

Deep Learning-Based Image Classification for Endoscopic Images

Fachbereich II – Ingenieurinformatik

Erster Betreuer: Prof. Dr. rer. nat. Erik Rodner

Zweiter Betreuer: Dipl.-Ing.(FH) Holger Müller

Abdelrahman Elsharkawi | B.Sc. Ingenieurinformatik | Finale B.Sc. Präsentation



**Hochschule für Technik
und Wirtschaft Berlin**

University of Applied Sciences

Overview

- 1 Einführung
- 2 Arbeitsziel
- 3 Methode
- 4 Organ Classifier
- 5 Quality Classifier
- 6 Fazit

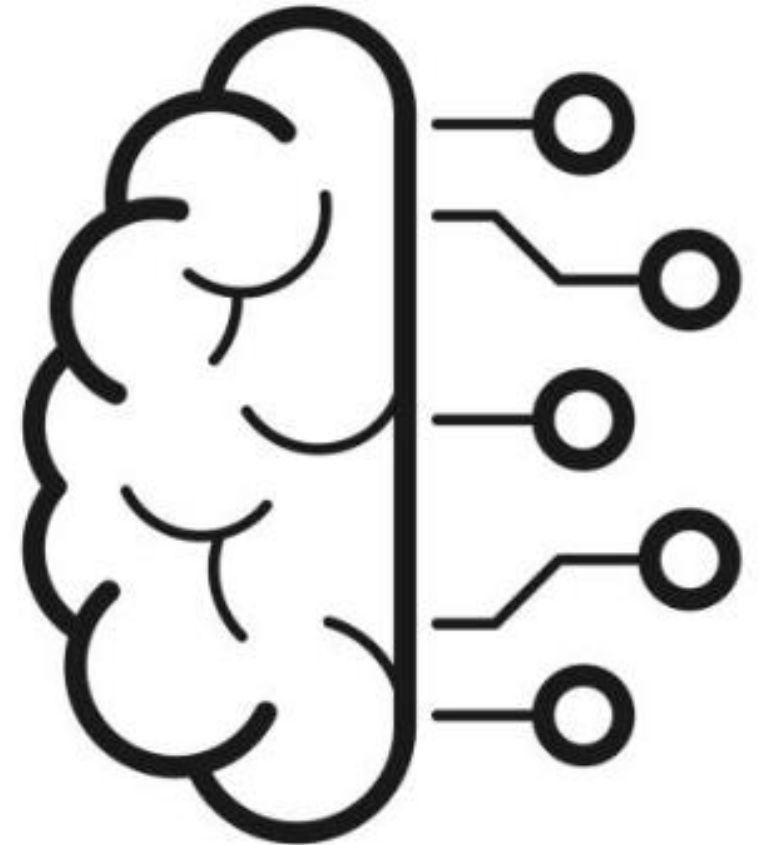
Einführung

- Was ist ein Endoskop?
- XION entwickelt sowohl starre als auch flexible Endoskope.
- Deep-Learning in der Endoskopie.



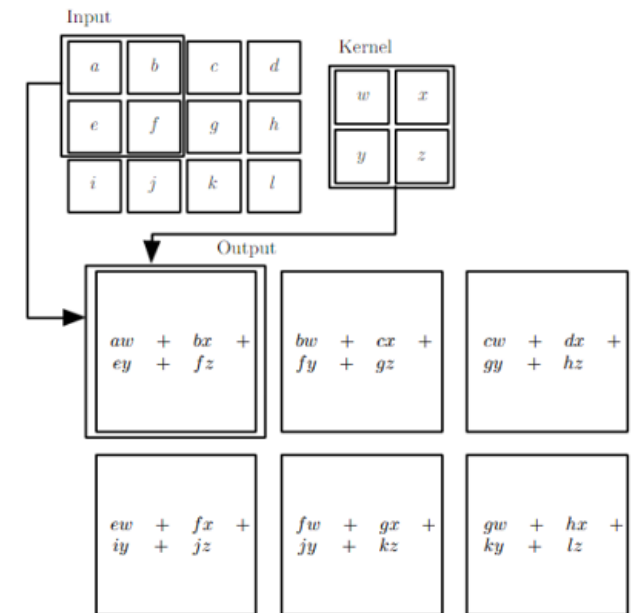
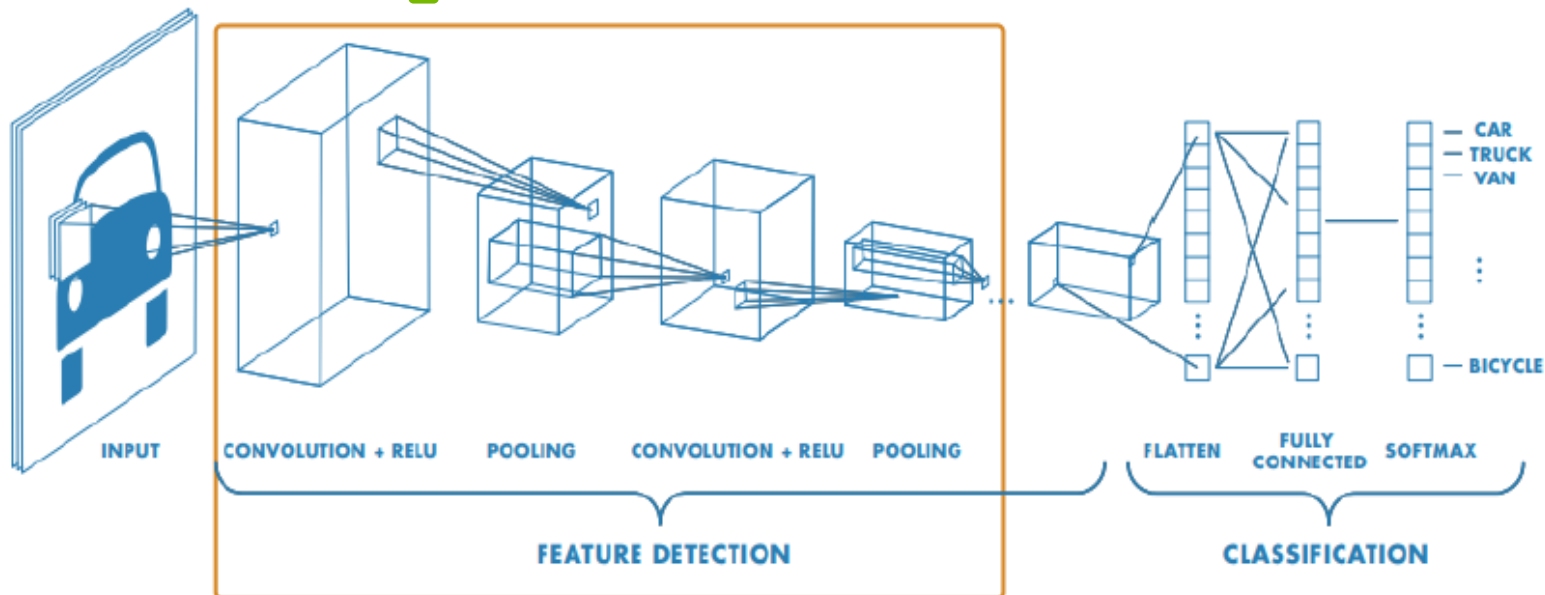
Deep Learning

- Was ist Machine Learning?
- Unterschied zwischen Machine Learning und Deep Learning.
- Warum habe ich Deep Learning benützt?

