

# 1. 서론

## 1) 개요

본 문서는 Vision TransGAN 기술을 도입한 Video Inpainting 이 포함된 동영상 편집 프로그램 웹 서비스(VIVE : Video inpainting Video Editor) 개발에 대한 상세 요구사항을 명세하는 문서다.

## 2) 목적

기억하고 싶은 순간을 동영상으로 촬영했는데 원치 않는 사물이나 사람이 같이 찍히는 경우가 있다. 그러한 물체를 제거하고 변형된 이미지를 제공해 편집하는 시간 단축, 동영상에 대한 만족도, 초상권 문제 회피 그리고 동영상 결함 복구로 영상의 품질을 향상시키는 것에 목적을 두고 있다. 핵심 기술인 Video Inpainting 에 경우 비디오 품질 평가하는데 사용하는 지표인 PSNR, SSIM 을 기준으로 기술을 이해하기 위해 참고한 논문(Deep Video Inpainting)에서 발표한 ViNet 모델보다 높은 성능(DAVIS dataset 기준 PSNR : 28.96, SSIM : 0.9411)을 넘는 걸 목표로 한다.

## 3) 개발 의의

새로운 기술 Video Inpainting 을 제공하는 것뿐만 아니라 기본적인 편집기술을 제공해 현 동영상 편집 시장에서 경쟁력을 가지고 추후에 다른 기술들을 적용할 기반을 만들어 서비스에 대한 발전 가능성을 열어 둔다.

## 4) 대상

우리는 대표적인 무료 동영상 편집 프로그램을 대체해 해당 서비스를 사용하는 사람들과 Video Inpainting 기술에 대한 흥미를 가지거나 해당 기술을 필요로 하는 사람들을 대상으로 한다.

## 5)참고

[1] Dahun Kim, et al., "Deep Video Inpainting", arXiv:1905.01639,  
<https://arxiv.org/abs/1905.01639>

[2] Kwanyoung Park, et al., "Per-Clip Video Object Segmentation",  
arXiv:2208.01924, <https://arxiv.org/abs/2208.01924>

## 2. 기능

### 2.1 시스템 인터페이스

서비스는 프로그램으로 배포하고 프로그램을 다운 받은 경우 Inpainting 의 최저 사양인 그래픽카드 RTX 2060 이상이 탑재된 환경에서 사용 가능하다. Pyqt 를 활용해 프로그램의 UI 를 제작하고 OpenCV, Moviepy 를 이용하여 기본 편집 기능들(동영상 불러오기, 저장, 나누기, 오디오, 자막 추가)을 제공하며 Per-Clip 을 활용해 Tracking & Masking 을 작업해 원하는 물체를 Masking 하고 Torch 의 환경에서 학습된 CNN 모델을 이용하여 Inpainting 을통하여 기본 편집 기능과 다른 기능을 제공한다.

### 2.2 사용자 인터페이스

사용자는 Inpainting 을 쓸 때 원하는 물체를 선택하여 지워졌는지 확인한다. 그 후 기본 편집 기능들을 활용해 원하는 방향으로 영상을 편집한다. 이후 편집이 완료된 영상을 결과로 받을 수 있다.

### 2.3 제약 조건

- PyQt 로 UI 를 설계한다.
- 하나의 화면에 구현 가능한 기술 버튼들을 제공하여야 한다.
- 추후 다른 기능을 추가할 수 있는 확장성을 고려한다.
- Data set 은 DAVIS 에서 제공하는 것으로 training 을 한다.

## 2.4 학습 모델 및 기능

### 2.4.1 Tracking & Masking

삭제하고자 하는 물체가 나오는 프레임에서 직접 색칠하면서 Masking 을 하면 Per-Clip 기술을 사용해 그 물체가 나오는 프레임들을 Tracking 해서 Masking 을 해준다. 이때 각 프레임별 이미지들이 Masking 이 된 상태로 Output 이 된다.

