Identity-CBAM ResNet 기반 얼굴 감정 식별 모듈

오규태[1], 김인기[2], 김범준[2], 곽정환[3,*]
[1]한국교통대학교 컴퓨터공학과
[2]한국교통대학교 교통에너지융합학과
[3]한국교통대학교 소프트웨어학과
alswo740012@naver.com, cv2@kakao.com, zhfkskseh@gmail.com,
jgwak@ut.ac.kr

Face Emotion Recognition using ResNet with Identity-CBAM

Gyutea Oh[1], Inki Kim[2], Beomjun Kim[2], Jeonghwan Gwak[3,*]
[1]Dept. of Computer Engineering, Korea National University of Transportation
[2]Dept. of IT·Energy Convergence, Korea National University of Transportation
[3]Dept. of Software, Korea National University of Transportation
[*]Correspnding Author

요 약

인공지능 시대에 들어서면서 개인 맞춤형 환경을 제공하기 위하여 사람의 감정을 인식하고 교감하는 기술이 많이 발전되고 있다. 사람의 감정을 인식하는 방법으로는 얼굴, 음성, 신체 동작, 생체 신호 등이 있지만 이 중 가장 직관적이면서도 쉽게 접할 수 있는 것은 표정이다. 따라서, 본 논문에서는 정확도 높은 얼굴 감정 식별을 위해서 Convolution Block Attention Module(CBAM)의 각 Gate와 Residual Block, Skip Connection을 이용한 Identity- CBAM Module을 제안한다. CBAM의 각 Gate와 Residual Block을 이용하여 각각의 표정에 대한 핵심 특징 정보들을 강조하여 Context 한 모델로 변화시켜주는 효과를 가지게 하였으며 Skip-Connection을 이용하여 기울기 소실 및 폭발에 강인하게 해주는 모듈을 제안한다. AI-HUB의 한국인 감정 인식을 위한 복합 영상 데이터 세트를 이용하여 총 6개의 클래스로 구분하였으며, F1-Score, Accuracy 기준으로 Identity-CBAM 모듈을 적용하였을 때 Vanilla ResNet50, ResNet101 대비 F1-Score 0.4~2.7%, Accuracy 0.18~2.03%의 성능 향상을 달성하였다. 또한, Guided Backpropagation과 Guided GradCam을 통해 시각화하였을 때 중요 특징점들을 더 세밀하게 표현하는 것을 확인하였다. 결과적으로 이미지 내 표정 분류 Task에서 Vanilla ResNet50, ResNet101을 사용하는 것보다 Identity-CBAM Module을 함께 사용하는 것이 더 적합함을 입증하였다.

