

MPIIFaceGaze 데이터셋 기반 시선 예측 모델의 효율적인 전처리 기법 및 증강 기법

김나영¹·인다운¹·조시은¹·박태정^{2*}¹덕성여자대학교 사이버보안전공 학부과정²덕성여자대학교 디지털소프트웨어공학부 교수

Efficient Preprocessing and Augmentation Techniques for Gaze Prediction Models Based on the MPIIFaceGaze Dataset

Nayoung Kim¹ · Dawoon In¹ · Sieun Jo¹ · Taejung Park^{2*}¹Bachelor's Course, Department of Cybersecurity, Duksung Women's University, Seoul 01369, Korea²Professor, School of Computer Science, Engineering and Converged Technology, Duksung Women's University, Seoul 01369, Korea

[요약]

최근 인공지능과 컴퓨터 비전 분야의 빠른 발전에 따라 시선 추적 기술의 중요성이 증가하고 있다. 시선 추적 정확성 향상을 위해 제한적인 학습 데이터를 보완하기 위한 다양한 증강 기법들이 널리 활용되고 있으나 기존의 시선 추적 데이터 증강 연구들은 최신 시선 추적 기술의 특성을 올바르게 반영하지 못하고 있다. 따라서 본 연구에서는 현재 상용화된 최신 시선 추적 기술에 적합한 이미지 전처리 및 증강 기술의 조합을 제시한다. 다양한 환경 하에서 시선 추적 학습을 위해 촬영된 MPIIFaceGaze 데이터셋을 기반으로 통계적 이미지 전처리 기법(HE, CLAHE)과 함께 전형적인 이미지 증강 기술(Albumentation, Transform, Light-Enhancer)이 시선 추적 성능에 미치는 영향을 분석하였다. 이미지 전처리와 증강 기술을 적용하지 않는 경우도 포함하여 총 12가지 조합을 실험한 결과, CLAHE 전처리 기법과 이미지 증강 기술을 적용하지 않은 None 조합이 가장 우수한 성능을 보였다. 실험 결과, 과거 연구에서 크게 주목하지 않았던 통계적 이미지 전처리 기법으로 인한 성능 개선 효과를 얻었다.

[Abstract]

As artificial intelligence and computer vision rapidly advance, gaze tracking technology is becoming increasingly important. To address the limitations of training data, various augmentation techniques are commonly used to enhance gaze tracking accuracy. However, existing research on gaze tracking data augmentation has not fully considered the characteristics of the latest technologies. In this study, we propose suitable combinations of image preprocessing and augmentation techniques for modern gaze tracking systems. Using the MPIIFaceGaze dataset, we analyzed the effects of statistical image preprocessing methods (Histogram Equalization and Contrast Limited Adaptive Histogram Equalization) and typical augmentation techniques (Albumentation, Transform, Light-Enhancer) on gaze tracking performance. The experiments tested 12 combinations, including the absence of preprocessing and augmentation. The results showed that the combination of CLAHE preprocessing without augmentation achieved the best performance. The experimental results showed performance improvements due to statistical image preprocessing techniques that were not heavily emphasized in previous research.

색인어 : MPIIFacegaze, 이미지 전처리, 이미지 증강, 시선추적, Angular Error**Keyword** : MPIIFacegaze, Image Preprocessing, Image Augmentation, Gaze Estimation, Angular Error<http://dx.doi.org/10.9728/dcs.2024.25.10.2871>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 19 August 2024; Revised 20 September 2024

Accepted 18 October 2024

*Corresponding Author; Taejung Park

Tel: +82-2-901-8339

E-mail: tjpark@duksung.ac.kr

1. 서 론

1-1 연구배경

최근 인공지능과 컴퓨터 비전 분야에서 시선 추적 기술의 중요성이 대두되고 있다. 시선 추적 기술은 사용자의 시선을 파악하여 교육, 마케팅, 심리학 연구 등에 활용할 수 있다. 이는 다양한 환경 조건에서도 사용자의 시선을 정확히 추적하는 것이 핵심이다[1]. 시선 추적은 크게 model-based 방식과 appearance-based 방식으로 나뉜다. 특히 사용자의 이미지를 직접 활용하는 appearance-based 방식에 딥러닝이 도입되면서 정확성이 크게 향상되면서 실용성이 증가하였다[2]. 이러한 배경 하에 딥러닝 연구의 발전으로 대규모 데이터셋에 대한 필요성이 점점 커지면서 다양한 공개 데이터셋가 제안되었으나 대체로 한정된 환경에서 수집된 데이터셋은 다양성과 수량 부족, 일상 환경에서 수집된 데이터셋의 품질 저하 문제로 인해 활용성이 제한된다. 이러한 제약으로 인해 결과적으로 딥러닝 모델의 일반화 성능이 저하된다.

따라서 앞서 논의한 한계, 즉, 양적인 측면과 질적인 측면에서 학습 데이터의 부족을 극복하기 위해서 데이터를 강화할 수 있는 다양한 방식들이 제안되었고 실용화되었다. 대표적인 방식은 데이터셋 증강(dataset augmentation)[3]-[5]이 있으며 좀 더 나아가 같은 범주(class)에 해당되는 데이터를 증강시켜서 표현 공간(representation space)에서 같은 범주끼리는 가깝게 배치하고 다른 범주끼리는 멀리 배치하는 대비 학습(contrastive learning)[6] 기법이 활발하게 연구되고 있다.

