Position Constraint Loss for Fashion Landmark Estimation

Presenter: Meijia Song

2020. 05

北京大學
PEKING UNIVERSITY

OUTLINE

INTRODUCTION

PART ONE

METHODS

PART TWO

EXPERIMENTS

PART THREE

1 INTRODUCTION

What is Fashion Landmark Estimation?

- **Human Pose Estimation**

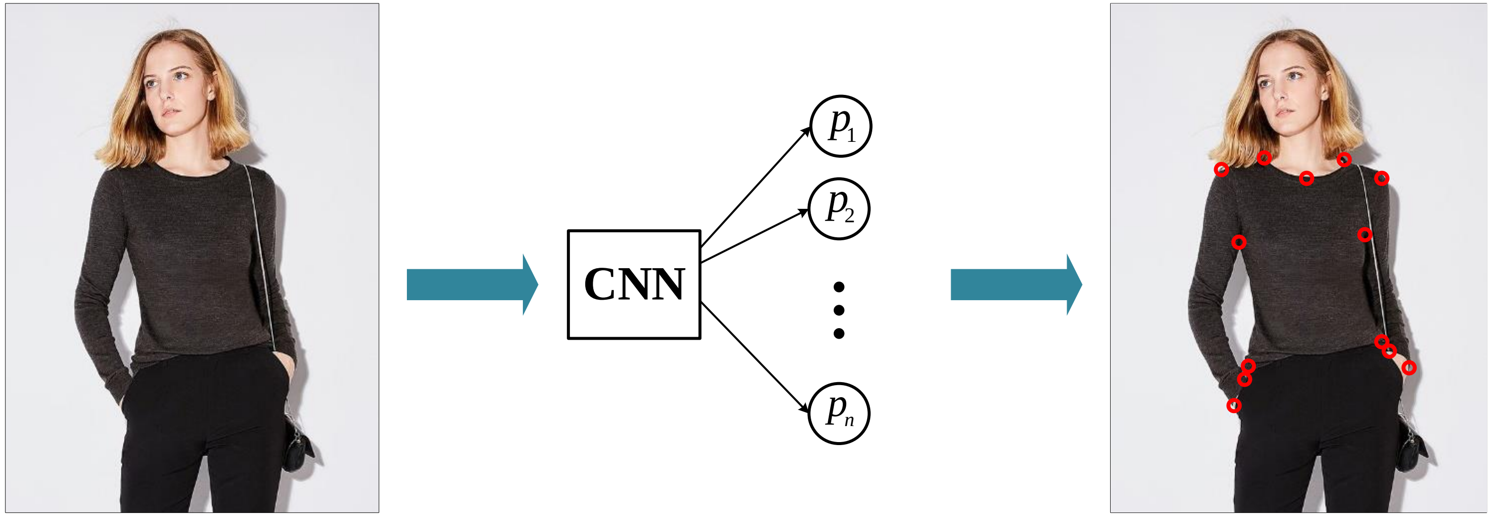
Human Joints

- **Fashion Landmark Estimation**

Clothing Landmarks

Common Methods

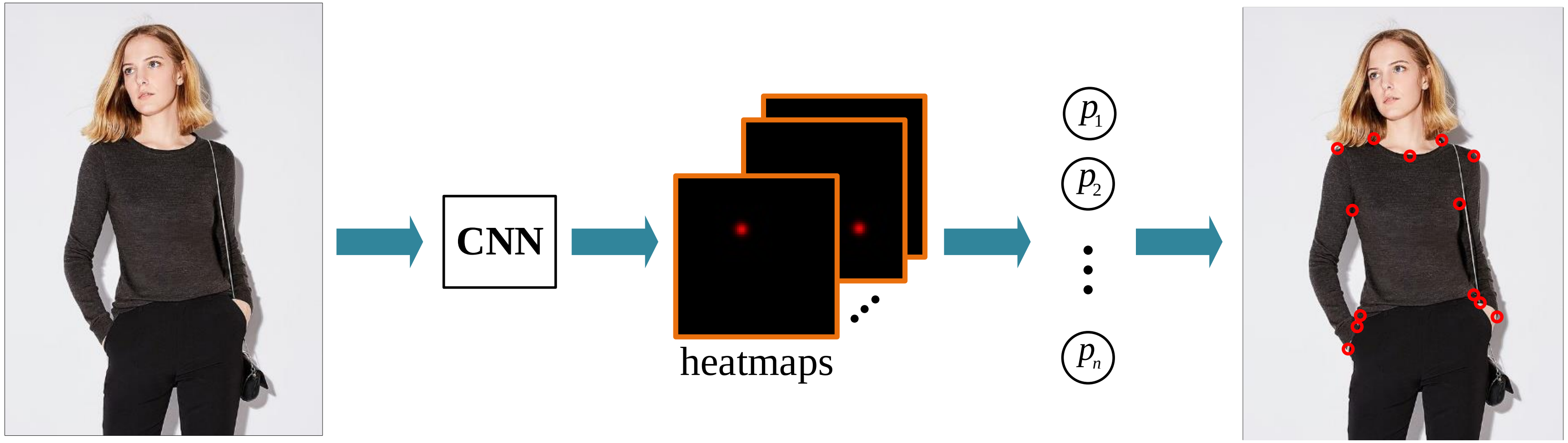
- **Regression based methods**



Obtain landmark coordinates from network outputs directly

Common Methods

- **Heatmap based methods**



- Predict the heatmap result for each landmark
- Heatmap: a confidence map of positional distribution for the landmark

