Graph Deep Learning in Medical Imaging (2022SoSe)

Introduction to Semantic Scene Graphs

21.06.2022

Chair for Computer Aided Medical Procedures and Augmented Reality

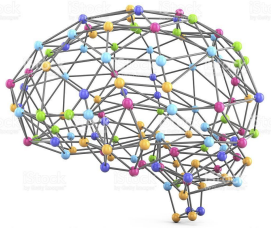
Technische Universität München + Johns Hopkins Whiting School of Engineering

Ege Özsoy
PhD Candidate
ege.oezsoy@tum.de

Felix Holm
PhD Candidate
felix.holm@tum.de

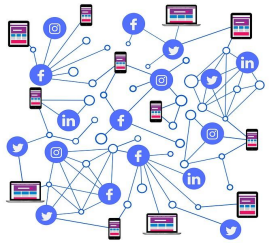
What is a Scene Graph vs. other Graphs?

Graph



Brain connection

Chemistry



Social network

GDL Methods

GraphSage,
GAT,
etc.

Result

Classification,
Regression,
Updated Graph,
etc.

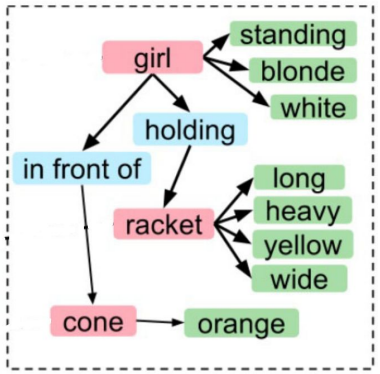
Image sources: istockphoto.com

What is a Scene Graph vs. other Graphs?

Visual Scene

Scene Graph Generation

Scene Graph



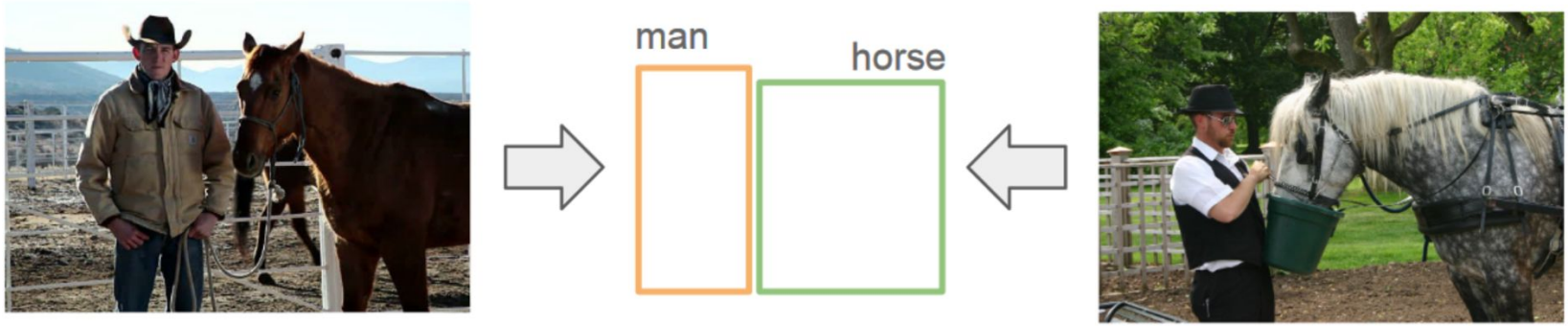
GDL Methods

Classification, Regression, Updated Graph, etc.

- Scene Graphs are representations of Visual Scenes (Images, Videos, Pointclouds)
- Graph based methods can be applied to them

R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Why represent Scenes as Graphs?



- Today's visual models are narrow (i.e. task-specific) and not able to extract general or more complex information from scenes

D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. "Scene Graph Generation by Iterative Message Passing". IEEE 2017.

Example: Why Scene Graphs?

More complex task: Recognize “woman holding umbrella”

Image sources: istockphoto.com, dissolve.com, wallpaperup.com, create.vista.com

Example: Why Scene Graphs?

Classifier to predict “woman holding umbrella”

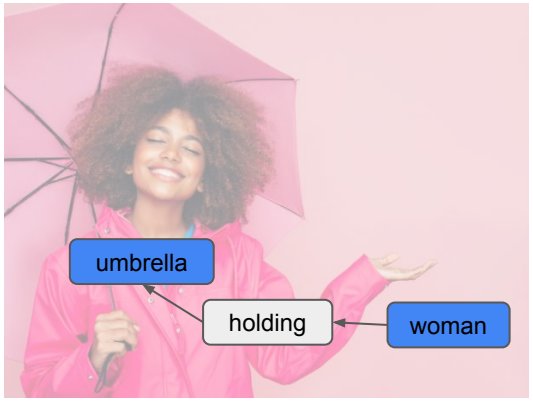
Image sources: istockphoto.com, dissolve.com, wallpaperup.com, create.vista.com

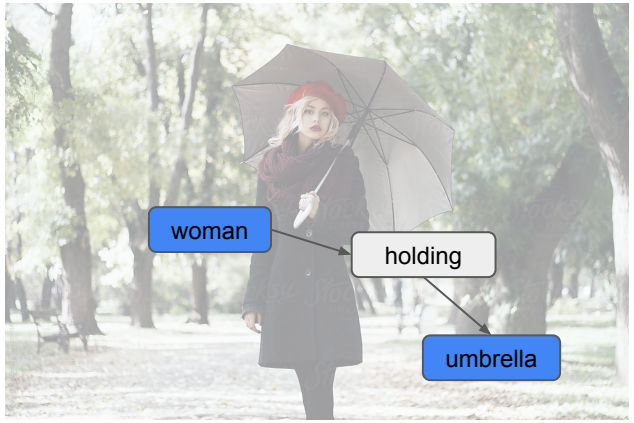
Example: Why Scene Graphs?

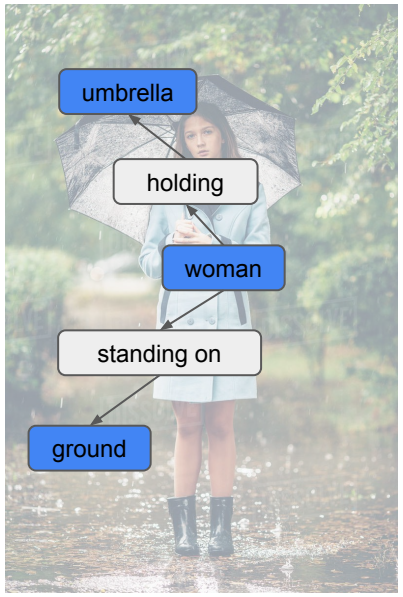
Classifier for “woman holding umbrella” cannot recognize “man holding umbrella” → narrow

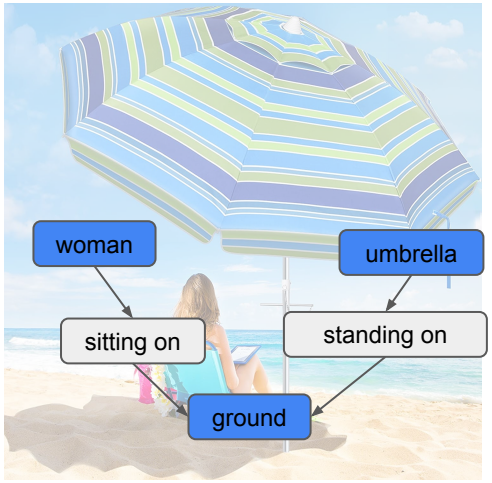
Image sources: istockphoto.com, dissolve.com, wallpaperup.com, create.vista.com

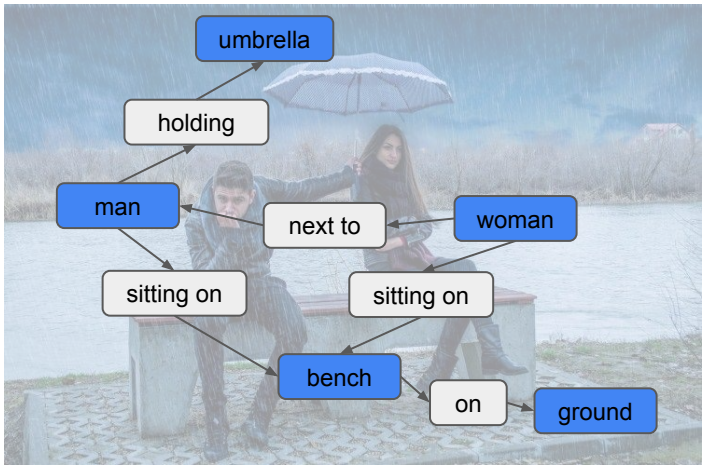
Example: Why Scene Graphs?

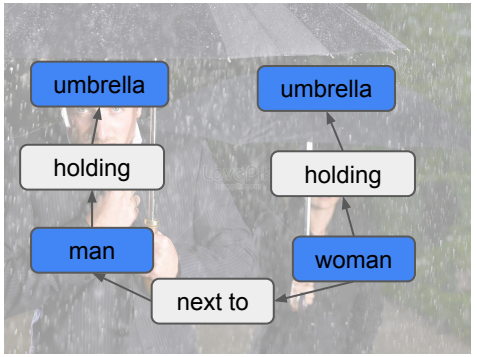












More aspects of the image are described to solve more complex tasks

object

relationship

Image sources: istockphoto.com, dissolve.com, wallpaperup.com, create.vista.com

Why not Natural Language?

