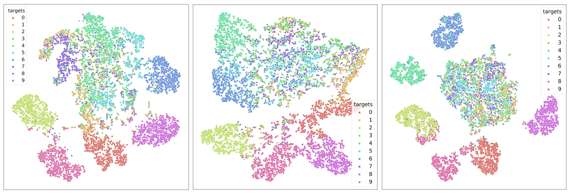
Qualitative analysis

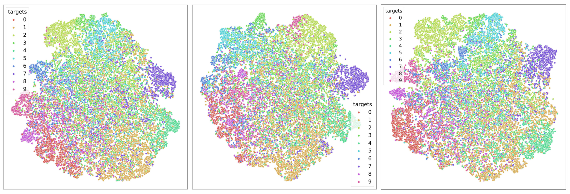
We present the T-SNE visualization of features on the STL-10 test dataset with a 40-label split in Figure 12a,12b,12c. The visualization are using trained models from FixMatch, FlexMatch, and SequenceMatch. SequenceMatch shows better feature space than FixMatch and FlexMatch with less confusing clusters.

We also visualize the T-SNE visualization of features on the SVHN test dataset and CIFAR-10 test dataset with a 40-label split in Figure 13a,13b,13c and Figure 14a,14b,14c, respectively.



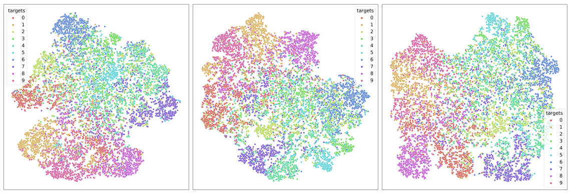
(a) FixMatch (b) FlexMatch (c) SequenceMatch

Figure 12. T-SNE visualization on STL-10 dataset with 40 labels.



(a) FixMatch (b) FlexMatch (c) SequenceMatch

Figure 13. T-SNE visualization on SVHN dataset with 40 labels.



(a) FixMatch (b) FlexMatch (c) SequenceMatch

Figure 14. T-SNE visualization on CIFAR-10 dataset with 40 labels.

Hyperparameters setting

For reproduction, we show the detailed hyperparameter setting for each method in Table 6 and Table 7, for algorithm-dependent and algorithm-independent hyperparameters, respectively.

Ablation study on KL loss

The additional medium augmentation requires the model to adjust to the new distribution. This is not the case with strong augmentation since the strongly augmented sample is heavily distorted, making it impossible to retrieve relevant information. As a consequence, SequenceMatch without and with KL loss obtains 5.01% and **4.80%** for CIFAR-10-40, respectively. Furthermore, after 150k iterations, the pseudo-label accuracy is 81.1%, 81.50%, and **83.20%** for FixMatch, FlexMatch, and SequenceMatch, respectively. SequenceMatch enhances the pseudo-label accuracy while improving the hard-to-learn class-wise accuracy simultaneously. This clearly demonstrates that employing medium augmentation and KL loss can reduce divergence and eliminate a confirmation bias.

Detailed results

We also report the mean error rates of the last 20 checkpoints for various methods in Table 10. It can be seen that while most of the algorithms are overfitting to the training data at the end of the training process, our proposed method still maintains its robustness.

ImageNet detailed results

In this section, we show the detailed results of Table 5 for ImageNet dataset on 10% labeled data. We could see that SequenceMatch outperforms previous methods in both scenarios where self-supervised pre-trained weights are included or not.

Ablation study on medium augmentation

For the medium one, we conduct ablation studies on various types of augmentation and report the results in Table 9. We systematically test different types of augmentation for medium one, such as MoCo [15], SimCLR [8], CTAugment [3], and a combination of weak augmentation with 1 or 2 randomly selected strong augmentation. As can be seen in Table 9, using the combination of weak augmentation with 1 random strong augmentation results in the best performance.

List of data transformations

We used the same sets of image transformations used in FixMatch [44]. For completeness, we listed all transformation operations for these augmentation strategies in Table 12.

We visualize the weak, medium, and strong augmentation examples in Figure 15 for a better understanding of the differences among the three augmentations. As we can see, the mediumly augmented examples are different from the weakly augmented ones but they are not heavily distorted like the strongly augmented ones.

Table 6. Algorithm dependent parameters. **'F-Match'** indicates FixMatch, FlexMatch, and FreeMatch.

Algorithm	UDA	SequenceMatch	F-Match
Unlabeled Data to Labeled Data Ratio (CIFAR-10/100, STL-10, SVHN)	7	7	7
Unlabeled Data to Labeled Data Ratio (ImageNet)	-	1	1
Pre-defined Threshold (CIFAR-10/100, STL-10, SVHN)	0.8	0.95	0.95
Pre-defined Threshold (ImageNet)	-	0.7	0.7
Temperature	0.4	0.5	-

Table 7. Algorithm independent parameters.

