



Herausforderungen und Lösungsvorschläge
in der Auswahl relevanter Daten für
supervised learning Trainings

1.10.2024 | LUVIS AI GMBH
Nicolas Hoyer

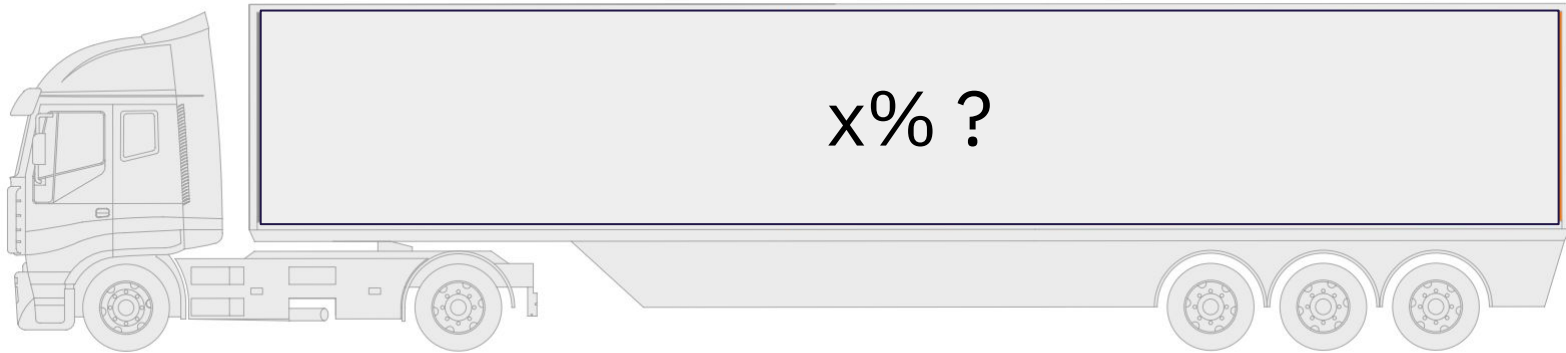
Overview

1. Some background:
What's our product / what are we actually doing?
2. Our approach to minimizing AI Operations Gap and what is it in our case
or: how we monitor training space..

Everything I'm showing is something we've learnt and may be applicable only in our specific usecase.

Problem

Die ungenutzte Ladekapazität beträgt in Deutschland laut Kraftfahrt-Bundesamt ...?

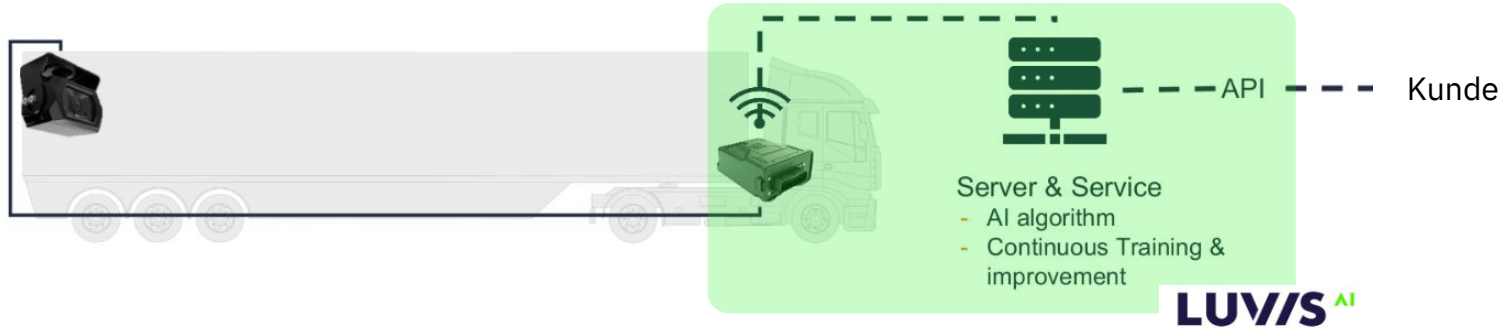


Unnötige jährliche Belastungen: 11 Mrd. Euro / 12 Mio. Tonnen CO₂

Wir möchten dies verbessern.

