

それぞれの AI 物語通信 No.15

2024年7月21日

画像解析に関して

クラウドでの画像解析は Google Cloud Vision が、OpenAIより先に出していました。基本的にどちらも有償ですが、オリジナルの深層学習での初期経費(GPUサーバー、学習人件費等)との兼ね合いによります。Google Cloud もやってみました。API が非常に使いづらく(設定がややこしい)、GTP-4o(オーまたはオムニ)が出たので、こちらで実装しました。

<AIトレーナー玲子:画像解析>



プロンプト

```
"role": "system",
"content": "あなたは山ガールです。山ガールの口調で質問に答えてください。",
},
{
"role": "user",
"content": [
{"type": "text", "text": "提供された画像について説明してください。"},
{
```

回答

おお～、美味しそうな料理ですね！提供された画像には、いくつかの美味しそうな日本料理が見えますよ。

左側には、茶碗に入ったスープまたは煮物が見えます。野菜やお肉が入っているみたいで、温かそうで美味しそうですね～。山で食べると格別に美味しそう！

中央には、赤い小皿に盛り付けられたサラダのような料理が見えます。ネギやコーンなどの新鮮な野菜がトッピングされていますね。シャキシャキしていて、さっぱりとした味が楽しめそうです。

右側には、揚げ物(天ぷらや唐揚げのようなもの)が見えます。タレがかかっている、とても食欲をそそりますね！山でこんな揚げ物を食べたら、エネルギーが一気にチャージされそう。

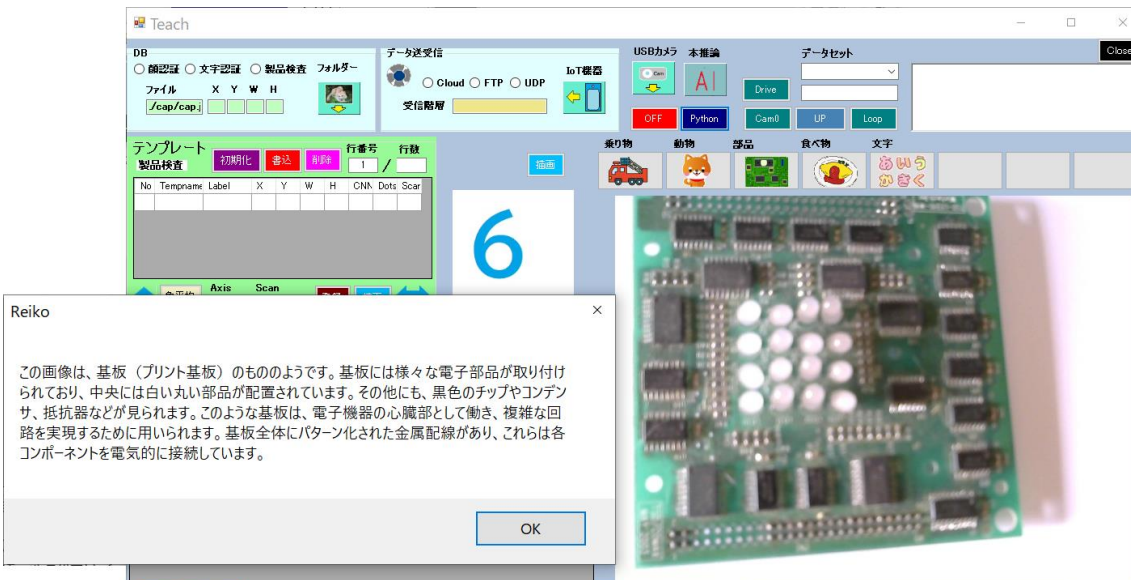
鍵盤も少し見えますので、オフィスでのランチかもしれませんね！おいしいごはんを食べて、午後も元気に頑張ってくださいね～。

その他の解析例

オリジナルの深層学習だと学習がものすごく大変です。PCのUSBカメラで撮影しました。gpt-4oでは茶トラを薄いオレンジ色と白の猫と言っていますが、これなら使えるでしょう。

料金は1枚当たり1.7円(レート150円/\$1)で、自分で学習すると自分の人件費や従業員の給与ははるかに超えていくでしょう。価格はどんどん落ちていきますよね(gpt-4oの料金はgpt-4の半分)。とはいえ画像も文章も、深層学習は進めていきます。これはオリジナル日本語LLM実装時に生きてくるからです。

