털 합성 데이터 증강을 통한 피부영상 털 제거 심층학습 기법: 기존 방법의 적용성 평가

이영찬[1], 박범진[1], 최재영[2], 이범식[3], 유원상[O,1]

[1]선문대학교 공과대학 정보통신공학과

[2]한국외국어대학교 컴퓨터공학부, [3]조선대학교 IT 융합대학 정보통신공학부

lyc19960618@sunmoon.ac.kr, parkbj24@sunmoon.ac.kr,

jychoi@hufs.ac.kr, bslee@chosun.ac.kr, wyou@sunmoon.ac.kr

요 약

피부영상에서 털 제거는 자동화된 피부진단을 위한 필수적인 전처리 과정이다. 기존의 털 제거 기법은 대부분 영상처리에 기반한 방식이었지만, 털이 많은 피부에는 성능이 크게 저하되는 문제점이 있다. 다른 접근방식으로서 심층학습 모델에 기반한 인페인팅 기법을 고려할 수 있는데, 털이 많은 피부영상 학습 데이터의 부족으로 인하여 심층학습 기반 털 제거 모델은 아직 보고된 바 없다. 본 논문에서는 털이 많은 피부 합성영상을 생성함으로써 털 제거 심층학습 모델의 훈련을 위한 학습 데이터를 증강하는 기법을 제안한다. 제안된 데이터증강 기법은 Poisson image editing (PIE)을 사용하여 털이 있는 피부영상으로부터 털 영역을 분할하여 털이 없는 피부영상에 혼합하는 방식이다. 털 제거를 위한 인페인팅 심층학습 모델로서 Shift-Net 과 DeepFill 을 개선하여 PIE 기반 합성영상 증강 데이터를 사용할 수 있도록 하였다. 합성영상에 대한 털 제거 성능을 평가한 결과, 털 제거 이미지의 구조적 유사도(SSIM) 지수와 실감성에 대한 평균의견점수(MOS)에서 Shift-Net 이 DeepFill 및 기존 영상처리 기법인 Fast marching method 보다 높았다(SSIM=0.924±0.029, MOS=3.73±1.02) 실험결과는 제안된 털 합성 데이터 증강 기법이 피부영상 털 제거 심층학습의 성능을 향상시키는데 사용될 수 있는 가능성을 보여준다.

1. 서론

최근 괄목할 만한 성장을 보인 인공지능 기반 피부영상 진단 기술에서, 털 제거는 피부병변의 인식과 분류작업을 위한 필수적인 전처리 과정이다[1]. 피부영상 진단기술은 주로 털이 많지 않은 사람의 피부를 대상으로 발전되어 왔기 때문에, 털 제거는 영상처리에 의한 방법만으로도 유효한 성능을 보여왔다.

그러나 기존의 영상처리 기반 방식은 털이 많이 덮여 있는 피부에서는 털 제거 성능이 저하되는 문제점이 있었다. 이러한 털 제거의 기술적 어려움은 특히 동물 피부영상 분석에서 가장 큰 장애요인으로 작용한다.

이러한 문제를 극복하기 위해 심층학습 모델에 기반한 영상 인페인팅 기법을 고려할 수 있다. 털 제거 심층학습 모델을 훈련하기 위해서는 털이 많은 피부영상과 털 영역분할 정답 마스크 등으로 구성된 대량의 학습 데이터가 필요한데, 이러한 데이터셋은 확보가 어렵다. 이러한 이유로 심층학습 기반 털 제거 모델은 아직 보고된 바 없다[1].

본 연구에서는, 털이 많은 피부 합성영상을 생성함으로써 털 제거 심층학습 모델의 훈련을 위한 학습 데이터를 증강하는 기법을 제안한다. 제안된 데이터증강 기법은 털이 많은 피부영상으로부터 분할된 털 영역을 Poisson image editing (PIE)을 사용하여 다른 피부에 혼합하는 방식이다[2]. 피부영상으로부터 털 제거를 위한 인페인팅 심층학습 모델로는 Shift-Net 과 DeepFill 을 사용하였다[3,4]. 기존 모델과 달리, 본 논문에서는 PIE 기반 합성영상 증강 데이터에 포함된 털 영역 마스크를 사용하여 훈련할 수 있도록 하였다.

