

CNN과 Transformer를 결합한 하이브리드 딥러닝 기반의 가시광-근적외선 초분광 영상 가스 식별 연구

A Hybrid Deep Learning Approach Combining CNN and Transformer for Gas Classification in Visible - NIR Hyperspectral Image

김 은 엽¹, 허 석 행¹, 조 용 진¹, 방 혜 원¹, 김 성 호^{1,*}
 (Eunyeob Kim¹, Seokhaeng Heo², Yongjin Jo², Hyewon Bang², and Sungho Kim^{1,*})

¹Department of Robotics and AI Convergence Yeungnam University

²LIG Nex1 Co.

Abstract: Hyperspectral gas classification plays a critical role in chemical threat detection, industrial safety, and environmental monitoring. However, conventional RGB imaging is inherently limited for gas discrimination since many gases appear transparent or indistinguishable in the visible spectrum. In this study, fog, phytoncide, and H₂O vapor were used to construct gas environments that cannot be reliably differentiated using RGB imagery alone. In contrast, visible - near-infrared (VIS - NIR) hyperspectral imaging reveals distinctive spectral signatures that enable practical gas discrimination. To exploit these characteristics, this paper presents HyMixer, a CNN - Transformer hybrid model for VIS - NIR hyperspectral gas classification that jointly learns spatial and spectral representations while prioritizing efficient full-image inference. Experimental results show that VIS - NIR hyperspectral data provide clear separability among gas types. Compared to a strong baseline model, HyMixer achieves substantially improved inference efficiency, with a slight tradeoff in classification accuracy. This study provides a practical foundation for VIS - NIR hyperspectral gas identification in time-sensitive scenarios and can be extended to the identification of hazardous gases, including real chemical, biological, and radiological (CBR) agents, in future applications.

Keywords: hyperspectral imaging, gas identification, Visible - NIR spectroscopy, CNN - Transformer hybrid, spectral - spatial feature extraction

1. 서론

가스 식별(gas classification)은 군사, 산업, 환경 등 다양한 분야에서 화학물질 유출 감지, 오염 모니터링, 화학방(chemical biological radiological, CBR) 위협 대응 등 중요한 역할을 수행하는 핵심 기술이다. 특히 유해 가스의 조기 탐지와 식별은 인명 피해를 최소화하고 상황 인식을 향상시키는 데 필수적이다[1,2]. 그러나 기존의 RGB 영상 기반 감시 시스템은 가시광 영역에서 대부분의 가스가 투명하거나 균일한 백색 형태로 나타나기 때문에, 가스의 종류나 농도 차이를 구분하기 어렵다는 한계가 있다.

이러한 한계를 극복하기 위한 대안으로 초분광 영상(hyperspectral imaging, HSI) 기반 가스 식별 기술이 주목받고 있다. 초분광 영상은 수백 개의 연속적인 스펙트럼 밴드를 통해 물질의 화학적 조성 및 분자 구조에 따른 미세한 스펙트럼 차이를 분석할 수 있어, 기존의 단일 밴드 영상으로는 불가능

했던 물질 식별이 가능하다[3,4].

특히 가시광-근적외선(Visible-NIR, 470-900 nm) 영역의 초분광 영상은 반사 및 산란 특성을 기반으로 다양한 물질의 고유한 스펙트럼 변화를 감지할 수 있어 실시간 모니터링 및 저비용 센서 기반 시스템에 적용하기 적합하다[5,6]. 그러나 초분광 가스 식별 연구는 대부분 중적외선(MWIR, 3-5 μ m) 또는 장적외선(LWIR, 8-12 μ m) 영역에서 이루어져 왔으며, 가시광-NIR 대역에서 가스 식별을 다룬 연구는 매우 제한적이다[7,8]. VIS-NIR 영역은 비용이 낮고 센서 확보가 용이하다는 장점이 있음에도 불구하고, 분광 차이가 미세하고 환경 영향(산란·반사)도 커서 연구 접근 자체가 어려웠기 때문이다. 이러한 점에서 가시광-NIR 초분광 영상 기반 가스 식별 모델을 구축·평가하는 본 연구는 기존 문헌에서 거의 다뤄지지 않은 새로운 영역에 해당한다.

한편, 최근 딥러닝 기반 초분광 영상 분석 연구가 활발히

*Corresponding Author

Manuscript received December 16, 2025; revised December 22, 2025; accepted January 16, 2026

김은엽: 영남대학교 로봇AI융합학과 대학원생(yup4943@yu.ac.kr, ORCID[®] 0009-0009-3905-0236)

허석행: LIG Nex1 복합무인통제연구소 수석연구원(seokhaeng.heo@lignex1.com, ORCID[®] 0000-0003-0820-8220)

조용진: LIG Nex1 복합무인통제연구소(yongjin.jo@lignex1.com, ORCID[®] 0000-0002-3749-7570)

방혜원: LIG Nex1 복합무인통제연구소(hyewon.bang@lignex1.com, ORCID[®] 0009-0002-6787-5330)

김성호: 영남대학교 로봇AI융합학과 교수(sunghokim@yu.ac.kr, ORCID[®] 0000-0002-5401-2459)

※ 이 연구는 LIG NEX1 산학협력과제지원으로 연구되었음.

진행되면서, 물체 및 재질 식별 분야에서 우수한 성능이 보고되고 있다[3]. CNN (Convolutional Neural Network)은 공간적 특징을 추출하는 데 효과적이며, 스펙트럼-공간 융합 구조를 통해 지역적 패턴을 학습하는 연구들이 제안되었다[9,10]. 그러나 CNN 기반 모델은 긴 분광 차원에서의 전역적 스펙트럼 의존성을 충분히 반영하기 어렵다는 한계를 가진다[11]. 이를 보완하기 위해 ViT 구조가 도입되어 밴드 간 상관관계를 고려한 전역적 문맥 정보를 학습하는 방식이 제안되고 있으며[11, 12], CNN과 Vision Transformer (ViT)를 결합한 하이브리드 구조 또한 활발히 연구되고 있다[13]. 이후에 논문에서 언급하는 Transformer는 모두 ViT이다.

