

Advanced Deep Learning for Computer Vision

The Team

Lecturers



Prof. Dr. Laura
Leal-Taixé



Prof. Dr. Matthias
Niessner

Tutors



Dave
Chen



Ji
Hou



Tim
Meinhardt



Maxim
Maximov

What is this course about

- Presentation of advanced Deep Learning methods for various Computer Vision tasks
- Focus on new methods, some of them presented only this year! There will be extra references, many opportunities for you to dig deeper into the topics
- Research-oriented course

While we go over new methods...

- You have to come up with your own ideas to solve a specific vision problem!
- Strong focus on the practical side: semester-long project where you can put all the knowledge to practice

Course organization

About the lecture

- Theory: 12 lectures
 - Every Monday **10:00-11:30h**
 - Seminar Room, 02.13.010
- Practical:
 - Project to be done in groups of 2 (non-negotiable!)
 - Presentations during the semester
 - Wednesdays **14:00-15:30h** (Seminar Room, 02.09.023)
 - Final poster presentation

<https://dvl.in.tum.de/teaching/adl4cv-ws19/>

Grading system

- Exam: 27th February, 13:30-14:30
- Review: 2 review sessions
- Practical part = $\frac{2}{3}$ of the grade
- Exam = $\frac{1}{3}$ of the grade

<https://dvl.in.tum.de/teaching/adl4cv-ws19/>

Project deadline

- 21.10., today: project presentation
- 23.10.: project assignments (projects <-> TAs)
- 30.10., midnight: deliver a 1 page abstract of your idea for the project.

- Until 6.11.: Evaluation of the project and feedback

Project evaluation

- Presentations: everyone needs to attend!
- First presentation: first results, challenges
 - 04.12.: Groups #1
 - 11.12.: Groups #2

Project evaluation

- Presentations: everyone needs to attend!
- Second presentation: almost final results, new things you tried
 - 08.01.: Groups #1
 - 15.01.: Groups #2

Project evaluation

- Presentations: everyone needs to attend!
- 04.02.: final deadline on report (deadline noon)
 - Max 4 pages using CVPR template
- Final presentation = POSTER
 - Date 05.02. 13:00-16:00

Grading system

- Exam = $1/3$ of the grade
- Practical part = $2/3$ of the grade
 - Presentations (2 oral pres. + 1 poster) = $1/3$
 - Final report = $1/3$
 - Code/submission = $1/3$

Following up with the projects

- Each project will be assigned to a TA and you will have weekly office hours to discuss the progress
- These will be announced after the projects are approved

Slides

- Moodle is set up! Lecture will NOT be recorded.
- Slides will be posted on Moodle and on the website:

<https://dvl.in.tum.de/teaching/adl4cv-ws19/>

- Questions regarding organization of the course:

adl4cv@dvl.in.tum.de

- Emails to our individual addresses will not be answered!

Teams

- Teams of two per project!
- Moodle is set up!
- If you do not have a team
 - Chat after the lecture
 - Post it on Moodle

Project Ideas / Directions

3D Scene Understanding

Ji Hou

Projects Directions

3D Detection/Segemntation/Instance/Completion on various 3D data



Project Directions

- 3D Detection on Single RGB-D Image.
 - Song, Shuran, and Jianxiong Xiao. "Deep sliding shapes for amodal 3d object detection in rgb-d images." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
 - Qi, Charles R., et al. "Frustum pointnets for 3d object detection from rgb-d data." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
 - Qi, Charles R., et al. "Deep Hough Voting for 3D Object Detection in Point Clouds." *arXiv preprint arXiv:1904.09664*(2019).

Project Directions

- Lifting 2D detection to 3D
 - Srivastava, Siddharth, Frederic Jurie, and Gaurav Sharma. "Learning 2D to 3D Lifting for Object Detection in 3D for Autonomous Vehicles." *arXiv preprint arXiv:1904.08494*(2019).
 - Kulkarni, Nilesh, et al. "3D-RelNet: Joint Object and Relational Network for 3D Prediction." *arXiv preprint arXiv:1906.02729*(2019).
 - <http://www.cvlibs.net/datasets/kitti/>

Project Directions

- Instance Segmentation/Completion on 3D reconstruction
 - Hou, Ji, Angela Dai, and Matthias Nießner. "3d-sis: 3d semantic instance segmentation of rgb-d scans." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
 - Hou, Ji, Angela Dai, and Matthias Nießner. "3D-SIC: 3D Semantic Instance Completion for RGB-D Scans." *arXiv preprint arXiv:1904.12012* (2019).

