

1. はじめに

大規模な事前学習モデルでは、下流タスクに適応する方法としてモデル全体を再学習することが難しい。そこで、入力に学習可能なパラメータであるプロンプトを追加し、プロンプトを学習するプロンプトチューニングがある。Visual Prompt は、画像にパディング状のプロンプトを付加して学習する手法である。これにより、プロンプトのみを学習するだけで、モデル全体を再学習する手法に準じる性能を達成できる。一方で、Visual Prompt が学習にどのような影響を与えているか明確ではない。そこで本研究では、Visual Prompt を付加する位置を変更したときの影響を調査する。具体的には、分類精度の変化とアテンションマップによる評価を行う。

2. Visual Prompt

Visual Prompt [1] は、CLIP [2] を下流タスクに適応させるために、入力画像に対して学習可能なパラメータを持つ共通のプロンプトを挿入する手法である。図 1 に Visual Prompt の概要図を示す。Visual Prompt を付加した画像を重みが固定された事前学習済みモデルに入力し、正解ラベルとの尤度を最大化するように Visual Prompt を学習する。

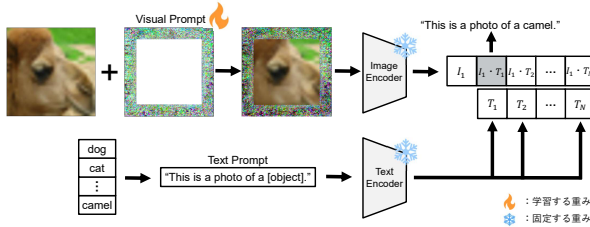


図 1: Visual Prompt の概要

3. Visual Prompt の改善

Visual Prompt には問題点がある。例として図 2 に Grad-CAM を用いて Visual Prompt を付与する前後のアテンションマップを可視化した結果を示す。図 2(a) のように Visual Prompt を付与する前は識別対象に注視していたのに対し、図 2(b) のように Visual Prompt を付与した後は識別対象に注視していないことが確認できる。

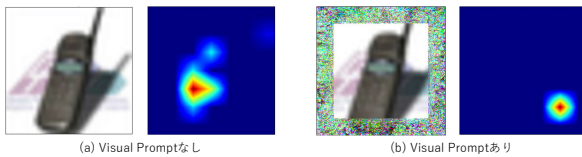


図 2: アテンションマップの可視化の結果

Visual Prompt を画像に重ねることで、図 3(b) のように識別対象の物体の一部が Visual Prompt により隠れてしまうことがある。これにより、図 2 のように認識に有効な特徴量が抽出されなくなると考えられる。そこで、図 3(c) のように Visual Prompt を画像の外側に付与することで、Visual Prompt が対象物体に重ならないようにして、プロンプトチューニングを行う。



図 3: Visual Prompt の改善

4. 実験

本研究では、Visual Prompt を画像の内側または外側に付与した場合の分類精度を比較する。加えて、それぞれの場合のアテンションの可視化結果を比較する。

4.1. 実験条件

データセットに CIFAR100 を用いて、CLIP で事前学習した ResNet50 と ViT-B/32 を評価する。学習するエポック数は 1,000、バッチサイズは 256、最適化は SGD を使用する。

4.2. 実験結果

表 1 より、Visual Prompt を画像の外側に付与すると、モデルに ResNet50 を用いた場合は精度が 5.90pt、ViT-B/32 を用いた場合は精度が 4.93pt 向上した。このことから、Visual Prompt は画像の外側に付与すると良いことを確認した。

表 1: 各 Visual Prompt における分類精度

Image Encoder	Visual Prompt の位置		
	プロンプトなし	従来手法	提案手法
ResNet50	40.47	47.28	53.18
ViT-B/32	63.11	75.09	80.02

Visual Prompt を付与した各位置におけるアテンションマップの可視化結果を図 4 に示す。図 4(a) のように ResNet50 の場合、画像の外側に Visual Prompt を付与することで物体を注視するようになった。一方で、図 4(b) のように物体を注視しないことも確認された。ViT-B/32 の場合、画像の外側に Visual Prompt を付与することで、図 4(c) のように画像領域内の物体に注視した。これにより、Visual Prompt の位置が精度向上に影響することを確認した。

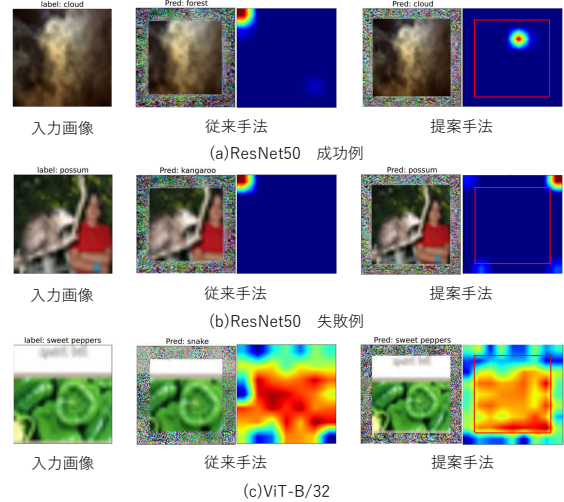


図 4: プロンプトの位置の変更に伴うアテンションマップ

5. おわりに

本研究では、Visual Prompt の位置が認識精度及び解釈性に与える影響を分析した。実験により、Visual Prompt を画像の外側に付与することで精度の改善が確認された。一方、アテンションマップは十分に改善されていない画像も確認できた。今後は、さらなるアテンションマップの改善を目的とした Visual Prompt を考案する予定である。

参考文献

- [1] H. Bahng, *et al.* “Exploring Visual Prompts for Adapting Large-Scale Models”, arXiv, 2022.
- [2] A. Radford, *et al.* “Learning Transferable Visual Models From Natural Language Supervision”, ICML, 2021.