

DA-LoRA: 植物病害診断における ドメイン特徴の視覚的分析のための敵対的画像生成

畔上 達希^{1,a)} 有馬 祐二^{1,b)} 川田 拓朗^{1,c)} 彌富 仁^{1,d)}

概要

植物病害の画像診断において、撮影環境などのドメイン変化に十分な精度が得られないドメインシフトが問題となっている。しかし、敵対的学習などによるドメイン依存特徴の除去は、必ずしも精度向上に寄与しない。本研究では、画像生成モデルにドメインラベルを用いたドメインの分離を行う DA-LoRA を提案し、ドメイン依存、ドメイン中立特徴を画像として分離することで、それぞれの特徴が分類に与える影響を分析する。

1. はじめに

機械学習を用いた植物病害診断では、対象病害の多様性に対して学習データの多様性が不足するため、ドメイン(圃場や撮影条件)の変化に伴う分布の変化により、十分な診断能が得られないことが問題となっている。これまで多様な研究において高い診断性能が報告されたが、その多くは同一ドメイン内の評価による過大評価であり、学習データと別環境における検証では十分な精度が得られないことが指摘されている [12, 17]。この課題に対し、古典的データ拡張 [1, 6] や生成的データ拡張 [8] の他、敵対的学習などによるドメイン適応手法 [4] が提案されたが、これまで植物病害診断では十分な汎化性能は得られていない。古典的なデータ拡張(回転、色変換など)は、画像の多様性向上に広く用いられ、機械学習モデルの基本的な性能確保に一定の効果を挙げているが、識別に寄与する特徴の本質的な多様性は向上しないため、単体での適用では限界がある。これに対し、Generative Adversarial Networks (GAN) [5] や拡散モデルを用いた生成的データ拡張による多様性の補完は、学習データの本質的な多様性を増やす手法として近年特に注目を集めている。潜在拡散モデルを用いた生成的拡張手法は、テキスト指示による柔軟な生成的拡張を可能とする。これは学習データの分布外にある多様な画像生成が実現でき、この問題の本質を改善できる可能性を有した

技術であるが、まだ本課題で十分な成果は報告されていない。異なるアプローチとして、ドメイン適応・調和技術は、複数ドメインの分布差を緩和することで、解析対象ドメインでの識別性能向上を目指す手法がある。この技術は、ドメインにより変化しない低次元表現の獲得を目指す距離学習 [13] や対照学習 [9] に代表される手法群と、敵対的学習により、分類能の維持とドメイン除去の両立を目指す Domain Adversarial Neural Networks (DANN) [4] や Conditional Domain Adversarial Networks (CDAN) [11] 等の手法に大別できる。しかし、一般的な条件では評価環境の情報が利用できないため、学習環境と評価環境の差が大きいタスクにおいて、いずれの手法も学習ドメイン内の適応による評価環境での精度向上効果は限定的である [14]。事前に対象ドメインの情報を学習に利用できる場合は、少ないデータであっても大きな効果を実現する [10] が、その設定が利用できる状況は限られている。

また、植物病害診断は背景や撮影環境の影響を強く受けるタスクであるが、同時に病害特徴が生育環境の影響により変化するため、病害を分類するための特徴自体がドメインに強く依存することが予想される。その状況下で、DANN のようなドメイン除去を行う場合、ドメイン依存の特徴を除去すると同時に、病害診断に寄与する病害特徴まで除去されてしまうリスクがある。この問題は、従来のドメイン汎化手法におけるドメインに非依存な特徴の抽出が汎化性能に直結するという前提と、植物病害におけるドメイン依存性の食い違いに起因する。しかし、従来の識別モデルを用いたアプローチでは、特徴空間における病徴情報の消失を直接確認できず、分類精度の観測のみではこれら进行分析することは困難である。そのため、ドメイン依存・ドメイン非依存の特徴がそれぞれ持つ病害特徴を分析する方法が求められる。

