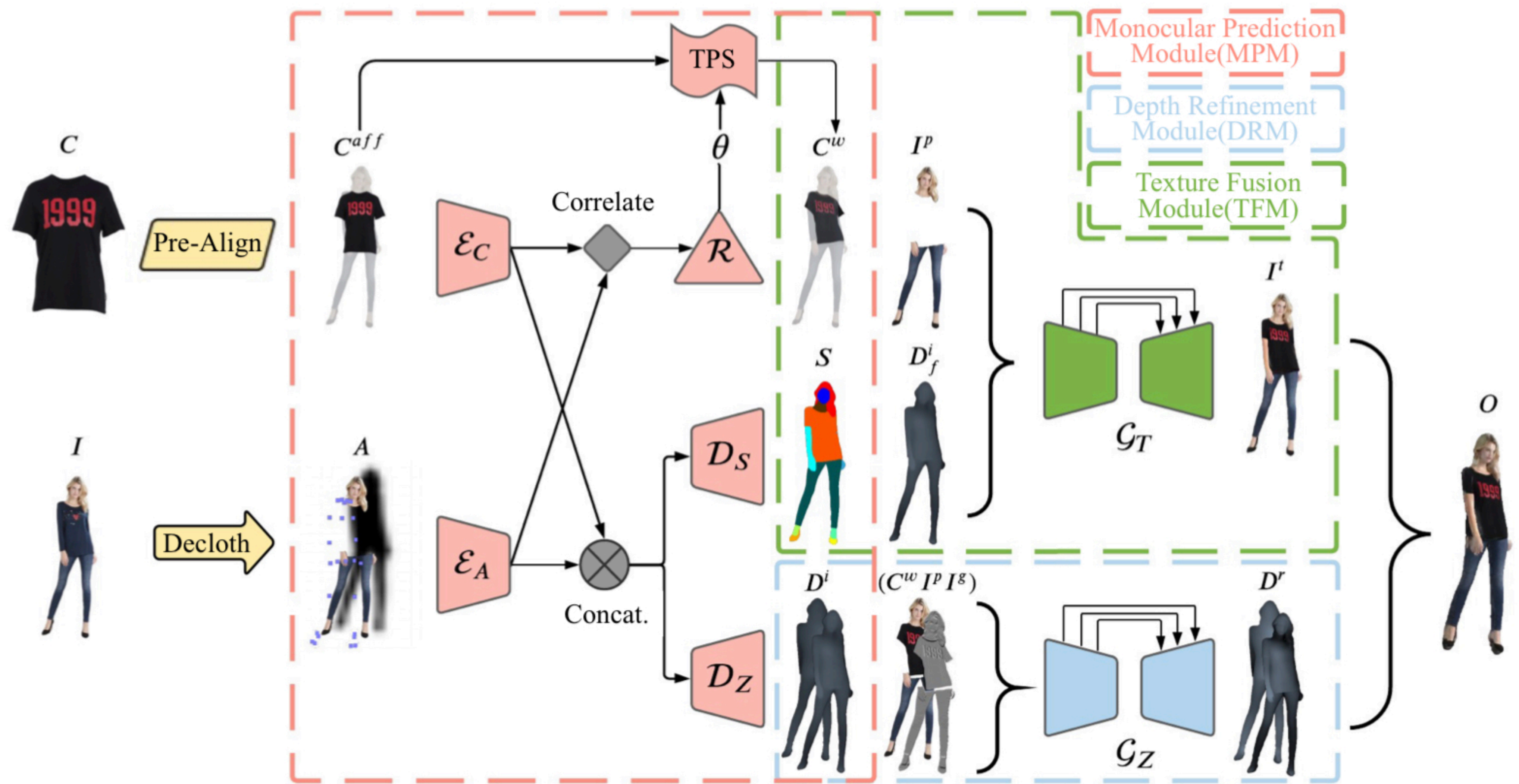


M3D-VTON Processing

전체 과정