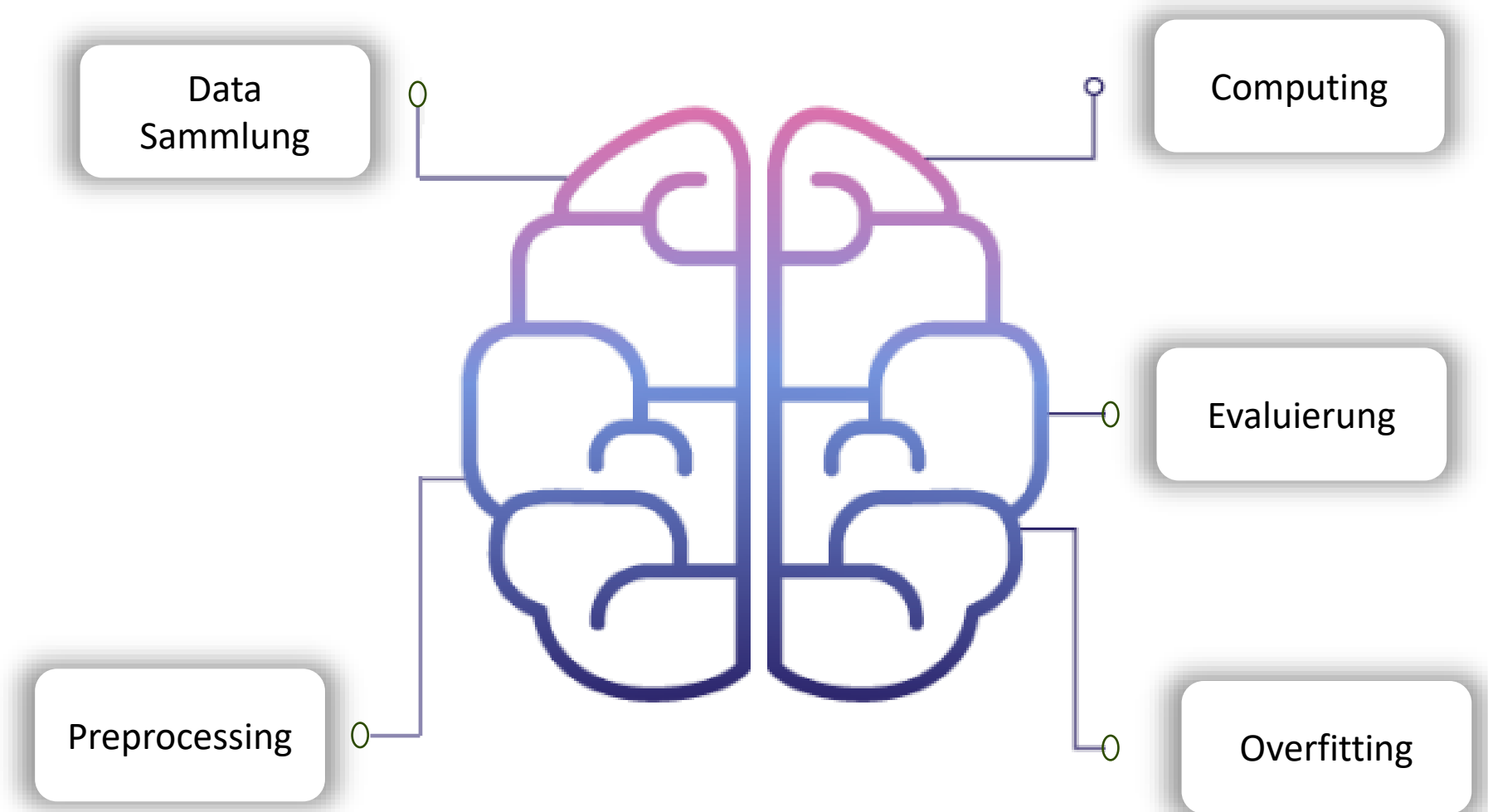


CNN

- Was ist die Convolution (Faltung)?
- Convolutional neural networks.
- Hyper Parameter.
- EfficientNetB0.
- Warum EfficientNetB0?
- Transfer learning.



Deep learning challenges



Arbeitsziel

Ziel:

Zwei ML-Modelle entwickeln, die die folgenden Aufgaben erfüllen:

- Klassifizierung abgebildeter Organe (Ohr, Nase oder Stimmlippen)
- Klassifizierung der Bildqualität (gut oder schlecht)

Nutzen:

Verbesserung der User Experience für XION-Kunden

- Fokussierung auf einen bestimmten Videoabschnitt.
- Entfernung von Bildern mit schlechter Qualität.



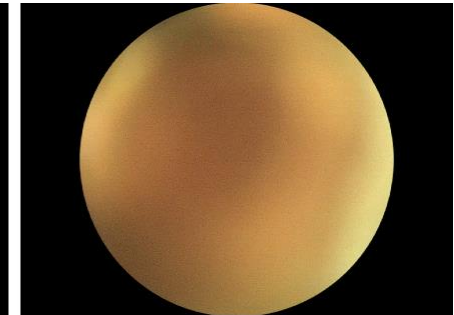
Pred: Nose | Prob: 0.994



Pred: bad | Prob: 0.937



Pred: Ear | Prob: 1.000



Pred: bad | Prob: 0.964



Pred: Vocal Folds | Prob: 1.000



Pred: bad | Prob: 0.925

Methode

Python, PyTorch

Datasets

Video → Frams
Folder Dataset
Bilder Bearbeiten
Tensor Dataset

CNN

EfficientNetB0

WandB

Datasets loggen
Modelle loggen

Train

Train the model

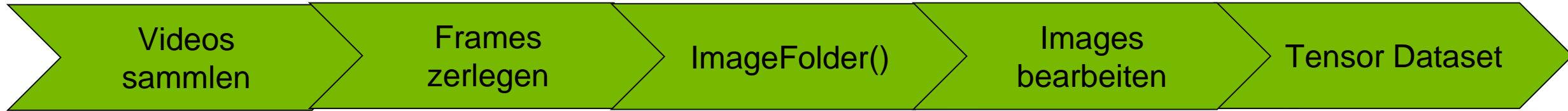
WandB

Tainierte Modelle
loggen
Ergebnisse loggen

Evaluate

Reports erstellen
Ergebnisse
Vergleichen

Datasets



- Sammlungsrunde1 882 Frames.
- Sammlungsrunde 21418 Frames.
- Videos aufnehmen
- XION Archive

