

스타일 전이 Diffusion 모델의 백색 이미지 유도 기법을 통한 족적 데이터 증강

권이진¹, 한상범², 이민식^{1,2}

¹한양대학교 ERICA 전자공학부, ²한양대학교 전자공학과

rnmjsdlwls01@hanyang.ac.kr, gkstkdqja88@hanyang.ac.kr, mleepaper@hanyang.ac.kr

Footprint Data Augmentation via White Image Guidance in Style Transfer Diffusion Models

Ijin Kwon¹, Sangbeom Han², Minsik Lee^{1,2}

¹School of Electrical Engineering, Hanyang University ERICA,

²Department of Electrical and Electronic Engineering, Hanyang University

요약

본 논문에서는 diffusion 모델 기반 스타일 전이 기법을 통한 데이터 증강으로 족적 템색 성능 향상 기법을 제안한다. 본 연구 분야는 현장에서 촬영된 족적 이미지와 사전에 수집된 족적 이미지들을 딥러닝 모델을 활용해, 현장에서 촬영된 족적 이미지가 사전에 수집된 족적 이미지 중 어떤 신발 모델에 해당하는지를 유사도 거리 기반으로 판단한다. 하지만, 모델의 학습에 사용할 현장 족적 이미지의 수가 부족하고 사전에 수집된 족적 이미지와 비교해 현장에서 촬영된 족적 이미지가 훼손이 심하여 모델의 학습 난이도가 높아 성능을 확보하기 어렵다. 이를 극복하기 위해, 사전에 수집된 족적 이미지를 현장 이미지와 같은 다양한 훼손 정보를 전이하는 방법으로 데이터를 증강시켜 최종적인 족적 템지 성능을 향상한다.

I. 서 론

최근 딥러닝 기술이 큰 발전을 이루며, 족적 수사 분야에 딥러닝 기술을 도입하는 다양한 연구가 이루어졌다. 신발의 족적 이미지(reference 이미지)를 학습된 딥러닝 모델에 통과시켜 특징 벡터를 구한다. 그리고 사건 현장에서 촬영한 족적 이미지(query 이미지)를 딥러닝 모델에 통과시켜, 특징 벡터를 구한다. 이렇게 구한 query 이미지의 특징 벡터와 reference 이미지들의 특징 벡터 간의 유사도 순위를 측정하여, 가장 유사한 신발을 찾는다. 하지만 신발 데이터가 부족하고 훼손이 심하여, 이미지 간의 도메인 차이로 인해 모델의 학습이 어렵다.

이 문제를 해결하기 위해, 자기지도 학습 방법을 사용해 라벨링이 없는 원시 데이터들로 학습을 진행하는 방식[1]을 사용하였다. 하지만, 이런 방식들도 query 이미지의 훼손에서 오는 reference 이미지와의 차이를 보완하기는 어렵다.

본 논문에서는 reference 이미지를 사건 현장의 query 이미지로 스타일 전이하고 모델 학습 데이터 세트를 증강하는 방법을 제안한다. 세부적으로 diffusion 모델[2] 기반의 흰색 이미지를 활용한 스타일 전이 기법을 통해 데이터 증강을 수행한다. 특히, 사전 학습된 diffusion 모델을 사용하면서 각 스텝 별 결과로 얻은 이미지를 단순 흑백 이미지 변환하는 기법을 도입한다.

II. 본론

II-1. Diffusion 모델의 흑백 이미지 변환 기법

현재 사용되는 diffusion 모델의 경우, 컬러 이미지에 대하여 사전 학습되어 있기 때문에, 흑백의 족적 이미지를 스타일 이미지로 하여 스타일 전이시키면 원치 않는 이상 색상 현상이 발생한다.

따라서, 본 논문에서는 diffusion 모델의 매 스텝마다 결과를 3채널 평균값을 취하는 방법을 제안한다. 이는 족적 이미지 스타일 전이 결과 이미지에 이상 색상이 나타나는 현상을 효과적으로 방지할 수 있다.

II-2. CLIP and White Image Guided Style Disentanglement Loss

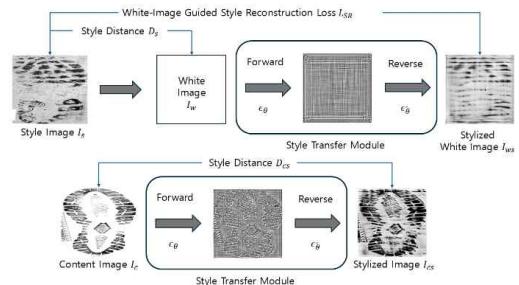


그림1. 제안한 스타일 전이 모델의 pipeline

CLIP-based Style Disentanglement[3] 손실함수는 CLIP 모델[6]을 사용하여 스타일 거리를 측정한다. 이는 다음과 같이 정의된다.

$$D_s = E(I_s) - E(I_{sc}) \quad (1)$$

$$D_{cs} = E(I_{cs}) - E(I_{cc}) \quad (2)$$

E 는 CLIP 모델의 이미지 encoder, I_s 는 스타일 이미지, I_{sc} 는 I_s 에서 스타일 제거한 결과, I_{cs} 는 스타일이 전이된 결과, 그리고 I_{cc} 는 스타일 전이할 콘텐츠 이미지를 스타일 제거한 결과다. 이를 기반으로 다음과 같은 손실함수로 구성된다.

$$L_{SD}^{L1} = \|D_{cs} - D_s\|, \quad L_{SD}^{dir} = 1 - \frac{D_{cs} \cdot D_s}{\|D_{cs}\| \|D_s\|} \quad (3)$$

$$L_{SD} = \lambda_{L1} L_{SD}^{L1} + \lambda_{dir} L_{SD}^{dir} \quad (4)$$

L_{SD}^{L1} 는 추출된 스타일 정보와 콘텐츠 이미지의 스타일 정보가 서로 일치하도록 최적화하는 것이다. 또한, 추출된 두 스타일 정보가 동일한 방향을 향하도록 유도하는 L_{SD}^{dir} 손실함수를 도입한다. 다음과 같은 새로운 손실함수를 정의한다.

$$D_s = E(I_s) - E(I_w) \quad (5)$$

본 논문에서는 흰색 이미지 I_w 와 스타일 이미지 간의 스타일 거리를 측정하여, 스타일 이미지의 색감과 화풍뿐만 아니라 스타일 이미지의 전체적인 정보가 온전히 전이되도록 하였다.

