Week	Heading	Main Paper	Additional Readings
Week 1	**LeNet**	Gradient-Based Learning Applied to Document Recognition	All about convolutions : A guide to convolution arithmetic for deep learning
	Deep model problems	Understanding the Difficulty of training Deep feedforward Neural Networks	Activation Functions: Comparison of trends in Practice and Research for Deep Learning
Week 2	**AlexNet**	ImageNet Classification with Deep Convolutional Neural Networks	Visualizing and Understanding Convolutional Networks
			Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification
	VGGNet	Very Deep Convolutional Networks for Large-Scale Visual Recognition	Dropout: A Simple Way to Prevent Neural Networks from Overfitting
Week 3	**Deep-supervision**	Deeply supervised nets	1. Layer normalization, 2. instance normalization, 3. group normalization, 4. weight standardization
	Net-in-Net	Inception v2: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,	Network in Network
	Knowledge Distillation	Distilling the Knowledge in a Neural Network	
Week 4	**Inception Models**	GoogLeNet: Going Deeper with Convolutions	Multi-scale context aggregation by dilated convolutions
		Inception v3: Rethinking the inception architecture for computer vision	Deformable convolutional networks
Week 5	**ResNet**	Deep Residual Learning for Image Recognition	
	Identity mapping	Identity Mapping in Deep Residual Networks	
Week 6	**Wide ResNet**	Wide Residual Networks	Deep Networks with Stochastic Depth
	ResNeXt	Aggregated Residual Transformations for Deep Neural Networks	ResNet strikes back: An improved training procedure in timm
Week 7	**DenseNet**	Densely Connected Convolutional Networks	Memory efficient implementation of DenseNets
	U-Net	U-Net: Convolutional Networks for Biomedical Image Segmentation	Inception-ResNet, indeption-Densenet
Week 8	**Transformer**	Attention is all you need	
	Vision Transformer	An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale	https://www.youtube.com/watch?v=3B6q4xnuFUE&list=PLcp6ZnH4WYlbSPL-zrFNOygUiYV6HN6sK&index=3&ab_channel=AIBites
Week 9	**DieT**	Training data-efficient image transformers & distillation through attention	
	Swin Transformer	Swin Transformer: Hierarchical Vision Transformer using Shifted Windows	https://www.youtube.com/watch?v=tFYxJZBAbE8&ab_channel=AIBites
			https://sieunpark77.medium.com/swin-transformers-the-most-powerful-tool-in-computer-vision-659f78744871
	Spatial Attention	Twins: Revisiting the design of spatial attention in vision transformers	Pyramid Vision Transformer
Week 10	**convolution in ViT**	CvT: Introducing Convolutions to Vision Transformers	Xception: Deep Learning with Depthwise Separable Convolutions
	Locality in ViT	Local ViT: Bringing Locality to vision transformers	Demystifying local vision transformers: sparse connectivity, weight sharing, and dynamic weight
Week 11	**Self-attention Nets**	Exploring self-attention for image recognition	
	No attention	Do you even need Attention? A stack of feed-forward layers does surprisingly well on ImageNet	
	Feature Shifting	When Shift Operation Meets Vision Transformer: An Extremely Simple Alternative to Attention Mechanism (partial shift operation)	All you need is a few shifts: designing efficient convolutional neural networks for image classification
			Active shift: Constructing fast network through deconstruction of convolutions
			Shift: A Zero FLOP, Zero Parameter Alternative to Spatial Convolutions (grouped shift)
Week 12	**All MLP models**	MLP-Mixer: An all-MLP Architecture for Vision	Feedforward networks for image classification with data-efficient training
		cycleMLP: A MLP like architecture for dense prediction	S2MLP: Spatial-shift MLP architecture for vision
Week 13	**convNext**	A ConvNet for the 2020s	
	CNN, Transformer, MLPs	A Battle of Network Structures: An Empirical Study of CNN, Transformer, and MLP	
Week 14	**Efficient Nets**	EfficientNet	EfficientNet v2: smaller models and faster training

Week 14	**Sqeeze-and-excite**	Squeeze-and-excitation Networks	Constructing Fast Networks through Deconstruction of Convolution