딥러닝 기반 동영상 객체 분할 기술 동향

□ 고영준 / 충남대학교

요약

동영상 프레임 내 객체 영역들을 배경으로부터 분할하는 기술인 동영상 객체 분할(video object segmentation)은 다양한 컴퓨터 비전 분야에 활용 가능한 연구 분야이다. 최근, 동영상 객체 분할과 관련된 연구 내용으로 CVPR, ICCV, ECCV의 컴퓨터 비전 최우수 학회에 매년 20편 가까이 발표될 정도로 많은 관심을 받고 있다. 동영상 객체 분할은 사용자가 제공하는 정보에 따라 비지도(unsupervised) 동영상 객체 분할, 준지도(semi-supervised) 동영상 객체 분할, 인터렉티브(interactive) 동영상 객체 분할의 세 카테고리로 분류할 수 있다. 본 고에서는 최근 연구가 활발하게 수행되고 있는 비지도 동영상 객체 분할과 준지도 동영상 객체 분할 연구의 최신 동향에 대해 소개하고자 한다.

I. 서 론

동영상 객체 분할은 동영상 프레임 내 객체 영역들을 배경으로부터 분할하는 기술을 의미한다. 동영상 객체 분할 기술은 동영상 요약, 동영상 검색, 행위 인식, 객체 클래스 학습, 3차원 객체 모델링 등의 많은 컴퓨터 비전 분야에 적용 가능한 중요한 연구 분야이다. 동영상 객체 분할은 사용자의 관여 정도에 따라 크게 비지도 동영상 객체 분할, 준지도 동영상 객체 분할, 인터렉티브 동영상 객체 분할 세 카테고리로 분류할 수 있다. 비지도 동영상 객체 분할은 객체에 대한 어떠한 사용자 주석(user annotation) 정보 없이 배경으로부터 객체를 분할하는 것을 목표로 한다. 준지도 동영상 객체 분할은 첫 프레임에서 사용자가 제공한 타겟 객체에 대한 정확한 분할 영역을 이용하여 이 후 프레임에서의 타겟 객체 분할을 수행한다. 인터렉티브 동영상 객체 분할은 사용자와의 상호작용을 통해 객체 분할 결과를 개선한다. 사용자가 개선 영역에 대한 정보를 반복적으로 제공해야 되기 때문에, 인터렉티브 동영상 객체 분할에서는 빠르게 제공 가능한 점

(point click)이나 scribble 등의 형태를 갖는 사용자 주석 정보를 이용하여 동영상 객체 분할을 수행한다. 〈그림 1〉은 지도(supervision) 수준에 따라 분류된 동영상 객체 분할 방식을 도시한다. 준지도 방식에서 필요한 첫 프레임에서의 정확한 타겟 객체 분할 영역 정보는 사람이 획득하는데, 대략 79초 이상의 시간이 소요될 만큼 많은 작업량을 요구한다. 반면, 인터렉티브 방식에서 많이 사용되는 scribble은 타겟 객체를 지정하는데 3초 이내의 시간밖에 소요되지 않으나, 실험 결과를 지켜본 후, 수정에 필요한 추가적인 scribble을 제공해야 한다.

본 고에서는 최근 활발히 연구되고 있는 비지도 방식과 준지도 방식의 최신 기술 동향을 살펴보고자 한다. II장에서는 비지도 동영상 객체 분할과 준지도 동영상 객체 분할의 최신 연구를 순차적으로 소개한다. III장에서는 최신 동영상 객체 분할에 대한 성능 비교 결과를 제시하고, 마지막으로 V장에서 결론을 맺는다.