Project Directions

- 3D Detection on Multi-Views
 - Chen, Xiaozi, et al. "Multi-view 3d object detection network for autonomous driving." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
 - Single View + Merging

Project Directions

- How to combine geometry and color (and radar)
 - Dai, Angela, and Matthias Nießner. "3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
 - Jaritz, Maximilian, Jiayuan Gu, and Hao Su. "Multi-view PointNet for 3D Scene Understanding." *arXiv preprint arXiv:1909.13603* (2019).

Project Directions

- 3D Reconstruction from RGB image(s)
 - Choy, Christopher B., et al. "3d-r2n2: A unified approach for single and multi-view 3d object reconstruction." *European conference on computer vision*. Springer, Cham, 2016.
 - Fan, Haoqiang, Hao Su, and Leonidas J. Guibas. "A point set generation network for 3d object reconstruction from a single image." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

3D vision and NLP

Dave Z. Chen

Project Directions

- 3D Cross-modal Retrieval: Bridging the Gap between 3D Objects and Natural Language Descriptions
 - Chen et al. "Text2Shape: Generating Shapes from Natural Language by Learning Joint Embeddings" *ArXiv Preprint*, 2018.
 - Han et al. "Y2Seq2Seq: Cross-Modal Representation Learning for 3D Shape and Text by Joint Reconstruction and Prediction of View and Word Sequences" *The AAAI Conference on Artificial Intelligence*, 2018.
 - Tutor: Dave Z. Chen
 - Contact: zhenyu.chen@tum.de

Project Directions

- Automatic Description Generating for 3D CAD models
 - Xu et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
 - Lu et al. "Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* . 2017.
 - Tutor: Dave Z. Chen
 - Contact: zhenyu.chen@tum.de

Project Directions

- Scan2Cap: Generating descriptions for objects in 3D scenes
 - Xu et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
 - Lu et al. "Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* . 2017.
 - Tutor: Dave Z. Chen
 - Contact: zhenyu.chen@tum.de

Project Directions

- Object Localization in 3D scenes using Natural Language
 - Hu et al. "Natural Language Object Retrieval" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
 - Hu et al. "Segmentation from Natural Language Expressions" *Proceedings of the IEEE European Conference on Computer Vision*. 2016.
 - Tutor: Dave Z. Chen
 - Contact: zhenyu.chen@tum.de

Project Directions

- Grounding referring expressions in 3D scenes with multimodal data
 - Hu et al. "Natural Language Object Retrieval" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
 - Dai, Angela, and Matthias Nießner. "3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation." *Proceedings of the European Conference on Computer Vision*. 2018.
 - Tutor: Dave Z. Chen
 - Contact: zhenyu.chen@tum.de

Segmentation and tracking

Tim Meinhardt

Project Directions

- Video object segmentation (single/multiple objects)



