7T-07

Black-Box Visual-Prompting による Few-Shot 教師なしドメイン適応

豊岡 真知† 宮井 淳行† 郁 青† 入江 豪‡ 相澤 清晴†
東京大学† 東京理科大学‡

1 はじめに

研究の進展に伴って人工知能モデルはパラメータ数と学習データ量の点で大規模化の一途を辿り，巨大企業らによる大規模人工知能モデルが API という形で提供される機会が増加している．それらモデルは学習データに含まれているプライバシーや法的・商業的観点からモデルのパラメータ等を公開しないことがしばしばある．

従って，人工知能モデルが API の形で提供されている場合や，メモリ容量の観点からモデルのパラメータに自由にアクセスできない状況がこれから増加していくと推測できる．このようにユーザが自由にアクセスできないモデルのことを以下ではブラックボックスモデルと呼ぶ．これまでブラックボックスモデルを適応する手法は検討されてきたが，それらは下流タスクにおいてラベルのついたデータが与えられることを前提としており，ラベルの付与には人為的コストが大きくかかるため実用性の観点で課題がある．

このことから我々はより現実的な問題設定として，下流タスクで与えられるデータがラベルのない少量データであることを仮定し，これに対して分類タスクを行うブラックボックス画像モデルの適応を試みる．特に画像領域でプロンプティングを行う Visual-Prompting を用いた手法の有効性を検証する．Visual-Prompting は，画像情報のみを用いることでモデルの入力となる画像埋め込み空間や，出力する潜在空間にアクセスする必要がなく，より広義のブラックボックスモデルに対して適用できる．

2 提案手法

本手法ではまず，モデルの適応のため Black-Box Visual Prompting（Black-VIP）を行う（小節 2.1）．Black-VIP は教師あり学習手法であるため，ラベルなしデータに対してラベルを付与する Pseudo Labeling を行う必要がある．Pseudo Labeling で得られるラベルはノイズが多く含まれているため，そのようなラベルを適切に扱うための手法を小節 2.2 で説明する．最後に，全体の学習方法について小節 2.3 で説明する．

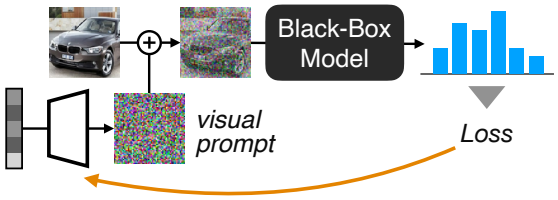

図 1 Black-VIP．Visual Prompt 生成機構のパラメータを損失が小さくなる方向へ更新し，プロンプトを付与した際のモデルの分類精度を向上させる．

2.1 Black-VIP

Visual Prompting [1](VIP) とは，画像として扱うことのできる Prompt を学習し，入力画像に加算してモデルへ入力することでモデルのパラメータを変更することなく下流タスクへ適応する転移学習手法である．VIP をブラックボックスモデルに応用したものが Black-VIP [2]（図 1）である．Visual Prompt はニューラルネットワークにベクトルを入力することで画像形式で生成され，入力画像とピクセルレベルで加算された後，ブラックボックスモデルへ入力し出力確率分布から損失を計算する．続けて，数値微分で近似した損失勾配を基に損失を最小化するアルゴリズムを用いて Visual Prompt 生成機構を学習する．モデルが API 等で提供されている場合，モデルへの入力は画像形式データのみであるため，そのような場合において Black-VIP は有効であり，本手法における転移学習手法にこれを採用する．

2.2 Learning with Noisy Pseudo Labels

Pseudo Labeling は，ラベルのないデータに対して人為的操作を伴わず擬似的なラベル (Pseudo Label) を付与し，それを用いた教師あり学習を可能にする手法である (図 2)．学習対象となるプロンプト生成機構のパラメータを $\boldsymbol{\theta}$ とする．n 枚，c クラスの学習データ

Black-Box Visual-Prompting for Few-Shot Unsupervised Domain Adaptation
Mashiro Toyooka†, Atsuyuki Miyai†, Qing Yu†, Go Irie‡ and Kiyoharu Aizawa†
†The University of Tokyo
‡Tokyo University of Science
本研究の一部は，JST JPMJCR22U4 の支援を受けた