Common Methods

- **Regression based methods**
  - Advantages: differentiable and can be trained end-to-end.
  - Problems: lack of spatial information, worse locating ability

- **Heatmap based methods**
  - Advantages: utilize more spatial information to obtain higher locating accuracy
  - Problems: not differentiable and exist quantization error

2

METHODS

Motivation

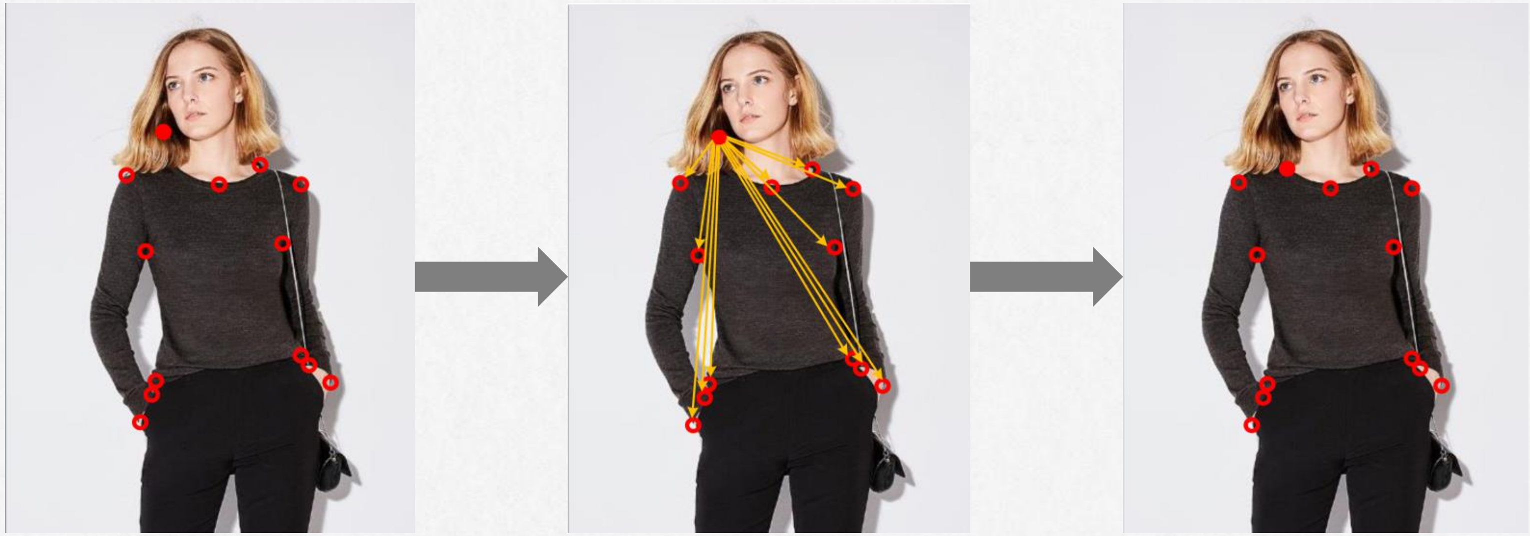
● Outliers

The estimator is confused by the high response area of background or occlusion, resulting in some outliers in the predicted result.

● Duplicate Detection

The estimator repeatedly detects a certain keypoint due to large deformation or weak outline information in clothing.

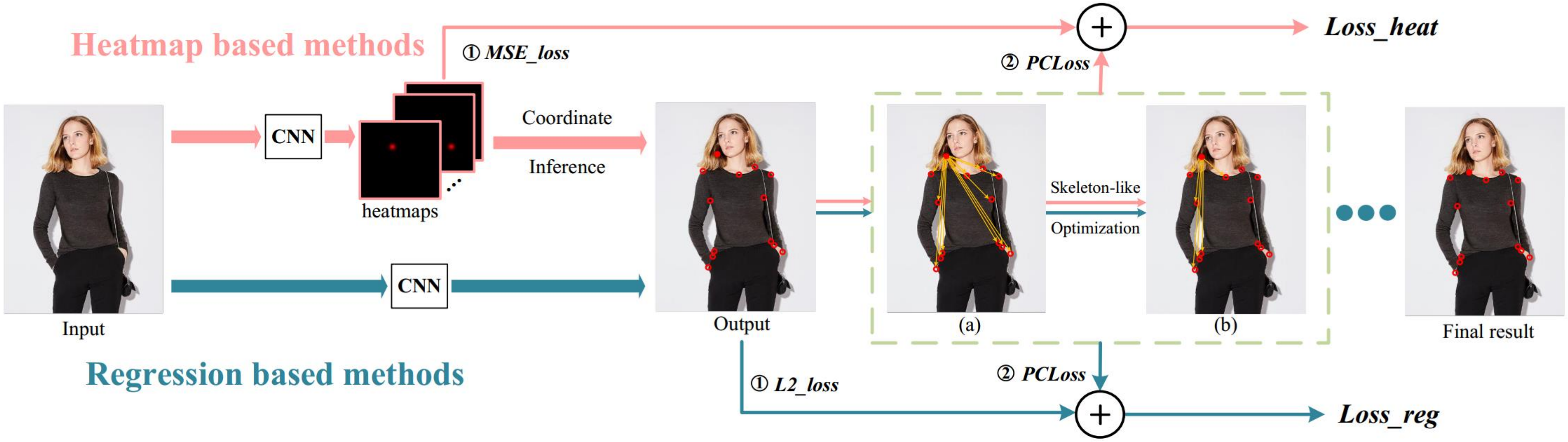
Position Constraint Loss



- To solve the problems, we proposed an efficient solution, **Position Constraint Loss (PCLoss),** to regularize relative positions of landmarks during training only.

