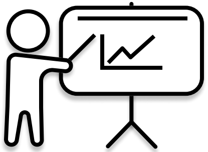
Deep and Generative Processing and Coding of Visual Data
Kick-Off Meeting

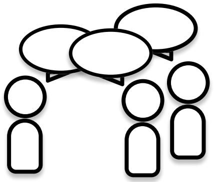
Ferienakademie 2026 – Course 7

TUM FAU LMS University of Stuttgart Germany

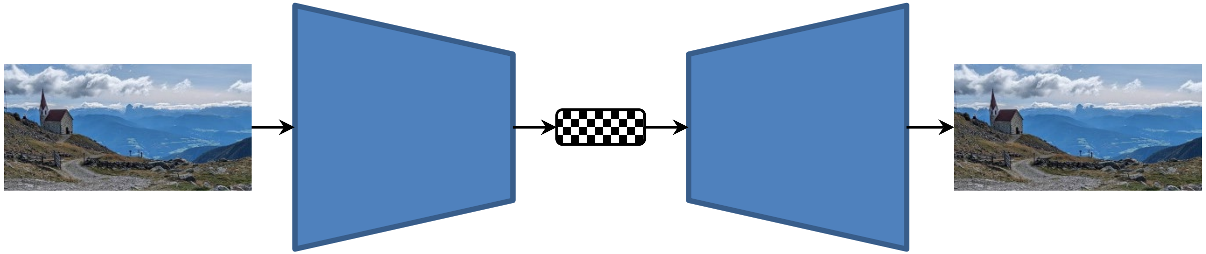
Course Organization

- Theoretical part
  - Presentations on selected topics
  - 45 min slots
    - ➢ 30 min presentation
    - ➢ 15 min discussion





- Practical part
  - Comparison of different image codecs and processing tools
  - Self-captured test sequences



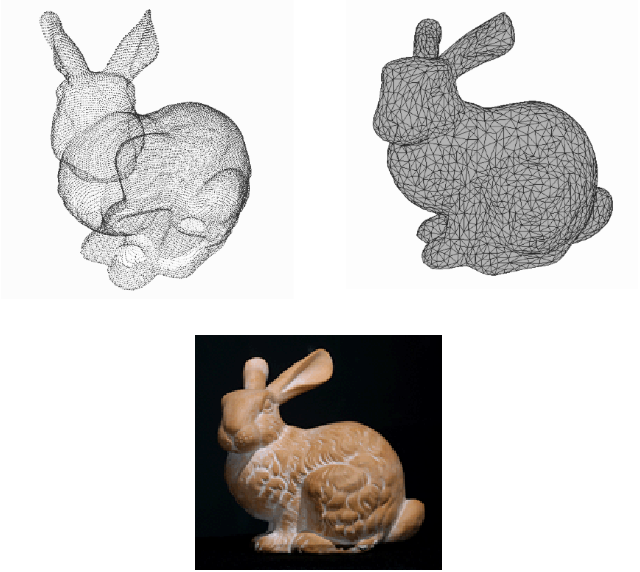
University of Stuttgart
Germany

Topic List

1. Visual Data Representation and Quality Assessment Metrics
2. Deep Learning: Basics
3. Deep Learning: Training and Risks
4. Self-Supervised and Contrastive Learning
5. Transfer Learning
6. Reinforcement Learning
7. Physics-Informed Neural Networks
8. Gaussian Splatting
9. Implicit Neural Representations
10. Object Detection and Classification in the Pixel-domain
11. Image Processing with Latent Representations
12. Depth Perception and Estimation
13. Human Action Recognition
14. Optimal Transport for Generative Models
15. Deep Generative Models
16. Deep Fakes
17. Inverse Problems
18. Learning-based Image Compression
19. Learning-based Video Compression
20. Semantic and Generative Coding for Visual Compression
21. Processing and Coding of 360°-Videos
22. Perception-based Quality Evaluation
23. Processing and Compression of Astronomical Data

1 – Visual Data Representation and Quality Assessment Metrics

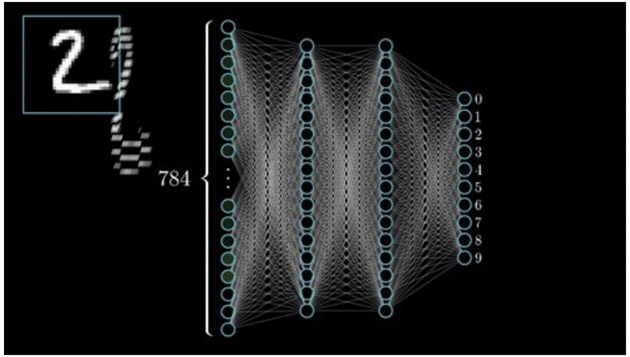
- Mathematical foundation for visual data
  - Image, video, point cloud, multispectral, …
  - Statistical properties: manifold hypothesis
  - Natural vs. synthetic data
  - Frequency transformations
- Coding of visual data
  - Rate-distortion theory
  - Classical evaluation metrics (PSNR, SSIM)
  - Evaluation metrics for other data formats



