

SONY



Vision and Sensing Application SDK モデル学習 機能仕様書

Copyright 2023 Sony Semiconductor Solutions Corporation

Version 0.2.0

2023 - 1 - 30

AITRIOS™、およびそのロゴは、ソニーグループ株式会社またはその関連会社の登録商標または商標です。

目次

1. 更新履歴	1
2. 用語・略語	2
3. 参照資料	3
4. 想定ユースケース	4
5. 機能概要、アルゴリズム	5
6. 操作性仕様、画面仕様	8
7. 目標性能	14
8. 制限事項	15
9. その他特記事項	16
10. 未決定事項	17

1.更新履歷

Date	What/Why
2023/01/30	初版作成

2. 用語・略語

Terms/Abbreviations	Meaning
MCT	モデルを量子化するためのオープンソースソフトウェア
Keras	AIモデルのフォーマットの一つ
TFLite	TensorFlow Liteのこと AIモデルのフォーマットの一つ
イテレーション	(1回あたりの)学習

3. 参照資料

- Reference/Related documents (関連資料)
 - Model Compression Toolkit (MCT)
 - https://github.com/sony/model_optimization

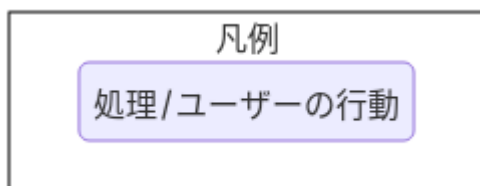
4. 想定ユースケース

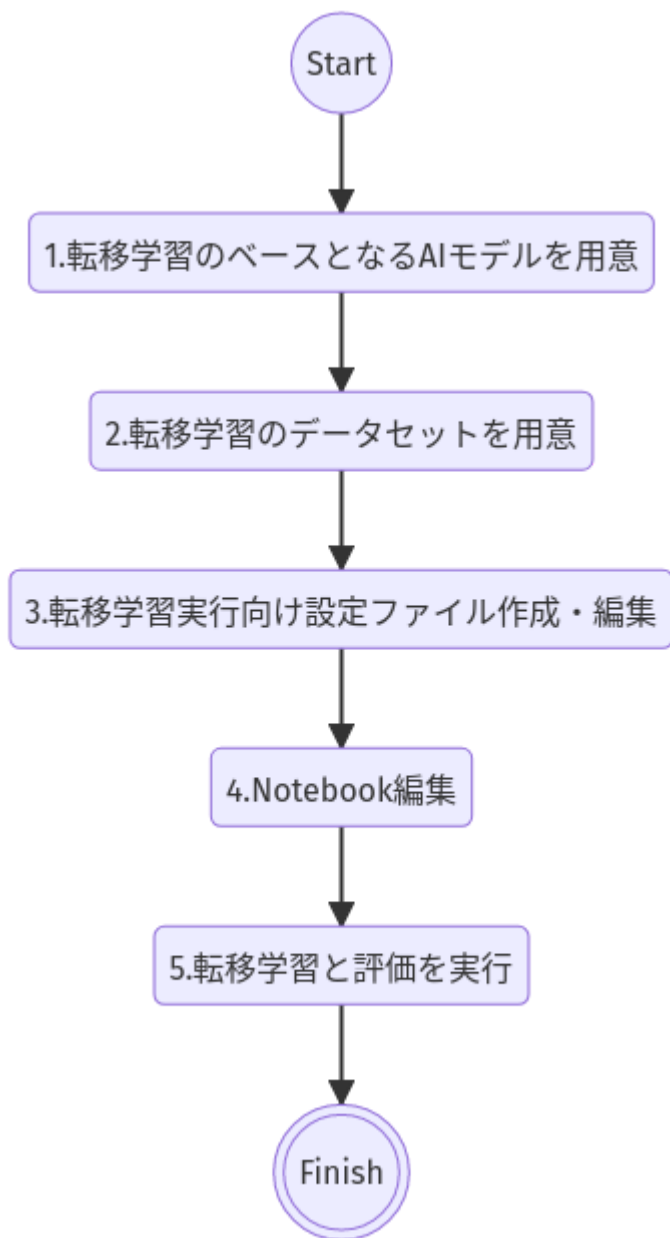
- 転移学習を実行したい
学習の過程で、推論実行し精度を確認したい

5. 機能概要、アルゴリズム

Functional Overview

- SDKにて下記のフローでImage ClassificationのAIモデル(Keras)を転移学習できる
- 転移学習したAIモデルで推論実行し、推論実行結果の統計値(Top1 accuracy)を取得できる
- SDKにてサポートするAIモデルは、MCTの [supported-features](#) に準拠する
- SDKにてサポートする画像フォーマットはJPEGとする
- フロー概要





