

# 지도 학습 기반 CLIP 을 활용한 레이더 신호 스펙트로그램 식별

윤재혁, 남해운  
한양대학교

serp82@hanyang.ac.kr, hnam@hanyang.ac.kr

## Radar Signal Spectrogram Identification Using Supervised Learning-Based CLIP

Jaehyeok Yoon, Haewoon Nam  
Hanyang Univ.

### 요 약

저피탐 레이더 파형 식별은 낮은 신호 대 잡음비 환경에서 기존 단일 모달리티 모델의 성능 한계에 직면한다. 본 논문에서는 이를 극복하기 위해 시간-주파수 이미지와 속성 기반 텍스트로 구성된 멀티모달 데이터 쌍을 구축하고, 지도 학습 기반의 대조 손실을 적용한 식별 기법을 제안한다. 제안하는 기법은 텍스트의 의미론적 정보를 활용하여 클래스 간 분리도를 극대화함으로써, 극한 잡음 환경에서도 강건한 식별 성능을 달성하였다.

### I. 서 론

최근 레이더 신호의 자동 변조 식별은 전자전뿐 아니라 자율주행, 드론, 스마트 교통체계 등 다양한 분야에서 그 중요성이 빠르게 증가하고 있다. 특히 자율주행 시스템에서는 악천후나 야간과 같은 가시광 기반 센서(CCTV, LiDAR 등)의 성능이 저하되는 상황에서, 레이더는 환경 인식의 마지막 보루로 기능한다. 이에 따라 레이더 기반 신호 식별 기술은 미래 지능형 센서 플랫폼의 핵심 요소로 주목받고 있다.

그러나 최근에는 탐지를 회피하기 위한 저피탐 레이더가 등장하면서, 기존 신호 분류 알고리즘의 적용이 한계에 직면하고 있다. LPI 신호는 넓은 대역폭과 낮은 출력, 복잡한 변조 방식 등을 통해 일반 수신기로는 식별이 어렵도록 설계된다. 이러한 신호는 특히  $-10$  dB 이하의 낮은 신호 대 잡음비(signal to noise ratio, SNR) 환경에서는 그 고유한 특성이 잡음에 묻히거나 심하게 왜곡되어, 기존의 CNN 기반 이미지 분류 모델로는 신뢰할 만한 식별이 어렵다 [1]. 실제로 대부분의 기존 연구들은 1차원 레이더 신호를 시간-주파수 변환을 통해 이미지로 변환한 후 CNN 계열의 모델에 입력하는 방식을 사용해 왔으며 [2], 이는 이미지의 시각적 특징에 과도하게 의존하는 한계를 보였다.

이에 본 논문에서는 이러한 이미지 중심의 단일 모달리티 기반 접근의 한계를 극복하고자, 시각 정보와 신호 속성 정보를 결합한 멀티모달 학습 기반 레이더 신호 식별 모델을 제안한다. 특히 인간 전문가가 레이더 신호를 식별할 때 사용하는 물리적, 구조적 지식을 반영하기 위해, 속성 기반 텍스트 설명을 이미지와 함께 학습에 활용한다. 본 연구에서는 이러한 텍스트 설명을

Vision-Language Model(VLM) 프레임워크에 통합하고, 기존 이미지-텍스트 페어 학습을 수행하는 기존 CLIP [3] 모델의 비지도 대조 손실 함수 대신, 클래스 레이블 정보를 활용한 지도 대조 손실 함수를 도입함으로써 낮은 SNR 환경에서도 강건한 신호 식별 알고리즘을 제안한다.

### II. 본론

본 논문에서는 시간-주파수 도메인 정보를 효과적으로 표현할 수 있는 고해상도 스펙트로그램과, 신호의 물리적 특성을 설명하는 텍스트 데이터를 결합하여 멀티모달 입력을 구성하였다.

