

## 서론 및 배경지식

- 최근 메타버스와 소셜 VR 플랫폼의 발전으로 인해 사용자가 직접 생성한 콘텐츠인 **UGC(User Generated Content)**가 증가하고 있다[1].
- UGC**는 사용자가 직접 콘텐츠를 제작할 수 있는 자유를 제공하지만, 성적이거나 폭력적인 유해 콘텐츠의 무분별한 제작 및 배포 문제를 발생시킨다[2].
- 관련 연구에서는 VR 플랫폼 상에서의 괴롭힘을 감지하기 위해 음성, 제스처를 분석하였다[3].

본 연구에서는 소셜 VR 플랫폼 내의 성적으로 유해한 콘텐츠를 실시간으로 탐지 및 마스킹하여 사용자, 특히 아동과 청소년을 보호하는 솔루션인 **VRMask**를 소개한다.

## VRMask

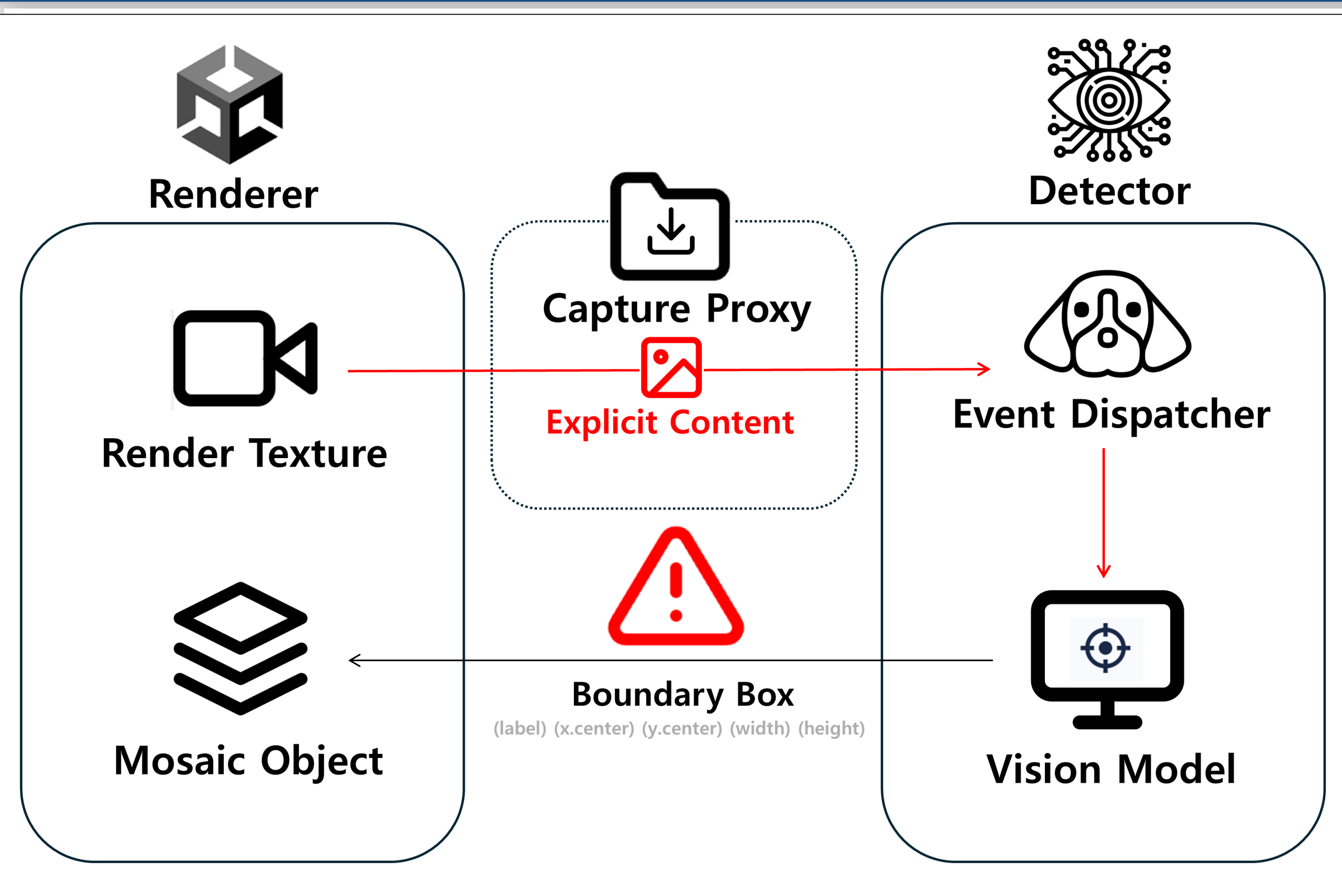


Fig.1 VRMask System overview

### Renderer

- 사용자 시점에서 렌더링된 화면을 캡처한다. 또한 유해 콘텐츠 탐지를 위한 프레임을 생성하여, Capture Proxy에 기록한다.
- 마스킹이 유해 콘텐츠 탐지를 방해하는 현상을 방지하기 위해서, 별도의 Renderer를 통해 마스킹을 하지 않은 유해 콘텐츠를 전달하며, 추론을 지연시키지 않게 구성하였다.