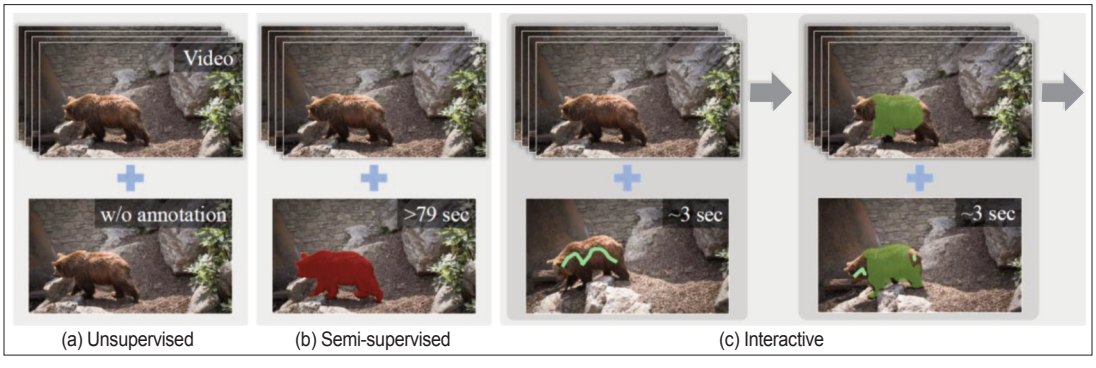
II. 동영상 객체 분할 기술 동향

이번 장에서는 먼저 비지도 동영상 객체 분할 기술을 소개하고, 이 후 준지도 동영상 객체 분할의 최신 기술 동향을 살펴본다.

1. 비지도 동영상 객체 분할 기법

비지도 동영상 객체 분할은 객체에 대한 어떠한 사용자 주석(user annotation) 정보 없이 배경으로부터 객체를 분할하는 것을 목표로 한다. 사용자 주석 없이 동영상 객체 분할을 수행하는 것은 매우 어려운 일이기 때문에, 비지도 동영상 객체 분할 기법들은 동영상에서 가장 빈번하게 출현하는 주요 객체를 타겟으로 설정한다. 이 관점에서 비지도 동영상 객체 분할은 동영상에서 색상과 움직임 측면에서 많은 관심을 받는 중요한 객체를 자동적으로 탐지하는 동영상 중요 객체 탐지(video salient object detection) 기술과 연관성이 깊다.

딥러닝 기술이 도입되기 전의 비지도 동영상 객체 분할은 주요 객체에 대한 시각적 단서를 찾기 위해 움직임 경계선(motion boundary)[1], 객체 제안 영역(object proposal)[2], 영상 중요도 맵(image saliency map)[3] 등을 활용하였다. 대부분의 비지도 동영상 객체 분할 기법들은 이와 같은 시각적 단서들을 이용하여 주요 객체에 대한 초기 영역을 결

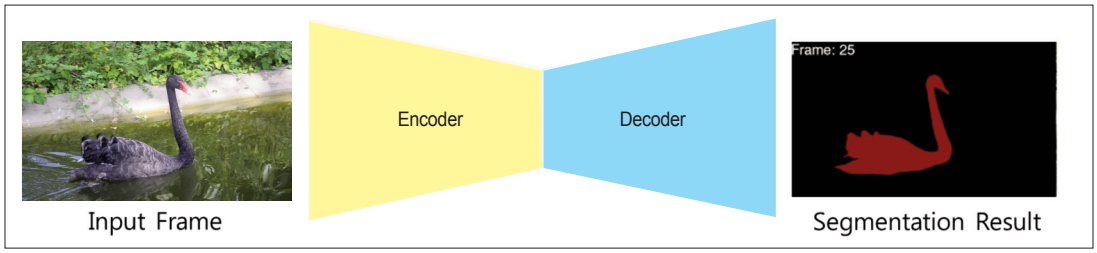


<그림 1> 지도 수준에 따라 분류한 동영상 객체 분할 기법: (a) 비지도 동영상 객체 분할, (b) 준지도 동영상 객체 분할, (c) 인터렉티브 동영상 객체 분할

정한 후, 초기 영역의 픽셀 정보들을 이용하여 객체 모델 및 배경 모델을 구축하고, Markov random field와 같은 최적화 방식을 이용하여 동영상 내 각 픽셀의 클래스를 주요 객체 또는 배경으로 이진화 하였다.