- Ordner für jedes Patient
- Excel Sheet
- Organ
 - Patient_1_1
 - Patient_1_2

- Quality Dataset
- Erste Gruppe Organ Dataset
- Zweite Gruppe Organ Dataset

- Die Schwarzen Randen ausschneiden
- Schlecht qualität Images generieren

- Normalisieren
- Tupel(Tensor(3,H,W), Tensor(int))



CNN > Train

Train

```
print(f"\n-----run:{run_number}-----\n")
torch.manual_seed(42)
model_ = model_setups.efficientNet_b0(False, 2, device)
training_dict = dict(project_name="Quality-Classifier-1",
                    artifact_name="tr_v0_32_5e5",
                    artifact_discription="hyperparameter experiments",
                    dataset="quality_0:v0",
                    download_data=False,
                    batch_size=32,
                    run_number=run_number,
                    model_name="efficientNet_b0:v0",
                    model_filename="initialized_model.pth",
                    continue_training=False,
                    optimizer=torch.optim.Adam(params=model_.parameters(), lr=0.00005),
                    loss_fn=torch.nn.CrossEntropyLoss(),
                    epochs=30,
                    class_names=["bad", "good"],
                    device=device)

trained_model = engine.train_and_log(model_, training_dict)
```

Code Beispiel für ein Training Run

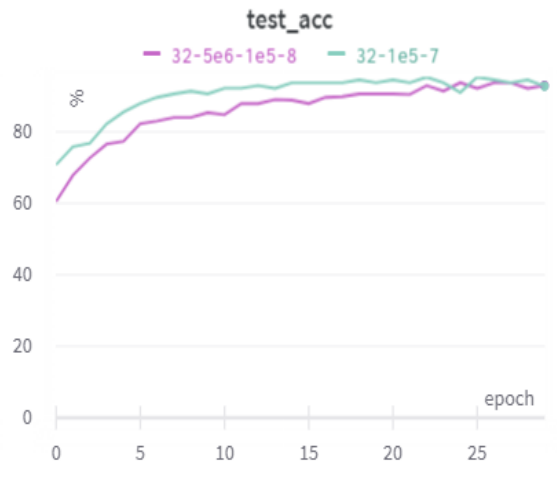
WandB

Evaluate

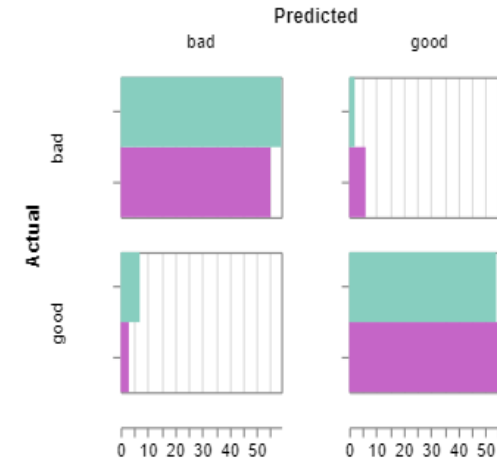
Evaluate



1 - Test loss



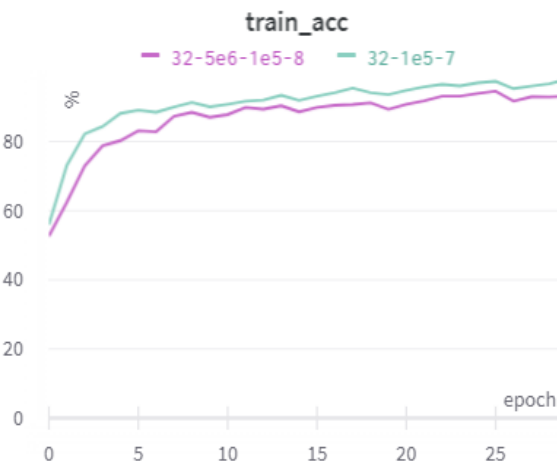
3 - Test accuracy



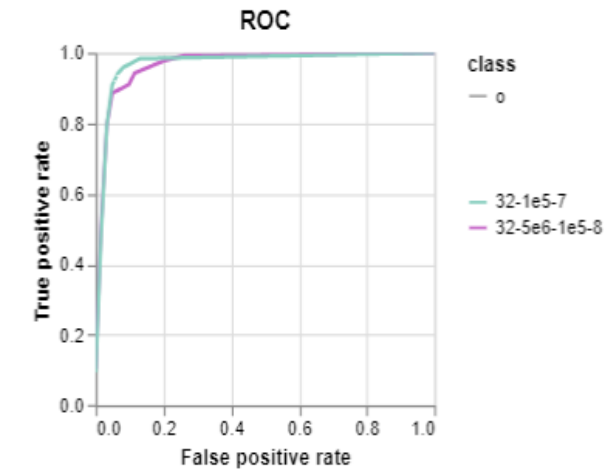
5 - Confusion Matrix



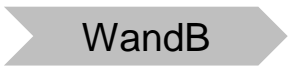
2 - Train Loss



4 - Train accuracy

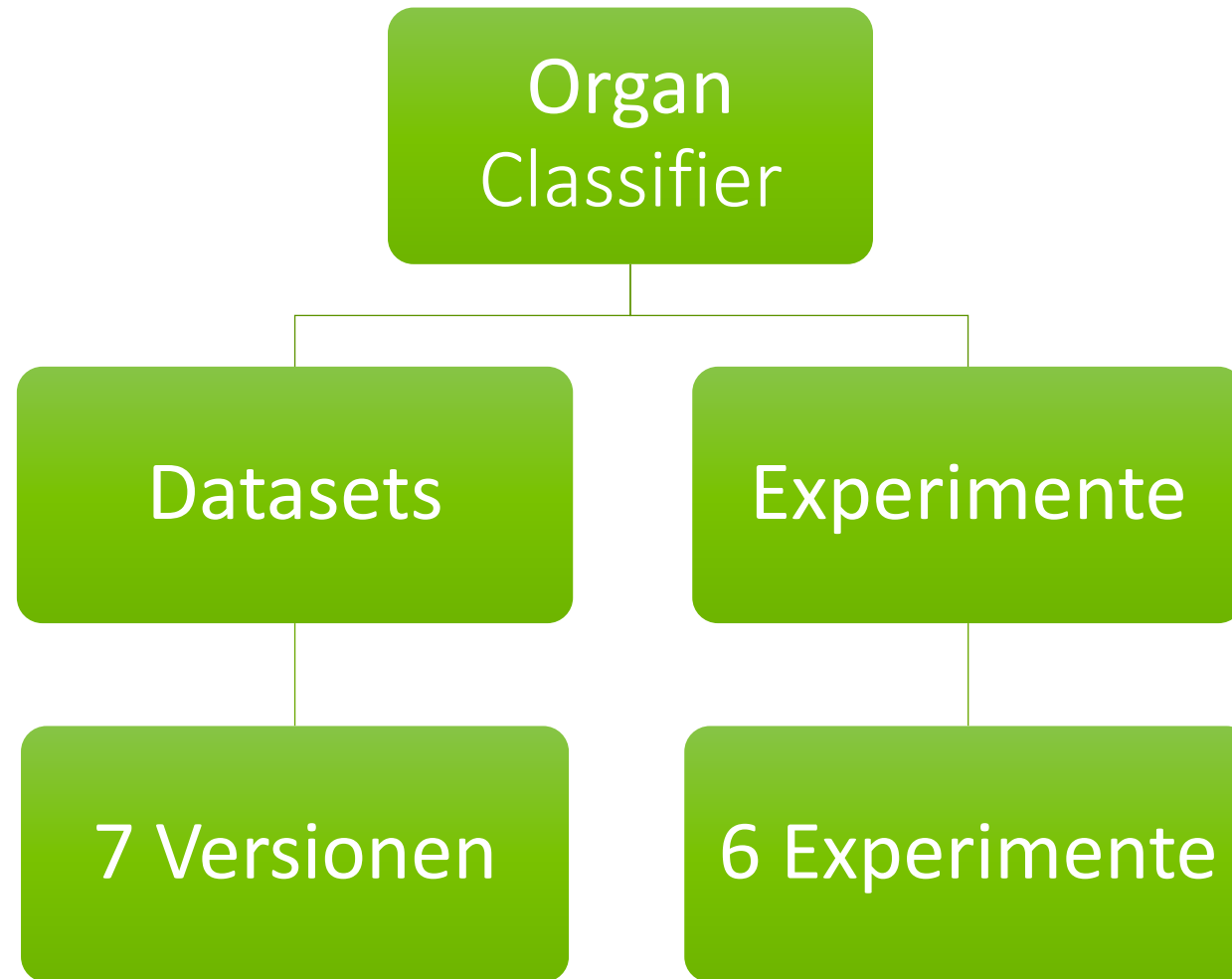


6 - ROC Kurve



WandB

- Weights and Biases.
- MLOps.
- Datasets Versionierung.
- Ergebnisse der Experimente.