Dataset	CIFAR-10	CIFAR-100	STL-10	SVHN	ImageNet
Model	WRN-28-2	WRN-28-8	WRN-37-2	WRN-28-2	ResNet-50
Weight Decay	5e-4	1e-3	5e-4	5e-4	3e-4
Batch Size	64				128
Learning Rate	0.03				
SGD Momentum	0.9				
EMA Momentum	0.999				
Unsupervised Loss Weight	1				

Table 8. KL loss and results on CIFAR-10 with 40-label split.

	with KL loss	Top-1
SequenceMatch	×	5.01
	√	4.80

Table 9. Augmentation results on CIFAR-10 with 40-label split.

Method	Top-1
Weak augmentation + 1 random strong augmentation	**4.80**
Weak augmentation + 2 random strong augmentation	4.91
MoCo augmentation [15]	5.97
CTAugment [3]	4.85
SimCLR augmentation [8]	5.32

Table 10. Mean error rates of last 20 checkpoints of all methods. There are 1000 iterations between every two checkpoints.

Dataset	CIFAR-10			CIFAR-100			SVHN			STL-10	
# Label	40	250	4000	400	2500	10000	40	250	1000	40	1000
UDA	10.65±4.97	5.67±0.28	4.58±0.07	99.0±0.0	99.0±0.0	99.0±0.0	2.5±0.54	2.06±0.02	2.01±0.03	90.0±0.0	34.88±38.98
MPL	8.2±1.9	8.7±1.22	4.78±0.03	48.72±0.46	29.02±0.46	22.39±0.38	11.06±6.45	2.45±0.08	2.29±0.06	44.63±7.16	7.51±0.19
MixMatch	51.5±17.51	22.14±2.83	66.57±5.38	95.87±0.24	97.88±0.36	99.0±0.0	48.86±14.71	10.16±2.7	30.09±2.42	64.99±2.3	59.5±2.48
ReMixMatch	8.5±0.6	6.59±0.18	4.97±0.13	42.1±1.35	26.19±0.15	**20.57±0.14**	21.41±12.26	10.69±0.73	11.44±1.91	34.12±5.27	6.99±0.08
FixMatch	12.85±4.51	5.26±0.08	4.43±0.02	48.87±2.48	28.84±0.4	22.93±0.14	3.5±1.05	2.06±0.01	2.11±0.02	46.71±5.25	6.14±0.25
FlexMatch	5.53±0.28	5.24±0.08	4.49±0.04	47.56±2.68	27.62±0.11	22.88±0.17	18.55±8.46	19.17±4.63	12.93±1.85	51.15±15.35	6.34±0.37
Dash	9.96±3.45	5.38±0.29	4.6±0.12	50.37±1.77	28.61±0.38	22.85±0.15	5.39±2.05	2.08±0.02	2.16±0.09	44.19±6.07	6.61±0.52
CoMatch	7.2±1.77	5.64±0.17	4.52±0.24	60.43±8.27	31.41±0.2	23.94±0.28	13.63±5.27	3.16±0.95	2.1±0.03	**17.88±5.09**	6.07±0.02
SimMatch	5.55±0.03	5.51±0.06	4.64±0.06	42.17±0.62	30.2±0.17	23.77±0.13	14.4±0.31	3.89±2.45	2.15±0.05	27.95±6.5	6.39±0.56
AdaMatch	5.33±0.22	5.34±0.05	4.71±0.02	**40.19±1.63**	28.08±0.39	22.91±0.18	11.49±3.52	2.22±0.06	2.12±0.08	36.46±5.49	6.43±0.13
SequenceMatch	**5.03±0.11**	**5.07±0.11**	**4.43±0.02**	44.52±1.01	**27.16±0.23**	22.90±0.16	**2.01±0.43**	**1.89±0.01**	**1.86±0.01**	40.21±6.11	**5.88±0.14**

A. Algorithm

We present the complete algorithm for SequenceMatch in Algorithm 1.

Table 11. Accuracy results on ImageNet with 10% labeled examples using [24] and [54] source code.

Self-supervised Pre-training	Method	Top-1	Top-5	Params (train/test)	Epochs
None	FixMatch	71.5	89.1	25.6M/25.6M	$\sim$ 300
MoCo-EMAN [6]	FixMatch-EMAN [6]	74.0	90.9	30.0M/25.6M	$\sim$ 1100
None	CoMatch [24]	73.6	91.6	30.0M/25.6M	$\sim$ 400
MoCo V2 [9]	CoMatch [24]	73.7	91.4	30.0M/25.6M	$\sim$ 1200
None	SimMatch [54]	74.4	91.6	30.0M/25.6M	$\sim$ 400
None	**SequenceMatch**	**75.2**	**91.9**	25.6M/25.6M	$\sim$ 400