2017년부터 영상 인식 분야에서 뛰어난 성능 향상을 입증한 심층신경망 기술을 비지도 동영상 객체 분할에 적용하려는 시도가 시작되었다. 기본적으로 〈그림 1〉과 같이 동영상의 각 프레임 영상을 신경망에 입력하면, 주요 객체에 대한 확률 영상이 출력되는 인코더-디코더(encoder-decoder) 구조를 기반으로 하고 있다. 동영상에서 주요 객체의 움직임 정보를 활용하기 위해 심층신경망의 인코더는 RGB 색상 영상과 인접한 두 프레임에서 예측한 광흐름(optical flow)의 크기 영상을 모두 수용하는 멀티 스트림 신경망 구조를 가지고 있다[4,5]. 색상과 움직임의 2 종류의 영상으로부터 주요 객체 영역을 분할하기 위한 의미 있는 특징을 효과적으로 추출한 후, 디코더에서는 정의된 손실 함숫값을 최소화하도록 특징으로부터 주요 객체 확률 맵을 출력한다. 디코더에는 일반적으로 U-Net 구조와 같이 영상 해상도가 작아지는 과정에서 잃어버리는 영상 디테일을 복원하기 위해 Skip connection이 적용된다. 광흐름 기법을 통해 예측한 움직임 정보를 이용하여, 독특한 움직임을 갖는 주요 객체의 영역에 대한 분할 정확도를 높일 수 있으나, 정지하고 있는 주요 객체에 대한 분할 정확도는 떨어지는 문제가 발생한다.

이 문제를 극복하기 위해 최근 비지도 동영상 객체 분할 기법들은[6,7] 움직임 정보에 의존하지 않고, 주요 객체가 동영상에서 매우 빈번하게 출현한다는 성질을 이용하여 프레임간 정합(matching)을 통해 주요 객체의 확률 맵을 예측하였다. 구체적으로, 인코더로부터 임의로 선택된 두 프레임에 대한 특징 맵을 추출한 후, 같은 위치의 특징 벡터간 내적이 계산되도록 행렬 연산을 수행함으로써 두 프레임간 정합 결과를 획득할 수 있다. 정합을 수행하는 두 프레임이 동일한 주요 객체를 포함하고 있다면, 정합 결과에서 주요 객체에 해당하는 위치에 높은 반응 값을 보인다. 최종적으로, 프레임간 정합 결과를 프레임 특징 맵에 연결한(concatenate) 후, 디코더를 통해 주요 객체 확률 맵을 출력함으로써, 두 프레임에 모두 존재하는 주요 객체에 대한 정확한 픽셀 영역을 분할할 수 있다. 프레임간 정합에 기반한 비지도 동영상 객체 분할 기술에는 단순히 두 프레임간 정합을 통해 주요 객체 분할을 수행하거나[6], 한 프레임의 주요 객체 분할 결과를 예측하기 위해 여러 프레임을 이용하기도 한다[7]. 또



<그림 2> 심층신경망 기반 동영상 객체 분할의 인코더-디코더 구조

한, 동영상 내 주요 객체는 사람의 시선을 많이 받는 성질을 이용하여 사람의 시선 추적 정보를 활용한 비지도 동영상 객체 분할 기술도 존재한다[8]. 이와 같은 비지도 동영상 객체 분할 기법들은 동영상 내 단일 주요 객체에 대한 분할 능력은 뛰어나지만, 다수의 주요 객체 분할에 대처할 수 없는 문제가 있다.

2. 준지도 동영상 객체 분할

준지도 동영상 객체 분할은 첫 프레임에서 사용자가 제공한 타겟 객체에 대한 정확한 분할 영역을 이용하여 이 후 프레임에서의 타겟 객체 분할을 수행한다. 첫 프레임에서 타겟 객체에 대한 정보가 주어지기 때문에, 이를 활용한 심층신경망 학습이 용이하여, 최근 동영상 객체 분할의 세 카테고리 중 가장 많은 논문이 발표되고 있다. 이러한 준지도 동영상 객체 분할 기술에서 사용되는 심층신경망 구조는 기본적으로 비지도 기술과 같이 인코더-디코더 구조를 갖는다. 그러나, 준지도 동영상 객체 분할 기법들은 비지도 방식과는 달리, 심층신경망 학습을 위해 첫 프레임에서 제공되는 타겟 객체에 대한 분할 마스크 정보를 다양한 방법으로 이용한다.

