

작품명 : AI를 활용한 비접촉 의류 착용 시스템

팀명 : 유진경은호

지도교수 : 전윤호

참여학생 : 진경은, 이유진, 신은호

1. 서론(작품 개요)

현재 온라인 쇼핑몰의 반품률은 약 30%로 오프라인 매장의 10%에 비해 3배 높은 수준이다. 이러한 차이는 주로 실제 착용감을 미리 확인 할 수 없어 발생하는 사이즈 불일치 문제로 인한 것으로, 물류 비용 증가와 소비자 만족도 저하의 주요 원인이 된다. 이러한 문제를 해결하기 위해 VTON(Virtual Try-On Network) 기술에 대한 연구가 활발히 진행되고 있다. VTON은 사용자의 이미지와 선택한 의류를 자연스럽게 합성하여 실제 착용과 유사한 가상 이미지를 제공함으로써 구매 전 착용감을 미리 확인 할 수 있도록 하는 기술이다.

특히 최근 Diffusion 모델 기반의 VITON 시스템들이 높은 품질의 합성 이미지를 생성하며 주목받고 있다. 그러나 기존의 Diffusion 기반 VTON 시스템들은 주로 서버 환경에서의 추론을 전제로 하고 있어, 사용자의 개인 이미지가 외부 서버로 전송되는 과정에서 개인정보 유출 문제가 발생 할 수 있다. 이에 Latent Diffusion 모델인 CatVTON을 On-device 환경에 적합한 형태로 모델을 최적화하여 사용자 이미지를 Latent Vector 형태로 암호화해 서버로 전송하거나 전체 추론과정을 On-device에서 수행하여 개인정보 유출 문제를 해소한다.

2. 프로젝트 목표

실제 매장 환경에서 활용할 수 있도록 Jetson Orin Nano 개발자 키트를 활용해 모든 추론 과정을 On-device로 수행한다.

사용자가 어디서든 편리하게 서비스를 활용할 수 있도록 S25의 Adreno 830(GPU) 프로세서를 활용해 사용자 이미지를 암호화하고 이를 서버로 전송해 암호화된 상태에서 합성을 수행함으로써 개인 정보 유출 문제를 최소화한다.

3. 본론(제품 소개, 작품 추진과정)

1) CatVTON

기존 VITON 시스템의 높은 모델 복잡도와 추론 시간을 개선하기 위해 제안된 모델로 Reference U-Net, CLIP, DINOv2와 같은 외부 모듈을 제거함으로써 전체 파라미터 수와 연산량을 크게 줄인 단일 U-Net 구조를 사용해 추론하는 모델이다.

