반려동물 피부질환 검출을 위한 피부영상 털 영역분할에 대한 크로스 도메인 지식 증류 기법

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

Cross-domain Knowledge Distillation Approach to Hair Segmentation in Pet Skin Images for Skin Disease Detection

Youngchan Lee[1], Bumjin Park[1], Jae Young Choi[2], Bumshik Lee[3], Wonsang You[1]

[1]Information and Communications Engineering, Sun Moon University

[2]Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies

[3]Information and Communication Engineering, Chosun University

요 약

인공지능 기반 반려동물 피부진단을 위한 전처리 과정으로서, 피부영상으로부터 털을 제거하기 위한 털 영역분할 딥러닝 모델이 최근 제안되었다. 그러나, 학습 데이터의 부족으로 인하여 모델의 데이터 편향성 문제가 제기되어 왔다. 본 논문에서는 사람 또는 합성 피부영상으로 학습된 Teacher 네트워크의 털 특징정보를 사용하여 적은 양의 반려동물 피부영상 데이터로 훈련되는 Student 네트워크의 성능을 향상시키는 크로스 도메인 지식 증류 기법을 제안한다. 실험 결과, 기존 LadderNet 모델과 비교하여 LadderNet을 백본 네트워크로 하는 Teacher-Student 모델을 사용하였을 때 털 영역분할 정확도가 개선되었다(IoU=78.1>76%).

1. 서 론

인공지능 기반 피부영상진단 기술은 최근 사람 피부를 중심으로 급속도로 발전되어 온 반면[1], 반려동물 피부에 관한 연구는 전세계적으로 극히 미진하다. 반려동물을 키우는 인구가 늘어나면서 반려동물 피부관리 기술에 대한 관심과 수요가 증가하고 있는 산업발전의 추세를 보았을 때, 반려동물 피부영상진단을 위한 인공지능 기술의 의학적, 산업적 가치는 매우 높다고 보여진다.

피부영상진단을 위해서는 피부영상으로부터 털을 제거하는 전처리 과정이 필수적이다. 털 제거를 위해서는 털의 영역을 검출하는 털 영역분할 과정이 선행되어야 한다. 털 밀도가 낮은 사람 피부영상에서는 영상처리 기법만으로도 털 영역분할에 좋은 성능을 보여왔다. Tim Lee et al.은 형태학적 연산에 기반한 Dullrazor[2], Kimia Kiani et al.은 Radon transform과 Prewitt 필터를 사용한 E-shaver[3], Ihab Zaqout는 블록 기반의 형태학적 연산 기법을 제안하였다[4].

반려동물 피부영상에서는 털 밀도가 높고 종류도 다양하여 영상처리 기반 털 영역분할 방법은 좋은 성능을 보이지 않는다. 이러한 한계를 극복하기 위해, 최근 반려동물 피부영상에 최적화된 딥러닝 기반 털 영역분할 방법이 제안되었다[5,6]. Lee et al.은 반려동물 피부영상 데이터셋을 구축하고 U-net의 연쇄적 확장인 LadderNet 모델을 훈련하여 기존 영상처리 기법보다 털 영역분할의 정확도를 20.4 % 향상시켰다. 그러나, 모델 훈련에 사용 가능한 반려동물 피부영상 데이터셋의 크기(n=115)가 매우 작아 데이터 편향의 위험이 있다.

본 논문에서는 사람 또는 합성 피부영상으로 사전 학습된 털 특징정보를 사용하여 모델의 훈련 성능을 향상시키는 크로스 도메인 지식 증류 기법(cross-domain knowledge distillation)을 제안한다[7]. LadderNet을 백본 네트워크로 하는 Teacher-Student 네트워크 구조로서[8], Teacher 네트워크 에서는 사람 또는 합성 피부영상으로 사전 학습된 털 특징정보가 메모리 뱅크로 축적되고, Student 네트워크에서 입력된 반려동물 피부영상의 털 특징벡터를 메모리 뱅크의 털 특정정보와 비교하는 대조손실함수를 통하여 Student 네트워크의 성능이 향상된다. 또한, Teacher 네트워크의 훈련에 사용되는 학습 데이터의 크기를 증가시키기 위해 피부 텍스처와 털 합성을 통한 데이터 증강 기법을 사용하였다[9].

본 논문은 반려동물 피부영상에서 털 영역분할을 위해 사람 털 영역분할에 관한 학습 정보를 활용하는 크로스 도메인 지식 증류 기법을 처음 적용하였다는 점에서 독창성이 있다.