Lösung

Bessere Auslastung der LKW durch Kamera-Monitoring

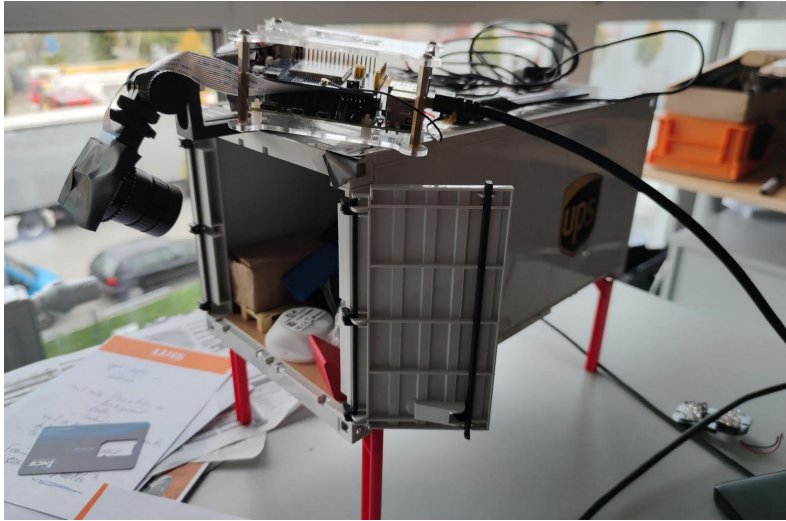


Wir bieten ein **SaaS Produkt** an, bestehend aus:

- embedded code zur optimierten Auslösung eines Bilds
- KI-Analyse der Laderaumfotos
- API

Wir integrieren die Software auf allen nötigen Systemen und definieren zusammen mit dem Kunden die Schnittstellen.

From desk demonstrator to many installed trailers



2019



Beispiel-Visualisierung unserer Daten (für LUIS ähnlich)

Smart Scan

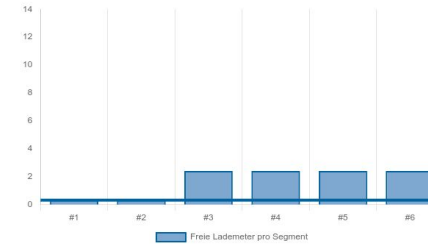


Smart Load



Smart Capacity

KI Analyse



Freie Ladefläche [m]: 0.3

Freie Fläche [m²]: 4.2

Freie Fläche [%]: 13



zwei Themen mit unseren, sehr subjektiven lessons learnt

AI Operations Gap

“

Data is the new oil"

Clive Humby

and it also doesn't come for
free...



ordinary training process



There's a lot to read here on how to optimize, but..

ordinary training process

but ..



we've had max 18k images / day → 100% labeling → 250k€/month

Inherently important question: Which images do need to be labeled?
What images will be helpful for training?

- Where was the AI result insufficient?
- Which image is showing something “new”?

Our answer for a long time: **manual selection**, we've built a nice toolset around it, that we're no longer using.