2. 관련 연구

피부영상으로부터 털을 제거하는 문제는 영상 인페인팅(image inpainting)의 문제로 이해할 수 있다. 영상 인페인팅은 입력 영상의 훼손된 영역을 자연스럽게 채워 영상을 복원하는 기법으로서, 영상처리에 의한 접근방법과 심층학습에 의한 접근방법으

로 구분할 수 있다[5].

피부영상에서 털 제거를 위한 기존 인페인팅 방법은 DullRazor, E-shaver 를 비롯하여 Fast marching method (FMM), Block-based algorithm, PDE-based method, linear interpolation method, threshold-set model 등 주로 영상처리에 의한 기법이 적용되어 왔다[6-16]. 특히, FMM 기법은 손실된 화소영역의 경계로부터 영역 내부를 점진적으로 채우는 방식이다[17]. 그러나, 기존 영상처리 방법은 털이 적은 사람 피부영상 데이터에 주로 적용되어 온 반면, 털이 많은 피부영상에서는 성능이 입증된 바 없다. 본 논문에서 제안된 방법은 털이 많은 피부영상에서 털을 제거하는 것을 목적으로 한다.

최근 영상 인페인팅을 위한 다양한 심층학습 모델이 발전되어 왔는데, Shift-Net, DeepFill, GMCNN, PartialConv, LBAM 등을 포함한다[3,4,18-20]. 이러한 모델은 대부분 원본 이미지에 마스크를 내부에서 무작위로 생성하여 복원하는 방식으로 훈련이 진행된다. 그러나 이와 같이 임의 형태의 마스크를 사용한 방식은 매우 가늘고 긴 형태의 털 영역 뒤에 가려진 피부를 복원하는데 적합하지 않을 수 있다. 본 논문에서 제안된 방법은 기존 모델과는 달리 입력 피부영상에 대응하는 털 영역 마스크를 외부에서 삽입하여 모델을 훈련하는 방식이다.

이를 구현하기 위해서는, 털이 없는 원본 피부영상과 이에 대응하는 털 영역 마스크의 쌍으로 구성된 학습 데이터셋이 필요하다. 이러한 데이터셋은 구축하는데 어려움이 있으므로, 대안으로 영상합성을 통한 데이터 증강 기법을 고려할 수 있다. 관련 연구로서, 흉부 X-ray 합성영상을 사용하여 훈련된 모델이 실제 X-ray 영상을 사용하여 훈련된 모델보다 성능이 20% 향상되었다는 연구결과가 최근 보고되었다[21]. 최근에는 생성적 적대 신경망(GAN)을 사용하여 피부질환 합성영상을 생성하는 기법이 제안되었다[22]. 기존 영상합성 기반 데이터 증강 기법이 피부병변의 합성영상을 이미지 전체로 생성하는 방식인 반면, 제안된 방법은 털이 많은 피부 합성영상을 생성하는 방법으로서, 털이 있는 피부영상으로부터 털 영역을 분리한 후 여러 털 영역을 타겟 피부영상에 혼합하는 방식이라는 점에서 독창성이 있다.

3. 방법

3.1 털 합성 기반 데이터 증강

본 논문에서 제안된 피부영상 털 제거 심층학습 모델의 훈련을 위한 데이터 증강 기법은 그림 1 에 요약되어 있다. 제안된 기법은 (1) 털이 있는 피부영상 데이터셋에서 털 영역을 분할하는 털 영역분할 단계와 (2) 분할된 털 영역을 털이 없는 피부영상에 혼합하는 털-피부혼합 단계로 구성된다.

