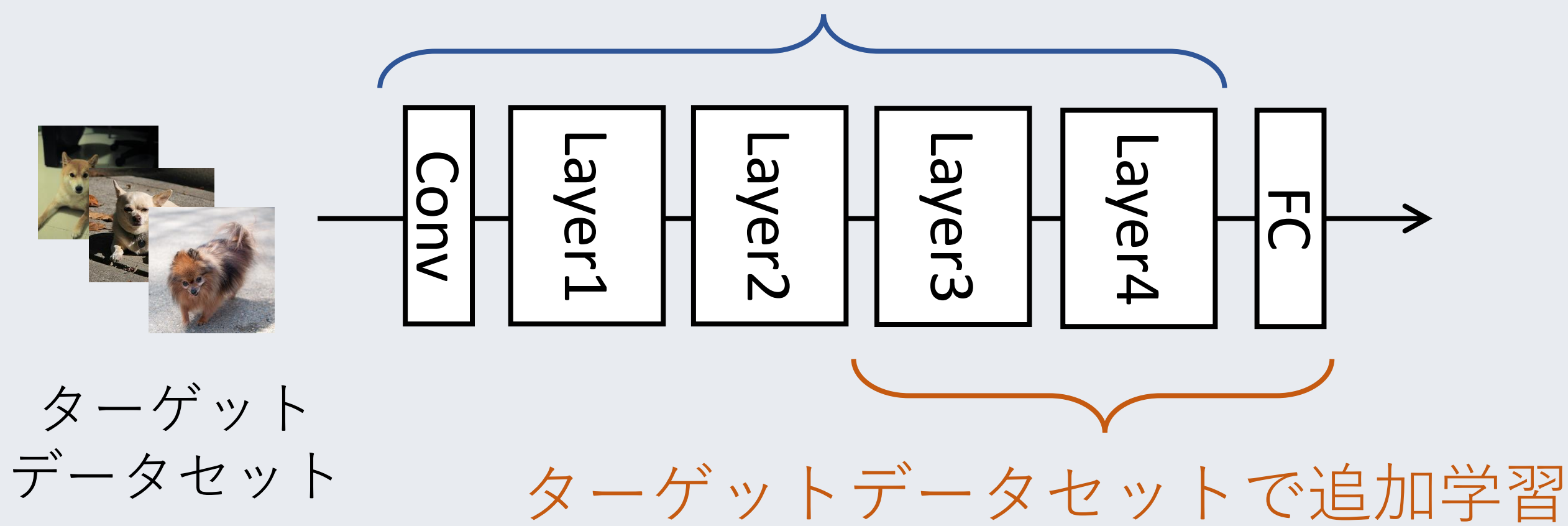


1. 背景と本研究の貢献

背景 学習データの少ない画像分類タスクでは、別の大規模データセットで学習したモデルを再利用し、全部または後半の一部の層を追加学習する転移学習が主流

別のデータセットで事前に学習済み



しかし、「どの層を追加学習するべきか」がターゲットデータセットによって曖昧である

→ 学習する層の選択に試行錯誤を要する

貢献 ランダムに1層ずつ選択して学習するだけの、**簡易な**転移学習手法を提案。多くのデータセットにおいて、従来手法よりも大幅に性能を改善することを示す。

2. 提案手法

複数の層を持つ深層ネットワークに対して、学習ステップ毎に以下の手順で学習を行う

- ① 学習データからミニバッチを取り出す
- ② ネットワークからランダムに一層選択し、その層に属するパラメータには1を、それ以外の層のパラメータには0を、学習率に乗じる
- ③ パラメータを更新する

$$\theta_l^{(t)} = \theta_l^{(t-1)} - m_l^{(t)} \eta \nabla_{\theta_l} \mathcal{L}$$

