1	**Digital Cameras** Cameras are the basis of most perception systems in Computer Vision. A fundamental understanding of the underlying concepts of digital cameras is thus crucial. This topic shall cover the basics of digital cameras, introduce different camera models such as the pinhole or fisheye model, and explain the methodology of camera calibration, including intrinsic and extrinsic parameters, distortion coefficients, and common calibration techniques like checkerboard calibration or Zhang's method, which are essential for accurate measurements and image rectification.
2	**Depth Perception** Depth perception is crucial for accurate scene understanding, object placement, and navigation in image and video processing applications. This topic shall introduce the basics of depth perception including monocular and binocular cues, explain the epipolar geometry and its relation to stereo depth sensing, and cover common algorithms for monocular and stereo depth estimation. The topic shall also briefly touch upon the calibration of such systems.
3	**Deep Learning: Basics** Deep Learning algorithms form the basis of most state-of-the-art Computer Vision systems. Hence, a basic understanding of the theory behind Deep Neural Networks is essential for anyone working in this field. This topic will introduce the basics of Neural Networks by first explaining neurons, activations, and essential layers such as convolutional layers, pooling layers, and fully connected layers. The presentation will then cover how these networks can be trained using gradient-descent algorithms with the help of backpropagation. These foundational concepts are crucial for building and understanding advanced deep learning models.
4	**Deep Learning: Training and Risks** Effective training of deep neural networks is essential for achieving high performance in various applications. This topic shall introduce the fundamental principles and methodologies for training neural networks, focusing on different optimization algorithms such as Stochastic Gradient Descent (SGD), momentum, Adagrad and Adam, and explain how these algorithms build upon each other to improve training efficiency and convergence. Additionally, essential training techniques, including batch normalization, dropout, and learning rate schedules shall be covered, and common risks associated with training and deploying neural networks, such as class imbalance or adversarial attacks shall be addressed briefly.
5	**Object Detection and Classification** Object detection and classification are foundational tasks in computer vision, enabling systems to identify and categorize objects within images and videos. This topic shall introduce the fundamental concepts of object detection and classification, cover key frameworks such as YOLO and Faster R-CNN, and explain their architectures, strengths, and applications. The topic shall also provide a brief overview of common object tracking methodologies such as SORT (Simple Online and Realtime Tracking), Deep SORT, or tracking-by-detection.
6	**Image Segmentation** Image segmentation is a crucial technique in computer vision that involves partitioning an image into segments, making it easier to analyze and process. Unlike classical object detection, which provides rectangular bounding boxes, image segmentation offers pixel-accurate masks for different objects in a scene. This topic shall introduce the fundamental concepts of image segmentation, differentiating between semantic segmentation and instance segmentation, and cover popular segmentation networks such as Mask R-CNN and DeepLabV3+. The topic shall also cover a basic overview of the recent Segment-Anything-Model (SAM) that is able to produce masks for arbitrary image content.

7	**Autoencoders** Autoencoders are used to learn efficient representations of data, typically for dimensionality reduction or generative purposes. They consist of an encoder that compresses the input into a latent space representation, and a decoder that reconstructs the output from this representation. This topic shall introduce the basic concept of autoencoders, explain their architecture and training procedure, and highlight different variants such as Denoising Autoencoders, Contractive Autoencoders, Variational Autoencoders, and Vector-Quantized Variational Autoencoders in the context of image and video signal processing.
8	**Generative Adversarial Networks** Generative Adversarial Networks (GANs) are a class of machine learning frameworks designed by letting two neural networks compete against each other to generate new, synthetic instances of data that can pass for real data. This topic shall introduce the fundamental concept of GANs, explaining the roles of the generator and discriminator, how to train them, and how they compete in a zero-sum game. Furthermore, the challenges in training GANs, such as mode collapse and convergence issues, as well as techniques to mitigate these problems shall be discussed.
9	**Diffusion Models** Diffusion models are advanced generative machine learning algorithms that excel in creating high-quality synthetic data, especially images. They operate by gradually adding noise to training data, then learning to reverse this process. High resolution generation was made possible with Latent Diffusion Models. Conditional Diffusion Models allow the user to control the output with text or other forms of input. Consistency Models and Progressive Distillation tries to speed up the slow generation process.
10	**Transformers** Transformers, originally designed for natural language processing, have revolutionized image and video processing tasks. This architecture uses self-attention mechanisms to capture long-range dependencies in data. In vision tasks, images are divided into patches and treated as sequences. Transformers excel in object detection, image classification, and video understanding. Their ability to process global context and handle variable-length inputs makes them powerful for tasks like image generation and video captioning. Windowed Attention and Linear Attention helps to scale these models for large inputs.
11	**Text-Conditioned Image Generation** Text-conditioned image generation is a powerful deep learning task that creates images based on textual descriptions. This technique has seen significant advancements through both self-supervised and diffusion-based approaches. Self-supervised methods, like DALL-E and CLIP, learn text-image relationships without explicit labels. Diffusion models, such as Stable Diffusion, generate high-quality images by iteratively denoising random noise guided by text prompts.
12	**Conditional Video Generation** Conditional video generation creates videos based on specific inputs or constraints. This deep learning task encompasses various approaches, including text-to-video, image-to-video, and video-to-video generation. Models like GANs (ImaGINator), VAEs, and diffusion-based architectures (LFDM, Imagen-Video) are employed to synthesize realistic and coherent video sequences. These techniques can generate facial expressions, human actions, or complex scenes guided by textual descriptions or initial images.