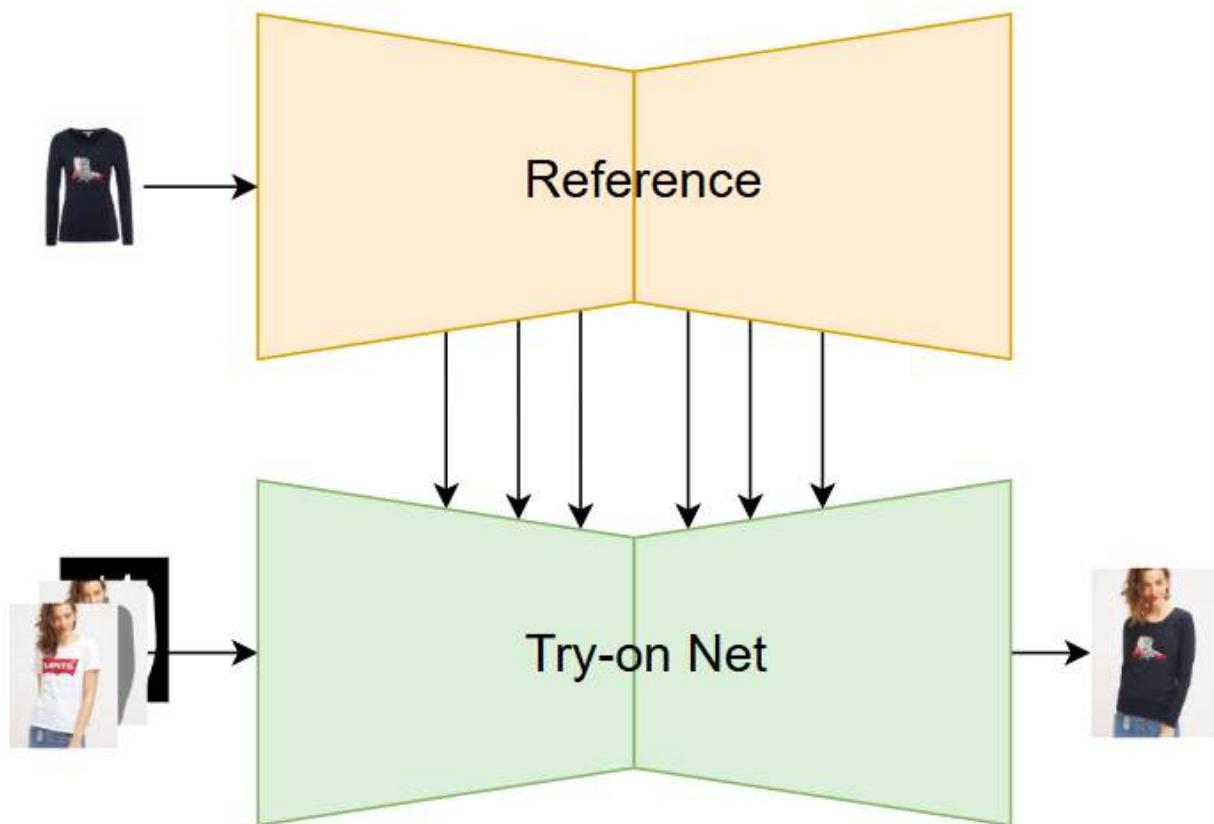


그림1. 기존 모델 구조

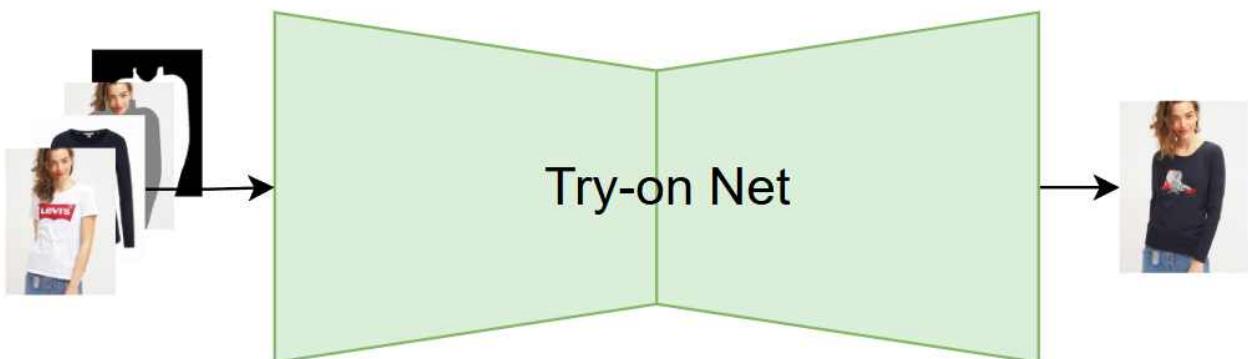


그림2. CatVTON 모델 구조

2) Reference Network 제거

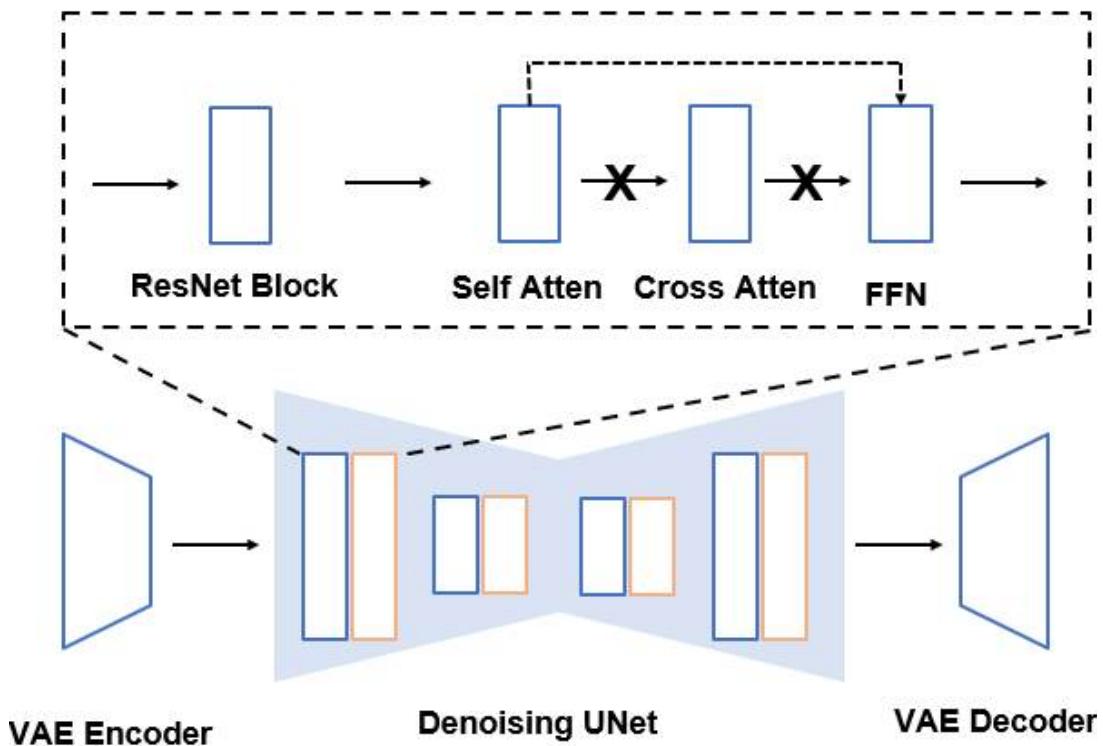


그림3. Reference 모듈 제거

기존 VTON 모델에서 성능 향상을 위해 별도의 Encoder를 두어 Cross-attention을 활용해 조건을 주입한다. CatVTON의 경우 조건과 입력을 하나의 모델로 처리하기 때문에 Cross-attention이 필요로 하지 않는다. 연산량 감소와 속도 향상을 위해 Cross-attention layer를 제거 한다.

3). 마스크 생성 모델 설계 및 구현

본 프로젝트에서는 Segmentation Models Pytorch 라이브러리를 활용하여 ResNet-50 백본 기반의 U-Net 세그멘테이션 모델을 구현하였다. Dice Loss와 IoU 평가 지표를 사용하여 VITON-HD 데이터셋으로 학습을 진행하였으나, 모든 이미지가 흰색 배경으로 구성되어 있어 인터넷 이미지나 실제 촬영 이미지에서 마스크 예측 성능이 크게 저하되는 문제가 발생하였다. 이는 모델이 배경의 다양성에 대해 강건하지 못하다는 것이 원인으로 파악되었다.

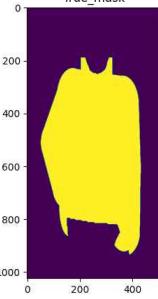
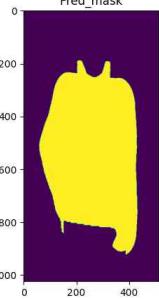
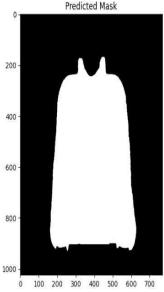
data	Image	True mask	Pred mask
VITON-HD			
인터넷 이미지			
흰색 배경 이미지			

