

딕셔너리 증류 기법을 적용한 얼굴 초해상화

조병호, 박인규, 홍성은

인하대학교 전기컴퓨터공학과

{byunghojo12@gmail.com, pik@inha.ac.kr, csehong@inha.ac.kr}

Dictionary Distillation in Face Super-Resolution

Byungho Jo, In Kyu Park, Sungeun Hong,

Department of Electrical and Computer Engineering, Inha University

요 약

본 논문에서는 지식 증류 (knowledge distillation) 기법을 적용한 얼굴 초해상화 모델을 제안한다. 제안하는 기법은 최근 얼굴 복원 분야에서 좋은 성능을 보여준 얼굴 영역의 딕셔너리 (dictionary) 정보를 사용한 모델을 선생 모델로 선정하여 적대적 (adversarial) 지식 증류 기법을 통해 효율적인 학생 모델을 구축하였다. 본 논문은 테스트시 얼굴의 사전 정보가 초래하는 추가적인 비용이 필요 없는 얼굴 초해상화 방법을 제시하고, 제안하는 기법과 다양한 기존 초해상화 기법과의 정량적, 정성적 비교를 통해 우수성을 보인다.

1. 서론

얼굴 초해상화 (face super-resolution)의 목적은 저해상도의 얼굴 영상을 고해상도의 얼굴 영상으로 복원하는 것으로서 얼굴 영상 분석에 있어 중요한 전처리 기법으로 사용 된다. 최근 다양한 CNN Block 구조 [1, 2] 및 얼굴에 특화된 사전 정보 [3]를 활용한 초해상화 연구가 제안되었으며 우수한 복원 성능을 보여 주고 있다. 그러나 깊은 네트워크, 얼굴 랜드마크 정보 및 참조 얼굴 영상과 같은 사전 정보를 사용할 경우 추가적인 메모리 및 시간을 소모하는 단점이 존재한다. 최근 지식 증류 기법을 통해 랜드마크 정보를 추출하는 깊은 모델로부터 얇은 모델을 구축하여 이를 사용하는 초해상화를 수행하는 연구 [4]가 제안되었다. 그러나 부가적인 메모리를 요구하는 단점은 그대로 존재한다.

본 논문에서는 적대적 지식 증류 기법을 사용하여 기존에 존재하는 단점을 극복할 수 있는 효율적인 얼굴 초해상화 모델을 제안한다. 얼굴 영역의 딕셔너리 [3]를 사용하는 선생 모델로부터 지식 증류를 통해 학생 모델이 사전 정보 없이 초해상화를 수행하도록 구축하였다. 본 연구는 32 x 32 해상도의 입력 얼굴 영상을 256 x 256 해상도의 영상으로 복원하도록 실험을 하였다.

2. 제안하는 기법

제안하는 기법은 DFDNet [3]에서 제안한 얼굴 랜드마크 정보를 바탕으로 구축된 눈, 코, 입 영역의 딕셔너리 정보를 담고 있는 선생 모델의 특징 맵 (feature maps)과 학생 모델의 특징 맵이 가지는 분포의 차이를 줄이도록 적대적 지식 증류 기법을 사용하여 학습을 진행하였다.

2.1 특징 맵 기반의 적대적 지식 증류 기법

본 연구는 최근 제안된 적대적 지식 증류 기법 [5]을 참고하여 오프라인 지식 증류 학습 방식으로 진행되었다. 사전 훈련이 완료된 선생 모델로부터 말단 층의 특징 맵을 추출하여 구분자 (discriminator)에게 실제 데이터로서 입력이 되며 학생 모델의 특징 맵은 구분자에게 가짜 데이터로 입력되어 적대적 학습이 진행된다. 이를 통해 학생 모델은 특징 맵 기반의 적대적 지식 증류 기법을 통해 얼굴 영역 기반의 딕셔너리 정보를 담고 있는 선생 모델이 가지는 특징 맵의 분포를 학습하도록 수행된다. 선생과 학생 모델은 동일한 모델 구조를 가지며 encoder-decoder 구조를 가진다. 구분자의 경우 간단한 CNN 모델 구조를 가진다.



