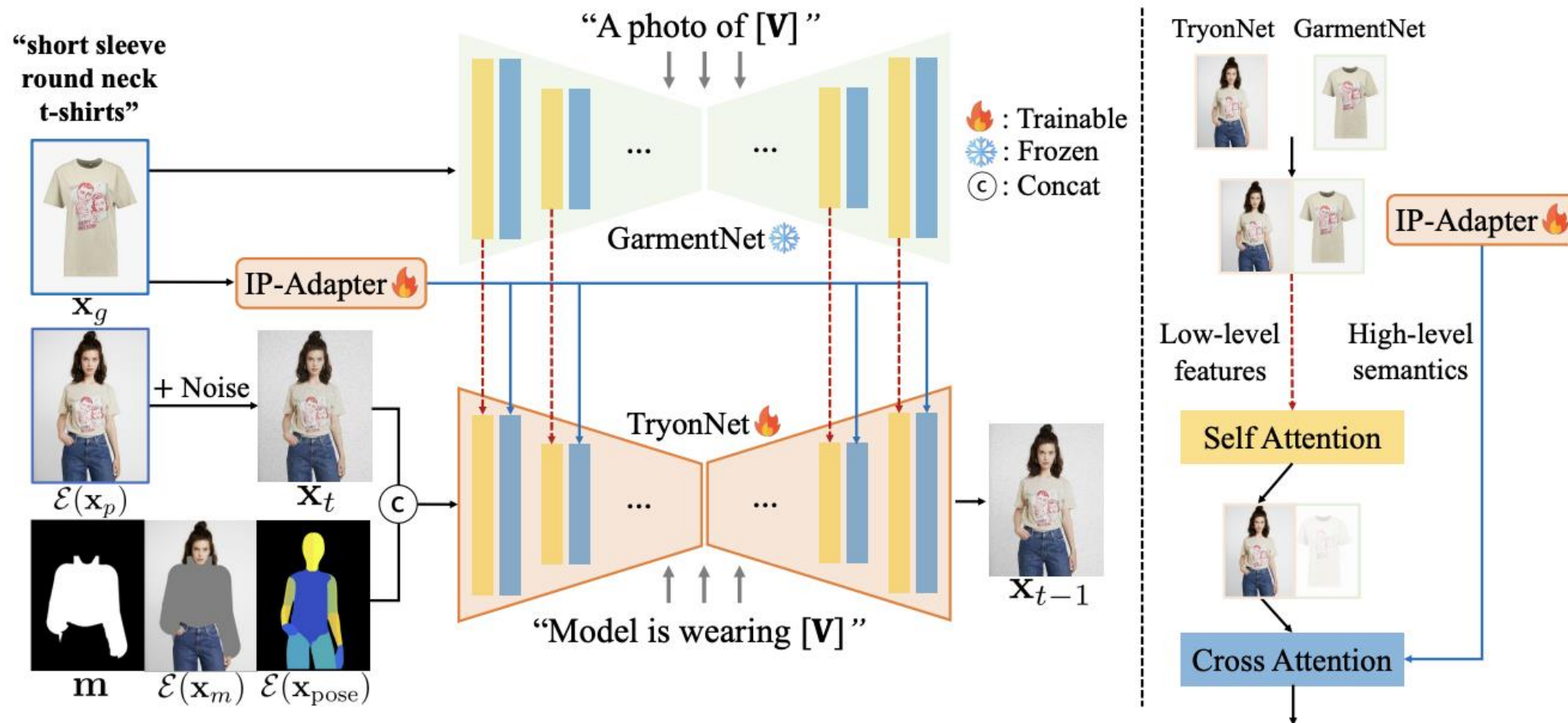


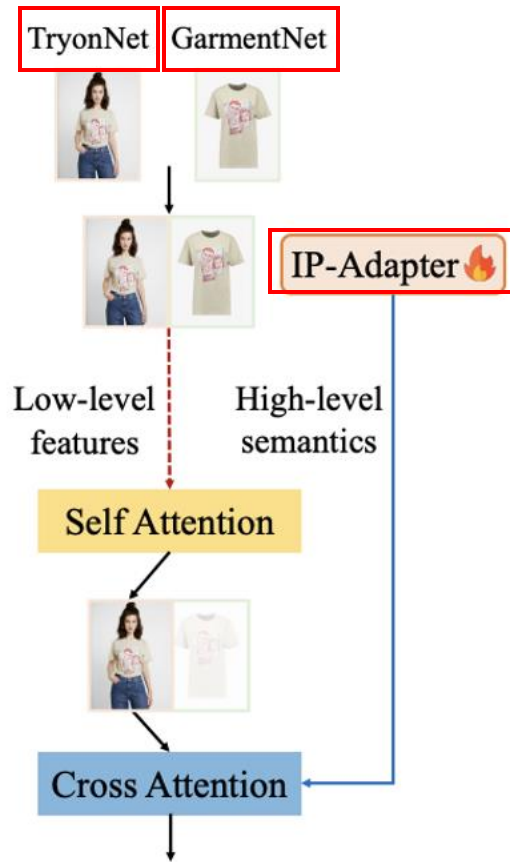
IVM –VTON 핵심원리

논문 정리

IDM-VTON 모델 개요



