DIGITAL - Institute for Information and Communication Technologies

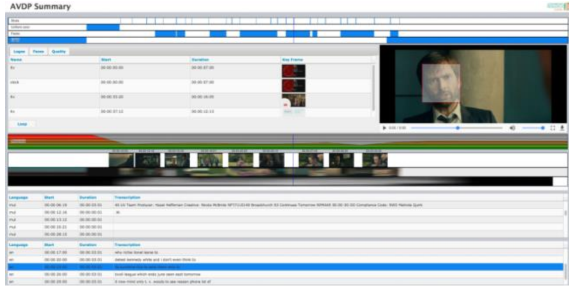
Employing Deep Learning for Automatic Analysis of Conventional and 360° Video

Hannes Fassold

2019-03-20

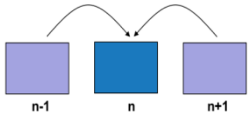
Our research group

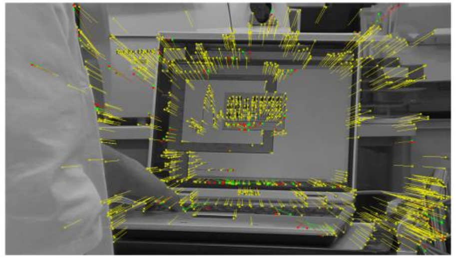
- GPU-accelerated algorithms / applications @ CCM / JRS
  - Connected Computing research group, DIGITAL – Institute for Information and Communication Technologies, JOANNEUM RESEARCH (**JRS**), Graz, Austria
  - Content-based quality analysis & restoration of film and video
    - http://vidicert.com
    - http://www.hs-art.com
  - Real-time video analysis
    - Brand monitoring
    - Object (faces, persons, ….) detection, tracking & recognition
  - Surveillance / traffic video analysis
  - Standardization activities
    - MPEG: Compact neural networks, CDVA, …
  - GPU research & development since 2007

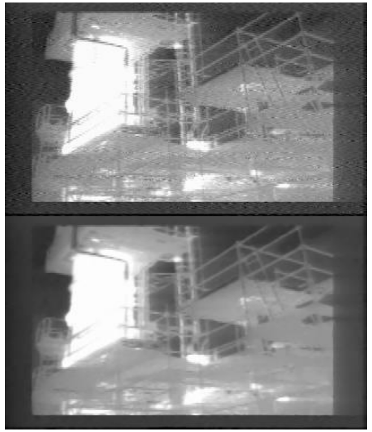


Our GTC history

- NVISION 2008 – the „start“
  - 1000 (?) attendees, 45 sessions, 19 posters
- GTC 2018
  - 8500 attendies, 700 sessions, 150 posters
- Our presence at GTC (San Jose)
  - NVISION 2008 (visitor)
  - All years except 2011 & 2017 ☺
  - Gave 6 sessions, 3 posters
  - Feature point tracking, inpainting, optical flow, SIFT features, wavelets, …

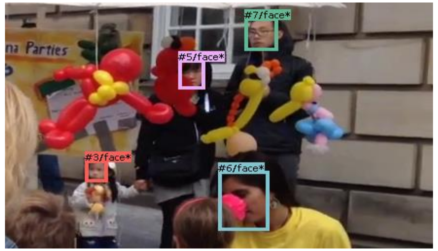


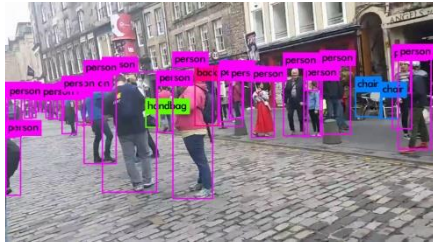




Presentation overview

- Building / Deployment of AI Frameworks
  - Frameworks & platforms
  - Docker container & cloud
- JRS Face Framework
  - Face detection & recognition (FaceNet)
  - Face synthesis (GANs)
  - Application: Anonymization of training data
- JRS Object Framework
  - Object detection (YOLOv3) & tracking (Yoco)
  - Application: Camera path from 360° video
- Standardization activities & Outlook





