Capstone Project

Steel Defect Detection

Daniela Stürmer, Fabio Teichmann, Michael Funke

Heart For Steel

Intro to Challenge

Steel Defect Detection for Severstal	
The Stakeholder	- **Severstal** is a major producer of steel (11.3 mil tonnes in 2020) - Leads the charge in efficient steel mining and production
The Material	- **Flat iron sheets** - Many different process steps (heating, rolling, drying, cutting) during production
The Challenge	- **Localise** and **classify defects** (*segmentation*) from images - Measure success through **Sørensen-Dice coefficient**

Economic Relevance

Short Term: Quality Management

- Monitors production quality
- Detects and localize defects
- Avoid complaints from customers

Long Term: Process Improvement

- Gather statistical data
- Detect defect patterns
- Increase product processing quality
- Proactive maintenance

The Team

Fabio Teichmann

Business Administration & International Management

Daniela Stürmer

Economic Engineering

Michael Funke

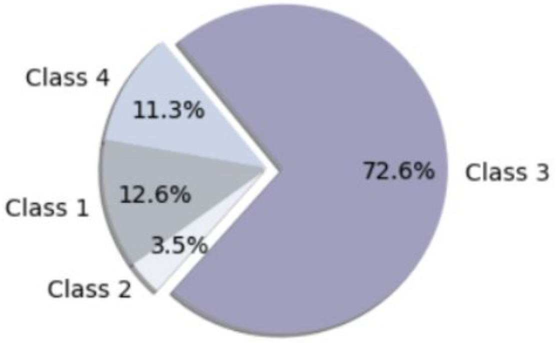
Computational Engineering

About the Data Set

Number of Images

Labelled Images: ~6,700

Overall Defect Class Distribution

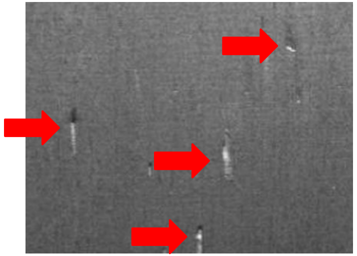


- 4 defect classes
- Data unbalanced ➜ Class 3 over-represented
- Class 2 very small