stanford.edu/~rqi/pointnet/docs/cvpr17_pointnet_slides.pdf

2 – Deep Learning: Basics

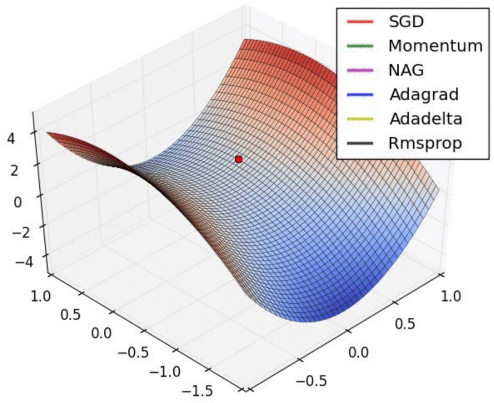
- Basis of most state-of-the-art computer vision systems
- Basics of neural networks
  - Neurons and activations
  - Essential layers: fully-connected, convolutional, pooling, …
  - Training using gradient-descent and backpropagation



youtube.com/watch?v=aircAruvnKk&t=112s

3 – Deep Learning: Training and Risks

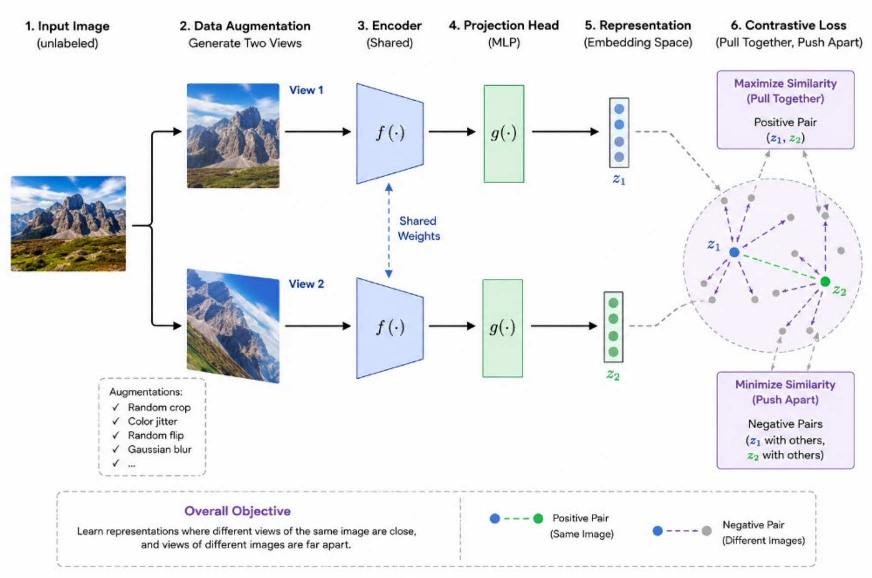
- Key principles of training Deep Neural Networks
  - Supervised, semi-supervised, unsupervised
  - Optimization algorithms
  - Batch-normalization, dropout, learning rate schedules, …
- Risks of application in real-world systems
  - Class imbalance, uncertainty, adversarial attacks



ruder.io/optimizing-gradient-descent/index.html

4 – Self-Supervised and Contrastive Learning

- Learning useful representations from unlabeled data
- Contrastive learning: pulling similar samples closer and pushing dissimilar samples apart
- Representatives: SimCLR, CPC, DINO, MAE
- Applications in image/video understanding, detection, segmentation, denoising, and super-resolution



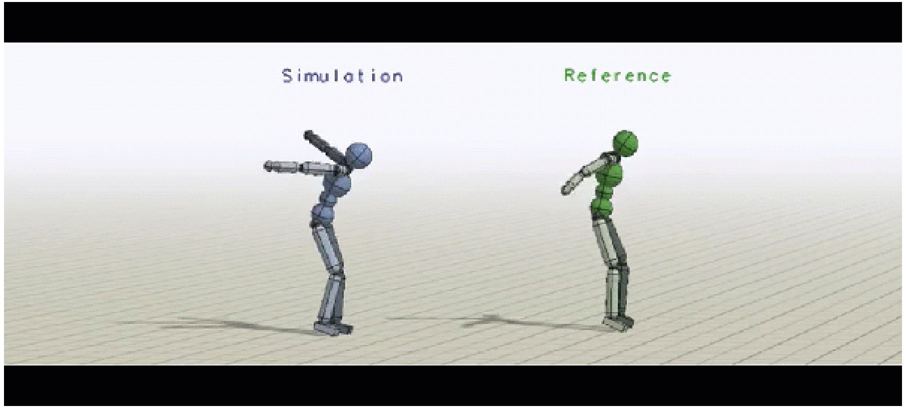
5 – Transfer Learning

- Reuse knowledge from a pre-trained model for a related new task
- Fine-tuning and parameter-efficient fine-tuning (LoRA)
- Reduced training time and data requirements
- Few-shot learning, zero-shot learning, and domain adaptation
- Related concepts: knowledge distillation and multi-task learning