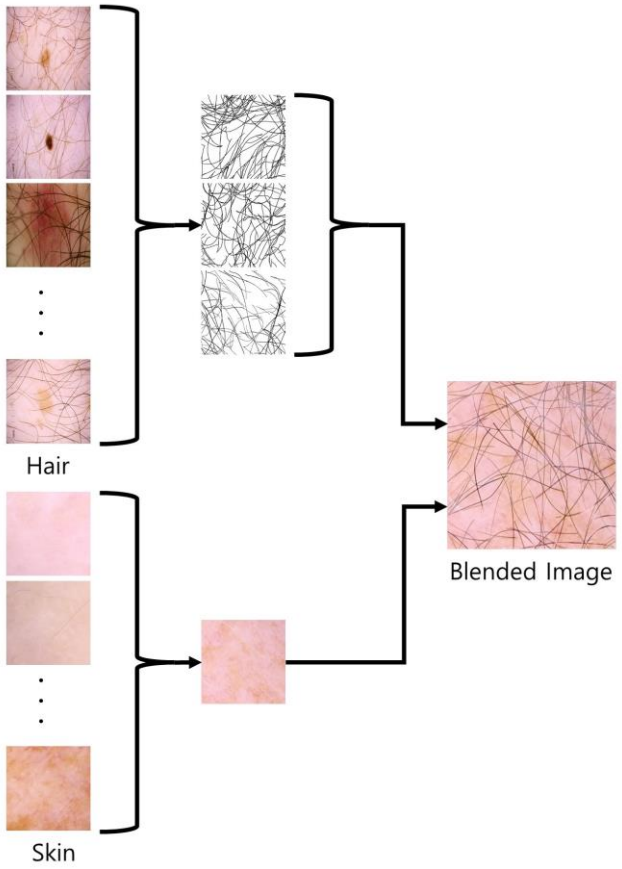

그림 1 털 합성영상 생성 과정

털 영역분할 단계에서는 피부영상 데이터셋으로부터 털 영역 정답 마스크를 생성하였다. 털-피부혼합 단계에서는 털이 있는 피부영상 데이터셋에서 여러 피부영상을 무작위로 선택한 후 해당 털 영역을 털이 없는 피부영상에 순차적으로 혼합하였다. 영상혼합 기법으로서 PIE 기법을 사용하였다[2]. PIE 기법은 그림 2 와 같이 배경영상에 객체 영상을 혼합하는 기법으로 픽셀 변화도(gradient)에 기반하여 배경과 객체가 자연스럽게 혼합되도록 한다.

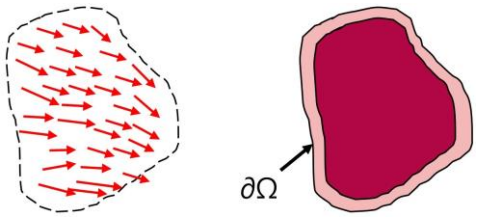

그림 2 Poisson image editing 영상혼합 기법

3.2 피부영상 복원

피부영상에서 털 영역을 제거하고 피부를 복원하기 위한 심층학습 인페인팅 모델로서 Shift-Net 과 DeepFill v2 를 사용하였다. 그림 3 과 같이, 두 모델

은 털 영역 마스크를 참조하여 털이 있는 피부영상으로부터 털을 제거한 이미지를 출력한다. 털 영역 마스크를 생성하기 위한 털 영역분할을 위해 형태학적 연산 기법이 사용되었다. Black hat 필터링을 통해 털 영역의 특징을 추출하고, connected component labeling 을 통하여 잡음 영역을 제거하였다[23,24].

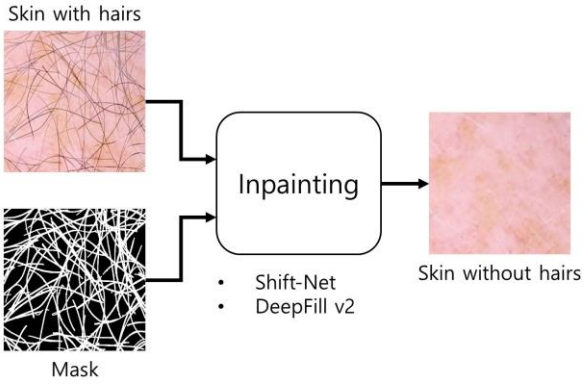

그림 3 심층학습 기반 털 제거 기법

Shift-Net 은 그림 4 와 같이 U-net 구조에 shift 연결 레이어를 더한 것으로, 알려진 영역의 인코더 특성과 훼손된 영역의 디코더 특성을 연결하여 훼손된 영역의 전체적 형상 구조와 로컬 텍스쳐의 세부 내용을 모두 개선할 수 있도록 한다[3].

