Representation Learning for Semantic Image Manipulation Using Scene Graphs

General Info

Project Title: Representation Learning for Semantic Image Manipulation Using Scene Graphs

Contact Person: Azade Farshad, Yousef Yeganeh

Contact Email: azade.farshad@tum.de, y.yeganeh@tum.de

Project Abstract

The goal of this project is to improve the quality of generated images in the image manipulation framework (SIMSG[1]). There are two objectives: 1) Learning disentangled features to decompose pose and appearance features in the GAN using variational inference or state-of-the-art disentangled representation learning methods, 2) Measuring model uncertainty in image generation and minimizing the model uncertainty considering self-attention mechanism and bayesian approaches.

Background and Motivation

Image manipulation is an important task for different situations. Scene graphs are useful tools for easier manipulation of the scenes. Scene graphs have been recently used in modelling surgery rooms for action recognition. Learning a good feature representation is an important aspect of image manipulation and generation. In this project, we focus on two aspects of representation learning for image manipulation:
1) Disentangled representation 2) Attention / Uncertainty -based representation.

Technical Prerequisites

- Good background in statistics
- Good background in machine learning, deep learning
- Good skills in Python
- Good skills in PyTorch

Benefits:

- Weekly supervision and discussions
- Possible novelty of the research
- Possible publication

Students' Tasks Description

Students' tasks would be the following:

Groups 1 & 2:

- Understanding the underlying methods
- Evaluation on Visual Genome / Clinical dataset
- Testing and documentation.

Groups 1:

- Familiarize with disentanglement concepts
- Implement disentangled representation learning in SIMSG framework

Groups 2:

- Familiarize with uncertainty concepts
- Implement uncertainty measurement + minimization in SIMSG framework

- **Work-packages and Time-plan:**

	Description	#Students	From	To
WP1	Familiarizing with the literature.	4	22.10	29.10
WP2	Familiarizing with the required frameworks. Come up with a detailed time-plan (gantt)	4	29.10	06.11
WP3	Implementing disentanglement and adapting it to SIMSG	2	06.11	27.11
WP4	Implementing uncertainty/attention and adapting it to SIMSG	2	06.11	27.11
WP5	Evaluation of the implemented method	4	27.11	03.12
WP6	Comparison to related work + Preparing midterm presentation	4	03.12	10.12
M1	Intermediate Presentation II	4	12.2021	
WP7	Familiarizing with clinical data, data pre-processing	4	10.12	17.12
WP8	Implement and Evaluate WP3/WP4 & WP6 on medical data	4	17.12	15.01
WP9	Testing and Documentation	4	15.01	26.02
M2	Final Presentation	4	02.2021	

References

1. Dhamo, H., Farshad, A., Laina, I., Navab, N., Hager, G. D., Tombari, F., & Rupprecht, C. (2020). Semantic image manipulation using scene graphs. In CVPR.
2. Saatci, Y., & Wilson, A. (2017). Bayesian gans. In NeurIPS.
3. Özsoy, E., Örnek, E. P., Eck, U., Tombari, F., & Navab, N. (2021). Multimodal Semantic Scene Graphs for Holistic Modeling of Surgical Procedures. arXiv preprint arXiv:2106.15309.
4. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., ... & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion.
5. Patel, D. V., & Oberai, A. A. (2020). GAN-based Priors for Quantifying Uncertainty. arXiv preprint arXiv:2003.12597.

6. Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019, May). Self-attention generative adversarial networks. In ICML.
7. Pandey, A., Fanuel, M., Schreurs, J., & Suykens, J. A. (2020). Disentangled Representation Learning and Generation with Manifold Optimization. arXiv preprint arXiv:2006.07046.
8. Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., & Bachem, O. (2019, May). Challenging common assumptions in the unsupervised learning of disentangled representations. In ICML.
9. Zhu, X., Xu, C., & Tao, D. (2020, August). Learning disentangled representations with latent variation predictability. In ECCV.