Algorithm 2 Fashion Landmark Detection

Input: Fashion image dataset
Output: Fashion landmarks and bounding boxes
Model: EfficientDet with modifications
- Backbone: EfficientNet
- Feature Pyramid Network: BiFPN
- Prediction Head: 3x3 convolution followed by convolution with filter size 3, stride 1, and padding 1
- Number of anchors: 9
- Number of classes: 13
- Number of landmarks: 294
- Loss Function: Focal Loss, Complete IoU Loss, Rooted Mean Squared Error

Procedure:
1. Pretrain EfficientNet backbone
2. Modify EfficientDet with BiFPN structure
3. Design prediction head for fashion landmarks and bounding boxes
4. Train the model using Focal Loss, Complete IoU Loss, Rooted Mean Squared Error

Loss Function:
- Focal Loss for classification:
  $L_{cls} = -\alpha_t(1-p_t)^\gamma \log(p_t)$
  $\alpha = 0.25, \gamma = 2$

- Complete IoU Loss for bounding box regression:
  $V = \pi^2(\arctan(h) - \arctan(h_{gt}))$
  $\alpha = V$ if IoU < 0.5, else $\alpha = 1 - IoU + V$
  $L_{bbox} = 1 - IoU + c_w^2 + c_h^2 + \alpha V$

- Rooted Mean Squared Error for landmark prediction:
  v is the visibility of clothing landmarks. If v does not exist, it is not reflected in the loss.
  $L_{landmark} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}i)^2}$ if $v > 0$, else $Llandmark = 0$

- Total Loss:
  $L_{tot} = L_{cls} + L_{bbox} + \lambda_{size}L_{landmark} + \lambda_{off}L_{off}$
  $\lambda_{size} = 0.1, \lambda_{off} = 1$