### 2.4.2 Inpainting

Deep video Inpainting 논문에서 사용된 ViNet 모델은 CNN 을 기반으로 영상 프레임의 Feature 를 파악하여 Inpainting 을 진행하였다. 그러나 현재 컴퓨터 비전 분야는 Vision Transformer(이하 ViT)의 발명으로 CNN 을 능가하는 성능을 보이고 있다. 우리 또한 Inpainting 모델에 ViT 를 사용하여 동영상 복원 성능을 더욱 높이고자 한다.

우리의 모델은 시계열 데이터인 동영상 프레임 사이의 정보를 파악하기 위해, 여러 개의 프레임을 패치 단위로 나누어 ViT 모델에 입력하고, 이미지 생성 모델인 GAN 을 활용하여 복원된 동영상을 생성한다. GAN 은 크게 생성기와 판별기로 구성되어, 서로가 경쟁하며 스스로를 훈련시킨다. Inpainting module 의 개괄적인 훈련 프로세스는 다음과 같다. 우선 마스크에 의해 손상된 이미지가 Input Data 로 들어가고, GAN 의 생성기는 해당 데이터와 원본 이미지 사이의 데이터 오차를 줄이는 방향으로 학습을 진행한다. 생성기가 모델을 학습하는 동안, 판별기는 생성기에 의해 합성된 이미지와 원본 이미지를 판별하는 역할을 한다. 이 때, 판별기가 만들어진 이미지와 원본 이미지를 더 이상 구별할 수 없을 정도로 모델이 훈련되었을 때 트레이닝을 종료한다.

이후 업데이트 된 가중치(Weight)를 .pth 파일로 저장한 뒤, 새로운 동영상 데이터에서 인페인팅을 진행하여 영상 내의 물체를 지울 수 있다.

### 3. 요구사항

ID	정의
DT	Deep learning – Technology
BOT	Based - OpenCV – Technology

#### 1) 기능 요구사항

##### (1) 편집 동영상&사진 불러오기

기능	편집 동영상&사진 불러오기
ID	BOT_001
설명	편집할 동영상 또는 사진 파일을 불러온다.
상세 요구사항	1. 동영상 플레이어를 누르면 파일 탐색기 실행 2. 파일 탐색기에서 원하는 파일을 선택 시 불러오기 실행 3. 트랙 파일 목록에 추가
핵심 기술	- Python - OpenCV - PyQt

(2) 타임라인 구현

기능	타임라인 구현
ID	BOT_002
설명	편집하고 있는 파일들을 보기와 순서 이동을 편하게 만든다.
상세 요구사항	1. 파일을 불러왔을 때 타임라인에 파일 추가 2. 현재 편집 상태 표현
핵심 기술	- PyQt - OpenCV

(3) 편집 동영상 플레이어 제공

기능	동영상 플레이어 제공
ID	BOT_003
설명	편집 도중 영상 확인용 플레이어
상세 요구사항	1. 편집 도중 트랙 시간에 맞춰 재생되는 플레이어 2. 오디오까지 제공
핵심 기술	- OpenCV

(4) 동영상 나누기

기능	동영상 나누기
ID	BOT_004
설명	동영상에서 원하는 부분을 나누어 관리
상세 요구사항	1. 원하는 부분에서 동영상을 구간으로 나누어 편집 가능 2. 타임라인에 동영상 구간이 구분됨
핵심 기술	- OpenCV

(5) 오디오 추가

기능	오디오 추가
ID	BOT_005
설명	동영상이 재생될 때 같이 재생될 오디오 추가
상세 요구사항	1. 불러오기 버튼을 누르면 파일 탐색기 실행 2. 파일 탐색기에서 원하는 음악 파일을 선택 시 불러오기 실행 3. 트랙 파일 목록에 추가
핵심 기술	- moviepy - OpenCV

(6) 자막 추가

기능	자막 추가
ID	BOT_006
설명	동영상이 재생될 때 위에
상세 요구사항	1. 자막 추가 2. 자막 글꼴/색상 변경 3. 트랙에 추가 4. 띄우는 시간 조정
핵심 기술	- OpenCV