준지도 동영상 객체 분할을 위해 첫 프레임의 타겟 객체 분할 영역을 이용한 심층신경망의 fine-tuning 방식이 처음으로 시도되었다[9]. 심층신경망을 fine-tuning한 후, 이 후 프레임에서 fine-tuning한 심층신경망을 이용하여 타겟 객체에 대한 확률 맵을 출력하는 방법이다. 첫 프레임에서 주어지는 단 한 장의 타겟 객체 분할 정보로는 심층신경망을 학습하기에 부족하기 때문에 데이터 증강을 통해 데이터 부족 문제를 해결하려고 하였다. 그러나, 이러한 fine-tuning 방식은 데이터 부족과 첫 번째 프레임과 멀어질수록 변형되는 타겟 객체에 대응하지 못하는 문제가 있어 성능의 한계를 보인다.

이전 프레임의 타겟 객체 분할 결과를 보정(refine)하여 현재 프레임의 결과를 예측하는 방식들이 제안되었다[10,11,12,13]. 보정하기 앞서 이전 프레임의 결과를 현재 프레임으로 전파(propagation)시키기 위해 광흐름이나 객체 추적 등의 움직임 정보에 활용되었다[10,11]. 움직임 정보가 부정확한 경우, 전파된 타겟 객체 분할 결과가 부정확할 수 있으며, 이를 극복하기 위해 전파된 분할 영역을 보정(refine)하는 심층신경망이 개발되었다[10]. 첫 프레임 영상과 분할 영역을 가이드 정보로 활용하여, 현재 프레임의 전파 결과를 보정하도록 심층신경망을 학습함으로써, 프레임이 진행됨에 따라 타겟 객체 분할의 정확도가 떨어지는 문제를 줄일 수 있었다[11]. 전파 방식이 움직임 정보에 의존적인 것을 극복하기 위해, 분할 영역 전파를 수행하지 않고, 이전 프레임의 결과로부터 현재 프레임의 타겟 객체 분할 영역을 예측하는 방식이 제안되었다[12,13]. 이 방식에 기반한 심층신경망은 현재 프레임을 입력으로 받는 인코더와 첫 프레임과 타겟 객체 마스크를 입력으로 받는 인코더로 구성된 Siamese 구조를 이루고 있다. 두 인코더로부터 추출된 특징 정보들을 연결한 후, 디코더를 통해 현재 프레임의 타겟 객체 분할 영역을 획득한다. 이러한 움직임 정보 없이 보정만으로 현재 프레임의 분할 영역을 예측하는 방법들은[12,13] 빠른 속도와 높은 정확도를 보인다.

최근에는 embedding space에서 타겟 객체와 유사한 특징을 갖는 픽셀들을 분할 결과로 도출하는 방식들이 시도되고 있다[14,15,16]. PML[14] 기법은 인코더를 통해 첫 프레임과 현재 프레임의 특징