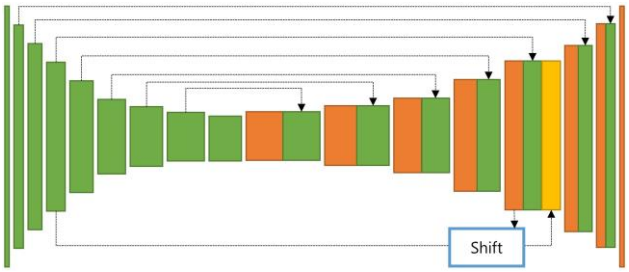

그림 4 Shift-Net 구조

DeepFill v2 는 그림 5 와 같이 Contextual attention 레이어를 사용하여 훼손된 영역의 주변 이미지 특성을 참조하여 인페인팅 네트워크를 훈련한다[4]. 네트워크는 두개의 단계로 구성되는데, 첫번째 단계는 훼손된 영역을 대략적으로 복원하기 위한 합성곱 네트워크이며, 두번째 단계는 정밀하게 복원하기 위한 네트워크이다. 특히, 자유 형태의 이미지 인페인팅을 위해 출력 이미지의 각 점에서 손실함수를 계산하도록 하는 SN-PatchGAN 이라 불리는 패치 기반 GAN 프레임워크에서 훈련된다[25].

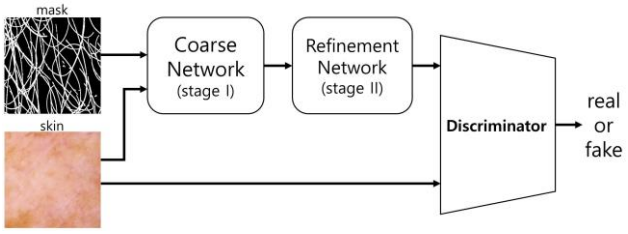

그림 5 DeepFill 구조

영상 인페인팅을 위한 두 심층학습 모델은 훈련 과정에서 내부에서 입력영상에 전경 마스크(foreground mask)를 무작위로 생성하여 훈련한다. 본 연구에서는 그림 6 과 같이 데이터증강 기법을 통하여 생성된 털 영역 마스크를 외부에서 삽입하여 훈련하도록 Shift-Net 과 DeepFill v2 모델을 개선하였다.

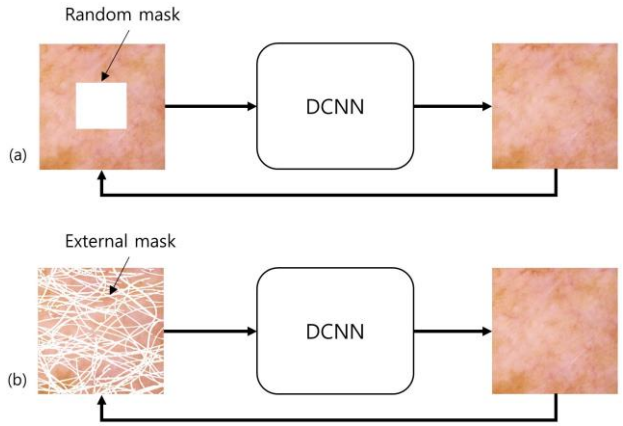

그림 6 (a) 무작위 마스크 생성에 기반한 기존 영상 인페인팅 모델, (b) 증강된 털 영역 마스크에 기반한 영상 인페인팅 모델

4. 실험 결과

털이 있는 피부 합성영상을 생성하는 단계에서, ISIC 2020 Dataset 에서 털이 있는 피부영상 48 개를 선택하여 각 이미지로부터 털 영역의 정답 마스크를 생성하였다. 또한, 털이 없는 피부영상 5 개를 선택하여 털 영역을 혼합하기 위한 타겟 피부 이미지로 사용하였다.

무작위로 털이 있는 피부영상 3 개의 이미지를 선택한 후, 무작위로 선택된 털이 없는 피부영상에 순차적으로 영상혼합기법(PIE 또는 DIB)을 사용하여 털이 많은 피부 합성영상을 생성하였다. 털 혼합 과정에서, 털이 있는 피부영상에 0, 90, 180, 270 도 중에서 회전각을 무작위로 적용하여 털 패턴의 다양성을 향상시켰다. 또한 털이 없는 피부영상에는 Color jitter 를 통해 lightness, hue, saturation 을 조절하여 피부의 색상을 다양하게 변형하였다.