Project Directions

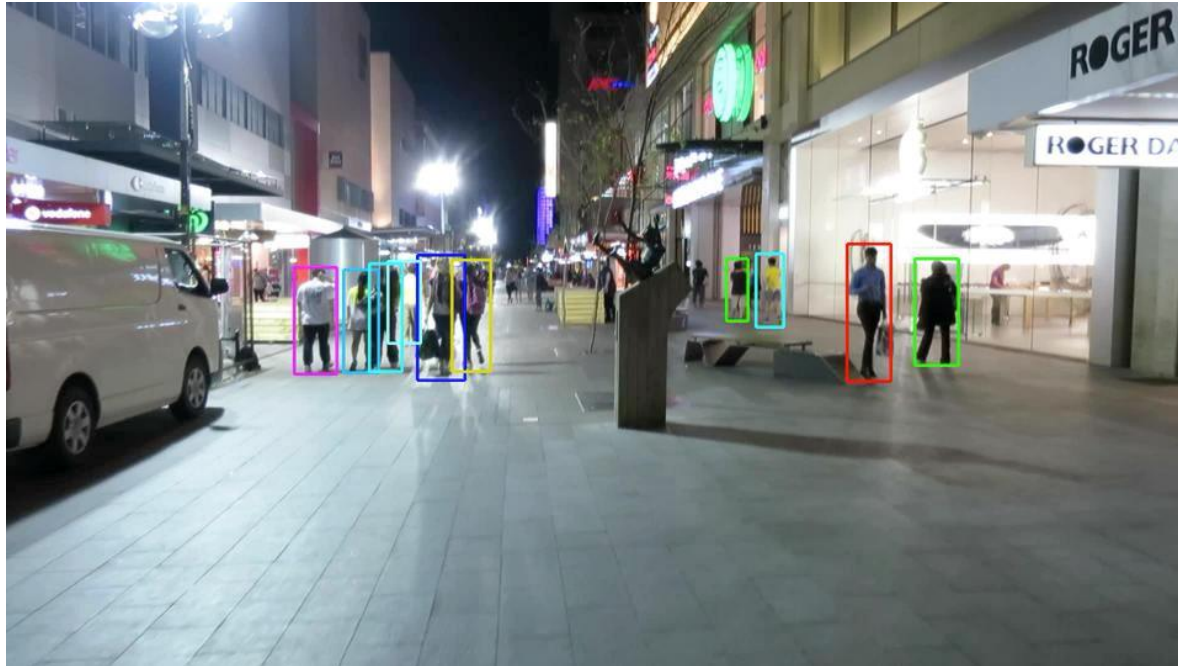
- Video object segmentation (single/multiple objects)
 - Bringing OSVOS to real world pedestrian tracking scenarios:
 - [One-Shot Video Object Segmentation](#). *S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixe, D. Cremers, and L. Van Gool*.
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
 - [Video Object Segmentation Without Temporal Information](#).
K.-K. Maninis, S. Caelles, Y. Chen, J. Pont-Tuset, L. Leal-Taixe, D. Cremers, and L. Van Gool.
Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.
 - Related work:
 - OnAVOS: Online Adaptation of Convolutional Neural Networks for Video Object Segmentation. *P. Voigtlaender, B. Leibe*, BMVC 2017.
 - Datasets: [MOTS: Multi-Object Tracking and Segmentation](#)
Paul Voigtlaender, Michael Krause, Aljoša Ošep, Jonathon Luiten, Berin Balachandar Gnana Sekar, Andreas Geiger, Bastian Leibe. CVPR 2019
- Tutor: Tim Meinhardt
- Contact: tim.meinhardt@tum.de

Project Directions

- Video object segmentation (single/multiple objects)
 - Enhancing OSVOS for multi-object segmentation:
 - Related work:
 - [One-Shot Video Object Segmentation](#). *S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixe, D. Cremers, and L. Van Gool*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
 - [Video Object Segmentation Without Temporal Information](#). *K.-K. Maninis, S. Caelles, Y. Chen, J. Pont-Tuset, L. Leal-Taixe, D. Cremers, and L. Van Gool*. Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.
 - OnAVOS: Online Adaptation of Convolutional Neural Networks for Video Object Segmentation. *P. Voigtlaender, B. Leibe*, BMVC 2017.
 - CINM: CNN in MRF: Video Object Segmentation via Inference in a CNN-Based Higher-Order Spatio-Temporal MRF. *L. Bao, B. Wu, W. Liu*, CVPR 2018.
 - Tutor: Tim Meinhardt
 - Contact: tim.meinhardt@tum.de

Project Directions

- Multiple object tracking in real-world scenarios



Project Directions

- Multiple object tracking in real-world scenarios
 - Meta-learning for:
 - [Tracking without bells and whistles](#).
Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixe.
IEEE International Conference on Computer Vision (ICCV), 2019.
 - Related work:
 - Collaborative Deep Reinforcement Learning for Multi-Object Tracking.
Liangliang Ren, Jiwen Lu, Zifeng Wang¹, Qi Tian, Jie Zhou. ECCV 2018.
- Tutor: Tim Meinhardt
- Contact: tim.meinhardt@tum.de

Project Directions

- Multiple object tracking in real-world scenarios
 - Building an appearance model for:
 - [Tracking without bells and whistles](#).
Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixe.
IEEE International Conference on Computer Vision (ICCV),
2019.
 - Tutor: Tim Meinhardt
 - Contact: tim.meinhardt@tum.de

Image Post-Processing, Rendering and Interpretability

Maxim Maximov

Project Directions

- Topic: Stereo Matching
- Problem: Matching blurry images
- Main Points:
 - How partially blurry images can be matched with sharp ones (for different tasks)
 - Estimate between 2 images: disparity map OR camera localization OR some other metric
- Related Work\Info:
 - [“The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”](#)
 - [“Efficient Deep Learning for Stereo Matching”](#)
 - [“Cascade Residual Learning: A Two-stage Convolutional Neural Network for Stereo Matching”](#)
 - [“DeMoN: Depth and Motion Network for Learning Monocular Stereo”](#)
 - [“Learning Monocular Depth by Distilling Cross-domain Stereo Networks”](#)
 - Other works for stereo matching
- Tutor: Maxim (maxim.maximov@tum.de)



Project Directions

- Post-processing methods: de-blurring, stabilization, stylization etc.