하지만 대부분의 기존 딥러닝 기반 초분광 분류 모델은 패치 기반(patch-based) 학습 구조를 사용하며, 입력 패치 1개로 오직 단일 픽셀만을 예측하는 pixel-by-pixel 추론 방식을 채택한다. 이러한 방식은 전체 이미지 추론 시 각 예측 위치마다 입력 패치를 구성하여 모델을 반복 호출해야 하므로, 결과적으로 영상 해상도 $H \times W$ 에 비례하여 연산 시간이 증가하는 구조적 병목을 가진다. 가스 식별과 같이 영상 전체를 빠르게 분석해야 하는 응용에서는 이러한 한계가 치명적일 수 있다. 이에 본 연구에서는 단일 패치(11×11) 입력으로 $N \times N$ 영역의 다중 픽셀을 병렬적으로 예측할 수 있는 새로운 추론 구조를 도입하여, 모델 호출 횟수를 기존 pixel-by-pixel 방식 대비 현저히 감소시킴으로써 전체 영상 추론 속도 문제를 근본적으로 완화하고자 한다.

본 연구에서는 CNN의 지역적 특징 추출 능력과 Transformer의 전역적 스펙트럼 상관관계 학습 능력을 결합한 CNN-Transformer 하이브리드 모델 HyMixer를 제안한다. 제안된 모델은 가시광-NIR 초분광 영상을 입력으로 하며, 서로 다른 가스 환경에 대해 정교한 분광-시공간 특징을 동시에 학습하도록 설계되었다. 또한 본 연구에서는 안전상의 이유로 실제 위험 가스를 직접 취급하지 않고, 이를 대체할 수 있는 세 가지 위험 가스 유사 물질인 Fog, Phytocide, H_2O 를 실험 대상으로 사용하였다. 이 세 물질은 RGB 영상에서는 모두 백색 연무 형태로 나타나 구분할 수 없지만, 초분광 영상에서는 서로 다른 스펙트럼 반사 특성을 나타낸다.

본 연구의 핵심 기여도는 다음과 같이 정리할 수 있다.

- RGB 영상 대비 초분광 영상의 가스 식별 능력 검증
- 국내 최초의 가시광-NIR 초분광 기반 가스 식별 연구
- $N \times N$ 다중 픽셀 병렬 예측 구조를 통한 패치 처리 모델의 추론 속도 개선.

2장에서는 제안하는 방법을 설명하고, 3장에서는 제안하는 데이터셋 구축 방법을 설명하고, 4장에서는 실험 및 결과를 분석하며, 5장에서는 결론을 맺는다.

II. 제안하는 방법

본 연구에서는 가시광-근적외선(Visible-NIR) 영역의 초분광 영상을 입력으로 하여 공간적·분광적 정보를 동시에 학습할 수 있는 CNN-Transformer 기반 하이브리드 모델 HyMixer를 제안한다. HyMixer는 Zhang et al.의 CTMixer [13]에서 제안된 dual-branch 구조 (CNN Branch + Transformer Branch)를 계승하되, 초분광 가스 영상의 특성과 실제 환경에서 요구되는

전체 이미지 추론 속도를 동시에 개선하기 위해 내부 모듈을 전면적으로 재설계한 구조이다. 특히 기존의 패치 기반 초분광 식별 모델이 입력 패치로부터 단일 픽셀만을 예측하는 pixel-by-pixel 방식을 사용함으로써, 고해상도 영상 처리 시 동일한 모델을 수천 번 반복 호출해야 하는 구조적 병목을 갖는데 반해, HyMixer는 단 한 번의 패치 입력으로 $N \times N$ 영역의 다중 픽셀을 병렬적으로 예측할 수 있는 고속 추론 구조를 중심 설계 목표로 설정하였다. 이러한 구조적 목표를 실현하기 위해, HyMixer는 다음의 두 핵심 모듈을 기반으로 설계되었다. 그림 2는 HyMixer의 전체 Architecture를 나타낸다.

1. ConvLSTM 기반 백본 특징 추출

기존의 합성곱 기반 모델은 주로 공간적 특징에 집중하였으나, 본 연구에서는 스펙트럼의 연속성을 함께 반영하기 위해 모든 합성곱 계층을 ConvLSTM (Cell) 구조로 대체하였다. ConvLSTM은 합성곱 기반의 게이트 연산을 통해 공간적 및 분광적 특징을 동시에 처리하며, 초분광 영상의 밴드 간 상호작용을 학습할 수 있다.

그림 1은 ConvLSTM의 내부 연산 구조를 나타낸다. ConvLSTM은 입력 특징 F_t , 이전 은닉 상태 H_{t-1} , 그리고 셀 상태 C_{t-1} 을 입력으로 받아, 입력 게이트(i_t), 망각 게이트(f_t), 출력 게이트(o_t)를 통해 시점 t 의 은닉 상태 H_t 와 셀 상태 C_t 를 산출한다. 각 단계의 연산은 다음과 같다.

$$\begin{aligned} i_t &= \sigma(W_{x_i} * X_t + W_{h_i} * H_{t-1} + W_{c_i} * C_{t-1} + b_i) \\ f_t &= \sigma(W_{x_f} * X_t + W_{h_f} * H_{t-1} + W_{c_f} * C_{t-1} + b_f) \\ C_t &= f_t * C_{t-1} + i_t * \tanh(W_{x_c} * X_t + W_{h_c} * H_{t-1} + b_c) \\ o_t &= \sigma(W_{x_o} * X_t + W_{h_o} * H_{t-1} + W_{c_o} * C_t + b_o) \\ H_t &= o_t * \tanh(C_t) \end{aligned} \quad (1)$$