serokell.io/blog/guide-to-transfer-learning

6 – Reinforcement Learning

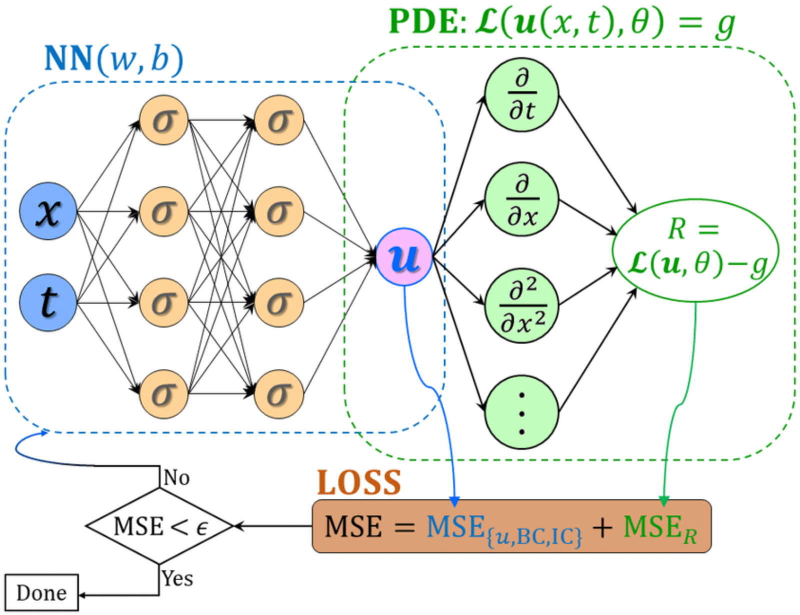
- Learn decision-making through interaction with an environment
- Agent, environment, state, action, reward, and policy
- Rewards guide the training process
- Sequential decision making modeled by MDPs/POMDPs
- Representative methods: Q-learning, DQN, REINFORCE, PPO, A2C



Peng, X. B., Abbeel, P., Levine, S., & Van de Panne, M. (2018). Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, *37*(4), 1-14.

7 – Physics-Informed Neural Networks

- Integrate physical laws into neural network training
- Use governing equations, conservation laws, or PDEs as constraints
- Combine data-driven learning with physics-based supervision
- Applications in smoke, water, fire, cloth, and deformable object simulation
- Improve physical consistency, but may face high computational cost and optimization difficulty

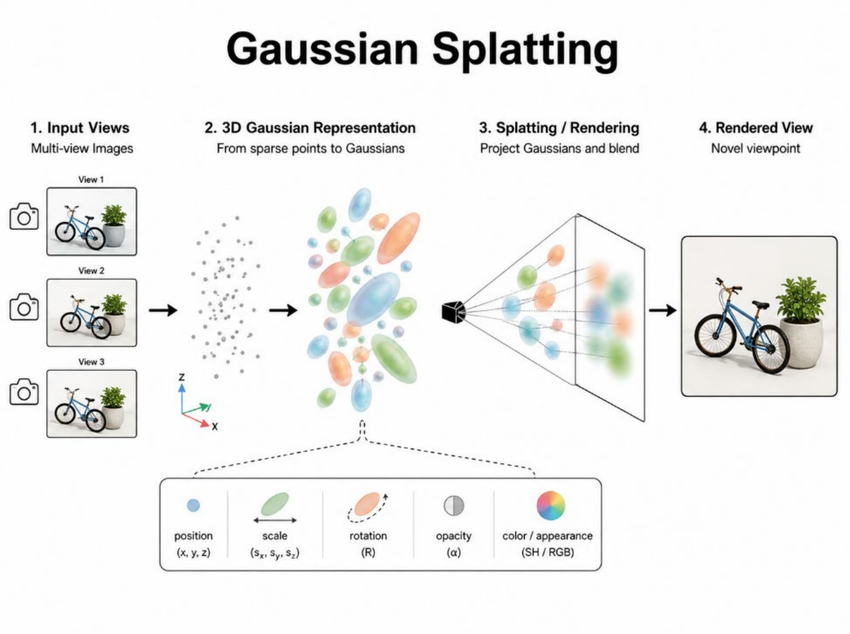


Meng, X., Li, Z., Zhang, D., & Karniadakis, G. E. (2020). PPINN: Parareal physics-informed neural network for time-dependent PDEs. *Computer Methods in Applied Mechanics and Engineering*, *370*, 113250.

