ML classification: Amazon Rekognition or Amazon SageMaker?

Computer vision problems have been tackled using neural networks in recent years, obtaining unprecedented results and continuously raising the bar of accuracy to near-human performances. In this article, the focus is set on image classification in an uncommon context related to a Neosperience customer that provided the opportunity to compare two different approaches: Amazon Rekognition Custom Labels and Amazon SageMaker custom model.
Both approaches have advantages and could find their spot in a given context, but Amazon Rekognition Custom Labels offer an interesting tradeoff between time-to-market and cost.

Alisea

Machine Learning applications are steadily shifting from research domains to industry, opening a wide range of applications from simple object detection to people tracking in dangerous environments.
In this scenario, brands decide to innovate their target market, introducing smart products with features made possible by modern machine learning applications.
Alisea has led the Heating, Ventilation, and Air Conditioning (HVAC) systems sanitization market for almost two decades with over 3000 customers in Italy and abroad. Back in 2005, Alisea was born with a single mission: to offer the market the best HVAC hygienic management service with state-of-the-art innovations, without compromise.

Alisea customers range from corporate offices to malls, supermarkets, hotels, and many more. In recent years, in the US, private houses also adopted HVAC technologies to control air temperature due to better efficiency than standard air conditioning systems.
Every HVAC plant requires periodic inspection, usually once a year, to provide a minimum hygiene level and reduce health risks.
Unfortunately, in crowded environments, this cannot provide good enough guarantees about risk biological management: hazardous agents such as bacterias, dust, and viruses tend to reform faster than expected.
To ensure the safest environment for people, air quality and plant status need to be checked frequently to avoid the spread of respiratory diseases, but it is an expensive procedure.

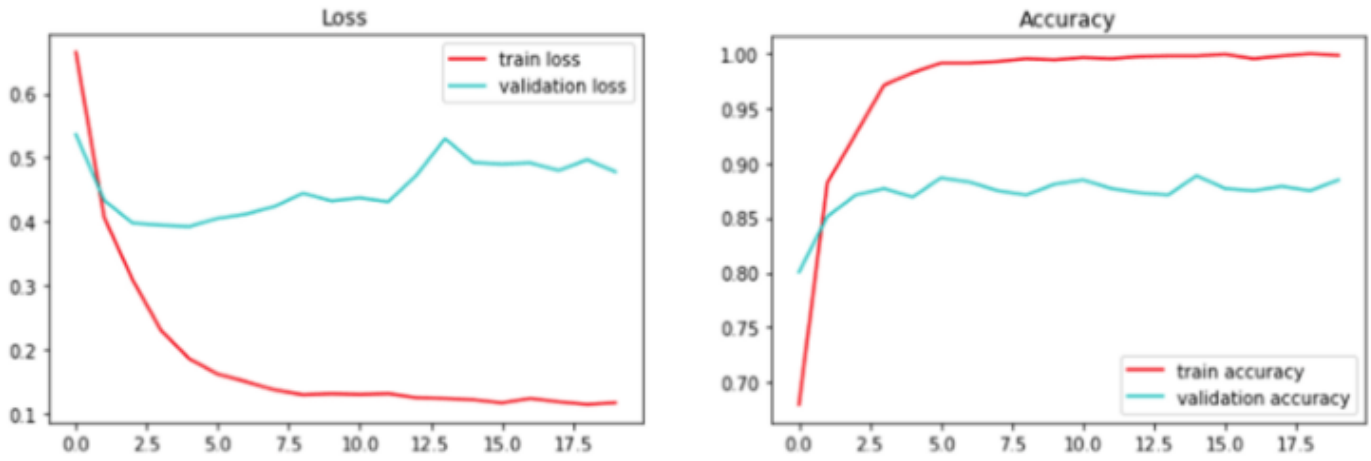
Examples of dirty HVAC systems

Introducing Remotair

A few years ago, Alisea decided to innovate its market, start developing a new smart product, able to detect near real-time duct status, and trigger alarms when air quality and cleanliness fall below a risk threshold.
The system comprises a hardware board with many sensors and two cameras. It leverages at edge and cloud ML models to fulfill its requirements and a dedicated cloud-native architecture to process IoT data. Snapshots of duct status are taken periodically and are processed by a machine learning model to understand whether or not the plant is dangerous to people. This feature has been called **Visual Clean.**

Implementation on Amazon SageMaker

The first implementation of the Visual Clean machine learning model consisted of a neural network, trained using PyTorch and FastAI library with transfer learning and loss function adjustments.
Also, considerable effort was required through the phases of data preparation.
The training phase required a couple of hours and a hundred epochs to reach a final accuracy score of 0.90 ~ 0.92.



