# Attention Lattice Adapter: 視覚言語基盤モデルのための説明生成

Attention Lattice Adapter: Visual Explanation for Vision-Language Foundation Models

平野 愼之助　飯田 紡　杉浦 孔明
Shinnosuke Hirano　Iida Tsumugi　Komei Sugiura

慶應義塾大学
Keio University

In the modern era where deep learning is applied across a wide range of fields, the explainability of models is of paramount importance. However, existing methods are not optimized for vision-language foundation models, leading to lower explanation quality for such models. Therefore, this study proposes the Alternative Adapter Model, an explanation generation model tailored to vision-language foundation models. By introducing a Side Branch Network connected to the vision-language foundation model, the proposed method extracts features suitable for explanation generation. Furthermore, by implementing the Alternative Epoch Architecture, which dynamically changes the outputs of modules and the layers to be frozen, we address the issue of overly narrow focus areas. To evaluate the proposed method, experiments were conducted using the CUB-200-2011 dataset. The results demonstrate that the proposed method surpasses existing methods in mean IoU, Insertion Score, Deletion Score, and Insertion-Deletion Score, which are standard metrics for visual explanation generation tasks.

## 1. はじめに

深層学習が幅広い分野に応用されている現代において，深層学習モデルの説明性は重要である [Shrikumar 17, Ribeiro 16]. 例えば，理論が未解明な自然現象の予測に深層学習を用いた場合，視覚的説明による重要な部分の可視化を通して，理論の洞察を与えることができる．一方で，複雑な深層学習モデルにおいては，判断根拠を説明することが困難であり，誤った根拠をもとに分類しているかどうかを見分けることが難しい．例えば，クレバーハンス効果 [Pfungst 07] のように，モデルが本質的な特徴ではなく，無関係な特徴に基づいて分類を行い，汎化性能の低下をもたらす可能性がある．そのため，判断根拠を明確化し，深層学習モデルの説明性を向上させることは有益である．本研究では，分類結果に対する判断根拠の視覚的説明生成タスクを扱う．特に，視覚言語基盤モデルに対する視覚的説明を生成するタスクに焦点をあてる．この視覚的説明は，画像内の分類対象物体に対するマスクと考えることもできる．この場合，セグメンテーションのマスクが正解マスクとして与えられず，クラスラベルのみを用いてマスクを生成するため，本タスクは image-level weakly supervised semantic segmentation タスクとみなすことができる．

本タスクは正解マスクを利用しない弱教師あり学習タスクであり，モデルの特徴や構造によって適切な説明生成手法は異なる．そのため，本タスクは正解マスクを利用せずに過不足なく適切な領域に注目する必要のある，困難なタスクである．

畳み込みニューラルネットワークを基盤とするモデルにおいて，視覚的説明の生成に関する研究は数多く提案されている [Selvaraju 17, Zhang 21, Petsiuk 18]．これらの手法は，既定の計算方法により説明を生成する．このような手法はモデルの構造に依存しないが，複雑なモデル構造に特化した説明の生成が難しく，不適切な領域に注目する場合がある．また，説明生成のための専用モジュールをブランチとして組み込んだ手法として，Attention Branch Network [Fukui 19] , Lambda Attention Branch Network [Iida 22] や PonNet [Magassouba 21] などが存在する．しかし，これらの手法は説明生成のための特徴を単一の層から抽出するため，性能が不十分であるという問題がある．

連絡先: 平野愼之助，慶應義塾大学，神奈川県横浜市港北区日吉 3-14-1，shinhirano@keio.jp

図 1: 本タスクにおける視覚的説明の例

本研究では，視覚言語基盤モデルである CLIP image encoder に対する視覚的説明モデルを提案する．既存手法との違いは，視覚言語基盤モデルである CLIP image encoder に対する視覚的説明モデルを提案し，モジュールの出力および freeze する層を動的に変更する Alternating Epoch Architect を導入した点である．本タスクはラベルのみを用いて学習をしていることから注目領域に対して直接損失を計算することができないため，注目領域の大きさを制限できない．その結果，少数のピクセルのみに注目してしまい，注目領域が狭くなるという問題がある．そこで，Alternating Epoch Architect を導入し，モジュールの出力および freeze する層を動的に変更することでこの問題を解消することができる．提案手法の新規性は以下の通りである．

