25.04.2025

Thesis Proposal

Title	Context-Aware Emotion Recognition from Pictures using Frozen CLIP
Student	Vinzenz Dewor (22995021)
Course of study	Data Science B.Sc.
Supervisor	Mathias Zinnen
Affiliation	Friedrich-Alexander University of Erlangen-Nuremberg Department of Computer Science Pattern Recognition Lab Computer Vision Research Group

This thesis is inspired by previous work[1] at the Pattern Recognition Lab using the EMOTIC dataset, originally aimed at transfer to artworks. The focus of this bachelor thesis is to develop and test a new approach for estimating the emotions of people in images, while taking the entire scene into account. This approach similarly uses machine learning, but relies much more on general-purpose pretraining, specifically the popular CLIP model[2].

The goal is to achieve competitive results, mainly evaluated on EMOTIC with mean average precision on discrete emotion categories, but also beyond.

Mandatory requirements:

- different image crops for independent feature extraction
- specific face branch and comparison with facial emotion recognition
- zero-shot classification evaluation, including some prompt optimization
- fusion of different crop features and comparison of fusion strategies
- training and evaluation on EMOTIC, CAER-S, HECO, FindingEmo

Optional requirements:

- evaluation on ODOR-emotions
- test context debiasing methodology[3]
- valence-arousal estimation
- improve generalization assessed by cross-dataset performance

[1] V. Patoliya, M. Zinnen, A. Maier, and V. Christlein, 'Smell and Emotion: Recognising emotions in smell-related artworks', arXiv [cs.CV], 05-Jul-2024.
[2] A. Radford et al., 'Learning transferable visual models from natural language supervision', arXiv [cs.CV], 26-Feb-2021.
[3] D. Yang, K. Yang, M. Li, S. Wang, S. Wang, and L. Zhang, 'Robust Emotion Recognition in Context Debiasing', arXiv [cs.CV], 09-Mar-2024.