

Stable Diffusion のプロンプト文予測 (CLIP) をデータ拡張に活用 (generate_prompt.py と settings.ini について)

プロンプト文を生成する仕組み

Stable Diffusion で検出対象物の背景部分を画像生成し、検出対象物の画像と組み合わせることで画像を出力します。これに役立つ機能としては、第 2 部でも紹介した CLIP モデルです。この機能を使うと、入力した画像からプロンプト文が予測されます。

今回は、このプロンプト文の生成をデータ拡張に活用します。以下の図の赤点線で囲った部分が、ここで作成する Python プログラム (generate_prompt.py) です。また、その下の青点線で囲った部分は、Inpaint でデータ拡張に使用する処理です。赤点線の範囲で作られた予測されたプロンプト文を利用したり、参考にしたりして背景画像用のプロンプトを作ります。このようにプロンプト文を連携することで、データ拡張を簡単にします。

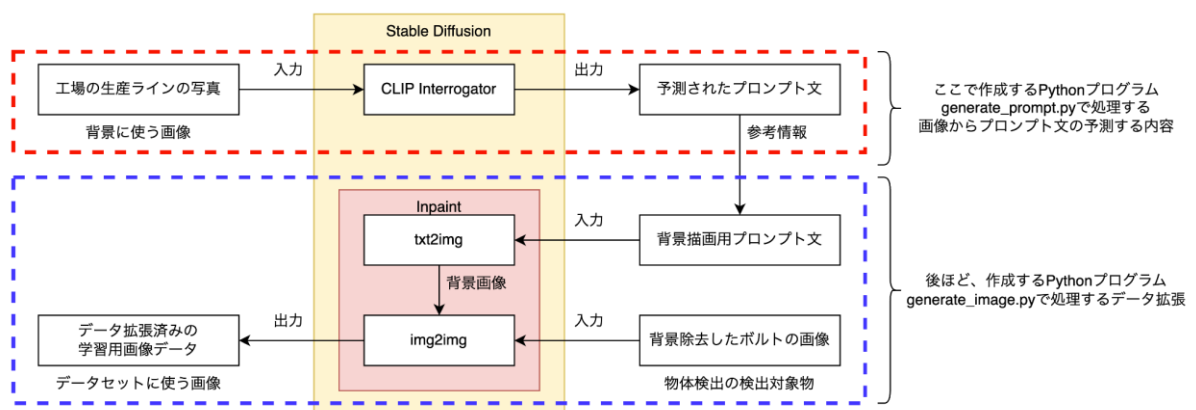


図 1 CLIP を利用してプロンプトを生成

プロンプト文を作成する Python プログラム

CLIP モデルを利用した Python プログラムで、画像からプロンプト文を作成します。自分で撮影した画像を、この Python プログラムに入力して利用できるようにします。Python プログラムの処理イメージを次の図 2 に示します。

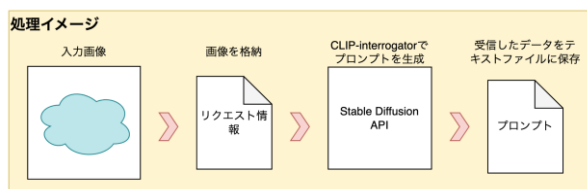


図 2 処理イメージ

Python プログラム (generate_prompt.py) の処理フローを図 3 に示します。Python プログラムから画像を読み込み、WebUI の API にリクエストを送信し、返答をテキスト・ファイルに保存する処理のみです。このプログラムは仮想環境 env-da から実行し、入出力ファイルはプログラムと同じ user フォルダに格納します。

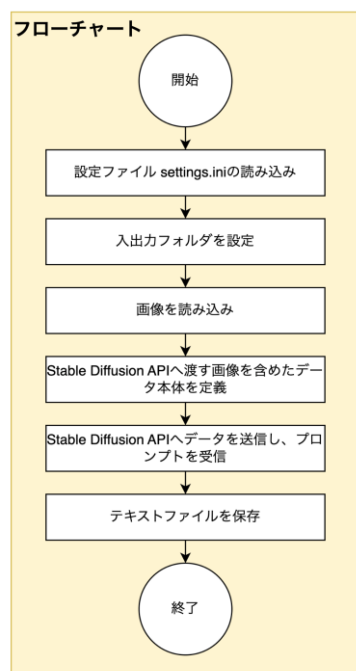


図 3 処理フローチャート

プログラム実行時の注意

プログラムの実行時の注意点として、実行前に仮想環境 **env-ga** を有効にすることと、**webui-user.bat**（サポートページの環境設定の資料参照）を実行して、Python プログラムからの API リクエスト送信を受診できるようにします（図 4）。

