박물관 넘어 도망친 화가들

김현지†, 송지언†, 여화선†, *강제원
이화여자대학교 엘텍공과대학 전자전기공학과
{guswl8033, 96catherine, emiliehue}@ewhain.net, *jewonk@ewha.ac.kr

Painters who Climbed Out the Museum and Disappeared

Hyeonji Kim†, Jiuhn Song†, Hwaseon Yeo†, and *Je-won Kang
Department of Electronic and Electrical Engineering Ewha Womans University

요약

본 팀은 웹캠으로 촬영한 영상에서 원하는 물체를 선택하여 텍스처를 선택한 이미지의 스타일로 변환하는 프로젝트를 수행했다. 영상을 세그멘테이션하고 원하는 물체만을 원하는 텍스처로 변환하여 최종 아웃풋을 얻는다. 제안하는 네트워크는 물체를 다양한 스타일로 바꾸는 것이 가능한데, 이 중에서 이미지에 명화의 화풍을 입히는 것을 중점으로 하여 데모를 구현했다.

빠른 속도로 네트워크를 실행하기 위해 기존 연구들에 비디오 처리의 관점을 접목했다. 여러 프레임을 묶어 옵티컬 플로우를 생성하고, 첫 번째 프레임을 인스턴스 세그멘테이션한 후 마스크를 추출했다. 이후 마스크 영역만 뽑아낸 이미지를 새로운 입력으로 하여 스타일 트랜스퍼를 거치고, 이 첫번째 프레임과 나머지 프레임들의 옵티컬 플로우로 나머지 프레임들의 세그멘테이션과 스타일 트랜스퍼를 예측하여 다시 비디오 프레임으로 만들어 주었다.

본 알고리즘은 옵티컬 플로우 설정으로 네트워크의 계산량을 줄이며 속도를 개선했다. 빠른 데이터 처리로 사용자가 원하는 물체의 텍스쳐가 바뀔 수 있게 되었고, 이는 현실 세계가 실제로 바뀐 듯한 느낌을 들게 한다. 또한, 컴퓨터 비전에서 활발하게 연구되었던 분야를 AR로 끌어와 두 분야의 융합 가능성을 열었다.

현재 코로나의 영향으로 집에서 취미생활을 즐기는 인구가 많아졌다. 본 연구를 통해 많은 사람에게 집에서 쉽게 명화의 감성을 즐기고 느낄 수 있는 양질의 콘텐츠를 제공해주려 한다. 또한, 박물관과 미술관 등의 기관에서도 이 기술이 활용될 수 있다. 명화를 느낄 수 있는 다양한 콘텐츠를 이용하여 박물관이나 미술관의 홍보 효과도 기대할 수 있다.

1. 작품의 개발 동기 및 필요성

본 팀은 웹캠으로 촬영한 영상에서 원하는 물체를 선택하여 빠른 속도로 텍스처를 변환하는 프로젝트를 수행했다. 영상을 세그멘테이션 후, 원하는 물체만을 원하는 텍스처로 변환하여 최종 아웃풋을 얻는다. 본 과제는 물체를 다양한 스타일로 바꾸는 것이 가능한데, 이 중에서 이미지에 명화의 화풍을 입히는 것을 중점으로 하여 데모를 구현했다.

기존 연구의 알고리즘[3]은 이미지 처리를 수행하기에는 속도가 현저히 느린 경우가 많다. 전체 이미지에 대한 스타일 트랜스퍼의 연구는 있지만, 특정 물체에만 스타일을 적용하는 연구는 거의 없다. 원하는 물체를 검출하기 위해서는 세그멘테이션을 거쳐야 하므로 보다 긴 시간이 소요된다. 이러한 배경을 바탕으로 본 팀은 네트워크의 수행 속도를 높이기 위해 기존 연구들에 비디오 처리의 관점을 접목했다. 여러 프레임을 묶어서 옵티컬 플로우를 생성함으로써 네트워크의 계산량을 크게 줄여 처리 속도를 높였다. 또한, 세그멘테이션 과정에서 마스크를 추출하여 해당 부분만 스타일 트랜스퍼를 거치며 다시 한번 계산량을 줄이도록 하였다.