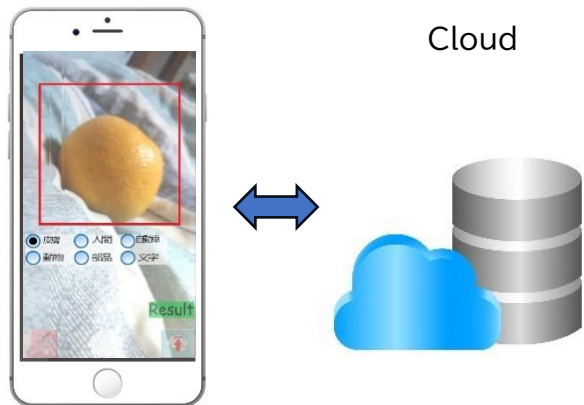




AIトレーナー玲子では2つの解析を用意しています。

1. オリジナルデータ ⇒ 深層学習で行います。
数種類の学習済データセットを用意しています。

2. GPT-4o による解析
Cloud 動作になります。
費用が発生します。
(後半で説明しています。)



プロンプト あなたは季節の画像の解説者です。質問に教えてください。提供された画像について説明してください。



この画像は、澄んだ川や湖の上にかかるアーチ型の橋を映しています。橋の材質はコンクリートでできており、堅牢な造りがうかがえます。橋のアーチ構造が水面に映り込み、ほぼ完全な円形を描いています。橋の両側には石が積まれた護岸があり、周囲の草木や木々が冬を迎えて葉を落としています。この景色から、澄んだ青空の下での穏やかな冬の日が感じられます。背景には一部電柱や電線も見え、都市部か郊外に位置していることを示しています。



この画像には、干支に関する飾り物が写っています。背景には金色の屏風があり、前面には白色と緑色の干支の動物の置物が並んでいます。これらは新年や正月の飾りとしてよく見られ、日本の伝統的な文化を象徴しています。中心の置物は獅子舞を模しているように見え、他の置物も細かいデザインが施されています。全体的に華やかで縁起の良い印象を与える画像です。



この画像には、雪が降っている中で暖かそうな格好をした女性が映っています。彼女はベージュのコートと緑色のマフラーを身に着けており、手を合わせて顔の近くにかざしています。彼女の表情は穏やかで、寒さを感じながらも楽しんでいるように見えます。背景にはぼんやりとした光が見え、冬の夜や夕方の雰囲気が漂っています。これは冬の寒い時期を表現した美しいシーンです。

Cloud 動作なので、自動運転には向きません。自動運転や Cloud を利用できない環境では、オリジナル 深層学習(画像解析、文章解析)と、オリジナル LLM をローカルに配置します。この時は(当方では)FPGA で開発し、ASIC という ROM のような IC に落として安価にします。ECU(電子制御ユニット)として、車載し、ローカルですべてを行います。

gpt-4o 料金

モデル	入力	出力
gpt-4o	\$5 / 1M tokens	\$15 / 1M tokens
gpt-4o-2024-05-13	\$5 / 1M tokens	\$15 / 1M tokens
gpt-4-turbo	\$10 / 1M tokens	\$30 / 1M tokens
gpt-4-turbo-2024-04-09	\$10 / 1M tokens	\$30 / 1M tokens

入力

$1,000,000 \div 765 \text{ トークン} = 1,307 \text{ 枚} \Rightarrow \$5 = 750 \text{ 円}$

1枚あたり 1.7円 レート 150円/\$1

画像の料金

- 1024 x 1024 の正方形の画像は detail: high 765 トークンかかります
 - 1024 は 2048 より小さいため、初期のサイズ変更は行われません。
 - 最短辺は 1024 なので、画像を 768 x 768 に縮小します。
 - 画像を表すには 4 つの 512 ピクセルの正方形タイルが必要なため、最終的なトークン コストは $170 * 4 + 85 = 765$ 。
- 2048 x 4096 の画像は detail: high 1105 トークンかかる
 - 2048 の正方形内に収まるように、画像を 1024 x 2048 に縮小します。
 - 最短辺は 1024 なので、さらに 768 x 1536 に縮小します。
 - 512px のタイルが 6 個必要なため、最終的なトークン コストは $170 * 6 + 85 = 1105$ 。
- 4096 x 8192 の画像は detail: low ほとんどの場合 85 トークンかかります

参考: 深層学習時経費

1. GPU サーバー 約 120 万円 (RTX4090 x 2 枚) 学習、推論両使い
2. 学習人件費 「1 画像 x 100 (単純画像) ~ 10,000 (キズ検査時) / 1 種類」 x 必要な種類

※今の方法では始めは GPT-4o などを使った方が無難ではないかと思われます。できるようになってから、オリジナルに移行すればよいこととなります。

ファインチューニングに関して

結論としてですが、Step1 から Step3 までで使用するデータフォーマットは同じなので、まずはこれを順番に進めていきます。最終的にオリジナル日本語LLMも自分のデータが専有されることにより自分のビジネスAIが初動します。

Step1 専門用語・文章の準備 ⇒ プロンプトで指示 **これが効果的と思われこれで開始します。**
 投資やビジネス構築時のノウハウなどの用語や文章の整理。
 初期の専門用語の登録(概念辞書編集)。



AIトレーナー玲子の文章解析で、辞書と文章を登録し、編集、保守していきます。

Step2 ファインチューニング(ChatGPT 作業)
 登録したファインチューニングデータでチューニングします。
 概念辞書や深層学習は引き続き編集し、精度をさらに上げます。