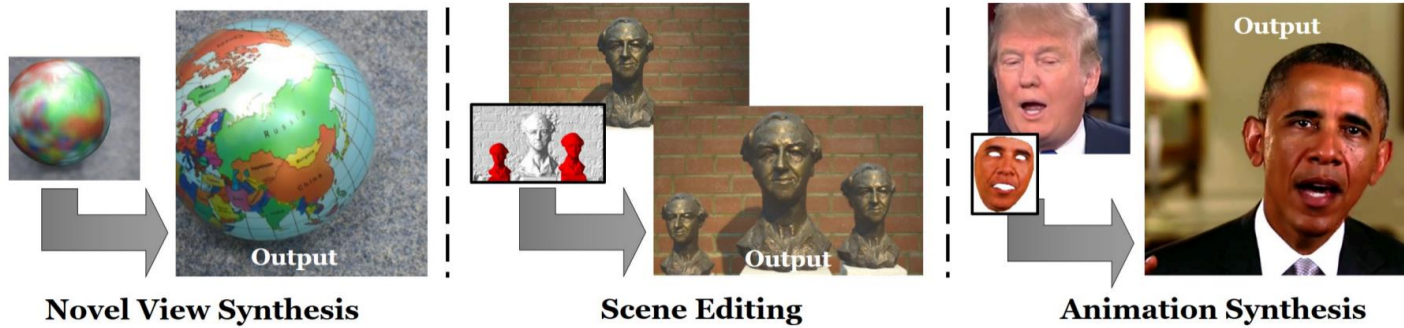


Project Directions

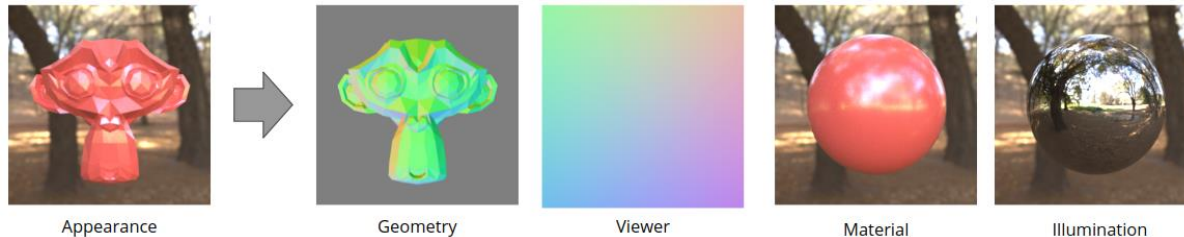
- Topic: Image Processing
- Problem: Video Stabilization
- Main Points:
 - Temporally coherent & sharp
 - Only from videos
- Related Work\Info:
 - "[Burst Image Deblurring Using Permutation Invariant Convolutional Neural Networks](#)"
 - [Google Approach](#)
 - "[Deep Online Video Stabilization](#)"
 - "[Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring](#)"
 - "[DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks](#)"
 - Other motion deblurring papers
- Tutor: Maxim (maxim.maximov@tum.de)

Project Directions

- Use CNN to render 2D Images (Video)



