Self-supervised learning in the ‘wild’

Péter István

Bartos-Elekes Miklós

Tárnok Márton

Task

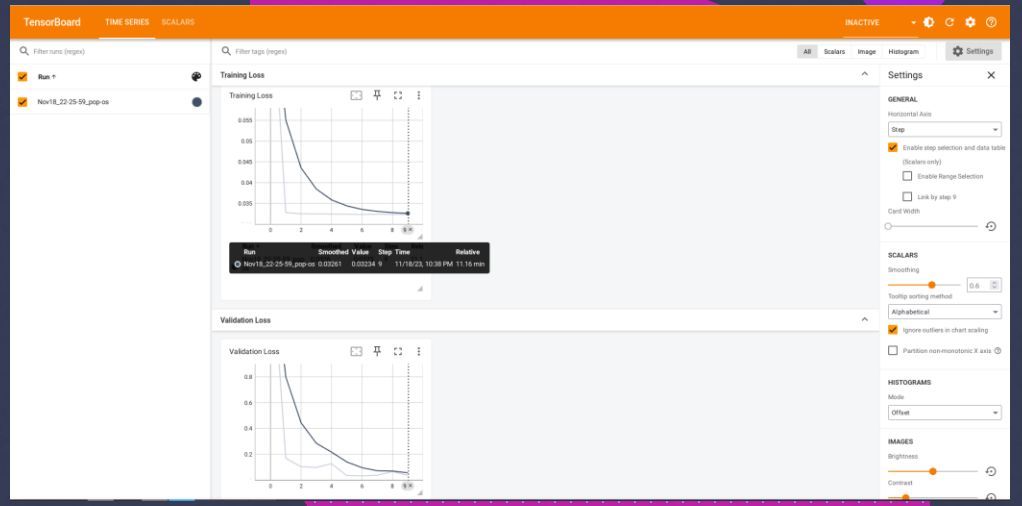
- Investigate various open-source self-supervised learning methods
- Select one or two classification datasets unrelated to ImageNet
- Utilize the selected SSL methods to pretrain models on your chosen datasets.
- Conduct a linear benchmark evaluation using a portion of ImageNet
- Compare the performance of your pretrained models against those publicly available and trained on ImageNet

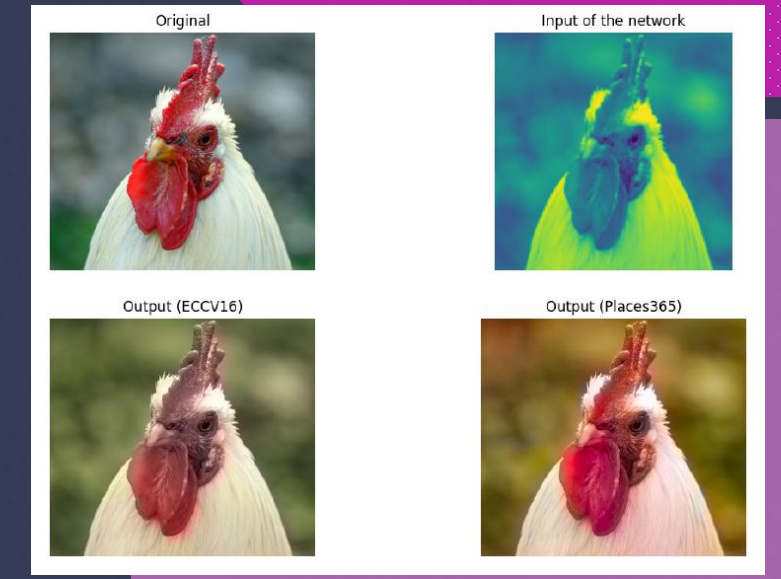
Concept

- Two different methods:
  - Colorization
  - Jigsaw
- Various training techniques and hyperparameter settings
- Train a baseline model (VISSL framework [4])

