ACR – Nvidia reference implementation

Connectivity to ACR Models using ACR TRIAD

TRIAD is a tool developed by the ACR to increase site participation in clinical studies, accreditation, and registries. It can integrates with local systems to search, extract, anonymization, and send data to various ACR programs. ACR's AI-Lab will be built on top of the TRIAD foundation to allow AI models to be safely moved between participating sites. From there the sites can chose to create new AI models, customize AI models to local needs, create data sets for training, participate in cross institutional AI research, or participate in the validation of AI models.

In the simplest case, the users would activate the interface, upload the model and metadata, and choose the institutions with which the model can be shared for training, validation, or strictly inference. Receiving institutions, can download the models via the user interface and place them in their appropriate training, validation, or inference environments. Optionally, TRIAD may also support (through the data available in DART) the validation of models on relevant annotated data using TLT.

Configuring TRIAD for Sharing Models

This section assumes

- All necessary data-sharing agreements are in place and valid
- Both hospitals have TRIAD deployed

Do not proceed unless these are complete.

Steps:

1. Before beginning, a custom “research project” will set up by the ACR in which both sites are enrolled.
2. At the sending facility, the model and transforms are compressed into a single file using ZIP.
3. Using the TRIAD client, the sending facility pushes the compressed file.
4. Once the upload is complete, the sending facility notifies the receiving facility.
5. Using the TRIAD client, the receiving facility retrieves the compressed file.

6. The receiving facility decompresses the file, and then uses these files for performing validation, fine-tuning, and inference, as desired.

Annotation and Training using NVIDIA Clara Train SDK

Deep learning models are sensitive to the data they are trained on. Factors such as varying scanner configurations, age differences of patients, and so on, need to be considered. This makes it hard to train the deep learning models on a specific dataset and deploy them to be used on a different dataset for annotation reproducing same accuracy.

The Clara Train SDK provides several tools to address these needs in a hospital system; these are defined in the following sections.

AI-Assisted Annotation toolkit

Having annotated data is key to training a model and strengthening the specificity and sensitivity, but it is a time-consuming and tedious ordeal. In order to accelerate the annotation process, a hospital system or solution provider (like a PACS or viewing tool) can leverage Clara Train SDK’s AI Assisted Annotation SDK.