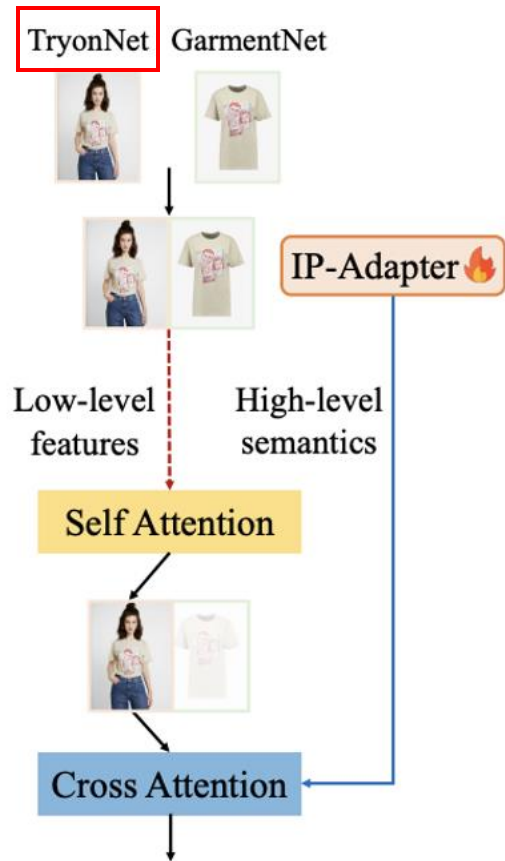
IDM-VTON 모델 개요



- 크게 3가지로 나눌 수 있음

1. TryonNet
2. GarmentNet
3. IP-Adapter(Image Prompt Adapter)

IDM-VTON 모델 개요 - TryonNet