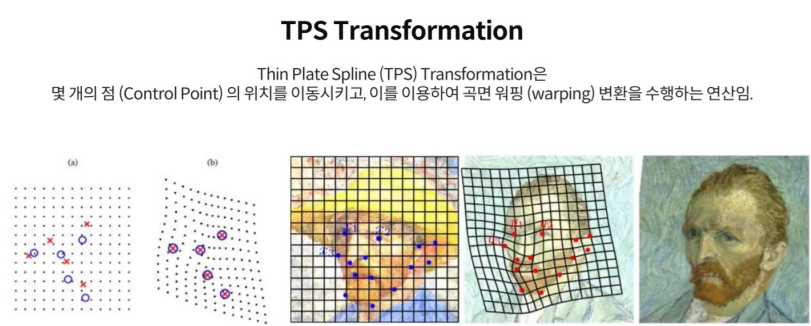
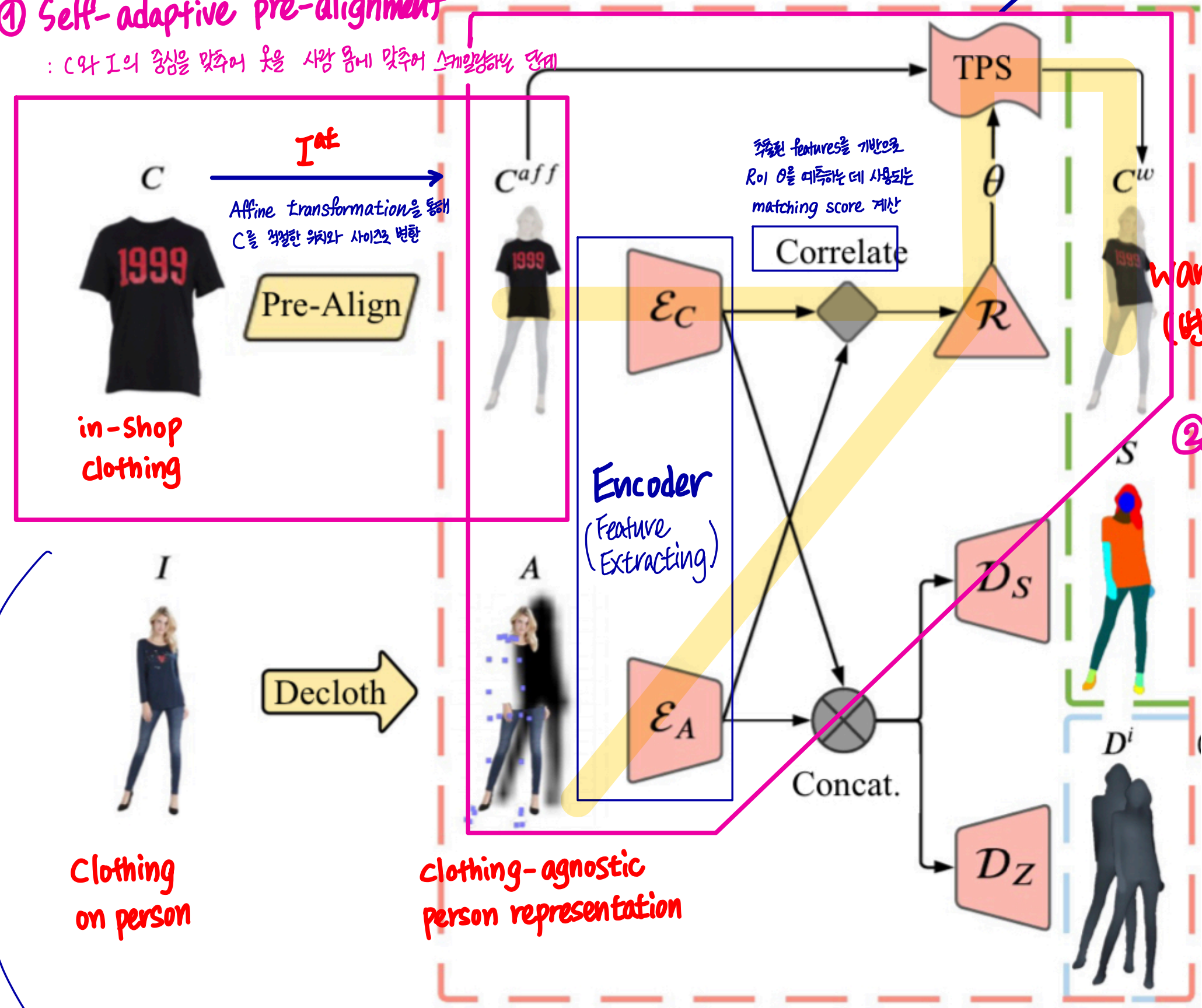