University of Stuttgart
Germany

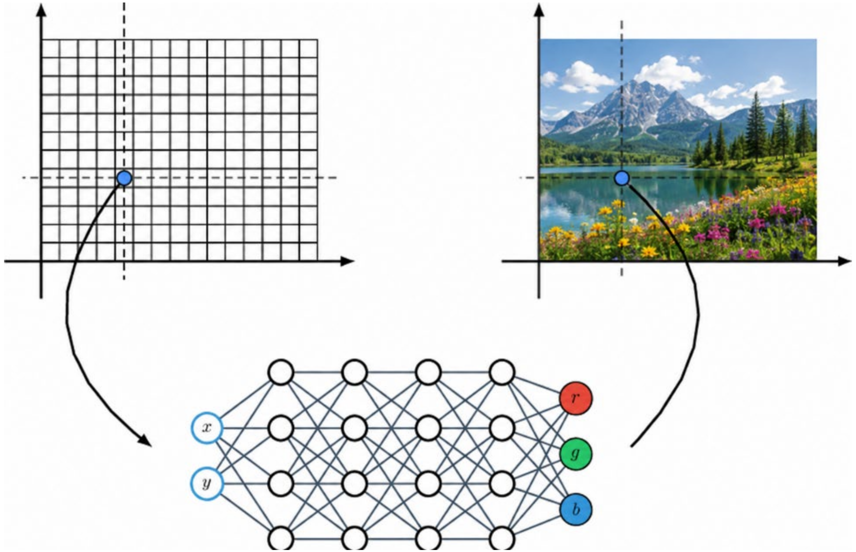
8 – Gaussian Splatting

- Explicit 3D scene representation using learnable Gaussian primitives
- Each Gaussian stores position, scale, rotation, opacity, and appearance information
- Efficient differentiable rendering for real-time novel-view synthesis
- Extensions to dynamic/video scenes and compression of Gaussian parameters



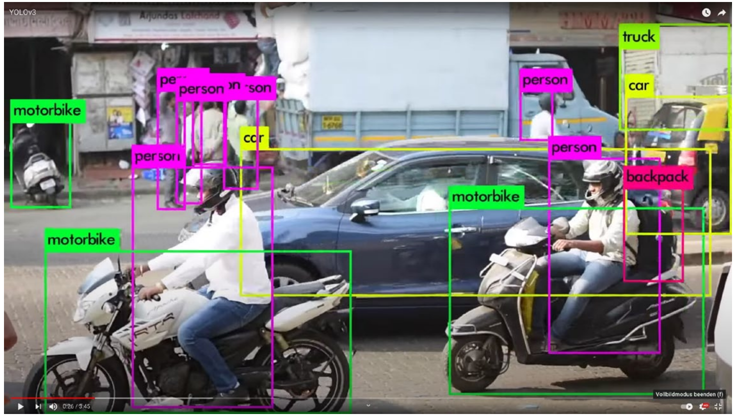
9 – Implicit Neural Representations

- Represent signals as continuous functions
- Map input coordinates to signal values
- Resolution-independent representation for images, audio, videos, 3D shapes, and radiance fields
- Representative methods:
  - SIREN
  - Cool-Chic



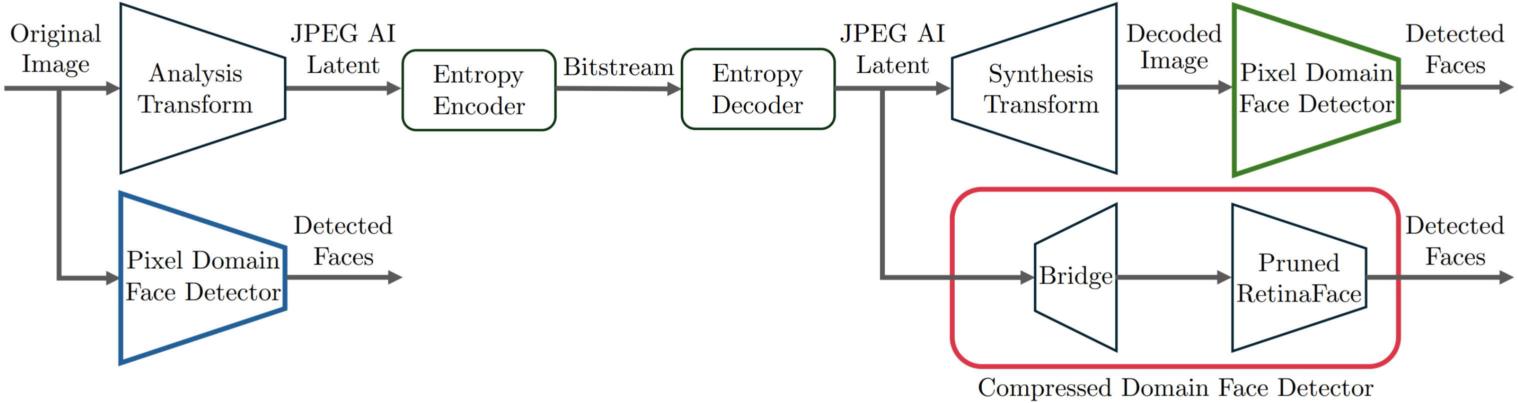
10 – Object Detection and Classification in the Pixel-domain