이때 σ 는 시그모이드 함수, $*$ 는 요소별 곱(element-wise product)을 의미한다. ConvLSTM은 순환 구조를 통해 밴드 간의 연속적 상호작용을 학습할 수 있어, 초분광 데이터의 분광적 연속성을 효과적으로 반영한다. ConvLSTM의 입력 채널은 그룹 수($S=8$)로 나누어떨어져야 한다. 따라서 ConvLSTM 백본은 1×1 Conv를 통해 밴드 수를 $B = 150 \rightarrow n_b = 160$ 으로 확장한 뒤, 3단의 ConvLSTM 블록을 통해 채널을 점진적으로 축소한다($160 \rightarrow 122 \rightarrow 86 \rightarrow 48$). 이와 같은 차원 축소는 단순히 파라미터 절감을 위한 것이 아니라, Transformer의 분광 임베딩 단계에서의 연산 복잡도 감소를 목표로 한다. Transformer의 MHSA (Multi-Head Self-Attention) 연산량은 입력 토큰 수 N 에 대해 다음과 같은 수식으로 증가한다.

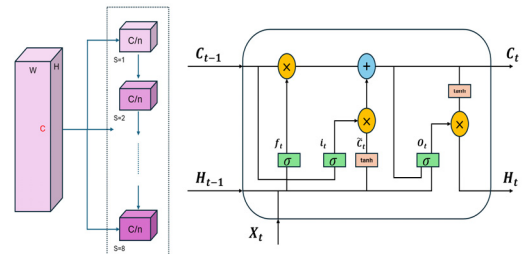


그림 1. ConvLSTM Architecture [10].

Fig. 1. ConvLSTM Architecture [10].

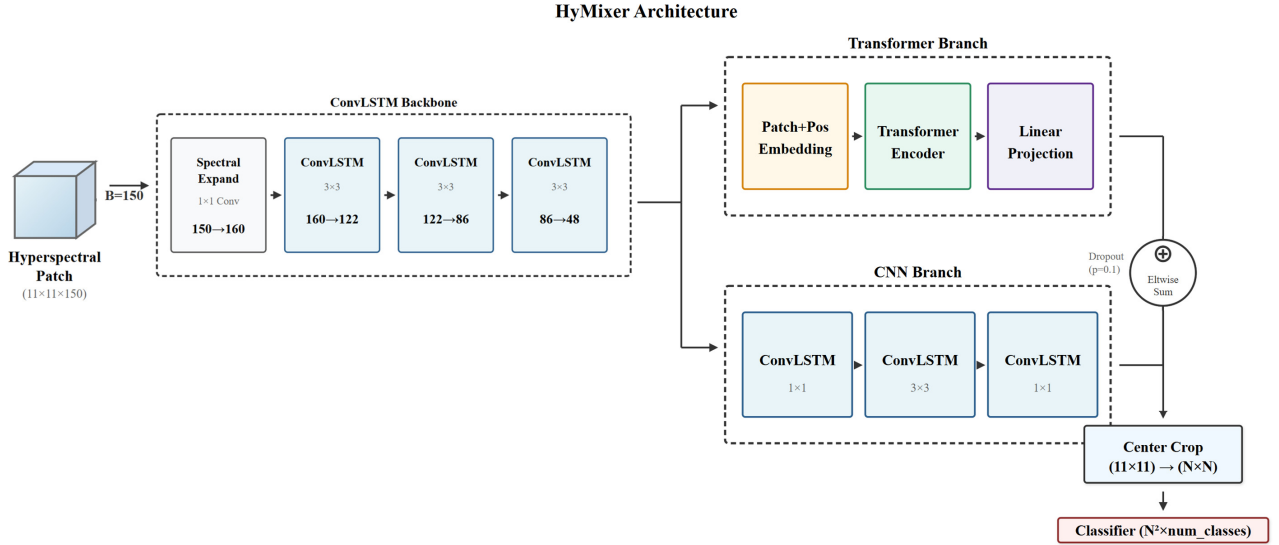


그림 2. 제안된 HyMixer Architecture.

Fig. 2. Proposed HyMixer Architecture.

$$\text{Self-Attention 연산량} = O(N^2 * D) \quad (2)$$

(※D는 임베딩 차원)

즉, 밴드 차원(토큰 수)에 비례하여 제곱적으로 연산량이 증가하기 때문에, 고밴드 HSI 입력의 경우 계산 비용이 급격히 커진다.

이에 본 연구에서는 ConvLSTM을 이용하여 밴드별 특징을 압축함으로써, 스펙트럼 정보를 손상시키지 않으면서 연산량을 효율적으로 감소시켰다. 실제 구현 실험에서 ConvLSTM 차원 축소 적용 시, 기존 연산 시간의 약 3/8 수준으로 줄어 들었다(80.530ms → 31.030ms). 이는 Self-Attention의 연산량이 N^2 에 비례함을 고려할 때, 밴드 수 감소가 직접적으로 연산 효율 개선에 기여함을 의미한다. 즉, ConvLSTM은 단순한 특징 압축기가 아닌, 분광적 연속성을 유지한 상태에서 Transformer 연산의 병목을 완화하는 핵심 전처리 블록으로 작동한다. 따라서 ConvLSTM을 통해 얻어진 압축된 특징 맵은, 이후 Transformer 단계에서 분광 임베딩 학습을 수행하기 위해 활용된다.

2. 분광 임베딩 기반 Transformer

ConvLSTM 백본을 통해 얻어진 출력은 Patch Embedding과 Positional Embedding을 거쳐 Transformer 입력으로 변환된다. 이때 Transformer는 공간 방향이 아닌 분광 축(spectral dimension)을 따라 밴드 간의 전역적 상관관계를 학습하며, 스펙트럼 내

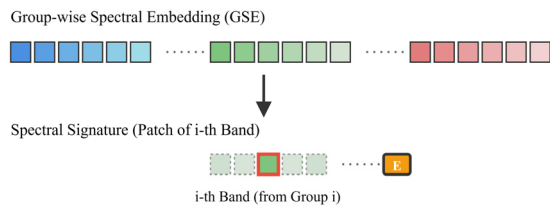


그림 3. SpectralFormer [11]의 GSE (Group-wise Spectral Embedding).