What we're doing to identify important images

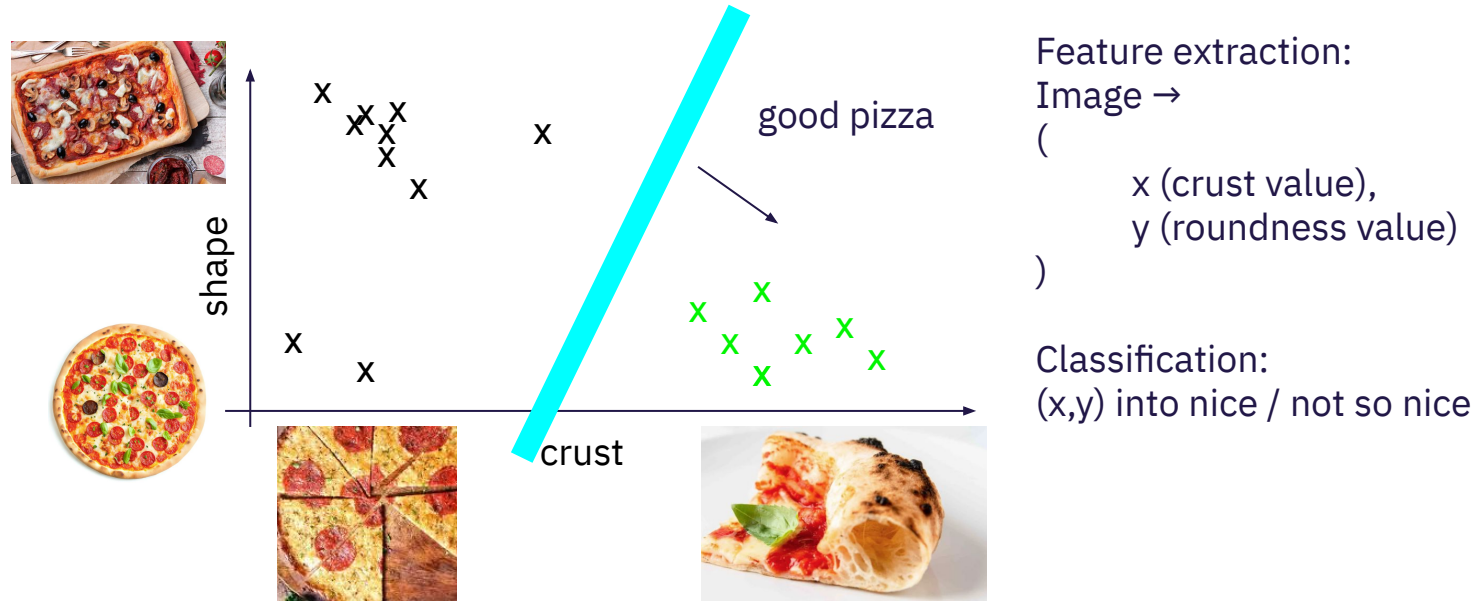
Let's have a look at pizza / how to distinguish between good and bad pizza?



What we're doing to identify important images

manage the training space

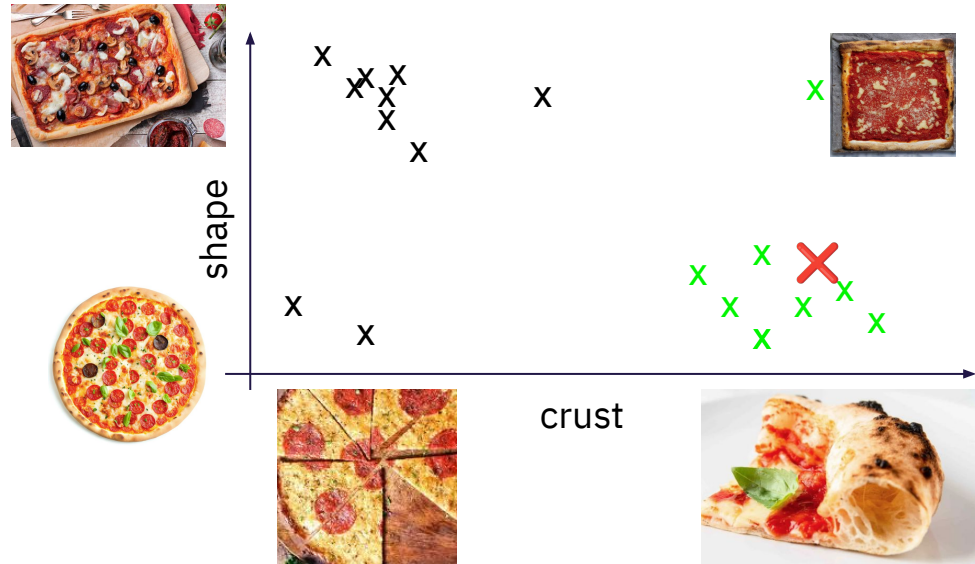
Let's have a look at a good pizza / bad pizza classifier with a 2D feature space



