TASTI

Application-TAilored SynThetic Image generation

Labelled in Xecs Call 1, a EUREKA cluster

Xecs Project Number 2022-005

Initial framework description

Due date of deliverable: M12

Start date of project: 1 January 2023		**Duration:** 36 months
Organisation name of lead contractor for this deliverable:		Philips
Author(s):	Stefan Schalk	
Status:	Draft	
Version number:	1.0	
Submission Date:	05-Jan-2024	
Doc reference:	TASTI_D1.3 - Initial framework description - v1.0	
Work Pack./ Task:	WP1 – Task 1.3	
Description: ***(max 5 lines)***	Initial description of the TASTI framework for synthetic image generation, including reference architecture and expected building blocks.	

Nature:	Report		
Dissemination Level:	**PU**	Public	**x**
	PP	Restricted to other programme participants	
	RE	Restricted to a group specified by the consortium	
	CO	Confidential, only for members of the consortium	

DOCUMENT INFO

Author

Author(s)	*Company*	*e-mail*
Stefan Schalk	Philips	stefan.schalk@philips.com
Co Melissant	MonkeyProof Solutions	co.melissant@monkeyproofsolutions.nl
Raffaele Imbriaco	Sioux	raffaele.imbriaco@sioux.eu

Documents history

Document version #	*Date*	*Change*
V0.1	9 Nov 2023	Initial version for internal review
V0.2	18 Dec 2023	Added hardware to infrastructure. Some typos corrected.
V1.0	5 Jan 2024	Version submitted to portal. Changes: - Added blocks “auto-encoder-decoder”, “NF”, “Image Deformation” to Figure 1 - Added description for Image Deformation in Section 2.3 - Added reference data/images as domain-specific input in Section 2.2. - Mentioned idea of having an abstraction layer for hardware in Section 2.7

Document data

Keywords		Synthetic image generation framework
Editor Address data	Name: Partner: e-mail:	Stefan Schalk Philips stefan.schalk@philips.com
Delivery date		05-Jan-2024

Table of Contents

Glossary

Abbreviation / acronym	Description
AI	Artificial Intelligence
Dx.x	Deliverable x.x (project deliverable)
GAN	Generative Adversarial Network
HW	Hardware
IQ	Image quality
NF	Normalized Flow
SW	Software

1. Introduction

The TASTI framework aims to simplify synthetic image generation with properties that can be tailored to the applications for which they are used. For some applications it may be important to have very realistic images with the right noise characteristics. In other cases, a broad range of scenarios or image artifacts and augmentations are required, for example to add robustness to computer vision applications. In simulators, the computation time at which images are generated could be more important than noise characteristics. Finally, assessment of the quality of the generated images against the requirements of the target application will also be included in the framework.

The TASTI framework will provide an architecture and technical elements to generate and assess synthetic images. The technical elements will come from the technical work packages 2 to 4 and are designed to be reusable across applications with a degree of customization. We aim to demonstrate the TASTI framework in each application through the demonstrators in work package 5. It is not expected that the entire framework will be used in all applications, but each element should be demonstrated in at least one use case.

This document provides the first version of the reference architecture of the TASTI framework for synthetic image generation. The term “first version” is used here because the architecture is expected to change in the course of the project. The intention is not to set a framework to which all use cases should comply, but rather extract the framework from the use cases that are presented by the industrial partners. As a result, learnings from those cases may lead to modifications of the framework.

This first version has been based on the use case descriptions and requirements drafted in deliverables D1.1 and D1.2. Via a series of workshops with use case and technology providers, a first architecture of the framework has been created, which is presented in this document.

