초분광 영상 기반 얼굴 위변조 탐지 및 신원 인식을 위한 CNN-MoE 프레임워크

김준성, 최영인, 이흥노
광주과학기술원

wnstjd123@gm.gist.ac.kr, young_in@gm.gist.ac.kr, heungno@gist.ac.kr

An HSI-based CNN-MoE Framework for Face Anti-Spoofing and Identification

Kim Jun Sung, Choi Young In, Lee Heung No
Gwangju Institute of Science and Technology

요 약

본 논문은 얼굴 위변조 공격에 대한 안정성을 향상시키기 위해 초분광 영상 (Hyperspectral Imaging, HSI) 기반 얼굴 위변조 탐지 프레임워크인 CNN-MoE FAS 를 제안한다. HSI 는 실제 피부와 위조 매체를 구분하는데 유용한 재질·분광 단서를 제공하지만, 데이터의 고차원성과 파장 간 높은 상관관계로 인해 계산 비용이 증가하고 중복 정보가 많아지는 문제가 있다. 이를 해결하기 위해 VGG-style CNN 백본과 희소 게이팅 Mixture-of-Experts (MoE) 모듈을 결합하고, 샘플별로 상위 k 개 전문가만 선택적으로 활성화하는 입력 적응형 라우팅을 통해 연산량을 제한하면서도 표현력을 확장한다. 제안한 프레임워크는 희소 전문가 선택이 위변조 분류에 필요한 재질·분광 단서와 신원 인증에 필요한 개인 고유 특성을 동시에 반영하는 표현 학습에 도움이 됨을 보여준다.

Ⅰ. 서 론

얼굴 인식은 효과적인 개인 인증 기술이나 RGB, 깊이, 적외선 카메라를 결합하는 시도에도 여전히 고품질 3D 마스크 공격에 대해서는 안전하지 않다 [1, 2].

이를 보완하기 위해 최근 연구들은 HSI 기반 얼굴 위변조 감지를 탐색해왔다[2, 3]. 초분광 센서는 여러 파장대에서 분광 특성을 포착하므로 재질 구분에 유효한 단서를 제공하여 실제 피부와 실리콘 마스크와 같은 위조 매체를 구분하는데 기여한다. 하지만 HSI 데이터는 차원이 높고 밴드 간 상관관계가 높기 때문에 계산 비용이 증가하고 추가적인 처리 과정을 요구한다 [4].

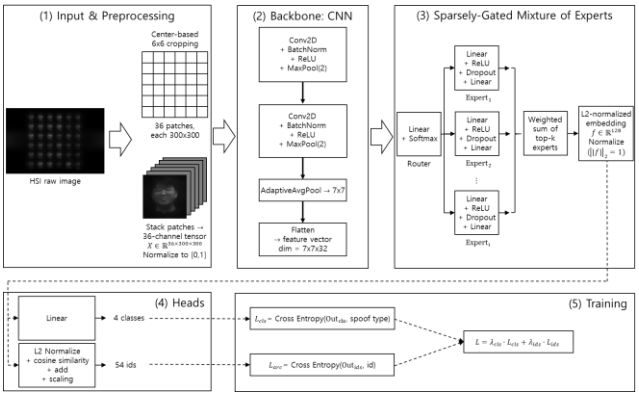
본 논문은 이를 해결하기 위해 CNN 백본과 희소 게이팅 MoE 모듈을 결합한 프레임워크를 제안하여, 연산 효율성과 다양한 위변조 공격 매체에 대한 안정성을 동시에 향상시켰다.

본 논문의 주요 기여는 (1) HSI 얼굴 위변조 탐지를 위한 희소 게이팅 CNN-MoE 프레임워크인 CNN-MoE FAS 를 제안하고, (2) HSI 얼굴 위변조 데이터셋을 구축하여 CNN-only 베이스라인 대비 성능 향상을 실험적으로 보였다.

Ⅱ. 제안 프레임워크

제안하는 프레임워크는 [그림 1]과 같다. 입력으로 원본 HSI 영상으로부터 추출한 6 x 6 비중첩 패치를 채널 방향으로 쌓은 7 x 7 x 128 차원의 분광 큐브를 사용한다.

이를 공유되는 VGG-style CNN 백본에 통과시켜 7 x 7 x 128 차원의 공유 특징 벡터 $x \in \mathbb{R}^D$를 생성한다. 이후 경량 게이팅 네트워크가 라우팅 확률을 기반으로 상위 k 개의 전문가를 선택하여 샘플 적응형 게이팅을 수행한다.



[그림 1] CNN-MoE FAS 개요

게이트는 먼저 $h(x) = xW + b$ 로 로짓(logit)을 계산한다. 여기서 W와 b는 각각 게이트의 가중치 행렬과 편향 벡터를 의미한다. 계산된 로짓은 softmax 를 통해 N 개의 전문가에 대한 확률 분포로 변환된다.

$$p_i(x) = \frac{e^{h_i(x)}}{\Sigma_{j=1}^{N} e^{h_j(x)}}, i = 1, \ldots, N.$$

희소 활성화를 수행하기 위해 상위 k 개의 확률만 유지하고 나머지 항목은 마스킹한다.

$$g_i(x) = \begin{cases} p_i(x), \text{if } i \in \text{TopK}(p(x), k) \\ -\infty, \text{otherwise} \end{cases}$$

여기서 $\text{TopK}(p(x), k)$ 는 $p(x)$ 에서 값이 큰 상위 k 개 성분의 위치를 반환하며, $g(x)$ 는 그 결과로 얻어지는 희소 게이팅 벡터를 나타낸다.

