Fictiva Case Study Analysis

By: Andrew Rover

Introduction

Fictiva, a fictitious data annotation company has been chosen by a European automotive company to provide ground truth annotations to train their machine learning models used on L3 automation features.

Unfortunately, after a few months of delivering annotations for this customer, it has been found that this annotation project has been suffering from poor performance. Current quality reports indicate that this project has been reaching a 90% quality SLA, 5% percent below the 95% threshold set by the customer.

This case study analysis aims to pinpoint the factors contributing to this underperformance and in turn prioritize features in our product roadmap to address them, fulfilling the mandate of getting the project into high customer satisfaction and high profitability.

What we know

The first step in approaching this case study is defining what we know about the annotation process and the problems it may be facing. The annotation process is made up of the two following phases:

1. **Production Phase:** Where Production Associates annotate all required objects.
2. **Quality Phase:** Where Quality Associates review the work of the Production Associates, ensuring it meets the customer's 95% quality level.

The reported problems facing the annotation process, indicated both formally in our quality reports and via internal feedback from the Fictiva operations team, are summarized below.

- **Problem #1:** We are 5% below the customer's SLA quality level (90% vs 95%).
- **Problem #2:** Some QAs are spending too much time correcting Production work.
- **Problem #3:** Some QAs are consistently delivering below the 95% quality SLA.
- **Problem #4:** The production step has been highlighted for operational inefficiency.
- **Problem #5:** Not all platform features are being used properly.
- **Problem #6:** We don't have strong feature analytics to help on performance tracking.

Lastly, the LoaA (Life of an Associate) report details the total number of hours each Production Associate and Quality Associate, dedicated to each step in the annotation process:

LoaA Report				
Step	**Bucket**	**Hours Spent**	**Bucket Description**	**Cost ($)**
Quality Step	Image navigation	1998	Time spent zooming/panning the image to search for objects and reposition the tool for annotation	29970
Quality Step	Object list navigation	534	Time spent going through the object list and selecting objects	8010
Quality Step	3D annotations	475	Time spent annotating 3D cuboids	7125
Quality Step	Point cloud navigation	411	Time spent zooming/panning/orbiting the 3D point cloud to search for objects and reposition the tool for annotation	6165
Quality Step	2D annotations	269	Time spent annotating 2D rectangles	4035
Quality Step	Object attributes	232	Time spent choosing the right attributes from the large taxonomy requirements	3480
Quality Step	Quality feedback	109	Time spent adding feedback to the quality issues encountered on the task	1635
Production Step	Image navigation	3728	Time spent zooming/panning the image to search for objects and reposition the tool for annotation	37280
Production Step	3D annotations	2914	Time spent annotating 3D cuboids	29140
Production Step	2D annotations	1519	Time spent annotating 2D rectangles	15190
Production Step	Point cloud navigation	1422	Time spent zooming/panning/orbiting the 3D point cloud to search for objects and reposition the tool for annotation	14220
Production Step	Object list navigation	1329	Time spent going through the object list and selecting objects	13290
Production Step	Object attributes	1172	Time spent choosing the right attributes from the large taxonomy requirements	11720
Production Step	Quality feedback	0	Time spent adding feedback to the quality issues encountered on the task	0

As the report lists that a Production Associate's hourly rate is $10/hour and that a Quality Associate's hourly rate is $15/hour, we have additionally calculated the total cost of each step in the annotation process.

Combining our insights drawn from a qualitative analysis of these reported symptoms and a quantitative analysis of the LoaA report will allow us to diagnose the underlying problems facing this annotation project. From there, we can identify what features we believe will fix these problems and how we will go about prioritizing them in our product roadmap.

Problems

Problem 1: Lack of metrics on feature performance makes it difficult to diagnose problems

Not having strong metrics on how features on the Fictiva platform are being used by Fictiva Associates is preventing Fictiva from identifying the source of operational inefficiencies and making informed product decisions.

For example, there have been reports that Fictiva Associates have been improperly using features on the Fictiva platform, making them slower. Pinpointing which features are being misused using our current data however proves to be difficult.

Additional Time PAs Spend on Each Task Relative to QAs	
Bucket	**Hours Spent**
3D annotations	PAs spend 6.13x more time
2D annotations	PAs spend 5.65x more time
Object attributes	PAs spend 5.05x more time
Point cloud navigation	PAs spend 3.46x more time
Object list navigation	PAs spend 2.49x more time
Image navigation	PAs spend 1.87x more time

Currently the tasks of 2D and 3D annotation have the largest time discrepancy between PAs and QAs.

On one hand, this can be used as evidence that 2D and 3D annotation features are the features most likely to be misused by Production Associates. Production Associates are misusing these features, causing them to take significantly more time to complete these tasks compared to QAs.