本研究では、画像生成モデルに対する敵対的学習を用いたドメイン特徴の分離を行う Low Rank Adaptation (LoRA) [7]、DA-LoRA を提案し、分離した特徴の可視化を行う。DA-LoRA は、ドメイン分類不能な特徴を保持する LoRA と、ドメイン分類可能な特徴を保持する LoRA を並列に学習し、ドメインの影響を分離した画像生成を試みる。

¹ 法政大学大学院

a) tatsuki.azegami.7f@stu.hosei.ac.jp

b) yuji.a.20010416@gmail.com

c) takuro.kawada.3g@stu.hosei.ac.jp

d) iyatomi@hosei.ac.jp

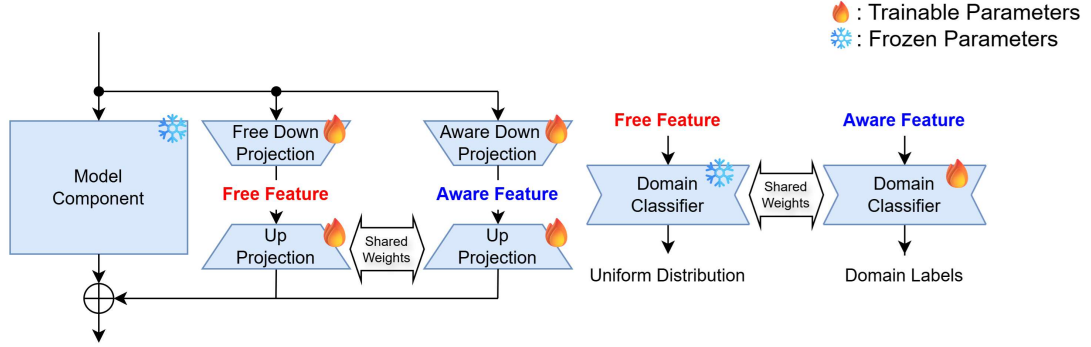


図 1: DA-LoRA の概要.

生成画像の背景, 病徴を比較することで, ドメインの影響を分析し, ドメインに依存しクラス分類に寄与する特徴の存在を確認する.

2. DA-LoRA

本研究では, 画像生成モデルを用いて学習データにおけるドメイン依存の特徴を持つ画像 (ドメイン画像) と, ドメインに非依存の特徴を持つ画像 (中立画像) を生成する DA-LoRA を提案する. 生成画像のドメイン中立性と画像特徴を評価することで, ドメインシフトの要因を分析し, その改善に向けた知見を得る.

モデル構成. 図 1 に DA-LoRA の構成図を示す. DA-LoRA は並列な 2 つの LoRA と 1 つのドメイン分類器で構成され, これを Latent Diffusion Model (LDM) [15] などに適用することで画像生成に利用する. ここで, 図 1 に示す Model Component (MC) は, ベースモデルにおける Attention 機構などの構成要素を示す. この 2 つの LoRA は, ドメイン分類器で分類可能な特徴 (Aware Feature) を保持する Domain Aware LoRA (Aware-LoRA) と, ドメイン分類器で分類不能な特徴 (Free Feature) を保持する Domain Free LoRA (Free-LoRA) である. これらの重みは DANN などの既存手法と同様, ドメイン分類器を利用した敵対的学習により獲得される. ドメイン分類器は LoRA の中間特徴を入力とし, 学習画像に対応したドメインラベルを推論する単純な MLP である.

DA-LoRA ではこの 2 つの LoRA を並列で配置し, Decoder の重みを共有する. 通常の LoRA と, DA-LoRA の動作は, それぞれ式 (1), 式 (2) のように表される.

$$h = W_0x + BAx \quad (1)$$

$$h = W_0x + BA_{\text{aware}}x + BA_{\text{free}}x \quad (2)$$

ここで, W_0 は MC の重み, A, B はそれぞれ Decoder, Encoder を表し, x, h は入力と出力を示す. また, それぞれの LoRA の出力は, 等倍で MC 出力に加算される. ドメイン分類器は Aware Feature を入力として学習されるため, Free Feature が Aware Feature と異なる意味空間を持

つ場合, ドメイン分類器の Free Feature に対する分類能が保証されず, 敵対的学習によるドメイン中立性の獲得が困難となる. そこで DA-LoRA では, 2 つの Encoder で作られる内部表現は, 共通の Decoder で MC の出力空間へ写像することで, Aware Feature と Free Feature の意味空間の共有を促し, 両特徴空間の乖離を抑制する.