A girl with curly hair has a pink jacket on. Her umbrella is matching her jacket. She is holding it over her shoulder.

A girl with blonde hair and a red hat is carrying a gray umbrella.

A woman is sitting on the beach. She is using a beach umbrella to protect herself from the sun.

?

?

?

"Woman holding umbrella"?

Major Challenges
- Irrelevant Details and Filler words
- Contextual words
- Ambiguities

Natural Language is semantically rich, but less structured

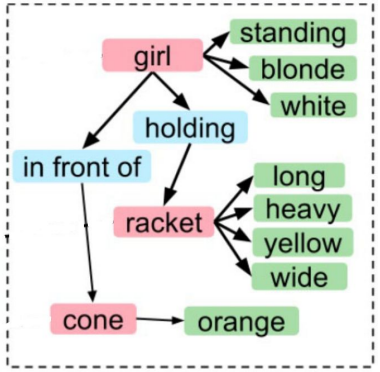
Image sources: istockphoto.com, dissolve.com, wallpaperup.com, create.vista.com

Why Scene Graphs?

Visual Scene

Scene Graph Generation

Scene Graph



Task-independent Representation of Visual Scene

- High-level
- Semantically rich

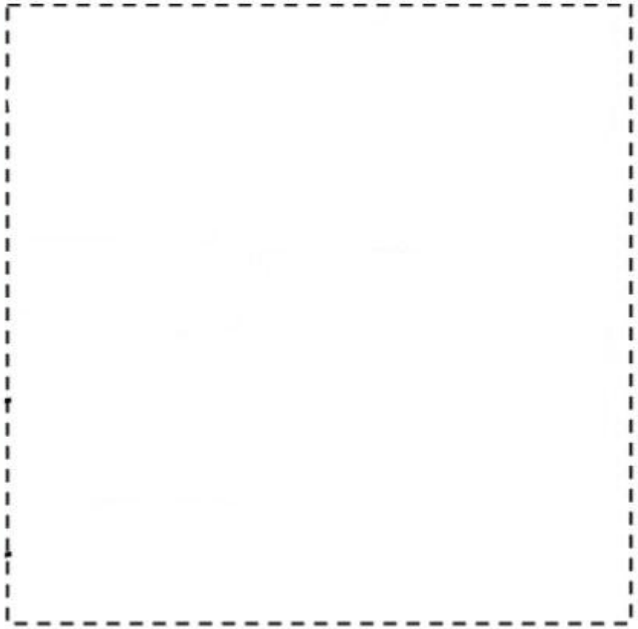
Improve performance on previous tasks
Classification, Regression, etc.

Enable more complex tasks
Image Captioning, Image Retrieval, Visual Question Answering, Relationship modeling, Image Generation

R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Scene Graphs

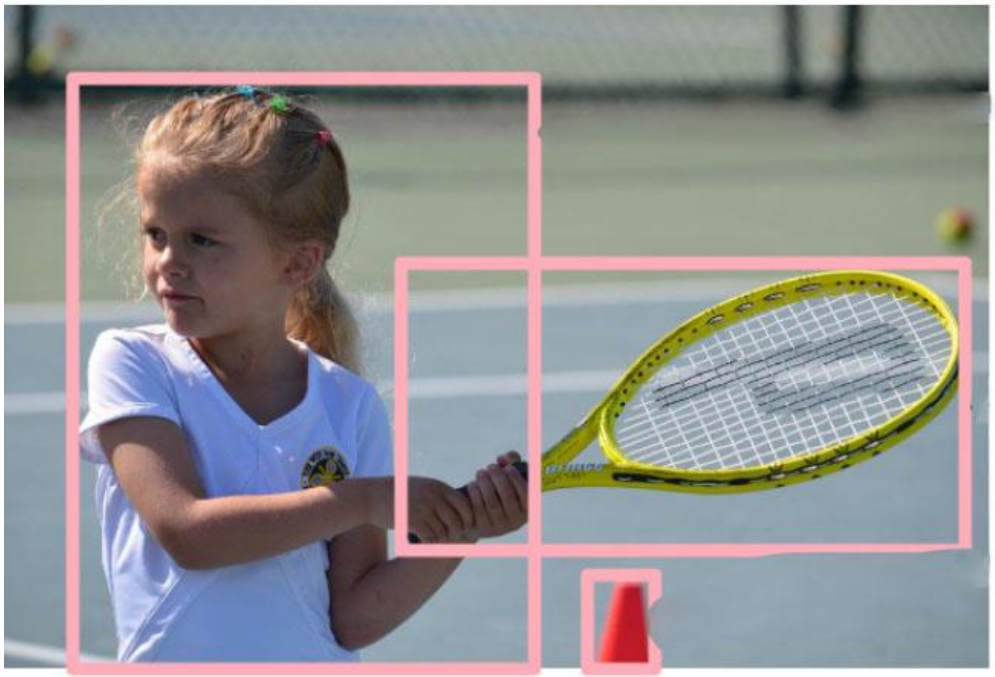
- Describes:
  - objects in a scene (nodes)
  - the relationships between them (edges)

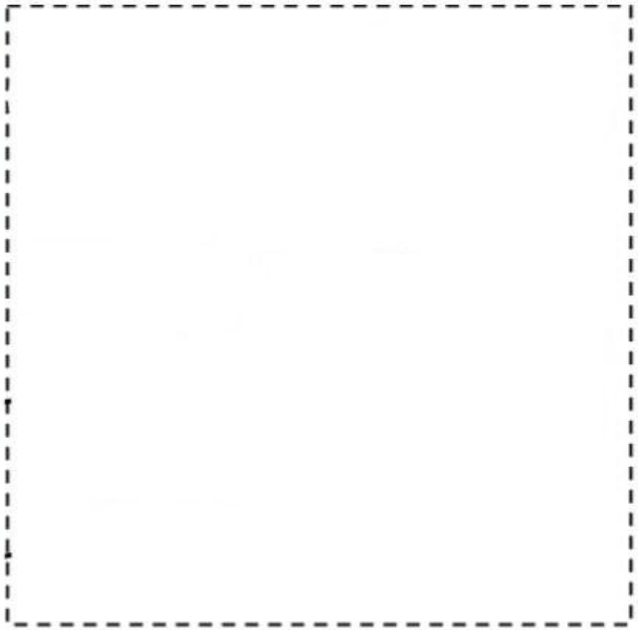


R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Scene Graphs

- Objects as Nodes (from Object Bounding Boxes)