2. Reference architecture

2.1 Framework overview

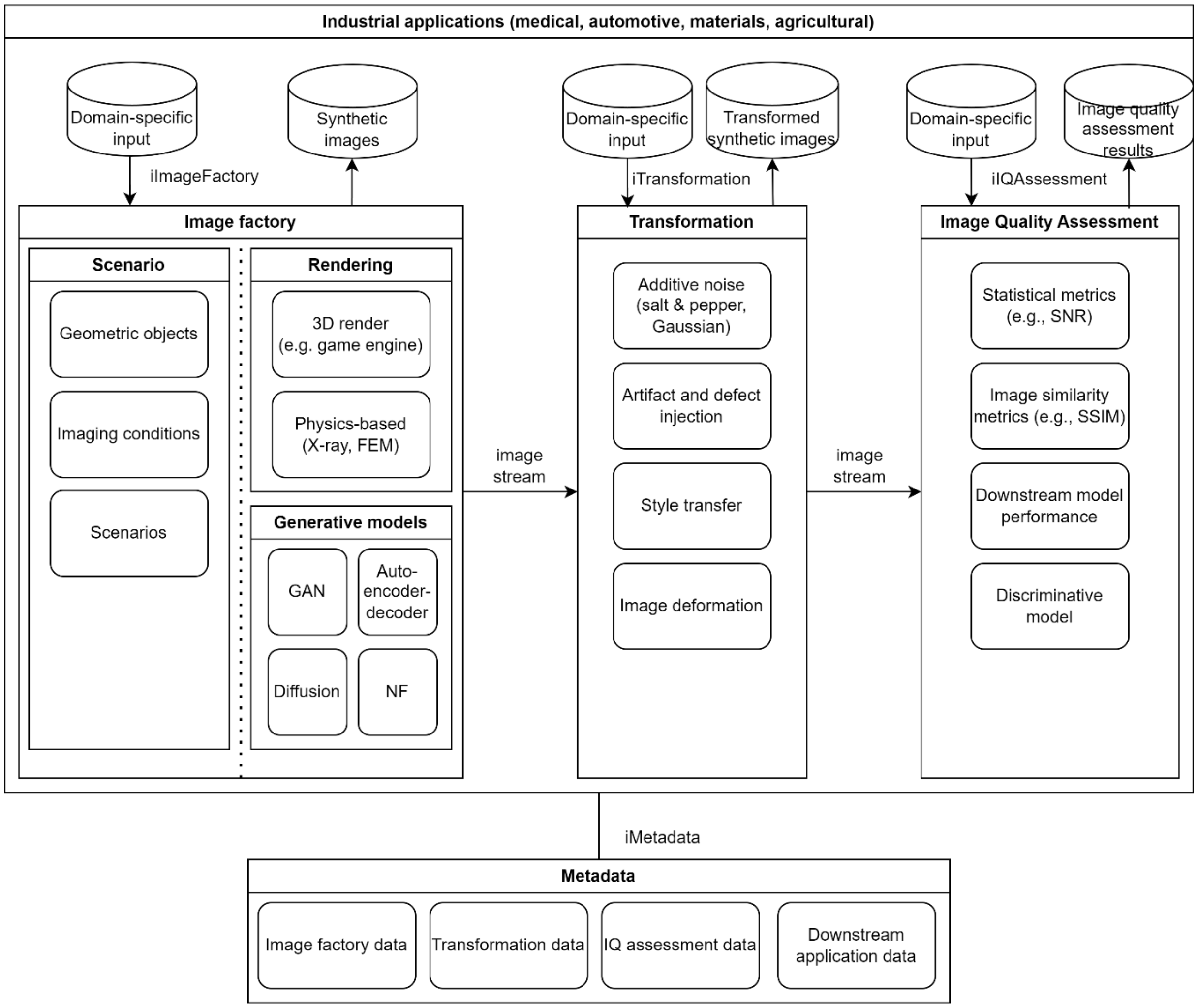

Figure 1: Reference architecture of the TASTI framework. More elements may be added to the blocks later when new insights are gained.

Figure 1 depicts the reference architecture of the TASTI framework. The framework consists of three main blocks: an image factory, a transformation block, and an image quality assessment block. These are the blocks that generate and assess the images. A fourth block in the framework is the metadata block, which processes the information that should be shared across blocks and can be used to reproduce and analyse results. Each block contains several elements that could be reused and customized based on the target application's needs. Note that the collection of elements in this overview is not exhaustive but gives an indication of the elements that can be expected per block. During the project, when the use cases further develop, more elements will be added to the framework. The various blocks and the elements in it are discussed in the next sections.

The industrial application determines the elements that are used within each block and the order in which they are executed. Relevant results are stored at a location specified by the industrial application. This is further described as part of the infrastructure in Section 2.7. Hardware architecture and strategy are also considered to be part of the infrastructure of the framework.

