PhD subject:

Deep 3D garment semantics

Hosting institute

ICube Laboratory (The Engineering science, computer science and imaging laboratory) at the University of Strasbourg is a leading research center in Computer Science, with more than 300 permanent researchers, with the recently opened AI graduate school supported by the French government.

Work place and salary

The thesis work will take place in the MLMS (Machine Learning, Modeling & Simulation) research team of the ICube laboratory (The Engineering science, computer science and imaging laboratory) of the University of Strasbourg, a leading research center with more than 300 permanent researchers. The workplace is located on the hospital site of the laboratory, a 10-minute walk from the heart of downtown Strasbourg, listed as a UNESCO World Heritage Site.

2 135,00 € gross monthly

Supervisors

– director: Hyewon Seo (ICube, Univ. Strasbourg)

– co-supervisors: Cédric Bobenrieth (ICAM, Strasbourg)

Staring date

October 2024 – January 2025.

Context

In various scenarios, such as providing daily assistance to the elderly, aiding individuals recovering from severe injuries, or supporting commercial and recycling tasks, one can envision a robotic assistant handling garments by fetching, folding, and organizing clothes. In this project, we will focus on a key technical prerequisite for realizing this challenging scenario: advancing current computer vision techniques to enhance understanding and reasoning about the garment being manipulated. Specifically, we will address the challenging tasks of recognizing a garment and its semantic parts in the visual content. The deep semantic understanding of garments is particularly challenging as garment objects depict complex topological and physical behaviors, involving strong self-occlusions and deformations of fabrics.

Recent successful image segmentation techniques in computer vision, such as SAM (Segment Anything Model) [KMR*23] do not easily generalize to the detection of semantic parts, since the boundary assumption in object detection does not apply in semantic part detection where the presence of explicit boundaries is only optional. While its adaptation through prompting to specific domains like medical imagery have been demonstrated, SAM has shown exceptional proficiency only in certain objects and modalities, exhibiting inadequacy or ineffectiveness in other contexts [MHL*24]. Advancing beyond FCN (Fully Convolutional Network) based per-pixel labeling [LSD*15], recent developments have introduced specialized context modules [CPK*18, ZSQ*17] and various self-attention mechanisms [SGL*21, ZLZ*21], enabling the incorporation of contextual information. However, these models may not be particularly beneficial in our targeted settings, where the context is often limited. Methods specially developed for garment semantics mostly deal with 2D static images depicting clothes close to their canonical states seen by clear views, and focus on the segmentation of individual clothes rather than their semantic parts [II22, XHD*23].

Work description

We will ground our 2D segmentation models on paired 3D objects, considering the highly flexible, self-occluding, and nonlinear dynamics nature of the garment object. Our 3D segmentation models will leverage temporal information during the transition to/from the garment's canonical form, to effectively handle the highly deformed state. We will proceed with the following tasks:

1. **Detection of garment types and semantic parts in canonical garments with partial occlusions, from 3D image input:** A first step is to investigate the problem in relatively well-defined, simple settings: detection of garment categories and semantic part-segmentation in their canonical shapes. Although we will initially focus on static garments, utilizing static images, we also intend to explore dynamic garments in specific scenarios, using videos.
2. **Detection of garment types and semantic parts in non-canonical state with significant deformations and occlusions**: The main challenges occur when clothes are in their non-canonical states, due to a high level of deformation, which induces drastic shape changes and significant self-occlusions. To make the problem feasible, we will assume that a single garment is depicted in the image input. Two different approaches are considered: One, we start with identification of a garment in its canonical shape and develop methodology to transfer its semantics to the non-canonical form, by estimating nonrigid alignment between them. Two, we develop a bending-invariant method for the detection tasks, so that consistent results are obtained regardless of the garment state.

Constraints

The thesis is within the framework of a collaborative project with a predefined time schedule, so the delivery of the work should conform to this global schedule. The garment mesh and video dataset will be provided by a project partner. We anticipate one or two domestic trips per year for project workshops or meetings to collaborate and discuss with project partners.

Supplementary information

It would be ideal to start with the master 2 internship linked to this thesis.

Candidate profile

− Master degree in Computer Science, Electronic & Electrical Engineering, or in Applied Mathematics

− Solid programming skills: Python/C++

− Background in Geometric Modeling

− Experience in Deep Learning

− Good communication skills

Application

Send your CV and academic records (Bachelor and Master) to seo@unistra.fr

Bibliography

[CPK*18] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. PAMI, 2018.

[II22] S. Ishikawa, T. Ikenaga, Image-based virtual try-on system with clothing extraction module that adapts to any posture, Computers & Graphics, Volume 106, 2022.

[KMR*23] Kirillov, A. & Mintun, E. & Ravi, N. & Mao, H. & Rolland, C. & Gustafson, L. & Xiao, T. & Whitehead, S. & Berg, A. & Lo, W.-Y. & Dollár, P. & Girshick, R., (2023). Segment Anything, https://arxiv.org/pdf/2304.02643.pdf.

[LSD*15] Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

[MHL*24] Ma, J., He, Y., Li, F. et al. Segment anything in medical images. Nature Communications 15, 654 (2024).

[SGL*21] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In ICCV, 2021.

[XHD*23] Z. Xie, Z. Huang, X. Dong, F. Zhao, H. Dong, X. Zhang, F. Zhu, X. Liang. GP-VTON: Towards General Purpose Virtual Try-On via Collaborative Local-Flow Global-Parsing Learning. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023.

[ZLZ*21] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In CVPR, 2021.

[ZSQ*17] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, 2017.