Data Challenge

Klassifikation von Münzenbildern vom Corpus-Nummorum Projekt*

*https://www.corpus-nummorum.eu/

Cahide Heidemann & Öykü Gercek

Ziele von Zwischenpräsentation

- Finetuning bei Transformer Modelle

- Vergleich der Leistung:
  - ➢ Typerkennung: Vor- und Rückseite zusammen in einem Bild

  - ➢ Mints: Vor- und Rückseite zusammen in einem Bild

  - ➢ Coins: Vor- und Rückseite getrennt

- Test mit verschiedenen Datasets, z. B. Image Augmentation

Die Münze des Monats: Telesphoros mit Weintraube*

Werkzeuge

- Schnittstellen von Hugging Face fürs Training & Evaluieren
- Pytorch für Grafikkarte & Preprocessing von Daten
- Roboflow für Data Split & Augmentation
- Weights & Biases für Logging der Metriken

Hugging Face

roboflow

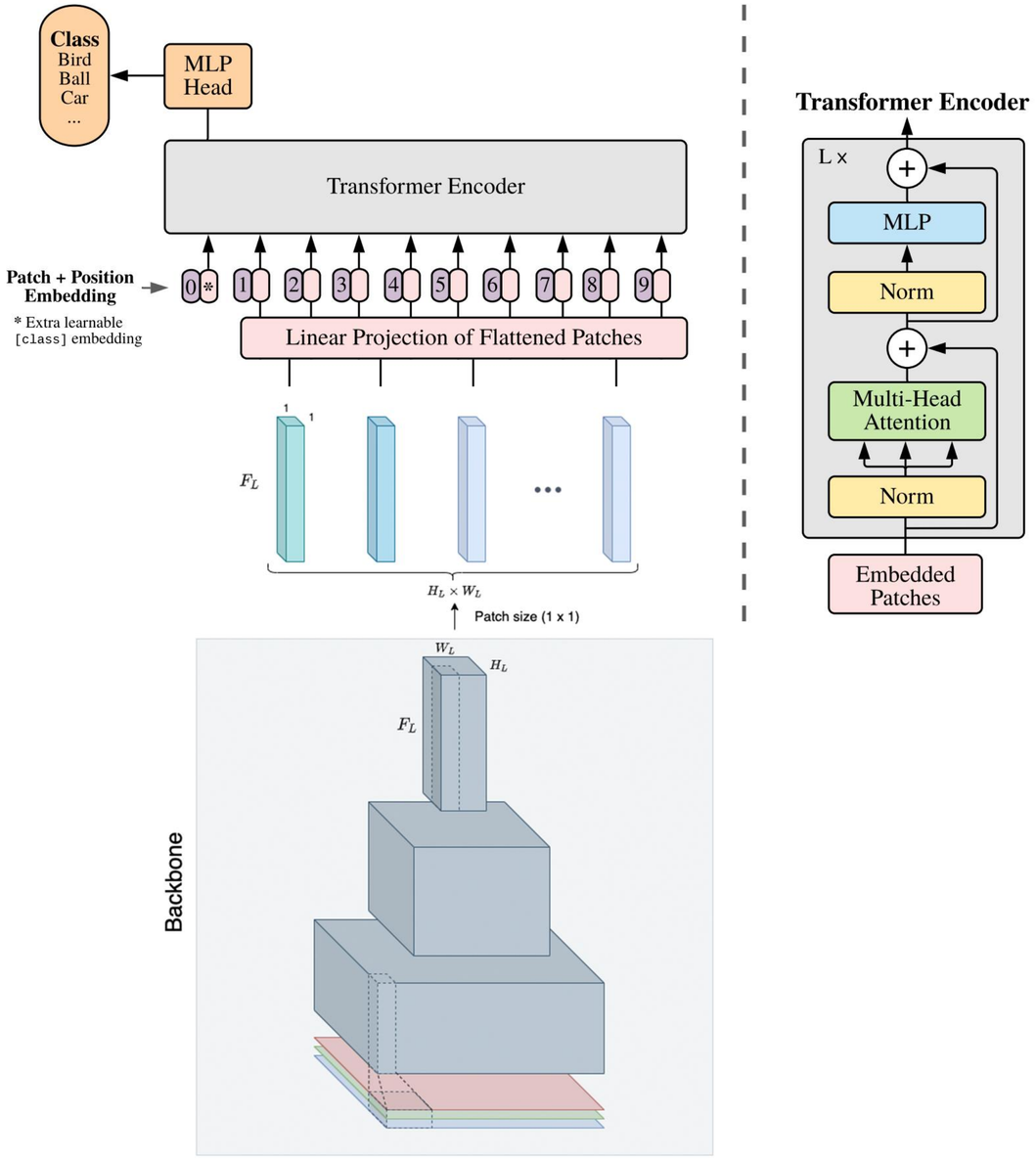
Weights & Biases

Werkzeuge

- Zwei Server des AI Systems Engineering Lab
- Je vier NVIDIA Tesla V100 GPUs (á 32GB) and A100 GPUs (á 40GB)

Transformer-Modelle

- ViT: Vision Transformer
- DeiT: Data-efficient Image Transformer
- BEiT: BERT Pre-Training Image Transformer
- LeViT: LeNet Vision Transformer
- ViThybrid: Vision Transformer hybrid
- Swin: Shifted Window Transformer
- CvT: Convolutional Vision Transformer
- BiT: Big Transfer Transformer
- EfficientFormer: Efficient Vision Transformer



Quelle: https://sh-tsang.medium.com/review-beit-bert-pre-training-of-image-transformers-c14a7ef7e295