II-3. White Image Guided Style Reconstruction Loss

Style Reconstruction Loss[3]는 스타일 제거된 스타일 이미지를 다시 스타일을 복원한 이미지 I_{ss} 와 원본 스타일 이미지 I_s 간의 차이를 구하는 손실함수이다. 다음과 같이 손실함수를 정의한다.

$$L_{SR} = \|I_{ss} - I_s\| \quad (6)$$

이와 달리, 본 논문에서는 족적 이미지의 훼손 정보를 온전히 전이시키기 위해 백색 이미지를 도입한다. 그러하여 백색 이미지를 기반으로 스타일 전이시킨 이미지 I_{ws} 와 원본 스타일 이미지 I_s 간의 차이를 구하는 손실함수를 사용한다.

$$L_{SR} = \|I_{ws} - I_s\| \quad (7)$$

III. 실험 결과

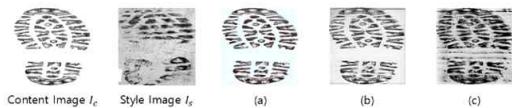


그림 2. 스타일 전이 비교 결과

그림 2에 대해서, (a)의 결과는 Style Diffusion[3]을 활용하여 스타일 전이한 결과, (b)의 결과는 DDIM의 각 시간 스텝에서 흑백으로 변환한 기법으로 적용한 결과, (c)는 본 논문에서 제안한 기법을 모두 적용한 결과이다. Style Diffusion과 비교하여, 본 논문에서 제안하는 방법이 이상 색상 요소를 억제하고, 사건 현장의 query 이미지의 특징이 효과적으로 reference 이미지에 전이된 것을 확인할 수 있다. 그림 3은 본 논문이 제안한 기법의 추가 결과들이다.

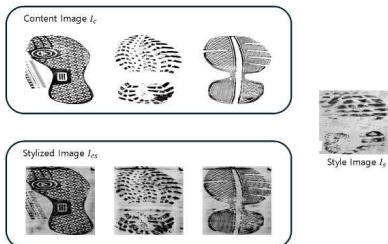


그림 3. fid-300[5]에 대한 스타일 전이 결과

족적 검색 모델 학습에 사용하기 위하여, reference 이미지를 사건 현장의 query 이미지에 대해 스타일 전이를 수행하였다. 족적 검색 모델은 ResNet-50을 사용하였으며, Cross Entropy Loss를 사용하여 특징 벡터를 추출하도록 학습을 진행하였다. 학습과 평가는 fid-300[5] 데이터 세트가 사용되었다.

	Top 1%	Top 5%	Top 10%
(1)	14%	38%	56%
(2)	15%	40%	50%
(3)	17%	49%	63%

표 1. fid-300의 족적 템색 성능 결과

표 1에 대해서, (1)는 이미지 증강 없이 학습을 진행한 결과, (2)는 Rand Augment[4]을 사용하여 1장씩 증강하여 학습한 결과, (3)는 본 논문에서 제안한 방법을 사용하여 1장씩 이미지 증강하여 학습한 결과다. 본 논문에서 제안한 방법은 모든 수치에서 더 높은 성능을 기록하였다.

IV. 결론

본 논문에서는 족적 검색 모델의 성능 향상을 위하여, 스타일 전이를 통한 이미지 증강을 제안하였다. 기존 이미지 증강에 많이 사용되는 Rand Augment와 비교하여, 유의미한 성능 향상을 확인하였다. 제안된 방법을 통하여, 족적 이미지 사건 현장 족적 이미지 간의 도메인 차이를 줄이고, 데이터 세트의 부족 문제 해결에 기여할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 지원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임
(IITP-2026-RS-2020-II201741)

참 고 문 헌

- [1] 이창엽, 김동주, 서영주, 황도경, “효과적인 족적 감정 수사를 위한 자기지도 학습 및 이진화 기반 이미지 검색 기술”, 한국정보기술학회논문지, vol. 22, no. 7, pp. 1-11, 2024.
- [2] Jiaming Song, Chenlin Meng, and Stefano Ermon, “Denoising Diffusion Implicit Models”, in International Conference on Learning Representations (ICLR), 2021.
- [3] Zhizhong Wang, Lei Zhao, and Wei Xing, “StyleDiffusion: Controllable Disentangled Style Transfer via Diffusion Models”, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 7677-7689.
- [4] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le, “RandAugment: Practical Automated Data Augmentation With a Reduced Search Space”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020, pp. 702-703.
- [5] Adam Kortylewski, Thomas Albrecht, and Thomas Vetter, “Unsupervised Footwear Impression Analysis and Retrieval from Crime Scene Data”, in Asian Conference on Computer Vision (ACCV), 2014, pp. 644-658.
- [6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al., “Learning Transferable Visual Models from Natural Language Super Vision”, in International Conference on Machine Learning (ICML), 2021, pp. 8748 - 8763.