본 과제는 다양한 영상 처리 기술을 접목하여 기존 연구의 속도의 한계점을 개선 시켰다. 이를 통해, 사용자는 원하는 물체의 텍스쳐가 빠른 속도로 바뀜을 확인할 수 있으며, 이는 현실 세계가 실제로 바뀐 듯한 느낌이 들게 한다.

†해당 저자들은 동일하게 기여함

2. 선행 기술 조사

컴퓨터 비전에서 세그멘테이션과 스타일 트랜스퍼는 오랜 시간 연구되어온 분야이다. 세그멘테이션에는 여러 방식이 존재하는데, 본 과제에서는 배경을 제외하고 사물들을 각각 구별하는 인스턴스 세그멘테이션을 활용하려 한다. 이는 Mask R-CNN[1]을 통해서 처음으로 나온 세그멘테이션 방식이며, 최근에는 YOLACT++[2]와 같은 발전된 성능의 연구도 나왔다. 스타일 트랜스퍼는 이미지의 스타일을 바꾸어주는 것이다. 한 쌍의 이미지를 인풋으로 넣어주면, 원하는 스타일의 이미지에 맞춰 변형하려 하는 이미지의 스타일을 바꿔주는 방식을 취한다. Perceptual Losses for Real-Time Style Transfer and Super- Resolution[3]이 2016년에 빠른 속도의 스타일 트랜스퍼에 대한 연구를 발표하였고 이후, 스타일 트랜스퍼의 속도를 높이고자 연구를 하는 논문들이 출간되었다. 하지만, 실시간으로 스타일 트랜스퍼를 진행하는 연구는 다양하지 않다. 그중에서 Learning Linear Transform for Fast Image and Video Style Transfer[4]는 최근 연구 중, 가장 빠른 속도로 스타일 트랜스퍼를 수행한다. 그리고, 본 과제의 목적과 유사하게, 세그멘테이션과 스타일 트랜스퍼를 융합한 연구도 존재한다. CBS(Class-Based Styling)[5]는 기존의 네트워크를 이용해 원하는 클래스만을 스타일 트랜스퍼를 진행한다. 하지만, 속도가 느려 본 과제에 적용하기 위해서는 네트워크의 개선이 필요하다. 비디오 처리 및 압축에 많이 쓰이는 방법