2. 방 법

2.1 LadderNet 모델

LadderNet은 의료영상분할에 사용되어 온 대표적인 모델인 U-Net을 확장한 네트워크로, 두 쌍의 인코더-디코더 구조를 가진다[10]. U-net의 연쇄적인 연결을 통해 네트워크 내에 여러 정보 흐름을 형성되어 학습 성능이 향상된다. 반면, 기존 U-net에 비하여 학습 파라미터의 수가 증가하므로, 모델 복잡도를 줄이기 위해 두 합성곱층 사이에 가중치를 공유하는 잔차 블록(shared-weights residual block)을 사용한다.

2.2 Teacher-Student 모델

Teacher-Student 네트워크 모델은 그림 1과 같이 사전학습된 Teacher 네트워크와 데이터로 훈련될 Student 네트워크로 구성된다[8]. Teacher 네트워크는 사람 피부영상 또는 합성 피부영상 데이터셋을 사용하여 사전에 학습되고 추출된 피부 및 털 특징벡터 정보가 메모리 뱅크(memory bank)에 저장된다(초록색 화살표).

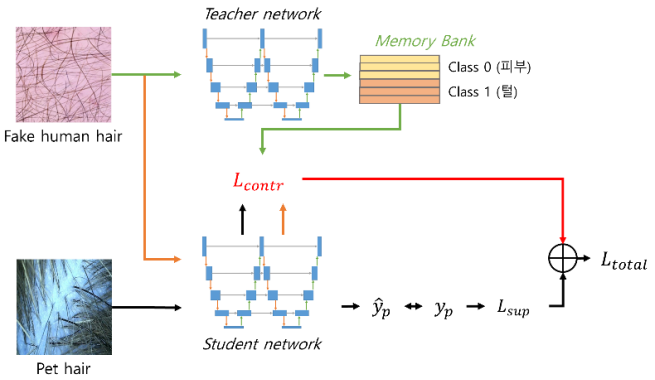

그림 1. Teacher-student 모델

Student 네트워크는, 지도손실함수(supervised loss, L_{sup})와 함께, 식 1과 같이 정의되는 대조손실함수(L_{contr})를 통하여 반려동물 피부영상 데이터셋으로부터 예측된 피부/털 특징벡터(p_c)와 Teacher 네트워크의 메모리 뱅크에 저장된 피부/털 특징벡터(z'_c) 사이의 유사도($D(p_c, z'_c)$)가 최소화되도록 훈련된다(검정색 화살표)[9].

$$L_{contr} = \frac{1}{C}\frac{1}{N_{p_c}}\frac{1}{N_{z'_c}}\sum_{c=1}^{C}\sum_{p_c \in P_c}\sum_{z'_c \in Z'_c} D(p_c, z'_c) \qquad (1)$$

대조손실함수(contrastive loss function)는 정답 레이블이 없는 데이터를 사용한 학습이 가능하도록 한 준지도학습(semi-supervised learning)에 주로 사용되나, 본 논문에서는 다른 학습 모델로부터 지식을 추출하여 사용하는 크로스 도메인 지식 증류에 활용되었다[10]. 마찬가지로, Teacher 네트워크 훈련에 사용된 사람 또는 합성 피부영상 데이터도 Student 네트워크에 입력되어 피부/털 특징벡터가 추출되어 대조손실함수를 통하여 메모리 뱅크의 피부/털 특징벡터와 비교되는 동일한 과정을 거친다(주황색 화살표).

Teacher 네트워크는 사전 학습되어 사용되지만, Student 네트워크의 학습이 진행됨에 따라 Teacher 네트워크의 가중치(θ_t)가 student 네트워크의 가중치(θ_s)로 점진적으로 바뀌도록 식 2에 따라 업데이트된다.

$$\theta_t = \tau\theta_t + (1-\tau)\theta_s \qquad (2)$$

Teacher-student 모델의 백본 네트워크로서 LadderNet 모델이 사용되었다(그림 1). Teacher 및 Student 네트워크의 훈련을 위해, 입력 영상을 48x48 크기의 190,000개 패치로 무작위 샘플링한 후, 학습될 패치의 10%를 교차검증을 위해 사용하였다. Batch size는 1024, 최대 epoch은 250로 설정하였고, 학습율은 0.001에서 0.0001까지 감소되도록 하였다. 최적화 기법으로는 Adam을 사용하였다.

2.3 학습 데이터

모델의 훈련 및 평가에 사용된 데이터로서, 사람 피부영상 100개, 반려동물 피부영상 115개 (반려묘 76, 반려견 39), 사람 합성 피부영상 3,000개, 반려동물 합성 피부영상 1,000개가 사용되었다. 사람 피부영상 데이터는 ISIC 2020 공개 데이터셋으로부터 다양한 배율(10배율 등)로 추출하였고, 반려동물 피부영상 데이터는 20배율 확대경을 사용하여 고배율 피부영상을 취득하였다. 취득된 사람 및 반려동물 피부영상 데이터로부터 털 영역 정답 마스크를 생성하였다.

