Advanced Deep Learning for Computer Vision

The Team

Lecturers

Prof. Dr. Laura Leal-Taixé

Prof. Dr. Matthias Niessner

Tutors

Dave Chen

Yawar Siddiqui

Guillem Brasó

Ismail Elezi

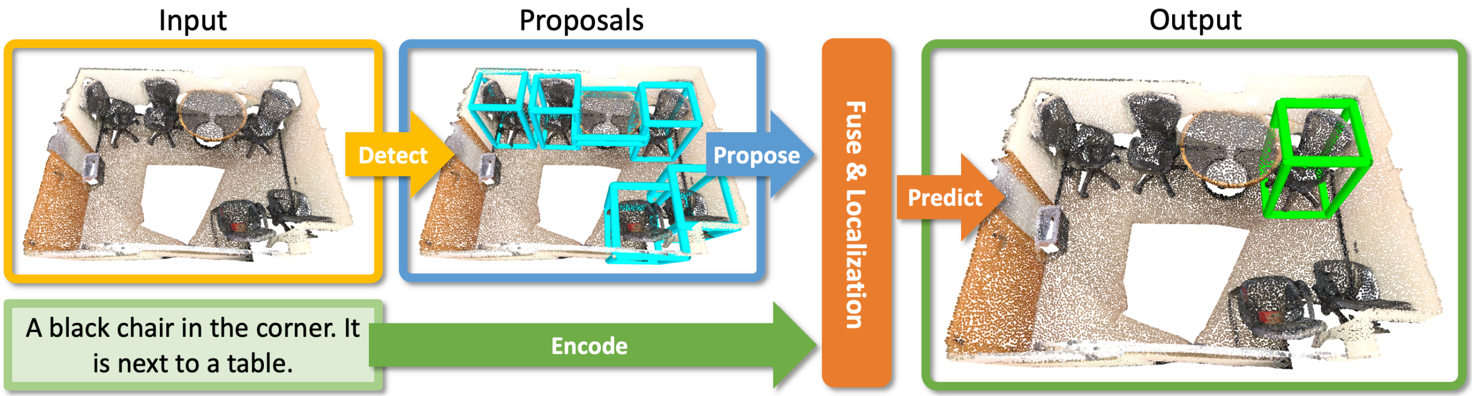
Project Ideas

3D vision and NLP

Dave Z. Chen

Project Directions

- TASK: 3D Visual Grounding

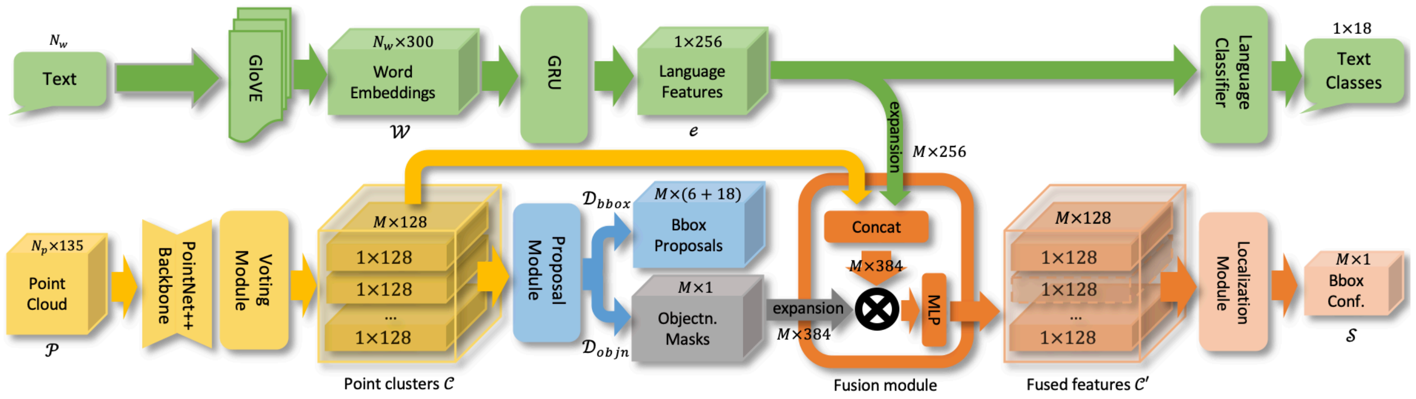


  - Input: a 3D scene + a description for an object
  - Output: a bounding box for the described object

  - Chen et al. "ScanRefer: 3D Object Localization in RGB-D Scans using Natural Language " *The 16th European Conference on Computer Vision (ECCV). 2020.*

Project Directions

- 3D Visual Grounding with SparseConv

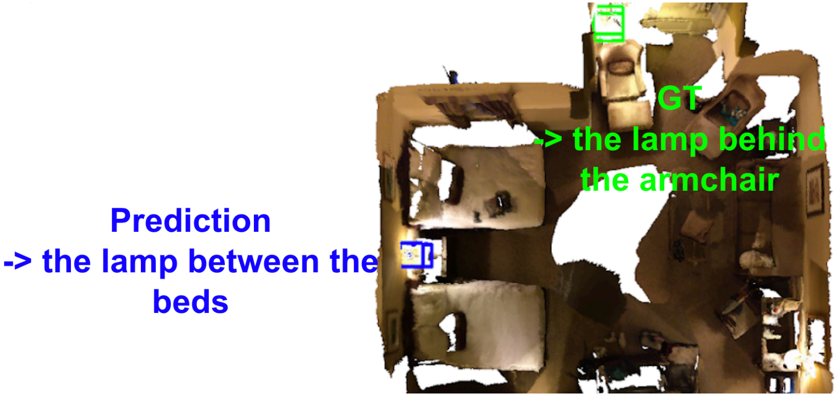


– Jiang et al. "PointGroup: Dual-Set Point Grouping for 3D Instance Segmentation" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2020.

