

Chaoxi Xu

+86-166-1971-1367 | xcx@ruc.edu.cn
xuchaoxi.github.io

EDUCATION

Renmin University of China

2017.09 - 2020.07

Master, Computer Application Technology, School of Information

Beijing

- GPA: 3.74 / 4.0
- Awards: Second Prize scholarship (2017), First Prize scholarship (2018)
- Related courses: artificial intelligence algorithms and applications, statistical machine learning, modern statistical methods

Renmin University of China

2013.09 - 2017.06

Bachelor, Computer Science and Technology, School of Information

Beijing

- Related courses: data structure and algorithms, operating system

RESEARCH EXPERIENCE

Publications

- **Chaoxi Xu**, Xiangjia Zhu, Wenwen He, Yi Lu, XiXi He, Zongjiang Shang, Jun Wu, Keke Zhang, Yinglei Zhang, Xianfang Rong, Zhennan Zhao, Lei Cai, Dayong Ding, Xirong Li (2019): Fully Deep Learning for Slit-lamp Photo based Nuclear Cataract Grading. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2019 (early accept)
- Xirong Li, **Chaoxi Xu**, Gang Yang, Zhineng Chen, Jianfeng Dong (2019): W2VV++: Fully Deep Learning for Ad-hoc Video Search. In: ACM Multimedia, 2019 (CCF-A)
- Jianfeng Dong, Xirong Li, **Chaoxi Xu**, Shouling Ji, Yuan He, Gang Yang, Xun Wang (2019): Dual Encoding for Zero-Example Video Retrieval. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019 (CCF-A)
- Xirong Li, **Chaoxi Xu**, Xiaoxu Wang, Weiyu Lan, Zhengxiong Jia, Gang Yang, Jieping Xu (2019): COCO-CN for Cross-Lingual Image Tagging, Captioning and Retrieval. IEEE Transactions on Multimedia (TMM), 2019 (IF=5.452)
- Jianfeng Dong, Xirong Li, **Chaoxi Xu**, Gang Yang, Xun Wang (2018): Feature Re-Learning with Data Augmentation for Content-based Video Recommendation, ACM Multimedia, 2018 (Grand challenge paper) (CCF-A)

Deep learning based cross-media representation learning

2017.08 - Present

Master, AI & Media Computing Lab, Renmin University of China

Beijing

- To tackle cross-lingual cross-modal retrieval task, we use the enhanced W2VV model with the Attention mechanism to enhance the text representation, and with contrastive loss function as the optimization target. In combination with MSCOCO and COCO-CN data, we use this scheme to simultaneously project Chinese and English sentences into the same visual feature space, and thus boost the performance of image retrieval by cross-lingual information. COCO-CN dataset and source code are released at: <https://github.com/li-xirong/coco-cn>
- In a zero-example video retrieval scenario, the user expresses his query in the form of natural language sentences. We use a new concept-free model W2VV++, which uses the improved margin ranking loss as the optimization target for multimodal matching in the common space and applies hard negative mining. Simultaneously, the mean pooling method is used to utilize the information of all hidden layers of the GRU, and the visual features and the text encoding are jointly projected to the subspace to measure the similarity. This method won the first place in the TRECVID 2018 Ad-hoc Video Search (AVS) task and was accepted by ACM Multimedia 2019.

- The current video relevance prediction in most online services is based on user behavior, which will inevitably leads to a "cold start" problem. Based on this background, Hulu (Beijing) and ACM Multimedia 2018 jointly launched a content-based video relevance prediction contest, hoping to effectively utilize the image, sound, text and other source contents contained in the videos for video relevance prediction. In view of the insufficiency of existing deep learning features, we propose a new Feature Re-Learning with Data Augmentation scheme. The official evaluation shows that our recommended accuracy of the program is significantly higher than the baseline system provided by the organizers. This method won the first place in Movies relevance prediction track.
- In order to improve the cross-modal representation and similarity calculation between sentences and videos, we propose a new dual encoding network, which uses mean pooling to encode global information, GRU to encode timing information and CNN to encode local information, and it is the first time to use a multi-layered encoding network with a similar architecture to simultaneously represent sentences and videos, surpassing existing results on multiple challenging evaluation benchmarks. This work has been accepted by CVPR2019.

International Challenges

- Top performer of the TRECVID 2018 Video-to-Text (VTT) Matching and Ranking task
- Top performer of the TRECVID 2018 Ad-hoc Video Search (AVS) task
- Winner of the ACM Multimedia 2018 Hulu Content-based Video Relevance Prediction Challenge

INTERNSHIP EXPERIENCE

Visionary Intelligence Ltd. (Beijing)

2018.11 - Present

Algorithm Engineer, AI Lab

Beijing

- Nuclear Cataract Automatic Grading, core development member

Cataract is clouding in the lens of eye and is the leading cause of blindness. According to the LOCS III, the level of the opacity of the nuclear is from 1.0 to 6.0 with one decimal place. The manual grading process not only requires well-trained eyes but also time-consuming. While automating the grading process is much in demand, challenges exist. Firstly, it is difficult to reliably localize the nuclear region in a relatively complex background. Another challenge is how to derive a vectorized representation of the ROI because of both intra-grade divergence and inter-grade similarity in the photometric appearance of the nuclear area. To tackle these challenges, we use fully deep learning solution including Faster R-CNN to detect and localize the nuclear region and ResNet-101 to represent the ROI and for final grading. To deal with the problem of imbalanced data, we use the batch balance strategy to ensure that the input of mini-batch is balanced in each level interval. This work improved nuclear cataract automatic grading process in a great margin and was accepted by MICCAI2019.

LEADERSHIP EXPERIENCE

Outdoor Association of Renmin University of China

2015.09 - 2016.09

Vice President, Climbing Captain

Beijing

- As the climbing captain, I organized our school climbing team to reach the main peak of Animaqing(6282m).

SKILLS

- C++/Python programming skills, proficient in Pytorch framework, familiar with mxnet
- Strong English reading, writing and presentation skills

CERTIFICATIONS & OTHERS

- **Languages:** English(CET-6)