표1. Segmentation model 결과

배경 강건성 문제를 해결하기 위해 BG-20K 데이터셋을 활용한 이미지 합성 알고리즘을 개발하였다. SAM 모델의 예측 마스크로 전경을 추출하고 위치 검증 후, 침식 연산으로 외곽선을 제거하는 후처리를 수행하였다. 배경 이미지를 전경과 동일한 크기로 리사이즈한 뒤 알파 채널 기반의 alpha_composite 연산으로 합성을 완성하였으며, 초기 오합성 현상은 전경 위치 검증 로직 추가로 해결하였다. 아래의 순서도는 최종 이미지 합성 알고리즘의 전체 과정을 나타낸다.



그림4. 최종 합성 이미지

개선된 합성 데이터로 재학습된 모델은 다양한 배경을 가진 인터넷 이미지와 직접 촬영한 이미지에서 안정적이고 정확한 마스크 예측 성능을 보였다. 배경 강건성이 크게 향상되어 실제 환경에서도 효과적으로 의류 영역을 분할할 수 있음을 확인하였다.



그림5. 인터넷 이미지로 추론한 결과

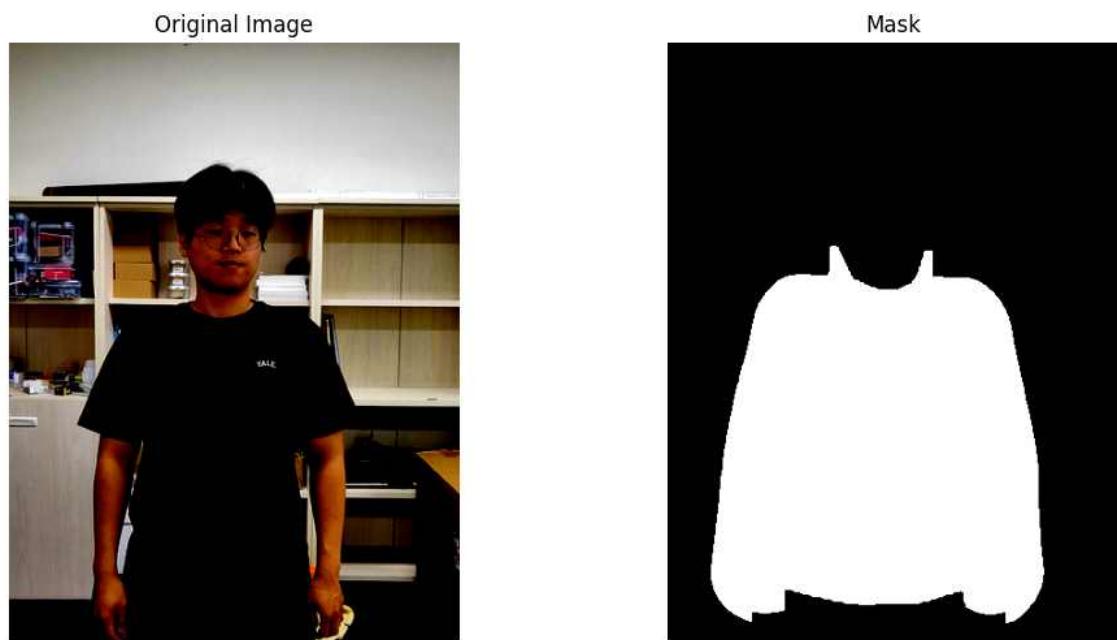


그림6. 자체 이미지로 추론한 결과