Transformer-Modelle

Modelle	Merkmale
ViT	erster Transformer, der für die Bilderkennung und –klassifizierung trainiert wurde
DeiT	"distillierte" Version von ViT, der effizienter arbeitet, um mit weniger Daten gute Ergebnisse zu erzielen
BeiT	inspiriert von BERT, benutzt selbst-überwachtes Lernen beim Training statt überwachtes
LeViT	Convolutions werden vor den Transformer-Teil geschaltet (Self-Attention + Classifier)
ViThybrid	Vision Transformer, der mit CNN Feature Extractor kombiniert wurde, um genaure Daten aus Patches zu generieren
Swin	"Shifted Windows"; um Self Attention effizienter auf nicht überlappenden Patches zu fokussieren
Cvt	Versucht, Eigenschaften von CNNs auf Transformer zu übertragen und die Attention-Mechanismen zu behalten
BiT	Fokus auf Pre-Training auf großen Datasets, um spezifische Finetune Tasks zu verbessern
Efficient-Former	Kleinere Architektur um Ressourcen zu sparen; gemeint für Mobilgeräte

Vorgehen beim Testen

Datensatz Split in Train / Test / Validation

Datensatz Augmentation

Parameter Anpassung (Batch-size & Learning rate)

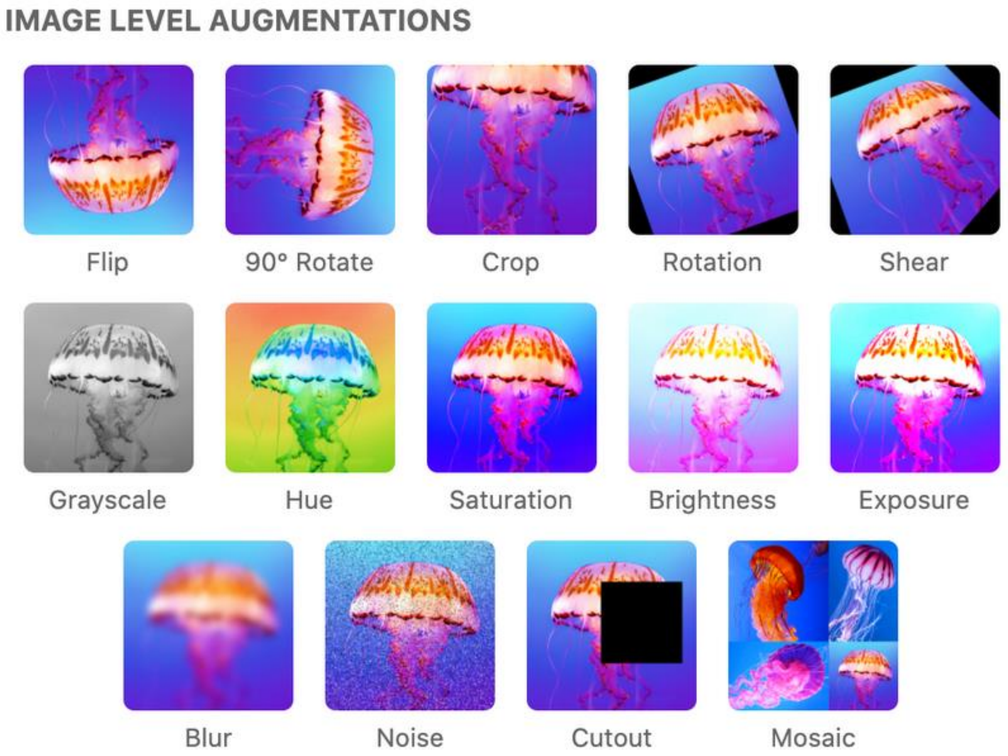
Erneuter Test: das beste Ergebnis mit unterschiedlichen Parametern

Testverfahren

- Insgesamt 250 Tests ausgeführt

- Transformer-Modelle (erster Transformer, schnellste Transformer, Transformer für kleine Datensätze, Transformer mit Convolutional NN – Eigenschaften)

- Augmentations:
  - Blur
  - Brightness
  - Exposure
  - Grayscale
  - Hue
  - Noise
  - Saturation