맵을 추출한 후, 현재 프레임의 각 픽셀과 첫 프레임의 타겟 객체 영역과 배경 영역과 특징 거리를 측정한 후, 최근접 이웃 분류(nearest neighbor classifier)를 통해 각 픽셀이 타겟 객체인지 배경인지 결정한다. 즉, PML 기법은 첫 프레임의 타겟 객체 영역과의 정확한 거리 측정이 가능하도록 embedding space를 학습함으로써 현재 프레임의 각 픽셀의 클래스 예측의 정확도를 향상시켰다. FEELVOS[15] 기법 embedding space에서 첫 프레임과 현재 프레임과의 거리를 측정하는 전역적 정합(global matching) 방식에 인접한 프레임 간의 지역적 정보를 추가적으로 고려하기 위한 지역적 정합(local matching)을 수행하였다. 전역적 정합과 지역적 정합 결과를 모두 활용함으로써 기존 embedding space 기반 정합 방법보다 우수한 결과를 보였다. 최근에는 동영상 내 여러 프레임들의 정합 반응 값을 고려하는 비지역적(non-local) 방식[17]을 활용한 준지도 동영상 객체 분할 기법이 제안되었다[16]. 기존 embedding space 방식의 기술들이 두개 혹은 세 프레임 사이의 정합만을 수행했던 반면, [16] 기법은 non-local 신경망[17] 기술을 채용하여 여러 프레임의 정합 결과를 효율적으로 활용하는 심층신경망을 구축하였다. 그 결과로 속도 대비 가장 우수한 결과를 도출하였고, 현재 가장 높은 성능을 보이고 있다.

이 밖에도 준지도 동영상 객체 기법에는 각 프레임의 객체 검출 기법을 활용하거나[18,19], 동영상 내 프레임들의 정보들을 기억하기 위한 순환신경망(recurrent neural network)을 활용한 심층신경망 구조들이 제안되었다[20,21].

Ⅲ. 동영상 객체 분할 성능 비교

이번 장에서는 다양한 동영상 객체 분할 기법들에 대한 성능을 분석하고자 한다. 동영상 객체 분할에 널리 사용되는 데이터셋으로 DAVIS[22] 데이터셋이 있다. DAVIS 데이터셋에는 동영상마다 단일 객체만을 분할 대상으로 여기는 DAVIS2016 데이터셋과 다수 객체들을 목표로 하는 DAVIS 2017 데이터셋을 포함하고 있다. 본 고에서는 DAVIS 데이터셋의 DAIVS2016, DAVIS2017 데이터셋들을 이용하여 동영상 객체 분할 기법들의 성능을 비교한다. 성능 평가를 위해 사용되는 메트릭은 대표적으로 분할 영역 유사도(Intersection-over-Union)를 측정하는 J score와 객체 경계선 정확도를 측정하는 F score가 있다. 〈표 1〉은 비지도 동영상 객체 분할 기술들의 DAVIS-2016 데이터셋에 대한 성능을 비교한다. 비지도 방식은 단일 주요 객체 분할을 목표로 하기 때문에 대부분의 비지도 동영상 객체 분할 기법들은 DAVIS-2016 데이터셋에 대한 결과를 제시하는 상황이며, 최근 몇 편의 논문[8,21]이 다수의 주요 객체가 존재하는 DAVIS-2017 데이터셋에 대한 성능을 제공한다. 〈표 1〉의 FSEG[4]는

<표 1> 비지도 동영상 객체 분할 기법들의 DAVIS-2016 데이터셋에 대한 성능 비교

메트릭	FST[1]	FSEG[4]	MG[5]	COSNET[6]	AnDiff[7]	AGS[8]	MATNet[22]
J score	55.8	70.7	81.4	80.5	81.7	79.7	82.4
F score	51.1	65.3	81.0	79.5	80.5	77.4	80.7

심층신경망이 처음 도입함으로써, 심층신경망이 도입되기 이전 기술인 FST[1]에 비해 14.9 J score의 비약적인 성능 향상을 이끌었다. 그 후, 심층신경망이 더욱 발전하고, 학습에 사용할 동영상 객체 분할을 위한 데이터셋이 증가됨에 따라 성능이 점차 향상되어 최근에는 J score기준 82.4의 성능까지 도달하였다. 현재 높은 성능을 보이는 대부분의 기술들은[6,7,8,22] 프레임 간 정합 방식에 기반하고 있는 추세다.