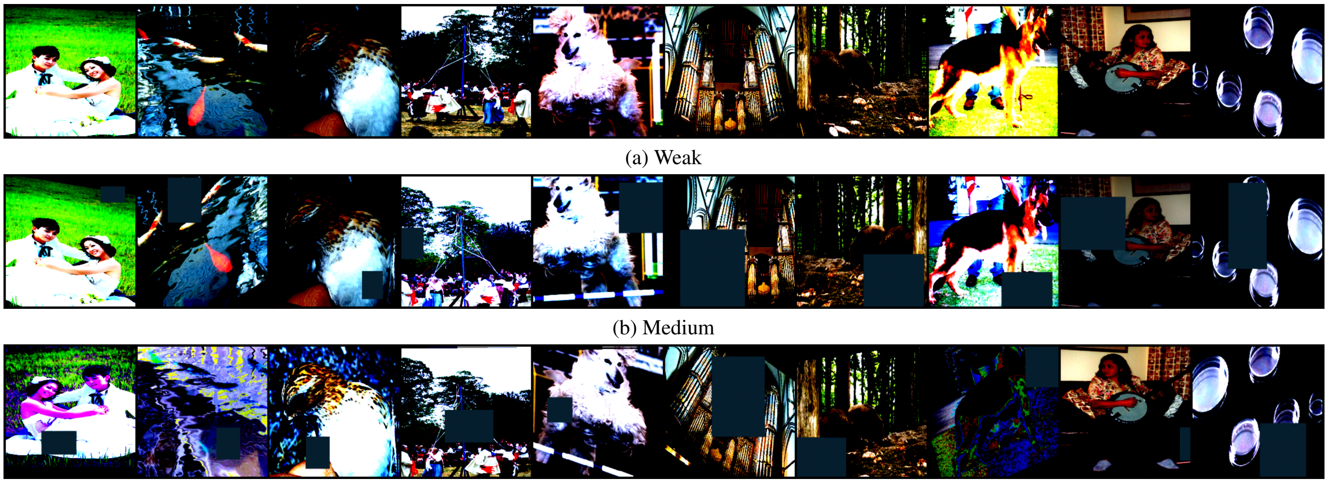
Table 12. List of transformations used in RandAugment

Transformation	Description	Parameter	Range
Autocontrast	Maximizes the image contrast by setting the darkest (lightest) pixel to black (white).		
Brightness	Adjusts the brightness of the image. $B = 0$ returns a black image, $B = 1$ returns the original image.	B	[0.05, 0.95]
Color	Adjusts the color balance of the image like in a TV. $C = 0$ returns a black & white image, $C = 1$ returns the original image.	C	[0.05, 0.95]
Contrast	Controls the contrast of the image. A $C = 0$ returns a gray image, $C = 1$ returns the original image.	C	[0.05, 0.95]
Equalize	Equalizes the image histogram.		
Identity	Returns the original image.		
Posterize	Reduces each pixel to B bits.	B	[4, 8]
Rotate	Rotates the image by θ degrees.	θ	[-30, 30]
Sharpness	Adjusts the sharpness of the image, where $S = 0$ returns a blurred image, and $S = 1$ returns the original image.	S	[0.05, 0.95]
Shear_x	Shears the image along the horizontal axis with rate R.	R	[-0.3, 0.3]
Shear_y	Shears the image along the vertical axis with rate R.	R	[-0.3, 0.3]
Solarize	Inverts all pixels above a threshold value of T.	T	[0, 1]
Translate_x	Translates the image horizontally by ($\lambda\times$image width) pixels.	λ	[-0.3, 0.3]
Translate_y	Translates the image vertically by ($\lambda\times$image height) pixels.	λ	[-0.3, 0.3]

B. Long-tailed issue

To further prove the effectiveness of SequenceMatch, we evaluate SequenceMatch on the imbalanced SSL setting. We conduct experiments on CIFAR-10-LT, SVHN-LT, and CIFAR-100-LT with different imbalance ratios. Following [23,37,48], we use WRN-28-2 as the backbone. We consider long-tailed (LT) imbalance, where the number of data points exponentially decreases from the first class to the last, i.e., $N_k = N_1 \times \lambda^{-\frac{k-1}{L-1}}$, where $\lambda = \frac{N_1}{N_k}$. For CIFAR-10, we set $\lambda = 100, N_1 = 1000$, and $\beta = 10\%, 20\%$, and 30%, respectively. Similarly, we set $\lambda = 100, N_1 = 1000$, and $\beta = 20\%$ for SVHN. And for CIFAR-100, we set $\lambda = 20, N_1 = 200$, and $\beta = 40$. The results are recorded in Table 13 with an average of three different runs.

Surprisingly, SequenceMatch boosts the performance by a large margin when used with ABC [23]. With an accuracy of 85.4%, SequenceMatch outperforms ABC with an 8.2% improvement when β equals 10%.