그러나 이러한 강화 기반 기법들은 대부분 분류(classification) 문제라는 맥락에서 개발된 것이며 시선 추적과 같은 회귀(regression) 문제에 그대로 적용하면 기대했던 성능을 얻기가 힘든 경우가 많다. 본 연구에서는 시선 추적 모델의 성능 향상을 목표로 주요한 몇 가지 전처리 및 이미지 증강 방법을 조합하고 그 결과를 제시한다. 구체적으로 데이터 전처리 또는 증강 기법을 적용하지 않은 데이터를 포함한 총 12가지의 다른 조합을 실험하고 그 결과를 비교해 최적의 방법을 검토하며 평가지표는 예측된 시선 벡터와 정답 시선 벡터의 차이를 각도로 나타낸 Angular Error를 사용하였다. 실험 결과, 분류 문제와는 다르게 몇몇 이미지 증강 방식에서는 오히려 시선 추적 성능이 저하되는 결과를 얻었다. 본 논문에서는 이러한 성능 저하의 원인을 분석하고 실험 과정에서 얻은 경험을 통해 시선 추적과 같은 회귀 문제에서 데이터 증강을 적용하는 경우의 의의와 한계에 대해 논의한다.

1-2 관련 연구

1) 딥러닝 기반 시선 추적 기법

Gaze Capture[2]가 그전까지 인간의 안구에 대한 해부학

적인 구조와 기하 정보를 바탕으로 전용 장비를 이용해서 시선 추적을 수행하던 model-based 시선 추적 기법의 한계를 극복하고 딥러닝 기술로 모바일 기기 사용 중인 사용자의 얼굴 사진만으로(즉, appearance-based) 시선 추적이 가능하다는 사실을 증명한 이후, 딥러닝 네트워크를 활용한 여러 기술들이 발표되었다. 이러한 딥러닝 네트워크를 활용한 appearance-based 시선 추적 기술은 사용자의 얼굴 이미지의 특징을 추출하는 backbone network로 거의 대부분 CNN 계열의 네트워크를 사용한다. 그 이후 appearance-based 시선 추적 기법 연구는 대부분 개선된 네트워크 구조와 함께 여러 환경에서 디지털 기기를 사용 중인 사람들의 얼굴 이미지와 화면 상의 좌표를 제공하는 데이터셋들이 함께 공개되었다. 예를 들어 MPIIGaze[7]는 기존 시선 추적용 데이터셋이 주로 실험실 환경에서 수집되었던 한계를 극복하기 위해 15명의 사용자가 몇 달 간 다양한 환경에서 노트북 컴퓨터를 사용하면서 전면 카메라를 통해 촬영한 얼굴 이미지 37,667장과 눈 주변 지점, 입 주변 지점, 동공의 중심 지점 위치 정보를 제공한다. 또한 ETH-XGaze[8]는 통제된 실험실 환경에서 촬영한 시선 추적 데이터셋이지만 정확한 얼굴 각도, 조명 환경 관련 정보를 제공하며 동시에 기존 이미지와는 달리 18대의 SLR 카메라를 이용해서 110명의 사용자들을 다양한 카메라 각도에서 촬영한 6000 x 4000 크기의 고 해상도 이미지를 제공한다. 이러한 대규모 시선 추적 데이터들이 제공되면서 appearance-based 시선 추적 기법에 대한 실용화가 진행되었으나 사용자마다 안구의 해부학적 구조가 다르기 때문에 보다 정확한 시선 위치를 예측하기 위해서는 사용 전에 이미 학습이 종료된 네트워크 가중치를 사용자의 특성에 맞게 조정해야 한다. FAZE(Few-shot Adaptive Gaze Estimation)[9]은 비교적 적은 수의 데이터만으로도 학습이 가능한 few-shot learning[10]을 사용자 맞춤 시선 추적 조정 기술에 응용한 사례이다. 이러한 딥러닝 기반 시선 추적 기술이 레저, 장애인용 전동 휠체어 조종 등에 실용화되면서 그 전까지 대두되지 않았던 안전 문제가 중요해지기 시작했는데, 특히 눈을 감은 상태 또는 사용자의 얼굴이 카메라에 포착되지 않을 때에도 2차원 화면 상에서 시선 위치를 예측하는 appearance-based 방식의 한계를 극복하기 위해 매우 어두운 조명 하 또는 역광 상태에서 촬영한 어두운 얼굴 이미지에서도 사용자가 눈을 감았는지의 여부를 정확하게 예측할 필요가 발생하였다. 이러한 문제 해결을 위해 BEAT(Blink Estimation with Domain Adversarial Training) 네트워크[11]는 매우 어두운 조명 하에서도 여러 배경과 사용자의 인종적인 차이를 극복하고 높은 정밀도와 짧은 시간 내에 눈깜박임 여부를 판별한다.