4) Jetson Orin Nano

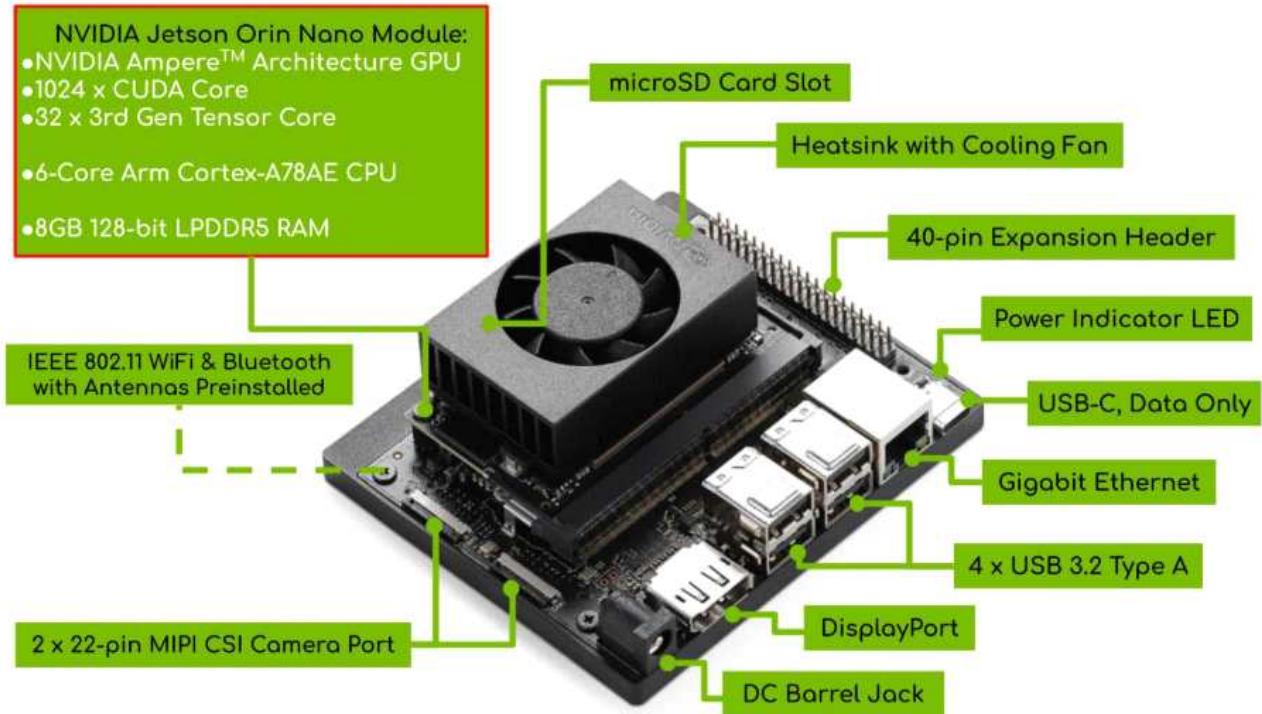


그림7. Jetson Orin Nano, GPU : 472G FLOPS (FP32 기준)

Jetson Orin Nano의 경우 NVIDIA JetPack SDK를 지원하여 기존 서버환경과 동일한 형태로 추론이 가능하다. CatVTON 모델을 Jetson Orin Nano에서 실행 할 경우 아래와 같은 표2 와 같은 Latency가 발생하며 단일 추론에 30초 이상 소요되어 사용자 응답성이 떨어진다. 이는 Noise prediction 과정이 주요 병목이다. 기존 DDIM Scheduler의 경우 DDPM Scheduler를 1차 근사해 구현한다. DPM++의 경우 고차 방정식 형태로 근사해 미분의 근사값 까지 활용 함으로써 더 적은 Step 만으로도 정확한 역 확산을 수행한다. 이러한 특성을 활용해 DDIM Scheduler를 DPM++ Scheduler로 대체하고, FP32 형태의 모델을 BFP16으로 압축하여 이를 활용해 표3과 같이 최적화한다.