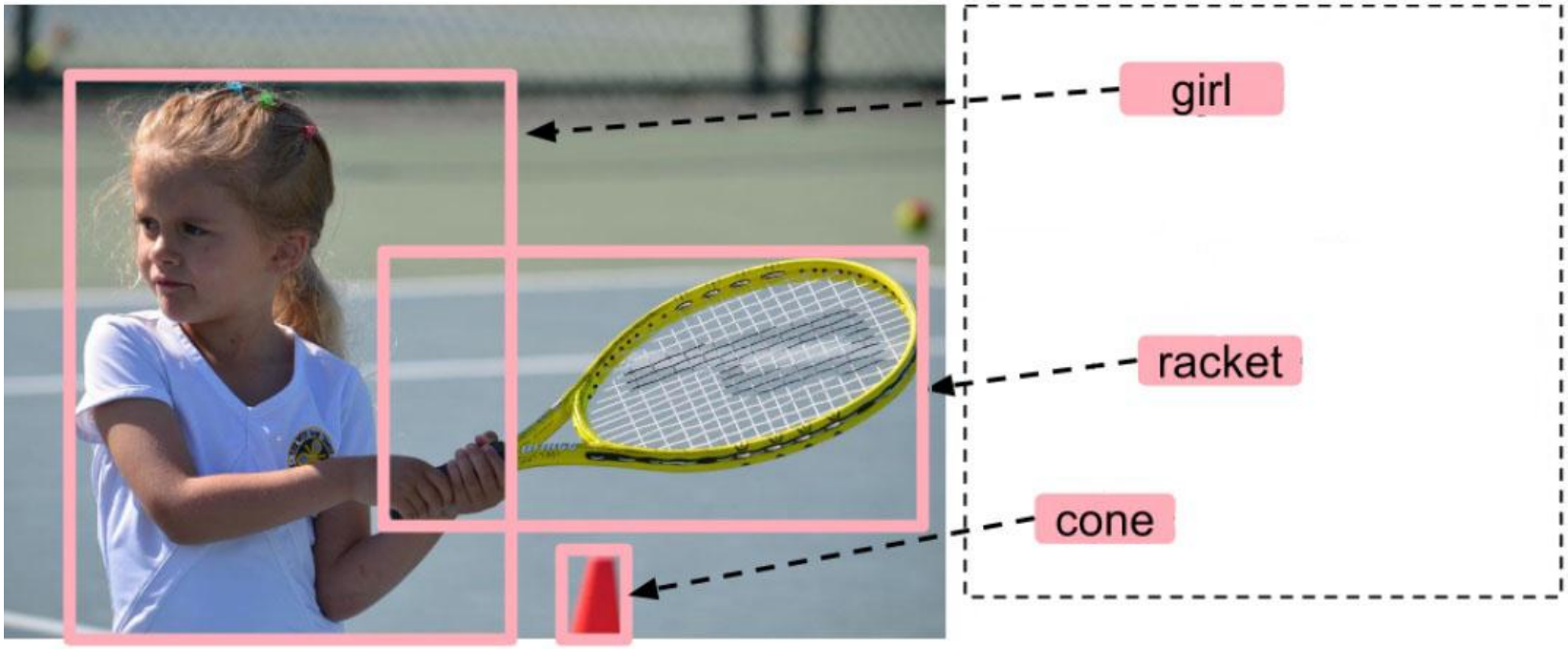




R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Scene Graphs

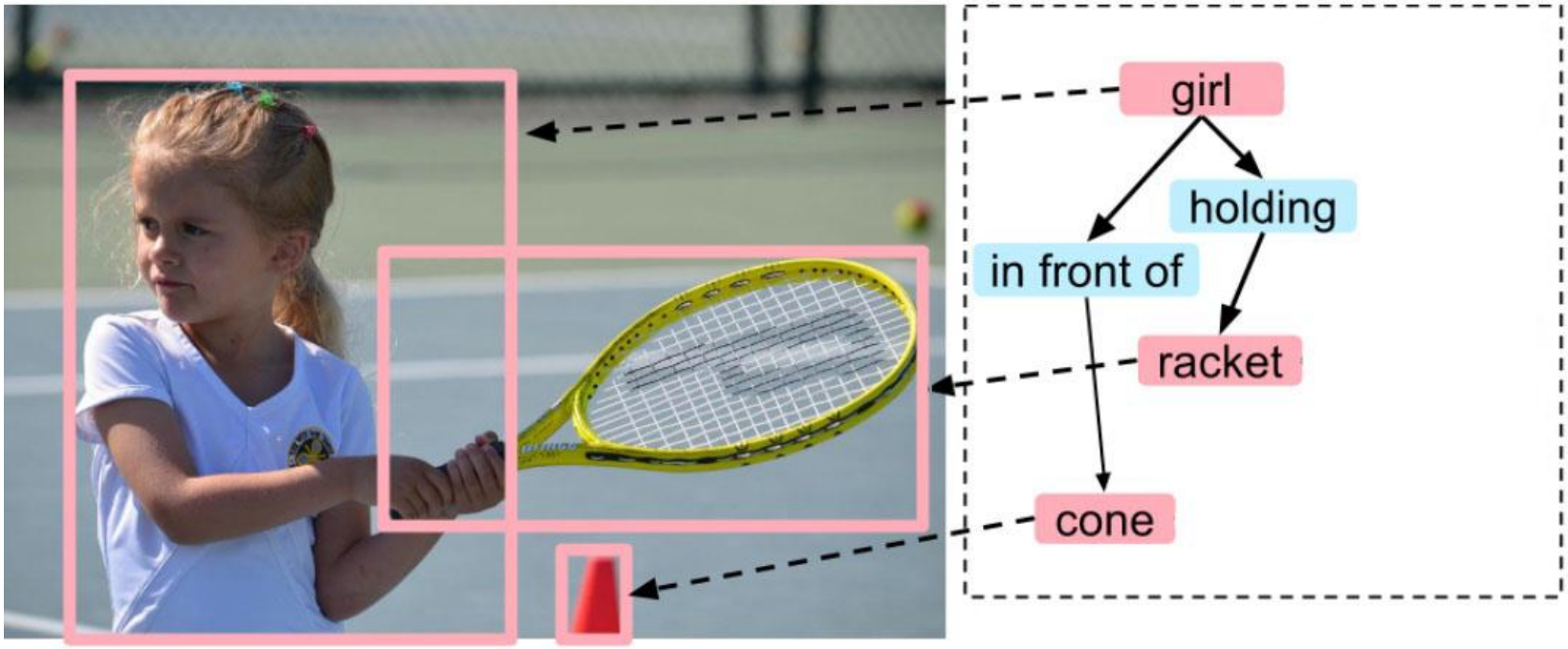
- Objects as Nodes (from Object Bounding Boxes)



R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Scene Graphs

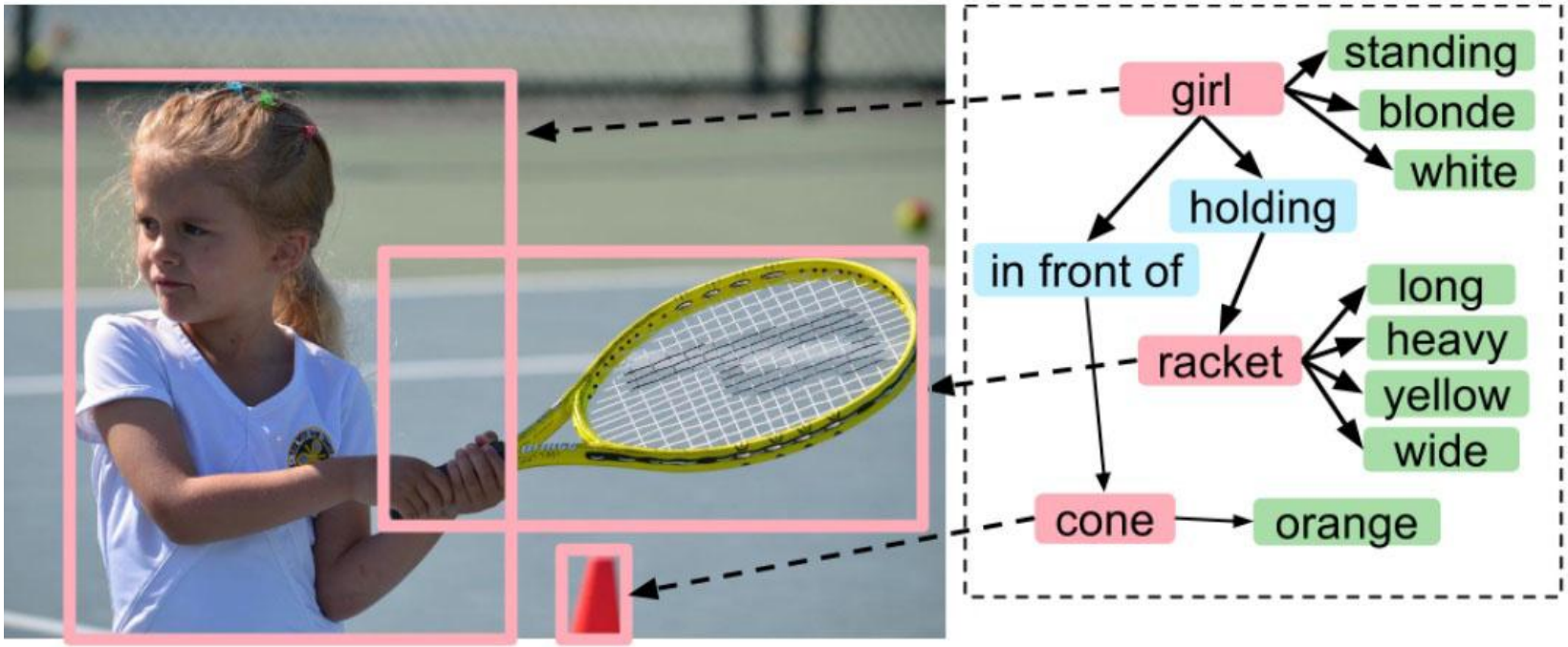
- Objects as Nodes (from Object Bounding Boxes)
- Relations as Edges



R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Scene Graphs

- Objects as Nodes (from Object Bounding Boxes)
- Relations as Edges
- Attributes as Features



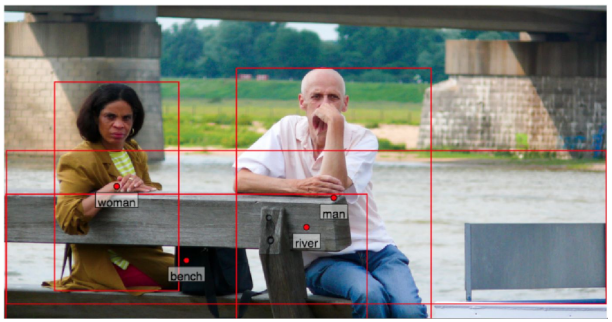
R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Visual Genome

- First really big semantic scene graph dataset
- 2D Images

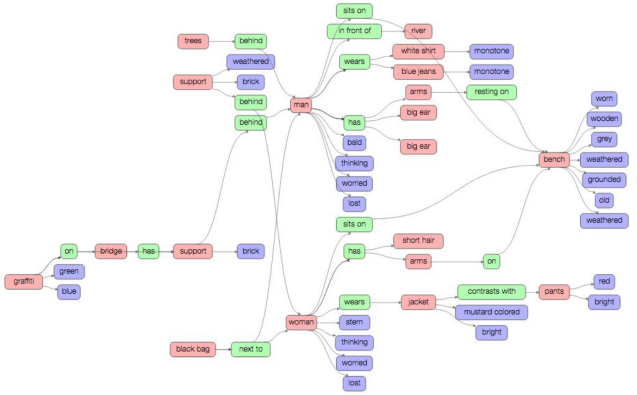
R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Visual Genome