### Capture Proxy

- Renderer에서 생성된 프레임을 저장하며, 마스킹이 적용되지 않은 콘텐츠를 Detector에게 전달할 수 있는 컴포넌트이다.

### Detector

- YOLOv5 모델을 사용하여 유해 콘텐츠를 탐지하고, 탐지된 영역에 대해서 마스킹을 위한 Boundary Box를 생성한다.

## 평가 및 탐지 지연 시간

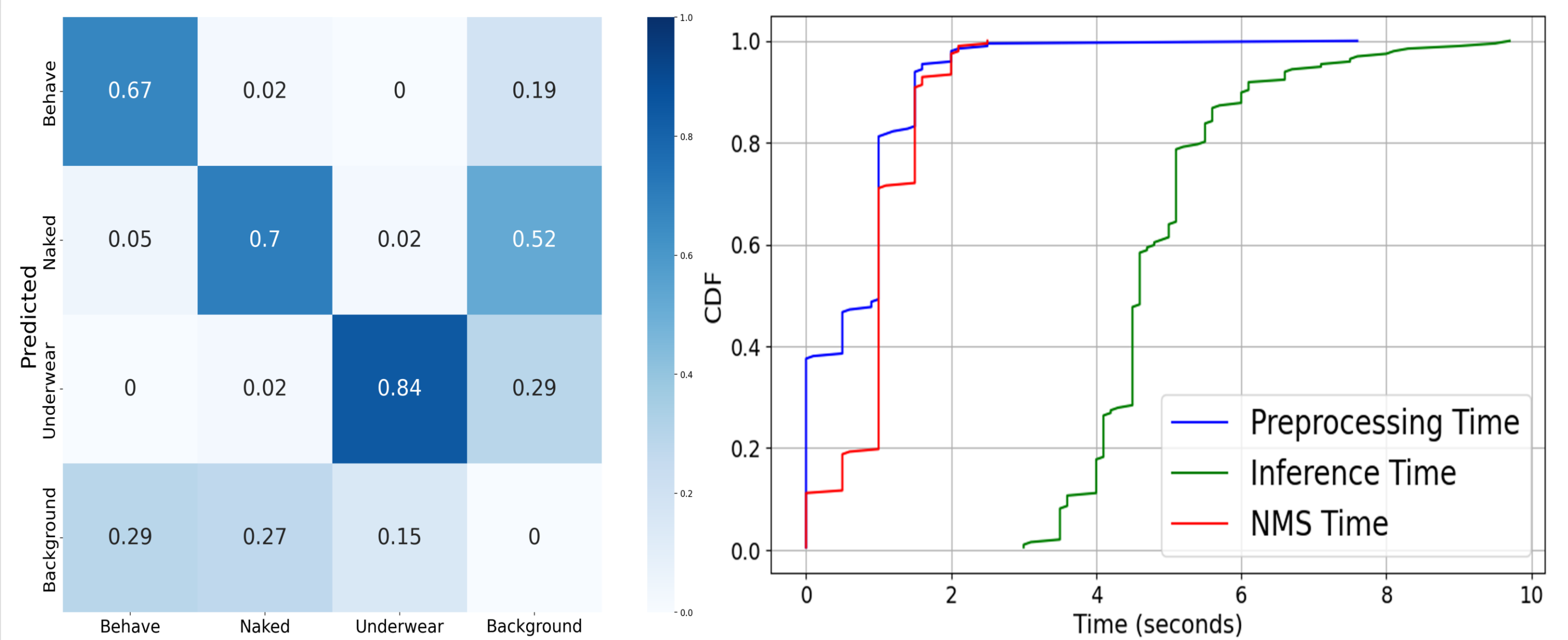


Fig. 2 Confusion matrix & Latency of YOLOv5 Vision Model

### 평가

- 데이터셋
  - 인기 VR 소셜 게임인 VRChat에서 수집한 800여 개의 유해 콘텐츠 데이터를 직접 수집하였다.
  - 수집된 데이터를 4가지의 class인 Behave, Naked, Underwear, Background로 각각 라벨링하여 학습에 사용하였다
- 탐지 성능
  - Underwear과 Naked 클래스가 각각 **0.842, 0.695**로 준수한 성능을 보인다[Fig. 2].
  - Behave는 0.608로 낮은 성능을 보이고 있는데, 이는 모션을 탐지하는데 어려움이 있었기 때문이다. 또한 클래스 간 혼동은 0.2~0.5로 낮은 수치를 보인다.