学習機構. DANN などの敵対的学習で用いられる Gradient Reversal Layer (GRL) は, ドメイン分類器に対するクロスエントロピーを最大化するような学習を行う. しかし, 勾配消失や学習の不安定性を招きやすいため, DA-LoRA では均一分布をターゲットとしたクロスエントロピーの最小化を行う. Aware LoRA の学習では, Aware Feature をドメイン分類器に入力し, 式 (3) に示すドメインラベルとのクロスエントロピー誤差を損失関数として Aware LoRA, ドメイン分類器の学習を行う.

$$\mathcal{L}_{\text{aware}} = - \sum_{k=1}^K y_k \log(p_k) \quad (3)$$

ここで, K はドメインの総数, y_k は正解ドメインを示す one-hot ベクトル, p_k はドメイン分類器が出力したクラス k に対する予測確率である. Free-LoRA の敵対的学習では, Free LoRA の中間表現をドメイン分類器に入力し, 式 (4) に示す, 一様分布ラベルとモデル出力のクロスエントロピー誤差を損失関数として Free LoRA の学習を行う.

$$\mathcal{L}_{\text{free}} = - \sum_{k=1}^K \frac{1}{K} \log(p_k) \quad (4)$$

この際, ドメイン分類器の重みは固定し, Free-LoRA のパラメータのみを更新する. これにより, DANN と同様に, ドメイン分類器による識別が困難な特徴の獲得を促す.

ドメイン画像とドメイン中立画像の生成. LoRA はベースモデルの重みを固定したまま, 学習データに適応するために必要な重みの差分を学習する. したがって, 理想的な特徴分離が行われていれば, Aware-LoRA は事前知識とドメイン特徴の差分, Free-LoRA は事前知識とドメイン中立特徴の差分を学習していると解釈できる. 両 LoRA はベースモデルから着脱可能なため, 推論時に Aware-LoRA のみ

表 1: データセットの内訳.

domain	群馬①	群馬②	富山①	富山②	山梨	京都①	京都②	高知	鹿児島	三重
Healthy	448	236	82	1,570	0	32	32	251	0	631
V-Wilt	440	502	0	0	0	0	0	0	240	259
B-Wilt	554	222	0	1,065	448	0	0	0	0	926

V-Wilt: Verticillium Wilt. B-Wilt: Bacterial Wilt.

を適用することでドメイン画像の生成が, Free-LoRA のみを適用することでドメイン中立画像の生成が可能である.

3. 実験

データセット. 本実験では, 表 1 に示す 10 ドメイン, 3 クラス (Healthy, Verticillium Wilt, Bacterial Wilt) からなる, ナスの植物全体を撮影した画像データセットを用いる. このデータセットは, 専門家でも目視での診断が困難であり, ドメイン間の差分が非常に大きい上, 植物全体に現れる症状を映す必要があることから, 画像あたりの背景が占める割合が大きいことが特徴である. それぞれの画像に対し, 健康クラスには "Photo of a e-plant with green leaves growing in (Domain)-field", 病害クラスには "Photo of a e-plant with green leaves infected with (Disease)-disease growing in (Domain)-field" をプロンプトとして割り当てる. ここで, (Disease) は病害クラス, (Domain) はドメインを表す文字列であり, 意味のある単語を使わないことで, 事前知識の汚染を防ぐ. 評価には三重県の画像を用い, その他のドメインを学習に用いる.