– Tutor: Dave Z. Chen
– Contact: zhenyu.chen@tum.de

Project Directions

- 3D Visual Grounding with graph and attention



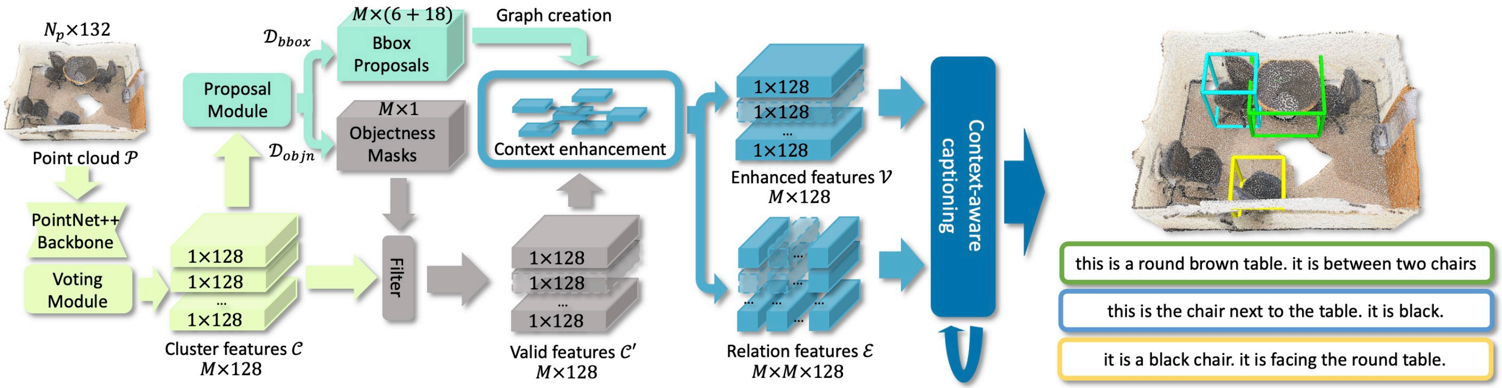
Query: a tall reading lamp sitting in the reading nook behind the armchair. it is turned on and is on the left side of the nook.

- Wang et al. " Neighbourhood Watch: Referring Expression Comprehension via Language-guided Graph Attention Networks " *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2017.

- Tutor: Dave Z. Chen
- Contact: zhenyu.chen@tum.de

Project Directions

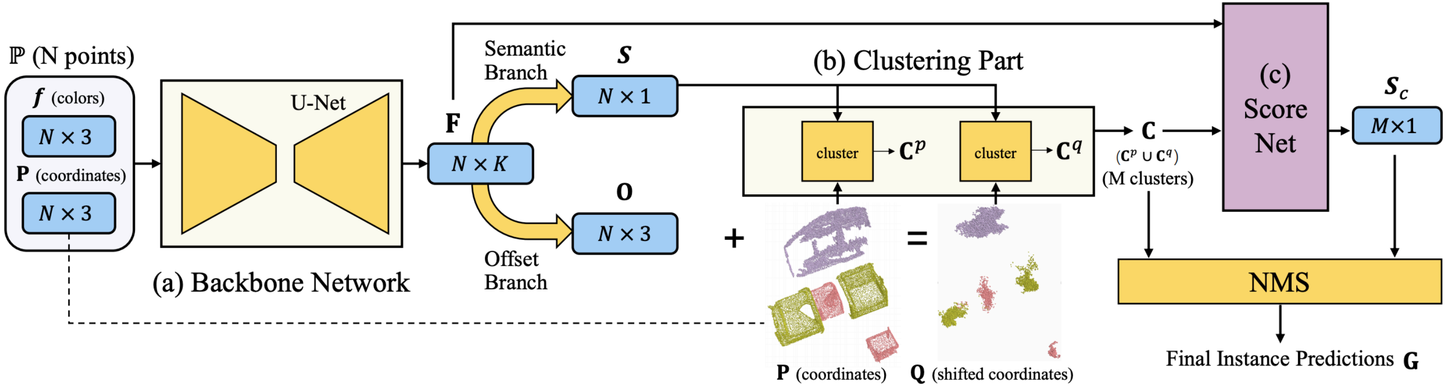
- TASK: Dense Captioning for 3D scenes



- Input: a 3D scene
- Output: bounding boxes and descriptions for all objects

Project Directions

- Dense Captioning for 3D scenes with SparseConv

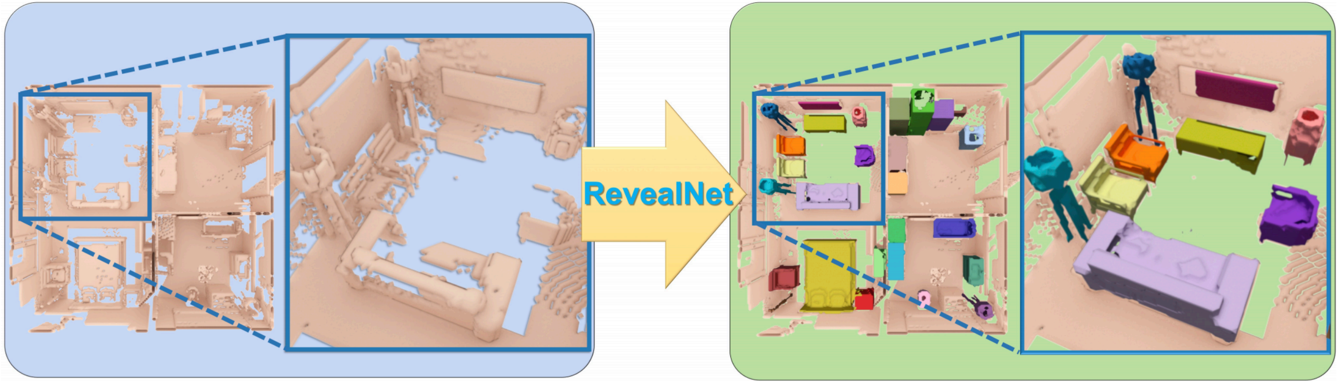


- Jiang et al. "PointGroup: Dual-Set Point Grouping for 3D Instance Segmentation" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2020.