MPM(Monocular Prediction Module, 단안 예측 모듈)

1) Clothing Warping Branch, 옷 뒤틀림 분기 : 옷과 사람 몸을 맞추어 입히는 과정

① Self-adaptive pre-alignment

: C와 I의 중심을 맞추어 옷을 사람 몸에 맞추어 스케일링하는 단계



warped clothing
(변형된 옷 이미지)

$$L_w = || C_w - I ||$$

② TPS warp

: Geometric matching network

[Clothing warping branch]



① Pre-align

: 옷 중심 맞추고 사이즈 변환

② TPS warp

: cloth warping 수행해 사람에게 옷 입히기

TPS를 직접 적용하면
과도한 변형 & 제대로 안 휘어진 옷 발생
사전 정렬을 통해
안정한 변형 & 정확하게 뒤틀린 옷 획득
(사전정렬에 따라 뒤틀린 옷 & 옷 부위 사이의
IOU(얼마나 일치?) 증가)

[Affine transformation]

$$C^{aff} = \begin{bmatrix} R & 0 \\ 0 & R \end{bmatrix} C + \begin{bmatrix} x_{I^{at}}^c - x_C^c \\ y_{I^{at}}^c - y_C^c \end{bmatrix}$$

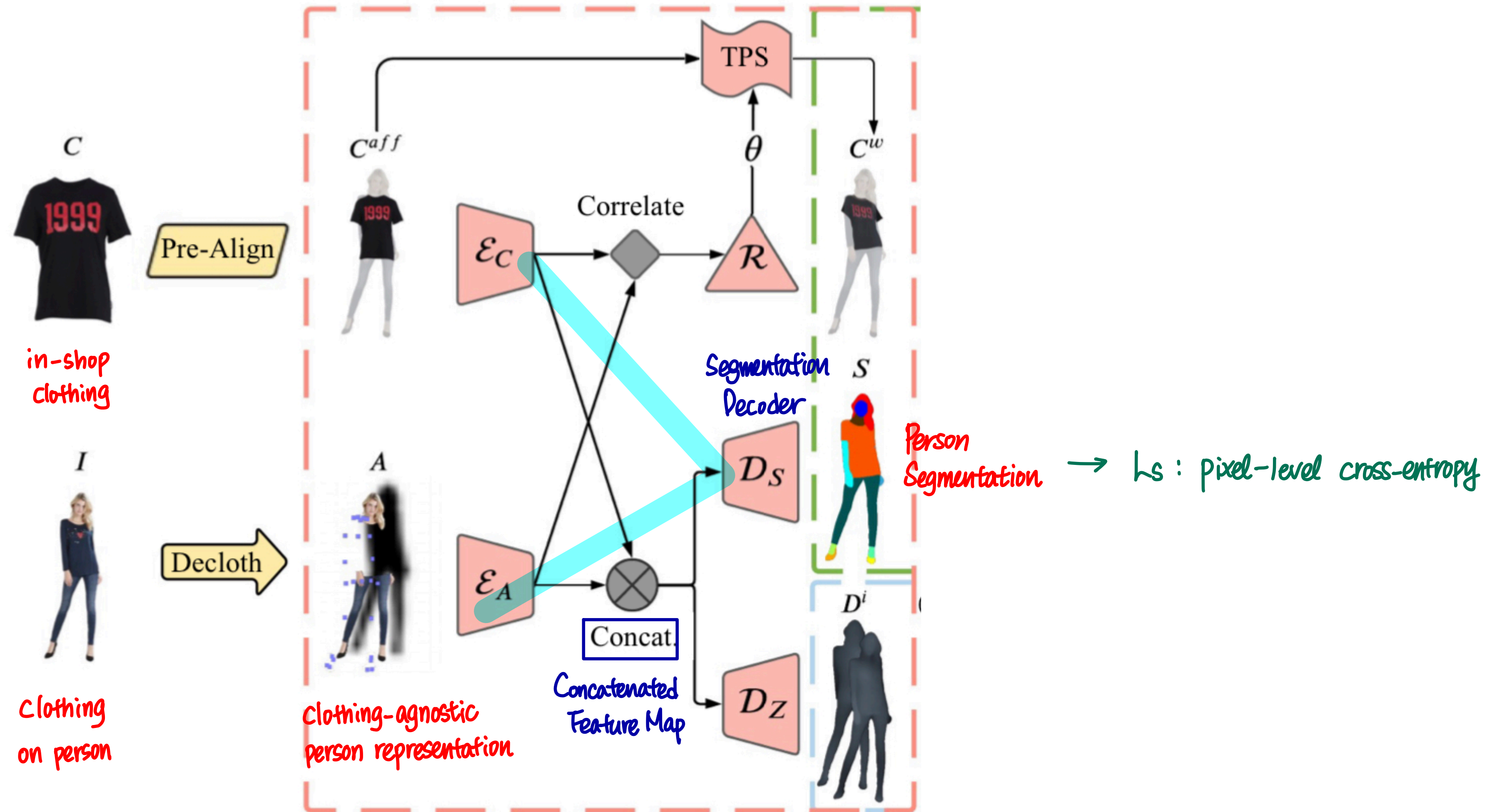
I^{at} 중심 C 중심

$$R = \begin{cases} \frac{h_I^{at}}{h_C}, & \frac{w_C}{h_C} \geq \frac{w_I^{at}}{h_I^{at}} \\ \frac{w_I^{at}}{w_C}, & \frac{w_C}{h_C} < \frac{w_I^{at}}{h_I^{at}} \end{cases}$$

정렬된 옷이 팔걸이 위치보다 크거나 같지 않게 하기 위한 가로, 세로 비율 비교

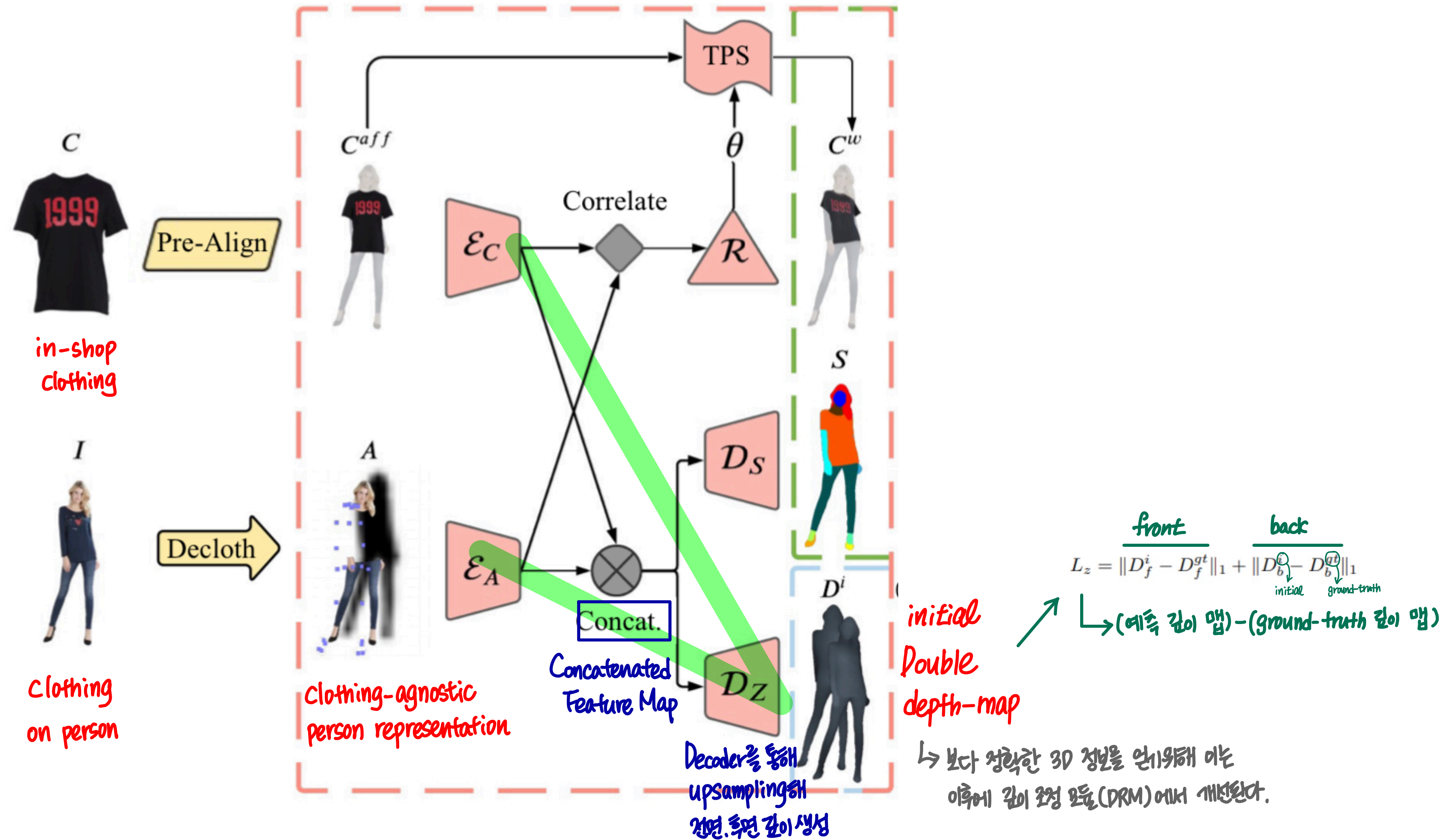
MPM(Monocular Prediction Module, 단안 예측 모듈)