R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Evaluating Scene Graphs



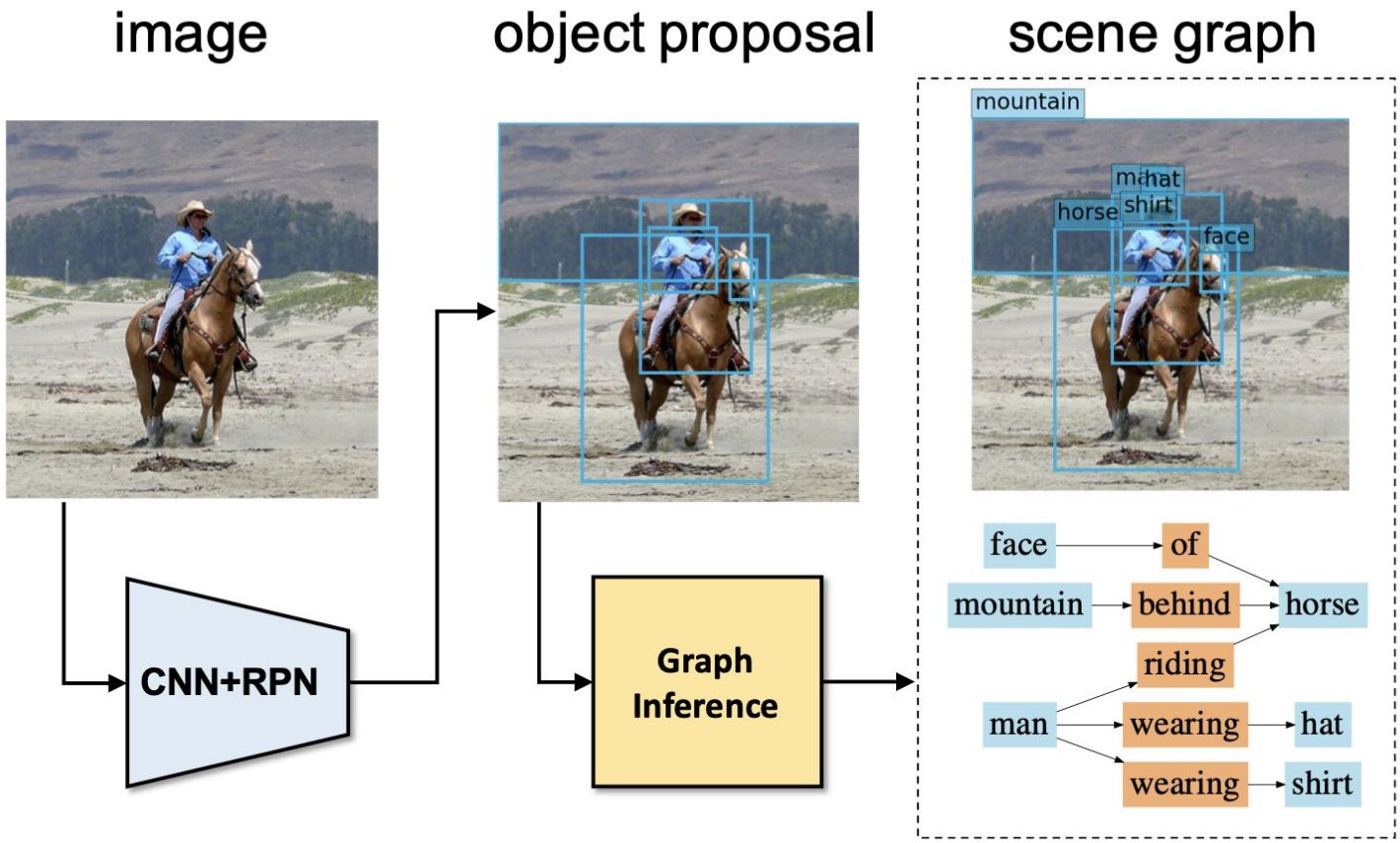
- Three setups:
  - Predicate Classification (PredCls):
    Predict predicates of relationships given localized objects
  - Scene Graph Classification (SGCls):
    Predict predicates and object categories given localized objects
  - Scene Graph Generation (SGGen):
    Detect and predict predicates and categories simultaneously
- Scene Graph labels are sparse → Recall as metric instead of mAP
  - R@k: Recall for top k predictions (e.g. R@50)

D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. “Scene Graph Generation by Iterative Message Passing“. IEEE 2017.

R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei. “Visual genome: Connecting language and vision using crowdsourced dense image annotations“. ICJV 2017.

Scene Graph Generation by Iterative Message Passing

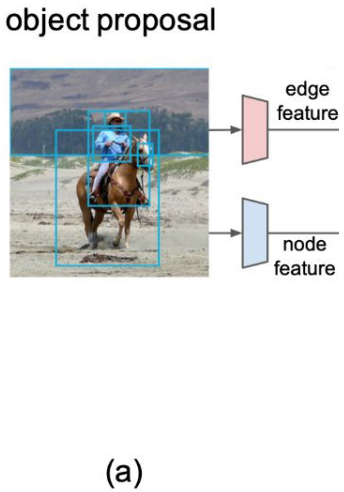
- Seminal work on learning based SSG prediction
- Object Detection → Pairwise relationship prediction
  - Quadratic cost



D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. “Scene Graph Generation by Iterative Message Passing“. IEEE 2017.

Scene Graph Generation by Iterative Message Passing

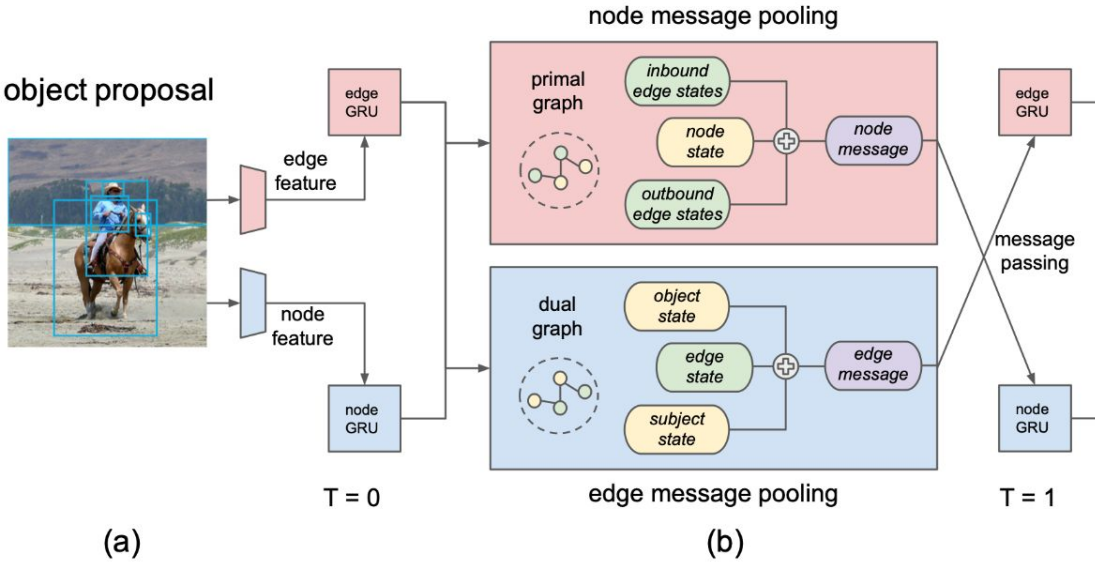
a. extracts visual features of nodes and edges



D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. “Scene Graph Generation by Iterative Message Passing“. IEEE 2017.

Scene Graph Generation by Iterative Message Passing

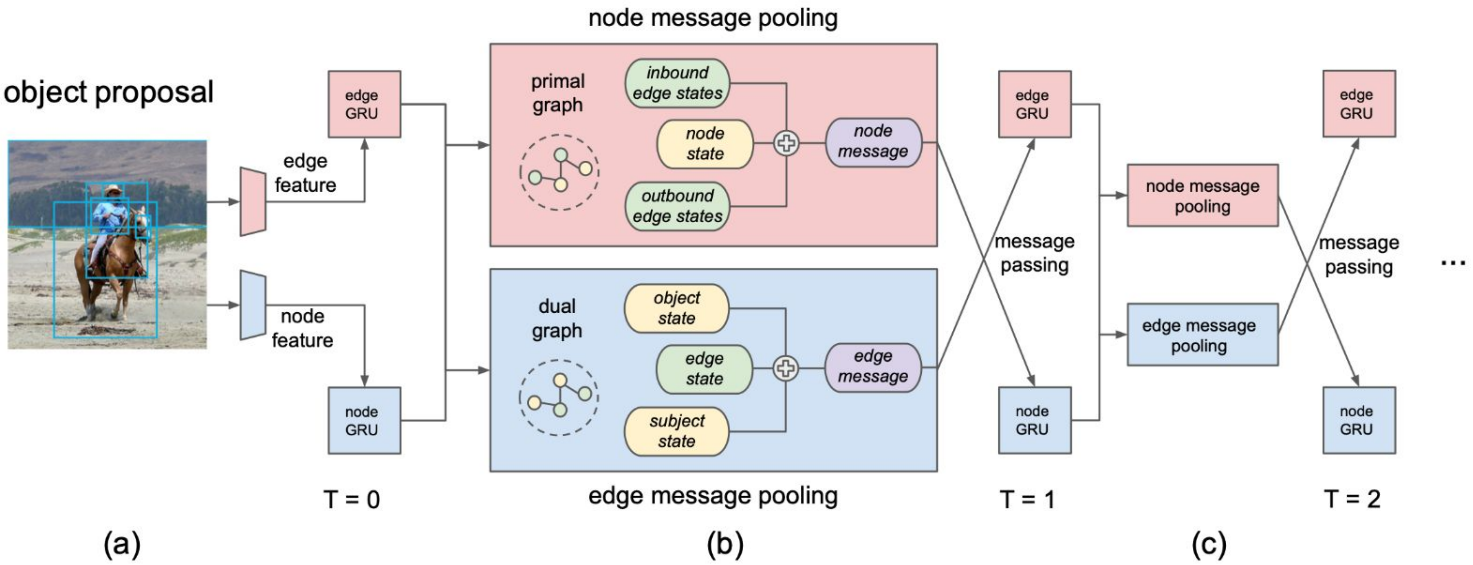
a. extracts visual features of nodes and edges
b. node and edge message pooling functions compute messages that are passed to the next GRU's



D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. "Scene Graph Generation by Iterative Message Passing". IEEE 2017.

