UNO

$$D^l = \{(x_i^l, y_i^l);\ i = 1 \dots N\}$$

$$D^u = \{x_i^u;\ i = 1 \dots M\}$$ from c^u cluster.

a priori

$$y_i^l \in \mathcal{Y}^l = \{1, 2, \dots c^l\}$$

$$\Rightarrow y \in \mathcal{Y} = \{1 \dots c^l, c^l+1, \dots c^l + c^u\}$$

Network f_θ:

$$f_\theta(x) = \{p(y|x);\ y \in \mathcal{Y}\}$$

two head on top of f

- h [classification] L
- g [projection U]

$$\ell = [\ell_h, \ell_g]$$

$\mathcal{Y}$: $p = \sigma(z/\tau)$

CE loss: $\ell(x, y) = -\sum_{c=1}^{C} y_c \log(p_c)$

y: one-hot

Depend on $x \in \mathcal{D}^l$ or $x \in \mathcal{D}^u$

$$y = \begin{cases} [y^l, \mathbb{0}_{cu}] & ; x \in \mathcal{D}^l \\ [\mathbb{0}_{cl}, \hat{y}] & ; x \in \mathcal{D}^u \end{cases}$$ Disjoint.

$\hat{y}$ = most key contribution.

Multiview & Pseudo-labeling:

① $(x, y^l) \in \mathcal{D}^l$

find v_1, v_2 where $y_1 = y_2 = [y^l, \mathbb{0}_{cu}]$

② $x \in \mathcal{D}^u$

$v_1 \rightarrow \hat{y}_1$

$v_2 \rightarrow \hat{y}_2$

} pseudo-labeling.

swapped prediction task:

$$\ell(v_1, y_2) + \ell(v_2, y_1)$$ // swapped.

Obtain $\hat{y}_i$ given v_i : use the $g(\epsilon_i)$

[Avoiding degenerate solution]

Network predicted

Batch size.

$$L = \left[l_g^1, l_g^2, \ldots l_g^B \right]$$

$$\hat{Y} = \left[\hat{y}_1, \ldots \hat{y}_B \right]^T$$

// rows are unknown pseudolabel.

Solve the following to get $\hat{Y}$

hyperparam.

$$\hat{Y} = \max_{Y \in \Gamma} \text{Trace}(YL) + \epsilon H(Y)$$

Entropy

scatter the pseudo label

can be fixed ?!

by alter Y & L

$$\Gamma = \{ Y \in \mathbb{R}_{+}^{C^u \times B} \mid Y 1_B = \frac{1}{C^u} 1_{C^u} \; ; \; Y^T 1_{C^u} = \frac{1}{B} 1_B \}$$

{ is there a dimension mismatch ??

En force each cluster $\frac{B^u}{C^u}$ times. → $\hat{y}_i$

↙

use in the swaped assignment.