- Tutor: Dave Z. Chen
- Contact: zhenyu.chen@tum.de

Project Directions

- Dense Captioning for 3D scenes with Completion



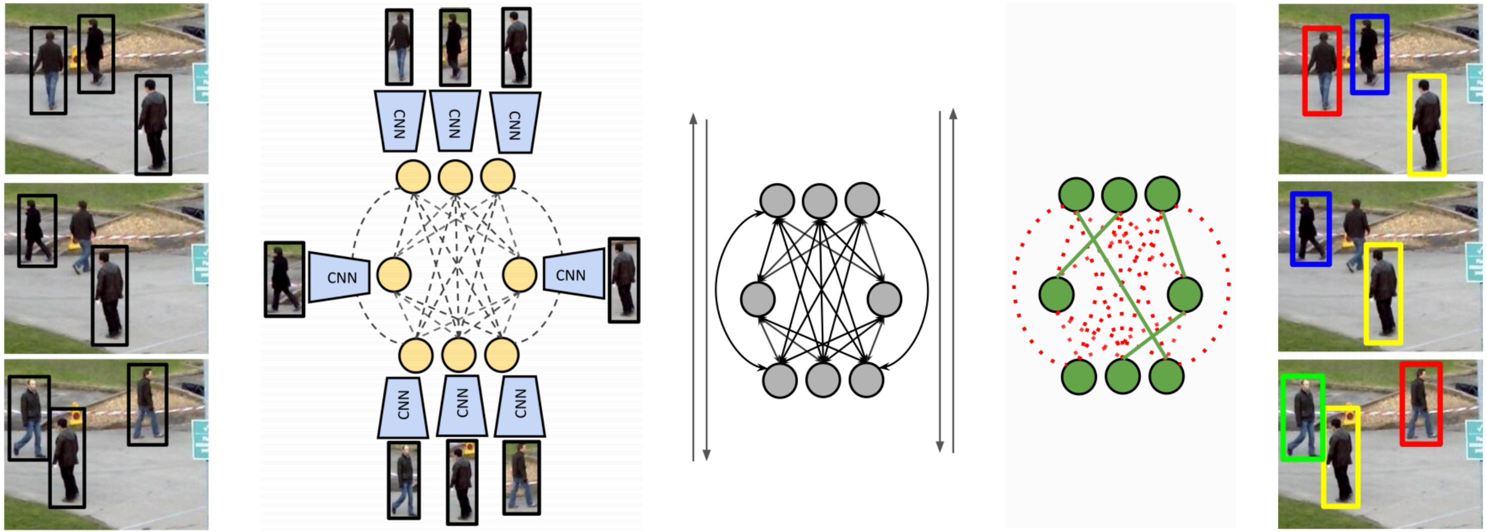
  - Hou et al. "RevealNet: Seeing Behind Objects in RGB-D Scans" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2020.

  - Tutor: Dave Z. Chen
  - Contact: zhenyu.chen@tum.de

Graph-structured representations for Video Understanding

Guillem Brasó

Tracking with MPNs

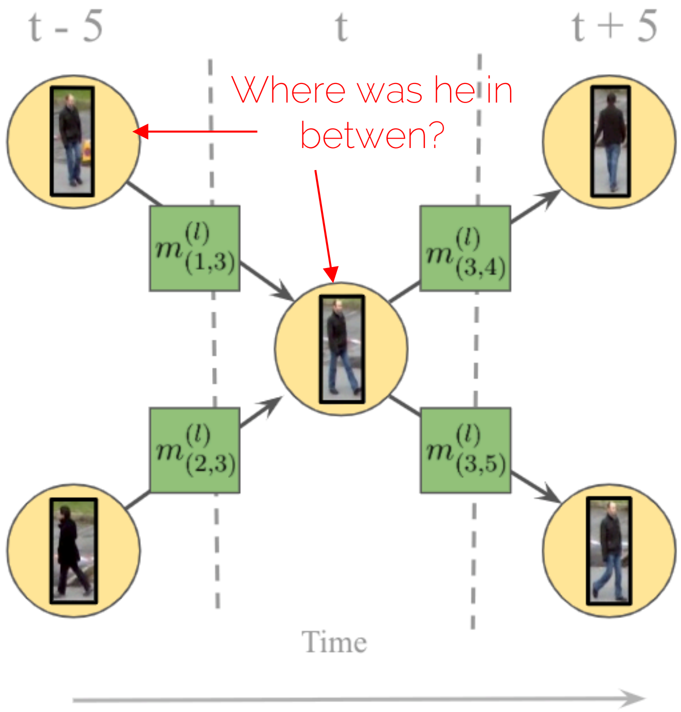


(a) Input (b) Graph Construction + Feature Encoding (c) Neural Message Passing (d) Edge Classification (e) Output

Brasó and Leal-Taixé. "Learning a Neural Solver for Multiple Object Tracking". *CVPR 2020.*

Project Directions

- Project 1 : Learning a Motion Model for MOT

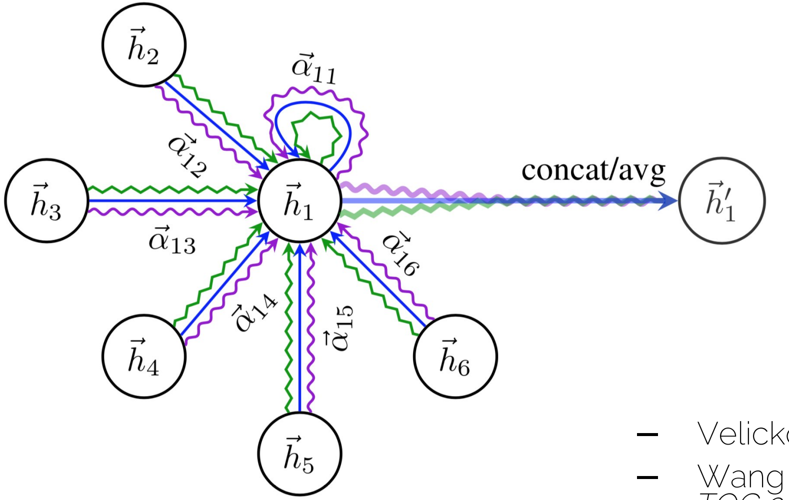


- <u>Main idea</u>: utilize node embeddings produced by the MPN to predict the locations of objects in missing frames.
- Could be used to make our tracker online.