〈표 2〉는 본 고에서 살펴본 기술들의 DAVIS-2016과 DAVIS-2017 데이터셋에 대한 J score, F score, 프레임당 소요되는 시간(frame per seconds)을 비교한다. 심층신경망이 도입된 초기 논문들은 동영상마다 단일 타겟 객체만을 분할하는 DVAIS-2016 데이터셋에 대해서만 실험 결과를 제공한다. 2017년 DAVIS-2017 데이터셋이 제공된 후, 2018년 이후부터 다수의 타겟 객체에도 분할이 가능한 심층신경망 기술들이 제안되었다. OSVOS[9] 기술은 첫 프레임의 타겟 객체 분할 영역 정보를 기반으로 심층신경망의 fine-tuning을 수행하기 때문에 많은 시간이 소요된다. 또한, 전파 없이 분할 영역 보정을 수행하는 기술들이[12,13] 전파 방식의 방식과[10,11] 비교했을 때 움직임 예측을 필요로하지 않기 때문에 더 빠른 처리 속도를 보이는 것을 확인할 수 있다. 또한 embedding space에서의 정합 방식들이[14,15,16] 속도 대비 높은 정확도를 보인다. STM[16] 기술은 현재 준지도 동영상 객체 분할 기술들 중 가장 좋은 성능을 보이고 있으며, 프레임당 0.16초의 속도를 보이고 있어, 심층신경망 연산량도 적다. 특히, DAVIS-2017 데이터셋에서 비약적인 성능 향상을 이끌어 J score 기준 80에 가까운 정확도를 보이고 있다. PReMVOS[18] 기술은 2019년도 중반까지 가장 좋은 정확도를 보였으나, 높은 연산량으로 인해 많은 시간이 소요되어 실효성이 떨어진다. 순환신경망 기술이 적용된 YouTube[20] 기술은 낮은 정확도를 보이고 있어, 동영상 객체 분할을 위한 최적의 순환신경망 기술이 아직 개발되지 않았다고 추정할 수 있다.

<표 2> 준지도 동영상 객체 분할 기법들의 DAVIS-2016과 DAVIS-2017에 대한 성능 비교

기법	DAVIS-2016			DAVIS-2017	
	J score	F score	시간/프레임	J score	F score
OSVOS[9]	79.8	80.6	9	-	-
CTN[10]	73.5	69.3	1.5	-	-
MGCR[11]	84.4	85.7	0.73	-	-
RGMP[12]	81.5	82.0	0.13	64.8	68.6
AGSS[13]	-	-	<0.10	64.9	69.9
PML[14]	75.5	79.3	0.28	-	-
FEELVOS[15]	81.1	82.2	0.45	69.1	74.0
STM[16]	88.7	89.9	0.16	79.2	84.3
PReMVOS[18]	84.9	88.6	>30	73.9	81.7
MHP-VOS[19]	85.7	88.1	-	71.8	78.8
YouTube[20]	79.1	-	9	-	-

Ⅳ. 결 론

본 고에서는 심층신경망을 이용한 비지도 동영상 객체 분할 기술과 준지도 동영상 객체 분할의 최신 동향에 대해 살펴보았다. 비지도 동영상 객체 분할의 경우, 어떠한 사용자 정보 없이 동영상 내 주요 객체를 분할하기 위해, 움직임 정보를 활용한 심층신경망 기술이 개발되었고, 최근에는 프레임간 정합을 통해 반복적으로 출현하는 주요 객체 영역을 분할하려는 시도가 활발히 수행되고 있다. 준지도 동영상 객체 분할을 위해 fine-tuning, 분할 영역 전파, 분할 영역 보정(refine), embedding space에서의 프레임 정합, 객체 검출, 순환신경망 등의 다양한 기술들이 개발되고 있으며, 현재 프레임 정합 방식이 높은 정확도와 빠른 연산 속도를 보이고 있는 실정이다.

참 고 문 헌

[1] A. papazoglou and V. Ferrari, “Fast Object Segmentation in Unconstrained Video,” ICCV, 2013.

[2] Y. J. Koh and C.-S. Kim, “Primary Object Segmentation in Videos Based on Region Augmentation and Reduction,” CVPR, 2017.

[3] A. Faktor and M. Irani, “Video Segmentation by Non-Local Consensus voting,” BMVC, 2014.