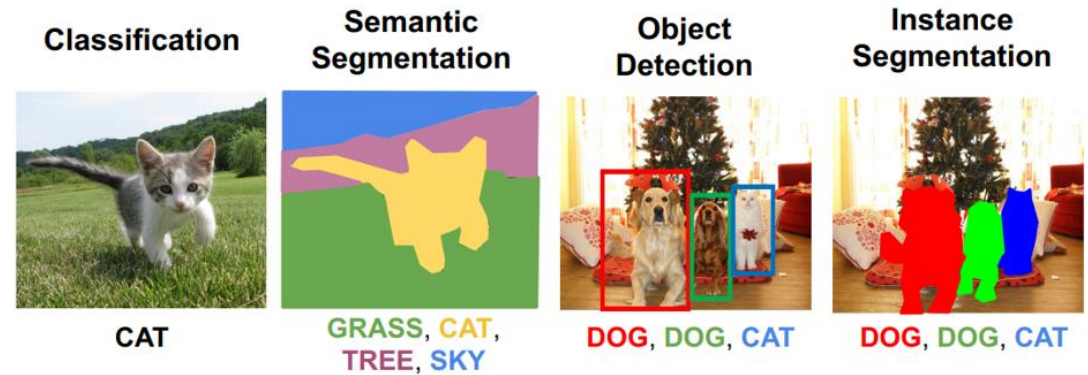


TryonNet

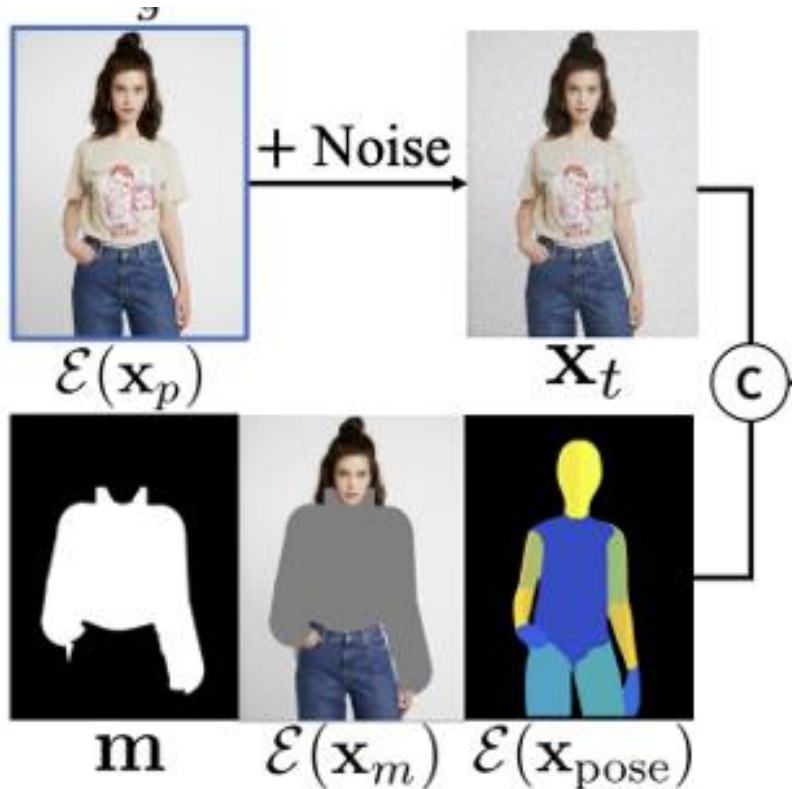
- 사람 이미지를 처리하는 주요 UNet

-> UNet이란?