기존의 Short-Time Fourier Transform(STFT)는 시간 및 주파수 해상도 간의 트레이드오프 문제가 있어, LPI 파형의 미세한 구조를 포착하는 데 한계가 있다. 이를 극복하기 위해 본 연구에서는 Smoothed Pseudo Wigner-Ville Distribution (SPWVD)를 활용하여 1차원 레이더 신호를 2차원 시-주파수 이미지로 변환하였다. SPWVD는 높은 시간-주파수 해상도를 제공하면서도, STFT에서 발생하는 교차항 간섭을 효과적으로 억제할 수 있어, 저 SNR 환경에서도 신호의 패턴을 선명하게 유지할 수 있다. 레이더 신호의 클래스를 단순히 파형의 종류로 지정하는 대신, 본 연구에서는 각 클래스의 구조적 특성과 물리적 의미를 언어적 설명으로 변환하여 활용하였다. 각 클래스는 기하학, 질감, 물리 속성으로 구분되며, 이를 바탕으로 “The frequency changes linearly over time.”, “It appears as a scattered pattern of blocks.”와 같은 의미 기반 텍스트 설명을 구성하였다. 이처럼 생성된 언어 기반 라벨은 스펙트로그램 이미지와 함께 이미지-텍스트 페어로 입력되어, 모델이 저 SNR

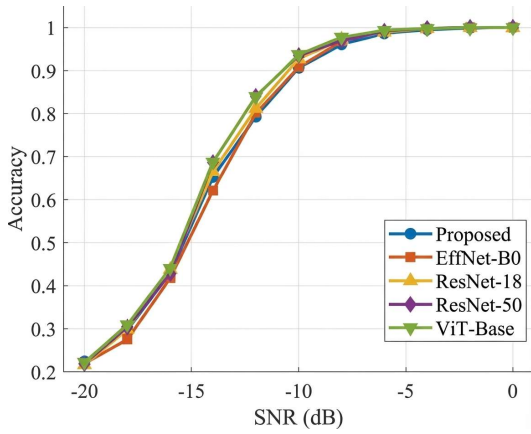


그림 1 다양한 SNR 환경(-20dB ~ 0dB)에서의 제안하는 VLM 과 기존 분류 모델들간의 식별 정확도 비교

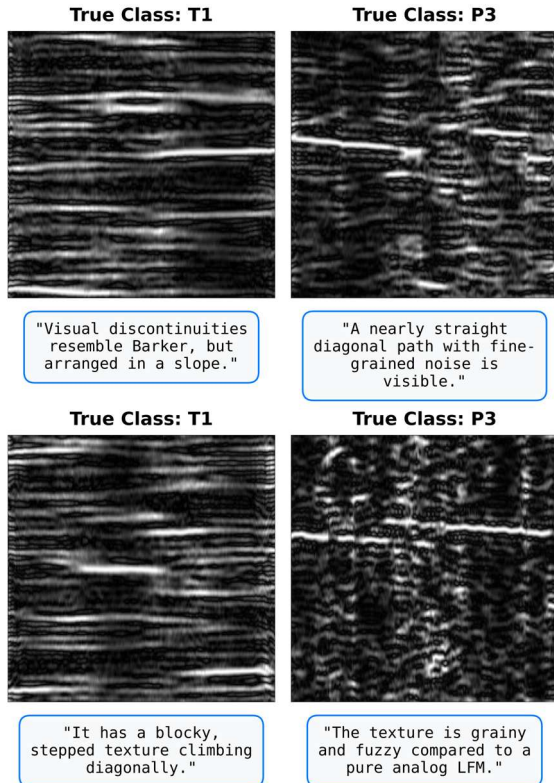


그림 2 제안하는 모델의 정성적 분석 결과 예시. 각 레이더 파형의 SPWVD 이미지에 대해 모델이 예측한 가장 높은 유사도의 텍스트 설명을 시각화하여, 모델의 판단 근거에 대한 설명 가능성을 입증함.

환경에서도 중요한 특징에 집중할 수 있도록 유도한다.