2) 이미지 데이터 증강 기법

가장 기초적인 이미지 증강 기법은 이미지에 이동(translation), 회전(rotation), 확대/축소(scale), 반전(flip) 등의 선형 변환(transform)을 적용하는 것이다. PyTorch[12]에

서는 구성 모듈 중 하나인 torchvision을 통해 이러한 기초적인 선형 변환을 지원한다. Cutout[3]은 CNN(Convolutional Neural Network)에서 과적합(overfitting)을 막기 위해서 이미지 내 무작위 영역을 삭제함으로써 데이터를 증강시키는 기법이다. 이 기법은 학습된 인공지능망의 가치치 일부를 연산 중간에 제거하는 dropout[13] 기법처럼 데이터의 일부를 의도적으로 제거함으로써 정규화(regularization)를 적용한다. CTAugment[4]는 주어진 문제의 특성을 보다 적응적으로 잘 반영할 수 있도록 모델 학습 과정 중에 가장 적합한 증강 방식을 학습한다. RandAugment[5]는 모델 학습 복잡도를 증가시키는 증강 단계에서의 탐색 과정을 제거하고 다양한 모델 크기와 데이터 규모에 대응하기 위한 방안을 제시하였다. Dijk은 본 연구와 유사하게 데이터 증강 기법이 CNN 기반 시선 추적 성능에 미치는 영향을 연구하였다[14]. Dijk의 연구에서는 데이터 증강을 위해 반전, 회전, 이동과 같은 일차 변환과 노이즈 추가, 색상 변화(color jitter), 블러 등을 적용하였다. 그러나 이러한 증강 방식은 최근 이미지 처리를 위한 딥러닝 기술에서 학습 전 이미지 데이터의 통계적인 특성을 정리하기 위해 적용되는 사전 처리 기술(preprocessing)은 적용하지 않고 비교적 단순한 색상 변경 등만 적용하는 한계가 있다. 이러한 한계를 극복하기 위해서 본 연구에서는 다양한 이미지 증강 기법들을 보다 세밀하게 조정하고 조합하여 적용할 수 있는 Albumentations[15]을 적용하였고 통계적 전처리 기법들의 영향을 적용하고 그 영향을 제시한다. 또한 Dijk은 실험 결과에서 가장 좋은 성능을 얻은 증강 기법이 반전과 이동이라고 결론을 내렸는데 실무에서 널리 사용되는 시선 추적 기술(예. SeeSo SDK[16])에서 기본적으로 이미지 반전이 널리 적용되고 있을 뿐만 아니라 사용자의 얼굴 이미지를 사용자의 특징 위치(예를 들어 사용자 이미지의 오른쪽 눈)에서 일정한 거리에 가상 카메라 좌표의 원점을 일치시키고 사용자 정면에서 눈을 바라보도록 정규화(normalization)를 수행하는 방식[17]이 일반적이기 때문에 이미지 이동을 적용하는 데이터 증강 방식은 이미 실용화된 시선 추적 기술에서는 큰 의미를 가지지 못한다. 따라서 본 연구에서는 이미지 반전과 이동을 제외한 데이터 증강이 시선 추적 성능에 어떠한 영향을 주는지 실험을 통해 분석한다.

3) 준지도 학습(SSL) 기법

이미지 증강 기법의 발전과 함께 이러한 증강된 이미지들을 활용하기 위한 여러 방법들도 제안되었는데 준지도 학습 기법(SSL, Semi-Supervised Learning)이 대표적이다. 일반적으로 SSL 기법은 입력 데이터에 증강 등을 통해 어느 정도의 변화가 적용되더라도 출력이 일정하게 유지(consistency regularization)되도록 하는 한편 범주(class)의 분리를 보다 확실하게 보장하기 위해 레이블이 없는 데이터를 활용해서 모델이 예측한 범주의 entropy가 낮은 수준(즉, confidence가 높도록)으로 유지되도록 학습을 수행한다.

대표적인 SSL 방식들을 살펴 보면, ReMixMatch[4]는 분포 정렬(distribution alignment)와 증강 고정(augmentation anchoring)을 통해 출력의 일관성과 레이블이 없는 데이터의 출력 entropy를 낮은 수준으로 유지하며, FixMatch[18]는 이동, 확대 등과 같은 단순한 선형 변환을 적용하는 약한 증강(weakly augmented)과 경계선 필터링 등을 적용하는 강한 증강(strongly augmented)을 활용하여 레이블이 적용되지 않은 데이터를 이용한다. 그러나 이러한 SSL에서 적용하는 여러 증강 기법들이 분류 문제에는 적합하지만 appearance-based 방식의 시선 추적과 같은 회귀 문제에서는 적합하지 않은 측면이 있다. 예를 들어 appearance-based 시선 추적에서는 모바일 기기나 노트북 PC 등에 설치된 전면 카메라를 이용해서 현재 해당 기기를 사용 중인 사용자의 얼굴 이미지를 획득(그림 1)해서 현재 사용자가 기기의 화면 상에서 바라보고 있는 2차원 지점의 좌표를 예측하는 방식이기 때문에 훈련 이미지 데이터를 이동, 회전 등과 같은 약한 증강을 적용하는 경우 기하 정보 자체가 왜곡될 수 있다. 따라서 본 논문에서는 시선 추적 성능 개선이라는 목표의 달성을 위해서 개개의 증강 방식들의 영향을 평가하고자 한다.

대비 학습(contrastive learning) 방식에서는 증강을 활용해서 표현 공간(representation space)에서 같은 범주(class)의 데이터들은 서로 가깝게 배치하고 다른 범주의 데이터는 서로 멀리 배치하는 학습 과정을 거친다. SimpleCLR[6]이 제안된 이후 구체적인 레이블 또는 범주를 지정하기 어려운 SSL(Self-Supervised Learning)의 한계를 극복하기 위한 방식인 지도 학습 대비(Supervised Contrastive) 방식[19]도 제안되었다. 데이터 증강 기법이 주로 이미지를 대상으로 활발하게 연구되어온 영향으로 대비 학습이 주로 이미지 분류에 사용되어 왔으나 SoundCLR[20]처럼 사운드에 대비 학습을 적용하는 사례도 있다. 그러나 SoundCLR 같은 경우 1차원 사운드 데이터에 바로 데이터 증강 기법을 적용하지 않고 사운드 정보를 2차원 이미지(mel-spectrogram)로 변환해서 이 이미지에 이미지 증강 기법을 적용한다는 측면에서 이미지 기반 대비 학습 기법과 유사하다.