- Detecting and classifying objects
- Introduce and explain common object detection networks
  - R-CNN, Fast R-CNN, Faster R-CNN
  - YOLO
- Object tracking methodologies
  - SORT, DeepSORT
  - Tracking-by-detection



ruder.io/youtube.com/watch?v=MPU2HistivI&t=18s/index.html

11 – Image Processing with Latent Representations

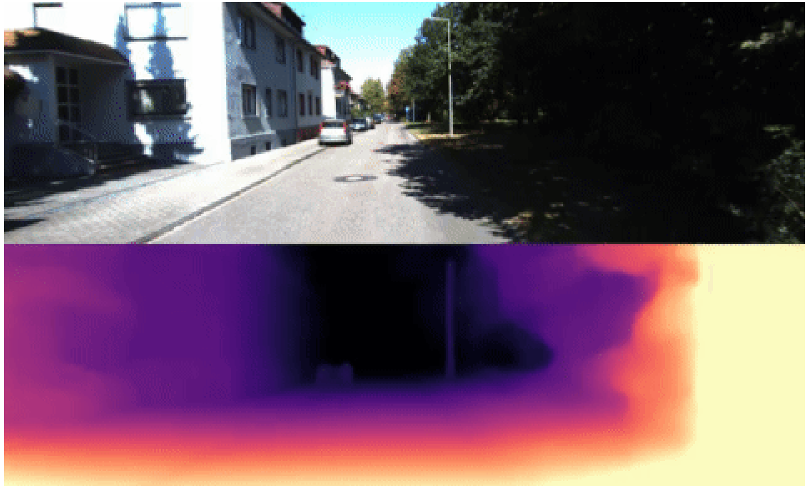


Alkhateeb, A. & Pereira, F. (2024, October). Jpeg ai compressed domain face detection. In 2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.

- Transform images into latent feature representation instead of operating directly on raw pixels
- Latent representations capture semantic, structural, or textural information
- Latent-space processing can improve efficiency, robustness, and generalization
- Applications: texture prediction, image classification, face detection, and super-resolution

12 – Depth Perception and Estimation

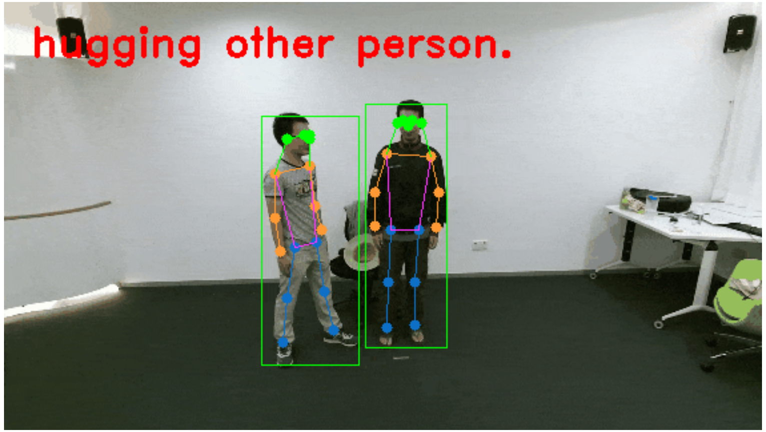
- Understand 3D scene structure and object positions in the environment
- Hardware-based depth sensing: LiDAR, radar, time-of-flight, and structured-light cameras
- Multi-camera depth estimation using stereo matching and geometric triangulation
- Monocular depth estimation and sensor fusion for robust real-world perception



