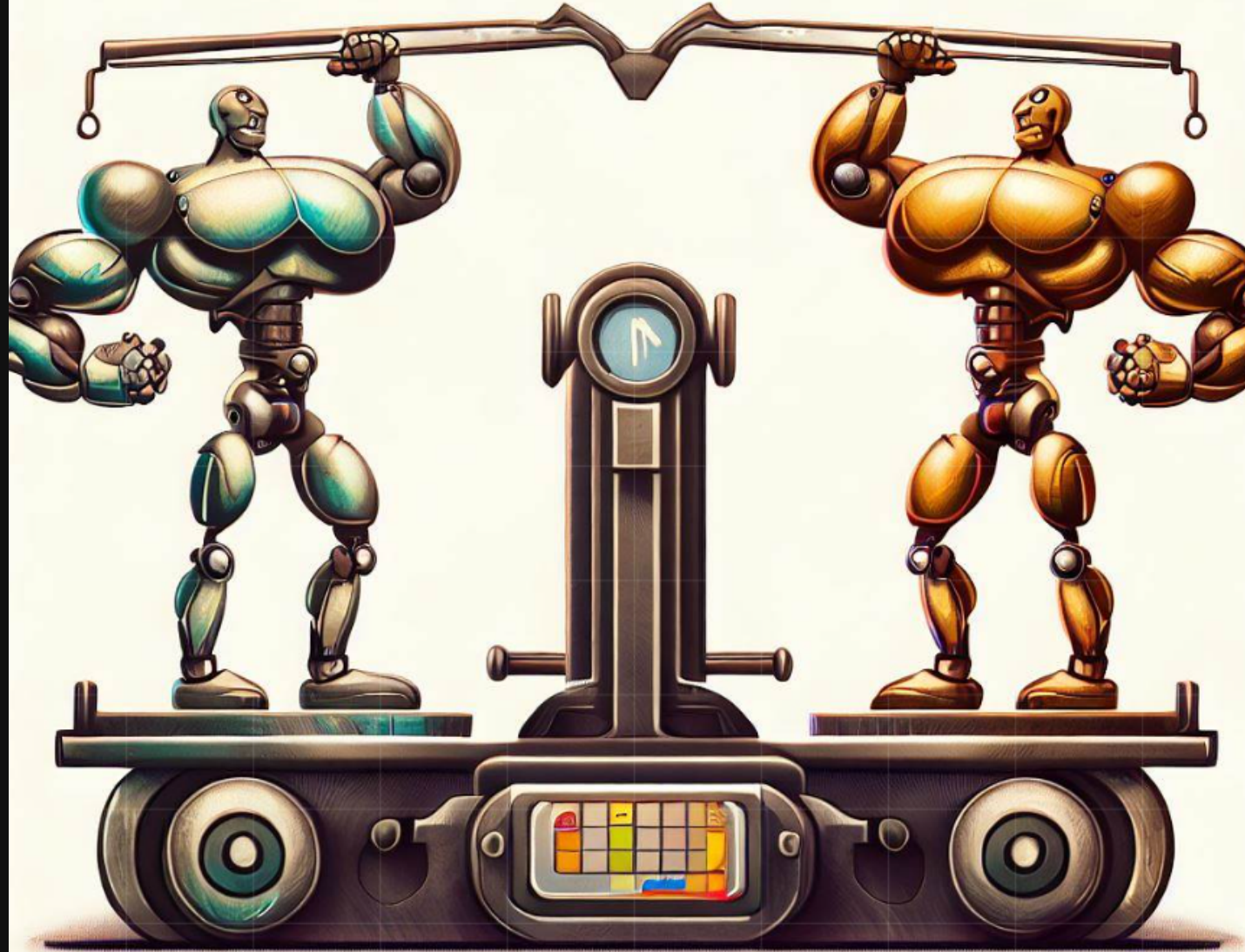


Evaluieren



https://www.youtube.com/embed/nwya_rjfd0T4?enablejsapi=1



Evaluieren

<https://www.youtube.com/embed/cRRMPGqJKUs?enablejsapi=1>

Wie gut und stabil ist mein Model?

- Performance Messen
- Test auf unbekannten Daten
- Robust gegen Änderungen
- Vergleich mit Mensch?