II. 본 론

2-1 데이터셋

앞서 논의한 대로 여러 데이터 증강 방식이 딥러닝 기반 시선 추적 기술의 성능에 미치는 영향을 연구하기 위해 일반적으로 널리 사용되는 MPIIFaceGaze 데이터셋[7]을 사용하였다. 본 연구에서는 사용자 1명당 얼굴 이미지 3,000장씩, 총 15명의 사람 이미지 45,000장을 사용하였다. 본 논문에서는 시선 좌표 정보만 사용하며 효율적인 학습을 위해 얼굴 중앙으로 잘라낸 448x448 픽셀 크기의 정규화[17]된 이미지

를 사용한다. 앞서 논의한대로 이 이미지 정규화 기법은 디지털 기기를 사용하고 있는 사용자의 얼굴이 이미지 내에서 다양한 지점에 위치하더라도 일관성 있는 방식으로 얼굴 위치를 조정하기 때문에 유사한 기존 연구[14]에서 가장 높은 효과를 얻은 이동 증강을 무효화시킨다. 따라서 본 연구는 [14]와는 달리 이미 실용화된 최신 시선 추적 기술에 적합한 데이터 증강 기법의 영향들에 초점을 맞추었다.



그림 1. 전처리 기법이 적용된 이미지
Fig. 1. Preprocessed images

1) 이미지 전처리(Preprocessing)

실용화된 시선 추적 기술에서는 이미지 데이터를 사용하여 모델 학습을 하는 경우 이미지 데이터를 모델에 입력하기 전에 조정하고 변환하는 과정을 거치는 것이 일반적이다. 모델에 입력되는 이미지들의 크기를 일정하게 맞추거나, 이미지 픽셀 값들을 정규화하는 전처리 기법들을 통해 계산 효율성을 향상시키고 모델 학습 및 예측을 더 빠르게 진행할 수 있다. 또한 전처리 기법은 모델의 학습에도 훨씬 효과적이다.

히스토그램 평활화(Histogram Equalization, HE)는 이미지의 그레이 스케일(grayscale) 값이 특정 값에 치우쳐 있을 때 균일하게 분산시키는 이미지 처리법이다. 이 방식에서는 영상의 픽셀 값들의 누적분포함수를 이용하여 영상을 개선한다. 히스토그램 평활화를 하면 밀집된 그레이 스케일 값은 분산되고, 분산된 그레이 스케일 값은 밀집되어 결과적으로 이미지의 명암 대비가 증가한다[21].

대비 제한 적응 히스토그램 평활화(Contrast Limited Adaptive Histogram Equalization, CLAHE)는 평활화 과정에서 대비가 과도하게 증폭되는 것을 막기 위한 히스토그램 평활화(HE)를 개선한 방법이다. 대비 제한 적응 히스토그램 평활화는 일정한 크기를 가진 작은 블록(tile)으로 영상을 분할한다. 각 타일에 대해 개별적으로 히스토그램을 계산한다. 그 과정에서 인접한 tile 사이에 양방향선형보간을 사용하여 tile 경계에서 급격한 변화가 일어나는 것을 방지한다 [22]. 그림 1에서는 HE와 CLAHE를 MPIIFaceGaze 내 이미지에 적용한 사례를 제시한다.

2) 이미지 증강(Augmentation) 기법

딥러닝 모델 학습에서 데이터셋이 충분하지 않으면 과적합 발생 가능성이 높아진다. 그러나 충분한 크기의 데이터셋을 수집하기 위해서 많은 비용과 시간이 소요된다. 따라서 이러한 과적합을 막고 비용을 절약하기 위해 이미지 증강을

적용하는 것이 일반적이다. 이미지 증강을 통해 수렴성 개선, 일반화 성능 개선, 분포 외 샘플(outlier)에 대한 안정성을 향상시킬 수 있다[23]. 앞서 논의한 대로 다양한 맥락 하에서 발생하는 복잡한 목표들을 달성하기 위해 여러 증강 기법들이 연구되고 있으나 본 논문에서는 실용적으로 실무 현장에서 널리 적용되는 Albumentations 기법과 TorchVision에서 제공되는 Transform, Light-Enhancer를 대상으로 여러 hyperparameter들을 조정하고 적용했을 때의 시선 추적 성능에 미치는 영향을 살펴 보고자 한다.

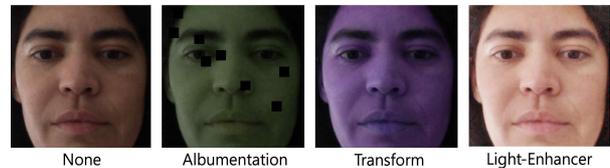


그림 2. 이미지 증강 기법 적용 이미지
Fig. 2. Image of applying augmentation technique

Albumentations은 여러 이미지 증강 기법 중 하나이다 [15]. 본 논문에서는 랜덤한 검정 박스 8개를 생성하는 coarse dropout 옵션[3]과 랜덤한 brightness, contrast, saturation, hue 값을 주는 colorjitter 옵션을 사용하였다. Albumentations 버전은 1.3.1을 사용하였다.