Scene Graph Generation by Iterative Message Passing

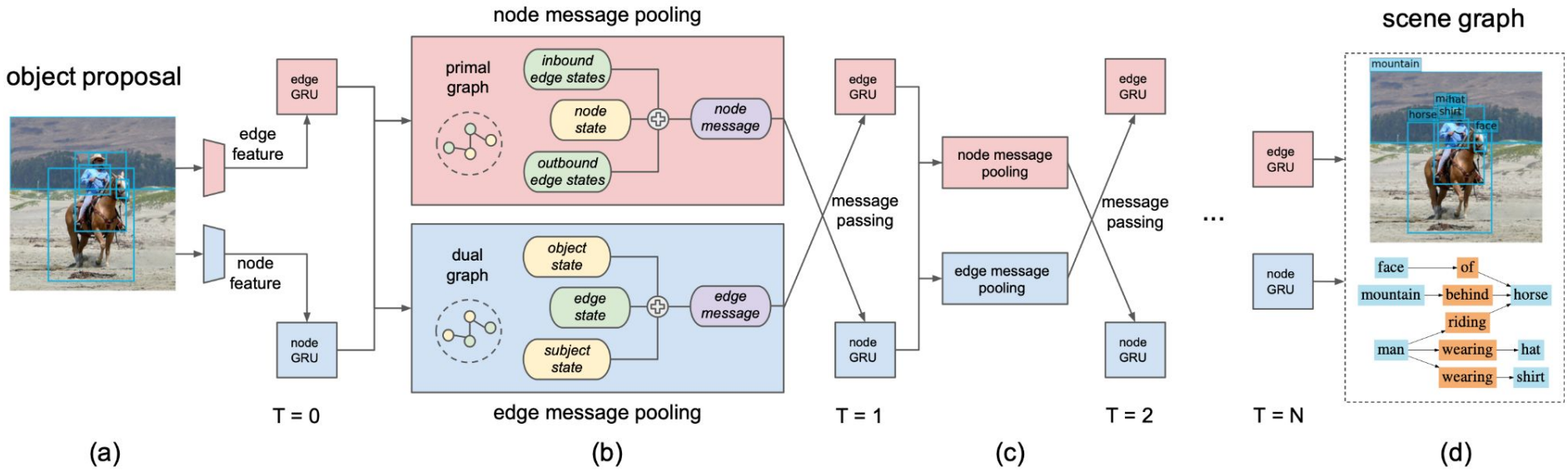
a. extracts visual features of nodes and edges
b. node and edge message pooling functions compute messages that are passed to the next GRU's
c. updates the hidden states of the GRUs



D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. "Scene Graph Generation by Iterative Message Passing". IEEE 2017.

Scene Graph Generation by Iterative Message Passing

a. extracts visual features of nodes and edges
b. node and edge message pooling functions compute messages that are passed to the next GRU's
c. updates the hidden states of the GRUs
d. the hidden states of the GRUs are used to predict object categories, bounding box offsets, and relationship types

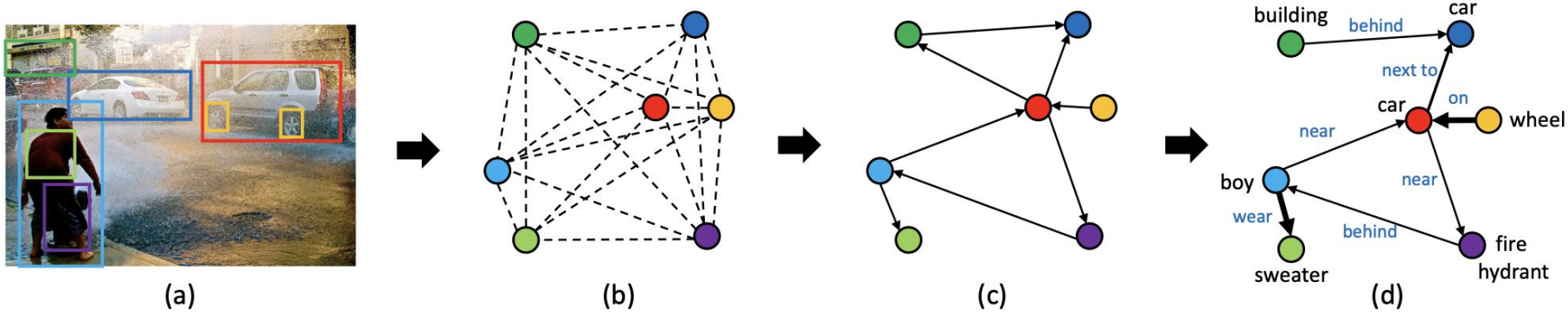


D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei. "Scene Graph Generation by Iterative Message Passing". IEEE 2017.

Graph R-CNN for Scene Graph Generation

Main contributions:

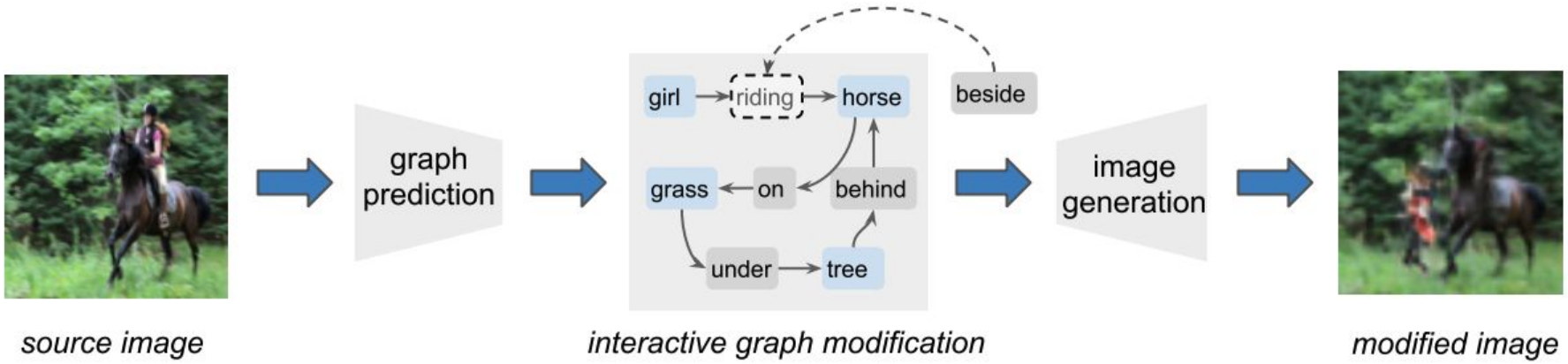
- Attentional GCN: Integrate context better
- Relation Proposal Network: Less Computation



