Using Lightly in Infrastructure Inspection: Swiss Railway Company

*Learn how SBB achieved a **10% better performance of their visual inspection system** while **saving 50% of data related costs**.*

SBB's Goals

SBB's goal is to provide the safest and most reliable transportation service possible. At the same time, it is essential to them to offer affordable prices and run an efficient operation. SBB's wants to improve the following metrics:

- **Few delays**: Trains should be as reliably on time as possible
- **Efficiency**: Railway maintenance should be as efficient as possible
- **Safety**: Consistent monitoring of infrastructure to avoid incidents for passengers and rail construction workers

SBB aims to use Machine Learning to achieve these goals through better infrastructure maintenance.

Figure 1: Manual visual inspection at SBB

Company Profile Swiss Railway Company (SBB)

SBB is the national railway company in Switzerland and is currently the **largest rail and transport company of Switzerland**.

- **Heavily collaborates** with most other transport companies of the country, **to provide fully integrated timetables with cyclic schedules**
- **Ranked first** among national European rail systems in the **European Railway Performance Index**, BCG conducted in 2017, for its quality of service, the intensity of use, and safety
- While many rail operators in continental Europe have emphasised the building of high-speed rail, **SBB has invested in its conventional rail network's reliability and quality of service**
- SBB operates **cargo and freight rail service** through its subsidiary SBB Cargo and has **large real estate holdings** in Switzerland

Using Lightly

SBB has partnered with Lighty to **improve their visual inspection** through better data. Using Lightly's data curation platform, SBB can easily enhance the dataset's quality by reducing data bias. In other words: Lightly helps to achieve higher model performance by improving the data the models are trained on.

Additionally, **SBB wants to save labeling costs** by setting up its data pipeline more efficiently. At SBB, the machine learning team was able to use Lightly to promptly find and filter thousand of images and build workflows to improve the efficiency of the labeling process.

In summary, SBB was able to:

- View their unlabeled datasets at a glance, plot the class distribution, and query for data of interest within Lightly's platform
- Quickly filter their data with Lightly to find the image samples that hurt their model's performance
- Integrate Lightly's data selection and labeling workflows into SBB's in-house labeling system, allowing the team to reduce label costs

Data errors that have been resolved with Lightly include: (1) Nearby duplicates that introduced additional weight on the sample during training, leading to overfitting on specific samples and hurting generalization. (2) Different annotations for nearby duplicates introduced a "blurring" of the decision boundary, which can hurt generalization and accuracy (e.g., in the case of bounding box detection, the localization and classification are affected).

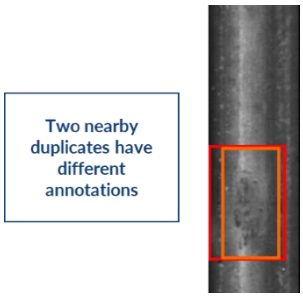

Figure 2: Label mistakes detected by Lightly

Results

Using Lightly to accelerate labeling processes and provide more high-quality data, SBB was able to ***decrease labeling costs by 50%.*** Additionally, SBB significantly increased their machine learning model's accuracy among certain classes by **at least 10%**.

Furthermore, the SBB machine learning team was able **to cut both engineering and labeling time in half** by using Lightly's streamlined data management interface, yielding significant cost savings quickly.

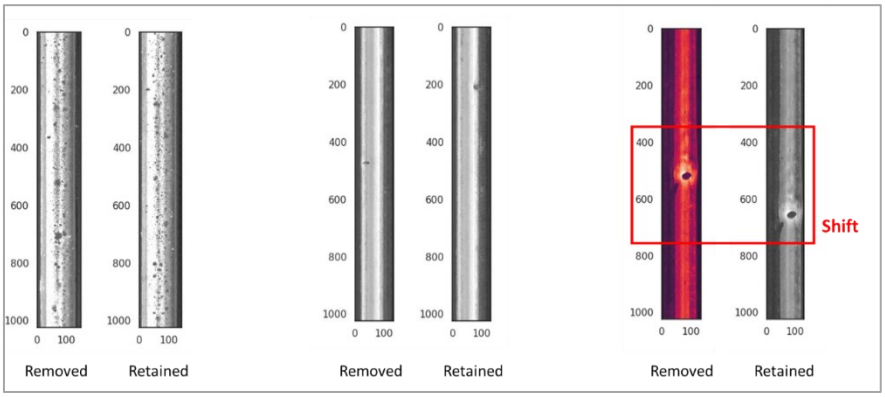

Figure 3: Removed vs. kept data by Lightly

These improvements allowed SBB to provide a more valuable service to their customers:

- More reliable train schedule by avoiding unnecessary railtrack closures causing fewer delays
- Fewer accidents through reduced human involvement in the infrastructure maintenance as well as better maintained infrastructure
- Saving costs in the data labeling process as well as infrastructure maintenance through a better visual inspection system and earlier detection

Those cost savings and quality increase ultimately lead to a safer, more reliable, and more affordable/profitable railway system in the country of Switzerland.

Figure 4: Showing the highest accuracy with 50% of the data

```
df['class'].value_counts() * 100 / df['class'].count()

ok                39.109620
surface_defect    32.996440
welding           10.016776
unclear            9.587135
squat              4.181841
joint              4.050902
wheel_slip         0.057285
Name: class, dtype: float64

df_filtered['class'].value_counts() * 100 / df_filtered['class'].count()

ok                37.905568
surface_defect    30.520289
welding           12.430504
unclear           10.048959
joint              4.555639
squat              4.480956
wheel_slip         0.058086
Name: class, dtype: float64
```

Class balance before filtering

Class balance after filtering (50%)

Figure 5: Class rebalancing by Lightly