Platforms & frameworks

- AI Frameworks for Rapid Prototyping
  - **TensorFlow**, MxNet, PyTorch, .. (Python)
- AI Frameworks for Deployment
  - TensorFlow (*C++ API*)
  - Darknet (C API)
- Platforms & Build Tools
  - **Windows**, CentOS 7, Ubuntu 16.04, …
  - CMAKE for generating native ‚project files‘
  - C++ Compilers – VS 2013/2017, GCC 4.8 / 5.3 / …

https://pjreddie.com/darknet/

https://cmake.org/

TensorFlow C++ API

- Building TensorFlow C++ library
  - *Bazel* build tool
  - Very complex to build TF with all dependencies
    - Lot of 3rdparty contributions, with multiple Eigen & protobuf versions, …
    - High risk of conflict of TF dependencies with dependencies of our own software libs
- Porting TensorFlow Python DL Models to C++
  - TensorFlow C++ API contains only subset of TF Python framework
    - Only inference-related functionality is available, no creation or (re)training of graphs
  - *Numpy* functionality must be substituted with C++ library
    - Blitz++
    - XTensor (recent C++ 11 capable compiler necessary,
      not working for VS 2013 / GCC 4.8)

Darknet C API

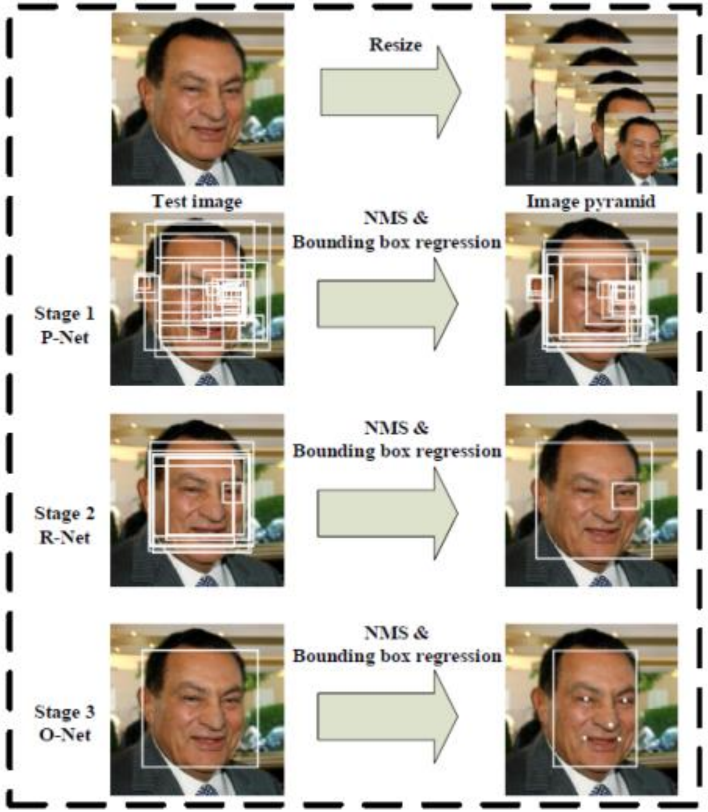
- Darknet
  - https://github.com/pjreddie/darknet
  - Small, self-contained and fast C library
    for 2D DNNs and RNNs
  - Missing: 3D CNNs, <newest-superfancy-tensorflow-contrib-stuff>
  - Contains all versions of SoA *Yolo* object detector (more later)
- Building Darknet C library on Windows
  - Significant code adaptions necessary (GCC vs. VS 2013)
  - Windows replacement for *Pthreads* Linux system library was necessary

Docker & cloud deployment

- We use NV-Docker (version 2.0)
- Platforms
  - CentOS 7
  - Container Linux (Core OS) for Amazon ECS
- Issues
  - Out-of-the-box Amazon ECS instance did not work well with NV-Docker
    - Reason: Driver issues, 8 GB default size of attached storage is easily exceeded for DL containers
    - **Workaround**: Create own Amazon EC2 image (with CoreOS) for use with ECS
  - Docker-compose and NV-Docker did not work together well
    - Compose is a tool for defining and running multi-container Docker applications
    - **Workaround**: Employ own startup-script instead of docker-compose