처리단계	이미지 전처리	VAE Encoder	denoising (50step)	VAE Decoder	Safety Checker
시간 (초)	0.006초	0.06초	29.5초	0.05초	0.34~0.36초

표 2 Jetson Orin Nano CATVTON 모델 단계별 추론 시간

항목	기존 모델 (FP32, DDIM, 50step)	최종 모델 (BFP16, DPM++, 20step)	변화
추론 시간	약 30초	11초	-19초
연산 정밀도	FP32	BFP16	-
입력 해상도	512×384 (정상 작동)	512×386 (정상 작동)	-
모델 크기	대용량 FP32 기반	양자화 제외, BFP16 + 타임스텝 축소	구조 최적화로 간접 축소

표 3 최종 모델 성능 비교

5) S25 - Qualcomm Adreno 830

S25의 경우 Executorch, TensorFlow Lite 등 다양한 Model Optimizer를 지원한다. 본 프로젝트에서는 pytorch model에 최적화 되어 현재 갤럭시 핸드폰에 환경에 맞게 개발중인 Executorch를 활용한다. Executorch의 경우 모델의 연산 그래프를 캡쳐하여 Target-device 에 맞는 IR로 변환해 원하는 device에서 실행 가능하도록 변환해 주는 역할을 수행한다. 이를 위해 기존 성능을 유지하면서 모델을 Lowering 가능한 형태로 편집을 수행한다.



그림8. Adreno830, GPU : 3686.4 GFLOPS (FP32 기준)

