## A. Algorithms

We present the DC-SSL algorithms with training-free and training-based strategies in Algorithms 1 and 2. Due to their simplicity, both strategies are easy to be implemented with only minor changes to FixMatch.

**Algorithm 1** DC-SSL with Training-free strategy

**Input**: labeled batch $\mathcal{X} = \{(x_b, y_b)|(x_b, y_b) \in D_x\}_{b=1}^{B}$, unlabeled batch $\mathcal{U} = \{u_b|u_b \in D_u\}_{b=1}^{\mu B}$
**Parameter**: unlabeled ratio $\mu$, momentum coefficient $\alpha$, threshold $\tau$, loss weight $\lambda_u$

```
Calculate supervised loss $\mathcal{L}_x$ on $\mathcal{X}$ using Eq. (2).
for $b = 1$ to $\mu B$ do
    Obtain $f$'s predictions $p_b^{w,f}$ and $p_b^{s,f}$ for $u_b^w$ and $u_b^s$
    Obtain $g$'s prediction $p_b^{w,g}$ for $u_b^w$
end for
Estimate the PCD $q^f$ using Eq. (4)
Estimate the RCD $q^g$ using Eq. (5)
Obtain revised pseudo-labels $\bar{p}_b^{w,f}$ using Eq. (6)
Calculate consistency loss $\mathcal{L}_u$ using Eq. (7)
return $\mathcal{L}_x + \lambda_u \mathcal{L}_u$
```

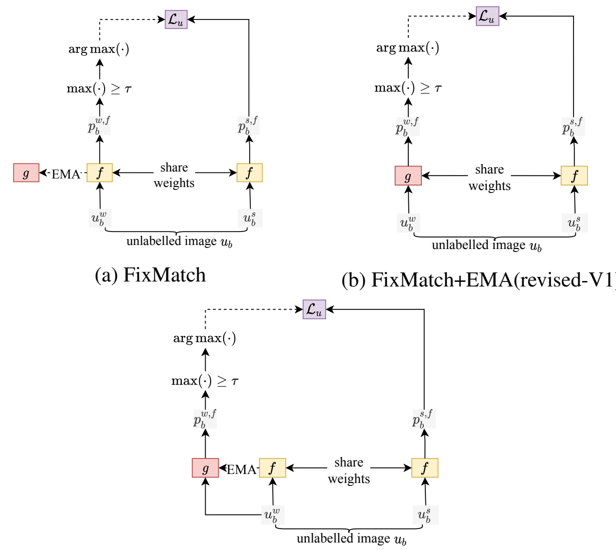
**Algorithm 2** DC-SSL with Training-based strategy

**Input**: labeled batch $\mathcal{X} = \{(x_b, y_b)|(x_b, y_b) \in D_x\}_{b=1}^{B}$, unlabeled batch $\mathcal{U} = \{u_b|u_b \in D_u\}_{b=1}^{\mu B}$
**Parameter**: unlabeled ratio $\mu$, momentum coefficient $\alpha$, threshold $\tau$, loss weight $\lambda_u$ and $\lambda_d$

```
Calculate supervised loss $\mathcal{L}_x$ on $\mathcal{X}$ using Eq. (2).
for $b = 1$ to $\mu B$ do
    Obtain $f$'s predictions $p_b^{w,f}$ and $p_b^{s,f}$ for $u_b^w$ and $u_b^s$
    Obtain $g$'s prediction $p_b^{w,g}$ for $u_b^w$
end for
Estimate the PCD $q^f$ using Eq. (4)
Estimate the RCD $q^g$ using Eq. (5)
Calculate instance-consistency loss $\mathcal{L}_u$ using Eq. (9)
Calculate distribution-consistency loss $\mathcal{L}_d$ using Eq. (8)
return $\mathcal{L}_x + \lambda_u \mathcal{L}_u + \lambda_d \mathcal{L}_d$
```

## B. EMA model for pseudo-label generations

In this section, we investigate two different ways to using the EMA model's predictions as pseduo-labels, as shown in Figs. 5b and 5c. The revised-V1 directly replaces the $f$ with $g$ to produce pseudo-labels, while the revised-V2 retains the FixMatch structure and generate pseudo-labels on $g$. As shown in Tab. 7, the revised-V2 can achieve relatively higher test accuracy, simply because it ensures $f$ can also view the weakly-augmented images, which can slightly benefit the Batch-normalization layers of $f$. However, as we have discussed in Sec. 3.2.1, the accuracy in both ways decreases sharply compared to the original FixMatch.