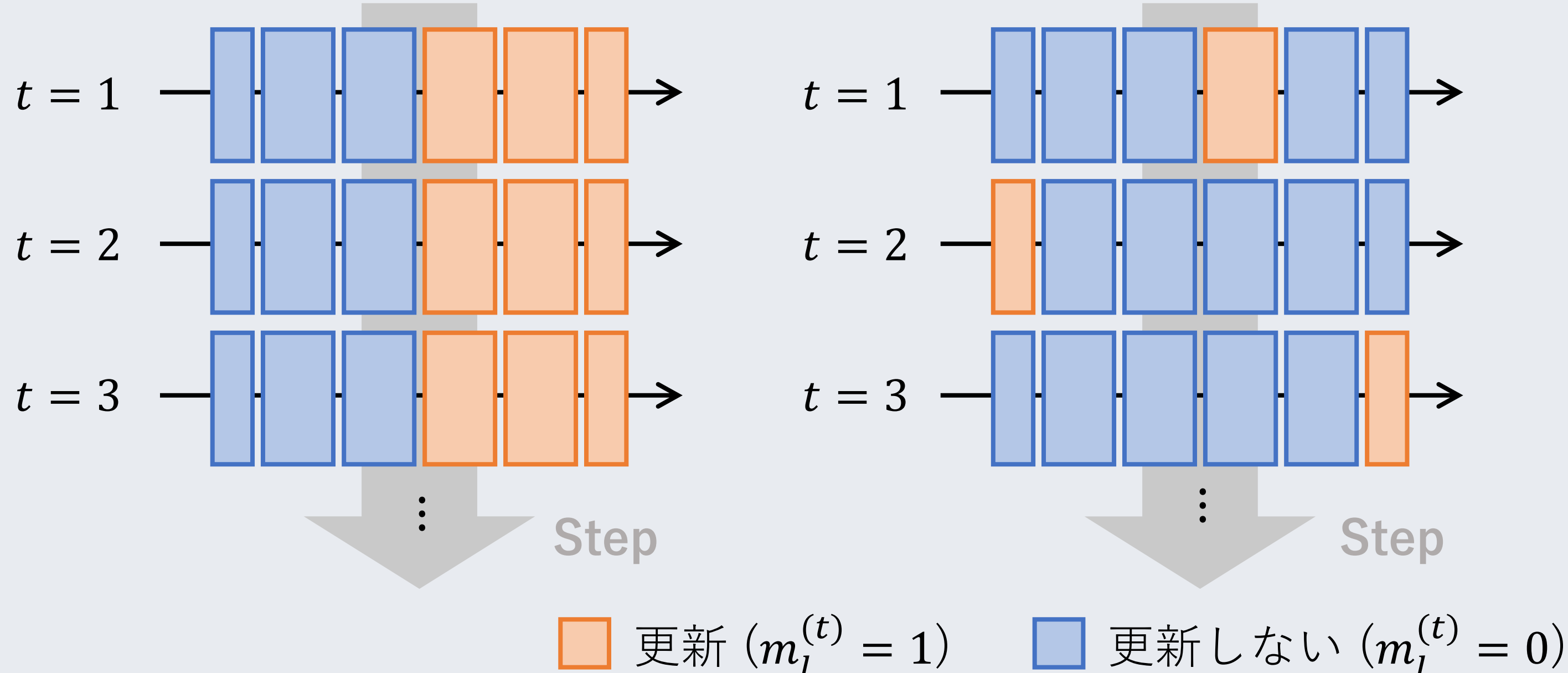
θ_l : l層目のパラメータ η : 学習率 $m_l^{(t)} \in \{0,1\}$

従来手法

更新する層は常に同じ

提案手法

更新する層をランダムに選択



3. 評価条件

- WideResNet-28-10をImageNetで250エポック学習した事前学習モデルを使用
- 最終層 (FC層) はターゲットデータセットのクラス数に応じて付け替え、ランダムに初期化
- 9つのターゲットデータセットに対して、モーメントSGDで50エポック追加学習
- 学習率は実験を通して0.01で固定

