

시선의 학습 난이도를 고려한 특징 변환 기반 시선 전향 학습 방법

박상진 / 인하대학교 컴퓨터비전 및 영상처리 연구실

본 연구의 목표는 딥러닝 모델을 기반으로 시선 전향 문제를 통해 미세한 시선 표현을 학습하는 것과 적대적 공격에 취약한 신경망을 방어하는 것이다. 먼저, 시선 전향은 얼굴 또는 눈 영상의 시선을 원하는 방향으로 조정하는 task이며 신경망 기반 생성기를 통해 얼굴 영상의 시선 방향을 학습하는 것을 목표로 한다. 본 연구는 잠재 특징 공간 상에서 다양한 시선 방향에 상응하는 시선 특징을 제공하는 특징 단계 변환과 시선 특징의 차별 학습을 위한 새로운 손실 함수를 제안한다. 제안된 방법은 GazeCapture 데이터셋에서 2도의 전향 오차를 보여 주었다. 이 수치는 최신 기법인 STED보다 10% 우수한 성능이다. 게다가, 다양한 특성들의 잠재 특징이 차별화되어야 하는 이유가 활성화 맵 시각화를 통해 제시된다.

고전적인 시선 전향 기법들은 일반적인 생성 모델 및 영상 변환 기법을 통해 시선 전향을 수행하였다. 최근, 잠재 특징과 입력 영상 간 등변 매핑을 학습하는 변환 오토인코더가 시선 전향 기법에 사용되어 신뢰할 만한 성능을 보여주었다. 이 기법들은 원천 영상의 시선 방향을 목표 영상의 시선 방향으로 변환하는 오토인코딩 과정을 학습한다. 하지만, 기존 기법들은 제한된 수의 시선 방향

을 갖는 영상만을 활용했기 때문에 실제 환경에서 시선 방향을 정확하게 학습할 수 없었다. 시선 방향은 연속적인 정보이기 때문에, 제한된 영상으로 학습된 모델은 학습에 사용되지 않은 시선 방향을 추론할 수 없다. 또한, 잠재 특징들의 특징 공간에서 얽힘은 시선만을 표현하는 특징을 학습을 제한한다. 만약, 학습된 시선 특징에 다른 정보가 포함되어 있다면 이 특징을 활용한 하류 태스크의 성능을 저하시키는 결과를 초래한다.

본 논문은 더 강인한 시선의 표현을 학습하기 위해 다음 두 가지 문제가 해결되어야 한다고 주장한다. 먼저, 입력 영상에 의해 표현될 수 없는 시선 방향이 (잠재) 특징 공간에 적절히 반영되어야 한다. 선행 연구들은 입력 영상 즉, 원천과 목표 영상의 한정된 시선 방향만을 학습 동안의 감독으로 사용하였기 때문에 학습 도중 보지 못한 시선 방향의 표현을 학습하기 어려웠다. 두 번째로, 시선은 머리 자세와 외모와 같이 시선과 관련이 적거나 거의 없는 몇 가지 사람에 관련된 요인들과 강하게 얽혀있다. 따라서, 시선, 머리 자세와 외모 등이 특징 공간에 얽혀 있다면 시선을 완전히 표현할 수 있는 특징을 학습하는 것은 매우 어려울 것이다. 시선 특징과 머리 자세 특징과

졸업논문 소개

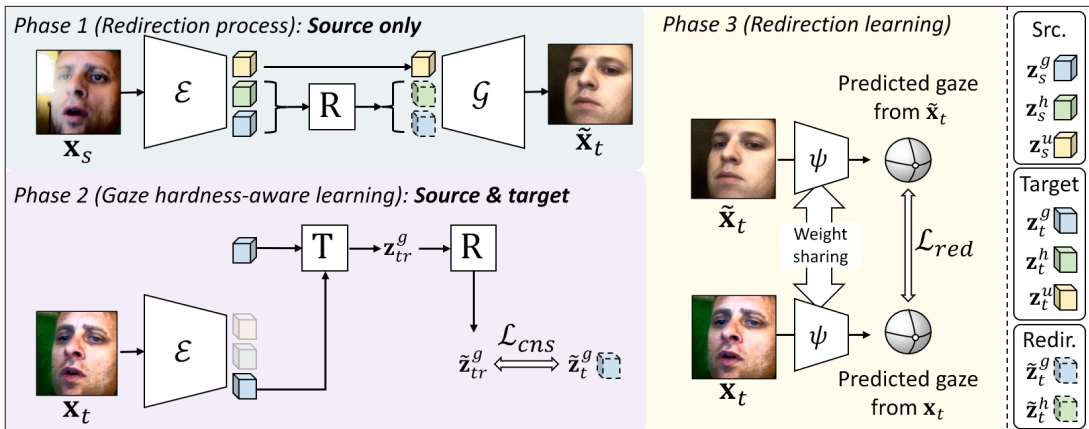
같은 중요하지 않은 특징의 차별화 학습, 즉 특징 간 관계의 학습은 아직까지 시도된 바 없다.