(a) FixMatch (b) FixMatch+EMA(revised-V1) (c) FixMatch+EMA(revised-V2)

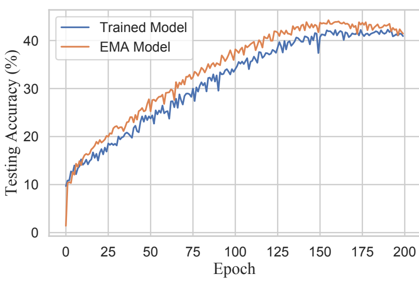
Figure 5. (a) shows the standard structure of FixMatch while (b) and (c) show two different ways to use the EMA's predictions as pseudo-labels on top of FixMatch, respectively.

| Method | Accuracy (%) | Pseudo-labels | |
|---|---|---|---|
| | | Quality (%) | Quantity (%) |
| FixMatch | 82.5 | 83.4 | 96.9 |
| revised-V1 | 43.6 | 44.7 | 93.1 |
| revised-V2 | 45.31 | 46.3 | 90.1 |

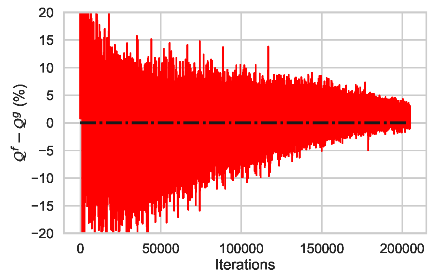
Table 7. Test accuracy on different ways to use the EMA model's prediction as pseudo-labels. The model configurations of "revised-V1" and "revised-V2" are shown in Figs. 5b and 5c, respectively. "Quality" represents the accuracy of high-confidence pseudo-labels while "Quantity" represents the ratio between the amount of high-confidence pseudo-labels to that of total pseudo-labels.

## C. Benefits of the EMA model

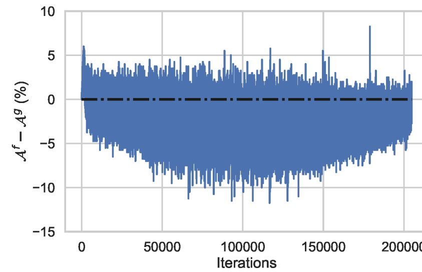
In our paper, based on the analysis on CIFAR10 with 40 labels, we observe that **the EMA model can achieve a higher accuracy of pseudo-labels on all unlabeled data but a lower accuracy on high-confidence ones**. As shown in Fig. 6, we find the same observations on MiniImageNet with 1000 labels. In addition, we investigate the performance of the EMA model in a mismatched distribution setting on CIFAR-10 with $|D_x| = 40$ and $\gamma_u = 50$. It can been seen from Fig. 7 that $g$ can outperform $f$ throughout the training process in terms of the accuracy on all pseudo-labels, yet with lower accuracy on high-confidence ones.




(a) Test accuracy

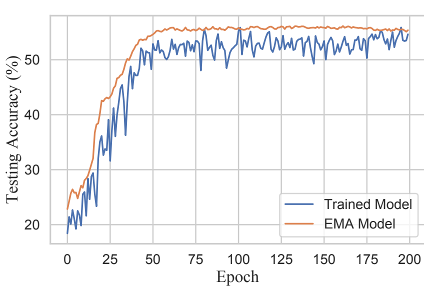



(b) Accuracy difference on high-confidence pseudo-labels

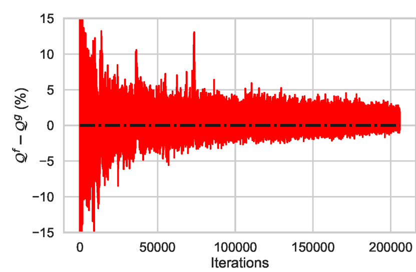



(c) Accuracy difference on all pseudo-labels

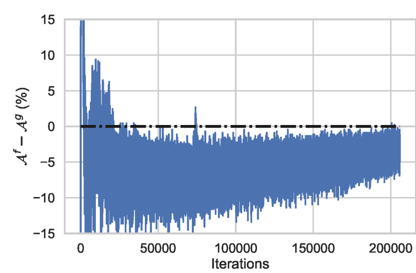
Figure 6. Following the explorations in Fig. 4, we observe same findings on MiniImageNet with 1000 labels. (b) In terms of the accuracy difference ($Q^f - Q^g$) of the high-confidence pseudo-labels in a mini-batch, $g$ obtains a lower accuracy than $f$ at about 61% iterations. (c) However, the accurate difference ($\mathcal{A}^f - \mathcal{A}^g$) of all pseudo-labels between $f$ and $g$ shows that the model $g$ can generate more accurate pseudo-labels in 88% iterations.




(a) Test accuracy




(b) Accuracy difference on high-confidence pseudo-labels




(c) Accuracy difference on all pseudo-labels

Figure 7. Following the explorations in Fig. 4, we investigate the EMA model's performance on CIFAR10 in a mismatched distribution setting as in Fig. 1b. (b) In terms of the accuracy difference ($Q^f - Q^g$) of the high-confidence pseudo-labels in a mini-batch, $g$ obtains a lower accuracy than $f$ at about 67% iterations. (c) However, the accurate difference ($\mathcal{A}^f - \mathcal{A}^g$) of all pseudo-labels between $f$ and $g$ shows that the model $g$ can generate more accurate pseudo-labels in 97% iterations.