4. 評価

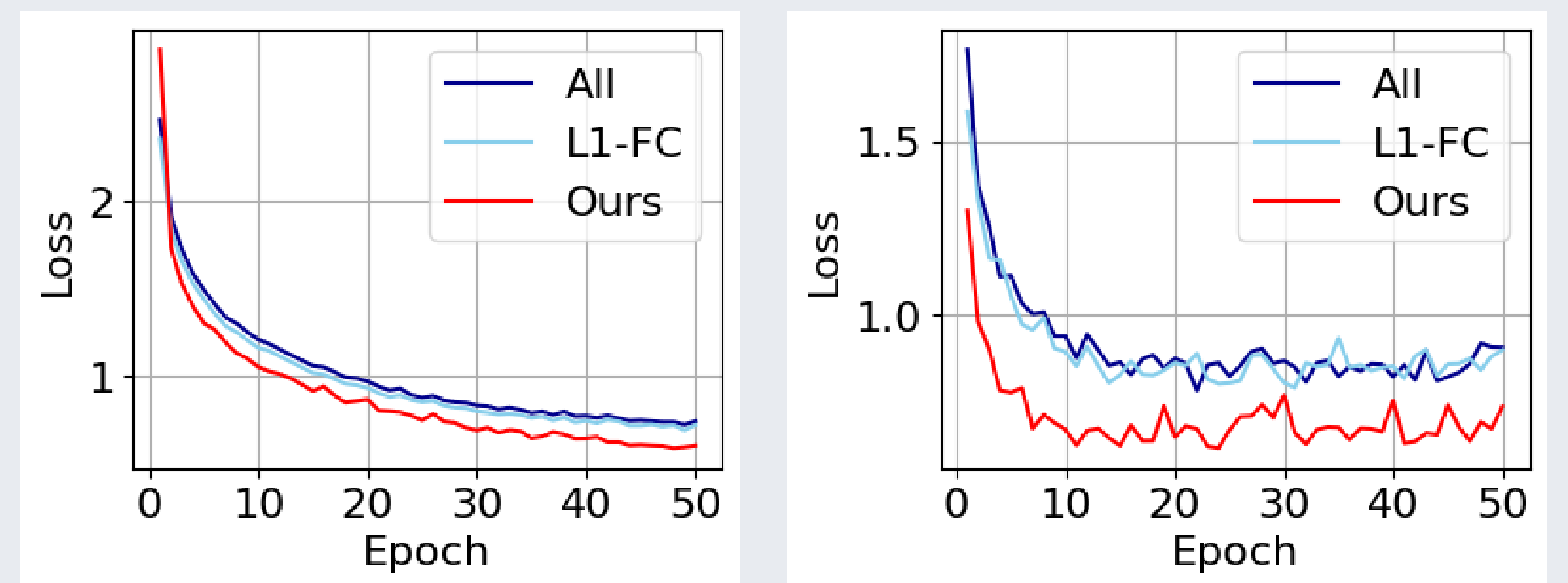
● 従来手法とのテストエラー率の比較

ターゲットデータセット	従来手法 (特定の層のみを更新)						Ours
	All	L1-FC	L2-FC	L3-FC	L4-FC	FC	
Flowers	7.65	6.67	6.37	5.88	4.61	12.84	5.78
Pets	15.14	13.24	11.96	10.29	8.33	8.22	6.29
DTD	40.37	38.78	37.23	34.57	31.60	35.32	28.35
Aircraft	19.34	18.32	16.28	16.22	17.72	49.40	14.60
Food	18.01	17.05	16.52	15.84	16.97	35.45	14.55
SUN	35.54	35.18	34.23	33.21	31.77	36.72	30.84
Caltech101	13.65	12.53	11.14	10.46	8.06	8.47	6.19
CIFAR10	5.19	5.15	5.33	5.91	9.77	34.11	3.63
CIFAR100	22.07	21.65	21.89	23.06	31.93	56.87	17.52
AVE	19.66	18.73	17.88	17.27	17.86	30.82	14.19

※下線は従来手法での最良値、太字は全手法での最良値

- ✓ Flowers以外のデータセットで従来手法の最良値を上回る
- ✓ フルファインチューニング(All)に対しては、全データセットで性能が大幅に向上

● 従来手法との学習曲線の比較 (CIFAR100)



(a) Train loss

(b) Test loss

- ✓ 訓練データとテストデータの両方において、提案手法の方が損失が小さい
- ✓ 提案手法は1層あたりの更新回数が6分の1で少ないはずだが、収束速度はほとんど変わっていない

● なぜ性能が上がるのか?

各ステップで一層のみが更新されるため、パラメータ更新の際に他の層の影響を受けない

→ 層間の依存(Co-Adaptation)を防いでいると推測