위와 같은 방법으로 털이 많은 피부 합성영상 3,000 개를 생성하였고, train/validation 및 test 를 위해 각각 2,500 개와 500 개로 구분하였다. 그림 7 은 실제 사람 피부영상 보다 많은 양의 털을 가지고 있는 합성된 피부영상의 예시를 보여준다.

털 합성영상에서 털을 제거하는 단계에서, Shift-Net 모델의 훈련을 위한 일괄처리량(batch size)은 1, 학습율(learning rate)은 0.0002, 반복횟수(epoch)는 100 으로 각각 설정하였다. 또한, DeepFill v2 모델의 훈련을 위한 일괄처리량은 1, 학습율은 0.001, 반복횟수는 100 으로 각각 설정하였다. 훈련 결과, Shift-Net 과 DeepFill v2 의 최종 손실함수는 각각 1.415 와 -0.471 로 측정되었다.

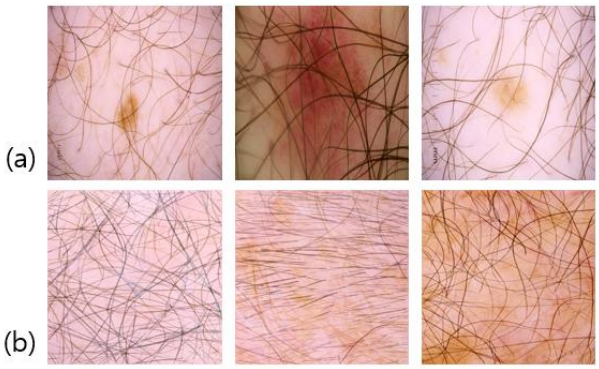

그림 7 (a) 실제 사람피부 이미지, (b) 털-피부 혼합 기법을 통해 합성된 피부 이미지

털 제거 성능 평가를 위해 Shift-Net 과 DeepFill v2 모델을 기존 FMM 기반 영상처리 기법과 비교하였다. 그림 8 은 훈련된 두 심층학습 모델과 FMM 을 사용하여 털 합성영상으로부터 털을 제거한 예시를 보여준다.

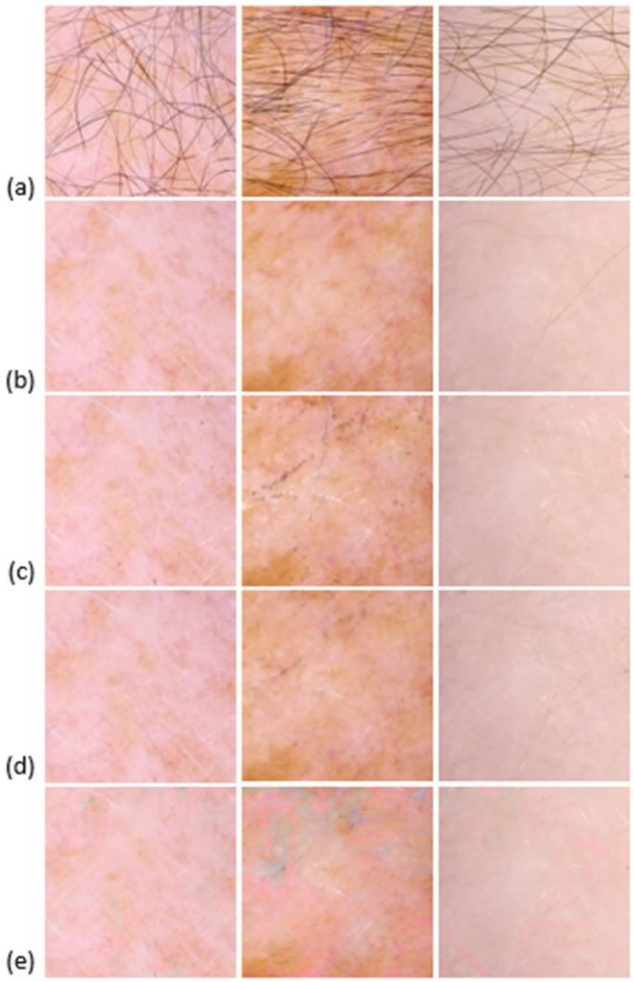

그림 8 (a) 털 합성 영상, (b) 정답 영상, (c) FMM 기반 영상처리 기법, (d) Shift-Net, (e) DeepFill v2

그림 9 는 털이 많은 실제 사람 피부영상으로부터 털을 제거한 예시를 보여준다.

