**2. Related Work.**

**2-1. Key Frame Extraction.**

동영상에서 하이라이트 이미지를 추출해내기 위한 방법을 제시하는 논문. Key Frame Extraction 방법으로 6가지 종류가 있으며, 해당 논문은 clustering을 사용한 방법을 제시하는데 구현이 쉽고 계산 속도가 빨라 실시간으로 적용할 수 있다는 이점이 있다.

Key Frame Extraction하는 과정은 다음과 같다. Global Level Histogram을 사용하여, 프레임 간의 색상 유사도를 측정하고, GLCM(Grey-Level Co-occurrence Matrix)을 사용하여 프레임 간의 텍스처(질감) 유사도를 측정한다.

두 유사도를 가중치를 곱하여 결합, 최종 유사도 값을 얻는다.

최종 유사도 값을 사용하여 Clustering을 진행하는데, 이때 임계 값 을 사용하여 각 Cluster의 밀집도를 제어한다. 가 높을수록 Cluster의 개수가 많아진다. 최종 유사도 값이 임계 값보다 낮다는 것은 이는 기존 Cluster에 추가되기에는 유사도가 낮다는 것을 의미하므로, 새로운 Cluster가 생성된다.

Clustering이 끝나면, 각 Cluster에서 대표하는 Key Frame을 선택한다. 해당 논문에서는 Cluster가 가진 Frame 수가 전체 Frame 수의 10%를 넘으면 해당 Cluster가 충분히 큰 Cluster라고 판단하여 해당 Cluster의 Key Frame을 선택한다. 해당 Cluster에서 유사한 이미지의 와 의 중심에 가장 가까운 Frame을 Key Frame으로 선택한다.

**2-2. InstaGAN**

기존의 GAN 모델은 이미지 내의 다수 인스턴스를 변화시킬 때, 혹은 이미지 내의 인스턴스의 모양을 변화시킬 때 낮은 성능을 보이는 단점을 가지고 있다. 대표적인 예로, 말이 존재하는 이미지 내에서 말을 얼룩말로 변환하는 경우 기존 GAN으로 충분히 가능하였으나, 말을 기린으로 변환하는 경우에는 기존 GAN이 제대로 변환하지 못하고, 이미지가 깨지는 결과를 보인다.

이에 대한 해결책으로 해당 논문에서는 이미지에서 변환하고자 하는 인스턴스의 Segmentation Mask를 사용하여 인스턴스의 모양을 변화시키는 성능을 높였다.

기존의 GAN 모델은 생성자와 판별자를 사용하여 생성자를 사용한 Fake Image와 Real Image를 판별자가 Real Image를 판별하게 하여, 생성자는 원하는 결과물을 생성해내도록, 판별자는 Fake Image와 Real Image를 더 잘 분별할 수 있도록 학습한다.

InstaGAN에서는 이러한 기존의 GAN 모델에서 발전하여, 해당 생성자와 판별자를 학습하는데 하나의 Input이 아닌 Input 쌍을 사용한다. 실제 이미지와 해당 이미지에서 변화하고자 하는 인스턴스의 Segmentation Mask 이미지를 쌍으로 하여 학습에 사용한다.

즉 기존의 GAN 모델의 생성자와 판별자가 다음과 같았다면,

InstaGAN 모델의 생성자와 판별자는 다음과 같다.

이를 이용하여 손실함수를 정의하면 다음과 같다.

는 기존 GAN에서 사용한 손실함수와 동일한 식으로 판별자가 이미지를 판별한 결과와 실제 결과의 차이를 오차 제곱합으로 구하는 식이다.

는 각각 생성자를 두 번, 한 번 매핑하여 나온 결과를 실제 결과의 차이를 구하는 식으로, 는 을 통해 와 의 차이를 나타내고, 는 을 통해 와 의 차이를 나타낸다.

는 Context를 얼마나 보존했는가를 나타낸 손실 함수로, Segmentation Mask Image에서 각 픽셀이 배경(Background)을 나타낼 땐 1로, 인스턴스를 나타낼 땐 0으로 설정하여 이진 값으로 나타내면, 생성자를 통해 원하는 이미지로 변경하면, 인스턴스의 모양이 변화하여 각 픽셀 값의 변화도 생길 것이다. 이를 이용하여 얼마나 Context를 보존했는가를 나타낸다. 는 각 픽셀 이미지에서 최솟값을 1에서 뺀 것으로, 동일한 위치의 픽셀이 둘 다 1이라면 0, 아니라면 1일 것이다. 연산자는 각 행렬의 성분 곱으로 같은 크기의 행렬을 같은 위치의 성분끼리 곱하는 연산자이다.

위의 손실 함수들을 종합하여 최종 을 생성하고 이를 최소화하도록 학습하여 최종적으로 높은 성능의 인스턴스 모양 변화 모델을 구현하는 것이 InstaGAN의 최종 목표이다.

**2-3. CartoonGAN**

CartoonGAN은 실제 현실 이미지를 만화 스타일의 이미지로 변환하는 것을 목표로 하는 모델로, GAN 모델을 기초로 한다.

만화 스타일의 이미지는 간단하게 표현되며, 추상적인 특징을 가지고, 클리어한 윤곽선과 부드러운 질감을 갖는 것이 특징이다. 그러나 현재 널리 사용되는 GAN 모델의 손실함수로는 이를 처리하기 힘들기 때문에 기존 GAN 모델은 만화 스타일로의 변화에 좋은 성능을 보여주지 못한다.

해당 논문은 이를 위해 기존 GAN이 사용하는 손실함수에 기초를 두어 새로운 손실함수를 정의했는데 이는 다음과 같다.

는 기존 GAN 모델의 손실함수인 Adversarial Loss에 새로운 오차 계산식을 추가한 식이다. 부분을 살펴보면 는 학습을 위한 이미지 데이터 에서 윤곽선을 제거한 이미지로 만화 스타일의 이미지에서 윤곽선은 매우 중요한 특성을 나타내지만, 이미지 전체에서 윤곽선이 차지하는 비율은 매우 적기 때문에 기존 GAN 모델은 이를 정확하게 학습하지 못하였다. 이를 해결하기 위해 손실함수를 위처럼 변형하여 생성자가 윤곽선의 비중을 높이도록 학습하게 하였다.

는 실제 이미지를 만화 스타일의 이미지로 변환 시에 이미지를 충분히 보존하면서 스타일을 변형하도록 하는 손실함수로 이를 위해 사전에 학습된 VGG Network를 사용하여 얼마나 보존되었는지 차이를 구할 수 있었다.

추가적으로, 기존 GAN 모델은 비선형이기 때문에, 무작위로 초기화할 경우 지역 최소점에 갇혀 원하는 방향으로 학습되지 않을 가능성이 크다. 이를 방지하기 위해, CartoonGAN에서는 semantic content loss, 을 사용하여 사전에 생성자를 학습하고, 이것이 결과에 유의미한 차이가 있음을 보였다.