Quelle: https://blog.roboflow.com/introducing-grayscale-and-hue-augmentations/

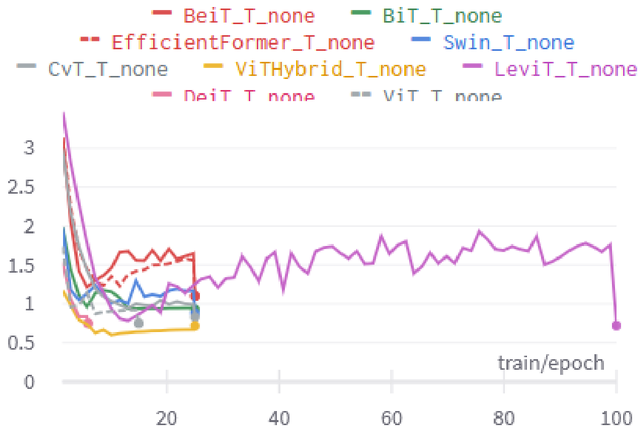
Werte mit unveränderten Bildern: Types

Längste Dauer

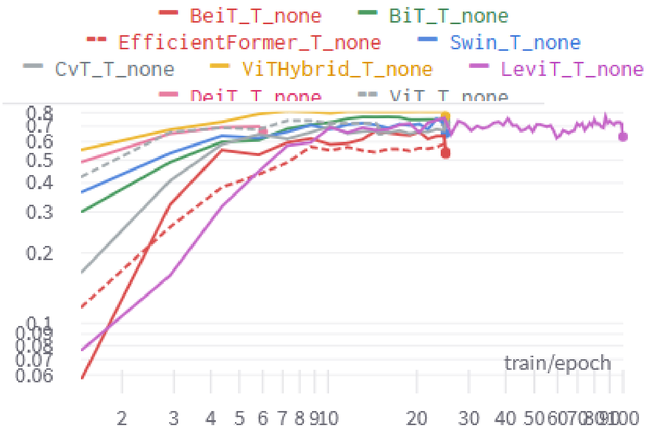
ViTHybrid_T_none

226.629

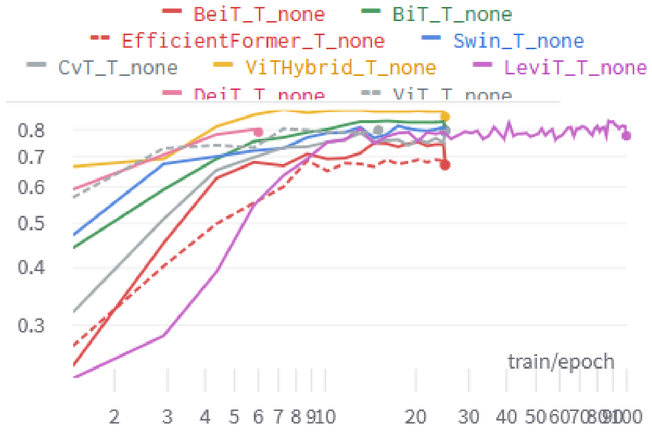
eval/loss



eval/f1



eval/accuracy



Min of Accuracy

EfficientFormer_T_none

0.6695

Max of Accuracy

ViTHybrid_T_none

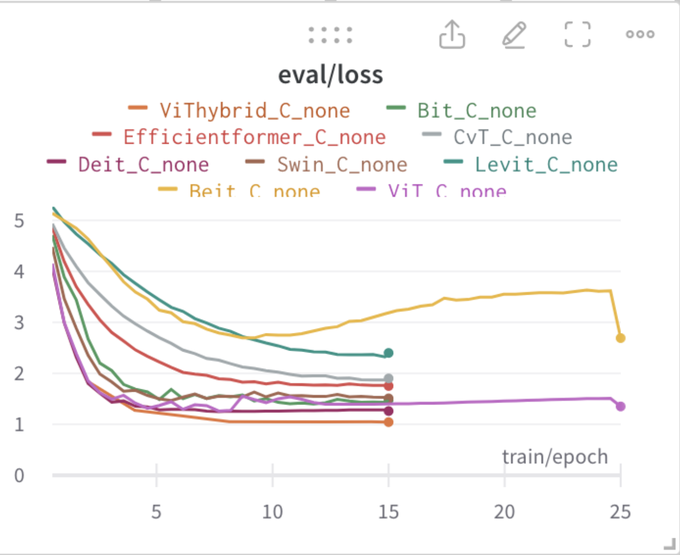
0.8534

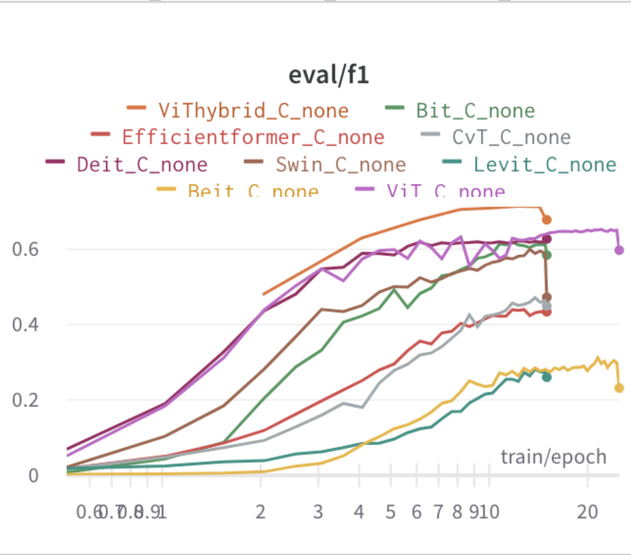
Werte mit unveränderten Bildern: Coins

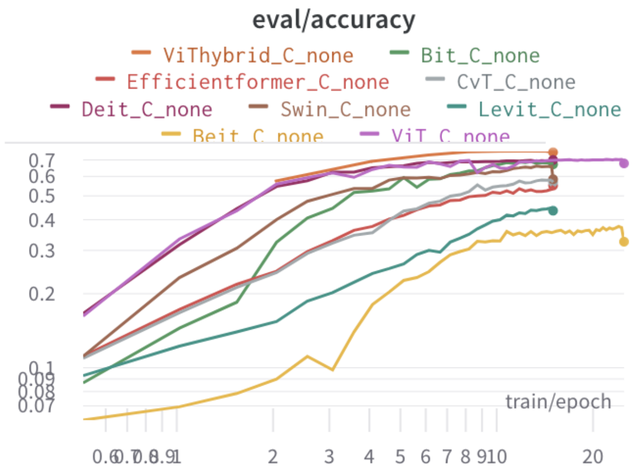
Längste Dauer

ViT_C_none

34999.688







Min of Accuracy

Beit_C_none

