

Beyond Deep Learning: Selected Topics

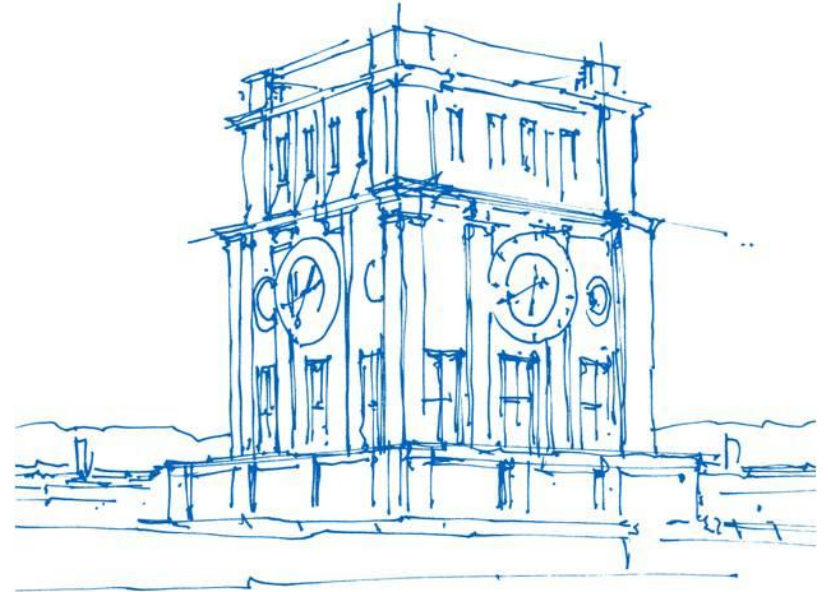
Felix Wimbauer

Technical University of Munich

Chair of Computer Vision and Artificial Intelligence

Garching,

July, 11th, 2024



Uhrenturm der TUM

Agenda

- Who am I
- What are the topics we will cover?
- How is the course organized?
- How to apply?

Felix Wimbauer

Background:

- 3rd year PhD student
- TUM, University of Oxford

Research Interests:

- (Dynamic) 3D Reconstruction, Object-centric learning, Diffusion Models, Bayesian Approaches, MCMC

Website:

