Project Machine Learning

Mina Jamshidi, Farnoush Rezaei-Jafari, Laure Ciernik, Khaled Kahouli

Machine Learning Group

October 15, 2024

Outline

1 Overview
2 Organization
3 Natural Language Processing
- Transformer Models

4 Generative Modeling
- Generative Adversarial Network (GAN)
- Vector-Quantised Variational Autoencoder (VQ-VAE)
- Flow Matching (FM)
- Diffusion Posterior Sampling for General Noisy Inverse Problems

5 Self-Supervised Learning
- Emerging Properties in Self-Supervised Vision Transformers
- SSL from Images with a Joint-Embedding Predictive Architecture

6 Few-Shot Learning
- Low-Rank Few-Shot Adaptation of Vision-Language Models

7 Computer Vision
- Attention-based Deep Multiple Instance Learning
- TransMIL: Transformer-based multiple instance learning
- U-Net: CNN for Biomedical Image Segmentation
- Set Learning for Accurate and Calibrated Models

1

Overview

Overview

Milestone 1: Prototyping

- October 15 – November 22
- Inspect the data, understand the problem, search the literature, build a prototype

Milestone 2: Model Assessment

- December 26 – January 3
- Train the model and validate it, verify the original results

Milestone 3: Experiments

- January 7 – January 31
- Additional experiments: put the original claims to the test, work out limitations and surprises

Organization

Organization

- Each milestone gives up to **10 points**, the final score is the sum.
  - Written **report** ($\leq$ 10 pages)
  - **Implementation** in Python, preferably PyTorch
- **Register** via MTS or PA **before the first submission**, each report is a Prüfungsequivalente Studienleistung.
- After each milestone, a selection of groups presents their results (15-20 minutes).
- Share codes with your supervisor on GitHub.
- Access to the computer cluster
  - Use the command `ssh-keygen` to generate a key pair on your computer.
  - Send the public key file `id_rsa.pub` as an E-Mail attachment to Dominik Kühne (**dominik.kuehne@tu-berlin.de**). The subject of the email must be "PML"
  - You will receive an E-Mail when your cluster account has been activated.
  - Follow the instructions on the **"Cheat Sheet for ML Cluster" PDF** on ISIS

Project Selection

- We have set up a fair allocation process on the ISIS page, which opens on 16.10.2024 at 23:59. Students can choose up to four topics they are interested in. Afterward, they will be assigned to a group.

Project Choice

Opens: Wednesday, 16 October 2024, 11:59 PM **Closes:** Saturday, 19 October 2024, 11:59 PM

Students are kindly asked to select one of the topics from the provided list and avoid suggesting their own. If more than three students choose the same topic, efforts will be made to prioritize selections fairly, ensuring the best possible outcome for all students.

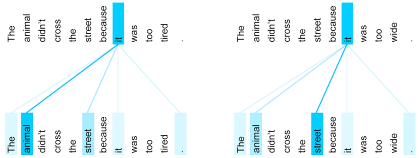
The selection process will conclude on **19.10.2024**.

Natural Language Processing

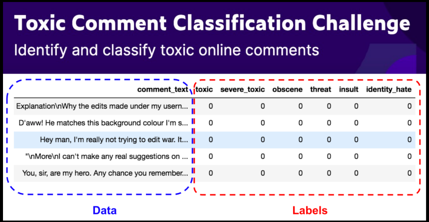
Transformer Models for Different NLP Applications

Overview

- Uses **self-attention** mechanism.
- More **efficient** and **parallelizable** across multiple GPUs.
- Can better capture **long-term dependencies.**



The encoder self-attention distribution for the word "it" from the 5th to the 6th layer of a Transformer trained on English to French translation (one of eight attention heads).



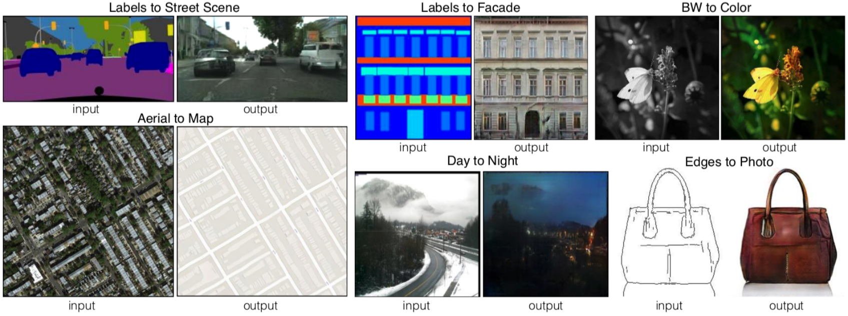
comment_text	toxic	severe_toxic	obscene	threat	insult	identity_hate
Explanation\nWhy the edits made under my usern...	0	0	0	0	0	0
D'aww! He matches this background colour I'm s...	0	0	0	0	0	0
Hey man, I'm really not trying to edit war. It...	0	0	0	0	0	0
"\nMore\nI can't make any real suggestions on ...	0	0	0	0	0	0
You, sir, are my hero. Any chance you remember...	0	0	0	0	0	0

Generative Modeling

Generative Adversarial Network (GAN)

Overview

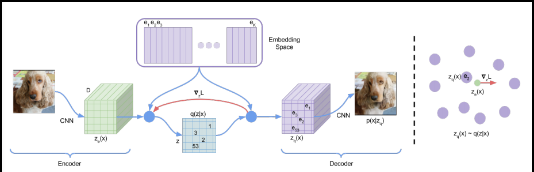
- It is composed of two parts: **Generator** and **Discriminator**
- The generator network learns to generate data.
- The discriminator network learns to discriminate fake from real.