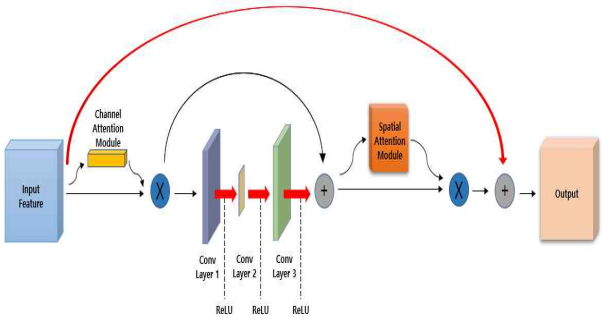
1. 서론

인공지능 시대에 들어서면서 개인 맞춤형 환경을 제공해주기 위하여 사람의 감정을 인식하고 교감하는 기술들이 많이 발전되고 있다[1]. 사람의 감정을 인식하는 방법으로는 얼굴, 음성, 신체 동작, 생체 신호 등이 있지만 이 중 가장 직관적이면서도 쉽게 접할 수 있는 것은 표정이다. 감정에 따른 표정을 분석하기 위해서는 안면 근육의 움직임에 따른 주름, 눈, 입 모양 등 이미지 속 작은 특징들과 위치 정보를 파악할 수 있어야 정확한 분석이 가능하다. 현재 영상 기반 감정분류 기반 딥러닝 모델의 기법들은 이미지 속 작은 특징점들을 명확히 추출하고 학습에 이용하기에 한계가 있어 이를 보완하기 위하여 영상 정보와 함께 다른 부가 정보를 통합하는 방식으로 연구가 진행되고 있다. 이는 이미지 이외에 다른 특징 정보들을 함께 사용하여야 하기에 많은 데이터를 요구한다. 따라서 본 논문에서는 다른 특징 정보들을 함께 사용하지 않고 이미지 데이터만을 이용하여 영상에서 나타나는 감정 분류 기반 딥러닝 모델이 주름, 눈, 입 모양 등 이미지 속 작은 특징들과 위치 정보를 파악할 수 있도록 해주는 Identity-CBAM Module을 제안한다. Convolutional Block Attention Module(CBAM)[2]을 기반으로 한 Identity-CBAM Module은 이미지에서 핵심 특징점의 위치와 중요도를 파악하여 기존 Input에 Attention을 부여할 수 있는 Module이다.

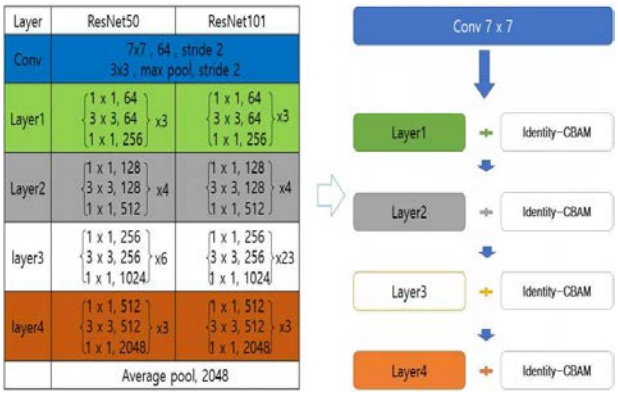
2. 제안하는 모듈

그림 1과 그림 2는 본 논문에서 제안하는 Identity-CBAM

Module의 전반적인 Architecture이다. Identity-CBAM Module의 경우 감정에 따라 다르게 나타나는 표정의 핵심 특징을 잘 추출하기 위해 Channel Attention Module[2]을 이용하여 학습한 특징에서 중요 특징 정보를 선별하고 Spatial Attention Module[2]을 통해 위치 정보를 고려해 중요 정보를 선별하여 Attention 효과를 부여한다. 그 결과, 기존의 딥러닝 모델로는 추출하기 힘들었던 세밀한 특징들을 더 잘 추출할 수 있다. 또한, 기존의 CBAM에 Bottleneck Residual Block[3]을 추가하여 계층을 통과하면서 추출된 핵심 특징점을 적은 파라미터를 통해서도 학습할 수 있도록 하였다. 더불어 Skip connection[4]을 추가하여 많은 layer를 거치게 되면 발생하는 기울기 소실 및 폭주[5]를 해결하고자 하였으며, Input에 Attention을 부여하였다.



(그림 1) Identity-CBAM Module Architecture



(그림 2) ResNet based Model + Identity-CBAM Module

3. 실험 및 결과

본 논문에서 제안하는 Identity-CBAM Module을 학습 및 평가하기 위해 사용된 데이터 세트는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 "한국인 감정 인식을 위한 복합 영상 데이터 세트"를 활용하여 수행하였으며, 총 6개의 감정 클래스를 사용하여 감정을 분류하였다. 각 클래스당 4,160장을 이용하여 모델을 학습시켰으며, 검증과 테스트 데이터 세트에서는 각 클래스당 360장을 사용하여 총 이미지 27,120장을 크기 256×256으로 실험을 진행하였다. 또한 실험 환경은 RTX 3080, PyTorch 1.11.0, CUDA 11.3을 사용하였으며, Identity-CBAM Module을 적용하기 위한 모델은 이미 성능이 검증된 SOTA(State-Of-The-Art) 모델 ResNet50, ResNet101[4] 모델을 선정하여 그림 2와 같이 사용하였다. 표 1과 표2를 통해 F1-Score 기준 ResNet50, ResNet101에서 약 0.4~2.7%의 향상, Accuracy 기준 약 0.18~2.03%의 성능 개선을 확인하여 본 모듈이 성능 향상 및 신뢰성 개선에 도움이 됨을 입증하였다.