vision.in.tum.de/members/wimbauer



Topics

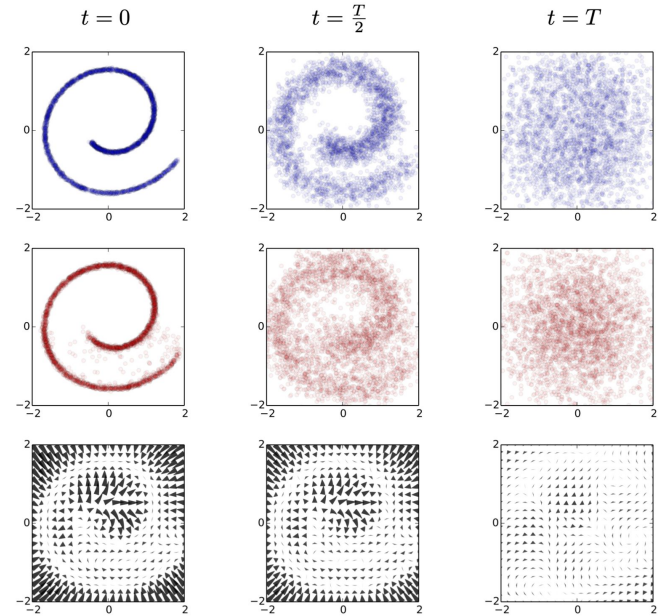
Diffusion Models



The forward trajectory
 $q(\mathbf{x}_{0:T})$

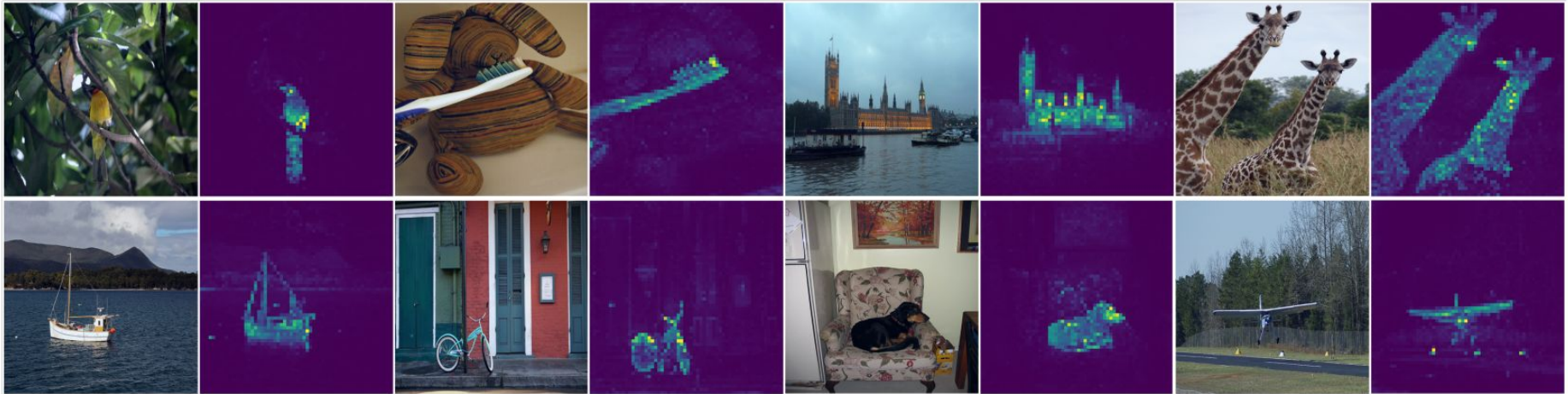
The reverse trajectory
 $p_{\theta}(\mathbf{x}_{0:T})$

The drifting term
 $\mu_{\theta}(\mathbf{x}_t, t) - \mathbf{x}_t$



- Dall E 2, Ramesh et al, 2022
- Dhariwal et al 2021
- ...

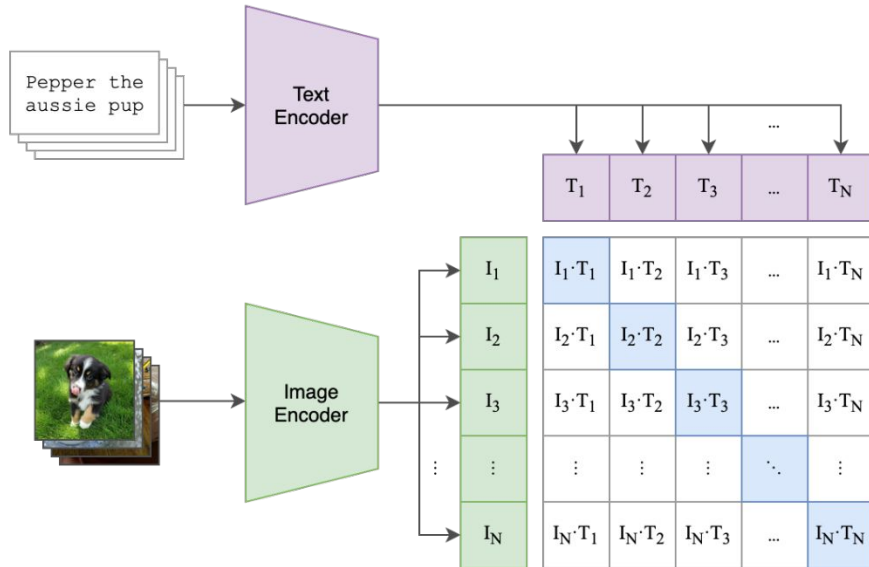
DINO - Student-teacher models for self-supervised rep. learning