github.com/nianticlabs/monodepth2

13 – Human Action Recognition

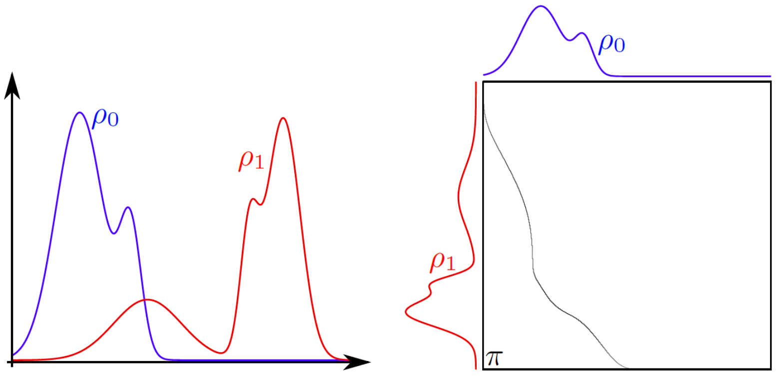
- Recognize and understand human activities from videos, images, or sensor data
- Model both spatial appearance and temporal motion information
- Action localization: temporal start/end detection and spatial action localization
- Skeleton-based recognition with GCNs, and challenges such as occlusion, viewpoint variation, and untrimmed videos



github.com/open-mmlab/mmaction2

14 – Optimal Transport for Generative Models

- Mathematical foundation for Generative Models
- Distance metrics between probability density functions
  - Kullback-Leibler divergence
  - Wasserstein distance
- Connection to generative models

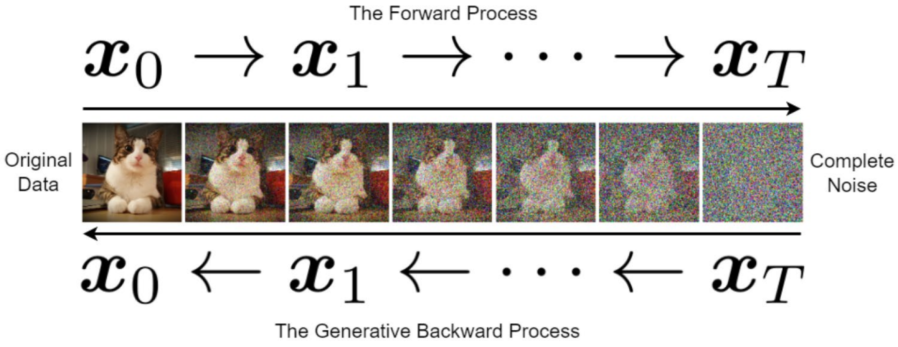


(a) Source and target (b) Transport map

Optimal Transport on Discrete Domains, Solomon

15 – Deep Generative Models

- Overview over generative models for visual data
  - Generative adversarial networks
  - Diffusion models
  - Image-to-image translation
  - Text-to-image models
- Introduce fundamental concepts and training procedures



Diffusion Models: Tutorial and Survey, Ghojogh et al.

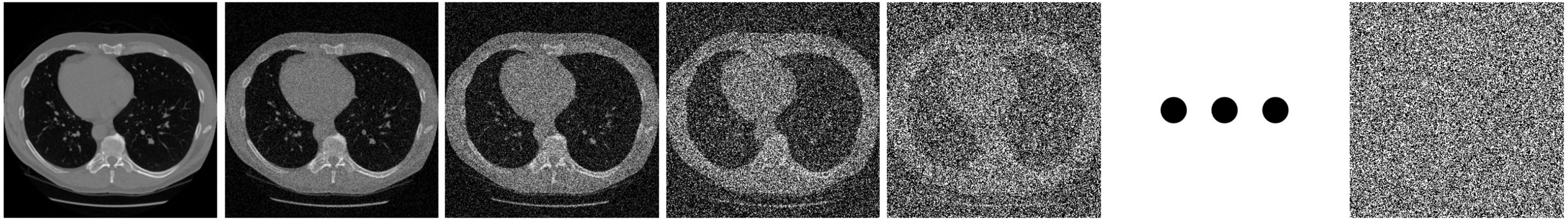
16 – Deep Fakes

- Overview of generative models
- Security risks and guidelines
- Deep fake detection

MegaPortraits: One-shot Megapixel Neural Head Avatars, Drobyshev et al.

17 – Inverse Problems

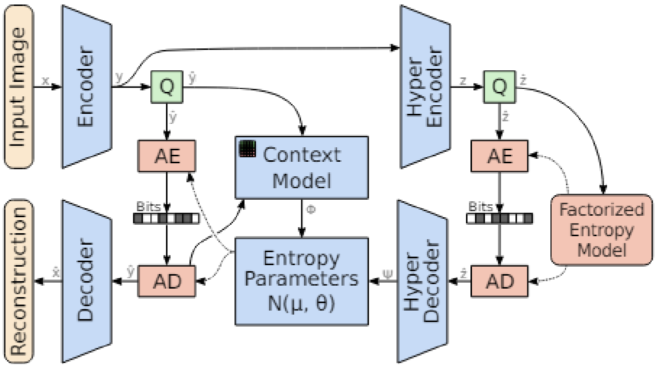
- Overview of inverse problems and solutions
  - Denoising
  - Deblurring
  - Image restauration
- Applications in medical imaging



Solving Inverse Problems in Medical Imaging with Score-Based Generative Models, Song et al.