実験条件. はじめに, 多くの分野で成果を上げる画像分類モデルの Vision Transformer (ViT) [2] に対し, このデータセットを以下 2 つの方法で学習することで, ドメインシフトの影響と, 単純なドメイン除去が診断に与える影響を評価する: (1) ViT: ViT に拡張前のデータセットを学習. (2) DANN: ViT に DANN を適用し, 拡張前のデータセットを学習. 実験はそれぞれ 5 回行い, 平均したスコアを評価する. また, クラスごとの枚数の偏りを緩和するために, 損失関数にはクラスごとの枚数で重み付けしたクロスエントロピー誤差を用いる. このデータセットに対する, 中立画像, ドメイン画像の生成には Stable Diffusion 3.5 (SD3.5) [3] を用い, 病害クラスごとに学習, 生成を行う. このとき, 少数データによる過学習抑制のため, 正則化項として DreamBooth [16] の Prior Preservation Loss を導入する. また, それぞれの LoRA の rank は 64 とする. 生成画像の比較には, DA-LoRA (Aware+Free), Free-LoRA, Aware-LoRA でそれぞれドメインごとに 1,000 枚, 通常の LoRA でクラスごとに 1,000 枚の生成を行い, 比較を行う. このとき, DA/Aware/Free-LoRA による生成では, 学習時に用いたプロンプトによる生成を行う. また, 通常の LoRA による生成では, 学習時のプロンプトから (Domain) を取り除き, 生成に用いる. これは, プロンプト制御によるドメ

表 2: ViT と DANN の分類精度.

	accuracy	precision	recall	macro F1-score
(1) ViT	0.771	0.755	0.672	0.686
(2) DANN	0.597	0.652	0.597	0.564

表 3: 生成画像に対するドメイン分類結果.

	accuracy	macro F1-score	entropy
Free-LoRA	0.197	0.170	0.590
LoRA	-	-	0.369

イン除去を目的としている.

評価指標. 分類精度の評価には, accuracy, macro F1-score を用いる. 生成した中立画像の評価には, ドメイン分類器として学習した ViT に対し, 生成画像を入力した際の分類精度と, モデル出力のエントロピーを用いる. ドメイン分類器の学習には, train/test ドメインを合わせた 10 ドメインを学習に用いる. 学習には 10 ドメインを用いるため, ドメイン分類器の精度が 0.10 に近づき, モデル出力のエントロピーが $\log 10 \simeq 2.3$ に近づくほど, ドメインに依存しない特徴であると評価する. ここで, エントロピーの増加は, 分類モデルの予測の不確実性の増加を意味し, ドメイン分類器のエントロピー増加は, ドメイン中立性の一つの指標となる. このとき DA-LoRA はドメインごとの生成を行うが, LoRA はクラスごとの生成を行うため正解ラベルを持たないことから, LoRA による中立画像の評価には, エントロピーのみを用いる. また, 表 1 のデータセットで学習した病害診断モデルに生成した中立画像を入力し, 画像が持つクラス特徴を評価する.

4. 結果

分類精度. 表 2 に, それぞれの手法による分類精度を示す. DANN を適用した分類結果では, ViT 単体による分類結果と比較して, accuracy が 17.4%, macro F1-score が 12.2% 低下した.

生成画像の定量評価. 表 3 にはドメイン分類器による生成画像の分類結果, 表 4, 5 には病害診断モデルによる分類結果と混同行列を示す. 表 3 に示すように, DA-LoRA (Free-LoRA) によって生成された中立画像は, 通常の LoRA による生成画像と比較して, ドメイン分類器出力のエントロピーが 0.221 増加した. 表 4 に示した病害診断モデルによる生成画像の分類結果から, LoRA で生成した中立画像は高い精度を記録したが, DA-LoRA では精度が 0.248 低下し, F1-score は 0.307 低下した. また, 表 5 から健康クラスは正常に分類されたが, 病害クラスは平均で約 49.4% が健康に偏ったことが読み取れる.

生成画像の定性評価. 図 2 にそれぞれの生成画像を示す. LoRA による生成画像では, 背景にビニールハウスや支柱の

表 4: 中立画像のクラス分類結果.

	accuracy	precision	recall	macro F1-score
LoRA	0.990	0.990	0.990	0.990
DA-LoRA	0.742	0.846	0.648	0.683

表 5: 中立画像のクラス分類における混同行列.