LadderNet의 훈련에는 반려동물 피부영상 데이터가 사용된 반면, Teacher-Student 네트워크의 훈련에는 사람 피부영상 또는 합성 피부영상 데이터가 사용되었다. 합성 피부영상은 Poisson image editing (PIE)을 사용하여 사람 또는 반려동물 피부영상으로부터 추출된 털 영역을 털이 없는 피부 텍스처에 혼합하는 방식으로 합성하였다[9].

3. 결 과

모델 성능의 정량적 평가 지표로서 IoU(Intersection over Union), F1-score, Accuracy 및 AUC(Area Under Curve)를 사용하였다. 표 1은 LadderNet 및 Teacher-Student 네트워크 모델의 털 영역분할 결과를 정량적으로 보여준다. 특히, 사람 피부영상(Human), 사람 합성 피부영상(Fake human), 반려동물 합성 피부영상(Fake pet)으로 각각 훈련하여 사전학습된 Teacher 네트워크를 사용한 결과를 비교하였다.

표 1. 털 영역분할 정량적 평가 결과

모델	학습데이터	IoU	F1	ACC	AUC
LadderNet	-	0.567	0.714	0.784	0.862
Proposed	**Fake human**	**0.575**	**0.724**	0.781	0.860
	Human	0.56	0.712	0.774	0.854
	Fake pet	0.558	0.71	0.767	0.849

LadderNet을 백본으로 하는 Teacher-Student 네트워크 모델은 Teacher 네트워크의 훈련에 사람 합성 피부영상을 사용하였을 때 기존 LadderNet에 비하여 IoU와 F1-score가 향상되었다 (IoU= 0.575>0.567, F1=0.724>0.714). 반면, 사람 피부영상 또는 반려동물 합성 피부영상을 사용한 경우, 성능의 향상은 두드러지지 않았다.

그림 2는 반려동물 피부영상으로부터 털 영역분할의 예시를 보여준다. LadderNet에서는 털과 피부의 경계가 모호한 영역이 많은 반면, LadderNet을 백본으로 한 Teacher-Student 모델은 피부와 털의 영역을 더욱 명확하게 구분하고 있음을 볼 수 있다. 이러한 영역분할 성능 개선 효과는 얇은 털의 영역에서 더욱 두드러진다.

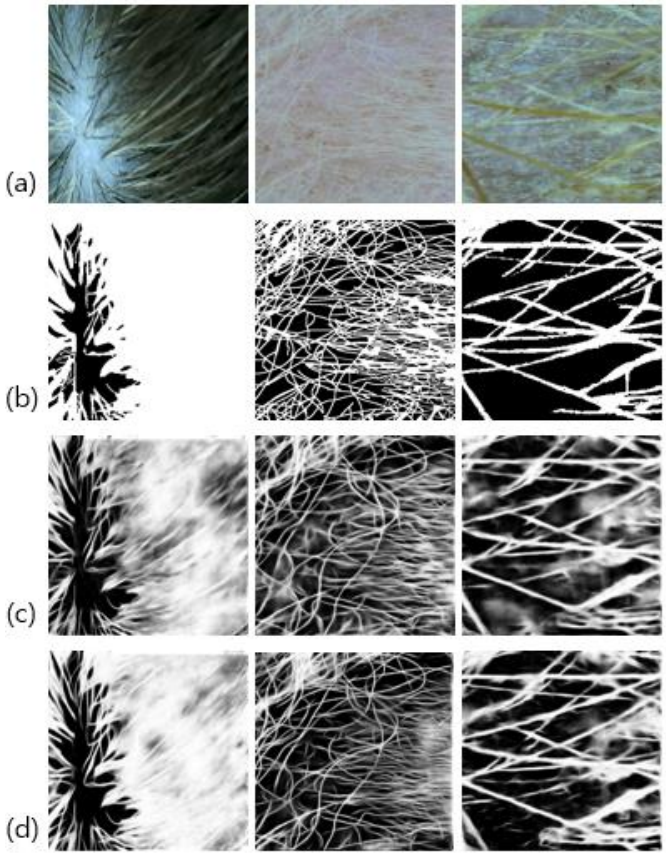

그림 2. 반려동물 피부영상 털 영역분할 결과. (a) 원본 반려묘 피부영상(좌측 2)과 반려견 피부영상(우측 1)에 대한 (b) 정답 마스크, (c) LadderNet, (d) Teacher-Student 모델.