J. Yang, J. Lu, S. Lee, D. Batra, D. Parikh. “Graph R-CNN for Scene Graph Generation“. ECCV 2018.

Semantic Image Manipulation Using Scene Graphs

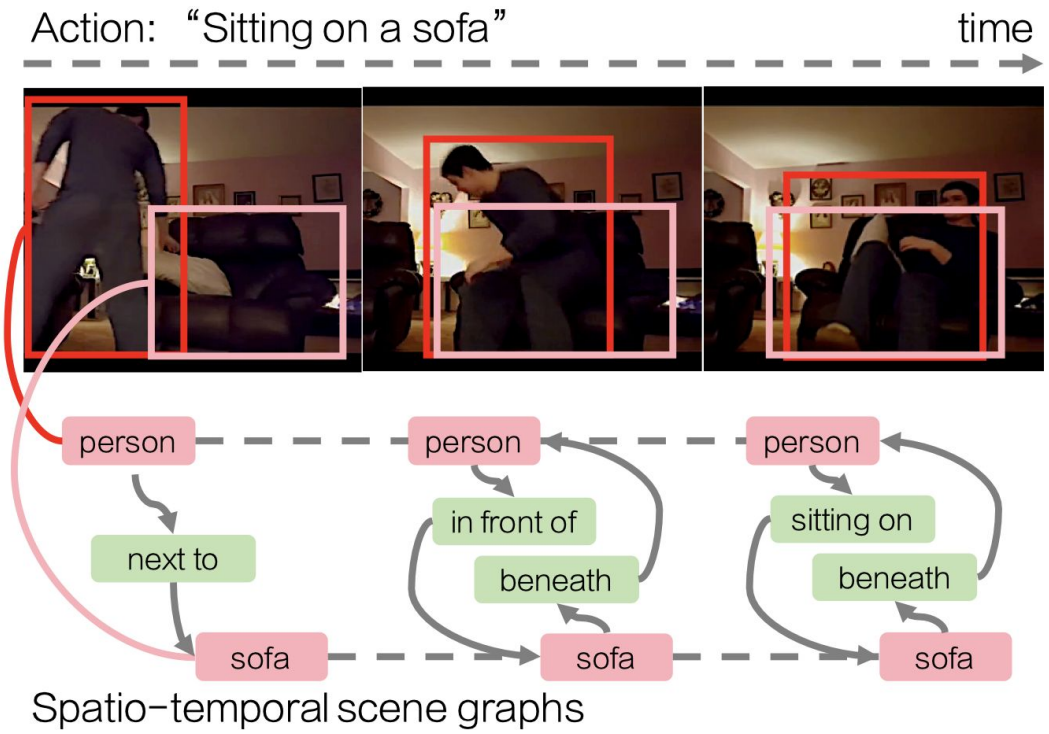
- Predict scene graph
- Allow user to modify
- Generate image from new scene graph



H. Dhamo, A. Farshad, I. Laina, N. Navab, G. D. Hager, F. Tombari, C. Rupprecht. Semantic Image Manipulation Using Scene Graphs. CVPR 2020

Action Genome: Actions as Composition of Spatio-temporal Scene Graphs

- Temporal Dataset
- Scene graphs evolve over time
- Baseline SSG prediction model using temporality (3D CNN)



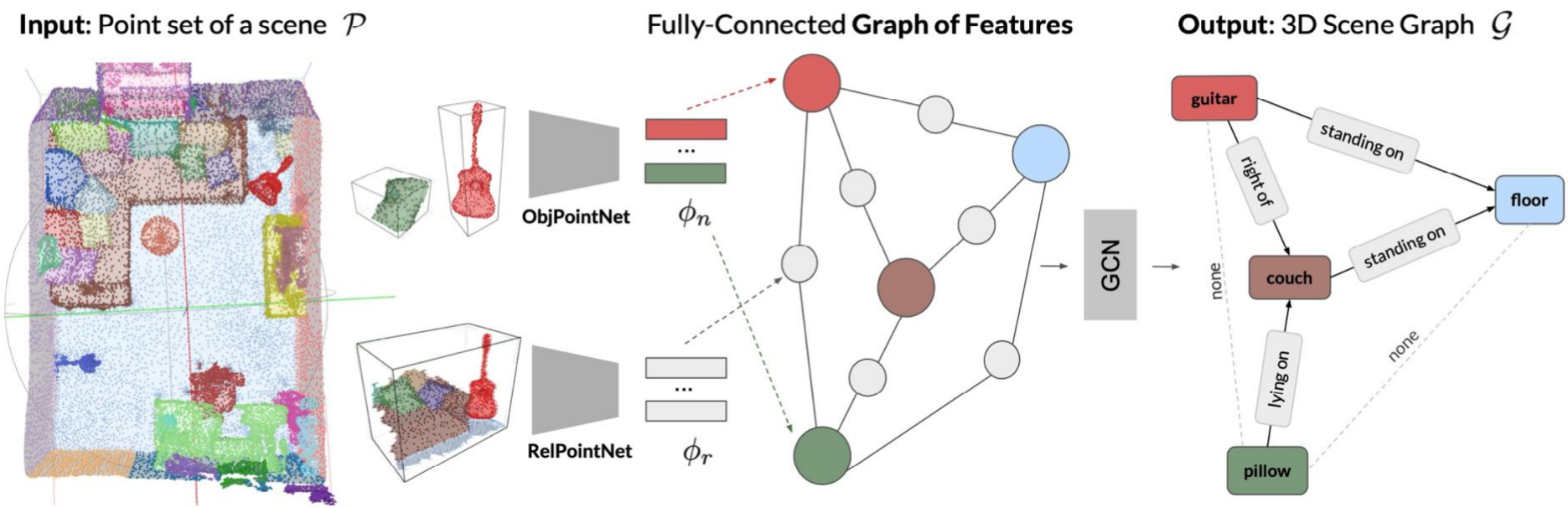
J. Ji, R. Krishna, L. Fei-Fei, J. C. Niebles. "Action Genome: Actions as Composition of Spatio-temporal Scene Graphs". CVPR 2020.

Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions

- ~1000 scenes & 3D dataset: has object and relationship labels
- Objects: static (no humans)
- Relationships: geometric (eg. left right up down) / semantic (eg. cleaner than)
- Scene graph prediction: 3DSSG. Input: Point Cloud, Instance Segmentation

J. Wald, H. Dhamo, N. Navab and F. Tombari, “Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions”, CVPR 2020.

3DSSG Architecture



J. Wald, H. Dhamo, N. Navab and F. Tombari, “Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions”, CVPR 2020.

Downstream Examples

- Image Retrieval
- Visual Question Answering
- Image / 3D Scene Synthesis
- Captioning

(a) Results for the query on a popular image search engine.

(b) Expected results for the query.

Figure 1: Image search using a complex query like "man holding fish and wearing hat on white boat" returns unsatisfactory results in (a). Ideal results (b) include correct *objects* ("man", "boat"), *attributes* ("boat is white") and *relationships* ("man on boat").

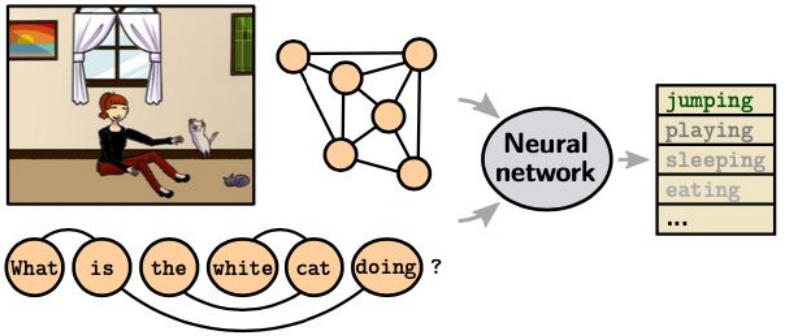
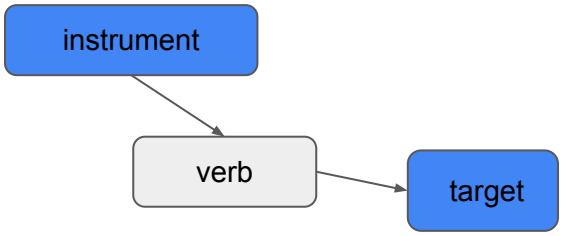


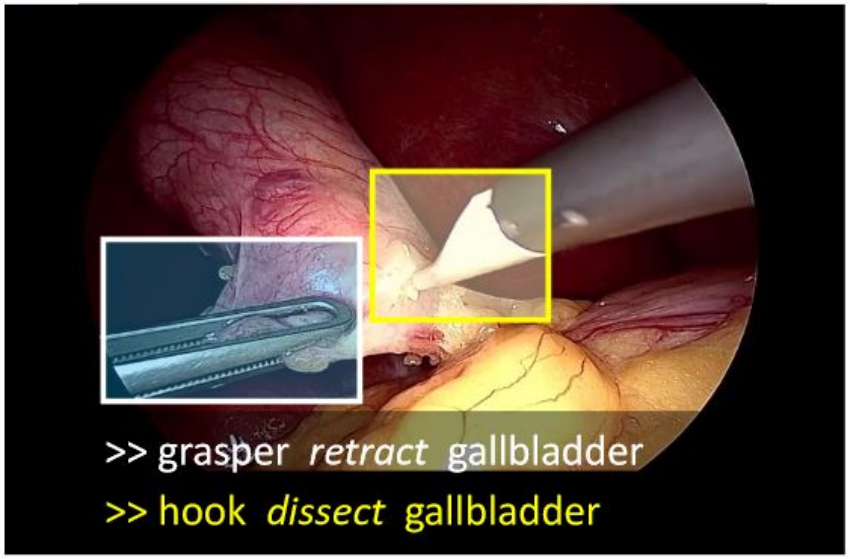
Figure 1. We encode the input scene as a graph representing the objects and their spatial arrangement, and the input question as a graph representing words and their syntactic dependencies. A neural network is trained to reason over these representations, and to produce a suitable answer as a prediction over an output vocabulary.

