UNIVERSITÄT BONN AIS

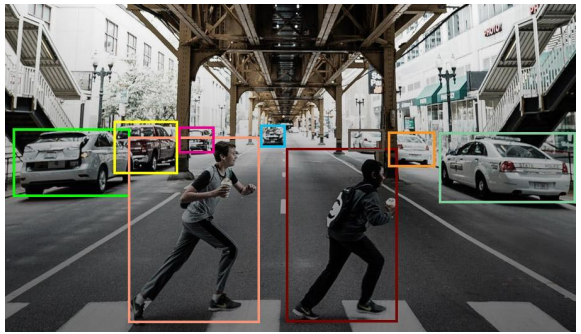
# Seminar Vision Systems MA-INF 4208

03.02.2023

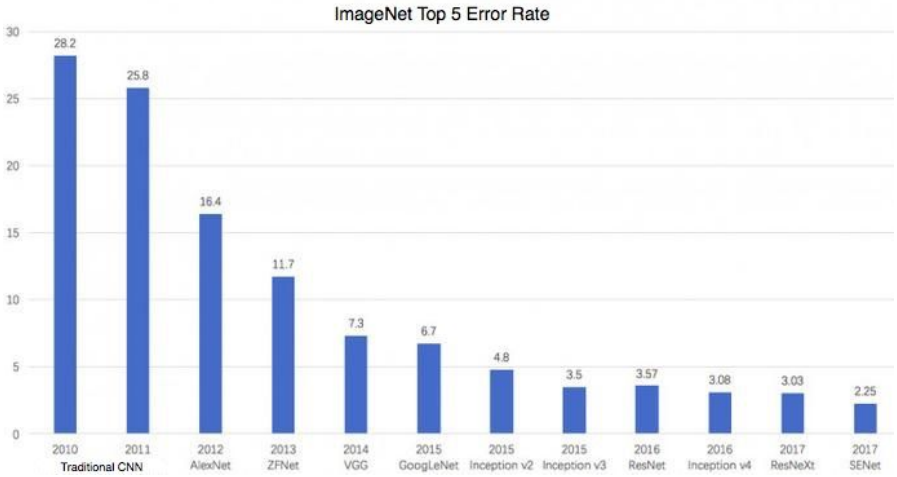
PROF. SVEN BEHNKE, ANGEL VILLAR-CORRALES