Fig. 3. Group-wise Spectral Embedding (GSE) module from SpectralFormer [11].

상호작용을 토큰 간 Attention 연산으로 모델링한다. Transformer 기반 분광 토큰 생성 단계에서는 그림 3과 같이 SpectralFormer에서 제안된 GSE (Group-wise Spectral Embedding) 기법을 적용하였다[11]. GSE는 입력 분광 밴드를 일정한 그룹 단위로 묶고, 각 밴드가 속한 그룹 내에서 인접한 전·후 밴드의 정보를 함께 고려하여 임베딩을 수행하는 방식이다. 이를 통해 단일 밴드가 갖는 스펙트럼 정보를 주변 밴드의 분광 맥락과 함께 표현함으로써 더욱 풍부하고 연속적인 분광 표현을 학습하도록 돕는다. Transformer의 출력은 선형 변환(linear projection)을 통해 임베딩 차원 E에서 공간 해상도 11x11로 복원되며, 이후 CNN Branch의 출력과 동일 해상도 상에서 요소별 덧셈(eltwise sum)으로 융합된다. 이 과정을 통해 모델은 전역적 스펙트럼 문맥과 지역적 공간 패턴을 통합적으로 학습할 수 있다.

III. 데이터셋 취득 및 촬영 구성

1. 가스 선정 및 촬영 환경 제약

본 연구에서는 가스 식별 연구를 안전하게 수행하기 위하여, Fog, Phytocide, H₂O의 세 가지 가스를 실험 대상으로 선정하였다. 이들은 모두 가시광 영역에서는 백색 연무 형태로 관측되어 RGB 영상 기반 구분이 어렵지만, 분자 조성, 굴절률, 입자 크기 분포가 서로 다르기 때문에 VIS-NIR(470-900 nm) 영역에서는 상이한 분광 패턴을 나타낸다. Fog는 프로필렌글리콜(C₃H₈O₂) 및 트리에틸렌글리콜(C₆H₁₄O₄) 기반의 glycol 계열 에어로졸(aerosol)이며, Phytocide는 C₁₀H₁₆ 계열의 유기

표 1. 가스 3종 종류 및 특징.

Table 1. Characteristics of the three gas types used in the experiment.

가스 종류	특징
Fog	프로판디올 C ₃ H ₈ O ₂ , 트리에틸렌글리콜 C ₆ H ₁₄ O ₄ 등
Phytocide	C ₁₀ H ₁₆ 등이 포함된 유기 화합물
H ₂ O	드라이아이스로 기화된 H ₂ O

화합물 에어로졸로 확산되며, H_2O 는 드라이아이스 기화 시 CO_2 가스 자체는 투명하지만, 냉각 효과로 인해 공기 중 수증기가 응축되어 생성된 물 기반 에어로졸로 관측된다. 따라서 Fog, Phytoncide, H_2O 는 모두 에어로졸 형태라는 공통점을 가지지만, 각각의 액적 조성·광학 상수·입자 크기 분포의 차이에 의해 초분광 영상에서 상이한 반사 및 산란 특성이 나타난다 [14-16].

2. 초분광 카메라 시스템(Imec Snapscan VNIR)

데이터 취득에는 Imec사의 Snapscan VNIR 초분광 카메라를 사용하였다. 해당 시스템은 470-900nm의 가시광-근적외선(Visible-NIR) 영역을 커버하며, 약 150개 이상의 스펙트럼 밴드를 제공한다.

이 카메라는 그림 4에 나타난 것처럼 센서 내부 이동형 구조를 가진다. 그림 4에서 파란색, 녹색, 빨간색 블록은 각각 서로 다른 분광 밴드에서 획득된 단일 파장 이미지를 의미한다. 실제 실험에서 사용된 초분광 영상은 총 150개의 연속적인 분광 밴드로 구성되지만, 본 그림에서는 초분광 데이터의 획득 원리와 시간적 특성을 직관적으로 설명하기 위해 대표적인 3개 밴드만을 예시적으로 시각화하였다. 각 색 블록은 서로 다른 파장에서 순차적으로 획득된 프레임을 나타내며, 이들이 결합되어 하나의 초분광 데이터 큐브를 형성한다. 즉, 카메라와 피사체가 모두 정지한 상태에서 센서가 카메라 내부에서 선형으로 이동하면서 장면을 순차적으로 촬영하고, 이 프레임들을 조합하여 하나의 초분광 데이터 큐브를 생성한다. 이 구조는 높은 공간적·분광적 해상도를 달성할 수 있다는 장점이 있지만, 프레임 취득 과정에 시간이 개입한다는 특징을 가진다. 따라서 실험 대상이 정적이지 않을 경우, 시간 축에서의 변화가 스펙트럼 축에 영향을 미칠 수 있다. 즉, 움직이는 가스 환경에서는 가스의 스펙트럼 특성보다 분사 패턴이나 농도 분포의 시공간적 변화가 더 지배적으로 반영될 가능성이 있다. 이는 곧 딥러닝 기반 분류 모델이 스펙트럼 정보를 학습하기보다 가스 확산의 형태적 차이를 기준으로 식별할 위험을 내포한다.

