

2024年 1月23日 情報工学実験II 研究発表会

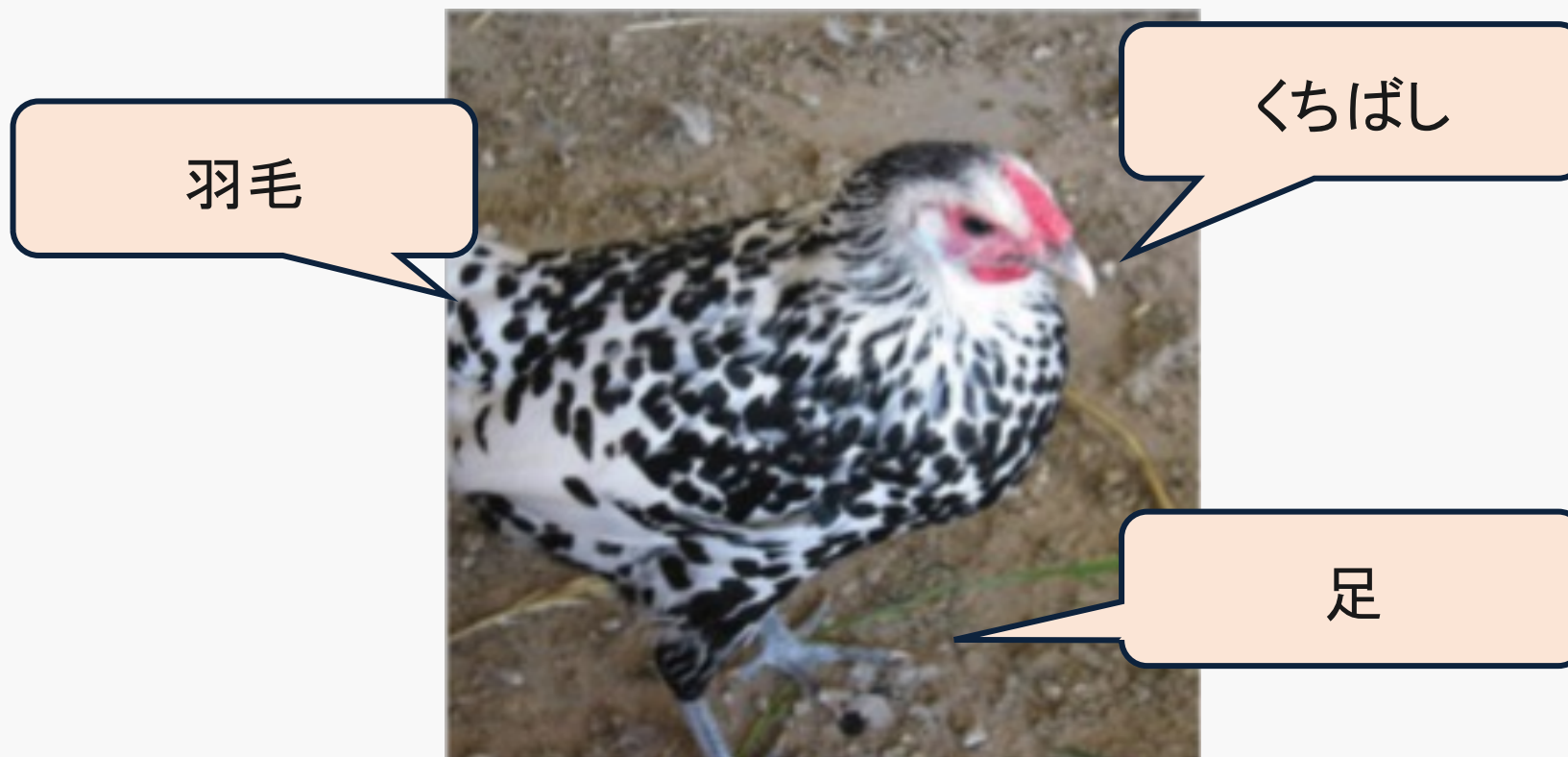
大規模言語モデルを用いた 説明可能性の高い画像分類 手法の再現

第3グループ 川越 壮

- 研究背景
- 先行研究
- 提案手法
- 実験
- まとめ

- 研究背景
- 先行研究
- 提案手法
- 実験
- まとめ

Q：人はどこに注目してめんどりと認識するか？



人が言葉で説明するのは簡単だが、そのような特徴を生成したり活用することは深層学習にとっては**重要な課題**

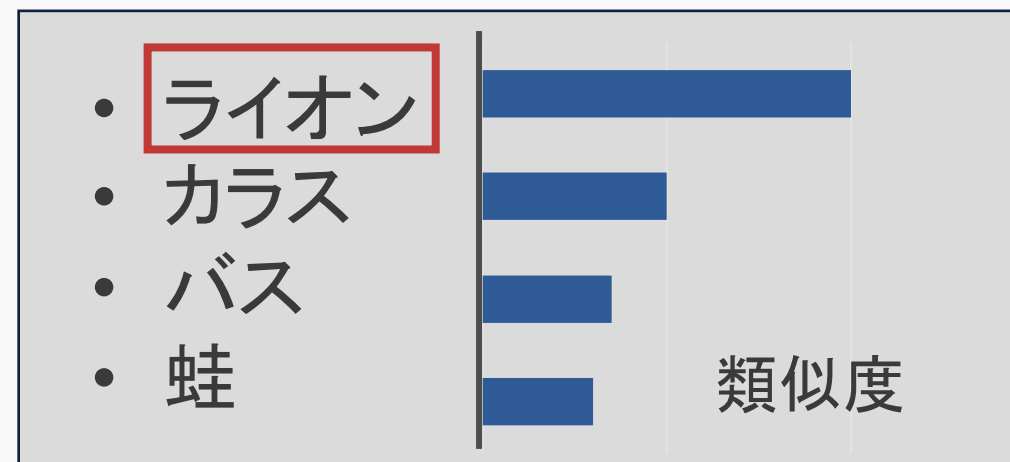
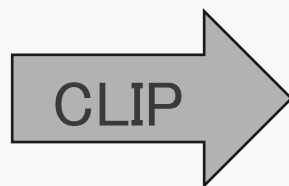
画像分類手法

Contrastive Language-Image Pre-training (CLIP) [Alec, ICML2021]

- Web上の4億組の画像とテキストのペアを学習
- 画像とカテゴリ名の類似度計算をし、高いものを選択



推論画像

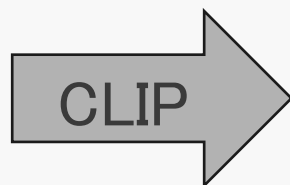


候補文

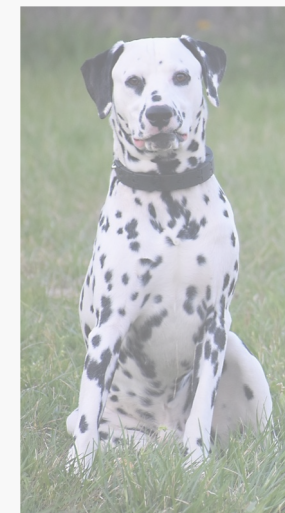
深層学習を用いた画像分類手法の課題

6

- 不合理な間違いをする
- 説明可能性が低い
→モデルの判断要因がブラックボックス化



Dalmatian (ダルメシアン)

