

BO DAI

Site: daibo.info

Email: bdai@ie.cuhk.edu.hk / doubledaibo@gmail.com

Address: Rm 702, HSB, The Chinese University of Hong Kong, Hong Kong

EDUCATION

Ph.D. Aug 2014 - Aug 2018

- **The Chinese University of Hong Kong (CUHK)**
- Major: Information Engineering
- Advisor: Prof. Dahua Lin

B.Eng. Sept 2010 - Jun 2014

- **Shanghai Jiao Tong University (SJTU)**
- ACM Honor Class
- Major: Computer Science and Technology

WORK EXPERIENCE

The Chinese University of Hong Kong (CUHK)
PostDoc Oct 2018 - Present

- Helping supervisor lead a group of students to conduct research in computer vision and machine learning, including generative models, rl, cross-modality, and video analysis.

University of Toronto (UofT)
Visiting Student Sep 2017 - Dec 2017

- Working with Prof.Sanja Fidler
- Focusing on improving evaluation metrics and model structures for image captioning

Microsoft Research Asia (MSRA)
Research Intern Jul 2013 - Jan 2014

- Working with Dr.Liang Jeff Chen
- Focusing on developing a mechanism for storing semi-structured data in structured databases

Center for Brain-like Computing and Machine Intelligence (BCMI)
Research Assistant Jun 2012 - Jun 2013

- Working with Prof.Liqing Zhang
- Focusing on developing trademark retrieval systems

PUBLICATIONS

FineGym: A Hierarchical Video Dataset for Fine-grained Action Understanding

Dian Shao, Yue Zhao, **Bo Dai**, Dahua Lin.

(**Three Strong Accepts**) IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020.

- A new dataset for fine-grained gymnastics action recognition.
- High-quality visual data collected from professional competitions.

- High-quality annotations organized hierarchically, acrossing several semantic/temporal granularities.
- Revealing the gaps between coarse- and fine-grained action recognition.

Temporal Pyramid Network for Action Recognition

Ceyuan Yang, Yinghao Xu, Jianping Shi, **Bo Dai**, Bolei Zhou.

IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020.

- A generic module for action recognition.
- A feature-level temporal pyramid that fuses temporal dynamics captured at different tempos.
- Robustly handling actions with varying visual tempos, which means the speed of an action.

Scene De-occlusion: Ordering Recovery, Amodal and Content Completion without Corresponding Annotations

Xiaohang Zhan, Xingang Pan, **Bo Dai**, Ziwei Liu, Dahua Lin, Chen Change Loy.

IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020.

- An unsupervised framework that de-occludes scenes.
- Recovering orders of occluded objects by deciding their pair-wise relationships.
- Providing amodal completion and succeeding manipulations based on recovered orders.

Intra- and Inter-Action Understanding via Temporal Action Parsing

Dian Shao, Yue Zhao, **Bo Dai**, Dahua Lin.

IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020.

- Proposing a new task, temporal action parsing, for action understanding.
- Revealing the internal structures of actions by parsing them into sub-actions without labels.
- Revealing the connections between different actions via common sub-actions.

Real or Not Real, that is the Question

Yuanbo Xiangli*, Yubin Deng*, **Bo Dai***, Chen Change Loy, Dahua Lin.

(**Spotlight**) International Conference on Learning Representations (ICLR) 2020. (* = equal contribution)

- Extending GAN into a more general form by treating realness as a random variable.
- Training GAN with the loss of maximizing KL divergence.
- Training DCGAN at 1024^2 in one run, without progressive tricks.

Recursive Visual Sound Separation Using Minus-Plus Net

Xudong Xu, **Bo Dai**, Dahua Lin.

IEEE International Conference on Computer Vision (ICCV) 2019.

- A framework for visual-guided sound separation.
- Recursively separating out the most salient sound and removing it from the mixture.
- Preventing less salient sounds sound like noises in the context of salient sounds.

Feature Intertwiner for Object Detection

Hongyang Li, **Bo Dai**, Shaoshuai Shi, Wanli Ouyang, Xiaogang Wang.

International Conference on Learning Representations (ICLR) 2019.