– Yu et al. "Spatio-Temporal Graph Transformer Networks for Pedestrian Trajectory Prediction". *ECCV 2020.*

Project Directions

- **Project 2:** Learning a better architecture for MOT

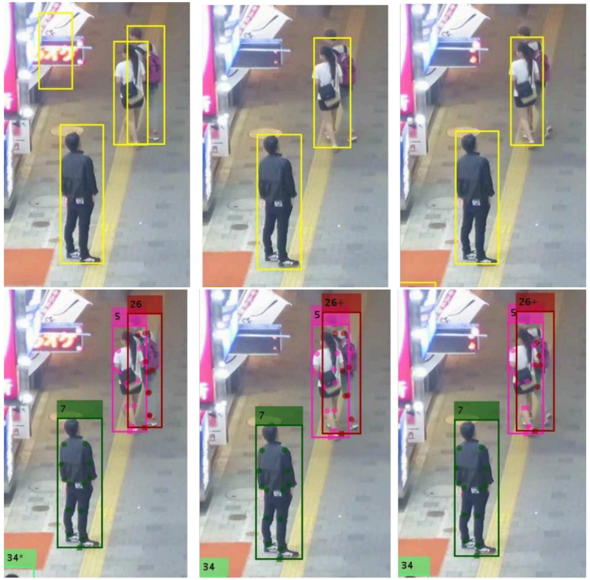


- Some "architecture" choices to explore:
  - Attention Mechanisms
  - Dynamic Connectivity
  - Hierarchical aggregation

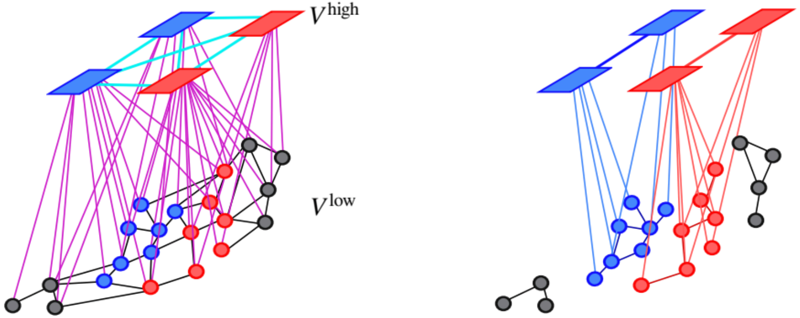
- Velickovic et al. "Graph Attention Networks". *ICLR 2020.*
- Wang et al. "Dynamic Graph CNN for Learning on Point Clouds". *ACM TOG 2019.*
- Ying et al. "Hierarchical Graph Representation Learning with Differentiable Pooling". *Neurips 2018.*

Project Directions

- **Project 3:** Exploiting body joints/low-level features for MOT



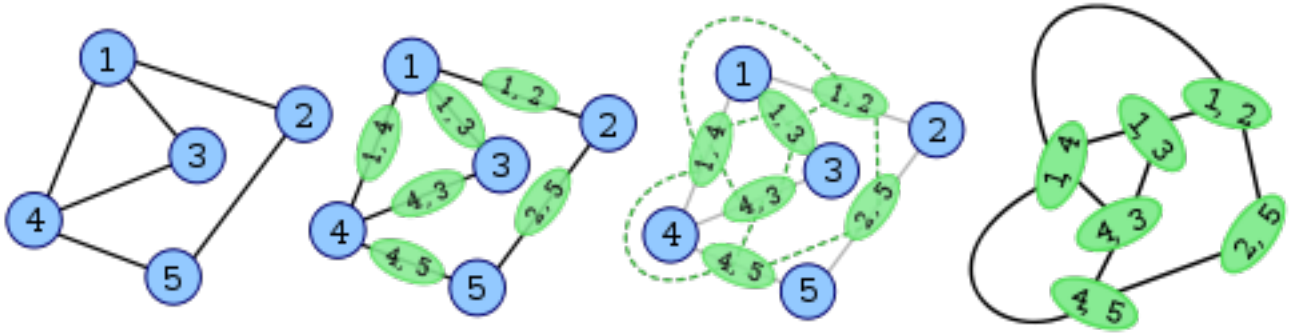
frame 19 frame 22 frame 25



- Henschel et al. "Multiple Object Tracking Using Body and Joint Detections". *ICLR 2020.*
- Keuper et al. "Motion Segmentation & Multiple Object Tracking by Correlation Co-Clustering.". *TPAMI 2020.*
- Revaud et al. "DeepMatching: Hierarchical Deformable Dense Matching". *IJCV 2016.*

Project Directions

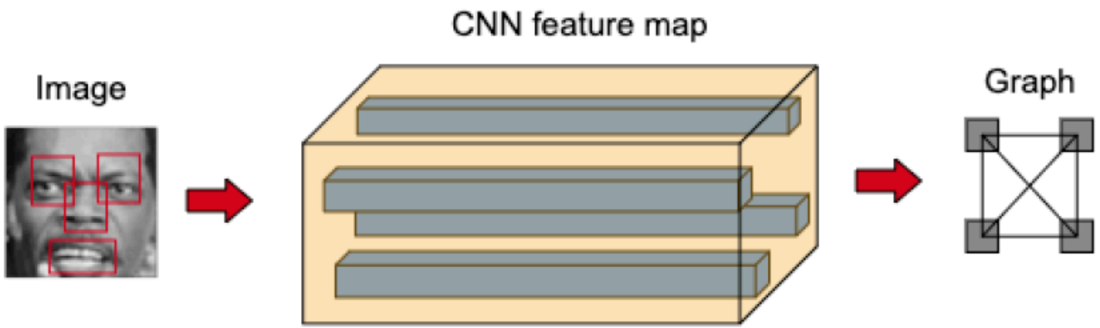
- **Project 4**: Exploring alternative graph formulations for MPN-based MOT



  - Monti et al. "Dual-Primal Graph Convolutional Networks". *arXiv 2018.*
  - Brendel et al. "Multi-Object Tracking as Maximum Weight Independent Set.". *CVPR 2011.*
  - Tang et al. "Subgraph Decomposition for Multi-Target Tracking". *CVPR 2015.*

Project Directions

- **Project 5:** Graph-structured representations for emotion recognition from videos.