본 논문은 상기 언급한 문제들을 모두 다루는 시선 이해의 새로운 개념을 제시한다. 먼저, 본 연구는 원천과 목표 영상 쌍으로부터 다양한 시선 특징을 생성하기 위한 Gaze Hardness-aware Transformation (GHT)를 제안한다. GHT와 본 연구의 시선 전향 과정은 <그림 1>과 같다. GHT는 원천과 목표 시선 특징 간 선형 보간으로 정의된다. 변환된 특징은 원천과 목표 영상 만으로 표현될 수 없는 시선 방향의 수를 증대시키는 추가적인 감독의 역할을 하며 제안하는 시선 일관성 손실 함수의 입력이다. 추가적으로, GHT는 시선 표현 학습의 어려움 정도를 증대시키도록 설계되었기 때문에, 학습 후반에서 과적합 문제를 완화할 수 있다. 두 번째로, 본 논문은 시선 특징과 중요하지 않은 특징 간 차별화 학습을 위한 소위 structured gaze (SG) 손실 함수를 제안한다. 본 연구는 시선과 중요하지 않은 특징들을 부정 쌍으로 다른 시선 특징들은 긍정 쌍으로 정의한다. 이렇게 정의된 쌍들에 기반한 SG 손실함수는 미니 배치 내에 긍정과 부정 샘플 간 다양한 조합을 통해 구조화된 특징 유사도를 계산한다. 여기서, 메트릭 학습의 내재적인 과적합 문제를 완화하기 위해, 기존

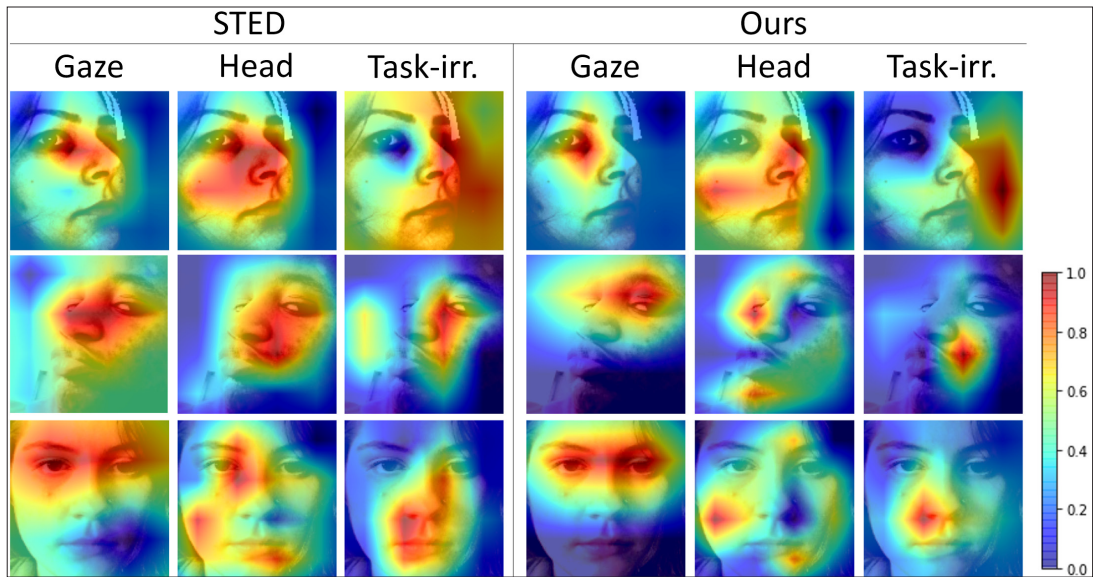
메트릭 학습 연구에서 사용한 어려운 부정과 긍정 샘플들이 추가적으로 활용된다. 그러므로, SG 손실 함수는 소위 ‘밀고 당기기’ 전략에 기반한 특징 내 관계를 학습한다.

본 논문의 효과를 증명하기 위해 대표적인 시선 및 머리 자세 데이터셋인 GazeCapture에 대해 성능을 평가하였다. 제안 기법은 기존 시선 전향 기법들에 비해 높은 성능을 달성하였으며 모든 평가 메트릭에서 평균 10% 앞선 성능을 보인다. 예를 들어, 제안 기법은 1,884의 시선 전향 오차를 달성하였으며 이는 STED 대비 14.2% 향상된 수치이다. 이 결과는 잠재 특징들의 consistency와 disentanglement 성질이 TA의 auto-encoding 학습에 중요하다는 것을 시사한다.

본 연구는 제안하는 차별화 학습의 효과를 입증하기 위해 최신 시각화 기술인 ContraCAM을 사용하였다. 활성화 맵을 계산하기 위한 연속적인 형태의 GT를 활용하는 ContraCAM은 클래스 확률 점수가 아닌 연속적인 시선 또는 머리 자세를 GT로 활용하는 제안 기법에 더 적합하다. <그림 2>는 제안 기법과 STED의 특징 맵을 시각화한다. STED의 시선 특징은 시선과 관련이 적은 눈이 아닌 영역에 더 집중하고 있다. 반면, 제안 기법의 시선 특징은 오직 눈 영역에만 초점을 맞추며, 중요하지 않은 특징들이 시선



<그림 1> 제안 기법의 알고리즘



<그림 2> 제안 기법과 비교 기법의 ContraCAM 시각화 결과

특징과 독립적인 영역을 가리킨다.

본 연구는 GHT를 통해 다양한 시선 방향을 포함한 시선 특징을 증강 및 조작하는 데 성공하였다. 생성된 시선 특징은 추가 감독 역할을 하여 시선 전향의 일반화 성능을 향상시킬 수 있다. 앞으로 GHT는 많은 어노테이션 비용이 필요한 시선 표현 학습에서 다양한 용도로 사용될 수 있다. 또한 특징의 차별 학습을 위한 구조적 시선 손실

함수는 얼굴 감정이나 제스처 인식과 같은 다른 컴퓨터 비전 태스크로 확장될 수 있다.



박상진

- 2021년 : 인하대학교 전자공학과 학사
- 2023년 : 인하대학교 전기컴퓨터공학과 석사
- 주관심분야 : 딥러닝, 시선 추정, 생성 모델