18 – Learning-based Image Compression

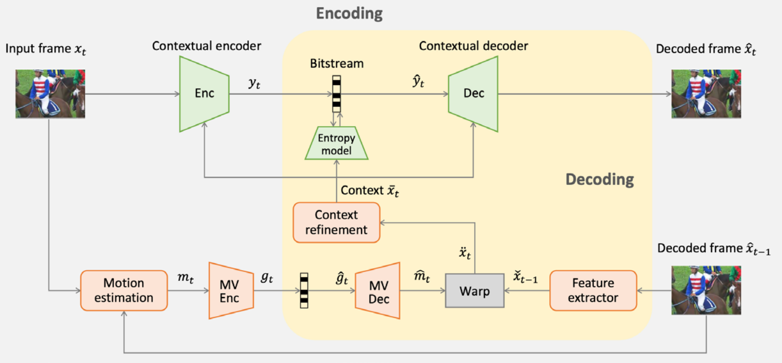
- How to represent an image with fewer signals without losing too much quality?
- Find and exploit redundancies in an image
- Autoencoder-based algorithms
  - Ballé, Minnen, Cheng
  - Contextformer, MLIC
- Implicit neural representation-based codecs
  - SHACIRA
  - COIN++
  - Cool-Chic



Joint Autoregressive and Hierarchical Priors for Learned Image Compression, Minnen et al.

19 – Learning-based Video Compression

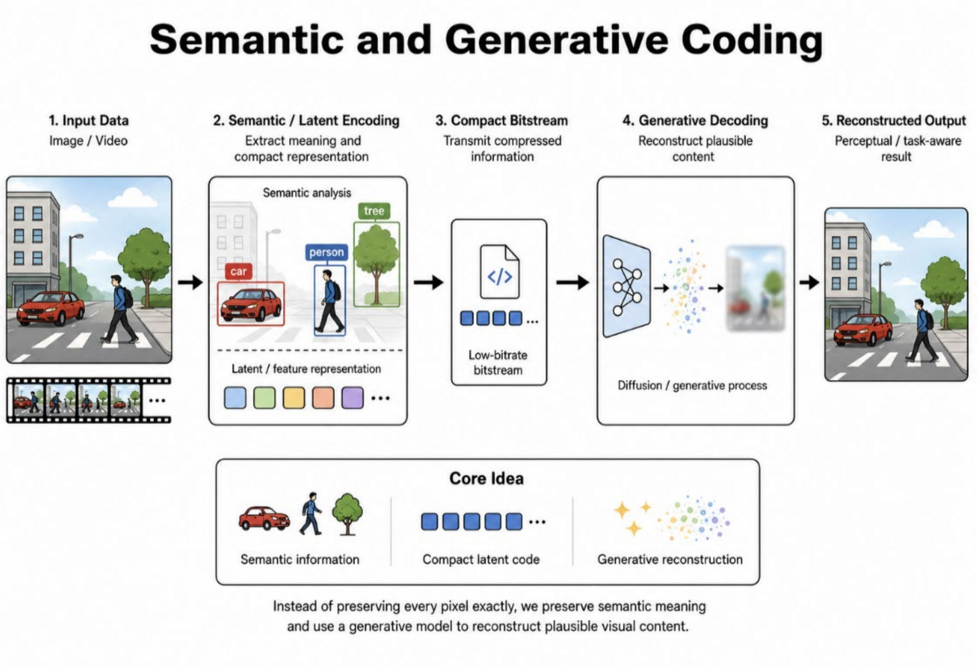
- Exploit temporal correlations
- Introduce optical-flow-based motion estimation and motion compensation
- Overview over most influential networks
  - Deep Video Compression (DVC)
  - Deep Contextual Video Compression (DCVC) and extensions: TCM, HEM, DC, FM, RT



Deep Contextual Video Compression, Li et al.