3. 데이터 촬영 환경 구축

촬영은 초분광 센서가 가시광-근적외선(Visible-NIR) 대역의 분광 정보를 활용하기 때문에, 태양광이 차단된 실내 환경에서는 NIR 대역의 스펙트럼을 확보할 수 없어 실험 수행이 불가능하였다. 반면 공간 제약이 없고, 바람이나 미세한 공기

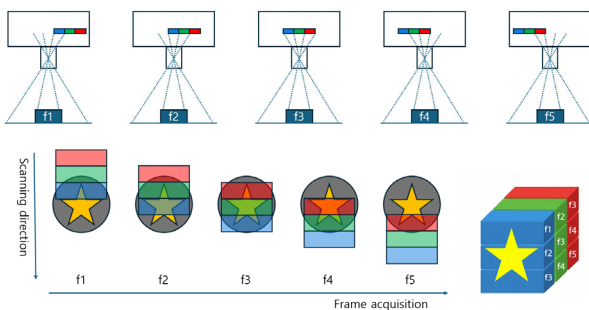


그림 4. Snapscan VNIR 초분광 카메라의 작동 원리.

Fig. 4. Operating principle of the Snapscan VNIR hyperspectral camera.

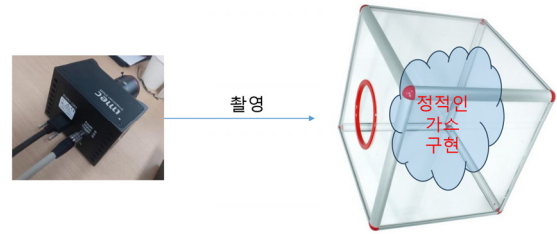


그림 5. 초분광 데이터 취득 시 가스를 정적인 상태로 유지하기 위해 제작된 투명 아크릴 챔버.

Fig. 5. Transparent acrylic chamber designed to maintain static gas conditions during hyperspectral data acquisition.

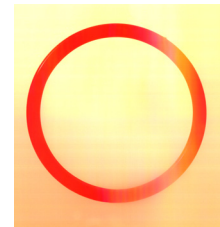


그림 6. Fog 촬영 예시.

Fig. 6. Example of fog environment.

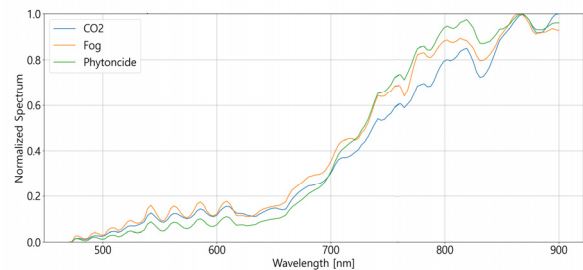


그림 7. Fog, Phytoncide, H_2O 스펙트럼

Fig. 7. Spectral signatures of Fog, Phytoncide, and H_2O

흐름과 같은 환경 요인을 통제할 수 없는 실외 환경에서는 가스 분사기의 분사량의 한계로 가스가 충분 시간 동안 머무를 수 있는 안정적인 공간을 확보하기 어려웠다. 이러한 이유로, 각 가스를 외부 환경 요인의 영향을 받지 않는 상태에서 안정적으로 유지하기 위해 그림 5와 같이 투명 아크릴 박스 내부에 가스를 가두어 정적인 환경을 조성한 뒤 실험을 진행하였다. 이와 같은 폐쇄형 구조를 통해 외부 광원은 태양광 조건을 유지하면서도, 가스의 흐름이나 확산 패턴에 의한 교란 없이 순수한 가스의 분광 패턴을 취득할 수 있도록 하였다.

데이터 취득은 3종의 가스를 모두 합해 총 18개의 초분광 이미지로 구성하였다. 그림 6은 Fog 환경에서 촬영된 실제 초분광 프레임의 예시로, RGB 영상에서는 구분되지 않던 미세한 차이가 초분광 스펙트럼에서는 뚜렷하게 나타남을 확인할 수 있다.

이와 같이 정적인 가스를 생성하기 위해 아크릴 박스를 활용함으로써, 가스의 스펙트럼 특성을 왜곡시키는 시간적 요인을 최소화하였고, 각 가스의 반사·산란 기반 분광 특성을 안정적으로 확보할 수 있었다. 그림 7은 각 기체별 스펙트럼을 나타낸다.

본 연구에서는 이 3개의 스펙트럼의 고유한 특징을 시각적으로 찾기는 어렵지만 이를 딥러닝으로 판별하고자 한다.

IV. 실험 및 결과

본 장에서는 제안한 HyMixer의 성능을 정량적으로 평가하기 위해 RGB 영상과 초분광 영상(HSI)을 각각 입력으로 사용하여 가스 식별 실험을 수행하였다. 또한 최근 초분광 영상 분류 분야에서는 다양한 딥러닝 기반 모델들이 제안되고 있으나, 가스 식별을 대상으로 하면서 가시광-근적외선(VIS-NIR) 영역의 초분광 영상을 고려한 연구는 극히 드물다. 이러한 배경에서 본 연구에서는 CNN과 Transformer를 결합한 대표적인 초분광 영상 분류 모델인 CTMixer를 주요 비교 대상으로 선정하였다. CTMixer는 최근 초분광 분류 연구에서 널리 활용되는 CNN-Transformer 하이브리드 구조를 기반으로 하며, 패치 기반 픽셀 추론 방식을 채택하고 있어 제안한 모델과 구조적 비교가 가능하다. 따라서 실험에서는 CTMixer와 HyMixer를 동일한 조건에서 비교함으로써, 제안한 다중 픽셀 병렬 추론 구조와 ConvLSTM 기반 분광 채널 축소 설계가 분류 성능과 추론 효율 측면에서 기존 구조 대비 어떠한 이점을 제공하는지를 분석하였다.