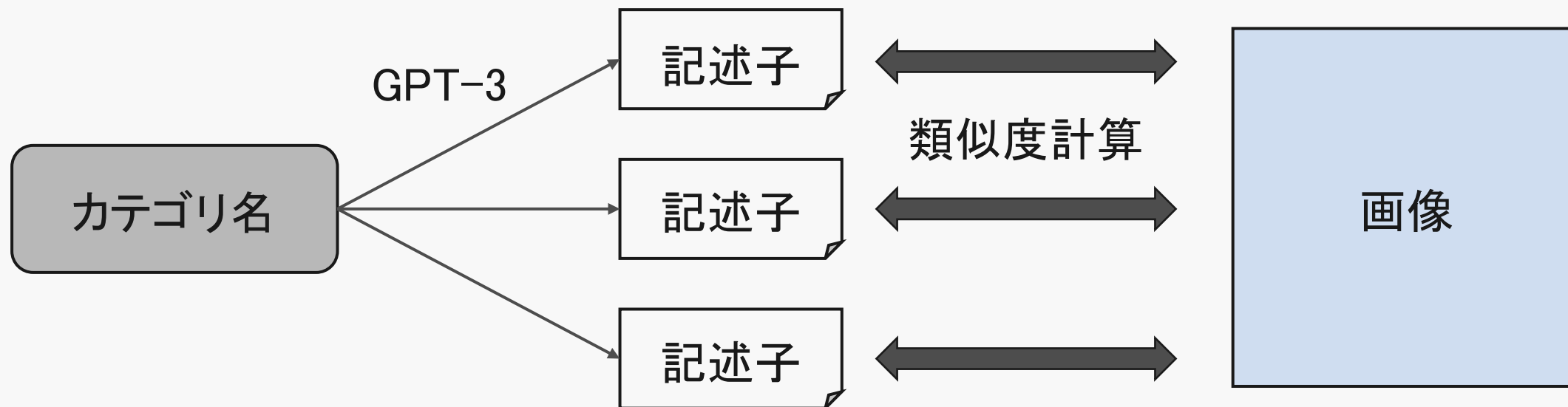


この課題を解決する手法の再現, および追加実験を行う

- 研究背景
- 先行研究
- 提案手法
- 実験
- まとめ

Classification By Description (以下CBD) [Sachit, ICLR2023] 8

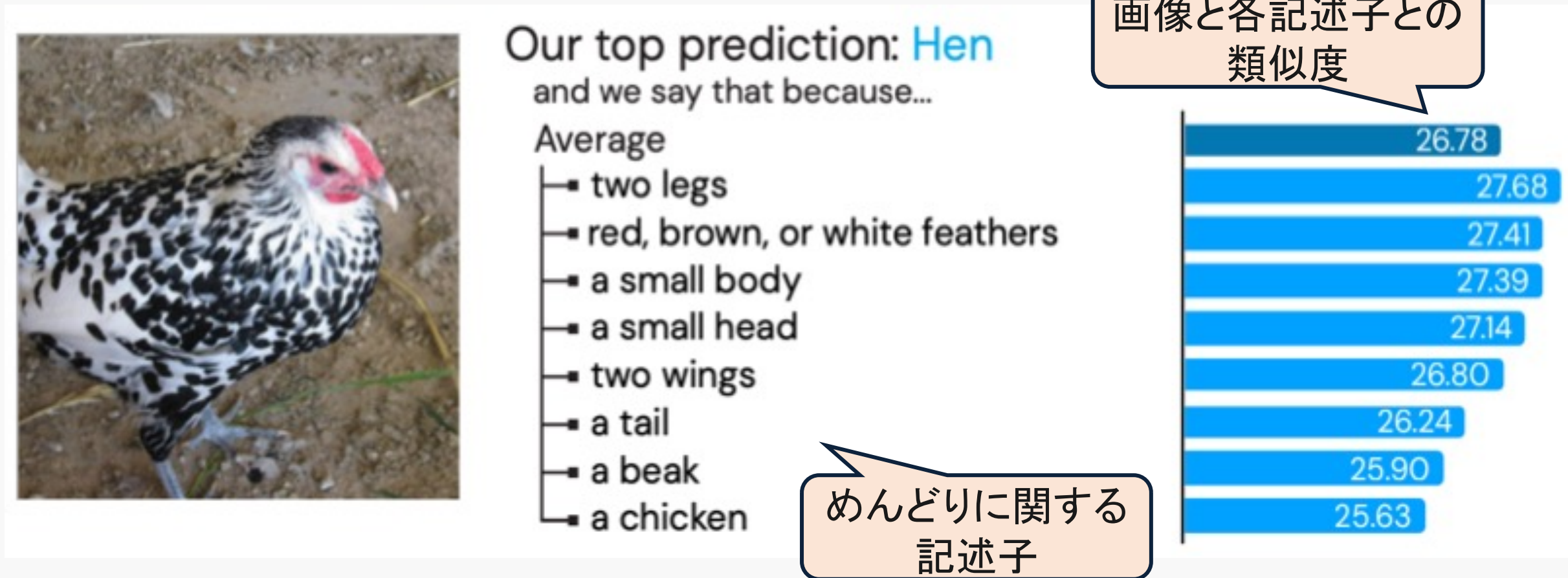
- カテゴリ名と画像の類似度を計算する流れはCLIPと同じ
- カテゴリ名だけでなく識別特徴(記述子)と類似度計算
 - 類似度の平均が最も高いカテゴリを選択
 - 記述子の生成にはGPT-3を用いる