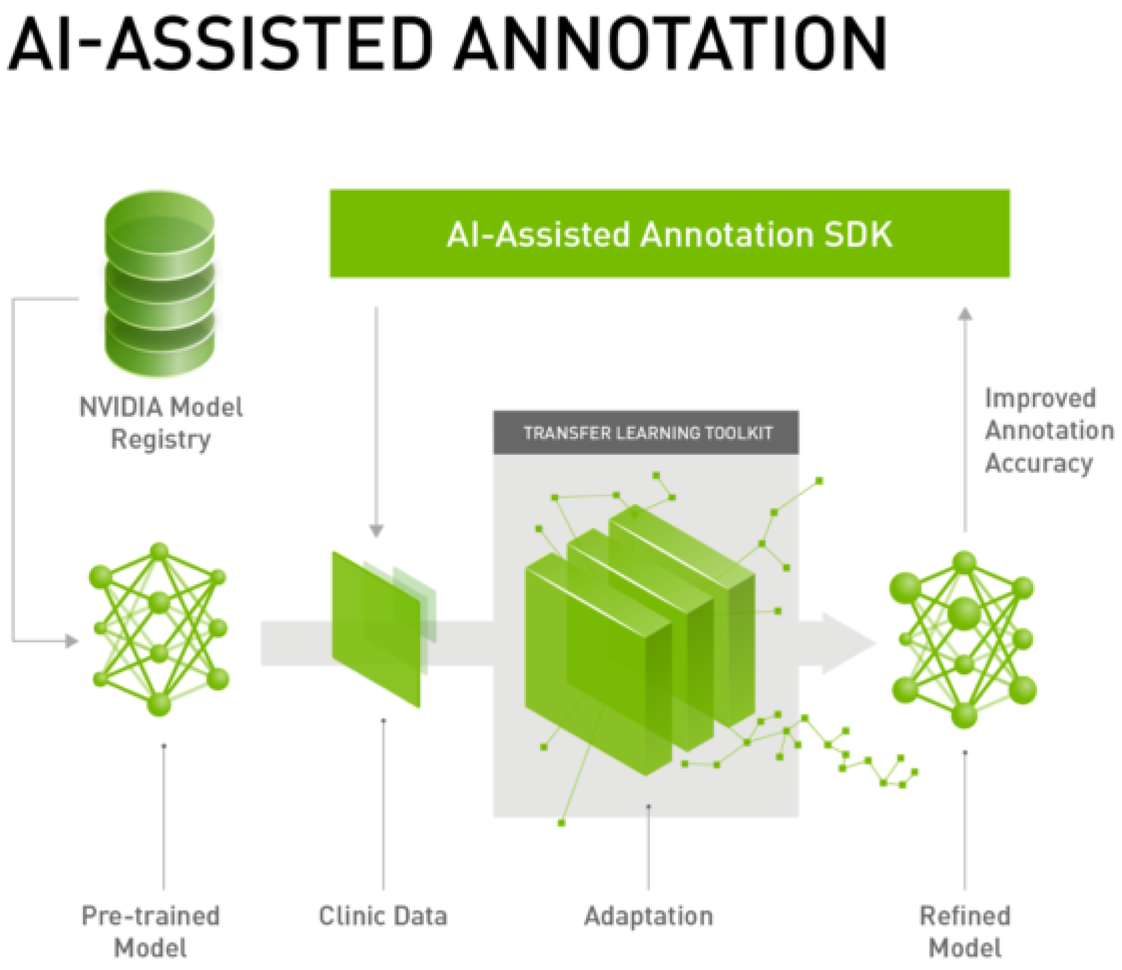

Figure 2: AI-Assisted Annotation SDK. Image courtesy NVIDIA Corporation.

The AI-Assisted Annotation SDK enables application developers to integrate the deep learning tools built into the SDK with their existing medical imaging applications, such as MITK. This is accomplished using a simple API and requires no prior deep learning knowledge.

Transfer Learning toolkit (TLT)

Transfer learning fits very well in medical image analysis. Given that medical image analysis is often seen as a computer vision task, Convolutional Neural Networks (CNNs) represents one of the best performing methods for this. Getting a large well-annotated dataset is considerably harder in the medical domain compared to the general computer vision domain because of a variety of issues, including:

- Availability and cost of the domain expertise required to annotate the medical images
- Legal and ethical concerns with accessing patient medical data

Reducing the number of medical imaging studies necessary for a hospital system to leverage an AI model reduces the barrier to adopting AI. This makes transfer learning a natural fit for medical image analysis, as using pre-trained CNN on larger databases and then transfer-learning to a target domain of medical images with limited availability.

The Transfer Learning Toolkit (TLT) enables participating institutions to share data pipelines, and train and deploy models in their own environment. TLT provides an intuitive interface for model training, where the programming is abstracted from the user via configurations files that, in turn, allow for easy integration of GUI-based training/inference configuration pipeline.

Specifically, hospitals use TLT to accomplish the following activities:

- Combine a trained model with locally acquired annotated datasets to improve local specificity and sensitivity (tlt-train)
- Perform validation of models prior for use with local data (tlt-validate)
- Export a trained model for archive, or to share with the ACR or another institution (tlt-export)

Exported models from TLT can be used in inference workflows (e.g. via Clara SDK, TensorRT Inference Server, or TensorFlow Serving). In addition, TLT describes the data pipeline and model training hyperparameters in a JSON configuration file allowing for the portability of models and data pipelines across institutions.

DICOM Adapters and Pipelines using NVIDIA Clara Deploy SDK

Once a quality-assured neural network becomes available, the Clara Deploy SDK provides the framework and tools required to define an application workflow based on the algorithm developed/adapted during the Training phase. The Clara Deploy SDK provides a container-based development and deployment framework for building AI-accelerated medical imaging workflows. The SDK uses Kubernetes under the hood, enabling developers and data scientists the ability to define a multi-staged container-based pipeline.

The core capabilities of Clara Deploy SDK include:

- Data Ingestion: Includes a containerized DICOM Adapter interface to communicate with hospital PACS and other imaging systems (both to receive and transmit data)
- Pipeline Manager and Core Services: Provides container-based orchestration, resource management & services for TensorRT based inference and Rendered Images Streaming
- Sample Deployment Workflows: Includes capabilities to define and configure container based workflows using sample workflow with user defined data or modified with user-defined-AI algorithms
- Visualization Capabilities. Enables the ability to monitor progress and view results