

---

**Algorithm 2** Fashion Landmark Detection

---

- 1: **Input:** Fashion image dataset
  - 2: **Output:** Fashion landmarks and bounding boxes
  - 3: **Model:** EfficientDet with modifications
  - 4:   - Backbone: EfficientNet
  - 5:   - Feature Pyramid Network: BiFPN
  - 6:   - Prediction Head: 3x3 convolution followed by convolution with filter size 3, stride 1, and padding 1
  - 7:   - Number of anchors: 9
  - 8:   - Number of classes: 13
  - 9:   - Number of landmarks: 294
  - 10:   - Loss Function: Focal Loss, Complete IoU Loss, Rooted Mean Squared Error
  - 11:
  - 12: **Procedure:**
  - 13:   1. Pretrain EfficientNet backbone
  - 14:   2. Modify EfficientDet with BiFPN structure
  - 15:   3. Design prediction head for fashion landmarks and bounding boxes
  - 16:   4. Train the model using Focal Loss, Complete IoU Loss, Rooted Mean Squared Error
  - 17:
  - 18: **Loss Function:**
  - 19:   - Focal Loss for classification:
  - 20:     
$$L_{cls} = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$
  - 21:     
$$\alpha = 0.25, \gamma = 2$$
  - 22:
  - 23:   - Complete IoU Loss for bounding box regression:
  - 24:     
$$V = \pi^2(\arctan(h) - \arctan(h_{gt}))$$
  - 25:     
$$\alpha = V \text{ if } IoU < 0.5, \text{ else } \alpha = 1 - IoU + V$$
  - 26:     
$$L_{bbox} = 1 - IoU + c_w^2 + c_h^2 + \alpha V$$
  - 27:
  - 28:   - Rooted Mean Squared Error for landmark prediction:
  - 29:     
$$v \text{ is the visibility of clothing landmarks. If } v \text{ does not exist, it is not reflected in the loss.}$$
  - 30:     
$$L_{landmark} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \text{ if } v > 0, \text{ else } L_{landmark} = 0$$
  - 31:
  - 32:   - Total Loss:
  - 33:     
$$L_{tot} = L_{cls} + L_{bbox} + \lambda_{size} L_{landmark} + \lambda_{off} L_{off}$$
  - 34:     
$$\lambda_{size} = 0.1, \lambda_{off} = 1$$
-