[4] S. Jain, B. Xiong, and K. Grauman, “FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos,” CVPR, 2017

[5] H. Li, G. Chen, G. Li, and Y. Yu, “Motion Guided Attention for Video Salient Object Detection,” ICCV, 2019.

[6] X. Lu, W. Wang, C. Ma, J. Shen, L. Shao, and F. Porikli, “See More, Know More: Unsupervised Video Object Segmentation with Co-Attention Siamese Networks,” CVPR, 2019.

[7] Z. Yang, Q. Wang, L. Bertinetto, W. Hu, S. Bai, and P. H.S. Torr, “Anchor Diffusion for Unsupervised Video Object Segmentation,” ICCV, 2019.

[8] W. Wang, H. Song, S. Zhao, J. Shen, S. Zhao, S. C. H. Hoi, and H. Ling, “Learning Unsupervised Video Object Segmentation through Visual Attention,” CVPR, 2019.

[9] S. Caelles, K.K. Maninis, J. Pont-Tuset, L. Leal-Taixe, D. Cremers, and L. Van Gool, “One-Shot Video Object Segmentation,” CVPR, 2017.

[10] W.-D. Jang and C.-S Kim, “Online Video Object Segmentation via Convolutional Trident Network,” CVPR, 2017.

[11] P. Hu, G. Wang, X. Kong, J. Kuen, and T.-P. Tan “Motion-Guided Cascaded Refinement Network for Video Object Segmentation,” CVPR, 2018.

[12] S. W. Oh, J.-Y. Lee, K. Sunkavalli, S. J. Kim, “RGMP: Fast video object segmentation by reference-guided mask propagation,” CVPR, 2018.

[13] H. Lin, X. Qi, and J. Jia, “AGSS-VOS: Attention Guided Single-Shot Video Object Segmentation,” ICCV, 2019.

[14] Y. Chen, J. Pong-Tuset, A. Montes, and L. Van Gool, “PML: Blazingly Fast Video Object Segmentation with Pixel-Wise Metric Learning,” CVPR, 2018.

[15] P. Voigtlaender, Y. Chai, F. Schroff, H. Adam, B. Leibe, and L.-C. Chen, “FEELVOS: Fast End-To-End Embedding Learning for Video Object Segmentation,” CVPR, 2019.

[16] S. W. Oh, J.-Y. Lee, N. Xu, S. J. Lim, “Video Object Segmentation using Space-Time Memory Networks,” ICCV, 2019.

[17] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local Neural Networks”, CVPR, 2018.

[18] J. Luiten, P. Voigtlaender, and B. Leibe, “PReMVOS: Proposal-generation, Refinement and Merging for Video Object Segmentation,” ACCV, 2018.

[19] S. Xu, D. Liu, L. Bao, W. Liu, and P. Zhou, “MHP-VOS: Multiple Hypotheses Propagation for Video Object Segmentation,” CVPR, 2019.
[20] N. Xu, L. Yang, Y. Fan, J. Yang, D. Yue, Y. Liang, B. Price, S. Cohen, and T. Huang, “YouTube-VOS: Sequence-to-Sequence Video Object Segmentation,” ECCV, 2018.
[21] C. Ventura, M. Bellver, A. Girbau, A. Salvador, “RVOS: End-to-End Recurrent Network for Video Object Segmentation,” CVPR, 2019.
[22] T. Zhou, S. Wang, Y. Zhou, Y. Yao, J. Li, L. Shao, “Motion-Attentive Transition for Zero-Shot Video Object Segmentation, AAAI 2020,” AAAI, 2020.

필 자 소 개

고 영 준

– 2018년 2월 : 고려대학교 전기전자과 박사
– 2018년 2월 ~ 2018년 9월 : 고려대학교 박사후연구원
– 2018년 10월 ~ 2019년 1월 : UCSD 박사후연구원
– 2019년 3월 ~ 현재 : 충남대학교 컴퓨터융합학부 조교수
– 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습