13	**Implicit Neural Representations** Implicit neural representations (INRs) are a novel approach to parameterizing signals as continuous functions using neural networks. INRs map input coordinates directly to signal values, enabling resolution-independent representations of images, audio, video, and 3D scenes. Key advantages include memory efficiency, infinite resolution, and the ability to incorporate priors through network architectures. Popular variants like SIRENs use periodic activations for improved performance. INRs have shown promise in tasks like novel view synthesis (NeRF) and compression (Cool-Chic, NeRV).
14	**Image Super Resolution** Image super-resolution is aimed at enhancing low-resolution images. Key techniques include convolutional neural networks (CNNs), generative adversarial networks (GANs), implicit neural representations (INRs) and diffusion-based models. Notable algorithms such as SRCNN, ESRGAN, and RCAN have pushed boundaries in this field. Recent advancements incorporate transformer architectures and diffusion models. While some deterministic algorithms produce the same output for the same input, there are also stochastic solutions where there can be multiple super resolution candidates for a given low resolution image. The challenge lies in balancing perceptual quality, computational efficiency, and faithful reconstruction.
15	**Learned Image Compression** Learned image compression applies deep learning techniques to compress images more efficiently than traditional methods. This approach typically uses neural network architectures like autoencoders or transformers to learn optimal representations of image data. Key components include entropy models for latent representations and encoding/decoding networks. Learned compression methods have shown superior rate-distortion performance compared to classical standards like JPEG or BPG, often preserving more details at lower bitrates. Recent advancements incorporate attention mechanisms, context modeling, and hybrid CNN-Transformer architectures to further improve compression efficiency.
16	**Learned Video Compression** Learned video compression extends learned image compression techniques by exploiting temporal correlations. This topic shall cover the basics of optical flow-based motion estimation and motion compensation, essential for effectively compressing video by predicting and encoding motion between frames. Furthermore, the concepts of residual coding and conditional coding shall be introduced among an overview over the most influential networks in the field, including Deep Video Compression (DVC), Deep Contextual Video Compression (DCVC), as well as its extensions DCVC-TCM, DCVC-HEM, DCVC-DC, and the state of the art DCVC-FM.
17	**Image to Image Translation** Image to Image Translation is a versatile technique in computer vision that enables the transformation of images from one domain to another, making it useful for applications like style transfer, image restoration, and data augmentation. This topic shall introduce the concept of image to image translation, cover the difference between paired and unpaired training, and discuss common methodologies and networks used in image to image translation, such as Pix2Pix, StyleGAN and CycleGAN along their individual strengths and weaknesses.
18	**Self-Supervised Learning** Self-supervised learning in deep learning enables models to learn from unlabeled data, discovering patterns and representations without human annotations. These methods may be expanded into contrastive representation learning. Applications in image and video processing, such as image denoising, super-resolution, and object detection, showcase the benefits of self-supervised learning, including reduced labeling costs and improved model robustness. SIMCLR and Contrastive Predictive Coding are some prominent models in this domain.

19	**Transfer Learning** The power of transfer learning in deep learning lies in fine-tuning pre-trained models for new tasks, leveraging knowledge from one task to improve performance on related tasks, reducing training time and improving accuracy. Applications in image and video processing, such as object detection, image classification, and image generation can benefit from transfer learning. Few-shot and zero-shot learning may employ this technique. Parameter efficient fine-tuning algorithms like LoRA are also a form of knowledge transfer between domains.
20	**Reinforcement Learning** Reinforcement learning techniques are applied to tasks like object detection, segmentation, and action recognition in images and videos. Markov Decision Processes (MDPs) and Partially Observable MDPs (POMDPs) model sequential decision making. Optimal actions are learned from rewards through Q-learning and Deep Q-Networks (DQNs). Policies are directly learned through policy gradient methods like REINFORCE. Stable policy learning is enabled by Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C). RL has been extensively used in robot action planning for visual tasks like navigation, manipulation, and interaction.
21	**Human Action Recognition** Human action recognition techniques are explored for recognizing and localizing human actions in videos. Temporal action localization detects the start and end times of actions, while spatial action localization identifies the spatial extent of actions. Two-stream architectures combine RGB and optical flow for effective motion modeling. 3D convolutional networks enable spatiotemporal feature learning, while recurrent neural networks like LSTMs or Transformers model long-range dependencies. Graph convolutional networks operate on skeletal data for action recognition. Weakly-supervised learning leverages video-level labels, and few-shot learning recognizes actions from limited examples. There are still many challenges to be solved like untrimmed videos, occlusions, and viewpoint variations.
22	**Visual Simultaneous Localization and Mapping** Simultaneous Localization and Mapping (SLAM) is essential to enable applications such as Augmented and Virtual Reality, Robotics, or Autonomous Driving. Visual SLAM (vSLAM) refers to SLAM algorithms that perform the localization and mapping process based on visual data. This topic shall introduce the main processing steps of SLAM algorithms including an explanation of Kalman filtering that is a crucial part of one of the first practical SLAM algorithms called EKF-SLAM. The topic shall also cover how camera sensor information (image information) can be incorporated into vSLAM, and how deep learning based methods can benefit these systems.
23	**Lane Detection and Motion Planning** Lane detection is an essential task in assisted and autonomous driving to understand the vehicle's environment and find valid motion trajectories. This topic shall introduce learning-based lane detection approaches such as LaneNet or SCNN, and provide an overview over common motion planning techniques such as Rapidly-exploring Random Trees (RRT), Probabilistic Roadmaps (PRM), or Model Predictive Control (MPC), whereby the difference between path planning and trajectory planning shall be made clear.