2.2 Image factory block

The image factory contains the elements that are required to generate basic synthetic images. The image factory is based on two main parts: scenario elements and image generation elements. The scenario description contains the boundary conditions under which the images are to be generated. The boundary conditions are used by the elements that generate the images. Typically, domain-specific input is required to form the scenario descriptions or as reference data (e.g., images, shape statistics) for synthetic images to be generated by generative models.

Scenario elements are:

Element	**Description**
Geometric objects	Describes the geometric objects to be considered, such as CAD drawings of vehicles or patient phantoms.
Imaging conditions	Settings specific for the sensing element (such as camera setting), providing the conditions under which an image is generated.
Scenario	Placement of objects in a scenario, contains for example positioning of geometric objects, light sources, imaging angle and boundary conditions for the scenario.

The image generation elements in the image factory create the images based on the input from the scenario elements, either based on physics-based rendering or using more general generative models.

Image generation elements are:

Element	**Description**
3D rendering	Software that renders a 3D scenario, such as game engines
Physics-based simulation	Finite element models, X-ray simulators
GAN	Generative adversarial networks, generative model
Diffusion models	Generative model
Normalizing flow (NF)	Generative model
Auto-encoder-decoders	Generative model

2.3 Transformation block

The transformation block takes the raw synthetic images as input and applies transformations to them. The resulting enhanced images should better fit the requirements of the target applications. Enhancements could be focused on adding realism, for example by adding noise or applying a style transfer, but could also augment images by injecting defects or imperfections.

Elements in the transformation block are:

Element	**Description**
Additive noise	Adding realistic noise to the image. Examples are salt & pepper noise and Gaussian noise.
Artifact and defect injection	Injecting some imperfection in the image, like a glare or disturbance. This requires image context, so it needs to be customized to the target application.
Style transfer	Deep learning technique to mix a simulated image with the style of other images (typically real images).

Image deformation	Deformation of (parts of) images, for example used to augment AI training sets.

2.4 Image quality assessment block

The image quality assessment block contains elements to evaluate the quality of the generated images. Quality does not have a clear general definition, but depends specifically on the target application. Criteria could be realism, sharpness, statistical properties of noise, richness of scenarios and more. In some cases, the output of this block could serve as verification evidence.

Elements in the image quality assessment block:

Element	**Description**
Statistical metrics	Classical metrics to define the quality of an image, such as signal-to-noise ratio (SNR). These metrics could be compared with the same metrics from (a set of) ground truth images to assess image realism.
Image similarity metrics	Compares the generated image with a reference image and quantifies its similarity, using metrics such as structural similarity (SSIM).
Downstream model performance	Assesses the quality of the synthetic images based on the performances of the application for which they are used. This element is different from the previous two in that it cannot be executed on the images alone but should be evaluated separately after the synthetic images have been applied. Evaluation criteria could be accuracy (improvement) of a computer vision application trained by synthetic images or scoring of realism or usefulness by experts.
Discriminative model	Model that tries to discriminate synthetic from real images. Success rate is a measure of realism of the synthetic images.

2.5 Metadata block

The purpose of the metadata block is to store all relevant information to be able to reproduce results and to use the synthetic images downstream. The elements in this block are not producing outputs but are mere data structures in which relevant data should be stored. Each of the blocks connects to the metadata block to store the required data.

Elements in the metadata block:

Element	**Description**
Image factory data	Chain of elements used in the image factory block. Data to include: - Elements with respective input parameters - Version numbers - If applicable: randomization seed - Data useful for AI applications, like object labelling
Transformation data	Chain of elements use in the transformation block. Data to include: - Elements with respective input parameters - Version numbers - If applicable: randomization seed
Image quality assessment data	Chain of elements use in the image quality assessment block. Data to include: - Elements with respective input parameters - Version numbers

	- If applicable: randomization seed
Downstream application data	Any data required to use synthetic images in the target application. An example is labelling that can serve as ground truth for classification or segmentation problems. Such labels normally come from the image factory block.