Train/validation loss and accuracy

The result has been a deep learning model, served through Amazon SageMaker, and invoked as an HTTP endpoint.

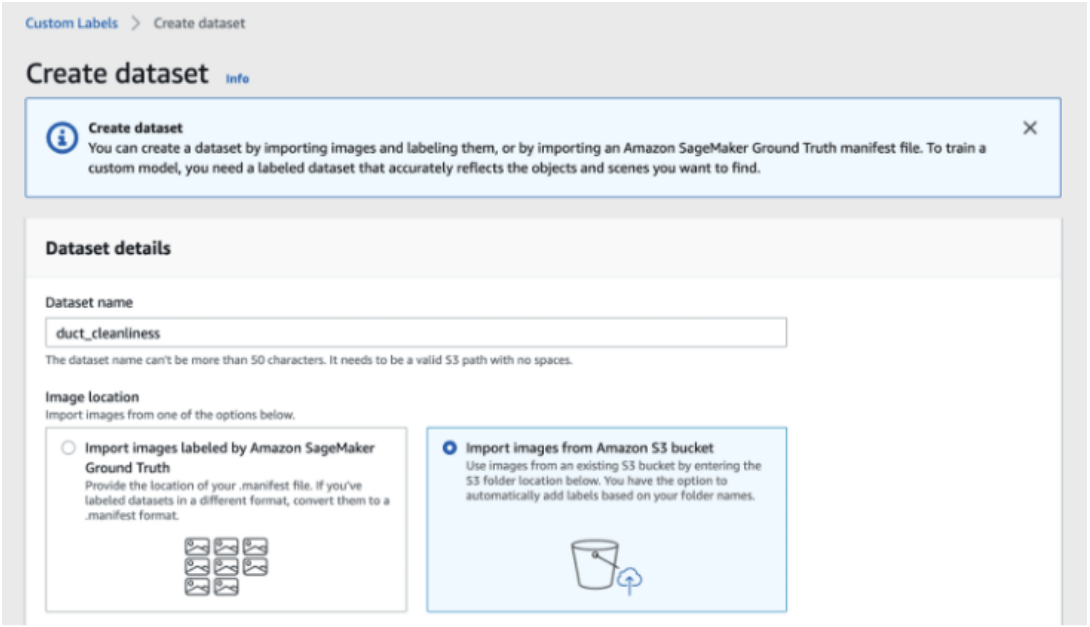
It required having data scientists available to handle the task, prepare the dataset, build the network, and fine-tune the training phase to reach that score.

Amazon Rekognition Custom Labels

As soon as AWS released Rekognition Custom Labels, we decided to compare the results to our Visual Clean implementation to the one produced by Rekognition.

Upload images

The first step to create a dataset is to upload the images to S3 or directly to Amazon Rekognition.