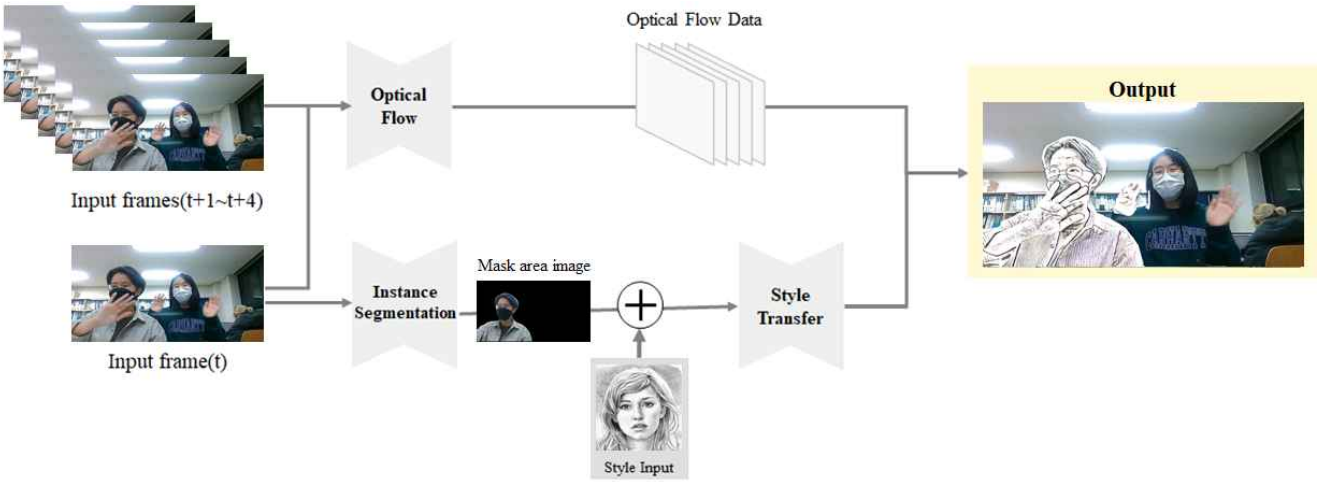

그림 1. 본 과제의 Flow Chart

의 하나인 옵티컬 플로우는 비디오에서 앞뒤 프레임 간의 픽셀들의 변화량이라고 할 수 있다. 전통적으로 여러 가지 방법이 있지만, 최근 딥러닝으로도 이를 구현하는 연구도 나오고 있다. 대표적인 연구로는 FlowNet2.0[6]이 있다.

3. 과제 해결방안 수행과정

본 과제에서는 네트워크의 수행 속도를 높이기 위해, 기존 연구들에 비디오 처리 기술을 접목한다. 그림 1을 참고하면, 일정한 개수의 프레임 단위로 옵티컬 플로우를 검출함과 동시에, 프레임 묶음 중 한 개의 프레임에만 세그멘테이션을 적용해 나머지 프레임의 세그멘테이션 결과를 예측한다. 세그멘테이션 수행을 하는 동안 옵티컬 플로우를 백그라운드에서 검출함으로써 세그멘테이션과 스타일 트랜스퍼에 들어가는 프레임 수를 1/5로 줄여 수행 속도를 보다 높일 수 있다. 또한, 세그멘테이션 마스크 영역만을 추출하여 나머지 부분의 정보를 없애줌으로써 스타일 트랜스퍼를 거칠 때의 계산량을 줄인다.

3.1 과제의 개념 및 설계

여러 프레임들 중 첫 프레임만 인스턴스 세그멘테이션과 스타일 트랜스퍼를 거치면 나머지 프레임들의 결과는 옵티컬 플로우를 활용하여 예측할 수 있다. 이를 통해 복잡한 네트워크에 들어가는 프레임의 수를 크게 줄여 전체 네트워크의 계산량을 줄이고 속도를 빠르게 한다. 이러한 과정을 위해 본 과제에서는 웹캠으로 촬영한 영상을 다섯 프레임 단위로 나눠 연속한 두 프레임 사이의 옵티컬 플로우를 각각 추출한다. 이렇게 얻어진 4장의 옵티컬 플로우와 첫 번째 프레임의 아웃풋의 와핑을 통해 나머지 4개 프레임의 결과를 예측한다. 전체 네트워크가 한번 통과될 때마다 5개의 프레임의 결과가 얻어지며 같은 방식으로 입력 영상의 모든 프레임을 추정하여 동영상 결과를 얻을 수 있다.

3.2 구체적 프레임워크

옵티컬 플로우를 얻기 위해서 FlowNet2.0[6]의 FlowNetS 모델을 사용했으며, 원하는 클래스의 물체를 스타일 트랜스퍼 하기 위해 먼저 인스턴스 세그멘테이션 YOLACT++[2]로 객체를 검출하고 마스크 이미지를 얻어냈다. 그림 2를 참고하면, YOLACT++를 이용하여 추출한 마스크 이미지와 해당 영역의 입력 이미지를 확인할 수 있다.

이후 인스턴스 세그멘테이션의 이미지를 스타일 트랜스퍼[4]의 입력으로 넣는다. 해당 논문은 비디오 프레임의 결과들이 프레임 간 차이가 안정적이라는 장점이 있다. 본 논문에서는 인코더의 relu3_1, relu4_1 레이어에서 각각 스타일 이미지의 특징을 추출하였으며, 본 팀은 성능을 위해 더 깊은 relu4_1 레이어에서 특징을 추출하여 사용하였다. 이후 스타일 트랜스퍼 된 이미지와 마스크 이미지를 곱하고 원본 이미지에 덮어씌움으로써 최종 결과 영상을 도출하였다.