Face framework
Face detection & landmark extraction

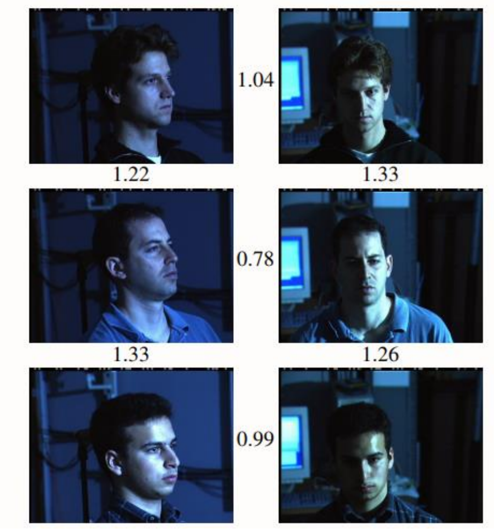
- Face detection & facial landmark extraction
  - Via multi-task cascaded CNNs [Zhang2016]
    - 3 stage approach
    - Employs specialized CNN for each stage (P-Net, R-Net, O-Net)
    - TensorFlow implemention employed
- Algorithm stages
  - Proposal generation (bounding box candidates)
  - Refinement (false positive reduction, NMS, …)
  - Facial landmark detection (5 points)



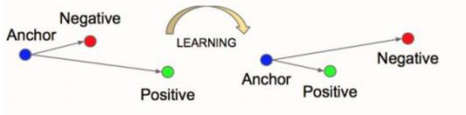
Multi-task cascaded CNNs
Image courtesy of [Zhang2016]

Face framework
Face recognition

- Face recognition
  - Via FaceNet algorithm [Schroff2015]
  - TensorFlow implemention employed
- FaceNet
  - DNN learns ‚optimal' mapping from face to 128-dimensional face descriptor
  - *Triplet* loss function is employed
  - Highly robust against variations in pose & illumination
  - SoA recognition performance
    - 99.63 % on LFW, 95.12 % on Youtube Faces DB



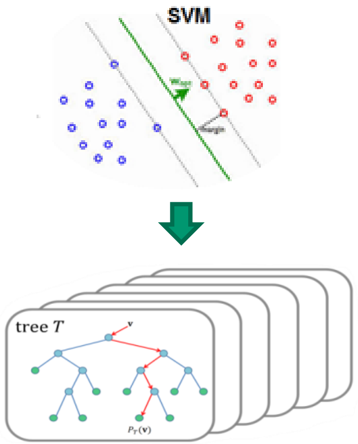
Distance between face descriptors. Image courtesy of [Schroff2015]



Triplet loss. Image courtesy of [Schroff2015]

Face framework
Own extensions

- JRS Extensions to face pipeline
  - Incremental / automatic learning
  - Face tracking
- Incremental / auto-training
  - Allows to add new faces on-the-fly without full re-training
  - Auto-training of faces newly appering in content
  - **Online random forests** (with significant adaptions) instead SVM for classification



- Face tracking
  - Increases robustness of face recognition
- Demo video - courtesy of *Tools On Air*, www.toolsonair.com

Face framework
Face synthesis / GANs

- Generative adversial network (GANs)
  - State of the art for **image synthesis**
  - Two competing networks
    - Generator – Discriminator
    - Generator trys to generate a synthetic image which ‚fools' the discriminator
  - Have reputation of being hard to train (but see [Salimans2016])
- Face synthesis algorithm
  - Employs Deep Convolutional GANs [Radford 2015]

