

深層学習/DL 基礎 2024 最終課題 VQA 工夫点レポート

Omnicampus アカウント名：nekodaisuki

Omunicampus アカウント ID：7078

方針

ベースラインの README にあった「考えられる工夫の例」にあった四つの例を、一つずつ main.py に適応していく。

工夫点① 質問文の前処理

ベースラインコードの main.py にすでに実装済みだったと思われるので、そのまま使いました。

工夫点② 回答の出力候補の変更

README に示されていた class_mapping を利用して、回答の分類先の候補数を増やしました。

工夫点③ 画像の前処理

元々前処理として書いてあった Resize 以外に、以下の処理を追加しました。

1. ランダムな水平反転 (RandomHorizontalFlip): 画像をランダムに左右反転させます。
2. ランダムな回転 (RandomRotation): 画像をランダムに最大 10 度まで回転させます。
3. ランダムなアフィン変換 (RandomAffine): 画像をランダムに平行移動します。移動範囲は画像の 10%以内です。
4. カラー調整 (ColorJitter): 画像の明るさ、コントラスト、彩度、色相をランダムに変更します。各パラメータの変動幅は 0.1 です。
5. ランダムリサイズクロップ (RandomResizedCrop): 画像をランダムに切り取り、224x224 ピクセルのサイズにリサイズします。切り取る範囲は、元の画像の 80%から 100%の範囲内です。
6. ランダムグレースケール (RandomGrayscale): 画像をランダムにグレースケールに変換します。変換される確率は 10%です。
7. ランダムパースペクティブ変換 (RandomPerspective): 画像にランダムなパースペクティブ変換を適用します。歪みの度合いは 0.1 で、適用される確率は 10%です。