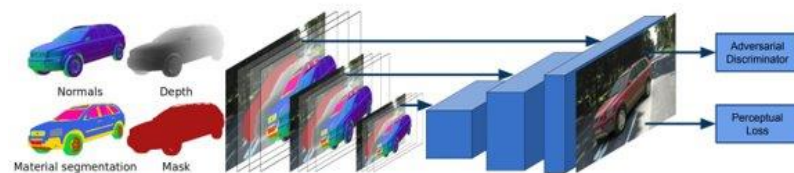
- Estimate components of rendered image



Project Directions

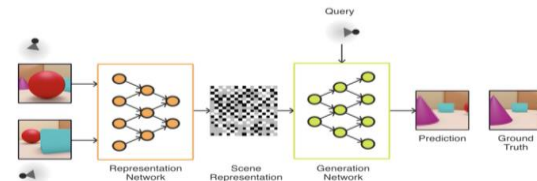
- Topic: Neural Rendering
- Problem: Image Rendering from intermediate renders
- Main Points:
 - “Render” RGB image based on masks, normals, depth or RGB or etc
 - Make it realistic (appearance, shadow)
 - Different options (regular approach, focus on light\shadows, GAN, video)
- Related Work\Info:
 - [“Geometric Image Synthesis”](#),
 - [“Photographic Image Synthesis with Cascaded Refinement Networks”](#),
 - [“IGNOR: Image-guided Neural Object Rendering”](#)
 - [“NVS Machines: Learning Novel View Synthesis with Fine-grained View Control”](#)
 - Other Image synthesis papers

- Tutor: Maxim (maxim.maximov@tum.de)



Project Directions

- Topic: Representation and 3D Scene Understanding
- Problem: Representation of a scene reconstruction network
- Main Points:
 - How to fuse representations from different viewpoints
 - Open topic
- Related Work\Info:
 - "[Neural scene representation and rendering](#)"
 - "[Inverting Visual Representations with Convolutional Networks](#)"
 - "[Learning to Generate Chairs, Tables and Cars with Convolutional Networks](#)"
 - "[Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling](#)"
 - "[Neural Discrete Representation Learning](#)"
 - "[DeepVoxels: Learning Persistent 3D Feature Embeddings](#)"
 - Other 3D Reconstruction papers with latent representation
- Tutor: Maxim (maxim.maximov@tum.de)



Project Directions

- Topic: Inverse-rendering
- Problem: Illumination estimation
- Main Points:
 - Use RGB (+ optionally Depth)
 - How *mirror ball* would look like given an image
 - Or\and estimate shadow map
- Related Work\Info:
 - "[Neural Inverse Rendering of an Indoor Scene from a Single Image](#)"
 - "[DeepLight: Learning Illumination for Unconstrained Mobile Mixed Reality](#)"
 - "[What Is Around The Camera?](#)"
 - "[LIME: Live Intrinsic Material Estimation](#)"
 - "[Neural Inverse Rendering of an Indoor Scene from a Single Image](#)"
 - "[Learning to Reconstruct Shape and Spatially-Varying Reflectance from a Single Image](#)"
 - [AR selfie method](#)
 - Other inverse-rendering papers
- Tutor: Maxim (maxim.maximov@tum.de)

Project Directions

- Topic: Interpretability\Generalization
- Problem: Deep Learning Models Interpretability
- Main Points:
 - Analysis + Visualization
 - Common Problems
 - Open topic
 - Many related work
- Related Work\Info:
 - [Github - pytorch-cnn-vizualization](#)
 - [Building blocks of interpretability](#)
 - [“The elephant in the room”](#), etc
 - Many other papers
- Tutor: Maxim (maxim.maximov@tum.de)



General topics

Fake image generation/detection

- Generative adversarial networks for video generation
 - Vondrick, Carl, Hamed Pirsiavash, and Antonio Torralba. "Generating videos with scene dynamics." *Advances In Neural Information Processing Systems*. 2016
 - Kalchbrenner, Nal, et al. "Video pixel networks." *arXiv preprint arXiv:1610.00527* (2016)
 - Wang, Ting-Chun, et al. "Video-to-Video Synthesis." *arXiv preprint arXiv:1808.06601* (2018).

Fake image generation/detection

- DeepFakes++: forgery generation and detection
 - Rössler, Andreas, et al. "FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces." *arXiv preprint arXiv:1803.09179* (2018).
 - Kim, Hyeongwoo, et al. "Deep Video Portraits." *arXiv preprint arXiv:1805.11714* (2018).
 - DeepFakes

Project Timeline

- Oct 21st Project Introduction (second half of lecture)
- Oct 23rd Projects Assignments (to TAs)
- Oct 30th Abstract Submissions (midnight)
- Until Nov 6th -> Feedback Projects
- 4th + 11th December -> First presentations (group #1, #2)
- 8th + 15th January -> Second presentation (group #1, #2)
- Feb 4th -> Deadline report (noon)
- Feb 5th -> Poster Presentation (Regular ex time slot)

Next lectures

- Monday 4th November, 10-11:30h: Siamese Networks
 - No lecture next week !!! (ICCV)
- This Wednesday, 14-15:30h: Meeting here to assign projects!

See you on
Wednesday 😊