Class InferBase

Table of contents

- Defined in File infer.hpp

Inheritance Relationships

Derived Types

- `public holoscan::inference::OnnxInfer` (Class OnnxInfer)
- `public holoscan::inference::TorchInfer` (Class TorchInfer)
- `public holoscan::inference::TrtInfer` (Class TrtInfer)

Class Documentation

class InferBase

Base Inference Class.

Subclassed by holoscan::inference::OnnxInfer, holoscan::inference::TorchInfer, holoscan::inference::TrtInfer

Public Functions

virtual ~InferBase() = default

Default destructor.

inline virtual InferStatus do_inference(const std::vector<std::shared_ptr<DataBuffer>> &input_data, std::vector<std::shared_ptr<DataBuffer>> &output_buffer)

Does the Core inference.

Parameters

- **input_data** – Input DataBuffer
- **output_buffer** – Output DataBuffer, is populated with inferred results

Returns

InferStatus

inline virtual std::vector<std::vector<int64_t>> get_input_dims() const

Get input data dimensions to the model.

Returns

Vector of values as dimension

inline virtual std::vector<std::vector<int64_t>> get_output_dims() const

Get output data dimensions from the model.

Returns

Vector of input dimensions. Each dimension is a vector of int64_t corresponding to the shape of the input tensor.

inline virtual std::vector<holoinfer_datatype> get_input_datatype() const

Get input data types from the model.

Returns

Vector of input dimensions. Each dimension is a vector of int64_t corresponding to the shape of the input tensor.

inline virtual std::vector<holoinfer_datatype> get_output_datatype() const

Get output data types from the model.

Returns

Vector of values as datatype per output tensor

inline virtual void cleanup()