모든 실험은 Ubuntu 20.04 LTS 운영체제에서 Python 3.12.4 및 PyTorch 2.5.1+cu121 환경을 기반으로 수행되었다. 실험에 사용된 하드웨어 환경은 AMD Ryzen Threadripper 3970X (32-core CPU), 24 GB VRAM을 갖는 NVIDIA TITAN RTX GPU, 그리고 DDR4 (2666 MT/s) 256GB 시스템 메모리로 구성되었다. 제안한 모델은 PyTorch 딥러닝 프레임워크를 이용하여 구현되었으며, 학습 과정에서는 Adam 옵티마이저를 사용하고 초기 학습률은 $1e-3$, 가중치 감쇠 계수는 $1e-4$ 로 설정하였다. 손실 함수로는 교차 엔트로피 손실 함수(cross-entropy loss)를 사용하였다.

본 실험에서 RGB 기반 가스 식별에는 SegNet을, 초분광 기반 가스 식별에는 제안한 HyMixer 모델을 각각 사용하였다 [17]. 그림 8은 Fog, Phytioncide, H₂O 세 가지 가스 환경에 대해 RGB 영상과 초분광 영상 각각을 입력으로 했을 때의 분류 결과를 시각적으로 비교한 것이다. 표 2는 그림 8의 결과에 대한 정량적 성능 지표를 요약한 것으로, RGB와 초분광 영상에 대해 OA (Overall Accuracy), AA (Average Accuracy), Kappa 계수를 비교하였다. 초분광 영상 기반 식별의 경우 세 가스를 사실상 완벽하게 구분하는 성능을 보였다. 이는 가시광 영역에서는 거의 동일한 백색 연무로 보이지만, 가스마다 서로 다른 분자 구성과 입자 크기 분포에 의해 초분광 영역에서 미세하게 달라지는 분광 반사·산란 특성이 존재하기 때문이다. 반면 RGB 기반 식별의 경우 전반적인 분류 성능이 매우 낮게 나타났다. 여기서 47% 수준의 OA가 나타난 것도 모델이 실제로 가스의 종류를 구분한 것이 아니라, 투명 아크릴 박스의 테두리·유리 부분과 배경을 일정 부분 구분한 결과가 성능 지표에 반영된 것이다. 세 가스 본체 영역만을 기준으로 보면 RGB 영상에서는 서로를 안정적으로 구별하기 어렵고, 시각적으로도 거의 동일하게 보인다. 또한 본 연구에서는 학습 데이터를 구성할 때 각 초분광 이미지에서 전체 GT 픽셀 중 클래스별 5%만을 무작위로 샘플링하여 학습에 사용하였다. 이는 과도

한 학습 데이터 의존성을 방지하고 최소한의 데이터로도 초분광 기반 식별이 가능한지를 검증하기 위한 설정이다. 추가로, 초분광 영상 기반 모델이 매우 높은 성능을 보임에 따라 과적합 가능성을 검증하기 위해 입력 초분광 데이터에 Gaussian Noise($\sigma = 0.1$)를 주입하여 견고성(robustness)을 평가하였다. 그 결과 초분광+Noise 조건에서도 초분광 원본 대비 거의 동일한 수준의 성능을 유지하였다. 이는 HyMixer가 훈련 데이터의 특정 국소 패턴에 과적합 된 것이 아니라, 가스 스펙트럼 전반의 구조적 차이를 안정적으로 학습하고 있음을 의미한다.

따라서 그림 8과 표 2의 결과는, RGB 영상 기반 접근은 가스 식별 문제에 적합하지 않으며, 초분광 영상만이 세 가스 간의 분광적 차이를 활용하여 안정적인 식별이 가능함을 분명히 보여준다.

HyMixer는 CTMixer [13]의 dual-branch 구조를 계승하면서도, 하나의 입력 패치로 $N \times N$ 픽셀을 동시에 예측할 수 있는 다중 픽셀 추론 구조와 ConvLSTM + GSE 기반의 효율적인 분광 임베딩 구조를 도입하여 전체 추론 시간을 줄이는 것을 중요한 설계 목표로 삼았다. 표 3과 표 4에서는 이러한 구조 개선이 실제 구현 속도에 미치는 영향을 정량적으로 비교하였다.

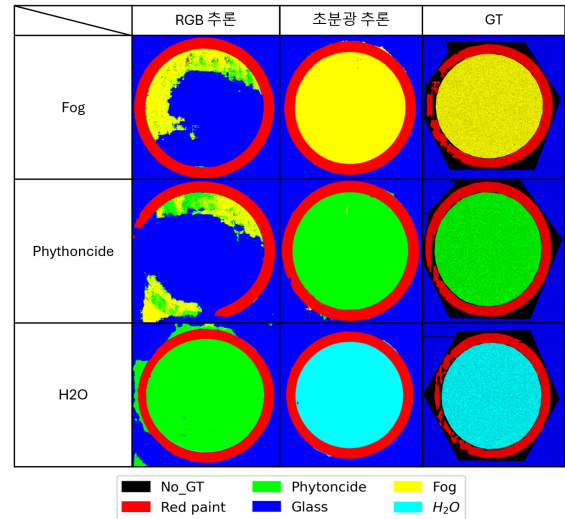


그림 8. RGB와 초분광 추론 결과 시각화.

Fig. 8. Visual comparison of RGB-based and hyperspectral-based inference results.

표 2. RGB와 제안된 초분광 모델 추론 성능 지표(%).

Table 2. RGB and proposed hyperspectral inference performance metrics (%).

	RGB	초분광	
	SegNet [17]	HyMixer (proposed)	HyMixer(proposed) + Noise
OA	47.38	99.93	99.90
AA	24.02	99.88	99.86
Kappa	24.78	99.88	99.82

표 3. CTMixer와 HyMixer의 H=1129, W=1087 해상도 이미지 추론 시간(초) 및 정확도(%).

Table 3. Inference time(sec) and accuracy(%) of CTMixer and HyMixer for images with resolution H=1129, W=1087.

	CTMixer [13]	HyMixer(다중 출력 사이즈: N×N)			
		3x3	5x5	7x7	9x9
시간	128	45	24	15	11
OA	99.98	99.96	99.95	99.94	99.93
AA	99.97	99.94	99.92	99.90	99.88
KAPPA	99.97	99.93	99.91	99.89	99.88

표 4. CTMixer와 HyMixer의 단일 추론 시간(ms), 파라미터 수, GFLOP.