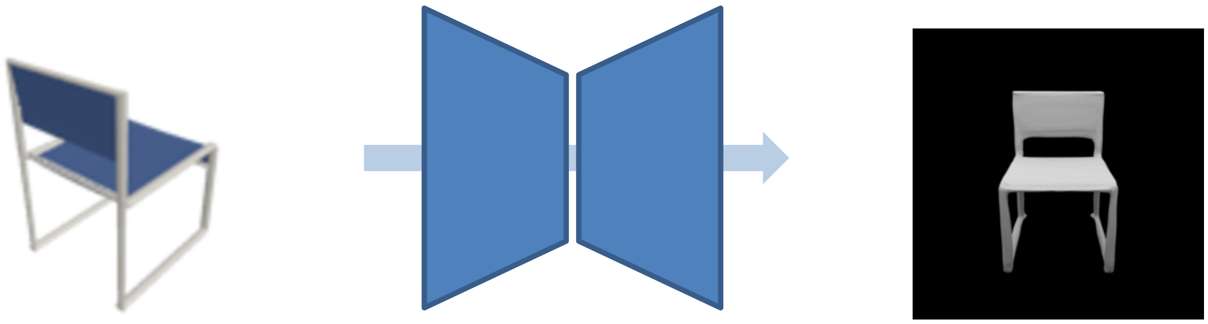
Main idea: Exploit interactions among different face parts across time in order to build better models for emotion recognition

3D Shape & Scene Understanding

Yawar Siddiqui

Project Directions

- 3D Reconstruction from Single RGB Image
  - Going from a 2D image to a 3D mesh



  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

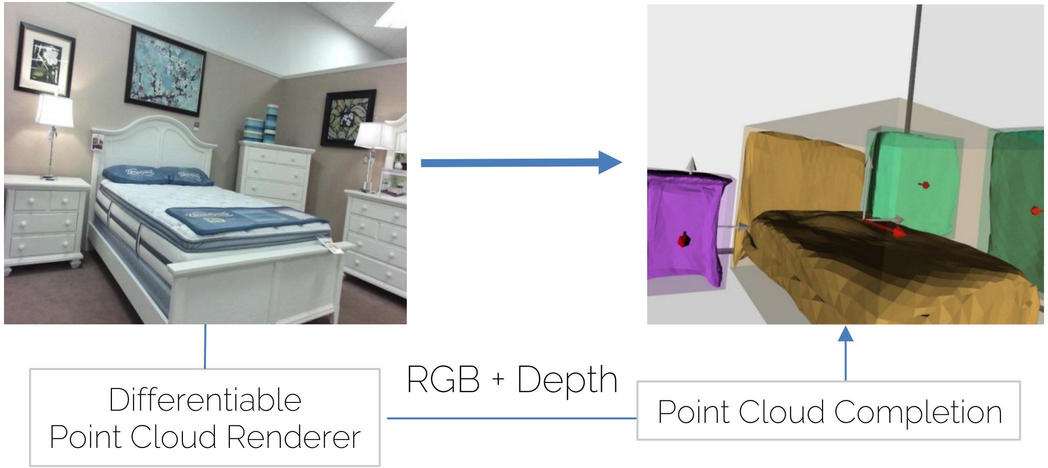
Project Directions

- 3D Reconstruction from Single RGB Image
  - Going from a 2D image to a 3D mesh

  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Project Directions

- 3D Reconstruction from Single RGB Image
  - Going from a 2D image to a 3D mesh



  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

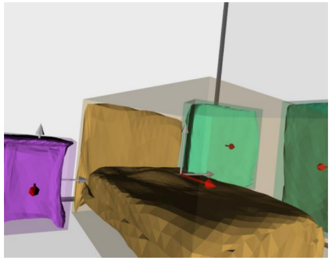
Project Directions

- 3D Reconstruction from Single RGB Image
  - Going from a 2D image to a 3D mesh

  - Related Work:
    - Nie et al., Total3DUnderstanding: Joint Layout, Object Pose and Mesh Reconstruction for Indoor Scenes from a Single Image. CVPR 2020
    - Denninger et al., 3D Scene Reconstruction from a Single Viewport, ECCV 2020
    - Gkioxari et al., Mesh R-CNN, ICCV2019

  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Input RGB



3D Scene

Depth Prediction

RGB + Depth

Point Cloud Completion

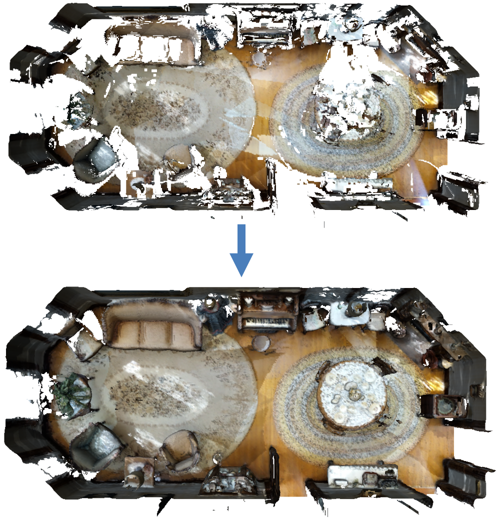
Project Directions

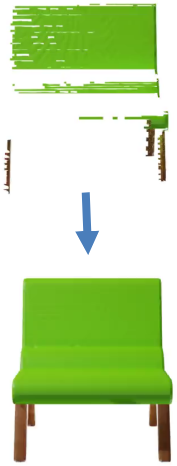
- Shape & Texture Completion from RGBD Image(s)
  - Given an incomplete colored 3D shape, complete both the shape and color