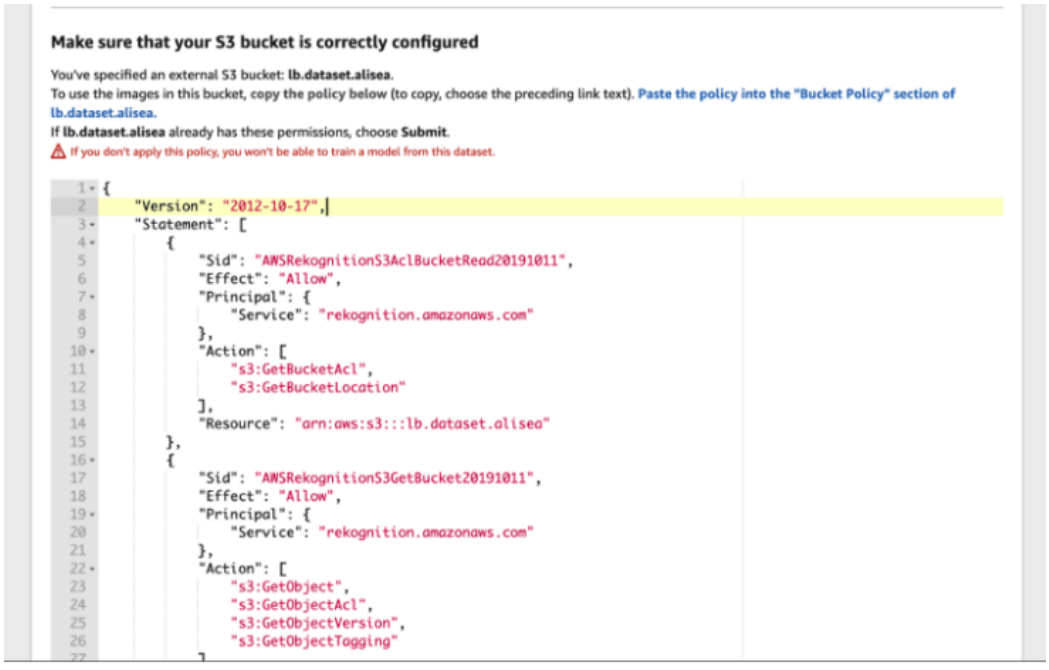
Select the source for your data before any operation.

When data folders respect a clean structure, Amazon Rekognition supports automatic labeling.

Recently, the capability to upload images into the console has been added. However, reproducibility is fundamental in an industrial project, so we decided to upload data into an S3 bucket, divided into a folder for each label to predict it is **dirty** and **clean**. Massive data uploads can be achieved through the command line with

```
aws s3 sync <source_path> s3://<destination_bucket>
```

Once images have been uploaded, the access policy must be set on the bucket to ensure Amazon Rekognition will access data. This can be done by adding to the bucket policies (in bucket properties within S3):



The bucket should be accessible by Rekognition on your behalf to read data for model training.