- Demo.



Datasets

version	Train split			Test split			description
	Ear	Nose	Vocal folds	Ear	Nose	Vocal Folds	
0	220	307	111	37	128	78	No cropping
1	220	307	111	37	128	78	Train dataset is cropped
2	220	307	111	37	128	78	Train and test dataset are cropped
3	652	924	379	107	124	111	Uniformly cropped, noisy, and imbalanced
4	379	379	379	98	98	98	Uniformly cropped, noisy, and balanced
5	370	370	370	93	93	93	Uniformly cropped, clean, and balanced
6	370	370	370	98	98	98	Uniformly cropped, train clean and test noisy
7	379	379	379	93	93	93	Uniformly cropped, train noisy and test clean

Datasets

version	Train split			Test split			description
	Ear	Nose	Vocal folds	Ear	Nose	Vocal Folds	
0	220	307	111	37	128	78	No cropping
1	220	307	111	37	128	78	Train dataset is cropped
2	220	307	111	37	128	78	Train and test dataset are cropped
3	652	924	379	107	124	111	Uniformly cropped, noisy, and imbalanced
4	379	379	379	98	98	98	Uniformly cropped, noisy, and balanced
5	370	370	370	93	93	93	Uniformly cropped, clean, and balanced
6	370	370	370	98	98	98	Uniformly cropped, train clean and test noisy
7	379	379	379	93	93	93	Uniformly cropped, train noisy and test clean

Experimente

1. Transfer Learning vs Learning from Scratch
2. Finding Suitable Hyper Parameters for The First Datasets Group
3. Uniform Preprocessing
4. Finding Suitable Hyper Parameters for The Second Datasets Group
5. Balanced vs Imbalanced Classes
6. Clean vs Noisy Dataset

Was kann hier schief laufen?

Bilder von den unterschiedlichen Kameras haben unterschiedliche Formen



Otoskope



Nasopharyngoscope



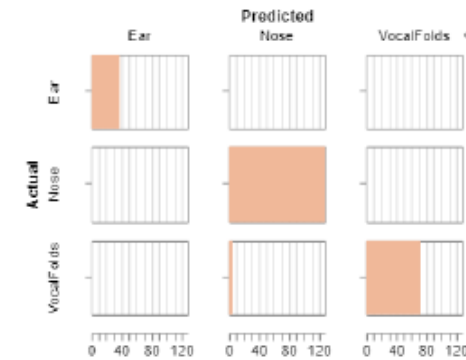
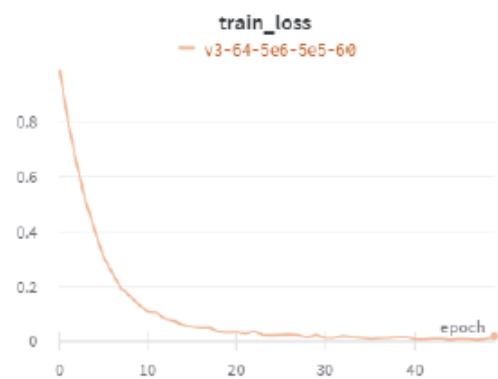
Nasopharyngoscope



Laryngoskope

Uniform Preprocessing

Ergebnisse durch Trainieren mit nicht beschnittenen Bildern:



Pred: ear | Prob: 0.642



Pred: nose | Prob: 0.997



Pred: ear | Prob: 0.502



Pred: nose | Prob: 0.803

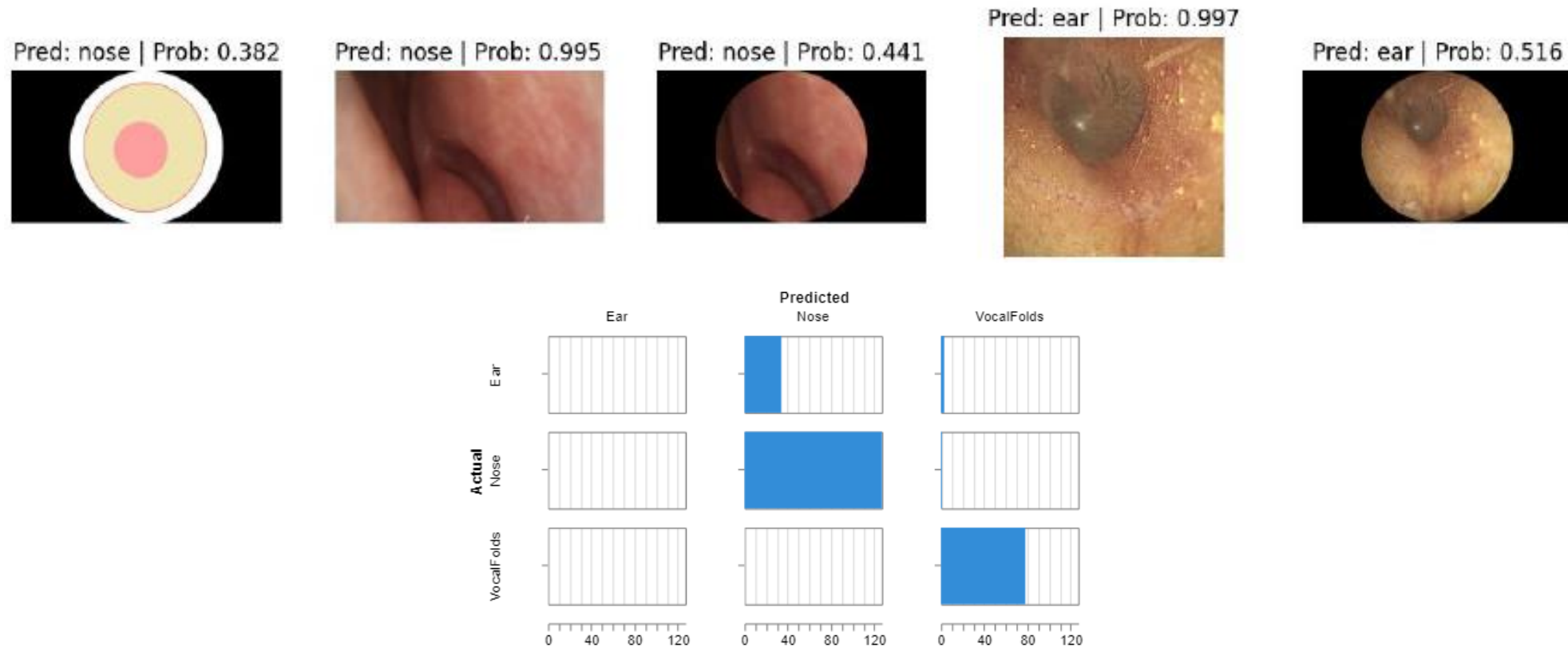


Pred: ear | Prob: 0.997



Uniform Preprocessing

- Schwarze Umrandungen im Training Dataset manuell weggeschnitten.
- Schwarze Umrandungen im Test Dataset vorhanden.



Uniform Preprocessing

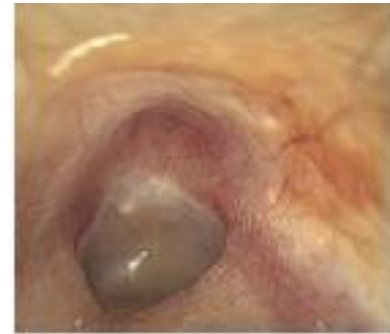
Methode zum automatischen Beschneiden der schwarzen Umrandung.



