Deep Learning

Recap

Danijel Skočaj
University of Ljubljana
Faculty of Computer and Information Science

Academic year: 2022/23

Content

- Basic concepts
- Training neural networks
- Optimazing training
- Convolutional Neural Networks
- Computer vision
- Recurrent Neural Networks
- Transformers
- Natural Language processing
- Transformers in comp. vision
- Generative models

Theory

Applications

SOTA

Contents

Contents

Computer vision: Visual information. Computer vision tasks. Classification. Localistaion. Semantic segmentation. Fully-convolutional approach. Upsampling. Transpose convolution. FCN. Deconvolution network. SegNet. Encoder-decoder architecture. U-net. Shortuct connections. PSP-Net. Pyramid Pooling. DeepLab. Atrous convolution. Atrous Spatial Pyramid Pooling. Semantic segmentation overview. Beyond segmentation. Surface-defect detection. Counting. Obstacle detection. Image enhancement. Spatially Adaptive Filter Units. Semantic edge detection. Object detection. Detection of multiple objects. Two-stage object detectors. Region proposal methods. R-CNN. Fast R-CNN. RoI pooling. Faster RCNN. Region Proposal Network. Instance segmentation. Mask R-CNN. RoI Align. Task specific heads. Human pose detection. Panoptic segmentation. Feature Pyramid Network. Panoptic Feature Pyramid Networks. Single-stage object detectors. SSD. Anchors. Yolo. RetinaNet. Focal loss. FCOS. Traffic sign detection. Face detection. Object detection arhitectures overview. Performance of object detectors. Classification peformance measures. Object detection performance measures. Semantic segmentation performance measures.

Recurrent Neural Networks: Sequential data. CNN-based approach. Requirements. Recurrent Neural network. Unfolding RNN. Multilayer RNN. Reccurence formula. Computational graph. Backpropagation through time. TBPTT. Character-level language model. Predicting the next letter. Interpreting the hidden cell values. RNN trade-offs. BPTT problems. Long Short Term Memory. LSTM gradient flow. Gated Recurrent Units. RNN variants. Bidirectional LSTM. LSTM examples. Attention mechanism. Attention in RNNs. Image captioning with attention.

Transformers and NLP: Transformer architecture. Encoder. Self-attention. Multi-head attention. Positional encoding. Decoder. Encoder-decoder attention. Masked self-attention. Decoding. RNNs vs. Transformers. Transformer-XL. BERT. BERT pretraining. BERT fine-tuning. RoBERTa. ALBERT. DeBERTa. BERT examples. T5. GPT. Unsupervised generative pretraining. Supervised discriminative fine-tuning. GPT-2. GPT-3. InstructGPT. Aligment problem. ChatGPT. Power and considerations of modern LLMs. Speech recognition. Whisper. Music transformer.

Contents

Transformers in computer vision: Image transformer. ViT Vision transformer. DeiT. MViT. Swin transformer. Swin V2. SeMask. CvT. CoAtNet. DETR. Deformable DETR. UP-DETR. MaskFormer. Mask2Former. DINO. DINOv2. MLP-Mixer. Segment Anything.

Generative models: Gaussian and GMM. Principal Component Analysis. Robust projection. Autoencoders. Masked autoencoders. Autoencoders for anomaly detection. Generative vs. discriminative approaches. Variational Autoencoders. VAE generating images. VAE in music. Generative Adversarial Networks. Training GAN. DCGAN. LSGAN. Wasserstain GAN. Improved Wasserstain GAN. Progressive GAN. Pix2pix cGAN. CycleGAN. BigGAN. SinGan. StyleGAN. F-AnoGAN. GANomaly. C-VTON. Normalizing flows. Real NVP. Flow++. Glow. DifferNet. FastFlow. PixelCNN. VQ-VAE. CLIP. DALL-E. Diffusion models. DDPM. Stable Diffusion. GLIDE. DALL-E2. ControlNet. Midjourney.