- 視覚言語基盤モデルに Side Branch Network を導入した説明生成手法を提案する．
- 学習時，モジュールの出力および freeze する層を動的に変更する Alternating Epoch Architect を提案する．

## 2. 問題設定

本論文では，分類問題における判断根拠の視覚的説明生成を扱う．本タスクでは，モデルが正しく予測するために貢献し

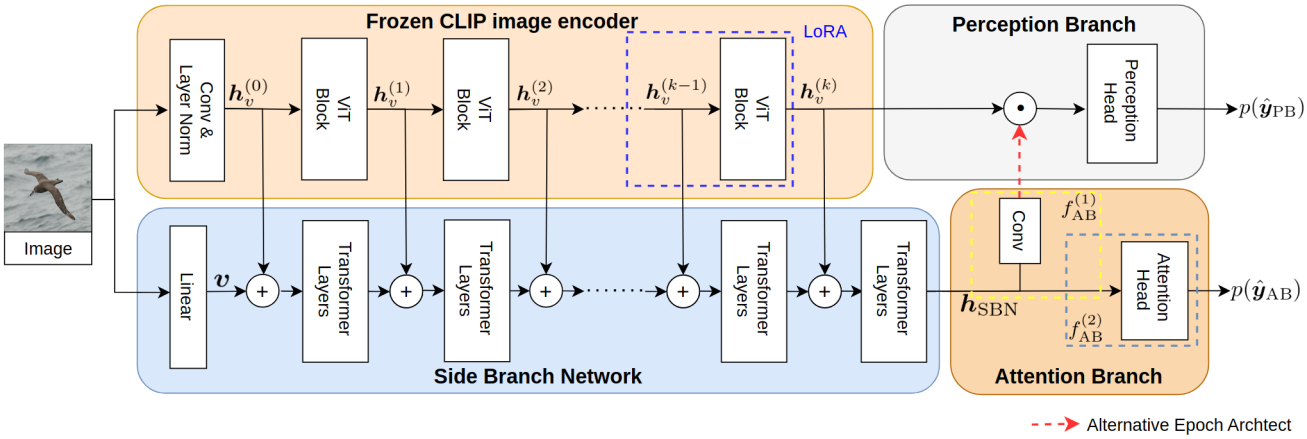



図 2: 提案手法のモデル構造

た画素を重要とする視覚的説明が望ましい．

本タスクでは，モデルが正しく予測するために貢献した画素を重要とする視覚的説明が望ましい．図 1 に本タスクにおける分類問題の例を示す．左図が入力であり，右図はモデルの注目領域を入力画像に重畳した画像である．入力は画像 $\boldsymbol{x} \in \mathbb{R}^{c \times h \times w}$ である．ここで，$c, h, w$ はそれぞれ入力画像のチャネル数，縦幅，横幅を表す．

出力は，入力画像がどのクラスに属するかの確率の予測値 $p(\hat{\boldsymbol{y}}) \in \mathbb{R}^C$ である．ここで，$\hat{\boldsymbol{y}}$, $C$ は，それぞれ 1-*of*-$K$ 表現とクラス数を表す．また，視覚的説明として画像中の各画素に重要度を割り当てた attention map $\boldsymbol{\alpha} \in \mathbb{R}^{h \times w}$ を利用する．

本研究では，基盤モデルとして事前学習済みの CLIP [Radford 21] を使用することを前提とする．本タスクの評価指標は mean IoU, Insertion Score, Deletion Score および Insertion-Deletion Score を用いる．

## 3. 提案手法

本論文では Side Adapter Network [Xu 23] に Attention Branch Network (ABN) [Fukui 19] を導入し，拡張した視覚的説明生成モデルを提案する．