Nicht geschnitten

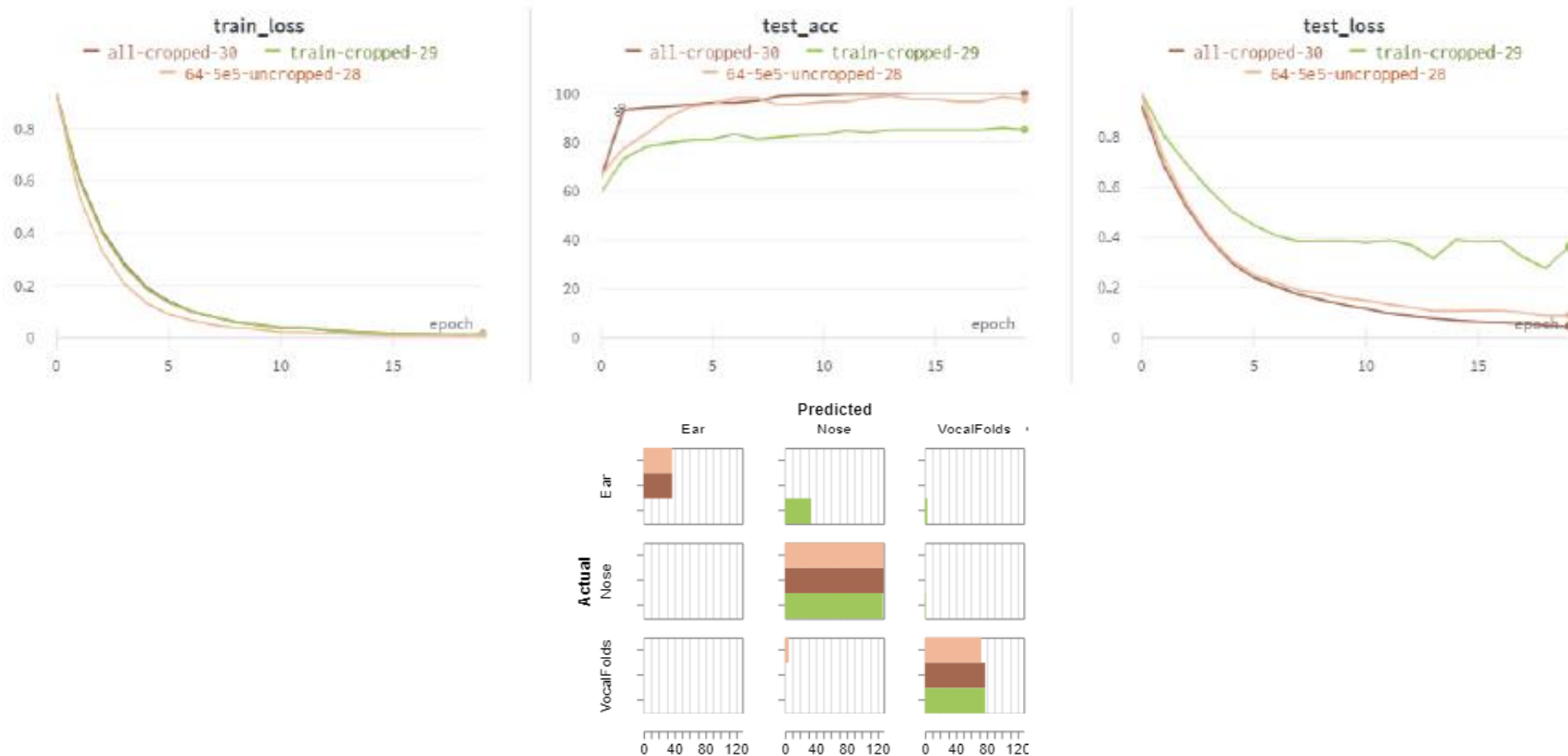


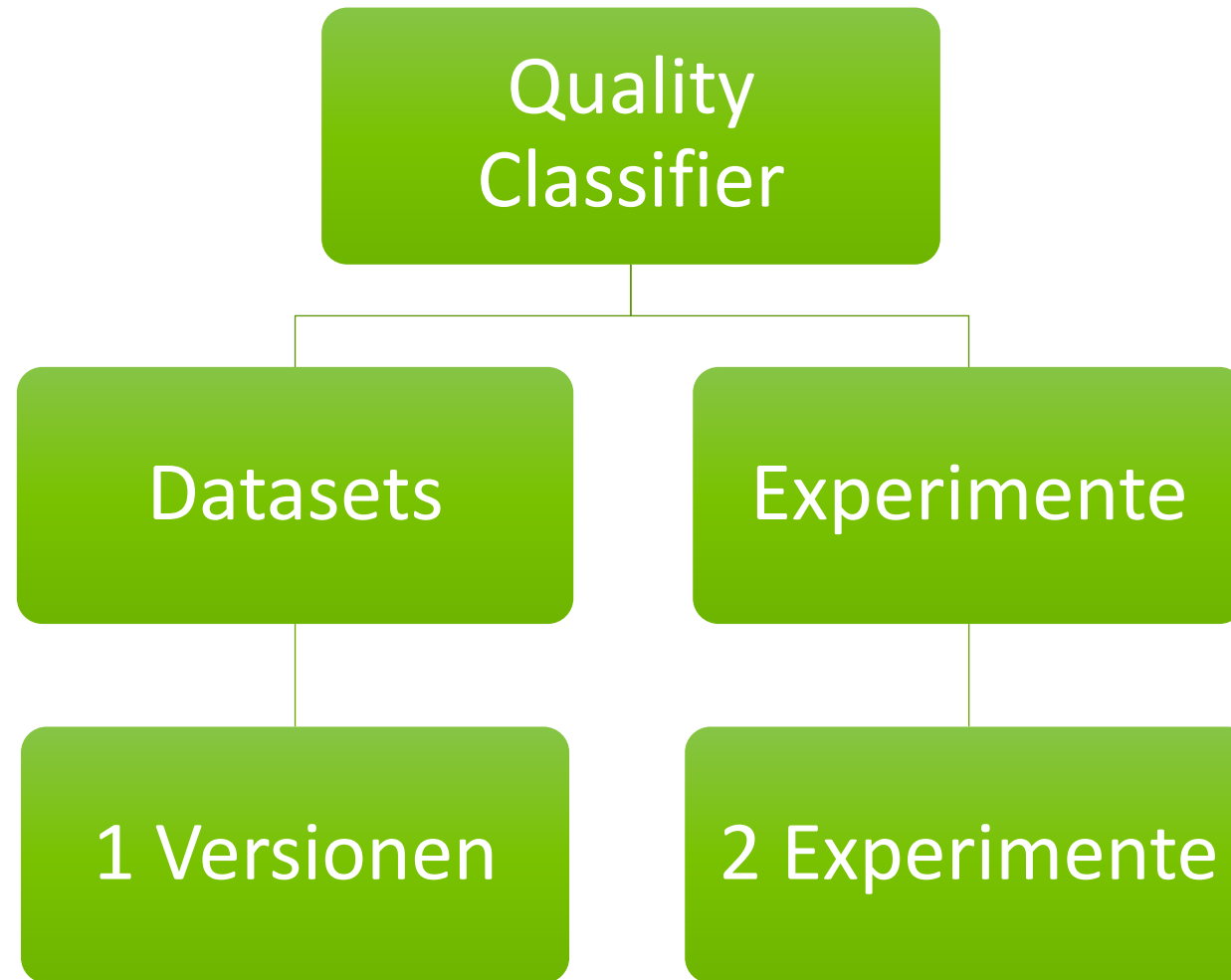
Schwarze Spalten
wurden beschnitten



Kreis ins Rechteck
transformiert

Uniform Preprocessing





Dataset-Labeling

- Ziel ist zwei balanced Klassen (bad|good) mit 339 Bilder in Training und 61 Bilder in Test.
- Es gibt keine genügenden schlechten Bilder.
- Schlechte Bilder aus guten Bildern generieren.
- Welche Bilder können bearbeitet werde?
- In der Excel Datei sind die Bilder in sechs Kategorien eingeteilt:

Kategorie	Sammlungsrunde	Anzahl der Bilder
Good	Erste	440
Very_good	Erste	166
Bad	Erste	222
Very_bad	Erste	45
Very_good	Zweite	61
Very_bad	Zweite	61

path	camera	quality	organ	data collection	split
../data/allData/Ear/Ear_1_1/Ear_1_1_000.png	Otoscope	good	Ear	first	train
../data/allData/Ear/Ear_1_1/Ear_1_1_001.png	Otoscope	very_bad	Ear	first	train
../data/allData/Ear/Ear_1_1/Ear_1_1_002.png	Otoscope	very_bad	Ear	first	train

```
test_ear_bad = find_cells_by_keyword(file_path,
                                     {2:"very_bad", 3:"Ear", 4:"second"})
```

```
test_ear_good = find_cells_by_keyword(file_path,
                                       {2:"very good", 3: "Ear" ,4: "second"})
```