20 – Semantic and Generative Coding for Visual Compression

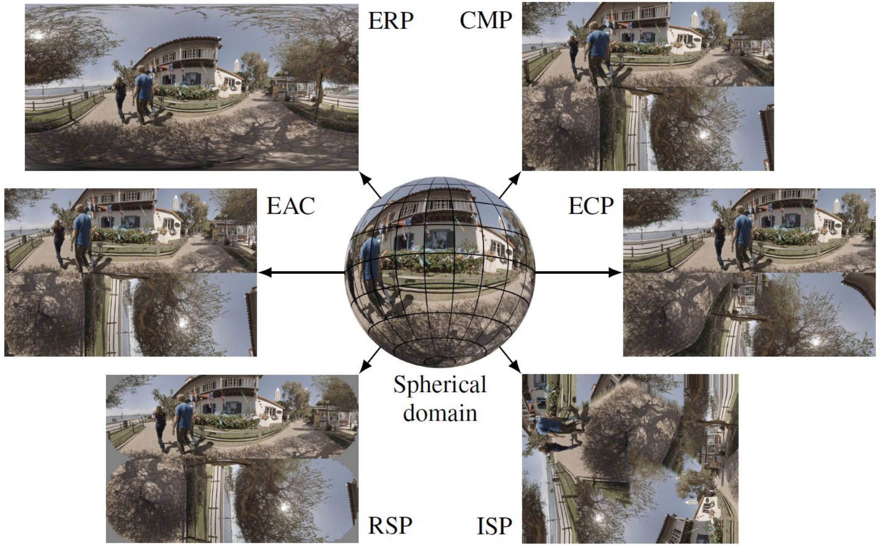
- Compression paradigms beyond pixel-level reconstruction
- Semantic coding preserves task-relevant information, e.g., objects, actions, scene structure, and features
- Generative coding reconstructs plausible content from compact latents, semantic descriptions, or sparse information
- Applications: ultra-low-bitrate compression, video coding for machines, and visual communication



TUM FAU LMS University of Stuttgart Germany

21 – Processing and Coding of 360°-Videos

- Introduction and challenges of 360°-video processing
- Projection formats
  - Equirectangular (ERP)
  - Cubemaps (CMP, EAC, GCP)
  - Cylindrical projections
- Coding of 360°-videos



Analysis of Neural Video Compression Networks for 360-Degree Video Coding, Regensky et al.

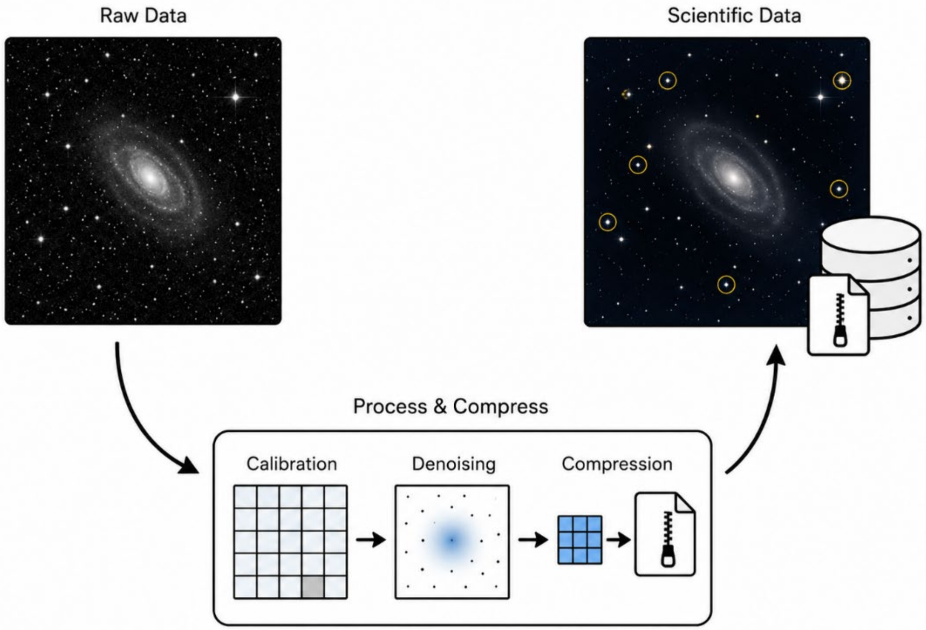
22 – Perception-based Quality Evaluation

- How to represent subjective quality with a single number?
- Limitations of classical objective evaluation metrics
  - No-reference and full-reference metrics
  - PSNR, MS-SSIM, BRISQUE, NIQE
- State-of-the-art approaches towards perception-based quality evaluation
  - LPIPS, VMAF, …

r0k.us/graphics/kodak

23 – Processing and Compression of Astronomical Data

- Extract scientifically meaningful information from large-scale astronomical observations
- Special data properties: high dynamic range, low signal-to-noise ratio, sparsity, and sensor artifacts
- Compression is important due to massive image, spectral, time-series, and radio data
- Methods include FITS compression, Rice/Hcompress, learned compression, and task-aware compression



Practical Part

- Apply selected algorithms on self-captured images and videos
- Possible topics:
  - Group 1: Learned image compression
  - Group 2: Learned video compression
  - Group 3: Hybrid codecs
  - Group 4: Style transfer
  - Group 5: Super resolution
- Open for project proposals

Topic Distribution

- Topic list available at **lms.tf.fau.de/ferienakademie-2026**
- Send your topic priorities via mail to **paul.wawerek-lopez@fau.de**
  - Highest → lowest (at least 5)
  - Example: [21, 4, 7, 19, 12]
  - Deadline: 19th of July (Sunday), 11:59 pm (midnight)