	Healthy	Verticillium Wilt	Bacterial Wilt
Healthy	99.67	0.07	0.26
Verticillium Wilt	57.70	38.10	4.20
Bacterial Wilt	41.13	2.25	56.63

ような特徴が残り、葉の形などは Prior に示す SD3.5 の事前知識とは大きく異なっている。DA-LoRA では、地面の黒いビニールや支柱などがみられ、Verticillium Wilt では葉の全体に黄変が出て、葉の一部が茶色く萎れており、Bacterial Wilt では葉だけでなく植物全体が萎れるような特徴が現れた。Aware-LoRA では、地面に黒いビニールなどがみられ、Verticillium Wilt では葉全体にわたる黄変、Bacterial Wilt では植物全体の萎れが現れた。Free-LoRA では、背景には地面と短い草が見られ、Verticillium Wilt では葉が局所的に黄色く萎れ、Bacterial Wilt では植物全体に萎れが見られる。また、Bacterial Wilt は Aware/Free-LoRA で大きな病徴の違いが見られなかった。

5. 考察

分類精度. 結果から、ViT に DANN を適用する場合、accuracy, macro F1-score がどちらも大きく低下することが確認された。これは、ドメインに依存する特徴を除去したことで、学習ドメインへの過学習が悪化したためであると考えられる。DANN は、クラスの分類に利用できる特徴の中から、ドメインに依存する特徴を除去するため、クラスの分類に必要な情報がドメインに依存して変化する場合、十分な精度を得ることができないと考えられる。この結果は DANN がドメインに依存する特徴を除去する際に、クラスの分類に関わる特徴が除去されたことを示唆する。

生成画像の定量評価. 結果から、ドメイン分類器の出力するエントロピーの向上が確認された。エントロピーの最大値が 2.30 であることから、これは十分な値と言えないが、accuracy や macro F1-score の大幅な低下を伴っている。したがって、Free-LoRA の生成画像は理想的な中立画像ではないが、ドメインに依存した特徴が分離されたと考えられる。また、診断モデルを用いた推論では、病害画像の約半数が健康として推論された。これは、中立画像の持つ病徴を診断モデルが認識できていないためであると考えられる。このことは、DA-LoRA で明示的なドメインの分離を行った結果、ドメインの情報は低減されたが、同時にクラスを分類するための特徴が失われた、もしくは残されたクラスの特徴が微かであることを示している。

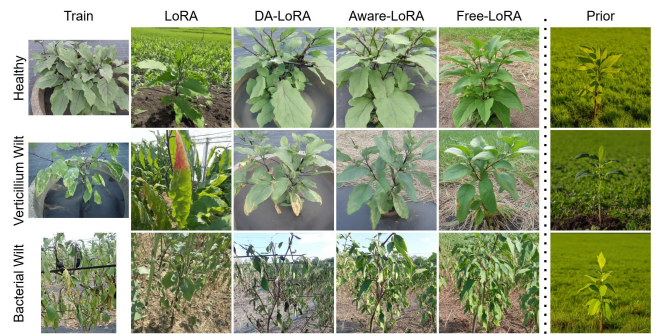


図 2: 画像生成結果 (Train は学習画像, Prior は LoRA を適用しない SD3.5 の生成画像).

生成画像の定性評価. LoRA による画像生成では背景にビニールハウスなどが残ったが、これはタスクに対して LoRA の rank が大きく、ドメインに過学習したと考えられる。DA-LoRA による生成では、地面に黒いビニールなどの背景、強い病害特徴が見られたが、これは Train の示す学習データの持つ特徴と似通っており、DA-LoRA は学習ドメインの画像を再現するために十分な特徴を持っていると考えられる。Aware-LoRA による生成では、ドメインに依存する背景だけでなく、病徴も同時に見られた。これは、敵対的学習によりドメイン分類可能な特徴が Aware-LoRA 側に蓄積されたが、その時に背景と同時に病徴の一部が Aware 側に学習されたことを示している。このことから、ドメインに依存する病徴が存在し、DANN などの手法の効果を妨げる要因であることが考えられる。また、ここで分離されたドメインに依存する背景と、ドメインに依存する病徴を分離することは、ドメインラベルのみを用いる学習では困難である。Free-LoRA による生成では、事前知識による生成画像に近い背景とクラスで異なる病徴のつき方が見られた。このことから、Verticillium Wilt ではドメインに依存しない病徴が局所的に現れ、Bacterial Wilt では植物全体に現れることが考えられる。また、Bacterial Wilt の病徴は、Aware と Free でほとんど変化しなかったが、これはドメイン差に病徴の強弱が含まれる場合、単純なドメイン除去では、ドメイン中立な病徴が減衰することが考えられる。