Defect Classes to Detect

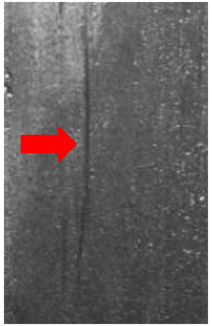
1 Small Holes

- Pimple-like pattern
- Tiny but deep
- Widely distributed



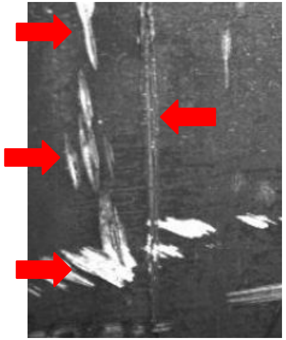
2 Light Scratches

- Not very deep
- Mostly vertically
- often solitary



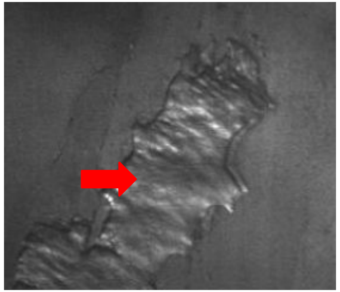
3 Deep Scratches

- Deep marks
- Chaotic patterns
- Widely distributed

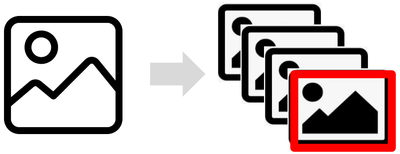


4 Craters

- Flat craters
- Good to discover
- Often very big



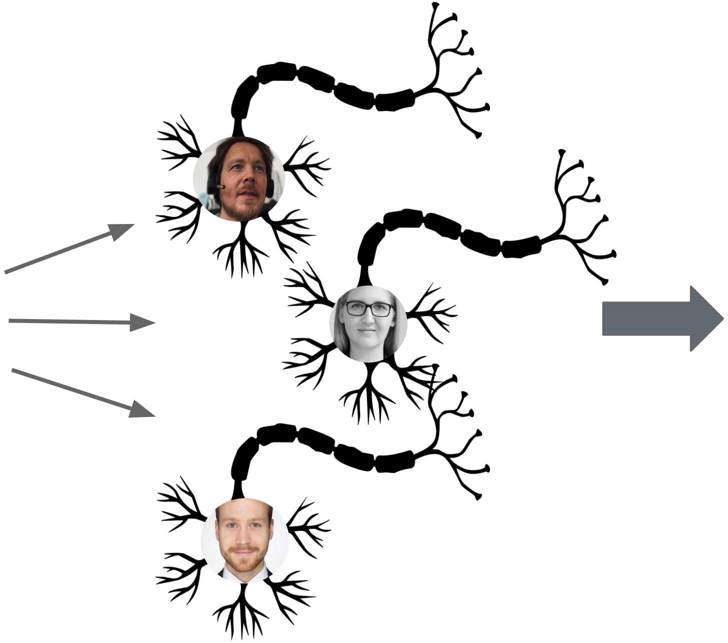
Capstone Approach

Idea	Classification	Localisation & Classification
		
Setup approach to classify each image and build pipeline to localise defect on each image	**Output: defect classes**	**Output: exact localisation of defect and its respective class**

“3 Neurons” - Training with 240 Images

Baseline Model - Train the Team

20 images per class per Person = 80 images per Person

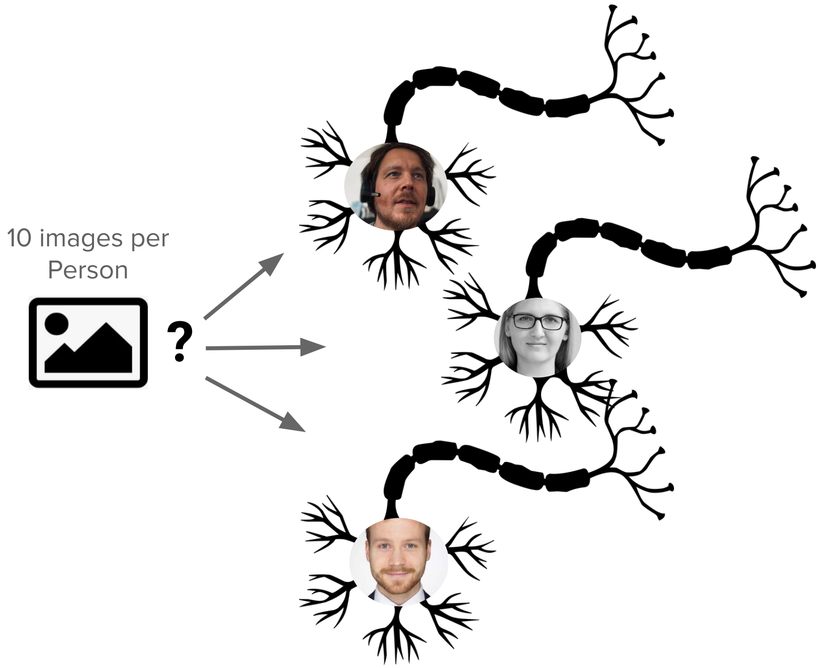


Findings

Class	average “black content” in %	Count
1	19,42	60
2	47,92	60
3	13,83	60
4	3,30	60
Sum	21,12	240

“3 Neurons” – Prediction for 30 Test Images

Baseline Model - Test the Team



Result

Accuracy for Classification

57 %

True label /Class	*Prediction*	count
1	False	4
	True	4
2	False	1
	True	0
3	False	8
	True	12
4	False	0
	True	1
Total		**30**

Overview of Classification Results

Nr.	Model	Input Data	Model Accuracy (all classes)
1	k-Nearest-Neighbour	Initial images	**81%**
2	k-Nearest-Neighbour	HOG Features	**79%**
3	Human “3 Neurons”	20 images per defect class	**57%**

- Excellent match for **class 3** defects (imbalanced dataset)
- Identification of **class 4** defects challenging
- **Class 1** and **class 2** achieve moderate results
- **Augmented images** improve scores of underrepresented classes, but **increase incorrect assignment** to class 3
- Proportion of black pixels per image may inhibit classification for some classes