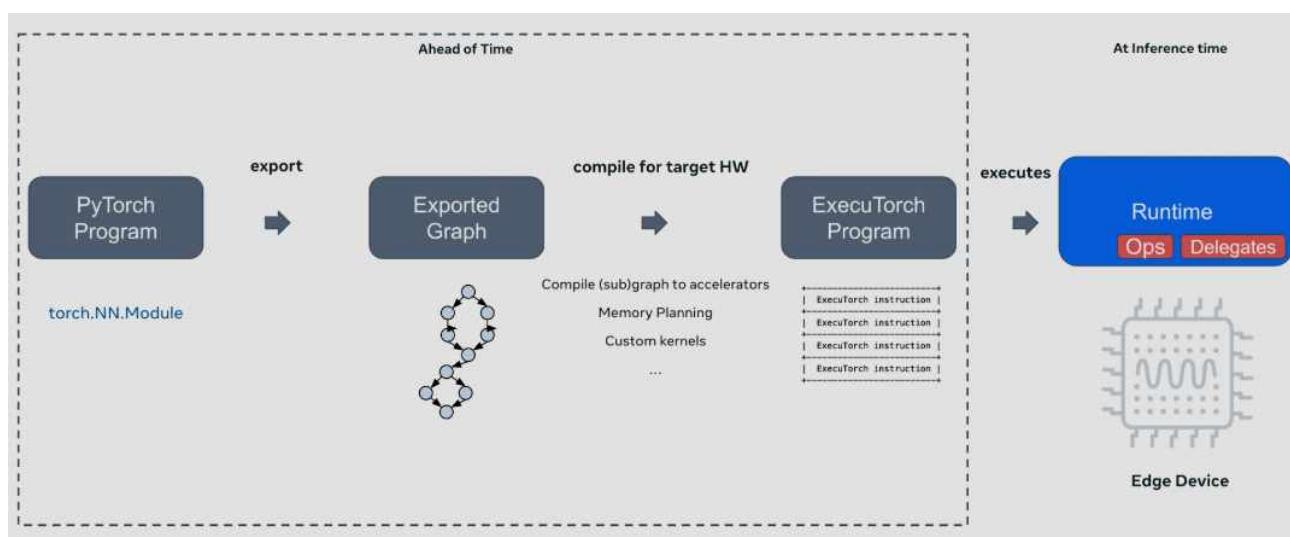


그림9. Executorch Module Lowering

6) 서비스 구현

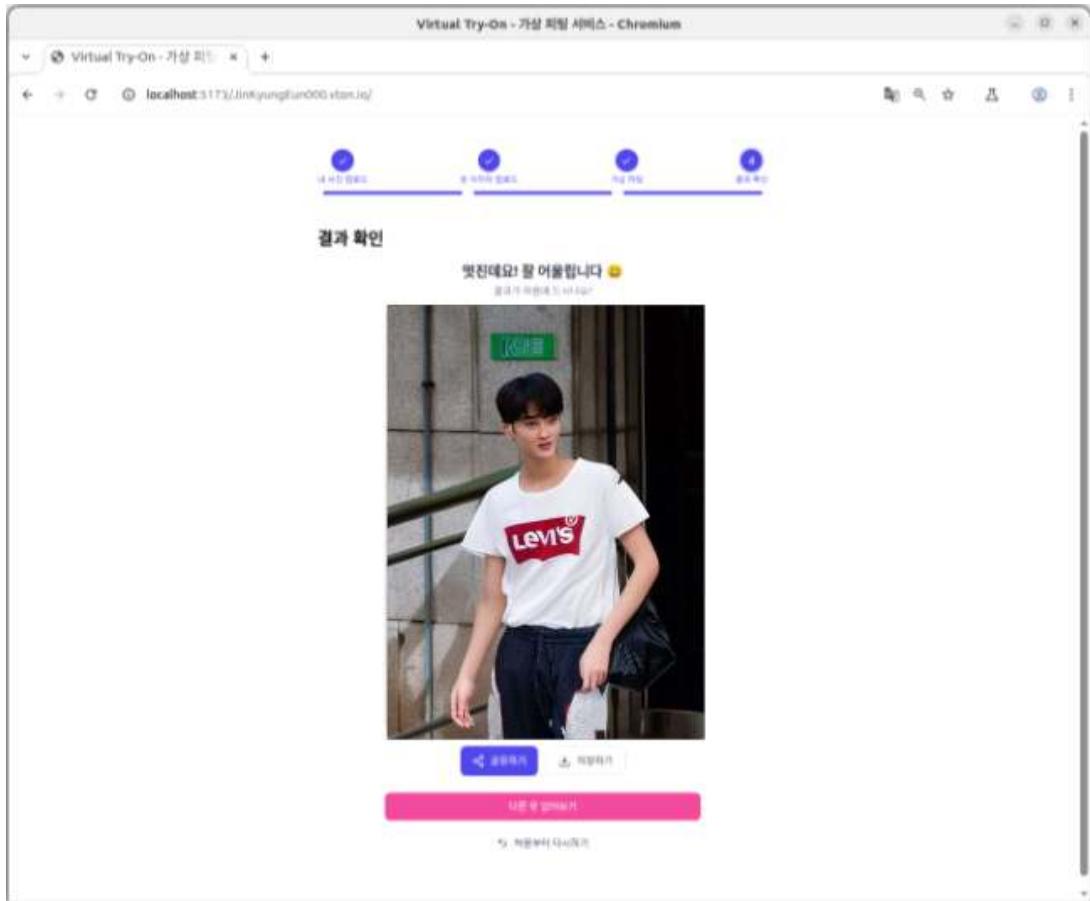


그림10. 서비스 디자인

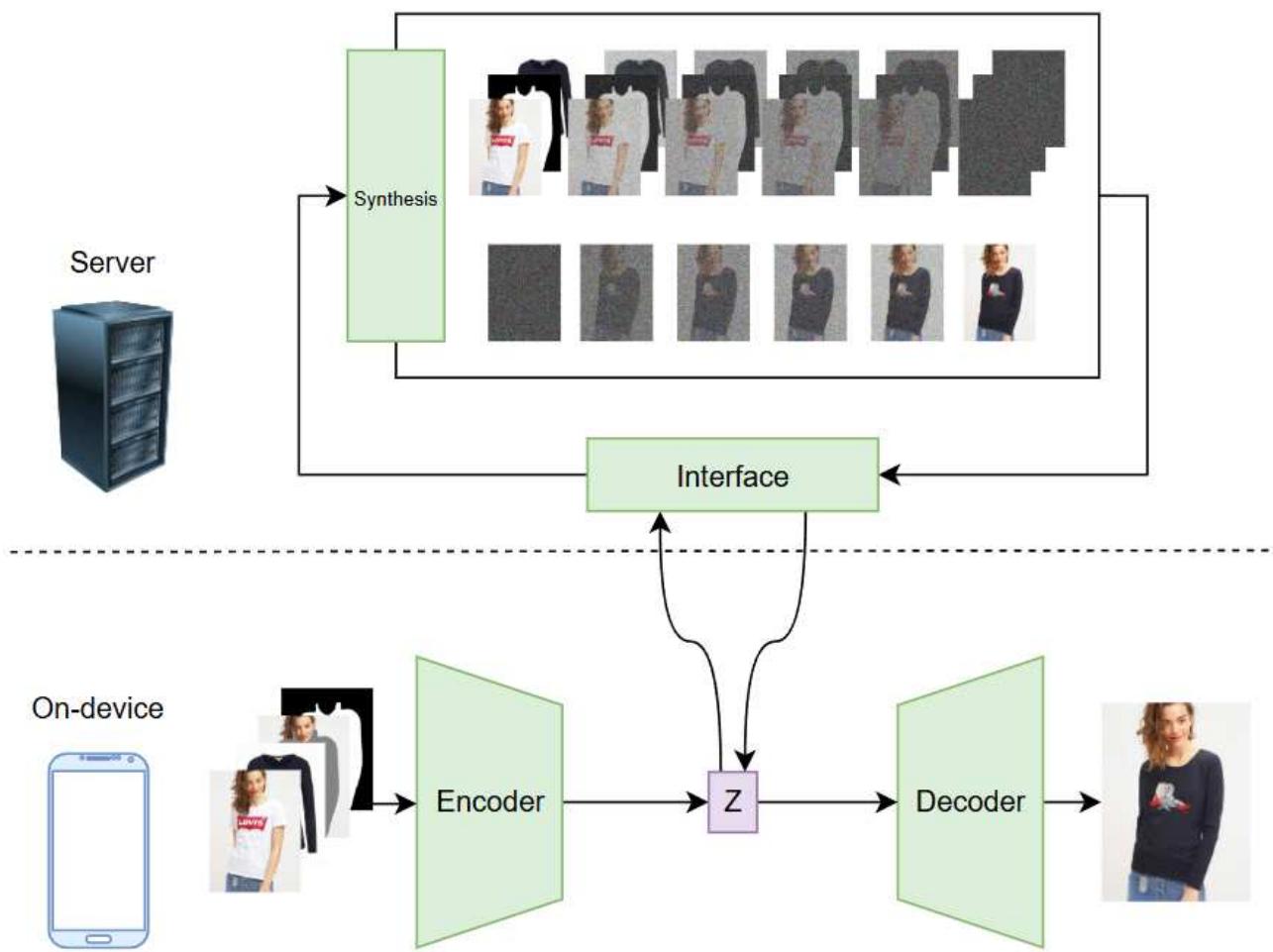


그림11. 서비스 구조도

4. 결론

본 프로젝트는 기존 서버 기반 VTON 시스템의 개인정보 유출 문제와 높은 추론 시간 문제를 동시에 해결하기 위해, CatVTON 모델을 On-device 환경에 최적화한 비대면 의류 착용 시스템을 구현하였다.

Jetson Orin Nano에서는 FP32 → BFP16 변환 및 DPM++ 스케줄러 적용을 통해 추론 시간 단축(30초 → 11초)을 달성하였으며, 모바일 환경인 Galaxy S25(Adreno 830)에서도 Executorch를 활용해 추론이 가능함을 확인하였다. 본 시스템은 개인 이미지 암호화 및 On-device 추론 수행으로 개인정보 유출 위험을 최소화하고, 경량화된 Diffusion 기반 VTON 모델을 통해 고품질의 가상 착용 이미지를 빠르게 생성하며, Jetson 및 모바일 디바이스 환경에서도 실시간 응용이 가능한 수준으로 최적화했다.

5. 기대효과

비대면 환경에서도 안전하고 실감나는 의류 가상 착용 서비스를 제공할 수 있는 실질적 솔루션을 제시하였으며, 향후 스마트 미러, 모바일 쇼핑앱, AI 피팅룸 등 다양한 산업 응용으로 확장 가능성을 가진다.