Transform도 여러 이미지 증강 기법 중 하나이다. 주로 기하학적 변화를 줄 때 사용하지만, 본 논문에서는 데이터셋 특성상 기하학적 변화를 주게 되면 시선 좌표에 변화가 생기는 문제가 있었다. 따라서 Transform에서 기하학적 변화를 주는 옵션을 제외한 colorjitter와 특정 확률로 선명도를 변경하는 RandomAdjustSharpness 옵션을 사용하였다. torchvision.transform 버전은 0.15.2를 사용하였다.

본 논문에서 사용한 Light-Enhancer는 이미지의 밝기를 동일한 정도로 향상시킨다. 따로 제공되는 라이브러리를 사용하지 않고 자체적으로 제작한 함수를 사용하였다. 그림 2에서는 Albumentation과 Transform, Light-Enhancer를 MPIIFaceGaze 내 이미지에 적용한 사례를 제시한다.

2-2 실험 과정 및 결과 분석

1) 실험 파이프라인 및 성능 평가 방법

본 연구에서 사용된 실험 파이프라인은 그림 3에서 볼 수 있다. MPIIFaceGaze 데이터셋에 전처리 및 증강 기법을 적용한 후 DenseNet121 모델에 입력된다[24]. 성능은 Angular Error로 평가한다.

전처리 기법은 HE, CLAHE을, 증강기법은 Albumentation, Transform, Light-Enhancer을 사용한다. 전처리 기법을 적용하지 않은 'Original'과 증강 기법을 적용하지 않은 'None'을 포함한 총 12개의 조합을 실험하였다. 각 조합에 대해 6회씩 반복 실험을 수행하였다. Light-Enhancer 의사 코드는

그림 4에서 자세히 설명한다. 이후 모델을 통해 출력된 결과 값은 원본의 시선 벡터 값과 비교하여 Angular Error로 평가된다.

Angular Error는 실제 시선 벡터와 예측 시선 벡터 사이의 거리를 측정하는 값이다. 시선 벡터는 눈동자와 머리의 각도를 고려하여 계산된다[25]. Angular Error의 계산은 실제 시선 벡터와 예측 시선 벡터 사이의 코사인 유사도를 기반으로 한다. 따라서 Angular Error 값이 낮을수록 높은 성능을 낸다고 할 수 있다. Angular Error를 성능 지표로 사용하는 이유는, 3D 및 2D 시선 방향 간의 차이를 직관적이고 명확하게 나타내기 때문이다.

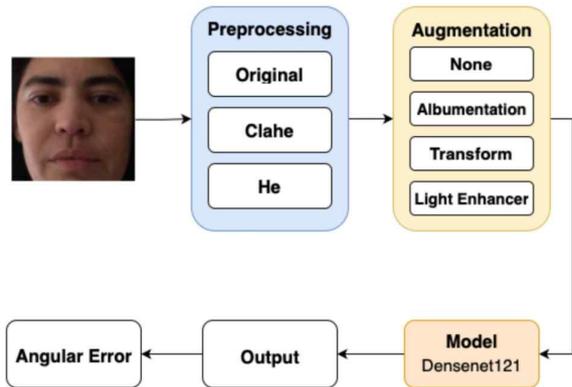


그림 3. 이미지 전처리 및 증강을 통한 DenseNet121 모델 기반 시선 추적 파이프라인

Fig. 3. Pipeline of eye tracking DenseNet121 model using augmentation technique

Algorithm 1 Light Enhancement Algorithm

Require: Low-light image *image*, Scale factor *scale_factor*

Ensure: Enhanced image *enhanced_image*

- 1: Normalize the image by converting to *float32* and scaling by $1/255$
- 2: Convert the normalized image to a PyTorch tensor
- 3: Compute height *h* and width *w* as:
- 4: $h = (image.height \div scale_factor) \times scale_factor$
- 5: $w = (image.width \div scale_factor) \times scale_factor$
- 6: Resize the image to dimensions $h \times w$
- 7: Permute the image dimensions to match the input format of the deep learning model
- 8: Move the image tensor to the computation device (e.g., GPU) and add a batch dimension
- 9: Pass the image tensor through the deep learning model *DCE_net* to obtain:
- 10: $enhanced_image, params_maps = DCE_net(data_lowlight)$
- 11: Remove the batch dimension from the *enhanced_image* tensor
- 12: Convert the *enhanced_image* tensor back to a PIL image
- 13: Convert the PIL image to a NumPy array
- 14: return *enhanced_image*

그림 4. Light-Enhancer 의사 코드

Fig. 4. Light-Enhancer pseudocode

2) 실험 결과 분석

실험 결과는 표 1과 같다. 실험의 특성상 이상치가 발생할 가능성이 높기 때문에 결과 비교를 위해 중앙값을 사용하였다.