0.3255

Max of Accuracy

ViThybrid_C_none

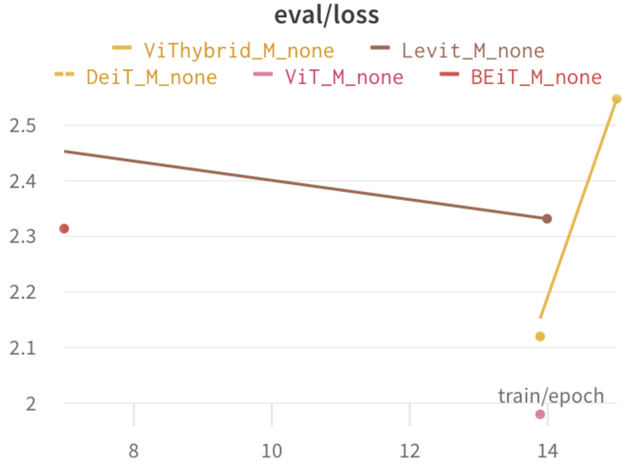
0.7523

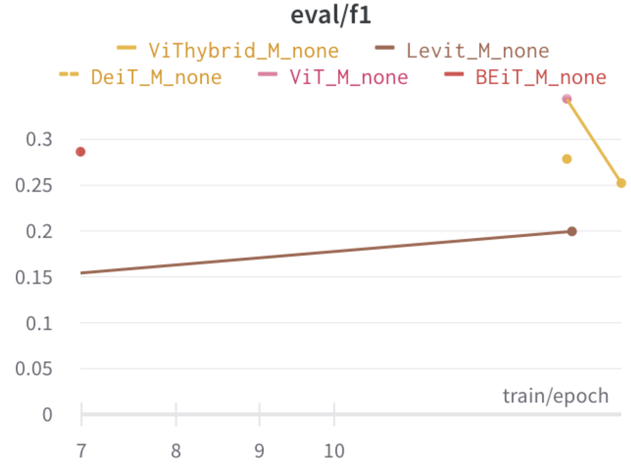
Werte mit unveränderten Bildern: Mints

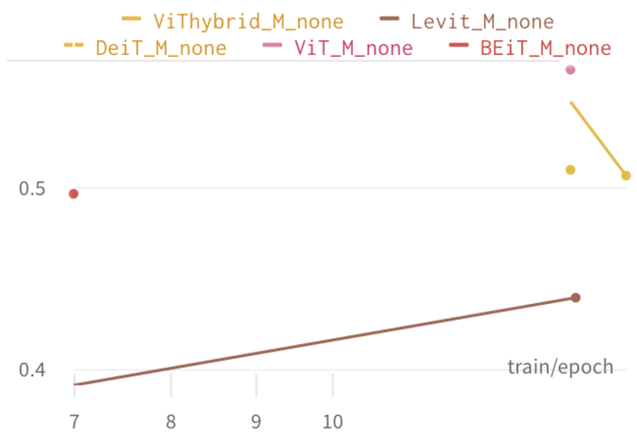
Längste Dauer

ViThybrid_M_none

978.895







Min of Accuracy

Levit_M_none

0.437

Max of Accuracy

ViT_M_none

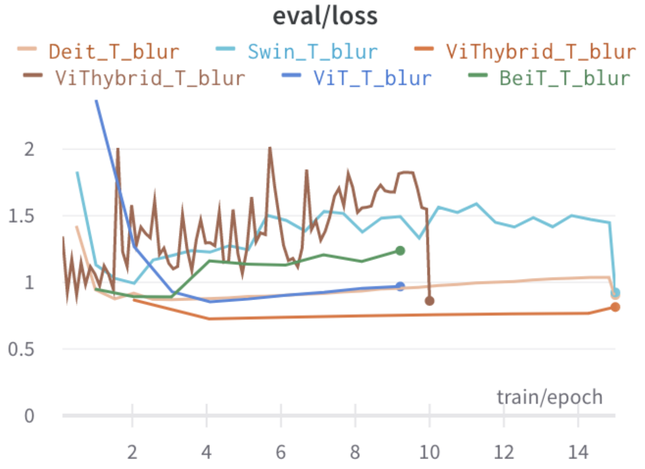
0.5778

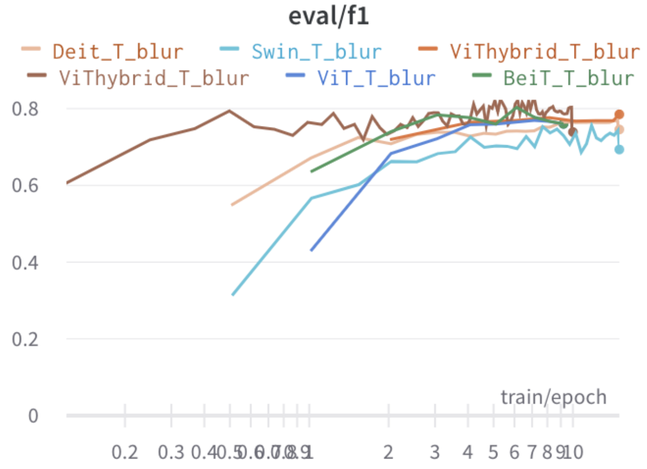
Werte mit veränderten Bildern: Blur

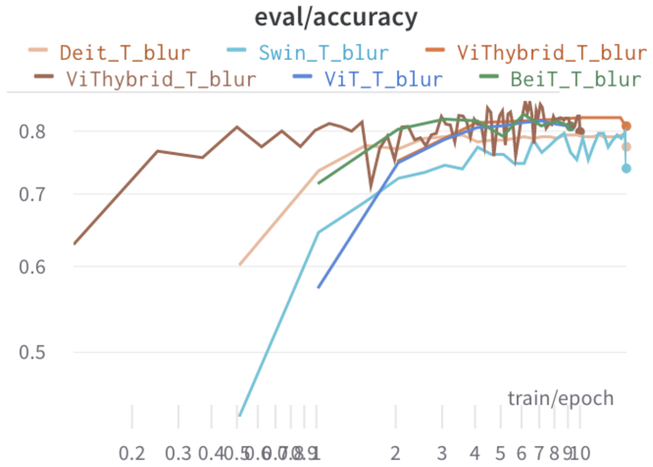
Längste Dauer

ViThybrid_T_blur

85959.408







Min of Accuracy

Swin_T_blur

0.7391

Max of Accuracy

ViThybrid_T_blur

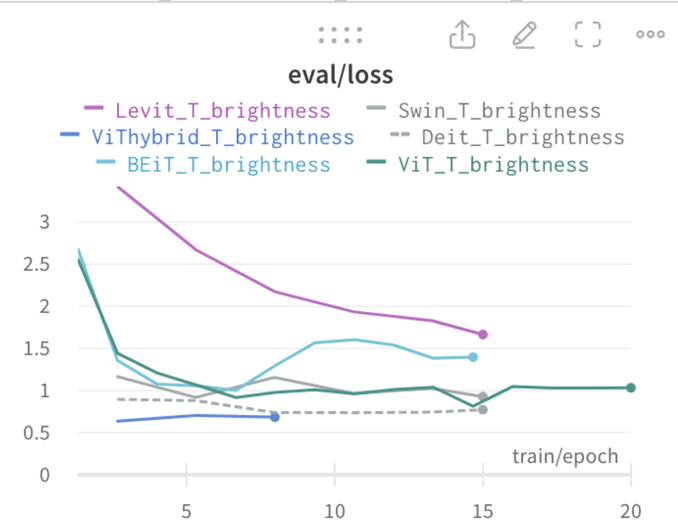