- フロー詳細

1. 転移学習のベースとなるAIモデルを用意

- 転移学習のベースとなるAIモデル(Keras)を用意する

2. 転移学習のデータセットを用意

- 転移学習するためのデータセット画像とそのlabel情報を用意する

3. 転移学習実行向け設定ファイル作成・編集

- 設定ファイル`configuration.json`を作成、編集してNotebook実行時の設定を行う

4. Notebook編集

- ベースとなるAIモデルがTop(output)レイヤーを含んでいる場合は、Notebook内の`remove_top_layer_if_needed()`の実装を修正する

5. 転移学習と評価を実行

- 転移学習を実行し、推論評価するNotebookを実行する

6. 操作性仕様、画面仕様

How to start each function

1. SDK環境を立ち上げ、Topの **README.md** をプレビュー表示する
2. SDK環境Topの **README.md** に含まれるハイパーリンクから、 **tutorials** ディレクトリの **README.md** にジャンプする
3. **tutorials** ディレクトリの **README.md** に含まれるハイパーリンクから、 **3_prepare_model** ディレクトリの **README.md** にジャンプする
4. **3_prepare_model** ディレクトリの **README.md** に含まれるハイパーリンクから、 **develop_on_sdk** ディレクトリの **README.md** にジャンプする
5. **develop_on_sdk** ディレクトリの **README.md** に含まれるハイパーリンクから、 **1_train_model** ディレクトリの **README.md** にジャンプする
6. **1_train_model** ディレクトリの **README.md** に含まれるハイパーリンクから、 **image_classification** ディレクトリの **README.md** にジャンプする
7. **image_classification** ディレクトリの各ファイルから各機能に遷移する

転移学習のベースとなるAIモデルを用意

1. 転移学習のベースとなるAIモデル(Keras)を用意する
 - 転移学習のベースとなるAIモデル(Keras)を、SDK実行環境に格納する

転移学習のデータセットを用意

1. 転移学習のためのデータセット画像とlabel情報を用意する

- [ImageNet 1.0形式のフォルダ構成](#) のアノテーションデータを転移学習用と評価用の2つのフォルダで作成し、SDK実行環境に格納する
- tutorials/_common/datasetフォルダ内に格納する場合は、下記のように格納する

```
tutorials/  
└─ _common  
   └─ dataset  
      ├── training/ (1)  
      │   ├── 画像の分類名/  
      │   │   └─ 画像ファイル  
      │   ├── 画像の分類名/  
      │   │   └─ 画像ファイル  
      │   └─ ...  
      ├── validation/ (2)  
      │   ├── 画像の分類名/  
      │   │   └─ 画像ファイル  
      │   ├── 画像の分類名/  
      │   │   └─ 画像ファイル  
      │   └─ ...  
      └─ labels.json (3)
```

(1) 転移学習時に使用するデータセット

(2) 転移学習後の評価時に使用するデータセット

(3) label情報ファイル

- label情報ファイルのフォーマットは下記のようにlabel名とそのid値が記載されたjsonファイルとする

```
{"daisy": 0, "dandelion": 1, "roses": 2, "sunflowers": 3, "tulips": 4}
```