표 1. 전처리 기법과 증강 기법을 적용한 시선 추적 모델 실험 결과

Table 1. Experimental results of eye tracking model using augmentation techniques

	None	Albumentation	Transform	Light-Enhancer
Original	7.0153	7.0945	7.1921	6.9366
HE	6.9300	7.3529	7.4963	7.2878
CLAHE	6.8517	7.4668	7.3391	7.1761

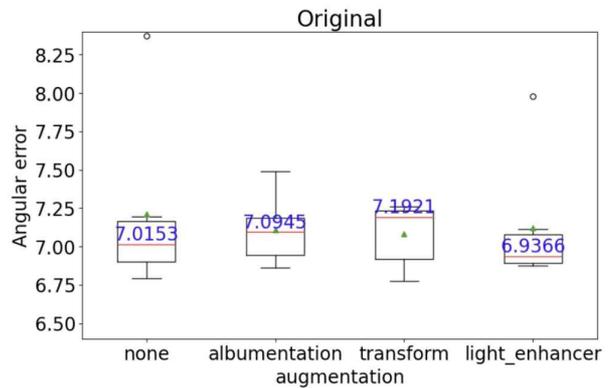


그림 5. Original과 증강 기법 조합의 Angular Error 중앙값을 나타내는 그래프

Fig. 5. Graph of the median angular error for combinations of original and augmentation techniques

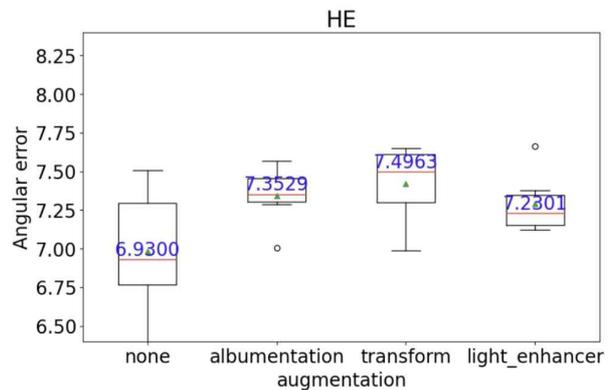


그림 6. HE와 증강 기법 조합의 Angular Error 중앙값을 나타내는 그래프

Fig. 6. Graph of the median angular error for combinations of HE and Augmentation techniques

표 1을 분석하면 가장 효율적인 전처리-증강 조합은 Angular Error가 6.8517인 CLAHE-None이다. 가장 성능이 낮은 조합은 Angular Error가 7.4963인 HE-Transform이다. 종합적으로 분석할 때, 전처리 기법을 사용하면서 증강 기법을 적용하지 않는 경우(HE-None, CLAHE-None)나, 전처리 기법을 사용하지 않고 Light-Enhancer만을 적용하는 경우(None-Light-Enhancer)가 성능 면에서 우수한 것

으로 나타났다. Alumentation과 Transform에서 사용된 Colorjitter 옵션이 시선 추적 학습에 도움이 되지 않았음을 알 수 있다. 이 옵션은 이미지의 색상을 변화시키는데, 이로 인해 사용자의 피부색이 일반적인 범위를 벗어나 성능 저하가 발생한 것으로 추정된다. 또한 HE와 CLAHE를 적용한 후 Light-Enhancer를 사용하면 최대 밝기를 초과하는 현상이 발생하여 성능이 저하되는 것으로 관찰되었다.

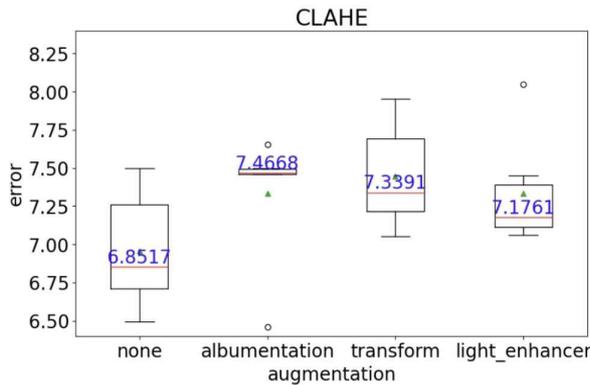


그림 7. CLAHE와 증강 기법 조합의 Angular Error 중앙값을 나타내는 그래프

Fig. 7. Graph of the median angular error for combinations of CLAHE and augmentation techniques

III. 결 론

본 연구에서는 MPIIFaceGaze 데이터셋을 활용하여 시선 추적 모델의 성능 향상을 위한 다양한 전처리 및 이미지 증강 기법들의 조합을 실험하고 그 영향을 분석하였다. CLAHE와 같은 특정 전처리 기법은 Angular Error를 감소시켜 성능을 개선하는데 기여했으며, 이미지 증강 기법 중 Light-Enhancer만을 적용한 경우가 가장 효과적이었다. 이는 얼굴의 명확한 인식과 세부 사항의 강화에 중요한 역할을 했다. 그러나, 특정 색상 변화를 주는 Alumentation과 Transform의 Colorjitter 옵션은 시선 추적 학습에 부정적인 영향을 미쳤다. 이 결과는 이미지의 형상에 따른 분류를 수행하는 일반적인 분류(classification) 문제에서는 다양한 색상 채널 값들의 변화를 통한 데이터 증강이 일반적으로 분류 성능 향상에 영향을 미치지 않지만[18], 시선 추적과 같은 회귀(regression) 문제에서는 성능 향상에 기여하지 않거나 실험 결과와 같이 오히려 부정적인 영향을 미치는 것으로 해석할 수 있다. 다양한 색상을 적용하는 학습 이미지 데이터 증강은 변화하는(variant) 다양한 조명 조건 하 또는 같은 이미지의 다양한 색상(예를 들어, 다양한 색상을 가지는 같은 모델의 자동차) 정보 속에서 불변하는(invariant) 대상의 범주(class)를 정확하게 파악하는 작업 성능을 강화하기 위한 데

이터 증강으로 볼 수 있다. 그러나 시선 추적의 경우, 인간 사용자의 피부색이 인종적으로 약간의 차이가 있다고 하더라도 다른 사물의 경우에 비해 그 차이가 크지 않으며 사용자가 디지털 기기를 사용하는 환경 내 조명이 극단적으로 달라지는 양기 때문에 이러한 결과가 도출되었다고 추정한다.

