LoRA-ViT 를 활용한 아동 그림 감정 인식

안양 [1], 조경은 [2]
[1] 동국대학교 자율사물지능학과
[2] 동국대학교 첨단융합대학 컴퓨터 AI 학부

2022126763@dgu.ac.kr, cke@dongguk.edu(교신저자)

Applying LoRA-ViT for Emotion Recognition in Children's Drawings

Yang An[1], Kyungeun Cho [2],
[1] Dept. of Autonomous Things Intelligence, Dongguk University
[2] Dept. of Computer Science and Artificial Intelligence,
College of Advanced Convergence Engineering, Dongguk University

요 약

본 연구는 아동 그림의 자동 감정 인식을 위한 머신러닝 기반 접근을 탐구하여, 정서 상태 진단의 객관성과 정확성을 향상시키고자 하였다. 이를 위해 ViT 기반 감정 인식 시스템에 LoRA 기법을 활용하였으며, ViT 가 갖는 높은 성능에도 불구하고 방대한 파라미터와 소규모 데이터셋에서의 한계를 보완하고자 하였다. 실험 결과, LoRA-ViT 기반 접근은 아동 그림 감정 인식 과제에서 기존 모델보다 우수한 성능을 보였다. 또한 학습에 앞서 ViT 에 대한 사전학습을 수행하여 시각적 특징 표현 능력을 강화하고, 이후 주 과제에서 모델의 표현력과 일반화 성능을 향상시켰다.

1. 서론

어린이의 그림은 언어 표현이 서툰 시기에 감정을 드러내는 중요한 수단으로, 색상과 구성을 통해 기쁨, 두려움, 불안 등을 표현한다. 따라서 그림 분석은 아동 정서 진단 방법 중 하나로 주목받고 있으며[1], 특히 언어적 표현이 어려운 시기에 감정 표현의 대체 수단이 된다.

오늘날 그림을 활용한 심리 검사는 아동의 성격 분석에서 가장 널리 쓰이는 도구이나[2], 기존 해석은 전문가의 경험에 의존해 주관적 오류와 편견의 한계가 있다. 이에 따라 아동의 감정 상태를 보다 객관적으로 진단하기 위해 아동 그림을 자동 분석하는 연구가 중요해졌으며, 이를 위해 지능형 정보 처리 기술을 활용할 수 있다 [3].

그래서 본 연구는 아동 그림을 활용한 감정 평가 프레임워크를 구축하고, LoRA 가 적용된 ViT 기반의 머신러닝 자동 인식 및 진단 시스템을 활용하는 것을 목표로 한다. 이를 통해 심리학자, 치료사, 보호자가 아동의 감정을 조기 선별하고 지속적으로 모니터링할 수 있도록 지원하고자 한다.

2. 관련 연구

아동 그림 감정 분석은 약 100 년의 역사를 지니며 [4], 아동은 흔히 그림으로 두려움, 기쁨, 슬픔을 표현한다[5]. 초기에는 전문가의 주관적 판단에 의존해 한계가 있었으나, 최근 딥러닝 도입으로 자동화와 객관화의 가능성이 열렸다[6].

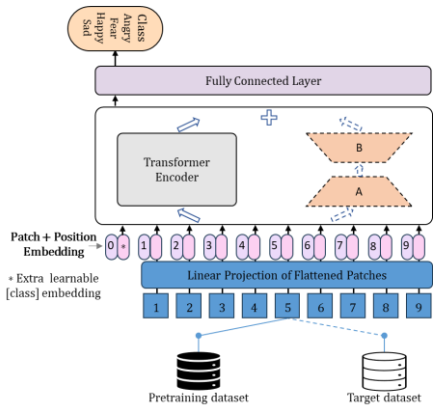
Lee 등 [7]은 고전적인 집-나무-사람(HTP) 검사를 기반으로 사전학습된 모델을 활용하여 공격성, 사회적 불안, 우울, 낮은 자존감과 관련된 심리적 특징을 분석하였다. Kissos 등 [8]은 CNN 구조를 사용하여 아동 자화상에서 성적 학대의 징후를 탐지하였으며, 그 결과 심층 모델이 잠재적 심리적 외상을 효과적으로 식별할 수 있음을 입증하였다.

또한, 연구자들은 그림과 발달 및 정서적 특성 간의 관계를 탐구하였다. 예를 들어 [9]에서는 CNN 을

활용하여 2- 9 세 아동 그림의 표상 발달 차이를 분석하였으며, Khan 등 [10]은 Transformer 모델을 아동의 필기 및 그림 감정 탐지에 적용하여 장기 의존성과 인지 과정 포착에서의 잠재력을 보여주었다.