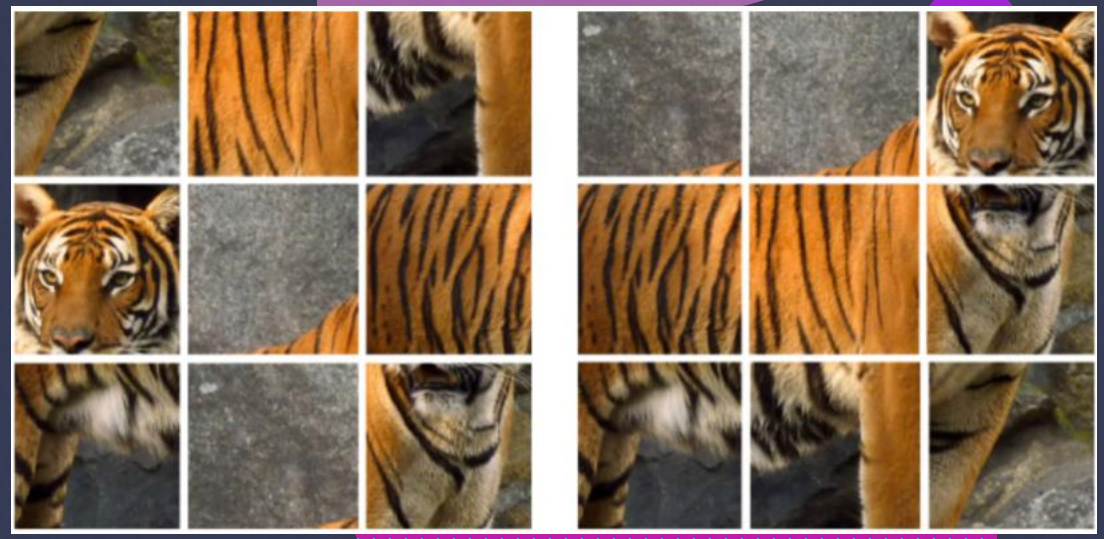
Colorization

- Colorful Image Colorization [1]
- Places365 dataset [2]
- Resolution:128x128
- 35 hours of training, 10 epochs
- Split the model into backend and head, frozen the weights and merged with the ResNet50