0.8087

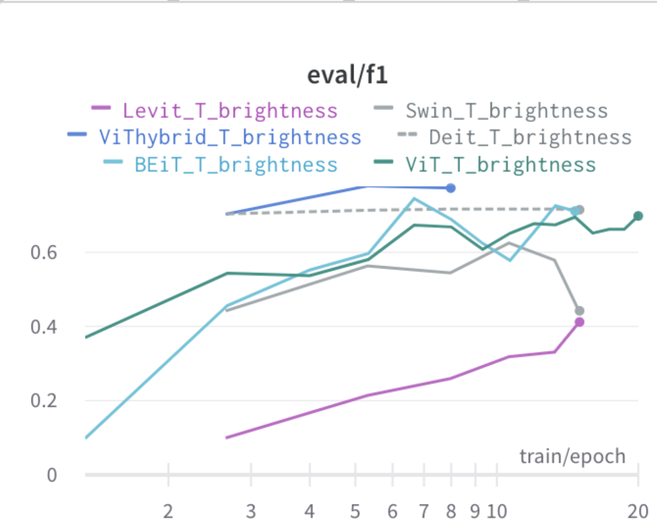
Werte mit veränderten Bildern: Brightness

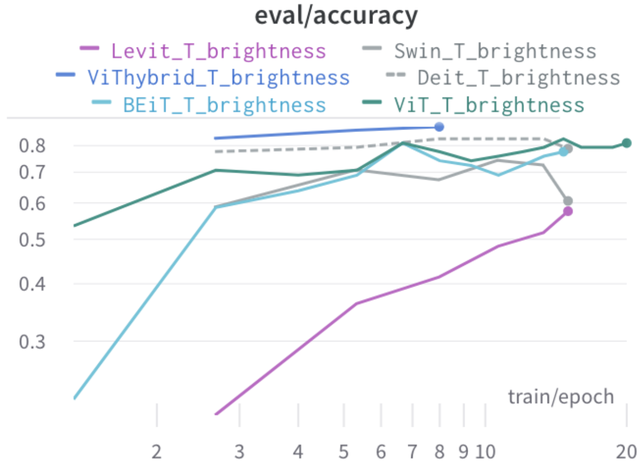
Längste Dauer

BEiT_T_brightness

5116.495







Min of Accuracy

Levit_T_brightness

0.5758

Max of Accuracy

ViThybrid_T_brightness

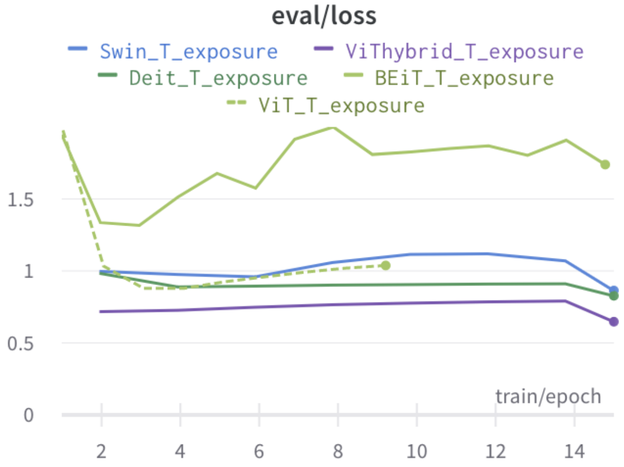
0.8793

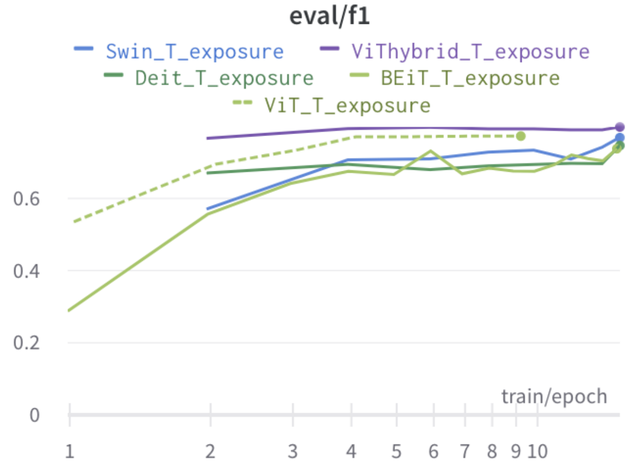
Werte mit veränderten Bildern: Exposure

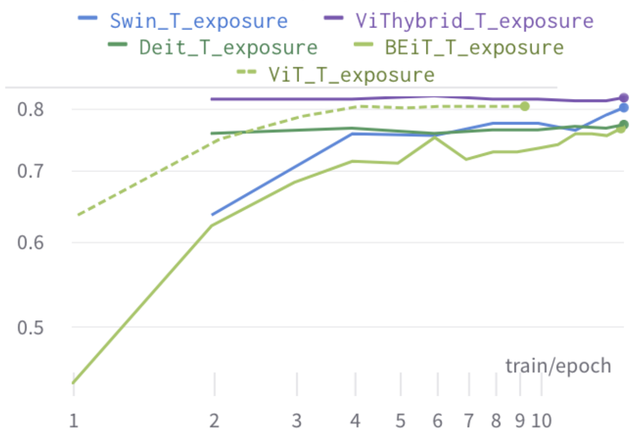
Längste Dauer

BEiT_T_exposure

7003.683







Min of Accuracy

BEiT_T_exposure

0.7674

Max of Accuracy

ViThybrid_T_exposure

0.8203

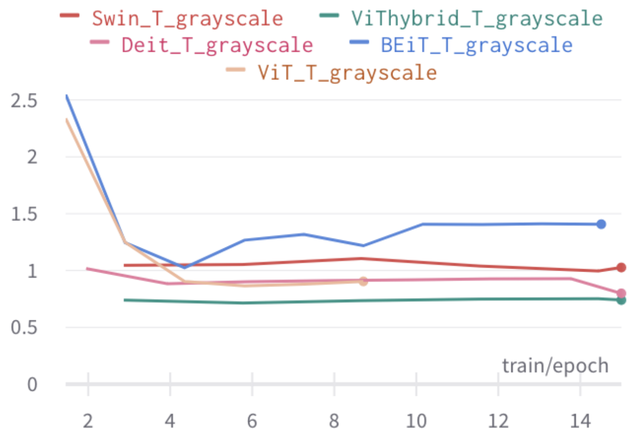
Werte mit veränderten Bildern: Grayscale