本研究では，CLIP image encoder にアダプターを接続したセグメンテーションモデルである Side Adapter Network を拡張し，基盤モデルである CLIP image encoder に対する視覚的説明の生成を行う．そのため，CLIP image encoder を用いる手法全般に適用可能である．提案手法の新規性は以下の通りである．

- 視覚言語基盤モデルに Side Branch Network (SBN) を導入した説明生成手法を提案する．
- 学習時，モジュールの出力および freeze する層を動的に変更する Alternating Epoch Architect を提案する．

提案手法は，Frozen CLIP image encoder, SBN, Attention Branch (AB), Perception Branch (PB) の 4 つのモジュールから構成される．入力は $\boldsymbol{x}$ である．図 2 にモデルの構造を示す．

### 3.1 Frozen CLIP image encoder

本モジュールでは CLIP image encoder の中間特徴量を抽出する．CLIP image encoder を用いることで open-vocabulary な入力に対応した特徴量を抽出することができる．$\boldsymbol{x}$ を事前学習済みの $k$ ブロックからなる CLIP image encoder (ViT-B/16) [Radford 21] に入力し，1 ブロック目への入力 $\boldsymbol{h}_v^{(0)}$ および $i$ 個目のブロックにおける中間特徴量 $\boldsymbol{h}_v^{(i)}$ を得る．この時，より説明生成に適した特徴量を抽出するため，CLIP image encoder の $m$ 層目に Low-Rank Adoptation (LoRA) [Hu 22] を適用し，パラメータの差分のみを低ランク近似により学習する．LoRA によるパラメータ更新の式は以下で表される．

$$W = W^{(0)} + BA$$

ここで，$W$ および $W^{(0)}$ はそれぞれ更新後の重みおよび学習済みの重みを表す．また，$B \in \mathbb{R}^{d_1 \times r}$, $A \in \mathbb{R}^{r \times d_2}$ であり，パラメータの差分を低ランク近似する．

### 3.2 Side Branch Network

本モジュールでは $\boldsymbol{x}$ および CLIP image encoder の中間特徴量から説明生成のための特徴量を抽出する．ABN のように入力特徴の抽出に単一の層のみを用いて説明生成を行う場合，特徴抽出に適切な層を恣意的に決める必要があった．そこで，CLIP image encoder の 複数の階層から取得した特徴量を用いる．複数の階層から取得した特徴量を用いることにより，CLIP image encoder が学習中に獲得した知識を画素単位で適切に組み合わせることができるため，説明生成の精度が高まる．$\boldsymbol{x}$ を $w_1 \times w_2$ のパッチに分割後，全結合層に入力し，視覚特徴量 $\boldsymbol{v}$ を抽出する．その後，$\boldsymbol{v}$ を $l$ 層の transformer layer [Vaswani 17] に入力する．この時，各 $i$ 層 ($i = 1, 2, 3, 4$) の入力前に $\boldsymbol{h}_v^{(i-1)}$ を加算する．最終的に，transformer layer の $l$ 層目の出力 $\boldsymbol{h}_{\mathrm{SBN}}$ を SBN の出力とする．また，後述する注目領域が狭くなる問題を解決するため，エポック数 $e = 2n$ ( $n \in \mathbb{N}$) の時，SBN を freeze する．

### 3.3 Attention Branch

AB は $\boldsymbol{\alpha}$ を計算し，適切な領域に注目した説明生成を行うことができる．$\boldsymbol{\alpha}$ の計算と attention loss を計算するための確率の予測値 $p(\hat{\boldsymbol{y}}_{\mathrm{AB}}) \in \mathbb{R}^C$ を同時に行うことで，分類と結びついた説明を生成することができる．既存の ABN と同様にこの $\boldsymbol{\alpha}$ を $f_{\mathrm{AB}}^{(1)}$ の出力として使用することも可能である．しかし，ラベルのみを用いて学習をしていることから注目領域に対して直接損失を計算することができないため，注目領域の大きさを制限することが困難である．その結果，少数のピクセルにのみ注目してしまい，注目領域が狭くなるという問題がある．ここで，注目領域を拡大させることを促す正則化項を損失に含めることを考えられるが，適切な領域の大きさはタスクに大きく依存してしまうため，不適切である．そこで，本手法では Alternating Epoch Architect を導入し，エポック毎に以下を