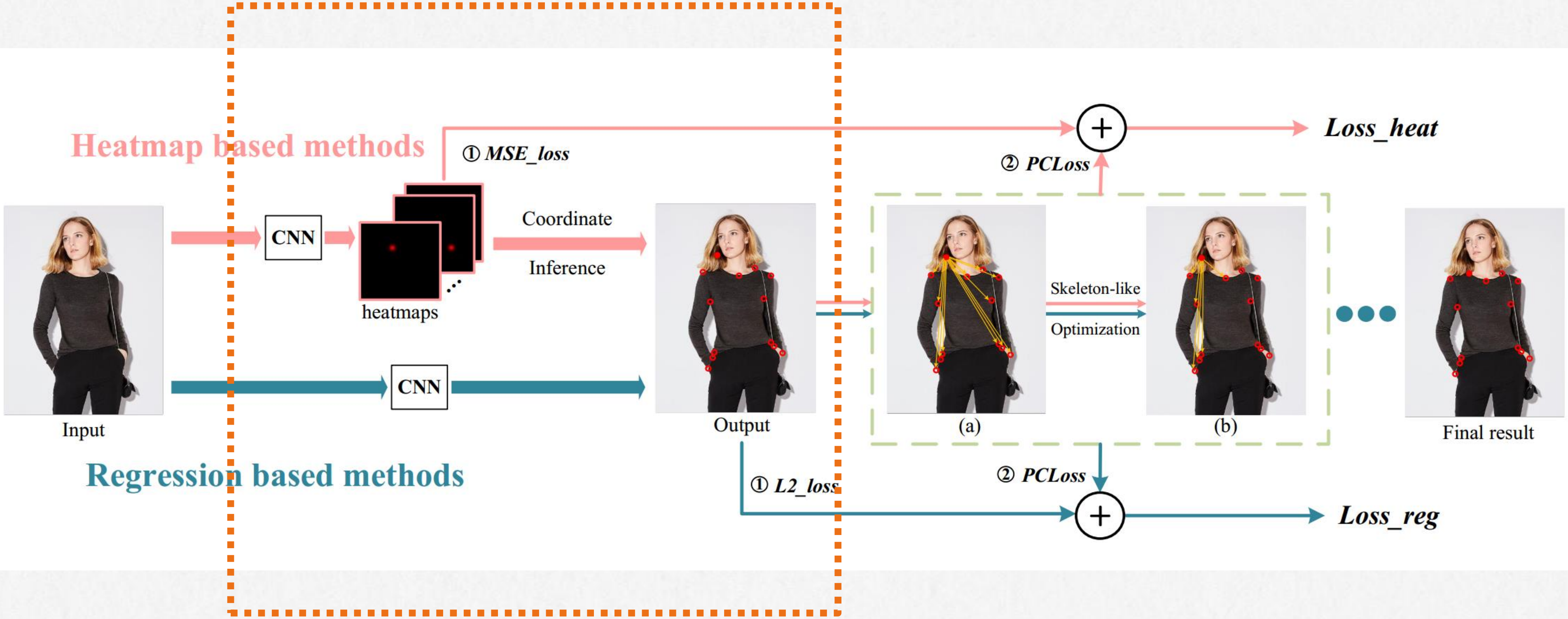
- It adds a regularization term for each landmark by **loss function** to correct error points, which can be easily applied to **both regression and heatmap based models**.

Pipeline of the Method



- We evaluate the effectiveness of PCLoss in two kind of methods: regression based methods and heatmap based methods.
- In both types of methods, landmark coordinates should be acquired first. Then, coordinates will be used to calculate PCLoss with skeleton-like optimization.

Coordinate Inference for Heatmaps



Coordinate Inference for Heatmaps

- In regression based methods, we can easily obtain landmark coordinates for PCLoss and train them end-to-end. But in heatmap based methods, coordinates can not be acquired directly.

- Argmax operation:

$$p_i = \arg\max_{l} H_i(l)$$

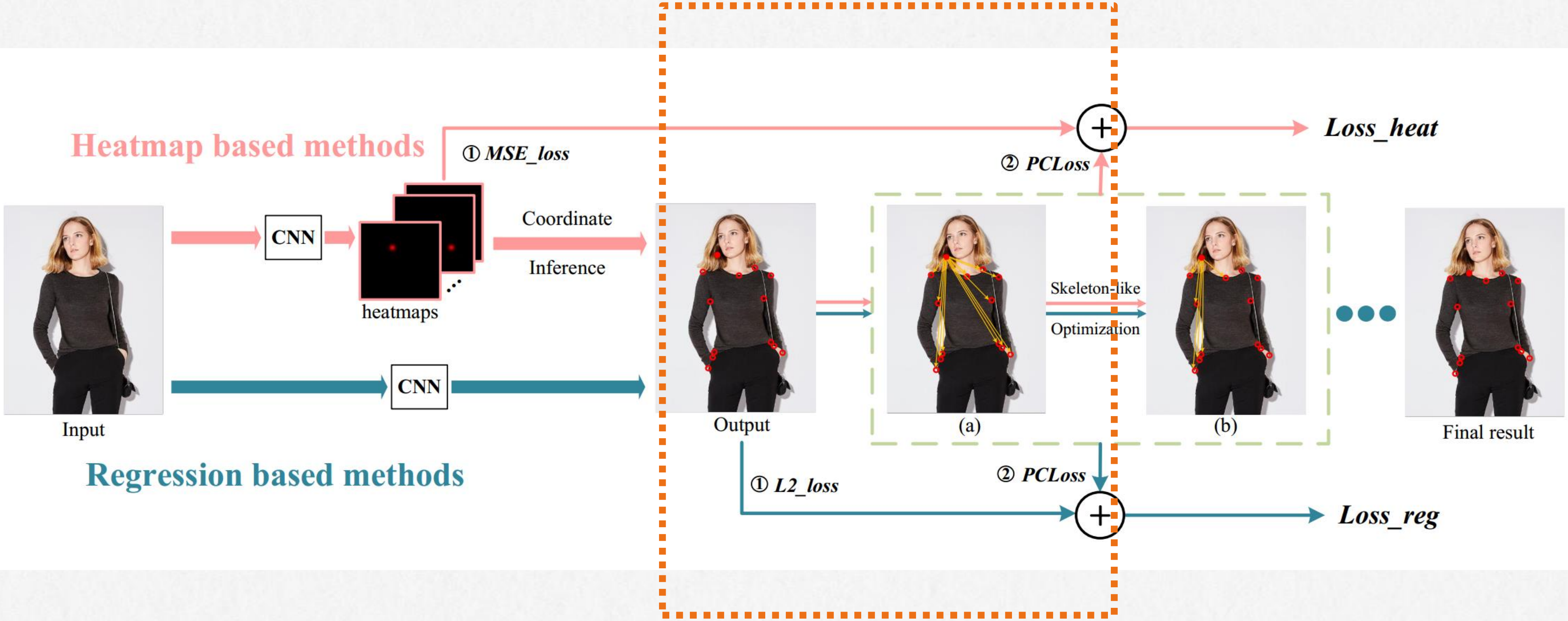
- Integral operation:

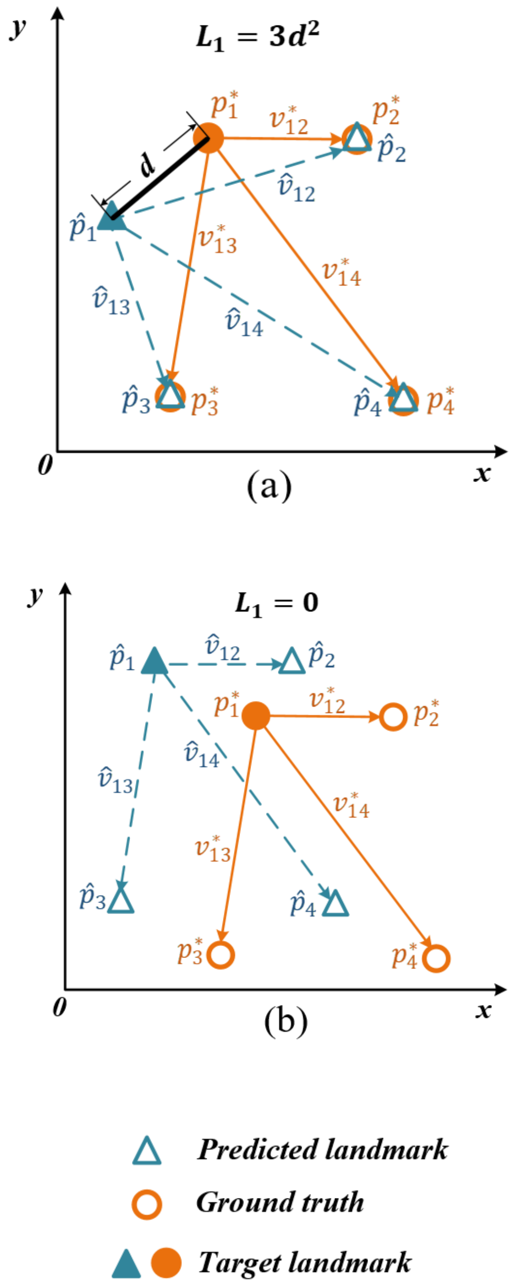
$$\tilde{H}_i(l) = \frac{e^{\alpha \cdot H_i(l)}}{\sum_k e^{\alpha \cdot H_i(k)}},$$

$$p_i = \sum_{l_y=1}^{H} \sum_{l_x=1}^{W} l \cdot \tilde{H}_i(l),$$

Reference: *" Integral human pose regression ", ECCV, 2018*

The definition of PCLoss





- Assume there are N landmarks on clothes, the goal of fashion landmark estimation is to predict the position P for all landmarks as