Table 4. Single inference time (ms), the number of parameter comparison between CTMixer and HyMixer, GFLOP.

	CTMixer[11]	HyMixer
Conv Backbone	23.537	19.292
Transformer Branch	14.918	5.570
CNN Branch	1.897	5.578
Pooling & Classification	0.487	0.589
총 실행 시간	40.839	31.030
파라미터 수	623,708	278,426
GFLOP	68.03	39.12

표 3은 해상도 H=1129, W=1087을 갖는 초분광 이미지 18장에 대해, CTMixer [13]와 HyMixer를 각각 적용했을 때의 평균 추론 시간과 분류 정확도를 나타낸다. CTMixer [13]는 pixel-by-pixel 방식으로 전체 이미지를 추론하기 때문에, 한 장의 영상을 처리하는 데 약 128초가 소요되었다. 반면, 제안하는 HyMixer는 하나의 입력 패치로 N×N 영역을 동시에 예측하는 구조를 기반으로 하며, 추론 영역 크기를 3×3, 5×5, 7×7, 9×9로 변화시키면서 실험을 수행하였다. 그 결과, HyMixer의 평균 구현 시간은 패치 단위 예측 영역이 커질수록 전체 영상 추론 시간이 많이 감소하였다. 흥미로운 점은, 이와 같은 시간 감소에도 불구하고 OA, AA, Kappa는 모든 설정에서 99.93% 이상을 유지하며 사실상 거의 동일한 수준의 분류 성능을 보였다는 점이다. 즉, HyMixer는 CTMixer [13] 대비 최대 약 11배 이상(128초 → 11초)의 전체 구현 시간 단축을 달성하면서도, 가스 식별 정확도는 손실 없이 유지한다는 것을 표 3이 보여준다. 이는 III장에서 설계한 다중 픽셀 기반 추론 구조가 단순 이론적 이점에 그치지 않고, 실제 전체 이미지 단위 추론에서 매우 큰 효율 향상을 가져온다는 것을 의미한다.

표 4는 11×11 패치 1024개를 배치로 입력했을 때 CTMixer와 HyMixer의 모듈별 추론 시간을 비교한 결과로 대규모 패치를 병렬 처리하는 상황에서 두 모델의 구조적 효율 차이를 직접적으로 보여준다. 특히 Transformer Branch에서의 연산 시간이 크게 감소하였는데, 이는 ConvLSTM 기반의 채널 축소

와 GSE 기반 분광 임베딩을 통해 Self-Attention에서 처리되는 토큰 수가 줄어들어 연산량이 대폭 감소한 결과이다. 반면 CNN Branch는 HyMixer에서 다소 증가하였으나 전체 실행 시간에는 상대적으로 작은 영향을 미치며, Transformer Branch의 효율 향상이 더 크게 작용하여 최종적으로 약 24%의 추론 시간 감소가 나타났다. 또한 기존 CTMixer는 병렬 Convolution Branch를 구성한 뒤 element-wise summation으로 결합하는 2단 Backbone 구조를 사용한다. 반면 HyMixer는 ConvLSTM으로 스펙트럼 채널을 압축한 단일 Backbone으로 설계되어 전체 파라미터 수가 크게 감소하였다.

V. 결론

본 연구에서는 가시광-근적외선(VIS-NIR) 초분광 영상을 활용하여 서로 다른 가스 환경을 정밀하게 식별할 수 있는 하이브리드 딥러닝 모델 HyMixer를 제안하였다. 기존 CTMixer의 dual-branch 구조를 계승하면서도, 초분광 가스 영상의 특성과 실시간 활용 가능성을 고려하여 ConvLSTM 기반 분광-시공간 특징 추출, SpectralFormer의 GSE 기반 분광 임베딩, N×N 영역 병렬 예측 구조 등을 통합함으로써 모델의 정확도와 효율성을 동시에 향상시켰다.

실험 결과, RGB 영상은 세 가스(Fog, Phytocide, H₂O)를 안정적으로 구분하지 못했으나, 초분광 영상 기반 HyMixer는 OA·AA·Kappa 모두 99.9% 수준의 매우 높은 식별 성능을 달성하였다. 추론 속도 측면에서도 제한한 HyMixer는 pixel-by-pixel 방식의 CTMixer 대비 큰 성능 향상을 보였다. 전체 해상도(H=1129, W=1087) 영상 18장에 대한 평균 추론 시간은 CTMixer의 약 128 s에서 최대 11 s까지 감소하였으며, 11×11 패치 단위 입력 실험에서도 Transformer Branch 중심의 연산 효율 개선으로 총 실행 시간이 약 24% 단축되었다. 이는 HyMixer의 다중 픽셀 기반 병렬 추론 구조가 실제 전체 영상 처리 과정에서 매우 높은 실용적 이점을 제공함을 의미한다.

본 연구는 세 가지 가스 유사 물질(Fog, Phytocide, H₂O)을 활용하여 기본적인 가스 식별 가능성을 검증한 초기 단계 연구이지만, 향후 실제 화재방(CBR) 물질 및 유해 화학 가스 식별로 확장될 수 있는 기반을 마련하였다. 향후 연구에서는 더 다양한 가스 종류, 농도 변화, 외란 환경(안개, 바람, 조도 변화 등)을 포함하는 복합 환경 데이터셋 구축과 모델 및 하드웨어 최적화 연구를 진행할 예정이다.