Performance Messen

<https://www.youtube.com/embed/w7QXn-VJq2A?enablejsapi=1>

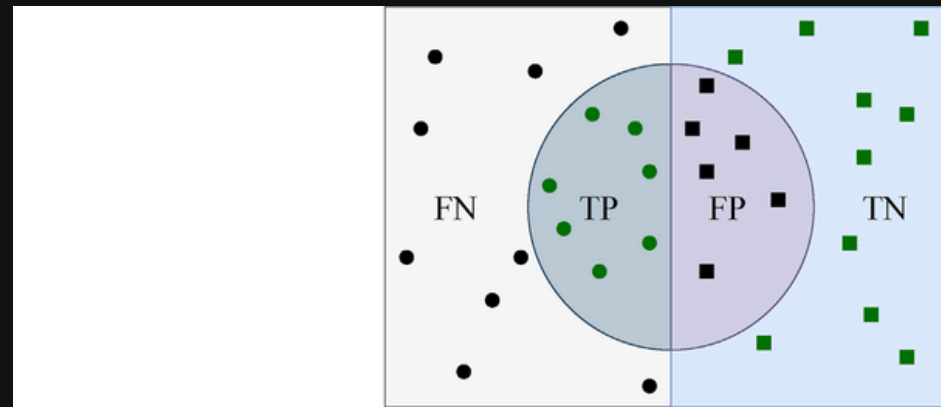
- Kosten vs Accuracy
 - Kosten: Training, nicht interpretierbar
 - Accuracy: verständliches Mass der Leistung



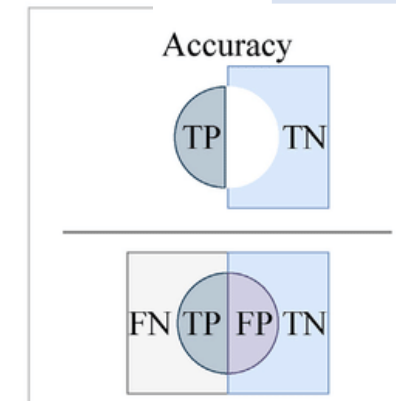
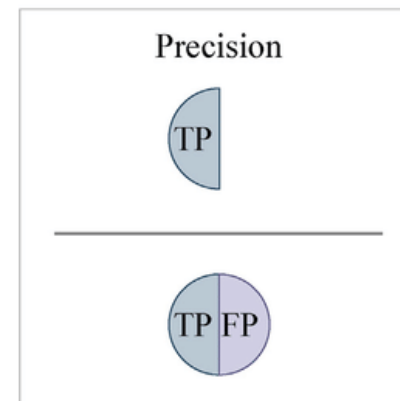
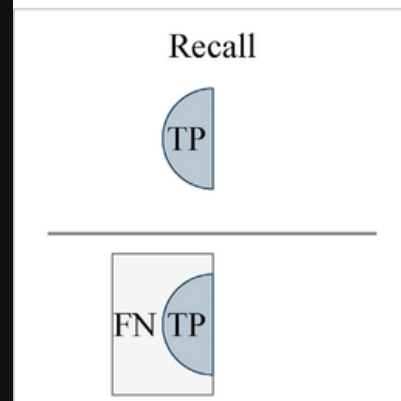
Performance Messen

<https://www.youtube.com/embed/ADa7mhbSZ68?enablejsapi=1>

- Kosten vs Accuracy
 - Kosten: Training, nicht interpretierbar
 - Accuracy: verständliches Mass der Leistung
- Klassifizierung:
 - Accuracy
 - Precision & Recall



		Prediction	
		0	1
True Label	0	48 true negatives	8 false positives
	1	4 false negatives	37 true positives