On the other hand it can also be argued that this time discrepancy indicates the work produced by the Production Associates during the 2D and 3D annotation step rarely needs to be corrected by the Quality Associates. As such, Quality Associates need to spend less time on that task relative to Production Associates. If this is the case, it is unlikely that PAs are misusing Fictiva platform features during this task, considering the quality of the work they are outputting for this task.

Without any strong feature metrics it is difficult to have confidence in identifying how the use of platform features correlates to operational efficiency and the likelihood of an Associate generating ground truth. This in turn makes prioritizing value delivering features to Associates especially difficult. Investing in enabling feature metrics should thus be prioritized.

Problem 2: Quality Associates may not be spending enough time providing quality feedback

Quality Associates only allocate 2.71% of their time to providing quality feedback.

LoaA Report (Sorted By Hours)		
Step	Bucket	Hours Spent
Production Step	Image navigation	3728
Production Step	3D annotations	2914
Quality Step	Image navigation	1998
Production Step	2D annotations	1519
Production Step	Point cloud navigation	1422
Production Step	Object list navigation	1329
Production Step	Object attributes	1172
Quality Step	Object list navigation	534
Quality Step	3D annotations	475
Quality Step	Point cloud navigation	411
Quality Step	2D annotations	269
Quality Step	Object attributes	232
Quality Step	Quality feedback	109
Production Step	Quality feedback	0

In theory, quality feedback submitted to the Production Associates should allow them to learn from their mistakes and improve their likelihood of submitting correct annotations. Allocating such a small percentage of their time to the single task that most leverages their increased level of expertise, is questionable.

With there being a significant amount of rejected ground truths constantly needing to be corrected, it can be seen how QAs deprioritize their feedback step, which takes longer to realize its value to the annotation process.

Consequently, ensuring QAs can allocate time to submitting quality feedback may have long-term positive effects on the quality level achieved by the annotation project.

Problem 3: Image Navigation is extremely costly

The first observation made in the LoaA (Life of an Associate) report is that Image Navigation by Production Associates (PAs) and Quality Associate (QAs) are the #1 and #2 most costly steps in the entire annotation process, costing Fictiva $37280 and $29970 respectively, leaving significant cost reduction opportunities.

Image Navigation tasks additionally ranked first and third respectively in terms of total hours spent on a task by either Production Associates or Quality Associates, taking 3728 hours and 1998 hours respectively.

LoaA Report (Sorted By Cost)				
Step	Bucket	Hours Spent	Bucket Description	Cost ($)
Production Step	Image navigation	3728	Time spent zooming/panning the image to search for objects and reposition the tool for annotation	37280
Quality Step	Image navigation	1998	Time spent zooming/panning the image to search for objects and reposition the tool for annotation	29970
Production Step	3D annotations	2914	Time spent annotating 3D cuboids	29140
Production Step	2D annotations	1519	Time spent annotating 2D rectangles	15190
Production Step	Point cloud navigation	1422	Time spent zooming/panning/orbiting the 3D point cloud to search for objects and reposition the tool for annotation	14220
Production Step	Object list navigation	1329	Time spent going through the object list and selecting objects	13290
Production Step	Object attributes	1172	Time spent choosing the right attributes from the large taxonomy requirements	11720
Quality Step	Object list navigation	534	Time spent going through the object list and selecting objects	8010
Quality Step	3D annotations	475	Time spent annotating 3D cuboids	7125
Quality Step	Point cloud navigation	411	Time spent zooming/panning/orbiting the 3D point cloud to search for objects and reposition the tool for annotation	6165
Quality Step	2D annotations	269	Time spent annotating 2D rectangles	4035
Quality Step	Object attributes	232	Time spent choosing the right attributes from the large taxonomy requirements	3480
Quality Step	Quality feedback	109	Time spent adding feedback to the quality issues encountered on the task	1635
Production Step	Quality feedback	0	Time spent adding feedback to the quality issues encountered on the task	0

Finding a solution that can reduce the number of hours Associate's spend on Image Navigation would increase the project's outputted quality / Associate hour worked, improving both customer satisfaction and project profitability.

Prioritized Features

In this section we will propose features that aim to solve the problems identified above. We will then decide how we will weigh the priority of developing these features against existing technical items from our engineering team.

We will use the ICE scoring method to determine the value and thus prioritization of each proposed feature and technical item, combining them into one complete prioritized roadmap.

The ICE scoring method is a prioritization technique used to evaluate and rank features based on three key factors: Impact, Confidence, and Effort.

1. **Impact:** Assesses the magnitude of the effect the proposed feature will have.
2. **Confidence:** Reflects the level of certainty in the Impact and Effort estimate.
3. **Effort:** Represents the resources and time required to implement the proposed feature.

Each factor may be assigned a numerical score ranging from 1-10. The formula for calculating the ICE score is: Impact x Confidence / Effort. The higher the ICE score, the higher the priority of the proposed feature. This helps prioritize by considering both a feature's potential impact and its feasibility of implementation.