REFERENCES

- [1] S. Kumar, C. Torres, O. Ulutan, A. Ayasse, D. Roberts and B. S. Manjunath, "Deep remote sensing methods for methane detection in overhead hyperspectral imagery," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020.
https://openaccess.thecvf.com/content_WACV_2020/html/Kumar_Deep_Remote_Sensing_Methods_for_Methane_Detection_in_Overhead_Hyperspectral_WACV_2020_paper.html
- [2] H. Park, H. Jang, J. Kim, S. Kim, and H. Oh, "EV-infotaxis: Extended vision-language model-based infotaxis for efficient gas source term estimation and search," *Journal of Instit*

- ute of Control, Robotics and Systems (in Korean), vol. 31, no. 6, pp. 643 - 653, 2025.
https://doi.org/10.5302/J.ICROS.2025.25.0049
- [3] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi and J. A. Benediktsson, "Deep learning for hyperspectral image classification: an overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690-6709, 2019.
https://doi.org/10.1109/TGRS.2019.2907932
- [4] J.-Y. Heo, U.-C. Jung, Y. Hwang, and Y. Choi, "A study on application of self-supervised learning for multispectral pedestrian detection," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 31, no. 5, pp. 526 - 535, 2025.
https://doi.org/10.5302/J.ICROS.2025.25.0034
- [5] Q. He, Z. Fang, O. Shoshanim, S. S. Brown, and Y. Rudich, "Scattering and absorption cross sections of atmospheric gases in the ultraviolet - visible wavelength range (307 - 725 nm)," *Atmospheric Chemistry and Physics*, vol. 21, no. 19, pp. 14927-14940, 2021.
https://doi.org/10.5194/acp-21-14927-2021
- [6] K. B. Ferreira, A. G. G. Oliveira, A. S. Gonçalves, and J. A. Gomes, "Evaluation of hyperspectral imaging visible/near infrared spectroscopy as a forensic tool for automotive paint distinction," *Forensic Chemistry*, vol. 5, pp. 46-52, 2017.
https://doi.org/10.1016/j.forc.2017.06.001
- [7] R. D. P. M. Scafutto and C. R. De Souza Filho, "Detection of methane plumes using airborne midwave infrared (3 - 5 μ m) hyperspectral data," *Remote Sensing*, vol. 10, no. 8, p. 1237, 2018.
https://doi.org/10.3390/rs10081237
- [8] L. A. Zimmerman and J. P. Kerekes, "Comparison of methane detection using shortwave and longwave infrared hyperspectral sensors under varying environmental conditions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2517-2531, 2023.
https://doi.org/10.1109/JSTARS.2023.3247246
- [9] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: exploring 3-D - 2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277-281, 2019.
https://doi.org/10.1109/LGRS.2019.2918719
- [10] Q. Zhu, W. Deng, Z. Zheng, Y. Zhong, Q. Guan, W. Lin, L. Zhang, and D. Li, "A spectral-spatial-dependent global learning framework for insufficient and imbalanced hyperspectral image classification," *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 11709-11723, 2021.
https://doi.org/10.1109/TCYB.2021.3070577
- [11] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, and A. Plaza, "SpectralFormer: rethinking hyperspectral image classification with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-15, 2021.
https://doi.org/10.1109/TGRS.2021.3130716
- [12] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, "HSI-BERT: hyperspectral image classification using the bi-directional encoder representation from transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 165-178, 2019.
https://doi.org/10.1109/TGRS.2019.2934760
- [13] J. Zhang, Z. Meng, F. Zhao, H. Liu, and Z. Chang, "Convolution transformer mixer for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022.
https://doi.org/10.1109/LGRS.2022.3208935
- [14] Y. Wu, V. Everett, E. Thomsen, and A. Blakers, "Investigation of the temperature dependence of the optical properties of thermal transfer fluids for hybrid CPV-T systems," *Proceedings of the AIP Conference Proceedings*, vol. 1556, no. 1, 2013.
https://doi.org/10.1063/1.4822205
- [15] D. Jakubczyk, G. Derkachov, K. Nyandey, S. Alikhanzadeh-Arani, and A. Derkachova, "Chromatic dispersion and thermal coefficients of hygroscopic liquids: 5 glycols and glycerol," *Scientific Data*, vol. 10, no. 1, p. 894, 2023.
https://doi.org/10.1038/s41597-023-02819-3
- [16] B. J. Sumlin, Y. W. Heinson, N. Shetty, A. Pandey, R. S. Pattison, S. Baker, W. M. Hao, and R. K. Chakrabarty, "UV - Vis - IR spectral complex refractive indices and optical properties of brown carbon aerosol from biomass burning," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 206, pp. 392-398, 2018.
https://doi.org/10.1016/j.jqsrt.2017.12.009
- [17] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.
https://doi.org/10.1109/TPAMI.2016.2644615



김 은 엽

2024년 영남대 전자공학과 졸업. 2024년~현재 동 대학원 석사과정. 관심분야는 딥러닝 기반 초분광 영상 분석 및 식별.



허 석 행

2009년~ LIG넥스원 기계연구소 연구원
2017년~ LIG넥스원 유도무기연구소 선임
2019년~현재 LIG넥스원 복합무인통제
연구소 수석연구원. 관심분야는 지상/수
중/공중 무인로봇, 인공지능 기반 의사결
정, 무기체계 M&S.



조 용 진

2013년 한양대학교 전자시스템공학과 대학원 석사. 2013년~현재 LIG넥스원 복합무인통제연구소 수석연구원(PM) 관심분야는 Image Enhancement, CNN, Autonomous Driving, Object Detection& Tracking.



방 혜 원

2021년 숭실대학교 스마트시스템 소프트웨어학과 졸업. 2023년 동 대학원 석사. 2024년~현재 LIG넥스원 복합무인통제연구소 연구원. 관심분야는 영상처리, 객체탐지, 자율주행, 딥러닝.



김 성 호

2000년 고려대 전기전자전파공학부 졸업. 2002년, 2007년 KAIST 석사/박사. 2007~2010년 ADD 선임연구원, 2010년~현재 영남대학교 전자공학과 교수. 관심분야는 EO/IR/HSI/LiDAR/RADAR 딥러닝 응용.