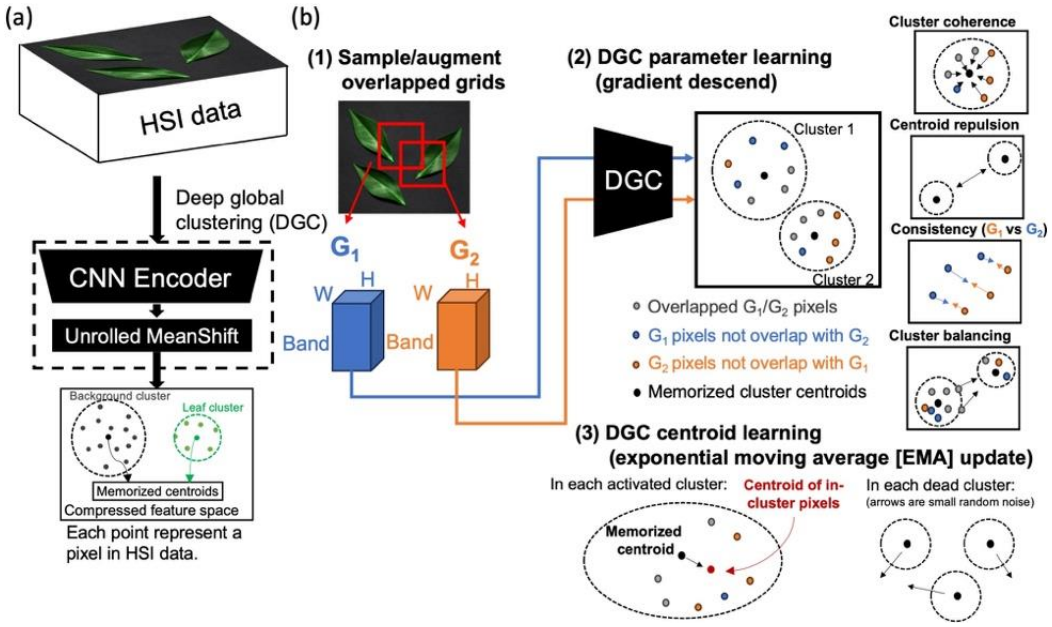

Fig. 2. Deep global clustering (DGC) overview: (a) modules; (b) training flow.

Training uses overlapping local patches to learn global clusters (Fig. 2b). Each step samples two 64 × 64 patches from the same HSI data with overlap, applies data augmentation, and feeds both through the encoder and UMS. Independent mini-batches may drift to different clusterings, so the model is constrained to produce consistent, well-separated clusters. The loss combines four terms: (i) intra-cluster coherence (compactness), (ii) centroid repulsion (orthogonality), (iii) cross-view (overlap) consistency, and (iv) balanced cluster usage. After backpropagation, the centroid update via exponential exponential moving average. Training is monitored by mutual information between predictions from consecutive epochs. AdamW (lr 1e-4) and batch size 4 are used (Loshchilov & Hutter, 2017). This requires about 10 GB VRAM in training.

RESULTS & DISCUSSION

Two configurations with two and four learnable clusters, are reported—DGC-2 and DGC-4 (Fig. 3a). DGC-2 produced clean B–T separation, although parts of lesions were labeled as backgrounds. In the pseudo-image these lesions appear nearly black, close to the background spectrum, so a spectral-driven unsupervised model can confuse them. DGC-4 separated the background into two types that reflect flat surface and texture, and split tissue into healthy and lesion-affected areas. Using manual masks to evaluate B–T separation, DGC-2 achieved IoU 0.972 (background) and 0.878 (tissue), mean 0.935; DGC-4 achieved IoU 0.944 (background) and 0.780 (tissue), mean 0.857 (Fig. 3b). The two-cluster model delivered higher B–T agreement; the four-cluster model provided finer semantics at lower scores. Disease areas formed coherent clusters without supervision, demonstrating unsupervised disease detection

capability.

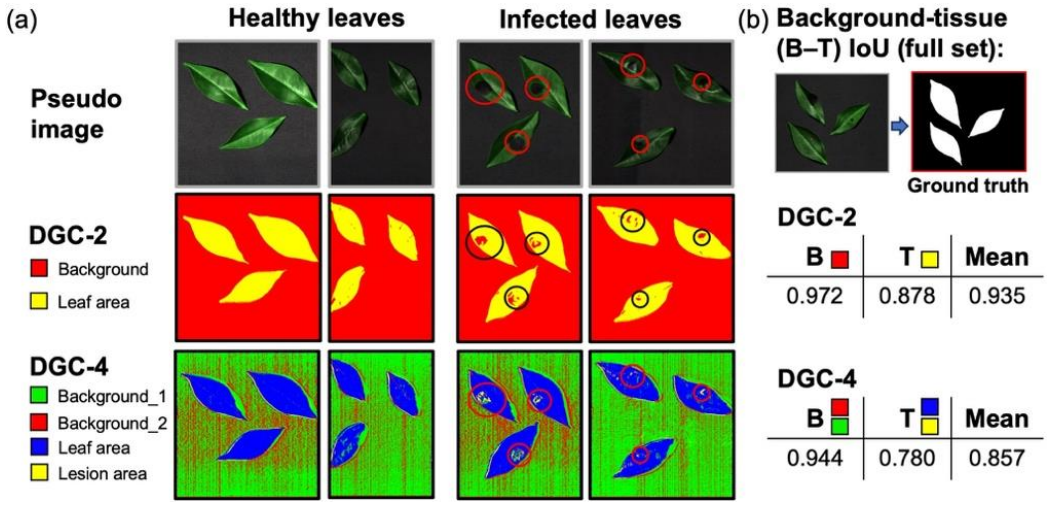

Fig. 3. DGC: 2- and 4-cluster results—pseudo-segmentation and B–T IoU.

Centroid initialization affected convergence; poorer starts required more iterations to activate all centroids. Even with this delay, patch-based training avoided full-dataset passes and reduced wall-time. The sampler processes two overlapping 64 × 64 patches per step and holds memory near 10 GB VRAM on a single consumer GPU. VRAM usage remains largely constant as image size grows, so large scenes and datasets are practical for applying DGC.

Two issues remained: residual noisy pixels and boundaries. Increasing the number of mean-shift iterations removes many stray pixels but can over-smooth boundaries in the latent space. Boundary errors likely stem from the encoder's limited receptive field; enlarging the field might improve edges but increases VRAM demand. Illumination variation across captures is another challenge; pixels with different apparent spectra cannot be grouped reliably by an unsupervised spectral model. When illumination shifts are strong, supervised calibration or radiometric normalization is recommended.

CONCLUSIONS

Deep Global Clustering (DGC) provides effective unsupervised HSI segmentation through patch-based training. The implementation runs on a consumer-grade GPU (~10 GB VRAM) and completes training within one hour. Key advantages include elimination of annotation requirements, scalable memory usage, and rapid convergence. Although the initialization sensitivity and boundary noise require further refinement for robust deployment, DGC provides a practical workflow for rapid exploration and early-stage analysis of HSI datasets.

REFERENCES

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(5), 603-619.

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.