-> 이미지의 모든 픽셀이 어떤 카테고리에 속하는지 분류하는 것

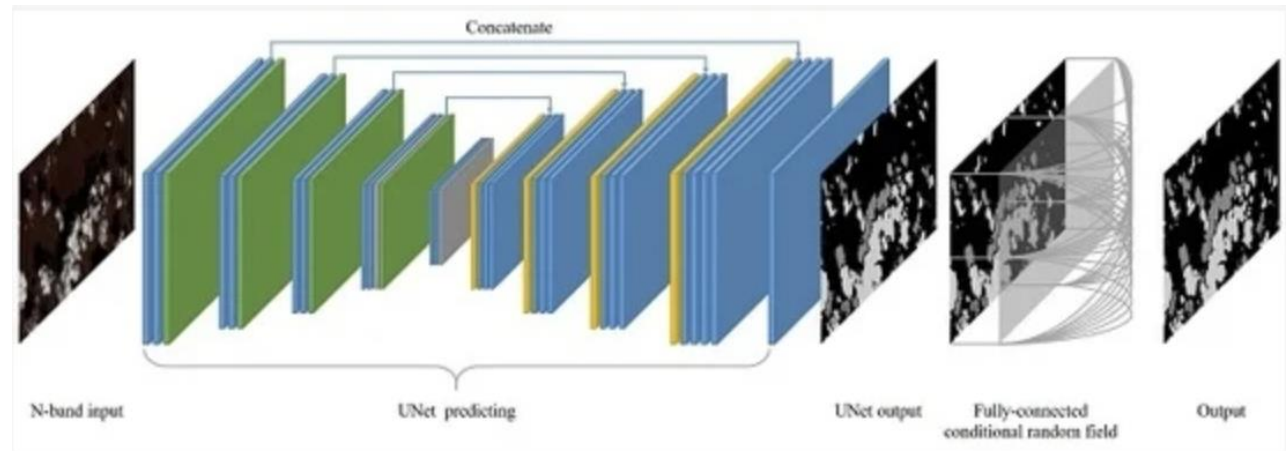


IDM-VTON 모델 개요 - TryonNet

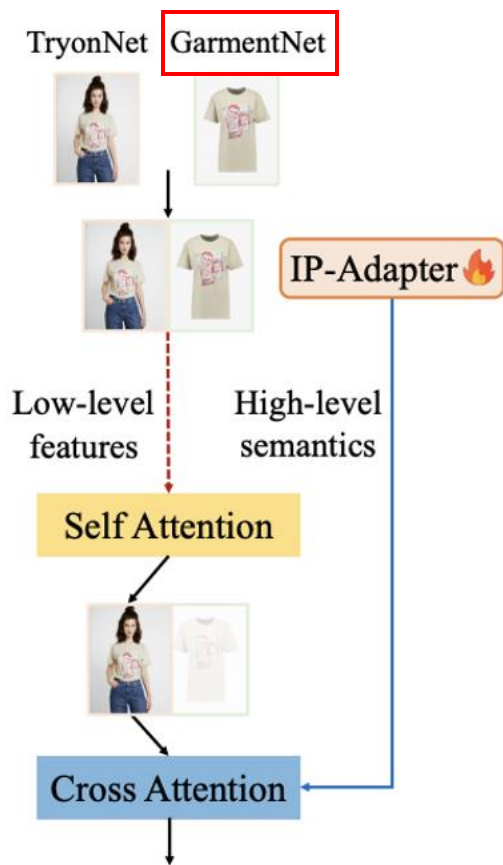


TryonNet의 입력 정보

- 인물 이미지의 잠재 변수 $\mathcal{E}(x_p)$: 인물 이미지의 기본 특성
- 리사이징된 마스크 m : 인물 이미지에서 의류를 제거하는데 사용
- 마스크가 적용된 인물 이미지의 잠재 변수 $\mathcal{E}(x_m)$: 마스크를 통해 변경될 인물 이미지의 잠재 변수
- 사람 이미지의 Denspose $\mathcal{E}(x_{pose})$: 인물의 포즈 정보를 포함