선택된 각 전문가 $E_i(\cdot)$는 공유 표현을 저차원 임베딩 공간으로 투영한다. 이후 전문가 출력은 게이팅 가중치를 이용해 다음과 같이 집계된다.

$$y = \sum_{i=1}^{N} g_i(x) E_i(x)$$

마지막으로 ℓ_2 정규화를 적용하여 최종 표현을 얻는다.

$$z = \frac{y}{\|y\|_2}$$

III. 실험

3.1 구현 세부 사항

CNN-MoE FAS 프레임워크는 전문가 4 개를 사용하여 총 20 epoch 동안 학습하였으며, 배치 크기는 64 로 설정하였다. 재현성을 확보하기 위해 랜덤 시드는 43 으로 고정하였다. 손실 함수에서 분류 항의 가중치 λ_{cls} 와 신원 인증 항의 가중치 λ_{ids} 는 모두 0.8 로 설정하였다. 따라서 전체 목적 함수는 $L = \lambda_{cls} \cdot L_{cls} + \lambda_{ids} \cdot L_{ids}$로 정의된다.

3.2 데이터셋

평가를 위해 본 연구에서는 [5]에서 제안된 초분광 카메라를 사용하여 54 명으로부터 총 4,374 장의 이미지로 구성된 HSI 얼굴 위변조 탐지 데이터셋을 구축하였다. 각 피험자에 대해 형광등, 전면 LED, 좌측 LED 의 3 가지 조명 조건에서 데이터를 수집하였으며, 마스크, 모자, 선글라스의 3 가지 액세서리 조건을 포함하였다.

실제 샘플의 경우, 각 액세서리 착용 조건은 5 가지 포즈로 촬영하였고, 액세서리 미착용 조건은 7 가지 포즈로 촬영하였다. 위조 샘플은 종이, iPhone, iPad 의 3 가지 공격 유형을 고려하였으며, 각 공격 유형은 5 가지 포즈로 촬영하였다.

3.3 평가 지표

분류 작업의 성능 평가는 위변조 제시 분류 오류율 (APCER), 정상 제시 분류 오류율 (BPCER), 평균 분류 오류율 (ACER), 그리고 정확도를 사용하였으며, 신원 인증 작업은 정확도로 평가하였다. APCER, BPCER, ACER 은 다음과 같이 정의된다.

$$\text{APCER} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \text{BPCER} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$

$$\text{ACER} = \frac{\text{APCER} + \text{NPCER}}{2}$$

여기서 FP, FN, TP, TN 은 각각 거짓 양성, 거짓 음성, 참 양성, 참 음성을 의미한다.

3.4 실험 결과

	Classification				Identification
	APCER	NPCER	ACER	Accuracy	Accuracy
CNN	97.53%	0.00%	48.77%	82.79%	73.75%
ours (CNN-MoE FAS)	12.96%	0.00%	6.48%	93.68%	75.93%

[표 1] 실험 결과

본 연구에서는 MoE 블록을 포함하지 않은 동일한 VGG-Style CNN 백본 모델을 비교 기준으로 사용하였다. 표 1 에 나타난 바와 같이 제안한 프레임워크는 APCER 을 84.57%p 감소시키고, 분류 정확도와 신원 인증 정확도를 각각 10.89%p. 2.78%p 향상시켰다. 이러한 결과는 희소 활성화된 전문가 게이팅 방식이 위변조 유형 분류에 필요한 재질·분광 단서와 신원 인증에 필요한 개인별 고유 특성을 동시에 반영하는 표현을 학습하는데 도움이 됨을 보여준다.

IV. 결론 및 향후 방향

본 연구에서는 VGG-style CNN 백본에 MoE 블록을 결합한 HSI 기반 얼굴 위변조 탐지 프레임워크를 제안하고, MoE 를 적용하지 않은 동일한 구조의 베이스라인과 비교하였다. 특히 FP 가 0 으로 나타나 제안 프레임워크의 높은 탐지 성능을 확인하였다. 하지만 이는 제한된 데이터셋 환경의 결과일 수 있기에 향후 다양한 위변조 매체와 피험자를 바탕으로 대규모 데이터셋을 구축하여 프레임워크의 일반화 가능성을 추가 검증하고자 한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2026-RS-2021-II211835) 그리고 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (RS-2025-22932973)

참 고 문 헌

[1] Shao, H., & Zhong, D. (2022). Towards open-set touchless palmprint recognition via weight-based meta metric learning. Pattern Recognition, 121, 108247.

[2] Rao, S., Huang, Y., Cui, K., & Li, Y. (2022). Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor. Optica, 9(11), 1253-1259.

[3] Song, C., Hong, Y., Lan, J., Zhu, H., Wang, W., & Zhang, J. (2024). Supervised Contrastive Learning for Snapshot Spectral Imaging Face Anti-Spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 980-985).

[4] Kurz, W., Wang, K., Bektas, F., Zhu, C., Kariper, E., Dong, X., ... & Koch, A. W. (2025). Dimensionality reduction in hyperspectral imaging using standard deviation-based band selection for efficient classification. Scientific Reports, 15(1), 34478.

[5] Kim, C., Ni, P., Lee, K. R., and Lee, H.-N. “Mass production-enabled computational spectrometers based on Multilayer thin films,” Scientific Reports, vol. 12, Art. No. 4053, Mar. 2022