2) Conditional Segmentation Estimation Branch, 조건부 분할 추정 분기 : 사람 분할(S) 예측

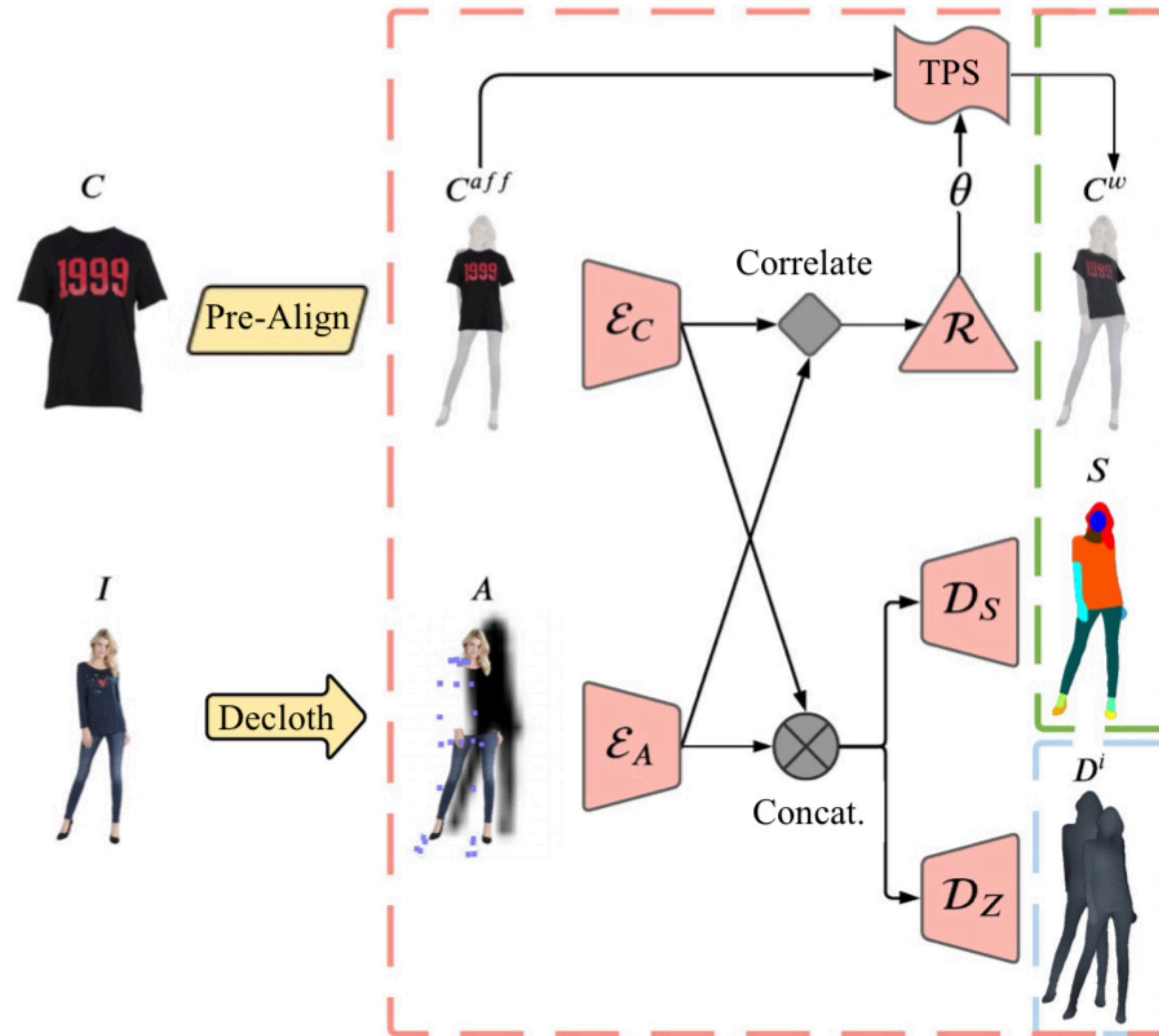


MPM(Monocular Prediction Module, 단안 예측 모듈)

3) Depth Estimation Branch, 깊이 추정 분기 : 인물의 기본 3D 형상(초기 이중 깊이맵 D_i 추정)



MPM(Monocular Prediction Module, 단안 예측 모듈)