4. 과제 결과 및 분석

본 실험은 Intel(R) Xeon(R) CPU E5-1620 v4 @ 3.50GHz와 gpu Nvidia GTX TITAN XP 1개로 진행하였다.

그림 3을 참고하면 최종 결과 영상의 다섯 프레임을 확인할 수 있다. 앞쪽의 사람만이 스타일 트랜스퍼 되었으며, 첫 프레임은 원본, 나머지 네 프레임은 첫 프레임과 옵티컬 플로우를 이용한 와핑된 영상이다.

과제에 사용한 방법의 성능 향상을 비교하기 위해 와핑을 사용했을 때와 와핑없이 모든 프레임을 스타일 트랜스퍼 했을 때의 프레임당 소요 시간을 비교, 평가하였다.

<표 1> 와핑 유무에 따라 소요되는 프레임 당 시간(sec/frame)

	프레임당 시간 (sec/frame)
옵티컬 플로우 有	0.43
옵티컬 플로우 無	0.65

<표 1>은 입력 영상을 넣었을 때 결과 영상이 나오는 결과 시간을 비교한 것이다. 옵티컬 플로우를 사용하여 와핑 했는지에 따라 시간을 비교하였다. 본 과제에서 사용한 방법이 그렇지 않을 때보다 1.5배가량 빨라 효율적임을 알 수 있다.

5. 기대효과 및 향후 연구 방향

본 과제는 웹캠에서 촬영한 영상에서 사용자가 원하는 물체의 텍스쳐를 바꿀 수 있고, 이는 현실 세계가 실제로 바뀐 듯한 느낌을 들게 한다.

그림 2. 중간 결과 영상(왼쪽부터 차례대로 입력 영상, 옵티컬 플로우 영상, 마스크 영역 영상, 마스크 영상)

그림 3. 본 과제의 수행 최종 결과(위: 첫 번째 결과 원본 프레임(빨간색 네모)과 나머지 와핑 된 프레임 영상들, 아래: 연속된 두 프레임 사이의 옵티컬 플로우)

또한, 컴퓨터 비전에서 활발하게 연구되었던 분야를 AR로 끌어와 두 분야의 융합 가능성을 열었다. 현재 코로나의 영향으로 집에서 취미생활을 즐기는 인구가 많아졌다. 본 연구를 통해 많은 사람에게 집에서 쉽게 명화의 감성을 즐기고 느낄 수 있는 양질의 콘텐츠를 제공해주려 한다. 또한, 박물관과 미술관 등의 기관에서도 이 기술이 활용될 수 있다. 예를 들어, 물체를 작품들의 화풍으로 변형해보는 체험용 전시관을 만들 수 있을 것이다. 이러한 콘텐츠를 이용하여 박물관이나 미술관의 홍보 효과도 기대할 수 있다.

현재 인스턴스 세그멘테이션에서 한 클래스 안의 물체가 프레임마다 다르게 분류되는 경우가 있다. 향후 객체 추적 등을 통해 성능을 개선할 예정이다.

Acknowledgement

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음" (IITP-2020-0-01460)

5. 참고문헌

[1] Kaiming He, Georgia Gkioxari, Piotr Dollar and Ross Girshick(2017), "Mask R-CNN", ICCV

[2] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee(2020), "YOLACT++: Better Real-time Instance Segmentation", https://arxiv.org/pdf/1912.06218.pdf

[3] Justin Johnson, Alexandre Alahi and Li Fei-Fei(2016), "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", ECCV

[4] Xueting Li, Sifei Liu, Jan Kautz and Ming-Hsuan Yang(2019), "Learning Linear Transformation for Fast Image and Video Style Transfer", CVPR

[5] Lironne Kurzman, David Vazquez and Issam Laradji(2019), "Class-Based Styling: Real-time Localized Style Transfer with Semantic Segmentation", ICCV

[6] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, Thomas Brox(2017), "FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks", CVPR