CBDの推論例

9

記述子ごとの類似度を見ることでモデルの判断基準が分かる



Q: What are useful visual features for distinguishing a lemur in a photo?

A: There are several useful visual features to tell there is a lemur in a photo:

- four-limbed primate
- black, grey, white, brown, or red-brown
- wet and hairless nose with curved nostrils
- long tail
- large eyes
- furry bodies
- clawed hands and feet

} 記述子

} 1つ目の凡例
(キツネザル)

Q: What are useful visual features for distinguishing a television in a photo?

A: There are several useful visual features to tell there is a television in a photo:

- electronic device
- black or grey
- a large, rectangular screen
- a stand or mount to support the screen
- one or more speakers
- a power cord
- input ports for connecting to other devices
- a remote control

} 記述子

} 2つ目の凡例
(テレビ)

Q: What are useful features for distinguishing a {category_name} in a photo?

A: There are several useful visual features to tell there is a {category_name} in a photo:

-

} 推論したいカテゴリ

- 研究背景
- 先行研究
- 提案手法
- 実験
- まとめ

- ① 凡例のカテゴリ名をデータセットに特化したカテゴリにする
 - GPT-3に渡す凡例は分類したいカテゴリとは無関係であった
- ② 背景の特徴についての記述子を生成させる
 - 記述子は対象物の特徴のみであった

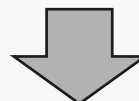
鳥のデータセットで実験する場合の例

13

Q: What are useful visual features for distinguishing a lemur in a photo?

A: There are several useful visual features to tell there is a lemur in a photo:

- four-limbed primate
- black, grey, white, brown, or red-brown
- wet and hairless nose with curved nostrils
- long tail
- large eyes
- furry bodies
- clawed hands and feet



①カテゴリを鳥の品種に

Q: What are useful visual features for distinguishing a Sooty Albatross in a photo?

A: There are several useful visual features to tell there is a Sooty Albatross in a photo:

- sark grey or sooty plumage overall
- thinner, darker bill compared to other albatrosses
- white rings around the eyes
- long, narrow wings with a slight upward curve at the tips
- shorter tail than other albatross species
- often a lighter grey or white throat area
- solitary or in small groups, usually found over open ocean

②背景に関する記述を追加

- 研究背景
- 先行研究
- 提案手法
- 実験
- まとめ

- ベースモデル
CLIP
- データセット
CUB for fine grained classification of birds (Wah et al., 2011)
(鳥類200種の画像5,794枚)
- CLIPのモデルサイズ
'ViT-B/32', 'ViT-B/16', 'ViT-L/14', 'ViT-L/14@336px'
- 評価指標
Zero-Shot Accuracy, Zero-Shot Top-5 Accuracy

Zero-Shot Accuracy

| モデルサイズ | 提案手法 | CBDの再現値 | CLIP |
|----------------|--------------|--------------|-------|
| ViT-B/32 | 54.04 | 53.14 | 52.00 |
| ViT-B/16 | 58.78 | 57.77 | 56.42 |
| ViT-L/14 | 65.22 | 64.69 | 63.07 |
| ViT-L/14@336px | 66.10 | 65.40 | 63.29 |