4. 결론 및 향후 연구

본 논문에서는 반려동물의 고배율 피부영상으로부터 털 영역분할을 위해, LadderNet을 백본으로 하는 Teacher-Student 네트워크 모델을 기반으로 사람 피부영상으로부터 학습된 털 특징정보를 사용하여 반려동물 피부영상의 네트워크 훈련에 사용하는 지식 증류 기법을 사용하였다.

실험결과, Teacher 네트워크의 사전학습 데이터로서 사람 합성 피부영상을 사용하였을 때, 기존 LadderNet보다 향상된 성능을 보여주었다. 이러한 성능 개선은 얇은 털 영역에서 두드러졌는데, 털이 얇은 사람 피부영상으로부터 학습한 지식이 반려동물 피부영상의 털 영역분할의 성능을 향상시키는데 기여한 것으로 보인다.

이러한 성과에도 불구하고 본 연구에는 몇 가지 한계가 있다. 첫째, Teacher 네트워크의 훈련을 위해 합성 피부영상을 사용하였지만, 실제 피부영상에 비하여 실감성이 떨어지고 반려동물 고유의 털 패턴을 반영하지 못한다. 둘째, 털의 특징벡터가 털의 구조적 특징을 반영하는데 한계가 있다. 셋째, 털이 촘촘하게 뭉친 영역의 경우 털 영역을 전체적으로 분할하는 성능이 저하된다. 넷째, 학습 데이터의 종류에 따른 모델 성능을 비교하는 과정에서 데이터셋 크기 등 학습환경의 동일성을 엄밀하게 확보하지 못하였다.

이러한 문제를 개선하기 위해, 반려동물 털 패턴을 실감성 있게 생성하는 데이터 증강 기법을 개량하고, 털의 구조적 특징을 반영하도록 털 특징벡터의 구조를 개선하며, 모델 성능 비교를 위한 실험환경의 동일성을 확보하고, 어텐션 레이어를 사용하여 털의 밀도가 적은 부분을 자동으로 검출하여 성능을 개선하도록 할 계획이다. 향후 연구에서는 털 영역분할과 피부복원 인페인팅을 동시에 학습하는 딥러닝 모델로 확장할 계획이다. 본 연구는 피부영상 기반 반려동물 피부진단에 대한 의학적 연구와 인공지능 기반 반려동물 헬스케어 기술의 발전을 위한 중요한 초석을 놓는다.

참고문헌

[1] L. Talavera-martinez et al., “Hair Segmentation and Removal in Dermoscopic Images Using Deep Learning," in IEEE Access, vol. 9, pp. 2694-2704, 2021.

[2] T. Lee et al., “Dullrazor®: A software approach to hair removal from images”, Computers in Biology and Medicine, vol. 27, no. 6, pp. 533 - 543, 1997.

[3] K. Kiani et al., “E-shaver: An improved Dullrazor® for digitally removing dark and light-colored hair in dermoscopic images”, Computers in Biology and Medicine, vol. 41, no. 3, pp. 139 - 145, 2011.

[4] I. Zaqout, “An Efficient Block-Based Algorithm for Hair Removal in Dermoscopic Images”, Computer Methods and Programs in Biomedical Signal and Image Processing, IntechOpen, 2020.

[5] 이영찬, 박범진, 유원상, “LadderNet 기반 반려동물 피부영상 털 영역분할에 관한 탐색적 연구”, 대한전자공학회, vol. 2021, no. 11, 2021.

[6] 이영찬, 박범진, 최재영, 이범식, 유원상, “반려동물 피부영상에서의 질환 검출을 위한 딥러닝 기반 털 영역분할 기법: 기존 방법의 적용성 평가”, 한국차세대컴퓨팅학회 논문지, vol. 18 no. 1, 2022.

[7] H. Kim et al., “Multi-domain Knowledge Distillation via Uncertainty-Matching for End-to-End ASR Models”, Interspeech, pp.1311 - 1315, 2021.

[8] I. Alonso et al., “Semi-Supervised Semantic Segmentation with Pixel-Level Contrastive Learning from a Class-wise Memory Bank”, CVPR, 2021.

[9] 이영찬, 박범진, 유원상, “털 합성 데이터 증강을 통한 피부영상 털 제거 심층학습 기법: 기존 방법의 적용성 평가”, 영상처리 및 이해에 관한 워크샵(IPIU), 2022.

[10] Z. Zhuang, “LadderNet: Multi-path networks based on U-Net for medical image segmentation”, arXiv:1810.07810, pp. 2 - 5, 2018.