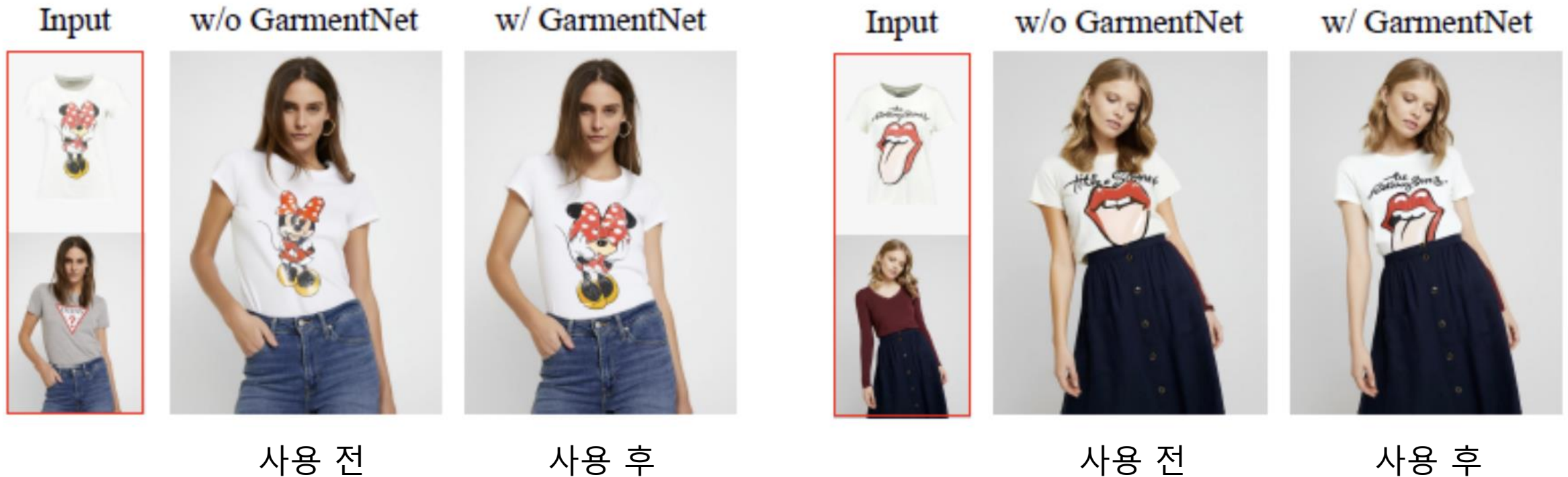
IDM-VTON 모델 개요 – GarmentNet



GarmentNet

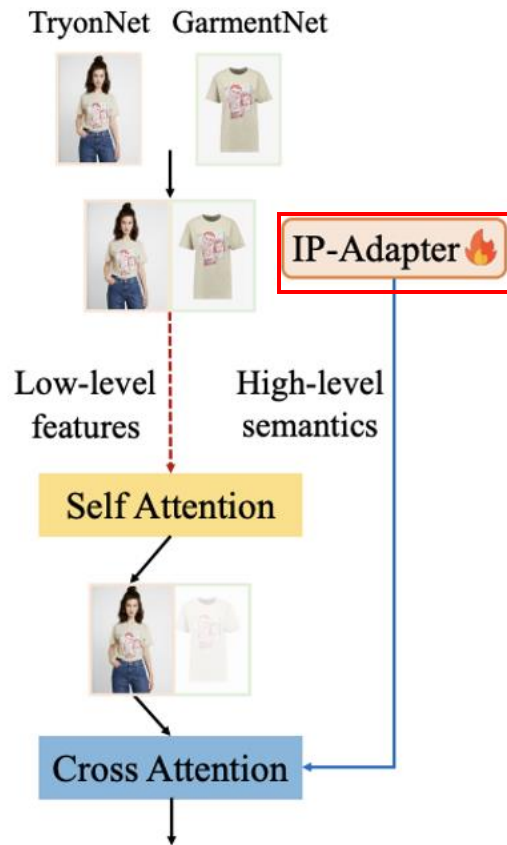
- 의류 이미지에서 저수준 특성을 인코딩 -> 현실적인 착용 모습을 위해 필수적인 질감, 패턴과 같은 세밀한 부분에 중점
- 추가 UNet 인코더를 사용하여 의류에서 세부 사항 강화 (의상 UNet인코더)
- GarmentNet에서 나온 특성은 TryonNet에서 나온 특성과 결합되어 Self Attention 및 Cross Attention 계층에서 추가 처리를 거쳐 최종 가상 이미지를 생성