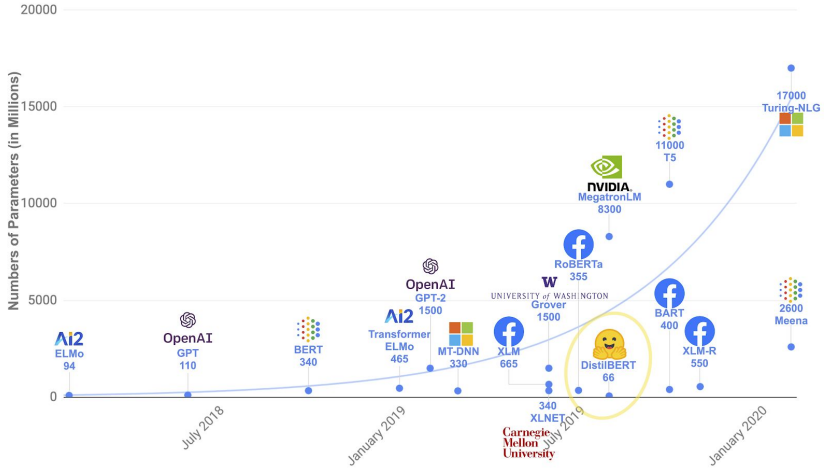
Contact: villar@ais.uni-bonn.de

# The Age of Deep Learning



# The Age of Deep Learning

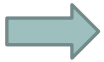


# The Age of Deep Learning

# In this seminar...

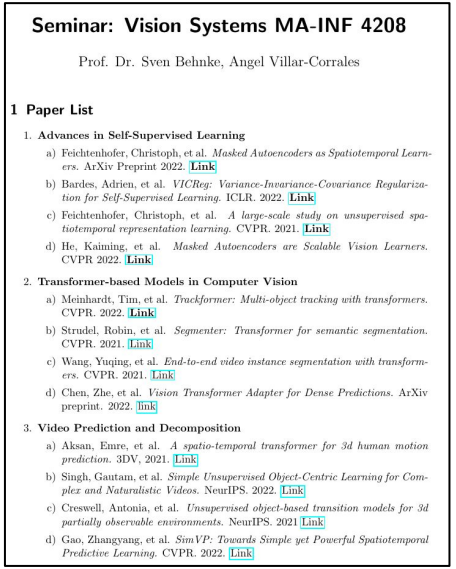
- Acquire/improve ability to:
  - deal with scientific publications (e.g., papers)
  - write a scientific report
  - present a scientific topic to an audience
  - engage technical topics



Important skills for Master Thesis!

# In this seminar

- Discuss trending topics in deep learning and computer vision
- We will cover the following topics
  - Self-Supervised Learning
  - Transformers in Computer Vision
  - Unsupervised Video Prediction and Decomposition



**Seminar: Vision Systems MA-INF 4208**

Prof. Dr. Sven Behnke, Angel Villar-Corrales

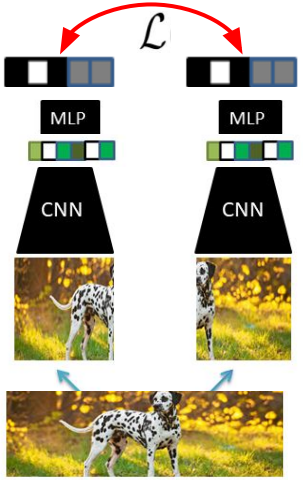
**1 Paper List**

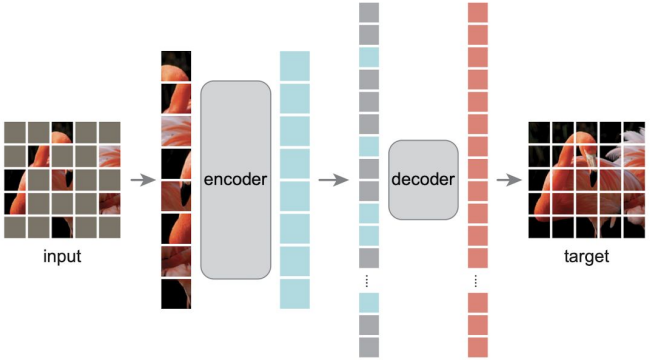
1. **Advances in Self-Supervised Learning**
   a) Feichtenhofer, Christoph, et al. *Masked Autoencoders as Spatiotemporal Learners.* ArXiv Preprint 2022. Link
   b) Bardes, Adrien, et al. *VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning.* ICLR. 2022. Link
   c) Feichtenhofer, Christoph, et al. *A large-scale study on unsupervised spatiotemporal representation learning.* CVPR. 2021. Link
   d) He, Kaiming, et al. *Masked Autoencoders are Scalable Vision Learners.* CVPR 2022. Link
2. **Transformer-based Models in Computer Vision**
   a) Meinhardt, Tim, et al. *Trackformer: Multi-object tracking with transformers.* CVPR. 2022. Link
   b) Strudel, Robin, et al. *Segmenter: Transformer for semantic segmentation.* CVPR. 2021. Link
   c) Wang, Yuqing, et al. *End-to-end video instance segmentation with transformers.* CVPR. 2021. Link
   d) Chen, Zhe, et al. *Vision Transformer Adapter for Dense Predictions.* ArXiv preprint. 2022. link
3. **Video Prediction and Decomposition**
   a) Aksan, Emre, et al. *A spatio-temporal transformer for 3d human motion prediction.* 3DV, 2021. Link
   b) Singh, Gautam, et al. *Simple Unsupervised Object-Centric Learning for Complex and Naturalistic Videos.* NeurIPS. 2022. Link
   c) Creswell, Antonia, et al. *Unsupervised object-based transition models for 3d partially observable environments.* NeurIPS. 2021 Link
   d) Gao, Zhangyang, et al. *SimVP: Towards Simple yet Powerful Spatiotemporal Predictive Learning.* CVPR. 2022. Link