(a) Weak

(b) Medium

(c) Strong

Figure 15. Weak, medium, and strong augmented examples.

```
Algorithm 1: SequenceMatch algorithm
  Input: Labeled batch $\mathcal{X} = (x_b, p_b) : b \in (1, \dots, B)$, unlabeled batch $\mathcal{U} = u_b : b \in (1, \dots, \mu B)$, confidence threshold
         $\tau$, unlabeled data ratio $\mu$, unlabeled loss weight $\lambda_u$, temperature $\mathbf{T}$, $\Omega$ is $\mathcal{A}_w$, $\mathcal{A}_m$ or $\mathcal{A}_s$
  /* Cross-entropy loss for labeled data                                                                  */
  $\mathcal{L}_s^{CE} = \frac{1}{B} \sum_{b=1}^{B} \mathrm{H}\left(p_b, \mathcal{A}_w\left(x_b\right)\right)$ for $b = 1$ to $\mu B$ do
      $q_b\left(y \mid \Omega\left(u_b\right)\right) = p_m\left(y \mid \Omega\left(u_b\right)\right)$ // Compute prediction after applying weak data augmentation of $u_b$
      $q_s = \frac{\exp\left(q_b / \mathbf{T}\right)}{\sum_k \exp\left(q_k / \mathbf{T}\right)}$. // Sharpen the output probability
  /* Cross-entropy loss with pseudo-label and confidence threshold for high-confidence unlabeled          */
  $\mathcal{L}_u^{\mathrm{CE}} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \left(\mathbb{1}\left(\max\left(q_b^w\right) \geq \tau\right) \mathrm{H}\left(\hat{q}_b, p_m\left(y \mid \mathcal{A}_s\left(u_b\right)\right)\right) + \mathbb{1}\left(\max\left(q_b^w\right) < \tau\right) \mathrm{H}\left(q_s^w \mid q_b\left(\mathcal{A}_s\left(u_b\right)\right)\right)\right)$
  /* Kullback-Leibler divergence loss with each pair of augmented examples                                 */
  $\mathcal{L}_{\mathrm{KL}}^{w-m} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbb{1}\left(\max\left(q_b^w\right) \geq \tau\right) \quad D_{\mathrm{KL}}\left(q_s^w \mid p_m\left(y \mid \mathcal{A}_m\left(u_b\right)\right)\right)$
  $\mathcal{L}_{\mathrm{KL}}^{m-s} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbb{1}\left(\max\left(q_b^m\right) \geq \tau\right) \quad D_{\mathrm{KL}}\left(q_s^m \mid p_m\left(y \mid \mathcal{A}_s\left(u_b\right)\right)\right)$
  $\mathcal{L}_{\mathrm{KL}}^{w-s} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbb{1}\left(\max\left(q_b^w\right) \geq \tau\right) \quad D_{\mathrm{KL}}\left(q_s^w \mid p_m\left(y \mid \mathcal{A}_s\left(u_b\right)\right)\right)$
  $\mathcal{L}_u = \mathcal{L}_u^{\mathrm{CE}} + \mathcal{L}_{\mathrm{KL}}^{w-m} + \mathcal{L}_{\mathrm{KL}}^{m-s} + \mathcal{L}_{\mathrm{KL}}^{w-s}$
  return $\mathcal{L}_s^{CE} + \lambda_u \mathcal{L}_u$
```

Table 13. Overall accuracy under the long-tailed setting

	CIFAR-10-LT			SVHN-LT	CIFAR-100-LT
ALGORITHM		$\lambda = 100$		$\lambda = 100$	$\lambda = 20$
	$\beta = 10\%$	$\beta = 20\%$	$\beta = 30\%$	$\beta = 20\%$	$\beta = 40\%$
VANILLA	-	55.3±1.30	-	77.0±0.67	40.1±1.15
VAT [30]	-	55.3±0.88	-	81.3±0.47	40.4±0.34
BALMS [37]	-	70.7±0.59	-	87.6±0.53	50.2±0.54
FIXMATCH [44]	70.0±0.59	72.3±0.33	74.9±0.63	88.0±0.30	51.0±0.20
W/ CREST+PDA [48]	73.9±0.40	76.6±0.46	74.9±0.63	89.1±0.69	51.6±0.29
W/ DARP [19]	-	73.7±0.98	-	88.6±0.19	51.4±0.37
W/ DARP+CRT [19]	74.6±0.98	78.1±0.89	77.6±0.73	89.9±0.44	54.7±0.46
W/ ABC [23]	77.2±1.60	81.1±0.82	81.5±0.29	92.0±0.38	56.3±0.19
SEQUENCEMATCH	**85.4±0.01**	**81.5±0.75**	**82.2±0.25**	**92.4±0.06**	**57.2±0.09**