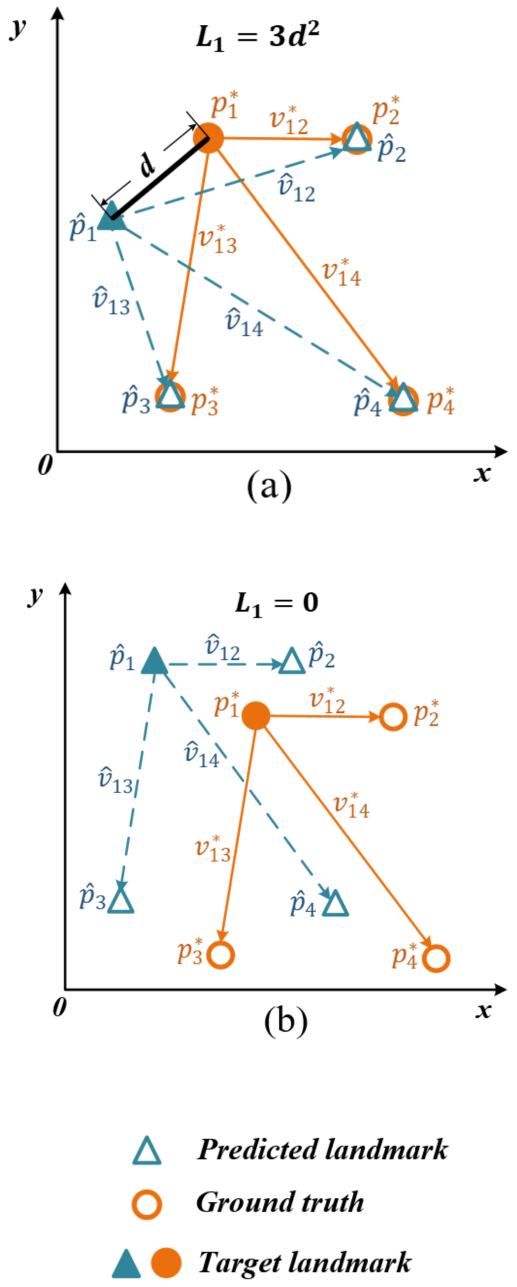
$$P = \{ p_1, p_2, \cdots, p_N \}$$

- The relative position vector between landmark i and j is given by

$$v_{ij} = p_i - p_j$$

- Then the PCLoss for the landmark i is defined as L2 loss between predicted and ground truth relative position vector, which is formulated as

$$L_i = \sum_{j=1}^{N} (\hat{v}_{ij} - v_{ij}^*)^2$$



- PCLoss for all landmarks L_{lan} can be described as

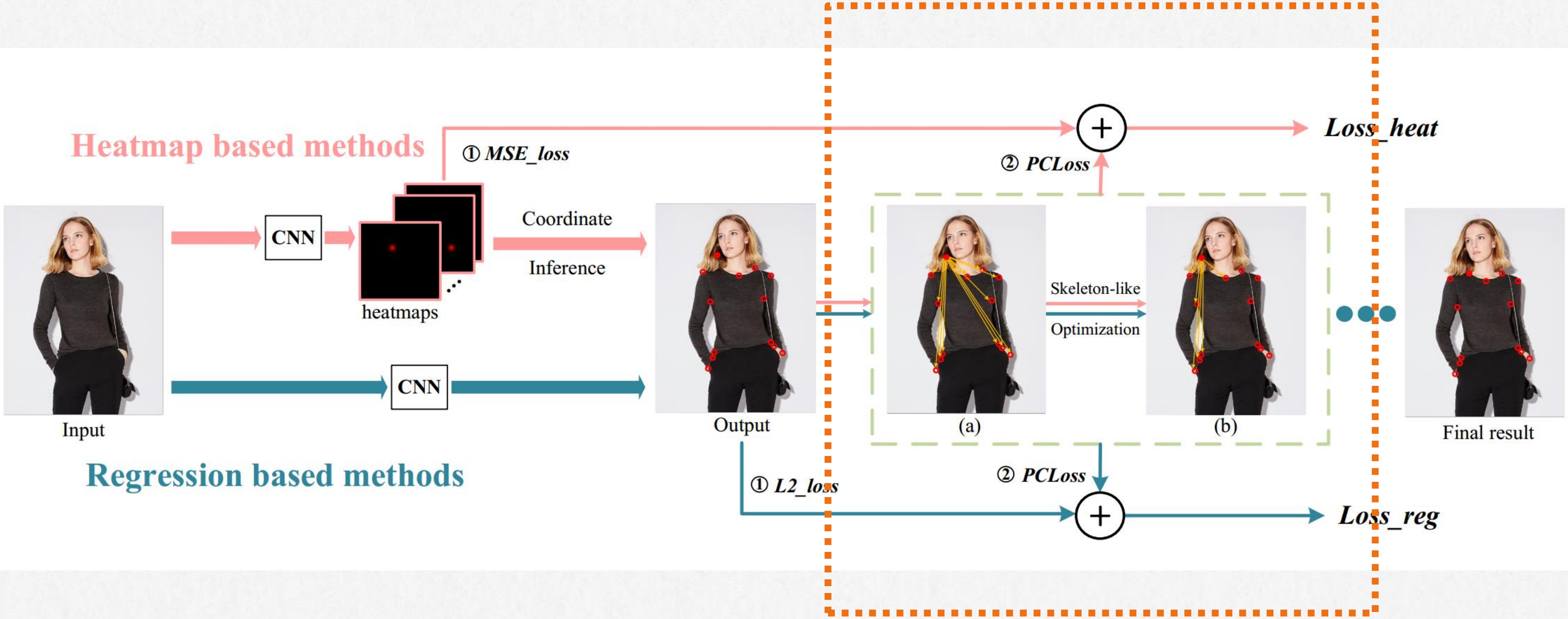
$$L_{lan} = \{ L_1, L_2, \cdots, L_N \}$$

- We only calculate k max PCLoss in L_{lan} like OHEM. Then, the final PCLoss can be formulated as

$$L_{PC} = f_k\{ L_{lan} \}$$

where $f_k\{\cdot\}$ is a function that calculates the average of k maxima of the set.