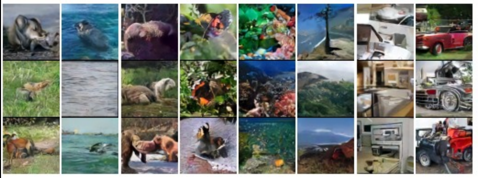


Vector-Quantised Variational Autoencoder (VQ-VAE)

Overview

- Differs from VAEs in two crucial aspects.
  - **Discrete** codes instead of continuous codes.
  - The prior is **learnt** rather than static.



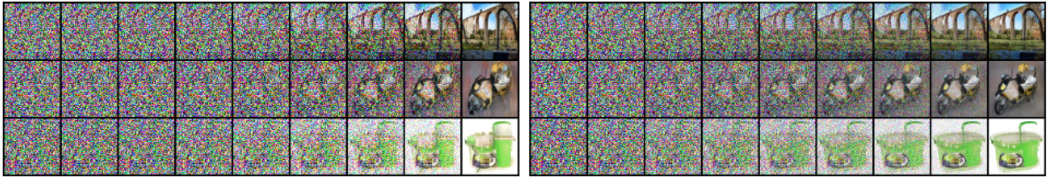


Flow Matching for Generative Modeling

Lipman et al. (ICLR 2023)

Overview

- FM combines aspects from **Continuous Normalising Flows** and **Diffusion Models**:
  - Simulation-free way to train CNF models at unprecedented scale.
  - More robust and stable alternative for training diffusion models.

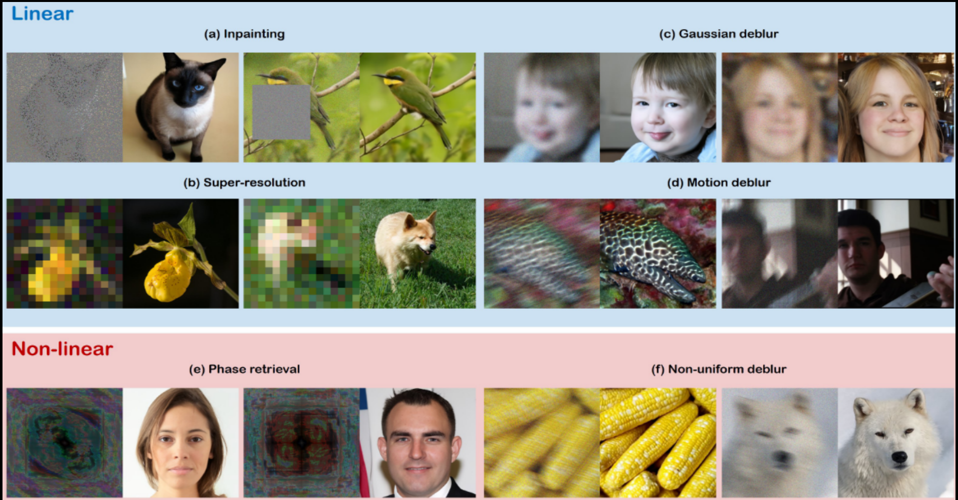


Flow Matching w/ Diffusion Flow Matching w/ OT

Diffusion Posterior Sampling for General Noisy Inverse Problems

Chung et al. (ICLR 2023)

- Diffusion models to approximate posterior sampling for **(Non)linear** inverse problems.



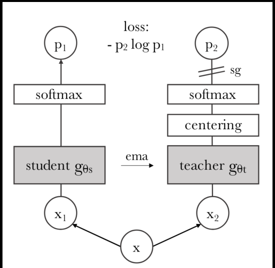
Self-Supervised Learning

Emerging Properties in Self-Supervised Vision Transformers

Caron et al., ICCV 2021

Overview

- A self-supervised learning method for vision transformers/ CNNs that uses a teacher-student architecture to learn visual representations without labels.
- Pertaining on ImageNet1k or smaller datasets with heavy data augmentations.