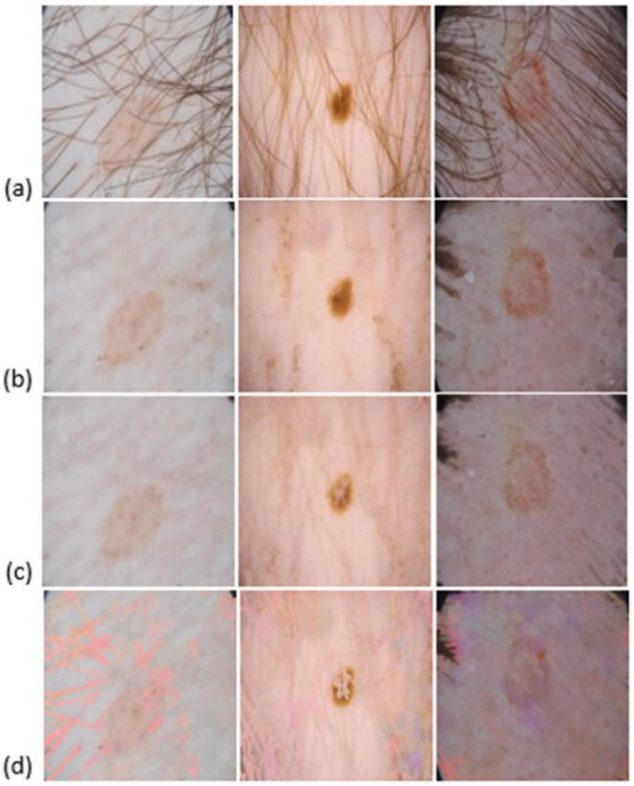

그림 9 (a) 털 합성 영상, (b) FMM 기반 영상처리 기법, (c) Shift-Net, (d) DeepFill v2

털 제거에 대한 정량적 평가결과는 표 1 에 요약되어 있다. 털 합성영상에 대하여, 구조적 유사도 지수(SSIM)는 Shift-Net 이 FMM 과 DeepFill 에 비하여 가장 높은 값을 갖는 반면, PSNR 은 FMM 이 가장 높은 값을 보였다. 대응표본 t-검정 (paired sample t-test) 결과 p-value 는 모두 0.05 의 유의수준보다 작으므로 위 결과는 통계적으로 유의미한 차이를 가진다. 이와 같은 결과는 Shift-Net 이 피부의 구조정보는 잘 보존하지만 blurring 효과로 인하여 피부의 미세한 텍스쳐 정보를 잃기 때문인 것으로 분석된다.

표 1 합성영상의 털 제거에 대한 정량적 평가

	SSIM	*PSNR*
FMM	0.907 ± 0.047	**38.8 ± 2.6**
Shift-Net	**0.924 ± 0.029**	36.3 ± 1.6
DeepFill	0.859 ± 0.056	34.9 ± 1.9

털 제거된 결과 이미지의 실감성에 대한 정성적 평가를 위해, 20 대 한국인 26 명을 대상으로 합성 피부영상 5 장과 실제 사람 피부영상 5 장으로부터 털 제거된 결과 이미지 샘플에 대하여 1~5 사이의 실감성 등급을 매기게 하여 평균의견점수(mean opinion score, MOS)를 계산하였다 [26]. 털 제거에 대한 정성적 평가결과는 표 2 에 요약되어 있다. 털 합성영상으로부터 털 제거 결과 이미지의 실감성 지표는 Shift-Net 이 FMM 과 DeepFill 과 비교하여 가장 높았다(MOS=3.73± 1.02, $p<0.05$). 반면, 실제 사

람 피부영상으로부터 털 제거 결과 이미지의 실감성 지표는 FMM 이 가장 높았고 통계적으로 유의미한 차이를 보였다(MOS=3.72± 1.19, $p<0.05$). 이 결과는 Shift-Net 이 털 합성영상에 대해서는 털 제거의 효과가 가장 높지만, 실제 피부영상에서는 털 제거의 효과가 기존 영상처리 기법에 비해서 크지 않다는 점을 지적한다.