6. 結論

本研究では、ドメイン依存の特徴を画像生成モデルを用いて分離する DA-LoRA を提案し、植物病害におけるドメインシフト問題の分析を行った。DA-LoRA を用いた画像生成の結果、理想的ではないが、ある程度ドメインの特徴を分離した画像生成が可能となり、ドメイン調和手法の影響を画像として可視化した。可視化による分析の結果、植物病害では撮影環境などのドメインに依存した特徴の影響を強く受けることに加え、ドメインに依存する特徴として、病害の特徴が含まれることが明らかになった。今後の展望として、より規模の大きいデータにおける実験を行い、上記の分析で得られた課題の検証と解決を目指す。

参考文献

- [1] Cubuk, E. D., Zoph, B., Shlens, J. and Le, Q. V.: RandAugment: Practical Automated Data Augmentation with a Reduced Search Space, *NeurIPS* (2020).
- [2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, *ICLR* (2021).
- [3] Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., Podell, D., Dockhorn, T., English, Z., Lacey, K., Goodwin, A., Marek, Y. and Rombach, R.: Scaling Rectified Flow Transformers for High-Resolution Image Synthesis, *ICML* (2024).
- [4] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M. and Lempitsky, V.: Domain-Adversarial Training of Neural Networks, *JMLR*, Vol. 17, No. 59, pp. 1–35 (2016).
- [5] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Networks, *NIPS* (2014).
- [6] Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J. and Lakshminarayanan, B.: AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty, *ICLR* (2020).
- [7] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models, *ICLR* (2022).
- [8] Huang, J., Wang, Z., Xu, M., Ma, L., Wu, W. and Cao, J.: Enhancing Few-Shot Plant Disease Classification with Diffusion Model, *CAC* (2024).
- [9] Kang, G., Jiang, L., Yang, Y. and Hauptmann, A. G.: Contrastive Adaptation Network for Unsupervised Domain Adaptation, *CVPR* (2019).
- [10] Kudo, S., Kagiwada, S. and Iyatomi, H.: Few-shot Metric Domain Adaptation: Practical Learning Strategies for an Automated Plant Disease Diagnosis, *AAAI Workshop on AI2ASE* (2025).
- [11] Long, M., Cao, Z., Wang, J. and Jordan, M. I.: Conditional Adversarial Domain Adaptation, *NIPS* (2018).
- [12] Mohanty, S. P., Hughes, D. and Salathe, M.: Using Deep Learning for Image-Based Plant Disease Detection, *Front. Plant Sci.*, Vol. 7, 1419 (2016).
- [13] Motiian, S., Piccirilli, M., Adjeroh, D. A. and Doretto, G.: Unified Deep Supervised Domain Adaptation and Generalization, *ICCV* (2017).
- [14] Nogami, T., Kagiwada, S. and Iyatomi, H.: Robust Plant Disease Diagnosis with Few Target-Domain Samples, *VCIP* (2025).
- [15] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models, *CVPR* (2022).
- [16] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. and Aberman, K.: DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation, *CVPR* (2023).
- [17] Shibuya, S., Cap, Q. H., Nagasawa, S., Kagiwada, S., Uga, H. and Iyatomi, H.: Validation of Prerequisites for Correct Performance Evaluation of Image-based Plant Disease Diagnosis using Reliable 221K Images Collected from Actual Fields, *AAAI Workshop on AIAFS* (2022).