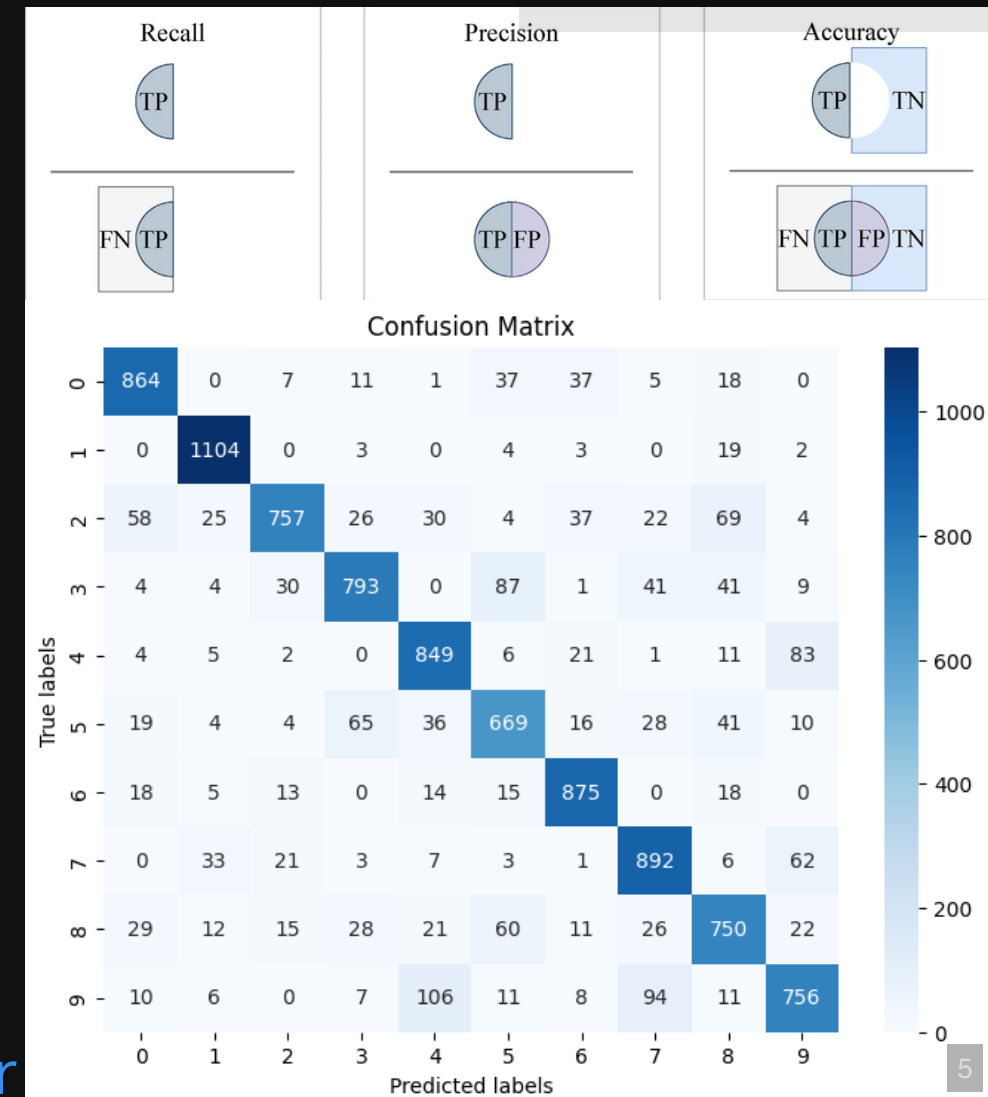
Performance Messen

<https://www.youtube.com/embed/-ym27X11wEU?enablejsapi=1>

- Kosten vs Accuracy
 - Kosten: Training, nicht interpretierbar
 - Accuracy: verständliches Mass der Leistung
- Klassifizierung:
 - Accuracy (Binäres label & Balancierte Daten)
 - Precision & Recall (Reinheit & Vollständigkeit)
 - Confusion Matrix (Interpretation)
 - F1 Score (vereint Precision & Recall)