IDM-VTON 모델 개요 – GarmentNet



-> GarmentNet 사용 시 의류의 세부 정보를 보존할 수 있음

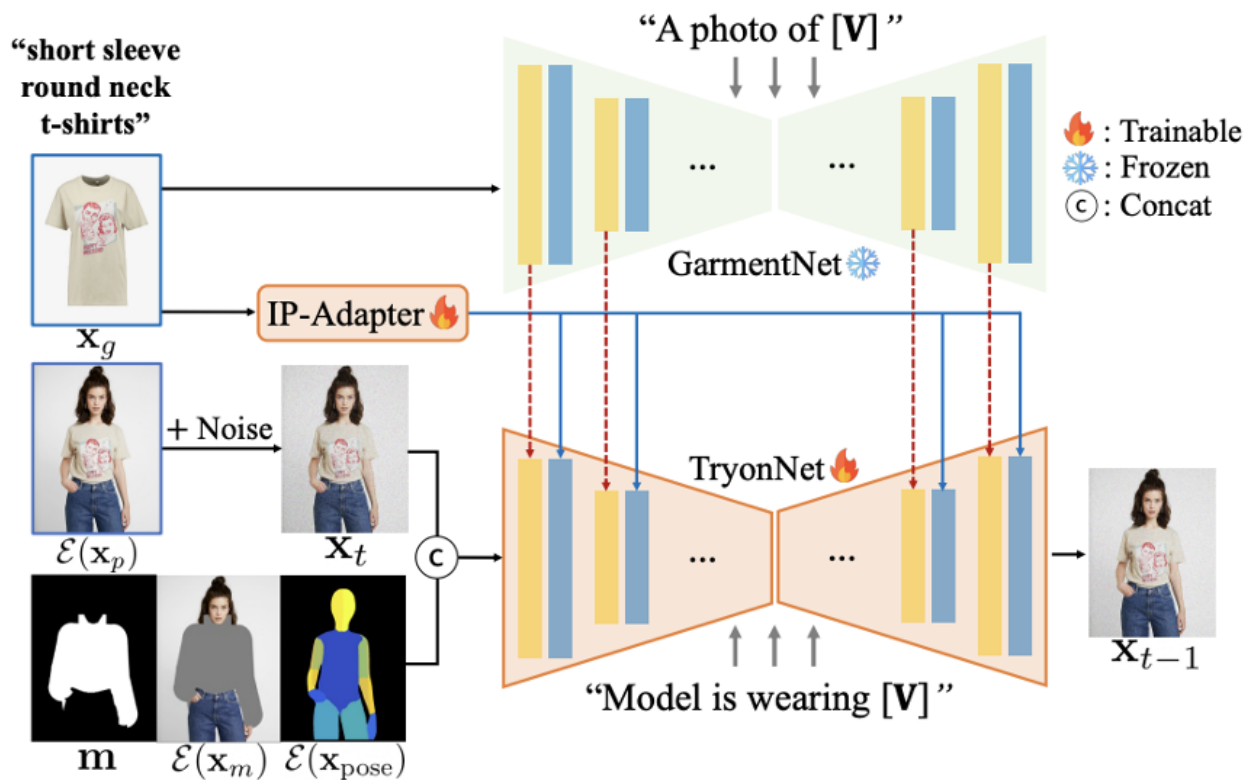
IDM-VTON 모델 개요 – IP-Adapter



IP-Adapter(Image Prompt Adapter)

- 의류 이미지의 고수준 의미론적 정보를 인코딩하는 데 사용
- 인코딩을 위해 frozen CLIP 이미지 인코더를 사용하여 특성 추출
- 추출한 특성들을 투영 계층을 사용하여 미세 조정
- 미세조정된 특성들을 Cross Attention 계층에 통합

IDM-VTON 모델 개요



정리하면

- TryonNet은 사람 이미지를 처리
- GarmentNet는 의류의 저수준 특성
- IP-Adapter는 의류의 고수준 특성 처리
- > 더 세밀하고 정확한 합성 이미지 생성