Längste Dauer

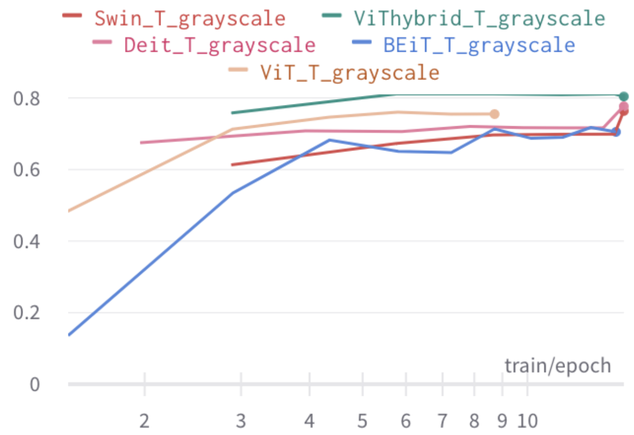
ViThybrid_T_grayscale

4931.252

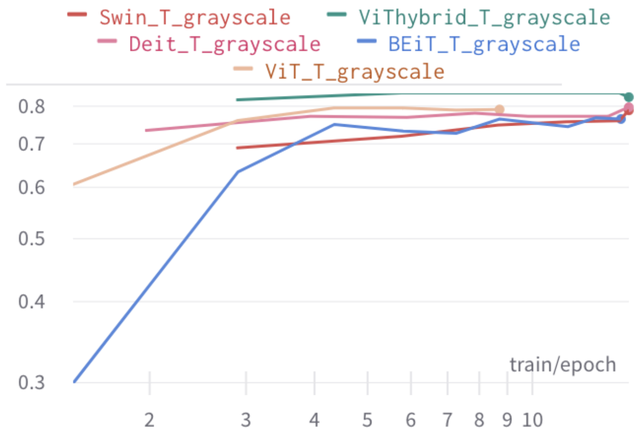
eval/loss



eval/f1



eval/accuracy



Min of Accuracy

BEiT_T_grayscale

0.7645

Max of Accuracy

ViThybrid_T_grayscale

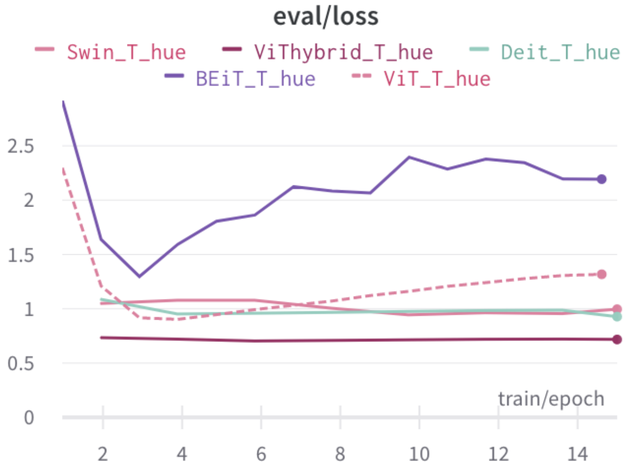
0.8261

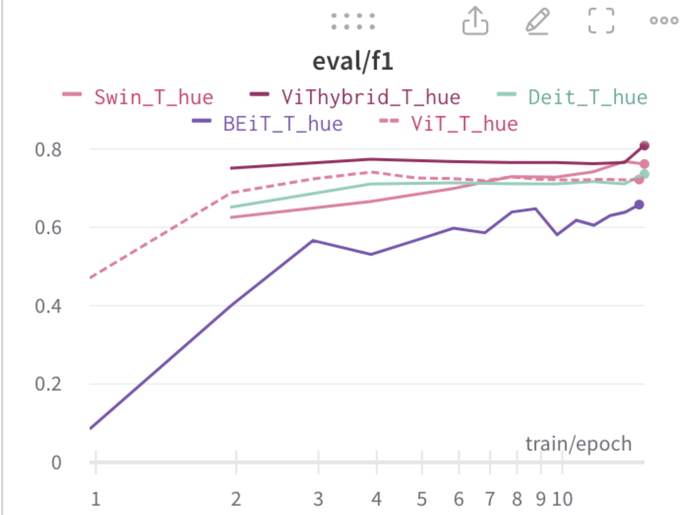
Werte mit veränderten Bildern: Hue

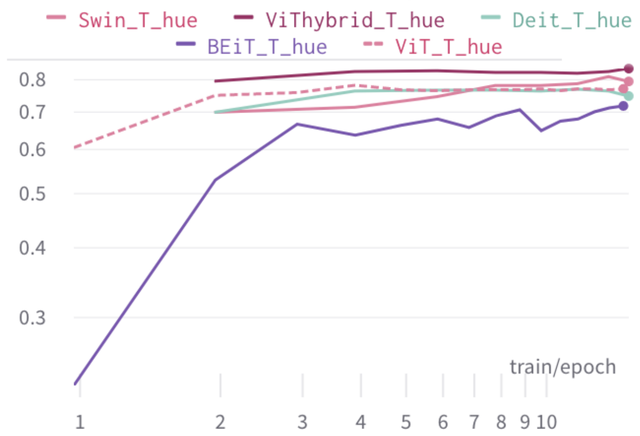
Längste Dauer

ViThybrid_T_hue

6679.319