* 다음 분들을 위한 준비작업

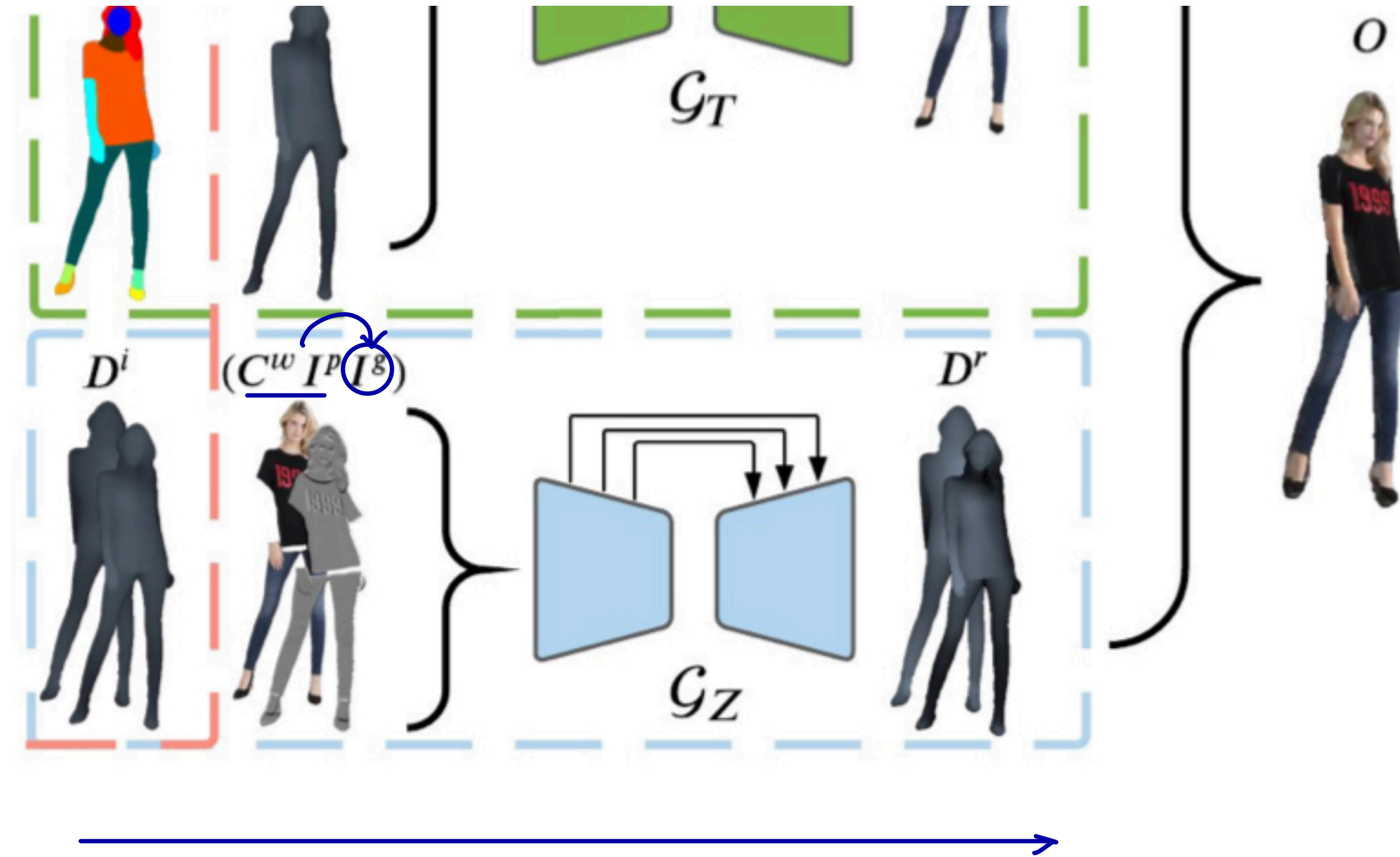
* Multi-target network에서 3branch 형태로 구현

Cloth-Warping Branch : 옷 warped
 Conditional Segmentation Estimation Branch : 사람 분할
 Depth Estimation Branch : 현재 3D 모양 베이스 예측

* Full Loss of MPM

$$L_{MPM} = L_w + L_s + L_z$$

DRM(Depth Refinement Module, 깊이 조정 모듈) : 고주파 깊이 세부 정보 추가(add high-frequency depth details)



D_i (초기 깊이맵) 개선

* I^g (이미지 기울기) & L_{grad} (기울기 제약) \Rightarrow 깊이 예측 개선 & 3D 형상의 더 복잡한 세부 사항 조각

① C^W, I^P 에 Sobel 연산자 적용

\Rightarrow 그레이스케일 이미지 concatenate 해서 I^g 획득

②

D^i (초기 깊이맵)

C^W (변형된 옷)

I^P (변형된 사람 부분)

I^g (기울기 정보)

Generator G_Z
(UNet)

$\Rightarrow D^r$

(최종 깊이맵)

옷褶, 얼굴 좌등
더 많은 디테일
생산

* Loss Function

$$\mathcal{L}_{\text{depth}} = \frac{1}{n} \sum_{i=1}^n \ln(\epsilon_i + 1)$$

초기/최종 깊이맵 픽셀들의 총합
i번째 깊이맵의 L1 Loss

$$\mathcal{L}_{\text{grad}} = \frac{1}{n} \sum_{i=1}^n (\ln(\nabla_x(\epsilon_i) + 1) + \ln(\nabla_y(\epsilon_i) + 1))$$

Sobel 연산자

\hookrightarrow 근접도를 더욱 무겁게 처벌해 복잡한 주변 세부사항에 더욱 관심을 맞춤

\hookrightarrow 깊이 추정을 더 강화하고 특히 인접한 산재 부위 경계에서 기하학적인 세부 사항을 정확하게 위해 깊이 기울기 손실 통합