앞서 논의한 대로 본 연구와 유사한 주제로 Dijk의 연구 [14]가 있다. 본 연구를 Dijk의 연구와 비교했을 때 가장 큰 차이점은 이미지 크롭 외에도 이미지 대비와 밝기 조정과 같은 이미지 전처리 기법을 사용한다는 점이다. 즉, 본 연구는 과거 연구[14]와 달리 이미지 전처리와 데이터 증강 방법을 결합하여 모델 성능을 더욱 향상시킬 수 있는 가능성들도 탐색하였다. 특히, 본 연구에서 제안하는 방식에서는 시선 좌표의 회귀 문제와 같이 좌표 변환이 어려운 문제에서 이미지를 증강시키는 외관 변환(Appearance Transformations)에 초점을 맞춘다.

[14]에서는 이미지에 선형 변환을 적용하여 이동, 회전, 반전 등의 기하학적 변환을 이용해 이미지를 증강시키고 시선 정보 추적에 적용하였으나 이러한 방식은 사전 처리[17]를 통해 이미지의 기하학적인 변동을 사전에 차단하는 최신 시선 추적 기법에서는 큰 의미가 없다. 마지막으로 [14]와의 세부적인 차이로는 본 연구에서는 DenseNet121 모델 하나만을 사용하는 반면, Dijk의 연구에서는 AlexNet과 ResNet18 두 가지 모델을 사용하여 이중 모델 접근 방식으로 데이터 증강 기법의 효과를 평가하였다. 이러한 백본(backbone) 네트워크의 세부적인 차이는 시선 추적 성능 자체에는 영향이 있을지 몰라도, 동일한 백본 네트워크 상에서 데이터 증강의 영향을 비교하고자 하는 목적에는 크게 영향을 미치지 않을 것으로 판단한다.

생성 모델인 GAN(Generative Adversarial Networks)과 Diffusion 모델은 이미지 증강 분야에서도 사용되고 있다. GAN은 적대적 학습을 통해 새로운 이미지를 생성한다. [26], [27]은 GAN을 사용한 증강이 전통적인 증강 기법보다 성능 향상에 효과적이었다는 결과를 보였다. Diffusion 모델은 노이즈 기반의 이미지 변환을 통해 고품질 이미지를 생성한다. [28], [29]은 데이터 증강을 위한 여러가지 Diffusion 모델을 제안한다. 본 논문에서 사용한 이미지 증강 기법 이외에 생성 모델을 사용하여 조명 변화 및 기타 환경적 변수를 반영한 증강 기법을 적용할 수 있어 데이터셋에 다양성을 크게 확장할 수 있다. 향후 연구에서는 이러한 증강 방안을 추가적으로 탐구할 예정이다.

결론적으로, 본 연구는 인공지능 기반 최신 시선 추적 알고리즘에 기초하여 기하학적 변화를 제외시키고 이미지 전처리 및 외관 변환을 통한 데이터 증강의 효과를 검증하였으며 향후 인공지능 시선 추적 기반 기술 개발의 기초 자료로 활용 가능할 것으로 판단한다.