Examples of Common Misclassifications

Misclassification as Class 3

Image ID: 19d892dd9.jpg | True ClassId: 4 | Predicted ClassId: 3

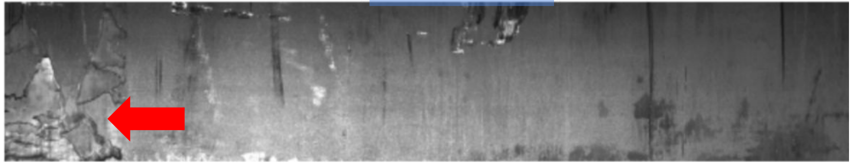
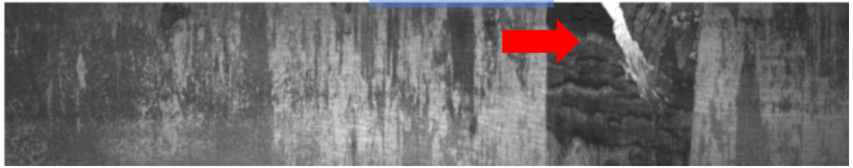


Image ID: ce02c322e.jpg | True ClassId: 4 | Predicted ClassId: 3



Misclassification of Class 3

Image ID: 4cc6b09fd.jpg | True ClassId: 3 | Predicted ClassId: 2

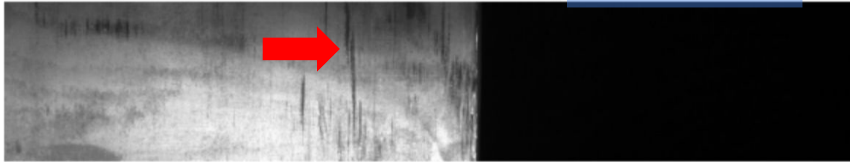
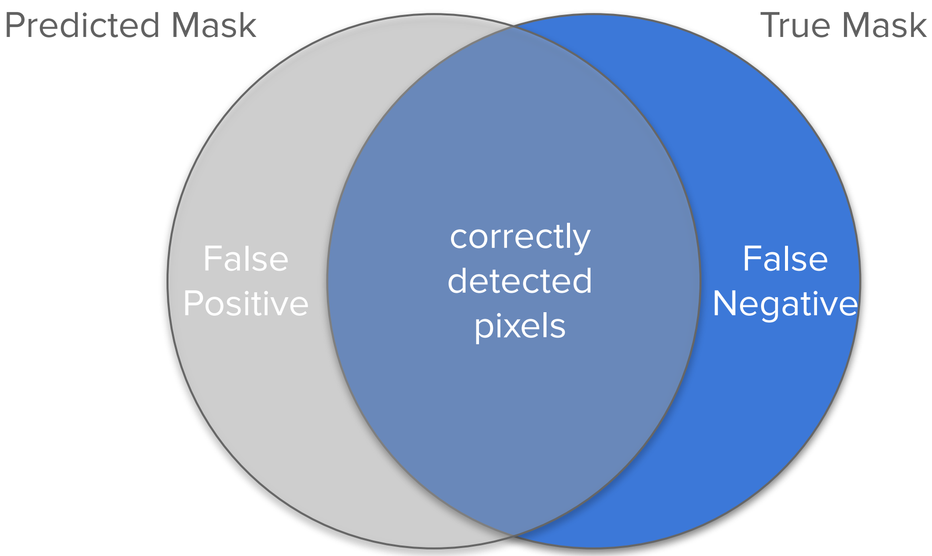


Image ID: 5bec4663f.jpg | True ClassId: 3 | Predicted ClassId: 1

Class 3 images do show **defects** that **look identical to other defect classes**
Quality of image labels for class 3 may not be sufficient

⇒ classification is complicated for the respective models

Segmentation Metric – Dice Coefficient

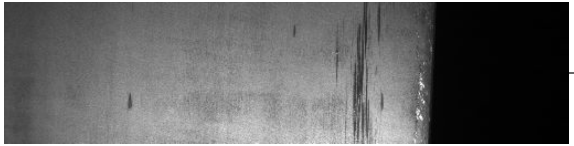


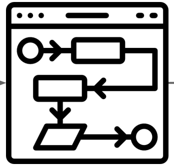
The Sørensen-Dice index can be seen as the **percentage of overlap** between two masks