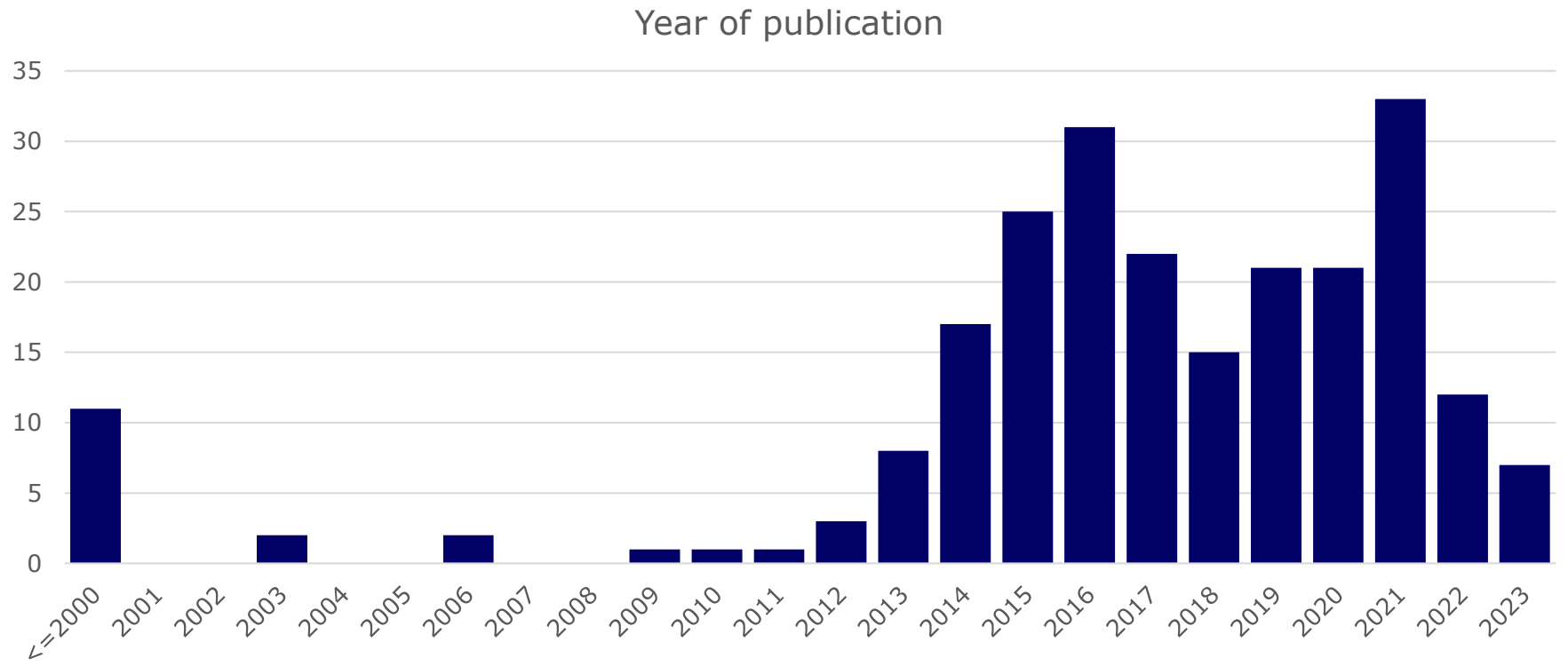
Conclusion: Deep learning. Key factors for fair deep learning. Learning regimes. Deep reinforcement learning. RL for local navigation. Built in vs. Learned. Beyond CV and NLP. Function approximator. Development, deployment and maintenance. Problem solving.

Level of details:

- Detailed knowledge
- Main concepts
- Main idea

Research papers

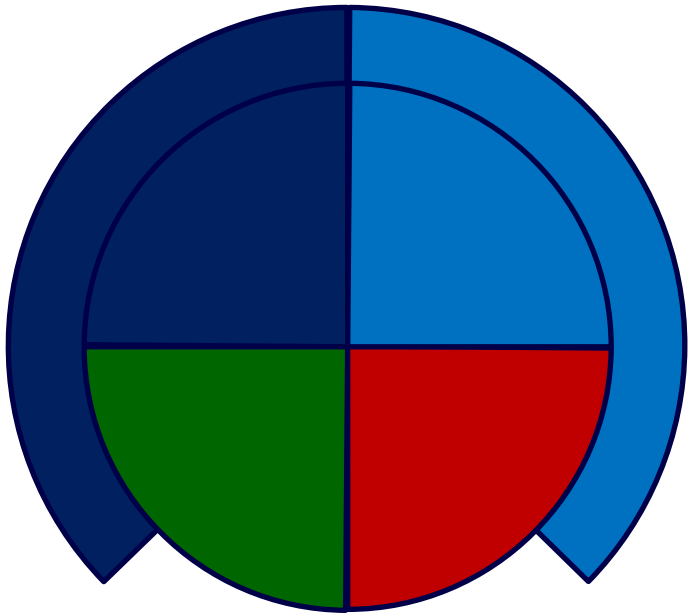
- 233 papers presented/mentioned
- All of them available online!
- Seminal papers and SOTA