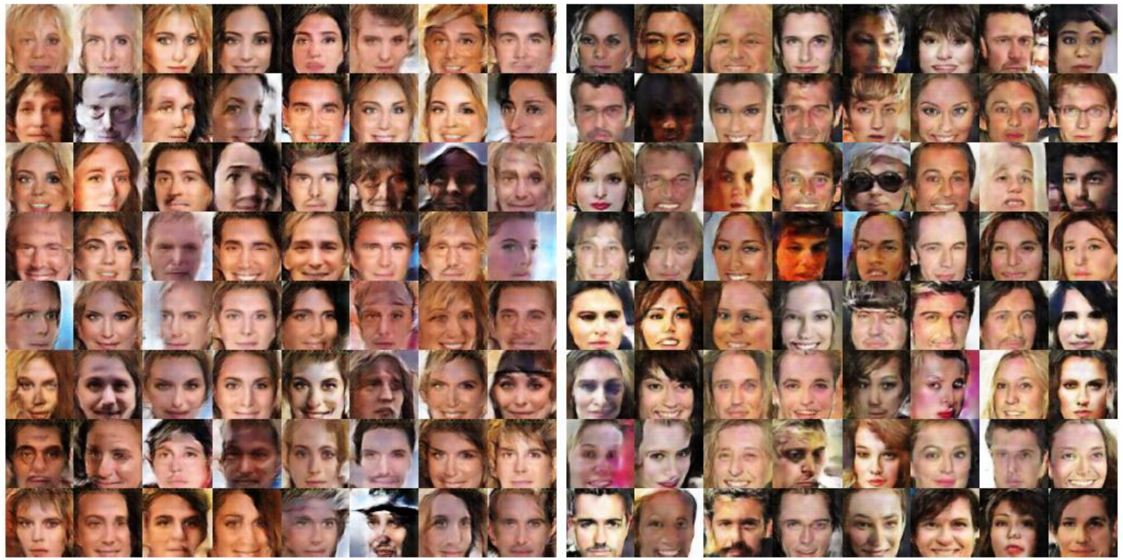


Fig. 1. Examples of generated faces, after two training epochs (left) and seven epochs (right).

Image courtesy of [Bailer2019]

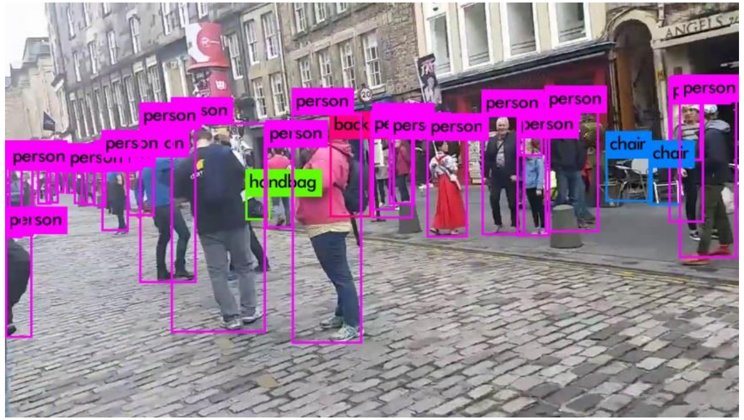
Application
Anonymization of training data

- Motivation
  - Privacy issues
  - EU General data protection regulation (GPDR)
- Face anonymization approach [Bailer2019]
  - Synthesize faces with GANs
  - Bad faces (‚zombie faces‘) are filtered out in a post-processing step
    - Our standard face detector is employed as ‚verificator‘
  - Face swapping in Python
    - https://github.com/wuhuikai/FaceSwap
    - Uses OpenCV & Dlib internally

Anonymized faces .
Images courtesy of [Bailer2019]

Object framework
YOLOv3 object detector

- YOLOv3 object detector [Redmon2018]
  - Very good compromise between detection quality & speed
  - Detects 80 object classes from MS COCO Dataset (person, handbag, car / truck, dog / cat, bottle, …)
- Algorithm principle
  - Single shot detector (no ‚region-proposal' phase employed like in Faster-RCNN)
  - Multi-scale detection at 3 different scales (13 x 13, 26 x 26, 52 x 52 grid)
  - Fully convolutional 106-layer network employed (ResNet-like)