감사의 글

본 연구는 덕성여자대학교 2023년도 교내연구비 지원에 의해 수행되었음

참고문헌

- [1] Z.-Y. Wang and J. Y. Cho, "Older Adults' Response to Color Visibility in Indoor Residential Environment Using Eye-Tracking Technology," *Sensors*, Vol. 22, No. 22, 8766, November 2022. <https://doi.org/10.3390/s22228766>
- [2] K. Krafska, A. Khosla, P. Kellnhöfer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye Tracking for Everyone," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2176-2184, June 2016. <https://doi.org/10.1109/CVPR.2016.239>
- [3] T. DeVries and G. W. Taylor, "Improved Regularization of Convolutional Neural Networks with Cutout," arXiv: arXiv:1708.04552v2, November 2017. <https://doi.org/10.48550/arXiv.1708.04552>
- [4] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel, "ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring," arXiv:1911.09785v2, February 2020. <https://doi.org/10.48550/arXiv.1911.09785>
- [5] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "RandAugment: Practical Automated Data Augmentation with a Reduced Search Space," arXiv:1909.13719v2, November 2019. <https://doi.org/10.48550/arXiv.1909.13719>
- [6] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," arXiv: arXiv:2002.05709, June 2020. <https://doi.org/10.48550/arXiv.2002.05709>
- [7] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 1, pp. 162-175, January 2019. <https://doi.org/10.1109/TPAMI.2017.2778103>
- [8] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges, "ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation," in *Computer Vision – ECCV 2020*, Vol. 12350, pp. 365-381, August 2020. https://doi.org/10.1007/978-3-030-58558-7_22
- [9] S. Park, S. De Mello, P. Molchanov, U. Iqbal, O. Hilliges, and J. Kautz, "Few-Shot Adaptive Gaze Estimation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9367-9376, October 2019. <https://doi.org/10.1109/ICCV.2019.00946>
- [10] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," in *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70, pp. 1126-1135, August 2017. <https://doi.org/10.48550/arXiv.1703.03400>
- [11] S. Hong, Y. Kim, and T. Park, "'Blinks in the Dark': Blink Estimation with Domain Adversarial Training (BEAT) Network," *IEEE Transactions on Consumer Electronics*, pp. 1-1, August 2023. <https://doi.org/10.1109/TCE.2023.275540>
- [12] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, ... and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, Canada, pp. 8026-8037, December 2019.
- [13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, Vol. 15, No. 56, pp. 1929-1958, 2014.
- [14] J. Dijk, Data Augmentation for Deep Learning-based Gaze Estimation, 2023 [Internet]. Available: <https://repository.tu-delft.nl/islandora/object/uuid%3A63ed04d9-8fcf-4df8-9c10-a0d491408a83>.
- [15] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, Vol. 11, No. 2, 125, February 2020. <https://doi.org/10.3390/info11020125>
- [16] VisualCamp. Seeso SDK [Internet]. Available: <https://visualcamp.seeso-sdk/>.
- [17] X. Zhang, Y. Sugano, and A. Bulling, "Revisiting Data Normalization for Appearance-Based Gaze Estimation," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*, Warsaw, Poland, 12, June 2018. <https://doi.org/10.1145/3204493.3204548>
- [18] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, ... and C. Raffel, "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence," arXiv:2001.07685v2, November 2020. <https://doi.org/10.48550/arXiv.2001.07685>
- [19] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, ... and D. Krishnan, "Supervised Contrastive Learning," arXiv:2004.11362v5, March 2021. <https://doi.org/10.48550/arXiv.2004.11362>
- [20] A. Nasiri and J. Hu, "SoundCLR: Contrastive Learning of

Representations for Improved Environmental Sound Classification,” arXiv:2103.01929, March 2021. <https://doi.org/10.48550/arXiv.2103.01929>

- [21] Y.-T. Kim, “Contrast Enhancement Using Brightness Preserving Bi-Histogram Equalization,” *IEEE Transactions on Consumer Electronics*, Vol. 43, No. 1, pp. 1-8, February 1997. <https://doi.org/10.1109/30.580378>
- [22] K. Zuiderveld, Contrast Limited Adaptive Histogram Equalization, in *Graphics Gems IV*, San Diego, CA: Academic Press, ch. 8.5, pp. 474-485, 1994. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
- [23] A. Mikołajczyk and M. Grochowski, “Data Augmentation for Improving Deep Learning in Image Classification Problem,” in *Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Świnouście, Poland, pp. 117-122, May 2018. <https://doi.org/10.1109/IIPHDW.2018.8388338>
- [24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu: HI, pp. 2261-2269, July 2017. <https://doi.org/10.1109/CVPR.2017.243>
- [25] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “It’s Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation,” arXiv:1611.08860v4, May 2023. <https://doi.org/10.48550/arXiv.1611.08860>
- [26] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, ... and D. Rueckert, “GAN Augmentation: Augmenting Training Data Using Generative Adversarial Networks,” arXiv:1810.10863v1, October 2018. <https://doi.org/10.48550/arXiv.1810.10863>
- [27] S. Motamed, P. Rogalla, and F. Khalvati, “Data Augmentation Using Generative Adversarial Networks (GANs) for GAN-Based Detection of Pneumonia and COVID-19 in Chest X-Ray Images,” *Informatics in Medicine Unlocked*, Vol. 27, 100779, 2021. <https://doi.org/10.1016/j.imu.2021.100779>
- [28] B. Trabucco, K. Doherty, M. Gurinas, and R. Salakhutdinov, “Effective Data Augmentation With Diffusion Models,” arXiv:2302.07944v2, May 2023. <https://doi.org/10.48550/arXiv.2302.07944>
- [29] Y. Fu, C. Chen, Y. Qiao, and Y. Yu, “DreamDA: Generative Data Augmentation with Diffusion Models,” arXiv:2403.12803v1, March 2024. <https://doi.org/10.48550/arXiv.2403.12803>



김나영(Nayoung Kim)

2021년~현 재: 덕성여자대학교 사이버보안전공 학사 과정
※ 관심분야 : Data Science, Machine Learning, Deep Learning



인다운(Dawoon In)

2021년~현 재: 덕성여자대학교 사이버보안전공 학사 과정
※ 관심분야 : Machine Learning, Deep Learning



조시은(Sieun Jo)

2021년~현 재: 덕성여자대학교 사이버보안전공 학사 과정
※ 관심분야 : AI Security, Machine Learning, Deep Learning



박태정(Taejung Park)

1997년 : 서울대 전기공학부 (공학사)
1999년 : 서울대 전기공학부 대학원 (공학 석사, 반도체 물리 전공)
2006년 : 서울대 전기컴퓨터공학부 대학원 (공학박사, 컴퓨터 그래픽스 전공)
2006년~2013년: 고려대학교 연구교수
2013년~2017년: 덕성여자대학교 정보미디어대학 디지털미디어학과 조교수
2018년~현 재: 덕성여자대학교 공과대학 사이버보안/IT미디어공학과 부교수
※ 관심분야 : 컴퓨터그래픽스, 인공지능, 수치해석, 3차원 모델링