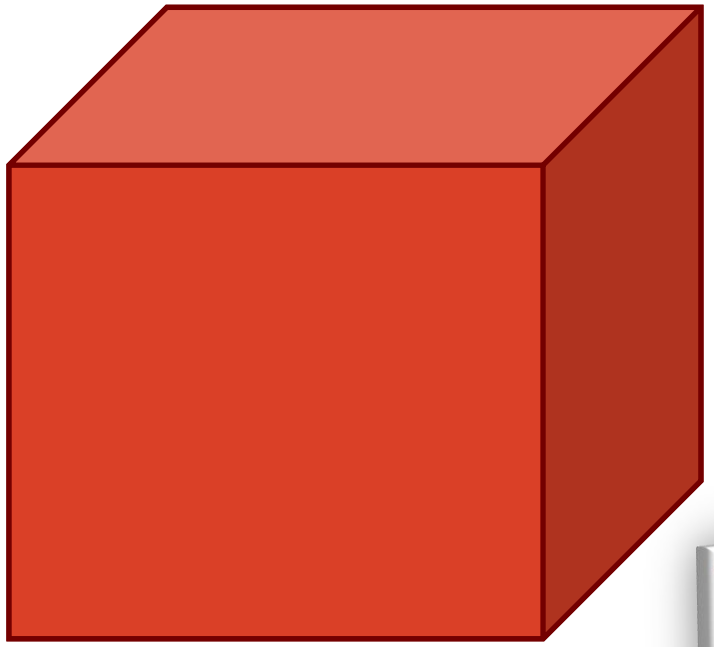
Grading

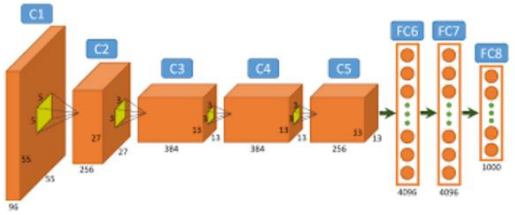
- Assignments: 25 pts
- Project: 25 pts
- Written exam: 25 pts
- Oral exam: 25 pts

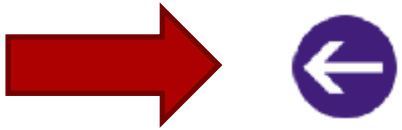
- The student has to collect at least 50% in every task



Understanding deep learning







ReLU η

FullyConnectedLayer

$$C = -\frac{1}{} \sum [y \ln a + (1-y)\ln(1-a)]$$

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

$$\delta^L = \nabla_a C \odot \sigma'(z^L)$$


```
net = Network([
    ConvPoolLayer(image_shape=(mini_batch_size, 1, 28, 28),
                  filter_shape=(20, 1, 5, 5),
                  poolsize=(2, 2)),
    FullyConnectedLayer(n_in=20*12*12, n_out=100),
    SoftmaxLayer(n_in=100, n_out=10)], mini_batch_size)
net.SGD(
    training_data, epochs, mini_batch_size, 0.1,
    validation_data, test_data)
```

```
for x, y in mini_batch:
    delta_nabla_b, delta_nabla_w = self.backprop(x, y)
    nabla_b = [nb+dnb for nb, dnb in zip(nabla_b, delta_nabla_b)]
    nabla_w = [nw+dnw for nw, dnw in zip(nabla_w, delta_nabla_w)]
self.weights = [w-(eta/len(mini_batch))*nw
                for w, nw in zip(self.weights, nabla_w)]
```

Using deep learning