Segmentation Model - Description

Input	Model	Output

Steel Image



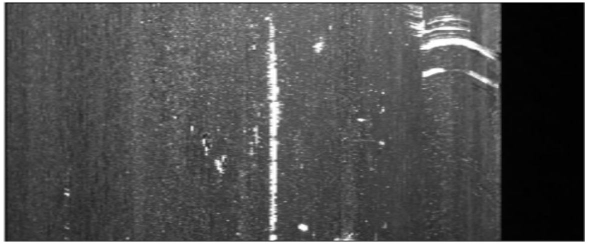
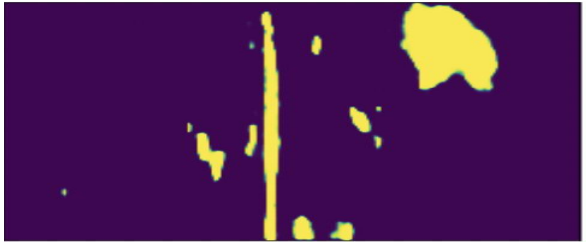


Mask image that shows defect areas

Specifications:

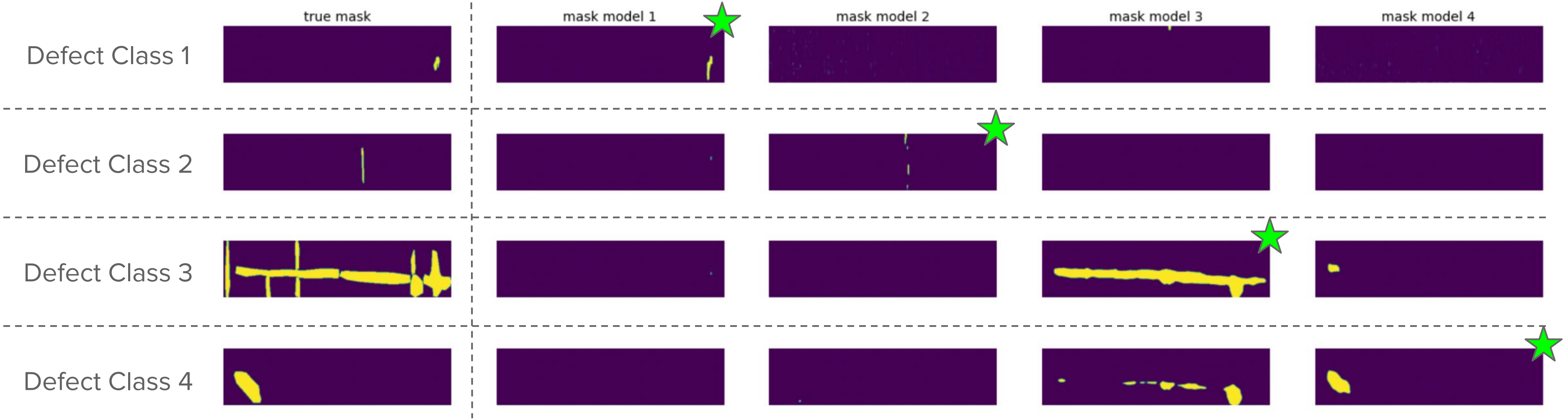
Architecture - Unet
Pre-trained - EfficientNetB5

Example of a Segmentation Mask

True Image	True Mask	Mask from Model
		

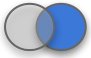
- True masks (used for model training) mark in many cases a much broader area than the visible defect
  ⇒ model training may be more difficult
- Masks generated by the model are **able to localize relevant pattern** that belong to a defect

Example Results from our Model

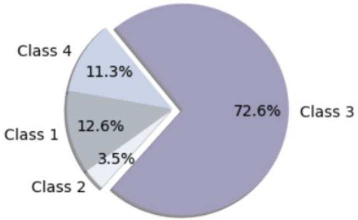


Segmentation by our model generates reasonable results across the 4 defect classes.

Prediction Performance Localisation

Defect Class	Prediction Performance Sørensen-Dice 
1	57%
2	61%
3	52%
4	72%

- Very good results regarding defect localisation
  ➜ Prediction masks with even sharper contour than provided “true masks”
- weighted classification accuracy through predicted masks: **76.1 %**



Our Achievements for Severstal

1	**Defect Classification:** - Classification models: 81% - Via segmentation models: 76%
2	**Defect Localisation:** - Segmentation models produce accurate defect masks - Defect masks of our models are partly more accurate than ground truth

→ We identified various approaches to **further improve** the overall **results**

Recommended Advancements

No.	Measure(s)	Potential Benefit(s)
1	Add downstream classifier to segmentation model	Model accuracy ~6 PP
2	Train last layers of pre-trained model with our steel images	Model accuracy ~4 PP
3	Individualise image augmentations per defect class	Model accuracy ~3 PP

Thank You for your Attention!