$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision}$$

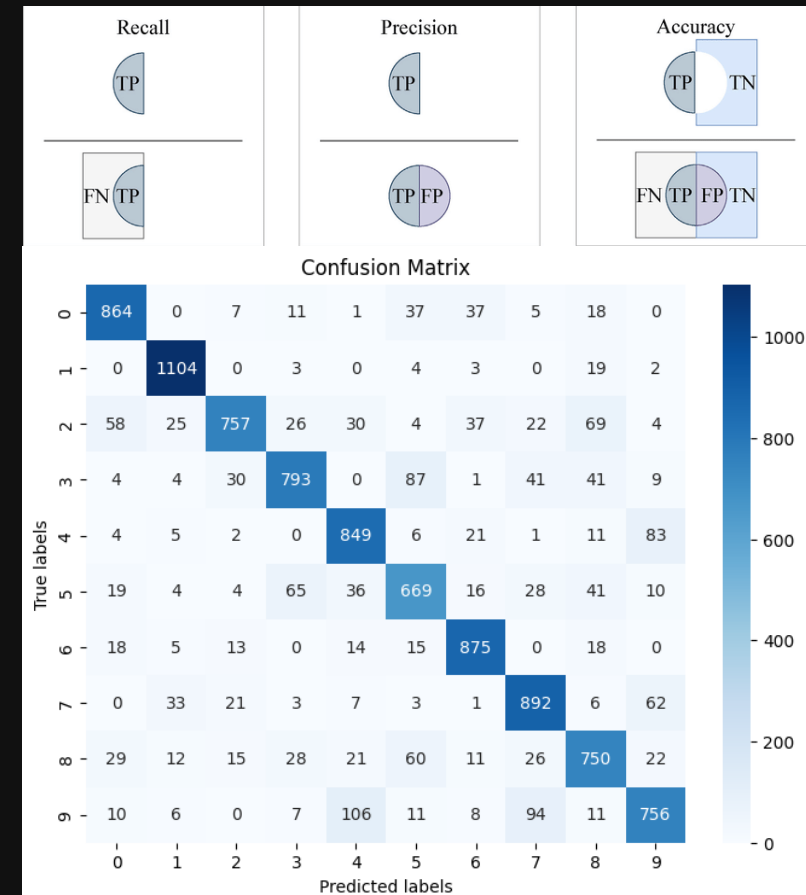
Mehr dazu [hier](#) und [hier](#)



Performance Messen

- Kosten vs Accuracy
 - Kosten: Training, nicht interpretierbar
 - Accuracy: verständliches Mass der Leistung
- Klassifizierung:
 - Accuracy (Binäres label & Balancierte Daten)
 - Precision & Recall (Reinheit & Vollständigkeit)
 - Confusion Matrix (Interpretation)
 - F1 Score (vereint Precision & Recall)

```
1 from sklearn.metrics import confusion_matrix, f1_score
2 conf_mat = confusion_matrix(all_labels, all_preds)
3 f1 = f1_score(all_labels, all_preds, average='macro')
```



$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision}$$

Unbekannte Daten

<https://www.youtube.com/embed/CC7BW1TnhLU?enablejsapi=1>

- Train-Test Split
- Training mit Train, Evaluation mit Test
- Verfolge bei Training:
für train & test je loss & accuracy