Incomplete 3D Geometry & Color

↓

Completed 3D Geometry & Color



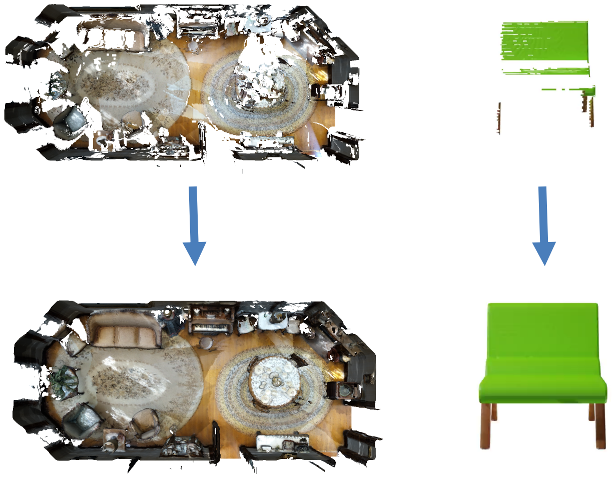


  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Project Directions

- Shape & Texture Completion from RGBD Image(s)
  - Given an incomplete colored 3D shape, complete **both** the shape and color
  - Related Work:
    - Dai et al., SPSG: Self-Supervised Photometric Scene Generation from RGB-D Scans, arxiv 2020
    - Oechsle et al., Texture Fields: Learning Texture Representations in Function Space, ICCV2019
    - Chibane et al., Implicit Feature Networks for Texture Completion from Partial 3D Data, ECCV2020
  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

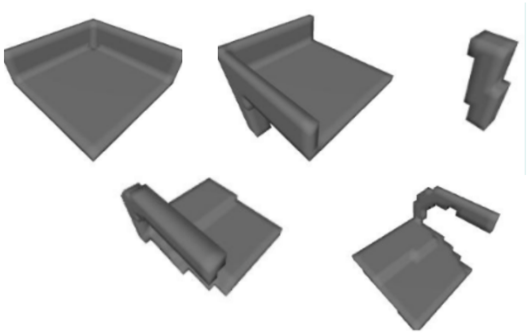
Incomplete Colored Scene/Shape



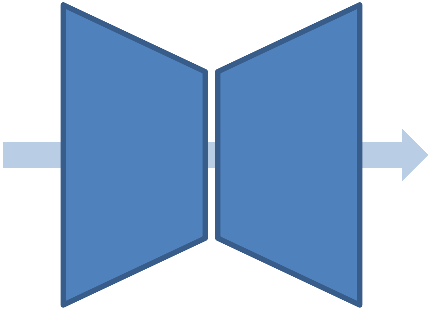
Completed Colored Scene/Shape

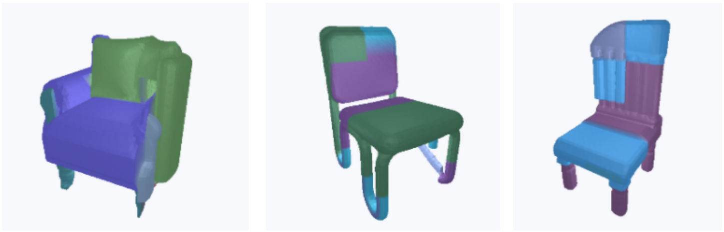
Project Directions

- Shape Synthesis by Patch Recombination
  - Given a database of patches - retrieve, transform, and synthesize joints between these patches to create new shapes



Patches



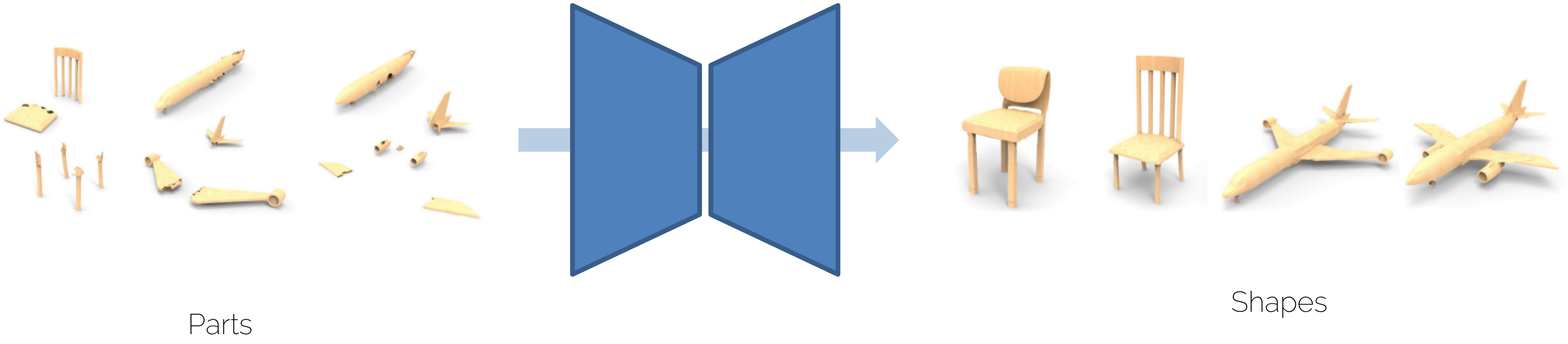


Shapes

  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Project Directions

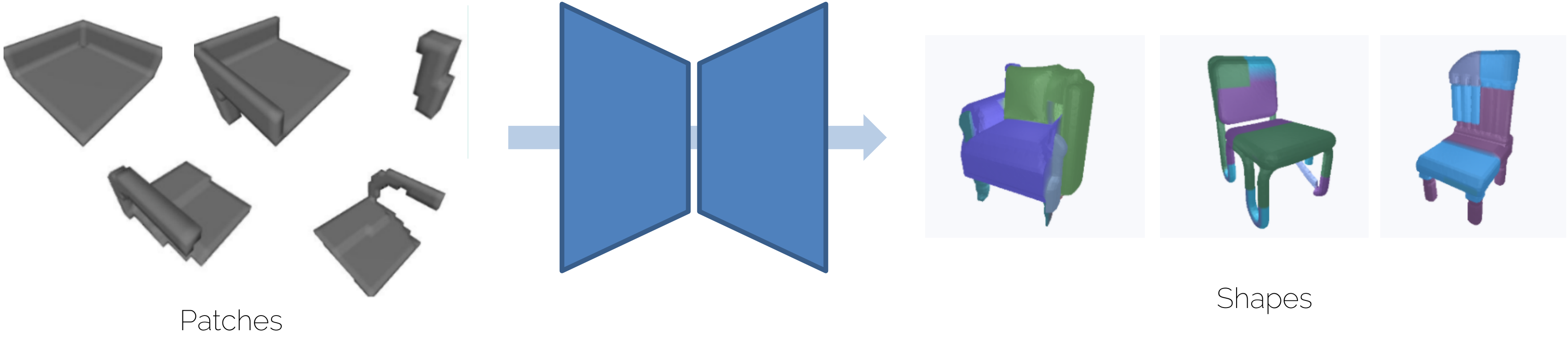
- Shape Synthesis by Patch Recombination
  - Given a database of patches - retrieve, transform, and synthesize joints between these patches to create new shapes