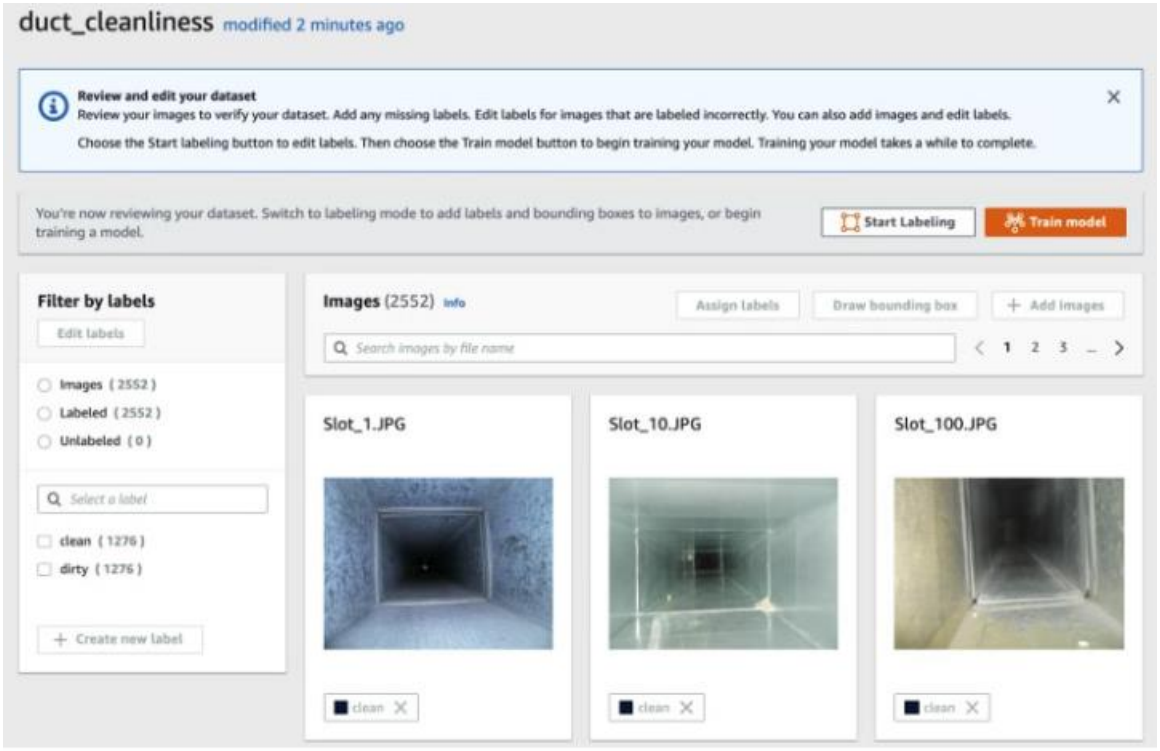
Create a dataset

Once images have been uploaded, a dataset can be created, providing a manifest file describing the available data type.
The manifest consists of JSON lines added to the same file (*note:* it is not a JSON file itself, but a plain text file containing JSON fragments). Each line has the structure:

```
{
    "source-ref": "s3://dataset/images/Slot123_clean.png",
    "visual-clean-dataset-cleanliness": 1,
    "visual-clean-dataset-cleanliness-metadata": {
        "confidence": 1,
        "job-name": "labeling-job/visual-clean-dataset",
        "class-name": "clean",
        "human-annotated": "yes",
        "creation-date": "2020-06-26T17:46:39.176",
        "type": "groundtruth/image-classification"
    }
}
```

To make things easy, images can be labeled through Amazon SageMaker Ground Truth or with a pre-defined folder structure (one for each label).

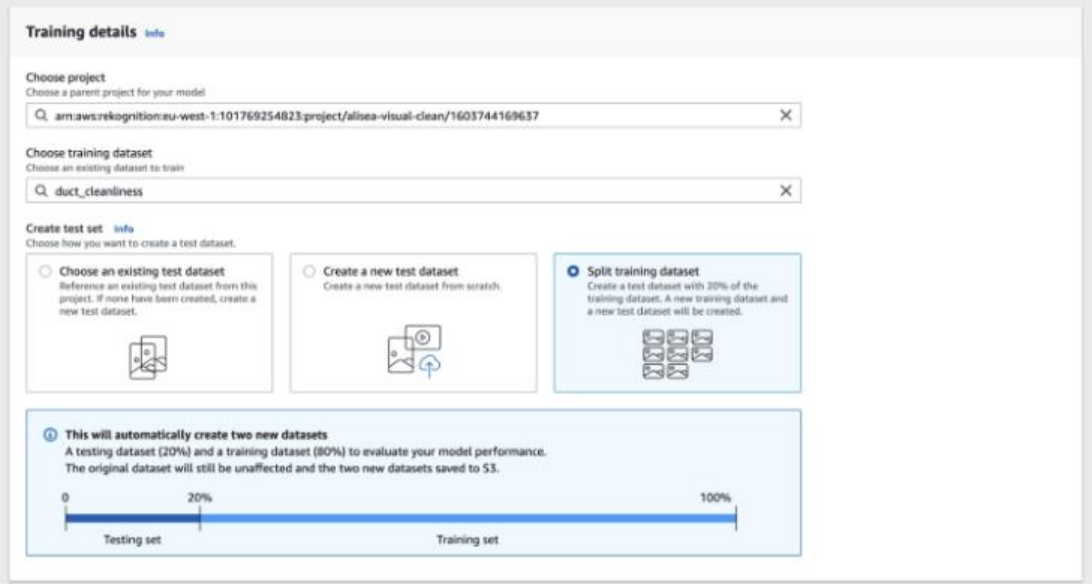
The overall result is shown as follows:



Once data has been loaded, it can be browsed and manually cleaned to remove outliers.