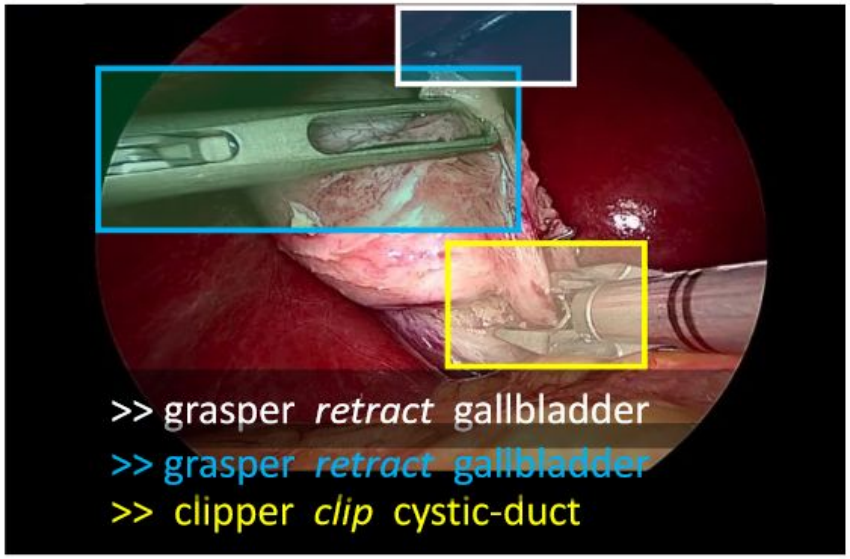
Johnson, Justin, Ranjay Krishna, Michael Stark, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. "Image Retrieval Using Scene Graphs." In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3668–78. Boston, MA, USA: IEEE, 2015. https://doi.org/10.1109/CVPR.2015.7298990.

CholecT50

- Laparoscopic cholecystectomy
- Dataset of 50 videos
- Detailed analysis of surgical procedure





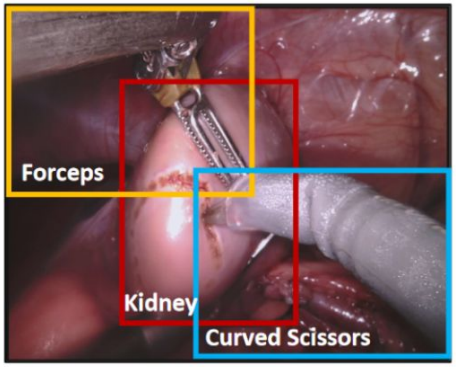


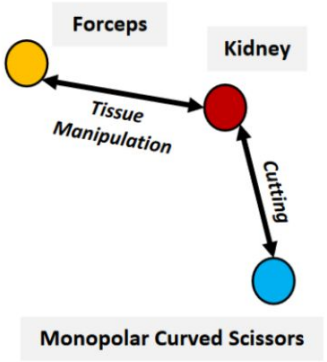
Nwoye, Chinedu Innocent, Cristians Gonzalez, Tong Yu, Pietro Mascagni, Didier Mutter, Jacques Marescaux, and Nicolas Padoy. “Recognition of Instrument-Tissue Interactions in Endoscopic Videos via Action Triplets.” In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*

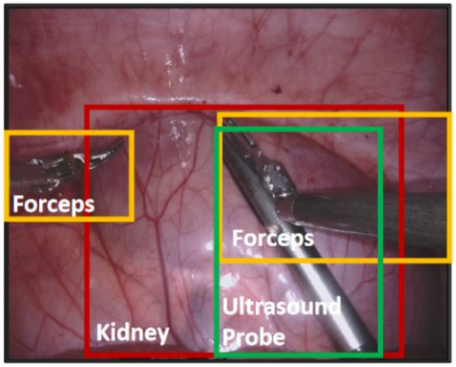
Nwoye, Chinedu Innocent, Tong Yu, Cristians Gonzalez, Barbara Seeliger, Pietro Mascagni, Didier Mutter, Jacques Marescaux, and Nicolas Padoy. “Rendezvous: Attention Mechanisms for the Recognition of Surgical Action Triplets in Endoscopic Videos.” *ArXiv:2109.03223 [Cs]*

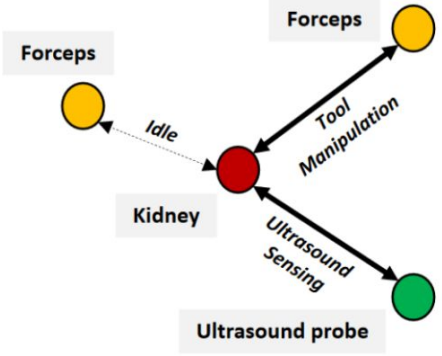
Endoscopic Scene Graph Dataset

- Robotic Nephrectomy
- Dataset of 15 videos





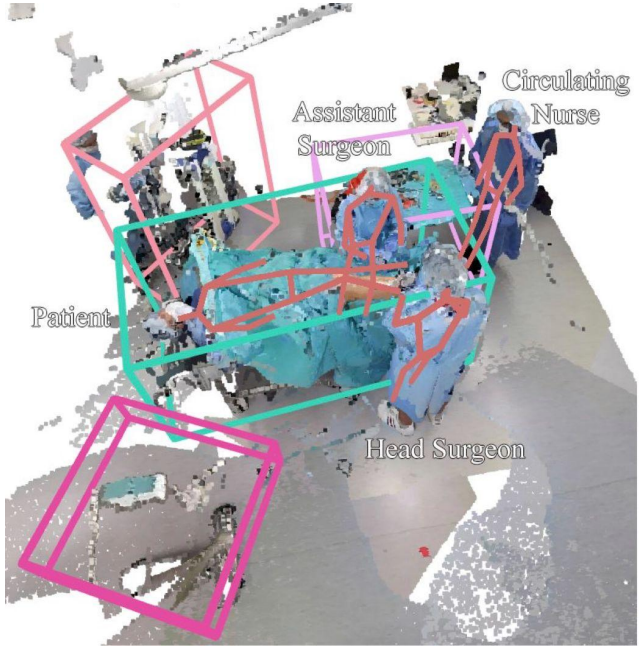


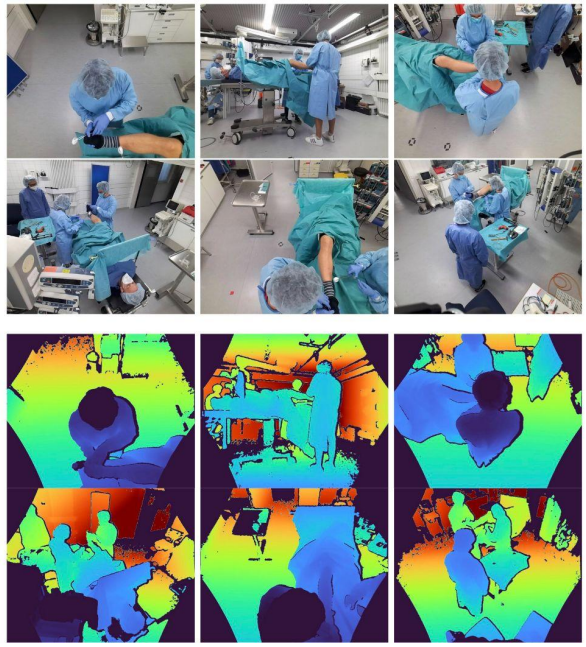


Islam, Mobarakol, Lalithkumar Seenivasan, Lim Chwee Ming, and Hongliang Ren. “Learning and Reasoning with the Graph Structure Representation in Robotic Surgery.” In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-59716-0_60.

4D-OR: Semantic Scene Graphs for OR Domain Modeling

- Multi-View RGBD Images -> 3D+time Point Cloud (1 FPS)
- 10 Simulated Total Knee Replacement Surgeries
- Human Tracks over Time
- 3D Human & Object Pose
- Semantic Scene Graph
- Clinical Roles