(7) 객체 Masking

기능	객체 Masking
IDB	BOT_ 007
설명	Tracking 을 할 객체가 나오는 첫 프레임을 색칠해 Masking 한다.
상세 요구사항	1. OpenCV 를 통해 이미지를 불러와서 색칠이 가능하도록 함 2. 지울 객체가 다 색칠되어야 함
핵심 기술	- OpenCV

(8) Tracking

기능	Tracking
ID	DT_001
설명	객체 Masking 에서 나온 Output 을 Input 으로 받아 해당 객체를 tracking 한다.
상세 요구사항	1. 객체 Masking 에서 나온 Output 을 Input 으로 입력한다. 2. 시간에 따라 물체가 이동하면 앞뒤로 따라간다. 3. 동영상을 프레임 마다 사진 파일로 저장한다.
핵심 기술	- Per-Clip - OpenCV

(9) Masking

기능	Masking
ID	DT_002
설명	Tracking 물체를 검정색으로 Masking 한다.
상세 요구사항	1. Tracking 을 하는 것으로 지정된 물체를 검정색으로 가림 2. 해당 동영상 물체를 Masking 하는 파일 생성 3. 동영상의 프레임 마다 Masking 된 사진 파일을 저장한다.
핵심 기술	- Per-Clip

(10) Video Inpainting (train)

기능	Video Inpainting (train)
ID	DT_003
설명	Inpainting 을 학습한 .pth 파일을 생성한다.
상세 요구사항	1. 선정된 Inpainting 의 학습 방법에 따라 weight 값이 계산된 .pth 파일을 생성한다.
핵심 기술	- Vision transfer



(11) Video Inpainting

기능	Video Inpainting
ID	DT_004
설명	Masking 으로 가려진 이미지들을 채워 넣는다. 시간에 따라 서로 연관되게 (학습을 평가할 때 서로를 평가 항목에 추가)
상세 요구사항	1. Tracking 과 Masking 을 걸쳐 만들어진 이미지를 채워 넣는다. 2. 동영상이기에 시간에 따라 서로 연관됨을 활용 3. train 을 평가할 때 서로를 평가 항목에 추가 4. 기술의 성능은 PSNR, SSIM 지표로 확인
핵심 기술	- Vision transfer

(12) Video Inpainting (Valid)

기능	Video Inpainting (Valid)
ID	DT_005
설명	성능 지표 확인
상세 요구사항	1. PSNR, SSIM 에 대한지표 확인할 수 있게 표현 2. 원본 이미지와 복원된 이미지의 차이(MSE)를 측정 3. 영상의 밝기, 대비 및 구조를 고려 유사성 측정
핵심 기술	- Pytorch, PSNR, SSIM

## 2) 비기능 요구사항

### (1) 시스템 장비 요구사항

기능	Inpainting 작동 최소사양
설명	Inpainting 이 돌아갈 최소사양 필요
상세 요구사항	1. RTX 2060 이상의 그래픽카드 필요 2. CPU i7 이상 필요
핵심 기술	- 학습된 .pth 파일과 video 를 계산 가능한 그래픽카드 탑재