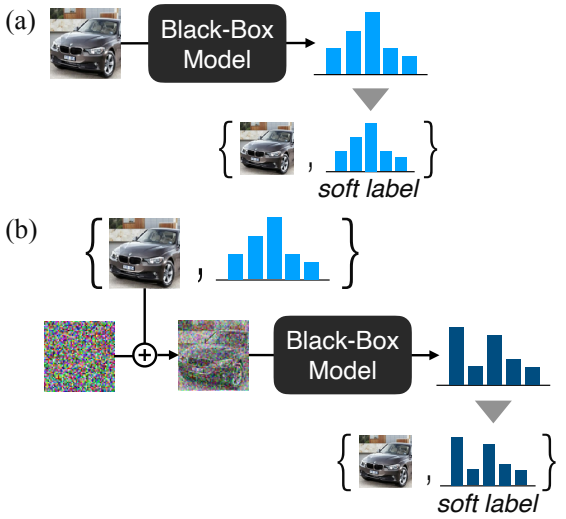

図 2　Pseudo Labeling．ラベルのない画像に対して擬似的なラベルを付与する．本手法では，モデルの出力確率分布をそのままラベルとする soft label を用いる．

$X = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]$ があり，プロンプトを付した画像に対するモデルの出力確率分布 $\boldsymbol{f}(\boldsymbol{x}|\boldsymbol{\theta}) \in \mathbb{R}^c$ が得られる．Pseudo Labeling による，各画像に対応するラベル $Y = [\boldsymbol{y}_1, ..., \boldsymbol{y}_n]$ は，soft label の場合

$$\boldsymbol{y}_i = \boldsymbol{f}(\boldsymbol{x}_i|\boldsymbol{\theta}) \tag{1}$$

で与えられる．

一般的に，Pseudo Labeling によって付与されたラベルはノイズが含まれており，Joint Optimization Framework [3](JOF) によるラベルの浄化が効果的であると考え，これを採用する．JOF はモデルとラベルを同時に更新する戦略と新たな損失を導入することで，ノイズを含んだ学習データが与えられた場合においても正常な学習を行うことを目的とする手法である．学習過程のある時刻 t において損失 $\mathcal{L}(\boldsymbol{\theta}^{(t)}, Y^{(t)}|X)$ が得られたとき

1. $\mathcal{L}(\boldsymbol{\theta}^{(t)}, Y^{(t)}|X)$ を用いて $\boldsymbol{\theta}^{(t+1)}$ を更新
2. 式 1 によってラベル $Y^{(t+1)}$ を更新

の手順で $\boldsymbol{\theta}^{(t+1)}, Y^{(t+1)}$ を得る．このようにしてモデルとラベルを同時に更新する．損失については，一般的な分類損失と [3] で提案された 2 つの損失を導入する．

2.3　学習

本手法では，まず，与えられたブラックボックスモデルを用いて Visual Prompt を適用せずに Pseudo Labeling を行い，画像のラベル付けを行う (図 2 (a))．そのようにして得られたラベルを用いて，しばらく Prompt 生成機構を学習する．学習初期ではランダムな Visual Prompt が生成されることで，プロンプトが付与された画像を入力すると何もしない場合 (Zero-Shot) に比べて分類精度が低下するが，上記のようにして，Zero-Shot で生成した Pseudo Label を用いて学習することで，Visual Prompt を付した認識精度が少なくとも Zero-Shot と同程度であることを保証する狙いがある．

しばらく学習した後，JOF の学習 (小節 2.2) に倣いプロンプト生成機構の更新（図 1）とラベルの更新 (図 2 (b)) を同時に繰り返し行う．

表 1　実験結果

Method	EuroSAT	DTD	Flowers	UCF101
Zero-Shot	50.3	43.6	70.8	66.0
Ours	54.0	43.8	69.3	67.2

3　実験結果・まとめ

下流データセットとして EuroSAT, DTD, Flowers, UCF101 の 4 つを選定し実験を行う．これらのデータセットに含まれる意味的領域は広く，本手法の有効性を多角的に測ることができると考える．学習時には，各クラスから 16 枚ずつをランダムにサンプリングしラベルのない学習データとして用いる．また，ブラックボックス画像モデルは OpenAI による事前学習済み CLIP [4] を用いる．結果を表 1 に示す．EuroSAT, Flowers, UCF101 の 3 つのデータセットに対しては，Zero-Shot と比べて 0.2〜3.7 ポイントの精度上昇が確認できる一方で，Flowers データセットでは Visual Prompt の付与によって精度が下がる結果が得られた．以上より，本手法によって事前学習モデルのパラメータにアクセスすることなくラベルのないデータに対してモデルを適応させることが可能であることが明らかになった．しかし，データセットによっては精度向上の度合いが低い．様々なパラメータの選択によりさらなる精度向上が可能であると考え，検証を継続している．

参考文献

[1] H. Bahng, A. Jahanian, S. Sankaranarayanan, and P. Isola. Exploring visual prompts for adapting large-scale models. *arXiv preprint arXiv:2203.17274*, 2022.
[2] C. Oh, H. Hwang, H.-y. Lee, Y. Lim, G. Jung, J. Jung, H. Choi, and K. Song. Blackvip: Black-box visual prompting for robust transfer learning. In *CVPR*, 2023.
[3] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa. Joint optimization framework for learning with noisy labels. In *CVPR*, 2018.
[4] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, 2021.