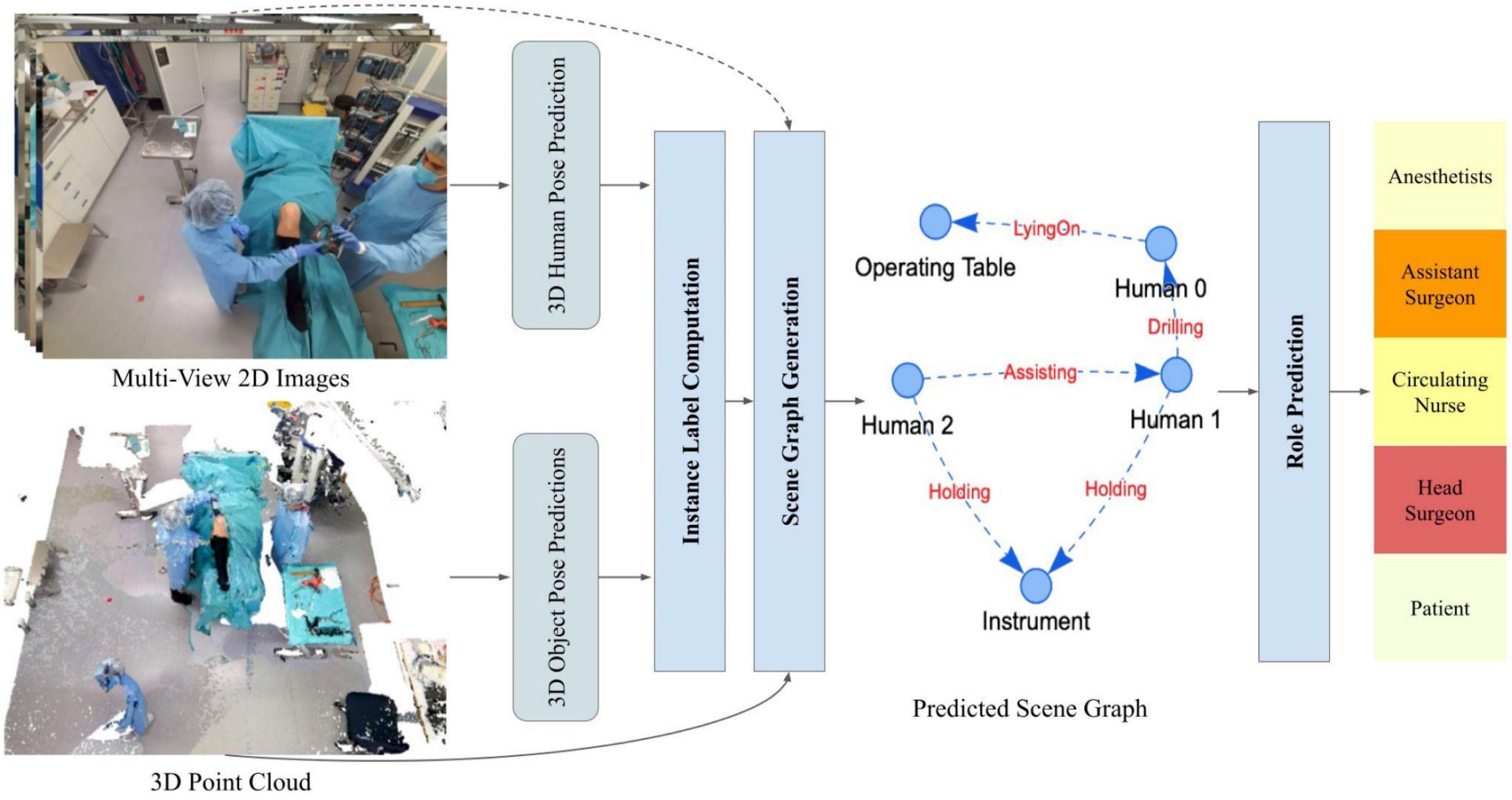


E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, N. Navab. 4D-OR: Semantic Scene Graphs for OR Domain Modeling. Accepted at MICCAI 2022

4D-OR: Semantic Scene Graphs for OR Domain Modeling

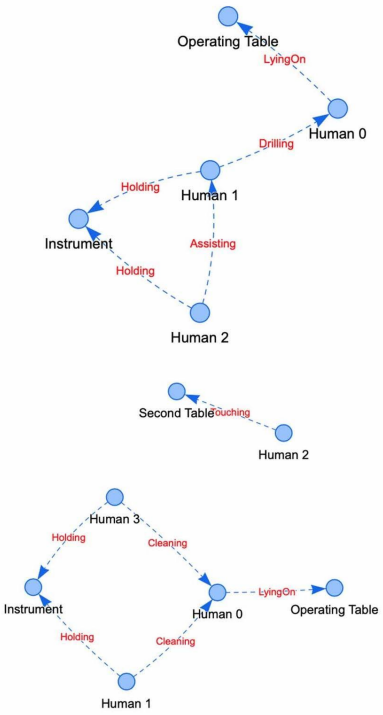
E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, N. Navab. 4D-OR: Semantic Scene Graphs for OR Domain Modeling. Accepted at MICCAI 2022

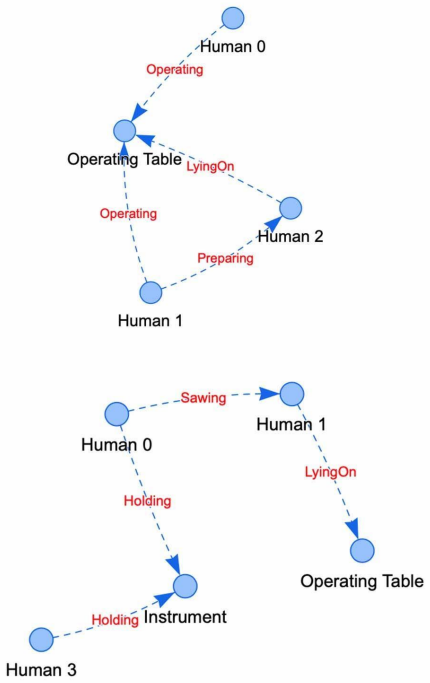
4D-OR: Semantic Scene Graphs for OR Domain Modeling



E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, N. Navab. 4D-OR: Semantic Scene Graphs for OR Domain Modeling. Accepted at MICCAI 2022

Qualitative Examples

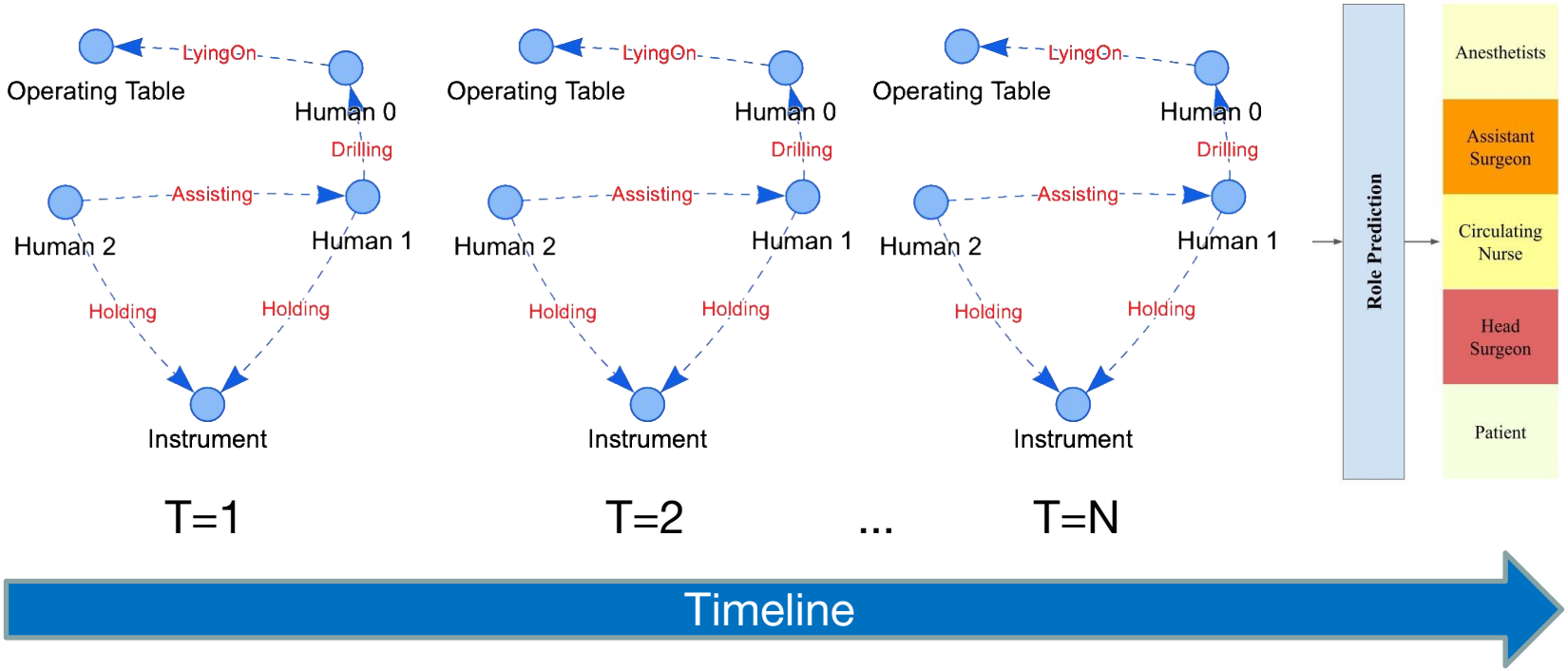




E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, N. Navab. 4D-OR: Semantic Scene Graphs for OR Domain Modeling. Accepted at MICCAI 2022

Downstream Task: Role Prediction

For each human track: Assign a probability for each role (head surgeon: 0.8, patient: 0.1) either heuristic-based or deep learning based



E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, N. Navab. 4D-OR: Semantic Scene Graphs for OR Domain Modeling. Accepted at MICCAI 2022

4D-OR: Semantic Scene Graphs for OR Domain Modeling

0:

E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, N. Navab. 4D-OR: Semantic Scene Graphs for OR Domain Modeling. Accepted at MICCAI 2022

Challenges and Future Work

- Significantly more labeling effort than classification or object detection
- SSG requires object detection
- Temporarily inconsistent results
- Limited to few domains, datasets, applications

Recap

- Scene Graphs are structured, semantically rich, and task-independent representations of visual scenes
- Scene Graphs consist of Objects, Relations and Attributes
- Scene Graph labels are sparse → Recall as metric instead of mAP
- Scene Graphs enable complex downstream tasks such as Image Retrieval, Visual Question Answering, Image Generation
- Scene Graphs are used with Images, Videos, Pointclouds and even 4D data
- Many challenges remain for future work

Questions? Looking for projects?

Ege Özsoy
PhD Candidate
ege.oezsoy@tum.de

Felix Holm
PhD Candidate
felix.holm@tum.de