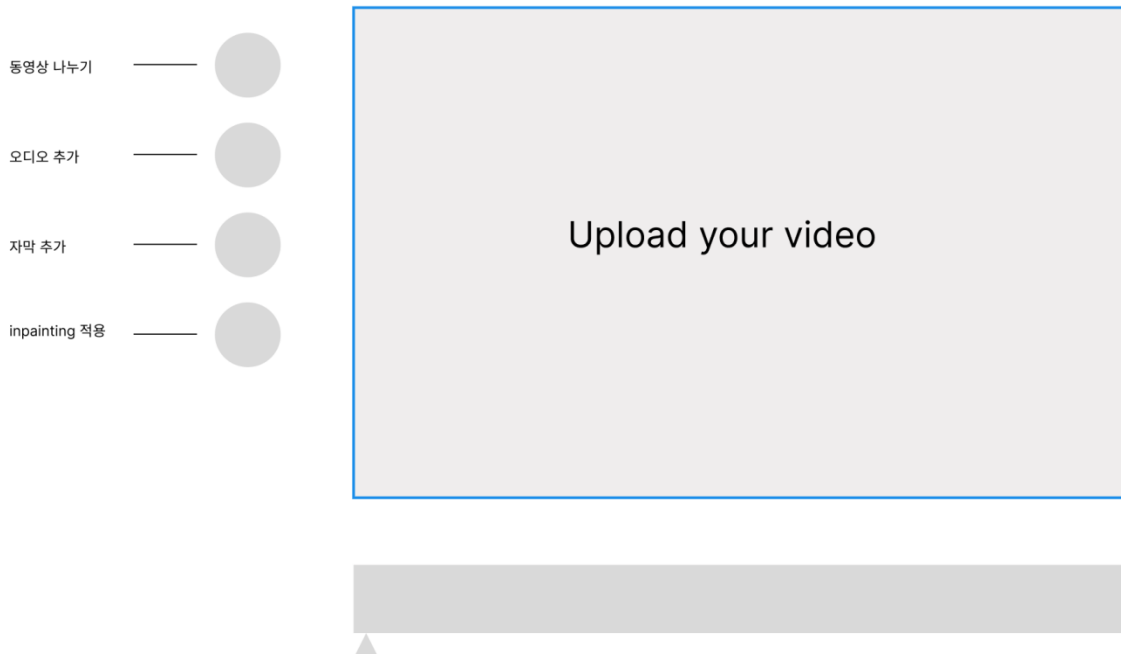
### (2) 성능 요구사항

기능	처리 속도
설명	Inpainting 편집 장면 처리 속도 제한
상세 요구사항	1. 편집 장면 1 초당 10 초 미만 요구 2. Inpainting 사용 후 동영상 편집
핵심 기술	- 그래픽카드에 따라 처리 속도가 변한다. - 연산 값을 최대한 낮춰야 한다.

### 3) 사용자 인터페이스 요구사항

#### 3.1) 메인 화면

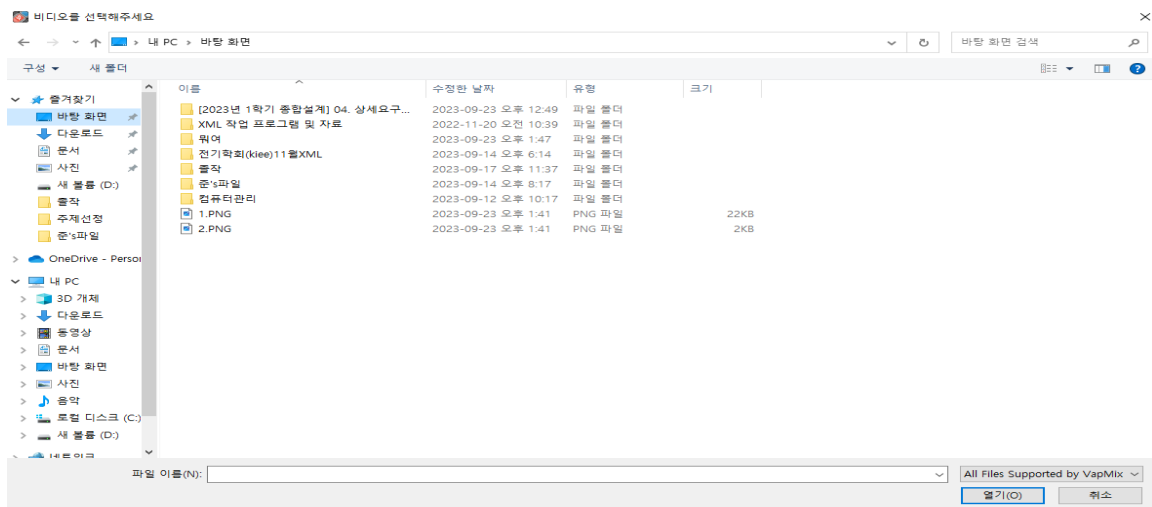
프로그램 실행 시[그림 1]과 같은 화면을 볼 수 있고, Video 화면을 클릭하면 파일 탐색기를 통해 동영상을 업로드 할 수 있다. 동영상 업로드 전 모든 기능은 사용할 수 없다.