Min of Accuracy

BEiT_T_hue

0.718

Max of Accuracy

ViThybrid_T_hue

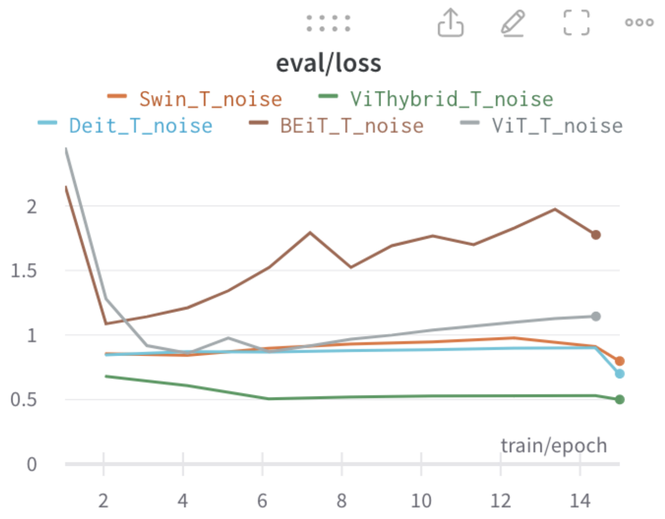
0.8377

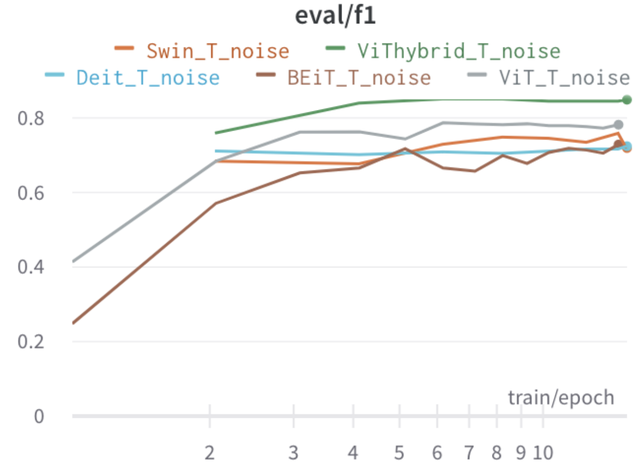
Werte mit veränderten Bildern: Noise

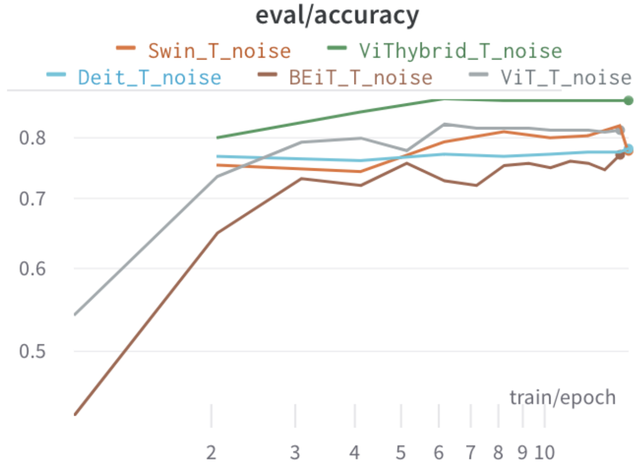
Längste Dauer

ViThybrid_T_noise

7025.21







Min of Accuracy

BEiT_T_noise

0.7706

Max of Accuracy

ViThybrid_T_noise

0.8686

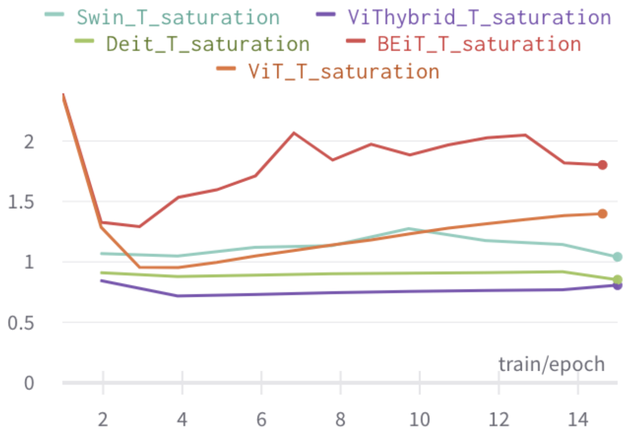
Werte mit veränderten Bildern: Saturation