表 1: ベースライン手法との比較および Ablation Study の定量的結果

| Model | LoRA | SBN | AEA | mean IoU ↑ | Insertion ↑ | Deletion ↓ | ID Score ↑ |
|---|---|---|---|---|---|---|---|
| RISE [Petsiuk 18] | | | | $0.390 \pm 0.014$ | $0.604 \pm 0.007$ | $0.086 \pm 0.002$ | $0.522 \pm 0.005$ |
| F-CAM [Belharbi 22] | | | | $0.550 \pm 0.017$ | $0.681 \pm 0.008$ | $0.034 \pm 0.001$ | $0.647 \pm 0.008$ |
| Ours (i) | | ✓ | ✓ | $0.520 \pm 0.041$ | $0.678 \pm 0.020$ | $0.019 \pm 0.002$ | $0.659 \pm 0.021$ |
| Ours (ii) | ✓ | | ✓ | $0.477 \pm 0.019$ | $0.595 \pm 0.054$ | $0.020 \pm 0.012$ | $0.575 \pm 0.045$ |
| Ours (iii) | ✓ | ✓ | | $0.495 \pm 0.008$ | $\mathbf{0.717 \pm 0.009}$ | $0.014 \pm 0.004$ | $\mathbf{0.702 \pm 0.011}$ |
| Ours (iv) | ✓ | ✓ | ✓ | $\mathbf{0.693 \pm 0.007}$ | $0.704 \pm 0.012$ | $\mathbf{0.007 \pm 0.002}$ | $0.697 \pm 0.011$ |

$f_{\mathrm{AB}}^{(1)}$ の出力とすることで，AB と PB をバランスし，この問題を解消する．

$$\boldsymbol{h}_{\mathrm{AB}} = \begin{cases} \boldsymbol{\alpha} & (e = 2n-1) \\ \mathbb{1} & (e = 2n) \end{cases}$$

ここで，$\mathbb{1}$ は $\boldsymbol{\alpha}$ と同じ形状で全要素が 1 の行列，$e$ はエポック数，$n$ は任意の自然数を表す．また，$e = 2n$ の時，AB を freeze する．$f_{\mathrm{AB}}^{(2)}$ は $\boldsymbol{h}_{\mathrm{SBN}}$ から attention loss を計算するための分類予測確率 $p(\hat{\boldsymbol{y}}_{\mathrm{AB}})$ を出力する．

### 3.4 Perception Branch

本モジュールは最終的なクラス予測を行うための分類を行う．$f_{\mathrm{PB}}$ は畳み込み層，プーリング層および全結合層より構成される．$f_{\mathrm{PB}}$ の入力は $\boldsymbol{h}_v^{(k)} \odot \boldsymbol{h}_{\mathrm{AB}}$ である．$\boldsymbol{h}_{\mathrm{AB}}$ は各画素に重要度を割り当てた attention map であるため，$\boldsymbol{h}_v^{(k)}$ と $\boldsymbol{h}_{\mathrm{AB}}$ のアダマール積を計算することで，予測に重要な領域を抽出することができる．$f_{\mathrm{PB}}$ の出力は分類確率の予測値 $p(\hat{\boldsymbol{y}}_{\mathrm{PB}})$ である．

### 3.5 損失関数

最終的なモデルの予測は以下の式で表される．

$$p(\hat{\boldsymbol{y}}_{\mathrm{PB}}) = f_{\mathrm{PB}}(\boldsymbol{\alpha} \odot \boldsymbol{h}_v^{(k)})$$

$$p(\hat{\boldsymbol{y}}_{\mathrm{AB}}) = f_{\mathrm{AB}}^{(2)}(\boldsymbol{h}_{\mathrm{SBN}})$$