Robust

<https://www.youtube.com/embed/cp4jpExRssw?enablejsapi=1>

- Training an verschiedenen Punkten starten

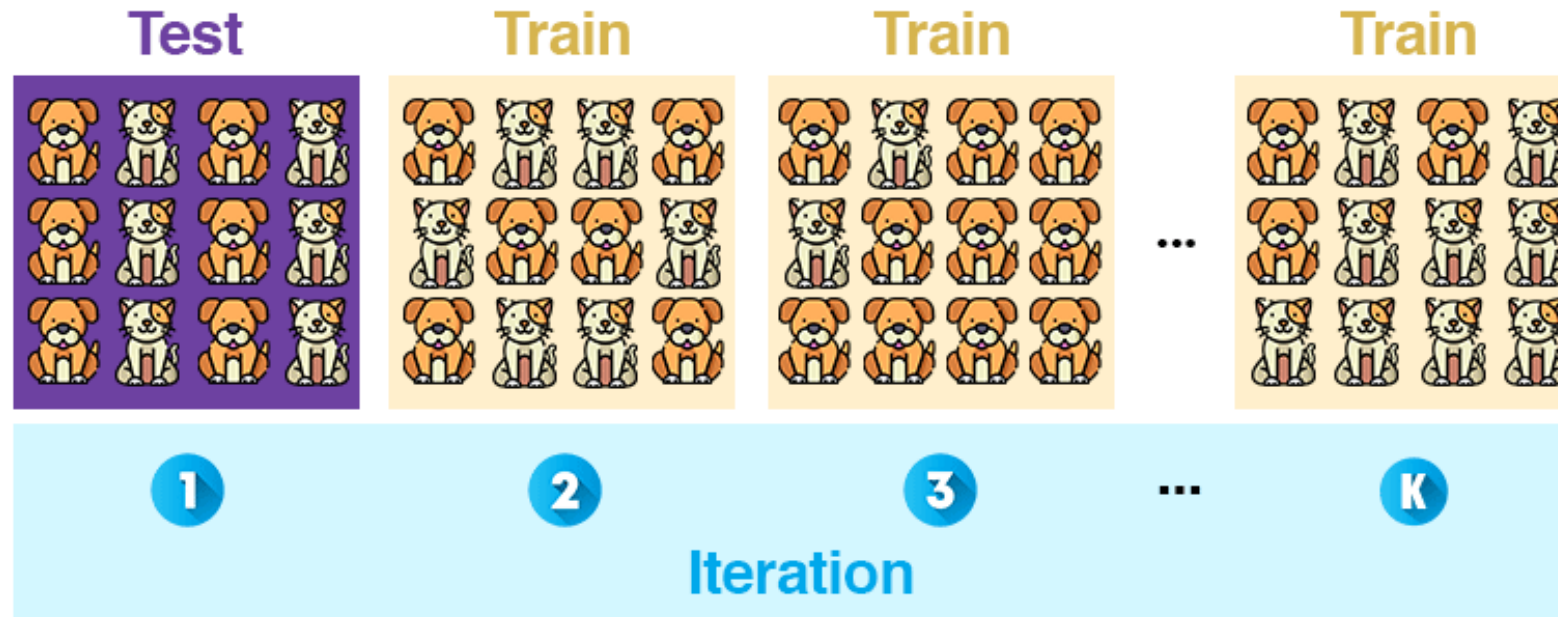


Robust

<https://www.youtube.com/embed/qXYHMEoowAk?enablejsapi=1>

- Training an verschiedenen Punkten starten
- k-fold cross-validation ([mehr dazu](#))