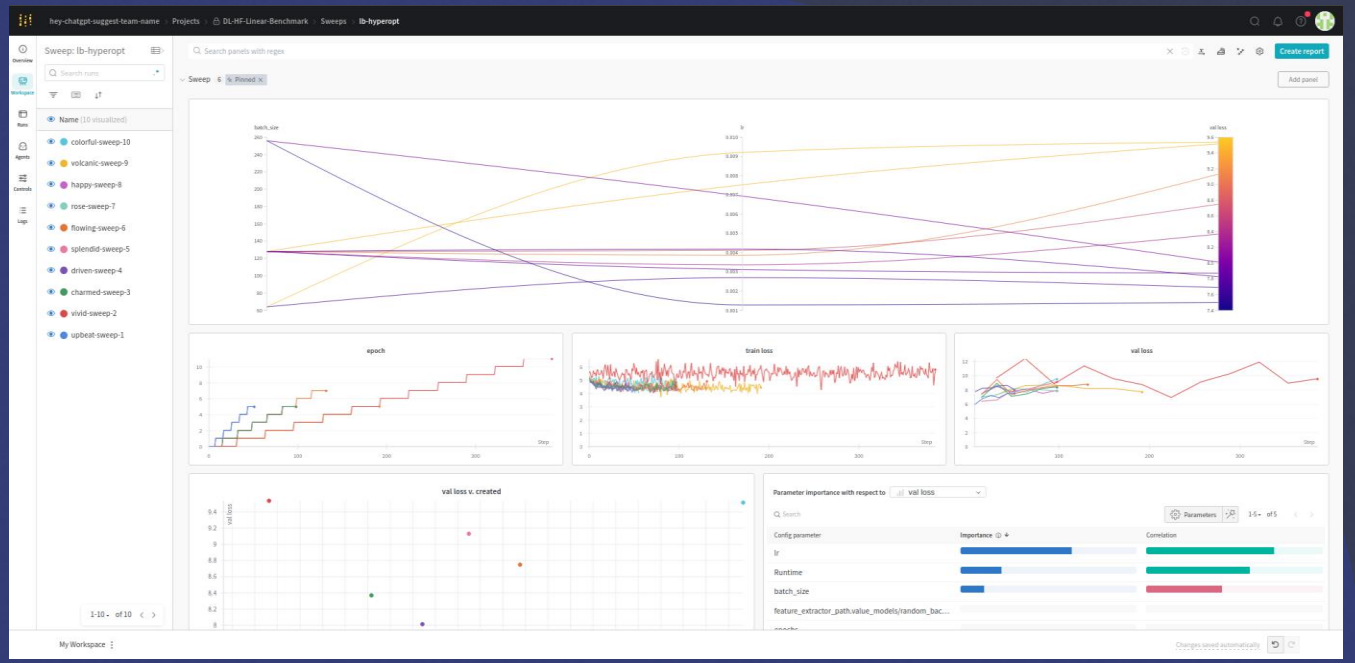
Jigsaw - 1

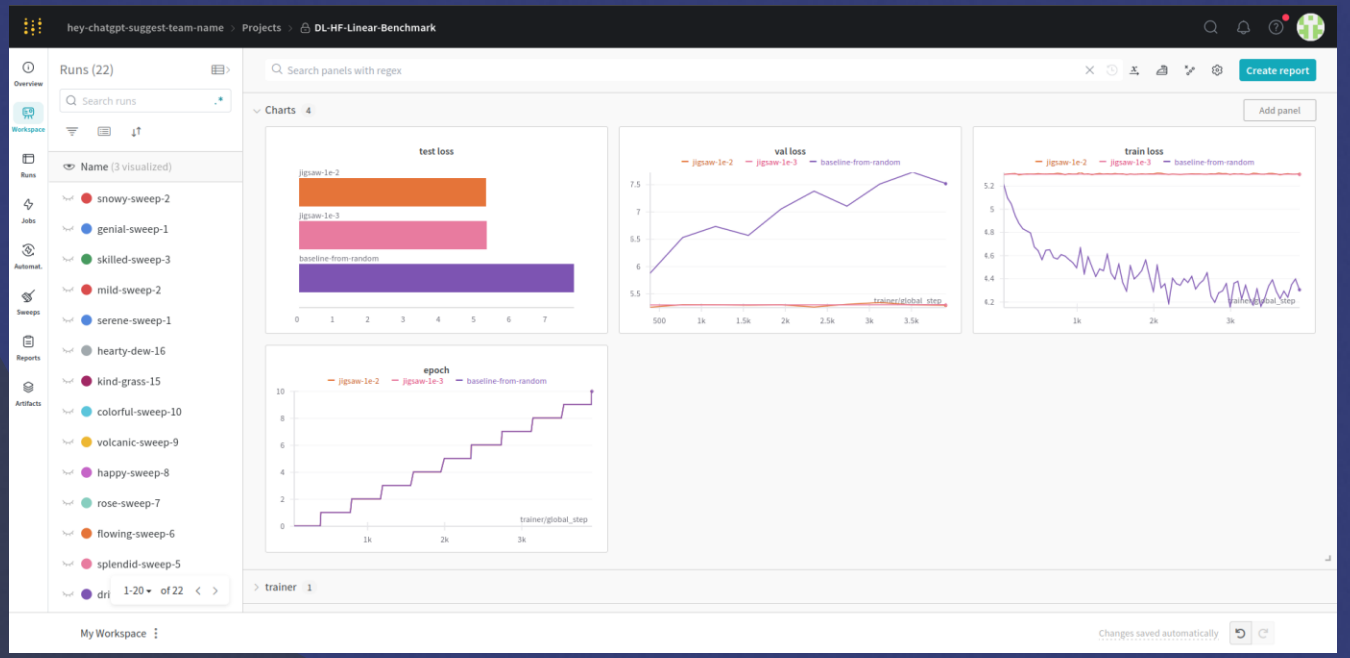


- Dividing into 3x3 tiles, random permutations
- Pretraining on non-ImageNet: Places365 [2]
- Linear benchmark evaluation:
  - Loading and freezing the pretrained weights
  - Fully connected classification layer
  - Tiny ImageNet dataset, 200 classes
- Initial experiments: VISSL (unsuccessful)