Zero-Shot Top-5 Accuracy

| モデルサイズ | 提案手法 | CBDの再現値 | CLIP |
|----------------|--------------|--------------|-------|
| ViT-B/32 | 84.35 | 84.17 | 82.59 |
| ViT-B/16 | 87.45 | 87.33 | 84.85 |
| ViT-L/14 | 92.44 | 92.01 | 90.44 |
| ViT-L/14@336px | 92.39 | 92.10 | 90.42 |

- 研究背景
- 先行研究
- 提案手法
- 実験
- まとめ

- CLIPにカテゴリの記述子を用いる説明可能性の高い画像分類手法を再現
- プロンプトを工夫することにより精度が向上

2024年 1月23日 情報工学実験II 研究発表会

大規模言語モデルを用いた 説明可能性の高い画像分類 手法の再現

第3グループ 川越 壮

大規模言語モデルを用いた 説明可能性の高い画像分類 手法の再現

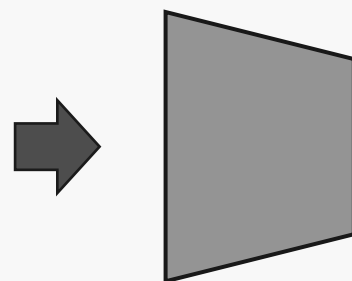
補足資料

B3 Deep Learning班 川越 壮

- 推論時

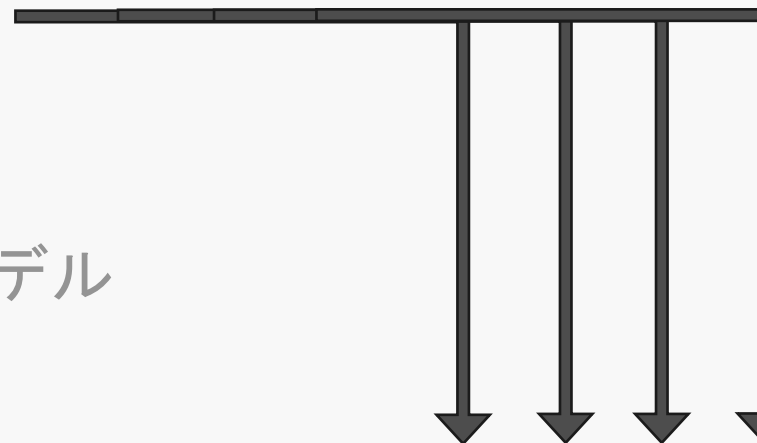


推論画像



特徴抽出モデル

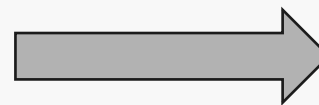
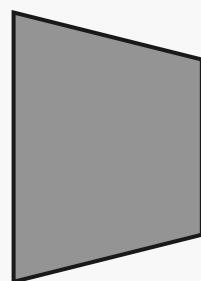
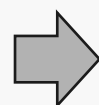
画像特徴ベクトル