3. 제안 방법

본 연구에서는 아동 그림 감정 인식을 위해 사전학습된 ViT(Vision Transformer) 구조에 LoRA(Low-Rank Adaptation) 기법을 적용한 학습 방식을 사용하였다. ViT 는 ImageNet 으로 사전학습된 가중치를 초기값으로 사용하여 전역 시각 표현을 유지하고, LoRA 는 Self-Attention 구조 내에 추가되어 모델의 표현력을 확장하였다.



(그림 1) **Extended E-LoRA-ViT 프레임워크**.

모델의 전체 구조는 그림 1 에 제시되어 있으며, 시스템은 세 가지 주요 구성 요소로 이루어진다.

1. 첫째, 이미지 인코더(ViT Backbone) 는 입력된 아동 그림을 패치 단위로 분할하고 Transformer 인코더를 통해 전역적인 시각 정보를 추출한다. 사전학습된 ViT-B/16 의 표현을 기반으로 하여 아동 그림의 색상과 형태적 특징을 안정적으로 포착한다.
2. LoRA 통합 학습 단계에서는 ViT 본체와 LoRA 모듈의 파라미터를 동시에 업데이트하였다. 이를 통해 사전학습된 표현을 유지하면서도 데이터셋 특성에 맞춘 세밀한 조정이 가능하였다. LoRA 는 Attention 가중치 경로에 삽입되어 ViT 의 기존 구조를 유지한 채 표현 범위를 확장한다.
3. 분류기(Classification Head) 는 인코더에서 출력된 [CLS] 토큰을 입력으로 받아 전결합층(MLP)을 통해 네 가지 감정 클래스(Angry, Fear, Happy, Sad)를 예측한다.

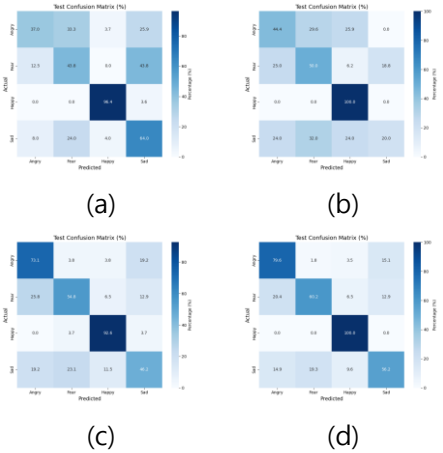
이와 같은 학습 구조를 통해 ViT 의 전역 표현력과 LoRA 의 적응 능력을 결합함으로써, 소규모 데이터 환경에서도 안정적인 감정 인식 성능을 확보한다.

실험에서는 아동 그림 데이터셋(1109 장)을 학습, 검증, 테스트(8:1:1)로 분할하여 LoRA 기반 미세조정을 수행하였다. 입력 그림은 ViT-Backbone 과 LoRA 모듈을 거쳐 특성 벡터로 변환되고, 최종 분류기에서 네 가지 감정 중 하나로 예측되었다.

4. 실험 및 결과

본 장에서는 LoRA 가 적용된 ViT(E-LoRA-ViT) 모델의 성능을 평가하기 위해 데이터셋, 비교 대상 모델, 실험 환경 및 결과를 상세히 기술한다.

실험에는 공개된 아동 그림 데이터셋(총 1109 장)을 사용했으며, 각 이미지는 Angry, Fear, Happy, Sad 네 감정으로 라벨링되었다. 모든 이미지는 256x256 크기로 리사이즈 후 RGB 로 변환되었고, 학습(80%), 검증(10%), 테스트(10%)의 비율로 분할하였다.



(그림 2) (a)ResNet , (b)ViT-B/16
(c) E-LoRA-ViT (d)Eextended E-LoRA-ViT

혼동 행렬 분석 결과, E-LoRA-ViT 모델이 ResNet-50 및 ViT-B/16 보다 전반적으로 우수한 분류 성능을 보였다. 기존 모델에서는 Fear 와 Sad 간 혼동이 뚜렷했으나, E-LoRA-ViT 는 Happy 와 Angry 클래스의 정확도가 향상되며 보다 균형 잡힌 예측을 보였다. 특히 ViT 사전학습(pretraining)을 적용한 모델은 모든 클래스에서 성능이 추가로 향상되어 일반화 능력이 강화되었다.