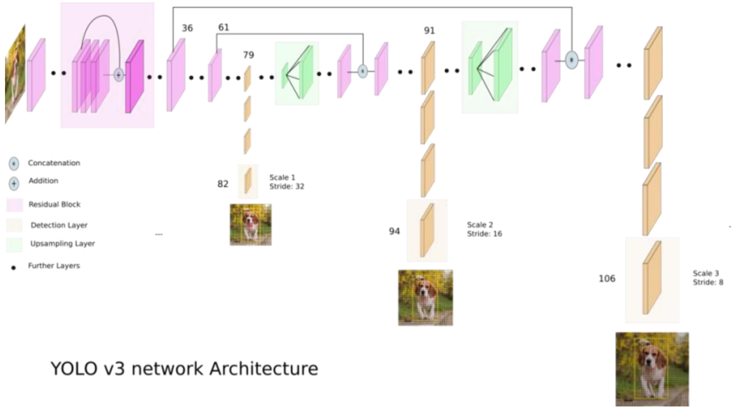
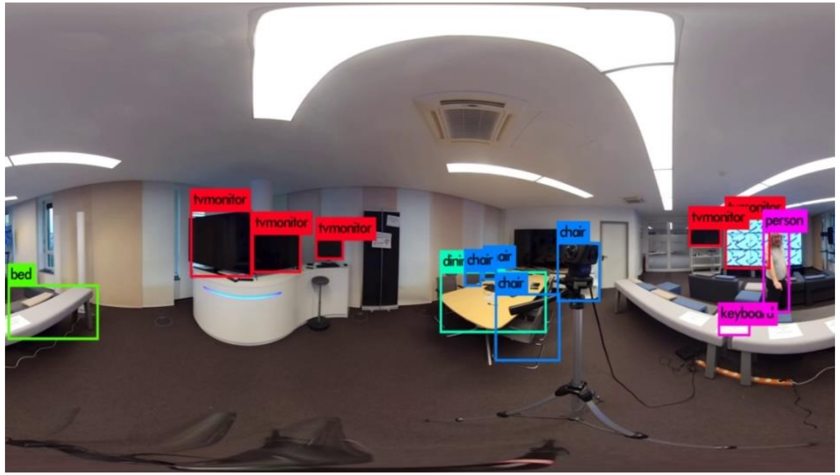


Image courtesy of
https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b

Object framework
YOLOv3 object detector (ct'd)

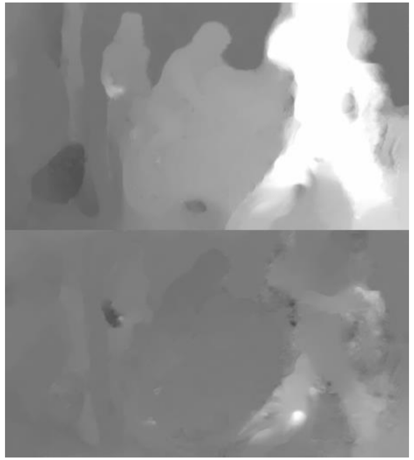
- Algorithm (ct'd)
  - Implementation from Darknet C library
  - Runtime ~ 50 milliseconds
    (608 x 608 pixel, Titan X Pascal)
  - ~ 58 % (mAP-50) detection capability
  - Works well also for images from 360° video



- JRS extensions
  - *Adaptive* size of receptive field (keep same aspect ratio as input image)
  - Do multiple inferences on a single GPU in *parallel* (via separate CUDA streams)
- << Demovideo 360° viewer object detector >>

Object framework
Yoco algorithm

- ***YO**LOv3* **co**mbined with optical flow
  - Detects and tracks all scene objects (persons, …)
  - Important semantic information for many tasks
- Combination of SoA components
  - YOLOv3 algorithm for object detection
  - High-quality GPU-based optical flow for motionfield calculation (TV-L1)
  - Hungarian algorithm for optimal matching
- << Demovideo Yoco algorithm >>



