

Bohua Peng

bp1119@ic.ac.uk marvinbp119@gmail.com +86 18800295676 <https://pengbohua.github.io/Pengbohua/>

EDUCATION

Imperial College London	Master	Sep.2020 – Nov.2021
Major: Artificial Intelligence and Machine Learning	Average Score: Distinction	72.6/100.0
Shanghai Maritime University	Bachelor	Sep.2011 – Jun.2015
Major: Computer Engineering	Average Score: 90.5/100.0	

SELECTED PUBLICATIONS

Bohua Peng, Islam Mobarakol et al. Bi-Link: Bridging Knowledge Graph Inductive Link Predictions from Text via Contrastive Learning of Transformers and Prompts (Under Review)

Bohua Peng, Islam Mobarakol, Mei Tu. Angular Gap: Reducing the Uncertainty of Image Difficulty through Model Calibration. (ACM Multimedia 2022)

Yujin Wang, Jing Zhang, **Bohua Peng** et al. SRCB at SemEval-2022 Task 5: Pretraining Based Image to Text Late Sequential Fusion System for Multimodal Misogynous Meme Identification. (ACL Semeval 2022)

SELECTED PROJECTS

- **WebCPM Browser-assisted question-answering for commonsense** 2022,06-Now
 - WebCPM can retrieve data from Bing like a human and answer questions like **WebGPT**, but with more actions.
 - Interactive QA involves, searching strategy, document retrieval, summarization, answer ranking.
 - I study information retrieval and answer ranking. Here is a [demo](#) for a **multi-hop QA data collection** system.
- **Misogyny detection (SemEval Task5 Top1 over 500+ teams)** 2021,12-2022,02
 - Background: Weak semantic consistency and domain shift exist in this image-text classification.
 - To handle statistical bias between train and test distribution, we aggregated **CLIP** and **XGBoost** as final submission.
 - I tried finetuning **CLIP** and Zero-Shot with Prompt Tuning. I also tried **UNITER** and **ELECTRA**.
- **Conditional visual question generation system** 2020,6 – 2020,12
 - To generate questions based on CEFR level, I used two VAEs to fuse questions, answers and CEFRs. I investigated posterior collapse with KL term, FreeBits and BN-VAE.
- **Evaluating humor in edited news (Semeval2020 Task7)** 2019,12- 2020,02
 - Background: Comparing the funniness of the edited versions of the original headline.
 - Data augmentation: I applied **NER** with **BERT** and **CRF**. I finetune a **GPT-2** to generate context.
 - I finetuned **RoBERTa** with MarginLoss and BCE against multi-head GRU, **XGBoost** and SVM.
- **Image segmentation uncertainty quantification (MICCAI2019 QUBIQ)** 2018,12- 2019,02
 - Background: Estimate the uncertainty from different delineations of human radiologists.
 - I implemented **cascade FCNN** (Chongyi Li 2018 et al.) from scratch.
 - I wrote **cascade random cropping** function to generate image crops for this specific model.
 - This work is further rendered using **anisotropic angle surface reconstruction** with C++.

WORK EXPERIENCE

Samsung Research	Resident; Software Engineer	Oct.2017 – August. 2019; Nov.2021 – Now
------------------	-----------------------------	---

PROGRAMMING&LANGUAGE

IELTS 7.5, GRE 316.

Proficient: C++, Python, Huggingface Transformers, SentenceBERT, Detectron, YOLO, timm, EasyOCR, PaddleOCR, PyTorch, Bash, Latex, Docker, Django, labelstudio, labelme.

Competitive: Virtuoso, ElasticSearch, Tensorflow, PaddlePaddle, CSS, html5, MySQL.

AWARDS

National Scholarship 2015 (Top 2%)

Shanghai Sunshine Dance Competition [Rehearsal Video]