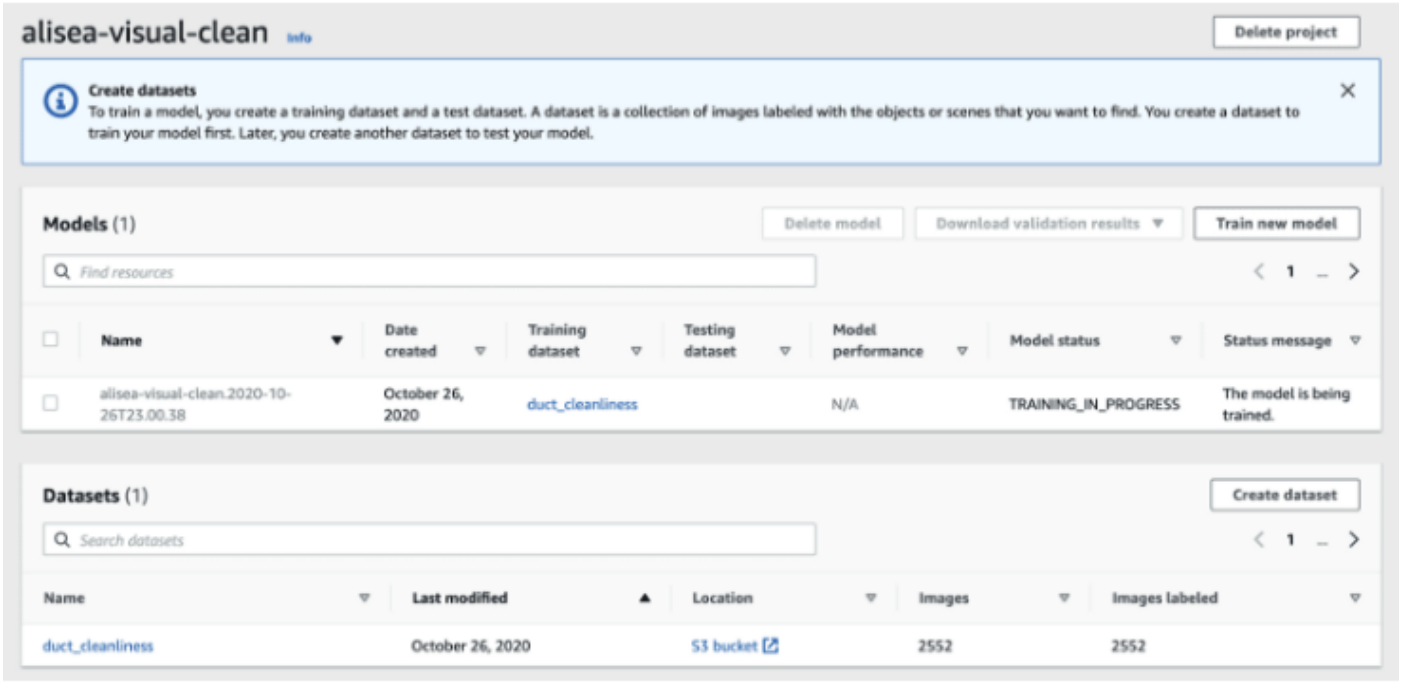
Model training and evaluation

Once a dataset has been built, just hitting the “Train Model” button triggers a training job, asking how to obtain a testing set. The simplest way is to let Rekognition split the existing dataset in 80% training and 20% testing.