[그림 1] 메인

#### 3.2) 동영상 업로드

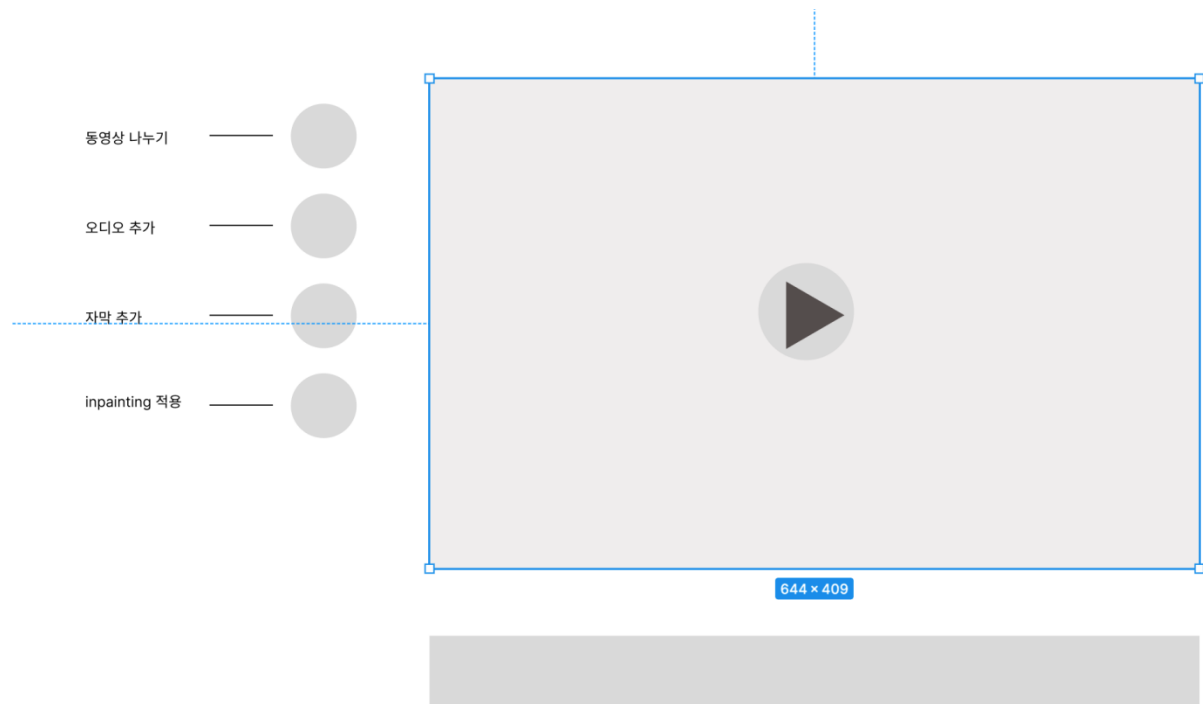
파일 탐색기를 통해 동영상을 업로드할 수 있다



[그림 2] 파일 탐색기

### 3.3) 동영상 업로드 후

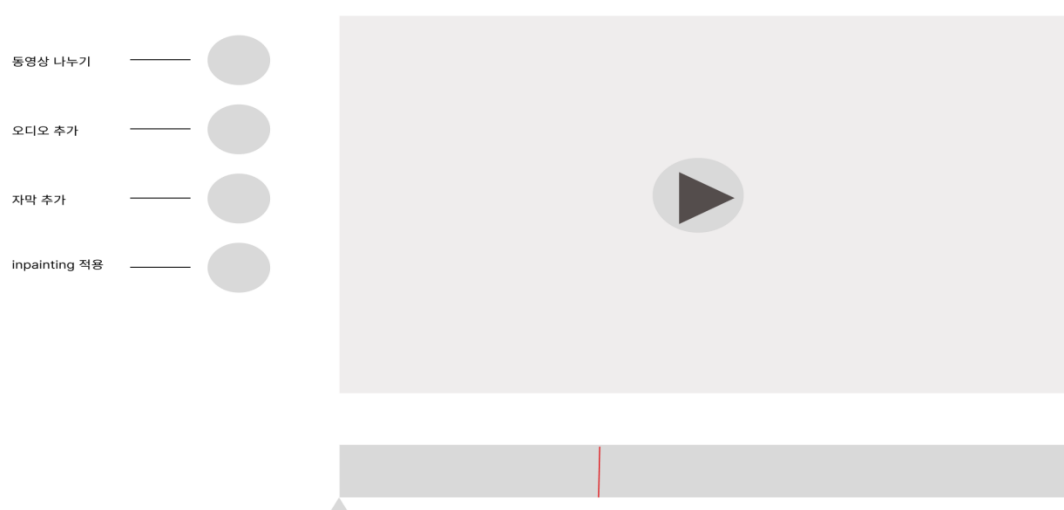