We will first start by giving ICE scores to the technical items the R&D team have requested to be prioritized, and then our own features.

Feature #1: ML-generated pre-annotations

ML-generated pre-annotations is a feature that can significantly expedite the overall annotation process with little added risk.

Pre-annotations are flexible as they can still be reviewed and corrected by our Fictiva Associates if they are incorrect, adding minimal downside risk.

In the best case scenario pre-annotated data requires almost no human intervention, consequently significantly reducing the time and cost of an annotation. In the worst case scenario, Fictiva Associates simply do not use the pre-annotations.

Additionally, we can have a reasonable degree of confidence that our internal ML models are quite performant due to the wealth of data we have from various customers that our models can be trained on.

The above factors result in an **ICE Score of 12**.

$$Impact = 10,\ Confidence = 6,\ Effort = 5$$
$$ICE\ Score = (10 \times 6) \div 5 = 12$$

Feature #2: Automatic sensor calibration correction

Automatic sensor calibration correction will enable Fictiva Associates to spend less time performing 2D annotation, given the projection of 3D objects into 2D being more accurate.

We can thus have high confidence in the fact that this would lower the time Fictiva Associates spend annotating 2D objects. However, as indicated from the R&D team, this feature would only enable an “improvement” in camera calibration. Accuracy projection from 3D to 2D will be improved, but corrections will still be needed.

Additionally, 2D annotation is one of the least expensive steps in the annotation process, resulting in this feature having a limited potential impact on increased Associate productivity.

Annotation Task Cost Distribution (High to Low)		
Bucket	**Cost ($)**	**Percentage of Total Cost**
Image navigation	67250	37.1%
3D annotations	36265	20.01%
Object list navigation	21300	11.75%
Point cloud navigation	20385	11.25%
2D annotations	19225	10.61%
Object attributes	15200	8.39%
Quality feedback	1635	0.9%

The above factors result in an **ICE Score of 7.2**.

$Impact = 4, Confidence = 9, Effort = 5$
$ICE\ Score = (4 \times 9) \div 5 = 7.2$

Feature #3: Investing in performance upgrades

Investing in performance upgrades mitigates the risk of technical regressions affecting the user experience, an especial risk for 3D applications

Image navigation taking both Production Associates and Quality Associates such a significant amount of time, compared to relatively more complicated tasks such as 3D annotation and navigation, may be indicative of performance regressions in image navigation.

In which case, investing R&D into performance upgrades for Image navigation can offer considerable productivity gains to the annotation project.

However, there are multiple other factors that can cause Image navigation to take Associates so much time. For one, as indicated by the ordering of the annotation steps, image navigation is the first step in the annotation process, so it makes sense for unseen data to take longer to navigate. Additionally, there are 3 streams of camera image data compared to the single stream of LiDAR data.

It is consequently difficult to say with any level of certainty the benefits that investing in performance upgrades will provide to the project as a whole, or even Image navigation specifically.

The above factors result in an **ICE Score of 4.8**.

$$Impact = 8,\ Confidence = 3,\ Effort = 5$$
$$ICE\ Score = (8 \times 3) \div 5 = 4.8$$

Feature #4: Stronger Performance Analytics

A common theme in this analysis is the difficulty in having confidence in identifying the specific underlying problems facing this project and how they can be solved. Having strong analytics on the performance of Fictiva features would solve this.

Strong performance analytics will allow the product team to better understand how the use of a feature by a Fictiva Associate correlates to the likelihood of generating ground truth. In turn, this will enable us to better understand what features deliver value to the Associates and why.

This is a technical item that will not have any immediate impact, however it is the building block to truly understanding how the issues facing this project can be fixed.

The above factors result in an **ICE Score of 11.25**.

$$Impact = 5,\ Confidence = 9,\ Effort = 4$$
$$ICE\ Score = (5 \times 9) \div 4 = 11.25$$

Feature #5: AI Copilot for Production and Quality Associates

One feature we can build out of the analytics generated from tracking the performance of features is an AI Copilot for Production & Quality Associates. This Copilot would be able to provide Associates with constant feedback throughout the annotation process.

For example, if a Production Associate is measured to be taking significantly longer in annotating a given 2D image, the AI Copilot can indicate to them that perhaps the tool they are using to annotate the object with, perhaps a sliding rectangle instead of a rectangle, is incorrect, and they should try something else.

This Copilot would additionally be able to help or even completely automate the process of QAs providing quality feedback to PAs

With strong feature analytics on feature use, the insights that can be recommended to them throughout the annotation process is vast.

The above factors result in an **ICE Score of 4**.

$Impact = 8, Confidence = 3, Effort = 6$
$ICE\ Score = (8 \times 3) \div 6 = 4$

Final Roadmap