Step3 オリジナル LLM の構築
 ChatGPT のより専用化された日本語 LLM を構築します。
 これにより生成 AI への費用負担はなくなります。

※ Step1のやり方を、そのフォーマットと AIトレーナー玲子での操作方法を説明します。

<AIトレーナー玲子:文章解析>

The screenshot shows the Notion interface. At the top, there's a chat window titled 'LLM数' with a 'Close' button. The chat contains a conversation between System, User, and Assistant. The System prompt is 'キャラクターは元気いっばいの福のキャラクターです。'. The User says 'おはよう!'. The Assistant responds 'おはよう! 朝から元気に♪!'. The User says 'キャラクターは元気いっばいの。 今日一日がんばろう!'. The Assistant responds 'わわっ、朝から大忙しに♪! ?でもがんばろう! 今日もお仕事頑張ろう!'. The User says 'キャラクターは元気いっばいの。 今日もお仕事頑張ろう!'. The Assistant responds 'ありがと〜、助かるに♪! ご主人さまと〜'.

Below the chat window, there's a '概念辞書' (Concept Dictionary) section. It has buttons for '初期化' (Reset) and 'ファイル' (File). Below that, there are three tables: '事例(イベント)' (Cases/Events), '普通の概念' (Normal Concepts), and '特殊な概念' (Special Concepts). Each table has columns for 'Sel', '1', '2', '3', '4', '5', '6', '7', '8', '9'. The '事例(イベント)' table has rows for 'A' through 'J' with columns for '名', '家', '車', '貯金', '仕事', '趣味', '結婚', '子'. The '普通の概念' table has rows for 'A' through 'J' with columns for '名', '家', '車', '貯金', '仕事', '趣味', '結婚', '子'. The '特殊な概念' table has rows for 'A' through 'J' with columns for '名', '家', '車', '貯金', '仕事', '趣味', '結婚', '子'.

Step1 専門用語・文章の準備 ⇒ プロンプトで指示

まず、Step1のプロンプトでのオリジナル化を実行します。ここではファインチューニングをせずに、ChatGPT を呼ぶ時にプロンプトでオリジナル情報をセットし、回答にオリジナル性を活用します。これは、プロンプトを工夫することで、ChatGPT の回答に特定のオリジナル情報や設定を反映させることができます。AIトレーナー玲子の文章解析でセットしますが、以下の方法になります。これらの方法を用いることで、特定の情報やコンテキストに基づいたよりの確な回答を得ることが可能です。

(1)背景情報の提供

最初に必要な情報をプロンプトの中に含めることで、モデルがその情報を考慮に入れて回答します。

例 1: 設定の説明

あなたは 1980 年代のロック音楽に非常に詳しいミュージシャンです。プログレッシブロックの名作について教えてください。

例 2: 特定の条件

今日の天気予報を知りたいのですが、晴れだった場合にオススメの屋外アクティビティも教えてください。

(2)オリジナルのシナリオ設定

特定のオリジナルシナリオを設定することで、モデルがそのシナリオに基づいた回答を提供します。

例1:オリジナルシナリオ

あなたは未来の都市で働く探偵です。最新の AI 技術を使った事件解決法を教えてください。

(3)会話の流れを指示する

モデルに特定のトーンやスタイルで回答するよう指示することも有効です。

例:

これからの質問には、ユーモアを交えつつ、親しみやすい口調で回答してください。

(4)必要な情報の具体的指示

具体的な情報やフォーマットをプロンプトで指示することで、より精密な回答を得ることができます。

例:

次の情報を含めた自己紹介を書いてください: 名前、年齢、職業、趣味。そして、最後にユーモラスな一言も添えてください。

(5)カスタムメッセージの使用

プロンプトの冒頭にカスタムメッセージを提供し、これに基づいて応えてもらう形式です。

例:

次のカスタムメッセージを考慮に入れて、対応してください。カスタムメッセージ:『ユーザーは映画ファンであり、特に 1970 年代の映画に詳しい』。ユーザーが好きそうな 1970 年代の映画を 3 つ紹介してください。

考察

最近によく新しい AI 商品が出てきますが、まがい物も多くありそうです。AI は人工知能。人間の脳の模倣です。GPT などの汎用クラウドも全体構築で使いながら(社会構造のようなもの)、オリジナル深層学習と LLM が個々の個性になって動きます。このバランスを上手く設計し、データを入れた人間が、新しい製品を生むことができます。

話題の「rabbit r1」は欠陥だらけ。「もはや救いようがない」と海外メディアが酷評

https://getnavi.jp/digital/958899/?fbclid=IwZXh0bgNhZW0CMTAAAR1qfu0A2H9cPXyUt0PTaekPUuh9IpvSljFAxNegJHAK0i2DCW7AHffSPyE_aem_DBunBmqKe_ltS932xmkvvg



以上