k-Fold Cross Validation

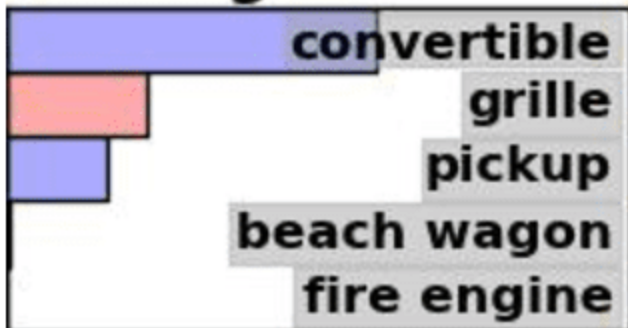


Vergleich mit Mensch

ImageNet Klassifikation



grille



mushroom



cherry



Vergleich mit Mensch

<https://www.youtube.com/embed/mzfiBtD5II4?enablejsapi=1>

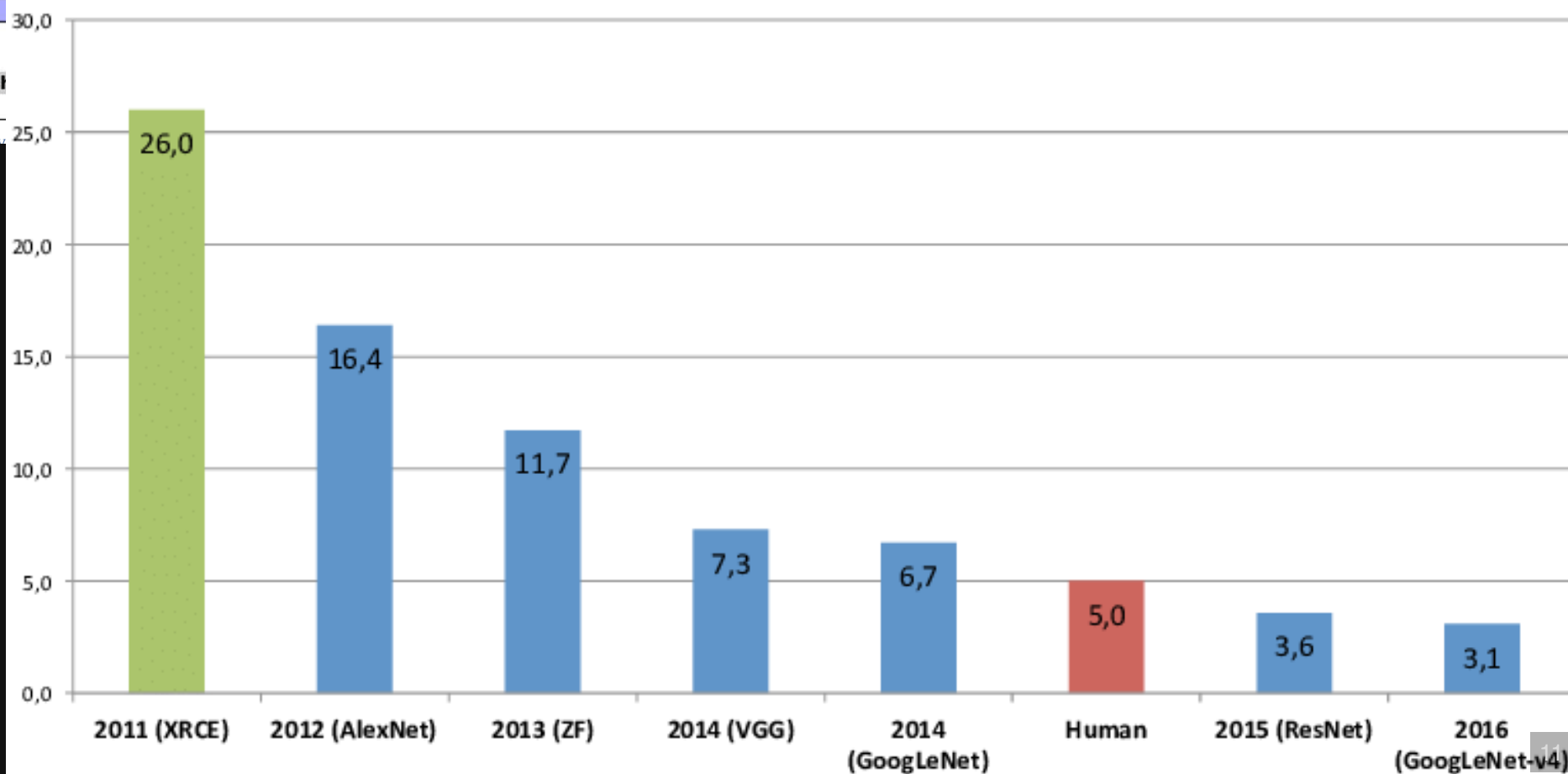
ImageNet Klassifikation



grille mushroom

convertible	agaric
grille	mushroom
pickup	jelly fungus
beach wagon	gill fungus
fire engine	dead-man's-fingers

ImageNet Classification Error (Top 5)