Längste Dauer

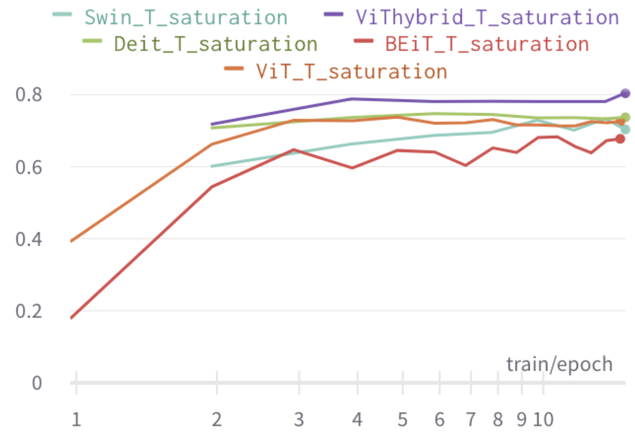
ViThybrid_T_saturation

18713.525

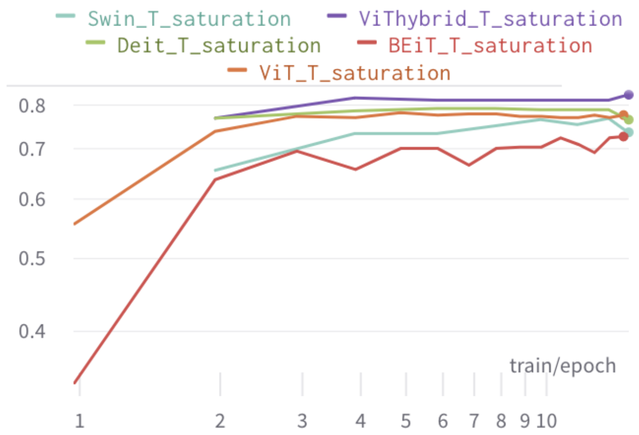
eval/loss



eval/f1



eval/accuracy



Min of Accuracy

BEiT_T_saturation

0.7267

Max of Accuracy

ViThybrid_T_saturation

0.8261

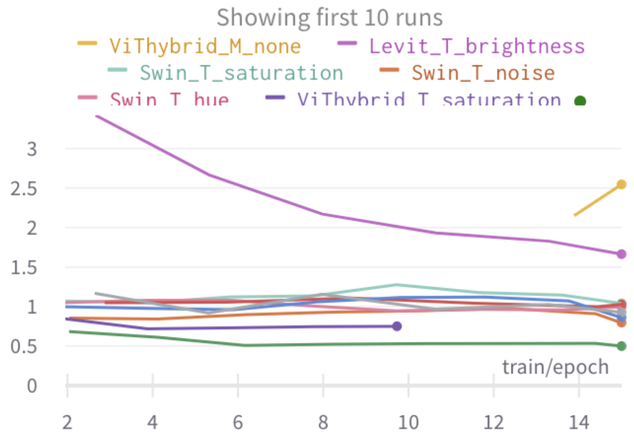
Beste vs. schlechteste Ergebnisse über alle Runs

Längste Dauer

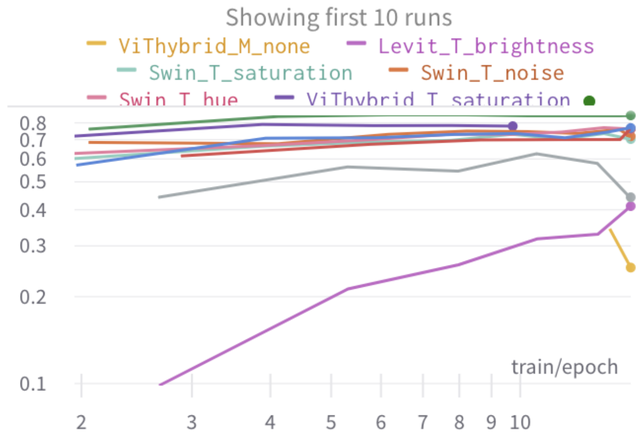
ViT_T_all

2781.436

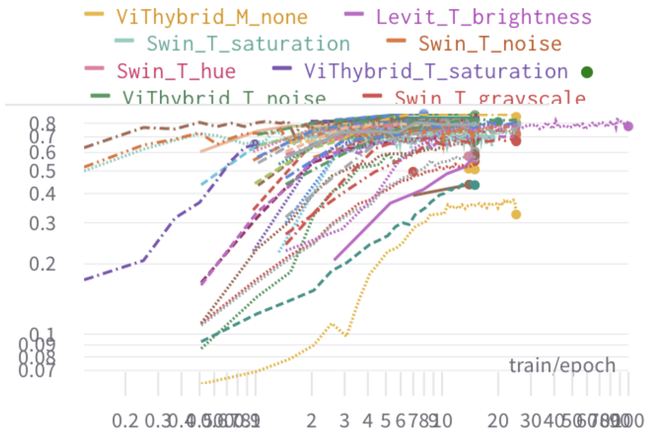
eval/loss



eval/f1



eval/accuracy



Min of Accuracy

Beit_C_none

0.3255

Max of Accuracy

ViThybrid_T_brightness

0.8793

Probleme

- Dependencies beim Ausführen
- Beim Data-Splitten wurden Bilder verdreifacht
  → 95% Accuracy
- Server-Maintenance
- Server-Belastung

Zusammenfassung

- Verschiedene Transformer-Modelle angewendet
- Typerkennung 85% Accuracy
- Typerkennung 87,9% Accuracy mit Augmentation (Brightness)
- Minterkennung 57,8% Accuracy
- Coinerkennung 75,2% Accuracy
- Beste Ergebnisse von ViThybrid (Convolutional Feature Extractor + Transformer Encoding & Classification)

Zusammenfassung

Alle Runs sichtbar unter:

https://wandb.ai/oeykue_cahide/huggingface/reports/Zusammenfassung-aller-Ergebnisse--Vmlldzo0OTM3NDA5

Trainings-Scripts (ohne Datasets):

https://github.com/cr-heidemann/Data-Challenges

Literatur

- Bao, Dong, Piao, Wei (2021): BEiT: BERT Pre-Training for Image Transformers.
- Kolesnikov, Beyer, Zhai, Puigcerver, Yung, Sylvain Gelly, Houlsby (2019): Big Transfer (BiT): General Visual Representation Learning.
- Wu, Xiao, Codella, Liu, Dai, Yuan, Zhang (2021): CvT: Introducing Convolutions to Vision Transformers.
- Touvron, Cord, Douze, Massa, Sablayrolles, Jégou (2020): Training data-efficient image transformers & distillation through attention.
- Li, Yuan, Wen, Hu, Evangelidis, Tulyakov, Wang, Ren (2022): EfficientFormer: Vision Transformers at MobileNet Speed.
- Graham, El-Nouby, Touvron, Stock, Joulin, Jégou, Douze (2021): LeViT: a Vision Transformer in ConvNet's Clothing for Faster Inference.
- Liu, Lin, Cao, Hu, Wei, Zhang, Lin, Guo (2021): Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.
- Sosovitskiy, Beyer, Kolesnikov, Weissenborn, Zhai, Unterthiner, Dehghani, Minderer, Heigold, Gelly, Uszkoreit, Houlsby (2020): An Image is Worth 16x16 Wirds: Transformers for Image Recognition at Scale.

Danke für die Aufmerksamkeit!