제안하는 모델은 CLIP 구조를 기반으로 한 VLM 형태로, 이미지 인코더와 텍스트 인코더로 구성된다. 이미지 인코더에는 Vision Transformer 를 사용하여 시간-주파수 이미지 내 전역적인 시각 특징을 추출하고, 텍스트 인코더에는 Transformer 기반 언어 모델을 적용하여 속성 기반 캡션으로부터 의미론적 임베딩을 생성한다. 각 인코더의 출력은 동일한 차원의 잠재 공간으로 매핑되며, 이를 통해 이미지와 텍스트 간의 의미적 유사도를 학습한다. 이러한 구조는 잡음이 많은 환경에서도 레이더 신호의 핵심 구조를 효과적으로 구분하고, 클래스 간 식별 성능을 향상시키는 데 기여한다.

기존 CLIP 모델은 이미지-텍스트 페어 간 유사도를 학습하기 위해 비지도 대조 손실을 사용하며, 자신의 페어만을 양성 샘플(positive)로 간주하고 나머지는 모두

음성 샘플(negative)로 처리한다. 그러나 명확한 클래스 레이블이 존재하는 경우, 같은 클래스에 속한 다른 샘플을 음성으로 간주하는 것은 학습 효율을 저해하고 클래스 내 분산을 증가시킬 수 있다. 이를 해결하기 위해 본 연구에서는 Supervised Contrastive Loss(SupCon) [4]를 적용하였다. 해당 손실 함수는 같은 클래스에 속하는 모든 샘플을 양성 샘플로 정의하여, 클래스 내부의 응집도를 높이고 클래스 간 경계를 명확하게 학습하도록 유도하며, 다음과 같이 정의된다.

$$L_{\text{SupCon}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)} \quad (1)$$

여기서  $z_i$  와  $z_p$  는 각각 앵커(Anchor)와 양성(Positive) 샘플의 투영 공간상의 정규화된 임베딩 벡터이고,  $\tau$  는 온도 매개변수이다. 또한  $I$  는 미니 배치 내 모든 샘플의 인덱스 집합,  $|P(i)|$  는 양성 샘플의 개수를 의미하며, 분모의  $z_a$  는 앵커  $i$  를 제외한 미니 배치 내의 모든 샘플의 임베딩 벡터를 나타낸다.

본 실험에서는 -20dB 에서 0dB 사이의 다양한 SNR 환경에서 제안하는 모델과 기존 최신 이미지 분류 모델들의 성능을 비교 분석하였다. 그림 1 에서 확인할 수 있듯이, 제안하는 기법은 전체 SNR 구간에서 베이스라인 모델들과 동등하게 우수한 식별 정확도를 달성하였다. 제안하는 모델은 단순한 클래스 분류를 넘어, 판단의 근거를 텍스트로 제시하는 설명 가능성을 제공한다. 그림 2 는 모델이 입력된 SPWVD 이미지에 대해 학습된 속성 기반 텍스트 설명 중 가장 적합한 문장을 매칭한 결과를 보여준다.

### III. 결론

본 논문에서는 저 SNR 환경의 레이더 신호 식별을 위해 SPWVD 이미지와 속성 텍스트를 결합한 지도 학습 기반 멀티모달 기법을 제안하였다. 실험 결과, 제안하는 모델은 극한 잡음 환경에서도 최신 베이스라인 모델과 동등한 수준의 정확도를 달성하였으며, 텍스트 매칭을 통해 판단 근거를 제시하는 설명 가능성을 입증하였다. 향후 퓨샷(few-shot) 및 제로샷(zero-shot) 학습으로 확장하여 실제 전자전 환경에서의 활용성을 높이고자 한다.

### ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 디지털분야글로벌연구지원 연구결과로 수행되었음(RS-2024-00428465).

### 참 고 문 헌

- [1] P. E. Pace, *Detecting and Classifying Low Probability of Intercept Radar*. Norwood, MA, USA: Artech House, 2009.
- [2] M. Zhang, L. Liu, and M. Diao, "LPI radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, p. 1682, 2016.
- [3] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 8748-8763.
- [4] P. Khosla *et al.*, "Supervised contrastive learning," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 18661-18673.