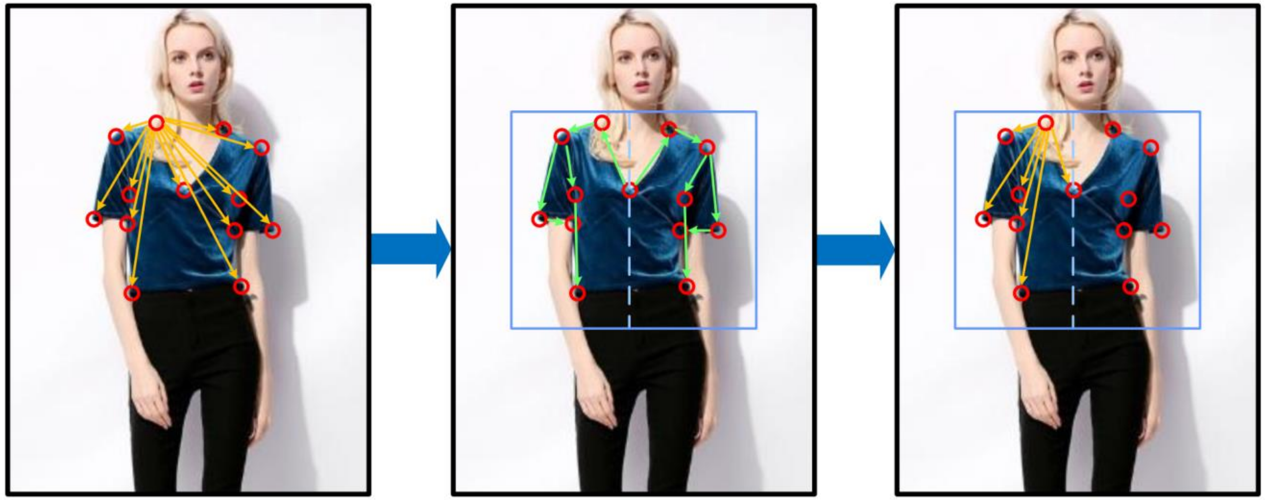
Skeleton-Like Optimization



Skeleton-Like Optimization

- Since each landmark in PCLoss is associated with all remaining points, once there are some error points, all landmarks will be affected due to position constraints.

- To reduce the influence of pose variation and error points on PCLoss, we propose the skeleton-like structural constrain mechanism, which associate the target landmark with only high-related points.



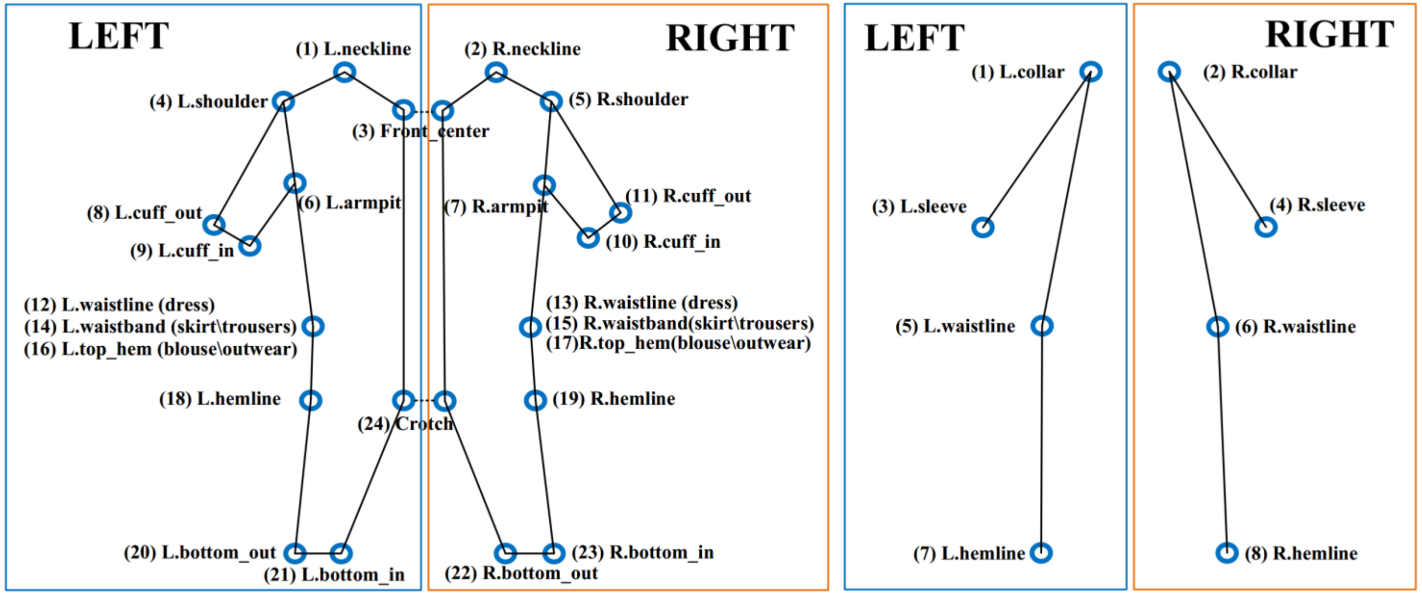
(a) (b) (c)

Skeleton-Like Optimization

- According to the skeleton-like relation, we divide fashion landmarks into two parts as follows

$$P_l = \{p_{l_1}, p_{l_2}, \ldots, p_{l_M}\}$$
$$P_r = \{p_{r_1}, p_{r_2}, \ldots, p_{r_M}\}$$

where P_l and P_r are landmark positions in left part and right part, and each part contains M landmarks.