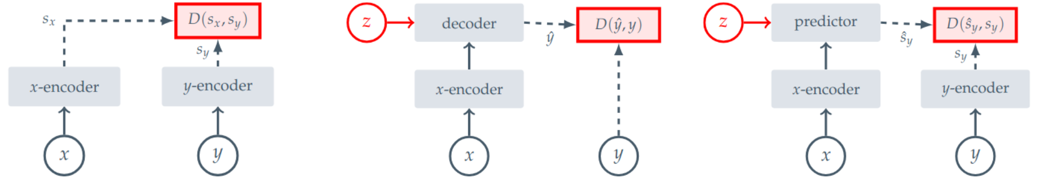


SSL from Images with a Joint-Embedding Predictive Architecture

Assran et al. (CVPR 2023)

Overview

- Semantic image representations without relying on hand-crafted data augmentation.
- Generative architecture in the latent space.



(a) **Joint-Embedding Architecture** (b) **Generative Architecture** (c) **Joint-Embedding Predictive Architecture**

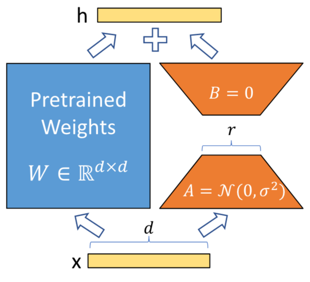
Few-Shot Learning

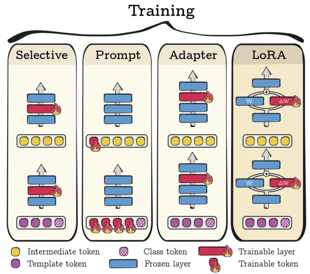
Low-Rank Few-Shot Adaptation of Vision-Language Models

Zanella, Ben Ayed, CVPR 2024

Overview

- Application of low-rank updates to the attention matrices of both the vision and text encoders in CLIP, allowing for efficient fine-tuning with limited data.
- Outperforms existing prompt-learning and adapter-based methods across 11 datasets
- Uses fixed hyperparameters across tasks





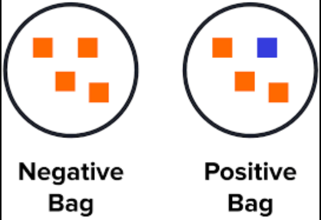
Computer Vision

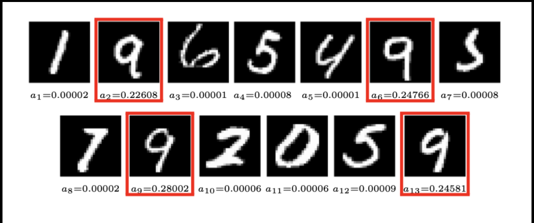
Attention-based Deep Multiple Instance Learning

Ilse, Tomczak, and Welling, PMLR 2018

Overview

- What is Multiple instance learning (MIL)?
- MIL with neural network, incorporating attention mechanism



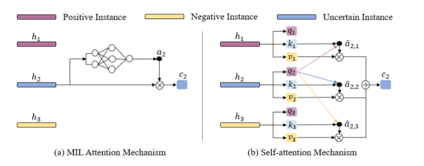


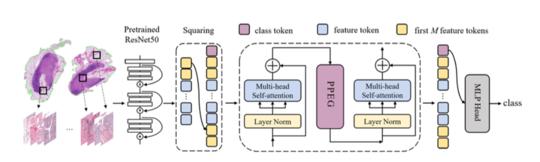
TransMIL: Transformer-based multiple instance learning

Shao et al, NeurIPS 2021

Overview

- using a Transformer as an aggregator in multiple instance learning
- data: histopathology whole slide images



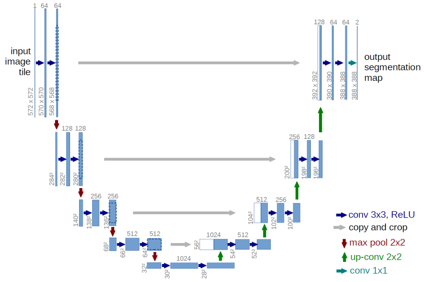


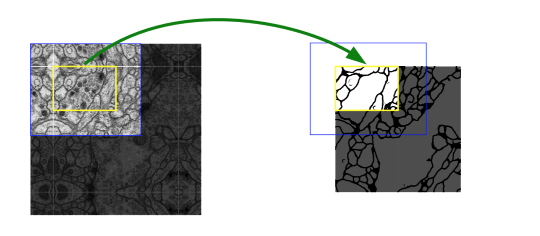
U-Net: Convolutional Network for Biomedical Image Segmentation

Ronneberger, Fischer, Brox, MICCAI 2015

Overview

- A U-shaped convolutional architecture for **image segmentation** and **translation**
- Requires **few training images**, learns invariances via **data augmentation**
- Reaches unprecedented performance in **localization** and **contextualization**





Set Learning for Accurate and Calibrated Models

Muttenthaler et al., ICLR 2024

Overview

- Supervised training objective based on sets of samples instead of single ones to achieve better-calibrated classifiers
- Method improves classification in heavy-tail settings.
- Data: CIFAR-10, CIFAR-100