$p(\hat{\boldsymbol{y}}_{\mathrm{PB}})$ は予測結果を出力するために利用する．$p(\hat{\boldsymbol{y}}_{\mathrm{AB}})$ は分類には直接用いないが，損失関数に導入することで説明の品質を向上させることができる．

損失関数 $\mathcal{L}$ として，以下を使用する．

$$\mathcal{L} = \mathrm{CE}(p(\hat{\boldsymbol{y}}_{\mathrm{PB}}), \boldsymbol{y}) + \lambda \mathrm{CE}(p(\hat{\boldsymbol{y}}_{\mathrm{AB}}), \boldsymbol{y})$$

ここで，$\boldsymbol{y}$，CE，$\lambda$ はそれぞれ正解ラベル，交差エントロピー誤差関数，損失関数の重みを表す．

## 4. 実験

### 4.1 データセット

実験では Caltech-UCSD Birds-200-2011 (CUB-200-2011) データセット [Wah 11] を用いた．CUB-200-2011 データセットは 200 種類の鳥の画像，鳥の位置を表す矩形領域およびマスク画像から構成される．CUB-200-2011 データセットは，視覚的説明生成タスクにおける標準データセットのため使用した．

本研究では，視覚的説明生成タスクの標準的な設定と揃えるため，CUB-200-2011 データセットのうち画像，マスク画像およびクラスを利用した．事前処理として画像の反転およびランダム切り抜きによるデータ拡張を行った．CUB-200-2011 データセットは 200 種類の鳥の画像 11,788 枚から構成される．各クラスの平均画像枚数は 58.9 枚である．訓練集合とテスト集合の分割に関しては，CUB-200-2011 データセットの標準の分割を使用した．加えて，訓練集合: 検証集合が 5:1 になるように訓練集合の一部を検証集合に割り当てた．最終的に，訓練集合，検証集合，テスト集合はそれぞれ 5,000，994，5,794 サンプルを含む．本論文では，訓練集合および検証集合をそれぞれパラメータの更新およびハイパーパラメータの選択に使用した．また，テスト集合を性能の評価に使用した．

### 4.2 定量的結果

表 1 にベースライン手法と提案手法との比較に関する定量的結果を示す．ベースライン手法として，RISE [Petsiuk 18] および F-CAM [Belharbi 22] を使用した．RISE は汎用的なモデルに適用可能な手法の中で標準的であるためベースライン手法とした．また，F-CAM は CUB-200-2011 データセットを扱う注目領域の可視化手法として標準的であるため使用した．本実験における評価尺度には， mean IoU, Insertion Score, Deletion Score， Insertion-Deletion Score (ID Score) を用いた．mean IoU， Insertion Score, Deletion Score, および ID Score は説明生成タスクの標準的な評価尺度であるため使用した．また，mean IoU を主要評価尺度とした．mean IoU は以下の式で定義される．

$$\text{mean IoU} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{IoU}(\hat{\boldsymbol{y}}_i, \boldsymbol{y}_i)$$

ここで，$N$ はサンプル数，$\hat{\boldsymbol{y}}_i$ , $\boldsymbol{y}_i$ は $i$ 番目のサンプルにおける予測及び正解マスク，IoU は 2 つのマスク間の Intersection over Union (IoU) を示す．