동영상 올린 후 화면으로 가운데 버튼을 통해 동영상을 재생할 수 있다. 타임라인을 통해 동영상을 원하는 위치로 이동할 수 있고 좌측 네가지 기능을 사용할 수 있다.



[그림 3] 동영상 업로드 후

### 3.4) 동영상 나누기

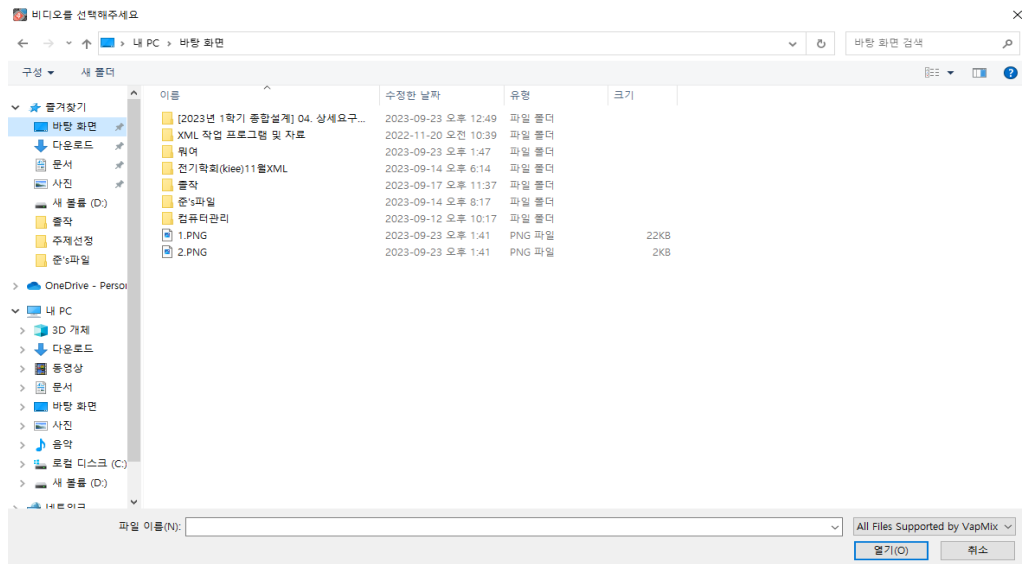
동영상 나누기 버튼을 클릭하고 타임라인 내에서 나누고자 하는 부분을 선택하면 동영상이 나뉜다.