<표 1> F1-Score based performance analysis

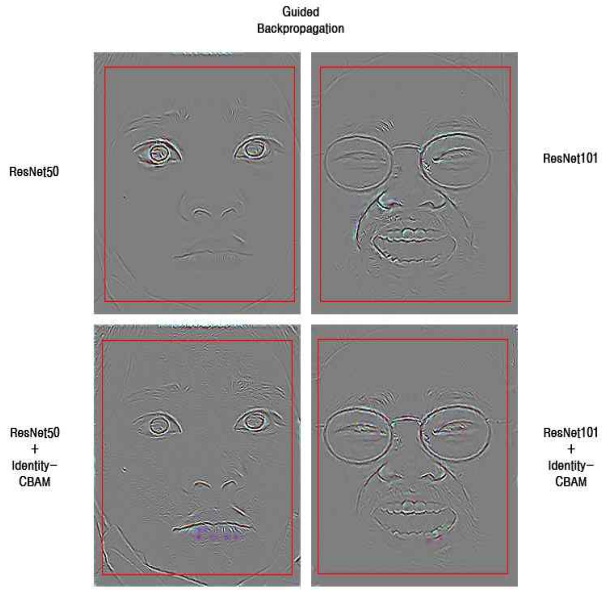
Class	Model			
	ResNet50	ResNet101	ResNet50 +Identity-CBAM	ResNet101 +Identity-CBAM
Happy	0.97	0.98	0.98	0.98
Anger	0.85	0.85	0.87	0.87
Unrest	0.69	0.71	0.73	0.72
Wound	0.61	0.64	0.69	0.66
Sad	0.79	0.80	0.81	0.78
Neutrality	0.93	0.92	0.92	0.91
Average	0.806	0.816	0.833	0.82

<표 2> Accuracy based performance analysis

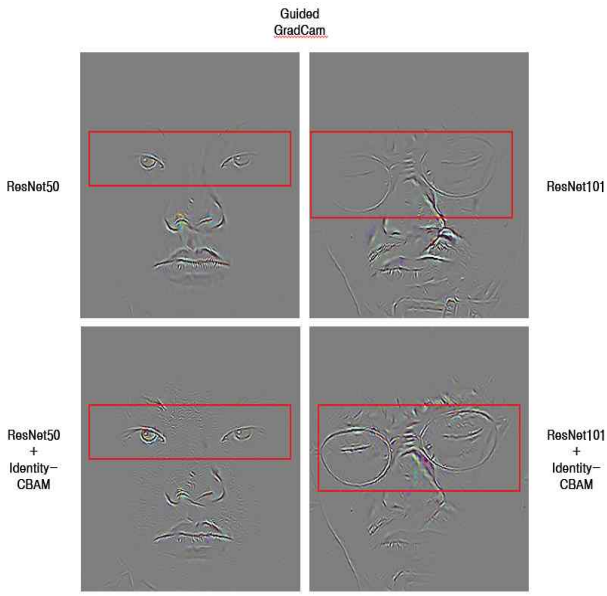
Model			
ResNet50	ResNet101	ResNet50 +Identity-CBAM	ResNet101 +Identity-CBAM
81.2	81.99	83.23	82.17