표 2 털 제거 피부영상에 대한 실감성 평가

	합성영상 (n=5)	*실제 피부영상 (n=5)*
FMM	3.39 ± 1.22	**3.72 ± 1.19**
Shift-Net	**3.73 ± 1.02**	3.27 ± 1.29
DeepFill	3.06 ± 1.19	1.88 ± 1.26

제안된 방법은 몇 가지 위협요소를 안고 있다. 첫째, 합성영상에 대한 정성적 평가를 위한 설문조사는 참여자의 주관적 편향성에 좌우될 수 있다. 이러한 문제를 경감시키기 위해 방법과 내용이 참여자에게 블라인드(blind) 되도록 하였다. 둘째, 털 제거의 성능은 피부의 종류에 따라 좌우될 수 있다. 이러한 위험을 줄이기 위해, 데이터 증강을 위한 털 합성영상 생성 과정에서 Color jitter 를 통해 피부의 색상을 다양하게 변형하였다.

5. 결론

본 연구에서는 털 제거 심층학습 모델의 훈련을 위해 털이 많은 피부 합성영상 생성을 통한 학습 데이터 증강 기법을 제안하였다. PIE 기법으로 생성된 털 합성영상은 털 제거를 위한 인페인팅 심층학습 모델인 Shift-Net 과 DeepFill 을 학습하는데 사용되었다.

본 연구는 학습 데이터로서 증강된 털 합성영상의 사용이 털 제거 인페인팅 모델의 성능을 향상시킨다는 가설을 입증하지는 않는다. 다만, 실험결과는 털 합성영상 증강 데이터셋으로 훈련된 심층학습 모델 Shift-Net 이 적어도 합성영상에 대한 실험에서는 기존 영상처리 털 제거 기법에 비하여 정량적, 정성적 평가지표에서 우수한 성능을 보인다는 점을 보여준다.

본 연구는 여러 기술적 한계와 개선점을 갖고 있다. 털 합성영상 증강 데이터셋으로 심층학습 모델을 훈련하였을 때 실제 털이 많은 피부영상에서 털 제거의 효과가 입증되지 못하였다. 이는 털 제거 과정에서 피부의 텍스처가 상당히 손상되었기 때문으로 분석된다. 피부의 텍스처를 보존하기 위한 손실함수를 추가하는 등의 방법으로 모델의 성능을 개선할 필요가 있다. 또한, Inception Score (IS), Frechet Inception Distan-ce (FID) 등 실감성에 대한 정량적 지표를 사용한 평가가 보완되어야 한다.

이와 같은 기술적 한계에도 불구하고, 털 합성기반 데이터 증강 기법이자 털 제거를 위한 심층학습 모델에 대한 최초의 시도라는 점에서 의의가 있다. 본 연구는 후속 연구를 통하여 털이 많은 사람 피부영상 진단 뿐만 아니라 반려동물 피부진단 분야에도 적용하여 폭넓게 활용될 수 있다. 털 제거 심층학습 모델을 위한 학습 데이터가 부족한 현실에서, 합성 피부영상 기반의 데이터 증강 기법과 심층학습 기반 털 제거 모델은 피부영상의 화질을 개선하여 피부진단 기술의 발전을 위한 중요한 돌파구가 될 것으로 기대된다.

감사의 글

본 연구는 정보통신기획평가원의 정보통신·방송 기반조성 및 R&D 바우처 사업의 연구결과로 수행되었음(IITP-2021-0-00995 반려동물 피부 진단 및 치료 기능을 가진 인공지능 빗 플랫폼 개발).

참고문헌

[1] Q. Abbas, M.E. Celebi, I.F. García, “Hair removal methods: A comparative study for dermoscopy images,” Biomedical Signal Processing and Control, 6:4, pp. 395-404, 2011.

[2] P. Pérez, M. Gangnet, A. Blake, “Poisson image editing,” ACM SIGGRAPH 2003, pp. 313–318, 2003.

[3] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, “Shift-net: Image inpainting via deep feature rearrangement,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11218 LNCS, pp. 3–19, 2018.

[4] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, “Free-form image inpainting with gated convolution,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-October, pp. 4470–4479, 2019.

[5] Z. Qin, Q. Zeng, Y. Zong, and F. Xu, “Image inpainting based on deep learning: A review,” Displays, vol. 69, no. March, p. 102028, 2021.