  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Project Directions

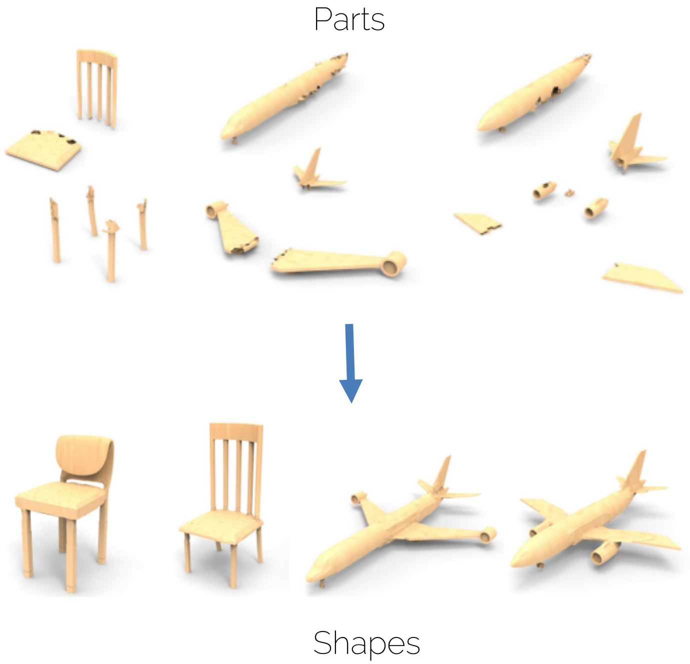
- Shape Synthesis by Patch Recombination
  - Given a database of patches - retrieve, transform, and synthesize joints between these patches to create new shapes



  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

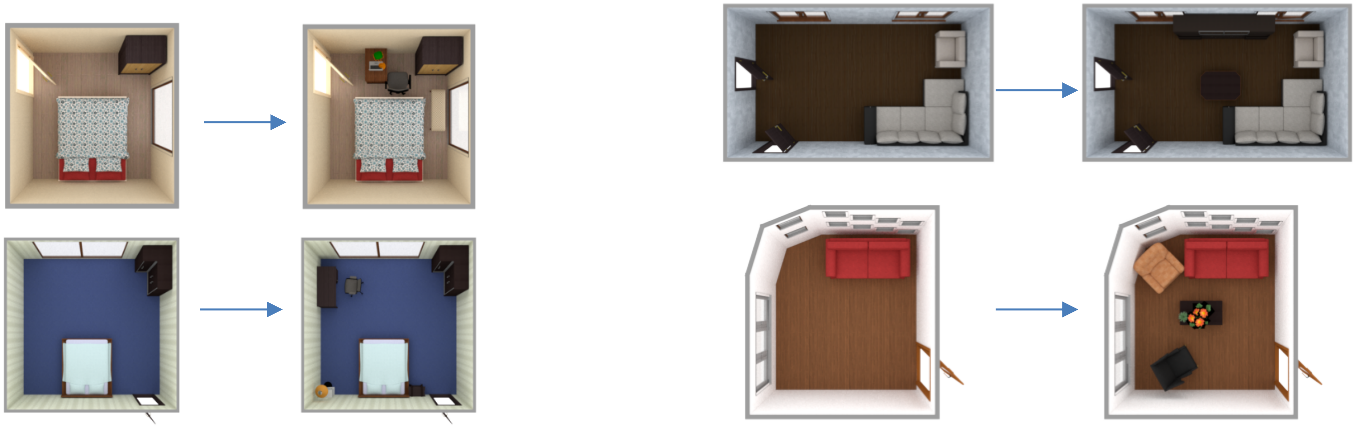
Project Directions

- Shape Synthesis by Patch Recombination
  - Given a database of patches - retrieve, transform, and synthesize joints between these patches to create new shapes
  - Related Work in "Part" domain:
    - K. Yin et al., COALESCE: Component Assembly by Learning to Synthesize Connections, arxiv 2020
    - Dubrovina et al., Composite Shape Modeling via Latent Space Factorization, ICCV2019
    - Zhu et al., SCORES: Shape Composition with Recursive Substructure Priors, TOG2018
  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)



Project Directions

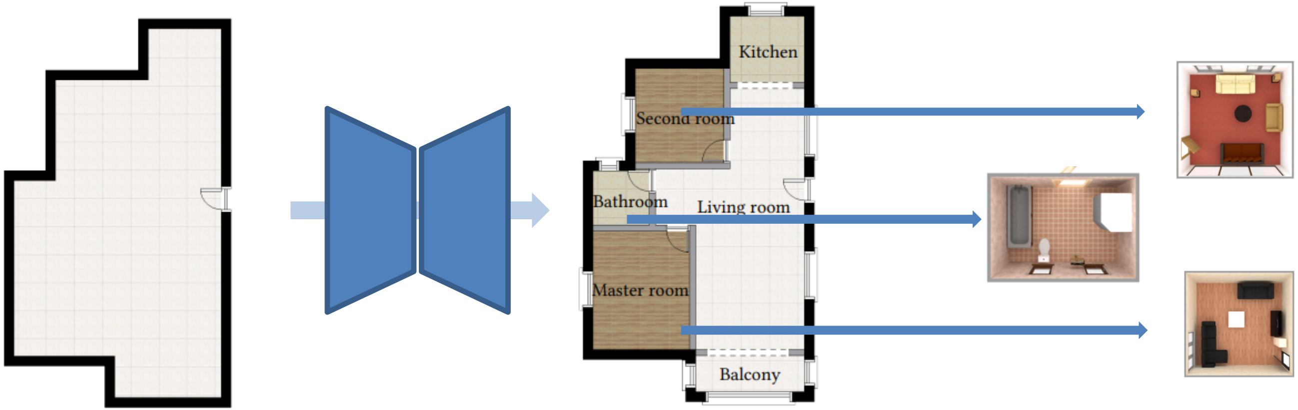
- Indoor Scene Synthesis
  - Learning to synthesize novel indoor scenes