**Paper List**: https://www.ais.uni-bonn.de/WS2223/SeminarVision/PaperList.pdf

# Self-Supervised Learning

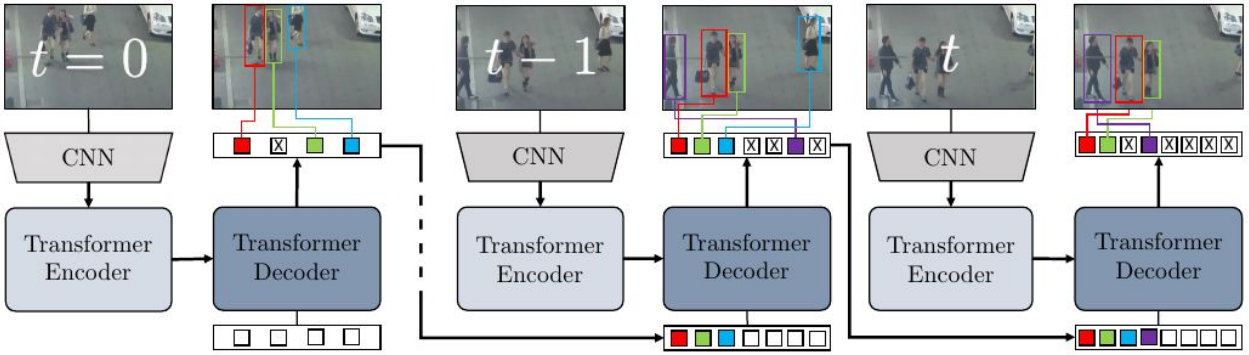
- Subcategory of unsupervised learning
- Use pretext task to train in a supervised fashion
- Hot-topic in deep learning community
  - Outperforms to supervised pretraining
  - No need for manual annotations



# Transformers in Computer Vision

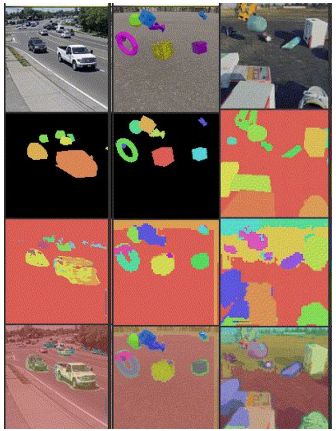
- Transformer-based architectures for computer vision tasks
  - Semantic and instance segmentation
  - Multi-object tracking
- State-of-the-art across many benchmarks

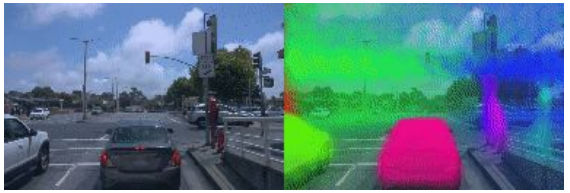


# Video Prediction and Decomposition

- **Video prediction**: given C seed frames, predict next N plausible frames
  - Useful for decision making for autonomous agents or for learning representations

- **Video decomposition**: decomposing a video sequence into object-centric components in an unsupervised manner





# Select your topic

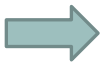
- Send me an email at villar@ais.uni-bonn.de
  - Your name
  - Matriculation number
  - Your two preferred topics/papers

- Upon my confirmation: Register in BASIS

# Deliverables (preliminary dates)

- Presentation: Thursday 24.03.2023
  - 30 min presentation
  - 15 min discussion

- Report: Thursday 31.03.2023
  - LaTeX template
  - 8-12 pages
  - Brief but readable and informative
  - BibTex citations



Arrange a meeting with me ≈2 weeks before the presentation to check the preliminary materials for the presentation and report.

# Report

- Well structured:
  - Abstract
  - Introduction, methods, results, conclusion, ...
  - Tables and figures
  - Correct citations

- Your own scientific opinion:
  - What are the weak points of the paper?
  - What is missing?
  - Are comparisons fair and believable?
  - Possible future steps?

We don’t want a copy of the paper!

# Grading

- 60%: Presentation
  - Quality of the presentation slides
  - Presentation skills
  - Ability to answer questions

- 40%: Report
  - Overall quality of the report
  - Critical thinking and own discussion
  - Understanding of the concept

# Slot Assignment Selection

- Six slots for students
  - Assigned at random

# Questions?