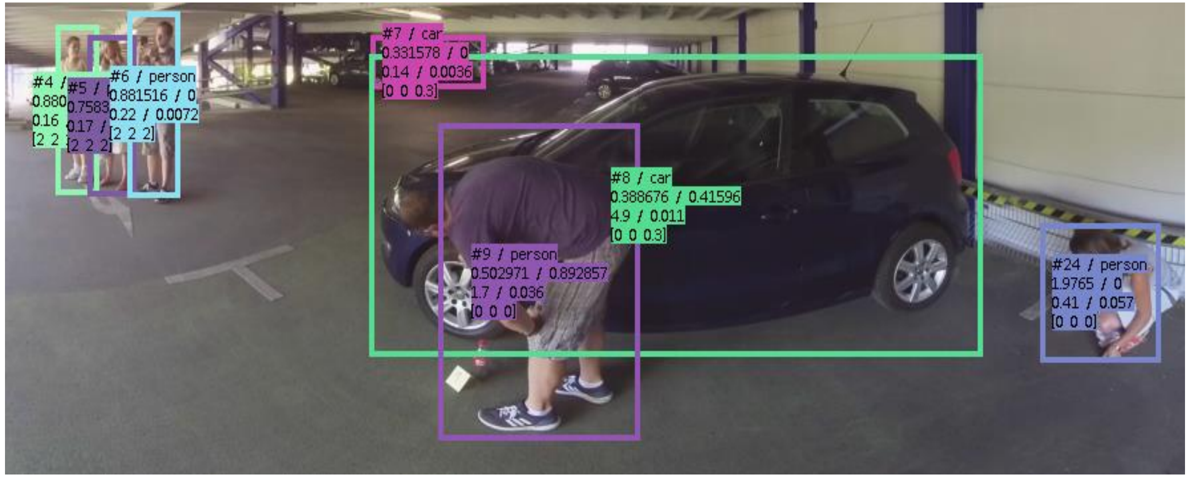
Visualized motionfield

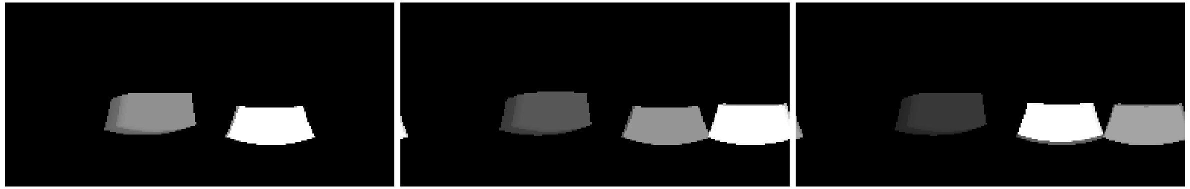
Application
Automatic camera path calculation

- Automatic camera path calculation
  - Provide a „lean-back“ experience for consuming 360° video
- Algorithm outline
  - Works iteratively, shot-per-shot
  - Detect and track all scene objects in shot
  - Calculate measures for each scene object
    - Size, motion magnitude, …
  - Calculate ‚visited map‘
    - Steers camera away from already seen areas of 360° video
  - Calculate saliency score for each object
  - Camera path = track most interesting object

Application
Automatic camera path calculation (ct‘d)

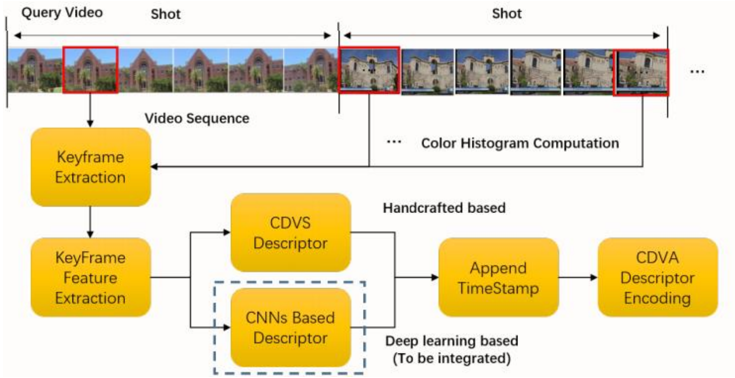
- Influencing factors for saliency score
  - Object class
  - (Average) object size
  - (Average) motion magnitude
  - Visited score
  - Neighborhood score
  - …
- << Demovideo ACP >>





Standardization activities
Our involvement

- MPEG-7 AVDP, EBU QC, FIMS, …
- MPEG-CDVA
  - Compact descriptors for video analysis
  - For efficient video matching & retrieval, …
  - Descriptor size is just a few KByte per second video
- MPEG activity on compact neural networks [1]
  - Goal: efficient and interoperable represention
  - Via compression, pruning, quantization, …
  - JRS co-organized a workshop on that topic [2] at NeurIPS 2018 conference, workshop at ICML 2019



Extraction of CDVA features.
Image courtesy of [Duan2017]

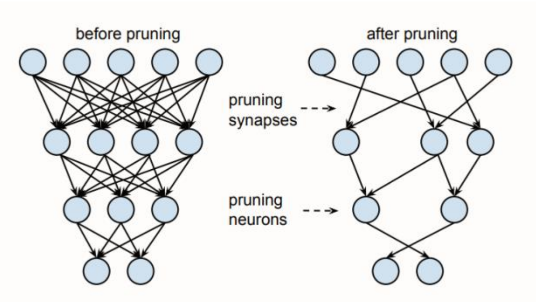


Illustration of pruning process.
Image courtesy of [Han2015]