<표 1> 아동 그림 감정 인식 데이터셋에서의 모델 성능 비교

모델(Model)	Accuracy ↑	Precision ↑	Recall ↑	F1-Score ↑
ResNet-50	0.5982	0.6153	0.5982	0.5935
ViT-B/16	0.5446	0.5184	0.5446	0.5112
E-LoRA-ViT	0.6636	0.6589	0.6636	0.6657
Extended E-LoRA-ViT	**0.6818**	**0.6751**	**0.6818**	**0.6751**

표 1 의 결과에서 E-LoRA-ViT 모델은 ResNet-50 과 ViT-B/16 보다 높은 정확도(0.664)와 F1-score(0.666)를 기록하였다. 특히 사전학습을 추가로 적용한 Extended E-LoRA-ViT 는 Accuracy 0.682, F1-score 0.675 로 가장 우수한 성능을 보였다. 이는 ViT 의 사전학습이 아동 그림 감정 인식에서 특징 표현의 안정성을 높이고, LoRA 와의 공동 학습을 통해 일반화 성능을 향상시킨 결과로 해석된다.

5. 결론

본 연구에서는 사전학습된 ViT-B/16 구조에 LoRA 기법을 결합하여 아동 그림 감정 인식을 수행하였다. ViT 백본과 LoRA 모듈을 함께 학습함으로써 소규모 데이터 환경에서도 안정적인 수렴과 향상된 인식 성능을 확보할 수 있었다. 실험 결과, E-LoRA-ViT 는 기존 ResNet-50 및 ViT-B/16 보다 높은 정확도를 보였으며, 사전학습을 적용한 Extended E-LoRA-ViT 를 적용한 학습에서 가장 높은 성능이 관찰되었다. 이를 통해 사전학습된 ViT 의 표현력과 LoRA 의 적응성이 상호 보완적으로 작용하여 감정 인식의 일반화 성능을 향상된 경향을 보였다.

참고문헌

[1] P.-Y. Brandt, Z. Dandarova-Robert, C. Cocco, D. Vinck, F. Darbellay, “When Childre n Draw Gods. Multicultural and Interdisciplinary Approa ch to Children's Representations of Supernatural Agents", Cham, Springer, 2023.

[2] U. Podobnik, J. Jerman, J. Selan, “Understanding analytical drawings of preschool children: the importance of a dialog with a child”, International Journal of Early Years Education, Vol. 32, No. 1, pp. 189-203, 2024.

[3] Nashva Ali, Alaa Ali Abd-alrazaq, Zubair Shah, Mohannad Alajlani, Tanvir Alam, and Mowafa Househ, “Artificial Intelligence-Based Mobile Application for Sensing Children Emotion Through Drawings”, Advances in Informatics, Management, and Technology in Healthcare, 2022, doi: 10.3233/SHTI220674.

[4] F. L. Goodenough, “Measurement of Intelligence by Drawings,’’ New York: Harcourt, Brace, & World, 1925

[5] C. Kaplun, ‘‘Children’s drawings speak a thousand words in their transition to school,’’ Australasian Journal of Early Childhood, vol. 44, no. 4, pp. 392–407, 2018.

[6] C. A. Jensen, D. Sumanthiran, H. L. Kirkorian, B. G. Travers, K. S. Rosengren, and T. T. Rogers, ‘‘Human perception and machine vision reveal rich latent structure in human figure drawings,’’ Frontiers in Psychology, vol. 14, p. 1029808, 2023.

[7] M. Lee, Y. Kim, and Y.-K. Kim, ‘‘Generating psychological analysis tables for children’s drawings using deep learning,’’ Data & Knowledge Engineering, vol. 149, p. 102266, 2024.

[8] L. Kissos, L. Goldner, M. Butman, N. Eliyahu, and R. Lev-Wiesel, ‘‘Can artificial intelligence achieve human-level performance? a pilot study of childhood sexual abuse detection in self-figure drawings,’’ Child Abuse & Neglect, vol. 109, p. 104755, 2020.

[9] A. Philippsen, S. Tsuji, and Y. Nagai, ‘‘Picture completion reveals developmental change in representational drawing ability: An analysis using a convolutional neural network,’’ in 2020 joint ieee 10th international conference on development and learning and epigenetic robotics (icdlepirob). IEEE, 2020, pp. 1–8.

[10] Z. A. Khan, Y. Xia, K. Aurangzeb, F. Khaliq, M. Alam, J . A. Khan, and M. S. Anwar, ‘‘Emotion detection from ha ndwriting and drawing samples using an attention-based transformer model,’’ PeerJ Computer Science, vol. 10, p. e1887, 2024.