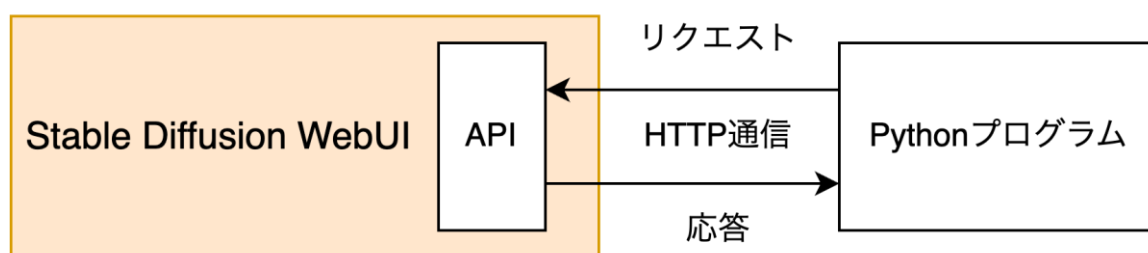


図 4 Python プログラムと API の関係

API との HTTP 通信を行うので、Web スクレイピングなどでよく使われる **urllib** ライブラリを使用しています。入力画像ファイル、出力テキスト・ファイルは、**settings.ini** の **GENERATE_PROMPT** で設定された内容を読み取って利用しています。Python プログラミング初心者でも、ソースコードを読むのに苦労しないはずで

プログラムで画像を API リクエストとして送信できるように、**Base64** エンコードで **ASCII** テキストに変換しています。ソースコードの「リクエスト・データを構成（**CLIP-img2txt**）」の部分では、**payload** 変数に格納する設定として、**Base64** エンコードされた画像データや使用する AI モデルを記述しています。これらを API に送信して返答される内容は、画像から予測されたプロンプト文のテキストになります。

Python プログラムの設定情報は、**settings.ini** に格納しています。**settings.ini** の中に、カギ括弧で囲われた文字列（例：**[PREPROCESSING]**）で記述されているのがセクションです。各 Python プログラムに対応したセクションに設定値などが記載されています。これを読み取って Python プログラムの中で利用します（図 5）。

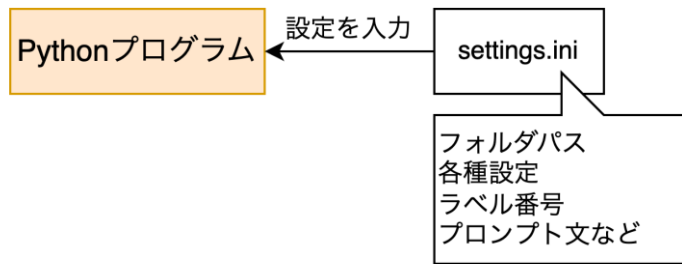


図 5 settings.ini の利用イメージ

主な記載内容は、入出力フォルダのパス、画像サイズや枚数などの各種設定、ラベル番号（クラス数）、プロンプト文やネガティブ・プロンプト文などです。必要に応じて適宜書き換えて利用します。Python プログラムは、settings.ini に記載されているセクション名と変数名で対象データを文字列を比較して識別します。

*

データ拡張では、複数のラベルの画像をデータ拡張すると思います。ラベルごとに一連のデータ拡張を処理しますので、異なるラベルのデータ拡張をするときには、事前に settings.ini の書き換えが必要な部分があります。

例えば、ラベル 0 のデータ拡張の処理が終わったら、ラベル 1 のデータ拡張をする前に settings.ini を修正します。修正点は、[ANNOTATION]セクションにある num_label = 0 を、num_label = 1 に書き換えてファイルを保存することです。