What we're doing to identify important images

manage the training space

Some examples



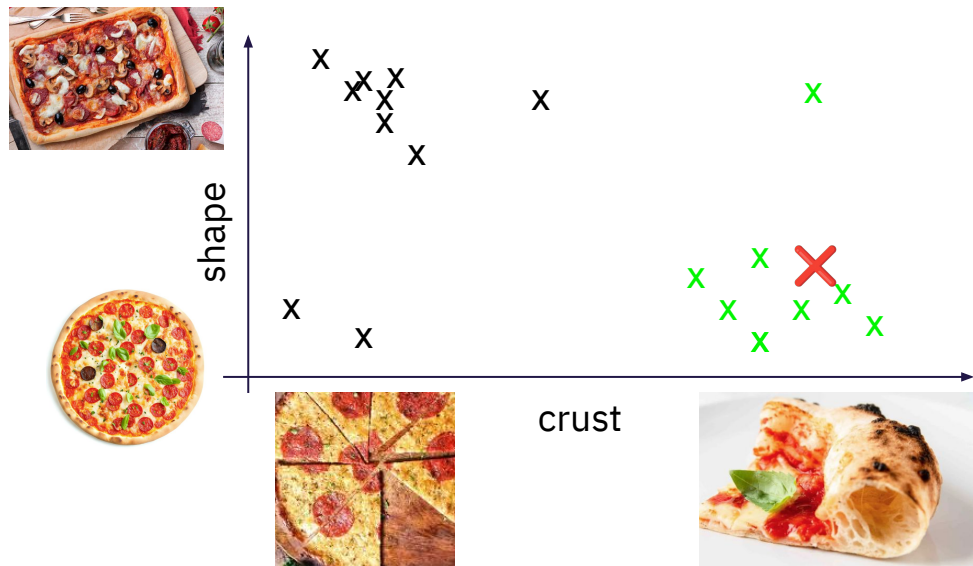
Where do we put this?



What we're doing to identify important images

manage the training space

Let's have a look at a good pizza / bad pizza classifier with a 2D feature space



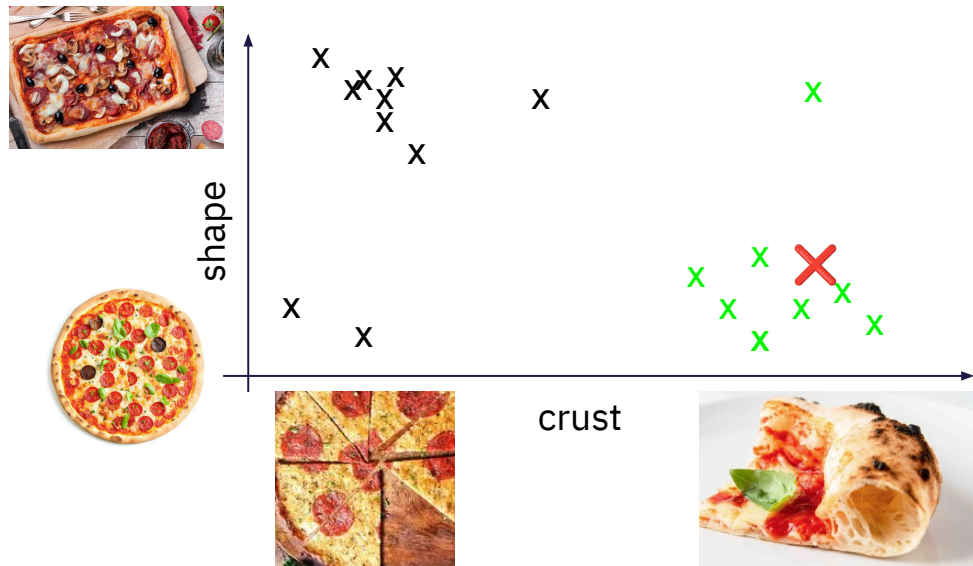
What we're doing

- remove outlier
- if possible: not through the feature space
- get the feature vector of every new data set
- check if it's "needed" in the training space
- we use pg_vector for this, it's nice and pretty fast
- and Triton REST queries can be configured in a way to give back not only the result but also interim layer data

What we're doing to identify important images

manage the training space

Let's have a look at a good pizza / bad pizza classifier with a 2D feature space



What feature space are we aiming for?

- well balanced (no clusters)?
- evenly distributed?
- in every dimension?

Has anyone used this or other metrics to identify / predict AI failing?

We've successfully used this to integrate new data into our model training and it's performing better than randomly picked data.

We've also done an analysis confidence / probabilities, but..

Thank you!

NICOLAS HOYER

+ 49 1637533638

hoyer@luvis.ai