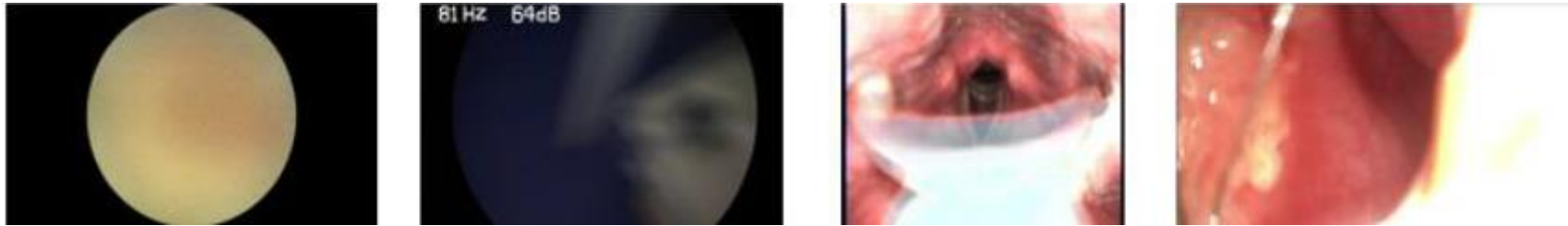
```
2 to_transform = bad_to_transform + good_to_transform
3 len(to_transform) # + 45 very bad images = 339 bad images in the training set
```

294

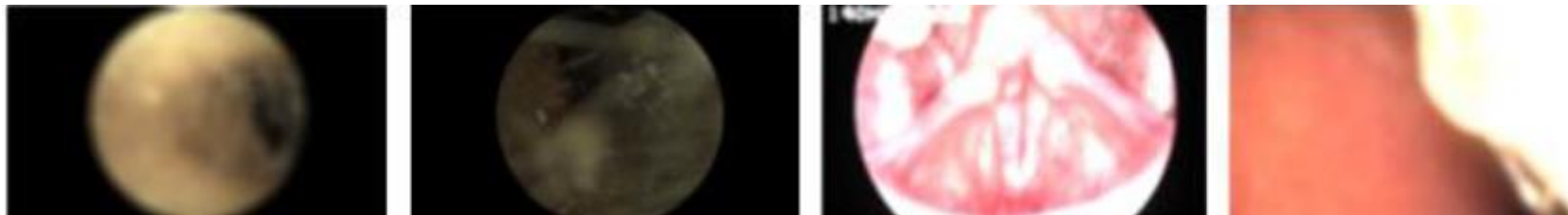
Schlechte Bilder generieren

Eins von Drei Transformationen wird auf jedes Bild in der Liste angewandt:

- Gaußsche Unschärfe mit einem zufällig ausgewählten Radius.
- Zufällige Helligkeitsänderung.
- kombiniert Unschärfe und Helligkeitsänderungen.



Raw schlecht



Generiert schlecht

Experimente

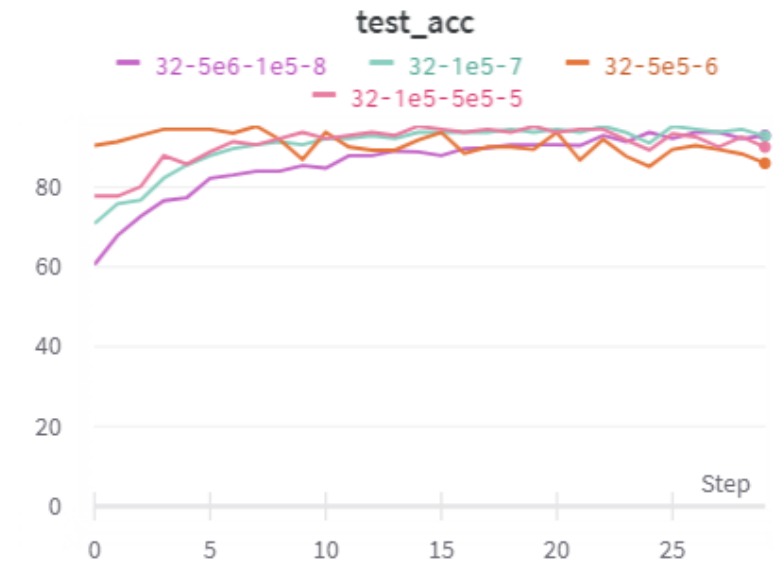
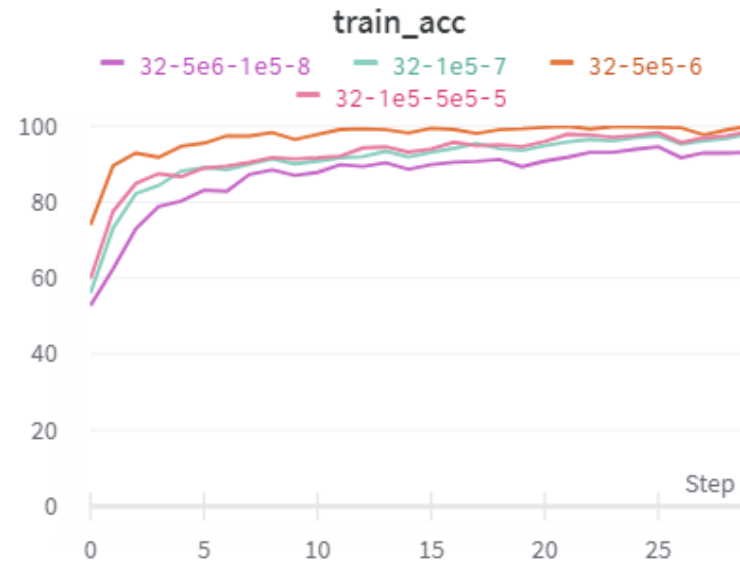
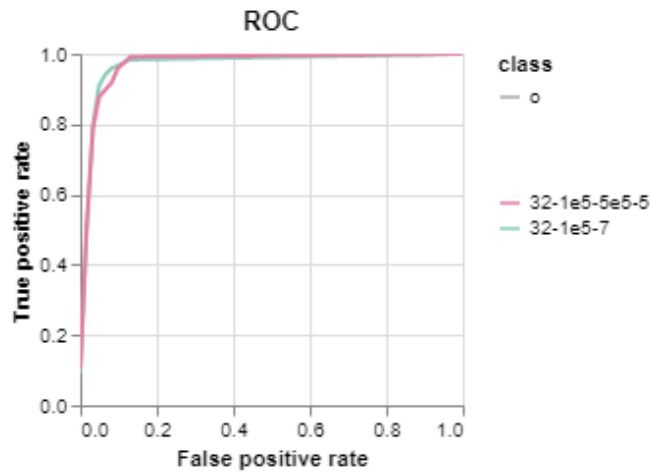
Alle Experimente in diesem Kapitel werden Transfer-Learning-Techniken anwenden und ein zugeschnittenes und ausgewogenes Dataset nutzen.