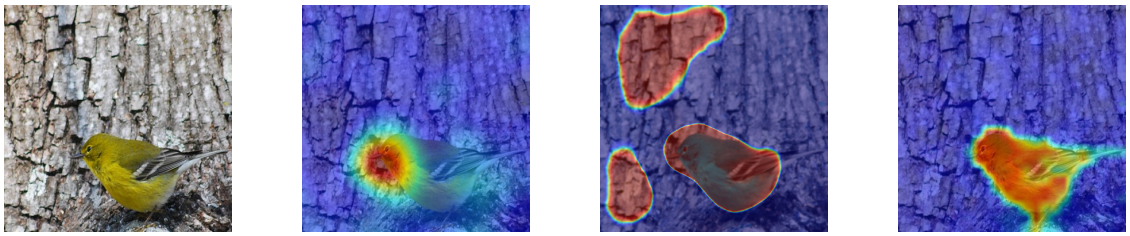
Insertion Score, Deletion Score は Insertion 曲線，Deletion 曲線の AUC で計算される．また，ID Score は Insertion Score と Deletion Score の差で定義される．ここで，Insertion 曲線，Deletion 曲線はそれぞれ $\boldsymbol{\alpha}$ を基に重要な領域を挿入, 削除した際の予測の変化を表す．詳細は以下で定義する．まず，$\boldsymbol{\alpha}$ の要素を降順に $\alpha_{i_1,j_1}, \alpha_{i_2,j_2}, \cdots, \alpha_{i_w,i_h}$ として，集合 $A_n$, $\boldsymbol{i}_n$, $\boldsymbol{d}_n$ を次のように定義する．

$$A_n = \{(i_k, j_k) | k \leq n\}$$

$$(\boldsymbol{i}_n, \boldsymbol{d}_n) = \begin{cases} (x_{ij}, 0) | (i,j) \in A_n \\ (0, x_{ij}) | (i,j) \notin A_n \end{cases}$$

ここで，$n$ は挿入・削除するピクセル数を表す．$\boldsymbol{i}_n$, $\boldsymbol{d}_n$ をモデルに入力した際の出力をそれぞれ $\boldsymbol{y}^{(\mathrm{ins},n)}$, $\boldsymbol{y}^{(\mathrm{del},n)}$ とする．このとき，$n$ と $\boldsymbol{y}_C^{(\mathrm{ins},n)}$, $n$ と $\boldsymbol{y}_C^{(\mathrm{del},n)}$ をプロットした曲線が，Insertion 曲線, Deletion 曲線である．ここで，$C$ は $\boldsymbol{x}$ が属するクラスを表す．

表 1 より，主要尺度である mean IoU において，RISE, F-CAM および提案手法はそれぞれ 0.390, 0.562 および 0.693 であり，提案手法はベースラインの中で最良であった F-CAM と比較して 0.131 ポイント上回った．また，ID score においては RISE, F-CAM および提案手法はそれぞれ 0.522, 0.647 および 0.697 であり，提案手法はベースラインの中で最良であった F-CAM と比較して 0.050 ポイント上回った．実験で使用したすべての評価尺度における性能差は統計有意であった（



(a) 元画像　(b) RISE [Petsiuk 18]　(c) F-CAM [Belharbi 22]　(d) Ours

図 3: ベースライン手法との定性的比較結果

$p < 0.05$).

### 4.3 定性的結果

図3 に定性的結果の成功例を示す．(a) は元画像を示し，(b),(c) はベースライン手法，(d) は提案手法によって生成した説明を元画像に重畳した結果を表す．図3(b) より，RISE によって生成された説明は鳥の領域の一部にのみ注目しており，不適切である．また，(c) より，F-CAM によって生成された説明は鳥の領域以外にも強く注目している．一方で提案手法は鳥の領域全体に注目しており，鳥の領域以外の注目度は低く，適切な説明を生成している．

### 4.4 Ablation Study

表1 に Ablation Study の定量的結果を示す．Ablation 条件として以下の 3 つを定めた．

1. Low Rank Adapter (LoRA) Ablation
   CLIP image encoder に適用している LoRA を取り除き，LoRA の有効性を調査した．表1 より，モデル (ii) における mean IoU および ID Score はそれぞれ 0.477 および 0.575 であり，モデル (iv) よりもそれぞれ 0.216 ポイントおよび 0.122 ポイント減少した．このことから，LoRA によって高品質な説明生成が促されたことが考えられる．