  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Project Directions

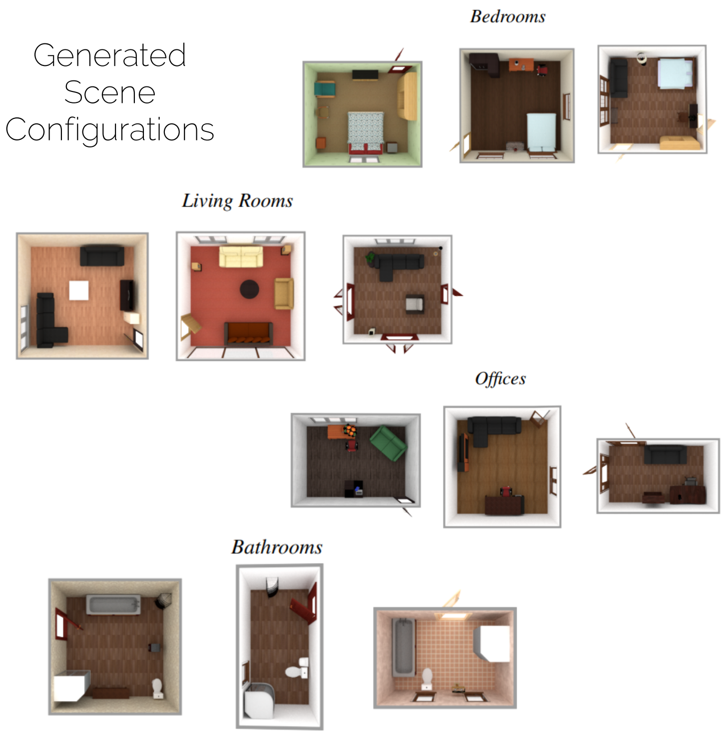
- Indoor Scene Synthesis
  - Learning to synthesize novel indoor scenes



  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)

Project Directions

- Indoor Scene Synthesis
  - Learning to synthesize novel indoor scenes
  - Related Work:
    - Ritchie et al., Fast and Flexible Indoor Scene Synthesis via Deep Convolutional Generative Models, CVPR2019
    - Keshvarzi et al, SceneGen: Generative Contextual Scene Augmentation using Scene Graph Priors, arxiv 2020
    - Wu et al., Data-driven interior plan generation for residential buildings, TOG2019
  - Tutor: Yawar Siddiqui (yawarnihal@gmail.com)



Project Ideas for ADL course

Tutor: Ismail Elezi - postdoc in DVL

Areas of research

1) Deep Metric Learning
2) Semi-Supervised Learning
3) Active Learning
4) Generative Adversarial Networks

NB: You are free to come up with your ideas of research in these topics. Feel free to ping me in order to arrange a chat so we can talk about your project idea.

Semi-Supervised Learning 1

Most of semi-supervised learning models are based on entropy regularization (alter one image, and see if the predictions of the real and altered image match). For object detection, such a paper is:

Jeong et al., Consistency-based semi-supervised learning for object detection, NeurIPS 2019

The altering of the image is done by simply flipping it. In this project, we want to go deeper by also adding Resizing and Aggressive Augmentations (RandAug).

Semi-Supervised Learning 2

We recently proposed a loss function for the task of metric learning.

We would like to explore if an idea based on Group Loss works for semi-supervised learning. In particular, we would like to make different versions of strong augmentations (similar to MixMatch, FixMatch) and classify them via Group Loss.

Elezi, …, Leal-Taixe, The Group Loss for Deep Metric Learning, ECCV 2020

Semi-Supervised Learning 3

Similar to the previous project, but now instead of using The Group Loss, we will use Message Passing Networks (MPNs) [1]. The work can be extended to Domain Adaptation.

During inference, we might need to explore different strategies:

a) use MPN with mini-batches in inference.
b) use a Teacher-Student model [2,3].

[1] *Battaglia et al., Relational Inductive biases, deep learning and neural networks. Arxiv 2018.*
[2] Hinton et al., Distilling the Knowledge in a Neural Network, NIPS workshops 2014
[3] Park et al., Relational Knowledge Distillation, CVPR 2019

Active Learning 1

The mainstream approach of doing active learning, is to train a classifier (neural network) and then to select for labeling the points which have high entropy.

We would like to add a Message Passing Network in the pipeline, where samples will be classified also based on the other samples. Then on inference, we choose the samples with the highest entropy.

During inference we either use a teacher-student model, or we might learn to predict the MPN loss as done in [1]

Active Learning 2

In this project, we want to explore a different approach towards active learning (applied to classification). The approach is similar to [1]. The main idea:

1) Train a network in the AL pool with a metric learning loss in addition to cross-entropy.
2) Find samples that belong to dense regions which contain many unconfident samples.
3) Label those samples.

Generative Adversarial Networks

In our recently proposed CIAGAN model, we achieved state-of-the-art face anonymization results. In this project, we would like to explore some guided ways of altering parts of the faces (for example, eyes, mouth etc) by using VAE or exploiting some smooth latent space.

Maximov, Elezi*, Leal-Taixe, CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks, CVPR 2020*

Next steps

- 06.11-12.11.: read the papers related to the projects you are interested in (pick more than 1 since it is hard everyone gets to be assigned to their favourite project).
- 13.11.: project assignments (projects <-> tutors) via an interactive zoom session
- 13.11-22.11: schedule meetings with the tutors to give shape to the project and make concrete plans.
- 23.11., midnight: deliver a 1 page abstract of your idea for the project.

Advanced Deep Learning for Computer Vision