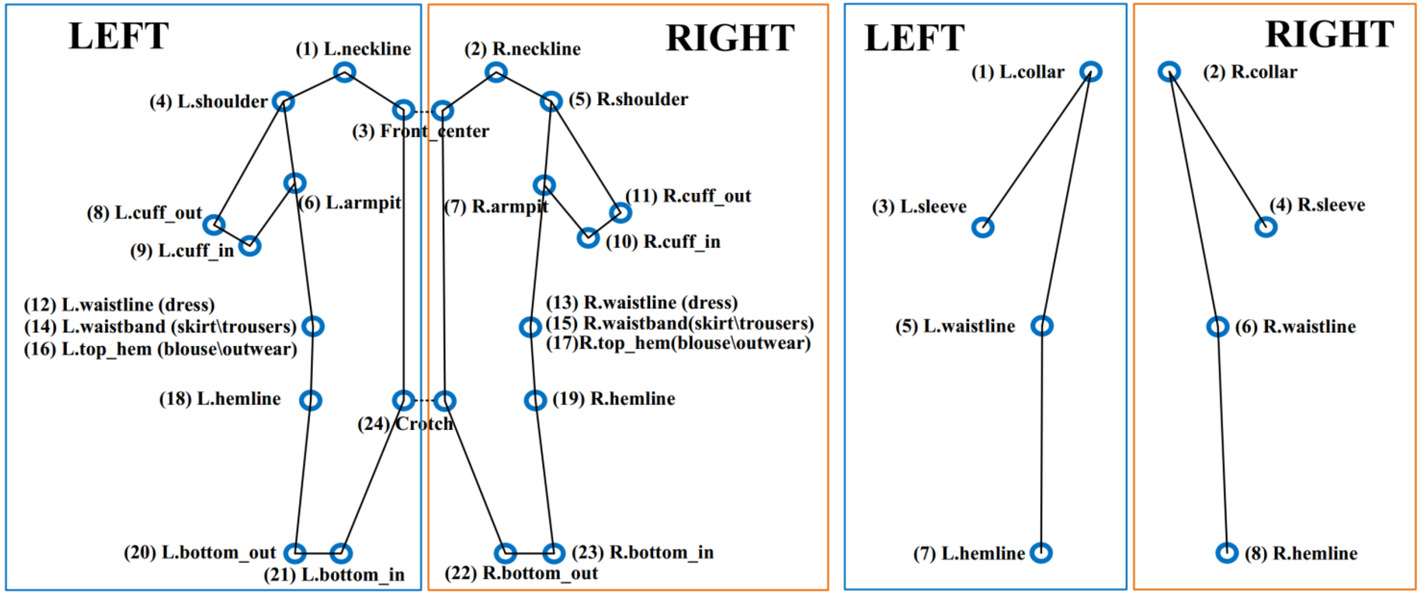


Skeleton-Like Optimization

- Then, the final PCLoss after optimization can be formulated as

$$L_{PC} = f_k\{L^l_{lan}\} + f_k\{L^r_{lan}\}$$

where L^l_{lan} and L^r_{lan} are the PCLoss for P_l and P_r .



Overall Loss Function

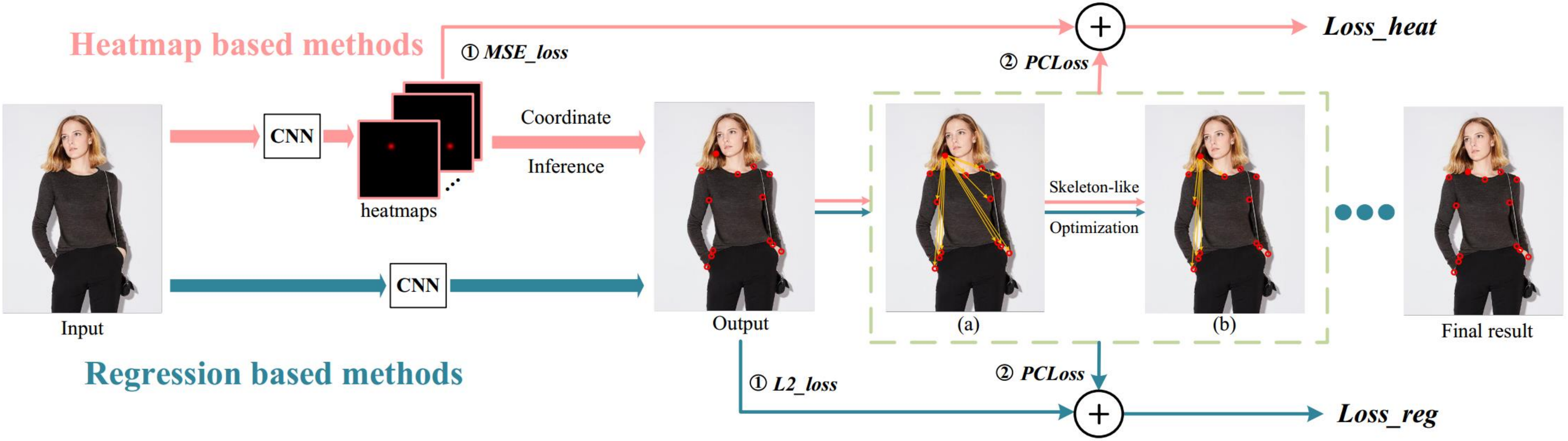
Overall Loss Function

- Regression based methods:

$$L_{reg} = \alpha \cdot L2(\hat{p}, p^*) + \beta \cdot L_{PC}(\hat{\boldsymbol{v}}, \boldsymbol{v}^*)$$

- Heatmap based methods:

$$L_{heat} = \alpha \cdot L_{MSE}(\hat{h}, h^*) + \beta \cdot L_{PC}(\hat{\boldsymbol{v}}, \boldsymbol{v}^*)$$



3

EXPERIMENTS

Experiments

Table 1. Comparison results on the DeepFashion dataset with Normalized Error

Model	L.Collar	R.Collar	L.Sleeve	R.Sleeve	L.Waist	R.Waist	L.Hem	R.Hem	NE (avg.)
FashionNet [8](reg)	8.54%	9.02%	9.73%	9.35%	8.54%	8.45%	8.12%	8.23%	8.75%
DFA [14](reg)	6.28%	6.37%	6.58%	6.21%	7.26%	7.02%	6.58%	6.63%	6.62%
DLAN [15](reg)	5.70%	6.11%	6.72%	6.47%	7.03%	6.94%	6.24%	6.27%	6.44%
Ours+FashionNet	**2.33%**	**2.40%**	**4.00%**	**4.17%**	**5.81%**	**6.01%**	**2.13%**	**2.20%**	**3.63%**
FGN [9](heat)	4.15%	4.04%	4.96%	4.49%	5.02%	5.23%	5.37%	5.51%	4.85%
FPN [29](heat)	2.19%	2.19%	3.28%	3.31%	4.81%	4.87%	2.09%	2.11%	3.11%
SPB [30](heat)	2.07%	2.10%	3.18%	3.13%	4.83%	4.83%	1.75%	1.77%	2.96%
Ours+FPN	**2.04%**	**2.05%**	**3.18%**	**3.19%**	**4.72%**	**4.81%**	**1.90%**	**1.96%**	**2.98%**
Ours+SPB	**2.03%**	**2.06%**	**3.00%**	**3.04%**	**4.64%**	**4.75%**	**1.50%**	**1.53%**	**2.82%**

Table 2. Comparison results on the FLD dataset with Normalized Error

Model	L.Collar	R.Collar	L.Sleeve	R.Sleeve	L.Waist	R.Waist	L.Hem	R.Hem	NE (avg.)
FashionNet [8](reg)	7.84%	8.03%	9.75%	9.23%	8.74%	8.21%	8.02%	8.93%	8.59%
DFA [14](reg)	4.80%	4.80%	9.10%	8.90%	–	–	7.10%	7.20%	6.98%
DLAN [15](reg)	5.31%	5.47%	7.05%	7.35%	7.52%	7.48%	6.93%	6.75%	6.73%
Ours+FashionNet	**3.86%**	**3.94%**	**7.46%**	**7.38%**	**7.70%**	**7.65%**	**5.00%**	**4.95%**	**5.99%**
FGN [9](heat)	4.63%	4.71%	6.27%	6.14%	6.35%	6.92%	6.35%	5.27%	5.83%
FPN [29](heat)	2.83%	2.89%	5.19%	5.22%	6.60%	6.57%	4.73%	4.50%	4.82%
SPB [30](heat)	2.88%	2.89%	5.11%	5.20%	6.53%	6.28%	4.48%	4.37%	4.72%
Ours+FPN	**2.83%**	**2.84%**	**5.21%**	**5.17%**	**6.56%**	**6.49%**	**4.31%**	**4.30%**	**4.71%**
Ours+SPB	**2.86%**	**2.84%**	**5.01%**	**5.05%**	**6.44%**	**6.28%**	**4.18%**	**3.95%**	**4.58%**

[1] Our methods are marked in bold. The label 'reg' means the model is based on regression methods, and 'heat' means the model is based on heatmap methods. '–' denotes the detailed results which are not released. Lower values are better.

Experiments

Table 3. Comparison results on FashionAI-val with NE

Model	NE (avg.)
FashionNet [8](reg)	10.17%
Ours+FashionNet [8]	**9.71%**
FPN [29](heat)	4.12%
SPB [30](heat)	4.10%
Ours+FPN	**3.91%**
Ours+SPB	**3.90%**

[1] Our methods are marked in bold. Lower values are better.

Table 4. Ablation studies on FashionAI-val with NE

Model	PCLoss	Skeleton	NE (avg.)	△
FashionNet	-	-	10.1724%	0.4646%
FashionNet	✓	-	9.8295%	0.1217%
FashionNet	✓	✓	**9.7078%**	-
FPN	-	-	4.1219%	0.2109%
FPN	✓	-	3.9646%	0.0536%
FPN	✓	✓	**3.9110%**	-

[1] △ denotes the difference value between our methods (marked in bold) and others.

THANK YOU