### 탐지 지연 시간

- 전처리 시간
  - 추론 시간
  - NMS(Non-Maximum Suppression) 시간
- Fig. 2의 측정된 Detector의 탐지 지연 시간은 **8ms 이하**의 지연 시간을 **80% 확률**로 가진다. 또한 Unity에서의 영역 마스킹 생성 시간은 탐지되지 않을 정도의 미미한 시간이다.
  - 총 지연시간은 8ms로 VR상에서 몰입감을 해치지 않고 유해 콘텐츠를 차단하기에 충분하다[Fig. 3].

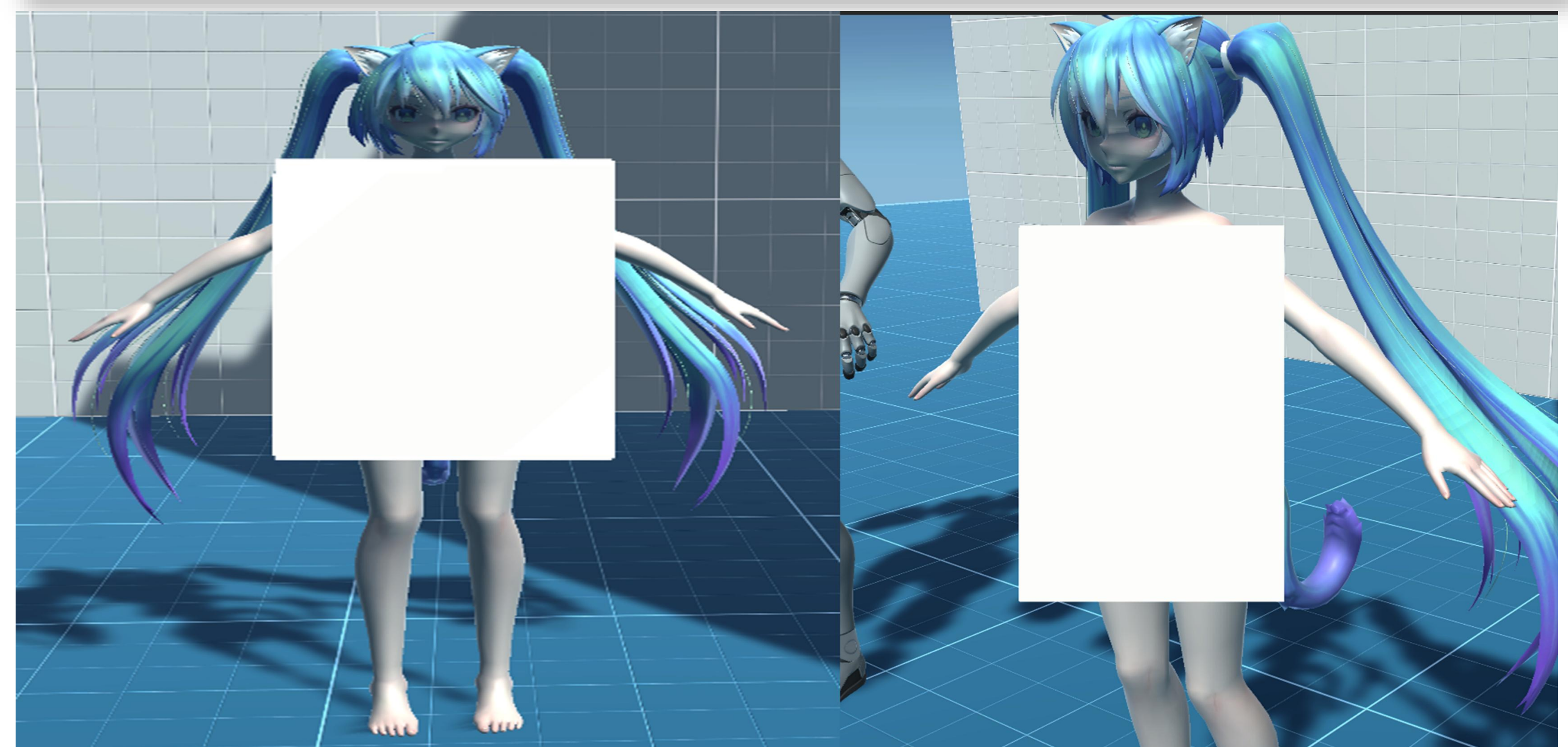


Fig. 3 The harmful avatars masked by VRMask

## 결론

본 논문에서는 직접 수집한 유해 콘텐츠 데이터셋을 기반으로 YOLOv5 모델을 학습시키고, 이를 통해 실시간 유해 콘텐츠 탐지 및 이에 대한 마스킹 처리를 통해 사용자를 보호하는 시스템인 **VRMask**를 제안했다.

## 참고문헌

- [1] Guo, Keyan, et al. "Moderating Illicit Online Image Promotion for Unsafe User-Generated Content Games Using Large Vision-Language Models." USenix Security Symposium (2024).
- [2] "Metaverse app allows kids into virtual strip clubs", <https://www.bbc.com/news/technology-60415317>
- [3] Wang, Na, et al. "HardenVR: Harassment Detection in Social Virtual Reality." IEEE Virtual Reality and 3D user Interfaces (2024)