CVATでアノテーションを行ったデータセットをエクスポートしSDK実行環境に格納する方法は、[CVAT画像アノテーション 機能仕様書](#) を参照。

転移学習実行向け設定ファイル作成・編集

1. 実行ディレクトリに設定ファイル(`configuration.json`)を作成し、編集する



特別な記載がある場合を除き、原則として省略は不可。



特別な記載がある場合を除き、原則として大文字小文字を区別する。



原則としてシンボリックリンクのフォルダパス、ファイルパスは使用不可。

Configuration	Meaning	Range	Remarks
source_keras_model	転移学習のベースとなるAIモデル(Keras)パス。KerasのSavedModel形式のフォルダまたはh5形式のファイルを指定する	絶対パスまたはNotebook(*.ipynb)からの相対パス	未指定の場合、Keras標準のMobileNetV2のAIモデルを使用する動作となる
dataset_training_dir	転移学習の入力用データセット画像パス。 ImageNet 1.0形式のフォルダ を指定する	絶対パスまたはNotebook(*.ipynb)からの相対パス	
dataset_validation_dir	転移学習後の評価用データセット画像パス。 ImageNet 1.0形式のフォルダ を指定する	絶対パスまたはNotebook(*.ipynb)からの相対パス	
batch_size	転移学習の入力用データセットと評価用データセットのバッチサイズ	1以上(2のn乗を推奨)	
input_tensor_size	AIモデルの入力テンソルのサイズ(画像の一辺のピクセル数)	AIモデルの入力テンソルに準拠	
epochs	転移学習時のepoch数	1以上	
output_dir	転移学習したAIモデルの出力先となるディレクトリ	絶対パスまたはNotebook(*.ipynb)からの相対パス	

Configuration	Meaning	Range	Remarks
<code>evaluate_result_dir</code>	推論実行結果の統計情報を保存するディレクトリ	絶対パスまたは Notebook(*.ipynb)からの 相対パス	

Notebook編集

1. 実行ディレクトリの転移学習実行用Notebook(*.ipynb)を開く
2. ベースとなるAIモデルがTop(output)レイヤーを含んでいる場合は、Notebook内の
`remove_top_layer_if_needed()`の実装を修正する

転移学習と評価を実行

1. 実行ディレクトリの転移学習実行用Notebook(*.ipynb)を開き、その中のPythonスクリプトを実行する

。その後下記の動作をする

- 実行ディレクトリの`configuration.json`存在をチェックする
 - エラー発生時はその内容を表示し、中断する
- `configuration.json` `source_keras_model`、`dataset_training_dir` の存在をチェックする
 - エラー発生時はその内容を表示し、中断する
- `configuration.json` の下記の内容を読み取り、TensorFlowへ必要な設定を行い、転移学習する
 - `configuration.json` `source_keras_model`
 - `configuration.json` `dataset_training_dir`
 - `configuration.json` `input_tensor_size`
 - `configuration.json` `epochs`
- TensorFlowなどの外製ソフトでエラー発生時は、外製ソフトが出力するエラーを表示し、中断する
- `configuration.json` `output_dir` に、KerasのSavedModel形式のAIモデルを出力する
 - `output_dir` で指定するディレクトリがなければ作成し、そこに出力する
- 学習中はNotebookに下記のような表示をする(`epochs` が10の場合)

```
Epoch 1/10
3/3 [=====] - 4s 1s/step - loss: 1.6911 - acc: 0.3000 - val_loss: 1.8147 - val_acc: 0.1500
...
Epoch 3/10
3/3 [=====] - 2s 769ms/step - loss: 1.0132 - acc: 0.6750 - val_loss: 1.5243 - val_acc: 0.4000
...
Epoch 10/10
3/3 [=====] - 2s 673ms/step - loss: 0.2634 - acc: 0.9625 - val_loss: 1.1520 - val_acc: 0.6000
```

- `configuration.json` `dataset_validation_dir` の存在をチェックする
 - エラー発生時はその内容を表示し、中断する

- `configuration.json` の下記の内容を読み取り、TensorFlowへ必要な設定を行う
 - `configuration.json` `dataset_validation_dir`
 - `configuration.json` `output_dir`
 - `configuration.json` `evaluate_result_dir`
- 転移学習したAIモデルで推論実行し、統計情報を表示する
- 統計情報を、`evaluate_result_dir` 配下に `results.json` ファイルとして保存する
- TensorFlowなどの外製ソフトでエラー発生時は、外製ソフトが出力するエラーを表示し、中断する
- AIモデルの推論実行中はTensorFlowライブラリによるログを表示する
- 処理中でもNotebook Cell機能のStop Cell Executionで中断できる

7. 目標性能

- SDKの環境構築完了後、追加のインストール手順なしに、転移学習を実行できること
- UIの応答時間が1.2秒以内であること
- 処理に5秒以上かかる場合は、処理中の表現を逐次更新表示できること

8. 制限事項

- データセットのサイズによってはCodespacesのMachine Typeが4-coreでも転移学習時にメモリ不足でエラーになるため、その場合は8-core以上のMachine Typeを選択する必要がある

9. その他特記事項

- MCT(model-compression-toolkit)、TensorFlowのバージョン確認方法について
 - SDK環境のルートフォルダにある requirements.txt を参照する

10. 未決定事項

- なし