- A component for object detection and more.
- Relying on the consistency between two sets of features, separated by scales, difficulty, etc.
- Guiding features in the less robust set with features in the other.

A Neural Compositional Paradigm for Image Captioning

Bo Dai, Sanja Fidler, and Dahua Lin.

Advances in Neural Information Processing Systems (NIPS) 2018.

- A nonsequential paradigm that generates captions hierarchically.
- Recursively composing incomplete phrases into longer phrases, fitting the properties of natural language.
- Strong ability in cross dataset generalization.

Rethinking the Form of Latent States in Image Captioning

Bo Dai*, Deming Ye*, and Dahua Lin.

European Conference on Computer Vision (ECCV) 2018. (* = equal contribution)

- Rethinking the representations of latent states in captioning models.
- Treating latent states as 2D maps, instead of 1D vectors.
- Visually revealing the dynamics of caption generation, and connections between visual and linguistic domain.

Move Forward and Tell: A Progressive Generator of Video Descriptions

Yilei Xiong, **Bo Dai**, and Dahua Lin.

European Conference on Computer Vision (ECCV) 2018.

- A framework that generates a descriptive paragraph for a video.
- Remembering involved events and generated sentences so that the paragraph is coherent and concise.

Neural Network Encapsulation

Hongyang Li, Xiaoyang Guo, **Bo Dai**, Wanli Ouyang, and Xiaogang Wang.

European Conference on Computer Vision (ECCV) 2018.

- A network structure and a loss that approximate the routing mechanism in Capsule Network.
- Scaling capsule-like networks up to large datasets.
- Adapting optimal transport as an objective.

Contrastive Learning for Image Captioning

Bo Dai, and Dahua Lin.

Advances in Neural Information Processing Systems (NIPS) 2017.

- A learning method for image captioning.
- Encouraging distinctiveness in generated captions.
- Introducing a reference model, inspired by Noise Contrastive Estimation.

Towards Diverse and Natural Image Descriptions via a Conditional GAN

Bo Dai, Sanja Fidler, Raquel Urtasun and Dahua Lin.

(**Oral**) IEEE International Conference on Computer Vision (ICCV) 2017.

- A learning method for image captioning that jointly learns a caption generator and a caption evaluator.
- Assessing captions in terms of semantical relatedness, rather than n-gram matching.
- Generating captions that are more natural and vivid.

Detecting Visual Relationships with Deep Relational Networks

Bo Dai, Yuqi Zhang, and Dahua Lin.

(**Oral**) IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

- A visual relationship detector based on visual appearances, spatial configurations and statistical relations.
- Core component, DR-Net, is a statistical net that captures statistical relations among variables.

- Serving as a component for structurally representing images with scene graphs.

ACADEMIC SERVICES

I reviewed papers for the following journals:

- IET Computer Vision
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- IEEE Transactions on Cognitive and Developmental Systems (TCDS)
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

I reviewed papers for the following conferences:

- AAAI Conference on Artificial Intelligence (AAAI)
- International Conference on Computer Vision (ICCV)
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- The North American Chapter of the Association for Computational Linguistics (NAACL)
- European Conference on Computer Vision (ECCV)
- Asian Conference on Computer Vision (ACCV)
- ACM Multimedia Conference (ACM MM)
- International Conference on Learning Representations (ICLR)

TEACHING EXPERIENCE

The Chinese University of Hong Kong

Teaching Assistant

- Image and Video Processing Spring 2018
- Problem Solving by Programming Spring 2018
- Probabilistic Modeling and Inference Spring 2017
- Information & Software Engineering Practice Fall 2016
- Problem Solving by Programming (Best TA) Spring 2016
- Linear Algebra and Vector Calculus for Engineers (Best TA) Fall 2015
- Information & Software Engineering Practice Spring 2015

Shanghai Jiao Tong University

Teaching Assistant

- Project Workshop of Database Fall 2014
- Project Workshop of Compiler Principles Fall 2013
- Programming Fall 2011

HONORS & AWARDS

- Star of Tomorrow, Microsoft Research Asia Jan 2014