2.6 Interfaces

The interfaces to be defined are (see Figure 1):

- iImageFactory
- iTransformation
- iIQAssessment
- iMetadata

These interfaces shall define data formats for the input and output images, input parameters, and output parameters (such as randomization seeds and software versions.)

2.7 Infrastructure

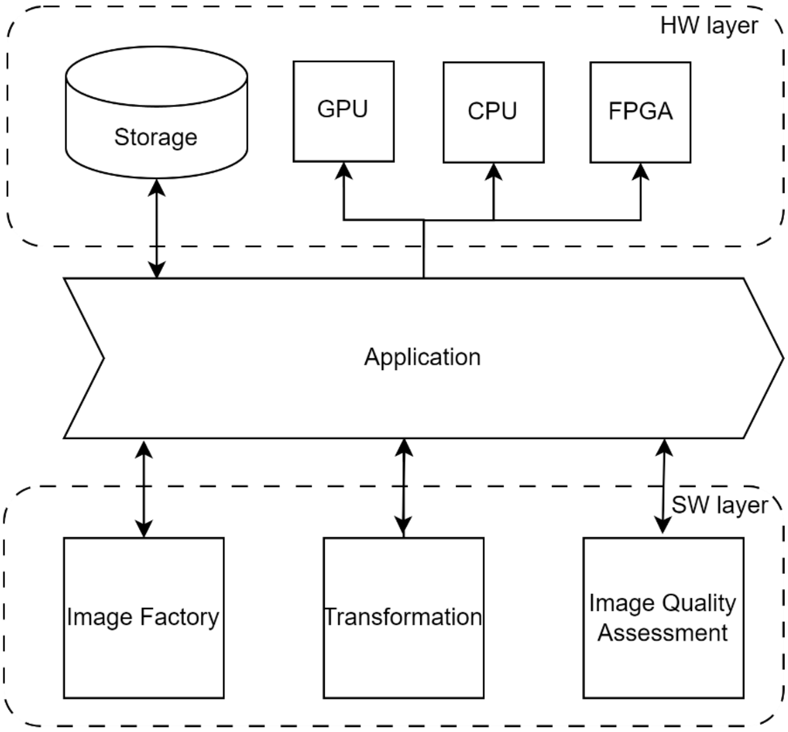

Figure 2: Infrastructure of an application that uses the TASTI framework.

A central application (arrow in Figure 2) controls the steps to be taken to create and assess synthetic images. It provides the input parameters, input data and elements to be used within each block and stores any output. The application must be able to communicate with the blocks and the elements in it, which may be stored remotely or locally. Similarly, the storage must be reachable by the application, but not necessarily by the blocks directly. In this way, it is possible for elements to run remotely and let the central application deal with resulting images and metadata. If the elements cannot run on a remote, dedicated server, the use of containers (such as Docker) should be considered to make it easier to ship elements (e.g., scripts, models) across the platforms on which they run.

The hardware layer shows that the TASTI framework allows for different hardware acceleration options to be used to execute the image synthesis application. Multiple different hardware platforms are investigated within the scope of the project, most notably, mainstream CPUs as well as acceleration using GPUs and dedicated hardware on FPGAs. The decision to choose for either of these platforms depends on the requirements of each application domains. These requirements relate to both functional properties such as latency and throughput, as well as non-functional requirements such as power consumption and system cost. The requirements for each application and the considerations to opt for one of the available hardware platforms is investigated in deliverable D4.1. Where applicable, framework elements should consider compatibility with multiple hardware platforms. Ideally, an abstraction layer would then take care of selecting the algorithm implementation that matches the available hardware.

2.8 Not in scope

The framework focuses on sharable technologies for synthetic image generation. For this reason, the target applications, such as computer vision models, are not part of the framework's scope.

3. Conclusions

Following a bottom-up approach, a concept architecture for the TASTI framework has been presented, which contains the main building blocks of the framework with a suggested set of elements. In the second project year, work will focus on:

- Further development of technical elements through the use cases, based on which the elements in the three main blocks will be specified
- Interfaces (API)
- Infrastructure
- Hardware abstraction

Pilots that will be executed based on each of the use cases will provide valuable feedback to reach a final framework description at the end of the project. The goal is to then demonstrate this framework in each of the use cases and show its exploitation value.