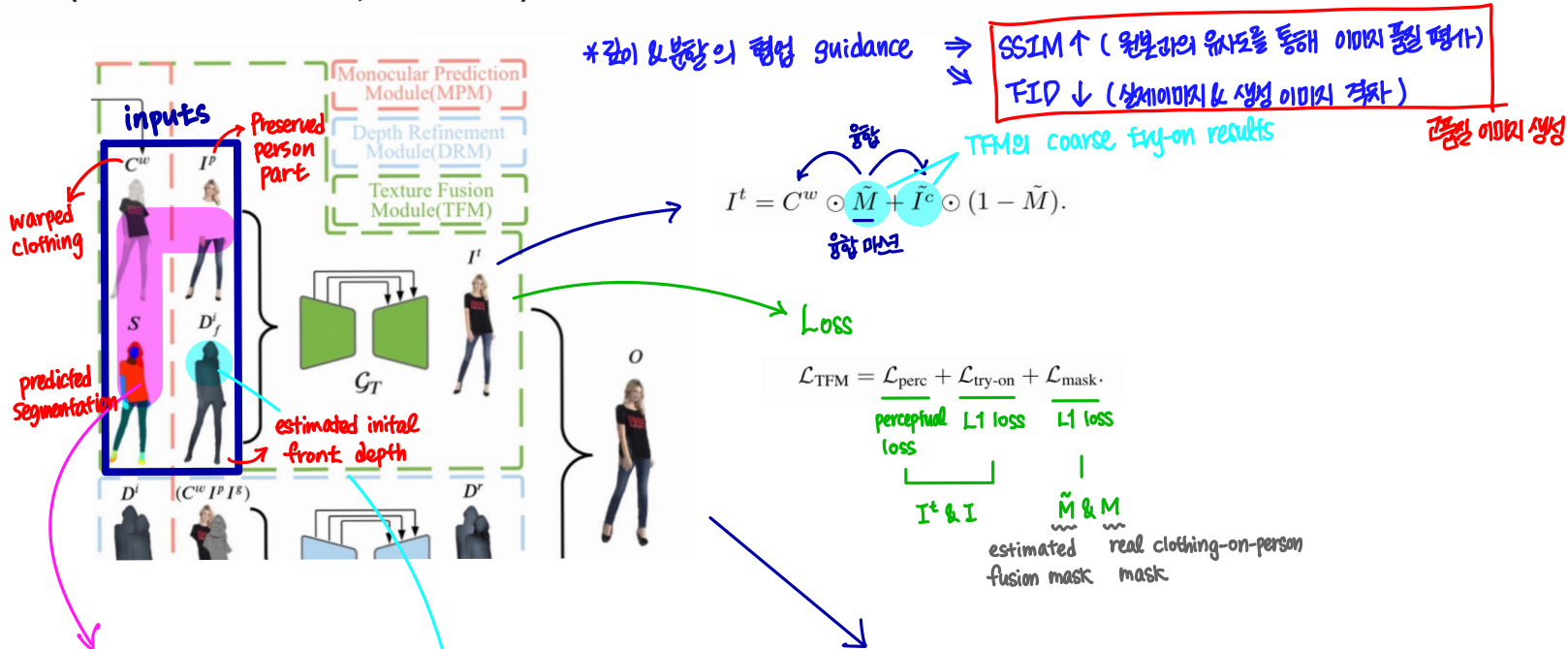
$$\mathcal{L}_{\text{DRM}} = \lambda_{\text{depth}} \mathcal{L}_{\text{depth}} + \lambda_{\text{grad}} \mathcal{L}_{\text{grad}}$$

\hookrightarrow 손실 "가중합제" 활용

\hookrightarrow 두 손실은 서로 다른 유형의 오차를 제약하는 보완 방식으로 작용

- $\left[\begin{array}{l} L_{\text{depth}} : z\text{-방향에 따른 일관성 보장} \\ L_{\text{grad}} : x, y\text{-방향에 따른 일관성 보장} \end{array} \right.$

TFM(Texture Fusion Module, 질감 융합 모듈) : 최종 3D 사람 mesh를 위한 사실적인 신체 질감 합성



2D clues provide ...

- 사람 외모
- 옷의 질감
- 분할 지판/지침

- Z축에 따른 다른 신체 부위의 공간정보
- 보다 더 정확한 형상 결과 도출

- DRM의 전면,후면 깊이맵 unproject
 \rightarrow 3D point clouds, triangulate
- TFM의 try-on 결과가 깊이맵과 공간적으로 정렬되어 있어 mesh 채색, 인페인팅을 통해 옷을 입은 3D 인물을 생성해낼 수 있다.

Conclusion

- 2D 및 3D 접근 방식의 장점을 바탕으로 2D 정보에서 3D try-on 메시를 생성하는 계산 효율적인 단안 3D 가상 트라이온 네트워크(M3D-VTON)를 제안
- 3D try-on 작업 = 2D try-on + 신체 깊이 추정
- 2단계 뒤틀림(wrapping) 전략, 분할 및 깊이 guidance를 통해 보다 사실적인 텍스처 융합 결과를 도출
- 새로운 깊이 기울기 제약 조건 도입하여 보다 자세한 깊이맵 생성
- monocular-to-3D virtual try-on 작업을 위한 더 빠르고 경제적인 해결책 제공