그림 1. VGGFace2 테스트 영상에 대한 정성적 비교

2.2 손실 함수

제안하는 학생 모델 M_S 와 구분자 D 에 대한 목적 함수 L_S , L_D 는 다음과 같다.

$$\min L_S = E_{SR,GT} [\|SR - GT\|_2] + E_{F_S, F_T} [(D(F_S) - 1)^2] + E_{SR,GT} [\sum_i \|f_i(SR) - f_i(GT)\|_2], \quad (1)$$

여기서 L_S 의 각 손실 함수는 학생 모델의 초해상화 결과 SR 과 정답 영상 GT 영상과의 픽셀 기반의 L2 거리, adversarial loss와 perceptual loss를 나타낸다. f_i 는 ImageNet 데이터셋에 학습된 VGG-19 모델의 i 번째의 층에서 추출된 특징 맵을 나타낸다.

$$\min L_D = E_{F_T, Real} [D(F_T - 1)^2] + E_{F_S, Fake} [D(F_S)^2], \quad (2)$$

L_D 는 선생 모델의 특징 맵 F_T 와 학생 모델의 특징 맵 F_S 를 입력으로 받아 진위 여부를 판별하는 LSGAN [6]을 사용한다.

3. 실험 결과

본 논문에서는 VGGFace2 [7] 데이터셋을 훈련 및 테스트에 사용한다. 학생 모델의 성능 측정을 위하여 선생 모델, 지식 증류 기법을 적용하지 않은 학생 모델 (S w/o KD)과 초해상화 분야에 우수한 성능을 보인 RCAN [1]과 SRFBN [2]을 사용하여 정성적 및 정량적 비교를 수행한다.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
RCAN	26.59	0.779	0.3191
SRFBN	26.399	0.761	0.3944
Teacher	26.74	0.771	0.2413
S w/o KD	25.837	0.75	0.2746
Student	26.623	0.769	0.2511

표 1. VGGFace2 테스트 영상에 대한 정량적 평가 결과

표 1을 통해 제안하는 학생 모델이 픽셀 단위의 평가 방식인 PSNR, SSIM과 perceptual 평가 방식인 LPIPS에서 선생 모델과 근접한 성능을 보이며 다른 초해상화 모델에 비해 우수한 성능을 보인다. 그림 1을 통해 학생 모델이 정성적으로 선생 모델과 비슷한 성능을 보이고 사실적인 얼굴 영상 복원 결과를 보인다.

4. 결론

본 논문에서는 특징 맵 기반의 지식 증류를 통한 효율적인 얼굴 초해상화 모델을 제안하였다. 실험 결과를 통해 사전 정보를 요구하는 기존의 단점을 극복하고 정량적 및 정성적으로 우수한 복원 성능을 보임을 확인할 수 있었다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2020-0-01389, 인공지능융합연구센터지원(인하대학교)). 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019R1A2C1006706).

참고문헌

- [1] Y. Zhang, et al., "Image super-resolution using very deep residual channel attention networks," In *Proc. of European Conference on Computer Vision*, pp. 294-310, 2018.
- [2] Z. Li, et al., "Feedback network for image super-resolution," In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3867-3876, 2019.
- [3] X. Li, et al., "Blind face restoration via deep multi-scale component dictionaries," In *Proc. of European Conference on Computer Vision*, pp. 399-415, 2020.
- [4] D. Kim, M. Kim, G. Kwon, and D. Kim, "Progressive face super-resolution via attention to facial landmark," In *Proc. of British Machine Vision Conference*, pp. 192, 2019.
- [5] I. Chung, S. Park, J. Kim, and N. Kwak, "Feature-map-level online adversarial knowledge distillation," In *Proc. of International Conference on Machine Learning*, pp. 2006-2015, 2020.
- [6] X. Mao, et al., "Least squares generative adversarial networks," In *Proc. of IEEE International Conference of Computer Vision*, pp. 2813-2821, 2017.
- [7] Q. Cao, et al., "VGGFace2: A dataset for recognizing faces across pose and age," In *Proc. of IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 67-74, 2017.