Experimente:

1. Varying Learning Rates with Constant Batch Size of 32.
2. Continue Training with Batch Size of 64.

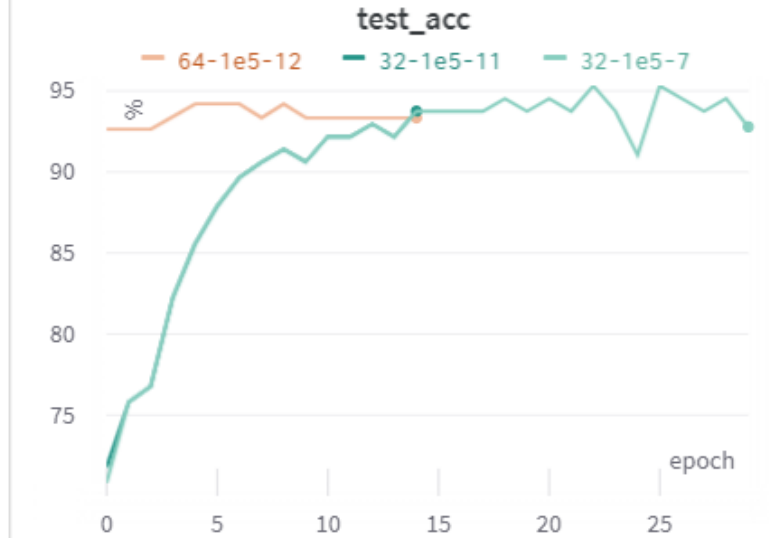
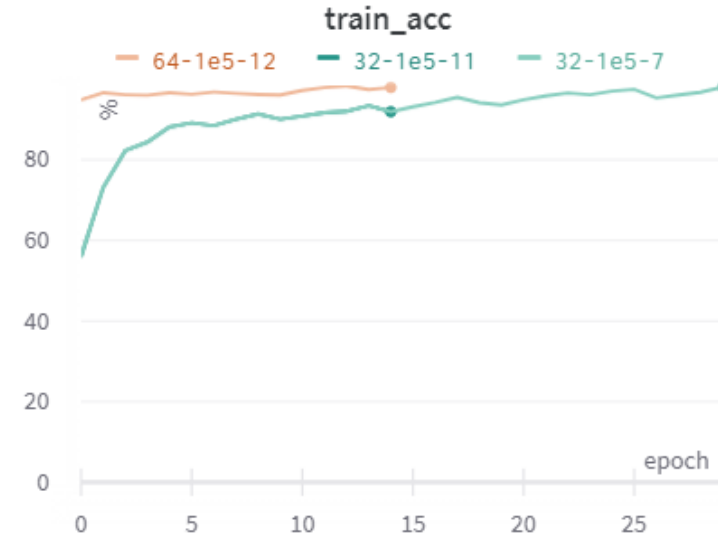
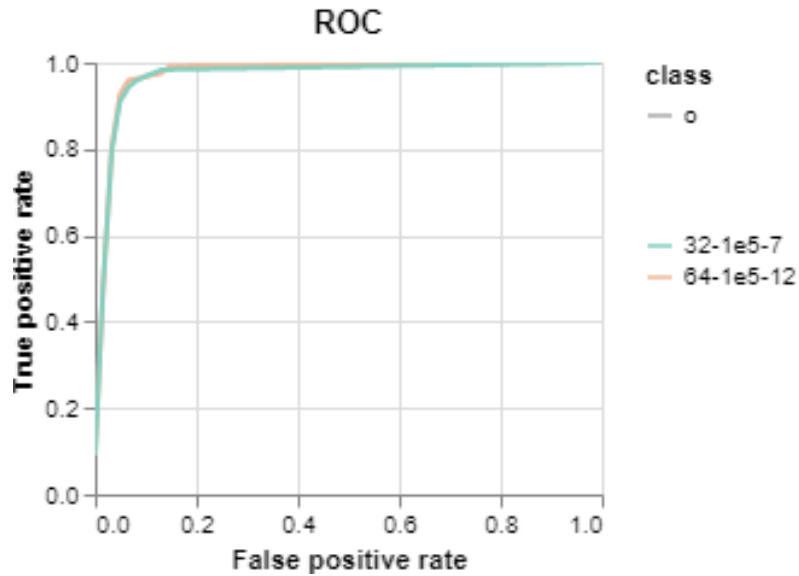
Varying Learning Rates with Constant Batch Size of 32.

best performing:
Batch-Size 32 mit Learning-Rate 1e-5.



Continue Training with Batch Size of 64.

- Kleinere Batches zeigen schnelleres lernen.
- Größere Batches konvergieren zu den globalen Optima.



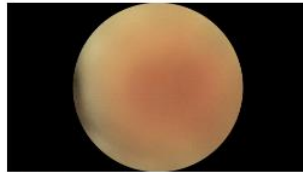
Flask-App

Endoscopy classifier

A bachelor thesis of Abdelrahman Elsharkawi at HTW-Berlin and XION

The image quality will be classified as either good or bad. In case it is good the organ in the image will be predicted

Uploader

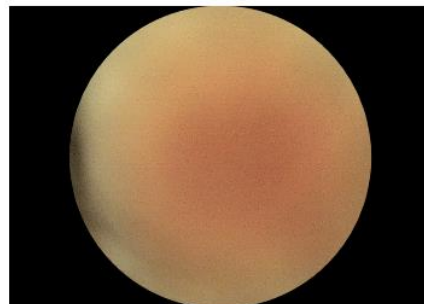


Upload your image:

Choose file | Ear_2_1_017.png

CLASSIFY!

Results



Pred: bad | Prob: 0.984



Pred: Ear | Prob: 0.999



Pred: Nose | Prob: 0.999

Fazit

- Organe und Bildqualität ließen sich gut mit DL erkennen (klassifizieren).
- Fazit der Experimenten:
 1. Transfer Learning konvergierte schneller als Learning from Scratch.
 2. Geeignete Hyper Parameter konnten für beide Datensammlungen gefunden werden.
 3. Uniform Preprocessing war zur Unterdrückung der Gerätemaskierung nötig.
 4. Balanced Datasets tendierten weniger zu Overfitting.
 5. Bildrauschen stört nicht beim Anlernen, sollte aber bei Nutzung des Modells entfernt werden.
 6. Die Verengung des Batch-Sizes hat den Ergebnis verbessert.
- Ergebnisse zeigen Machbarkeit der Anwendung von DL bei XION mit einer hohen Genauigkeit.
- Mögliche Verbesserungen.

Vielen Dank.

Abdelrahman Elsharkawi

Master Student Informatik in der Ingenieurwissenschaften

LinkedIn  : Abdelrahman Elsharkawi

www.htw-berlin.de



**Hochschule für Technik
und Wirtschaft Berlin**

University of Applied Sciences