[그림 4] 동영상 나누기

### 3.5) 오디오 추가

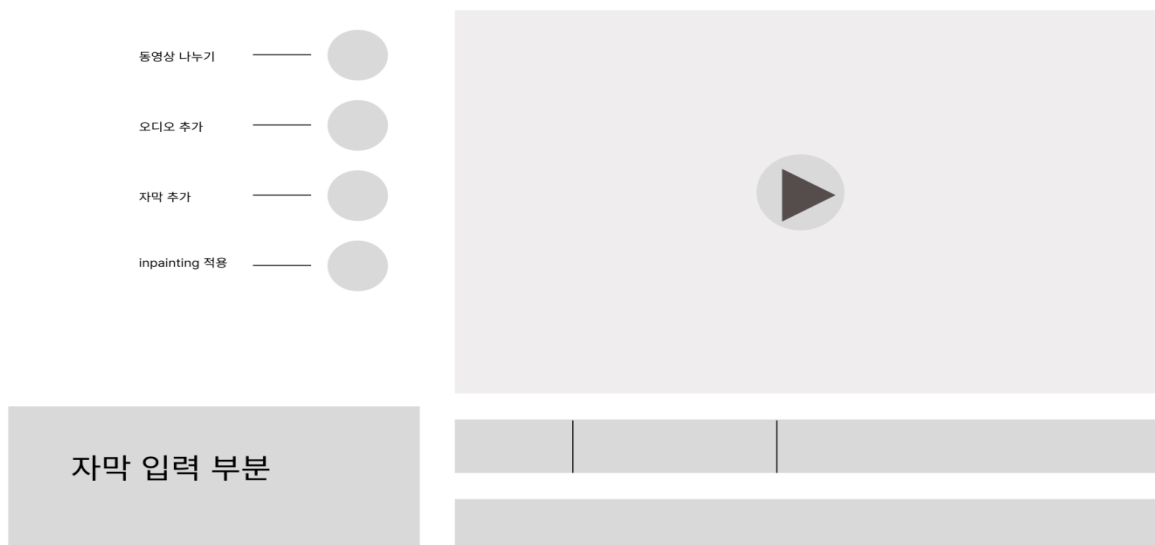
오디오 추가 버튼을 누르면 [그림 5]와 같이 파일 탐색기가 열리고 동영상에 추가하고자 하는 오디오 파일을 업로드할 수 있다.



[그림 5] 오디오 파일 탐색

### 3.6) 자막 추가

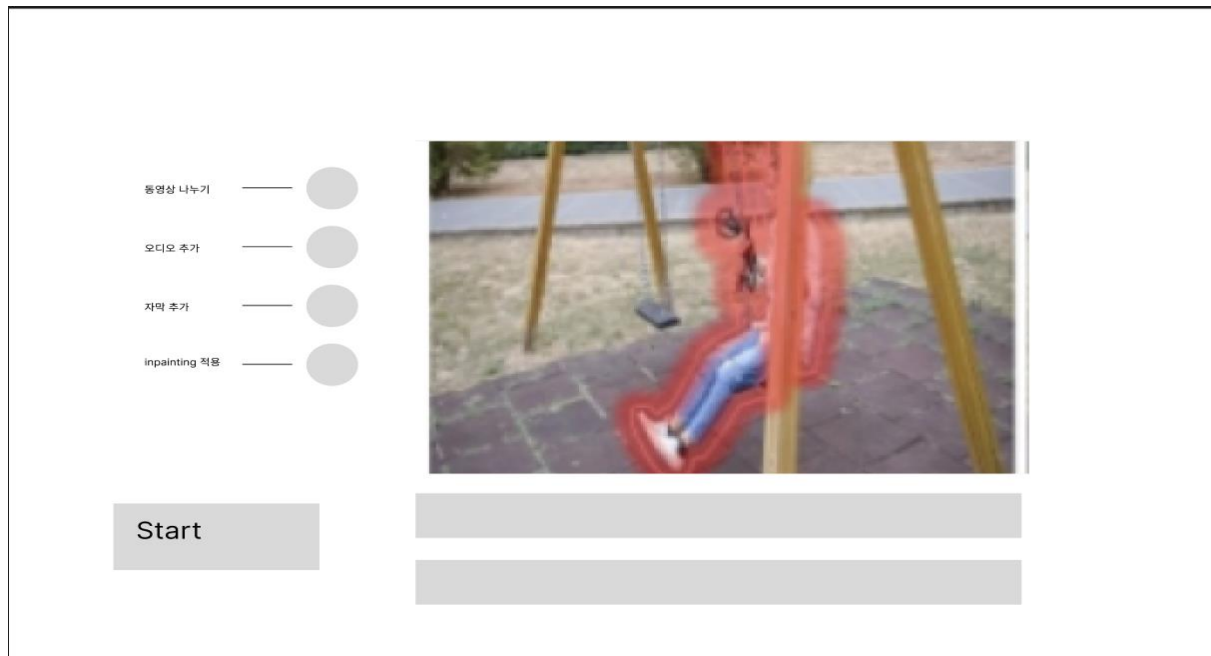
자막 추가 버튼을 누르면 동영상 시간과 같은 타임라인 하나가 더 생기게 된다. 두개의 커서를 통해 자막을 넣고 싶은 구간을 지정하고 자막 입력 부분에 텍스트를 입력하면 동영상의 같은 시간대에 자막이 나타나게 된다



