

Computer Vision

1. Image Classification

- VGG
- ResNet (both Basic - Bottleneck block)
- MobileNet
- DenseNet
- EfficientNet

2. Attention & Transformer

- Attention
 - For channel and spatial
- Transformer - Vision Transformer
- Swin Transformer

3. Object Detection

- Basic: Anchor boxes, IoU, NMS
- Label classifier - Bounding box regressor
- Multi-scale detection
- Feature map visualization
- Model:
 - SSD, RetinaNet, YOLO
 - EfficientDet, CenterNet
 - DETR
 - Pix2Seq

4. Video object tracking

- SORT, DeepSORT
- CenterTrack

5. Object Segmentation

- FCN
- U-Net
- Mask RCNN
- Masked YOLO
- DeepLab v1-3, NAS-FPN, HRNET?

6. Action Recognition

- 3DConv
- One/Two streams: early vs late fusion

7. Image Generation

- Neural Style Transfer
- GAN: Pix2Pix, ConditionalGAN, CycleGAN, StyleGAN
- Diffusion Model

Syllabus

ID	Module	Feature	Duration (days)	Note
1	Image Classification	VGG	1	
		ResNet	1	Basic &

				Bottleneck block
		DenseNet	1	
		MobileNet	1	v1-3
		EfficientNet	1	
2	Attention & Transformer	Attention	1	CBAM & SAM
		Transformer	1	
		Transformer for Vision	1	VíTs
		Swin Transformer	1	
3	Object Detection	Basics	1	
		SSD	1	
		RetinaNet	1	
		YOLO	3	v3 4 5 6 7
		EfficientDet	1	
		CenterNet	1	
		DETR	1	
4	Video Object Tracking	SORT	1	
		DeepSORT	1	
		CenterTrack	1	
5	Object Segmentation	FCN	1	
		U-Net	1	
		Mask RCNN	1	
		DeepLab	2	v1-3
		NAS-FPN	1	
		HRNet	1	
6	Action Recognition	3DConv	2	

		One/Two Streams	1	
7	Image Generation	Neural Style Transfer	1	
		Pix2Pix	1	
		ConditionalGAN	1	
		CycleGAN	1	
		StyleGAN	1	