2. Side Branch Network (SBN) Ablation
   frozen CLIP image encoder の中間特徴量を加算しないことで中間特徴量を加算することの有効性を調査した．表1 より，モデル (i) における mean IoU および ID Score はそれぞれ 0.520 および 0.659 であり，モデル (iv) よりもそれぞれ 0.173 ポイントおよび 0.042 ポイント減少した．このことから，中間特徴量を加算することは説明生成に有用な特徴量を抽出していることが示唆される．

3. Alternating Epoch Architect (AEA) Ablation
   損失を変更せず，動的にモジュールの出力および freeze する層を変更することの有効性を調査した．表1 より，モデル (iii) における mean IoU は 0.495 であり，モデル (iv) よりもそれぞれ 0.198 ポイント減少した．このことから，AEA は注目領域に寄与したことが示唆される．

## 5. おわりに

本研究では，分類タスクにおける判断根拠の視覚的説明生成を扱った．本研究の貢献は以下である．

- 視覚言語基盤モデルに Side Branch Network を導入した説明生成手法を提案した．
- 学習時，モジュールの出力および freeze する層を動的に変更する Alternating Epoch Architect を提案した．
- 本タスクの標準的な評価尺度である mean IoU，Insertion Score, Deletion Score および Insertion-Deletion Score において，提案手法がベースライン手法を上回った．

**謝辞**

本研究の一部は，JSPS 科研費 20H04269，JST CREST，NEDO の助成を受けて実施されたものである．

## 参考文献

[Belharbi 22] Belharbi, S., Sarraf, A., Pedersoli, M., et al.: F-CAM: Full Resolution Class Activation Maps via Guided Parametric Upscaling, in *WACV*, pp. 3490–3499 (2022)

[Fukui 19] Fukui, H., Hirakawa, T., Yamashita, T., et al.: Attention Branch Network: Learning of Attention Mechanism for Visual Explanation, in *CVPR*, pp. 10705–10714 (2019)

[Hu 22] 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, in *ICLR* (2022)

[Iida 22] Iida, T., Komatsu, T., Kaneda, K., et al.: Visual Explanation Generation Based on Lambda Attention Branch Networks, in *ACCV*, pp. 3536–3551 (2022)

[Magassouba 21] Magassouba, A., Sugiura, K., et al.: Predicting and Attending to Damaging Collisions for Placing Everyday Objects in Photo-Realistic Simulations, *Advanced Robotics*, Vol. 35, No. 12, pp. 787–799 (2021)

[Petsiuk 18] Petsiuk, V., Das, A., and Saenko, K.: RISE: Randomized Input Sampling for Explanation of Black-box Models, in *BMVC*, pp. 151–164 (2018)

[Pfungst 07] Pfungst, O.: *Das Pferd des Herrn von Osten: der kluge Hans. Ein Beitrag zur experimentellen Tier- und Menschen-Psychologie*, Barth (1907)

[Radford 21] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., et al.: Learning Transferable Visual Models From Natural Language Supervision, in *ICML*, pp. 8748–8763 (2021)

[Ribeiro 16] Ribeiro, M., Singh, S., et al.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in *KDD*, pp. 1135–1144 (2016)

[Selvaraju 17] Selvaraju, R., et al.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in *ICCV*, pp. 618–626 (2017)

[Shrikumar 17] Shrikumar, A., et al.: Learning Important Features Through Propagating Activation Differences, in *PMLR*, Vol. 70, pp. 3145–3153 (2017)

[Vaswani 17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al.: Attention is all you need, *neurIPS*, Vol. 30, (2017)

[Wah 11] Wah, C., Branson, S., Welinder, P., Perona, P., et al.: The Caltech-UCSD Birds-200-2011 Dataset, Technical Report CNS-TR-2011-001, California Institute of Technology (2011)

[Xu 23] Xu, M., Zhang, Z., Wei, F., Hu, H., and Bai, X.: Side Adapter Network for Open-Vocabulary Semantic Segmentation, in *CVPR*, pp. 2945–2954 (2023)

[Zhang 21] Zhang, Q., et al.: Group-CAM: Group Score-Weighted Visual Explanations for Deep Convolutional Networks, *arXiv preprint arXiv:2103.13859* (2021)