Jigsaw - 2

- Our own implementation loosely based on [5]:
  - Convolutional backbone for each tile
  - Fully connected layer per-tile
  - Global fully connected layer
  - Output: one-hot representation of a permutation
  - Additions: BN, poolings etc.
- Customizable: # of conv. filters, layer width
- Customizable: # of possible permutations, tiles (2x2 or 3x3)
- Implemented linear benchmark evaluation
- Torch, Lightning, WandB, hyperopt with Sweeps
- Heavy technical constraints due to HW limitations - POC





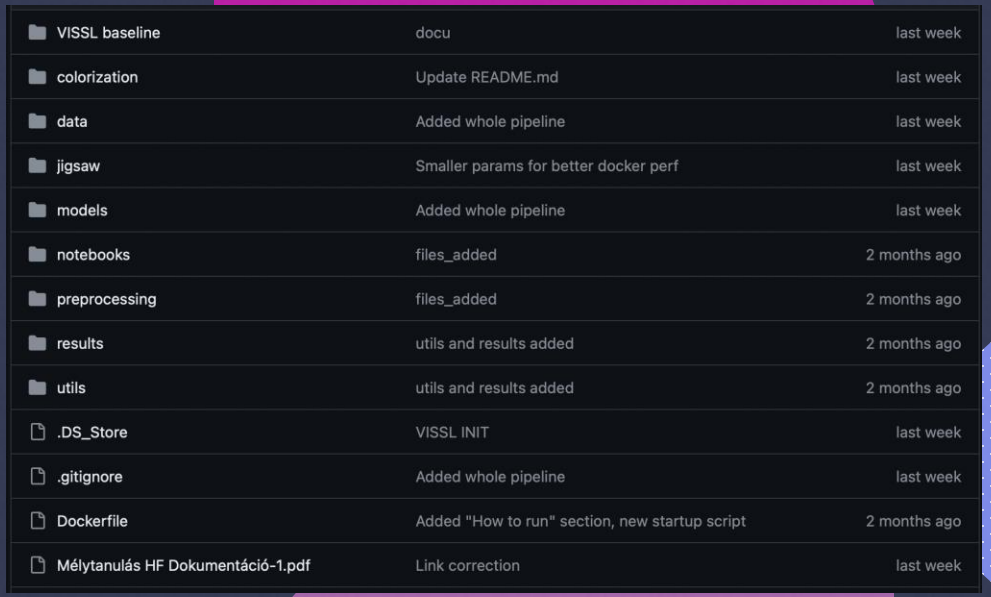
VISSL Baseline

- Facebook AI Research
- Model ZOO
- JIGSAW problem
- Configfile based setting
- Error during training

```yaml
config:
  VERBOSE: True
  LOG_FREQUENCY: 200
  TEST_ONLY: False
  TEST_EVERY_NUM_EPOCH: 1
  TEST_MODEL: True
  SEED_VALUE: 1
  MULTI_PROCESSING_METHOD: forkserver
  HOOKS:
    PERF_STATS:
      MONITOR_PERF_STATS: True
  DATA:
    NUM_DATALOADER_WORKERS: 5
    TRAIN:
      DATA_SOURCES: [data]
      LABEL_SOURCES: [data/labels]
      DATASET_NAMES: [imagenet1k_folder]
      BATCHSIZE_PER_REPLICA: 32
      TRANSFORMS:
        - name: RandomResizedCrop
          size: 224
        - name: RandomHorizontalFlip
        - name: ToTensor
        - name: Normalize
          mean: [0.485, 0.456, 0.406]
          std: [0.229, 0.224, 0.225]
      MMAP_MODE: True
      COPY_TO_LOCAL_DISK: False
      COPY_DESTINATION_DIR: /tmp/imagenet1k/
    TEST:
      DATA_SOURCES: [data]
      LABEL_SOURCES: [data/labels]
```