[6] J. Koehoorn, A. Sobiecki, P. Rauber, A. Jalba, and A. Telea, “Effcient and Effective Automated Digital Hair Removal from Dermoscopy Images,” Math. Morphol. - Theory Appl., vol. 1, no. 1, pp. 1–17, Jan. 2016.

[7] K. Kiani and A. R. Sharafat, “E-shaver: An improved DullRazor® for digitally removing dark and light-colored hairs in dermoscopic images,” Comput. Biol. Med., vol. 41, no. 3, pp. 139–145, Mar. 2011.

[8] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, “Dullrazor®: A software approach to hair removal from images,” Comput. Biol. Med., vol. 27, no. 6, pp. 533–543, 1997.

[9] F.-Y. Xie, S.-Y. Qin, Z.-G. Jiang, and R.-S. Meng, “PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma,” Comput. Med. Imaging Graph., vol. 33, no. 4, pp. 275–282, Jun. 2009.

[10] A. Huang *et al.*, “A robust hair segmentation and removal approach for clinical images of skin lesions,”

in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3315–3318, Jul. 2013.

[11] A. Nasonova, A. Nasonov, A. Krylov, I. Pechenko, A. Umnov, and N. Makhneva, “Image Warping in Dermatological Image Hair Removal,” in Lecture Notes in Computer Science, vol. 8815, 2014.

[12] I. Zaqout, “An Efficient Block-Based Algorithm for Hair Removal in Dermoscopic Images,” in Computer Methods and Programs in Biomedical Signal and Image Processing, vol. i, no. Cell Interaction-Regulation of Immune Responses, Disease Development and Management Strategies, IntechOpen, pp. 1–15, 2020.

[13] D. Borys, P. Kowalska, M. Frackiewicz, and Z. Ostrowski, “A Simple Hair Removal Algorithm from Dermoscopic Images,” in Bioinformatics and Biomedical Engineering: New Advances, pp. 262–273, 2015.

[14] Q. Abbas, I. F. Garcia, M. Emre Celebi, and W. Ahmad, “A Feature-Preserving Hair Removal Algorithm for Dermoscopy Images,” Ski. Res. Technol., vol. 19, no. 1, pp. e27–e36, Feb. 2013.

[15] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, “Hair detection and lesion segmentation in dermoscopic images using domain knowledge,” Med. Biol. Eng. Comput., vol. 56, no. 11, pp. 2051–2065, 2018.

[16] K. Zafar et al., “Skin lesion segmentation from dermoscopic images using convolutional neural network,” Sensors (Switzerland), vol. 20, no. 6, pp. 1–14, Mar. 2020.

[17] F. Bornemann and T. März, “Fast image inpainting based on coherence transport,” J. Math. Imaging Vis., vol. 28, no. 3, pp. 259–278, 2007.

[18] Y. Wang, X. Tao, X. Qi, X. Shen, and J. Jia, “Image inpainting via generative multi-column convolutional neural networks,” Adv. Neural Inf. Process. Syst., vol. 2018-December, no. NeurIPS, pp. 331–340, 2018.

[19] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, “Image Inpainting for Irregular Holes Using Partial Convolutions,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11215 LNCS, pp. 89–105, 2018.

[20] C. Xie et al., “Image inpainting with learnable bidirectional attention maps,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2019-October, pp. 8857–8866, 2019.

[21] H. Salehinejad et al., “Generalization of Deep Neural Networks for Chest Pathology Classification in X-Rays Using Generative Adversarial Networks”, in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), pp. 990-994, 2018.

[22] 정현석, 하성민, 심학준, 오상호, 장혁재, "데이터 불균형 해결을 위한 조건부 심층 적대적 생성 신경망 기반 피부질환 영상생성 기법," 대한전자공학회 하계학술대회, 2020.

[23] O. Marques, “Morphological Image Processing,” in Practical Image and Video Processing Using MATLAB®, Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 299–334, 2011.

[24] H. Samet and M. Tamminen, “Efficient component labeling of images of arbitrary dimension represented by linear bintrees,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 10, no. 4, pp. 579–586, 1988.

[25] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative Image Inpainting with Contextual Attention,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 5505–5514, 2018

[26] Aaron van den Oord et al., "Parallel WaveNet: Fast High-Fidelity Speech Synthesis," Proceedings of the 35th International Conference on Machine Learning, PMLR 80:3918-3926, 2018.