또한, 모델이 특징을 정확하게 추출했는지 검증 및 해석을 위해 딥러닝 모델이 추출한 특징들을 시각화하여 주는 Guided Backpropagation[6]을 사용하여 시각화를 진행하였다. 이를 통해 이미지 속 중요한 특징들을 잘 학습하여 분류에 이용함을 알 수 있었다. 그러나 Guided Backpropagation만으로 평가하기에는 부족함이 있기에 각 클래스에서 감정분류를 위한 특징들을 잘 학습하였는지 확인하기 위하여 Guided GradCam[6]을 추가적으로 적용하여 더 명확한 해석 결과를 도출하였다. 그림 3을 통해 기존 ResNet50, ResNet101 모델을 단일로 사용하는 경우보다 Identity-CBAM Module을 적용하여 Guided Backpropagation을 적용하였을 때 감정분류를 위한 특징을 전체적으로 더 뚜렷하게 추출하는 것을 확인하였다. 그림 4를 통해 추가적으로 진행한 Guided GradCam을 통해 시각화하였을 때는 이미지 속 눈동자를 더 명확하게 추출하는 것을 확인하였으며 특히, 안경을 쓴 경우 기존의 모델보다 더 명확하게 추출해 안경과 같은 폐색(Occlusion)에도 중요한 특징을 명확하게 추출하는 능력이 있음을 확인하였다. 결과적으로 얼굴 감정 분류 Task에 Vanilla ResNet50, ResNet101을 사용하는 것보다 중요 특징점을 효과적으

로 파악할 수 있는 특징을 가진 Identity-CBAM Module을 함께 사용하는 것이 적합함을 확인하였다.



(그림 3) Guided Backpropagation Visualization.



(그림 4) Guided GradCam Visualization.

4. 결론

본 논문에서는 정확도 높은 얼굴 감정 식별을 위해 CBAM의 각 Gate와 Residual Block, Skip Connection을 이용한 Identity-CBAM Module을 제안하였다. 실험 결과 Vanilla ResNet50, ResNet101 모델에 각 layer 당 1개의 Identity-CBAM Module 적용 시 약 F1-Score 기준 0.4~2.7%, Accuracy 기준 약 0.18~2.03%의 성능 향상을 보였다. Guided Backpropagation과 Guided GradCam을 통해 시각화할 때 이미지 내에서 추출한 특징점들이 더 세밀하게 표현되는 것을 확인하였으며, 폐색이 존재하여도 중요한 특징점을 추출할 수 있는 능력이 있음을 확인하였다. 결과적으로 Vanilla ResNet50, ResNet101을 사용하는 것보다 중요 특징을 강조해 줄 수 있는 Identity-CBAM Module을 함께 사용하는 것이 이미지 내 표정 분류 Task에 더 적합함을 확인하였다. 하지만 Neutrality 와 같은 클래스에서는 성능 향상이 이루어지지 않았기 때문에 이를 해결하기 위하여 추후 연구에서 Identity-CBAM Module 적용 위치 및 적용 개수 등을 추가로 실험하여 최적의 적용 위치 및 적용 개수 등을 찾아내 지금보다 더 고도화된 성능을 가지도록 하여 현재 발표된 다른 감정 식별 모델들과 성능 비교를 진행할 예정이다.

Acknowledgment

This work was supported in part by the "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE) (2021RIS-001 (1345341783))

참고문헌

[1] Hermann E. "Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective", New Media & Society, Vol. 24, No. 5, pp. 1258-1277, 2022

[2] SH Woo, JC Park, JY Lee and IS Kweon, "CBAM: Convolutional Block Attention Module", European conference on computer vision(ECCV), Vol. 3 No. 19 , 2018, pp. 3-19

[3] Zagoruyko S. and Komodakis, N, "Wide Residual Networks", arXiv preprint arXiv, Vol. 1605 No. 07146, 2016

[4] He K. Zhang and X Ren. S, "Deep Residual Learning for Image Recognition", Institute of Electrical and Electronics Engineers(IEEE), pp. 770-778, 2016

[5] Tan H. H, Lim K. H, "Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization", 2019 7th International Conference on Smart Computing & Communications (ICSCC), 2019, pp. 1-4

[6] Sekvaraju R. R, Das A. Vedantam, "Grad-CAM: Why did you say that?", arXiv preprint arXiv. Vol. 1611, No. 07450, 2016