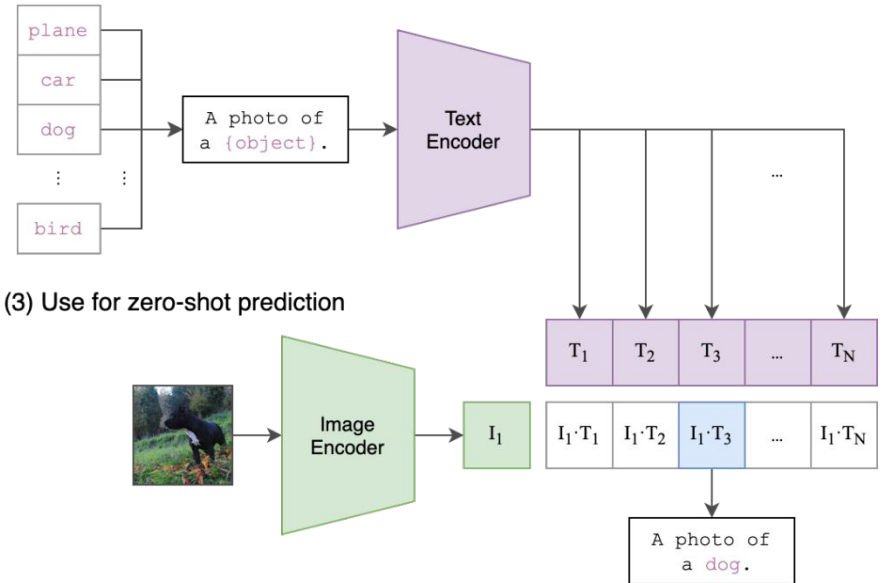
- DINO, Caron et al 2021
- ...

CLIP - Representation Learning for Text and Images

(1) Contrastive pre-training



(2) Create dataset classifier from label text



(3) Use for zero-shot prediction

- Clip, Radford et al., 2021
- SigLIP, Zhai et al., 2023
- ...

Multimodal Language Models

Mobile Manipulation



Human: Bring me the rice chips from the drawer. Robot: 1. Go to the drawers, 2. Open top drawer. I see . 3. Pick the green rice chip bag from the drawer and place it on the counter.

Visual Q&A, Captioning ...



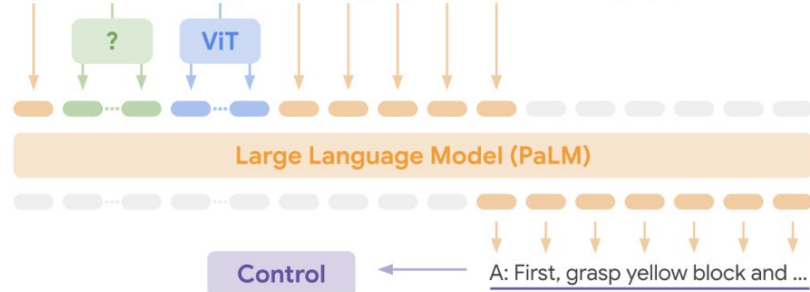
Given . Q: What's in the image? Answer in emojis.
 A: 🍏 🍇 🍌 🍎 🍓



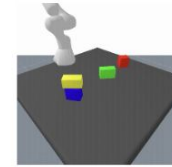
Describe the following :
 A dog jumping over a hurdle at a dog show.

PaLM-E: An Embodied Multimodal Language Model

Given ... Q: How to grasp blue block? A: First, grasp yellow block



Task and Motion Planning



Given Q: How to grasp blue block?
 A: First grasp yellow block and place it on the table, then grasp the blue block.

Tabletop Manipulation



Given Task: Sort colors into corners.
 Step 1. Push the green star to the bottom left.
 Step 2. Push the green circle to the green star.

Language Only Tasks

Q: Miami Beach borders which ocean? A: Atlantic. Q: What is 372 x 18? A: 6696. Q: Write a Haiku about embodied LLMs. A: Embodied language. Models learn to understand. The world around them.