類似度が最も高くなる候補を選ぶ

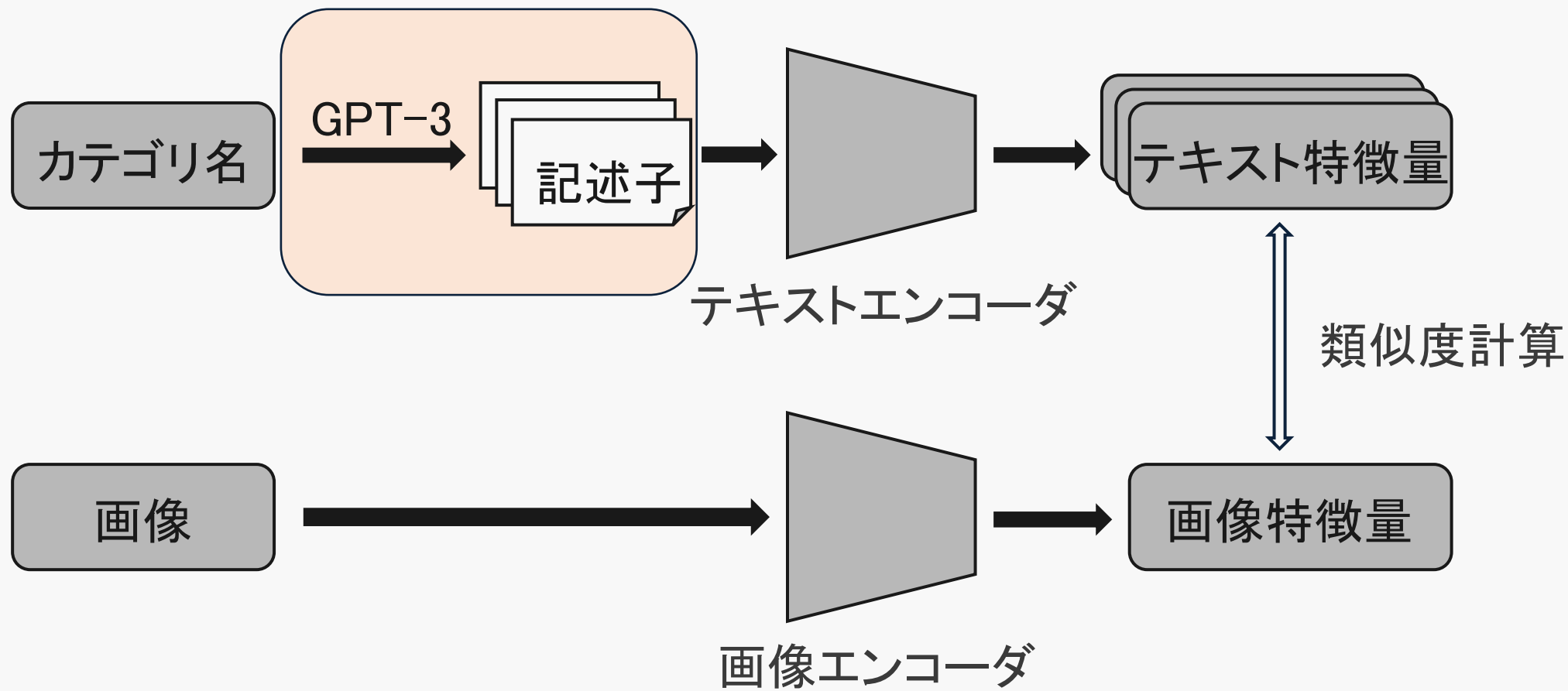
1. 草の上にいるライオン
2. 飛んでいるカラス
3. 黄色いバス
4. 水の中の蛙

候補文



テキスト特徴ベクトル





補足資料：背景に関する記述子を用いない場合の実験結果との比較 ²⁴

Zero-Shot Accuracy

| モデルサイズ | 提案手法 | 凡例のみ変えた場合 | CBD |
|----------------|-------|--------------|-------|
| ViT-B/32 | 54.04 | 53.57 | 53.14 |
| ViT-B/16 | 58.78 | 58.46 | 57.77 |
| ViT-L/14 | 65.22 | 64.60 | 64.69 |
| ViT-L/14@336px | 66.10 | 65.65 | 65.40 |

Zero-Shot Top-5 Accuracy

| モデルサイズ | 提案手法 | 凡例のみ変えた場合 | CLIP |
|----------------|-------|--------------|-------|
| ViT-B/32 | 84.35 | 84.24 | 84.17 |
| ViT-B/16 | 87.45 | 87.66 | 87.33 |
| ViT-L/14 | 92.44 | 92.22 | 92.01 |
| ViT-L/14@336px | 92.39 | 91.96 | 92.10 |

正解ラベル Horned Puffin (ツノメドリ)

予測ラベル Horned Puffin
Average

- often seen in large colonies
- habitat often includes rocky coasts and islands →
- bright orange legs and feet
- black and white plumage with a white face and chest
- bright yellow eyes with a black line running through them
- two black "horns" on the head