[1] https://mpeg.chiariglione.org/standards/exploration/digital-representation-neural-networks
[2] https://nips.cc/Conferences/2018/Schedule?showEvent=10941

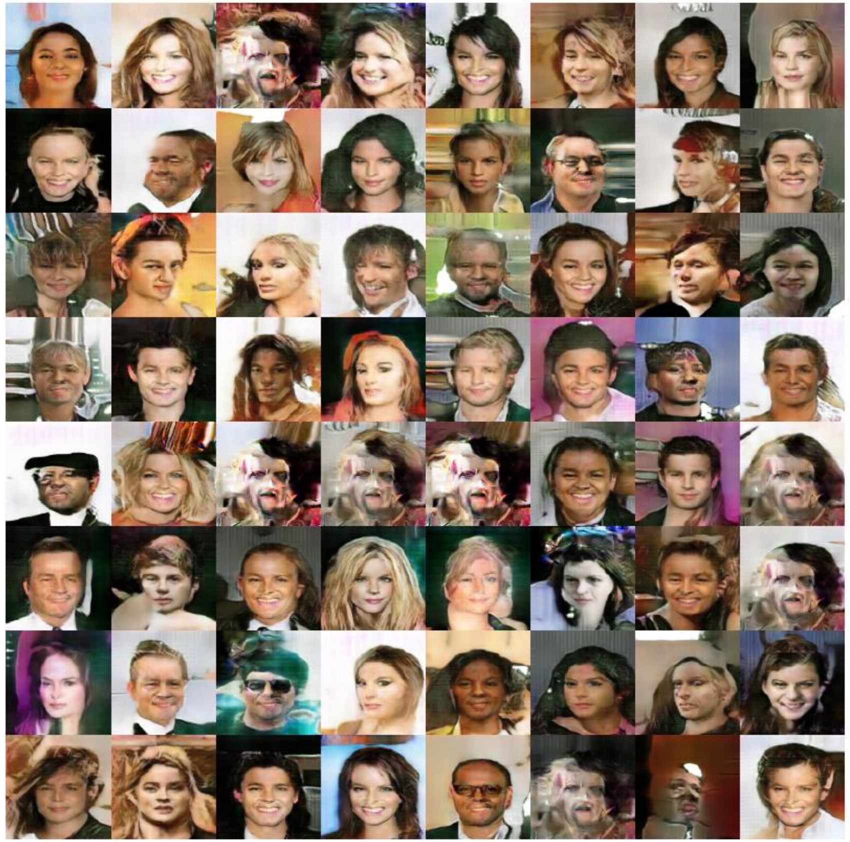
Trends / Outlook

- Software / Hardware
  - Training in the cloud in virtualized instances (Docker)
  - Inference on the edge (mobile phones, 5G base stations, …)
- Algorithmic trends
  - DNNs will continue to ***assimilate*** / **aggregate** successful concepts from the pre-DNN epoch
    - A trous algorithm (undecimated wavelet transform) → dilated convolution
    - Sparsity, transforms (Fourier, Gabor, …), nonlocal / k-NN filtering → DCFNet [Qiu2018], Gabor CNN [Luan2018], NN3D [Cruz2018], Neural Nearest Neighbors Networks [Ploetz2018]
    - Morphological operators, Sinc Filter, Box Filter → PConv [Masci2012], SincNet [Ravanelli2018], [Burkov2018]
    - Normalized cross correlation (NCC) → NCC-Nets [Subramaniam2018]
    - Robust statistics (M-Estimators, outlier rejection, …) → Deep robust regression [Lathuiliere2018]
    - Variational bayesian inference [1] → Bayes by Backprop [Blundell2015], Bayesian CNN [Shridhar2019]
  - More sophisticated optimization algorithms (second order [Bollapragada2018], nonlinear acceleration [Bollapragada2019], loss visualization [Li2019], …)

[1] https://kaybrodersen.github.io/talks/Brodersen_2013_03_22.pdf

Contact

- Interested in our **technologies** and/or **applications** ?
  - Contact me (hannes.fassold@joanneum.at)
  - Or contact Georg Thallinger (head of *Smart Media Solutions* Team) georg.thallinger@joanneum.at)