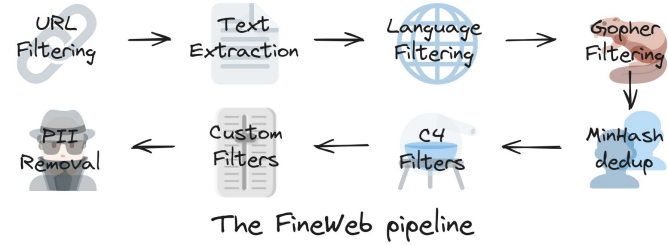
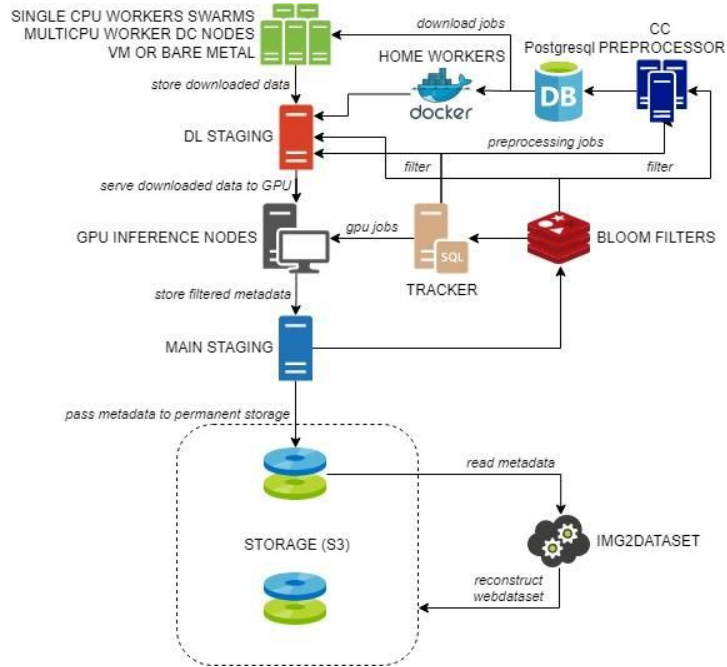
- PaLM-E, Driess et al., 2023
- Many more

Segment Anything and Follow-Ups



- Segment Anything, Kirillov et al., 2023
- GroundedSAM, UnSAM
- Many more

Datasets and Dataset Curation



Hugging Face



FineWeb

Course logistics

Course Organization

Course website: <https://cvg.cit.tum.de/teaching/ws2024/bdl>

Course email (for now): felix.wimbauer@tum.de

Course structure:

- Kick-Off Meeting with all the topics (default date: Oct 16th)
- Matching to the topics
- Read the papers and do a literature search and elaborate on the topic you are provided with
- Get optional help, if you did not understand the paper
- Send a first draft of the presentation and get optional feedback
- Presentations take place mid January
- Final report will be due one week after the presentations

Prerequisites

- Machine learning & deep learning knowledge:
Basic ML concepts and ML/DL models
Min. Requirement: passed one ML/DL related course (I2ML, I2DL, ADL4CV, PGM ...)
- Soft skills:
Manage regular workflow and communicate with tutors efficiently
- We also value:
 - solid basis & interest for maths
 - prior experience with ML/DL projects

How to apply

1. Apply via the **TUM Matching system** (until July 16th, 2023)
 - If you like our course, make sure to give it a high priority :)
 2. **Send us an email** to show your interest and fulfillment of prerequisites
 - Crucial for us to give you a priority
- The email should be sent to us **latest on July 16** with the title “[BDL] <Firstname> <Lastname>” and contain
 - Filled information form (template on course website, rename to “firstname_lastname.xlsx”)
 - Transcript
 - CV
 - Course Website: <https://cvg.cit.tum.de/teaching/ws2024/bdl>

Thank you! Questions?