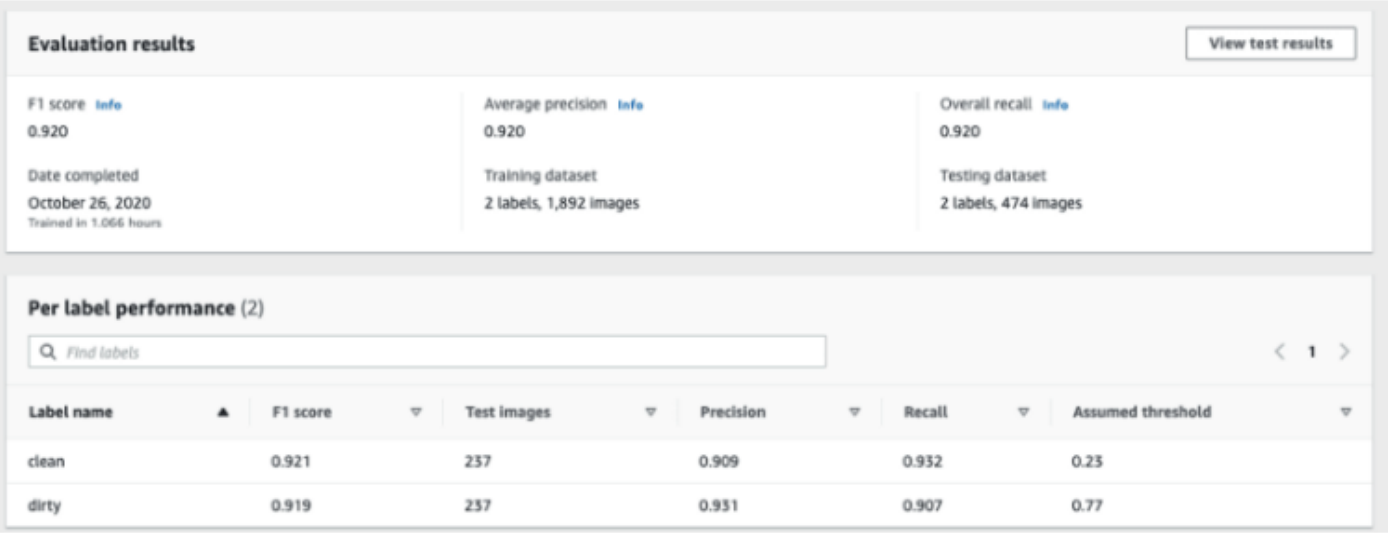
Set up train / validation dataset splits

After that, the training job starts and required a couple of hours to complete



Training job monitoring status

Once the model has been trained, it is reported into the **“Projects”** section of Amazon Rekognition Custom Labels with its evaluation.



Model accuracy is reported for each model version.

Discovering that the Rekognition Custom Label model reached the same accuracy score of **0.92** was mind-blowing. We obtained with both models the same accuracy, but this happened after we've prepared our dataset for training, increasing its quality. The overall training time stating that **1.066** hrs were used a clear insight that Rekognition tries many different models and parameters simultaneously to find the best one.

Once the model has been trained, it can be invoked directly through API, just passing an image to its endpoint:

```
aws rekognition detect-custom-labels \
  --project-version-arn "arn:aws:rekognition:eu-west-
1:XXXXXXXXXX:project/alisea-visual-clean/version/alisea-visual-
clean.2020-10-26T23.00.38/1603749XXXXX0" \
  --image '{"S3Object": {"Bucket": "dataset.alisea","Name":
"clean/Slot_620.JPG"}}' \
  --region eu-west-1
```

Conclusions

Amazon Rekognition offers a viable solution to machine learning model development every time a custom classification model (either binary and multi-class) is required.

It doesn't require a dedicated data scientist and can provide the same outcome when starting a new project, with comparable results. The pricing model is hourly based on a fixed cost of $1 for each training hour and $4 for each inference hour. Better profile cost can be achieved by optimizing the model active hours, thus reducing the working window.

Domain expertise (provided by the Alisea team) is nevertheless required to evaluate, compose, and prepare a suitable dataset, to be uploaded to the model. Amazon SageMaker still remains a viable solution to fine-tune the model for deeper estimation, o whenever labeling is not the focus of our machine learning task. Actually, to obtain the best of both worlds we deployed a mixed solution, leveraging these technologies.

Remotair is an innovative and proprietary technology by **Alisea** developed in partnership with **Neosperience** involving AWS technology.
More details can be found at www.remotair.com.

My name is **Luca Bianchi**. I am **Chief Technology Officer** at **Neosperience** and, the author of **Serverless Design Patterns and Best Practices**. I have built software architectures for production workload at scale on AWS for nearly a decade.

Neosperience Cloud is the one-stop SaaS solution for brands aiming to bring *Empathy in Technology*, leveraging innovation in machine learning to provide support for 1:1 customer experiences.