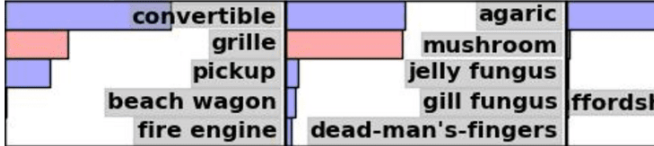
Vergleich mit Mensch

ImageNet Klassifikation

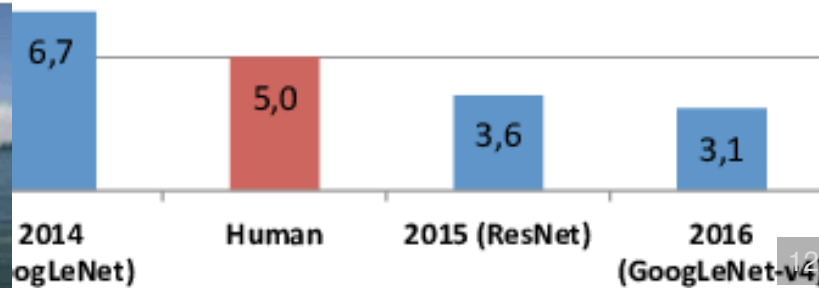
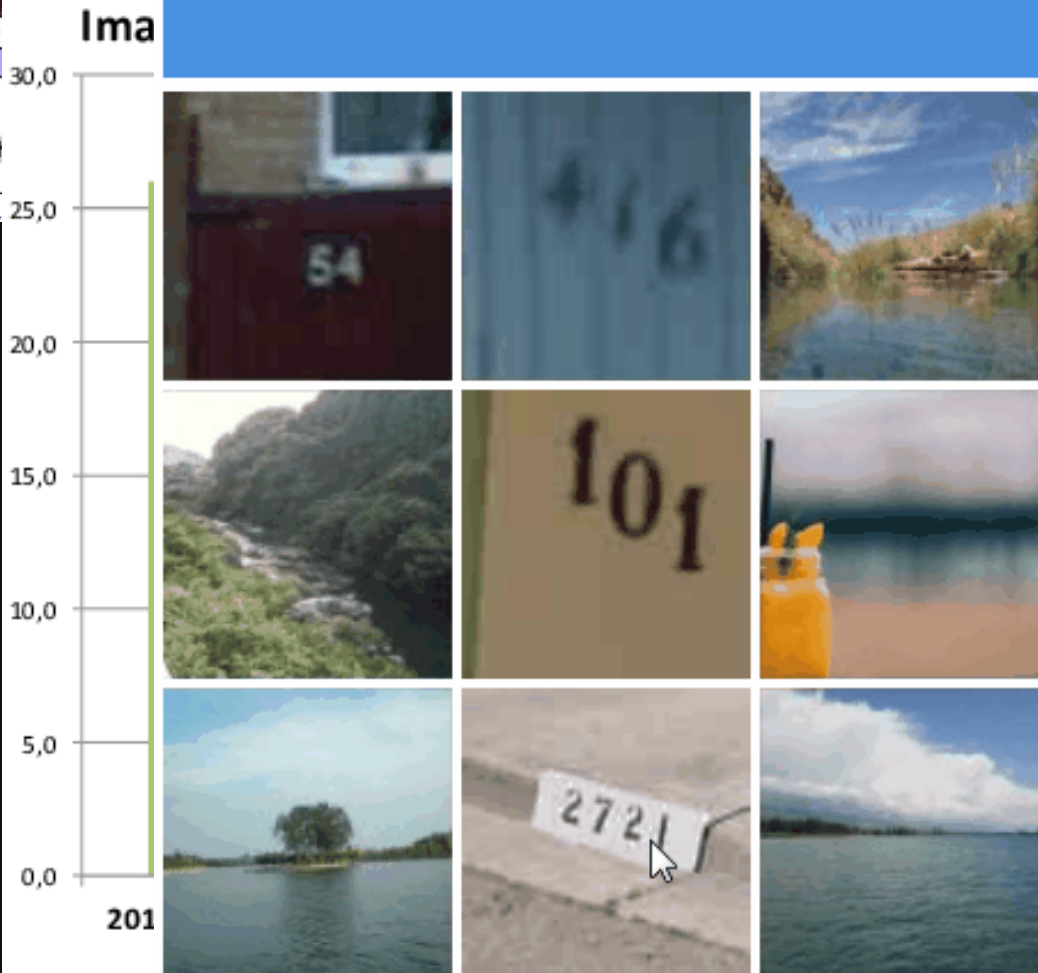
Select all images with **street numbers**.
Click verify once there are none left.



grille mushroom



Here 3 pictures are given as input to the network and the probability



Evaluiieren

Hands-On: MNIST Classifier

<https://www.youtube.com/embed/YDP65huTN98?enablejsapi=1>

Bearbeiten Sie [dieses Notebook](#)

- Berechnen Sie die Konfusionsmatrix sowie den F1 score auf einem beliebigen Datensatz
- Berechnen Sie diese explizit für ein untrainiertes Modell auf dem Testdatensatz

Die Lösung finden Sie in [diesem Notebook](#)