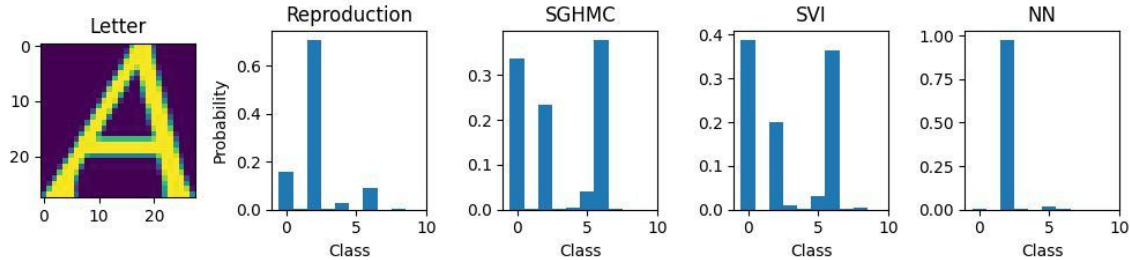
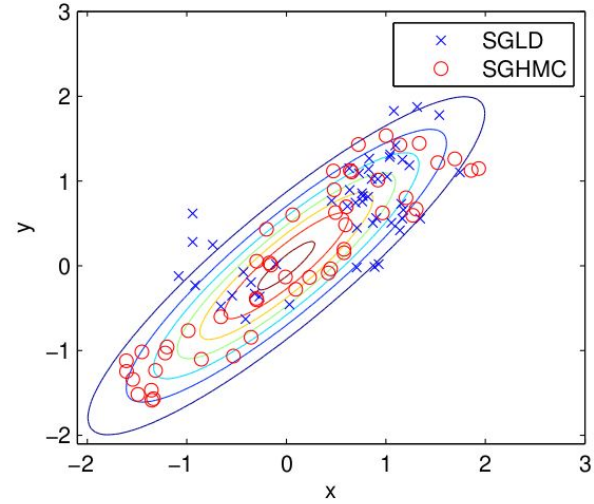
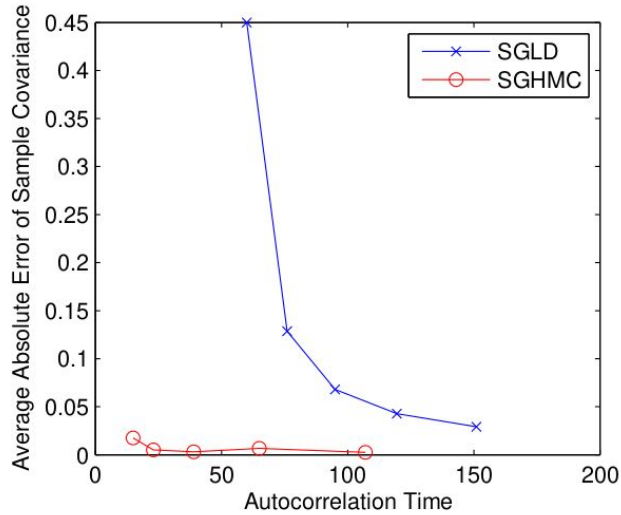


3D Aware Generative Models



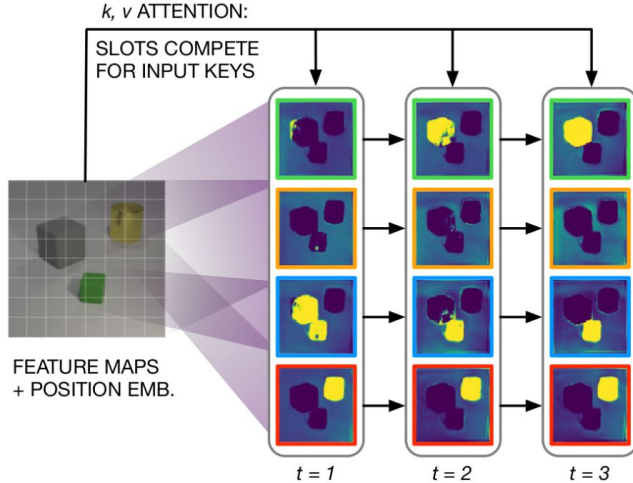
- EG3D, Chan et al 2022
- PiGAN, Chan et al 2021

Markov Chain Monte Carlo + Optimization

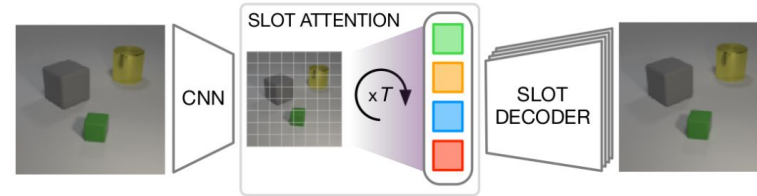


- SGHMC, Chen et al, 2014
- SGLD, Welling et al, 2011

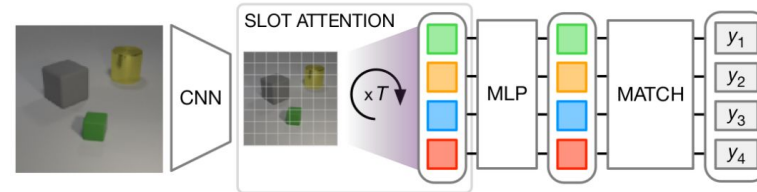
Object-centric learning with slot attention



(a) Slot Attention module.



(b) Object discovery architecture.



(c) Set prediction architecture.

- Locatello et al 2020
- Kipf et al 2021