References

- [Bailer2019] W. Bailer, "Face Swapping for Solving Collateral Privacy Issues", International Conference on MultiMedia Modeling, 2019
- [Burkov2018] E. Burkov, V. Lempitsky, "Deep Neural Networks with Box Convolutions", NIPS, 2018
- [Bollapragada 2018] R. Bollapragada, D. Mudigere, J. Nocedal, H. Shi, P. Tang, "A Progressive Batching L-BFGS Method for Machine Learning", arxiv preprint, 2018
- [Bollapragada 2019] R. Bollapragada, D. Scieur, A. d'Aspremont, "Nonlinear Acceleration of Momentum and Primal-Dual Algorithms", Arxiv preprint, 2018
- [Blundell2015] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, "Weight Uncertainty in Neural Networks", ICML, 2015
- [Cruz2018] C. Cruz, A. Foi, V. Katkovnik, K. Egiazarian, "Nonlocality-Reinforced Convolutional Neural Networks for Image Denoising", IEEE SPL, 2018
- [Duan2017] L. Duan, V. Chandrasekhar, S. Wang, Y. Lou, J. Lin, Y. Bai, „Compact descriptors for video analysis: the emerging MPEG standard", arXiv preprint, 2017
- [Han2015] S. Han, J. Tran, J. Pool, W. Dally, "Learning both Weights and Connections for Efficient Neural Networks", NIPS, 2015
- [Lathuiliere2018] S. Lathuiliere, P. Mesejo, X. Alameda-Pineda, "DeepGUM: Learning Deep Robust Regression with a Gaussian-Uniform Mixture Model", ECCV, 2018
- [Li2019] Li et al, "Visualizing the loss landscape of neural nets", NIPS, 2018
- [Luan2018] S. Luan, C. Chen, B. Zhang, J. Han, J. Liu, "Gabor Convolutional Networks", IEEE TIP, 2018
- [Masci2012] J. Masci, J. Angelo, J. Schmidhuber, „A learning framework for morphological operators using counter-harmonic mean", ISMM, 2012
- [Ploetz2018] T. Ploetz, S. Roth, „ Neural Nearest Neighbors Networks", NeurIPS, 2018
- [Qiu2018] Q. Qiu, X. Cheng, R. Calderbank, G. Sapiro, "DCFNet: Deep Neural Network with Decomposed Convolutional Filters", ICML, 2018
- [Radford2015] A. Radford, L. Metz, S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", CoRR, 2015
- [Ravanelli2018] M. Ravanelli, Y. Bengio, "Interpretable Convolutional Filters with SincNet", NIPS, 2018
- [Redmon2018] J. Redmon, A. Farhadi, "YOLOv3: An incremental improvement", arXiv preprint, 2018
- [Salimans2016], T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, "Improved techniques for training GANs", NIPS, 2016
- [Shridhar2019], K. Shridhar, F. Laumann, M. Liwicki, "A Comprehensive guide to Bayesian Convolutional Neural Network with Variational Inference", arxiv preprint, 2019
- [Schroff2015] F. Schroff, D. Kalenichenko, J. Philbin, "Facenet: A unified embedding for face recognition and clustering", CVPR, 2015
- [Subramaniam2018] A. Subramaniam, A. Mittal, "NCC-Net: Normalized Cross Correlation Based Deep Matcher with Robustness to Illumination Variations", WACV, 2018
- [Zhang2016] K. Zhang, Z. Zhang, Z. Li, and Yu Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks", IEEE SPL, 2016

Acknowledgments

- Thanks to NVIDIA for the technical support and the provided GPUs.
- Thanks to the Hyper360 project partners RBB, Mediaset, Fraunhofer Fokus, Drukka for providing the 360° video sequences for research and development purposes within the project.
- The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 761934 - Hyper360, under grant agreement No. 761802 – MARCONI and under grant agreement No. 732461 - ReCAP
  - http://www.hyper360.eu/
  - https://www.projectmarconi.eu/
  - https://recap-project.com

Hannes Fassold

hannes.fassold@joanneum.at

JOANNEUM RESEARCH
Forschungsgesellschaft mbH

Institute for Information and
Communication Technologies

www.joanneum.at/digital