Structure of the project

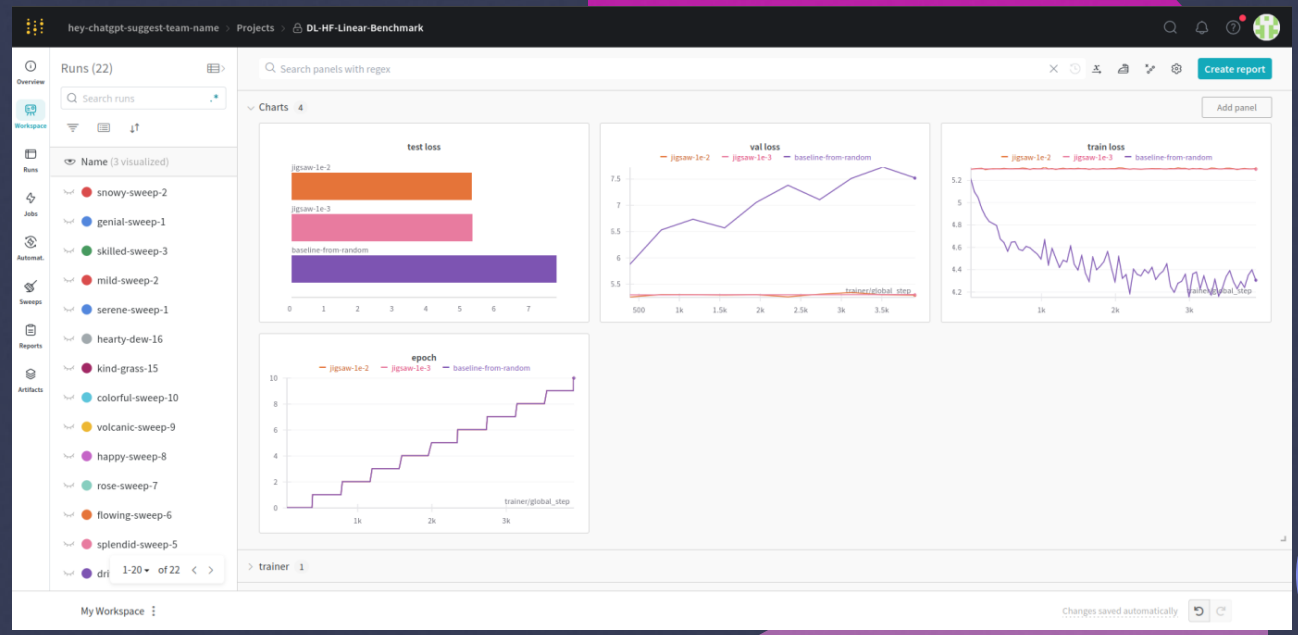
- Each project in different folders (Colorization, JIGSAW, VISSL)
  - Seperate:
    - Dataloading and pre-proccessing
    - Training
    - Evaluation
    - Environement (Docker)
- Some global features are in the root



VISSL baseline	docu	last week
colorization	Update README.md	last week
data	Added whole pipeline	last week
jigsaw	Smaller params for better docker perf	last week
models	Added whole pipeline	last week
notebooks	files_added	2 months ago
preprocessing	files_added	2 months ago
results	utils and results added	2 months ago
utils	utils and results added	2 months ago
.DS_Store	VISSL INIT	last week
.gitignore	Added whole pipeline	last week
Dockerfile	Added "How to run" section, new startup script	2 months ago
Mélytanulás HF Dokumentáció-1.pdf	Link correction	last week

Results

- Descending loss
- Proof of concept pipeline
- Trained models available for inference
- Containerization, reproducibility



Summary

- Gained experience in SSL
- Two different type of frameworks
- Future possibilities:
  - Integrating both methods
  - Concatenate both feature extractors: better "ensemble" features

Refrences

1. Zhang, R., Isola, P. and Efros, A.A., 2016. Colorful Image Colorization. Available at: arXiv:1603.08511
2. http://places.csail.mit.edu
3. Goyal, P., Mahajan, D., Gupta, A. and Misra, I., 2019. Scaling and Benchmarking Self-Supervised Visual Representation Learning.: *Available at: <https://arxiv.org/pdf/1905.01235.pdf>*
4. VISSL Team, n.d. VISSL: A library for state-of-the-art self-supervised learning from images. Available at: https://vissl.ai
5. Mehdi Noroozi, Paolo Favaro: Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. Available at: https://arxiv.org/abs/1603.09246v3