[그림 6] 자막 추가

### 3.7) Video Inpainting 적용

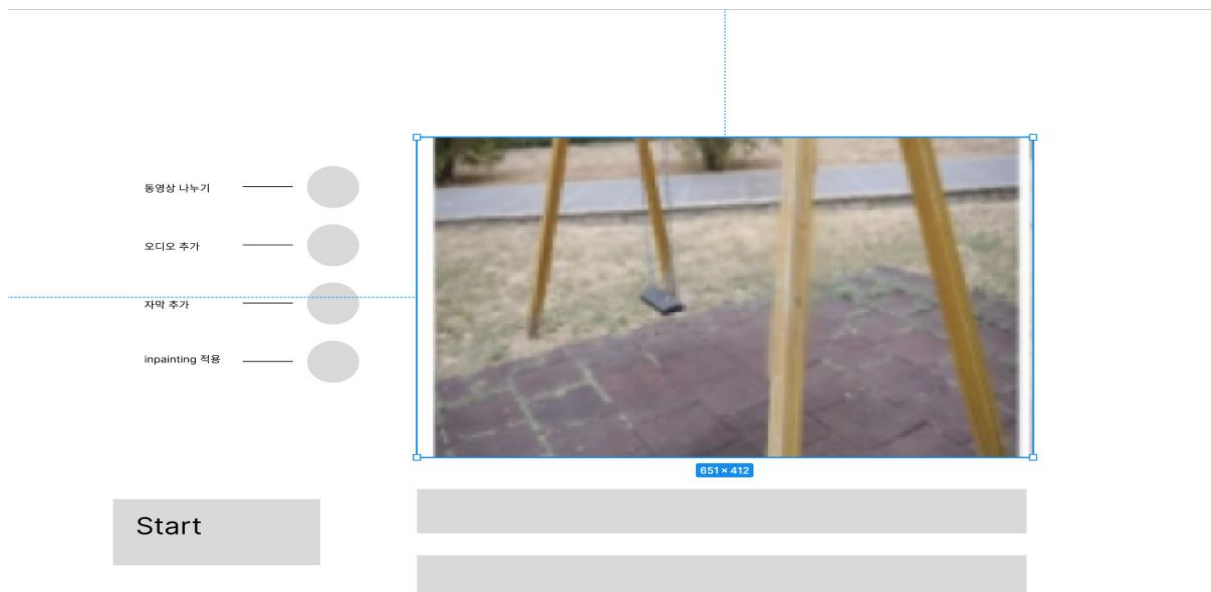
지우고자 하는 객체가 있는 시간대로 타임라인을 이동시켜 [그림 7]처럼 객체를 색칠한다. 이후 Start 버튼을 클릭하면 Inpainting 이 적용되어 해당 객체가 동영상에서 지워진다



[그림 7] 객체 Masking

### 3.8) Video Inpainting 적용 이후

[그림 8]처럼 지우고자 하는 객체가 지워진 동영상을 얻을 수 있다



[그림 8] Inpainting 이후