Confmix:

↳ Generate new samples by combining the regions of target image where the model is most confident.

↳ Introduced consistency loss to enforce coherent predictions

$$L_{consistency} = \| x_i - f(\hat{x}_i) \|$$

↳ motivates model to make same predictions across domains for the same object.

Steps:

1. Generate N_T set of psuedo detections on the target image & compute confidence per prediction using the model originally trained on only source data.

2. They use gaussian modelling, where they predict the mean, & variance intsead.
   Here mean is the most likely coordinate & variance is the uncertainty / how spread out the mean is.
   If variance is high, then the model is uncertain about the prediction.

3. Divide the target image in equal patches and then choose the one with the highest score to be joined with source image.

4. The consistency loss is computed by joining target pred & source pred in the same way as the mixed data

3.1 Confidence Score

$$C_{combined} = C_{det} \cdot C_{bbox}$$

$$C_{bbox} = 1 - \text{mean}\left(\text{Var}(\hat{b})\right)$$

→ As higher variance, more uncertainty so we use $1-$.

applied sigmoid to limit range b/w 0 & 1.

3.2 Mixing

↳ After dividing target image into patches, compute confidence like above & take mean of all pred.

↳ Choose region with highest score.

$$x_M = \underbrace{M^T \cdot x_S}_{\text{0 out where target patch will go.}} + \underbrace{(1 - M^T) \cdot x_T}_{\text{0 out what is not our target patch.}}$$

3.3 Loss

$$L_{box} = \frac{1}{N} \sum \left[1 - \text{mean}\left(\mathcal{N}\left(y^i \mid \hat{b}^i_{\mu}, \hat{b}^i_{\sigma^2} \right) \right) \right]$$

↓ How likely is our y^i given the μ & σ^2.

↓ captures uncertainty.

$$\mathcal{L}_T = \mathcal{L}_{det} + \underbrace{\gamma}_{\text{Hyparam}} L_{cons}$$

Confmix → Mix samples.

↳ Join local region of target image that corresponds to most confident pseudo detections with source image

↳ Apply consistently loss to adapt to the target data distribution.

Dqca

Daca

↳ Identifies region in target with highest confidence.

↳ Augment that region & then join them to compose new image.

Main idea:

↳ Generate high quality psuedo-labels from the target

↳ Use them to supervise the augmented versions of the target to allign model to target data distribution.

Steps:

1. Pass X_s through Detector to get D_s.
2. Pass X_T through Detector to get P_T.
3. Use this to get a composite of augmentations of the region with highest confidence score and get D_T.
4. Pass composite to get $\tilde{P}_T$.
5. Minimize $l_s(D_s \,\&\, h_s)$ and $l_T(D_T \,\&\, \tilde{P}_T)$ to avoid learned feature forgetting.

Why is H and W of bbox more important x & y of it?

Two neighbouring object mostly have same H & W. So, we smooth them to reduce localization error. The two objects can be located at different locations. So, we don't need to smooth x and y.

Why is the author using gaussian on the x & y coordinates?

↳ The author is using Gaussian modelling of uncertanity.

↳ Here, the model is also predicting how uncertain it is about the location.